VOLUMES ONE TO TWELVE

INTERNATIONAL ENCYCLOPEDIA OF HUMAN GEOGRAPHY

EDITORS-IN-CHIEF ROB KITCHIN AND NIGEL THRIFT

Senior Editors

Noel Castree
Mike Crang
Mona Domoah

Section Editors

Kay Anderson
Paul Cloke
Jeremy Crampton
Brian Graham
Costis Hadjimichalis
Phil Hubbard
Robin Kearns
Mei-Po Kwan
Loretta Lees
Sara McLafferty
Anssi Paasi
Chris Philo
James Sidaway
Katie Willis
Henry Yeung

INTERNATIONAL ENCYCLOPEDIA OF HUMAN GEOGRAPHY

EDITORS-IN-CHIEF

ROB KITCHIN
National Institute of Regional and Spatial Analysis (NIRSA)
John Hume Building, National University of Ireland
Maynooth, County Kildare
Ireland

NIGEL THRIFT
University House, University of Warwick
Coventry
UK

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

International Encyclopedia of HUMAN GEOGRAPHY

INTERNATIONAL ENCYCLOPEDIA OF HUMAN GEOGRAPHY

EDITORS-IN-CHIEF

ROB KITCHIN
National Institute of Regional and Spatial Analysis (NIRSA)
John Hume Building, National University of Ireland
Maynooth, County Kildare
Ireland

NIGEL THRIFT
University House, University of Warwick
Coventry
UK

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands
Linacre House, Jordan Hill, Oxford OX2 8DP, UK

First edition 2009

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Catalog Number: 2007927349

ISBN: 978-0-08-044911-1 (set)

Printed and bound in Italy

09 10 11 12 11 10 9 8 7 6 5 4 3 2 1

SENIOR EDITORS

Noel Castree
Geography Discipline
Arthur Lewis Building
School of Environment and Development
The University of Manchester
Manchester
UK

Mike Crang
Department of Geography
Durham University
Science Laboratories
Durham
UK

Mona Domosh
Department of Geography
Dartmouth College
Hanover, NH
USA

SECTION EDITORS

Kay Anderson
Centre for Cultural Research
University of Western Sydney
Penrith South DC, NSW, Sydney
Australia

Paul Cloke
Department of Geography
University of Exeter
Exeter
UK

Jeremy Crampton
Department of Geosciences
Georgia State University
Atlanta, GA
USA

Brian Graham
University of Ulster
Coleraine, N. Ireland
UK

Costis Hadjimichalis
Department of Geography
Harokopio University
Athens
Greece

Phil Hubbard
Department of Geography
University of Loughborough
Loughborough
UK

Robin Kearns
School of Geography, Geology and Environmental Sciences
The University of Auckland
Auckland
New Zealand

Mei-Po Kwan
Department of Geography
The Ohio State University
Columbus, OH
USA

Loretta Lees
Cities Research Group
Department of Geography
King's College London
London
UK

Sara McLafferty
Department of Geography
University of Illinois at Urbana – Champaign
Urbana, IL
USA

Anssi Paasi
Department of Geography
University of Oulu
Linnanmaa
Oulu
Finland

Chris Philo
Department of Geographical and Earth Sciences
University of Glasgow
Glasgow
UK

James Sidaway
School of Geography
University of Plymouth
Plymouth
UK

Katie Willis
Department of Geography
Royal Holloway, University of London
Egham
UK

Henry Yeung
Department of Geography
National University of Singapore
Singapore

GUIDE TO USE OF THE ENCYCLOPEDIA

Structure of the Encyclopedia

The material in the Encyclopedia is arranged as a series of articles in alphabetical order.

There are four features to help you easily find the topic you're interested in: an alphabetical contents list, a subject classification index, cross references and a full subject index.

1. Alphabetical Contents List

The alphabetical contents list, which appears at the front of the first volume, lists the entries in the order that they appear in the Encyclopedia. It includes both the volume number and the page number of each entry.

2. Subject Classification Index

This index appears at the start of Volume 1 and groups entries under subject headings that reflect the broad themes of Human Geography. This index is useful for making quick connections between entries and locating the relevant entry for a topic that is covered in more than one article.

3. Cross-references

All of the entries in the Encyclopedia have been exten sively cross referenced. The cross references which ap pear at the end of an entry, serve three different functions:

i. To indicate if a topic is discussed in greater detail elsewhere
ii. To draw the readers attention to parallel discussions in other entries
iii. To indicate material that broadens the discussion

Example

The following list of cross references appears at the end of the entry IMPERIALISM, CULTURAL

See also: Colonialism I; Colonialism II; Dependency; First World; Globalization, Cultural; Hegemony; Imperialistic Geographies; Neocolonialism; Orientalism; Postcolonia lism/Postcolonial Geographies; Poststructuralism/Post structuralist Geographies; Third World.

Here you will find examples of all three functions of the cross reference list: a topic discussed in greater detail elsewhere (e.g., Orientalism), parallel discussion in other entries (Imperialistic Geographies) and reference to en tries that broaden the discussion (e.g. Postcolonialism/ Postcolonial Geographies).

4. Index

The index includes page numbers for quick reference to the information you're looking for. The index entries differentiate between references to a whole entry, a part of an entry, and a table or figure.

5. Contributors

At the start of each volume there is list of the authors who contributed to that volume.

CONTRIBUTORS

VOLUME 1

M B Aalbers
Columbia University, New York, NY, USA

G J Andrews
McMaster University, Hamilton, Canada

H Asche
University of Potsdam, Potsdam, Germany

J O Bærenholdt
Roskilde University, Roskilde, Denmark

J Battersby-Lennard
University of Cape Town, Cape Town, South Africa

J Beall
London School of Economics and Political Science, London, UK

K B Beesley
Brandon University, Brandon, MB, Canada

K Besio
University of Hawaii at Hilo, Hilo, HI, USA

M Brayshay
University of Plymouth, Devon, UK

M M Breitbart
Hampshire College, Amherst, MA, USA

S Brentjes
Universidad de Sevilla, Sevilla, Spain

John R Bryson
The University of Birmingham, Birmingham, UK

H Buller
University of Exeter, Exeter, UK

K A Butler
University of Akron, Akron, OH, USA

K W Butzer
University of Texas, Austin, TX, USA

S Chaturvedi
Panjab University, Chandigarh, India

J Crush
University of Cape Town, Rondebosch, South Africa

C D'Ignazio
The Institute for Infinitely Small Things, Waltham, MA, USA

R J Das
York University, Toronto, ON, Canada

J Davidson
Queen's University, Kingston, ON, Canada

C Delano-Smith
University of London, London, UK

V Desai
Royal Holloway, University of London, Egham, UK

J D Dewsbury
Bristol University, Bristol, UK

J Dittmer
University college London, London, UK

H Fassmann
Universität Wien, Vienna, Austria

C Gibson
University of Wollongong, Wollongong, NSW, Australia

J R Gold
Oxford Brookes University, Oxford, UK

O Groza
Universitatea Al.I Cuza, Iasi, Romania

S W Hardwick
University of Oregon, Eugene, OR, USA

M Harrower
University of Wisconsin – Madison, Madison, WI, USA

H Hazen
Macalester College, St Paul, MN, USA

C Hughes
Carleton University, Ottawa, ON, Canada

G T Jóhannesson
University of Iceland, Reykjavik, Iceland

L Johnston
University of Waikato, Hamilton, New Zealand

R J P Kain
University of Exeter, Exeter, UK

R B Kent
California State University, Northridge, Northridge, CA, USA

M Kurtz
Open University, Ottawa, ON, Canada

M Kurtz
Open University, Milton Keynes, UK

R W Lake
Rutgers University, New Brunswick, NJ, USA

R Le Heron
The University of Auckland, Auckland, New Zealand

A A Lehtinen
University of Joensuu, Joensuu, Finland

Y Leung
The Chinese University of Hong Kong, Hong Kong, People's Republic of China

S Lloyd-Evans
University of Reading, Reading, UK

J Lossau
Humboldt-Universität zu Berlin, Berlin, Germany

A Malmberg
Uppsala University, Uppsala, Sweden

T Marsden
Cardiff University, Cardiff, UK

F Martinelli
Università 'Mediterranea' di Reggio Calabria, Reggio di Calabria, Italy

L Maxey
Swansea University, Wales, UK

D P McCormack
Oxford University Centre for the Environment, Oxford, UK

P Merlin

G A Myers
University of Kansas, Lawrence, KS, USA

R P Neumann
Florida International University, Miami, FL, USA

T Oakes
University of Colorado at Boulder, Boulder, CO, USA

U Oslender
University of Glasgow, Glasgow, UK

M Pacione
University of Strathclyde, Glasgow, UK

C Potter
Imperial College London, London, UK

R B Potter
University of Reading, Reading, UK

G Prevelakis
University of Pantheon-Sorbonne (Paris 1), Paris, France

M Purcell
University of Washington, Seattle, WA, USA

D Ramsey
Brandon University, Brandon, MB, Canada

S Reimer
University of Southampton, Southampton, UK

J Renes
Utrecht University, Utrecht, The Netherlands, and Vrije Universiteit, Amsterdam, The Netherlands

V Rey
Ecole Normale Supérieure Lettres et Sciences Humaines, Lyon, France

M-C Robic
CNRS, Paris, France

U Rossi
University of Cagliari, Cagliari, Italy

P Routledge
University of Glasgow, Glasgow, UK

J Siemer
University of Regina, Regina, SK, Canada

M W Skinner
Trent University, Peterborough, ON, Canada

T Slater
University of Edinburgh, Bristol, UK

L M Takahashi
University of California, Los Angeles, CA, USA

A Taylor
University of Canberra, Canberra, ACT, Australia

M E Thomas
The Ohio State University, Columbus, OH, USA

T M Vowles
University of Northern Colorado, Greeley, CO, USA

A M Warnes
University of Sheffield, Sheffield, UK

D Wastl-Walter
University of Bern, Bern, Switzerland

A Watson
University of Minnesota, Minneapolis, MN, USA

C Wilbert
Anglia Ruskin University, Chelmsford, UK

M Williams
Oxford University, Oxford, UK

S D Withers
University of Washington, Seattle, WA, USA

N M Yantzi
Laurentian University, Sudbury, ON, Canada

VOLUME 2

H Aay
Calvin College, Grand Rapids, MI, USA

S C Aitken
San Diego State University, San Diego, CA, USA and National Technical Norwegian University, Trondheim, Norway

L Albrechts
Catholic University of Leuven, Leuven, Belgium

G J Andrews
McMaster University, Hamilton, ON, Canada

N Ansell
Brunel University, Uxbridge, UK

D Atkinson
University of Hull, Hull, UK

M Batty
University College London, London, UK

J Baxter
University of Western Ontario, London, ON, Canada

M Bédard
Université du Québec à Montréal, Montréal, QC, Canada

D Bell
University of Leeds, Leeds, UK

D Bennett
Carleton University, Ottawa, ON, Canada

U Best
Technische Universität Chemnitz, Chemnitz, Germany

J Binnie
Manchester Metropolitan University, Manchester, UK

C R Bryant
Université de Montréal, Montréal, QC, Canada

R Capello
Milan Polytechnic, Milan, Italy

R Carvalho
Queen Mary, University of London, London, UK

G Chapman
Lancaster University, Lancaster, UK

L Cheshire
The University of Queensland, Brisbane, QLD, Australia

S Chilvers
York University, Toronto, ON, Canada

V Chouinard
McMaster University, Hamilton, ON, Canada

H Couclelis
University of California, Santa Barbara, CA, USA

A Currah
Oxford University Centre for the Environment, Oxford, UK

R J Das
York University, Toronto, ON, Canada

S Davoudi
Newcastle University, Newcastle upon Tyne, UK

E Delaney
US Deparment of the Interior, National Park Service, Washington, DC, USA

J Dempsey
University of British Columbia, Vancouver, BC, Canada

S J Elliott
McMaster University, Hamilton, ON, Canada

V Filippakopoulou
National Technical University of Athens, Athens, Greece

W Friesen
The University of Auckland, Auckland, New Zealand

C Gibson
University of Wollongong, Wollongong, NSW, Australia

A L Griffin
University of New South Wales – ADFA, Canberra, ACT, Australia

J W Harrington, Jr.
University of Washington, Seattle, WA, USA

M Hassler
Philipps-University Marburg, Marburg, Germany

G Helms
University of Glasgow, Glasgow, UK

J J Holloway
Manchester Metropolitan University, Manchester, UK

A Hughes
Newcastle University, Newcastle upon Tyne, UK

M Jay
University of Waikato, Hamilton, New Zealand

N Kliot
University of Haifa, Haifa, Israel

H Koskela
University of Helsinki, Helsinki, Finland

J Krygier
Ohio Wesleyan University, Delaware, OH, USA

M S Kumar
Queen's University Belfast, Belfast, UK

B Lambregts
University of Amsterdam, Amsterdam, The Netherlands

D Leslie
University of Toronto, Toronto, ON, Canada

A Lester
University of Sussex, Brighton, UK

N Lewis
The University of Auckland, Auckland, New Zealand

K D Lilley
Queen's University Belfast, Belfast, UK

T W Luke
Virginia Polytechnic Institute and State University, USA

L J C Ma
University of Akron, Akron, OH, USA

J Malczewski
University of Western Ontario, London, ON, Canada

D Martin
University of Southampton, Southampton, UK

E J McCann
Simon Fraser University, Burnaby, BC, Canada

C McIlwaine
Queen Mary, University of London, London, UK

K J Mee
University of Newcastle, Callaghan, NSW, Australia

S Millington
Manchester Metropolitan University, Manchester, UK

C J A Mitchell
University of Waterloo, Waterloo, ON, Canada

D R Montello
University of California, Santa Barbara, CA, USA

M Morad
London South Bank University, London, UK

B Nakos
National Technical University of Athens, Athens, Greece

D O'Sullivan
The University of Auckland, Auckland, New Zealand

J R Oppong
University of North Texas, Denton, TX, USA

M Phillips
University of Leicester, Leicester, UK

C Philo
University of Glasgow, Glasgow, UK

A C Pratt
London School of Economics, London, UK

S Punch
University of Stirling, Stirling, UK

B Ramírez
Universidad Autónoma Metropolitana, Mexico City, Mexico

G A Rice
University of Helsinki, Helsinki, Finland

K Robinson
McMaster University, Hamilton, ON, Canada

R Rundstrom
University of Oklahoma, Norman, OK, USA

S Rycroft
University of Sussex, Brighton, UK

N Schuurman
Simon Fraser University, Burnaby, BC, Canada

J P Sharp
University of Glasgow, Glasgow, UK

D Sibley
University of Leeds, Leeds, UK

T Skelton
National University of Singapore, Singapore

F Smith
Brunel University, Uxbridge, UK

J Sundberg
University of British Columbia, Vancouver, BC, Canada

J C Sweeney
National University of Ireland, Maynooth, Republic of Ireland

W S Tang
Hong Kong Baptist University, Kowloon Tong, Kowloon, Hong Kong, People's Republic of China

T Tasan-Kok
Delft University of Technology, Jaffalaan, The Netherlands

J-C Thill
University of North Carolina at Charlotte, Charlotte, NC, USA

J Timár
Center for Regional Studies, Hungarian Academy of Sciences, Békéscsaba, Hungary

P M Torrens
Arizona State University, Tempe, AZ, USA

G Waitt
University of Wollongong, Wollongong, NSW, Australia

J L Waters
University of Liverpool, Liverpool, UK

D Wood
Raleigh, NC, USA

M Woods
Aberystwyth University, Aberystwyth, UK

F Wray
Newcastle University, Newcastle upon Tyne, UK

H van der Wusten
Universiteit van Amsterdam, Amsterdam, The Netherlands

C Young
Manchester Metropolitan University, Manchester, UK

VOLUME 3

T J Barnes
University of British Columbia, Vancouver, BC, Canada

C Barnett
The Open University, Milton Keynes, UK

J Barnett
University of Melbourne, Melbourne, VIC, Australia

C J Barrow
Swansea University, Swansea, UK

G de Bastion
Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ) GmbH, Eschborn, Germany

J V Beaverstock
University of Nottingham, Nottingham, UK

S Bell
University of Saskatchewan, Saskatoon, SK, Canada

L D Berg
University of British Columbia, Kelowna, BC, Canada

L Bondi
University of Edinburgh, Edinburgh, UK

A Bonnett
Newcastle University, Newcastle upon Tyne, UK

L A Brown
Ohio State University, Columbus, OH, USA

R A Butlin
University of Leeds, Leeds, UK

A Cameron
University of Leicester, Leicester, UK

C Cocklin
James Cook University, Townsville, QLD, Australia

M P Conzen
University of Chicago, Chicago, IL, USA

K E Corey
Michigan State University, East Lansing, MI, USA

T Cresswell
Royal Holloway, University of London, Egham, UK

C D'Alessandro-Scarpari
Université Lumière Lyon 2, Lyon, France

S Dall'erba
University of Arizona, Tucson, AZ, USA

J Davidson
Queen's University, Kingston, ON, Canada

A Davies
Trinity College, Dublin, Republic of Ireland

C Dixon
London Metropolitan University, London, UK

R Dowling
Macquarie University, Sydney, NSW, Australia

S Eden
University of Hull, Hull, UK

S J Elliott
McMaster University, Hamilton, ON, Canada

G A Elmes
West Virginia University, Morgantown, WV, USA

H Ernste
Radboud University, Nijmegan, The Netherlands

J Evans
McMaster University, Hamilton, ON, Canada

J R Faulconbridge
Lancaster University, Lancaster, UK

D A Fennell
Brock University, St Catharines, ON, Canada

S Gopal
Boston University, Boston, MA, USA

S Hall
University of Nottingham, Nottingham, UK

E Hartwick
Framingham State College, Framingham, MA, USA

S Healy
Worcester State College, Worcester, MA, USA

M Hess
University of Manchester, Manchester, UK

J Holloway
Manchester Metropolitan University, Manchester, UK

T Jazeel
University of Sheffield, Sheffield, UK

R Johnston
University of Bristol, Bristol, UK

O Jones
CCRI, Cheltenham, UK

A Kalogeressis
University of Macedonia, Greece

A Kavanagh
National University of Ireland, Maynooth, Republic of Ireland

M C Keane
St Mary's University College, Belfast, UK

M Keating
European University Institute, Florence, Italy

J Kneale
University College London, London, UK

A Kobayashi
Queen's University, Kingston, ON, Canada

T Koch
University of British Columbia, Vancouver, BC, Canada

B Korf
University of Zürich – Irchel, Zürich, Switzerland

M S Kumar
Queen's University Belfast, Belfast, UK

M Kuus
University of British Columbia, Vancouver, BC, Canada

L Labrianidis
University of Macedonia, Greece

A Latham
University College London, London, UK

E Laurier
University of Edinburgh, Edinburgh, UK

A A Lehtinen
University of Joensuu, Joensuu, Finland

J I Leib
Old Dominion University, Norfolk, VA, USA

N Lewis
University of Auckland, Auckland, New Zealand

W Li
Arizona State University, Tempe, AZ, USA

D Linehan
University College Cork, Cork, Republic of Ireland

L Lo
York University, Toronto, ON Canada

R Longhurst
University of Waikato, Hamilton, New Zealand

M Low
London School of Economics and Political Science, London, UK

K McCracken
Macquarie University, Sydney, NSW, Australia

K McKittrick
Queen's University, Kingston, ON, Canada

P McManus
The University of Sydney, Sydney, NSW, Australia

P Meth
University of Sheffield, Sheffield, UK

J J Metz
Northern Kentucky University, Highland Heights, KY, USA

B E Montz
Binghamton University, Binghamton, NY, USA

D P Nally
University of Cambridge, Cambridge, UK

P M O'Neill
University of Western Sydney, Parramatta, NSW, Australia

C Pattie
University of Sheffield, Sheffield, UK

N A Phelps
University College London, London, UK

D R Phillips
Lingnan University, Tuen Mun, Hong Kong

C Philo
University of Glasgow, Glasgow, UK

A Pike
Newcastle University, Newcastle upon Tyne, UK

G H Pirie
University of the Western Cape, Cape Town, South Africa

P Plummer
University of Calgary, Calgary, AB, Canada

M Power
University of Durham, Durham, UK

P Rees
University of Leeds, Leeds, UK

S A Royle
Queen's University Belfast, Belfast, UK

A Schærström
Kometvägen, Täby, Sweden

R Scheyvens
Massey University, Palmerston North, New Zealand

J W Scott
University of Joensuu, Joensuu, Finland

D Seamon
Kansas State University, Manhattan, KS, USA

E Sheppard
University of Minnesota, Minneapolis, MN, USA

David Simon
University of London, Egham, UK

V F S Sit
University of Hong Kong, Hong Kong, People's Republic of China

E Skop
The University of Colorado at Colorado Springs, Colorado Springs, CO, USA

L Smith
Radboud University, Nijmegen, The Netherlands

M Smith
Queen's University, Kingston, ON, Canada

C Sneddon
Dartmouth College, Hanover NH, USA

H Southall
University of Portsmouth, Portsmouth, UK

J Sowers
Missouri State University, Springfield, MO, USA

E Stam
Utrecht University, Utrecht, The Netherlands

D Storey
University of Worcester, Worcester, UK

D Sui
Texas A & M University, College Station, TX, USA

J Sundberg
University of British Columbia, Vancouver, BC, Canada

E Swyngedouw
Manchester University, Manchester, UK

K E Till
Virginia Tech University, Alexandria, VA, USA

G A Tobin
University of South Florida, Tampa, FL, USA

S W Trimble
University of California, Los Angeles, CA, USA

S Turner
McGill University, Montréal, QC, Canada

M Tykkyläinen
University of Joensuu, Joensuu, Finland

T Unwin
University of London, Egham, UK

E Warwick
King's College London, London, UK

H D Watts
University of Sheffield, Sheffield, UK

M Watts
UC Berkeley, Berkeley, CA, USA

M Webber
University of Melbourne, Melbourne, VIC, Australia

G R Webster
University of Wyoming, Laramie, WY, USA

D Weiner
West Virginia University, Morgantown, WV, USA

J Wiles
University of Auckland, Auckland, New Zealand

K D Willis
Royal Holloway, University of London, Egham, UK

M I Wilson
Michigan State University, East Lansing, MI, USA

R Wilton
McMaster University, Hamilton, ON, Canada

H Winlow
Bath Spa University, Bath, UK

C W J Withers
University of Edinburgh, Edinburgh, UK

F Wu
Cardiff University, Cardiff, UK

N Xiao
The Ohio State University, Columbus, OH, USA

I Yamada
University of Utah, Salt Lake City, UT, USA

B S A Yeoh
National University of Singapore, Singapore

O Yiftachel
Ben Gurion University, Beer-sheva, Israel

E K Z Zygmunt
University of Silesia, Sosnowiec, Poland

VOLUME 4

D Abbott
University of Derby, Derby, UK

P Adey
Keele University, Keele, UK

P J Atkins
Durham University, Durham, UK

R Atkinson
University of Tasmania, Hobart, TAS, Australia

G Benko
Université Panthéon-Sorbonne, Paris, France

P Bennett
University of Edinburgh, Edinburgh, UK

L D Berg
University of British Columbia, Kelowna, BC, Canada

M Birkin
University of Leeds, Leeds, UK

S Blandy
University of Leeds, Leeds, UK

F J Bosco
San Diego State University, San Diego, CA, USA

S Bowlby
University of Reading, Reading, UK

H Bradley
Mary Immaculate College, University of Limerick, Limerick, Republic of Ireland

P Breathnach
National University of Ireland, Maynooth, Republic of Ireland

J Briggs
University of Glasgow, Glasgow, UK

J E Castro
Newcastle University, Newcastle upon Tyne, UK

J Chapman
University of Portsmouth, Portsmouth, UK

O Z Chaudhry
University of Edinburgh, Edinburgh, UK

G L Clark
Oxford University Centre for the Environment, Oxford, UK

G P Clarke
University of Leeds, Leeds, UK

K C Clarke
University of California, Santa Barbara, CA, USA

P Claval
Université de Paris-Sorbonne, Paris, France

N M Coe
University of Manchester, Manchester, UK

M Cope
University of Vermont, Burlington, VT, USA

David Crouch
University of Derby, Derby, UK

Ú Crowley
National University of Ireland, Maynooth, Co Kildare, Republic of Ireland

C Desbiens
Université Laval, Québec City, QC, Canada

P Dicken
University of Manchester, Manchester, UK

G Dijkink
University of Amsterdam, Amsterdam, The Netherlands

I Docherty
University of Glasgow, Glasgow, UK

M Duckham
University of Melbourne, Melbourne, VIC, Australia

M Duffy
University of Melbourne, Melbourne, VIC, Australia

S Elwood
University of Washington, Seattle, WA, USA

M F Goodchild
University of California, Santa Barbara, CA, USA

D Goodman
University of California – Santa Cruz, Santa Cruz, CA, USA

M K Goodman
King's College London, London, UK

M Goodwin
University of Exeter, Exeter, UK

L van Grunsven
Utrecht University, Utrecht, The Netherlands

J Häkli
University of Tampere, Finland

E Hall
University of Dundee, Dundee, UK

D J Hammel
University of Toledo, Toledo, OH, USA

B Hendrikx
Radboud University, Nijmegen, The Netherlands

M Hess
University of Manchester, Manchester, UK

J Hessler
Library of Congress, Washington, DC, USA

R Hudson
Durham University, Durham, UK

M Huxley
University of Sheffield, Sheffield, UK

L C Johnson
Deakin University, Geelong, VIC, Australia

R Kaiser
University of Wisconsin – Madison, Madison, WI, USA

A Kealy
University of Melbourne, Melbourne, VIC, Australia

M J Kraak
International Institute of Geo-Information Science and Earth Observation, Enschede, The Netherlands

A Lagendijk
Radboud University, Nijmegen, The Netherlands

N S Lam
Louisiana State University, Baton Rouge, LA, USA

A Latham
UCL, London, UK

J Lee
The University of Seoul, Seoul, South Korea

Y Leung
The Chinese University of Hong Kong, Hong Kong

A Leyshon
University of Nottingham, Nottingham, UK

J Little
University of Exeter, Exeter, UK

J Lovering
Cardiff University, Cardiff, UK

C Lukinbeal
Arizona State University, Tempe, AZ, USA

W A Mackaness
University of Edinburgh, Edinburgh, UK

A Maddrell
University of the West of England, Bristol, UK

J C Marshall[†]
University of Oxford, Oxford, UK

D R Meyer
Brown University, Providence, RI, USA

V Miller
University of Kent, Canterbury, UK

C M Moreno
San Diego State University, San Diego, CA, USA

K M Morin
Bucknell University, Lewisburg, PA, USA

R Morrill
University of Washington, Seattle, WA, USA

L Muscarà
Università del Molise, Isernia, Campobasso, Italy

T Nyerges
University of Washington, Seattle, WA, USA

S M Otterstrom
Brigham Young University, Provo, UT, USA

A Páez
McMaster University, Hamilton, ON, Canada

M Pavlovskaya
Hunter College – CUNY, New York, NY, USA

Linda Peake
York University, Toronto, ON, Canada

M Phillips
University of Leicester, Leicester, UK

J Pollard
Newcastle University, Newcastle upon Tyne, UK

L Price
Queen's University Belfast, Belfast, UK

B Pritchard
University of Sydney, Sydney, NSW, Australia

M Purvis
University of Leeds, Leeds, UK

M Raco
King's College London, London, UK

K N Rankin
University of Toronto, Toronto, ON, Canada

J Raper
City University, London, UK

L T Raynolds
Colorado State University, Fort Collins, CO, USA

P Reuber
University of Münster, Münster, Germany

P Richardson
University of Oxford, Oxford, UK

S M Roberts
University of Kentucky, Lexington, KY, USA

U Rossi
University of Cagliari, Cagliari, Italy

G Rushton
The University of Iowa, Iowa City, IA, USA

R B Sarma
Cotton College, Guwahati, India

A J Secor
University of Kentucky, Lexington, KY, USA

J Shaw
University of Plymouth, Plymouth, UK

T Slater
University of Edinburgh, Edinburgh, UK

K Strauss
Oxford University Centre for the Environment, Oxford, UK

[†]Deceased, 22 August 2007.

M Taylor
Mary Immaculate College, University of Limerick, Limerick, Republic of Ireland

M Taylor
University of Birmingham, Birmingham, UK

D Thien
California State University, Long Beach, CA, USA

J Tivers
Nottingham Trent University, Nottingham, UK

Y Underhill-Sem
University of Auckland, Auckland, New Zealand

J Wallace
University of Melbourne, Melbourne, VIC, Australia

F Wang
Louisiana State University, Baton Rouge, LA, USA

B Warf
University of Kansas, Lawrence, KS, USA

S J Whatmore
University of Oxford, Oxford, UK

D C Wheeler
Emory University, Atlanta, GA, USA

J Wiles
University of Auckland, Auckland, New Zealand

G Williams
University of Sheffield, Sheffield, UK

K D Willis
Royal Holloway, University of London, Egham, UK

D Wójcik
Oxford University Centre for the Environment, Oxford, UK

X Yao
University of Georgia, Atlanta, GA, USA

H W-c Yeung
National University of Singapore, Singapore

VOLUME 5

A G Aguilar
Universidad Nacional Autónoma de México, Mexico City, Mexico

J Agyeman
Tufts University, Medford, MA, USA

K Falconer Al-Hindi
University of Nebraska at Omaha, Omaha, NE, USA

N AlSayyad
University of California, Berkeley, CA, USA

G J Ashworth
University of Groningen, Groningen, The Netherlands

S Asthana
University of Plymouth, Plymouth, UK

P Barnett
University of Otago, Christchurch, New Zealand

R Barnett
University of Canterbury, Christchurch, New Zealand

D Bell
University of Leeds, Leeds, UK

J D Bohland
Hollins University, Roanoke, VA, USA

T Brown
Queen Mary, University of London, London, UK

I H Burnley
University of New South Wales, Sydney, NSW, Australia

E Cameron
Queen's University, Kingston, ON, Canada

E P Campuzano
Instituto Politécnico Nacional, Mexico City, Mexico

F Celata
University of Rome "La Sapienza", Rome, Italy

M Chapin
Center for the Support of Native Lands, Arlington, VA, USA

K Chapman
University of Aberdeen, Aberdeen, UK

S Chaturvedi
Panjab University, Chandigarh, India

E Clark
Lund University, Lund, Sweden

N Collins-Kreiner
University of Haifa, Haifa, Israel

J Corbett
University of British Columbia, Vancouver, BC, Canada

G DeVerteuil
University of Southampton, Southampton, UK

M Dodge
University of Manchester, Manchester, UK

D Dorling
University of Sheffield, Sheffield, UK

P J Duffy
National University of Ireland, Maynooth, Maynooth, Republic of Ireland

J Dunn
St Michael's Hospital, Toronto, ON, Canada

R Fish
University of Exeter, Exeter, UK

R Flowerdew
University of St Andrews, St Andrews, UK

R Foley
NUI Maynooth, Maynooth, Republic of Ireland

N Gallent
University College London, London, UK

C Gibson
University of Wollongong, Wollongong, NSW, Australia

L Gibson
Land and People Information Sharing Society (LAPIS), Parksville, BC, Canada

J Glassman
University of British Columbia, Vancouver, BC, Canada

A R Goetz
University of Denver, Denver, CO, USA

A Golan
University of Haifa, Haifa, Israel

W T S Gould
University of Liverpool, Liverpool, UK

M Greenwood
University of Northern British Columbia, Prince George, BC, Canada

A L Griffin
University of New South Wales – ADFA, Canberra, ACT, Australia

L Guelke
University of Waterloo, Waterloo, ON, Canada

E Hague
DePaul University, Chicago, IL, USA

R Haines-Young
University of Nottingham, Nottingham, UK

R Harris
McMaster University, Hamilton, ON, Canada

F Harvey
University of Minnesota, Minneapolis, MN, USA

G J D Hewings
University of Illinois, Urbana, IL, USA

R Honey
The University of Iowa, Iowa City, IA, USA

R Howitt
Macquarie University, Sydney, NSW, Australia

M Hoyler
Loughborough University, Loughborough, UK

R Hudson
University of Durham, Durham, UK

J S Humphreys
Monash University, Bendigo, VIC, Australia

R D Jacobson
University of Calgary, Calgary, AB, Canada

J P Jones III
University of Arizona, Tucson, AZ, USA

R Jones
Curtin University of Technology, Perth, WA, Australia

C Kelly
University of Greenwich, London, UK

A D King
State University of New York at Binghamton, Binghamton, NY, USA

S Kirsch
University of North Carolina, Chapel Hill, NC, USA

A Kobayashi
Queen's University, Kingston, ON, Canada

N Koch
Dartmouth College, Hanover, NH, USA

Z Kovács
University of Szeged, Szeged, Hungary

M-J Kraak
ITC – International Institute of Geo-Information Science and Earth Observation, Enschede, The Netherlands

S de Leeuw
University of Northern British Columbia, Prince George, BC, Canada

R Lewis
University of Toronto, Toronto, ON, Canada

R Longhurst
University of Waikato, Hamilton, New Zealand

D Mackinnon
University of Aberdeen, Aberdeen, UK

S P Mains
University of the West Indies-Mona, Kingston, Jamaica

J May
Queen Mary, University of London, London, UK

P Milbourne
Cardiff University, Cardiff, UK

C Minca
Royal Holloway, University of London, Egham, UK

J Mistry
University of London, Egham, UK

R Mitchell
University of Glasgow, Glasgow, UK

G Moon
University of Southampton, Highfield, UK

S Muller
Macquarie University, Sydney, NSW, Australia

S Neal
Open University, Milton Keynes, UK

M E O'Kelly
Ohio State University, Columbus, OH, USA

K R Olwig
Swedish University of Agricultural Sciences, Alnarp, Sweden

S Orford
Cardiff University, Cardiff, UK

M Pacione
University of Strathclyde, Glasgow, UK

M Parnwell
University of Leeds, Leeds, UK

B Parthasarathy
International Institute of Information Technology, Bangalore, India

J Patchell
Hong Kong University of Science and Technology, Kowloon, Hong Kong, China

R Peet
Clark University, Worcester, MA, USA

T Peil
Tallinn University, Tallinn, Estonia

N A Phelps
University College London, London, UK

J M Powell
Monash University, Melbourne, VIC, Australia

P Raento
University of Helsinki, Helsinki, Finland

G Rambaldi
Technical Centre for Agricultural and Rural Cooperation ACP-EU, Wageningen, The Netherlands

C Richmond
The University of Western Ontario, London, ON, Canada

E J Roe
Southampton University, Southampton, UK

U Rossi
University of Cagliari, Cagliari, Italy

J Rutherford
LATTS (Laboratoire Techniques, Territoires et Sociétés), Marne-la-Vallée, France

G Setten
Department of Geography, Norwegian University of Science and Technology, Trondheim, Norway

M Shaw
Bristol University, Bristol, UK

D Simandan
Brock University, St Catharines, ON, Canada

J M Smith
Texas A & M University, College Station, TX, USA

K B Smith
Monash University, Bendigo, VIC, Australia

M Sonis
University of Illinois, Urbana, IL, USA, and Bar Ilan University, Ramat Gan, Israel

M Sothern
University of St Andrews, St Andrews, UK

R Sternberg
University of Hannover, Hannover, Germany

S Suchet-Pearson
Macquarie University, Sydney, NSW, Australia

I Türeli
Brown University, Providence, RI, USA

J Timár
Center for Regional Studies, Békéscsaba, Hungary

D P Tolia-Kelly
Durham University, Durham, UK

H Tunstall
University of York, York, UK

D Vaiou
National Technical University of Athens, Athens, Greece

S M Walcott
University of North Carolina, Greensboro, NC, USA

J L Waters
University of Liverpool, Liverpool, UK

G W White
Frostburg State University, Frostburg, MD, USA

K Wilson
University of Toronto, Mississauga, ON, Canada

G M Winder
The University of Auckland, Auckland, New Zealand

G M Winder
Ludwig-Maximilians-Universität, Munich, Germany

K Woodward
University of Exeter, Exeter, UK

R Wright
Dartmouth College, Hanover, NH, USA

Y Zhou
Vassar College, New York, NY, USA

M Zook
University of Kentucky, Lexington, KY, USA

VOLUME 6

B T Asheim
Lund University, Lund, Sweden

A Aylett
University of British Columbia, Vancouver, BC, Canada

A Bain
York University, Toronto, ON, Canada

T J Barnes
University of British Columbia, Vancouver, BC, Canada

C Benner
University of California, Davis, Davis, CA, USA

M Brosseau
University of Ottawa, Ottawa, ON, Canada

K Browne
University of Brighton, Brighton, UK

R A Butlin
University of Leeds, Leeds, UK

C A Calder
The Ohio State University, Columbus, OH, USA

P Carmody
Trinity College, Dublin, Republic of Ireland

W Cartwright
RMIT University, Melbourne, VIC, Australia

P Collier
University of Portsmouth, Portsmouth, UK

S Conti
Università e Politecnico di Torino, Turin, Italy

P Cooke
Cardiff University, Cardiff, UK

K R Cox
The Ohio State University, Columbus, OH, USA

N Cressie
The Ohio State University, Columbus, OH, USA

A Cumbers
University of Glasgow, Glasgow, UK

D Delaney
Amherst College, Amherst, MA, USA

M Dodge
University of Manchester, Manchester, UK

J Dubow
University of Sheffield, Sheffield, UK

C Dwyer
University College London, London, UK

R Edsall
University of Minnesota, Minneapolis, MN, USA

F Eva
Universita' Ca' Foscari, Venezia, Treviso Campus, Italy

M Farish
University of Toronto, Toronto, ON, Canada

R Fincher
University of Melbourne, Melbourne, VIC, Australia

S M Freundschuh
National Science Foundation, Arlington, VA, USA

G Garofoli
Insubria University, Varese, Italy

J Gaspar
CEG, University of Lisbon, Lisbon, Portugal

P Giaccaria
Università e Politecnico di Torino, Turin, Italy

E Gilbert
University of Toronto, Toronto, ON, Canada

J R Gold
Oxford Brookes University, Oxford, UK

A E Green
University of Warwick, Coventry, UK

M Haldrup
Roskilde University, Roskilde, Denmark

F Harvey
University of Minnesota, Minneapolis, MN, USA

A Herod
University of Georgia, Athens, GA, USA

M Hesse
University of Luxembourg, Luxembourg

S Hoelscher
The University of Texas at Austin, Austin, TX, USA

E Holland
University of Colorado, Boulder, CO, USA

B Hooper
Radboud University, Nijmegen, The Netherlands

P Hopkins
Newcastle University, Newcastle, UK

M W Horner
The Florida State University, Tallahassee, FL, USA

B van Hoven
University of Groningen, Groningen, The Netherlands

R Howitt
Macquarie University, Sydney, NSW, Australia

B Hunt
Santa Barbara, CA, USA

J Jäger
University of Applied Sciences BFI, Vienna, Austria

P Jackson
University of Sheffield, Sheffield, UK

A E G Jonas
University of Hull, Hull, UK

A Jones
University of London, London, UK

R Jones
Aberystwyth University, Aberystwyth, UK

Cristóbal Kay
Institute of Social Studies, The Hague, The Netherlands

G Kearns
Virginia Polytechnic and State University, Blacksburg, VA, USA

G Kearns
Virginia Tech, Blacksburg, VA, USA

A Kerr
National University of Ireland, Maynooth, Republic of Ireland

J Krygier
Ohio Wesleyan University, Delaware, OH, USA

W Li
University of Connecticut, Storrs, CT, USA

E Mac Gillavry
Webmapper, Utrecht, The Netherlands

J Mansvelt
Massey University, Palmerston North, New Zealand

F Mizuoka
Hitotsubashi University, Tokyo, Japan

K M Morin
Bucknell University, Lewisburg, PA, USA

A T Murray
Arizona State University, Tempe, AZ, USA

C J Nash
University of Brighton, Brighton, UK

J O'Loughlin
University of Colorado, Boulder, CO, USA

F Owusu
Iowa State University, Ames, IA, USA

R B Parry
The University of Reading, Reading, UK

M W Pearce
Ohio University, Athens, OH, USA

C Perkins
University of Manchester, Manchester, UK

S Pinch
University of Southampton, Southampton, UK

S Ó Riain
National University of Ireland, Maynooth, Republic of Ireland

G A Rice
University of Helsinki, Finland

J-P Rodrigue
Hofstra University, Hempstead, NY, USA

T D Rutherford
Syracuse University, Syracuse, NY, USA

A Saito
National University of Singapore, Singapore

James D Sidaway
School of Geography, University of Plymouth, UK

T Skelton
National University of Singapore, Singapore

A Tay
University of Bristol, Bristol, UK

D P Tolia-Kelly
Durham University, Durham, UK

B L Turner II
Arizona State University, Tempe, AZ, USA

H Winlow
Bath Spa University, Bath, UK

S D Withers
University of Washington, Seattle, WA, USA

D Wood
Raleigh, NC, USA

P Wood
University College London, London, UK

C Zhang
University of Connecticut, Storrs, CT, USA

M Zook
University of Kentucky, Lexington, KY, USA

VOLUME 7

S C Aitken
San Diego State University, San Diego, CA, USA, and National Technical University of Norway, Trondheim, Norway

D H Alderman
East Carolina University, Greenville, NC, USA

N Argent
University of New England, Armidale, NSW, Australia

S Bell
University of Saskatchewan, Saskatoon, SK, Canada

M E Bonine
University of Arizona, Tucson, AZ, USA

W M Bowen
Cleveland State University, Cleveland, OH, USA

P Boyle
University of St Andrews, St Andrews, UK

G Bridge
University of Manchester, Manchester, UK

K M Brown
Macaulay Land Use Research Institute, Craigiebuckler, Aberdeen, UK

G Burgel
University of Paris 10, Nanterre, France

M A Busteed
University of Manchester, Manchester, UK

L Cadman
University of Glasgow, Glasgow, UK

I Casas
University at Buffalo – SUNY, Buffalo, NY, USA

D B Clarke
Swansea University, Swansea, UK

N Clarke
University of Southampton, Southampton, UK

P Claval
Université de Paris-Sorbonne, Paris, France

J Clayton
University of Sunderland, Sunderland, UK

P Cooke
Cardiff University, Cardiff, UK

N Curry
Countryside and Community Research Institute CCRI, Cheltenham, UK

K M Dunn
School of Social Sciences, Penrith, NSW, Australia

O J Dwyer
Indiana University, Indianapolis, IN, USA

R Earickson
University of Maryland, Baltimore, MD, USA

T Edensor
Manchester Metropolitan University, Manchester, UK

C Cindy Fan
University of California, Los Angeles, CA, USA

M Farish
University of Toronto, Toronto, ON, Canada

J Flint
Sheffield Hallam University, Sheffield, UK

S Fullagar
Griffith University, Nathan, QLD, Australia

J R Gold
Oxford Brookes University, Oxford, UK

G Grabher
University of Bonn, Bonn, Germany

R N Gwynne
University of Birmingham, Birmingham, UK

D Hill
University of Otago, Dunedin, New Zealand

H Hoernig
Université de Québec à Montréal, Montréal, QC, Canada

P Holland
University of Otago, Dunedin, New Zealand

P Hubbard
Leicestershire University, Leicestershire, UK

R D Jacobson
University of Calgary, Calgary, AB, Canada

W Jenkins
York University, Toronto, ON, Canada

O Jones
Countryside and Community Research Institute CCRI, Cheltenham, UK

R Jones
Aberystwyth University, Aberystwyth, UK

K Kafkoula
Aristotle University of Thessaloniki, Thessaloniki, Greece

D H Kaplan
Kent State University, Kent, OH, USA

J Kenny
University of Wisconsin – Milwaukee, Milwaukee, WI, USA

M J Kuby
Arizona State University, Tempe, AZ, USA

W Larner
University of Bristol, Bristol, UK

R Le Heron
The University of Auckland, Auckland, New Zealand

Sang-Il Lee
Seoul National University, Seoul, South Korea

K D Lilley
Queen's University Belfast, Belfast, UK

D Linehan
University College Cork, Cork, Republic of Ireland

J Lorimer
King's College London, London, UK

G MacLeod
University of Durham, Durham, UK

E J McCann
Simon Fraser University, Burnaby, BC, Canada

L McDowell
University of Oxford, Oxford, UK

S McDowell
University of Ulster, Coleraine, UK

A McGregor
University of Otago, Dunedin, New Zealand

J H McKendrick
Glasgow Caledonian University, Glasgow, UK

P Merriman
Aberystwyth University, Aberystwyth, UK

W E Murray
Victoria University of Wellington, Wellington, New Zealand

G Olsson
Uppsala University, Uppsala, Sweden

K R Olwig
Swedish University of Agricultural Sciences, Alnarp, Sweden

J Penrose
University of Edinburgh, Edinburgh, UK

C G Pooley
Lancaster University, Lancaster, UK

C Radel
Utah State University, Logan, UT, USA

Saraswati Raju
Jawaharlal Nehru University, New Delhi, India

S Ratick
Clark University, Worcester, MA, USA

T D Roberts
University of Colorado at Boulder, Boulder, CO, USA

G Schwarz
Universidad Nacional de Colombia, Bogota, DC, Colombia

G Setten
Norwegian University of Science and Technology, Trondheim, Norway

K Simonsen
Roskilde University, Roskilde, Denmark

M Sioh
DePaul University, Chicago, IL, USA

C J Smith
University at Albany, SUNY, Albany, NY, USA

M Tanskanen
University of Joensuu, Joensuu, Finland

S Tierney
University of Denver, Denver, CO, USA

C D Upchurch
University of South Carolina, Columbia, USA

M Walton-Roberts
Wilfrid Laurier University, Waterloo, ON, Canada

M Watts
UC Berkeley, Berkeley, CA, USA

A Wearing
University of Otago, Dunedin, New Zealand

D W Wong
George Mason University, Fairfax, VA, USA

R Woodward
Newcastle University, Newcastle upon Tyne, UK

Y-M Yeung
The Chinese University of Hong Kong, Hong Kong, People's Republic of China

VOLUME 8

O Ahlqvist
The Ohio State University, Columbus, OH, USA

S C Aitken
San Diego State University, San Diego, CA, USA, and Norwegian University of Science and Technology, Trondheim, Norway

G Backhaus
Loyola College in Maryland, Baltimore, MD, USA

A J Bailey
University of Leeds, Leeds, UK

D Bennett
Carleton University, Ottawa, ON, Canada

V Berdoulay
CNRS and Université de Pau, Pau, France

L Bondi
University of Edinburgh, Edinburgh, UK

M Boyle
National University of Ireland Maynooth, Co Kildare, Republic of Ireland

E Chacko
The George Washington University, Washington, DC, USA

X Chen
Trinity College, Hartford, CT, USA

A O Chimhowu
University of Manchester, Manchester, UK

P Collier
University of Portsmouth, Portsmouth, UK

D Collins
University of Alberta, Alberta, Canada

D Collins
University of Auckland, Auckland, New Zealand

K R Cox
The Ohio State University, Columbus, OH, USA

T Cresswell
Royal Holloway, University of London, Egham, UK

E Delmelle
University of North Carolina at Charlotte, NC, USA

J Dittmer
University College London, London, UK

D P Dixon
The University of Wales, Aberystwyth, UK

M Dunford
University of Sussex, Brighton, UK

S Elden
Durham University, Durham, UK

R Flowerdew
University of St Andrews, Fife, UK

R Flowerdew
University of St Andrews, St Andrews, UK

B Forest
McGill University, Montréal, QC, Canada

A Franklin
University of Tasmania, Hobart, TAS, Australia

N R Fyfe
University of Dundee, Dundee, UK

M Gren
Hólar University College, Sauðárkrókur, Iceland

M Haldrup
Roskilde University, Roskilde, Denmark

P Hall
University College London, London, UK

S Harrison
University of Exeter, Penryn, UK

S Hickey
University of Manchester, Manchester, UK

M Huxley
University of Sheffield, Sheffield, UK

J P Jones III
University of Arizona, Tucson, AZ, USA

R Jones
Aberystwyth University, Aberystwyth, UK

M Kesby
University of St Andrews, St Andrews, UK

F C Kessler
Frostburg State University, Frostburg, MD, USA

S Kindon
Victoria University of Wellington, Wellington, New Zealand

A D King
State University of New York Binghamton, Binghamton, NY, USA

P T Kingsbury
Simon Fraser University, Burnaby, BC, Canada

A Kobayashi
Queen's University, Kingston, ON, Canada

L Koefoed
Roskilde University, Roskilde, Denmark

U Kothari
University of Manchester, Manchester, UK

D M Lawrence
Virginia State University and Virginia Union University, Richmond, VA, USA

R Le Heron
University of Auckland, Auckland, New Zealand

J Lea
Lancaster University, Lancaster, UK

J Lorimer
King's College London, London, UK

D P McCormack
Oxford University Centre for the Environment, Oxford, UK

C McEwan
Durham University, Durham, UK

C Minca
Royal Holloway, University of London, Egham, UK

D Mitchell
Syracuse University, Syracuse, NY, USA

F Morton
Royal Scottish Academy of Music and Drama, Glasgow, UK

W E Murray
Victoria University of Wellington, Wellington, New Zealand

C J Nash
Brock University, St Catharines, ON, Canada

R P Neumann
Florida International University, Miami, FL, USA

A Nightingale
University of Edinburgh, Edinburgh, UK

J O'Loughlin
University of Colorado, Boulder, CO, USA

P M O'Neill
University of Western Sydney, NSW, Australia

A Paasi
University of Oulu, Linnanmaa, Finland

R Pain
University of Durham, Durham, UK

E Pawson
University of Canterbury, Christchurch, New Zealand

C Perkins
University of Manchester, Manchester, UK

M Perry
Massey University (Wellington), Wellington, New Zealand

J Nederveen Pieterse
University of Illinois, Urbana-Champaign, IL, USA

J D van der Ploeg
Wageningen University, Wageningen, The Netherlands

Saraswati Raju
Jawaharlal Nehru University, New Delhi, India

R Sanders
Temple University, Philadelphia, PA, USA

V R Savage
National University of Singapore, Singapore

B-M Shantz
University of Auckland, Auckland, New Zealand

S J Smith
University of Durham, Durham, UK

L A Staeheli
University of Edinburgh, Edinburgh, UK

J-F Staszak
Université de Genève Geneva, Switzerland

P E Steinberg
Florida State University, Tallahassee, FL, USA

D Storey
University of Worcester, Worcester, UK

L Sýkora
Charles University, Prague, Czech Republic

A R Tickamyer
Ohio University, Athens, OH, USA

K Walsh
University of Sussex, Brighton, UK

B Warf
University of Kansas, Lawrence, KS, USA

S Warren
Eastern Washington University, Cheney, WA, USA

G A Wilson
University of Plymouth, Plymouth, UK

M Woods
Aberystwyth University, Aberystwyth, UK

K Woodward
University of Exeter, Exeter, UK

G Yeung
National University of Singapore, Singapore

VOLUME 9

M B Aalbers
Columbia University, New York, NY, USA

M Abreu
University of Cambridge, Cambridge, UK

O Ahlqvist
The Ohio State University, Columbus, OH, USA

J O Bærenholdt
Roskilde University, Roskilde, Denmark

T J Barnes
University of British Columbia, Vancouver, BC, Canada

R Billen
University of Liege, Liege, Belgium

R Black
University of Sussex, Brighton, UK

J S Boggs
Brock University, St Catharines, ON, Canada

G Bridge
University of Bristol, Bristol, UK

K Browne
University of Brighton, Brighton, UK

T Butler
King's College London, London, UK

D Carr
University of California, Santa Barbara, CA, USA

M Charlton
National University of Ireland Maynooth, Maynooth, Republic of Ireland

J P Clark
Loyola University, New Orleans, LA, USA

M Cope
University of Vermont, Burlington, VT, USA

A Cumbers
University of Glasgow, Glasgow, UK

J Cupples
University of Canterbury, Christchurch, New Zealand

J Revilla Diez
Leibniz University of Hannover, Hannover, Germany

I Docherty
University of Glasgow, Glasgow, UK

M Doevenspeck
University of Bayreuth, Bayreuth, Germany

M Dunford
University of Sussex, Brighton, UK

D W Edgington
University of British Columbia, Vancouver, BC, Canada

C Fox
Dartmouth College, Hanover, NH, USA

M F Goodchild
University of California, Santa Barbara, CA, USA

K H Halfacree
Swansea University, Swansea, UK

L M B Harrington
Kansas State University, Manhattan, KS, USA

R Hayter
Simon Fraser University, Byrnaby, BC, Canada

A Holden
University of Durham, Durham, UK

J R L Howells
University of Manchester, Manchester, UK

R Johnston
University of Bristol, Bristol, UK

J-K Jung
University of North Dakota, Grand Forks, ND

M Kiese
Leibniz University of Hannover, Hannover, Germany

A Kobayashi
Queen's University, Kingston, ON, Canada

V Kolossov
Russian Academy of Sciences, Moscow, Russia

L Kong
National University of Singapore, Singapore

A Lord
University of Liverpool, Liverpool, UK

D MacKinnon
University of Aberdeen, Aberdeen, UK

G MacLeod
University of Durham, Durham, UK

K A MacTavish
Oregon State University, Corvallis, OR, USA

J Mennis
Temple University, Philadelphia, PA, USA

F Mizuoka
Hitotsubashi University, Tokyo, Japan

A Nayak
University of Newcastle, Newcastle, UK

H Neo
National University of Singapore, Singapore

P Nijkamp
Free University, Amsterdam, The Netherlands

E S Norman
University of British Columbia, Vancouver, BC, Canada

H G Overman
London School of Economics, London, UK

A Paasi
University of Oulu, Oulu, Finland

B Page
University College London, London, UK

J Pearce
University of Canterbury, Christchurch, New Zealand

R Peet
Clark University, Worcester, MA, USA

I M M Pires
Universidade Nova de Lisboa, Lisbon, Portugal

V Preston
York University, Toronto, ON, Canada

N M Rantisi
Concordia University, Montréal, QC, Canada

J M Read
Maxwell School of Syracuse University, Syracuse, NY, USA

P Robbins
University of Arizona, Tucson, AZ, USA

S Salamon
University of Texas at Dallas, Dallas, TX, USA

H V Scott
Aberystwyth University, Aberystwyth, UK

J Shaw
University of Plymouth, Plymouth, UK

N K Shortt
University of Edinburgh, Edinburgh, UK

I G Simmons
University of Durham, Durham, UK

T R Slater
University of Birmingham, Birmingham, UK

E W Soja
University of California, Los Angeles, CA, USA, and London School of Economics and Political Science, London, UK

M Sokol
Queen Mary, University of London, London, UK

M Tewdwr-Jones
University College London, London, UK

F Tödtling
Vienna University of Economics and Business Administration, Vienna, Austria

J Tomaney
Newcastle University, Newcastle upon Tyne, UK, and Monash University, Melbourne, VIC, Australia

M Torrado
Maxwell School of Syracuse University, Syracuse, NY, USA

A Treivish
Russian Academy of Sciences, Moscow, Russia

N Van de Weghe
Ghent University, Gent, Belgium

B Werlen
Friedrich-Schiller University, Jena, Germany

J Winders
Syracuse University, Syracuse, NY, USA

M Woods
Aberystwyth University, Aberystwyth, UK

N Wrigley
University of Southampton, Southampton, UK

VOLUME 10

A Albet
Autonomous University of Barcelona, Bellaterra, Spain

T A Arentze
Eindhoven University of Technology, Eindhoven, The Netherlands

R Atkinson
University of Tasmania, Hobart, TAS, Australia

M Azaryahu
University of Colorado at Boulder, Boulder, CO, USA

M Azaryahu
University of Haifa, Haifa, Israel

A Bebbington
University of Manchester, Manchester, UK

L Bian
University at Buffalo – SUNY, Buffalo, NY, USA

E Casetti
Ohio State University, Columbus, OH, USA

G P Chapman
Centre for Advanced Study, Oslo, Norway

M Coleman
The Ohio State University, Columbus, OH, USA

T J Cooke
University of Connecticut, Storrs, CT, USA

T W Crawford
East Carolina University, Greenville, NC, USA

M A de Souza
Universidade de São Paulo, Campinas, Brazil

A V Di Vittorio
University of California – Berkeley, Berkeley, CA, USA

M Duffy
University of Melbourne, Carlton, VIC, Australia

N Duncan
University of Cambridge, Cambridge, UK

S Elden
Durham University, Durham, UK

S Engel-Di Mauro
SUNY New Paltz, NY, USA

R Flowerdew
University of St Andrews, St Andrews, UK

K E Foote
University of Colorado at Boulder, Boulder, CO, USA

J D Gatrell
Indiana State University, Terre Haute, IN, USA

C Gibson
University of Wollongong, Wollongong, NSW, Australia

B Greenhough
Queen Mary, University of London, London, UK

D A Griffith
University of Texas – Dallas, Richardson, TX, USA

M Haldrup
Roskilde University, Roskilde, Denmark

A L Hansen
Lund University, Lund, Sweden

R Harris
Hamilton, ON, Canada

K E Haynes
George Mason University, Fairfax, VA, USA

P Howell
University of Cambridge, Cambridge, UK

R W Jackson
West Virginia University, Morgantown, WV, USA

A James
Queen Mary, University of London, London, UK

B Jessop
Lancaster University, Lancaster, UK

Ron Johnston
University of Bristol, Bristol, UK

M Jones
Aberystwyth University, Aberystwyth, UK

C Kim
University of Cincinnati, Cincinnati, OH, USA

R Kitchin
NUI Maynooth, Maynooth, Republic of Ireland

A Kobayashi
Queen’s University, Kingston, ON, Canada

E W LaFary
The University of Auckland, Auckland, New Zealand

N S Lam
Louisiana State University, Baton Rouge, LA, USA

A Latham
University College London, London, UK

J Lee
Kent State University, Kent, OH, USA

L Lo
York University, Toronto, ON, Canada

Y Lu
Texas State University – San Marcos, San Marcos, TX, USA

S M Manson
University of Minnesota, Minneapolis, MN, USA

D McCormack
Oxford University Centre for the Environment, Oxford, UK

J McGarrigle
Universidade de Lisboa, Lisbon, Portugal

J Mohan
University of Southampton, Southampton, UK

K Newman
Rutgers University, New Brunswick, NJ, USA

M E O’Kelly
The Ohio State University, Columbus, OH, USA

J Oldfield
University of Glasgow, Glasgow, UK

R Panelli
University College London, London, UK

D Pinder
Queen Mary, University of London, London, UK

M F Poulsen
Macquarie University, Sydney, NSW, Australia

M Ramutsindela
University of Cape Town, Rondebosch, South Africa

G M Robinson
Kingston University, Kingston upon Thames, UK

S M Ruddick
University of Toronto, Toronto, ON, Canada

G Rushton
The University of Iowa, Iowa City, IA, USA

A Saldanha
University of Minnesota, Minneapolis, MN, USA

N F Sayre
University of California – Berkeley, Berkeley, CA, USA

N Schuurman
Simon Fraser University, Burnaby, BC, Canada

T Schwanen
Utrecht University, Utrecht, The Netherlands

D M Scott
McMaster University, Hamilton, ON, Canada

D Sibley
University of Leeds, Leeds, UK

A Smith
Queen Mary, University of London, London, UK

Q Stevens
University College London, London, UK

P Sunley
University of Southampton, Southampton, UK

Eric C Thompson
National University of Singapore, Singapore

C Tiwari
The University of Iowa, Iowa City, IA, USA

D Turnbull
ACSIS, Melbourne University, Melbourne, VIC, Australia

L Twigg
University of Portsmouth, Portsmouth, UK

D J Unwin
Birkbeck, University of London, London, UK

A C Vias
University of Connecticut, Storrs, CT, USA

M Williams
Oxford University, Oxford, UK

K D Willis
Royal Holloway, University of London, Egham, UK

A C Winstanley
National University of Ireland, Maynooth, Republic of Ireland

D W Wong
George Mason University, Fairfax, VA, USA

M Yuan
The University of Oklahoma, Norman, OK, USA

P Zusman
University of Buenos Aires, Buenos Aires, Argentina

VOLUME 11

C G Alvstam
Göteborg University, Gothenburg, Sweden

D Banister
University of Oxford, Oxford, UK

C Berndt
Johann Wolfgang Goethe-Universitaet Frankfurt, Frankfurt am Main, Germany

K Button
George Mason University, Fairfax, VA, USA

M Cawley
National University of Ireland, Galway, Republic of Ireland

G P Chapman
Lancaster University, Lancaster, UK

J Connell
University of Sydney, Sydney, NSW, Australia

M Cope
University of Vermont, Burlington, VT, USA

T Cresswell
Royal Holloway, University of London, Egham, UK

V J Del Casino, Jr.
California State University, Long Beach, CA, USA

D Delaney
Amherst College, Amherst, MA, USA

M Denil
Washington, DC, USA

M Dijst
Utrecht University, Utrecht, The Netherlands

B Domański
Jagiellonian Univeristy, Krakow, Poland

T Edensor
Manchester Metropolitan University, Manchester, UK

J A Elliott
University of Brighton, Brighton, UK

W M Gesler
University of North Carolina, Chapel Hill, NC, USA

V Gidwani
University of Minnesota, Minneapolis, MN, USA

J Glassman
University of British Columbia, Vancouver, BC, Canada

R Grant
University of Miami, Coral Gables, FL, USA

M Gren
Hólar University College, Sauðárkrókur, Iceland

J Gutiérrez
Universidad Complutense de Madrid, Madrid, Spain

T Hall
University of Gloucestershire, Cheltenham, UK

M Henry
Massey University, Palmerston North, New Zealand

J Hine
University of Ulster, Newtownabbey, UK

J-Y Hsu
National Taiwan University, Taipei, Taiwan

S Huang
National University of Singapore, Singapore

P J Hugill
Texas A & M University, College Station, TX, USA

P F Kelly
York University, Toronto, ON, Canada

A Kent
University of Manchester, Manchester, UK

R D Knowles
University of Salford, Greater Manchester, UK

M Kurtz
Open University, Ottawa, ON, Canada

D Ley
University of British Columbia, Vancouver, BC, Canada

R Lippuner
Friedrich-Schiller-University Jena, Jena, Germany

S Lloyd-Evans
University of Reading, Reading, UK

B P Y Loo
University of Hong Kong, Hong Kong, People's Republic of China

A Mace
University of Westminster, London, UK

T Marsden
Cardiff University, Cardiff, UK

C McEwan
Durham University, Durham, UK

Nick Megoran
University of Newcastle upon Tyne, Newcastle upon Tyne, UK

G Mohan
The Open University – Milton Keynes, UK

P L Mokhtarian
University of California – Davis, Davis, CA, USA

L Mu
University of Georgia, Athens, GA, USA

D Mustafa
King's College London, London, UK

A Nightingale
University of Edinburgh, Edinburgh, UK

G Norcliffe
York University, Toronto, ON, Canada

K O'Connor
University of Melbourne, Parkville, Australia

J O'Loughlin
University of Colorado, Boulder, CO, USA

K Olds
University of Wisconsin – Madison, Madison, WI, USA

D T Ory
Federal Transit Administration, Washington, DC, USA

U F Paleo
University of Extremadura, Caceres, Spain

R Phillips
University of Liverpool, Liverpool, UK

D Pinder
Queen Mary, University of London, London, UK

J P H Poon
University at Buffalo – SUNY, Buffalo, NY, USA

R B Potter
University of Reading, Reading, UK

J M Powell
Monash University, Melbourne, VIC, Australia

M Power
University of Durham, Durham, UK

J Preston
University of Southampton, Southampton, UK

G M Robinson
Kingston University, Kingston upon Thames, UK

J Round
University of Birmingham, Birmingham, UK

T Schwanen
Utrecht University, Utrecht, The Netherlands

J P Sharp
University of Glasgow, Glasgow, UK

S-L Shaw
The University of Tennessee, Knoxville, TN, USA

M Sheller
Swarthmore College, Swarthmore, PA, USA

F M Smith
University of Dundee, Dundee, UK

J M Smith
Texas A & M University, College Station, TX, USA

R G Smith
Swansea University, Swansea, UK

D Straussfogel
Pennsylvania State University, DuBois, PA, USA

U Strohmayer
National University of Ireland, Galway, Republic of Ireland

A Sturt
University of Westminster, London, UK

D Thien
California State University, Long Beach, CA, USA

D J Unwin
University College London, London, UK

C von Schilling
University of Northern British Columbia, Prince George, BC, Canada

B Warf
Florida State University, Tallahassee, FL, USA

B Werlen
Friedrich-Schiller-University Jena, Jena, Germany

P White
University of Westminster, London, UK

M Whitehead
Aberystwyth University, Aberystwyth, UK

D Wilson
University of Illinois, Urbana, IL, USA

J Zhang
University of Minnesota, Tacoma, WA, USA

VOLUME 12

Fiona Allon
Centre for Cultural Research, University of Western Sydney, Australia

S Batterbury
University of Melbourne, Melbourne, VIC, Australia

M Batty
University College London, London, UK

D Bell
University of Leeds, Leeds, UK

G Bridge
University of Bristol, Bristol, UK

G Brown
University of Leicester, Leicester, UK

A Buttimer
University College Dublin, Dublin, Republic of Ireland

J Byrne
Griffith University, Gold Coast, QLD, Australia

B Christophers
The University of Auckland, Auckland, New Zealand

G P Clarke
University of Leeds, Leeds, UK

A Cochrane
The Open University, Milton Keynes, UK

D Conradson
University of Southampton, Southampton, UK

G Crow
University of Southampton, Southampton, UK

M Davidson
University of Western Sydney, Sydney, NSW, Australia

B Derudder
Ghent University, Gent, Belgium

J Dodson
Griffith University, Brisbane, QLD, Australia

B Gleeson
Griffith University, Brisbane, QLD, Australia

M F Goodchild
University of California, Santa Barbara, CA, USA

K H Halfacree
Swansea University, Swansea, UK

M G Hatvany
Université Laval, Quebec City, QC, Canada

C Haylett
University of Manchester, Manchester, UK

L Holt
University of Reading, Reading, UK

J Horton
University of Northampton, Northampton, UK

R Hudson
Durham University, UK

M Jayne
University of Manchester, Manchester, UK

N C Johnson
Queen's University, Belfast, Northern Ireland

D Kay
Aberystwyth University, Aberystwyth, UK

I M Keighren
University of Edinburgh, Edinburgh, UK

P Knox
Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

P Kraftl
University of Leicester, Leicester, UK

D Kunze
Pennsylvania State University, University Park, PA, USA

M Lauria
Clemson University, Clemson, SC, USA

R Lee
Queen Mary, University of London, London, UK

K D Lilley
Queen's University Belfast, Belfast, UK

F MacDonald
University of Melbourne, Melbourne, VIC, Australia

S C MacKian
Open University, Milton Keynes, UK

V Mamadouh
University of Amsterdam, Amsterdam, The Netherlands

C Mason
University of Strathclyde, Glasgow, UK

G Mercier
Laval University, Quebec City, QC, Canada

C Milligan
Lancaster University, Lancaster, UK

M Pavlovskaya
Hunter College – CUNY, New York, NY, USA

L Peake
York University, Toronto, ON, Canada

C Philo
University of Glasgow, Glasgow, UK

G M Robinson
Kingston University, Kingston upon Thames, UK

S Rodgers
The Open University, Milton Keynes, UK

T R Slater
University of Birmingham, Birmingham, UK

E Street
King's College London, London, UK

K Terlouw
Utrecht University, Utrecht, The Netherlands

C R Warren
University of St Andrews, St Andrews, UK

M Watson
University of Sheffield, Sheffield, UK

R K Whelan
University of Texas, Arlington, TX, USA

B Wisner
Benfield UCL Hazard Research Centre, London, UK

J Wolch
University of Southern California, Los Angeles, CA, USA

CONTENTS

VOLUME 1

A

C

VOLUME 2

C

VOLUME 3

D

VOLUME 4

F

[†]Deceased, 22 August 2007.

G

VOLUME 5

H

VOLUME 6

J

K

L

M

VOLUME 7

M

N

VOLUME 9

Q

R

VOLUME 11

S

VOLUME 12

U

SUBJECT CLASSIFICATION

CARTOGRAPHY AND GIS

Art and Cartography
Atlases
Biodiversity Mapping
Cartographic Animation
Cartography, History of
Cartography in Islamic Societies
Census Geography
Census Mapping
Children and Mapping
Color, Mapping
Computational Human Geography
Counter Mapping
Critical Cartography
Critical GIS
Digital Data, Historical Geography and
Digital Earth
Disease Mapping
Electoral Cartography
Feminism, Maps and GIS
Generalization
Geodemographics
Geodesy
Geographical Masking
Geomatics
Georeferencing, Geocoding
Geospatial Intelligence
Geovisualization
GIS and Cartography
GIS and Society
GIS, Mobile and Locational Based Services
GIS, Public Participation
GIScience and Systems
Global Positioning/GPS
Indigenous Mapping
Information Graphics
Integrated Spatial Data Infrastructure
Internet/Web Mapping
Map Hacking
Map Interactivity
Map Libraries and Archives
Map Perception and Cognition
Map Types
Mapping Agencies
Mapping, Commercial
Mapping, Cyberspace
Mapping, Distributed
Mapping, Non Western
Mapping, Philosophy
Mapping, Race and Ethnicity
Mapping, Topographic
Maps
Maps and Governance
Maps and Protest
Maps and the State
Oceanographic Mapping
Performative and Embodied Mapping
Photogrammetry/Aerial Photography
Projections
Qualitative Geographic Information Systems
Qualitative Spatial Reasoning
Quantitative Revolution
Representation Mapping
Science and Scientism, Cartography
Space Time Modeling
Spatial Databases
Spatial Data Mining, Geovisualization
Spatial Ontologies
Surveying
Symbolism, Iconography

DEVELOPMENT GEOGRAPHY

Aid
Brain Drain
Brandt Commission
Brown Agenda
Child Labor
Civil Society
Colonialism I
Colonialism II
Colonialism, Internal
Commodity Chains
Debt
Deforestation

Dependency
Desertification
Developmentalism
Development I
Development II
Digital Divide
East Asian Miracle
Empire
Empowerment
Eurocentrism
Export Processing Zones
Extended Metropolitan Region
Fair Trade
Famine
First World
Global Commodity Chains
Governance, Good
Green Revolution
Health and Development
HIV/AIDS in Developing Countries
Imperialism, Cultural
Indigenous Geographies
Indigenous Knowledges
Informal Sector
Intermediate Technology
Latin American Structuralist School
Livelihoods
Locality Debates
Migrant Workers
Modernization Theory
Neocolonialism
Neoliberalism and Development
Nongovernmental Organizations
Orientalism
Participation
Postcolonial Cities
Postcolonialism/Postcolonial Geographies
Postdevelopment
Poverty
Remittances
Resistance
Rio Summit*
Second World
Social Capital
Squatter Settlements
Structural Adjustment
Sustainable Development
Theocracy
Third World
Third World Cities
Trade, International
Transnational Corporations in Developing Countries
Tropical Geography
Vulnerability

DISCIPLINARY MATTERS

Anglo American/Anglophone Hegemony
Anthropology and Human Geography
Chinese Language Geography
Citation Geography
Communist and Post Communist Geographies
Critical Geography
Cultural Studies and Human Geography
Dutch Human Geography
Economics and Human Geography
Enlightenment Geography
Environmental Studies and Human Geography
Feminist Groups within Geography
Francophone Geography
Geographical Journals
German Language Geography
Historical Geography
Human Geography
Interdisciplinarity
Italian Language Geography
Ivy League and Geography in the US
Japanese Geography
Lusophone Geography
Medieval Geography
Military and Geography
National Schools of Geography
Nordic Geography
Oxbridge Geographies
Philosophy and Human Geography
Physical Geography and Human Geography
Redbrick University Geography in Britain
Research Funding Bodies
Russian Language Geography
Spanish Language Geography
Urban Planning and Human Geography

ECONOMIC GEOGRAPHY

Agglomeration
Business Services
Capital and Space
Capitalism
Capitalism and Division of Labor
Commodity Chains
Competitiveness
Consumption
Core Periphery Models
Corporate Responsibilities
Corridor and Axis Development
Creativity
Cultural Economy
Debt
De Industrialization
e Business and e Commerce
Economic Crises

HEALTH AND MEDICAL GEOGRAPHY

HISTORICAL GEOGRAPHY

METACONCEPTS

METHODS

Cellular Automata
Census Mapping
Chaos and Complexity
Choice Modeling
Complexity Theory, Nonlinear Dynamic Spatial Systems
Content Analysis
Cross Cultural Research
Diaries (Video, Audio or Written)
Discourse Analysis
Ecological Fallacy
Edge Effects
Embodied Knowing
Emotional Knowing
Entropy Maximising Models
Error (Propagation and Modeling)
Ethical Issues in Research
Ethnography
Evolutionary Algorithms
Experimental Design
Exploratory Spatial Data Analysis
Factor Analysis and Principal Components Analysis
Feminist Methodologies
Fieldwork
First Law of Geography
Focus Groups
Fractal Analysis
Fuzzy Set and Fuzzy Logic
Genealogy Method
Geocomputation
Geographical Masking
Geographically Weighted Regression
Georeferencing, Geocoding
Grounded Theory
Haptic or Touch Based Knowledge
Hub Network Location
Hypothesis Testing
Input–Output Analysis
Intensive/Extensive Research
Internet Based Measurement
Interviews: In Depth, Semi Structured
Kriging and Variogram Models
Landscape Iconography
Landscape Perception
Language and Research
Life Course Approaches
Location Analysis
Longitudinal Methods (Cohort Analysis, Life Tables)
Markov Chain Analysis
Masculinism
Mental Maps
Mixed and Multiple Methods
Modifiable Areal Unit Problem
Monte Carlo Simulation
Movies and Films, Analysis of
Multidimensional Scaling
Naturalistic Testing
Neighborhood Effects
Network Analysis
Neural Networks
Oral History
Oral History, Ecological
Overlay (in GIS)
Participant Observation
Participatory Action Research
Participatory Video
Performance, Research as
Photographs
Point Pattern Analysis
Polyvocality
Psychoanalysis
Q Method/Analysis
Qualitative Geographic Information Systems
Quantitative Data
Quantitative Methodologies
Questionnaire Survey
Regionalization/Zoning Systems
Regression, Linear and Nonlinear
Reliability and Validity
Remote Sensing
Representation and Re presentation
Sampling
Scale Analytical
Scientific Method
Segregation Indices
Selection Bias
Semiotics
Shift Share Analysis
Simulation
Situated Knowledge, Reflexivity
Sound and Music
Space Time Modeling
Spatial Analysis, Critical
Spatial Autocorrelation
Spatial Clustering, Detection and Analysis of
Spatial Data Mining, Cluster and Pattern Recognition
Spatial Data Mining, Geovisualization
Spatial Data Models
Spatial Expansion Method
Spatial Filtering/Kernel Density Estimation
Spatial Interaction Models
Spatial Interpolation
Spatially Autoregressive Models
Statistics, Descriptive
Statistics, Inferential
Statistics, Overview
Statistics, Spatial
Structural Equations Models
Subalternity
Subjectivity

Text, Textual Analysis
Thiessen Polygon
Time Geographic Analysis
Time Series Analysis
Time Space Diaries
Transcripts (Coding and Analysis)
Translation
Trend Surface Models
Triangulation
Uncertainty
Visualization, Feminist

NATURE/ENVIRONMENT

Agrarian Transformations
Agri Environmentalism and Rural Change
Animal Geographies
Biodiversity
Biodiversity Mapping
Climate Change
Conservation and Ecology
Culture/Natures
Deforestation
Desertification
Ecotourism
Environment
Environmental Hazards
Environmental Health
Environmentalism
Environmental Justice
Environmental Policy
Environmental Regulation
Environmental Security
Environmental Studies and Human Geography
Environment, Historical Geography of
Gardens and Gardening
Green Revolution
Land Change Science
National Parks
Natural Resources
Nature
Nature, Historical Geographies of
Nature, History of
Nature, Performing
Nature, Social
Natures, Charismatic
Natures, Gendered
Natures, Postcolonial
Plant Geographies
Political Ecology
Radical Environmentalism
Resource and Environmental Economics
Resource Management, Rural
Sustainability
Sustainable Development
Urban Habitats/Nature
Waste Management
Water Management
Wetlands and Reclamation
Wilderness

PEOPLE

Barnes, T.
Beaujeu Garnier, J.
Berry, B.
Bobek, H.
Bowman, I.
Christaller, W.
Claval, P.
Cloke, P.
Cohen, S.
Corbridge, S.
Cosgrove, D.
Cox, K.
Darby, H. C.
Dear, M. J.
Dicken, P.
Dudley Stamp, L.
Evans, E. E.
Garrison, W.
Golledge, R.
Gottmann, J.
Gregory, D.
Hagerstrand, T.
Haggett, P.
Harley, J. B.
Hartshorne, R.
Harvey, D.
Hettner, A.
Jackson, P.
Johnston, R. J.
Kolossov, V.
Kropótkin, P.
Lacoste, Y.
Ley, D.
Lowenthal, D.
Mackinder, H. J.
Massey, D.
McDowell, L.
Meinig, D.
Olsson, G.
Paasi, A.
Peet, R.
Pred, A.
Reclus, E.
Ritter, C.
Santos, M.
Sauer, C.
Scott, A.
Smith, N.
Soja, E.

Storper, M.
Taylor, G.
Taylor, P.
Thrift, N.
Tuan, Y. F.
Vidal de la Blache, P.
von Humboldt, A.
Watts, M. J.
Wilson, A.
Wreford Watson, J.
Wright, J. K.

PHILOSOPHY AND GEOGRAPHY

Activist Geographies
Actor Network Theory/Network Geographies
Anarchism/Anarchist Geography
Anthropogeography (After Ratzel)
Applied Geography
Avant Garde/Avant Garde Geographies
Behavioral Geography
Berkeley School
Chicago School
Christian Geography
Cognitive Geography
Computational Human Geography
Critical Geography
Critical Rationalism (After Popper)
Critical Realism/Critical Realist Geographies
Critical Theory (After Habermas)
Cultural Materialism
Cultural Turn
Darwinism (and Social Darwinism)
Deconstruction
Determinism/Environmental Determinism
Developmentalism
Dialectical Reasoning and Dialectical Materialism
Dialogism (After Bakhtin)
Ecology
Ethnomethodology/Ethnomethodological Geography
Existentialism/Existential Geography
Feminism/Feminist Geography
Feminist Geography, Prehistory of
Feminist Political Economy
Field Geographies
Fluidity–Fixity
Foucauldianism
Functionalism (Including Structural Functionalism)
Geography, History of
GIS and Society
Historical Geographical Materialism
Humanism/Humanistic Geography
Human Nonhuman
Idealism/Idealist Human Geography
Indigenous Geographies
Institutionalism/Institutional Geographies
Lamarck(ian)ism
Local–Global
Los Angeles School of Post Modern Urbanism
Marxism/Marxist Geography I
Marxism/Marxist Geography II
Military Geographies
Nature Culture
Non Representational Theory/Non Representational Geographies
People's Geography
Phenomenology/Phenomenological Geography
Philosophy and Human Geography
Physical Geography and Human Geography
Political Ecology
Political Economy, Geographical
Positivism/Positivist Geography
Possibilism
Posthumanism/Posthumanistic Geographies
Postmodernism/Postmodern Geography
Post Phenomenology/Post Phenomenological Geographies
Poststructuralism/Poststructuralist Geographies
Pragmatism/Pragmatist Geographies
Probabilism
Probability Models
Psychoanalytic Theory/Psychoanalytic Geographies
Psychotherapy/Psychotherapeutic Geographies
Quantitative Revolution
Radical Environmentalism
Radical Geography
Radical Political Economy
Rational Choice Theory (and Rational Choice Marxism)
Regional Geography I
Regional Geography II
Regionalisations, Everyday
Regional Science
Science and Scientism, Cartography
Self Other
Semiotics
Situationism/Situationist Geography
Social Studies of Scientific Knowledge
Society–Space
Space Time
Spatial Science
Structural Marxism
Structuralism/Structuralist Geography
Structurationist Geography
Structuration Theory
Surrealism/Surrealist Geographies
Symbolic Interactionism
Systems
Systems Theory
Time Geography
Vichianism (After Vico)
Welfare Geography

POLITICAL GEOGRAPHY

Activism
Activist Geographies
Ageing and Health
Anti Geopolitics
Apartheid/Post Apartheid
Biopolitics
Borderlands
Buffer Zone
Citizenship
Citizenship and Governmentality, Rural
Cold War
Colonialism I
Colonialism II
Communist and Post Communist Geographies
Critical Geopolitics
Cultural Politics
Democracy
Devolution
Electoral Cartography
Electoral Districts
Electoral Geography
Empire
Environmental Justice
Environmental Security
Ethnic Conflict
Eurocentrism
Fatherland/Homeland
Feudalism and Feudal Society
Geopolitics
Geopolitics and Religion
Gerrymandering
Globalization, Economic
Governance
Governance, Corporate
Governance, Good
Governance, Transport
Governance, Urban
Governmentality
Hegemony
Human Rights
Ideology
Imperialistic Geographies
Irredentism
Liberalism
Maps and Governance
Military and Geography
Nation
National Spatialities
Nationalism
Nationalism, Historical Geography of
Neoliberalism
Neoliberalism, Urban
Nongovernmental Organizations
Place, Politics of
Political Boundaries
Political Geography
Political Representation
Postcolonialism/Postcolonial Geographies
Postconflict Geographies
Protest, Rural
Public Good
Public Policy
Regionalism
Regulation
Representation, Politics of
Socialism
Social Movements
Sovereignty
State
State Theory
Superpower
Territory and Territoriality
Terrorism
War
War, Historical Geography and
World System

POPULATION GEOGRAPHY

Ageing and Mobility
Census Geography
Demography
Diaspora
Emigration
Fertility
Genetics
Immigration II
Migration
Migration, Historical Geographies of
Population Geography
Refugees and Displacement
Rural Populations
Segregation
Transnationalism

REGIONAL DEVELOPMENT

Aménagement du Territoire: Territorial Development
Cassa per il Mezzogiorno
City Region
Concentrated Deconcentration
Core Periphery Models
Corridor and Axis Development
Cumulative Causation
De Industrialization
De Localization
Edge Cities
Europe of Regions
Greenfield Development
Growth Poles, Growth Centers

RURAL GEOGRAPHY

SOCIAL & CULTURAL GEOGRAPHY

TRANSPORT GEOGRAPHY

Transport, Rural
Transport, Urban

URBAN GEOGRAPHY

Anti Urbanism
Cellular Automata
Central Business District
Central Place Theory
Chinese Urbanism
City Marketing
City Region
Counterurbanization
Defensible Space
Edge Cities
Financial Centers, International
Flâneur, The
Gated Communities/Privatopias
Gender in the City
Gentrification
Ghettos
Governance, Urban
Historical Geographies, Urban
Housing
Imperial Cities
Industrial City
Informational City
Islamic Urbanism
Land Rent Theory
Malls/Retail Parks
Mega Cities
Modern City
Multicultural City
Neighborhood Change
Neighborhoods and Community
Neoliberalism, Urban
Networks, Urban
New Towns
New Urbanism
NIMBY
Planning, Urban
Polycentricity
Postcolonial Cities
Postmodern City
Post Socialist Cities
Public Spaces, Urban
Redlining
Regeneration to Renaissance
Segregation Indices
Segregation, Urban
Situationism/Situationist Geography
Situationist City
Slums
Social Justice, Urban
Street Names and Iconography
Suburbanization
Sustainability, Urban
Third World Cities
Tourism, Urban
Transport, Urban
Underclass
Urban Architecture
Urban Design
Urban Growth Machine
Urban Habitats/Nature
Urbanism
Urbanization
Urban Modeling
Urban Morphologies, Historical
Urban Morphology
Urban Order
Urban Policy
Urban Regimes
Urban Representation/Imagination
Urban Village
Urban–Rural Continuum
Utopian Cities
Waterfront Development
World/Global Cities

FOREWORD

We should reflect more on the increasingly interconnected and interdependent world we live in. Places that a century ago would have taken months to reach by boat can be reached in hours on a plane and at a price affordable to many more people. Changes in financial markets on one side of the globe instantly ricochet around the planet. Decisions taken in one country or at a supra national scale affect jobs and well being in another. This is the world we have inherited sixty years after the adoption of the Universal Declaration of Human Rights, which reminds us of the work still needed to implement those shared values.

For while the Earth seems to be getting smaller in space and time, the relative differences between places are often growing apace. The situation in many developing countries is currently regressing with falling life expectancy, rising national debts, and weakening economies. In the developed North there continue to be large differences in the standards of living between rich and poor, core and periphery, rural and urban areas. Consequently, the mobility of people is increasing around the globe with millions of labour migrants and refugees on the move, seeking better lives elsewhere. In addition, humanity is beginning to recognise and address the significant global challenge of climate change which is in part created by the processes of development and globalisation, but will require to be addressed with principles of climate justice.

The *International Encyclopedia of Human Geography* documents and explains all of these issues, and many more besides. Focused through the spatial lens of a modern geography sensitive to how social, economic, political, cultural or environmental processes work within and between places, the entries cover the full spectrum of issues facing humanity today across the planet. Together the essays provide a fascinating overview of the diverse, complex and sometimes paradoxical relationships between people, places and environments, written in a style accessible to students and interested parties. As well, the vast array of methodologies and theories employed by geographers and others is documented, to make sense of the developments now occurring. Indeed, in the very fact that it contains 914 essays, written by 844 contributors from over 40 countries, it is itself a product of the way in which the geography of communication and cooperation has rapidly evolved in recent years!

The challenges facing all of us, whether they concern the present global economic downturn, survival in a country at war, managing environmental change, and a host of other pressing issues, require a broad and deep knowledge of the fundamental processes shaping our future. The *International Encyclopedia of Human Geography* provides a comprehensive overview of that knowledge and points to the tools needed to build planetary citizenship and to think through a more ethical version of globalisation. I hope that it will be used extensively by present and future generations so that, as the planet seemingly shrinks in size, so the problems we face and the differences between us shrink too.

Mary Robinson

Mary Robinson
President of Realizing Rights: The Ethical Globalisation Initiative
Honorary President of Oxfam International
United Nations High Commissioner for Human Rights, 1997–2002
President of Ireland, 1990–1997

FOREWORD

Over the past 100 hundred years or so, human geography has grown into one of the most vibrant university disciplines around the world. Expanding from its origins in the colonial pursuit of geographic knowledge, modern human geography has developed into a diverse collection of sophisticated, spatially inflected knowledges underpinned by a refined set of theoretical concepts and methodological tools. As a result, defining contemporary human geography is not an easy task, not least because what passes for geographic theory and praxis varies across time and space. For us, what marks the discipline from other social and natural sciences is its focus on the relationship between people and the world they inhabit, its core metaconcepts (such as space, place, landscape, nature, mobility, environment, and scale), and its use of geographically sensitive methods of data generation and analysis. Human geography, as the contents list of this encyclopedia amply demonstrates, focuses on key issues of the day, and opens up new and vital perspectives on questions that affect our everyday lives. It engages with and problematizes apparently unequivocal statements such as:

- the planet is large;
- the planet teems with life; and
- the planet is under threat.

Through the lens of human geography, these statements are revealed as equivocal and often paradoxical. What do we mean by 'large' in a period in which a scientific shift based on the discovery of fractals has expanded our perception of the length and complexity of lines at exactly the same time that transport and communications technologies shrink the world to such an extent that we can interact in real time with people on the far side of the planet and maps of the planet can be called up online by ordinary citizens and analyzed in diverse ways? What do we mean by 'life' in a period in which we keep discovering more of it in more and more unlikely places, while scientific and cultural changes are continually expanding our definition of what 'life' consists of? What do we mean by 'threat' when we are unable to agree on the nature of a threat, let alone a solution, when technologies that offer the promise of salvation all too often create new problems and when increasing populations offer new resources as well as consume them?

Human geography seeks to answer such questions not only in order to understand the world but also to make practical interventions. Unlike many disciplines, it is built on unsure foundations without much in the way of a canon. Some might say that that is its attraction; that it has perennially been driven by the changing world around it. The discipline is willing to go where other disciplines fear to tread because its history is loosely connected, more like a conglomerate than a series of well defined strata – it has not been wed to a single or limited theoretical approach but rather it has explored and drawn together all manner of approaches in order to address issues geographically. Equally though, it is born out of a longing for a planetary citizenship, with that longing sitting alongside a radical appreciation of how all the differences that make up the discipline both compromise and strengthen that goal. Whatever might be the case, at the heart of this endeavor has been the notion that geography matters – the spatial configuration of events is not a mere add on to other somehow deeper, more abstract aspatial processes, but rather is central to how the processes unfold.

These tensions between a discipline organized around understanding the world and a discipline consisting of and defined by its own approaches, are manifest in how geography tells its own history (and how one might organize an encyclopedia). One way would be to trace the changing approaches, epistemologies, and concerns of the discipline.

Thus one could track the multiple interpretations of the writing of Alexander von Humboldt, often cited as one of the key disciplinary ancestors. Alternately, one could chart the intellectual genealogies of ‘key’ texts, concepts, and places to provide counterpoints to more traditional chronological histories of various theoretical schools. Or, one could create accounts that organize the discipline via key processes and events in the world. To take other examples, there are presently attempts to write about differences in more satisfying and nuanced ways, ways that can bring about new means of living and experiencing the world through reinventing familiar categories like gender, race, and sexuality, categories that have their own complex geographies which are a very part of the process of reinvention. There are efforts to reconceive cities taking account of the affective, the mundane, the ongoing incompleteness, fuzziness, and unpredictability that make up urban life, and there are attempts to reimagine the economic as thoroughly infused with the cultural. Sometimes clumsy and awkward, sometimes plain inspired, these kinds of developments are surely worth following in a world too often characterized by division and despair. And follow them, this encyclopedia does.

The Encyclopedia

The *International Encyclopedia of Human Geography* is a multinational attempt to capture and trace the state of human geography as a discipline and as a description of the world as it exists today and as it changes its shape in the future. Its ambition, in other words, is to provide a major and continually updated resource that provides an always temporary but hopefully authoritative means of answering questions of the sort posed above, and many others like them. This it does by taking on the venerable format of an encyclopedia which we can understand both in its original meaning as a course of education – in this case with the description ‘human geography’ – or in its more modern meaning as treating a particular branch of knowledge comprehensively through the medium of articles arranged alphabetically, by subject.

Of course, producing an encyclopedia provides real opportunities, most especially the ability to stretch out explanations in a way not available in the more common dictionary form. Thus, our aim has not been to produce a portable, condensed summary or bite sized definitions of concepts, but rather clear, authoritative statements that set out the evolution and implications of geographic thought. Such an endeavor also produces some inevitable challenges. We want to use these challenges to explore how this encyclopedia has been put together, understanding that such challenges do not have to be understood just as undesirable or unpleasant choices between alternatives. They can equally well be understood as producing the means to fuel productive encounters.

The first challenge is an obvious one. There is an inevitable degree of arbitrariness about what is included and what is not. We have made an honest attempt to cover the whole range of what can possibly be treated as human geography in terms of issues on the ground and traditions within the discipline. This has been achieved through an intensive, iterative process involving all the editors, worked through at face to face meetings, conference phone calls, and e mail. Inevitably, there will be topics, methods, and thinkers considered important by some that we felt did not justify a stand alone entry. Moreover, we fully expect that as some issues grow or decline in importance – to the world and/or the discipline – we shall have occasion in the future to expand or contract the coverage in different fields. This dilemma of selecting what to cover brings with it other issues too. Foremost among these has been how to shape the coverage of each entry and title each entry in an informative yet pithy way. There are many bodies of conceptual and substantive knowledge that cannot easily be encompassed by a single term or phrase. The result is that some of the entries have somewhat contrived titles. Better that, though, than titles that are vague or oblique, especially in these days of voluminous information accessible across the Internet.

The second challenge is an authorial one. It was a guiding principle of this encyclopedia that we would attempt to extend authorship beyond the ‘charmed circle’ of Anglophone/Western geographers, both as a response to postcolonial critiques and as a response to the critiques of scholars from outside the Euro American zone who felt disenfranchised by what is possible to perceive as an Anglophone/Western ascendancy. We have then been attentive to the geographies of the discipline of geography, not least since scholars in diverse locales see the world differently in terms of what processes seem most important and what traditions of interpretation they use. We have not always been successful in achieving our goal of wide international authorship, partly because the geographical establishment is simply larger in some countries than others and partly because human geography still bears some of the marks of its own history, not least as a colonial enterprise. Inevitably perhaps, the historical geography of geography kept coming home to roost. Moreover, scholars from different locales had varying abilities to contribute due to issues of time, access to resources, and language. As a result, we have not been able to include some topics and perspectives. That said, entries have been solicited from 844 scholars located in over 40 countries around the world working within different traditions and we have managed to describe the different ideas and practices of geography in many countries/language groups. Again, this is a project that will be added to, over the coming years.

The third challenge was to draw on the expertise of the human geographical community in ways that ensured some degree of diversity. In particular, we are proud that we have been able to balance the voices of established scholars with those of younger writers. This has had salutary effects. For example, we have been able to trace the history of philosophical ideas in geography both from the point of view of those that have been deeply involved in the explosion of different conceptual possibilities that took place in the discipline from the late 1960s onward, as well as the point of view of younger scholars who have set out quite different agendas.

The fourth challenge was to ensure a relatively consistent standard of scholarship for each entry and to provide a balanced content with respect to ideas and geographical coverage. To that end, each entry was initially refereed by a section editor who had overall responsibility for a selection of related entries (e.g., political, urban, regional development, quantitative methods, people) and a senior editor. Each senior editor oversaw three sections in order to ensure sections were approximately commensurate in style, content, and length. To provide breadth as well as depth, authors were asked to draw on examples and traditions from around the world and not simply rely on charting an issue with respect to their own local circumstance.

The fifth challenge was to use the Internet in productive ways. One of the motivations for this encyclopedia was the production of a platform upon which we might then build a future memory for human geography. It is an attempt to begin to produce a living, breathing archive which will gradually evolve, by using the powers of the Internet. Though in this first edition, both print and Internet editions exist side by side, in later editions the Internet edition will exist singly, opening the way to a vision of geography which is in keeping with the age in which we now live. Since the articles are on the web, it will be possible to update them on a regular, rolling basis without having to wait for the revision of every other article. Equally, they will have all kinds of extra resources associated with them – illustrations of all types, including more and more videos as well as maps and diagrams, 'active' reference links that will take the reader straight to the listed journal article, and so on. Then, more articles will be added at regular intervals, both filling in gaps and supplementing existing articles. All articles will be left *in situ* – old articles will not be deleted – so that, in time, we will be able to produce a 'timeline' for many subjects, making it possible to see how thinking has evolved, thereby producing a real sense of historical accretion. Over time, we hope that the encyclopedia will become an institutionalized memory of human geography.

The sixth challenge was to produce an encyclopedia of human geography that was relevant to issues that must concern us all. We sought an encyclopedia that espoused responsibility to the planet and its people certainly. But we also became, however awkwardly and unsuccessfully, involved in thinking what that responsibility might mean. A difficult but ultimately productive instance of what this might entail – and how difficult it can be to think through – was provided by the proposed boycott of the encyclopedia by some authors concerned at Reed International's (a sister company to Elsevier) involvement in the arms trade, an affair settled when Reed withdrew from these activities. Here, we saw the same concerns being aired by many participants but radically different solutions being put on offer. We hope that the encyclopedia mirrors this diversity of response.

Bringing these six, and especially the last three, points together we hope and trust that this encyclopedia will be counted as a contribution to the global commons of knowledge. The production of a discipline depends on the goodwill of many who labor over ideas, who disseminate, review, and rework them. We are painfully aware that in the current global climate of academia, when what counts as valued academic practices is narrowing, writing authoritative and scholarly articles for publications such as this venture provides few rewards and attracts little institutional support. However, it is a vital component in the reproduction and development of a discipline. Rather than simply being understood as the creation of canonical knowledge by authorial communities we would like to thank all those who have written and reviewed material here for an activity that may be deemed a professional service but is one that, sadly, is seen as subverting the institutional priorities of a number of universities.

By Way of Conclusion

This encyclopedia is an attempt to summarize knowledge of the discipline using the encyclopedia format. In time, we hope that this encyclopedia will build into a comprehensive resource which constitutes both an archive of the discipline and, of course, of what it knows about the world. That knowledge, what facts are arraigned to support it, and what theories and methodologies are used as a means of making it credible, is, of course, a movable feast. Think of the notion of the 'world', which it is easy to demonstrate has changed its form many times over the course of history, as the development of the map shows all too well as Denis Cosgrove noted:

> There is no single map of the world, but a vast range of images that present different facets of the globe and its contents. … A map of the world is a global cultural artefact, an extraordinary human accomplishment, produced by contributions from many cultures. It is a highly sophisticated scientific achievement, and each advance in the various technologies that coalesce within it renders it more detailed, flexible, and widely available. Google Earth is certainly not the last stage in its evolution.

> Yet even [a] cursory survey of world mapping … reveals that science and technology are only a part of the story. Of equal importance are such affective aspects as imagination, faith, fear, and desire. We spot them immediately in the world maps of nonmodern cultures, and even in early examples of the modern world map. For us humans, the earth is always more than its physical form and nature; it is, indeed, a world. … Today that world seems wholly visible, even transparent. … Yet today's world maps are as hued by the contingencies of our own times as any previous ones, and we are 'mapped' into them as surely as Xiuhtechitli was mapped into the fifteenth century Mesoamerican world. Because every 'world' is social and imaginative as much as it is material, our own world maps will in due course come to seem as quaint as Jain mandalas or medieval mappae mundi seem to us today. (Denis Cosgrove, 2007: 112 113)

Or, indeed, consider the idea of the encyclopedia itself. Encyclopedias are often thought to be an invention of the West. Specifically, the general purpose, widely distributed, printed encyclopedia is usually considered to have been conceived in eighteenth century Europe with Chambers' Cyclopaedia, or Universal Dictionary of Arts and Sciences (1728), and the Encyclopédie of Diderot and D'Alembert (1751 onward). These volumes are often considered to be the first to take on a recognizably modern form, with a comprehensive list of topics, discussed in depth and organized in what we would consider a systematic way. But, once we take a broader view in which places and times are not necessarily weighted by particular prejudices, a very different account hoves into view in which all manner of authors in different parts of the world revealed an encyclopedic instinct, and from much earlier on. For example, a tradition of encyclopedia like volumes can be found in China from the eleventh century onward, and these were often massive undertakings involving as many as a thousand volumes.

The point that these two examples make is that all we can be certain of is that the contours of knowledge will constantly change. There can be no absolute certainty about what will be included in and what will be excluded from human geography in times to come. But with this encyclopedia, we believe we now have a tool that will be able to trace this evolution as it happens.

Rob Kitchin and Nigel Thrift

Further Reading

Cosgrove, D. (2007). Mapping the world. In Akerman, J. & Karrow, R. (eds.) *Maps. Finding Our Place in the World*, pp 65 158. Chicago, IL: University of Chicago Press.

A

Activism

L. M. Takahashi, University of California, Los Angeles, Los Angeles, CA, USA

Glossary

Civil Society Ordinary people, nongovernmental/nonprofit organizations, and/or businesses or private firms.

Coalitions Partnerships, alliances, and collaborations among different groups working together to change specific social, political, or economic issues.

Collective Action Activities taken on by groups that focus on group or community-level outcomes rather than individual outcomes.

Neoliberalism Political approach that assumes the market is better at managing economic activity than government, and consequently, aims at greater reliance on the market for governing and at reducing regulation and intervention by government at multiple scales.

Nongovernmental Organizations Institutions, agencies, and groups that provide services and advocate for resources and political change, but which are not formally part of local, regional, or national governments.

Political Opportunity Structures Opportunities in the current arrangement of political institutions, rules, and practices for disruption and change.

Spatial Scale Units of spatial analysis (such as local, regional, national, international) and also processes whereby social action is facilitated or disabled.

Introduction

Activism is most commonly associated with collective or group action by ordinary people, usually volunteers, who come together to change what they consider to be unacceptable or unfair circumstances. These individuals come together in a shared belief about the nature of the problem, and the means to best address the issue, using political opportunity structures to enable their influence on elected officials, policy makers, and institutions. Activism is also practiced by institutions and organizations, particularly in the nonprofit/nongovernmental and private sectors, as a means to apply economic, political, or moral pressures on communities, governments, and other organizations to stop specific practices or to encourage desired policies. Finally, geographers are also activists, both inside academic environments and in wider society.

The focus of action has spanned the political spectrum, from exclusionary efforts aimed at protecting property rights and the status quo (e.g., community opposition to homeless shelter siting in local neighborhoods) to inclusionary actions aimed at expanding access to political power (e.g., the US Civil Rights Movement); from reformist aims (e.g., redefining legal parameters to prevent hate crimes directed at specific groups) to radical goals (e.g., overthrowing or taking over government). These foci also have varying substantive aims, from very local issue based conflicts (e.g., trying to prevent evictions of renters from housing) to national level controversies (e.g., strengthening or weakening environmental regulations).

There are three primary reasons for the rising importance of activism in the twenty first century. First, governing has become less centered in formal government agencies, where the practical routines of managing political controversies, public financing, and resource allocation have moved increasingly into the realm of ordinary people, nongovernmental agencies, and private firms. Part of this trend globally is related to strong pressures from international donors and national governments for decentralizing decision making to more local levels, making communities, cities, and regions more able to make their own decisions about resource allocation, regulations, and planning for the future. Second, and related to the first trend, is the rise of civil society in public and private sector governing at the local, regional, national, and international levels. The expansion of civil society in public sector governing is due not only to the pulling back of formal government agencies in governing, but also to the increasing sense by decision makers that residents, groups, and businesses need to be involved in the design and implementation of policies to ensure their effectiveness. In the private sector, the rise of shareholder activism in corporate governance has become increasingly important since the 1980s and 1990s with the rise of large institutional investors, especially pension funds for public sector employees. And finally, activism has emerged as a challenge to larger economic and political pressures for decentralization and neoliberalism. Though the general notion of more autonomy by communities and cities by providing more responsibility for plans and policies may seem in general to be a way to make governing more responsive, there are drawbacks to this approach. The most important challenge is that resources (e.g., funding) to implement local ideas are not often transferred from central or national governments even when local responsibility is increased, and consequently, local communities and cities are being expected to use the private market, or commercial activities, as the way to deal with economic, social, and environmental challenges.

The linkages between geography and activism are twofold, and can be described as (1) the geographies of activism, and (2) activist geographies. The geographies of activism consist of different examples of activism, the distinct goals of activism, and the scales of activism. Activist geographies consider the more active links between activism and geography through research that challenges conventional ways of thinking, and through political action engaged in by geographers and other social scientists to change political practices, and social and economic systems.

Geographies of Activism

Types and Catalysts of Activism

Activism centers on a variety of different issues, can be reformist or radical in its objectives, and consists of distinct methods ranging from peaceful civil disobedience to violence. The catalysts for activism often center on visible and substantial change in social values, political debates and decision makers, economic activities, or urban land use patterns, but require awareness or knowledge of the problem and an affinity with the political position and methods of the groups that aim to change the unacceptable situation, practice, or policy. Activism responding to such catalysts has distinct goals, but often uses similar methods and processes, including mobilization and coalition building.

Land use change often sparks local activism or collective action. At the very local level, the Not In My Back Yard (NIMBY) syndrome comprises one prevalent and highly visible example. NIMBY activities usually revolve around a controversial proposed land use or facility type that is seen by residents, businesses, and/or local governments as threatening local quality of life, property values, or property rights. NIMBY is generally seen as conservative in its politics, in its attempts to preserve the status quo, and usually consists of relatively peaceful though highly vocal political acts, including petitions, demonstrations, and vocal participation in local government planning processes.

Economic shifts, such as industrial restructuring that has resulted in significant reductions in manufacturing jobs, downward pressures on wages, and reductions in benefits such as healthcare coverage, have led unions and community groups to engage in labor activism. Labor activism consists of collective efforts directed at businesses and government, such as worker strikes and organized marches, to strive for better working conditions, living wages, and expansions in benefits. With the decline in union membership nationally over the past few decades, labor activism has increasingly been spurred by community groups that are also working for affordable housing, income equality, and environmental justice rather than being solely or even primarily the purview of labor unions.

Changes in political debate or decision making bodies, such as transitions to and from majority neoconservative to liberal political parties, may also spark activism on the part of organizations and groups concerned with the domination of political debate by specific political views. Political advocacy organizations and political parties are highly active in working to influence the results of elections through public media campaigns, door to door canvassing, and debates among political candidates. Judicial activism may also be activated by political debate and public controversies, and is expressed through court decisions that enlarge or contract the powers of the government to manage public speech or activism. In other words, the courts may be activist in negative (preventing wrongs) and positive (promoting desirable outcomes) ways.

Individuals and groups that are deemed to be undesirable or socially unacceptable by the mainstream have also worked collectively to improve material living conditions, to challenge discrimination and prejudice, to expand rights and capacity to engage in public debates, to facilitate economic advancement, and to promote self determination. Such activism has been largely described as new social movements and identity politics. They are different from more classic social movements, which were largely based on worker or labor activism that focused largely on the uneven and unequal distribution of wealth. Instead, new social movements and identity politics activism include diverse efforts as the Civil Rights Movement (e.g., political struggle by the African American community to make racial segregation illegal and to challenge race based discrimination), feminism and gender equality (e.g., activism in the US to pass the Equal Rights Amendment and ongoing conflicts over Roe v. Wade, which legalized abortions in the US), immigrant rights efforts (e.g., Janitors for Justice, a largely immigrant movement striving for fair wages and better working conditions), and sexual identity activism (e.g., AIDS Coalition to Unleash Power (ACT UP), and gay and lesbian antiviolence efforts).

The interrelationships among issues driving activism lead many activist organizations and groups to form coalitions. Activist groups and organizations struggle to change existing social, political, and/or economic circumstances, but often, the issues are much larger than the individual organization's ability to change. Mobilization of ordinary people, groups, and institutions bring to bear visible support so that governing bodies may be influenced by activism, however, sometimes even greater numbers and larger presence are needed to create adequate pressure. Coalitions provide the opportunity for different groups and organizations to work together on specific issues of concern to provide a larger, more organized, and coordinated voice. Coalitions among labor groups, legal organizations, and housing nonprofit organizations, for example, have been vital in addressing the negative social impacts of gentrification, where redevelopment particularly in central city areas has resulted in rapidly escalating housing prices, often displacing low income residents and workers. Coalitions have been effective at devising strategies such as community benefits agreements that require minimum wage levels and hiring of local workers. Coalitions have also been important at the national and international levels, especially for transnational issues such as environmental degradation. Climate change, habitat protection, and deforestation have engaged scientists, environmental nongovernmental organizations, and governments to develop partnerships to measure and cope with the complex issues of urbanization, industrialization, and their negative environmental impacts.

Reformist and Radical Activism

In all of these examples, what is clear is that activism is not solely reactionary or progressive, exclusionary or inclusionary, conservative or liberal, or peaceful or violent. Both disadvantaged and wealthy activists use the language of social movements, self determination, and oppression in their efforts to change policies, procedures, and outcomes. Activism can take on different faces depending on its goals and strategies. In general, there are two basic approaches to activism: reformist and radical. Reformist activism sets out to redefine parameters within existing systems of governance, political interaction, or social practices, with the goal of incrementally changing these systems to address unacceptable or unfair circumstances, either in the processes of decision making or in the outcomes of decision making. Radical activism in contrast is focused on severe and significant change to governance, social norms, or economic systems; the goal for radical activists is structural change, where current institutions, the framework of decision making, social mores, or environmental conditions are seen as illegitimate or unacceptable.

The goals of reformist efforts emphasize change in the processes of decision making (such as expanding the number and types of participants in decision making, or including more and different kinds of stakeholders in a planning or policy design process), with some focus on adjusting outcomes associated with decision making (such as realigning funding priorities to alter resource distribution, shifting the distribution of negative impacts, or expanding the recipients of positive benefits). Examples of reformist efforts include shareholder activism (where owners of shares of corporations, especially large institutional investors that emerged as major corporate shareholders in the 1980s and 1990s, vote to change boards of directors, influence executive compensation, or alter investment patterns), neighborhood activism (where residents work together to improve local infrastructure conditions, or to expand government resources being invested in the community), judicial activism (where dissenting opinions of justices provide opportunities for questioning prevailing doctrine or legal precedents), or issue or place based activism (where organizations protest existing environmental policies, or advocate for expanded protection of sensitive habitat for endangered species). In essence, reformist activism aims at changing the processes that lead to particular emphases, beneficiaries, and distribution of costs or burdens associated with policies, economic activity, and social interactions without necessarily altering the basic or foundational structure of governance, economies, and social values. Because reformist activism does not aim at structural transformation, conflict and disagreement between activists and targeted decision makers may be avoided or minimized through strategies such as mediation, negotiation, or compensation.

Radical activism, in contrast, tends to not work within existing systems of dissent, such as through political systems' usual process of debate and decision making, because it seeks to alter the foundations underlying economic, social, and political life. Such activism is not always violent, but it does usually aim at significant and fundamentally different arrangements in distributions of wealth and resources, environmental conditions, political power, and/or social mores. Consequently, in contrast to reformist activism, the source of the problem for radical activism lies in the system itself rather than being attributable to particular individuals, procedures, regulations, or practices. In other words, rather than trying to identify specific people, rules, or issues as the source of economic, political, environmental, or social ills, radical activism sees the problem as one of systems and structures. Progressive social movements, such as the anticapitalist global movement, the environmental justice movement, and the animal rights movement, are critical of existing systems of governance, distributions of wealth and poverty, and relationships among people, economy, and the environment, and aim at fundamental shifts in the capitalist economic system, definitions of citizenship and rights within and across nation states (where citizenship is related not only to the ability to access public resources and protection, but also more broadly as full acceptance as legitimate and valued participants in society), and existing distributions of wealth and privilege. Existing systems of governance are structured so that government agencies and middle and upper income communities benefit from understanding and using existing political, economic, and social institutions. The formal and informal rules of governance create advantage for those privileged groups, to the disadvantage of low income, immigrant, low skilled, and other marginalized groups. For some individuals and groups, only complete transformation of daily life to a new arrangement of governance and social interaction will be acceptable. Radical religious fundamentalism is an example of such activism, with often violent results.

Scale of Activism

Though the types, focus, and methods of activism are critical to understanding what activism is, what is also important is understanding more clearly the role of space in activism, especially how scale affects the effectiveness and transferability of the impacts of activism from one place to other political jurisdictions. That is, activism is not just a social phenomenon, it is also a spatial one. Different types or foci of activism may be more effective or difficult because of the spatial scale defining the problem being addressed and the methods of activism being used. For example, anticapitalist activism requires that actions be conducted at spatial scales that match the global reach of capitalist forces. Consequently, though protests against global capitalist expansion take place in particular places at specific times, these protests are often directed at international organizations, such as the World Trade Organization (WTO), include coalitions of different activist organizations, and often consist of multi city protests activated around the world at the same time.

Spatial scale is a process, as well as a characteristic. In other words, space and location are not static containers of activism, but rather, are dynamic features that interact in important ways to enable or obstruct activist efforts. While technology, such as the Internet and cell phones, has reduced the importance of distance in coordinating activist efforts, physical distance, urban form, and the built environment remain significant in both facilitating and disabling activism. Urbanization patterns, the siting of specific land uses, and growth have served, for example, as catalysts for action to protect property rights, manage urban growth, and protect natural resources and sensitive habitat. Proximity to rapidly growing urban or suburban areas, the visibility of specific sites where controversial land uses and development are proposed, and the general increase in disamenities (such as traffic congestion and air pollution) are ready examples used by activist organizations to mobilize and expand participation by the populace for various purposes.

Space also defines how ordinary people, nongovernmental organizations, businesses, and governments understand and define the world around them, providing clues about what needs to be changed and how. Though the places where people live, work, and play provide the most readily available examples of the potential problems that might be influenced by activism, what is also important are the relationships between these familiar places and places that are unfamiliar, distant, and distinct. Those place relationships provide clear examples of inequality, marginalization, and oppression, and in so doing, become catalysts for action. In other words, the relational aspects of space serve to define inequality and marginalization, and in defining these inequities, space itself not only motivates action but also enables linkage among local efforts across the globe. There is a scaling up possibility to activism when this relational aspect is incorporated, providing opportunities for action to move from a grassroots, localized practice to a coordinated set of actions reliant on coalitions of individuals, communities, and organizations that take place at a transnational scale. This scaling up possibility is especially important in terms of issues that cannot be resolved in local places alone, because complex problems or issues are important at multiple spatial scales. For example, while issues such as climate change, environmental degradation, anticapitalism, and immigrant rights are important for individual groups, cities, regions, and nations, they cannot be addressed by one group, city, region, or nation alone because these issues intricately link the economies, politics, environments, and social values of varying places. Such global scale issues thus require scaling up of activism (or the transference of activism across spatial scales), as well as the coordination of activism across space (or the mutual planning and action across cities, regions, and nations) to have significant and sustainable impact.

The rise of transnational activism, focused on issues such as global climate change, citizenship, and economic growth, is linked to broader sociopolitical transitions. The global pressures for decentralization and neoliberalist policies in governance have highlighted the need for better understanding social relationships and resources available through such relationships. In contrast to activism, however, this emphasis by governments and nongovernmental agencies on what has been termed social capital (which has been seen to be a solution for a wide variety of urban problems, ranging from crime to inefficient government) emphasizes informal webs of relationships, the provision of needed monetary and other resources through these webs, and the addressing of community and regional problems through broader engagement and participation by the populace. For social capitalists, problem solving increasingly shifts from government to civil society, and communities that have more and better relationships within and outside their networks do better economically, socially, and politically across a myriad of measures. Activism, though also reliant on groups to solve problems, is fundamentally different from this social capital approach, in that actions are being taken to change the definition of problems and the ways that problems are solved.

Activist Geographies

Geographers have been activists in two primary ways. First, geographers have constructed new theories and explanations for social phenomena that have both substantially altered the ways that students and researchers understand problems, and the ways that academia evaluates research. Geographers, and other social scientists, have challenged the status quo in scholarly approaches, methods, and substantive areas by offering conceptual frameworks that have highlighted the structural dimensions of social life (e.g., Marxist geographies), empowered the participants in research (e.g., feminist geographies), and shifted the topics of research to socially relevant issues (e.g., critical praxis). Second, geographers have engaged in political action aiming at redefining the nature of problems, appropriate processes to address these problems, and alternative visions for the future. For activist geographers, scholarship, teaching, and political action are all necessary for a critical praxis that fundamentally challenges imbalances in power relations and leads to significant social change.

In developing new and innovative ways of understanding social life, politics, and economies, geographers have been on the forefront of highlighting social and spatial issues that have been left out of research deemed classic by scholars. In emphasizing and studying the plight of marginalized groups and communities, for example, women, people of color, disabled and impoverished populations, immigrants, and others, geographers and other social scientists studying such issues have not only brought new information to students and researchers, but have also expanded the ways in which such issues are studied by promoting new theories of oppression, marginalization, and justice. Many have also changed the way that issues are studied by promoting participatory action research, where the participants and academics together define the research questions and appropriate research methods, and together interpret the results of the research conducted. Such geographers, and other social scientists, have used their research findings to promote public debate, to influence public policy and governance, and to realign public perceptions about poverty, oppression, and marginalization to help bring about social transformation. They have strived to open public arenas of political debate and decision making to those who have been shut out of processes that directly affect their lives to promote a more just global society.

For activist geographers, theory and practice go hand in hand, not only providing alternative approaches in the classroom and among academics, but also bringing these approaches to bear in the political realm, advocating for justice, equity, and improved quality of life. For activist geographers, research and teaching alone are not sufficient for a critical praxis. Political action is also a central component, where research and teaching inform political action, and conversely, where activism defines research concepts, methods, and topics.

There are multiple obstacles to such a critical praxis in the contemporary academic environment, because of broader social, economic, and political pressures brought by neoliberalism and globalization. Two obstacles are especially important, and have affected academic institutions across the globe. First, there are rising pressures in colleges and universities to engage in applied research that is seen as beneficial by government agencies and businesses. Such applied research tends to support the current structure of governance and industry, striving for greater efficiencies in policy design and implementation. This focus on efficiencies clearly matches with neoliberalist governance approaches, tending to reinforce existing relations of power and privilege. Second, the expanding use of contract, part time, and nontenure track lecturers and instructors in universities and colleges creates challenges for activist geographers seeking to link their research and political action. The lack of tenured faculty positions results in a higher risk environment for those activist researchers engaging in politically unpopular actions in academic and/or public contexts because nontenure positions mean that activist geographers may lose their teaching positions if they engage in controversial actions. Moreover, academic institutions, by enlarging cadres of temporary and nontenure faculty, create groups of instructors who must take on multiple teaching positions at different institutions to obtain living wages. Consequently, such instructors have little time to engage in nonpaid labor associated with political action.

Even with these obstacles, geographers continue to make significant contributions to scholarly understanding of activism, and to provide clear and compelling linkages to researchers, students, communities, policy makers and elected officials, and broader society between scholarly work and political action.

See also: Civil Society; Communist and Post-Communist Geographies; Environmental Justice; Identity Politics; Neoliberalism; NIMBY; People's Geography; Resistance.

Further Reading

Alinsky, S. D. (1947). *Reville for radicals*. Chicago, IL: University of Chicago Press.

Kitchin, R. M. and Hubbard, P. J. (1999). Research, action and 'critical' geographies. *Area* 31(3), 195 198.

Kobayashi, A. (1994). Coloring the field: Gender, 'race', and the politics of fieldwork. *Professional Geographer* 46(1), 73 80.

Lake, R. W. (1994). Negotiating local autonomy. *Political Geography* 13(5), 423 442.

Mitchell, D. (2003). *The right to the city: Social justice and the fight for public space*. New York: Guilford.

Piven, F. F. and Richard, C. (1977). *Poor people's movements*. New York: Pantheon.

Pulido, L. (2006). *Black, brown, yellow, and left: Radical activism in Los Angeles*. Berkeley, CA: University of California Press.

Taylor, P. J. and Frederick, H. B. (1992). How do we know we have global environmental problems? Science and the globalization of environmental discourse. *GeoForum* 23(3), 405 416.

Relevant Websites

http://www.aclu.org
American Civil Liberties Union.

http://acorn.org
Association of Community Organizations for Reform

Activist Geographies

P. Routledge, University of Glasgow, Glasgow, UK

Glossary

Activist Geographies Geographies concerned with action, reflection, and empowerment in order to challenge oppressive power relations.

Advocacy Geography An approach to research and teaching developed by William Bunge and co-workers demanding that geographers are responsive to the needs of, and advocates on behalf of, poor, deprived communities.

Participatory Geographies Geographical inquiries marked by the embodied participation of researchers in the lifeworlds of their research subjects, and/or participation of those research subjects in the production of geographical research.

Introduction

Activist geographies are concerned with action, reflection, and empowerment (of oneself and others) in order to challenge oppressive power relations. As a fusion of theory and practice (praxis), activist geographers seek to forge mutual solidarity with resisting others – communities, groups, social movements, or nongovernment organizations who are challenging various practices of dominating power – through critical collaboration. The objectives can be varied: the creation of participatory spaces of action which are inclusive and antihierarchical; the nurturing of creative interaction with resisting others independent of electoral politics; and conducting direct action whereby 'we' devise a plan to do something, in collaboration with others, and then do it without recourse to external authorities. For an activist geography, or a human geography with an 'action oriented research' agenda, as Rachel Pain described it in 2003, 'relevance' and 'criticality' entail making certain political commitments or commitments to certain kinds of action. Here, commitment is to a moral and political philosophy of social justice, and research is directed both toward conforming to that commitment and toward helping to realize the values that lie at its root.

While the 'anarchist geography' of Prince Peter Kropotkin hinted at a nineteenth century version of activist geographies, committed to dispelling the territorial bases for unjust power relations associated with 'rulers' of all kinds, it is only relatively recently in the history of human geography that such an approach has become at all widespread. In the late 1960s and early 1970s, many geographers became deeply concerned about the seeming 'irrelevance' of the discipline to the multiple problems facing global society, whether environmental or to do with decaying cities, social unrest, Cold War threats and machinations, and so on. While some, often the spatial scientists like Brian Berry who had come to prominence in the 1960s, concluded that the discipline needed to become more policy relevant, working with and on the 'levers' of state and corporate power, others like David Harvey hoped for a more fundamental assault on all forms of 'establishments', whether in the academy (including standard theories and methods) or in wider society (government, big business, sedimented structures of class inequalities, etc.). Harvey argued for the development of historical materialist 'people's geographies'

> that open[ed] the way to the creation of new forms of society in which common people have the power to create their own geography and history in the image of liberty and mutual respect of opposed interests. (Staeheli and Mitchell, 2005: 359)

However, a cleavage of sorts swiftly emerged between those radical geographers who, in all sorts of ways, often brilliantly and compellingly, turned to Marxist and related philosophies and social theories to inform their critical scholarship, and a smaller banding of scholars who – while often alert to broader conceptual concerns – wished to apply themselves much more immediately and directly to help alleviate social injustices 'on the streets', 'on the reservations', and in diverse other spaces where disadvantaged peoples of all kinds were struggling to survive day by day.

Perhaps the earliest example of deeply engaged involvement 'on the streets', what might be taken as the real birthplace of activist geographies akin to the characterization above, was the 'advocacy geography' of William Bunge, practiced in the so called 'geographical expedition' to poor black inner city neighborhood of Fitzgerald in the US city of Detroit in the late 1960s and early 1970s. Here, a handful of professional geographers and some of their students effectively left the campus to work for the residents (the resisting others) of this multiply oppressed place. In this remarkable experiment, residents called upon Bunge and his co workers to seriously take up subject matters such as ghetto formation and the actions of unscrupulous slum landlords, and more particularly, they asked for assistance in preventing their children from being victims of traffic accidents or getting ill as a result of rat bites in vermin infested housing schemes. These early activist geographers duly researched the relevant patterns and processes, putting their skills in data acquisition, processing, and mapwork to the service of their local hosts, notably when acting as 'advocates' for them in dealings with the municipal authorities (who were prepared to listen to the academics, whereas they seemingly remained deaf to the residents themselves). However, this experiment did not last for long, as the University of Detroit, disturbed by this emerging constellation of academia and activism, put blocks in the way which eventually led to the departure of Bunge from the institution, and to his colleagues and their students returning from the streets to the library. Bunge himself became involved in trying to establish something equivalent in Toronto, while parallel moves briefly fluttered in other cities like London, but by the mid 1970s the promise of advocacy geography had largely petered out, albeit still being remembered today as a beacon of alternative possibilities and arguably an important precursor of more recent attempts at marrying activism with the geographical academy.

Toward More Fully Activist Geographies

Since the 1970s, Marxist, feminist, and, more broadly, radical and critical geographers have produced outstanding scholarship and theorizing, substantively researching sociospatially constituted oppressions, inequalities, and injustices. Yet, there has often remained an uneasiness about the relative disconnect between such work and the potential of the discipline to contribute genuine activist interventions aiming to right the wrongs of such oppressions and the like as grounded in particular in peoples, places, and issues. Moreover, as Lynn Staeheli and Don Mitchell note in their 2005 paper, while theoretical developments in the discipline have perhaps drawn geography closer to the heart of the main debates in the social sciences and humanities, they may have simultaneously drawn it further from the social movements, political formations, policymakers, and lay people that many geographers hope to reach. In a similar vein, writing a few years earlier, Noel Castree had worried about the dangers of 'domesticating critical geography', such that critical scholarship actually becomes a new trendy currency of the discipline, rewarded in the likes of research quality assessment exercises, but strangely divorced from real contexts of deprivation, struggle, and the need for change. As a result, and dating back at least to Nick Blomley's 1994 provocation, some geographers have lamented the separation between critical sectors of the discipline and activism going on 'outside' the academy. Hence, calls have been heard for geographers of a radical or critical persuasion to become politically engaged outside the academy in social movements, community groups, protests, and other moments where resisting others seek to take control of their own destinies.

One of the clearest articulations has been by Duncan Fuller and Rob Kitchin, in introducing their 2004 edited e volume on *Critical Theory/Radical Praxis*, who argued that academics have a social responsibility, given their training, access to information, and freedom of expression, to make a difference on the ground through activist research. Drawing upon feminist praxis, they see the role of the academic as primarily that of an enabler or facilitator, acting in collaboration with diverse communities. Radical activist praxis is thus committed to exposing the sociospatial processes that (re)produce inequalities between people and places; challenging and changing those inequalities; and then bridging the divide between theorization and praxis. Fuller and Kitchin bemoan the fact that there is still some scholarly distance between geographers' activism and their teaching, research, and publishing activities, with radical activist praxis consisting of little else beyond academic writing and pedagogy (not that the possibilities of these activities, critical pedagogy especially, should be neglected). They posit that the structural constraints arising from the desire to maintain the power of the academy in knowledge production, and the desire to shape the education system for the purposes of the neoliberal *status quo*, inevitably but debilitatingly work to (de)limit the work of radical and critical geographers.

Thus, a variety of activist geographies (not always explicitly named as such) have begun to appear. These have been practiced by various geographers operating from different institutional bases and with varying levels of involvement in different worldly situations of social injustice (in both the Global North and the Global South). Such initiatives include, but are not limited to, 'people's geography', as embodied in the People's Geography Project organized out of the University of Syracuse, with its effort at making research and geographical concepts relevant to social struggles 'for' (and to an extent 'by') the people, for example, through the Syracuse Hunger Project; 'critical development work', marked by the efforts to construct venues for community or organizational input on development and planning work, positioned by R. Howitt in 1993 as 'applied people's geography'; and 'autonomous geographies', as outlined by Jenny Pickerill and Paul Chatterton in 2006, not only studying but also actively seeking to contribute to the building of "spaces where there is a desire to constitute non capitalist, collective forms of politics, identity and citizenship" (p. 730).

Particular reference might also be made to 'participatory geographies', an emerging body of work centralizing the importance of geographical researchers more or less fully participating in the lifeworlds of research subjects, and/or the participation of those research subjects in the production of geographical research. In this work, questions about the researcher's entangled responsibilities to these research subjects are always foregrounded (and also discussed further on). Bunge's advocacy geography had urged this participatory moment, but so too, in a less overtly politicized manner, have various inputs to the discipline emerging from the ethnographic sensibilities of anthropologists and (some) sociologists. Coupled with a diversity of other conceptual, methodological, and ethico political impulses, not least derived from feminist and postcolonial literatures and practices, a trajectory of engagement (on occasion named as participatory geographies) can now be identified, leading to the major book on this theme in 2008 edited by Sara Kindon, Rachel Pain, and Mike Kesby.

The extent to which participatory and activist geographies converge or diverge is an important question, while there are clear overlaps in the geographers involved, and clearly many shared ethics and politicized senses of responsibility, there is arguably a distinction in the extent to which activist geographies commit to an overt and actively engaged politics of social justice (which also commonly if not always marks out the groupings with whom they elect to 'participate'). In other words, for some participatory geographers, it may be enough just to participate, with the danger of 'going native' and so closely empathizing with the peoples and places under study, being that the possibility of a critical perspective is largely eroded. For activist geographers, the challenge is to participate in enabling transformative social change. However, while sharing in others' everyday worlds, an onus remains to be alert to when social injustices – maybe practices of dominating power – arise within the communities hosting the participant researcher, particularly when such injustices mirror, or import, precisely those wider socio spatial processes of oppression that the activist researcher must remain committed to oppose (whenever and wherever they appear, however uncomfortable that might then render their immediate research relations).

Activist Methodologies, Collaborations, and Everyday Life

This eruption of activist and related geographies within the discipline is taking place at a time in which there is increased social mobilization and conflict around the world. Issues such as globalizing capitalism, trade agreements, failed development, neoliberalism, and the war on terrorism have been met by opposition ranging from the Zapatista rebellion in Chiapas, itself both locally played out and stretching its influence and effects worldwide, to what Paul Routledge and Andrew Cumbers term 'global justice networks' articulated through global days of action, social forums, and violent insurgencies (e.g., in Iraq). Hence, activism cannot be simply bounded off from other aspects of everyday life; our lives are always entwined with the lives of others – through the legacies of colonialism, through flows of capital and commodities, through modern telecommunications – which demand that we as academics become politically sensitive to the needs and rights of what Stuart Corbridge termed 'distant strangers' in his influential 1993 paper.

Activist geographies require a broad interpretation of activism. Not least because activist practices have frequently required the construction of allies and enemies and the negative identification of outsiders, thereby creating divisions rather than connections between diverse communities, as noted by Chatterton in his 2006 paper. Moreover, Ian Maxey, writing a few years earlier, had argued that activism is discursively produced within a range of sites, including the media, grassroots organizations, and academia, and this has led to a restrictive view of activism that emphasizes dramatic, physical, and even 'macho' forms of action. However, Maxey argues that the social world is produced through everyday acts and thoughts in which we all engage, and he therefore understands activism to be the process of reflecting and acting upon exactly this condition. This means that everybody is an activist, engaged in some way in producing the world, and reflexivity enables people to place themselves actively within this process and for geographers thereby to become overtly activist:

> By actively and critically reflecting on the world and our place in within it, we are more able to act in creative, constructive ways that challenge oppressive power relations rather than reinforce them. (Maxey, 1999: 201)

Such an interpretation opens up the field of activism to everybody, and serves to entangle the worlds of academia and activism, as Paul Routledge has insisted in his 1996 discussion of the 'third spaces' between academia and activism. There are no preconceptions about the forms that such an engagement might take; and the importance of this inclusive definition is that it is a palliative to the privileging of certain forms of activism over others, and to the exclusivist and domineering tendencies that occur in certain activist discourses. Indeed, a relational understanding of the activism that stresses connection and negotiation between diverse (activist, academic, and lay) communities is required to enable what Paul Chatterton terms "learning, acting and talking together on uncommon ground" (2006: 277).

As a minimal requirement, nonetheless, activist geographies must always stress the inseparability of knowledge and action, impelling them to be self consciously interventionist in approach. For a critical geography to be fully critical, one must therefore 'walk the walk' as well as 'talk the talk', that is, one must be critically engaged in some way with those others with whom one is conducting research. Such an approach grants no special privileges to the researcher and acknowledges that our understanding of the world as it is and our actions to achieve a world as it 'ought' to be are inseparable. Through such a broad approach, and through a variety of possible engagements, geography can be made relevant to 'real world concerns', producing a very different version of 'relevance' than that pedalled by the spatial scientists in the 1970s 'relevance debate' (as mentioned above) and their inheritors in more recent exchanges.

Activist methodologies are conceptualized with an eye to both communication and emancipation, confronting, and seeking solutions to, issues of social, economic, and environmental injustice. They are conceptualized and carried out in collaboration with activist (and sometimes academic) others – the precise contours of such collaboration being worked through in cooperation with those others, and including the possibility that some research will actually not be participatory at all, but carried out 'covertly' with certain powerful groupings in order to generate findings useful to resisting others. Such methodologies can be written about in a variety of ways in both academic and nonacademic forums, the extent of authorial distance from one's subject matter being a matter of personal choice rather than academic protocol. Critical engagement with issues of (in)justice can invest such narratives with political and emotional power, however, precisely because the activist researcher is writing from within a particular issue or struggle. Critical engagement also implies an active process of attempting to bring about specific outcomes (in practices, policies, etc.), and, while such involvement can raise concerns about the 'bias' of the researcher, particularly when viewed through standard protocols of academic knowledge production, this risk can be ameliorated through a critical reflexivity on behalf of the researcher.

Because the personal is political and relational, an activist geography implies a commitment to deconstruct at least some of the barriers that exist between academics and the lives of the people they professes to represent, so that scholarly work interprets and effects social change, again as had been the ideal for the advocacy geographers of the late 1960s and early 1970s. Critical collaborative engagement with resisting others must recognize that academics are entangled within broader powers of association and intellectual production – with the institutions that employ us and/or fund our research, and their location within a global hierarchy that privileges the West's economic systems, institutions, and policy 'experts' at the expense of those of the rest of the world. Such associations grant 'us' certain securities and advantages – for example, economic, political, and representational – that may not be enjoyed by those with whom we collaborate. Hence, academics frequently enjoy a range of privileges that may include mobility, funding, class, ethnicity, gender, and nationality.

Activist geographies seek to make these privileges work in political ways that attempt to effect social, environmental, and political change. This implies that the 'field' of our fieldwork becomes "located and defined in terms of specific political objectives" which "ideally work toward critical and liberatory ends" (Nast, 1994: 57). It further requires an ethics of struggle to be developed within academia, one that is 'with' resisting others as well as 'for' them, and that accepts moral and political responsibility as a crucial act of self constitution. As Doreen Massey notes, such responsibility is relational, embodied, and implies extension (i.e., not necessarily being restricted to the immediate or the very local). Taking on board such geographies of responsibility involves at least seven processes.

Responsible Activist Geographies

First, the motivation of activist geographies is to develop practices aimed at social transformation rather than merely the 'production of knowledge' and/or the 'solving' of 'local' problems. Activist research moves beyond the acquisition, cataloging, ordering, and the publishing of information toward jointly producing knowledge with resisting others to produce critical interpretations and readings of the world which are accessible, understandable to all those involved, and actionable, as Paul Chatterton *et al.* argue.

Second, activist geographies nurture a politics of affinity with others. The connections and solidarities forged with resisting others is a key part of activist research, implying a common identification of problems and desires among groups or individuals committed to social change, as well as a desire to work together to confront and reverse a set of issues which have a common effect on all the people concerned. Practically, affinity consists of a group of people who share common ground (friends, lovers, shared beliefs and dreams, etc.), and who can provide supportive, sympathetic spaces for its members to articulate, listen to one another, share concerns, emotions, hopes, and fears. The politics of affinity enables people to provide support and solidarity for one another. Ideally, such a politics of research should be built on consensual decision making, which is non hierarchical and participatory, embodying flexible and fluid modes of action. The common values and beliefs articulated within the politics of affinity constitute a 'structure of feeling' resting upon collective experiences and interpretations, which are cooperative rather than competitive, all predicated upon taking political action. The idea of consensus here is based upon the notion of mutual solidarity, constructing the grievances and aspirations of geographically and culturally different people as always interlinked – a key claim of Thomas Olesen's work on 'international Zapatismo' in 2005. Mutual solidarity enables connections to be drawn that extend beyond the local and particular, by recognizing and respecting differences between people, while at the same time recognizing similarities (e.g., in people's aspirations). It is also about imagining global subjectivities through similarities of experience, recognizing the shared opportunities, and techniques of struggle. Putting mutual solidarity into practice means co producing contextually relevant knowledges which are useful and accessible to groups in their struggles, perhaps taking the form of pamphlets, guides, and websites, which may be more readily used and understood by the general public.

Third, activist geographers must be attentive to, and negotiate, problematic power relations that exist between (research) collaborators. It is crucial to theorize and negotiate both the differences in power between collaborators and the connections forged through such collaboration, and it is important to note that these differences in power are almost always diverse and entangled. Collaborations between activist geographers and others are neither relationships of difference articulated through an objectifying distance, nor relationships of sameness based upon entirely commensurate backgrounds, interests, and ambitions, since "situating knowledge through [hypothetical] transparent reflexivity gives no space to understanding across difference" (Rose, 1997: 313). Rather than 'map' distance and difference between distinctly separate agents, activist geographers ask how difference – for instance, regarding power relations – is constituted, tracing its destabilizing emergence during the research process itself. Indeed, the activist, researcher, and 'public' are all essentialisms which obscure common agendas and links which could emerge between collaborators and negates more hybrid senses of self. What Henry Giroux called 'border pedagogy' in 1992 eschews fixed notions of 'us' and 'them', recognizing the many ruptures between groups, and embracing and questioning differences and newness.

Negotiating power relations means working with groups to uncover structures of power to empower people to take control of their own lives. The pedagogical project of the Brazilian educator, Paulo Freire, developed in the 1970s, has insisted on the dialectical relationship between the oppressor and the oppressed. It is only through this dialectical relationship that we can unpack relationships and causalities which structure injustice. Through this we can acknowledge that there is oppression and inequality, not merely external oppression of the 'other', but that 'we' too are subject to oppression and in turn subject others to it, thereby reproducing it. The double movement here requires both recognizing our own role in perpetuating inequality and injustices and larger examples of systematic oppression, as also argued, albeit from the differing coordinates of liberatory and post structurally inflected Christian theology, by Paul Cloke in his diagnosis for "living ethically and acting politically in human geography." However, in the collaborative politics of affinity, power accrues to different people at different times, depending on the context. Activist academics are frequently in a position of power by virtue of their ability to name the categories, control information about the research agenda, define interventions, and indeed to 'come and go' as so called research scientists. Yet, because power circulates, in difference and in unity, in relations of empowerment and disempowerment, there are still certain 'powers' that accrue to research subjects within the research process, ones that might empower as well as disempower the activist academic. Just as "we need to listen, contextualise, and admit to the power we bring to bear as multiply positioned authors in the research process" (Nast, 1994: 59), so we also need to be attentive to the power that our collaborators bring to the research process.

Through the conduct of action with others, bonds of association are forged that are crucial to the creation of common ground. Such considerations are intimately entwined with what Laura Pulido called 'the interior life of politics:' the entanglement of the emotions, psychological development, souls, passions, and minds of activist academic collaborators. An 'engagement with emotion' is thus the fourth process of activist geographies, not least because transformative encounters based on solidarity often come from our deep emotional responses to the world. Politically, emotions are intimately bound up with power relations and also with relations of affinity, and are a means of initiating action. We become politically active because we feel something profoundly, such as social injustice or ecological destruction. This emotion triggers changes in us that motivates us to engage in politics, and it is our ability to transform our feelings about the world into actions that inspires us to participate in political action. Affinity with others under such conditions creates intensive encounters, what Hakim Bey termed as 'seizure of presence', wherein practical politics – embodied, intersubjective, and relational – are practiced. Collaborative association necessitates interaction with activist others, through the doing of particular actions and the experiencing of personal and collective emotions, through creativity and imagination, and through embodied, relational practices that produce political effects.

Fifth, activist geographies attempt to engage in prefigurative action: that is, to embody visions of transformation as if they are already achieved, thereby calling them into being. There are many examples of postcapitalist ways of living already part of the present, as J. K. Gibson Graham have argued in their claims about building new postcapitalist ontologies, providing a spur for activist geographies to contribute to the creation of other realities, what Harvey terms 'spaces of hope'. Moreover, collaborations are not just about action in the research process, but also about how the research process can contribute to wider activism like protests, demonstrations, events, etc. In a similar vein, and sixth process, activist geographies attempt to create spaces for action: that is, physical spaces that can be created and occupied in building commonality and connection between different groups unmediated by consumer relations or profit, those autonomous geographies discussed by Jenny Pickerill and Paul Chatterton. These common places seek opportunities for transformative dialog, mutual learning, and (creative) conflict. They form participatory spaces for building modes of understanding, encounter, and action, which are inclusive, nurture creative interaction with others independent of electoral politics, and can lead to critical reflection and interventions.

Finally, activist geographies are concerned with relational ethics. Ethical considerations raise crucial questions concerning the role played by concepts of social (in)justice in geographical research, and the extent to which ethical conduct is desirable, definable, and/or enforceable in the practice of geography. Laura Pulido suggests there are three benefits to cultivating a dialog on ethics in political activism. First, we cultivate relations of honesty, truth, and interpersonal acknowledgement, and so it is important for activist academics to be open about their 'dual' positionalities of activist and academic when collaborating with others. Second, it allows us to build a genuine moral language, in which case Pulido argues that historically the Left has settled for making arguments based on policy, fiscal analyses, legal precedents, and history to the almost complete exclusion of ethics. Third, it contributes to us becoming more fully conscious human beings, in that, while political consciousness is distinguished by its focus on structures, practices, and social relations of societal and global power, self consciousness refers to self knowledge, including the understanding of one's past and present; one's motivations, desires, fears, and needs; and one's relationship with the larger world – hence nurturing a politics of both affinity and relationality.

The ethics of activist geography need to be relational and contextual, a product of reciprocity between collaborators, and negotiated in practice. Relational ethical positionalities need to be for dignity, self determination, and empowerment, acknowledging that any collaborative 'we' constitutes the performance of multiple lived worlds and an entangled web of power relationships. Collaboration can enable what J. K. Gibson Graham termed in 1994 a 'partial identification' between ourselves and resisting others, as well as an articulation of a temporary common ground wherein relations of difference and power (concerning gender, age, ethnicity, class, sexuality, etc.) are negotiated across distances of culture, space, and positionality in the search for mutual understanding. A relational ethics is based on the notion of difference in relation, constituted in an intersubjective manner where difference is neither denied, essentialized, nor exoticized, but rather engaged with in an enabling and potentially transformative way. A relational ethics is also about 'decolonizing' oneself, getting used to not being the expert, and it is about mutual solidarity through the process of mutual discovery and knowing one another. A relational ethics is attentive to the social context of collaboration and our situatedness with respect to that context, and it is enacted in a material, embodied way; for example, through relations of friendship, solidarity, and empathy.

Activist geographies seek to embrace a politics of recognition that identifies and defends only those differences that can be coherently combined with social and environmental justice. Hence, when the original proposal for this encyclopaedia was mooted, many activist geographers who were asked to contribute boycotted the project. This was because the publisher, Elsevier, was owned by a corporation, Reed International, which was actively involved in organizing arms trade fairs. It was only when Reed International agreed to end its involvement with the arms trade that some of those geographers who had boycotted the project finally agreed to participate.

Activism and Academia: Contradictory Spaces?

Many of the remarks above gesture to the complications of working in the 'third space' between activism and academia, not least because the academy is traditionally suspicious of highly engaged, participatory, and transparently ethico politically 'committed' inquiry, but also sometimes because activists are (often justifiably) suspicious of what these academics are 'really' after (ultimately career boosting or wishing to make a 'real' contribution?). Pursuing activist methodologies may entail problems and privations on behalf of (young) academics, for example, through delays in promotion, or failure to gain tenure, and such difficulties may be particularly acute within the increasingly neoliberal university system. However, being actively involved outside of the academy in order to effect some form of social change can be both a rewarding personal experience as well as bringing potential insights into one's academic work. There are many examples of geographers who have become involved, in different ways, ethico politically, and who have discussed the hazards and potentials of such engagement; notably the contributors to the 'Research, action and 'critical' geogaphies' theme issue of *Area* in 1999 and the recent e volume edited by Fuller and Kitchin (mentioned earlier).

Activist desires sit in an uneasy relationship to that academic desire to conduct research and to be published. The activist and the academic are usually identities performed on different stages and for different audiences, which necessitates a careful negotiation of the different powers at work in the collaboration process – 'our' own power as activist academics, the entangled relations of power between us and our collaborators, and the limitations, possibilities, and deployments of these vari ous powers. Within these entangled powers the import ance of reflexivity cannot be underestimated, since we miss an important dimension of self reflexivity if we neglect to examine our own capacity to facilitate change in the lives of others in addition to merely rep resenting them. Although critical collaboration enables academics potentially to play a greater role in effecting social change, the entangled powers that are entailed sit uneasily together, and thus we need to acknowledge an embodied politics that is committed, yet partial in per spective. Creating common ground with resisting others serves to highlight and 'ground' differences (in language, ethnicity, power, access to resources, etc.) in particular ways in particular places. When placed in such active proximity, difference (in ways of being, talking, and act ing) can be both recognized and negotiated.

Activist geographic praxis acknowledges the import ance of letting myriad flowers bloom in the academic imagination, while recognizing the importance of tem pering the privilege of academics to pursue personal research agendas with an ethics of political responsibility to resisting others. Such responsibility extends beyond teaching within the academy (to relatively privileged students), in order for academic research to "make its deliberations more consequential for the poorer eighty per cent of the population of the world" (Appadurai, 2000: 3). Activist geographies are concerned with grounding our theories and imaginations in the messy unpredictability of everyday life, and embodying radic alism that situates the geographical and academic im agination within the key political debates and actions of our time – thereby always politicizing the personal. For example, we might be actively resisting rather than be moaning or complying with the neoliberal restructuring of academia; or we might be practically living alternative geographical possibilities in addition to theorizing about the radicalizing of geography to engage actively with 'the wretched of the Earth'. In particular, activist geographies can contribute to the creation of new political spaces and potentially new forms of collaborative power. Myriad forms of engagement are possible: as activist academics, we can reclaim streets, blockade military bases, and barricade corporate offices; occupy land, warehouses, and unoccupied houses; use our skills in popular health and education outreach; participate in social centers, infoshops, guerrilla gardens, and independent media initiatives; culture jam and decommodify corporate/ private space; contribute to direct democracy and autonomist experiments in everyday living through practising consensus based politics; engage in the long term work of contributing to, and constructing, so cially just and environmentally sustainable communities. Ultimately, activist geographies prioritizes grounded, embodied political action, the role of theory here being to contribute to, be informed by, and be grounded in such action, in order to create and nurture mutual solidarity and collective action, a liberatory politics of affinity.

See also: Anarchism/Anarchist Geography; Applied Geography; Feminism/Feminist Geography; Field Geographies; Marxism/Marxist Geography I; Marxism/ Marxist Geography II; Participatory Action Research; People's Geography; Spatial Science.

Further Reading

Appadurai, A. (2000). Grassroots globalization and the research imagination. *Public Culture* 12, 1 19.

Bey, H. (1991). *T.A.Z.* Brooklyn, NY: Autonomedia.

Blomley, N. (1994). Activism and the academy. *Environment and Planning D: Society and Space* 12, 383 385.

Bunge, W. (1971). *Fitzgerald: Geography of a Revolution*. Cambridge, MA: Shenkman.

Castree, N. (1999). 'Out there'? 'In here'? Domesticating critical geography. *Area* 31, 81 86.

Chatterton, P. (2006). 'Give up activism' and change the world in unknown ways: Or, learning to walk with others on uncommon ground. *Antipode* 38, 259 281.

Chatterton, P., Fuller, D. and Routledge, P. (2008). Relating action to activism: Theoretical and methodological reflections. In Kindon, S., Pain, R. & Kesby, M. (eds.) *Participatory Action Research Approaches and Methods: Connecting People, Participation and Place*, pp 216 222. London: Routledge.

Cloke, P. (2002). Deliver us from evil? Prospects for living ethically and acting politically in human geography. *Progress in Human Geography* 26, 587 604.

Corbridge, S. (1993). Marxisms, modernities and moralities: Development praxis and the claims of distant strangers. *Environment and Planning D: Society and Space* 11, 449 472.

Cumbers, A. and Routledge, P. (2004). Alternative geographical imaginations: Introduction. *Antipode* 36, 818 828.

Friere, P. (1971). *Pedagogy of the Oppressed*. New York: Herder and Herder.

Fuller, D. and Kitchin, R. (eds.) (2004). *Critical Theory/Radical Praxis: Making a Difference beyond the Academy?* Victoria, BC: Praxis (e)Press.

Gibson Graham, J. K. (1994). Stuffed if I know! Reflections on postmodern feminist social research. *Gender, Place and Culture* 1, 205 224.

Gibson Graham, J. K. (2006). *A Postcapitalist Politics*. Minneapolis, MN: University of Minnesota Press.
Giroux, H. (1992). *Border Crossings*. London: Routledge.
Harvey, D. (2000). *Spaces of Hope*. Berkeley, CA: University of California Press.
Howwit, R. (1993). Social impact assessment as 'applied people's geography'. *Australian Geographical Studies* 31, 127 140.
Kindon, S., Pain, R. and Kesby, M. (eds.) (2008). *Participatory Action Research Approaches and Methods: Connecting People, Participation and Place*. London: Routledge.
Kitchin, R. and Hubbard, P. (1999). Editorial: Research, action and 'critical' geographies. *Area* 31, 195 198.
Massey, D. (2004). Geographies of responsibility. *Geografiska Annaler* 86B, 5 18.
Maxey, I. (1999). Beyond boundaries? Activism, academia, reflexivity and research. *Area* 31, 199 208.
Nast, H. (1994). Opening remarks on 'women in the field'. *The Professional Geographer* 46, 54 66.
Olesen, T. (2005). *International Zapatismo*. London: Zed Books.
Pain, R. (2003). Social geography: On action oriented research. *Progress in Human Geography* 7, 649 657.
Pickerill, J. and Chatterton, P. (2006). Notes towards autonomous geographies: Creation, resistance and self management as survival tactics. *Progress in Human Geography* 30, 730 746.
Pulido, L. (2003). The interior life of politics. *Ethics, Place and Environment* 6, 46 52.
Rose, G. (1997). Situating knowledges: Positionality, reflexivities and other tactics. *Progress in Human Geography* 21, 305 320.
Routledge, P. (1996). The third space as critical engagement. *Antipode* 28, 397 419.
Routledge, P. (2002). Travelling east as Walter Kurtz: Identity, performance and collaboration in Goa, India. *Environment and Planning D: Society and Space* 20, 477 498.
Routledge, P. and Cumbers, A. (2009). *Global Justice Networks: Geographies of Transnational Solidarity*. Manchester: Manchester University Press.
Staeheli, L. A. and Mitchell, D. (2005). The complex politics of relevance in geography. *Annals of the Association of American Geographers* 95, 357 372.

Relevant Websites

http://www.peoplesgeography.org
Peoples geography.

Actor-Network Theory/Network Geographies

G. T. Jóhannesson, University of Iceland, Reykjavik, Iceland
J. O. Bærenholdt, Roskilde University, Roskilde, Denmark

Actor-Network Theory

Material Relationalism

The origins of actor network theory (ANT) can be traced to science and technology studies (STS) and sociology of scientific knowledge (SSK) in the early 1980s. The approach has from the start been especially associated with the figures of Bruno Latour, Michel Callon, and John Law, who are much inspired by the French philosophers Michel Foucault, Michel Serres, Gilles Deluze, and Felix Guttari. ANT has spread to different areas of the social sciences and has, since the mid and late 1990s, increasingly been felt in some of the subdisciplines of human geography.

From the outset, the central theme of ANT has been the emergence of societal order. Thus many ANT studies have revolved around questions of how order is accom plished and made stable in time and space. In general, ANT can be described as a methodological device based on a particular worldview which aims at tracing the practices through which society is assembled. ANT can thus be framed as a practice based perspective, but compared to other such approaches it stands out due to its grounding in material relationalism or material semiotics, clearly expressed by the principle of 'general symmetry'. This is a methodological principle that states that researchers should refute all pre given distinctions between classes of possible actors (natural/social, local/global, and economic/cultural) and treat these categories as symmetrical effects of relational practices. Con sequently, ANT approaches the world as consisting of heterogeneous relations and practices through which humans and nonhumans alike are treated as possible actors. This basically means that we cannot take anything as given, as everything is an effect of relational practices. Actors are assembled and structures are arranged in a recursive process of networking or translation. Through the filter of ANT, the world is depicted as a mobile ar rangement. It thus seeks to abandon much of the con ceptualization of the modern episteme that rests on binary terms and dichotomous relations between purified categories such as nature and society. ANT has sought to highlight the frailty of the modernistic worldview and underline how the making of society demands association of diverse elements that never exist as pure categories cut off from wider fabric of relations. Two characteristic terms that cast light on this associational process are immutable mobiles and centers of calculation. The idea of immutable mobile refers to the general term of in scription that describes how an entity may be trans formed into a stable object. An immutable mobile is an object that holds its shape even when it is displaced in time space. As such it is a crucial component in the mechanism of the modernistic worldview. Centers of calculation refer to the sites of transformation, that is, the location where immutable mobiles are drawn together, assembled, and calculated.

To explicate the argument and the worldview of ANT, it is useful to run through the meaning of the key con cepts of ANT: actor, network, and theory. The principle of general symmetry is instrumental for a quite different understanding of each of the concepts from what is usual within the social sciences.

Actors

In ANT, an actor is a relational effect. Hence, agency is a matter of accomplishment or a collective achievement produced through the enactment of networks but not an inherent trait of particular subjects such as humans. This implodes the conventional distinction between subjects and objects. Instead of being cut off from each other, ANT views them as relational. Objects are thereby not passive components of our world, they are 'quasi objects', meaning that they are capable of agency as they affect all inter action. Most importantly, objects and materials are able to stabilize social interaction, rendering it more durable than pure social acts. Humans would only be naked bodies without their props and it is only through interaction – networking – with materials they become (human) actors. Hence, for ANT, the solid individual actor is nonexistent until it has been stabilized as such through relational ordering. This basically means that ANT treats all actors as hybrids and perceives agency as spun between different actants in networks. Thereby, there is no pure element but everything is part of bundles of heterogeneous networks. The point ANT is making is that the division that can be recognized in the world between the active human subjects and the passive material objects is not given in the order of things. It has to be worked upon, practiced, enacted, and re enacted in order not to crumble down.

Networks

The conventional view on networks depicts them as a sort of channel between nodes stretched across Euclidian space. The nodes can, for example, be people connected through social networks or places linked through a network of transportation. Within those channels, reciprocal exchange or transport of various sorts can take place, for example, of information, trust, goods, and money. The network concept undergoes changes on two fronts within ANT. First, it is not either social or material but made up of series of heterogeneous actants. This means that networks are always actor networks. Hence, the social is not some stuff that is transported along system of pipelines but rather a part of and an outcome of the relational practices of networking. Second, it follows that the work inherent in networks, which is carried out by heterogeneous actants is highlighted. Hence, network cannot be thought as distinct from practices as they necessarily emerge through those. One of the implications of this is that one cannot automatically trust networks as lines of transport. They emerge through series of transformative practices; they are translations. Actor networks may be ordered in ways which make them capable of transporting objects intact over distance but that demands work and effort by diverse actors, to make them immutable objects. Actor networks are thereby more fluid and insecure but also more material than most other networks in the social sciences.

Theory

The concept of theory does also shift its meaning in relation to ANT. The ANT approach should be understood as a theory of what to study rather than an interpretive framework of the world. The concept of translation can be used to cast light on this. Within ANT, translation refers in general to the relational practices through which actors come into being, that is, the work involved in actor networks. The concept highlights how actors must constantly work in relations for assembling an order to live by. Translation in this sense is a process of establishing communication or making connections between actants. The theory of ANT is nothing more and nothing less than this: in order to gain some insights into the different ways actors use to order their lives, it is necessary to follow their actor networks, that is, their translations. It is through translation that actors (and networks) can grow by recruiting ever more actants under their power and thus construct an order. However, translation is a precarious process, which demands a lot of work if it is to create what appears as a solid order and it follows that even though an order may seem durable, it should not be taken as permanent or ever reaching a final level stabilization.

These key concepts are instrumental for the study policy of ANT. ANT stresses an empirical investigation – the tracing of relations – and the description of these. Explanations should emerge out of detailed descriptions and are thus local to the network under study but not transcendent of it. As such, ANT seeks to illuminate the emergent patterns of order and/or disorder. In turn, the social cannot be used as explanation but is part of what has to be explained. Early ANT studies, because of this objective, sometimes had a penchant to functionalistic descriptions of relational configuration, where everything seems to turn into networks. In recent years many students of ANT have sought to distance themselves from such inclinations and given up the hope to reach the end point of network order.

Critical Remarks

At least three critical points to ANT have been raised: First, some would argue that ANT inspired approaches are not the most obvious choice for studies of cultural imaginations, geographical imaginations, nationalistic projects, and the like. While some ANT studies do not discuss the discursive framing of practices much, this is however not the case with all ANT inspired research. It can thus not be taken as an inherent weakness of the general approach. Somehow along the same line, questions of power, domination, oppression, and colonization are not always made explicit. This is in spite of the obvious inspiration from Michael Foucault, the familiarity with much work flowing from feminist studies on these matters, and the fact that some ANT researchers have worked explicitly with power relations. Clearly, however, normative positions are usually not spelled out by ANT proponents themselves. Finally, there is the critique that ANT makes fluids a fetish, appealing to almost liberalist notions of the freedom of movement. But while there is a risk that ANT can be used this way, this does not need to be the case as a variety of studies show. Here it may be noted that two of the main advocates of the approach, Latour and Law, have distanciated themselves from the ANT label after 1999 (the publishing year of *Actor Network Theory and After*) although the former has reclaimed it, most notably, in an introductory book to the approach published in 2005.

Moves into Geography – Bypassing Dualisms

The move of ANT into the realm of human geography happened parallel to a general conceptual development in the field. This was not least expressed by the cultural turn, the general rise of feminist studies to prominence, and a broad relational view as, for instance, framed in the non representational theory. The import of ANT perspective into the fields of human geography mainly took place along two interconnected routes, both of which were inspired by the principle of general symmetry. In both cases, the prime motivation for turning to ANT was to bypass what seemed to be insolvable dualism in geo graphical research practices.

First, during the mid 1990s geographers turned to ANT as a source of inspiration for moving beyond the traditional dichotomy between society and nature. For many, the relation between society, culture, and nature seemed too complex and messy to be easily fit into the conceptual framework of modernity. Geographers in Britain working along these lines included, for instance, Jonathan Murdoch, Nigel Thrift, and Sarah Whatmore. The borderline between what was natural and what was social or cultural came across as increasingly fuzzy in times of global warming, increasing mobilities, expanding use of information technologies, evermore refined cyber technology, and genetic sciences. The ANT approach promised a way to bypass the dualism between nature and society by shifting the focus away from *a priori* cat egories toward the associations and relations through which these were emergent. A very important element of ANT in this regard was that it sees nature and the ma terial as integral to the social and the cultural; the ma terial is already there but is not given the status of an add in variable in a cultural explanation of society as human geographers were prone to do. Key concepts in this regard are the hybrid and quasi objects. The argu ment made by ANT is that all things are relational or hybrids and if they seem to be solid bits and pieces it is only the effect of work in nets. It is exactly through such network that our modern world order that neatly sep arates nature and society has emerged. Importantly, the hybrid is not thought as a mix of two pure forms but simply as a condition of the world where nothing is thinkable outside relations. In the terminology of ANT, the world is like a seamless web and the challenge is to follow how this web is broken up and ordered in the form we recognize as, for instance, society and nature.

Second point of entry for ANT into the realm of human geography was configured by a general debate on philosophies of science. Roughly put, this discussion drew its energy from a rift between those adhering to realism (especially the natural sciences) and those advocating so cial constructionism (especially the social sciences). As the making of scientific knowledge was one of the earliest domains of ANT research, ANT found itself somewhat caught in the eye of the storm. It did not, however, fit easily to either side of the argument. The proponents of ANT framed the approach as much more material than social constructionism and much more discursive than realism. It thereby offered a middle way that could move beyond the realist–constructionist impasse.

This offer was welcomed by human geographers who saw the realist–constructionist discussion leading nowhere. Early examples may be seen in Thrift's *Spatial Formations*, published in 1996 and in the writings of Nick Bingham, Steve Hinchcliffe, and Jonathan Murdoch. ANT highlighted that hardcore social constructionism was the flip side of hard core realism. While the former saw society and culture as explaining everything, the latter gave the power of explanation to natural laws or technical deter minism. The principle of general symmetry is critical in this regard. In short, the theorem of general symmetry renders in principle all actants equal. The study policy of ANT then aims at tracing the practical orderings under lying our society and hence the ways power relations are formed and stabilized. It can thus be said that ANT sug gests a flat ontology with no taken for granted horizontal or vertical hierarchies; it seeks to trace the becoming of the ontological order we live by. Reality is thereby nothing one can think up independent of the crude and concrete effect of material presence. Cultural meaning is not in scribed in a straightforward way onto material objects but is emergent through relational practice. ANT thereby contends that reality is indeed constructed, but it shifts the focus away from the purely social to the practices of ontological construction of the world, which takes place through heterogeneous actor networks.

Economic Geography

Since the mid 1990s, ANT has been used in different contexts of geographical research. One subfield which has been affected by ANT is economic geography. For eco nomic geographers, it was not so much the idea of hy bridity or the ontological stress that was the most interesting part of ANT, but the centrality of the concept of network. During the 1990s, economic geographers paid more and more attention to the ways the economic was situated in a wider fabric of social relations and networks of cultural institutions in a search for more nuanced accounts than were possible with the traditional macro approach of political economy. The focus of an alysis was increasingly put on the interconnections be tween the cultural and the economic and how socioeconomic linkages worked to shape the spatial or ganization of the capitalistic market economy. Con versely, relational concepts such as chains and network gained ground. Often, conceptualizations of networks took the form of strong social networks inside a bounded region or place linked to a global scale through more sparse or weak connections.

As was explicated above, ANT endorses a different kind of network view than is usual within the social sci ences. This may have prohibited its spread or general acceptance in the field but it has also provided a particular edge to the approach in the context of economic geog raphy. ANT has thus not been taken on board by many economic geographers but is however being increasingly accepted and understood as a valuable device especially due to its capacity to bypass dualisms such as between the local and the global and the economic and the cultural.

ANT made it possible to recast economy relevant relations and their ordering in time and space. ANT sees scale as problematic and does not recognize a distinction between a local and a global level or a micro–macro distinction. There are only networks, which are of different lengths to be sure, but still only networks that emerge through practices. The implication of this is that ANT highlights the work and the processes underlying the more or less far reaching networks comprising economic activities. It offers a perspective to study the work of localizing and globalizing, understood as practices that render particular actor networks so stabilized and robust that they can extend over long distances, even worldwide, but still remaining local at all points. Not surprisingly, the key elements for the extension of local networks are material elements and nonhuman actors.

This approach makes macro theorizing of economic processes, conventional in political economy, problematic because that approach depends on a distinction between local life world and global world of system logic in explaining place specific effects of a globalized economy. The ANT perspective formulates both the life world and the structural logic as relational effects of actor networks. The ANT approach also blows up the usual framework of analysis used by institutional economic geographers, which depends on a distinction between social interaction on the microlevel, often seen as unfolding within a bounded place, and macrolevel connections between places. The reason is that more attention is given to diverse forms of connectivity of places and regions. Places are seen as coming into being through relations rather than taken as fixed points that stabilize economic networks. Similarly, markets are not taken for granted, but understood as assembled and materialized through the process of network. The geometric grid like surface of the world on which economic life has been projected is thereby increasingly being disrupted. Alternative spatialities of economic activities are being developed, which make use of topological thinking thus highlighting diverse rationalities and relational configuration of economic practice in time and space.

Network Geographies

ANT can be said to be proposing network geographies. There are two aspects of this proposal that need mentioning: first are the basic characteristics of the geographical imagination advocated by ANT, and second is the methodological implication of this perspective for geographical research

Topological World

Space and time are central to ANT. Actor networks necessarily depend on nonhuman actors for extension across space and durability in time. To trace the practices of actor networks conversely involves shifting in time space; space and time are entwined and emerge out of the enactment of heterogeneous relations of actor networks. In other words, ANT follows the networks and associations which configure and shape the sociospatial landscape of the world. Attention to time and space has often been an implicit feature of ANT studies although it has been central to the philosophies that ANT is based on, such as in the work of Serres. Questions of space and time are currently being increasingly addressed in ANT studies that express both an interest by practitioners of ANT in geography and by geographers in potentials of ANT.

Central to ANT is the process of translation, that is, transformative practices that describe the making of connections, assemblages, or associations. This is the productive force of actor networks. When translation is studied, it is evident that it enfolds combinations and recombinations of diverse objects or actors. This is a process that plays with – and goes beyond – the relation between proximity and distance. Actor networks thereby render distant things seem proximate and conversely the inclusion or exclusion from actor networks creates new distances between actors.

These moves disturb the conventional geometric conceptualization of space. The world is no longer a grid like surface on which it is possible to draw stable and enduring lines of proximities and distance. This kind of Euclidian geometry and the cartographic geography that builds on it is rather seen by ANT as one instance of 'spacing' and 'timing'. Euclidian space is a particular form of time space that does not have monopoly on spatial imagination. There are possibilities, other ways of spacing and timing through which the sociospatial landscape emerges.

ANT suggests a turn to topological ways of thinking. Topologies can describe the relational ordering of spaces for which geometry cannot give a good indication of proximity and distance. Topology is not dependant on linear time. It can deal with time spaces that come across as fluid, mutable, or even flickering as fire. These are spatialities that emerge through enactment of complex and heterogeneous relations. Topology grasps sociospatial realities that are beyond measurement but which are all based on relations.

A topological world calls for topological geography – a geography that can deal with complexity of relations and networks. Topological world is a world of multiple spaces. The Euclidian space is not written off but only thought of as one possible way to order spatial configuration of relations enacted through actor networks. Increasingly, the proponents of ANT and geographers alike are pursuing such a multiple approach that recognizes the interference between forms of time spaces. That demands an apprehension of how any object or any actor is part of and co producer of many time spaces. In other words, an actor is basically an intersection of multiple relations or to cite one of the old mantras of ANT: Every actor is a network.

Methodologies

ANT, as a study policy, is characterized by emphasis on description, where stress is put on following the re lational practices of actor networks. In line with ethno methodology, the researcher should trace the footsteps of the actors under study and describe their ways of raising the world. It follows that one should trust the actors so as not to render them subject to external explanations. The explanations coming from the study should be internal, emerging out of the field of practice under study.

In this regard, it is important to highlight the re lational worldview of ANT. The field is seen as always a relational achievement and many of the practices going on in every field revolve around its maintenance, or the continual enactment and re enactment of the field. This is a key point if the study is not to turn out to be a functionalistic exercise in filling up a blank field with descriptions of networks. It follows that the relations in and out of the field in question are as important as what goes on within it, as these constitute the field to a large extent and hence, the boundary of a field is primarily a practical achievement.

This relational understanding of the field has come as an inspiration for geographers grappling with the geo graphical dimensions of the field and fieldwork. The common sense notion of the field in human geography has been that of a bounded space, or at least a space that can be metrically defined and located in Euclidian space. Thinking of the field as an actor network that draws on and enacts diverse time spaces opens up new pathways for fieldwork. It becomes possible to move the focus from a description of the content of a field – often predefined by particular actors – toward the work and the contingencies involved in constructing the field. ANT is thus able to sensitize fieldworkers to their own role in constructing the field they are describing. This underlines that an ever present feature of fieldwork is that it partly creates the field it describes as it carves out situated knowledges of it.

In sum, network geographies acknowledge the ontol ogy of relational practices as a starting point of research. They also acknowledge that the implication of that is a slight confusion of ontology and epistemology, that is, what the world is like and what can be known about that world. Practitioners of network geography partly as semble the time spaces they describe, through their re search practice. The aim of such practice is not to find a fit between an account of a field and its reality but rather to find ways to move on, keep on course, or change it and thereby not least to reflect on alternative routes of travel.

ANT has attained a place in the toolbox of con temporary human geography. It has spread deeply into different subdisciplines often as a part of more general relational turn in the field. Taking the aforementioned critiques into consideration it seems however, imperative not to link ANT too tightly to other network approaches, simply because it is about much more than networks in the conventional sense of the word. ANT is thereby an important part of the inspiration to a broader new tra dition of network geographies as it seeks to keep the concept of network, fluid and mutable.

See also: Critical Realism/Critical Realist Geographies; Cultural Turn; Feminism, Maps and GIS; Political Economy, Geographical; Networks; Non-Representational Theory/Non-Representational Geographies; Regional Actors; Regional Development and Noneconomic Factors.

Further Reading

Bingham, N. (1996). Object ions: From technological determinism towards geographies of relations. *Environment and Planning D: Society and Space* 14, 635 657.
Bingham, N. and Thrift, N. (2000). Some new instructions for travellers: The geography of Bruno Latour and Michel Serres. In Crang, M. & Thrift, N. (eds.) *Thinking Space*, pp 281 301. London: Routledge.
Callon, M. (ed.) (1998). *The Laws of the Market*. Oxford: Blackwell.
Dicken, P., Kelly, P. F., Olds, K. and Yeung, H. W. C. (2001). Chains and networks, territory and scale: Towards a relational framework for analysing the global economy. *Global Networks* 1, 89 112.
Grabher, G. (2006). Trading routes, bypasses, and risky intersections: Mapping the travels of 'networks' between economic sociology and economic geography. *Progress in Human Geography* 30, 163 189.
Greenhough, B. (2006). Tales of an island laboratory: Defining the field in geography and science studies. *Transactions of British Geographers* 31, 224 237.
Hetherington, K. (1997). In place of geometry: The materiality of place. In Hetherington, K. & Munro, R. (eds.) *Ideas of Difference*, pp 183 199. Oxford: Blackwell.
Jóhannesson, G. T. (2005). Actor network theory and tourism research. *Tourist Studies* 5, 133 150.
Latham, A. (2002). Retheorizing the scale of globalization: Topologies, actor networks, and cosmopolitanism. In Herod, A. & Wright, M. W. (eds.) *Geographies of Power, Placing Scale*, pp 115 144. Malden: Blackwell.
Latour, B. (1993). *We Have Never Been Modern*. Cambridge: Harvard University Press.
Latour, B. (2005). *Reassembling the Social: An Introduction to Actor Network Theory*. Oxford: Oxford University Press.
Law, J. (ed.) (1986). *Power, Action, Belief*. London: Routledge & Kegan Paul.
Law, J. and Hassard, J. (eds.) (1999). *Actor Network Theory and After*. Oxford: Blackwell.
Mol, A. and Law, J. (1994). Regions, networks and fluids: Anaemia and social topology. *Social Studies of Science* 24, 641 671.
Murdoch, J. (1997). Inhuman/nonhuman/human: Actor network theory and the prospects for a nondualistic and symmetrical perspective on nature and society. *Environment and Planning D: Society and Space* 15, 731 756.
Murdoch, J. (1997). Towards a geography of heterogeneous associations. *Progress in Human Geography* 21, 321 337.
Thrift, N. (1996). *Spatial Formations*. London: Sage.
Whatmore, S. (2002). *Hybrid Geographies*. London: Sage.

Affect

J. D. Dewsbury, Bristol University, Bristol, UK

Glossary

Affect Affect is an experiential force and a material thing that exists as a relational experience and a phenomenological realm in a noninterpretative and impersonal manner. As such it is a state of becoming, given that it prestructures codified emotional responses to physical experiences.

Immateriality The immaterial is that which is not manifest in tangible physical form yet exists as instructive, channelling, and formative in forms such as dreams, intuition, ghosts, and imagination, etc. Its existence and power rests more on belief than certain knowledge but is nonetheless real in its affects.

Phenomenology Phenomenology is a conceptual approach that prioritizes embodiment by arguing that the role of the body is equal to that of the mind. As such knowledge is based on the perception of objects in the world and thus on what is apparent as mediated through the body by vision, touch, smell, etc., as much as the significations that make sense of such direct experience.

The Virtual The virtual is counterposed to the actual, being a realm of real experience that shapes what actually happens while also being suggestive of all the multiple worlds that don't get actualized. It is the pure past, being all that has happened creating the conditions for the emergence of the actual present, and an open future, given that what actually happens influences what is to come.

Introduction

Affect is at once an actual phenomenon and a virtual force, a material effect and an immaterial disposition. As a conception, it pairs dangerously close to our understandings of emotion and therefore exists as something familiar and seemingly knowable despite not having objective tangibility. As a result of this it is often rendered academically in fairly abstract and philosophically rich but obscurantist terms. It is then a troublesome concept to mobilize in the social scientific endeavor of doing human geography. It is at once a seemingly necessary and useful idea with which to confront and reveal key aspects of practice, embodiment, and performative based understandings of sociality, and at the same time steals our intent as academics, forcing us to rethink the habitual interpretations we live through, thus taking us into unfamiliar ways of thinking that are not easily written up fit for academic dissemination.

For human geography the study of affect comes at a particular time in the history of the discipline, a child of a fast and furious set of fashions emanating from the cultural, practice, and performative turns that have reverberated across the field in the last 10–15 years. As a phenomenon itself though it is of course nothing new sitting inventively at that juncture where we attempt to think through how we, both in the past and the present, perceive the world through the senses. Herein Christopher Woolgar's accounting of the senses in late medieval England halts any dismissal of affect's importance as merely faddish. In terms of theorization, affect places us close to philosophy's heart, namely at that first, although not originary, moment of intuitive comprehension; at the edge of becoming in a becoming world where we can never quite know exactly what is what and where we are. Therefore questions of idealism, realism, essentialism, and even materialism are recast as what was once assured interpretation housed in the phenomenology of intentional consciousness gets fragmented, displaced, and shaken up.

In a sense then affect seriously disorientates the modes by which we make academic accounts of the world. This is particularly so with the body, the center, and the origin of the logos in these affect based accounts, whether discerned passive or active, which no longer finds itself in relation to an outside world of lifeless matter for matter now permeates its very being. Here a concept of nature is presented in terms of a materialism whereby one substance, that exists both as matter as force and as force as matter, has many centers, perspectives, or folds prior to any sense of a perceiving subject. The humanist telos of a singular perspectival phenomenology of perception is undone by a constructivist doubling of sense wherein the sensible (what we feel) is rendered simultaneous with the intelligible (understanding and acting on that feeling). Consciousness and comprehension do not prejudge action as intentional, rather action takes place in the realm of the senses where the nerve snap of taste – 'disgusting!' for example – informs and infects the visceral cartography of the body before it is apprehended and understood – as 'sour'.

The philosophy of Gilles Deleuze is a significant resource for human geography's growing appreciation for affect. In particular, Deleuze opens up the question of the thought demanded in the empirical acceptance of affect, pushing us to think affect heterogeneologically and differently accessed by philosophy, science, and art; and in the excitement of all this experimentation, in our own bodies too. There are many research compounds here. Philosophy and science … Philosophy and art … Science and art ….

Affect is then a concept that tasks social science to think real experience again. Contra critiques that it is more an engineered abstraction than an earnest accounting of a grounded encounter; it is empirically messy and often explicitly sensuous. What is distinct is its removal from social constructions of everyday emotion, from ideas that specific movements and feelings of the sensible are intelligibly framed and thus tangible and thus understandable. Affect is then the surplus that is not socially constructed; it is instead the stuff of our being and not the semiotic material that enables us to understand our being. Ultimately affect attends to the fact that a lot of the world is unintelligible. It is then in excess of epistemology and of consciousness, existing as a substrate of potential bodily response, being auto motive or ontogenetic rather than willed.

Modes of Thinking Affect

Summarizing and working through the concept of affect, it is helpful to think of it in four different ways. The first two take affect to be a material thing, firstly understanding it as a phenomena and secondly as a force; then it can be addressed directly as a theory before finally understanding affect as a mode of expression. So in what follows each of these modes of thinking, affect will be discussed in turn and references will be made to the way in which affect has figured respectively in the work of human geographers.

Affect as a Phenomenon

In relating to affect through phenomenon, a lot of research into affect attends to the affective capacities of our bodies. One can categorize in this way the work done on dance movement therapy, on neuroscientific experiment, the appreciation of art, and so on. For this the work of cultural and communication theorist, and translator of Gilles Deleuze, Brian Massumi, has been the primary source of ideas with Derek McCormack being the main geographer working in this area. What this work does is to take affect as exactly the medium through which the body relates to the materiality of the world. Exampling this, one can think on dance, as McCormack does, as relations and movement that signal affect in the body's encounter with the volume of space, the proximity of touch, and the connective rhythms of life produced therein. In other words, the movement of air, the pump of blood, the flow of adrenalin as the dancer and dance unfold. The significant additive of thinking phenomenon as and through affect is that it enables us to take a step back from the intentional subject prevalent in the traditional phenomenological tradition of Husserl and Merleau Ponty (although not exclusively these two thinkers or indeed their work as a whole). Phenomena are thus encountered prior to any subjective framing and are rendered more in terms of felt intensity which produces a kind of understanding before it can be signified and articulated. Thus the material phenomenological state of being in the world is expressed in terms of a field of sensible experience full of unqualified intensity.

Affect as a Force

What you then have is a shift in vocabulary and thus conceptual purchase. Conceptual words such as intensity – with concomitant economies of increase and decrease, thresholds and reduction; and sensibility – with notions of tendency, acquiescence, and intuition – come to the forefront in accounts about affect. Such a shift is equally an emphasis toward rendering affect as a somewhat nebulous force rather than as 'something' that exists as a directly graspable object. No longer something definite like that finger pressing on the skin of another, but rather the sensation of the flesh of the finger as it encounters the sensible penumbra of other fleshy matter. Affect is then this penumbra, it is like a force field; something felt, something known to be there, but equally intangible in being not quite there. Affect is then understood as an invisible presence. Significantly this maintains that state of understanding the world in a realm prior to any heavy handed signification: thus it is not that your hand is being stroked by a doctor or a lover (significant significations altering the event of the stroking in your mind) but that the sensation itself is what matters, or at least exists in a less reductive state than being a medical examination or an amorous caress. Epistemologically the focus is modified from attending to interpretations of what things 'mean' toward an experimental endeavor to find expressive space for thinking and presenting what things 'do'.

These two taken together, affect as phenomenon and force, focus attention directly onto that of the body itself. What this means is that we are understood as subjects which are caught and situated as bodies within radiating ripples and circuits of feeling, intensity, response, and sensation, the flows of which wrap into and fold out of the body that we call our own. There is then no such thing as a singular subject but rather a series of potential subjectivities that are multiple and emergent. Affect places the body in different ways and toward different understandings of selfhood. It is helpful to think of the body in affective terms as a surface of intensities whose regions become folded to provide depth. The body is thus defined as the nexus of internal pain and external pleasure. As such the body is understood as a receptacle for affective experiences, as something that remembers such experiences as and through corporeal proclivities which have a long hold on constituting our being by telling us something about our past and aligning us to the present as more or less open or closed to the current sensorium in which we find ourselves. Illustrating this David Bissell's work, looking at how we acquiesce and become differentially present with respect to affective intensities in seemingly mundane practices, opens up interesting edges to the work of embodiment in geography. This comprehension of the body doesn't then fit neatly into the social interpretations we inevitably have and get given to us. Shuddering and quivering the body shatters expectation and opens up the obvious nature of seemingly self same encounters. As we are held within our bodies, being literally embodied, affect redirects our lives from their predictable journeys. However, this force is not just physical – when something affects us, let us think of this as two entities encountering each other, they become other than what they just were. Crucially this becoming is something between the two, outside the two. In this sense there is both a virtual and an agential dimension to affect given that the body is always in transition caught as it were on the waves of affect: 'You don't think so much as go with the flow'. Affect is therefore vectorial and not scalar: it crashes geography's propensity to map, replacing it with a passion for immediate diagramming.

The concept of becoming is crucial to the logic of affect and the quintessential example for thinking this through comes in the reinterpretation of Herman Melville's *Moby Dick* by Deleuze and Felix Guattari where Ahab becomes the whale in what they call zones of indiscernibility that are here manifested in intensities of whiteness and terror. Affect thus comes to be understood as the nonhuman becomings of humankind. On the one hand, there is the percept of the whale, this voluminous and inertially present or absent proximate mass – a nonhuman landscape of sea and animal – where it is the proximity and geometry that situates the becoming body. But crucially, and on the other hand, the percept does not operate alone, for there is the affect, that becoming other in the body in the encounter of this space of whale and sea. Both the percept and the affect are wrested from the perceptions and affections of the perceiving and feeling subject. It is for this reason that with an extreme suppleness the concept of affect opens doors in the impasse between subjectivity and objectivity. In combination, percept and affect produce blocks of sensation: visceral proximities across a whole host of media all explicitly material – "the smile of oil, the gesture of fired clay, the thrust of metal, the crouch of Romanesque stone," according to Deleuze and Guattari (1994: 166). John Wylie's phenomenological animation of landscape within geography has been literally the most affecting presentation of these ideas.

Affect as a Theory

Theorizing affect directly questions and seeks to expose how we relate to each other, and the world, being produced by the world in certain ways. Starting at the immediate level of the body emphasizes that before we signify affective forces as emotional feelings such as love, hate, envy, shame, pride, etc., we acknowledge that the force itself delivers us to embrace or spurn that which we encounter in the world. Love or hate, are thus registered on a visceral level as desire or abjection in a way that is more akin to just 'following your nose' than signifying and framing, even if subsequently such frames enhance the experience and even intensify the affect – for example, your love in performances such as love letters or the words 'I love you'. It is for this that affect as a theory is often referred to as an ethological one: it defines an ethos in terms of a pared down visceral way of behaving. The concept of Parallelism, drawn from the work of Baruch Spinoza and Deleuze's translation of his ideas, is useful in arguing that a movement of the body is simultaneously a movement in the mind such that thinking and doing equate to the same thing. Affect is then subsequently the active result of an encounter either as an increase or decrease in the affective capacity of the body and mind to act. Perhaps somewhat awkwardly for our social scientific minds, Spinoza depicts these resulting manifolds of affect in a romantic, poetic, and emotional vocabulary, using the notion of joy when one's capacity to act is enhanced, and that of sorrow, when it is diminished. Take the example of an arachnophobe whose hairs on the back of their neck register the affect resulting in their encountering of a spider and who then thus freeze in fear. The affective capacity at some molecular level in the arachnophobe's body produces a diminished capacity to act in the presence of those eight legged modes of nature. It is the relation between the terms, between the body and the spider itself, which presents the affect and not the predominance in any shape or form of one of the bodies over the other. Further, it is in this sense that affect speaks of an impersonal relation of being affecting and affected, and thus of something that exists between 'you' and it.

In this regard, the political theorist, William Connolly, engages body/brain research in its obvious synergy with the Parallelism of Spinoza's concept of affect. In unpacking the contemporary scientific, cultural, and political terrain of body–brain culture relays he utilizes recent scientific epistemological advances and ontological developments in the field of neuroscience. As a result he coins the term 'neuropolitics' to define and make visible the politics at large in the cultural formations by which specific body/brain processes are produced. In geography Nigel Thrift has propagated many lines of research into the neuropolitical, rendering them all in some extent upon an understanding of the increasingly systematic knowledges behind the creation and mobilization of af fect as precisely a means and resource for choreo graphing our everyday lives in more or less subconscious ways. In being a point of application between institution and individual, Thrift's work cautions the social sciences to pay attention to formations through which affect as such is being inducted politically for specific instru mental ends in ways that offer up "new means of ma nipulation by the powerful."

Here, affect nestles into the project of non representational theory and its practice and perform ance based accent, for what is signaled is the performance of ideology following Nietzsche's dictum that change is better gained by convincing the body than disciplining the mind. A politics thus plays out around the formal relation of the bodily capacity to be affected and to affect, addressing all manner of spaces as gen erative ecologies that solicit means of thinking as thought in action. For example, one can think of crowds or the 'team' mentality in sport or military contexts, and so from the event dimensions of dance floors to the football pitch to the battlefield where moving–thinking without con templation delivers ecstatic escapism, sublime craft, and saves lives. There is a cautionary note to be had here, for at work is a leveling of the human with nature and this can lead to an unintentional slip into grandiose claims for affect. It is too sweeping a concept for too little actual empirical detail. The untimely and alternative work of sociologist, Gabriel Tarde, again much used and valor ized in the thinking of Deleuze, counters this and is the current avant garde for expanding our appreciation of the affective contagion of micro social mimicry (as repetition) and innovation (as difference) in the scene of different social milieus.

Affect as Expressed

In terms of human geography's general uptake of the concept of affect, this space of distinction between affect and emotion has become both the key point of application for ideas in this area and at the same time the most contentious as can be seen by the exchange in the papers between McCormack, Deborah Thien, and Ben Anderson and Paul Harrison. The distinction can be grasped as a spectrum from the autonomy of affect in its open, impersonal, and unqualified form, to its capture and expression in the qualified forms pertaining to emotion.

Emotions are usefully understood as tangible mani festations of affect. In this regard, emotions are the most intense capture of affect in communicable and expres sive terms. What the theory of affect makes explicit in thinking through emotion is that there is always an ex cessive, unqualified aspect to the qualified articulation of the state of feeling – the emotion – such that although words never quite do, they do become the means for rendering visible the affects we are subjected to and through which we become who we are. It is not the words that count, it is the stuttering to articulate that which is unarticulable as such, that renders partially visible that which would otherwise remain painfully incommunicable. Emotions are then affects as clothed in language such that we can retrospectively communicate the affective dispositions we have been delivered unto and begin to understand their impact upon us with the help of others. Questions have been raised and distinct approaches have emerged in siding with either the pri oritization of affect over emotion or emotion over affect. In addressing affect first, the emphasis is in decon structing the tendency to treat emotions as an authentic register of something that can be captured and investi gated as such. Affect is mobilized often precisely in its move beyond biological essentialism and epistemological certainty. Further, in here affirming a virtual openness, the concept of affect stresses the politics of 'world making' on a whole ambit of scales, enacting both a vigilant stance against a reductionism in social–scientific thinking and an exposure to the potential for creative experimentalism.

On the other hand, taking emotion as the starting point accentuates that everyday life is that which needs to be addressed, that this is elided in conceiving of a world beyond or before the human that thinking through affect solicits, and that on a daily basis it is the social constructions of affect as emotion that are the ways in which people live and think through affect. One dir ection out of this is to follow Massumi and argue that the purchase of thinking the social through affect is taxing because affect itself is of a different order to the usual social scientific epistemological frameworks in that it is inassimilable and is unknowlable as such. Affect and emotion are both denigrated as a threat to reason: the question then is to what extent, and in what different ways, do affect and emotion disguise and distort reality, distracting us by appealing to shallow sensibilities? This move takes us beyond questions of obvious modes of judgment whereby the ethos of affect affirms the creative potential of a heightened awareness of such sensibilities. The event, the singular, and the beauty of the unusual are thus valorized. The other direction has been to square up more directly with the emotional side, and this has revealed itself most prominently in the uptake of a more psychoanalytical perspective through the work of Silvan Tomkins and the translation of that work into the fields of queer theory and performativity by Eve Sedg wick where desire scripts affect through emotionally framed narrations such as shame and biologically indi viduations such as smiling. Accommodating both these directions Clare Hemmings offers a critically aware overview.

Conclusion: The Politics of Affect

For human geography, and for our roles as researchers, thinking with and using the concept of affect plugs us into the qualitative means and strategies we have for investi gating the everyday worlds in which we live. Affect based research thus operates a strong political gesture in dir ecting attention to the affective affinities produced among forms of life where we can think of Jane Bennett's – and in geography Jamie Lorimer's charisma based – notion of nonhuman entanglements; or to affect based accounts of modes of living, for example, that of homosexuality in Sedgwick's work, best accounted for in geography by Jason Lim; or to therapeutic practices as outlined by Jen Lea; or of the material and historical power geometries of affect commented on by Divya Tolia Kelly and the Materialist ontologies of race argued for in Arun Saldanha's work. All of which moves us further toward communities of understanding that are not only produced but also ac knowledged as having been there already. As such, re search approaches are transformed and social normatives are stretched, exceeded, overcome, and replaced.

As the debate between affect and emotion shows, the ability to construct enduring and determinable political environments is an open and hotly debated question, and thus a burgeoning research frontier. In particular, we can see this in very everyday terms when identifying affect with particular ecologies that afford specific affectual practices. Affect therefore tasks human geographers to think of the very specific renditions of the material effect of affect. The point is to understand that different in stitutional, visceral, and ecological situations access the performative implications of affect in different ways: thus, for example, think of the affective difference present in (1) designed environments for work (be that medical and counseling spaces, offices, shopping foyers, or army training grounds); (2) our own body capacities (be they differently sexed, ethnically marked, or physically able); (3) ecological assemblages mediated by technology and information (in the political mobilization of fear in times of terror or scenario planning in preparation of the un foreseen); and (4) human nature interfaces (be they the domestication of animals, the cyborg folding of flesh and technology, or the environmental impact of the man made on equilibrium states). The sum effect is to sharpen us toward thinking of the different technological and visceral ways in which affect is achieved and is consti tutive of new worlds for living in, and that the intensi fication of affect in thought and practice, where affect literally breeds affect, leads us to map anew quite literally the human of which geographers write.

See also: Actor-Network Theory/Network Geographies; Becoming; Body, The; Dwelling; Feminism/Feminist Geography; Non-Representational Theory/Non-Representational Geographies; Performativity; Poststructuralism/Poststructuralist Geographies; Psychoanalytic Theory/Psychoanalytic Geographies; Sensorium; Soundscapes.

Further Reading

Anderson, B. and Harrison, P. (2006). Questioning affect and emotion. *Area* 38(3), 333 335.
Bennett, J. (2001). *The Enchantment of Modernity: Crossings, Energetics, and Ethics*. Princeton, NJ: Princeton University Press.
Bissell, D. (2007). Animating suspension: Waiting for mobilities. *Mobilities* 2(2), 277 298.
Clough, P. (ed.) (2007). *The Affective Turn*. Durham, NC: Duke University Press.
Connolly, W. (2002). *Neuropolitics: Thinking, Culture, Speed*. Minneapolis, MN: University of Minnesota Press.
Deleuze, G. (1988). *Spinoza: Practical Philosophy*. San Francisco, CA: City Lights Books.
Deleuze, G. and Guattari, F. (1994). *What Is Philosophy?* London: Verso.
Hemmings, C. (2005). Invoking affect: Cultural theory and the ontological turn. *Cultural Studies* 19(5), 548 567.
Lea, J. (2008). Retreating to nature: Rethinking 'therapeutic landscapes'. *Area* 40(1), 90 98.
Lim, J. (2007). Queer critique and the politics of affect. In Browne, K., Brown, G. & Lim, J. (eds.) *Geographies of Sexualities*, pp 53 67. London: Ashgate.
Lorimer, J. (2007). Nonhuman charisma. *Environment and Planning D: Society and Space* 25, 911 932.
Massumi, B. (2002). *Parables for the Virtual*. Durham, NC: Duke University Press.
McCormack, D. (2003). An event of geographical ethics in spaces of affect. *Transactions in Institute of British Geographers* 28, 488 507.
Saldanha, A. (2006). Reontologising race: The machinic geography of phenotype. *Environment and Planning D: Society and Space* 24(1), 9 24.
Sedgwick, E. (2003). *Touching Feeling: Affect, Pedagogy, Performativity*. Durham, NC: Duke University Press.
Spinks, L. (2001). Thinking the post human: Literature, affect and the politics of style. *Textual Practice* 15(1), 23 46.
Tarde, G. (1969). *On Communication and Social Influence*. Chicago, IL: University of Chicago Press.
Thien, D. (2005). After or beyond feeling? A consideration of affect and emotion in geography. *Area* 37(4), 450 456.
Thrift, N. (2004). Intensities of feeling: Towards a spatial politics of affect. *Geografiska Annaler* 86B(1), 57 78.
Tolia Kelly, D. (2006). Affect an ethnocentric encounter? Exploring the 'universalist' imperative of emotional/affectual geographies. *Area* 38(2), 213 217.
Tomkins, S. (1995). *Shame and Its Sisters: A Silvan Tomkins Reader*. Durham, NC: Duke University Press.
Woolgar, C. M. (2006). *The Senses in Late Medieval England*. London: Yale University Press.
Wylie, J. (2005). 'A single day's walking: Narrating self and landscape on the South West coast path.' *Transactions in Institute of British Geographers* 30(2), 234 247.

Relevant Websites

http://www.affective sciences.org
Affective Sciences.
http://www.affectjournal.co.uk
Affect A Journal for Alternative Geographical and Political Analysis.
http://www.tomkins.org
The Silvan S. Tomkins Institute.

Africa

G. A. Myers, University of Kansas, Lawrence, KS, USA

Introduction: Defining Africa

Africa is huge, diverse, and complex. Africa is identified as the world's second largest continent, covering more than 30 million square kilometers. Its diversity and complexity extend to its physical geography, climatology, and biogeography. Africa has many vast plateaus, five major basins, long stretches of fertile volcanic mountains not far from desolate wastelands below sea level, lakes that are among the world's largest (Lake Victoria), deep est (Lake Tanganyika), and one of the most rapidly shrinking (Lake Chad), and an array of features famous for their similarly dramatic contrasts. With its great variations in longitude (from 37 ° N to 35 ° S) and topo graphy, Africa is both the most tropical continent and one with extraordinarily diverse and complex climates and life forms.

Equally notable elements of Africa's diversity and complexity appear in its human geography. This land mass is divided into more than 50 countries, and its peoples speak more than 1000 separate languages. Afri cans divide themselves or find themselves divided into a set of regions. The highest order of division is that be tween North Africa (countries that lie largely north of the vast Sahara desert) and sub Saharan Africa (every other country, including – strangely – the countries where much of the Sahara desert can be found, since their southerly and more populated areas are south of the desert in a belt of land referred to as the Sahel). Sub Saharan Africa is typically subdivided into regions re ferred to as West Africa, Central or Equatorial Africa, Eastern Africa, and Southern Africa. Further subdivisions of these regions are common, most importantly between Atlantic West Africa and Sahelian West Africa.

The difficulties of categorizing countries in terms of their regional identity give us a clue as to why Africa's diversity and complexity problematize its easy definition as a 'continent'. To divide Africa into regions based on state boundaries ignores the fact that with a few limited exceptions these lines were devised by Europeans in the nineteenth century colonial conquests, without consulting Africans. The boundaries of today's independent African states ignore preexisting African political divisions of space, many of which still endure in the consciousness of African peoples, often bisecting countries, straddling across two or three states, or leaving one state with dozens of entities that think of themselves as separate nations. Then there is the complex question of where Africa begins and ends, wrapped up in human preoccupations with and misunderstandings of race throughout history. Separating North Africa and sub Saharan Africa culturally is dubious unless one draws lines simply on the basis of skin color (and even then, things quickly get far more complicated than some simple 'North is white, sub Saharan is Black' dichotomy). The Sahara has served as a 'sea' (most re cently of sand, but over the millennia of human history, of water, or grass, or trees) across which a high volume of economic trade and cultural exchanges flowed for thou sands of years.

The idea of 'Africa' gets very fuzzy at its edges. There are a number of island countries that identify with Africa, several of which (notably Mauritius) have majority populations of non African origin. Egypt is identified as a part of the African continent – except then what do we do with the Sinai Peninsula, part of Egypt that is not on the African landmass? A growing African population all across Europe has a history stretching back far longer than most other Europeans recognize or understand. Arab caliphates ruled North Africa and much of Europe's Iberian Peninsula for the better part of eight centuries (while Spain still rules two Moroccan cities as part and parcel of Europe). Most significantly, of course, people of African origins comprise a major proportion of the populations of most countries in North, South, and Central America as well as the Caribbean islands, as a result of the three and a half centuries of the transatlantic slave trade. Typically referred to as the African diaspora, this diverse and complex population, in combination with or separate from smaller populations of diasporic Africans in Europe, Asia, or Oceania, creates, in a variety of forms, its own ideas of 'Africa', and this African pres ence across the globe leads to an all encompassing ex pansiveness to the idea of where Africa begins and ends.

Hence, it comes to be that Africa's geography is hard to separate from representations of Africa's geography and the political and ideological uses to which these representations are put. Anglophone geography has itself weighed heavily on this representational politics. The shifts in the ways English language introductory texts have represented Africa over the last century, for in stance, are remarkable. The discipline has certainly moved away from depictions of Africa as a Dark Con tinent, where the Europeans brought light to the darkness through the Christianity, civilization, and commerce of David Livingstone – a depiction quite consonant with the need for justification of Africa's subjugation and ex ploitation by Europeans and Americans. From this co lonialist or colonial apologist geography, through a somewhat less ideologically driven phase of tropical geography, to the ubiquitous development geography of the discipline today, Anglophone geographers have steadily moved toward what is considered 'postcolonial' approaches to Africa. Yet, many introductory texts con tinue to use sensational representations of Africa that are redolent with the Orientalism of an earlier age, refracted through a geopolitics that remains heavily influenced by the early to mid twentieth century colonial order. It is therefore useful to trace out the deeper historical geog raphy of representations of Africa, in an effort to understand the trajectory that has brought Anglophone geography to this still problematic postcolonial moment in its depictions of the continent.

Representing Africa I: The Early Years

When non Africans talk about Africa today, it is often in the context of discussions of poverty, disease, warfare, or political instability. Africans are often represented as villains or victims. This was not at all the way the outside world's understanding of Africa began. Indeed, the whole notion of any generalized collective understanding of a landmass or entity called 'Africa' took a long time to develop in the world – including within Africa itself. Africans did not, in most cases, refer to themselves as Africans or to the landmass collectively as Africa until the nineteenth century.

The term, 'Africa', most likely derives from a local term for a region of Tunisia (Ifriqiya) that came to be utilized by Greek, Roman, and Arab conquerors to stand for a larger area – eventually a province in the Roman empire's dominions on the southern tip of the Medi terranean Sea, and the major Arab settlement zone under the Ummayad and Abbasid dynasties that ruled North Africa from 670 to 900 CE. The more commonly utilized term for the territories that Europeans left largely un explored to the south of Egypt, Libya, and Africa was Ethiopia, inhabited by people they called Ethiopians. In the fifth century BC, the Greek historian Herodotus wrote detailed second hand descriptions of 'Ethiopian' cities and their customs (likely to have actually been based on the northern Sudanese city state of Meroe). Ptolemy's geography of the second century CE provided more details, but again was limited to the middle and northern stretches of the greater Nile Valley. What is identified today as Ethiopia 'was' a heavily populated and highly sophisticated agrarian civilization by the Greco Roman era. But the Greeks and Romans knew little of it, and even less about any other region of what would become known as Africa. In a maritime trading guide to the Red Sea from the late first century or early second century CE, *Periplus of the Erythraean* [*Eritrean*] *Sea*, written in ancient Greek, we gain some understanding of what outsiders knew of ancient cities in the Horn of Africa and along the Swahili coast, but little or nothing about Africa's vast interior.

Europeans were not the only outsiders curious about the continent. Early medieval Arab and Chinese travelogs and cartographies broadened and deepened the larger picture of the landmass. Arab writers such as Ibn Khal dun and Ibn Battuta traveled widely in the northern portion of Africa and wrote extensively about the people they met along the way. The Chinese explorer and trader, Cheng Ho, visited the East African coast in the fifteenth century and reputedly returned to the court of the em peror he served with two giraffes, gifts from the Sultan of the city state of Kilwa in today's southern Tanzania.

Indigenous African geographies and understandings of broader spaces in the ancient and medieval world are given short shrift in most historical geographies of the continent. Many African peoples developed elaborate geographical knowledge of their immediate regions, and several, most famously the Dogon of today's Mali, had highly developed astronomic skills that included broad awareness of the whole of the world. Africa even pro duced its own explorers. It is vital to remember that Ibn Khaldun and Ibn Battuta, despite their common identi fication as 'Arab' explorers, can be considered Africans, since they were born and raised in what became Tunisia and Morocco, respectively. That neither considered himself an 'African' and that both occasionally used de rogatory language toward sub Saharan peoples, though, usually leads to them being classified as outsiders to 'Africa'. But the seafaring explorers of the medieval Malian empire and the Swahili sailors of the Indian Ocean are less likely to be stripped of their Africanness, and both saw quite a bit of the world beyond the continent.

Representing Africa II: Exploration, Colonialism, and Imperialism

The earliest European human geographies of Africa were geographies of exploration and conquest, and from the era of the Atlantic slave trade through colonialism (a span of exploitation of Africa stretching from about 1450 into the 1960s and 1970s), they shaped the way much of the world views Africa. Through the Atlantic slave trade, an estimated 11.6 million Africans were forcibly relocated to the Americas, while roughly an equal number are esti mated to have died along the way in what has come to be called the Middle Passage. Historical geographers of Africa also point to the massive impacts of the Atlantic slave trade 'within' Africa. Untold millions died in Africa itself as a result of slave raiding, the destruction of babies, the sick, or the elderly that were captured, and still millions more were forcibly displaced within the continent. For example, the route from the southern Congo basin's interior empires of Luba and Lunda to the Portuguese slave port of Luanda came to be known as the Way of Death because of the atrocities all along it. About two thirds of the population of today's Zambia can trace its geographical and cultural lineage to the Luba–Lunda region of Congo, their ancestors having fled the collapse of these empires due to the crushing blow of Portuguese slave raiding and enslavement. British, French, Dutch, Spanish, and eventually American slaving ships expanded upon the Portuguese beginnings of the trade, with their main interests in the Atlantic West African coast. African states, city states, or empires that collaborated with the slave trade expanded the reach of their powers, irreparably corrupting their relations with neighboring peoples, the most likely to fall victim to the slave raiders.

West Africa's fortunes in the period of the slave trade illustrate its lasting geographical impacts. From the eighth to the fifteenth century CE, a set of major empires rose and fell along the Sahelian belt of West Africa, capitalizing on their geographical position to be the middle agents of trade between West Africa's gold rich forest belt and the Arab north (and Europe – the metal for the gold coins of the Italian Renaissance came mostly from West Africa) across the Sahara. The empires of Ghana, Mali, and Songhai, in particular, had highly developed bureaucracies and hierarchies of administration, beautiful architecture, renowned universities, extensive libraries, rigorous legal codes, and immense wealth. When we look at the Sahel today, it is one of the world's poorest regions. This reversal of fortunes symbolizes the impacts that the slave trade, and the later period of formal colonial rule, had on the continent as a whole, and it is a lesson in fundamentals of economic geography. When the Portuguese and other Europeans began to trade along the Atlantic West African coast, they cut out the middle agent role that the Sahelian empires had enjoyed.

By the nineteenth century, opposition to the Atlantic slave trade, and to slavery in general, had grown quite strong, not only in Europe and the Americas, but also in Africa. Gradually, the slave trade gave way to what was termed the Legitimate Trade – as a tacit acknowledgement of how illegitimate the previous Euro American trade with the continent had been. But the legitimate trade hardly brought an end to Euro American exploitation of Africa – indeed, it accelerated it. Imperial rivalries and competition between European powers for access to and control of Africa's significant store of natural resources and its many market outlets for European manufactured goods gave rise to the era of Geography Militant, when European explorers and geographers embarked on the cause of cataloging the continent's resources, products, and markets for their home countries. Even while agreeing together to avoid selling the most advanced line of weapons to African states, the Europeans competed heavily with one another to set up exclusive trade deals, to redirect the productive forces of African farmers, and, occasionally and in specific settings, to control territories and implant colonists. In South Africa, the implantation of white settlers had begun a bit earlier, with the Dutch establishment of Cape Town in 1652 CE. White settlement intensified in the nineteenth century, particularly with the 1806 handover of Cape Colony to the British, and then the discoveries of the enormous diamond and gold deposits (in and around Kimberly and Johannesburg, respectively) later in the century. The French conquest of Algeria in 1832 led to French settlement there in relatively modest numbers. Elsewhere, only small white populations existed, chiefly in port cities or in areas of mining or missionary activity, but the Euro American presence became quite strong in other ways – as a controlling hand on trade, as an influence on religious and cultural development.

Eventually, the competition between Europeans over African resources and markets led the German ruler, Kaiser Wilhelm, to call for a conference of European powers to settle disputes by marking out spheres of influence. This conference, what we now call the Berlin Conference of 1884–85, did establish these spheres of influence, delineating the rightful claims of the British, French, German, Belgian, Italian, Portuguese, and Spanish. However, the fact that few of the spheres claimed were actually even remotely under the control of those who claimed them in Berlin set off the scramble for Africa that lasted until the 1914 outbreak of World War I. In this scramble, the Europeans set about the task of pacifying the territories they had laid claim to. In some areas, the decades from 1880 to 1910 were as a result bloodier and more disastrous for Africans than the four centuries of slave trading. Some 5–10 million people are estimated to have died in the ruthless and shocking Anglo Belgian India Rubber concession area of what Belgian King Leopold II callously called the Congo Free State (after 1908, this became the Belgian Congo, and after 1960, the Democratic Republic of Congo). The Germans nearly wiped out the entire Herero ethnic group of German Southwest Africa (today's Namibia). The Italian conquest of Libya killed a third of its population. Other transitions to formal colonial rule occurred more peaceably, or at least came about in the absence of genocide. In any case, by 1914, nearly all of the continent had fallen under European control. The only politically independent countries by that year were Ethiopia, Liberia, and South Africa. Liberia, though, was a *de facto* colony of the United States, and more particularly of the Firestone Tire and Rubber Company, whose enormous holdings in the territory gave rise to it being mockingly called the Republic of Firestone. South Africa had gained independence in 1910 – but as a white minority ruled state where the black majority were denied political and economic rights, along with Asian immigrants and mixed race peoples.

(Under)Developing Africa

In the nineteenth century, European geographers explored the African continent and then aided and abetted the colonial conquest. In the twentieth century, they mostly served the cause of establishing and maintaining colonial rule during the first half, and then worked in the cause that came to be called Africa's 'development', in combination with new generations of African professional geographers. This project of development – which in most definitions combines economic growth with improving social welfare – began in earnest largely with the close of World War II, but it started on highly doubtful terrain.

After all, European colonial rule more or less established the rules and parameters by which Africa would be incorporated into the world capitalist economy, in such a way that advancing within it would prove elusive for most African states once they achieved independence (for most colonies, this came in the 1960s and 1970s). In 1972, the Egyptian economist Samir Amin subdivided Africa's colonial economies as 'Africa of the Colonial Trade Economy', 'Africa of the Concessions', and 'Africa of the Labor Reserves', and proceeded to argue that one could in effect still see these three categories of economies at work in the underdevelopment of the continent. Much of his argument continues to hover over Africa's generally difficult relationship to the Western idea of development.

Briefly, the colonial trade economies – predominantly Atlantic West Africa – were those in which Europeans determined the trade, markets, and prices for a thin range of one or two commodities (often primary agricultural goods) that each colony produced for European industrial interests. African producers maintained control over their land and labor, but under severely constricted conditions that tied their economic fortunes to European capital. Thus Senegal and Gambia came to be almost exclusively dependent on the production and export of raw groundnuts, whereas in Cote d'Ivoire or Ghana it was raw cocoa beans – to be then processed in Europe or the US, where the manufacturing value added accrues. As a direct legacy of this colonial system, for instance, West Africa produces more then 70% of the world's unprocessed cocoa in a given year, and yet almost none of the world's chocolate bars.

In concessionary colonies like those of the Congo Basin, huge stretches of the territory were simply handed over to European companies to administer as they saw fit, for their profit. This commonly resulted in the same heavy dependence on one or two exports tied to the whims of European or American producers and consumers, but with the added elements of land alienation, severe labor exploitation, and a more limited interest of the colonial administrations in human development. Former concessionary territories often saw the most violent processes of decolonization, and they have had the most violent postcolonial experiences – most obviously in the cases of Mozambique and the Democratic Republic of Congo (DRC). In most respects, they remained tied to export enclave economies of resource extraction dominated by European and American transnational corporations.

Part of the story of the underdevelopment of Mozambique and the DRC, though, ties in to Amin's third category, that of the labor reserves. Labor reserve colonies typically, though not exclusively, were those that provided workers to European settlers or capitalists, on mines and plantations (Amin also placed much of the Sahel in this category, since Sahelian labor became central to the cocoa industry of Atlantic West Africa). In many labor reserve colonies, Europeans took the best lands for agriculture and the most valuable resources, leaving Africans to less fertile and densely crowded reserves. Much of southern African today bears witness to the enduring legacies of this model of 'development': more geographically extensive infrastructure for the territories, but embedded in a highly unequal system for the distribution of development across classes and races, as in the extreme case of South Africa under apartheid. Most former colonies of this variety retain some infrastructural and industrial advantages when compared with the rest of the continent, but with vestiges of the severe dualism (geographically and socially) enduring as well.

The net result of all three types of colonial economies was that most African countries came to independence with heavy dependence on a limited number of exports (the trading, marketing, and pricing of which lay out of their hands), with a shortage of indigenous capital and powerful presence of foreign capitalists, and a high degree of dualism dividing the parts of the country strongly linked with the world economy and those disaggregated from it.

The post independence development era began with grand ambitions, but 40 years or so on from its origins, the project of development has left Africa with a few 'miracle' cases (South Africa, Botswana, and Mauritius are often cited), several 'mirages' (countries like Kenya or Cameroon where the surface might look good, but it is a façade of development only), and much exclusion and poverty. The blame for the lack of development on the continent can be placed with both exogenous and endogenous factors. History, as we have seen, plays a key role as a largely exogenous factor working against development. The Western agenda for development intersected directly with the Cold War, and many resource rich or resource poor African countries (Angola, DRC, Uganda, Somalia, Ethiopia, Mozambique, Namibia, and others) were caught in the Cold War crossfire, as another exogenous factor. Declining producer prices for many African exports coincided with skyrocketing prices for petroleum in the 1970s to leave many African countries seriously in debt. Yet endogenous factors have also been at work. The first generation of independent leaders often had great ambitions to set out on alternative development paths, frequently with expensive showcase projects to highlight how they were different. In other cases, corruption, greed, mismanagement, or internal conflict siphoned millions of dollars of Western development aid or investment into unproductive channels that did little to improve the quality of life of ordinary Africans.

Regardless of the path toward development that African leaders chose, by the early 1980s, most countries had fallen into external debt so large as to force countries into signing on to restrictive plans created by the World Bank to recover its loans. These Structural Adjustment Programs (SAPs) yielded few tangible results and created havoc in many countries politically. By the 1990s, SAPs gave way to the Heavily Indebted Poor Countries (HIPCs) initiative, but HIPC itself became mired in its own inadequacy, morphing into aid projects that required the creation of Poverty Reduction Strategy Papers (PRSPs) in order to receive the money. The PRSPs themselves were largely the SAPs in friendlier language.

The development era has transformed Africa and its connectivity to global political economy, alongside the continuities from earlier eras. There are a few cases where economic development has led to a broader extension of a decent quality of life to more Africans. But what one more readily sees is a splintering of African countries in sociogeographic terms, such that some parts of each country – typically, well to do enclaves in cities – are highly integrated into the global economy and global cultural flows, while many areas – sometimes right next door to those enclaves – are witnessing increasing deprivation and dislocation.

Postcolonial Africa

The phrase, 'postcolonial Africa' is literally a reference to the continent in the period after the end of European colonialism. However, the usage of the phrase in this sense is actually inadequate, because the autonomy of African states is so circumscribed and the legacies of colonialism so great as to lead many to question how far African countries can be considered to be 'post' – (as in past) – colonialism. Two other uses of the phrase, postcolonial Africa, are worthy of consideration. The first is the use of the phrase to describe the material and psychological conditions of living with the legacies of colonialism. The second is the effort to think beyond them. In this latter regard, it is worth contemplating several examples of new ways to think about Africa and Africanity that lead to reconsiderations of Africa's human geographies. The first example is the very practical example of the efforts toward African unity through the transformation of the Organization of African Unity into the African Union (AU) at the beginning of the millennium. The AU has certainly proven to be a problematic entity in its first decade, as have related efforts like the New Partnership for African Development, but they deserve mention as efforts to move the continent beyond colonized modes of thought or being. The broad rubric of 'African solutions to African problems', implicated as it often is in neoliberal visions of development not too dissimilar from structural adjustment, can nonetheless here and there be credited with genuinely realigning the way Africans think about Africa, for the better.

A second example of postcolonial rethinking of Africa leads us into contemporary discussions of and interest in the development potentiality of new African diasporas in Europe, North America, and elsewhere. As a result of the transatlantic slave trade, hundreds of millions of people of African origin came to reside in the Americas, and significant numbers of the Africans of the Americas also resettled in Europe – along with Africans who came to Europe more directly. Intellectual, cultural, political, and even economic connectivities have often been strong across the ocean separating these various parts of 'Africa outside of Africa', fueling extraordinary movements like the Harlem Renaissance and Negritude literature or philosophy, and leading to ideas of Pan Africanism and of a Black Atlantic. In African American culture (people of African origin within the United States), there have been various important reconnections with Africa, such as the movement for Afrocentricity that draws upon the writings of Senegalese intellectual Cheikh Ante Diop to articulate an alternative vision of Africa and Africans as central to human history (from what is pointedly represented as an African ancient Egypt onward).

This older diasporic Africa and its visions of Africa are now broadening and diversifying still further, as more and more Africa born people of African origin reside in the rest of the world. Newer diasporas often develop a transnational sense of themselves as both African and something else. As fluidity, flexibility, interdependence, contingency, and disjuncture have been implicit touchstones for understanding Africa from within many African cultures, so now has the very fluidity and flexibility of African ness have stretched around the globe. Transnationality can be seen as creating a 'rhizomatic' geography of Africa, like the scattered shoots of a rhizome (e.g., a strawberry plant). The challenge may lie in re collecting these shoots and runners into a coherent 'Africa', but regardless of this challenge it is certainly true that we now see many examples of new visions of Africa, or of what it means to be African. At the 2002 and 2006 FIFA World Cup in men's football, for example, players born on the African continent lined up playing for Germany, France, Portugal, Poland, Saudi Arabia, and the United States, and dozens more players with African ancestry had their pictures taken in uniforms that said Brazil, Sweden, Japan, Trinidad, or England. The third black United States Senator in modern history was raised in Hawaii and Indonesia, but he has a Kenyan father. In any case, in the 2008 US election, presidential candidate Barack Obama, like millions of African Americans, can claim dual citizenship in Ghana if he wishes. The in creasingly dynamic interrelationships of Africans in Africa and global Africans continue to redefine the human geography and the potentiality of the continent everyday.

See also: Colonialism I; Colonialism II; Development I; Development II; Diaspora; Postcolonialism/Postcolonial Geographies; Race; Representation and Re-presentation.

Further Reading

Abrahamsen, R. (2000). *Disciplining Democracy: Development Discourse and Good Governance in Africa*. London: Zed Books.
Amin, S. (1998). *Africa and the Challenge of Development: Essays*. Ibadan: Hope Publications.
Appiah, K. A. and Gates, H. L. (1999). *Africana: The Encyclopedia of the African and African American Experience*. New York: Basic Civitas Books.
Aryeetey Attoh, S. (2005). *Geography of Sub Saharan Africa*. Upper Saddle River, NJ: Prentice Hall.
Cheru, F. (2002). *African Renaissance: Roadmaps to the Challenge of Globalization*. London: Zed Books.
Enwezor, O. (2002). *Under Seige: Four African Cities: Freetown, Johannesburg, Kinshasa, Lagos*. Ostiledern Ruit: Hatje Cantz.
Freund, B. (1998). *The Making of Contemporary Africa*. Boulder, CO: Lynne Rienner.
Hamilton, R. S. (2007). *Routes of Passage: Rethinking the African Diaspora*. East Lansing: Michigan State University Press.
Hochschild, A. (1998). *King Leopold's Ghost*. Boston, MA: Houghton Mifflin.
Iliffe, J. (1995). *Africans: The History of a Continent*. Cambridge: Cambridge University Press.
Newman, J. (1995). *The Peopling of Africa: A Geographic Interpretation*. New Haven, CT: Yale University Press.

Relevant Websites

http://igskmncnwb015.cr.usgs.gov
Africa Data Dissemination Service.
http://www sul.stanford.edu
Africa South of the Sahara: Selected Internet Resources.
http://www.afrika.no
Index on Africa for the Norwegian Council for Africa.
http://130.238.24.99
The Nordic Africa Institute.
http://www.africa.upenn.edu
University of Pennsylvania African Studies Center web links.

Ageing and Health

G. J. Andrews, McMaster University, Hamilton, Canada

Glossary

Ageing-in-Place The policy and practice priority to allow older people to stay in one form of accommodation for the longest possible duration and to avoid unnecessary relocation

Demographic Transition The historical process of a population moving from high birth and high death rates to low birth and death rates.

Gerontology The academic study of the biological, psychological, and social aspects of aging.

Introduction

When considering the potential importance of the geo graphical study of older people and populations, some basic facts are well worth remembering. Older people – typically defined by governments and other officialdom as being over 65 years of age – constitute a significant and growing proportion of the general populations of both developed and developing world counties; the former approximately 15%, the latter approximately 5% (but growing more rapidly). Although many characteristics, and activities of older people and populations are similar to other age groups, they also face unique circumstances and challenges. Older age is a stage of life associated with specific physical and psychological changes and states and, of all demographic groups, older people experience most health and mobility problems, are the greatest consumers of health and social care, and face a great deal of negative stereotyping and stigmatization. However, contrary to some popular beliefs, they also hold valuable positions in society and can, of course, live diverse, rich, and fulfilling lives. Given this scale and distinctiveness of aging related issues, human geography has long been, and continues to be, interested in studying older people in a variety of contexts from a variety of perspectives.

Broadly defined, the geography of aging (or what, in disciplinary terms is occasionally termed geographical gerontology) is concerned with the spatial characteristics and contexts of aging. These range from the distributive features of older populations, to the ways in which places reflect, older people and the ways in which places are experienced and negotiated by them. Not all of this field is focused on health issues and many studies consider wider social – and more recently cultural – aspects of older peoples' lives. However, concerns for health are evident in some form, and to some extent, in a substantial majority of studies in this field.

As a field of human geography, geographical studies of older people date back over 30 years and crosscut the interests of a number of subdisciplines including social geography, health geography, and more specialist var ieties such as population geography and development geography. Not only have the empirical subjects studied mirrored the changing interests of these subdisciplines, but they have also changed in line with underlying paradigm shifts. As well as crosscutting various sub disciplines of human geography, geographical studies of aging have also drawn from, and contributed to, other recognizable bodies of research in human geography. For example, it is not unusual to see a particular geographical study of older people described as a welfare geography, a geography of caring, or an emotional geography. Notably, many studies are not identified explicitly as geographies of aging. This might be because the author does not consider the age of their subjects to be particularly im portant and/or the dimension he/she wishes to focus on.

With such a long standing and broad based discip linary history, and such a substantive body of work pro duced to date, it is perhaps surprising that geographies of aging do not posses a higher and more distinct profile in human geography (in comparison, for example, to chil dren's geographies). To speculate, this might be due to the broad scope of the field. It arises right across human geography yet is not associated with, or essential to, any one single subdiscipline. Another reason might be be cause geographies of aging have generally followed or mirrored disciplinary progressions and innovations rather than forging them. A very different reason for a modest profile might simply be that, no matter how socially relevant and important aging is, it is not a particularly glamorous subject and indeed reminds all individuals – including geographers – of their own mortality. Indeed, despite four previous reports on the field published in *Progress in Human Geography* since 1980 the last – pub lished in 1995 – is now rather dated.

Outside human geography, reflecting the global im portance of aging and health and the academic attention it accords, geographies of aging have also made strong and sustained contributions to, and have also drawn from, other disciplines. Some of these are squarely aging fo cused such as clinical gerontology and social gerontology, while others are more general health categories such as epidemiology, demography, and various professionally focused research disciplines – including nursing and social work research – within which aging is one empirical interest among others. While adding variety to the field, this interdisciplinarity, has equally caused tensions among commentators. For example, while certain geographers have been eager for geographies of aging to draw on, contribute to, and reflect gerontology and the health sciences, others have been more concerned that connections are strengthened with human and cultural geography.

Population Aging

Since the 1950s, a long standing strand of aging and health research has been concerned with population aging. It is an area of interest dominated by the work of demographers and population geographers who have particular expertise in handling and analysing large population data sets that are essential sources for these studies. Researchers have identified the spatial patterning and characteristics of demographic transitions including declining fertility (due to social change) and increasing life expectancy (due to improvements in diet, living conditions, and healthcare) that underpin demographic aging in particular areas. Following this, they have also identified the national, regional and local consequences of population aging in terms of health, society and economics. For example, research has highlighted the substantial challenges that lay ahead for governments, health services, and nongovernmental organizations in meeting even the most basic needs of those developing world populations, which are both growing and aging most rapidly. Here, countries in sub Saharan Africa and the Asia Pacific region are particularly well researched.

With respect to developed world countries, many of whose populations are not growing significantly yet are still aging, research has highlighted the continued need for policies and planning so that health, social care, and welfare services are funded, and provided to the necessary levels and standards. These inquiries inform popular political and public questions, and are often connected to specific issues such as pensions or access to medical treatments, how to handle the needs of an aging population, and who is ultimately financially responsible for older people. One by product of this research is that, in demonstrating increasing life expectancy, it has highlighted the need to rethink older age. Most notably, instead of older adulthood being thought of as a single 'third age', increasingly it is now thought of in terms of being the third and fourth ages of life (i.e., older and most old). Other social scientific and geographical research has built on this recognition and has articulated how these groups have distinct qualities, needs, circumstances, experiences, and diversities. Indeed, rather than simply recategorizing older age, the idea of third and fourth stages of life helps recognize that many younger older people are active participants in labor markets, are active and powerful consumers in economies, and are a powerful force for social and cultural change. Notably, rather than being consumer's of health and social care, they often themselves provide care to their parents, well into their 60 s and 70 s.

Movements of Older Population

Over the last 20 years, a great deal of research has focused on movements of older population. At one level, research has focused on broad migration patterns. Other studies – often which are qualitative have investigated some important social contexts and consequences such as family networks and the pushes and pulls to and from different locations in migration decision making, as well as the outcomes of such migrations on groups, individuals and health and social care systems. This research has considered these issues in the context of semipermanent migrations between countries (such as British retirees to Spain) with respect to semipermanent migrations within countries (such as British retirement migration to coastal areas) and short term migrations within, around or between countries (such as North American 'snowbirds'/ seasonal migrants).

A number of studies describe the factors that have led to the increasing distancing of older adults from their children. Underpinning this trend has been changing social attitudes with regard to the need for close geographical proximity between families and also changing attitudes regarding responsibilities for providing care and social support. Indeed, what is commonly known as 'the loosening of family ties' is recognized to have a strong geographical dimension. At the same time, both older people and their children have been shown to be more actively mobile, due to their changed economic circumstances. For example, in comparison with previous generations, the current generation of older people has more money and resources that provide opportunities to relocate, while their children might be more likely to relocate for career purposes. Whatever the reason for increased distancing however, research identifies decay in immediate interactions and social and caregiving support as distances increase between families. This does not necessarily have to lead to a compete breakdown of communication between families, and geographers and others have highlighted the concept and emerging reality of 'intimacy at a distance'.

The Health of Older Populations

Another long standing strand of research is focused on the health of older populations. Population geographers and demographers, along with epidemiologists and health geographers have played central roles. At the national scale, life expectancy in specific countries is an important consideration and bodies such as the World Health Organization have produced a number of publications outlining rates and variations in expectancy. Health status and morbidity of older people in given locales is another important concern, and studies have considered rates of many health conditions related to aging, as well as their relationships to older peoples' ethnicity, gender, affluence, and access to health and social care. For example, localized rates of age related conditions such as dementia, have been investigated.

Mirroring current interests in health geography, in recent years a growing interest in place effects on older peoples' health has emerged, often at the neighborhood and community level. This work is particularly important with respect to older people because, more than any other group, they rely on local social support and facilities. Certain research explores the impact of social 'composition' on health; that being the characteristics of people in given areas (including affluence, class, and family composition). Other research explores the impact of social 'context' of health; that being the resources available to people locally (such as primary care facilities, various local shops, and services). Meanwhile, most recently, collective social dimensions to health have been explored; specifically the role of social capital and the ways communities work together – or not – for common goals and gains for their older members.

Negotiating Environments

Some of the earliest geographical research, which engaged with older peoples' changing relationships with environments, drew heavily on environmental psychology published in the 1960s and 1970s, and in particular on the pathbreaking work of gerontologist Powel Lawton. Lawton's ecological theory of aging was most famously articulated through his environmental docility hypothesis (also associated with terms such as aging and adaptation). Here older people are thought to adapt their behaviors to cope with the increasing challenges that negotiating the environment poses with increasing age (what is known as environmental press). Although these ideas provided popular frameworks for research for many years, along with psychologists, over time geographers began to question the assumption that older people assume a passive role in such adaptation. Following amendments to the theory forwarded by Lawton himself in the 1980s, geographers have begun to research and articulate older peoples proactivity in adaptive responses, specifically how older people use their environments as a valuable resource in their lives and experience and approach them positively.

Developing these early considerations of environment, geographers have focused on many aspects of living environments from the scale of local communities to the scale of very specific forms of residential accommodation. In particular, one line of research asks how specific environments – whether they be a dwelling, urban neighborhood, or rural area – can be enabling or disabling. Much of this work is based on qualitative methodologies, and presents a rich and critical understanding of peoples' lives lived in, and through, places. With regard to rural living, research has focused on many everyday issues related to housing, transport, and broader socioeconomic change. With regard to urban living, research has focused on the design and negotiation of public spaces and facilities. Indeed, the various obstacles associated with negotiating urban life in older age are argued to include poor access to public transport, inadequate public signage, and a lack of alternatives to staircases. Research here has focused on their consequences, such as falls and increasing social isolation, but equally on design and planning solutions to assist older people. Indeed a particular strength of this research is that it moves beyond a traditional focus on older people's functional limitations to find positive ways to make environments more enabling for them.

Residence and Care

It is typical for studies in this field to simultaneously capture elements of older peoples' residential (living) arrangements, and their health and social care. This is because in older age, many older people require both assistance with their everyday living activities and various health and social services. Consequently many places – including nursing homes, residential homes, assisted living arrangements, retirement communities, and even private dwellings – contain different elements and combinations of each.

Concerned with longer term movements between places of residence, the concept of 'aging in place' has emerged in the past decade in social and healthcare practice, policy, and research. As the term suggests, it articulates a priority to allow older people to live and be cared for in the place of their choice and to avoid their unnecessary relocation. Because of the concept's obvious geographical dimensions, geographers have naturally engaged with it and have offered more critical and theoretically informed explanations of the dynamics between aging, health, and place. Indeed, research by geographers shows the many ways in which places of residence provide attachments, meaning, and current and historical identity.

Geographers have engaged with the provision and experience of health and social care in a variety of ways. With an emphasis on the political economy of care, certain studies have indicated how policy changes, and the restructuring of health services, shift both the location of, and access to services, and older peoples' experiences of services. Indeed, with regard to the latter, studies highlight the potential tensions between the management of health services and older peoples' experiences of receiving them. In this way, commentators have been able to contribute to broader debates in the social and health sciences about what are, and what are not, the most appropriate places for older people to live and be cared for in. For example, the 'residential care versus community care debate' is implicitly geographical and has involved contributions from geographers.

On one side, the critique of residential homes has been twofold. One strand of criticism has focused on the inability of communal residential living environments to provide independence for residents and to protect their civil liberties, and the inability of regulating bodies to maintain adequate standards of environment and care. Meanwhile, a second strand of criticism views residential settings as places of stigma and social marginalization. Indeed, it is posited that regardless of whether quality standards can be improved, residential facilities will always carry negative associations that further marginalize older people through their association with them. In short, society will always dislike and think negatively of these settings as places of decay and decline.

On the other side, studies that support residential settings highlight that much of the aforementioned antiresidential literature is underpinned by an incorrect assumption that residential care is irrevocably institutional while people's homes are the only places to maintain personal control and freedom in older age. Moreover, commentators note that researchers who criticize residential settings often ignore the quality of older people's lives prior to them moving into such places. They argue that institutionalization occurs due to need regardless of the care setting, and residential settings cannot be blamed for generating dependency in older age. Others challenge the negative stereotyping of residential care and suggest that the residents of these settings are not institutionalized victims separated from society. Rather, residential settings are held up by their advocates as positive options for older persons that can be experienced by their residents as friendly, reliable, and safe communities.

Most recently, geographers have attempted to articulate the importance of place, to professional health care practices. Indeed, with respect to gerontological nursing, studies have attempted to identify the spatialities in everyday practice that have the potential to effect the experiences of older people. Drawing on Goffman's ideas on the creation of 'total institutions', a number of publications articulate the roles of nurses and their actions in the making of institutional places and place experiences. Meanwhile, other research attempts to connect geographical concepts and ideas to some fundamental concepts that drive professional clinical practice (such as client centeredness, autonomy, quality of care, and environment). Here, a geographical approach has helped rethink practices associated with physical and chemical restraint of people with dementia who often display challenging behavior. Various commentaries have articulated the spatial aspects and consequences of restraint and advocate the minimization of restraint and psychosocial interventions as practice solutions. Collectively, this emerging research not only connects the discipline of geography with health professional research disciplines, but also constitutes an attempt at providing a direct contribution to the evidence base for professional clinical gerontolgical practice.

Emerging Critical Perspectives

Reflecting the emergence of critical perspectives in both social gerontology and human geography, research published since the mid 1990s draws much more on social theory, focuses on constructions and representations of aging and challenges accepted social norms about aging. Moving away from treating older people as 'objects' of study, several studies engage with aging and the concept of embodiment, and talk in depth with older people to explore their perceptions of health and aging and their negotiations and changing relationships with spaces and places. Other studies attempt to understand how places reveal social attitudes and values about aging. For example, as highlighted in residential versus community care debate discussed above, retirement homes are recognized to mean something to both individuals and society as a whole but, in turn, feed back into social perceptions of what it is like to be old. On the other hand, commentators have also considered more positive places and activities of healthy aging, what the sociologist Andrew Blaikie refers to as part of new landscapes of later life. Examples here include retirement communities, RVing (campervan ing), and Universities of the Third Age. Geographers and others have identified the commodification of healthy aging and, as part of this, the radical fracturing of traditional spaces of aging and common understanding of them. All this is part of an understanding, noted earlier, that because of changes in health status, life expectancy, services, and greater wealth, older people are able to enjoy more varied and active personal and social lives than previous generations.

Another strand of recent geographical research has focused on the 'cataloging' of older age and places of old age in photography, painting, and the media. Concerned with ageism and stereotyping, these studies articulate who are included, not included, and how they are positioned in such materials and practices. With respect to ageism and terminology, it is also notable that researchers have gradually moved away from using terms such as 'the elderly' and 'elderly people', that might be perceived to categorize and problematize older age, to using such terms as 'older people' which are generally accepted to be less harmful. Commentators have also begun to deconstruct what Blaikie describes as imagined landscapes of aging identity. In this research, geographical metaphors are explored including life's late journeys, paths traveled, autumn years, and the sunsets of our lives.

Future Inquiries

Future geographical research on aging and health will need to be diverse. The development of traditional perspectives on older population health – in terms of dynamics, distributions, and movements – will no doubt continue, while emerging postmodern perspectives and qualitative approaches will be needed to more sensitively and critically investigate the complex relationships between older people and the varied places within which they live and are cared for. The future research challenges that lie ahead will involve finding appropriate ways to study the increasingly diverse lives of older people and articulate new cultural practices, social processes, and social and spatial relations that are emerging for older people. To meet these challenges most effectively, there has to be greater collaboration and debate within and between those who study aging and health, their various disciplines, and diverse empirical interests. Given that people are living longer and leading increasingly active and diverse lives, the study of aging and health is necessarily a growth industry.

See also: Ageism and Age; Ageing and Mobility; Care/Caregiving; Epidemiological Transition; Health Geography; Health Services Restructuring; Health Systems and Health Services; Medical Geography; Voluntary Sector.

Further Reading

Andrews, G. J., Cutchin, M., McCracken, K., Phillips, D. R. and Wiles, J. (2007). Geographical gerontology: The constitution of a discipline. *Social Science and Medicine* 65(1), 151 168.

Andrews, G. J. and Phillips, D. R. (eds.) (2005). *Ageing and Place: Perspectives, Policy and Practice*. London: Routledge.

Cutchin, M. (2003). The process of mediated ageing in place: A theoretically and empirically based model. *Social Science and Medicine* 57, 1077 1090.

Harper, S. and Laws, G. (1995). Rethinking the geography of ageing. *Progress in Human Geography* 19, 199 221.

Hockey, J., Penhale, B. and Sibley, D. (2001). Landscapes of loss: Spaces of memory, times of bereavement. *Ageing and Society* 21, 739 757.

Joseph, A. E. and Hallman, B. C. (1998). Over the hill and far away: Distance as a barrier to the provision of assistance to elderly relatives. *Social Science and Medicine* 46, 631 639.

King, R., Warnes, T. and Williams, A. (2000). *Sunset Lives: British Retirement Migration to the Mediterranean*. Oxford: Berg.

Kinsella, K. and Phillips, D. R. (2005). Global ageing: The challenge of success. *Population Bulletin* 60(1), 1 40.

Laws, G. (1993). 'The land of old age': Society's changing attitudes toward urban built environments for elderly people. *Annals of the Association of American Geographers* 83, 672 693.

Phillips, D. R. (ed.) (2000). *Ageing in the Asia Pacific Region*. London: Routledge.

Rowles, G. D. (1978). *Prisoners of Space?* Boulder, CO: Westview.

Rowles, G. D. (1986). The geography of ageing and the aged: Towards an integrated perspective. *Progress in Human Geography* 10, 511 539.

Smith, G. (1998). Residential separation and patterns of interaction between elderly parents and their adult children. *Progress in Human Geography* 22, 368 384.

Warnes, A. M. (1981). Towards a geographical contribution to gerontology. *Progress in Human Geography* 5(2), 317 341.

Warnes, A. M. (1990). Geographical questions in gerontology: Needed directions for research. *Progress in Human Geography* 14, 24 56.

Ageing and Mobility

A. M. Warnes, University of Sheffield, Sheffield, UK

Glossary

Age-Specific Migration Rate The percentage of a population in a specified age range living in a named area that move during a defined interval, usually 1 year or 5 years.

Gerontology The study of aging and old age in all forms of life. 'Social gerontology' focuses on the social circumstances and contexts of human aging.

Propensity to Migrate The likelihood or 'risk' of a person or group migrating over a defined period, normally a year. The propensity is highest around ages 18–25, and lowest in late middle age. It rises at the oldest ages, reflecting the rising likelihood of moving to seek care and support.

Residential Tourist A person (or a household) that resides far away from his/her main home for an extended period (e.g., for more than 2 months).

Retirement As an event, the age at which a person ceases paid work. As a life course stage, the state and period of life of having permanently given up work, usually in old age. In the 'official' or administrative sense it is the age at which a person is entitled to state or occupational pensions and other payments and concessions (as with public transport fares). In the 'social form' sense it is the actual age at which a person retires, which is commonly earlier than the official age.

Retirement Migration Relatively long-distance migrations just before or in the early years of retirement, generally not including short-distance housing-adjustment moves, or migrations in advanced old age.

Return Migrant A migrant who moves back to an origin area or settlement from which he/she earlier emigrated.

Seasonal Migrant A migrant who moves to a destination for part of the year and then returns to the origin area in at least 2 successive years. Seasonal migrants include agricultural workers, particularly harvesters, and retirees who generally move to warmer regions for the winter. There is no sharp distinction between a seasonal retired migrant and an older residential tourist.

Transnational Lives A lifestyle and activity pattern that is divided among two or more countries, involving more than holidays and occasional business or social trips, and characteristically with many, or close, family members and housing resources in multiple countries.

Introduction

Migration can be defined as a change in residence from one location to another. Such movements are immensely diverse, from short distance house changes to intercontinental moves. Many moves are short term, have little impact on the migrants' lives, and are reversible; others are permanent and life changing. As migration is a spatial process, it has been keenly studied by population geographers, whereas demographers are more interested in the dynamics of fertility and mortality. Migrations, however, also change the size, composition, and distribution of populations. Although most migrations are mundane, in aggregate and over time several types have radically transformed continental, national, and regional populations, prominent examples being the slave trade across the Atlantic Ocean, the colonization of Australia, and late twentieth century migration from the Caribbean, Turkey, the Middle East, and the Indian subcontinent to northwest Europe, and most widespread of all since 1950, rural–urban migration throughout the less developed world.

This article examines the connections between migration and old age, particularly in contemporary affluent societies. Most migrants are young adults, demonstrating the link between migration and the life course, or more precisely that migration most often occurs among people going through a transition from one life course stage to another. Both local and long distance moves are stimulated by many different transitions: becoming independent from one's parents, beginning a new career, marriage or a new partner, economic success or failure, retirement, and becoming chronically frail or sick. The first three of these 'life course transitions' explain why the likelihood of an individual migrating peaks in late adolescence and young adulthood. The last two explain a secondary peak around the ages at which people cease paid work, and why there is a rising propensity to migrate through the oldest ages, above 75 years. These relationships between age and migration have been demonstrated in North America, several European countries, and Australasia. The details vary and sometimes reflect social policy and economic conditions. If, for example, there is 'full employment', a low rate of divorce, and universal arrangements for retirement at a specified age, the 'retirement peak' is prominent.

Migration and Old Age

Although migration researchers focus on young adults' moves, for more than 25 years, population geographers and social gerontologists have studied old age migrations and migrants. The first interest, in the 1940s, was whether moves into nursing homes or between hospitals entailed a mortality risk. By the early 1980s, Pastalan, in Rowles and Otha in 1983, was able to review 34 US studies and identify the factors that influence the outcomes of older people's moves. He showed that among clinicians, mi gration was generally seen as a stressful life event with a real risk for survival among frail older people. Abstract or theoretical formulations went further, setting the migration experience in a psychosocial framework of personal life stage adjustment, which involves the ac ceptance of losses and the optimization of resources and abilities. This conceptualization recognized that migra tion brings new interests as well as stress. As Kahana and Kahana (in Rowles and Otha 1983: 221) put it, "the total stimulation of anticipating a move, moving, and environmental change can benefit those who are under stimulated relative to their capacities and ex haust the lesser capacities of others [and we need] to understand all relocation phenomena from [hospital] transfer trauma to eager globetrotting by the 'adventur ous aged'."

Another long standing interest has been in the mi nority of international migrations that have involved many elderly people, most of which have been 'impelled' or 'forced' displacements from natural disasters, famines, wars, political oppression, and racial enmity. Prominent recent examples have included the evacuation of Montserrat, 'ethnic cleansing' in the former Yugoslavia, and the emigration of Russian Jews to North America and Israel since 1973, when institutionalized anti semitism was relaxed in the Soviet Union, including punitive charges for exit visas. Studies of the latter dispersal have focused on the experience and welfare of the migrants in the destination societies, although there are accounts of the migration process. Many impelled migrations are poorly recorded and difficult to analyze: a recent example is the exodus from New Orleans following Hurricane Katrina.

Internal Retirement Migration

A third interest shared by population geographers and social gerontologists has been to understand the motiv ations, patterns, and consequences of retirement migra tion. Until the 1980s, almost all such moves were within various European countries, the United States, and Australia. The settlement of estranged aristocrats and retired senior soldiers and sailors to coastal resorts and spas was evident even before the railway era. Through the nineteenth century, the habit slowly spread from the very rich down the socioeconomic spectrum until, by the 1920s, a mass phenomenon was evident in Europe and the United States, particularly in Belgium, France, England, California, and Florida. By the 1970s, retirement migration was well established on both sides of the Atlantic in human geography research and teaching.

In the United States, long distance retirement mi gration developed vigorously during the 1940s, prompt ing the first commercial handbook guides and academic studies, while in Europe it was halted by World War II. By 1950, St. Petersburg on the west coast of Florida, and St Augustine and Miami on the Atlantic coast, had become fashionable resorts and places of permanent and winter residence for both affluent retired people and, more remarkably, retired black and white domestic servants. "Then after World War II, the development of the interstate highway system made distant locations much more accessible for vacationers and migrants alike. The changing state destinations can be charted in detail since 1960, because the decennial census counts the number who lived at a different address five years earlier. One of the defining characteristics of interstate retirement migration has been that migrants from all over the nation concentrate into only a few states. Nearly 60% (56% in 1990) of interstate migrants aged 60 or more years go to just 10 of the 50 states, and in all four census decades since 1960, Florida has attracted about a quarter of the total." (Longino and Warnes 2005: 538–9)

By 1985–90, however, a dispersal of the favored des tinations was clear, both among and within the states. The leading four destination states – Florida, California, Arizona, and Texas – all had lower percentages in 1990 than in 1980, although their ranking was unchanged (**Table 1**). Although the losses were initially small, the pattern has persisted and grown. Within Florida, housing developments had spread from the initial resort cities, first to form almost continuous coastal urban strips, and then inland. Many other favored destinations attract migrants primarily from adjacent states, for example, Cape Cod, Massachusetts, the New Jersey shore, the Pocono Mountains of northeastern Pennsylvania, and the Wisconsin Dells. Southern and western Nevada and areas in the Pacific Northwest are all recommended in retire ment guides as good places to retire.

International Retirement Migration

The first international retirement flow to attract the at tention of American researchers was of Canadians to Florida, and the first to attract the attention of European population geographers was the settlement of northern European retirees in various regions of southern Europe. The latter moves began to increase rapidly during the 1960s for a host of reasons: the new democratic regimes in Spain, Portugal, and Greece and the admission of these countries to the European Union, the growth of the mass

Table 1 UK state pensioners resident in four groups of countries, the USA and Ireland, March 2005

Year	*Country group (thousands)*					
	Den, Nor, Swe.	*Southern Europe*	*Caribbean*	*ACNZSA*	*USA*	*Ireland*
1995	1.7	67.8	23.8	357.9	85.6	71.2
2000	2.9	95.6	30.2	419.6	107.0	84.1
2005	4.8	153.8	31.5	471.1	123.9	97.7
Increase 1995 2005 (%)	182.4	126.8	32.4	31.6	44.7	37.2
Mean annual % increase	10.9	8.5	2.8	2.8	3.8	3.8

Notes: The country groups in the first four columns, respectively, are: (1) Denmark, Norway and Sweden; (2) Cyprus, France, Greece, Italy, Malta, Portugal and Spain; (3) Barbados, Dominica Republic, Grenada, Jamaica, St Lucia and Trinidad and Tobago; and (4) Australia, Canada, New Zealand and South Africa. The tabulation includes only countries with at least 1000 customers in 2005.
Derived from Department for Work and Pensions, Overseas Division, Newcastle upon Tyne, personal communication.

'package' holiday trade and its associated airports and modern 'infrastructure' facilities, and the spread of occupational pensions, early retirement and home ownership in the Northern countries (which enabled more people to purchase property abroad). Ever since, the proliferating studies have made clear that there is no clear break between seasonal, temporary, and permanent moves, or between 'residential tourism' and migration.

A warm climate is the main attraction, and most 'migration units' are couples aged in their late 50s and early 60s. Some move to the popular holiday islands or 'costas' and their hinterlands, as particularly in Spain; others disperse across rural areas of high landscape value, as particularly in the south of France and in Tuscany and Umbria in Italy. Many of these retirement migrants are relatively affluent – they epitomize the latest cohort of healthier and more active older people, commonly (if not always accurately) called the 'baby boomers', which some say are less family oriented and more concerned with 'quality of life' than their predecessors. The cultural, psychosocial, and behavioral transformations involved in retiring abroad are beginning to be documented. Like all international migrants, by moving to countries with different languages, customs, institutions, and social welfare and healthcare policies they are taking risks. One reason for giving close attention to their situation is that little is yet known about how the decrements of old age – in vigor, health, income, and social networks – and its diseases and disorders shape people's household situations and migration decisions. Studies of the phenomenon are rapidly diversifying and there are now impressive contributions from many European countries.

Aged Labor Migrants

A contrasting group of older migrants comprises labor migrants who have reached old age in the countries to which they moved in their early adult years. They are also socially diverse, but in Europe very many migrated from the Caribbean, north and west Africa, Turkey, the Middle East, and the Indian subcontinent to the cities of northwest Europe from the 1950s to the 1970s. Many came from depressed rural areas, had relatively little education, and acquired few formal or technical job skills. They are now some of the most disadvantaged and socially excluded older people in Western Europe. They are markedly heterogeneous, not just in origins and cultural and ethnic characteristics, but also in the extent to which they have raised children and formed social networks in their chosen country – important factors in their ability both to develop satisfying roles when no longer in work and, should their capacities decline, to turn to informal family and community support. There is also considerable diversity in their knowledge of, entitlements to, and utilization of state income, social housing, social service, and healthcare benefits and services.

Migrations by older people and their residential space–time patterns have become more diverse with the growth of second (or more) home owning and with the reduced costs of international air travel and telecommunications. Seasonal and transnational residence patterns are probably growing more quickly than clean break, 'total displacement' retirement moves, and it seems likely that the once clear distinction between 'retirement' and 'return' migrations is becoming blurred. There are multiplying reports from The Netherlands, Belgium, Norway, Sweden, and Switzerland that an increasing number of international labor migrants who reach retirement have maintained social contacts and residences (or residential opportunities) in both their adopted and their origin countries – and some in third countries too. Transnational lifestyles have recently attracted much attention from migration researchers, but only rarely with reference to older people. Ever since the studies in the early 1990s of Canadian retirees' 'snowbird' seasonal migration to Florida it has been known, however, that some skillfully manipulate their entitlements to health care, social security, and other benefits (and their tax liabilities) in both origin and destination countries by programmed moves between the two. Some also derive much satisfaction from having different social networks and activities in two locations.

Return and Family-Joining Older Migrants

Alongside the two main types of older migrants, there are others about which much less is known. The most apparent are 'return migrants', the labor migrants of long residence who return to their native regions and countries when they cease work. They are themselves diverse, and their migrations straddle internal and international moves. In all countries, some who moved to the capital or largest commercial cities of a country from rural provinces return to their native regions when they retire, and some make similar returns across an international boundary. Large numbers of Irish citizens return to Ireland after spending most of their working lives in the United Kingdom, and the same applies to Canadians whose careers were in the United States. Only a few of these moves have attracted systematic study and published accounts, but careful studies show that many returns are problematic, partly because the people and the settlements they left decades before have radically changed. The rate of return is influenced by whether the origin country is politically stable, has a strong welfare system, and a clement climate. Among European labor migrants in Australia, for example, there is a higher rate of return to Italy and Malta than to the United Kingdom or the former Yugoslavian Republic.

Social security and insurance agencies in Germany, the United States, and the United Kingdom publish statistics on the number of their clients receiving old age benefits that are resident in other countries. The total number of UK State Pensioners living abroad exceeded 1 million for the first time early in 2007 (approximating one in eight of all recipients). These data also show that other types of migration in later life are more voluminous than either 'amenity led' or 'return' migrations. For both the Germans and the British, the largest overseas beneficiary populations are in the United States, Canada, and Australia (**Table 1**). There are many German recipients in Austria, Switzerland, and Brazil, and there are many British recipients in Ireland and Germany. The clear inference is that there are substantial flows of family joining migrants who follow their children's earlier migrations, and that the dispersion is influenced by long established colonial, economic, and armed forces connections. While the processes and consequences of family joining migrations have been studied among recent intercontinental labor migrants into Europe and North America, the comparable flows out of north west Europe continue to be neglected.

Older Minority-Group Migrants in Rich Countries

In most developed countries, many surveys of the health and treatment needs of minority group older migrants are now undertaken by health and social services agencies, which of course have privileged access to their patients, and have the resources needed to translate questionnaires and employ bilingual interviewers. The findings of such surveys are largely descriptive and policy and practice oriented, and have been criticized as too ready to generalize about very diverse groups, for example in assuming that all minority groups have stronger family support networks than the host population. Overall, however, their findings confirm that the older migrants from peasant and near subsistence agricultural origins, who in the reception countries have been employed mainly in low skilled and low paid jobs, have high rates of sickness, poverty, social isolation and disconnection from the welfare system, and low entitlements to social security.

Academic studies have added understanding of older migrants' integration with the host societies and of the influence of the destination societies on their values, identities, and views. A particularly delicate matter is any disparity between their expectations and actual experience of intergenerational relations, and more specifically of the likely support and care from their children if and when they become dependent. Studies at opposite ends of the Earth, in Detroit, Michigan, and Brisbane, Queensland, have reached similar conclusions not only about the insensitivity of services to migrants' cultural differences and their resulting exclusion, but also about ambivalences and confusions in identities and values. Many Middle Eastern migrants in Detroit left their homes as very young adults and did not care for their parents when they became old. As the migrants reach their own old age, some have views about what their children 'should' do for them that are idealized and normative, but most are also keenly aware that their wish to be closely involved with their children and grandchildren is inconsistent with the latter's adoption of American ways, material ambitions, and need to work long hours. The cultural gap between the generations is in some cases a source of serious distress and dismay.

Turning to older Chinese migrants in Brisbane, Ip, Lui and Chui's (2007: 733) found that "social isolation and feelings of loneliness and confinement were common and particularly acute among women … and exacerbated by increasing dependence upon their adult children for transport. … Inadequate support from family members made it more difficult for them to take part in social activities or to access the amenities provided by the city for older people in general." Paradoxically, because of the common belief that Chinese older parents are usually well looked after, especially in middle class families, the predicament of many was unnoticed. In many developed countries, because Chinese immigration is long established, the population has many older people and they tend to be more dispersed than other migrant groups. Many came to work in restaurants and fast food shops, where they worked long hours in poor physical conditions; and many men never learnt the local language or married and in old age had little informal support. Across the world, Chinese expatriate communities have generated their own charitable welfare associations that provide vital links to the most isolated older migrants with formal health and social service agencies.

Migrants in Advanced Old Age

A particular concern of both older people's welfare organizations and the public authorities in areas with high concentrations of older migrants is that some, through poor planning, bad luck, or sickness, end up in situations in which they are isolated, impoverished, and out of touch with both formal and informal support. Most who move long distances for retirement, whether within a country or across an international border, do so as couples. In most cases they move away from their closest family, and some move to foreign countries where they do not speak the language and have reduced entitlements to public sector healthcare, personal social services, and 'social' housing. Retirement migrants who move in their 50s or early 60s are generally fit and active, but as time passes, health and income decline and a point may come when one of the couple becomes severely ill or dies. Others are duped by property developers or builders. While most avoid severe problems, or have family or friends to help, there are occasional cases of serious distress.

Recent Developments in Retirement Migration

The percentage of United States retirees who migrate abroad is very small, but given the very large population and the recent surge of international travel and communications, the absolute number is growing. The phenomenon has recently attracted a major study by the Migration Policy Institute in 2006 in Washington DC. This found that "the US born senior population (aged 55 or more years) grew substantially in both Mexico and Panama between 1990 and 2000, and continues to increase rapidly. Mexico saw the number of US born senior residents increase by 17 per cent during the 1990s, while the number in Panama increased by 136 per cent. Some municipalities in Mexico have seen much faster growth, such as Chapala (581%) and Los Cabos (308%). The number of US citizens obtaining pensioner visas in Panama more than tripled between 2003 and 2005." (Migration Policy Institute 2006: 2–3)

A recent survey of US retirees in the Lake Chapala region of central Mexico found, "important differences between the European and north American cases: in the former, the retirees move from a developed nation to another developed nation, which reduces the cost of living or financial advantage, and raises the relative importance of the environmental attractions. By comparison, the US retirees with their fixed incomes are mainly looking for affordable retirement locations. The sense of community and friendship at the retirement destination was also found to be a strong attraction. Many of the respondents had visited several possible retirement locations, primarily as tourists, but it was the areas at which they had developed social networks that became the most likely places for retirement. The local expatriate associations organize social and cultural activities and help the retirees, particularly the newcomers, during their transition to the new community. These new social networks give the migrants a sense of community and for many become more active components of their social life than kinship ties and relationships in the United States." (Sunil, Rojas, and Bradley 2007: 505)

In 2006, the UK Foreign and Commonwealth Office commissioned a wide ranging study of the circumstances of 'Britons Abroad' (Institute for Public Policy Research 2006). This was partly stimulated by the increasing load on consular staff as more and more nationals traveled and resided in foreign countries, and partly by moves in Brussels to increase the portability of healthcare entitlements among the member states of the European Union. From the inception of the EU, among its key principles have been a 'common market' and the 'free movement of labor'. The result has been millions of cross border moves, which at first were conceived as temporary labor placements but of course have led to cross national marriages and permanent settlement. Soon large numbers of workers and their families had complex multinational biographies and histories of payments to more than one country's social security system. This in turn has required binational policy assessments of the respective responsibilities of the social security administrations in the two countries for paying unemployment, sickness, and retirement benefits in individual cases, and bilateral policy agreements about how these estimates are made. These arrangements have raised the 'social protection' of millions of workers, but inevitably some complex biographies, many involving the breakup of the migrant family through estrangement or bereavement, mean that some individuals fall through the welfare 'safety net'.

There have also been studies of the migration of Western country retired men to Thailand. The major reported motives were low living costs, a warm climate, to escape a disliked home nation, like of the Thai lifestyle and culture, and the availability of attractive sexual partners. Most retirees took a Thai spouse or live in partner. The move apparently works out well for most, at least initially. They report positive wellbeing and feel assimilated, but most live with visa insecurity and socialize mainly with other foreigners. The major long term concerns of Western retired men in Thailand are their healthcare and welfare needs, income problems, the rise of negative local reactions to the influx of Westerners, and the possibility of visa cancellation that would enforce a move elsewhere.

Mobility and Older People's Quality of Life

Studies of migration tend to feature the more visible and radical forms, like retirement to Spain or Costa Rica, or the return from London or New York to Jamaica, when most migrations by older people are local housing ad justments – like those by adults of all ages. Many local moves make an important contribution to maintaining older people's independence and raising their quality of life. With more than one fifth of the population of most developed countries now aged 60 or more (and many in their 50s having retired), and average life expectancy close to 80 years of age, clearly 'old age' can extend for decades and be a substantial fraction of a person's life. It is a life course stage of many changes. Early in retire ment, some move to more attractive environments, where they can pursue nonwork interests, and others move to smaller homes or to live nearer their children and grandchildren. At older ages, many experience losses in health, vigor, partners, and income, and become no longer able to afford or drive a car, to climb stairs, to maintain a large garden or mow the smallest lawn, and to walk to and from shops. Several of these changes ser iously impair an older person's or a couple's quality of life, but can be ameliorated by a move to a more con venient home or location. One can say that impaired personal mobility encourages residential mobility. Some move to housing schemes designed for older people, for the surveillance or support they provide, or to enjoy more personal security. Population geographers have made substantial contributions to understanding the role of migration in older people's lives, but the work has been selective, and local moves and housing adjustments have been neglected. There is immense scope for more col laboration between geographers and gerontologists in understanding the role of residential mobility in older people's lives.

See also: Emigration; Immigration I; Immigration II; Migration; Population Geography; Transnationalism.

Further Reading

Ackers, L. and Dwyer, P. (2002). *Senior citizenship? Retirement, migration and welfare in the European union*. Bristol: Policy Press.
Ajrouch, K. J. (2005). Arab American immigrant elders' views about social support. *Ageing & Society* 25(5), 655 674.
Casado Díaz, M.A., Lundh, U. and Warnes, T. (2002). Bibliography on older migrants in Europe: Projects and sources. http://www.shef.ac.uk/sisa/esf/EW Bibliography.shtml (accessed on January 2008).
Hardill, I., Spradbery, J., Arnold Boakes, J. A. and Marrugat, M. L. (2005). Retirement migration: The other story, issues facing British migrants who retire to Spain. *Ageing & Society* 25(5), 769 783.
Howard, R. W. (2008). Western retirees in Thailand: Motives, experiences, wellbeing, assimilation and future needs. *Ageing & Society* 28(2), 145 166.
Ip, D., Lui, C. W. and Chui, W. H. (2007). Veiled entrapment: A study of social isolation of older Chinese migrants in Brisbane, Queensland. *Ageing & Society* 27(5), 719 738.
King, R., Warnes, A. M. and Williams, A. (2000). *Sunset lives: British retirement to the mediterranean*. Oxford: Berg.
Longino, C. F. Jr. (1995). *Retirement migration in America*. Houston, TX: Vacation Publications.
Longino, C. F. Jr. and Warnes, A. M. (2005). Migration and older people. In Johnson, M. L. (ed.) *The Cambridge handbook of age and ageing*, pp 538 545. Cambridge: Cambridge University Press.
Migration Policy Institute (2006). *America's emigrants: US retirement migration to Mexico and Panama*. Washington, DC: Migration Policy Institute.
O'Reilly, K. (2000). *The British on the Costa del Sol: Trans National identities and local communities*. London: Routledge.
Rogers, A. (ed.) (1992). *Elderly migration and population redistribution: A comparative study*. London: Belhaven.
Rowles, G. and Ohta, R. J. (eds.) (1983). *Aging and milieu: Environmental perspectives on growing old*, pp 189 203. New York: Academic.
Sunil, T. S., Rojas, V. and Bradley, D. E. (2007). United States' international retirement migration: Reasons for retiring to the environs of Lake Chapala, Mexico. *Ageing & Society* 27(4), 489 510.
Walters, W. H. (2002). Later life migration in the United States: A review of recent research. *Journal of Planning Literature* 17(1), 37 66.
Warnes, A.M. and Williams, A.M. (eds.) (2006). Special Issue: Older Migrants in Europe: An Innovative Focus for Migration Studies. *Journal of Ethnic and Migration Studies* 32(8), 1257 1427.

Relevant Websites

http://www.ippr.org
Institute for Public Policy Research, London, 'Brits Abroad' project.
http://www.migrationpolicy.org
United States Migration Policy Institute.
http://www.britishpensions.org.au
World Alliance of British Expatriate Pensioners.

Ageism and Age

L. Maxey, Swansea University, Wales, UK

Glossary

Adultism Systematic mistreatment and disrespect of younger people due to age. Ageism based on a binary distinction between the categories adult and child and privileging adults over children.

Ageism Any form of prejudice or stereotyping based on age.

Intersectionality The interaction between different forms of oppression and privilege, including gender, race, ethnicity, class, sexuality, disability, and age.

Life Course Dynamic and often unpredictable ways in which a person's life progresses, rather than along rigid or predetermined stages.

Performativity Emphasizes the ways in which ideas and identities such as those connected to gender and age are not fixed, but continually produced in each momentary performance of them.

Relationality An approach to social axes such as age which destabilizes rigid divisions and highlights interdependence. In the case of age, relationality emphasizes the coexistence of different ages on a shared continuum and the mutual reliance of one point on this continuum to all others.

Segregation The separation of people based on a particular trait or series of traits such as race, gender, and age. Segregation may be spatial, sociocultural, political, and psychological in its operation.

Younger People A term used to approach people in an age relational way, avoiding closing off subjects and creating an apparently rigid category as is commonly the case with 'terms' 'child' and 'youth.'

Introduction

Age is important. Alongside gender, race, ethnicity, sexual orientation, class, dis/ability, and other axes of social difference, age shapes the social world at every scale and in many different ways. In common with all lines of social difference, age is also the site of oppression and prejudice. Ageism, or prejudice based on age, takes many forms. It is one of the most prevalent types of oppression in that everyone is exposed to ageism at some point and often in multiple and shifting ways throughout their lives. Ageism operates and is experienced through a complex range of mechanisms which have only recently become the focus of academic research and social policy. This includes individuals' internalized ageist norms and the institutionalized ageism within political, legal, and spatial planning systems, for example. While there has been limited geographical analysis of age to date, the development of multidisciplinary understandings of age and governmental policies addressing age and ageism highlight the important ways space, place, and scale shape experiences of age and the (re)production of ageism.

The gap in geographical analysis of age has shrunk since the 1990s, with the flourishing of work on children and, to a lesser extent, older people. Despite this almost exclusive focus on the margins of the age continuum, an emerging body of work within children's geographies and older people's geographies demonstrates the importance of geographies of age more generally. Within these two emerging subdisciplines, geographical work offers valuable contributions to social science and policy debates, highlighting the ways space and scale intersect with age and other axes of social difference. Although geographical analysis has clustered around the extremes of the age spectrum, more recent work suggests this may be used to create a broader, more relational approach to age which explores the performativity of age and infuses human geography with a keener sensitivity to age more generally.

Children's and older people's geographies have a number of features in common. Each have their roots in pioneering studies carried out in the 1970s and in both cases this groundwork became reinvigorated in the 1990s and early twenty first century. A wider social science and political interest in both children and older people has enabled these subdisciplines to engage with debates and practices beyond geography. Notwithstanding these common features, there has been far greater attention to work at the younger end of the age spectrum and work in this field provides a suitable point from which to begin an exploration of geographies of age in a fuller sense.

Children's Geographies

There has been sufficient research, publication, and teaching activity in the field of children's geographies since Alison James' seminal piece in the journal *Area* in 1990 that a distinct subdiscipline has emerged, complete with its own journal (*Children's Geographies*, founded in 2003) and a specialty research group of the RGS IBG (Children, Youth and Family Working Group, formed in 2006). As the name of this group suggests, the focus within the subdiscipline has begun to broaden from childhood to include youth and family. This demonstrates a growing willingness within geography to engage with geographies of age more broadly. Also since 1990, there have been claims of an 'emergent paradigm' forming within the social sciences which recognizes children and childhood as a legitimate locus of concern and introduces new epistemological and methodological perspectives.

The new social studies of childhood have thus emerged as an interdisciplinary, social, and cultural paradigm within which children's geographies play a significant role. The interdisciplinary nature of this work is important as it has supported the development of epistemological and methodological critiques and approaches which have helped propel children's geographies since at least 1990. Sharing insights from, among others, history, sociology, anthropology, geography, education, childhood studies, psychology, legal studies, and policy studies has allowed the formation of a series of ethical, epistemological, and methodological approaches which now inform both children's geographies and the wider new social studies of childhood of which it forms part. Several insights stem from an ontological interrogation of terms such as child and childhood, highlighting the extent to which these terms are socially, culturally, and politically constructed:

1. Such apparently natural concepts vary considerably over time and space. This insight helps to locate the ideas and practices surrounding childhood temporally and spatially. Furthermore, it highlights the extent to which such concepts and customs are malleable. This means that geographers are able to contribute to understandings not only of how adultism, for example, operates, but also how it may be overcome in favor of greater equality and empowerment.
2. Work in this area has begun to unpick oppressive ways in which disciplines have theorized children and childhood, from structural sociology's view of children as passive subjects of socialization, as illustrated in the work of Talcot Parsons, to the uncritical ways in which Jean Piaget's developmental models have been used to produce a linear, predictable view of child development. Critiques such as these have paved the way for new theoretical approaches to childhood and supported attempts to address adultist forms of exclusion and oppression wherever they occur.
3. Theoretical and empirical work has demonstrated that there is no singular, universal childhood. Rather, childhood is produced and experienced in innumerable ways and geographers have been highly active in this regard, exploring, for example, the diverse ways in which childhood is experienced between and within Minority and Majority World contexts.
4. Stemming from all the above, children are not passive 'adults in the making' but competent social actors who, from birth, exercise their own agency to actively shape the world. This agency, however, is not fixed or given, but is as fluid and diverse as the experiences of younger people themselves.

A commitment to equality and addressing adultism, therefore, lies at the heart of children's geographies and the new social studies of childhood. Geographical work has highlighted several key issues impacting younger people including poverty, education, violence, health and nutrition, and participation. While these issues often relate to both Majority and Minority World contexts, to date there has been disproportionately more work on Minority World contexts. In the Minority World, for example, geographical research has shown how adultist practices and discourses have contributed to the erosion of children's independent mobility, through transport policies favoring the motor car, the growth of privatized civic space such as shopping centers and moral panics surrounding stranger danger and 'uncontrolled youth'. Increasingly, however, work in this field has begun to broaden its analysis both spatially and conceptually, so that age is studied across Minority and Majority World contexts and in a fuller, more relational sense.

Broadening the spatial analysis of children's geographies has in part been promoted by the United Nation's Convention on the Rights of the Child (UNCRC) (agreed 1989). The most widely ratified UN human right's instrument, this convention has been signed by all nations except the US and Somalia. It sets out specific nation state responsibilities and regular detailed reporting and accountability to the UN, thus creating an agreed global framework to raise the status of children. Over time this has had significant implications for children's geographies as each nation state where the UNCRC is in force adopts a rolling program to improve the rights and experiences of people under 18 years of age. While the UNCRC has been criticized for its Western, or Minority World bias and repeating outdated universalized understandings of childhood, it has helped place children on the global political agenda and paved the way for more recent work on older age.

The second, conceptual, broadening of approaches to children involves a relational turn in which the focus has shifted from addressing children as a discrete group to looking at younger people and the way their geographies intersect with others wherever they are placed on the age continuum. Rather than closing off childhood and essentializing 'the child', more recent work has begun to look at younger people in their specificity and brings to this analysis an awareness that age intersects with other identities and forms of oppression, including those based on gender, ethnicity, race, class, sexuality, and dis/ability. Using the term 'younger people' rather than 'children' draws upon work on older people where relationality is at least implicitly embedded in the very label used. Whereas the label 'children' may be used to carry the implication that the subjects are discrete and bounded, the term 'younger people' highlights the relationality of age and allows greater scope for broadening analysis to all actors, regardless of their position on the age spec trum. Younger people as a term, therefore, may be ap plied not only to children and youth, but also to locate younger adults vis à vis older ones.

As this discussion suggests, there are several com monalities between emerging subdisciplines of younger and older people, with both contributing to under standings of geographies of age and ageism. It is neces sary, therefore, to look at geographies of older people in their own right in order to further explore how geog raphies of age and ageism are understood and where their future contributions and developments may lie.

Geographies of Older People

While there has been significantly less geographical work on older than younger people, there is a growing will ingness among geographers to explore the ways space, place, and scale impact upon and are shaped by older people. Geographers have thus looked at ageism within spatial planning and the ways older people experience the built environment as well as how older age shapes experiences of healthcare, embodiment, public transport, memory, and bereavement. Work in this area has a number of features in common with younger people's geographies. Both address highly prevalent, yet deeply engrained forms of prejudice and both may be placed within broader academic and policy contexts. Critical gerontology, for example, emphasizes that older age may be considered a socially constructed category, rather than a fixed or natural given. It has also demonstrated that modernity's preoccupation with documenting chrono logical age is a relatively recent phenomenon driven by lawyers, bureaucrats, and legislators, who specified pre cise numerical ages for an increasing number of civil rights and duties. Gerontology, like the new social studies of childhood, is also a multidisciplinary field, drawing upon social and behavioral sciences, humanities, medi cine, health, housing, and social care as well as policy studies and those involved in voluntary agencies. Geo graphical analysis has important roles to play within gerontology, although to date it has been under represented.

The scope for older people's geographies to inform policy is considerable, particularly given the growing body of international and national legislation on age and age discrimination. Pivotal here are two interrelated trends. The first is a postmodern embracing of difference and respect for diversity. This can be seen in academic disciplines such as geography. Significantly, however, it also extends to wider social and political arenas. Within this climate it becomes progressively less tenable to maintain positions of prejudice and bigotry, at least through forms which are open to explicit and direct questioning. The second trend is demographic, that is, the rapid aging of the world's population. No longer confined to Minority World countries in their older states of industrialization, this trend offers significant chal lenges and opportunities, particularly in more populous and resource scarce Majority World countries.

Responding to these dual trends, key international agencies such as the UN and the World Health Organ ization (WHO) have developed a rigorous body of policy addressing ageism and creating a positive agenda for older people with parallels to the UNCRC. These pol icies include the WHO's Brasilia Declaration on Ageing (1996) and the UN's Second World Assembly on Ageing (Madrid, 8–12 April 2002). The latter unanimously adopted the Madrid Political Declaration and Inter national Plan of Action on Ageing. These policy devel opments can be criticized for their implicit reproduction of Western developmental norms. Indeed, there is a striking instrumentalism behind such policies intended to turn the so called 'problem' of a rapidly aging population into an asset which can contribute to economic growth and a mainstream development agenda. This instru mentalism parallels policies on Education for Sustainable Development and Global Citizenship which suggest that the world's economic and social challenges can be solved by educating children to avoid their parent's mistakes, rather than parents making the appropriate changes themselves. Notwithstanding their Western develop mental bias and instrumentalism, these policies promote an 'active aging' agenda which establishes older people's right to live in security, enjoy good health, and continue to participate fully in society.

Such policies place older age and the ageism associ ated with it firmly on international and national arenas and have led to recent anti age discrimination legislation such as that in the UK. The scale of the challenges these initiatives address and the scope for greater engagement with age by geographers can be seen in the suggestion that ageism is the most pervasive form of prejudice ex perienced in the UK population. While anti ageism le gislation such as the UK Employment Equality (Age) Regulations (2006) has mainly been regarded in relation to its implications for older people, it applies equally to younger people. Once the logic of addressing age based discrimination is accepted, it becomes increasingly dif ficult to avoid addressing all forms of ageism, including that against younger people. It may not be long, there fore, before there are renewed calls to reconsider the minimum voting age and the age at which younger people have the right to an independent income. Indeed, the ever broadening nature of the focus on age and aging is reflected in the WHO and National Governments', emerging research agendas looking at issues such as the gender dimensions of aging. The fact that women and men age so differently has significant social, political, and economic implications. The ratio of women to men over 80 years is 2:1, for example, while among super centenarians (those over 110 years) it is more than 10:1. Geographers are well placed to contribute to research and analysis on such age based phenomena.

Relational Geographies of Age

This section considers current trends and directions being addressed by and impacting upon geographies of age. Relational approaches to age help render the oper ation of age explicit within all geographical analysis and support greater awareness of age and ageism throughout human geographical research. This has strong parallels with the way understandings of gender and race now inform a vast range of geographical studies. Furthermore, relational approaches to age can help address ageism within the discipline as well as within society more generally.

Geographers have begun to demonstrate how norms and expectations associated with age are played out spatially, demonstrating how homes, streets, neighbor hoods, parks, and shopping centers acquire meanings of age and become sites at which ageism may be reproduced and challenged. Although geographers have contributed a large amount of research and analysis to ethnic or race based segregation, however, there has been very little attention to age based segregation, a highly significant, yet largely overlooked geographical issue. This is par ticularly striking given the growing attention given to age segregation within disciplines such as sociology and an thropology. Furthermore, the burgeoning policy interest in age noted above suggests a valuable role for geo graphical analysis of the ways age plays out in society generally and shapes segregation more specifically.

There are clear signs that age segregation is increas ingly manifest spatially as well as socially and eco nomically. Older age, for example, has been described as 'another country' by several commentators. Equally, as children's geographers have shown, younger people are increasingly segregated based on age not only from middle years and older adults, but also from younger people of different ages. This segregation operates through a range of practices and norms, including those surrounding housing and retail provision, education, transport, employment, and leisure. There has been very little work within geography, or indeed any discipline, looking at the ways age segregation affects both the young and old. The tendency to focus on either end of the age spectrum is in part the result of increasingly narrow research disciplines and subdisciplines re sponding to a increasingly competitive neoliberal agenda privileging international research. However, it is also indicative of the inbuilt ageism within academic discip lines such as geography and the wider social propensity to segregate based on age.

Beyond studies which attempt to bridge the gap be tween young and old, however, relational approaches to age suggest a broader and more inclusive agenda in which adulthood itself is directly interrogated. To develop re lational understandings of age and begin engaging with the root causes of ageism and segregation, it is necessary not only to look at the ways childhood, youth, and older age are produced and experienced, but also to explicitly look at the invisible middle years. In this way relational approaches to age can learn from feminist geographies which have begun to explore masculinities and engage with men as research subjects and similar moves to look at whiteness within geographies of race and ethnic re lations in the Minority World.

Addressing segregation is important as research sug gests age based prejudice increases as levels of segre gation increase. This has already led to calls for greater age integration. It is important to note, however, that age is always highly intersectional, differentiated by other social axes, including gender, class, and race. An inter sectional approach, which recognizes complex and overlapping identities, is central to relational geographies of age. While greater age mixing often creates oppor tunities for shared understandings and challenging the stereotypes upon which ageism is founded, many gaps remain in understanding how such processes operate. Indeed, evidence suggests that tokenistic and socially and culturally insensitive age mixing may actually increase prejudice.

A prerequisite in reversing ageism and age differen tiation is better understanding of how they operate, are experienced, and reproduced. This involves looking critically and sensitively at the range of contexts and scales at play, from the individual to the region. For ex ample, segregation operates on an individual level within an individual's own psyche as they internalize norms and practices which compartmentalize our life courses into different age based stages. These individual scale pro cesses contribute, in the post welfare state Minority World at least, to the reproduction of a tripartite segre gation of life course into periods of education, work, and leisure. As lifestyle and life course become more indi vidualized, this segregation may weaken. The rise of neoliberalism, and other contemporary trends have led to flexible employment, lifelong learning and reskilling, the extension of the retirement age and questions over the feasibility of pension funds, for example. These processes undermine established age based patterns and norms such as the tripartite life course and open up new pos sibilities which could reduce or exacerbate segregation. The difficulty in predicting the implications of such changes highlights the importance of continued research and analysis.

Another example of such trends is the increase in market segmentation used to sell goods and services. This can contribute to age segregation through the creation of nar rowly defined identities which often have rigid age based ordering, though this is not inevitably nor exclusively the case. Nevertheless, much work remains to be done by geographers, and others, employing relational understand ings of age to explore such processes. Furthermore, a gap remains when it comes to practical policies capable of in creasing age based mixing and reversing the trend toward age segregation and more holistic and relational approaches to age and segregation can help address this lacuna.

Sustainability, for example, is an imperative moving rapidly up the political and social agenda throughout the world as challenges posed by climate change, peak oil, and growing inequality become more apparent and urgent. The transition to a more sustainable society in volves significant changes in terms of social, environ mental, and economic organization. Therefore, these transformations will both benefit from and significantly contribute to greater age based awareness, equality, and integration. For example, sustainable settlement planning tends to involve mixed use developments integrating a full range of functions, including employment, housing, retail, education, recreation, energy generation, and other forms of production. This is a reversal of the zoning approach dominant within urban planning since 1945 and opens up considerable opportunities for greater age based integration to be placed at the heart of new approaches to settlement planning.

A more nuanced and relational approach to segre gation is paramount when one looks at other future trends, including the aging of society. Despite exceptions, the tendency within geographical studies of age to date has been for geographies of age to emerge from Minority World institutions, research Minority World contexts, and address Minority World concerns. While such a critique has been leveled at geography as a whole and is by no means unique to geographies of age, it is important that geographies of age address it and relational ap proaches can again support this. Much can be learnt from looking at the ways age is experienced and produced across different contexts around the world. The phe nomenon of society's aging, for example, is happening in both Minority and Majority World countries, and the implications of this phenomenon have both similarities and marked contrasts across these contexts. The trend toward increased age segregation in the Minority World has been driven in many respects by developmental pathways based on consumerism driven economic growth. As this pathway has been increasingly taken up in the Majority World since 1945, how is this shaping age relations in these diverse contexts? Relational geog raphies of age are well placed to help answer such questions and address the various gaps identified above. Perhaps key to this is a more integrated approach which relational approaches to age can help provide. This in volves a refusal to simplify either identities or the process at work which lead to differentiation and segregation.

Conclusion: Performing age

Recent work within geography has highlighted the per formed nature of identity. This approach has been very useful in deepening understandings of the ways gender and sexuality, for example, are not fixed or given but continually performed and endlessly capable of dis ruption and reinvention. There has been very little work approaching age in this way, although such an approach offers many insights into the relational nature of geog raphies of age. The imperative to address ageism informs geographies of age just as the imperative to address sexism informs feminist geographies. However, in high lighting the pernicious nature of age discrimination it is important that the geographers do not treat those most vulnerable to ageism as passive victims, but emphasizes their agency and competence. Engaging with the per formativity of age based identities and practices can help draw out this agency and the potential to disrupt as well as re inscribe age based norms. Participation also offers great potential to contribute to relational geographies of age, both highlighting the diverse ways in which actors participate in the performance of age and in terms of the contribution participatory research approaches can make to these geographies. Participatory research is particu larly well suited to the creation of knowledge which is relational, intersectional, and empowering and thus has an important role to play in developing relational geographies of age. Geographies which approach all people, irrespective of age, as different, yet equal and actively performing their identities may be better able to embrace, understand, and engage with the diversity and fluidity of childhoods, youthhoods, and adulthoods.

See also: Children/Childhood; Performance, Research as; Transport and Social Exclusion; Youth/Youth Cultures.

Further Reading

Bytheway, B. (1995). *Ageism*. Buckingham: Open University Press.
Harper, S. (2005). *Ageing Societies*. London: Hodder.
Henderson, S. J., Holland, J., McGrellis, S., Sharpe, S. and Thomson, R. (2006). *Inventing Adulthoods: A Biographical Approach to*

Youth Transitions. London: Sage in association with the Open University.
Holloway, S. and Valentine, G. (eds.) (2000). *Children's Geographies: Living, Playing, Learning*. London: Routledge.
Hopkins, P. and Pain, R. (2007). Geographies of age: Thinking relationally. *Area* 39, 287 294.
James, A., Jenks, C. and Prout, A. (1998). *Theorizing Childhood*. Cambridge: Polity Press.
James, S. (1990). Is there a place for children in geography? *Area* 22, 278 283.
Journal of Social Issues (2005). *Special Edition on Ageism* 61, 207 405.
Laws, G. (1997). Spatiality and age relations. In Jamieson, A., Harper, S. & Victor, C. (eds.) *Critical Approaches to Ageing in Later Life*. Milton Keynes: Open University Press.
Maxey, J. L. (2004). The participation of younger people within intentional communities: Evidence from two case studies. *Children's Geographies* 2, 29 48.
van Blerk, L. and Ansell, N. (2006). Children's experiences of migration: Moving in the wake of AIDS in Southern Africa. *Environment and Planning D: Society and Space* 24, 449 471.
Vanderbeck, R. M. (2007). Intergenerational geographies: Age relations, segregation and re engagements. *Geography Compass* 1, 200 221.
Ward, C. (1978). *Child in the City*. London: Architectural Press.

Relevant Websites

http://www.youthonboard.org
A project working with youth and adults to empower and foster learning around age based issues and their intersection with other lines of oppression.
http://youthfacts.org
A web based project setup by Mike A. Males, author of several books on younger people and adultism. The project aims to debunk myths and stereotypes about youth and provide balanced information.
http://childrens research centre.open.ac.uk
An example of an academic research institution which empowers younger people to engage in advanced research projects.
http://www.colorado.edu
Children, Youth, Environments is an academic, online journal available to all.
http://www.magicme.co.uk
Example of a project approaching age in a relational and intersectional way, using art to bring people of diverse ages and backgrounds together.
http://www.gcyf.org.uk
Geographies of Children, Youth and Families Working Group of the RGS IBG exists to explore current and future research, teaching, and practice in this field.
http://www.pygywg.org
Participatory Geographies Research Group of the RGS IBG. Participatory research has been integral to the development of children's geographies and can play a significant role in developing relational geographies of age.
http://www.who.int
The English language page of the WHO's website on its various aging programs.
http://www.iagg.com.br
The International Association of Gerontology and Geriatrics.
http://time space life course.ncl.ac.uk
Website of an Economic and Social Research Council Seminar Series on Time Space and Life Course. This 2 year project demonstrates the growing interest in relational geographies of age and will provide a resource for those interested in life course geographies beyond the life of the project.
http://www.freechild.org
Website run by Common Action, an youth empowerment organization.

Agglomeration

A. Malmberg, Uppsala University, Uppsala, Sweden

Introduction

The term agglomeration can be used both as a noun and as a verb. Dictionary entries of agglomeration as a noun will typically refer to a "collection or mass of things rudely or loosely thrown or huddled together," an "unmethodical assemblage," a "confused or jumbled mass," or merely a "clustering or cluster." As a verb, consequently, agglomeration refers to the act or process of collecting in a mass, of heaping together, or to combine mechanically without any adaptation of parts.

The Core Issue of Economic Geography

In the context of economic geography, agglomeration may refer to a concentration in space of some entity, or the process by which such a spatial formation is created. While the dictionary definitions signal that agglomeration is to be understood as an unsystematic, haphazard process, where some entities pile up together in space but are not really affected in any particular way, economic geographers tend to think of agglomeration as something that can be explained by some sort of rationale on behalf of the actors involved, and with profound effect on economic outcomes in terms of productivity, competitiveness, or prosperity.

There is reason to take the issue of agglomeration seriously. Much may be learned about the role of space and place in economic processes by trying to pinpoint the driving forces that make for agglomeration in space of economic activity. Indeed, many would argue that spatial agglomeration is the phenomenon that ultimately motivates and defines the existence of economic geography as an academic discipline.

> We would argue that the defining issue of economic geography is the need to explain concentrations of population and of economic activity: the distinction between manufacturing belt and farm belt, the existence of cities, the role of industry clusters. Broadly speaking, all these concentrations form and survive because of some form of agglomeration economies, in which spatial concentration itself creates the favorable economic environment that supports further and continued concentration. (Fujita *et al.*, 1999: 4)

To complicate things, the concept of agglomeration has come to denote several sociospatial phenomena as the above quotation illustrates. It can refer to the tendency for people and economic activity in general to concentrate in big cities or economic core regions. Remarkably large proportions of global value creation take place in a limited number of highly concentrated core regions. It can also refer to the fact that firms within the same or closely related industries often tend to gather at certain places, to form specialized clusters or districts.

Early attempts to define the economic advantages gained by agglomerative behavior were framed by a division into two sorts of agglomeration economies. Thus, urbanization economies came to denote those advantages to be gained by location in a large and dense urban area, while the notion of localization economies referred to advantages gained by locating close to other similar or related forms. While this distinction is well established, one should note that similar factors are often held to make up both mechanisms.

It is the latter aspect of agglomeration – the 'birds of a feather flock together' phenomenon – that makes up the main focus of this article. This sort of spatial agglomeration of similar and related firms and industries is most easily illustrated by pointing to some world known examples. Thus, most people cannot think of geographical notions like Hollywood, Silicon Valley, Detroit, or The City of London without associating them with the industries that define and dominate them (movies, information and communication technology (ICT), cars, and financial services, respectively). Indeed, in the most extreme cases, it is difficult to think of an industry or its products without thinking in parallel of the particular places where its core activities are assembled. **Figure 1** shows some more or less well known leading regional agglomerations, or clusters, in Europe.

Throughout the twentieth century, a literature proliferated which, taken together, contributes to the formulation of a theory of why industry agglomeration emerges, and in what ways location in proximity of similar or related firms contributes to the competitiveness of an individual firm. Among the 'classics' in this field of research can be mentioned Alfred Marshall, Alfred Weber, Edgar Hoover, Gunnar Myrdal, and Allen Pred.

After a period of relative neglect during the 1970s and 1980s, the period since the early 1990s has seen reemerging research efforts being devoted to analyzing and explaining localization economies. Major contributions to this revival have been signed by geographers like Allen Scott, alongside business strategists like Michael Porter and economists like Paul Krugman. Despite quite different points of departure and methodologies, they have

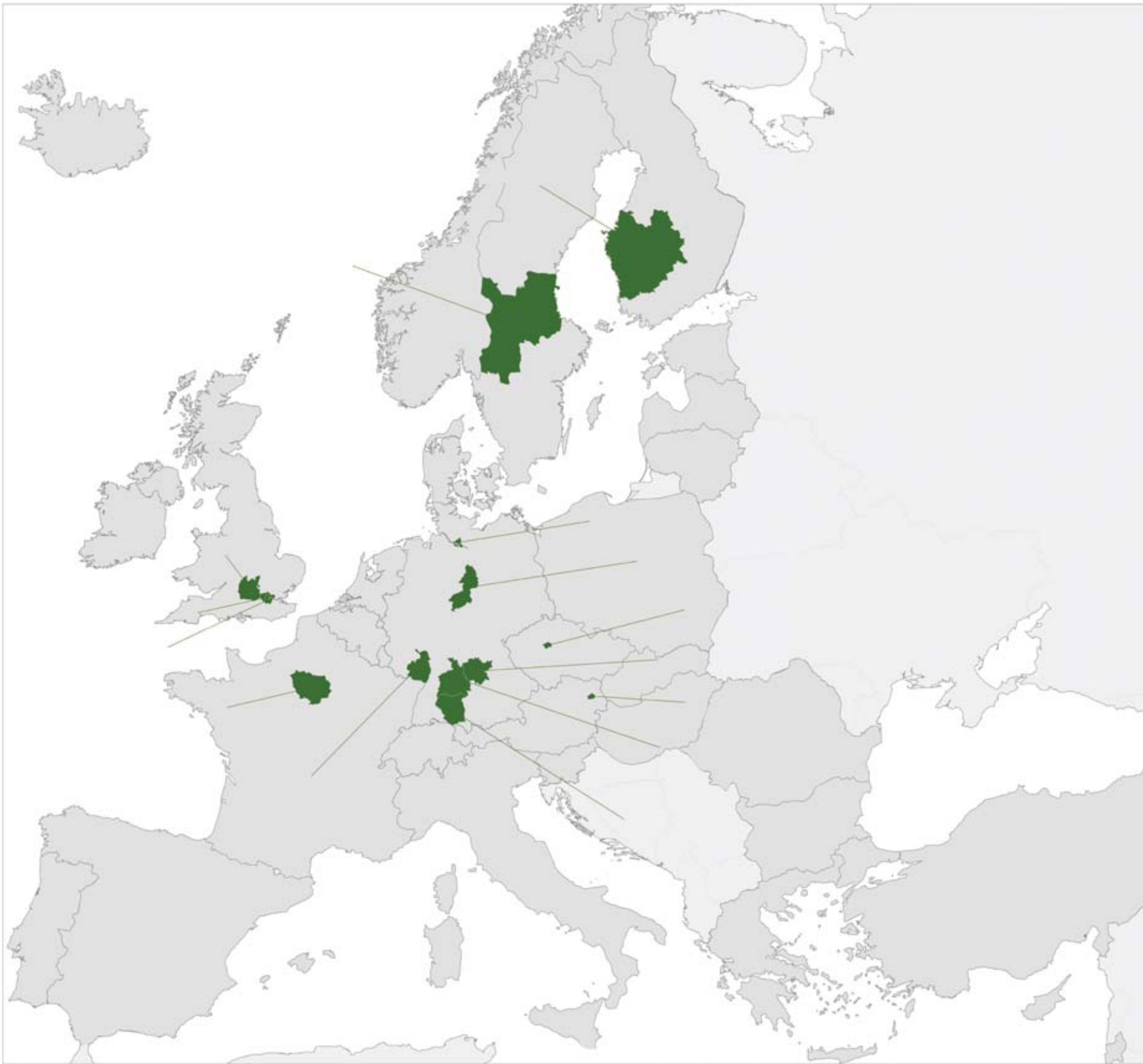

Figure 1 Some major industry agglomeratins in Europe. Source: European Cluster observatory (www.cluster.observatory.eu).

all come to regard spatial agglomeration of similar and related economic activity as a key starting point for understanding the workings and transformations of the world economy:

> The endemic tendency in capitalism for dense localized clusters of productive activity to appear at different lo cations on the landscape has major implications for growth and productivity. These clusters are constituted as transactions intensive regional economies which are in turn caught up in structures of interdependency stretch ing across the entire globe. (Scott, 1998: 398)

> Today's economic map of the world is dominated by (…) critical masses in one place of unusual competitive success in particular fields. Clusters are a striking feature of virtually every national, regional, state, and even metropolitan economy, especially in more economically advanced nations. (Porter, 1998: 78)

> Agglomeration the clustering of economic activity, created and sustained by some sort of circular logic occurs at many different levels, from the local shopping districts that serve residential areas within cities to spe cialized economic regions like Silicon Valley (or the City of London) that serve the world market as a whole. (Fujita *et al.*, 1999: 1)

While being pursued by scholars who represent dif ferent academic disciplines and traditions, and analyze the role of geographical space in economic process as part of rather disparate wider agendas, the point of de parture here is that there exists a common research agenda across various disciplines: that of understanding the role of space and place in industrial development in general, and the phenomenon of spatial agglomeration of similar and related economic activity in particular.

A note on terminology should be entered here. The agglomeration phenomenon is dealt with in the literature under different headings. Alfred Marshall's classical – and still in many ways unsurpassed – contribution to the topic uses the concept of localization to analyze the coming into existence of 'industrial districts'. In recent years, clusters and clustering have been the most widely circulated terms, following the work by Michael Porter. It has been argued by some that the agglomeration concept would refer exclusively to the spatial concentration of firms and industries, while the cluster concept would refer to spatial collocation in combination with functional linkage between the parties involved. For the purpose of this article, however, the terms agglomeration, local ization, and clustering are treated as synonyms.

False Paradoxes and 'Circular Logics'

The research problem attached to the agglomeration phenomenon may seem to be trivial. We have a phe nomenon that can be observed in reality. Firms in the same or related industries often – although far from al ways – tend to co locate in space, and this should rea sonably indicate that there should be some advantages connected to such a location pattern: agglomerations do exist and this may legitimately make us assume that such a spatial structure is in some sense efficient or rational. Still, the agglomeration phenomenon is often misconceived and the concept of agglomeration economies remains elusive.

A common misconception of the problem has to do with the relation between globalization and the spatial location of economic activity. All too often in con temporary economic geography, reference is being made to an alleged paradox: in an increasingly globalized world economy, localization seems to matter more, not less. The key to this alleged paradox can be found already in Adam Smith's famous dictum in his *Inquiry into the Wealth of Nations* in 1776, however, stating that the division of labor is determined by the extent of the market. The extension of the market explicitly refers to the spatial extension of the producers' market area. Smith's statement has a strong bearing on recent work on the logic of industrial agglomeration. If globalization means the emergence of truly integrated global markets, it follows that it is in creasingly possible to access these global markets from, in principle, any point of location. It should be easy to see the increased global economic integration opens up for more local/regional/national specialization. Thus, the structural effect of such globalization is that one or a few leading global production centers should in principle be able to supply the entire world market in each industry, where previously there were many of regional, national, or local production centers. In other words, increasing agglomeration of similar and related activity should be seen as a structural effect of globalization rather than some strange paradox.

The circular logic referred to in the quotation by Fujita *et al.* incurs more serious problems, however. In the terminology of economists, the agglomeration phenom enon is intimately related to increasing returns. (Eco nomic) geographers often prefer the concept of external economies of scale. Either way, we have a situation where initial events trigger subsequent action in a process of cumulative causation – or a circular logic, as Fujita *et al.* phrased it in the quote above. The existence of one or several firms in a certain place will lead to the attraction of more firms of the same kind, and the increased volume of activity will in turn exert an even greater attraction on further firms. Such processes tend to be self reinforcing, at least until some countering diseconomies kick in. There are several potential problems with this line of reasoning. One is that, unless it can be specified in some detail what the mechanisms of agglomeration economies are, there is the risk of ending up in circular reasoning. Paul Krugman cites the remark of a sarcastic physicist who had just heard a presentation on increasing returns: "So you're telling us that agglomerations form because of agglomeration economies." Models of agglomeration have always run such a risk: the existence of agglomer ations is explained by agglomeration economies. These agglomeration economies cannot really be specified or observed but their existence is proved by the existence of agglomerations! A related problem is that a proper ex planation of the agglomeration phenomenon should also have to specify the limits of agglomeration economies. If we specify a number of reasons why it is advantageous for firms to locate close to other similar or related firms or in big cities – which indeed many of them do – we also have to be able to say something about why not all firms locate in such settings, or why in some case dense agglomer ations seem to petrify and lose their dynamism.

The next section summarizes the factors held to make up agglomeration (or localization) economies. The final section then discusses in schematic terms how we can understand the development of agglomerations over time, by adopting an evolutionary approach.

What Factors Make Up Agglomeration Economies?

Identifying and analyzing those advantages that may accrue to firms located in close juxtaposition to other similar and related firms, rather than being located in isolation, is one way to account for the agglomeration phenomenon. The concept of agglomeration economies (or localization economies) refers to these advantages. The basic understanding of this concept was in place already a century ago. Thus, Alfred Marshall identified already in 1890 three different mechanisms – related to interfirm linkage, labor skills, and knowledge spillover that are still more or less valid. A few decades later location theorists like Alfred Weber added a fourth, pointing to cost savings derived from shared infrastructure.

Beginning with the latter, there are benefits to be gained from the possibility for agglomerated firms to share the cost for certain collective resources between several firms. This applies in particular to the cost of installing proper infrastructures (roads, sewage, electricity, and broadband networks) and institutions (education and legal systems). It need not be just a matter of cost savings, however. When an agglomeration of similar or related firms is established, there is also a potential to adjust the local infrastructure, the education system, and other types of collective goods after the needs of this particular industry.

Second, firms in agglomerations can reduce their transport and transaction costs as interfirm transactions and shipments are simplified when the distance between firms is negligible. The customer firm that can place an order with a supplier located down the street will gain an advantage in relation to a competitor that has to travel long distances to see its supplier or, alternatively, place the order by means of some communications system and then wait for the shipment to take place. Reduced transaction costs for co located trading partners may be the shorthand for this second mechanism.

Third, agglomeration makes for the development of a local labor market for specialized skills. The establishment of a local pool of skilled labor has been proposed as a major element of the localization economies ever since Alfred Marshall more than a century ago wrote so elegantly about the advantages of being located in an industrial district:

> Again, in all but the earliest stages of economic development a localized industry gains a great advantage from the fact that it offers a constant market for skill. Employers are apt to resort to any place where they are likely to find a good choice of workers with the special skill which they require; while men seeking employment naturally go places where there are many employers who need such skill as theirs and where therefore it is likely to find a good market. (Marshall, 1890/1916: 271)

Thus, it can theoretically be argued that local labor markets function better, from the point of view of both firms and employees, if there are several similar and related firms around. Well functioning markets for specialized labor skills can thus be added to the list of localization economies.

Fourth, an agglomeration of related firms forms the basis of a local milieu that may facilitate knowledge spillovers and stimulate various forms of learning and adaptation. This is the aspect of industry agglomeration that has attracted the bulk of research interest since the early 1990s. The general argument is that agglomerations tend to trigger processes that create not only dynamism and flexibility in general, but also learning and innovation. In such milieus, chances are greater that an individual firm gets in touch with actors that have developed or been early adopters of new technology. The flow of industry related information and knowledge is generally more abundant to the advantage of all firms involved. A local culture with specific norms, values, and institutions (formal and informal) makes it possible to transfer also more tacit forms of knowledge from one actor to another. Thus, agglomeration may also promote localized learning and knowledge spillovers between firms.

To sum up: spatial proximity between similar and related firms does make easier not just interfirm transactions in general, but in particular those interactions and knowledge spillovers which form the basis for innovation and learning. The generation of well functioning labor markets for specialized skills does also contribute both to efficiency and knowledge creation. Institutional and infrastructural factors add further to making location in an agglomeration beneficial for the individual firm.

How Do Agglomerations Emerge and Develop – Historical Approaches

Spatial agglomerations of similar and related economic activity often have deep historical roots. An alternative way to understand the agglomeration phenomenon is to study real agglomerations historically in order to analyze their birth, growth, saturation, decline, and possible reinvigoration.

The origin of the agglomeration in such accounts, that is, the event or action which triggered subsequent developments, sometimes turns out to be related to some more or less traditional factor of location, that is, some natural or social asset that at a certain point in time turned out to be an important location factor for a particular type of economic activity. Almost equally often, however, there seems to be a considerable amount of chance involved. Some entrepreneurial person did, for some reason, get the idea to start a certain type of economic activity. Analyses of Silicon Valley do almost without exception start out telling one particular story. This is the story of how Frederick Terman, who had moved from Massachusetts Institute of Technology (MIT) in Boston to a chair in electro technology at

Stanford University, encouraged his two students, William Hewlett and David Packard, to try to find a way to commercialize an instrument which Hewlett had designed as part of his work with his master thesis. Thus, in a garage in Palo Alto in the late 1930s, a business was set up that meant the start of a process which within a couple of decades had developed into the world's without doubt most important agglomeration in the computers, electronics, and software industry. The fact that a young woman named Catherine Evans decided to make a wedding gift in 1895 was one of the reasons why Dalton, Georgia, half a century later had become the primary US agglomeration of firms producing carpets. The literature is rather silent of the mechanisms causing these rather random historical moments which sometimes trigger the development of agglomerations, except for the fact that most new firms start out at the place of residence of their founder.

Once an activity has started in a particular place, other mechanisms come into play, contributing to subsequent developments. Thus, while agglomerations often originate in a rather random series of events leading to the start of a new firm at the place of residence of the entrepreneur; they are sustained by various forms of inertia and myopia, meaning that firms rarely relocate once they have been rooted in a place, but more importantly that a successful economic activity in one place is often followed by other similar or related activities. This could be the result from spinoffs, that is, former employees see a chance to start their own business, either as a competitor or a supplier to the original firm, but also through simple imitation. The fact that one firm succeeds is seen as a proof that this is a viable business in the particular location and it will inspire others in the neighborhood to try their luck in the same business.

In addition to pioneering, imitation, and spinoffs, various processes will tend to follow: the creation of networks between local firms; the creation of a local culture, supporting infrastructures and institutions adopted to the proliferation industry; the establishment of the place as a brand of the industry and subsequent attraction on resources from the outside (people, capital, firms). Ultimately, agglomerations of many similar and related firms will also tend to go through phases of consolidation – at some stage some firms tend to take on leading roles and often this means that they take over other firms in the agglomeration such that an initially small firm based agglomeration often ends up being dominated by a limited number of larger firms. Also, history tells us that most regional agglomerations, sooner or later, run into phases of crisis. Elements of petrifaction are often revealed at points in history when technological or other developments call for rapid restructuring. This may mean that what was once a leading center of dynamism within a given line of business ends up as an 'old industrial region', facing problems of renewal and finding itself being out competed by firms located elsewhere. Then again, there are numerous examples of how industry agglomerations manage to reinvent themselves also in this phase, such that they can actually retrieve some of their former greatness.

The above is not intended as a complete stage model of the rise and fall of industry agglomerations. It merely serves to establish as a fact that some of our knowledge on the agglomeration phenomenon comes from studying the historical development of agglomerations, and that it is possible through such accounts to gain insight into the processes which make for the evolutionary development of industry agglomerations. In addition, it is useful to keep in mind that the impact of agglomeration economies may vary considerably over time, throughout the life cycles of industry agglomeration.

Remaining Research Challenges

The agglomeration phenomenon has been a core research issue in economic geography for many years. Still there are unresolved questions that remain challenges for further research in the field. One issue stands out as being particularly important in this context. Will the continuous development of ICT eventually reduce the need for face to face interaction not just for transmitting standard information but also when it comes to complex problem solving and explorative activities? And will this in turn reduce the need for, and benefits following from, spatial agglomeration of similar and related activity?

See also: Competitiveness; Creativity; Entrepreneurship; Industrial Location; Innovation; Labor Market; Location Theory; Uneven Development.

Further Reading

Fujita, M., Krugman, P. and Venables, A. J. (1999). *The Spatial Economy. Cities, Regions and International Trade*. Cambridge, MA: The MIT Press.

Hoover, E. M. (1948). *The Location of Economic Activity.* New York: McGraw Hill.

Krugman, P. (1991a). *Geography and Trade*. Cambridge, MA: MIT Press.

Krugman, P. (1991b). Increasing returns and economic geography. *Journal of Political Economy* 99, 483 499.

Krugman, P. (1995). *Development, Geography and Economic Theory.* Cambridge, MA: The MIT Press.

Malmberg, A. and Maskell, P. (2002). The elusive concept of localization economies: Towards a knowledge based theory of spatial clustering. *Environment and Planning A* 34, 429 449.

Malmberg, A. and Power, D. J. (2006). True clusters/A severe case of conceptual headache. In Asheim, B., Cooke, P. & Martin, R. (eds.) *Clusters and Regional Development: Critical Reflections and Explorations*, pp 50 68. London: Routledge.

Marshall, A. (1916) *Principles of Economics*. *An Introductory Volume* (first published in 1890), (7th edn.). London: Macmillan.
Martin, R. and Sunley, P. (2003). Deconstructing clusters: Chaotic concept or policy panacea? *Journal of Economic Geography* 3(1), 5 35.
Myrdal, G. (1957). *Economic Theory and the Underdeveloped Regions*. London: Ducksworth.
Pred, A. (1966). *The Spatial Dynamics of US Urban Industrial Growth, 1800 1914. Interpretive and Theoretical Essays*. Cambridge, MA: The Harvard University Press.
Porter, M. E. (1990). *The Competitive Advantage of Nations*. London: Macmillan.
Porter, M. E. (1998). *Clusters and the New Economics of Competition*. Boston, MA: Harvard Business Review.
Scott, A. J. (1988). *New Industrial Spaces: Flexible Production Organisation and Regional Development in North America and Western Europe*. London: Pion.
Scott, A. J. (2006). *Geography and Economy. Three Lectures*. Oxford: Claredon Press.
Weber, A. (1929). *Theory of the Location of Industries* (first published in 1909). Chicago, IL: The University of Chicago Press.

Agoraphobia

J. Davidson, Queen's University, Kingston, ON, Canada

Glossary

Existentialism A philosophy of existence or being concerned with choice, freedom, and authenticity that holds individuals to be without predetermined essence. Rather, persons are responsible for their own essential being through their individual will and actions.

Panic Attack Sudden and overwhelming experience of intense and disabling anxiety experienced by phobia sufferers when confronted with their phobic object.

Phenomenology A philosophy that investigates the experience and subjective significance of phenomena – essential objects of perception as presented or disclosed to the subject. To facilitate purely intuitive analysis of phenomena, phenomenology holds that all naturalistic presuppositions and prior knowledge should be set aside or bracketed off; this suspension of judgment is termed the *epoché*.

Phobia Fear and avoidance of a particular object or circumstance, generally considered to be abnormal and irrational. Phobias range widely in severity from the mild and merely inconvenient to the severely debilitating forms that lead sufferers to seek professional help.

Subjectivity Often used in preference to identity for its less fixed and individualistic, more fluid, and inclusive connotations, subjectivity is concerned with the qualities of being a thinking–feeling self in context. The term is associated with approaches such as feminism that attend to individuals' complexly constituted subjective experiences of the world.

Introduction

Since the late 1970s, there has been a surge of interest in agoraphobia and panic among clinicians, and publications on the subject often begin by drawing attention to its contemporary prevalence. Agoraphobia is described by clinical researchers as one of the most common – as well as distressing – phobic disorders encountered. The significance of phobic anxieties for contemporary life has not gone unnoticed beyond clinical circles, and different disciplinary perspectives have helped shed light on the experience and meanings of agoraphobia. As this article shows, feminist geographers have made an important contribution to such studies, particularly through their insistence on rethinking boundaries and relations between gendered bodies and environments.

Definitions and Background

Contrary to popular opinion, the condition of agoraphobia actually relates to public and social, as opposed to open spaces. Although such widely held misconceptions are supported by dictionary definitions of the disorder, notes on the term's origins explain its derivation from the Greek *agora* or marketplace, providing substantial clues as to the realities of the condition for its many sufferers. It is precisely such populous gathering places that agoraphobics avoid to this day.

Clinicians' descriptions and diagnostic criteria have continued to evolve since the first usage of the term agoraphobia in 1871 by German psychiatrist Karl Otto Westphal. Current definitions tend to be drawn from the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders*. Here, the clinical features emphasized involve persistent anxiety, repeated experience of panic attacks (discussed below), and avoidance of places associated with the onset of these attacks.

Prevalence

Exactly how many sufferers there are is a matter of some debate, and estimates range from between 6 per 1000 and 6 per 100. It is widely agreed, however, that the vast majority of agoraphobia sufferers – almost 90% – are women. For all those who develop agoraphobia, the onset tends to occur between the ages of 18 and 35, and the condition tends to be chronic. Many sufferers attest to the fact that one can never be entirely cured from agoraphobia, but must rather learn to cope with it on a continuing, daily basis.

Causes and Treatment

Clinical research on causes of agoraphobia has not been conclusive, but there is evidence that higher than usual levels of stress and events that undermine the person's sense of safety tend to be experienced prior to its onset. In terms of treatment, reviews of recent literature suggest that a multifaceted approach capable of recognizing and tackling each problematic aspect of the disorder is most likely to be successful. For example, programs of cognitive therapy to unsettle harmful patterns of fearful thoughts can be combined with behavioral therapy, involving gradual exposure to avoided situations, and drug treatments that target physiological symptoms of anxiety. Approaches documented in clinical literature thus acknowledge much of the disorder's complexity, but representations tend to be simplistic in certain important respects, and the model of the subject employed tends to be somewhat philosophically naive. Also, while recognizing that agoraphobia is socially and spatially mediated, understandings of what this might mean for sufferers and how social and spatial theory might be of benefit are, as one might expect, limited.

A small number of nonclinical studies have argued that some of these limitations might be at least partially rectified by the careful inclusion of sufferers' own voices (conspicuously absent from clinical accounts) and resolved further with reference to a sympathetic theoretical perspective. Arguably, geographers – informed by feminist, philosophical and sociological theory – are particularly well placed to refine and improve existing definitions of agoraphobia. Employing a more spatialized and experientially influenced perspective, geographers can better account for and understand the mutually constitutive nature of person/place relations, and the implications of these relations for emotional health.

Embodying and Theorizing Agoraphobia

Placing Panic

Drawing on in depth interviews with agoraphobics, geographical studies have stressed that the phenomenological aspects of agoraphobia of overwhelming significance for its sufferers relate primarily to the experience of panic. A panic attack involves a wide range of terrifying mind–body symptoms, including: chest pain, palpitations, and shortness of breath; feelings of going crazy and losing control of one's thoughts; sensations of nausea, pins and needles, and intense shaking; a sense of fading away and merging with surroundings (clinically referred to as derealization and depersonalization). All of these symptoms can be experienced simultaneously and combine to make the sufferer fear that she will die or, at the very least, suffer irreparable damage and complete humiliation in public.

Existentialism and Being Bound

Geographical accounts have shown that these symptoms cause sufferers to question their ordinarily taken for granted sense of themselves and their bodily boundaries. Derealization and depersonalization, for example, are commonly understood to involve strange and disturbing feelings relating to both inner and outer reality. This aspect of panic entails a horrendous sense of dissolution of self into one's environs, and a simultaneous feeling of invasion by one's surroundings. It severely disrupts our unconscious sense of ourselves throwing the relation we have with our bodies into profound question. Some aspects of agoraphobic panic can be described as existentialist in nature, causing sufferers to question the very grounds of their existence. Existential philosophies – for example of Sartre and Kierkegaard – have thus been employed to shed light on the nature of agoraphobic disruption to the sense of being in the world, helping us understand how it feels and what it means to be agoraphobic.

Following the onset of anxiety as a way of life, studies suggest that there is a loss of trust in the integrity of the body. Sufferers fear being betrayed by it, and are constantly on the lookout for breaches of bodily boundaries that might be visible to others. These include, for example, minor discrepancies such as blushing, trembling, or sweating, in addition to more extreme examples, such as screaming out loud, or the loss of bowel or bladder control; manifestations of their fear on the surface of their bodies which, sufferers explain, would be mortifying. In agoraphobia, our sense of embodied subjectivity is inescapably conspicuous (as indeed it is with many other manifestations of chronic illness and disability). The intensely embodied nature of these emotional states necessitates a reconfiguration of the subject's conception of herself. The disorder seems to initiate a sense of separateness from one's body, not just momentarily, as in the fleeting depersonalization of the panic attack, but in the more general sense of creating a constant anxiety which hovers over the question of our bodily and mental identity.

Phenomenology, Subjectivity, and Space

From a phenomenological position (such as Merleau Ponty's) it is unsurprising that anxiety should manifest itself both bodily and mentally since phenomenology assumes a model of subjectivity which regards the body as the self expressed. Such phenomenological insights have been highly influential for geographies of agoraphobia, as the approach enables us to return to, and concentrate on, the accounts of women's actual lived and embodied experiences, to take seriously their own voices and the stories they tell, rather than requiring us to posit specific external causes. For feminist geographers, a nuanced understanding of the nature and constitution of social space is required if we are to understand the complex interactions between self and space that sufferers' narratives highlight again and again.

For agoraphobics and nonphobics alike, space is never simply abstract and detached from the self. It influences and is influenced by one's own and others' emotions and actions. The sufferer's apparent gut reaction to social space should be seen as an interaction, as an embodied, emotional exchange, arising from a complexity of meanings, meetings, and movements between particular selves and spaces. Geographical approaches to agoraphobic phenomena clearly move beyond the search for either internal (self) or external (space) explanations, and are open to relational reconfiguration of both. By con ceptualizing boundaries as relationally configured, per meable rather than fixed, geographers have been able to conduct sympathetic and meaningful explorations of agoraphobic landscapes, the spaces and places that par tially produce (and are partially produced by) feelings of panic in susceptible individuals. Geographical accounts often highlight sufferers' felt need to reinforce their weakened boundaries, to render self space relations more predictable and so take some control over their dis ordered experience.

Managing Social Space

Common Coping Mechanisms

Given the crisis state of spatial affairs entailed by agoraphobia, it is unsurprising that the subjects should attempt to find ways to protect themselves. Erving Goffman's sociology has been inspirational for feminist studies of agoraphobia because of his attempts to eluci date the nature of less pathological social anxieties (such as everyday embarrassments) and the tactics by which they are managed. Goffman demonstrates that non phobics feel a similar (though less pressing) need to ne gotiate boundaries between and distance from others, often using what he terms involvement shields, such as dark glasses or a raised newspaper to block eye contact or interaction in social spaces of, for example, public transport. Such tactics are invaluable for agoraphobics, who employ them far more frequently and less casually to manage anxiety and reinforce their weakened boundaries. Other phobic tactics include holding an umbrella or pushing a bicycle to create a sense of private and protective space around the fragile self that keeps others at a safe distance.

Being Housebound

Such tactics are not always, however, experienced as sufficiently protective, and some agoraphobics withdraw from social space entirely, becoming more or less tem porarily housebound. Agoraphobic vulnerability can mean the sufferer feels she has no choice but to remain exclusively within the home, assuming its protective boundaries as reinforcement and extension of the psy chocorporeal boundaries of the self. The agoraphobic thus incorporates her own four walls as an essential element of her ontological or existential security.

Feminist geographers have explored such issues in relation to distinctions between private and public spaces, arguing that, perhaps especially for the agoraphobic, the boundaries between these spheres are very much fluid and unfixed. This is evidenced by the fact that the home space traditionally conceived as private can itself become threatening and dangerous when, for example, relatives converge at Christmas. Likewise, the public space of the town center can be experienced as comfortable and safe when no one else is around. (Ironically, given the com mon and legitimate concerns of nonphobic women for their safety, this is often when darkness provides calm and cover from the socializing presence of others.)

The agoraphobic's experience of home can be am biguous in other ways; for some, home becomes so secure that they are rendered incapable of leaving, and it can thus simultaneously be experienced as both prison and asylum. But, despite the ambivalence of feelings, it is clearly the space of the home that is overwhelmingly perceived to be safe, the (anti)social space beyond its confines that menaces and disturbs.

Gendered Boundaries

In considering what it is about the gender(ing), situ ation(s) and treatment of women that entails their heightened susceptibility to agoraphobia, geographers have considered the potential significance of gender role stereotypes. For women, such stereotypes involve char acteristics not dissimilar to those found in agoraphobia; a woman's place is commonly described as in the home. Such notions about proper behavior and roles are cul turally pervasive to the extent of being clichés; they portray and arguably partially construct women as more dependent, emotional, passive, fearful, home based, etc. It thus seems relevant to our understanding of agora phobia that, despite the achievements of feminism in expanding women's horizons, the gendered identities of women are still more likely to be enmeshed with the fabric of the home than those of men. In many ways such stereotypical expectations make it easier, more normal, and acceptable, for women to fall into and develop a pattern of agoraphobic avoidance. Staying home is less deviant or unusual for women than it is for men, who may be more likely to receive direct challenges to and indeed opportunities to curtail signs of agoraphobic be havior. Women, however, are still culturally encouraged to experience home as a normatively feminine place of retreat and refuge, where gendered boundaries are pro tected and potentially reinforced.

Of those few social theorists who have offered gender sensitive insights into agoraphobia, some have attempted to contextualize its emergence, occurrence, and experi ence in terms of wider cultural, economic, and political circumstances. Issues of sex inequality and its effect on social institutions are, on such accounts, particularly relevant. Consequently, in accordance with feminist aims and objectives, putting the beginnings of a cure into place would involve the creation of real options for both sexes, in the home as well as the market place, and the wide scale reduction of sex inequality in institutional structures.

Conclusion

Feminist geographies of agoraphobia have stressed the importance of a relational and gender sensitive under standing of space and subjectivity. By conducting in depth research on the phenomenology and embodiment of agoraphobia, and the kinds of places perceived to threaten or protect the self, studies highlight the com plexity of interactions between gendered bodies and environments.

See also: Body, The; Embodied Knowing; Emotional Geographies; Emotional Knowing; Existentialism/Existential Geography; Feminism/Feminist Geography; Gender in the City; Home; Phenomenology/Phenomenological Geography; Private/Public Divide; Subjectivity.

Further Reading

Bankey, R. (2001). La donna e mobile: Constructing the irrational woman. *Gender, Place and Culture: A Journal of Feminist Geography* 8, 37 54.
Bankey, R. (2004). The agoraphobic condition: Review essay. *Cultural Geographies* 11, 347 355.
Brown, G. (1987). The empire of agoraphobia. *Representations* 20, 134 157.
Callard, F. (2006). Understanding agoraphobia: Women, men, and the historical geography of urban anxiety. In Berkin, C., Pinch, J. L. & Appel, C. S. (eds.) *Exploring Women's Studies: Looking Forward, Looking back*, pp 201 217. Upper Saddle River, NJ: Prentice Hall.
Callard, F. (2006). 'The sensation of infinite vastness'; or, the emergence of agoraphobia in the late 19th century. *Environment and Planning D: Society and Space* 24, 873 889.
Carter, P. (2002). *Repressed Spaces: The Poetics of Agoraphobia*. London: Reaktion Books.
da Costa Meyer, E. (1996). La donna e mobile: Agoraphobia, women, and urban space. In Agrest, D., Conway, P. & Weisman, L. K. (eds.) *The Sex of Architecture*, pp 141 156. New York: Harry N. Abrams.
Davidson, J. (2000). A phenomenology of fear: Merleau Ponty and agoraphobic life worlds. *Sociology of Health and Illness* 22, 640 660.
Davidson, J. (2003). 'Putting on a face': Sartre, Goffman, and agoraphobic anxiety in social space. *Environment and Planning D: Society and Space* 21, 107 122.
Davidson, J. (2003). *Phobic Geographies: The Phenomenology and Spatiality of Identity*. Aldershot: Ashgate Press.
de Swaan, A. (1981). The politics of agoraphobia: On changes in emotional and relational management. *Theory and Society* 10, 359 385.
Gournay, K. (1989). Agoraphobia: A woman's problem? The sex role perspective. In Gournay, K. (ed.) *Agoraphobia: Current Perspectives on Theory and Treatment*. London: Routledge.
Kirby, K. M. (1996). *Indifferent Boundaries: Spatial Concepts of Human Subjectivity*. London: The Guilford Press.
Reeves, J. B. (1986). Toward a sociology of agoraphobia. *Free Inquiry in Creative Sociology* 14, 153 158.
Strange, C. and Bashford, A. (2003). *Isolation: Places and Practices of Exclusion*. London: Routledge.
Vidler, A. (1991). Agoraphobia: Spatial estrangement in Simmel and Kracauer. *New German Critique* 54, 31 45.

Relevant Websites

http://www.anxietysupport.org
Agoraphobics Building Independent Lives (ABIL), Mental Health America Virginia.
http://www.adaa.org
Anxiety Disorders Association of America.
http://www.phobics society.org.uk
National Phobics Society (UK).
http://anxietybc.com/index.php
The Anxiety BC Website.

Agrarian Transformations

J. Renes, Utrecht University, Utrecht, The Netherlands, and Vrije Universiteit, Amsterdam, The Netherlands

Introduction

For thousands of years, landscape and society have been heavily influenced by developments in agriculture. In most preindustrial societies, agriculture occupied the majority of the population. Holland was an early exception, with the majority of the population already living in towns at the end of the sixteenth century. Most countries in Western Europe reached this point somewhere during the first half of the twentieth century, but for Europe as a whole even in the beginning of the twenty first century half the population is still considered 'rural'. Transformations – defined as qualitative changes or events that occur when something passes from one state or phase to another – in agriculture therefore were important developments that influenced large parts of the population, as well as the landscape.

Most agrarian historians, following pioneering work by, among others, the Danish historian, Esther Boserup, nowadays see population pressure as the main driving force for developments in the preindustrial countryside. This is elaborated by the geographer David Grigg in a model (**Figure 1**) that shows how population pressure can lead to different developments, not only changes in agriculture, but also a growth of nonagricultural activities and emigration, as well as feedback mechanisms such as birth control.

For this article, in particular the top row of the diagram is important. Reclamation of new fields is perhaps the most obvious result of a growing demand for food and is certainly the theme that has received most attention from historical geographers. In particular, the period of population growth from the tenth to the early fourteenth century led to reclamations in most parts of Europe. A later period of population growth, the so called 'long sixteenth century' (*c.* 1450 to *c.* 1650), was characterized more by innovations within the existing arable and pasture lands and is less clearly visible in the landscape. The diagram also shows effects such as regional specialization and growth of urban populations that led to a growing economic interdependence in times of population pressure.

When phases of population pressure led to reclamations and to a growing interdependence, periods of decline could bring the opposite: land desertion and a certain degree of economic de specialization and disintegration. Both phases

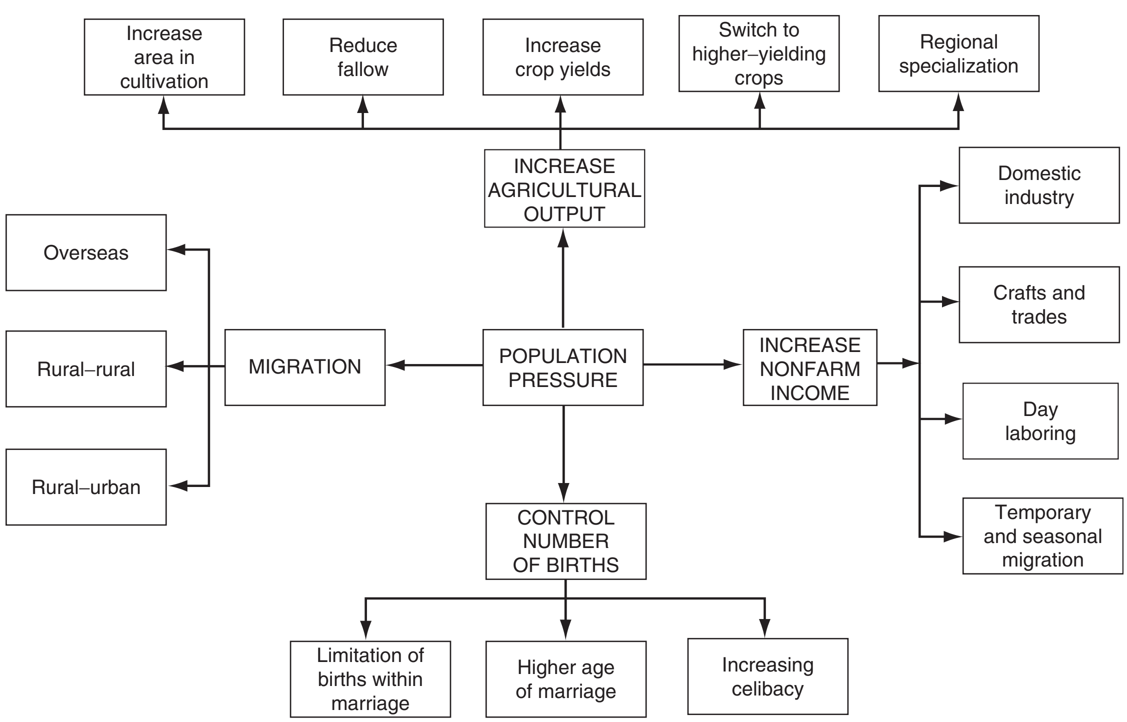

Figure 1 Production and demographic responses to population pressures. From: Grigg, D. B. (1980). *Population growth and agrarian change*; *an historical perspective*, p 46. Cambridge: Cambridge UP (Cambridge Geographical Studies 13).

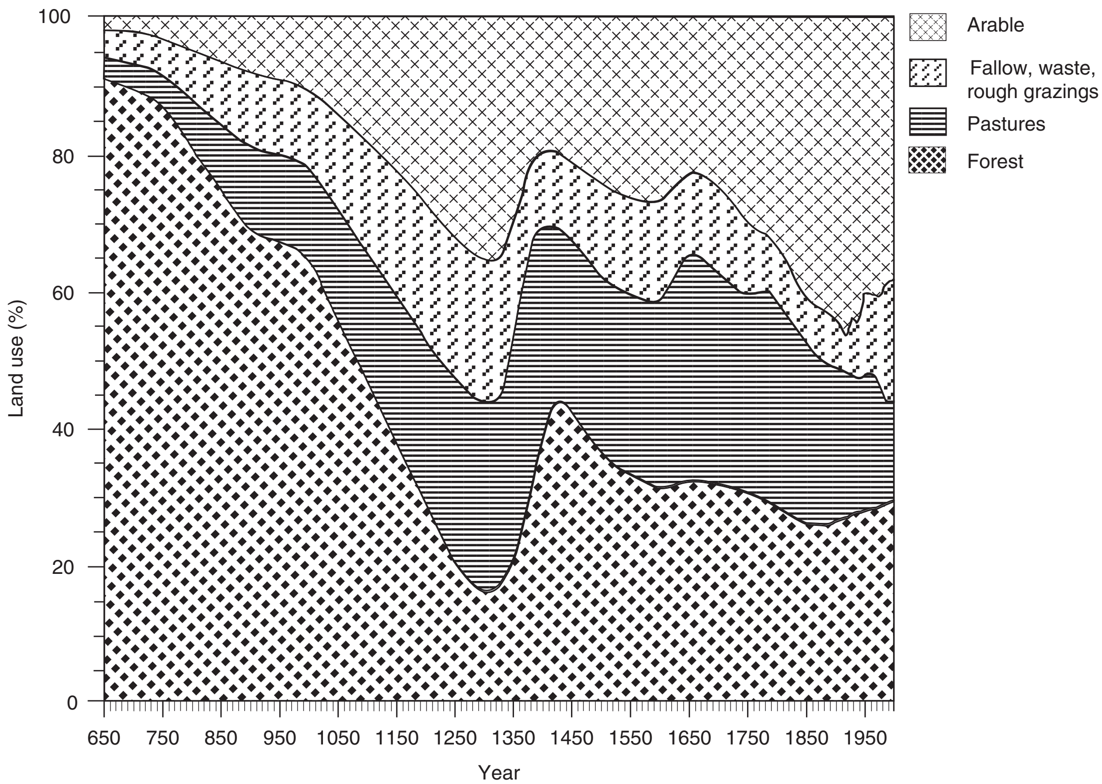

Figure 2 Land use in Germany (seventh to twentieth century). From: Bork, H.-R., Bork, H., Dalchow, C., Faust, B., Piorr, H.-P., and Schatz, T. (1998). *Landschaftsentwicklung in Mitteleuropa*, p 221. Gotha/Stuttgart: Klett-Perthes.

were characterized by different types of development, whereby times of crisis often showed the more funda mental, qualitative changes (transformations), making the conditions to profit maximally from a later period of renewed growth. In this article, we will give a short over view of the main stages in the development of the agrarian economy and landscape. The general picture of the de velopment of agrarian land use in Central Europe is given in **Figure 2**. Particularly the rise and fall of the size of the arable closely followed the development of popu lation size.

Early and High Medieval Europe

In many parts of Europe, within as well as outside the struggling Roman Empire, the third to fifth centuries AD were a period of crisis and de population. This was fol lowed by a slow recovery in the following centuries. During the seventh to ninth centuries, the peripheries of Europe, Scandinavia, Ireland, Muslim Spain, and the Byzantine Empire, showed the more dynamic develop ment. From the ninth century, the North Sea region and Central Europe recovered and went through a period of fast population growth, which led to large scale reclam ations of forests, heathlands, and wetlands. Most of the extensive forests and fen lands that still existed in the beginning of the period were reclaimed a few centuries later. Many of these reclamations were almost standard ized, comprising linear settlements in which each farm consisted of a single strip of land. A standard size of 14–16 ha was found in the Dutch fen lands and in other regions, giving an impression of what was seen as a standard family farm unit. As part of the expansion, people from northwestern Europe migrated to Ireland, Spain, and Eastern Europe, especially in the last men tioned region, parts of the landscape still show the traces of planned settlements from this period. During the twelfth and thirteenth centuries, a network of towns developed and the European economic system increas ingly centered around two emerging core regions, Northern Italy and Flanders, which were both charac terized by a large number of towns and a very intensive agriculture.

The most characteristic landscape of this period was that of the open fields, that developed from the ninth century onward. One of the results of a rising demand for food for the fast growing population was the growing emphasis on grain products, such as bread and porridge. Grain became the main staple crop and the most efficient system for producing grain consisted of open fields, large arable fields without hedges, on which the farmers from a village cooperated in a number of ways: ploughing, harvesting, management of the common lands, and using the arable as common pasture after the harvest. After a few centuries, these ‘open fields’ formed an east–west band through Europe, from central England via the Paris Basin through Central Europe, with outlying regions in southern Scandinavia and the Mediterranean – a region including the best soils of the subcontinent (**Figure 3**). High grain prices made subdivision of farms possible and during the early fourteenth century the fields were managed by an agrarian population that had grown immensely during the preceding centuries. Fragmentation of ownership made ever more cooperation between farmers necessary and many open fields were managed as a whole, making the villages almost being run as one large farm. When every single farmer owned a number (sometimes dozens) of parcels dispersed in the fields, the best location for the farm buildings was the village, at the center of the network of country roads. In most of the open field landscapes, villages grew into large, densely built settlements.

Quite different were the much more individualized ‘enclosed’ landscapes (**Figure 3**) that were characterized by mixed farming with often an emphasis on pasture and a large degree of flexibility. An example was the sand belt, the zone of poor sandy soils that runs from Flanders through the Netherlands and Germany into Denmark. This zone was characterized by poor arable soils (in individual enclosures, as well as in small open fields) and

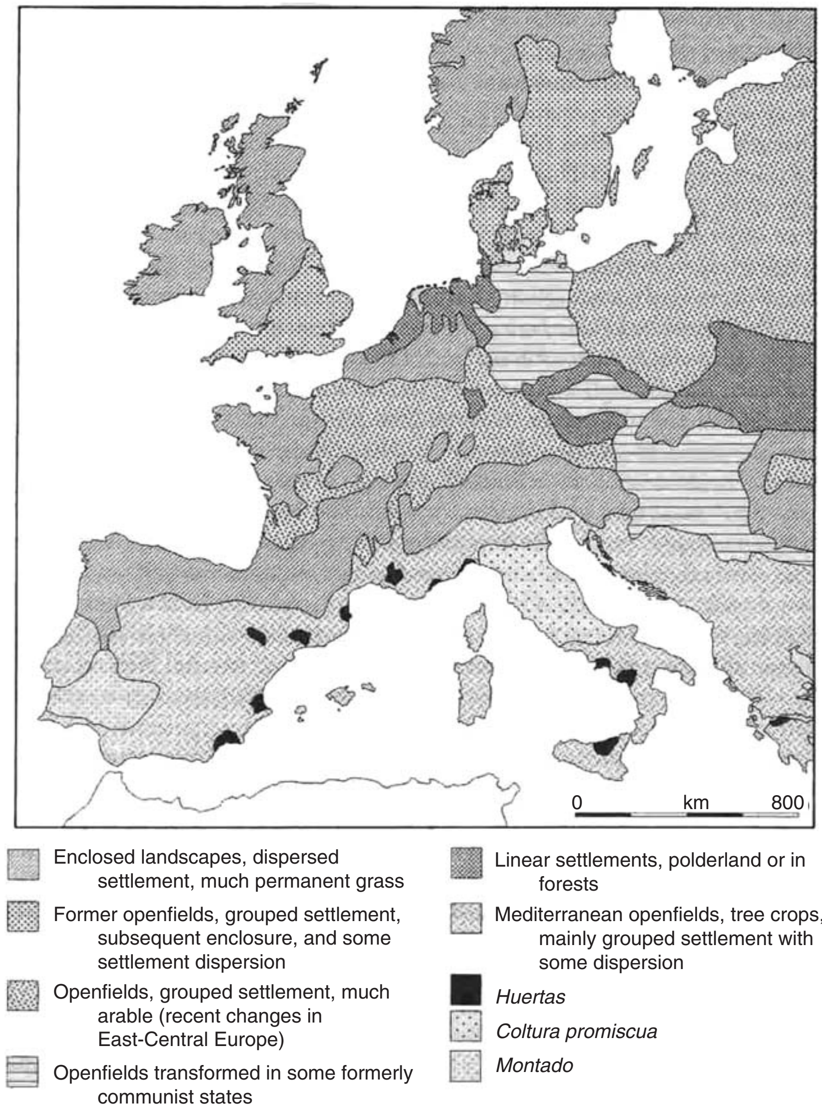

Figure 3 European cultural landscapes. From: Clout, H. D. (1998). The European countryside: Contested space. In Graham, B. (ed.). *Modern Europe*; *place, culture and identity*, p 289. London/Sydney/Auckland: © Edward Arnold Publishers Ltd. Reprinted by permission of Edward Arnold (Publishers) Ltd.

more than half of the village territories consisted of common forests, heathlands, and moors. One result of large scale reclamations was the shrinking of these common lands, that were used as pastures and supplied many products such as wood, game, wool, sand, building materials, etc. and were therefore important parts of the village territories. In an agrarian economy in which risk management rather than profit maximization was the main aim for farmers, the commons could even mean survival in years of bad harvests. In many regions the population pressure of the thirteenth century led to stricter rules for the use of the commons. In parts of the sandy regions, complex systems for management of the commons developed, in which the old established farms obtained shares according to the size of their arable. Newly founded farms had less rights on the commons and their owners belonged therefore to the growing class of cottagers.

The Late Middle Ages

During the first half of the fourteenth century, the population had reached a maximum. A population that had become increasingly dependent on grain, was hit hard by a succession of bad harvests and, in the years after 1346, by the large plague epidemic. Within a few decades, the European population dropped by almost a third. There were exceptions. In the Low Countries, the population size diminished relatively little and recovered fast. By large inputs of labor and manure, the small farmers in Flanders succeeded in getting almost incredible harvests from their tiny plots of land.

The population crisis led to a reconstruction of European agriculture. Especially the grain growing regions had a bad time. Some regions turned toward animal husbandry. An example is the Dutch fen lands, where subsidence of the soil made arable increasingly difficult and where, on the other hand, the growing towns created a demand for dairy products and meat. Elsewhere small regions succeeded in specializing on wine or other special crops. Many of the traditional grain growing regions, however, were not able to develop alternatives and here, much arable and thousands of settlements were abandoned. In parts of the Central European hills, agriculture never recovered.

In central England, a long process of change from an open field grain producing region to a sheep farming region started during this period, made possible by the growing demand for wool from the (protected) industry. Most English deserted villages differ from the Continental ones, because the English examples were not completely abandoned but contracted into a single (sheep) farm. The change from arable to pasture led to the first enclosures, in which some of the open fields with their fragmented landownership were transformed into a landscape of enclosed farms. Often the farm buildings were moved to their newly enclosed fields.

The Early Modern Period

A slow recovery started during the fifteenth century and continued during the 'long sixteenth century'. A new phase of economic integration started during this period, in which the European economies and landscapes regrouped around northwestern Europe, the only remaining core region. The main center of trade moved within the core region from Antwerp (sixteenth century) to Holland (particularly Amsterdam, seventeenth century) and England (eighteenth century). Around this core, Von Thünen like zones can be distinguished, with the most intensive agriculture, including dairy farming and horticulture in the core region itself, with large scale grain growing extending along the Baltic and with export oriented breeding of cattle and horses in regions such as Jutland and Hungary. Around the Baltic Sea, grain growing estates could compete with competitors closer to the urbanized core region by eroding the position of their tenants and so reducing their labor costs. Here, a countryside of small farmers developed into a land of large estates with forced labor. On a smaller scale, Von Thünen like zones can be distinguished around the urbanized region of northern Italy and within the Ottoman Empire (around the town of Istanbul).

In some parts of Europe large areas of land were reclaimed. In the Netherlands, for example, the draining of lakes started in the first half of the sixteenth century, but reached a summit during the first half of the seventeenth century. In the second half of that century, lower prices for agricultural products made new reclamations uneconomic and the draining of natural lakes stopped for two centuries. In Central and Eastern Europe a new wave of settlers, many of them fleeing for religious prosecutions, settled in thinly populated or partly de populated regions. Reclamations, but also the needs of a growing population, led to an increasing pressure on forests and common lands. All over Europe during the sixteenth century people complained about disappearing or degrading forests.

In most regions, however, the emphasis was on improvements in agriculture. This period showed an increasing specialization in many regions. In the Low Countries, around the staple market of Amsterdam, an intensive and commercial agriculture developed, in which large inputs and modern methods brought high yields per hectare. In the European core region, particularly in the Netherlands and in England during the seventeenth and eighteenth centuries, the spread of crops such as clover and turnips brought higher arable yields, as well as extra animal fodder. The introduction of new crops such as potatoes in northwestern Europe and maize in Mediterranean Europe gave high yields per hectare and made larger rural population densities possible. These innovations were part of what was later called the 'agrarian revolution'.

In some regions, very intensive and balanced types of farming developed. In the semiperipheral sand belt of the southeastern Netherlands and the adjoining parts of Belgium, Germany, and Denmark, the remaining com mons were more strongly integrated into the agrarian system, when the manure of cattle and sheep was mixed with sods – the so called 'plaggen' – and used for fer tilizing the arable. It led to a huge transport of organic and mineral materials from the commons to the arable. The system remained in use until the end of the nine teenth century.

Specialized crops, such as hemp, hop, and tobacco, became important for specific regions. Some regions, particularly close to urban centers, specialized in dairy farming or horticulture. In regions that changed from grain to cattle, the old open fields became dysfunctional and were often enclosed. Within a few decades open landscapes became characterized by hedges and dis persed farms. During the sixteenth and seventeenth centuries, this happened, for example, in parts of the Allgäu in southern Germany and in the country of Herve near the industrial town of Liège (Belgium).

But also for the remaining arable regions, the old open fields were no longer the optimal landscape. The more complex rotations led to a more individual management of the open fields. Most of the new reclamations were highly individualized, with enclosed fields and dispersed farms. Enclosures were now also being applied in open fields that remained in use as arable. In England, the enclosures speeded up during the phase of parliamentary enclosure (from 1720). Elsewhere in Europe, enclosures took place on a small scale during the fifteenth and six teenth centuries, but became large scale during the eighteenth and nineteenth centuries, particularly in southern Scandinavia (eighteenth century) and parts of Germany (eighteenth and nineteenth century).

Table 1 shows some interesting differences within northwestern Europe. Whereas in the UK the emphasis was on increasing labor productivity, the more densely populated and more urbanized Low Countries show particularly increases in land productivity. This is the main explanation for the early enclosures in the UK; in the Low Countries, a reconstruction of the field pattern was less urgent and enclosures (*ruilverkavelingen*) only started during the twentieth century. France is yet an other story, being characterized by a less intensive agri culture in terms of labor, as well as land productivity. Here also, the enclosures (*rémembrements*) only started during the twentieth century.

Table 1 Estimates of labour productivity and land productivity in agriculture in England & Wales, France, Belgium and the Netherlands, *c.* 1810. Source: Zanden, J. L. van (1994). *The transformation of European agriculture in the nineteenth century: The case of the Netherlands*, p 2. Amsterdam: Vrije Universiteit

	Production per laborer	*Production per ha*
England and Wales	939	104
Netherlands	910	159
Belgium	591	222
France	547	89

The Nineteenth and Early Twentieth Centuries

The Agrarian crisis of the 1880s brought yet another transformation. Again, this was mainly a grain crisis, caused by the competition of Russian and, particularly, American grain. The transport revolution – intercontin ental trains and ever larger transatlantic steamships – tipped the balance and within a few years American grain, cheaply produced on large farms, flooded the European market.

European countries reacted differently, dependent on internal balances of power. France, Germany, and Italy, for example, started protecting their farmers, but the urbanized and industrialized Britain chose for free trade, as did small trade oriented countries such as Denmark and the Netherlands. The last two countries developed into large exporters of agrarian products. Here, the crisis stimulated radical changes in agriculture. In the sandy landscapes of the southeastern half of the Netherlands, the farmers had for centuries kept animals mainly to get enough manure for their poor arable lands. Now they started specializing in intensive production of dairy products and meat for export, using the cheap grain from the arable as animal fodder. Another large change in the sand belt was the division of the commons, that still oc cupied more than half the total surface. In the 1890s, the availability of chemical fertilizers made it possible to turn most of the former commons into arable and pasture.

The next large crisis, during the 1930s, more or less repeated the earlier one. Again a number of countries, with Germany as the best example, turned to market regulations. Many countries even turned to import re strictions to protect their own farmers. This global crisis was followed by World War II, with an increasing state intervention in the occupied countries. During the last year of the war, scarcities of food occurred in some re gions, including a number of towns in Holland. Although a logistics rather than a production problem, the feeling of scarcity influenced post war agriculture, which be came characterized by an active role of the states.

The Second Half of the Twentieth Century

From the 1950s, agricultural policy was a main driving force within the growing European Economic Com munity (EEC). Already during the 1950s, agriculture in France and the Netherlands had recovered from wartime damage and searched for new export markets. Germany, that had lost many of its most productive agricultural regions, aimed at reestablishment of its position as an industrial nation and expected to remain an importer of agrarian products. These differences between France and Germany became the basis of the new EEC. In Eastern Europe, the implementation of socialist economic prin ciples in the Soviet sphere of influence brought far reaching changes. In Eastern Germany a short lived land reform divided the old estates into family farms, but after a few years the old estates were reestablished as col lective farms. Collective and state farms were also es tablished in other parts of Eastern Europe, most rigorously in the Soviet conquered Baltic States. Else where however, particularly in most parts of Poland, small scale farmers were left untouched.

The second half of the twentieth century has been an extraordinary period for European agriculture. During the Cold War, agriculture in the EEC, as well as in its Eastern European counterpart Comecon, was character ized by a strong influence of (supra) national govern ments on price levels within a protected market. The main reasons to protect the agricultural sector were the characteristic fluctuations in prices, for a product that was essential for survival. The fluctuations were mainly caused by external causes, such as weather and diseases, which were difficult to control. Moreover, the inelastic demand had in the past caused extreme fluctuations in prices. Support for agriculture therefore aimed at sta bilizing prices and, by doing so, production. Within the EEC, another factor was the effective lobby by farmers' organizations. Even after new technologies (pesticides, selection) had diminished fluctuations in production levels and surpluses became a growing burden on the budget, farmers' organizations still successfully lobbied for continuation of subsidies.

In fact, many of the uncertainties that had always characterized agriculture, now diminished, as production circumstances were increasingly controlled and pests seemed ever more controllable. Agriculture moved from risk management toward profit maximization. The land scape in Western Europe developed further in the dir ection of dispersed farms with enclosed land. This process was speeded up by enclosures, that now took place all over Western Europe. These rural restructurings (German: *Flurbereinigung*; French: *rémembrement*; Dutch: *ruilverkaveling*) became a highly planned activity, stimu lated and often executed by the national governments. The process took place in a period of relatively low world market prices and rising labor costs and strengthened the processes of scale enlargement and mechanization in European agriculture. In 2000 the 15 member states of the European Union together counted 9 million farms, and agriculture and forestry together employed some 9.5 million people. Fifty years earlier, in the same countries 20 million farms still employed 30 million people.

The extraordinary period of extreme state inter vention in agricultural production is now coming to an end. In Eastern Europe a phase of reprivatization is taking place since the collapse of the socialist system in 1989. It is certainly a radical transformation, but there is also a remarkable degree of continuity: locally, for ex ample, in parts of Eastern Germany, a direct line runs from the old country estates to the post war collective farms and to the present large scale capitalist farms.

The tendency toward a more open world economy, which is proclaimed by the World Trade Organization and the large non European agrarian exporters, is ex pected to lead to a gradual abolition of agrarian subsidies. One result of the successful process of enlargement of scale in agriculture is, that in most European countries less than 5% of the population is still involved in farm ing. The few remaining farmers lost most of their polit ical influence. The future of European agriculture is uncertain, and future transformations are to be expected.

With all transformations, there is also a degree of continuity. Actual maps of differences in productivity in agriculture still show – just like four centuries ago – the most intensive agriculture in the northwest European core region. Also, the main landscape types, that essen tially go back to the Middle Ages, are still recognizable.

See also: Field Systems and Enclosure; Historical Geographies, Rural; Post-Productivist and Multifunctional Agriculture.

Further Reading

Bartlett, R. (1994). *The Making of Europe; Conquest, Colonization and Cultural Change 950 1350*. London: Penguin.
Bork, H. R., Bork, H., Dalchow, C., Faust, B., Piorr, H. P., and Schatz, T. (1998). *Landschaftsentwicklung in Mitteleuropa*, p 221. Gotha/ Stuttgart: Klett Perthes.
Boserup, E. (1965). *The Conditions of Agricultural Growth; the Economics of Agrarian Change under Population Pressure*. London: Allen & Unwin.
Butlin, R. A. and Dodgshon, R. A. (eds.) (1998). *An Historical Geography of Europe*. Oxford: Clarendon.
Clout, H. D. (1993). The recomposition of rural Europe. *Annales de Géographie* 100, 714 729.
Clout, H. D. (1998). The European countryside: Contested space. In Graham, B. (ed.) *Modern Europe; Place, Culture and Identity*, pp 289. London/Sydney/Auckland: Arnold.
Grigg, D. B. (1980). *Population Growth and Agrarian Change; an Historical Perspective*, p 46. Cambridge: Cambridge UP (Cambridge Geographical Studies 13).

Grigg, D. (1992). *New Perspectives on the Past: The Transformation of Agriculture in the West*. Oxford: Blackwell.

Nitz, H. J. (1989). Transformation of old and formation of new structures in the rural landscape of northern central Europe during the 16th to 18th centuries under the impact of the early modern commercial economy. *Tijdschrift van de Belgische Vereniging voor Aardrijkskundige Studies* 58, 267 290.

Nitz, H. J. (1993). The European world system: A Von Thunen interpretation of its eastern continental sector. In Nitz, H. J. (ed.) *The Early Modern World System in Geographical Perspective*, pp 62 83. Stuttgart: Steiner (Erdkundliches Wissen 110).

Renes, H. (2004). Landbouw en cultuurlandschappen. In Pater, B., de Béneker, T. & Buunk, W. (eds.) *Europa: Ruimtelijke Samenhang en Verscheidenheid in de Europese Unie*, pp 69 92. Assen: Van Gorcum.

Spek, T. (1996). Die bodenkundliche und landschaftliche Lage von Siedlungen, Ackern und Graberfeldern in Drenthe (nordliche Niederlande); eine Studie zur Standortwahl in vorgeschichtlicher, fruhgeschichtlicher und mittelalterlicher Zeit (34 v.Chr. 1000 n.Chr). *Siedlungsforschung; Archäologie Geschichte Geographie* 14, 95 193.

Van Zanden, J. L. (1994). *The Transformation of European Agriculture in the Nineteenth Century: The Case of the Netherlands*. Amsterdam: Vrije Universiteit.

Agricultural Land Preservation

K. B. Beesley and D. Ramsey, Brandon University, Brandon, MB, Canada

Glossary

Covenant The voluntary surrender of property rights through legally binding agreements.

Easement It is also referred to as donated development rights, where a less than fee simple interest in the land is voluntarily donated or sold by a landowner to a unit of government or recognized nonprofit conservation organization for the purpose of protecting significant open space, recreation, ecological, agricultural, or historic resources.

Multifunctionality It is a concept that acknowledges that a particular resource base serves more than one purpose. In agricultural systems, for example, the economic, social, and ecological values of agriculture are recognized.

Purchase of Development Rights It refers to the voluntary sale of the development rights of one's land, a more permanent strategy for protecting land, but more costly than transfer of development rights.

Smart Growth It is a land-use planning concept, more generally referred to as growth management, in areas adjacent to urban centers.

Stewardship It is an effort to create, foster, and enable landowners and resource users to responsibly manage and protect land and natural resources.

Transfer of Development Rights It refers to transferring the development rights within an agricultural preservation district to a nonagricultural area, and is less costly than purchase of development rights.

Urbanization It refers to processes of spatial expansion of urban centers into the countryside.

Why Protect Agricultural Land

The success of today's agricultural sector is vital to the economy of most nations. In Canada, for example, agriculture is the third largest contributor to the gross domestic product and employs over 762 000 people, but its supply of agricultural land is limited and unevenly distributed, with only 11% of the land base capable of producing any agricultural product and less than 5% capable of producing crops. Agriculture also plays an important role in the American economy, accounting for 10% of all exports, employing over 3.1 million people on over 2 million farms.

The number of farms in most of the industrialized world has steadily decreased. Meanwhile, both the intensity of farming activities and the average farm size have increased, indicating the consolidation of smaller farms into bigger units. Today, farmers comprise less of the workforce; farms have become more specialized, more efficient, and more productive. These trends, and associated urbanization pressures, are stressing the land base and contributing to a decline in the availability of productive agricultural land.

Urbanization of Rural and Agricultural Land

In any country with limited productive agricultural land, the loss of farmland to urban uses is a serious concern. Urbanization of rural and agricultural land has various resource, social, and environmental implications. Many cities were established upon productive agricultural land and the characteristics that make land suitable for agriculture also make land appealing for development. As a result, pressure continues to be placed upon agricultural land for purposes other than agriculture, especially residential, industrial, commercial, and recreational uses. The emerging pattern on the landscape is characterized by low density urban development, particularly in North America. In the more densely populated countryside of the United Kingdom (UK) and Western Europe, there are greater pressures for development on productive agricultural lands.

The Effects of Urbanization on Agricultural Land Use

Urbanization has a variety of impacts on agricultural land use that combine to threaten the viability of the agricultural industry and the natural resource base. Urbanization influences agricultural land use most directly by converting agricultural land to urban uses: a permanent, irreversible consequence. Indirectly, urbanization influences agricultural land use through changes in the economic environment of urban fringe areas. As the demand for land in the urban fringe grows, the pressure on land prices to reflect urban values, rather than farm use values, increases. When farmers and rural landowners can receive more money by selling land for uses other than farming, they can no longer economically justify keeping the land for farm use. Alternatively, there may be abandoned farms, idled lands, or financial problems for those who continue to farm. As urban and suburban residents move into agricultural areas, conflicts inevitably arise which may threaten or restrict farm operations, such as nuisance complaints, vandalism, trespassing, and robbery. Urban encroachment onto farmland may also cause changes in farmers' status in their communities, increase land speculation, taxation, and demand for public services, and result in an increase in nonfarm ownership of land. Farmers facing these effects may respond by selling the land, intensifying production, or reducing investments in the farm operation.

Preservation for Food Security

The conversion of agricultural land to urban or nonagricultural uses is a critical threat to the long term health of the agricultural industry. A secure land base is essential for agriculture and can only be assured if mechanisms are in place to protect agricultural land.

Two opposing views exist over the need to preserve agricultural land for food and fiber production. On the one hand, economists question the need to preserve agricultural land because of food surpluses, low prices for the producer, and the bankruptcy of farms. Many argue that there is no need for farmland protection policies because the free market will effectively allocate land for various uses. On the other hand, preservationists support the claim that domestic sources of food should be protected because the extent of high quality agricultural land is limited, and the urbanization of farmland is irreversible and could lead to food shortages.

It is clear that lands with the highest agricultural potential are continually being converted to nonagricultural uses. Meanwhile, less productive, marginal land is being brought into production to offset the losses of prime agricultural land. Bringing new, lower quality lands into production involves substantially higher costs, reduced profits, and environmental damage. The need to preserve agricultural land for food security exists more because of the quality, not quantity, of land being converted to urban uses. Preserving farmland may reduce transportation costs associated with food imports, reduce dependency on foreign food supplies, and allow consumers to enjoy fresh, local produce.

Preservation for Open Space, Environmental Sustainability, and the Rural Economy

In addition to its function as a food and fiber producer, agricultural land provides wildlife habitat, areas for groundwater recharge, and open space for urban dwellers. Protecting agricultural land can also preserve farming as a pastoral ideal and protect the rural economy. Preserving farmland for amenity and natural environment reasons helps to protect the rural character of the countryside, environmentally sensitive areas, and visually significant resources. While strategies to preserve farmland may contribute to the protection of open space and natural resources, true farmland preservation strategies aim to protect land for agricultural use.

Approaches for Preserving Agricultural Land

Since the realization was made that the conversion of agricultural lands to urban uses in the post World War II era was having an impact on the ability of farmers to operate, and threatening future food security in the Western world, a range of approaches have been implemented. Some were designed specifically for, or dominated by, agricultural interests. Others are rooted more broadly in the desire to preserve or maintain natural resources and ecosystems within agricultural contexts. This section briefly outlines some of the primary approaches, which vary greatly throughout the Western world, based on three basic types: incentive programs, land use controls, and integrated programs.

Incentive Programs

Focused in agricultural areas of the industrialized world, and most commonly in the United States, incentive programs are designed to deter landowners from converting their land to nonagricultural uses. Incentive programs focusing on agriculture are of three general types. The most common type focuses on tax incentives, or differential tax assessments, which can include restrictive agreements, deferred taxation, and preferential assessment. Generally speaking, lands are taxed based on agricultural value rather than the general market value, thus reducing tax levels paid by farmers.

The second type of incentive program is agricultural districting. With this strategy, landowners voluntarily have their lands included in an agricultural district for a fixed, but renewable, period of time during which the landowners must maintain the land. This strategy is less costly than tax incentives and tends to be more popular where urban pressures, and therefore property market values, are lower. However, it is not a permanent solution; landowners are not required to participate, particularly given that financial incentives to participate do not always result in prevention of agricultural land conversion where pressures, and therefore market values, are greatest.

The third type of incentive refers to purchase of development rights (PDR) and transfer of development rights (TDR). PDR refers to the voluntary sale of the development rights of one's land, a more permanent strategy for protecting land. However, if the purchaser is a level of government, it can be quite costly. TDRs are a less costly version and include transferring the development rights within an agricultural preservation district to a nonagricultural area. Once purchased or transferred, lands are designated a preservation district for the present and future owners of said land.

Land-Use Controls

Land use controls are primarily a post World War II response to the loss of agricultural lands to other uses. The UK system of state planning, for example, was built on the recognized need to preserve high quality agricultural lands. The land use control approach for preserving agricultural lands is obtained through control mechanisms, including land use zoning, land banking, and right to farm legislation. Land use zoning for agricultural land preservation is more widely adopted in North America, focused on agricultural land uses, and located in areas adjacent to larger urban centers where development pressures are the greatest. The most widely referred to examples include the States of Oregon, Hawaii, and Wisconsin in the United States and the Provinces of British Columbia, Québec, and Newfoundland in Canada. Two types of zoning can be implemented: exclusive, whereby only agricultural activities are allowed within an agricultural zone, and, nonexclusive, which allows for limited nonagricultural development to take place. In theory, land use zoning controls are the most affordable because the appropriate level of government simply restricts particular lands from being developed to nonagricultural uses. This is similar to restricting residential development to areas zoned for that development in an urban center. However, in practice, such mechanisms can be difficult to maintain over time.

Land banks, by comparison, are more costly as they involve a particular level of government purchasing the land from farmers, who then retain the development rights. This mechanism can be either voluntary or compulsory. While an expensive action, land banking does encourage orderly growth and provides the opportunity to acquire agricultural lands when their values are lower. The first and most successful example is perhaps Saskatchewan, Canada.

Unlike the initiatives discussed thus far, land trusts are not focused solely on agricultural lands, but can include open spaces and areas of cultural or ecological value. A land trust is a private, nonprofit, citizen based organization that comes together to protect land through purchase or other legal mechanisms. Operating at a local or regional scale, land is purchased or donated. In North America, such initiatives are referred to as easements. In New Zealand, these types of arrangements are referred to as conservation covenants. These are defined as the voluntary surrender of property rights through legally binding agreements. In addition to the National Trust, in the UK, a similar initiative relates to set aside programs and both voluntary and compulsory set aside programs have been implemented.

In the United States, the value of the land donated can be claimed as a tax deduction. The land trusts, or conservation easements, can be for a range of purposes, including the preservation of agricultural, open space, ecological, recreational, historic, or scenic resources. The approach is implemented through the creation of conservation easements within particular land holdings, also referred to as donated development rights, where a less than fee simple interest in the land is voluntarily donated or sold by a landowner to a unit of government or recognized nonprofit conservation organization for the purpose of protecting significant open space, recreation, ecological, agricultural, or historic resources. The easement includes limits to activities on designated portions of a land holding and these limits are legally added to the land title.

Unlike the approaches discussed thus far, right to farm legislation does not protect the land from conversion, but rather protects farmers from complaints about their agricultural activities, often referred to as nuisance complaints. The implementation of such legislation is most pronounced near urban areas and in rural residential enclaves, where residents raise concerns about noise, smell, and dust resulting from farm activities. In fact, right to farm legislation can assist farmers who wish to expand their operations, including converting natural areas (e.g., forest clearing and draining wetlands) for agricultural purposes. Right to farm legislation has been enacted in various jurisdictions, primarily in areas adjacent to urban development; however, such legislation may offer farmers a false sense of security from litigation.

Integrated Programs

A range of integrated programs exist, including land stewardship, multifunctionality, and smart growth initiatives. Whereas land trusts and easements result in a restriction on development either in perpetuity or for a term limit, land stewardship is a voluntary program focused on how resources are managed. Stewardship programs are also not limited to agricultural land, but can include forested lands and areas of natural significance, such as wetlands. The more general nature of such initiatives is indicated in their efforts to create, foster, and enable landowners and resource users to responsibly manage and protect land and natural resources. Several examples can be cited, including: Canada's landowner contact programs, the UK's countryside stewardship scheme, and Australia's land care program.

One of the more recent approaches to protecting natural resources, and one that has been the center of recent World Trade Organization (WTO) debates, is multifunctionality – a concept first officially presented as part of Agenda 21 at the Rio Earth Summit in 1992. It centers on the complementarities between landscape preservation and food security. Multifunctionality in Europe and the UK marks a shift from planning designed to protect elite landscapes (e.g., national parks and greenbelts) toward encouraging the development of multifunctional agricultural enterprises. While originating in the UK and European countries, multifunctionality is beginning to be implemented in other jurisdictions, such as Australia and the United States. As the term connotes, multifunctionality seeks to acknowledge that a particular resource base serves more than one purpose. In agricultural systems, for example, the economic, social, and ecological values of agriculture are recognized. As a program of implementation, multifunctionality seeks to offer farmers remuneration for ensuring the protection of the biophysical aspects of their farms. Like other strategies, the actual implementation of multifunctionality varies. The implementation of the environmental dimension of multifunctionality is similar to the notion of land stewardship.

In contrast to easements and multifunctionality, smart growth, or more generally growth management, is the focus of areas adjacent to urban centers. From an American perspective, there are three main types of urban sprawl: expanding urban/suburban areas that push into the countryside at low to medium densities; linear commercial sprawl along major arterial roads leading to and from cities and suburbs; and scattered residential sprawl, or country residential development, outside established settlements at low densities.

One of the first attempts to control urban sprawl in the United States took place in Hawaii in 1961, when land was divided into four categories: urban, rural, agricultural, and conservation. However, given the pressures of growth in the tourism sector and the restructuring of the agricultural sector, the result was still land lost to urban sprawl. One of the arguments for instituting smart growth planning is not just to preserve agricultural land and natural resources, but to reduce the costs of servicing (water, sewer, and transportation) that is necessitated with low density urban development.

The Future of Agricultural Land Preservation

Government responses to the urbanization of agricultural land have recognized the need to develop programs to support agricultural viability in conjunction with agricultural land preservation policies. However, policies to protect agricultural land must account for the uncertainty of the agricultural economy. If farming becomes too unprofitable an occupation, it will not be economically viable for governments or private landowners to preserve farmland in anticipation of changes in the market. Similarly, if high quality farmland is not preserved, farmers will be forced to farm marginal lands that will be less sustainable and harm the resource base and economy in the long term. Clearly, agricultural land preservation programs must coexist with programs supporting the agricultural economy to ensure the sustainability of the land resource, the long term viability of the industry and the food supply. Farming is not just a lifestyle, it is an economic activity, and if it cannot be undertaken in a way that allows farmers an adequate life, then how much farmland has been saved from bricks and mortar matters little.

No single farmland preservation strategy has proven entirely effective and it is unlikely that a perfect strategy exists. Similarly, there is no existing set of universal characteristics that can guarantee farmland preservation and a viable agricultural industry. A strategy that works in one jurisdiction may fail in another. The success or failure of a strategy depends on a number of political, social, and environmental values and circumstances: land productivity, income levels of area residents, level of planning experience, skills and leadership of government officials, public attitudes toward land use control, land ethics, and the need, perceived or real, to preserve farmland.

Nevertheless, for a farmland preservation strategy to succeed, certain characteristics should exist. The presence of these characteristics may also be used to measure the effectiveness of farmland preservation policies and programs. A comprehensive list of criteria, which may be used to evaluate the effectiveness of a proposed, or existing agricultural land preservation program includes: does the program identify high quality agricultural lands to be protected? Does the program prevent or limit land use conflicts? Does the program provide certainty that significant agricultural lands will remain available for agricultural use? Does the program provide an integrated approach to agricultural land use planning and meet the objectives of economic and environmental sustainability? Can the program be implemented using existing resources? Can the program be implemented readily? Will it provide immediate results? Is the program supported by farmers, municipalities, and the public? Does it address the concerns of these groups? Is the program affordable? Will it negatively affect the agriculture and food industry in the long term? Is the program compatible with other land use planning and financial programs? These criteria are useful measures of policy effectiveness. While some argue that policies and programs to preserve agriculture should focus on land use planning rather than exclusively on other initiatives, such as tax incentives, others argue that preservation initiatives do not guarantee viable agriculture if they focus too much on physical land use planning.

In contrast to the criteria presented above, a purely economic view argues that farmland preservation strategies will not be effective unless they influence the rural land market so that farmers are not outbid by households seeking rural nonfarm lifestyles. In other words, preservation strategies must reduce or eliminate the impermanence syndrome and the speculative value of land; otherwise, farmers will sell their land and/or cease farming. In this context, preservation strategies must in corporate social and economic considerations and work in cooperation with farmers, families, and professional associations.

While many areas appear to account for various social, environmental, and economic issues by using two or more strategies to preserve farmland, these are rarely co ordinated into a larger framework. Legislation and policy influencing agriculture and land use are developed and enforced by various government departments and agen cies. Land use planning for agriculture and the economic health of the agri food industry are generally regarded as separate issues, yet they are very much connected. If farming cannot remain a viable industry, then farmers will choose other occupations, thereby freeing up land for consumption by other uses. In the meantime, if there is no intervention to protect the agricultural land base, then farmers may choose to make a quick profit by selling their land when development pressures escalate.

Agriculture is an important industry in most nations, and the most basic resource – land – should be a national concern. While the forces behind agricultural land con version are often found at a broader scale than that at which land use planning is normally carried out, it seems clear that all levels of government should play a role in slowing or preventing agricultural land conversion, and preserving the agricultural land base. While municipalities are primarily left with the often voluntary responsibility of protecting farmland through local land use planning (usually under provincial land use law or policy), most upper level governments, take charge of agricultural re search and financial support programs. Further, strategies to preserve farmland need to coordinate preservation policy with the development of a viable and environ mentally sustainable agricultural industry. These strategies should be part of an integrated effort within and among government departments and jurisdictions, because issues affecting agriculture and land use occur at various scales, often transcending political boundaries. Additionally, land use strategies which integrate conservation techniques involve farming and local populations, and combine land use planning with other forms of social and economic development to hold more promise for achieving farmland preservation, and social and economic goals.

Conclusion

In response to growing concerns about the loss of farmland to nonagricultural land uses, many jurisdictions have adopted a variety of approaches to preserve farm land. Policy and legislative tools are used to combat the loss of farmland in different ways with varying degrees of success. Most areas use land use planning controls to influence the use and allocation of agricultural land, and tax or other incentives to encourage farmers to keep their land in agricultural use. This combination of techniques attempts to protect land for agricultural use while keeping farming profitable for farmers. While farmland conversion has generally been slower than it would have been in the absence of farmland preservation strategies, it has not been eliminated. The future of farmland pre servation depends on the coordinated efforts of all levels of government and the integration of economic, social, and environmental issues into a comprehensive land use program designed to preserve farmland and the agri cultural industry.

See also: Agri-Environmentalism and Rural Change; Agriculture, Sustainable; Counterurbanization; Edge Cities; Resource Management, Rural; Suburbanization; Urban-Rural Continuum.

Further Reading

Beesley, K. B., Millward, H., Ilbery, B. and Harrington, L. (eds.) (2003). *The New Countryside: Geographic Perspectives on Rural Change*. Brandon: Brandon University, Rural Development Institute and Saint Mary's University.

Bryant, C. R. and Johnston, T. R. R. (1992). *Agriculture in the City's Countryside*. Toronto, ON: University of Toronto Press.

Bunce, M. (1998). Thirty years of farmland preservation in North America: Discourses and ideologies of a movement. *Journal of Rural Studies* 14, 233 247.

Daniels, T. and Lapping, M. (2005). Land preservation: An essential ingredient in smart growth. *Journal of Planning Literature* 19, 316 329.

Furuseth, O. J. and Lapping, M. (eds.) (1999). *Contested Countryside: The Rural Urban Fringe in North America*. Hampshire, UK: Avebury Press.

Gilg, A. W. (ed.) (1991). *Progress in Rural Policy and Planning*. London: Belhaven.

Hamin, E. M. (2002). Western European approaches to landscape protection: A review of the literature. *Journal of Planning Literature* 16, 339 358.

Woods, M. (2005). *Rural Geography*. London: Sage.

Relevant Websites

http://www.alc.gov.bc.ca/
The Agricultural Land Commission, British Columbia, Canada.
http://www.defra.gov.uk/
The Department for Environment, Food and Rural Affairs, UK.
http://www.farmland.uoguelph.ca/
The Ontario Farmland Trust.
http://www.farmland.org/
The American Farmland Trust.
http://www.farmlandinfo.org/
The Farmland Information Center.
http://www.farmlandpreservationreport.com/
The Farmland Preservation Report.

Agriculture, Sustainable

T. Marsden, Cardiff University, Cardiff, UK

High-Level Conceptualizations: Some Foundation Principles

Since the 1960s the concept of sustainable agriculture has been increasingly used to denote a variable but significant rupture with what has been termed the more dominant agri industrial model of agriculture and rural development. In general terms it can be defined as re ferring to the environmental or ecological soundness of the agricultural production processes or the wider agri food commodity chain. This implies that with respect to renewable nutrient cycles, such sustainable systems are those that are capable of being continued indefinitely, or at least over considerable generational time. In many zones we might argue that agricultural systems have long been sustainable (since Neolithic times), in that while they may have been subject to considerable adaptation and technical changes, they have sustained in reasonable shape the main factors of production needed for the continuance of food production both 'in' and 'of' place. In addition, they have managed to do this without significantly destroying the local and production based ecologies on which they are based.

It has also long been established that agriculture, as a production process, is quite distinctive in that it both produces and transforms nature, through the (re) pro duction of livestock and plants, at the same time as being physically and geographically reliant upon nature as a 'means' and 'condition' of production (most notably soils, micro climate, topography, and vegetation complexes). Agriculture, and especially a sustainable agriculture, thus efficiently transforms, as well as recreates its space or *terroir* as part and parcel of the production process. Usually the indefiniteness of sustainable agriculture makes primary reference to a set of ecological principles, and it is also linked to not so much an endpoint, *sui generis*, but rather as a continual and increasing 'contested process'. In this sense it needs to be recognized that there will always be ways in which agro food systems can be adapted, which make them more (or less) ecologically, economically, and socially sustainable. In general and abstract terms then the 'process' of sustainable agri culture must infer a 'coordinated relationship between humans, organisms, and environment, which is adaptive to location, season, creating biophysiologically or eco logically mutual benefits which give complementary status to the different components of the production system'.

Such explicit complementarities were a central part, for instance, of ancient Chinese classical works written about 3000 years ago. Here agriculture needed to be practiced as part of the harmony between heaven, Earth, and humans; and between the careful utilization of the five key elements of the universe: that is, metal, wood, water, fire, and earth. These principles and adaptive techniques were passed down under systems of 'inherited experience'. For instance, in the Tai Lake basin area of Southeast China, the grain–livestock–mulberry–fish in tegrated production system has existed for more than 1000 years, by means of a cycling based on ecological food chains and synergies between the different but complimentary production systems.

In many cases both now and in the past, therefore, we see examples of long standing agricultural sustainability. This has, however, increasingly been associated with those areas (often regarded under the twentieth century agri industrial modernization project) as being geographically marginal to the mainstream, lying beyond the productivist industrialization of agriculture. Sustainable agriculture is commonly associated in this general and abstract sense as being ecologically self sustaining with low inputs, eco nomically viable, of a sufficient scale and diversity which limits long term changes to the environment, and is eth ically and aesthetically acceptable. More recent writers often tend to define it in oppositional terms with the agri industrial model of development, involving a rejection of high external input systems and encouraging greater in volvement of local farmers, protecting their unique knowl edge systems, and in incorporating their active resource conservation techniques within a production framework.

These oppositional definitions have grown since the 1970s, as the environmental, social, and economic problems associated with the agri industrial model have come to the fore. Of particular importance here is the ability of the sustainable agricultural systems to maintain levels of prod uctivity in the face of periodic shocks or stress. Rather than using external inputs, like fertilizers and pesticides, that usually have to be purchased at prices outside of the control of the farmer, internal resources on the farm or in the community, such as natural parasites and predators of pest, algae, bacteria, and green manure providing nitrogen, agroforestry and cropping systems reducing erosion, underexploited wild tree and fish species, and indigenous crop varieties with tolerance to salts and toxins, provide the potential for being both renewable and ecologically more efficient.

As these principles have reemerged, differing stances on agricultural sustainability have developed between, for instance, farmer groups (stressing especially the sustainability of the farm business and family farming), and a widening array of environmental interests who stress 'deeper' ecological principles associated with reducing waste and enhancing systematic recycling and renewable practices. These divisions in thinking have led to different expressions of sustainable farming practice from organic certified farming, through to community supported agricultures, short food supply chains, and the rapid development over the past decade of different types of farmer driven 'farmers markets'.

Widening and Grounding Key Principles and Concepts

Up until the 1970s the agri industrial model, which had been superimposed upon much of the advanced world both in its capitalist and state socialist variants, had managed to control and legitimize itself concerning its longevity and general societal acceptance. It was as if the ideology of growth, specialization, intensity, mechanization, and increasing yields were to be never ending. From that point, however, and following Rachael Carson's trenchant critique of pesticides and insecticides in her book *Silent Spring* in 1962, it became clearer to the public at large that the agri industrial model which had helped to produce 'cheap foods' for the Fordist and urban growth machine, was largely about 'sustaining the unsustainable' rather than an ecologically efficient way of food production.

In ecological–economics terms it was increasingly being realized that the agri industrial system was falling victim to the second law of thermodynamics (entropy law); in that it was becoming a highly wasteful system in energy terms, with a continuous degradation of free (usable) energy into bound (unusable energy). However, this did not mean that the agri industrial model, as an unsustainable and entropy maximizing system, does not have an enormous amount of resistance to change or 'staying power'. Rather, as we have seen in the advanced economies of the past 20 years, and despite the depletion of the 'natural capital' of agroecosystems, the agri industrial/bioscience system has a remarkable longevity.

Nevertheless, the slow realization of its inherent unsustainability was taken up by environmental non governmental organizations (NGOs) and different publics and consumer groups, and became gradually institutionalized into government bodies at national and supernational levels (e.g., the European Union's series of Environmental Action Programmes from the 1970s onward). This period of 'ecological modernization' was embodied in the World Commission on Environment and Development's 'Bruntland' report, 'Our common future' in 1987; with agriculture being increasingly seen as a threatened 'renewable resource'. Writers have delineated several key features of the 'sustainability of unsustainability' thesis, which is implicit within the agri industrial model. These include: (1) the tendency toward monoculture/specialization of production systems and the creation of spatial homogeneity and subsequent declines in spatial and biodiversity; (2) the intensification of production through expanded use of external chemical, energy, and irrigation inputs; (3) the concentration of livestock in space, and the spatial separation of crop and livestock production; and, (4) the overall supply chain tendency for the intensification of the cost price squeeze on the farming and land based populations as a result of their increasing reliance, both upon upstream suppliers of external inputs, and downstream buyers of farm gate products (especially the corporate retailers).

The result of these trends is important in the context of defining the more contemporary and increasingly contested notion of sustainable agriculture, for two related reasons. First, the profound implementation of the agri industrial model has tended to distinguish with 'more clarity what is and is not sustainable'; and second, it has acted to widen and encourage the contested vector of what constitutes sustainable forms of agriculture as an 'oppositional concept', as well as a concept in its own right. Sustainable agriculture debates have thus come to the fore and their political and geographical vector has both broadened and diversified into different strands of both agricultural practice (e.g., organics, fair trade, integrated farm management, ethical farming, and agroecology); and alternative agri food supplychain geographies (including relocalized chains, farmers markets, organic box schemes, and slow food circuits). It is important to recognize that these movements have been given extra vibrancy because of 'both' the longevity and continuous crisis tendencies inherent in the agri industrial model.

One recent example of this has surrounded the highly contested introduction of genetically modified (GM) crops into global food supply chains. Introduced as the latest more 'sustainable' 'panacea' for the productivist–intensive system with regard to reducing costs of production, as well as removing some of the threats of pests and diseases, it has acted also to fuel the alternative sector, creating real boundary disputes between rival 'sustainable systems'. Here sustainable agriculture also becomes liable to various linguistic appropriations and regulatory pressures which either attempt to continue to deny its existence, or where it becomes more accepted and entrenched, to attempt to 'conventionalize' and dilute its central tenets. Increasing tensions have thus grown between the adherence to the basic principles of sustainable agriculture and serious attempts to dilute and generalize its very ontology. This 'battleground' is seen, for instance, in the agriculturally productive state of California, where growing consumer and producer demands have led to innovative and sig nificant growth in a whole vector of sustainable agri cultural developments (not least organics) since the 1980s. At the same time, many of these more successful initiatives potentially fall victim to economic and regulatory pres sures to conventionalize and dilute their founding principles.

At the same time as sustainable agriculture debates are becoming more vibrant and promoted (both by dif ferent groups of consumers, environmentalists, as well as producers), we also see new challenges and devaluation attempts to its overarching principles, as outlined in the first section of this article. The second part of this article will outline two related conceptual and methodological avenues of research and conceptual development which have been attempting to provide and to situate sustain able agriculture within a more robust and conceptually distinctive set of parameters. First, these surround agroecological and ecological modernization debates; and second, more specific to the rural development domain, concern the development of what some scholars term the 'new rural development paradigm'.

Agroecology and Ecological Modernization Perspectives

A particularly forceful oppositional response to the conventionalizing and entropy maximizing features of the agri industrial system has been the growth and sali ence of agroecological approaches. Agroecology emerged in the 1980s as an important attempt to establish a natural and social scientific basis for alternatives to industrialized agriculture which attempted to avoid its resource de grading tendencies.

Agroecological frameworks draw upon different in tellectual traditions and disciplines, including peasant studies, ecology, environmentalism, and development theory. Originating in developing countries it has quickly spread to the North American and European intellectual traditions. It provides some key conceptual building blocks for a much more normative direction to the an alysis of human–nature relations and rural development, and potentially links more broadly to the (mainly European) policy debates surrounding ecological mod ernization. Several writers have attempted to synthesize some of its key principles and features. Here we can begin to summarize and build upon these within the contemporary context:

1. *Alternative definitions of modernity*. Here, the process of emancipation from the strictly economic sphere, and the gradual re embedding of ecology in the insti tutions of economy, is a central aspect across different spatial scales (e.g., local, regional, national, and international); creating the spaces for a socio ecological, as well as economic rationality. This in volves aspects of co evolution which incorporates the autonomous development of socioecological practices and principles within wider political and institutional structures.
2. *Co evolution*. Agroecology also refers to the reliant co development or co evolution of society and natural factors. It is recognized that farming systems essen tially result out of co production, that is, the ongoing interaction, mutual transformation, and dependency between humans, animals, and nature; and that the agri industrial model has created a biotic, biological, and metabolic rift between these elements. The question then becomes how reversible processes can be put in place to regenerate interdependency over time, for instance, in recreating natural manuring, rearing, and feeding techniques. Also, more recently, co evolution also represents the new reconnections that groups of consumers are now making with co production at the farm level.
3. *Local farmers' knowledge and innovation systems*. Local peasant or indigenous knowledge needs to be seen (both in the North and the South) as significantly different from standard scientific knowledge in that it is embedded in local ecologies and is encoded in local and regional cultures rather than more abstract, uni versalizing, and reductionist notions. Dominant no tions of science tend to be independent of social and local contexts, yet agriculture is actually defined both by its biophysical context and its localized socio political elements. It is based upon the interdependent accumulation of local, natural, and social knowledge resources and practices. These are not just about maintaining 'old cultures', but they involve the re shaping of knowledge systems around, for instance, new embedded entrepreneurial and social networking and marketing skills.
4. *Endogenous potentialities*. While all agroecological sys tems have their own endogenous potential, a major factor concerns how these potentialities can become articulated and valorized/commodified through social and political mechanisms. For instance, the role of the cooperative Chianti wine consortium in Tuscany serves as a major social and economic mechanism for over 30 key producers of Chianti wines, simul taneously protecting, regulating, promoting, and ar ticulating the endogenous potential of the local producers. These potentialities and their realization are often social struggles involving local groups who attempt to resist, oppose, and actively construct al ternatives to industrial standardization and modern ization. Many examples in the South (in Kenya, for instance) involve local groups of farmers who are

attempting to protect their rights and practices concerning the production of local seed varieties in the face of their appropriation by corporate seed firms. Creating and sustaining these endogenous potentialities becomes a key feature of agroecology, and it relies upon the development of levels of cooperation and collaboration within farming and rural communities.

5. *Collective forms of social action.* Agroecology, therefore, also relies upon new or recreated collective forms of social action, not only between producers, but also producers and consumers, and producers and other key actors in the supply chains (such as processors and retailers). New associations are forged which enable reconnections between production and consumption, not only of foods but of amenity and rural experience.
6. *Systemic strategies.* New and more holistic understandings between the broad range of biophysical factors, such as water, soil, solar energy, and plant and animal species, must be conceived in ways in which they interact not only among themselves but also with social actions and practices. This involves an understanding, for instance, of energy, material, cash, and knowledge flows generated in the processes of production and exchange of goods and services associated with the farm. Moreover, ethical decisions, for instance, concerning animal welfare, human diets, are at the heart of progressing more systemic sustainable systems.
7. *Ecological, cultural diversity, and agricultural multifunctionality.* Agroecology aims not only to celebrate cultural and natural diversity, but also to progress and materialize it in new co evolutionary ways. It accepts that there are diverse pathways of agroecological development; and that this is based upon both local and regional forms of embeddedness, on the one hand, but complex social–natural relations and producer–consumer linkages, on the other. This places an emphasis upon both the multifunctionality of farming and related practices on the farm, at the same time as recreating multifunctional linkages between farms and the wider rural and urban communities. Such diversity and multifunctionality is seen as a central plank of moving toward a more sustainable society more generally. Moreover, taking much evidence from the South, it demonstrates the food security, ecological and social benefits of a much more diverse agriculture. It is now more easily accepted that agroecological methods do not necessarily result in lower yields. While global industrial agriculture serves the worlds population with 90% of its calories from a mere 15 species of crops, organic and agroecological farmers are providing a vital service in maintaining genetic diversity for the future. For instance, indigenous farmers in Peru cultivate more than 3000 different types of potatoes, and there are more than 5000 varieties of sweet potatoes cultivated in Papua New Guinea. In West Java, researchers have identified more than 230 species of plants within a dual cropping system, which includes 'agroforestry', home gardens, and outfields. In Mexico, the Huastec Indians manage a number of plots in which up to 300 species are cultivated. Areas around the farmhouses may contain between 80 and 125 useful species, many with medicinal properties.
8. *Encompassing the social ecology component.* While agroecology embraces both *de facto* and certified organic, both of which are now rapidly growing in terms of both production systems and markets (globally there are more than 100 different organic certification systems in place, but this is only the 'tip of the iceberg' with regard to *de facto* organic and agroecological farming systems, especially in the South); it contains a more explicit social component than the strictly organic approach, where the focus is often more strictly tied to verifiable technical standards. Many agroecological movements do not provide internationally recognized standards and are culturally and locally more specific; often tied to a more 'farmer first' approach. Nevertheless, it is important to view agroecological developments as complimentary systems with the more rules based organic approaches, as both form significant components of the widening vector of sustainable agriculture and agri food. The systems share common methodological and theoretical ground also in their use of participative approaches to agricultural and rural research and development.

While agroecological and multifunctional principles and practices have burgeoned in most advanced countries, including the US, it has been in Europe, over the past two decades where such principles have become more embedded into the formal institutional fabric and agricultural and environmental policy frameworks. It has been here that the benefits of a more diversified and 'green' agriculture have been articulated in agricultural reform, and where more of a resistance through, for instance, the 'green box' discussions in World Trade Organization (WTO) negotiations has been at least articulated on the global stage. However, as will be shown below, these European ecologically modernizing trends have been associated more with the growing consumer/citizen concerns regarding the quality and provenance of food and a reconnected green amenity, as they have with a more positive and coherent farmer movement. Moreover, as we shall see, the particular expression of agricultural sustainability in a grounded and contemporary European context is much more embedded in reconnecting its agricultures back into a wider and more holistic notion of rural sustainable development more generally.

The Sustainable Rural Development Paradigm in Europe

Since the 1990s a fledgling, sustainable rural develop ment model has been emerging, at least in a European context which has incorporated many of the principles of sustainable agriculture, specifically, agroecology. While building upon the multifunctional assumptions, but also adapting many of the agroecological principles de veloped in Southern countries, this model attempts to reintegrate agriculture as a multifunctional set of prac tices. These can hold the potential to enhance the interrelationships between farms and people both in rural and between rural and urban areas. While accepting the realities of much of the cosmopolitan 'consumption countryside' in a European context, the new rural de velopment paradigm distinguishes itself in the emphasis and autonomy it promotes for a reconstituted and en gaged agricultural and land based rural sector. This is seen as a potentially new driving force for rural devel opment, both in some policy, as well as academic fora and involves what a group of scholars refer to as broadening, deepening, and regrounding of farm based activities (including new on farm activities like agri tourism, en ergy production, value added food chains, and a variety of environmental schemes).

Recent studies in Europe estimate that at least 50% of all farmers are actively engaged in some of these new rural development practices. This reasserts the socio environmental role of agriculture and other land based activities as a major agent in sustaining rural economies and cultures. It reconnects a renewed priority of agri cultural production to the wider markets and social in novations and possibilities (such as aspects of 'retro innovation' of agricultural practices, local embeddedness, and new forms of ecological entrepreneurship). This is based upon assumptions of the wider associational eco nomics and institutional economics theories which em phasize the significance of 'trust and association based' relationships in new forms of economic development. And, it potentially places sustainable rural development in a more central role in understanding the mechanics and complexities of twenty first century economic reorganization.

In this context agriculture acquires a more com prehensive meaning and displays high integrative po tential in that it is recognized as a central feature of delivering real rural sustainable development benefits. This does not necessarily rely upon significant amounts of production subsidy; and can engage the State in new and innovative ways (e.g., in aspects of marketing and procurement policies). To do this it is necessary to create a radical rupture with the agroindustrial processes.

Sustainable agriculture must, in a variety of ways, attempt to find new political, social, and ecological platforms and spaces to distinguish (and exert energy for) itself from the conventional processes which tend to continue to devalue its base. Empirically we see many examples of this across Europe in what has been termed the 'relocalization' of agri food. While these re embed ding tendencies have proliferated in some countries (such as Italy and France) more than others (such as UK and the Netherlands), it is now clear that even in those countries where the agri industrial model has historically had the most profound impact, that new 'sociotechnical niches', new spatially embedded networks and structures, are clearly developing.

This is not just associated with the rise of (*de jure*) organic food production and consumption in Europe (which has indeed been significant). It is also associated with *de facto*: (1) rises of both locality based and new locally based short food supply chains and (2) the dif ferent responses and ruptures from the conventional system brought about by its periodic and severe crises tendencies (such as bovine spongiform encephalopathy (BSE), foot and mouth disease (FMD), and Avian Flu). For instance, in South West England new and more au tonomous relocalized networks of locally and regionally based beef, dairy, and cheese production have developed partly as some producers have attempted to create a 'rupture' with the past conventional systems. Rather than continue to struggle under the continual pressures of the (agri industrial) conventional cost price squeeze on the one hand, or face the burdensome regulatory costs of the hygienic bureaucratic state on the other, these emerging networks have managed to significantly detach them selves from these 'lock in effects' by developing new sets of relationships both between themselves and with downstream buyers (such as retailers and caterers). It is also clear that these new networks begin to return con siderable economic value added for the local rural economy.

There are many critics of the new rural development paradigm given the resilient power of the agri industrial and bioscientific orthodoxies both in research and policy circles. For instance, many agricultural economists still see it as really only another novel form of niche pro duction in the global and WTO context of liberalization and continued agri industrialism – an archipelago within the sea of corporate inspired cost reduction. Some question its more widespread benefits both socially and economically, arguing that such new networks are in themselves socially elitist, and indeed display quite neoliberal tendencies of their own; or that they can also reflect the more sinister features of 'defensive localism' and protected markets. More optimistically, some sug gest that this movement begins to represent a sort of postmodern 're peasantization', whereby, networks of landholders begin to develop new organizational systems of autonomous and endogenous rural development; and thereby recreate significant ‘room for maneuver’. It is an interesting question as to how the ‘peasant’ – often re garded as a pejorative concept under the twentieth century agricultural modernization project – now po tentially reemerges as an empowered ecological entre preneur within the new rural development paradigm.

Some Conceptual Building Blocks for the Sustainable Rural Development Paradigm

One common mistake about the new rural development paradigm (within the contemporary European context, at least) is to assume that it is simply a return to localism, and/or that it falls victim to the ‘local trap’ in general sustainability literatures. This assumes that there is an unproblematic set of elisions between the local embeddedness and overall sustainability. It is important to correct these sets of assertions by pointing out that a key feature of the rural development paradigm is the dual (i.e., with a strong agroecological root) socioecological process of rebuilding local resources (lets call these ‘countryside capitals’) at the same time as enlarging and deepening the interactions with the wider economy. Hence, it is not sufficient to simply read off socio economic or ecological actions from the particular geo graphical scale within which they may reside; nor to assume that local actions by themselves are the panacea. Rather, it is important to examine how new local (en dogenous) and external interactions, as processes and practices, become constituted and reconnected in and through space. Nevertheless, this does not deny the sig nificance of local and regional scales as potentially be coming new sites of resistance and innovation for the development of more sustainable actions and processes which can then be enlarged and diffused. What is crucial though is that new social, economic, and ecological processes begin to be put in place – to be rebuilt – which then connects in multifarious ways with the wider layers of the national and international economies.

These are important points of definition, not least because of the regular challenge that highlighting ‘the local’ also tends, by definition, to render it relatively marginal. In this sense the global and the macro market based perspectives, including those based around notions of competitiveness, have for too long been appropriated by those who, by definition, see the local as ‘caged’; as the marginal.

If we hypothesize, as many of the results of inter national comparative European research suggest, that the ‘capacity for local and regional forms of sustainable rural development are dependent upon enlarging the intensity and quality of their interactions with the wider economy’, we can take one further step which then specifies this with regard to some key parameters of embeddedness and interactiveness which are necessary. Hence, we can suggest a set of key composite variables which are re quired and which need to be enlarged in order to create sustainable rural development:

1. Defining rural development as an active structural and behavioral change in the rural economy, leading to a raising of its competitive capabilities in the face of ‘cost price squeezes’ and sustainability threats.
2. Creating new local and regional institutional frame works both involving the supply and demand man agement of rural goods and services.
3. Raising the social and entrepreneurial capability for multiple resource use and synergy. That is the ability to generate different economic values from the same resource through co production, cooperation, and co evolution of the resource base; economies of scope rather than economies of scale.
4. Increasing the quality, skills, and trust base of inter actions between local resources and the wider economy.
5. Increasing the number and density of the interactions between local resources and the wider economy.
6. Increasing the degree of local producer and marketing negotiating power residing in the interactions between local resources and the wider economy.

It is clear that these are still somewhat abstract concepts, which in combination (rather than separately), and in the aggregate rather than in the singular, begin to provide a basis for real shifts toward more sustainable rural development, taking many of the earlier agroeco logical principles as their point of departure. These bring together some of the main components identified in empirical rural and agri food research over the past decade. Of course the parameters need to be grounded in and through particular rural spaces, given the inherent differentiation of these, and given the particularity of local agroecological and social conditions.

Two redefined principles need to underpin these rural development parameters. First is a ‘reconfigured notion of competitiveness’. Competitiveness in this rural development context refers to not only price setting (although this can be both high and low depending on the variety of goods and services demanded by a growing quality based wider economy), but as importantly are the particular ‘terms of trade’ established between key actors and institutions in the supply and demand networks for these goods and services. In a world where markets, in a real sense, have been at least partially eclipsed by supply chain regulation, and where access to these chains becomes increasingly based around wider social and regulatory criteria, it follows that the overall com petitiveness of any local rural economy is increasingly based around the two main infrastructural/competitive ness parameters (i.e., points (2) and (3) above). That is, real competitiveness of the local rural economy becomes increasingly dependent upon the robustness and inclusiveness of new institutional frameworks, on the one hand, and the social and practical capacity to generate multiple value chains and products from the same resource base, on the other.

This latter point represents the second redefined principle: 'the relative capability for the local rural economy to do more than one thing at the same time from the same (and necessarily restricted) resource base'. We should remember that all historical forms of agrarian capitalism have attempted to meet this challenge. Early capitalism did it through the enclosure movement; Fordism did it through creating an increasing range of products from a very restricted agricultural input base; more recently 'post Fordism' in agri food, led by the internationalized retailers and processors, does it through the generation of over 25 000 food products in any one store and systems of logistics and category management which play havoc with our traditional (Fordist) notions of space–time.

The new rural development model, in this sense is no different, however alternative and sustainable be its overall objectives. Based in and as part of a revised capitalist logic, at the very same time as attempting to create and wrest as much autonomy from it, a key condition for rural development needs to be the actual intensification of time–space in such a way as to create a variety of goods and services with the intent to service a wider variety of external and quality constructed markets. The central feature here is to create synergies between the heterogeneity of production with the heterogeneity of marketing and demand; to create economies of scope and diversity, rather than just of scale. This is the very antithesis to the monocultural tendencies of agri industrialism.

Giving shape to concrete and functioning activities, in this rural development domain requires grasping revised notions of the competitiveness in which local rural economies are situated. In addition, it requires an improved understanding of how new horizontal and vertical institutional and entrepreneurial frameworks and platforms can be put into place.

These parameters of new rural development processes have to be underpinned by the three key features of interaction between the local and the wider rural economy. That is the quality, number/density, and the actual power geometries which exist within these interactions. These, of course are in interdependent relationship with the above more infrastructural factors, and they can be fostered and changed over time. In this sense this is a dynamic rather than a static parameterization.

In depth research in the rural regions of Italy, the Netherlands, and the UK exemplify and support the conceptualizations outlined here. In these studies we see the differential processes of rural development occurring with, for instance, new synergies developing between agri food, tourism, local and locality food chain developments. Here the specified use of the term multifunctionality in the context of rural development and the reconfiguring of resource structures is important. In Tuscany, for instance, we have witnessed the importance of new forms of institutional frameworks – such as agricultural consortiums and cooperatives – as important building blocks in enrolling producers and in managing the volatilities in quality supply and demand of new, or reconditioned local and (exportable) quality locality foods.

To a lesser extent, and with weaker institutional development and, weaker degrees of synergy we have witnessed still significant levels of rural development impacts from our studies in South West England and Wales in the UK. Here, however, the still dominant and somewhat inert agri industrial and postproductivist institutional structures and local/external relations can act as an impediment to the flourishing growth of real rural development alternatives. It is also clear, more generally from these researches that there tends to be an inverse relationship between the degree to which the rural development parameters outlined above can really become embedded and scaled up, with the institutional power and regional and local dominance of the agri industrial and postproducivist regulatory systems.

Wales is a case in point here. Defined by the national state since World War II as a cheap producer of essentially three agri food inputs – beef, lamb, and liquid milk – it has struggled to develop an alternative agri food and rural development strategy based upon endogenous development and the principles and parameters of rural development identified above. Finding space for alternative rural development strategies is ironically somewhat easier in those regions and localities which the agri industrial regulatory model tended to historically ignore or sidestep. Many upland and mountainous European regions, for instance, are becoming the centers for new forms of innovative rural development, while much of lowland North West Europe (e.g., Eastern England, Northwest Netherlands) struggles to disentangle itself from the 'lock in' effects of different combinations of agri industrialism and postproductivism.

Hence, the parameters outlined here serve as a way of theorizing and conceptualizing the inherent differentiation of the rural development model, as it confronts the still powerful forces of agri industrialism and post productivism. It has attempted to ask: what are some of the common parameters and principles involved in the process of rural differentiation? And, how might an improved theory and understanding be developed of this inherently differentiating and heterogeneous rural development process?

What these studies illuminate, however, are that practices such as 'retro innovation' – bringing back agricultural and land based practices from the past; 'ecological entrepreneurship' – whereby key actors begin to create and sustain new networks of action; can begin to build new supply chains and action spaces in which more value added can be captured at the local rural level. Moreover, it is clear that fostering 'urban based forms of procurement', not least in foods; and linking these to their rural hinterlands could have a radical impact upon generating rural development benefits both in the North and the South. For instance, the city of New York re quires 800 000 school meals per day. Innovations in bio energy and renewable energy, as well as food procurement are so far in their early stages. Thus the recalibration of urban demands with the reconstitution of the rural resource base is an important area for devel opment, and one which the rural development policy community needs to be attuned to.

The parameters of sustainable rural development outlined in the latter part of the article begin to assemble some of the building blocks which are relevant in har nessing the internal and external potentialities of rural localities and regions in a more mobile social and eco nomic context. These building blocks have their intel lectual roots in the earlier discussions concerning agroecology and sustainable agricultural principles.

What is clear (from our research evidence in Europe, as well as beyond) is that it is possible to rebuild rural development in ways which increase the interactions with the external (urban and consumer based) economy at the same time as maximizing the ways in which more economic and social value can be agroecologically fixed in specific rural spaces; that is, to fit more effectively rural spaces into the more mobile and consumer based worlds in which they are now necessarily situated. However, this will not occur through market mech anisms alone; and it relies upon rural actors and insti tutions being proactive in both distancing and escaping themselves from many of the existing and devaluing aspects of rural life, at the same time as reassembling and redefining resources and infrastructures in ways which carve out new diversified niches of sustainable production of goods and services.

Conclusions: Prospects and Challenges for Sustainable Agricultural and Rural Development

This article started by outlining some of the basic and time honored principles of sustainable agriculture. The proceeding synthesis, concerning the variant ways and vectors in which these principles have been (contestedly) grounded and conceptually developed by scholars, has demonstrated their continual and growing relevance. For much of the latter half of the twentieth century agri cultural sustainability debates concentrated on both the recognition, definition, as well as the oppositional sig nificance of these principles. This led to considerable vitality and energy in creative conceptual development. More recently, the political and economic vector for sustainable agriculture has dramatically widened and deepened into what some scholars have termed the 'real green revolution'.

There may now be a real shift occurring, not least in places like Europe, from an emphasis on oppositional and ontological politics toward really embedding and 'sus taining the sustainable' as opposed to 'sustaining the unsustainable'. That is, focusing upon the rebuilding of new spatial and ecological configurations for sustainable rural and agricultural development *sui generis*. This is becoming more probable and realistic not only because of, as we have seen, the gradual moves toward ecological modernization by the state apparatus (especially in Europe). Agricultural sustainability, or rather agri food sustainability (taking in whole complex supply chain factors rather that just the farm), is now recognized as a central dynamic in the broader political economy and governance of what we might call the macro eco economy.

In the first decades of the twenty first century: (1) the real effects of the growing scarcity of the carbon based resource are becoming clearer, and having both corporate and geopolitical effects; at the same time (2) the real effects of 150 years of the carbon based economy are being realized with regard to climate change. In this changing context the positioning and political power of sustainable agriculture, agroecology, and the rural de velopment paradigm can only grow in ways which should see these principles become a more powerful vehicle for reintegrating both sustainable production and con sumption, and also re equating the fast growing urban world with its diverse rural ecological hinterlands. However, in the face of reignited global concerns about food security, vast disparities in the volumes and types of food consumption, and the persistence of the agri in dustrial model in attempting to find new technological 'fixes', we should not expect this to be an easy or un problematic journey.

See also: Environment; Sustainability.

Further Reading

Altieri, A. (1987). *Agro ecology: The Scientific Basis of Alternative Agriculture*. Boulder, CO: Westview Press.
Gliessman, S. R. (1990). *Agro ecology: Researching the Ecological Basis of Sustainable Agriculture*. New York: Springer.

Guzman, E. S. and Martinez Alier, J. (2006). New social movements and agro ecology. In Cloke, P., Marsden, T. K. & Mooney, P. (eds.) *Handbook of Rural Studies*, ch. 34, pp 472 483. London: Sage.

Guzman, E. S. and Woodgate, G. (1997). Sustainable rural development: From industrial agriculture to agro ecology. In Redlift, M. & Woodgate, G. (eds.) *The International Handbook of Environmental Sociology*, ch. 5, pp 83 101. Cheltenham: Edward Elgar.

Marsden, T. K. (2003). *The Condition of Rural Sustainability, European Rural Development Series*. Assen: Van Gorcum.

Morgan, K., Mardsen, T. and Murdoch, J. (2006). *Worlds of Food: Power, Place and Provenance in the Food Chain*. Oxford: Oxford University Press.

Parrott, N. and Marsden, T. K. (2002). *The Real Green Revolution: Organic and Agro ecological Farming in the South*. London: Greenpeace Environmental Trust.

Van der Ploeg, J. D. (2003). *The Virtual Farmer*. Assen: Royal van Gorcum.

Agri-Environmentalism and Rural Change

C. Potter, Imperial College London, London, UK

Introduction

The realization that modern, intensive agriculture could be environmentally damaging was slow to gain ground among policy elites in industrialized countries. Traditionally portrayed as environmental stewards engaged in the production of essential staples, farmers in most developed countries since the 1930s have received substantial levels of government support in order to achieve food security and public good policy goals. The conventional wisdom is that farmers as a group deserve state assistance, not merely because their incomes tend to be lower and more volatile than those of others in society, but also because, without them, many rural communities and environments would no longer be sustainable. During the 1970s and 1980s, mounting evidence that agricultural intensification and the production of food on an industrial scale was depleting biodiversity, eroding soils, and contaminating the wider environment brought about a profound questioning of the idea that the rural environment can safely be left to the care of farmers.

The result has been a rethinking of policy models and approaches and arguably one of the largest scale experiments in environmental regulation and incentivization ever attempted in places like the European Union (EU) and the United States. With this has come a substantial body of rural geographical research concerned with documenting the nature and extent of the environmental threat, contributing to debates concerning the public policy and regulatory implications and reflecting on the sociocultural–political repercussions of the associated shifts in public attitudes toward farmers and farming. 'Agri environmentalism' is now an established feature of public policy discourse in most developed countries, indicative of a concern to ensure that farming practices are environmentally sustainable in a practical and largely policy driven sense. It also refers, however, to a larger and much more politically charged renegotiation of the relationship between farmers and the state in a period of policy change and market liberalization.

This article falls into three sections. The first section briefly reviews the nature of the environmental impact of modern agriculture and the different ways in which this has been reported, analyzed, and understood by geographers, followed by a section that describes the emergence and evolution of the agri environmental policies that have been put in place to manage the complex and spatially variable environmental problems brought about by agricultural changes. This section gives a flavor of the policy debates that continue to circulate within different policy jurisdictions about how best to reconcile agricultural restructuring with environmental sustainability. The final section places these agri environmental concerns in a broader international context by assessing the challenge to rural environmental sustainability of agricultural market liberalization under the auspices of the recent World Trade Organization (WTO) negotiations on agriculture.

Developing an Environmentalist Critique of Modern Agriculture

Environmentalist critiques of modern farming first emerged in places like Western Europe, the United States, and Canada during the late 1970s. Mounting evidence of the biodiversity, soil erosion, and off farm pollution impacts of agricultural intensification and specialization was brought to public attention by campaigning scientists and journalists and by conservation groups in both the EU and North America. In the classic former enclosed landscapes of lowland England and northwest Denmark and throughout the bocage landscapes of northern France and Belgium, for instance, the removal of hedgerows from the late 1960s onward to facilitate larger scale farming operations had brought about dramatic visual change. Southern member states of the EU such as Spain, Portugal, and Greece underwent a transformation of their farm sectors following accession in 1986. Here, intensification of production on the best land was accompanied by the marginalization and decline of traditional farming practices, and geographers and ecologists reported a steady erosion of biodiversity and a growing threat from groundwater contamination, soil erosion, and fire due to a complex double movement, which saw the expansion of intensive arable production on irrigated land and the loss of long established and environmentally sympathetic dryland arable production systems elsewhere. Soil erosion had long been recognized as a major environmental threat in the US before this date, but evidence published by the National Resources Inventory in the mid 1980s revealed a significant upswing in rates of soil loss since the early 1970s. Evidence marshalled by campaign groups like the Conservation Foundation suggested that off site damage resulting from the sedimentation of watercourses and from pollution far outstripped the on site costs of reduced soil productivity.

By the late 1980s it was becoming increasingly clear that high levels of agricultural subsidy offered to farmers by governments, and directly coupled to production through price support, were a major driver of these and other reported types of agricultural change and environmental damage. In the United States, it was noted that environmental concerns have been little considered when farmers had been encouraged through the United States Department of Agriculture's commodity programs to 'plant fence row to fence row' in pursuit of increased production and overseas markets. The result was an almost oceanic shift in land use that swept away many of the soil conservation practices previously in place and significantly exacerbated the soil erosion and diffuse pollution problem.

Critics of the EU's Common Agricultural Policy (CAP) were by now offering a similar farm policy explanation for many of the environmental problems associated with European agriculture. The high price guarantees available to farmers under the CAP since the early 1960s had encouraged a widespread intensification of production as producers strove to increase output by bringing hitherto unfarmed land into production, by applying more bought inputs like fertilizer and pesticide to every hectare of land in production, and by stepping up stocking rates. This essentially policy centric understanding of agri environmental change prevailed in the public debates of the time, not least because it offered agri environmentalists a policy target for their campaigns. Moreover, by attaching blame to policies rather than farmers this construction made possible a series of new alliances between agrarian and conservationist interests, and hence rendered politically feasible a series of attempts throughout the 1990s to green the farm policies of industrialized countries.

In the event, progress toward a fully sustainable agri environmental policy has been slow, reflecting the culturally embedded nature of agricultural support and the ability of farmers' groups to defend traditional policy entitlements. As a case study in environmental governance and policy experimentation, however, agri environmentalism offers geographers a rich object of study. Agri environmental policy development is interesting as much for what it tells us about the nature of the policy process and the still culturally embedded ways of constructing the countryside and the role of farming within it, as it is a route map for environmental reform. It is to these aspects that we now turn.

Toward Greener Farm Policies?

The idea that farmers deserve special treatment because they are farmers was never likely to be defensible as a long term justification for state assistance. Policymakers have presented public good and social equity justifications for shielding farmers from world market forces for many decades, but it was not until the late 1980s that an environmental public good rationale was articulated with any great persuasiveness. Conventional farm support, by now under challenge on grounds of its spiralling budgetary costs (the CAP was absorbing over 70% of the total EU budget in 1985), its bias in favor of larger producers (20% of the largest European farm operators were receiving 80% of the subsidy), and its trade distorting consequences for trading partners, began to look unsustainable in its current form.

Farm policy critics, geographers among them, sensed that a new policy settlement was possible which would rebalance agricultural support away from production and farmer income support *per se* and toward green payment and rural development schemes of various complexions. The US was among the first to experiment with this new approach, establishing a Conservation Title under the 1984 Farm Bill and creating a 20 million ha Conservation Reserve Program (CRP). This sought to tackle soil erosion and off site damage by paying farmers to take highly erodible lands out of intensive agriculture use. In the EU, farmers were also being offered agri environmental contracts that subsidized the adoption of agricultural stewardship and the management of landscapes and biodiversity. By 2000, one in every five European farmers was in receipt of an agri environmental payment and over 3 million ha had been enrolled in national programs. Rather than reserving land by taking it out of production, here the focus was on 'working lands' and a variety of subsidy schemes were deployed to influence the way farmers managed features like hedgerows, woodland, and semi natural grassland.

Despite important differences in emphasis and approach – Europeans have been more concerned to encourage farmers, many of them economically marginal, to manage established agricultural landscapes and semi natural habitats, whereas North Americans to minimize the soil eroding and polluting effects of intensive production – it has been observed that in both settings a 'provider gets' rather than a 'polluter pays' philosophy tended to prevail. Payments went to farmers as the client group because they possessed the property rights and occupied the land. They were also regarded as deserving recipients and the best people to manage and protect the habitats, landscapes, and resources of the countryside. Agri environmental contracts proved successful in enrolling large numbers of farmers into a new land stewardship community, and this appeared to revive the public view of farmers as providers of first resort of rural environmental goods. Moreover, the agricultural policy community seemed convinced by the tactical argument that paying farmers to produce public goods offered a publicly more defensible basis for agricultural support than income support to a supposedly beleaguered oc cupational group. Under this new public goods model, the role of government is not to support agriculture for its own sake but rather to contract farmers and other land managers directly to supply the environmental goods and services that markets, if left to their own devices, will fail to deliver.

Supporters of the agri environmental turn in farm policy point to the huge expansion of environmental management on farms that the expedient of 'paying farmers to produce countryside' has made possible. The EU, for instance, regards agri environmental programmes as one of the key mechanisms for achieving its bio diversity protection targets, while the US continues to invest significant resources in the CRP and successor programmes such as the Conservation Security Program. Australia, a relative newcomer in this field, appears similarly committed to a version of the stewardship model in its attempts to deal at state and regional level with a legacy of environmental degradation created by overexploitation and intensification (though see further discussion of the conflict with neoliberal priorities below).

Agri environmental policy is now an established fea ture of the international policy scene and the subject of a substantial academic literature devoted to better under standing and comparing how different incentive struc tures work, investigating the motives and behavior of the farmers who chose to be guided by them and assessing the environmental outcomes achieved as a result. Doubts remain, however, about the environmental effectiveness of the measures themselves and the advisability of con tinuing to pursue conservation goals within an agri cultural policy setting. A long standing criticism of agri environmental policy measures is that they have been grafted onto an apparatus of state support that remains fundamentally productivist in the way it operates and is justified. The decoupling of agricultural subsidies from farmers' production decision making that has been the hallmark of the farm policy reforms of the 1990s has eased this contradiction somewhat. The substitution of direct income transfers for open ended price support, for example, is generally agreed to have reduced the in centive for operators to maximize subsidy receipts by increasing output. Measures such as the EU's single farm payment (introduced as part of the Agenda 2000 reforms of the CAP) present other difficulties, with a complex system of 'cross compliance' now in place designed to build environmental safeguards into a measure which is actually about maintaining farmers' incomes rather than their environmental capacities.

So far as agri environmental contracts themselves are concerned, agriculture ministries continue to find it dif ficult to target incentives in ways that maximize en vironmental value for money and too many schemes have been operated as disguised income support measures for marginal farmers. Governments face a dilemma in de ciding how strictly prescribed agri environmental schemes should be – do they impose tightly prescribed entry and delivery conditions on payments and risk a reduced rate of uptake and a disproportionate burden of administrative costs under a 'narrow but deep' strategy or do they opt for a 'broad but shallow' approach which aims for maximum enrolments but seeks only modest changes to farming practice in return? This is one of many de bates concerning the 'fitness for purpose' of agri en vironmental policy tools in a period of intensifying competition for resources and growing public scrutiny of spending decisions.

Contested Models of Agri-Environmental Governance

At the same time, agri environmental concerns have been swept into an increasingly international debate about the broad nature of agricultural sustainability and the role of farmers in delivering this. The restructuring and relocation of production that is likely within an in creasingly globalized agro food system may prove as damaging to habitats and landscapes as earlier periods of agricultural change and there is a continuing debate about how far agri environmentalism can be squared with the dismantling of state support and the establish ment of market rule in agriculture. The convening of a new round of trade negotiations in 2000 under the aus pices of the WTO's Doha trade round has raised the possibility of a much more substantial liberalization of agricultural markets than might previously have been considered possible and led to an international debate about the trade distorting effects of green payment schemes which operate by keeping farmers on the land. Agriculture had been brought into the international framework of multilateral trade negotiations for the first time under the Uruguay Round (1986–93). The resulting agriculture agreement committed signatories to a long term agenda for agricultural trade liberalization that would entail the reform of domestic subsidies, improved market access, and the elimination of export subsidies. The successor Doha Round negotiations have yet to be concluded at the time of writing but it seems unlikely that the neoliberal project for agriculture has yet run its course.

The structural pressures on governments to further liberalize agricultural markets in any event remain im mense. In particular, the globalization of food processing, distribution, and retailing that has come to define a global agro food system means that a new set of corporate interests committed to a neoliberal agenda is exerting growing influence over an international policy agenda centered on the rules and procedures of the WTO. Representatives of these interests have emerged as fierce critics of the traditional model of state assistance to farmers and see the further liberalization of agricultural markets, particularly in places like the EU, as essential if they are to control costs and compete for new markets.

Working separately and through country groupings such as the Cairns Group of agro exporters, they have succeeded in constructing a powerful agenda for reform that has as its informing vision a global agriculture governed by market rule.

All of which are raising questions about the types of agri environmental governance that would be feasible within the framework of such a policy settlement. The important debate about 'agricultural multifunctionality' that has dominated early stages of the Doha round is motivated by a concern that the public good attributes of farming may become harder to defend if the market protection currently extended to large numbers of mar ginal producers in some industrialized economies has to be dismantled. A number of different understandings of the relationship between agriculture, rural society, and the environment are in play here, giving rise to a con tinuing debate about the extent to which the protection and management of rural environments should continue to be a state directed project.

Those who promote a multifunctional view appear to be advocating a 'working lands' model of agricultural governance, which seeks to uphold a joint production view of the relationship between agriculture and the environment. This construction says that not only be cause of the practice of farming, principally through the activity of grazing animals, but also because of the dif ferent land management practices that are put in place in order to retain fertility and yields essentially underpins the landscapes and habitats valued by the public in places like the European Union. Unless governments continue to underwrite agricultural incomes through state support, large numbers of vulnerable farmers will leave the land, extensive tracts of countryside risk being abandoned, and biodiversity values will decline. It is a model that has the support of many EU member states, notably France, which insists that it is agricultural production which defines the very qualities (or multifunctions) of rural space that most European citizens are interested in and willing to pay for as both taxpayers and consumers of food. With its affirmation of the importance of sustaining the territorial association between farming, food, and environment, this is a stance that suggests that the farming population is itself a public good. Despite strong criticism from trading partners, it is an approach that continues to be strongly defended in international and domestic policy forums and is arguably the informing concept behind the recent introduction of the EU's single farm payment scheme, a system of income supports apparently designed to keep farmers in large numbers on the land.

On the other hand are those supporters of a de coupled public goods model of agri environmental gov ernance who believe that the idea that governments need to support farming incomes and structures 'in order' to achieve environmental goals rather than doing so directly has far too many spillover effects into production and, thus, into international markets to be acceptable in the WTO. Their prescription is for payments to be linked as directly as possible to the environmental outputs that farmers (and other land managers) achieve. They tend to be more optimistic about the ability of farmers to adapt to world market conditions and point to the European experience of large numbers of family farmers managing to remain on the land through income diversification and off farm employment. This analysis and the policy pre scriptions to which it gives rise enjoys the support of WTO members such as Australia and, to a lesser extent, the US.

Among European countries, the UK has come closest to advocating the decoupling of policy being implied here. The contention of the UK government in recent years has been that further reform of the CAP should achieve a situation in which the majority of support to farmers is delivered through rural development and agri environmental policy programmes. This is presented within a framework of rural policy which seeks to move away from a focus on farming as the keystone of the rural economy in favor of a much wider range of economic activities. A component of this 'alternative paradigm' is a view of the countryside as an increasingly important consumption space in which the growing demand for quality products and the scope for branding goods, ser vices, and places will create new income streams for farmers and their families and enable governments to withdraw from their traditional role as an income guar antor. Agri environmentalists face a dilemma in deciding how far to give this latter, more decoupled model their full support. The idea that subsidies to farmers should be conceived more or less exclusively in public good terms is one that they have campaigned for since the beginning. However, there remain considerable uncertainties about the restructuring implications of such a radical policy shift and the inevitable net withdrawal of government support that would result.

Australian governments have long struggled to rec oncile neoliberal policy priorities with the need to tackle an agri environmental crisis that many critics believe requires some form of federal state intervention for their solution. The much vaunted National Landcare Pro gram, introduced in 1989, attempted to persuade farmers to undertake environmental management through better information and voluntary action. In the absence of fi nancial support, and already substantially exposed to world markets, farmers favored measures which im proved productivity and business viability such as tree planting for erosion control and shade or establishing permanent pasture. More radical options such as pro tecting biodiversity or revegatating saline groundwater recharge zones have been much less widely adopted. While some state level initiatives such as the Victoria's Rural Land Stewardship Project suggest a move toward the European 'public payment for public goods' model, Dibden and Cocklin argue that Australian farmers con tinue to lack the capital, knowledge, or income to gen erate effective change at the landscape scale. If, as James McCarthy suggests, struggles over multifunctionality are essentially about negotiating the revaluation of rural nature in the context of trade liberalization, then the issue at stake for policymakers in Europe particularly is how far it will be possible to perpetuate working farmed landscapes under world market conditions. Although interest is growing in the scope for 'rewilding' upland landscapes here in the wake of a contraction of farming activity, the idea that there may be other ways of working with nature than through farming has as yet been slow to appear in public debate.

Conclusions

At one level, the invention of agri environmental policy is arguably one of the most significant environmental policy achievements of recent years. Partly a reaction to the sweeping changes in farming practice of the 1960s, 1970s, and 1980s, the greening of agricultural policy re flects a profound public reassessment of the role of farmers as providers of public environmental goods. Rather than the use of regulatory instruments, however, the approach followed has been to subsidize farmers so that they can better fulfill their stewardship obligations. In the US, programs like the CRP have brought about land use change on an unprecedented scale and a very significant degree of new investment in conservation capital; in the EU, a large scale experiment is effectively in progress involving the use of environmental contracts in order to achieve landscape, biodiversity, and pollution control policy objectives. Critics point to the essentially second best nature of these reforms, however, and the mixed results that have been achieved in terms of im proved environmental outcomes and changes to the be havior and outlook of farmers. Having contributed both as advocates of domestic policy reform and as analysts of policy processes and implementation strategies, geog raphers are now increasingly interested in the way agri environmental concerns are being drawn into an inter national policy debate centered on the liberalization of agricultural markets and the continuing globalization of agricultural production. Countries like Australia and New Zealand, where farmers are already significantly more exposed to world market forces than their more protected European competitors, are meanwhile closely studied for what they can tell us about the perils and opportunities of agri environmental governance under more open market conditions. Further comparative work is increasingly valued for what it can tell us about the changing role of farmers and 'the nature of nature' in what seems likely to be a period of unprecedented agricultural change.

See also: Agrarian Transformations; Environmental Security; Rural Geography; Rurality and Post-Rurality.

Further Reading

Cocklin, C., Dibden, J. and Mautner, N. (2006). From market to multifunctionality: Land stewardship in Australia. *The Geographical Journal* 172, 197 205.

Dobbs, T. and Pretty, J. (2004). Agri environmental stewardship schemes and multifunctionality. *Review of Agricultural Economics* 26, 221 237.

Fish, R., Seymour, S. and Watkins, C. (2007). *Sustainable Farmland Management: New Transdisciplinary Approaches*. Wallingford, CT: CABI.

Hodge, I. (2000). Agri environmental relationships and the choice of policy mechanisms. *The World Economy* 23, 257 273.

Lobley, M. and Potter, C. (1998). Environmental stewardship in UK agriculture: A comparison of the environmentally sensitive areas programme and countryside stewardship. *Geoforum* 29, 41 52.

McCarthy, J. (2005). Multifunctional geographies: Reactionary or radical? *Progress in Human Geography* 29, 773 782.

Potter, C. and Tilzey, M. (2005). Agricultural policy discourses in the European post Fordist transition: Neoliberalism, neomercantalism and multifunctionality. *Progress in Human Geography* 29, 1 20.

Aid

V. Desai, Royal Holloway University of London, Egham, UK

Glossary

Civil Society The state's engagement with Organizations such as NGOs, trade unions, religious leaders.

Complex Political Emergencies Refers to intra-country conflict involving economic, political, and social destabilization, often combined with a 'natural' trigger mechanism such as a drought, flood, or earthquake.

Internally Displaced Persons (IDPs) People who have been forced from their homes as a result of armed conflict or violence, environmental degradation, large-scale development schemes, or economic crisis. Unlike refugees, whose movement across national borders provides them with special status in international law with rights specific to their situation, IDPs have no such entitlement.

Different Types of Aid

The main aim of aid (also known as 'official development assistance' – ODA) is to provide resources on concessional terms for the promotion of economic development and welfare in developing countries. Most often the transfer of these resources is in the form of grants and loans on concessional terms from the developed (mostly northern countries) to developing countries (mainly in the southern hemisphere). The US' Marshall Plan (1948–51) aid to economic reconstruction in Western Europe post World War II set a successful precedent of promoting development the objective of the multilateral organizations (such as the World Bank, International Monetary Fund (IMF) for, African Development Bank (ADB), the Asian Development Bank (ADB), and the Inter American Development Bank (IDB)) has been the transfer of knowledge as the solution to poverty and lack of development in developing countries. Development was seen as a technical issue until the mid 1990s. There are mainly two forms of aid: development aid and humanitarian aid (mainly dispersed to provide relief after disasters and conflicts). Substantial amount of military aid, which is very limited nowadays, was mainly given by the Soviet Union and the United States during the Cold War. Although there are different forms of aid, aid intended for one purpose can lead to another.

The Development Assistance Committee (DAC) was established to improve and coordinate international aid efforts in the 1960s. Over 95% of ODA now comes from DAC members comprising 22 countries plus the European Commission. Multilateral institutions occupy a dominant role in the global political economy and in the developing world as they provide both much needed loans and technical assistance (in projects on power, transport, and telecommunications). About 30% of development aid is managed by the multilateral donors, such as the World Bank, which also make interest free loans to the least developed countries. Many official donors also provide substantial additional amounts of humanitarian aid which is administered through nongovernmental organizations (NGOs). Besides these official donors there are (NGOs) such as Oxfam who provide assistance in the region of $6 billion annually. Northern based NGOs have branches based in southern countries in which they work. Often they run projects in partnership with NGOs in the south, effectively bypassing those of the state. NGOs have become increasingly globalized, and are working through transnational networks, forging new enduring alliances and networks to maximize their impact and engage in global policy processes. Along with bilateral aid, multilateral aid through agencies such as the World Bank, UN, EU, and NGOs has gained a higher profile in formulation and implementation of development policy, becoming key actors in the political economy of development. Over the period 2001–05, sub Saharan Africa was a recipient of 31% of all ODA, which has steadily increased over the years (see **Tables 1** and **2** for detail). On the other hand, aid to China and India has steadily declined in importance. Smaller countries are treated more favorably on a per capita basis and are among the most aid dependent, for example, Bangladesh and Nicaragua.

Reasons for Giving Aid

The main objective of giving aid, since decolonization, is to eradicate poverty from developing countries. Although over the last 50 years other objectives have also emerged which influence donors. For example, around two thirds of Britain's aid has gone to Commonwealth countries to maintain close historical relationships. A has also been given to secure economic and political interests such as cheap labor, expansion of markets, and natural resources. During the Cold War, aid was motivated by political ideology with capitalist governments providing aid to stem the spread of communism.

Table 1 ODA net disbursement received in $billion by recipient country, 2001 05

Iraq	28.78
Congo Dem.Rep. (Zaire)	10.48
Afghanistan	8.24
Vietnam	8.24
Pakistan	8.22
Nigeria	7.79
China	7.76
Ethiopia	7.75
Tanzania	7.47
Indonesia	7.14
Serbia	6.85
Mozambique	6.71
India	6.46
Bangladesh	6.08
Egypt	5.93
Palestinian adm.areas	5.69

Source: OECD. http://www.oecd.org/statistics.

Table 2 ODA net disbursement received in $billion by recipient region, 2001 05

All developing countries	368.17
Africa	129.45
Africa North of Sahara	12.48
Africa South of Sahara	114.50
America	29.98
North and Central America	14.46
South America	13.43
Asia	123.96
Far East Asia	33.41
South and Central Asia	44.94
Middle East	43.57
Europe	19.51
Oceania	4.38

Source: OECD. http://www.oecd.org/statistics.

Over the years, the main motivation of US aid has been national security and superpower rivalry; this led to geopolitical allocations, while Germany and Japan have had very economic objectives. The sup port for aid is declining within the US. Post 9/11, re flecting on the origins of international terrorism has provided new objectives to giving aid. The post Cold War disorders such as violent conflicts have seen the reemergence of new kinds of security threats, while countries like Canada, the Netherlands, and Scandinavia attach great importance to the objective of international humanity.

Changing Trends in Development Aid

In the early decades, development aid was given on an economic basis mainly to enhance productive capacity and bring about modernization of developing countries (influenced by modernization theories). Many develop ing countries in the initial phase of development, post decolonization, wanted to industrialize their econ omies. This led to imports of technologies (such as hybrid varieties of crops, fertilizers, imports of machinery, and fuel) or building large scale infrastructural projects such as highways and dams, which would support develop ment in other sectors. Many developing countries then borrowed on lower interest for these imports.

Over the last decade, grants have become the dom inant form of bilateral aid to the poorest countries. Many of these countries over time are facing difficulties in re paying their loans which were previously borrowed on concessional terms. Iraq and Nigeria received debt for giveness in 2005 accounting for $14 billion and $5 billion. This is of course an exception, but development aid is set to rise as long as the donors keep their promise, which, some critics think, is unlikely.

There has been increased collaboration both with governments and aid agencies based on a growing belief over the period that the promotion of NGOs could offer an alternative model of development and play a key role in processes of democratization. NGOs were seen as more administratively flexible, closer to the poor, in novative in problem solving and more cost effective than corresponding state partners. Donor pressure toward structural reform and privatization underlies the in creased interest in NGOs as 'service deliverers' – part of a wider and explicit objective to facilitate productive NGO–state partnerships. In such an environment, the important role for NGOs in the last two decades has been in mitigating the adverse costs of structural ad justment (liberalization, privatization, export promotion, and reshaping the welfare state) and promoting donor reform packages in offering insurance against a political backlash towards harsh adjustment regimes. NGO in volvement in implementing World Bank financed pro jects grew rapidly during the 1980s and 1990s as many governments' service delivery capacities were shrinking. The World Bank report in 2007 highlighted that 70% of World Bank financed projects in 2006 had NGO in volvement. There is a realization among donor countries that aid is becoming ever more complex, with new in struments and players, with increasing criticisms over its effectiveness.

The expansion in partnership between northern and southern NGOs originated in changing attitudes in the north in the 1960s and 1970s. A view took hold that merely transferring resources in the form of tools or funds was not an adequate response to poverty when that was rooted in structural problems. Indeed, such transfers could just preserve the situation by creating financial dependency. The establishment of research departments and policy units in northern NGOs marked this change in approach. Northern NGOs respond to emergencies, short term re lief, and long term rehabilitation, such as victims of war and of natural or man made disasters. They raise money in the north from the general public, private sector, and governments to pay for their work, and to share as much as possible with their southern counterparts in order to help in building the capacity of southern NGOs and educate their own constituencies in the north about the underlying causes of poverty, and draw people into active lobbying and campaigning for change.

NGOs have pressed the World Bank to become more generous, egalitarian, and responsive to gender and mi nority concerns; more transparent and open to effective civil society participation; and more sensitive to natural resource use and human rights standards. Resistance and confrontational strategies to World Bank financed infra structure projects have been the most visible form of NGO criticism, and have forced the Bank to give at tention to environmental and rights policies, social issues involving major dams (e.g., Nepal's Arun River), energy projects (e.g., the Chad Cameroon pipeline), forestry, water, and major infrastructural projects. NGOs have won a seat at many World Bank policy discussions. NGOs will now have to build the local capacity to monitor and force implementation of the policy com mitments won at Bank headquarters and increase the voice and control to NGOs based in borrowing countries.

Aid Conditionality

There was an increase in condition based lending in the 1980s based primarily on institutional reform and eco nomic policies with strong neoliberal principles of mar ket led development in developing countries. These economic conditions came to be known as the Wash ington Consensus. Recipient countries had to fulfill re quirements such as structural adjustment programs (SAPS) which became a major feature to persuade countries which were defaulters of debt to their house in order. The multilateral agencies embodied the view that the debt crisis of any given country was caused by do mestic, economic mismanagement. The basic com ponents involved in the program were currency devaluation, tax and public spending cuts, and incentives to support export led growth. The content of con ditionality evolved over time. The practice of adding conditions to loans given became central to the oper ations of the IMF, the World Bank and was later followed by other development agencies and governments of major industrial nations.

These economic conditions led to two developments. The donors began to believe that to make aid effective, sound economic management needs to be in place sup ported by a conducive, political environment. This also led to further political conditionalities in the 1990s focusing on good governance (bringing the state back in as a key agent of development). The principles of good governance were mainly to promote human rights, economic growth (market led economy), democracy, accountability of public finance, and to reduce cor ruption. Development is seen not only in the economic context of global capitalism but also in the political context. The role of the state is to provide an enabling environment for development by other agencies. The World Bank and the IMF, and also many bilateral ODA programs, require recipient countries to adhere to certain policies. Aid allocation from then onwards went to se lected countries which had demonstrated a commitment to sound economic development policy and good gov ernance and those who were credit worthy. This mar ginalized those countries which desperately needed help and support. In the later half of the 1990s, a further condition was added requiring countries to adopt pro poor policies most notably requiring poverty reduction strategy papers (PRSPs), to which aid and debt relief were increasingly tied. These were based on partici patory processes of consulting various local stakeholders, especially civil society, generating a sense of citizen ownership and acceptance of state initiated development, among the aid recipients. Forty nine countries had ap proved PRSPs as of 2007. An increasing amount of aid is given to restore democracy, especially by United States Agency for International Development (USAID), United Nations Development Program, and the EU estimated at reaching US$5 billion. It is very difficult to assess if democratization specifically benefits the performance of development aid. The good governance agenda requires civil society to have a sufficiently independent position with respect to the state. To attain this condition, donors will need to channel finance outside of government structures.

Neither the IMF nor the World Bank has been able to show connection between their own programs and im provements in economic policies. There is still a great deal of political influence on the process of aid allocation. There are programs that have failed or have taken longer to implement than originally planned. Some programs have been poorly implemented in the recipient countries. Donors have ignored failures in some countries and have pretended that conditions have been met when the reality is otherwise. The 1995 World Bank report high lights that generally adjustment lending has promoted good policies, but achieved weak program results. In most cases, the implementation of programs has failed due to internal political pressures within the country rather than the donor pressures. Donors also do not have much in fluence in bringing about political change.

Aid conditions were heavily criticized, for under mining the sovereignty of aid recipient countries and were regarded as interfering in economic decisions and policy making. Some donors believe that it gives them legitimate interest in the policies of recipient countries.

Conditions were difficult to enforce too. Some critics also regarded them as coercive, because of the intrinsic im balance in negotiating strength between rich donor agencies and poor recipient governments. Government performances change over time, as the present is not necessarily a good predictor of the future. Recipient governments have little incentive to accept unwanted policies if they can borrow from elsewhere or their economy is strong enough to manage without the loan or grant, while countries, which have not had adequate economic growth or have experienced instability, rely heavily on borrowing through grants and loans on lower interest rates than the markets. Also, noncompliance of conditions has been rarely punished; rather, governments have been able to negotiate new credit similar to those who have complied. There is usually a lot of pressure on donors to continue aid as there are always fears of damaging macroeconomic conditions within the country and affecting people at the grassroots. Most World Bank officials would accept that both economic and political conditions have not really worked; some bilateral donors have moved away from these conditions and want to maintain flexible conditions. Most donors are in search of alternative strategies, in selecting recipients as partners for aid allocation through mutually agreed programs and outcomes. It is difficult to imagine an equal partnership given the nature of intrinsic imbalance of power that exists between the donors and the recipient countries.

Humanitarian Aid

Since the early 1990s, there has been both an increase in the number of disasters and a change in the nature of emergencies leading to a substantial increase in hu manitarian assistance. Emergencies have changed in na ture from mainly being natural disasters such as floods, famines, and droughts to complex emergencies such as the 'war on terror' post 11 September 2001, and the interventions in Afghanistan and Iraq have created new challenges for the implementation of humanitarian as sistance. Due to climatic changes, natural disasters have grown in number, for example, floods, hurricanes, typhoons, and droughts. Since the end of the Cold War, complex emergencies especially violent conflicts have increased, which are characterized by the collapse of the rule of law, destruction of livelihoods, and displacement of people. All these complex emergencies are brought into the public spotlight by the international media, which has led to significant awareness among people. The World Disasters Report in 2001 estimated that in the 10 years (1991–2000), 2.3 million people lost their lives in conflict situations. The increasing frequency and chan ging face of emergencies has caused humanitarian ex penditure to rise to an all time high of US$7.8 billion in 2003. This figure does not record the number of charit able donations from individuals or groups or of non western assistance of various kinds, including those from religious organizations.

Humanitarian aid is the assistance given to people in distress by individuals, organizations, or governments with the core purpose of preventing and alleviating human suffering. Humanitarian interventions are basic ally guided by needs, but it follows three main principles of humanitarian law as expressed in the Geneva Con vention of 1949 emphasizing impartiality, neutrality, and independence. Impartiality means no discrimination on the basis of nationality, race, religious beliefs, class, gender, or political opinions. Neutrality demands that humanitarian agencies do not take sides in either hos tilities or ideological controversy. Independence requires that humanitarian agencies retain their autonomy of action.

The Coordination of Humanitarian Assistance (OCHA), on behalf of the United Nations, provides leadership and coordinates the prompt and smooth de livery of relief assistance. It is supported by the World Food Programme (WFP) responsible for emergency food delivery; the United Nations High Commission for Refugees (UNHCR) responsible for shelter; the United Nations Children Fund (UNICEF) responsible for nu trition and water and sanitation; and the Food and Agricultural Organization (FAO) responsible for food and emergency agriculture, while the Red Cross and international NGOs like Merlin provide medical interventions.

Humanitarian aid is mainly a short term intervention to provide relief. The transition from relief to rehabili tation and development is difficult and the long term aspects of development need to be assessed with other development agencies from the initial phase of providing relief. In some countries – Angola, Sierra Leone, Somalia, to name a few – humanitarian aid was provided over a long period, but the system is not equipped for long term intervention leading to rehabilitation and development. Donors at times have used humanitarian aid to avoid engagement with undemocratic states and their ap proaches were often driven by political interest rather than according to the need.

Humanitarian operations are now accompanied by military interventions (peace keeping operations) espe cially in the last two decades in the context of violent conflicts. These have been termed as complex political emergencies (e.g., East Timor, Kosovo, and Burundi). The World Bank has thus placed considerable emphasis on responding to conflict, particularly where it notes that one fifth of Africans now live in countries severely disrupted by conflict. These have accompanied pro liferation of small arms and landmines and particularly sexual violence and exploitation of women and children leading to high risk of HIV/AIDS. In 2003, three million people died of HIV/AIDS and it is turning into a large scale disaster. The UNHCR, in 2006, recorded that there were 8.4 million refugees worldwide (31% in Africa and 36% in Asia) and 23.7 million internally displaced persons (IDPs), who were not only displaced from their home but also their livelihoods. According to the UNHCR, in 2006, there were also a further 688 000 asylum seekers mostly in Europe and North America – who had claimed refugee status but whose claims were not yet recognized. In Sudan, it is estimated that more than 400 000 people have been killed and 2.5 million have been displaced in the genocide in Darfur. In 2005, world military expenditure totalled to $1118 billion, of which 47% was accounted for by the US. The response to terrorist attacks in the west, known as the 'war on terror', has also led to increase of expenditure. All these events pose new challenges to the neutrality of humanitarian action and the aid workers themselves. The rise of terrorism and the range of counterterrorism initiatives by governments have also generated new protection issues especially in relation to human right laws. Can humanitarian assistance be independent of foreign aid policy in the contemporary world. Many questions are raised with regard to the accountability and performance of aid agencies and United Nations OCHA, as the lead global humanitarian agencies are heavily criticized for poor performance. On the other hand, if we did not have an effective coordinating agency like the OCHA, we would not be able to provide relief and assistance at all. Since 1996, the humanitarian sector has undertaken many initiatives to improve accountability and performance such as the Good Humanitarian Donorship initiative to ensure that donors' responses are equitable, effective, and consistent with the humanitarian principles as explained above. The humanitarian agencies have developed the Active Learning Network and the Humanitarian Accountability Partnership.

Critical Contemporary Aid Debates

Many critics have doubted the effectiveness of aid especially in the context of reducing poverty. Throughout the 1970s, the dependency theorists viewed aid as a source of exploitation and dominance by the northern countries. In the 1980s, ODA was challenged by the rise of the neoliberal agenda in the west as aid was now held to contribute to excessive government and harm economic markets.

Currently, aid is viewed as a means of promoting donors' perceptions of 'good governance' and 'sound' economic practices, leading many analysts and politicians to become very critical of aid. Research has indicated that economic growth has taken place post structural adjustment and reforms. Good governance is defined as sound management of a country's economic and social resources for development. What is 'sound' for the World Bank and others holding the view that 'democratization stimulates development' is a range of management techniques that are believed to work well within a standardized liberal democratic model. Critics contend that there are, and ought to be, different paths for development; they are not opposed to 'good governance' but urge that this is compatible with alternatives to liberal democracy in poor countries with different institutional contexts. Despite various campaigns led by various advocacy groups such as Jubilee 2000 and Make Poverty History urging for cancellations, progress on this front has been very slow. The campaign for debt cancellation has been based on the grounds that the original sums borrowed have been paid many times over, because of the interest payments attached to the principal sum. According to a report prepared by Oxfam, in 1998, African countries have spent four times more in debt payments to northern countries than on health and education. Not all countries in sub Saharan Africa have high debt–GNP ratios, for example, Malawi, Chad, and Burkina Faso have generally been under 50% since the early 1990s. The contemporary language among donors is of building partnerships and building capacity within developing countries and with institutions rather than resting the burden of responsibilities on the recipient countries. Various initiatives such as the Heavily Indebted Poor Countries (HIPC) initiative were announced at the G7 summit in 1997. It is believed that governance is the key to development. This is prominently highlighted in 2005 in the Commission for Africa report, *Our Common Interest* launched before the G8 summit at Gleneagles, Scotland, which pledged to double aid by 2010, including an extra $25 billion for Africa and debt write offs for 18 of the world's poorest countries. The EU pledged to increase aid to an average of 0.5% of GDP of member states; this challenges governments to absorb the cost of aid allocation (see **Table 3** for current figures). Critics believe that this ODA is preventive investment in and beyond failed or fragile developing countries and aid is becoming part and parcel of a comprehensive security strategy for a donor's own country.

Institutional reform in developing countries, increasing the role of the private sector in the 1990s in public service delivery, has generated significant opposition and protest. It emphasizes the renewed interest of the markets in the development process, in the context of the state's failure and a possible to responds to the vast needs of the low income citizens of the developing countries. The search continues for combining the assumed efficiency of the private sector and the democratic accountability of the public bodies. In combining market economics and liberal democratic politics, NGOs are simultaneously

Table 3 ODA as percent of Gross National Income by donor DAC countries, 2005

Australia	0.25	Japan	0.28
Austria	0.52	Luxembourg	0.86
Belgium	0.53	Netherlands	0.82
Canada	0.34	New Zealand	0.27
Denmark	0.81	Norway	0.94
Finland	0.46	Portugal	0.21
France	0.47	Spain	0.27
Germany	0.36	Sweden	0.94
Greece	0.17	Switzerland	0.44
Ireland	0.42	United Kingdom	0.47
Italy	0.29	United States	0.22
Total			0.33

Source: OECD. http://www.oecd.org/statistics.

viewed as market based actors and central components of civil society. NGOs fill gaps left by the privatization of state services as part of a structural adjustment or donor promoted reform package. Governments need NGOs to help ensure their programs are effective, well targeted, and accessible to the poorest of the poor. It is now widely accepted that domestic ownership of policy reforms is crucial especially through local stakeholders and domestic leaders for effective implementation and that donor pressures lead to nationalistic resentment of donor interference and undermines the legitimacy of externally promoted policy reforms.

Nongovernmental development assistance organizations have delivered assistance, primarily at the local scale, to disadvantaged communities in developing countries. Their main aim is to contribute to improved quality of life for people in those communities. However, rather than continuing to concentrate on the 'development project' in order to provide services and alternative livelihood strategies, increasingly since 1990s, northern NGOs have devoted their energies and commitment to advocacy campaigns engaging. They have involved with global institutions and affecting on their policies to highlight wider structural factors which perpetuate poverty, for example, entrenched unfair terms of trade, poor commodity prices, or oppressive debt obligations. NGOs are lobbying for further cancellation of debt (Jubilee Debt Campaign) or eliminate poverty (Make Poverty History). NGOs are forging alliances and increasing their engagement in global networks and decision making processes.

While some see multilateral institutions such as the World Bank playing an important role in eradicating poverty, others see them as part of the problem. Multilateral institutions have to satisfy different constituencies such as the borrowing countries, donor countries, and increasingly NGOs and civil society. Hence, multilateral institutions are unable to articulate competing views. Gaining consensus is not easy and becomes an objective in itself while the kind of consensus established is based on the power relationships that prevail in the institutions themselves. Development is no longer just a matter of finding the right solution to the problem of persistent poverty in developing countries. In the contemporary world today, the challenge is to construct some sort of consensus around an increasingly politicized agenda constituted around a whole range of new crosscutting themes such as governance and indigenous people. It is no longer in any credible way possible to define development solely in a technical manner. An increasingly political agenda will make the process of political maneuvering between donor and recipient countries and other stakeholders (civil societies and the private sectors) increasingly difficult for multilateral institutions. They have been forced to face the criticism from a wide range of civil society stakeholders and have to respond to the emerging changes and challenges.

The achievement of development aid is difficult to assess. Many recent studies have tried to assess the effectiveness of aid. The purpose of giving ODA and its deployment is more open to scrutiny. One response has been increased funding for NGOs, generally thought more able to reach the local grassroots level.

Conclusion

There have been significant changes in the way aid is given, in the last 50 years. Our expectations of giving aid and what we understand about it has changed too over this period.

Aid in real terms and expressed as a percentage of donor wealth is declining, in contrast to increasing flows of private international finance. Issues such as global climatic changes, environmental threats, asylum seekers, internationally organized crimes drugs and people trafficking, international migration, and human suffering all have drawn attention to aid allocation policies. It is clear from the present instability, insecurities, and wars, that world leaders have to make a conscious effort in maintaining the multicivilization characteristics of global politics. Early warning systems for complex emergencies have to be adhered to and actions taken immediately. It remains to be seen if these wide ranging objectives can be achieved. Considerable efforts are being made to improve the effectiveness of aid but the outcomes are as yet unclear.

All aid agencies see poverty alleviation as the main priority. The problems it hopes to resolve are very complex which require multidisciplinary approaches. Perhaps we need to lower our expectations on what really is achievable through development aid.

See also: Civil Society; Debt; Development I; Development II; Governance, Good; HIV/AIDS in Developing Countries; Livelihoods; Modernization

Theory; Neoliberalism and Development; Nongovernmental Organizations; Participation; Poverty; Structural Adjustment.

Further Reading

Bebbington, A. (2005). Donor NGO relations and representations of livelihood in nongovernmental aid chains. *World Development* 33(6), 937 950.

Bøås, M. and McNeill, D. (2003). *Multilateral Institutions: A Critical Introduction*. London: Pluto Press.

Browne, S. (2006). *Aid and Influence. Do Donors Help or Hinder?* London: Earthscan.

Burnell, P. and Morrissey, O. (eds.) (2004). *Foreign Aid in the New Global Economy*. Cheltenham: Edward Elgar.

Cliffe, L. and Luckham, R. (1999). Complex political emergencies and the state: Failure and the fate of the state. *Third World Quarterly* 20(1), 27 50. (Complex Political Emergencies, Special Issue of *Third World Quarterly* 20(1)).

Collier, P., Guillaumont, P., Guillaumont, S. and Gunning, J. W. (1997). Redesigning conditionality. *World Development* 25(9), September.

Commission for Africa (2005). Our common interest. *Report of the Commission for Africa*. London: Penguin.

Crawford, G. (1997). Foreign aid and political conditionality: Issues of effectiveness and consistency. *Democratization* 4(3), 69 108.

DeMars, E. D. (2005). *NGOs and Transnational Networks: Wild Cards in World Politics*. London: Pluto Press.

Dollar, D. and Burnside, C. (1998). *Assessing Aid: What Works, What Doesn't Work and Why*. Washington, DC: World Bank.

Easterly, W. (2006). *The White Man's Burden*. New York: The Penguin Press.

Goldman, M. (2005). *Imperial Nature: The World Bank and Struggles for Social Justice in the Age of Globalization*. New Haven: Yale University Press.

Hopkins, R., Powell, A., Roy, A. and Gilbert, C. L. (1997). The World Bank and conditionality. *Journal of International Development* 9(4), June.

Killick, T. (1997). Principals, agents and the failings of conditionality. *Journal of International Development* 9(4), June.

Killick, T., Gunatilaka, R. and Marr, A. (1998). *Aid and the Political Economy of Policy Change*. London: Routledge and Overseas Development Institute.

Koeberle, S., Bedoya, H., Silarszky, P. and Verheyen, G. (eds.) (2005). *Conditionality Revisited: Concepts, Experiences and Lessons*. Washington: World Bank.

Overseas Development Institute (2003). *Humanitarian NGOs: Challenges and Trends*. Humanitarian Policy Group, Briefing Paper, no.12, July 2003.

Rugendyke, B. (ed.) (2007). *NGOs as Advocates for Development in a Globalising World*. Abingdon and New York: Routledge.

Tvedt, T. (1998). *Angels of Mercy or Development Diplomats? NGOs and Foreign Aid*. London: James Currey.

Van Rooy, A. (ed.) (1998). *Civil Society and the Aid Industry: The Politics and Promise*. London and the North South Institute, Ottawa: Earthscan.

Weiss, G. and Collins, C. (2000). *Humanitarian Challenges and Intervention*. Oxford: Westview Press.

Woods, N. (2006). *The Globalizers: The IMF, the World Bank and Their Borrowers*. Ithaca: Cornell University Press.

Relevant Websites

http://www.dfid.gov.uk
British Government's Aid Agency, the Department for International Development.

http://www.jubileedebtcampaign.org.uk
Jubilee Debt Campaign.

http://www.makepovertyhistory.org
Make Poverty History.

http://www.oecd.dac
Site of the Development Co operation Directorate (DAC) of the Organization for Economic Co operation and Development (OECD).

http://www.usaid.gov
United States Agency for International Development.

http://www.worldbank.org
World Bank, including the International Development Association.

Aménagement du Territoire: Territorial Development

P. Merlin, Université de Paris I, Magistère d'Aménagement et Master d'Aménagement et Urbaniste, Paris, France

It can be considered neither as a science nor as a technique, or as an art, but as a voluntary intervention or as the territorial aspect of planning. Therefore, it is a *praxis* (a set of actions aiming to the same objective) and a practice. It is fed less with theoretical rules or *savoir faire* than with empiricism. Therefore, it needs cooperation of specialists of various scientific fields.

Territorial development has to face several conflicts between objectives: economic development versus territorial equality, efficiency versus equity, liberalism versus planning, development versus environment, and centralization versus decentralization.

The first experiments took place in Netherlands (since centuries), and in Soviet Union, Britain, and Italy (between the two world wars). After World War II, the most active countries for territorial development have been Soviet Union, Netherlands, Britain, and France. Few developing countries (e.g., Lebanon or Morocco) have tried to define a territorial development policy, but priority to economic development makes its implementation difficult.

It is not clear what can be the English equivalent of the French *Aménagement du territoire* or of the German *Raumordnung*. Many expressions are used: regional development, territorial development, and town and country planning. The last expression has been chosen for the ministry created in Britain after World War II. Regional development may be considered as limited to the regional scale, while territorial development concerns various spatial scales.

What Is *Amenagement du Territoire* or Territorial Development?

We suggest the following definition of *Aménagement du territoire* (territorial development): action and practice (rather than science, technique, or art) of disposing with order, through the space of a country and in a prospective vision, people and its activities, amenities, and means of communication they can use, taking in consideration natural, human, and economic (and even strategic) constraints in order that functions and relations between the people can be exerted in the most easy, economic, and harmonious way.

Each term of this definition has its utility:

- action and practice: territorial development implies an operational intervention and can be considered neither as a science nor as a technique, or as an art;
- dispose with order through space (…): this expression, apparently insignificant, emphasizes the voluntary character of the intervention, which seeks to institute an ordered situation, for that preferred to *laissez faire* attitude;
- the space of a country: one mentions currently other scales than the national one, mostly local ones, but the actions at the national scale contributing to the organization of the country constitute the specific field of territorial development;
- people and its activities, amenities and means of communication they can use: territorial development is global, even if one may consider specific aspects (e.g., rural development, development of tourism, road development, etc.) which participate in territorial development;
- natural, human, and economic (and even strategic) constraints: here are the three dimensions of sustainable development (economic development that cares for the environment and for the preservation of resources for future generations); and
- the most easy, economic, and harmonious way: this second trilogy underlines that neither attractiveness nor economic development, neither esthetic search nor excellence of interindividual relations can separately be sufficient and that all these goals must be pursued simultaneously; therefore, territorial development is in essence interdisciplinary and cannot be claimed by any particular discipline (geography, economy, sociology, architecture, etc.).

Even if geography, no more than any other discipline, can claim a priority in territorial development, space is at the heart of such an approach. It is the same for time, despite the fact that this term is only present, in our definition, in an implicit way and this through a third trilogy:

- the past, the historic time: the space to organize is not virgin and the planner must take in account what has been 'disposed' by previous generations. He must be imbued with historic knowledge and culture;
- the future: the living environment that will result from the actions of territorial development is marked for a period of several decades (e.g., for leisure amenities), generations (housing and activities), or even centuries (transport infrastructure). Territorial development is therefore inseparable from prospective reflection (necessity reinforced by the objective of sustainable development), which cannot be joined with simple projections or 'futurologist' imaginings; and
- the present: the intervention on space will be only a wavering, caricature of willfulness, if it is not inserted

in the state of the society, which does not exclude a debate to let it evolve. This supposes fine preliminary analysis to know and understand the values and the attempts of the society, which will lead to a precise diagnosis, to hypothetic solutions, and to a reasoned choice between them.

Which Scientific Status for Territorial Development?

Aménagement du territoire (territorial development) cannot be considered as a science, which is a coherent ensemble of knowledge based on specific concepts and on theories and developing rules proved either by a rigorous dem onstration or by experimentation.

It can no more be qualified of being a technique, that is a *savoir faire*, a set of processes, of means, and of methods applied to an activity. Indeed, there is no controlled know how or tested processes. Methods used are diverse and changing and do not receive generalized recognition.

It is no more an art, which is "an ensemble of processes, of knowledge and of rules concerned with the practice of an activity or an action and aims at searching an ideal of beauty." Processes and knowledge are imperfect, rules are fluctuating, and the ideal of beauty rarely reached.

It seems more careful to bring back territorial devel opment to its character of voluntary intervention, of territorial aspect of planning. Therefore, it is a 'praxis' (a set of actions aiming to the same objective). It is fed less with theoretical rules or *savoir faire* than with empiricism. This praxis is exercise of application, execution, way of doing, and confrontation with complex and multiple realities, and hesitation.

Interventions of territorial development need many different specialists: the researcher, interdisciplinary, alone, or within a team who collects the information, analyzes the data, prepares a diagnosis, proposes a choice of alter native solutions for the future; the elected member in charge of decision by choosing between these possible solutions; the administrative official who has to implement this solution; the lawyer who translates it under juridical terms; the engineer who builds the infrastructures; and the architect who designs the buildings. Territorial develop ment can be efficient only through continuity in action and consensus between these different actors.

First Experiments

Netherlands: Struggle against the Sea

Due to the exiguity of their territory (now 34 000 km^2 for 16 million inhabitants) and to the permanent menace of maritime invasions (a large part of the territory is below sea level), Netherlands had a practice of territorial development of polders since the twelfth century, before the term existed. This country, which has been the first to make compulsory the establishment of plans for towns (act of 1901), has initiated a policy of planning of its major metropolis; their municipalities acquiring an actual monopoly on building land, in particular, in the case of Amsterdam with the developments prepared by Berlage and Van Eesteren (plan of 1936).

Soviet Union: Equilibrium West–East

Around 1930, after a decade of ideological debate, the Soviet Union set up a policy of rigid economic planning (quinquennial plans) and the spatial translation of these plans (distribution of activities, in particular of manu facturing large plants, infrastructures, housing, and public amenities). The policy of economic planning has been associated with a policy of territorial planning in order to reduce disparities between the European and Asian parts of the Union. Indeed, in the first one were concentrated three quarters of the population and 80% of the agricultural and industrial production; while, the second represented three quarters of the territory, 80% of raw material and of hydroelectrical potential, 90% of the coal reserves, etc. It was decided to create high capacity infrastructures west to east (railways and pipelines), but as well, to found, close to resources of raw material in Siberia and Central Asia, manufacturing complexes and, for this purpose, about 1000 of new towns.

Italy: Equilibrium North–South

At a different scale, Italy had to face development dis parities between the northern part, rich and industrial ized, and the southern one (*Mezzogiorno*), poor and essentially agricultural, due to geographical, historic, and social reasons (the south being dominated by an oligarchy of landowners and by a criminal organization). The gap, in terms of income, was from one to two. The result was an intense migration from south to north.

The Mussolini government had initiated a policy of financial subsidies to reduce this gap. But he did not try to transform the economical and social structures.

It was only during the 1950s that the *Cassa per il Mezzogiorn* was created to undertake a land reform, to create infrastructures, and to develop manufacturing activity. Despite the money engaged, an important gap is still remaining.

Britain: Decentralization

After Ebenezer Howard had written a book in 1898 proposing 'garden cities' to avoid huge metropolitan development, Letchworth (from 1903) and Welwyn (from 1919) were built in the north of London (**Figure 1**).

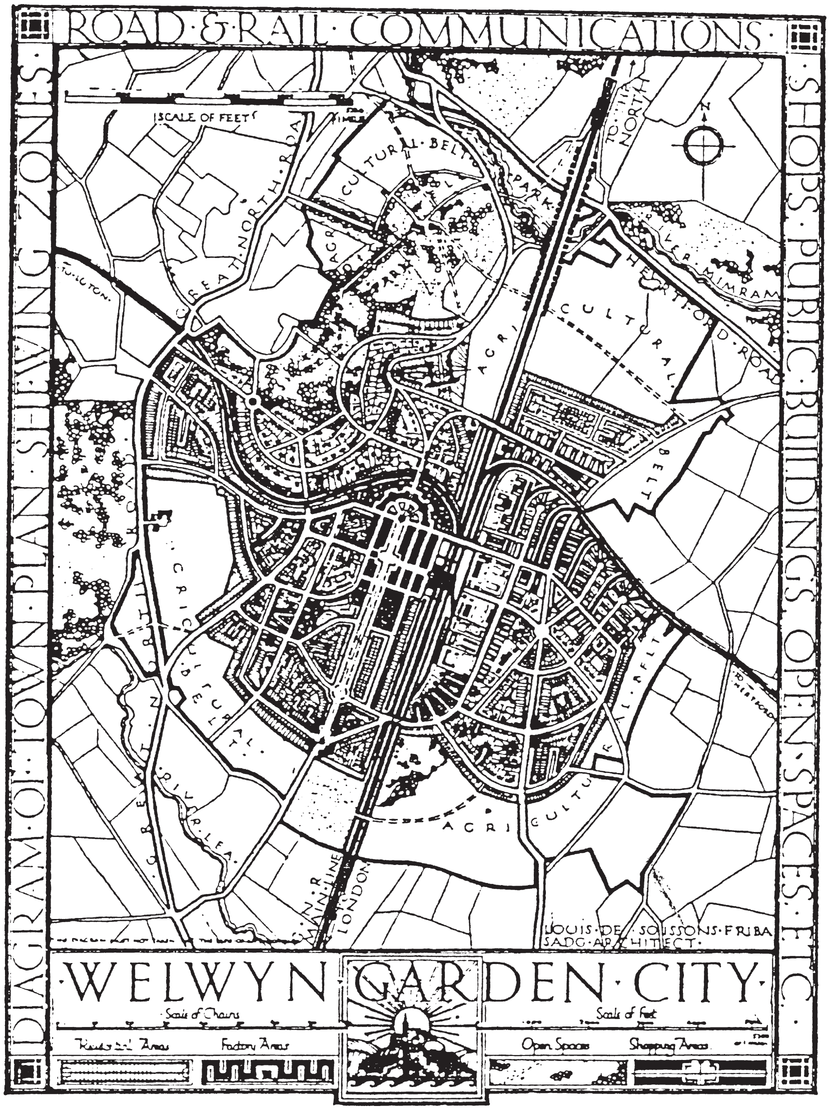

Figure 1 The plan for Welwyn Garden City by Louis de Soissons (1919).

During the 1930s, before other European countries, Britain, where industrialization had been more precocious, felt the first signs of the aging of traditional industry (metallurgy, mechanics, and textile) and, during the great economic crisis, the rise of underemployment. The Special Areas Act was voted. It delivered public funds to create industrial estates and to build factories ready to let.

In 1937, it decided to prepare a national territorial plan (the Barlow Report, published in 1940), which advocated for activities of decentralization from London region to 'development areas', created in 1945. Consequently, a Greater London Plan (Abercrombie Plan, 1944) was prepared for the London region. This plan has been the origin of the policy of activities overspill from central London to the outer part of the conurbation and of the construction of 'new towns' and 'expanding towns'.

The Five Great Dilemmas of Territorial Development

Territorial development is at the heart of many conflicts between objectives.

Economic Development versus Territorial Equality

The ordered disposition sought by territorial development does not necessarily correspond to an economic optimum, which would suppose a priority development of settlements and regions already favored by their localization, their assets in terms of manpower, image, etc. Territorial development, on the contrary, has, among its objectives, to recreate an equality of chances between the inhabitants, whatever the place where they live: opportunities of jobs and good salaries, access to public (especially education) and private (shopping and services) amenities. This often needs a public intervention to make up territorial handicaps. This is true in particular for developing countries, where the easiest solution is to give a priority to regions already developed (generally, the capital city, which grows to become a 'primatial' city, and frequently the seashore). In these countries territorial development policies generally do not exist or are purely formal. Even in developed countries, the support to territorial development policies is more vigorous when the economic growth is high than in periods of economic crisis.

Liberalism versus Planning

As a voluntary intervention, territorial development is a spatial planning, complementary of the economic planning, which 'spatializes' its perspectives. Moreover, the setting up of economic planning often takes place in those countries which have felt necessary the territorial development. The best example is the Soviet Union, even if the economic planning appeared to be too rigid and if territorial development was clearly subordinate to economic planning.

Territorial development, spatial section of planning, is naturally conflicting with the *laissez faire*, therefore with liberal ideas. Supporters of the last ones (including the European Union) do not cease to claim the winding out of constraints and of financial aids linked with territorial development, which they consider as concurrence distortions.

Efficiency versus Equity

Territorial development has as well to contribute to the solution of social disparities, which accompany often, but not always, spatial disparities. It is important to question who will be the final beneficiaries of subsidies given to regions receiving priority. A good example is the case of tourism, which beneficiates sometimes more to the investors, often foreigners to the region, than to local population (especially in the case of developing countries).

Territorial development, as concerned with equity as it can be, must however reach economic efficiency, in order not to be rejected. It cannot appear as a brake upon modernization in the name of maintaining traditional ways of doing, even if the last can represent a qualitative extra for the local economy (craft production, open air agriculture, sea fishing, etc.).

Development versus Environment

Choices of territorial development often conflict with environmental concerns. It is often the case for infrastructures, in particular, for motorways which favor emissions of greenhouse gas, noise, and pollution. It is where space is the rarest, therefore, the most desired and the most fragile that this dilemma has the greatest chance to create a conflict; this concerns notably not only the littoral, but also towns and often the mountains.

The development of tourism acutely appeals arbitration, search of a balance between development (construction of lodgings, creation of jobs, generation of receipts, and reception of currencies), protection of the environment (limitation of the frequentation and preserved areas), and social objectives (favor the opportunity of holidays for the largest part of the population).

Centralization versus Decentralization

Territorial development policies have this in particular that they are conceived at the national scale – even if it is concerned by various scales – and that they are expressed through mechanisms and actions exerted only on a part, sometimes limited, of the territory.

Orientations of territorial development are generally defined in a very centralized way. Their implementation is more often due to local authorities (regions or provinces, *départements* or counties, municipalities, etc.). A solution of this contradiction may be searched through procedures of contractualization for the length of a period of planning (now 2007–13 in the European Union) between state, regions, towns, local areas, etc.

Few Examples of Territorial Development Policies

Netherlands

It has already been said that Netherlands practiced territorial development before this expression was used. However, it has not been before 1962 that an act instituted territorial plans at the national scale. There has been, up to now, five such plans, plus revisions of some of them. This means that it is an evolutive planning, indicating priorities that can evolve:

- in the second plan (1966), the limitation of the growth of the Randstad, large conurbation in the west of the country (Amsterdam, Rotterdam, Den Haag, Utrecht, etc.) (**Figure 2**);
- in the third plan (1974–77), the constraints of the environment and a policy of new towns (*groeikernen*) for the overspill of big cities;
- in the third plan revised (1983–85), the policy of the 'compact city' appeared;
- in the fourth plan (1988), the search of quality and the reinforcement of large cities;
- in the fourth plan 'extra' (1991), a policy of localization of activities according to their relation with the transportation networks; and
- in the fifth plan (2001), urban reorganization, intensification of the use of land, and search for a greener space.

Figure 2 The second plan on physical planning for Netherlands (1966).

These national plans have been declined through the provincial and local plans.

Parallel to this huge effort of planning, spectacular projects were developed:

- the development of the Zuyderzee gulf by closing it with a 30 km long dyke (1932) and the creation of a vast lake (Ijsselmeer) behind the dyke and, in this lake, the construction of four polders (166 000 ha) devoted to agriculture with new villages and towns (**Figure 3**); and
- the closing of the delta, common to Rhine, Maas, and Scheldt through a double range of obstacles: bridges *c.* 25 km from the shore and of dams right at the shore.

Britain

Territorial development has been a priority after the Barlow Report (1940) and until the end of the 1970s. It had two main lines:

- the control of the growth of large conurbations, par ticularly London, with the Abercrombie Plan (1944) and the construction, around London and other big cities, of new towns (New Town Act, 1946) and of expanding towns (Town Development Act, 1952); and
- the decentralization of manufacturing (necessity to get an industrial development certificate, from 1947) and later of tertiary (Office permit, 1963) activities outside the London region, preferentially toward under industrialized or declining areas ('development areas', created in 1945), where aids, financial or others were attributed.

This policy, very willful, has been however braked down by the Labour government after the oil crisis (1973–74), then given up by the Thatcher government in the early 1980s, to be replaced by a policy of economic (Urban programs, 1977, City grants, 1982) and spatial (Urban development corporations, particularly for the renewal of London Docklands and of the Merseyside in

Figure 3 The plan for Almere New Town under construction in the Zuyderzee Polders.

Liverpool) revitalization using methods and capital of the private sector.

France

It is only after World War II that France has attempted to have a territorial development policy. As in Britain, the departure point was the concentration, considered as excessive, of population, activities, amenities and, decision of power in and around Paris.

The first measures (*Agrément* to create or develop factories, and later any activity building, instituted in 1955), financial (*prime d'aménagement du territoire*) and fiscal aids to the firms creating activities in other regions, aimed to decentralize activities and to develop other big cities, which were to become *métropoles d'équilibre* (1964) (**Figure 4**). The *Délégation à l'aménagement du territoire et à l'action régionale* (*DATAR*) was created in 1963 to implement these policies.

However, the latter has been inflected, then reoriented, notably after the oil crisis (1974) and new priorities (economic, social, and environmental) appeared:

- planning and development (nonagricultural activities and amenities) in the rural areas;
- preservation of the environment: creation of national and regional parks, of natural reserves, conciliation, on the littoral, and in mountainous areas, of the objectives of economic and development of tourism and of protection; and
- correction of the physical, economic, and social degradation in the collective housing estates (*grands ensembles*) built between 1955 and 1975.

At the same time, the institutional context has been modified:

- administrative decentralization (1982–83, reinforced in 2004) by transfer of competences from the state to regions, *départements*, and municipalities; and
- encouragement (acts of 1992 and 1999) to the grouping of municipalities (they are 36 784) by creation of *communautés urbaines* (for large conurbations), *communautés d'agglomération* (for smaller ones), and *communautés de communes* (in mostly rural areas).

An Example of a Developing Country: Lebanon

As Netherlands, Lebanon is a good example of the necessity of a territorial development. Beyond its contrasted

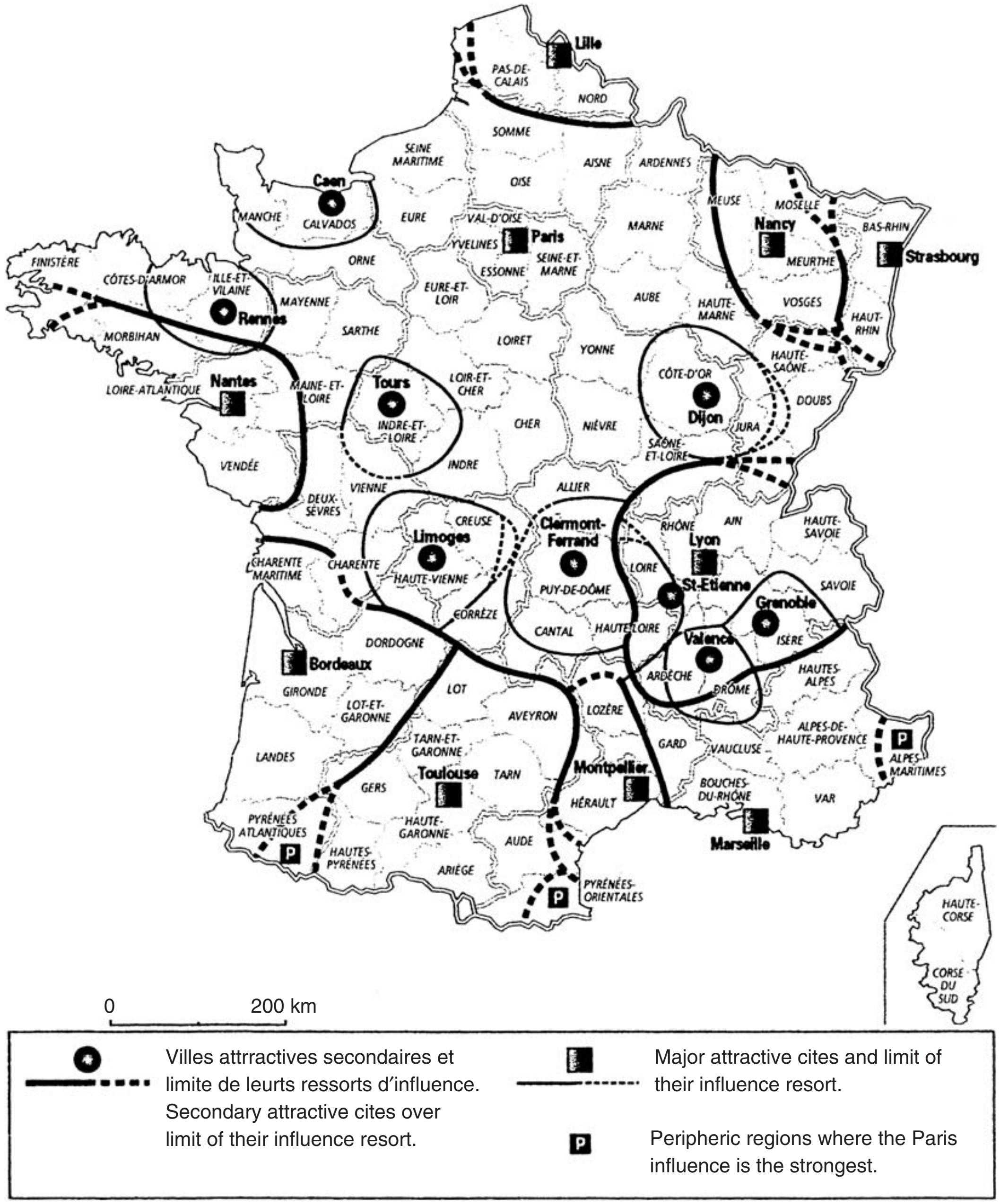

Figure 4 The definition of French *Métropoles d'équilibre* (1964).

relief, its exiguous territory (11 000 km^2 for 4 million inhabitants), the opposition between Beyrouth metropolis and the rest of the country, Lebanon is split from the religious point of view. The 1943 constitution and the Taëf agreement (1992), and also the long Civil War (1975–90), had for result that this religious division be came the base of the institutional organization of the country. Therefore, it is an absolute necessity that ter ritorial development contributes to cement the national unity. But, as different from Netherlands, there is, in Lebanon, no consensus around territorial development.

A *Schéma directeur d'aménagement du territoire libanais* has however been published. Its proposals aim at a unitary and integrated development of the national territory and at a rationalization of the use of resources, including water. For that purpose, this plan, beyond a proposed distribution of land, for which it specifies the rules to apply, advocates a reorganization of the territory around powerful urban poles, an economic development balanced between all the regions, the creation of a performing transportation net work, and an oriented distribution of important amenities. The implementation of this plan has, however, to face the maintenance of religious and political conflicts, the difficulty to assert national independence, the habit of non respect of the rules, and the weight of private capital not very concerned with the general interest.

Conclusion

Besides the five dilemmas that any policy of territorial development must face, the four examples presented underline the risk of a lack of continuity in the actions, new priorities tending constantly to replace the previous ones.

A policy of territorial development, not to be a simple alibi, needs a strong will of intervention to orientate the economic, social, and spatial choices; an agreement to reject the primate of economy, despite the globalization induces at it; a will of spatial and social equity; and a concern with care for nature.

These principles must be understood independently of any ideological preference, except the search of the general interest. However, they suppose a wide consensus around societal values to give an answer to the three major crises (energy, climate, and society) of the twelfth century. A policy of territorial planning will then encounter conditions that will make possible that it participates, at all the scales, to a real sustainable development.

See also: Planning, Urban; Regional Development Models; Regional Development Theory; Regional Geography I; Regional Geography II; Regional Science.

Further Reading

Cullingworth, B. and Nadin, V. (2001). *Town and Country Planning in the UK*. London: Routledge.
Gravier, J. F. (1947). *Paris et le Désert Français*, 421p. Paris: Le Portulan.
Hall, P. (2002). *Urban and Regional Planning*. London: Routledge.
Harrington, J. W. and Barney, W. (1995). *Industrial Location: Principles, Practice and Policy*, 236p. London: Routledge.
Institut d'aménagement et d'urbanisme de la région Ile de France et Dar el Handasah (2004). Schéma directeur d'aménagement du territoire libanais, mai, 233p.
Lloyd, P. E. and Dicken, P. (1990). *Location in Space: Theoretical Perspectives in Economic Geography,* (3rd edn.), 431p. New York: Harpers Collins.
Merlin, P. (2002). *L'aménagement du Territoire*, 448p. Paris: PUF.
Merlin, P. (2006). *Le Tourisme en France, Enjeux et Aménagement*, 160p. Paris: Ellipses.
Merlin, P. (2007). *L'aménagement du Territoire en France*, 188p. Paris: La Documentation française.
Merlin, P. (2008). *Tourisme et Aménagement Touristique: Des Objectifs Inconciliables?* (2nd edn.), 228p. Paris: La Documentation française.
Merlin, P. and Choay, F. (eds.) (2005). *Dictionnaire de L'urbanisme et de L'aménagement,* (4th edn.), 986p. Paris: PUF.
Scott, A. J. (1988). *New Industrail Spaces: Flexible Production, Organization, and Regional Development in North America*. London: Pion.

Americas

R. B. Kent, California State University, Northridge, Northridge, CA, USA
K. A. Butler, The University of Akron, Akron, OH, USA

The Americas is an immense geographical region spanning two continents, North and South America, and includes the islands of the Caribbean Basin. While the Caribbean Basin is a part of the Americas, that subregion is covered elsewhere in this encyclopedia. The Americas lies fully within the Western Hemisphere. The region spans nearly 150° of longitude extending from the eastern coast of South America at 35° W to westernmost reaches of the Aleutian Islands at 180°. The latitudinal extent of the Americas runs from Arctic islands north of 80° N southward to Cape Horn at 55° S. Encompassing the third and fourth largest continents, North America and South America, respectively, the land area of the Americas extends over more than 42 million square kilometer.

The first recorded reference to the term 'America' appeared in 1507 on a map produced by the German cartographer, Martin Waldseemuller. The name derived from the Latinized spelling of the first name of Americo Vespucci. Vespucci was a sixteenth century Italian explorer and merchant who made two voyages (1499 and 1502) to the eastern coast of South America. He was among the first European explorers to recognize that they had not reached India but had discovered a new continent instead. Use of the place names (toponyms) 'America' and 'the Americas' can be problematic. As noted above the Americas is usually defined as a macro geographical region encompassing North America, South America, and the Caribbean basin. However, the terms 'America' and 'Americans' in the common vernacular are synonymous with the United States of America and its citizens, an association that often draws howls of protest from those living in Central and South America who also claim that they are 'Americans' too.

Regional Perspectives

Conceptually, the Americas can be subdivided into subregions using distinct criteria. From an environmental or geographical perspective, it is often divided into three macro divisions – the continental landmasses of North America and South America, and the small narrow land bridge, Central America, which joins the two continents. The human geography of the region is better understood and conceptualized differently. At the broadest scale, two principal cultural regions can be identified – Anglo America and Latin America. Anglo America corresponds essentially to those countries, the United States and Canada, which trace their colonial history and their principal cultural heritage and dominant language to England (Great Britain). The regional delimitation of Latin America is also determined from the region's colonial heritage, and pertains to those countries which were colonies of Spain or Portugal. 'Iberoamerica' is also a term used to describe the same region. The term originates from the fact that in Europe, both Spain and Portugal occupy the Iberian Peninsula. Latin America is further subdivided among those areas which experienced Spanish colonial rule, Spanish America or Hispano America, and those areas that were under the colonial control of the Portuguese, Luso America.

The contemporary human geography of these macro regions, Anglo America and Latin America, is by no means homogenous. The French speaking region of Canada, Quebec, along the St. Lawrence River, is a notable anomaly in the Anglo America realm. Francophone Quebec is home to approximately 7.5 million Canadians and comprises nearly 24% of Canada's total population. Quebec has a distinct culture within the greater Canadian society and separatist sentiment has been strong at times. In the 1960s, Quebec enacted provincial laws to protect and strengthen its identity as a separate and unique culture. The official language is French and nearly all place names, business activities, and education are conducted in French. With significant political clout and the Canada's second largest provincial economy, Quebec, has used its power to secure its identity as a distinct 'nation' within the Canadian state. Other French speaking enclaves are found in Latin America as well, most notably Haiti and French Guiana, an overseas territory of France (**Figure 1**).

Contemporary migration streams, largely from Mexico and other Latin American countries into the US's southwest and its major urban areas, are also creating anomalies in Anglo America's cultural landscape.

Amerindian peoples maintain distinct cultural identities in rural settlements and reservations across the United States and Canada. Many of these extant groups are few in number and geographically restricted, but others like the Navajo peoples in the southwestern United States, number in the 100 000 s and occupy large geographical areas. In Latin America, the survival of Amerindian peoples as well as their cultures and languages represent the most significant exceptions to the

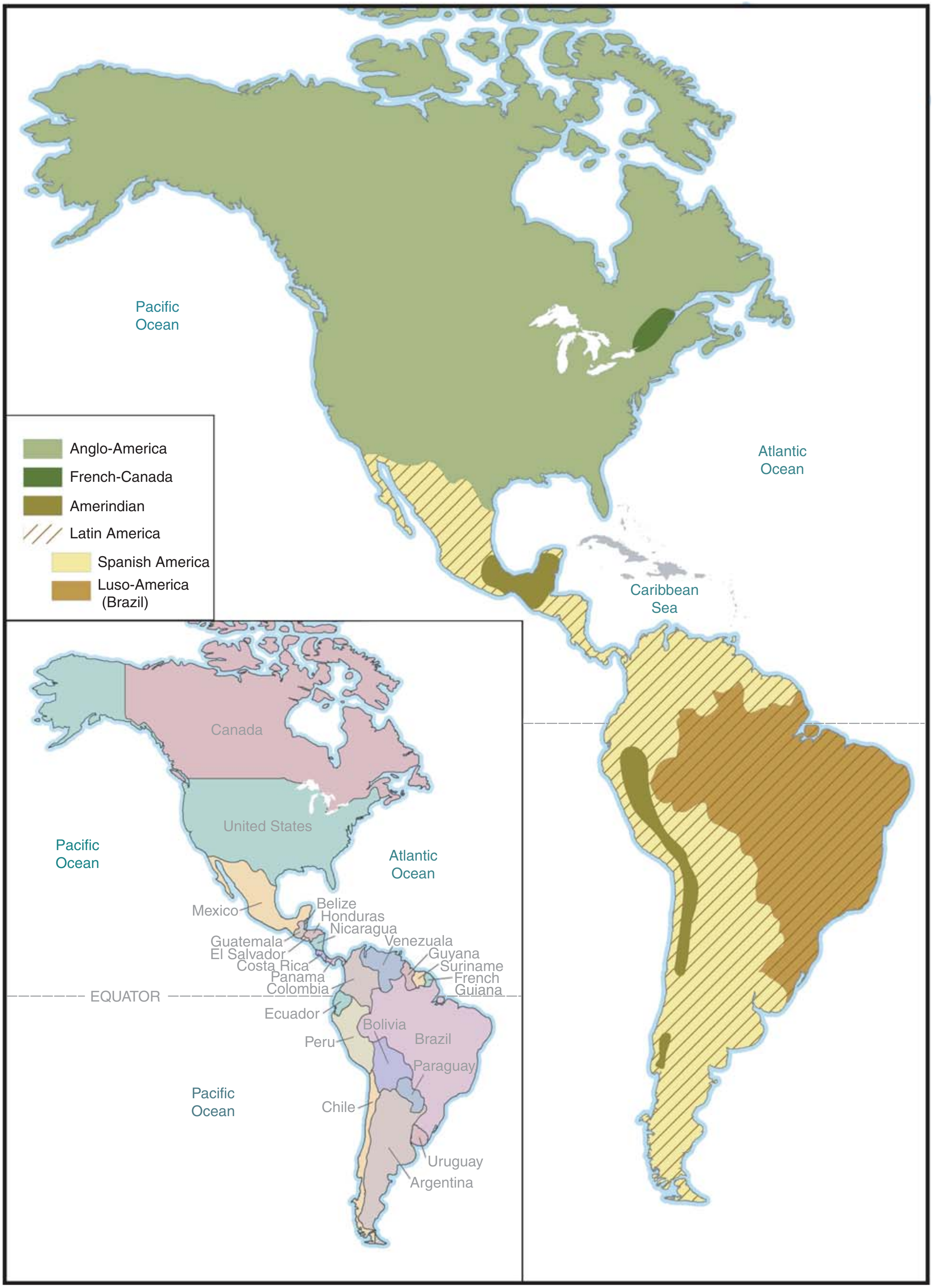

Figure 1 Principal cultural regions and nation-states of the Americas.

region's cultural homogeneity. These Amerindian cul tures are most significant, and number in the millions in three or four principal areas in Latin America. These include the Nahuatl and Maya in southern Mexico and northern Central American, the Quechua and Aymara in the Andes, Guarani peoples in the Amazon Basin and Paraguay, and Mapuche in southern Chile. Other cultural anomalies exist in Latin America, but generally these are small and geographically and numerically insignificant. Despite these exceptions, when compared to many other world regions, especially at a continental scale, the Americas exhibit a strong degree of cultural and social homogeneity in terms of human geography.

African immigration, most of it forced through the institution of slavery between the sixteenth and nine teenth centuries, has had a profound impact on the human geography of the Americas. At present, peoples of African descent represent a substantial proportion of the population in Brazil (30–40%) and in the United States (close to 15%). Populations are also locally dense along the Pacific coastal areas of northern South America and on the margins of the Caribbean in both South and Central America. Regionally distinct Afro American cultures are important in Brazil, the United States, and other countries with significant populations. Their cul tural contributions are especially notable in music (e.g., blues and jazz), cuisine, and religious beliefs and prac tices (e.g., candomblé in Brazil). All told, peoples of Af rican descent account for approximately 150 million of the inhabitants of the Americas (excluding the Carib bean) – nearly 20% of the region's total population.

Human Antecedents

Prior to the arrival of Europeans and their colonization of the Americas at the beginning in the early sixteenth century, Amerindian peoples had occupied much of the Americas for at least 10 000 years. The level of social and economic development among these peoples varied tre mendously. Pre agricultural societies dependent on hunting and gathering, early agricultural communities, and highly developed advanced civilizations which do mesticated plants and animals all flourished in the Americas prior to their conquest and subjugation by Europeans. In the southwestern portion of North America, advanced agricultural societies evolved with populations that numbered in the hundreds of thousands. However, the region's most advanced societies, truly high civilizations, evolved in two places – on the Mexican Plateau and the adjacent southern highlands and low lands, in a cultural region known as Mesoamerica, and in the central Andes. In both regions, a series of cultures had evolved through the preceding millennium culminating in sophisticated societies. In Mesoamerica, early cultures included the Olmec, Zapotec, and Toltec among others … culminating with the Aztec and Maya societies en countered by the Spanish conquistadores at the begin ning of the sixteenth century. In the Andes and on the adjacent Pacific coastal plain a similar pattern prevailed and early civilizations included Moche, Chimu, Wari, and Tiahuanaco culminating with the Inca civilization in the two centuries preceding the European conquest. In these areas of high civilization, population densities were high and population numbers are estimated to have ranged as high as 25 million in the Mesoamerica region and around 15 million in the Andes.

Cultural Commonalities

Although the human geography of the Americas is sometimes described as diverse, in fact the peoples of this region share many common cultural characteristics. These include similar colonial histories, cultural orien tation, religious traditions, political values and systems, and economic organization. With the exception of Aus tralia, the contemporary human geography of the Americas is one of the least varied in the world when examining regions on a continental scale.

The Colonial Experience

Culturally and politically, all of the countries and peoples of the Americas share a strong colonial legacy. For nearly two centuries, the seventeenth and eighteenth, England and France occupied and controlled much of the eastern half of North America leaving a permanent and indelible imprint on the present day cultures and societies of that part of the continent, the United States, and Canada. Spanish colonial control extended over much of the rest of the Americas including the western and southern parts of North America (the modern day US states of Cali fornia, Texas, and Mexico), Central America, and all of South America except the present day country of Brazil, which was a Portuguese colony. Colonial control in the Spanish and Portuguese areas lasted from the early to mid sixteenth century to the beginnings of the nine teenth, almost a century longer than in the French and Anglo areas in North America.

The colonial experience and legacy varied in different spheres of influence. In the English and French colonies, Amerindian populations were small and the European settlers soon subjugated, and sometimes eliminated, these peoples, eventually overwhelming them as seemingly endless waves of new settlers arrived from Europe. French colonial control and influence had been elimin ated from North America by the end of the eighteenth century. In the areas that eventually became the United States and Canada, the colonists and later settlers, after political independence, established new societies that replicated many of the social, economic, and political systems and practices of England and other northwestern European countries.

In the colonies controlled by the Spanish and the Portuguese, a different colonial experience unfolded. In these regions, direct migration from the home countries to the colonies was insignificant as opposed to the experience in the US and Canada. Another key difference was that in some regions, especially those controlled by the Spanish Crown, dense populations of Amerindians survived the colonial experience despite conquest and exposure to devastating epidemics caused by Old World diseases. These peoples became the colonial subjects of the Spanish and Portuguese Crowns and suffered the imposition of a colonial bureaucracy. Colonial control was deep and longlived and it transformed indigenous social, economic, and political systems and practices and imposed models from the Iberian Peninsula in southern Europe forming the cornerstones of the region's societies today.

Cultural Orientation to Western Europe

The present day human and political landscape of the Americas reflects strong cultural and historical ties to Western Europe. Religious affiliation and orientation reflects one of the strongest and most profound ties of the Americas to the Europe. Christianity is absolutely dominant throughout the Americas. Catholicism predominates in those parts of the Americas once colonized by the Iberian powers of Spain and Portugal (Latin America) and is the official religion in some countries there. Within Latin America, evangelical Protestants have made stunning inroads on the religious landscape since the 1950s and at the beginning of the twenty first century, perhaps as much as 20% of the region's population professes an evangelical Protestant faith. In the Anglo America realm of the Americas (US and Canada), a more variegated pattern of religious beliefs occurs. Christianity is still undisputed as the dominant religion, but the number and variety of its denominations is more diverse here than perhaps anywhere else in the world. Perhaps as much as one quarter of the population professes no religious affiliation. Among those that do adhere to a clear religious faith, Catholics comprise a sizeable proportion, perhaps as many as one half, while traditional Protestant denominations (Lutherans, Presbyterians, Episcopalians, and Methodists), Baptists and Southern Baptists, and a myriad of evangelical Protestant denominations account for most of the remainder. Other distinctive Christian denominations which have prospered in Anglo America include the Mormons, Amish, and Mennonites. A small minority, perhaps totaling just 2% of total population, practices a range of other world religions – Hinduism, Judaism, and Islam, among others.

Political ties to Europe are also significant in some cases. Canada and the United States are closely tied to Europe through their membership in the North Atlantic Treaty Organization (NATO). This alliance, established in the immediate post World War II period, serves to insure the security and freedom of its 29 member nations, all but two of which are located in Europe. In recent decades, Spain has increasingly ratcheted up its efforts to reconnect with its former colonies in the Americas, investing considerable resources in fostering closer cultural, economic, and political ties with those states.

Dominant language usage represents another significant unifying characteristic of human landscape of the Americas. The predominant languages today, and almost without exception, the languages of government, business, commerce, and social life, are European in origin. Furthermore, three of the four, Spanish, Portuguese, and French, are closely related Romance Languages sharing a common root in Latin. English, the other dominant tongue, is Germanic in origin, but also shares significant influence from Latin.

Indigenous languages have survived on a limited scale in the Americas, but most survivals are limited to small numbers of speakers in geographically limited areas. There are some exceptions. In Mexico, nearly 10 million people, 10% of the nation's population, speak an indigenous language. Nahuatl and Maya speakers are the most numerous and they number in the millions. Maya speakers predominate over a broad area of southern Mexico and neighboring Guatemala, where they account for 40–50% of the population. In South America, about 10 million Quechua speakers are distributed throughout the Andes Mountains, running from southern Colombia into northern Argentina, with the highest concentration of speakers found in Bolivia. Aymara speakers represent perhaps 1 million people in Peru and Bolivia. Despite a Spanish colonial heritage, in Paraguay, Guarani, the language of preconquest Amerindians there, survived and is spoken by nearly the entire population of almost 7 million, although almost all are bilingual in Spanish. In southern Chile, about 1 million people speak Araucarian (Mapuche), the language of the region's people prior to the European conquest. Despite these and many other exceptions to the dominant use of European languages in the Americas by Amerindians, most Amerindian individuals are at least somewhat bilingual in one or another of the region's dominant European languages.

Political Values and Orientation

A widely held commitment to democracy and representative government is another common characteristic of the nation states and peoples of the Americas. With an end of colonial rule beginning at the end of the eighteenth century, and inspired by the American and French

Revolutions and informed by the experience of the Enlightenment, the newly independent governments of the region enshrined democratic values in constitutions and political systems. The distance between theory and practice was often immense and democracy and representative government often took a back seat to dictators and military rulers for much of the postcolonial period. The United States and Canada are important exceptions to this generalization. Nevertheless, by the last decades of the twentieth century, democratic values and institutions had become more firmly entrenched both politically and culturally (notably within the Latin American portion of the region). At the present, in the first decade of the twenty first century, all of the region's states can be reasonably described as representative democracies. No significant military intervention in civilian governance anywhere in the region has occurred in recent decades.

For nearly two centuries, the United States of America has cast a long shadow over much of the Americas, and especially Latin America where it has intervened, often militarily, in the internal affairs of its neighbors. This has been justified by a long standing policy known as the Monroe Doctrine, first enunciated in 1823 by then president James Monroe. Monroe asserted that the Americas were no longer to be the object of European colonization or interference, and that when necessary the United States would intervene to protect the sovereignty of other states in the Americas. And it has done so directly on many occasions … sending its military into Haiti, Dominican Republic, Mexico, Nicaragua, Grenada, Cuba, and most recently Panama. Protecting US economic interests has usually provided the justification for these incursions and the prospect of European intervention has almost never been an issue.

Economic Organization and Market Systems

Today, capitalism is the principal economic ideology and it dominates the political economy of the nation states of the region. Generally, a neoliberal capitalist economic model prevails in one form or another. Free and open markets, a reduction in state participation or regulation of the economy, transparency in financial dealings, and a reduction of trade barriers are among the model's key pillars. This has not always been so, and during the twentieth century many nation states in Latin America experimented with a range of economic models. Most of these models were essentially capitalistic in one way or another, crony capitalism and state capitalism, for example. These practices usually involved a favored elite class with strong intervention of the state in the economy and often included state ownership of key industries, a practice still common in some Latin American countries – Mexico, Venezuela, and Bolivia, for example. A few experiments with more socialist modes of economic organization occurred, as in Chile and Peru, but these were generally short lived. Chile's experiment with socialism died in a coup supported by the United States in 1973, while Peru's transitioned peacefully into a capitalistic democracy.

Contemporary Human Geography

The territory of the Americas is demarcated into some 22 sovereign nation states and one overseas territory, Guyane (French Guiana). The geographical size, population, economic clout, and role in the world political system of these nation states, varies tremendously. The United States, as one of the world's most powerful nation states, is both the region's most populous nation, with close to 300 million inhabitants, as well as its most powerful. In terms of political and economic geography, the America's clearly fall within the sphere of influence of the United States, but rapidly growing trade with China and its increasingly active diplomacy and economic investment in the region are challenging long standing US hegemony. After the United States, Brazil and Mexico stand out as the most significant regional powers in terms of areal extent, population, and economic activity. Argentina has at times projected itself successfully as a regional power and leader, while Venezuela, bolstered by oil revenues and the vision of a charismatic leader, has recently attempted, mostly unsuccessfully, to do so as well.

Population Characteristics

Overwhelmingly, the people of the Americas are urban dwellers, living in towns, cities, and major metropolitan centers, and their values and lifestyles reflect that fact. Throughout the region some 75–80% of the total population is urban, although in some countries, especially in Central America, the average is slightly lower. A few of the world's most important urban places, as well as some of its most populous ones, are found here. New York, a city of about 10 million, is one of the world's top tier metropolitan centers and a global center for financial activity. Other world class metropolitan centers whose populations near or exceed 10 million include Chicago, Los Angeles, Toronto, Rio de Janeiro, and Buenos Aires, while Mexico City and São Paulo, whose populations exceed 20 million, are among the world's most populous urban places.

Despite many shared cultural characteristics and similar economic systems, the nation states of the Americas vary tremendously in terms of their population numbers and socioeconomic status. The population of the region stands at about 850 million and represents slightly less than 15% of the world's population. Population distribution is highly uneven, and most of the region's population is concentrated along or near the coastal margins of the continents with vast interior regions which are sparsely populated (e.g., the Great Plains, the Pampas, the Amazon Basin). Overall, population densities in most countries are comparatively low and pressure on the natural resources modest. Central America is a notable exception, and there population densities are high and pressure on natural resources is also great. Among the various nation states, population distribution is also un even and there is considerable variation between coun tries. The United States, with 300 million inhabitants, is the most populous country and alone accounts for well over one third of the entire regional population. Brazil and Mexico, with populations of about 190 and 105 mil lion, respectively, are the next most populous nation states in the region. Nearly three fourths of the region's total population resides in just these three countries. No other countries have populations that exceed 45 million in habitants, and about half of the region's nation states have populations of less than 10 million. A handful, are really mini states, with populations less than 1 million – Belize, Guyana, and Suriname.

Some parts of the Americas, notably Latin America, experienced rapid population growth during the twen tieth century. In Brazil and Mexico, population numbers increased by a factor of five during the course of the century and similar patterns pertained in many Latin American countries. In the more economically developed nations of Chile, Argentina, and Uruguay population growth was more modest. By the end of the twentieth century, population growth rates had begun to moderate as populations became increasingly urban, social norms and values changed, and birth control and abortion both became more widespread. Overall, the rate of natural population growth across Latin America stands at about 1.5%, although it is higher in Central America (>2.0) and less in the historically developed countries of South America (<1.0). In Anglo America, natural population growth rates are very low – 0.6 in the United States and 0.3 in Canada. In both cases however, pro immigration policies have lead to steady and appreciable increases in total population numbers since the 1960s as large num bers of both legal and illegal immigrants have migrated to those two countries.

Socioeconomic Conditions

Compared to many other major world regions, espe cially at a continental scale, for the vast majority of the population, socioeconomic conditions in the Americas are adequate and basic human needs are largely met. However, that said, income levels and socioeconomic conditions across the region are quite uneven. One way to explore the level of economic development of a re gion is to examine its gross national income (GNI) per capita. But, because the cost of the same goods or ser vices can vary dramatically from one country to another, economists adjust the GNI using a measure called purchasing power parity (PPP). This measure, GNI PPP, measures how much in goods and services can be purchased in a particular country. Thus, it can be used to effectively compare the true purchasing power of money from one country to another and the standard of living.

Dramatic disparities in GNI–PPP are evident be tween Anglo America and Latin America. The United States and Canada are far and away the most eco nomically prosperous nation states of the Americas, where per capita GNI–PPP stands at US$44 260 and $34 610, respectively. In comparison to affluent Anglo America, there is a precipitous drop in the per capita GNI–PPP in Latin American countries. Surprisingly, despite a myriad of economic problems and economic policy blunders, Argentina has the highest per capita incomes in Latin America, US$15 000. Only a handful of other countries have per capita rates that exceed US$10 000 – Mexico, Chile, Uruguay, and Costa Rica. Many of the region's remaining nation states have per capita GNI–PPP that is below US$5000, while the re gion's poorest country is Bolivia with a rate of US$2900. Nevertheless, in some ways these data may misrepresent living conditions on the ground. Bolivia, for instance, is indeed a poor country, but hunger and abject poverty are not at all the rule.

Continuing advancements in communications tech nology and the spread of these technologies to all sectors of society will continue to transform the Americas pro moting both economic and social integration. Tele phones, cellular telephones, and Internet access have increased the level of social and political interaction throughout the Americas. In both the United States and Canada, there are more phones than people and tele phone and cell phone service is essentially ubiquitous. Nearly 70% of Anglo America has access to the Internet, with both Canada and the United States enjoying similar levels of access.

In Latin America, access to telecommunications technology is more restricted especially in poorer countries and in rural environments, but in most urban environments and among at least the middle class, access to telephones, cell phones, and the Internet is common. In Mexico and Costa Rica, two comparatively well off countries, only about 25% of the population have access to either a telephone or a cell phone and this climbs to a high of over 40% in Chile. At the other end of the spectrum are countries like Guyana and Nicaragua where less than 10% of the population has access to a telephone or a cell phone. Access to computers and the Internet follows a similar pattern. About 20% of the population has access to these technologies in Mexico

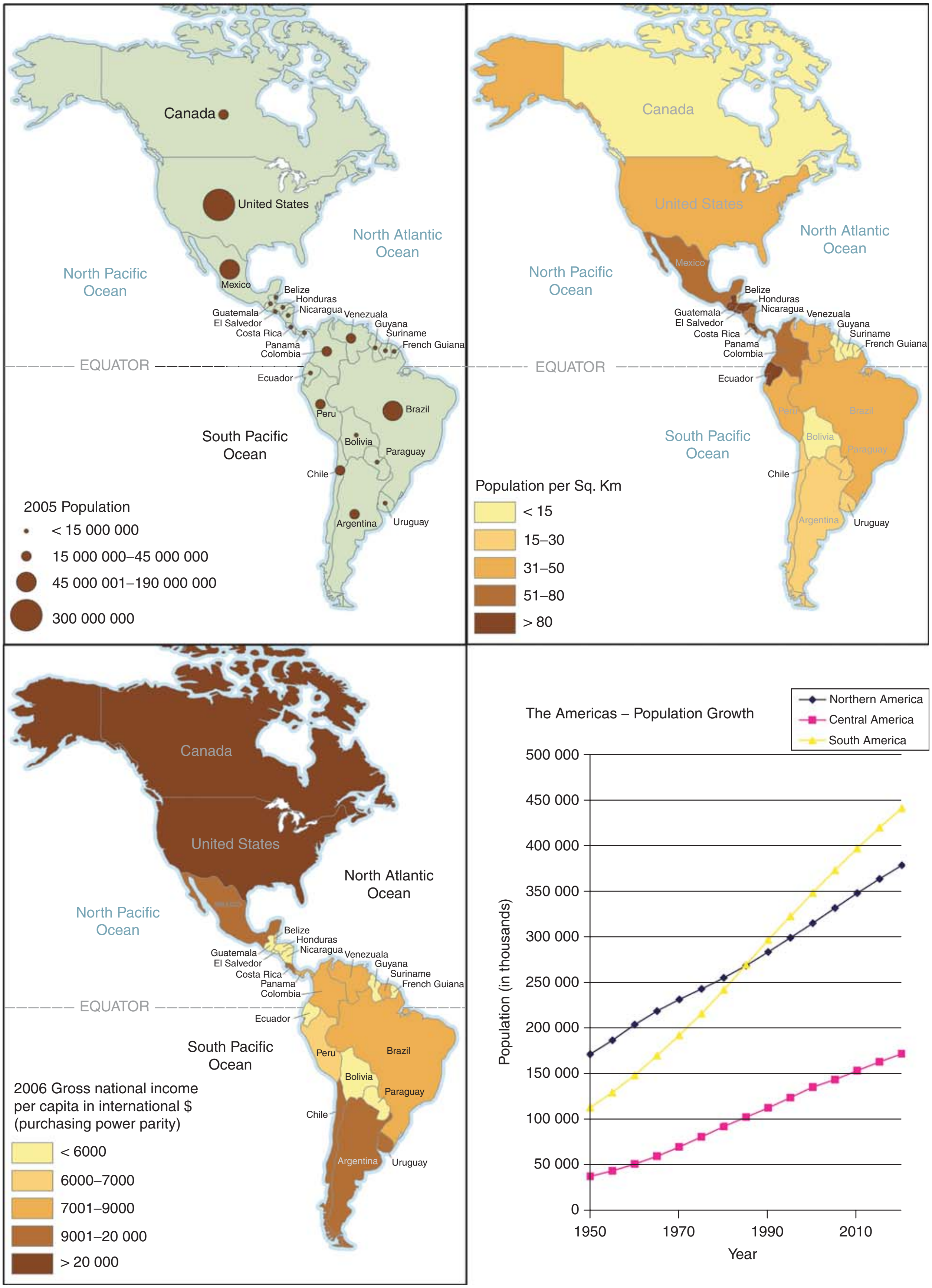

Figure 2 Contemporary demographic and economic conditions in the Americas.

and Costa Rica, and this figure increases to 30–40% in Chile, Argentina, and Uruguay. Nevertheless, in the region's poorer nations, Bolivia, Paraguay, and Ecuador, this statistic does not exceed 10% (**Figure 2**).

Future Directions

The twenty first century will bring greater cultural and economic integration to the Americas and increasing homogeneity in terms of the region's human geography. Throughout the twentieth century, the United States has been the region's hegemonic power, exerting a tremendous influence over both culture and economy. This dominance is unlikely to be seriously challenged in the coming century, although China's growing trade and investment in the region represents an important counterpoint to the US's influence here. To some extent the increasingly ubiquitous presence of a plethora of electronic communications technologies will work to allow continued US penetration in Latin America's cultures and economies. On the other hand, low birth rates, demand for both unskilled and skilled labor, and weak immigration laws in the United States have and will continue to encourage the flow of immigrants from Latin America to the US and to a lesser degree to Canada. These migration flows, numbering perhaps 1 million annually, are transforming the cultural geography of Anglo America, especially the United States, making it more 'Latin' as these immigrants settle permanently. Substantially higher birthrates among Latin American immigrants multiply their impact. In the United States, Hispanics, those who trace their heritage to Latin America, now account for 15% of the total population and will likely account for 25% within 50 years.

Economic integration and the removal of trade barriers between nation states is one of the central tenants of contemporary capitalist world economy. Free trade, comparative advantage, and open markets are among the objectives of these efforts. In the Americas, early efforts at economic integration began in the middle of the twentieth century with the formation of groups like the Andean Pact and the Central American Common Market. But these efforts were largely ineffectual. Subsequent efforts have proved far more successful. The most notable example of these free trade agreements in the Americas is the North American Free Trade Agreement (NAFTA). This agreement between the United States, Canada, and Mexico became effective in 1994. NAFTA removed many of the tariffs – duties charged on imported goods – on popular imports such as automobiles, electronics, and textiles. Agricultural tariffs, long the subject of heated debate in the US agricultural Midwest, were negotiated bilaterally and tariffs are being phased out rather than eliminated outright. NAFTA has resulted in substantial increases in trade between the member countries. While there is no consensus on whether NAFTA has benefited the economies of the United States, Canada, and Mexico public policy in all three nation states suggest that economic and political relations between them will continue to expand and deepen.

While NAFTA is perhaps the most well known multilateral free trade agreement, others exist in the Americas. MERCOSUR – *Mercado Común del Sur* or Southern Common Market – is a free trade agreement between Brazil, Argentina, Uruguay, and Paraguay established in 1991. Chile and Bolivia later joined MERCOSUR as associate members. There has been revitalized interest in MERCOSUR in direct opposition to the predominantly United States' backed plans to expand NAFTA into the Free Trade Area of the Americas (FTAA). There has been significant resistance to the concept of FTAA in some countries, but the United States and other partners in the region are moving incrementally toward the goal, often through bilateral agreements between the United States and individual countries (e.g., Chile).

By the middle of the twenty first century, the members of NAFTA and MERCOSUR, as well as the remainder of nation states of the Americas are likely to be joined in some form of customs union, or perhaps even a common market similar to that envisioned by the FTAA. Given the region's historical ties and strong cultural links to Europe, it is not inconceivable that this economic block would be joined or allied in some way with the European Union.

Further Reading

Collier, S., Skidmore, T. E. and Blakemore, H. (1992). *The Cambridge Encyclopedia of Latin America and the Caribbean* (2nd edn.). Cambridge: Cambridge University Press.

Garreau, J. (1981). *The Nine Nations of North America*. Boston, MA: Houghton Mifflin Company.

Gastil, R. D. (1975). *Cultural Regions of the United States*. Seattle, WA: University of Washington Press.

Ignatieff, M. (1994). *Blood and Belonging: Journeys into the New Nationalism*. New York: Farrar, Straus, and Giroux.

Lewis, M. W. and Wigen, K. E. (1997). *The Myth of Continents: A Critique of Metageography*. Berkeley, CA: University of California Press.

Meing, D. W. (1989 1998). *The Shaping of America: A Geographical Perspective on 500 Years of History*, Vols. 1 3. New Haven, CT: Yale University Press.

Paige, J. A. (1995). *The Brazilians*. New York: Addison Wesley.

Riding, A. (1984). *Distant Neighbors: A Portrait of the Mexicans*. New York: Vintage Books.

Skidmore, T. E. and Smith, P. H. (2001). *Modern Latin America* (5th edn.). Oxford: Oxford University Press.

Winn, P. (1995). *The Americas: The Changing Face of Latin America and the Caribbean*. Berkeley, CA: University of California Press.

Zelinsky, W. (1994). *The Cultural Geography of the United States* (2nd edn.). Englewood Cliffs, NJ: Prentice Hall.

Relevant Websites

http://www.internetworldstats.com
Internet usage data from nation states of the world.
http://www.glenmary.org
The Glenmary Research Center provides data, maps, and published reports on religious demographics for the United States.
http://pdba.georgetown.edu
The Political Database of the Americas provides extensive information on political processes, branches of government, and elections throughout the Americas.
http://www.fas.usda.gov
The United States Department of Agriculture site provides broad background information on NAFTA and reports on its progress and impacts on United States agriculture.
http://landcover.usgs.gov
The United States Geological Survey provides background and maps for land cover from the local to world scale.
http://www.worldbank.org
The World Bank provides a variety of detailed standardized economic and population data for nation states of the world.
http://www.worldvaluessurvey.org
The World Values Survey is a comprehensive long term assessment of values and culture throughout the world.

Anarchism/Anarchist Geography

M. M. Breitbart, Hampshire College, Amherst, MA, USA

Glossary

Decentralism The revolutionary social anarchist philosophy and organization of space necessary to support new cooperative modes of living and working.

Federation A union comprised of a number of self-governing territories or organizations who freely constitute themselves to cooperate for a number of varied purposes (economic, social, cultural, etc.).

Psychogeography The study of the effects of the environment on the behavior and emotions of people in order to better understand and subvert the forces that control the circumstances of everyday life.

Social Ecology Recognition of the social basis for environmental problems: a philosophy and activist practice that attributes imbalances in human/environment relations to hierarchies and forms of domination in social relationships (and vice versa).

What Is Anarchism?

Anarchism is a philosophy that argues against statism and a political order based on authority or hierarchy. It opposes the use of power to secure the privileges of a few, restrict individual freedom and deny the rich diversity of social life. While anarchism shares with socialism a critique of capitalism and the desire to replace repressive economic structures with common ownership of the means of production and distribution according to need, it stresses the importance of eliminating authoritarian relationships wherever they arise. It also promotes the need for parity of means and ends in the social change process. This includes direct action and efforts to bring into being, in the present, environments that experiment with new communal forms and promote greater popular participation, equality of condition, freedom, and social justice.

Anarchists maintain a belief in the capacity of people to base their economic, social, and political lives upon cooperation and federation, and to function without the imposition of structures of domination. While individualist anarchists stress the overarching importance of personal autonomy, social anarchists seek to demonstrate the importance of basing personal freedom and creative development on social responsibility, a strong collective foundation, and supportive social environments.

The roots of these ideas trace back to several nineteenth century activists and theorists, including some prominent geographers. Many of their ideas were revived within the pages of *Antipode: A Radical Journal of Geography* in the 1970s, a time when a growing number of geographers demanded greater relevance and attention to social inequality from their discipline. The actual term 'anarchist geography' likely emerged during this time.

The Historical Relationship of Anarchism and Geography

The confluence of anarchism as a political philosophy and geography as an emerging discipline occurred in the latter half of the nineteenth century largely in response to European nation building and growing social inequities. As taught in the halls of European military colleges, geography was often complicit in promoting imperial conquest by mapping resources to benefit European nations. Two prominent dissidents within the field, Elisée Reclus (1830–1905) and Peter Kropotkin (1842–1921), employed their geographic knowledge to counter these tendencies. Their efforts contributed not only to the progressive development of the discipline, but to revolutionary movements for social justice and the foundation of social anarchism as a philosophy and practice. In the effort to pursue their profession as geographers, Kropotkin and Reclus uncovered principles of social and physical life that underscored the need for revolutionary change in their society. They also believed that a revolution in social and economic relationships would necessitate the creation of totally new or renovated land use patterns and built environments. Spatial concepts are thus key to social anarchist philosophy and strategies to promote radical social change.

The Social Geography of Elisée Reclus and Peter Kropotkin

One can best understand the evolving relationship between anarchism and geography, its theories and practical applications, in the lives and body of writing generated by two prominent geographers, Elisée Reclus and Peter Alexeyevich Kropotkin. Their contributions to geography in the form of new theories about human ecology, decentralization, and experiential education were clearly aimed at halting the use of geographic research for imperialistic ends and toward the discovery of models of better balanced and more humane social relationships and relations between people and their environments. Their notoriety stemmed from a different source – a growing commitment to radical social change and their role in the development of a social anarchist philosophy and practice.

Development of a radical theory of social ecology

In a comprehensive analysis of the contributions of Reclus to social geography and ecology, John Clark claims that both Reclus and Kropotkin shared a particular optimism based on their broad love and understanding of the natural world and people, but also on their belief in the power of science to provide a better understanding of the determinants of the social world and examples of more balanced human/environment relations. Both geographers believed that this knowledge would provoke critiques of social inequality and the abuse of nature that would result in the creation of people who were more active, responsible agents in their own liberation. Reclus devotes considerable attention in his six volume work, *L'Homme et la Terre* (1905–08), to analyses of the effects of the human exploitation of nature and examples of how resources can be used to promote social well being. Establishing the basis of a contemporary radical theory of social ecology, both geographers believed that imbalances in nature reflected imbalances in human relationships and suggested that people base their use of the natural world on a respect for, and understanding of, its key properties. Kropotkin and Reclus assumed that sustainable human/environment relations could only be initiated through social transformation and fundamental changes in human values that would promote the demise of capitalism, racism, the modern State, gender inequities, and other forms of social hierarchy. They perceived that these changes would then be supported by a progressive sense of place, greater human interaction, and the centrality of love. While Kropotkin tried to develop a basis for higher moral standards from the natural world, Reclus assumed that moral development would come from the growing scope of our knowledge and attachment to key life systems.

As students of the natural world, Reclus and Kropotkin found unity and progress in diversity. They believed that people would derive their understanding of both progressive and flawed social relationships from the active and deliberate exploration of their communities at the ground level. Some of their initial ideas about the power of particular kinds of egalitarian social and human/environment relationships were based on their observations of life among the Swiss Jura watchmakers in the late nineteenth century who overcame many social and environmental obstacles to progress through complex cooperative social and spatial networks. Kropotkin also developed his concept of 'mutual aid', in part, as a reaction against the rise of Social Darwinism in the 1890s, which argued for the importance of competition over cooperation in human relations. Using counter examples drawn from the natural and social world to establish the basis for a moral standard, he hoped to support cooperative impulses and encourage people to replace freedom blocking institutions with those that support collaborative work. Much of this basis for a theory of social ecology and municipal democracy was taken up in the late twentieth century and developed further by anarcho feminist ecologists and social anarchist Murray Bookchin.

Decentralism: The sociology and geography of anarchism

Kropotkin also believed that certain kinds of human–environment relations and social and spatial modes of interaction were more conducive to development of the human potential. Decentralism is the revolutionary social anarchist philosophy and socio spatial form of organization that he believed would comprise the foundation of a new cooperative anarchist mode of living. These ideas are developed in several pieces of writing, much of which is considered prophetic in its anticipation of contemporary problems of food security and critiques of large scale enterprise, the dehumanization of work, and the lack of imagination in learning. Kropotkin's version of 'decentralism' envisions a multitude of associations federated for the purposes of trade and production, intellectual, and artistic exchange, as well as communal villages and urban neighborhoods federated across geographic and thematic boundaries for the purposes of distribution, consumption, and innovation.

Both Reclus and Kropotkin anchored their ideas for communal decentralism not in fanciful utopias but in direct human observation, scientific analysis, and experimentation with new social and spatial forms of living. They also encouraged people to engage in active, participatory experiential learning from a young age, believing that this approach to education would capitalize on the natural imagination and curiosity of young people and encourage a critical examination and analysis of both the past and the present. These ideas were best articulated in a moving plea for social relevance in education entitled 'What geography ought to be', written by Kropotkin in 1885. Here, he describes the importance of building on the passions and natural curiosities of children to want to know about different people and places. He also suggests a potential role for geography as a discipline to redirect its prior support for imperialist ventures toward helping to dissipate national rivalries, racial prejudice, and class interests.

Anarchist Decentralism in Spain, 1936–39

Nowhere did the ideas generated by Kropotkin and Reclus about a communal space economy have more extensive application than in Spain in the late nineteenth and early twentieth centuries when anarchism emerged as a strong revolutionary movement. When Franco and his fascist generals attacked the newly elected Republic in July 1936, thousands of industrial workers and peasants responded with militias and also with a massive collectivization of land, factories, transportation systems, and public services. Collectivization encompassed more than one half of the total land area of Republican Spain, affecting the lives of nearly eight million people. Large cities like Barcelona were transformed into federations of neighborhoods, while in many parts of the Republican held countryside, new irrigation systems and well organized federations of communes allowed peasants to bring new land under cultivation, expanding and diversifying production. Social landscapes accommodated new educational, cultural, and health facilities. Massive regional exchange networks formed by federations of collectives starting at the local level and working their way up to districts and provinces, linked cities with the countryside for the purposes of distribution and consumption, extending transportation and health services into areas that had never been serviced before. A revolution, which began by creating more communal and egalitarian relationships among people, resulted in the creation of highly efficient and environmentally sensitive new spatial formations.

None of these developments would have taken place without decades of preparation for self management and the elaborate organizing strategies used by anarcho syndicalist trade unions that acknowledged the importance to workers and peasants of an affiliation based on both trade and locale (pueblo). These tactics encouraged people operating under extremely oppressive conditions, to become critical observers and analysts of the economic system under which they labored and the political structures that secured inequality and unequal divisions of power. Strikes and pueblo seizures, however temporary, also provided the space and opportunity for Spanish workers to experiment with many alternative communal forms of workplace and community management. It is possible to trace many efforts to decentralize economic and social space, and create whole new communal systems of regional communication and exchange, directly to the writings of Kropotkin and the ideas spread by the travel of Elisée Reclus and his brother Elie in Spain prior to the start of the Spanish Civil War and social revolution in 1936. Decentralist practices similar to these have emerged more recently in Argentina, in the struggles of the Zapatistas in Mexico, and in contemporary anti globalization struggles.

The Early Influence of Social Anarchism on Land-Use Planners

Anarchist theories of social ecology and decentralism developed by geographers Kropotkin and Reclus have also influenced many practitioners and theorists in the spatial disciplines who sought solutions to a variety of social problems partly through environmental planning. For example, it is claimed that his close reading of Kropotkin, heavily influenced the more communal residential forms and innovative ideas for linking agriculture and industry, found in Ebenezer Howard's early and more radical formulation of the Garden City alternative.

Paul Goodman and his architect brother, Percival, also drew heavily on Kropotkin's decentralist concepts and examined their implications for planning in their groundbreaking book, *Communitas: Means of Livelihood and Ways of Life* (1947). Here, they decried the dehumanizing effects of bureaucracy and social engineering while demonstrating through hypothetical community planning scenarios, how built environments could display and reinforce differing social priorities from efficiency to communality. Encouraging people to reject externally imposed designs for living, the Goodmans echoed Kropotkin's call for people to transform themselves into active change agents and to decentralize decision making in planning wherever and whenever possible.

Patrick Geddes, a prominent Scottish biologist and town planner, also drew heavily upon social anarchist philosophy and Kropotkin's concept of decentralism in the promotion of participatory, experiential education, and urban regional development. To encourage responsible citizenship, Geddes devised and built the Outlook Tower in Edinburgh to promote active environmental exploration of the social networks and the physical environment of the city by ordinary citizens. Drawing on the ideas of Kropotkin and Reclus, he sought to awaken students' imaginations by moving them beyond book study into the field as the site for active inquiry based learning. The Outlook Tower became the hub of these activities and the repository of visual findings. With a *camera obscura* on the roof, the building was open to all, and encouraged shared insights and accumulated knowledge.

In the 1970s, British architect and anarchist, Colin Ward, founder of Freedom Press and former Director of Education for the Town and Country Planning Association, reinvented this concept of a place to train young explorers in participatory design and planning in the many Urban Studies Centres he inspired throughout the UK. He also wrote several books on how to encourage young people and adults to become actively involved in their cities and towns. Ward reissued an edited and updated version of Kropotkin's *Fields, Factories and Workshops* in the 1970s and wrote extensively on dweller control of housing, squatting, and numerous efforts to reclaim space in the city and countryside for the purposes of survival and free public use. In the years since, he has continued to advocate for open minded, flexible planning and extensive resident involvement in housing.

In addition to those traditions that emphasize social planning, it is important to recognize other aspects of anarchist philosophy promoted in the 1950s and the 1960s by groups such as the avant garde Situationist Internationale, whose philosophies influenced the worker and student riots in France in May 1968. The Situ ationists, and particularly the writings of Guy Debord, demonstrate how the urban environment and everyday sensory experience are conditioned by the 'society of the spectacle', driven by the ethos of commodified capitalism and the need to sell both products and lifestyles. Part of their revolutionary response was to utilize the techniques of psychogeography such as the 'dérive' (unplanned drifting through the city to awaken real desire) and 'détournement' (an artistic technique to subversively reappropriate space through the transformation of mes sages found on such artifacts as billboards). These tactics were meant to contest the passivity of the spectacle, break the hold of commodities, and create terrains of resistance. Hakim Bey's promotion of temporary au tonomous zones (TAZs) as sites of free expression and the encouragement of imaginative and experimental short term uses of space is a more recent manifestation of this approach.

Anarchist Geographies: The Spatial Foundation of Anarchist Theory and Practice

Space is important to the social anarchist revolutionary project in several key respects. Social injustice, op pression, inequality, and the sources of power that underlie these relationships are often reflected in, or reveal themselves through, the structuring and appear ance of the landscape and built environment. Anarchists also recognize the importance of creating social and physical spaces in which to develop strategies of resist ance to the exercise of political and economic power, and for experimentation with alternative social formulations. Finally, the transformation of space plays a key role in anarchist conceptions of a new society built upon a foundation of cooperation and exchange rather than hierarchy.

Inciting Protest, Building Desire: Spatial Reflections of Domination and Inequality

Under capitalism, the study of the underlying forces behind the structuring and appearance of space reveal how many social needs are ignored, and ecological bal ances disturbed, by land use patterns that promote the rights of property over need or the public good. An archists believe that 'reading' a landscape in such a fashion enables conceptual connections to be drawn be tween social inequities, the exercise of power through public policy, and space. It is thought that the stark visibility of injustice inscribed in the environment can motivate a desire for greater knowledge of the underlying causes, as well as spark protest and direct action.

In pre Civil War Spain, inequities and manifestations of the centralized power of the state and church were clearly reflected in the landscape. Peasants with few means of survival were crowded into small villages while large landowners left most of the countryside un cultivated. In other regions, a limited capitalist agrarian and industrial economy produced massive unemploy ment, housing crises, and the need for the presence of the Guardia Civile in every neighborhood of most large cities. A more contemporary example of how space can be used in organizing to draw attention to conditions of domination and inequality involves the use of public art and memory in anti gentrification struggles. In present day Liberty City, Miami, images of once viable working class neighborhoods and the names of the residents displaced by urban renewal strategies were recently dis played in the spaces undergoing change. Housing acti vists and residents believed that the placement of such disruptive elements in a gentrifying landscape that itself depended so heavily on projecting an image of upscale transformation could subvert the redevelopment process by revealing and tainting hidden agendas and stirring public protest.

Strategies of Resistance and Transformation: Insurgent Place Making

Efforts to create alternative, place based development scenarios follow closely from the kinds of tactics of re sistance described above. Social anarchists do not draw precise blueprints for a new society or the built en vironment that supports it, believing that alternative social and spatial formations must emerge from precise historical circumstances as well as local preferences, needs, and assets. The broad objectives of a decentralist vision are nevertheless clear to: underscore the import ance of personal freedom within the context of social responsibility; allow the free development of self while supporting the freedom of others; and enable continuous growth and change in environments that have the eco logical and social capacity to sustain present and future generations.

To move closer to this ideal, anarchists encourage the creation of insurgent citizens and organized grassroot mobilizations that are designed to subvert the agendas of the state or private developers by claiming rights to such necessities as affordable housing, quality education, healthcare, and other social provisions. Anarchists also seek to identify positive tendencies and assets that al ready exist in a society (or place), and to enlarge the space for these tendencies to grow and percolate even while situated within an otherwise oppressive set of social constraints. In social anarchist theory, it is believed that the contradictions unleashed by creating ‘free’ spaces that nurture resistance, hope, and experimentation with alternative ways of living and working, will contribute to productive dissidence and the spatial imaginary. It is then hoped that the contradictions between ‘what is’ and ‘what could be’, as well as the visibility of well functioning collective social institutions and noncapitalist work settings, will move people beyond their more limited experience, and support their belief in the possibility for change. The alteration of space is, in this way, meant to become a key part of a process of bringing new visions of the future into the present and motivating action.

Spatial Alternatives to Hierarchy: From Small Scale to the Mobile and Networked

Bringing imagined spaces of active participation, freedom, and cooperative production into being within the hierarchical strictures of present day society necessarily places emphasis on the local and small scale. The idea is that by reducing the scale of institutions and enabling people to establish collective control over them, more intimate and egalitarian social relationships, and caring approaches to the environment, will result. Less understood but equally important to the theory and practice of anarchist decentralism is the utilization of the geographic constructs of mobility and networking in the formation of more efficient and sophisticated alternatives to hierarchical structures for organizing business and the social and economic base of community life. Social anarchism challenges many notions of economic efficiency by demonstrating how various forms of technology, divisions of labor, and workplace organization are not based on scientific principle, but rather on property interests and modes of production that allow enormous discrepancies in wealth and power to exist. Smallness of scale, self management, networks, and federations are posed as alternative ways to both structure the economic landscape on a more egalitarian basis and produce alternative avenues to efficiency in systems of production, distribution, and consumption.

As applied in Spain during the anarchist social revolution of the 1930s, villages and neighborhoods were not expected to become self sufficient entities. Rather, they were made part of elaborate space economies that promoted interregional exchange and cooperation. While greater regional self sufficiency was sought to reduce dependency and facilitate more equal relationships among people, limited specialization became part of the active interchange of information, products, knowledge, and even labor as facilitated by multiple federative structures. To avoid cultural isolation and enhance the environment for innovation, many extra territorial linkages of people were also established to enable the collaborative exchange of ideas and cultural endeavors. In a contemporary vein, while the potential of transnational networks of resistance and collective enterprise may not be fully realized, anti globalization efforts have recognized and begun to utilize new communication technologies to forge ever more effective extra local connections among relatively autonomous groups in their organizing strategies.

Anarchist Epistemologies of Freedom: Challenging and Reconstructing Geographic Knowledge and Planning Practice

Neither Kropotkin nor Reclus abandoned their geography to become political and social anarchists. Instead, they set out to extend the idea of revolution and social critique from the realm of politics alone into the realm of geographic inquiry. Many contemporary geographers and planners have since taken on this project and applied anarchist social and spatial principles to both a critique of spatial planning and the acquisition of geographic knowledge and teaching.

Questioning Knowledge Hierarchies

Fundamental to the anarchist vision of social change is the importance of consistency between means and ends. Social anarchists do not believe that authoritarian driven struggles can produce anti authoritarian ends. Much attention is therefore devoted to the importance of decentering knowledge and promoting education for self management. Transgressing knowledge boundaries often produces anxiety in academic circles and yet anarchist geographers and planners argue forcefully for democratizing education and promoting the free flow of ideas and ‘chance encounters’. They call for a more liberating, critical, and reflexive form of learning to be taught in the schools and in venues outside the classroom. Besides challenging conventional ways of knowing and transgressing the boundaries set by individual disciplines and professions, they advocate for transgressive relational thinking and practice as well as the examination of community life from its margins.

This process is intimately connected to the formation of liberatory pedagogies that question all hierarchies of knowledge and support forms of mutual learning that facilitate choice rather than impose ideas or mold character. The implications for epistemologies and pedagogies of higher education are profound. Kropotkin and Reclus were among the first to call for geography as a discipline to be taught both inside and outside the classroom by teachers who would be drawn from all walks of life. Today, community based learning and engaged scholarship are sanctioned by even the most traditional colleges and universities who see the value of extending learning venues outside the academy and into the 'real' world, enabling students to augment and apply what they are learning to pressing social issues. While anarchists would applaud such approaches, they would caution students and faculty to take their direction in such settings from the grassroot organizations they are working with. They would suggest the need to build relationships upon a foundation of reciprocity, account ability, reflection, and sustainability (i.e., extending proj ects beyond academic semesters and over time through long term partnerships).

Promoting Participatory Design and Insurgent Planning Practices

In anarchist geography, the emphasis on decentering knowledge has also been applied to a critique of the professional realm of spatial planning. In this light, an archists devote considerable thought and experimen tation to uncovering the methods by which people can transform themselves into the kinds of individuals who become active agents of change in their own lives and who function without submitting to authority. In the spatial disciplines, this has taken the form of calls for a return to the active exploration of local environments by residents, and more participatory planning and design. In the US, demands for greater citizen involvement in planning followed urban riots in many cities in the late 1960s, and tumultuous protests of the disastrous impacts to low income and working class residents displaced by urban renewal policies of the mid twentieth century. As in the pedagogical debates alluded to above, epistemo logical questions were raised about professional versus indigenous knowledge and about monopolies on ex pertise, including who has the right to apply their knowledge to address social problems.

Since the 1960s, various forms of critical and par ticipative practice have been incorporated into local planning, including efforts to actively involve recent immigrants, and children and youth in the planning process. These approaches involve the application of a range of methods designed to uncover local assets and facilitate the study and representation of problems that have social and environmental consequences. Environ mental exploration, storytelling, and community based and asset mapping are key approaches used by anarchist geographers and radical planners to create more engaged citizens and generate new models of communication to better negotiate people's hopes and desires. The impacts of these practices have infiltrated the academy and the spatial disciplines in the form of new discourses and in the invention of new approaches to research. Most no table is participatory action research, a methodology that seeks to directly involve community residents in the framing of problems and research questions, the choice of study methods, data collection, the interpretation of re sults, and the design of actions in response. Anarchist geographers are actively involved in promoting partici pative approaches to research that have political impli cations and shift power downward in a genuine way. However, they caution practitioners to be aware of the challenges presented by people's complicated attach ments to place and by a nonlinear process that requires a great time commitment and can produce unexpected results. Anarchists also argue that the encouragement of greater participation in planning be seen as one of several ways to create room for negotiating needs and trans forming spaces that allow residents to imprint their surroundings and better articulate and meet their needs.

Dismantling the 'Ordered' Landscape

One significant roadblock that anarchist geographers see to greater citizen participation in innovative planning and urban design are regulations and laws pertaining to the use of space. They argue that these are often unrelated to people's needs and are solely there to support particular power relationships and uneven development. Anarchist geography thus challenges overt efforts to manipulate behavior or deny free access to space through design, while also disputing the practice of 'master planning' and all efforts to divide space into neat categories or impose overly determined restrictions on use. Many geographers have pointed to the basis in fear of spatial policies de signed to contain perceived disorder (e.g., the Haussma nization of Paris that replaced thousands of working class neighborhoods with wide boulevards or the spatial for tressing policies that restrict the mobility of the homeless today). Anarchist geographers move beyond this critique to argue forcibly against all forms of planning that utilize spatial codes for the purposes of social control or the containment of diverse life styles.

David Sibley's studies of the urban gypsies in Hull, England provide one example of how planned spaces can contain embedded and false assumptions about people and social structures, thus inhibiting free association. He describes how planners have attempted to use restrictive classifications of space to control and alter the activities of nonconforming groups. Sibley believes that such ef forts to homogenize and 'purify' space through the pre cise assignment of its use violates the rights of people to survive through the integration of activities in space (e.g., living and working). Rather than support culturally di verse dwellers in their efforts to enhance autonomy and self sufficiency, he argues that planners tend to define these groups as 'deviant'. Sibley also suggests that much of the mathematical spatial modeling that went on in geography in the mid twentieth century – Central Place theory and the geometry of Walter Christaller, for ex ample – was an effort to bring order to a disorderly physical world while building the power and prestige of its practitioners. Critics of earlier Garden City or New Town movements as well as more recent critics of the New Urbanism draw heavily upon this opposition to order for order's sake and the stifling of creative ex pression through over regulation of the uses and ap pearances of public and private space.

The Inevitability of Uncertainty: Planning Flexible Spaces to Accommodate Ambiguity, Diversity, and Change

Accompanying the call for a more informed and involved citizenry, and a belief in the social intentionality behind spatial design, is the insistence by anarchists that planning processes and regulations be less restrictive in order to meet changing needs and diverse constituencies. Anarch ists eschew artificial and prescribed utopias, and do not believe that modifications to the built environment can substitute for radical change in the fabric of social life. They do nevertheless believe that design can help to create spaces in which positive social tendencies and the fulfillment of human needs can incubate. In opposition to the modernist project of state centered planning with no acknowledgment of contradiction and conflict, anarchists argue for the return of a visionary reality, activism, and the social imagination to planning and design. This would take the form of new models of regionalism and the design of more flexible spaces to accommodate integrated and multiple uses of space that are open to the possibility of alteration. Anarchist planning would also promote the practice of using existing structures differently over time and of recycling spaces as needs change.

In an era of globalization, where cities have come to house many transnational communities, and neighbor hoods are often characterized by the constancy of change, anarchist geographers would argue as well for greater acceptance of heterogeneous cultural geographies. They would suggest that spaces remain fluid to accommodate the conflict and ambiguity of actual social life. This in cludes attention to the profound adjustments necessitated by the movement of people across borders and to the changing meanings and attachments to place that result. While not antiplanning, anarchists oppose state centered planning and emphasize the need to decentralize formal decision making to the most local level possible given historically and situation specific circumstances. They also encourage the creation of many informal social centers that bring disparate groups together to secure everyday spaces for experimentation and the development of co operative work related, residential, and cultural settings. They believe that occupations of unoccupied buildings and the creation of informal sites, such as a growing number of Occupied Social Centres in Italy and the rest of Europe, can support survival and promote practical alli ances among residents in their struggles with local au thorities around a range of issues that affect the quality of daily life. Openness to changing plans and access to spaces that bring different people together are considered crucial if social diversity is to become a basis for a new collective politics, and if resistance to the continued erosion of social welfare wrought by neoliberal political agendas is to be mounted on a local level.

Charting a Course: The Potential Bricolage of Anarchism and Geography

The revolution in geography that Kropotkin and Reclus initiated well over a century ago remains tied to an as yet unfulfilled social revolution based upon anarchist prin ciples and practices. Their prescient insights have nevertheless altered the practices of contemporary geographers and planning professionals in profound ways, and their ideas continue to present a challenge to all of the spatial disciplines and their practitioners. In one sense, the challenge is the same as that posed by Kro potkin a century ago: the need for geography to re invigorate active exploration of local environments as a means of fostering more critical social analysis, de veloping more effective networks of resistance, and spurring the creative incubation of alternative social and work environments to meet real needs and unmet desires. Questions that remain concern how best to accomplish this mission. How, for example, can geography better employ social anarchist principles to catalyze the active involvement of people at the local level in the creative processes of environmental study? What methods might enable the study of space and its transformation to sup port struggles for social justice and ecological responsi bility while extending free and creative expression? How do we deploy geographic knowledge to invite rather than stifle dialog among people with differing perspectives and experientially derived expertise? How can the powerful spatial tools of federations and networks be better used to link social movements across space? These are questions that would have interested Kropotkin and Reclus greatly.

See also: Darwinism (and Social Darwinism); Ecology; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Radical Environmentalism; Radical Geography.

Further Reading

Amin, A., Massey, D. and Thrift, N. (2003). *Decentering the Nation: A Radical Approach to Regional Inequality.* London: Catalyst.

Bey, H. (1991). *T.A.Z.: The Temporary Autonomous Zone, Ontological Anarchy, Poetic Terrorism*. Brooklyn, NY: Autonomedia.
Blunt, A. and Wills, J. (2000). *Dissident Geographies: An Introduction to Radical Ideas and Practice*. London: Pearson Education.
Bookchin, M. (1982). *The Ecology of Freedom: The Emergence and Dissolution of Hierarchy*. Palo Alto, CA: Cheshire Books.
Breitbart, M. (1975). Impressions of an anarchist landscape. *Antipode* 7, 44 49.
Breitbart, M. (1978). *The Theory and Practice of Anarchist Decentralism in Spain, 1936 1939: The Integration of Community and Environment*. PhD Dissertation, Clark University.
Breitbart, M. (ed.) (1979). *Special Issue: Theme Issue on Anarchism and Geography. Antipode* 10.
Breitbart, M. (1979). Spanish anarchism: An introductory essay. *Antipode* 10, 60 70.
Breitbart, M. (1979). Anarchist decentralism in rural Spain, 1936 1939: The integration of community and environment. *Antipode* 10, 83 97.
Breitbart, M. (1981). Peter Kropotkin, the anarchist geographer. In Stoddart, D. (ed.) *Geography, Ideology and Social Concern*, pp 134 153. Oxford: Basil Blackwell.
Clark, J. (2004). *Anarchy, Geography, Modernity: The Social Thought of Elisée Reclus*. Lanham, MD: Lexington Books.
Dolgoff, S. (ed.) (1974). *The Anarchist Collectives: Workers' Self Management in the Spanish Revolution, 1936 1939*. New York: Free Life Editions.
Dunbar, G. S. (1978). *Elisée Reclus: Historian of Nature*. Hampden, CT: Archon Books.
Goodman, P. (1947). *Communitas: Ways of Livelihood and Means of Life*. Chicago, IL: University of Chicago.
Kropotkin, P. (1885). What geography ought to be. *Nineteenth Century* 18, 940 956.
Kropotkin, P. (1899). *Fields, Factories and Workshops Tomorrow* (edited, introduced by Colin Ward 1985). London: Freedom Press.
Kropotkin, P. (1902). *Mutual Aid: A Factor in Evolution*. London: Heinemann.
Mac Laughlin, J. (1986). State centered social science and the anarchist critique: Ideology in political geography. *Antipode* 18, 11 38.
Meller, H. (1993). *Patrick Geddes: Social Evolutionist and City Planner*. New York: Routledge.
Pickerill, J. and Chatterton, P. (2006). Notes towards autonomous geographies: Creation, resistance and self management as survival tactics. *Progress in Geography* 30, 1 17.
Pickles, J. (1986). Geographic theory and educating for democracy. *Antipode* 18, 136 154.
Reclus, E. (1905 08). *L'Homme et la Terre*, (6 vols). Paris: Librairie Universelle.
Sibley, D. (1995). *Geographies of Exclusion: Society and Difference in the West*. London: Routledge.
Sibley, D. (1998). Sensations and spatial science: Gratification and anxiety in the production of ordered landscapes. *Environment and Planning A* 30, 235 246.
Ward, C. (1973). *Anarchy in Action*. London: Allen and Unwin.

Relevant Websites

http://peoplesgeography.com/
Peoples Geography (Reclaiming Space).
http://Autonomousgeographies.org
Autonomous Geographies.
http://dwardmac.pitzer.edu/Anarchist Archives/
Dana World's Homepage.
http://www.lib.umich.edu/spec coll/labadie/labadie.html
University of Michigan Library.

Anglo-American/Anglophone Hegemony

M. B. Aalbers, Columbia University, New York, NY, USA
U. Rossi, University of Cagliari, Cagliari, Italy

Introduction

One of the two authors of this article was born and raised in Italy, and is presently also working in an Italian uni versity. The other was born and raised in the Nether lands, and has temporarily moved to an American university after having worked in a Dutch university for a number of years. Both of us grew up with English as a second language, and we have had the opportunity to live in an English speaking country for a short period of time. Also, both of us have had the opportunity to work in each other's country for a short period of time. But although we have had to try to 'overcome' Anglophone or Anglo American hegemony, our situation is different in many other respects. In Italy, within the human and social sciences, publishing in English is valued, but is not considered as something unavoidable. In the Nether lands, on the other hand, publishing in English language journals is compulsory, and particularly important if one's contract is up for renewal or extension. Journals in which Dutch scholars are supposed to publish – the 'international' journals – are in general also relatively easy to access; most libraries carry many journal sub scriptions on paper as well as digitally. This is different in Italy where access to 'international' journals – whether in printed or in digital form – is more scarce and limited to more prosperous or large scale universities. In short, language is one important factor in shaping hegemonies in academic publishing, but access to 'international' publications is another one. Local culture is a third factor that clearly intervenes with access, since access to 'international' work is considered less important when a local or national academic environment is more inward looking and less outward looking. Of course, it should also be noted that in English speaking academic circles, there is also a tendency to look inward: to publications from their own country and to publications from other English speaking countries. In many cases English pub lications from nonnative speakers are virtually ignored, or reduced to 'case studies from the Other'.

In recent years, many authors have debated the in creasing international hegemonic position that Anglo phone countries, and most notably the US and the UK, have acquired in human geography over the last decades. This situation is not unique to geography. In fact, it is typical of social sciences more generally, because his torically a strong geographical bias has affected the production of social scientific knowledge. Yet, more than in other social sciences, a lively body of literature in human geography has started to discuss the so called Anglo American or Anglophone hegemony. The central premise in this literature is that within human geography English has become the dominant language used in the production, reproduction, and circulation of knowledge. As a result, what are considered to be 'relevant' or 'in fluential' international journals are almost exclusively English language journals in which predominantly native English speakers publish. Protagonists in this debate are geographers pursuing an intellectual argument which is sensitive to the critiques of cultural Euro centrism and rationalism and the related modes of discourse that have been developed since the late 1970s onward by thinkers and social scientists such as Edward Said, Homi Bhabha, Donna Haraway, and Pierre Bourdieu. It is to be noted that, with the notable exception of Pierre Bourdieu, all these leading theorists and critics of the Anglo American and, more generally, Western intellectual hegemony are members of British or North American universities, and are thus integrally part of the system they criticize. To some extent, this is also true for the proponents of the Anglo American hegemony thesis in human geography. This literature is indeed firmly positioned 'within' and not 'outside' the Anglo American geographical academia, and those in the forefront of this critical position who are not working in Anglo American universities in many cases have been trained in these universities.

This article first reports on the quantitative evidence offered about the extent of and reasons behind the sup posed Anglo American hegemony in human geography and then outlines the direction in which the debate has developed in recent years. Next, some underlying rea sons are discussed, most importantly the different roles of language, translation, 'Othering', and access. Finally, some critiques are discussed.

Is There an Anglo-American Hegemony in Geography Journals? The Empirical Evidence

It is easier for native English speakers to publish in English language journals, just like it is easier for native French speakers to publish in French language journals, or for any other native speaker publishing in journals in her/his language for that matter. Yet, the situation is different because English is the dominant language in social science and so being a native speaker of English does become a major advantage. Hence, it is not sur prising that most papers in the so called 'international' journals, that is, an important share of English language journals, are written by native speakers.

In a survey supporting the Anglo American hegemony thesis, Gutiérrez and López Nieva have shown that of the papers published in a selected number of major English language geography journals, 38.3% are written by US based authors and 35.1% by UK based authors. Two other predominantly English speaking countries follow at 8.6% and 3.2% (Canada and Australia, respectively). Israel is the first non English speaking country in the list with a meager 1.5% and is directly followed by two pre dominately English speaking countries. All in all, less than 14% of all papers appear to be written by authors based in countries where English is not the dominant language. Of course, there are differences between journals as well, but none of these 'international' journals features more than 25% authors from non English speaking countries. Like wise, the editorial boards of these journals are predomin antly made up of academics based in English speaking countries. At 41.0% and 37.8%, the US and the UK again stand out. Singapore (1.6%) and Japan (1.1%) are the only non English speaking countries that supply more than 1% of editorial board members. This often concerns native speakers or natives trained in English speaking countries. The differences between the different journals are very large with the *International Journal of GIS* standing out with 40.9% of its editorial board members from non English speaking countries, while about one third of the selected journals (7 out of 19) only have editorial board members from predominantly English speaking countries.

This situation does not only exist in human geography, other social sciences also show the predominance of authors from English speaking countries. The big differ ence between geography and other social sciences is the position of the UK: in geography, UK based authors publish almost as much as US based authors, while in the other social sciences the UK is far behind the US. On the other hand, Rodríguez Pose has demonstrated that a journal often attracts most authors from the country of publication and from other countries in which the same language is spoken as the language of publication (which is not necessarily the same as the language of the country of publication as two English language geography journals from respectively Sweden and the Netherlands show) and this would prove that there is no Anglo American he gemony, but merely a parochial division of papers over journals. Although empirically sound, this argument overlooks the fact that many English language journals are considered and consider themselves without exception, thus including journals published by national geography associations, – 'international' journals while most non English language journals are not considered – even when they consider themselves – 'international' journals. More over, it is rare to find non English journals that have been admitted to the ISI rankings, further undermining the position of non English journals and making it harder for these journals to acquire 'international' status.

Publication Practices and the Production of the 'Other'

One obvious observation is that with English being the dominant language it is easier for native English speakers to publish in English language journals. But this is not simply because non native speakers have a language disadvantage and native speakers a language advantage, but also because the most important journals – or, in fact, almost all English language journals – are primarily run by native speakers. Thus, other native speakers have a gatekeeper and network advantage, and a cultural ad vantage too. Arguably, editors and referees of these journals will be better persuaded by submitted papers that find their origins in the same scholarly environment. Moreover, it has been pointed out how contemporary debates taking place within 'international' journals are essentially internal to the Anglo American geographical community, and thus the discourse which allows access to these outlets is one that should be considered, to all extents and purposes, a national one. As a result, 'Other' traditions or, more simply, works originating in other countries run the risk of being considered irrelevant or incorrect or, at best, being regarded as interesting, 'exotic', or 'local' case studies incapable of producing innovative theory or framing the epistemological boundaries of disciplinary practices. In addition, few endeavors are still made by most English language aca demic institutions (journals and conference organizers) to assist non native speakers to cope with the linguistic gap.

What follows from all this is that scholars based in the leading Anglophone countries play an unchallenged role in the development of academic debate. However, these scholars are able to do so not just because they have a competitive edge in publishing in the 'relevant' journals, but also because they have the opportunity to set the research agenda. In fact, non native speakers are allowed to participate in this debate as long as they adhere to the conventions of the native speakers, cite the 'right' – that is, Anglo American – literature, and show the relevance of the argument of their 'Otherness' to the core of native speakers. While scholars who are non native speakers of English are supposed to keep up with the literature in their own language as well as in English, scholars who are native speakers of English not only frequently ignore the non English literature, they often neglect pieces from authors outside the UK and US even when they are written in English. Indirectly, this neglect has another consequence: non native speakers are required to intro duce the reader to the geographical 'context' of their research in great detail because so many UK and US based referees are unfamiliar with research that deals with another context than the Anglo American one, in cluding the research contexts of peripheral countries within the Anglophone world (Ireland, New Zealand, etc.) or even those of the more peripheral regions within the Anglo American countries themselves. Ironically, this has the unwanted effect of making the work of non US/UK writers more conventional and descriptive and, thus, of reproducing the existing division of labor be tween native and non native scholars as theorists on the one hand and providers of case studies on the other.

The fact that US and UK based scholars dominate international publications in social sciences and, even more visibly, in geography, is therefore a well known and also widely shared argument within the debate on Anglo American hegemony. In addition, this matters even more in a discipline like human geography, whose peculiar history and developing pathway has left much room, in many respects greater than in other social sciences, for national publications and, at the same time, also for the preservation of 'other' international academic circuits. At the same time, however, it can also be argued that the customary divide between an Anglo American geography and, for instance, a 'continental European geography' is blurred when one looks not only at the long term evo lution of the discipline (as proved by the enduring influ ence of French regional geography on Anglophone geography in many stages of the discipline's life) but also at the more recent critical developments in the discipline drawing on the same philosophical foundations, as a consequence of the large influence exerted today by the 'continental thought' on contemporary Anglo American geography (most notably, through the work of prominent thinkers such as Foucault, Derrida, Lefebvre and, in more recent times, Latour, and of other authors such as Badiou, Agamben, and Rancière). Beyond the existence of separate national traditions what then needs to be emphasized in order to capture the geography of knowledge production is the peculiar circulation of ideas and academic theories, characterized by relentlessly different timing and spacing, highly dependent on the laws and the conventions of the capitalist market in ideas and the other intellectual commodities.

It Is Not Only a Matter of Publishing: The Issue of Access

The power geometries in academic publishing are not only dependent on 'immaterial' factors such as language, discourse, academic power, and cultural traditions, but also on 'material' access to scientific knowledge and, most notably, to academic journals. Thus, even if we would be able to challenge the Anglo American hegemony with respect to language and the shaping of a more truly international discipline, we would still have to resolve the issue of access.

Some journals have already taken steps in this direction by making their full contents available on the Internet without charging readers for downloading papers. As suming a scholar has access to the Internet (and preferably through a fast connection), this is a large improvement. Recently, human geography has witnessed the establish ment of independent, freely accessible academic journals. These journals seek to disrupt Anglo American hegemony in international academic publishing not only by adopting a policy of free electronic accessibility, but also by ex plicitly opening up these journals to non native speakers and by taking a 'critical' stance in publishing and refer eeing procedures. In doing so, these outlets show their awareness of the intimate relationship between issues of 'material access' to scientific knowledge and more imma terial factors related to the production of discourse in academic work. Yet, to break down the geographical dis parities in access to journal more steps need to be taken, starting from already existing journals, in order to make them more easily accessible to scholars, especially to those working in less favored countries and regions. The debate that has developed around the issue of 'open access' can be illustrative in this regard.

As a point of departure, contributors to this debate generally recognize the fact that it is nearly impossible for a library to carry subscriptions to every single journal that 'may' be relevant. The Internet has in some ways been a liberating development in this regard, but in most cases it has not solved the crucial issue that withholds access: money. With some exceptions, such as the newly established *PLoS Biology*, the most successful journals have been well established publications that switched from exclusively hard copy formats to online access. Furthermore, while Internet access does not solve the accessibility problem for everyone, and certainly not in an equal way, as access to computers and the Internet can be a serious problem in some countries, the Internet seems to provide better opportunities for access than printed formats, especially when electronic access is free.

An 'open access' strategy would undoubtedly have beneficial effects on the community of scholars and other science practitioners, considering the fact that freely accessible articles are downloaded and cited two to six times more often than those that require subscription (as both Lawrence, and Harnad and Brody have demon strated) because they attract new readers, increase usage, maximize impact, and speed up scientific progress. There is also a more ethical reason for charging authors or *persuading publishers* to provide free access: most research is publicly funded and should thus be accessible, while dissemination, which is increasingly seen as an integral part of a research project, should be included in research costs. In short, the debate on access to academic journals shows how the goal of internationalizing human geog raphy must take into account the issue of access: access is not a sufficient condition for scientific progress, citation, or geographically equal opportunities, but it is a neces sary one. Currently a (r)evolution is taking place in scholarly communication and it seems that most human geographers are not in the forefront of these changes that would benefit them both as writers and as readers of academic papers.

What about the Excluded? Critical Stances on the Anglo-American Hegemony Thesis

Critics of the 'Anglo American hegemony' literature in human geography have built on a theoretical and polit ical position that can be defined as 'post structuralist'. Post structuralist thinking, and notably its radical ver sions, finds its origins in a variety of intertwined strands of research and reflection, including feminism, post colonial theory, and cultural studies, which have exerted growing influence during the past decades on established social science disciplines such as sociology, geography, and anthropology. The distinctive feature of this litera ture is to be found in the reflexive and critical character of its approach to the understanding of knowledge pro duction: knowledge is viewed as a situated process and practice, which is to be related to varying contextual, social, and political–institutional factors. The awareness of the uniqueness of the research context has therefore led scholars to place an increasing emphasis on the specific and differentiated positionalities of the re searcher in the undertaking of his/her research endeavors.

Albeit engaging with critical reflections on the posi tionality of the researcher, those stressing and debating the Anglo American hegemony in human geography have not attempted – according to some critics – to offer a more balanced analysis of the factors behind the uneven geographies of scientific production: not only an analysis of the 'external' factors (the hegemonic role of Anglo American countries) explaining the subaltern position of many countries in the knowledge production, but also an analysis of the 'internal' factors explaining such a sub altern position (the weaknesses and the contradictions of the national university systems in many countries) is central to the understanding of the contemporary un evenness of geographical knowledge production. Along these lines, the benefits of using English as a global *lingua franca* are presented, in contrast to those views that re gard English as a sheer factor of cultural homogenization or even domination. According to this view, having a *lingua franca* recognized as such by the majority of the scholarly community ensures a stronger level of cohesion and is indeed a preferable situation to one characterized by features of fragmentation and isolation of the various national 'traditions'. From this perspective, the use of a global *lingua francer* can be seen as a way to preserve geographical diversity and to enhance scholarly com munication at the international level under the current context of globalization.

This is frequently linked to arguments about the 'in ternal' factors undermining scholars' productivity outside the Anglo American world. In fact, while critics of the Anglo American hegemony have focused their attention on the unequal power relations existing between the 'center' (the Anglo American countries) and the 'per iphery' (more or less, albeit to differing degrees, the rest of the world), they have left aside from their reflections a critical consideration of the 'endogenous' contexts in which scholars operate outside the Anglo American countries: the either too 'precarious' or too static aca demic positions of researchers; the hierarchical power structures limiting the autonomy of researchers; and, probably most importantly, the lack of institutional in centives aimed at improving the research performances of university departments.

This critique of the Anglo American hegemony thesis expresses a point of view that can be considered 'liberal institutionalist' in the ways in which it points the attention on the institutional conditions preventing researchers from being competitive at the international level. Whatever political and theoretical orientation is behind this position, this critical contribution is of interest here because it touches on a crucial aspect in the Anglo America hegemony debate: the issues related to the governance of research production and, most notably, to the introduction of assessment procedures and audit cultures aimed at stimulating university systems to fi nesse and enhance research performance and, in general, to more competitive behaviors. While more rationalist accounts stress the potential changes following in the wake of possible transformations in the governance structures of the university systems, critical scholars ex press a concern that the growing importance of assess ment systems and other accountability procedures has consequences not only on the quantity but also on the quality, the meaning, and the substance, of scholarship. Attention is devoted, from this perspective, to the process of neoliberalization of higher education systems taking place in many non English speaking countries. The introduction of centralized or multilevel assessment systems and of the related governance structures, along the lines of the Research Assessment Exercises adopted in Britain since the mid 1980s, brings about profound changes in the ways in which research is undertaken by human geographers in an increasing number of countries: academic entrepreneurialism and publishing practices centered on the central role of impact factor journals are the distinguishing features of the changing university systems in Western Europe and many other countries inclined to adopt the neoliberal recipes of societal governance.

In light of these reflections, it can be argued that the earlier formulation of the Anglo American hegemony thesis is not sufficient to address the current develop ments taking place within the structures, the institutions, and the intellectual technologies governing social sci ences and the higher education systems within advanced liberal countries. The debate on this subject needs to take more closely into account the institutional factors underlying scholars' publishing practices and strategies. Beyond the Anglo America/rest of the world divide is the increasing neoliberalization of higher education sys tems and the multiple techniques of 'government at distance' affecting scholars' working practices in an ever larger number of countries that demand careful scrutiny by critical geographers today.

Conclusions

The debate about the Anglo American hegemony in human geography should be more concerned with the kind of vantage point to be adopted in the discussion. Issues related to the corporatization of the university system, to the increasing productivity being demanded to researchers, and to the commodification of academic labor have so far been analyzed mostly with reference to the UK and the US. From this perspective, the 'peripheries' have been seen either as mere importers of dynamics and processes which have been created elsewhere in the 'center' or as spaces being excluded from the globalization of social science and more specifically of geographical knowledge. The non English language geographer has somehow been confined to the role of 'subaltern, co herently with the postcolonial approach inspiring the critical literature about the Anglo American hegemony, while s/he has not been considered as a potentially em powered subject, pursuing autonomous strategies and practices of knowledge production.

Following from this, it can be noted how little is known at the moment about the ways in which geo graphical knowledge is produced and transmitted to the wider community of scholars and other readership out side the Anglophone world. In particular, we still know little about the ways in which non English language human geographers disseminate their scientific results and are already accustomed to deal with multilevel, 'international', publishing spaces. Today, within the context of globalization, features of unification and cul tural domination can coexist with those of plurality and heterogeneity giving shape to a more heterarchical geography of geographical knowledge than the one we know from existing accounts of human geographers' publishing practices. Thus, along with tendencies leading to homogenization of university systems at the world scale related for instance to the aforementioned adoption of research assessment procedures, there are still multiple academic subjects performing a host of situated practices and strategies of academic publishing addressed to the contemporary globalized world.

In this context, future research needs to be committed to deal with the lack of knowledge about the machineries of geographical knowledge production outside the Anglo American universities, including those that have the potential to challenge the Anglo American hegemony in human geography. Along these lines, forthcoming research efforts will then take into account the relationship existing between researchers' publication practices, the adoption of exogenous approaches to the governance of knowledge production, the consequent changing character of national policies and institutions, and the ways in which academic communities react and adapt themselves to these multi level institutional pressures.

See also: Citation Geography; Critical Geography; Cultural Turn; Hegemony; Language; Language and Research; Other/Otherness; Geographical Journals.

Further Reading

Aalbers, M. and Rossi, U. (2007). A coming community: Young geographers coping with multi tier academic spaces across Europe. *Social and Cultural Geography* 8, 283 302.
Berg, L. D. and Kearns, R. A. (1998). America unlimited. *Environment and Planning D: Society and Space* 16, 128 132.
Bourdieu, P. and Wacquant, L. (1999). On the cunning of imperialist reason. *Theory, Culture and Society* 16, 41 58.
Braun, B., Vaiou, D., Timar, J. and Minca, C. (2003). Guest editorials. *Environment and Planning D: Society and Space* 21, 131 168.
Castree, N., Aspinall, R., Berg, L. D. *et al*. (2006). Forum: Research assessment and the production of geographical knowledge. *Progress in Human Geography* 30, 747 782.
Garcia Ramon, M. D. (2003). Globalization and international geography: The questions of languages and scholarly traditions. *Progress in Human Geography* 27, 1 5.
Gutiérrez, J. and López Nieva, P. (2001). Are international journals of human geography really international? *Progress in Human Geography* 25, 53 69.
Harnad, S. and Brody, T. (2004). Comparing the impact of open access (OA) vs. non OA articles in the same journals. *D Lib Magazine* 10(6).
Kitchin, R. (2005). Disrupting and destabilizing Anglo American English language hegemony in geography. *Social and Cultural Geography* 6, 1 15.

Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature* 411, 521.
Minca, C. (2000). Venetian geographical praxis. *Environment and Planning D: Society and Space* 18, 285 289.
Paasi, A. (2005). Globalization, academic capitalism and the uneven geographies of international journal publishing spaces. *Environment and Planning A* 37, 769 789.
Rodríguez Pose, A. (2004). On English as a vehicle to preserve geographical diversity. *Progress in Human Geography* 28, 1 4.
Rodríguez Pose, A. (2006). Is there an 'Anglo American' domination in human geography? And, is it bad? *Environment and Planning A* 38, 603 610.
Strathern, M. (ed.) (2000). *Audit Cultures. Anthropological Studies in Accountability, Ethics, and the Academy*. London: Routledge.
Yeung, H. (2001). Redressing the geographical bias in social science knowledge. *Environment and Planning A* 33, 1 9.

Animal Geographies

C. Wilbert, Anglia Ruskin University, Chelmsford, UK

Glossary

Intra-Actions This is opposed to the everyday term of interactions, one that presupposes an *a priori* distinction between separate worlds of, for example, nature and culture and technologies, interacting in zones of contact. Intra-action refers to such natures, cultures, technologies as are always already mixed up in everyday practices.

Zoogeographies The mainly nineteenth-century approach to mapping the major regions of the world by their animal species.

Introduction

Animal geographies might, at first, be thought a misnomer by those coming across the term for the first time. Current geographies of animals are not *just* about animals; how could they be? Instead, current animal geographies seek to challenge entrenched assumptions about animals in modern societies and in the historical developments of humanities disciplines. Recent animal geographies have engaged in three overlapping themes that, to some extent, reflect differing political and theoretical leanings. First, they have focused on identities and meanings. That is of the ways certain human identities have been constructed in relation to certain animals in terms of practices such as eating, working, or hunting that stress social categories such as class, race, gender, and religion as important in construction of identities in relation to some animals. Second, studies have focused on human–animal intra actions in terms of the political economy or political ecology of animal productions in agriculture, or struggles over resources in the developing world, but increasingly, in developed regions as well. Third, animal studies have been encouraged by developments in social theory that acknowledge a wider range of actors than just (some) people. Such approaches have pointed to very complex geographies of intra action between people and animals in ways that questions and rethinks the human centeredness of human geography. Such a perspective is allied to certain aspects of posthumanism, but also to nonrepresentational theory, and especially actor network theory (ANT), though it seeks to further develop the spatialities, politics, and ethics of human–nonhuman associations.

Histories of Animal Geographies

Animals have long had a presence in geography as a discipline. In the modern period, two closely related approaches characterized a distinct study of animals in geography. The first became known as zoogeography and sought to map evolutionary distributions of animals and was closely allied to what was to become a more physical geographical approach, allied with zoology and the (then) emerging disciplines of ecology and biogeography. The other related approach focused more on animal domestication as part of studies of the Neolithic agricultural revolution and of distributions and diffusions of domesticated animals. The latter was referred to as cultural geography but overlapped with cultural ecology (and human ecology in some studies) and became closely associated with archaeology and historical anthropology.

Zoogeographies

What became known as zoogeography emerged at a time of unheralded exploration, colonial conquest and Empires, collections of animal and plant specimens to be returned to nations such as Britain and France, trophy hunting as an elite leisure pursuit, alongside, and inimical to, the development of scientific disciplines such as biology, geography, and zoology. As such zoogeography, or regional biogeography, should be viewed as emerging out of distinctly geopolitical concerns. A figure such as P L Sclater (for much of his life the Secretary of the London Zoological Society) can be seen as an early and important cartographer of animal distributions from the many specimens he was sent to catalog and classify (writing over 1000 scientific and popular tracts) which were provided by freelance entrepreneurs, naval and military personnel, and those employed by institutions to collect from around the colonized world. For some 70 years before Darwin's *Origin of Species*, those who worked with animals, plants, or fossils developed theories to explain patterns, dispersals, and habitats in relation to collected species. P L Sclater's 1858 paper on animal distributions (based on avian samples) was seen by many of his contemporaries, most notably Alfred Russell Wallace, as developing a new approach to animal geographies. In this work, Sclater set to map out the main geographical regions of the Earth according to how animals arranged themselves across the Earth's surface, or at subregional scales to establish patterns of spatial covariation between animals and other environmental factors. These regions were much debated and fine tuned by Wallace and J A Allen, amongst others, and still retain some provenance today, having become known as the Sclater–Wallace regions. Few geographers have actually continued to work in this field, or even with zoo geographical studies in biogeography, these now being more associated with biological and ecological discip lines. However, the taxidermic specimens that emerged from this time of exploration and hunting are now be coming the focus of new animal geographies in terms of developing their biographies in another aspect of animal geographies.

Early Cultural Geographies of Animals

In contrast to zoogeographical work, early twentieth century geographies of animals sought to examine how some animal distributions were affected by human groups through animal domestication and introductions of animals into new regions. Studies of animal domesti cation were rejuvenated by evolutionary theory in the late nineteenth and twentieth centuries – viewing animal domestication as one of the main qualities and practices that distinguished 'man' from animals. It is, of course, unsurprising that animal and plant domestication have had such a central interest in human geographies. Agri culture, in its many varieties, but particularly, modern capitalist industrial agriculture, has, and continues to exhibit enormous scope for transforming landscapes, the fleshy bodies of animals, and the socioecological relations between people and animals in myriad ways. However, it is the origins of domestication that were dominant in early geographies, pioneered by geographers such as Friedrich Ratzel and Eduard Hahn in the 1880s and 1890s, though before this many other studies of do mestication were undertaken focusing on sequences of human development. Much of this early work on the origins of animal domestication can be seen to dovetail with what came to be called cultural ecology.

But it was Carl Sauer, with his work in the 1950s, who became most associated with what became known as cultural geography or cultural ecology approaches to agricultural origins, dispersals, and how humans influ ence animal numbers and distributions. Here the atten tion was focused on place, region, and, above all, the cultural landscape. Whilst animals were not central to Sauer's work, he argued for 11 main centers or 'hearths' of animal and plant domestication and for religious or ritualistic motivations for animal domestication. Sauerian cultural geography has been dominated by the view that domestication of plants necessarily preceded that of animals, bringing about more leisure and storable abun dance than hunter gathering. This, in turn, was seen to give rise to sacrificial religions where animals were kept for ritual reasons as sacrifices to help secure this new abundance and its security. Here, it was argued that only subsequently did 'man' begin experimenting with breeding domesticated animals as a food source.

Debate and controversy has long accompanied both cultural ecology and Sauerian cultural geography. Ac cording to some, this work failed to adequately address ethnographic work that contradicted the view that plant tending gave rise to more abundance and leisure time than hunter gathering. Within the cultural ecology tra dition others have recently re investigated empirical evidence to argue that materialist economic reasons were more powerful motivations for animal domestication across the regions seen as the main hearths of domesti cation. Still others have argued that the cultural evo lutionism of such approaches does not adequately historicize the politics of animal domestication and spe cies alteration, or the variable and changing ways that notions of animality have been construed to inform no tions of the human. Instead, for cultural ecology/cultural geography, domestication was viewed as a factor all humans share, helping to make them distinct, through culture, from animals. As such, human–animal dis tinctions were seen as evolutionary and naturalized, with a tendency to view domestic animals as, somewhat pas sive, human enacted transformers of landscapes through a superorganic notion of culture that appeared to func tion independently of human individuals. Some cultural ecological studies went further than this; for example, those on 'transitional species' that lie between the wild and domestic, that emphasize qualities of animals that have made them seemingly 'uncooperative' to domesti cation and hence, alluding to a possibility of nonhuman agency. Yet, this latter point remained largely un developed in this body of work.

New Animal Geographies

Many of the renewed animal geographies that emerged in the mid 1990s sought to put some distance between themselves and cultural ecology and Sauer's cultural geography, though some writers do point to a few con tinuities with that tradition. Such work emerged out of the new cultural geographies that took more materialist approaches to cultural interactions, post structuralist theories that decentered human subjects (therefore, stressing possibilities of other than human agency), as well as environmentalist and eco feminist developments in ethics and politics.

These renewed animal geographies take a wide var iety of forms. Nearly all new approaches retain the strong empirical focus found in earlier cultural geographies but with a focus on more posthumanist ontologies and epistemologies, geographically dispersed notions of ethics, and more innovative and politically engaged methodologies.

Theoretical engagements in this renewal of animal geographies have also been varied. To simplify, we might divide these into sociocultural approaches that focus on identities, class, gender, and race; political economic approaches, and more vitalist inflected more than human approaches. However, the boundaries between these theoretical and empirical approaches are un surprisingly porous as animal geographies develop. As such, they are divided here simply as a rough guide to this field of work.

Sociocultural Animal Geographies: Identities and Place

Much of the focus of sociocultural animal geographies has been on animal roles in construction and ordering of culture and individual human subjects, and wider links between human and animal identities in particular times and places. Some animal geographies have experi mentally approached animals as something akin to a marginal social group that emphasizes unequal power relations between animals and people, and then also between differing groups of people. Historically, in London for example, farmers, or drovers who brought animals into the city to be sold and slaughtered, in the nineteenth century, came to be defined as more animal istic in terms of having an inclination to wildness, de viancy, and sexual excess because of their closeness to cattle and other animals. The animal practices revolving around live – meat markets and slaughterhouses clashed with new urban identities associated with standards of civility, public decency, and norms of compassion and sexual propriety. This ultimately led to both animals and drovers being excluded from the city.

Such work also suggested that a continuum of inclu sions and exclusions of certain animals in, or of, certain spaces can be found, with, for example, pets currently seen as inclusive members of some cities, homes, and families, and other animals such as larger carnivores as excluded to the periphery of human lived spaces – with many points in between these. Yet, many examples of animal practices do not fit easily into such a range, and it is a range complicated by social, cultural, and historical factors of diverse forms.

Other sociocultural animal geographies informed by feminism and environmental politics have dwelt on links between race, gender, class, and certain animal practices in modern or late modern spaces. Much of the focus here has been on re conceptualizing the human–animal divide from one of oppositional dualism into networks of in tricate interdependencies focused on kinship. In terms of connections between race and representations of animality, empirical work with different ethnic groups in American cities showed that attitudes to animals are complex, but also that certain animal food practices are, and have long been, used to other, and often to demonize, certain ethnic groups in particular societies.

Such work has to a great extent been experimental, theoretically and methodologically, in order to search for new ways of bringing animals back in to social theory and practice. Yet, just how far the marginal group approach has useful durability, rather than becoming a standard social explanation of certain situations, is open to some debate. Moreover, there may be residual humanist as pects to this work that emphasizes social categories, though this is a wider difficulty of social and cultural animal geographies.

To some extent, the role of animals in the making of places and landscapes can be incorporated in this section, though it does bleed out into other approaches. Such work has sought to reverse the view of earlier cultural ecology by demonstrating the active roles of animals in helping to produce certain landscapes. For example, work on the early twentieth century English Broadlands demonstrated how some resident animals of these land scapes helped define relational human identities. This work has variously showed how differing technologies and corporeal expert practices – looking, touching, lis tening, testing, hunting, killing, tasting animals – pro duced differing claims to authority concerning local conservation and right ways of behavior towards animals of these landscapes. Similarly, work on whale and dolphin watching has pointed to how such eventful spaces work as landscapes of performance that are co constituted be tween tourists, technologies of envisioning, sea life, and wider processes such as weather conditions. These ex cursions may not always be successful – whales may not appear, weather may be bad, or other factors may intervene. When they are successful, these performances are seen to be, often, highly emotional for tourists and are discussed by them in ways that emphasize the transfor mative capacities of the experience.

Political-Economic Animal Geographies

Animal geographies that have utilized political economic analyses have mainly been found in rural and agricultural geographies. Such studies have at times focused on the development of certain intensive food production re gimes as emblematic of current structures of capitalism. For example, studies have focused on chicken and hog intensive production systems in the USA, amongst other matters. Here, animals were always seemingly in the background being acted upon, having their bodies transformed to suit the needs of these regimes, but always part of much wider processes of power and constant economic restructuring of production. However, the biological being of these animals were also seen to set certain limits to what could be done within these systems of production.

The advantage of such approaches is that they focused on power, on one of the main ways in which animals are engaged by humans – in food production – and that they analyzed meat production in terms of transport, feed and pharmaceutical interventions, techniques, technologies, and labor processes involved in slaughter, and other factors besides these. Arguably, this has given quite rounded approaches to intensive animal production, though one still with very humanistic concerns. Yet, it might be argued that the animals have also been some what lost in this whole process. But then that may per haps be a good reflection on where such animals are, that is, lost in often gigantic systems and practices of power and technological systems being focused upon them with only their biological basis as increasingly genetically uniform beings creating some limits to these powers, and with only fears of, or actual, diseases creating some interruption in the speeded up life processes of agro industrial production.

Similar political ecology approaches to animals have emerged more directly out of cultural ecology and focus on geographic and historically specific transformations of natures under capitalism. Such work emphasizes the ac tive roles played by subjects (including animals) in the ongoing fluid processes of landscape changes that have, for example, accompanied introductions of animals into colonized regions. The work here is somewhat new and emergent, though specific animals do tend to be treated again as already constituted entities, although capable through their biological makeup of affecting extractive or productive sectors of economies. Similarly, animals understandably tend to be grouped together as, for ex ample, 'fisheries' as contested resources struggled over within changing political ecological practices.

Arguably, a residual humanism is retained in many political economic and political ecological works. Moreover, the somewhat overarching analysis of some work has led to moves to develop theoretical and em pirical approaches that appreciates, in more detailed ways, the heterogeneous character of social life, but also the roles of nonhumans, as well as the many people and technologies, in ordering and disordering seemingly overarching practices that political economic approaches can take.

More-than-Human, or Hybrid, Animal Geographies

Other animal geographies have, in recent years, sought to problematize the use of orthodox social categories in their theoretical and empirical approaches to the spati alities of human – nonhuman relations, encounters, and co productions of spaces. This includes ditching many aspects of political economy approaches whilst holding on to a focus on the wider networks in which animals are engaged and enacted. Here the focus has been on recognizing the human of human geography and to more fully accommodate the nonhuman (including animals) in the innermost fabric of societies – by developing hybrid, or more than human, geographies that focus on the liv ing spaces and bodily entanglements of associational lives.

Here, ANT has been most inspirational to more than human geographies. Such work refuses to see animals (and humans) as centered preexisting beings, thereby, refusing *a priori* distinctions of separate worlds such as nature/culture or human/animals. As such, it steps aside from debates around animal rights on notions such as sentience and an extension of rights, yet, also seeks to grant a positive sense of the existence of animals but not by means of attributing human qualities to them. Instead, these approaches seek a nonanthropocentric animal geography that emphasizes how the social is populated and co constituted by myriad nonhumans, importantly including animals, in which these other entities do things in the particular configurations they are engaged, enac ted, or resistant, or indeed do other things. As such, much of this work has sought to develop an ethics and politics that take seriously the entangled imbroglios of hybrid geographies or human–animal intra actions, and chal lenges policy, politics that sees nature, environment, or animals as matters of fixed identities.

This sense of seemingly stripping animals of any in herent qualities or agency has raised some deep concerns. Yet, it may be argued that more than human geo graphical work seeks to de essentialize a notion of the animal species as if *all* animals share exactly the same characteristics in all places. Instead, it may be seen that in terms of the material assemblages they are enacted within that a race horse is radically different from a draft horse that pulls a plough or cart, or an elephant raised in a zoo is very different from the wilder elephants studied by conservationists in a National Park in Kenya. Instead, it sees that there are varied relational forms of elephant ness or horse ness that are made and remade within complex configurations. Animals might be seen to have some 'inherent' characteristics or affordances in terms of the ways they may be enacted (therefore making some links with older cultural geographies in terms of transi tional species). But such characteristics are treated as established by prior ordering and disordering material practices (that always involve animals as, in some ways, co producers of these material knowledges, though in often highly unequal ways) that are subject to change, though these can also be very robust. Therefore, the centered, known animals (as species) we have come to know through television documentaries, natural history guides, zoos, theme parks, and more, are rendered as themselves in need of explanation as to how they come to be dominant, generalized views of how certain animals are known. As such, these animal geographies seem to point to how animal's characteristics differ according to the spatialities of their engagements or participations within certain assemblages.

As in the more humanist political economy and pol itical ecology approaches, some claim that the cor poreality of animal bodily doings seem to be lost among a miasma of stretched out subjects made and remade in complex intra actions with myriad other things in these hybrid geographies. Though these approaches are reso lutely material, there is some concern that the fleshiness of animal bodies can easily be lost in the haze of entangle ments invoked. This might be partly due to the theoretical bent of some work, but may partly be addressed by a deeper empirical focus on animals. However, more prob lems surround the move to evoke the corporeal aspects of relational animal lives in work that comes out of all of these theoretical approaches to animal geographies. In this, animal geographies will need to continually develop not just more experimental methodologies, but also more innovative narrative approaches to the study of animal–human intra actions.

Conclusions and Futures

The theoretical and empirical approaches discussed here are not the only ones on the block in terms of animal geographies. Moreover, the three generalized approaches outlined here in terms of new animal geographies are not wholly separable as they have been developed and engaged together and against each other in significant ways. Yet, a division between sociocultural studies that focus more on identities and culture and those that seek to develop more radically posthumanistic, lively, accounts of intra active worlds of encounter where subjectivities are dispersed and not seen to adhere in bodies and minds, can be found in this emerging literature. As such, animal geographies are party to the continuing displacements of the human centeredness of human geography. Yet, animal geographies have many places still to go and much to do in developing new politics and geographical ethics. As yet, few studies have been done on less visible animals, such as insects, that are involved in human lives in myriad important ways.

See also: Actor-Network Theory/Network Geographies; Cultural Geography; Non-Representational Theory/Non-Representational Geographies; Posthumanism/Posthumanistic Geographies.

Further Reading

Anderson, K. (1997). A walk on the wild side: A critical geography of domestication. *Progress in Human Geography* 21(4), 463 485.
Bingham, N. (2006). Bees, butterflies, and bacteria: Biotechnology and the politics of non human friendship. *Environment and Planning A* 38, 483 498.
Boyd, W. and Watts, M. (1997). Agro industrial just in time: The chicken industry and postwar American capitalism. In Goodman, D. & Watts, M. (eds.) *Globalising food*, pp 192 225. London: Routledge.
Buller, H. (2004). Where the wild things are: The evolving iconography of rural fauna. *Journal of Rural Studies* 20, 131 141.
Cloke, P. (2005). Cetacean performance and tourism in Kaikoura, New Zealand. *Environment and Planning D: Society & Space* 23, 903 924.
Jones, O. (2004). 'The restraint of beasts': Rurality. Animality, actor network theory and dwelling. In Cloke, P. (ed.) *Country visions*, pp 450 487. London: Pearson.
Philo, C. and Wilbert, C. (eds.) (2000). *Animal spaces, beastly places*. London: Routledge.
Philo, C. and Wolch, J. (eds.) (1998). Special theme issue: Animals and geography. Society & Animals 6(2).
Sauer, C. (1952). *Agricultural origins and dispersals*. New York: American Geographical Society.
Whatmore, S. (2002). *Hybrid geographies: Natures cultures spaces*. London: Sage.
Wilbert, C. (2006). What is doing the killing? In The Animal Studies Group (eds.) *Killing animals*, pp 30 49. Chicago: University of Illinois Press.
Wolch, J. and Emel, J. (eds.) (1998). *Animal geographies*. Verso: London.

Relevant Websites

http://www.wku.edu/%7Esmithch/biogeog
Early classics in Biogeography, distribution and diversity studies: to 1950.

Animal Welfare, Agricultural

H. Buller, University of Exeter, Exeter, UK

Glossary

Anthropomorphism The representation or treatment of animals as if they shared human appearance, characteristics, and/or behavior.
Ethics A system of accepted beliefs which have an impact upon human behavior.
Farm Animal An animal kept to produce food.
Welfare State of physical and mental health.

Introduction

Though arguably not a traditional area of geographical concern, agricultural animal welfare is currently at tracting the interest of a growing number of human geographers. The (humanist) social sciences in general have long ignored farm animals, and, one might even claim, all animals, rendering them largely invisible and their agency unaccounted for in the analysis of human society. Yet, geography's relatively recent reinterrogation of society–nature relations, drawing upon cultural and post structural theory in particular, has pushed open the gates, at least in part, and has, to use a phrase of Jennifer Wolch and Jody Emel's, sought to bring the animals back in. Today, animal geographies are proving to be a flour ishing arena for the investigation of interspecies rela tionality, symbolic co constitution, and posthuman, nonhuman, or 'more than human' ontologies, to name but a few, drawing on science and technology studies and such innovative conceptual approaches as actor network theory, hybridity, and dwelling. Agricultural animal welfare, as an issue and a concern that is defined both by science and by society within a set of very distinct and overt, but nevertheless fluid and heterogeneous, socio natural boundaries (from farmyards and attenuated lives to notions of cruelty and 'humane' killing), sits at a crucial interface between these theoretical concerns and what has been often distinct geographies of care and ethics. For geographers, the current popular, scientific, and political engagement with farm animal welfare is far more than merely the latest rung on the ladder of con temporary society's ethical ascendancy. Critically, it challenges not only the representation of animals as sets of scientific 'facts' or as corporal commodities with their resultant invisible subjectivity, but also the mechanisms by which the welfare of these bodies and sentiencies are assessed and acted upon, through the multiplicity of sociomaterial practices, by a widening array of different actors from veterinary scientists and farmers to retailers and food consumers.

The Difficulty of Animal Welfare

The notion of farm animal welfare is directly linked, first, to the status of farm animals as chattels and commodities, belonging to and traded by farmers and others as a source of revenue and, second, to the fact that their lives, from birth and before, to death, are almost entirely controlled, adapted, and modified by human actors for the purposes of consuming them and the products derived from their bodies. The welfare of agricultural animals is therefore bounded by the economics of production, by the de mands of human metabolism, and by the inevitable interruption of their lives. For however much we might seek, for ethical reasons, to improve their quality of life, that life will be constrained and violently ended by processes entirely out of their control. Farm animal welfare is, by definition, a moral and ethical compromise, yet one with which a great many people are entirely comfortable.

As a concern, farm animal welfare is immediately distinguishable from animal liberation. The latter, as championed by writers such as Singer and Regan, seeks to free animals from all human induced suffering, including husbandry and slaughter, on the grounds that to inflict such suffering is to engage in an oppressive (and un necessary) anthropocentric specisism that exploits animals for human use and denies them any subjective existence or rights. Farm animal welfare, by contrast, largely accepts human consumption of animals and animal products, and the constraints this imposes on animal freedoms, but strives to minimize the suffering that this entails.

The inevitable anthropocentrism inherent in animal husbandry places farm animals themselves in an equivocal position. They are the archetypical hybrids lying between, on the one hand, the naturality of their evident categorization as nonhuman animals and, on the other, the artificiality of their domestication, their breeding, their genetic modification, and, ultimately, their lives and deaths. Like pets and companion animals, they are the acknowledged and vital animalian co constituents of human worlds, yet unlike pets and domestic animals, their existence is defined by the fact that they are there to be materially consumed. For this reason, certain rad ical environmentalist thinkers like Baird Callicott, for whom the individualism of welfarism conflicts with the bio collectivism of ecocentrism, have likened farm animals to pieces of furniture, entirely human constructions with little or no subjective legitimacy as occupants of the 'natural' world. Indeed, were they to be 'liberated' and humans to stop eating them, it is often claimed that large numbers of farm animal species might disappear altogether, incapable of surviving and reproducing without human agency in any newly naturalized environment.

Human concern for the welfare of farm animals is thus a complex, often incoherent and similarly hybrid, engagement that draws upon and juxtaposes an ethic of care for our dependent animalian *confrères* with their physical exploitation as economic resources; a humanist utilitarianism that values humans above animals with a moral absolutism that nevertheless unequivocally rejects the deliberate torturing of animals; and an affective relationality that acknowledges their co constitutive vitality with a readiness to consume them.

As a normative concept, farm animal welfare suffers from three major inconveniences. First, farm animals cannot verbally communicate their discomfort or their suffering to human beings. Neither can they demand improved conditions. Second, farm animal welfare is less an absolute than a relative concept. Not only is there an inherent problem of definition (what 'is' farm animal welfare and how does it differ from, say, cruelty?) but also the determination and measurement as 'good' or 'bad' welfare is highly contingent upon a wide range of material, epistemological, cultural, ethical, esthetic, and other circumstances which imply both a moral and a scientific commitment of those engaged in doing so. Third, notions of animal welfare are frequently captured by anthropomorphic interpretations of animal suffering or animal happiness that may, despite their widespread popular appeal, have either little basis in scientific assessment or little relevance to the actual physical and/or mental state of the animals concerned. These three inconveniences translate into three practical issues: how to identify and recognize animal suffering; how to establish parameters and indicators of good and 'bad' welfare; and how to reconcile scientific objective and cultural subjective differences in approaches to welfare and its achievement.

Because farm animals are a source not only of food, clothing, and any number of other by products, but also of wealth, status, and tradable assets, their welfare has always been an important component of animal husbandry. As such, it is most clearly associated with animal productivity, and hence value. This functional and essentially mechanistic approach to welfare, whose principal indicators have thus been both physiological and input based, has been the indispensable accompaniment to modernism. Indeed, it has been designed as such through breeding and selection procedures that ensure that farm animals are fit solely for the productive purpose to which they have been ascribed. An efficient animal will convert feed to output at a desired rate, in desired volumes, and at a desired cost, according to their genetic potential and their state of health. Farm animal welfare thus becomes the fulcrum point between sustaining animal capacity on the one side and animal productivity on the other, establishing the minimal conditions to ensure, or to augment, productivity and economic viability.

The apogee of this mechanistic and functional approach to farm animals and their welfare has been the 'factory farming' techniques that became widespread in the second half of the last century and are still the norm for much farm animal production. Enshrining the Cartesian notion of animals as intricate machines and 'natural automata' devoid of any state of conscious mind, intensive livestock farming at a massive scale has perfected the de animalization and desubjectivization of cows, pigs, and poultry, to the point where, in many cases, their lives have become both individually irrelevant (as acceptable wastage rates increase) and so spatially and temporally compressed that their productivity limits are no longer seen in (whole) animal life terms but rather as a series of somatically interconnected material events. Broiler chickens, pigs, and beef cattle, for example, are designed to reach their slaughter weight in increasingly shorter times irrespective of the longer term (and thereby irrelevant for they have no 'longer term') health problems such accelerated growth might engender. Dairy cattle are commonly inseminated so as to produce calves, and thereby maintain lactation, at a rhythm of around one per year. As such, their 'natural' life span of over 20 years is reduced to around 5 years, much of which is spent in calf.

Without a doubt, it has been the development of factory farming techniques that has prompted and stimulated the growing scientific, political, and popular concern for farm animal welfare that has emerged over the last 50 or so years. The publication of Ruth Harrison's book *Animal Machines* in 1964 is often held to be a landmark in that growth. Harrison purposefully moved the focus away from the more traditional areas of legislative control over farm animal welfare, such as deliberate mistreatment and cruelty, to the more mundane but far more prevalent welfare issues associated with animal confinement and 'factory' farming methods. The essence of her book was to reject the notion, implicit in the rationale of factory farming, that farm animals are simply machines. On the contrary, they are conscious beings, capable of experiencing suffering and deprivation and to whom humankind owes an ethic of care.

Redefining Animal Welfare

It is at this point that the conceptualization of animal welfare shifts from a purely instrumentalist notion, where welfare is equated for all intents and purposes with productive efficiency, to a re animalized notion based upon the physical health needs of the individual (in cluding hunger, acute metabolic disease, discomfort, chronic pain, and exhaustion), its mental well being, and its natural behavior. The Brambell Committee, set up by the British government in 1965, partly in response to Harrison's book and the debate it engendered, advocated the establishment of an independent farm animal welfare advisory body in Britain (set up in 1967) and the de velopment of a set of broad farm animal welfare prin ciples. Farm animals, they argued, should have the freedom to stand up, lie down, turn around, and under take natural activities, such as grooming and stretch their limbs. These were later recast by the newly established Farm Animal Welfare Council in Great Britain (founded in 1979) as the Five Freedoms: freedom from hunger and thirst; freedom from discomfort; freedom from pain, in jury, or disease; freedom to express normal behavior; and freedom from fear and distress.

The Five Freedoms, which have since become widely adopted internationally, represent a set of ideal (some might say unrealistic) states of welfare among farm ani mals. Critically, they combine resource based aspects of welfare (e.g., room to move) with the more innovative and complex concept of natural behavior (such as contact with other members of the same species). An animal benefiting from such freedoms may be said to be enjoying good welfare. Much current farm animal welfare legis lation takes them on board. In Great Britain, the Agri cultural (Miscellaneous Provisions) Act of 1968 made it an offence for those engaged in animal farming to cause or allow unnecessary pain or unnecessary distress, whe ther by commission or omission. The Act did not lay down any statutory requirements of minimum welfare provision, nor did it seek formal definition of unnecessary pain or distress, preferring to avoid the costs and impli cations of compliance monitoring as well as giving farmers a degree of interpretive freedom. It was, none theless, accompanied, in 1971, by a series of Codes of Recommendation which farmers were recommended to follow. That Act has since been superseded both by subsequent British legislation and by European Union legislation. Offering a broad ethical framework, the Animal Welfare Protocol of the European Union's 1997 Amsterdam Treaty specifically acknowledges the sen tience of animals, charging member states with the re spect and protection of the welfare of animals. Today, within the UK, the member states of the European Union and many other countries, there is a significant array of farm animal welfare legislation, setting specific physical requirements such as minimum cage sizes for battery hens or stall sizes for calves, outlawing certain practices such as tethering and tail docking (in certain countries), insisting on management practices such as regular access to fresh air and evoking a general duty of care toward farm animals whether on the farm, at the market place, or in the slaughterhouse. According to many animal welfare organizations, there nevertheless remains significant gaps in the general reach of farm animal welfare legislation, and in levels of compliance, made all the more visible as a result of the numerous disease outbreaks in animal farming over the last decade.

Despite this growth in regulation and legislation, or indeed because of it, the definition and monitoring of farm animal welfare remains a complex and contested issue. Three complementary approaches might be iden tified, each of which occupies a fairly distinctive ethical and ontological ground. The first takes as its focus the animal's state and is often associated with the work of people like the veterinary scientist Donald Broom for whom welfare is defined as the physical and mental state of an animal as it seeks to cope up with its environment and meet its needs. Failure to cope, or merely difficulty in coping, can generate stress, injury, illness, and suffering, thereby amounting to negative welfare. Crucially for Broom and others, these can all be measured and assessed scientifically, placing animal welfare assessment firmly in the camp of the natural sciences. This, for Broom, makes the assessment of welfare independent of moral con siderations. Ethical judgments only become relevant when considering what actions are to be undertaken where welfare is deemed insufficient.

The second approach, sometimes referred to as the 'feelings' approach, is associated with the work of Marion Dawkins, Ian Duncan, and others and takes the sentience of farm animals as its starting point. It argues that, as sentient beings, their welfare is defined by subjective feelings, from pleasure to suffering, and that these feel ings guide actions. Good welfare is allied with positive feelings, poor welfare, or suffering, with negative emo tional conditions, themselves caused by an animal's in ability either to achieve a desired state or to escape from an undesired one. Dawkins herself has led the way in identifying animal motivation through animal preference testing, while Temple Grandin in the United States had championed innovative approaches to improving welfare founded upon an innate understanding of the feelings of farm animals. Here the ethical position is slightly dif ferent. Although the 'feelings' approach is similarly em bedded in experimental science, its acknowledgment of animal sentience, and hence preference (and even, for some ethologists, emotion), raises the ethical stakes considerably.

A third approach takes a slightly different perspective in that it seeks to reemphasize the naturality of farm animals, their animalhood, and autonomy. Welfare is as sociated with the extent to which they perform, not as beasts of burden and farm animals, but as cows, sheep, pigs, and so on, exhibiting 'natural' behavior and having access, as far as is practically feasible within the context of contemporary agriculture, to 'natural' environments. As writers such as Appleby and Fraser have demon strated, there are clear contradictions between certain forms of normal behavior and the requirements of animal confinement. Yet, animals that are designed to walk should be able to walk. Animals that forage in the wild should be allowed to forage in domesticated contexts. The belief that farm animals have had their naturality bred out of them has been described as one of the great misconceptions of animal welfarism. The animal rights philosopher Bernard Rollin advocates a consensual and communautarianist ethic toward animals which trans lates, in welfare terms, to allowing them to function and live according to their distinctive animalian natures or 'telos'. Although many people have reservations about both the concept of natural behavior (particularly its ready embrace of ontological division between 'nature' and 'culture') and its pursuit in a farming context, the 'naturalness' of farm animal lives is increasingly emerging both as a criteria of welfare assessment and as a selling point within quality food chains.

One consequence of these evolving interrogations as to what animal welfare is and how to measure it has been the gradual move away from the more immediate, post Brambell concern of relieving evident animal suffering toward the active improvement of farm animals' quality of life. This has entailed, in theory at least, a shift toward more animal centered methods of welfare assessment and improvement. Of course, most codes of practice and evaluation procedures still concentrate either on resource provision (following on from the Five Freedoms) or on the interdiction of certain practices in certain countries (such as tail docking, beak trimming, and castration without anesthetic). As such, they remain at least partly anchored in the physiological and productivist tradition. The growing call for animal based measures, from wel fare scientists and animal welfare organizations, offers a different standpoint by assessing the animals themselves, their behavior, any injuries they may have sustained, their responses, and their general quality of life. To date, governments, certification bodies, farming lobbies, and other organizations have generally shied away from the extensive use of animal based assessments on the grounds that they are more costly, more time consuming, and less practical than production and resource based assess ments. Moreover, the scientific basis for certain assess ments is not always considered as robust as for resource measures. For some commentators, this reluctance also testifies to an unacknowledged acceptance that such as sessments would reveal generally lower levels of animal welfare than mere compliance to resource provision re quirements would suggest. A recent attempt by the European Union to introduce animal based assessment protocols relating to the foot pad health and mortality levels of the EU's annual production of around 4 billion meat chickens as part of the proposed Broiler Directive was hotly disputed and eventually withdrawn prior to the directive's approval in 2007. Nevertheless, the recent European Union Action Plan on Animal Welfare seeks to shift the regulatory emphasis increasingly toward animal based outcomes, and there are, at the present time, a number of experimental animal based assessment protocols currently being developed and tested. Whether they will ultimately replace resource based measures is a moot point. Though one obvious reason for their de velopment is that with growing regulatory harmoniza tion, not only within regional regimes like the European Union but also at the global level, as demonstrated by the World Trade Organization's increasing attention to the trade implications of differential animal welfare rules, it becomes increasingly difficult to set fixed resource based standards across widely varying geographical and en vironmental contexts. Time spent on outdoor pastures, for example, may be a welfare benefit in areas of tem perate climate but become a distinct disadvantage in areas of extreme heat or cold.

A Shift of Responsibility

Agricultural modernization arguably took the issue of animal welfare out of the traditional moral realm of stewardship and into the domain of science. Empiricism and the scientific method, under the paradigm of agri cultural progress, furnished animal science with a refuge from overt and explicit ethical engagement. Although the contested anthropocentrism of Enlightenment thinking contested the moral acceptability of the cruel treatment of animals, largely as a means of containing socialized violence, and led ultimately to the foundation of bodies such as the Society for the Prevention of Cruelty to Animals in 1824, growing awareness of the potential for animal suffering was, in practice, soon superseded by a technological mastery over animal lives and an economic rationalization of animal bodies in which animal science was entirely complicit and society, largely unquestioning. With the growth of farm animal welfare concerns, be ginning in the latter decades of the last century, three things have happened to alter the prevailing balance of responsibility.

First, there has been mounting contestation of the natural sciences' long standing claims of ethical neu trality in matters of animal welfare definition and as sessment. This is partly evident in the growing number of committed farm animal scientists that have become ac tively engaged in promoting greater welfare for farm animals. With the gradual acceptance by animal scientists of animal sentience as a critical element in the under standing of welfare, this contestation has extended to the natural sciences' quasi exclusive claim to 'speak' for nonhuman animals. Ethologists and philosophers engaged in the welfare issue, such as Tannenbaum, Sandoe, and Rollin, have all called for greater dialog between social and natural sciences, arguing, first, that the normative determination of levels of animal welfare implicitly requires value judgments and, second, that attempts to identify and act upon the subjective experiences of animals necessitates an interpretive approach that extends beyond the empirical realm. More recently, researchers drawing on science and technology studies have begun to expand the repertoire of social and natural science dialog in the field of animal welfare by bringing animal bodies, scientists, instruments, farmers, and other material and semiotic elements together in actor networks, arguing that farm animal welfare is, above all, simultaneously socially and materially constructed, while the multiple ontologies surrounding welfare itself are inherently political.

Second, there has been a significant growth in popular awareness of, and concern for, the welfare of farm animals in many countries. There are now a large number of animal welfare organizations operating all over the world, for example, the Royal Society for the Prevention of Cruelty to Animals in the UK; the international body, Compassion in World Farming, the Humane Society, and the Humane Farming Association in the US. Certain husbandry practices have led to militant action and sustained popular protest, for example, the transport of veal calves in Great Britain and, more generally, the use of veal crates across the European Union. Others have been the subject of mounting popular objection (such as the fabrication of *foie gras*). Increasingly, polls and public surveys, like those undertaken by Eurobarometer, are revealing widespread public engagement with the issue of farm animal welfare. This growing popular concern raises two important issues. First, it discloses what are, in places, significant discrepancies between, on the one hand, scientific approaches to the definition and assessment of farm animal welfare and, on the other hand, popular conceptions of what is good and what is bad for farm animals. Here, anthropomorphic and naturalist notions are often prevalent with welfare being clearly associated with such elements as outdoor and barn rearing (over battery rearing), straw bedding, communal housing, and grass feed. The science, on the other hand, is often far more equivocal on these issues pointing to higher mortality rates, aggressive behavior, greater exposure to variable environmental conditions, and so on. Second, the growth of popular awareness of farm animal welfare nonetheless exposes the considerable distance that persists, for a variety of reasons, between expressed concern, on the one side, and individual or collective action on the other. While surveys show that many people disapprove of certain contemporary husbandry practices, research systematically reveals that relatively few act upon these concerns when acquiring food. In part, this is because farm animal welfare has to take its place in the increasingly pluralist ethical habitus of today's consumers. In part too, however, farm animal welfare, and its ultimate association with animal death, is something that the majority of people do not wish to actively think about when buying animal products, preferring to pass the ethical buck to legislators and food chain actors to act ethically on their behalf. This has major implications for the way in which farm animal welfare is regulated.

The third change in the shifting balance of responsibility for farm animal welfare is therefore its growing commodification. For alongside the expanding legislative regulation of farm animal welfare at the national and international scale, food chain actors, from producers and manufacturers to retailers, are increasingly using animal welfare criteria as an element of market segmentation and added product value. In a retail food sector dominated by competition over price, segmentation by quality is opening up a far more differentiated market and is responding to more diversified consumer demands. The use of animal welfare criteria in such segmentation strategies is exemplified by egg categorization and the enormous growth in sales of free range eggs. Retailers, manufacturers, and producers are also employing welfare criteria over and above the minimum legislative requirements, such as outdoor rearing, longer life, absence of hormones, and so on, to a widening range of other foodstuffs and are distinguishing these products on supermarket shelves through labeling information. Although, with the noted exception of free range eggs, segmentation and labeling on the sole basis of animal welfare is relatively rare (the principal exception being the Royal Society for the Prevention of Cruelty to Animals' Freedom Foods label), an expanding number of food quality assurance and certification schemes, operated largely by retailers in some countries, and also by manufacturers and producers in others, include animal welfare requirements either as an entry point to retailer shelves (in which case they are not necessarily labeled as such but form part of the corporate responsibility of the food actor) or as a component in a broader classification (such as organic or 'quality' labeled products). The market, therefore, is currently demonstrating its potential for playing a significant role in raising welfare standards, often over and above regulatory minima. In recognition of this, there are currently moves within the European Union to develop an animal welfare label for products derived from higher welfare production systems, similar in its objectives to the existing organic labeling scheme.

For many observers, the growth in consumer awareness of animal welfare issues, and the market response to that awareness, represents a significant step forward that has the potential to go well beyond mere legislative standards. As such, welfare is bundled together with other contemporary consumer concerns such as the environ ment and health into a package of newly commodified ‘goods’. The introduction of this new and growing set of actors in the debate, retailers as well as consumers, ul timately messes up the concept of animal welfare even more and further politicizes the competing welfare ontologies that exist within science. Yet, in doing so, they collectively counter the long standing disavowal of ani mal lives and animal realities in the processes associated with providing humans food, and this is surely a progress. There is no easy politics to the question of animal wel fare. Any animal happiness that may result from heightened welfare provision still ends just as suddenly with premature death. The ethical impasse remains.

See also: Human-Nonhuman; Posthumanism/ Posthumanistic Geographies.

Further Reading

Appleby, M. C. and Hughes, B. O. (eds.) (1997). *Animal Welfare*. Wallingford, CT: CAB International.

Appleby, M. C. and Sandoe, P. (2002). Philosophical debate on the nature of well being: Implications for animal welfare. *Animal Welfare* 11, 283 294.

Armstrong, S. J. and Botzler, R. G. (eds.) (2003). *The Animal Ethics reader*. London: Routledge.

Broom, D. M. (1991). Animal welfare: Concepts and measurement. *Journal of Animal Science* 69, 4167 4175.

Buller, H. and Morris, C. (2003). Farm animal welfare: A new repertoire of nature society relations or modernism re embedded? *Sociologia Ruralis* 43(3), 216 237.

Callicott, J. B. (1980). Animal ethics: A triangular affair. *Environmental Ethics* 2, 311 338.

Dawkins, M. S. (1990). From an animal’s point of view: Motivation, fitness, and animal welfare. *Behavioural and Brain Sciences* 13(1), 1 9.

Dawkins, M. S. (2006). The scientific basis for assessing suffering in animals. In Singer, P. (ed.) *In Defense of Animals. The Second Wave*, pp 26 39. Oxford: Blackwell Publishing.

Duncan, I. (2006). The changing concept of animal sentience. *Applied Animal Behaviour Science* 100(1 2), 11 19.

Duncan, I. J. H. (1993). Welfare is to do with what animals feel. *Journal of Agricultural and Environmental Ethics* 6 (supplement 2), 8 14.

Fraser, D., Weary, D. M., Pajor, E. A. and Milligan, B. N. (1997). A scientific conception of animal welfare that reflects ethical concerns. *Animal Welfare* 6, 187 205.

Grandin, T. and Johnson, C. (2005). *Animals in Translation*. London: Bloomsbury.

Harrison, R. (1964). *Animal Machines*. London: Robertson and Watkins.

Lund, V., Coleman, G., Gunnarsson, S., Appleby, M. C. and Karkinen, K. (2006). Animal welfare science Working at the interface between the natural and the social sciences. *Applied Animal Behaviour Science* 97(1), 37 49.

Regan, T. (1996). Animal rights and welfare. In Borchert, D. M. (ed.) *The Encyclopedia of Philosophy Supplement*, pp 58 60. New York: Macmillan Reference, Simon and Schuster and Prentice Hall International.

Regan, T. (2004). *The Case for Animal Rights* (3rd edn.). Berkeley, CA: University of California Press.

Rollin, B. (1995). *The Frankenstein Syndrome. Ethical and Social Issues in the Genetic Engineering of Animals*. Cambridge, MA: Cambridge University Press.

Rollin, B. (2003). *Farm Animal Welfare* (2nd edn.). Ames, IA: Iowa State University Press.

Sandoe, P. and Simonsen, H. (1992). Assessing animal welfare: Where does science end and philosophy begin? *Animal Welfare* 1, 257 267.

Singer, P. (1975). *Animal Liberation: A New Ethics for Our Treatment of Animals*. New York: Random House.

Singer, P. and Mason, J. (2006). *The Ethics of What We Eat*. New York: Rodale.

Tannenbaum, J. (1991). Ethics and animal welfare: The inextricable connection. *Journal of the American Veterinary Medical Asoociation* 198(8), 1360 1376.

Webster, J. (2005). *Animal Welfare: Limping Towards Eden*. Oxford: Blackwell.

Wolch, J. and Emel, J. (eds.) (1998). *Animal Geographies*. London: Verso.

Relevant Websites

http://www.fawc.org.uk/
Farm Animal Welfare Council.

http://www.rspca.org.uk/servlet/Satellite?pagename=RSPCA/RSPCARedirect&pg=FreedomFoodHomepage
RSPCA Freedom Food.

http://ec.europa.eu/food/animal/index en.htm
The EU Animal Welfare site.

http://www.hsus.org/
The Humane Society.

http://www.welfarequality.net/everyone
The Welfare Quality research project.

Antarctica

S. Chaturvedi, Panjab University, Chandigarh, India

Introduction

With a total area of about 5.5 million square miles (or some 14 million km^2) constituting one tenth of the world's land surface, Antarctica is far larger than India and China put together. Right through the nineteenth century and up to World War II, human presence and activity in the Antarctic was rather patchy, both spatially and temporally. Thereafter, and particularly from the International Geophysical Year (IGY), during 1956–57, permanent scientific stations were established and sea sonal marine mammal harvesting continued. But it was not until the late 1970s that the southern polar region was subjected to significant modern commercial interest. From this period (marked by resource geopolitics and diplomacy) onward, Antarctica has been increasingly integrated into global systems and highly capitalized actors and forces of the globalized economy have arrived on the scene. Acting as a major catalyst for this trans formation are of course the technological, political, and attitudinal transformations in the wider international system.

One can identify the following major geographical representations of the Antarctic, which broadly coincide with different periods of the present century: (1) a par tially filled space, with seven countries (Argentina, Aus tralia, Chile, France, Great Britain, Norway, and New Zealand) staking territorial claims (1900–40), (2) a site implicated in the Cold War politicking (1945–60), (3) the 'greatest natural laboratory of science' on the planet (1960s), (4) a place of galvanizing resource potential (1970s to 1980s), and (5) a 'natural reserve' devoted to science, peace, and environmental protection, increas ingly subjected to commercial tourism (since the 1990s). These varied, and in some respects overlapping, repre sentations of the Antarctic carry important policy implications.

Antarctica in the Imperial Age (1900–40)

During the course of the late nineteenth century, discoveries and exploration of the last unclaimed and unstated spaces on the globe, Antarctica was subjected to a predominantly state centric and power political geopolitics, based on the premise that territory and ter ritorial control necessarily implied more power, prestige, and security.

In the second half of the twentieth century, Antarctica was gradually transformed from a smooth and un differentiated space into a differentiated landscape on sheet maps, especially those of the claimant states, with the help of expanding pool of cartographic and geo graphical knowledges. Modern technologies associated with aerial photography and aviation were fruitfully employed in the pursuit of 'knowing', 'naming', and 'mastering' the Antarctic environment. In short, throughout the early part of the twentieth century, the Antarctic continent was being imagined and approached by both the explorers and imperial administrators as an unexplored place that not only had to be 'discovered' and transformed into a territorial object (through the pro motion of expedition) but also owned politically (through annexation), economically (through the control of whaling), and epistemologically (understood within im perial knowledge systems).

Antarctica in the Cold War Ideological Geopolitics (1950s and 1960s)

The dominant spatial representations of the Antarctic during the 1950s were affected by the Cold War dis courses. Earlier on, there were clashes between Britain and Argentina over the issue of territorial ownership. By late 1947, Antarctic affairs assumed immense significance in foreign policy considerations of the US, and the im petus for this shift in focus had more to do with the Soviet Union than with events in the southern polar region *per se.* The US Department of State in mid Feb ruary 1949 suggested to interested governments (ex cluding the Soviet Union) that perhaps the best way of effectively dealing with the problem of conflicting claims for the sake of science in the Antarctic was to put in place some form of internationalization.

With the East–West rivalry for power and influence now extending to the south polar region as well, the US proposal to internationalize Antarctica was motivated by the broader goals of the containment strategy: keeping the Soviet Union out of Antarctica and its affairs. It was also about containing Argentine, Chilean, and British rivalry in the Antarctic Peninsula. For its part, the Soviet Union characterized the US initiative as a "fig leaf to cover the American imperialists' design to seize the whole of the Antarctic" and proved the Western hopes of excluding the Soviets from the Antarctic negotiations rather naïve.

IGY (1957–58) and the Discursive Transformation of the Antarctic

It was science, especially the IGY held during 1957–58, which transformed Antarctic politics and laid down the groundwork for the Antarctic Treaty of 1959. More than 40 research stations and observatories were established during this period, and thousands of scientists were located on the continent. There was more or less a tacit agreement among the participating countries that political problems regarding the Antarctic ought to be shelved for the duration of the IGY. Yet, some kind of a 'red threat' was perceived by the United States and its allies from Soviet polar program. The IGY was instrumental in giving rise to the dominant representation of Antarctica as a laboratory for fundamental science and accorded the politicization of Antarctica a totally new, unprecedented direction. The geopolitical imperative to preserve it as such and in its pristine form now became apparent, bordering some kind of urgency, in the given Cold War conditions pervading the world.

Against this background, President Eisenhower announced on 3 May 1958 that the US had invited 11 other countries, this time including the USSR, to confer with US in seeking an effective joint means of keeping Antarctica open to all nations to conduct scientific or other peaceful activities there under joint administrative arrangements. The invitation avoided suggesting a prominent position or role for the United Nations in the envisaged arrangement. Instead, USA would prefer an agreement to disagree that would preserve a legal *status quo* on Antarctica, not only in regard to the claims of sovereignty by some nations but also in regard to the rights of the other interested parties accruing from their stakes and activities in the Antarctic, including, of course, her own.

In case of the claimant states, what so obviously entered into their willingness to meet and discuss a political arrangement for Antarctica was not just the nagging uncertainty over the future of their claims in the event of a concerted occupation of the continent by the United States and the then Soviet Union, but also the utter impracticality of defending national interests in Antarctica by conventional military means. Both Argentina and Chile, probably others too, concluded on the basis of pragmatic geopolitical reasoning that their Antarctic claims would be advanced more effectively within rather than outside the treaty. Apparently, the interests of both USA and the USSR too had converged. Both found in the proposed treaty, therefore, the best possible way of avoiding a confrontation, while at the same time guaranteeing and safeguarding one's own strategic position, lest it be diminished by a strategic advantage going to the other party.

Antarctica as a Continent of Science and Peace: 1950s and 1960s

The Antarctic Treaty (cited hereafter as the Treaty), with a preamble and 14 articles, was signed on 1 December 1959 by the representatives of the 12 IGY participating countries – Argentina, Australia, Belgium, Chile, France, Japan, New Zealand, Norway, South Africa, United Kingdom, United States of America, and USSR. It took effect on 23 June 1961. The preamble to the Treaty underlines that, "it is in the interest of all mankind that Antarctica shall continue forever to be used exclusively for peaceful purposes and shall not become the scene or object of international discord." Article I of the Treaty prohibits all activities of a military nature, such as establishment of military bases and fortifications, carrying out of military maneuvers, as well as the testing of any weapons. Article V prohibits any nuclear explosion in Antarctica and the disposal of radioactive waste material.

Article VII provides for wide rights of inspection to all areas of Antarctica by the Antarctic Treaty Consultative Parties (ATCPs). According to Article VI, the Treaty's provisions apply to the area south of 60° South latitude. At the same time, however, Article VI disavows any incursion upon high seas rights as recognized by international law. The Treaty in no way attempts to define explicitly just which maritime areas of the Southern Ocean below 60° South latitude are, or should be, considered high seas. The primary reason for this ambiguity is that five of the seven who claim sovereignty over portions of the Antarctic continent demarcate their claims in such a manner so as to include areas of the Southern Ocean.

The innovative containment of sovereignty positions through Article IV of the Treaty is the major reality of Antarctic geopolitics. The article explicitly declares that "nothing contained in the present Treaty shall be interpreted as: a renunciation by any contracting party of previously asserted right or claims to territorial sovereignty." The Antarctic Treaty in general and Article IV in particular have escaped a critical scrutiny with regard to the manner in which they 'rewarded' colonial occupation and annexation. The legal 'freezing' of territorial claims for the duration of the Treaty (no specific termination date is being mentioned) is therefore much more than a carefully crafted diplomatic solution to the thorny issue of claimant and nonclaimant states; it protects and promotes a particular vision of the continent anchored in the colonial past.

The Antarctic Treaty parties (as of June 2008, there are 46 parties to the Treaty, of which 28 are the consultative parties), including the original members of the Antarctic Treaty, meet annually (previously it was biennially) in order to discuss recommendations for the Antarctic continent or the implementation of existing agreements. Although a good deal has been said and written of late about the authority, legitimacy, and ef fectiveness (legal and political) of the Antarctic Treaty System (ATS), much less critical attention has been paid to the fact that beyond the apparent egalitarian frame work of the Treaty, considerable asymmetries of know ledge–power nexus continue to persist. The structure of the ATS continues to 'reward' those consultative parties that have demonstrated 'substantial scientific research' with full voting rights. In contrast to the obligation of the new acceding states to carry out 'substantial scientific research activity' in Antarctica, understood to mean the establishment of scientific stations or the dispatch of scientific expeditions (Article IX), the original signatories are not obliged to do so.

The power–knowledge nexus is perhaps most clearly witnessed in the annual meetings of the ATCPs, which are overwhelmingly dominated by English speaking parties with established histories and geographies of en gagement with Antarctica. This is consolidated by the role played by specialist agencies/bodies such as Scien tific Committee on Antarctic Research (SCAR), the Committee on Environmental Protection, and the Sci entific Committee of Conservation of Antarctic Marine Living Resources (CCAMLR), which further reinforces the connection between scientific knowledge production and political influence and decision making.

As pointed out by some scholars, the role of science and the production of scientific knowledge have been pivotal in the colonization of Antarctica. Science con tinues to be the key currency of geopolitical influence in the southern polar region, also deployed for the purposes of the assertion as well as denial of territorial claims. In the light of burgeoning research in cultural histories of science, a far more serious and systematic research is needed to investigate how various scientific discourses, and the practices flowing from them, contribute to the consolidation and exercise of geopolitical power in the Antarctic.

During the 1950s and 1960s, it became commonplace for political leaders and scientists alike to frequently describe the Antarctic as either a 'laboratory of science' or a 'continent of peace and science'. Such a represen tation of the Antarctic was driven in part by the attempt to secure further funding for research and/or to preserve Antarctica's status as a zone of peace. Furthermore, the deployment of the term 'laboratory' helped create a specific place as a spectacle or an experiment, around which a consensus or consent among the participants of the ATS could possibly be generated. The conventions and procedures of the Antarctic Treaty parties too have played an important role in establishing controls on ac tions within that space and in establishing criteria for what qualified as proper scientific knowledge or sub stantial scientific research.

Growing Focus on the Resource Geopolitics: Southern Polar Region during 1970s and 1980s

The 1970s were a landmark in the discursive transfor mation of the Antarctic. By this time a rather impressive outline of Antarctic resources had emerged as a result of extensive geological and biological research of the region. This was soon filled in by perceptions of the burgeoning population in terms of consumption and depletion of resources in the wake of the oil crisis. During the 1970s, the fact of non Treaty actors taking interest in the Ant arctic resources was enough in its own right to galvanize the ATCPs into action. The ATCPs accordingly per ceived the need to have a regulatory framework in place before the krill fishery, the dynamic of which were not yet fully understood, was actually overexploited. More over, in the mid 1970s, Northern Hemisphere countries such as the USSR, Poland, and Japan – also South Korea and Taiwan – were excluded from their traditional fish ing grounds due to the proclamation of the 200 mile fishing zones.

Against the backdrop of such a curious mix of political compulsions and ecological imperatives, the ATCPs formalized the convention on the CCAMLR at Canberra in May 1980. After due ratification, it came into force in 1982. Although created at least in theory outside the ATS, in reality, CCAMLR is organically linked to the latter. The geopolitical consensus underlying Article IV of the Treaty is not only retained but also explicitly re iterated along with its concomitant ramifications for the freedom of the high seas. Article IV of CCAMLR con tinues to maintain the moratorium on territorial claims. The steadily creeping jurisdiction of the ATS into the seas 60° South latitude is further extended to the Ant arctic convergence. Second, political preeminence of the ATCPs within the ATS is preserved.

It is worth noting that like other Antarctic Treaty provisions, the convention is not enforceable, even on members, and relies on voluntary compliance. Hence, CCAMLR suffers from tremendous disadvantage when it comes to the enforcement of its conservation measures. In recent times, Antarctic finfish stocks, particularly Patagonian toothfish (*Dissostichus eleginoides*), have been heavily fished around the sub Antarctic islands within the CCAMLR area.

The 1970s were a landmark in the discursive trans formation of the Antarctic. By this time, a rather im pressive outline of Antarctic resources had emerged as a result of extensive geological and biological research of the region. This was soon filled in by perceptions of the burgeoning population in terms of consumption and depletion of resources in the wake of the oil crisis. In the prevailing economic, technological, and geopolitical cli mate at the time, mining the Antarctic was simply inconceivable. When the ATCPs turned to the question of Antarctic minerals in 1970, to be joined by India and Brazil as consultative members in1983, it was commonly felt, however, that once commercial curiosity was excited, interests awakened, or some mineral deposits on commercial scale found, the question of territorial sovereignty would inevitably surface, thus jeopardizing the ATS. Under the circumstances, it was just as well that full knowledge of the minerals map of the continent was not to hand.

The division between claimant and nonclaimant states that emerged at these discussions was sharper even than expected. The key question appeared to be whether the legal *status quo* under the Treaty could be reconciled with an equitable plan to develop mineral resources. India, Brazil, China, and Uruguay, after being admitted as consultative members, formed a lobby of countries in their condition to watch and promote Third World interests in the minerals negotiations. No wonder that the ATCPs, faced with such dilemmas, felt obliged to reiterate that mineral resource regime should not prejudice their respective positions on the question of territorial claims in Antarctica. Any agreement that may be reached on a regime for mineral exploration and exploitation in Antarctica elaborated by the ATCPs should be acceptable and be without prejudice to those states which had previously asserted rights of or claims to territorial sovereignty in Antarctica.

In January 1988 when CRAMRA seemed a virtual certainty, dispatch of an expedition to service the Greenpeace's World Park base at Cape Evans highlighted the persistent and well articulated campaign of the Greenpeace movement, along with many other national and international environmental groups, against the very idea of mining the Antarctic. The prospects of CRAMRA's disappeared eventually, when, in May 1989, the Government of Australia announced that it would not sign since it now felt strongly committed to the view that no mining should take place at all in and around Antarctica and a comprehensive environmental protection convention within the framework of the ATS should be pursued instead. French support for the Australian position promptly followed.

In order to enter into force, CRAMRA needed to be ratified by all the countries having territorial claims in Antarctica. The consensus within ATS was obviously threatened. The 'U' turns by Australia and France on CRAMRA brought into question the collective understanding of the ATCPs to abide by the norms of the system, and seriously undermined the capability of the ATS to resolve intrasystem conflicts.

The announcement, on 4 July 1991, of the US decision to sign the protocol on environmental protection to the Antarctic Treaty marked, in a way, the end of the most critical trial of the inner strength and viability of the ATS. The protocol on environmental protection to the Antarctic Treaty (hereafter cited as the Protocol) was concluded by consensus on 4 October 1991 at Madrid.

'Question of Antarctica' in the United Nations: The Rise and Decline of Alternative Visualization of the Antarctic

The interest of the 'outsiders' in the icy continent arose as early as early 1980s and somewhat in direct proportion to the origins and evolution of minerals issue within the ATS. Malaysia was to emerge as the most vociferous and dissatisfied among the critics of the ATS. The Malaysian position reflected, in part or whole, that of the most of the developing nations, including Antigua, Barbuda, Pakistan, Bangladesh, Cameroon, Cape Verde, Egypt, Ghana, Nigeria, the Philippines, Sri Lanka, and Zambia, to name a few. All were critical of the allegedly exclusive nature of the system, the membership of South Africa (no longer an issue in the UN debate), and the distribution of Antarctic resource benefits. From 1984 to 1987, Malaysia and the ATCP's positions over the 'question of Antarctica' were polarized. Malaysia was bent upon ensuring that Antarctica and its resources were managed in the interests of all mankind. In other words, Malaysia was persistent in pursuing the interests of developing countries and demanded that the concept of 'common heritage of mankind' should be applied to Antarctica.

The ATCPs, on the other hand, have rejected as 'misconceived' and 'misrepresenting' the accusation in the past that ATS is anachronistic, discriminatory, harbors colonial territorial claims, is exclusive, and thus should be replaced by the common heritage of mankind principle. Whereas the growing membership of the ATS (including the accession to the Antarctic Treaty by India, China, and Brazil, among others) is underlined to refute the charge of exclusiveness, the "widely observed principle in international relations whereby those countries primarily engaged in particular activity are responsible for management and decision making" is being emphasized as 'sensible and working' for the Antarctic. From 2002 onward, the debate in the UN on the 'question of Antarctica' can be interpreted as 'constructive engagement'. Malaysia too has increased its direct and indirect Antarctic scientific effort in the late 1990s and early 2000s. Malaysia seems to have accepted the *de facto* presence of the ATS and joined the SCAR.

Antarctica as a 'Natural Reserve' Devoted to Peace and Science: Since the 1990s

Since the 1990s, the Antarctic has increasingly been represented in global ecological terms. The growing body of evidence purporting to link the polar continent to possible global climatic change has been used by scientists and scientific organizations to explain and legitimize their endeavors. Antarctic is being increasingly seen not only as a barometer of global environmental and resource debates but also as a key point of reference to other concerns such as atmospheric circulation, sea level changes, and ozone depletion.

The 1991 Protocol to the Antarctic Treaty on Environmental Protection designates Antarctica as a 'natural reserve devoted to peace and science' and binds its present and future signatories to total protection of the Antarctic environment – its intrinsic and extrinsic worth, including its wilderness, esthetic value, and its value as an area for scientific research, especially that which is essential to understanding global environment. It categorically prohibits any activity relating to mineral resources, 'other than scientific research'. The Protocol sets out some basic environmental principles to govern all human activity in Antarctica – be it scientific, tourism related, governmental, nongovernmental, or related to logistic support. According to the Protocol, activities in the Antarctic Treaty area shall be planned and conducted on the basis of information sufficient to allow prior assessments of, and informed judgments about, their possible impacts on the Antarctic environment and dependent and associated ecosystems and on the value of Antarctica for the conduct of scientific research.

Organically linked to the Antarctic Treaty and other components of the ATS, the Protocol in no way alters the 'special legal and political status of Antarctica'. Still, it does break new ground, while raising certain pertinent questions about the protection of the Antarctic environment in the wake of mounting global awareness and interests of all kinds, growing capability of certain nations to devise new uses or find new values of the area, and as advances in high tech expand general levels of science and capability.

Imagining Antarctica as a World Park: Regulating Tourism

Managing tourism in extremely scenic but ecologically sensitive Antarctic, whose chief export remains scientific data, has emerged in recent years as one of the key questions facing the ATS, characterized by steep annual increases, diversification, and geographical expansion. The increasing number of tourists visiting specific or select, more accessible, and logistically considered as safe areas on the continent or its surrounding islands has led to serious concern over their likely environmental, legal, as well as geopolitical implications in the ATS. Although Antarctic tourism began in the late 1950s, it remained at low levels until the early 1990s when it took off. From a base of 4698 tourists in the 1990–91 summer, annual numbers of tourists landing from sea borne vessels in the Antarctic Treaty area have risen to approximately 32 198 in the summer of 2007–08.

What makes Antarctic tourism different from tourism in other parts of the globe is the peculiar geographical and legal–geopolitical setting of the region. Despite revolutionary improvements in the technology and logistics in recent years, it is the physical geography of the area that dictates the when and 'where' of Antarctic tourism. As a result, tourist visits tend to concentrate at the most accessible parts of the Antarctic (such as the northern Antarctic Peninsula and the sub Antarctic islands, particularly South Georgia, Macquarie Island, South Shetland Islands, and New Zealand sub Antarctic islands) and coincide – even clash – with the most productive period for scientific research, that is, the Antarctic summer.

Jurisdictional and liability related uncertainties surrounding Antarctic tourism are further complicated when third parties are involved. The dilemma that the ATPs continue to face in terms of all third party activities in the Antarctic – but more so in the case of tourism, due to its multinational character – arises out of generally accepted principle in international law that no treaty creates obligations for any third party without its consent. If that is so, then what kind of 'appropriate efforts' should the ATPs take (under Article X of the Antarctic Treaty) to ensure that tour companies and/or tourists do not engage in "any activity contrary to the principles or purposes of the Antarctic Treaty." Even though tourism related activities in the Antarctic are being accepted by the ATPs as a peaceful use of the Antarctic, many potential problems remain.

According to the Association of Southern Ocean Coalition (ASOC), Antarctica is not a theme park, but a region protected according to the basic principles of a World Park – a conception of Antarctica emerging from the 1972 World Park Congress. The World Park Antarctica concept is based on the protection of the Antarctica wilderness values and wildlife, and in the maintenance of Antarctica as a zone of peace, free of all weapons, and devoted to international scientific cooperation. The representation of Antarctica as a 'World Park' has been contested by a number of leading Antarctic Treaty parties, who would prefer to visualize Antarctica as a 'natural reserve' devoted to science and peace.

Tourism has the potential to reinforce territorial claims on Antarctica, using tourism, and control of tourist activity, to support a particular claim to sovereignty. One might argue that, the primary reason for Argentina and Chile promoting tourist activity on the Antarctic Peninsula is that it provides support for their territorial claims. In the light of the issues raised above, the claimant states face a serious challenge to their claims and assertions of sovereignty. What remains obvious under the prevailing circumstances is that regulation of Antarctic tourism will continue to take place at three important levels: (1) the ATS level, (2) the governmental level, and (3) tourism industry level. Undoubtedly, all the three have their respective contributions to make in an effective regulation of Antarctic tourism.

Growing Focus on Biological Prospecting

Biological prospecting is taking place in the Antarctic, and some Antarctic life forms have already been ‘protected’ with patents. The US Patent Office database contains more than 300 references to Antarctica and 92 applications for patents that refer to Antarctica, while a similar survey of European Patent Office records contains at least 62 patents that rely on Antarctic biodiversity. The question of whether companies should be able to profit from species unique to this region is complicated. In principle, provided there was a proper regulatory regime in place, one could say it could be treated like other activities. As Antarctica is set aside under the environment protocol to the Antarctic Treaty as a protected area dedicated to open science and environmental protection, allowing a free for all on bioprospecting does not appear consistent with those values. There seems little doubt that allowing a commercial bioprospecting to develop without a regulatory framework will tend to undermine one of the most important aspects of the way the Treaty system has operated – sharing all scientific information freely.

Conclusion

Today, the prime purpose of the ATS is no longer the containment of the East–West confrontation of the Cold War. Technological developments and steadily expanding economic/commercial interests imply that certain actors are looking to realize economic benefits from the region. So long as human activity in the Antarctic was of a limited scale, science oriented, and perhaps largely under the direct control of partner governments, practical challenges to the legitimacy, authority, and effectiveness were minimal as well as manageable. Whereas the ATS continues to provide regionally focused responses to various current and emerging activities in Antarctica, there is a growing realization that the period of Antarctic exceptionalism is over and effective regulation requires a greater engagement with global trends, instruments, and norms.

The most significant continuity and the legacy of the colonial past in the Antarctic relates to the array of territorial claims on the continent which, despite Article IV of the Antarctic Treaty, seems to have acquired new dimensions of complexity. The claimant states have asserted sovereign rights over their claimed territories, by way of issuing postage stamps, increased activity by their national scientific expeditions, increasing their Antarctic budgets, and the application of domestic or territorial laws to the claimed sectors. In clear defiance of the end of the nation state thesis, the unresolved and obviously intractable sovereignty dispute continues to hamper the evolution of the ATS. It is not without reason that, in all the instruments negotiated and concluded by the Antarctic Treaty powers thus far, “the special legal and political status of Antarctica” is invariably highlighted as a reminder to those within or outside the ATS that every policy pursued must avoid prejudicing the position of claimants or nonclaimants in Antarctica.

Within this context, the perceptions of Antarctica in individual Southern Cone countries are strongly conditioned by their respective geographical locations and overall geopolitical visions, and are frequently different from those held by not only more distant nations but also other Southern Hemisphere countries, for example, Australia. A frequently overlooked source of conflict in the Antarctic is the geopolitical thinking that has flourished in the past quarter century, if not more, in South America – especially in Argentina and Chile, where Antarctica has already become an integral part of national consciousness and identity. An appreciation of the manner in which South American geopolitical thinking influences attitudes to Antarctica will continue to be an invaluable guide to understanding of both perceptions and realities for the quadrant located between the Greenwich and 90° W meridians.

Whereas the supremacy of the member states within the ATS is hard to ignore, political pressures exerted in recent years on the member governments, individually and collectively, to acknowledge and cooperate with a range of NGOs, especially ASOC and International Association of Antarctic Tour Operators (IAATO), is beginning to produce results. The success of the NGOs in opposing CRAMRA, and in forcing some of the ATPs to reverse their position on this issue, has contributed significantly toward making the ATS more tolerant of the presence of the NGOs on the Antarctic geopolitical scene. However, the co option of some of the NGOs could be made possible only by changing the core values of the Antarctic Treaty itself and by introducing a new environmental ethic into Antarctic geopolitics. Whether these changes would eventually lead to more dramatic institutional change in the ATS, so that some of the NGOs would be granted consultative status, is difficult to predict. What does appear certain is that Antarctic geopolitics can no longer be as state centric as in the past.

During the past decade, a number of human geographers have critically examined the cultures of imperial exploration, mapping, and administration. Some such studies have examined the roles of geographical societies and agencies (such as the Royal Geographical Society) with regard to the production of knowledge and imperial networks, which contributed to the colonization of spaces (both material and discursive) outside Europe and North America. The term 'postcolonial', therefore, does not refer to the ending of European imperialism in the post 1945 period; rather, it invites critical attention to how systems of colonial domination (whether in the form of varied geographies of colonialism or knowledge pro duction) persist in the contemporary era and continue to resist the demand for decolonization of imaginations. There is no reason why the critical gaze of 'postcolonial' research agenda should not penetrate the realm of Ant arctic studies.

See also: Arctic; Environmental Security; Postcolonialism/ Postcolonial Geographies; Tourism.

Further Reading

Bush, W. M. (1982). *Antarctica and International Law: A Collection of Inter State and National Documents*, vols. 1 and 2. London: Oceana Publications.
Bush, W. M. (1992). *Antarctica and International Law: A Collection of Inter State and National Documents,* Booklet AT91C (Issued November 1992). London: Oceana Publications.
Chaturvedi, S. (1996). *Polar Regions: A Political Geography*. Chichester: Wiley.
Chaturvedi, S. (2001). Ocean governance and the polar regions: Geopolitics, law and sustainability. *Ocean Yearbook* 15 (University of Chicago Press).
Dodds, K. (2002). *Pink Ice: Britain and the South Atlantic Empire*. London: I.B. Tauris.
Dodds, K. J. (1997). *Geopolitics in Antarctica: Views from the Southern Oceanic Rim*. Chichester: Wiley.
Dodds, K. J. (2006). Post colonial Antarctica: An emerging engagement. *Polar Record* 42(220), 59 70.
Elliott, L. M. (1994). *International Environmental Politics: Protecting the Antarctic*. London: St. Martin's Press.
Hemmings, A. D. (2007). Globalization's cold genius and the ending of Antarctic isolation. In Kriwoken, L. K., Jabour, J. & Hemmings, A. D. (eds.) *Looking South: Australia's Antarctic Agenda, Sydney*, pp 176 190. Annandale, NSW: The Federation Press.
Rothwell, D. R. (1996). *The Polar Regions and the Development of International Law*. Cambridge: Cambridge University Press.
Sahurie, E. J. (1992). *The International Law of Antarctica*. New Haven, CT; Dordrecht: New Haven Press; Martinus Nojhoff Publishers.
Sidaway, J. (2000). Postcolonial geographies: An exploratory essay. *Progress in Human Geography* 24, 573 594.
Tepper, R. and Haward, M. (2005). The development of Malaysia's position on Antarctica: 1982 to 2004. *Polar Record* 41(217), 113 124.
Verhoeven, J., Sands, P. and Bruce, M. (1992). *The Antarctic Environment and International Law*. London: Graham & Trotman.
Watts, A. (1992). *International Law and the Antarctic Treaty System*. Cambridge: Grotius Publications Limited.

Relevant Websites

http://www.asoc.org
Antarctic and Southern Ocean Coalition (ASOC).
http://www.ats.aq
Antarctic Treaty Secretariat.
http://www.ccamlr.org
Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR).
http://www.state.gov
Handbook of the Antarctic Treaty System, US Department of State.
http://www.iaato.org
International Association of Antarctic Tour Operators (IAATO).
http://www.scar.org
Scientific Committee on Antarctic Research (SCAR).
http://www.spri.cam.ac.uk
Scott Polar Research Institute.

Anthropogeography (After Ratzel)

J. Lossau, Humboldt-Universität zu Berlin, Berlin, Germany

Glossary

Causal Explanations These are applied by researchers who assume that phenomena have causes (and can be explained by these causes). According to this rather rigid logic, science is aimed at finding causes and not purposes or motivations.

Ecumene The part of the Earth which is inhabited and cultivated on a permanent basis.

Environmental Determinism It is the belief that society (i.e., human agencies, practices, and institutions) is causally shaped and constrained by the natural environment.

Evolution It states the gradual transformation of patterns toward growing complexity. The term is often applied to life forms (simple organisms but also humankind).

Lebensgebiet The habitat ('living space') of a given species. In biologist approaches to human life, the term also refers to the area where a 'volk' lives.

Lebensraum See. Lebensgebiet.

Regional Geography It is the study of an area. Traditional regional geography assumed that a 'natural' arena (e.g., *Landschaften*) was characterized by a unique and harmonious interplay between natural and cultural components.

Social Darwinism It is a broad array of ideas based on predominantly Lamarckian beliefs, stating that human societies must compete with each other in close interaction with their respective natural environments in order to survive.

Volk A term that refers to a collection of people who share the same culture, language, and history. While contemporary notions assume that a given 'volk' represents an imagined and socially constructed community, a 'volk' was thought to be constituted by natural forces in Ratzel's time.

Wagner's Law of Migration The naturalist, explorer, and evolutionary theorist Moritz Wagner aimed at integrating the significance of migration and geographical isolation into the concept of evolution. According to his 'law of migration', it is due to migration and isolation that biological divergence and specification is induced and preserved.

Introduction

Anthropogeography, or the geography of Friedrich Ratzel, is widely acknowledged as the groundwork of human geography. Ratzel (1844–1904), who had studied, among other subjects, zoology in Heidelberg and Jena, was a German pharmacist, zoologist, and journalist. After having traveled over much of Europe as well as of North and Central America, he became established as an academic geographer first in Munich and then at Leipzig. Despite his education in the natural sciences, he developed a special interest into the relations between nature and humankind which led him to integrate humans firmly into his systematic pursuit of academic geography. Although he is often criticized today for his environmental determinism, it should be acknowledged that Ratzel led the way to a new form of human geography in which cultural, as well as natural, phenomena are subjected to systematic – and not only regional – study. His work, however, is characterized by a paradigmatic ambivalence which, in turn, is eventually responsible for ongoing discussions about anthropogeography's emplacement in the history of the discipline.

Friedrich Ratzel

Having published more than 1200 papers and books, not including book reviews, Ratzel represents an exceptional academic and a rather abundant writer. A big part of his oeuvre dates back to his earlier years when he worked as a journalist for the *Kölnische Zeitung*. In order to help finance his academic education, he had begun to write popular natural scientific reports of his travels and field excursions. Since his journalistic activity proved to be successful, he soon started to work full time for the newspaper, and in 1873 he was sent for 2 years as a correspondent to North and Central America. The popularity his travelogs enjoyed with their academic and indeed nonacademic readership at least partly stemmed from a remarkable gift for prose composition, revealing Ratzel's passion for artistic beauty and esthetics more generally. From his days as a journalist up to his death, he devoted himself to the depiction of nature and landscape, trying to reconcile scientific reasoning with esthetic philosophy.

Despite his success in writing journalistic, popular scientific texts, Ratzel (**Figure 1**) resigned from the *Kölnische Zeitung* in 1875, taking up an academic career in the then young discipline of geography. After his return from the Americas, he was appointed at the Technical University of Munich; in 1886, he followed Ferdinand von Richthofen on the chair of geography at the University of Leipzig, where he remained professor until his death in 1904. As several commentators have

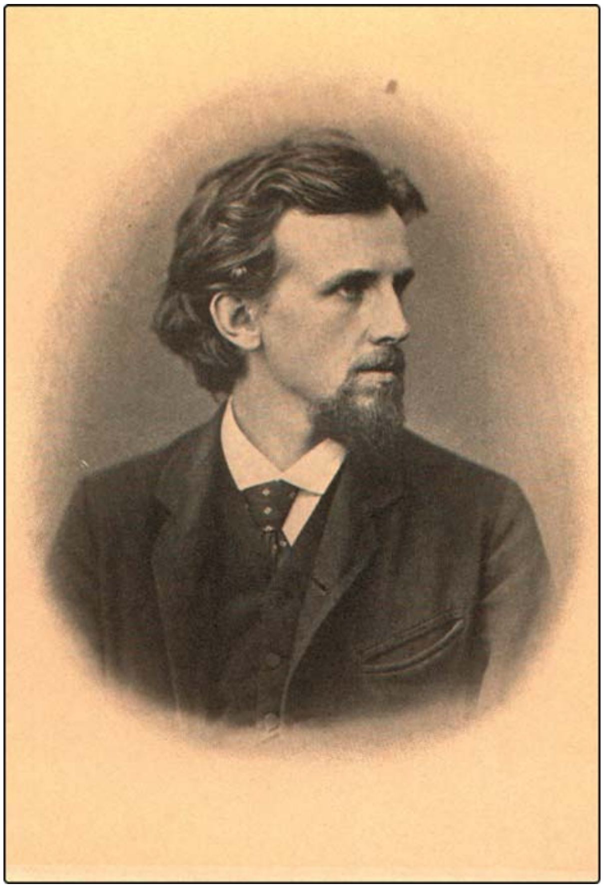

Figure 1 Friedrich Ratzel in summer 1807. From Helmolt, H. v. (ed.) (1906). *Kleine Schriften*, Bd. 1. Munich: Oldenbourg.

noted, it had been his travels which had turned Ratzel into a geographer. Especially the 2 years that he spent in the United States, Cuba, and Mexico had a deep impact on his thinking and writings. His analyses of the rapid development of North American society indicate the shift of interest in his writings from the natural world to the human realm. This shift, which is constitutive of Ratzel's *Anthropogeographie*, is embodied in the three academic publications which are widely regarded as his core contributions to the development of modern geography.

Tracing Anthropogeography in Ratzel's Work and Life

The first of these publications was a two volume regional geography of the United States, consisting of *Physikalische Geographie und Naturcharakter der Vereinigten Staaten von Amerika and Culturgeographie der Vereinigten Staaten von Amerika.* While the latter, according to Wolfgang Natter, contained a somewhat 'totalitarian' approach considering "every side of cultural life in North America" (Natter, 2004: p. 175), it was not before the publication of his second major publication – *Anthropogeographie* – that Ratzel presented a more systematic, albeit rather general, statement of his ideas on the principles of human geography. The first volume of *Anthropogeographie* appeared under the subtitle *Anthropo Geographie oder Grundzüge der Anwendung der Erdkunde auf die Geschichte* (*Characteristics of Geography's Application to History*, 1882, revised 1899); the second volume, which was published in 1891, was subtitled *Die Geographische Verbreitung des Menschen (The Geographical Distribution of Mankind).* The new science of anthropogeography pursued three principal aims of investigation: (1) to describe the regions of the ecumene, and to research the distribution and groupings of mankind over it; (2) to research human migrations in their 'dependency on the land'; and (3) to analyze the effects produced by the natural environment on individuals and societies.

The determinism insinuated in the third aim is often said to surface even more explicitly in Ratzel's third major work. In his *Politische Geographie, oder die Geographie der Staaten, des Verkehres und des Krieges* (1897; second edition 1903), Ratzel outlined a theory of expansionism which he developed out of the central concept of *Lebensraum*, defining the state as an 'earth bound organism': "The distribution of mankind (…) over the earth' surface has all the distinctive characteristics of a versatile body which expands and contracts alternatively in regression and progression, building new connections and disrupting old ones, and thereby assuming shapes which, to the closest, resemble the shapes of other (…) versatile bodies on the earth surface" (Ratzel, 1897: p. 1, transl. JL). While the question of precisely how Ratzel's political geography can be regarded as deterministic is discussed in more detail below, it remains to be said here that Ratzel's interest in questions of political theory was reflected by a thorough engagement in the political life of his time. Besides being a founding member of both the *Kolonialgesellschaft* and its successor, the *Kolonialverein*, he participated in the establishment of the nationalist conservative *Alldeutscher Verband* (Pan German League), advocating not only the need for overseas expansion but also developing an imperative for the country to develop a competitive navy.

It is at least partly due to such political considerations and obligations that Ratzel's work came to lack appreciation after the end of World War II. Especially in German academic discourse, Ratzel is widely regarded as the 'grandfather' of German *Geopolitik* which has been charged of being complicit in the development of the national socialist ideology. Another reason for why Ratzel is not commonly appreciated, at least in the Anglo American world, is that he is hardly translated into English, leaving English speaking scholars dependent mostly on the writings of one of his disciples, Ellen Churchill Semple. As a consequence, in the post war era, one had to turn to other disciplines, namely ethnography, to find Ratzel's work, as Robert E. Dickinson writes, "properly" appreciated (Dickinson, 1969: 64). It is indeed only recently that there is a renewed interest in the work of Ratzel who, despite his imperialist *Weltanschauung*, is widely acknowledged as "the true founder and greatest single contributor to the development of modern human geography" (Beckinsale, 1975: 309).

'Is Geography an Independent Discipline?'

Why is Ratzel's geography regarded so central for the development of the discipline? In order to answer this question, an investigation into the broader academic context in which anthropogeography evolved is insightful. During the last third of the nineteenth century, that is, by the time Ratzel was appointed as an academic geographer in Germany, geography underwent its institutionalization as an academic discipline. Although geography had been taught at universities for a long time before, the respective scholars had not been 'professionalized', that is, they did not have chairs in academic geography. This situation changed after 1870 when "the young German national state under Prussian hegemony sought to educate loyal citizens" (Brogiato, 2004: 53) and, therefore, promoted school subjects with a nationalistic bias, like history, German language, and literature or geography. As a consequence of the discipline's academic institutionalization, the question arose of how to delineate the new subject from already existent ones.

In order to legitimize geography as a true science with a distinct object of study, many scholars, at least in Germany, focused on the surface of Earth which was supposed to be explained in causal terms and by genetic arguments. As a consequence, geomorphology became the central approach, making Ferdinand von Richthofen one of the most important figures of this time. Richthofen, who had studied geology, defined geography as "the science of the earth's surface and the phenomena which are causally and reciprocally related to the former" (Richthofen, 1883: 5, transl. JL). Due to the natural scientific character of their frameworks, the time of Richthofen and like minded colleagues, for example, Oscar Peschel, is referred to as the 'geological period' of geography as an academic discipline.

It is in the context of the 'geological period' that the significance of Ratzel's geography becomes obvious. Ratzel, too, sought to justify the newly institutionalized discipline, reasoning about its position among other fields. In the opening pages of *Anthropogeographie*, he asks: "Is geography actually nothing more than a heap of twigs effectively belonging to other trees that have been thrown together for (...) practical reasons, and are there boundaries to be drawn capable of separating geography sharply from neighbouring disciplines? Put differently, is geography an independent discipline?" (Ratzel, 1882: 16). The latter question, however, seems to be of rather rhetorical character, since Ratzel makes it clear that, to him, geography enjoys a very privileged viewpoint to explore the world: "And this viewpoint is the synopsis of the earth's surface and the life that belongs to it as whole combined by the most manifold reciprocal relations" (Ratzel, 1882: 17).

Compared to many of his colleagues, Ratzel thus paid greater attention to the 'human factor', that is, to the relations between mankind and nature or the Earth's surface. "Ratzel's innovation," Wolfgang Natter puts it, "was to have developed a perspective for researching the reciprocal historical geographical development of humans and their social, cultural and political emplacements in relation to the earth's surface" (Natter, 2005: p. 179). By putting "humans back into the systematic pursuit of a general geography" (ibid.), Ratzel's anthropogeography challenged the discipline's preoccupation with geology and physical geography more generally as it had been developed in the writings of scholars like Richthofen or Peschel. As one observer noted in *Petermanns Mitteilungen*: "The publication of the second edition of the first volume of Ratzel's anthropogeography clearly indicates that the days in which geography was inclined to bend itself completely towards the side of the natural sciences, or indeed to become a branch of geology, are over" (Hahn, 1900: 140).

Darwinian Concepts

All of this is not to say, however, that Ratzel's perspective turned geography, in any way, into a social science. Although anthropogeography (re)introduced human aspects into a predominantly physical geography, it firmly rested on a framework of causal explanations, cause–effect relationships, and natural laws characteristic for the sciences of his era. As a consequence, Ratzel was criticized by his French contemporary Emile Durkheim, the founder of academic sociology. Durkheim, in his journal *L'année Sociologique*, reviewed both Ratzel's *Anthropogeographie* and *Politische Geography*. On the one hand, he appreciated Ratzel's contributions in that they had underlined the significant role that the 'geographical factor' played for human life. On the other hand, Durkheim clearly disapproved of the idea that there were direct relations between the Earth and mankind. Ultimately interested in how the transformation of society – and more specifically the division of labor in the industrialization process – shaped the conditions of human life, Durkheim aimed at explaining spatial patterns by human agency, and not vice versa. Ratzel, by contrast, kept insisting that geographical knowledge had to be derived from the Earth or the ground (*Boden*).

In so doing, Ratzel was influenced by Darwinian concepts of evolution through natural selection and the struggle for survival. Not only had he, as a student, at tended lectures of Ernst Heinrich Haeckel, whose ex cessive and overstated *Sozialdarwinismus* encompassed a series of evolutionary theories, for instance – and very prominent – the neo Lamarckian principle of direct heritable adaptation. He had also incorporated ideas of the naturalist, explorer, and evolutionary theorist Moritz Wagner who aimed at integrating the significance of migration and geographical isolation into the concept of evolution. According to Marc Bassin, the relationship with Wagner, whom Ratzel had met in the winter of 1871–72 in Munich, represented "probably the single most significant influence on Ratzel's intellectual devel opment, for Wagner's 'Law of Migration' provided the foundation on which Ratzel was to develop his anthro pogeography" (Bassin, 1987a: 124). Taking up Wagner's idea of the *Lebensgebiet* as the territorial base of a given species, Ratzel developed his concept of *Lebensraum*, or living space, which he continually emphasized as being of elemental significance for his work.

To Ratzel, every living organism required a specific, definite area from which to draw sustenance in order to come into existence. This was not only true for individual living organisms but also included entire populations, such as herds of animals. Even more importantly, and by means of a consistent biological analogism, Ratzel went on to apply his concept of *Lebensraum* to human society, conceptualizing the political state as the corresponding organism or 'aggregate organism'. It is precisely the idea of the political state as organic or biogeographical unit de pending on, and at the same time struggling for, *Lebens raum* that represents the source of competing geographical imaginations standing at the heart of anthropogeography.

Competing Geographical Imaginations

Ratzel argued that the state as spatial organism 'naturally' seeks to grow beyond a given area. Due to an increase in population which he regards as being ultimately dictated by the laws of nature, the space need of a state organism would grow. The resulting tendency of the state to re quire – and acquire – new territory is a core idea in Ratzel's anthropogeography which led him to formulate his 'law of the growing areas' (*Gesetz der wachsenden Räume*, varied as *Gesetz der Zunahme der politischen Räume*, *Gesetz des räumlichen Wachstums der Staaten und Völker*). At the same time, however, Ratzel realized that the surface of the Earth was limited, offering virtually not enough space for never ending and unbounded growth. Reflecting on the contradiction between the 'law of the growing areas', on the one hand, and the fact that the Earth's territory is limited, on the other hand, Ratzel believed in the exist ence of a constant 'struggle for space':

> Between the flow of life, which is never at rest, and the territory of the earth, which remains constant, there is a contradiction. Out of this contradiction the struggle for space is born. Life quickly subdued the ground of the earth, but when it had reached its limits it flowed back, and since this time life struggles with life for space everywhere and ceaselessly. The much misused, and even more misunderstood expression 'the struggle for existence (*Dasein*)' actually means first of all a struggle for space. For space is the very first condition of life, and it is space which serves as measure for all other con ditions, above all aliment. (Ratzel, 1901, 165 168, transl. JL)

Against the background of the 'struggle for space', Ratzel never lost sight of a second, and rather different, geographical imagination. Both in line with and contra dictory to the 'law of the growing areas', his anthro pogeography entails the idea of 'natural lands' (*Länder*) which, in turn, represent the demarcated territories of (supposedly) socially homogenous entities or nation states. The idea of a 'natural land' figures, for instance, in Ratzel's definition of the state as an organic whole which, for him, developed out of the interaction between a specific group of people and the discrete territory the group occupies – *Ein Stück Menschheit und ein Stück Boden*, as he famously wrote (Ratzel, 1903: 4). According to this more bounded imagination, the bonding of the group depended not so much on ethnic or racial kinship but on the shared relationship to the land. This is an important point which is further underlined in Ratzel's definition of the *Volk* as "a politically united group made up of groups and individuals, who need neither to be related ethnically nor linguistically, but who through their common terri tory (*Boden*) are also spatially linked together" (Ratzel, 1903: 5).

In accordance with the second imagination, Ratzel disapproves of what can be called arbitrary boundaries, that is, boundaries which violate the laws of nature. While the passion for expansion, for Ratzel, necessarily leads to a transgression of natural boundaries, it is nevertheless impossible to fully overrule the related natural land: The "right to particular [i.e. individual] development as it is given in the forms of the earth's surface" is said to counter the striving for growing areas of exchange as well as growing states (Ratzel, 1903: 207). For Ratzel, "the multi structured earth wants multi structured political entities, and the earth's many limbs, with silent violence, continuously break through the uniformity aiming at covering the heterogeneity of the

Figure 2 Cover illustration of Ratzel's book on Germany. The illustration allegorically depicts the German landscape with its lowlands, highlands, and high mountains. From Ratzel, F. (1898). *Deutschland*. Leipzig: Grunow.

earth" (ibid.; see **Figure 2** for an example of a natural *Land*).

Anthropogeography as Imperialist Transformation of the Classical Paradigm of Geography

It is against the background of the competing geographical imaginations that Hans Dietrich Schultz sees a "cycle of growth and defeat, of progress and collapse" at work in Ratzel's writings (Schultz, 1998: p. 140). Since the surface of the Earth is limited, the natural impulse of growth necessarily leads to a form of spatial implosion which, in turn, represents the starting point for another process of growth and, again, defeat and collapse. According to Schultz, this cycle indicates a transformation of the classic concept of geography which is both theoretical and political in nature.

With the *Lebensraum* concept having a decisively expansionist dimension, Ratzel theoretically departs from the classical paradigm of geography which, in turn, rested upon the idea of a close relationship between humankind, on the one hand, and nature, on the other hand. Revolving around the concept of *Ganzheit* or 'wholeness' as it had been developed by the philosopher Johann Gottfried Herder and reproduced, for example, by the early geographer Carl Ritter, the classical paradigm regards the region (*Land* or *Landschaft*) as individual and unique part of the Earth's surface. As **Figure 3** shows, the rationale behind the classical paradigm assumes a harmonization between a given 'natural land', its *Volk*, and the related state.

While this harmonization is to a certain extent reproduced in Ratzel's writings (e.g., when he writes about the Earth's many limbs fighting uniformity), it is at the same time clearly disrupted in his 'law of the growing areas'. By assuming an eternal 'struggle for space', Herder's idea of a 'fraternal community' or a 'brotherhood of nations' in which all nations stick to their own limited national territory, competing peacefully in the pursuit of humanity, is dismissed. Instead, Ratzel adapts the classical paradigm – which had been customized to the process of nation building – to the political circumstances of his own era, that is, the age of imperialism. In so doing, he develops a rather bellicose concept which, to Schultz, represents nothing less than a shift from "eternal peace" to "eternal war" (Schultz, 1998: 127; see **Figure 4**).

It does not come as a surprise, therefore, that the expansionist imagination has political effects. The applicability of Ratzel's theoretical ideas to late nineteenth century German *Tagespolitik* becomes obvious, for instance, in a passage written in 1898 on the naval question and Germany's position as a world power: "There will always be ruling nations and serving nations [*Völker*]," Ratzel maintained, suggesting an essential choice for a *Volk* as being "either a hammer or an anvil:"

> Whether they [*Völker*] become one or the other depends on their recognising in good time the demands which the world situation presents to a nation which is striving to rise. It has been a different exercise for the Prussia of the eighteenth century to attain its position as a major power [*Großmachtstellung*] in the middle of the European con tinental powers than it is for Germany in the nineteenth century to assert itself among the world powers. This task can no longer be solved in Europe alone; only as a world power can Germany hope to secure for its people the land which it needs for growth. Germany must not remain apart from the transformations and redistri butions taking place in all parts of the world (…) if it does not want to run the risk again, like in the 16th century, of being pushed into the background for a series of generations. (Ratzel, 1906: pp. 377 378)

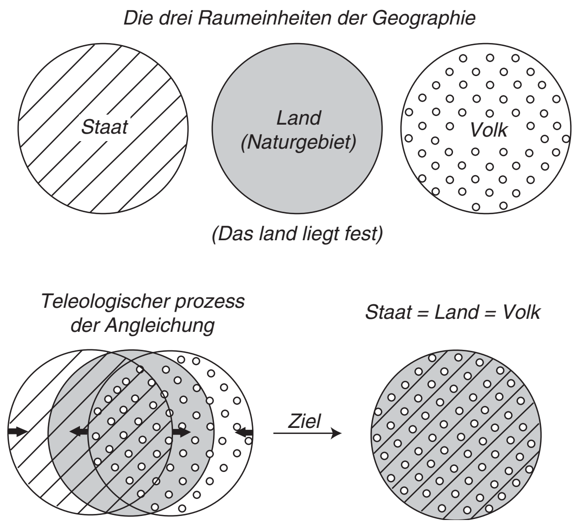

Figure 3 The classical paradigm of geography. The figure shows three spatial entities: state, land (natural area), and nation (*Volk*). According to the classical paradigm of geography, these entities, in the course of history, strive toward teleological congruency. From Schultz, H.-D. (1998). Herder und Ratzel: Zwei Extreme, ein Paradigma? *Erdkunde* 52, 127 143.

Political Impact and Deterministic Content of Ratzel's Geography

Against the background of such statements most com mentators, after the end of World War II and up until today, have come to think of Ratzel's geography in rather ambivalent terms. While the central role Ratzel played for the development of the discipline is mostly recog nized, his anthropogeography has been criticized for being deeply environmental deterministic. Ratzel, so the general opinion, is to be blamed for having suspended the will of humans in favor of eternal spatial laws governing the fate of mankind. Such judgments about the de terministic nature of Ratzel's writings often feed in with judgments about the effects of Ratzel's political engage ment, connecting his spatial determinism to the imperi alist considerations of his time, if not to the ideological excesses of the national socialist era.

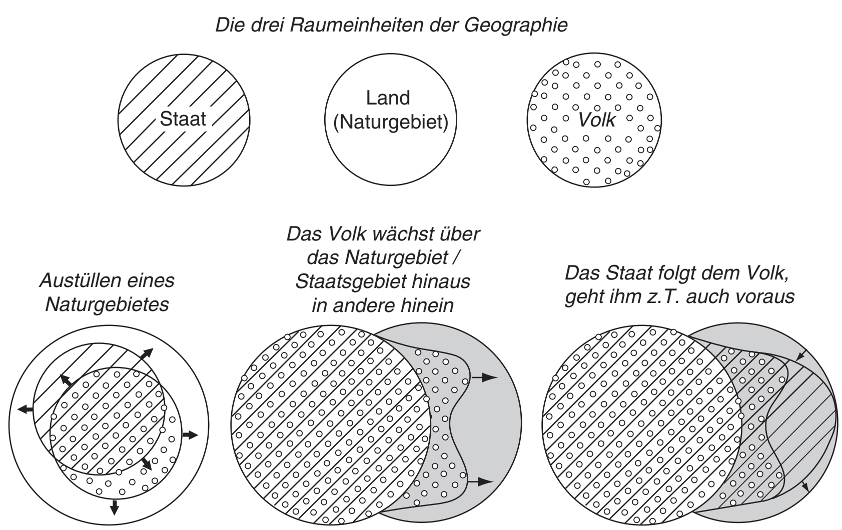

Figure 4 The imperial transformation of the classical paradigm of geography. The figure shows how the idea of teleological congruency, in Ratzel's work, is replaced by the idea of expansion. Due to the 'law of the growing areas', the three spatial entities of the classical paradigm are characterized by incongruent movements of transgression and regression. From Schultz, H.-D. (1998). Herder und Ratzel: Zwei Extreme, ein Paradigma? *Erdkunde* 52, 127 143.

More recently, however, scholars like Wolfgang Natter or, in the German speaking context, Ulrich Eisel, have bemoaned what they regard as a certain one sidedness in the appraisal of Ratzel's work. Underlining Ratzel's interest in the globalizing tendencies of the world market, it is suggested that Ratzel had developed a dynamic and modern theory of space which precisely did away with the static and environmental deterministic approach pursued by classical geography. Contemplating the merits of Ratzel's oeuvre on North America, Natter even maintains that "Ratzel had identified at least part of the dynamic of uneven development and what David Harvey has described as the spatial fix" (Natter, 2005: 175). Moreover, and with regards to the political implications of anthropogeography, Marc Bassin has detected in Ratzel's writings a "genuine and deep rooted humanism" (Bassin, 1987a: 128), arguing that Ratzel's political orientation cannot simply be read into German geopolitics after 1918, let alone into Nazi war aims.

In order to try and finally assess the deterministic content of Ratzel's geography, it is helpful, in a first step, to distinguish between two different forms of deter minism, both relating to one of the geographical im aginations inherent in Ratzel's work. The first one is connected to the imperialist imagination and suggests that there are immanent natural forces governing the fate and conduct of humankind (e.g., the 'law of the growing areas'); the second one is more in line with the geographical imagination of the classical paradigm, drawing upon the physical environment as a causal in fluence on the character and development of a given *Volk*. For Ratzel, the fate of humankind and states is thus determined by both the programs of nature (as in the imagination of 'growing areas') and the laws of the ground (as in the imagination of 'natural lands'). While this duality eventually explains the ongoing discussions about how deterministic or how modern Ratzel's geog raphy actually had been, it should be noted that, in both cases, 'nature' is clearly conceptualized as a causally determining force.

Moreover, 'nature', for Ratzel, had normative power in that it represented the irrevocable authority over right or wrong. As a consequence, anthropogeography could be used to provide an apparently scientific justification for political aims, serving, for instance, as a source of in spiration for geopoliticians, such as Karl Haushofer. Al though Bassin has argued that the influence of German *Geopolitik* on the ideology of National Socialism has been overrated since the post war years, it is with some plausibility that Schultz maintains connectivity between Ratzel's geography, on the one hand, and Hitler's ideol ogy of *Lebensraum*, on the other hand. According to Schultz, affinities do not so much concern the level of concrete political decision making but rather relate to more general aspects of political reasoning which could be understood as timeless stage directions in the drama of international relations. While there is no straightforward connection between Ratzel's geography and Hitler's war on *Lebensraum*, anthropogeography certainly belongs to the prehistory of the latter.

See also: Geography, History of; Geopolitics; Harvey, D.; Idealism/Idealist Human Geography; Lamarck(ian)ism; Region; Ritter, C.; Uneven Development.

Further Reading

Bassin, M. (1987a). Friedrich Ratzel. 1844 1904. *Geographers: Biobibliographical Studies* 11, 123 132.

Bassin, M. (1987b). Imperialism and the nation state in Friedrich Ratzel's political geography. *Progress in Human Geography* 11, 473 495.

Bassin, M. (1987c). Race contra space: The conflict between German Geopolitik and national socialism. *Political Geography Quarterly* 6, 115 134.

Beckinsale, R. P. (1975). Friedrich Ratzel. *Dictionary of Scientific Biography* 11, 308 310.

Brogiato, H. P. (2004). Geschichte der deutschen Geographie im 19. und 20. Jahrhundert ein Abriss. In Schenk, W. & Schliephake, K. (eds.) *Allgemeine Anthropogeographie*, pp 41 81. Gotha: Klett Perthes Verlag.

Dickinson, R. E. (1969). *The Makers of Modern Geography*. London: Routledge and Kegan Paul.

Eisel, U. (1980). *Die Entwicklungen der Anthropogeographie von einer "Raumwissenschaft" zur Gesellschaftswissenschaft*. Kassel: Urbs et Regio 17.

Hahn, F. (1900). Litteraturbericht. Ratzel, F.: Anthropogeographie. 1. teil. *Petermanns Mitteilungen* 10, 140 141.

Helmolt, H. v. (ed.) (1906). *Kleine Schriften*, bd. 1. Munich, Berlin: Oldenbourg.

Kost, K. (1988). *Die Einflüsse der Geopolitik auf Forschung und Theorie der politischen Geographie von Ihren Anfängen bis 1945*. Bonn: Bonner Geographische Abhandlungen 76.

Muller, G. H. (1996). *Friedrich Ratzel (1844 1904): Naturwissenschaftler, Geograph, Gelehrter. Neue Studien zu Leben und Werk und sein Konzept der "Allgemeinen Biogeographie"*. Stuttgart: Verlag.

Natter, W. (2005). Friedrich Ratzel's spatial turn. Identities of disciplinary space and its borders between the anthropo and political geography and the United States. In Houtum, H. v., Kramsch, O. & Zierhofer, W. (eds.) *B/Ordering Space*, pp 171 186. Aldershot: Ashgate.

Peet, R. (1985). The social origin of environmental determinism. *Annals of the Association of American Geographers* 75(3), 309 333.

Ratzel, F. (1882). *Anthropo Geographie*. Grundzuge der Anwendung der Erdkunde auf die Geschichte, Bd. 1. Stuttgart: Engelborn.

Ratzel, F. (1897). *Politische Geographie* (1st edn.). Munich: Oldenbourg.

Ratzel, F. (1898). *Deutschland*. Leipzig: Grunow.

Ratzel, F. (1901). Der Lebensraum. Eine biographische Studie. In Bucher, K. *et al*. (eds.) *Festgaben für Albert Schäffle zur siebenzigsten Wiederkehr seines Geburtstages am 24 Februar 1901*, pp 101 189. Tubingen: Verlag des Laupp'schem Buchhandlung.

Ratzel, F. (1903). *Politische Geographie* (2nd edn.). Munich, Berlin: Oldenbourg.

Ratzel, F. (1906). Flottenfrage und Weltlage. In Helmolt, H. v. (ed.) *Kleine Schriften*, bd. 2. Munich: Oldenbourg.

Richthofen, F. v. (1883). *Aufgaben und Methoden der heutigen Geographie. Akademische Antrittsrede*. Leipzig.

Schultz, H. D. (1995). Was ist des Deutschen Vaterland? Geographie und Nationalstaat vor dem ersten Weltkrieg. *Geographische Rundschau* 47, 492 497.
Schultz, H. D. (1998). Herder und Ratzel: Zwei Extreme, ein Paradigma? *Erdkunde* 52, 127 143.
Schultz, H. D. (2004). Nach 100 Jahren immer noch im Gesprach: Friedrich Ratzel (1844 1904). *Petermanns Geographische Mitteilungen* 148, 94 95.
Wanklyn, W. (1961). *Friedrich Ratzel. A Biographical Memoir and a Bibliography*. Cambridge: University Press.

Anthropology and Human Geography

K. W. Butzer, University of Texas, Austin, TX, USA

An Attenuated Separation

Interdisciplinary relationships are a reminder that traditional university disciplines are historical constructs that do not necessarily promote innovative teaching and research. Some disciplines have long shared a measure of organic integrity, but the 'newer' social sciences did not. Geography, together with anthropology, sociology, and political science, occupied a poorly articulated middle ground between the natural sciences, history or philosophy, and language studies. That intermediate space, never a monopoly of the academy, remained open or accessible to educated elites that drew from and contributed to a reservoir of information and ideas.

Under the banner of 'geographia', in its etymological sense, ancient Greek, Roman, and Islamic travelers and writers shared their experiences and knowledge with an educated public, intensely curious about the 'world'. From Herodotus to Idrisi, they created a durable genre of 'places, people, and events', that met the intrinsic human need to know about 'how the world looks', about 'what other peoples do and how they behave', and about 'where things happened'. Contributing to a sense of roots, lineage, and identity, such reflections developed into culturally grounded metanarratives that cast the individual as an observer of similarities and differences in physical and human landscapes.

Given the exponential growth of educational interest and activity during the late nineteenth century, 'geographia' benefited from broad interest, sponsorship, and applicability to new commercial and political needs. The scientific survey of Napoleon in Egypt (1799–1801) led to the formation of geographical societies in major cities, with meetings that drew increasing attendance and funding for university positions in geography. Governments and wealthy patrons financed a spate of expeditions around the globe, to 'discover' and explore, with predictable feedbacks for popular enthusiasm and geopolitical strategies. Business travelers abroad led to the appearance of ponderous 'handbooks' and slim primers about foreign places and peoples that became part of school curricula.

The emerging discipline was empowered by adventurous travelers who calculated geographical coordinates, measured temperatures in rainforests, recorded geological structures, pressed plant sheets, and brought back storerooms full of ethnographic 'curios'. Not all were missionaries, and some 'shot their way across Africa'. Leopold II of Belgium, who presided over the real world set for *The Heart of Darkness*, used part of his tainted wealth to build the incomparable Africana collection of Tervuren. Distasteful or unethical as the process and some of its actors were, the era of Victorian exploration produced a wealth of empirical data. It also placed a premium on fieldwork and engaged the academy in outreach via regular public lectures, serious travel publications, and a strong body of university courses presenting distant lands or unfamiliar but intriguing proximal regions.

During the 1880s, it became apparent that the classical era curriculum could not do justice to either the systematics of human geography or the accumulating wealth of ethnographic material. Then, in a space of two decades, different solutions were found. In Germany, geography diversified to accommodate human geography and, temporarily, ethnography (*Völkerkunde*). But in the USA, a separate field of anthropology emerged with the momentum of a German expatriate.

The Bridge between Human Geography and Anthropology

A major figure in this transition was Friedrich Ratzel (1844–1904), trained as a zoologist and an admirer of Darwin. He was appointed to the chair of geography at Leipzig (1886) but carried out his most influential work in what today is anthropology. Traveling widely, he worked as an itinerant reporter in the USA (1872–75), publishing *Urban and Cultural Images of North America* in 1876. With a first appointment in geography, he wrote his *Anthropogeographie* (1882–91) in which volume 1 organized and explicated the new subfield of human geography. Volume 2 attempted to deal with global cultural complexity by offering a novel social framework with concepts equivalent to migration, geographical speciation, adaptive radiation, and response to information exchange (diffusion). Evolutionary and historical, this work anticipated a systemic approach and some of its ideas influenced US anthropology and Berkeley over the next 60 years. This was accompanied by a great compendium on ethnography (1886–88; translated as *A History of Mankind*, 1896), and further writings on political geography.

Ratzel was an innovative, synthetic, and theoretical scholar who first attempted to create order from the chaos of a vast body of information on material culture and human difference. In the spirit of the time, he was Eurocentric, but occasional claims that he was a determinist, let alone a Nazi precursor, are based mainly on linguistic and cultural misreadings. It is now recognized that his views on culture areas and political structures were paralleled in French geography, under different labels.

The second key personality was Franz Boas (1858–1942) who, by comparison with Ratzel, was empirical and skeptical of theory. Based on long term fieldwork in the Arctic and Pacific Northwest, and complemented by extensive museum studies, he sought an 'insider' perspective on indigenous peoples and espoused a stance of cultural relativism, in *The Mind of Primitive Man* (1911). Moving almost directly from his doctorate in physics and geography in Germany to the USA in 1888, he became an influential professor of anthropology at Harvard in 1899. He strongly supported the clustering of cultural anthropology, together with cognitive linguistics, archaeology, and physical anthropology in a single discipline. While his ideas were scattered in papers, reports, and lectures, his dedication to indigenous studies and his interest in archaeology were imparted to his major students. Of these, Alfred Kroeber (1876–1960) had a significant impact on the creation of the Berkeley tradition and subsequently on some of the basic concepts of cultural ecology.

Anthropology and the Berkeley Tradition

Kroeber, who became professor of anthropology at Berkeley in 1919, was instrumental in bringing Carl Sauer (1889–1975) to the University of California, to found its department of geography in 1923. Yet Sauer's early signature publication, *The Morphology of Landscape* (1925), focused on an ongoing discourse in German geography and does not yet reveal the close connection that had developed between Sauer and Kroeber. But Kroeber's long term preoccupation with the indigenous peoples of California and his interest in archaeology did have a major impact on Berkeley geography. Sauer and his students turned to Latin America, however, where they established a durable tradition of Latin Americanist geography. That focus continued to evolve with sophistication and complexity among several generations of students, and Sauer's own views and emphasis on material culture are in part exemplified in *The Personality of Mexico*. The new Berkeley school diverged from the North American orientation of its 'midwestern' counterpart, and was critical in defining the legitimacy of American geography as a field oriented, international discipline.

While Kroeber did not share Boas' intellectual capacity, 'Berkeley geography' was partly rooted in Boas' empirical penchant and his non Eurocentric perspective. This served to give American geography, as a whole, its multicultural credentials. Sauer's enthusiasm for diffusion, a process he never explicated, was directly or indirectly indebted to Ratzel, who did give considerable thought to the agents of information exchange. Several deductive papers on human prehistory and its landscape implications reveal Sauer's abiding interest in archaeology, and he encouraged his students to pursue archaeological questions. Here the exchanges between Sauer, Kroeber, and other anthropologists at Berkeley must have been important, although Sauer's strong historical bent was evident long before he went to California.

Sauer's vague concept of culture has been criticized as 'superorganic' or *sui generis*, but most American anthropologists before the 1950s also did not delve into the nature of culture. Sauer's notion of 'culture hearth' (roughly equivalent to culture core or culture area), as a center of clustered and interconnected material phenomena, from which traits were dispersed, goes back to Ratzel, via Kroeber's essay on Clark Wissler.

A critique of the Berkeley tradition by Harold Brookfield is symptomatic of an impending sea change. Brookfield had spent years on New Guinea working with indigenous people, in collaboration with the British social anthropologist Paula Brown and several US anthropologists. He faulted the absence of participant observation and protracted interaction with indigenous groups by the Berkeley students. The trajectories of Brookfield's own PhDs from the New Guinea project, including Larry Grossmann and Eric Waddell, bear out that a new era of cultural and political ecology was dawning.

Cultural Ecology as a Reunion

In the wake of New Guinea, anthropologists and geographers began to rediscover common ground in what was then called cultural ecology. This approach concentrated on processes rather than material culture – subsistence, work, reproduction, and resource use – and their interrelations – as embedded in the rules and values of a particular society. The configurations could be functional, behavioral, or structural.

The functionalist method emphasizes a multiplicity of social, economic, and environmental variables for understanding decision making and problem solving, and views cultural adaptation as a reflection of human creativity. This emphasis is exemplified in the work of geographers Bill Denevan and Gregory Knapp, or the anthropologists Donald Hardesty and Andrew Vayda.

The premise of the behavioral approach is that decision making and the selection of alternative strategies or solutions are based on community choice, anchored in traditional values as much as they are in economic rationalization. This requires more intensive person to person field research and a degree of 'insider' under standing of perceptions, attitudes, beliefs, and mental processes. Here we mention the anthropologist Robert Netting, or the geographers Phil Porter and Karl Butzer, both of whom have worked closely with anthropologists, the latter also collaborating with archivist Elisabeth Butzer.

The structuralist perspective shifts the emphasis from the community to the socioeconomic or sociopolitical constraints in which individuals or communities can make decisions. Here, traditional values are subordinated to a political ecology, marked by different dialectics be tween technology and development, or between pro ductive sector and multinationals, access to markets and services, and limited choices for marginalized communities.

These three forms of cultural ecology are comple mentary but embrace different questions, subjects, and methodologies, to the point where epistemologies may differ. The functional is to varying degrees systemic and normative; the behavioral is non normative and some times postprocessual; while the structural is critical, more deductive, and increasingly focused on development studies. Each perspective yields different insights, that reflect several scales of analysis, and diachronic versus future concerns. All are pertinent to understanding en vironmental behavior, with the behavioral approach of particular importance for global environmental change, since it can provide historical monitoring of processes and trends.

Cultural ecology draws heavily from both anthro pology and geography, but it is not a logical extension of the Berkeley tradition, and is only nominally linked to the cultural ecology of anthropologist Julian Steward, a student of Kroeber. Instead, anthropology and geography have come to recognize common interests, sharing methods and objectives, and applying new levels of so phistication to the problems of change and transforma tion that defied adequate comprehension a century ago. This is apparent in a renewed exploration of the methods of ethnography on the part of human geography.

Geoarchaeology, Environmental History, and Causality

A very different intersection between anthropology and geography came about in the form of geoarchaeology, another integrative approach. Although geoscience ap plications to problems of dating, environmental re construction, and spatial organization in archaeology are common (mainly as 'archaeological geology'), and rapidly gaining in popularity, direct attention to the nature of the archaeological record is not. That record is a proxy for human biological and cultural evolution which identifies subsistence and behavioral patterns, encoded in artifacts and the like, that are embedded in sediments. Geoarch aeology can provide expertise to help evaluate the in tegrity of such associations, and place sites into a changing physical and human environment. Prehistoric and historical activities may not only disturb the en vironment, but they play out in space and time, at dif ferent scales and amid unstable conditions. The site and the region become the context for aggregate activity and of transformations such as domestication, land use, or adaptation, so as to restrict the realm of speculation about how change 'happens'.

Two corollaries follow. For one, parallel experience from cultural ecology can be applied to transcend de ductive 'accepted truths' as to how 'degradation' is per ceived and managed in other cultures. This raises caveats about purported cause and effect relations, specifically: (1) what is or is not a disturbed environment; (2) when do or do not people practice destructive land use; and (3) how judgmental is a 'conclusion' about cause and ef fect, given several environmental, social, and political factors that favor similar outcomes? A second question concerns the convergence of possible mismanagement, not with climatic shifts, but with unusual and recurrent wea ther events such as episodes of excessive precipitation. Do climate and weather cause, precondition, or trigger soil erosion, with or without nonconservationist land use?

The potential applications of geoarchaeology to in terpret environmental history, evaluate longer term trends and predict scenarios for sustainability are con siderable. Such research ultimately depends on close collaboration with archaeologists and promises new perspectives for the practice of environmental history. It also identifies interactive opportunities between cultural and political ecology.

See also: Anthropogeography (After Ratzel); Berkeley School; Diffusion; Land Change Science; Political Ecology; Sauer, C.

Further Reading

Brookfield, H. C. (1964). Questions on the human frontiers of geography. *Journal of Economic Geography* 40, 283 303.

Butzer, K. W. (1982). *Archaeology as Human Ecology*. Cambridge: Cambridge University Press.

Butzer, K. W. (2005). Environmental history in the Mediterranean world: Cross disciplinary investigation of cause and effect for degradation and soil erosion. *Journal of Archaeological Science* 32, 1773 1800.

Butzer, K. W. (2008). Challenges for a cross disciplinary geo archaeology. *Geomorphology*. 101, 402 411.

Crang, M. and Cook, I. (2007). *Doing Ethnographies*. London: Sage.

Knapp, G. W. (1991). *Andean Ecology: Adaptive Dynamics in Ecuador*. Boulder, CO: Westview.

Kroeber, A. L. (1931). The cultural area and age area concepts of Clark Wissler. In Rice, S. A. (ed.) *Methods in Social Science*, pp 248 265. Chicago: University of Chicago Press.

Mercier, G. (1995). La région de l"état selon Friedrich Ratzel et Paul Vidal de la Blache. *Annales de Géographie* 583, 211 235.

Sauer, C. O. (1941). The personality of Mexico. *Geographical Review* 31, 353 364.

Sauer, C. O. (1971). The formative years of Ratzel in the United States. *Annals of the Association of American Geographers* 61, 245 254.

Stocking, G. W. Jr. (ed.) (1974). *A Franz Boas Reader: The Shaping of American Anthropology 1883 1911*. Chicago: University of Chicago Press.

Anti-Geopolitics

U. Oslender, University of Glasgow, Glasgow, UK

Origin and Context of the Term Anti-Geopolitics

The term anti geopolitics emerged in the English speaking literature in the mid 1990s. Although the term is briefly mentioned by Ó Tuathail in 1996 (then spelt without hyphen as antigeopolitics), it was only in 1998 that it was conceptually developed in the first edition of the *Geopolitics Reader*, edited by Ó Tuathail, Dalby and Routledge. Subsequently, Routledge has published slightly edited versions of this first elaboration. Various authors have since made reference to the term anti geopolitics, yet mostly without engaging conceptually with it in any great detail. Anti geopolitics should be regarded as a recent and historically contingent concept that can still be developed further. It must also be seen in the wider context of the evolution of geopolitical thought and practice.

As its prefix anti suggests, it relates critically to dominant geopolitical thinking and discourse. As such, the notion of anti geopolitics must be placed within the school of critical geopolitics that itself emerged in Anglo American geographical writing in the 1980s. This school of thought offers a radical critique at the ways in which geographies of global politics are constructed through discourses and representational practices of statecraft intellectuals. It sees itself as a way of dealing critically with the often silenced imperial and colonial heritage of the discipline of geography and the ways in which it has been conscripted into political service. Even though many practitioners of critical geopolitics mentioned below do not employ the term anti geopolitics, it is important to understand the context and intellectual soil out of which this notion arose.

Lacoste, *Hérodote*, and Critical Geopolitics

Geography and geographical reasoning are central to political power, military strategy, and warfare. As the French geographer Yves Lacoste memorably put it in a polemic publication in 1976 – seen by many as a revolutionary manifesto for geography – *la géographie*, *ça sert*, *d'abord*, *à faire la guerre*, the purpose of geography is above all warfare. What this provocative title and book intended to challenge was the commonplace assumption that academic geography was essentially about the production of objective, universal knowledge that would provide the citizenry with basic geographical facts, while the discipline's military logic was hidden or ignored. Breaking with this lopsided understanding, Lacoste (1985) argued that geography should rather be about knowing how to think about space in order to know how to organize and fight there ("*savoir penser l'espace pour savoir s'y organiser*, *pour savoir y combattre*").

Lacoste was one of many radical voices based at the experimental University of Paris VIII at Vincennes that arose as a result of the student revolt of 1968 and encouraged critical, Marxist, and anarchist thinking. In the same year, 1976, he founded the radical geographical journal *Hérodote*. The launch of this journal was to transform French geography away from the distinct Vidalian tradition of regional geography toward more critical perspectives on the role of geographical reasoning in policy circles and military strategy. Its first issues were thus dedicated to liberation struggles and revolutionary movements in the Third World. Although Lacoste did not use the term anti geopolitics, these are the very realms of anti geopolitical struggle, according to later enunciations by scholars from the Anglophone critical geopolitics school. There is then an important overlapping of concerns between the French radical geographical vision *à la* Lacoste and *Hérodote* on the one hand, and Anglophone critical geopolitics on the other, even though a mutual lack of engagement has to be deplored between these two radical geographical schools of thought.

Modern Geopolitical Discourse, the Cold War, and Popular Geopolitics

Anti geopolitics in its broadest meaning has been suggested to embrace all kinds of resistances to the practices of geopolitics. As such it does not only concern itself with the material counter hegemonic struggles of anticolonial, anti imperial, independence, and liberation movements. Since geopolitics deploys representations of the world that serve the interests of the ruling political and economic elite, anti geopolitics pretends to challenge the very discourses and visualizations of these representations. The beginnings of modern geopolitical discourse can be traced back to encounters between Europeans and non Europeans since the fifteenth century. As Agnew and Corbridge argue, modern geopolitical discourse is characterized by its representation of others as backward and permanently disadvantaged. This discursive strategy fixes non Europeans in a constant state of inferiority and savagery, out of which they can only emerge by changing into something they are not and adapting European traits. Such a powerful discursive separation of the world into Europeans and non Europeans underlies much racist, xenophobic, and antisemitic thought until today.

Agnew later went on to suggest that the most dis tinguishing feature of the modern geopolitical imagina tion is a global visualization without which world politics would not be possible. During the Cold War, for example, Western geopolitical discourse divided the world into three broad categories: 'the East', which stood for Com munism and totalitarianism, portrayed as an essentially negative (even evil) world view; 'the West' was equated with freedom, democracy, and a positive image; and 'the Third World', which lay somewhere in between as the space where the battle between 'East' and 'West' was played out. Such a simplification of the spatial organ ization of the world may appear almost too naïve to be true. Yet, it constituted a powerful global visualization that oriented foreign policy and justified political, eco nomic, and military interventions by 'the West' and 'the East' elsewhere. Moreover, these practices of dividing the world into a simplified pattern of easily understandable opposites were not restricted to political elites and their advisors. They were also reproduced in the geopolitical cultures of the everyday. A popular geopolitics emerged that reproduced such commonsense geopolitical rea soning in films, television, novels, newspapers, and journals.

The War on Terror and Moral Boundaries

Anti geopolitics intends to provide an interrogation of such a global visualization and the ways in which these representational practices produce physical and moral boundaries that divide the world into 'us' and 'them'. Today such moral divisions can, maybe best, be seen in the discourses of the so called 'war on terror' that have revived in many ways the dichotomous discursive sep aration practices of the Cold War period. US President George W. Bush's announcement in September 2001 – 9 days after the terrorist attacks on the Pentagon and the World Trade Center – that "either you are with us or you are with the terrorists" set the tone for the current geo political discursive practices of drawing physical and moral boundaries. Bush's geopolitical speech act forces the audience to take sides and face the consequences of doing so.

The global reach of the geopolitical terror discourse has led to significant discursive changes in the repre sentation of political conflicts in other parts of the world, too. In South America, for example, Colombian President Uribe declared the country's largest guerrilla movement, the Revolutionary Armed Forces of Colombia (FARC), to be narcoterrorists. This categorization has subsequently been used by the mainstream media at home and abroad to such an extent that it has become a commonplace assertion to link this rebel movement with illegal drug trafficking and terrorist activities. While some of these links undoubtedly do exist, this discursive shift has re duced a very complex political conflict in Colombia to one in which the State seemingly fights drug traffickers who use terrorist means. Gone in this geopolitical rea soning and representation are the historical roots and continued relations of social injustice and exploitation that lie at the heart of the political struggle in Colombia, one of Latin America's longest running conflicts. An anti geopolitical perspective would challenge the ideologic ally motivated narcoterrorist discourse employed by Colombia's government and expose the complexities underlying the political and social conflict in Colombia.

Anti-Geopolitics as (Radical) Politics

The policies and strategies of empires, colonial powers, dictatorships, and states have rarely gone unchallenged by those who have been subjects to various degrees of domination and exploitation. Quite on the contrary, dominating power has almost always faced resistance from various forms of counter hegemonic struggles. Proponents of anti geopolitics focus on the latter as enacting a geopolitics from below that emanates from subaltern positions within society that challenge the military, political, economic, and cultural logic of hege monic power. Those contestations can be violent or nonviolent. They include armed struggle against op pressive outside forces or state structures, as well as peaceful mass demonstrations organized by social movements. Anti geopolitics intends to be a wide open container in which all kinds of protests are considered that organize material and discursive challenges to dominant geopolitical power and representations. Par ticular attention has been paid to anticolonial and anti imperial struggles, nationalist rebellions, and revolutions, as well as to more recently emerging globalizing resist ance movements and transnational activism networks.

Anticolonial and Anti-Imperial Struggles

The imperial and colonial geopolitical order implied a spatialization of the world organized around the interests of Europeans. Although implicitly a violent form of domination that was established and maintained through the physical control, coercion, and co optation of the colonial subjects, colonialism and empire building were seen by the dominating powers as a legitimate, necessary, and even beneficial enterprises for the world at large. Colonialism was not only about the appropriation of natural resources in the colonized lands, but it also saw itself as a civilizing mission on earth to rescue the native savage from ignorance, sin, and superstition. It constituted a vision for the world, in which Europeans would spread civilization into the darkest corners of the planet. The brutal missionary zeal of the Spanish Empire in the Americas, for example, attests to this global visu alization of the modern geopolitical imagination under lying the colonial and imperial logic.

Algeria and Frantz Fanon

Anticolonial and anti imperial struggles were thus not only directed against the physical presence of the col onizer on colonized soil. They also challenged the geo political vision and the spatialization of the world that underlies colonialism. As the Martinique born French writer, medical doctor, and psychiatrist Frantz Fanon explained in his incisive critique of colonialism, in de colonization there is the need of a complete calling into question of the colonial situation. While working as Head of the Psychiatry Department at the Blida Joinville Psychiatric Hospital in Algeria in 1953, and confronted with the brutalities and torture of the French colonial system in Algeria, Fanon developed his theory of the psychology of colonial domination. In 1956, he resigned his post with the French colonial authorities to work for the Algerian independence movement from Tunisia and Ghana. Fanon's life history epitomizes the two fronts of anti geopolitical struggle: (1) the material challenge of dominant geopolitical power and (2) contestations of geopolitical representations of the spatialized normative world order. First, he actively collaborated with the armed struggle of the Algerian National Liberation Front (FLN), as Ambassador to Ghana for the Provisional Al gerian Government he helped to establish a southern arms supply route. Second, he vociferously denounced the geopolitical vision of the French imperial project through his writings and publications. Fanon was par ticularly inspired by the victory of the Vietnamese Viet Minh Communist revolutionary forces against the French colonial power in the battle of Dien Bien Phu in 1954, which effectively ended the First Indo China War and ushered in French withdrawal from its Indo Chinese colonies. Yet, Fanon would not live to see Algerian in dependence, as accorded between France and the FLN in March 1962. He died of leukemia in December 1961. His writings have inspired anticolonial liberation movements worldwide.

India and Mohandas Gandhi

Another influential anticolonial struggle, though of a different kind, emerged on the Indian subcontinent. Although the Indian independence movement involved a wide array of political organizations, it became most associated with the figure of Mohandas (or Mahatma) Gandhi. Rejecting armed struggle as an option, Gandhi advocated nonviolent mass civil disobedience against British colonial rule, a strategy of resistance that he termed *Satyagraha* (a Sanskrit compound meaning truth and firmness). A British educated lawyer, Gandhi first gained experience in peaceful civil disobedience in South Africa during the civil rights struggles of Indian communities there. On his return to India, he helped to mobilize poor farmers and laborers against widespread discrimination and oppressive taxation. As leader of the Indian National Congress, Gandhi led nationwide campaigns for India's independence from foreign dom ination, which was eventually achieved in August 1947. Gandhi remained committed to nonviolence throughout his life. The form of protest he advocated was a very embodied one, employing long fasts as a form of self purification and resistance. His life and teachings have been an inspiration for many civil rights leaders and social movements to come. From an anti geopolitical perspective one may argue that not only did the Indian anticolonial movement achieve independence from Britain, a significant material victory, but also the phil osophy and practice of nonviolent resistance provided a model of subaltern politics that successfully challenged military geopolitical power and the underlying dominant geopolitical vision and logic of maintaining colonial order by force. More so than the armed struggle in Algeria maybe, *Satyagraha* is an expression of an anti geopolitical perspective put into practice, as it chal lenges the very geopolitical logic of warfare and violence.

Nationalist Rebellions and Revolutions

Cuba

To proponents of anti geopolitics, postcolonial nation alist rebellions and revolutions also provide a focus for analysis, in particular those enframed by Cold War geopolitics. The Cuban revolution of 1959 is a case in point. Led by Fidel Castro, *Che* Guevara, Camilo Cien fuegos, and others, a group of guerrilla fighters who had landed on the shores of Eastern Cuba in the yacht *Granma* in December 1956 began an armed struggle against Cuban dictator Fulgencio Batista. Helped by local peasants and with support from Batista opponents in Havana, the revolutionaries finally drove out the dictator who fled to the Dominican Republic on 1 January 1959. The initial aim of the *barbudos*, the bearded guerrillas, was to rid Cuba of a brutal dictatorship and affirm national, political, and economic interests against the oppressive interference of the United States. As such the revolution posed a significant material challenge to US foreign policy and the geopolitical world order in the Americas. In his amendment to the original Monroe Doctrine of 1823, US President Theodore Roosevelt had asserted in 1904 the right of the United States to intervene in the countries of the Americas in order to stabilize their economies. In reality, however, the Roosevelt Corollary was little short of being a license for the US to practice its own form of colonialism. Subsequent interventions in the first half of the twentieth century in Cuba, Nicaragua, and Haiti were justified by recourse to the Roosevelt Corollary. The Cuban Revolution of 1959 therefore marked a clear opposition to the US imposed geo political order.

Nicaragua and El Salvador

Inspired by the success of the Cuban revolution, similar challenges were raised by guerrilla movements in Nicar agua and El Salvador. In Nicaragua, the Sandinista Na tional Liberation Front (FSLN) overthrew the dictatorship of Anastasio Somoza in 1979. They saw themselves as concluding the unfinished business of revolutionary leader Augusto César Sandino who opposed US occupation of Nicaragua in the early twentieth cen tury before he was assassinated by the US trained Na tional Guard in 1934. In El Salvador, the Farabundo Martí National Liberation Front (FMLN) was founded in 1980 as a union of various guerrilla movements with the aim of overthrowing the US supported government. After a ser ies of military offensives a stalemate was achieved in 1989 that forced negotiations between the FMLN and the Salvadoran government, leading eventually to Peace Ac cords and a ceasefire in 1992. Since then the FMLN has been a political party aiming at taking control of the country through democratic elections.

Cold War geopolitical battles

Both the FSLN in Nicaragua and the FMLN in El Sal vador inherited the radical tradition of the Cuban revo lutionaries to challenge US geopolitical dominance in the Americas. As such, they may be seen as examples of an anti geopolitical approach to US hegemony in the region. However, all three revolutionary experiences were sim ultaneously enframed by Cold War geopolitics. The US response to these challenges was to declare them a Communist threat to be opposed. In the case of Cuba, the US imposed economic sanctions and a trade embargo on the Caribbean island in 1962, which are still in force today. This, in turn, drove Fidel Castro's vision of a free and independent Cuba into the geopolitical arms of the Soviet Union's sphere of influence. In the case of Nicaragua, US President Ronald Reagan accused the Sandinistas of im porting Cuban style socialism. The US consequently provided significant military and financial assistance to the Contras, a group opposing the FSLN, which resulted in a long and bloody civil war. Reagan's support for the Con tras formed part of his wider strategy to actively support and promote movements that opposed Soviet influenced governments. The nationalist struggles of Cuba and Nic aragua were thus drawn into the geopolitical battles of the Cold War period.

Social Movements and Transnational Activism Networks

A further field of anti geopolitical struggle is said to exist in the myriads of social movements, human rights groups, land rights movements, environmental groups, neigh borhood associations, and others that challenge the power of the state, transnational companies, and the worst ex cesses and effects of global neoliberalism. Different from anticolonial struggles and nationalist rebellions, these challenges are mostly located in the sphere of civil so ciety. They do not generally aim at overthrowing gov ernments, but are concerned with achieving significant reforms within an existing governmental system. In the social movement literature, many of these struggles have been grouped under the term 'new social movements', because they are frequently cultural struggles over the practices and meanings of everyday life.

Globalization from below

While most movements and groups mobilize around a specific issue – from housing for the poor in Douala, preservation of indigenous cultures in the Amazon, land rights for landless peasants in Brazil, protests against dam construction in India, to an immigration bill in the US – many also connect to wider issues. They recognize that their particular grievance is just one cog in the machine of a neoliberal economic globalization that favors trans national capital at the expense of workers' rights, com munity values, and the environment. As these groups make these links and articulate their particular struggles against the more structural macroeconomic logic of late capitalism, it has been argued that these resistances embody a globalization from below. This entails a dif ferent vision of the world, one that challenges the geoeconomic discourse of transnational liberalism and the geopolitical representations underlying it. As such, these resistances have also been seen as a form of geo politics from below, or anti geopolitics of protest.

The World Social Forum

The World Social Forum, held annually since 2001 in Porto Alegre, Brazil (except in Mumbai in 2004), is maybe one of the most exciting developments in this transnationalization of protest beyond borders. The Forum effectively constitutes a convergence space for globalizing anticapitalist resistance that connects a large number of social movements, activists, NGOs, and trade unions across space and beyond state boundaries to ar ticulate protests globally against the privatization of every aspect of life, and the transformation of every ac tivity and value into a commodity. It draws up alternative visions to the dominant global neoliberal project and aims to show that another world is possible. The success in mobilization has been attributed to the decentralized, nonhierarchical and web like structure of such a movement of movements. This mobilization has been significantly facilitated by the use of the internet, which allows for local, regional, and national real life experiences on the ground to connect and cooperate.

Another important worldwide coordination network of (anti geopolitical) resistance to the global market is Peoples' Global Action (PGA). PGA was launched in February 1998 in Geneva by grass roots movements from all continents as an alliance of struggle and solidarity against free trade and the World Trade Organization, and as a global tool for communication and coordination to build local alternatives to globalization. PGA defines itself via five hallmarks that include a clear rejection of capitalism, imperialism, and all forms of domination and discrimination, and a call for direct action and civil disobedience. Its organizational philosophy is based on decentralization and autonomy.

Anti-Geopolitics as Perspective

Geopolitics has always been a contentious form of knowledge and practice, and there have been many critiques to dominant geopolitical reasoning. These range from the orthodox Marxism of the German communist Karl Wittfogel in the 1920s to the poststructuralist interventions of the critical geopolitics school of the 1980s. More recently, feminist geographers have called for a feminist geopolitics. They expose the historical reasoning of geopolitical thought and practice as masculinist and advocate an embodied geopolitics that rewrites the lived experiences of women and other marginal voices back into geopolitical arguments.

These critiques have given rise to different ways of thinking geopolitically, or to anti geopolitical forms of knowledge. Since geopolitics is not only about politics *per se* but also about the representation of the world and a normative world view, anti geopolitical discourse also intends to disrupt these representations. In other words, anti geopolitics not only sees itself as radical politics opposed to dominant geopolitical practices, but also as a differential perspective undermining traditional geopolitical reasoning. Some of these perspectives are briefly reviewed here.

Dependency Theory and Development Critique

Dependency theory emerged in the 1960s as an influential, complex body of theoretical concepts with structuralist and Marxist roots that explained Latin America's underdevelopment in terms of a structural logic inherent in the development of global capitalism. It built on arguments developed in the Economic Commission for Latin America that the global system was not a uniform marketplace but divided structurally between rich and poor economies – a center of industrialized nations and a periphery of primary producers. Through this system, all of the benefits of technology and international trade would accrue to the center. Dependency theorists regarded the economic development of the periphery within such an unequal world system a nearly impossible task. According to André Gunder Frank, one of its principal exponents, the metropolis (or the core) creates and exploits peripheral satellites, from which it expropriates economic surplus for its own economic development. The satellites remain underdeveloped for lack of access to their own surplus and as a consequence of the exploitative contradictions that the metropolis introduces and maintains in the satellite's domestic structure.

Dependency theory has been critiqued for its overly structuralist explanatory approach to uneven development. Yet it is hard not to agree with its principal insights over the highly structured unevenness of global economic development and its underlying capitalist logic in the production of space. As a sustained theoretical critique of global capitalism, dependency theory constituted an important intellectual challenge to US hegemony and political and economic intervention in the Americas. It thus also disrupts the geopolitical and geoeconomic vision of US sponsored global capitalism. In this sense dependency theory can be regarded as an anti geopolitical perspective.

The Anti-Geopolitical Eye

One of the most prolific writers of the critical geopolitics tradition, Gearóid Ó Tuathail coined the term anti geopolitical eye in 1996. With that he refers to an alternative way of seeing and representing that disturbs the all seeing eye usually evident in geopolitical arguments. The anti geopolitical eye acknowledges its view from somewhere specific. It represents a situated view of the world that rejects the detached perspectives of statesmen. Instead it travels at ground level to observe, feel, and transmit a moral proximity to those observed and described. Ó Tuathail considered the anti geopolitical eye a provisional category, which he introduced to analyze the impassioned reports of the war in Bosnia that the British journalist Maggie O'Kane sent back to the readership in Britain from the frontlines. Her reporting was based on an eyewitness approach of multiple perspectives. The victims of the war in Bosnia were present in her visceral descriptions of the horror of war. The suffering of the victims entered graphically into her reporting, which accused, took sides with the victims, and did not try to hide behind the high minded abstractions of traditional geopolitical discourse. The anti geopolitical eye breaks down the distance between observer and observed. It reveals a moral proximity that accuses and asks awkward questions enquiring about ethical responsibilities that are often ignored.

Feminist Geopolitics

With its emphasis on marginal voices that are usually silenced in the big geopolitical stories and its embodied vision of lived experience on the ground, the category of the anti geopolitical eye has attracted the attention of feminist geographers. Some consider it a first step in critically addressing the marginalization of such voices and in recovering viewpoints that are otherwise hidden in geopolitical meta narratives. They see the anti geo political eye as a way of embodying geopolitics and to give it that visceral feeling of life histories transmitted from close up. It is out of this necessity that the call for a feminist geopolitics emerged at the turn of the millen nium, challenging the intrinsic masculinist character of geopolitical arguments and reasoning. A feminist geopolitics perspective argues that women's bodies are caught up in international relations, but often at the mundane level of the everyday. As a result, they have not been written into the texts of geopolitical discourse. Analytical attention grounded in the everyday would therefore recover women's experiences and those of other marginalized people, it is argued. Specific attention in this perspective is paid to the scale of analysis. Feminists argue that geography's understanding of space has been significantly transformed by examining the world through the scale of the body. A feminist geo politics thus stresses the need to analyze geopolitical relationships at different scales, privileging the body as the site of performance in its own right. It therefore provides a radical departure from traditional geopolitical perspectives.

Critique of Anti-Geopolitics

The notion of anti geopolitics is a very recent one, barely 10 years old. Some of its associated ideas have attracted analytical attention, such as the anti geopolitical eye in feminist geopolitics. The term anti geopolitics in itself, however, has not found widespread application to date. Its use has mainly been restricted to a handful of writers from the Anglo American critical geopolitics perspective. It has occasionally been referenced in articles, mostly, however, without a deeper analytical engagement. In a few cases, it has also led to very disparaging and confused interpretations.

This reception may partly be explained by the fact that the prefix anti suggests a dichotomy to geopolitics that appears to be rather unhelpful. The protest move ments and differential perspectives gathered under the term anti geopolitics do not oppose geopolitics *per se*, but rather dominant geopolitical practices, discourses, and representations. In fact, they themselves are geopolitical. Other terms that have described this oppositional atti tude include counter hegemonic geopolitics, geopolitics from below, countergeopolitics, alternative geopolitics, anti imperialist geopolitics, geopolitics of and for social movements, other geopolitics, and so on. Whereas the adjectival notation of anti geopolitical can be useful – for example in the notion of the anti geopolitical eye – the noun anti geopolitics is ambiguous and can be mislead ing. As a historically contingent concept of geopolitical thought and practice, it requires further critical devel opment and elaboration.

See also: Cold War; Critical Geopolitics; Geopolitics; Lacoste, Y.; Postcolonialism/Postcolonial Geographies; Resistance; Social Movements.

Further Reading

Agnew, J. (2003). *Geopolitics: Re visioning world politics* (2nd edn). London: Routledge.
Agnew, J. and Corbridge, S. (1995). *Mastering space: Hegemony, territory and international political economy*. London: Routledge.
Agnew, J., Mitchell, K. and Ó Tuathail, G. (eds.) (2003). *A companion to political geography*. Oxford: Blackwell.
Dodds, K. and Atkinson, D. (eds.) (2000). *Geopolitical traditions: A century of geopolitical thought*. London: Routledge.
Dowler, L. and Sharp, J. (2001). A feminist geopolitics? *Space & Polity* 5(3), 165 176.
Fanon, F. (1963). *The wretched of the Earth* (trans. Farrington, C.). New York: Grove Press.
Frank, A. G. (1969). *Latin America: Underdevelopment or revolution: Essays on the development of underdevelopment and immediate enemy*. London: Monthly Review Press.
Girot, P. and Kofman, E. (eds.) (1985). *International geopolitical analysis: A selection from Hérodote*. Paris: L'Harmattan.
Lacoste, Y. (1985). *La géographie, ça sert, d'abord, à faire la guerre* (3rd edn). Paris: Éditions La Découverte.
Ó Tuathail, G. (1996). An anti geopolitical eye: Maggie O'Kane in Bosnia, 1992 93. *Gender, Place and Culture* 3(2), 171 185.
Ó Tuathail, G. (1996). *Critical geopolitics: The politics of writing global space*. London: Routledge.
Ó Tuathail, G., Dalby, S. and Routledge, P. (eds.) (2006). *The geopolitics reader* (2nd edn). London: Routledge.
Routledge, P. (2003). Anti geopolitics. In Agnew, J., Mitchell, K. & Ó Tuathail, G. (eds.) *A companion to political geography*, pp 236 248. Oxford: Blackwell.
Routledge, P. (2006). Anti geopolitics: Introduction to part five. In Ó Tuathail, G., Dalby, S. & Routledge, P. (eds.) *The geopolitics reader* (2nd edn). pp 233 248. London: Routledge.
Wittfogel, K. (1985) (1929). Geopolitics, geographical materialism, and Marxism (trans. Ulmen, G. L.). *Antipode*. *A Radical Journal of Geography* 17(1), 21 72.

Relevant Websites

http://www.agp.org
Official website of Peoples' Global Action, a worldwide coordination network of grass roots movements and resistance to free trade and the World Trade Organization.
http://www.antipode online.net
Official website of *Antipode*, a radical journal of geography. Launched in the late 1960s, Antipode is still (or once again) the flagship for radical (anti geopolitical) thinking and debate in Anglophone geography.
http://www.forumsocialmundial.org.br
Official website of the World Social Forum, convergence space for

globalizing anti capitalist resistance that connects social movements, activists, NGOs and trade unions across space and beyond state boundaries to articulate protests globally.

http://www.herodote.org

Official website of *Hérodote*, French journal of geography and geopolitics, launched in 1976 by Yves Lacoste and named after the Greek geographer and historian Herodotus. It was pivotal in the critical regeneration of geopolitical discourse in France. Its first issues in particular were devoted to the study of post colonial wars and revolutionary movements in the so called Third World.

http://www.indymedia.org

Indymedia is a global, alternative, Internet based media outlet that promotes radical political activism. Its origin lies in the temporary alternative Independent Media Center (IMC) that emerged during the mobilization against the World Trade Organization in Seattle in 1999 to counter distortions of the demonstrations in the mainstream media. It quickly evolved into a global network of over 160 local and thematic IMCs worldwide.

Anti-Urbanism

T. Slater, University of Edinburgh, Bristol, UK

Glossary

Anomie A term popularized by the sociologist, Emile Durkheim in his 1897 book *Suicide*, and used to describe a condition in individuals characterized by an absence or diminution of standards or values (referred to as 'normlessness'), and an associated feeling of alienation. He believed that 'anomie' is common when the surrounding society has undergone significant changes in its economic fortunes, whether for good or for worse and, more generally, when there is a significant discrepancy between the ideological theories and values commonly professed and what was actually achievable in everyday life.

Anti-Urbanism A discourse of fear of the city, produced and reproduced through a variety of negative literary, artistic, media, cinematic, and photographic representations of urban places.

Ash Can School An artistic movement that came into prominence in the United States during the early twentieth century, best known for works portraying scenes of daily life in poor, crowded urban neighborhoods. The movement is most associated with a group of urban Realists known as 'The Eight', comprising Robert Henri, Arthur B. Davies, Maurice Prendergast, Ernest Lawson, William Glackens, Everett Shinn, John Sloan, and George Luks.

Discourse A network of signs, symbols, and practices through which we make our world(s) meaningful to ourselves and to others.

Film Noir An American cinematic genre with roots in the fiction of Depression America, tabloid photojournalism, the bombed landscapes of World War II cities, and the anti-urban visions of European directors, cinematographers, and writers who emigrated to America in the 1930s. Film noir cities were usually blighted, dark, empty, threatening, miserable, and dystopian.

Realism A mode of representation in the visual arts and literature which was particularly prominent in Europe and North America in the nineteenth and early twentieth centuries. In painting, the aim of Realists was to give a truthful, objective, and impartial representation of the real world, based on meticulous observation of contemporary life.

Transcendentalism A group of ideas in literature, religion, culture, and philosophy that emerged in New England in the early- to mid-nineteenth century. The ideas were rooted in the philosophy of Immanuel Kant (and of German Idealism more generally), which the New England intellectuals of the early nineteenth century embraced as an alternative to the doctrines of the Unitarian church. Transcendentalists desired to ground their religion and philosophy in principles not based on, or falsifiable by, sensuous experience, but deriving from the inner, spiritual or mental essence of the human. Nature was central to transcendentalist thought, and was viewed to have the powers, status, and authority traditionally attributed to an independent deity.

Introduction: 100% Urban Proof

In early 2007 Nissan launched a new model of sport utility vehicle (SUV), the Qashqai. The massive in your face marketing campaign described the Qashqai as "100% Urban Proof" – posters were designed showing a triumphant tank like vehicle protected from gritty looking streets, with the backdrop of walls plastered in menacing graffiti (**Figure 1**). The website takes the 'urban proof' theme even further, showing it roaring around a deserted industrial district accompanied by sinister music, where creepy characters in the graffiti come to life to throw out all kinds of obstacles (oil slicks, road bumps) for the Qashqai to negotiate. It does so with ease, for Qashqai "claims the urban environment as its own playground," and its "dynamic profile, flared wheel arches and durable black cladding protect it from the trials and tribulations of urban life." Furthermore, there is an optional "electronic

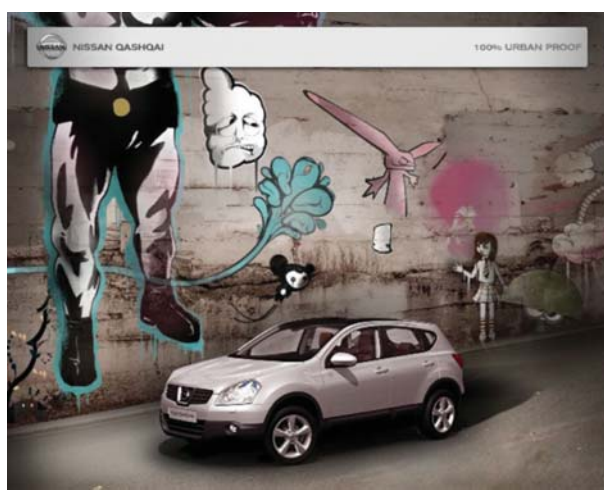

Figure 1 The Nissan Qashqai 100% Urban Proof. Reproduced with permission of the Nissan Corporation.

stability program," which "enhances this surefootedness by keeping on course when the unexpected happens."

The Qashqai is a pure exemplar of the discourse of anti urbanism materialized into a product that attempts to provide safety for consumers perceived as afraid of urban places. The message from Nissan is clear: in the dark and dangerous corners of the urban jungle, you will be fine inside this quasi tank off roader! The sense of being in control is amplified by the section of the website entitled "My Qashqai," where you can even design your 'own' urban background for your vehicle. Urban fears can be eased or erased when you go out and buy a Qashqai – just wait and see!

Anti urbanism is best defined as a discourse of fear of the city, and something fuelled by the impact of images of urban dystopia we see in a variety of media, cinematic, literary, artistic, photographic – and in the case of the Qashqai, corporate – representations of urban places. It is a discourse that has been around for a long time, in con junction with the emergence of the industrial city, and often constructed in relation to the 'good city' of the an cient Greeks, and especially the perceived virtues of rural life. Anti urbanism is particularly advanced in the United States in a variety of guises, from the celebration of rural small town kinship and community to the fact that Los Angeles has been completely destroyed 138 times in vari ous motion pictures from 1909 to 1999! Critical analyses of anti urbanism are vital if the material consequences of widespread urban fears are to be exposed and challenged. As cultural geographers have argued for a long time now, if we leave powerful representations unquestioned, then supposedly fixed 'evidence' about how a society is organ ized can very easily become treated as overwhelming evi dence of how it 'should', or 'must' be organized.

Origins: The Good Life

In order to understand where anti urban sentiment came from, we must appropriately invert one of the key prin ciples of the natural sciences (especially geology), uni formitarianism, often summed with the statement: 'the present is the key to the past'. For our purposes, the past is the key to the present, and therefore we must acknowledge the widespread and deeply entrenched pro urban sentiment of the past to grasp any anti urban sentiment of the present.

Athens (*c.* 500–400 BC) was not only the birthplace of democracy, it was the birthplace of the arena in which democracy was born – the city state or the *polis*, captured by Aristotle as a self sufficient and self governing group of villages in a narrow and closed region lying around an urban center:

> The partnership finally composed of several villages is the city state; it has at last attained the limit of virtually complete self sufficiency, and thus, while it comes into existence for the sake of life, it exists for the good life ….it is clear that the city state is a natural growth, and that man is by nature a political animal, and a man that is by nature and not merely by fortune citiless is either low in the scale of humanity or above it. (P. Hall, 1998: 2)

For our purposes we are less interested in the morphology of the city than the moral geography of this statement – the city as the "natural" and "good" unit for the development of human society; indeed, Aristotle went on to say that anyone who lives outside such a unit must be "either a god or a beast." He was not alone – Socrates and Plato before him had both treated Athens as the arena where 'citizens' where born and nurtured, a place where public and private life were in rhythmic interplay for the first time, and a place which supported an unprecedented flowering of art, music, intellect, politics, and love. The Athenians demonstrated for the first time the capacity, richness, and depth of the human mind in the theater of the city. It didn't matter that Athens also harbored extremes of rich and poor – the possibilities for transcending inequality were legion be cause of the possibilities the city offered for human innovation.

The urban concentration of indisputable genius ex hibited by the Athenians had a lasting impact on how future societies throughout human history treated cities. No city in history has ever been devoid of human suf fering, but that hard reality was for centuries trumped by the fact that no city that followed the Athens example has ever been starved of the same basic conditions for soci etal advancement brought about by the clustering of citizens with divergent thoughts, abilities, talents, and politics under a democratic model. From Imperial Rome to Renaissance Florence to Elizabethan London, pro urban sentiment was articulated and reproduced through artistic, dramatic, and musical representations. The good city for the good life, an immensely powerful Athenian invention, diffused through time and space and left a legacy of pro urbanism difficult to dislodge.

Dislodging the Good Life, Disrupting the Rural

Industrialization proved to be the force that dislodged widespread pro urbanism. The good city for the good life morphed into Frederick Engels' Manchester, Henry Mayhew's London, and Upton Sinclair's Chicago, to name but a few representations of the industrial city as dirty, dark, crowded, anonymous, threatening to the weak, a jungle of brick, stone, and smoke. These repre sentations were powerful in part because their audiences could see how far things had fallen, how much the in dustrial city had become the brutal opposite of the

Athenian model. Gone was the city of intellectual discovery, artistic achievement, and political progress; in its place was L. S. Lowry's city of mighty factories sapping the life out of its faceless citizens (**Figure 2**). However, another factor was in place in conditioning antiurbanism – it is impossible to trace the emergence of urban fears without considering 'pro rural' sentiment. Even though the Industrial Revolution transformed both urban and rural areas, the burgeoning system of capitalist accumulation in the nineteenth century and its manifestation in intense urbanization was viewed with dismay by subscribers to traditional rural values (and it is important to note that, prior to the nineteenth century, the vast majority of humans lived in rural areas), who saw the cities as places which encouraged a rejection of the basic human values behind the formation of an inherently moral existence.

For example, in the United States, a negative discourse of the city, which began with the pastoral musings of Thomas Jefferson and was furthered significantly by the transcendental contemplations of Ralph Waldo Emerson, grew stronger and became embedded in social life through powerful representations of urban malaise in late nineteenth and early twentieth century American literature, art, and social theory. Far from being celebrated as signs of industrial and economic progress, American cities were often viewed as dirty and diseaseridden arenas of degenerate, immoral, and corrupt behavior, the exact and unruly opposites of small town and rural America, and therefore places which middle class Americans would be wise to avoid for their own wellbeing. Raymond Williams' *The Country and the City*, a magisterial survey of the dialectical tensions between the country and the city in English literary representations, reminds us that the celebration of the rural in contrast to the castigation of the urban is not something confined to America, and through centuries of celebrated poetry, fiction, and philosophy this contrast has become extraordinarily pervasive, to the extent that it is "one of the major forms in which we become conscious of a central part our experience and of the crises of our society." While Williams was careful to point out that the rural–urban contrasts were varied and sometimes inverted, for example, by Marx's observations of 'rural idiocy' and Baudelaire's city of 'feverish joys,' arguably the most powerful and lasting representations were those of rural tranquillity and urban hostility.

Figure 2 *Coming from the Mill*. L. S. Lowry (1930). Reproduced with permission of the Salford Museum.

It was to a sizeable transatlantic body of anti urban literary sentiment, deriving its power from the contrast to the pro urbanism of the past and pro ruralism of the age, that European social theorists such as Ferdinand Tonnies, Emile Durkheim, and Georg Simmel delivered their classic studies of the social and psychological effects of rampant urbanization. While it would be incorrect to describe their immensely influential and diverse perspectives on the changing patterns of social life in the city as somehow anti urban, it would be equally incorrect to downplay the role they had in reinforcing anti urban sentiment. Tonnies's elaborations of the impersonal and superficial human associations of an urban *gessellschaft* society, the opposite of community (*gemeinschaft*); Durkheim's plaintive projections of the urban condition of 'anomie' and the deviant behavior which might accompany social isolation and confusion; and Simmel's accounts of individuals protecting themselves from the startling disruptions and discontinuities of the metropolitan milieu (leading to what he called 'psychic overload') all added fuel to the legacy of urban fear left by Emerson, Thoreau, and the transcendentalists in the US; and Thomas Hardy, William Wordsworth, George Gissing, and the Victorian writers in England. The ideas and attitudes of the sociologists may have differed, but their writings on what can happen to societies and individuals when cities expand did little to reverse the tide of antiurban thinking which had penetrated the public imagination, even if the city was not always portrayed as the maelstrom of noise, suffering, and individual isolation like it was by those who came before them.

But why does this discourse of anti urbanism matter, and why is it of interest to urban geographers? The answer is simple – if cities are repeatedly depicted as places devoid of compassion, care, and community, or worse, as dreadful, dark, and dangerous, then it becomes all too easy to treat them as places which must be avoided, or even places which must be tamed and seriously modified in the image of those possessing the appropriate powers. Nowhere is this more apparent than in the United States. In 1962, Morton White and Lucia White wrote a landmark book entitled *The Intellectual Versus the City*, a remarkable survey of what they called the 'powerful tradition of antiurbanism' emanating from the pens of several generations of leading American writers. They concluded that the collective criticism of urban life coming from distinguished American intellectuals such as, *inter alia*, Thoreau, Edgar Allan Poe, and Henry James, "helps ex plain America's lethargy in confronting the massive problems of the contemporary city in a rational way (p. 200)." In 1962, White and White were not able to write with the benefit of insights from cultural geography, which have taught us to understand how landscapes are part of a system of social regulation and reproduction due to their dual properties of material form and discursive sign. However, in their book we begin to detect how anti urban discourse becomes materialized into a particular 'way of seeing' the American city, and a way of treating it. A case study involving representations of New York serves to illustrate the power of the discourse of anti urbanism.

Case Study: Edward Hopper's New York

> Great art is the outward expression of an inner life in the artist, and this inner life will result in his [sic] personal vision of the world. Edward Hopper, 1953 (quoted in Hobbs, 1987: 64)

Edward Hopper (1882–1967), an artist with a huge posthumous following (a traveling retrospective of his work in the summer of 2004 drew massive crowds to galleries worldwide), is often seen as the quintessential American 'Realist', not least because he studied with Robert Henri, the teacher of the famous Ash Can Realists at the New York School of Art, which has often led to erroneous associations of Hopper with the Ash Can school. Hopper was certainly a Realist, yet his brand of Realism was unique, and very different from his con temporaries who sought to fulfill Henri's injunction to his students to paint the crowded life of the modern city, implicitly that of its poor and immigrant communities. Hopper never painted urban crowds – quite the opposite, as we shall see. A brief biographical sketch helps us understand the foundations of the anti urbanism visible in Hopper's art.

Hopper was a recluse who spent a lifetime reading poetry, literature, drama, and philosophy, absorbing what he learned from books and observed from life, and ap plying it to canvas. His childhood was spent in an upper middle class Dutch settlement, Nyack in upstate New York, a world away from the expanding metropolis to the south where he would make his name and live permanently from 1910 to his death in 1967. The roots of his approach to the city can be found in the place in which he was raised, one where nature, community, religious education, patriotism, and duty were seen to lead to a solid moral existence. The 1890s, the decade of Hopper's adolescence, marked the acceleration of urban and industrial development that eroded traditional ways of life and challenged that solid moral existence of rural and small town America. Gail Levin, Hopper's most prolific biographer, identified this period as the root of the conflicts that Hopper lived throughout his life be tween traditional and modern, rural and urban, Ameri can and foreign ways; conflicts he would explore again and again in his work. Crucial to the development of Hopper's attitude toward the city were the writings of Thoreau and (particularly) Emerson, two individuals known for their dislike and distrust of city life in favor of romantic notions of the entwinement of humans with nature, and with whom Hopper felt close intellectual affinity. On Emerson, Hopper once said: "I admire him greatly. I read him a lot. I read him over and over again." Hopper held Emerson in such high esteem that many have argued that he sought to express the Emersonian vision through his art. In *The Intellectual Versus the City*, White and White refer to Thoreau's classic *Walden* (1854) as a

> bible of anti urbanism… The values it espouses are es sentially those of the isolated individual, living in nature and free of social attachments. (1962: 30)

The connections are striking – most of the people in Hopper's urban paintings are isolated individuals who appear out of place, detached from the city both socially and spatially as it changes around them, and seemingly bewildered by the threat to 'nature' posed by the built environment. Another writer who harbored an approach to the city to which Hopper could relate was Henry James, whose *The American Scene* (1907) was full of negative portrayals of a New York placed in cultural opposition to European cities. Hopper spent his early twenties in Paris, which he described to be "very graceful and beautiful…after the raw disorder of New York," and the year James' book was published was the same year Hopper returned to New York from Paris. *The American Scene* was avidly consumed by Hopper as it viewed ur banization with suspicion, and returning to New York after a lengthy absence made rapid urban change all the more tangible and unsettling to the artist. The appear ance of skyscrapers in New York's cityscape was seen by James and then Hopper as representative of encroaching, unwelcome modernity – a theme which is never far from a Hopper painting and usually depicted with trepidation and uneasiness. Take the following observation, again by Levin:

> He rarely represented skyscrapers at all, and when he did, he reduced them to fragmentary glimpses or intru sions on the cityscape…His recurrent visual ironies on the manifestations of modern life suggest his highly ambivalent attitude toward the changes occurring in twentieth century society; it is his profound alienation from contemporary life that makes his art so character istic of modernity itself. (1995a: 229)

Skyscrapers, to Hopper, were signs of unruliness and dislocation – serious violations of all that he had been brought up to believe, that humans should be in harmony with nature and situated away from disruptions to this almost puritan way of existence. The workings of modernity were antithetical to a man who disapproved of social and structural change, of overcrowding, of disorder. Although not an outright misanthrope, Hopper was a deeply private man lost in the worlds of his art and his passion for reading. He lived in New York City for most of his adult life, yet it is hard to find a Hopper painting where this city is celebrated or loved, or presented with any optimism. Much of this arose from his observations of social life in New York during the inter war years of Prohibition and Depression; strikes, unemployment, protest, poverty, and uncertain futures had a profound effect on his vision of the city. Hopper saw what could happen to a city when its growth was explosive, when its economy collapsed, when some of its people were left behind and struggled to make sense of the transforming world in which they lived.

Hopper's *Sunday* (1926) provides an instructive illustration (**Figure 3**). It shows a solitary man in contemplative mood sitting in front of a store. Hunched forward and appearing bored and disconsolate, it is a forlorn and lonely scene which deals with themes of emptiness, isolation, and loss. Although the man is shown on a Sunday, his day of rest, the store behind him is empty, lacking window displays or any other suggestion that it is ever open for business. We are left wondering if this is the storeowner, his livelihood in tatters, who is living his life in perpetual Sunday, reflecting on his losses and worrying about what to do next. While the painting predates the Depression, it is a

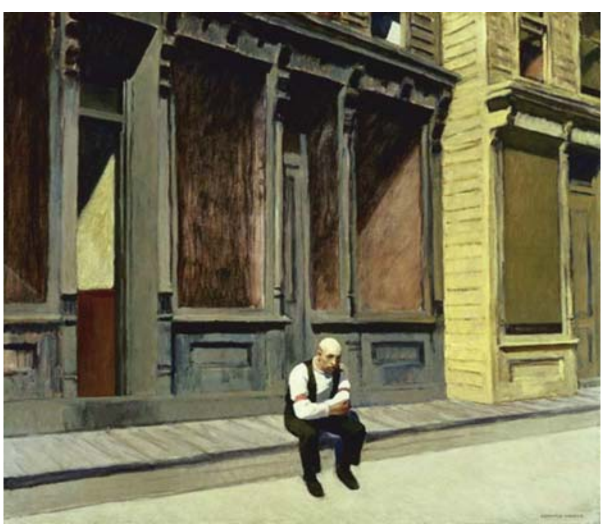

Figure 3 *Sunday*. Edward Hopper (1926). Oil on canvas. Reproduced with permission of The Phillips Collection, Washington, DC.

portrayal of an individual bypassed by the optimism and opulence of F. Scott Fitzgerald's 'Jazz Age', someone struggling to make sense of what has happened to him and his business as the world has changed around him. Hopper's profound suspicion of modernity is suggested by the nineteenth century storefronts – they have become the walls of a twentieth century ghost town, and like the man who sits before them, relics of an age which Hopper preferred, an age of small, modest businesses and architecture of a more human scale. While it could be argued that the painting has a small town feel, it was in fact based on Hopper's frequent sojourns to the city of Hoboken, New Jersey, just across the Hudson River from New York.

Early reactions to *Sunday* were notable for their tendency to describe its 'Americanness' in both place and subject. Perhaps this is because the title has universal significance, or because the location is unidentifiable – this could be any city in the USA. Most relevant to this discussion, perhaps the Americanness comes from the depiction of a city street that has come to exclude and alienate one of its residents. The storekeeper is left helpless and idle, exhibiting vulnerability rather than confidence by staring into a space to which he no longer belongs, and the viewer is left almost wanting to console him. It is this side of the painting which tells us much about Hopper's unique brand of Realism. The viewer becomes more involved than in earlier American (Ash Can) Realism for two reasons. First, the city is completely devoid of any optimism, depressing rather than liberating, lamented rather than celebrated; and second, city people appear emotionally weak and lost in their own melancholy thoughts, detached from the world around them in a near catatonic state. Is this man socially isolated because he is suffering from Durhkeim's condition of 'anomie'? In *Sunday*, we begin to imagine the inner situation of the man in the painting, how the city may have fashioned that situation, and reflect upon why this man appears to have been left behind by the optimism of the time. This is concentrated Hopper – lonely, silent, reflective, melancholy, empty.

The same themes can be detected in the New York night that is the setting for Hopper's most famous work, *Nighthawks* (1942) (**Figure 4**). Hopper contested those viewers who looked beneath the surface, saying that the work showed little more than "a restaurant on Greenwich Avenue where two streets meet," but was later forced to admit that "[u]nconsciously, probably, I was painting the loneliness of a large city." Some key indicators help us to generate disquiet in a scene which has a calm surface appearance, and situate the composition firmly within the anti urban mould. It is a scene of contrasting qualities of light – the only light for the empty streets comes from inside the restaurant, amplifying the motif of darkness

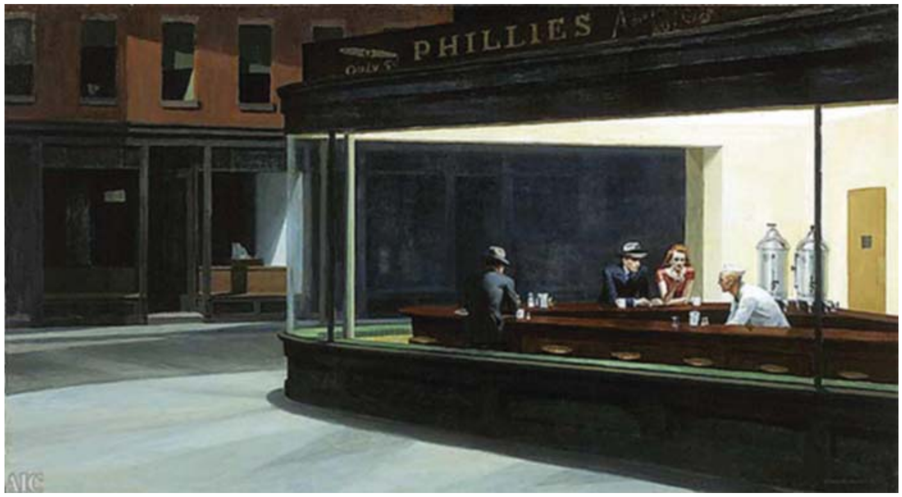

Figure 4 *Nighthawks*. Edward Hopper (1942). Oil on canvas. Reproduced with permission of the Art Institute of Chicago.

outside and intensifying the painting's communication of loneliness. Nowhere is this more effective than in the darkness behind the couple at the counter, which sits uneasily with the restaurant's bland cream interior and its glaring lighting. Hopper scholar Robert Hobbs contends that it is the use of light that is responsible for the mood of the entire scene:

> Circular in form, this building is an island that beckons and repels; and the fluorescent lighting is intimidating, alienating, and dehumanising. It creates an unreal and artificial feeling of warmth, an atmosphere that is clinical and more in tune with a laboratory than a restaurant. (1987: 129)

There is contrast within the restaurant too – the third character at the counter with his back to us, nearest the street, seems more in tune with outside than inside relative to the others, and less of an extra on a deserted stage, perhaps because of his positioning and solitude. This creates a sense that the couple facing us are alienated from their surroundings, literally out of place in a space which stands for little other than isolation.

Hopper was fascinated with cinema, and it is no co incidence that *Nighthawks* was painted at a time when the film noir genre was beginning to penetrate American public discourse. The genre's dark, dramatic scenes and generally pessimistic, almost paranoid outlook on life appealed to Hopper, and the Hopper city has much in common with the noir city as articulated by Frank Krutnik's subtle reference to Marx:

> Dark with something more than night, the noir city is a realm in which all that seemed solid melts into the shadows, and where the traumas and disjunctions ex perienced by individuals hint at a broader crisis of cultural self configuration engendered by urban Amer ica. (Krutnik, 1995: 99)

Film noir embodied and fed the anti urban senti ments of increasingly suburbanized moviegoers through its depiction of cities as disorderly, corrupt, dark, abysmal, threatening, and most relevant to this essay, curiously empty. While many of Hopper's urban rep resentations could be stills from a noir thriller, it is in *Nighthawks* where references to the noir city can be discerned with most ease – the strong theatrical light, the curiously empty streets, the mysterious figure with his back to us, the dark, forbidding spaces of the corner and buildings behind the diner. The title of the painting also suggests noir themes of people preying on others in the dark spaces of the unsafe city. The interlocking scripts of alienation, isolation, loneliness, fear, and a suggestion that something disorderly might occur out side immediately generated a sympathetic response from art critics and the public alike – the painting was soon recognized as an important American artifact and sold to the Art Institute of Chicago soon after its completion, where it remains today. *Nighthawks* also demonstrates a technique which Hopper used again and again in his work. He could have painted a far more threatening city, with more dubious characters, with litter on the streets, signs of crime, more suffering – yet he refrained, preferring to avoid extremities and thus 'suggesting' that these scenes might be around the cor ner by way of the emptiness of the cityscape. An empty city at night, captured at standstill, is always more threatening, more sinister, than an animated or extreme portrayal of urban fears.

Nighthawks is one of the most famous paintings in the history of American art, and has been reproduced on countless formats from coffee mugs to mouse mats.

Levin perfectly described the painting as communi cating a "nocturnal urban disquiet;" opening up *Night hawks* to closer iconographic inspection exposes the symbolic meanings of the painting to cement its pos ition as a landmark of twentieth century anti urbanism. Like *Sunday*, it condenses (with a little more suspense) all Hopper's influences and views on modern cities into a single image. Broadly defined, then, Hopper's urban work is a joining of nineteenth century transcenden talist beliefs with the tendency of Realists to document and interpret city scenes in a relentless quest for a hidden 'truth'. The products are 'stills' of the American city that evoke senses of loss, loneliness, and alienation – mournful commentaries on the unhappy material consequences of rampant, erosive modernity. Hopper's besmirchment of the entire modern age attracted at tention because it appealed to the anti urban imaginings of much of middle class America, most glaringly ex hibited in the paranoia and angst that led to the mass evacuation of the central city known as 'white flight', and the clearance of 'slum' neighborhoods now known as 'urban renewal'.

Hopper's work also struck a chord with Americans seeking to express their nostalgic yearnings for past times and places – a desire to see something American, and by inference, something virtuous, in an idealized landscape being steadily eroded. The American landscape idealized by Emerson, Thoreau, and their many followers was disturbed by modernity's leading edge of urbanization, and in the visual arts it is the work of Edward Hopper that documents this process with the most regular and solemn introspection. Whether it was Hopper's intention to tap into the anti urbanism of his viewers is open to debate, and indeed something we may never know, but there can be little doubt that Hopper 'fed' the American anti urban imaginary with forlorn scenes (all of Hopper's places are in fact anonymous 'nonplaces') that speak volumes about pro rural, small town suspicions of the modern metropolis.

Conclusion

In his magisterial book *Voices of Decline*, Bob Beauregard (1993) traced what he called the "discourse of decline" affecting urban America following World War II and, like White and White 30 years earlier, concluded that the sheer volume of negative journalistic and intellectual representations of city life was inhibitive toward a sens ible, sustained agenda of tackling America's urban problems. The "100% Urban Proof" Nissan Qashqai with which this article began is the latest example of that anti urban discourse materialized – the message from Nissan is clear: don't bother solving problems in the urban jungle, just protect yourself from them in this tank! It is quite simply Nissan's reaction to an ever more ex treme form of anti urbanism penetrating cities in a time of heightened hysteria created by the 'War on Terror'. There is a term, 'urbicide' (coined by Marshall Berman to describe the destruction caused by 'urban renewal' in the Bronx in the 1960s, and used more recently by pol itical urban geographers analyzing US and Israeli mili tary tactics) to capture what happens when extreme anti urbanism is taken to its logical conclusion. Discourses of urban fear, if left to grow and spread, can have these sorts of annihilation effects.

Derek Gregory has helpfully defined a discourse as:

> the ways in which we communicate with one another, …that vast network of signs, symbols, and practices through which we make our world(s) meaningful to ourselves and to others. (1994: 11)

One of the key contributions of post structuralist theory to human geography has been sustained attention to the ways in which discourses are laden with the power to affect social change, and in this case, to affect the way we think about urban places. Urban fears feed into policy from the right, but more recently, it has been argued that some respected leftist writers on urban af fairs are implicated in reproducing this anti urbanism discourse. Tom Angotti's scathing critique of Mike Davis' recent book *Planet of Slums* is notable for its ob servation that the book's apocalyptic rhetoric "feeds into longstanding anti urban fears about working people who live in cities," and the book itself is essentially a "windshield survey of cities in the South by a stranger from the North." If Angotti is onto something here, then two lessons can be learned: first, that writers concerned about urban problems ignore entrenched anti urbanism at their peril (they play into the hands of conservative political agendas if they do not think carefully about how they describe urban problems), and second, that intensive, long term ethnographic immersion coupled with sensitive analyses is a possible way to dispel anti urbanism, and the worrying implications of widespread fear of the city.

See also: Discourse; Industrial City; Landscape; Poststructuralism/Poststructuralist Geographies; Representation, Politics of; Urban Representation/ Imagination.

Further Reading

Angotti, T. (2006). Apocalyptic anti urbanism: Mike Davis and his planet of slums. *International Journal of Urban and Regional Research* 30, 961 968.

Beauregard, R. (1993). *Voices of Decline: The Post War Fate of US Cities*. Oxford: Blackwell.

Farish, M. (2005). Cities in shade: Urban geography and the uses of noir. *Environment and Planning D: Society and Space* 23, 95 118.

Gregory, D. (1994). *Geographical Imaginations*. Oxford: Blackwell.

Hall, P. (1998). *Cities and Civilization: Culture, Innovation and Urban Order*. London: Weidenfeld & Nicolson.

Hayward, K. (2004). *City Limits: Crime, Consumer Culture and the Urban Experience*. London: Glasshouse Press.

Hobbs, R. (1987). *Edward Hopper*. New York: Harry N. Abrams Publishing.

James, H. (1993). The American scene. In Howard, R. (ed.) *Henry James: Collected Travel Writings* (first publ. 1907), pp 351 736. New York: Library of America Series.

Kazin, A. (1983). Fear of the city 1783 to 1983. *American Heritage Magazine*. February, 6 11.

Krutnik, F. (1995). Something more than night: Tales of the noir city. In Clarke, D. (ed.) *The Cinematic City*, pp 83 99. London: Routledge.

Levin, G. (1995). *Edward Hopper: An Intimate Biography*. New York: Alfred A. Knopf.

Levin, G. (1995). Edward Hopper: His legacy for artists. In Lyons, D. & Weinberg, A. (eds.) *Edward Hopper and the American Imagination*, pp 109 128. New York: W.W. Norton.

Mumford, L. (1961). *The City in History: Its Origins, Its Transformations, and Its Prospects*. London: Penguin.

Renner, R. (1990). *Edward Hopper 1882 1967: Transformations of the Real*. Hamburg: Taschen Books.

Simmel, G. (1950). The metropolis and mental life. In Wolff, K. (ed.) *The Sociology of Georg Simmel* (first publ. 1903), pp 409 424. New York: Free Press.

Slater, T. (2002). Fear of the city 1882 1967: Edward Hopper and the discourse of anti urbanism. *Social and Cultural Geography* 3, 135 154.

Theisen, G. (2006). *Staying up Much Too Late: Edward Hopper's Nighthawks and the Dark Side of the American Psyche*. London: St. Martin's Press.

Thoreau, H. D. (1998). *Walden* (first publ. 1854). Boston, MA: Beacon Press.

White, M. and White, L. (1962). *The Intellectual Versus the City: From Thomas Jefferson to Frank Lloyd Wright*. Cambridge, MA: Harvard University Press.

Williams, R. (1973). *The Country and the City*. London: Chatto & Windus.

Relevant Websites

http://www.nissan.co.uk
Nissan Qashqai Website.

http://www.tate.org.uk
Tate Modern Pages for the 2004 Edward Hopper Retrospective.

Apartheid/Post-Apartheid

J. Battersby-Lennard, University of Cape Town, Cape Town, South Africa

The apartheid era in South Africa existed for 46 years, between 1948 and 1994. However, its historical reach can be traced considerably further back into the country's history and forward into present lived experiences.

Apartheid (literally: 'state of being apart' in Afrikaans) was perhaps most clearly defined in the 1985 United Nations International Convention Against Apartheid in Sport as, "a system of institutionalized racial segregation and discrimination for the purpose of establishing and maintaining domination by one racial group of per sons over another racial group of persons and system atically oppressing them." Essentially, the policy sought to shape the country's spatial, economic, and political life to enable 'separate development' of the four racial groups into which each resident of the country was categorized: white, Coloured, Indian, and black African (labeled 'Bantu', 'Native', and 'black' during apartheid). Inherent in the system was a racial hierarchy, reinforced through all spheres of state activity and at all spatial scales.

In the post apartheid era there are ongoing efforts to overcome the spatial, economic, social, and political legacy of apartheid. There are claims of other 'apartheids' throughout the world (e.g., Northern Ireland, Israel/Palestine, Brazil, the USA), drawing on similar legacies of segregation and dispossession, although without the in stitutional legality of South Africa's apartheid. This art icle examines the origins and nature of apartheid, factors leading to the end of apartheid, transformation within the post apartheid era, and provides a discussion of these other 'apartheids'.

From Segregation to Apartheid

Many of the building blocks of apartheid were in exist ence by the time the National Party was elected to power in 1948, leading some commentators to stress the con tinuity from pre apartheid to apartheid, from *de facto* to *de jure* segregation.

The Union of South Africa was formed in 1910 from the British Colonies of Natal and the Cape, and the Boer Republics of Transvaal and the Orange Free State. By the time of Union, the country had already largely been divided up along racial lines, in both rural and urban areas. The 1913 Land Act and 1936 Native Trust and Land Act apportioned 13% of the land in the Union to the black African population, who constituted around 75% of the total population. These Acts prohibited the purchase of land by black African people outside of the scheduled areas. It also made provision for the prevention of other means of independent black access to land, through squatting, leasing, share cropping, or labor tenancy. These Native reserves and Trust lands became the basis for the 'homelands', as illustrated in **Figure 1**.

In urban areas, too, there was already significant racial segregation. As early as 1847 the Cape Colonial gov ernment promulgated regulations for the establish ment of separate locations close to town where members of indigenous groups (including Afrikaans speaking Coloured people) were required to live, if they were not housed by their employers or were not independent property owners in their own rights. In the light of increased black African urbanization, the Native (Urban Areas) Act was passed in 1923 to establish sep arate locations for the black African population, and to place limitations on migration to towns. The Slums Act of 1934 was also used as a means of removing black African people from the inner urban areas. This Act also allowed for the expropriation of Coloured and In dian land. These various means of racial segregation in urban areas, and racially stratified economic realities, gave rise to highly segregated cities, as **Figure 2** illustrates.

In addition to spatial segregation, there were a number of other racially discriminatory laws and practices in existence before 1948. The black African population did not have voting rights, and the Coloured population in the Cape had their limited franchise diluted by laws passed in the 1930s. Through acts such as the Industrial Conciliation Act of 1924 and the Wages Act of 1925 certain jobs were reserved for white workers. Education was also highly segregated and funded differentially along racial lines.

While segregation had existed since the establishment of the Cape, there was an increase in legislation to ensure racial difference after the formation of the Union of South Africa in 1910. Many of the laws passed in the period between 1910 and 1948 were designed to guarantee the upliftment of poor whites. In the 1920s between 30% and 50% of all Afrikaners were in or close to the 'poor white' category. In 1939, almost 40% of adult male Afrikaners were unskilled. It has been argued that the National Party came to power in the light of an upsurge in poor white numbers after World War II at the same time as a black urban middle class was emerging, and in the light of the development of black opposition.

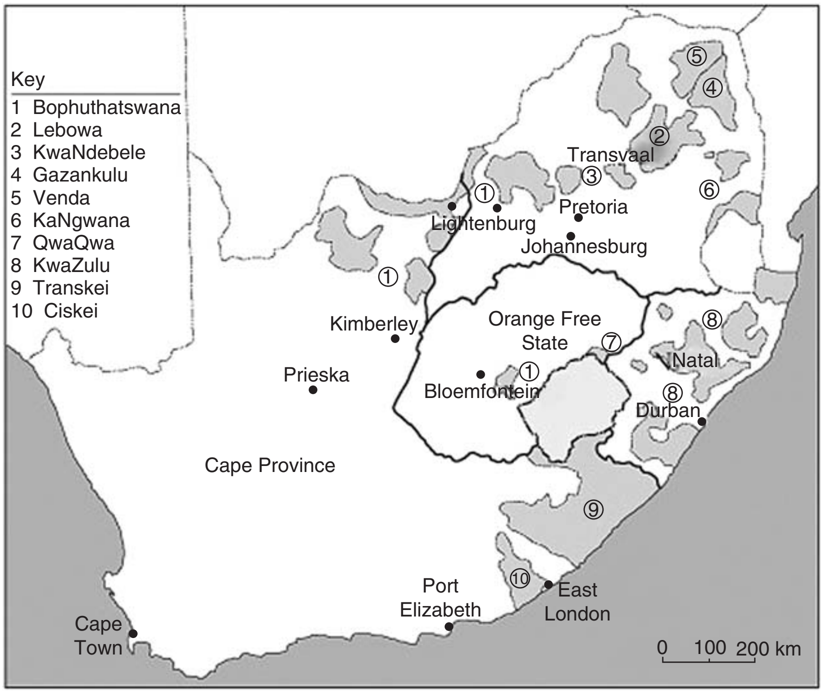

Figure 1 Apartheid Era Homelands. From Martin, S. and Boshoff, H. (2003) *'Volk', Faith and Fatherland: The Security Threat of Posed by the White Right*. ISS Monograph No 81. Institute for Security Studies, P. 38.

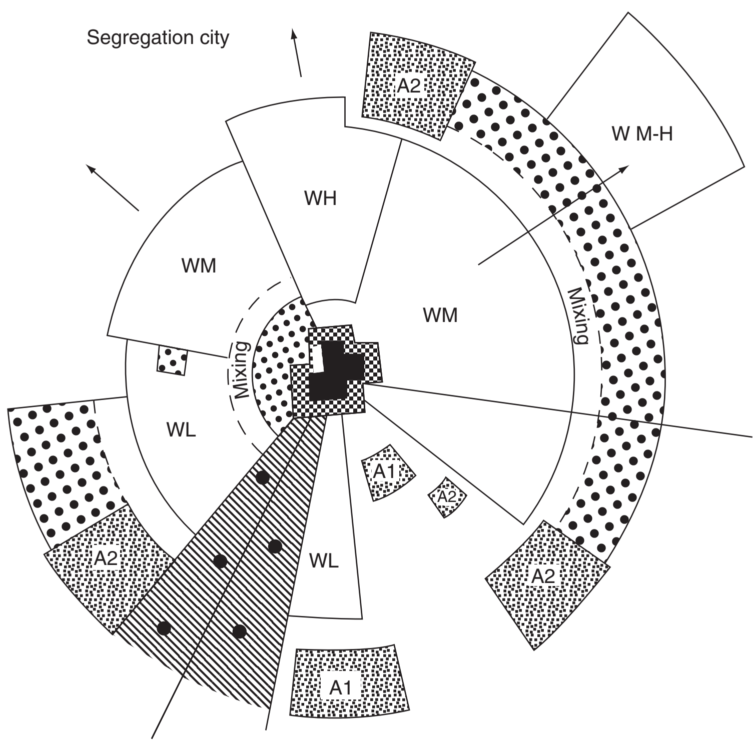

Figure 2 Model Segregation City. White central business district (C.B.D.); Indian C.B.D.; Frame C.B.D.; Industrial. Residential areas: White: Economic status: H High; M Middle; L Low. Indian and/or Coloured: Economic status not differentiated. African; ● Barracks compounds; A1 Municipal township; A2 Informal housing; Zones of racial mixing Mixing; Domestic servant quarters not shown. From Davies, R.J. (1981) The spatial formation of the South African city. *GeoJournal 2* (supplementary issue), 59 72.

The Nature of Apartheid

The degree of segregation in existence before 1948 has led some to argue that apartheid was simply a continuation of what was already in process. However, the election of the National Party brought a far more comprehensive kind of segregation to what had gone before. Between 1948 and 1951 a series of acts were passed to lay the foundation for complete segregation on grounds of race. H. F. Verwoerd, Prime Minister of South Africa from 1958 to 1966, said in 1948 that "[n]obody has ever contended that the policy of apartheid should be identified with 'total segregation'. The apartheid policy has been described as what one can do in the direction of what you regard as ideal. Nobody will deny that for the Natives as well as for the European complete separation would have been the ideal if it had developed that way historically." Verwoerd's identification of the limits of total segregation derived from the dependence of the white economy on black labor, thus preventing the desired complete separation.

In order to achieve its objectives, the apartheid state passed a series of laws which maintained and extended existing segregation. The core policies, which encapsulated the ideology and intent of the state were rapidly passed into law, namely the Prohibition of Mixed Marriages Act (1949), the Group Areas Act (1950), Population Registration Act (1950), the Immorality Act (1950), and the Bantu Authorities Act (1951). These five key policies were interlinked and created their own internal logic.

The Group Areas Act provided for areas within the urban environment to be declared for the exclusive use of one particular racial group. People were only allowed to own or rent residential property or businesses in areas classified for their racial group. Unlike previous acts, which proclaimed new areas for particular races, the Group Areas Act was also retrospective in its planning. It allowed for the rezoning of existing residential and business areas for particular race groups. It was estimated that by the end of 1976 although 32% of Chinese and Indian families and 20% of Coloured families had been spatially disqualified, just 0.2% of white families had been disqualified. The high proportion of Chinese and Indian families removed under the Act is a reflection of their previous level of integration in predominantly white areas as a wealthy subsection of the black majority. It has been suggested that one of the motivations for the Act was to restrict the further rise in status of these non white elite groups.

Black African South Africans were not provided for within the Group Areas Act, as they were not recognized as legitimate urban dwellers as a result of pre apartheid legislation. Black Africans were only allowed in cities as 'temporary sojourners', their 'home' being in the lands demarcated under the 1913 and 1936 Land Acts. The passing of the Bantu Authorities Act brought in a governmental system based on chiefs and 'tribal' authorities as the only political representation for black Africans. All black Africans were therefore assigned a 'homeland' according to their record of origin and were therefore effectively retribalized and ceased to be considered South African citizens.

These two grand spatial plans (national and urban) could not be implemented effectively without a rigid racial categorization. The Population Registration Act therefore required all citizens to be identified and registered as belonging to one of four core racial groups: white, Coloured, Indian, Bantu (black African). This classification influenced all spheres of life: political, economic, social, and spatial. It also facilitated other policies, for example, the Group Areas Act, to reinforce apartheid ideology. Apartheid could not allow these racial boundaries to blur and insisted on racial purity. The Prohibition of Mixed Marriages and Immorality Acts therefore banned sexual contact between people of different races in an attempt to maintain these racial boundaries.

Once these core laws were in operation, the apartheid state then implemented a series of other laws reinforcing the unequal separation of people in South Africa, including the Abolition of Passes Act (1951), Reservation of Separate Amenities Act (1953), Bantu Education Act (1953), Natives Resettlement Act (1954), Natives (Urban Areas) Amendment Act (1955), Industrial Conciliation Act (1956).

In its implementation, apartheid was fundamentally spatial. As David Smith stated in 1982, apartheid was perhaps the "most ambitious contemporary exercise in applied geography." The power of apartheid largely derived from its control of people's personal and community geographies at all scales of existence.

At the macro scale, South Africa was divided into black African homelands and the rest of the country. These ten homelands were nominally sovereign nations. Black Africans were denied citizenship of South Africa, as they were considered citizens of their allocated homeland. The homelands were economically, and often agriculturally, marginal and overcrowded places. Black African South Africans remained economically dependent on white capital as laborers, yet without the rights of citizens. At the mesoscale, the Group Areas Act divided urban areas along racial grounds residentially, economically, and recreationally, refining the 'segregation city' and reinforcing the racial hierarchy through spatial means (see **Figure 3**).

At the micro scale, the daily lives of people in public spaces and private spaces were dictated by laws designed to reflect the apartheid ideology. Separate benches, buses, hospitals, schools, etc, were all set in place. Businesses employing people of more than one race were obligated to provide separate toilets for people of each race. Public buildings, like post offices, had separate entrances for people of different races. Within this segregation, there was a clear racial hierarchy. Within education, for ex ample, the ratio of per capita expenditure on education in South Africa for black Africans, Coloureds, Indians, and whites was 1:3.4:4.9:8.3 in 1984. This pattern of separate and unequal provisioning was reflected in all spheres of life. The interlinking of these policies at different spatial scales created a systematic web of oppression.

However, despite the best efforts of the state, the system was neither watertight nor internally coherent. Throughout the apartheid era cracks appeared in the system, most obviously in the form of protest, but also through periodic economic crises. The following section discusses some of the reasons for the ending of apartheid.

Towards Ending Apartheid

Although apartheid officially ended in 1994, when South Africa held its first democratic election, it had been in a period of transition for a number of years, possibly as far back as the early 1970s. Reasons for the decline of apartheid can be broadly divided into agency based and structure based arguments.

Agency based arguments reflect the popular narra tives of internal political activism and the international anti apartheid movement. These arguments range from those around individual agency, such as the role of key figures like Nelson Mandela and F. W. de Klerk, to the role of wider internal and international protest. The structure based arguments tend to focus on local and global economic crises which de stablized the system, the failure of import substitution industrialization as a growth model, and other defects within the system. It is vital to recognize the impact of both in the nature and temporality of the end of apartheid.

It can be argued that apartheid had at its core a ten sion or contradiction, that its downfall was therefore in evitable, but dependent on agency to reveal and exploit these structural weaknesses. The National Party came to power through playing on white fears of the emergent black middle and working classes. It was therefore in a permanent position of attempting to stem the tide of black economic and political development. The internal struggle was such that in order for South Africa to be economically viable it needed to take advantage of black labor, but needed to do this without conceding political

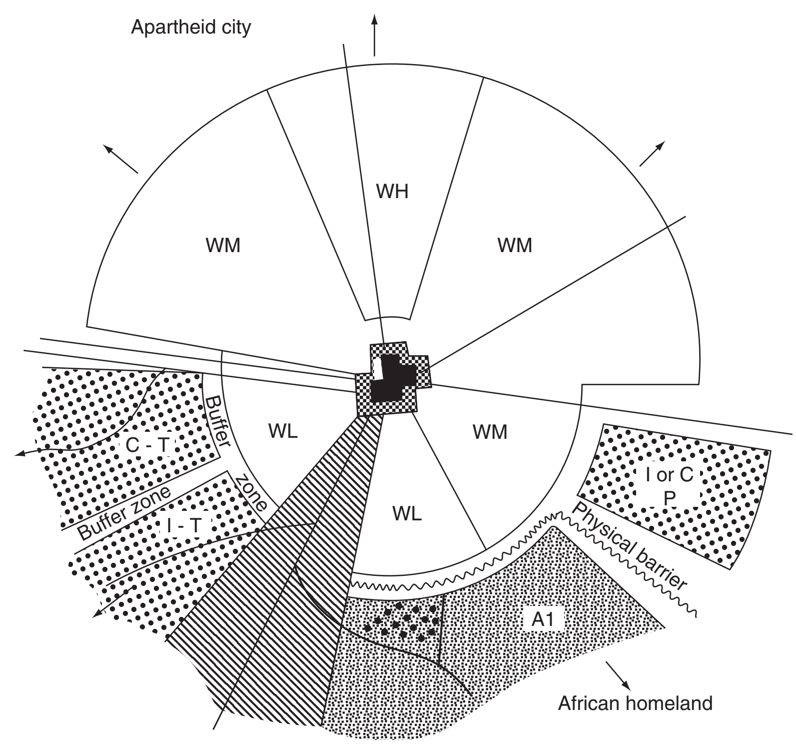

Figure 3 Model Apartheid City. ■ White C.B.D.; □ Indian C.B.D.; ⊡ Frame C.B.D.; ▨ Industrial. Residential areas: □ White: group area: Economic status: H High; M Middle; L Low. ▩ Indian and/or Coloured group areas: I Indian; T Township; C Coloured; P Privately developed; A1 Municipal township. ▦ Hostels; ● African group areas; Economic status of Black group areas not differentiated. Domestic servant quarters not shown. From Davies, R. J. (1981). The spatial formation of the South African city. GeoJournal 2 (supplementary issue), 59 73.

franchise. As the South African economy developed, this contradiction became more marked. The constraints placed upon black African access to urban areas began to impact economic efficiency. The economy was changing: requiring higher skilled workers in the manufacturing and service sectors, and yet this workforce could not feasibly emerge from the disenfranchised temporary urban sojourners educated by the Bantu education sys tem to be little more than manual labor. By 1977 there was a skills shortage of 99 000 in the professional, semi professional, and technical grades, up from 47 000 in 1969. These contradictions did not just manifest them selves in economic crises, but also by means of popular protest, both locally and internationally.

In the light of these issues, the state therefore began a program to reform apartheid through measures such as the legalization of black trade unions in 1979. This was in part an attempt to contain rising black labor militancy. In 1983, the government introduced the Tricameral Par liament system, which afforded a limited vote to Col oureds and Indians. This reform did not impact black Africans who were not recognized as citizens of South Africa, but of their allocated homelands. Under the tri cameral system, parliament had three houses, one for each race. Each of the three chambers had powers of various 'own affairs' issues, such as education, housing, and cultural, but 'general affairs', such as finance, defense, and foreign affairs, were beyond the control of individual houses. However, the balance of power always rested with the (white) House of Assembly, both numerically and in terms of the responsibilities of the houses. A third at tempt at reform was the 1986 White Paper on Urban ization, which reduced control of access to urban areas for black Africans, with the objective of allowing a compliant black middle class to develop and therefore provide political and economic stability. However, through these attempts to ensure stability, greater in stability was generated and protests increased.

In the 1980s, as the state implemented reforms de signed to make apartheid acceptable internally and internationally without disturbing the fundamentals, the perversity of the system became more apparent. There fore, when F. W. de Klerk became President in 1989, the year that the Cold War ended (therefore dissipating the fear of black led communism that many white South Africans held), the political and economic climate was right for moves to be made by the state to end apartheid. In 1990 the African National Congress (ANC), Pan Af ricanist Congress, South African Communist Party, and other liberation organizations were unbanned. In 1991, key political leaders from both sides began formal con stitutional negotiations at the Convention for a Demo cratic South Africa (CODESA) meetings. In 1992 a referendum of white voters was held, with 68.6% voting 'yes' to the continuation of the reform process. The Interim Constitution was passed in 1993 and in April 1994 the first democratic elections in South Africa were held.

These changes were not a sudden transformation, but part of a long term process of de stablizing the apartheid state through interlinked structural and agency led pro cesses. The nature of the post apartheid state was the result of a negotiated settlement between the apartheid state and the liberation movement. This process has largely been recognized as playing a vital role in the peaceful transition and the economic stability of the country. However, there are many who argue that the compromises made to ensure this stability have impacted the pace of social and economic transformation in the post apartheid era.

Post-Apartheid

The end of apartheid as a political force heralded the start of the post apartheid era. However, post apartheid South Africa is in many ways still attempting to liberate itself from the legacy of apartheid and precursors. The multiscalar spatial reach of apartheid impacted upon every area of life for individuals and communities in a complex, interlinked manner. Therefore, the success of efforts to overcome the legacy of apartheid in one sector of South African life are contingent on addressing many other aspects at a range of scales. For example, equal, or even redistributive, funding of education will not result in all schools having equal economic and educational resources. Historical underfunding of education has left many Coloured and black African schools with a sig nificant proportion of underqualified teachers, particu larly in rural areas. The spatial legacy of apartheid places black schools at a disadvantage in terms of access to valuable teaching resources and amenities in the city centers. The economic legacy leaves schools in poorer areas with less ability to raise school fees and additional funds to supplement state education provision. The legacy of lack of investment in black areas leaves the school community resource poor in terms of libraries and sports facilities. The educational legacy leaves schools contending with undereducated parents unable to pro vide the educational support for their children that children in areas of historic privilege have.

Returning to the idea that apartheid's power used spatial reordering at many scales to reinforce a political ideology, it is possible to observe the legacy of apartheid. At the macro scale, the state has stated its objective to redistribute 30% of South Africa's agricultural land to black farmers by 2014. The purpose of this is to break down the historical legacy of the homelands and reserves policies, which forced the black African population onto just 13% of South Africa's land, enforcing racial segregation. Land reform is critical for promoting equity, addressing past injustices, and providing sustainable rural livelihoods. However, at the time of writing, just 4% had been redistributed. This slow rate of change can be at tributed to the protracted negotiations that take place within the current 'willing buyer, willing seller' system. There have been allegations that white farmers are only offering their worst land for sale and asking way over market value. There have been counter allegations that the state is not offering reasonable payment. As yet, land redistribution and land restitution have failed to signifi cantly disrupt apartheid spatiality. In 2006 the govern ment announced the first expropriation of land as a result of the failure of negotiations between a farmer and the Department of Land Affairs.

At the mesoscale, urban areas in South Africa are still largely constrained by the pre existing spatial framework. Individual movement across the Group Areas Act demar cations has been fairly limited for a number of reasons. An initial reason was the overcapitalization that had taken place in wealthier Coloured and Indian areas due to a shortage in housing stock. After the Group Areas Act was repealed in 1991 the value of these properties dropped, therefore making it difficult for people who wanted to buy into his torically white areas. Houses of equal quality in historically white areas are still considerably more expensive than those elsewhere, therefore the economic legacy of apartheid makes racial desegregation difficult. In addition, many people do not wish to move from areas where they have developed a sense of community, into areas where they fear hostility from the existing racial community, part of the apartheid legacy of a racialized mindset. This limited de segregation is less evident in newer areas and in Johan nesburg's wealthy Northern Suburbs, reflecting the rise of a considerable black middle class in Johannesburg.

Government has attempted to encourage desegregation through means such as the land restitution program. This program allowed individuals and communities dispossessed of land after 1913 to make a claim for the land from which they were removed, alternative land, or financial compen sation. This program has had little impact in urban areas as, for various reasons, most urban claims have been settled through financial compensation, rather than land. The post apartheid city has been characterized by continued pe ripheralization of poor households. Since 1994, 2.3 million houses have been delivered by means of a capital subsidy toward site and service and a starter house. There is cur rently an estimated backlog of 2.4 million houses. This figure far exceeds original estimates of housing need and is driven by migration and reduced household sizes.

The vast majority of the new government provided houses have been located on the periphery of urban areas, largely as a result of the high cost of urban land. It has also been asserted that urban spatial policy was less radical than had been desired due to factors such as weak local government and pressure to demonstrate urban economic stability for international investment. In 2004, the Depart ment of Housing identified a new policy focus, Breaking New Ground, which seeks to address this peripheralization. However, it is likely that the majority of poor households will continue to live in peripheral areas, whether in formal or informal housing, as a result of the chronic housing shortfall and limited budgetary capacity of the state.

In the post apartheid era the role of the country within the southern African region and the continent as a whole has been transformed. South Africa is now a crit ical political influence and economic driver within the region. At the same time as South Africa is realizing its new role in the regional and global economy, it is now also finding that its ability to overcome the apartheid legacy is profoundly influenced by the challenges of the global economy.

South Africa has undergone radical transformation in many spheres since the end of apartheid; however, given the continued impacts of many aspects of the system on the experiences of citizens today, it is not yet possible to conceive of the country without reflecting on apartheid. Apartheid and post apartheid remain fundamentally linked.

Other Apartheids

There have been numerous claims of other apartheids around the world, in countries as diverse as Northern Ireland, the USA, Israel/Palestine, and Brazil. If we ac cept the United Nations definition of apartheid used at the start of this article, it is debatable whether these can be considered apartheid equivalents. What constitutes institutionalized racial segregation and discrimination? There is indeed segregation within these nations and in some cases extreme inequality on grounds of race. In addition, the commonalities of experience, can allow critical dialog to take place between citizens of these countries. However, the legal framework and total ideology purposefully created under apartheid, marks South Africa's experiences as fundamentally different.

See also: Africa; Citizenship; Colonialism I; Colonialism II; Imperialism, Cultural; Inequality; Postcolonialism/ Postcolonial Geographies; Race; Segregation; Segregation, Urban; Third World Cities; Urbanization.

Further Reading

Beall, J., Crankshaw, O. and Parnell, S. (2002). *Uniting a Divided City: Governance and Social Exclusion in Johannesburg*. London: Earthscan.

Beinart, W. (2001). *Twentieth Century South Africa*. New York: Oxford University Press.

Bond, P. (2000). *Elite Transition: From Apartheid to Neoliberalism in South Africa*. Pietermaritzburg: University of Natal Press.
Christopher, A. J. (1994). *The Atlas of Changing South Africa*. London: Routledge.
Davies, R. J. (1981). The spatial formation of the South African city. *GeoJournal* 2(supplementary issue), 59 73.
Lemon, A. (1983). *Apartheid in Transition*. Aldershot: Gower.
Lemon, A. (ed.) (1995). *The Geography of Change in South Africa*. Chichester: Wiley.
Lester, A., Nel, E. and Binns, T. (2000). *South Africa: Past, Present and Future*. Harlow: Prentice Hall.
Posel, D. (1991). *The Making of Apartheid 1948 1961*. Oxford: Clarendon Press.
Saff, G. R. (1998). *Changing Cape Town: Urban Dynamics, Policy and Planning during the Political Transition in South Africa*. Lanham: University Press of America.
Smith, D. M. (1992). *The Apartheid City and Beyond: Urbanization and Social Change in South Africa*. London: Routledge.
Western, J. (1996). *Outcast Cape Town* (2nd edn.). Berkeley, CA: University of California Press.

Relevant Websites

www.sahistoryonline.org.za
South African History Online

Applied Geography

M. Pacione, University of Strathclyde, Glasgow, UK

Glossary

Applied Geography An approach based on the philosophy of relevance or social usefulness that focuses on the application of geographical knowledge and skills to advance the resolution of real-world social, economic, and environmental problems.

Marxist Geography The study of geographical questions using the analytical insights, concepts, and theoretical frameworks of historical materialism.

Postmodernism A broad-based movement in philosophy, the arts, and social science characterized by skepticism toward the 'grand theory' of the modern era, stressing instead an openness to a wide range of views in social enquiry.

Public Policy The actions of government and the intentions that determine these actions.

Quality of Life The state of social well-being of an individual or group, either perceived or as identified by observable indicators.

Social Justice A broad principle concerned with the 'fairness' of the distribution of society's benefits and burdens among individuals and groups.

Social Policy The part of public policy that aims to improve human welfare and meet human needs.

The rationale underlying applied geography is based on the particular philosophy of relevance or social usefulness that focuses on the application of geographical knowledge and skills to advance the resolution of real world social, economic, and environmental problems. Rather than being considered as a 'subarea' of geography, (akin to economic, social, or historical geography), applied geography refers to an 'approach' that crosscuts artificial disciplinary boundaries to involve problem oriented research in both human and physical geography. Applied geographers are active across the human physical geography divide and in most subareas of the discipline. Applied geography is a socially relevant approach to the study of the relationship between people and their environments.

The relevance and value of applied geographical research have never been more apparent given the plethora of problem situations which confront modern societies – ranging from extreme natural events (such as floods, droughts, and earthquakes) through environmental concerns (such as deforestation, disease, and desertification) to human issues (such as crime, poverty, and unemployment). An applied geographical approach has the potential to illuminate the nature and causes of such problems and inform the formulation of appropriate responses.

The Concept of Useful Knowledge

The concept of useful knowledge is of central importance for applied geography but is problematic for some practicing geographers. Those who do not see themselves as applied geographers may interpret the concept as indicating a corollary in the shape of geographical research that is less useful or even useless. This would be a misinterpretation. The concept of useful knowledge should be seen as an expression of the fundamental ethos of applied geography rather than a design to alienate 'nonapplied' geographers. The use of the concept of useful knowledge makes explicit the view that some kinds of research are more useful than other kinds. This is not the same as saying that some geographical research is better than other work – all knowledge is useful – but some kinds of research and knowledge are more useful than other kinds in terms of their ability to interpret and offer solutions to problems in contemporary physical and human environments. For 'nonapplied' geographers, the idea of applied geography or useful research is a chaotic concept which does not fit with the cultural turn in social geography or the postmodern theorizing of recent years.

The philosophical disjuncture between applied geographers and other geography practitioners is illustrated most clearly by comparing the applied geographical approach with the alternative postmodern perspective. One of the major achievements of postmodern discourse has been the illumination of the importance of difference in society as part of the theoretical shift from an emphasis on economically rooted structures of dominance to cultural 'otherness' focused on the social construction of group identities. However, there is a danger that the reification of difference may preclude communal efforts in pursuit of goals such as social justice. A failure to address the unavoidable real life question of 'whose is the more important difference among differences' when strategic choices have to be made represents a serious threat to constructing a 'practical politics' of difference. Furthermore, if all viewpoints and expressions of identity are equally valid, how do we evaluate social policy or, for that matter, right from wrong? How do we avoid the segregation, discrimination, and marginalization which the postmodern appeal for recognition of difference seeks to counteract? For some commentators the failure to address real issues would seem to suggest that the advent of postmodernism in radical scholarship has done little to advance the cause of social justice. Discussion of relevant issues is abstracted into con sideration of how particular discourses of power are constructed and reproduced. Responsibility for bringing theory to bear on real world circumstances is largely abdicated in favor of the intellectually sound but morally myopic premise that there is no such thing as reality. Decoupling social critique from its political–economic basis is not helpful for dealing with the shifting realities of life. Some postmodern writers have acknowledged the dangers of cultural theorists retreating from the empirical world into an 'in house' dialog set within an 'immaterial world' of limited relevance to 'material world' events. Notwithstanding such cautionary codas, in terms of real world problems, for many applied geographers, much postmodern thought would appear to condemn us to inaction while we reflect on the nature of the issue. As we shall see below, a similar critique may be leveled at the Marxist critique of applied geography which was preva lent during the 1970s and 1980s.

These views do not represent an attempt to be pre scriptive of all geographical research but are intended to indicate clearly the principles and areas of concern for applied geography. It is a matter of individual conscience whether geographers study topics such as the icon ography of landscapes or the optimum location for health centers, but the principle underlying the kind of useful geography espoused by most applied geographers is a commitment to improving existing social, economic, and environmental conditions. In this there can be no com promise – no academic fudge; some geographical re search 'is' more useful than other work. This is the focus of applied geography.

There will continue to be divergent views on the content and value of geographical research. Such debate is healthy and raises a number of important questions for the discipline and for applied geography in particular. The concept of 'useful research' poses the basic questions of useful for whom?, who decides what is useful?, and based on what criteria?. The related questions of values in research, and the nature of the relationship between pure and applied research are also issues of central importance for applied geography.

The Relationship between Pure and Applied Research

Applied research in any discipline is best understood in contrast with basic, or pure, research. For some com mentators basic research in geography aims to develop new theory and methods that help explain the processes through which the spatial organization of physical or human environments evolves. In contrast, applied re search uses existing geographic theory or techniques to understand and solve specific empirical problems.

While this distinction is useful at a general level it overplays the notion of a dichotomy between pure and applied geography, which are more correctly seen as two sides of the same coin. There is, in fact, a dialectic re lationship between the two. Applied geography uses the principles and methods of pure geography but is different in that it analyzes and evaluates real world action and planning and seeks to implement and manipulate en vironmental and spatial realities. In the process, it con tributes to, as well as utilizes, general geography through the revelation of new relationships. The conjuncture between pure and applied research is illustrated clearly in geomorphology where, for example, attempts to ad dress problems of shoreline management have contrib uted to theories of beach transport, the difficulties of road construction in the arctic have informed theories of permafrost behavior, while problems encountered in tunneling have aided the development of subsidence theory. Applied research provides the opportunity to use theories and methods in the ultimate proving ground of the real world, as well as enabling researchers to con tribute to the resolution of real world problems. More generally, theory is essential in applied geography at two levels. First, it provides the framework for asking ques tions about the substantive relationships embodied in a problem (e.g., here a model of a hydrological catchment illuminates the potential effects of a proposed flood prevention scheme). Second, social theory provides a normative standard against which current and future social conditions may be judged in terms of defined moral goals (which may address issues such as whether a min imum wage and basic standard of living should be a legal entitlement in advanced capitalist societies).

There is little merit in pursuing a false dichotomy between pure and applied research. A more useful dis tinction is that which recognizes the different levels of involvement of researchers at each stage of the research and specifically the greater engagement of applied geographers in the 'downstream' or postanalysis stages. The applied researcher has a greater interest than the pure researcher in taking the investigation beyond an alysis into the realms of application of results and monitoring the effects of proposed strategies. Researcher participation in the implementation stage may range from recommendations in scholarly publications or contracted reports (a route favored by most academic applied geographers, though not exclusively) to active involvement in implementation (more usually by applied geographers employed outside academia). Between these positions lie a variety of degrees of engagement, in cluding acting as expert witnesses at public enquiries,

Table 1 Cycles of pure and applied geography

First applied period (late nineteenth century)	Geography created as an applied discipline to serve the political, military, and commercial interests of the state.
First pure period (early twentieth century)	Based around the holistic philosophy encompassing both physical and human phenomena and focused on the core concept of the region and regional synthesis.
Second applied period (inter-war)	A period of war, followed by depression, and war again demanded geography demonstrate its usefulness in fields such as land-use planning.
Second pure period (post-1945 boom)	Rejection of ideographic regionalism replaced by spatial science and the quantitative revolution; demise of holistic approach and emergency of subfields within the discipline.
Third applied period (mid-1980s)	Extension of the concept of useful research into new areas of concern relating to social, economic, and environmental problems; applied geographers working both in academic and in public and private sectors. Applied geography as an approach rather than a subfield crosscuts the artificial boundary between physical and human geography and emphasizes the dialectic relationship between pure and applied research. Acknowledgement of the role of human agency and values in research and environmental change, and the need for a pluralist view of science.
Third pure period (?)	Characteristics unknown but speculatively a return to a more holistic philosophy reflecting the growing importance of environmental issues and the combinatory perspective of applied geography.

From Pacione, M. (1999). *Applied Geography: Principles and Practice*. London: Routledge.

dissemination of research findings via the media, field involvement in, for example, landscape conservation projects, and monitoring the effects of policies and strategies enacted by governmental and private sector agencies.

The balance between pure and applied research within a discipline varies over time in relation to the prevailing sociopolitical environment. When external pressures are at their greatest disciplines will tend to emphasize their problem solving capacity while during periods of national economic expansion 'more academic' activity may be pursued in comfort. Some have equated these cycles with the long waves of the world economy, and identified three periods in which applied geography was in the ascendancy (in the late nineteenth century, inter war era, and mid 1980s), separated by two periods of pure geography (in the early twentieth century, and during the post 1945 economic boom) (**Table 1**).

The Value of Applied Geography

A fundamental question for those working within the framework of applied geography concerns the value of a problem oriented approach. We have examined this issue already in our discussions of useful knowledge and the relationship between pure and applied research but we return to it here to address the specific critique of applied geography that has emanated from Marxist theorists. While the power of the Marxist critique has been much reduced by its own success in exposing the value bases of research, it still offers a useful perspective on the value of applied research.

The essence of the Marxist critique of applied social research is that it produces ameliorative policies that merely serve to patch up the present system, aid the legitimation of the state, and bolster the forces of capit alism with their inherent tendencies to create inequality. For these radical geographers participation in policy evaluation and formulation is ineffective since it hinders the achievement of the greater goal of revolutionary social change.

In terms of praxis, the outcome of this perspective is to do nothing short of a radical reconstruction of the dominant political economy (a position which, as we have seen, may also be reached from a different direction by some postmodernist theorists). Although the analytical value of the Marxist critique of capitalism is widely ac knowledged, its political agenda, and in particular, op position to any action not directed at revolutionary social change, finds little favor among applied geographers. To ignore the opportunity to improve the quality of life of some people in the short term in the hope of achieving possibly greater benefit in the longer term is not com mensurate with the ethical position implicit in the problem oriented approach of applied geography.

Neither does the argument that knowledge is power and a public commodity that can be used for good or evil undermine the strength of applied geography. Any knowledge could be employed in an oppressive and discriminating manner to accentuate inequalities of wealth and power but this is no argument for eschewing research. On the contrary, it signals a need for greater engagement by applied geographers in the policymaking and implementation process provided, of course, that those involved are aware of and avoid the danger of co optation by, for example, funding agencies.

Furthermore, access to the expertise and knowledge produced by applied geographical research is not the sole prerogative of the advantaged in society, but can be equally available to pressure groups or local communities seeking a more equitable share of society's resources.

Examples of the participatory tradition of applied re search in urban geography range from the advocacy planning practiced in several US cities whereby pro fessional planning expertise is provided to minority groups lacking the financial ability to purchase such services, to nongovernmental organization (NGO) sponsored grassroots social action by marginalized com munities in cities in the Third World.

Applied research involves the formulation of goals and strategies and the testing of existing institutional policies within the context of ethical standards as criteria. This should not imply a simple system maintenance approach to problem solving. Indeed, it is often necessary to take an unpopular anti establishment position, which can re sult in a major confrontation. For practical examples of this we may refer to the recommendations of the British community development projects of the 1970s, which advocated fundamental changes in the distribution of wealth and power and which led to conflict with both central and local government; as well as more recent policy oriented analyses of poverty and deprivation in which the identification of sociospatial patterns is used to advance a critique of government policy.

Values in Applied Geography

At each stage of the research process the applied geog rapher is faced with a number of methodological and ethical questions. Decisions are required on defining the nature of the problem, its magnitude, who is affected and in what ways, as well as on the best means of addressing the problem. All of these require value judgments on, for example, the acceptability of existing conditions (what is an acceptable level of air pollution? or of infant mal nutrition?). Values are also central to the evaluation and selection of possible remedial strategies, including com parative analysis of the benefits and disbenefits of dif ferent approaches for different people and places. In some cases, the applied geographer may seek to minimize such value judgments by enhancing the objectivity of the research methodology (e.g., by employing a classification of agricultural land capability to inform a set aside pol icy). In most instances, however, it is impossible to re move the need for value judgment. The impossibility of objective value free research is now axiomatic.

One issue of particular concern refers to the values that condition the selection, conduct, and implemen tation of research, a dilemma highlighted by the aphor ism 'he who pays the piper calls the tune.' Some advocates of public policy research by applied geog raphers, express this in terms of doubts over whether government departments will commission necessary re search into the effectiveness and consequences of their own policies and a real danger that constraints will be imposed over publication, especially if this contains criticisms of the sponsors or explores politically sensitive areas. Applied geographers must beware of any re strictions imposed by research sponsors and be aware of the ways in which their research results may be used. Applied geographers must seek to ensure that their work contributes to human welfare. In practice, this goal may be approached by careful selection of clients and research projects, by ensuring freedom to disseminate results and, where possible, through engagement in the implemen tation and monitoring of relevant policies or strategies.

Applied Geography and Public Policy

The practical value of the applied geographical ap proach has been demonstrated in the foregoing dis cussion. Applied geographers are actively engaged in investigating the causes and ameliorating the effects of 'natural' phenomena including acid precipitation, land slides, and flooding. Key issues of environmental change and management also represent a focus for applied geographical research with significant contributions being made in relation to a host of problems ranging from the quality and supply of water, deforestation, and desertification to a series of land use issues including agricultural de intensification, derelict and vacant land, and wetland and townscape conservation. Applied geographers with a particular interest in the built en vironment have, in recent decades, directed considerable research attention to the gamut of social, economic, and environmental problems which confront the populations of urban and rural areas in both the developed and developing realms. Problems of housing, poverty, crime, transport, ill health, sociospatial segregation, and dis crimination have been the subject of intense investi gation, while other topics under examination include problems ranging from boundary disputes and political representation to city marketing. The application of techniques in applied geographical analysis is of par ticular relevance in relation to spatial analyses where the suite of problems addressed by applied geographers ranges from computer mapping of disease incidence to simulation and modeling of the processes of change in human and physical environments.

The list of research undertaken by applied geog raphers is impressive but there are no grounds for complacency. While applied geographers have made a major contribution to the resolution of real world prob lems, particularly in the context of the physical en vironment, in terms of social policy formulation in the post war era the influence of applied geography has been mixed and arguably less than hoped for by those socially concerned geographers who engaged in the relevance debate a quarter of century ago.

Several reasons may be proposed to account for this. The first refers to the eclectic and poorly focused nature of geography and the fact that 'geographical' work is being undertaken by 'nongeographers' in other disciplines. This undermines the identity of geography as a subject with something particular to offer in public policy debate. The very breadth of the discipline, which for many represents a pedagogic advantage, may blur its image as a point of reference for decision makers seeking an informed input. Geographers wishing to influence public policy must compete with other more clearly identified 'experts' working on similar themes.

A second reason for the relatively limited influence on public policy may be the apparent reluctance of (human) geographers to 'get their hands dirty' – an attitude redolent of the eighteenth century distinction between gentlemen, who derived a livelihood from the proceeds of landownership, and those who earned a living through trade. This applies less to research in physical geography where a basis in empirical science and positivist methodology has ensured that applied research has attracted support and acclaim more readily both from within the discipline and from external agencies. Significantly, the growth of environmentalism and accompanying convergence of the philosophy and methodology of physical and human geography has gone some way to bridge the gap between the two major subareas of the discipline and may represent a route for applied geographers to increase their policy influence.

The changing content and shifting emphases of human geography during the last quarter of the twentieth century represents a third factor underlying the limited social impact of applied geography. Over the period the replacement of the earlier land use focus in applied human geography by questions relating to the geography of poverty, crime, healthcare, ethnic segregation, education, and the allocation of public goods brought applied geographers into direct confrontation with those responsible for the production and reproduction of these social problems. Unsurprisingly, since policymakers are resistant to research which might undermine the legitimacy of the dominant ideology, social policy remained largely impervious to geographical critique, particularly that which emanated from the Marxist analysis of capitalism.

The failure of applied geography to exert a major influence on social policy, however, does not signal the failure of applied geography to promote any significant improvement in human well being, which, as we have seen, may be achieved by means other than via public policy. Any assessment of the contribution of applied geography to the resolution of real world problems must balance the limited success in the specific area of social policy against the major achievements of applied geographers in the large number of other problem areas outlined above. Rather than dwelling on the limited impact to date of applied geographical research in the field of social policy applied geographers can draw encouragement from their unwillingness to compromise a critical stance in return for public research funds or public acceptability of research findings. Furthermore, much of the applied social research undertaken achieves the goal of addressing real world problems via its emancipatory power to expose the structural underpinnings of contemporary sociospatial problems and by encouraging exploration of alternative social arrangements.

See also: Activism; Activist Geographies; Critical Geography; Critical Realism/Critical Realist Geographies; Public Policy; Radical Political Economy; Social Justice, Urban.

Further Reading

Berry, B. (1972). More on relevance and policy analysis. *Area* 4, 77 80.
Blomley, N. (1994). Activism and the academy. *Environment and Planning D: Society and Space* 12, 383 385.
Castree, N. (2000). Professionalism, activism and the university: Wither critical geography? *Environment and Planning A* 32, 955 970.
Chisholm, M. (1971). Geography and the question of relevance. *Area* 3, 65 68.
Cooper, S. (1966). Theoretical geography, applied geography and planning. *Professional Geographer* 18, 1 2.
Coppock, J. T. (1974). Geography and public policy: Challenges, opportunities and implications. *Transactions of the Institute of British Geographers* 63, 1 16.
Dickenson, J. and Carke, C. (1972). Relevance and the newest geography. *Area* 4, 25 27.
Folke, S. (1972). Why a radical geography must be Marxist. *Antipode* 4, 13 18.
Frazier, J. W. (1982). *Applied Geography: Selected Perspectives*. Englewood Cliffs, NJ: Prentice Hall.
Harvey, D. (1974). What kind of geography for what kind of public policy. *Transactions of the Institute of British Geographers* 63, 18 24.
Kenzer, M. (1989). *Applied Geography: Issues, Questions and Concerns*. Dordrecht: Kluwer.
Murphy, A. (2006). Enhancing geography's role in public debate. *Annals of the Association of American Geographers* 96(1), 1 13.
Pacione, M. (1990). What about people? A critical analysis of urban policy in the United Kingdom. *Geography* 75, 193 202.
Pacione, M. (1999). *Applied Geography: Principles and Practice*. London: Routledge.
Pacione, M. (2009). *Urban Geography: A Global Perspective* (3rd edn.). London: Routledge.
Palm, R. and Brazel, A. (1992). Applications of geographic concepts and methods. In Abler, R., Marcus, M. & Olsson, J. (eds.) *Geographies Inner Worlds*, pp 342 362. New Brunswick, NJ: Rutgers University Press.
Peet, R. (1977). *Radical Geography*. London: Methuen.
Prince, H. (1971). Questions of social relevance. *Area* 3, 150 153.
Sant, M. (1982). *Applied Geography: Practice, Problems and Prospects*. London: Longman.
Smith, D. M. (1971). Radical geography the next revolution? *Area* 3, 153 157.

Archives

M. Kurtz, Open University, Ottawa, ON, Canada

Glossary

Archive A place in which records are deposited and preserved (in print or digital form).

Discourse A dispersed set of representations that produce and authorize/contest meaning.

Document A record on paper or in printable digital form, which furnishes evidence or information. In a history of the modern fact, this objective sense for the word superceded a more performative (now obsolete) meaning in English: "an instruction or lesson; a warning."

Inscription The practices and techniques through which things are rendered onto paper as diagrams, numbers, notes, etc; the processes through which a representation is produced.

Materiality Bodies, objects, and things of consequence, and also the concrete presences and ephemeral qualities that partake in the formation of phenomena.

Memory Shared recollections of the past which, as social activities and material artifacts, forge a sense of identity, place, and social differentiation.

Introduction

Document based and archival research has been quietly undergoing a wholly remarkable change over the last decade. Once largely set aside in human geography as a means to an end, documents and archives are no longer seen simply as a goldmine for the extraction of select elements and valuable information, or as the source of data about some other interesting phenomenon (a person, place, or event, popular iconography, a growth sector, or a government project). Increasingly, they are seen as fascinating artifacts in themselves. Following on the work of Kenneth Foote, Joan Schwartz, Gillian Rose, and Charles Withers, one now finds geographers writing about photograph collections, departmental archives, military archives, imperial archives, postcolonial archives, local history archives, weather records, and even ice cores as nature's own 'archive' of environmental change. In each case, the collections of material that earlier generations of geographers referenced in their footnotes as sources for their data have, instead, become one of the central objects of interest, a dynamic site where knowledge and power have been and continue to be negotiated.

The following pages first sketch several reasons for this development. Among them is the fact that 'the archive' holds a provocative position through its links to influential notions of discourse and to responses of dissatisfaction with discourse analysis in geography. These explanations fold into a sketch of some work that geographers are beginning to draw from in their reconceptualization of the document as a cultural artifact. This, in turn, feeds into an outline of some major theoretical works that inform geographers' interest in archives as curious subjects of study in themselves, a growing interest that might be called (with some trepidation) 'archive fever'. As a caveat, the article does not offer a categorical survey or any broad toolkit of techniques that can be used for archival or document based research on a range of geographical topics. Rather, it offers a brief diagnosis of the archive fever that appears to be slowly spreading among geographers.

Documents of a Material Culture

The fascination with archives and documents as cultural artifacts has arisen in geography for a number of reasons. First, digital publication software and the globalized workforce that assembles cameras, scanners, and other inexpensive mechanical reproduction technologies have enabled scholars to produce facsimiles of particular documents, such that they can then pay close attention to the form and quality of a document through the publication of its image within their research articles. Second, an interest in the history of geography has intensified over recent decades. That literature has often drawn researchers to look at the various modern projects and cultural productions of the discipline, including the documents and archives that earlier geographers produced. Third, there is widespread concern with the consequences of an information economy. This is subject to broad variation, from Castells' broad sweep to an army of economic geographers tracking changes in specific urban service sectors, yet these formulations brush up against archival practices now and again in a variety of ways. Fourth, there are those interested in articulating more precisely the relationship between knowledge production and politics. This has been theorized as a power/knowledge nexus in Foucaultian terms, and as conceptual practices of power in one variant of the feminisms that transformed thinking in geography in fundamental ways and which are now, perhaps, only begrudgingly acknowledged. In these and other cases, arguments that the uneven production, preservation, dissemination, and interpretation of knowledge are in herently political processes have given the emergent analyses of archives a critical edge.

Fifth and most importantly, the growing interest in archives and documents follows from an abundance of work in geography using textual semiotics and discourse analysis. In one sense, this is entirely straightforward: if many more geographers began to tease out the ways in which discourses vary over space as well as how spatiality is discursively produced, then it was a small step from the study of texts and popular media to the discursive ana lysis of forms, reports, and letters to be found in a wide variety of archives. At the same time, these analytical techniques have been subject to criticism among some geographers, from northeast England to northwest North America, who suggest that postcolonial discursive ana lysis (at least) is too cultural and 'textual' in its emphasis, that it fails to make explicit the 'material' practices of power and inequality in a society. But as Clive Barnett argues, postcolonial discourse theorists have been con cerned as much with the economics of publishing and politics of education as with uncovering the hidden meanings of a text. Thus, the oft repeated claim in geography that discourse analysis has been troublingly 'immaterial' may be overhardened. Yet in that sense, the gathering interest in archival material acquired its energy both from the straightforward popularity of discourse analysis (which was also applied to primary documents), and from the ricochet against critiques that such studies have been overwhelmingly textual – too immaterial – in their focus.

In the wake of the increasing fashionability and dis satisfaction with discourse analysis, the work of Bruno Latour has gained numerous allies. This is evident in the literature with which he is closely associated, that is, actor network theory. His work provides a provocative vocabulary with which to think through the materiality of documents one comes across in archival research. Consider Latour's essay, 'Drawing things together'. Here he seeks to replace grand theories of the scientific revolution with more mundane explanations for the power of science, explanations that "take writing and imaging craftsmanship into account." His major argu ment is that we need to focus on two processes that work together: the 'inscription' of material and the 'mobiliza tion' of allies. By inscription, he refers to the technologies through which the world is translated into print material: measuring devices and the making of texts, diagrams, tables, maps, or photos. By mobilization, he means the practices that encourage others (scientists and non specialists alike) to believe in new facts and to behave in new ways. For Latour, the documents found in archival research about, say, a nineteenth century social survey in East London carry potentially important qualities that are all too easily overlooked: (1) these 'inscriptions' of the place are small, flat, and easily handled; (2) they are easily reproduced; (3) they are mobile across space and time, preserving their form while in transit and in storage; and (4) they are readily translated into other texts and documents, some of which can often be found in the archives. In short, his work provides an account of the materials and mundane processes through which one understanding of a place assembled power over the knowledge of others – including those who lived there. Indeed, more than a few geographers have mobilized Latour's writing in their response to the 'all too imma terial' critiques of discourse analysis.

Engagements in the materiality of documents are moving into theaters beyond science studies as well. Here the work of Annelise Riles is illustrative. While many economic, political, and development geographers have turned to ethnographic methods to understand the spa tializing practices of firms, states, and nongovernmental organizations (NGOs), Riles and other anthropologists have been reconsidering the very nature of ethnographic inquiry. Riles asks specifically what can be made of eth nography when the subjects of anthropological research often look all too familiar, when informants are pre occupied with the same sorts of information (interviews, surveys, images, reports, and so on) as anthropologists are themselves. Riles' response is to "find ways to apprehend what is already too familiar" by focusing on the esthetics of bureaucratic practice. Her move to esthetics is not a romanticist step away from materiality. Rather, it deepens an analysis of what Latour had simply described as 'dramatic visual effects', for, amidst the Fijian activist network she works with, Riles pursues that which makes certain documents effective and moving – the agency of nonhuman material in the generation of dimension and depth. Her study stands at considerable distance from Max Weber's work on bureaucratic organization too. Riles' work is not about the formalization of knowledge and practice. She does not assume, like Weber, that documents strip away deeper social contexts; instead, the context of anthropological concern is the making of documents. The shaping of lines on a page, the persua siveness of form, the production of documents: these are becoming more the subject of ethnographies in anthro pology while, in the wake of the 'all too immaterial' critique of discourse analysis, ethnographic methods seem to be used increasingly for research in human geography.

Riles' ethnography may not be one that many geog raphers care to perform. Long standing conventions in document based research distinguish between form and content, and then pay closer attention to the latter. Documents are thus mined for their original content. As primary source materials, their analysis can add to the prestige of the scholar who translates previously hard to find data into a publishable form, provided it is effectively made to speak to the new audience. If the contents include pages of quantitative data, statistical techniques sometimes can be used to produce a few numbers and graphs (averages, variations, trends – all of which are more amenable for inclusion in the space of an article), along with cautions about overextending the inferences drawn from the dataset. Qualitative material, such as images, letters, and manuscripts, can be analyzed using techniques from discourse theory or semiotics. The contexts in which the documents were produced are almost always worth exploration, and contextual evidence can be gathered from the same material as well as other sources. Yet, an evaluation of the utility of the material for one's research objectives (to better understand trends in a local economic sector, or the reasons for migration, etc.) should always come first: how relevant are the documents to that which one is trying to understand? Do the data they present triangulate with other sources? Was its author notable, representative, and reliable? What interests did their author bring to these papers, and what biases are inherent within them? What can be learned by reading against the grain for absences and evasions in this material?

These sorts of questions are invaluable. Yet, many acknowledge that they help researchers move forward largely by putting a series of oppositions under foot: content/form, researcher versus the subject, representation/reality, (general) methods versus (specific) data. This is not to say that all binary oppositions must be renegotiated, nor that this particular set of dichotomies is more problematic than others. Rather, it is only to suggest that the best questions with which one 'begins' to evaluate a set of documents may not be, in every instance, the traditional ones. In some cases and for some purposes, a more effective research strategy may perhaps start by interrogating what differences are assumed with the conventional distinction between form and content, and why. That may in turn provoke the formulation of another field of initial questions for a document based research project, with questions like: how did the material shape and constitute the very thing that one is trying to understand?

Archive Fever?

Compared to studies of documentation and the material of documents in geography, the disciplinary effort that has been put into reconceptualizing the archives – the locations where such documents are preserved – is far more explicit and fecund, making a notable change. Up to the 1980s, any mention of archives was usually confined to footnotes. When they were discussed, this would take place within a methodological treatise in which a rodent model of research prevailed: researchers were like squirrels who would forage for acorns of 'evidence' to be collected from the contents, taken home, and stored for later analysis; and the more experienced among them would instruct others (through a list) as to which were the more promising trees for a process of collection. In the 1990s, the signal work of Kenneth Foote and Joan Schwartz began to rethink the archives as peculiar places in themselves, places that extend the spatial and temporal range of communication, yet exercise remarkable power over what people remember and whose knowledge is marginalized. At the same time, a body of work started to percolate from social theory in which 'the archive' was a central figure. In this material, the archive remained an abstraction. For Michel Foucault, it was the level of practice that enabled statements to survive; in the hands of Edward Said and Thomas Richards, it became a whole body of literature, media, and more that constituted 'the cultural archive' of a society; with Jacques Derrida, the archive represented a figure of interpretive authority. It was not until the turn of the millennium that geographers began to follow the leads of Foote and Schwartz in appreciable numbers, to think through the specificities of particular archives (e.g., this library's collection of photographs, that record of testimony in South Africa) as a place of materiality, memory, and power, and as a peculiar cultural artifact.

If that brand of archival research has become contagious, there are reasons to be cautious before naming it 'archive fever'. First by implication, the label seems to locate its origins in 1996 with the publication of Derrida's book. That text, *Archive Fever*, is almost always cited in the new geography of archives. Second, as Carolyn Steedman says, many "have assumed that *Archive Fever* has something to do with archives (rather than with psychoanalysis, or memory, or finding things)," but the book is not about archives. Derrida begins with an etymology of the word 'archive' (material on the second page is often quoted in geography), but his foremost concern – which consists of a reading of Yerushalmi's reading of Freud – becomes a matter of interpretation. In the edited volume, *Reconfiguring the Archive*, Derrida even reinterprets *Archive Fever*, so as to "leave the book and just discuss burning things here" at the seminar in South Africa where he was speaking. In that text, he positions the book as a homage to (and failed discussion with) Yerushalmi at the Freud Museum. Third then, what is archive fever? In Derrida's rereading, it is the tension that comes from "something in us … driven to destroy the trace without any reminder," a death wish to destroy the material archives and silence the riot of relationships that murmur within them. But in Steedman's interpretation, archive fever is the impossible and insatiable desire to find the beginning of things, to recover their singular moment of inception. She sees a little irony, accordingly, in the 'great assiduity' of historical attention given "to looking for and finding what wasn't there" – that is, an origin – in Derrida's text.

If we were to outline the emergence of recent enthusiasms for research about archives in geography without locating its source in seminal material by a French philosopher, we could do worse than to build the sketch, again, in relation to interests in materiality. The analysis of archives in the discipline followed in the wake of critiques against discourse theory as all too textual and immaterial. Moreover, archives have proven to be a more substantial topic than documents alone, due in part to the localization of material archives. As Steedman points out, archives name "the many places in which the past (which does not now exist, but which once did actually happen; which cannot be retrieved, but which may be represented) has deposited some traces and fragments, usually in written form." Because they are actual places, archives are finite. They cannot hold every trace, nor can they serve every patron at once. Some selection is necessary, which means the archives on which (and in which) geographers conduct research are inherently political. Yet they are not wholly the instruments of human agency: "the archive is made from selected and consciously chosen documentation from the past, and also from the mad fragments that no one intended to preserve and that just ended up there … It is indexed and catalogued, and some of it is not indexed and catalogued, and some of it is lost." As a finite location, an actual place, "you cannot be shocked at its exclusions" (so Steedman soberly tells us), because the archive's "condition of being deflects outrage: in its quiet folders and bundles is the neatest demonstration for how state power has operated, through ledgers and lists and indictments, and through what is missing from them."

Along with materiality, state power is a second major vector with which the interest in archives emerged in geography. A historian and anthropologist, Ann Stoler effectively maps the major benchmarks in an essay titled 'Colonial archives and the arts of governance', published in the volume cited earlier, *Refiguring the Archive*. She sketches a move from archive as source to archive as subject across several fields: historical methods, archival science, cultural theory, the history of science, and colonial ethnography. As a student of colonialism, Stoler's own archival research has led her to look at archives not as 'inert sites of storage' for old documents, but as 'technologies of rule' and governmentality, as instruments of knowledge production that became critical to the administration of European colonies overseas over the course of the nineteenth century. She argues that colonial states were information hungry machines that accrued power from the quantity (not the quality) of the knowledge that their administrations produced, and that their archives "ordered (in both the imperative and taxonomic sense) the criteria of evidence, proof, testimony, and witnessing" to construct such knowledge. Like Riles, Stoler is interested in the form this material takes, as well as its content. Among her examples are commission inquiries, which "organized knowledge, rearranged colonial categories, and prescribed" what colonial officials were charged to know. Anthropologists and historians need these archives for their ethnographic content as they often have, she suggests, "but also for the content in their form", for those forms mark out 'populations' that warrant the interest of the state, offer up historiographies of colonial research, and most of all, serve as monuments to the importance of history to officialdom.

The methodological program that Stoler recommends is rippling across the discipline, as more geographers return to archival research. A recent article in the *Journal of Historical Geography* is indicative. There, Ruth Cragg explores the imaginative geographies of empire produced by the Royal Empire Society Library and the practices of acquisition, classification, display, and study within its walls in London. Stoler's counsel frames Cragg's work, which turns from an extractive to a more ethnographic project in archival research so as to consider the archive itself as a cultural artifact. For instance, the layout of this library's book collection served to reproduce the regionalization and hierarchy of the empire: the form of the archive reproduced a vision of imperial heartland and peripheries. Cragg also argues that the library was imagined in ways akin to what Latour calls a center of calculation (more so or less so by era), that it was understood as a space of imperial knowledge in the heart of empire. But her final caveat about the inadequacy of the terms 'library' and 'archive' relative to the individual topography of this institution follows and completes the warrant that Stoler offers near her own conclusion: "to understand an archive one needs to understand the institutions that it served. One needs to understand what subjects are cross referenced, what parts are rewritten, … which [categories] become common sense and then fall out of favour …."

In geography, recent developments in archival and document based research attend more to the form than the content of an empire of inscriptions. They read documents and archives as important cultural artifacts in themselves rather than as inert sources of other information. These new methodologies have drawn from theoretical innovations in other fields (with work by Latour, Riles, and Stoler, highlighted above), and they also collude with the influences and critiques of discourse analysis in recent years. These research projects are further energized by the dynamic interest in materiality and state power across the discipline. Yet the strongest enthusiasm that animates these fresh approaches to archival research – and one that saw little direct diagnosis in the pages above – is perhaps the open politics that they bring to issues of memory, knowledge, and power.

See also: Actor-Network Theory/Network Geographies; Discourse Analysis; Governmentality; Material Culture; Semiotics.

Further Reading

Craggs, R. (2007). Situating the imperial archive: The Royal Empire Society Library, 1868 1945. *Journal of Historical Geography* 34, 48 67.
Derrida, J. (2002). Archive fever in South Africa. In Hamilton, C., Harris, V., Taylor, J., Pickover, M., Reid, G. & Saleh, R. (eds.) *Refiguring the Archive*, pp 38 81. Dordrecht: Kluwer Academic Publishers.
Foote, K. (1990). To remember and forget: Archives, memory, and culture. *American Archivist* 53, 378 392.
Gagen, E., Lorimer, H. and Vasudevan, A. (eds.) (2007). *Practicing the Archive: Reflections on Method and Practice in Historical Geography*. London: Historical Geography Research Group.
Harley, J. B. (1982). Historical geography and its evidence: Reflections on modelling sources. In Baker, A. & Billinge, M. (eds.) *Period and Place: Research Methods in Historical Geography*, pp 261 273. Cambridge: University of Cambridge Press.
Latour, B. (1990). Drawing things together. In Lynch, M. & Woolgar, S. (eds.) *Representation in Scientific Practice*, pp 19 68. Boston, MA: MIT Press.
Riles, A. (2006). *The Network Inside Out*. Ann Arbor, MI: University of Michigan Press.
Schwartz, J. and Cook, T. (2002). Archives, records, and power: The making of modern memory. *Archival Science* 2(1 2), 1 19.
Steedman, C. (2002). *Dust: The Archive and Cultural History*. New Brunswick, NJ: Rutgers University Press.
Stoler, A. (2002). Colonial archives and the arts of governance. In Hamilton, C., Harris, V., Taylor, J., Pickover, M., Reid, G. & Saleh, R. (eds.) *Refiguring the Archive*, pp 83 101. Dordrecht: Kluwer Academic Publishers.

Relevant Websites

http://www.ica.org
International Council on Archives (ICA). The ICA promotes the preservation and use of archives around the world, and it serves as a channel of professional expertise for archivists. The link provides access to a searchable database of most of the major professional societies for archivists.

http://readingarchives.blogspot.com
Reading Archives. A blog by Richard Cox, a leading scholar in the archival sciences. He offers book reviews and critical observations on the scholarly and popular literature that analyzes the nature of archives or that contributes to an understanding of archives in society.

http://isites.harvard.edu
Social History of Archives. The site features resources of interest for those who study the social history of archives, including bibliographies, papers in preprint form, conference schedules, and calls for papers.

http://www.unesco ci.org
UNESCO Archives Portal. The site provides access to the websites for many archival institutions around the world. It is also a gateway to resources related to records and archives management and to international cooperation in this area.

Arctic

M. Kurtz and A. Watson, Open University, Milton Keynes, UK, and University of Minnesota, Minneapolis, MN, USA

Introduction

With regularity, studies of the Arctic begin with comments about the difficulties of defining the region. Its name is derived from the Greek word, *Arktos* (meaning bear), for the constellation around the Pole Star. In its earlier meaning, the Arctic Circle defined that part of the sky, but it now refers to a line of latitude (currently 66°33′ N) north of which the Sun will remain above the horizon for all 24 h on the June solstice. This shift in English, when the word began to designate a region of the Earth rather than part of the sky, occurred around the sixteenth century when Mercator first circulated a map of high latitudes. Based on information in a book by an Englishman who had traveled North in the 1300s, Mercator's 1569 map popularized the polar projection, a cartographic technique that is still used in textbooks to visualize the Arctic.

If most arctic studies begin with problems of definition, they usually settle on one of five delineations for the region. The best known is the Arctic Circle, that imaginary line of latitude that has been variously calculated and variably drawn across a long humanized landscape. Four other delineations are also often used: the extent of permafrost, tree line, the 10 °C July isotherm, and 60° N latitude. While no definition is suitable for all research projects, the tenacity of these five definitions suggests that, compared to 'the Orient', the Arctic is understood primarily in physical, rather than cultural, terms. While the region's environmental materialities are hardly the product of fantasy, the emphasis on its physical geographies reflects three research trends. First, the parameters of solar heat are understood to set the Arctic – as a homogeneous region – wholly and physically apart. Second, the abundance of geophysical research provides common definitions that are then often adopted elsewhere. Third, fewer geographers have had reason to render the regional coherence of the Arctic as problematic as that of, say, Africa or Asia.

Nonetheless, there is a growing body of work in the discipline that defines the Arctic as an object of knowledge. The 1920s had already seen a profusion of books – by Diamond Jenness, Knud Rasmussen, and Vilhjalmur Stefansson among others – as well as the first two geography textbooks in English: one by R.N. Rudmose Brown, the other by Ludwig Mecking. Against earlier narratives, this material circulated as a new historiography, a network of statements that summarized the accounts of explorers and corrected earlier knowledge about the Arctic. Stefansson was the most adamant in his pursuit for correction. In *Geographical Review* and beyond, he argued that the variety of conditions in "the friendly Arctic" was poorly represented amidst popular misconceptions.

What for Stefansson were the 'standard errors' of arctic discourse became the raw data for John K Wright's landmark essay about the role of imagination in geographical studies. Wright traced popular nineteenth century arguments that claimed an open sea existed around the Pole. His interest in the imaginative geographies of the Arctic gained more momentum after the 1970s, when substantial scholarship revealed the many different ways people imagined the Arctic: as empty or barren, as a rich frontier, as a key transportation route, as a site of scientific discovery, as a strategic military front, as a resource development zone, as a fragile wilderness, and as an ancestral homeland.

Postcolonial studies inform more recent scholarship on the Arctic. Like Edward Said's 1978 classic, *Orientalism*, such work examines the discourses of 'nordicity' or 'borealism' through which the Arctic is represented, along with the shifting grids of colonial power that produced these discourses. Yet studies that ask how 'the Arctic' was constructed as a coherent regional entity in the first place have only begun to appear. Such work suggests that the relations of production for imagining the Arctic involved epistemic violence, and that categories like the Arctic obscure the diverse histories and fractured connections of the people and lands that these categories were made to contain.

Migration, Expedition, and Inscription

In his address to the Association of American Geographers in January 1937, W.H. Hobbs began a familiar historiography of 'Exploration within the Arctic Region' around the year 330 BC, when a Greek geographer named Pytheas began a 6 year voyage to the north coast of Britain. There, Pytheas heard about a land called Thule, which he then claimed to reach after a 6 day voyage. Although discredited, the name 'Thule' has since gained mythic status as the designation for an unknown country at the northern end of the world.

The degree to which that inscription continues to fuel fantasies of a strange world awaiting discovery is remarkable, given that we have overwhelming evidence that Pytheas was not the first to discover the Arctic. North Sea trade was common in his time, and archaeological sites now date human presence in northern Scandinavia to 8100 BC. In North America, a few stone tools left in the Yukon Territory date to 10 000 BC. On Russia's Wrangell Island, the extinction of the last mammoths followed the first arrival of hunters in roughly 3000 BC. At the same time, Tuniit hunters (whose tools resemble those found in Chukotka) began to leave camp remains in Alaska, then east across the Arctic to Greenland, where they likely became the first indigenous people the Norse encountered. Tuniit culture was later displaced by a new Inuit culture, a complex society of large coastal villages found originally around the Bering Strait. At this bustling nexus between continents, an Inuit culture emerged and came to control the flow of iron products – possibly from China – into the Americas more than 2000 years ago. Inuit people soon created the technologies needed to hunt large bowhead whales, and archaeologist Robert McGee believes it may have been a desire for smelted metal that drew Inuit adventurers to a second new source, the Norse in Greenland.

Between Pytheas and Inuit travelers, scholars face potent differences in the way each communicated their discovery. Yet researchers can readily overload the explanatory burden of writing, because there are sometimes ways around communicative barriers. With secondhand records, recent essays present accounts of the Inuit who first traveled to Europe, as well as whalers who crafted material records well in advance of the usual litany of explorers. Such work has moved the scholarship away from stories of technological progress and heroism, to examine instead the context within which certain narratives and images – and not others – gained circulation, privilege, and trust. Mercator's polar projection could have disappeared in 1569, for instance, had it not been useful (as scholars now suggest) to British and Dutch merchants who were seeking a new route to East Asia. While most of its coastlines were speculatively drawn, Mercator's new map nonetheless made it a simple matter to consider three much discussed alternatives – a northwest passage around North America, a northeast passage around Siberia, or straight across the Pole through an open polar sea – with the almost effortless movement of hands and eyes if not ships.

If the context of document production served to privilege some images, it also obscured other knowledge. The opening of the White Sea route to Moscow after 1553 boosted the fur trade between Europe and Siberia, especially for sable. That became a major source of wealth for the Muscovite state, and local depletions of these resources fuelled Russian conquests eastward across Siberia. Small vessels began to traverse many of Siberia's north–south rivers and its arctic coast in a long distance trade network. By the 1640s, Cossack fur traders operated from the Kolyma River. In 1648, one started to work his way further east along the coast. With six small boats, Simon Dezhnev passed through the Bering Strait, eventually to reach his destination at the Anadyr River. There, a century before Vitus Bering sailed in the opposite direction, Dezhnev filed his report with a company officer, where it disappeared from circulation, only to be discovered later in Yakutsk. It serves to illustrate a context in which one narrative about the Arctic remained obscure, while that of Bering gained wide recognition.

Nationalism and the popular press were later forces to privilege some narratives of arctic exploration over others, and this is the context in which the 1845 Franklin Expedition is understood. By then, it had become clear that potential profits were too slim, and risks too costly, for privately funded expeditions to locate a northwest passage to the Pacific. After the Napoleonic wars, the British Navy took up that ambition as a matter of national prestige. Pride was not the only impetus. From accounts of earlier explorers, the Arctic had come to be associated with mystery, grandeur, and austere beauty in British literature. In 1845, that mystery merged with patriotism in support of Britain's best equipped expedition to date. The disappearance of Franklin's ships, and the 40 rescue expeditions that followed, coincided with the emergence of sensationalist journalism. Such newspapers were in wide circulation when reports surfaced about the fate of Britain's finest. The news was doubly disturbing because it challenged two discourses at once: that of British 'mastery' (especially when Inuit testimony suggested the crew had resorted to cannibalism), as well as that about the glorious and sublime landscapes of the far North. Both challenges generated volumes of press in Great Britain about the Arctic.

In the nineteenth century, imaginative geographies of the Arctic emerged in juvenile literature as well. In *The young fur traders*, for instance, Robert Ballantyne used the Canadian Northwest as the setting for his first adventure story for boys. The nephew of Sir Walter Scott, Ballantyne's first job with the Hudson Bay Company had taken him to Canada. Fifteen years later, he embellished his own experience with tales of trapping to create a rambling story of boyish exuberance, masculinity, and mastery. Amidst news about the Franklin Expedition, *The young fur traders* was a success and by 1884, Ballantyne was ranked among the ten most popular writers of juvenile literature. Arctic landscapes often served as the setting in which he constructed his vision of adventure, and with Ballantyne, Jack London, and other writers, adventurous white masculinity also started to become the popular lens through which the Arctic was understood.

That literature inspired another generation of explorers to a quest for the North Pole, notably Frederick Cook and Robert Peary in 1909. In the absence of commercial value, their privately funded expeditions were legitimated as the pursuit for scientific knowledge. Yet Peary's endeavor remained co constitutive with na tional conquest as he proclaimed the Anglo Saxon male body as a virtuous and rational scientific device, at the same time his desire to plant an American flag at the pole projected an older imperial agenda. Framed through a language of nationalism, masculinity, and science in the US press, Peary then positioned himself – not his Afri can American partner, nor the Inuit guides whose vital knowledge he reduced to that of mere laborers – at the leading edge of scientific enterprise, literally clothed in American colors.

Science and State Colonialisms

In the early twentieth century, arctic science continued its service to governments and nations. Little was left of the *terra incognita* that Wright studied 40 years later. Rather than being unknown, many narratives of arctic lands had accrued in imperial archives. Most shorelines had been mapped and aerial surveys would soon be introduced. Sublime landscapes that had once leant currency to grand explorations became reconstituted as political territory through the repeated mapping and construction of boundaries. These ventures helped to produce the regionally partitioned Arctic of today: Greenland; the Canadian Arctic (now divided into three territories); Alaska (purchased from the Russians in 1867 by the United States); Russian Siberia; northern Scandi navia (whose boundaries stabilized after World War II); and Iceland (often included as part of the Arctic).

Reinforcing this transformation were scientific reports that sustained images of an arctic 'wasteland' – simul taneously an empty space, and one filled with challenges for what was deemed the 'northward progress of civil ization'. Except for Russia, governments left the spon sorship of extensive exploration to private enterprises. States instead sponsored intensive exploration – geo logical inquiries like that of George Dawson in the Yukon in 1887, or the US Geological Survey in Alaska under Alfred Brooks in 1903. Such projects produced new vi sions of subterranean space that enabled the North to be seen as a three dimensional reservoir of geologic re sources, visions that obscured any detailed understanding of existing human ecologies that crisscrossed their sur faces. Throughout the Arctic, such investigations ordered up a territory in relation to the sustenance of large in dustrial populations elsewhere.

They also rendered the northern part of the country intelligible as a landscape of raw resources. Swedish plant geographer Gunnar Andersson, for instance, configured the North as a rich timber resource, visualizing its rivers as the perfect transportation device for efficient industry. Dawson's work, republished in 1897, helped to bring an influx of prospectors into the Klondike, and huge min eral based migrations soon followed in search for gold, copper, coal, and oil. The Gold Rush provided material for writers like Robert Service and Jack London, as well as impetus to contest and resurvey the Alaska–Canada border. In northern Siberia, the discourse of mineral wealth and economic development followed after the 1917 Russian Revolution. Amidst a throng of prospectors displaced by the war, the Evenk people refused to serve as guides for the government geologists. Nonetheless, the first Five Year Plan created a state corporation in 1928 to develop heavy industry on Siberia's arctic coast, which brought a million migrants by 1932.

Over the same years, social sciences studied arctic aboriginals. Moscow's Institute of Geography produced ethnographies that represented the Nenets, Evenk, and Chukchi as 'backward' people, but the school soon di vided between liberal ethnographers who saw these indigenous groups as primitive national cultures that needed gradual reform, and Marxist anthropologists who saw them as communal societies in need of rapid as similation with Soviet socialism. Yale geographer Ells worth Huntington developed a theory of climatic determinism that depicted arctic peoples as inferior to those from temperate climates. In the *Annals* in 1927, Clark University's Elmer Ekblaw used a similar moral climatology to ascribe the "homogeneity of Eskimo character, culture, and language" to the environment they inhabited. Such broad brush treatment of the Arctic contributed to an image of aboriginal people as a van ishing race, which justified a regional colonization by a 'more advanced' civilization.

While environmental determinists often suggested the Arctic was a homogeneous region, that image was dis rupted in practice, since the narratives science produced had differentiating effects on both national and arctic identities. As the volume *Narrating the Arctic* illustrates, even the Scandinavian Arctic was heterogeneous in its national identities, its scientific practices, and in the ef fects these narrations had on indigenous peoples. While the practices of Danish scientists valued both fieldwork and the knowledge of Greenlandic Inuit, Swedish sci entists developed a network of scientific institutions that devalued both fieldwork and the Saami people. But whether or not research positioned Native peoples as an inferior culture, field researchers invariably used the help of indigenous community members, inscribing local knowledges into their reports as they erased the visibility of native contributions.

This epistemic violence cut Native peoples out of the decision process in setting research agendas and devel opment goals. The Saami case is telling. Since 1747, Swedish law legally defined Saami identity as 'primitive', and scientists affixed Saami identity to the reindeer they herded. This 'traditional' designation constrained their ability to influence the narrative of the Swedish nation and continues to affect Saami political power, unable to manage other resources – like fish – that were once considered part of their subsistence regime. Indigenous peoples across the Arctic resisted the new political order resulting from these migrations and researches. Else where, however, geographer Kirk Stone advised on pol icies to populate 'the last frontier' in Alaska. Others conducted applied work to overcome 'limits' that the climate imposed on development, through herding and agricultural experiments in boreal and tundra eco systems. Such work attempted to 'advance' Inuit society by transforming them from hunters into herders and farmers, including the introduction of reindeer herds in Alaska and the expansion of the Canadian northern agricultural frontier. At the same time, many in the south of Canada turned to the northern landscape for a sense of distinction and unity. For instance, the concept of 'Nor dicity' – articulated in the 1970s by geographer Louis Edmond Hamelin – described the embrace of the North as essential to Canadian national identity.

World War II prompted researchers to create new arctic narratives in which the region could be understood as a military theater in which the tensions between two nations were expressed. Its strategic importance promp ted what Matthew Farish calls "wars on geography," since military personnel treated the Arctic – contra Stefansson – as a hostile environment. During World War II, Isaiah Bowman stressed that the Arctic was a key geopolitical arena as others gathered aerial photos and used new cartographic technologies for its analysis. In the Office of Strategic Services, Richard Hartshorne helped change the organization of research from disciplinary configura tions to interdisciplinary regional topics, further enabling military narratives to construct the Arctic as a cohesive theater of operation, despite its environmental and cul tural diversity. Following the 1942 Japanese occupation of two islands in the Aleutians, the US established an archipelago of bases across northern North America, from the Alaska Command in Anchorage to the air base at Thule, Greenland. In the Cold War, the Distant Early Warning system strung remote communications stations across the high latitudes. In Canada, the military used Inuit as stand ins for 'the enemy' in strategic exercises, sometimes resulting in real casualties. By 1990, Russia had built over 250 military installations in northern Si beria, shifting the location of many industries to eastern Siberia as well. These activities contribute to a troubled military environmental legacy, the most disturbing of which is perhaps in Murmansk on the Kola Peninsula, where the Soviets dumped nuclear waste on the floor of the Arctic Ocean or left it in crumbling concrete tanks.

Militarization also changed the Arctic's political and cultural geographies. The Korean War added 30% to Alaska's population, catalyzing rapid urbanization and the statehood movement. The territory became a state in 1959, and tensions between Alaska's rural and urban populations continue to the present day. Greenland ex perienced similar tensions to a greater degree after the Danish government relocated Native populations to its urban centers. An attempt to modernize and integrate the Greenlandic peoples into the global economy and create 'productive' self supporting populations, the new high rise buildings disrupted social networks and traditional subsistence economies, creating a legacy of dependence despite a home rule government.

'Nature', the Economy, and Changing Research Initiatives

Post war governments utilized the Arctic as nuclear lab oratory, but a contrary ecological consciousness emerged amidst efforts of 'big science' to control nature. While the Soviets relocated Nenets to conduct tests on Novaya Zemlya, the US Atomic Energy Commission (AEC) tested bombs on Amchitka Island, Alaska. This then in spired the organization of Greenpeace. The AEC's peacetime efforts had included plans to excavate a harbor on the northwest coast of Alaska. Dubbed 'Project Chariot', the plan exemplified science as spectacle, a demonstration of the ability of nuclear technologies to overcome limits posed by arctic environments. Yet the local Inuit community rallied national opposition against Project Chariot, and so too did a few academics. One of the dissidents was geographer Don Foote, employed by the AEC to complete an ethnography of Inuit com munities. When he learned that the AEC altered his reports to make the experiment seem favorable, he be came an activist researcher, informing Inuit leaders and the media about the project's dangers. The episode ex emplifies two concurrent and sometimes contradictory changes in the practice of arctic sciences: first, research methods became more inclusive of aboriginal people's own interests and environmental knowledges; but, sec ond, research agendas began to imagine yet another arctic – a 'wilderness' that overlooked its human history.

The wilderness paradigm also became a new focus for arctic development: nature as a tourist commodity. Arctic tourism is hardly a new phenomenon. By the 1870s, it had become profitable to run steamship tours to one of continental Europe's northern extremities, the North Cape – which a quarter million people were visiting annually by the 1990s. The first packaged tour from Norway to Svalbard, an isolated island group north of 77° latitude, ran in 1871, and the decade also saw regular steamship tours to Glacier Bay in Alaska. Changes in transportation technologies have since had profound ef fects. The 1929 railroad to 'polar bear capital' Churchill, Manitoba; the Al Can Highway, completed in 1943;

aviation technologies and military airfield construction; all allowed a commercial tourist industry to offer more affordable access to small settlements. Geographers have responded to these changes in a number of ways. In the textbook and journal literature before the 1980s, they framed arctic tourism research within a rubric of modernization: the need for economic development and the promises and challenges that the growing tourism sector posed. As cultural tour and later eco tour packages to remote settlements increased, and as the discipline's analytic paradigms focused more on cultural politics, geographers were producing a substantial community focused literature about arctic tourism by the 1990s.

The US Wilderness Act, passed in 1964, considered humans to be destructive of natural environments, and activists and scholars argued that such spaces should remain 'untouched' by humans. Naturalist Barry Lopez, along with humanist geographers like Yi Fu Tuan, may have inadvertently strengthened perceptions of the Arctic as a 'wilderness' wrapped in a silence where "no events ever happen," even while situating these borrowed quotes in an explicitly perceptual geography. Moreover, in the 1970s Tuan would also draw frequently from ethnographies that rehearsed the trope of the vanishing Indian, with only minimal reminders that Inuit people were contemporary agents of knowledge.

Yet indigenous activism did have some impact. Although propelled by a desire to build the Trans Alaska oil pipeline, the Alaska Native Claims Settlement Act became law in 1971, soon followed by the environmentally conscious Alaska National Interest Land Claims Act in 1980. However, since their political autonomy was tempered by colonial legacies and multiscalar social movements, such claims would not be easy for indigenous peoples to settle to their liking. For example, tribes worked with environmental lobbies to prevent 'overdevelopment' in wilderness areas. They thus protected subsistence resources against Project Chariot and oil drilling in the Arctic National Wildlife Refuge in Alaska. Yet the definition of wilderness is not compatible with indigenous subsistence lifestyles. An emerging environmentalist consciousness also contributed to the collapse of the modern fur trade economy. Perhaps the greatest economic devastation came when Greenland's economy collapsed in the face of a Greenpeace protest over the killing of young seals.

This is not to claim that there has been no cause for alliances between environmentalists and indigenous groups around the Arctic. In the early 1990s, the legacy of 'big science' was widely exposed in Northern media: the ecological perils of sunken Russian submarines; deliberate radiation experiments associated with Project Chariot; and the Air Force's radio iodine experiments on uninformed subjects in arctic Alaska. This history contributed to growing interests in critical theory in arctic social sciences, and common methodologies among geographers shifted to studies of the construction and contestation of these landscapes as pristine wildernesses, as sites of a traditional culture, as well as a canvas for a nationalist myth of the North. Such studies are clearly informed by a new cultural geography, but more recent work may be moving the conceptual center of the literature toward post structural geographies of subject formation and science studies around the Arctic.

With these theoretical tools, researchers from multiple disciplines began to question the origination of scientific agendas and develop agendas that were more inclusive of aboriginal interests. Because of the strong applied tradition amongst Arctic geographers, many have maintained access to North American arctic communities with increasing tribal power in the research process. Others, having developed a literature of vulnerability through study of natural hazards, began to articulate theories of environmental change, describing how arctic communities bear the brunt of climate change. Related work contributes to the new literature on socioecological systems, articulating ways that people contribute to ecological processes. Instead of seeing aboriginal knowledge as anecdotal and unscientific, this research assumes that their knowledges are commensurable and potentially compelling. Rather than objects of research, indigenous people are becoming 'partners' in arctic research. Yet George Wenzel has argued that "traditional ecological knowledge" (or TEK) cannot encompass the diversity of local knowledges, and while the signs are hopeful, invocations of TEK still seem to occur more in grant writing than in actual research practices.

Another major response to the environmental and social legacies of 'big science' was to create new governmental institutions that grant aboriginal peoples more political power. In 1978 Greenland achieved Home Rule Government, and in 1999 Canada recognized the Inuit territory of Nunavut. Other nongovernmental organizations (NGOs) represent aboriginal groups that extend across national borders: the Inuit Circumpolar Conference, the Aleut International Association, the Arctic Athabaskan Council, the Gwich'in Council International, the Saami Council, and the Russian Arctic Indigenous Peoples of the North (or RAIPON). While these new political bodies suggest a fracturing of the Arctic's regional identity, the legacies of environmentalism and science continue to assemble the Arctic as a coherent region – especially with the growing recognitions of human induced climate change and its measured affect upon arctic landscapes. The Arctic Council, formed in 1996, aims to provide 'environmental security' by creating an organization that facilitates cooperation between the 'Arctic nations' as well as indigenous member organizations. And with its emphasis on geophysical scientific research, the fourth International Polar Year (or

IPY) in 2007–08 is again indicative of the degree to which the Arctic is defined by its physical parameters rather than in diverse sociocultural terms. In the twenty first century after the many developments described above, social scientists still faced a small struggle to get 'people' included as a thematic component in the IPY's inter disciplinary research agenda.

Although diverse, the Arctic is still an object of knowledge. A rich body of work now traces how the Arctic was envisioned in formative ways and, increas ingly, how 'the Arctic' came to be constituted as a co herent regional entity. When the terms through which the Arctic has been imagined (as wasteland, resource, wilderness, etc.) are brought into question, the fact that each term is not wholly accurate is not the only issue. It is also problematic that the Arctic is simultaneously ren dered up – through many processes of regionalization – as a singular object of analysis, which can then be con sidered comparable to another object. It is partly through such processes that many things 'Arctic' become paired with Antarctica like penguins and polar bears.

See also: Antarctica; Exploration; Indigenous Geographies; Postcolonialism/Postcolonial Geographies; Regional Geography I; Regional Geography II.

Further Reading

Birkeland, I. (2005). *Making place, making self: Travel, subjectivity, and sexual difference*. Aldershot: Ashgate.
Bravo, M. and Sorlin, S. (eds.) (2002). *Narrating the Arctic: A cultural history of Nordic scientific practice*. Canton, MA: Science History Publications.
Chance, N. (1990). *The Iñupiat and Arctic Alaska: An ethnography of development*. Fort Worth, IN: Holt, Rinehart, and Winston.
Farish, M. (2006). Frontier engineering: From the globe to the body in the Cold War Arctic. *Canadian Geographer* 50(2), 1 20.
Ford, J., Smit, B. and Wandel, J. (2005). Vulnerability to climate change in the Arctic: A case study from Arctic Bay, Canada. *Global Environmental Change* 16(2), 145 160.
Grace, S. (2002). *Canada and the idea of North*. Montreal: McGill Queen's University Press.
Hall, C. M. and Johnston, M. (eds.) (1995). *Polar tourism: Tourism in the Arctic and Antarctic regions*. New York: Wiley.
Keskitalo, E. C. (2004). *Negotiating the Arctic: The construction of an international region*. London: Routledge.
Kirsch, S. and Mitchell, D. (1998). Earth moving as the 'measure of man': Edward Teller, geographical engineering, and the matter of progress. *Social Text* 16(1), 101 134.
Loomis, C. (1977). The Arctic sublime. In Knoepflmacher, V. C. & Tennyson, G. B. (eds.) *Nature and the Victorian imagination*, pp 95 112. Berkeley: University of California Press.
McGhee, R. (2004). *The last imaginary place: A human history of the Arctic world*. New York: Oxford University Press.
Slezkine, Y. (1994). *Arctic mirrors: Russia and the small peoples of the North*. Ithaca: Cornell University Press.
Stern, P. and Stevenson, L. (eds.) (2006). *Critical Inuit studies: An anthology of contemporary Arctic ethnography*. Lincoln: University of Nebraska Press.
Wenzel, G. (2004). From TEK to IQ: Inuit qajimajatuqangit and Inuit cultural ecology. *Arctic Anthropology* 41(2), 238 250.
Wright, J. K. (1953). The open polar sea. *Geographical Review* 43, 338 365.

Relevant Websites

http://www.nativescience.org
Alaska Native Science Commission (1994) an organization aiming to facilitate more equitable relations between scientists and Native communities through providing information and referral services to researchers and communities desiring to develop projects in Alaska.
http://www.acia.uaf.edu
Arctic Climate Impact Assessment (2000) a project initiated by the Arctic Council and the International Arctic Science Committee to evaluate and synthesize knowledge about climate change in the Arctic.
http://www.arctic council.org
Arctic Council (1996) a high level intergovernmental forum to address the common environmental, social, and economic concerns of its eight member states.
http://www.arcticnet ulaval.ca
ArcticNet (2003) a network of scientists and managers in the natural, human health, and social sciences, along with partner Inuit organizations, northern communities, government agencies, and private sector groups, for the study of climate change in the Canadian Arctic. ArcticNet provides international news about the arctic environment and circumpolar politics.
http://www.iassa.gl/index.htm
International Arctic Social Sciences Association (1990) an association representing arctic social scientists that promotes the communication and coordination of research.
http://www.ipy.org
International Polar Year (2007) the fourth international science initiative promoting an intensive, interdisciplinary program of research to provide a 'snapshot' of the state of the Earth's polar regions, following from the third International Polar Year in 1957 58.
http://www.inuitcircumpolar.com
Inuit Circumpolar Conference (1977) the nongovernmental organization that represents approximately 150 000 Inuit people across the Arctic.
http://www.uarctic.org
University of the Arctic (2005) a cooperative network of universities, colleges, and other organizations that share resources, facilities, and expertise so as to build knowledge networks and forge a shared identity between northerners who learn and work together.

Art and Cartography

C. D'Ignazio, The Institute for Infinitely Small Things, Waltham, MA, USA

Glossary

Dérive A term coined by the *Situationist Internationale* to denote a mode of experimental behavior or way of passing through urban spaces. In practice, *dérives* are often ways of walking that attempt to research, through direct observation and/or individual experience, diverse facets of the spatial environment.

Détournement A term coined by the *Situationist Internationale* to denote a process that involved removing cultural signs and media elements (painting, literature, film, words, and gestures) from their original context and recontextualizing them in a new context, often for the purposes of social and political critique.

Psychogeography Entails the study of the precise laws and specific effects of the geographical environment, consciously organized or not, on the emotions and behavior of individuals.

Introduction

In the last century, and especially in the last 30 years, artists have made maps, subverted maps, performed itineraries, imagined territories, contested borders, charted the invisible, and hacked physical, virtual, and hybrid spaces in the name of cartography. Numerous names have been suggested for various strains of this intersection: 'psychogeography', 'locative media', 'experimental geography', 'site specific art', 'new genre public art', 'critical cartography' (slightly different from the academic versions), and 'critical spatial practices'. How did we get to this point?

Locating the intersection of art and cartography in the last 100 years does not mean that we should ignore the achievements of artists such as Vermeer (**Figure 1**) or works like the Vatican's *Galleria delle Carte Geografiche*, commissioned by Pope Gregory XIII. But prior to the twentieth century, maps were in the background of paintings, and paintings were insets or illustrations on maps. Art and cartography related to each other via visual cues like insets and tiny representations that made the separation between each distinct.

It is in the early twentieth century, directly after a wave of globalization (1870–1914), that artists began to engage seriously with cartography in numerous ways. Indeed, globalization plays no small part in the contemporary 'spatial turn' of the arts, equally as cause, artistic subject matter, and the professional condition of a field characterized by rapidly proliferating international biennials, conferences, and art fairs. The accelerated accumulation and circulation of capital, conflict, and people around the globe is a phenomenon that required (and is still requiring) diverse societies to develop visual and cultural mechanisms for articulating their relationships to the 'whole' world – a world which, economically and technologically speaking, is already right in their backyard.

For the sake of clarity, this article discusses *three mapping impulses* that, against the backdrop of the social and economic changes to everyday life wrought by war and globalization, cut across the twentieth and twenty first centuries and characterize contemporary art–map practices. These are loose groupings with numerous overlaps, not rigid categories. The artists chosen to represent these impulses are several out of a field of many hundreds of practitioners across the last 100 years.

1. Symbol saboteurs: Artists who use the visual iconography of the map to reference personal, fictional, utopian, or metaphorical places.

Figure 1 *The Art of Painting* by Johannes Vermeer, 1667, Vienna, Kunsthistorisches Museum.

2. Agents and actors: Artists who make maps or engage *in situated*, locational activities in order to challenge the *status quo* or change the world.
3. Invisible data mappers: Artists who use cartographic metaphors to visualize informational territories such as the stock market, the Internet, or the human genome.

Symbol Saboteurs: Artists Who Use the Visual Iconography of the Map to Reference Personal, Fictional, Utopian, or Metaphorical Places

As a result of advances in printing and image repro duction technologies, by the late nineteenth century, maps and globes had become affairs of everyday life in Western nation states. Atlases, such as Blackie's *Imperial Atlas of Modern Geography* (1860), were standard fare in children's education in the late nineteenth century. The classroom globe served to instruct pupils as to the shape of their nation and its role in an all encompassing geo political space. In the first half of the twentieth century, touring maps sponsored by advertisers accompanied the rise of the automobile. Between the 1920s and 1960s, over 5 billion maps were given away at US gas stations alone. In the early twenty first century, with the increasing availability and accuracy of civilian geographic infor mation system (GIS) data, online services such as Map Quest and GoogleEarth offer free, searchable databases composed of millions of satellite map images. Denis Wood estimates that 99.99% of all maps have been made in the last 100 years.

The ubiquity of the map as a means of locating one's place in relationship to the rest of the world created a unique opportunity for artists to exploit cartography's language, symbols, and strategies. Political boundaries became iconic shapes, legible visual markers of identity and belonging that were ripe for artistic distortion, sub version, and reimagination.

Raoul Haussman's *A Bourgeois Precision Brain Incites World Movement* (1920), also known as *Dada triumphs!*), is an early example of the symbol saboteur impulse. In this piece, the artist claims the entire world as the empire of the short lived Dada movement. The movement was in formal and internationalist in organization. Its period, roughly 1916–23, coincided with World War I and rejected capitalist logic, efficiency, and esthetics in favor of cele brating chaos, destruction, and 'anti art'. In *Dada Tri umphs!*, the letters 'DADA' stretch across a map of the world at the top of the photomontage. In an absurd, utopian gesture, Haussman uses the recognizable shapes of the continents and the authoritative power of naming these shapes to appropriate the entire Northern Hemi sphere under the international empire of a tiny art movement. The title and territorial ambitions of the im agery suggest a call to revolution, but the revolution ap pears to have happened in the name of nonsense. Nevertheless, this act of imaginary conquest calls attention to the naming conventions of cartography and the way that a map makes a world.

Contemporary artist Nina Katchadourian works dir ectly with printed maps in order to subvert and distort their iconography. In her cutout map series, she removes everything except the roads from maps of *Austria* (1997; **Figure 2**), Finland (*Finland's Longest Road*, 1999), and the New York Subway system (*Handheld Subway*, 1996). These works leave a fragile, messy tangle of paper behind, disrupting both the legibility of the map and our ten dency to conflate the map's symbols with reality. Rather than referring naturalistically to the roads it represents, the resulting object points back to its material origins as a designed, constructed, and printed artifact, subject to transformation and dissent.

Given the ubiquity of maps in educational settings, national boundaries alone provide rich iconic sources for artistic interrogation. In the case of the US, there are numerous refigurings of the US shape, as it is commonly

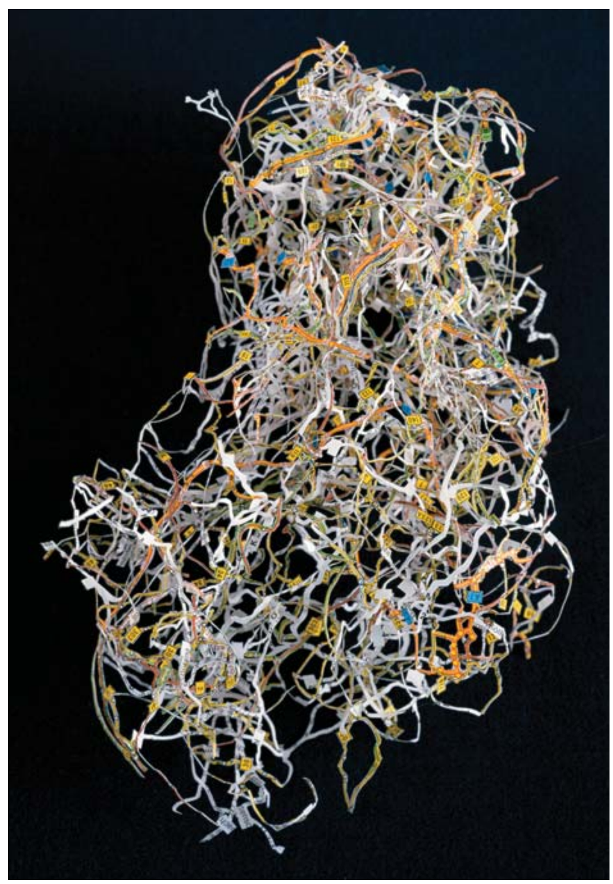

Figure 2 *Austria* (1997). Reproduced with permission from Nina Katchadourian.

understood on the geopolitical Mercator projection classroom map. For example, in 1991, Kim Dingle asked Las Vegas teenagers to draw their country, and then she painted the results (*United Shapes of America*, *Maps Drawn by Las Vegas Teenagers*; **Figure 3**). Through their similar but slightly different bulges and appendages, the teen agers' (see **Figure 3**) drawings indicate the ubiquity of the country's shape in our imagination of place. At the same time, the multiplicity of suggested shapes chal lenges the notion of 'right' and 'wrong' representations of place.

There is also a rich history of artists who use the language of maps to chart emotional, interpersonal, or imaginary territories. These artists use the symbology of the road map or geopolitical map for metaphorical pur poses; as methodology for locating the individual subject within vast psychological, interpersonal, territories; and/or to draw geographical parallels between the vastness of the Earth and the scope of the human psyche. As far back as 1772, there is a map of love (*New Map of the Land of Matrimony*, Artist Unknown; **Figure 4**) that depicts the 'Ocean of Love', 'Bride's Bay', and the 'Land of Matri mony'. Contemporary works in this vein include 40 maps by Wim Delvoye, which allude to human anatomy and reference imaginary territories such as 'Izuch', 'See of FioxXanilla', 'Gnody', and 'Gulf of Doj' (from *Atlas I*, 1992; **Figure 5**).

Finally, there was an explosion in the late nineteenth and twentieth centuries of cartographies of literary landscapes, such as Tolkien's Middle Earth and Faul kner's Yoknapatawpha County, a phenomenon extensive enough to be worthy of a separate investigation.

Agents and Actors: Artists Who Make Maps or Engage in Situated, Locational Activities in Order to Challenge the Status Quo or Change the World

In the twentieth century, particularly during the de struction, chaos, and geo scrambling of World War I, avant garde artists began not only to take on the icon ography of the map but also to envision themselves in the roles of mapmakers, which is to say, capable of leveraging the authority of the map to change the shape of the world.

The Great War left millions dead, caused the dis integration of four empires, created new nations, such as Yugoslavia and Czechoslovakia, granted independence to others (the Baltic states, Canada, etc.), and brought back old nations, such as Poland. During this redrawing of national borders, Tristan Tzara, a central organizer of the Dada art movement, wrote, "The world has gone insane; the artist makes fun of insanity – a gesture very sane, indeed. Throw away the old rules. Manipulate your chance. Dada is a virgin microbe that will get into your brain only in the places where the conventional is not present!" Dada reinforced the project of the artistic avant garde that had begun more than 50 years earlier: to envision new futures, political utopias, and radical spiritual alternatives to the existing order. In their pro fessed ideals, avant garde projects valorize the new, critique established forms (of art making, politics, gov ernance, culture, and class), while seeking to unify or transform art and 'everyday life' in varying degrees. In their numerous manifestations since Dada, avant garde

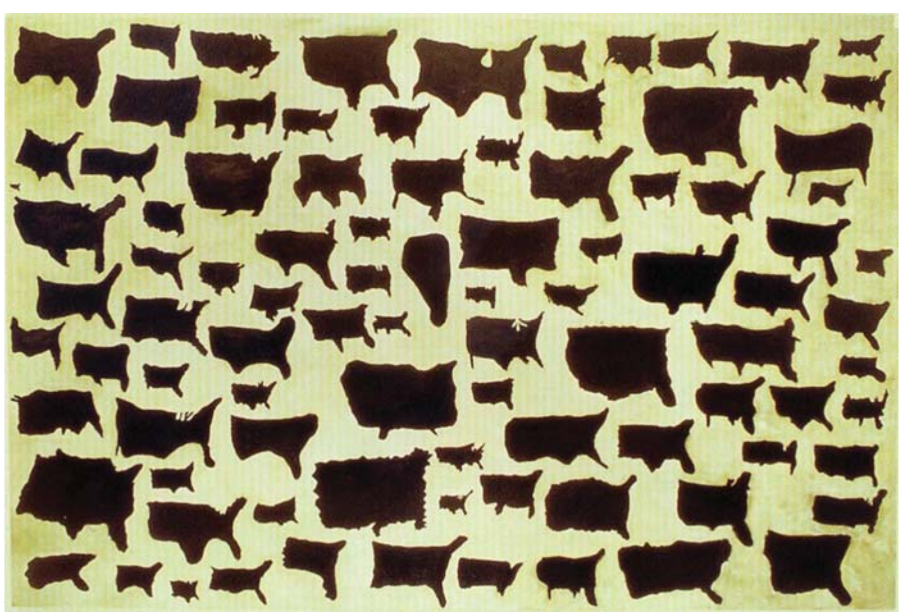

Figure 3 *United Shapes of America, Maps Drawn by Las Vegas Teenagers*, by Kim Dingle, 1991.

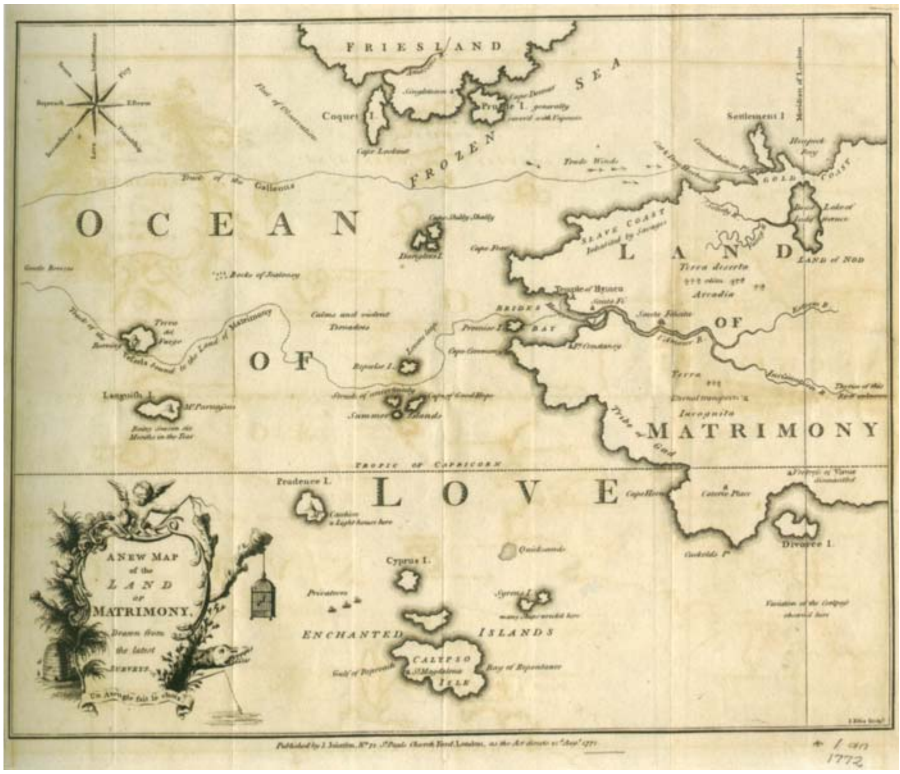

Figure 4 *New Map of the Land of Matrimony*, Artist Unknown, 1772. Courtesy of Map Collection, Yale University Library.

Figure 5 Atlas I (detail), by Wim Delvoye, 1992.

projects have consistently been absorbed by the art establishment and upheld as examples of fine art while simultaneously being critiqued as symbolic, bourgeois, homogeneous attempts at social change and revolution.

The artists discussed as 'agents and actors' pick up on the *avant garde* project to leverage culture for social change. They use cartography from the standpoint of critical cartography, which is to say that they are aware of the power of maps and leverage that authority strategically in order to reshape the world from a social, political, or cultural standpoint.

Artists using maps in their work in the first half of the twentieth century often used them to critique or comment on war, inequality, and other geopolitical realities. For example, Dada artist Hannah Hoch made use of maps in her photomontage works that combined photos, illustration, and typography from mass media sources. Hoch's *Cut with the Kitchen Knife* Dada *through the Last Weimar Beer Belly Epoch of Germany* (1919–20; **Figure 6**) inserts a cutout map – of European countries that planned to give women the right to vote – in the lower right hand corner of the work among icons of industrial modernization, leaders of the Weimar Republic, and images of modern women, such as poets and athletes. Max Ernst's *Europe After the Rain I* (1933; **Figure 7**) depicts an abstract European shape destroyed by war. Likewise, Picasso makes use of found maps and wallpaper in *Women at Their Toilette* (1938; **Figure 8**), where a collaged woman is depicted wearing a dress made of

Figure 6 *Cut with the Kitchen Knife* Dada *through the Last Weimar Beer Belly Epoch of Germany* by Hannah Hoch (1919 20).

Figure 7 *Europe After the Rain I* by Max Ernst (1933).

Figure 8 Women at their Toilette by Pablo Picasso, 1938.

continents. Critics have read this piece as an allegory for the world at the edge of World War II.

Contemporary figures continue this tradition of geo political commentary and critique. Lize Mogel's project, *The Privatization of War* (based on research by Dario Azzellini; **Figure 9**), geographically maps relationships among contemporary wars in Iraq and Colombia, the private military contractors hired to fight them, and the countries, often poor, from which these mercenaries are recruited. On this stark, black and white map, terri tory is represented by segmented cells occupied by corporations. The intent of the map is journalistic: to educate the public about the geographic flows of capital and labor involved in the Iraq War.

In a similar vein, Joyce Kozloff, who has a dis tinguished career using cartographic strategies in her artwork, created the project *Targets* (2000; **Figure 10**) that consisted of a large, walk in globe wallpapered with detailed paintings of US military maps of numerous 'enemy' countries such as Sudan, Libya, and Cambodia. Not only does this work problematize cartography as a tool of conquest and domination, but *Targets* (see **Fig ure 10**) also engages critically with a history of life size globes at international exhibitions – from Wylde's Globe at London's Great Exhibition in 1851 to the Unisphere at the New York World's Fair in 1964–65. Where the latter globes are meant to delight viewers with a visible world, Kozloff's globe intimidates with its large scale, all en compassing threat of military violence.

But activist mappers of the last 100 years also use humor, inversion, and play to denaturalize cartography and to strategically provoke their audiences. The *Sur realist Map of the World* (1929; **Figure 11**) depicts the world with many imperial powers missing (noticeably, the US, France, Canada, and Great Britain). Easter Island is bigger than Australia, Paris belongs to Germany, and the Middle East is entirely absent. Joaquín Torres Gar cía's *Upside down* Map (1943; **Figure 12**) became an icon for the School of the South. It shows the South American continent inverted, with the southern tip at the top of the map and the equator at the bottom in a gesture in de fiance of North–South hierarchical relationships: "the North is now below," declared Torres García. The *Upside down Map* (see **Figure 12**) set a course for nu merous other artists, geographers, educators, and others to question the default orientation and projection of the world map, notably Buckminster Fuller's *Dymaxion Air Ocean World Map* and Jasper Johns' 1967 painting of the same title.

Fluxus maps had an additional political twist in that they called on the viewer/reader to 'complete' the mapmaking actions. Yoko Ono created several versions of *Map Piece* that consisted of instructions such as "Draw a map to get lost" (1964) and "Draw an imaginary map of your dreams" (2001). An early version in 1962 called on readers to draw an imaginary map and use it to navigate actual streets in an actual town. *Map Piece* resonates with the contemporaneous practices of the *Situationist Inter nationale* (SI). Drawing on this work in *You Are Not Here* (2006; **Figure 13**), a collaborative project created by Thomas Duc, Kati London, Dan Phiffer, Andrew Schneider, Ran Tao, and Mushon Zer Aviv, a guide leads tourists on walking audio tours of the war torn Baghdad through the New York City streets. Participants use maps printed with New York streets on one side and Baghdad streets on the other. In this case, the Fluxus poetry of using one place to navigate another is infused with pol itical urgency, that is to say, the need to understand ex perientially the scope of human destruction in an era where images of remote wars resemble video game consoles.

Even though they did not produce much in the way of maps, it was the SI – a group spawned by late Surrealism that broke away from the French Lettrist movement – that had the most significant influence on map practices to come in the next 50 years. Guy Debord, leader of the SI, coined the terms 'psychogeography', '*dérive*', and '*détournement*' to denote critical spatial practices that could be put to use by an individual in encountering and

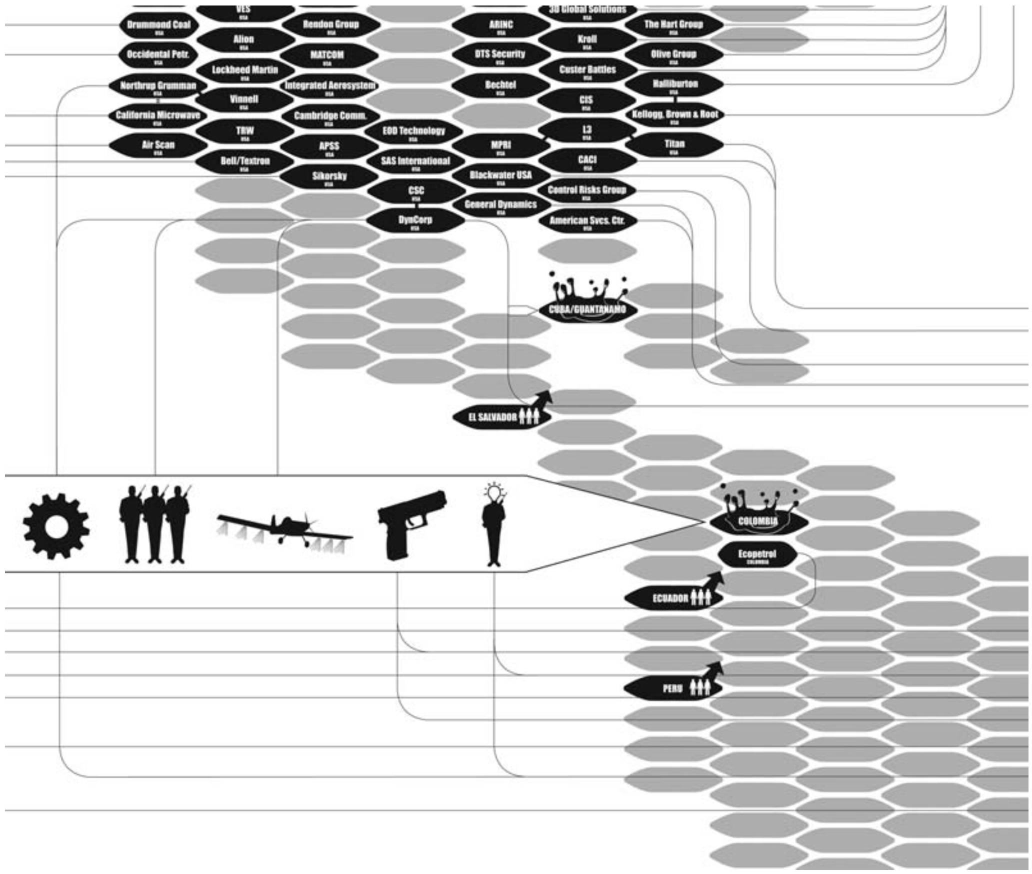

Figure 9 *The Privatization of War* (2007). Reproduced with permission from Lize Mogel and Dario Azzellini.

changing the rationalized, urban environment and the 'society of the spectacle'. Together with Asger Jorn, Debord created *The Naked City* (1957; **Figure 14**), a map of Paris that envisioned its spaces in relation to psycho geographic energies, attractions, and repulsions. But the SI gave up creating art around 1962 and, thereafter, dedicated its energies to explicitly political projects, playing a key role in the French demonstrations of 1968. The key contribution from the SI, in relation to cartog raphy, politics, and art, is that they set the stage for 'mapping' as an activity that was 'performed' through the individual human body in action in public spaces such as streets, parks, and plazas. Not only were artists taking on the role of mapmaker, but they were also taking on the roles of the surveyor, the photogrammetrist, and the data collectors, albeit in iconoclastic, idiosyncratic ways.

While many of the performative, activist cartographic practices that have followed the SI have multiple influ ences and operated without knowledge of the movement, many of these can be conceptually linked to the original questions that the SI posed about the individual or col lective body encountering social and political space. These projects range from meditative to the explicitly political and concern not only urban space, but also suburban, rural, and uninhabited landscapes. One trajectory that persists today is the impulse to map 'the radically specific' – the very small, the hyperpersonal or the overlooked facets of the environment. Richard Long's art practice (late 1960s to present), for example, consists of solitary, multiday walks in locations around the world. He considers the walks 'sculptures' and documents them via short texts, photographs, and gallery installations using natural materials that measure the individual's temporal presence up against a vast landscape.

Likewise, in Teri Rueb's *Choreography of Everyday Movement* (2001; **Figure 15**), dancers carry global pos itioning system (GPS) receivers during the course of their everyday activities. Each person's daily movement created a real time drawing on the web. The artist later printed these drawings on acetate and stacked them be tween sheets of glass, overlaying 1 day on top of the next so viewers could see how a particular person's daily it inerary through the city had changed. In recent years, this trajectory of the radically specific has led to year long research projects to map a single city block in New York City (*One Block Radius* by Glowlab, 2004; **Figure 16**), a map of pumpkins on porches in the neighborhood of Boylan Heights, NC (*Boylan Heights pumpkin map* by Denis Wood, 1982; **Figure 17**), and a map of silent places in London (*Silent London*, Simon Elvins, 2005; **Figure 18**). In

Figure 10 Targets by Joyce Kozloff, 2000. Courtesy: DC Moore Gallery, New York.

Figure 11 The *Surrealist Map of the World*, anonymous, 1929.

their own quiet way, these projects make a political case that challenges the authority, embedded value system, and perceived utility of the map by displacing our attention to things that are definitively small, everyday, and personal.

Figure 12 Joaquín Torres-García's *Upside-down Map*, 1943.

A second contemporary trajectory influenced by the Situationists might be called 'experimental geography', a term coined in 2002 by Trevor Paglen, artist and geographer at UCLA and the title of an exhibition curated by Nato Thompson for the Independent Curators International (2008). With an explicitly social and political orientation, these projects pick up on the Situationist's use of performance and on Fluxus strategies of participation to map the complex territories of a globalized, informationalized world. Paglen's own project consists of a long term, multifaceted investigation into the Central Investigation Agency (CIA)'s 'black world': a shadowy underworld of secret prisons, illegal torture, and classified operations (**Figure 19**). Exhibitions of his work display artifacts collected on his investigative journeys around the world: signatures of people who do not exist at particular places, logos of false companies with fake addresses, and photos of planes that are not supposed to be where he finds them. Paglen's work challenges the notion, popularized since GoogleEarth, that the whole world is now visible, mappable, and knowable. He charts precisely those geographic places that are deliberately hidden from the public eye.

Other projects interrogate land use, ownership, inequality, borders, and diasporic notions of identity and belonging. In 2006, Lauren Rosenthal created a new atlas of the USA entitled *Political/Hydrological: A Watershed Remapping of the Contiguous United States* (see **Figure 20**) that reimagines state boundaries around freshwater

Figure 13 *You Are Not Here* by Thomas Duc, Kati London, Dan Phiffer, Andrew Schneider, Ran Tao and Mushon Zer-Aviv, 2006. Participants use the double-sided map with Baghdad tourist locations on one side and New York City streets on the opposite as they make their way by foot to the Baghdad, Firdos Square audio tour site. Reproduced with permission from Thomas Duc, Kati London, Dan Phiffer, Andrew Schneider, Ran Tao and Mushon Zer-Aviv.

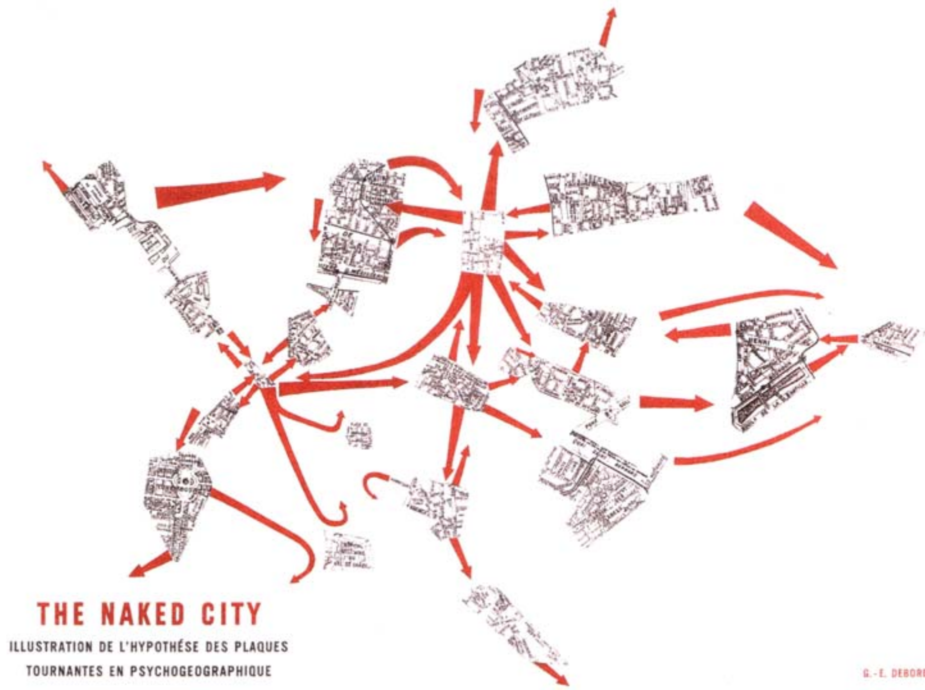

Figure 14 *The Naked City* by Guy Debord and Asger Jorn, 1957.

Figure 15 *Choreography of Everyday Movement* (2001). Reproduced with permission from Teri Rueb.

Figure 16 *One Block Radius* by Glowlab (2004). Reproduced with permission from Christina Ray.

systems. In this ecocentric vision, watershed divides act as territorial borders, allowing citizens to locate them selves within the river networks upon which they depend. In Emily Jacir's project *Where We Come From* (2001–03; **Figure 21**) the Palestinian American artist invited Pal estinians living abroad, who had limited or no access to their homeland, "If I could do something for you, any where in Palestine, what would it be?" She then used her American passport to fulfill their requests, which ranged from playing soccer with children to placing flowers on a mother's grave. The exhibition consists of a moving collection of photographs and texts that document these experiences.

Other contemporary art projects that fall in the cat egory of experimental geography take the form of tools. The Institute for Applied Autonomy, for example, cre ated *iSee* (2001–present; **Figure 22**), a web based soft ware application that would chart 'the path of least surveillance' through New York City. Users need only plug in the beginning and end coordinates of their journey through Manhattan, and the application spits out a printable map that takes the traveler past the fewest number of surveillance cameras. Numerous artists, community groups, and activists have created collectively authored maps with free mapping tools provided by Internet companies such as Google and Flickr. These

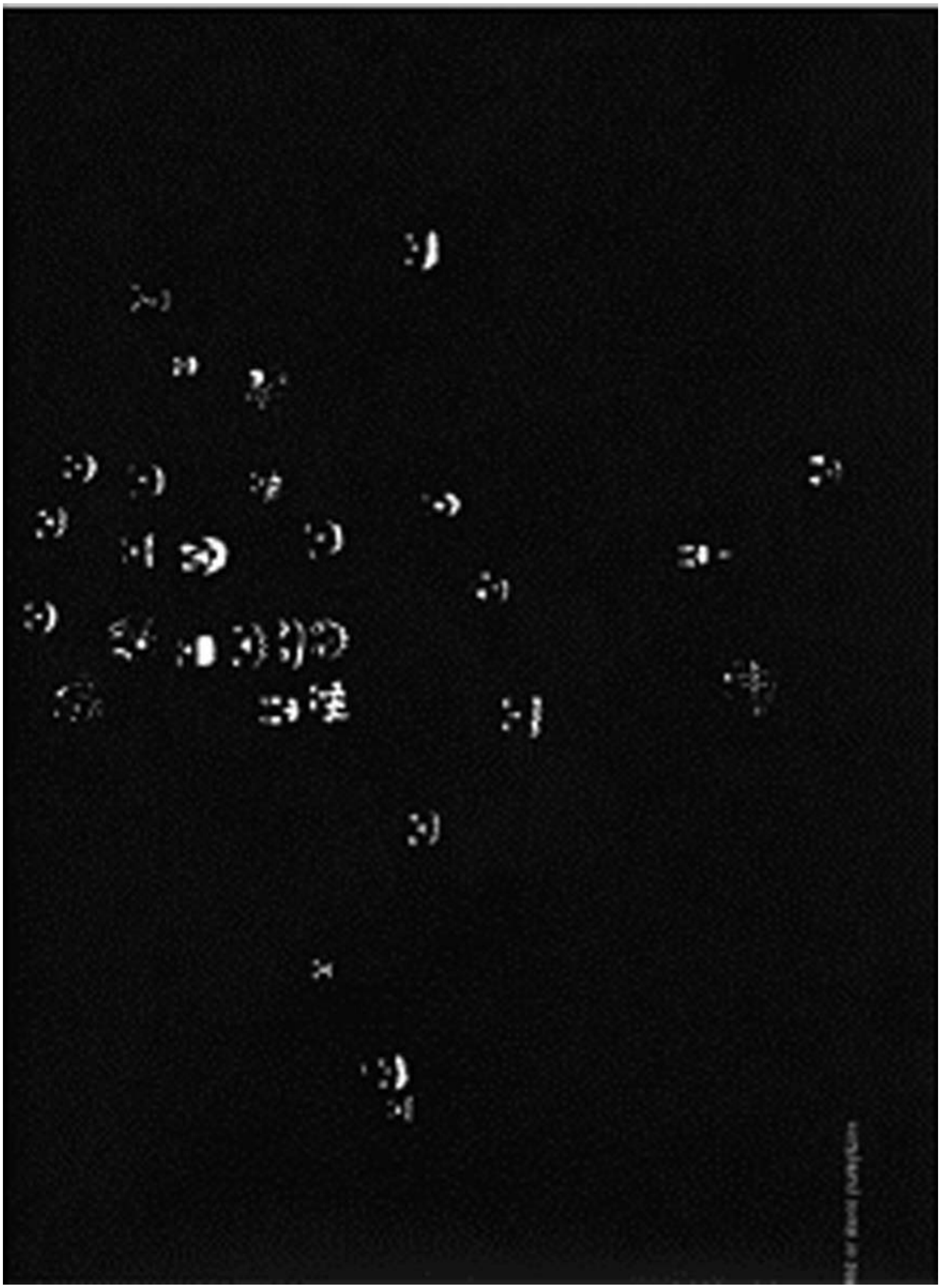

Figure 17 *Boylan Heights* pumpkin map (1982). Reproduced with permission from Denis Wood.

chart everything from Romantic places to Greenpeace expeditions to international graffiti sites.

Many action oriented projects exist at the borders of cartography, less directly referencing the visual language of maps but nonetheless consisting of collections of geographic places with a specific focus (a process of 'mapping'). In *The Pansy Project* (2005–present; **Figure 23**), Paul Harfleet plants pansies wherever people

Figure 18 *Silent London* (2005). Reproduced with permission from Simon Elvins.

Figure 19 Control Tower (Area 52), Tonopah Test Range, NV, Distance ~20 miles by Trevor Paglen, 2006.

report being victims of homophobic verbal or physical abuse, a way of reclaiming both homophobic language ('pansy') and the geographic site of trauma or abuse. Using his body, Alex Villar stages *Temporary Occupations* (2004; **Figure 24**) of private spaces that border on public spaces in New York City. The video shows a series of clips of the artist gracefully jumping fences and slipping behind boundaries of private areas that are adjacent to the New York sidewalks, calling into question the pur pose of these lines of demarcation. The project *Mapping the Working Coasts of Maine* (2005; **Figure 25**) by the collective spurse was designed to interrogate the chan ging demographics, ecosystems, and industries of the Maine coasts, as more people travel there for tourism or to live there permanently. To this end, spurse traveled up the coast of Maine in a boat (their 'laboratory') interviewing coastal workers and residents and asking them to draw psychogeographic maps of the changing social, cultural, and natural landscape around them. In

Figure 20 Mississippi/Mobile-Tombigbee, Pascagoula, Pearl, (a detail from Political/Hydrological: A Watershed Remapping of the Contiguous United States) (2006). Reproduced with permission from Lauren Rosenthal.

many of the process oriented mapping projects, public conversation, dialog, and community building are in separable from the art. Indeed, many of these projects use anomalous, idiosyncratic actions (planting pansies, jumping fences, putting a conversation lab on a boat, etc.) in order to provoke public engagement, raise awareness about the larger issues at stake, and, ultimately, catalyze transformative, collective action.

Invisible Data Mappers: Artists Who Use Cartographic Metaphors to Visualize Informational Territories Such as the Stock Market, the Internet, or the Human Genome

The twentieth century witnessed many technological and social changes as people developed new ways of sensing the world, communicating with each other, killing each other, and moving through space. These changes helped spawn what might be called 'informational territories' – virtual, invisible, infinitely small or large, multidimen sional, time based, and even cultural and political 'spaces'. The Internet, the stock market, the human genome, the electromagnetic spectrum, and global cor porate power all serve as data landscapes that can be mined, visualized, and experienced in different ways. Edward Tufte has traced the graphic representation of

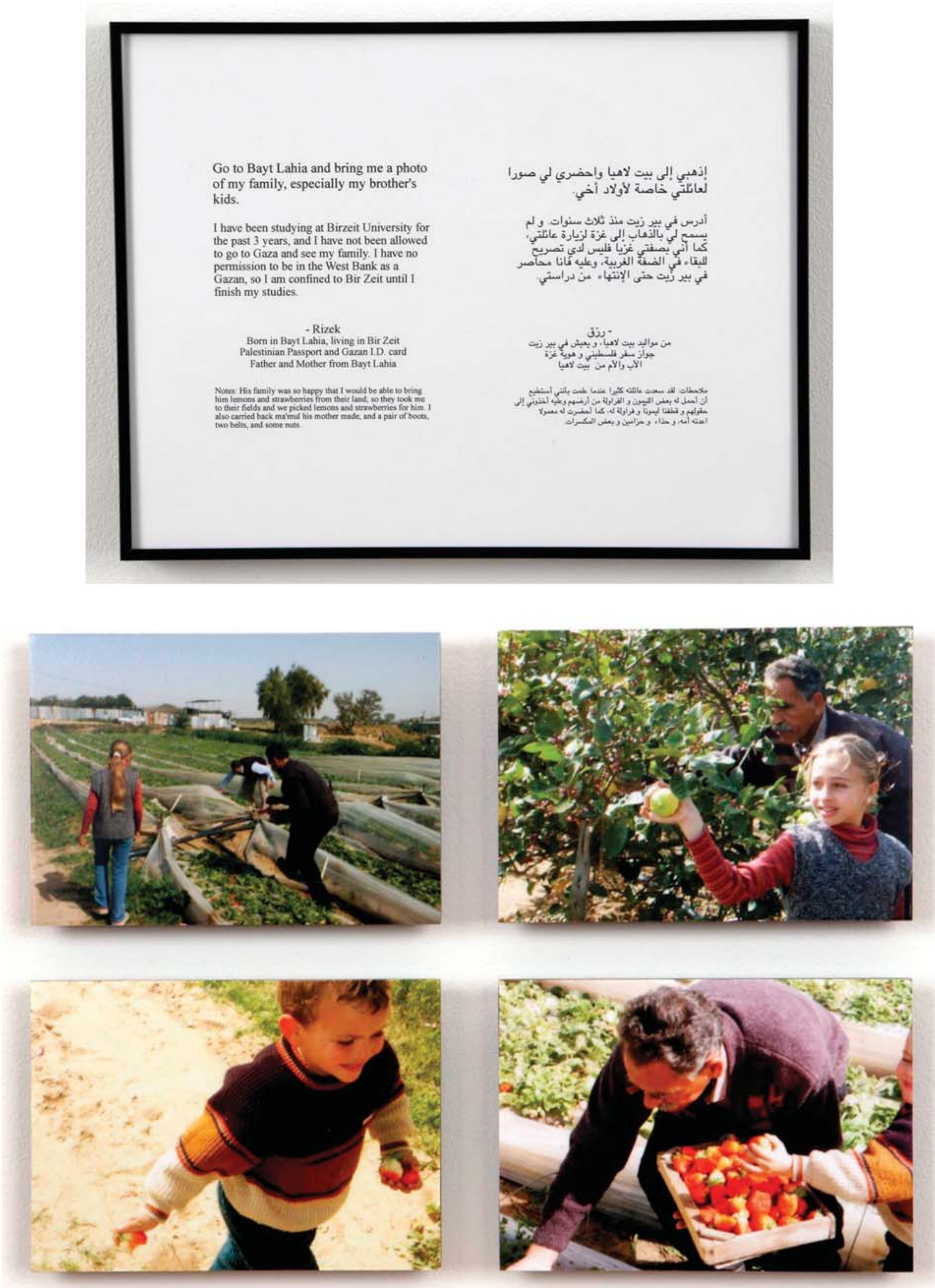

Figure 21 *Where we come from* by Emily Jacir, 2001 3.

quantitative information back to the nineteenth century, but the cartographic metaphor – the idea of data as 'space' and creating visual relationships from nonvisual phenomena as 'mapping' – came later and has been accentuated and popularized in recent years.

Figure 22 iSee by the Institute for Applied Autonomy, 2001 3 present. Reproduced with permission from John Henry, Institute of Applied Autonomy.

What the 'invisible data mappers' have in common is that they use cartographic terminology, typically reserved for discussing the surface of the Earth, and apply it metaphorically to these informational 'frontiers'. An operative principle, descended from information theory, cybernetics, and popular computing culture, is that all the world can be treated as data, ready for selecting, categorizing, visualizing, and revisualizing in infinite ways. This, of course, opens into a 'new politics' of data mapping which looks very similar to the present politics of the geographic map: What data is mapped? What data is left out? Who makes the map, for whom, and for what purpose?

Some projects that map 'the invisible' overlay data on physical or virtual landscapes and are clearly linked to cartographic concerns in familiar ways. For example, Ingo Günther's long term project *Worldprocessor* (1988–2005) is a series of over 300 globes that represent different views of the Earth with datasets artfully overlaid on the surface. In the case of each globe, the dataset is selected carefully to visualize the world with a different intent. For example, *Landlocked Nations* shows only those countries that do not border an ocean or sea, and the wall text discusses the economic and social impact of the absence of a body

Figure 23 The Pansy Project (2005 to present). by Reproduced with permission from Paul Harfleet.

Figure 24 Temporary Occupations (2004). Reproduced with permission from Alex Villar.

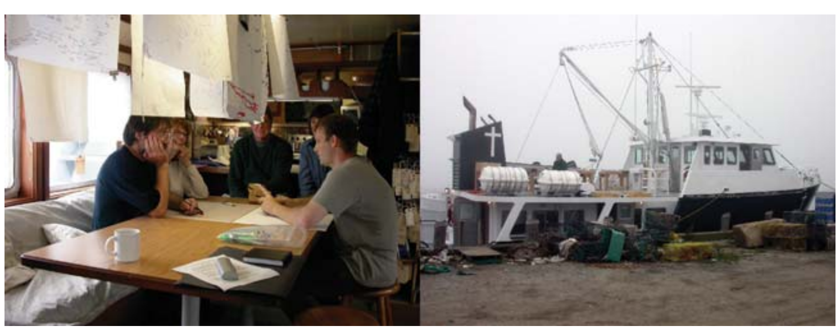

Figure 25 Mapping the Working Coasts of Maine by Spurse, 2005. Reproduced with permission from Spurse.

of water. Other datasets include life expectancy, the world according to Chinese geography, and *Statistical Challenges* (**Figure 26**), which maps elusive, invisible qualities, such as 'Happiness', 'Jokes created per year', and 'Intensity of Dreams', onto a blank globe with no geopolitical lines. Each of these datasets has an associated globe, which is lit from within.

Many projects are entirely dissociated from geographic conceptions of physical space and are preoccupied with visualizing other spaces. In *Genome Valence*, (2000; **Figure 27**), Ben Frye maps genetic data into a delicate, spherical computer graphic that one can search and zoom through. Technically, Genome Valance is a visualization of the BLAST algorithm, the most common way that scientists have of searching through genomic data to see if a particular sequence of letters is found in the genes of other organisms. Fry calls this series of work 'genomic cartography'.

Artists have also mapped distributed global power as its own kind of informational space. In *They* Rule (2001; **Figure 28**), Josh On created an interactive online application that allows the user to map the interlocking relationships between individuals who sit on boards of different global corporations. Using publicly available data gathered from Securities and Exchange Commission (SEC) filings and corporate websites, the application

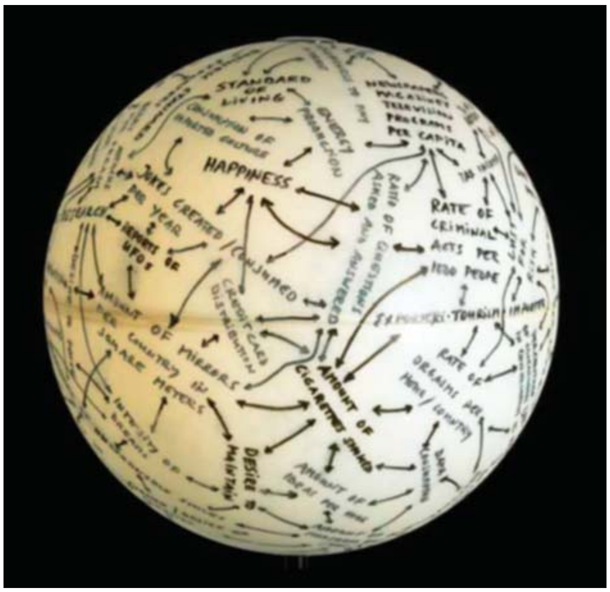

Figure 26 Statistical Challenges, from the World Processor series by Ingo Günther, 1988 2005.

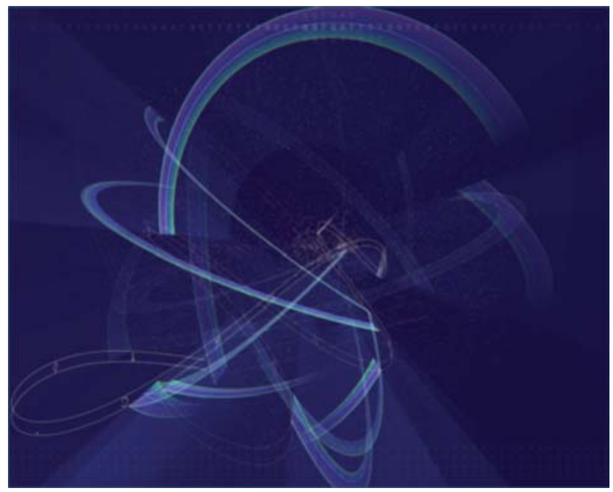

Figure 27 Genome Valence by Ben Fry, 2000.

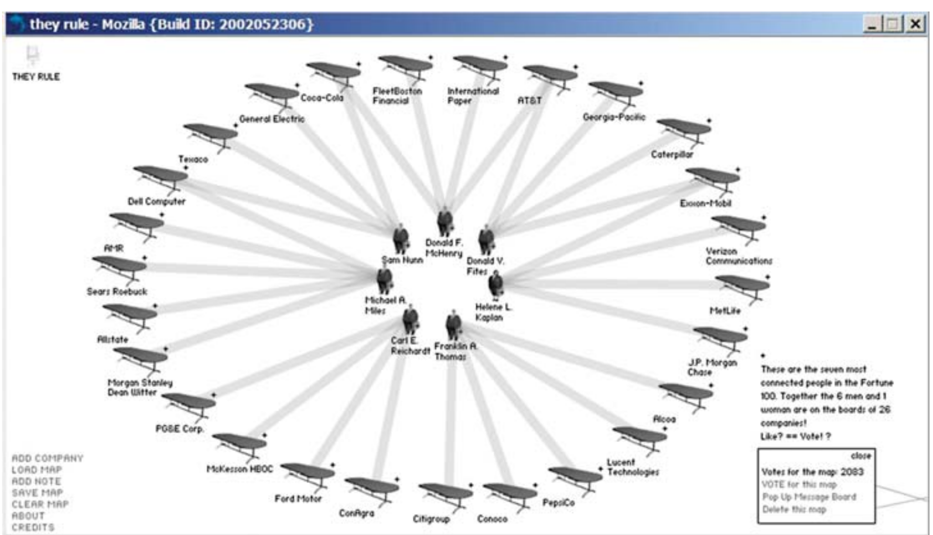

Figure 28 One of the users of *They Rule* by Josh On, 2001.

Figure 29 *1:1 Interface : Every (IP)*, 1999. Source: http://jevbratt.com/1 to 1.

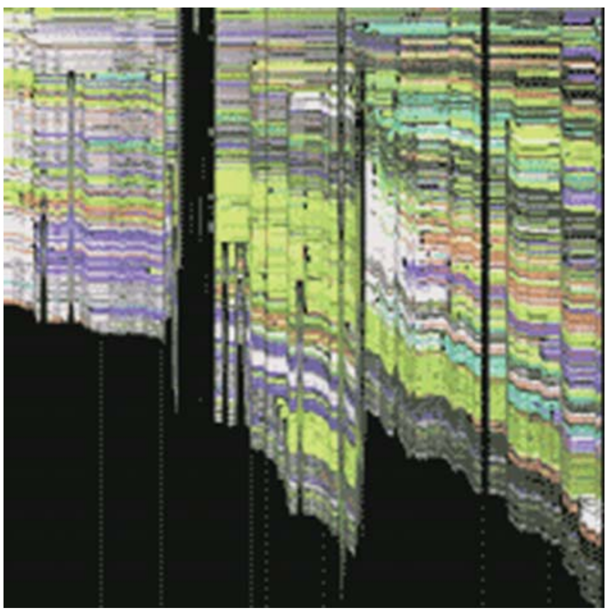

Figure 30 History Flow by Martin Wattenberg, 2003.

demonstrates the unsurprising overlap among corporate boards of directors and raises questions about the ruling power of an elite class of citizens.

Since the mid 1990s and the introduction of 'cyber space' into popular imagination, the Internet has inspired numerous artistic visualizations. Inspired by a Jorge Luis Borges quote, in *1:1 Interface: Every* (1999; **Figure 29**), Lisa Jevbratt visualizes the dataset of Internet protocol (IP) addresses on the Internet (everything from 000.000.000.000 to 255.255.255.255) color coded as to whether or not they are occupied by an actual website. The results of this process are output as an enormous, billboard sized print, which is both physically beautiful and overwhelming. Martin Wattenberg, an artist and interface researcher at IBM, has created visualizations of

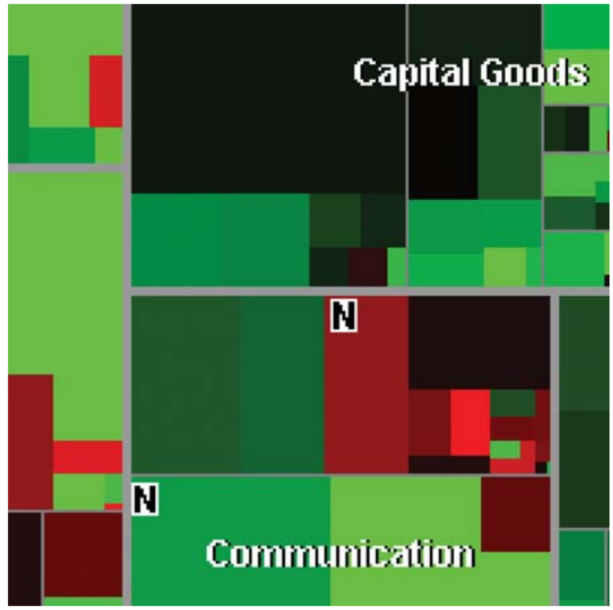

Figure 31 Map of the Market by Martin Wattenberg, 1997.

the history of Internet art (*Idea Line*, 2001) and the entire database, over time, of Wikipedia, the collective en cyclopedia project (*History Flow*, 2003; **Figure 30**).

Finally, the stock market, as data source, has been an inspiration to a number of artists. Wattenberg was com missioned by Money.com to produce *Map of the Market* (1998; **Figure 31**), an online interface that presents an immense amount of real time stock performance data in a single interactive graphic.

Though these data scapes of cyberspace, the stock market, and corporate power do not correspond to physical geography, they borrow spatial metaphors to represent complex informational phenomena.

Conclusion

Though art and cartography have always been in dialog, the last 100 years constitute a veritable explosion of artwork that takes on maps and mapping in order to critique, subvert, re imagine, or simply envision geo graphic and informational territories. There are three loose groupings of important mapping impulses that have characterized the artistic appropriation of cartographic strategies, both literally and metaphorically, from the early twentieth century to present times: 'symbol sabo teurs', 'agents and actors', and 'invisible data mappers'.

Both the Enlightenment project of cartography as a way of accessing fundamental truths about reality and the more recent critical cartographic project in which visual representations are constructed in order to transform the world remain powerful inspiration to artists. Cartography – the idea that we can, should, or must map the world in particular ways – retains an ever growing hold on the artistic imagination. This is especially the case in a world characterized by complexity, inequality, war, and global ization. We live in a world where the idea of place is complicated and where maps (critical, contestational, hacked, technological, and imaginary) are more essential than ever.

See also: Counter-Mapping; Critical Cartography; Critical GIS; Map Hacking.

Further Reading

Abrams, J. and Hall, P. (2006). *Else/where: Mapping New Cartographies of Networks and Territories*. Minneapolis, MN: University of Minnesota Design Institute.

Berry, I., Bender, S., Possidente, B. and Gallery, F. Y. T. T. M.a. A. (2001). *The World According to the Newest and Most Exact Observations: Mapping Art and Science*. Saratoga Springs, NY: Tang Teaching Museum.

Careri, F. (2005). *Walkscapes: El Andar Como Practica Estetica (Land and Scape)*. Barcelona: Editorial Gustavo Gili.

Cosgrove, D. (2003). *Apollo's Eye: A Cartographic Genealogy of the Earth in the Western Imagination*. Baltimore, MD: The Johns Hopkins University Press.

Debord, G. (1995). *The Society of the Spectacle*. Cambridge, MA: Zone Books.

Dziewior, Y. and Montmann, N. (2005). *Mapping a City*. Ostfildern: Hatje Cantz Publishers.

Harmon, K. (2003). *You Are Here*. New York: Princeton Architectural Press.

Hopkins, M. E. and Buscher, M. (1999). *Language of the Land: The Library of Congress Book of Literary Maps*. Washington, DC: Library of Congress.

Kanarinka. (2006). *42 or 363 Definitions of Cartography*. Waltham, MA: Free Press/Sal Randolph.

Kozloff, J. and Kushner, R. (2003). *Boys' Art*. Boston, MA: Charles Rivers Publishing Co.

Kygier, J. and Wood, D. (2006). Art and mapping. *Cartographic Perspectives, Journal of the North American Cartographic Information Society* 53.

Long, R. and Hooker, D. (2005). *Richard Long: Walking the Line*. London: Thames and Hudson.

Wollen, P. (1989). *On the Passage of a Few People through a Rather Brief Moment in Time: The Situationist International 1957 1972 (Inst of Contemporary Art, Boston)*. Boston, MA: The MIT Press.

Wood, D. and Fels, J. (1992). *The Power of Maps (Mappings)*. New York: Guilford Press.

Relevant Websites

http://criticalspatialpractice.blogspot.com
Critical Spatial Practices. A blog run by Nicholas Senn.

http://www.glowlab.com
Glowlab. A Brooklyn based bimonthly psychogeography magazine and annual festival.

http://spaceandculture.org
Space and Culture Blog, A cross disciplinary journal of cultural studies.

http://www.mcwetboy.net
The Map Room, A blog about maps, Category: Art.

Artificial Intelligence and Expert Systems

Y. Leung, The Chinese University of Hong Kong, Hong Kong, People's Republic of China

Glossary

Artificial Intelligence The study of how to make machines perform tasks requiring human intelligence.

Evolutionary Computation Computational methods that mimic the mechanisms of biological evolutions in order to solve complex problems.

Expert System Computer system that solves domain-specific problems much like human experts with high level of performance.

Expert System Shell General purpose tool that provides complete environments for efficient development of knowledge-based domain-specific expert systems.

Fuzzy Logic Nonstandard logic for approximate reasoning with imprecise premises.

Knowledge Representation Formalisms for the representation of facts, truth, and knowledge entities in a machine understandable form.

Neural Network Simple brain-like computational system made up of interconnected processing elements called neurons that process information, compute, and infer by its dynamic response to system inputs.

Spatial Decision Support System Computer system that supports decision making by coupling human intelligence with the capabilities of the computer.

Introduction

With our ability to amass information from various sources, geographers are encountering information explosion in their analyses. Facing an almost unabated growth of information, it is absolutely necessary to apply geographical knowledge and expertise to streamline, analyze, manage, digest, visualize, and integrate information in an efficient and effective manner. Without knowledge, information becomes static and perhaps useless in our analysis of human behaviors or physical processes in space and time. Geographic information system (GIS) renders an efficient way to manage our data. Geographical analysis however goes beyond what GIS has to offer. It is not just the mechanical storage and retrieval of information, and the dazzling display of data. It is rather the kind of analysis that calls for the integrative utilization of information and geographical knowledge to solve structured, semistructured, or unstructured problems with varying degrees of complexity.

In broad terms, the information side deals with the collection, representation, storage, retrieval, processing, and display of data. It, in general, involves the handling, processing, and organization of data suitable for calculating and measuring, as well as for reasoning with and updating of knowledge. The knowledge side comprises the acquisition, representation, and storage of knowledge, as well as the use of knowledge in inference and analysis. In general, it deals with the handling of the body of truth and principles acquired through experience or association. It serves as a basis for inference and analysis. Furthermore, knowledge can be used to guide us to monitor, visualize, explore, integrate, and create information. The interplay of knowledge and information greatly expands the role of information technology in geographical analysis. Apparently, the synergy surpasses the capabilities of spatial information systems which concentrate more on the mechanics of data manipulation than analysis and inference. It is thus absolutely essential that we can fully utilize the complementarity roles of knowledge and information in decision making. Artificial intelligence is thus a way for achieving such goal.

Spatial Knowledge Representation and Inference

Geographical knowledge may be structured or unstructured. We may organize our knowledge in a highly structured form so that problems can be solved by systematic and rigid procedures. Mathematical models, statistical methods, and heuristic procedures are knowledge in procedural form. They follow a rigid framework for the representation and analysis of structures and processes in space and time. Procedural knowledge is effective in system specification, calibration, analysis, scenario generation, and forecasting of well specified and structured problems. Through research, we have accumulated, over the years, a wealth of procedural knowledge which can be effectively utilized for geographical analysis in spatial information systems.

A majority of our knowledge, however, is loosely structured. Subjective experience, valuation, intuition, and loosely structured expertise often cannot be appropriately captured by rigid procedures. They are declarative in nature and can only be represented by more flexible frameworks. Making inference on such knowledge structures cannot follow strict procedures championed by procedural knowledge. Problem solving by if–then arguments is a typical example of using declarative knowledge for decision making. This type of knowledge is effective in solving unstructured or semistructured problems. It is suitable for inference with concepts, ideas, and values. Similar to the use of procedural knowledge, declarative knowledge can be utilized in spatial decision making, especially with spatial information systems. Declarative knowledge is, however, ineffective to solve highly structured problems. Consequently, procedural and declarative knowledge have to be used in synchrony throughout a decision making process.

Once a spatial structure or process is understood and can be specified in a formal and structured manner, we can always capture it by a mathematical model or procedure. The representation of loosely structured knowledge is, nevertheless, not as straightforward. Declarative knowledge representation and inference have thus become a main concern in building spatial reasoning systems with artificial intelligence. To be able to understand and to reason, an intelligent machine needs prior knowledge about the problem domain. To understand sentences used to describe geographical phenomena, for example, natural language understanding systems have to be equipped with prior knowledge about topics of those phenomena. To be able to see and interpret scenes, spatial vision systems need to have, in store, prior information about objects to be seen. Therefore, any intelligent system should possess a knowledge base containing facts and concepts related to a problem domain and their relationships. There should also be an inference mechanism which can process symbols in the knowledge base and derive implicit knowledge from explicitly expressed knowledge.

Knowledge representation formalism consists of a structure to express domain knowledge, a knowledge representation language, and an inference mechanism. Conventionally, its duty is to select an appropriate symbolic structure to represent knowledge in the most explicit and formal manner, and an appropriate mechanism for reasoning.

Through evolution and civilization, human beings have developed a sophisticated way, namely our natural language systems, of representing knowledge. We are able to name and to describe facts by natural language sentences, and are able to reason and to infer with facts and beliefs. This mechanism of representing facts and inferring with knowledge may be captured by logic. Though logic is not a theory on knowledge representation, it provides formalism for reasoning about beliefs. It consists of a syntax, a semantic, and a proof theory which can be utilized for knowledge representation and inference. Propositional logic and predicate logic are typical paradigms.

However, human knowledge is usually imprecise and our inference often consists of a certain level of uncertainty. While uncertainty is of various sources, the one that stems from imprecision is rampant in human systems. To represent and infer with such knowledge, we need a logical system which can handle imprecision. Among existing paradigms, fuzzy logic appears to be instrumental in processing imprecision. Fuzzy logic is a nonstandard logic for approximate reasoning. It is a formalism for drawing possibly imprecise conclusions from a set of imprecise premises. A premise is fuzzy if it has imprecise predicates. In the narrow sense, fuzzy logic is a formalism of approximate reasoning. In the broad sense, fuzzy logic is the theory of fuzzy sets.

Though first order predicate logic gives a powerful mathematical tool to represent knowledge, its theorem proving mechanism is not too efficient and flexible to handle spatial reasoning consisting of a large decision tree or a long chain of inference involving a large set of if–then statements. A system containing an ordered set of if–then rules is called a production system. It can be employed to perform tasks involving deep knowledge reasoning.

Unlike logic and production rules which store knowledge independent of each other and with no interconnections between them, a semantic network is a highly interconnected hierarchical representation of knowledge consisting of a set of nodes connected to each other by a set of directed labeled links. Mathematically speaking, it is a labeled, directed graph. The nodes represent objects which can be facts, events, situations, actions, concepts, sets, individuals, propositions, predicates, terms, descriptions, and procedures. The links represent relationships between the objects. A semantic network provides a structure with which knowledge can be formally represented and efficiently retrieved. Retrieval and inference thus become a search for paths between nodes or a match of patterns in the network. Though semantic networks were developed for the purpose of representing the English language, it turns out to be a rather pictorial and effective tool for representing binary spatial relationships and perhaps human associative memories.

Similar to semantic networks, frames are hierarchical representations of knowledge. They in a way can be interpreted as complex semantic networks with internal structures. Frames are generally used to represent prototypical knowledge or knowledge with well known characteristics. Differing from logic and production rules which store knowledge in small independent trunks, frames store knowledge in larger interconnected chunks (conceptual entities). A knowledge base is then a collection of frames.

A weakness common to semantic networks and frames is that knowledge representation lacks a well defined structure. Knowledge abstraction, encapsulation, and modularity which are considered as salient features of knowledge representation are not effectively realized in semantic networks and frames. These characteristics however can be captured by the object oriented ap proach. Under the object oriented framework, the real world is represented as a hierarchical set of objects linked by some protocols of communication among them. An object is an independent entity consisting of its own at tributes (variables) and methods (operations, procedures, actions) that work on them. The attributes depict the state of an object and the methods are used for intra object controls and inter object communications. The object oriented approach possesses the advantages of knowledge encapsulation and inheritance. It can make spatial knowledge compact and facilitate inference.

In addition to the major formalisms discussed above, other paradigms such as conceptual dependency, script, CYC, logical programming, and case based reasoning are also potentially useful frameworks for spatial knowledge representation and inference. Though each of the for malisms has its unique structure, they are in some ways related and can be made complementary to each other. It is possible to develop hybrid knowledge representation languages which can take advantages of the salient fea tures of individual languages.

Expert Systems for Domain-Specific Problems

The major challenge in the design of intelligent spatial reasoning systems lies on our ability to build into a system mechanisms to memorize and use knowledge extracted from domain specific experts, and to automatically ac quire knowledge from voluminous but incomplete infor mation through learning by examples. Such system can facilitate machine reasoning in a commonly encountered environment where knowledge (in terms of explicitly specified rules) and information (in the form of raw data, digitized maps, or remotely sensed images) are mixed together. The situation is equivalent to human reasoning with previously taught or acquired knowledge that sits in our memories, and knowledge to be acquired by self learning from our everyday experience.

In the recognition and classification of remotely sensed images, for instance, there may be a set of rules obtained from remote sensing experts that can be used to accomplish the task. Thus, reasoning with a mixture of rules and data is a general phenomenon rather than an exception. The advancement of artificial intelligence at tempts to design various means for the achievement of such purpose. Expert systems are products of such undertaking.

Building expert systems from scratch is very involving and time consuming. The general approach is to develop a generic expert system development tool, called a shell, with which expert systems for various problems can be effectively constructed. The shell can assist domain ex perts to build expert systems to solve specific spatial problems. Using a fuzzy logic based expert system shell (FLESS) as an example, it not only can manage rule based inference under certainty and uncertainty, but can also utilize procedural knowledge, GIS, and remote sensing operations in an interactive and integrative manner.

Specifically, the shell facilitates the construction of rule based spatial expert systems with intelligence and decision making capabilities. Any mix of fuzzy and nonfuzzy terms and uncertainties in rules and facts are allowed in the system. It can employ fuzzy logic to handle approximate reasoning and fuzzy numbers to handle imprecision and certainty factors in rules.

The key features of the shell are knowledge base development, tracing, dynamic link library (DLL) tech nology, and operations. Knowledge base development is the part of the shell which handles the construction of domain specific knowledge bases necessary for securing knowledge of a specific problem. Knowledge base is built by the knowledge acquisition subsystem of the shell. It is responsible for storing knowledge entities such as objects, rules, and fuzzy terms acquired through the knowledge acquisition subsystem. Knowledge can be hard encoded from domain specific experts. The acquired knowledge entities, representing expertise, provide knowledge for the inference engine to perform consultations. The knowledge base development part of the tool consists of management modules for objects, fuzzy terms, rules, and inference options.

Tracing is the part which provides inferential strat egies and review management facilities. After defining a knowledge base, consultation can be performed. The system will trace the rule base according to the goal and the tracing method set by users. There are three basic components in the consultation driver. They are the in ference engine, the linguistic approximation routine, and the review management module. The inference engine of the shell supports both forward (data driven) and back ward (goal directed) reasoning. Linguistic approximation is a process that maps the set of fuzzy subsets onto a set of linguistic values or expressions used in decision making. The review management module monitors and traces relevant rules and facts (objects with inferred values) at any time during consultation. It is also responsible for tracing the reasoning chain when explanations are re quired. The system provides two types of explanations. Users can ask 'why' a fact is required by the system and 'how' a fact is established. This module can also handle 'what if' reviews which find out what conclusions will be deduced if certain facts are changed. The feature is es pecially useful for decision makers to evaluate different spatial options or scenarios. The ability to provide con sultations is an important part of an intelligent system and FLESS is equipped with such capability.

DLL technology is the part which manages com munications with the outside environment such as ex ternal libraries and databases. In FLESS, function calls in rules are implemented by the methods of DLL, and it provides a mechanism so that data can be exchanged between an application (e.g., a GIS application involving the use of mathematical models; automatic knowledge acquisition by neural networks or genetic algorithms) and the shell. The application can pass the data required for rule based inference to the shell. The shell will imple ment the inference based on the predefined knowledge base and the given data. After making an inference, the shell will display the result or pass it back to the appli cations for further analysis.

Operations is the part that manages systems oper ations and file manipulations. There is an assortment of operations in the shell. Pull down menu and toolbox are major operations for the manipulation of knowledge bases and inference. Under pull down menu, file oper ations of the knowledge base include the building of a new knowledge base, retrieving an existing knowledge base, editing a knowledge base, saving a knowledge base, and setting system preferences for external editors. Furthermore, operations for consulting a knowledge base throughout an inferential process are implemented. Help menus for objects, rules, fuzzy terms, and helps are also provided.

Acquisition of Spatial Knowledge

Symbolic approaches to spatial knowledge representation and inference are discussed and applied to construct expert systems in the previous section. Logic (fuzzy and nonfuzzy), production systems, semantic networks, frames, object oriented programming, and their hybrids all belong to symbolic systems in which knowledge is modeled by symbols. Intelligence is realized by a sym bolic structure in which symbols can be manipulated and reasoning can be made. The advantages of the symbolic approaches are that they provide a structured represen tation of knowledge so that processing elements corres ponding to meaningful concepts and inference can be traced and explained. The approach is thus a top down approach which gives consensus knowledge to a system by instructing it what to feel and respond without having to gain knowledge through experience. It corresponds to our learning of spatial knowledge from a domain specific expert.

The symbolic approaches are however insufficient or inappropriate to construct intelligent spatial decision support system (SDSS). Symbolic systems are usually intolerant to faults and inefficient in automatic know ledge acquisition and learning via sensation and experi ence. Intelligence is not attained through evolution, like human race, of these systems. Modeled after the human brain, the neural network approach, on the other hand, provides a mechanism for knowledge acquisition. A neural network is a massively parallel structure in which a large number of processing units (called neurons), each performing simple computational tasks, are connected together to represent and acquire knowledge, and to make inference. Geographical information is distributed to the neurons, and knowledge is encoded by connection strengths and acquired through a learning process. It is, more or less, a bottom up approach and is regarded by some as an efficient model for recognition, content addressable memory, and associative reasoning. Differing from the symbolic systems, neural network models ap pear to have stronger learning capability and higher tolerance to faults. Their dynamic behaviors and rea soning processes are, however, difficult to explain. At tainment of intelligence generally takes a longer time and tends to fail in complex situations.

Parallel to neural networks, evolutionary computation which imitates biological evolution can also be used for automatic spatial knowledge acquisition through learning by examples via a parallel multipoint optimal search mechanism. We can, for example, learn rules by evo lutionary computational models.

One may wonder why we need neural networks or evolutionary paradigms for geographical analysis. The answer to the question is mainly twofold. First, as dis cussed above, the neural network or evolutionary para digm provides an alternative framework for spatial knowledge representation and inference. Problems such as spatial feature extraction and pattern recognition which cannot be appropriately modeled by the symbolic approaches may be effectively captured by these dis tributive models. Second, they may serve as a means to acquire spatial knowledge (which may be in symbolic form) through automatic learning.

Taking information, structured and unstructured knowledge as a whole in geographical analysis, their coordination should take on the format depicted in **Figure 1**. Sitting on top of the hierarchy is the deep knowledge which captures complicated spatial relation ships that are, at the present moment, not easy to be automatically acquired through experience or learned by examples as advocated by neural networks, evolutionary computations, and other machine learning models. This is the type of domain specific knowledge which can be effectively instilled by experts as symbolic or highly structured knowledge. Knowledge in the second level down the hierarchy is of lower level of intelligence. It is responsible for tasks which do not require deep thinking but fast extraction or recognition of spatial

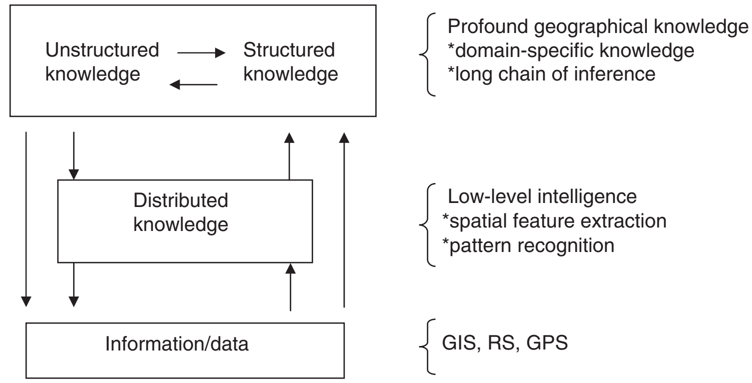

Figure 1 Coordination of knowledge and information.

features or images. It belongs to the perceptual and cognitive functions of our audio and visual faculties which efficiently detect structures and motions from data. It provides fundamental information for our in depth thought processes. Such level of intelligence can be rather effectively imitated by neural or evolutionary computations.

Apparently, the bottom up process is similar to our unsupervised learning process which acquires knowledge from information and examples. The top down process, on the other hand, is our learning process which uses accumulated knowledge to orient our faculties to look for useful things contained in information or our everyday experience so that new knowledge can be formed.

To recapitulate, humans often reason with intuitions, values, experiences, and judgements. We tend to organize our knowledge with loosely defined concepts and struc tures. Reasoning may not require highly structured mathematical models but loosely structured common sense that has been extensively studied in artificial in telligence. Nevertheless, we generally need both to solve complicated problems.

Intelligent Decision Support System – An Integration

In geographical analysis, the demand on domain specific knowledge, technical know hows, and accessibility to data is tremendous. It is difficult to make good decisions without powerful systems to provide, in an integrative way, supports in various phases of the decision making process. Taking the three phase process as an example, the system needs to support, for instance, problem diagnosis, access and scanning of databases, interpretation and monitoring of information in the intelligence phase; generation of alternatives and prediction in the design phase; and analysis of scenarios (e.g., what ifs), explan ation, and justification in the choice phase. It is almost useless if such a decision support system has no intelligence to handle information efficiently and to apply the right kind of knowledge to assist the decision making process in a user friendly manner.

For a system to be able to support decision making, it has to possess a certain level of intelligence. Therefore, an SDSS should be able to reason with structured and loosely structured knowledge. They should be able to manage data and user communication efficiently and effectively. Their development calls for the utilization of artificial intelligence and knowledge engineering meth ods to represent and infer with spatial knowledge; soft ware engineering techniques to manage systems development, information and control flows of models and data; and spatial information system technologies to process and display data. All these have to be integrated in a seamless manner.

To be generic and economical, instead of building an SDSS for a specific problem from scratch, we need to develop an SDSS development environment (shell or generator) so that experts can use it to build effectively and efficiently a variety of domain specific SDSSs. That is, we should have a general development tool which decision makers can use to customize, modify, adapt, and evolve SDSS for solving specific spatial problems.

The general architecture of such SDSS shell is de picted in **Figure 2**. The core of the system is an expert system shell, which, standing alone, can be used as an expert system development tool. When used in an SDSS, the expert system shell directs control flows and infor mation flows. It provides facilities to represent and store domain specific knowledge acquired from experts or learning examples. It can also contain meta knowledge for inference control, systems and user interface, and external communication. The shell has in its possession inference mechanisms for reasoning with loosely struc tured spatial knowledge. It is the brain of the SDSS.

To utilize spatial and nonspatial data, the expert system shell has an interface with external database management system (DBMS) such as GIS, relational databases, and remotely sensed information systems. The

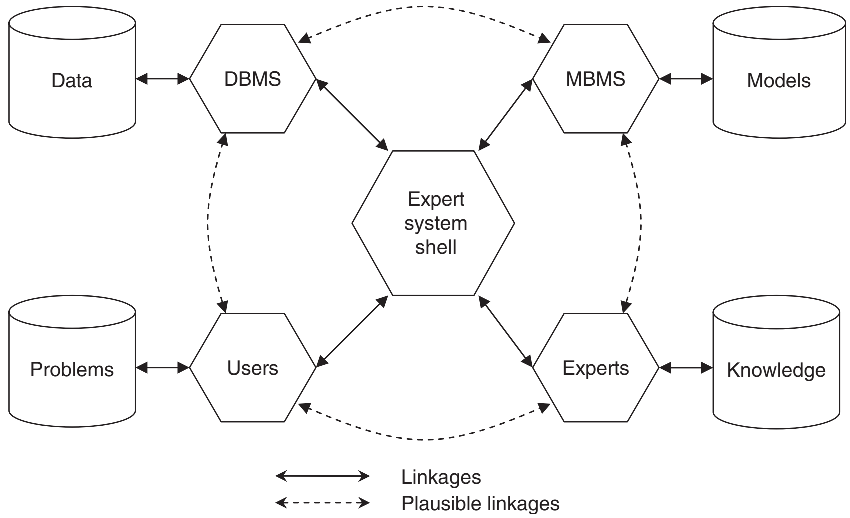

Figure 2 General architecture of an intelligent spatial decision support system.

communication between the expert system shell and the DBMS can be carried out by intelligent expert system database communication methods.

To facilitate the utilization of externally stored pro cedural knowledge such as algorithms, statistical pro cedures, and mathematical models, an interface with a model base management system (MBMS) should be in corporated into the expert system shell. Parallel to DBMS, MBMS organizes procedural models into an easy to use structure. Calls to MBMS can be invoked by meta knowledge in the expert system shell.

In addition to linkages to DBMS and MBMS, friendly user interface and knowledge acquisition modules are essential parts of the expert system shell for human–machine interaction. Communication between DBMS and MBMS, users, and experts should also be considered.

Based on the general architecture, an SDSS shell has been developed, implemented, and applied to construct SDSS to assist spatial decision making tasks. The FLESS is a general SDSS development environment constructed for the purpose of building SDSS to solve specific spatial problems in an effective and efficient manner. To facili tate learning by examples, automatic knowledge acqui sition functionalities materialized as the neural network and genetic algorithm modules have been built into the system. Whenever necessary, neural networks and/or genetic algorithms may be utilized by FLESS to extract spatial features or rules from data which can sub sequently be stored as a knowledge base for some domain specific problems. Of course, neural networks and genetic algorithms can also be utilized directly to calibrate models or to process data in GIS.

Over the years, a number of SDSSs have been de veloped for solving practical problems. Typical examples are decision support system for the management of water pollution in tidal river networks, classification systems for remotely sensed images based on fuzzy logic, neural networks, genetic algorithms, and hybrids, as well as system for flood simulation and damage assessment. All of these systems are built fully or partly by the generic SDSS shell discussed above, and they all integrate spatial and nonspatial data, as well as structured and/or un structured geographical knowledge. Though the appli cations are problem and site specific, the knowledge and technology are transportable and can be made useful to other places of the world.

Similar to the above approach, artificial intelligence has been employed to construct expert systems for landfill design and management, environmental man agement, landslide hazard monitoring, geographical database management, and spatial analysis.

Conclusion

In the future, geographical analysis will increasingly rely on information and the associated technology. We will have to find spatial structures and processes from very large databases which are multisourced, multiscaled, het erogeneous, imperfect, and dynamic. Creating spatial data mining and knowledge discovery capability is thus a ne cessity in the development of intelligent SDSS. Through the electronic super highway, information and knowledge will take on a highly distributive form. The globalization of information and decision making is rapidly taking place. How to discover useful and nontrivial knowledge in such environment is thus crucial in this knowledge age. Effective utilization and integration of information from various sources with geographical knowledge acquired through our in depth analysis of the human and physical processes in space and time will greatly enhance our understanding of the ever changing environment and chart out courses of action for the sustainable develop ment of human race. Artificial intelligence, in cooperation with geographical knowledge, will continue to play a major role in the development of intelligent systems to support our spatial decision making process.

See also: Fuzzy Set and Fuzzy Logic; Neural Networks; Remote Sensing; Spatial Data Mining, Cluster and Pattern Recognition; Spatial Data Mining, Geovisualization.

Further Reading

Arbib, M. A. (ed.) (1995). *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA: MIT Press.

Back, T., Fogel, D. and Michaelewicz, Z. (eds.) (1997). *Handbook of Evolutionary Computation*. New York: Oxford University Press.

Cortés, U., Sànchez Marrè, M., Ceccaroni, L., R Roda, I. and Poch, M. (2000). Artificial intelligence and environmental decision support systems. *Applied Intelligence* 13, 77 91.

Filis, I. V., Sabrakos, M., Yialouris, C. P. and Sideridis, A. B. (2003). GEDAS: An integrated geographical expert database system. *Expert Systems with Applications* 24, 25 34.

Fischer, M. M. (1994). Expert systems and artificial neural networks for spatial analysis and modeling: Essential components for knowledge based geographical information systems. *Geographical Systems* 1, 221 235.

Leung, Y. (1997). *Intelligent Spatial Decision Support Systems*. Berlin: Springer Verlag.

Leung, Y. and Leung, K. S. (1993). An intelligent expert system shell for knowledge based geographic information systems: 1, the tools; 2, some applications. *International Journal of Geographical Information Systems* 7, 189 199; 201 213.

Leung, Y., Leung, K. S. and He, J. Z. (1999). A concept based object oriented geographical information system. *International Journal of Geographical Information Science* 13(5), 475 498.

Minsky, M. (1975). A framework for representing knowledge. In Winston, P. (ed.) *The Psychology of Computer Vision*, pp 211 277. New York: McGraw Hill.

Rich, E. and Knight, K. (1991). *Artificial Intelligence*. New York: McGraw Hill.

Simon, H. (1977). *The New Science of Management Decision*. Englewood Cliffs, NJ: Prentice Hall.

Zadeh, L. A. (1990). The birth and evolution of fuzzy logic. *International Journal of General Systems* 17, 95 105.

Asia

T. Oakes, University of Colorado at Boulder, Boulder, CO, USA

Introduction – Asia: Idea, Ideology, and Project

As the largest of the world's continents, comprising a third of the earth's land area and some two thirds of the earth's population, Asia's size and diversity have always represented a challenge to the categories and conventions of regional geography. Overcoming this challenge has preoccupied a great deal of work within and beyond the discipline. It has become standard practice, for instance, to carve Asia into more manageable subunits such as East Asia, Southeast Asia, and South Asia, in order to convey some sense, however awkward, of the geographical commonality that must lie at the heart of any regionalization scheme. These regional subunits are mirrored by supranational institutions such as ASEAN (Southeast Asia), APEC (Pacific Rim), ECO (Central and Southwest Asia), and SAPTA (South Asia). Yet the need to denote a relatively precise continental boundary remains troubled by the obvious problems of working out where, for example, 'Asia' ends and 'Oceana', the 'Middle East', or 'Europe' begins. And most popular mental maps of the continent seem to imagine a core that is somehow culturally 'Asian' while simply ignoring those marginal areas that don't quite fit, such as Siberia, and the eastern shore of the Mediterranean Sea. At the heart of this problem lies the persistence of Asia as a meaningful signifier of something that inherently defies regional geography. Asia's heterogeneity is not the issue here (for heterogeneity is a problem shared by *all* regionalization schemes). Rather, it is the binary epistemology of Euro American modernity that continually breathes ontological life into an idea so seemingly flimsy that it would otherwise collapse under the weight of its own absurdity. Asia, then, exists not as a continent or a region, but as 'the other' of the West.

However, this argument – that the 'idea' of Asia has derived its strength by virtue of its role as the other of Euro American secular modernity – is itself deeply problematic for several reasons. Ultimately, one must acknowledge that were Asia merely part of an ideological apparatus propping up the superiority of the West, it would have been the first casualty of the iconoclastic nationalisms that swept across the continent during the early and middle twentieth century. Instead, Asia was embraced by intellectuals and political leaders throughout much of the continent. Thus, while it would be easy to dismiss the idea of Asia as merely a European intellectual and geopolitical construction, a complete understanding of the term must approach it as a product of dialog, negotiation, and struggle within a global system of intellectual, cultural, economic, and political change. Regardless of its origins, Asia remains an idea of great importance for many people living in Asia. Thus, after outlining the ways Asia should be understood as 'the other' of Euro American modernity, this entry will explore the complexities of Asian appropriations and reinventions of the idea of Asia. Such appropriations and reinventions might thus be viewed as postcolonial 'projects', in which Asian identity is being reconstituted in alternative ways.

Asia as Other

What is Asia?

As Edward Said observed in *Orientalism* (1978), Asia has played a crucial role in the stories Europe has told about itself. It has done so by always occupying the other side of Europe in a set of binary couplets: rational–emotional, reason–faith, analytical–aesthetic, and so on. Perhaps the two pairings among these which have had the most purchase have been secular–spiritual, and democratic–despotic. Discourses of Asian spirituality and despotism have long served to make Western secularism and democratic freedoms knowable. It is often noted that all of the world's major religions originated in Asia, and while this bestows Asia with great historical significance, religion has been often viewed as a challenge to be overcome if Asia is to achieve a kind of rational modernity on par with the West. Spirituality then becomes a sort of pathology, this being a perspective seen, for example, in Paul Masson Oursel's *La Pensée en Orient* (1949), in which he wrote that the Asian is "above all homo religiosus." At the same time, however, Asia has long been looked to as a repository of spiritual values long lost in the West, with James Hilton's *Lost Horizon* (1933) offering perhaps the most famous example of this.

The theme of 'oriental despotism' has an even deeper genealogy in European thought, dating at least to Herodotus. Aristotle wrote of Asian despotism in *Politics*, and these ideas were reintroduced into Europe when *Politics* was translated into Latin in the thirteenth century. Later, Montesquieu compared French absolutism to the despotism of Asia, and as the modern European state developed, Asia was increasingly regarded as the example of everything backward and despotic that European states would liberate their people from. Perhaps the most well known articulation of Asian despotism comes from the Marxist idea of an 'Asiatic Mode of Production' and the 'hydraulic civilizations' thesis it inspired. This will be covered in greater depth in a later section. But here it should be noted that an important element of Asian despotism is its link to environmental deter minism. Despotism was, according to this thinking, a necessary result of large scale water control projects for irrigated agricultural production. Whereas Europe's environment was generally seen as offering a series of possible options for human engagement and modifi cation, Asia's was viewed as offering only one possible outcome, and that was centralized, authoritarian, and despotic rule.

Where is Asia?

Given Asia's role as a representational foil in the devel opment of European thinking about the meaning of 'Europe' and 'the West', the boundaries of Asia itself have always been ambiguous. There are a number of different claims advanced regarding the origin of the term. Some argue that it derives from the Greek *asu*, meaning 'sunrise'. But it also possibly derives from the term used by the Hittites in the second millennium BCE for the area where they lived, that is, Anatolia. Toynbee noted that it was Greek mariners who first used the terms 'Europa' and 'Asia' to describe the lands on either side of the Aegean Dardanelles Bosporus divide. More to the point, perhaps was Toynbee's observation that the Greeks themselves spanned this divide, but used the term to mark the inferiority of their Ionian kin who suffered the influence of the Persians on the east side of the divide. Still, Toynbee points out, there remained confusion among the Greeks about whether they were 'European' or not, with Aris totle, for example, arguing that Greece lay in between Europe and Asia.

In *The Myth of Continents* (1997), Lewis and Wigen note that the Romans generally maintained the Greek convention of equating Asia with Anatolia. But they point out that as cartographic skills and geographical knowledge expanded in Europe, the convention of dividing Europe from Asia along the divide established by the Greeks, which extended through the Black Sea, the Sea of Azov, and up the Don River, was increasingly questioned. Strahlenberg proposed the Ural Mountains as an alternative boundary and this was supported by Russian intellectuals eager to link Peter the Great's modernization program with Europe while marking Siberia as Asian and thus subject to colonization and exploitation. Here the boundary has essentially remained, although it is interesting to note that Soviet cartographers put the boundary just east of the Urals.

When is Asia?

Perhaps more significant than the precise geographical demarcation separating Asia from Europe is the role of Asia in European conceptions of history. Dipesh Chak rabarty has observed in *Provincializing Europe* (2000) that European historicism made modernity and capitalism seem global, by making them seem to become global 'over time', originating in Europe and then spreading to the rest of the world. Fabian (1983) has called this move "the denial of coevalness," and it follows from the idea that whereas the West is dynamic and progressing, Asia is static and despotic. For Hegel, Asia lay 'outside history'. Herder saw no historical progress but only static un changing civilization, and John Stuart Mill equated Asia with the unchanging village. Such a view was conveyed as late as 1965, when Romein wrote in *The Asian Century*, "Thus lived the Asian masses, generation after gener ation, century after century, living in AD 1500 just as they had lived in 1500 BC, uncoveting, undesiring, and so renouncing the effort to gain an existence worthy of a human being." To the extent that change did occur, it was viewed as the 'surface storms' of rebellion and upheavals among elites that never reached into the calm depths of agrarian society.

Indeed, the agrarian nature of Asian society – in which 'the peasant' was viewed as the ultimate organic unit of a deep rooted social structure – suggested an Asia governed by the rhythms of the natural world, rather than the linear progression of history. For Hegel, this quality placed Asia at the very beginning of history. And to the extent that historical change was acknowledged, for ex ample, in the development of religions such as Hinduism or Islam, it remained limited to the specific cultural subregions of Asia and not connected with 'world his tory'. History with a capital H, bringing Asia as a co herent space into existence and into the world, arrives only with European trade, colonialism, and imperialism.

The European conception of Asia as an organic ex tension of its environment yet unable to act upon and 'conquer' that environment found its most well known expression in the concept of an 'Asiatic Mode of Pro duction'. This term defined, in a very general sense, an agrarian system of centralized social control necessitated by intensive agricultural production and immune to the social disruptions that produce dynamic historical change. And while it was elevated to the level of scientific truth and promoted as late as the 1950s in the work of Karl Wittfogel, the Asiatic Mode essentially reflected both Europe's eighteenth and nineteenth century ignor ance of Asian society as well as Europe's ongoing need to use an invented conception of Asia as unchanging and despotic to mark itself as dynamic and liberal.

It was Marx who articulated the idea of an Asiatic Mode most clearly. Working with received notions of Asia as stagnant, he saw a distinctive pattern in Asia's lack of private property in land, large scale irrigation works, and omnipotent states. In *A Contribution to the Critique of Political Economy*, Marx mentioned the Asiatic Mode as one of four major societal formations, in which the village was the primary social unit, land was held communally, and social organization was conditioned by large scale irrigation works that required centralized authoritarian control. This explained the predominance of despotism and meant that local societies lacked autonomy. But Marx characteristically devoted scant analytical attention to the concept, and indeed seems to have contradicted some of its tenets in his other work (i.e., his celebration of China's Taiping Rebellion). The Asiatic Mode is significant less because of its ignorance of the specifics of Asian societies (most of its assumptions about, for instance, private property and the lack of local autonomy are incorrect) and more as a conception that allowed enormous parts of the world to be brought into Marxist conceptions of world history. It thus explained Asia's *lack* of history in the only way such a lack could be explained from a European perspective.

It is in this context that Asia has also inspired a belief that upon entering the stream of world history, a 'slumbering dragon' would awaken from the 'stupor of tradition' and become, finally, a force to be reckoned with. From a Western perspective, Asia has experienced any number of 'awakenings', including Japan's early twentieth century victory over Russia, China's post Mao economic reforms, and Central Asia's rebirth in the post Soviet era.

Asia Area Studies

A final chapter of Asia's construction as 'other' comes in the form of post war area studies in United States academic institutions. The study of Asia in the US has always been closely linked to its geopolitical significance, particularly during the years separating World War II and the Vietnam War. But US defeat in Vietnam and the normalization of diplomatic relations with China meant the relative eclipse of Asia Pacific as a geopolitical priority by concerns over its integration into the global economy. As a result, the study of Asia has experienced some significant readjustment to account for the trans national border crossing impulse of a new so called 'Pacific Century'. Once a key battleground for Cold War geopolitics, Asia was reconfigured as an emerging market and key hub of globalization.

One result of this shift has been a questioning of the territorial basis for our understanding of Asia. Given the problems of defining the continent's boundaries, there have been proposals for a more maritime oriented understanding of Asia as merely part of a broader network of connection across the Pacific and Indian Oceans. One example is Sugata Bose's history of the Indian Ocean, *A Hundred Horizons* (2006). Another is the recent 'Oceans Connect' project at Duke University, funded by the Ford Foundation's 'Crossing Borders: Revitalizing Area Studies' initiative. Similarly, in a recent Association of Asian Studies Presidential Address, Barbara Andaya argued that, "If we insist that the sea and those who live with the sea deserve a more prominent place in the study of Asia …this framework will go a long way toward overcoming the confines of so called area studies while redressing the scholarly preoccupation with land based societies that has so informed the presentation of Asian cultures" (Andaya, 2006: 669–690).

Yet while the area studies tradition has thus been moving toward a much more interregional understanding of Asia as more of a node in an ever changing network of connections, it is perhaps premature to assume that a geopolitical interest in territorial based understandings of Asia died with the military fortunes of the United States in Vietnam. The profoundly ironic 'end of history' celebrated by Francis Fukiyama was replaced by the Bush administration's 'war on terror', with Asia, once again, emerging as a key battleground. There remains great interest in United States area studies scholarship in coming to terms with Asia (particularly Islamic Asia) as a cultural–religious phenomenon, and a place still in need of secular modernity and democracy.

Asia as an Asian Intellectual Project

Does Asia Exist?

In 1963, provoked by the appearance of yet another Western book questioning the validity of Asia as a meaningful label for such a vast and diverse space, Indian journalist Romesh Thapar regarded the question of Asia's existence with understandable indignation. But Thapar did concede that the "image" of Asia is "blurred." "Yet if we act with good sense," he continued, "it is possible to salvage the best of what is essentially our common heritage." While it may be possible to deconstruct Asia as a mere invention, and a convenience of Eurocentric epistemologies, the term has also expressed a profound geography of hope among many Asians themselves. This hope has at times been conceived in terms that still privilege 'the West' as the key historical agent, as illustrated here by Dick Wilson in *Asia Awakes* (1970): "The wounds of Western imperialism still throb in the minds of Asians, and the idea of an Asian community, an Asian coherence, remains a favorite medicine to stifle the pain." Such a view risks the assumption that whereas Asians are busy trying to build the impossible (an 'Asian coherence') we in the West now know better. If Asia is to be appre ciated as an ongoing project of dialog, negotiation, and struggle within a global system of intellectual, cultural, economic, and political change, Asian engagement with this project deserves a more thorough treatment. This section will discuss the challenges of Asian appropri ations of Western conceptions of Asia, while the fol lowing section will take up the issue of alternative conceptions advanced as explicitly non Western versions of Asian identity.

Putting Asia into 'World History'

Wang Hui (2005) has argued that the Asian idea in Asia is conditioned by two opposing historical processes: Jap anese colonialism and Asian nationalist–socialist liber ation movements. By the late 1960s, Western observers were viewing socialism as the very 'Asian coherence' sought by the likes of Romesh Thapar. For many of them, like Gunnar Myrdal, socialism only reconfirmed an Asia mired in despotism. But for others, socialism marked the culmination of Asia's integration into world history and united the continent's distinct cultural subregions like never before. For many Asian nationalists and intel lectuals, however, the revolutionary potential of socialism was tempered by the problem of adopting a conception of Asia which was based largely on European ignorance. That is, the idea of an 'Asiatic Mode of Production' re quired significant reconfiguration in order to explain what Asian intellectuals insisted on accounting for: Asia had indeed undergone significant change in the context of 'world history'. How then would socialists reconcile an idea that presumed an Asia 'outside history' with the need for an historical materialist understanding of Asian history that justified socialism as the inevitable outcome of historical progress in Asia itself?

A common response to this problem was to delimit the Asiatic Mode to a specific period of Asian history, one which had long since been left behind, but which still lingered in various forms of class based oppression. Using Marx's own derivation of linear history – culled from the nineteenth century evolutionary theories of Lewis Henry Morgan and E.B. Tylor – leftists in Asia tended to locate the Asiatic Mode in ancient primitive communal, slave owning, or early feudal societies. This enabled a rejection of the idea of Asia as unchanging without questioning the rightness of Marxist politics. And the acknowledgment of the Asiatic Mode's existence was strategic for revolutionaries who saw in the imperial state systems of Japan and China its oppressive vestiges. Others saw the Asiatic Mode lurking in the countryside, claiming that real change in Asia only came in the form of industrial urbanization and urban based revo lution. Asking why such a clumsy and ambiguous idea as the Asiatic Mode gained such attention in Asia at all, Fogel argued that there was a need for the "ontological security" of historical materialism as a foundation for an incipient narrative of national identity as well as a pan Asian commonality that could not otherwise be found.

Another significant appropriation among Asians is the idea of Asia as an essentially spiritual identity. More than historical materialism, the idea of Asia as the birthplace of the world's great religions, and of Asia as still har boring a spiritual essence lost in the secularism of the modern West, has been a resource for dreams of a pan Asian community. Thus, Okakura Kakuzo wrote that "Asia is one." The divisions of physical geography that separated Asia's distinct cultural regions cannot, he continued, "interrupt for one moment the broad expanse of love for the Ultimate and Universal, which is the common thought of every Asiatic race." Rabindranath Tagore evoked Asia as a primarily "spiritual civilization," and Swami Vivekananda reserved for Asia a monopoly on the spiritual contributions to world history. It is signifi cant to point out here that such pronouncements ef fectively turned European conceptions of Asia as (inferior) 'other' on their heads. Okakura thus contrasted Asia's pursuit of Universal Truth with a narrow Euro pean preoccupation with the Particular. And Viveka nanda contrasted Asia's spiritual contributions with Europe's political and scientific contributions. Such ar ticulations, of course, offered little in the way of alter natives to the categories – 'secular', 'spiritual', 'science', 'religion', and so on – whose dominant discursive forms were themselves the product of European modernity, colonialism, and imperialism.

Postcolonial Asia

It is precisely this conundrum that has come to define Asian postcolonial projects. Recognizing that the very categories of modern scholarship are the products of a particular historical and geographical context, post colonial scholarship has been committed to engaging the universals of modernity – citizenship, state, civil society, public sphere, human rights, equality, the individual, the subject, democracy, popular sovereignty, social justice, rationality – in a way that acknowledges the specific context of their origins and subjects them to sustained interrogation and critique rather than merely turning them on their head. Central to this project has been, not surprisingly, a rethinking of Asia as a space of difference, and a rethinking of Asian history not as occupying a particular place in a (European dominated) world his tory, but as something representing more than a static past buried under the onslaught of modernity. Chakra barty puts it this way: "One result of European colonial rule in South Asia is that the intellectual traditions once unbroken and alive in Sanskrit or Persian or Arabic are now only matters of historical research for most – perhaps all – modern social scientists in the region. They treat these traditions as truly dead, as history." European traditions, by contrast, are never dead in quite the same way.

While the intellectual project of postcolonialism thus raises questions about the universal presumptions of an epistemology which conceived Asia as a mirror for European conceptions of the world, on a political level, postcolonialism represents, in the words of Kishore Mahbubani, "an effort [by Asians] to define their own personal, social, and national identities that enhances their sense of self esteem in a world in which their immediate ancestors had subconsciously accepted the fact that they were lesser beings in the Western universe." It is in this context that the political discourse of 'Asian values' or 'the Asian way' may be viewed. In March 1993 an Asian Summit was held in Bangkok to determine specifically Asian approaches to human rights. It was concluded that there was no single definition of human rights either within Asia or the wider world. This reflected a broader postcolonial project of reconfiguring universal values as local, fragmented, and particular. This discourse marks an effort to not simply appropriate and invert Eurocentric thinking about Asia, but to re imagine Asia as a viable alternative to the West. Such an effort has accompanied the rising economic power of Asian states as well as the eclipsing of Cold War geopolitics by globalization as the dominant reference point for asserting the existence of Asia as a coherent space of some kind.

Asia as Political Ideology and Populist Movement

"My Asiatic blood has always called to other Asiatics. In my dreams, both sleeping and waking, I have heard the voice of Asia calling to her children… . Today…I seem to hear…the same voice of Asia gathering her children together. It is the call of our Asiatic blood. This is not the time to think with our minds; this is the time to think with our blood… ." Thus claimed Burmese nationalist Ba Maw in a moment of Asiatic reverie at the Greater East Asian Conference of 1943. While it is not surprising that nationalism should thrive by articulating a distinctive national cultural heritage and identity as a legitimizing narrative for the nation's existence, Asian nationalisms have often added to this a transnational narrative of Asian commonality. One might assume from this that the legitimacy of independent Asian nations is wrapped up in the existence of a broader ideology of Asian coherence. Indeed, as Wang Hui argues, both socialism and Japanese colonialism have provided the context for just such a coherence around which Asian nationalisms have emerged. It was, after all, under the umbrella of Japanese colonialism that Ba Maw made his remarks. But the impulse to tie national identity to a broader discourse of 'Asianness' now sees its clearest articulation in terms of an Asian way of modernity, and Asian values as alternative to the universal pretensions of Anglo American modernity.

Asian Values and the Dream of an AU

The idea of an Asian way of modernity, based on distinctive Asian cultural and philosophical traditions, was promoted by leaders such as Lee Kwan Yew of Singapore and Mahathir Mohamed of Malaysia. Perhaps its most strident articulation has been under the guise of 'Confucian capitalism' which posits 'Confucian values', kinship, and native place ties as generating an alternative center for global capitalism. Yet, Confucian capitalism, like the broader discourse of Asian values is in fact highly limited to a relatively small region within Asia, and expresses more of a cultural invention than an economic reality. The more significant issue is less the divergence of the Asian values discourse from reality and more the fact that Asian values emerge as the product of a dialog across a space not limited to Asia. The Asian way of modernity reflects a global conversation about modernity, rather than any supposed cultural essence bounded within Asia itself.

It is significant to note, in this context, the obvious fact of intense intellectual and ideological traffic between Asia and the West since the nineteenth century. At various moments in history Western intellectuals have enjoyed enormous popularity in Asia. The celebrity status enjoyed by reformers like Samuel Smiles in Japan, or artists like George Bernard Shaw in China are but two examples. Indeed, Marxism itself is probably the most profound example of this traffic. Given this situation, then, it seems that Asian values, like 'Western values' before them, should be regarded as the product of a long term global dialog. And such a view is perhaps all the more necessary given the polemics of some, such as Anwar Ibrahim, Malaysia's former Deputy Prime Minister, who has argued that, "it is altogether shameful, if ingenious, to cite Asian values as an excuse for autocratic practices and denial of basic rights and civil liberties." At the same time, Anwar has appealed to the nebulous formation of 'Islamic values' to combat both despotism and 'Westernization' in Asia.

The 1997 financial crisis brought about a renewed questioning of the idea of an 'Asian way' of modernity. While the crisis offered an opportunity for Western critics to rail against the idea of Asian exceptionalism, it also spurred efforts in East and Southeast Asia to move toward the founding of an Asian Union on par with the EU. Japan's early proposal for an Asian Monetary Fund was resisted (perhaps in remembrance of Japan's im perialist history in the region), but in 1998 an 'East Asian Vision Group' met in Hanoi to propose steps toward greater Asian cooperation in all spheres. Such cooper ation could include a common market, a common cur rency, and common citizenship.

Conclusion: Beyond Orientalism

It remains to be seen, perhaps, whether a continent wide idea of Asia will be increasingly shadowed by more manageable regionalisms based in south, southeast, east, or even southwest Asia. Most movements developing some degree of cultural, political, and/or ideological coherence in the name of Asia have in fact been re gionally limited. And so it may be that Asia in practice means little more than a disjuncture of regionalisms. At the very least, however, it can be claimed that Asia represent far more than the 'other' of the West. It is most accurately described as an ongoing project of identity fueled by engagement with global patterns of politics, culture, economics, and governance.

See also: Cold War; East/West; Europe; Historical-Geographical Materialism; Indian Ocean; Middle East and North Africa; Nationalism, Historical Geography of; Nationalism; Orientalism; Other/Otherness; Pacific Rim; Peasant Agriculture; Postcolonialism/Postcolonial Geographies; Region.

Further Reading

Andaya, B. W. (2006). Oceans unbounded: Transversing Asia across 'area studies'. *Journal of Asian Studies* 65, 669 690.

Chakrabarty, D. (2000). *Provincializing Europe*. Princeton: Princeton University Press.

Embree, A. and Gluck, C. (eds.) (1997). *Asia in western and world history: A guide for teaching*. Armonk, NY: M.E. Sharpe.

Fabian, J. (1983). *Time and the other: How anthropology makes its object*. New York: Columbia University Press.

Fogel, J. (1988). The debates over the Asiatic mode of production in Soviet Russia, China, and Japan. *American Historical Review* 93, 56 79.

Lewis, M. and Wigen, K. (1997). *The myth of continents*. Berkeley: University of California Press.

Mark, E. (2006). 'Asia's' transwar lineage: Nationalism, Marxism, and 'Greater Asia' in an Indonesian inflection. *Journal of Asian Studies* 65, 461 493.

Meyer, M. W. (1997). *Asia: A concise history*. Lanham, MD: Rowman & Littlefield.

Myrdal, G. (1968). *Asian drama: An inquiry into the poverty of nations*. New York: Pantheon.

Said, E. (1978). *Orientalism*. New York: Pantheon.

Shrestha, N., Lewis, M., Cohen, S. and McDonald, M. (2002). State of Asian geography in America, 1988 2000. In Gaile, G. & Wilmot, C. (eds.) *Geography in America at the turn of the century*, pp 625 649. London: Oxford University Press.

Wang, H. (2005). Reclaiming Asia from the West: Rethinking global history. *Japan Focus* (23 February), http://www.japanfocus.org/products/details/1781 (accessed on Nov 2007).

Relevant Websites

http://people.uleth.ca/~wei.xu/AAG China/cgsg.htm
AAG China Geography Specialty Group, Personal Web Site Index.
http://www.uky.edu/AS/Geography/AGSG/welcome.htm
AAG Asia Geography Specialty Group, University of Kentucky.
http://www.aasianst.org
Association for Asian Studies.
http://www.asiasource.org
Asia Source.
http://www.iias.nl
International Institute for Asian Studies.
http://coombs.anu.edu.au/asia www monitor.html
The Asian Studies WWW Monitor, The Australian National University.
http://www.eastwestcenter.org
University of Hawaii East/West Center.

Atlases

J. Siemer, University of Regina, Regina, SK, Canada
H. Asche, University of Potsdam, Potsdam, Germany

Glossary

Atlas Collection of maps with uniform formal structure and design.

Atlas Information System Digital collection of maps with formal structure and design on specific region or topic with interactive features.

Electronic Atlas Digital atlas, usually available on electronic medium such as CD ROM or DVD as well as online.

National Atlas Complex regional atlas referring to a national state.

Physical Map Map depicting the physical environment of an area.

Regional Atlas Atlas portraying a specific geographically or culturally defined area and its physical and human environment.

Thematic Atlas Collection of thematic maps depicting the spatial distribution of one or more attributes of the physical or human environment other than topography of a specific geographically or culturally defined area.

Thematic Map Map depicting the spatial distribution of one or more thematic attributes of the physical or human environment other than topography.

Introduction

Every systematic application specific collection of maps with uniform formal structure and design is termed an atlas after a book collection of world maps by the German Dutch cartographer Gerardus Mercator more than four centuries ago. Since Mercator's time the term atlas implies not only a specific map content, a system of map models based on a common narrative, but also the presentation medium, a map product in book form or paper maps assembled in a map case. The change to computer assisted map construction and production brought about increasing numbers of electronic atlases to be designed and developed for on screen use. Characteristic of any electronic presentation is the screen map allowing the user to interact with the map graphic in various ways. When linked with a geographic database, electronic atlases offer extensive interaction and manipulation features of their graphic and nongraphic components. Atlases and their maps, regardless of the presentation medium, offer functions for viewing, data exploring, and knowledge acquisition of a certain geographic area or theme. As maps are a medium of communication or language, atlases can be regarded as essays in communication.

Historical Background

The development of atlases can be divided into different stages.

The first stage of atlas development is defined by Ptolemy's Guide to Geography (Geographike hyphegesis), which dates back to the second century AD. In volume 8 of the guide Ptolemy included instructions on how to show the Earth in maps and also included 26 maps. The guide was unknown until it was translated into Arabic around 800 BC. Not until it was translated into Latin in 1409 in Italy did it have an impact on the development of cartography and atlases in particular. Later (1477) other printed versions of the Guide to Geography were available, which also included the maps. In the sixteenth century, mainly in Italy, later also in Germany and France, more map collections were published in book form, which followed the order of maps suggested by Ptolemy. The Guide to Geography can thus be seen as an early example of an atlas.

The first consistently structured atlas (Theatrum orbis terrarum) was produced by the Dutch cartographer Abraham Ortelius and published in 1570, though its maps only shared a consistent format and print technique (etching) but lacked uniform orientation, content, and visualization. The work of the Dutch German cartographer Gerardus Mercator (Atlas sive cosmographicae meditationes de fabrica mundi et fabricati figura, posthumously published in 1595) introduced the term atlas for a collection of maps in book form, which was also more consistent than earlier map collections.

New developments in the field of atlas cartography took place during the seventeenth and eighteenth century in France. Advances in the quality of atlases are mainly based on higher accuracy in land surveying techniques and tools and therefore more accurate maps.

In the nineteenth century new printing and reproduction technology facilitated the production of various types of atlases. Prominent examples are general reference atlases like the German Stieler Handatlas or the American Rand McNally Atlas.

Current developments in computer technology have led to new types of maps and atlases. Not only has the production of atlases changed due to new technologies but also the use of maps and of atlases is strongly affected. Digital technology applied in atlas production and use offers new ways to communicate spatial information.

Types of Atlases

Atlases, whether in conventional or electronic form, portray (1) a clearly defined region geographically or (2) a specific topic or theme.

General-Reference Atlases

Atlases can depict a specific area or cover the entire Earth. The latter are often referred to as general refer ence atlases and contain mainly small scale physical maps to emphasize locations of places and their spatial relationships. General reference atlases typically also contain additional noncartographic information such as tabular data, (air) photos, satellite images, explanatory texts, and, most importantly, indexes of place names and geographic objects linked to coordinates (e.g., latitude and longitude or an internal locational grid consisting of numbers and letters along the sides of the atlas maps). General reference atlases, which portray a specific area rather than the entire Earth, typically comprise large to medium scale topographic maps, depending on the size of the region shown, and are often referred to as topo graphic atlases. Two examples of general reference at lases are the Times Atlas of the World and the Atlas of the World by Oxford University Press.

Regional and National Atlases

Some atlases concentrate on a specific region, for ex ample, a specific country, a province or state, or any other geographically or culturally defined region and portray it in a more comprehensive manner. Atlases, portraying a region, province, or state rather than a specific country, are referred to as regional atlases. Other than general reference atlases, regional atlases are not limited to dis playing the physical environment and topography but document a region more broadly. In doing so topics can vary and cover all subject areas of physical and human geography. A region or area can vary from a relatively small area such as a city to a group of states or even a continent. Most regional atlases are also considered to be thematic atlases as they focus on one main theme or subject area. Examples of regional atlases among many others are the Atlas of Oregon and the Atlante Socio economico della Regione Insubrica (Socio economic Atlas of the Insubrica Region), produced in the Swiss–Italian cooperation.

Special types of regional atlases are national atlases. As complex regional atlases referring to a national state they display current thematic information on the physical and human environment, usually in a series of medium to small scale maps. At the same time, they depict the status of geographic visualization and geographic information technology employed in their production and can therefore be seen as a country's official carto graphic business card. Typically, government agencies or research institutions edit and publish national atlases, which are predominately geared toward research, and are of a problem oriented nature. National atlases can also be seen as comprehensive collections of current geographic knowledge of the countries they document. Depending on the design and structure, many national atlases also address the general public to raise awareness of a country's specific characteristics. Additionally, national atlases play more than one role: on the one hand, they provide a comprehensive summary of a country's physical and human geography by displaying current results of geographic and cartographic research and achievements; on the other hand, they are a symbol of a country's unity and political independence. Hence, national atlases have played an important role in nation building (e.g., the National Atlas of Finland was pub lished in 1899 before Finland became an independent state in 1917).

Due to the complexity and the large number of pos sible themes and related workload, national atlases are often published in a number of separate volumes, each portraying one single theme (e.g., demography, physical environment, and climate). Other national atlases are published as a series of single map sheets, which are successively assembled in a map case.

Thematic Atlases

Contrasting to general reference or topographic atlases, thematic atlases depict one selected geographic theme (or a small number of related topics). The number of themes is almost unlimited. Themes can range from geology, climatology, demography, land use, planning, etc.

A special type of thematic atlas is a school atlas. It is based on didactics, following the principle from familiar to unknown, from near to far away, rather than the re verse, more common order of maps and topics found in general reference atlases. Traditionally, school atlases focused on small scale physical maps but are increasingly complemented by thematic maps on various topics. School atlases usually are published according to official school curricula.

Thematic atlases can be produced and distributed as conventional paper version or as electronic version. Selected examples of thematic atlases are the Historical Atlas of Canada, the German Weltseuchenatlas (Atlas of Epidemic Disease), and the digital Dutch Health Atlas (**Figure 1**).

Dissemination Medium

Paper Atlases

Traditionally, atlases have been and are still published in book form or as single map sheets assembled in a map case. Such conventional atlases offer the user a clear structure. Although paper atlases allow for browsing through the different chapters, they also guide the user by their sequential setup. Atlas chapters and themes are arranged to facilitate the best possible way to com municate spatial information. Further advantages of paper atlases are the easy access (i.e., independency of computer technology) and the nearly unlimited atlas format and size, which allows for customizing maps to a certain region or area portrayed. On the other hand, paper atlases do not support advanced geovisualizations like animated maps and other multimedia features. Nowadays, many conventional paper atlases are com plemented by a digital version, which can offer those features.

Electronic Atlases

More and more atlases are nowadays produced as elec tronic versions. Whether an atlas is produced as a con ventional paper volume or an electronic medium is dependent on a number of limiting factors affecting its production and use. Prominent among those are acces sibility of potential atlas users by the presentation and distribution medium selected, production costs, finan cing, and user access to computers. Nowadays, electronic atlases are predominantly distributed on CD ROM, DVD, or are available online. All three types of electronic atlases, to a different extent, allow for a combination of various visualization formats such as static and animated maps, videos, photos, 3D models, etc., and can therefore also be referred to as multimedia atlases. Electronic at lases are not limited to a predefined order of chapters and maps like paper atlases but allow for random access of maps based on the user's needs. They also offer a wide range of digital visualizations of spatial information for on screen display and use.

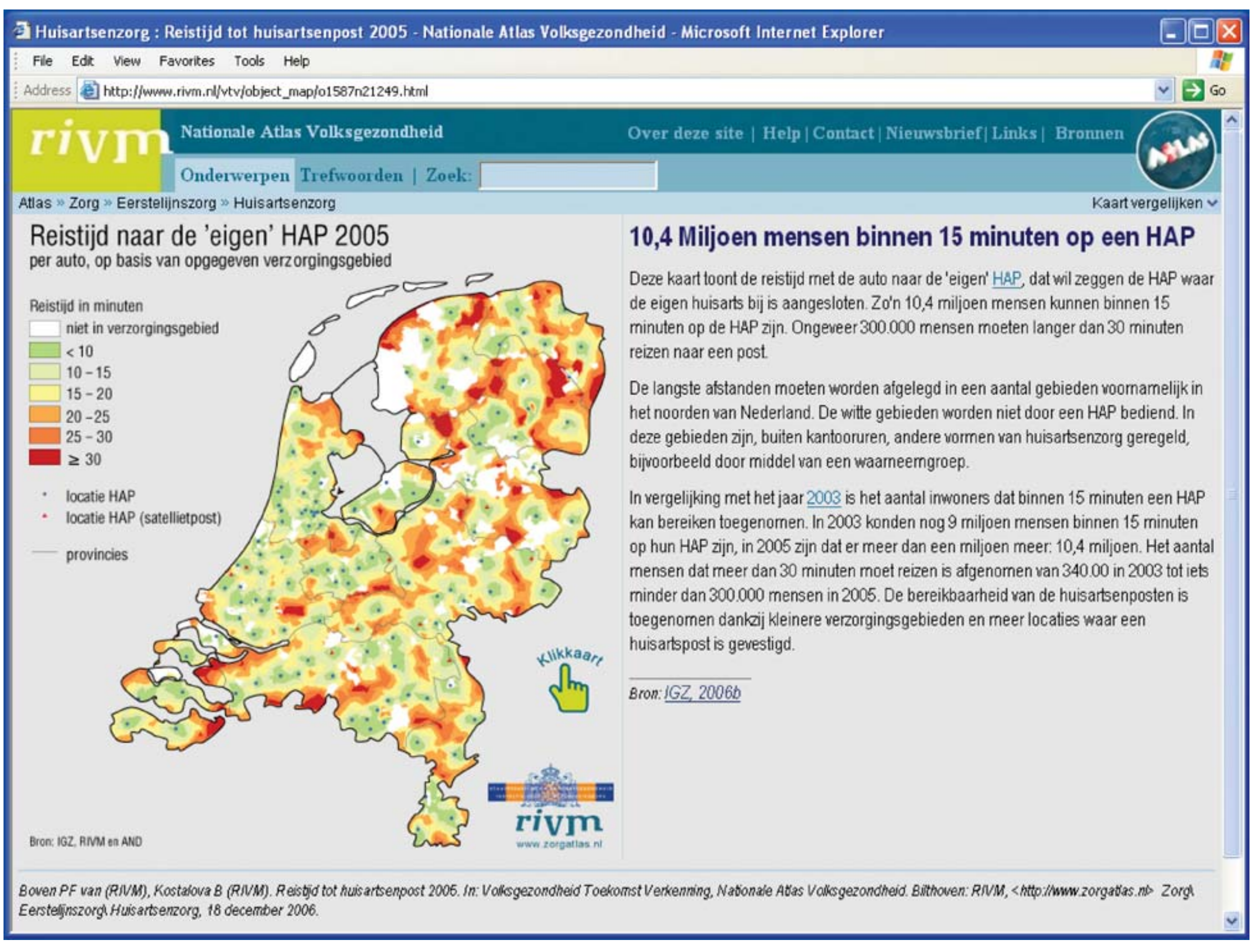

Figure 1 Example of a thematic atlas (Online Dutch Health Atlas, screenshot, www.rivm.nl/vtv/object document/o4235n21143.html). Produced with permission of Rijksinstituut voor Volksgezondhied en Milieu (RIVM).

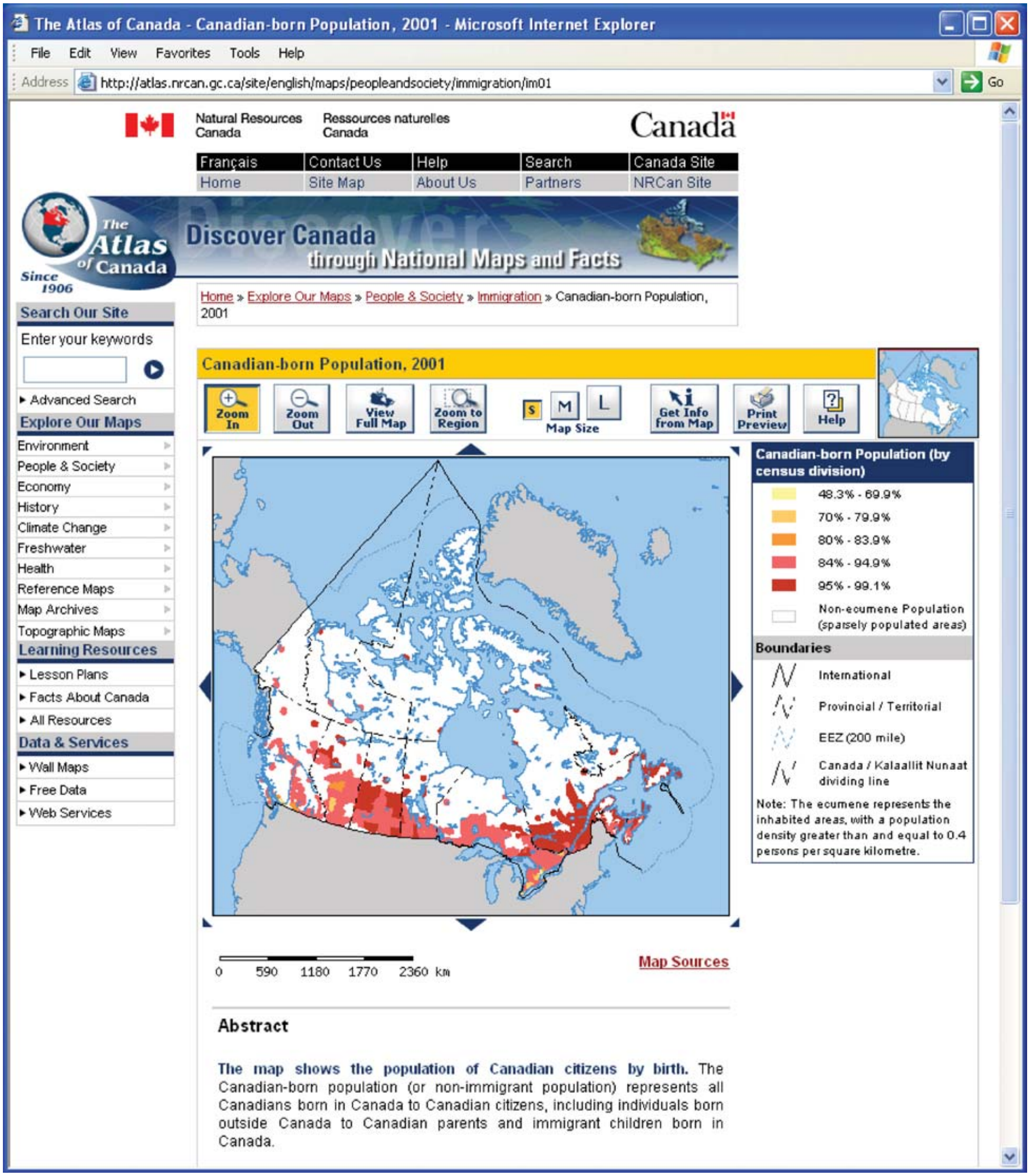

Figure 2 The National Atlas of Canada, original map data provided by The Atlas of Canada http://atlas.gc.ca/ © 2007. Produced under license from Her Majesty the Queen in Right of Canada with permission of Natural Resources Canada.

Depending on the level of interactivity, a distinction between three types of electronic atlases can be made. (1) View only atlases contain maps that can be randomly displayed but do not allow for any other interactivity than navigating through the atlas. (2) Interactive atlases allow for interactive map use by means of manipulation of the data sets contained. In an interactive atlas the user can change colors, choose different symbols, adjust classification methods, change the number of classes, etc. An example of an interactive atlas is the Atlas of Switz erland Interactive, published on CD ROM and DVD. (3) Analytical atlases allow for the highest level of interactivity facilitating additional object query functions and are not limited to predefined map topics. The user can choose themes depending on individual needs, combine different themes, and conduct computations on areas and themes. Designed to its full potential, an ana lytical atlas offers major functionalities found in geo graphic information systems (GISs).

Most interactive and analytical atlases offer the ad vantages of the following interactive features: (1) meas urements of geometrical information like distances and areas, (2) changing the map extent by moving it (panning) and scaling it (zooming), (3) searching for places, objects, and locations by means of search engines or hyperlinks, (4) production of customized maps in regard to display and visualization of themes and geographic objects, (5) use of multimedia features such as photos, videos, sounds, and animations, and (6) visualization options that allow the user to aggregate thematic data to different spatial units or to change the type of data (e.g., absolute and relative values), offering new views and possibly new insight into the topic under investigation.

In the age of digital information and communication, atlases are increasingly being produced as electronic atlases for online use, like the recent national atlases of Canada (**Figure 2**), the United States of America, or Sweden. These atlases can easily be updated and there fore can function as current information services or geographic data warehouses. These services facilitate easy access to official geographic data for the public.

Atlas Information Systems

As stated before, atlases can be considered comprehen sive spatial information systems of a particular region irrespective of their production and presentation media. They represent the complex human and physical en vironment in a narrative based collection of application specific, synoptic, and significant map and other graphics. This may include static data presentations like static maps, tables, text, and images, as well as dynamic pre sentations like animated maps, sound, video, etc. Viewed from information and communication theories, atlases thus represent highly compressed knowledge bases of all kinds of spatial information. The current term atlas in formation system (AIS) stresses the affinity of electronic or digital atlases with GISs. Like their conventional counterparts, AISs offer extensive functions for viewing, exploring, and knowledge acquisition, however, based on screen maps. AISs combine the powerful potential of nongraphic data management and manipulation like se lection, aggregation, or spatial join with a flexible pre sentation of spatial information.

See also: Cartographic Animation; Geovisualization; Map Types; Mapping, Topographic.

Further Reading

Buckley, A. (2003). Atlas mapping in the 21st century. *Cartography and Geographic Information Science* 30, 149 158.
Hurni, L., Bar, H. R. and Sieber, R. (2001). The atlas of Switzerland as an interface multimedia atlas information system. In Cartwright, W., Peterson, M. P. & Gartner, G. (eds.) *Multimedia Cartography*, pp 99 111. Heidelberg: Springer.
Keller, C. P. (1995). Visualizing digital atlas information products and the user perspectives. *Cartographic Perspectives* 20, 21 28.
MacEachren, A. M. (1994). Visualization in modern cartography: Setting the agenda. In MacEachren, A. M. & Taylor, D. R. F. (eds.) *Visualization in Modern Cartography*, pp 1 12. Oxford: Pergamon.
Morrison, J. L. (1995). A personalized national atlas of the United States. *Cartographic Perspectives* 20, 40 44.
Ormeling, F. J. (1995). New forms, concepts, and structures for European national atlases. *Cartographic Perspectives* 20, 12 20.
Ormeling, F. J. (1993). Ariadne's thread structure in multimedia atlases. In Mesenburg, P. (ed.) *Proceedings of the 16th International Cartographic Conference Cologne*, pp 1093 1100. Bielefeld: DGfK.
Slocum, T. A., McMaster, R. B., Kessler, F. C. and Howard, H. H. (2009). *Thematic Cartography and Geovisualization* (3rd edn.). Upper Saddle River, NJ: Prentice Hall.

Relevant Websites

http://atlas.nrcan.gc.ca
Natural Resources Canada. National Atlas of Canada.
http://www.sna.se
National Land Survey of Sweden. National Atlas of Sweden.
http://www.zorgatlas.nl
Rijksinstituut voor Volksgezondheid en Milieu. Nationale Atlas Volksgezondheid.
http://www.nationalatlas.gov
United States Geological Survey. National Atlas of the United States of America.

Australasia

C. Gibson, University of Wollongong, Wollongong, NSW, Australia

Glossary

Colonialism It is a policy or intention to invade, possess, and control overseas territories.

Ecoregion This is a way of describing and classifying world regions by ecological characteristics. An ecoregion usually consists of land or water areas containing a geographically distinct assemblage of natural communities and characteristic combinations of soil and landform.

Etymology The practice of tracing and historically verifying the origins of words and their meanings.

Geographical Imagination Images and ideas about places; human perceptions of places and their characteristics formed at various scales (from the body, the local, and the regional to the national and global).

Wallace Line A cartographic line separating the two ecoregions of Australasia and Indomalaya.

Introduction

Australasia is a geographic place name for a large region of the southern part of the world. As to which countries are considered a part of Australasia has not been universally defined. Different countries are included depending on the use of the term, and throughout its history, Australasia has often been used interchangeably (and usually no more accurately) with others such as Oceania, the Antipodes, the South Pacific, the Sealing Islands, the Southern Continent, the South Seas, and even the colloquial 'down under'. In its most simple contemporary use, it is a generalist term used to describe just Australia and New Zealand, but in times past and in modern scientific disciplines, such as botany, it has had more specific meanings and definitions, and has been taken to include parts of Indonesia, Papua New Guinea, and other Pacific Islands. In New Zealand, the term is unpopular and considered redundant. But in business, science, and administration, it still lingers. In this manner, the use of 'Australasia' provides a classic example of how geographical place names – even those with long histories covering vast parts of the Earth's surface – are fluid, imprecise, and contested. Since its origin over 250 years ago, the idea of Australasia has had a life of its own, coming to mean different things to different people. From French scholarly origins, 'Australasia' has become a phrase bound up in colonial imaginations of distant lands; a label considered by some to be benign, to others offensive; and a moniker surviving today as a word of convenience in science and bureaucracy.

Colonial Origins and Imaginations

Australasia is an invented name, the product of European desires to scientifically describe and classify the world's geography. The name Australasia is the English translation of *Australasie*, first coined by Charles de Brosses, an aristocratic French magistrate and politician (he became president of the Burgundy parliament), who collected, synthesized, and translated, into French, accounts of all the voyages made to the South Seas by Spanish, Portuguese, Dutch, and British mariners. His work was originally published as the two volume *Histoire des navigations aux terres australes* in 1756. The first English version appeared in 1766.

De Brosses derived *Australasie* from the Latin *austrālis* (from *auster*, *austr* , meaning south) to describe, literally, everything 'south of Asia'. He saw *Australasie* as distinct from *Polynésie* ('many islands', referring to those of the Pacific Ocean between *Australasie* and America) and *Magellanie* (the territories south of the line running from the Strait of Magellan in Tierra del Fuego to the Cape of Good Hope; see **Figure 1**). Included in de Brosses' original description of *Australasie* were the islands of the Indian Ocean, New Guinea, New Britain, Espiritu Santo (a land mass imagined on maps as physically attached to today's Queensland, but actually an island of what would become Vanuatu), Van Diemen's Land (later to be called Tasmania), New Holland (later to be called Australia), and New Zealand. He argued that the Pacific was divided by three submerged mountain ranges – one running north–south down the east coast of Asia, the other two spanning the Pacific between Asia and America – whose peaks created the thousands of islands and thus sociocultural distinctions between world subregions. Australasia was one of these. The term fell into regular subsequent use by geographers, politicians, colonists, and bureaucrats – especially in its English translation, Australasia – although, as is seen in this article, not without variation or contestation.

Recognition of de Brosses' contribution to world geography went far beyond authoring a place name. Indeed, his synthesis of explorers' writings indelibly shaped the way Europeans interpreted and engaged with the South. Although his work (and others by his contemporaries such as Rousseau and Buffon) was driven by

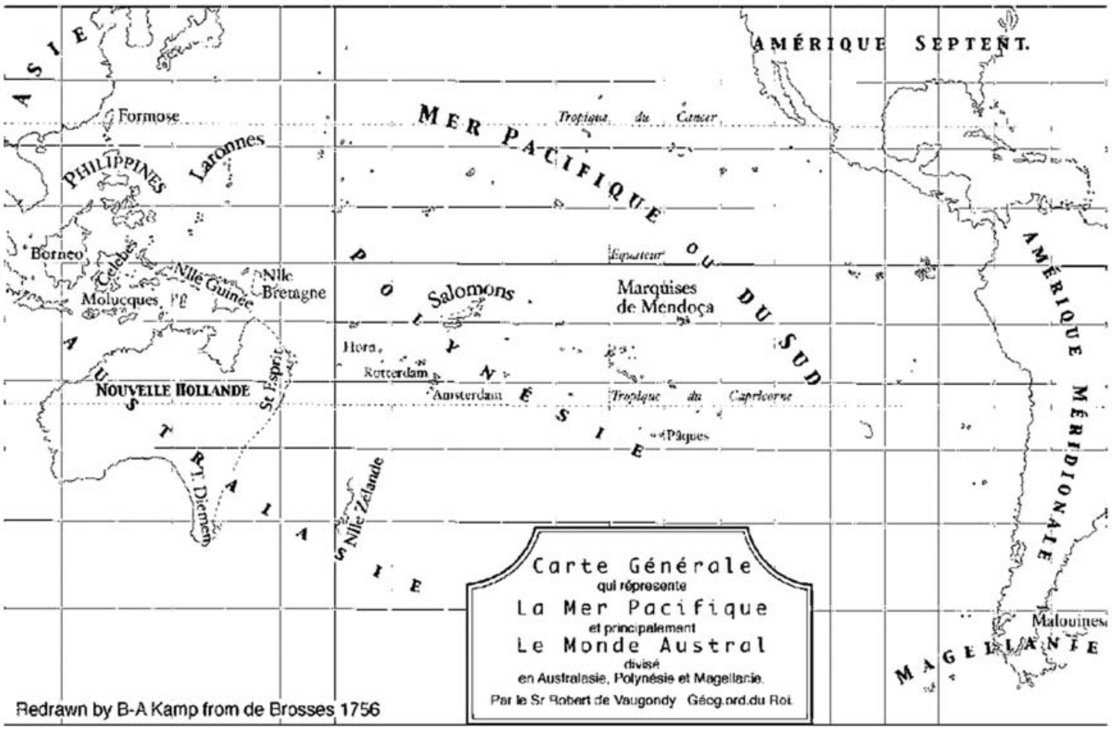

Figure 1 *Australasie*, according to de Brosses, 1756. Source: Ryan, T. (2002). Le Président des Terres Australes: Charles de Brosses and the French Enlightenment beginnings of Oceanic anthropology. *Journal of Pacific History* 37, 157 186.

a desire to map and scientifically classify the world's geography and humanity, in hindsight, it reveals much more about the context of European imaginings of other world regions, and the superiority and power Europeans assumed for themselves. The naming of Australasia was not just an act of benign description. It semantically distinguished and labeled a world region that had already, for centuries, been presumed to be distinct, disconnected, backward, and capable (if not 'in need') of invasion and civilization. This was at a time when "the Pacific became something of a laboratory for the [European] testing of scientific methodologies" (Livingstone, 2000: 246). Even before they were fully traversed and mapped by Europeans, the Antipodes were considered an inversion of northern civility, nature, and norms, a "place of perversity" (Ryan, 1996: 105). Writing at around the same time as de Brosses, French astronomer and mathematician Maupertuis (regarded by some as the first to use the term 'race' in describing peoples of the world) exemplified this European imagination of *Australasie* as disconnected and different:

> Everyone knows that in the Southern Hemisphere there is an unknown space where there may be situated a new part of the world greater than any of the other four… nowhere else on the globe is there a space as vast as this, but rather than being totally occupied by a continuous sea, there is much more probability that one will find there land… It would be difficult to make unfounded conjectures on the productions and on the inhabitants of these lands… but one suspects that one will find there things very different from those one finds in the four other parts of the world. (1752, trans. in Ryan, 2002: 170)

The mid 1700s was a time of revolutions in science and industry, and both England and France sought to expand their empires. The lure was an imagined great southern continent – *terres australes* – believed to span the globe beneath the continents of Africa, America, and Asia. Conjecture was rife about the places and societies of the South, with stories of untold riches, potentially vital strategic geographical sites, as well as "innumerable speculations on giants, Patagonians, savages, natives, and monsters supposedly residing to the far east, west, south, and north of Europe" (Said, 1978: 117). Such decidedly fantastical predictions grew and were transformed throughout the Enlightenment period. Amid this conjecture, de Brosses' text sought to amalgamate otherwise scattered and fragmented accounts from prior explorers, and more accurately categorize places, peoples, landscapes, and biota, redefining *terres australes* into three sections, in terms of underlying physical geography. Behind de Brosses' writing was a view to the capacity of these places to be explored; of the natives to be pacified and civilized; and of the suitability of land for coloniza tion by Europeans. The naming of Australasia thus, in many respects, signaled a pivotal moment in the history of European colonialism, as speculation about the little known South culminated in an act of scientific synthesis and classification, with accompanying proclamations of intent to systematically dominate new territories.

De Brosses also wrote about the peoples of the South Seas in the context of a wider debate about whether there was one united human race (with shared genetics), or instead a series of 'races', 'varieties', or 'kinds', of which Europeans were considered superior (indeed, a different, more evolved species). *Australasie* became the place name for a largely unknown southern part of the globe, onto which a series of European fantasies were projected about their (equally unknown) human inhabitants. For instance, Maupertuis argued that in contrast to other world re gions, such as Europe, Asia, Africa, and America – where people and animals had always been in communication and thus shared many traits – in *Australasie* distinct spe cies of all life forms, including a kind of 'man ape', may yet be discovered:

> It is certain that they [the islands of Australasia] are absolutely isolated, and that they form, so to speak, a world apart, so that one is not able to anticipate what will be found there… it is in these islands in this sea that voyagers assure us they have seen wild men, hairy men, with tails; a species midway between monkeys and our selves. (trans. in Ryan, 2002: 171)

Australasie was not just a convenient place name for a little known part of the world. Instead, the act of naming *Australasie* was a mechanism for formalizing, in the European imagination, a distinct sociobiological region. It was remote, disconnected, and savage – the archetypal product of the European Orientalist gaze. In this way, *Australasie* as a category represented a spatial manifest ation of European ideas about race, evolution, and superiority.

The translation of *Australasie* into the English *Austra lasia* occurred in 1766 when John Callander, a minor writer, translated de Brosses' original *Histoire des Navi gations* into English without crediting the original author (de Brosses happened to leave his name off the cover of his opus). Claiming the extensive work as his own, Callander translated the full text (renaming it *Terra Australis Cognita*), including place names, into English. Thus *Polynésie* became Polynesia and *Australasie* became Australasia. Maps were stolen and redesigned with An glicized nomenclature, and "associated ideas of diverse 'savage' racial groups and societal types coexisting in the Pacific, formally entered English language and thought" (Ryan, 2002: 180). Far from being a noble exchange of scientific ideas, the wholesale theft of de Brosses' work was representative of the manner in which colonial ideas, concepts, maps, and names were hotly contested and claimed between France and England. The two rivals fought wars in the middle decades of the 1700s, and competed for territorial expansion in North America and in the Pacific. In the intervening years between the publication of de Brosses' original and Callender's copy, France and England fought the Seven Years War and England had deprived France of its colonies in Canada and India. De Brosses text was written to inform a French colonial expedition in the midst of this rivalry. But cru cially, the translation of new geographical knowledges into English provided information to support the English crown's invasions of that part of the world.

Thus began the British colonial use of Australasia, to define the parts of the South Pacific it would come to dominate through colonial invasion. Captain James Cook's voyage of the 1760s and 1770s shortly followed, during which Aotearoa (as New Zealand was known to Māori) was surveyed; Fiji and the New Hebrides (Vanuatu) visited; New Caledonia named; and New Holland claimed. Of little surprise, in light of the pre colonial fantasies described above, native inhabitants were variously described as 'degraded' or 'savages', oc cupying a place 'at the edge of humanity' (Anderson, 2007). Australasian botany was unusual and was described as inverted, twisted, and abnormal. This way of imagining the continent's indigenous peoples and plant and animal life had already been framed prior to colonization by a European geographical classification of Australasia as a 'world apart', inhabited by man apes and bizarre creatures.

Research, exploration, and writings of British colonial geographers maintained the fiction of Australasia – but also further confused it. In 1804, Matthew Flinders cir cumnavigated New Holland, and used the term Australia on resulting maps, thus ushering in the formal use of that term to describe just the mainland of one part of Australasia. His naming of Australia built on centuries old uses of variants of that word in the writings of European explorers and geographers, to refer to the mythic 'great southern land'. Flinders did not invent 'Australia', but he did pin the name formally to the one land mass. It eroded claims that the Dutch once had to the continent (as New Holland), but was also a direct acknowledgement of the importance of de Brosses' work in mapping and describing the Antipodes. In addition, the naming of Australia created a semantic slippage with the word 'Australasia' – such that one part of the world re gion shared part of its name (indeed, 'Austral', the dominant part) with an individual country. This would later confuse the meaning of 'Australasia' and complicate its usage. 'Australia and New Zealand' was often used instead as a *de facto* hyphenated place name (as with

'Trinidad and Tobago'), because it encompassed all the 'seven colonies' of Britain in the region (none of which had yet been federated into unified nation states). It has even been said that in the late 1800s, New Zealanders were comfortable with 'Australia' as a shortened place name encompassing all the region – including their own colony – rather than just one landmass, reflecting the loose geopolitical categories of the time. Only adding to this confusion, a few years after Flinders' voyage, John Pinkerton in *Modern Geography*, in 1807, defined Australasia, Polynesia, and the 'Asiatic Islands' as the primary geographical divisions in the Pacific, and at around the same time, another English geographer, Conrad Malte Brun, coined the terms 'Océanique'/'Oceanica' (later to morph into 'Oceania') to describe the unknown 'fifth part of the world' (as de Brosses described it). These alternatives would make even more fluid the naming of this part of the world and lay the foundations for later misunderstandings.

Although beaten by the British in colonizing much of the Pacific, the French continued to invest in the scientific quest to further and more accurately describe and categorize world regions, as precursor or companion to actual colonial annexation. Major colonial voyages (often disguised in the name of science) were led to *Australasie* by Louis Antoine de Bougainville (1766–69) and Jean François de Galaup, the Comte de la Pérouse (1785–88). On them, France claimed numerous Pacific possessions including *Nouvelle Cythere* (Tahiti) and Vanuatu (New Hebrides); indeed, many islands were claimed on Bougainville's journey from the bow of their passing ship, without him or his crew having to disembark. In contrast, on his voyage, De Galaup had the ignominious misfortune of arriving at Botany Bay, Sydney, a mere 5 days after Captain Arthur Phillip had established Britain's first permanent colony in New South Wales, and thus just missed out on the opportunity to claim Australia for France. Departing rather despondently after experiencing what one commentator has since called 'a fluke of cosmic proportions', de Galaup headed for New Caledonia, but never arrived.

Anglo French rivalries in Australasia continued into the nineteenth century, both in the development of scientific geography, and as competing powers extending their tentacles of empire. Grégoire Louis Domeny de Rienzi presented his paper to the *Société de géographie* in Paris in 1831, in which he further classified *Océanie* (Oceania) into five geographical regions and four principal 'races'. At the same time, (and to the same audience) Jules Sébastien César Dumont d'Urville argued that the region was made up of four geographical divisions – *Polynésie*, *Micronésie*, *Malaisie*, and *Melanésie* (including Australia) – and two key racial divisions, one belonging to the first three geographical regions, the other to the last. Explicit in these classifications were hierarchies of culture, as peoples were organized into fixed racial types, positioned along continuums of civilization, and partitioned by geographical space. Further French expansion soon followed; the Marquesas were annexed in 1843 and 10 years later *Nouvelle Calédonie* (New Caledonia) was made a penal colony in an attempt by Napoleon III to rival the British colonies in Australia and New Zealand.

Britain too reconfigured its role in the Pacific in the mid 1800s. No longer penal encampments or remote geopolitical outposts, the large Australasian colonies became important contributors of raw materials to a newly industrializing Britain, while the smaller island states became largely ignored. The textile mills, heavy industry, and shipbuilding yards of rapidly growing towns like Liverpool, Leeds, and Bradford all required vast quantities of inputs transported from Australia and New Zealand. Pastoral and mining activities were expanded (or more accurately, were 'allowed' to expand in a largely unregulated and invariably damaging manner), and Britain poured capital into the building of capital cities, ports, and railways. By contrast, British colonial authority over its smaller island territories was held onto rather reluctantly; investment was rare, and Australia and New Zealand assumed much of the colonial responsibility over them. New power relations and patterns of uneven development within the region were being established. In Australia and New Zealand, populations grew, cities were built, and this new form of industrial capitalism delivered the fruits of Antipodean colonialism long promised as part of the myth of *Terra Australis*. Accompanying this was the persistence of a more deeply exploitative racist undercurrent. Indigenous peoples continued to be denigrated, presumed to be subhuman, less advanced, or in the process of becoming extinct, and thus their sovereign rights were downplayed or ignored. Throughout the 1800s, wars were fought by soldiers, 'settler' farmers, and civilian populated militia against most of the approximately 250 Aboriginal nations on the Australian and Tasmanian land masses; and British descendant Australians and New Zealanders lined up together to fight Māori in New Zealand. Widespread dispossession of Australian Aboriginal peoples and Māori accompanied the re imagining of Antipodean spaces as productive suppliers of raw materials for the colonial centers. At the same time, neighboring islands were often excluded from flows of capital and commerce, and faded from the center of the British colonial imagination in Australasia.

Australasia as Ecoregion

The apex of the use of Australasia as a place name stemmed from its acceptance as a scientific term to describe a unique ecoregion. This emphasis on ecological uniqueness was, along with theories about racial difference, a central concern of Europeans in their at tempts to distinguish and classify world regions. Early evidence of a distinction between Asia and lands to the south was noted by Antonio Pigafetta, who identified biological contrasts between the Philippines islands to the west during the voyage of Ferdinand Magellan in 1521. Much later, more formal ecological boundaries of 'Australasia' were mapped, as writers and explorers sought to describe the unique flora and fauna in the re gion, differentiated from that of Asia. Most significant was Alfred Russel Wallace, who, in the 1860s, observed key differences in bird distributions between the islands of Bali and Lombok in the Indonesian archipelago. Birds observed on Bali were clearly related to those observed on islands to the west (such as Java and Sumatra), while those on Lombok were related to species observed in New Guinea and Australia.

Wallace charted a dividing line between these islands and, in doing so, buttressed European imaginations of Australasia as distinct from Asia. Now known as the 'Wallace line' in his honor (and subsequently augmented and re drawn by other biogeographers observing transi tions in distributions of flora and fauna among those is lands), this cartographic device established the basis of a supposedly fundamental division in world regions that survives into contemporary scientific discourse (**Figure 2**). The Australasian ecoregion is thus now considered distinct, with its own common evolutionary history and ancestry. It includes Australia, New Guinea/ West Papua, Timor, the Indonesian islands of Sulawesi, Lombok, Sumbawa, Sumba and Flores, Vanuatu, the Solomon Islands, and New Caledonia. New Zealand and its surrounding islands are considered a distinct sub region, while the rest of Indonesia is part of the Indo malayan ecoregion.

Australasia has since been continuously used as a re gional category in botany and taxonomy – hence, the currency of the name in recent scholarly articles, such as 'Two newly described polypores from Australasia and southern South America' and 'A new family of *Muscoid Diptera* from Australasia, with sixteen new species in four new genera'. It has become a largely uncontested term in science linked to a commonly accepted understanding of a unique world region.

A Contemporary Australasia?

If the link between colonial imaginations of Australasia as a distinct world region can be seen to neatly translate onto contemporary scientific classification and nomen clature, the story is not nearly so simple in the spheres of politics and society.

Most simply, there is no clear wisdom on what countries are contained in the geopolitical entity known as Australasia, and whether it – rather than other groupings such as Oceania – is the most accurate moniker for the 'sixth' continent (after Asia, Africa, Europe, North and South America). Many textbooks and maps depict Australia alone as a continent, leaving other parts of the Pacific continent less. Others deploy Australasia or Oceania as 'catchall' phrases to refer to

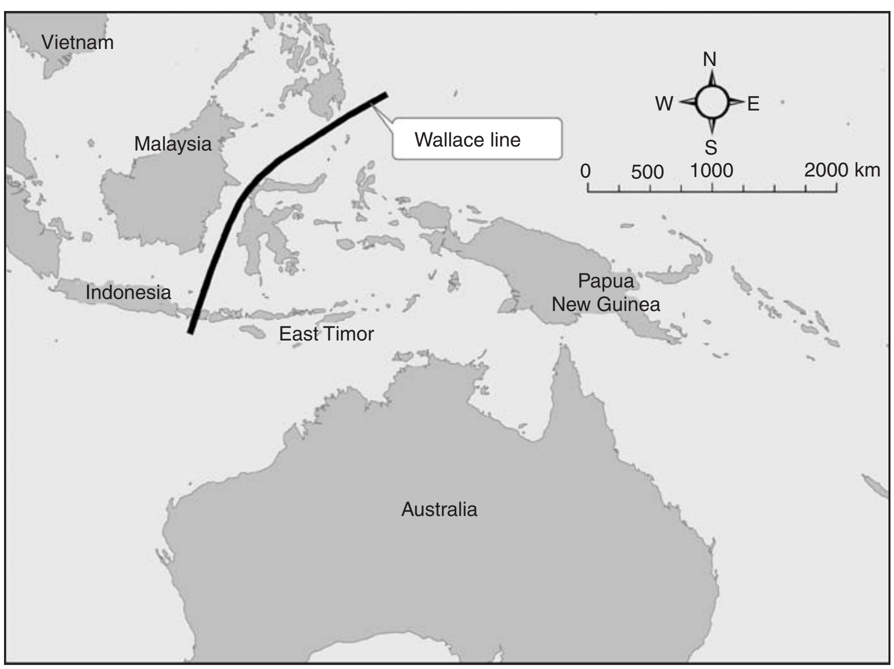

Figure 2 The Wallace line an ecological definition of Australasia.

Australia, New Zealand, and other landmasses of the Pacific not subsumed as part of other continents. This fluidity reflects changing meanings over the past 250 years.

What was known as *Australasie* in de Brosses' age took on another, more specific meaning in the late 1700s and early 1800s, as its more expansive definition ('south of Asia') contracted in ordinary parlance to become 'the seven British colonies of Australasia', meaning New South Wales ("the Mother Colony of Australia," as described in one school textbook from the early twentieth century), Victoria, Queensland, South Australia, Western Australia, Tasmania, and New Zealand (**Figure 3**). This was the dominant meaning for 'Australasia' up until the 1901 Federation of Australian colonies (from which the contemporary Commonwealth of Australia derives). Indeed, up until it became a separate colony in 1840, New Zealand was administered as part of New South Wales, and even after this time it was considered part of a single British presence in the Antipodes (see **Figure 4**). Thus, New Zealand was included in conventions held in the 1890s in which federation of all these colonies was discussed. New Zealand bowed out of such discussions and became an autonomous country. There are numerous theories as to why: New Zealanders held fears at the time of being ignored or exploited if governed from the Australian mainland at such a distance; there was also growing nationalism and a sense of superiority over the Australian colonies (a result of the self perception of racial purity, stemming from the lack of convicts in New Zealand), as well as questions about the benefits of international, rather than inter colony trade.

After Australia and New Zealand became separate sovereign nation states, the term Australasia largely fell from use within those countries, especially in New Zealand. Separate trajectories of nationalism and development had emerged and new kinds of geopolitical distinctions and logics were set in train. The Royal Geographical Society of Australasia, established as an offshoot of the Royal Society of New South Wales – and which at its height sponsored expeditions to Antarctica and New Guinea – fragmented into branches for each of the distinct colonies. Separate national geographical organizations were established for Australia and New Zealand (the Institute of Australian Geographers and the New Zealand Geographical Society), and a very different set of goals and intellectual endeavors emerged, driven instead by new metropolitan geography departments, and national and state based curricula. School textbooks taught distinct and unique histories and geographies, and rivalries grew in sport, academia, and the arts (indeed, people promptly forgot that New Zealand and Australia fielded combined teams at the Olympic Games of 1908 and 1912, under the banner 'Australasia'). The newly imagined separation was so effective that some scholars have since argued that 'Australasia' – abandoned as a meaningful geopolitical category in the twentieth century – became little more than a moniker for a 'repressed' shared memory, as New Zealand and Australia each subsequently sought to emphasize distinct national cultures.

Hence, in *The Geography of New Zealand: Historical, Physical, Political and Commercial* (*c.* 1904) by P. Marshall, 'Australasia' did not earn an index entry. Its historical

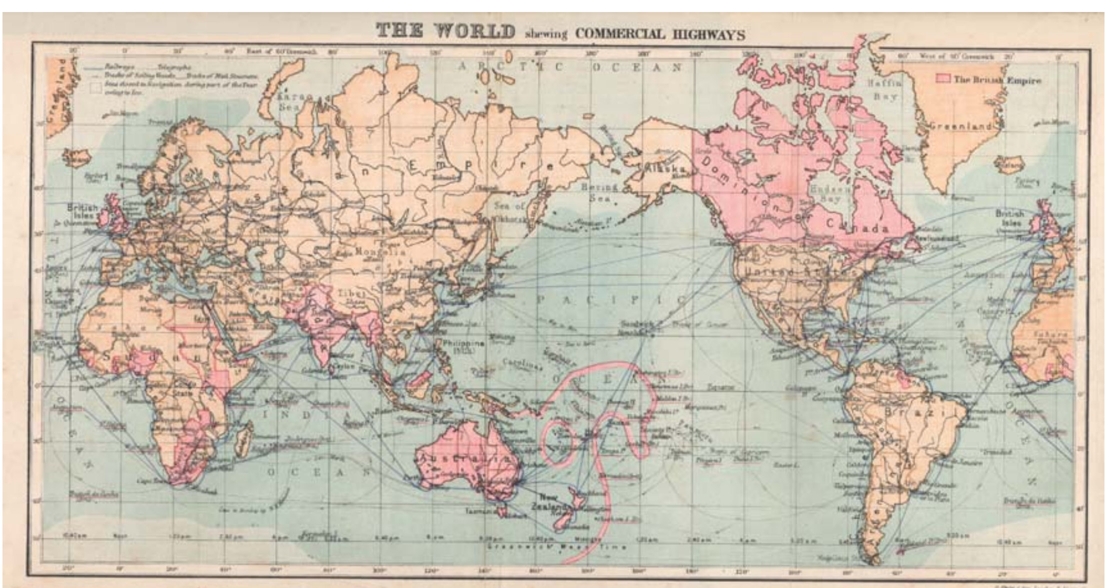

Figure 3 Map of the British Empire, c. 1900. It demonstrates the interpretation of Australasia as 'the seven British colonies' (with French colonies excluded by re-drawing regional boundaries).

Figure 4 Advertisement, Bank of New South Wales, 1935. It plays on the lack of general public knowledge even at the time of its publication of the common colonial history of Australia and New Zealand.

section described New Zealand's 'discovery' and colonization separate from that of Australia; and stressed the physical geographical uniqueness of New Zealand: "New Zealand is a continental island that has been isolated from neighbouring continents for an enormous lapse of time. In consequence of this its flora and fauna are widely different from those of any other land in existence" (P. Marshall, 1904: 247); and "two thirds of the indigenous plants are absolutely confined to New Zealand in their distribution" (P. Marshall, 1904: 255).

Indeed, the term Australasia has somewhat become taboo in contemporary New Zealand, because it conflates it with Australia and renders it invisible in the wider regional descriptor. Subsequent differences in the interpretation of the term are easily understood by comparing dictionary entries from New Zealand with those in

Britain and Australia. The British *Oxford Dictionary of Current English* describes Australasia as "a region consisting of Australia and islands of the SW Pacific." Its sister publication, the *Australian Oxford Dictionary* describes it as "Australia, New Zealand, New Guinea, and neighbouring islands of Pacific." In contrast – and in an example of both affirmative action and linguistic resistance – the *Oxford Dictionary of New Zealand English* defines Australasia as "New Zealand with Australia and their outlying islands… often felt as a patronizing term and replaced by locutions as Australia and New Zealand, New Zealand Australia(n), Anzac, etc." In its further description of the etymology of 'Australasia', the *Oxford Dictionary of New Zealand English* cites examples of its use as a word New Zealanders "dislike intensely," because it "implies a close connection between New Zealand and her bullying barbaric neighbour;" one that had been "banned for many years – informally… but by common consent – in the in New Zealand press."

Yet as a descriptive term for this part of the world, the term Australasia has lingered on in different ways. In Britain, a long way from Australia and New Zealand and their respective nationalisms, the term has not seemed to lose its usefulness as a general descriptor and was continually employed throughout the twentieth century without regard for its redundancy in the South. Administrative matters were often still conducted with regard to 'Australasia', and records of trade, exports, and imports through the colonial network still referred to 'British Australasia' well into the twentieth century. Colonial imaginings of Australasian peoples and natures as perverse or despoiled also lingered; thus school textbooks such as the *Class Book of Modern Geography* described Australian Aboriginal tribes as "among the most degraded members of the human race. A few roots and berries, with shell fish, insects, grubs, and other repulsive objects, form the food resources of the Australian savage, who will eat almost anything – lizards, snakes, and frogs being especially esteemed, while most of the wilder tribes are also cannibals, not from necessity, but from choice… they are gradually dying out, although it probably will be a long time before they are entirely extinct" (p. 396). British ignorance of local circumstances on the other side of the world combined with persistent myths of Australasia as inversion of European civilization.

In Australia in the 1990s the name Australasia took on another meaning in the context of an increasing 'push into Asia'. This was spearheaded by then prime minister Paul Keating, who courted Asian leaders, encouraged closer military and economic ties with key states (notably Japan, Singapore, Thailand, and Indonesia), and altered immigration laws in ways that enabled more extensive flows of new settlers and overseas students from Asia. Australasia had crept back into partial use as a means to describe a more fluid world region of exchanges and interactions between an ex British, postcolonial Australia and modern, rapidly developing 'Asian Tiger' nations. Although the subsequent conservative government reversed the focus on Asia when elected in 1996 (ushering in a more inward looking, even xenophobic, perspective), many Australians continued to think that 'Australasia' meant 'Australia as a part of Asia'. Indeed, the *Australian Oxford Dictionary* still wrongly cites the origin of Australasia as a conjunction of "Australia + Asia," reflecting this misnomer.

Mostly, though, Australasia is a word that people in the Pacific know of, but which has fallen from their regular use. In formal arenas, it has largely been replaced by Oceania, including in international diplomacy and sport (as in *Fédération Internationale de Football Association* (FIFA)'s football confederations); and debates about whether Australia by itself is or is not a continent have largely subsided, remaining unresolved. For many residents in those countries that have at one time or another been officially considered part of Australasia, it matters little whether there is an overarching name for a continent to which they belong – or what it might be. People in this part of the world do have a sense of belonging to a South Pacific region including large and small countries, culturally diverse groups, and thousands of islands. But this sense of regionalism is rather loose and contradictory. The Australian and New Zealand governments have, at different times, intervened heavily in the geopolitical affairs of region, sending peacekeeping forces, police staff, and diplomats in times of crisis, ethnic tensions, and political instability (as happened recently in East Timor, the Solomon Islands, and Bougainville). John Howard, the Australian prime minister until 2007, for instance, described Australasia as "our patch, we have a special responsibility in this part of the world, I believe very strongly and this is our region" (quoted in Connell, 2006: 111). But by and large, Australians (and perhaps less so Kiwis) do not hold the same feelings of belonging to a coherent larger continent in the way that Africans or Europeans might. Fijians or Samoans might hold allegiance to pan Pacific island identities, but probably would not to an Australasian identity. Kanaky from New Caledonia have more of an allegiance to other Melanesian nations than to English speaking Australasia (or indeed to France), and other tribal identities remain much stronger than wider regional ones.

Where it is still used, Australasia has become a simple referent for just Australia and New Zealand. It is now most common as an adjunct to the names of hundreds of nongovernment and representative organizations who happen to have offices and activities in the two countries, whether scholarly (Australasian Quaternary Association, Australasian International Studies Association, etc.), leisurely (Professional Divers Association of Australasia), mundane (Plywood Association of Australasia), or bizarre (Biggles Association of Australasia). Uneven contemporary uses of ‘Australasia’ are thus matters of convenience and applicability. But they also reveal traces of a colonial past, and a shifting geographical im agination influenced by increasing internationalization and interconnectivity.

See also: Biodiversity Mapping; Colonialism I; Colonialism II; Empire; Exploration; Mapping, Race and Ethnicity.

Further Reading

Anderson, K. (2005). Australia and the ‘state of nature/native’. *Australian Geographer* 36, 267 282.

Anderson, K. (2007). *Race and the Crisis of Humanism*. New York: Routledge.

Belich, J. (2001). *Paradise Reforged: A History of the New Zealanders from the 1880s to the Year 2000*. Auckland: Penguin.

Chalkley, B. and Winchester, H. (1991). Australia in transition: Cultural diversification and geopolitical reorientation. *Geography* 76, 97 108.

Connell, J. (2006). ‘Saving the solomons’: A new geopolitics in the ‘arc of instability’? *Geographical Research* 44, 111 122.

Dalby, S. (1996). Continent adrift? Dissident security discourse and the Australian geopolitical imagination. *Australian Journal of International Affairs* 50, 59 75.

De Brosses, C. (1756). *Histoire des navigations aux terrres australes: Contenant ce que l’on sçait des mœurs et des productions des contrées découvertes jusqu’á ce jour: & où il est traité de l’utilité d’y faire de plus amples découvertes, & des moyens d’y former un établissement*. Paris: Durand.

Dixon, C. and Drakakis Smith, D. (1995). The Pacific Asian region: Myth or reality? *Geografiska Annaler* 77B, 75 91.

Healey, J. (2002). The origins of the name ‘Australia’. *South Australian Geographical Journal* 101, 94 109.

Livingstone, D. (2000). Exploration. In Johnston, R. J., Gregory, D., Pratt, G. & Watts, M. (eds.) *The Dictionary of Human Geography*, pp 245 249. Oxford: Blackwell.

MacLeod, R. and Rehbock, P. H. (eds.) (2004). *Darwin’s Laboratory: Evolutionary Theory and Natural History in the Pacific*. Honolulu: University of Hawaii Press.

Mayell, P. (2004). Beyond the ‘outer crescent’: The Mackinder century in New Zealand geopolitics. *The Geographical Journal* 170, 368 376.

Pinkerton, J. (1807). *Modern Geography: A Description of the Empires, States, and Colonies; with the Oceans, Seas, and Isles; in all Parts of the World*. London: Cadell and Davies.

Ryan, S. (1996). *The Cartographic Eye: How Explorers Saw Australia*. Melbourne: Cambridge University Press.

Ryan, T. (2002). ‘Le Président des Terres Australes’: Charles de Brosses and the French enlightenment beginnings of oceanic anthropology. *Journal of Pacific History* 37, 157 186.

Said, E. (1978). *Orientalism*. New York: Pantheon.

Smith, P. M. and Hempenstall, P. (2003). Australia and New Zealand: Turning shared pasts into a shared history. *History Compass* 31, 1 8.

Relevant Websites

http://www.ugm.ac.id/eng
Gadja Mada University, Yogyakarta, Indonesia home to the largest geography faculty in the region if not the world.

http://www.gsnsw.org.au
Geographical Society of New South Wales.

http://www.iag.org.au
Institute of Australian Geographers.

http://www.nzgs.co.nz
New Zealand Geographical Society.

http://www.rgsq.gil.com.au
Royal Geographical Society of Queensland.

http://www.rgssa.org.au
Royal Geographical Society of South Australia.

http://www.geo.fio.usp.ac.fj
School of Geography, University of the South Pacific.

Autobiography

M. Purcell, University of Washington, Seattle, WA, USA

Glossary

Autobiography A method of inquiry in which the researcher narrates his or her own life.
Overdetermination The idea that a given effect has multiple causes that overlap in intricate ways.
Phenomenology Tradition in geography that values what we can learn from studying subjective experience. It takes seriously not just reason, but also emotion, intuition, and perception.
Positivism The argument that inquiry through the scientific method is the best way to know the world.
Reflexivity A tradition in geography that values critical reflection on the researcher's role in producing academic knowledge.

Introduction

As I understand it, the idea of an article like this is for the editors to choose an expert on a topic, and ask that expert to call upon his or her vast existing knowledge to com municate what we currently know about the topic. I am not an expert on the use of autobiography in geography, or I do not think of myself that way. However, I have been thinking seriously about it recently. I wrote an auto biography, or really an account of my 4 years in pro fessional limbo, between getting my PhD and finally landing a tenure track job. The article was published in *Antipode*, and I argued that we need to begin a serious critical examination of the issue of non tenure track faculty in geography. One thing such an examination will reveal, I suggested, is that when we look at power re lations in the academy from the perspective of the marginalized, a whole host of assumptions that justify those power relations – and therefore prop them up – will crumble. There is something very valuable, I claimed, about seeing the world and writing about it from a particular perspective. Writing the article was at once the most frightening, cathartic, and rewarding experience of my career. It was some of the best academic work I have done. It also elicited, by far, the largest response I have ever received. Some few (who made an appearance in the narrative) were angry, but the overwhelming majority were extremely appreciative. The article really affected people, in a variety of ways. As the responses rolled in, I came to understand that autobiography has great power. I also appreciated even more fully the wealth of ethical and intellectual debates swirling around the form. Because autobiography is so powerful, we must take those debates very seriously. In writing the *Antipode* article, engaging the response, and in preparing this article, I spent much energy grappling with those debates. In what follows, I report on what I have learned.

Four Forms

In trying to pinpoint exactly what is an autobiography, it is worth trying to understand its relationship to similar forms of inquiry. Biography, ethnography, and auto ethnography are perhaps the closest cognates. On the surface, there are differences clear enough one might even be tempted to put them in a matrix (see **Table 1**). It appears that autobiography is what a person writes about one's own life. In biography a person's life is written by another. Ethnography involves a researcher writing about a group of people, while in autoethnography the group writes about itself. All forms share the 'graph' root: they seek to 'write' some human experience, to represent it to an audience. In addition, for most practitioners, such 'graphing' does not imagine it is producing a scientific, objective understanding; rather it accepts that represen tations are necessarily subjective and involve complex relations between representer and represented.

Those complex relations are also what blurs the lines among the four forms. The differences among them rest on the assumption that there are clear distinctions to be made between self and other, and between individual and group. However, various currents of recent thought – both modernist and postmodernist – have muddied such distinctions. Derrida, for example, in his notion of the constitutive outside, suggests that one's identity is necessarily constituted in part by its outside, by what it is not. The lines between self and other, between insider and outsider, are therefore difficult to draw neatly, since they are necessarily imbricated in each other. Similarly, both liberal and postliberal political philosophy have accepted that each individual is significantly composed of the social relations in which one is embedded. It is not

Table 1 Relationship between autobiography and biography, ethnography, and authoethnography

	Individual	*Group*
Self-authored	Autobiography	Autoethnography
Other-authored	Biography	Ethnography

so simple, therefore, to distinguish clearly between individual and society. An autobiography, for example, ostensibly aims at representing an individual. But if the individual is necessarily constituted in part by one's social context, then writing the life of an individual is always also, in part, writing the life of one's society. For autobiography, therefore, the goal is never to elucidate 'only' the self, since no self exists in isolation. To make matters more complex, yet another set of relations are inherent in the act of writing: those between author and audience, mediated by the text. As a result of those critical destabilizations, most writers in all four traditions have come to accept that they confront a complex and mutable set of interrelationships between self and other, individual and society, and author and audience.

The issue of power is yet another element of the mix. Recent work in political philosophy has stressed that all social relations are imbued with relations of power. The relations between self and other, individual and society, and author and audience that characterize all four traditions are therefore also shot through with relations of power. Any project to represent a social reality is inescapably shaped by (and an intervention in) relations of power. Autoethnography, for example, is in many respects an attempt by colonized peoples to reclaim from the colonizer the authority to represent their culture. It is both a reaction to existing power relations and a conscious attempt to change them.

Autobiography in Geography

The place of autobiography in geography is currently in flux. On the one hand, it is fair to say that "autobiographical approaches are not widely accepted as a standard method for geographic research" (Moss, 2001a: 190). In part, that lack of acceptance can be attributed to a lingering positivism that devalues autobiography because it does not produce 'objective' data that can be counted. However, even the rising respect for qualitative methods in geography has not yet led to a concomitant recognition of autobiography. *The Dictionary of Human Geography*, for example, has entries for ethnography, interviewing, participant observation, psychoanalysis, and focus groups, but does not have an entry for autobiography (or biography). Despite that lack of acceptance, however, others have noted that very recently autobiography is becoming increasingly common in the discipline. While that new work probably does not yet add up to the 'autobiographical turn' that Daniels and Nash proclaim, it may be that the current spate of autobiographical work is beginning to undermine the traditional disrespect the form has received.

Autobiography as Data

In taking stock of autobiography in geography, I want to distinguish the two ways it is used. Autobiography as data is when geographers use an autobiography written by someone else as data for their research project. One of the most well known examples of that approach is the use of the journals of Victorian women travelers. Here the draw of autobiography is that it is told from a particular perspective. It articulates how women in that time saw the world and British colonialism. That perspective was systematically excluded from public discourse in Victorian Britain, and so autobiography provides researchers with information they would not otherwise have. Of course, in this case, recapturing women's perspectives also contributes to a feminist project of interrogating historically unequal power relations between men and women.

Another reason geographers have used autobiography as data is because it offers information that is emotional, experiential, and subjective. For example, migration studies tend to be highly quantitative, reducing each person's complex history to a set of commensurate data. So Findlay and Li used what they called an 'auto biographical approach' to better understand why migrants choose to move. Though they called their approach autobiography, they actually used interviews to prompt subjects to tell some of their life story. The quasi autobiographical stories they gathered allowed Findlay and Li to understand migration decisions in a way they could not have using traditional quantitative data. In that respect, autobiography is not much different from other qualitative methods: geographers use it to collect rich information about people's experience, emotions, and everyday life that quantitative techniques do not handle well.

Autobiography as Method

Autobiography can also be used as a method. That is, a researcher can narrate elements of his or her own life in order to achieve a particular intellectual goal. There are many ways that project can be pursued. In order to make some sense of that diversity I ariculate four different approaches. Those four are not intended to be exhaustive, rather, I hope they begin to suggest some of the different intellectual projects that autobiography can serve. It is important to stress that there is much overlap among the various forms. The differences are not stark breaks but variations in the degree of emphasis each places on particular intellectual goals.

The 'great man'

The most traditional approach grows out of the European Enlightenment and comes to us through St. Augustine and Rousseau. It is much more willing than the other three to accept that individuals are self contained and relatively autonomous. It portrays individuals mak ing key choices and fashioning themselves through those choices. The narratives tend to be comprehensive; that is, they do not focus on one aspect of the life (e.g., one's sexuality) but aim at a full account of an individual's life. That is because the goal of the approach is more than anything to make sense of the life itself: not as a case that can shed light on larger trends, but as worth under standing on its own terms. Therefore the subject of such autobiographies has usually attained a measure of social status, some import that would cause readers to want to understand the story behind the great personage (who is usually a man). Bill Clinton's *My Life* is an archetypical example. In geography, the pieces in Peter Gould and Forrest Pitts' *Geographical Voices* are good illustrations, as is Gould's *Becoming a Geographer.* The reader's interest is assumed to be in the figure himself, and so the narratives aim to help us understand the paths that led to the preeminence of each figure (all 14 of whom are men in *Geographical Voices*). In such accounts, one can often pick up an element of apologia – of self vindication – as when David Harvey's excellent essay in Gould and Pitts' book narrates, with some spleen, his failed attempts to publish, in *Society and Space,* a response to postmodern criticisms of his work. There is also an element of 'life and times' in this form. To an extent, the narratives are intended to shed light on the wider history of the discipline. But in the main, the goal of this kind of autobiography is what might be called "enshrining'. It writes about a life in order to codify an individual's position in a social pantheon.

Phenomenology

Another approach, what we might call phenomeno logical, is interested in the individual not as a great man of history but as someone with rich human experience. Here the aim is to better understand everyday life, and so the 'ordinary' person is generally more prized as subject than the extraordinary one. The approach is typified by Yi Fu Tuan's *Who Am I ?*, the goal of which is to sift through the artifacts of everyday experience to better understand the emotion, mind, and spirit of one life. "Just as no human life is negligible," writes Tuan, "so no human life story is negligible, not worth telling." The subjective nature of autobiography is particularly im portant here, because it is the only way to properly perceive and convey what it is like to feel, to experience, and to live. For phenomenologists, the attempt to understand what it means to be human, to use thick description to convey the rich details of a human life, is the most important kind of academic inquiry. The choice of self as subject is therefore often driven by access: an author's own life is the one she or he knows most in timately, from which the richest detail can be drawn. In Tuan's case, one might justifiably suspect that there is an element of enshrinement and/or apologia at play here, given that he is nearing the end of his career and has seen his approach to geography become unfashion able. But one searches mostly in vain for self justification. What comes through much more clearly is the earnest desire to take seriously and to capture a rich human experience.

Reflexivity

A third approach is heir to the post structuralist critiques of the past several decades. Those critiques have stressed the importance of reflexivity, whereby researchers crit ically examine their own subject position. Advocates of reflexivity reject the idea that knowledge is preexisting, to be collected and reported by the researcher. Rather knowledge is taken to be actively 'produced' by the re searcher (in collaboration with many others). Whenever research is reactive in some way (i.e., surveys, eth nography, interviews, and participant observation), the researcher enters into complex relations with those being researched. Those relations play a central role in pro ducing the knowledge the research creates. Therefore, we must critically examine the multiple social and political relations that influence the production of academic knowledge. That imperative prompts researchers to write what might be called critical mini autobiographies de signed to explore their positionality and their relations with their research subjects. They do not aim at a com prehensive account of a life, since that life is not the focus of the inquiry (unlike in the 'great man' approach). Ra ther the autobiography is what could be called targeted: it presents an autobiographical account designed primarily to examine reflexively the production of academic knowledge. In this approach, autobiography is therefore a buttressing strategy. It is taken seriously, but it is not the main focus of the research, which is usually some form of ethnography.

However, this work stresses that such distinctions are tricky. The researcher in many ways becomes part of the community, in the sense of having entered into a web of relationships with those she or he is studying. The more embedded the researcher is, the greater the degree of incorporation will be, and the more the autobiography will bleed into ethnography. It will become harder to make distinctions between self and other, between insider and outsider. As a result, this approach undermines the idea of the self contained, autonomous individual. Rather it sees individuals as overdetermined by multiple social factors. A good example of the approach is the work of David Butz. During his intensive and extended fieldwork in a remote village in Pakistan, it became clear to him that his presence was significantly changing how the village saw and engaged the world. They actively worked to incorporate him and his research project into their social structure (albeit in limited and strategic ways). Moreover, his experience in the village changed him, particularly in how he wrote his findings. He increasingly became a self conscious advocate rather than a neutral observer. As Butz' life became increasingly intertwined with the history and culture of the village, distinguishing his autobiography from the village's ethnography became increasingly messy.

This movement toward reflexive mini autobiographies is by far the most popular of the autobiographical ap proaches presented here. Reflecting critically on one's subject position has become almost compulsory in writ ing up qualitative research in geography. Any claims about an 'autobiographical turn' in geography, therefore, can be traced to the waxing of this approach.

One subelement of the move toward reflexivity has been attempts at a more critical version of the 'great man' approach. An increasing number of scholars have begun to reflect critically on how they came to hold the intel lectual convictions they do. They explore the life ex periences that have led them to adopt a particular intellectual approach (e.g., queer geography). Unlike the more traditional approach, the goal here is not to en shrine the self in the pantheon, although some measure of that does inevitably happen; rather it is more fully to critically question how academic ideas come to be ac cepted. It suggests that scholars come to champion in tellectual arguments not because of the crystalline truth of those arguments, but because scholars' embodied ex periences resonate with the values those arguments propound. The approach therefore continually under mines all received wisdom – for example, of Marxism, of feminism, and of queer theory – by suggesting that the ascendancy of a particular idea is due to the contingent ways it intersects with the life histories of practicing academics. Rather than a positivist academia in which scholars set out to progressively uncover truth through the scientific method, this approach to autobiography posits a political struggle. It sees academia as an ongoing contest among multiple arguments, and those arguments are always necessarily subjective, embodied, and partial. That argument is of course part of the wider (originally feminist) claim that the personal is political, that one's subjective experience should be made absolutely central to public and academic discourse.

Insurgent autobiography

That political understanding of the personal is best ex pressed in a fourth tradition, what I call insurgent autobiography. It is an emerging trend in autobiography that shares much with conventions of *testimonio* and autoethnography. The idea of insurgent autobiography, as the name implies, is to use autobiography to actively destabilize ruling assumptions. Autobiography is valued as a way to give voice to the voiceless, to articulate the perspective of the marginalized, the dispossessed, and the oppressed. That perspective is commonly devalued, ig nored, or silenced by a dominant culture. In Guatemala, for example, the *testimonio* method is a way that victims of systematic violence can tell their story of rape, torture, and mass killings, a story that is aggressively suppressed by those in power. Voicing the perspective from the margins can subvert the dominant logics that help mar ginalize subordinate populations. In that vein, auto ethnography is very much akin to insurgent autobiography. It is an attempt by the colonized to re claim control of how we understand colonized cultures, the legacy of colonial rule, and current development politics. One goal of insurgent autobiography, therefore, is to help liberate oppressed people.

Insurgent autobiography has also been used by a few geographers in order to examine social relations in the academy. Mona Domosh narrates how dominant mas culine norms in the academy served to devalue her interpersonal style and limit her job opportunities. Laura Pulido talks about the enormous emotional im pact of being a Latina in a dominantly white discipline. She uses her story to make that impact apparent to a white discipline that cannot see it. My own auto biographical piece pursued a similar agenda from the perspective of non tenure track faculty, and Ian Cook did so from the perspective of a graduate student. That work on the academy, taking its cue especially from feminist work of the 1980s and queer theory from the 1990s, often emphasizes the details of everyday experi ence in order to make clear the destructive impact structures of inequality have on the lives of marginalized people.

As the connection to autoethnography implies, there is often an element of collective representation involved in insurgent autobiography, especially in *testimonio*. That is, the author often claims that she or he speaks not only for oneself, but also on behalf of the whole community of which she or he is a part. That collectivization is not so much hubris, nor is it shoddy scholarship. *Testimonio* re fuses the assumption that autobiography must be indi vidual and cannot be collective. It insists rather that individual stories are necessarily imbricated in and in separable from collective histories. That is not to say they are identical, but that the lines between autobiography, ethnography, and autoethnography are impossible to draw clearly. The lines are always transgressed – both unintentionally and creatively – in the writing process. *Testimonio* presents the collective experience of op pression, and it therefore aims at liberation not only for a person, but for a whole group.

Alternatively, insurgent autobiography can be more willing to accept the convention of bounded individuals. My autobiography claimed to speak authoritatively only about my own experience as a non tenure track faculty member, and to make no claims about the experience of others. I hoped my story would make it easier for others to tell their stories, and in the aggregate, many individual stories would make up a wider narrative about the pathology of academic hierarchy. However, at the same time, because it is only one of a few such stories, to some extent it will likely be read as representative, as standing in for a whole category of people.

Another goal, complementary to liberation, concerns healing. Just the act of writing your story, of publicly revealing the violence, pain, or humiliation you have endured, can, for many, be therapeutic. It can prompt a catharsis that lifts a heavy psychological burden. Moreover, it can begin a process of healing through solidarity. That is, it often reveals connections among seemingly isolated life stories. Another's autobiography can seem eerily similar to your own, and you can begin to understand that your experience of marginalization or oppression is shared by others. When you are system atically devalued by a dominant culture, knowing you are not alone can help begin the process of reclaiming your self esteem, of healing the wounds your position has inflicted. Of course such connections can be the start of a more functional political solidarity, whereby people with similar autobiographies realize that they share a common subjugation, and a common agenda for liber ation. For all those reasons, insurgent autobiography can be a powerful strategy for marginalized populations, both to cope and to remake the conditions of their disadvantage.

Conclusion

Certainly my heuristic categorization separates out what are interlaced approaches to autobiography. All four valorize a subjective method of academic inquiry, one that values emotion, situated knowledge, perception, and even, as Tuan has it, spirit. Autobiography is therefore firmly part of the rising respect for qualitative methods in geography. Moreover, all four are concerned, to different degrees, not with the 'auto' alone but with the social context in which that subject is embedded, and more specific connections are evident as well. Phenomenology shares with insurgent autobiography a concern for everyday experience. Reflexivity and the 'great man' approach both strive to understand how an individual came to be who she or he is. So it is important to re member these approaches are deeply intertwined, even as we parse out their various differences.

Perhaps the most important point to stress in closing is that autobiography remains very much emergent in geography. Despite the increasing acceptance of situated knowledge, and despite the obvious quality and success of Pamela Moss' edited volume on the form, the place of autobiography is not secure in the discipline. The desire of the great man approach to enshrine rugged indi viduals, while consonant with current popular culture, cannot flourish in an academy that largely rejects its assumptions about the self contained individual. While feminist and queer geographers have produced some compelling work that critically updates the 'great man' tradition, only a few have been published so far. The phenomenological tradition of autobiography seems to be fading, although critical geographers who em phasize the importance of everyday life may well return to phenomenological autobiography in the future. But while the reflexive turn has brought self examination to the fore, that is not quite the same thing as the rich detail and quiet humanity of a traditional phenomeno logical autobiography like Tuan's. Lastly, insurgent autobiography is really just getting off the ground in academic geography. Since it is written from the per spective of the marginalized, and since academics doing research are generally in positions of power, in surgent autobiography by academics would usually mean accounts by marginalized populations inside the academy. Following Domosh, Pulido, Cook, and Purcell, insurgent autobiographies could provoke an extended and thoroughly critical examination of the social hierarchies of the academy. But even though such an examination is desperately needed, it is never easy for the less powerful to write (and publish) their stories.

It is therefore critical to consciously advocate for the fledgling form of autobiography. If it is to mature and become an accepted qualitative method, we must con tinue advocating, and writing, more autobiography in academic geography.

See also: Auto-Photography; Autoethnography; Ethnography.

Further Reading

Butz, D. (2001). Autobiography, autoethnography, and intersubjectivity. In Moss, P. (ed.) *Placing Autobiography in Geography*, pp 149 166. Syracuse, NY: Syracuse University Press.
Cook, I. (2001). You want to be careful you don't end up like Ian. He's all over the place. In Moss, P. (ed.) *Placing Auto biography in Geography*, pp 99 120. Syracuse, NY: Syracuse University Press.
Daniels, S. and Nash, C. (2004). Lifepaths: Geography and biography. *Journal of Historical Geography* 30, 449 458.
Domosh, M. (2000). Unintentional transgressions and other reflections on the job search process. *Professional Geographer* 52(4), 703 708.
Findlay, A. and Li, F. (1997). An auto biographical approach to understanding migration: The case of Hong Kong Residents. *Area* 29(1), 34 44.
Hanlon, C. and Shankar, F. (2000). Gendered spaces of terror and assault: The Testimonio of Remhi and the Commission for

Historical Clarification in Guatemala. *Gender, Place, and Culture* 7(3), 265 286.
Harvey, D. (2002). Memories and desires. In Gould, P. & Pitts, F. (eds.) *Geographical Voices*, pp 149 188. Syracuse, NY: Syracuse University Press.
Moss, P. (2001). Engaging autobiography. In Moss, P. (ed.) *Placing Autobiography in Geography*, pp 188 198. Syracuse, NY: Syracuse University Press.
Moss, P. (2001). *Placing Autobiography in Geography*. Syracuse, NY: Syracuse University Press.
Pulido, L. (2002). Reflections on a white discipline. *Professional Geographer* 54(1), 42 49.
Purcell, M. (2007). Skilled, cheap, and desperate: Non tenure track faculty and the delusion of meritocracy. *Antipode* 39(1), 121 143.
Tuan, Y. (1999). *Who Am I?* Madison, WI: University of Wisconsin Press.

Autoethnography

K. Besio, University of Hawaii at Hilo, Hilo, HI, USA

Glossary

Transculturation This term refers to the ways that subjugated groups may creatively adopt and transform cultural forms from dominant groups. The term is distinguishable from acculturation, deculturation and cross-culturation.

Introduction

Autoethnography, like ethnography, is both method and text. That is, it is a textual record of people and places, and way of doing qualitative research that relies on participant observation or using the ethnographic method of embedding the researcher in settings in order to write about social groups unlike those of the re searcher. In contrast to ethnography, autoethnography is distinguishable by its explicitly self referential mode of writing, whereby the divide between participant and observer is blurred and even ignored, and the text pro duced refers explicitly to the author. In autoethnography, there are two ways of conceptualizing the ethnographer/author: as an ethnographer writing autobiographically, or an ethnographic subject writing ethnographically about her or his culture. These two authorial positions com plicate researchers' understanding of the term and in the application of it as a method. This article examines these applications and complex yet intertwined understandings of autoethnography.

The two ways that geographers understand auto ethnography are discrete but closely interlinked. First, autoethnography is a type of ethnographic representation that is a blend of autobiographical and ethnographical data. In this use, the 'auto' refers to ethnographer, who incorporates biographical data. Second, autoethnography is a text, however, loosely defined, which is produced by an author who has more often been the subject of ethnographic writing. For Western trained ethnographers, this form of autoethnographic text is one that they wish to study, and thus ethnographers may engage in those autoethnographic moments by cultivating an 'autoethno graphic sensibility'. An autoethnographic sensibility relies upon the recognition by the ethnographer that a subject is representing him or herself using the idiom of eth nography, whether textual or performative, perhaps to engage dialogically with those who would represent them. These two modes are usually considered separate from one another and build upon the previous understandings of autoethnography within anthropology and sociology, where ethnography and autoethnography have been used more widely. While it is useful to consider each as dis crete modes, they increasingly inform one another, and what unites them is their insistence upon a rethinking how a socially constructed self produces knowledge about others.

One potential of autoethnography is how it might contribute to a more nuanced understanding of trans cultural interactions in postcolonial research settings specifically, although not exclusively. Transcultural interactions are the material and discursive manifest ations of interactions between groups of people positioned in asymmetrical relations of domination and subordination. In that the research process may be conceptualized as dialog between researcher and researcher which becomes a textual representation of that dialog, their interactions are also necessarily con textualized by a myriad power relations that must be carefully considered. It is the potential of auto ethnography to work toward a more fully articulated postcolonial research method in human geography that considers these transcultural relations of power, which makes it an especially significant method for geographers' consideration, a point that the conclusion of this article takes up in more detail.

Autobiography and Ethnography

In the first and perhaps most intuitively accessible understanding to researchers, autoethnography refers to a type of ethnographic writing that seeks to combine elements of 'autobiography' with 'ethnography'. The 'auto' in autoethnography, the 'self', is the researcher, most often a Western trained researcher, who draws upon their own biographies to foreground social relations. However, there are also 'native' ethnographers who work from within their cultural frameworks, who draw upon this type of biographical and culturally contextualized writing. Common to both non native and the so called native researchers is an understanding that autoethnography is a reflexive writing strategy, whereby the researcher/author positions her or himself prominently within the text in a way that is self consciously author/researcher centered. This form of autoethnography is directly concerned with representing the researcher and her or his situatedness within the research process or, alternately, in situations related to the research setting.

Sociologist Carolyn Ellis is one of the most often cited autoethnographers. She describes autoethnography as "an autobiographical genre of writing and research that displays multiple layers of consciousness, connecting the personal to the cultural." Ellis uses autoethnography to great effect in demonstrating how to do autoethnography by drawing upon conversations and dialog, demonstrating how researchers analyze the social through the personal. In this representational mode of autoethnography, the researchers' embeddedness in the research process is part of the authors' concerns, as well as making apparent the links between an individual author and the social and cultural structures that implicitly produce the research setting.

This autobiographical form of autoethnography is receiving increased attention across the social sciences, although as anthropologist Deborah Reed Danahay states autoethnography is not new to anthropology. Anthropologists have been combining biography and ethnography for some time, using 'experimental' writing strategies to produce ethnographies. Although geographers use experimental writing strategies as well, they do not seem to do so as extensively as anthropologists, perhaps in part because of what Mike Crang calls geographer's use of more 'pluralistic methodologies' than solely ethnography. There is increased interest in combining autobiography with geography, with Pamela Moss's edited volume, *Autobiography and Geography*, a good starting point for those interested in the links between autobiography and geography writing.

Ethnography, as the hallmark method within social and cultural anthropology, occupies a more solid position within the discipline and much more attention has been paid to the 'writing of culture' than geographers have done. However, with that said, researcher centered writing is experiencing a resurgence in academic literature in qualitative methods in geography, not only in response to the social sciences 'cultural turn' late in the last century, but also is in response to feminist scholars, such as Sandra Harding's and Donna Haraway's, insistence upon situated research and the now nearly mandatory recognition of reflexive and reflective authorship. Researchers working ethnographically are now much more cautious about how they position themselves in their texts, aware of potential responses from those they research to how they are represented, and of the inherent power imbalances in the research process.

While an author centered autoethnography makes more transparent the author/ethnographer's positioning in the field through the text, it nevertheless still draws suspicion and often warranted criticism that it is too indulgently on the 'auto' side of ethnography. That is, autoethnography may overstate the ethnographer's positioning at the expense of illuminating broader social interactions and the flows and machinations of power. This remains a significant concern for autoethnographers for a number of reasons, but two important points bear further consideration. First, work that illuminates the researcher's position at the expense of cultural structures may falter in its attempts to foster understanding of wider cultural relations, what Ellis calls the movement of in and of out between the researcher's lens as she or he moves their analysis from self to other. Second, work that does not move beyond the individual researcher may further marginalize more experimental writing strategies, leaving readers increasingly weary of reflexive research and autoethnography as yet one more way that the Western subject resituates the authorial gaze toward Western produced knowledge. In feminist geography, a critique of reflexive writing strategies is mounting. Currently, there seems to be some backlash regarding such author centered representational strategies as yet another way to allow the Western researcher the most prominent position in ethnographic writing, once again marginalizing those they research.

An Autoethnographic Sensibility

One response to these criticisms may come from employing an autoethnographic sensibility, the second way human geographers might approach autoethnography. Geographers David Butz and Kathryn Besio outline what they call an autoethnographic sensibility, drawing from the work of literary critic Mary Louise Pratt. Pratt conceptualizes autoethnographic texts as those produced by indigenous others in order to represent themselves to their others – in Pratt's conceptual dyad, their colonial others – in an idiom that is not their own. Pratt's mode of autoethnography is one that understands the 'auto', the author, in autoethnography as the colonized or formerly colonized subject, who seeks to position her or himself in metropolitan produced texts. The purpose of these autoethnographic representations is to intervene dialogically in representations that would otherwise write them out. She states, "if ethnographic texts are a means by which Europeans represent themselves to their (usually subjugated) others, autoethnographic texts are those the others construct in response to or in dialogue with those metropolitan representations." In Pratt's conceptualization of autoethnography, it captures something of the dialectic between accommodation and resistance: accommodation in terms of idiom – writing ethnographically – resistance in terms of intent – writing to make one's voice heard. Pratt uses the example of a manuscript written in Quechua and Spanish by an Incan writer in 1613 that rewrites Christianity to include an indigenous worldview. What Pratt outlines as autoethnography does not include Western trained ethnographers writing as ‘autoethnographers’. For Pratt, the autoethnographer is an indigenous author whose ethnographic representation highlights the transcultural nature of knowledge, that is, a product of unequal cultural exchange in a contact zone.

Drawing upon Pratt’s substantial insights into ethnographic representation and its production, Butz and Besio outline what they call an ‘autoethnography sensibility’, in which autoethnography is something that research subjects do but that ethnographers/researchers want to understand. Certainly, one of the goals of ethnography is to understand lifeworlds, usually those dissimilar to researchers’ own, by immersing themselves in the activities of others and participating and observing their lives. In this view of ethnography, the ethnographer is a kind of scribe who merely writes down what he or she sees. In contrast, an autoethnographic sensibility would attend to ethnographic interactions as strategic, intentional, and possibly interventionist, in which ethnographic selves are always already representing themselves as selves to their Western others. Researchers working within an autoethnographic sensibility must cultivate ways to understand when their ethnographic others engage them in producing their autoethnographic texts through the ethnographers’ text. As researchers, we might attune our research to texts and those moments where research subjects use the idiom of the researcher to make a point that is more theirs than the researchers’. An autoethnographic sensibility could be better cultivated among researchers to pay more attention to moments of self representation by those engaged in the research process. It is in this understanding of an autoethnographic sensibility that much of the potential of an anticolonial research practice comes to the foreground, because it more actively supports research subjects’ political projects that researchers are interested in studying.

Like the autobiographical form of autoethnography, there are significant critiques of an autoethnographic sensibility that must be heeded. First, ethnographers rarely can represent the interests of everyone in a group or community, and their work is often a representation of ‘key informants’ views’, whose views stand in for the whole. Recognizing the agency and intentionality of information ‘given’ to the researcher as strategic has the potential to rescript researchers’ perceptions of the researcher researched dynamic. However, these autoethnographic representations are nevertheless partial or as partially true as ethnographic representations.

Second, given different spaces and places of research, some subjects have more representational tools at their disposal to produce textual autoethnographies. One of the problems of confining autoethnography to written texts, maps, and other forms of printed representations is that there is diversity of representational tools among those who have access to making these sorts of representations. Certainly not all autoethnographers produce texts but nevertheless represent their interests in ways that engage with ethnographic representation. There is room within an autoethnographic sensibility to consider the ways that research subjects may perform their ethnographic understandings in spaces outside of ‘texts’ or in those ethnographic moments, such as interviews, life stories, case studies, etc., that make up the ethnographic data. In understanding autoethnography as more performative, ethnographers with an autoethnographic sensibility might acknowledge those moments in which those who are usually represented assert and represent themselves as selves to the ethnographer, strategically inserting themselves into ethnographic texts.

Autoethnographies

A strategy of combining autoethnography that is committed to self reflexivity with an autoethnographic sensibility allows for more complex representation and analysis of the research setting. Representing the research setting as one in which knowledge is co produced, in which researchers use autoethnography of the first type along with recognizing the many moments in which ethnographic subjects disclose knowledge, may go some ways toward greater transparency and ethical ways of doing qualitative research. Additionally, this might allow researchers to position themselves within the text without losing sight of empirical and political detail. Postcolonial field research methods are still finding their footing between the text and the ground. As postcolonial research in geography has evolved from more textual based analyses to analyses of material settings, postcolonial geographers have yet to define how its methods differ from its forbearers. What constitutes postcolonial geographies is an undeniably diverse body of work, but for many geographers that draw upon its insights, it is about issues surrounding social and economic development in formerly colonized and currently occupied places. In these settings, indigenous knowledge and its contributions to the production of knowledge that was once considered ‘Western knowledge’ need to be made more clearly. Autoethnography has a potential to contribute to this process.

See also: Autobiography; Auto-Photography; Ethnography; Indigenous Geographies; Postcolonialism/Postcolonial Geographies; Situated Knowledge, Reflexivity.

Further Reading

Besio, K. (2005). Telling stories to hear autoethnography: Researching women’s lives in northern Pakistan. *Gender Place and Culture* 12(3), 317 332.

Besio, K. and Butz, D. (2004). Autoethnography: A limited endorsement. *The Professional Geographer* 56(3), 432 438.

Butz, D. (2001). Autobiography, autoethnography and intersubjectivity analyzing communication in northern Pakistan. In Moss, P. (ed.) *Placing Autobiography in Geography*. Syracuse, NY: Syracuse University Press.

Butz, D. and Besio, K. (2004). The value of ethnography for field research in transcultural settings. *The Professional Geographer* 56(3), 350 360.

Cloke, P., Crang, P. and Goodwin, M. (1999). *Introducing Human Geographies*. London: Arnold.

Crang, M. (2005). Qualitative methods: There is nothing outside the text? *Progress in Human Geography* 29(2), 225 233.

Ellis, C. and Bochner, A. (eds.) (1996). *Composing Ethnography: Alternative Forms of Qualitative Writing*. Walnut Creek, CA: Alta Mira.

Moss, P. (ed.) (2001). Writing one's life. *In Placing Autography in Geography.* Syracause, NY: Syracuse University Press.

Pratt, M. L. (1992). *Imperial Eyes: Travel Writing and Transculturation*. London: Routledge.

Pratt, M. L. (1992). Transculturation and autoethnography: Peru, 1615/1980. In Barker, F., Hulme, P. & Iversen, M. (eds.) *Colonial Discourse/Postcolonial Theory*. Manchester: Manchester University Press.

Reed Danahay, D. (ed.) (1997). Introduction. In *Auto/Ethnography: Rewriting the self and the Social,* p 1 20. Oxford. Berg.

Auto-Photography

M. E. Thomas, The Ohio State University, Columbus, OH, USA

Glossary

Ethnography The study of culture and social groups based largely on field research methods.

Photo Elicitation The use of photographs to evoke verbal discussion.

Politics of Representation The debate about who can represent whom, who has authority to speak about and for whom, and what power relations shape and influence these questions.

Visual Methodology The collection of methods used in research to understand and interpret images; the philosophy of how researchers practice and utilize these methods to create knowledge.

Introduction

Auto photography usually entails the following steps: first, a researcher gives cameras to research subjects and asks them to photograph particular places and people that are relevant to the topic of the research project. (In some cases, cameras may be distributed with no prior information about the research given, however.) These instructions may be prefaced by an initial interview or other participatory research activities. For example, a geographer interested in children's experiences of urban parks would give youngsters disposable cameras to take to the park for an afternoon. The researcher may already have been observing children at play in parks, or perhaps the researcher interviewed children about their idealized play space. The researcher may or may not give detailed instructions to the children about taking photos; perhaps he/she just tells the children to photo graph what they feel is important about their play space. After the children take their photos, the researcher would then develop the film to see what sorts of environments, places, objects, and perspectives the children captured. These photos would be considered to be data for the research, as would any follow up interview or con versation with the children discussing why they took the particular photographs they did, the meanings the places had for the children, and their ideas about these places or even alternative imaginary play spaces. The analysis of photos and interview or narrative data then takes a variety of forms, from counting and comparing themes to treating each photo as a lesson on what objects, places, and people represent something significant for the subject.

Auto photography is an ethnographic field research method that attempts to 'see the world through someone else's eyes'. Of course, seeing from another person's position is a fraught process, since gaining the true perspective of someone else is impossible. Auto photography, however, provides a tool in qualitative and ethnographic research projects that moves a step toward understanding what qualities of environments and places are important for research subjects in their daily lives. Human geographers have begun to use this tool more and more as photographic technology has become affordable and easier to use. Of course, because auto photography relies on the technology of camera and film development, the history of its use is relatively recent. In geography, the practice of auto photography is closely tied to the development of the inexpensive dis posable camera, a one time use camera that operates with film and became more popular and affordable in the 1990s. The cost of disposable cameras allows scholars with research funds to gather visual data somewhat easily, although for new scholars and those without substantial funding, auto photography may be prohibitively costly, especially when combined with film development. Per haps the popularity of digital photography will widen the use of auto photography, given the many numbers of photos that can be taken with digital cameras, even one time use digital cameras. Self made videos will also become more popular with researchers as the cost of equipment continues to fall.

In human geography, auto photography has largely been used by scholars studying children's geographies throughout the world. The technology is easy to use and allows a self representation from subject groups – children and youth – who might otherwise find highly articulate and reflexive verbal explanations more difficult or intimidating in a research situation. Children might also be drawn to the 'hands on' aspect of taking photo graphs. The use of auto photography is not restricted to work on children and youth, however. Scholars are also using auto photography to study identity, time–space geographies, and human–environment interactions. Par ticipatory methodologies also often utilize auto photography methods since they seek to actively engage research participants in the production of knowledge.

Genealogy of Auto-Photography

Auto photography is an ethnographic research method that has its historical roots in nineteenth century anthropology. In many ways, ethnography and photographic methodologies were born together. Photos were often taken in the field and then presented to academic and popular audiences to show the 'savage' culture far away in lands (or the 'natives' on colonized lands like in North America and Australia) and to illustrate how these strange people were alien and undeveloped in comparison to a Western audience. Today, such photography is largely seen as deeply problematic with colonial, racist, sexist, and Eurocentrist motivations. The anthropologist who photographed his native subjects sought to show the 'strangeness' of the other sociocultural group and accomplished this through poses of those studied that emphasized spectacle and racial–geographic difference. These visual presentations were markedly different from written depictions of alien cultures, although both served primarily to describe and represent societies and cultures different than the ethnographer's own. Photographs were intended by anthropologists to document cultural practices and racialized and sexualized bodies, although ironically, the vast majority of photographs from early anthropologists were staged and posed according to anthropologists' viewpoints and their Eurocentrist stereotypes of the 'savages'. Photographs were less realist reflections of life and people than highly rehearsed and constructed scenes stemming from the ethnographer's interests.

Over time, the camera has become synonymous with fieldwork in ethnographic research, and it became a staple field tool. It remains so today, even though the photographs fieldworkers take have been treated as actual data only recently in anthropology, sociology, and geography. Educating students on gathering and analyzing visual data is still rare in the social sciences, despite the ubiquity of the camera and photography in fieldwork research. Instead, photographs are seen as evidence supporting the central textual analysis, and their importance as visual data are not often mined except by a few experts of visual methodologies.

Auto photography specifically arose as a concept in the 1960s and early 1970s when ethnographers began to question how their own subject positions fundamentally shaped the representations of those whom they studied. This new reflexivity suggested that the scholar's own gender, race, and location were central factors in the types of research methods, fieldwork practices, and analysis that he or she conducted. One of the first uses of auto photography took these issues to heart, although the authors suggest that such questions are often unanswerable in full. The bulk of this project was actually film work, not still photography. The two scholars were named Sol Worth (a film and communications researcher) and John Adair (an anthropologist), and in 1966 they gave 16 mm cameras to Navajo Native Americans in the state of Arizona in the United States. They taught the Navajo participants to use the cameras and edit film, and the resulting work became a benchmark in the participatory activities of research subjects in ethnography. Worth and Adair analyzed the films of their participants by characterizing them as 'Navajo' ways of seeing and experiencing the world. This emphasis placed the Navajo filmmakers' race and ethnicity as the central social categories influencing their work in creating film. While Worth and Adair's project broke boundaries of established research methodologies, its emphasis on ethnicity and racial difference above all other social identities maintained the anthropological practice of highlighting racial and ethnic difference between researcher and researched. Their ideas, however, became a fundamental moment in the development of auto photography and its methodological emphasis on 'ways of seeing', the visual, and knowledge production. It also showed that research participants with no experience of using film technology could rapidly become competent with cameras, editing equipment, and technique.

In the 1970s and 1980s, instamatic cameras, also known by their brand name, 'Polaroids', allowed researchers to have their research subjects take photos and have instant results. These instant photos stimulated immediate discussion with research subjects, and much work revolved around questions about social identity. However, the camera equipment itself was costly, and researchers were not often able to give many participants the cameras to go off alone to photograph. Usually the researchers themselves accompanied subjects on photography excursions. Auto photography as a field method became more widely used and practical when disposable, one time use cameras were marketed in the late 1980s and early 1990s.

The connection between the use of the method and the changing, more affordable, and easy to use technology is an important one. While use of auto photography existed prior to the relatively inexpensive, 'disposable' camera, in the 1990s the wide availability of one time use cameras allowed this research technique to be more widely realizable. The disposable camera allowed researchers to give a camera to participants, rather than loan out expensive cameras at great financial risk or accompany research subjects on short photographic endeavors. Disposable cameras meant that researchers could have participants take photos over longer time periods and on their own without direct intervention of the researcher. In tandem, publishing photos in academic journals was also more likely. As the 1990s progressed, the spread of desktop computing, the development of the World Wide Web, Internet publishing, digital photography, and digital imaging exploded the use of visual methodologies.

In human geography, auto photography arose through this technological possibility. Probably the first geographers to advocate the use of auto photography were Stuart Aitken and Joan Wingate. Their study of children's geographies asked how youth understand their environments, and how auto photography could aid in helping these adult researchers comprehend children's different views of their neighborhoods. Their work inaugurated a methodological approach to children's geographies that emphasized children's local and everyday lives, the impact of social differences like race, ethnicity, and income on children's mobility and experiences of their environments, and the ways that children's places are made through social relationships (e.g., play spaces like a friend's house). One conclusion the authors derived through the auto photography project was that children with cerebral palsy who experienced restricted mobility photographed family more often than children whose mobility was not as hampered. The children with cerebral palsy also took more 'action' shots of other children playing, and the authors concluded that watching other children at play was an important way for the disabled children to participate. Their photos and follow up interviews indicated that they often considered themselves a part of children's playing even if they were only spectators.

Most studies to date in human geography have utilized auto photography with children, perhaps because of its hands on, interactive nature. With time, as visual methodologies become more theorized and prevalent in the discipline, visual field methods will be more widely taught to students and thus be more popular and understood. For instance, more recently, human geographers have drawn on the benefits of the auto photographic method for studying urban geography, social identity, and spatial mobility.

Purposes of Auto-Photography

Photographs derived from the auto photographic field method contribute to ethnographic research in numerous ways. First, because photos are seen, they are considered visual data. In geography the visual has always been an important dimension of scholarship and the production of knowledge, for example, through cartography and landscape analysis. Photographs taken by research subjects consist of a radically different genre of visual data than maps or landscapes, however. Probably the most important aspect of auto photographs as visual data is how the photographer participant oriented him or herself to the spatial context that they photographed. While the visuals images contain symbols, that is, images that can be interpreted, there are many different theoretical trajectories for how interpretation proceeds (from semiotics to psychoanalysis). The researcher would analyze the visual data of the photos by considering how social, economic, and political aspects helped to frame a scene's production. These processes are also central to how we look at photos, and how researchers are able to incorporate visual data like photographs. The objects, spaces, or people in the pictures must also be considered for how the photographer perceived them. For example, a photograph (see **Figure 1**) of a doorway with '300' above appears to be a poorly snapped picture, even complete with finger covering part of the lens. Placed in context of a research project about the social and spatial segregation of different youth groups at a high school in California, the building number takes a different meaning: the teenage girl who took this picture did so to show a particular 'territory' of one ethnic group at the school. Thus, this visual data shows the particular investment this girl has of ethnic social difference and a 'place' in the 300 building for her social–peer group at her high school. In another example (see **Figure 2**), a photograph of a police car outside of the high school could be placed within a geography of police and adult power, the political and economic aspects of policing urban public schools, and the gendered relationships between the police officers at the school (who are men) and the young woman who photographed the car. This visual data, along with her other photographs, helps a researcher understand the specific social–racial geographies of youth at the school. The police car is a visual symbol to be interpreted by the researcher.

Photos are the production of the research subject and reflect her or his intent or position when snapping the picture. Thus researchers can glean different social motivations by considering why certain scenes were captured when they were; each photo is a lesson about the subject who took that particular picture. When a person takes a picture knowing that it is for a research project, they have made a decision to represent themselves through the visual scene they frame in the camera. However, this should not be overstated, since many times people take for granted social situations, relationships, and spaces, so their decisions about photos may not be as conscious or as intended as the previous sentence implies. To continue the example above, the high school auto photography project was framed by initial interviews about social, racial–ethnic, and peer 'cliques' at school, so the research participant photographers were influenced by these conversations with the researcher. The photos they took included a documentary style framework to 'show' to the researcher the places they had discussed. The pictures, in other words, were considered as proof to illustrate what the girls had already indicated in the initial interview, when the camera was handed out. The photo of the schoolyard, for example (see **Figure 3**), is notated with regions for each social group that one girl had described in the interview. Her intent was to show to the researcher, visually, what had only been described verbally at first. She might not have had the obvious

Figure 1 A teenage participant in a study about segregation at a California high school took a photo of this school building marked as the '300' building with a one-time-use camera. She said that the 300 building is where the Latino 'gangsters' hang out. They are her friends, thus she indicated that this building is part of her territory at school. Notice the finger in the corner. Poor image quality is a common issue with auto-photography using one-time-use cameras.

Figure 2 This police car is on the street outside of the high school. Police presence is usual at this urban school. The girl who took this photo used it to show the researcher that police are always around campus because the threat of racial fighting at the school is common. She may be indicating a resistance to police presence, or she may agree that there should be police surveillance at her school; it all depends on how she talks about the photo with the researcher. It may even depend on her mood that day.

intent of giving further evidence for how segregated cliques at school are by ethnicity and income, though. The 'preps' were typified in interviews with girls as having nicer clothes and cars. While the photo showed the school's territories, the photographer had said nothing directly about income and class differences at school. The photo helps the researcher to think about the social processes of economic disparity for peer groupings and segregation at this school, however.

Thus, the process of the photography project elicits different conversations and interactions between researcher and researched than would have occurred with no auto photography. For example, the researcher may sit with the research participant with the photos spread

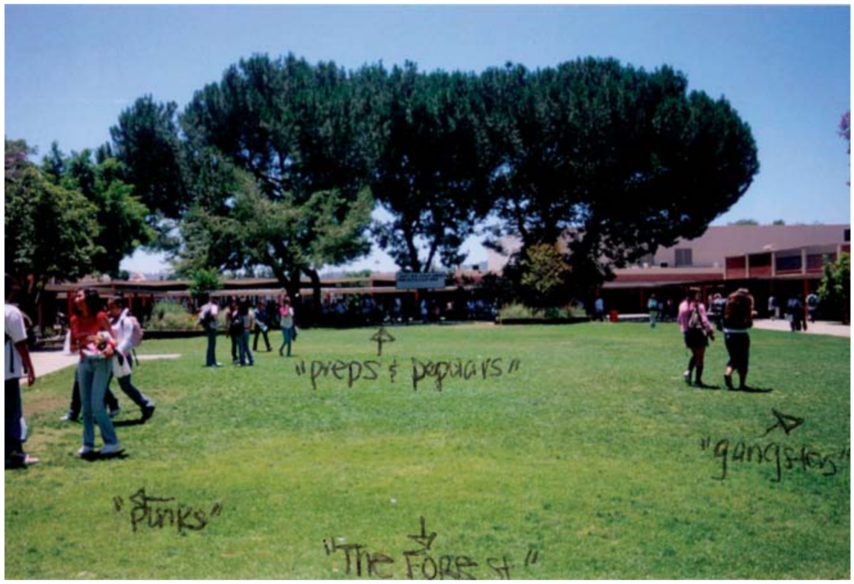

Figure 3 The campus view of the high school is annotated by the research participant's writing on the photograph itself. She noted where on campus different groups 'hang out' everyday. Thus the geography of different cliques is clearly displayed for the researcher. Interviews with the research participant indicated that these groups of friends are also largely racially segregated. The 'forest' is the property of Armenian students, for example, while 'gangsters' are Latinos.

before them. Talking about each photo will bring new information and knowledge to their research relationship than might have been the case with no visual products. This is known as photo elicitation. This process might even highlight contradictions between what the participant might have said before the auto photography method was used, and what they photographed or said afterwards. Auto photography can help the researcher to understand how social identities and practices are always changing and unstable. Photo elicitation also indicates that what is spoken verbally is just as important as the visual data produced in the form of the photographs. It is rare to find auto photography in geography or other social sciences without interviews, participant observation, or other research methods combined with it. For example, one photo (see **Figure 4**) from the high school project shows a blurred shot of a backpack in the foreground on the left, and some kids standing under a tree, to the right. This photo was taken by a girl secretly, as she walked by these boys; her friend's backpack served to hide the camera as she took the picture. She discussed the photo with the researcher in order to talk about these men as dangerous 'thugs'. An important point is that these boys are a different ethnicity than her, and she talked about the danger of these boys in tandem with the fact that they are Latino (the girl is white). The photo then served to open up a discussion about gangs at school, and it indicated to the researcher the racism that some white youth at the school have for Latinos. while very few youth would openly admit to racist sentiments, the photographs of the youth indicated that some white girls at the school were afraid of Latino boys, even though none had been harmed by the boys in any way. They represented the boys in their photos and in the follow up interviews about the boys, as dangerous gangsters that should be avoided, and who should be photographed stealthily. If the researcher had just looked at this photo with no follow up interview, it may never have become clear that the backpack served as camouflage because the girl photographer considered the subjects of her photo to be dangerous. It might just look like a badly taken picture!

Photographs taken through the auto photography method can therefore illuminate contradictions between what people say and what they represent visually. While a girl could say she is not racist, perhaps her photographs indicate that she is overwhelmingly concerned with racialized others at school. This is a contradiction that the researcher can then interpret. Photos can also indicate the desires of participants that they might not articulate in interviews. An image (see **Figure 5**) shows a group of 'rockers' at the high school in California. The teenage girl who took this photo explained that 'rockers' (which another girl referred to as 'punks', seen in the schoolyard photo) were able to integrate racially more easily, because they shared an identity around music, rather than race or ethnicity. This may or may not be true in actuality. What is important, instead, is that the girl expressed a longing for racial integration in her discussion of the photograph. She was not herself a 'rocker' and did not belong to this group of kids, but she took a

Figure 4 This blurry photo was taken by a white girl who wanted to hide the fact that she took a photo of these Latino high school students. The photo was taken after school, and her friend's backpack provided cover as she snapped the picture with a one-time-use camera. The angle of the camera was pointed up, indicating that she held the camera out of view at a low angle. The angle also makes the bright sky darken the image in the lower part of the picture, so that the boys are harder to see. The image quality is not as important as the effort she took in order to show the researcher these 'gangster' Latinos hanging out after school. The girl photographer found these boys to be dangerous, and she said she would never hang out at the corner with them.

Figure 5 'Rocker' youth at school are identified by the chains on the boys' belts, the girl's fishnet stockings, and her handbag style. The girl who took this picture found typical rocker kids in order to show the researcher how youth at school form their own cliques. The photographer indicated that rocker kids are less likely to segregate by race or ethnicity because they unify around their interests in punk and rock-'n'-roll. She does not belong to this clique.

photo of them to express an approval of their supposed racial–ethnic integration.

Finally, the content of photos cannot be taken for granted by geographic researchers. In other words, the two dimensionality of the photograph must be interrogated for meaning that is not so flat and for the spatial dynamics that help to construct the image's con text as well as the perspective of the photographer. The social identities of the photographer are just as important as the objects or images that the photos indicate. This

Figure 6 The graffiti marker is a symbol for gangs, and the photographer targeted this image to indicate the possible presence of gangs nearby her high school. She did not know who 'tagged' this wall, or what gang affiliation it might mark, but the symbolism of the graffiti allowed her point about danger to be represented and communicated. Visual symbols are important devices to relay themes and messages through auto-photography.

point means that any photograph is not merely a neutral reflection of what is captured in the frame of the picture. Taking a photo of a police car (see **Figure 2**) might be an easy mark of youth resistance to adult authority, but the race, age, gender, ethnicity, and life experience of the photographer can indicate an ambivalent attachment to the authority, violence, and power that the police represent in American society.

There are as many ways to interpret a photograph as there are researchers using auto photography. However, there are general approaches that researchers use to code and analyze auto photographs and their contents. The first is to gather the photos into different themes, to code these themes, and then to count how many themes come up across all the photos. For example, perhaps a researcher studying a high school's social spaces would code photos for images of peer groups, particular buildings on campus, local neighborhood streets, friends, family members, and adult authority figures. If there were 100 photos from the high school project, and 75 showed images of graffiti around the school, then this overwhelmingly would indicate that the students were concerned about gangs at school (see **Figure 6**). It might also indicate that the researcher asked every student about gangs in the neighborhood around campus, and that participants wanted to respond to this issue by proving the gangs' existence. Counting and comparing themes or general types of landscape photographs still require an analysis of why those particular groupings occurred.

It is never enough just to indicate the numbers of themes that arose without also interpreting the processes that lead to these groupings. The auto photographer communicates something specific by taking pictures, but the production of a photograph results from the complicated spaces of social life. It is the geographer's job to consider not only the content of the photographs, but how the research relationship put the photograph into production in the first place, how the complex spatiality of the image was produced over time, how the research participant's subjectivity and identities influenced what they decided to photograph, and importantly, how to represent the auto photographs in any research product that results from the project.

Challenges

There are several practical challenges for using auto photography in human geographic research. These include the cost of cameras and film development, the likelihood that photo quality will be poor and inconsistent, and the possibility that participants will drop out of the research project over time. Conceptual challenges also exist. In human geography there is very little theoretical or conceptual work on auto photography or analysis of photographs stemming from use of this method. Challenges do arise as geographers integrate this method into their research. First, although the photograph is a two dimensional object, ways of seeing and the visual itself are never neutral or flat. The photographs are representations of complex spatialities and subjects that are imbued with social regulation and power.

The photograph is a moment in time and space, but that space is unbounded, multidimensional, and even contradictorily experienced and produced. Second, although the auto photograph is produced by a research subject, it is important to remember that the product of this method is still largely controlled by the researcher. The question of who holds the authority of the data is rarely questioned, and this is an area where further debate is needed in geography. Third, researchers should remember that not everyone has the same cultural, social, and economic background in using photography or other technologies of the visual. They must think about the social and geographic contexts of images, photography, viewing, and technological knowledge. Whatever the challenges that exist, however, visual methodologies are rapidly becoming widely utilized, studied, and debated in human geography. As the development of digital photography and video progresses, and digital technologies become more affordable, the use of auto photography will continue to expand the boundaries of human geographic research.

See also: Diaries (Video, Audio or Written); Ethnography; Situated Knowledge, Reflexivity.

Further Reading

Aitken, S. and Wingate, J. (1993). A preliminary study of the self directed photography of middle class, homeless, and mobility impaired children. *The Professional Geographer* 45, 65 72.

Ball, M. and Smith, G. (2001). Technologies of realism? Ethnographic uses of photography and film. In Atkinson, P., Coffey, A., Delamont, S., Lofland, J. & Lofland, L. (eds.) *Handbook of Ethnography*, pp 302 319. London: Sage Publications.

Bateson, G. and Mead, M. (1942). *Balinese Character: A Photographic Analysis*. New York: New York Academy of Sciences Special Publication.

Dodman, D. (2003). Shooting in the city: An autophotographic exploration of the urban environment in Kingston, Jamaica. *Area* 35, 293 304.

Pink, S., Kurti, L. and Afonso, A. I. (eds.) (2004). *Working Images: Visual Research and Representation in Ethnography*. London: Routledge.

Rose, G. (2001). *Visual Methodologies: An Introduction to the Interpretation of Visual Materials*. London: Sage Publications.

Rose, G. (2003). On the need to ask how, exactly, is geography 'visual'? *Antipode* 35, 212 221.

Thomas, M. (2005). Girls, consumption space, and the contradictions of hanging out in the city. *Social and Cultural Geography* 6, 587 605.

Worth, S. and Adair, J. (1972). *Through Navajo Eyes: An Exploration in Film Communication and Anthropology*. Bloomington, IN: Indiana University Press.

Ziller, R. C. (1990). *Photographing the Self: Methods for Observing Personal Orientations*. Newbury Park: Sage Publications.

Avant-Garde/Avant-Garde Geographies

J. D. Dewsbury, Bristol University, Bristol, UK

Glossary

Avant-Garde A theoretical, artistic, or political movement that is 'ahead of its time' and which specifically has the political aim of offering up and instituting new ways of ordering the life of ideas, forms, and people. Given this critique of the *status quo,* it is by its very nature often based on a minority group and short lived.

Esthetic A specific sensory experience that is both instituted and created out of artistic forms and the arts of living, and which heralds particular values and modes of living for individuals and communities.

Situationism The practices and ethos of the Situationist International movement based in Paris from 1957–72 whose prime aim was to change everyday life. The ethos rejects conventional forms of artistic expression and politics utilizing more performative, experimental, psychoanalytical, and symbolic activities.

Introduction: Context and Thought

The *avant garde* is predominantly a modernist term for a movement in art, culture, and politics that cuts at the vanguard of ideas both in terms of their mode of expression and the social impact that they have for everyday living. As a term, it does however have roots in the Renaissance for signaling an advanced position in the arts, and further gains its political purchase in the time of the French Revolution where its kinship to Jacobin politics married it to the general suggestion of an assault on authority and institution. As commonly understood today, it indicates a movement that challenges and goes beyond existing frameworks for considering the relations between art, culture, and politics. As such, the *avant garde* is heterogeneous because it unites groups from very different origins in opposition against the homogenous dominant position or *status quo* of the time. An *avant garde* group was usually made up of just a few people, often those who had different backgrounds and skills, and who cohered only for the duration of the immediate novel event or project that brought them together in the first place. Such ephemeral consistencies usually derived from living in shared neighborhoods, working in the same institutional and artistic spaces, or in being introduced through bohemian social networks as friend or lover. The promiscuousness of such connections and the heterogeneity of background meant that the *avant garde* often amounts to no more than a fragmented set of practices.

Geographically, its genealogy as a term and a practice comes from European social cultural movements, particularly in the context of England, France, Germany, and Italy, leading up to, around, and for the next 30 or so years after, the turn of the twentieth century. Synonymous with such *avant garde* practices were distinct communities set in bohemian areas within the major urban centers of these countries. These were made up of people who knew of each other purely to be either with fellow *avant garde* newcomers or united against the dominant competitors being challenged; as such the overall ethos was one of competition where vying for novelty and heresy was key. In moving against the custodians and bastions of tradition and social propriety, the *avant garde*, although mainly thought of in artistic terms, was equally a pseudomilitaristic endeavor. The *avant garde* is therefore very much 'anarchistic'. Thought of as a war machine, the *avant garde* is productive of the means rather than the end, at ease with war but decidedly queasy in victory, being an act devoid of any determination other than its artistic intent; in other words, existing only and incessantly as the vanguard. As such, the epochal days of the *avant garde* often came in periods of economic indifference (either decadence or nihilism) where riskier positions of intellectual or artistic endeavor could be and were accommodated. As David Harvey has argued in geography, the *avant garde* is perpetually undermined in its political aspirations for its failure to ground and enact its esthetic link between knowledge and action in historical terms.

If scaled up, first in terms of time such that an *avant garde* group can be said to have gained a history and reputation, and second in terms of space through the range and coverage of its spatial impact, the *avant garde* can be seen as part of more general modernist utopian schemas indicative of early twentieth century thinking. One can think here of the way in which Henri Lefebvre's production of space gets clipped under such aphorisms – "we need machines for living in" – that peopled the expressions for new futures in the construction of places like the Bauhaus. Although this comparison goes against the characteristic of the *avant garde* for countering dominant ideas of the time, being institutionalized as part of a dominant cultural elite, it does serve to show, as David Pinder carefully argues, that such modern utopias themselves contained complex counterdiscourses and are thus productive within themselves of further oppositional utopianisms. More generally, the *avant garde* as a notion itself did gain the attention of, and was indeed articulated as such by, the intelligentsia becoming a key point of intellectual debate both in the informal scholarly insti tutions fueling coffee house conversations on places like the Left Bank in Paris, as well as in the more formally instituted school of thought like that of the Frankfurt School. Wherever it was articulated, its political ser iousness was forefront, the latter institution mindful to distinguish its place and role in society as distinct from any notion of kitsch and mass culture objectification.

Now in the twenty first century, it is debatable whe ther or not there can be an *avant garde* as such given that the notion is diminished in the wake of a plethora and pluralism of different *avant gardes*, and given the global ization and speed of ideas which have diluted the fix of any development of thought and action, resulting in an overwhelming diversity of emerging social movements. Overall, human geography has looked at the geography and spatialities of the *avant garde* movements in the last 150 years with perhaps the key *avant garde* movement to gain the most attention by geographers being that of the Situationist International founded in 1957 which, under the 'command' of Guy Debord, argued for 'the free construction of daily life' through the investigation of more explicitly revolutionary critiques of everyday spaces and practices than the *avant garde* movements that preceded it. It was disbanded in 1972, having achieved a total membership of no more than 70. Despite this, the ideas it propagated continue to live on and influence urban geographers' thinking, with particular emphasis being placed on the argument it made for the necessary role of spatial transformation for the success of any social revolutionary change.

One can of course argue that there has been an *avant garde* within academic geography itself, taking the work of Gunnar Olsson, a self appointed heretic, as the ex emplary instance of a geographer who has positioned himself outside the standard practices of writing by, not least in his excellent book *Abysmal*, uniting personal and philosophical, artistic, and political modes of expression in examining experimentally the art of carto graphy across the genres of quantitative and qualitative registers. Taking the breath away of our discipline's critics, much as Marcel Duchamp's urinal entitled 'Fountain' did to the artistic community in 1917, is a classic *avant garde* result – 'Is Olsson's work geography?' Another indication of an *avant garde* of sorts within geography itself comes in the artists in residence now peopling the contingent of geography staff in many a UK geography department. This at least raises the potential for *avant garde* experimentation as the lan guages of laboratory and studio come into closer proximity co producing alternative conversations and points for debate, even if, having become almost com monplace, the presence of an artist in residence is in itself no longer *avant garde per se*.

Avant-Garde: Translating the Esthetic into Politics

Addressing the conceptual purchase of the *avant garde*, one can combine the relationship of the esthetic to the political by unpacking the arts into three key categories – that of experience, performance, and expression – while distilling and summarizing their respective influence as manifested through, and housed within, the spatialities of particular institutions and communities. Within this telling of the *avant garde*, it is also possible to discern six interlacing themes: a philosophical affiliation with 'ni hilism'; an 'agonistic sense of politics' and ethics; concern with the realm of 'technological science' in projects like Futurism; a 'decadent lifestyle' with a parallel sense of morality; an 'attention to architecture', the best example seen in the De Stijl movement; and a concern with 'everyday practice' for the achievement of its aims (reaching its apogee with the Situationalists).

Experience

The *avant garde* is particular in the way its experimen tation with cultural forms united conscious practice with unconscious transformations. It is very much about thinking life literally by materializing it in different forms. As such its currency is experience itself. Jonathan Crary shows how, from the designs of the Italian Fu turists, to the philosophical refiguration of matter, thought, and time in Henri Bergson, and to the social reorganization of the self in an age of technological and esthetic re imagining in the expressionist but equally academic observations of Walter Benjamin, the *avant garde* gathers together a whole set of experiments that shift the general conception of the body toward a more pliable composition that acts as the interface between technology and corporeality.

Taken forward to perhaps its last coherent formation in the guise of the Situational International, what is paramount in the social impact of *avant garde* thinking is that such thoughts explicitly 'matter' – in other words, are material enactments themselves. Thus, an *avant garde* movement distils and enacts its ideas in an embodied and performative manner soliciting immediate, everyday, and thus potentially political, action. Explicit in these for mulations of the *avant garde*, according to Pinder, is an attention to the ways in which experiential aspects of our individuality, such as feelings and desires, are affected by the everyday environments in which we live. Given the avant garde's placement within a particular historical period, what is really being talked about here is the maelstrom of architected stimuli prevalent in modern urban living. However, these understandings were not thought of in asymmetrical terms, and attention was equally placed on the way our individual experiential dispositions – of mood and emotion – also affect and give form to the geography of those urban landscapes.

In its translation into human geography, the *avant garde*, as seen in experiential terms, is rendered more as a subversive rather than straightforwardly oppositional attitude to geographical thinking. This is particularly, taking Kevin Lynch's work as an example, illustrated through the cartographic sensibilities that are evoked in its name. Esthetics and politics come together in an *avant garde* mapping that refigures the map as a questioning of how we know ourselves and hence our place in society. The map here, as depicted by Simon Sadler, both directs our practice and our placement in the environment while at the same time maps are produced through the experiential register of those sensibilities, moods, and derives solicited from the placements themselves. As such, it is easy to see why psychogeography is seen as an evolution of this way of thinking, and one which still holds sway today. Nonetheless, psychogeography's status of being *avant garde* is somewhat moot, being just the latest genealogical trace of the *avant garde* principles after the Situationists, and as such speaks the old ground rather than the frontier, both literally and in the conceptual thinking behind the practices it invokes. So in a sense geographers, like Steve Pile, have side stepped the *avant garde* in turning to psychogeography to iron out the indifference and ultimate ineffectualness of the incessant novelty of the *avant garde* enterprise of what is in effect an incessantly refolding mapping, looking instead to the richness, and the rich lessons, of the past for the root of its social politics.

In sum, specific creative activities draw attention to the experience of everyday living, politically trying to fashion its direction by facilitating the production of new modes of behavior and social–spatial awareness.

Performance

In terms of performance, it is important to distinguish between street and stage, or artistic studio and gallery, author's intent and reader experience. At its heart the *avant garde* is about merging art and life, and taking as the example, the cultural institution of the theater, such artistic spaces and institutions were in effect the laboratories for forging new forms of praxis for everyday existence. What makes theater stand out is its explicitly embodied and immediately present presentation of ideas, and as such it exemplifies how the public, as a spectator of an art form, here intensified within the heightened communal affect of an audience, come to see themselves as part of a historical process through which the metonymic depictions on stage embolden them to experience their transformative potential to effect change in the less reified ambits of the everyday life of the street. Of course, theater is an illusion, the key word here is potential, and any resultant actual transformative action peters out all too quickly. However, in two key ways, theater does gain critical impact and is perhaps the most successful translation of *avant garde* sentiment into actual political action. First, taking the work of Bertolt Brecht as an exemplary *avant garde* practitioner, theater enacts a spatial interperformativity by both drawing on the collective knowledge and memory of shared experience – key adjuncts for the production of community – and in explicitly using the audience as the link between the dialectical relation between performance and reality by getting them to 'think' aside from the theatrical illusion. The art of emotional distance, that kept the actor and audience aware of the character's individual relationship with society rather than with itself or wider metaphysical questions, staged the critical distance required for an *avant garde* practice not to slip into an indiscernible merging of art with life. Second, theater, in its *avant garde* tendency to distil its principles to their most purified form, developed what came to be known as the Theater of the Absurd as coined by Martin Esslin's 1991 book of the same name. Here, the use of language, images, forms, rhythms, sounds, and movements crafted pared down vistas on to metaphysical ideas, and in effect set about staging questions about self perception and representation, language and identity, and body and space. The exemplary figure here is Samuel Beckett, whose influence today stretches into dance, film, sonic, and video installation art as well as theater, literature, and painting.

In broader terms everyday terms, questions of style, fashion, and manner were considered by *avant garde* groups to envisage a subconscious translation of the creative values of the art form to the quotidian tasks carried out by the worker soldiers of public life. To different degrees and to different levels of success, *avant garde* movements removed the distance between artist and plebiscite. However, the aporetic irony at the heart of the notion of the *avant garde* is that in terms of its political purchase, very much the drive of its ambition, the *avant garde* can never mobilize a large number of followers; by its very definition, it is always a minoritarian enterprise. As history shows, Nietzsche's belief in the aristocratic nature of the artist holds true and the source of *avant garde* politics, the artistic, can never, if to be critically effective, merge with the everyday aspirations and demands of the street.

Expression

In the realm of expression, be that in word, movement, paint, image, or form, the avant garde once again enacts its political gesture by opening up how these different cultural fields can become autonomous of religious, political, and economic forces. One can think of how

Charles Baudelaire was prosecuted by the political establishment of Second Empire Paris by offending public morality through his poetry. Geographically, as illustrated in Harvey's book *Paris: The Capital of Modernity*, and also in Jill Fenton's recent research, such *avant garde* movements have had strong linkages to particular areas and communities; specifically here to the urban village of Montmatre in Paris, ghosted by the insurgent spirit of the Paris Commune, and the artistic studios and bohemian haunts peppered throughout its narrow streets.

Of course, our most relied upon mode of expression, the very art and artifice of writing, when considered in the light of the covert philosophical work of Maurice Blanchot, in particular his book *The Infinite Conversation*, turns it into the ultimate and incessant act of avantgardism. Thus, like Blanchot, the author takes on the very substance and identity of that which engages her, becoming permanently on the vanguard of infinite conversation. In opening up how to write experience, and in questioning the boundaries of what was to be judged as art, the *avant garde* is then about moving beyond both the form and content of expression. In terms of form, stylistic questions of competence where slighted, and alongside the new technological advances – as in both new phenomena to be witnessed (railways in Impressionistic art for example) and as new equipment for capturing the world (photography) – experimental and thus surreal depictions of everyday life came to the fore; hence their disturbance of, and the consternation they caused for, the *status quo*. Ironically, but purposefully, for the manifestation and thus recognition of an *avant garde* movement, galleries (the inaugural exemplar being the alternative of Salon dé Refuses to the Paris Salon of 1863) were the key institutional sites for the perception, appreciation, canonization, and eventual secularization of the *avant garde*. With such an inevitable teleology toward secularization, the *avant garde* has a short half life; perhaps that is why its endgame figure, Samuel Beckett, insisted so much on failing again and failing better to perpetually avoid such a 'dead end'.

In terms of content, the traditional patronage of the arts by the elite political, economic, and religious groups loses its grip on censoring the values depicted in the artworks themselves. Thus, the defense of beauty, truth, integrity, propriety, and responsibility conveyed in stock images, perfected proportions, etc. are dispensed with or consciously subverted. On these grounds, one finds the quintessential expression of a philosophical nihilism at the heart of *avant garde* practice. By and large, expression took a more emotional rather than educational form, and often this was achieved through the aim of integrating the subject matter depicted with the form in which it was expressed. Whatever was to be shown was to be depicted merely through the presentation of its being. It is for this reason that the term absurd fits the *avant garde* movement of theatre. So it was not through artifice but more through instinct and intuition that dramatists such as Beckett, Jean Genet, Eugene Ionesco, and Harold Pinter aimed to solely present life as it is on stage. Five minutes of watching someone do nothing on stage might be true to life, but is it art? Five minutes of someone doing nothing framed as art is absurd. Of course, the representational additive of the frame (it being theater) was played with, but its nonrepresentational ethos gave it the *avant garde* characteristic of being heretical. Likewise in art, following the work of Martin Jay, one can think of Marcel Duchamp's 'readymades' in the same light. Eventually, the tendency toward a purification of the principles and presuppositions of the *avant garde* whatever its form – be it theater, poetry, art etc. – results in a pared down epitome of form that in postmodern times becomes easy parody. That is not to say that one cannot still see successful *avant garde* tendencies in art movements today; the work of Dogma 95 in the cinema is a good example.

Conclusion: Politics

The question of distance between art and life is crucial in evaluating the *avant garde*, and significantly an *aporia* lies at the heart of the *avant garde* project. If creative principles are to be merged with everyday practice, it is a difficult politics to achieve given that a certain social distance is necessary for any political action. Ultimately, the sublimation of art into political praxis is both impossible and undesirable, although not without merit, given its catalytic capacity for engendering novelty, invention, and originality. It is for these reasons that when most successful in political and artistic terms, the *avant garde* is about exposure and failure.

As the twenty first century finds its feet and heads into its teenage years, the concept of the *avant garde* is really a defunct one, and one that no prefixing of neo or post is going to salvage. As Hal Foster has argued, with novelty so fast and furious, or spent, the space of the *avant garde* is reversed in polarity, no longer ahead but looking behind to reconnect with a lost practice so as to disconnect with the workings of the present. And further, if we believe we have never been modern, and as nano, biological, and intelligent technology grows apace, it is not perhaps a question of art synthesizing with life, but with life becoming by nature more artistic itself. The undiscovered country of *avant garde* thought has been found in what Gilles Deleuze has called, after Beckett, the 'exhausted', but the experimentation with the thinking once there is only just beginning.

See also: Anarchism/Anarchist Geography; Critical Theory (After Habermas); Situationism/Situationist Geography; Surrealism/Surrealist Geographies.

Further Reading

Beckett, S. (1983). *Worstward Ho*. London: Calder Publishers.
Benjamin, W. (2002). *Selected Writings,* Vol. 3: 1935 1938. Cambridge, MA: Harvard University Press.
Bergson, H. (2004). *Matter and Memory*. Dover Publications.
Blanchot, M. (1993). *The Infinite Conversation*. Minneapolis, MN: University of Minnesota Press.
Bonnett, A. (2006). The nostalgias of situationist subversion. *Theory, Culture and Society* 23(5), 23 48.
Bourdieu, P. (1993). *The Field of Cultural Production*. Cambridge, UK: Polity Press.
Crary, J. (1999). *Suspensions of Perception*. Cambridge, MA: MIT Press.
Deleuze, G. (1997). *Critical and Clinical*. London: Verso.
Esslin, M. (1991). *The Theatre of the Absurd*. London: Penguin Books.
Fenton, J. (2004). A world where action is the sister of dream: Surrealism and anti capitalism in contemporary Paris. *Antipode* 36(5), 942 962.
Foster, H. (1996). *The Return of the Real: The Avant Garde at the End of the Century*. Cambridge, MA: MIT Press.
Harvey, D. (1989). *The Condition of Postmodernity*. Oxford: Blackwells.
Harvey, D. (2003). *Paris, Capital of Modernity*. London: Routledge.
Jay, M. (1994). *Downcast Eyes: The Denigration of Vision in 20th Century French Thought*. Berkeley, CA: University of California Press.
Lefebvre, H. (1990). *The Production of Space*. Oxford: Blackwells.
Lynch, K. (1960). *The Image of the City*. Cambridge, MA: MIT Press.
Olsson, G. (2007). *Abysmal: A Critique of Cartographical Reason*. Chicago, IL: University of Chicago Press.
Pile, S. (2005). *Real Cities: Modernity, Space and the Phantasmagorias of City Life*. London: Sage.
Pinder, D. (2005). *Visions of the City*. Edinburgh: Edinburgh University Press.
Sadler, S. (1999). *The Situationist City*. Cambridge, MA: MIT Press.

Relevant Websites

http://www.goatislandperformance.org/home.htm
A key US/UK theatre group using Beckett and Deleuze producing avant garde textual documentations.
http://www.cddc.vt.edu/sionline/
A Situationist International archive site.

Aviation

T. M. Vowles, University of Northern Colorado, Greeley, CO, USA

Introduction

The study of air transport geography can be divided into three sections of inquiry: airlines/industry, airports, and aircraft. Air transport geographers are generally concerned with examining the role these different components have, either separately or in combination, in the geographic distribution of human activities across space.

The majority of air passenger and cargo traffic is concentrated in and between three regions: North America, Europe, and East Asia. **Figure 1** shows the entire top 30 airports in terms of passenger enplanements that are located in these regions with North America having 18 and Europe and East Asia six each. The top 30 airports in terms of cargo handled, shown in **Table 1**, shows a similar geographic spread with 12 of the top 30 in North America, 11 in East Asia, six in Europe, and one, Dubai, in the Middle East.

Within North America the geography of air transport is focused primarily around flowing passengers through the many hubs operated by the numerous carriers in the market. These hubs serve two purposes: first to act as multidirectional hubs for connecting traffic and second to act as regional collection points for traffic from smaller centers. The primary flow directions are north–south flows along both the Atlantic and Pacific coast and east–west transcontinental flows.

European air traffic is heavily focused on a band running southeast from London and Amsterdam in the north through Paris and Frankfurt down to Milan and Rome in the south. The liberalization of the European markets has left two strong geographic footprints. The first is the emergence of four strong hubs at London Heathrow, Amsterdam, Frankfurt, and Paris De Gaulle, which in turn has led to a shift away from national net works, shown in **Table 2**. The second geographic impact of liberalization is the emergence of low cost carriers, led by the highly successful Ryanair, which in turn has led to the appearance of a strong network of secondary airports concentrating on providing accessibility for holiday makers in the northern parts of the continent to leisure destinations in the southern part.

The East Asian market is the fastest growing market in the world. Flows here are focused around three areas: the Southeast Asian hubs of Singapore, Kuala Lumpur, and Bangkok; China's gateways of Hong Kong and Shanghai; and the North Asian hubs of Tokyo and Seoul. Global

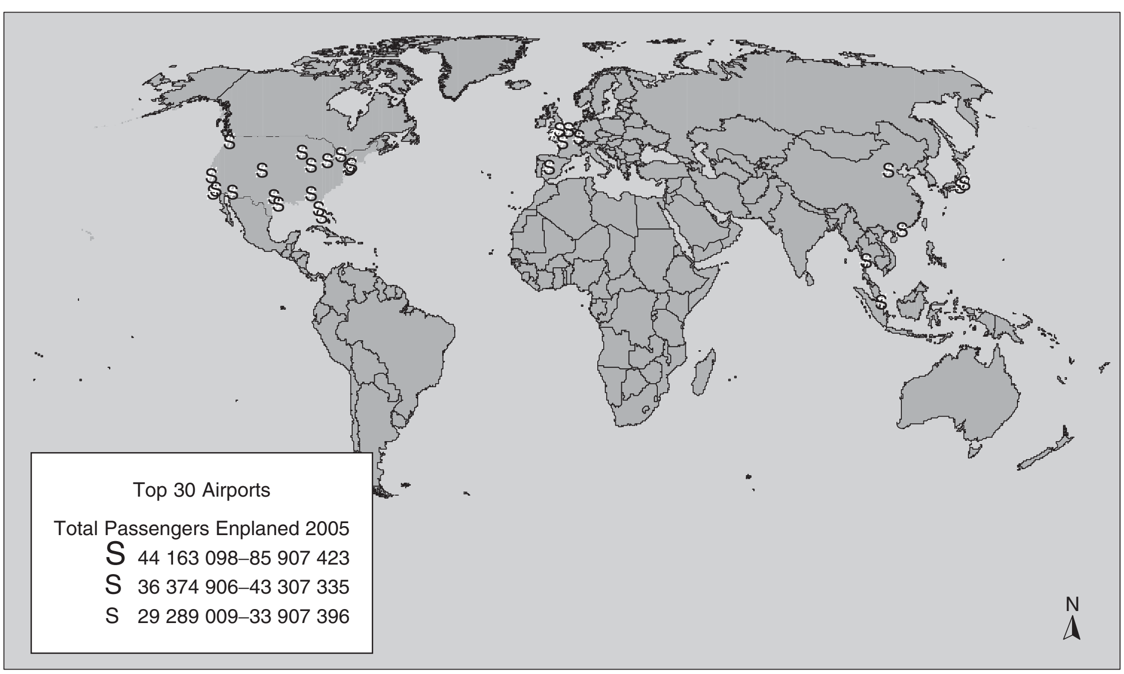

Figure 1 Top 30 airports in terms of passenger enplanements in 2005.

Table 1 Top 20 cargo airports measured by tonnes handled

Rank	*Airport*	*Traffic (tonnes)*
1	Memphis (Mem)	3 554 577
2	Hong Kong (HKG)	3 132 449
3	Tokyo Narita (NRT)	2 373 133
4	Anchorage (ANC)	2 372 903
5	Seoul Incheon (ICN)	2 133 433
6	Los Angeles (LAX)	1 902 547
7	Frankfurt (FRT)	1 839 084
8	Singapore (SIN)	1 795 646
9	Miami (MIA)	1 778 897
10	Louisville (SDF)	1 739 029
11	Taipei (TPE)	1 701 020
12	New York (JFK)	1 683 073
13	Chicago O'Hare (ORD)	1 681 835
14	Shanghai Pudong (PUG)	1 642 062
15	Paris De Gaulle (CDG)	1 637 610
16	Amsterdam (AMS)	1 467 205
17	London Heathrow (LHR)	1 412 031
18	Dubai (DXB)	1 169 256
19	Bangkok (BKK)	1 058 145
20	Indianapolis (IND)	1 000 263

Table created from CargoWorld.com data.

Table 2 Top 20 European airports in terms of passengers enplaned in 2002

Rank	*Country*	*Airport*	*Passengers 2002*
1	United Kingdom	London/Heathrow	63 041 754
2	France	Paris/ CharlesDeGaulle	48 257 964
3	Germany	Frankfurt Am Main	48 078 824
4	Netherlands	Amsterdam/Schiphol	40 587 562
5	Spain	Madrid/Barajas	33 696 101
6	United Kingdom	London/Gatwick	29 509 921
7	Italy	Roma/Fiumicino	24 204 778
8	France	Paris/Orly	23 143 632
9	Germany	Munchen	22 877 714
10	Spain	Barcelona	21 164 324
11	United Kingdom	Manchester/Intl	18 605 651
12	Denmark	Kobenhavn/Kastrup	18 189 580
13	Spain Palma	De Mallorca	17 758 972
14	Italy	Milano/Malpensa	17 330 080
15	Sweden	Stockholm/Arlanda	16 636 815
16	United Kingdom	London/Stansted	16 044 864
17	Ireland	Dublin	14 836 153
18	Germany	Düsseldorf	14 588 642
19	Belgium	Bruxelles	14 315 810
20	Austria	Wien/Schwechat	11 911 246

Table created from Eurostat data.

flows into and out of the region are focused primarily around Singapore, Hong Kong, and Tokyo. It is predicted that this region will continue its enormous growth rates due to the expansion of Chinese markets both domestically and internationally.

Globally, airlines around the world serve 14 000 city pairs with over 75 000 flights daily. Estimates are that 40% of the value of all goods traded are transported by aircraft and the industry employs over 4 million people around the world and generates over $400 billion in economic output.

Airlines

The early international airlines began their services through the mandate of national governments and were used, in the case of some European countries, to serve and reach outlying colonial outposts. It was the location of these outposts and colonial holdings that dictated the early linear networks operated by these international airlines. Early American airlines were used primarily as mail carriers and some of their geographic areas of domination today can be traced back to these mail routes.

As international air travel became more common, certain general air service agreements were reached between international governments. The first five of these air freedom rights, shown in **Figure 2**, were decided upon in the Chicago Convention of 1944. The First Freedom is simply the authorization to fly over another country's territory while on route to a third country. The Second Freedom permits an airline flying from its home country to a third country to stop in another country for maintenance, fuel, or in emergencies. The Second Freedom however does not allow the carrier to pick up or discharge any passengers or cargo. Third and Fourth Freedom rights allow a carrier to carry passengers and/or cargo to and from another country. Third Freedom rights cover the outbound trip and Fourth Freedom covers the inbound journey. Fifth Freedom rights allow airlines to carry passengers from the home country to another country and then continue on to a third country with the right to pick up passengers or cargo in the intermediate country.

Subsequently, four additional air freedom rights were eventually agreed upon and it is these air freedom rights that serve as the basis of all bilateral and multilateral air service agreements between countries. These four freedoms are more complicated than the first five and are the result of the increased complexity and economic importance of international air transportation. Sixth Freedom Rights allow the carrier of a country to carry passengers between two separate countries with a stop in the airline's home country in between. The establishment of this freedom has allowed places like Amsterdam and Singapore to become important hubs sites in international air transport and allow carriers such as KLM and Singapore Airlines to become important international airlines. Seventh Freedom Rights allow a carrier to operate flights between two separate countries that do not include the home country of the operating carrier. An

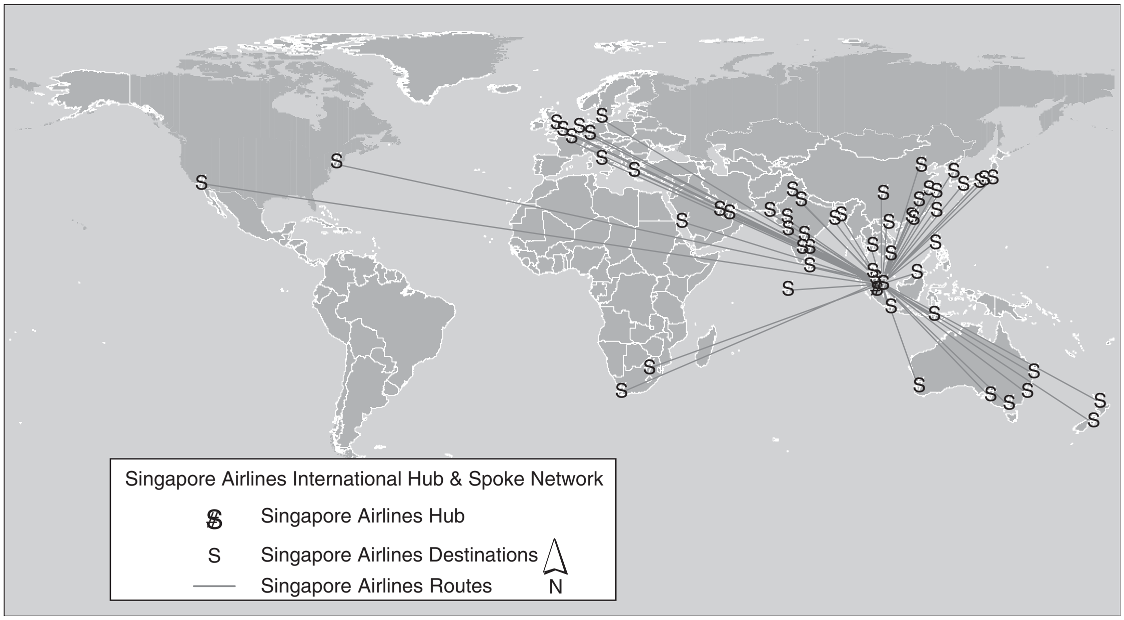

Figure 2 Singapore Airlines international hub and spoke network.

example of this type of practice is Emirates operating service between Auckland and Sydney. Eighth Freedom Rights, also known as 'cabotage' allow a carrier to carry passengers or cargo from the carrier's home country to another country that uses more than one stop along which passengers may be loaded and unloaded. The final air freedom is the Ninth Air Freedom, which allows a carrier from a foreign country to operate between two points in another country without having to have the flight serve the carrier's home country. This is also referred to as 'full cabotage' or complete 'open skies'.

The major airlines of the world, shown in **Table 3**, are headquartered in the three major air transport regions of the world – North America, Europe, and Asia. The majority of the traffic carried by North American carriers is domestic traffic, while that of the other carriers is international, though heavily focused on intraregional traffic. This is due to the geographic size of the domestic markets for most of the European and Asian carriers, which are relatively small in comparison to the domestic United States market, which accounts for 33% of all passenger air travel in the world. The largest international markets are the intra European and intra Asian markets which account for 23% and 16% of the total number of passengers carried globally. The largest market between regions is between North America and Europe, accounting for nearly 4% of all the air passengers in the world.

In the past 30 years, the air transport industry has undergone immense changes and none has had a greater impact than the gradual deregulation and liberalization of the industry. The United States led the deregulation push by deregulating its domestic market in 1978. This

Table 3 Top 10 airlines in terms of passengers carried in 2004

Rank	*Airline*	*Thousands*
International		
1	Lufthansa	34 407
2	British Airways	28 567
3	Air France	26 904
4	KLM Royal Dutch Airlines	20 286
5	American Airlines	18 841
6	Singapore Airlines	15 877
7	Cathay Pacific Airways Ltd.	13 623
8	Thai Airways International	13 277
9	Scandinavian Airlines System (SAS)	12 823
10	Japan Airlines International	12 732
Domestic		
1	Delta Air Lines	79 289
2	American Airlines	72 729
3	United Airlines	61 868
4	Northwest Airlines	46 525
5	All Nippon Airways	42 845
6	Japan Airlines	39 004
7	US Airways	37 949
8	China Southern Airlines	36 374
9	Continental Airlines	30 883
10	America West Airlines	20 152
Total		
1	American Airlines	91 570
2	Delta Air Lines	86 783
3	United Airlines	71 236
4	Northwest Airlines	56 429
5	Japan Airlines International	51 736
6	Lufthansa	48 268
7	All Nippon Airways	46 450
8	Air France	45 393
9	US Airways	42 400
10	Continental Airlines	40 548

Table created from International Air Transport Association's World Air Transport Statistics data.

allowed airlines to determine for themselves the markets they would serve and the fares they would charge. Initial results of deregulation included an immediate rise in competition among existing carriers, entrance of a sub stantial number of new carriers, and the eventual bank ruptcies of several new and old carriers. A number of different factors, both industry and nonindustry related, led to the failure of these competitors. Among the in dustry related factors were the relatively high labor costs of the incumbents, fleet fuel inefficiencies, and fleet mismatch in the new environment of hub and spokes, and dense markets attracting new entrants eating into the incumbents' profits.

The geographic impacts of deregulation and liberal ization are seen in two major areas, types of airlines and operational systems of airlines. As deregulation and lib eralization of air transport spread around the globe, a dramatic increase in the number of competitors in the market has occurred. The new airlines, usually low cost or low fare carriers, are widespread in North America and Europe and are increasing in importance and num ber in Asia and other regions around the world. They generally serve major metropolitan centers with service to other metropolitan centers and/or leisure/holiday destinations and are intracontinental in their geographic service range. A large number of them such as Southwest in the United States and Ryanair in Europe are extremely successful in terms of both network coverage and eco nomic results. Part of this success is attributed to having lower costs than their larger rivals through practices such as high fleet utilization, use of secondary airports where fees and congestion are lower, and lower labor costs. These cost savings allow low cost carriers to offer lower fares to their passengers putting pressure on the legacy carriers to match or lose market share.

The hub and spoke system expanded greatly following deregulation. The larger carriers began to rely on the hub and spoke system as a way to increase the number of markets they could serve by connecting thinner markets that were not feasible for direct service through a hub. As liberalization expanded globally, the hub and spoke sys tem took on an international scope with carriers such as Singapore Airlines and KLM using Sixth Freedom Rights to flow international passengers through their hubs (shown in **Figure 3**). This practice allows these carriers to be successful despite not having domestic markets of any consequence.

An additional outcome of the industry liberalization and the resulting globalization was the introduction of alliances. Alliances within the air transportation industry are defined as a set of agreements between two or more airlines to work in conjunction in a particular matter while still possessing their independence. There are nine different types of alliances that airlines can utilize: cost sharing ventures, asset pools, pro rate agreements, code sharing, feeder, marketing alliance, joint ventures, inte grated feeder, and equity stakes. Those of particular interest to geographers include code sharing, feeder, in tegrated feeder, and joint ventures.

Three major global alliances, Star Alliance, One World, and Sky Team, operate in the world today and the major airlines of the world are connected to one of the three in some form. One of the cornerstones of all three alliances is the ability of aligned carriers to participate in code shar ing, an agreement between two or more airlines which allows the airline that does not operate a particular flight to put its two letter identification code on the flight of the operating carrier for the purpose of marketing and selling tickets in that particular market. Code sharing benefits not only the airlines but also their passengers and the airports they serve. Benefits include increased service options, seamless transfer of passengers and luggage between car riers, an increase in marketable city pairs, increased linkages to global markets, and the ability to circum navigate some air service agreements. An example of code sharing is Singapore Airlines and Air New Zealand, who code share on flights from Auckland to Frankfurt. A ticket can be purchased from Air New Zealand for all legs of the journey but the actual operating carrier varies for each leg. From Auckland to Singapore, Air New Zealand operates the flight and on the Singapore–Frankfurt leg Singapore Airlines is the operating carrier.

Another form of alliance prevalent at primarily the domestic level is the feeder or integrated feeder agree ment. In these agreements a carrier, generally a small one using small equipment, feeds traffic to another carrier from regional airports to the larger carrier's hub oper ations where they are then transferred to the larger carrier's network to reach their final destination. These types of agreements allow the larger carrier to access markets they might not be able to serve due to such issues as equipment or market size. Passengers benefit by gaining access to the air transport network and can be somewhat assured of the quality of the carrier providing service under the larger carrier's brand. The difference between feeder and integrated feeder is that the inte grated feeder generally operates its service under the 'colors' of the larger carrier and rarely if at all operates its own separate service. An example of this type of agree ment is Air New Zealand's agreement with Air Nelson, Mount Cook Air, and Eagle Air who all fly under Air New Zealand Link title. These carriers operate regional service in New Zealand with smaller turboprop aircraft under the Air New Zealand Airlink brand and feed re gional traffic to/from hubs at Auckland, Wellington, and Christchurch.

Dedicated cargo airlines such as FedEx, UPS, and DHL are among the most successful airlines in the world. The geography of air cargo service is very similar to air passenger service with North America and Europe being

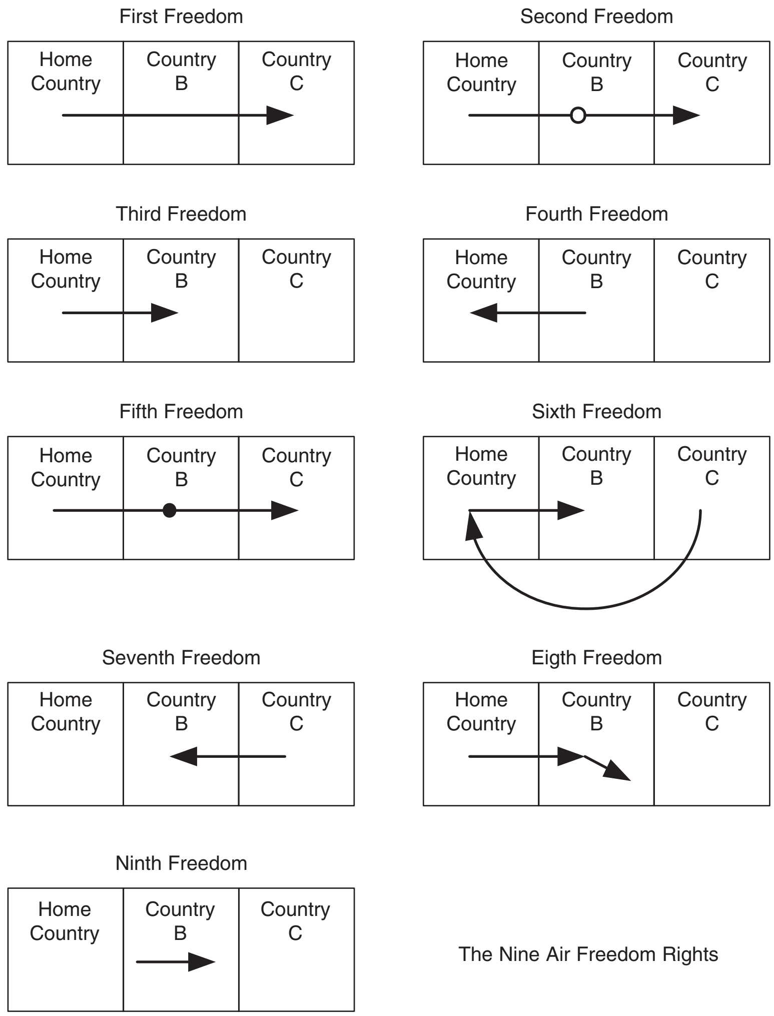

Figure 3 The nine Air Freedom Rights.

the key destinations for cargo both inter and intra regional and the Asia Pacific region being the primary origin of air cargo due to its manufacturing expertise. Four of the five fastest growing airports for air cargo are in China and ten out of 20 are in Asia. The top cargo airports in the world show geographic similarities to those top passenger airports with a clear Asian influence and the emergence of smaller passenger airports in the United States having significant air cargo influence. Airports such as Memphis (FedEx's primary hub) and Louisville (UPS's primary hub) derive more of their business from air cargo services than they do passenger services. The continuing globalization of markets around the world will continue to highlight air cargo services.

Airports

The geography of airports can be broken into two key areas: the study of airport sites and the study of airport functions. The geography of an airport site looks at the airport at three different scales: international/regional/national scale, local scale, and site specific scale. At the international/regional/national scale the airport's lo cation and function within a larger air transportation network is examined. Among the questions that may be asked include: is the airport a hub or spoke in an airline's network? What sort of market does the airport serve? The level of accessibility of an airport within the metropolitan area it services is at the core of geographic analysis of airports at the local level. The various components of the airport such as terminal design, gate placement, and services are the focus of site specific enquires, the third scale.

Physical geographic considerations are important in understanding the sites selected for airport construction. Large tracts of flat, obstruction free parcels of land are necessary for the construction of airports. A large amount of land is needed for such uses as aircraft operations (3000 m runway for 747 to takeoff), terminals (some terminals are in excess of 2 km in length), parking (aircraft and automobiles), associated land uses (fuel farms, maintenance hangers, cargo storage, fire stations, etc.), and buffers to separate uncomplimentary land use (limit noise and other pollutants). Climatic variables such as wind direction and temperature extremes are other important geographic considerations. Finding large enough tracts of land to support airports is proving to be more and more difficult as urban areas grow. In addition to operation considerations airport developers and operators are cautious of other land use types surrounding the airport. As large tracts of land suitable for airport sites become scarcer, particularly in close proximity to urban populations, airports are being located farther out in the periphery of urban areas leading to issues involving accessibility. In some extreme cases, lack of suitable land leads developers to take such dramatic steps as creating man made islands in open water locations near Osaka, Japan and Hong Kong, China for airport creation.

Airports are important components of not only the air transport network but also in urban systems serving as growth poles, drawing commercial, industrial as well as residential developments. In those cities where new airports have been built the old sites open up new development opportunities. Airports Council International estimates that airports are economic engines accounting for more than $500 billion of economic activity in the United States in 2001, including 1.9 million direct and 4.8 million indirect jobs.

The function airports possess in air transport networks is an area of study of great importance to geographers. The various roles include hub airport, spoke airport, secondary airport, and niche airport. Depending on the dimensions of the network being studied airports may have multiple functions. The hub and spoke system and the role a particular airport plays in the network is one of the most important topics studied by air transport geographers.

There are two primary definitions of hubs currently used by geographers. The first defines a hub as a geographical area (airport) classified by the percentage of total passengers that enplane there (airport enplanements/total country's enplanements). The second definition is that a hub is a subjective corporate decision to concentrate resources with the object of interconnecting passengers to maximize profits. The goal of a hub and spoke system from this second viewpoint is to bring a large number of passengers from outlying points, assemble them at the hub and then fly them on to their final destination. In this definition, hubs are used to enhance the overall efficiency of a carrier's service and increase the number of serviced destinations. From a mathematical standpoint the use of a hub and spoke system by an airline allows the carrier to increase the number of destinations served from a hub airport geometrically instead of arithmetically.

Hub and spoke systems have both advantages and disadvantages when compared to a point to point system. Among the advantages is the ability of the hubbing carrier to offer a large number of destinations from hub airport, frequent service, large numbers of passengers flowing through the airport, and higher load factors. Disadvantages of this type of operational system include highly variable work demands, hub premium on fares, missed connections or delays impeding the entire network, high congestion at hub airports, and longer travel time.

The increase of importance of low cost carriers around the world in the last 20 years has led to the emergence of a class of airports termed 'secondary airports'. These secondary airports are usually within the same metropolitan region as larger, better known airports or were previously the primary airport for a region until a newer airport was built. Global examples of secondary airports include Chicago Midway, Dallas Love, Glasgow Prestwick, London Stansted, and Frankfurt Hahn. These airports are successful because their operations are focused on being uncongested and low cost, creating a distinct geographical network not generally duplicated by full service airlines.

Aircraft

The technological advances of aircraft over the past century of manned flight have led to monumental advances in the geographic coverage of air transport. Propeller driven piston engines on the first commercial aircraft limited the range, speed, and number of passengers that the aircraft were able to accommodate. One of the first aircraft to enter commercial service was the Douglas DC 3 in 1935. The DC 3 had a capacity of 30 passengers and when full could fly at a speed of nearly 350 km/h and had a range of just over 500 km. Compare that to the newest commercial aircraft to enter service, the Airbus A 380, which depending on configuration can accommodate in excess of 555 passengers, has a top speed of 980 km/h and a range of 14 800 km, putting nearly any two points in the world within nonstop distance of each other.

The technical capabilities of the aircraft used by early airliners played one of the most important roles in determining network structure. Due to the limited range of early aircraft, routes were linear in nature and included a large number of intermediate stops along the way. Flights that today are measured in hours were instead measured in days. An example is the London–Cape Town route. Originally the route had as many as 27 stops and took nearly 11 days. Today, the two cities are connected with nonstop flights taking less than 12 h. As aircraft technology allowed aircraft to fly further with larger payloads, passenger network structures evolved allowing carriers to bypass the intermediate stops that were his torically important. At present there is a wide range of aircraft available allowing carriers to match the aircraft in their fleets to the particular conditions of the markets they serve.

Today's aircraft ranges can be divided into three geographic realms: regional, international, and inter continental. Examples of jet aircraft serving the regional realm include the Boeing 737 series and the Airbus 320/319 range. These aircraft have a range of less than 4000 km and are primarily used on high density regional routes such as London–Paris or New York–Chicago. Aircraft that fall into the international geographic scope are twin aisle aircraft like the Airbus 330 or the Boeing 767. These aircraft have a longer range, up to 10 000 km, and can also carry a greater number of passengers. These are used in markets such as Los Angeles–New York and Paris–Chicago. The aircraft in the intercontinental class have the ability to connect almost any two cities in the world. Examples include the Boeing 747 and the Airbus A340. The range of these aircraft is over 14 000 km and can carry in excess of 400 passengers. Popular routes for these aircraft include Sydney–Los Angeles and Singa pore–London.

The two primary views on the immediate future of aircraft development are fragmentation and consoli dation. The fragmentation approach is defined by hub bypassing, market development, and increased frequency on thin routes. Fragmentation is the preferred operation strategy of the airlines' most frequent and profitable traveler, the business traveler. The consolidation strategy focuses on being more cost effective by utilizing a hub dominated global network where major hubs are linked together through airline alliance formation.

Boeing is taking a fragmentation approach to the fu ture of passenger aircraft, based on the belief that the airlines of the world, and in turn their passengers, want an aircraft able to connect a large number of origin–destination pairs covering great distances and avoiding hub connections. This belief and the opinion that the global airline industry will continue to liberalize is leading Boeing to construct the 787, which will have a range of over 16 000 km, and is able to carry between 200 and 300 passengers.

Airbus is taking a consolidation approach with its newest aircraft, the Airbus 380. This double decker air craft is the largest passenger aircraft ever built and will be able to carry over 550 passengers up to 15 000 km. The freighter version of the aircraft has triple deck capacity. It is Airbus' belief that the airline industry will consoli date globally around a number of global hubs and the A380 will allow carriers to move large amounts of pas sengers or cargo between these points reducing con gestion in the air transport network. Airbus also believes the fragmentation strategy being pursued by Boeing is practical and has developed the A350, which has similar characteristics to the 787.

Conclusion

The diversity and dynamic nature of the air transport industry provide a rich research vein for geographers to tap and explore. The very geographic nature of the in dustry allows the viewpoints of geographers to be sought after by researchers in other disciplines and practitioners within the industry. The three sections of air transport discussed are rich with areas of geographic inquiries and as the industry continues to adjust to political, social, economic, and environmental stimulus spatial knowledge of the industry will grow along with the industry.

See also: Trade, Transport and Communications, Historical Geographies of; Transport and Deregulation; Transport and Globalization.

Further Reading

Burghouwt, G., Hakfoort, J. and Ritsema van Eck, J. (2003). The spatial configuration of airline networks in Europe. *Journal of Air Transport Management* 9, 309 323.
Dempsey, P. S., Goetz, A. R. and Szyliowicz, J. S. (1997). *Denver International Airport: Lessons Learned*. New York: McGraw Hill.
Goetz, A. R. and Graham, B. (2004). Air transport globalization, liberalization and sustainability: Post 2001 policy dynamics in the United States and Europe. *Journal of Transport Geography* 12(4), 265 276.
Goetz, A. R., Ralston, B. A., Stutz, F. and Leinbach, T. R. (2004). Transportation geography. In Gaile, G. & Wilmott, C. (eds.) *Geography in America at the Dawn of the 21st Century*. Oxford: Oxford University Press.
Graham, A. (2003). *Managing Airports: An International Perspective* (2nd edn.). London: Elsevier.
Graham, B. (1995). *Geography and Air Transport*. Chichester: Wiley.
Graham, B. (1998). International air transport. In Hoyle, B. & Knowles, R. (eds.) *Modern Transport Geography* (2nd edn.). pp 311 336. West Sussex: John Wiley & Sons.
Leinbach, T. R. and Bowen, J. (2004). Airspaces: Technology, geography, and air transport. In Brunn, S. D., Cutter, S. & Harrington, J. W. (eds.) *Geography and Technology*, pp 285 314. Dordrecht: Kluwer.
O'Kelly, M. (1998). A geographer's analysis of hub and spoke networks. *Journal of Transport Geography* 6(3), 171 186.
Rodrigue, J. P., Slack, B. and Comtois, C. (in press). *The Geography of Transport Systems.* London: Routledge.
Vowles, T. M. (2006). Geographic perspectives of air transportation. *Professional Geographer* 58(1).

Relevant Websites

http://www.airlineroutemaps.com
Airline Route Maps – an enormous collection of route maps from airlines around the world.
http://www.airports.org
Airports Council International (ACI) the international association of the worlds' airports, with the prime purpose of advancing the

interests of airports and to promote professional excellence in airport management and operations.

http://epp.eurostat.cec.eu.int

Eurostat Air Transport Data an enormous amount of air transport statistics for the European Union including statistics on airlines, airports, and countries.

http://gc.kls2.com

Great Circle Mapper – allows the user to find and map the great circle distance between any set of airports in the world.

http://www.iata.org

International Air Transport Association (IATA) the global trade organization of the air transport industry representing 265 airlines around the world.

http://www.icao.int

International Civil Aviation Organization (ICAO) an agency of the United Nations dedicated to the safe and orderly development of international civil aviation.

http://www4.passur.com

New York JFK International Airport Air Traffic – view in real time the air traffic into and out of one of the busiest air corridors in the world. Allows the user to select different flights and see approach landing and takeoff patterns.

http://ostpxweb.dot.gov

Office of the Assistant Secretary for Aviation and International Affairs (USDOT) a wealth of information on the United States air transport system including data and links to policy, airfare data, and financial information.

http://www.oag.com

Official Airline Guide (OAG) – website of the leading passenger and cargo schedule and information producing company. Databases include information on over 1000 airlines and 3500 airports.

B

Balkans

V. Rey, Ecole Normale Supérieure Lettres et Sciences Humaines, Lyon, France
O. Groza, Universitatea Al.I Cuza, Iasi, Romania

Glossary

Bektachism Mystic Sufi Moslem brotherhood originating in Anatolia, the world center of which is in Albania; major supporter of Albanian nationalism at the start of the twentieth century.
In-Between Between-Ness A space where developments are governed by processes of outside interference and influence that are stronger than inside forces. Its history is composed of recurrent adversity, while lack of durability prevents accumulation and causes return to previous states, contributing to a cyclic, discontinuous experience of time that does not favor integration of what has been previously acquired.
Karst It refers to the various forms of geographical relief connected with the action of water in limestone terrain. It originates from the name of a region in the Dinaric area of Slovenia.
Polje A large, closed karstic depression. The term is Serbo-Croatian and refers to a plain.

The Meaning of the Word ‘Balkan’, the Border Issue

Defining what the Balkans are is a complex question from the outset, because of the interweaving of fact and representation. *Balkan* is a Turkish word meaning ‘mountain’ referring to the Stara Planina, the east–west mountain range that cuts across Bulgaria. The German geographer Zeune used the term *Balkan peninsula* to identify the eastern Europe peninsula that extends into the Mediterranean. The northern limit is conventionally formed by the Danube and the Drava, the two rivers that do not possess particular geographical significance. From a historical viewpoint, the Balkan space corresponds to the area of Europe that has been culturally impregnated by the Ottoman domination. References to the Ottoman past appeared at the end of the nineteenth century when wars arose between local peoples seeking to establish their territories in areas left by the withdrawal of the Ottoman occupation. The term ‘balkanization’ became synonymous with the geopolitical instability resulting from lasting fragmentation into small, unstable, hostile political units, involving the risk of war for an area reaching beyond the conflict zones. Balkanization has become a generic term with a negative connotation that the inhabitants of the region dislike, and it is at present the subject of critical examination of how mental representations are constructed and deconstructed, involving consideration of the negative ‘invention’ or projection of ‘otherness’, possibly characteristic of Western thought in the construction of its civilization of domination.

The terms Balkan and balkanization in relation to geopolitical realities and mental representations have had a contrasted history in the last half century, disappearing during the Soviet phase, and reappearing after 1990: following the geopolitical demise of Eastern Europe, Central and Balkan Europe recovered their respective identities, and the wars in Yugoslavia reactivated the negative connotation of the word Balkan, which it is more appropriate to analyze than to criticize (**Figure 1**).

In a regional geography approach, the question of delimitation is a means of encompassing identity, the spatial core, and the blurred fringes. The central core articulates around Belgrade, Sarajevo, Tirana, Salonica, and Sofia, with its predominantly continental and mountainous structure; the northern fringes are the Slovenian Alps, the middle and lower Danube basin, and to the south maritime Greece and the Turkish area of Thrace. But the particularly marked fluctuations of borders suggest that it is important to distinguish between the geographical approach and the geopolitical approach, a key to the volatile and dangerous configurations of the region.

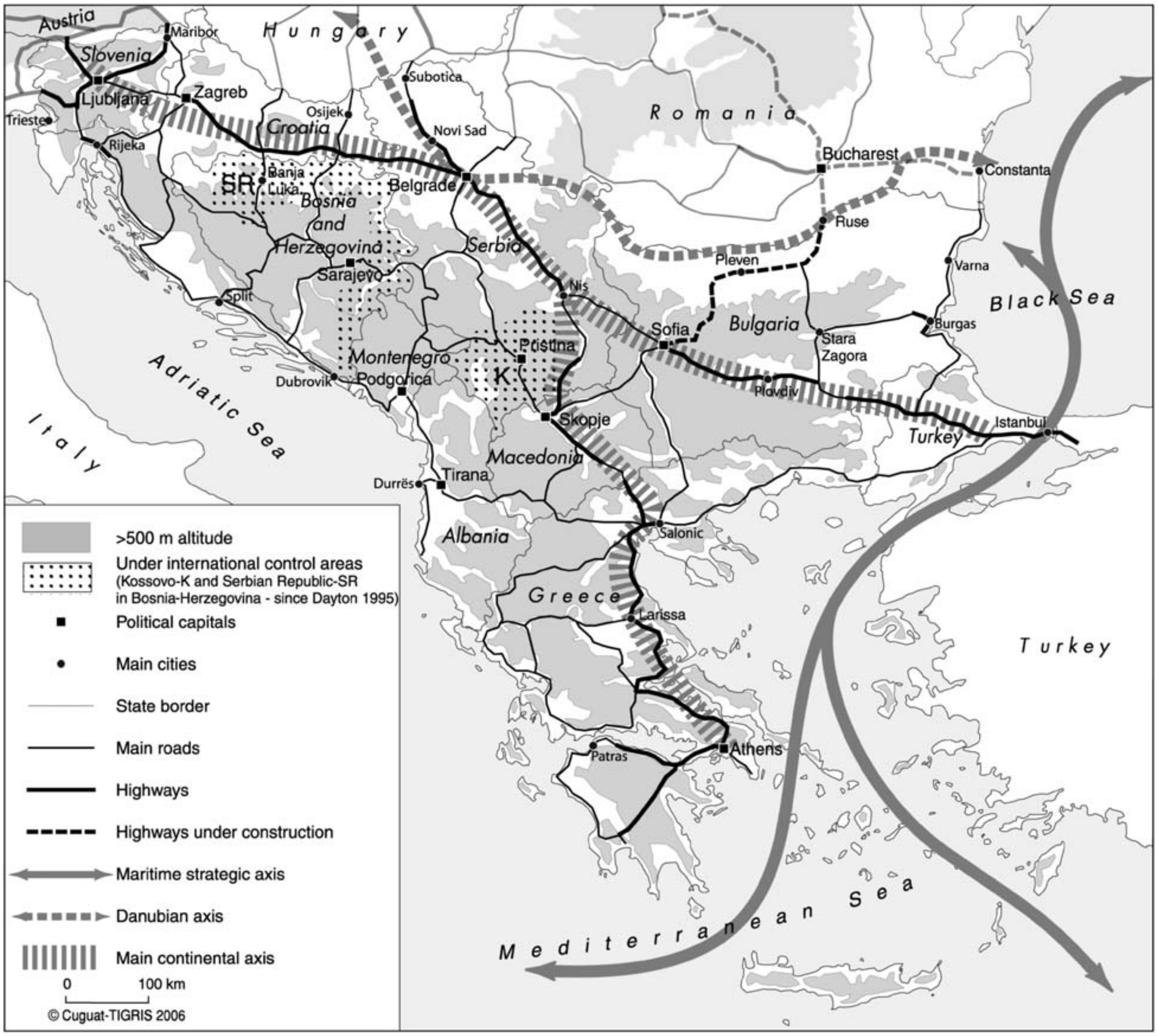

Figure 1 Geographical structures and major roads.

A Geopolitical Region Where Major Powers Meet: The East, the West, an 'In-Between' Europe

On 1 January 2006, the peninsula comprised eight states: Greece, Albania, Bulgaria, and, resulting from the dismantling of Yugoslavia, Macedonia, Serbia Montenegro, Bosnia Herzegovina, Slovenia, and Croatia. Turkey also has a fragment, and its first city, Istanbul. Romania is also sometimes classified as a Balkan state. The decomposition of what was the Yugoslav territory until 1991 is still underway: Bosnia Herzegovina has been under international mandate since 1995, Montenegro became independent in May 2006 and Kosovo is gaining for the same status on 17 February 2008, but not yet recognized by the whole international community. At each stage, the Serbian state is shrinking. There is on the one hand a discrepancy between the geographical distribution of the peoples who are claiming recognition as nations and the state borders, and on the other hand an exceptionally marked fluctuation of these borders, and these two phenomena focus on the 'in between' spatial mechanisms that have modeled the Balkan states from the outset.

Two Millennia to Form a Linguistic, Ethnic, and Religious Mosaic

If we return to antiquity it can be seen how the superimposition of civilizational strata is still influential in contemporary geopolitics (for instance, the Greek objection to the name of Macedonia). In antiquity the peninsula was divided between the influence zones of Rome and Constantinople, the first manifestation of the confrontation between West and East. This division is rooted in the languages and their alphabets, Latin to the West, Greek to the East; it was later reactivated by religious identity, when Latin and Greek (and later Slavonic) became the liturgical languages of the two Christian churches. On this cultural dividing line, stable over two millennia, religious competition was intense (**Figure 2**).

From the sixth century Slav migratory flows brought new populations, but in a territory that was already markedly 'humanized' by groups who maintained their languages (Greek and Albanian): the first kingdoms of Bulgaria and Serbia had to gain settlement space by opposing the Byzantine Empire.

The long Ottoman episode introduced new population patterns, such as Turkish populations, populations converted to Islam and acting as soldiers on the marches of the Empire (Slav Mohammedan Bosnians, Pomaks Bulgarian Mohammedans, Albanian janissaries, etc). From the eighteenth century, the Hapsburg Empire produced similar clusters of peasant soldiers (Serbs in Croatia, villages of all nationalities in Vojvodina). With the fluctuations of the borders between these imperial fronts, major migrations created shatterbelts with their unstable populations. Sephardi Jews found refuge in the Ottoman Empire, which was open to these city dwelling populations with talents for complex social and economic organization. The linguistic, ethnic, and religious mosaic that underpinned social and territorial patterns was the result of an intertwining of outside influences and the confrontation between large powers whose centers were not within the peninsula. In each episode the 'in between' mechanism produced breaks that further fragmented the Balkan area.

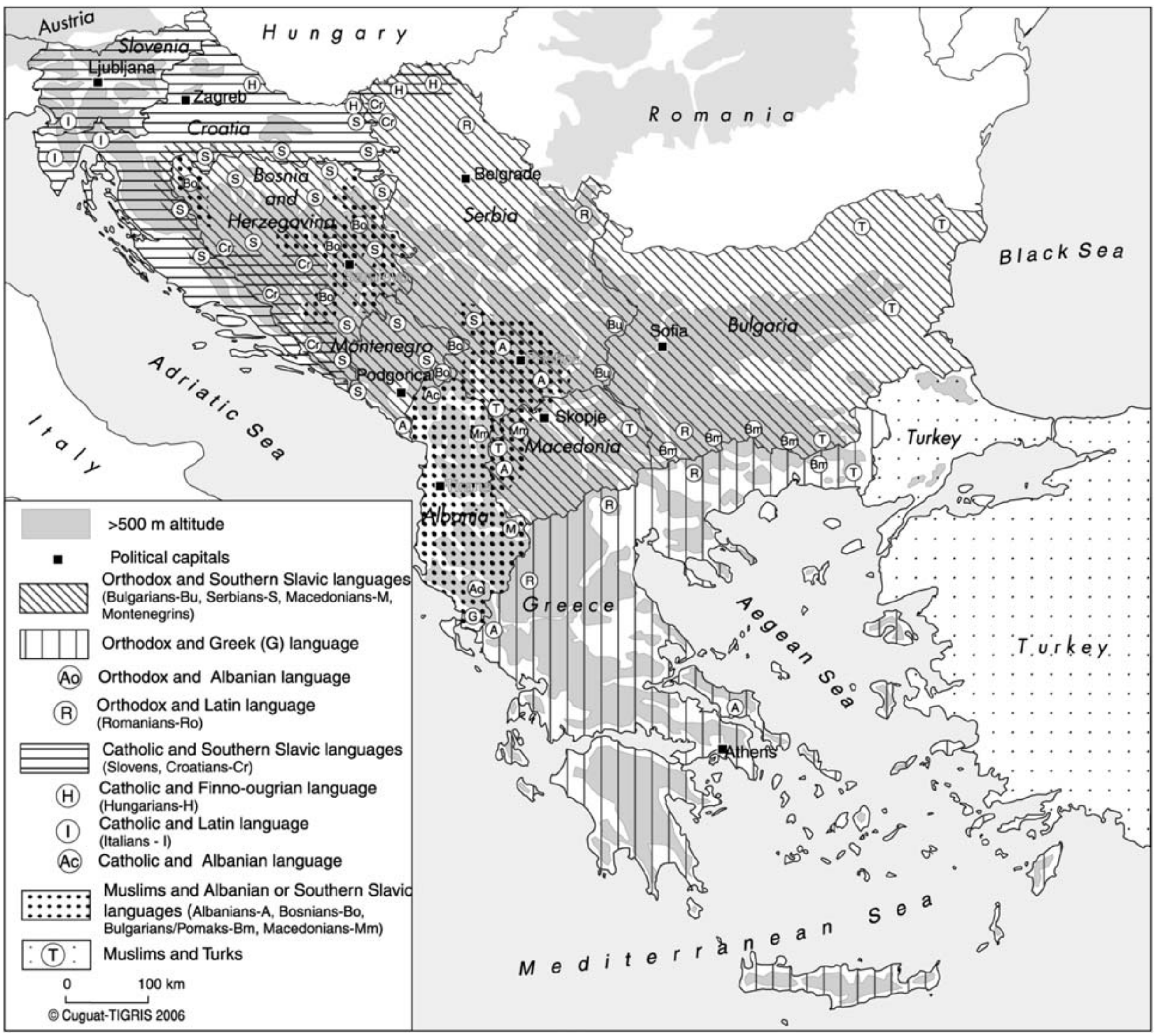

Figure 2 Religions and languages in the Balkans.

The Dilemma of Configurations on the Nation-State Model

The political map of the nation states that became es tablished in the course of the nineteenth century is a major event with the territorial configuration of the peninsula, which thus far had been managed on imperial principles. This resulted from the weakening of the Ot toman power, and from the ambitions of neighboring Hapsburg and Russian empires, which were aimed at the peninsula and the control of the Bosphorus. It was also the result of the rise of national ideologies, with the support of Western powers, which initiated awareness and 'inventions' of nation. To escape from the Ottoman system, in which social distinction was functional, na tional construction was founded on language and litera ture, substituting the ethnic category for the social hierarchy. The modern nation states broke away from the imperial territory, Greece in 1830, Serbia in 1878, joined by the provinces of Croatia and Slovenia in 1921 (pro ducing Yugoslavia), Bulgaria in 1878, and Albania in 1913. The creation of these four states brought a new group of actors onto the geopolitical map, with their problems as newly independent states (hiatus between nationality and nation). These creations did not give in dependence to all the communities involved. They also needed political institutions recognizing a certain degree of individuality for the minorities (language, independent local administration, etc). It is in the setting of this dif ficult articulation between state and community legit imacies that conflicts developed successively. The 'civil' wars were mainly the result of internal tension, even if outside powers intervened. Yugoslavia, broken up into six states in 2006, is the typical example: neither the unitary configuration (1921–1940) nor the federal configuration established by Tito to counterbalance Serbian hegemonic ambition (1944–1991) was successful.

The Major Powers Still Present in Balkan Europe on the Borders of the European Union

Competition for influence over these new states persisted throughout the twentieth century: the sparking of World War I by the Sarajevo assassination in 1914; the Yalta agreement in 1944 which put the whole of the peninsula under Soviet control, except for Greece and Turkey which became NATO bases, signaling the arrival of US influence; then, after 1945, the interplay of influence of the major foreign powers within the then communist zone: Tito's Yugoslavia found Western support in his conflict with Moscow, turned to the Arab countries and initiated an opening up of the 'non aligned' countries in the Third World; Albania used China to escape from Russian domination.

The end of the Cold War saw a very marked re activation of Western influence in the Balkans, and a return to former historical alliances. Germany and Aus tria, retrieving their areas of influence, supported the breaking away of Slovenia and Croatia from Yugoslavia in 1991; although Slav, these two countries had only shared a short period of the past with other Slavs in the south in comparison with their long period of belonging to the central European empire. Russia, although weak ened by the collapse of the Soviet system, reestablished its cultural influence by using its affinities, both linguistic and religious, in Bulgaria and above all Serbia, where it supported the military nationalism. The position of Moslem countries toward Moslem minorities has been complex: it has been moderate, targeting the political and economic fields for Turkey, where the priority has been to obtain entry into the European Union (EU); it has been religious and ideological for the Arab countries, which have financed infrastructures (mosques, schools, and preachers) and the creation of the Eastern European Islamic Council as early as 1991. It was outside military intervention, NATO and then UN forces and the EU (ASA – agreement on stabilization and association, 2003) that put an end to the conflict surrounding the break up of Yugoslavia; US influence partly explains Kosovo's unilateral independence.

Because the Balkans is an 'in between' space, they are areas where geopolitical transition takes the form of ab rupt and violent territorial reorganization.

Balkan Civilization – Similarity Is Not Synonymous with Cohesion

The Balkan civilization is difficult to apprehend on account of the negative connotations of the term balkani zation. When endeavoring to understand the fragmen tation that is the main characteristic, there are two obvious pitfalls: determinism of the natural environment and determinism of history and culture, including repeated violence between groups that are identified sometimes by lifestyle, sometimes by position with respect to political authority, and sometimes by a sense of ethnic, national, or religious belonging. What is left today of the social organization considered typical to the Balkans, after the sudden arrival of modernity imposed by the communist system (Greece excepted), and the upheavals of post communism?

Physical Diversity, Underpinning Human Diversity

The fragmented relief of the Balkan Peninsula is difficult to penetrate. Between tectonic plates, it is a geological mosaic with scattered deposits of rare polymetals. The relief divides the area into small regions: basins, gorges, large lakes, corridors (Vardar Morava and Maritsa), the enormous Dinaric *karst* area with its *polje* and caverns:

these exceptional natural features produce small human units lacking easy communication with one another, and favoring withdrawal into representations of identity and stronghold status. Since invaders arrived up the valleys, the mountains were more densely settled than the plains, and this reversal of settlement patterns only disappeared during the twentieth century.

The peninsula is Mediterranean on its coastal fringes, while the interior is characterized by continental climate, with marked rainy seasons and high temperatures. There is severe erosion and long standing deforestation. Transhumance practices increase risks of flooding, and deposition reduces infiltration downstream, generating marshy plains that have long been infested with malaria. Irrigation in depressions has enabled intensive agriculture, favoring crops suited to hot climates which are the specialty of Bulgarian and Turkish farmers. Farming settlement was traditionally concentrated in the *podgora* area (piedmont). Here, villages of homogeneous ethnic and religious composition are juxtaposed, forming a diverse territorial mosaic. Because of the relief, crossroads and traffic corridors acquire fundamental strategic importance and localize cities, often of imperial origin: in the interior, Belgrade, Sarajevo, Skopje, or Plovdiv, and on the coasts, giving maritime access, Rijeka, Split, Durrës, Athens, Thessalonica, Burgas, and Istanbul as the gateway to Asia. The geographical distribution of the centers and their terrestrial communication links give a fairly uniform mesh with no central pole.

The bimodal nature of settlement patterns – towns and cities on the natural traffic routes, scattered village settlements away from the roads – constitutes the hidden paradox of the Balkans, suspended between the potentialities of communication intersections and very real partitionings. The relatively similar structural patterns across the peninsula and the absence of a center phenomenon are obstacles to any integration process.

Social Systems: Permanence and Mobility, Villages, and Towns

The Balkans are readily viewed as museums of archaism and ideal hunting ground for ethnologists and folklore enthusiasts. Within Europe, they stand apart on account of social structures that have remained markedly organized around the extended family and the pre eminence of men. Family and religious celebrations, which are widely observed, punctuate the community year. Repeated changes in border demarcation, displacement of groups to form peasant soldier colonies on the imperial marches, as well as pastoral movements over large distances from mountain to plain (Vlaks), traveling traders, and massive returns (Greeks from Anatolia) have all contributed to forming a social relationship with space that carries the mark of mobility. Network relationships and the sense of belonging to a group have the upper hand on belonging to a territory – across distance, solidarity, whether empathetic or economic, plays an essential part. Contemporary economic migration perpetuates this link, known as the diaspora culture. The construction of a state based civic identity is hindered by its existence. The energy expended in removing all trace of previous occupiers to discourage any return is a sign of it.

This characteristic is shared by rural and urban populations alike, even if the latter are of more cosmopolitan composition. There are, on the one hand, ports and trading towns and cities on the Mediterranean, which are better interconnected than with the hinterland (Dubrovnik Ragusa). On the other hand, there are the cities that the Ottoman power developed into a network supporting the garrisons, administrations, and craft activities (Sarajevo, Skopje, Thessalonica, and Plovdiv). Their populations were multiethnic, as each trade was related to an ethnic group in a given quarter, with strong links from one city to another, which favored diaspora type identification. Only those cities that became capitals when national states were created (Sofia, Tirana, and even Athens) have been characterized by reference to national identity.

Religions: The Main Identity Substrate

Religious identity sets the Balkans apart in Europe and segments populations within the area. Although religion does not contribute any cultural determinism any more than the geographical environment or history, it is a major component of identity. Oriental Orthodox Christianity covers most of the peninsula. The Roman Catholic church has a foothold only in Slovenia and Croatia. The antagonism between the two branches of Christianity, which formed in the course of the first millennium, is used by historians to explain the moderate resistance toward the Turks of Byzantium, to avoid falling under the papal authority, felt to be more of a threat to identity than submission to the Ottoman Porte.

The orthodox religion has no supreme authority. The autocephalous organization of religion is the ancient matrix of the Balkan State. The Ottoman Empire increased the secular strength of the orthodox religion by entrusting the patriarchs with responsibility for the Christian communities, so that Greeks, Serbs, or Bulgarians were culturally more than religiously orthodox. Religious and political powers supported one another, sharing organic functions. Spiritual life became rooted in the numerous monasteries scattered across orthodox territory, which were very autonomous since they were not organized into orders. Thus the orthodox component has a paradoxical structure, whereby the theocratic model of the sociopolitical whole is manifested in a scatter of communities and territories.

Islam concerns around 8 million inhabitants: they are mainly native Balkan populations, Albanian, Slav (and later Rom) – its first feature. The second concerns location: the Islamic populations occupy the front zone between the Catholic and Orthodox areas, from the mountains of Bosnia (formerly a Bogomil zone) to Albania and Macedonia in the north, and also the Rhodope Mountains and the Dobruja district in Bulgaria. The third feature is the diversity of Balkan Islam, in particular the Bektachi group in Albania. There is some degree of organizational similitude between the two ecumens, Orthodox and Moslem, in particular the nonseparation of religion from politics, which deprives the present democratic construction of a previous experience of independent institutions.

When apprehended in this manner, the basis and configuration of Balkan culture yields complex geographical constructions, which belie the 'essentialist' representation that is often attached to the notion of culture.

Modernity and Development

The Legacy of Delayed Development

The scale of the differences between West and East in the Ottoman period is widely considered to be the reason for the present lack of development of the peninsula; the front line between two civilizations is seen as having acted as a barrier, retarding the diffusion of the innovations of the modern world initiated by Western Europe. In addition, in the absence of large deposits of iron and coal, Western capital has solely funded port and railway infrastructures. The determination to obliterate the Ottoman cities by modern geometrical urban layouts expresses a will to 'europeanize' (**Table 1**).

The communist system aimed to implement modernization on a different basis. It made a break with the long dominant, very poor rural societies. It succeeded in initiating a remarkable economic rise on an industrial base, and introduced urbanism targeted on the medium sized towns and cities. The present postcommunist crisis has somewhat obscured the progress achieved in living standards in Balkan populations up to the 1980s. However, as the system was based not so much on exchanges and complementarities as on making good use of scattered resources *in situ*, it contributed to perpetuating spatial partitioning and constituted an obstacle to competitiveness – all the economies were experiencing severe difficulties by 1989.

Socioeconomic Disasters and Spatial Trends

The identity crisis and the systemic postcommunist transition crisis are mutually supportive. The economies have collapsed; infrastructures have deteriorated as a result of the virtual disappearance of the means of the

Table 1 Social and economic indicators of the Balkan states: Inequalities of development

Countries Capitals	*Area (km^2)*	*Population inhab. (1991)*	*Population inhab. (2002)*	*Urban pop. (‰)*	*Natality rate (‰)*	*Mortality rate (‰)*	*Infantile mortality (‰)*	*HDI*	*GDP/ inhab. $*
Slovenia	20 200	2 000 000	2 000 000	48	9	10	4	0.895	10 400
Ljubljana		*330 000*	*259 000*[a]						
Croatia	56 500	4 760 000	4 300 000	59	10	12	7	0.830	4540
Zagreb		*737 000*	*740 000*[a]						
Bulgaria	111 000	8 990 000	7 500 000	68	9	14	14	0.796	1770
Sofia		*1 107 000*[a]	*1 092 000*[a]						
Greece	132 000	10 000 000	11 000 000	61	9	11	5	0.902	11 600
Athens		*3 010 000*	*3 188 000*						
Albania	28 700	3 360 000	3 100 000	44	17	6	22	0.781	1450
Tirana		*244 000*	*343 000*[a]						
Bosnia H	51 100	4 370 000	3 900 000	44	12	8	15	0.781	1310
Sarajevo		No data	*552 000*[a]						
Macedonia	25 700	2 030 000	2 100 000	60	14	9	22	0.793	1710
Skopje		*563 000*	*485 000*[a]						
Serbia-Montenegro	102 000	10 630 000	10 700 000	52	12	12	16	No data	1400
Belgrad		*1 229 000*	*1 281 000*						
Montenegro	No data	No data	650 000						
Podgorica		*115 000*	*132 000*						
Kosovo	No data	No data	2 000 000						
Pristina		*184 000*	*292 000*						
Total countries	525 200	46 140 000	44 600 000						

[a] Districts excluding suburbs. **HDI**: human development indicator; **GDP**: gross domestic product.
Sources: National Yearbooks.

state; liberal privatizations have led to decomposition, thus facilitating the rise of informal economies and corruptions, reducing the influx of the foreign investments that are required for modernization; debt and extreme dependency on outside aid have become structural features. Severe poverty has taken a strong hold, therefore affecting a quarter of the population; it has led to massive (often illegal) emigration, temporary or otherwise. The backdrop is a longstanding, worsening demographic decline, the accentuation of density contrasts, and whole zones devastated by war. The rural areas, sometimes abandoned, face ageing populations and declining agriculture. The large cities attract rural populations and those displaced by conflict; their social composition change and is again segregated.

There is a degree of economic development under international control concerning the construction of housing, the reopening of mining industries and services; it urgently requires road communications to reinforce regional links. Because there is a lack of local energy resources, the challenge will be crucial, requiring common policies on international energy infrastructures. Any economic future also depends on good use of the major tourist potential.

Political Decomposition and Fragile Territorial Recomposition

Three regional subunits can be distinguished, according to their different trajectories, how far they have opened up to the outside, and how far they enter into integration process with EU.

In the northwest, Slovenia and Croatia are close to Western Europe, favoring development, and also entertaining relations with Central Europe. Slovenia, which has Alpine affinities and is culturally homogeneous (with 88% Slovenes), has bolstered its economy, which was 8 times stronger than that of Kosovo in 1989. Croatia possesses part of the Pannonian agricultural plain, the north of the Dinaric Alps and the Dalmatian coast (tourism). When it left Yugoslavia, Croatia became more homogeneous (from 78% to 90% Croatian between 1991 and 2001), but with an economy that has suffered from the aftermath of war and from its standby status for entry into the EU.

In the east and the south, the main feature is the giant Istanbul which has returned to its position as an economic metropolis for the peninsula. Bulgaria and Greece are both neighbors of Turkey, thus having an immediate horizon opening onto the Oriental and Russian zones, and onto the geopolitical and economic organization of the Black Sea (pipelines). Bulgaria, lacking sufficient impetus of its own, has lapsed from relative prosperity into a major crisis. Its spatial organization is on the center–periphery model, inherited from the construction of the modern state. Opposition is growing between the declining outer fringe areas and the central zone (Sofia region, Thrace depression, and the tourist coast), where most of the population, activity, and foreign investments are to be found. Greece, which has been a member of the EU since 1981, is regaining its Balkan dimension after having put its stakes on the Mediterranean, with international tourism and the merchant navy (ranked number one in the world) as its two economic mainstays. It is in the process of developing its role as a geopolitical power relay, growing closer to Turkey. After being a country characterized by emigration, it has now become a destination for post socialist immigration from the peninsula, often illegally. After a very marked period of center–periphery organization, the patterns of distribution of activities have improved, mainly on the north–south Thessalonica–Athens axis. Athens attained the status of a major metropolis with the Olympic Games in 2004.

Albania, Macedonia, and the fragmented Serbia make up the Western Balkans, under the cloud of tragic tension between Moslem groups and the Orthodox Serbian pole, compounded by poverty. The issue is how to construct a political order of Sovereign states. Should the way forward be the ethnic solution, or some form of multiethnic cohabitation?

The breaking up of the Yugoslav political entity is an enduring process, for lack of a strategy for recomposition, and on account of the very marked ethnic interweaving. Macedonia, independent since 1992 and the key to circulation within the peninsula, has been paralyzed by tension between Macedonians (64%) and Albanians (25%) following the uprising in 2001. Montenegro, which shares language and religion with Serbia, achieves independence in 2006 to negotiate its own entry into the EU, which it will thus deprive Serbia of its access to the Adriatic. The population is reputed to be homogeneous, yet only 43% state that they are Montenegrin. Tourism is its only asset. Bosnia Herzegovina, after 3 years of war and 11 years of international supervision, still lacks an effective political solution. With three nations – Bosnian Slav of Moslem culture, Croatian, and Serbian – and two political entities – the Croatian Moslem federation and the Serbian Republic that desires integration with Serbia – this federation does not function according to a multicultural logic, even if there are signs of a new beginning. Serbia has been enclosed within the Danube horizon, with the highly multiethnic Vojvodina province and the region dominated by the agglomeration of Belgrade.

Kosovo, densely populated, is soon to become exclusively Albanian. Till 2008, part of Serbia, but in fact completely autonomous, is awaiting recognition by some states; its example holds the key to the political future of the whole area. Albania is the youngest and poorest country in Europe. Corruption, organized crime, and funds from the 700 000 emigrants (a quarter of the population) are the basis of the economy, and the EU does not expect to integrate it before 2020. The center–periphery phenomenon is acute in the central plain around Tirana.

The processes of homogenization (for Croatia, Serbia, and Kosovo) and of diversification (for Macedonia, Montenegro, and Greece) are continuing: the political miniaturization that results from identity grouping is not simple to represent. The extreme contrasts in demo graphic dynamics is one of the keys to explaining the hiatus between representations of rights linked to iden tity established in the past, and the reality of population numbers in each country. On the one hand there are populations of Moslem culture that are prolific and very young, and on the other are ageing Slav populations suffering from major demographic decline. Whatever the debate on the causes of the contrast – economic and/or cultural – the result is a reversal of the relative weight of the two groups, which completely alters the terms of recomposition and requires finding new ways of living together. After the exacerbation of conflicts rooted in identity, poverty, and inequalities are bringing develop ment issues to the fore.

What can be the future for so many tiny countries where the hostility of the neighbors and the small vol umes of proximity exchange predominate? In a world where globalization is the order of the day, the handicap of small size will probably not be as great as it was in the previous century, provided that these countries can move beyond the autarchy model, toward a logic of inter dependency on a democratic foundation.

See also: Borderlands; Ethnic Conflict; Ethnicity; Geopolitics; National Spatialities; Nationalism; Religion/Spirituality/Faith.

Further Reading

Allcock, J. B. (2002). Rural urban differences and the break up of Yugoslavia. *Balkanologie* IV, 101 125.
Bjelic, D. and Savic, O. (2002). *Balkan as Metaphor. Between Globalization and Fragmentation*, 382p. Cambridge, MA: MIT Press.
Brunnbauer, U. and Pichler, R. (2002). Mountains as "lieux de mémoire," highlands values and nation building in the Balkans. *Balkanologie* IV, 77 100.
Carter, F. W. and Norris, H. T. (1996). *The Changing Shape of the Balkans*. Boulder, CO: Westview Press.
Cvijic, J. (1918). *La Péninsule Balkanique. Géographie Humaine*. Paris: Armand Colin.
Hadjimichalis, C. (2004). The changing geography of Europe: A Balkan and southern European view. *European Urban and Regional Studies* 11, 363 366.
Hall, D. and Danta, D. (1996). *Reconstructing the Balkans. A Geography of the New Southeast Europe*. Chichester: Wiley.
O'Thuatail, G. and Dahlman, C. (2004). The effort to reverse ethnic cleansing in Bosnia Herzegovina: The limits to returns. *Eurasian Geography and Economics* 45(6), 439 464.
Petrakos, G. (2002). The Balkans in the New European economic space: Problems of adjustment and policies of development. *Eastern European Economics* 40(4), 6 30.
Populations in the Balkan Regions (2004). Espace Populations Sociétés, no. 3, pp 459 693.
Poulton, H. (2000). *Who Are the Macedonians?*, 226p. London: Hurst and Company.
Prévélakis, G. (1994). *Les Balkans, cultures et géopolitique*, 220p. Paris: Nathan.
Rey, V. (1996). Europes orientales, Europe balkanique. In Brunet, R. & Rey, V. (eds.) *Géographie universelle: Europes orientales, Russie, Asie centrale*, vol. 1, pp 8 156. Paris: Belin Reclus.
Todorova, M. (1997). *Imagining the Balkans*. New York: Oxford University Press.
Transitions balkaniques (2004). Méditerranée, no. 103, 104p.

Relevant Websites

http://www.afebalk.org
Association Française d'études sur les Balkans.
http://www.ssees.ac.uk
Centre for South East European Studies, UCL School of Slavonic and East European Studies.
http://www.brill.nl/seeu
Journal Southeastern Europe.

Barnes, T.

S. Reimer, University of Southampton, Southampton, UK

Trevor Barnes (**Figure 1**) was born in London, England in 1956. Having grown up in Cornwall, he studied economics and geography at University College London between 1975 and 1978. Barnes completed MA and PhD degrees in geography at the University of Minnesota and since 1983 has taught at the University of British Columbia in Vancouver, Canada. He was promoted to full professor in 1994. Barnes has been awarded a wide range of professional distinctions including the American Association of Geographers' Presidential Award for Professional Achievement in 2006. Barnes' work extends across theories of economic value; analytical political economy; and flexibility and industrial restructuring. Later he has explored the 'theoretical histories' of Anglo American economic geography. Drawing upon insights from science studies, Barnes has critically interrogated the changing position of economic geography through the twentieth and into the twenty first century.

Although perhaps giving the appearance of a relatively divergent set of themes, there are strong threads of continuity running through Barnes' research and writing interests. He has long been captivated by the work of the Italian economist Piero Sraffa, for example. In Barnes' view, Sraffa's terse expositions on value in *The Production of Commodities by Means of Commodities* (via a set of simultaneous production equations) usefully speak to both rigorous abstract theorists and scholars who are more interested in the contextual and the concrete. Barnes views antiessentialist accounts such as Sraffa's as a useful means of critiquing both classical Marxist accounts of the labor theory of value (which essentialize the role and nature of labor power) as well as neoclassical utility theory.

Figure 1 Photo of Trevor Barnes.

Together with Eric Sheppard (University of Minnesota, US), Barnes sought to ground political economy in space and place by developing analytical approaches. Engaging with, but also building a substantial critique of analytical Marxism, such approaches employ statistical testing, mathematical reasoning, and logic to understand the production and functioning of socioeconomic geographies. Although Barnes and Sheppard's work employs formal mathematical language, it does not deal solely with the role of space and place in the abstract. Rather, they would suggest that analytical approaches can be useful in assessing and understanding concrete and contextual worlds.

Barnes' engagement with debates surrounding flexible production has drawn upon research on the forestry industry in British Columbia conducted with Roger Hayter (Simon Fraser University, Vancouver, Canada). While many accounts of flexibility through the 1980s and 1990s centered upon developments in 'new industrial spaces', Barnes and Hayter extended conceptualizations of flexibility through a consideration of *in situ* restructuring within a marginal resource economy. Barnes also has drawn upon and reworked the Canadian economic historian Harold Innis' 'staples model'. The staples model emphasizes the key role of staples production for export within Canadian economic development, which gives rise to relations of dependency and marginality within the international economy. Throughout his writing in this area, Barnes' explicit concern has been to confront the profound devastation of lives and communities wrought by the decline of the British Columbian forest products sector. Most recently, Barnes and Hayter have reiterated that 'local models' such as Innis' – which explicitly reject unilinear causal logics, universalist assumptions, and formal reasoning – are vital to understanding the complex socioeconomic geographies of resource peripheries.

From the late 1990s onwards, Barnes became interested in tracing the social and political connections that produced the spatial scientific narratives which came to dominate geography – and particularly economic geography – during the 1950s and 1960s. He has suggested that one of the most important aspects of the so called quantitative revolution was not in fact its use of new methodologies, but rather the impact it had upon dominant theoretical sensibilities. In fact, Barnes argued that geography had been a quantitative discipline since the nineteenth century. While the practices of geography might have changed in the mid twentieth century (in cluding computerization, the use of complex statistical methods, and new 'scientific' vocabularies), the most significant impact of the quantitative revolution for Barnes lays in its explicit championing of foundational understandings of the world in which the 'truthfulness of representation' was guaranteed.

In researching connections between and among spatial scientists in North America, Barnes reflected upon the socially embedded nature of geography's quantitative revolution. Crucially, transformations in geographical thinking emerged as 'local affairs' within particular in stitutional sites. Locational analysis, for example, became dominant precisely through the continuing use and transmission of ideas via teaching and research in par ticular places. Informed by 'philosophy of science' lit eratures, Barnes argues that there is no universal principle which shapes the development of scientific knowledge, but rather our understandings of the world are intricately connected to the local context in which geographical ideas develop. His interest in the nature of geographical knowledge production derives from a desire to be conscious of the social power and interests that shape such knowledge. Importantly, geography re searchers' involvement in the academy quickly becomes inextricably bound up with networks of power and ar rangements of competing interests.

One of the key contributions provided by Barnes' examination of the histories of economic geography is his provision of a more nuanced story of the discipline than narrations of strict succession and progression of know ledge generally imply. The notion of a quantitative 'revolution' in geography obviously suggests a move be yond preexisting theoretical perspectives – and certainly post spatial science approaches such as Marxism, femi nism, locality studies, and accounts of flexible production represent attempts to challenge previous ways of think ing. Nonetheless, Barnes suggests that economic geog raphy through the 1970s and 1980s did not in fact move far beyond an Enlightenment view which sought cer tainties and foundations. Despite seeking to distance themselves from both the language and the practice of spatial science, he argues that most analysts ultimately were unable to escape the legacy of the seventeenth century.

Barnes has exhorted researchers to strive for a more edgy engagement with economic geography. Rather than pursuing a singular order, he has suggested that – like favorite punk bands Sham 69 and the Sex Pistols – economic geographers should engage in messy, energetic, and multiple approaches. Commentators should seek to disrupt, rework, and even work 'against' existing styles of thought. Discrete conceptual categories such as capital, labor, culture, economy, or 'the consumer' similarly should be interrogated. Barnes' work has contributed substantially to the reconfiguration of economic geo graphical approaches in ways which seek new theoretical understandings of space and place but which also allow for experimentation and the search for new possibilities.

See also: Anglo-American/Anglophone Hegemony; Economic Geography; Quantitative Revolution; Resource Industries; Social Studies of Scientific Knowledge.

Further Reading

Barnes, T. J. (1989). Place, space and theories of economic value: Contextualism and essentialism in economic geography. *Transactions of the Institute of British Geographers* 14, 299 316.

Barnes, T. J. (1996). *The Logics of Dislocation: Models, Metaphors and Meanings of Economic Space*. New York: Guildford.

Barnes, T. J. (2001). Lives lived and lives told: Biographies of geography's quantitative revolution. *Environment and Planning D: Society and Space* 19, 409 429.

Barnes, T. J. (2001). Retheorising economic geography from the quantitiative revolution to the cultural turn. *Annals of the Association of American Geographers* 91, 546 565.

Barnes, T. J. (2002). Never mind the economy. Here's culture. Economic geography goes punk. In Anderson, K., Domosh, M., Pile, S. & Thrift, N. J. (eds.) *Handbook of Cultural Geography*, pp 89 97. London: Sage.

Barnes, T. J. and Duncan, D. S. (eds.) (1992). *Writing Worlds: Texts, Discourses and Metaphors in the Interpretation of Landscape*. London: Routledge.

Barnes, T. J. and Faris, M. (2006). Between regions: Science, militarism, and American geography from World War to Cold War. *Annals of the Association of American Geographers* 96, 807 826.

Barnes, T. J. and Hayter, R. (1992). The little town that did: Flexible production and community response in Chemainus, BC. *Regional Studies* 26, 647 663.

Barnes, T. J. and Hayter, R. (2005). No Greek letter writing: Local models of resource economies. *Growth and Change* 36, 453 470.

Sheppard, E. and Barnes, T. J. (1990). *The Capitalist Space Economy: Geographical Analysis after Ricardo, Marx and Sraffa*. London: Unwin Hyman.

Beaujeu-Garnier, J.

M.-C. Robic, CNRS, Paris, France

Jacqueline Beaujeu Garnier (1917–95) was the first woman in France to obtain a *Doctorat d'Etat* in geography (1947) and the first female geographer to become a university professor (in 1948 at Lille and from 1960 to 1986 in the Sorbonne). She was also incredibly active in numerous academic institutions and editorial bodies in French geography from the 1960s onward. An honorary member of the International Geographical Union (IGU) from 1988, she owed her reputation in France and in the international community to her never tiring curiosity about the changes taking place in the world, her remarkable energy, and her very great openness of mind toward both ideas and people. She was intellectually pragmatic, and also demonstrated the value of articulating scientific research with spatial planning, so that her work this field earned her the highest official recognition in France.

Beaujeu Garnier was very representative of her generation of geographers: she followed a scientific career in which she explored issues ranging from geomorphology to regional and human geography, and she soon found herself confronting issues of regional planning. Promoting goegraphy was, throughout her life, the horizon of her university functions and of her initiatives in the community field. She did not, however, enter the controversy of the 1960s between the defendants of applied geography such as Michel Phlipponneau and Jean Labasse, and the proponents of a critical, antitechnocrat and antiliberal perspective, particularly advocated from a Marxist viewpoint by Pierre George. Thus, starting from the experiences of modernization of a former industrial region and planning in the Paris metropolitan area, she undertook the creation of a series of regional atlases of France. She succeeded in assembling teams of geographers in this task, and obtained funding from the *Délégation à l'Aménagement du Territoire et à l'Action Régionale* (DATAR; French planning authority created in 1963). On this impetus, she supported the collaboration of geographers with territorial bodies, administrations, public statistics departments, and young researchers engaged in contract research. Despite the competition from other areas such as economics, sociology, urbanism, and other professional corporations, in particular the civil engineers trained in the French *Grandes Ecoles* (University equivalent higher education in France), Jacqueline Beaujeu Garnier contributed to legitimizing geographical, regional, and thematic approaches to spatial intervention and planning.

Very early on, she grew aware of the need for regular statistical data on the countries of the world, and she also can be credited with introducing a worldwide view of the demographic and economic changes underway during the second half of the twentieth century. Her books on population provide a synthetic approach that is both demographic and geographic to the issues of demographic transition, first underlining the decrease in mortality in the developed countries, and then the rapid densification of the planet as a result of the demographic boom in Third World countries. Her pioneering interest in urbanization led her not only to write manuals on urban geography, but above all to analyze the major phenomena affecting the city on different scales: research on the commercial structure of cities, on city centers, on transport systems, on underground city developments, etc.

She was a launcher of new ideas and a great traveler, and she had a talent for synthesizing. Throughout her life she generated national and international cooperation, and this has given her ideas a wider and more enduring audience. This is true of the working group on the great metropolitan cities she presided in IGU from 1980 to 1988 (she was one of the very few women in her generation to occupy such a position), and of the work on the geography of commercial activities which she initiated in France and later in IGU.

In many respects, Beaujeu Garnier was a classical geographer, on account of the importance she long attached to physical geography, and her faith in regional geography to guarantee unity in the discipline. However, she implemented a dynamic style of research, preferring to clear the way and to move forward rather than devote herself to the theoretical, historical, or epistemological reflection which was the preference of her contemporary, Philippe Pinchemel. She opened up several fields of human geography, such as the geography of health, which she approached via the medical geography developed in the 1950s by Max Sorre. She supervised a whole range of studies in human geography as it developed from the 1950s: food, industry, services, decentralization, spatial diffusion, leisure, the elderly, scientific research, public space, gender, etc. In this way she showed her awareness of contemporary changes in ideas and practice, and her marked adaptability. The same flexibility is found in the methodological and epistemological field. In particular, Beaujeu Garnier defended the notion of system, perhaps because it enabled the modernization of the notion of combination which her tutor, André Cholley, had promoted in the 1940s. Her 1971 *Methods in human geography* plead for the modernization of geography espoused solutions that the 'new geography' had set out in the course of the previous decade. However, her diagnosis of a malaise in the discipline went hand in hand with an eclectic viewpoint. Thus at the time of the crisis that French geography went through in the 1970s, she ap peared as a reformer trying to reconcile the classic French tradition with a form of modernity derived from English language authors.

In fact, Beaujeu Garnier was a universal geographer, in the words of her colleagues, by way of the thematic extension of her work, the magnitude of her fields of application, and the international networks in which her work founds its place. She was aware of the challenges facing the world in terms of population growth and urban concentration, she was very attached to the feasibility aspects of scientific knowledge, and she gave impetus, both in the national and in the worldwide context, to a pragmatic form of human geography that was open to diversity.

See also: Aménagement du Territoire: Territorial Development; Atlases; City-Region; Francophone Geography; Health Geography; Population Geography; Regional Development Theory; Urbanism; World/Global Cities.

Further Reading

Bataillon, C. (2006). Six géographes en quête d'engagement: Du communisme à l'aménagement du territoire. Essai sur une génération. *Cybergeo*. *European Journal of Geography*, http://www.cybergeo.eu/index1739.html (accessed on Nov 2007).
Beaujeu Garnier, J. (1950). *Le Morvan et sa bordure: étude morphologique*. Paris: PUF.
Beaujeu Garnier, J. (ed.) (1959). *Atlas du Nord*. Paris: Berger Levrault.
Beaujeu Garnier, J. (1965). *Trois milliards d'hommes. Traité de démo géographie*. Paris: Hachette. (*Geography of population* (1978), 2nd edn. London/New York: Longman).
Beaujeu Garnier, J. (1971). *La géographie, méthodes et perspectives*. Paris: Masson (*Methods and perspectives in human geography* (1976). London: Longman).
Beaujeu Garnier, J. (1975). Les géographes au service de l'action. *Revue Internationale des Sciences Sociales* 27 (2), 290 302.
Beaujeu Garnier, J. (1980). *Géographie urbaine*. Paris: Armand Colin.
Beaujeu Garnier, J. (1983). Autobiographical essay. In Buttimer, A. (ed.) *The practice of Geography*, pp 141 152. New York: Longman.
Beaujeu Garnier, J. (1987). Bibliographie géographique de Jacqueline Beaujeu Garnier. In *Régions*, *villes et aménagement*. *Mélanges jubilaires offerts à Jacqueline Beaujeu Garnier*, pp 633 652. Paris: CREPIF, Société de géographie.
Beaujeu Garnier, J. (1993). *Paris: Hasard ou prédestination? Une géographie de Paris*. Paris: Association pour la publication d'une histoire de Paris.
Beaujeu Garnier, J. (1996). *une géographe universelle.* Paris: PRODIG.
Beaujeu Garnier, J. and Bastié, A. (eds.) (1967). *Atlas de Paris et de la région parisienne*. Paris: Berger Levrault.
Beaujeu Garnier, J. and Chabot, G. (1964). *Traité de géographie urbaine*. Paris: Armand Colin (*Urban geography* (1967). London: Longman).
Beaujeu Garnier, J. and Delobez, A. (1977). *La géographie du commerce*. Paris: Masson (*Geography of Marketing* (1979), 2nd edn. London: Longman).
George, P., Guglielmo, R., Kayser, B. and Lacoste, Y. (1964). *La géographie active*. Paris: Armand Colin.
Labasse, J. (1966). *L'Organisation de l'espace*. Paris: Hermann.
Phlipponneau, M. (1960). *Géographie et action*. Paris: Armand Colin.
Pinchemel, P. and Pinchemel, G. (2005). *Géographie. Une intelligence de la terre*. Paris: Arguments.
Robic, M C. (coord.), Gosme, C., Mendibil, D., Orain, O. and Tissier, J. L. (2006). *Couvrir le monde*. *Un grand vingtième siècle de géographie française.* Paris: ADPF (http://www.culturesfrance.com/adpf publi/folio/couvrirlemonde/index.html).
Sorre, M. (1943 1952). *Les fondements de la géographie humaine,* 3 vols. Paris: Armand Colin.

Relevant Websites

http://www.persee.fr
Persee Ministére de l'Enseignment supérior et de la Recherche.

Becoming

D. P. McCormack, Oxford University Centre for the Environment, Oxford, UK

Glossary

Becoming, Theory of A conceptualization of the world as a dynamic and open-ended system of relational transformations.

Being The existence and essence of an entity.

Duration A conception of time as a lived process of intensive differentiation.

More-Than-Human Geographies Geographies that take seriously the role of nonhuman agents in shaping space and place.

Ontogenetic A term derived from developmental biology but now often used to describe the kind of continuous differentiation associated with becoming.

Ontology The philosophical inquiry into the nature of existence and being.

Spacing Spatiality conceived as an ongoing process rather than a noun.

Virtual Something which is real (insofar as it produces an affect) without being actual (insofar as it cannot be located within or defined as an object or discrete entity), for example, memory.

Introduction

Becoming designates a process based ontology of movement, in which the world is conceived of as a dy namic and open ended set of relational transformations. Becoming is frequently distinguished from and some times opposed to being, with the latter suggesting es sential and relatively enduring entities (real or ideal) upon which existence, thought, ethics, etc., can be grounded. For philosophies of becoming, in contrast, such categories are generally understood to be the provisional outcome 'of', rather than the precondition 'for', change.

Given how becoming suggests flux, change, and con tinuous transformation, it might appear to be something about which human geography has always been con cerned: after all, geographers have often defined their disciplinary interest in terms of a concern with how humans have transformed their natural environments. Yet while much of human geography has clearly involved inquiry into processes of spatial transformation, such inquiry has often proceeded via a belief in the ne cessity of establishing certain fixed ontological and epi stemological reference points which themselves are unchanging. The most obvious of these ontological reference points is space, often understood as a kind of preexisting, three dimensional container within which things and events take place. On an intuitive level this seems relatively unproblematic. Yet it has an important limitation: space itself is not bound up in the process of change – it exists instead as a kind of fixed grid within which transformation unfolds. This ontological claim is linked with an epistemological one: namely, that know ledge is the purview of a knowing subject with the cap acity to survey the world from a position of objective independence. Much geographical knowledge is prem ised upon such claims: insofar as this is the case, such knowledge has allied itself with philosophies of being. The primacy and legitimacy of such claims have more recently been questioned by human geographers inter ested in trying to understand space itself as a process: one best understood in terms of becoming rather than being. Reconceptualizing space in this way has implications for the theoretical and methodological practice of human geography. Most importantly, it has encouraged human geographers to understand the very activity of geo graphical thinking as the folding of theory and practice into one another as a process of ongoing differentiation.

A Brief Genealogy of Becoming

The increasing visibility of ideas of becoming in human geography might easily be read as a conceptual response to the spatiotemporal disjunctions of modernity, post modernity, or globalization. Becoming, more than being, seems to capture the dynamic logics of worlds where the coordinates of identity and place seem increasingly fluid. Yet philosophies of flux are not mere symptoms or re flections of the contemporary condition: they have a much longer genealogy. Sketching the outlines of this genealogy provides a useful and necessary way of con textualizing the more recent re emergence of the con cept of becoming within the theoretical and empirical commitments of human geography.

The respective works of the pre Socratic philosophers Heraclitus and Parmenides provide a useful (if also thoroughly Western) point of departure for such an ex ercise insofar as their writings effectively exemplify the tension between being and becoming as philosophical categories. For the former the world is best understood as flux and motion, akin to the ceaseless flow of a river into which it is never possible to step more than once. Fur thermore, this flow is not uniform: it is a lively mix of often contrasting and conflicting forces. The Heraclitean view of the world as flux contrasts sharply with the ideas attributed to Parmenides for whom the claim that change is the fundamental condition of the world is based upon an illusory or false perception. The truth of the world lies instead in the apprehension of those enduring things which have always and will always exist.

The difference between Heraclitus and Parmenides provides an early and oft quoted expression of the distinction between a philosophy of being, in which everything that truly exists has a fixed, eternal essence, and a philosophy of becoming in which the universe is defined by change. Much of Western philosophy has been dominated by the former tradition, a dominance that can be traced through some of the key thinkers of this tradition. It is present, for instance, in Plato's argument for the existence of unchanging ideal forms, an argument indebted to the writing of Parmenides. It is present in the effort by much of the Western philosophical canon to establish the existence of an enduring, transcendent God. And it is present in the effort to establish the foundations for a rational philosophy of mind. Perhaps the most important articulation of such a philosophy is contained in the work of René Descartes, who posits the mind (and/or soul) as a nonextensive center of pure understanding in the midst of a world of mobile and motile bodies, to which the mind is loosely and conveniently connected yet nevertheless fundamentally distinct. The significance of this move is that it excludes the mind from any claims about the becoming of the world. In this philosophical model, the world may change as much as it likes: the independent and indivisible substance of which the mind consists is the one thing upon which thinking can depend.

Philosophies placing primacy on the establishment of categories of being (such as the subject) have achieved a degree of dominance in Western thinking: yet they have not gone unchallenged. This challenge has come from a number of 'minor' strands of thinking, each of which might be said to foreground becoming rather than being. The first consists of the work of a number of 'minor' philosophers whose thinking runs against the grain of the tradition outlined above by challenging the assumption that the mind is independent from and unaffected by transformations of the body. Perhaps the most important figures here are Spinoza and Nietzsche. For Spinoza, mind and body are conceived as two attributes of the same substance, and individual entities, such as persons, are understood as complex, relational orderings between bodies and environments. It makes little sense therefore to speak of the soul or mind as a kind of transcendent entity unaffected by the becoming of the world of which it is a part. Nor would such a claim make much sense to Nietzsche, according to whom the very idea of 'being' represents an illusory image of the world. In this respect Nietzsche echoes and indeed draws directly upon the ideas of Heraclitus, defining the world by the fact of its becoming, by the ongoing play of conflicting forces and energies that never crystallize into a final shape.

A second important impetus to the foregrounding of becoming in Western philosophy can be found in theories of duration and process, exemplified in the respective works of Henri Bergson and Alfred Whitehead. Both thinkers grapple with some of the philosophical implications of changing conceptions of space and time emerging at the end of the nineteenth and beginning of the twentieth century. The force of Bergson's critique is directed against those theories of change which tend to 'spatialize time' by subordinating it to a logic of extension – precisely the move made by thinkers such as Descartes. Bergson insists instead on the necessity of a conception of duration allowing for the existence of pure qualitative differentiation. If this seems to render temporality and movement abstract it is because, to paraphrase Whitehead, the real task of philosophy is not to arrive at explanations of the concrete, but to explain abstractions. Influenced in part by Bergson, Whitehead develops a process based ontology in which the very essence of being is the fact of its becoming. Or as he himself puts it, "the actual world is a process, and that process is the becoming of actual entities."

The work of Bergson and Whitehead is related to a third influence on the foregrounding of becoming as a philosophical imperative. This revolves around efforts to place primacy on the relation between thinking and the experience of pragmatic involvement in the world. Such efforts are best exemplified by two traditions, the first of which is the variety of North American pragmatism associated with William James and John Dewey. Key to the work of both thinkers is the idea that the world is not given in advance but must be made and re made through ongoing activity. A second and distinctly more European tradition can be discerned in the various strands of post Cartesian phenomenology, most notably the work of Martin Heidegger and Maurice Merleau Ponty. While their respective works are very different, they both advocate a kind of philosophical inquiry into the category of being. On one level this might place them at odds with philosophies of becoming, and in some respects this is the case. Nevertheless, both insist on practical relational involvement in the world as the moving condition for existence. The influence of Merleau Ponty's work on flesh as a kind of intimate corporeally intertwining with the world has been particularly significant in this respect.

A fourth set of influences moves beyond the phenomenological tradition to consider the 'more than human' becomings from which the apparent individuality of the human subject emerges. Such post phenomenological arguments can be discerned, with different degrees of intensity, in the work of thinkers such as Michel Foucault, Gilles Deleuze, Felix Guattari, and more recently in the writing of Brian Massumi. These thinkers owe a debt to many of the figures whose work has already been mentioned above. For instance, drawing upon Nietzsche, Foucault develops a critique of history as the progressive and linear unfolding of time, and of the very status of the human as a subject with the capacity to survey history from without. Deleuze is similarly critical of philosophies of being in which the subject is established as a kind of *a priori* entity. Rather, and taking support from thinkers including Spinoza and Bergson, Deleuze argues that being is already multiple, already defined by the difference immanent to its own creative evolution. In collaboration with Guattari, Deleuze's ideas are reworked as part of an explicitly cartographic geophilosophy emphasizing the creative and affective movement of life as a process of operating below and transversal to the scale of distinct, organized bodies and territories. Perhaps the most emphatic recent articulation of this line of thought is to be found in the writing of Brian Massumi, who articulates and practices a thinking space which takes seriously the moving fact of its own becoming.

A fifth and final set of influences on the foregrounding of becoming in Western thought can be traced through feminist theories of embodiment and difference that challenge the primacy of sex and gender as fixed categories of identity. Here the work of thinkers such as Judith Butler, Donna Haraway, and Elizabeth Grosz is of particular significance. Clearly, individual feminist theorists differ in the extent to which they configure the relation between the sociocultural and the biological, and the degree to which the category of woman is essentialized. Yet their collective importance consists in the opening up of the spaces and politics of identity (and desire) to questions of becoming.

Of course these developments do not add up to a coherent philosophical framework: they are better described as a set of conceptual tendencies with all sorts of contradictions and conflicts. Furthermore, it would be a mistake to claim that each of the thinkers above sharply juxtaposes the idea of being with that of becoming. So for Merleau Ponty, for instance, it is possible to hold in tension 'both' the idea of phenomenology as the philosophical enquiry into essences 'and' the claim that this inquiry can only proceed if it acknowledges its own relational involvement in the world. And while sharply critical of elements of Merleau Ponty's work, Deleuze nevertheless retains the category of being in some of his own writing. However, in doing so, he seeks to complicate the apparent unity of this category. Put another way, for Deleuze it is only possible to speak of the event of being as a process of becoming that is always already differentiated.

Becomings/Geographies

Although diverse, and often divergent, the philosophical tendencies outlined above have a number of important implications for the theoretical, empirical, and political commitments of human geography. First, and perhaps most obviously, they precipitate sustained reflection on the nature of space. To think of space in relation to becoming is to challenge the familiar spatial coordinates of Euclidean geometry. Space is not an unchanging, abstract container within which extensive objects are located and within which activity takes place. Nor is it prior to the positionality of an object. Rather, space, or more accurately, spacing, is better conceived in dynamic terms as a process of folding and refolding productive of what Marcus Doel (1996: 421–439) calls "scrumpled geography": a world of "continuous variation, becoming, and chance, rather than one of constancy, being and predictability." To think of space in relation to ideas of becoming – as Doel does with particular energy – is therefore to confront the fact that space is not an ideal form transcending the processual nature of involvement in the world. Space is better conceived as verb rather a noun.

Second, and relatedly, to think of space through the logics of becoming is to problematize the separation of space from time. While this claim might seem obvious, it is potentially contentious in a discipline whose distinctiveness is bound up with understanding of space and spatiality. Such a claim also disturbs the argument that the social sciences need to rediscover space and spatiality, particularly in the wake of the collapse of grand historical narratives. The real question is not to rethink or reequilibrate the relation between space and time, but to think of both as part of the same material semiotic process of becoming. This is not without difficulty. Hence our fondness, as Bergson noted, for spatializing time through certain habits of thinking that locate temporality within particular objects and entities or cut up the world into discrete quanta of representational content. There are however conceptual vehicles through which the temporal intensities of extensity can be made palpable: rhythm is one. In the work of social and spatial theorist Henri Lefebvre rhythm is used to support a way of thinking through and apprehending the dynamic (and for him discontinuous) becoming of space–time.

Third, if becoming precipitates a rethinking of space–time as a kind of rhythmic flux, it also demands a careful reassessment of the space of materiality. More specifically, notions of becoming complicate any attempt to 'ground' the abstractions of theory in a relatively stable world of material entities and objects. Any spatially inflected theory of materiality, if it is not to collapse under the sheer fact of its own self coincidence, must be able to account for the virtuality of durational differentiation outlined by thinkers such as Bergson, Deleuze, and Massumi. This is a crucial question as far as the issue of spatial transformation is concerned. If everything is actual then no possibility exists for transformation – or rather change is only possible in relation to already formed entities. Taking becoming seriously means that the material of space is not that around or through which becoming occurs. Rather, becoming – and its relative speeds and rhythms – is part of what makes the materiality of space matter. In response to this claim geographers are beginning therefore to develop ways of thinking through the transformative materiality of space and about how this transformation is facilitated by a range of technologies and technological practices. Consider, for example, software code. On one level this might be understood as information that flows from one site to another. But the flow of code can also be understood to be part of how space becomes insofar as it provides for new ways of connecting, ordering, and inhabiting worlds. In this sense code is not simple representational: it is also generative of new kinds of space.

Such conceptual agitations of the spatiotemporal and material entanglements of geography by definition have important implications for the very 'doing', or practice, of geography. Not least of these is the fact that the conceptual element of geographical thinking is redefined as a practice caught up in the becoming of the world, and not as something surveying this flux from a position of distanced abstraction. This has important consequences for how basic geographical activities such as fieldwork are understood. Following Deleuze and Guattari, fieldwork can be rethought as a kind of 'becoming minor': a series of modest, performative interventions between the conceptual and the empirical. Such conceptions of fieldwork resonate with recent work within the discipline about how the becoming of space is entangled in questions of performance and performativity. It is also supported by the claims of nonrepresentational styles of work, particularly in the writing of Nigel Thrift, for whom theory is refigured as a conceptually charged set of mobile practices through which to apprehend the movement of an open ended and ontogenetic world.

Informed by such theoretical reorientations we can trace becoming through the geographies of very ordinary practices. Consider, for instance, gardening. While such an activity might be dismissed as politically unimportant, it can also be more generously understood as practice of relational spacing that moves in the ambivalent zone between being and becoming, between the predictability of habit and the unexpectedness of potential for change, however minor. Focusing on such mundane practices also poses an important challenge to the primacy and fixity of the human as a kind of unchanging reference point for geographical thought. In this respect theories and practices of becoming overlap with and supplement efforts to produce cartographies for a more than human world, involving animals, devices, and artefacts of various kinds that have always been hybrid. As Sarah Whatmore (2002) has argued, the task of such hybrid cartographies is to articulate "the fluxes of becoming that complicate the spacings timings of social life."

If the idea of becoming complicates the boundaries between the human and nonhuman as distinct registers of being, it also creatively confuses the subdisciplinary boundaries that usually separate human and physical geography. Indeed, one of the key implications of taking seriously ideas of becoming is precisely such boundary confusion. Following Deleuze and Guattari, we can think of particular places as 'blocks of becoming' composed of physical, social, technological, and political–economic forces. It makes little sense – conceptually and or empirically – to try to abstract such forces from one another. In this respect the idea of becoming provides an important element of a conceptual vocabulary that might better allow geographers to rethink complex questions that transcend the strict division between the human and physical side of the discipline – climate change being an obvious example. Here the task is not simply to link natural, social, and technological systems. Rather, each needs to be understood as a series of generative processes through which the very relations that constitute the environment are transformed.

However, the challenge posed by philosophies of becoming is not just that they point to new ways of thinking about space. Such philosophies also demand a recasting of practices of geographical knowledge production. Put another way: to take seriously the becoming of the world is to cultivate a commitment to enactive styles of thinking and doing. This does not necessarily mean devising strange new practices or sites of eventful encounter, although it can. Rather, generating a sense of relations of becoming necessitates the cultivation of modes of address and techniques of writing that amplify the moving resonance of our ongoing involvement with the world.

Conclusion

It is of course possible to raise various cautionary points about thinking space through ideas of becoming. Critics might very well claim that to foreground becoming is to cleave to an overly optimistic view of the potential for change. Nor are the ethicopolitical implications of such ideas always clear. Theories of becoming provoke suspicion insofar as they seem to erode the coherence and distinctness of agencies of and for change. Put another way, of what value is differentiation without agency if this differentiation cannot in some way become the focus for tactics and techniques of intervention? One might counter this critique with the observation that to foreground becoming is to extend agency across assemblages and blocks of space time. It is also about harnessing the transformative potential immanent to open ended spacings – and in this sense it is intensely political. Crucially, any spatial politics of becoming is poorly served by thinking of spaces in ontological terms – ontology is insufficient to the task of thinking through geographies of becoming. What is required is an 'ontogenetic' form of thinking space – a form of thinking open to and attuned to its own capacity for variation, cultivating in the pro cess an affective orientation not so much of optimism, but of hope.

See also: Fluidity-Fixity; Non-Representational Theory/ Non-Representational Geographies; Performativity; Philosophy and Human Geography; Pragmatism/ Pragmatist Geographies; Space I.

Further Reading

Anderson, B. (2006). Being and becoming hopeful: Towards a theory of affect. *Environment and Planning D: Society and Space* 24(5), 733 752.

Bergson, H. (1988). *Matter and memory*, Paul, N. M. and Palmer, W. S. (trans.). London: Zone Books.

Bonta, M. (2005). Becoming forest, becoming local: Transformations of a protected area in Honduras. *Geoforum* 36(1), 95 112.

Bonta, M. and Protevi, J. (2004). *Deleuze and geophilosophy: A guide and glossary.* Edinburgh: Edinburgh University Press.

Crouch, D. (2003). Spacing, performing, and becoming: Tangles in the mundane. *Environment and Planning A* 35(11), 1945 1960.

Deleuze, G. (1994). *Difference and repetition*. Patton, P. (trans.). London: Athlone.

Deleuze, G. and Guattari, F. (1988). *A thousand plateaus*, Massumi, B. (trans.). London: Athlone.

Doel, M. (1996). A hundred thousand lines of flight: A machinic introduction to the nomad thought and scrumpled geography of Gilles Deleuze and Felix Guattari. *Environment and Planning D: Society and Space* 14(4), 421 439.

Dodge, M. and Kitchin, R. (2005). Code and the transduction of space. *Annals of the Association of American Geographers* 95(1), 162 180.

Grosz, E. (ed.) (1999). *Becomings: Explorations in time, memory, and futures*. London: Cornell University Press.

Katz, C. (1996). Toward minor theory. *Environment and Planning D: Society and Space* 14(4), 487 499.

Lefebvre, H. (1991). *The production of space*. Nicholson Smith, D. (trans.). Oxford: Blackwell.

Lefebvre, H. (2004). *Rhythmanalysis: Space time and everyday life*. London: Continuum.

Massey, D. (2005). *For space*. London: Sage.

Massumi, B. (1996). Becoming Deleuzian. *Environment and Planning D: Society and Space* 14(4), 395 406.

Massumi, B. (2002). *Parables for the virtual: Movement, affect, sensation*. London: Duke University Press.

May, J. and Thrift, N. (2000). *Time space: Geographies of temporality*. London: Routledge.

Nietzsche, F. (1968). *The will to power*. Kaufmann, W. and Hollingdale, R. J. (trans.). New York: Vintage Books.

Spinoza, B. (1989). *Ethics*. Parkinson, G. H. R. (trans.). London: Everyman.

Thrift, N. (2004). Summoning life. In Cloke, P., Goodwin, M. & Crang, P. (eds.) *Envisioning human geographies*, pp 81 103. London: Arnold.

Thrift, N. (1999). Steps to an ecology of place. In Allen, J. & Massey, D. (eds.) *Human geography today*, pp 295 321. Cambridge: Polity Press.

Whatmore, S. (2002). *Hybrid geographies: Natures, cultures, spaces*. London: Sage.

Whitehead, A. N. (1978). *Process and reality*. New York: The Free Press.

Wylie, J. (2002). Becoming icy: Scott and Amundsen's polar voyages, 1910 1913. *Cultural Geographies* 9(3), 249 265.

Behavioral Geography

J. R. Gold, Oxford Brookes University, Oxford, UK

Glossary

Cognitive Behavioralism An approach which holds that people's spatial behavior depends on how they understand (perceive, cognize) the world around them.
Cognitive Mapping The processes by which individuals acquire, store, and recall information about the places in the environments with which they come into contact.
Cognitive Science The branch of the science that studies the cognitive processes which underpin the acquisition, storage, and use of knowledge.
Ethology The comparative study of animal behavior.
Gated Community A residential community protected by physical barriers or symbolic means in order to exclude intruders.
Humanism A broad set of philosophies that focus on human reason, actions, and motives without reference to supernatural phenomena.
Natural Hazard An extreme event in the physical environment, of geophysical or biological origin, that is injurious to human life and destructive to property.
Satisficing Seeking to achieving a satisfactory outcome, rather than the optimal.
Territoriality It refers to any form of behavior displayed by individuals and groups seeking to establish, maintain, or defend specific bounded portions of space.

Introduction

Defining terms is an unusually important activity in the case of behavioural geography, since it has two historically contingent meanings. The first views it *sensu lato* as a movement with multidisciplinary leanings that enjoyed its greatest influence between the years 1965 and 1980. Broadly coterminous with what others variously referred to as 'environmental perception', 'behavioral and perceptual geography' or 'image geography', behavioral geography emphasized the role of cognitive processes in shaping decision making and behavior, for which reason its underlying approach was known as 'cognitive behavioralism'. In essence, its proponents argued that people's spatial behavior depended on how they understood (perceived, cognized) the world around them, but researchers varied markedly in how they conceived and tackled their subject matter. Some embraced 'cognitive science', examining regularities in human spatial cognition and behavior and using their findings as a basis from which to generate theories about how people make decisions and act in geographic space. Others looked to 'humanistic' approaches. Critical of what they saw as the reductionist tendencies stemming from scientific inquiry in general and mainstream psychology in particular, they more often sought to understand human imagination and experience holistically than wishing to make overt connections with behavior.

The second and more contemporary meaning of behavioral geography defines it *sensu stricto* as a subdiscipline of human geography. To some extent, this definition arose by default. Tensions between behavioral geography's 'cognitive science' and 'humanistic' streams eventually led to these two schools of thought parting company by the early 1980s. Henceforth, the term 'behavioral geography' described the work of those previously associated with the 'cognitive science' wing and was largely confined to North America, where it maintained an accepted but increasingly marginal presence in geographical research and in the undergraduate teaching curriculum.

At first glance the existence of these two overlapping definitions would seem of little real consequence, but the failure to differentiate between them lies at the heart of the prevailing historiographic misrepresentations of behavioral geography. By the 1990s at least, the prevailing view was, first, that behavioral geography was primarily a limited extension of spatial science and, second, that the limitations of its underlying positivist philosophy led to it being challenged and replaced by alternative approaches. By failing to recognize its true scope as a broad movement that provided a home for nonpositivist as well as positivist approaches during the 1960s and 1970s, it becomes difficult to recognize behavioral geography's contribution as a forum that once nurtured what was later termed 'humanistic geography' and, indeed, indirectly helped to lay the foundations of the 'new cultural geography'.

This article, which is arranged chronologically, starts by looking at behavioral geography *sensu lato*. After examining its precursors, the early sections analyze its principal conceptual frameworks, before dealing thematically with three foci for cognitive behavioralist research in geography, namely: cognitive (mental) mapping, natural hazards, and attachment to place. The ensuing section notes the emergence of powerful bodies of criticism from radical and humanistic geographers, which progressively eroded the uneasy coexistence between positivist and nonpositivist approaches. The penultimate section of this article notes the continuing existence of behavioral geography *sensu stricto*, but points to its in creasing peripherality and remnant status. The conclu sion considers the limited prospects for revitalization of behavioral geography *per se*, despite the continuing relevance of some of its central tenets.

Early Development

The 1960s' lasting reputation as a decade of rapid and revolutionary change is mirrored in many fields of in tellectual and creative endeavor in which practitioners seized the opportunity to challenge doctrinal orthodoxy and question established practices. Geography was no exception. Restlessness with the prevailing idea of geography as a compendium of factual material about regions of the world combined with an intellectual curiosity that drew strength from a sudden and decisive weakening of disciplinary boundaries contributed to the advent of the theoretical quantitative revolution and the rise of spatial science. These developments indelibly reshaped the discipline, influencing what geographers did and the modes of analysis that they adopted. Yet almost as soon as the new agenda of quantification and model building dawned, a significant groundswell gathered pace that had different preoccupations and priorities. Geog raphers following these new directions also gained in spiration from multidisciplinary convergence, but while quantitative geography looked to mathematics, systems theory, and positive economics, other nascent areas of human geographical inquiry turned instead to the social sciences, the humanities, the arts, planning, architecture, and design.

Behavioral geography was a case in point. Whether viewed in narrower or broader senses, it stemmed from recognizable, if isolated, antecedents from the first half of the twentieth century. Early interest focused on maps and orientation. In 1907, F. P. Gulliver read a paper to the Association of American Geographers on the way that schoolchildren perceived issues involving direction finding in their use of maps, a theme furthered by the physicist C.C. Trowbridge in 1913 in an essay on maps and imagination. They found that orientation was more often achieved by an egocentric reference system than by points of the compass, with Trowbridge recognizing the significance of the home as an anchoring point in these reference systems.

During the interwar years, the writings of American geographer Carl Sauer and the work associated with the Geography Department at the University of California – Berkeley (which Sauer headed) helped to counter pre vailing theories of environmental determinism by stressing the way in which human beings, through the agency of culture, also molded their environment. His compatriot John Kirtland Wright emphasized the importance of the imagination and intellectual curiosity in shaping the environment and suggested that geog raphers might concern themselves with geosophy – the study of geographical knowledge from any or all points of view. In the Francophone world, Georges Hardy's *La Géographie Psychologique* (published in 1939) drew attention to the significance of human psychology when con sidering the relationship between people and their environment.

The 1950s saw further contributions of considerable importance that, while also initially isolated initiatives, exerted a direct influence on later research. In 1952, the British historical geographer William Kirk produced a model based on Gestalt theory to argue that people's behavior was rooted in a behavioral environment (the world as perceived) rather than the objective environ ment (the world of 'facts'). In 1956, the British born social scientist Kenneth Boulding produced a monograph en titled *The Image*, which suggested that human knowledge was centered around 'images' – organized impressions of the world that people develop through their experience and which act as the basis for their behavior. In itself, this work might have remained an obscure discussion piece if the architect Kevin Lynch had not taken up its central idea in his book *The Image of the City*. Lynch explored the visual quality of three US cities (Los Angeles, Boston, and Jersey City) by reference to the way that the physical features of their cityscapes were perceived by small samples of their residents. He showed that people understood their surroundings in consistent ways, by forming 'mental maps' that had five recurring features (paths, edges, districts, nodes, and landmarks). The extraordinary popularity of Lynch's book helped to es tablish Boulding's concept as a key part of the frame works employed in early behavioral geographical research (see **Figure 1**), even though the concept bore no

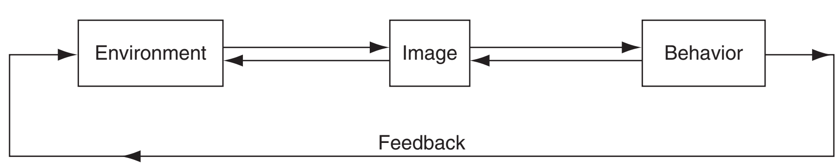

Figure 1 A simplified model of human environment relations, *c.* 1970. Source: Gold, J. R. (1980). *An Introduction to Behavioural Geography*, p 38. Oxford: Oxford University Press.

direct relationship to the notion of imagery used by psychologists.

In their different ways, these works in effect foreshadowed the humanistic and cognitive science streams that fueled the growth of behavioral geography in the 1960s. The former gained impetus from David Lowenthal's paper 'Geography, experience and imagination', published in 1961, which drew on J. K. Wright's agenda to recognize that people live in parochial worlds, shaped by selective perceptual filtering of environmental information but bound into the common realm of experience through culture. For his part, the cultural geographer Yi Fu Tuan moved beyond his early geomorphological interests in the landforms of New Mexico to reflect on the emotional connections between human beings and the physical environments that they encountered. These reflections appeared first in articles published in *Landscape*, the journal founded and edited by the writer J. B. Jackson, then in a series of monographs that explored the relationship between cosmology and feelings toward landscape, and finally in his 1974 book *Topophilia*. At the end of the 1960s, Anne Buttimer supplied an important introduction to non Anglophone sources on human–environment relations through her discussion of French theories of social space and their implication for the subjective configuration of urban areas. She cautiously applied these perspectives in a subsequent study of a residential area in Glasgow (Scotland).

For their part, those who founded the 'cognitive science' stream were initially interested in issues concerning the knowledge held by decision makers. Peter Gould's use of game theory in 1963 to simulate the probable planting strategies of African farmers in the face of the uncertainties of their environment indicated a wish to relax the typical assumptions made by spatial analysts about perfect knowledge and economic rationality. In 1964, Julian Wolpert used a sample of Swedish farmers to test whether real world decision making accorded to satisficing (guided by the goal of seeking a sufficient output) or optimizing principles. Given that the results showed that crop yields were well below the theoretical optimum – in certain areas as low as 40% of the possible yield – Wolpert concluded that the satisficing was more descriptively accurate of the behavioral pattern of the sample population. Nonetheless, there was no necessary implication of irrationality in these findings. While they might signify lack of knowledge about potential yields and the alternatives that were available, they might also reflect perception of risk, aversion to uncertainty, and culturally transmitted values that placed other benefits above profit maximization. Similarly, when analyzing industrial location decision making, Allan Pred developed a powerful critique of normative economic models, pointing to their illogicality (e.g., it is impossible for competing decision makers to arrive at optimal locational decisions simultaneously) and the failure to consider motivation (in that decision makers might well be satisfied with a location that is satisfactory in the long run rather than optimal in the short run). As an improvement, Pred proposed a 'behavioral matrix', which included the quantity and quality of information on one axis with the ability to use that information on the other, pointing to the unlikelihood of decision makers managing to achieve the optimum on both counts.

These types of studies represented a significant first step. By pointing to likely variations in behavior instead of being able to rely on assumed optimisation, their authors stressed the need to include 'black box' variables within the analysis – in other words, incorporate those nonobservable elements that could only be captured by surrogate means. It also heralded the future close alignment between behavioral geography and psychology. In a published version of a talk given to the Michigan Inter University Community of Mathematical Geographers in 1965, Peter Gould pointed to the need to know about the 'maps' that people have in their heads with regard to the world around them. He accepted that individual cognitions were unique but felt that general statements might be made with a technique that could externalize personal preferences and the values that underpinned them. He illustrated this point by reference to studies of residential desirability in the USA, Sweden, the Federal Republic of Germany, Ghana, and Nigeria, whereby student respondents were asked to rank regions of their home country on the basis of lists of administrative areas (such as states or counties). By aggregating the scores for individual respondents, it was possible to map residential preferences for the national groups of students and make comparisons between students living in different nations.

Other researchers looked to mainstream psychological methodology to investigate the cognitive domain. In 1966, for example, Thomas Saarinen published an account of his doctoral research on the behavior of farmers in the drought prone Great Plains region of the USA. Using personality measures such as the Thematic Apperception Test as part of his methodology, he found from their responses that successes were recalled more frequently than failures. Farmers consistently underestimated the frequency of years when drought was experienced and were optimistic about the number of good years and about the size of crops in such years. Perceptions of the drought hazard were also influenced by age and previous experience, especially whether farmers remembered the drought conditions of the 'dusty thirties' (the years of the formation of the Dust Bowl).

Collectively, these early studies indicated that the emerging field of research served as both a corrective and a catalyst. As the former, it countered the dominant thrust of spatial science by embracing such considerations as learning, imperfect information, and bounded rationality in order to improve the diagnostic and predictive power of locational models. As a catalyst, behavioral geography quickly moved beyond being an adjunct to spatial scientific inquiry to become a channel for work that expanded research horizons and reconceptualized the relationships between people and their environments in quite different ways from those conventional in geographical discourse. Benefiting from the breadth of perspectives available from its two component intellectual streams, behavioral geography became a wide ranging field that revealed a notable flowering of creativity, with almost any scale of cognitive behavioral problem, from the interior spaces of buildings to understandings of the cosmos, becoming of interest to geographers.

Consolidation

As with many embryonic fields of study, the work of the pioneers was followed by extensive efforts to consolidate knowledge. A supportive infrastructure of discussion groups and symposia developed first in North America, with the subsequent foundation of newsletters, journals, and edited collections allowing wider dissemination of materials. For example, the first symposium on 'environmental perception and behavior', held at the Annual Meeting of the Association of American Geographers (AAG) at Columbus (Ohio) in April 1965, gave rise to an influential collection of papers. Edited by David Lowenthal, it immediately symptomized the rampant plurality of the field, containing essays on attitudes toward environment, the development of psychological tests to analyze spatial symbolism, human adaptation to Arctic conditions, the study of storm hazards, and the perceptual implications of travel along major urban highways. A further volume, entitled *Behavioural Problems in Geography*, grew out of the 1968 Annual Meeting of the AAG in Washington, DC. Edited by Kevin Cox and Reginald Golledge, it attempted to bring some measure of overview based around an analytic approach that emphasized theory building and positivist research methodologies.

Individually authored surveys of the state of the art quickly followed. Thomas Saarinen's primer *Perception of the Environment*, published under the imprimatur of the Association of American Geographers in 1969, offered a simple organizing framework for research findings arranged around spatial scale. His analysis began with personal space – an area of space around the body defined by culturally observed rules of appropriate interpersonal distance into which intrusion by others is resented and often resisted. The survey then moved up through a spatial hierarchy that extended from the home to the neighborhood, city, region, and nation, culminating in world views. Saarinen also gave expression to the widely held conviction that cognitive behavioralist research could assist practical planning and design by offering critical insight into the relationships between people and their environment.

With hindsight, this text expressed two abiding features of behavioral geography. First, it had a powerful, if seldom recognized, ideological dimension. Behavioral geographers followed spatial analysts in their practice of asserting that the new expert knowledge and associated skills were of great value to society. The basic principle was that cognitive behavioralist research could assist planning and design by offering new understandings and appropriate techniques for studying people's relationships with their environment. In the longer term, however, it became apparent that the results of that research more often revealed evidence of the complex ways in which spaces and places were perceived rather than generating results capable of practical implementation.

Second, and related, the idea of an egocentric hierarchy of bounded spaces had a compelling appeal. It seemed to offer geographers a lucid way of thinking about the world and implicitly suggested that space was an important variable in environmental cognition. The idea was avidly adopted by other researchers. For example, a 1971 monograph by Brian Goodey suggested a similar hierarchy of perceived spaces as a heuristic device (**Figure 2**). At its core were personal space and the intimately known places that arouse a sense of belonging. The hierarchy then included places visited with some regularity, places visited less frequently, and extended outwards to 'far places'. There was, however, no direct relationship between distance and knowledge. 'Far places' were often spatially distant but not necessarily so. Furthermore, while individuals may never have visited these places, they could still hold strong views about them on the basis of information derived from other people, from the press or broadcasting, or through the operation of cultural stereotypes.

Over time other, more complex frameworks emerged. **Figure 3**, for example, shows an example dating from around 1977 that purported to identify the elements and relationships linking spatial cognition and behavior. Broadly speaking, it depicts two forms of 'representation' – images and spatial schemata – as the basic cognitive elements that mediate behavior. Individuals are part of two worlds: the objective and behavioral environments. The 'images' that they hold are the mental pictures called to mind when objects, people, places, or areas are not part of current sensory information. 'Spatial schemata' are frameworks within which people organize their knowledge of the environment, containing the residue of past experience and accommodating current sensory information. Both images and spatial schemata are influenced by cognitive processes, personality, and socio cultural

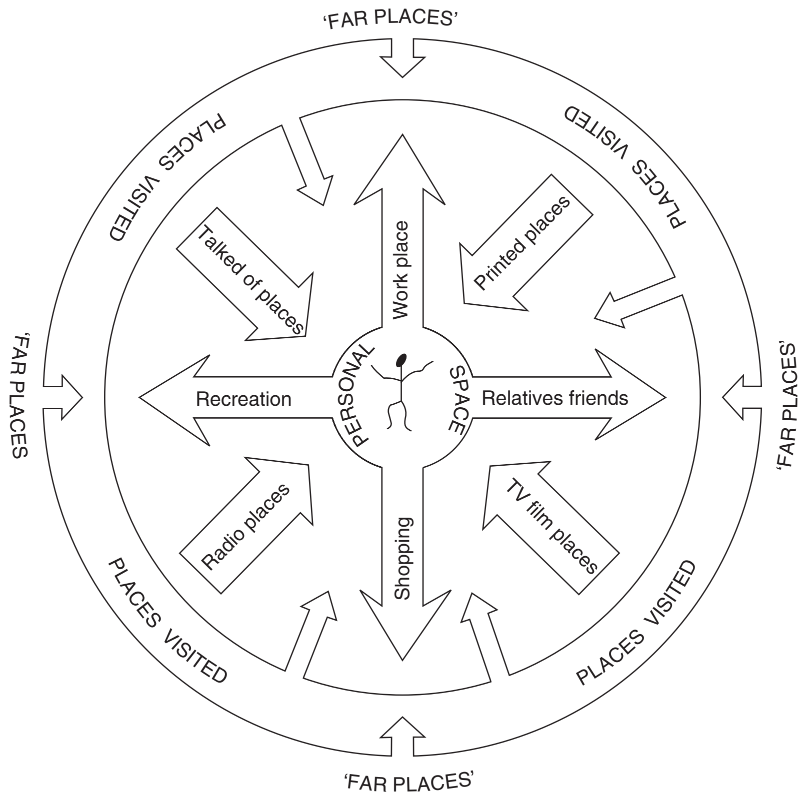

Figure 2 The egocentric spatial hierarchy. Source: Goodey, B. (1971). *Perception of the Environment*, Occasional Paper 17, p 7. Birmingham: Centre for Urban and Regional Studies, University of Birmingham.

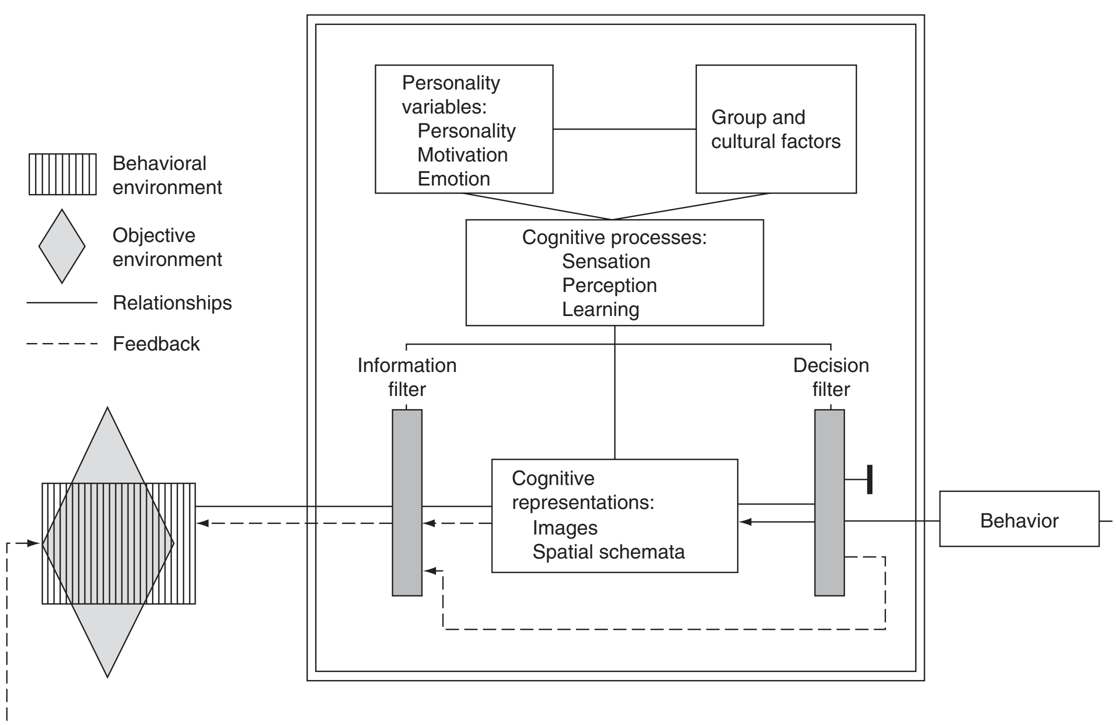

Figure 3 A schematic framework for individual spatial cognition and behavior, *c.* 1977. Source: Gold, J. R. (1980). *An Introduction to Behavioural Geography*, p 42. Oxford: Oxford University Press.

variables. These collectively contribute to filters that af fect the flow of information and shape the decisions made. Feedback loops at all levels serve to indicate learning, assign value to experience, and to correct future behavioral strategies.

At one level, this type of framework had a seductive simplicity, suggesting permissively broad lines within which to integrate research. Yet despite superficially grafting on terminology and perspectives more closely aligned to the discourse of psychology, in reality such frameworks made only passing reference to established cognitive theory or even to the particular psychological concepts that it included. In retrospect, it now appears as symptomatic of the conceptual weakness that bedevilled behavioral geography. Moreover, this conclusion, which is true for behavioral geography in general, was also true for many of the component areas addressed by re searchers in the years up to 1980 – as the following sections indicate.

Cognitive Maps

The term 'cognitive map' derives from a paper written by Edward Tolman in 1948. Entitled 'Cognitive maps in rats and men', although mainly about rats, it put forward the idea that the rodents built up so called 'cognitive maps' over time that enabled them to steer their way about their environments. Tolman compared the rat's brain to the central office of a map making agency, using the prin ciples of psychological field theory to explain how the rat selectively extracted information from the environment in order to build up a field like map of that environment. This, of course, was merely an analogy, but the idea of a 'map in the mind' and the corresponding processes of 'cognitive mapping' – by which individuals acquire, store, and recall information about the places and environ ments with which they come into contact – were eagerly adopted by behavioral geographers. Broadly speaking, the resulting research centered on three main areas.

The first comprised inquiries into the construction and organization of cognitive maps, examining the way in which maps acquire stability and how they accommodate change through learning. An immediate and lasting focus comprised studies of the spatial knowledge of children at particular stages in their development. *Inter alia*, re searchers examined the differences encountered when comparing children of different ages, the world views of schoolchildren, the effect of ethnic group and gender upon the known world, the role of socioeconomic status, the impact of the journey to school, preferential cognitions and perceptions of danger, and the role of fantasy and imagination. Attention was paid to the rep resentational strategies adopted and to methods by which to externalize spatial knowledge. A broad distinction became apparent between those that considered that spatial knowledge was accumulated incrementally and those that believed that qualitatively different abilities were involved at different ages. This distinction mirrored that found in developmental psychology between incre mentalists and constructivists. Influenced particularly by the ideas of Jean Piaget and Heinz Werner, the latter group believed that spatial learning was not just a ques tion of acquiring more spatial information (as did the incrementalists), but that it also involved a sequence of stages in which the child developed progressively more sophisticated schemata through which to organize this information.

The second area of research centered on cognition of the urban environment. A large corpus of research that examined perceived distance within cities, for example, generated four variables that helped to explain the dis crepancies between perception and actuality, namely: direction (overestimates of distances towards the city centre compared with the opposite direction); the at tractiveness of the location; familiarity (especially in terms of length of residence); and the directness of the route. A substantial volume of research also focused on the character and function of the urban neighborhood. Typical topics for investigation included comparisons between perceived neighborhood and the spatially de fined 'neighborhood units' then popular with planners, cognition of neighborhood quality, studies of linkages between residential satisfactions and neighborhood sta bility, neighborhood preferences, and intra urban mi gration. Studies of spatial decision making with regard to retailing and services also showed pronounced 'neigh borhood effects' in choice of destination and resulting travel behavior.

The most significant body of work on cognition of the urban environment, however, was stimulated by Lynch's studies of urban legibility. *The Image of the City* showed how readily understood methods (sketch maps and verbal descriptions) could be used to explore urban imagery and provided the fivefold typology by which to classify elements of the perceived city. Although Lynch neg lected questions of meaning, the speed with which other workers replicated his approach indicated that a viable research paradigm had been created – within 15 years, there were studies consolidating his findings in other cities of the USA, the UK, France, the Netherlands, West Germany, Lebanon, and Venezuela amongst others. Elsewhere, Donald Appleyard applied Lynch's ideas to the development of a new town (Ciudad Guayana, Venezuela) and gave credence to the assertion that cog nitive behavioralist research might have policy relevance. In addition, Appleyard contributed to a project with John Myer and Lynch on the 'view from the road'. This used a study of motorists' travel along freeways into four American cities (Boston, Hartford, New York, and

Philadelphia) to explore road travel as a conscious aesthetic experience. Understandably, this study emphasized the visual component of spatial cognition, but the same technique of using a set route as a transect was employed elsewhere to explore the sonic and haptic environments of cities.

The third area of cognitive mapping research explored the wider scale of regional and world images. In their book *Mental Maps*, for example, Peter Gould and Rodney White used the technique of rank ordering regions from lists of national administrative units (see above) to inquire into the nature of preferential cognition. Their text contains studies of images of Britain; environmental preferences and regional images in the United States and Canada; analysis of patterns of ignorance, information, and learning; and the implications of mental maps for administration and education. Gould and White mapped preferences by means of principal components analysis and depicted the results by differential shading on a base map. Subsequent researchers used a rubber sheeting metaphor, producing a three dimensional impression of the degree of distortion of the cognitive map compared with the cartographic map. The theme of place and regional stereotypes was further developed by researchers interested in the way that people cope with environmental complexity, noting that places and regions become classified on the basis of gross generalizations that may contain only a grain of truth. Correspondingly, those studying the images of negatively stereotyped locations also drew attention to the use of place promotion – the conscious use of publicity and marketing to communicate selective images of specific localities or areas to a target audience – to carry out rebranding in order to make places more attractive to potential industrial, commercial, or residential migrants.

Natural Hazards

The fact that people continue to live in places at risk from hazardous events, where 'objective analysis' might be presumed to suggest safer alternative locations, provided the rationale for another strand of cognitive behavioral research by geographers. Natural hazards research combined the geographer's interest in physical environmental processes that apparently lie outside human control with the desire to understand complex patterns of behavioral response. It focused on a range of highly infrequent, but extreme geophysical (climatological and geological), events that cause major hardship and possible loss of life for the communities affected. Natural hazards research owed its origins to Gilbert F. White, whose work on response to flooding in the USA had pointed to the complexities of individual and societal response to flooding. From initial focus on river floods, the scope of this research broadened in its coverage to include other forms of flooding (e.g., coastal floods), drought, hurricanes, snow and avalanche, tornadoes, volcanic eruption, and tsunamis. Natural hazards research also increasingly acquired a cross national remit in the 1970s with an American led program of studies carried out under the auspices of the International Geographical Union, which involved researchers from 15 nations examining 9 different types of hazard.

The most significant results from this research emphasized a view of human beings as possessing bounded rationality. Extreme events themselves were seen as being outside the control of individuals but, since disaster occurred with relative infrequency, the threat was seen as dormant and remote rather than immediate and real. What might appear as misperception when disaster does occur normally helps to compensate for the fact that making precautionary adjustments to cope with these events would require major changes in the way of life of those under threat, which, for a variety of reasons, they are reluctant to make. However, most researchers stressed that this was not a simple case of people ignoring unequivocal evidence. In most cases, information from the hazardous environment is ambiguous and provides a paucity of reliable cues on which to base action. As a result, judgments may well err on the side of an optimistic scenario.

Considered collectively, these points hint at the role of three underlying factors. First, previous experience of the magnitude and frequency of hazards will influence how people assess the likelihood of possible recurrence, the potential impact on their lives, and the appropriate measures that they consider need to be taken. Extreme events tend to act as a fixed point in experience, obliterating memories of earlier occurrences and acting as a standard against which later ones will be compared, although the poignancy of the recollection will fade if the extreme event happens only very rarely. Second, researchers often argued that personality was a significant variable, with speculative hypotheses that people living in hazard prone zones have personalities similar to gamblers in that they trade off the odds of losses caused by natural hazards against the prospects of profitable and untroubled living. Very little concrete evidence, however, was ever offered in support of such arguments aside from anecdote, with the suspicion that the third factor – the influence of culture – was more important. Culture undoubtedly influences attitudes toward nature, with the responses of different societies to the same form of hazard covering the full spectrum from fatalism in the face of nature to belief in human mastery over it. Equally, response to natural hazards is mediated by culturally transmitted attachments to place. While some writers seem genuinely surprised that anyone should be foolish enough to reside in hazard prone zones, areas such as the sea shore and river basins have had records of settlement that extend unbroken back into antiquity. Symbolic and emotional attachments of these places, coupled with the economic advantages of living there, are more than sufficient to offset the risks involved.

Attachment to Place

Interest in attachment to place was not confined to natural hazards research. Cognitive mapping studies had repeatedly shown the significance of attachment to place by revealing the way in which the home and the home neighborhood served as anchoring points in patterns of spatial knowledge and activity. Attention to attachment to place was also an important point of convergence for the cognitive science and humanistic streams of behavioral geography, in particular by encouraging researchers to explore the emotional content of place, including issues involving belonging, identity, security, and sense of self. This led, in turn, to several wider elements of rethinking that had profound implications not just for behavioral geography but also for human geography as a whole. First, it produced a reappraisal of 'place' at a time when attention to that concept ran counter to the dominant wisdom imparted by the rise of spatial science. Second, it led to the revalorization of two concepts that had dwindled in importance in geographical research: 'landscape', an ambiguous notion originally derived from art that an earlier generation had rejected as a unit of geographical analysis in favor of 'region'; and 'territoriality', a term with little currency outside political geography.

Rethinking on the nature and continuing meaning of place within everyday life was epitomized by Edward Relph's 1976 book *Place and Placelessness*. This examined the lived experience of place at a time where trends in urban and environmental design were fundamentally altering the everyday environment. Adopting a phenomenological viewpoint, Relph pointed to the insensitivity to the significance of place that was causing the casual eradication of distinctive places and the making of standardized landscapes. Taking examples that included the design of tourist resorts, theme parks, malls, commercial strips, new towns, and suburbia, he argued that what was occurring was not just superficial loss of physical distinctiveness but also the loss of meaning, including the replacement of the 'authentic' by the 'inauthentic'.

The question of meaning was also central to work that saw landscapes, past and present, as media that expressed human feelings for and attachment to place. Landscape was conceived as a repository of human endeavor that comprised three elements: the tangible and physical features of an area, the record of human activities, and the meanings imposed by human consciousness (communicated through symbols). It was the third dimension, its symbolic meaning, which gave landscape the character of art in addition to being an artifact and hence made it amenable to the textual and iconographic analyses typical of art history and literary criticism. Essays by writers on both sides of the Atlantic interpreted American and British landscape tastes as reflections of cultural norms, with interest latterly turning to analysis of the meaning of heritage landscapes. Others studied imaginary, fictional, or utopian landscapes, which frequently contained coded longings for the return of vanishing elements of place and landscape. They examined the pleasurable feelings associated with loved places ('topophilia') as well as its converse as expressed in 'landscapes of fear'. Mindful of the debate over the role of instinct as opposed to learning in shaping human cognition, there was speculation that the experience of landscape could be rooted in atavistic traits, where the reason that people find themselves attached to some landscapes and avoid others is taken to have innate roots.

The realm of atavism was even more strongly expressed in research on territoriality. The works of writers such as Edward Hall and Desmond Morris had helped to establish cross species similarities in behavioral patterns, with the suggestion that ethological observations offered at least analogous similarities with the behavioral characteristics found in human society. The principle of territoriality, a behavioral strategy found across a wide variety of species, was seen as having particular promise in this respect. Territoriality was taken here to refer to any form of behavior displayed by individuals and groups seeking to establish, maintain, or defend specific bounded portions of space. While it immediately conjured up visions of aggressive defence, territoriality was more often a fundamental expression of social organization rather than its antithesis. Even when applied to human societies thought likely to need territories as a way of addressing primary needs, (such as hunter gatherers), the anthropological evidence showed that territorial defence was seldom required. Territories were rarely challenged once they are established. Despite often serving deep rooted social and physical needs, therefore, human territoriality was seen as creating a reliable background for everyday life rather than supplying a forum for conflict.

These ideas were relevant when exploring the significance of place attachments in the lives of city residents. Work on the dwelling, for example, identified four needs thought to have a territorial dimension: as a retreat that affords security; as a controlled place in which to balance residents' wishes for stimulation and sociability against their needs for privacy and independence; as a physical framework for the spatial and temporal organization of domestic activities; and as a microcosm that can be molded and ordered to satisfy the need for self expression. In addition, it was suggested that the dwelling served as a territorial core that helped to bolster and reinforce identity (perhaps by being surrounded by known and familiar things, which personalize the living space and convey a sense of the continuity of the present with the past). Ability to alter the exterior of the dwelling could also serve as a medium by which to communicate symbolic messages, although interpretation of those symbols depends upon the culture concerned.

Another subject area for territorial perspectives on attachment to place centered on the neighborhood. Influenced by ecological and community study approaches, which saw specific groups as rooted in particular districts ('niches') of a city, behavioral geographers explored territorial relations of groups in cities where there were intercommunal hostilities. Studies from, *inter alia*, Chicago, Philadelphia, and Belfast showed that residential segregation was matched by activity segregation to the extent that the main criterion in many movement decisions appeared to be the desire to avoid hostile territory. In the absence of major changes in external circumstances, such territorial hostilities were perpetuated through socialization. Children were taught to fear adjacent groups and to avoid their territory at all costs but, except in cases of extreme tension, the notion of defense was interpreted in the broadest sense. People might cluster together for security, but also did so for reasons that include preservation of culture or other shared values. Moreover, research from east coast US cities, which prefigured the modern explosive growth of 'gated communities', showed how the long established 'gilded ghettoes' of the rich were sometimes maintained counter to the general tendency for suburban and ex urban flight – showing that territorial defense could as easily arise from restrictive covenants and symbolic displays of ownership as from walls and overt defensive actions.

Critique

These brief summaries of three representative areas of research within behavioral geography by no means exhaust a field of inquiry that addressed problems of cognition and behavior at all levels from microspace to cosmology in environments ranging from polar tundra to inner areas of metropolitan cities. Not surprisingly given this diversity and the uneasy coexistence of separate schools of thought, there was an endemic climate of critical debate even during behavioral geography's halcyon years. Discussion about the validity of methodology, for example, was pervasive, especially with regard to cognitive mapping. It was commonly assumed that 'maps in the head' may be externally represented by asking people to draw cartographic maps or to rank areas from supplied lists. Yet, in reality, the abilities to draw maps, comment upon cartographic material, or to carry out ranking exercises depend heavily upon education. A subject's responses to mapping exercises may thus imperfectly represent spatial cognition. Moreover, there was no compelling evidence that the idea of a 'map in the mind', which Tolman had used as an analogy, meant that people actually have cartographic representations of the environment in their heads, regardless of the frequency with which that idea was asserted.

Similarly, there was concern about the social context of cognitive behavioralist inquiry. Although, for instance, many observers applauded natural hazards research for its coherent and cross cultural basis and for an approach that consciously built on previous work, others worried about its implicit agenda. A group of radical geographers then based at Clark University argued that it had primarily responded to parochial American interests and needed refocusing to give greater attention to third world nations, which suffered disproportionately more human casualties but smaller loss of property than countries in the 'Western world'. There was also a fundamental objection to the use of the word 'natural' in relation to hazards on the grounds that in many cases it was poverty and landowning policies that forced large concentrations of people to be living in hazard prone zones rather than any deficiencies of knowledge.

The tenor of these arguments accorded with the root and branch criticisms of behavioral geography in general delivered by a group of radical, primarily Marxist, geographers in the 1970s. The earliest coherent critique came from Richard Rieser, then a graduate research student at the London School of Economics. Rieser, who was working on the cognitive maps held by residents of the isolated communities of London's Isle of Dogs in the years prior to the Docklands regeneration, became progressively more disillusioned with the direction of his research. Arguing that it was conceptually and methodologically flawed, he turned his attention to the broader picture to warn that cognitive behavioralism quarried dangerous and dehumanising theories (e.g., ideas about territoriality derived from ethology), led to psychologism (the fallacy by which social phenomena are explained purely in terms of facts and doctrines about the mental characteristics of individuals), and obscured the objective economic and social conditions that operate independently of the individual. Other radical critics, who had equally shown initial support for cognitive behavioralism, denounced its practices on similar lines. Notably, behavioral geography became branded as 'bourgeois thought' that had the ideological function of supporting the *status quo* in that, inadvertently or otherwise, its emphasis on the individual made it unlikely that the data collected would throw light on the true nature of social relations.

The directness of these comments and the standing of their authors as ‘insiders’ in the behavioralist movement meant that the force of the criticisms struck home, even if the extremity of certain pronouncements and, more specifically, their Marxist associations meant that it was possible for their thinking to be treated as the product of political polemic and therefore dismissed. That strategy proved less viable with those geographers identified with the humanistic stream. Similarly well regarded within human geography, these were individuals whose work, a few years earlier, had fitted into the broader behavioralist movement. In the late 1970s, however, they promoted ‘humanistic geography’ as a new *omnium gatherum* that loosely embraced diverse moral and philosophical standpoints, including idealism, existentialism, and phenomenology. Humanistic geographers now attacked the materialist framework of positivism said to underpin much of behavioral geography and disputed the prevailing dichotomies between subject/object and between fact/value. They argued that ‘cognitive’ denoted deliberate and conscious thought processes and excluded consideration of the realm of the ‘precognitive’. Worse still, they argued that behavioral geography’s positivist basis depersonalized and dehumanized the people and places that were studied.

Signs of a schism were readily apparent and given formal expression in 1978 by Derek Gregory’s influential book *Ideology, Science and Human Geography*. Gregory categorized phenomenology as a critique of positivism. Following an agenda which emphasized the importance of structure and agency, Gregory helped to propagate a view that treated ‘humanistic geography’ as a free standing branch of geography, with roots deep in the heart of the discipline. Its erstwhile tactical connections with the cognitive science strand of behavioral geography were firmly severed.

As this interpretation gained support, cognitive behavioralism was treated by default purely as an off shoot of spatial science. Commentators on the development of geography reinforced this view by identifying behavioral geography with the style of scientific analysis reminiscent of mainstream experimental psychology. In particular, they repeatedly conflated ‘behavioralism’ with ‘behaviorism’ – the branch of psychology that reduced behavior to stimulus response and had associations with Skinnerian behavioral reinforcement and modification. The attribution of this dubious moral baggage also supplied grounds for identifying cognitive behavioralism with the worst aspects of scientific practice – reductionist, mechanistic, repetitious, and socially manipulative. According to this version of history, ‘humanistic geography’, the province of maverick geographers, had stood out against this trend and would act as the precursor of the new cultural geography. By contrast, behavioral geography could only look back to its brief heyday as a corrective to specific behavioral deficiencies of spatial science. Although this strategy possessed elements of caricature, there was no denying that, intentionally or otherwise, the new historiography was devastatingly effective in marginalizing behavioral geography.

Continuity and Marginality

Naturally, the process of marginalization was not immediate. Behavioral geography, albeit now defined *sensu stricto* as the ‘cognitive science’ strand, retained its adherents. Its future within academic geography after 1980, however, varied from country to country according to factors such as national research traditions and the extent of institutional support. In Europe, for instance, behavioral geography had only ever enjoyed patchy support outside of the British Isles, Scandinavia, and the Netherlands. In the Francophone areas, for instance, it never had a strong foothold in research or pedagogy despite an intellectual environment that included the heritage of Piaget and Merleau Ponty and drew strength from the cognitive tradition within French psychology. At best, behavioral geography’s relatively brief period of support in the early 1980s was confined to individual researchers based at widely scattered universities, including Lausanne and Geneva in Switzerland, Liège in Belgium, and Montpellier, Rouen, Grenoble, and Strasbourg in France. In Great Britain, behavioral geography had reached its zenith by the mid 1970s. Never strongly represented outside of a handful of major university geography departments (notably University College London, Durham, and Bristol), it fared badly when researchers at these institutions either lost interest or turned to other preoccupations. Pedagogically, too, behavioral geography was ousted from the curriculum during the 1990s by the ‘cultural turn’, which saw the new cultural geography and reinvigorated historical geography courses covering elements of the same agenda.

The situation in the USA was different. The larger size of the academic community and lack of competing traditions – with the entrenched traditional version of cultural geography providing no direct rival – saw behavioral geography continue to occupy a significant place in both research and the curriculum. Specialty groups and the existence of thriving associations that promulgated multidisciplinary links, such as the Environmental Design Research Association, helped to sustain a critical mass of researchers, at least in the medium term. Behavioral geography also retained its presence in geographical education. For example, a 600 page textbook by Reginald Golledge and Robert Stimson, which appeared as late as 1997, offered students a striking new synthesis of material in the mold of the cognitive science stream, with a particular focus on spatial decision making and choice behavior. Yet imposing as the range of material cited might seem, first impressions could be deceptive. Even a cursory review of the contents made it clear that the crucial sources had appeared at least a decade, and often two decades, before that book's publication. In reality, behavioral geography might maintain a niche, but the dwindling amount of new scholarship generated showed that it no longer represented the research frontier of the discipline.

Looking at the content of research since 1980, it is clear that what has appeared has continued long established trends, although with some notable innovations. The early 1980s, for example, saw two significant developments that might have improved the conceptual basis of behavioral geography. The first was an epistemological initiative that favored transactionalism – an approach to cognition which held that the values that people attach to different landscapes evolve from the interactions between people and landscape. The second was time geography. Based on ideas originally developed at the University of Lund by a group of Swedish geographers led by Torsten Hägerstrand, time geography developed the idea of a path that showed the movement of an individual over space and through time, subject to three sets of constraints: capability (including the need for food and sleep); authority (limits to the places where an individual is allowed); and coupling (places where an individual is required to be due to the requirement for interaction with others).

In principle, both perspectives had much to offer. Transactionalism provided a means of drawing together disparate approaches to human–landscape relationships, but came to prominence too late to attract researchers willing to develop the basic approaches through empirical inquiry and critical reflection. Time geography, on the other hand, had considerable indirect influence on geography, especially through its impact on Anthony Giddens' theories of structuration, with Giddens explicitly relating his analysis of the function of localities to time–space paths. By the time that Giddens' work appeared, however, there were few remaining that would wish to develop the underlying principles under the banner of behavioral geography or even to make connection with that area of inquiry.

Rather more success was experienced in specialist areas able to function with a handful of active researchers. Cognitive behavioral studies of children's geographies continued to flourish, yielding impressive amounts of cross cultural research. The topics now developed included the child's access to and use of the outdoor environment, attraction and avoidance behavior, the development of environmental cognition and its pedagogic implications, and policy oriented findings on improvement of children's play spaces. At the other end of the age spectrum, gerontological studies by geographers focused on usage of services and facilities, residential preferences, and the contracting activity space of the elderly, especially in relation to fear and perceived threat. A further cluster of research centered on issues of disability, including real and perceived barriers to access in the built environment, social inclusion, and auditory and haptic environmental interpretation for the visually handicapped. Finally, cognitive mapping maintained its attraction, with the development of highly sophisticated quantitative methodologies and cultivation of linkages with GIS – both of which opened new avenues for multidisciplinary collaboration.

Prospect

These active areas of research reflect at least some degree of continuity with the project that began in the 1960s as a corrective to spatial science and that became an important catalyst for new waves of research that sought to put the 'human' back into 'human geography'. The current outlook for behavioral geography *sensu stricto*, however, reflects its undeniable marginality to the discipline of geography. The loss of connection with the humanistic stream, the relegation of its status through the writing of history, and the unanswered challenge posed by the 'cultural turn' left behavioral geography with the reputation of being outmoded in outlook, peripheral to present day interests, and primarily occupied by an older generation of researchers. It was also undermined by the loss of its once thriving relationship with psychology, which due to the rise of humanistic approaches and fading attractions of environmental psychology, no longer offers the basis for partnership that it did in the 1970s.

Nevertheless, there remains merit in the study of many elements that were once included in the ambit of behavioral geography, regardless of whether or not it has a recognizable future as a separate subdiscipline. Despite the reappraisal of research agendas in the wake of the cultural turn, recent interest in embodiment, sensory geographies, emotional geographies, nonrepresentational theory, actor network theory, performativity, and ethnomethodology, amongst others, has turned attention back, in diverse ways, to human behavior. Many of these developments engage with the territory of behavioral geography without making, or indeed seeing the need to make, any such allusion.

In saying this, of course, there is no suggestion that the typical perspective on human behavior favored by cognitive behavioral researchers – still heavily influenced by the intellectual climate of the 1960s – directly accords with more recent concerns with 'discourse', 'practices', or 'performance'. The shifting currents of ideas over the last half century have done more than just change the lexicon of terms deployed when studying the material and symbolic relationships between people and their en vironments, since they also challenge the conceptual frameworks in which human agency is situated, question the methods by which information is gathered, and invite scrutiny of the positionality of the researcher when gath ering data. Yet it is undeniable that there is only so much that can be learned about spatially embedded everyday practices by forms of analysis that proceed primarily at the group level. As such, there is still a need to reintegrate consideration of the individual and the cognitive back into geographical inquiry, even if pursuit of that goal will not prove an easy task. Certainly, contriving any form of rapprochement between cognitive behavioralist and cul tural perspectives will require a great deal more attention to building conceptual foundations than behavioral geography exhibited during its heyday.

See also: Cognitive Geography; Feminism/Feminist Geography; Landscape Perception; Map Perception and Cognition; Spatial Science; Time Geography.

Further Reading

Appleyard, D. (1976). *Planning a Pluralist City: Conflicting Realities in Ciudad Guayana*. Cambridge, MA: MIT Press.

Appleyard, D., Lynch, K. and Myer, J. (1964). *The View from the Road*. Cambridge, MA: MIT Press.

Burton, I., Kates, R. W. and White, G. F. (1978). *The Environment as Hazard*. New York: Oxford University Press.

Cox, K. R. and Golledge, R. G. (eds.) (1969). *Behavioural Problems in Geography: A Symposium.* Studies in Geography, 17. Evanston, IL: Northwestern University.

Gold, J. R. (1980). *An Introduction to Behavioural Geography*. Oxford: Oxford University Press.

Gold, J. R. (1992). Image and environment: The decline of cognitive behaviouralism in human geography and grounds for regeneration. *Geoforum* 23, 239 247.

Golledge, R. G. and Stimpson, R. J. (1997). *Spatial Behaviour: A Geographic Perspective*. New York: Guilford Press.

Goodey, B. (1971). *Perception of the Environment*. Occasional Paper 17, p 7, Birmingham: Centre for Urban and Regional Studies, University of Birmingham.

Gould, P. (1966). *On Mental Maps*. Discussion Paper 9, Ann Arbor: Michigan Inter University Community of Mathematical Geographers.

Gould, P. and White, R. R. (1974). *Mental Maps*. Harmondsworth: Penguin.

Gregory, D. (1978). *Ideology, Science and Human Geography*. London: Hutchinson.

Johnston, R. J. and Sidaway, J. D. (2004). *Geography and Geographers: Anglo American Human Geography since 1945* (6th edn.). London: Hodder Arnold.

Kitchin, R. and Blades, M. (2002). The cognition of geographic space. I B Tauris, London.

Lowenthal, D. (1961). Geography, experience, and imagination: Towards a geographical epistemology. *Annals of the Association of American Geographers* 51, 241 260.

Lowenthal, D. (1967). *Environmental Perception and Behaviour*. Research Paper 109, Chicago: Department of Geography, University of Chicago.

Lynch, K. (1960). *The Image of the City*. Cambridge, MA: MIT Press.

Relph, E. (1976). *Place and Placelessness*. London: Pion.

Rieser, R. L. (1973). The territorial illusion and behavioural sink: Critical notes on behavioural geography. *Antipode* 5, 52 57.

Saarinen, T. F. (1966). *Perception of Drought Hazard on the Great Plains*. Research Paper 106, Chicago: Department of Geography, University of Chicago.

Saarinen, T. F. (1969). *Perception of the Environment, Commission on College Geography*. Resource Paper 5, Washington, DC: Association of American Geographers.

Tuan, Y. F. (1974). *Topophilia: A Study of Environmental Perception, Attitudes and Values*. Englewood Cliffs, NJ: Prentice Hall.

Wolpert, J. (1964). The decision process in spatial perspective. *Annals of the Association of American Geographers* 54, 537 558.

Relevant Websites

http://www.edra.org
Environmental Design Research Association.

http://epbg.blogspot.com
Environmental Perception and Behavioural Geography Specialty Group, Association of American Geographers.

http://www.iaps association.org
International Association for People Environment Studies.

http://www.landscaperesearch.org
Landscape Research Group.

Belonging

A. Taylor, University of Canberra, Canberra, ACT, Australia

Glossary

Counter-Hegemonic Identities Those that challenge the normalization of dominant identities.

Diaspora Refers to a critical mass of dispersed and displaced nationals.

Dissensus Refers to the productive tensions that result from dialog around rather than suppression of internal group differences.

Essentializing Identity A way of making sense of identity that assumes it to be intrinsic, innate, predetermined, fixed, natural, and normal. Essentialist approaches to identity tend to privilege scientific explanations of biological causes over social and cultural considerations. Essentialist approaches to identity can also draw upon spiritual authority.

Identity Politics Refers to the spectrum of power relations, negotiations, and struggles that take place within and between groups based on sociocultural differences (such as gender, ethnicity, socioeconomic status, religion, etc.).

Multiscalar Refers to the articulation of global, regional, national, and local scales. A multiscalar perspective draws attention to the ways in which each scale is simultaneously constituting and constituted by the other.

Palimpsest An ancient manuscript that has been overwritten. Layers of previous underwritings remain in varying degrees of erasure and visibility. It is often used as a metaphor for the transformative relationship of culture and place across history.

Politics of Difference Refers to the relations of power that structure difference, often through the construction of unequal binary categories. It also refers to the generation of new forms of difference through the contestations of these binary orders and their power regimes.

Postcolonial Used more as a critical reflection on the ways in which the colonial past inheres in the present than as an indicator of a complete temporal break with bygone colonial eras.

Relational Ethics A philosophy that stresses human interdependence and a responsibility for others. It challenges us to think beyond the notion of the self-directed autonomous individual who exercises complete agency.

Territorialization A term that underscores the political currency of territory. It refers to the ways in which territory is both invested with political significance and mobilized (defended, demarcated, staked out, etc.) within power struggles.

Transnationalism It is both a philosophy that promotes the transcendence of national interests and a form of international collaboration.

The Politics of Belonging

The fact that geographers now consider belonging as an issue of disciplinary interest is indicative of the increasingly fluid, uneven, and above all contested terrains of identity politics and social relations in the early twenty first century. Along with the unsettling of fixed and coherent identities associated with the relatively predictable concerns of early modernity, it is becoming more and more rare for belonging to any group, place, or community to be unquestioningly accepted as an unproblematic given. Late (or post) modernity is now characterized by mobile, multiple, and hybrid identity formations, the complexities of which belie a simple membership formula. Belonging to any social group is now much more likely to be considered as a political negotiation than a predetermined state.

This shift toward considering the political dynamics of belonging represents a deepening engagement with the politics of identity and the politics of difference. This is not to assert that belonging, like identity, has only just become a field of power because it is becoming increasingly unstable and complex. Even those societies typified by stratification and regulated by the assumption of natural order, of everyone being in their rightful place, were (and in some cases still are) inherently political in their naturalization of social inequality and their prescriptions of social belonging. However, it is the increasing challenges to these hitherto assumed to be natural orders and their corresponding unearned entitlements and structural inequalities that trigger public dissent over identity and throw wide open the question of belonging.

There is a significant productive momentum to these intensifying debates over the naturalness of predetermined social orders, the essentializing of their corresponding identity categories and structured patterns of belonging, which fuel the grounds of identity politics. Dissensus over identity and belonging actually creates new spaces of possibility for the emergence of differences, for the generation of new identity formations, as well as new modes of belonging. Not all of these are geographically located. New social identities are constructed in the process of resistance, dissent, and even parody. Some of them are hybridized identities (for instance, transnational tribal youth identities and metrosexual identities) or based upon a resistance to the essentializing of difference (for instance queer identities). New modes of belonging are forged through developing counter hegemonic identities through affinities or affili ations with people in other places.

The politics of belonging are characterized by recurring debates over questions such as – who belongs where; who decides who belongs where; on what basis is this belonging determined; who is considered to be in place or out of place; and who is authorized to represent place and community. Questions such as these draw attention to the ways in which claims and counterclaims over rights to belong take place within an ongoing series of related material and symbolic struggles. Struggles over the own ership, stewardship, occupation, and management of real places are often connected to struggles over the expression, status, recognition, and authenticity of group identities, as well as to struggles over membership and representation within both formally constituted groups and more loosely imagined communities. There are correlations between social power, material wealth, and the authorization of belonging on the one hand and contestations over social marginalization and exclusions on the other.

Many of the inclusions and exclusions that characterize struggles over belonging are underpinned by the socio cultural politics of difference and played out across the intersections of racialized, ethnic, socioeconomic, religious, sexual, and gender identities and affiliations. This means that the experiences of inclusion/belonging or exclusion/ not belonging are directly related to sociocultural iden tities and their differential social positionings. Scales of belonging come into play here. Those who experience a sense of exclusion from larger social formations, for ex ample, those who feel alienated from the national com munity on the basis of race or ethnicity, might well feel a heightened sense of affinity with others who are similarly excluded. Their sense of belonging is thus likely to be more tightly focused upon a minority group identity. Thus ironically, oppositional, marginalized, or resistant identity based belonging is at least partially constituted and framed by the shared experience of social exclusion.

Although the politics of belonging takes place across the scales of local, national, and global affiliations and alliances, the inclusionary and exclusionary processes of citizenship and nation building are of particular interest to geographers. This is because the acceleration of globalization, increase in diasporas, and rise of transna tionalism (for instance the proliferation of multinational corporations) in late modernity have been perceived as a potential threat to the sovereignty of the modern nation state. Some governments have responded by renewed efforts to fuel nationalism and firm up national boundaries. Studies of the macro politics and micro practices of national belonging throw light on these effects and responses. Consideration of the intersections, layer ings, and potentially conflicting loyalties between national identities on the one hand and ethnic, class, racialized, gender, sexual, and religious affiliations on the other, further fleshes out the complexities and multiple posi tionings involved in a politics of belonging.

Locating Geographies of Belonging

Like many of the currents of contemporary social issues, the politics of belonging is not exclusively studied within human geography. Neither is it represented within this, or any other discipline, as a discrete body of work. Rather, issues of belonging emerge from a slip stream of inter disciplinary conversations. They are threaded within and across the related topics of identity, citizenship, migra tion, globalization, postcolonialism, and transnationalism; and as such form part of a fabric of intellectual endeavors to understand the nature of contemporary societies within a rapidly changing world. Across these inter secting themes, studies of belonging within human geography have been both informed by and themselves informed thinking across a number of disciplinary fields, such as history, sociology, anthropology, political science, cultural studies, feminism, psychoanalysis, postcolonial studies, and multicultural studies. Embedded within this milieu, human geography nevertheless offers a particular set of analytic lenses for retheorizing belonging that distinguishes it, to some extent, from other disciplinary perspectives. The position that human geography takes on matters of belonging is necessarily informed by its disciplinary preoccupation with theorizing space, place, and nature.

Picking up on 'the spatial turn' most often associated with Michel Foucault's situated genealogies, human geographers have continued to theorize the ways in which space decenters or disperses history, or in other words, the difference that geography/space/place makes. Such work is exploring the intrinsic spatialities of power and the power relations that flow between space, place, and difference.

When considering the politics of belonging, therefore, it is not surprising that geographers foreground its spa tialities: the physical, as well as the identity related or subject position based proximities and distances that shape our sense of belonging.

This involves maintaining a steady focus on the spatial and sociocultural differences that fragment and particu larize experiences of belonging. It also involves con sidering the ways in which difference based, as well as geographically based proximities can spawn new constellations of spatial association, or new modes of belonging. Human goegraphy's spatial foregroundings thus help us to better understand the diversity of conditions, manners, and positions through which we negotiate belonging.

Two arenas in which human geography has foregrounded the spatiality of belonging are taken up below. The first of these is the arena of identity and place. Cognizant that the spatial turn has widely influenced thinking across the social sciences and humanities, human geographers have taken a lead in interrogating the politics of place and the place of identity and belonging within this politics. The second arena is that of globalization. Within widespread cross disciplinary debates about the significance of globalization to belonging, and in particular the relationship between global, national, and local concerns and loyalties, human geographers have closely charted the spatialities of our multiscalar world.

Place, Identity, and Belonging

Geographers approach the politics of place and relate it to identity and belonging in a double move. First, the proliferation of situated studies of identity events in different places provides an ongoing insistence and reminder that place matters. Underlying this insistence is the understanding that things turn out differently in different places, or in other words, that place makes a difference. The second move is to problematize place. Rather than accepting places as self evident and static backdrops, sets, or stages on which human events happen, places are recast in dynamic relationship with the people that inhabit them. The very materiality of place is seen as both acted upon by humans and affecting human action. Much attention is similarly drawn to the dynamic ways in which place exceeds its physicality, becomes an idea or an imaginary construct and in this way is culturally construed. Such interrogations of the human interface with place underscore the ways in which place and identity have come to be regarded as mutually constituting, and the politics of belonging are implicated in the place/identity making process.

Most of us come to know ourselves at least partially through some level of identification and relationship with the places we inhabit. For some, however, place is centrally defining, or even elided with identity. Such is the case of many indigenous people whose conceptualization of human, nonhuman animal, plant, and physical landscapes are often inseparably interdeterminant. Owing to the increasing mobility of many people (including those precipitated by the forced removals and dispossessions of indigenous peoples), the connections between place and identity might arguably diminish. An alternate view is that the experience of displacement can function to nostalgically sharpen this identity/place relationship, and to render it much more complex. Although fractured by distance and time, embodied memories of lost or left places persist and through narration, these memories can be passed on as part of the cultural maintenance work of identity. As in the case of many migrant groups, cultural practices and associations with places of origin are often strategically maintained. For many people, place still matters a lot.

History also reminds us of the gravitas and high stakes of place and belonging. There is no denying that struggles over place have always and still do evoke harsh materialities. Real places are not only the sites but also the subjects/objects of corporeal struggles. Human lives are lost, built and natural environments are damaged or destroyed in physical struggles over the ownership, custodianship, and governance of homelands and territories. On all continents there are plentiful examples of bloodletting identity related struggles over real places, such as those that take the form of ethnic cleansing, colonial invasions, wars of independence, neighborhood race riots, etc.

At the same time, struggles over place and belonging are also symbolically enacted. An oft cited example of this is the powerful stroke of the imperial cartographer's pen: charting, naming, and inscribing 'new world' places from a specifically Eurocentric viewpoint. The claiming of territory, as in this example of mapping as a discursive imperial strategy, can be as much a symbolic act of power as an act of physical force. All places that are struggled over are somehow mediated. Like palimpsests, they are symbolically inscribed with layers of sociocultural meanings: made, erased, and remade through consecutive, as well as multiple concurrent stories and imaginings. In their symbolic making, in the process of being inscribed with meaning, they are also being claimed by their meaning makers. The multiplicity of imaginings of place, reflecting the variant partialities, interests, and investments of their meaning makers, constitute the symbolic grounds of struggle over place and belonging.

Imaginings of place are thus entwined on a number of levels, with understandings of self and of one's place in the world. Nations, for example, although bordered geographical territories, are much more likely to be imagined, idealized, and abstracted as a set of iconic characteristics than they are to be exclusively evoked as a geographical landmass. These abstracted characteristics of the nation are also associated with the identity of its people.

It is well recognized that national identity is framed by a sense of belonging to the imagined space of the nation or to the imagined national community. As mentioned earlier, however, nominal group membership alone, which in this case is citizenship, does not necessarily guarantee a strong sense of national identity and belonging. Within the spatialized imaginary of the nation, certain citizens tend to be regarded or regard themselves as belonging more or less than others. This is linked to another aspect of the national imaginary, and that is the personification of the nation. The nation is personified through the elisions between the character or identity of the national subjects and the place itself. This personification is mobilized in common references to the nation having a 'heart' or a 'core' essence. Spatially abstracted and relatively measured, certain citizens will see themselves and others as more or less centrally or peripherally positioned in relation to the 'heart of the nation' or its 'core'. This differential regarding of some citizens as somehow more authentically national than others, or closer to the 'real' essence of the nation, arises out of the spatialized and naturalized imaginings of self and nation.

The territorialization of this national imaginary results when those who position themselves as core nationals, and therefore self authorized to manage its spaces, set out to delineate and patrol designated spaces – ranging from small locales to much larger scopes of territory. This border maintenance work is undertaken in order to regulate the national space and exclude or expel those who are perceived to be out of place in it. Exclusionary practices, for example, might be found in the local school community's marginalization of same sex families, as well as in the range of measures governments take to prevent asylum seekers crossing national borders. Expulsions, which resemble purification rituals, can be witnessed in neighborhood vigilante attacks, as well as in the relocation of public housing tenants from rapidly gentrifying urban areas. Moral panics are often generated in order to justify such incidents of expulsion and exclusion, closely articulating the notion of natural or authentic belongings with idealized imaginings of pure and vulnerable territory.

Globalization and Belonging

Consideration of the impact of globalization on our sense of belonging in the world necessitates a multiscalar analysis of concurrent and yet radically uneven processes that take place across global, national, and local contexts.

For instance, it is increasingly common for the compression of time and space (a key feature of globalization that geographers address) to be linked with a shift away from a primary identification with local places and toward a more worldly and abstracted sense of locatedness and connectivity. The premise here is that the experience of the world simultaneously shrinking and speeding up fundamentally alters our sense of belonging within it. Such a shift is significant, but far from universal and complete. In fact, shifting scales of belonging remain contingent, relative, and subjective; and intractably linked to material resources, wealth, status, and power. A confident sense of global connectivity and belonging is arguably much more indicative of privilege than it is of a widely shared experience of globalization. Associated with the large scale imagining of global citizenship, global belonging is most likely to be assumed by those privileged few who have continual access to global information and communication technologies, are well networked within transnational flows of information and capital exchange, and who productively exercise a high level of international mobility. On an imaginary, as well as an everyday lived register, most people occupy much smaller horizons of belonging.

On the other hand, there are very few people whose identities remain completely unaffected by Western dominated global communications. Through such media, foreign cultural influences also permeate the lives of those who remain well outside of the profitable circuits of global exchange. One example of this is the popularity and emulation of African American urban 'hip hop' music and fashion amongst some Indigenous Australian youth living in remote desert communities. Transnational cultural exposure such as this, while clearly altering the identity repertoires and esthetic desires of these youth, is unlikely to engender in them a sense of belonging that is akin to African American youth in the Bronx, or in any way comparable to that of the world's privileged 'global citizens'. A global sense of belonging does not automatically follow from the adoption of selected foreign cultural practices. While such transnational subcultural borrowings may enable these young Aboriginal men to express their indigenous and black identities in new ways, it is unlikely to alter their fundamental sense of place and belonging to their traditional desert country and communities. Despite the penetrating scope of cultural globalization, and the new kinds of hybrid identities that result from it, transnationally inflected identities continue to be expressed within distinctively local formations. For most of the world's population, local places and communities remain the locus of belonging.

These kinds of wide variations in the scope, reach, and experience of globalization point to its radical unevenness and inherent contradictions. The rapid increase in the numbers of people moving across national borders is another often noted feature of globalization. Nowhere are these contradictions and differences more marked than within the demographies of concurrent but radically different kinds of cross national border movements. Ranging from global citizens such as transnational executives, to migrant workers, to stateless refugees, these very different kinds of diasporas represent the continuum of privilege and inequality that characterize the spatialities of belonging. While global citizens exercise their mobility as a form of agency on a world wide stage, assuming to belong everywhere; and international migrant workers seek to enhance their agency through relocating to secure better life opportunities, trading displacement for the possibility of establishing a more productive locus of belonging; refugees are compelled to move around the world as stateless people, seeking asy lum from foreign states but belonging nowhere. Such stark contrasts in the agencies of global mobility under score the harsh paradoxes of global belonging.

These same paradoxes also reveal the pivotal dis criminatory function of the nation state within the global system of relations. In a historical moment when some predict the undermining of national sovereignty and the withering away of the nation state, selective national gatekeeping is in high gear. The spatialities of this gatekeeping process are strategically territorial and draw heavily upon a geographic rhetoric. National borders are porous when it comes to participating in the global market economy, are selectively managed when it comes to satisfying national population and labor force interests, and are protected when it comes to regulating or prohibiting the entry of asylum seekers.

When justifying the control of refugee movements across national borders, this rhetoric is explicitly ex clusionary. For example, through the tags of 'legal'/ 'process compliant' and 'illegal'/'queue jumpers', refu gees are differentially positioned as either worthy victims to be rescued by the host nation state or undesirable threats to its security. The rhetoric of scale is also utilized as further justification for exclusionary practices, for example, when inferences of being 'swamped' by unauthorized and uninvited refugees mobilize majority national fears about being outnumbered. The reassertion of national sovereign rights to actively exclude asylum seekers is inscribed through a new spatial lexicon full of terms such as 'firming up national boundaries', 'border protection', 'border patrol', and 'excision of territories for the purposes of immigration'. Those who do manage to evade border patrols and cross national borders are then criminalized and isolated through the reiteration of such policies as 'mandatory detention'. In this way, they can remain as demonized outsiders even while located within the nation. The geographies of selective inclusion and strategic exclusion are fundamental to the ways in which nation states manage the politics of national belonging in an increasingly mobile world.

Of course, mass movements of people across the world predate the current rise of migrant and refugee diasporas. Most significantly perhaps, legacies of the colonial era (from sixteenth to nineteenth centuries) still inhere in the present and continue to shape the geographies of be longing. Present day struggles over global belonging are not only the product of twentieth and twenty first century immigration and warfare displacements and diasporas. The reconfiguration of populations caused by colonial invasions and indigenous dispossessions cen turies ago continue to be felt in the present through movements within and between postcolonial nations and their ex imperial centers. For postcolonial nations, the politics of belonging are infused with questions about the recognition of indigenous custodianship and ownership of land; its current legal status and stewardship; and the responsibilities of current governments to reconcile and compensate for the dispossession and disenfranchisement of indigenous peoples. The question of postcolonial na tional belonging is complicated and problematized by such questions. Furthermore, many indigenous or first nations people's appropriation of nationhood as a tribal affiliation (for instance as members of the Cherokee Nation or the Wiradjuri Nation) directly challenges the status of citizenship as the primary determinant of national loyalty and belonging.

The displacements and dispersals that constellate in all local places that have been effected by geohistorical macro events, such as colonization and migration, render identity increasingly diffuse and complex. Being born in one country and growing up in another, for instance, might result in a dual or hyphenated national identity. Being a third generation child of immigrant parents from different countries of origin, or being of mixed immigrant and indigenous heritage might result in the foregrounding of one of these identities or in the adop tion of a hybridized identity. Increasingly fragmented and multiple ethnicities and racialized identities produce multiple affiliations and allegiances and thus fracture and disperse loyalties. Globalization, in this sense, can be seen as spawning hybridized and multiple identities that de stabilize unitary notions of national belonging. This raises the possibility, at least, for new forms of transnational or postnational belonging that no longer cohere in one na tion state, but articulate across the local, national, and global scales of belonging.

Conclusion

Recently human geographers have been moving toward developing a relational ethics of belonging. This involves focusing upon the relationships that are produced through complex networks of global/national/local associations between people, nonhuman animals, places, events, and things. Acknowledgment of the interplay between all these facets evokes a new conception of belonging, one that is network based and foregrounds interdependency. A relational ethics also involves paying more attention to the corporeal and affective dimensions of these relationships. It recognizes that belonging is experiential and driven by desires for sociality and connection.

Because belonging is an embodied, relational, and affective experience, it exceeds nominal designations of group membership. This goes some way to explaining why citizenship, for instance, although an important prerequisite, is not enough to ensure a sense of belonging to a nation. A sense of belonging is based in the lived experience of inclusion and affinity, and undermined by experiences of exclusion, marginalization, and rejection. The gauge of belonging is a sensory and emotional one, based in relationship and measured through comparative perceptions and judgments of the inclusions or exclusions of others.

Thus the final thing to note about belonging is that it is always an affective and embodied form of politics. It cannot be adequately addressed without acknowledging our visceral responses to social inclusions and exclusions, to our geohistorical legacies, to the places we inhabit and leave, and to the efforts of negotiating the power relations of our complex everyday lives.

See also: Affect; Body, The; Citizenship; Embodied Knowing; Migration.

Further Reading

Anderson, B. (1983). *Imagined communities: Reflections on the origins and spread of nationalism*. London: Verso.

Appadurai, A. (1996). *Modernity at large: Cultural dimensions of globalization*. Cambridge: Polity Press.

Anderson, K. and Taylor, A. (2005). Exclusionary politics and the question of national belonging: Australian ethnicities in multiscalar focus. *Ethnicities* 5(4), 460 475.

Castles, S. and Davidson, A. (2000). *Citizenship and migration: Globalization and the politics of belonging*. London: Macmillan Press.

Jacobs, J. M. (1996). *Edge of empire: Postcolonialism and the city*. London: Routledge.

Keith, M. and Pile, S. (eds.) (1993). *Place and politics of identity*. London: Routledge.

Massey, D. (1993). Power geometry and a progressive sense of place. In Bird, J., Curtis, B., Putnam, T., Robertson, G. & Tickner, L. (eds.) *Mapping the futures: Local cultures, global change*, pp 59 69. London: Routledge.

Migdal, J. S. (2004). *Boundaries and belonging: States and societies in the struggle to shape identities and local practices*. Cambridge, UK: Cambridge University Press.

Pile, S. and Thrift, D. (eds.) (1995). *Mapping the subject. Geographies of cultural transformation*. London: Routledge.

Probyn, E. (1996). *Outside belongings*. London: Routledge.

Sibley, D. (1995). *Geographies of exclusion: Society and difference in the west*. London: Routledge.

Thrift, N. (1999). Steps toward an ecology of place. In Massey, D., Allen, J. & Sarre, P. (eds.) *Human geography today*, pp 295 322. Cambridge: Polity Press.

Westwood, S. and Phizacklea, A. (2000). *Trans nationalism and the politics of belonging*. London: Routledge.

Whatmore, S. (2002). *Hybrid geographies: Natures, cultures, spaces*. London: Sage Publications.

Relevant Websites

http://www.geog.leeds.ac.uk
Citizenship and Belonging, Research Clusters, School of Geography, University of Leeds.

Berkeley School

M. Williams, Oxford University, Oxford, UK

Glossary

Culture The objects and ideas of humans through time.
Cultural Ecology Micro-level studies of human adaptation and their cultural context, informed by theory.
Cultural Landscape The natural landscape transformed by human activity.
Determinism The philosophical idea that human actions are conditioned and governed by physical geographical conditions.

Defining the Berkeley School

The Berkeley School is a collective term applied to the geographers and their publications that emanated from the work of Carl Sauer (*quod vide*) and the colleagues and postgraduate students taught by him during his long term as head of the department of geography at Berkeley from 1923 to 1955. By extension it must also include those taught by Jim Parsons, a student and colleague of Sauer's, and subsequently head of the department from 1960 to 1966 and from 1975 to 1979.

A hint of the flavor of the 'School' comes from the names of some of the 27 postgraduates who completed their PhDs under Sauer. They included Homer Aschmann, Robert Bowman, Jan Broek, Henry Bruman, Andrew Clark, George Carter, Leslie Hewes, Fred Kniffen, John Leighly, Marvin Mikesell, Jim Parsons, Joseph Spencer, Fred Simoons, David Sopher, Dan Stanislawski, Philip Wagner, and Wilber Zelinsky. Of the 38 PhD recipients under Parsons, we can include William Denevan, Robin Doughty, David Harris, Paul Starrs, Thomas Veblen, Bert Wallach, and Karl Zimmerer. Kindred spirits come in the names of David Lowenthal who received his Master's degree at Berkeley, Yi Fu Tuan who received his PhD in geomorphology under Kesselli, and Clarence Glacken, the author of the unsurpassed and magisterial *Traces on the Rhodian Shore* (1967), written when a member of the department and thoroughly enmeshed in the Sauer tradition, though with the new dimension of classical scholarship. The School could be said to have ended with the death of Parsons in 1997, or perhaps sooner with his last PhD in 1994, or it may have lingered on until Barney Nietschmann's death in 2000.

Sauer had an original, independent, and wide ranging mind, and did not adopt a doctrinal position and expect his students to tread a 'narrow path'. But it was inevitable that they were influenced by his broad horizons and attitudes, especially as he encouraged them to take courses in kindred departments like biology, plant sciences, anthropology, Latin American history, and palaeography, as well as foreign languages in order to read German, French, and Spanish geographical and historical works in the original. They were made well aware of the antiquity and continuity of geography. Moreover, he was an inspiring and persuasive teacher and supervisor. Consequently, they focused on the relationship between human communities and the natural world, investigating the transformations of the natural landscape into cultural landscapes, especially in Mexico and Lower California, with American Indian cultural survivals, plant and animal domestication and diffusion. There was an emphasis on the elements of material culture that gave character to an area. None focused on Sauer's other and later interests, for which he is probably more well known, such as land degradation, the entry of early man into the Americas, global plant and animal domestication, or the impoverishment of the cultural and natural environment through colonial expansion, and modern imperialism.

His lifelong colleague in Berkeley, John Leighly, said that he "always disclaimed any intention of founding a 'school' or of shaping students over a common last. He expected the individual to find his own objects of interest and his own ways of satisfying his curiosity." Jim Parsons said, "There was no 'program', no formal requirements that graduate students were aware of.... We were on our own, an assorted group of aspiring scholars immersed in our work, seldom questioning or concerned about how geography might be defined by others." All this was perfectly in accord with Sauer's deeply held conviction that the individual scholar knew best what questions were worth asking. What Sauer did require, however, was that 'a thesis must have a thesis'; in other words, it could not be an endless compilation of facts (an important consideration in the regionally orientated geography of the 1920s–50s) but must formulate a problem or unanswered question, the solution to which came through thorough archival work and field work to see what the landscape and its inhabitants offered. Parsons was even more freewheeling than Sauer (if that were possible), and equally concerned with subjects in cultural and historical geography, the human impact on the environments, biogeography, and especially Latin America, but with a far greater emphasis on the contemporary.

The catholicity of the School is emphasized by the fact that it contained studies on urban geography by John

Leighly and Warren Thornthwaite, and ranged in location well away from Mexico and Baja California and into, for example, New Zealand (Andrew Clark), Southeast Asia (David Sopher), Morocoo (Marvin Mikesell), Ethiopia (Fred Simoons), and Kenya – Tanzania (Lee Talbot), together with five studies in and around the Caribbean. The commonality of these studies was characterized as much by their nontheoretical approach, intellectual ini tiative, direct observation, and 'a good honest job of re porting' as by topic. All were major works of scholarship that synthesized impacts of humans on the environment.

Sauer's Ideas

The importance and influence of the School was out of all proportion to the work actually done. It has been suggested by Cole Harris that this was partly because Sauer and his associates were among the very few groups who cultivated 'an intellectual environment that kept the past open' in American geography during its anti historical, regional, and positivist decades. It is difficult to pigeonhole Sauer and be sure of his methodology be cause he seems to dart about the geographical scene like some intellectual Voortrekker who moved on when he saw the other man's methodological smoke. He recog nized his own shifting position; "I have the idiosyncrasy that once having written something I do not refer to it again myself, except to refresh my mind as to statements of fact…I thus escape commitment to previous opinions and conclusions of mine and am therefore not obliged to defend my past self," he told Kinvig in 1953.

The 'idiosyncrasy' was not difficult to understand. Like any contemplative and wide reading scholar of worth, he clarified, extended, and redefined his ideas during the long 64 writing years of his 85 years of life. Thus his own continuing intellectual development, together with his distaste for the increasing professionalization of American academic life, and disenchantment with aspects of American society and world events, all affected his writing. The world and its many questions were simply too exciting and interesting to commit oneself to a single approach over a long period. Second, he disliked meth odological writing. He told Richard Hartshorne that he was not competent in, nor did he like, logic and its application in dialectics. Any such writing he did was pragmatic; merely a 'sort of habilitation', in order to work his way through a literature and draw attention to topics that were worth looking at.

Nevertheless, he wrote so authoritatively and engagingly that one of his earliest 'habilitations', 'Morphology of landscape' (1925), which was basically a rebuttal of the then prevailing environmental determin ism as an explanation to human action, set American human geography off on a new road of exploration. 'Morphology' has been taken by some as the hallmark statement of the Berkeley School, especially its oft quo ted dictum, "the cultural landscape is fashioned from a natural landscape by a culture group." But it is doubtful if any students other than perhaps Jan Broek in his study of the Santa Clara Valley, California, ever attempted to do simply that. Sauer was far more concerned that his stu dents engaged in voyage of discovery and exhibited curiosity, restricted only by competence, as to what a landscape offered, and employ a methodological plural ism. 'Morphology' certainly emphasized the role of cul ture(s) in the landscape, and was probably the first introduction of the German concept of a 'cultural geography' into the English speaking world.

But the most abiding characteristic of the School was its emphasis on human processes of change – 'evolution istic', 'genetic', and 'historical' were all words that he used to describe this. His 'Foreword to historical geography' (1941) was another of his 'successive orientations' as he pursued his wider theme of the study of humans on earth through time. For Sauer, all geography was essentially historical, and human geography had little meaning and few valid labels unless it dealt with cultural origins, and the mechanisms of their spread, growth, and extinction.

Parsons confirms that 'morphology' was seldom re ferred to in Berkeley, least of all by Sauer, who was convinced that several Midwestern geographers must have spent more time reading it than he had in writing it. It saddened him that few seemed to have taken equal notice of his substantive historical works on Baja Cali fornia and Mexico that had more thoroughly engaged him throughout the 1930s and early 1940s. Indeed, it was Sauer's discovery of the Hispanic lands to the south, and his cooperation with the anthropologists Alfred Kroeber and Robert Lowie, which gave his work and that of the students who accompanied him in the field such a marked Central American tinge. These included Peveril Meigs (1932), Donald Brand (1933), Henry Bruman (1940), George Carter (1942), Dan Stanislawski (1944), Robert West (1946), Philip Wagner (1954), Homer Aschmann (1954), and Brigham Arnold (1954).

After 1954, Sauer tended to switch focus when he be came the advisor to the Office of Naval Research (ONR), and used its funds to support postgraduate work in the Caribbean and surrounding lands, for example, Edwin Doran (1953), Burton Gordon (1954), Gordon Merrill (1957), Donald Innes (1958), and Carl Johannessen (1959).

The Parsons Era

It was inevitable that every one of those students who passed out under Sauer took something of his approach and attitude with them. Some went on to found pres tigious postgraduate departments of their own, for example, Fred Kniffen in Louisiana State University, and, particularly, Andrew Clark at Madison, Wisconsin, who employed a very similar attitude to Sauer on the importance of the historical approach and the need for cross disciplinary approaches. The roll call of his students is impressive in its own right and included Roy Merrens (1962), David Ward (1963), Cole Harris (1964), James Lemon (1964), Terry Jordan (1965), Sam Hilliard (1966), Jim Gibson (1967), Clarissa Kimber (1969), and Robert Mitchell (1969). But to include them within the definition of the Berkeley 'School' would be tenuous and misleading.

However, the students that emanated from the teaching of Jim Parsons, a graduate student of Sauer's, who completed his work on the colonization of western Colombia in 1948, and stayed on in Berkeley as professor and then head of department, are well within the definition of the Berkeley School.

Parsons admired Sauer but had too independent a mind to be his intellectual clone. He certainly did work on ancient agricultural systems, agricultural origins and dispersions, but these were not the main emphases of his work. Parsons' studies were generally focused on the present, but with a good appreciation of what had gone before to produce that 'present'. He also encouraged a more 'cultural ecological' approach. Prolonged field research, far greater in breadth and depth than Sauer's, in a variety of locations and environments was a hallmark of his work, as was also his tendency in later years to return to old field sites in order to chart changes and transformations. A good example of his innovative approach was his 1972 paper 'Spread of African pasture grasses to the American Tropics'. The beginning was pure Sauer in that it traced the origins and spread of Guinea, Para[/], Molasses, Jaragua[/], Kikuyu, and Pagola grasses from Africa to Latin America from the seventeenth century onward. But the conclusion was ground breaking as the paper was probably the first alarm call about the extent and rapidity of Central and Latin American tropical deforestation, about 20 years before the term 'tropical deforestation' had even been coined and became the 'hot' environmental topic it is today.

"The once limitless forests of humid tropical America are rapidly being converted to grass," he said, and drawing on the changes he had seen on his many field visits in previous decades he continued,

> Where once stood great tracts of lowland forest along the Pan American Highway in Mexico and Central America, on the northern coast of Colombia, on the Andean spurs of eastern Venezuela, in the interior of Brazil, on the islands of the Greater Antilles today one sees pasture lands stretching to the horizon, interrupted only by scattered palms, remnant woodlots or rows of trees planted originally as live fences. (Parsons, 1972: 12 17)

Outstanding geographical contributions from his students came from David Harris (1963) on plant domestication and biogeography, William Denevan (1963) on agricultural systems in the Americas, Thomas Veblen (1975) on biogeography, and Karl Zimmerer (1988) on cultural and political ecology, rural land use and development, and the Andes, and there were many more.

The Berkeley School and Cultural Geography

During the 1980s, a reaction occurred in which the scholarship and striking achievements of the Berkeley School were belittled. A false and unnecessary contrast was made between Berkeley's 'traditional' cultural geography and the 'new' cultural geography of the post 1970s as the new cultural geography defined itself in part by contrast with what went before. The Berkeley School was typified as empiricist, parochial, obsessed with relict artifacts, and having an antiquarian 'object fetchism' in landscapes, and not concerned with the inner workings of culture. It was also represented as ossified and showing little variation over the years, with little concern for indigenous landscape interpretation and contemporary change. Only occasionally were its multiple concerns with environmental problems and human ecological processes acknowledged. Its political overtone was said to be conservative if not reactionary. The new cultural geography, on the other hand, proclaimed that it was all that the old was not, and more concerned with patterns of significance in the landscape and the reflexive role in moulding social relations on the spatial patterns of race, class, and gender.

In a spirited rebuttal of this 'reinvention' of cultural geography Price and Lewis (1993: 1–17) showed how the Berkeley School had been, in their words, "subjected to misrepresentation," "selective amnesia," and "gross generalization." Equally, they showed that the generalizations did scant justice to the variations and innovations in the School. In particular, they demolished the charge of 'object fetichism' by showing how few of Sauer's students actually focused on artifacts. From a list of 1148 articles by Sauer's students in *Sixty Years of Berkeley Geography,* only 49 (or 4.3%) are concerned with artifacts in a narrow sense. If the works of Kniffen and Doran are excluded, the number drops to a mere 2.5%. Other departments in the United States, totally unconnected to Berkeley, adopted the artifactual approach, and even contemporary cultural geographers in any continent did not neglect houses, fields, barns, shopping malls, memorials, and the like.

Another forgotten factor in the criticism of artifacts is that Sauer's concept of a human artifact included plants and animals, and human modifications of landscape, and it emphasized the interrelationships between humankind and the environment. In later years, Sauer's interests moved more from the past to the present, and even to the future, which gave him 'the cold horrors'. In the con cluding sentences of his Bowman lectures on *Agricultural Origins and Dispersals* (1952), he pointed out that humans would for ever remain part of the organic world and therefore needed to develop a respect and understanding for nature.

> …as we intervene more and more decisively to change the balance and nature of life, we have also more need to know, by retrospective study, the responsibilities and hazards of our present prospects as lords of creation. (Sauer, 1952)

Also, Sauer's conservatism must not be confused with simplistic political meanings of that word. He was con vinced (and with good reason) that academic life was becoming overbureaucratized, and that individual curi osity in research was being supplanted by social engin eering and complex research programs dreamt up by social science theorists distant from the places and people studied. The individual was being squashed. Similarly, his distrust of Eurocentric viewpoints led to a prescient appreciation of, for example, the true magnitude and diversity of the New World population numbers. Al though regarded initially with suspicion, the point about 'ecological imperialism' causing massive population de cline in the western hemisphere is now generally ac cepted. Again, his insistence on the primacy and persistence of the role of fire in altering landscapes is now firmly embedded in ecological thinking. Finally, Sauer's rejection of modernity and capitalist development was deeply felt, and during the 1950s he went on record voicing the then wholly revolutionary idea that capacity to produce and capacity to consume were the twin spirals of the new age that had no foreseeable end. He also had a distaste for the elimination of durable native systems of living in the name of progress. These fears align him more clearly with modern day radical environmentalists than with the conservatives.

Since Price and Lewis's article, the pendulum has swung again, and while the new cultural geography cer tainly has, and still, defines itself in part by contrasting itself with what went before, few contemporary prac titioners would now entirely disown the contributions of the Berkeley School.

Parsons's contribution and those of his students has been entirely underestimated when analyzing the Berkeley School. Two thirds of the 36 PhDs supervised by Parsons were based on field studies in nonwestern countries, and transcended any Eurocentric perception. Some have mentioned already but many pioneered a new variation of Berkeley human–environment geography in the form of 'cultural ecology'. This was a profound move, not wholly by the later Berkeley group to be sure (others were Carl Butzer and Harold Brookfield) but one which maintained the Sauerian landscape and historical inter ests but refocused them within a more scientific framework. This approach emphasized micro level studies of human adaptation and their cultural contexts, but also used theory to informed problems, systematic data collection, and attempted to contribute to general concepts and explanations. Prominent among the Berkeley graduates in this innovative approach were Bill Denevan (and his student Billy Lee Turner) and Karl Zimmerer. Additionally, studies founded on political issues, which had been absent from the work of the older Sauerian students, loom large in John Wright's and Paul Starr's works on the political economy of the Rockies and western states.

Assessment

If we accept Billy Lee Turner's recent conceptualization that geography's 'contested identities' have fluctuated around variants of two basic approaches that have dom inated the field during the twentieth century – (1) the spatial–chorological and (2) the human–environment – then the Berkeley School in all its different developments and manifestation is firmly a major component and practitioner of the latter. One can certainly agree with Pierce and Lewis that Pred's judgment that cultural geography should cast aside "the heavy ballast of Sauer influenced landscape study" which must be seen for what it is, a premature and needlessly bitter, and surprisingly naïve, view of what the individual members of the loosely defined Berkeley School were doing. A perusal of many of the theses emanating from the Berkeley department during recent years and deposited in the departmental library show that the gulf is nowhere as great as sup posed. Making allowances for the changes that 75 years would make to any subject, the commonalities are closer than suspected. One most not fall into the all too com mon trap that anything old in geography (say before 1960) is worthless.

Even if some geographers find the Berkeley ethos and emphasis does not answer their questions, we can be sure that environmental historians have grasped its themes with both hands and find it helps to explain the com plexities of a modern world. Sauer's project of elucidating the cultural landscape, suggests Livingstone, was "as much as anything else, an exercise in moral recovery. It was a double directed vision – back to the wisdom of the past and forward to posterity." Sauer's plea at the end of his contribution to *Man's Role in Changing the Face of the Earth* (1956) that "what we need …… is an ethic and aesthetic under which man, practicing the qualities of prudence and moderation, may indeed pass on to posterity a good earth" has (the gendered language aside) a decidedly contemporary and relevant ring to it.

See also: Anthropology and Human Geography; Cultural Geography; Cultural Turn; Sauer, C.

Further Reading

Callahan, R. (ed.) (1981). *Selected Essays 1963 1975: Carl O. Sauer*. Berkeley, CA: Turtle Island Foundation.

Denevan, W. (1989). *Hispanic Lands and Peoples: Selected Writings of James J. Parsons. Dellplain Latin American Studies, No. 23*. Boulder, CO: Westview Press.

Harris, C. (1978). The historical mind and the practice of geography. In Ley, D. & Samuels, M. S. (eds.) *Humanistic Geography: Prospects and Problems*, pp 123 137. Chicago, IL: Maaroufa Press.

Leighly, J. (ed.) (1963). *Land and Life: A Selection from the Writings of Carl Ortwin Sauer*. Berkeley, CA: University of California Press.

Leighly, J. (1979). Drifting into geography in the twenties. *Annals of the Association of American Geographers* 69, 4 9.

Livingston, D. N. (1993). *The Geographical tradition: Episodes in the history of a contested enterprise*. Oxford: Blackwell.

Parsons, J. J. (1972). Spread of African grasses to the American tropics. *Journal of Range Management* 25, 12 17.

Parsons, J. J. (1979). The later Sauer years. *Annals of the Association of American Geographers* 69, 9 15.

Pred, A. (1991). Review of The city as text: The politics of landscape interpretation in the Kandyan kingdom. *Journal of Historical Geography* 17, 115 117.

Price, M. and Lewis, M. (1993). The reinvention of cultural geography. *Annals of the Association of American Geographers* 83, 1 17. See also commentaries and replies in 83, 515 522.

Sauer, C. O. (1925). The morphology of landscape. *Publications in Geography*. Berkeley, CA: University of California 2(9), 271 302.

Sauer, C. O. (1941). Foreword to historical geography. *Annalsof the Association of American Geographers* 31, 1 24.

Sauer, C. O. (1952). *Agricultural Origins and Dispersals. Bowman Memorial Lecture Series 2*. New York: American Geographical Society.

Sauer, C. O. (1956). The agency of man on earth. In Thomas, W. L. (ed.) *Man's Role in Changing the Face of the Earth*, pp 49 69. Chicago, IL: University Press of Chicago.

Turner, B. L. II. (2002). Contested identities: Human environment geography and disciplinary implications in a restructuring academy. *Annals of the Association of American Geographers* 92, 52 74.

Williams, M. (1983). "The apple of my eye": Carl Sauer and historical geography. *Journal of Historical Geography* 9, 1 28.

Berry, B.

R. W. Lake, Rutgers University, New Brunswick, NJ, USA

Berry, Brain J. L. (1934–)

Brian Joe Lobley Berry (**Figure 1**) was born in Stafford, England on 16 February 1934, received his BSc in eco nomics and geography from University College, London in 1955 and his MA (1956) and PhD (1958) in geography from the University of Washington, and is the Lloyd Viel Berkner Regental Professor and Dean of the School of Economic, Political and Policy Sciences at the University of Texas at Dallas. Prior to arriving at the University of Texas in 1986, Berry held appointments as Assistant Professor, Associate Professor, Professor, and Irving B. Harris Professor of Urban Geography at the University of Chicago (1958–76); Williams Professor of City and Re gional Planning at Harvard University (1976–81); and University Professor of Urban Studies and Public Policy and Dean of the School of Urban and Public Affairs at Carnegie Mellon University (1981–86). He is a member since 1975 of the National Academy of Sciences; was the first geographer and one of very few social scientists to serve on the Council of the National Academy of Sciences (1999–2002); is a Fellow of the American Academy of Arts and Sciences, the American Association for the Advancement of Science, and the British Acad emy; is the recipient of many distinguished awards and honors; and was named geography's Vautrin Lud Laureate in 2005.

Figure 1 Photo of Brian Joe Lobley Berry.

Brian Berry is widely acclaimed as one of a small number of innovative and highly influential scholars who inaugurated the application of quantitative methods of spatial analysis in Anglo American human geography in the middle of the twentieth century. The hallmarks of spatial analysis include a human geography based on the adoption of scientific method; the search for empirical regularities and law like statements; the use of hypothesis testing, quantitative analysis, and methods of statistical inference; and identification of universally applicable generalizations leading to a general theory of spatial structure. The analytical human geography advanced by Berry and others (primarily the cohort of graduate stu dents trained by William Garrison at the University of Washington in the late 1950s and early 1960s) brought modern science into geography (and vice versa) in the second half of the twentieth century. The project of spatial analysis explicitly sought to transform the sub stantive, epistemological, and methodological practices of the discipline. In subject matter, the spatial–analytic ap proach rejected the long standing tradition of areal dif ferentiation through regional description involving an encyclopedic accounting of physiographic, material, and social characteristics of places and sought, instead, ab stract characterizations of the spatial organization of economy and society. This substantive shift repudiated the epistemological realism of regional description in favor of a commitment to theory, which Berry and other proponents of spatial analysis took to indicate the possibility of characterizing spatial structure in terms of empirical generalizations and universalistic laws applicable across space and time. The accompanying methodologies of inferential statistics applied to large quantitative data sets with the aid of newly emerging high speed computational technologies replaced tradi tional methods of extended fieldwork, detailed on site observation, and cartographic compilation and display.

These changes were far reaching, transforming the identity of geographers from intrepid fieldworkers to abstract theorists and bestowing on geography the patina of analytical objectivity and the assumed legitimacy of science. Theoretical clarity, analytical rigor, and replic ability are recurrent themes throughout Berry's writing, in which the avowed goal has always been the develop ment of a systematic approach to human geography. Berry has observed that "Theory, not memory, was what counted.... We began to perceive that we were revo lutionaries trying to develop a *science* of geography with a theory of its own, tested and refined using statistical inference, and with the essential spatial patterns captured by mathematical models." The search for empirical regularities and inductive generalizations represented a radical departure from the prevailing understanding of human geography as the study of exceptionalism and regional uniqueness. In Berry's approach to spatial analysis, theoretical abstraction replaced factual compilation, empirical objectivity replaced subjective interpretation, symbolic expression via mathematical notation replaced narrative description, and statistical manipulation of secondary data replaced direct observation in the field.

Berry's earliest work launched a substantive focus on the spatial analysis of urban economic systems while explicitly pursuing his agenda setting project of disciplinary transformation, twin themes that he has continued to pursue throughout his professional career. His first scholarly journal publication, titled 'Alternative explanations of urban rank size relationships', co authored with William Garrison and published in the *Annals of the Association of American Geographers* in 1958, inaugurated the practice of combining substantive argument and disciplinary agenda setting. Stemming from work reported in Berry's 1956 master's thesis, the paper appears as the last article in the issue, as if denoting its tenuous standing in then current human geography, and follows articles in the same issue describing a Japanese settlement on Brazil's Paulista frontier, suitcase farming in Sully County (South Dakota), and activities in the port of Matadi, 80 miles up the estuary of the Congo River. Explicitly rejecting such descriptive exercises in idiographic uniqueness (described by Berry as the 'rote regional geography of the time'), Berry and Garrison challenge the discipline in the first sentence of their article, which reads in full 'Pick any large area'. The authors then observe the regularity of city size distributions across such large areas, assess possible explanations suggested in the work of early systematizers including Zipf, Christaller, and Simon, and conclude by remarking on the absence in the geographic literature of 'a general theory of city size, function, and arrangement'. The publication on urban rank size relationships was quickly followed by extended analyses of the central place hierarchy, the spatial structure of urban business patterns, and regularities in city systems and in the internal structure of cities ('cities as systems within systems of cities'), eventually to include more than 500 books, articles, research papers, and other publications on topics ranging from residential segregation to long wave rhythms in economic development. A 1970 compilation of work initiated or influenced by Berry's approach to urban spatial analysis, titled *Geographic Perspectives on Urban Systems*, became a benchmark in the subdiscipline of urban geography. An additional measure of his influence on the field is the more than 130 doctoral dissertations he supervised in geography and public policy between 1960 and 2000. Complementing his prodigious scholarly production, Berry has placed great importance on linking theory to practice and application; his contributions to policy and planning span his career, from civil defense disaster planning for the State of Washington in the mid 1950s to consulting on national development planning in Indonesia, India, Brazil, Chile, Sri Lanka, and elsewhere.

Berry's project of disciplinary transformation encountered formidable opposition and continues to provoke reaction. His earliest attempts at publication in the principal geography journals were rejected as mathematics and 'not geography', and his 1958 articles were effectively published as an experiment to test the readership's acceptance. A critical response quickly appeared in the form of a 1960 guest editorial in *Economic Geography*, in which the author observed that 'statistics are but half of life', the other half being 'understanding and imaginative interpretation'. Berry replied within the year with a guest editorial entitled 'The Quantitative Bogey Man', repeating his rejection of uniqueness and exceptionalism as unconducive to the search for the generalizations he favored as the object of empirical research. The early criticisms focused on precisely those elements of spatial analysis that Berry promoted as its strengths – the transposition of 'facts' from the object of regional description to a means for construction of theory and the pursuit of universal laws – and on his dogmatic rejection of other approaches.

By the early 1970s, Berry's spatial analytic project had entered the mainstream of human geography, with criticism now emanating from a different source, namely those expressing political values fomented by the war in Vietnam and supporting Marxist approaches in urban economic geography. These critiques targeted the central premises of Berry's positivist epistemology: the existence of an objective external reality, the possibility of value free research, and the conflation of empirical regularity with underlying causality. Berry debated these themes in an extended, heated, and often *ad hominem* exchange with David Harvey in the early 1970s, primarily conducted through reviews of each other's books, in which Harvey challenged Berry's claims to political neutrality and 'the liberal virtue of objectivity in an ideological world'. An enduring effect of this debate has been to equate all forms of quantitative analysis with ideological conservatism, an effect that adherents of methodological pluralism in geography contend with to this day.

Berry's opportunity for rejoinder came in his 1980 Presidential address to the Association of American Geographers (AAG). Under the title 'Creating future geographies', he reprised the transformative agenda he had initiated nearly a quarter century earlier and characteristically sought to articulate a future direction for the discipline. In Berry's view, the ferment that allowed spatial analysis to replace regional description as the discipline's 'unitary mainstream' had since 'degenerated into license that threatens the future of the profession'. His recipe for 'retracking American geography' involved the complementary development of theory and practice in pursuit of a disciplinary agenda reflecting enduring American values of individualism and opportunity. American geography, he argued, should be 'a creative American social construct', that extolled the efficacy of individual entrepreneurship and private markets in solving social problems.

Berry's insistence on rooting American geography in American values revealed an ideological predisposition deeply rooted in his personal experience. Born to a working class family in the English West Midlands, Berry transcended the strictures of a rigid class system to attend university and, by his own account, was determined to outperform others so as to be recognized at least as an equal. From this formative experience came a tenacious belief in both the necessity for and the possibility of individual achievement. The centrality of this belief found repeated expression in Berry's writing over several decades. In his 1972 exchange with David Harvey, Berry challenged Harvey's structural Marxism by locating the 'driving force of the American mainstream (in) the individual pursuit of status through achievement in the material world in a mobile society that places the greatest values upon individual initiative, private decision making, and competitive determination of success'. Berry stressed in his 1980 AAG Presidential Address that American geography should be predicated on particular American values, notably the "utilitarian belief that solutions to needs must be fashioned … out of individual efforts in the private sector, where 'the stimulus of gain induces the hard work that is essential on the part of anyone wanting to succeed'. Ten years on, Berry used a 1993 editorial in *Urban Geography* to observe that 'I was brought up in a philosophical tradition that values individualism and universalism." These enduring values, signaled in the motto (*Nihil Sine Labore*) on the Coat of Arms he was granted in 2005, impelled both his personal trajectory as a dominant figure in twentieth century human geography and the spatial analytic style of human geography that is his enduring legacy.

See also: Central Place Theory; Christaller, W.; Economic Geography, Quantitative; Harvey, D.; Historical Geographies, Urban; Scientific Method.

Further Reading

Barnes, T. (2001). Lives lived and lives told: Biographies of geography's quantitative revolution. *Environment and Planning D: Society and Space* 19, 409 429.

Berry, B. J. L. (1959). Ribbon developments in the urban business pattern. *Annals of the Association of American Geographers* 49, 145 155.

Berry, B. J. L. (1960). The quantitative bogey man. *Economic Geography* 36, 283.

Berry, B. J. L. (1964). Cities as systems within systems of cities. *Papers and Proceedings of the Regional Science Association* 13, 147 163.

Berry, B. J. L. (1967). *Geography of Market Centers and Retail Distribution*. Englewood Cliffs, NJ: Prentice Hall.

Berry, B. J. L. (1972). Revolutionary and counter revolutionary theory in geography: A ghetto commentary. *Antipode* 4, 31 33.

Berry, B. J. L. (1980). Creating future geographies. *Annals of the Association of American Geographers* 70, 449 458.

Berry, B. J. L. (1993). Geography's quantitative revolution: Initial conditions, 1954 1960, a personal memoir. *Urban Geography* 14, 434 441.

Berry, B. J. L. (1993). Should UG become PC? *Urban Geography* 14, 315 316.

Berry, B. J. L. and Garrison, W. L. (1958). Alternative explanations of urban rank size relationships. *Annals of the Association of American Geographers* 48, 83 91.

Berry, B. J. L. and Garrison, W. L. (1958). Functional bases of the central place hierarchy. *Economic Geography* 34, 145 154.

Berry, B. J. L. and Horton, F. E. (1970). *Geographic Perspectives on Urban Systems*. Englewood Cliffs, NJ: Prentice Hall.

Berry, B. J. L. and Wheeler, J. O. (eds.) (2005). *Urban Geography in America, 1950 2000: Paradigms and Personalities*. New York: Routledge.

Johnston, R. (1991). *Geography and Geographers: Anglo American Geography since 1945* (4th edn.). London: Edward Arnold.

Spate, O. H. K. (1960). Lord Kelvin rides again. *Economic Geography* 36, 95.

Biodiversity

R. P. Neumann, Florida International University, Miami, FL, USA

Glossary

Biopiracy The unauthorized, unlicensed, and uncompensated appropriation of biological wealth from the territories of less powerful entities (e.g., states of the Global South, indigenous minorities) by more powerful ones (e.g., corporations of the Global North).

Bioprospecting The systematic collection and assessment of the genetic, chemical, and mechanical qualities of living organisms for potential biotechnological uses.

Biosphere The totality of life on earth and the space in which it exists. One of the four components of the physical earth, including the lithosphere, hydrosphere, and atmosphere.

Biotechnology The use of living organisms or their components (e.g., DNA) in production processes.

Conservation Biology An interdisciplinary field that directs scientific analysis toward the problems of biodiversity maintenance, loss, and restoration.

Genetic Engineering The process by which an organism is genetically modified by the insertion of DNA extracted from another organism in a way that cannot occur in nature.

Global Commons Natural resource systems that function on a global scale and upon which all humanity depends, such as the atmosphere and oceans, but which fall outside capitalist property regimes.

Introduction

Biodiversity, a contraction of biological diversity, is a recently invented term whose origins are readily traceable. The invention of the term 'biodiversity' coincides with the emergence, in the 1980s, of two new scientific fields, genetic engineering and conservation biology. Technological advances in genetic engineering resulted in the commodification of genetic material as a natural resource, the raw material for a new industry. Conservation biology emerged from an effort to structure conservation practices on the scientific foundation of evolutionary biology, with its emphasis on the relationships among genetic variation, genetic exchange, species population sizes, and species extinction rates. The term biodiversity originated from within the latter field in 1986. Walter Rosen, then senior program officer for the Board of Basic Biology at the National Science Research Council, is credited with coining the term in organizing the National Forum on Biodiversity. It is an all encompassing concept that refers to the diversity within the entirety of the world's biosphere, as measured at various scales, including between individuals, populations, species, communities, and ecosystems. Biodiversity thus encompasses genetic, species, and habitat diversity.

The invention of biodiversity was an act of political advocacy as much as one of scientific reasoning. Scientists, principally biologists and natural resource specialists, sought to rally political support for what they perceived as a general and global scale crisis of mass extinctions and habitat loss. Biodiversity thus served as a scientific sounding catchphrase that conservation biologists could employ to educate politicians and the general public. E. O. Wilson cemented the term in both scientific discourse and public consciousness when, in 1988, he produced the edited volume, *Biodiversity*, from the proceedings of the National Forum. Biodiversity has since supplanted several terms once used by biological scientists and conservationists to describe the object of their labors, including game, flora and fauna, wildlife, nature, and wilderness.

Planners, scientists, policymakers, and development experts rapidly incorporated biodiversity into national development plans, research programs, global conservation agendas, and international treaties. Only a year after the publication of Wilson's volume, the Costa Rican government established the National Biodiversity Institute to coordinate the conservation, harvesting, and marketing of biodiversity resources within its territory. In 1992, 150 countries signed the Convention on Biological Diversity (CBD), Article 2 of which defines biodiversity for the purposes of global governance. One of the Global Environmental Fund's (GEF) – which the World Bank created in 1991 as a way to finance environmental initiatives in developing countries – first projects was the preparation of 24 biodiversity country studies. In 1988, biodiversity was not listed as a keyword in the 'biological abstracts'. Currently, there are hundreds of entries annually, while the search of a typical research university's holdings yields over 1000 entries containing biodiversity in the title.

The Biodiversity Crisis

Scientists, conservationists, and hunters have been concerned about species extinction, primarily vertebrate species, at the local and regional scales for decades. Toward the end of the nineteenth century in North

America and Europe, bird plumage on hats was a dominant middle class fashion statement. Exotic and colorful bird feathers were in great demand. As a consequence, many bird species were hunted to near extinction. By the turn of the century, public outrage at the scale of the slaughter prompted local and national governments to crack down on the trade and led to the passage of regulatory laws and to the formation of conservationist organizations, notably the Audubon Society. At the height of European colonialism in the early twentieth century, hunter naturalists based in Europe feared that certain game animals were being pushed to the brink of extinction in Africa and South Asia. Hunters and conservation activists pressurized colonial governments to pass restrictive hunting laws and establish protected areas, such as game reserves and national parks. By the second half of the century, scientists increasingly conceptualized the problem of species extinction as a global one. This sentiment crystallized at the National Forum on Biodiversity, where extinction was framed as a global biodiversity crisis. In framing extinction as a worldwide crisis, conservation biologists constructed biodiversity as a global commons problem.

Trends in Global Biodiversity

The basic unit of analysis for assessing biodiversity trends at any geographic scale is the species. Thus, the number of species measured per unit area is taken as a representative measure of biodiversity. Conversely, species extinction is taken as a representative measure of biodiversity loss. Two important sources for assessing biodiversity at the global scale are the Secretariat of the CBD and the International Union for the Conservation of Nature and Natural Resources (IUCN; World Conservation Union). The Secretariat, among other duties, collects and disseminates information, notably through a series of reports titled, 'Global Biodiversity Outlook', first published in 2001. In publishing these reports, the Secretariat does not generate data from its own research, but rather serves as an information clearinghouse. Data come from individual scientists, governments, and major international nongovernmental organizations (NGOs). The IUCN, in the 1960s, began evaluating and cataloging species that have become extinct since AD 1500 or that were currently at risk of global extinction. It publicizes its findings through its trademarked Red List of Threatened Species, the most recent edition (2004). Similar to the CBD Secretariat, the IUCN derives its assessments of biodiversity loss from data collected by others (including partner organizations and members of the Red List Consortium) much of which is previously published in monographs, edited volumes, and web based databases.

Based on the IUCN and the Secretariat's most recent published assessments and their supporting sources – which, it must be stressed, are continually reevaluated as new data are obtained – it is widely accepted that the global rate of species extinction has increased. The increase is largely due to human activities. Of the estimated total of 14 000 000 species of all living organisms on Earth, approximately 1 900 000 have been described within scientific (Linnaean) taxonomies. Of the described species, 20% of mammals are threatened with extinction, 12% of birds, and 32% of amphibians. The Red List contains 784 documented extinctions since AD 1500 and, based on known extinctions of birds, mammals, and amphibians in the past 100 years, estimates that current extinction rates are 50–500 times higher than extinction rates in the fossil record. Since it is difficult to prove that a species extinction has occurred, scientists created a category of 'possibly extinct' species. If possibly extinct species are included, current extinction rates are 100–1000 times above average or 'background' extinction rates. Background extinction rates are determined by tracing rates of speciation and extinction of life forms in the fossil record. The evolutionary trend has been toward higher levels of biodiversity (i.e., speciation rates higher than extinction rates), and it is estimated that current global biodiversity is near an all time high in planetary history. The fossil record also shows, however, that global biodiversity has fluctuated greatly over time, punctuated by periods of relatively sudden (1–10 million years) mass extinctions. Scientists have identified five mass extinction events where a high proportion of species across a wide range of taxa disappeared from the fossil record, the first occurring 440 million years ago and the most recent, 65 million years ago. Some biologists suggest that the planet may be experiencing a sixth episode of mass extinction, largely induced by human activities.

Though many scientists are convinced that the world is facing a biodiversity crisis, there is a great deal of uncertainty involved in quantifying trends. First, there is a paucity of baseline data, beginning with the absence of a comprehensive species inventory. For example, it is estimated that only 13.6% of all existing species have been described and cataloged in scientific taxonomies, although the actual number of existing species is unknown. For some taxa, the difference between what is known and what (probably) exists is even more extreme. For example, only 0.4% of the estimated total bacteria species have been classified. The lack of basic data means that biodiversity assessments vary wildly by two or three orders of magnitude between what is known and what is estimated as possible or likely. Second, the methodologies used for assessment can vary, so that results are not comparable or only partially compatible. For example, the Secretariat's 'Global Biodiversity Outlook' considers bacteria species and cultivars in its assessment, while the IUCN's Red list does not. More troublingly, there are large epistemic differences between science based inventories and locally based bodies of knowledge that can make them incompatible for the purpose of bio diversity assessment. Third, there is little time series data that would allow scientists to quantify trends in bio diversity over a significant geographic area. Most of the data on species have been collected on an *ad hoc* basis in static, one time inventories. Finally, it is difficult to know the degree to which our assessment of biodiversity re flects actual trends in the Earth's biosphere or is an ar tifact of improved monitoring.

Proximate and Ultimate Causes of Biodiversity Loss

Beyond a general agreement that increased biodiversity loss is associated with human activities – ranging from the human colonization of the Americas and Australia 15 000–25 000 years ago to the industrial revolution – there is a great deal of debate about precise, causal mechanisms. A typical list of causes includes disparities in wealth and poverty, insecure property rights, overpopulation, habitat loss, overconsumption, and, in creasingly, climate change. Differences in assessing causation are driven partly by whether one is considering proximate or ultimate causes. For example, a common proximate causal explanation of the decline in global biodiversity is habitat loss driven by population pressure, especially in the Global South. An ultimate causal ex planation might point to government land tenure and taxation policies that promote the concentration of land ownership and the expansion of large scale capitalist agriculture at the expense of poor family farmers who are forced to clear new land that is agriculturally marginal, but biodiverse. Thus proximate explanations, driven by a neo Malthusian ideology, often single out demographic forces, while ultimate explanations might suggest com plex political economic processes through which class interests are pursued. Sometimes, the latter are referred to as root causes – processes that drive biodiversity loss, but which are spatially and temporally distant from where the loss occurs. An analysis that seeks to look beyond proximate causes, therefore, must examine cul tural, social, and political economic processes operating at multiple temporal and geographic scales.

The Geography of Biodiversity

Geography is critical to understanding biodiversity. First, biodiversity is unevenly distributed over the surface of the Earth. In general, species diversity increases toward the equator, with the world's most species rich habitats occurring in the belt of tropical moist forests that falls between the Tropics of Capricorn and Cancer. Tropical moist forests are the most species rich terrestrial habitats, covering only 7% of the Earth's surface, but estimated to contain 50–90% of all species. Coral reefs, occurring only between 30° north and south of the equator, comprise about 0.25% of the area of the world's marine environ ment, but contain 25% of its total number of species. Furthermore, the geographic distribution of many spe cies, particularly in the tropics, is extremely limited. Such species, often referred to as endemic or restricted range species, are found in a relatively small area and nowhere else on earth.

Second, our knowledge of biodiversity is greatly in fluenced by its geographic distribution in the biosphere. In general, the more inhospitable and inaccessible to humans a habitat is, the less we know about its biodiversity. For example, while oceans and seas by far make up the greatest area of the Earth's surface, they are also the least understood ecosystems and the biodiversity they contain is poorly documented. Indicative of this lack of knowledge is the fact that the most widely used maps of global bio diversity typically include only terrestrial ecosystems. Some marine habitats, such as those surrounding the hydrothermal vents in areas of seafloor spreading, are accessible only with the most advanced technology, yet they support a diverse community of species. Unique life forms, such as bacteria that use hydrogen sulfide as an energy source, have only recently been discovered near seafloor vents. Similar problems of accessibility exist in the terrestrial portion of the biosphere. While tropical moist forests are species rich, the most extensive are relatively inaccessible and difficult to traverse. Hence, scientists and conservation biologists have assessed only a small fraction of the tropical forests where most of the world's bio diversity is estimated to exist. In addition, most of the world's moist tropical forests are found in developing countries, where pervasive poverty, lack of infrastructure, and political and social instability may discourage sys tematic biodiversity assessments.

Third, certain physical geographic features, such as islands, mountaintops, and coral reefs, play key roles in global biodiversity. Islands are important to global bio diversity because many have been genetically isolated for millions of years, allowing for speciation and genetic variation to occur separately from that on continents. Many species are thus endemic to a single island or archipelago. In addition, the genetic isolation of islands creates evolutionary dynamics that produce unusual phenomenon such as dwarfism and flightlessness. Mountaintops are similarly, though not identically, iso lated geographic features. The steep terrain, altitudinal climate zones, and fragmented landscapes of mountain regions create conditions for high rates of species en demism. Limestone ridges formed by living corals in shallow tropical seas create the conditions for concen trations of marine biodiversity. A wide variety of organ isms, including, oysters, clams, crabs, sponges, turtles, and fish find shelter and food in the underwater architecture of the reefs.

Finally, biodiversity loss is strongly influenced by geography. For example, the three geographic features described above together account for a disproportionate amount of recent species extinctions and the world's most threatened species. The great majority of extinctions since AD 1500 have occurred on oceanic islands, where many species, having evolved in isolation, are relatively vulnerable to habitat disturbance and predation. How ever, this pattern has begun to change with the number of recent extinctions on continents equaling those of islands. Most of the species of birds, mammals, and amphibians currently threatened with extinction are located in the tropical latitudes. Tropical islands particularly stand out as having a disproportionate amount of the world's threatened species. Tropical moist forests also stand out in global assessments of threatened species, because they contain high concentrations of biodiversity, but are rap idly being converted to intensive agriculture, such as large scale corporate plantations. There is also a great deal of geographical variation in the extinctions and threats of extinctions among different taxonomic groups. For example, global mollusk extinctions are concentrated in North American rivers, whereas Australia has the largest concentration of mammal extinctions.

Given that both the levels of biodiversity and numbers of species threatened with extinction are often geo graphically concentrated, some conservation biologists base protection strategies on biodiversity hotspots. Nor man Myers coined the term 'hotspots', in 1988, to iden tify areas in the biosphere that have high levels of species diversity, endemism, and habitat loss. This exercise was driven by practical conservationist concerns, the idea being to identify areas where biodiversity protection is most urgent and where the payoff, measured as numbers of species protected, would be greatest. Since the term was introduced, conservationists have identified 34 hot spots around the world in both terrestrial and marine environments. Biodiversity hotspots have become a major focus for conservation programs and funding, with Conservation International building its program agenda around the concept and the World Bank and MacArthur Foundation providing hundreds of millions of dollars of financing. Subsequent research, however, suggests that focusing on hotspots identified under this system may not be an appropriate strategy. For example, in the case of the world's terrestrial mammals, little geographic overlap was found between species richness on the one hand, and endemism and threats of extinction on the other. Furthermore, it is becoming increasingly clear that global climate change is causing shifts in the geographic distribution of species across many taxa, decreasing the effectiveness of the hotspots approach to biodiversity loss.

Key Issues and Critical Perspectives

Biodiversity touches nearly every aspect of human affairs in the twenty first century, including environmental sustainability, poverty and social justice, medicine and healthcare, agriculture and food, and economic global ization. As a consequence, a wide range of critical as sessments of the concept has emerged, ranging from questions of who should own, control, and benefit from biodiversity to questions of how to inventory and catalog the world's life forms. Among the myriad scientific, political, economic, and social questions raised by the invention of biodiversity, three areas of concern stand out: social justice and proprietary rights; uncertainties in the study of biodiversity; and the effectiveness of bio diversity protection efforts.

Social Justice and Proprietary Rights

The construction of the biodiversity crisis as a global commons problem suggests that the solution lies with restructuring property relations from the top down. There hence has been a worldwide movement toward the enclosure and commodification of biodiversity, endorsed by and enforced through such international arrangements as the CBD and the GEF and state governments in concert with various mainstream international conser vation NGOs. Such a tragedy of the commons under standing of biodiversity loss assumes that the world's biodiversity exists 'out there' in nature without propri etary claims as an uncaptured natural resource. Many social justice issues arise from such a perspective. Critics, such as Vandana Shiva, argue that much of the world's biodiversity and knowledge about it are, in fact, the products of the labor generations of agrarian com munities. These critics argue that what passes for legit imate bioprospecting in the Global South is in reality biopiracy; that private interests (e.g., northern multi nationals) take freely the biodiversity produced in the Global South and commodify it as intellectual property to be sold for a profit in the form of improved seeds, medicines, and other biotechnology products. Others point out that the local commons, which rural com munities manage and depend upon for their livelihoods, are rescaled as global commons and enclosed for the benefit of an abstract 'humankind'. The consequent dis placements of communities in the interests of bio diversity conservation raise justice questions regarding who bears the costs of and benefits from enclosures.

Uncertainties in the Study of Biodiversity

Practitioners have labeled conservation biology a 'crisis science', meaning that in a context of great uncertainty, as with certain branches of medicine, the risks of inaction are greater than those of action. The question, what proportion of the world's biodiversity can be lost before the planet's human life support systems collapse, cannot be answered with certainty given the current state of scientific knowledge. Waiting for that question to be answered with scientific certainty is to risk, quite literally, the end of the world as we know it. Generally, it is this sort of unanswered empirical, scientific question that conservation biologists refer to when writing of uncertainty. But there are also more fundamental ontological and ideological sources of uncertainty in understanding biodiversity. These include questions regarding how species are inventoried and classified, how well scientific models reflect reality, and the anthropogenic sources of biodiversity.

At the center of the scientific response to the biodiversity crisis is an effort to inventory and classify as much of the world's biodiversity as possible and then to set priorities for its conservation. Critical observers of this strategy suggest that this is not as straightforward and empirically objective as may appear. First, there is a great deal more attention to charismatic life forms (e.g., large mammals) that appeal to scientists for subjective reasons, thus creating a biodiversity inventory skewed toward a limited number of taxa. Second, the structures of classification systems themselves are subjective constructions that vary geographically and historically. Third, as biodiversity is incorporated into human affairs in the form of commodities, scientific inventories are increasingly shaped by the demands of political economy. The prevailing model of nature in the scientific study of biodiversity is that of a steady state system of long duration, which, when disturbed, returns to stasis through an orderly set of stages (see section titled 'Effectiveness of biodiversity protection efforts'). Thus habitats of high biodiversity, such as moist tropical forests, are viewed as pristine remnant ecosystems threatened by anthropogenic disturbance. Increasing evidence from both the natural and social sciences suggests that many of the world's most diverse habitats, such as the forests of the Southeast Asian archipelago, are not ancient and pristine, but relatively recent systems that have been subject to widespread and sometimes frequent cycles of disturbance, particularly from fire. Evidence for greater resiliency and less fragility in tropical ecosystems has also challenged the prevailing stasis model of nature. Moreover, presumed remnant, pristine habitats have been found to be stages in cycles of human food production systems and that disturbances caused by such systems can have the effect of increasing biodiversity. This latter finding raises questions about the role of humans in shaping the world's biodiversity. Studies from virtually all regions of the globe suggest that humans have played an important role in shaping biodiversity at multiple scales, from habitat diversity to genetic diversity. Finally, some critical analysts, building on the work of Bruno Latour, argue that the scientific project to inventory and catalog species for conservation purposes will ultimately produce the world's future biodiversity in its own image.

Effectiveness of Biodiversity Protection Efforts

In the face of evidence of increasing extinction rates, conservationists and scientists of the 1950s and 1960s laid the conceptual foundation upon which twenty first century global biodiversity protection efforts are built. Major international NGOs and global governance institutions organized to create a worldwide network of protected areas, such as national parks, that would serve as 'nature islands' where wild species would be given refuge from human exploitation. As the fields of conservation biology and biotechnology emerged, protected area establishment became the centerpiece of a global scheme of *in situ* genetic diversity conservation. The CBD and GEF have adopted this framework in directing international efforts to stem biodiversity loss, focusing funding on the expansion and management of a global network of protected areas. Scientific and political interest in protected areas has risen tremendously because they are now viewed as the primary containers of the world's extant and future biodiversity. National governments have implemented the greatest expansion of protected areas in history, creating more parks and reserves between 1970 and 1990 than in all the previous decades. Most of the increase has occurred in the Global South where poverty levels are high and the majority of the populations are rural and directly dependent on the land for their livelihoods.

Critics of the nature island approach to global biodiversity conservation have questioned its effectiveness, observing that even as governments have exponentially increased the area of the Earth's surface under formal protection, extinction rates and the number of threatened species have risen. The reasons are complex. Proponents of the approach argue that many protected areas in the Global South are not really protected, but rather exist only as 'paper parks' with insufficient funds for planning and management. Thus, the solution is to funnel more funds for protected areas to poor tropical countries through such instruments as the GEF. More critical observers suggest that the problem lies in the very conceptualization of biodiversity protection in conservation biology. They note that the historical and philosophical foundations of the prevailing protected area model equate biodiversity with wilderness and presume that the normal ecosystem dynamics tend toward steady state or stable equilibrium conditions. Equating biodiversity with wilderness discounts the positive role of human activities in biodiversity, requires the displacement of human populations from protected areas, and fails to address the fact that the vast majority of the world's biodiversity exists outside of protected areas. The stable equilibrium concept of ecosystem dynamics has been increasingly challenged by theoretical advances and empirical findings in ecological science that suggest that instability, disequilibria, and chaotic fluctuations are far more common and perhaps 'normal' conditions. There fore, attempting to stop ecological change by bounding populations within protected areas is proving to be an ineffective response to biodiversity loss. A more com prehensive response, one that challenges the prevailing political economic model of global capitalism, will likely be required to staunch the loss of biodiversity.

See also: Deforestation; Ecology; National Parks; Privatization; Resource and Environmental Economics; Rio Summit; Sustainable Development.

Further Reading

Baillie, J. E. M., Hilton Taylor, C. and Stuart, S. N. (eds.) (2004). *IUCN Red List of Threatened Species: A Global Species Assessment*. Gland: The IUCN Species Programme.

Bowker, G. (2000). Biodiversity datadiversity. *Social Studies of Science* 30(5), 643 683.

Castree, N. (2003). Bioprospecting: From theory to practice (and back again). *Transactions of the Institute of British Geographers NS* 28(1), 35 55.

Goldman, M. (ed.) (1998). *Privatizing Nature: Political Struggles for the Global Commons*. New Brunswick, NJ: Rutgers University Press.

Robertson, M. (2006). The nature that capital can see: Science, state, and market in the commodification of ecosystem services. *Environment and Planning D: Society and Space* 24, 367 387.

Secretariat of the Convention on Biological Diversity (2001). Global biodiversity outlook. *Report of the Secretariat of the Convention on Biological Diversity*. Montreal, QC: Secretariat of the Convention on Biological Diversity.

Shiva, V. (2000). *Tomorrow's Biodiversity*. New York: Thames and Hudson.

Takacs, D. (1996). *The Idea of Biodiversity: Philosophies of Paradise*. Baltimore, MD: The John Hopkins University Press.

Wilson, E. O. (ed.) (1988). *Biodiversity*. Washington, DC: National Academy Press.

Zerner, C. (ed.) (2000). *People, Plants, and Justice: The Politics of Nature Conservation*. New York: Columbia University Press.

Relevant Websites

http://www.iucnredlist.org
IUCN's Red List of Endangered Species.

http://www.navdanya.org
Research Foundation for Science, Technology and Ecology, Navdanya.

http://www.biodiv.org
Secretariat of the Convention on Biological Diversity.

http://www.conbio.org
Society for Conservation Biology.

http://www.wrm.org.uy
World Rainforest Movement.

Biodiversity Mapping

H. Hazen, Macalester College, St. Paul, MN, USA

Glossary

Biodiversity The variety of life forms found in a particular area.

Ecoregion Geographical zones of regional extent that represent groups of similar ecosystems.

Endemic Native to just one region.

Gap Analysis The use of geographic information systems (GIS) to identify gaps in the protection of a particular region for the purposes of conservation.

Hot Spot Analysis A technique for prioritizing areas for conservation protection, which focuses on those areas considered to be the most biodiverse.

Representativeness Analysis A technique for prioritizing areas for conservation protection, which focuses on including as wide a variety of different ecosystems or habitats as possible in an array of protected areas.

Surrogate or Proxy A substitute; in this context, an easily quantified measure such as number of bird species is used as a substitute for the less quantifiable notion of biodiversity.

Taxonomy The science of classification of living organisms.

Introduction

The term 'biodiversity mapping' encompasses a variety of techniques used to represent and analyze patterns of biodiversity, and is almost always used within the framework of environmental conservation. Although maps of biodiversity can be useful simply as a way to represent complex environmental data in an easily interpretable form, usually biodiversity mapping involves further analysis and manipulation of the data in order to answer specific conservation questions. Out of necessity, some areas must be prioritized over others in terms of their conservation value. Until recently, this was done on a largely *ad hoc* basis with ecosystems receiving protection largely because of dramatic landscapes, recreational value, or the absence of alternative economic uses of the landscape. Over the past two decades, however, there has been a growing effort to protect important landscapes according to scientifically rigorous criteria; biodiversity mapping has been a critical tool in this effort.

Biodiversity mapping involves the cartographic representation of species distributions and/or other environmental characteristics as proxies for biodiversity. Further data, such as information on human use of the land, can also be incorporated in the analysis. These layers of data are typically entered into a geographic information system (GIS) in order to generate maps that can be used to answer the conservation questions at hand. There remain many barriers to producing useful biodiversity maps: the very notion of biodiversity is not easily quantifiable; there is a profound lack of environmental data for certain elements of ecosystems and many parts of the world, and often the quality of available data is poor. Much of the debate surrounding biodiversity mapping focuses on how to approach the problem of imperfect data and how the notion of biodiversity can best be represented. In particular, at what scale, and with what surrogate measures, can we best quantify biodiversity in ways that provide meaningful answers to conservation questions?

History

The mapping of species distributions has a long history in geography. As far back as the early 1800s, Alexander von Humboldt described the latitudinal and altitudinal distributions of vegetation zones, and others continued such practices until methods were eventually formalized within the subdiscipline of biogeography. In other disciplines, too, particularly biology, the notion that place matters in ecological processes and patterns was reinforced, perhaps most famously by Macarthur and Wilson in their 1967, *Theory of Island Biogeography*. More recently, in the face of rapid biodiversity loss, attention has been increasingly focused on the application of these ideas to the challenge of trying to protect biodiversity.

Modern biodiversity mapping began in the 1970s when several researchers began to devise schemes to divide the landscape into functional units of significance for conservation and/or resource development. It was hoped that these new ecological land classification systems would provide policymakers with boundaries that would be more suited to making resource management decisions than are political boundaries. For example, ecologist Raymond Dasmann identified approximately 200 biotic provinces worldwide, while biologist Miklos Udvardy devised a global scheme with a similar number of biogeographical units; at about the same time Robert Bailey of the US Forest Service created maps of the ecoregions of North America. Such schemes have since been further refined, with more focused and nuanced regional analyses as well as efforts to improve global approaches. One of the most extensive efforts in this

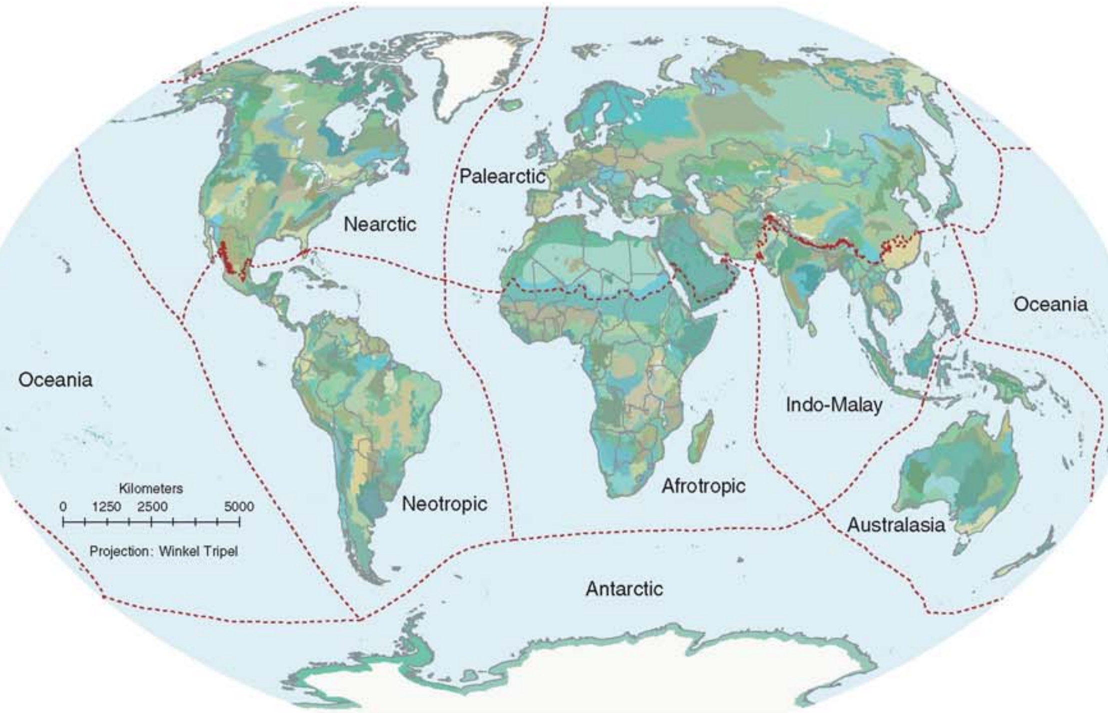

Figure 1 Terrestrial ecoregions of the world: the ecoregions are categorized within 14 biomes and eight biogeographic realms to facilitate representation analyses. Reprinted by permission, copyright American Institute of Biological Sciences. Olson, D., Dinerstein, E. Wikramanayake, E. *et al.* (2001). Terrestrial ecoregions of the world: A new map of life on Earth. *Bioscience* 51, 933–938.

respect has been the World Wildlife Fund's ecoregion project, which has now developed a comprehensive global classification system of ecoregions specifically for the purposes of conservation (**Figure 1**).

More recently, conservation activities have moved away from emphasis on the protection of individual species toward approaches that explicitly recognize the role of ecosystems and regional- or landscape-level processes. This has stimulated an interest in examining spatial patterns of biodiversity, which have become pivotal in developing new conservation strategies. The concept of gap analysis, originally developed by J. Michael Scott, was critical in this context. First used in the 1980s in Hawaii, this technique relies on overlaying maps of the distributions of different species with maps of protected areas in order to identify gaps in protection. The technique has subsequently been practiced elsewhere in the US, and is now promoted by several federal and state agencies. It has also been used in other countries, including Australia, Costa Rica, and Italy. Gap analysis remains one of the most common applications of biodiversity mapping.

Spurred on by technological innovations, particularly in the realms of GIS and remote sensing, the techniques of biodiversity mapping developed rapidly in the 1980s and 1990s, as many of the theoretical underpinnings of the idea were refined. The ability to easily overlay multiple layers of information in a GIS has proved critical to incorporating the large amounts of information required for an effective analysis. Today, biodiversity mapping is being widely implemented, and is supported by a large body of literature that has emerged particularly over the last 10 years. Most major conservation organizations currently use biodiversity mapping to some extent, including Conservation International, The Nature Conservancy, and the World Wildlife Fund, as well as many intergovernmental and governmental organizations, such as the United Nations Environment Program (UNEP), and the United States' Geological Survey and Forest Service.

A dramatic increase has been seen in the area protected for conservation between 1970 and 2000, with an approximately tenfold expansion in areal coverage globally. Further expansion of protected areas is predicted and biodiversity mapping will likely continue to play an important role in assessing conservation priorities. Increasingly, it may also play an important role in monitoring efforts as biodiversity mapping procedures become simpler and cheaper and the time span over which they have been applied grows.

Types of Biodiversity Mapping

Biodiversity mapping has evolved into a variety of forms since its inception, reflecting different approaches to measuring biodiversity. As already noted, biodiversity is notoriously difficult to measure or estimate. A particular challenge is that biodiversity is hierarchical, with meaning at a variety of scales from the molecular to the ecosystem and even global level. Maintaining biodiversity at all scales is a primary goal of conservation planners, and yet how this can be achieved is highly controversial. Should biodiversity be assessed at the cellular, species, or landscape level; and will striving for maximum biodiversity at one scale necessarily correlate with the protection of maximum biodiversity at other scales? Many reports suggest not, and so the scale at which biodiversity mapping takes place becomes a critical decision for the mapmaker.

Related to the challenges posed by scale, it is impossible to address effectively all elements of biodiversity in any single map. For reasons of feasibility and cost, surrogate (or proxy) measures of biodiversity must usually be employed as indicators of a region's overall biodiversity. A variety of different surrogate measures have been suggested and implemented in different schemes, including (1) species richness using an indicator group, (2) richness of higher taxa, and (3) landscape level environmental proxies.

The traditional approach has been to measure the species richness of a particular indicator group and then use this as an estimate of regional biodiversity; for obvious reasons the indicator group selected is usually one that can be readily inventoried, such as birds or amphibians. Although this technique is relatively easy to implement and can sometimes make use of the large volume of data that already exists on species distributions, it has been widely criticized for failing to provide sufficient evidence that species richness in one group necessarily correlates to species richness in other groups. For instance, the fact that a particular area of forest has a large variety of amphibian species does not necessarily mean that that same area is also host to diverse plant or invertebrate communities.

A second approach is to analyze higher taxa, such as genera or families. It is argued that simple number of species is not as relevant to preserving biodiversity as is taxonomic uniqueness. In this context, for instance, the two species of the rare tuatara (a reptile found in New Zealand), which represent the last remaining species of an entire taxonomic family, could be considered a more important conservation target than a large number of mouse species, each of which is taxonomically very similar. Along similar lines, endemic species are often considered especially valuable for protection as they are unique facets of the biodiversity of a particular region.

These first two approaches are often used to provide the information required for a hot spot analysis (**Figure 2**). This approach, devised by environmental scientist Norman Myers in the late 1980s, calls for the protection of hot spots of biodiversity, initially assessed by simple species counts, on the understanding that targeting these areas for protection would save the largest number of species at the lowest cost. As a result, hot spot analysis is sometimes referred to as a form of conservation triage. While the approach has received considerable attention and support, critics have pointed to the fact that it tends to focus attention on tropical areas at the expense of other less biodiverse but equally important regions, and that much of the biodiversity in a particular region may be taxonomically very similar. In response to such criticisms, many more recent hot spot analyses by Myers and others have incorporated notions of taxonomic richness and endemicity in addition to simple number of species in the selection of priority areas. In 1996, Conservation International engaged this more complex notion of hot spots, adding the idea of vulnerability of sites to the selection procedure in an effort to develop a more nuanced reading of the notion of hot spots.

The third approach to biodiversity mapping is to consider landscape level environmental characteristics as proxies for biodiversity. This may be done using particular environmental characteristics such as rainfall and temperature, but is most commonly accomplished using remotely sensed maps of vegetation coverage. Vegetation data are used to strive for the representation of as many different ecosystems as possible in an array of reserves. This representation approach contrasts the hot spot approach. While both approaches share the goal of protecting maximum biodiversity, proponents of the representation approach argue that this can be done most effectively by protecting portions of as wide a variety of different ecosystems as possible, rather than focusing on just those with the most species. In this way, it is hoped that broader goals of taxonomic uniqueness will be captured as well as high species counts. Opponents argue that representativeness approaches are less politically viable than hot spot approaches because they dilute conservation targets over an unfeasibly large number of sites. Furthermore, taken to its ultimate conclusion, a representativeness approach could be used to imply that a large number of small protected areas are more valuable than several large ones, contradicting many ecological studies which suggest that small reserves are particularly prone to species loss and damage from stochastic events. Most practitioners accept this limitation and do not take the technique to such an extreme.

The idea of representativeness is also an important component of gap analysis. In a gap analysis, information about different species' ranges and the current array of protected areas in the region are entered into a GIS. With appropriate processing, maps are then produced that can highlight species and communities that are not adequately protected – where there are 'gaps' in the reserve system. Here, the graphic representation of the map is critical to highlighting meaningful patterns from large

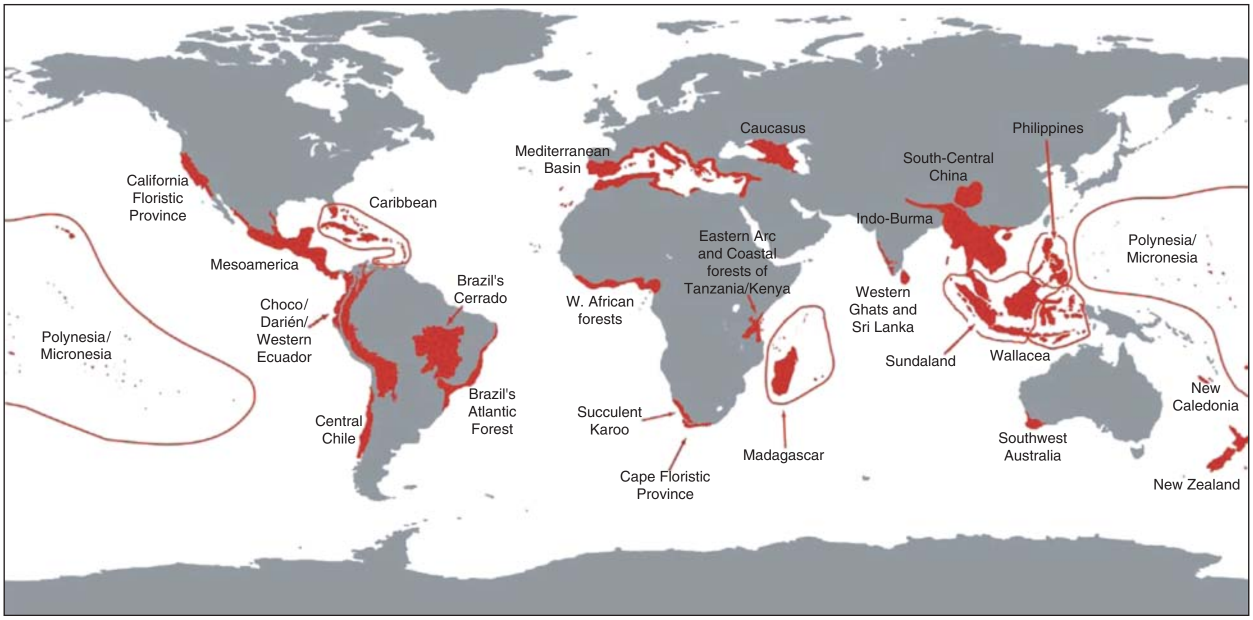

Figure 2 Regions identified as conservation priorities by Myers *et al.*'s hot spots approach. Reprinted by permission from Macmillan Publishers Ltd: *Nature* (Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. and Kent, J. (2000). Biodiversity hot spots for conservation planning. *Nature* 403, 853 858), copyright (2000).

quantities of spatial data. Critics of gap analysis point to the difficulty of producing meaningful results from multiple sources of data of questionable quality. Pro ponents argue that it is reasonable to use available data with appropriate precautions (such as ground truthing), even if the resulting maps are not perfect.

Although these techniques use diverse theoretical approaches, considerable effort has recently been made to try to combine some of these approaches to achieve the best of all worlds. The World Wildlife Fund's Global 200 program is one example of such a combination ap proach. The Global 200 identifies 238 target regions for conservation in order to try to achieve coverage of as many different ecosystems as possible (a representative ness approach). In addition, the areas selected are judged according to rareness of habitat type worldwide, total number of species present, number of endemic species, presence of higher taxa, and unusual ecological or evo lutionary phenomena, thereby also incorporating aspects of a hot spot approach.

Biodiversity mapping elicits considerable controversy over the best techniques for creating meaningful maps. While detailed information on species distributions and ecosystem patterns would be ideal for making conser vation decisions, this information is rarely available. To a large degree, biodiversity mapping projects must there fore not only rely on proxies for estimating biodiversity, but must also be mindful of the potential pitfalls of these approaches and strive to minimize error. Many techni ques have been suggested for addressing these limitations. Some efforts focus on ensuring that data are as accurate as possible, as with ground truthing remotely sensed data. Others attempt to refine the decision making pro cesses through the use of algorithms. Some of these key challenges within biodiversity mapping are discussed below.

Key Challenges within Biodiversity Mapping

Despite important advances in biodiversity mapping, there remains a variety of critical debates surrounding biodiversity mapping approaches. A key issue is the challenge of collecting and combining appropriate data with which to make conservation decisions. A lack of appropriate data has led to questions over the validity of results, particularly in the contexts of developing coun tries and the biodiverse tropical regions where data are especially sparse or poor. Even where data are relatively abundant, decisions must still be made over which measures to select to estimate regional biodiversity. There is a great need for examination of these predictive and estimative techniques to ensure that they are effective.

Identifying valid ecological regions is a second area of controversy. For instance, aerial photographs are often used to characterize vegetation types and yet identification of different plant communities via remote sensing is never completely accurate. Additionally, fuzzy boundaries make identifying the spatial extent of com munities highly contentious. Even if vegetation can be successfully identified, there remains considerable doubt over the degree to which this is an effective proxy for that region's animal communities, again challenging the validity of the ecoregion that has been identified. For instance, large predators commonly cross different vegetation communities, while other species may have far more specialized ranges or needs than can be delineated by a vegetation proxy.

Another critical point of controversy is the scale at which biodiversity mapping should be undertaken. A central scientific question is whether species level, habitat level, or landscape level conservation is most effective. With the advent of biodiversity mapping, many conservation programs have looked increasingly toward regional and global scale approaches, which have become feasible with recent technology and data availability. However, most conservation decisions are made at national or local scales, and so it is questionable how effective these data are for facilitating decision making. Additionally, biodiversity mapping has been challenged for potentially missing considerable detail because of its coarse scale of analysis. In this sense, it is often argued that biodiversity mapping should not replace traditional approaches to species or landscape conservation, but should be employed as a complementary technique.

Finally, biodiversity mapping has also become embroiled in political debate. The quantitative, scientific nature of biodiversity mapping is often criticized for neglecting to consider broader political issues, particularly the feasibility or likelihood of success of an area's protection. In other words, biodiversity mapping can certainly point us toward worthy targets for conservation, but further work must go into assessing the likelihood of persistence of protection of that area, considering political and economic factors. Political considerations are also brought to the fore in the context of intellectual property. Biodiversity now has recognized commercial value and as a consequence many countries have become wary of sharing information on biological resources. Much biodiversity is concentrated in the Global South (and in many cases also on indigenous lands), while information on it is concentrated in the affluent North. Biodiversity mapping has therefore been criticized for encouraging flows of intellectual property to affluent industrialized countries. On the other hand, some have suggested that mapping technologies may hold the potential to democratize conservation activities by giving local communities the tools to manage their own resources more effectively. Whichever the case, it is clear that biodiversity mapping is not only a scientific, but also a political, endeavor.

See also: Biodiversity; Ecology; National Parks; Remote Sensing; Wilderness.

Further Reading

Bailey, R. G. (1983). Delineation of ecosystem regions. *Environmental Management* 7(4), 365 373.

Bowker, G. C. (2000). Mapping biodiversity. *International Journal of Geographic Information Science* 14(8), 739 754.

Conservation International. (2005). Biodiversity Hotspots. http://www.biodiversityhotspots.org/xp/Hotspots/pages/map.aspx. (Accessed in May 2008.)

Conservation International. (2007). Biodiversity Hotspots. http://www.biodiversityhotspots.org. (Accessed in May 2008.)

Convis, C. Jr. (ed.) (2001). *Conservation Geography: Case Studies in GIS, Computer Mapping and Activism*. Redlands, CA: ESRI Press.

Dasmann, R. F. (1973). *A system for defining and classifying natural regions for purposes of conservation*. Morges (Switzerland): International Union for Conservation of Nature and Natural Resources. IUCN Occasional Paper no. 7.

Johnson, N. C. (1995). *Biodiversity in the balance: Approaches to setting geographic conservation priorities*. Biodiversity Support Program: WWF, The Nature Conservancy, and World Resources Institute.

Luoto, M. and Heikkinen, R. (2003). Recent developments in spatial methods and data in biogeographical modeling advantages and pitfalls. *Fennia* 181(1), 35 48.

Macarthur, R. H. and Wilson, E. O. (1967). *The Theory of Island Biogeography*. Princeton, NJ: Princeton University Press.

Miller, R. I. (1994). *Mapping the Diversity of Nature*. London: Chapman and Hall.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. and Kent, J. (2000). Biodiversity hot spots for conservation planning. *Nature* 403, 853 858.

National Geographic. (2001). Global 200: Priority Areas for Conservation. http://www.worldwildlife.org/science/ecoregions.cfm. (Accessed in May 2008.)

Natural History Museum (London). (2007). Biodiversity and WorldMap. http://www.nhm.ac.uk/research curation/projects/worldmap/. (Accessed in May 2008.)

Olson, D. M. and Dinerstein, E. (1998). The Global 200: A representation approach to conserving the earth's most biologically valuable ecoregions. *Conservation Biology* 12, 502 512.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D. *et al*. (2001). Terrestrial ecoregions of the world: A new map of life on Earth. *Bioscience* 51, 933 938.

Savitsky, B. (1998). Overview of gap analysis. In Savitsky, B. & Lacher, T. (eds.) *GIS Methodologies*, pp 151 157. New York: Columbia University Press.

Schmidt, K. (1996). Rare habitats vie for protection. *Science* 274, 916 918.

Udvardy, M. D. F. (1975). *A classification of the biogeographical provinces of the world*. IUCN Occasional Paper no. 18. Morges, Switzerland: IUCN.

United States Department of Agriculture Forest Service. (2006). Natural Resource Assessment and Analysis: Ecoregions Center http://www.fs.fed.us/rm/analytics/publications/ecoregionsindex.html. (Accessed in May 2008.)

World Wildlife Fund; National Geographic. (2001). Global 200: Priority Areas for Conservation. http://www.worldwildlife.org/science/ecoregions.cfm. (Accessed in May 2008.)

World Wildlife Fund; National Geographic. (2001). Terrestrial Ecoregions of the World. http://www.nationalgeographic.com/wildworld/terrestrial.html. (Accessed in May 2008.)

World Wildlife Fund. (2007). Conservation Science: Ecoregions. http://www.worldwildlife.org/science/pubs/g200.pdf. World Wildlife Fund. (Accessed in May 2008.)

Relevant Websites

http://www.biodiversityhotspots.org
Biodiversity Hot spots.

http://www.nhm.ac.uk
Natural History Museum (London).
http://www.fs.fed.us
United States Department of Agriculture Forest Service.
http://gapanalysis.nbii.gov
United States Geological Survey. The Gap Analysis Program.
http://www.worldwildlife.org
World Wildlife Fund.
http://www.nationalgeographic.com
World Wildlife Fund; National Geographic.
http://www.worldwildlife.org
World Wildlife Fund; National Geographic. Global 200: Priority Areas for Conservation.

Biopolitics

A. A. Lehtinen, University of Joensuu, Joensuu, Finland

Glossary

Biopolitical Geography The branch of human geography that links the particular concerns of land and life with the matters of population policy and demographic governance.

Biopower The power of and over life, born with the inclusion of people and their intimate sociospatial relations in the calculations of dominating order.

Displacement The act of unsettling the taken-for-granted or generally accepted. Sudden change in sociospatial conditions.

Horizontal Fix Inaccuracy caused by an exclusive concentration either on the apparent order (e.g., abstract topology, postlocalism) or on the textual dependencies of its representations (e.g., textualization without explicit spacing).

Multitude The unruly state of humans, the masses that resist or show indifference to obedience and established order. The always-present critical potential for challenging the dominating sociospatial unity.

Postlocalism An approach that considers local and place-bound views outdated and misleading due to their nonreflexivity to the current globalization grounded on continuous thorough displacement.

Rhizome A spatial metaphor that refers to the multitude that is made of occasional combinations of thoroughly interconnected semiotic and social relations.

Sovereign The elected board, person, or authority whose power rests – according to Thomas Hobbes – on a social contract covering the duties of both the sovereign (to work for the community) and the community (to show obedience to the sovereign).

Biopolitics and Place

Biopolitical interpretation of society is connected to the debate on the changes of life and work conditions in a world where precarious social relations prevail. It underlines the radical mixing of private and public spheres of life, emerging for example in the form of constant reorganization of families and communities under the pressure of neoliberal globalization. Human beings, both as embodied individuals and population in general, have in this view become units of immediate social transformation. It is therefore argued that no spatial categories are needed in describing the embodied social change, co linking us all, now and everywhere, at the level of population.

The general social pressure becomes concrete in the biopolitical view through endless flows of contests and comparisons between unaccompanied actors, which results in increasing social insecurity. Individuals are under permanent threat of being replaced, both as laborers and family members. Social life grows now into full life at the level of audience and masses, always under the pressure of all encompassing and thorough changes. More and more attractive stimulus and excitement, offered and renewed everywhere, simultaneously, and continuously, compress humans into one universal population, or a multitude, characterized by chaotic heterogeneity.

Social change emerges in the biopolitical view without any stable spatial coordinates or location in a particular cultural arena. Multitude is universal and therefore it cannot be found under the banner of any nation or people and it does not exist in any specific neighborhood or factory. It neither gathers in the streets, nor in the market places for campaigns or demonstrations. Accordingly, the governing of masses does not succeed with the help of local or territorial measures but it is instead developed via the flows of mass communication and media, and is intimately linked to individual human minds and bodies.

Local positions are regarded as biased in biopolitical thinking as they rest on a false dichotomy between the global and the local. Local views are also considered misleading in their belief in an authentic inside that is valued as uncontaminated by the more general changes in society. Accordingly, local perspectives and strategies are criticized as they tend to mask the real risks of social change. Localism is accused of obscuring the alternatives and potentials for liberation from the dominating forms of globalization. Instead, participating in the local–global flows in all their complexity is therefore preferred.

The multitude, referring to the masses, population, and audience, is among the key words of biopolitical thinking, and it has also become a central concept in today's globalization critique. This is no wonder as it clearly helps to explain and cope with the ongoing local–global changes in work, production, consumption, and politics.

The concept of multitude also links today's globalization critique to the centuries long politico philosophical debate on the masses and their political organization stimulated by Thomas Hobbes in *Leviathan*, published in 1561. For Hobbes, multitude refers to the original state of human beings that are in struggle with each other. It is characterized by a plurality of voices and it precedes organized society. Hobbes' multitude does not, however, remain a presocietal characteristic of the human past but is instead understood as an integral element of social renewal. It is seen as both a threat to the dominating order and a potential source for change. Hobbes' political community, the assembly of individuals, is established in the meetings of the multitude where matters of representation and participation are decided. The plurality of voices is united by giving up individual rights to the sovereign, the col lectively chosen board or person who promises to work for the community. Multitude grows in Leviathan into the founding force of community, the primary cause for pol itical renewal.

This multitude can be interpreted as a preorganiza tional element in human history and even the opposite of ordered society, but it can also be seen as the ever pre sent critical condition of social change. Consequently, the concept of primary multitude refers to the unruly state of humanity before the common rules of representation and participation are established, whereas the concept of political multitude links the power of the multitude with critical alternatives and civic pressure against the estab lished order. The two interpretations diverge in their definitions of space and place.

The multitude can take the shape of a universal rhi zome but it can also function as the critical condition for sociospatially diversified community building. The power of the multitude can thus emerge both through occasional or chaotic combinations of individual reactions and in the form of civic gatherings on the basis of specific sociospatial concerns. The latter alternative is, for example, explicated by Gilles Deleuze and Michel Foucault, two of the most refereed examiners of biopower and the multitude, who jointly argue in *Les intellectuelles et pouvoir* (*The Intellectuals and Power*), initially published in 1972, that it is not wise to fight against power in general but, instead, do it on the basis of local confrontations and personal resistance. For them, various types of local–regional theories can help us analyze how the relations of power are articulated.

Horizontal Fix

The idea of nonlocal biopolitics seems reasonable as a provocative means to influence the attitudes toward globalization and especially its neoliberal features. It functions as a warning for the direction that the humanity as a whole has taken. However, nonlocalism also seems to rise from the type of thinking that connects spatiality with traditional or early modern territorial practices of power. The spatial is then attached to the organizational dy namics of the past, expressing an attitude that clearly ignores the radical potentiality and process type openness inherent in sociospatial approaches. The strict division between the past and the present, which is an attractive option in post prefixed thinking, easily guides us toward this type of historical selection.

This setting helps us to understand the doubts some geographers have cast on the so called horizontal fix in biopolitical thinking. Horizontal fix refers in this context to a kind of sociospatial inaccuracy that comes from an exclusive focusing on the ruptures and incoherence in the debate on social change. This emphasis carries the risk of overlooking the multiple human co associations in space and time. The problem emerges, for example, when temporal and spatial are separated by defining the latter as a characteristic of the past. The ignoring of the constant becoming of space is then unavoidable. This results in a lack of sensitivity to emerging new initiatives from different corners of society. The af firmative option of co construction is lost under the pressure of endless deconstruction, fueled by an emphasis on fragmentation and dislocation. The language of dis ruption governs the debate and at worst results in textual overplaying. Instead of co constructing arguments for societal renewal and resistance, one remains im prisoned by the idea of decomposing the new initiatives presented.

If we thus accept that the power of the multitude can constantly emerge in and through multiple sociospatial formations, we can recognize the critical and creative potentials of active relocation under and against the general pressure of displacement. Place can then take a shape as a location of personal or collective denial or it can turn into a moment of selective participation. It can also grow into a core of symbolic fights or a specific motive for political meetings. It might for example ap pear as a context of globalization critique.

The value of place is now connected to the under standing of the spatial as an inseparable part of both general and particular social changes. Places are seen as collections of sociospatial trajectories and spatiotemporal events. The contested co formation of private and public spaces, combined by a proactive stance in building per sonal and collective coping strategies, is now the found ing forum and event of biopolitical critique. This is the place of the political multitude. The politics of place become associated with the conditions and settings where people and their stories meet and it also covers the processes these meetings co produce. Places turn into sites where the lines between the private and the public are repeatedly met and put under question. The politics of place emerge wherever the concern about the con ditions of participation, representation, exclusion, and indifference is presented. It is the place of the everyday – linked to the rest of the world.

We locate ourselves whenever we make choices, both small and big, in the practices of our daily lives. Placing is then reactive or proactive and relocation takes place wherever we identify alternatives under the pressure of the dominating order – in the junctions of the occasional and the dependent. The power of the multitude surfaces both in chaotic combinations and also through individual and collective choices or initiatives.

Spatially sensitive biopolitical research thus pays at tention to the activity, and also to the passivity, of indi viduals and their communities amidst social changes. The changes in the everyday practices and routines are not regarded as a direct outcome of social reorganization introduced as general necessities, but as sociospatial dy namics modified by choices in the intimate sphere of the communities themselves.

Geography and Biopower

The sociospatial view of biopolitics is based on the under standing of power as a web of dependencies emerging in and through the constantly evolving settings of human communities and associations. Power grows from within individual and collective initiatives, reactions, and proac tions under the fluid conditions of general change. Power has in this view thoroughly permeated our social practices and it emerges both in macro and microspheres of our sociospatial being. It is an immanent and unavoidable element of social coexistence, taking shape via continuous processes of re negotiations. Power becomes activated both through submission and reaction. This type of endless contest forms multilayered regimes with networks of reciprocal ties and an endless series of inclusions and exclusions. Biopower becomes manifested in those situations where both our bodily existence and our in timate sociospatial relations become increasingly inte grated into the general technologies of population governance, more or less reproduced in our daily decisions by our willing selves or through our respective reactions and pro actions.

This type of biogeographical approach links our everyday practices with broader matters of population policy and demographic development. In political geog raphy, studies of changing security regulations have be come popular in the 2000s and critical analyses of biometric border control innovations have for example much enriched the geographical debate. The prefix bio guides us to concerns of life and conditions of land and life. General social change is now indivisible from changes in the living conditions of individuals and their communities or networks. These ground level actors can obstruct, promote, and initiate change in society. Socio spatial variation is accordingly seen as constitutive to biopolitical renewal. It is inseparable from human change and governance.

The geographical emphasis in biopolitical thinking helps us notify the multilayered needs and motives of in dividual actors anchored in the multilayered dynamics of sociospatial projections. This line of thinking has been developed in the analyses of historical conditions of social space and power but it has also been examined in the studies of the more unarticulated spheres of human ex istence. In these studies geographers have developed their personal and professional views in close connection to the multiple practices of land and life, ranging from global biogeopolitics to individual geobiographies of the everyday.

Geographers with biopolitical concerns are, due to the spatial lessons from the disciplinary past, broadly aware of the risks of horizontal fixations. Abstract spatialism of the 1960s was followed by radical nonspatialism in the 1970s, and it is therefore feared that the current en thusiasm in postlocalism distances us from individual and shared concerns of life under the concrete conditions of social change. This tendency would undoubtedly con strain our ability to renew ourselves as researchers in dialog with the subjects of our study focus. We might then lose the opportunities to build a reciprocal rela tionship between the subject partners of research and perhaps even ignore the changing social conditions of the lay reality of everyday.

The emphasis of biopolitical practices has been a welcome step forward in developing human geography's sensitivity to endlessly fragmented matters of sociospatial change. The biopolitical renewal remains, however, par tial and elitist in geography if not co developed in close contact with the drama and the dynamics of human communities and networks in their concrete sociospatial settings. We could call this an invitation to move from the cartography of biopolitics toward more participatory studies of particular biogeographies or, alternatively, a leap from intertextual choreogeographies to critical and co constructive research profiles.

See also: Borderlands; Census Geography; Citizenship; Demography; Empire; Environmental Security; Governance; Governmentality; Hegemony; Political Boundaries; Postconflict Geographies; Territory and Territoriality.

Further Reading

Birkeland, I. (2005). *Making Place, Making Self*. Aldershot: Ashgate.
Bouchard, D. F. (ed.) (1977). *Language, Counter Memory, Practice: Selected Essays and Interviews by Michel Foucault*. New York: Cornell University Press.
Deleuze, G. and Guattari, F. (1988). *A Thousand Plateaus*. London: Athlone Press.
Duncan, N. (ed.) (1996). *BodySpace*. London: Routledge.
Hanninen, S. and Vahamaki, J. (eds.) (2000). *Displacement of Politics*. Jyvaskyla: SoPhi.
Hardt, M. and Negri, A. (2000). *Empire*. Cambridge, MA: Harvard University Press.
Hardt, M. and Negri, A. (2004). *Multitude*. New York: Penguin Press.
Hobbes, T. (1991). *Leviathan* (first pub. 1561). Cambridge: Cambridge University Press.
Kymalainen, P. (2005). *Geographies in Writing. Re Imagining Place*. Oulu: Nordia Geographical Publications.

Lehtinen, A. (2006). *Postcolonialism, Multitude, and the Politics of Nature*. Lanham: University Press of America.
Macnaghten, P. and Urry, J. (eds.) (2000). *Special Issue: Bodies of Nature*. *Body and Society* 6(3 4), 1 196.
Massey, D. (2005). *For Space*. London: Sage.
Penttinen, E. (2004). *Corporeal Globalization*. Tampere: Tapri.
Pile, S. (1996). *The Body and the City*. London: Routledge.
Virno, P. (2004). *A Grammar of Multitude*. Cambridge, MA: Semiotext(e) Foreign Agents Series.

Relevant Websites

http://www.keele.ac.uk
Biopolitics of Security Network.
http://www.framework.fi
Framework. The Finnish Art Review.
http://multitudes.samizdat.net
Multitudes. Revue Politique, Artistique, Philosophique.
http://www.vidyaasram.org
Vidya Ashram.

Bobek, H.

H. Fassmann, Universität Wien, Vienna, Austria

Bobek, Hans (1903–1990)

Born in Klagenfurt, Bobek's (**Figure 1**) studies at the University of Innsbruck (1921–26 – geography, history, social sciences) established two key areas of research. On the one hand his dissertation about Innsbruck (*Innsbruck, eine Gebirgsstadt, ihr Lebensraum und ihre Erscheinung*) es tablished his later interest in urban research as well as in central place theory. The other field of research derived from his occupation as Johann Sölch's assistant, at that time one of the best geomorphologists. The pre occupation with issues of the physical geography, glacial periods, and formation of the Alps led to an enduring interest in a holistic analysis of landscape, research fo cusing on the quaternary as well as an understanding of the discipline as an integrative 'bridge' discipline using socio scientific as well as natural scientific approaches.

After his promotion and the publication of his dis sertation, Bobek accepted an assistantship offered by Norbert Krebs in Berlin in 1931. Norbert Krebs was the successor of Albrecht Penck and a leading professor at the time. This provided the chance for Bobek to work at a leading geographical institute at the time and to be ac cepted as a member of a circle of young scientists who essentially determined the development of the geographical profession after the war. Furthermore, the location in the capital Berlin offered the possibility of international research. Hans Bobek selected the Orient, in particular Iran, which he had already visited several times in the 1930s and which he got to know more closely during a visit as a guest professor. This interest was to remain during the remainder of his career.

Figure 1 Hans Bobek.

In Berlin Bobek proceeded with his geomorphologic studies and 'habilitated' in 1935 with a work about the *Formentwicklung der Zillertaler und Tuxer Alpen*. He re ceived his *venia legendi* (entitlement to teach) as late as 1937 after he was allowed to repeat the Dozentenkurs he had failed 2 years previously. At this course, the young Dozenten had to present themselves observed by older colleagues: here, the specialist in German studies, Otto Höfler, described Bobek in a secret dossier, as the typical intellectual with the need to excel at every discussion, who disavowed national socialism, and who was not seen as popular by his colleagues due to his complacency and preciosity.

In 1939, Bobek finally obtained a position as lecturer and was enlisted in 1940, completing his military duty as member of the war administration board. In 1944 he became extraordinary professor and member of the re search echelon of the SS, where he participated with Schmitthüsen, Ellenberg, Walter, and other experts in Russia, Italy, Yugoslavia, and Czechia. During the war, Bobek was, together with Troll, a key promoter of geo graphic research into aerial photography.

His first call in the post war period led Hans Bobek to Freiburg (1946–48). There he wrote, in his forties, the pioneering articles 'The position and relevance of social geography', 'Social space formation exemplified by the Middle East', and 'The landscape in the logical system of geography'. After a short intermezzo at the Vienna University of Economics (1949–51), he was appointed as a professor for geography and as successor of Hugo Hassinger by the University of Vienna where he taught until 1971.

Bobek always sought to connect academic geography to planning and politics, hence linking theory and prac tice, with his time in Berlin important in this respect. In 1955, when the allies moved out of Austria and when full sovereignty was reestablished, Hans Bobek, together with Erik Arnberger, started publishing a national atlas of the Austrian Republic. Despite limited means, he managed to reel in the intellectual elite to contribute to the atlas of the Austrian Republic. Hans Bobek accomplished with this atlas and the related publications a singular docu ment about the Second Republic. In the mid 1960s he published, together with Elisabeth Lichtenberger, an al most exemplary analysis about the physical, social, and urban morphology of the city of Vienna (with Lichten berger) which provided an important foundation for urban morphological studies. In the end of the 1970s he published, together with Maria Fesl, a theoretical and empirical study about the Central Places of Austria, which is still current due to its fundamental methodical and theoretical approach.

Furthermore Bobek was always concerned about the conceptual underpinnings of the discipline and partici pated in the theoretical discourse about the subject area. As early as 1949, working together with Schmithüsen, he had developed the 'logical system of geography', which emphasized the integrative approach which geographers needed to adopt in the interpretation of a landscape. For Bobek, a geographical analysis of the landscape requires consideration of geological conditions as well as flora, fauna, and climate. Evidently, society (an aggregation of various social groups with uniform ways of life) is a factor shaping the landscape, equally or perhaps even more than the physical–geographical conditions. Bobek understood geography as one of the few integrative scientific dis ciplines, combining natural, as well as social and hu manistic scientific approaches.

Nevertheless his particular achievement consisted of the advancement of a modern social geography. Bobek effectively centered the human being as the key agent who creates and modifies the cultivated landscape. As such, he overcame the dominant perception of the nineteenth and early twentieth century, whereby the natural, elemental basis of the environment is seen to shape societal struc tures and individual action. Bobek deferred, arguing space does not determine human and society, but conversely, human and society produce space. To Bobek, the idea of a space with natural borders and a predetermined folkish settlement, which his predecessor Hugo Hassinger had proposed, was pure myth. Bobek's major contribution was hence to suggest that social forces (economy, population, culture, and politics) structure space, rather than being determined by space.

See also: German-Language Geography; Landscape; Social Geography; Urban Morphology.

Further Reading

Bobek, H. (1948). Stellung und bedeutung der sozialgeographie. *Erdkunde II* S, 118 125.

Bobek, H. (1967). Die theorie der zentralen orte im industriezeitalter. *Tagungsberichte und Wissenschaftliche Abhandlung des Deutschen Geographentages Bad Godesberg* 199 213.

Bobek, H. (1972). Die entwicklung der geographie kontinuitat und umbruch. *Mitteilungen der Österreichischen Geographischen Gesellschaft* 114, 3 18.

Bobek, H. (1974). Zum konzept des rentenkapitalismus. *Tijdschrift voor Economische en Sociale Geografie XV* 2, 74 78.

Bobek, H. (1975). Osterreichs regionalstruktur im spiegel des atlas der republik osterreich. *Mitteilungen der Österreichischen Geographischen Gesellschaft* 117, 116 164.

Bobek, H. and Fesl, M. (1978). Das system der zentralen orte osterreichs eine empirische untersuchung. *Schriften d. Komm. f. Raumforschung* 3, Graz Koln.

Bohm, H. (2001). Bobek, Hans. *Lexikon der Geographie in vier Bänden*, Bd 1, 183 184. Berlin: Spektrum.

Fassmann, H. (1955). Geographie an der universitat Wien. Grandner, Margarete, Gernot Heiss, Oliver Rathkolb (Hrsg.): Zukunft mit Altlasten. Die Universitat Wien 1945 1955. *Querschnitte 19. StudienVerlag, Wien* 273 289.

Lichtenberger, E. and Bobek, H. (1966). Wien. Bauliche gestalt und entwicklung seit der mitte des 19. Jahrhunderts (= Schriften der Kommission fur Raumforschung der Osterreichischen Akademie der Wissenschaften, Band 1). *Wien* 2.

Lichtenberger, E. (2001). Geographie. Acham, Karl (Hrsg.): Die Geschichte der osterreichischen Humanwissenschaften ein zentraleuropaisches Vermachtnis. Bd. 2: Lebensraum und Organsimus des Menschen. *Wien: Passage Verlag* 71 148.

Body, The

L. Johnston, University of Waikato, Hamilton Aotearoa, New Zealand

Glossary

Corporeality A recognition of the importance of the physical body (that cannot be removed from discourse) as to understanding lived experiences.

Dualisms Two factors (e.g., mind/body, work/home, and public/private) are assumed to be distinct and mutually exclusive.

Essentialism The belief that socially constructed differences (e.g., gender, race, and age) are determined by biology; therefore, bodies are deemed to have fixed properties or essences.

Identity Relates to social categories such as class and gender but does not assume that these are fixed essences. Identities are considered to be multiple, fluid, unstable, and constituted by space/place.

Performative The notion that identities if both bodies and places are produced through a repetition of particular acts with a regulatory framework.

Social Constructionism Often considered opposite 'essentialism', social constructionism is an approach that examines ideas which may appear to be natural and the norm to those who accept them, but in reality is an invention of a particular culture or society.

Subjectivity Relates to ideological concepts of the subject. Understanding subjectivity involves the negotiation of a whole series of interconnected terms such as the body, the self, identity and the person.

The Other Refers to a person that is different or opposite to the Self. The Other marginalized from the Self and/or excluded from particular places.

Introduction

The body is a crucial site of sociospatial relations, representation, and identities. It is the place, location, or site of the individual. It is also a site of pain, pleasure, and other emotions around which social definitions of wellness, illness, happiness, and health are constructed. The body is the location of social identities and differences such as gender, sexuality, race, ethnicity, age, (dis)ability, size, shape, appearance, and so on. These identities or subjectivities may form the basis for oppression and exclusion, meaning that the body as a space is bound up in knowledge and power. The body is, therefore, a site of struggle and contestation. Where a body can go, what it can do, who can get access to it, is regulated and controlled. The body – materially, socially, and psychologically – marks a more or less impermeable boundary between Self and Other.

There is no universal agreement about what the body is or where it begins and ends. The diversity and complexity of embodied experiences demonstrate that bodies are constructed through different social, cultural, and economic networks and through different spaces and places. The sociopolitics that surround bodies and space are increasingly being critiqued and scrutinized. By examining bodies, geographers are able to understand further the ways in which places are embodied and bodies are emplaced.

Often referred to as the 'geography closest in', the body became the subject of social science enquiry from the late 1970s when humanistic geographers were concerned with the way bodies move through and occupy space. This early work was influenced by phenomenological approaches and geographers explored the body as a subject that can approach and connect with its surroundings and thus bring space into being. For many, the phenomenological approach was an important starting point for embodying geography. Humanistic geographers, however, were criticized for not considering the subjectivities or identities of bodies under investigation. It was not until the 1990s, and in response to criticisms of disembodied geography by feminist geographers that the body became an explicit part of the geographical agenda. Turning to 'the body' was also part of the 'cultural turn' in the discipline. Over the past decade or so the growing attention to consumption and lifestyle issues such as food, fashion, and leisure, meant that geographers could no longer ignore either embodied experiences or representations of bodies. It has been shown that people's desires to be involved in particular practices or activities is intimately connected to their own body image. It has also been suggested that certain types of bodies may seek to exclude Other bodies from their environments. As such, feminist, social, and cultural geographers are critical of the mind/body dualism and are exploring notions of identity, power, and embodiment.

Mind/Body Dualism

The seventeenth century philosopher Rene Descartes established a dualistic concept of mind and body. Western rationalist tradition entails a radical separation of body and mind that accords primacy to the mind. Descartes argued that the mind had the conceptual power of intelligence and hence selfhood. The corporeal body was deemed to be a vehicle which could be directed by the mind. This Cartesian division – the subordination of the body to the mind and the emphasis placed on dualistic thinking – had a profound impact on Western thought. In particular, medical practices established the body as an object that could be measured, mapped, and experimented on.

Geography, as well as other social sciences, has been built on the mind/body dualism which posits the mind and body as separate entities and acting on each other.

Not surprisingly, dualisms have shaped geographers' understandings of place and people. Dualistic thinking in geographical research tends to create distinctions between culture/nature, production/reproduction, public space/private space, work/home, Western/Orient, North/South, and state/home.

The mind/body dualism is gendered with the mind traditionally connected to positive terms such as reason, rationality, subject, culture, public, Self, and masculinity. The body, however, has been associated with negative terms such as passion, irrationality, object, nature, private, Other, and femininity. These dualisms are, however, very mobile. When the dominant subject (white, heterosexual, and masculine) is challenged, discourses shift to accommodate both sides of its constitutive oppositions.

Feminist geographers have argued that rational knowledge has been defined as masculinist which assumes a knower who believes 'he' can separate himself from 'his' body, values, emotions, and past experiences so that 'he' is autonomous, objective, and free of context. This kind of objective rationality – supposedly free of bodily identity and experience – is assumed to be universal and the only form of legitimate knowledge. Ironically, this type of rationality means that certain subjectivities become dominant and normalized – for example, white, bourgeois, heterosexual men become defined as the Self and all other subjectivities become 'his' Other.

This dualistic thinking has played a key role in the construction of geographical knowledge. There is a history of white heterosexual men excluding or marginalizing women as producers of geographical knowledge as well as what are considered 'women's topics'. Topics such as embodiment and sexuality were regarded as inappropriate topics to teach and research. In some places, this is still the case, and the body remains Othered in geography. What constitutes appropriate issues and legitimate topics to teach and research in geography comes to be defined in terms of reason, rationality, and abstraction as though these can be separated out from passion, irrationality, messiness, and embodied sensation.

Many geographers have, however, challenged the privileging of the mind over the body within the discipline which has resulted in an array of new topics (those topics previously considered too 'dirty' or 'risque') being put on the geographical agenda. Geographers themselves are beginning to think about ways of researching which recognize that all knowledges are embodied and situated.

The body has acted as geography's Other in different ways at different times and places. Philosophers from the periods of ancient Greeks and Romans, Judaeo Christian thought, and the Renaissance to the present day have been fascinated with the body. While there is no one consensus on what the body is, it is useful to explore what seems to be a simple question: what is the body?

What Is the Body?

'The body' is often treated as a taken for granted entity that needs no explanation. We all have bodies; however, how and where we use our bodies varies greatly. Bodies are used to ground personal identity and to recognize other people's identity. Bodies are used for the assignment of various roles, tasks, and strategies, for practical action, for reproduction, and for the expression of moral judgments. Bodies may also be used as artwork or as a surface upon which meaning is inscribed. With the myriad of meanings associated with the body it becomes almost impossible to define what the body is.

Some theorists who have attempted to define the body remain puzzled as to what it actually is. The body has been thought of as a corporeal place and even as *terra incognita.* At best, the matter of the body might be thought of as: flesh, organs, muscles, nerves, skeletal structure, and concrete material organized and animated through social and psychical inscription. Bodies become human (as opposed to animal bodies) when a body coincides with the 'appropriate' shape and space of a psyche. In psychoanalytic terms, the body may define the limits of experience and subjectivity. This 'thing' – the body – is complicated and its meanings have changed over time and in different places. It may not be very useful, therefore, to attempt to offer an absolute or fixed definition of the body.

In trying to understand what the body is, human geographers have drawn on many different approaches from a variety of disciplinary areas such as feminist and gender studies, cultural studies, sociology, philosophy, social anthropology, and geography. As noted above, geography is a discipline that has traditionally ignored the body. During geography's engagement with positivist thought, bodies were quantified out of existence to the point that people of different shapes, colors, sizes, ages, and gender sexualities were reduced to points, flows, and movements in space.

An early exception to this type of disembodied quantitative geography was 'time geography'. Developed by Swedish geographers in the 1960s, time geography was a contextual approach to understanding human spatial behaviors based on the idea that space and time are resources on which bodies can draw in order to participate in personal projects or activities. Based on the assumption that a body's ability to do more than one thing at a time is limited and that people cannot be in two places at once this approach sought to examine the interweaving of bodies and activities in logical blocks of space and time. Life paths were mapped to show that a daily path was the sum of bodily movements. Cumu lative maps of life paths were deemed to represent both the end result and contest of people's actions. The ex ploration of time–space routines became an important concept in the development of structuration theory. Time geography has been criticized as offering a limited understanding of humanity because they failed to rep resent a body with feelings, memories, knowledge, in formation, and identities. Structuralist accounts were similarly disembodied as they reduced humans to lifeless units of labor.

Phenomenology was one of the first approaches that offered a more human centered geography. Humanistic geographers in the 1970s, intent on understanding the relationship between human identity and place, con sidered the prediscursive lived body. One of the most useful contributions to come out of the phenomeno logical approach to embodiment is that subjectivity is not located in the consciousness but the lived body. The body becomes a locus of experience that exists prior to all conscious reflection and knowledge. Bodily awareness becomes inseparable from the world of perception. This type of sensuous corporeality means that it is also the object of spatial relations – without the body there would be no space. Using this approach, the body has its own intentionality and capacity to direct behaviors and function as a subject which expresses itself in a pre conscious habitual and involuntarily way. A body may build up time–space routines where a series of behaviors or rituals are habitually repeated without conscious thought. Some humanistic geography scholarship re jected abstract theorization and modeling, yet other hu manistic geographers developed experiential frameworks for understanding the relationships between people and place. The 'triad of environmental experience' em phasized the importance of movement, rest, and en counter. Routine and repeated movements in familiar places can be thought of as choreographed. The concept of 'place ballet' was used to describe the choreographed but complex movements of several bodies simultaneously. Humanistic geographies of 'life worlds' provided an in triguing forerunner to an investigation of differently embodied subjectivities in space, yet they failed to rec ognize embodied differences such as gender/sex, race/ ethnicity, dis/ablities, age, class, and so on.

More recently, geographers with converging interests in embodiment and nonrepresentational theory have encouraged a re exploration of movement, practice, and senses. Nonrepresentation theory is concerned not so much with representations and textual discourse as the mundane and everyday practices and senses that shape bodies and places. Sensuous geographies usually involve considering some combination of touch, taste, hearing, smell, and sight although geographers have tended to privilege the visual over other senses. Attention to the surveillance gaze has opened up areas of research that involve understanding how and why bodies are inter preted and disciplined. Moving beyond the visual, how ever, has meant that geographers can understand the management of embodiment through other sensual in formation. Studies that focus on the social construction of disability and illness highlight embodied experiences of pain, what it means to be healthy and/or sick.

Hence, attention to things that words and represen tations cannot express is necessary to understand the rarely spoken of but constantly performed practical ex periences of ordinary people. Exploring the way people 'do' subjectivities, there is now a large geographical lit erature on the way people dance, swim, walk, or sit their identities, as well as the way identities are performed through bodily gestures, styles of dressing, and modes of eating. Theories about the development of selfhood and identity through embodiment are combined with ideas about the way the body becomes caught up in collective experiences and power relations.

There is a growing area in human geography that studies how power is articulated through geographies of exclusion. Some bodies are deemed to be 'out of place' in particular spaces and the exclusion of certain people happens at a variety of scales. The effect of this exclusion is often understood as Othering. To explain further, bodies are considered as markers of difference between Self and Other in ways that maintain distinctions between valued landscapes and marginalized landscapes. Psycho analytic theory is another important approach to under stand the relationship between the Self and the society, why feelings of repulsion and fascination attach to par ticular bodies, and how some bodies are labeled as Other.

Psychoanalytic approaches tend to focus on the un conscious, rather than the conscious, and this has enabled some geographers to try to understand the relationship between the individual and the external world. It is often argued, and hotly contested, that the unconscious is formed from individual's sexual drives and residues of infantile experiences, and hence the experiences of early childhood are generally accepted to be critical. Children may cope with early painful and upsetting experiences by repressing them from their consciousness. It is from this repression that a split is formed between the conscious and unconscious. Theories of socialization suggest that people suffer crises of identity if they do not conform to the positions described by the 'Oedipus complex'. This Greek story suggested that both male and female children regard their father as an opponent and competitor for the exclusive love of their mother. The female child supposedly experiences 'lack' which can only be resolved through the displacement of desire from mother to father.

Some feminist geographers interested in the ways in which bodies become sexed/gendered have adopted this psychoanalytic work. In particular, the idea of the Oedipal phase has been taken up to rethink the formation of patriarchal culture, spaces, and sexed bodies. Some psychoanalytic thought offers theories of how people's subjectivities – such as mother, father, daughter, and son – are formed. Psychoanalytic approaches are useful to understand how certain subjectivities and identities – women, children, ethnic minorities, gay men, and lesbians – may be both socially and spatially marginalized. Ideas from object relations psychoanalytic theory have been used by geographers to help understand how some groups of people are feared or loathed and how this attitude to cultural difference is then played out in space, with particular attention to how groups have been subject to spatial segregation and regulation. Cultural and social values in Western societies construct particular groups as abject – 'dirty' or 'polluting'. People may respond to these 'abject Others' with hatred and attempt to create social and spatial boundaries to exclude or expel them. This approach is useful to understand how exclusionary landscapes are developed in different times and places.

Another approach is to conceptualize the body as a site of cultural consumption and inscription upon which social meaning is inscribed. The body becomes significant mainly in terms of the social systems or discourses that construct it and make it intelligible. The philosopher Michel Foucault noted there are a myriad of techniques through which bodies are subjugated, disciplined, and populations controlled. In relation to sexuality, disciplinary techniques include laws concerning the policing of particular sexual practices so that same sex practices are repressed or censored. Ideas of morality, social laws, and values are inscribed on the surfaces of bodies. Bodies are 'disciplined' so that certain sexual practices are deemed pleasurable (and acceptable) while others are deemed unacceptable and in need of reform.

This approach shows that bodies are historical constructs that have been disciplined, regulated, and controlled. The individual body is the effect of an endless circulation of power and knowledge and using this approach it is possible to understand how whole populations are controlled. Many geographers have used this approach to understand bodies and spaces in, for example, the asylum, the workplace, the prison, and the state.

Within this far reaching 'body as a site of cultural construction' approach some geographers have emphasized the active and transformative role the body plays in relation to the capitalist processes that produce it. Focusing on the connections between appearance, esthetics, and capital show how the body is tied into hegemonic social identities. The body, it has been shown, can be conceptualized as an accumulation strategy, being used by people to acquire money and status. The relationship between social status and modes of dressing is complex and geographers have researched, to name just a few, merchant bankers who 'power dress' in the City of London, young Muslim women who prefer to wear 'Asian' clothing but conform to Western clothing in UK secondary schools, as well as body projects such as plastic surgery, body building, tattooing, and piercing.

Feminist geographers taking this constructionist approach to embodiment are able to trace the history of the way genders and sexualities have been discursively produced. It is possible to identify how heterosexual relations have been produced as the norm and other forms of sexuality have been classified as Other or deviant. As such, constructionist feminists argue that bodies are produced by discourses and that essentializing discourses – references to biology and physicality – tend to naturalize that which is social difference. A constructionist approach opens up opportunities for geographers to consider how place is written on the body and vice versa.

A focus on the body as discourse may ignore the materiality of embodiment, and it is not necessarily useful or indeed possible to conceptualize constructionist and essentialist approaches as distinct and separate. These opposing positions, it has been argued, are inseparable. The physical reality of material biology (a sexed body) is often considered by constructionists as irrelevant to women and men's gendered positions in particular places. The distinction between sex (biological differences between men and women) and gender (masculine and feminine) does not hold and is a product of historically specific Western dualistic thought which laid the foundations for the development of modern science. Philosophers have examined this dualism, plus others, such as mind/body, culture/nature, reason/passion, white/black, and positive/negative.

Corporeal Geographies

The growing scholarship in corporeal geography is evident in the vast range of studies and approaches to understanding the body. One of the most popular approaches, for geographers and other social scientists, continues to be the cultural inscription approach which draws on post structuralist theories to understand the body as a surface to be etched by cultural and social systems. While our bodies make a difference to the experience of places, we might also think of bodies and spaces as mutually constituted. Instead of thinking about space and place as preexisting sites in which bodily performances occur, some studies argue that bodily performances themselves constitute or reproduce space and place. Geographers have looked at the way in which bodies are gendered, sexualized, racialized, aged, and so on by, for example, workplaces, schools, leisure spaces, homes, suburbs, cities, and nations. Hence, the mutually constitutive relationship between bodies and places can be examined at a variety of scales, from the space of the body itself, to intimate geographies such as the bathroom, bedroom, changing room, and other home spaces.

Some geographers argue that bodies and spaces are performative, that is, they have no ontological status or fixed characteristics. In this way, space is not just a backdrop for bodies rather it plays an important role in constituting and reproducing social relations. Bodies and spaces are simultaneously material, imaginary, symbolic, and real.

Understanding bodies and spaces as performative offers geographers new ways of thinking about power relations. One such way is to think in terms of space as paradoxical. Thinking of space and bodies in this way means working with the idea of simultaneously occupy ing center and margin (masculine and feminine, hetero sexual and homosexual, young and old, and so on). Bodies cannot be understood as just 'man' or just 'woman', but rather as sexed bodies that are contingent upon time and place. The commonly held belief that we all have to fit into the model of one sex or another has been contested as there are many bodies that do not fit the two sex model. For example, some bodies have both male and female genitalia (usually known as 'intersexed'), trans sexuals, people with XXY chromosomes, and even female body builders. These corporeal geographies may queer, and make paradoxical, space at a variety of scales from the home to international tourism sites.

Performative bodies and spaces are central to the newly emerging 'emotional geographies' scholarship. Important studies in emotional geographies have argued that people and places are imaged, felt, and experienced in ways that are radically entwined with each other. Feelings of shame, pride, guilt, and pleasure which are bound up in exercise, dietary, and cosmetic regimes re veal that bodies are emotionalized sites and hence an important focus for geographers. These lived experiences are entwined in multiple power relations realized in space/place. In the future, it is likely that human geog raphers will increasingly interrogate bodies (their own and others) as multiple, fluid, fractured, and constituted in particular spaces.

The body has been increasingly embraced in some sections of the discipline as a scale for considering social and cultural difference, so much so that in some cases 'the body' has come to mean the expression of fluid subjectivities. It is important to note, however, that the body still remains geography's Other and that the weighty materiality of bodies – particularly as conceptualized in and of biology – has yet to be fully explored. The taken for granted notion that biology stands outside culture, is one of the primary reasons why geographers have neg lected the body as an object of study and despite this recent surge of interest in bodies, the 'leaky', 'messy' zones of bodies and their resulting spatial relations, remain largely unexamined in the discipline.

See also: Affect; Ageism and Age; Cross-Cultural Research; Deconstruction; Disability and Chronic Illness; Embodied Knowing; Emotional Geographies; Emotional Knowing; Feminism/Feminist Geography; Feminist Methodologies; Foucauldianism; Gay Geographies; Identity Politics; Kolossov, V.; Lesbian Geographies; Masculinities; Non-Representational Theory/Non-Representational Geographies; Other/Otherness; Patriarchy; Performative and Embodied Mapping; Phenomenology/Phenomenological Geography; Poststructuralism/Poststructuralist Geographies; Pregnancy and Childbirth; Psychoanalysis; Psychoanalytic Theory/Psychoanalytic Geographies; Queer Theory/Queer Geographies; Racism and Antiracism; Self-Other; Sensorium; Sexuality; Situated Knowledge, Reflexivity; Soundscapes; Time Geography; Whiteness.

Further Reading

Bell, D. and Valentine, G. (1995). *Mapping Desire: Geographies of Sexualities*. London: Routledge.
Butler, J. (1993). *Bodies that Matter: On the Discursive Limits of Sex*. New York: Routledge.
Butler, R. and Parr, H. (eds.) (1999). *Mind and Body Spaces: Geographies of Illness, Impairment and Disability*. London: Routledge.
Davidson, J., Bondi, L. and Smith, M. (eds.) (2005). *Emotional Geographies*. Aldershot: Ashgate.
Grosz, E. (1994). *Volatile Bodies: Towards a Corporeal Feminism*. St Leonards, NSW: Allen and Unwin.
Hubbard, P., Kitchin, R., Bartley, B. and Fuller, D. (2002). *Thinking Geographically*. London: Continuum.
Johnston, L. (2005). *Queering Tourism: Paradoxical Performances at Gay Pride Parades*. London: Routledge.
Longhurst, R. (1997). (Dis)embodied geographies. *Progress in Human Geography* 21, 486 501.
Longhurst, R. (2001). *Bodies: Exploring Fluid Boundaries*. London: Routledge.
Nast, H. and Pile, S. (eds.) (1998). *Places through the Body*. London: Routledge.
Pile, S. (1996). *The Body and the City: Psychoanalysis, Space and Subjectivity*. London: Routledge.
Rich, A. (1986). *Blood, Bread and Poetry: Selected Prose 1979 1985*. London: Norton and Co.
Rose, G. (1993). *Feminism and Geography: The Limits of Geographical Knowledge*. Cambridge: Polity.
Sibley, D. (1995). *Geographies of Exclusion*. London: Routledge.
Teather, E. (ed.) (1999). *Embodied Geographies: Space, Bodies and Rites of Passage*. London: Routledge.

Relevant Websites

http://www.medienkunstnetz.de/themes/cyborg bodies/
Cyborg bodies.
http://www.iep.utm.edu/d/dualism.htm
Dualisms.
http://plato.standord.edu/entries/feminism self/bib.html#2
Feminisms and the body.
http://www.queerbodies.org/
Queer bodies.
http://www.bodymod.org/home.aspx
The practices of body modification.
http://www.museum.upem.edu/new/exhibits/online exhibits/body modification/bodmodintro.shtml
The University of Pennsylvania Museum of Archaeology and Anthropology 'Bodies of Cultures' exhibition.

Borderlands

D. Wastl-Walter, University of Bern, Bern, Switzerland

Glossary

Borderlands Physical locations or discursive spaces along historically developed political and material borders, where differences are constructed especially between nation-states, First and Third Worlds, and colonizer and colonized.

Cross-Border Cooperation (CBC) It aims to develop cross-border social and economic centers through common development strategies between adjacent regions within the European Union.

Frontier A fraction of space that refers either to the political division between two countries or to a part of a country that lies at the limit of the settled area.

Interreg An initiative started by the European Community, in 1989, that aims to stimulate interregional cooperation in the EU.

NAFTA North American Free Trade Agreement.

Permeability of Borders It refers to the degree of their openness. International relations and treaties define this permeability.

SAR Special Administrative Region An administrative division of the People's Republic of China Administrative Region. There are currently two, Hong Kong and Macao.

Schengen City in Luxembourg where the much-publicized treaty abolishing systematic border controls between participating countries of the EU was signed.

Territoriality It refers to the need of an individual, a social group, local economy, or state to claim authority over a segment of space.

Territory It describes a socially constructed division of space that is controlled by some form of authority. A person, group, local economy, or state can occupy territories.

Borderlands are spaces where normative systems meet. This can mean that the border is a barrier, hindering and controlling cross border activities and contacts, which has a clear impact on the adjacent borderlands. Borders can also function as bridges with the implication for the residents in the borderlands that they can benefit from the differences in the neighboring states and develop specific economic, social, and cultural cross border activities. A special setting for such opportunities exists when there is an ethnic minority that speaks the language and is familiar with the culture of the other side.

Borderlands are also functional spaces, where the asymmetries and differences between the neighboring states can be used for the benefit of at least one of the areas.

Borderlands with permeable borders are used by the local population as an integrated region for their everyday activities where they choose freely between the two nations for specific purposes. In this way, they develop an identity as border people, which can be stronger and more persistent than national identities.

Borders can function as barriers or bridges, and their function and meaning change over time. However, empirical research has demonstrated that even a change in the legal definition or function does not bring about an immediate change in people's minds: In the collective memory of a population, the meanings of a border persist to such an extent that the population's activities, especially interactions and perceptions of others, change very slowly.

Over the years, the field of border studies has also undergone major changes, which are outlined briefly in the following section. During the late nineteenth and early twentieth centuries, a period marked by imperialism and World Wars I and II, the primary motivation of the first generation of border scholars was to delimit, allocate, and classify state borders according to their morphology, their natural features, origin, and history. In the 1960s, borders were still seen as lines of (natural) differentiation containing (natural) entities. This period was primarily influenced by Ladis Kristof's 'The nature of frontiers and boundaries' in *Annals of American Geographers* (1959), as well as by Julian Minghi's classic review of boundaries studies in political geography, published in 1963 and by V. Prescott's work on the geography of frontiers and boundaries, published in 1965. In the 1970s, the paradigm changed and border scholars discarded the conceptualization of borders as a given, natural phenomenon, replacing it with a conceptualization of borders as 'artificial', social and political constructs. This shift can be understood in the context of the politicization of 'natural' or 'organic' characteristics of borders influenced by Friedrich Ratzel during the first half of the twentieth century.

Since the 1980s, border studies have taken on an increasingly interdisciplinary character and have started to focus more on the postmodern 'how' perspective. The new multidisciplinary generation of border scholars has turned away from the rather stringent analysis and typology of the functions of boundary lines toward the study of both how borders influence the perceptions, relations, and (inter)actions of groups living in borderlands and how borders affect the evolution of territorial identities.

International relations also affect borders, whose function is both defined by international treaties and dependent on central governments' treaties with neigh boring countries. However, most central governments are not situated in the borderlands, and this has a number of consequences. For example, most issues affecting people and their lives in borderlands such as cross border re lations, their local and regional economies, and border land security are – even in highly federalized states – matters of foreign policy. Thus decisive matters are de termined outside the borderland without considering the residents' specific situation and needs thereby reducing the population's influence. Borderland populations must adjust to (changing) border regimes and relations. These shifts impact inhabitants as they organize their spaces of (inter)action both on their own side of the border and across it.

Most borderlands are considered peripheries within their states, characterized by geographical marginality, the limited presence of state institutions, large proportions of indigenous populations or ethnic minorities, and a weak socioeconomic structure. If they are also economically and politically oriented toward their state centers, it is referred to as a 'back to back' relation between the borderlands. In contrast, some borders that are open and easy to cross can function as bridges. Cross border networks develop, and complementarity characterizes the relations. In such cases, people on both sides of the border share a mutual interest in working together – be it because of cultural ties or for economic reasons – which is termed a 'face to face' relation between borderlands.

Worldwide, borderlands are usually asymmetrical and therefore complementary. This is the case in contexts where the development stage of two neighboring states and borderlands is quite similar, for example, between European Union (EU) member states. In other cases, however, when the asymmetries of wealth and living standards are pronounced, the results are often tight se curity measures, particularly to control and regulate migration. This occurs, for example, when the North meets the South, along the US–Mexican border, or around the Mediterranean, where Europe meets Africa. Nevertheless, even when borderlands differ with respect to economic or social standards, their standards tend to merge toward those of the more highly developed side, which means that the less developed borderland usually develops rapidly, causing an eventual increase in dis parities within the respective nation.

The Permeability of Borders

Permeability, the degree of openness of borders, is of crucial importance for people living in borderlands. The twentieth century saw each type of border regime: closed, semipermeable, and permeable. Some border regimes even ceased to exist, at least in one certain re spect, for example, in economic terms.

In Europe, the Iron Curtain was experienced as an insurmountable barrier. Inhabitants on both sides of the border lived at dead ends, which meant that no contacts between them were possible, and the state hindered or tightly controlled crossing, which was only possible at specific border crossings. The Iron Curtain separated not only borderlands and century long neighboring popu lations but also opposing political and economic systems. It symbolized the demarcation between the West and the East. This type of demarcation still exists in other parts of the world. In Korea, for example, contacts between people from North and South Korea (see **Figure 1**) are still nearly impossible. On the state level, tentative meetings have recently started to overcome the barriers. On the level of the local populations, however, it is un likely that they will experience an opening of the tight border regime in the near future.

In other border contexts, control is strict and security tight, although borderlands have also become eco nomically dynamic areas. The US–Mexican border is a case in point, because goods can easily be moved back and forth between the US and Mexico. The Border In dustrialization Program (BIP), initialized by the Mexican government in 1965, has shaped this border landscape and its cross border relations: Industrial plants (*maquilas*) built on the Mexican side of the border guaranteed both lower cost assembly than on the US side and jobs for thousands of people living in the less developed Mexican borderland. Today the well developed Mexican border land attracts migrant workers from the southern part of the country and Central America. However, the border regime specifically seeks to prevent the free movement of people from the South to the North, thereby ignoring cultural ties, social links, and economic dependence be tween the local populations, for example, in the twin cities Ambos Nogales in Arizona, USA, and Sonora, Mexico. In contrast to the impact of this strict border regime on cultural and economic ties, it is economic interdependence that affects the population living in and around Melilla and Ceuta, two Spanish cities on the African side of the Straits of Gibraltar. Since the EU erected a high security fence to demarcate the border between Spain and Morocco and to prevent South–North migration, the local populations can no longer interact or continue their traditional trading.

Within Europe, discussions on borders as obstacles to the free movement of goods dominated national dis courses for decades. Moreover, the perspective that borders are barriers to mutual understanding between nations resulted in the joint effort of several European states that has become known as 'European integration'.

Figure 1 South Korean wire entanglement and luminous advertising heading to North Korea on the southern limit of the demilitarized zone between the two Korean states.

Even when border regimes are tight, efforts are made to facilitate cooperation or at least peaceful coexistence between neighboring communities. In communist times as well, people on the Italian–Yugoslav border were allowed to cross the border with special permits to work on the other side. In the 1990s, the EU started programs to overcome barriers and differences in structures, to compensate for disadvantages pertaining to location and unequal development, and to design border regions that allow the local populations in the borderlands to improve economically and to interact culturally and socially.

In Africa and other contexts, where borders were introduced without taking account of local and traditional situations, borderlands became sites of conflict in the twentieth century – either due to governmental questioning of the border, or state borders not coinciding with cultural borders. Although the asymmetry between the two sides there is also economically determined, it is more often determined by political circumstances that motivate people to flee one side of the border to seek a life of security and peace on the other side. In these borderlands, most people live in refugee camps.

Finally, ecological problems, which do not stop at national borders, very often challenge tight border regimes and borderlands. Borderlands thus face a double burden: A population may face ecological problems produced across the border and then not have access to the required financial and logistical resources to handle these problems effectively. There are few examples where adjacent borderlands have agreed to cooperate in case of ecological catastrophe, and where adjustments to laws have been made on the national level to equalize ecological standards (**Figure 2**). In most cases, borderlands are the most directly affected, yet least supported areas within states.

The Dependency of Borderlands on Binational Relations

Central governments usually oversee borderlands, which means that border related issues are considered a matter of foreign affairs. Several exceptions, however, such as the Italian Trent South Tyrol region, enjoy a certain degree of autonomy. Even though borderland populations are those who deal with their neighbors on a daily basis, they usually have few or no rights to decide on issues such as border crossing opening hours, number of border crossings, joint economic projects, and common efforts to overcome ecological challenges. As a result, the freedom of action of these populations is not only highly dependent on their political status within their states but also on the political relations and interdependencies between their own and the neighboring country.

To consider this dependency in more depth, the province Vojvodina in Yugoslavia provides an interesting example. The Vojvodina was integrated into the Hungarian region of the Austrian–Hungarian Empire until 1918. It became part of Yugoslavia after the end of World War I, making the Hungarian population a minority within the newly formed state. However, the population has always been multiethnic; today up to 29 different ethnic groups live there. In the 2002 census, about 300 000 inhabitants referred to themselves as Hungarian out of a population of about two million. Hungarians in the Vojvodina have historically kept their close ties to Hungary and they have been guaranteed distinct rights.

Figure 2 Modern US border barrier with concrete columns in the San Diego section, allowing small animals to pass.

Even in communist times, both central governments agreed to support the Hungarian minority: Hungarians in Yugoslavia had a special passport, only valid for the border district, which allowed them to keep in contact with their families. After 1989, residents could cross the border between Hungary and Yugoslavia without a visa, making Hungary a gateway to Europe also for ethnic Serbs and others. This changed in 2008 when Hungary joined the Schengen group, because crossing the border is now only allowed with a Schengen visa. Only Hungarians living in the Vojvodina are able to cross, based on a 'national visa' valid for 90 days, which has been issued by Hungary since January 2006.

Binational relations, and increasingly international agreements, not only affect minorities living in border lands but also and specifically the overall economic situation. Most borderlands are marginalized within their states and thus rely on, for instance, shoppers from across the border or the possibility to work on the other side. In some contexts, this interdependency guarantees a min imum income and in others, even results in economic boom. Ambos Nogales, twin cities straddling the US–Mexican border, offers one example of economic inter dependency. In other borderlands, particularly in Europe, for example, between Poland and Ukraine, the Nether lands and Germany, and in the three border area Germany–France–Switzerland (Regio Basiliensis), the population depends on its neighboring borderland(s). Residents commute back and forth, working on one side and living on the other, and borderlands like the Regio Basiliensis have a well developed and long tradition of interaction. However, all these areas are affected by de cisions made at the national level and how relations be tween the nations develop.

Recently, it has been the EU that has strongly im pacted the development of cross border relations. Taking the case of the Polish–Ukrainian borderland, it benefited from the open border for only a short period. When Poland moved politically closer to and finally joined the EU, the border regime tightened and local cross border interaction again became restricted. In Europe, relations between states are more or less free of friction, and international treaties and binational agreements that also consider the interests of borderlands regulate binational relations. There are, however, other contexts like the Middle East, South Asia, and Africa where problems and conflicts – including military disputes – between states are specifically anchored in borderlands. The most ex treme consequence is the militarization of a borderland, through which its population faces restrictions in its everyday activities and experiences the border as a clear demarcation line. Kashmir Region is a case in point. The total population of the three parts controlled by India (Jammu, Kashmir, and Ladakh) amounts to slightly more than nine million people. About three million people live in the Pakistani parts (Azad Kashmir and Northern Areas); the third part is Aksai Chin, occupied by China. When India became independent in 1947, Jammu and Kashmir, one of the princely states of India, had the option of joining either India or Pakistan. The majority of the population was Muslim; in the Valley of Kashmir, this proportion reached more than 95%. Military disputes were followed by the UN effort to mediate a truce that finally resulted in the partition of Jammu and Kashmir. The situation has not stabilized since; two nuclear powers, India and Pakistan, meet at the checkpoint, which the Indian side wants to transform into a per manent international border. Kashmiris themselves argue that Pakistan and India are interested solely in the geo political role of Jammu and Kashmir, not in its population and development.

Borderlands as Normative Spaces

Borderlands are territories that are subordinated to a certain national regime and are therefore subject to the norms of its specific political system. Hence along a border, two extremely different political systems may meet if subjected to extremely different political norms. This is best illustrated through the example mentioned above of North and South Korea. However, in numerous cases borderlands are also zones of cultural overlap where national identities (or a people's self concept) become blurred in the population. This blurring of identity oc curs for reasons of proximity and history: Different peoples live in each other's vicinity and borders are highly dynamic, shifting over the course of history. On the Italian side of the Italian–Austrian border, elderly people may have lived in different states with different norms and values over the course of their lifetimes. They became used to learning and speaking different lan guages, holding different passports, and continuing to find themselves in different neighborhoods. These people and their descendants, like those living in the Vojvodina, may have a clear concept not of their territorial be longing and national identity, but rather of their regional belonging and identity. Thus, these people often have a normative knowledge of both sides.

In many borderlands, political regimes deliberately or unwittingly exercise influence on the other side of the border. Along the US–Mexican border, norms and standards in addition to cultural values are transferred across the border. Companies in the Mexican borderland frequently orient themselves toward US corporate gov ernance principles. In other contexts, cooperation be tween organizations is possible only if one partner adapts to the norms of the other. In the Dutch–German bor derlands, companies can decide whether they want to adapt to German or Dutch laws, regardless of which nation they are located in. Moreover, within the EU, political, social, and economic norms are actively propagated within and across EU boundaries through a variety of different programs initiated in the late 1980s. In 1989, the EU started to fund Interreg cross border cooperation (CBC) projects with the aim of assisting adjacent regions on the internal and external borders of the EU. The initiative's purpose was to form partnerships in order to work together on common projects designed to strengthen economic and social cohesion throughout the EU. With the launching of the 'Poland and Hungary: Assistance for Restructuring Their Economies' (PHARE) program in 1989, the EU went beyond its boundaries by beginning to exert political and economic influence on Central and Eastern European Countries. The aim of the PHARE program was to assist applicant countries in their preparations for joining the EU. It was replaced by the program 'Community Assistance for Reconstruction, Development and Stabilization' (CARDS) in 2001.

These examples thus demonstrate that political systems have a major influence on borderlands and their popu lations. Such influence is persistent in the collective memory of borderland populations and in business inter actions and can, in certain cases, even manifest itself in the landscape. Finally, values and norms are represented in borderlands through practices and a range of ways in which the material world is organized. One illustration is the militarization of the border and distinctive checkpoint procedures, as along the US–Mexican border. A second is the frequent location of camps for asylum seekers in borderlands. Another is the type of special agreement that allows police to cross the border when a crime has been committed. While in Europe the border landscape has changed dramatically, in other contexts, national power and norms are manifested at the border and in the bor derlands through, for instance, checkpoints, signs with warnings, and the presence of an army.

Borderlands as Functional Spaces

Borderlands are not always marginal or peripheral areas that are poorly developed with bleak economic per spectives. They can also be places that function dynam ically, with a high density of shopping centers and industrial production facilities resulting in job opportun ities for locals and migrants. Further, borderlands are frequently characterized by a high degree of inter dependency among regions within their own country and even more frequently with their neighboring borderlands.

In China, the 'special administrative regions' (SARs) (**Figure 3**) on the internal Chinese border between Mainland China and both Hong Kong and Macao again demonstrates how state policies determine the permea bility of borders and the economic dynamics. The city of Shenzhen, which has today about 12 million inhabitants, was built up as a 'special economic zone' with regard to the neighboring city Hong Kong within the last 27 years. Along the US–Mexican border, *maquilas* – usually situ ated on the Mexican side – attract people from per ipheral areas of Mexico. Such factories take advantage of, for example, low labor costs and lower environmental standards. This idea of relocating the assembly process from one side of the border to another is not unique to the US–Mexican context. The area around Bratislava, Slovakia, adjacent to Austria, has become an attractive partner for selected European car manufacturers. Even when Slovakia became a member of the EU, it main tained its booming character. The Regio Basiliensis, re ferred to previously, is primarily a functional space based on a common interest to support the various border crossing activities of company owners and employees. Today, the Regio Basiliensis constitutes a highly integrated space, particularly in functional terms. Elsewhere as well, inhabitants live on one side of the border and work on the other, commuting back and forth as part of their daily routine.

However, economic activity alone does not produce a functional space. An exchange of know how between local authorities, the joint use of technical and educational infrastructure, and common environmental efforts are also required. The borderlands around Lake Constance (Germany, Austria, and Switzerland) have agreed on common ecological standards to protect the lake from pollution, thus promoting the border region as a single integrated tourist region. In Frankfurt (Oder) and Słubice on the German–Polish border, the University of Viadrina on the German side and the Collegium Polonicum on the Polish side were conceptualized as an integrated center for higher education. In other contexts, agreements have been signed to allow the fire brigade to cross the border in case of emergency and to give health center access to people from both borderlands. Agreements have also been signed for the joint use of technical infrastructure such as sewage plants, highway corridors, and special facilities for trucks.

Borderlands as Everyday Spaces

Above, borderlands were discussed primarily as cross border spaces, because the (inter)dependency between both sides of the border is usually very high, which is also true in everyday life. In borderlands, distinct everyday (inter)action spaces can be observed. The population copes with its given realities, which are greatly determined by a number of factors: binational relations, national border policies, the overall economic situation, cultural conditions, and personal experience. Everyday spaces in borderlands expand to become cross border spaces if the border regimes are sufficiently open.

Everyday spaces may also be understood as social, overlapping political and functional spaces. These spaces are continuously produced and reproduced in single actions, networks, and carrying out projects, thus ignoring state borders. Social spaces are relational, that is, they are produced and exist for as long as people interact with each other. They do not constitute clear demarcations or barriers inscribed in space, although political borders may be considered clear barriers. Social spaces mirror social and cultural links as well as economic dependencies. This has best been portrayed in Ambos Nogales on the US–Mexican border, where people on both sides of the border are socially integrated. There, people have clearly constituted social spaces and experience living across ‘the line’. Even after 9/11, when border enforcement tightened, their understanding of interactions and their production of (social) spaces did not change. They commute back and forth to shop, meet friends and family, and use the entertainment facilities on the Mexican side of the border. Along the Finnish–Swedish border, functional and everyday spaces are to

Figure 3 Border crossing station of the new western corridor between Mainland China (Shenzhen) and Hong Kong with co-located checkpoints on reclaimed land. This fourth crossing point with a capacity of about 60 000 vehicles and passengers per day was opened on 1 July 2007.

some extent highly integrated. Residents create a single integrated borderland in the course of their daily activities; they do not reflect on territorial distinctions as they go about shopping on one side and meeting friends on the other.

However, in everyday spaces, the two populations also interact with each other. If these interactions are considered social space, then the construction of 'we' and 'the other' requires discussion. It is primarily the border itself that demarcates two distinct settings and populations. The populations in adjacent borderlands live in different national contexts with diverse norms, values, and standards, which may affect everyday activities and allow each population to highlight the differences between itself and the neighboring population. Furthermore, borderland populations may also distinguish themselves from those in the rest of their country. In their special situation, inhabitants have developed an understanding of their area and neighbors; they assume parallels in understanding and perspective and express less interest in their respective central governments.

Borderlands and Border Identities

History and the collective memory of a population are sometimes stronger reference points in the construction of identity than regulations and restrictions. Above, the regional identity of the Austrian–Italian borderland population was highlighted. While the border and with it the political regime changed several times over the past century, the population continued to identify with local customs and was also able to protect their local dialects and customs. A similar situation prevailed in Lwiw (Lemberg; Lwow), a city in the central part of western Ukraine that was part of the Austrian–Hungarian Empire in the nineteenth century. Those born there at the beginning of the twentieth century experienced various regimes. They were Austrians before World War I and became Poles afterward. The city was occupied by the Soviets in 1939, by the German army in 1941, and from 1944 until 1991, it was under Soviet control again. Since 1991, it has been the largest city in western Ukraine. Today, most inhabitants speak Ukrainian in Lwiw, while in Kiev the inhabitants speak Russian. A walk through the city provides an experience of these changing political and cultural influences, which have been codified in street names and signs and architecture as well as in less material forms such as poems and narratives. Although it was strictly forbidden in Soviet times to speak German and Polish and to refer to their eras, the language and customs from both traditions have still survived and flourished. As in Ambos Nogales, such populations develop a regional identity as 'border people', which can be stronger than the changing national identities.

Current Research Issues

Over the past 20 years, we have witnessed an exponential growth of border studies, predominantly in Europe and North America. The enlargement of the EU in 2005 brought about a change in the meaning and permeability of the Union's internal and external borders, while 9/11 resulted in putting the meaning of the outer borders of the US into a new geopolitical framework.

In Europe, borders have been an even greater source of discussion since the enlargement of the Schengen space in 2008, which means a larger internally 'borderless' Europe with the new member states enforcing the outer borders of 'fortress Europe'. This has led to critical discussion in the previous member states and needs more research on the new neighborhoods.

In Africa, borderlands are under tremendous pressure by the great numbers of refugees, who flee as a result of civil war or environmental dangers. Here, not much research has been undertaken yet, so we are in need of a greater understanding of the processes of inclusion and exclusion in borderlands.

Another challenge for borderlands is ongoing climatic and environmental change, which involves the entire planet, regardless of borders. Nevertheless, state responses in the form of regulations will differ along the borders, thus creating substantially different economic environments.

In addition, both the special restrictions and potential advantages of borderlands are being researched. A partial list of research issues includes the constant confrontation between regional and national identities and the subsequent questioning of the nation state and its norms; the level of development as a function of national law, international relations, and asymmetries; borderlands as both lived and perceived spaces; and borderlands as socially and discursively constructed regions.

See also: Citizenship; Cold War; Economies, Borderland; Ethnic Conflict; Geopolitics and Religion; Hegemony; Nation; Political Boundaries; Regionalism; State; Territory and Territoriality.

Further Reading

Anzaldua, G. (1987). *Borderlands/La frontera: The New Mestiza*. San Francisco, CA: Aunt Lute.

Kolossov, V. (2005). Border studies: Changing perspectives and theoretical approaches. *Geopolitics* 10, 606 632.

Kristof, L. (1959). The nature of frontiers and boundaries. *Annals of the Association of American Geographers* 49, 269 282.

Martinez, O. J. (1994). *Border People. Life and Society in the US Mexico Borderlands*. Tucson, AZ: University of Arizona Press.

Mattingly, D. J. and Hansen, E. (2006). *Women and Change at the US Mexico Border. Mobility, Labor and Activism*. Tucson, AZ: University of Arizona Press.

Minghi, J. (1963). Boundary studies in political geography. *Annals of the Association of American Geographers* 53, 407 428.
Newman, D. (2002). Boundaries. In Agnew, J., Mitchell, K. & Toal, G. (eds.) *A Companion to Political Geography*, pp 123 137. Oxford: Blackwell.
Newman, D. and Agnew, J. (eds.). *Geopolitics*. Philadelphia, PA: Taylor and Francis Group LLC.
Newman, D. (2006). The lines that continue to separate us: Borders in our 'borderless' world. *Progress in Human Geography* 2, 143 161.
Paasi, A. (1996). *Territories, Boundaries and Consciousness: The Changing Geographies of the Finnish Russian Border*. New York: Wiley.
Paasi, A. (2005). Generations and the 'development' of border studies. *Geopolitics* 10, 663 671.
Patrick, M. J., Van Houtum, H. and van der Velde, M. (eds.) *Journal of Borderland Studies*.
Prescott, J. R. V. (1965). *The Geography of Frontiers and Boundaries*. Chicago, IL: Aldine Publishing Company.
Van Houtum, H. (2005). The geopolitics of borders and boundaries. *Geopolitics* 10, 672 679.
Wastl Walter, D. (ed.) (2009). *Research Companion to Border Studies*. Aldershot: Ashgate Publishing Limited.
Werlen, B. (1997). *Sozialgeographie Alltäglicher Regionalisierungen. Band 2 Globalisierung, Region und Regionalisierung*. Stuttgart: Franz Steiner Verlag.

Relevant Websites

http://www.aebr.net
Association of European Border Regions.
http://www.ccbres.sdsu.edu
California Center for Border and Regional Economic Studies, San Diego State University, California.
http://www.crossborder.ie
Centre for Cross Border Studies, Amargh, Northern Ireland.
http://www.qub.ac.uk
Centre for International Border Research (CIBR), Queen's University, Belfast, Northern Ireland.
http://www.nmsu.edu
Center for Latin American Border Studies, New Mexico State University.
http://www.ifg.dk
Danish Institute of Border Region Studies, Aabenraa, Denmark.
http://www.dur.ac.uk
International Boundaries Research Unit, University of Durham, UK.
www.kun.nl
Nijmegen Centre for Border Research, University of Nijmegen, The Netherlands.
http://www.ctc.ee
Peipsi Centre for Transboundary Cooperation, Tartu, Estonia.
http://www.sandiego.edu
Trans Border Institute, University of San Diego, California.

Bowman, I.

J. Dittmer, University of college of London, London, UK

Glossary

Geopolitics The application of geographic principles to political behavior.
Internationalism A political ideology that seeks engagement between states, ostensibly for the betterment of everyone involved.
Peneplanation The progressive erosion of a mountain system into an elevated plain.
Self-Determination A political principle that people ought to be able to determine their own form and structure.
Turner Frontier Hypothesis A hypothesis that America's unique identity was forged on the western frontier, a site of conflict and rugged individualism. The closing of the frontier in the late 1800s led to a great deal of concern about America's continuing moral fiber.

Bowman, Isaiah (1878–1950)

Isaiah Bowman was arguably the most influential American geographer of the twentieth century, serving as the ambassador of the discipline to policymakers and the nation at large (**Figure 1**). Born in 1878 and educated at Harvard and Yale, he began his career with a research focus on geomorphology but soon became interested in human–environment interactions, particularly in the Andes Mountains. He led several expeditions to the Andean regions, in 1907, 1911, and 1913. The purpose of his 1907 expedition was to investigate the hypothesis of William Morris Davis that peneplanation was responsible for many Andean landscapes. In this expedition he ini tially followed the shores of Peru and Chile before heading inland to Bolivia, Lake Titicaca, and Peru. Fol lowing the expedition, Bowman believed Morris's hypotheis to be true. Bowman's 1911 expedition was relatively short, taking place entirely in Peru, visiting Mollendo, Arequipa, and the Urubamba river valley. The goal of this expedition was to map the topography of the region from Cuzco to the coast. In the aftermath Bowman was forced to moderate his earlier claims about pene planation, including glaciation and volcanic activity among the list of factors. This expedition was largely eclipsed by the simultaneous 'discovery' of Maccu Piccu by Bowman's colleague and fellow traveler, Hiram Bingham. The 1913 expedition was a wide ranging trip through Chile, Argentina, and Bolivia, beginning in Buenos Aires. It was less significant in regards to his conclusions than the previous two trips.

At the time of Bowman's expeditions, the American Geographical Society (AGS) was a social club that dab bled in the financing and fêting of explorers. With the exploration of the Earth's poles nearing its conclusion, the organization hired Bowman as its director in 1915. He immediately refocused the organization as an institution of academic research, doubling its membership and in creasing its public presence.

Bowman's directorship of the AGS enabled him to become a public face for the discipline. President Wil son's need for geographic information during World War I with which to construct a post war peace led him to create the House Inquiry (1917–19), an idealistic attempt to institute self determination for the peoples of Eastern Europe through detailed mapping of cultural and other variables. As the public face of geography, Bowman was invited to participate, bringing with him the facilities and skills of the AGS. Bowman's contributions to the

Figure 1 Isaiah Bowman.

Versailles settlement that concluded World War I are many and various, but he is perhaps best known for championing the existence of a Polish state. The 'scien tific' answer to geopolitical problems would remain a hallmark of Bowman's career. After the war Bowman worked more as a political geographer than as a physical geographer, although he continued to focus on the rela tionship between landscape and people. Bowman was included among President Wilson's advisors in Paris after the war; however, his unhappiness with the failure of the US Senate to ratify the League of Nations led him to help found the Council on Foreign Relations, an organ ization of academics and intellectuals dedicated to pro moting an internationalist foreign policy for the United States. His highly successful post war text, *The New World* in 1921, portrayed the geography of the world as open to American hegemony and political dominance. In 1931 he was elected president of the International Geographical Union and in 1933 he became chair of the National Research Council.

Selected to be the fifth president of Johns Hopkins University in 1935, he inherited a university with a large deficit, and one of his major accomplishments during his tenure was to improve the university's finances through a capital raising campaign, and also through his alignment of university research agendas with the needs of the United States' military during World War II. Johns Hopkins scientists provided major support to the military through two contributions in particular: the proximity fuse used in antiaircraft ammunition and support in the Manhattan Project, the project to develop the atomic bomb. During the war, Bowman further aligned his vision of American hegemony with the disciplinary project. He raised funds for the creation of a school of geography at Johns Hopkins with the goal being to create a continual source of geographic knowledge to be fed into the pol icymaking apparatus of the state. The school of geog raphy, however, never matured, as personnel changes left the school always in a state of flux, and when Harvard closed its geography department in 1948, Johns Hopkins followed suit shortly thereafter.

During World War II, Bowman played a critical role once again as an advisor for President Roosevelt, assisting in a project to find frontier areas for refugees to colonize. Bowman's own inter war research had been on frontier settlements, in which he argued that although the Turner Frontier Hypothesis was true in a historical sense, there were still internal frontiers in which human adaptation to, and modification of, the environment could lead to expanded settlement. The Turner thesis had argued that the closing of the western frontier presaged the end of the process by which American identity had been forged. Bowman argued that the process could continue by turning inward. Arguing that this research could be ap plied to the refugee problem during World War II, in cluding the question of a Jewish homeland, Bowman led the M Project, which sought geographical solutions for population shifts. Later Bowman became incorporated into Roosevelt's inner circle of advisors in the negotiation of the post war order. Although he argued futilely against the division of Germany, Bowman was very influential in the shaping of the United Nations, again (as in his time with Wilson) arguing for an internationalist foreign policy. Fiercely anticommunist, Bowman viewed the Cold War stymieing of his post war architecture for peace and American prosperity as a product of Soviet intransigence.

After the war, he played controversial roles in the establishment of the National Science Foundation and in the disestablishment of the Harvard Department of Geography that he himself had graduated from. In his personal life, Bowman was known by acquaintances to be polite but aloof. Ambitious and hard working, Bowman had few hobbies, save playing table tennis with his daughter, Olive, as well as listening to music and going to the theater. He died in January 1950.

See also: Geopolitics; Irredentism; Ivy League and Geography in the US; Neocolonialism; Political Geography; Popular Culture; Tropical Geography.

Further Reading

Bowman, I. (1911). *Forest Physiography: Physiography of the United States and Principles of Soils in Relation to Forestry*. New York, NY: John Wiley.

Bowman, I. (1916). *The Andes of Southern Peru: Reconnaissance along the Seventy Third Meridian*. New York, NY: Henry Holt and Company.

Bowman, I. (1921). *The New World: Problems in Political Geography*. Yonkers, NY: World Book.

Bowman, I. (1924). *Desert Trails of the Atacama*. New York, NY: American Geographical Society.

Bowman, I. (1931). *The Pioneer Fringe*. New York, NY: American Geographical Society.

Bowman, I. (1936). *A Design for Scholarship*. Baltimore, MD: The Johns Hopkins University Press.

Bowman, I. (1937). *Limits of Land Settlement: A Report on Present Day Possibilities*. New York, NY: Council on Foreign Relations.

Bowman, I. (1939). *The Graduate School in American Democracy*. Washington, DC: U.S. Government Printing Office.

Bowman, I. (1942). Geography vs. *geopolitics. Geographical Review* 32, 646 658.

Livingstone, D. N. (1992). *The Geographical Tradition: Episodes in the History of a Contested Enterprise*. Oxford: Blackwell.

Martin, G. J. (1980). *The Life and Thought of Isaiah Bowman*. Hamden, CT: Archon Books.

O'Tuathail, G. (1994). The critical reading/writing of geopolitics: Re reading/writing Wittfogel, Bowman and Lacoste. *Progress in Human Geography* 18, 313 332.

Smith, N. (2003). *American Empire: Roosevelt's Geographer and the Prelude to Globalization*. Berkeley, CA: University of California Press.

Brain Drain

J. Crush, University of Cape Town, Rondebosch, South Africa
C. Hughes, Carleton University, Ottawa, ON, Canada

Glossary

Brain Drain Emigration of educated and skilled labor power or intellectuals from their native country.
Diaspora Individuals living outside their countries of origin.
Emigration Exit from a country of origin or residence.
Immigration Entry into another country.
International Migration Movement of individuals across international boundaries.
Remittances Monies sent from migrants abroad to their countries of origin, usually to family members.
Return Migration The return of an emigrant to his/her country of origin.

Introduction – Not Just Drains and Gains

The globalization of international migration has greatly enhanced the international mobility of people with skills. In a global skills market, skills (rather than citizenship and nationality) are the passport to movement from one country to another. Skills do not move evenly across space, however. Underlying current geographical pat terns of skills migration are profound economic, political, and social differences between countries. These differ ences are usually viewed in the aggregate as involving two major blocs of states: the North and the South, the developed and the developing worlds. People with skills are portrayed as relocating from South to North due to a variety of push and pull factors. These are almost always seen as a logical response to differentials in wages and income, lifestyles, personal security, political partici pation, prospects for career and professional advance ment, children's futures, and so on. In the developed world, graying populations and consequent skills short ages are seen as major factors precipitating a new global hunt for skills from the developing world. The repre sentation of skills migration in such binary terms underwrites the equally simplistic notion that one bloc of countries is suffering an irreversible 'brain drain' and another is reaping the rewards of an inevitable brain gain.

Virtually every country is both a sender and a re cipient of migrants. Within the so called North there are major movements of skilled migrants between countries. Similarly, within the South there is growing evidence of skills migration from one country or region to another. Further complicating the dualistic model is the fact that skilled migrants no longer conform to the old migration model of a one time relocation from country of emi gration to country of immigration. The opportunities for temporary relocation have grown exponentially. Most first generation skilled immigrants maintain strong eco nomic, social, and cultural ties with 'home'. The phe nomenon of transnationalism – of migrants literally and symbolically living their lives across space between two or more countries – is growing exponentially. Return migration from destination to origin seems to be on the rise in certain traditional sending countries such as Ire land. Finally, some countries in the 'South', notably the Philippines and India, have deliberately sought to de velop a skills base explicitly for export in the belief that skills 'loss' will actually produce a gain. The overall interplay between the negative and positive implications of skills migration is specific to each country and to various sectors and groups therein. The phenomenon of skills migration is certainly complex and analysis is hampered by the paucity of reliable data.

In this context, the concept of brain drains and gains, as currently conceptualized, may appear to be overly simplistic, more a concept for inflamed political rhetoric than a serious analytical device for understanding the causes and consequences of skills migration. Yet in international policy debates around the relationship be tween migration and development, the concept of the brain drain continues to have considerable resonance, if not pride of place. This article examines the common understanding of the concept, the evidence that is usually marshaled to demonstrate its reality and the migration dynamics that render it an increasingly blunt analytical concept.

Toward a Definition

Among the terms often used to describe skills migration, brain drain is the most prevalent. The term skills mi gration is employed in this article instead of the more commonly used skilled migration in order to indicate a migration or movement of embodied skills instead of incorrectly suggesting that migration has or can have skills. The term brain drain was coined by the Royal Society of London in a 1963 report to refer to the exodus of British scientists to the United States and Canada following World War II. Since then, with the rise of skills migration from less to more developed countries, the term is usually used to refer to migration from Africa, Asia, and Latin America to North America and Europe, but the concept is not confined to these contexts or directionalities. For example, the movement of skilled professionals from Canada to the United States is often represented as a brain drain by scholars and the media. A sound and inclusive definition of brain drain is the emigration of educated and skilled labor power, professionals or intellectuals outside of their native country, be it developing or developed. A few variations persist in the way the term is defined, including just the movement (especially permanent) of highly skilled migrants from the developing to developed world, to the actual overall depletion of skills from developing countries, to clear evidence that migration flows have had adverse consequences for the sending economy. Brain drain is not used to describe migration from colonial powers to colonies, international movement of talent within a transnational firm, or migration of individuals with relatively little education or skills training.

A difficulty in providing an easy definition of brain drain is how to determine whether or not someone is skilled or highly skilled. Most authors use these terms to refer to individuals with some university or post secondary education, but sometimes highly skilled refers only to science and technology workers. Others have suggested that any individual who is part of the functional core of the economy is potentially a brain drain candidate. Brain drain concerns have been extended to include public sector professionals such as nurses and teachers, as well as students. Indeed, the term is now used most often in the context of the health and education sectors.

Global Patterns of Brain Drain

Each region, country, and sector exhibits different patterns of skills loss and gain, but some global patterns have been discerned in the statistics. The emigration rate among the tertiary educated has been estimated at 41% for the Caribbean region, 27% for Western Africa, 18% for Eastern Africa, and 16% for Central America (see **Table 1** for figures by African countries).

Some estimates suggest that one in ten tertiary educated adults born in Africa, Asia, and Latin America resided in America, Australia, or Western Europe in 2001, but this figure rises to between 30% and 50% for individuals trained in science and technology.

The Blame Game

Many African, Asian, and Latin American countries blame Europe and North America for the loss of their skilled personnel, claiming that skills are being 'raided' or 'poached' in increasing numbers, especially from the healthcare sector. The implicit argument is that without active government and private sector recruiting, skills out migration would be negligible. In a globalized world, this assumption is highly problematic. What is true is that Europe and North America have developed skills based immigration policies over the last two decades, which make it much easier for skilled individuals from Africa, Asia, and Latin America to emigrate there. The temptation to admit skills paid for by someone else has proved almost irresistible.

Table 1 Estimates of the brain drain from Africa: emigration rates for tertiary educated, 2000

Percentage of nationals with university education living abroad	
<50	Cape Verde, Gambia, Seychelles, Somalia
25 50	Angola, Equatorial Guinea, Eritrea, Ghana, Guinea Bissau, Kenya, Liberia, Madagascar, Mauritius, Mozambique, Nigeria, Sao Tome and Principe, Sierra Leone
5 25	Algeria, Benin, Burundi, Côte d'Ivoire, Cameroon, Chad, Comoros, Congo, DRC (formerly Zaire), Djibouti, Ethiopia, Gabon, Guinea, Malawi, Mali, Mauritania, Niger, Morocco, Rwanda, South Africa, Senegal, Sudan, Swaziland, Tanzania, Togo, Tunisia, Uganda, Zambia, Zimbabwe
<5	Botswana, Lesotho, Burkina Faso, Central African Republic, Egypt, Libya, Namibia

Source: Kapur, D. and McHale, D. (2005). *The Global Migration of Talent: What Does It Mean for Developing Countries?* Washington: Center for Global Development. http://www.cgdev.org/content/publications/detail/4473/

Impacts of Brain Drain

What does it mean for Asian, African, and Latin American countries that lose skilled professionals to developed countries? Traditionally, skills out migration has been seen as an unqualified negative for these countries; the very term brain drain suggests as much. Certainly the general tenor of current policy thinking is that skills migration produces a general 'brain strain' on development. While information technology (IT) migrants have spurred economic development in a few source countries through the knowledge they acquire, skills migration from other sectors – most notably health and education – has done considerable damage to countries' abilities to provide basic services.

Absence Effects

The 'absence' effects of brain drain are generally viewed as being very negative. Economically speaking, a country loses an emigration surplus, the difference between the value an emigrant was adding to the economy and that which the emigrant was being paid. This is a direct money measure of the economic loss of the emigrant's absence. Other potential effects include higher skill premiums that increase inequality through greater remuneration to people with increasingly scarce skills; fiscal losses through lost tax revenue; diminished scale economies; loss of role models and spillover knowledge from most skilled to lesser skilled individuals; loss of entrepreneurs; and weakened comparative advantage. Overall, human capital is lost which has implications for gross domestic product (GDP), entrepreneurialism, training the next generation, flows of foreign direct investment, provision of health and education services, and a country's capacity to build or strengthen critical domestic institutions, such as governments, schools, and hospitals. The absence of skilled personnel impacts every sector of the economy and society, and undermines the feasibility of meeting the Millennium Development Goals as the health, education, and private sectors, along with government, suffer from a lack of much needed skills and knowledge. With respect to skill sets lost, neoclassical theorists point out that those who remain behind can experience a loss of welfare if a loss of scarce skills causes externalities, such as poorer healthcare provision.

'Revisionist' approaches to skills migration take issue with the concept of the brain drain itself and suggest that skills migration should not be conceptualized according to winners and losers; the reality for many countries is a mix of positive and negative effects. Factors determining the country and sector specific impact of skills loss include the volume of cumulative loss, educational selectivity, the specific skill sets of emigrants, the timing of and permanency of migration, and the characteristics of the global economy.

Considered on its own, how much harm can absence do? Global data suggest that it largely depends on how many people leave (cumulative loss or CL) and what levels of education they have (educational selectivity or ES). One study found that CL was by far the greatest in North America at 15%, followed by Africa with 7%, Asia with 5%, and South America with 3%. The regional picture differs when ES is examined. The percentage of emigrants with tertiary education is largest for Africa at 75%, followed by South America at 48%, Asia and Pacific at 20%, and North America at 17%. So the proportions of highly educated people who have left are largest in the least developed parts of the world, which arguably have the most to lose from skills migration.

Feedback Effects

Adherents of the revisionist approach emphasize the positive potential of skills out migration for Africa, Asia, and Latin America. The loss of the highly skilled may not always mean a loss for economic development. The revisionists assert that brain drain is a temporary stage and that a brain gain may actually result. Along with the migrant and receiving country, the developing country could be one of the three beneficiaries of a 'win win win' situation if skills migration is managed correctly. To understand how benefits could possibly be seen as accruing to source countries, it is necessary to consider what are referred to in the literature as 'feedback effects'.

Human Capital Development

One such effect is enhanced levels of human capital formation. In countries where skills emigration is prominent, these levels positively correlate with the probability of emigration. In other words, the prospect of emigration can act as a spur to personal skills development and training. Yet, the prospect of emigration often does not translate into actually leaving, resulting in what has been termed an optimal brain drain. Research on emigration potential in Southern Africa shows a major gap between the numbers of individuals thinking about or expressing interest in emigration and the numbers of those who actually leave. As long as the majority of better educated people do not emigrate, a net increase in education and human capital results. A study of 111 countries between 1960 and 1990 found that a 1 year increase in the average education of a nation's workforce increases the output per worker by between 5% and 15%.

Remittances

Research has clearly demonstrated that in a globalized migration system, those that leave do not automatically cut their ties with home. Indeed, the phenomenon of skills migration underlies the massive increase in the global volume of remittances in recent years. Remittances represent a significant portion of many nations' GDP, and globally account for twice the value of foreign aid. One study estimates that each remittance dollar 'multiplies' into \$2 or \$3 of gross national product (GNP) for the receiving country. The potential benefit of remittances is dependent on several factors, including how much and how often emigrants send them, what form they take, and where they are directed.

While they may remit less often than lower skilled migrants, skilled migrants are likely to send more money than lower skilled migrants because they tend to earn more. They also are more likely to direct remittances toward more productive (as opposed to consumptive) purposes, such as the purchase of bonds, the establishment of foreign currency accounts, or investment in business creation or philanthropic ventures. However, remittances tend to have greater developmental impact when they go to rural areas, which is not where highly skilled emigrants tend to originate.

Econometric studies suggest that, overall, remittances probably do not offset the adverse effects of skills migration. Furthermore, they may contribute to the dependency of developing countries on foreign sources of income. They amount to substantial and important portions of income for many countries, however, and should continue to be considered a loss mitigating factor.

Networks and Social Capital

While financial remittances should be considered, so should the networks that migrants and governments set up between home/sending and destination countries. Through these networks, migrants build social capital and transfer knowledge, which are critical for economic dynamism and institutional change. New technologies, management strategies, and trade opportunities can be diffused, contributing to the generation of new possibilities and economic growth. Emigrants can link domestic residents to international social networks, use their wealth to invest in home country projects, and act as transnational entrepreneurs. These connections and flows of ideas may have more significant long term consequences than remittances.

Some analysts, however, question the effectiveness of networks in meaningfully impacting economic development, pointing to the paucity of evidence of the success of expatriate knowledge networks in generating real net economic gains. The extent to which these networks can be established and utilized may be constrained by state limits in the receiving country on the power and scope of a diaspora's public activity because of security interests and by companies limiting the exchange of technical knowledge. Furthermore, knowledge exchange through diasporas about, for example, job markets and opportunities, may serve to further encourage skilled individuals to emigrate.

Whatever the actual impacts of diaspora behavior, migrants and immigrants do maintain closer ties with home than ever before leading to the phenomenon of 'transnationalism'. In a globalized world, many transnational migrants never really 'leave', literally living their lives between two places. Today's migrants are less likely to stay elsewhere permanently. With the help of improved transportation and communications technologies, many are able to travel back and forth relatively easily, to retain close contact with their home countries and communities and to build relations with other communities in third countries. The implications of transnational behavior for development in sending countries have yet to be fully realized or studied.

Return Migration

Finally, the potentially positive feedback effects of return migration should be considered. Many emigrants do eventually return to their country of origin, raising the possibility that the time spent away can be turned to the advantage of the home country. Indeed, some literature points to shifts from permanent to temporary migration and intrinsic intentions to return.

Some economists argue that return migration may be more effective in boosting development and wages than foreign assistance. This advantage can be garnered from the human, entrepreneurial, financial, and social capital that migrants may have accumulated while away. They may return with greater education and financial wealth, technological knowledge, different experiences and changed expectations, new ideas and international connections, and new investment channels, all of which can help spur development. Some benefits can accrue even if the return is temporary.

The benefits of return migration depend on a number of factors, however. First and most obviously, the number of returnees as a percentage of those that left is an important figure to consider, and the incidence of return varies greatly across regions and countries and itself depends on various factors. For the migrant, integration experiences and changes in migration plans play a role, as do the abilities of developing countries to generate the incentives required to entice their skilled emigrants to return home.

Other factors influencing the impact of return migration on development include: which migrants return (how skilled they are); the skills and knowledge acquired while away and the ability to transfer these upon return; how long a migrant has been away and how long he/she remains back home; and reasons or motivations for return.

With all the absence, prospect, and feedback considerations factored in, the brain drain picture that emerges is quite complex. Piyasiri Wickramasekara presents a helpful brain drain balance sheet that offers a quick overview of the potential positive and negative effects for sending countries, which is reproduced here (**Table 2**).

Can countries in Africa, Asia, and Latin America turn a brain drain into a brain gain through these feedback effects? On a balance sheet, for most countries, the answer would probably be no. Benefits tend to accrue more to large, relatively better off countries that have deliberate labor export policies, and to elites in these countries. The feedback effects of brain drain do not generally relieve pervasive poverty and inefficient use of human resources, or promote more equitable long term development in every developing country.

Policy Prescriptions and Possibilities

With the international mobility of highly skilled workers forecast to increase, many countries and regional groupings of states are taking steps to coordinate and implement policy responses. Most authors agree that

Table 2 Brain-drain balance sheet

Positive effects	*Negative effects*
• Provides rewarding opportunities to educated workers not available at home • Inflow of remittances and foreign exchange • Induced stimulus to invest in domestic education and individual human capital investments • Return of skilled persons increases local human capital, transfer of skills, and links to foreign networks • Technology transfer, investments, and venture capital by diasporas, aided by ICT • Circulation of brains promotes integration into global markets	• Net decrease in human capital stock, especially those with valuable professional experience • Reduced growth and productivity because of the lower stock of human capital • Fiscal loss of heavy investments in subsidized education • Remittances from skills migration may taper off after some time • Reduced quality of essential services of health and education • Students educated at government expense imply further drain • Opportunities for short-term movement is seriously constrained by immigration policies of developed countries • Causes increasing disparities in incomes in country of origin

Source: Wickramasekara, P. (2003). Policy responses to skilled migration: Retention, return and circulation. *Perspectives on Labour Migration,* 5E. Geneva: International Labour Office. Copyright © 2003 International Labour Organization.

countries must do away with a *laissez faire* approach. Strategic immigration and development aid policies must be developed, aimed at effective management of human capital that produces a win win situation for both sending and receiving countries. Skills migration cannot be prevented, but it can, in theory, be managed to the benefit of all parties concerned.

In general, policy options have been categorized into six types, or the 6 Rs. These are reparation, restriction, resourcing, recruitment, retention, and return. In short, the six responses suggest that: (1) destination countries should financially compensate source countries for the loss of skilled personnel; (2) sending countries should restrict who leaves and/or receiving countries restrict who enters; (3) countries experiencing a skills loss should source missing personnel from other countries; (4) receiving countries should make changes to their immigration and recruitment policies; (5) origin countries should address the conditions that prompt skilled people to emigrate; and (6) sending countries should focus on creating the conditions and incentives to retain skilled personnel and on attracting back those who have left.

A considerable political divide exists between countries of origin and destination as to which policies should be implemented. Many countries in Africa demand a cessation of recruiting, closing of the doors to skilled migrants by Europe and North America, and financial compensation for skills loss. The beneficiaries of skills migration have largely rejected the idea of direct monetary compensation, and argue that denying entry to people from certain countries violates the human rights of the individual who wishes to leave. They also maintain that origin countries ignore the downstream benefits of skilled professionals leaving, especially remittances and knowledge transfer through exchanges and return migration. Of the six policy responses, retention is generally seen as the best long term policy option, but many countries cannot create the necessary conditions on their own.

Global Cooperation

Despite the conflict of interests at the table, there are signs of cooperation. The United Kingdom and the Commonwealth have developed protocols to address the recruitment of healthcare workers and teachers from developing countries. As well, the General Agreement on the Trade in Services (GATS) may open the door to a more coordinated and mutually beneficial international mobility of skilled workers. The main objective of this agreement is the liberalization in the trade of services, which would ostensibly involve greater international movement of service providers. Many developing countries have a comparative advantage with their supply of inexpensive high quality labor, so such liberalization could provide a major opportunity for them, provided that the GATS terminology is nondiscriminatory and developed countries' immigration laws are in line with the agreement so as not to unnecessarily hamper the movement of skilled labor from developing countries.

As a sign of the urgent need to pay better attention to brain drain and related migration and development issues, the United Nations convened a High Level Dialogue on Migration and Development in September of 2006. Brain drain and other facets of the relationship between migration and development were discussed in order to identify appropriate ways and means to maximize its development benefits and minimize its negative impacts. These issues will be carried forward multilaterally at the new Global Forum on International Migration and Development, established at the High Level Dialogue.

Lack of Data

Hampering the ability to understand global patterns of skills migration and to develop policy responses is a lack of reliable information. There is a critical absence of hard data on migrant stocks and flows, their human capital content, and the actual impacts (negative and positive) of skills migration from particular developing countries. A recent exception is an econometric study on the impacts of brain drain that covered 59 countries. Prescriptions were made as to whether specific countries should allow the emigration of skilled personnel. In general, it was argued that most countries with low levels of both human capital and skills emigration are positively affected by brain drain, whereas it tends to have negative growth effects in countries where the skills emigration rates are higher than 20% and/or the proportion of the highly educated in the total population is above 5%. While a useful starting point, this study did not consider the feedback effects of migration or the sector specific impacts of brain drain in particular countries. There is an urgent need for better data collection systems and further studies, especially at individual country levels.

Conclusion

It is generally agreed that skills migration is likely to increase in the future, despite attempts to control and/or manage it. The policy debate on the brain drain is cur rently hampered by an overly simplistic conceptual ization of the problem itself. Not only is more and better data urgently needed, so too is a better understanding of the complexity of the problem itself. Viewing the world as two great blocs engaged in a zero sum game is all too prevalent in current international migration discourse. Researchers, geographers included, have an important responsibility to disaggregate the phenomenon of skills migration and to understand its specific impacts on in dividual sectors, countries, and regions.

See also: Diaspora; Emigration; Migration; Remittances; Transnationalism; Transnationalism and Labor Geography; Transnationality.

Further Reading

Beine, M., Docquier, F. and Rapoport, H. (2003). *IZA Discussion Paper No. 819: Brain Drain and LDCs' Growth: Winners and Losers*. Bonn: Institute for the Study of Labour (IZA).

Carrington, W. and Detragiache, E. (1998). How big is the brain drain? *Working Paper of the International Monetary Fund*. Washington: IMF.

Commander, S., Kangasniemi, M. and Winters, L. A. (2003). *IZA Discussion Paper No. 809: The Brain Drain: Curse or boon?* Bonn: IZA.

Crush, J. (2002). The global raiders: Nationalism, globalization and the South African brain drain. *Journal of International Affairs* 56, 147 172.

Docquier, F. and Marfouk, A. (2004). *Policy Research Working Paper 3381: Measuring the International Mobility of Skilled Workers (1999 2000)*. Washington: World Bank.

Faini, R. (2003). *Discussion Paper No. 2003/64: Is the Brain Drain An Unmitigated Blessing?* Toyko: World Institute for Development Economics Research, United Nations University.

Goldin, I. and Reinert, K. (2006). *Globalization for Development*. Washington: World Bank.

Kapur, D. and McHale, D. (2005). *Give Us your Best and Brightest*. Washington: Center for Global Development.

Kapur, D. and McHale, D. (2005). *The Global Migration of Talent: What Does It Mean for Developing Countries?* Washington: Center for Global Development.

Lowell, L. and Findlay, A. (2002). *Migration of Highly Skilled Persons from Developing Countries: Impact and Policy Responses. Synthesis Report*. Geneva: ILO.

Lowell, L., Findlay, A. and Stewart, E. (2004). *Asylum and Migration Working Paper 3: Brain strain: Optimising Highly Skilled Migration from Developing Countries*. London: Institute for Public Policy Research.

McDonald, D. and Crush, J. (eds.) (2002). *Destinations Unknown: Perspectives on the Brain Drain in Southern Africa*. Pretoria: Africa Institute and Southern African Migration Project (SAMP).

Straubhaar, T. (2000). *HWWA Discussion Paper No. 88: International Mobility of the Highly Skilled: Brain Gain, Brain Drain or Brain Exchange*. Hamburg: Hamburg Institute of International Economics.

Tanner, A. (2005). *Emigration, Brain Drain and Development: The Case of Sub Saharan Africa*. Helsinki: East West Books.

Wickramasekara, P. (2003). *Policy Responses to Skilled Migration: Retention, Return and Circulation. Perspectives on Labour Migration, 5E*. Geneva: International Labour Office.

World Bank (2006). *International Migration, Remittances and the Brain Drain*. Washington: World Bank.

Relevant Websites

http://www.cmd.princeton.edu
Center for Migration and Development, Princeton University.

http://www.migrationdrc.org
Development Research Centre (DRC) on Globalisation, Migration, and Poverty.

http://www.ilo.org
International Migration Programme, International Labour Organization (ILO).

http://www.iom.int
International Organization for Migration (IOM).

http://www.migrationpolicy.org
Migration Policy Institute (MPI), Washington DC.

http://www.queensu.ca/samp
Southern African Migration Project (SAMP).

http://www.un.org
United Nations High Level Dialogue on Migration and Development.

Brandt Commission

R. B. Potter and S. Lloyd-Evans, University of Reading, Reading, UK

Glossary

Brandt Line The bold line shown on the Peters projection map of the world depicted on the front cover of the two Brandt Commission Reports, showing the division of the world into the Global South and the Global North.

Development It is a frequently used term, but in the social and economic realm it basically refers to improve for improving the quality of life and material conditions experienced by people around the world. While the term applies globally, it is frequently used to refer to improvements in the conditions of life experienced in relatively poor countries.

Global Commons Derived from English law, the 'commons' are land areas that are shared by all, but owned by no one individual or group. The term 'global commons' is applied in a similar way to those parts of the atmosphere, hydrosphere, lithosphere, and biosphere that we all share (e.g., the air and the oceans).

International Monetary Fund The International Monetary Fund was created at the Bretton Woods conference and is closely associated with the World Bank Group.

Neoliberalism A term used to refer to new forms of economic liberalism, arguing that the path to development is best achieved by means of free trade and the deregulation of market systems. The approach was particularly associated with the New Right movement in the 1980s and with implementation in the form of Structural Adjustment Programmes (SAPs) introduced by the World Bank and the International Monetary Fund.

South or Global South Collective name given to the poorer countries of the world by the Brandt Commission. These are generally situated to the south of the globe. This term was used in preference to the then common collective term, the Third World.

Third World A term coined in the 1960s to pinpoint those primarily former colonial countries that were emerging as independent states, in contrast to the free market First World and the state socialist Second World.

United Nations The United Nations was formed as an international body or 'league of nations' to help ensure peace and stability and to promote international change. Its charter was signed in San Francisco in June 1945.

World Bank Formed at a conference held at Bretton Woods in 1944 in an effort to establish a new framework for world economic stability after the crisis of the 1930s. It is based in Washington.

Introduction: Background to the Independent Commission on International Development Issues (the Brandt Commission) and Its Report (the Brandt Report)

The Brandt Commission, properly known as the Independent Commission on International Development Issues (ICIDI), was established in 1977 with the express aim of examining the world's development needs in the 1980s and beyond. In the words of the Commission's Report, the intention was to emphasize the "…belief that the two decades ahead…may be fateful for mankind (sic). We want responsible world citizens everywhere to realize that many global issues will come to a head during this period." The overall tone, however, was at once both optimistic and globally mutual and inclusive: "The Commission agreed on the necessity for a thorough rethinking to create a new type of relationship which could accommodate all nations" (p. 10).

The suggestion that such a Commission should be created under the chairmanship of Willy Brandt had been made in the first instance by the then president of the World Bank, Robert McNamara, in a speech he made in Boston early in 1977 and he again referred to this in his address to the Annual Meeting of the International Monetary Fund and the World Bank in Washington later in the same year. Although, from the very first, the Commission was designed to be independent in its standing, the then secretary general of the United Nations, Kurt Waldheim, showed great interest in the formation of the Independent Commission, and accordingly it was established at the outset that the first copy of the Commission's Report would be presented to him.

The terms of reference of the Commission are reproduced on page 296 of its first and best known report, *North–South: A Programme for Survival*, and were notably broad and all embracing. Notably, they specifically included mention of the critical issues of economic and social disparities and absolute poverty:

> The task of the Independent Commission on International Development Issues is to study the grave issues arising from the economic and social disparities of the

> world community and to suggest ways of promoting adequate solutions to the problems involved in developing and attacking absolute poverty. As an independent commission it is free to raise any aspects of the world situation which the Commission considers pertinent and to recommend any measures it finds in the interest of the world economy.
> The Commission should pay careful attention to the UN resolutions on development problems and other issues explored in international fora in recent years. It should seek to identify desirable and realistic directions for international development policy in the next decade, giving attention to what their mutual interest in both the developed and developing countries should do (p. 296).

In pursuing these objectives, it was envisaged that the Commission's work would encompass direct consideration of the following:

1. The past record of development – specifically covering "the record of development in the Third World" (p. 296) and the "widening disparities of *per capita* income" (p. 296).
2. Prospects for the world economy – the Commission was to "examine relevant trends in the world economy, particularly for the 1980s but also looking further into the future" (p. 296). The Commission was to pay particular attention to expenditures on armaments and the political conditions affecting them.
3. Roads to a new international economic order – the Commission saw its role as above all, to convince decision makers and members of the general public "(t)hat profound changes are required in international relations, particularly international economic relations" (p. 297), including the need to restructure international trade, international finance, regulate multinational corporations, and to enhance the involvement of all countries in international development efforts.

A secretariat was established in Geneva during January 1978, marking the formal start of work of the Commission. The Executive Secretary, Professor Goran Ohlin (Professor of Economics, Uppsala University), and the Director of the Secretariat, Dragoslav Avramovic (who had held various Senior Economic Staff positions at the World Bank 1965–77) were responsible for draft proposals and other documentation prepared for the use of the Commission.

In respect of funding, the Dutch government enabled the work of the Commission to start by making a pledge that guaranteed the total costs involved in its operation. In the final analysis, the Dutch government financed about half the total expenditure of US$1.1 million, with the governments of Denmark, Finland, India, Japan, Republic of Korea, Norway, Saudi Arabia, Sweden, and the UK making substantial contributions. Regional organizations such as the Commission of the European Communities and the OPEC Special Fund also gave support. While, on the final page of *North–South*, it is stated that no formal follow up was envisaged after the publication of the Report, interestingly, again with the assistance of the Dutch government, an ICIDI office was established at the Institute of Social Studies in The Hague, so that comments and requests could be dealt with. The Commission was formally disbanded in 1983.

The first and best known Report of the Brandt Commission, *North–South*, was published in 1980 with the subtitle, *A Programme for Survival* (**Figure 1**). A sequel was published in 1983 as *Common Crisis: North–South Cooperation for World Recovery* (**Figure 2**). As explained in Annex 2 of *North–South*, the Commission's purposeful existence was announced on 28 September 1977 when Willy Brandt, the former chancellor of the then Federal Republic of Germany, announced at a press conference held in New York that he was about to launch and chair the independent commission. It is also noted in Annex 2 of the Report on how:

> This announcement had been preceded by a large number of private consultations and discussions. The

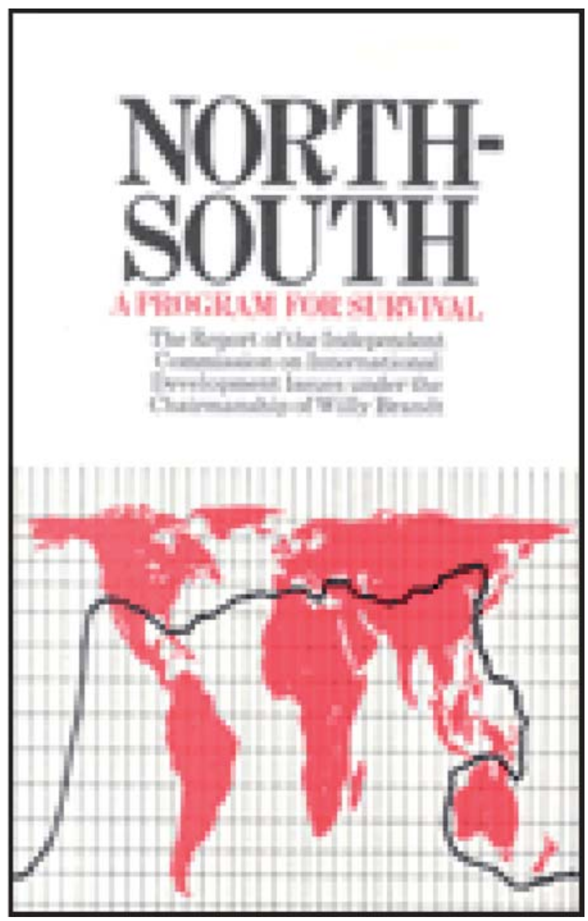

Figure 1 The cover of *North South: A Programme for Survival.*

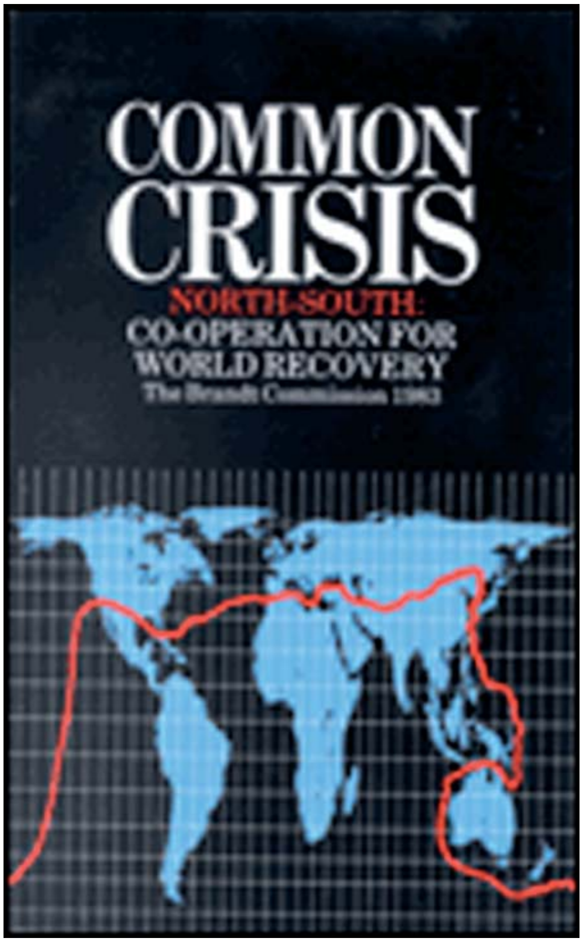

Figure 2 The cover of *Common Crisis: North South Co-Operation for World Recovery*.

> Chairman (sic) made a point of emphasizing that the Commission would not interfere in any way with governmental negotiations nor with the ongoing work of international organizations. Rather, such a group would have a supplementary function, to present recommendations which could improve the climate for further deliberations on North South relations (p. 293).

This brief summary of the Commission and its work points to several of its most important characteristics. On the positive side, the Commission emphasized the need for the so called northern and southern portions of the world to work together in tackling the interdependent development problems faced by the world. But the assurance that the deliberations of the Commission would in no way trespass on the roles of either governments or international development organizations meant that from the first, it might be left making meso level recommendations for policy action that governments and wider agents had little interest in implementing.

Nearly 30 years on from the publication of the Report, it seems fair to suggest that it is as famous for the iconic nature of its cover design (see **Figures 1** and **2**), as for its specific recommendations. The map on the front covers showed the world based on the Peters projection, rather than the then more familiar Mercator projection. Inside the title page it was noted that the Peters projection represented an important step away from the prevailing Eurocentric geographical and cultural conception of the world. The map showed by means of a single thick black line, the division of the world into the North or prosperous world, and the South or poor world. In wiggling its way across the globe, from the Pacific west of the Americas and back to the Pacific east of Oceania, the line served to emphasize just how large the Global South is in comparison with the Global North. Although the Commission has been criticized for reducing the diversity of development down to two categories, it is easy to understand that the motive for this was to emphasize the sharp polarity that characterized the global condition.

Membership of the Commission and the Schedule of Meetings

Willy Brandt regarded it as essential that the Commission was independent and that its members, no matter their background, should be invited to serve in a strictly private capacity. The Commission also wanted to ensure that the majority of its members were drawn from what were frequently described at the time as Third World (Southern) countries. The eighteen commissioners are listed in **Table 1** and consisted primarily of former politicians and diplomats, including three former heads of government, Willy Brandt (federal chancellor of Germany 1969–74), Edward Heath (prime minister of the United Kingdom 1970–74), and Olof Palme (prime minister of Sweden 1969–76). The commissioners were drawn from five continents and included members from Kuwait, Colombia, Upper Volta, Chile, Tanzania, India, Malaysia, Indonesia, Japan, Guyana, and Algeria, in addition to those from Germany, USA, UK, Canada, Sweden, and France (**Table 1**).

In the introduction to *North–South*, Willy Brandt provides an interesting assessment of his own credentials for the job of chairing the Commission, reporting quite candidly that "(i)n all frankness, my own record did not necessarily qualify me for the job. But learning from one's own shortcomings sometimes helps in addressing one's fellow citizens" (p. 8). As a young journalist, Brandt observes that he was aware of the problems of colonialism, the fight for independence, and the need to oppose dictatorship. Willy Brandt was awarded the Nobel Peace Prize in 1971 for his role in the *Ostpolitik* policies that achieved *détente* between the Soviet Union and the nations of the NATO alliance. But in a revealing passage referring to the short termism and state based orientation of national political leadership, Willy Brandt notes that "as a head of government, other priorities took up

Table 1 Members of the Brandt Commission (Independent Commission on International Development Issues)

Willy Brandt, Federal Republic of Germany
Abdlatif Y Al-Hamad, Kuwait
Rodrigo Botero Montoya, Colombia
Antoine Kipsa Dakoure, Upper Volta
Eduardo Frei Montalva, Chile
Katherine Graham, United States of America
Edward Heath, United Kingdom
Amir H Jamal, Tanzania
Lakshmi Kant Jha, India
Khatijah Ahmad, Malaysia
Adam Malik, Indonesia
Haruki Mori, Japan
Joe Morris, Canada
Olof Palme, Sweden
Peter G Peterson, United States of America
Edgard Pisani[a], France
Shridath Ramphal, Guyana
Layachi Yaker, Algeria

[a] Replaced Pierre Mendes, France who resigned from the Commission during the summer of 1978 for personal reasons.

Table 2 Schedule of meeting of the Commission

First meeting: December 1977	Gymnich, Germany (terms of reference established)	
Second meeting: March 1978	Mont-Pelerin, Switzerland	
Third meeting: May 1978	Bamako, Mali	
Fourth meeting: August 1978	Tarrytown, United States of America	
Fifth meeting: November 1978	Kuala Lumpur, Malaysia	
Sixth meeting: February 1979	Mont Pelerin, Switzerland	
Seventh meeting: May 1979	Annecy, France	
Eighth meeting: July 1979	Vienna, Austria	
Ninth meeting: October 1979	Brussels, Belgium	first editing session
Tenth meeting: December 1979	Leeds Castle, Kent	second editing session

most of my time and kept me away from realizing the full importance of North–South issues. I certainly did not give enough attention to those of my colleagues who at that time advocated a reappraisal of our priorities" (p. 9).

After the initial opening meeting held in Gymnich, Germany in December 1977 which included both open and closed sessions, four further meetings were held in 1978, followed by five in 1979, making ten formal meet ings in all (**Table 2**). At the last two meetings held in October 1979 and December 1979, drafts of the report were edited. An editorial group advised by experts was set up under the leadership of Edward Heath and Shridath Ramphal and prepared a complete draft of the report for submission to the tenth meeting, where the final text and its recommendations were adopted. In addition to the formal meetings listed in **Table 2**, Annex 2 of *North–South* lists details of talks in India and Japan, African contacts, Arab and other Middle East contacts, the EEC and European capitals, Soviet Union and Eastern Europe, other North American contacts, other contacts in the Asian and Pacific region, discussions in China, contacts with heads of international and regional organizations, and other contacts in Latin America and the Caribbean.

The Principal Recommendations of the Brandt Commission

In a passage that today seems remarkably prescient, the Commission's Report *North–South* clearly pinpointed what it saw as the principal threats to global order:

> If reduced to a simple denominator, this Report deals with peace. War is often thought of in terms of military conflict, or even annihilation. But there is a growing awareness that an equal danger might be chaos as a result of mass hunger, economic disaster, environmental catastrophes, and terrorism. So we should not think only of reducing the traditional threats to peace, but also of the need for change from chaos to order (p. 13).

If this quotation is read in the light of climate change and global warming, mass starvations, and the realities of global terrorism in the past 20 years, it appears highly prophetic. But despite the enormity of the issues that it was being suggested had to be faced if progress was to be made, from the outset, the measured and optimistic tone of the Report was clearly discernible:

> the moral dangers threatening our children and grand children can be averted; …we have the chance whe ther we are living in the North or South, East or West if we are determined to do so, to shape the world's future in peace and welfare, in solidarity and dignity (p. 7).

The main analyses and associated recommendations of the Commission were set out in the 17 chapters that make up *North–South*. Usefully, a full summary of the Commis sion's principal recommendations is provided as Annex 1 (pp. 282–292), under each of the main chapter headings. In presenting an outline of the principal recommendations of the Brandt Commission the account provided here follows that included in Annex 1 of *North–South*.

The Poorest Countries

The Commission called for an action program consisting of emergency and long term measures to assist the least developed countries, particularly those of Africa and Asia. Such a program would require additional financial assistance in excess of US$4 billion per annum at 1980 prices for the 20 year period from 1980 to 2000, in order to provide loans for regional projects in healthcare, dis ease eradication, water, soil and afforestation projects, industry, and transport among others.

Hunger and Food

Ending mass huger and malnutrition was identified as a major target. Low income countries should increasingly meet their food requirements, through expanded financial flows for agricultural development, thereby reducing food imports. Special attention should be given to agrarian reform, irrigation, agricultural research, enhanced storage, fertilizers, and fisheries development. International food security needs to be improved via international arrangements.

Population: Growth, Movements, and the Environment

The Commission stated its belief that development policies should include national population programs aiming at a satisfactory balance between population and resources. Family planning should be freely available to all. Migrant workers should be assured fair treatment and the rights of refugees to asylum and legal protection strengthened.

Disarmament and Development

This was a major theme and attention was drawn to the terrible danger to world stability that was posed by the arms race, and of the resources that are thereby diverted from peaceful development. Every effort should be made to secure international agreements to prevent the proliferation of nuclear weapons. A globally respected peacekeeping mechanism is needed to be built up, strengthening the role of the United Nations. The international community should become aware of the consequences of arms transfers and the export of arms producing facilities and reach agreement to restrain such activities to areas of conflict and tension.

The Task of the South

It was noted that social and economic reforms were also required within developing countries in order to complement the international environment for development, which itself needs to be more favorable. The redistribution of productive resources and incomes was called for where this had not already taken place. It was envisaged that broader changes would include expanded social services for the poor, agrarian reform, and the stimulation of small scale enterprises. The development potential of the informal sector needs to be appreciated and indigenous technological capacity strengthened.

Commodity Trade and Development

The Report stated that the commodity sector should contribute more to economic development through the greater participation of developing nations in the processing, marketing, and distribution of commodities. It was further urged that as a matter of urgency, action should be undertaken to stabilize commodity prices at remunerative levels. Tariff and other trade barriers against developing countries' products should be removed.

Energy

A call was made for the orderly transition from high dependence on increasingly scarce nonrenewable energy sources, and it was noted that prices that reflect long term scarcities will play an important role in this transition. Orderly and predictable price changes are important in facilitating a smooth development of the world economy. Special arrangements including financial assistance should be made to ensure supplies to the poorer developing countries.

Industrialization and World Trade

The industrialization of developing countries was seen as providing increasing opportunities for world trade and should therefore be facilitated as a matter of international policy. It was argued that protectionism threatens the future of the world economy and is inimical to the long term interests of developing and developed countries alike. Protectionism by industrialized nations against the exports of developing nations needs to be rolled back.

Transnational Corporations, Investment, and the Sharing of Technology

The Brandt Commission called for effective international codes of conduct and national laws to govern the sharing of technology, to control restrictive business practices, and to provide a framework for the activities of transnational corporations. Among other details, the investment regime called for by the Commission required reciprocal obligations on the part of host and home countries, covering foreign investment, transfer of technology, and the repatriation of profits, royalties, and dividends.

The World Monetary Order

It was argued that reform of the international monetary system needed to be undertaken urgently by all interested parties. These should involve improvements in the exchange rate regime, the reserve system, the balance of payments adjustment process, and the overall management of the system, which should permit the participation of the whole international community. Specifically, the participation of developing countries in the staffing, management, and decision making of the International Monetary Fund needs to be advanced.

A New Approach to Development Finance

This was central to the Commission's proposals, arguing that there must be a substantial increase in the transfer of resources to developing countries in order to finance various schemes. An international system of universal revenue mobilization, based on a sliding scale linked to national income, in which all but the poorest countries would participate, was called for. The Commission also stated that timetables should be agreed to increase official development assistance from industrialized countries to the level of 0.7% of gross national product (GNP) by 1985, and to 1.0% before the end of the century. The introduction of automatic revenue transfers through international levies on some of the following was called for: international trade, arms production or exports, international travel, and the global commons, especially seabed minerals. The establishment of a new international financial institution, referred to as the *World Development Fund*, was mooted. This would have universal membership, and decision making within it would be more evenly shared between lenders and borrowers. Lending policies and practices needed to be diversified.

International Organizations and Negotiations

It was central to the Brandt Commission that policies, agreements, and institutions in the field of international economic, financial, and monetary cooperation should be guided by the principle of universality. It was recommended that the United Nations system needed to be made more efficient and strengthened. Further, it was suggested that the performance of the various multilateral organizations in the field of international development should be monitored regularly, by means of a high level advisory body. Increased attention needed to be paid to educating public opinion and the younger generation, in particular, about the importance of international cooperation in the field of development. It was further suggested that limited occasional summit meetings of world leaders should be held to advance the cause of consensus and change.

An Evaluation of the Recommendations of the Brandt Commission

Behind the individual measures called for by the Brandt Commission, there were two key dimensions. The first was a need for the large scale transfer of resources to the countries of the South. The second was the recognition of the global mutuality of development and the need to act on such interdependence. In both these senses, the Brandt Commission's recommendations were visionary. The perspective of the Commission was of impending crisis and ultimate breakdown of the world system unless such mutuality and interdependence were clearly recognized and comprehensive reforms undertaken. The recommendations of the Brandt Commission thereby amounted to a strategy of global reformism, with the Commission talking of a new concept of global responsibility and new types of international relation that could accommodate all nations. The *North–South* report referred to the need for understanding, commitment, and solidarity between people and nations.

Critics, on the other hand, have argued that the Commission was unrealistic, idealistic, and naive, specifically in believing that Northern governments would easily be persuaded to pass on the recommended financial support to Southern governments, who in turn would pass assistance on to those most in need. A major issue here involves identifying who are the most effective agents of change, and historically, sovereign states have been dominant in this regard.

A further critique runs that although hailed as the first internationally recognized proposals for development reform, the initial enthusiastic response was soon overshadowed by the recession of the early 1980s. The father of dependency theory, Andre Gunder Frank, maintained that the impetus behind this 'new concern' was the economic crisis faced by the developed world and its search for new markets. Others commented that in addition to underestimating the apathy and intransigence of governments, the Brandt recommendations also underestimated the power of transnational corporations. The 1983 Report noted that the Cancun Summit of world leaders held in October 1981 as a direct outcome of the 1980 report made no tangible contribution to resolving the problems of developing countries and made no suggestions for putting new procedures in place. The 1983 Report acknowledged that increasing hardship had characterized the period from 1980 to 1983.

A further point is that the Brandt proposals were eclipsed by the rise of the New Right wing in the mid 1980s, spearheaded by President Reagan in the United Sates and Prime Minister Thatcher in the United Kingdom. By the end of the 1980s, the so called neoliberal agenda was taking shape. As part of this, the World Bank and the International Monetary Fund were promulgating private sector led development initiatives and decreasing public sector expenditures in the form of the conditions placed on loans as part of structural adjustment programs.

The *Brandt 21 Forum*, which seeks to keep the Brandt recommendations alive, comments that the Brandt Report Card makes for depressing reading. For example, globally official international development assistance has most definitely not increased to 0.7% and then to 1.0% of gross domestic product (GDP), it has fallen from 0.35% in 1980 to 0.21% in 2000. Governments have shied away from global solutions to global problems of poverty and development – a case in point being the call for global forms of taxation (e.g., on currency speculation as in the so called Tobin Tax), which governments dismiss, arguing that taxation is the concern of sovereign states alone. Some point to the charity initiatives such as *Band Aid*, *Live Aid*, *Drop the Debt*, *Make Poverty History*, and others as having achieved more than the actions of states. While there have been moves to write off debt, thus far progress can only be described as partial. While the Millennium Development Goals may be interpreted as truly Brandtian, they are showing just how much needs to be achieved and how little was achieved from 1980 to 2000. In short, it is suggested that since the Brandt Commission in the 1980s, the international community has not responded in any meaningful way and the inequalities pointed to by the Brandt Reports have widened conspicuously, not reduced.

See also: Dependency; Development I; Development II; Neoliberalism; Third World.

Further Reading

Brookfield, H. (1975). *Interdependent Development*. London: Methuen.
Desai, V. and Potter, R. B. (eds.) (2008). *The Companion to Development Studies* (2nd edn.). London: Hodder Arnold and Oxford University Press.
Frank, A. G. (1980). North south and east west paradoxes in the Brandt Report. *Third World Quarterly* 2, 669 680.
Hettne, B. (1995). *Development Theory and the Three Worlds: Towards an International Political Economy of Development* (2nd edn.). London: Longman Scientific and Technical.
Independent Commission on International Development Issues (1980). *North South: A Programme for Survival: Report of the Independent Commission on International Development Issues*. London: Pan Books.
Independent Commission on International Development Issues (1983). *Common Crisis: North South Co Operation for World Recovery.* London: Pan Books.
Potter, R. B., Binns, T., Elliott, J. and Smith, D. (2008). *Geographies of Development* (3rd edn.). Harlow: Pearson Prentice Hall.
Quilligan, J. B. (2002). *The Brandt Equation: 21st Century Blueprint for the New Global Economy.* Philadelphia, PA: Brandt 21 Forum.
Singer, H. (1980). The Brandt Report: A north western point of view. *Third World Quarterly* 2, 694 700.

Relevant Websites

http://www.brandt21forum.info
Centre for Global Negotiations.
http://www.stwr.net/content/view/43/83
Sustainable Economics to end Global Poverty.

Brown Agenda

J. Beall, London School of Economics and Political Science, London, UK

Glossary

Ecological Footprint The term 'ecological footprint' refers to how much land and water area a human population requires to produce the resources it consumes and to absorb its wastes under prevailing technology.

Global Commons The global commons are those things that no one person or state may own or control, and which are central to life; these include the atmosphere, biosphere, and hydrosphere, all of which are threatened by pollution and climate change.

Organization for Economic Cooperation and Development (OECD) The OECD was established in 1961, bringing together the governments of countries committed to democracy and the market economy to support sustainable economic growth and stability, boost employment and living standards, assist the economic development of other countries, and contribute to the growth in world trade. There are 30 full members, although views and expertise are shared with a much larger number.

Solid Waste Management (SWM) SWM is a generic term that refers to the collection, transport, processing, recycling, or disposal of waste materials in an effort to reduce their effect on human health or local aesthetics or amenities.

United Nations Development Programme (UNDP) UNDP is the UN's global development network, advocating for change and connecting countries to knowledge, experience, and resources to help people build a better life. They work on the ground in 166 countries, aiming to help people find their own solutions to global and national development challenges.

United Nations Educational, Scientific, and Cultural Organization (UNESCO) UNESCO promotes international cooperation among its 192 member states and six associate members in the fields of education, science, culture, and communication. The organization aims to create the conditions for genuine dialog based on respect for shared values and the dignity of each civilization and culture.

United Nations Human Settlements Programme (UN-HABITAT) UN-HABITAT is the United Nations agency for human settlements. It is mandated by the UN General Assembly to promote socially and environmentally sustainable towns and cities with the goal of providing adequate shelter for all.

Urban Management Programme (UMP) UMP is a joint effort established in 1986 by UN-HABITAT/UNDP and aimed at strengthening the contribution that cities and towns in developing countries make toward economic growth, social development, and the alleviation of poverty.

Water Supply and Sanitation Collaborative Council (WSSCC) WSSCC seeks to accelerate the achievement of sustainable sanitation, hygiene, and water services to all people, with special attention to the unserved poor, by enhancing collaboration and coordination of all sector stakeholders, nationally and internationally.

Introduction

With the increasingly high profile of global environmental concerns from the 1980s, the urban agenda for developing countries has seen progressive shifts toward a focus on urban environmental issues, collectively termed the 'brown agenda'. The most critical and immediate development issues facing cities of the developing world – those of water supply, sanitation, and waste management – were reconceptualized through an environmental lens rather than being seen solely in terms of ensuring urban service delivery and meeting the basic needs of the urban poor. This shift in emphasis was clearly evident in the discourses of the United Nations and World Bank in the early 1990s, with the latter heralding a 'new urban agenda'. Given the global scope of the 'green' movement, by this time, international development agencies, national governments, and NGOs found the concept of a 'brown' movement a useful vehicle for deflecting blame for environmental ills from urbanization and cities and for focusing on the local and immediate environmental concerns of urban dwellers. This environmental turn constituted both a potential threat to a focus on the needs and rights of poor people in cities and an opportunity to reprioritize urban development issues by feeding brown agenda concerns into wider environmental debates.

The issues that constitute the brown agenda are by and large not new; however, the relabeling of these issues as specifically environmental has a number of implications. From a development perspective, there are worries that shifting the focus of attention from the basic needs of poor people to the urban environment might undermine efforts at poverty reduction. From an environmental angle, the question is raised as to whether brown agenda issues ought to be thought of as separate from the concerns of the broader green movement or as localized expressions of the same agenda; if thought of as separate, there are implications about potential trade offs between brown and green issues. Adrian Atkinson has argued, for example, that the idea of a brown agenda can distract from the critical broader green one and result in a prioritizing of the short term over the long, and the local over the planet itself. Ultimately, while the codification of urban environmental issues into a clear agenda has been useful in highlighting the relationship between poverty and the environment at the local level, the brown agenda is best seen as a means of integrating these issues with wider green and development problems respectively, rather than as a separate or opposing agenda.

What Is the Brown Agenda?

In brief, the brown agenda refers to issues of safe water provision, sanitation, and drainage; inadequate solid and hazardous waste management; and air pollution including uncontrolled emissions from motor vehicles, factories, and low grade domestic fuels. The brown agenda emerged as a matter of concern and debate, in part, because of a perceived lack of attention to the specifically urban environmental problems of the developing world. This lack of a coherent urban environmental agenda was due to both the rural focus of many development debates and interventions, and the emphasis in environmental debates on natural resources and the global commons. In contrast, the gaze of the brown agenda is firmly trained on problems faced by cities of the global South, including rapid urbanization and growing urban poverty alongside poor urban management and inadequate service delivery. These problems in turn are inextricably linked to environmental degradation at city level.

In many respects, the brown agenda is an attempt to place developing countries at the forefront of both environmental and development debates and to confront the disconnect between global environmental 'green' problems and the problems confronting cities. The justification for an urban focus is clear. The United Nations estimates that in 2007 the world's urban population, as a percentage of the total world population, will surpass the rural population and by 2030, the proportion of the world's population living in cities is expected to reach 61%, and 79% of these urban dwellers will live in less developed countries. Cities loom large as 'engines of economic growth' and as perpetrators of environmental degradation, while also themselves being subject to some of the severest impacts of poor environmental management. Moreover, the environmental problems that cities face have very immediate and dramatic consequences. In this respect, the brown agenda serves to highlight environmental problems that are not necessarily the concern of future generations but are urgent problems whose impacts are being felt by millions of people already, problems such as the absence of safe and affordable drinking water or unnecessary mortality and morbidity from preventable diarrhoeal diseases. Certainly, problems such as inadequate water provision and poor sanitation facilities are not new in the developing world, but there is a consensus that urban environmental degradation has rapidly exacerbated these problems in recent decades.

That the global and regional problems of climate change, atmospheric pollution, ocean acidification, and land degradation are worsening is now well known and is often perceived as an inevitable consequence of unregulated economic development. But why problems such as basic sanitation, water provision, and waste management should be worsening is less clear. In part, the explanation can be found in the political and economic approaches championed in the developed world and increasingly adopted in the developing world, which have had negative and localized effects on cities in poor countries. Finding themselves in huge amounts of debt, many governments in developing countries have been encouraged to make investments that are likely to damage the environment in the long term. This has been accompanied by a 'retreat of the state' under neoliberal economic development policies, leading to failure on the part both of states and the private sector to secure urban environmental services for the poor. In part, however, the answer also lies in the dramatic increase in urbanization and urban poverty, which has had huge knock on effects for urban environments in which the urban poor effectively destroy their own environment through processes of survival. It tends to be the case that the poor create localized problems in their own environment, while the rich create problems impacting on a wider public.

In contrast to cities in developing countries, which predominantly degrade their 'own' immediate environment with negative health impacts for their own people, cities in the developed world often have an impact on the environment and global futures in ways that do not always affect their own populations in greater measure. Hence environmental degradation can take place on many different levels, but it is hard to expect wider green issues to be addressed adequately in contexts where brown agenda problems have such a damaging, inescapable impact on daily life. The following section briefly examines each of the major elements of the brown agenda in turn.

Main Elements of the Brown Agenda

Water Supply

Access to safe water and sanitation are two of the most urgent problems associated with the urban environment in the developing world. Globally, 1.1 billion people do not have access to adequate water supplies and more than double that figure – 2.4 billion – lack adequate sanitation facilities. These problems are particularly acute for the urban poor in many cities. For instance, UN HABITAT estimates that the number of African urban dwellers without access to adequate water provision is between 100 and 150 million or 35–40% of the population. Even where water is available, it may only run for a few hours a day via a pipe shared by hundreds of households. In Rajkot, India, a city with a population of 600 000, for example, piped water runs for only 20 min each day. Where water supplies are inadequate or unreliable, people are forced to purchase water, which can be financially devastating. In Mumbai, for example, the urban poor are often forced to buy water from informal vendors who may charge as much as forty times more per liter than the piped connections enjoyed by the wealthier households. Furthermore, access to water does not mean it is safe to drink. Wastewater treatment is so inadequate in many cities that the problem of contaminated water is an enormous environmental and health burden. Of India's 3700 cities and large towns, only 17 have full wastewater treatment facilities. As a result, the vast majority of wastewater is dumped into rivers, lakes, and coastal waters without treatment. This of course highlights one of the many linkages between the brown agenda and the health burden of the urban poor in cities of the global South.

Sanitation

Commanding less attention than water supply, the situation with regard to sanitation is nevertheless abysmal in most cities of the developing world, and again Africa suffers most of all. In cities such as Addis Ababa, Dar es Salaam, Kigali, Brazzaville, and Harare, more than 90% of the population live in homes with no connection to a sewer. Worse still, the city of Lagos – the largest city in Sub Saharan Africa with a population of over 13 million – has literally lost infrastructure over recent decades and now less than 1% of households is linked to a closed sewer system, this proportion largely consisting of hotels and wealthy compounds. Kinshasa, with a population approaching 10 million has no waterborne sewerage system at all. In these cities, pit and bucket latrines are often shared between hundreds or thousands of people, and open defecation is common. The situation in many parts of Asia is not much better, with sewer connections in Manila at below 10% and a study of slums in India, finding, in one case, just 19 latrines for 102 000 people. In some cities, communal toilet blocks for low income areas have become a focus of community organized activity, with successful examples from organized slum and pavement dwellers in Mumbai having inspired similar initiatives elsewhere. The Orangi Pilot Project in Karachi used community participation and self help strategies to help low income households in informal settlements acquire sewerage connections. However, while sanitation is an issue on which local community mobilization can bring about substantial improvements, it does of course crucially depend on connections into wider infrastructure for water supply and sewerage.

Solid and Hazardous Waste Management

Cities generate huge amounts of solid waste or garbage, with the amount of waste generated tending to increase with income. Nevertheless, it is the poorer cities that have the greatest problems dealing with waste collection and disposal. In Dar es Salaam, for example, with 75% of the population living in unplanned areas, only 10% of the 2000 tons of solid waste generated daily is actually collected. In Sambalpur, India, a survey indicated that 98% of households simply threw their rubbish out in the open as there were no rubbish bins available in the area. Very often waste is thrown into rivers and lakes with huge knock on effects for water pollution. In São Paolo, for instance, half the city's 'favelas' are located on the banks of the reservoirs supplying water to the metropolis, and those living in the slums throw their waste directly into this reservoir or the brooks running into it.

Inadequate waste disposal can also exacerbate problems of drainage and poor drainage, waste management, and sanitation interacting together can be devastating on the urban environment and health – piles of rubbish block drains leading to stagnant pools of water, breeding disease and attracting mosquitoes that spread malaria and dengue fever. Uncollected domestic waste is the most common cause of blocked urban drainage systems in Asian cities. In these situations, urban waste also often mixes with excreta, spreading pathogens and waterborne diseases around communities.

Low income areas are often so cramped as to be completely inaccessible to refuse collection vehicles. One method often resorted to is the burning or burying of rubbish, both of which can be environmentally damaging. Even where waste is collected, toxic and hazardous wastes generated by industry are often dumped untreated at the same disposal sites as domestic solid waste. Often considered an issue most pertinent to the wealthiest countries, the problem of toxic waste is also rapidly extending beyond the countries of the global North; according to current trends, the amount of toxic metals being generated in countries such as China, India, and Turkey could reach levels comparable with those of the UK and France within the next decade. However, many of these countries do not have adequate regulations for the disposal of toxic wastes, while the risk posed by the discharge of hazardous chemicals is being exacerbated by the transfer of certain 'dirty industries' from the North to the developing world. The brunt of these problems is borne by low income communities, many of which have been established in close proximity to damaging waste disposal sites.

Urban Energy Consumption

The way in which urban communities consume energy can have serious impacts not only on their own imme diate environment but on that of the surrounding region. The most obvious way this occurs is through the gen eration of air pollution. The rapid growth of coal fueled industry and of motor vehicles in cities of the developing world has obvious environmental consequences. In India, between 1990 and 2000, for example, the number of motor vehicles more than doubled. Indeed, for most of the cities examined in a study conducted by the Urban Management Programme, the motorized fleet is growing faster than the population and has significantly out stripped the capacity of the city's roads. The vehicles used are also often old and emit much higher level of pollutants than the newer models usually used in in dustrialized countries. The same is often true of the equipment used by industry, particularly in the case of smaller unregulated firms. The severity of atmospheric pollution is such that 1.5 billion urban dwellers world wide are exposed to levels of ambient air pollution above their recommended maximum levels. The main health threatening pollutants produced by industry, power sta tions, vehicles, and other activities in cities are sulfur dioxide, carbon monoxide, suspended particulates, hydrocarbons, nitrogen oxides, ozone, volatile organic compounds, and airborne lead. To take one extreme example, the Chinese city of Guangzhou has a remark able seventeen times the amount of sulfur dioxide in London. Dependence on low quality coal for industry is a particularly acute problem in the cities of China and Eastern Europe.

Another major problem relating to energy con sumption in developing world cities is that of indoor pollution. Often considered more of a rural problem, this is increasingly a huge hazard in cities – cottage industries and those working in the informal economy are par ticularly vulnerable, but even domestic energy use poses severe dangers. In Accra, for example, the high reliance on burning biomass fuels such as charcoal and wood for basic household activities can result in serious levels of indoor pollution. Here women and children are the most vulnerable; indeed, women and young children face far greater exposure to pollutants from indoor than outdoor air. It is estimated that half a million annual deaths worldwide can be attributed to particulate matter and sulfur dioxide alone, and while it is hard to distinguish the effects of indoor from outdoor pollution, in 1992, the World Bank identified indoor air pollution as one of the four most critical global environmental problems.

The effects of urban energy consumption do not end with air pollution; it can also have a huge impact on surrounding water and land. Where water pollution is severe, the impact on agriculture and fisheries can be enormous – average fish yields in polluted zones of one river in India are around a sixth of those in unpolluted zones. Air pollution from urban energy creation can also result in acid rain, which can devastate agricultural practices. Moreover, in Jakarta, the use of wood as source of urban energy has been an important cause of peri urban deforestation. In this respect, the environmental impacts of energy consumption in cities both on the cities themselves and their surrounding regions can be very wide ranging.

Natural Features and Housing

All of the urban environmental problems discussed above are affected by natural factors such as climate. Extreme temperatures can greatly enhance the effects of a par ticular environmental hazard. For example, a hot climate is likely to foster particular disease vectors, and air pol lution problems are likely to rise with temperature ex tremes; in cold areas, more coal is burned for heating, and in hot parts of the world contributions to ozone for mation increase due to the release of hydrocarbons and nitrogen oxides. Other aspects of geography and topo graphy can also determine the nature of environmental hazards; Mexico City, for example, is especially plagued by air pollution not just because of the number of cars per capita, but because it is bounded by mountains that prevent the dispersal of pollutants.

The location and quality of housing is a further source of urban environmental problems. Poorly constructed shelter can exacerbate environmental problems such as indoor air pollution, while problems of overcrowding can contribute to the spreading of disease. The poor general health of many low income groups and high levels of malnutrition combine with cramped living conditions to increase susceptibility to highly contagious diseases such as tuberculosis, influenza, and meningitis. Moreover, in formal settlements are often established on marginal land due to issues of affordability and the low probability of eviction. However, this also puts low income com munities at risk. Rio de Janeiro provides a trenchant example; the majority of its favelas are built on the steep hillsides surrounding the city, and many are destroyed by landslides and have to be rebuilt. Often environmental safety is traded for security of tenure; indeed, poor groups are usually very aware of the dangers, and choose these dangerous sites to live in because their low mon etary cost and proximity to work and urban livelihoods meet their most pressing needs.

Variables Shaping the Brown Agenda

There are a huge number of variables that determine the particular combination of environmental problems in any given city, but one of the most significant is its wealth. While richer cities and better off residents are more concerned with hazardous wastes, ambient air pollution, and lack of green space, the most pressing environmental concerns in poor cities and among low income groups tend to be those associated with basic sanitation, water supply, and indoor pollution. Population size and the rate of urbanization can also influence the nature and severity of urban environmental problems, with faster growing cities being less able to deal with service delivery issues such as managing solid waste, sanitation, and water sup ply. The size and economic significance of the informal economy is an important variable, although it can have both beneficial and detrimental impacts. For example, in India and Pakistan, the volume of waste requiring in cineration or disposal in landfill sites is reduced as a result of waste recovery and recycling activities in the informal economy. Conversely, difficulties in regulating the activities of small scale informal sector and cottage industries and their tendency to use cheap low quality fuel means that the informal economies of many cities pose hazardous waste problems. Overall, cities with high levels of informal or illegal industry tend to suffer add itional environmental damage.

An additional variable is the collective action of local actors who may be organized either informally or through statutory channels and who can mitigate en vironmental impacts through community organizations. There is a tendency for such organizations to proliferate among the better off with NIMBY (not in my backyard) objectives. However, the urban poor can be highly or ganized and have achieved impressive results through federations of the urban poor.

Tensions and Complementarities between the Green and Brown Agendas

Tensions clearly do exist between the brown agenda and the wider green movement. Indeed, it was partly due to a perceived neglect of local urban environmental issues by green organizations that urban development authors and practitioners started using the term brown agenda. However, solutions to brown agenda problems can be problematic from a green perspective – it is sometimes the case that when a city begins to cope more effectively with brown issues, the knock on effects can be damaging to the wider regional or global environment. For example, if a city manages to create a functional waterborne sewage system, thereby massively reducing the exposure of its residents to fecal matter and biological pathogens, sewage might not be treated properly before being discharged into bodies of water in the wider region. Even when this is not the case, waterborne sewerage escalates water use demands dramatically, increasing a city's ecological footprint. In other words, solving a brown problem is not necessarily a step in the direction of solving a green one and rigidly separating the two agendas or dealing with the issues of one without paying adequate attention to the other, can be counterproductive.

Recently, the tension has been framed as one between 'green' environmental concerns on the one hand, and 'brown' issues emerging from economic development pressures on the other. In some ways this is useful, serving as a reminder of the fact that brown issues are in fact fundamentally development issues that would previously have been referred to in the language of growth, poverty reduction, and addressing basic needs. In this context, the common green assertion that economic growth is over valued and environmentally detrimental is not particu larly helpful – there is no denying that many brown problems are perpetuated not by rapid economic growth but by chronic underinvestment and persistent poverty in a city. In many ways, brown agenda issues reflect the hazards of underdevelopment, while green issues reflect the opposite, a level of development bloated by high level mass consumption. However, to simplistically interpret brown issues in relation to development and green issues to environment is to undermine some of the positive achievements of creating a brown agenda in the first place and of linking development and the environment. The brown agenda serves to highlight the fact that the most pressing environmental problems facing people across the world are not global but local in their impact, especially as they affect poor people living and working in cities.

It is, in part, in recognition of the interrelation of the two agendas that the idea of sustainable development has become so significant. The sustainable development approach is in many respects an attempt to reconcile the two agendas, aimed at ensuring that meeting urban demands for resources does not cause irreversible harm to ecosystems on both the local and global level. In other words, and as laid out by the Brundtland Commission in 1987, the goal of sustainable development is to meet the needs of the present without compromising future gen erations' ability to meet their own needs. The challenge for those concerned with developing countries is that the 'sustainability' element of the contemporary environ mental discourse does not eclipse a focus on poverty re duction and basic needs. Here, it is interesting to observe a few of the ways in which the current high profile of certain green problems can interact with the brown agenda.

The 2006 Stern Review Report on the economics of climate change provides a few interesting examples of how a now highly salient global green issue is also of relevance to the brown agenda. First, it notes the ways in which global climate change could exacerbate brown problems, for example, by having a disastrous impact on water availability in many regions, with areas that are already dry likely to experience further decreases in water availability. This is, according to the Stern Report, likely to change the water status of billions of people with the result that considerably more will need to done on top of current efforts to meet people's demand for water. Furthermore, in other parts of the world, melting glaciers and mountain snow, as well as rising sea levels, will in crease flood risk, and floods very often conspire with brown issues of poor sanitation, drainage, and solid waste management to spread disease rapidly among dense urban populations. Similarly, the effect of heat waves in cities will be very dangerous, leading not only to extreme temperatures but to more dangerous air pollution inci dents and the spread of vector borne diseases; the report highlights the likelihood of a particularly severe impact on urban slums in this regard.

Also evident in the report, is the extent to which brown issues in the developing world are contributing to climate change itself, despite the relatively low levels of indus trialization and comparatively high emissions of the in dustrialized North. The International Energy Agency (IEA) has estimated that there are currently 2.5 billion people using traditional biomass for cooking and heating, while in developing economies there are strong pressures to expand energy supplies as quickly as possible. This has been especially apparent in China, where power com panies have rapidly been investing in new coal fired power stations, and where 16 out of the 20 most polluted cities in the world are located. On the positive side, China has established a goal to reduce energy intensity between 2006 and 2010; the Indian Ministry of Power is working to remove market barriers to low carbon home cooking sys tems and Mexico City has removed carbon intensive oil plants and replaced them with high efficiency gas turbines.

What is noticeable about many of these efforts is that, while they have an impact on global climate change, so lutions are often left to municipalities to initiate rather than national authorities. Municipalities in turn are mo tivated as much if not more by local and brown concerns as by the global green agenda. As the Stern Report reminds us, the fact that many developing countries are already making efforts to reduce their greenhouse gas emissions in the long term is often a result of their awareness of local co benefits. The basic needs of the urban poor and even the urban environment are far from a central focus of the Stern Report. The term 'brown agenda' does not even appear once in its 600 odd pages. However, the analysis demonstrates how futile it is to try and separate brown and green agenda issues, whether in terms of causes of en vironmental problems at all levels, their local impact, and in relation to the means of addressing them.

Conclusion: Dichotomy or Continuum?

The brown agenda elucidates urban environmental problems that are often sidelined by the green movement and this is welcome. Yet, it is clear that, in many cases, they positively interact and it is misleading to separate the two. The notion of a city's ecological footprint is important, whereby people outside cities sometimes suffer even more from the impact of the activities of a city than city dwellers themselves. The International Institute for Environment and Development (IIED) consider urban environmental burdens to arise at three main scales: those that have an immediate impact around the home and workplace, those that have wider impacts in and around urban centers, and those that are global but largely caused by urban activities. In this respect, the brown and green agendas do not give rise to a clear dichotomy but form part of a continuum. Acknowledging this, the way the brown agenda codifies a particular set of issues, nevertheless, serves one valuable purpose, that is, to clearly point up and demarcate issues that affect cities in the poorest parts of the world – inadequate water supply and poor sanitation – whether or not they have much effect on anyone else or the global green agenda. It is important to see and tackle environmental and de velopment problems holistically; but if the two agendas are too closely merged, the issues that will dominate will be those of obvious global concern, and those that threaten the more powerful industrialized North.

In sum, the brown agenda might best be thought of as an expression of the most basic and essential environ mental problems. As such, it should perhaps underwrite all efforts towards environmental improvement on any scale; the notion of a brown agenda exists to vocalize those elements of a broader environmental agenda that in an increasingly urbanized world, affect a disproportion ate number of people daily. As it is, the brown agenda is often spoken of only in a whisper. The challenge remains to turn up the volume and to make the transition from 'green' politics that sometimes favor the special interests of the better off to a genuine form of environmental justice that gives voice to the urban poor.

See also: Climate Change; Development I; Environment; Participation; Poverty; Squatter Settlements; Sustainable Development; Urbanization.

Further Reading

Atkinson, A., Davila, J. D., Fernandes, E. and Mattingly, M. (eds.) (1999). *The Challenge of Environmental Management in Urban Areas*. Aldershot: Ashgate.

Bartone, C., Berstein, J., Leitmann, J. and Eigen, J. (1994). *Toward Environmental Strategies for Cities*, Policy paper 18: *Urban management programme*. Washington DC: The World Bank.

Beall, J. (2006). Dealing with dirt and the disorder of development: Managing rubbish in urban Pakistan. *Oxford Development Studies* 34(1), 81 97.

Beall, J., Crankshaw, O. and Parnell, S. (2000). Victims, villains, and fixers: The urban environment and Johannesburg's poor. *Journal of Southern African Studies* 26(4), 803 855.

Brundtland Commission (1987). *Our Common Future*. Oxford: Oxford University Press.
Davis, M. (2006). *Planet of slums*. London: Verso.
D'Cruz, C. and Satterthwaite, D. (2005). *Building Homes, Changing Official Approaches: The Work of Urban Poor Organizations and Their Federations and Their Contributions to Meeting the Millennium Development Goals in Urban Areas*. London: IIED Human Settlements Programme.
Fazal, S. (2006). Addressing congestion and transport related air pollution in Saharanpur, India. *Environment and Urbanization* 18(1).
Hardoy, J. E., Mitlin, D. and Satterthwaite, D. (2001). *Environmental Problems in an Urbanizing World*. London: Earthscan.
Hardoy, J. E. and Satterthwaite, D. (1989). *Squatter Citizen*. London: Earthscan.
Gandy, M. (2006). Planning, anti planning and the infrastructure crisis facing metropolitan Lagos. *Urban Studies* 43(2).
Leitmann, J. (1994). *Rapid Urban Environmental Assessment: Lessons from the Cities In The Developing World*. Policy paper 14: *Urban Management Programme*. Washington DC: The World Bank.
UNDP (1991). *Cities, People And Poverty: Urban development Cooperation For The 1990S. A Strategy Paper*. New York: United Nations Development Programme.
UNESCO (2003). *Water for People, Water For Life. UNESCO World Water Assessment Programme*. Paris: UNESCO/Berghahn Books.
UN HABITAT. (2003). *Water and Sanitation In The World'S Cities*. UN HABITAT. London: Earthscan.

Relevant Websites

http://www.earthinstitute.columbia.edu
Earth Institute, Center for Sustainable Urban Development (2006).
http://www.iied.org
International Institute for Environment and Development (IIED) (2005).
http://www.oecd.org
OECD (2002) A DAC Reference manual on Urban Environmental Policy.
http://www.rec.org
Regional Environmental Center (REC) What is a sustainable city?
http://www.rrojasdatabank.org
Rojas, R. (1996) *World Resources 1996 97*. A joint publication by The World Resource Institute, The United Nations Environment Programme, The United Nations Development Programme, and the World Bank.
http://www.hm treasury.gov.uk
Stern Review Report on the Economics of Climate Change (2006).
http://www.unhabitat.org
UN HABITAT.
http://www.wsscc.org
Water Supply and Sanitation Collaborative Council (WSSCC).

Buffer Zone

G. Prevelakis, University of Pantheon-Sorbonne (Paris 1), Paris, France

Glossary

Chaordic An order that appears chaotic (chaos + order) because it is not founded on hierarchical and territorial structures but functions through networks, cultural codes, and decentralized initiatives.

Demilitarized Zone An area between two territories in conflict functioning as a temporary physical separation with the perspective of a permanent territorial settlement.

Geopolitik School A group of geographers, mostly German, active between the two world wars under the leadership of Karl Haushofer, who challenged the Versailles territorial settlements.

Iconography The set of symbols that unite a people and link it to a portion of space transforming it thus to a territory.

Marches Territories organized to protect the early medieval Europe from the Moors, Celts, Scandinavians, and Slavs.

Mitteleuropa The area between the Atlantic, the Mediterranean, and the Russian parts of Europe, characterized by dominant German economic and cultural presence.

Reticular In geography, referring to a network of territories, contiguous or not.

Shatter Belt An area where struggle between outside powers takes place openly and violently, with limited risk of generalization of conflict.

Introduction

In its simplest definition, a buffer zone is an area between two or more adversarial powers, whose function is to physically separate them. It is thus conceived as an instrument of peace and stability. The word buffer conveys the idea of absorption of tensions, of cushioning. A buffer zone can exist at various geographical scales, in accordance with the scale of the powers it is separating. If, for example, the buffer zone serves to separate two ethnic communities, it may have a depth of a few kilometers or less; if it separates two major continental or maritime powers, it can cover many state territories.

The term buffer zone made its appearance at the end of the nineteenth century in the context of the competition between the British and the Russian Empire in the north of the Indian peninsula. During the twentieth century, this term found extensive use. The '*Geopolitik*' school included the term of *Pufferzone* to its often used vocabulary. In the French literature, the equivalent *zone tampon* was somewhat less frequently used.

The success of the term in the political geography of the first part of the twentieth century is related to the strong influence of the environmental determinism tradition. Since it was thought that physical space conditioned human affairs, the use of physical separation seemed the most appropriate means to preempt conflict situations. What this approach underestimated was the role of human action in the creation and the survival of a buffer zone. Buffer zones are as much the causes as the results of human affairs. Their study involves the whole spectrum of the relationship between space and politics.

The term buffer zone found various uses in scientific and nonscientific literature. It has reached outside political geography to the fields of physical planning and environmental protection to designate any zone separating land uses that may be considered as incompatible or dangerous for certain species. In political geography, the limits of the term are ill defined. It is often used in the place or in combination with other terms like frontier, march, shatter belt, demilitarized zone, glacis, *cordon sanitaire*, etc. It would be vain to introduce precise distinctions to differentiate these terms. The realities they describe shade into each other; it is therefore more useful to examine them together. They all express geographical phenomena of spatial transition from one territory to another. At the same time, they are themselves situated in an intermediary position, between the concept of the boundary, a linear division, and that of the territory. Thus, for example, the Cyprus buffer zone becomes a boundary if examined at the scale of Greek and Turkish spaces. On the other hand, Belgium, during a long period a buffer zone between France and Germany, was and is a territory by itself.

Buffer zones constitute therefore a gray area of political geography. They are difficult to grasp because they are inherently unstable, changing, challenging the territorial status quo. In the established political geography, they are at its margins. In a renewed political geography, less centered on hierarchical center–periphery approaches and focusing more on issues of in between, the phenomena they encompass will probably occupy a more central position.

Buffer Zones as Peripheries of Territories

Ideally, buffer zones are neutral spaces between two or more territories. However, in many cases, buffer zones were organized by one power in order to cope with challenges or in order to take advantage of opportunities coming or existing in spaces dominated by less organized powers. In such situations, buffer zones are in fact peripheries of the territory of the more structured power.

At a primitive stage, spatial separation between different human communities was a gift of nature. Uninhabited spaces in the form of deserts, marchlands, forests, or seas created sufficient obstruction to circulation from one territory to another to make human made obstacles unnecessary. At the other extreme, the densely and continuously inhabited post Renaissance Europe had to devise a sophisticated territorial system defining with absolute precision the transition from one area of sovereignty to another. Between the two extremes, range a series of separating systems that organized the transition from one power field to another according to gradients rather than in the form of an abrupt spatial discontinuity.

Premodern Buffer Zones

Those premodern buffer zones resulted from asymmetrical power situations. They were organized by relatively well structured societies facing the challenges of outside infiltration or invasion that threatened to destabilize them. The most known examples of organized zones of transition from the 'civilized' to the 'barbarian' realm are the Roman 'limes' and the Chinese Great Wall systems. Although both terms convey the image of a linear separation, they were in fact zones, with sufficient depth to serve as protecting buffers. Those imperial protections were part of the imperial territory: they constituted its periphery. However, they also had a life of their own. Their inhabitants, in contact with the 'barbarians', were influenced by their culture. They developed their own leadership which could sometimes challenge the imperial center. These buffer zones had therefore many characteristics of a territory and often evolved to states. During the Middle Ages, buffer zones between what remained of the Roman Empire and the pressure coming from outside, from the sea, or from the steppe, took the form of special areas named marches in the West and 'akra' (singular: 'akron') in the East (Byzantium). Many States of today carry in their names the memory of this terminology, like Denmark (the Danish march) or Ukraine (akron). Some of them continued to function as buffers in modern times, but in the form of buffer states.

Modern Forms

The stabilization of the West European political map after the Renaissance excluded the practice of peripheral buffer zones, however, a new field was opened for them in America. This time buffer zones were not organized for defensive purposes, but, on the contrary, as a means of expansion into new lands. Those new peripheral territories became known as the American frontier. With the European marches/akra of the Middle Ages, they had in common their specific culture that differentiated their inhabitants from those of the core. The American frontier served as a buffer zone between the established and law abiding American society and the lawlessness in the Western lands it was taking away from the indigenous populations. It was thus instrumental in the legitimation of what was in fact a colonial situation.

Another form of peripheral buffer zone is represented by the glacis. The term comes from the French verb *glacer*, which means to slip. The glacis was the outside part of medieval fortifications. A strong slope protected the fortifications from artillery and exposed the assailants to the defensive fire. Following this analogy, a glacis is therefore a buffer zone surrounding a territory that it helps to defend. This metaphor has been used by Stalin to legitimate the Soviet domination of Eastern Europe. This East European Soviet glacis was presented as necessary in order to protect the Soviet Union from capitalist infiltration. A parallel can be found in the US containment strategy, that is, the construction of a chain of alliances with states surrounding the Soviet Union in order to contain the expansion of communism. In the same family should be placed the French idea of *cordon sanitaire*: a series of anticommunist buffer states surrounding the Soviet Union to protect Europe from the communist ideological epidemic after the Soviet Revolution.

The Israeli Security Zone in Southern Lebanon, set up in 1982 against Palestinian infiltration and missile fire, is an example of a modern buffer zone in the form of a territorial periphery in an asymmetrical situation that reminds those of the Middle Ages.

Buffer States

A buffer state is a relatively weak state lying between two or more antagonistic powers of comparable potential.

Conditions for Buffer State Status

The geographical condition seems paramount. The buffer state must be in contact with the territory of the other powers and must separate them. However, this separation may not be complete. Thus, for example, the fact that France and Germany had a common boundary did not cancel the role of Belgium as a buffer state before 1914 and between the two world wars.

Buffer states do not have as unique function to constitute a cushion between rival states. Very often they include in their territories essential strategic resources, especially in the form of major circulation corridors and crossroads. In such a case, their role is to neutralize them by depriving both rival states from their military use. It is obvious that for this role it suffices to cover only the portion of the in between space in which the important crossroads or axes are situated.

In addition to contiguity, conditions of foreign policy orientation must be met. The powers surrounding the buffer state are usually two, however, in rare cases they can be more, like in the case of Poland before its partition at the end of the eighteenth century (Austria, Prussia, Russia). Those surrounding states must be in rivalry. If antagonism ceases, the buffer state loses its *raison d'etre*. However, if the rivalry becomes too acute, it can also lead to the disappearance of the buffer status. At the beginning of the Cold War, the tension between the US and the Soviet Union was such that many preexisting or potential buffer states were exposed to efforts for unilateral control, that often ended in solutions in the form of partition. Stalin managed to take control of all of Eastern Europe, a possible buffer zone. Korea and then Vietnam have been divided into two and became areas of East–West confrontation. Before them, Germany, also a potential buffer state, shared the same fate. Thus, the Cold War, by diminishing the antagonisms inside each alliance through the discipline imposed by Washington and Moscow, by accentuating ideological rivalry around the international frontier dividing the two worlds, and also because of the balance of nuclear terror that made territorial arrangements more or less irrelevant, reduced the role of buffer states. Belgium and Korea, two historical buffer states ceased thus to function as such. Later, the Vietnam conflict led to the disappearance of two more buffer states: Laos and Cambodia.

Therefore, the foreign policy orientation condition cannot be reduced to the rivalry of the surrounding states. A certain balance in the intensity of the antagonism is necessary. The paradigmatic case of Afghanistan perfectly illustrates this situation. In fact, it is only in 1907, when Russia and England felt that they needed to water down their rivalry in Asia in order to cope with the new geopolitical challenges arising in Central Europe, that a compromise was found which consolidated the buffer state role of Afghanistan. Before that, and in spite of the use of the term buffer state for Afghanistan already in 1883, both powers were trying to monopolize the whole area.

The third condition for buffer state status has to do with capability distribution. The buffer state must be considerably weaker than the surrounding states. Thus, for example, Prussia in post Napoleonic Europe could hardly be considered as a buffer state between Russia and Austria, in spite of the fact that its location satisfied the geographical condition and that the two other powers did develop rivalries.

If the buffer state must be weak, the stronger states must have a comparable potential. The buffer system power field thus defined creates the possibility for buffer state neutrality through a series of strategic calculations from the part of the major actors, the surrounding antagonistic powers. According to this reasoning, the power differential between each of the surrounding powers and the buffer state creates a favorable situation for the absorption of the buffer state in the zone of influence of either, as a result of diplomatic pressure or invasion. However, any effort in that direction would provoke the intervention of the other strong state for which such an evolution would constitute a threat. The alliance of the second strong state with the smaller, buffer state, would tip the balance of power again and make the operation of the takeover risky or even condemned to failure. This triangular situation puts the power challenging the buffer state at a disadvantage, creating thus the necessary dissuasion and securing the buffer state's independence.

The Stability of the Buffer State

However, this situation is inherently instable, since the balance between the surrounding powers can fluctuate, creating windows of opportunity for the one or the other powers to intervene and abolish the state's sovereignty. During the late 1970s, when the US influence in Asia was at its lowest after the Vietnam debacle and the Iranian Revolution, the Soviet Union has had the opportunity to turn Afghanistan from a buffer state to a Soviet satellite.

Ideally for a buffer state, the antagonism of the surrounding states would vanish and therefore the danger of becoming an arena of confrontation disappear. The European construction has thus saved the West European buffer states that had suffered during the previous centuries from the French–English and then the French–German rivalries. However, when this optimal situation is unattainable, the preservation of a buffer state status is preferable to its collapse which can lead to wars, civil wars, partition, foreign domination, etc. Thus, in addition to the role of the international environment, the capacity of the buffer state to preserve its status by mobilizing its own resources is also to be taken in consideration.

The geographical factor comes again in the foreground. Although, by definition, a buffer state cannot be as strong as the surrounding states, the more it is capable of resisting challenges to its sovereignty, the more it can contribute in stabilizing the power equation. Thus when one of the surrounding states weakens, a relatively strong buffer state might be able to compensate the disequilibrium by mobilizing its own forces. This capacity is related to the geographical conditions often encountered in buffer states territories. Many buffer states developed in difficult natural environments, very often in mountainous areas. Those areas were of little interest for human establishment and therefore remained for long periods on the margins of state formation. With the gradual expansion of the surrounding states however, their peripheries met, creating buffer conditions at their intersection. Such areas are very often crossroads of important axes of circulation and, because of their intermediary geographical location, are inhabited by mixed populations whose cultural geography expresses the transition from one realm to another.

It is difficult to say if and when those conditions favor resistance to outside intervention. The mountainous character offers undoubtedly a critical advantage to defense. However, the cultural mix and the crossroads function can constitute elements of weakness by facilitating external influences and by promoting centrifugal forces. The difficulties of Lebanon to survive as a buffer state between Syria and Israel, largely because of the internal divisions of its population, illustrate the geographical handicap of many buffer state territories. On the other hand, in spite of cultural and geographical fragmentation, the Afghans united again and again against external challenges, making thus efficient use of the defensive character of their terrain. Thus, geographical conditions that can be encountered in many buffer state situations do not determine in a predefined way the capacity of their population to defend their state.

Leadership and diplomatic experience are also important factors. The city state of Dubrovnik/Ragusa not only survived the tensions between the Ottoman Empire and its Latin enemies, but also founded its prosperity and its culture on its role as a hinge between the two Mediterranean realms because of the extraordinary diplomatic skills of its leaders. However, at certain historical moments, the geopolitical stresses are so intense that no diplomatic talent can avert disaster. Sihanouk, the charismatic Cambodian leader, had realized that many years before the US and the South Vietnamese troops invaded his country and put an end to a history of buffer state status going back whole centuries.

Buffer states are spaces in between, that have thus to cope with the instability of their geopolitical environment in order to survive as sovereign states. Sometimes their buffer state status is determined from the outside, as a result of a power field that favors neutralization. Those situations, however, last as long as the conditions that created them. On the other hand, there are cases where a combination of geographical conditions and human factors lead to the emergence of buffer state iconographies that can stabilize the influence of the external power fields and thus extend the life span of the buffer state. The extraordinary stability of Switzerland as a European buffer state is based on a strong Swiss iconography that has been able to overcome the geographical and cultural diversity of this country and to create a determination to defend it that has gained the respect of all European powers.

As situations of geographical in between, buffer states can be the 'loci' of extraordinary material and cultural wealth, but at the same time of terrible suffering and destruction. The fate of Lebanon, an Eastern Mediterranean success story that turned into a nightmare, is the perfect illustration of the contradictory nature of many buffer states.

Buffer Zones and the International Frontier

Global and Large-Scale Outlooks

At a high level of geopolitical generalization, buffer zones appear as separating large areas unified by religion, empire, ideology, or culture. With the rise of Islam, the Mediterranean Sea became a buffer zone between Christian Europe and Muslim North Africa. Charlemagne's empire, divided in three parts at Verdun in AD 843, led to a historical buffer zone between what would become France and Germany. Many of the European buffer states between the two countries originated from Lotharingia, the Third Kingdom between the Eastern (future French) and the Western (future Germanic) Frankish realms. The Cold War separated the world on the basis of ideology, creating a complex buffer zone out of the Soviet glacis, the American containment zone, and the shatter belt. With the end of the Cold War, Samuel Huntington's 'Clash of Civilizations' hypothesis introduced a new division between large civilizational areas, separated by persistent historical frontiers, like the one between Western and Eastern Christianity (the 1054 Schism line).

On such a scale, buffer zones are usually perceived in negative light, as areas of conflict and political fragmentation, marginal in comparison to the civilization of the power centers. This vision has been recently theorized by Violette Rey. Her concept of 'in between' (*entre les deux*), inspired by her study of the area that Jean Gottmann named "the tidal lands of Europe" (Central Europe), describes the state of an area submitted to overwhelming, conflicting, and constantly changing outside influences. The resulting historical discontinuities constitute an obstacle to accumulation. A feeling of futility demoralizes elites and reinforces opportunistic attitudes. However, the argument is sometimes returned upside down. The nineteenth century German concept of *Mitteleuropa* promoted the idea of centrality of the German space, being in the middle of Europe, between East (Russia) and West (France, England). The partition of Germany after World War II transformed this potential European center into a borderline, part of the Iron Curtain. A parallel can be drawn with Samuel Huntington's frontier between East and West in Europe in comparison to Dimitri Kitsikis' 'Intermediary space'.

Kitsikis considers the area around the Aegean (today's Greece and Turkey, the former core of the Byzantine and the Ottoman empires) as the center of a large historical buffer zone between the two major civilization areas of Eurasia, the West (Europe), and the East (India, China, Southeast Asia, Japan). If the Schism linear icon represents Huntington's conflict hypothesis, Kitsikis' intermediary buffer zone indicates the possibility of dialog between civilizations.

The above examples illustrate the use of the buffer zone theme to convey geopolitical representations that can influence politics.

Demilitarized Zones

In many cases of violent conflict, various forms of sanctuaries are set up in order to offer protection to noncombatants, diplomats, hospitals, etc. Also secure zones are organized as meeting places for negotiation or exchange of prisoners. Finally, when fighting ceases, bands of land can be neutralized from a military point of view in order to separate the population and avoid the spontaneous start of hostilities. Such agreements are negotiated through the mediation of the United Nations or diplomats of the major powers. The control of their respect often falls under the responsibility of the United Nations. Thus, since their creation, the United Nations have organized an important number of demilitarized zones, often called buffer zones. Among the most important and long lasting are: the buffer zone separating North and South Korea since 1953 and the Cypriot demilitarized zone established in 1974 between the territory under Turkish control and the rest of the island. Many other demilitarized zones have been set up during the various stages of the Israeli–Arab struggle, as well as in the Balkans, in Africa, and elsewhere.

Such zones are part of a larger system that may be termed as the international frontier. This frontier is made up of a variety of international arrangements, past or present, including international protectorates, buffer states, neutralized zones, areas under international administration, international canals and waterways, international mandates and trusteeship areas, etc. The international frontier, managed more or less successfully by the international community, has been very active after World Wars I and II. At the peak of the Cold War, it has been superseded by the frontier dividing the two blocks. The bipolar system left few areas of the world outside either the rigid stability inside the blocks or the violent clashes in areas of confrontation, the shatter belts. The eastern Mediterranean and the Middle East are among the exceptions to this double rule and this is the reason for which the international community was especially active in those areas.

New Issues

The international frontier carries memories of past tensions that are related to the persistence of geographical factors or to the inertia of cultural geography. With the end of the Cold War, many parts of this international frontier became destabilized again. The return of war to Europe has led to ethnic cleansing, diminishing again the potential of many areas to function as buffer zones between different national groups. The cases of Croatia, Bosnia, and Kosovo clearly indicate the trend to reinforce the separation of nationalities and to diminish the intermediary function of residual mixed communities that had somehow survived the homogenizing processes in the Balkans. At the same time, demilitarized zones have been multiplied in order to cope with renewed conflict.

On the other hand, the new forms of terrorism do not express territorial conflicts. The Islamist challenges to American hegemony are not organized in the framework of a continuous territory with a hierarchical structure of organization behind it. Rather they function in a chaordic way, based on reticular spaces that penetrate deeply inside the opponent's territory. Does this shift to conflicts that make territory less relevant also neutralize the utility of buffers? Should we conclude that the interest of the buffer zone is or will become only historical? As it is difficult to hope that geographical inequalities – in the form of differences in demographic growth, development, stability, or even, in the urban sphere, of spatial segregation – will vanish, buffer zones will always exist, even if they have to take new forms.

In our globalized world, the new reticular political geography will continue to focus on territory, but at a larger variety of scales than in the past. It will be interested in combinations of territories forming networks that will be defined less and less by the logic of topographical distance. The cultural factor will become more and more important. The predominance of the territorial nation state, both as a reality and as a geopolitical representation, will be challenged and, with it, the geopolitical forms and phenomena that have accompanied its rise. In this sense, buffer states may in fact become less relevant.

However, buffer zones separating the wealthy inhabitants of Western metropolises from the immigrant communities may unfortunately emerge as new subjects of study. In cities where conflicting 'civilizational' groups coexist, buffer zones may also be needed to absorb the clashes. On a global scale, airports and airplanes are also turning into a reticular buffer zone. Like the maritime realm, they become spaces with their own rules that function as cushions between the societies of the developed world and the universe of the illegal immigration.

The geographical inequalities between Europe and Africa have led to various Euro Mediterranean projects, the latest of which is to create a Mediterranean Union on the model of the European Union. Those projects cor respond to a new form of buffer zone, a new European march to face the challenges of African poverty, under development, and instability, to regulate a relationship that might become explosive. North America facing the challenges of South America is also experimenting with various forms of buffers or marches.

It will be necessary to invent the reticular equivalents of the buffer zone at continental and at global scales in order to respond to the buildup of tensions related to the rapid unification of the world. It is essential, however, that these real or virtual buffer zones will not only have the role of partitioning the new geographical space but also of functioning as meeting points, as crossroads of exchange and dialog. Diasporas already play this role. Between the developed world and countries like China or India, they constitute in between reticular spaces that share cultural characteristics of both worlds. Their role has been in strumental in regulating population flows, encouraging the exchange of products, capital, and innovation, and promoting dialog and understanding. However, diasporas can also become unstable spaces, just as the traditional buffer zones. Modern terrorism and organized global crime are the ugly faces of those modern buffers.

See also: Balkans; Borderlands; East/West; Political Boundaries; Sovereignty; Territory and Territoriality.

Further Reading

Bailey, S. D. (1980). Non military areas in UN practice. *The American Journal of International Law* 74, 499 524.
Chay, J. and Ross, T. E. (eds.) (1986). *Buffer states in world politics*. Boulder: Westview Press.
Gottmann, J. (1980). Les frontières et les marches: cloisonnement et dynamique du monde. In Kishimoto, H. (ed.) *Geography and its frontiers. In memory of Hans Boesch*, pp 53 58. Bern: Kummerly and Frey.
Hall, D. H. (1948). Zones of the international frontier. *Geographical Review* 38, 615 625.
Heffernan, M. (1998). *The meaning of Europe. Geography and geopolitics*. London: Arnold.
Huntington, S. P. (1993). The clash of civilizations? *Foreign Affairs* 72, 22 49.
Kitsikis, D. (1994). *L'Empire ottoman* (3rd edn). Paris: Presses Universitaires de France.
Maier, C. S. (2006). *Among empires*. Cambridge, MA: Harvard University Press.
Partem, M. G. (1983). The buffer system in international relations. *The Journal of Conflict Resolution* 27, 3 26.
Rey, V. (1996). Europes orientales, Europe balkanique. In In Brunet, R. & Rey, V. (eds.) *Géographie universelle: Europes orientales, Russie, Asie centrale* vol. 1, pp 8 156. Paris: Belin Reclus.
Romer, J. C. (ed.) (2004). *Face aux Barbares, marches et confins d'empires*. Paris: Tallandier.
Sinnhuber, K. A. (1954). Central Europe Mitteleuropa Europe centrale. An analysis of a geographical term. *Transactions and Papers (Institute of British Geographers)* 20, 15 39.
Taillard, C. (1989). *Le Laos, stratégies d'un etat tampon*. Montpellier: RECLUS.

Business Services

John R. Bryson, The University of Birmingham, Birmingham, UK

Glossary

Business Services This term covers a broad spectrum of services that are predominantly traded in business-to-business transactions.

Emotional Labor The process by which employers manage employees' feelings during social interaction in the workplace.

Externalization The term used to describe the process by which in-house services are transferred to external suppliers.

Goods-Related Business Services A subdivision of the category of business services that describes firms that provide inputs to clients that directly support operational activities (cleaning, catering, and facility management).

Internalization The term used to describe the activities of a client who decides to replace a service obtained from an external provider with an in-house resource.

Knowledge-Intensive Business Services (KIBS) A subdivision of the category business services that describes firms that provide information, expertise, and often strategic knowledge as intermediate inputs into private and public sector clients.

Outsourcing It describes an in-house process (service or product) that has been transferred to an external supplier usually located in the same country.

Second Global Shift It describes the development of a new international division of labor in which service activities and processes are incorporated into the international economy. The 'shift' represents the transfer of service functions and employment to what may be lower-cost locations.

Service Offshoring/Global Sourcing These terms are refinements of the concept of outsourcing. They describe services that have been outsourced to suppliers located in foreign countries. The term 'service offshoring' was developed in the USA and UK and is problematic as services can be 'offshored' to foreign suppliers located in the same continental landmass. The alternative concept of global sourcing removes the Anglo-American bias from this terminology.

SIC The Standard Industrial Classification of Economic Activities is a measure of economic activity constructed by the UK National Statistics agency to classify business establishments and other statistical units by the type of economic activities they are engaged in.

Introduction

Developed market economies have been experiencing a process of structural change over the last two decades that has involved two related alterations. First, a process of de industrialization in which manufacturing's share of employment and output has been declining and, second, compensating growth in employment and output in services and especially in business services. Business services add value to their clients by providing operational and strategic inputs. Most business service firms are orientated to national markets and this reflects the importance of small firms in this sector. Nevertheless, transnational business service firms have developed albeit these tend to be created through linking national partnerships into an international corporate framework.

Business service firms play an important role in developing new process and product innovations. They develop new products (management/organizational ideas, marketing concepts, etc.) and sell them to clients, and this can have a major impact on the local and increasingly international organization of capitalist activities. It is important not to underestimate the role business service firms play in the formation of new economic geographies. Alterations in the economic system are the result of a series of unrelated and related changes made by independent private and public sector organizations. Central to these alterations are the activities of a group of business service professionals: management gurus and management consultants. Business services played an important role in the development of Fordism through the activities of F. W. Taylor, the American management guru and consultant, and more recently in encouraging clients to introduce e commerce, outsourcing, and offshoring into their business models.

Defining Business Services

The academic literature on business services has its foundations in a ground breaking monograph published by Harry Greenfield in 1966. In this classic work, Greenfield argued that the consumer/producer dichotomy with respect to goods could also be applied to services. Consumer goods are produced to satisfy final demand while producer goods (machine tools, etc.) are intermediate inputs that contribute to the further production of output. He argued that a category of consumer services existed, as well as producer services with the latter providing intermediate inputs into the activities of private and public sector organizations. The new term 'producer service' described all services that business firms, not for profit institutions, and governments provide and sell to producers rather than consumers. The category of producer service includes all intermediate services ranging from finance, insurance, and real estate (FIRE), business services, and professional services. Business services are a subcategory of producer services as the term excludes FIRE and consumer services that can be used for final and intermediate consumption (**Figure 1**).

The term business services covers a broad spectrum of services that are predominantly traded in business to business transactions. In practice, a number of business services are used by both producers and consumers, for example, an accountant can complete tax forms for individuals or firms. The category of business services is further divided into two types. First, goods related services directly support operational activities required to produce goods and services. This subcategory includes security, catering, and facility management. Second, knowledge intensive business services (KIBS) provide information and often strategic knowledge interventions into the activities of their clients. This subcategory includes computer services, industrial design, technical services, market research, executive search, and management consultants. It also includes a further subcategory of 'professional services' that includes legal and accountancy services. In all cases, business services are identified as activities that complement, substitute, or develop the in house service functions of their clients.

Measuring Business Services in National Economies

The ongoing structural realignment being experienced by economies makes it impossible for governments to ensure that their national economic statistics are an adequate reflection of economic activity. This has always been the case. The UK Standard Industrial Classification of Economic Activity (SIC) is a measure of economic activity, but essentially it is a backward looking measure; the SIC cannot be constantly amended to take into consideration ongoing developments in the division of labor. New functions are created and firms are

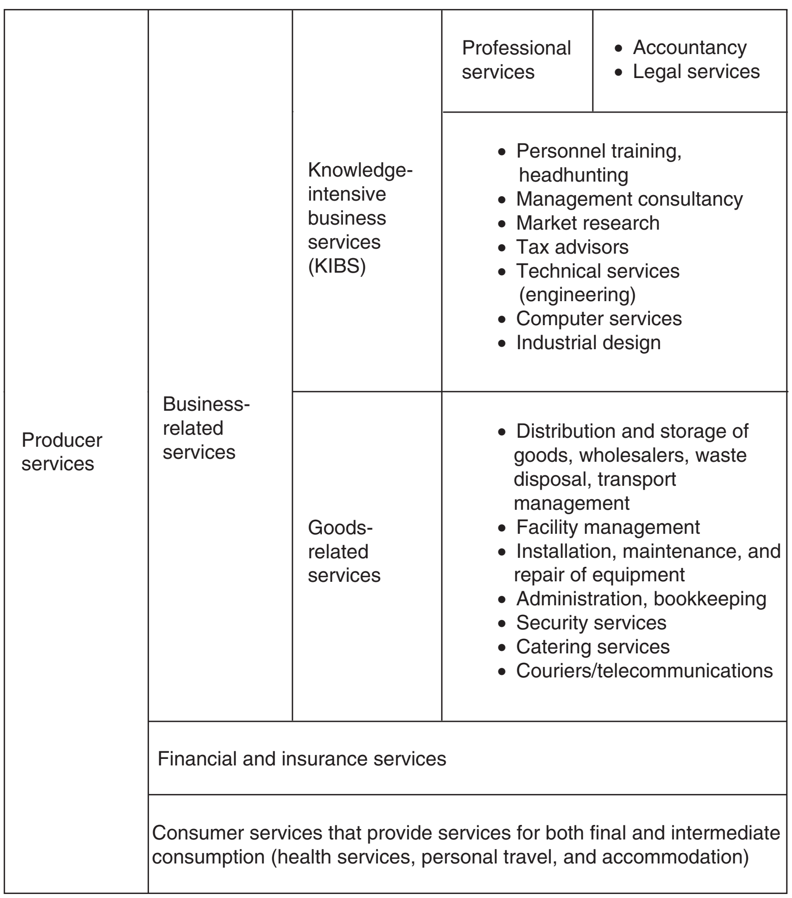

Figure 1 Defining producer, business, and professional services.

established that deliver new types of products and services that do not fit with the existing SIC. It takes some time for the SIC and academics to take into consideration these new types of economic activity and to alter the existing classification system. The United Kingdom's SIC has a long history of periodic change as it attempts to mirror the current structure of the economy; it was first introduced in 1948 and the classification was revised in 1958, 1968, 1980, 1992, and 1997.

During the late 1980s it became apparent that something rather interesting was happening in both the British and American economies and this was the rapid and unexpected growth of a heterogeneous group of activities that came under the label business services. The SIC was less than helpful in monitoring this development as it was designed to measure an economy centered on manufacturing rather than services. The SIC has always provided an excellent tool for measuring manufacturing industries but has until relatively recently failed to capture the diversity of the service side of the economy. An indication of this problem is that initially geographers working on the growth of business services had to rely on SIC 8395 (other business services not elsewhere specified) as the primary measure for business services. The number of employees in this category doubled in only 6 years, 1981–87 (up by 160 000 workers, to 316 000 in 1987). At this time, this SIC code was problematic as it included heterogeneous assortment of activities including management consultants, market research and public relations consultants, document copying, duplicating and tabulating services, and other services 'primarily engaged in providing services to other enterprises' such as employment agencies, security services, debt collection, press agencies, freelance journalists, translators, and typing services.

The growth in the number of firms categorized under SIC 8395 was noticed by some academics and policy makers and alerted them to the structural realignment that was been experienced by developed market economies. This is an important point in that it draws attention to emergent economic sectors that are hidden by the current composition of the SIC. It is essential that economic geographers explore the 'miscellaneous' SIC groupings as these include new high growth emergent activities. It is perhaps worth noting that the difficulty of classifying business services still exists. The European Union's nomenclature générale des activités économiques dans le communautés européennes (NACE) classification categorizes most business services under the residual category 'Other Business Services' (NACE 74). Both the SIC and NACE are constructed using a negative or residual classification of services. In this approach, miscellaneous categories are deployed to account for economic activities that cannot be accounted for by other SIC or NACE codes.

Characteristics of Business Service Firms

The growth in employment in business services was also reflected in extremely high rates of new firm formation. Business services were the main contributors of UK small firm service sector growth in the 1980s and 1990s. Not only were business service growth rates generally slightly higher than those for consumer service sectors, but business services generated nearly 3 times as many extra businesses (52 300) as consumer services (18 800) (1985–89). Particularly outstanding was an expansion of over 11 000 new computer service firms, but substantial growth rates were also experienced by management consultants, professional and technical services (surveyors, architects, consulting engineers, and draughtsmen), and market researchers.

The structural composition of all business service sectors reflects a bipolar distribution with a high number of very small firms and smaller numbers of large firms. This reflects a firm formation process that has comparatively few barriers to entry. In this context, it is important to distinguish between goods related services and KIBS. The goods related service providers supply relatively standardized support services often involving manual workers. They tend to have higher barriers to entry with regard to capital and equipment. They also tend to be much larger companies and in many cases these markets are dominated by large transnational service suppliers. A good example is Integrated Service Solutions (ISS), a Danish provider of goods related business services. This company was established in 1901 to provide security services, but has diversified into contract catering, cleaning, maintenance, facility management, and office support. Since 1998, ISS has made over 350 acquisitions and since 2005 the group has introduced a new strategy to transform itself from its core business as an international provider of cleaning services to a supplier of integrated service solutions customized to client needs. Compared to goods related services KIBS firms have very low barriers to entry (capital and equipment) as many of these firms sell expertise or knowledge founded upon the established reputations of individuals. Successful business service professionals create individual brands, and in some cases are transformed into iconic individuals, that are recognized by clients. Low barriers to entry means that from the 1980s the growth of small business service firms stands out as the single most important component of the rapid growth of small service sector business.

The majority of business service activities have yet to be transformed into professional occupations. The process of professionalization represents an attempt by an industrial sector to restrict entry by establishing a system of formal examinations or an accreditation process. Professionalism represents a form of social closure in that it is designed to maintain the integrity of an industry's knowledge base and also ultimately to limit supply and consequently enhance profit margins. The primary business service professions are represented by lawyers and accountants, but other business services are trying to transform themselves into recognized professions. Management consultancy has tried to pro fessionalize but failed due to the difficulties of identi fying a single branch of knowledge that can be defined as 'management consultancy'. Fundraising, a neglected business service targeted at not for profits, has also tried to professionalize though the establishment in the UK of the Institute of Fundraising in 1983, but this has so far failed to develop into a profession, in part, due to a shortage of fundraisers.

Explaining the Growth of Business Service Firms

The tendency for firms to externalize or outsource functions and activities previously produced in house (internalized) is one of the best known hypotheses used to explain the rapid relative growth of business services. It has its origins in research on the importance of com plementarity and economies of scale for decisions by firms about whether to externalize or internalize service functions. The problem facing all firms, irrespective of sector, concerns whether to 'make' or to 'buy' specified service inputs. Few firms can avoid making these de cisions because the inputs provided by business services can improve organizational efficiency and add value to a good or a service at different stages in the production chain. The more advanced or complicated the production process, the more significant the decisions about business service inputs. By choosing to 'buy', usually on a con tractual basis negotiated in advance, the firm is likely to be able to command the services of an expert employed by a firm specializing in a particular activity such as advertising, website construction and maintenance, or advanced computer software installation and configur ation for firm specific tasks.

Changing business practices and enhanced com petition have also been important drivers behind the growth of business services. Changing business practices as an explanation is, of course, related to the process of externalization if only because strategic downsizing and cost cutting especially by large client companies meant that many firms no longer had sufficient staff to in ternalize business service functions. The result was the creation of large numbers of new independent business service firms. Much of the research into the drivers be hind the growth of business services has revealed that cost driven externalization was not a major factor behind the growth of business service firms. The most important factor was demand for specialized technical expertise combined with a myriad of noncost and cost driven factors. Drawing upon a range of studies that have been undertaken on business services over the last 15 years it is possible to identify ten factors that have contributed to the growth of this sector of the economy (the first four factors list 'cost driven considerations' and the last six points list 'noncost considerations'):

1. *Transaction costs.* According to the transaction cost literature firms will turn to the marketplace for a range of services in cases where external provision is less expensive than internal provision. This explana tion for employing business service firms represents a 'pure' form of cost driven externalization. The price comparison between internal and external provision of business service expertise may be based on an actual comparison of cost differentials or be a per ception that external is cheaper than internal provision.
2. *Flexibility.* Firms employ business service firms where there is a requirement for different types and quan tities of expertise, information, and knowledge. In this instance it makes comparatively limited sense for the firm to employ full time staff, but instead to depend on external provision.
3. *Risk reduction.* By employing external providers of expertise clients are not exposed to the risks associ ated with full time employees. Clients do not have to provide training and invest in building or maintaining in house functions and during economic downturns clients can reduce their use of external expertise.
4. *Concentration on core skills.* This is the well known de bate concerning firms concentrating on what they do best rather than trying to run in house departments designed to provide peripheral business service ex pertise; peripheral in this instance to the company's core activity.
5. *Lack of expertise.* A firm may not have the required expertise. In this case the firm can decide to establish an in house department or draw upon external ex pertise. The difficulty and cost of establishing in house provision may encourage companies to consider employing external experts.
6. *New services.* The developing market for the supply of business services produced new skills and expertise that were rapidly incorporated into client's business requirements. Related to this is an increased awareness by clients of the types of expertise available.
7. *Third party expertise.* In some circumstances a client will require an independent evaluation of work undertaken in house or by another firm. This may take the form of an evaluation of a program of work or confirmation that an internal decision is correct.
8. *Growing complexity of management.* As business becomes increasingly competitive and global, client companies,

to remain competitive, have to access sophisticated design, consultancy, and marketing expertise. Enhanced global competition and increasing complexity of products and services has resulted in an increased demand for expertise that increasingly does not reside inside firms.

9. *New technology*. New technology frequently requires alterations to working practices that may be beyond the expertise and experience of a company's employees.
10. *New regulations*. New regulations (health and safety, employment, environmental) may be introduced that require expertise that is only required for a short period of time.

Business Services – the Body and Emotional Labor

Successful business service entrepreneurs possess three important attributes: professional expertise, an existing reputation (sometimes iconic), and a network of client contacts. These essential requirements for competitive success explain the concentration of founders' previous employment in same sector or client firms, and into larger rather than smaller companies. The founders of business service firms acquire their reputations, expertise, and client contacts while working for either large supply or client companies. This partly accounts for concentrations of business service firms in major cities.

Professional reputations rather than knowledge drive the formation process for business service firms. This is because an imperfect market place for business service expertise exists. Part of this imperfection is the difficulties clients have in determining the quality of a supplier's expertise. The problem is wrapped up in a product – knowledge/expertise – that is intangible and whose quality is difficult to measure or assess. One way of overcoming an imperfect market is to employ individuals and companies that the client already knows either directly or indirectly. Personal knowledge of individual professionals rather than of companies is thus extremely important in the client search process. The implication is that the balance of power in the labor market is held by individuals with established professional reputations and the value of this expertise is rising. This means that consultants, advertising executives, lawyers, investment bankers, and financial analysts, in effect, rent their brains, as well as reputations. The implication is that graduates enter large business service companies as trainees to acquire knowledge and a reputation amongst clients. Once these have been acquired, there is nothing to prevent the young professionals from leaving the firm, taking clients with them, and establishing their own practice. Business service activities have very low capital barriers to entry, but high human or reputational entry barriers.

The importance of presentation, communication, and display in service work implies that such work cannot be conceptualized solely as an economic performance. Business service work is a 'hybrid' form of work in which the economic and the cultural are blurred; this means that this form of work is a qualitatively different form of work to manufacturing employment. In effect, the provision of a business service to a client represents a dramatic act or a form of impression management. Impression management is a key feature of the work of KIBS as it is an essential part of the ways in which they present themselves and interact with clients. The accountants Coopers and Lybrand, as well as Ernst and Young hire image consultants to provide seminars in personal presentation for their audit teams. These seminars examine ways of increasing credibility and projecting the right image when undertaking client audits and when pitching for new business. Price Waterhouse employed image consultants to instruct potential partners in dining etiquette and in the art of looking, acting, and sounding like a partner of a major global accountancy company. This implies that KIBS represents a form of 'emotional labor', a term which describes the process by which employers manage employees' feelings during social interaction in the work process. Much of the literature on emotional labor concentrates on exploring relatively low paid service work in which employers are actively engaged in managing the hearts and minds of their employees. In business service work, the professional institutions, the employers, and most importantly individual professionals are implicated in emotional labor in order to succeed in this economic sector.

Business Service Geographies – Global Cities, Service Offshoring, and the Second Global Shift

Global Cities

Business services have a complex geography that reflects the existence of concentrations of clients in major cities or in the global cities of London and New York. Business services have been growing rapidly not only in the major cities, but also in locations further down the urban hierarchy and even in rural locations. Conventionally, it has been assumed by economic geographers that the major cities are the locations for business service firms that provide the best expertise for client companies while firms located in other locations provide adequate expertise. This is an assumption that has yet to be tested by empirical research. Concentrations of business service firms have been used as one of the most important indicators to identify world or global cities, but this research design has not addressed differences in the quality of available expertise between firms located in major cities and those located elsewhere. The difficulty is that the global cities literature prioritizes cities in the research design while neglecting to explore the ways in which expertise is produced and delivered by business service firms to their clients. The global cities may just be the access node used by clients for expertise that exists elsewhere in the urban hierarchy. It is perhaps worth noting that the global cities literature is rooted within the 'central place theory' and especially with a concern with understanding the location of high and low order services.

Rural Areas – Lone Eagles and High Fliers

Research by economic geographers into the activities of business service firms has been dominated by studies of firms located in major cities. This is unfortunate and it means that geographers have comparatively limited understanding of the knowledge economy and the activities of KIBS in smaller cities, market towns, and rural areas. Research into rural based business and professional services (BPS) is dominated by the American studies undertaken by William Beyers. He identified that the American Midwest has experienced significant growth rates in business service firms and importantly some of these companies are engaged in interregional and even international exports. Beyers classified exporter's rural business service firms as 'lone eagles' (one person proprietorships) or 'high fliers' (firms with one or more employees). By the very nature of the kinds of services offered by lone eagles and high fliers it is unlikely that they will be producing outputs intended primarily for consumption in local markets. There will, for example, be very limited demand for high level management consultancy or advice on the installation and maintenance of corporate IT systems in the small and dispersed communities typical of rural areas in the US, Canada, Australia, or Europe. Much of the output of these advanced services, and indeed those located in the much larger and longer established urban centers, are of necessity traded between localities and regions.

Service Offshoring and the Second Global Shift

Business services are usually supplied and consumed locally in a process of face to face co production. This relationship is based on embodied expertise centered on the ability of service providers to enter into interactive relationships with clients. Heavily embodied services have conventionally been considered to have low productivity levels as it is difficult to substitute embodied expertise with machines. The relationship between embodied expertise, machines, and proximity between the supplier and consumer of a business service has been challenged with developments in computer programs and information and communication technologies (ICTs).

New technology has the potential to create entirely new services or to challenge the relationship that conventionally exists between the co production of a service in time and space. Initially, replacing some employees with machines increased the efficiency and profitability of service firms, but local expensive labor was still a central component of these production systems. Nevertheless, the development of new technology and associated production systems offered the possibility of relocating some service functions from high to low cost production locations (India, Eastern Europe, and the Caribbean). Some business services are thus following manufacturing in the development of a 'global shift'. The dynamics and resulting geographies of the 'services shift' are so different to that of manufacturing that Bryson has argued that what is occurring is a 'second global shift' or the development of a new international division of 'service' labor (NIDSL). During the first global shift, branch plants in developing or less developed countries were associated with the assembly of products designed by and for the developed world. During the second global shift, service functions are transferred from developed market economies to other locations. This shift could be designed to reduce costs, to provides services for a foreign market, to develop 24/7 provision, to reduce exposure to country risk, or to access skilled labor. The geography of the second global shift is different to that of the first. It is more constrained by language, as well as cultural nearness, or the ability of foreign producers of services to relate to customers located in other countries. An additional factor influencing the location of offshore service facilities is the requirement to provide a 24 h service to customers or an extended service beyond standard working hours. The 'second global shift' is a recent development and academics and policymakers are just beginning to explore this process. This is a complex issue that embraces economic, as well as political geography. Data processing has now become a global industry and governments are increasingly sending sensitive data to be processed abroad. The UK government, for example, in 2007 entered into a data processing contract with the German company Siemens. Siemens is inputting every record of births, deaths, and marriages in the UK dating back to 1837, but this data has been transferred to Chennai, India, for processing. The NIDSL now involves the routine processing of government data in foreign countries that has the potential to expose citizens to identity theft.

A number of factors lie behind the decision to send a particular service activity offshore. First, the service must be capable of some degree of standardization that does not require face to face interaction with clients. Second, the inputs and outputs required to deliver a service must be capable of being traded or transmitted with the assistance of ICT. Third, some service activities are not fixed in space and can be provided either as a form of foreign trade or by the temporary relocation of a service worker to a client's premises, for example, management consultancy or various forms of auditing.

Conclusion

The growth in employment in business services and in the number of business service firms represents one of the most important structural realignments to be ex perienced by developed market economies. This growth reflects the shift toward competitiveness based upon ex pertise and education. Education and the effectiveness of a country's educational system are essential for the continued development of business services. This inter play between business service firms and the educational system requires further research. Academics must ex plore the ways in which business service professionals acquire, maintain, and develop their expertise. This re quires research into the educational system, and also into the activities of the professional bodies and the role they play in forming 'educated' professionals. It also involves research into identifying the 'soft' personality related skills that are so important in the client–business service provider relationship.

It is important to remember that business service firms sell expertise, but that this involves an important trans formation that requires considerable further research. Business service firms commercialise expertise by trans forming it into a product that has value. The existing literature overemphasizes the quality and nature of the expertise provided by many business service firms. Comparatively few academics have explored the types of knowledge/expertise provided by business service firms to their clients. Many business and professional service firms provide forms of recipe or standardized knowledge that is customized to meet client needs and expectations. This is an important point – the expertise is readily available, but the added value is formed during a process of commercialization and customization. While the knowledge or expertise may be standardized and largely generic, the creativity may be found in the process of commercialization rather than in the actual expertise. This raises important questions regarding the nature of the creativity that occurs in the 'creative' side of the economy. In many instances, creativity may just be a reflection of a business service professional's commercial or entrepreneurial skills rather than being an indication of creative expertise.

See also: Economic Geography; e-Business and e-Commerce; Innovation; Local Economic Development; Services, Professional.

Further Reading

Beaverstock, J. V. (2007). Transnational work: Global professional labour markets in professional service accounting firms. In Bryson, J. R. & Daniels, P. W. (eds.) *The Handbook of Service Industries*, pp 409 432. Cheltenham: Edward Elgar.

Beyers, W. B. and Lindahl, D. P. (1996a). Explaining the demand for producer services. *Papers in Regional Science* 75, 351 374.

Beyers, W. B. and Lindahl, D. P. (1996b). Lone eagles and high fliers in rural producer services. *Rural Development Perspectives* 12, 2 10.

Bryson, J. R. (1997). Business service firms, service space and the management of change. *Entrepreneurship and Regional Development* 9, 93 111.

Bryson, J. R. (2007). A 'second' global shift: The offshoring or global sourcing of corporate services and the rise of distanciated emotional labour. *Geografiska Annaler* 89B(S1), 31 43.

Bryson, J. R. (2007). Lone eagles and high flyers: Rural based business and professional service firms and information communication technology. In Rusten, G. & Skerratt, S. (eds.) *Being Rural in a Digital Age*, pp 36 60. London: Routledge.

Bryson, J. R. and Daniels, P. W. (eds.) (2007). *The Handbook of Service Industries*. Cheltenham: Edward Elgar.

Bryson, J. R., Daniels, P. W. and Warf, B. (2004). *Service Worlds: People, Organizations and Technologies*. London: Routledge.

Bryson, J. R., Taylor, M. and Daniels, P. W. (2008). Commercializing 'creative' expertise: Business and professional services and regional economic development in the West Midlands, UK. *Politics and Policy* 36(2), 306 328.

Bryson, J. R., Wood, P. and Keeble, D. (1993). Business networks, small firm flexibility and regional development in UK business services. *Entrepreneurship and Regional Development* 5, 265 277.

Greenfield, H. I. (1966). *Manpower and the Growth and Producer Services*. New York: Columbia University Press.

Marshall, J. N., Wood, P., Daniels, P. W. *et al*. (1988). *Services and Uneven Development*. Oxford: Oxford University Press.

Perry, M. (1990). Business service specialisation and regional economic change. *Regional Studies* 24, 195 209.

Rubalcaba, L. and Kok, H. (eds.) (2007). *Business Services in European Economic Growth*. London: Palgrave MacMillan.

Rusten, G., Bryson, J. R. and Gammelsæter, H. (2005). Dislocated versus local business service expertise and knowledge and the acquisition of external management consultancy expertise by small and medium sized enterprises in Norway. *Geoforum* 36(4), 525 539.

Sassen, S. (2001). *The Global City: New York, London, Tokyo*. Princeton, NJ: Princeton University Press.

Taylor, M., Plummer, P., Bryson, J. R. and Garlick, S. (2008). The role of universities in building local economic capacities. *Politics and Policy* 36(2), 216 231.

Relevant Websites

http://www.census.gov/econ/www/servmenu.html
American Government website that provides information on the U.S. service industry.

http://www.statistics.gov.uk/methods quality/sic/contents.asp
British government website that provides details of the UK Standard Industrial Classification of economic activities (SIC).

http://www.epp.eurostat.ec.europa.eu/portal/page? pageid=1996, 45323734& dad=portal& schema=PORTAL&screen=welcomeref& open=/industry&language=en&product=EU MAIN TREE&root=EU MAIN TREE&scrollto=0
Eurostat database that provides statistical information on services across the European Union.

http://www.reser.net/
The website of the European Association for Research on Services.

C

Capital and Space

R. J. Das, York University, Toronto, ON, Canada

Introduction

Tools, machines, and other types of means of production are considered to be capital in many conventional eco nomic geography texts. In this view, capital is not only a tangible thing but also a transhistorical thing. In a slightly more general view, capital is defined as any income generating asset that generates some benefit, including income. Using capital metaphorically, scholars talk about natural capital, produced capital (means of production produced by human activity), human capital, social capital, and cultural capital. Although in this view, social relations are included in the definition of capital (at least in case of social capital), transhistoricality of capital is still implied: for example, relations of reciprocity as a form of capital have always existed.

Like means of production, money as such is wrongly considered by many to be capital. When I spend my money, which I earn by selling my services to my em ployer, to buy my books, the money I use is not capital. Money has existed before capital. But money 'can' be come capital. Indeed, money is capital's first form of appearance. There are two ways of looking at money as capital. When a sum of money is advanced by a merchant or a moneylender to make a profit on it, it can be cor rectly called capital. Merchant and moneylending/usurer forms of capital appeared earlier than the third form of capital. This form of capital arises when the owner of money lays out money on wages and means of production to produce a new commodity which is sold for more money than what is initially advanced. This view of capital establishes capital's theoretical specificity as a social relation. As Marx says in *Capital vol 3*: capital is not a thing, it is rather a definite social relation of production between owners of means of production and labor, which simply takes the form of a thing. Means of production are no more capital in themselves than gold as money. They become capital under the relation between capital and labor. Capitalism is a form of society where the capital as money (investing money as capital to make more money) becomes the dominant relation, and more specifically when the third form of capital becomes (tendentially) the dominant form of capital relation.

Capital can be looked at in terms of capital in general (total capital in the global society or within a country or region) and capitals (as fractions of capital in general), which compete with one another for market, for support from the state, and so on. Capital in general appears as, and is reproduced through, various fractions of capital com peting with one another, including those that are based in specific spaces (e.g., regional scale capitalist in a country).

The remainder of the discussion is divided into five parts. First, we explain how capital in the sphere of production works, and in particular the relation between capital and labor. Then capital's space, both in terms of its normal workings and in terms of the crisis situation, is discussed. Subsequently, we discusse capital's space in the sphere of consumption. The discussion finally is sum marized and some (understressed) points for further re flection are raised.

How Does Capital Work?: Mechanisms of Capital

To understand how capital works, the author returns to Marx's general formula for capital as he outlines in chapter 4 of *Capital vol 1*: M–C–M′, which can be ex panded into: M … C [MP, LP] … P … C′… M′ (see **Figure 1**).

The general formula means that: the owner of capital invests money (M) to buy commodities (C) which include means of production (MP) (e.g., machine, energy, and raw material) as well as labor power (LP) in order to produce a new commodity (C′) which is then sold for M′. M′ includes the initial money invested (M) and a surplus (′). The owner of capital keeps a part of M′ as profit and another part is shared with bankers (interest), and land owners (rent) (if necessary). A part of the profit is spent on private consumption (including luxury consumption) and another part is reinvested in the second cycle to enlarge the scale of production. This process of reinvestment of

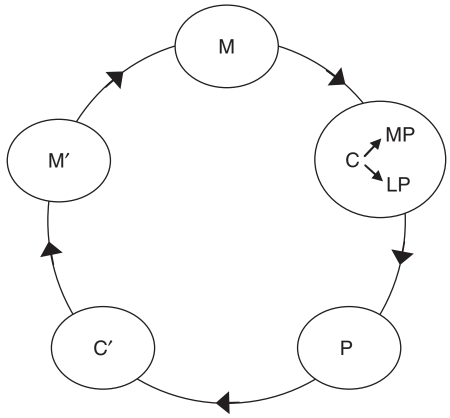

Figure 1 The circuit of capital.

profit is known as accumulation. M, C, … C′… M′ form the circuit of capital.

In the process of production, means of production simply transfer their value (the amount of labor time necessary to produce a commodity under average con ditions) to the final product. The exercise of the power to labor transfers more value than the labor embodied in it. This is because wages that labor (= laborer) gets paid are below the value that she produces for capital. This pro cess is the process of exploitation producing surplus value which is the source of profit. The relation of ex ploitation has spatial implications as discussed later.

Capital is indeed value in motion. Capital has a restless drive for making money out of money. Accu mulation, a positive rate of growth, must happen. But this is not guaranteed. For accumulation is prone to over accumulation crisis which slows down the rate of accu mulation (consider the extremely slow rate of economic growth in the currently advanced countries). Why? In dividual capitals are subjected to coercive laws of com petition. To continue being in business and making money, each individual capitalist has to reduce costs below the social average cost. This happens partly through the use of technology which, in turn, displaces labor. But there is a contradiction here. It is the laborers who buy and consume the commodities that are pro duced. It is they who create profit for capital. Also, each fraction of capital sees its own employees as a cost to minimize, so it keeps wages low, which depresses the market and contributes to the overaccumulation crisis (more on this later). Both accumulation and over accumulation of capital are spatial processes.

Capital and Space: Capital's 'Normal' Spatiality

Each phase in the circuit of capital has a corresponding spatial form. There are thus spaces of production, spaces of realization (capital realizes the surplus or profit when the product is sold), and spaces of accumulation (where a part of the surplus/profit is reinvested). An interesting aspect of these spaces is the mutual incongruence be tween these spaces: commodities produced in California (spaces of production) may be sold in China (spaces of realization) and the profit made may be invested in Canada (spaces of accumulation). Value can be produced in a less developed country through exploitation of its labor but may be transferred to advanced countries, leading to unequal development between nations.

The logic of value – the logic coercing every capitalist to produce commodities at a competitive price – pre supposes a spatial network. While concrete useful labor produces use values at a particular place (e.g., labor producing shirts), acts of exchange spatially connect these particular place specific instances of concrete labor. It is indeed only through the exchange can one know what the average cost of production is or what the so cially average labor (or value) is. It is not enough to say that capital 'requires' a spatial framework. We must understand how capital also 'produces' this spatial framework. Let us return to the capital circuit. The cir culation of capital from M to M′ has to be completed within a definite time span, which has two components. One is the time needed to actually produce a commodity. Another is the time spent in the spatial circulation of commodities necessary for the completion of the circuit of capital. Competition among capitals generates pres sure to reduce these two components through techno logical and organizational changes. This means that not only production time is to be shortened but also the spatial circulation time. Raw materials, finished products, and money have to move in space, and move they must fast. Marx beautifully expressed this idea in *Grundrisse*: this is that while capital must on the one hand strive to overcome every spatial barrier to exchange, it strives on the other hand to annihilate this space with time. The imperative to accumulate implies the imperative to overcome spatial barriers (the 'friction of space').

Like most processes under the rule of capital, the production of space is contradictory. Transformation in the cost, speed, regularity, and efficiency of movement over space alters the relative distances of places of pro duction from the larger markets. This can potentially lead to the deterioration of old and rise of new centers of production and a perpetual shifting and relocation of places of production, and of markets as a result of changes in their relative positions, and with it the relative fortunes of capitalists – as well as that of labor – based in different locations also changes. It is this sort of 'constant' changes in spatial relations that Storper and Walker call "the inconstant geography of capitalism." The chronic instability in spaces of capital is something that no amount of state regulation can permanently overcome.

Capital creates a physical landscape in its own image at a given time and undermines/destroys it at a later time.

Production of space is a contradictory affair in at least two more ways. A part of the total social capital at a given point in time has to be immobilized in the form of social and physical infrastructures, in the form of the built environment (e.g., transport network), for the remainder of the capital (e.g., capital in the form of raw materials or finished products) to be mobile. The viability of capital invested in the production and maintenance of the immobile infrastructure can be assured only if the remaining capital generates the intended profit. If, for example, sufficient traffic fails to be generated to make a railroad profitable, then capital invested in this can be potentially devalued. The production of space is contradictory in an additional manner. The production of space shrinking technologies and efficient spatial configurations open up more 'distant' markets, new sources of raw materials and new opportunities for the employment of labor under social relations dominated by capital. But this, however, increases the spatial circulation time (unless there are compensating improvements in the speed of circulation).

Space has both mobility and fixity aspects. Capital's production of space is characterized by its requirement for spatial mobility (which has been stressed so far) as well as spatial fixity. Although capital seeks to overcome space through the use of space shrinking technologies, technology defined spatial constraints exist (and this applies to multinational companies), but some fractions (e.g., real estate capital) are more spatially constrained than others (e.g., money capital). The need to reduce the cost/time of movement between spaces of production, between the spaces of production and spaces of consumption, and between workplaces and living places remains, which means that the production and reproduction of capital is indeed location bound. In terms of production only, capital (in the form of means of production) and labor (of different types) must be brought together in space. Spatial expression of this is imperative, for example, a factory at one scale and an industrial city at another and an industrial region at a higher scale. It is true that from the standpoint of capital as a whole (i.e., capital in general), free spatial mobility of labor is a necessary condition, so that labor is available in the spaces where capital needs them. But the situation is different for individual fractions of capital (e.g., individual firms) located in specific places. They need a stable reliable captive labor force which is accessible on a daily basis. Capital is also dependent on an industrial reserve army (an army of unemployed and underemployed people) *in situ* in order to control those who are employed and keep local wages down. All this means that capital 'may' support basic reproduction processes such as education and health, geared toward production and preservation of labor supply 'within' a given space. This also means that capital may support state policies to constrain free mobility of labor in space. Like capital, labor (as a variable capital) has a need for certain degree of fixity. It is locally dependent for the satisfaction of many of its needs. Labor needs to find and work for an employer 'within' a circumscribed space; and women and child laborers tend to be particularly spatially circumscribed. Labor makes use of social and physical infrastructures, including institutions of community and social network and state regulatory institutions, none of which is easily substitutable and portable in space. Thus, urbanization of capital–spatial concentration of productive forces – including social and physical infrastructures in specific places – is an inherent tendency in capital. The need to minimize the cost of movement of commodities promotes agglomeration of production and spatial concentration of specific fractions of capital such as textile capital or IT capital within a few large urban centers (the concentration effect of capital). We know this from the attempt of different capitals to locate close to one another to minimize costs of movements of intermediate products, to take advantage of a large labor pool within a definite space and to share knowledge. Agglomeration is about reducing cost and time of spatial movement.

There are limits to spatial concentration/agglomeration, however, in part because spatial concentration of workers is conducive to labor struggle which threatens profit rates, and this may lead to capital moving out. Also, revolution in technologies can liberate production and consumption from their local roots, so the spatial constraints of the past become less stringent. This makes the boundaries of urban/regional economies porous and unstable. In addition, transformation in capitalist organization – multinational corporations (MNCs), rise of financial capital, etc. – may permit control over progressively larger spaces by capital. To the extent that the production of a distinct efficient spatial landscape for/of capital itself promotes greater profit making activities, this produces surplus of capital, the overaccumulation problem (discussed later) putting pressure on some capital to leave for profitable opportunities elsewhere or even to pull in laborers from elsewhere to depress local wages by increasing the labor supply. Labor also follows a similar logic. If wages are not adequate here, labor may want to move there, a type of behavior that fits in with the needs of capital in general to coordinate supply and demand of labor power in space but that may hurt specific fractions of capital which tend to be more spatially constrained. Thus while capital and labor require and realize fixity, they, beyond a point, are vulnerable to the imperative of mobility.

In order to address this tension between the need for fixity and the need for mobility on the part of capital and labor, a regional cross class alliance between them may develop. It is usually organized through the state to protect the investment in place bound physical and social infra structures and to promote further accumulation within given territories (e.g., attracting extra local capital). Dif ferent fractions of capital and labor have differential stake in the territory depending on the nature of their assets. Land developers, like state functionaries dependent on local tax base and local votes, have greater stake in the regional space economy, because they are less mobile and may be forced to buy local labor peace and skills through compromise over wages and working conditions. Fractions of labor that have through struggle or out of scarcity managed to create spaces of privilege will also support the alliance. The community, regional and national level boosterism may be used as an ideology to strengthen the alliance. What is observed is spatial competition between cities, regions, and nations, that is, between different spaces of capital accumulation at different scales, for in ward investment, which threaten to supplant (inter national) class struggle between capital and its oppositional other. Capital's action/logic thus leads not only to spaces of economy but also to spaces of politics.

Capital and Space: Capital's Crisis-Spatiality

Capital accumulation and its relation to space cannot be looked at without a serious attention to the contradictory character of capital. There are four major types of con tradiction. Each of these is a dimension of the main contradiction: the contradiction between social relations and productive forces under capitalism. The use of space is crucial to the partial alleviation (or temporary dis placement) of these forms of contradiction.

First, capital accumulation requires the environment as a source of raw materials and energy as well as space to dump its waste, but capital accumulation leads to serious degradation of the environment because of capital's blind desire for profit. To be noted is the fact that much of the environmental problem can be explained partly in terms of the need for shrinking space: consider the pollution caused by cars and other types of vehicles. Capital seeks a spatial fix here: ecological problems at home can be transferred to other spaces (consider how floriculture activities which caused serious environmental pollution in the Netherlands have been relocated to the less de veloped countries). But ecological problems do not know spatial boundaries. Pollution there affects us here. A spatial fix will not work. The second contradiction is between capital and 'noncapitalist economic practices'. These may be spatially separated from spaces of capital accumulation (e.g., subsistence production in the tribal areas), and thus they are, in a sense, 'outside' of the spaces of capital. But insofar as they also supply labor (season ally) to centers of capitalist accumulation, and thus hold wages down and enhance capital's profits, and insofar as they are connected to capital's spaces through exchange (e.g., by buying some of the things they need or selling a part of what they produce), they are very much articu lated to the spaces of capital. It is this articulation that some of the recent economic geographical literature in fluenced by postmodernism tends to understress, and tends to valorize these spaces as 'alternative economic spaces', as spaces of empowerment. To the extent that these spaces become spaces of the primitive accumu lation (through mercantile, colonial, and other activities), they can, of course, be potential terrains of resistance against this process, and therefore their use for capital is not to be taken for granted. But these spaces are 'not' automatically anticapital spaces. The third contradiction is between capital and labor, the most important con tradiction politically. Accumulation of capital is simul taneously reproduction of domination of capital over labor. In its struggle against labor, space is used by capital: just as cost cutting technology is used to obtain relative surplus value, relocation to spaces of greater rate of profit serves the similar purpose. Space and location can be active sources of above average profit to individual cap italists. Relocation of capital is possible given a prior spatially uneven surface of profit opportunities. But this process only globalizes possibilities of class struggle. If capital uses space to deal with working class struggle in various ways (e.g., getting different parts of a commodity made in different places to extricate itself from de pendence on labor in any given space, and to restrict working class struggles to the local scale), labor also strives to use space and globalize its struggle, albeit with less success than capital so far.

Attendant to, and as a result of, the labor–capital contradiction is the contradiction of overaccumulation. Accumulation – reinvestment of a part of profit to en large the scale of production – and consequent devel opment of productive forces is a product of intercapitalist competition. In the process, commodities are produced without regard to the limits of the market, because human needs and the methods to meet these are not subject to collective democratic decision making. While, as suggested earlier, capitalists produce increasing mass of commodities they also try to maximize profits by keeping wages down, thus restricting the purchasing power of the society (for nonluxury goods/services). This leads to the overproduction or the overaccumulation crisis. There is a surplus of capital relative to profitable investment of opportunities. Factories and means of production stand but remain unused or underused (ex cess capacity).

The overaccumulation crisis is sought to be resolved through the spatial fix. The spatial fix can take many forms.

One is the production of spatial landscape in the form of bulky immobile investments. These are fixed capital (physical plants, and immobile social and physical infrastructure, etc.) which can contribute to the expansion of accumulation and can increase the rate of profit. The nature of the investment in these areas is such that the return of benefits (benefit = contribution to expansion of accumulation) is spread over many years so they can absorb surplus capital over a long period of time and alleviate the overaccumulation problem. The solution is problematic, however. If the built environment contributes to profit making, it contributes to overaccumulation. If it does not, the capital sunk in it is subject to devaluation.

Another form of spatial fix is to export the overaccumulation problem from one place to another, including a foreign country and, if necessary, through the use of coercive state power of (new) imperialist states. But this form of spatial fix is not unproblematic. Surplus capital can be used to create new regional economies abroad which may address the overaccumulation problem for long time periods. But these space economies also produce their own overaccumulation problem and may resort to the spatial fix and compete with the mother country as the USA did to the UK. Interestingly, this is what is happening now. China is contributing to the crisis of global overcapacity. As a huge amount of MNC investment has gone into China, China's idle capacity in key sectors such as steel, automobile, cement, aluminum and real estate has been soaring since the mid 1990s, with estimates that over 75% of China's industries are currently plagued by overcapacity.

Or, to avoid the situation of being out competed, a dependent space economy can be created which will only absorb the overaccumulated surplus, and it can be made to produce, again, with the use of imperialist state power, whatever the home country needs and in the needed quantities. The problem is that such a space economy cannot develop freely and therefore cannot absorb enough surpluses. There is clearly a contradiction here. If the new region develops freely, it may compete with the home country in disposing of its own surplus, and if it does not, it may not absorb enough surplus from the home country. The effect is to generalize capital's problem worldwide.

The overaccumulation crisis cannot be avoided. The process of capital trying a spatial fix resolves into an economic and political struggle between nation states, to the struggle over who is to bear the brunt of the crisis. Export of unemployment, inflation, and idle productive capacity becomes the stakes in an ugly game. Trade wars, dumping, tariffs and quotas, restrictions on capital flow, immigration policies, (neo)colonial conquest of and subjugation of dependent economies, and finally physical destruction through war – all these become a part of the attempt at the crisis resolution. Less powerful, poorer countries suffer the most. Capital creates not just spaces of accumulation and wealth but also spaces of destruction. The history and geography of spatial fix (the use of space by capital in its own interest) – like that of primitive accumulation and technological change – is written in letters of blood and fire.

Capital's Consumption Spatiality

Capital makes space not just for accumulation. Spaces of consumption are also important without which value will not be realized. But spaces of consumption are also use values that labor demands. For example, housing is an important working class demand. Capital, however, uses it as a means of investment: housing and the like can absorb a lot of surplus capital staving off the overaccumulation crisis. Homeownership can also create political stability (through the cultural discourse of homeownership).

Spaces of accumulation also influence spaces of consumption. As Marx observed of cities of his time: the greater the centralization of the means of production, the greater is the corresponding 'concentration of workers within a given space'; and therefore the faster the capitalistic accumulation takes place, the more miserable the working class housing situation. Indeed capital's production of space in its own image leads to particular implications for working class spaces of consumption. 'Improvements' of towns accompanying the increase of capital accumulation take the form of the demolition of badly built districts, the widening of streets for business traffic, for luxury carriages and so on, and all this obviously drives the poor away into even worse and more crowded corners, with state's complicity, of course. Housing estates in New York City and slums in Mumbai are great testimony to this.

The production and consumption of commodities produce and presuppose a spatial form. By exploiting the world market, capital gives a cosmopolitan character not only to production but also to consumption in every country. Our old wants, satisfied by products of our own countries, are replaced by new wants which are satisfied by the products of distant lands made by 'distant strangers'. Spaces of commodity circulation are arguably connected to the spaces of commodity fetishism. Commodity fetishism occurs when relations between persons appear as relations between products they exchange. A fundamental reason for this is the lack of labor's control over the production of commodities. The spatial form of commodity exchange contributes to, although it is not a fundamental cause of, commodity fetishism because commodity markets conceal social–geographical information and relations. We cannot tell from looking at a commodity whether it has been produced by exploited child or women laborers in a poor country, or by wage laborers protected by adequate labor legislation in a rich country. In a sense, the specifically capitalist character of commodity–space contributes to a certain erasure of the footprint of space on commodities, especially when commodities represent ever more complex encounters involving ever more people over ever greater distances. The spatiality of commodity fetishism is capital's 'spatial veil', through which the rule of capital is culturally reproduced. If it is true that as Lefevre argued, the very survival of capitalism depended on the creation of a socially mystified spatiality, the latter must include the spatiality of the fetishism of the commodity.

Conclusion

Capital is the name given to a 'process' of production of value and of exploitation of labor by owners of capital. Capital is also a process manifested as things, as means of production, and as commodities. It is also a 'relation' between labor and owners of capital. Capital is both capital in general, the total social capital as well as individual capitals, including those whose operations tend to be based in specific regions and sectors.

The inner contradiction of capital is expressed through the restless formation and reformation of the spatial landscape. In order to overcome spatial barriers, that is, to annihilate space with time, spatial structures are produced which themselves ultimately act as a barrier to further accumulation of capital. These spatial structures are partly expressed in the fixed and immovable form of transport facilities, plant, and other means of production and consumption which cannot be moved without being (partially) destroyed/devalued.

Implicit in the foregoing discussion is the fact there are two moments to the production of space under the rule of capital. Capital produces space, its built environment (this is the first moment of capital's production of space) but in a spatially uneven way thus producing space in the sense of spatial unevenness (the second moment of capital's production of space). This second moment of production of space – spatial unevenness – happens due to uneven penetration of capitalist relations, which happens partly because of a geography of class struggle against the penetration of capital into noncapitalist spaces. Even the current tendency towards 'accumulation by dispossession', including the process of dispossession of people from their land, progresses unevenly. Further, the physical transformation of the landscape (cities, transport system, factories, shopping malls, etc.) also occurs unevenly over space. And this second moment of the production of space – the spatial unevenness of capital relations and development of productive forces under the rule of capital – happens at different scales, leading to uneven spatial development at local, regional, national, and global scales. The actual forms of spatially uneven development and their mechanisms must be investigated.

The production of space under the rule of capital is not entirely an economic affair. Indeed social relations of gender and race and sexuality, and subsistence based economic relations and local cultural traditions influence the ways capital produces and makes use of space. For example, at an abstract level, an owner of labor has one unit of labor power and whether, therefore, a labor is a man or woman, whether the owner of capital is a man or woman, makes no difference to the mechanism exploitation. But the gendered or racialized nature of labor, for example, 'does' make a difference at the level of competing capitals, to the rate of exploitation and to the possibilities of political opposition to it. The location of particular kinds of capital is partly determined by the social–ideological construction of women labor as being docile and as having soft and nimble fingers (useful in concrete labor practices in units producing and processing new agricultural commodities such as flowers, and in assembly plants). Indeed, specific fractions of capital based in specific places and in mutual competition for excess profits can appropriate above average profits by using women labor and racially minority workers socially construed as vulnerable and who indeed are made vulnerable; consider tribal laborers or non white immigrant laborers in advanced countries. The ways in which the production of space occurs through the intersection of the logic of capital with noncapitalist logics, the logics of race and gender, need detailed analysis. The spaces that capital creates are most definitely gendered and racialized.

Not only does capital produce space. The spatial structure of capitalism produces positive benefits for capital as well, as indicated earlier. Capitalists in the developed regions/countries use superior technology and cheapen their goods and services below the average cost and thus sell these above the value in competition with commodities produced in technology poor regions. This will allow them to obtain excess profits from the regions using inferior technology. Individual fractions of capital make use of spatial differentiation – differentiation in rate of profit and wage rates, in sharing technology, in incidence of class struggle, etc. – to increase profits. Center–periphery relations between powerful developed nations and less developed less powerful nations – a specific form that capital's space takes at the global scale – are reproduced through new imperialist practices, and this requires serious analytical attention. Very briefly, analogous to technology, control over low cost locations in the periphery is a strategy employed by core country capitals backed by their states militarily to cut costs, derive excess profit, and remain globally competitive.

The process of production of space by capital – spatially uneven development, for example – has also an import ant ideological function for capital. Uneven development means that some spaces are developed and others are not. It is the developed spaces of the spatial equation that creates the idea that other spaces currently under developed 'can' achieve development and if they are less developed now, the reasons must be 'in these spaces themselves'. Capital's space acts as a veil for it.

Capital revolutionizes productive forces, including spatial relations. But one must not lose sight of the ir rationality of the space making processes of capital. On the one hand, we observe the overaccumulation of space shrinking technologies and spatial infrastructures (air ports, highways, etc.), but at the same time millions of people are in absolute and relative isolation from one another, from centers of employment, and so on. This realization is an important precondition for struggles against capital's space making project.

See also: Capitalism; Core-Periphery Models; Harvey, D.; Radical Political Economy; Regional Inequalities; Regional Planning and Development Theories; Space I; Space II; Uneven Development.

Further Reading

Bello, W. (2006). The capitalist conjuncture: Over accumulation, financial crises, and the retreat from globalisation. *Third World Quarterly* 27(8), 1345 1367.
Castree, N. (2001). Commodity fetishism, geographical imaginations and imaginative geographies. *Environment and Planning A* 33(9), 1519 1525.
Gibson Graham, J. K. (2006). *A Postcapitalist Politics*. Minneapolis, MN: University of Minnesota Press.
Hanson, S. and Pratt, G. (1995). *Gender, Work and Place*. London: Routledge.
Harvey, D. (2001). *Spaces of Capital: Towards a Critical Geography*. New York: Routledge.
Harvey, D. (1982). *Limits to Capital*. Chicago, IL: Chicago University Press.
Hudson, R. (2004). Conceptualizing economies and their geographies: Spaces, flows and circuits. *Progress in Human Geography* 28(4), 447 471.
Leyshon, A., Lee, R. and Williams, C. (eds.) (2003). *Alternative Economic Spaces*. London: Sage.
Marx, K. (1977). *Capital: Vol. 1. A Critique of Political Economy*. New York: Vintage.
Marx, K. (2005). Chapter 1. In Gasper, P. (ed.) *The Communist Manifesto*. Chicago, IL: Haymarket Books.
Massey, D. (1995). *Spatial Divisions of Labor*. London: Macmillan.
Soja, E. (1989). *Postmodern Geographies*. London: Verso.
Storper, M. and Walker, R. (1989). *The Capitalist Imperative: Territory, Technology and Industrial Growth*. New York: Blackwell.
Yeung, H. (1998). Capital, state and space: Contesting the borderless world. *Transactions of the Institute of British Geographers* 23(3), 291 309.

Capitalism

R. Le Heron, The University of Auckland, Auckland, New Zealand

Glossary

Accumulation The prime goal of capitalist production is to expand capital, through reinvestment of profits, into new or more capital; this ongoing process is accumulation.

Class Groupings of people with similar economic positioning, this impacts on contributions to production processes and on the distribution of returns from production.

Creative Destruction It conveys tendencies in capitalist production to encourage behaviors that lead to the development of the forces of production and consumption. Investment decisions around production, exchange, and consumption simultaneously create new capital and imperil existing capital.

Crises Interruptions in the reproduction of economic, cultural, social, and political life. Capitalism is seen as tending toward crises, from which new bases of accumulation and regulation are developed.

Division of Labor The separation of tasks in the labor process around distinctions such as: product sectors (social division of labor, ship building, designer fashion), production tasks (technical, managers, technicians), gender (different jobs assigned to men and women), and ethnicity (Philippinos as irrigation specialists).

Fictitious Capital Under capitalism some commodities such as land, labor, space, and money are put into circulation as commodities, though they are not produced as commodities.

International Division of Labor Changing international patterns of investment and labor processes arising from developments in capitalist processes, discussed in terms of countries and increasingly regions in different parts of the world.

Labor Process A descriptor for what is worked on, how work is organized, including intellectual and symbolic contributions.

Regulation Capitalism has exhibited different arrangements of social, political, and cultural mediation and organization of economic activity, the arrangements influencing possibilities for accumulation.

Spatial Division of Labor The historically specific patterning of sectors, tasks, gender, and ethnicity in production in specific geographic areas.

Introduction

You are invited to ask the next person you meet after reading this article what they understand capitalism to be. Chances are, you will be quite shocked. In most peoples eyes capitalism is a dangerous word. The freight of politics past and present makes it a very sticky word. It carries so many negative connotations and imaginaries that to say the word is to mark one as either some sort of radical or an old school thinker using outmoded language – and that reputation could stick. In the anxious times of the early twenty first century, few want to talk about capitalism, confront the dense language that accompanies the word, and the contradictions and dilemmas that both advocates and critics hold to be so central to any discussion of the subject. Yet most readers of this article will be living in a capitalist society, or market economy, somewhere on the globe and are probably successful in capitalist terms, though wondering and even deeply concerned about why the global economic and planetary situation appears bleak despite a half century of commitment to making capitalism work.

We argue that capitalism needs to be continually reunderstood since it is implicated in how we shape everyday lives and how our activities seem to have impacts on the many others, near and far, though we may never intend that such outcomes might arise. This is not to suggest that knowing capitalism is sufficient to understand all global woes. Rather it is to propose that as efforts are being made to imagine and develop alternative modes of living on the Earth we must be attentive to developments in capitalism, as the embedded rules of capitalist organization often get in the way of alternative projects. The disposition here is fundamentally geographical. It is an attempt to gain a global sense of what is going on, an awareness to use Nigel Thrift's phrase, derived from considering dimensions of capitalism from our own context, wherever that might be in the world and thinking constructively about how we are connected to others (human and nonhuman and attendant associations) on the planet. Indeed human geographers have played a disproportionately large part in updating the international academy on capitalist developments. They have done so through the lens of chiefly Marxist political economy, at a time ironically when there has been a rapid decline in interest by economists in the field of political economy. In political economy the role of human geographers is dramatically changing. Until recently the narratives from geography have been mostly masculine, privileged, Anglo American, unwittingly capital centric, and not designed to keep an eye on capitalism while also seeking to generate new imaginaries through the use of the rich collective resource of human geography. Now human geographers have available a powerful range of intellectual resources that enable the scrutiny of capitalism to be situated differently and to take place in new and additional ways.

Finding the message lines about capitalist develop ments and devising strategies to communicate them has never been easy for human geographers. This article faces similar challenges. There are at least three stumbling blocks that need to be confronted if we are to get a glimpse of capitalism's inner workings. The first is that there are plural political economies or para digmatic narratives in circulation about capitalism – simplified here as only two – neoclassical and Marxist. This dichotomy is a little forced but the lines of difference between the two are so crucial to under standings about capitalism that it is used to organize the discussion at various points. One unfortunate upshot of this plurality is that particular words have sharply different meanings in the two camps. Another is that efforts to illustrate how the different paradigm stories overlap and intersect have been put aside as too hard. The second problem is that for the visual and increasingly more sensate cultures of this century dia grams help with mapping relationships. Writings on capitalism should meet this demand. Geographers have shown great resourcefulness in pictorially representing conceptual, theoretical, spatial, and historical dimen sions of capitalism and spotting when others have utilized diagrams to simplify and heighten their mes sage. The third is the difficulty of attempting to translate formal and abstract political economy ideas into what it means for everyday human and nonhuman life, research practice, and political and ethical strat egies. This is as much about geographical interven tions, in multiple sites and certainly beyond, as well as in the academy, as it is about teaching and curricu lum. Human geography textbooks contain fascinating and accessible examples that aid learning and pre pare the reader for the heavier stuff of scholarly writing. The response to such challenges should be to reaffirm the importance of some abstraction relating to capitalism and capitalist processes, providing substantive issues are not separated from theorizing. The article is thus primarily about abstractions that relate to three things: the basis under which capital and the compelling interest in profits for profits sake exist and what that means in terms of pressures on behavior of economic entities, the circulation of capital in cluding the moment of increasing globalization, and new dimensions to capitalism that deserve close scrutiny.

Looking back on the past quarter century insight from especially Marxian informed study of the political economy of capitalism, capitalist production, and capitalist labor arrangements greatly influenced the direction, content, and practices of economic geography in the 1970s and 1980s. This flowering of geographical political economy was itself a product of emerging crisis conditions in both the UK and US, countries, homes to sizeable communities of university geographers. Sig nificant achievements included recognition that the built environment and urban processes should be on the agenda of economic geographers, extension of Marxist theorization to include the post World War II long boom and the aftermath of widespread restructuring and exploration of the changing and invidious geog raphies and complexities of uneven development and so on. If there was a conscious strategy by geographers, it was to reveal through comparison with existing geo graphical practice that there was more to geography than the study of the space economy. By the mid 1990s the abstract and theoretical discussion on capitalism had been vigorously challenged and largely rejected as narrow (focusing on the economic), singular (one ac count fits all), and totalizing (shutting out other lines of inquiry) by critics adopting postmodern, feminist and post structuralist, and other framings. Attention had also shifted to globalization and then neoliberalization, arguably two developments whose wider significance is discernible because of new insights by human geog raphers into capitalist processes. A key aspect that has driven the re engagement has been the rediscovery of relational ontologies in Marxian and other philo sophical thought.

Trying to Understand a Changing World

Academic interest in capitalism as a category of social organization stems from attempts to summarize how peoples at different times and places in history have sought to organize activities aimed at ensuring their survival, reproduction, and prosperity when faced with uneven availability of and access to resources. Categories such as capitalism both derive from and encourage politically laden questions either indirectly by contesting the basis of knowledge underlying distributional ar rangements or directly from a politics of emancipation. For example, 'How do or should societies in different territories attempt to use their economic and other ac tivities to achieve agreed or desired outcomes?' 'What are the main mechanisms available to extend existing dir ections or make changes?' 'How do the relationships amongst different actors influence who gets what, where, and when, and what might be done about things?' Regardless of political persuasion a consensus position is that investment decisions are paramount. This is where it gets tricky. While investment is undeniably important, opinions fly to the extreme over how to decide on investment priorities – markets or politics. In the twenty first century we must continue to ask such questions, and be open to the conditions in which the questions are being given voice, and especially by whom, and moreover, what the values and goals of the questioning are. In the revisioning of the academy underway at present human geographers have a responsibility to pose questions about the global scene. Against the backdrop of a century of expansion and elaboration of capitalist processes, how has most of the world's population (human and nonhuman) fared and what futures might they have? Those delving into these concerns argue that since the 1960s international spatial inequalities have become more rather than less pronounced. We know so little about nonhuman populations. Looking forward, the focus is on how humans and nonhumans cohabit in the nexus of bio physical processes. What questions need to be put? What might new questions mean for how we interrogate capitalist processes?

As a noun, the word capitalism tends to imply a fixity and finality of features and form that belies the attempts by researchers and scholars, especially those working in the political economy tradition, to highlight the nature and implications of the dynamics of capitalist processes. Framed in this manner it is amenable to abstraction. Depending on which literatures are read, however, diametrically opposing claims are made and conclusions drawn. This distinction is not trivial. Attempts to investigate economic processes and their institutional settings can be traced back as far as Adam Smith. Schools of thought dominating thinking over the past century have diverged around four main concerns: the degree to which the object of interest is the economy alone (neoclassical economics) or the economy and its social and political foundations (Marxian and Weberian analyses); a contest over the relative importance of legitimacy of profits (neoclassical economics) or needs and inequalities (Marxian and Weberian analyses) as guiding principles for individual, community, and social investment; the extent to which self regulation of markets, industries, firms, and individuals (neoclassical economics) or political processes (Marxian and Weberian analyses) should guide the allocation of resources and distribution of goods and services; the role of choice (neoclassical economics) versus democratic processes (Marxian and Weberian analyses) and; whether inquiry should attempt to make the world look like the model (neoclassical economics with an emphasis on perfecting markets) or whether the focus should be understanding how the world actually works (Marxian and Weberian analyses with its emphasis on acting to change the world).

Contesting Abstractions

Neoclassical Economics

In the abstractions of neoclassical economics, exchange is regarded as equivalent in and amongst all commodity markets, including the labor market. Consumers and producers have equivalent status as markets are paramount. The general structure of commodity exchange in terms of prices and relationships involving price is regarded as sufficient to characterize the whole economy. Buyers and sellers freely enter the frictionless frameworks of markets, armed with perfect rationality and perfect knowledge, respond to price signals generated by demand and supply schedules, to produce a general equilibrium. Note how this abstracted world operates – it tends toward equilibrium. As a consequence, neoclassical thought has privileged modeling (general equilibrium) of pricing and exchange, seeking mechanisms to obtain more efficient market operations on the assumption that the more efficient are markets the better will the economy perform.

So what does abstract capitalism look like when painted through this lens? Conventional economic texts see the world as now largely featuring a capitalist free enterprise system. For example, Paul Samuelson's classic textbook economics portrays (**Figure 1**) the competitive price system of capitalism solving the basic economic problem of what to produce, how, and for whom in an elegantly simple manner. This was spotted by geographer, David Smith who in the fourth Dictionary of Human Geography reflected that the characterization is illustrative of how far the assumptions of neoclassical thought guide the construction of the world and the choice of words to anchor the world. Two groups of participants are recognized – the public (though acting in their own self interest do so without conflict) and business (government is left outside the framing); the public takes its income into the markets for consumer goods and services, expressing its preferences in the form of what Samuelson calls 'dollar votes', an implicit analogy with the electoral process so evoking the principles of democracy in support of free market mechanisms.

Such a characterization is far removed from the conclusions of Marxist political economy. In the neoclassical account the value of anything and everything is established through price, the commodity is innocent, and people in their specific geographical and historical contexts and circumstances do not figure. Equally there is no hint of sharp lines of difference around the rules of engagement in the capitalist market economy, what start up conditions might mean for the terms of participation, and how individual and social outcomes may diverge greatly.

Influences of Marxian Political Economy

In marked contrast to the idea of equilibrium, Marxist political economy stresses the dynamism of capitalism.

Regimes of accumulation	Circuits of internationalization			Nation-state (fractions)	Social relations of production in nation-states		
	Production	*Realization*	*Reproduction*		*Capital (C) C-C*	*Labour (L) C-L*	*State (S) C-S*
Extensive		●		N I-C			
Restructuring crisis			●	N I-C			
Intensive ('Fordist')	●			N I-C M-C (G)			
Restructuring crisis	●		●	M I-C M-C G			
Integrated	●	●	●	N I-C (M-C) G			

Global geography — Spatial division of labor

Geography of accumulation

Figure 1 A model of globalisation. Source: Fagan, R. and Le Heron, R. (1994). Reinterpreting the geography of accumulation: The global shift and local restructuring. *Environment and Planning D: Society and Space* 12, 265 285, slightly modified from Figure 1, A model of globalisation.

The obvious question of course is where does the dynamic come from? This is where abstraction is especially helpful. The notion of a capitalist mode of production, is which a specific form of economic and social organization, conveys the idea that capitalism can be understood as involving particular behaviors in economic organization that have profit seeking as the main socially acceptable goal. While orthodox definitions pick up on any definition as being a construction of political and intellectual traditions, they rarely extend to considering the contextual or place laden shapings and the processual nature of theorizations about capitalism. Karl Marx's legacy was his abstraction of the key social relationships and interactions that were characterizing dramatic changes in the countryside and cities of mid nineteenth century Europe and Asia. This was a geographically bounded visioning (he spent most of his life in Germany and Great Britain). In that context he was able to comprehend both the internal relationships necessary to the circulation of capital and the contingent geographic reach of capitalist processes, at that time.

Capitalist Processes in the Making

Creation and Circulation of Capital

Within the contemporary world economy, the dominant sets of relationships are increasingly capitalist in nature. Further, through rapid globalization a globalizing capitalist economy is emerging. This landscape means abstract understanding of capitalism can be deployed to, first, understand the relational basis of capital's existence and the pressures for particular behaviors that spring from crucial interactions amongst the relationships, and second, the implications of the dynamic of capital circulation that spring from the existence of a globalizing capitalist economy.

In simplified terms capitalism and the emergence of capital originates from specific conditions. David Harvey, the geographer whose work has been hugely influential, phrases a sequence of interlinked conditions and consequences. He is quite sweeping in his purview, from the initial conditions that create capital to geopolitical inferences. His argument, drawing on Marx, is that capitalism is a historical and geographical process that continually creates value, when people enter daily into capitalist relations, though not always of their choosing. Thus, in the first instance direct producers (laborers) are separated from the ownership of the means of production and the ownership of the product of the labor process. This fundamental insight was grounded empirically from examination of the enclosure movement at the time and is still a crucial process occurring in the twenty first century. Direct producers must sell their labor power to capitalists (who own the means of production) being transformed into a commodity. They are bought and sold in a labor market regulated by price signals. Capitalists make every effort to secure a surplus of value from the difference between the costs of hiring labor and using other inputs and the realized value of product sold, so creating surplus value and profits. Realization at a profit means capital – value – is available and potentially in circulation. Capitalists try to retain and expand their wealth by putting capital back into circulation, in order to accumulate. A class relation exists between capitalists (who buy rights to labor power in order to gain a profit) and laborers (who sell their labor for a wage in order to live). The class relation implies conflicts over pay and work conditions. Capitalists seek to be competitive by changing the means of production by developing the forces of production such as technological change, materials improvements, new knowledge, changing labor skills. The organization of production occurs through both the technical and social division of labor. Pressures from intercapital competition and class struggles encourage technological and organizational change, which intensifies work and usually requires investment of capital and labor power. The circulation of capital is unstable as expansion depends upon the application of living labor in production whereas technological change typically displaces workers. Absorbing surplus capital (unused production capacity) and labor (unemployment) is not automatic, leading to crises and political unrest. Surpluses that are not absorbed are devalued, sometimes destroyed. Share markets collapse, inflation runs rife, default rates on debts climb, property prices fall, and so on. The potentialities of capital in circulation are aptly captured in Joseph Schumpeter's famous phrase, creative destruction.

A more user friendly statement of the contradictory position around the investment–workplace relation would highlight professional managers for capitalists and wage and salary earners rather than laborers. Today the prime movers of capital are corporations operating in business, government, and community shares. Harveys's chapter 5 in *The Limits to Capital* gives a glimpse into the constraints and possibilites of corporate operations by fusing understandings from Alfred Chandler's company studies with political economy logic. He shows that contrary to accepted wisdom the atomistic world of perfect competition offers enormous obstacles to capital circulation. He suggests that the divisionally structured organization (this includes the government sphere if run under commercial performance criteria) internalizes comparisons of performance (division to division instead of company to company). External benchmarking and monitoring systems simply widen the net and horizons of calculation. Since these systems are being rapidly adopted knowledge of both relative and absolute positioning of small and large entities, in multiple evaluative frames (e.g., industries, regions, best practices) provide signals on many performance measures. Double entry accounts consisting of income and expenditure and assets and liabilities ledgers keep track of how individual entities fare. In the increasing 'round about ness' of capital circulation, in configurations such as supply chains, many sites exist to generate and appropriate value and profits. Who gains and loses is a timeless and ever geographic question.

Paths of Capital Circulation

Toward the end of *The Limits to Capital* David Harvey diagrams the paths of capital circulation (**Figure 2**). This refreshed Marxian political economy, by putting a set of general theoretical observations into a single figure that both revealed and reframed the research agenda of social scientists, as well as geographers. The shift in register accomplished by the diagram was threefold; it reinstated the importance of circuits as a way to understand the dynamics of capitalist processes, it revealed multiple circuits, and the different temporality to the circuits. It also hinted at geographical dimensions in new ways – adding in the built environment for the first time. Despite its many positives, **Figure 2** still had a productionist and industrial stamp and was highly abstract. But as Harvey would be quick to remind us, Marx said capital is value in motion, and the diagram shows where that value might flow. **Figure 2** can also be seen as a Marxist political economy counter to the neoclassical view seen in **Figure 1**.

Interconnecting and Supportive Processes

Contemporary economic geographers such as Kathie Gibson, Julie Graham, Wendy Larner, Roger Lee, and others have been especially diligent in their attempts to

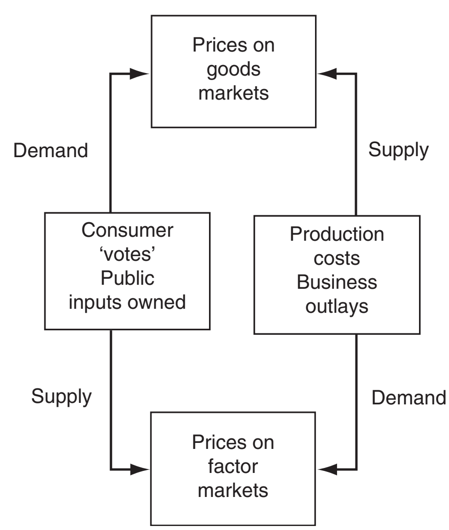

Figure 2 A neo-classical view of the economy. Source: Samuelson, P. and Nordhaus, D. (1989). *Economics*. New York: McGraw-Hill.

outline how capitalism is a constituted process, that is, the development of the key relationships is not in dependent of geographical and historical processes. Re flecting back on the prolific output of geographers inspired in the 1980s–1990s by political economy a number of threads are discernible as significant contri butions to the abstracted understandings of capitalism. The mapping includes:

- Processes that have aided the separation of people from their ability to produce for subsistence, including movements of enclosure, privatization schemes, and the active creation of markets or marketization.
- Processes exposing how capitalist processes are im plicated in alterations in household divisions of labor, the separation of residence and workplace, the femi nization of some areas of work, and changes in the participation rates of men and women.
- Processes connected with unions, labor processes, and labor markets, including under regional and inter national trade arrangements.
- Processes that have developed the forces of pro duction, especially mechanization and industrializa tion in agriculture, manufacturing, and services that have transformed human and nonhuman worlds. One political economy book title from the 1980s read by many geographers gives a feel for this – *From Farming to Biotechnology*.
- Processes that have altered the regulatory and gov ernance frameworks, including the rise of food re gimes and resource import complexes in the globalizing economy.
- Processes that have changed the spatial arrangements of where and how people can live, particularly ur banization, suburbanization, peri urban developments, and city regions.
- Processes that have shifted the focus of infrastructure in economies especially between production and consumption, strengthened by the work of geog raphers interested in cultural economy, including highway systems, industrial and science parks, water front developments, shopping malls, airports.
- Processes that extend the depth of calculation and surveillance, that are increasingly at a distance tech nologies giving new capacities to know competitors.

Regimes of Accumulation and Regulation

The study of the institutional foundations of markets was led early by sociologists, especially Max Weber and Karl and Michael Polyani who emphasized the importance of (1) the legal and political framework established by the nation state that provided unprecedented structured possibilities for capitalist development and (2) the bases upon which actions were decided, which have been steadily developing through successive generations of comparative practices anchored in the principle of ra tionality, for example, audit, standards, benchmarking, best practice, certification, and so on.

These interests have been replayed by geographers drawing on 'the regulation school' from European political economy. This school argues that two distinctive regimes of accumulation with particular regulatory modes had already appeared in capitalism. The first, extensive accu mulation, rested on actors foraging the world for fresh resources and products. The emphasis was on luxury and capital goods for the emergent economies of the UK, Germany, and USA. The second, intensive accumulation, involved mass production to cater for mass consumption stimulated by state assistance. This underpinned the post World War II long boom.

Geographies of Accumulation and Regulation

Introducing accumulation into the picture raises the issue of economic restructuring. Geographers quickly spotted de industrialization, redundant spaces, and job loss in national economies, but it was not until the mid 1990s that the significance of restructuring was addressed in theore tical terms that tied back into capital circulation. Eco nomic geographers working in the rapidly restructuring settings of Australia and New Zealand sought to show how particular episodes of restructuring were implicated in and constitutive of new expressions of capital circulation. The strategy was to devise a schema (**Figure 3**) to illustrate how capital circulation could be simultaneously examined as internationalization, as impacting on capital centered organizations and as reworking established capital–capital, capital–labor, and capital–state relationships. The schema acknowledges that capital moves in three ways: from money to production to commodities (the circuit of pro duction); through their sale at a profit, commodities are realized as money capital (the circuit of realization); money capital is reinvested in expanded production so that capital accumulation takes place (the circuit of repro duction). Capital fractions were introduced for two rea sons, to reveal how the nation state institutionally mediates capital circulation and to disclose the moment, that is, restructuring through state action, when market constrained (M C) and investment constrained (I C) fractions dissolved into global and local fractions. At all times in capitalist history internationalization of capital has been central to the accumulation process. The circuit of realization was internationalized early as foreign trade. In the late nineteenth century foreign investment poured into colonies of European powers. London played a crucial role in these international circuits a century ago. Between World War I and World War II and especially after 1950

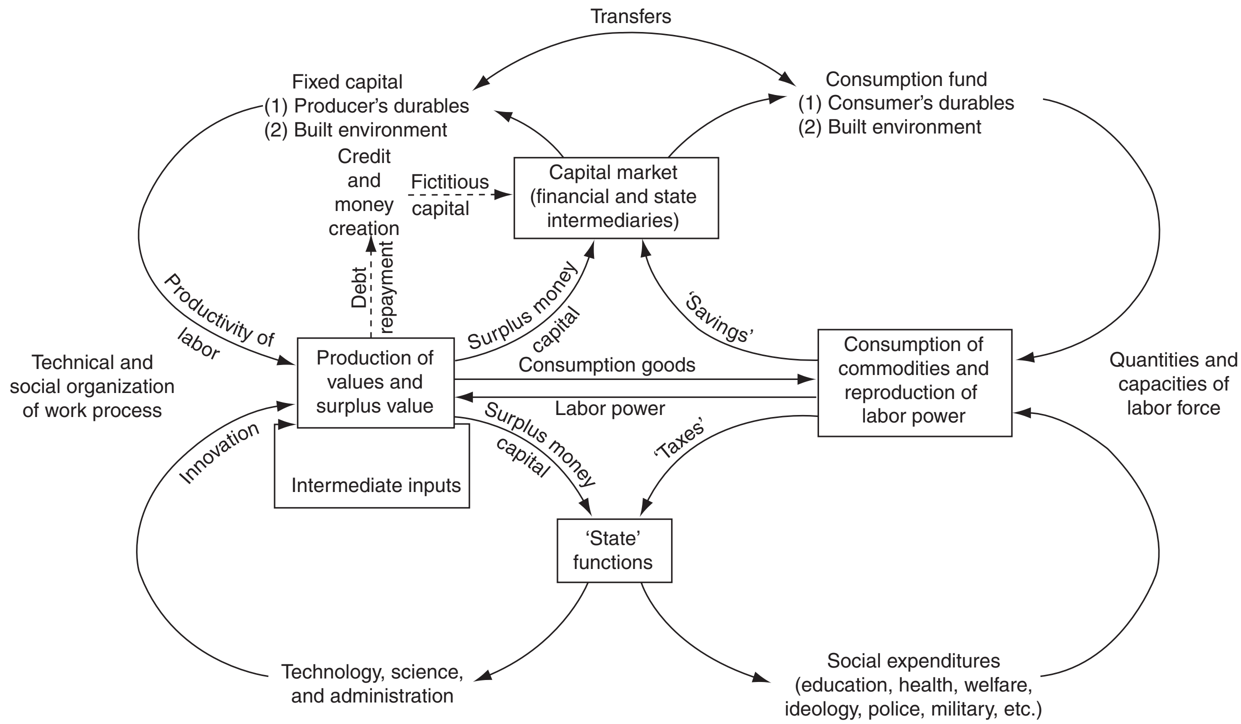

Figure 3 The path of capital flow. Harvey, D. (1982). *The Limits to Capital*. Oxford: Blackwell, Figure 12.1 The path of capital flow, p 408.

offshore location of transnational corporation (TNC) branch plants expanded globally the circuit of production, establishing production systems inside major markets.

The diagram, though depicting relationships in ab stract, provided one strategy for bridging global–local and analyzing the complexities of corporations in very different ways, and visualizing how globalization as a political project was opening up new prospects of accu mulation by all fractions of capital.

Contemporary economic geographers steeped in the political economy tradition would say the creation of value, or a surplus, takes place in and through specific economic spaces. Stated another way, capital always circulates in activities in territory. From this we can move to associations among places and be cognizant of temporality. Geo graphical and historical contexts and conditions matter.

Challenges … and Opportunities

It would be tempting to conclude that the globalization project, with neoliberalization, a technology of con tinuing reorganization of increasingly global connections, amounts to a solution for the crises of capitalism. The position taken here has been to probe what we encounter as we follow through abstraction the emergence of capital circulation. In this section aspects of translation that spring directly and insistently from knowledge of capital circulation have been touched upon.

The DNA of Capitalism

However, where one might ask is the commodity in the story? We owe much to Michael Watts for sharing an astonishing anecdote in *Introducing Human Geographies* (2004). He recounts how he sometimes tosses a cold and clammy chicken onto the lectern, to pronounce after moments of disbelief and jest, that it is a bundle of social relations. As he says, by starting with a trivial thing, the chicken as a commodity for sale, you can end up with a history of post war US capitalism. His way of thinking highlights the industrialization of agriculture as disclosed by the chicken. It repays a read! His intervention is a convention breaker. It goes to the heart of the commodity – the economic DNA or secret code of modern capitalism.

The Headquarters of Capitalism

Nearly a century ago Marxist political economy fore grounded the powerful disciplining effects of finance. The international circulation of massive amounts of money, a form of fictitious capital, continues to gather pace. There is much to watch – the daily rhythm of the London, New York, and Tokyo stock markets; the rise of bonds, deriva tives, options, futures, and other forms of financial in struments; the appearance of mutual funds, linked closely to pension funds; the volatility of floating exchange rates and so on. The crucial issue is the disruptions and havoc that can arrive suddenly, unfathomably, and heartlessly, when financial disciplines bear upon existing and proposed investments. The phenomenon of concentrated national capital (the bane of governments in the twentieth century) is reappearing on an enlarged scale, through TNC mergers and acquisitions. This has major implications for how entities operating under capitalist rules, regardless of size, will approach issues about conditions of existence for life on Earth.

A New Focus – From Capital-Labor to Capital-Life

Nigel Thrift suggests that quite extraordinary changes are occurring in how managers are incorporating workers into capitalist processes. His argument is that labor is only truly productive (i.e., producing surplus value and profits) when the full range of psychophysical powers of the human body are tapped into. He cites a range of evidence indicating efforts to squeeze value out of the system by increasing the rate of innovation and invention, especially through promoting connection. However, the most strategic aspect to Thrift's recent work is his call for a fundamental shift in how we think about capitalism, by replacing the capital labor antagonism with a capital life relation. My concern is that this be a broad shift, to encompass human and nonhuman life.

New Politics – of Post-Capitalist Kinds

A very powerful and illustrative critique of capitalism was mounted by Kathie Gibson and Julie Graham with their 1996 book, *The End of Capitalism* (*as We Knew It*). Their mission was to encourage a complete rethink of political economy, putting more politics into economy and more economy into politics. This was a marked switch away from the abstract. They grounded their critique in everyday circumstances, argued that new relationships and therefore interactions could be created by refusing to engage in capitalist relations. Their project has gathered momentum over the past decade around the metaphor of the diverse economy, something that coexists with capitalist activities, but has not been made explicit until very recently in geographical political economy. In their latest book, *A Post Capitalist Politics*, they show a way with some skilful depictions revealing invisible and alternative economic relationships.

Conclusion

Questions about where and how capital circulates are being revisited with respect to the incorporation of biophysical processes and the capital life relation in the rapidly developing field of political ecology, the implications of financialization for economies, the tensions amongst owners, managers, and workers in neoliberalizing contexts, the contradictions and crises of global property markets and affordable housing, the commodification of reproductive processes (e.g., egg and sperm banks), food quality, safety, and security, to which human geographers are contributing. The abstract knowledge available about capitalism is a crucial resource for empirical research into the specific expressions of capitalist processes in different geographical contexts and historical conditions.

See also: Firms; Fordism, Post-Fordism and Flexible Specialization; Foreign Direct Investment; Globalization, Economic; Neoliberal Economic Strategies; Neoliberalism and Development; Political Economy, Geographical.

Further Reading

Fagan, R. and Le Heron, R. (1994). Reinterpreting the geography of accumulation: The global shift and local restructuring. *Environment and Planning D: Society and Space* 12, 265 285.

Gibson Graham, J. K. (1996). *The End of Capitalism (As We Knew It)*. Oxford: Blackwell.

Gibson Graham, J. K. (2006). *A Postcapitalist Politics*. Minneapolis, MN: University of Minnesota Press.

Glassman, J. (2006). Primitive accumulation, accumulation by dispossession, accumulation by 'extra economic' means. *Progress in Human Geography* 30, 608 625.

Harvey, D. (1982). *The Limits to Capital*. Oxford: Blackwell.

Harvey, D. (1985). The geopolitics of capitalism. In Gregory, D. & Urry, J. (eds.) *Social Relations and Spatial Structures*, pp 128 163. New York: St. Martin's Press.

Jessop, B. (2000). The crisis of national spatial temporal fix and the tendential ecological domination of globalising capitalism. *International Journal of Urban and Regional Research* 24(2), 323 360.

Johnston, R., Gregory, D., Pratt, G. and Watts, M. (eds.) (2000). *The Dictionary of Human Geography*. Oxford: Blackwell.

Larner, W. and Le Heron, R. (2002). From economic globalization to globalizing economic processes: Towards post structural political economies. *Geoforum* 33(4), 415 419.

Le Heron, R. (1993). *Globalised Agriculture: Political Choice*. Oxford: Pergamon.

Massey, D. (1985). *Spatial Division of Labour*. London: Macmillan.

Murray, W. (2006). *Geographies of Globalization*. London: Routledge.

O'Neill, P. and Fagan, R. (2006). Geographical takes on three decades of economic reform in Australia. *Geographical Research* 44(2), 204 219.

Smith, N. (1991). *Uneven Development: Nature, Capital, and the Production of Space*. Oxford: Blackwell.

Thrift, N. (2006). Reinventing invention: New forms of capitalist commodification. *Economy and Society* 35, 279 306.

Walker, R. (1985). Is there a service economy? The changing capitalist division of labor. *Science and Society* 49, 42 83.

Capitalism and Division of Labor

M. Brayshay, University of Plymouth, Devon, UK

Glossary

Agrarian Capitalism An intermediate stage of capitalism occurring in late-Medieval England that laid foundations for the emergence of merchant capitalism and, later, industrial capitalism. Feudal tenancies were replaced by leases at costs that obliged peasant farmers to specialize and produce competitively for the market.

Capitalism A system where goods and services are produced for profitable exchange in a free market. The means of production are privately owned, usually by independent companies operating competitively, and accumulated capital is reinvested in further production. Goods and services are produced by labor working for wages, and money is used to mediate the processes of distribution and exchange.

Circulating Capital It refers, for example, to raw materials or intermediate products, as one of the two types of capital among the factors of production. The other factors of production are land, labor, and entrepreneurship (see fixed capital).

Commodification The process whereby every element in an economic system is assigned a price and is able to be bought and sold.

Communism (see Marxism) A highly centralized political system and centrally planned command economy in which private enterprise and free trade are eliminated, and agriculture and industry are collectivized. Communist regimes have also aimed to acquire military power as a bulwark against the perceived threat of the capitalist world.

Comparative Advantage The theory developed to explain why it can be deemed beneficial for two countries to engage in trade even though one of the parties might be easily capable of producing all the goods involved itself.

Division of Labor The 'functional' subdivision of production processes into a series of separate operations, each carried out by workers possessing relatively few skills, whose wages cost less than those of fully skilled individuals capable of producing the entire product. It also means the 'geographical' subdivision of production whereby the components of a particular product are made wherever efficiency and cost effectiveness can be maximized and subsequently brought together for final assembly. A 'gendered' division of labor is a wider concept covering both socially determined gender roles within societies, and the assignment of particular kinds of employment to men and women within an economy.

Fascism/Nazism Right-wing authoritarian, centralist political philosophies promoting militaristic nationalism, which favor strong state economic control and regulation, and are suspicious of Western liberal capitalist democracy.

Feudalism It describes a range of socioeconomic systems, generally comprising intricate linkages between lords, vassals, and land resources that existed in Medieval Europe as well as elsewhere in the world.

Fixed Capital It refers, for example, to the plants or factories where goods are made, as one of the two types of capital that forms a factor of production (see circulating capital).

Fordism A form of industrial capitalism (taking its name from Henry Ford's pre-World War I Detroit motor car factory) that came to dominate Western economies, especially between 1945 and the 1970s, characterized by mass production through the assembly line, relatively high industrial wages leading to (through the virtuous circle) mass consumption.

Free Market All economic decisions regarding the buying and selling of goods, services, and money take place voluntarily and are free from any coercive government influence. Individuals trade, bargain, cooperate, and compete in an unfettered and decentralized manner, and the role of government is to defend market freedoms and the rights of an individual to determine the price at which he or she will sell their property.

Futures Trading It involves speculating on the future price of derivatives (financial instruments based on the value of shares, bonds, currency relativities, and commodities). If a trader expects a price will rise, then a fixed-term 'future' in it will be purchased, which means that when the term expires the original agreed purchase price will fall due. Assuming prices have risen, resale at the new prevailing value will yield a profit to the trader. However, if prices fall, losses will be incurred. 'Options' on futures are more sophisticated deals that allow traders flexibility about when to commit themselves to a futures deal. While futures can be a means of reducing uncertainty and risk, they have become a form of 'casino capitalism' comprising more or less informed guesses about forthcoming movements in prices.

Gentlemanly Capitalism It describes the small elite group of London-based individuals who managed investment in, and the running of, Britain's colonial economy from the late 1600s until the 1930s. The theory identifies financial capitalism as the key driver and determinant of British colonial activity and imperialism.

Imperialism The creation, extension, or maintenance of political and/or cultural, and/or economic control of territory beyond that of the imperial power. The term has acquired a pejorative, negative meaning and it is important to note that, like colonialism, different connotations were implied by using the term in the past.

Industrial Capitalism The development of the so-called factory system of production, characterized by complex and intricate divisions of labor, both within and between production processes, associated strongly with the proletarianization of the laboring classes. The industrialist takes the place of the merchant as the key player in the capitalist system.

Islamic Revival Complex and fundamental cultural, philosophical, political, and ethical objections to those aspects of the Western way of life that are regarded as degenerate. Includes opposition to the conspicuous consumption and profligate spending of the super-rich and the stark social inequalities that pervade some of the strongest Western democracies and offend Muslim teachings on charity and social justice. There are also Muslim ethical objections to Western bank interest charges that have led to the development of Islamic banking.

Keynesian Economics The ideas of the English economist, John Maynard Keynes (1883–1946), that capitalism is compatible with government economic intervention regarding taxes, interest rates, market regulation, and regional policy aimed at promoting higher levels of demand, countering unemployment, and combating deflation.

***Laissez-Faire* Economics** It describes a system in which individuals trade, bargain, cooperate, and compete in a free and unregulated manner. Governments defend market freedoms and safeguard the rights of private property ownership.

Means of Production Resources that are necessary for any form of production to occur. Usually taken to include money capital, land, raw materials, machines, labor, knowledge and information, and entrepreneurial acumen.

Merchant Capitalism/Mercantilism A phase of capitalist evolution in which merchants engaged in a spatially expanding network of trade in staple commodities and manufactured goods (the latter still mostly produced by artisans and craftsmen). Mercantilism involved state controls, subsidies, and monopolies that regulated and protected the activities of merchant traders.

Mixed Economy Systems in which some of the means of production and distribution are state controlled and operated primarily as public services rather than for profit, but other parts of the economy are in private ownership and operation.

Monetarism The theory developed by the US economist, Milton Friedman, that the health of the economy of a country is determined by the amount of money in circulation.

Opportunity Costs Costs of producing an item or a service in terms of the opportunity foregone (or the benefits lost) of not producing the next most valuable alternative.

Post-Fordism Emerging since the 1970s, this form of production emphasizes flexible production, often made possible by the development of versatile, programmable machinery, capable of responding rapidly to changing consumer demand and market conditions.

Proletarianization The process whereby feudal peasants became landless wageworkers within the capitalist system.

Proletariat Those obliged to sell their labor in return for wages in order to earn a living within a capitalist system.

Stagflation Economic conditions in which high levels of price and wage inflation coincide with abnormally low levels of economic growth, resulting in high rates of unemployment.

Transnational Corporations Very large companies that run plants and offices in several different countries and often take business decisions and make profits on a global basis.

Welfare State Governments levy taxes in order to raise revenues to pay for a large measure of social provision such as unemployment benefit, healthcare, and pensions for the elderly.

World Systems Theory A dynamic structural explanation of the evolving spatial characteristics of the capitalist world economy in which a fluid mix of core, peripheral, and semi-peripheral territories is defined. Core regions, with higher wages, advanced technology, and diversified production, exploit both the semi-periphery and the periphery. With low wages, simpler technology, and limited production, peripheral regions are economically subordinated. Semi-peripheral regions are capable of exploiting the periphery, but are themselves exploited by the core.

Introduction

Capitalism is a socioeconomic system in which goods and services are produced for profitable exchange in a free market. The means of production are mostly in private ownership. As capital is accumulated, it is reinvested in further production. Investment decisions are usually made privately, and production and distribution are mainly controlled by independent companies that operate in competition with each other. The work required to pro duce goods and services is supplied by a labor force working for wages, and money is used to manage the distribution and exchange of goods, services, and labor. The division of labor refers to the functional subdivision of production processes into sequences of separate oper ations, each carried out by workers who possess relatively few skills, and whose wages cost less than those of fully skilled individuals who are capable of producing the entire product. Though discernible earlier, the adoption of a division of labor was a particular feature of the factories and mills of eighteenth century Britain. Moreover, its principles were ultimately adopted most fully in the Fordist assembly lines of industrial capitalism devised in Detroit before 1914 but which became dominant in Western economies especially in the period from the end of World War II until the 1970s. In more recent times, the term division of labor has also been employed to describe the geographical phenomenon whereby the components of a product are manufactured in plants in widely dispersed localities.

In the process of supplying the requirements of so ciety, capitalism relies on the efforts of wage earning workers to yield profits for the owners of capital. Theo retically, the two groups – the capitalist entrepreneurs and the wage earning workers – are separate; in practice, however, a clear separation is rare. The capitalist system is, nonetheless, a highly distinctive approach to the supply of society's material needs in which the core ob jective is the production and self expansion of capital; its operation has given rise to the creation of recognizable geographies of production, consumption, and exchange. Human geographers study the processes of capitalism in order to understand and explain its social, economic, and political impacts and the ways in which its related landscapes are created, used, and changed.

While they are undeniably dominant in the world today, institutionalized capitalist economic practices are in fact a relatively recent phenomenon. Thus, although some features of the system are discernible in the ancient world, and there are still strongly contested academic debates about precisely when capitalism evolved, most authorities believe that it began to emerge in the early post Medieval period as feudalism declined. Moreover, it is suggested that capitalism developed first in England. However, it soon appeared in other countries of Western Europe. Thriving merchant capitalism – focused on in vestment in trade and centered on cities such as London, Amsterdam, Barcelona, Venice, and Genoa – emerged as a major feature of the economic geography of early modern Europe. By the eighteenth century, trade, col onization, and imperialism had spread this form of cap italism to the rest of the globe. Moreover, by then, industrial capitalism had developed in Europe sowing the seeds of further, far reaching global impacts. Indeed, notwithstanding some strong ideological challenges, capitalism has profoundly changed the world and its in fluence has become pervasive.

Theoretical and Conceptual Ideas

The Human Imperative to 'Truck, Barter, and Exchange'

Among many writers who have offered theoretical ex planations for the workings of capitalism, the first was Adam Smith, the Scots born philosopher and historian. Largely as a result of his 1776 book, *An Inquiry into the Nature and Causes of the Wealth of Nations*, Smith is regarded as the intellectual father of classical economics. He famously believed in the innate human imperative to 'truck, barter, and exchange' and he developed the theory of the invisible hand of the market. Production and trade take place, argued Smith, in order to secure a profit, and surplus profits are accumulated as capital. Smith said that while people are naturally driven by a selfish desire for gain, ultimately the 'good of society' is better served by the efforts of self interested individuals than it is by al truistic collective action aimed at securing the 'common good'. Best known for his 1798 *Essay on the Principle of Population*, the political economist Thomas Malthus broadly shared the ideas of Adam Smith. Malthus focused primarily on consumer demand in relation to the finite supply of resources and, in his 1803 *Principles of Political Economy*, he also argued that the adoption of capitalist approaches in agriculture provided a floor in living standards that was absent in traditional peasant farming economies. Other theorists of the late eighteenth and early nineteenth centuries include David Ricardo, John Say, and John Stuart Mill who each published studies of the workings of capitalist economies. Ricardo's 1817 book, *The Principles of Political Economy and Taxation*, de fined fixed capital (e.g., a factory or mill where goods are made) and circulating capital (e.g., raw materials or intermediate products). Together, these types of capital represent one of the four factors of production; the others being land, labor, and entrepreneurship.

The Commercialization Model

Arising from the ideas of Adam Smith and others, the orthodox explanation for the emergence of capitalism – sometimes called the commercialization model – maintains that rational, self interested individuals have been engaging in trade since prehistoric times. If anything, therefore, systems such as feudalism were aberrations rather than the norm. Capitalism (re)emerged as trade became increas ingly specialized, technical advances were achieved that yielded improved productivity, and there was an evolving division of labor. Classical economic theory sees the capitalist system as the maturing of long embedded commercial practices and their liberation from political and cultural constraints.

Commodification, the Market, and Comparative Advantage

Commodification is integral to a capitalist system. According to Immanuel Wallerstein, the historical emergence of capitalism involved the "widespread commodification of processes – not merely exchange processes, but production processes, distribution processes, and investment processes … in the course of seeking to accumulate more and more capital, capitalists have sought to commodify more and more of these social processes in all spheres of economic life". In short, capitalism involves the commodification of 'everything': all the ingredients in an economic system are capable of being assigned a price and traded.

Markets have, of course, existed throughout human history. The earliest hunting and gathering societies evidently operated a bartering system. However, the introduction of money in antiquity facilitated more sophisticated trade, which then developed its full potential across Europe in the Middle Ages. Trade was highly regulated and restrictions prevented the emergence of truly free markets. A free market, in which all economic decisions regarding transfers of goods, services, and money are made voluntarily, is generally considered to be an essential feature of capitalism. Ideally, individuals trade, bargain, cooperate, and compete in an unfettered and decentralized manner, and the key role of government is not direct intervention or regulation, but simply one of defending market freedoms. Thus, in a capitalist system, the private ownership of property is both acknowledged and legally protected, as is the right of owners to sell their goods and property at a value which they determine. Such an approach is known as *laissez faire* economics.

Patterns of trade occurring in a free market can reflect a recognition of comparative advantage. The theory of comparative advantage was first propounded by David Ricardo in 1817 when he demonstrated the idea by examining wine and cloth production in Portugal and England. Both commodities could be produced in Portugal more easily than in England. However, the relative cost of production of the two goods also differed within each country. In England, wine production was extremely difficult but cloth could be manufactured more readily. It could therefore be said that there were high opportunity costs for England to produce wine and it was better to focus on cloth making. Although in Portugal it was easy to produce both commodities, there was a comparative advantage in concentrating on the production of a surplus of wine to meet England's demand, and trade it for English surplus cloth. Specialization enabled economies of scale to be secured and investment meant a high quality product could be delivered. England also benefited from the arrangement because, while the cost of its cloth production remained the same, it enjoyed access to a supply of wine produced more cheaply than was possible at home.

Critics point to inherent weaknesses in Ricardo's theory such as its failure to factor in transportation costs and the implied assumption that the advantages of increased production outweigh other problems including environmental impacts and social inequalities. The theory also assumes that capital is immobile. In reality, political restrictions on the flow of capital countered the tendency of entrepreneurs to invest in neither wine nor cloth manufacture in England (where both cost more to produce) and instead to focus their attention only on Portuguese manufacturing. In the absence of such controls, and in situations where the balance of competition is perfect, it is argued that there is a tendency toward absolute advantage whereby one country, or one entity, is able to produce more efficiently than all others.

World Systems Theory

Wallerstein's world systems theory has provided a highly influential dynamic structural explanation of the evolving spatial characteristics of the merchant and industrial capitalist world economy. He identifies a core, a periphery and a semi periphery. Between the 1600s and the late 1800s, regions such as northwest Europe maintained a dominant core position, while others, such as Africa, remained in the periphery. Other regions, for example, the United States, moved from the semi periphery into the core. In other cases, for example, the Balkan states that had emerged by the nineteenth century as the Ottoman empire collapsed, the move was from semi periphery to periphery. Core regions, where wages were consistently higher, technology was more advanced, and production was diversified, exploited both the semi periphery and the periphery. In peripheral regions, wages were low, technology remained simple, and production was limited in range. Between the two, semi peripheral regions exhibited a mix of the characteristics of the other two regions and while capable of exploiting the periphery, they were themselves exploited by the core.

The Division of Labor: Entrepreneurs and Proletarianized Wage Earners

A feature of the demise of feudalism, especially in later Medieval England, was the emergence from the small scale farming peasantry of what some have termed an intermediate stage of agrarian capitalism. Within this stage, two groups may be identified. First, as old feudal laws and customs – under which tenants had once held land from their lord – gradually decayed, they were replaced by economic leases that reflected market conditions. A new lease holding or land owning peasant farmer class was thereby established which was obliged increasingly to respond to market imperatives, to specialize and to produce competitively. Second, those unable to lease or purchase land formed a property less peasantry – a proletariat – that must sell its labor in order to secure a livelihood. The labor that an individual offered became a commodity that could be precisely evaluated and built into calculations about the costs of production and the price at which products must be sold to yield a profit.

Determining price levels is central to the capitalist system. In a free market, prices depend on the supply of, and demand for, all the means of production (i.e., the resources required for any form of production). In order to maximize profits, entrepreneurial capitalists seek to minimize the costs of production and thereby outcompete other producers. Wages will therefore be subject to downward pressure.

The division of labor, whereby production processes are reduced to a series of simple, repetitive operations, performed by unskilled, low paid workers, offer a means within capitalist systems of lowering production costs. The factory system that developed rapidly in Britain in the eighteenth century was peculiarly well suited to this kind of division of labor. Moreover, in some circumstances, tasks requiring the least skill might be assigned to female labor (one example of the gendered division of labor) or child employees, for whom rates of pay could be set even lower, giving a further opportunity to the owners of capital to secure an enhanced competitive edge in the market.

Marxist Ideas and Interpretations

The notion of the commercialization model at the heart of classical economic theory that the capitalist drive is somehow a natural human trait, and that the emergence of capitalism as the world's dominant socioeconomic system was inevitable, is strongly rejected by several theorists. Most prominent among these is Karl Polanyi whose challenge, articulated in his 1944 book *The Great Transformation*, to Adam Smith's concept of 'economic man', and his natural tendency to trade and accumulate capital, is very well known. Today's Marxian economists argue that the expansionary drive of capitalism was not a result of an innate human trait, but the product of its own historically specific internal laws of motion. Such critiques depend crucially on the founding body of radical doctrine, commentary, and critical analysis concerning capitalism developed by Friedrich Engels and Karl Marx, which the latter articulated in *Das Kapital* (1867).

Marx regarded capitalism as a mode of production characteristic of a historically specific phase occurring between the decline of feudalism and (what he saw as) the inevitable proletarian revolution that would usher in communism. He argued that the extraction by the bourgeois owners of capital of the surplus value of goods yielded as a result of the efforts of the labor force would ultimately trigger a power struggle in which the laboring classes were bound to prevail. The sequence that Marx predicted depended on his theory of value, which asserts that labor is the sole source of all wealth and profit – an idea which is, however, strongly contested by today's mainstream, and even some Marxian, economists.

Gentlemanly Capitalism

Interconnections between colonialism and imperialism and the development of the capitalist system have attracted considerable academic attention. London and, to a lesser extent, Paris became the preeminent centers of colonial finance, investment, and marketing. The twin iconic symbols of London's dominance were the Royal Exchange (founded in 1565 by Sir Thomas Gresham at the junction of Cornhill and Threadneedle Street) and the Bank of England (set up by Scots born William Paterson in 1694) (**Figures 1** and **2**). European manufactured goods were traded for primary products and for the output of plantation agriculture such as sugar, cotton, rice, and tobacco. Plantation agriculture depended on slave labor drawn from west Africa. By 1750, as a result of the impetus of capitalism, nearly 4 million Africans had already been taken to the Americas, thereby creating both new colonial economic geographies and a permanent reconfiguration in the cultural geography of two continents.

The notion of gentlemanly capitalism recognizes the pivotal role of economic drivers in European colonialism and imperialism. The term describes the small elite group of influential, London based individuals who managed investment in, and the running of, Britain's colonial economy from the late seventeenth century until the interwar years of the twentieth century. Gentlemanly capitalists controlled the finance and services sector, which are seen as the vital focal point of British imperial power and were the crucial underpinnings that linked the UK with the rest of the globe. Comprising a blend of the old landed elite and the *nouveaux riches* of the city of London, cemented by marriage connections, old school tie links, and later by membership of the same social and sporting clubs, the gentlemanly capitalists exerted immense power and influence across the rest of the world. As a radical revisionist theory of imperialism and capitalism, the thesis has proved highly controversial, but its importance lies in the emphasis it places on finance capitalism as the key motivation for, and determinant of, imperialism.

Figure 1 London's first exchange was founded on the site by Sir Thomas Gresham in 1565 to serve as a center of commerce for the city. The original building was based in design on the Antwerp bourse. When officially opened by Queen Elizabeth I on 23 January 1571 it was accorded the title Royal Exchange: a name that perfectly encapsulates the mercantilist philosophy of capitalism of which England was then a major center. The Great Fire of London in 1666 destroyed the first Royal Exchange and a second was built in 1669 to the designs of Edward Jarman. This building was also destroyed by fire in 1838 and the replacement, opened by Queen Victoria in 1844, was designed by Sir William Tite. Though not used as a center of commerce since 1939, London's Royal Exchange is nonetheless an iconic symbol of Western capitalism.

Figure 2 The Bank of England, located in London's Threadneedle Street (and referred to as 'The Old Lady of Threadneedle Street'), is the central bank of the United Kingdom. Up to 2006, there have been 119 governors. Among the most notable was Montagu Norman whose tenure as governor (1920 44) saw the bank's decisive shift away from commercial banking toward a role as a nationalized central bank. Two years after Norman's governorship, the bank was nationalized. Founded by William Paterson in 1694, the bank first occupied a location at Walbrook and moved to its present location only in 1734 when a new building was designed by Sir John Soane. The present early twentieth-century neoclassical building was designed by Sir Herbert Baker and its facade is one of the world's most familiar images representing Western capitalism.

Geographies of Early Capitalism

Within the context of the theoretical and conceptual ideas outlined so far, it is now appropriate to examine the historical development of the capitalist system and the division of labor in the period since the late Medieval decline of feudalism.

From Feudalism to Capitalism

Though the term itself was coined only in the seven teenth century, feudalism describes a range of Medieval socioeconomic systems comprising intricate linkages between lords, vassals, and land resources that existed in Europe as well as in other parts of the world. Lords were nobles who held land, and vassals were granted access to portions of that land in exchange for military service and meeting other obligations. The king was the supreme lord who granted the land of his realm to the nobility in return for their fidelity, counsel, and willingness to an swer a royal call to arms. Feudal vassals were permitted to collect revenues generated by the land that they held. The nobility granted portions of land to lesser lords and they, in turn, gave protection to peasants who, in return for labor services and rents (usually a combination of a sum of money and some agricultural produce), were allowed access to parcels of land for farming. The system of land grants in return for defined obligations was thus repeated from the apex to the base of the social hierarchy. While the roots of the decline of European feudalism may be traced to the 1200s, as will be shown below, its demise was slow. Importantly, in England, the vacuum left by the gradual decay of feudalism was filled by nas cent forms of modern capitalism. As noted already, from England capitalism spread to other parts of Western Europe and to the rest of the world. However, there is still considerable debate among scholars regarding the precise timing, the location, and the transition processes that saw its emergence.

A key question is why feudalism declined at all. It began as a social contract but, over the course of time, feudal lords could no longer provide new lands to their vassals and the enforcement of their right to reassign land diminished as holdings became *de facto* hereditary prop erties. Moreover, from the thirteenth century, the econ omies of countries in Western Europe began gradually to change from predominantly agrarian systems to ones that were increasingly money based and mixed. While land remained a principal source of income, and defined an individual's social status, even the high nobility de veloped a need for liquid capital – to pay for luxury goods or to fund wars against their enemies. The process of change was boosted by the after effects of the Black Death that decimated the population of Europe in the late 1340s and further undermined the social and cultural controls of feudalism. Indeed, for most scholars, the fourteenth century marks the end of the true feudal age of dispersed lordship thus paving the way for the de velopment in the 1500s of strong monarchies, nation states, and centralized control.

Merchant Capitalism and Colonial Expansion

Although some authorities suggest that, as noted already, an intermediate 'agrarian' form of capitalism emerged in England as feudalism decayed, for writers like Max Weber, Fernand Braudel, and especially Henri Pirenne, the key development in later Medieval Europe was the growth of largely self governing cities and towns devoted to manufacture, commerce, and trade. In these towns, an increasing and autonomous bourgeoisie managed the development of merchant capitalism and this was the essential prerequisite for the rise of the modern capitalist system. Mercantilism or finance capitalism involved the investment (or 'adventuring') of capital in a spatially expanding network of trade in both staple commodities, such as timber and corn, and in an increasingly broad range of manufactured goods. The latter were still, however, mostly those made by traditional handicraft and artisan methods. Between the later fifteenth and the eighteenth centuries, the growing merchant capitalist class – at first dominated by the Portuguese (strongly backed by Venetian and Genoese investment) and the Spanish, but soon superseded by those of England, The Netherlands, and France – sought new markets and made geographical discoveries well beyond Europe. The col onization of the Americas and the development of col onies for the purpose of conducting trade with India, the East Indies, and the Far East were tangible geographical expressions of merchant capitalism.

Many companies of investors, often obtaining royal charters that recognized their right to trade in particular regions of the world and notionally afforded the pro tection of the state, were formed in Europe during this period. Among the most famous are the Dutch and the English East India companies. The latter secured its charter in December 1600 and, by the end of 1606, the company had financed two expeditions to the East Indies that returned with cargoes of pepper, nutmegs, and cloves. Although one of the company's four ships was lost at sea when returning from the second trip, when the spices were sold, the shareholders had nonetheless se cured a 95% profit on their initial investment. This basic premise of financial investment whereby shareholders organized in companies staked their money in trading enterprises in the hope and expectation of substantial profits lay at the core merchant capitalism.

Protected by state controls, subsidies, and monopolies, merchant traders were able to amass capital on a spec tacular scale. Contemporary supporters of mercantilist doctrine strongly emphasized the importance of the state as both the regulator of the economy and the source of charters that accorded monopoly trading rights and offered state protection against foreign competition. Overseas conquest and the acquisition of colonies were, moreover, regarded as legitimate and indeed essential ambitions for mercantilist economies. Colonies were valuable sources of raw materials and lucrative markets for finished products. By the early eighteenth century, Britain's colonial activity was beginning to eclipse that of her rivals and London was the primate city of the world capitalist system.

Although the spread of mercantilism across Western Europe is easy to understand, it has proved more of a challenge to explain why the leap to a modern industrial capitalist economy first occurred in England. Mercantilism lasted much longer in some of Europe's less well developed economies (e.g., Prussia, part of today's Germany), but it was declining in England by the mid eighteenth century. Factors cited to explain England's lead toward industrial capitalism include the centralized character and political unification achieved by the sixteenth century, the Reformation and suppression of the monasteries that saw the enormous transfer of the church's land and political influence to secular interests, and the emergence of a new social order from the seventeenth century conflict between Crown and Parliament during the Civil War and Commonwealth, Restoration and Glorious Revolution. By 1700, the capitalist bourgeois classes had triumphed and were therefore able to determine the running of England's economy. Moreover, a further confluence of conditions allowed industrial capitalism to emerge. Most significant among the further conditions was the availability of the wealth accumulated during the merchant phase of capitalism which could now be invested in mechanized production.

Industrial Capitalism and Its Geographies

Industrial capitalism saw the rapid development of the factory system of production, characterized by those complex and intricate divisions of labor, both within and between production processes, to which reference has already been made. This process, which led to the proletarianization (and, he argues, exploitation) of labor, is explored in depth in E. P. Thompson's classic study, *The Making of the English Working Class*. Thompson shows that in this new turn of the economic revolution, industrialists replaced merchants as the key players in the capitalist system. Handicrafts and artisanship, the powers of the urban based regulatory guilds, and the traditional networks of masters, apprentices, and journeymen were all eclipsed. The ethos of industrial capitalism also began to challenge mercantilist doctrines of trade barriers and protective monopolies and, by the mid nineteenth century, Britain had fully embraced *laissez faire* economics. A clear manifestation of this new thinking occurred in the 1840s when the duty on grain imported to Britain was abolished by the repeal of the Corn Laws. Liberalism and competition in trade, and the development of a free market economy, were now the central political and economic philosophies of the era.

The capitalist transformation of Europe's textile production from the fifteenth century onwards has received the bulk of academic attention devoted to explanations of the emergence of industrial capitalism. Studies have focused on the increasing mechanization of the manufacturing process and its spatial concentration to enable the new technologies to be fully and efficiently harnessed. Geographers note the apparently relentless tendency toward regional specialization and concentration in the textile industries that peaked in England during the later eighteenth century. Although the manufacture of some products remained relatively domestic in scale well into the twentieth century, others (china and pottery, iron and steel making, brewing, etc.) were transformed in a similar way. The landscape impacts of the changes in production methods and geographical location were profound. Steam powered technologies – hugely enhanced by the engineering inventions of Matthew Boulton and James Watt in the early 1770s – inexorably prompted industries with high energy demands to concentrate on or near coal fields; and, in these areas, in addition to the construction of mills and factories, settlements dominated by high density industrial housing grew rapidly and the transport infrastructure was transformed as new networks of canals and turnpike (i.e., toll) roads were built. While traditionally characterized as the specific manifestations of the industrial revolution, such geographical transformations may also be seen as the tangible, virtually indelible landscape signature of the history of the wider capitalist system both in Britain and throughout much of the world.

By the nineteenth century, Britain was seen (and saw herself) as the workshop of the world, but the driving forces of industrial capitalism, comparable to those experienced in the United Kingdom, were also exerting an impact on landscapes in other parts of Europe and North America. However, the processes involved were complex, varied, and geographically uneven and there has been much debate about the precise timing and character of the transformation in the case of particular industries and places. Parallels with the British model of change have been drawn, for example, with the development of heavy industry in Germany's Ruhr region and in western Pennsylvania in the United States.

Steam technologies underpinned the invention in the early nineteenth century of steam powered locomotive engines and the development of railways. The impact of the introduction of railways was dramatic in Europe and, arguably, even more so in North America and in imperial colonies. Railways made possible the integration – politically, culturally, and economically – of vast continental spaces. The British began to build railways in India in the 1850s. In North America, the first transcontinental railroad opened in 1869. The transport costs of raw materials and finished products were significantly reduced, new sources of supply became accessible, and new markets were opened up.

Geographies of Modern Capitalism

Though progress has been neither uninterrupted nor straightforward, the period since the last quarter of the nineteenth century has seen the global triumph of capitalism. Indeed, the pace of change accelerated markedly during the twentieth century: the transformation of Western economies and societies that occurred during the 1900s as a result of immense technological advances is utterly without precedent. But there have also been major setbacks and catastrophes. Though providing a short term economic stimulus while conflict ensued, the wake of both the world wars of 1914–18 and 1939–45 saw severe disruption. Indeed, the global economic depression between 1929 and 1933 put a sharp brake on progress. Notwithstanding severe post war dislocation, after 1945, the motor of capitalism soon burst into renewed life and growing prosperity lasted, almost uninterrupted, until 1973. Alongside, and extending beyond, the catastrophic impacts of war and economic slump, capitalism also faced other mighty challenges in the 1900s: the end of empire, the rise of communism and fascism, and (more recently) the Islamic revival. It is to this more recent period in the history of capitalism, namely that since the last quarter of the nineteenth century, that attention is now turned.

Large Corporations, Political Regulation, Fordism, Mass Production, and Economies of Scale

In the last quarter of the nineteenth century, the control and direction of large areas of industry came increasingly into the hands of banks and financiers. This era has thus come to be dominated by finance capitalism, characterized by the growth of large corporations – some of them developing into mammoth industrial cartels. The development of increasingly complex systems of banking and equity markets, the corporate holdings of capital through stock ownership, and a widening separation between the owners of capital and the production process are further key features of this phase of capitalism. The emergence of large industrial trusts provoked legislation in the USA aimed at reducing the monopolistic tendencies then prevailing. The US federal government thus assumed a widening role in this period in the regulation of industrial standards for key industries of special public concern and in the passing of antitrust laws. Such regulation in the late 1800s and early 1900s marked a move away from *laissez faire* economics.

In 1913 Henry Ford famously introduced the assembly line in his automobile manufacturing plant at Dearborn, Detroit. Drawing on both the earlier experience of highly organized working methods adopted in the Chicago meatpacking industry and the time and motion ideas of F. W. Taylor, Ford took the concept of the division of labor to a new level whereby every manufacturing task was carefully timed and the production process was accordingly split into a meticulously devised sequence of equal duration operations – each usually requiring less than a minute to perform. Fordist principles were adopted in other US manufacturing districts, in the British West Midlands, and in the German Ruhr. Thus, the term Fordism, denoting efficiency, economies of scale, assembly line productions, standardization, and effective marketing, was coined to encapsulate the ingredients of the approach.

Liberal Capitalism under Pressure

Economic depressions and boom and bust business cycles became a recurring problem in the late nineteenth and twentieth centuries. The so called Long Depression of the 1870s and 1880s and the Great Depression triggered by the stock market crash in 1929 affected almost the entire capitalist world and prompted discussions about the long term survival prospects of capitalism as a system. During the Great Depression, Marxists confidently predicted the complete collapse of capitalism and pointed to the ability of the Soviet communist regime (which is discussed below) to avoid the worst extremes of the global economic downturn. However, economic recovery after the Depression and World War II undermined communist forecasts of capitalism's imminent demise. Nevertheless, this era saw yet another round of state intervention as governments tried to soften the harshness of the spatial and social impacts of capitalism. Notwithstanding the importance in classical economic theory of a *laissez faire* approach, it was claimed by economists such as John Maynard Keynes (1883–1946) that capitalism could be made compatible with interventionist government. Against the background of the 1930s Depression, Keynesian economics urged state intervention (e.g., through adjustments to personal and corporate taxation levels and to interest rates) to promote higher levels of aggregate demand, to fight unemployment, and to counter deflation. Together with the inexorable rise in the ratio of government expenditure to gross national product (GNP), further new terms were coined by commentators and politicians attempting to explain and understand the workings of evermore tightly managed forms of capitalism: these terms include mixed economy, welfare state, and state capitalism.

The lengthy post war boom came to an abrupt halt amid problems of a sudden embargo imposed in 1973 by the Middle East producing states on oil supplies that led to a leap in crude oil prices and resulted in prolonged stagflation – an episode that reignited interest among some economic theorists in nineteenth century *laissez faire* capitalism. Neoliberal policy prescriptions were urged as the best means of curing the ills of late twentieth century capitalism. Though retaining adherents, the ideas of Keynesian economics were most decisively countered in this time of oil crises, inflation, high unemployment, and low economic growth. Theorists such as Milton Friedman (1912–2006) pointed to a contraction in the money supply (controlled by the US Federal Reserve Bank), rather than a lack of investment, as a key cause of similar conditions that had prevailed in the 1930s slump. Friedman's monetarist ideas urged a return to the free market principles of classical economic theory as the solution to the problems current in the 1970s. Monetarism – a theoretical alter native to Keynesian regulation and interventionism – emerged as a policy approach more compatible with classical liberal capitalism and it was embraced in the 1980s in some parts of the capitalist world, most notably in the USA and the UK.

The Challenge of Command Economies and Communism

In the twentieth century, Marxist style communism un doubtedly posed the most significant challenge to capit alism. Although Karl Marx predicted that communism would find support among the exploited industrial working classes of core economies such as Britain and the USA, paradoxically, the doctrine secured its first victory in 1917 in the least industrialized of the great powers, namely Russia. The Bolsheviks who led the Russian Revolution were therefore faced with the task of intro ducing a Marxist socialist system not among a pre dominantly urbanized proletariat, but in a largely peasant rural society. In the newly created Soviet Union, a highly centralized political system was built on the ashes of the Czarist regime and, by the end of the 1920s, a centrally planned, command economy had been developed in which private enterprise and free trade were abolished and agriculture and industry were collectivized. Military power was strengthened against the perceived threat of the capitalist world. Victory in World War II (achieved at an appalling cost in loss of life), and the subsequent opportunistic territorial expansion of communism across most of Eastern and Central Europe, thereafter divided the world into two starkly contrasting ideological, polit ical, and economic entities – capitalist and communist – separated by Winston Churchill's famously evocative iron curtain. With both sides possessing a growing arsenal of nuclear weaponry, a confrontation arose that was char acterized by an equal and mutual threat, which came to be known as the Cold War.

Communism spread elsewhere, perhaps most notably to China but also, for example, to parts of Southeast Asia. Once again, these recipients did not fit Marx's capitalist–proletarian model and the ideology therefore required adaptation according to particular circumstances. Not withstanding the lack of Marxist purity in its application, communism was regarded as a highly dangerous chal lenge to both the philosophy and the practical operation of capitalism. It became clear, however, that the stifling and inflexible constraints of the command economies of Soviet Bloc countries severely encumbered their ability to compete with capitalism in terms of productivity and innovation. By the early 1980s, the communist world was in economic and political crisis. Unable to prevent the secession from communism of its client states in Central and Eastern Europe, in 1991 the Soviet Union itself crumbled into 15 separate states – once more becoming open to capitalist penetration. The chill of the Cold War began to thaw.

Intriguing debates ensued on the reasons for the collapse of communism in its pioneering heartland. Some argue that Marx's teachings are fundamentally flawed and communism was bound to fail; others blame Lenin for his inept attempt to build communism in a country that was an underdeveloped peasant farming society. Those who champion the notion of the irresistible forces of capitalism see communism's demise as a victory for Adam Smith's notion of the invisible hand of the market. Whatever the truth, the reintroduction of capitalism into the social and economic spaces of former communist countries has engendered a raft of severe social and economic problems that have not yet been resolved. Moreover, while communist rule survives, the forces of capitalism have more recently been unleashed in China with extraordinary, paradoxical, and startling con sequences for the entire world. Again, the full range of impacts and outcomes has yet to emerge.

The Challenge of Fascism and Nazism

Other potent twentieth century challenges to capitalism include fascism and Nazism. Under Benito Mussolini, a fascist government ruled Italy from 1922, and Adolf Hitler led the Nazi (National Socialist) government in Germany from 1933. Both regimes were militarily de feated in 1945. Although the reactionary political phil osophies and ideologies of fascism and Nazism extended across many spheres, one key aspect of its challenge was against the individualism, liberal commercialism, and unchecked materialism of the decadent Western capitalist democracies. While the economic principles of capitalism were not rejected by Europe's fascist and Nazi regimes, strengthening central controls were steadily imposed. Especially in Germany, new kinds of highly regulated capitalist enterprises were therefore developed in the years preceding the outbreak of World War II.

The Challenge of the Islamic Revival/Radicalism

While a full discussion, from a human geography point of view, of the Islamic revival is beyond the immediate scope of this article, it is important to note that the powerful dislike of what is seen as the secularism, colonialism, and social modernization of Western style liberal capitalism is a major feature of a movement that dates particularly from the second half of the twentieth century. Though there are immensely complicated and fundamental cultural, philosophical, political, and ethical issues involved in the Islamic challenge to those aspects of the Western way of life that are regarded as degenerate, there are also some clearly identifiable, tangible objections to the tenets of capitalism. For example, the conspicuous consumption, profligate spending, and the stark social inequalities that pervade some of the strongest capitalist democracies offend Muslim teachings on charity and social justice. There are also important ethical objections to the interest charges imposed on borrowers by the West's capitalist banking system. Rather like the current penetration, at breakneck speed, of capitalism in communist China, the Islamic revival is a rapidly changing and ongoing phenomenon about which it is far too early to predict a final outcome.

Post-Fordism, Globalization, and the Anticapitalist Backlash

While the phenomenon arguably has much deeper chronological roots, the term globalization is commonly invoked to explain the changes that have affected the world's economic geography in the past two or three decades. A startling revolution in all forms of communication has undeniably made the world a much more accessible place. Markets have become truly internationalized and finance now moves across the world in an unfettered way. Speculative trading in 'futures' has become a significant feature of finance capitalism. There has been a rise to prominence of transnational corporations (TNCs), which are vast commercial conglomerates, whose immense commercial strength has severely diminished the power of individual states to regulate their own economies. With their vast wealth and growing political potency, TNCs – usually headquartered in the great cities of the traditional core countries – freely switch their activities and investments from one country to another as economic circumstances demand. Much industrial production has been relocated to the Third World where TNCs seek the benefits of lower costs, cheap labor, and fewer restrictions. The exploitation of workers in Third World countries producing goods for TNCs in poor factory conditions for very low wages may be seen as the latest manifestation of the spatial division of labor. For the West, the corollary of these processes has been the painful experience of deindustrialization and a rapid shift into the services sector and the so called sunrise industries.

The increasing power and sophistication of computing and information technologies have further aided and intensified the development of more flexible production, capable as never before of responding swiftly to changing market demands and fashions. Thus those divisions of labor that once energized the factories and mills of the industrial revolution now occur in new forms on an international scale whereby the bespoke component elements of a product are often manufactured in plants spread across several continents. The rigid, standardized concept of manufacture where uniform products spilled with regularity in great quantity from the Fordist production lines of the past has given way to flexible, market led approaches that, perhaps inevitably, have come to be known as post Fordist.

Conclusion

Globalized capitalism is an ongoing phenomenon. The social changes that it has wrought across the world have already proved to be seismic in their power. In the years after World War II, Western capitalist democracies experienced an economic revival and a commercial boom; new state sponsored welfare systems provided an unprecedented level of security; average standards of living soared to new heights. The benefits of mass production and consumerism were reflected in massively rising car ownership, house purchase and the building of suburbs, international travel, and the purchase of an endless array of consumer goods. But, in the post Fordist era, the uncertainties caused by de industrialization, unemployment, and social inequalities have multiplied. Some commentators, noting the growing vociferousness in the past few years of anticapitalist protesters, have coined another term that rivals post Fordism: we are now, they argue, in an era of disorganized capitalism. While it is hard to foresee with clarity the human geographies that will be created as a result of this phase of disorganized capitalism, we can nonetheless be certain that they will be no less fascinating for human geographers than all those engendered during previous iterations of this most remarkable socioeconomic system.

See also: Feudalism and Feudal Society; Globalization, Economic; Marxism/Marxist Geography I; Marxism/Marxist Geography II.

Further Reading

Braudel, F. (1984). *Civilization and Capitalism 15th 18th Century V III: The Perspective of the World*, (English trans.). London: Collins Fontana Press.

Brenner, R. (1993). *Merchants and Revolution: Commercial Change, Political Conflict, and London's Overseas Traders, 1150 1653*. Cambridge: Cambridge University Press.

Cain, P. J. and Hopkins, A. G. (2002). *British Imperialism, 1688 2000*. Harlow: Longman.

Harvey, D. (1990). The political economic transformation of late twentieth century capitalism. In Harvey, D. (ed.) *The Condition of Postmodernity*, pp 121 200. Cambridge: Blackwell.

Pirenne, H. (1969). *Medieval Cities: The Origins and the Revival of Trade*. Princeton, NJ: Princeton University Press.

Polanyi, K. (1957). *The Great Transformation*. Boston, MA: Beacon Press.

Scott, J. (1997). *Corporate Business and Capitalist Classes*. Oxford: Oxford University Press.

Wall, D. (2005). *Babylon and Beyond: The Economics of Anti Capitalist, Anti Globalist and Radical Green Movements*. London: Pluto Press.

Wallerstein, I. (1976). *The Modern World System, 1: Capitalist Agriculture and the Origins of the European World Economy in the Sixteenth Century*. London: Academic Press.

Wallerstein, I. (1980). *The Modern World System, 2: Mercantilism and the Consolidation of the European World Economy, 1650 1750*. London: Academic Press.

Wallerstein, I. (1989). *The Modern World System, 3: The Second Great Era of Expansion of the Capitalist World Economy, 1730 1840s*. London: Academic Press.

Wood, M. E. (2002). *The Origin of Capitalism: A Longer View*. London: Verso Editions.

Relevant Websites

http://www.dowjones.com/
Dow Jones News Corporation.

http://www.imf.org/external/index.htm
International Monetary Fund.

http://www.imf.org/external/pubind.htm
International Monetary Fund publications.

http://online.wsj.com/public/us
Wall Street Journal online.

http://www.ingentaconnect.com/content/oup/wber
World Bank Economic Review.

http://www.gatt.org/
World Trade Organisation.

http://www.federalreserve.gov/
US Federal Reserve Bank.

Care/Caregiving

N. M. Yantzi, Laurentian University, Sudbury, ON, Canada
M. W. Skinner, Trent University, Peterborough, ON, Canada

Glossary

Care and Caregiving Care that is required due to an individual's short- or long-term health condition and/or disability. It involves diverse and complex webs of health interventions, relations, and practices that occur between people in various settings. Care activities include those for both health and social care such as administering medication and looking after life-sustaining technologies, bathing, toileting, feeding, dressing, communicating, mobility, and caregiver respite. Institutional care is provided in healthcare settings that are built for the express purpose of providing healthcare (e.g., hospitals, psychiatric or mental health facilities, nursing homes, or long-term care facilities). Home and community care is provided in nontraditional healthcare settings that are not built for the sole purpose of providing healthcare (e.g., homes, schools, workplaces).

Community Voluntary Organization A local, community-based organization that provides health and social care services such as meal programs, transportation, respite, and homemaking. These organizations primarily operate outside the formal healthcare system but often hold service-specific contracts with the government, and feature a significant but not exclusive degree of volunteer participation.

Formal/Paid Caregiving Care that is provided by workers who are reimbursed for their time.

Healthcare Restructuring An orchestrated set of policies and practices aimed at reducing the financial responsibility of the government for providing healthcare services. This frequently involves the separation of funding and delivery roles of the government, shifting the place where care is provided to nontraditional healthcare settings, closing or downsizing healthcare institutions in favor of community care (known as de-institutionalization), and transferring responsibility for healthcare from trained healthcare professionals to individuals and their families.

Informal/Unpaid Caregiving Care that is usually provided by close family members but can also be provided by extended family, friends, neighbors, and volunteers who do not receive financial reimbursement.

Publicly Funded Home and Community Care An ensemble of services that are provided by the government (nursing, physical and occupation therapy, personal support, and home-care worker) to enable an individual to live at home, often with the effect of preventing, delaying, or substituting long-term or acute-care alternatives.

Introduction

Care and caregiving involves a diverse and complex web of interventions, relations, and practices that occur between people in various traditional and nontraditional health and social care settings – all of which are intrinsically geographical. In this article, we examine geographical contributions to understanding care and caregiving beginning with a review of its roots within human/health geography with specific reference to changes in contemporary health policy, and a consideration of the reasons why this area of inquiry has grown substantially in the past decade. Next, we examine the diverse theories and concepts that inform geographical perspectives on care and caregiving, which is followed by a consideration of research on the various places and spaces in which, and from which, care and caregiving occur. We conclude with a discussion of the limitations of current knowledge, and we echo recent calls in the literature for a more inclusive approach to understanding care and caregiving.

Geographies of Care and Caregiving

The importance of care and caregiving as a field of study can be understood as part of the general shift toward greater emphasis on the social aspects of healthcare, and as part of the response within the social sciences to broader changes in contemporary health policy and practice in Western developed countries such as Canada, New Zealand, and the United Kingdom. Specific concern for the geographical dimensions of care and caregiving has its beginnings in research on the spaces and places of mental healthcare. Key studies in the late 1970s and 1980s critiqued the de institutionalization of mental health services from asylums to community based care and the implications for individuals coping with mental illness. This body of work focused on the gaps between often nonexistent spaces of care in the community and the needs of people with mental illness. Concerns were raised about how de institutionalization, in combination with the lack of service quality and quantity in the community, left many ex patients poorer, homeless, drug addicted, and even prisoners in the criminal system.

Following this early work in the mental health sector were four changes within the subdiscipline of medical/health geography that have led to the growing engagement with research questions and issues concerning care and caregiving. First, there has been a broadening of scope from focusing solely on the biomedical to considering issues of health, healing, and well being. Second, theories and concepts originating in social, cultural, and feminist geography that examined the body, inclusion and exclusion, and the links between people and their economic, social, and political contexts informed and shaped the way that health and healthcare were studied. Third, the interest in developing a situated understanding of the complex relations surrounding healthcare provided an impetus for studying the intersection between health, place, and the activities, routines, and relationships involved in care and caregiving. Finally, the use of techniques such as interviews to research the experiences and meanings of care and caregiving paralleled the increased interest, use, and validation of qualitative methods in human geography. Key studies in the geography of care and caregiving illustrate the strengths of qualitative methods for capturing the rich experiences and in depth meanings of health and healthcare in everyday contexts. The contributions of other social and health science disciplines such as nursing, health policy, gerontology, and sociology have also helped shape and direct this area of inquiry.

Underlying the growing work on the geographies of care and caregiving are broader concerns about changes in health policy and practice. Since the 1980s, concerns in Western developed countries about rising fiscal deficits, an aging population, and growing healthcare costs led to a period of massive healthcare restructuring and policy change. Strategies to deal with these issues coupled with advances in medical technology, and increasing life expectancies for individuals with disabilities and chronic illnesses, have affected where care is provided, how care is delivered, and who performs caregiving. Research on de institutionalization has been broadened to include acute and long term care. For example, the lengths of hospital stays for acute care have decreased, and therefore patients are released from hospital sicker, often requiring ongoing treatment at home. Advances in healthcare technology have made it possible for many treatments and interventions that were previously administered in hospital settings by healthcare professionals to be managed in homes and communities by individuals and their families. For some people, the home has become a place where complex, sophisticated, long term care is provided. Another cost containment strategy is the restructuring of healthcare work. This involves the transferring of responsibilities and activities to workers who are not as highly skilled or paid, and in the case of informal caregivers are not paid at all. For example, bathing was once the domain of nurses but now responsibility for this activity has been transferred to nursing assistants and home care workers. In addition, restructuring of state provided, publicly funded, home care services is also underway, resulting in stricter eligibility criteria to receive services, cut backs in the amount of hours of care, and greater restriction in the type of caregiving that workers can perform. Ultimately, these strategies have combined to change the places where care occurs, which shifts from healthcare institutions such as hospitals to homes, and the people who are responsible for caregiving as services are downloaded from trained healthcare professionals onto community volunteers, families, and individuals.

Theorizing and Conceptualizing Care and Caregiving

Within geography, a central concern has been to elucidate the spatial manifestations and characteristics of care and caregiving in people's everyday lives with the explicit purpose of challenging taken for granted assumptions within prevailing health policies and practices. There are, however, many different and often contested ways to think about care and caregiving, and main contributions have come from, among others, the political economy, cultural, and feminist perspectives.

Political/Institutional Change

Since the 1990s, the restructuring of health and social care services has had a tremendous impact on care relationships and roles. Key studies have revealed that the shift in where care is provided and how care is delivered affects the experiences and meanings of home and community. Drawing on political economy, much of this work seeks to understand how institutional, political, and economic structures shape experiences and realities of receiving and providing care. These studies help to explain the unequal distribution of care resources between groups of care providers and receivers, between different regions, and between various care settings. In other words, healthcare policies affecting the provision and receipt of care are entrenched in specific political, economic, cultural, and social contexts. From this perspective, care is viewed as a political concept that involves the allocation of public resources, and issues of justice, rights, and citizenship, and researchers have used this approach to examine care and caregiving relations and practices in the home, community, and hospital. Ultimately, this work reveals the consequences that healthcare restructuring has on various groups such as care recipients, informal caregivers, health professionals working in the home and community, and individuals affected by the closure of their community hospital.

Landscapes of Care and Caregiving

The long standing geographical concern for under standing the human experience of place has inspired work on the social and cultural dimensions of care and caregiving. Of particular note is the concept of therapeutic landscape, which draws on the humanistic interest in sense of place, and the attachment and deep meaning that places hold for people. Using this concept, researchers have theorized how particular attributes of the physical and social environment, and the meanings associated with place identity can have a profound im pact on health and well being, and on care and caregiv ing. This concept informs understanding of care providers' and receivers' attachment to particular lo cations. It is particularly useful for understanding care and caregiving in the home, which is a space of care that is deeply symbolic, familiar, and where individuals may experience profound attachment and place identity.

Gender, Care, and Caregiving

Approaching care and caregiving from a gendered per spective raises questions about how the caring work performed by women is undervalued and often invisible in society. The majority of caregiving, both paid and unpaid, is the responsibility of women (see **Table 1** for international data on unpaid caregiving). The profound changes in healthcare policy and delivery affect women the most. Care and caregiving activities that were once the responsibility of skilled professionals now are the responsibility of volunteers and family members in the home and community, the majority of whom are female. Research in this area has contributed to critiques about the often invisible and neglected work done by women in the home, and their critical role in maintaining the health of their family members, neighbors, and friends. An important area of care and caregiving re search centers on the connections between gender, paid work, and caregiving. Drawing on feminist perspectives, work on gender, care, and caregiving examines the travel and activity patterns of women between their home, paid work place, and caregiving location. These explorations of space–time mobility patterns and travel requirements illustrate the difficulties that women experience in bal ancing the competing demands and responsibilities of their households, children, work, and caregiving.

Ethics of Care and Caregiving

Underlying the diverse perspectives described above is a growing concern to develop an ethical and progressive politics of care and caregiving. Recent commentaries highlight the need for a more careful consideration of the ideologically driven settings in which, and from which, care and caregiving occur. For example, studies question whether 'community' is an appropriate resource for caregiving as opposed to institutions, the state, or the market. The changes in where care is delivered and the practice of providing care have led to a blurring of the boundaries between healthcare settings and non traditional healthcare settings; researchers are question ing the dualism and polarity of concepts such as formal and informal care, and private and public. For some in dividuals and their families, the home's physical and social space can go through some drastic changes. Med ical equipment, strict medication and care schedules, and paid care workers coming into and moving about the home can change the way a home looks and feels. This has led to some researchers discussing the institutional ization of the home space, which in some cases, begins to look and feel like a hospital. There is a blurring in the type of care as family members caring for individuals with very complex and intense needs may be providing the same level of care that trained health professionals provide in the hospital setting. The only difference be tween the care provided by family and friends (informal) and that provided by trained health professionals (formal) is that the latter group is paid.

An ethics of care and caregiving is informed by con cepts of social and/or compassionate justice, which ad vocate an inclusive approach in which all groups involved in care are included in debates about healthcare re structuring and caregiving. Work in this area is charged with setting the agenda for how the politics of care and caregiving might be played out, and critiquing the as sumptions and ideas that are framing the healthcare policy that impacts care and caregiving. In other words,

Table 1 Percentage of women among caregivers receiving a care benefit or caring for a recipient of such a benefit

Australia	59.9% of recipients of carer's payment are women[a]
Austria	80% of carers of persons receiving an attendance allowance are women
Canada	Not available
Finland	80% of carers of recipients of an informal carer's allowance are women
France	99% of those employed in the context of family services (*emplois familiaux*) are women
Germany	97% of registered informal carers are women
Japan	NA

[a] A married woman with an employed husband providing care is not eligible due to the means test. In addition, part of eligible carers over 60 65 would be on an age pension.

From Jenson, J. and Jacobzone, S. (2000). Care allowances for the frail elderly and their impact on women care givers. OECD Labor Market and Social Policy Occasional Papers. No. 41, p. 49.: OECD Publishing.

the ethics of care and caregiving moves beyond analytical exercises in description.

Places and Spaces of Care and Caregiving

There is a growing body of work that explores the meanings and experiences of care and caregiving in various settings including healthcare institutions, communities, homes, and the body. The characteristics, meanings, and experiences associated with these four spaces of care and caregiving affect the organization and delivery of care, and the relationships, experiences, and practices involved in providing and receiving care. At the same time, the organization and delivery of care, and the relationships, experiences, and practices that are involved may affect the social and spatial organization and characteristics of these spaces. While for clarity these spaces of care are discussed separately, they are intertwined in the everyday lived experiences of people, and it is crucial to recognize the interrelated scales and spaces of care and caregiving.

The Institution

The two main institutions for care and caregiving are hospitals including psychiatric facilities, and long term care facilities such as residential and nursing homes. Research on hospitals as spaces of care and caregiving has focused on the changing social, physical, and service environments associated with healthcare restructuring and the implications for patients and local communities. One aspect of this work critically examines the increasing commodification of healthcare institutions, and changing financial relationships and partnerships between the government and private market interests as strategies for dealing with soaring healthcare costs in Western developed countries. Another avenue of research centers on building understanding of the geographical variation in the availability and use of hospital services across time and place, and in relation to socioeconomic status, gender, and cultural and ethnic groups.

Residential and nursing homes are spaces of care and caregiving for individuals who are not acutely ill and require hospital care, but are not able to manage either by themselves or with assistance in their own home. Individuals living in these settings include frail elderly people, and individuals of all ages with physical, cognitive, and mental illnesses and disabilities. Key studies provide a more nuanced understanding of why long term care residential institutions are viewed as negative, while the home is viewed as a positive setting in which to provide or receive care. The research reveals differences in people's attachment and belonging to these two different spaces of care. In addition, there is a stark contrast between the familiarity and comfort of the home's physical environment such as the furniture, pictures, and the memories and stories that occurred there, and the control over the schedules and activities of care, and the unfamiliar, often sanitized, space of a nursing home, and the need to renegotiate caring roles and relationships.

The Home

An increasing number of studies are examining and critiquing the consequences of situating care and caregiving in the home. The home, as the principal setting for daily life, is one of the most important zones of familiarity and meaning. Current research demonstrates that situating healthcare in peoples' homes affects the spatial and temporal organization of the home, raises issues concerning privacy, and affects individual family members' social networks. Some of the key issues include the immense amount of work and effort involved, the impacts on the spatial mobility and social life of the families and individuals providing and receiving care, and the differing abilities of caregivers and care recipients to access resources and support networks. Key studies have determined that the ability to obtain resources and supports to assist in providing and receiving care varies due to distances between caregiver and care receiver, differences between urban and rural locations in terms of availability of public transportation and in home care services, and varying ability to pay for required services that are not publicly funded.

The profound relationship between the experiences and meanings of disability and/or illness and the physical and social environment of the home has been demonstrated. Research shows that people experience many changes in their home when it becomes a place where care activities, schedules, relationships, and processes occur. Some of these changes include moving to a more suitable environment, structural alterations such as entrance ramps, grab bars, and bath lifts, moving the bedroom to a more accessible location in the house, reordering domestic work including delegating tasks and pacing of the work, hiring private help, and using publicly funded home care workers. Both the characteristics of the home and family including accessibility, ownership, socioeconomic status, and family structure come together and create different meanings and experiences of care in the home.

Research on the impact of the changing context of publicly funded home care on individuals and families has revealed concerns about stricter eligibility criteria, cuts in the maximum number of hours, and restrictions in the types of tasks allowed. Service inflexibility, the ability to access the type and amount of services that are required, and high turnover of home care staff results in the inevitability that informal caregivers will have to look after most of the care activities. The example of the change in what publicly funded home care workers are permitted to do clearly shows the intersection between healthcare policy, homes, and individuals and families. Efforts to reduce home care budgets has led to workers being restricted to providing personal care such as baths and not other activities such as light house work, grocery shopping, or preparing meals. If individuals or families are not able to look after these activities they must hire someone to help, or more likely they rely on informal caregiving.

The Community

In addition to care and caregiving in the home, elderly people and individuals with disabilities also receive care in the community. Care and caregiving research focuses on the intersection between care and community settings such as workplaces, schools, counseling settings, and community centers. Of particular concern are the im plications of the restructuring of both acute and long term healthcare and the withdrawal of governments from the provision of services, which are characterized by the downloading of responsibilities for providing health and social care onto community based voluntary organiza tions. Numerous studies have examined the challenges facing these voluntary sector organizations as they relate to the uneven nature of restructuring, the relationships between health professionals, volunteers, and individuals who receive care. Key studies have also examined 'non traditional' spaces of care in the community such as day centers, drop in centers, and soup kitchens. Research on these spaces emphasizes the importance of caring prac tices that focus on bodily necessities such as food, shelter, and social relationships. Most of this research links back to the consequences of the de institutionalization of those with mental healthcare needs and the lack of ap propriate services in the community.

The Body

The body is an important, yet until recently neglected, focus of care and caregiving research. Research in this area concentrates on how caring relations and practices are experienced both physically and emotionally, and imposed, on the bodies of caregivers and care recipients. The majority of the activities associated with care work such as feeding, changing clothes, bathing, toileting, ad ministering medication, and monitoring healthcare technology are focused on the body. In essence, care giving activities revolve around preserving, maintaining, restoring, and comforting the body. These necessary care activities can either be done by the individuals them selves, known as self care, or with the assistance of a paid or unpaid caregiver. Studies have examined the extra ordinary investment of time and energy that is required in performing care work, the scheduling of care activities around fundamental bodily activities such as eating or sleeping, and the modifications that are required to the built environment as bodies grow, develop, deteriorate, and change. Bodily boundaries concerning privacy and control may have to be transgressed and renegotiated as many care activities are extremely personal and intimate.

Care and caregiving is also emotional work. Studies have shown that caregivers and care receivers experience complex and conflicting emotions such as joy, sorrow, guilt, anger, frustration, anxiety, hopefulness, and hope lessness. A myriad of roles and relationships may be re versed, such as an adult child providing care for the mother or father, or not expected, such as a parent having to provide more assistance to the child with a disability even though he or she is getting older. Drawing on psychoanalytic approaches, recent research has focused on the emotional implications of care and caregiving, and emotional connections to different care settings. The acknowledgement of the relationship among individuals' bodies, caregivers' bodies, places, and policy signals the importance of developing an ethics of care among health policy makers, researchers, and practitioners.

Toward an Inclusive Geography of Care and Caregiving

In this short article, we examined the dynamic and changing geographies of care and caregiving. This field of study evolved out of the combination of subdisciplinary developments in medical/health geography, and out of broader concerns for the implications of changes in healthcare policy and practice in Western developed countries. We reviewed the diverse set of theories and concepts that inform geographical approaches to care and caregiving, and we discussed the different institutional, home, community, and body spaces in which and from which care and caregiving occurs.

As with most emerging fields of study, there are sev eral gaps in existing knowledge about care and caregiv ing. Specifically, most research on care and caregiving has focused on older people, especially women, and urban places in Western developed countries. The experiences of children, youth, and younger adults, male caregivers, and individuals who practice self care are largely absent from the literature. The experiences of individuals living in developing countries, and the unique conditions and requirements of care and caregiving in rural areas are also largely under represented in the care and caregiving literature. These gaps in knowledge, however, have been the focus of several agenda setting developments in care and caregiving research, including the recent publication of special issues in international geography journals that call for more inclusive geographies of care and caregiv ing. The latter represents, arguably, the most important and potentially fruitful avenue for future research on care and caregiving.

We have shown that changes within healthcare policies and practices affect the places and spaces of care and caregiving, and people's place specific care experiences. Research on the geographies of care and caregiving is aimed at improving the lives of care recipients and caregivers, and this underpins the growing commitment to developing an ethical and progressive politics of care. In this pursuit, care and caregiving researchers are making rigorous, applied contributions to healthcare policy and practice, are engaging with numerous communities of practice, especially policy decision making structures and processes, and are informing larger debates within medical, health, and human geography.

See also: Ageing and Health; Body, The; Disability and Chronic Illness; Emotional Geographies; Gender and Health; Health Geography; Health Services Restructuring; Mental Health; Therapeutic Landscapes; Voluntary Sector; Welfare Reform.

Further Reading

Conradson, D. (ed.) (2003). *Special Issue on Geographies of Care*. *Social and Cultural Geography* 4(4), 451 547.

Gleeson, B. and Kearns, R. (2001). Remoralising landscapes of care. *Environment and Planning D: Society and Space* 19, 61 80.

Jenson, J. and Jacobzone, S. (2000). Care allowances for the frail elderly and their impact on women care givers. *OECD Labor Market and Social Policy Occasional Papers.* No.41, p. 49.: OECD Publishing.

Milligan, C. (2001). *Geographies of Care: Space, Place and the Voluntary Sector*. Aldershot, UK: Ashgate.

Parr, H. (2003). Medical geography: Care and caring. *Progress in Human Geography* 27(2), 212 221.

Staeheli, L. A. and Brown, M. (eds.) (2003). *Special Issue on Geographies of Care and Welfare*. *Environment and Planning A* 35(5), 771 870.

Twigg, J. (2000). *Bathing: The Body and Community Care*. London: Routledge.

Cartographic Animation

M. Harrower, University of Wisconsin – Madison, Madison, WI, USA

Glossary

Interactive/Interactivity Allows the user to change or manipulate the display.
Pace The amount of change per unit time.
Spatiotemporal Data Geographic data that has both spatial and temporal coordinates or components.
Temporal Granularity/Resolution The finest temporal unit resolvable.
Temporal Interpolation The creation of intermediate or missing data between moments for which data are known.

Introduction

Animated maps, sometimes called movie maps or change maps, are primarily used to depict geographic change and processes. Cartographers have long been interested in the potential of map animation because representing complex geographic processes on static maps is a significant challenge. Put simply, animated maps are appealing because we live in an animated world. Humans also possess an eye–brain system that is finely tuned to seeing (perceptually) and understanding (cognitively) motion and change. Static paper maps do not permit the direct representation of time since they do not change, although they can be used to indirectly represent time using either a single map (e.g., lines of equal travel time) or a series of maps (e.g., small multiples). By comparison, cartographic animations use time to represent time allowing us to depict dynamic phenomena directly on our maps. Thus, there is a strong conceptual congruence between dynamic phenomena and dynamic representations of those phenomena.

Research has shown that map animation can be used to generate new insights into geographic processes that had been previously unrecognized using either static maps or statistical analysis. The need for animated maps is further underscored by the recent and unparalleled explosion of digital geospatial data – much of it time series data. Fortunately, the simultaneous and rapid development of the World Wide Web and affordable, high performance desktop PCs have made animated maps far easier to make, distribute, and view. As a result, animated maps have gone from being painstakingly crafted novelties to commonplace digital documents. Making animated maps interactive further extends their utility and allows the reader to 'go into the map' and (in theory) change the map to suit their needs. The amount of interactivity, or control, the reader has over the animated map has steadily increased over time as technology permits (**Figure 1**), and ranges from none (e.g., weather map on TV) to near total (e.g., geovisualization system for epidemiologists).

Almost any kind of map can be animated: weather maps of air temperature, choropleth maps of population density, graduate symbols maps of income, and so on (**Figure 2**). Major research topics within map animation over the last 15 years include: (1) methods for animating time series data, (2) methods for animating across attributes of data, (3) methods for representing uncertainty and data quality in animated maps, (4) designing effective temporal legends and controls, (5) identifying the visual variables of animation, and (6) methods for temporal interpolation and smoothing of sparse datasets.

History of Animated Maps

The history of animated maps is very much tied to the history of computers. Early examples of animated maps were produced in the 1930s using hand drawn, cel based techniques borrowed from cartoon animation. Because these manual techniques were exceptionally time consuming, academic cartographers and computer scientists started experimenting with the potential of computer assisted animation surprisingly early (late 1950s) and by 1970 Waldo Tobler had created a path breaking oblique perspective 3 D animation of anticipated population growth in Detroit. Even more remarkable was that

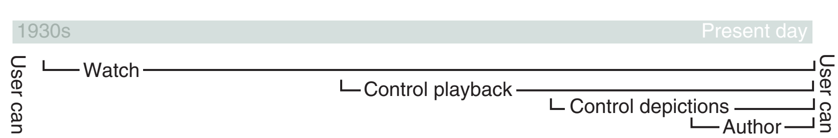

Figure 1 The history of animated maps and what users can do with them is in large part dictated by the storage medium/technology available at the time: celluloid film, VCR, CD-ROM, and the web. Like many other areas of GIScience, the move to computers in the 1980s allowed for much more interactive and versatile animated maps, while a decade later the migration to web-based systems potentially gave users the ability to make animated maps on-demand (from distributed and/or real-time data).

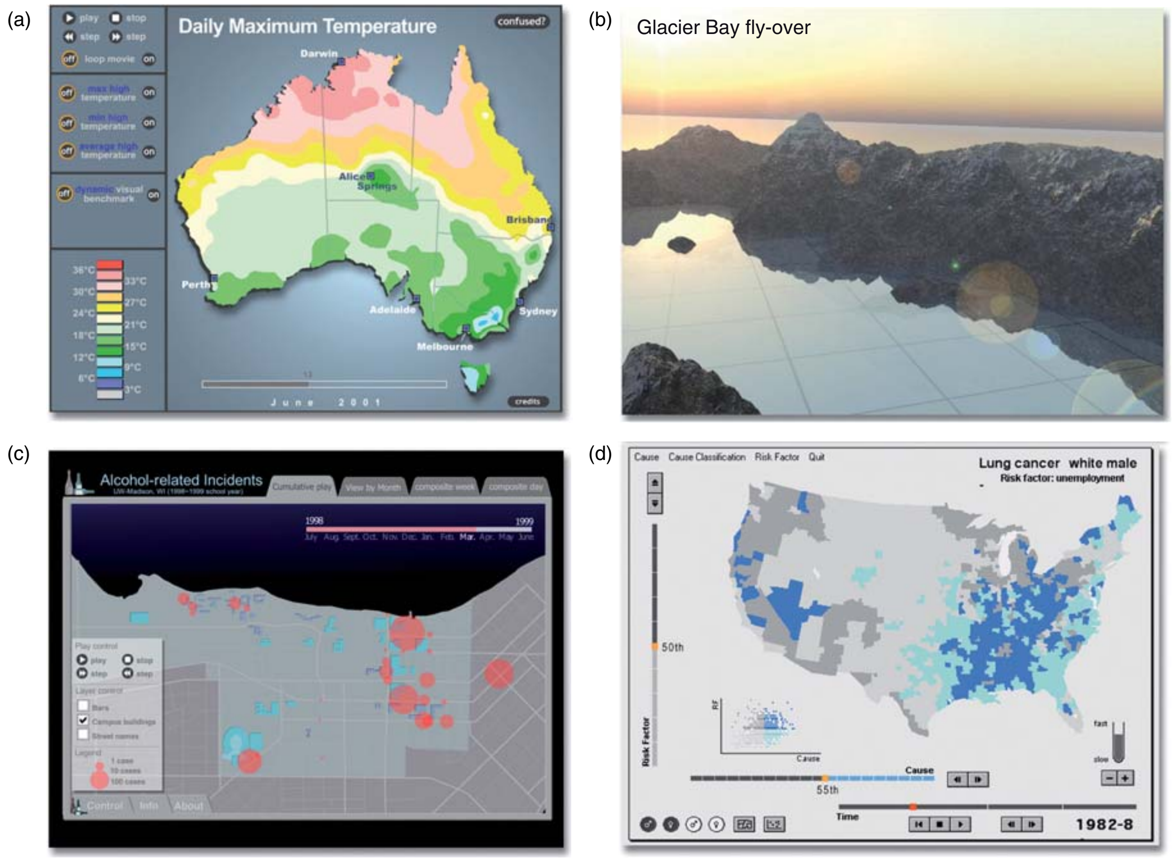

Figure 2 Four examples of animated maps The temperature map of Australia (a) demonstrates our ability to see coherent motion in map animation as warm and cold air move across Australia. The 3D oblique-view map (b) takes viewers shows how animation can create highly relaistic and immersive experiences. The alcohol-incident map (c) contains temporal re-expression controls allowing users to animate through the data by composite hour of day, day of the week, and month (Created by Jian Liu and Feng Qi at the University of Wisconsin-Madison, (2003). The cancer-rates map (d) is a focusable bivariate choropleth map allowing users to investigate the correlation between kinds of cancer and associated risk factors (across both time and space). (Created by Alan MacEachren, Pennsylvania State University, 1996.)

Tobler's map was not created to present known facts, but rather was a research tool to better understand the pro cess itself – an exploratory activity that today would be recognized as geovisualization.

In the 1970s, the advent of VCR technology meant that animated maps could now be copied and distributed easily and cheaply (compared to celluloid film). It also allowed map viewers rudimentary control over playback. However, it was not until the PC revolution of the late 1980s that animated maps finally came into their own: The desktop PC – with its graphical user interface, real time interactivity, standardized file formats, and off the shelf software – finally armed cartographers with a powerful tool for making animated maps, and viewers a means for watching and interacting with them. The last piece of the puzzle, product distribution, was solved by the rise of the Internet in the 1990s. No longer tied to physical media, web based animated maps could now be 'made' on demand and sent around the world in a blink of an eye at minimal cost to either the producer or the consumer. The drive to automate the production of maps (including animated maps) has been a significant topic of research in Geographic Information Science (GIScience) in the last decade, although a number of technological and conceptual hurdles must still be addressed (the latter being significantly harder to address than the former).

Temporal Animated Maps

Static maps present all of their information simul taneously; animated maps present information over time. Thus, animated maps have an additional representational dimension that can be used to display information. While it obviously takes time for a reader to scan a paper map, such as retrieving specific rates, the overall or holistic impression of the map happens at a glance. Moreover, the reader does not have to wait for the information to ap pear as with an animation. Increasing the running time of an animation increases the total amount of data that can be represented, but at a cost to the user. As the length of the animation increases so too does the difficulty of re membering each frame of the animation. Put another way, although the amount of data that can be represented within an animation is virtually unlimited, there is a fi nite amount of information the user can distill from the animation and store in his/her short term visual memory. Overwhelming the map reader with information does more harm than good. As a result, most animated maps are short and few run longer than a minute.

Just as static maps have a spatial scale, temporal ani mated maps have a temporal scale. This can be expressed as the ratio between real world time and movie time. For example, 5 years of data shown in a 10 s animation would have a temporal scale of 1:157 million. Although some authors have suggested varying the temporal scale of the map as it plays – to focus on important moments, or blur out others – most animated maps keep a constant tem poral scale. Additional aspects of the map include its temporal granularity/resolution (the finest temporal unit resolvable) and pace (the amount of change per unit time). Pace should not to be confused with frames per second (fps): an animation can have a high frame rate (30 fps) and display little or no change (slow pace). Per ceptually, true animation occurs when the individual frames of the map/movie no longer are discernable as discrete images. This occurs above roughly 24 fps – standard frame rate of celluloid film – although frame rates as low as 5 fps can generate a passable animation effect. Higher frame rates yield smoother looking ani mations, although modest computers will have trouble playing movies at high frame rates, especially if the map is a large raster file.

Temporal legends should be included in any animated map of temporal data. **Figure 3** shows three different kinds of temporal legends: digital clock, cyclical time wheel, and linear bar. The advantage of graphical tem poral legends, such as the time wheel and linear bar, is

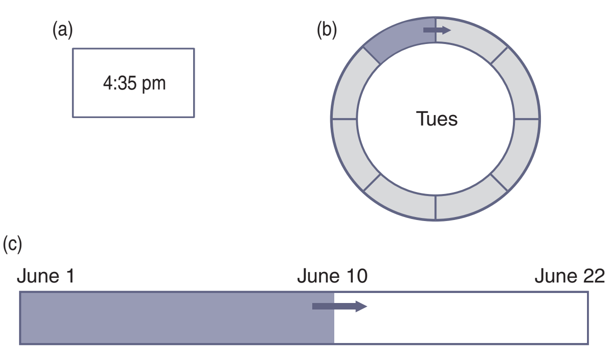

Figure 3 Three most common kinds of temporal legends: (a) digital clock, (b) cyclical time wheel, and (c) linear bar. The graphical cyclic and bar legend can communicate at a glance both the specific instance and the relation of that moment to the whole.

that they can communicate at a glance both the current moment (e.g., 4:35 pm) and the relation of that moment to the entire dataset (e.g., halfway through the animation). Because it is often assumed these kinds of legends sup ports different map reading tasks, designers often include more than one on a single map. Cyclic legends, for ex ample, can foster understanding of repeated cycles (e.g., diurnal or seasonal) while linear bars may emphasize overall change from beginning to end.

By making a temporal legend interactive, the reader can directly manipulate the playback direction and pace of the movie, or jump to a new moment in the animation (known as nonlinear navigation). This has become a common interface action in digital music and video players and many map readers now expect to be able to directly interact with temporal legends to control the map. One unresolved problem with legend design is split attention: because animated maps, by their very nature, change constantly, the moment the reader must focus on the temporal legend they are no longer focusing on the map and may miss important cues or information. If they pause the animation to look at the legend, they lose the animation effect. The more the reader must shift atten tion between the map and the legend, the greater the potential for disorientation or misunderstanding. Pro posed but untested solutions to the problem of split at tention include audio temporal legends and embedded temporal legends that are visually superimposed onto the map itself.

Temporal Re-Expression, Filtering, and Interpolation

Although the most obvious way to animate spatio temporal data is to simply play the data in the same sequence as it was collected (e.g., day 1 = frame 1, day 2 = frame 2, and so on), this is not always very insightful because: (1) the data may be very complex and contain multiple meaningful patterns, and/or (2) the data may be very noisy.

Temporal re expression is one of the most powerful cartographic transformations we can apply to better understand spatiotemporal data (the term was coined by David DiBiase in the early 1990s, although early ex amples appeared in the 1970s). Temporal re expression is displaying the same dataset using different temporal units in which the data are aggregated to create com posite frames (or moments) in the animation. For ex ample, in **Figure 2** the 360 days of drinking related incidents can be animated over a composite day (24 frame animation) to show how these might differ ac cording to time of day allowing us to answer questions such as, 'do more incidents occur at 8 am than at 8 pm?' Because the data are mapped, the spatial patterns of accidents and where those accidents occur (and by what cause factor) can be investigated allowing the reader, for example, to ask 'even if the number of incidents are the same, does where they occur differ between 8 am and 8 pm?' The map also depicts the incident data as a composite week (e.g., 'Do more incidents occur on Mondays than on Fridays?'), and composite month (e.g., December vs May).

Temporal aggregation – changing the temporal granularity of the map – allows viewers to create temporal slices of different thickness. At ballotbank.com (**Figure 4**) financial campaign contributions are initially displayed monthly (the finest temporal unit available), although the viewer can aggregate and average the data on the fly over long intervals. In **Figure 4**, the temporal granularity of the map has been increased from 6 months to 12 months (each frame now shows a 12 month period). Because campaign contributions are closely tied with tax season, as well as election cycles (both 2 and 4 years), this flexibility can help to uncover subtle behaviors in political participation that the most temporally disaggregate data obscures. The website ballotbank.com also incorporates temporal focusing which lets the map reader set the start and end dates of the animation.

Many animated maps employ temporal interpolation to 'fill in' missing time periods or to create a smoother looking animation. In some published examples, the interpolated data outnumbers the 'real' data 10 to 1. Although popular, the cognitive effectiveness of temporal interpolation has only been partially studied (results suggest subjects prefer maps with interpolation, although do not perform any better on map reading tasks). Various linear and nonlinear algorithms and techniques have been

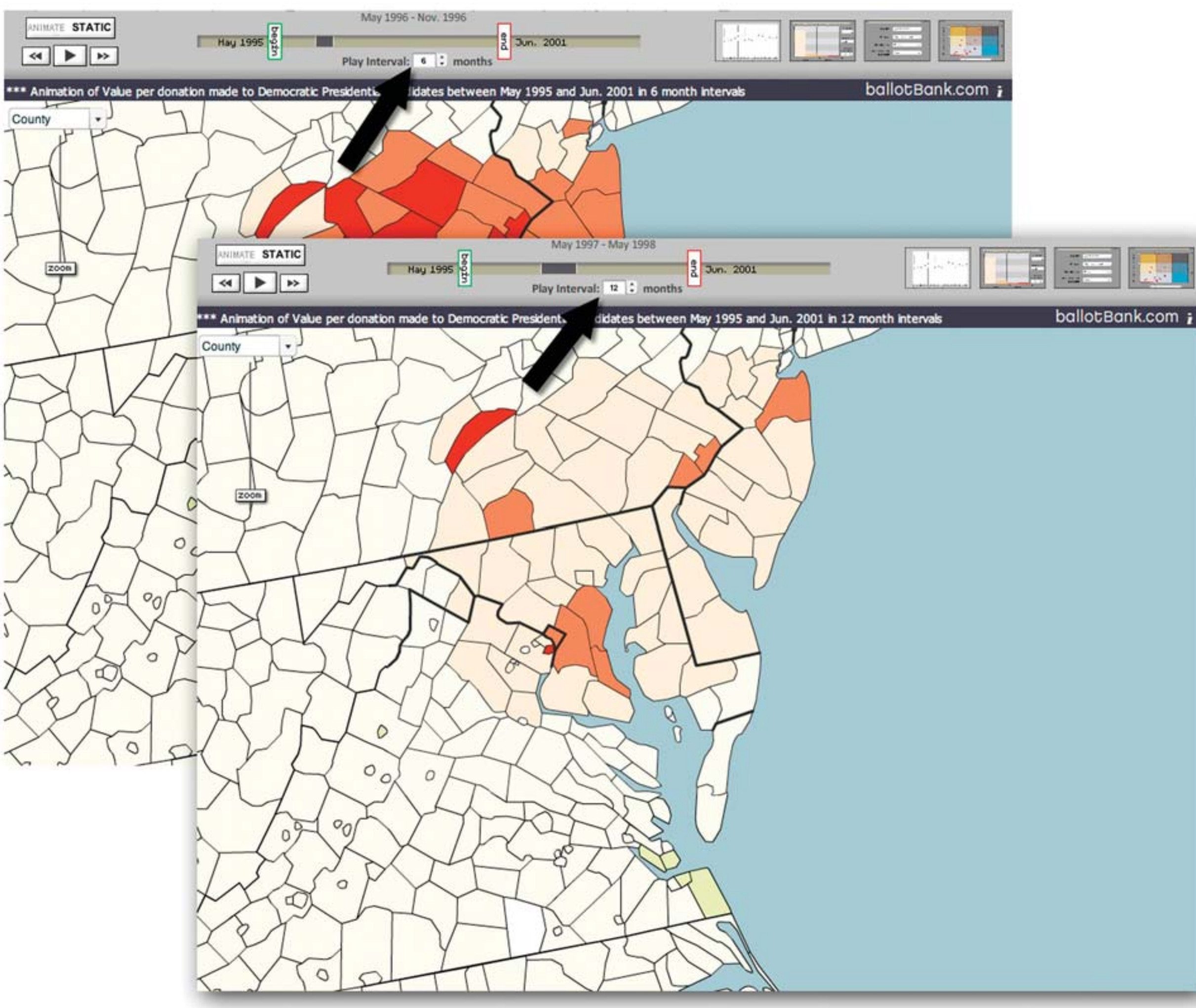

Figure 4 Ballotbank.com was built with powerful temporal controls including temporal focusing (i.e., set the start and end of the animation) and temporal aggregation (i.e., one frame represents from one to 'n' months) so that users can change the 'temporal granularity' of the map and animate through different intervals and durations of time. (Created by David Heyman, University of Wisconsin-Madison, 2007.)

proposed for creating quality interpolated temporal data, some of which are quite sophisticated mathematically. Nonetheless, some authors worry that such data manipu lation will create false and highly misleading patterns, such as an abrupt change in the location of a feature shown as a gradual spatial transition from point A to point B.

Nontemporal Animated Maps

Animated maps do not have to depict temporal data. One kind of a nontemporal animated map is a fly over map that simulates the experience of 'flying' over a digital landscape or zooming into or away from a location (as is possible with Google Earth). Most often with fly overs the map itself is static (e.g., the colors of the map do not change); it is the vantage point of the map reader that is animated. It is generally assumed that animating both the vantage point of the user and the underlying map/data can quickly exceed the information processing capabil ities of most map readers and hence few maps attempt to do both. While many fly over maps show an actual topographic surface (usually with strong vertical ex aggeration), any geographic data surface can be used (e.g., precipitation, income). With fly over maps, the flight path may be either predetermined or the reader may have some to complete interactive control of the flight path. While allowing users to explore these virtual spaces on their own allows for more flexible map use, it places an increased cognitive burden on them since they must: (1) learn how to control their navigation (usually in 3 D), and (2) remember where they are at risk becoming dis oriented. Most of the research to create more effective fly over maps and virtual spaces has been done in com puter science and psychology.

A second kind of a nontemporal animation is a map that cycles through some sequence of data in which each frame of the animation shows a different 'slice' or treat ment of the data. For example, an animated choropleth map might show voting patterns in a city by age, in which each frame maps a different age cohort (18–27, 28–37, and so on). By playing these frames in sequence as an ani mation, differences in spatial voting patterns between these age cohorts can be easily seen. Another example of application would be to run an urban simulation model that produces a single static map of anticipate growth in the next 25 years. By running the model many times and changing the parameters or assumptions in each run, an animation could be built up frame by frame from each model run and then played in rapid succession to see what areas remain consistently flagged for growth. Because the eye–brain system is especially attuned to seeing even very slight changes in color/position (such as a single pixel among millions flickering on and off) this is in fact a common technique for visually searching through massive amounts of data in scientific visualization.

Problems and Limitations of Animation

Beyond the problem of split attention mentioned above, psychologist Barbara Tversky and her colleagues have identified four major challenges when people watch and learn from animated graphics (including maps) as com pared to the static graphics of the same data:

1. *Attention:* In any visually rich display, especially in an animated map where many things can be happening simultaneously, readers are often unsure of where to look (a problem with the maps) or what to do (a problem with the interface). Potential solutions in clude guided tours, short tutorials, and voice over.
2. *Disappearance:* By their very nature animated maps change, often quite dramatically from moment to moment. As a result, there is always the potential that the map reader will miss important information or cues. Because of disappearance (i.e., blink and you miss it), many basic map reading tasks can be very difficult, such as estimating the size of symbols or areas, matching colors to a legend, comparing one symbol to another, or reading text labels.
3. *Complexity:* With (1) the astonishing level of computing and video processing power available today in the basic desktop computer, coupled with (2) the equally astonishing size and complexity of digital databases, it is entirely possible to make an animated map that simply overwhelms readers with data and 'saturates' their visual processing. Data filtering tools and tech niques that can help us cope with this complexity have been a major avenue of research in GIScience for 20 years now and many journals and international con ferences are devoted to this topic.
4. *Confidence:* Early testing has shown that while people often like looking at animated maps, they are some times less confident of the knowledge they acquire from the animation than from static graphics. In re sponse to this blanket concern, others have pointed out that people have far more exposure and experi ence using static maps, and thus there is naturally a learning curve to animated maps that needs to be understood (and systematically targeted) before we can conclude they do not work. Moreover, there are obviously some representational tasks for which ani mated maps are poorly suited (and a static map or table are inherently more efficient) and to test ani mated maps under those conditions is unfair and not especially informative.

It should be noted that none of these challenges are in surmountable and many graphic design solutions have been proposed (and tested) for each of these, primarily relying on increased and/or new forms of interactivity, better interface controls (e.g., more intuitive, less obtru sive), and a willingness to restrain the design of our animated maps so they respect the cognitive and per ceptual limitations of the eye–brain system, such as avoiding well understood limitations in human vision such as optical illusions and 'change blindness'. As a way forward, researcher Sara Fabrikant and her colleagues are doing promising work around the notion of 'visual sali ence' and understanding how people see/understand map animation in order to increase their effectiveness.

Anecdotally, many cartographers will confess they make animated maps because 'they look cool' and, to their credit, the public seems to enjoy watching these maps. Today, animated maps seem to be all around us, from the weather channel, to NASA press releases about lost planetary probes, to nightly television news stories that begin with an animated map that zooms in from space to some foreign locale, thus, connoting an air of (1) currency, (2) powerful panoptical surveillance, and (3) scientific authority to whatever may follow. What is less clear, however, is simply how useful animated maps are when static alternatives are both easier to make and ponder since they do not change. Research that compares the effectiveness of learning from animated and static maps has been mixed and, thus, academic cartographers have an incomplete answer to even the most basic question: 'For what kinds of map reading tasks are ani mated maps useful, and when should they be avoided?' This does not mean animated maps should be avoided, but rather, like any form of representation, they have their strengths and weaknesses that researchers are work ing to understand. Animated maps should not be seen as replacements to static maps, but rather as complimentary. They broaden our representational possibilities espe cially when data have a temporal component.

See also: Geovisualization; Internet/Web Mapping; Map Interactivity; Mapping, Distributed.

Further Reading

Cartwright, W., Peterson, M. and Gartner, G. (eds.) (1999). *Multimedia cartography*. Berlin: Springer Verlag.

DiBiase, D., MacEachren, A. M., Krygier, J. B. and Reeves, C. (1992). Animation and the role of map design in scientific visualization. *Cartography and Geographic Information Science* 27(4), 201 214.

Dorling, D. and Openshaw, S. (1992). Using computer animation to visualize space time patterns. *Environment and Planning B: Planning and Design* 19, 639 650.

Fabrikant, S. I. and Goldsberry, K. (2005). Thematic relevance and perceptual salience of dynamic geovisualization displays. *Proceedings*, 22th ICA/ACI International Cartographic Conference, A Coruna, Spain, Jul. 9 16: (CDROM).

Fairbain, D., Andrienko, G., Andrienko, N., Buziek, G. and Dykes, J. (2001). Special Issue: Representation and its relationship with cartographic visualization: A research agenda. *Cartography and Geographic Information Science* 28(1), 13 28.

Griffin, A. L., MacEachren, A. M., Hardisty, F., Steiner, E. and Li, B. (2006). A comparison of animated maps with static small multiple maps for visually identifying space time clusters. *Annals of the Association of American Geographers* 96(4), 740 753.

Harrower, M. (2004). A look at the history and future of animated maps. *Cartographica* 39(3), 33 42.

Harrower, M. and Sheesley, B. (2007). Utterly lost: Methods for reducing disorientation in 3 D fly over maps. *Cartography and Geographic Information Science* 34(1), 17 27.

Hearnshaw, H. M. and Unwin, D. J. (eds.) *Visualization in geographic information systems*. Wiley: New York.

MacEachren, A. M. (1995). *How maps work: Representation, visualization and design*. New York: Guilford Press.

Peterson, M. P. (ed.) (2003). *Maps and the internet*. Cambrige: Elsevier Press.

Slocum, T. A., McMaster, R. B., Kessler, F. C. and Howard, H. H. (2003). *Thematic cartography and geographic visualization* (2nd edn.). Upper Saddle River, NJ: Prentice Hall.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46, 234 240.

Tversky, B., Morrison, J. B. and Bétrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human Computer Studies* 57, 247 262.

Relevant Websites

http://www.ballotbank.com
BallotBank is an assisted, map based, visualization system designed to help political scientists and political professionals explore Federal Election Commission campaign contribution data.

http://www.geography.wisc.edu
Department of Geography (University of Wisconsin Madison, USA).

http://www.geo.unizh.ch/gis
GIScience Center (University of Zurich, Switzerland).

http://www.kartoweb.itc.nl
ICA Commission of Visualization and Virtual Environments.

http://www.karto.ethz.ch
Institut fur Kartografie (ETH, Switzerland).

http://www.ncgia.ucsb.edu
National Center for Geographic Information & Analysis (University of California, Santa Barbara, USA).

http://www.geovista.psu.edu
The GeoVISTA Center (Penn State, USA).

Cartography in Islamic Societies

S. Brentjes, Universidad de Sevilla, Sevilla, Spain

Glossary

Al-Baḥr al-Muḥīṭ The Encompassing Sea.
Barzakh A small, narrow land barrier that separates the two main oceans of the maps, that is, the Indian Ocean and the Mediterranean Sea. The main Quranic meaning of *barzakh* is, however, not geographical, but eschatological. It denotes the life of the soul after the death of the body until the Day of Judgment.
Dār al-funūn House of the Disciplines/Sciences/Arts, new school of higher education for the Qajar army and administration, founded in 1851; taught European military methods, theories, technologies, sciences including geography and cartography and languages.
Ḥadīth A body of religious teachings about what the Prophet and his companions had said and done.
Ḥudūd Boundaries, the word has also other meanings such as definitions.
'Ilm al-hay'a Mathematical cosmography.
Isolarii Books about islands, produced in Italy.
Mi'rāj Ascent to the heavens.
Mount Qāf Often drawn on world maps around the Encompassing Sea; this is derived from an interpretation of Sura 1.1 according to which God created a mountain chain at the edge of the earth out of a single green emerald. This chain is called Mt. Qaf.
Qibla Indicates direction toward Mecca, hence prayer direction, also direction regulating other social activities such as the way in which a dead body is laid to rest.
Sīlā Island Korea.

In Islamic societies, cartographic activities were part of courtly patronage and diplomatic exchange, educational literacy and scientific study, religious rituals and reminiscence, entertainment and the arts, and to some extent of war, seafaring, and administration.

Many of the extant cartographic products are parts of manuscripts. Words, narratives, numbers, tables, and geometrical figures organize, structure, delineate, and interpret them. Books with maps form several topical categories as well as genres. They include cosmography, astronomy, history, and geography on the one hand and pilgrimage books, encyclopedias, dictionaries, and miscellanies on the other. The most important types of cartographic products are world maps, regional maps of the Islamic world, and maps and diagrams of the prayer directions, the *qibla*. The first printed maps appeared in the early eighteenth century in Istanbul, first as single sheet maps, but quickly followed as parts of a book on the Old and the New World, that is, *Ḥajjī Khalīfa's* (d. 1067 h/1657) *Cihānnümā* (version II). These printed maps adopted and adapted certain cartographic conventions, methods, and fields of interest that dominated cartography in early modern Catholic and Protestant societies in Europe. This was, however, neither the first nor the last encounter between the various cartographic traditions and practices as pursued in various societies around the Mediterranean basin. The length and depth of the history of these cross cultural encounters renders the notion of a history of cartography in well defined geographical and cultural spaces that formed separate entities difficult to sustain.

Cartographic products independent of manuscripts or printed books appeared primarily in cultural fields outside the sphere of education and knowledge production. Navigation, war, water administration, architecture, landscape design, religious ritual, and gift giving were such domains where self contained cartographic products are known to have existed in several Islamic societies or are preserved in libraries, museums, and collections across the world. Military maps gained increasing importance during the nineteenth century. In the twentieth century, private and state institutions for cartography and surveying were founded in Iran, Turkey, and Arab countries. They produced a broad variety of thematic maps, among them street maps, historical maps, meteorological maps, geological maps, and maps for tourism. These new cartographic products were supported by new methods and techniques of surveying, displaying, researching, measuring, and representing.

Cultures of Mapmaking

Mapmaking in Islamic societies took its themes, methods, techniques, and formats from several societies that either existed before the emergence of Islamic societies or parallel to them. The regions where these societies flourished were Greece, Egypt, Iran, India, China, Spain, Italy, the Netherlands, and France. With the shifts and changes in Europe since the sixteenth century, further European societies became relevant to cartographic activities in various Islamic societies, among them Russia, German countries, and Great Britain.

The two earliest appropriations of cartographic products and practices from other cultures, which shaped mapmaking in Islamic societies for more than 500 years, occurred during the eighth and ninth centuries in Abbasid Baghdad and other regions of the empire. One of the two cross cultural encounters is represented through the translations of Ptolemy's *Geography* into Arabic during the ninth century and their subsequent appropriation and adaptation to the interests of scholars. These translations were made in the circle of the Abbasid courtier and son of the Arab governor of Kufa, the philosopher Abū Yūsuf Ya'qūb b. Isḥāq al Kindī (d. *c.* 260 h/874), and the Sabean moneylender, mathematician, astronomer, philosopher, and theologian Thābit b. Qurra (d. 288 h/901). The other encounter had its basis in a series of maps of possibly Sasanian origin that were reproduced in Arabic, although it is unclear when this took place. These maps were adapted during the ninth and tenth centuries to the new political and religious powers and structures in Iran, the Arabian Peninsula, and Northern Africa. Step by step the original set of maps was enriched by maps of new regions of the Islamic world and accompanied by descriptions of the mapped territories and their cultures.

The two encounters led to different views on how to organize and represent the world and its parts. Maps in the tradition of Ptolemy's *Geography* worked with projections, latitudes and longitudes, grids, occasionally scales, and the theory of astronomically defined climates, mostly seven in number. They often contain short references to uninhabitable parts of the earth due to excessive cold or heat derived from Aristotelian physics and meteorology. The geographical space represented in such maps was first and foremost the entire world. It was mapped as a sphere or as a semisphere. Over time, such world maps focused more and more on the Islamic world. But certain regions of the world outside of Islamic societies continued to be reported about such as China, Śri Lanka, Byzantium, or the land of the Slaves. In addition to the world maps, maps of oceans, seas, rivers, and islands are extant from this cartographic encounter. These local maps show clear non Ptolemaic features. They may be the result of the mapmakers' efforts to translate textual and oral geographic information into images. The people involved in this cartographic tradition were primarily scholars of the mathematical sciences which comprised in Islamic societies in continuation of earlier Greek models number theory, geometry, theory of proportions, and astronomy. The most important representatives of this tradition in Islamic societies were the lost maps produced in the early ninth century for the Abbasid caliph al Ma'mūn (r. 198–218 h/813–33), the maps extant in Muḥammad b. Mūsā al Khwārazmī's (fl. late second early third century h/late eighth early ninth century) *Ṣūrat al 'arḍ* (image/form of the earth), the maps extant in copies of the new Arabic translation of Ptolemy's *Geography* made by Georgios Amirutzes (d. 1475) and his son for the Ottoman Sultan Mehmet Fatih (r. 849 h/1444 and 855–86 h/1451–81), and the maps found in the geographical chapter of texts on *'ilm al hay'a* (mathematical cosmography) such as Niẓām al Dīn al Nīsābūrī's (d. 730 h/1329) *Tawdīḥ* or *Sharḥ al Tadhkira* (explanation of or commentary on the 'Memoir') (**Figure 1**).

Maps derived from possible Sasanian models worked with geometrical figures such as circles, semicircles, crescents, triangles, rectangles, ovals, and straight lines to represent regional units, directions, and distances. Their focal points are the postal and pilgrimage routes within the Abbasid Empire and other Islamic states, for instance, in al Andalus and the main parts that formed the Abbasid caliphate. The views on what these parts were and which rank each part assumed in relationship to the others depended largely on the outlook the mapmaker and commentator subscribed to – a preference for Iranian cultural values, a preference for Islamic values, or a preference for his own environment and travel experiences. The set of maps created within this tradition included as a rule a world map and in most cases 21 regional maps. Most of the world maps in the extant manuscripts show the world as a full circle; but there are also manuscripts from as early as the tenth century and as late as probably the seventeenth or eighteenth century that show the world as a bird. The distribution of the regions over the body of the bird varies. China or Arabia can be found as the head. India, the land of the Khazars in the Northern Caucasus Mountains, Asia, or Africa fill the wings. The tail is taken up by Northern Africa or Europe. The world maps show cosmological concepts such as *al Baḥr al Muḥīṭ* (the Encompassing Sea) also known from much earlier cultures and religious concepts specific to the Qur'ān and other sources of Muslim belief such as the *barzakh*, a small, narrow land barrier that separates the two main oceans of the maps, that is, the Indian Ocean and the Mediterranean Sea. The main Quranic meaning of *barzakh* is, however, not geographical, but eschatological. It denotes the life of the soul after the death of the body until the Day of Judgment. The provinces and regions in these world maps are mostly registered in form of rectangular tables or stripes of boxes. They also include countries outside the Islamic world such as India or China. A few cities such as Mecca or Alexandria, rivers and occasionally their sources, islands and tribes are also noted. The regional maps focus entirely on the Islamic world. They map Iran, Central Asia, parts of the Caucasus Mountains, the Arabian Peninsula, Iraq, the Fertile Crescent, Northern Africa, and three seas (the Persian Sea that is the Indian Ocean; the Mediterranean Sea; the Caspian Sea). The sequence of the mapped regions and their relative delineation differs from my naming them in the sense that most extant manuscript maps show the following order and content: Arabia; Persian Sea; al Maghrib (Northwest Africa); Egypt; Syria; Mediterranean Sea; al Jazira (Northern Iraq and Southeast Turkey); Iraq (Central and South Iraq); Khuzistan; Fars; Kirman (parts of Iran); Sind (part of Pakistan); Armenia, Arran, Azerbaijan (part of Caucasus Mountains, Azerbaijan, and Iran); Jibal; Daylam, Rayy, Tabaristan (parts of Iran), Caspian Sea;

Figure 1 Ptolemy, *Geographike Hyphegesis*, *Arabian Peninsula*, Arabic translation by Georgios Amyrutzes, MS Istanbul, Süleymaniye Kütüphanesi, Aya Sofya 2610, pp 198 199. Courtesy of Süleymaniye Kütüphanesi.

Persian Desert; Sijistan; Khurasan (parts of Iran, Turk menistan, Pakistan, Afghanistan); and Mā warā' al nahr (part of Uzbekistan) (**Figure 2**).

The people creating maps and writing texts in this cartographic tradition were mostly members of the Ab basid administration for which they worked as officers of the postal routes or of the finances, philosophers, and scholars of religious disciplines, or wealthy, independent travelers and merchants. After its first famous prac titioner, the geographer, mathematician, astronomer, physician, philosopher, and student of al Kindī, Abū Zayd Aḥmad b. Sahl al Balkhī (d. 322 h/934), this tradition is called the Balkhī school of geography. Other major representatives of this school were Abū Isḥāq Ibrahīm b. Muḥammad al Iṣṭakhrī (fourth century h/tenth century), Abū l Qāsim Muḥammad b. Ḥawqal (d. *c.* 367 h/977), and Abū 'Abdallāh Muḥammad b. Aḥmad al Muqaddasī (d. *c.* 390/1000). The earliest maps sur viving from this body of works belong to a manuscript of Ibn Ḥawqal from 479 h/1086 extant in the Topkapı Sarayı Müzesi Kütüphanesi in Istanbul.

The most pervasive feature of maps in Arabic, Persian, and Ottoman Turkish after the centuries of translation, adoption, and adaptation is, however, their combination of more than one cultural tradition of cartography. World maps attached to geographical, cosmographical, or his torical texts combine between the eleventh and the sixteenth century elements of the Ptolemaic tradition with the components of the Balkhī school to which their makers added further elements taken from the Qur'ān and contemporary developments in politics, commerce, war, or demography. *Mount Qāf*, another Quranic concept, appears for instance as a mountainous ring enclosing the Encompassing Sea in maps joined to geo graphic manuscripts such as Abū 'Abdallāh Muḥammad b. al Sharīf al Idrīsī's (495– *c.* 560 h/1100– *c.* 1165) *Nuzhat al mushtāq fī 'khtirāq al āfāq* (*The Pleasant Journey of Those Who Desire to Travel through Faraway Lands*), to texts seen as popularizing scientific ideas such as Zakariyā' al Qazwīnī's (d. 682 h/1283) *Kitāb 'ajā'ib al makhlūqāt fī gharā'ib al mawjūdāt (Book of Marvels of Things Created and Miraculous Aspects of Things Existing*), and to texts of a

Figure 2 *Ṣurat al-Jazıra* (Image of the land between the Tigris and the Euphrates), MS Istanbul, Topkapı Sarayı Müzesi Kütüphanesi Ahmet III.1285, f 39a (dated 1285). Courtesy of Topkapı Sarayı Müzesi.

more religious character such as the *Kharīdat al ʿajāʾib wa farīdat al gharāʾib* (*The Unbored Pearl of Wonders and Precious Gem of Marvels*) attributed to Sirāj al Dīn Abū Ḥafṣ ʿUmar b. al Wardī (d. 861 h/1457). The maps to *The Book of Curiosities*, a twelfth century manuscript on astronomy, geography, commerce, and wondrous things copying a work of the eleventh century, include already the famous circular map ascribed to al Idrīsī. It also has a rectangular world map and maps of the Indian Ocean, the Mediterranean Sea, Cyprus, Sicily, maps of great rivers such as the Nile, the Euphrates, and the Tigris, maps of the former Fatimid capital al Mahdīya and the Egyptian town of Tinnīs, and maps of lakes. The author, in all likelihood a Fatimid custom officer, combined conceptual and technical elements of the Ptolemaic tradition such as climate zones, geographical coordinates, a scale, and the configuration of the Old World with symbolic features from al Khwārazmī's maps such as the depiction of the Mountains of the Moon as sources of the Nile and conceptual and representational elements of the Balkhī school such as the representation of rivers by straight lines or lakes by circles to which he added information from travelers, texts, and perhaps administrative documents. He chose some unique and not yet well understood forms of representation for the Mediterranean Sea and the Indian Ocean (an oval) and for Cyprus (a rectangle divided into smaller units of rectangles and text) (**Figure 3**).

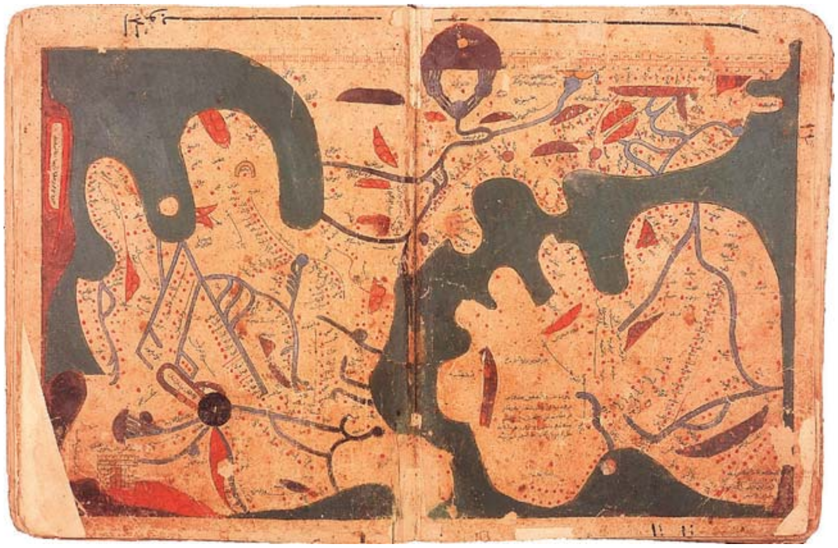

Figure 3 *Book of Curiosities*, Rectangular World Map, MS Oxford, Bodleian Library, Ar.*c*.90, ff 23b 24a. Courtesy of Bodleian Library.

Al Sharīf al Idrīsī and his geographical and cartographic work represent another kind of cross cultural encounter since he lived in more than one of the Mediterranean cultures. He was born in Ceuta, received his education in al Andalus, and traveled widely in the Islamic world. Then he worked for two Norman kings, Roger II (r. 1097–1154) and William I (r. 1154–66) at their court in Palermo. At the end of his life, he returned to the Islamic world, possibly to Ceuta. Conceptually, he followed Ptolemaic cartographic theory and methods. He produced two sets of new maps. One was a circular world map engraved on a silver plate, which is lost. The other is a set of 70 regional maps drawn on paper. They are extant in various manuscripts across the world. Al Idrīsī planned to combine them into one large rectangular map of the known world stretching from 180° in the east in the Sīlā Island (Korea) to the prime meridian in the west running through the Fortunate Isles. Presenting Roger not only as a successful conqueror and master of politics, he described him as the head of the cartographic project who decided on what geographical and cartographic truth was. One method applied for sorting out conflicting information was to trace with iron instruments on a drawing board all data on longitudes, latitudes, and distances. Al Idrīsī wrote two, perhaps even three geographical texts, all of them lavishly illustrated with regional maps portraying the entire world in a combination of Ptolemaic concepts with ideas and symbols taken from the Balkhī school. He also had access to non Arabic oral and perhaps written sources and provided fresh information about the physical and cultural geography of European lands, not available previously in Arabic.

Almost all of Ottoman mapmaking is best described as a highly flexible merging of cartographic and artistic languages from all major cultural traditions the Ottoman Empire drew upon, cooperated and traded with, and fought against. While these mergers are often seen as a reflex of Ottoman cultural and scholarly inferiority, the cultural dynamics and force that brought forth such mergers cannot be overrated. If nothing else, the people involved in producing such eclectic maps were capable of reading and understanding different geographical concepts, philological descriptions, and iconographic representations. A fair amount of concrete cross cultural work was needed in order to interpret foreign maps and make their copies and translations into Ottoman Turkish acceptable for the local consumer. As is well known and can be demonstrated for a number of the preserved Ottoman maps designed from the sixteenth to the eighteenth centuries, even the copyists participated in this process of cross cultural adaptation.

Four major cultures of mapmaking arose in the Ottoman realm. Two of these cultures, the making of portolan charts and the creation of maritime handbooks with maps of Mediterranean islands, coasts, and fortresses, were closely interconnected. They combined several previous mapmaking traditions from different Mediterranean cultures, mainly Arabic, Catalan, and Italian portolan charts, regional maps, and *isolarii*. They were apparently primarily designed by members of the Ottoman navy. Scribes and painters from imperial as well as commercial workshops participated in executing the fine copies. Specialized mapmaking workshops existed in the middle of the seventeenth century in Istanbul. Almost all extant specimens were made for acquiring courtly patronage or to be given away as gifts. In particular the maps illustrating the various copies of Pīrī Re'īs' (*c*. 875–961 h/1470–1554) two versions of the *Kitāb i Baḥrīye* (Maritime Handbook) are wonderful products of art. Some of them show the inclusion of Renaissance painting techniques, while others are splendid experiments in colors and symbolic interpretations of physical objects (**Figure 4**).

The two other main Ottoman mapmaking cultures were situated directly in the courtly sphere. One of them consists of town views and itineraries in official histories mapping Ottoman military campaigns, adversaries, and imperial architecture. Their execution was the task of the painters of the imperial painting workshops. As a result, many are superb exemplars of Ottoman fine arts. (**Figure 5**).

The other mapmaking culture began to appear during the sixteenth century when Arabic and Persian models of world maps became to be replaced gradually by models from Italy and then from the United Provinces, France, and German countries. Partial or full translations of books on the New World, a Latin edition of Ptolemy's *Geography*, Abraham Ortelius' *Theatrum Orbis Terrarum*, Gerard Mercator's and Henricus Hondius' *Atlas Minor*, Willem J. and Joan Blaeu's *Atlas Maior*, and other geo graphical and cartographic works produced in different Catholic and Protestant countries of Europe were undertaken by Ottoman scholars, often in collaboration with members of the religious minorities living in the Ottoman Empire or with converts from Catholic or Protestant Europe. The maps of these books had a pro found impact on Ottoman mapmaking. The changes, however, did not constitute a unilateral Westernization,

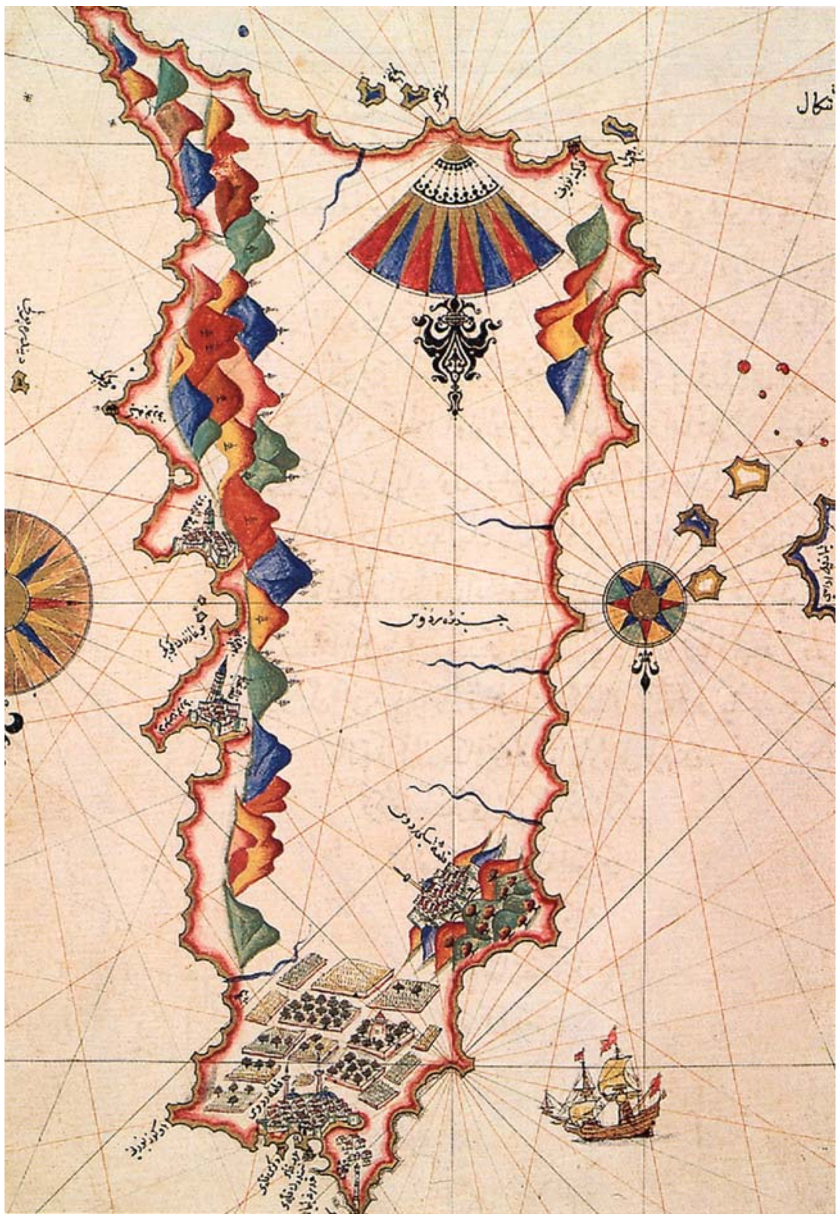

Figure 4 Piri Re'is, *Kitab-i Baḥrıye* (Maritime Handbook), The Island of Rhodes, MS Istanbul, İstanbul Üniversitesi, T6605, f 100a. Courtesy of İstanbul Üniversitesi.

Figure 5 Szigetvár, Ahmed Feridun, *Nüzhet Esrar el-Ahbar der Sefer-i Sigetvar* (Present of the Secrets of Information on the Journey to Szigetvár), MS Istanbul, Topkapı Sarayı Müzesi Kütüphanesi H.1339, ff 32b 33a (copied in 1568-69). Courtesy of Topkapı Sarayı Müzesi.

Figure 6 Willem and Ioan Blaeu, *Atlas Maior*, Translation ascribed to Abu Bakr al-Dimashqı, *Asia*, MS Istanbul, Süleymaniye Kütüphanesi, Nuruosmaniye 2995, before f 104a. Courtesy Süleymaniye Kütüphanesi.

to use an anachronistic term. Translators, scholars, copyists, calligraphers, painters, and trainees, and tea chers of the Ottoman army transformed the translated maps in a variety of ways into maps accessible to Otto man spectators and situated them firmly within Ottoman visual, geographic, and literary cultures (**Figure 6**).

Cartographic Genres and Practices

In addition to the world and regional maps, portolan charts, and handbooks for sailing directions discussed in the previous section, a number of other cartographic formats were developed in Islamic societies. Early in the ninth century, or perhaps even in the eighth century, the question of how to determine from any given place the direction toward Mecca was raised. David A. King claims that there were two cartographic worlds in Islamic societies with respect to the *qibla.* Mathemat icians, astronomers, and mathematical geographers inhabited one of these two worlds. They developed trigonometric, projective, arithmetic, and other exact or approximate methods for solving the problem (**Figure 7**).

Legal scholars, students of the religious traditions, descriptive geographers, folk astronomers, and cos mographers who focused on miracles, wonders, and narrated histories populated the other. They relied on *ḥadīth*, a body of religious teachings about what the Prophet and his companions had said and done, on astronomical and meteorological alignments ascribed to the Kaʿba since pre Islamic times, and on practices of travelers.

Among the various methods used for mapmaking, grids, coordinate frames, and scales played roles that are not always easily discernible. Some mapmakers applied them as mathematical tools, while others apparently saw in them primarily symbolic statements about the possi bility of mapping the world mathematically. Grids were used in rectangular and circular maps. Rectangular grids were applied in world maps such as the one made for al Ma'mūn and by Suhrāb (fl. first half fourth century h/ tenth century), for regional maps and world maps illus trating historical works or encyclopedias such as the *Masālik al abṣār fī mamālik al amṣār* (*Ways of Perception Concerning the Most Populous [Civilized] Provinces*) compiled by the Mamluk judge and administrator Ibn Faḍlallāh al ʿUmarī (691–749 h/1301–49). Suhrāb described his

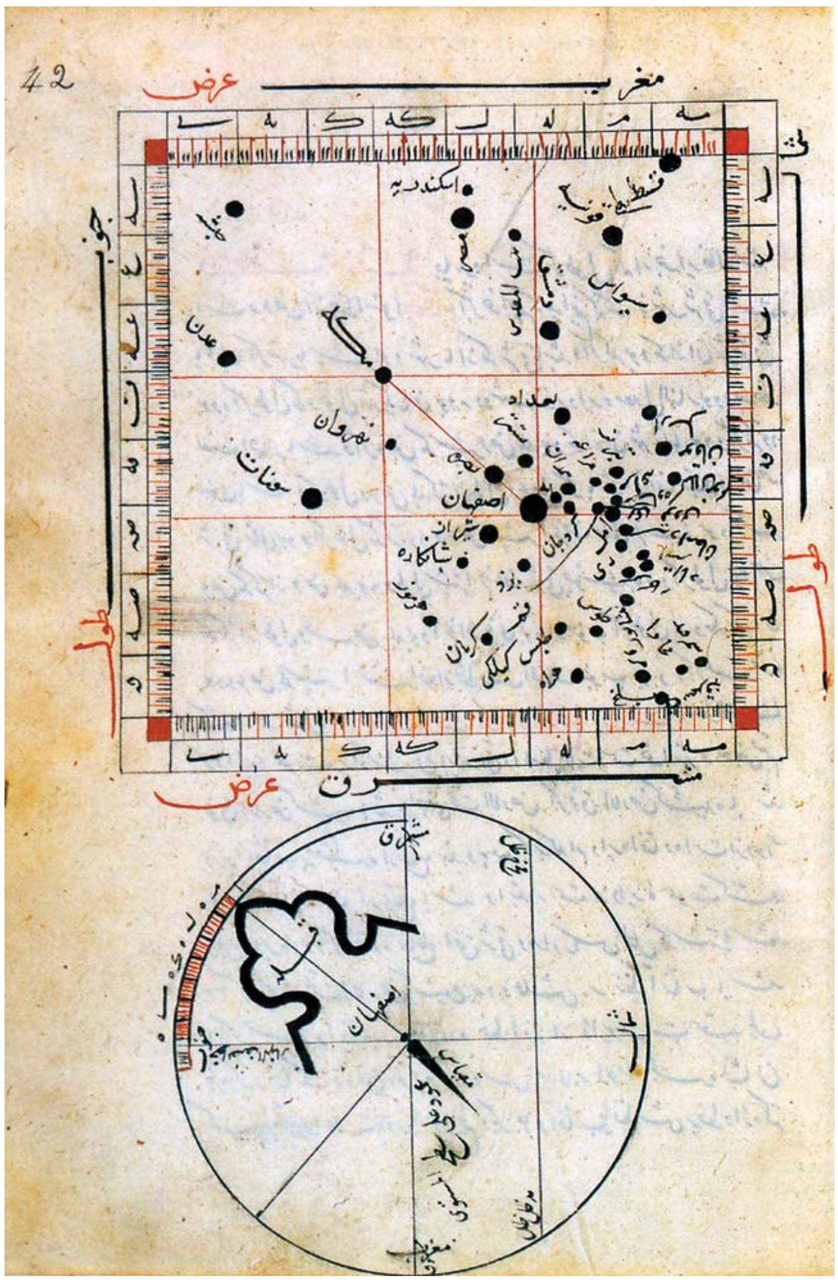

Figure 7 *Qibla* Map, MS Paris, BnF Perse 169, f 42a. Courtesy Bibliothèque nationale de France.

method as starting with a rectangle, followed by dividing the edges into degrees and marking the equator. Then he drew the horizontal lines that separate the climates. For positioning the localities he stretched two threads over the rectangle, one horizontally, the other vertically. Lost world maps made before 1500 may have worked with curvilinear grids as does the only extant exemplar of such a type attached to a manuscript of al ʿUmarī's work. The meaning of this grid has been interpreted differently by Fuat Sezgin who believes the grid to be a stereographic projection and the map to represent the map made for al Ma'mūn and by David A. King who points to the several errors in the grid and al ʿUmarī's explanation and refers to historical evidence contradicting the identification of this map with that made for al Ma'mūn.

Coordinate frames and scales appear early on maps made in Islamic societies. Several maps, however, are inscribed merely with one set of the coordinates, mostly longitudes, and the placement of the equator on which they are noted varies. Scales too show contradictory in formation. Such features are often taken to illustrate a lack of mathematical knowledge on the side of the mapmakers. The quality and care invested in some spe cimens of this kind and the people involved in the pro duction of the manuscripts where they are found make such a reading unlikely. An example is the world map attached to one of the anthologies produced for the Timurid prince Iskandar Sulṭān (r. 812–17 h/1409–14).

Simple circular as well as quadratic schemes were used for organizing geographical knowledge for religious and divinatory purposes. Such diagrams could evolve into small scale maps since they linked either the Kaʿba or major religious monuments in Mecca with the indi vidual sectors of the circle and the regions or towns in scribed in them. They often group a multitude of localities behind one specifically chosen town whose *qibla* serves as a marker or representative for all the other directions. The effect of such an arrangement was di dactic and mnemonic in the same time. The earliest known scheme of this kind is found in a manuscript copy of Ibn Khurradādhbih's (third century h/ninth century) *Kitāb al masālik wa'l mamālik* (*Book of the Routes and Prov inces*). It consists of four sectors of the world, each one of them being connected with a part of the Kaʿba – Maghrib, Egypt, Syria, al Jazīra; Iraq, Armenia, Kashmir; Manṣūra, Tibet, China; Yemen. Over time, the schemes became more elaborate and could comprise as many as seventy two divisions. In addition to the schemes, maps and in struments for finding the *qibla* were created. The most sophisticated, Mecca centered map was engraved on three astrolabes made by Safavid instrument makers in the late seventeenth century.

Climatic zone charts are a further type of diagram matic schemes that were widely used in geographical, cosmographical, astronomical, and historical treatises and dictionaries for illustrating the theoretical division of the earth as taught by natural philosophy and astronomy. Its widespread appearance across the disciplines and literary genres is but one indicator of the depth at which the intellectual outlook in Islamic societies was permeated by scientific elements of ancient Greek origin. A very simple exemplar of these kinds of diagrams is found in Sharaf al Dīn al Masʿūdī's (545–after 613 h/1150–after 1216) *Jahān Dānish* (*Knowledge of the World*), a book on astronomy, as trology, and geography written in (549 h/1154–55). Three unnamed circles enclose eight parallel, equidistant lines identified at the right as demarcating the seven climates. Geographical features such as oceans, islands, and countries are sparsely marked by words alone. Nu merous Arabic, Persian, and Ottoman circular and semicircular world maps also show such a scheme of climatic zones. The division into climates became the most important format of structuring the physical world in maps as well as texts. It was applied to maps of the Ptolemaic tradition, maps working within the tradition of the Balkhī schools, and maps derived from early modern European maps. Examples are the semicircular world map in ʿAyn al Zamān's Ḥasan b. ʿAlī Qaṭṭān Marwazī's (465–548 h/1072–1153) *Gayhān Shenākht* (*Constructing the World*) and the circular maps for Iskandar Sulṭān, the *Book of Curiosities* and Mehmet b. ʿAlī Sipāhī zādeh's (d. 997/ 1589) Turkish translation of his *Awḍaḥ al masālik fī maʿrifat al buldān wa'l mamālik* (*Explanation of the Routes about the Knowledge of Places and Provinces*).

Functions and Purposes

Maps, charts, plans, diagrams, and tables served a variety of functions and purposes in Islamic societies although the bulk seems to have spent their life in manuscripts, that is, in the spheres of educated readers. The earliest usage of maps seems to have happened when the armies of the newly converted Arab tribes invaded regions of Iran and Central Asia. The governor of the eastern parts of the Umayyad Empire, al Ḥajjāj b. Yūsuf (d. 95h/714), is said to have ordered a map of Daylam, a mountainous region south of the Caspian Sea, before attacking it in *c.* 83 h/702. Five years later, when planning to lay siege at Bukhara, he ordered a plan of the city. Plans or sketches are also reported for planning in 141 h/758 the new capital Baghdad of the Abbasid caliphate and possibly for waterworks in the swamps near Basra made for the sec ond Abbasid caliph al Manṣūr (r. 136–58 h/754–75). The next time that sources report of maps, sketches, or plans for military purposes occurs in the Ottoman Empire. Laying siege to cities and fortresses such as Constan tinople, Szigetvār, Vienna, Baghdad, or the fortresses of the Order of St. John at Rhodes and Malta included the drawing of sketches and plans, while the actual warfare made use of compasses for tunneling the walls. It is only after the treaty of Karlowitz in 1699, that maps became an important tool for the Ottoman army for settling boundary disputes and negotiations over lost territories.

Tax collection and market regulation, on the other hand, were subjects that apparently rarely worked with maps, and the manuals written about these themes were seldom illustrated with more than the simplest diagrams. This is, according to Ahmet T. Karamustafa, even true for the Islamic society with the broadest use for maps and the widest array of types of maps – the Ottoman Empire. Different bodies of the Ottoman state used maps and plans that visualized waterways and channels, architecture, and navigation in the Mediterranean and the Black Sea.

A central function of maps in Islamic societies was to indicate religious beliefs, support the believer in carrying out her or his religious duties, and help her or him remember the visits to the Holy shrines. A further religious theme illuminated by maps relates to the Prophet's *mi'rāj* (ascent to the heavens). Maps functioned as instruments, tools, and pictorial narratives. They were turned into souvenirs and certificates (**Figure 8**).

Maps also served as objects of art and diplomacy. Fatimid world and regional maps were produced on costly silks embroidered with gold and colored silk threads. Consequently, they were stored with other valuables in the princely treasury rooms. The anonymous author of the *Kitāb al Hadāya wa'l Tuḥaf* (*Book of Gifts and Rarities*) when describing the treasures stored by the Fatimid rulers exclaimed that among them several hundreds, close to a thousand silk drapes were found that represented kingdoms with their kings. He marveled at the precious material used for this display of royal splendor and possession and emphasized that these drapes showed a portrait of every king, his name, the span of his reign, and an account of his life. At least four more courts are known where maps were either painted on silk or engraved on silver plates – the court of the Khwarazmshah Sultan 'Alā' al Dīn Muḥammad (r. 596–617 h/1199–1220) in Khiva, the court of the Ilkhanid ruler Ghāzān Maḥmūd (r. 694–703 h/1295–1304) in Tabriz, the Norman court of Roger II (d. 1154) in Palermo, and the Ottoman court in Istanbul (857–1341 h/1453–1922). Except for Ottoman maps on silk, none of these maps are extant.

In various contexts, maps were treated like miniature paintings. Outstanding examples are the world map for Iskandar Sulṭān mentioned above, maps in a manuscript of the *Beyan i Menazil i Sefer i 'Iraḳeyn i Sulṭān Süleymān Ḫān* (*Explanation of the Stops in Sultan Süleyman Khan's Travel to the Two Iraqs*) (944 h/1537–38) by Maṭrakçī Naṣūḥ (d. 971 h/1564) commemorating Sultan Süleyman's (r. 926–74 h/1520–66) successful military campaign of 941 h/1534–35 against his Safavid neighbor Shāh Ṭahmāsb (r. 930–84 r/1524–76), and maps in eighteenth century manuscripts of Ḥajjī Khalīfa's *Cihānnümā* (version II).

A fourth function of maps was their capacity to memorize historical events and portray princely power and splendor. The most splendid examples of maps as constructs of historical memory are found in books describing and eternalizing the feats of the Ottoman sultans. Many of the views of cities and ports found in Ottoman histories since the early sixteenth century celebrate either their conquest by an Ottoman sultan or an attack by a fleet in Ottoman service. According to J. M. Rogers they were also meant to visualize "the pomp and circumstance of the Ottoman court". The composition of the views usually includes a representation of an abstract landscape – some mountains, some flowers, some herbs, occasionally also trees and animals. This theme of the landscape, its elements, and its colors is borrowed from miniature painting. The major difference between the maps and the miniatures is the complete lack of human beings in almost all maps of this genre.

A different, but equally splendid construction of imperial allegory can be found in miniatures produced for the Mughal rulers Jahāngīr (r. 1014–37 h/1605–1628) and Shāh Jahān (r. 1037–1068 h/1628–1658). Both rulers were keen to express their imperial claims by relying symbolically on the image of a globe with or without depictions of lands and seas. A series of images painted by Muslim and Hindu court painters between 1614 and 1658 stress the legitimacy of the ruler as a dutiful son, just defender of the interests of his subjects, and pious follower of the saints. Jahāngir also used the globe for depicting his relationship to his most powerful Muslim neighbor outside of India, the Safavid Shāh 'Abbās I (r. 986–1038 h/1587–1628). The portrait, called *Jahāngīr's Dream*, shows Jahāngīr as the superior and benign ruler who is connected by friendship with his smaller neighbor. Both stand on a world map that reflects knowledge of Western cartographic developments. Two animals rest at the feet of the two rulers, a lion and a lamb. The juxtaposition of lion and lamb symbolized the peaceful coexistence of natural enemies, of irreconcilable contradictions, hence the highest good a just ruler could achieve. The symbolic globe also figures in other portraits of Jahāngīr, which describe his successful squelching of Khurram's rebellion, the later Shāh Jahān, and his desire to destroy his political and military rival Malik 'Ambar, minister to the rulers of Aḥmadnagar from 1596 to 1626. The minister valiantly opposed Mughal imperial ambitions in the Deccan and pursued a policy of steady recovery of territory conquered by Shāh Akbar late in the sixteenth century. It was only 'Ambar's death in 1626 that opened the Deccan to Mughal invasion. The miniature of Jahāngīr shooting at a beheaded and speared Malik 'Ambar while standing on a complex cosmological allegory composed of a terrestrial globe seated both on an

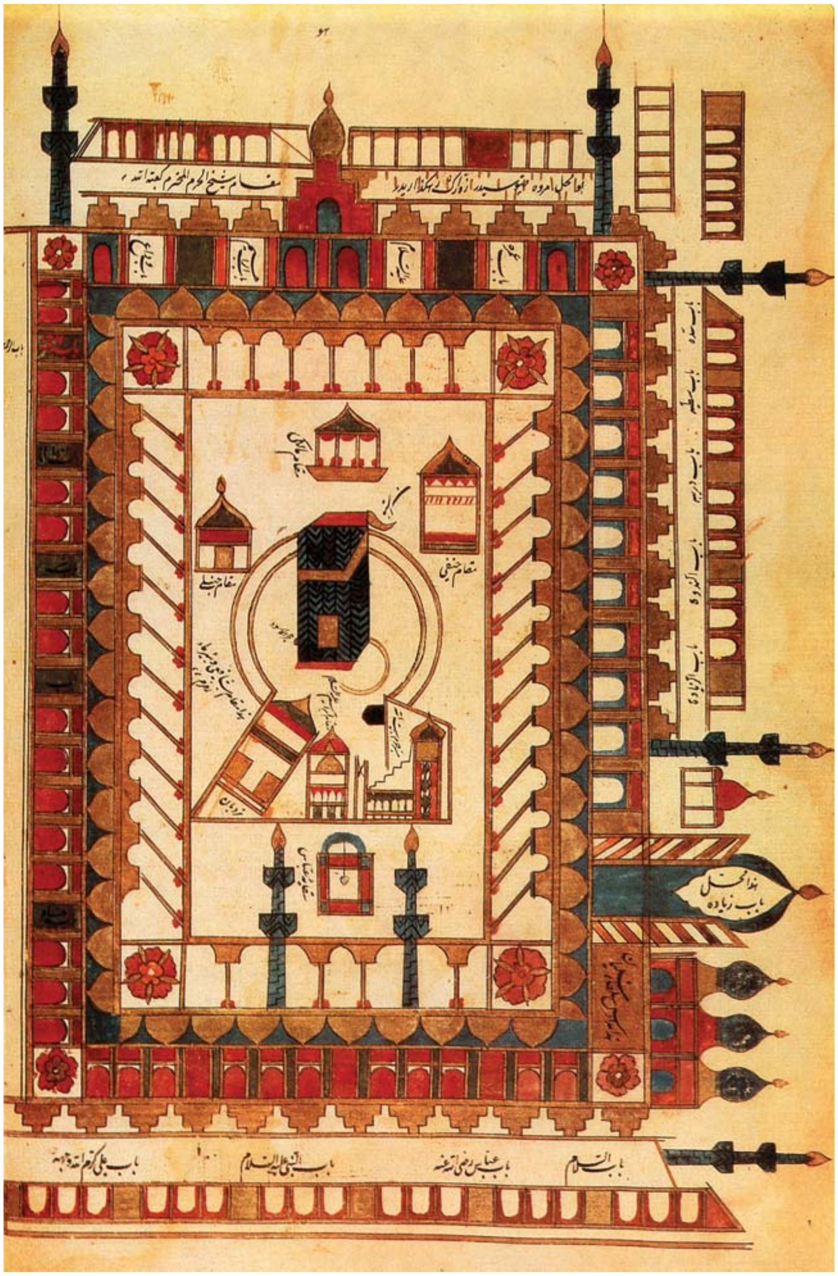

Figure 8 View of Mecca and its pilgrimage stations, MS Jena, University Library Prov. q. 215, f 6b. Courtesy of the University Library of the University of Jena.

elephant and a huge whale surrounded by a wealth of other symbolic elements expresses the magnitude of emotions roused in Jahāngīr by the obstacle the Deccani minister of Ethiopian origin posed to fulfilling his imperial dream.

Most Islamic maps before the eighteenth century ignore political space in favor of cultural space. Elements that may signal political entities such as boundaries, emblems, names of dynasties, flags, or monuments rarely occur at all on Islamic maps before 1700. Maps of the so called Balkhī school indicate boundaries (*ḥudūd*) explicitly, but they are not boundaries separating different states or kingdoms, but boundaries dividing the Islamic world into administrative subunits. In contrast to geographical writings, which even stressed the linkage between geography and history, maps produced in Islamic societies before the sixteenth century did not express this disciplinary neighborhood.

A similar reluctance to deal with the political composition of the world can be seen in maps illustrating the manuscripts of Ḥajjī Khalīfa and Abū Bakr al Dimashqī (d. 1102 h/1691). In the autograph of the translation of the *Atlas Minor*, Ḥajjī Khalīfa and his French partner Mehmet Ikhlāṣī followed closely the views about the political division of Europe and Asia as propagated by the Dutch cartographers Gerard Mercator and Henricus Hondius. In the autograph of Ḥajjī Khalīfa's own geo graphical work for which he had engaged in the project of translation, the *Cihānnümā* (version II), the Ottoman Empire is not mapped at all and the Safavid Empire has disintegrated into a series of maps reflecting views of Iranian provinces held in pre Safavid geographical lit erature. The fine copy of Abū Bakr al Dimashqī's trans lation of the *Atlas Major* does not contain any map of the Ottoman or the Safavid Empires. Later paraphrases of the work map the two Muslim neighbor states as found in Western atlases, that is, the Ottoman Empire is clipped to Anatolia and the Safavid Empire has borders that fit more the situation of the early nineteenth than that of the late seventeenth or early eighteenth centuries when these maps were painted in Istanbul. Possibly the first Ottoman map that names Anatolia and even parts of Iran *Memleket e devlet e ʿaliye* (*Kingdom of the Supreme Dynasty*), but con tinues to disregard the matter of precise boundaries of the empire in contrast to other states in Europe, is a silk map painted in Istanbul in 1182 h/1768. This map too follows without further reflection the structuring of the political space as seen in Catholic and Protestant Europe. A rupture with this disrespect for the political status and extension of the Ottoman state in Ottoman mapmaking occurred only in the nineteenth century, when the modernized Ottoman army and its Western advisors took a fresh interest due to military and economic reasons in mapping the Ottoman territories.

Maps of nineteenth century Qajar Iran likewise do not show any explicit interest in mapping the political space of the country. Their authors mostly focused on mapping particular provinces and their physical and cultural properties. Qajar maps show clearly that by now mapmaking in Iran followed the map culture of Europe. In the first half of the nineteenth century, British and Russian officers toured the country for mapping several of its provinces, mountains, and shores, partly on order of the Qajar court. In the second half of the nineteenth century, Iranian mapmakers who were trained in modern mathematical and cartographic methods and concepts at the *Dār al funūn* in Tehran started to work. The *Dār al funūn* was a new type of school for higher education founded in 1851 by Mīrzā Taqī Khān Amīr i Niẓām (1186 – assassinated in 1231 h sh/1807 – assassinated in 1852), a diplomat and highly disputed prime minister of the Qajar ruler Naṣīr al Dīn Shāh (r. 1227–75 h sh/1848–96). The new school was part of efforts to reform the Qajar army. European teachers taught new military methods, theories, technologies, sciences including geography and cartography, and languages. The major cartographic works undertaken by Austrian professors and Iranian students and later professors of the *Dār al funūn* were explorations, triangulations, descriptions, and drawings of boundaries between Iran, the Ottoman Em pire, Central Asia, Afghanistan, and Northwest India (today Pakistan) and various kinds of mapping enterprises about Tehran and the settlements in its immediate en vironment. A number of the engineers involved in these different projects studied for some time in France or England and worked for different departments of the government and the army. In addition to mapmaking, they wrote books about geography and mathematics (**Figure 9**).

Cartography and Surveying in Iran in the Twentieth Century

In the twentieth century, cartography developed in Iran in three sectors – private cartographic and surveying companies, the army, and governmental civil institutions. The first private cartographic company, Sahab Geo graphic & Drafting Company, was founded in 1315 h sh/1936 by ʿAbdolghāsem Saḥāb (1265–1335 h sh/1886–1956) and led for most of the century by his son ʿAbbās Saḥāb (1300–79 h sh/1921–2000), who was trained as an engineer. It began with publishing maps of Iranian history and physical geography. After 1945, it also produced road maps, maps for tourism, globes, maps of Iran's neighbors, atlases, posters, and school books. Ebādolāhe Bakhtiārī (1306–62 h sh/1927–83), the husband of a sister of ʿAbbās Saḥāb, founded in 1351 h sh/1972 the second private cartographic publishing house in Iran, Gitashenasi Geographical and Cartographic Institute. Being today the more active of the two, it has a similar range of products, including political, physical, historical, and topographical maps of Iran, its provinces and cities, road and street maps, maps for tourism and outdoor activities, maps of Asian countries, globes, books, and postcards. Since some time, the company produces its maps with computer technologies.

In the field of geophysics and surveying, two men shaped the face of modern practices and products in Iran – Maḥmūd Ḥisābī (1282–71 h sh/1903–92) and Taqī Riyāḥī (d. after 1989). Ḥisābī, a student of Albert Einstein, was educated as a civil engineer in Beirut and Paris and did his PhD in physics at the Sorbonne. Starting in 1928, he surveyed Iranian coastal roads and was a founder or co founder of a number of institutions important for Iranian cartography, among them the country's first weather sta tion (1931), Tehran University (1934), the Institute of Standards of Iran (1954), and the Geophysical Institute of Tehran University (1961). Ḥisābī worked for Mossadegh's government as the first general manager of the Iranian National Oil Company. Riyāḥī was a civil engineer who had studied at the *École Polytechnique* in Paris. In 1952, he

Figure 9 *Naqsh-i Faris* (Map of Persia), MS Berlin, Preussischer Kulturbesitz, Staatsbibliothek, Kartensammlung D9515; Courtesy of Preussischer Kulturbesitz, Staatsbibliothek.

became Mossadegh's chief of staff. After the CIA coup in 1953, Riyāḥī was sentenced to 2 years of prison. In 1979, he worked for the head of the interim government after the fall of the Shah, Mahdī Bāzargān (d. 1995), but left the country soon after to live until his death in France. In 1956, Riyāḥī founded with a partner the first private surveying company of Iran, *Cherkate e Nesbi Naghsheh Bardari*. The company began with land surveying in the year of its foundation, but quickly expanded into other fields, among them photogrammetry (1959) and terrestrial photogrammetry. It cooperated closely with the French company *La Foncière* registered in the French protectorate of Morocco. About 20 French engineers and technicians came to Iran providing expertise and bringing with them needed instruments for surveying and mapmaking. The company sent *c.* 25 of its employees for education to France, among them Naṣr Ghazālī, today's head of the company, who had to leave Tehran University due to political difficulties under the Shah; five or six to the Netherlands; and three to Switzerland. Renamed in 1967 as Rassad Surveying Company, it moved into hydrography, microgeodesy, orthophoto mapping, digital mapping, GIS design and implementation, tunnel drilling control, and industrial control surveying. Twice, it mapped half of Iran, the first time using photogrammetric methods, the second working with satellite material. Its main clients are oil and road companies, the municipalities of Tehran, Yazd, and Kish, and various ministries.

In 1921, the army founded its first group for cartography. The first tasks the engineers of this department carried out consisted in translating, correcting, and completing Russian and English maps of Iran and adjacent countries. Fourteen years later, the first books on Iranian geography illustrated with military maps were produced for educational purposes. Subsequently, the cartographic organization of the army began creating its own military maps. As in the nineteenth and early twentieth century, a major focus of military mapmaking was the exploration of the borderlands. As the National Geographic Organization, it is now the most powerful cartographic body in Iran and possesses the most advanced technology, for instance, for area photography.

In 1953, Iran founded its first governmental institution for civil cartography, the National Cartographic Center (NCC). It started in 1948 as an engineering department of the Plan and Budget Organization. Working with foreign consultants and army officers, it carried out surveying work and planned the structure for the new governmental institution for mapping Iran. It focused its activities on three major domains. The first domain embraces the creation of terrestrial maps of various provinces, regions, and cities of Iran and works for producing a comprehensive map of the entire Iranian territory. NCC brought foreign engineers and technicians, mainly from France and the Netherlands, for surveying the country and trained students of mathematics and physics in cartographic methods and techniques. The second domain covers the charting of Iranian coastal and maritime areas, in particular in the Persian Gulf. The third major domain lies in creating the intellectual infrastructure for cartography in Iran. NCC created a research journal, organized conferences, published textbooks and the series National Atlas of Iran, and sponsored the translation of foreign publications. It has two special responsibilities – the supervision of all government contracts and area photography.

From 1941 onward, state bodies such as the Mining Administration, the Ministry of Agriculture, or the Ministry of the Roads ordered engineers to map the entities under their responsibility. Other governmental organizations, in particular institutes of Tehran Uni versity, the Geological Survey of Iran, and the Me teorological Survey of Iran, founded mostly in the second half of the twentieth century, produce specialized thematic maps on geology, geomorphology, geophysics, geochemistry, meteorology and soil or land use.

See also: Art and Cartography; Mapping, Non-Western; Maps and the State; Medieval Geography; Middle East and North Africa; Orientalism.

Further Reading

Ahmad, S. M. (1992). Cartography of al Sharıf al Idrısı. In Harley, J. B. & Woodward, D. (eds.) *History of Cartography, vol. 2, Bk. 1: Cartography in the Traditional Islamic and South Asian Societies*, pp 156 174. Chicago, IL: University Press of Chicago.

‘Ayn al Zaman’s Ḥasan b. ‘Alı Qaṭṭan Marwazı, *Gayhan Shenakht*. Qom: Vaḥed e entesharat e kitabkhane ye bozorg e Ayatollah al ‘uẓma Mar‘ashı Najafı.

Brentjes, S. (2004). Mapmaking in Ottoman Istanbul between 1650 and 1750: A domain of painters, calligraphers, or cartographers? In Imber, C., Kiyotaki, K. & Murphey, R. (eds.) *Frontiers of Ottoman Studies*, vol. 2, pp 125 156. London: I. B. Tauris.

Brentjes, S. (2007). Naṣr Ghazalı, Rassadiran, Interview, Tehran, March, 25, 2007.

Edson, E. and Savage Smith, E. (2004). *Medieval Views of the Cosmos: Picturing the Universe in the Christian and Islamic Middle Ages*. Oxford: The Bodleian Library, Oxford University.

Ghada al Hijjawi al Qaddumi, (1996). *Book of Gifts and Rarities*. Harvard Middle Eastern Monographies XXIX, Harvard UP, Boston.

Goodrich, T. D. (1990). *The Ottoman Turks and the New World: A Study of Tarih i Hind i Garbi and Sixteenth Century Ottoman Americana*. Wiesbaden: Harrassowitz Verlag.

Hagen, G. (2003). *Ein Osmanischer Geograph bei der Arbeit. Entstehung und Gedankenwelt von Katib Čelebis Ğihannüma*. Berlin: Klaus Schwarz Verlag.

Karamustafa, A. T. (1992). Military, administrative, and scholarly maps and plans. In Harley, J. B. & Woodward, D. (eds.) *History of Cartography, vol. 2, Bk. 1: Cartography in the Traditional Islamic and South Asian Societies*, pp 209 227. Chicago, IL: University Press of Chicago.

King, D. A. (1999). *World Maps for Finding the Direction and Distance to Mecca. Innovation and Tradition in Islamic Science*. Leiden: Koln: Brill.

King, D. A. and Lorch, R. (1992). Qibla charts, Qibla maps, and related instruments. In Harley, J. B. & Woodward, D. (eds.) *History of Cartography, vol. 2, Bk. 1: Cartography in the Traditional Islamic and South Asian Societies*, pp 189 205. Chicago, IL: University Press of Chicago.

MS Istanbul, Suleymaniye Library, Mihrişah Sultan 304.

Orientalische Buchkunst in Gotha (1997). Ausstellung zum 350jahrigen Jubilaum der Forschungs und Landesbibliothek Gotha. Spiegelsaal 11. September 1997 bis 14. Dezember 1997. Forschungs und Landesbibliothek Gotha, 1997.

Rogers, J. M. (1992). Itineraries and town views in Ottoman Histories. In Harley, J. B. & Woodward, D. (eds.) *History of Cartography, vol. 2, Bk. 1: Cartography in the Traditional Islamic and South Asian Societies*, pp 228 255. Chicago, IL: University Press of Chicago.

Sezgin, F. (1987). *Contribution of the Arabic Islamic Geographers to the Formation of the World Map*. Frankfurt am Main: Institut fur Geschichte der Arabisch Islamischen Wissenschaften.

Sharaf al Dın Muḥammad b. Mas‘ud Mas‘udı, (2003). *Jahan Danish* (edited and introduced by Zanjani, D. A.). Tehran: Miraş e Maktub.

Soucek, S. (1992). Islamic charting in the Mediterranean. In Harley, J. B. & Woodward, D. (eds.) *History of Cartography, vol. 2, Bk. 1: Cartography in the Traditional Islamic and South Asian Societies*, pp 265 284. Chicago, IL: University Press of Chicago.

The National Atlas of Iran. Maps. vol. 18, The Islamic Republic of Iran, Management and Planning Organization, National Cartographic Center, Tehran, 1371 sh.

Tibbets, G. R. (1992). Later cartographic developments. In Harley, J. B. & Woodward, D. (eds.) *History of Cartography, vol. 2, Bk. 1: Cartography in the Traditional Islamic and South Asian Societies*, pp 137 155. Chicago, IL: University Press of Chicago.

Tibbets, G. R. (1992). The Balkhı school of geographers. In Harley, J. B. & Woodward, D. (eds.) *History of Cartography, vol. 2, Bk. 1: Cartography in the Traditional Islamic and South Asian Societies*, pp 108 136. Chicago, IL: University Press of Chicago.

Tibbets, G. R. (1992). The beginnings of a cartographic tradition. In Harley, J. B. & Woodward, D. (eds.) *History of Cartography, vol. 2, Bk. 1: Cartography in the Traditional Islamic and South Asian Societies*, pp 90 107. Chicago, IL: University Press of Chicago.

Relevant Websites

http://cosmos.bodley.ox.ac.uk
Medieval Views of the Cosmos, The Book of Curiosities.

http://www.ncc.org.ir
National Cartographic Center.

Cartography, History of

C. Delano-Smith, University of London, London, UK
R. J. P. Kain, University of Exeter, Exeter, UK

The ubiquity of maps today is beyond dispute. They are treasured as decoration in private and public places, and enjoyed in exhibitions. Whether in manuscript or print, or on television, maps are the workhorses of weather forecasters, reporters, political and cultural commen tators, cartoonists, textbook writers, scientific researchers, administrators, and planners alike. As printed road atlases (or, increasingly, electronic satellite navigation systems), they are used in almost every car owning household. City dwellers use metro maps and street guides. Visitors to the British countryside plan their days out with Ordnance Survey sheets. Consciously or unconsciously, everybody draws or gestures a sketch map in explanation of a route or a location.

At no other time in the history of maps have so many kinds of maps been recognized and used by so many people from so many walks of life. It is all the more paradoxical, then, that in academic geography the use of maps in traditional format has sharply declined, and the study of the history of maps and mapping has almost totally been abandoned by geographers. Geographers still create and use maps, but mainly in electronic format. Instead of the visual comparison of distributions on a series of maps on paper, spatial data are manipulated and stored electronically in geographical information systems (GISs). An increasingly large proportion of maps pro duced in the twenty first century are impermanent con structions, virtual maps that are liquidated when no longer needed. After 200 years of publication on paper, large scale Ordnance Survey topographical maps are distributed in digital format for printing in a high street shop on a scale specified by the customer, who also selects the point he or she wants to be in the center of the print.

The general heading 'history of cartography' embraces two main groups of maps: those 'compiled' from a sometimes eclectic range of information sources, which may include measured data (such as the coordinates of latitude and longitude for a selection of places) (**Figures 1** and **2**); and those 'structured' on modern scientific principles (involving Euclidean geometry and precise measurement) (**Figures 3** and **4**). The distinction may be implied through the use of words such as 'mapping', 'mapmaker', 'map compiler' (and, for manuscript maps 'artist' or 'scribe'), and the avoidance of 'cartographer' and 'cartography', for nonmodern, scientifically con structed maps, but in practice the 'history of cartography' is widely accepted as incorporating premodern maps and mapmaking.

The history of cartography is no stranger to paradigm shifts. Indeed, it would be surprising were a history that can be measured in millennia, and that goes back to prehistoric times, not to reflect some of great markers of world history. Factors such as changes in the political structure of Europe or of individual countries, the theological implications of the expansion of geographical knowledge, the rediscovery in medieval Europe of the Roman geographer, and astronomer Claudius Ptolemy's text on projections and of the world and regional maps drawn from his coordinates of latitude and longitude, the social impact of printing, the growth of the educated classes, and the invention of a timepiece for the meas urement of longitude – all, among many other events large or small, general or local, had a role in accounting for the content or the presentation of contemporary maps. The inevitable variety of early maps, the multi plicity of types and subtypes (genres), and the manner in which one type may seem to have been prominent at one time in the past, invisible at another time, and visible again at a later date, means that the imposition of fixed classifications is doomed to failure. Only when the focus remains on the individuality of each map is it possible to recognize order in that diversity.

Approaches to the History of Cartography

For geographers, the relationship between maps and history has taken several forms. The terms used to de scribe each form are not synonymous; the 'history of cartography' is not the same thing as 'historical cartog raphy' or the 'history of mapping', although in practice the boundaries between the categories of activities noted below are neither precise nor impermeable. Many studies conducted under one heading might almost equally well be placed under another heading. By individualizing the different approaches, however, both the variety of work on early maps and mapping and the breadth of the subject as practiced may be more easily appreciated.

Maps in the 'Unrolling of the World'

For many British geography students in the mid twentieth century, a first encounter with early maps came through introductions to the history of discovery and exploration in their university curricula. As they learned about the pro gressive expansion over the centuries of the European's map of the world out from the Old World heartland, first

Figure 1 *A new description of the island of Britain which now contains the kingdoms of England and Scotland with neighbouring Ireland.* The ear est extant pr nted map of the Br t sh s es was comp ed by George L y n Rome from a var ety of sources. Copperp ate. West s at the top. (Reproduced w th perm ss on from the Br t sh L brary.)

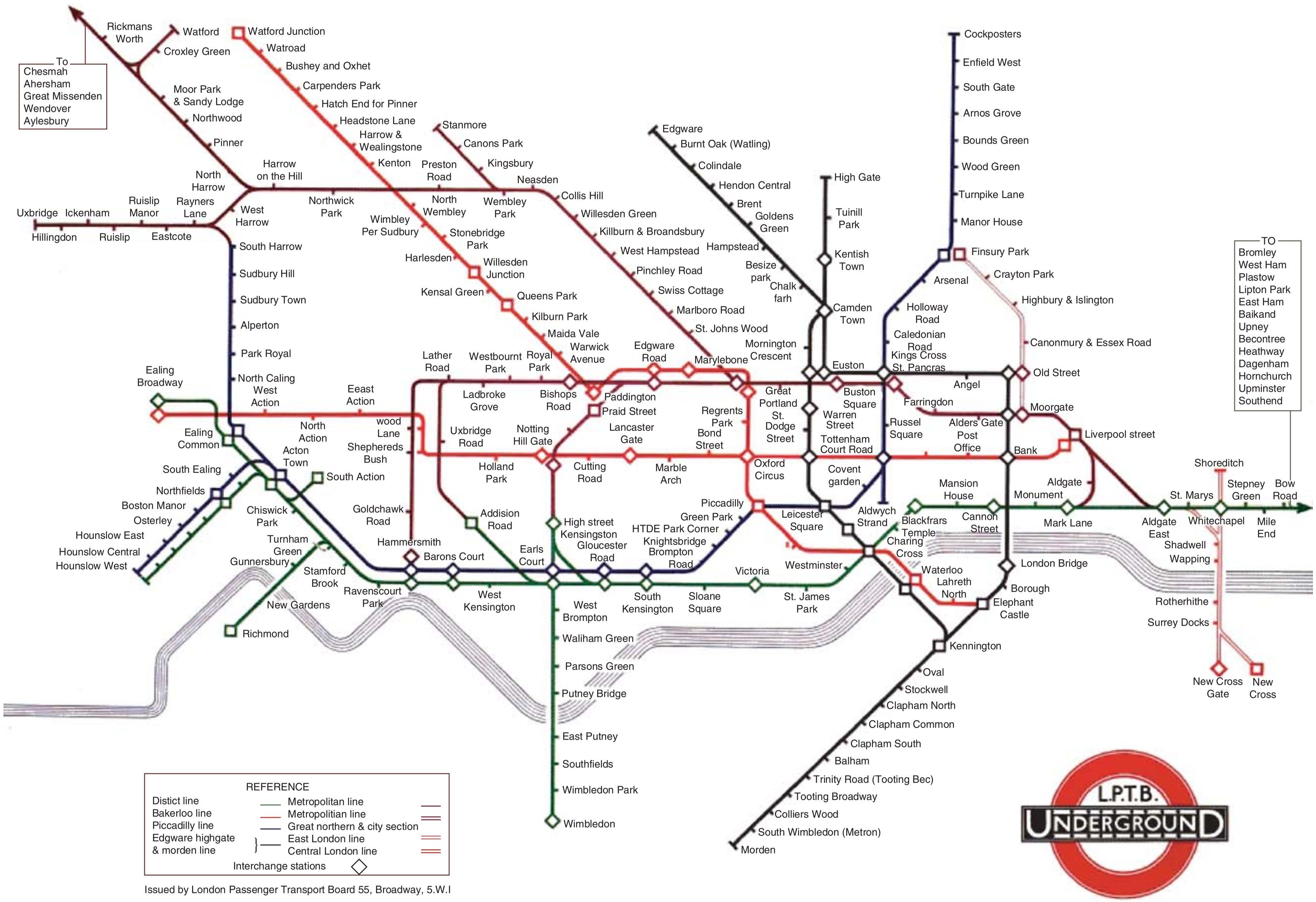

H.C. Beck

Figure 2 A modern examp e of a topo og ca map: Harry Beck's F rst Fo der map of the London Underground. n a topo og ca map, the most bas c express on of a spat a d str but on, the key pr nc p e s cont gu ty. D rect on and d stance (.e., as n Euc dean geometry s rre evant. Topo og ca maps have been drawn throughout h story and are a un versa and commonp ace form of mapp ng. (Reproduced w th perm ss on from London Transport Museum.)

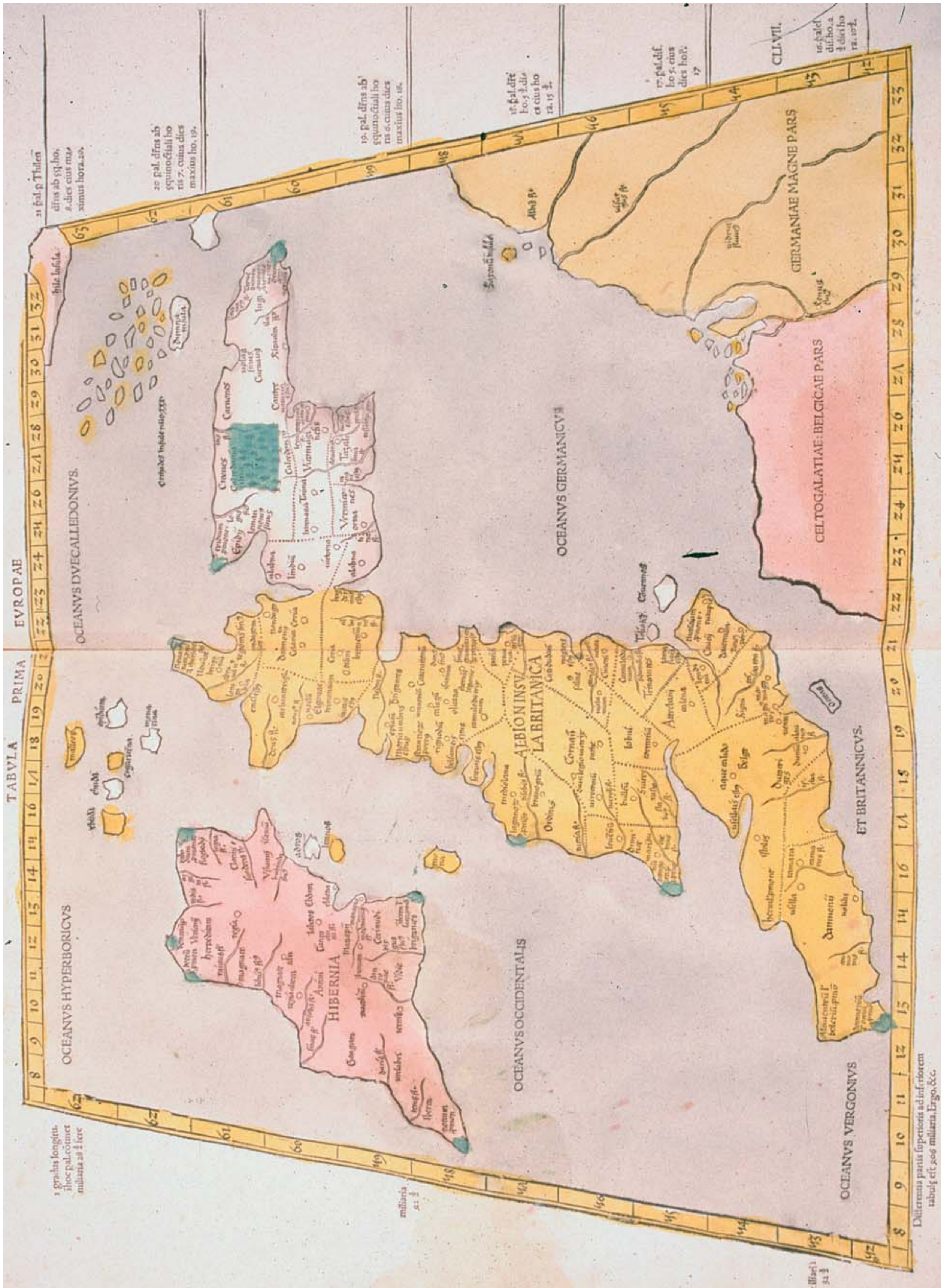

Figure 3 Roman cartographic science: Map of the British Isles constructed from the coordinates listed in Claudius Ptolemy's *Geography* (second century AD). Printed in Strasbourg in 1513. North is at the top. (Reproduced with permission from the British Library.)

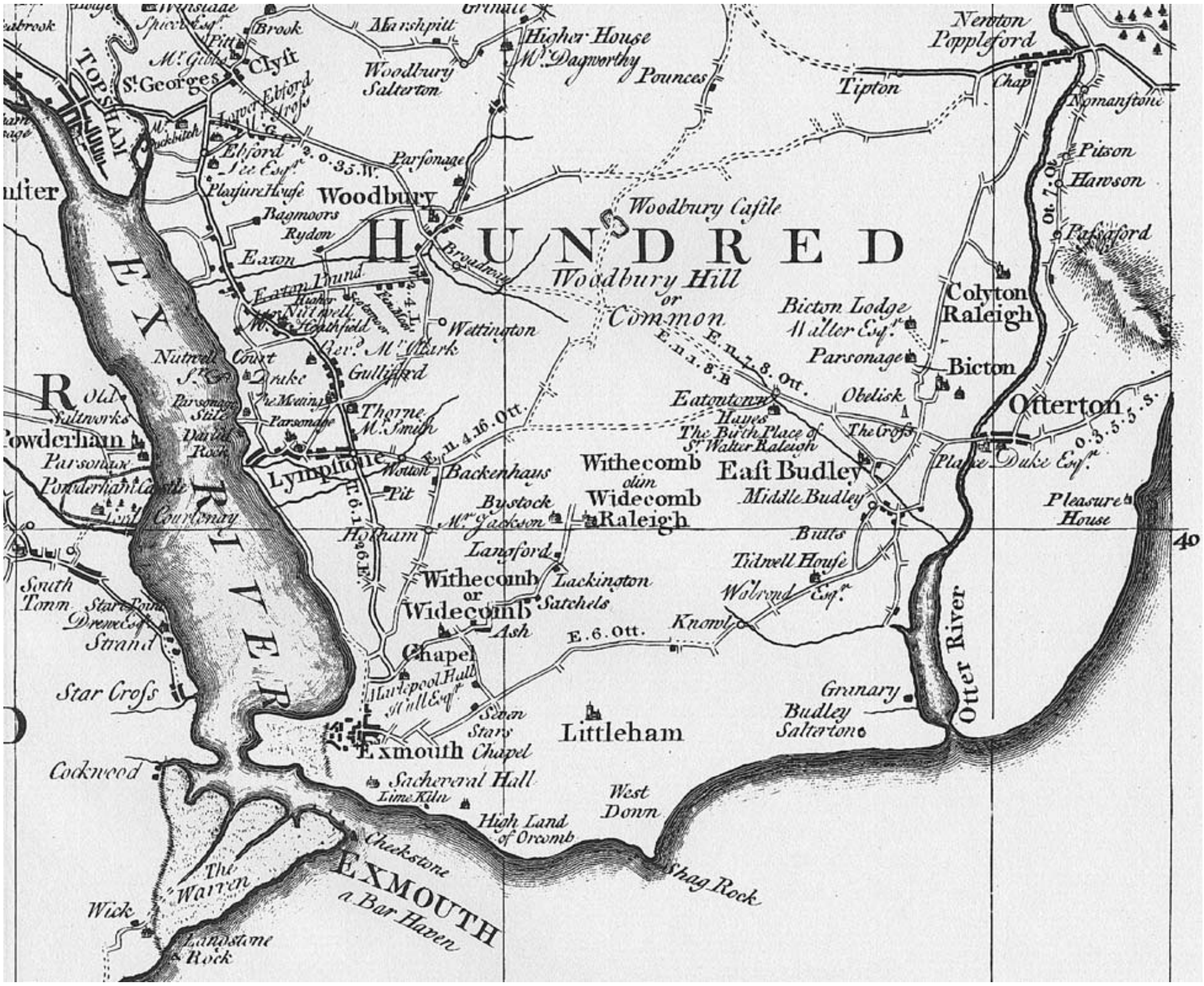

Figure 4 Benjamin Donn's 12-sheet map of the County of Devon was the first map to be awarded the prize, offered in 1759 by the Society of Arts for an English county map based on triangulation, that set new standards for scientific accuracy. Copperplate. North is at the top. (Reproduced with permission from the University of Exeter.)

eastward and southward, to the tip of Africa, then westward, and finally poleward, so they were introduced to contemporary maps recording the new geographies to inform those at home and to guide further exploration for trade and colonization. Likewise, as early medieval scholars wondered what lay on the 'other side' of the Earth's sphere, some of their maps depicted a hypothetical 'southern continent' (as on Beatus of Lièbana's map of the world from the eighth century, for example). As later medieval traders worked their way back to Venice along the Silk Route bringing with them new information about the distant East, so the Ptolemaic world was extended in that direction; about 1450, Fra Mauro explained that he had omitted Ptolemy's parallels, meridians, and degrees from his circular map of the world because these were too constraining in the light of what was now known about the world, and he noted that the extension of the map eastward meant that it was difficult to keep Jerusalem in the center.

After 1492, and the European discovery of the New World, world geography could be described in terms of Eastern and Western hemispheres, instead of the largely inhabited Northern and largely uninhabited, and uninhabitable, Southern sphere of the ancients and medievals. In the Renaissance, fragmentary and uncertain outlines on the maps were slowly completed and adjusted as European explorers and traders inched their ways around the coasts, mapping and charting as they went. Many of these maps are now lost, perhaps a reflection of their strategic value to the government, but even before the sixteenth century was out, Richard Hakluyt was collecting written accounts, and such maps as he could acquire, of the voyages of discovery as part of his and his contemporaries' (among whom was John Dee) promotion of commercially viable colonization of the 'new' lands. Today, the Hakluyt Society exists to advance the modern study of the political, social, and commercial consequences of the exploration of the world through the editing or reediting of contemporary writings and the maps that accompanied them. One fruitful line of research into maps and charts connected with the exploration and navigation of the world focuses on the toponyms recorded on such maps. While such studies may be used in the marshalling of arguments for partisan 'discovery' debates, they can also provide a shaft of light into the mind of the original mapmaker.

Historical Cartography

A second form of geographical and historical investigation about the past involves maps compiled from data relating to an earlier situation. Interpreted generously as the writing of history with maps rather than words, the practice is older and the reasons for it more diverse than is generally described. In the thirteenth and fourteenth centuries, for example, maps of the world were compiled to portray the history of human time. As God's creation, the inhabited Earth was depicted at the center of the cosmos and the map packed with vignettes summarizing the major historical events from the beginning of time (marked by the Earthly paradise) to the end of time (marked by Jerusalem, site of the anticipated apocalypse) (**Figure 5**). The content of these world maps (*mappae mundi*) is a systematic mixture of historical and contemporary geography, and myth and fantasy inherited from Classical texts, within the outlines of the three continents and their main natural features that provide the essential reference points. History (as related in the Bible) rather than geography was the key factor in this kind of map. When a Latin translation of Ptolemy's *Geography* became available in the West, about 1400, it was apparent that the Roman geographer's maps referred to a past situation and, from the middle of the fourteenth century, 'modern' versions of the regional maps began to be added to the manuscript copies and, in due course, the printed editions. Biblical geography remained of interest in the sixteenth century, especially to Protestants; Peter Laicksteen and Christian Sgrooten's map of the Holy Land distinguishes surviving, lost, and ruined places. In 1573, Abraham Ortelius published the first edition of his *Parergon*, an atlas of 'ancient geography', with 32 topographical maps drawn to help the student of the ancient world in the same way as his *Theatrum orbis terrarum* could help the student of the modern world and its regions. Maps showing regions and countries as they once had been continue to be produced.

A slightly different conception of the historical map is that of the map drawn to display the processes of history. The temptation to appropriate map history to make a political statement or for purposes of a nationalistic claim has never been entirely resisted. Maps reconstructing earlier territories, with their settlements, were produced in the Middle Ages, often in a national genealogical context, but they acquired a more explicit pedagogical function with the emergence in the sixteenth century of history as a formal subject in the modern sense. The growing interest from the eighteenth century in school atlases, and aids for geographical education such as globes, wall maps, and school atlases, testifies to the ways in which geographical knowledge was transmitted to the wider population beyond those of wealth and high social standing with access to private libraries. According to Walter Goffart, historical atlases in the modern sense had to await the emergence of a well defined sense of chronology and so were not found before the seventeenth century. Instead of showing the extent of territory in earlier times, the historical map *sensu stricto* represented periods. Gathered into atlas format, change could be demonstrated in a sequence of maps. Unless imported or translated into English, British readers probably saw few of the works produced in France, the Netherlands, and Germany; the first issue of an ambitious English project, the *Universal History from Ancient Times* by George Sale and T. Salmon appeared only in 1736, with the maps bound separately from the text. From then on, the impetus to construct cartographical images to show the past accelerated. By the end of the twentieth century, Jeremy Black, another historian commenting on historical mapping, analyzing in particular the products of the nineteenth and twentieth centuries, and pointing to the way the content of such maps changed after World War II, reflected sadly on how little attention is paid to historical maps and atlases by historians (and, he could have added, geographers) despite the considerable influence they have in "creating and sustaining notions of historical situations" and "advancing our understanding of space and spatial relationships." Modern geographers are familiar with the genre through the maps in Clifford Darby's seven volume reconstruction of the geography of England and Wales in 1086, from data recorded in *Domesday Book*. The millennial year 2000 saw a number of new publications in the by then well established genre of county historical atlases – the 'as we look forward, so we look back' genre – while the *Historical Atlas of Canada* (conceived in the early 1970s and Volume 1 published in 1987) represents the most ambitious construction of a national history through maps.

History from Maps

A particularly common approach to early maps assumes that the early map is a repository of information relating to the time when it was compiled, a simple quarry awaiting exploitation by the modern researcher. Early topographical maps, the printed regional or county maps of the sixteenth to eighteenth centuries, the precursors of modern official national surveys such as Britain's

Figure 5 The map of the world now in Hereford was compiled, probably in Lincoln and Hereford, by 1300. Its content is a mixture of empirical geographical knowledge (e.g., of the local area or gained through travel), textual information (e.g., from ancient classical geographies and histories, and the Bible), and traditional mythology. The primary purpose of the map was to display the Christian view of human history (time) within a geographical context (space). Manuscript. East is at the top. (Reproduced with permission from the Trustees of Hereford Cathedral.)

Ordnance Survey (founded 1791), are often packed with geographical, political, social, and administrative details as well as natural features. Probably, no maps are richer in their content of data from the past than large scale town plans with buildings shown in elevation and used as sources for the topographical history of towns. Maps are also used to study the changing courses of rivers – some, indeed, were drawn at the time for that purpose, with lift up flaps to indicate newly cut off meanders – and maps made originally for navigation are scrutinized in the study of environmental change for the coastal and estuarine features (human and natural). However, none of these maps should be used without a thorough understanding of their quality as evidence. Without the essential verifi cation of the map as historical document, the tendency of modern historians to relegate maps, like paintings and other nonverbal sources, to some lower division of evi dence than the written word is fully justified.

The Map in History

The arrival of the 'new' history of cartography (see below) gave life to yet another way of looking at early maps. The emphasis placed in the last decades of the twentieth century on attempting to understand the map in its contemporary context – how it was made, why it was made, for whom, and to what purpose or effect – encouraged (especially in the 1980s and 1990s) the study of early maps as expressions of power and as instruments of control, either intentional or subliminal. While there are circumstances in which maps have indubitably had a significant and tangible impact on people's lives – in border definition (or lack of definition), in the consoli dation of state control over land through cadastral mapping (or private control over landed property through estate mapping), in the manipulation of maps for political or religious (or, these days, commercial) propaganda – there has been a tendency to assume, rather than document, the power of maps and to overlook the point that every map, like every word, has some power of communication, whether in a textbook, exegetical trea tise, or in a newspaper or on television. Indeed, a ju diciously simplified, or diagrammatic, map may be more effective than the written text.

The History of Cartography

The division of the history of cartography into 'old' and 'new' came about in the penultimate decade of the twentieth century. Examples of the use of the word 'cartography' in Germany as early as the second decade of the nineteenth century have been found, but the idea of a formal history of cartography for the study of old, that is, nonmodern, maps is usually attributed to the Portuguese scholar Manuel Francisco de Barros e Sousa in 1839. Inevitably, the nineteenth and early twentieth century practitioners judged maps in the light of their own, post Enlightenment, times. The application of current thought to a position in the past to which it does not apply (presentism), the idea that civilization always advances in a desired direction (progressivism), the privileging of technology, and the acceptance of the invincibility, objectivity, and accuracy of science – only some of the constituents of a way of thinking that has yet to die out completely – meant that whereas the earliest maps might be discussed in a preliminary paragraph or chapter, they were not to be admitted to the pantheon. They were considered too 'crude', 'inaccurate' (mathematically), and oversimplified, or having too many exaggerated and out of scale features, to qualify as maps. cases, with exaggerated (i.e., out of scale) features in other cases to qualify as maps. A map was considered to be (as the *Shorter Oxford English Dictionary* continues to put it)

> A representation of the earth's surface or a part of it, its physical and political features, etc., or of the heavens, delineated on a flat surface of paper, etc., according to a definite scale or projection. (*Shorter Oxford English Dic tionary*, 1986, Vol.1, p. 1276)

Today, even when the progressivist notion, at least, has been abandoned by most scholars, the implicitly scientific connotation of the word cartography is held by others to justify a persistently exclusive attitude to the history of maps. The key factors in map production are seen as technical; the validity of the projection used for the representation of a sphere or parts of a sphere on a flat surface, and of the geodetical precision of the co ordinates of latitude and longitude. In short, their pre occupation is with maps drawn to mathematical scale, on which both distance and direction are measurable. For a majority of maps produced from the eighteenth century onward, these are valuable and important preoccupations, but they are not the only issues of relevance to the wider discipline of history.

Studies conducted in the perspective of the post Enlightenment paradigm have made, and continue to make, contributions of importance. They include, for example, work on the characteristics of projections, and on the science of surveying the heavens, the oceans and seas, and the land at large and small scale. They examine the nature of conventional signs used on nineteenth and twentieth century maps, and the processes of map re production and printing. They liaise with historians of science to document the way contemporary scientific knowledge was applied to surveying instruments for use on land or sea. This focus on the technical aspects of map construction and assemblage of a map's contents, however, leaves current practitioners of the history of 'cartography' impatient with maps that fail to meet the expected cri teria; while those who would like to place the emphasis on the 'history' of cartography remain unhappy with the curtailing, as they see it, of the lengthy and rich global past of maps. Maps from antiquity (not to mention pre history), the European Middle Ages, and many Renais sance maps tend to be structured on, or contain elements of, topological – not Euclidean – geometry; they are maps in diagrammatic style, for example, or they have selected place signs of exaggerated dimension in reflection of their perceived importance. To accommodate these maps, and the humanistic factors that account for their 'artistic' nature, a different approach is needed.

The History of Maps and Mapping

That new approach emerged in the 1980s, propelled by factors both external and internal. External to – one might say foreign to – the world of cartography was the French theoretical and philosophical post structuralist movement led by Michel Foucault and Jacque Derrida that had al ready found fertile ground in the humanities (including geography) in Europe in the late 1970s. Internal to the history of cartography were two factors: the foundation in 1981 by J. B. Harley and David Woodward of The History of Cartography Project, devoted to the publication of a six volume history of maps; and the personality and publications of Brian Harley himself. While the first, path breaking, volume of the *History of Cartography*, published by the University of Chicago Press, moved slowly toward completion through the decade (see below), Harley, a British historical geographer, had abandoned his insistence on empiricism and was applying the theories he found not only in Foucault and Derrida, but also in the work of "prominent sociologists, anthropologists, art historians, and literary and cultural scholars." Matthew Edney lists 15 authors used by Harley in an avowed attempt to move the study of early maps away from the ahistorical grip of cartography and toward academic autonomy. The sub stantial corpus of theoretical essays and articles Harley wrote between 1980 and his premature death in 1991 were the most influential of his career. By the close of the millennium, Harley's writings combined with his and Woodward's History of Cartography Project – which involved not only selecting authors for the three volumes published to date, but also, in the 1980s especially, per suading them to interpret maps as human constructs from a specific historical context – had, by the close of the millennium, ensured the desired shift of emphasis from the study of cartographical processes and products to in clude the social history of maps.

The new history of cartography (despite the private inclinations of Harley and Woodward, the traditional label proved too difficult to dislodge entirely), unlike the old, is inclusive, accommodating the humanistic and scientific aspects of the creation, compilation, and production of premodern maps. Accuracy, for example, is still an important criterion, but different kinds of accuracy are recognized as well as that of the map's overall con struction. In the Middle Ages and not only on the maps for Ptolemy's *Geography*. In the Middle Ages as in Sixteenth century, and not only on the maps for Plolemy's Geography, in individual localities might be positioned on a map according to contemporary measurements of lati tude (reasonably correct) and longitude (which remained difficult to measure until the eighteenth century), while other places were interpolated; those accurately plotted were identified by a special mark or place sign. Meas urements and distances might be given in writing on the map next to the features to which they applied. Another form of accuracy was that of the map's content. This might concern the correct classification of the local settlement hierarchy, the showing of towns as walled or unwalled (vital knowledge for anyone planning a journey), or the presence of a bridge at a major river crossing.

By the middle of the first decade of the twenty first century, the alternative term 'history of maps and map ping' was being generally welcomed as a more com modious term. To a few, wedded to the conception of the map as a technical construct, new definition offered they Harley and Woodward in 1987 —

> Maps are graphic representations that facilitate a spatial understanding of things, concepts, conditions, processes, or events in the human world.' (Harley and Woodward, 1987: xvi)

—is an invitation to overstretch the idea of what constitutes a map to extremes. It is true that the meta phorical use of the word 'map' has become extremely popular, with titles such as *The Mind Map Book* and *Mapping the Fiction of C F Cubas*. In the view of the ma jority of serious scholars now contributing to the ad vancement of our understanding of how and why maps were made and used at different times and under dif ferent circumstances in the past, however, there is no doubt that the subject has benefited considerably from the wider definition. Always open to input from a range of viewpoints, but traditionally fueled by geographers, historians, librarians, and collectors, those now writing articles and books or presenting papers at international and national conferences on, or involving the history of maps, include art historians, literary scholars, historians of science, social psychologists, and military historians; in truth, map history is a genuinely interdisciplinary endeavor.

In seeking, with Harley, to steer the 'history' of car tography away from an unbalanced preoccupation with the technical mapmaking process, Woodward continu ously stressed that a more holistic interpretation of early maps could be achieved by a systematic approach to three key aspects: the map as artifact, as image, and as vehicle of communication fashioned in the likeness of the society which created it. The order in which the three aspects are taken is unimportant; what matters is to arrive at an appreciation of the relationship among all three as a means of seeing the map in its historical context. The questions guiding researchers are deceptively simple – not only how was the map originally produced, but also why? For whom? How was it used? Not only what does this map show (or omit) but also what did the map mean to those for whom it was made (or who saw it sub sequently and under possibly quite different circum stances)? What were the messages conveyed overtly or subliminally by the map image? What knowledge (or misinformation) is embodied in each map? Such questions also direct attention to the need for evidence; analogy may occasionally be admissible to help illuminate the history of maps, as of other artifacts, but the emphasis is always on high standards of documentation from contemporary sources. The task is monumental, given the sheer quantity of extant material (in absolute, not relative terms; much has been lost over time, which makes the intellectual reconstructions all the more elusive) and its variety in medium, format, and function. The goal is to be able to treat early maps as individuals, not mere categories, and to illuminate the place of each one within the society that produced it, in relation to other genres of maps circulating at the time as well as to all aspects of the cultural, social, economic, and technical circumstances of the day. To this end, the concentrated scholarly efforts of researchers from every cognate field are combined in the furtherance of a worldwide history of maps and mapping practices. Harley and Woodward, and those who have taken their 1987 redefinition of 'map' to heart, have largely succeeded both in closing the methodological gap between the subdiscipline of the history of cartography and the broader discipline of history and in drawing together the conceptual approaches of the humanities and the social sciences academy. The history of cartography, in our opinion now more properly labeled map history, has come of age.

Cartometrics

Cartometrics, which involves the detailed assessment of the mathematical structure (geodetical accuracy) of a map, is still in its infancy as a potentially powerful tool in the historical study of early maps. The modern emphasis on understanding the social context from which a map emerged does not in any way exclude the need to understand also the technical processes of map production, compilation, and/or surveying. The matter of the map as physical artifact, and the medium on which its is drawn or printed, the inks and colors used, the nature of the paper, its dating by watermark or other means, is a separate, but no less legitimate, concern, but questions such as the nature of the projection, the use of triangulation, and the nature of that process (whether it was a purely local system, as in the case of most sixteenth and seventeenth century English county maps, or whether it was linked to a geodetic baseline) are still essential issues to be resolved, as they are also in the case of early charts for navigation: how were these compiled and was any projection involved? What instruments were used, how were they used, and what were their limitations? All such questions rest on the matter of the accuracy of the mathematical map (wrongly assumed to be the sole criterion of cartographical accuracy) and on the calculation of a map's geodetic errors. The arrival of digital cartography greatly increased the number of calculations that could be realistically done for each map, and the ease with which the distortion grids created by cartometric analysis can be applied to early maps, together, or instead of the vectors and circles used in predigital days. At present, though, cartometrics not only has technical problems to resolve (e.g., different programs are likely to produce different results even from the same map) but also, and more importantly, has yet to refine its approach to the selection of historical data or identification of a specific historical problem.

Cartobibliography: The History of Printed Maps

The indispensable step in approaching a map (or anything) to be used as evidence is to check its publication history and its credentials as a source in the same way as a bibliographer examines a book. The discriminating description of maps is known as cartobibliography, a term first used in 1914 by George Fordham in the title of his book *Studies in Carto bibliography.* Fordham neither repeated the word in the seven essays forming that book nor did he make any systematic attempt to establish any principles for the descriptive listing of maps; that had been done, up to a point, in 1838 by the Dutch map collector Bodel Nijenhuis, who did indeed comment on and evaluate early maps according to specified criteria. Unlike bibliography for books, however, for which a large number of societies exist to keep a watchful eye on standards of practice, the cartobibliographies produced throughout most of the twentieth century tended to come from individuals working independently in their own way. The outcome is a shelf full of regional and thematic cartobibliographies compiled to different specifications. An outstanding few were universally acclaimed to have set the heights to be achieved henceforth. Donald Hodson's *Printed Maps of Hertfordshire*, which was followed by his *County Atlases of the British Isles*, 1704 1789 are examples of the most comprehensive and detailed cartobibliographies to have been completed to date. Just as it is often alleged that Britain is the best mapped country in the world, so it may be said that it is also the richest for its county and other cartobibliographies.

The tasks of unraveling the origin of each individual map, recognizing different printings and issues of the same map (through the identification of a sequence of 'states' from altered imprints, coats of arms, sometimes tiny changes of content, or damage to the wood block or copper plate), checking what the given date actually represents (date of completion of draft, of printing, or of original publication) or ascribing a date for those lacking one, discovering who made it (and why, and for whom), chronicling its publication history (including reuses and derivations) and the postpublication history of the block or plate it was printed from, indexing its contents, es tablishing a bibliography, and much else, combines aca demic scholarship with the expertise of specialists, such as map collectors, map dealers, and auctioneers. In this context, cartometrics has already demonstrated its value in confirming families of maps. Setting the pace as well as the standards for modern cartobibliography is the group founded by Professor Günter Schilder in the Faculty of Geosciences, University of Utrecht, which has seen the publication of Schilder's own *Monumenta Cartographia Neerlandica* (eight volumes since 1986) and Peter van der Krogt's ongoing revision of Cornelius Koeman's *Atlantes Neerlandici. Bibliography of Terrestrial, Maritime and Celestial Atlases and Pilot Books, Published in the Netherlands to 1880* (1967 1971).

Infrastructure

Despite the relevance of maps to many fields of historical study in the humanities and the social sciences, academic responsibility for the study of early maps has tended to be nobody's business. Full length university courses are rare (there are at present only two professorial chairs in the subject in the world, both in the USA). A promising new development at postgraduate level are the short but intensive, taught history of maps and mapping courses in London under the aegis of the London Rare Book School of the University of London. Map historians are geo graphically dispersed in the world and heterogeneous in terms of professional background, but are held together by three international institutions – the flagship scholarly journal Imago Mundi, a series of biennial international conferences (22 till date), the already mentioned History of Cartography Project – and specific Internet portals.

From 1994, the journal *Imago Mundi* has been subtitled *The International Journal for the History of Cartography*. When it was founded in Berlin in 1935, by the *émigré* Russian map collector Leo Bagrow (later author of a substantial book entitled, *History of Cartography*, published in German in 1951, in English in 1964 and, with R. A. Skelton, 1966) the journal's title was *Iarbuch der alten Kartographie. Yearbook of Old Cartography*. The second issue, published in London in 1938, was titled *Imago Mundi: A Periodical Review of Early Cartography*. After a peripatetic existence, with publication successively in Stockholm, Leiden, The Hague and Amsterdam, *Imago Mundi* re turned to London in 1974, under the directorship of Imago Mundi Ltd., where it has since remained. Bagrow was the editor until 1957, but the character he gave his journal continued to shape it even after his death that year; his interests as a collector, like his definition of 'early maps' (nothing after 1800) still limited the journal's concern to the 'externals of maps' and the exclusion of "any examination of their content, of scientific methods of mapmaking, of the way the material is collected, or of the compilation of maps" (Bagrow, 1951: 13; Bagrow and Skelton, 1964: 22). Changes did come eventually, and the 59 volumes published up to 2007 can be taken as a bar ometer of the subject in general. In 1986, Harley criti cized *Imago Mundi* for "its antiquarian and bibliographical bias," for its "lack of philosophical and methodological direction," and for its bias toward description rather than interpretation. In keeping with the shift in the subject outlined above, and measured against papers presented at the international conferences (for which the journal's owner, Imago Mundi Ltd. acts as the coordinating body) as well as on its own contents pages, the decade strad dling the turn of the millennium has witnessed something of a sea change. In external appearance, and in basic internal structure (refereed illustrated articles and shorter articles, book reviews, obituaries, chronicle, bibliography, and notices), it remains much the same as in Bagrow's day and that of his successor as editor, Professor Eila Campbell, a London University geographer. More significant than the current editor's addition of abstracts in four languages (English, French, German, and Spanish; which in some measure compensates for the fact that all articles are in English) is the journal's request for em pirical, interpretative, and theoretical contributions, and the recognition that most of the twentieth century now is, or can be presented as, history; Bagrow's 'early' now extends to any kind of noncurrent map. Authors con tributing to the 14 volumes published between 1994 and 2008, wrote from 22 different countries (57% from the United Kingdom and the United States of America) and represented 16 subject areas (history accounting for 44%; geography, 42%; and history of art, 10% of the con tributors professional base). They covered periods from antiquity (a recently discovered map on papyrus from the early first century BC, and revisionary thoughts on the later Roman Peutinger map) to maps in World War II and in post war Israel. While articles on maps made before 1600 continue to account for more than half the total published in *Imago Mundi*, interest in maps made in the twentieth century as 'history' is slow to increase.

The first International Conferences on the History of Cartography (ICHC) were held in London in 1964 (on the occasion of the UK Congress of the International Geographical Union) and in 1967. Thereafter, meetings have taken place every 2 years in a different country. Insofar as the directors of Imago Mundi Ltd. are re sponsible for the academic standing of the conferences, as well as their continuity, the papers and posters presented at each conference parallel the articles published in the journal *Imago Mundi* as a reflection of the subject. In all other respects, each conference is autonomous. Four meetings have been held in the twenty first century (Madrid in 2001; Cambridge, Massachusetts, and Port land, Maine, in 2003; Budapest in 2005; and Berne in 2007). At the last, participants from 38 countries attended, presenting more than a hundred papers and posters. A *Short History and Bibliography of Papers* presented at all ICHC meetings since 1964, compiled and regularly updated by Douglas W. Sims and Peter van der Krogt are available on the *Universiteit Utrecht* website.

The History of Cartography Project was officially launched in Madison, Wisconsin, in 1981. As Woodward recounted, its inception dated back to May 1977, when he and Harley walked near the latter's home in Newton Abbot, Devon, discussing an older project of Harley's (a four volume text on the *Mapping of North America*). Dis suaded from the latter, Harley agreed to a joint venture to edit instead, with Woodward, four volumes of a worldwide *History of Cartography*, covering: (1) the prehistoric, classical, and medieval periods together with cartography in the non Western world; (2) the Renaissance; (3) the En lightenment; and (4) the nineteenth century. By 1982, well into preparations for the first volume, it had become clear that a separate volume would be needed to account for the richness and diversity of the non Western material. It was also accepted that a sixth volume would be needed for the twentieth century. By the time of Harley's sudden death, only *Volume 1 – Prehistoric, Ancient and Medieval Europe and the Mediterranean* – had been published, to widespread acclaim (it was awarded the best book in the humanities by the Professional and Scholarly Publishing Division of the Association of American Publishers). It involved ten authors from three countries (seven from the UK). The three books of *Volume 2 – Cartography in the Traditional Islamic and South Asian Societies* (awarded the R. R. Hawkins Award for the best scholarly book by the Association of American Publishers), *Cartography in the Traditional East and Southeast Asian Societies*, both under the names of Harley and Woodward, and *Cartography in the Traditional African, American, Arctic, Australian, and Pacific Societies* (awarded the Henry Breasted Prize by the American Historical Associ ation), edited by Woodward and G. Malcolm Lewis – includes essays by 34 authors from nine countries. *Volume 3 – Cartography in the European Renaissance*, was edited by David Woodward, although published posthumously in 2007, to which, 64 authors from 14 countries have contributed.

Underpinning the entire project is a broadly chronological framework adapted as necessary to the requirements of the individual volumes. It was *Volume 1* that, by announcing a new definition, broadcast to the wider world the new emphasis on the history of maps and mapping, rather than on the technical aspects of cartography, as much as Harley's theoretical essays were doing more or less concurrently. *Volume 2* demonstrated the liberating impact of that definition in terms of the nature of the material dealt with – many maps in Asia were created in the context of religious belief, and represent the imagined structure of the cosmos, the path from this world to the next, or depict pilgrimage routes to holy places – or corrected long established myths in the anthropological and history of cartography literature about, for example, the function of the stick charts of the Pacific Marshall Islanders. In focusing on European maps produced between *c.* 1470 and *c.* 1640, *Volume 3* deals with more familiar subject matter, but here too Woodward continued to promote the new approach to the history of maps and mapping by defining the general aim as "to describe the many levels on which maps became a central means to structure and understand the world, and how maps offered the means to articulate a cultural and political understanding of the state" (Woodward, 2001: 26). A similar epistemology is shaping the remaining three volumes (*Cartography in the European Enlightenment*; *Cartography in the Nineteenth Century*; *Car tography in the Twentieth Century*) for which changes al ready in train concern the format of the entries, not the intellectual approach (they are to be encyclopedic entries, instead of the long regional and thematic essays of previous volumes); the list is incomplete at the time of writing, but already 180 authors from 22 countries are signed up for *Volume 4* (edited by Matthew Edney and Mary S. Pedley). Whereas by no means all those who have contributed to *The History of Cartography* would describe themselves specifically as historians of maps, all are scholars who have addressed maps in their own work and whose interests lie in describing the cultural, social, and intellectual influence that maps gained as both as tools and visual icons in their contexts.

One of the challenges anticipated by Woodward in connection with writing the history of the twentieth century is "the affinity of the first two thirds of the century with much of the nineteenth century in terms of cartographic compilation, representation, and repro duction," meaning that the division of a 'long nineteenth century' between two volumes could be considered ar tificial; a more significant break, he observed, would be between a predigital and a postdigital age. A problem of a different order comes from the diversity and vulner ability of maps in this century. Ironically, although a greater number of maps have been produced in the twentieth century than in any previous period (it was said in 1999 that between 10 and 20 million maps are made everyday), the rate of loss has been, arguably, also un paralleled, either through the catastrophic physical de structiveness of the two world wars, which obliterated entire archives and libraries and many unique examples of early maps (the huge medieval Ebstorf map, lost in the bombing of Dresden, springs to mind) or simply because, increasingly in the digital age, maps are not intended to last. The corollary of the digital cartographers' 'con tinuous revision' is continuous obsolescence; is the twenty first century to be the first without the materials for writing its cartographic history?

On a daily basis, the Internet provides a number of services for anybody interested in the history of cartography. The web allows information to be presented in a dynamic rather than a static form, which means that updates that used to take many months to be published can be implemented immediately. It also attracts a wider audience than could be reached by traditional methods of publication, and has greatly increased awareness of, and interest in, the issues raised. The primary point of entry is the permanent site (i.e., listed in the 'Internet Archive') *Map History/History of Cartography: THE Gateway to the Subject*, which documents activities and resources relating to the history of cartography, and provides links to where these are described. The Map History site also contains the only systematic bibliography of relevant web publi cations, including two journals that are now available only online, and others that are available in searchable form online as well as in print, and provides access to a number of major external initiatives. These include the 700 member MapHist Internet discussion list (presently monitored from Utrecht); the regularly updated 'Calen dars' of events and exhibitions maintained by John Docktor (USA); and Matthew Edney's ongoing list of theoretical work, Recent Trends in the History of Car tography: A Selective, Annotated Bibliography to the English Language Literature. To be included is the tenth edition of The International Directory of Current Re searchers in the History of Cartography, in preparation for exclusive online publication for the first time.

Whatever the future may hold for the twenty first century, the challenge of the maps still packed into na tional archives, state and other institutional libraries, and private collections remains. The 'discovery' of a hitherto uncataloged or unrecognized early map is an almost daily occurrence. Not all the private great houses, in the UK and elsewhere, open their doors to researchers, and one wonders how many missing links in the history of maps lie hidden therein. One thing is clear, the history of maps and mapping has come to maturity with a holistic agenda. It is no longer a purely antiquarian pursuit, or hand maiden to modern mapmaking, but a respectable and respected subfield of history.

Further Reading

Bagrow, L. (1951). *Greschichte der Kartographie*. Berlin, Sefact: Verlag.

Bagrow, L. and Skelton, R.A. (1985). *History of Cartography* (2nd English edn.) Chicago IC: Precedent Publishing, Inc.

Black, J. (1997). *Maps and History. Constructing Images of the Past*. New Haven, CT: Yale University Press.

Cosgrove, D. (2007). Epistomology, geography and cartography: Matthew Edney on Brian Harley's cartographic theories. *Annals of the Association of American Geographers* 97(1), 202 209.

Delano Smith, C. (2001). The grip of the enlightenment: The separation of past and present. In Woodward, D., Delano Smith, C. & Yee, C. (eds.) *Plantejaments i objectuis d'una història universal de la cartografia/Approaches and Challenges in a Worldwide History of Cartography*, pp 283 297. Barcelona: Institut Cartogràfic de Catalunya.

Delano Smith, C. and Kain, R. J. P. (1999). *English Maps. A History.* London: The British Library.

Edney, M. (2005). The origins and development of J. B. Harley's cartographic theories monograph 54. *Cartographica: The International Journal for Geographic Information and Geovisualization* 40(1 2), 143pp.

Goffart, W. (2003). *Historical Atlases. The First Three Hundred Years, 1570 1870*. Chicago, IL: Chicago University Press.

Harley, J. B. (1986). Imago Mundi The first fifty years and the next ten. *Cartographica* 23(3), I 15.

Harley, J. B. (2001). *The New Nature of Maps*: *Essays in the History of Cartography* (Laxton, P. (ed.) with an Introduction by Andrews, J. H.). Baltimore, MD: The Johns Hopkins University Press (A reprint of six of Harley's theoretical articles, with a complete listing of all his publications).

Harley, J. B. and Woodward, D. (1987). *The History of Cartography, Vol. 1: Cartography in Prehistoric, Ancient and Medieval Europe and the Mediterranean*. Chicago, IL: University of Chicago Press.

Harley, J. B. and Woodward, D. (1992). *The History of Cartography, Vol. 2, Book 1: Cartography in the Traditional Islamic and South Asian Societies*. Chicago, IL: University of Chicago Press.

Harley, J. B. and Woodward, D. (1994). *The History of Cartography, Vol. 2, Book 2: Cartography in the Traditional East and Southeast Asian Societies*. Chicago, IL: University of Chicago Press.

Kain, R. J. P. and Delano Smith, C. (2003). Geography displayed: Maps and mapping. In Johnston, R. J. & Williams, M. (eds.) *Geography in the Twentieth Century*, pp 371 427. Oxford: Oxford University Press and The British Academy.

Sims, D. W. (2001). A short history and a bibliography of papers, The International Conferences on the History of Cartography. http://cartography.geog.uu.nl/ichc/index.html (accessed Jul. 2008).

Talbert, R. (2000). *The Barrington Atlas of the Greek and Roman World*. Princeton, NJ: Princeton University Press.

Woodward, D. and Lewis, G. M. (1998). *The History of Cartography, Vol. 2, Book 3: Cartography in the Traditional African, American, Arctic, Australian, and Pacific Societies*. Chicago, IL: University of Chicago Press.

Woodward, D. (ed.) (2001). 'The origins and history of *The History of Cartography*', 'Theory and *The History of Cartography*'', and 'Challenges: the recent history of cartography and the possible use of *The History of Cartography* volumes in the 21st century'. In Woodward, D., Delano Smith, C., and Yee, C. (eds.) *Plantejaments i objectuis d'una història universal de la cartografia/Approaches and Challenges in a Worldwide History of Cartography*, pp 23 29, 31 48, and 299 312. Barcelona: Institut Cartogràfic de Catalunya.

Woodward, D. (ed.) (2007). *The History of Cartography, Vol. 3: Cartography in the European Renaissance 2007*. University of Chicago PressChicago, IL.

Relevant Websites

http://cartography.geog.uu.nl
A Short History and Bibliography of Papers presented at all ICHC meetings since 1964.

http://www.sunysb.edu
Coordinates (online journal of map and geography round table, American Library Association), Stony Brook University.

http://www.maphistory.info
Site Map, Map History/History of Cartography.

Case Study Approach

S. W. Hardwick, University of Oregon, Eugene, OR, USA

Glossary

Case Study A type of research that focuses on the intensive analysis of a particular place, group, or specific issue that is most often conducted using a mixed methods approach.

Collective Case Study The study of a group of interrelated case studies conducted by a single scholar or several researchers working collaboratively. Data from more than one site are often gathered, analyzed, and synthesized collaboratively by a team of researchers using a comparative method of analysis.

Data Triangulation Use of different data sets, researchers, or data sources to confirm or corroborate research findings.

Descriptive Case Study Approach to doing research that requires that an overarching descriptive or interpretive theory be developed before beginning the study.

Explanatory Case Study Useful when doing causal investigations that contribute to testing predictive models.

Exploratory Case Study Research conducted to gather and analyze foundational data to be used for more expanded work to be carried out after these preliminary pilot data have been assessed.

Heterolocalism Theory that immigrants settling in North American cities in recent years are finding new ways to maintain their ethnic identities without residential propinquity.

Intrinsic Case Study Approach to doing research wherein the researcher has a (personal or professional) interest in the project and thus as an 'insider' may play the role of a relatively subjective observer.

Mixed Methods A research methodology often used in case study research that may encompass an overlapping suite of qualitative and quantitative research approaches to data gathering and analysis.

Qualitative Methods Research methods such as interviews, participant observation, and textual analysis useful in documenting and analyzing social structures or individual experiences and relationships.

Quantitative Methods Statistical and mathematical modeling approaches used by geographers and other scholars to analyze cultural, social, economic, and physical relationships.

Validity The extent to which a method of data collecting produces consistent and reliable results when applied at different times by different researchers.

Introduction

The case study approach in human geography is useful in both research and teaching, particularly when an in depth investigation is needed. The study of a single place, a particular group, or a specific issue in one location is helpful in narrowing down research topics. Such topics are often more manageable and meaningful for both early career geographers and more experienced scholars. This approach tends to be selective, with an emphasis on one or two issues that are fundamental to understanding larger research questions. Case studies are most appro priate when a researcher wishes to utilize a set of mixed methods of data collection and analysis in order to bring out the viewpoints of multiple participants in the study.

However, what criteria should be used to define and delimit a particular case study and how are research and teaching questions best framed when applying the case study approach in human geography? Answers to these and other related questions about the usefulness and the limitations of using case studies for research and teaching in geography provide the overarching structure for this article.

Types of Case Studies

At least five different types of case studies have been discussed in the geography and social science literature: intrinsic, collective, explanatory, descriptive, and ex ploratory. In each of these types, there may be single case or multiple case approaches used. An intrinsic case study is defined as research wherein the researcher has a (personal or professional) interest in the project. In this approach, the insider scholar may play the role of a relatively subjective observer (instead of working from a more objective outsider perspective). Collective work refers to the study of a group of interrelated case studies conducted by a single scholar or several researchers working collaboratively. Data from more than one site are often gathered, analyzed, and synthesized collaboratively by a team of researchers using a comparative method of analysis. Explanatory case studies are best used when doing causal investigations since they lend themselves to analysis and, at times, even make contributions to pre dictive models. Descriptive case studies require that an overarching descriptive or interpretive theory be de veloped before beginning the study. Finally, exploratory case studies are conducted to gather and analyze foundational data to be used for more expanded work and a set of larger questions to be pursued and carried out after these preliminary pilot data have been assessed.

In human geography, all five of these types of case study approaches have been used in teaching and research. None are mutually exclusive although the use of single case, exploratory, and then explanatory approaches is often the most reliable and valid. For example, a focused case study of Russian immigrants in one metropolitan area (as shown in **Figure 1**), began with a single case exploratory study and then moved into becoming an explanatory case study in later stages of the project as immigration theory ultimately was tested by a collaborative group of geographers. Key to understanding the application of the theory of heterolocalism in a North American metropolitan area, for example, was a follow up study that spatially analyzed religious networks of the same group in the same study area (as shown in **Figure 2**). This example, and many others published in the discipline's flagship journals in recent years, attest to the usefulness and popularity of the case study approach in conducting geographic research.

Methods Useful in Case Study Research

Particular attention must be paid to considering how best to test for reliability and validity when assembling the methods used to conduct case study research. One of the best ways to accomplish this is to use multiple sources of evidence including the use of such approaches to data gathering as interviews, survey questionnaires, field observation, textual analysis, participant observation, analysis of government and other statistical records, and spatial analysis. In fact, case study research most often involves a multimethod approach that is well grounded in data triangulation. This approach helps confirm the validity of findings on the particular case study and may employ an overlapping set of both qualitative and quantitative methods.

During the past 15 years or so, the use of strictly qualitative approaches has gained popularity in human geography as feminist, postmodern, Marxist, and other research approaches in the field have increasingly come to depend on the use of qualitative methods in their work. Deconstructing the more nuanced findings and meanings from data gathered through the use of

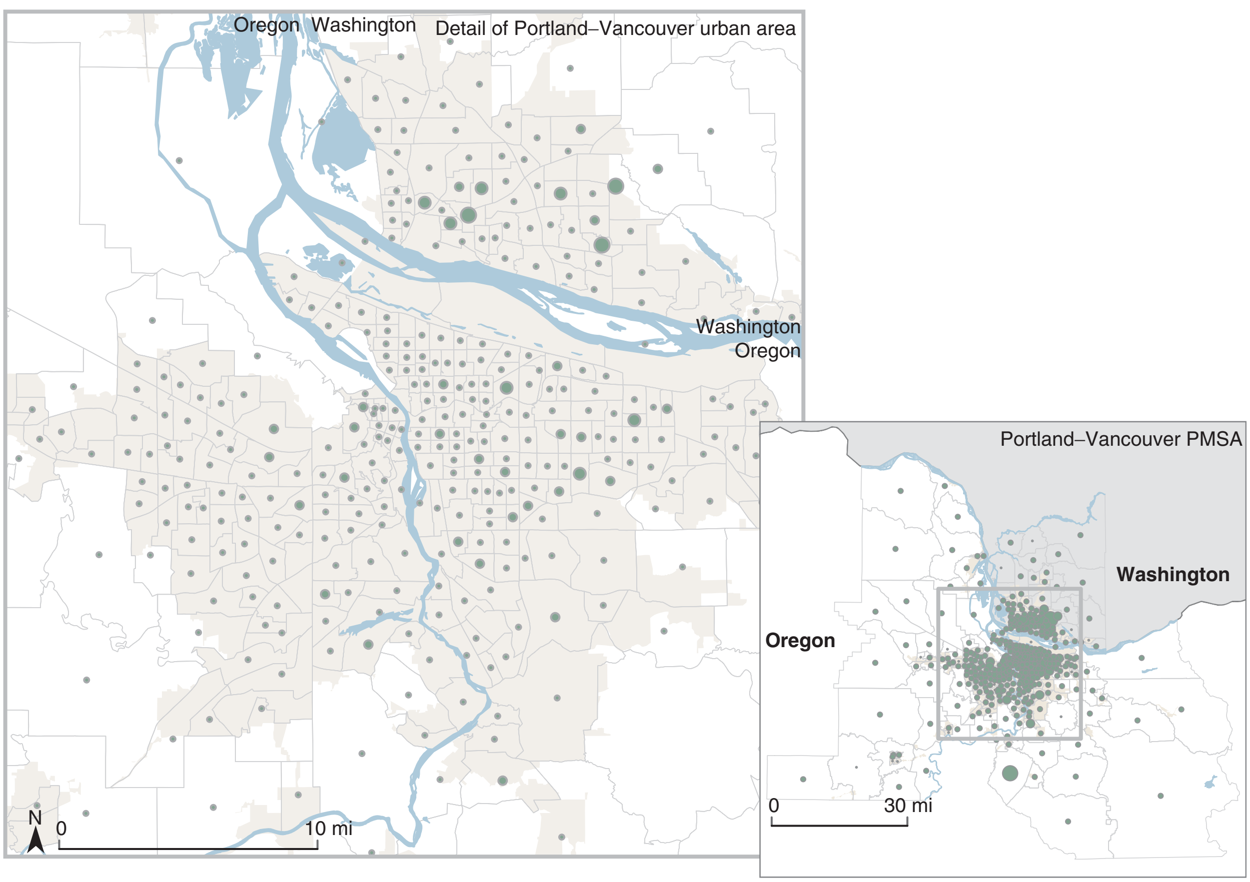

Figure 1 Russian first ancestry, 2000. Percent of all Russian first ancestry in each census tract. ●, 1.6% 2%; ●, 1.1% 1.5%; •, 0.6% 1.0%; •, 0.1% 0.5%; , Urban Area. From Census and TIGER\Line 2000 data.

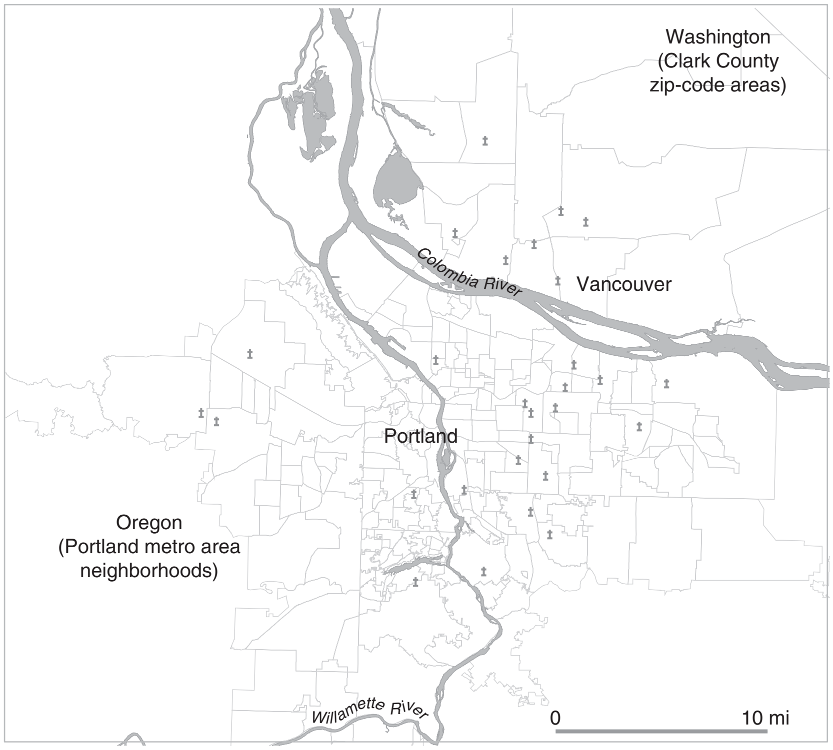

Figure 2 Portland Vancouver area churches serving Russian-speaking residents. ▭, Neighborhood boundary in Portland, or and zip-code boundary in Clark County, WA; ☨, Russian or Ukrainian church. From Immigrant and Refugee Community Organization, Census TIGER\Line 2000 and Slavic Directory, 2002.

unstructured and structured interviews, participant ob servation, and textual analysis often form the heart of this type of work. Feminist scholars in the field, in particular, have depended on ethnographic methods centered in the use of interviews to learn more about the lived experi ences of people on the ground in particular places.

In contrast to the early work accomplished by feminist geographers that developed from radical or Marxist ap proaches (and was thus strongly antipositivist and thus rejected the use of quantitative approaches), other types of human geographers, such as population geographers, have also depended upon various types of quantitative data as a key part of their mixed methodologies. In this type of work, data sources such as census records may be used to help corroborate qualitative findings when using the case study approach. Population geographers also often focus more heavily on spatial analysis in their case study work.

Using the Case Study Approach to Teach Human Geography

Case studies are useful not only in research, but also in classroom instruction in human geography. One of the many strengths of this type of teaching and learning is that it is one of the most effective ways to link research and teaching in geography classrooms at all levels of instruction. During the past decade, a spate of books and articles in geographic education journals have docu mented some of the many ways case studies can be useful in helping students understand and apply the geographic perspective. Using case studies as an instructional method encompasses many different types of interactive learning and thus encourages educators to move away from a teacher centered classroom to creation and nur turing of a student centered learning environment.

Scholars and educational practitioners have long argued for the benefits of interactive pedagogies that encourage students to become active contributors to the learning process. This approach moves geography faculty and classroom teachers away from traditional lecture approaches to instruction toward interactive and collaborative learning. Numerous prior studies have established that student knowledge and skills mastery are retained longer when they are encouraged to become active participants in the learning process. Long term assessments of learning outcomes have also ascertained that the majority of students improve their content knowledge, skills mastery, and affective and attitudinal learning in classrooms that are more student centered.

Use of the case study approach to teaching human geography requires students to actively engage with course content by reading, analyzing, comparing, and critiquing a set of cases that are issues based and often link local to global scales of learning. This way of teaching differs dramatically from the more traditional use of cases as supplementary reading assignments to illustrate key points in economic, cultural, or even physical geography lectures. The case study method instead refers to an approach that is grounded in an overarching educational practice with the primary objective of active, centered learning. In this ap proach, students are asked to become members of a col laborative group of learners whose task is to explore, analyze, synthesize, and potentially criticize specific case studies framed by real world issues. Case study materials useful in human geography instruction can come from book chapters, academic and popular articles, newspaper articles or other documents, and the teacher's own experiences.

Perhaps most importantly, this highly collaborative pedagogical approach provides faculty with the oppor tunity to embed their own research into classroom learning. The use of data gathered on site in the field doing case study research is only one way to integrate research into classroom instruction. Students can also be asked to conduct their own research projects modeled after the work of faculty or classroom teachers that in cludes development and refinement of research ques tions, discussion of appropriate research methods to be used for the study, practice in using various types of data compilation and data analysis, and writing up and then defending the final outcomes of the project in oral pre sentations. All of these methods are best conducted using local or regional case studies that can be scaled up to larger geographic questions under discussion in course work in geography.

One of the new sites of case study teaching and learning now underway in US classrooms is occurring in advanced placement human geography courses all across the nation. With the support of ample supplies of pub lished and online curricular materials developed espe cially for novice and experienced classroom teachers expanding exponentially as the number of students in these advanced secondary level classrooms expands each year, the use of case studies as a preferred approach continues to gain credibility. High school students (along with many of their teachers) often enter their first high level geography classroom unaware of even the most basic content, skills, and perspectives of human geography. Using case studies that illustrate issues, places, and spaces at a variety of scales helps deepen their understanding of core concepts and skills and also increases teacher and student confidence in applying the geographic perspective in geography classes and beyond.

The use of the case study approach in educational settings also provides a discussion forum for decon structing and rethinking some of the ways the world is changing. The study of geography, as well as teaching and learning itself, are contested spaces of change where fluidity is the norm. However, despite this ongoing cli mate of rapid change, many of the now outdated concepts in our field are the norm in geography classrooms in many parts of the world. The use of case studies can help expand these now outdated approaches and content to enliven students' interest and engage them in their own learning.

Issues and Challenges of the Case Study Approach

Use of the case study approach in both teaching and research has been criticized for more than five decades. Indeed, unlike this reference book, many research and pedagogy guidebooks fail to include the case study ap proach in their discussions. The reasons for this are twofold. First, case study research has been criticized for: (1) its unscientific nature (because findings cannot be replicated) and (2) reliance on overgeneralizable findings. Key to overcoming this first limitation is triangulating a rigorous set of mixed method approaches to data col lection and analysis and maintaining a chain of evidence to argue a case. The second criticism is best mitigated by using the findings from the case study to address and contribute to larger questions, issues, and theories in human geography.

Since almost all case studies involve the use of interview methodologies or ethnographic work, one of the strongest arguments for their validity is to emphasize that larger scale data sets often overlook or blur the significance of individual stories. Therefore, case study research has the potential to capture and analyze the lived experiences of people, and understand more about particular places on the ground. In sum, despite criticism related to studies that focus on specific places, groups, or issues, scaling up the findings from small scale projects to respond to larger research questions makes case study research and teaching critical in helping link local issues to larger global challenges.

See also: Mixed and Multiple Methods; Place.

Further Reading

Feagin, J., Orum, A. and Syoberg, G. (eds.) (1991). *A Case for Case Study.* Chapel Hill, NC: University of North Carolina Press.
Grant, R. (1997). A claim for the case method in the teaching of geography. *Journal of Geography in Higher Education* 21, 115 171.
Hancock, D. and Algozzine, B. (2006). *Doing Case Study Research*. New York: Teachers College Press.
Hay, I. (ed.) (2005). *Qualitative Research Methods in Human Geography.* Oxford: Oxford University Press.
Limb, M. and Dwyer, C. (2001). *Qualitative Methods for Geographers: Issues and Debates*. London: Arnold.

Ragin, C. C. and Becker, H. S. (1992). *What is a Case?* Cambridge: Cambridge University Press.
Robinson, G. (1998). *Methods and Techniques in Human Geography.* New York: John Wiley.
Stake, R. (1995). *The Art of Case Research*. Thousand Oaks, CA: Sage Publications.
Strauss, A. (1987). *Qualitative Analysis for Social Scientists*. New York: Cambridge University Press.
Telly, W. (1997). Application of a case study methodology. *The Qualitative Report*. http://www.nova.edu/ssss/QR/QR3 3/tellis2.html (accessed April 2008).
Velenchik, A. D. (1995). The case method as a strategy for teaching policy analysis to undergraduates. *Journal of Economic Education* 26, 29 38.
Yin, R. (1994). *Case Study Research: Design and Methods*. Thousand Oaks, CA: Sage Publishing.

Relevant Websites

http://www.apcentral.collegeboard.com
The College Board Online Forum for Advanced Placement Teachers.
http://www.geom.uiuc.edu
The Geometry Center: The Chance Project.
http://www.cases.bham.ac.uk
University of Birmingham, Metallurgy and Materials Case Studies Teaching and Learning Resources: Case Study Theory.
http://www.uky.edu
University of Kentucky: Teaching with Online Case Studies.
http://v.webring.com
Web Ring: Qualitative Research.

Cassa per il Mezzogiorno

F. Martinelli, Università 'Mediterranea' di Reggio Calabria, Reggio di Calabria, Italy

Glossary

Cassa per il Mezzogiorno The Cassa per le Opere Straordinarie di Pubblico Interesse per il Mezzogiorno ('fund for extraordinary works of public interest for the Mezzogiorno'), in short, Cassa per il Mezzogiorno, was the agency established by the Italian government in 1950 with the task of implementing the Intervento Straordinario, a special program of public investment in support of Southern Italy. The Cassa had planning and financial autonomy and was reauthorized several times, until its termination in 1984. Although the experience of the Tennessee Valley Authority in the USA was well known in Italy, it did not inspire the Cassa's institutional architecture, which was rather in continuity with the Fascist practice of special institutions.

Intervento Straordinario (IS) The term means special or extraordinary intervention and refers to the set of financial provisions and policy measures established by the Italian government – in addition to ordinary public spending – to support the development of Southern Italy. Established in 1950, it was terminated in 1992. Its chief institution until 1984 was the Cassa per il Mezzogiorno, often confused with the IS itself. After the Cassa, the IS operated with a more articulated institutional architecture regulated by the Law 64/1986.

Mezzogiorno The word means noon in Italian and refers to the Southern part of the country. It includes 8 of the 20 Italian administrative regions accounting for about 40% of the Italian territory and 35% of the national population. The roots of its underdevelopment lie in the very different historical trajectory of the region, which had a mostly colonial status up to the eighteenth century. Even after the political unification of the Italian peninsula in 1861, and throughout World War II, the Mezzogiorno was characterized by a semi-feudal economic and social structure, which hindered any bourgeois revolution and capitalist development. Because of the magnitude of the problem and means deployed, the Mezzogiorno and the policies established to support its development have become a paradigmatic case in regional studies.

Southern Question The term (Quistione Meridionale in Old Italian) synthesizes the multiple issues related to the underdevelopment of the Mezzogiorno. It became internationally known through the works of Antonio Gramsci (1891–1937), who provided a Marxist but very original analysis of the historic social bloc – an alliance between the Italian northern industrial bourgeoisie and the southern landed aristocracy – that contributed to reproduce southern underdevelopment after the unification of the country and throughout the Fascist regime.

Third Italy The term was used for the first time by Italian sociologist Arnaldo Bagnasco in his 1977 book *Le tre Italie* (*The three Italies*). Bagnasco stressed that, in between the traditional dualistic portrait of the Italian economy – the industrialized north based on large firms and the lagging agricultural south – a third model, that is, the Third Italy, was to be acknowledged in the north-eastern and central regions, based on a diffused and articulated system of small firms. His work together with that of Italian economists Giacomo Becattini and Sebastiano Brusco on industrial districts became internationally known in the 1980s, spurring a whole new approach to regional development. However, while the above authors cast the highly specific territorial models of production and reproduction of the Third Italy into an articulated institutional approach, their work was vulgarized by other scholars, who put forward a universal normative local development model.

Introduction

The Cassa per il Mezzogiorno was a public agency created by the Italian government in 1950 to carry out the Intervento Straordinario (IS), a special program of additional public investment directed to the historically lagging southern region (**Figure 1**) of the country (40% of the national territory and 35% of national population). Initially funded for a 10 year period, the Cassa was refinanced several times, becoming the main arm of Italian regional policy for over 30 years, until its suppression in 1984. The IS outlived the Cassa, with a different institutional architecture, until 1992, when dedicated national financial support for the Mezzogiorno was altogether ended.

Because of its magnitude and duration, the IS attracted significant international attention and the Cassa became a renowned model of regional developed agency. However, its impact on the Mezzogiorno has lately become controversial. Since the early 1980s, mainstream opinion portrays the Cassa – and the overall IS – as an epochal failure. Not only did it supposedly fail to bridge the north–south economic divide, but it also allegedly helped create in the South a distorted, inefficient, and corrupt regional development model, based on the clientelistic and parasitic redistribution of public expenditures. The termination of the Cassa in 1984 was thus saluted as a

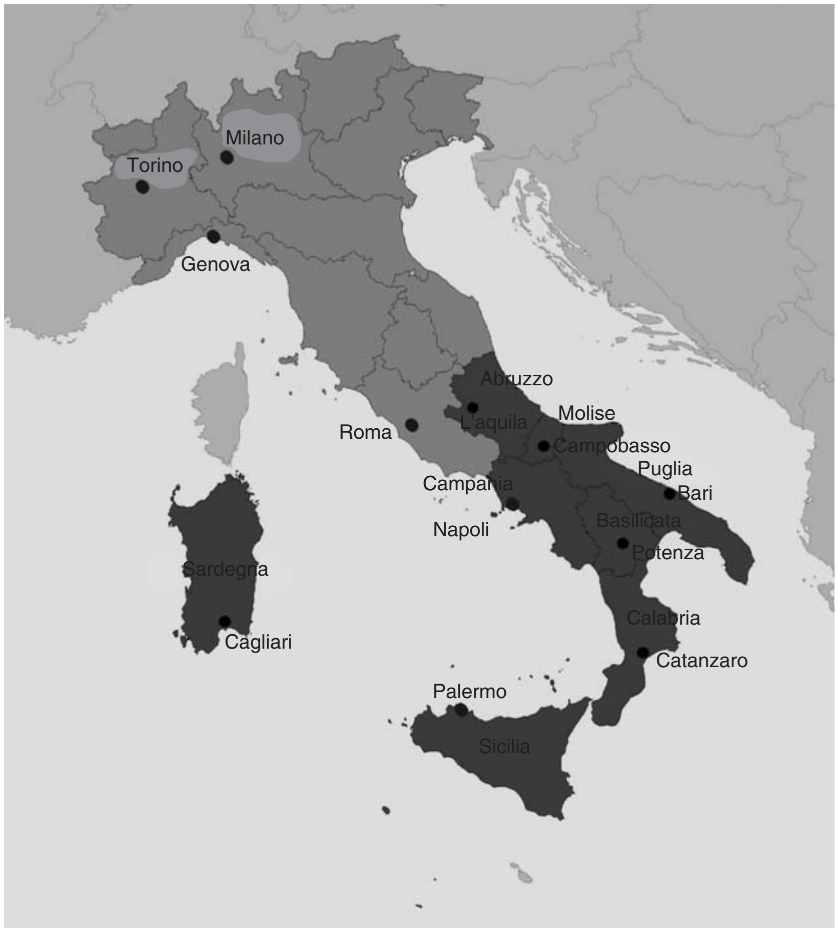

Figure 1 The Mezzogiorno of Italy.

necessary step to restore market mechanisms and make the South take its destiny in its own hands.

Such a reading, however, is higly reductionist and ideologically biased, as it is carried out through the lens of the new local development theories and discourses. Policies must be evaluated in their specific historical and geographical context and looking at their actual objectives. To this purpose, the article that follows re traces the evolution of the Cassa per il Mezzogiorno, taking into account the development paradigms of the time, the national conditioning context in terms of both regulation and accumulation, the aims of policy, as well as the changes in the southern economy and society. In this articulated perspective, the evaluation of the Cassa's impact on the Mezzogiorno turns out inevitably more nuanced and rather positive.

Rise and Decline of State Intervention in Southern Italy: The Cassa per il Mezzogiorno

Three main policy phases can be identified in the evo lution of the Cassa per il Mezzogiorno: (1) the pre industrialization phase (1950–57); (2) the active indus trialization phase (1958–75); and (3) the crisis years (1976–80).

The Pre-industrialization Phase (1950–57): Agrarian Reform and Public Works

In the first phase, policy for the Mezzogiorno was limited to sustaining regional employment and income, while improving the external conditions of accumulation.

The economic and political debate

At the end of World War II, a major focus of the eco nomic and political debate in the newborn Italian re public was the role of the state in a perspective of trade liberalization and international capitalist integration. The Right wing parties were against any form of State intervention, whereas the Left wing coalition was in favor of planning and of structural institutional reforms. With regard to the Mezzogiorno, except for the extreme right, most political forces favored some form of public action to deal with the severe unemployment and pov erty problems of the region and to modernize its very backward productive structure. Partly drawing on Gramsci's analysis of the Southern Question, the Left and some of the Moderates stressed the need for an agrarian reform and for an industrialization program. The latter view was especially advocated by the SVIMEZ – Associazione per lo Sviluppo dell'Industria nel Mezzogiorno, a research association funded in 1946 by prominent Italian intellectuals and economists and still active in promoting studies and initiatives.

Politics and policies

The first Right Center Republican government, elected in 1948, adopted a noninterventionist approach, but could not avoid tackling the problem of the South, where episodes of land occupation and bloody clashes with the police had occurred. In 1950, two important measures were authorized: a partial agrarian reform and the Intervento Straordinario (IS), a special program of public works. Both measures had economic, but also political consensus objectives. It was in these early years that the Italian style of developmental state was established, through what has been called a 'compromise without reforms': no structural changes of the ordinary State apparatus were implemented, whereas an extensive use was made of special public institutions.

The agrarian reform aimed at modernizing the feudal structure of southern agriculture, by expropriating the most unproductive portions of the large estates (latifundia) and redistributing them to landless peasants in the form of small holdings. It also contributed to defuse the potentially subversive aggregation of such peasants, transforming them into a new class of small landowners.

The IS was initially conceived as a true Keynesian program of public works, a compromise with the northern industrial interests, which opposed any strategy of southern industrialization. It had two major aims: (1) sustaining employment and income, thereby activating local multiplier effects; and (2) improving the environment for private investment through infrastructural investment and agricultural modernization. The operational architecture of the IS was highly centralized: a special State agency was created, the Cassa per il Mezzogiorno, with the exclusive authority to design, finance, and implement such investment programs, in addition and in close coordination with ordinary public expenditures.

Policy implementation

Throughout the 1950s, regional policy for the Mezzogiorno was rather efficient and effective, also because of its independence from cumbersome ordinary spending procedures. Within the agrarian reform, about 650 000 ha were expropriated and assigned to landless peasants. With regard to the IS, the quasi totality of the Cassa's expenditures in this phase went to public works (**Table 1**), the majority of which was in support of agricultural modernization (land reclamation, irrigation systems, rural infrastructure) and the rest for civilian infrastructure (transport, aqueducts, sewage). Investment programs were relatively integrated, extraordinary expenditures effectively additional to ordinary spending, and infrastructures of a high engineering quality.

Economic and social transformations

In these years, the Mezzogiorno experienced important changes, although not as radical as those of the subsequent decades.

The agrarian reform contributed to relieve the unemployment pressure and to increase agricultural production and productivity, although it could not prevent a massive exodus of rural population from the mountains and the most marginal areas. The program of public works was in turn quite successful in sustaining both

Table 1 Expenditures by the Intervento Straordinario in the Mezzogiorno 1950 92

Periods	*Direct expenditures by the Cassa and the Agency (Law 64/1986)*			*Transfers to regional and local administrations (%)*	*Total expenditures (%)*	*Average annual expenditures (billion liras at constant prices 1992)*	*Total expenditures by the Intervento Straordinario as a % of Italian GDP (current prices)*
	Direct investment in infrastructure (%)	*Financial incentives to investment (%)*	*Other measures (%)*				
1950 57	84.8	4.8	10.4		100.0	2295.8	0.70
1958 65	55.8	16.7	27.5		100.0	3582.3	0.74
1966 70	42.8	33.8	23.4		100.0	4893.4	0.70
1971 75	52.7	34.1	13.2		100.0	8226.9	0.90
1976 80	66.3	26.1	7.6		100.0	9753.0	0.90
1981 86	72.8	19.7	7.6		100.0	8051.4	0.65
1987 92	47.5	24.2	12.4	15.9	100.0	7708.5	0.53

From Marciani, G. E. (1993). La spesa della Cassa e dell'Agenzia per il Mezzogiorno nei quarantatre anni dell'Intervento Straordinario. *Rivista Economica del Mezzogiorno* 3, 673 689; SVIMEZ (1993). *Rapporto 1993 sull'economia del Mezzogiorno.* Bologna: Il Mulino.

employment and income. It also achieved important multiplier effects, as the increased local demand spurred a temporary growth in the local manufacturing of con sumer goods, still sheltered from external competition.

In fact, despite the rapid growth of the north, the gap between the two parts of the country did not increase: throughout the 1950s, the region's per capita GDP and per capita consumption remained a stable proportion of the north central values (respectively 53–55% and 61–62%). Per capita investment actually increased from 55% to 62%, as a result of public works (**Table 2**). Southern employment remained stable (**Table 3**): the heavy losses in agriculture were compensated by significant increases in the construction sector (owing to public works) and in manufacturing (as a consequence of the increased local demand).

However, southern unemployment pressures were mostly relieved by out migration: between 1951 and 1961, 2 054 000 people (12% of the southern population!) left the Mezzogiorno, initially bound to North America, Australia, and Northern Europe and later in the decade to feed the 'economic miracle' in Northern Italy.

While supporting the economy of the Mezzogiorno, the IS in this period achieved a major historical trans formation of the southern social basis: the unproductive agrarian aristocracy – a pillar of the post unification historic social bloc described by Gramsci – was substi tuted by a new social basis of consensus, formed by the new class of small landowners, the petty bourgeoisie of the expanding retail and service sector, as well as the local administrative bourgeoisie, subordinate to Central State policy choices.

In these years, the south began to be integrated in the national accumulation model, albeit with a still passive role: as a reservoir of labor force, as an outlet for con struction materials and equipment manufactured in the north, and as a basis of consensus for the hegemonic national political parties.

Table 2 Per capita GDP, consumption, and investment in the Mezzogiorno, as a % of north-central Italy, 1950 1996

Years	*Per capita GDP in the Mezzogiorno as a % of N C Italy (current prices)*	*Per capita household consumption in the Mezzogiorno as a % of N C Italy (constant prices)*	*Per capita gross investment in the Mezzogiorno as a % of N C Italy (constant prices)*
1951	53.4	61.4	55.2
1956	55.2	61.2	62.2
1960	53.1	62.0	62.0
1964	56.5	64.2	72.8
1969	58.0	66.1	77.6
1971	61.3	69.9	91.8
1974	60.7	71.3	84.2
1980	57.8	69.7	71.6
1986	58.0	70.6	73.3
1992	58.0	69.7	63.5
1996	55.6	66.7	57.8

From Del Monte, A. and Giannola, A. (1978). *Il Mezzogiorno nell'economia italiana*. Bologna: Il Mulino; SVIMEZ (1993). *I conti economici del Centro Nord e del Mezzogiorno nel ventennio 1970 89*. Bologna: Il Mulino; SVIMEZ (1999). *Rapporto 1999 sull'economia del Mezzogiorno*. Bologna: Il Mulino; SVIMEZ (2006). *Rapporto 2006 sull'economia del Mezzogiorno*. Bologna: Il Mulino.

The Industrialization Phase (1958–75): Growth Poles and Exogenous Fordist Investment

At the end of the 1950s, a major turn occurred in regional policy, with the beginning of an active industrialization strategy. This shift was explained by several concurrent factors: new developments in the economic debate, the changed national political conditions, and the take off of the Italian economy – the so called 'economic miracle'.

The economic and political debate

Starting in the late 1950s, the international economic development and policy debate was increasingly framed in the positivist paradigm of growth, industrialization, and modernization. At the same time, the Central State began playing a growing role in many European coun tries, actively supporting national accumulation through measures well beyond traditional macroeconomic policy – direct ownership in strategic sectors, infra structural investment, support of national industries – and featuring what has been called the developmental state.

In the same years, the issue of regional inequalities also came to prominence through the works of econo mists such as Myrdal and Hirschman, while theories such as the 'big push' by Rosenstein Rodan and the 'growth pole' by Perroux strongly informed regional policies. In the course of the 1960s, many regional development agencies and programs were established throughout Europe.

Politics and policies

In Italy, a Center Left government coalition was formed in the early 1960s, creating the conditions for imple menting such new approaches. The 1960s and 1970s were thus characterized by the national economic program ming experience, a growing role of State holdings in strategic industrial sectors (steel, chemicals, energy) fueling the economic miracle, and the direct support of southern industrialization, despite the continued resist ance of northern industrial groups. The economic mir acle, in turn, allowed for greater public expenditures and increased allocations to the IS.

The 1960s also witnessed a progressive strengthening of the labor movement, with two major waves of conflicts

Table 3 Variations (%) in employment, 1951 80

Periods	*Total employment*		*Manufacturing Employment*		*Construction employment*	
	Mezzogiorno	*North-central Italy*	*Mezzogiorno*	*North-central Italy*	*Mezzogiorno*	*North-central Italy*
1951 59	0.5	5.8	19.4	17.6	63.7	44.1
1959 69	7.0	3.8	3.0	13.7	32.2	10.3
1970 75	2.1	4.2	10.9	3.4	4.7	14.6
1975 80	6.5	6.8	11.0	4.5	4.9	0.5

From Del Monte, A. and Giannola, A. (1978). *Il Mezzogiorno nell'economia italiana*. Bologna: Il Mulino; SVIMEZ (1993). *I conti economici del Centro Nord e del Mezzogiorno nel ventennio 1970 89*. Bologna: Il Mulino.

and strikes: the first in 1962–63, leading to the estab lishment of collective bargaining procedures at firm level and some wage increases; the second in 1969–70 (the so called Hot Autumn), characterized by a much larger social mobilization for social reforms.

The new industrialization strategy for the Mezzogiorno was enacted through several legislative steps (**Box 1**). The Cassa per il Mezzogiorno was reauthorized and its funding were further increased. In addition to infrastructural in vestment, a robust program of very generous financial incentives to industrial investment and a growth poles strategy were launched. Initially geared to small and me dium firms, the incentives were progressively extended to large corporations.

A major institutional change occurred in 1970 with the establishment of the long delayed regional governments. This administrative decentralization involved a slight re organization of the IS, shifting some project formulation responsibility from the Cassa to the newly established southern regional governments (**Box 1**).

Policy implementation

The increased state effort in sustaining southern indus trialization was clearly reflected in the Cassa's expen ditures. Average annual spending almost tripled and increased also as a share of total national resources. The lion's share now went to industrial incentives, but spending on infrastructure remained relevant in absolute terms (**Table 1**). On the other hand, the additional character of the IS expenditures in this period signifi cantly declined, as public works financed by the Cassa ended up substituting for ordinary infrastructural in vestment. Moreover, starting in the late 1960s, the man agement of both the IS and the state holding system began losing autonomy and was progressively con ditioned by national politics.

Economic and social change

Throughout the 1960s, the Italian economy achieved impressive growth rates, which brought the country among the top five industrialized nations of the world. Italian industrial capital was now characterized by a bi polar structure – the large private northern industrial groups – mostly operating in the north west – and the increasingly powerful state holdings – which followed a more articulated locational strategy.

Both the economic miracle and the industrialization policy had a powerful impact on the Mezzogiorno. This was the period of the 'great transformation', when the southern economy and society were rapidly modernized and fully integrated into the Italian capitalist model – when the gap between the region and the rest of the country, according to basic macroeconomic indicators, was significantly reduced for the first and last time (**Table 2**). The years 1960–75 can indeed be considered the golden age of southern industrialization. From an average 16% of national industrial investment in the 1950s, the Mezzogiorno's share reached a peak of 33% in 1970–73 (**Table 4**). Manufacturing investment was prevailingly of exogenous origin, but its composition evolved over time. Two main waves can be identified.

In the 1960s, it was mainly the State holdings – which were locationally footloose and needed capital to fuel their expansion – that exploited the financial incentives for very large investment projects in the steel, chemical, and petrochemical sectors. However, in spite of such massive investment and the thousands of jobs created in the new factories, southern manufacturing employment in this decade declined (− 3% compared to + 14% in the rest of the country), as a consequence of the crisis of small local firms (**Table 3**). In fact, with the improvement of transport infrastructure, the more competitive con sumer goods manufactured in the north flooded the south and wiped out local production. The continued expan sion of the construction and service sectors only partially compensated for these losses. In this decade, although international migration declined, internal migration continued: between 1961 and 1971 another 2 000 000 people left the Mezzogiorno and migrated to the large cities of north central Italy.

The second industrialization wave occurred in the first years of the 1970s. This time, large private northern firms operating in the automobile, electromechanical equipment, and other modern sectors began to exploit industrial incentives for relocating production in the Mezzogiorno. It was a late Fordist strategy to evade the labor conflicts in the north and exploit the less unionized

Box 1 Italian regional policy for the Mezzogiorno 1950–92.

The Cassa per il Mezzogiorno (1950–80)

1950 57. The pre-industrialization phase: Public works and agrarian reform.

The first phase of regional policy for the Mezzogiorno the so-called Intervento Straordinario was regulated by the Law 646/1950, which established the Cassa per il Mezzogiorno, and by the subsequent Law 949/1952. The Cassa was to design, finance, and implement a 10-year program (later extended to 12 years) of extraordinary public works. The program was aimed at the modernization of agriculture (land reclamation, hydrogeological stabilization, irrigation systems, creation of rural centers), in close connection with the agrarian reform, and at the creation of basic civilian infrastructures (roads, sewage, water supply).

1958 75. The active industrialization phase: Industrial incentives and industrial development areas

In 1957, the Law 634/1957 extended the life of the Cassa until 1965 and almost doubled its budget. It also introduced important new measures in support of industrialization:

- a system of financial incentives to industrial investment (grants and subsidized loans to industrial firms investing in the Mezzogiorno); initially geared to small and medium firms, the incentives were later extended to any size of investment (Laws 555/1959 and 1462/1962), in order to accommodate large firms' investment projects;
- the promotion of industrial development areas (Aree di Sviluppo Industriale and Nuclei di Sviluppo Industriale) among consortia of municipalities, which would be privileged in the granting of financial incentives to industrial investment, as well as infrastructural investment;
- a reserve clause, that is, the obligation for all state-controlled firms and institutions to place at least 40% of their total investment and 60% of their new investment in the south, as well as the obligation for public institutions to purchase at least 30% of their inputs from firms located in the south.

The legislation also established a new Ministries Committee for the Mezzogiorno, with the task of coordinating extraordinary intervention with ordinary expenditures.

In 1965 the Law 717/1965 extended the life of the Cassa until 1980 and re-financed its activities for further five years. Industrial incentives and industrial development areas were confirmed, but action in the Mezzogiorno was to be integrated into the national economic programming framework, therefore brought under the coordination of the CIPE (Interministry Committee for Economic Programming). Moreover, starting in 1969, very large industrial investment projects had to be negotiated with the Interministry Committee and could obtain incentives above the usual standard. Finally, since the Cassa spending over the last period of operation had ended up substituting for ordinary expenditures, the new legislation emphasized that the Intervento Straordinario was to be additional to the normal share of public spending in the south.

In 1971 the Law 853/1971 refinanced the Cassa for another 5 years and introduced a number of important innovations, among which was the transfer of some policy responsibilities to the newly formed regional governments:

- the link with national economic planning was re-enforced; the Ministries Committee for the Mezzogiorno was abolished and its coordinating duties taken up by the CIPE (Interministry Committee for Economic Programming);
- large industrial groups had to obtain the CIPE approval for any large scale investment project on the national territory (the first attempt to establish concertation between large firms and the state, through the contractual agreement formula);
- infrastructural projects received new emphasis in the form of Special Projects, that is, large, intersectoral, integrated projects of regional or interregional interest.
- the task of designing Special Projects and proposing them to the CIPE for approval was entrusted to regional governments; to this end, a new Committee of the Presidents of Southern Regional Governments was established, while the Cassa was to provide technical assistance.

1976 84. After the crisis

In 1976 the Law 163/1976 refinanced the Cassa per il Mezzogiorno until 1980. Among the new features were:

- further fiscal exemptions for industrial firms investing in the south;
- a new form of financial incentive to support the creation of industrial employment, that is, a 10-year full subsidization of social security payments for any new industrial job created;
- a further focus on special projects;
- the explicit mandate to the Cassa to provide technical assistance and consulting to those regional government that requested it.

The Law 64/1986 (1985–93)

The Law 64/86 defined the new organic discipline and resources of the Intervento Straordinario in the Mezzogiorno for 9 years (1985 93). In organizational terms, it introduced a distributed system of responsibilities:

1. A new Department for the Mezzogiorno was to be established within the Ministry of the Mezzogiorno, with the following tasks: strategic formulation of sectoral priorities and budget allocation; evaluation and approval of investment projects submitted for financing; and coordination of extraordinary and ordinary expenditures.
2. A new Agency for the Mezzogiorno was also established, with the limited task of managing the funds allocated by the Department. Once programs and projects were approved by the Department, the Agency had to draft contracts with the entitled subjects and proceed to payments.
3. Regional and local governments were fully entrusted with the task of formulating local development programs and drafting investment projects, as well as implementing such programs and projects once approved by the Department and funded.

Content-wise, the new legislation significantly enlarged the domains of policy intervention, including industrial restructuring, business services, research and development, technological innovation, and training.

The existing toolkit of financial incentives was mostly confirmed: individual firms could apply for financial incentives, but big firms could negotiate large scale investment directly with the Department, through the *Contratti di Programma*. With regard to major infrastructural projects, involving different government levels, *Accordi di Programma* could be drafted and submitted to the Department.

The reserve clause for state enterprises and institutions was also reconfirmed.

Table 4 Investment in the Mezzogiorno as a % of Italy, 1951 1999

Periods	Gross total investment in the Mezzogiorno as a % of Italy	Gross industrial investment in the Mezzogiorno as a % of Italy[a]	Gross manufacturing investment in the Mezzogiorno as a % of Italy
1951 59	24.5	15.8	
1960 64	26.2	23.5	
1965 69	28.1	25.4	
1970 73	32.0	33.0	30.6
1974 80	30.8	27.7	23.5
1981 92	28.7		17.6
1993 99	23.4		17.1

[a] Includes the construction industry.

From Del Monte, A. and Giannola, A. (1978). *Il Mezzogiorno nell'eco nomia italiana*. Bologna: Il Mulino (current prices); SVIMEZ (1993). *I conti economici del Centro Nord e del Mezzogiorno nel ventennio 1970 89*. Bologna: Il Mulino (constant prices 1985); SVIMEZ (2000). *Rapporto 2000 sull'economia del Mezzogiorno*. Bologna: Il Mulino (constant prices 1990); SVIMEZ (2006). *Rapporto 2006 sull'economia del Mezzogiorno*. Bologna: Il Mulino (constant prices 1995).

workforce of the South. This time, manufacturing in vestment yielded a quite relevant employment increase (+11% between 1970 and 1975), for the first time higher than in north central Italy (**Table 3**).

In this policy phase, the Mezzogiorno performed two new functions within the national accumulation model: it represented a subsidized investment terrain for the ex panding and, later, relocating national Fordist corpor ations, as well as a market outlet for northern consumer products. The southern social structure was further profoundly transformed: a modern working class was created in the new Fordist factories, while a service middle class formed as a consequence of the further expansion of both collective and private services. By the mid 1970s, the south was fully integrated into the na tional affluent consumption model. Particularly relevant was the rise of relatively powerful local social blocs – made of the local political and administrative bourgeoisie, professionals, real estate and construction interests – lobbying for central state funding and sub ordinately ingrained in the growing political manage ment of the IS, through the national party system.

After the Crisis (1976–80): The End of Late-Fordist Industrialization

The economic crisis of the mid 1970s, as we now know, marked an epochal change. It was the end of the Fordist–Keynesian model of accumulation, based on mass pro duction, industrial growth, and expanding public expenditures.

Politics and policy

In Italy, the radicalization of social conflicts and the end of the Fordist growth starting with the first oil shock of 1973 eventually sapped the national economic pro gramming ambitions and the IS, bringing down the overall accumulation model. However, the structural character of the crisis was not immediately perceived.

At the national level, depressive macroeconomic policies were implemented to reduce inflation, which caused a major economic downturn. Later programs to support industrial restructuring were mostly used by firms in declining sectors to subsidize redundant em ployment. The state holdings system was burdened with rescue operations – that is, the acquisition of collapsing private firms – and employment support.

With regard to the Mezzogiorno, the Cassa was re financed in 1976 for the last time. The only relevant in novations were a new subsidy to employment creation and a further involvement of regional governments (**Box 1**). In 1980, the Cassa was due to expire, but its operations were repeatedly extended until 1984, when it was eventually terminated. The IS was to continue through a radically revised institutional architecture.

Policy implementation

Despite the economic crisis, the Cassa's expenditures in the last part of the 1970s further increased (**Table 1**). They mostly went into infrastructural investment, whereas industrial incentives, after the peak of 1971–75, declined and mostly concerned the completion of pre vious projects. On the organizational side, although since 1971 the Cassa was supposed to shift some policy for mulation responsibilities to the regional governments and evolve into a technical assistance agency, the transitional norms limited the transfer of competences by allowing the Cassa to complete all previously approved projects. On their side, the new regional administrations exhibited little technical capabilities and political willingness to take over a more propositional role.

Economic and social transformations

The years following the 1974–75 downturn – the worst since the Great Depression – were characterized in the entire industrialized world by the beginning of a pro found industrial restructuring process.

In the Mezzogiorno, the crisis meant the abrupt interruption of the Fordist industrialization process led by exogenous manufacturing investment. On the other hand, partly as a consequence of the industrialization policy, a revival of local entrepreneurial capabilities was observed, with the start up of several new small and medium firms, manufacturing either consumer goods or components for the exogenous plants.

In fact, the 1970s were, on the whole, a very positive decade for southern employment (**Table 3**). Between 1971 and 1981, while out migration virtually ended, about 150 000 manufacturing jobs were generated, 40% of which in the large establishments (>500 employees) created by Fordist investment. Employment in services continued to expand, especially in the public sector, also as a consequence of the new regional administrations, whereas construction employment declined.

Social change in the south during the second half of the 1970s was also relevant. The national political mobilization for greater democracy and social justice, that since 1968–69 had spread from factory workers to groups of the growing middle class, had reached its peak in the mid 1970s and involved for the first time also the Mezzogiorno. The success of the Left in the 1975 and 1976 elections was achieved to a large extent with the support of the traditionally conservative south.

But precisely when the above processes of social modernization and entrepreneurial awakening were gaining momentum, they were undermined by the escalation of terrorism, on the other hand, and the economic crisis, on the other hand. The end of growth and the degeneration of politics in the subsequent decade would shatter the new found mobilization of southern society.

In this last policy phase, the functional role of the south in the national accumulation model was significantly reduced: no longer a reservoir of labor, neither a subsidized terrain for industrial expansion, the Mezzogiorno remained a mere pool of electoral consensus. The further empowerment of regional governments, in a context of faltering market mechanisms, contributed to reinforce the local power blocs and their political management of public expenditures. Hiring procedures, the granting of permits, and public budget allocations were increasingly used to strengthen local political control.

State, Regional Policy, and Southern Development in the Cassa's Years: An Assessment

Between 1950 and 1980 the Mezzogiorno experienced the most dramatic transformation of its history, at a pace and to an extent that can only be understood now, in a historical perspective. In 30 years – less than half a lifetime – the Mezzogiorno evolved from an agricultural, semi feudal, poverty stricken society into a modern, partially industrialized, and relatively affluent society. The region was fully integrated – although still in a dependent fashion – into the national model of capitalist accumulation, mass consumption, and welfare. The gap with the rest of the country in terms of production, consumption, investment, physical and social infrastructure, although it remained important, was significantly reduced (**Table 2**).

How much was such an impressive change the result of the Cassa per il Mezzogiorno action? Evaluating the latter's impact is difficult, for two reasons. First, in policy evaluation it is not easy to isolate the actual effects of programs from what would have happened independently of them; second, the current debate is highly ideologically biased.

The widely accepted view – even in Left wing circles – is that the Cassa per il Mezzogiorno – and the IS in general – was an epochal mistake: at best, it was a waste of public resources, which created cathedrals in the desert, prevented the release of the region's endogenous entrepreneurial potential, and failed to generate a self sustaining development model; at worst, it was the chief cause of the parasitic, clientelistic, and corrupt degeneration of southern society. Only few observers and public actors – considered nostalgic – stoutly defend the action of the Cassa and the intervention of the state in the Mezzogiorno.

Between these two extremes, there is room for a more dispassionate and nuanced assessment, that is, contextualizing the Cassa's operations with regard to the specific historical period, the contingent problems of the Mezzogiorno, and the actual objectives of regional policy. Three aspects shall be considered: (1) the Cassa's approach and tools (2) its impact, and (3) its organizational structure.

In what concerns the toolkit, the Cassa per il Mezzogiorno was an advanced product of its time. Its approach to development and its actual programs were coherent with the contemporary scientific and policy paradigm, centered on growth, infrastructural investment, industrialization, and modernization. Most importantly, they were consistent with the contemporary Fordist accumulation model, based on transport infrastructure and vertically integrated mass production. Also, the use of state holdings as growth engines was in line with practices adopted in many European countries.

In what concerns the actual impact of the Cassa's programs, despite the above mentioned methodological difficulties, several results are above dispute. In the first 20 years, infrastructural investment was very effective in improving the economic and social environment of the Mezzogiorno, in reducing communication and transportation disparities, and in logistically integrating the two parts of the country, while fulfilling the Keynesian objectives of employment and income support. In a period when the north was growing at an accelerated pace, the IS was able to pull the Mezzogiorno along, so that the divide did not increase.

As to the industrialization strategy, policy was undoubtedly successful in creating a modern manufacturing basis. Industrial investment in the Mezzogiorno steadily increased in absolute and relative terms throughout the 1960s, reaching a peak in the early 1970s and bringing the productive structure of the region very close to that of the rest of the country: at the end of the 1970s, industrial employment had reached 29% of total southern employment, compared to 39% in north central Italy, whereas services accounted for 48% and 52%, respect ively. Some observers have criticized the limited support to local firms, but without the investment by state holdings and exogenous Fordist firms, an industrialization process of that magnitude could not have occurred. The fact that such a late Fordist industrialization model came to an end, then, can hardly be blamed on the Cassa per il Mezzogiorno, since the crisis of the accumulation regime involved the whole country and factories closed down in the south as well as in the north. The south was worst hit because of its still young and fragile socioeconomic structure.

From the point of view of the institutional archi tecture, the excessively independent and centralized structure of the Cassa has been blamed because it was prone to discretionary behavior and prevented the em powerment of local administrations. This was actually a conscious choice, dictated by the urgency of the southern problems after World War II and the need for agile and coordinated action, which could not have been ensured through the ordinary administration of the state, or by the weak local authorities. The Cassa was the natural and most effective solution within the developmental state approach of the Fordist–Keynesian regime, as proved also by its implementation records.

Finally, in evaluating the impact of the Cassa one should also consider that the amount of additional state resources devolved to the Cassa, for the Mezzogiorno were never truly additional. Repeatedly – in the 1960s and 1970s – the Cassa's investment in the Mezzogiorno substituted for the due share of infrastructural investment by the ordinary administration of the state.

This said, a number of negative side effects must also be mentioned. First, the success of a special institution such as the Cassa was contingent upon the commitment and technical capabilities of its staff. When technocrats began to be substituted by political appointees and pol itics began prevailing over economic strategies, the Cassa's expenditures – as in other state institutions – lost effectiveness. Second, the choice of a special, centrally managed state did contribute to de legitimize ordinary state procedures and to consolidate the passive and subordinate role of the local administrations it was to surrogate for. When politics took over in the late 1970s, the centralized institutional architecture of the Cassa – although it cannot be considered a cause – certainly became functional to the formation of a clientelistic local administrative culture.

In conclusion, regional policy implemented by the Italian government through the Cassa per il Mezzogiorno was a powerful and quite effective financial and technical effort, consistent with the theoretical and policy para digms of the time. Most importantly, the IS was coherent with the Italian accumulation strategies of the Fordist–Keynesian period, in which the Mezzogiorno did have a functional role. Without such an effort the Mezzogiorno would certainly have been left behind. On the other hand, the very exogenous character of the Cassa's action, while functional to quickly producing convergence – via infrastructure, industrialization, modernization, and in tegration – also undermined the sustainability of the 'great transformation' it had engineered. The end of the national accumulation model that had fueled this con vergence occurred too early to allow the socioeconomic model established in the Mezzogiorno to survive on its own legs.

Epilog

In the whole Western world, the 1980s was a period of transition, marked by the demise of the developmental state approach, the rise of neoliberalism, and the search for a new balance between regulation and accumulation. Squeezed in between heightened international com petition and increased internal conflicts, and failing to identify a new national accumulation strategy, the Italian state abdicated its developmental role. Economic policy became defensive and piecemeal. State holdings were increasingly politically controlled, betraying their ori ginal strategic role. The overall management of public expenditures was progressively bent to gain political consensus, rather than support accumulation. It is in this decade that corruption in the assignment of public con tracts reached unprecedented levels – in the north and the south alike – and a huge public debt was accumu lated, eventually leading to the Tangentopoli scandal and the fiscal crisis of the State in the early 1990s.

As to regional policy, 2 years after the termination of the Cassa and after a difficult parliamentary debate, the Law 64/1986 was approved, establishing new rules for the IS. It was the last – half hearted – attempt to reform Italian regional policy, trying to take into account both the changed features of post Fordist accumulation and the need to fully decentralize policymaking to regional and local governments, while maintaining central co ordination (**Box 1**).

Although sound on paper, the new IS legislation was the death rattle of Italian regional policy. Delays in the organizational restructuring and systematic underfunding undermined the implementation of the new policy framework, while, on their side, regional and local gov ernments exhibited a protracted resistance to give up their passive roles of recipients of central state transfers and to take over the new managerial role expected from them.

By the end of the 1980s, as the neoliberal ideology gained consensus with its corollaries of deregulation and privatization, the IS had become the culprit for all the Mezzogiorno's evils. Pressures to abolish national top down regional policy and establish decentralized bottom up programs in tune with the new local development theories came from within and from without. On the one hand, small and medium firms operating in central and northeastern Italy – the so called Third Italy, of which industrial districts were the nobler and most popular version – had substituted Fordist corporations in both the Italian accumulation model and development discourses, endorsing the rising paradigm of local development. The northeastern component of small firms had actually found a political voice in the new federalist party Lega, which strongly opposed any redistributive policy in favor of the south. On the other hand, the 1988 reform of the European structural funds and the beginning of European cohesion policies undermined the legitimacy of national regional policies deemed to distort competition.

In this hostile economic and political context, made all the more difficult by the fiscal crisis of the state and the constraints imposed by the Maastricht Treaty, the Mezzogiorno no longer fulfilled any functional role. In 1992 – before its natural expiration – the IS was altogether terminated. The term Mezzogiorno disappeared from Italian official documents and was substituted by the Euro term 'lagging areas'. The southern question ceased to be a national question and became a local issue, to be dealt with by ordinary or European local development programs, which were extended to the overall national territory.

The effect of this abrupt end of additional public investment in the Mezzogiorno was severe: by the mid 1990s the gap between the two parts of the countries had resumed pre industrialization levels (**Table 2**). Most importantly, unemployment soared and reached rates above 20%, that is, three times that of the rest of the country. This time it was no longer unskilled rural unemployment, but educated youth and women.

National concern for the Mezzogiorno resumed in 1998 – once Italy had achieved the required conditions to enter the European Monetary Union – with the creation of the department for development policies (DPS), a structure within the Ministry of Economy and Finance charged with the task of coordinating structural funds resources and local development programs. But, 15 years after the termination of the IS and the decentralization of regional policy, the north–south divide has not yet abated.

See also: Developmentalism; Fordism, Post-Fordism and Flexible Specialization; Growth Poles, Growth Centers; Industrial Districts; Industrialization; Local Development; Regional Inequalities; Regional Planning and Development Theories; Uneven Regional Development.

Further Reading

Barbagallo, F. and Bruno, G. (1997). Espansione e deriva del Mezzogiorno. In Barbagallo, F., Barone, G., Bruno, G. *et al*. (eds.) *Storia dell'Italia repubblicana*, vol. 3, Tome II, pp 399 470. Torino: Einaudi.

Barca, F. (ed.) (1997). *Storia del capitalismo italiano dal dopoguerra ad oggi*. Rome: Donzelli.

Cafiero, S. (2000). *Storia dell'Intervento Straordinario nel Mezzogiorno (1950 1993)*. Rome: Piero Lacaita Editore.

Cafiero, S. and Marciani, G. E. (1991). Quarant'anni di Intervento Straordinario nel Mezzogiorno (1950 1989). *Rivista Economica del Mezzogiorno* 2, 249 274.

D'Antone, L. (ed.) (1996). *Radici storiche ed esperienze dell'Intervento Straordinario*. Napoli: Bibliopolis.

Del Monte, A. and Giannola, A. (1978). *Il Mezzogiorno nell'economia italiana*. Bologna: Il Mulino.

Dunford, M. (1988). *Capital, the State and Regional Development*. London: Pion.

Gramsci, A. (1978). Some aspects of the southern question. In *Selections from Political Writings 1921 26*, pp 441 462. London: Lawrence & Wishart.

Graziani, A. (1978). The Mezzogiorno in the Italian economy. *Cambridge Journal of Economics* 2, 355 372.

Graziani, A. and Pugliese, E. (eds.) (1979). *Investimenti e disoccupazione nel Mezzogiorno*. Bologna: Il Mulino.

Marciani, G. E. (1993). La spesa della Cassa e dell'Agenzia per il Mezzogiorno nei quarantatre anni dell'Intervento Straordinario. *Rivista Economica del Mezzogiorno* 3, 673 689.

Martinelli, F. (1985). Public policy and industrial development in Southern Italy: anatomy of a dependent economy . *International Journal of Urban and Regional Research* 9, 47 81.

Perna, A. (1994). *Lo sviluppo insostenibile*. Napoli: Liguori.

Podbielski, A. (1976). *Twenty Five Years of Special Action for the Development of Southern Italy*. Milano: Giuffré.

SVIMEZ (1993). *I conti economici del Centro Nord e del Mezzogiorno nel ventennio 1970 89*. Bologna: Il Mulino.

SVIMEZ (2000). *Rapporto 2000 sull'economia del Mezzogiorno*. Bologna: Il Mulino.

SVIMEZ (2006). *Rapporto 2006 sull'economia del Mezzogiorno*. Bologna: Il Mulino.

Trigilia, C. (1992). *Sviluppo senza autonomia. Effetti perversi delle politiche nel Mezzogiorno*. Bologna: Il Mulino.

Relevant Websites

http://www.dps.mef.gov.it
DPS Dipartimento per le Politiche di Sviluppo, Ministero dello Sviluppo Economico.

http://www.formez.it
FORMEZ.

http://sviluppolocole.formez.it
FORMEZ.

http://europa.formiz.it/
FORMEZ.

http://www.svimez.it
SVIMEZ.

Categorical Data Analysis

S. D. Withers, University of Washington, Seattle, WA, USA

Glossary

Binomial Distribution A discrete probability distribution associated with situations in which only two outcomes are possible, for example, a coin toss.

Contingency Table A cross-tabulation of counts corresponding to two or more variables measured at nominal or ordinal levels used to analyze their conditional relationships and association.

Logistic Regression A regression model for a binomially distributed response (dependent) variable used to determine the probability of an event occurring as a function of other factors (dependent variables).

Logit The natural logarithm of the odds; also called the log-odds.

Loglinear Models A method of assessing the conditional relationships among three or more categorical variables.

Maximum Likelihood Estimation A mathematically intense procedure, usually conducted by computers, to determine the parameters that maximize the probability (likelihood) of the observed sample data.

Multinomial Distribution A discrete probability distribution associated with the situation where more than two fixed outcomes are possible, for example, a die toss.

Odds The ratio of the probability that an event will occur divided by the probability that the event will not occur. Unlike probabilities, odds can have values greater than 1.

Odds Ratio The comparison (ratio) of two conditional odds.

Parsimonious Model A loglinear model that includes as few effects (parameters) as needed to provide a goodness-of-fit with observed data. It is the opposite of the saturated model that includes all possible effects.

Poisson Distribution A probability distribution for discrete random variables, typically a count of the number of events that occur in a certain time interval or spatial area.

Probability The number of outcomes of a specific type expressed as a proportion of the total number of outcomes possible. Probabilities have values ranging from 0 to 1.

Probit A regression model similar to logistic regression but based on the cumulative normal probability distribution rather than the log-odds as in the logit.

Saturated Model A loglinear model that includes all possible main and interaction effects among variables in a contingency table. Such a model is said to be saturated since all available information (every cell) is specified.

The Origins and Foundations of Categorical Data Methods

Virtually all introductory quantitative methods courses begin with an introduction to the levels of measurement: nominal, ordinal, interval, and ratio. Expressed in this order, they range from little to considerable detail in the information provided. The most important distinguishing factor, of course, is that nominal and ordinal measurements are categorical and discrete, whereas the interval and ratio measurements are continuous. Part of the elementary importance of discerning the levels of measurement is that different measurement levels are associated with different probability functions and methods for hypothesis testing and prediction. This article describes methods appropriate for the study of associations and relationships among categorical variables.

Categorical methods are of particular interest to human geographers for a number of reasons. Many of the phenomena of interest to human geographers are discrete and categorical, and frequently do not lend themselves to measurement on a continuous scale. For example, gender is a dichotomous variable often used by geographers that is discrete rather than continuous. Continuous measures can be scaled down to ordinal or nominal measures when analytically appropriate. But more importantly, many of our theoretical and conceptual ideas lend themselves to empirical validation through the use of distinct categories, for example, distinctions such as urban or rural; above or below the poverty level; employment status that is described as being full time, part time, or unemployed; or economies that are described as traditional, command, market, or mixed. In addition, geographers are frequently interested in being able to understand or predict outcomes such as the probability of an event occurring or not occurring. Examples include events such as residential mobility, migration, shopping behavior, and travel mode selection, as well as many other discrete choices whose outcomes may be dichotomous (binary) or polytomous (unordered or ordered). Lastly, for quite some time, human geographers have been using research methodologies that include primary data collection, as well as secondary data analysis from surveys. Frequently, the responses and information available from surveys are qualitative or grouped, making them unsuitable for traditional statistical methods such as multiple regression modeling. Ordinary least squares (OLS) regression probably is the most commonly known quantitative method to human geographers, but it requires levels of measurement at the continuous scale and parametric assumptions about the normal distributions of the populations from which the data have been sampled. Qualitative measures demand alternative modeling approaches.

Categorical data methods have their origins in the synergies surrounding the integration of statistical methods and the social sciences. As the social sciences embraced quantitative methods, it became clear that the study of social phenomena represented challenges. Discrete measurement levels challenged traditional statistical methodologies, especially the constraints stemming from the normal probability distribution. In the late 1960s, statisticians, such as Cox and Goodman, developed integrative statistical approaches appropriate for categorical data analysis and other probability distributions. Further development and dissemination of categorical methods continued throughout the 1970s, most notably by statistical scholars such as Bishop, Feinberg, Haberman, Holland, and McFadden. These developments in categorical data analysis were embraced quickly by human geographers such as Neil Wrigley and Larry O'Brien – geographers in search of statistical techniques more appropriate for the study of qualitative data and discrete measurements. Wrigley broadly disseminated these methods among human geographers in a 1979 manuscript in *Progress in Human Geography*, just a few years after these innovative methods were published in the statistical journals. Throughout the 1980s and the 1990s categorical data methods came to be widely used in human geographic research, particularly in the fields of migration, residential mobility, discrete choice modeling, and retail and transportation behavior modeling. Now, a generation later, categorical data methods such as logistic regression are commonplace in quantitative human geography and regional science research.

Categorical Methods: Associations and Likelihoods

The simplest form of categorical data analysis is a 2×2 contingency table that represents the cross tabulation of two categorical variables. For example, variable I is cross tabulated with variable J, and each of these variables has two categories (**Table 1**). The most common approach to analyzing such a table is to test for independence between the two variables using a chi square statistic. In a chi square test the observed values in a cross tabulation are compared with the values that would be expected if

Table 1 The 2×2 contigency table and chi-square

Variable I	*Variable J*		
	C	*D*	*Row total*
Observed			
A	80	40	120
B	20	60	80
Column total	100	100	200
Expected			
A	60	60	120
B	40	40	80
Column total	100	100	200

Chi squre, 33.333; significance, 0.000; degrees of freedom, 1; Cramer's V, 0.480; significance, 0.000.

there was no association between the two variables. The expected values are determined by the marginal totals. Each cell's expected value is determined from the marginal row total (R), the marginal column total (C), and the grand total (N, sample size) distributed throughout the table such that:

$$E_{ij} = \frac{R^{*}C}{N} \qquad [1]$$

Then, the chi square statistic is calculated in the following manner:

$$\chi^2 = \sum_{i}^{r}\sum_{j}^{c}\left(\frac{\left(O_{ij} \quad E_{ij}\right)^2}{E_{ij}}\right) \qquad [2]$$

with degrees of freedom equaling $(R-1)^{*}(C-1)$. If the measured chi square value is greater than the critical value given on a chi square table, the null hypothesis of no association can be rejected and there is a statistically significant association between these two categorical variables. This popular method tells us if there is a statistical association between these two variables. The strength of the relationship is often measured using Cramer's V. The same approach is used for variables with more than two categories.

To give this example substantive meaning let us consider gender and journey to work. If we let variable I be mode of travel (where A = car driver and B = bus rider) and variable J be gender (where C = men and D = women), then this study tells us that there is a statistically significant association between mode of travel to work and gender. More men take cars and more women take the bus according to these figures. However, further categorical data analysis can reveal considerably more.

Detailed information about the relationship between two variables and their dichotomous measures can be assessed by calculating probabilities and odds (**Table 2**). A probability is the number of outcomes of a specific type expressed as a proportion of the total number of

Table 2 2×2 Contingency table, marginal[a], and conditional odds[b,c]

Mode of travel	*Gender*		
	Men	*Women*	*Row total*
Car driver	f_{11}	f_{12}	$= f_{11} + f_{12}$
Bus rider	f_{21}	f_{22}	$= f_{21} + f_{22}$
Column total	$= f_{11} + f_{21}$	$= f_{12} + f_{22}$	$= f_{11} + f_{12} + f_{21} + f_{22}$

[a] Marginal odds odds of being in category A, 1.50 (120/80); odds of being in category B, 0.67 (80/120); odds of being in category C, 1.00 (100/100); odds of being in category D, 1.00 (100/100).
[b] Conditional odds odds of A being in category C, 2.00 (80/40); odds of A being in category D, 0.50 (40/80); odds of B being in category C, 0.33 (20/60); odds of B being in category D, 3.00 (60/20); odds of C being in category A, 4.00 (80/20); odds of C being in category B, 0.25 (20/80); odds of D being in category A, 0.67 (40/60); odds of D being in category B, 1.50 (60/40).
[c] Odds ratio $6.00 = (f_{11}/f_{21})/(f_{12}/f_{22})$.
The observed values of Table 2 are same as the observed values of Table 1.

outcomes possible. Probabilities have values ranging from 0 to 1. The ratio of probabilities is called the odds. The odds are measured as the ratio of the probability that an event will occur divided by the probability that the event will not occur. Unlike probabilities, odds can have values greater than 1. **Table 2** provides the same observed values as those listed previously in **Table 1**. Marginal odds and conditional odds have been calculated. Marginal odds tell us about the distribution of frequencies across the table, and are of limited interest. Conditional odds, on the other hand, are quite informative since they provide the odds of being in a specific category of a specific variable. In our substantive example, a car driver is twice as likely to be a man as a woman, whereas a bus rider is thrice as likely to be a woman as a man. In addition, a man is 4 times more likely to be a car driver than a bus rider, and a woman is 1.5 times more likely to be a bus rider than a car driver.

By comparing two conditional odds we get the odds ratio. In this example the odds ratio is:

$$\text{Odds ratio} = (f_{11}/f_{21})/(f_{12}/f_{22}) \qquad [3]$$

When an odds ratio equals 1, the two conditional odds are equal. But, when an odds ratio is greater than 1 the covariation between the variables is direct, and when an odds ratio is less than 1 it indicates an inverse covariation. The direction of covariation is particularly meaningful if the variables are measured at an ordinal level. Since the categories are nominal in this example the interpretation depends merely on the placement in the cross tabulation. Thus, the odds of car driving among men are 6 times greater than the car driving odds of women.

This example has introduced the important concepts of cross classification, probabilities, marginal odds, conditional odds, and odds ratios. The odds ratio is the heart of categorical data analysis. However, the example is limited in its application since it includes only two variables, with only two nominal groupings. Usually we are interested in the relationships among many variables. In addition, we frequently have polytomous groupings. To extend the calculation of conditional odds to more than two variables we turn to two common methods of categorical data analysis: loglinear modeling and logistic regression.

Loglinear Modeling

Loglinear modeling (also referred to as multiway frequency analysis) is appropriate for the study of the association of three or more categorical variables. In a loglinear model the conditional relationship between two or more categorical variables is analyzed by taking the natural logarithm of the cell frequencies within a contingency table. It is appropriate for Poisson distributed data. To continue using our substantive example, an additional categorical variable indicating employment status is added to our two variables, mode of travel and gender. This variable is measured as full time or part time (**Table 3**). We have previously measured the statistical association between mode of travel and gender as being significant ($\chi^2 = 33.33$). Embedded within this larger table are two additional two way relationships. There is a statistically significant relationship between mode of travel and work status ($\chi^2 = 20.00$), as well as between gender and work status ($\chi^2 = 43.22$). The question that remains is which of these relationships are the most salient. Loglinear modeling strategies are designed to answer this question.

Loglinear models do not distinguish between independent and dependent variables. The aim of loglinear modeling is to determine how closely one can reproduce the distribution of frequency counts within the contingency table without including all of the effects. There are three variables within this contingency table: gender (G), mode of travel (T), and employment status (E). From these three variables there are several possible effects of association: three main effects (G, T, and E), three two way interaction effects (GT, GE, and TE), and one

Table 3 Contingency table for three categorical variables

Mode of travel	*Employment type*	*Gender*		
		Men	*Women*	*Row total*
Car driver	Full-time	70	20	90
	Part-time	10	20	30
Bus rider	Full-time	15	20	35
	Part-time	5	40	45
Column total		100	100	200

three way interaction effect (GTE). If we fit a log linear model with all of these terms it would reproduce the exact table provided in **Table 3**. This is called the 'saturated' model in that all of the information available was used. Every cell in the table was fully represented. We do not learn anything from the saturated model. We are trying to model expected cell frequencies to fit observed cell frequencies, but we are searching for a parsimonious way to represent the associations in the table.

The notation used to represent the saturated loglinear modeling structure is as follows:

$$\mathrm{Ln}(F_{ijk}) = \mu + \lambda_i^{\mathrm{G}} + \lambda_j^{\mathrm{E}} + \lambda_k^{\mathrm{T}} + \lambda_{ij}^{\mathrm{GE}} + \lambda_{ik}^{\mathrm{GT}} + \lambda_{jk}^{\mathrm{ET}} + \lambda_{ijk}^{\mathrm{GET}} \quad [4]$$

where $\mathrm{Ln}(F_{ijk})$ represents the log of the expected cell frequency for each cell, μ is the overall mean of the natural log of the expected frequencies, and λ's are parameter estimates representing the effects of the variables on the cell frequencies. Loglinear modeling strategies are usually hierarchical. This means that an interaction term is not included in a model without the main effect being present as well. Beneath the saturated model, there are eight models in the hierarchy when using all three main effects:

1. $\mathrm{Ln}(F_{ijk}) = \mu + \lambda_i^{\mathrm{G}} + \lambda_j^{\mathrm{E}} + \lambda_k^{\mathrm{T}} + \lambda_{ij}^{\mathrm{GE}} + \lambda_{ik}^{\mathrm{GT}} + \lambda_{jk}^{\mathrm{ET}}$ (all two way interactions),
2. $\mathrm{Ln}(F_{ijk}) = \mu + \lambda_i^{\mathrm{G}} + \lambda_j^{\mathrm{E}} + \lambda_k^{\mathrm{T}} + \lambda_{ij}^{\mathrm{GE}} + \lambda_{ik}^{\mathrm{GT}}$ (two two way interactions),
3. $\mathrm{Ln}(F_{ijk}) = \mu + \lambda_i^{\mathrm{G}} + \lambda_j^{\mathrm{E}} + \lambda_k^{\mathrm{T}} + \lambda_{ij}^{\mathrm{GE}} + \lambda_{jk}^{\mathrm{ET}}$ (two two way interactions),
4. $\mathrm{Ln}(F_{ijk}) = \mu + \lambda_i^{\mathrm{G}} + \lambda_j^{\mathrm{E}} + \lambda_k^{\mathrm{T}} + \lambda_{ik}^{\mathrm{GT}} + \lambda_{jk}^{\mathrm{ET}}$ (two two way interactions),
5. $\mathrm{Ln}(F_{ijk}) = \mu + \lambda_i^{\mathrm{G}} + \lambda_j^{\mathrm{E}} + \lambda_k^{\mathrm{T}} + \lambda_{jk}^{\mathrm{ET}}$ (one two way interaction),
6. $\mathrm{Ln}(F_{ijk}) = \mu + \lambda_i^{\mathrm{G}} + \lambda_j^{\mathrm{E}} + \lambda_k^{\mathrm{T}} + \lambda_{ik}^{\mathrm{GT}}$ (one two way interaction),
7. $\mathrm{Ln}(F_{ijk}) = \mu + \lambda_i^{\mathrm{G}} + \lambda_j^{\mathrm{E}} + \lambda_k^{\mathrm{T}} + \lambda_{ij}^{\mathrm{GE}}$ (one two way interaction),
8. $\mathrm{Ln}(F_{ijk}) = \mu + \lambda_i^{\mathrm{G}} + \lambda_j^{\mathrm{E}} + \lambda_k^{\mathrm{T}}$ (main effects only).

Each of these models depicts a distinct representation of the conditional relationships among the variables. Most statistical software packages determine the expected frequency of a model by using an iterative proportional fitting algorithm, which produces maximum likelihood estimates (MLEs) of the expected cell frequencies for a hierarchical model. All models with fewer effects than the saturated model are compared for goodness of fit between the model and the observed data. To determine the goodness of fit of the model based on the MLEs, a likelihood ratio statistic (L^2, also frequently called G^2) is calculated such that:

$$\mathrm{L}^2 = 2\sum f_{ijk} \ln\left(\frac{f_{ijk}}{F_{ijk}}\right) \quad [5]$$

where f_{ijk} are observed frequencies and F_{ijk} are expected frequencies, and the summation is over all cells in the table. However, since our aim is to try to parsimoniously estimate expected frequencies that are similar to the observed frequencies we are aiming for a small L^2. The null hypothesis is one of no difference between observed and expected, hence the desire for a low L^2 measure. The larger the L^2 the greater the discrepancy between our observed and expected frequency counts. Using the data in **Table 3**, the model with main effects alone was a poor fit, as were most of the models that had only one of the interaction effects. The best model included all three two way interactions, which is not terribly surprising. However, the model which excludes the interaction between travel mode and employment status was also an excellent fit with the data, suggesting that the most salient relationships are those related to gender. When more than one model fits the data well, we calculate the difference in the L^2 for the difference in the degrees of freedom between two nested models. If the contribution to the L^2 is significant based on the chi square distribution given the difference in degrees of freedom, it is considered a better model. After selecting the best model, rich information regarding conditional likelihoods can be calculated from the parameter estimates for each level of a categorical variable, controlling for other parameters in the model. Loglinear models provide a wealth of substantive interpretation of the association of categorical variables. They are well suited to detailed analysis of the conditional relationships among discrete qualitative data.

Logistic Regression and Logit Modeling

Whereas loglinear analysis makes no distinction between independent and dependent variables, logistic regression modeling does. Logistic regression is also known as logit modeling. Many geographers are familiar with ordinary least squares regression methods, yet such methods are only appropriate when the dependent variable is continuous. When the dependent variable is categorical, a multiple regression model is not suitable since the probability distribution of the dependent variable is not continuous. Binomial logistic regression has a dichotomous dependent variable, and multinomial logistic regression extends the approach for situations where the independent variable has more than two categories. Like loglinear analysis, logistic regression is based on probabilities, odds, and odds ratios. In the case of a binomial logit model, the odds ratio is defined as the ratio of the odds of being classified in one category of the dependent variable for two different values of the dependent variable.

The logit model is specified such that:

$$\ln\left(\frac{p_i}{1 \quad p_i}\right) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} \qquad [6]$$

where p_i is the probability that $y_i = 1$, for k explanatory variables, and $i = 1, \ldots, n$ observations. The left hand side of this equation is referred to as the logit, or simply the log odds. There is no random disturbance term in the logit model. Since we want to determine p_i, a common representation of this equation is:

$$p_i = \frac{\exp(\alpha + \beta_i x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})}{1 + \exp(\alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})} \qquad [7]$$

In a logistic regression the probability of having one outcome or another is based on a nonlinear model re sulting from the best linear combination of predictors. Predicted values are obtained by computing the prob ability of a particular case using MLE. If the probability is less than 0.5 it is categorized as 0, else it is categorized as 1. Another expression of the model provides the odds of the event or outcome such that:

$$\text{odds}(p_i) = e^{(\alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})} \qquad [8]$$

Based on the substantive example in **Table 3**, we could use a logistic regression to ask whether gender and employment status help predict the likelihood that someone drives a car or rides the bus. The mode of travel is then a dichotomous dependent variable. If we structure our interest in terms of predicting car driving, then the logit model becomes:

$$\ln\left(\frac{\text{drives}}{1 \quad \text{drives}}\right) = \alpha + \beta_1 \text{Gender}_{i1} + \beta_2 \text{Employment}_{i2} \qquad [9]$$

One of the advantages of logistic regression is flexi bility. Unlike multiple regression methods, there are no assumptions about the distribution of the predictor variables (such as normality, linearity, or equal variances), so this method particularly is useful when relationships are nonlinear. All types of independent variables are suitable, such as continuous, discrete, or dichotomous measures. Wald statistics are provided to test the sig nificance of the individual coefficients. Among early adopters there was resistance to logit models since the coefficients could not be interpreted as readily as in an OLS regression model. In the latter, a coefficient of 0.5 indicates that for every 1 unit increase of the in dependent variable the dependent variable increases by half as much. Since logit models assume a nonlinear relationship between the dependent variable and the independent variables the coefficient indicates the log odds. Since a log odds is not readily interpretable, the exponentiated value of the coefficient (e^{β}) provides the odds ratio for the independent variable, controlling for other variables in the model. It was common in early software packages that only the coefficients were listed, but now it is standard for the exponentiated values of the coefficients to be reported as well.

In a logistic regression the goodness of fit of a model can be assessed by the likelihood ratio statistic eqn [8]. This is the ratio of the maximum probability of an out come under two different hypotheses. The numerator corresponds to the maximum probability of an observed result under the null hypothesis, the denominator under an alternative hypothesis. Lower values mean that the observed result was less likely to occur under the null hypothesis, so again the desire is for low values of the likelihood ratio statistic. Many software packages also report $-2\log$ likelihood. This value is chi square dis tributed enabling a particular result to be compared to a chi square value corresponding to a desired statistical significance. A low value indicates a good fit with the observed data. A perfect model has a value near zero. A measure called Cox and Snell R^2 provides an estimate of the proportion of variability accounted for by all of the variables and is used in a similar manner to the co efficient of multiple determination in an OLS regression model.

Additional Categorical Methods and Approaches

Loglinear models and logit models (logistic regression) are common in geographic research, particularly in fields involving discrete choice and outcomes analyses. The binomial logit is without a doubt the most commonly used method of categorical data analysis. This model can also be extended to the multinomial context for situ ations in which there are more than two discrete outcomes.

A recent popular extension to a logistic regression model is a conditional logit model in which the dependent variable is still a dichotomy, but from within a given alternative of a discrete set. These models are sometimes referred to as nested logits, since the second outcome is nested within (conditional on) the first out come. These models are popular since they enable a more complex and perhaps a realistic approach to modeling decision making. Also common are multi nomial logit models. These are similar to binomial logit models but the MLE procedure applies to a polytomous dependent variable. Sometimes, this method is referred to as polytomous logistic regression. To fit a multnomial logit model, one of the categories of the outcomes is coded zero and serves as the baseline category, the ref erence group. The model estimates the outcomes for all the categories relative to the baseline simultaneously, providing more accurate measures than those provided by a series of binomial specification. The latter approach does not address the dependence among the outcomes.

Probit regression is very similar to logit. In both, the dependent variable is concerned with the probability of an outcome or event ($Y=1$). The difference lies in the assumption of the underlying distribution. The logit model assumes that the categorical dependent variable is binomial reflecting a qualitative distinction. In contrast, the probit assumes that the underlying distribution of the categorical dependent variable is truly quantitative and continuous, and uses the cumulative normal distribution. The idea is that the categorical measure is serving as a proxy for a true continuous phenomenon. So, while the logit model uses the log of the odds as the functional relationship, the probit model uses the inverse of the standard normal cumulative distribution as the functional relationship. Most texts suggest that the two methods provide very similar substantive results. However, the probit output provides probabilities, whereas the logit provides log odds. Probit models can be tailored for ordinal data and for multinomial outcomes. Probit models are not as common in geographic research as logit specifications.

The loglinear methods previously described are used to determine counts. These methods can be readily extended to consider rates. Lograte models are commonly applied to the study of rare events or to event histories. Lograte models have a similar specification to loglinear models with one fundamental difference. Instead of the cell values being a count, they are a rate. For example, if we are interested in a rare disease, the cells of a lograte model would be the count divided by the population at risk. These models are particularly applicable in medical geography, epidemiology, and demographic research.

A Promising Spatial Direction for Categorical Data Methods

During the 1980s and much of the 1990s the methods discussed above were used in a vast variety of research topics by geographers, and they have come to dominate certain fields such as migration, residential mobility, retail and transportation studies, as well as decision making models. A search of the literature indicates that recent research tends to use the more sophisticated versions of these categorical methods. However, there has been very little explicit spatial analysis until recently. This is true of much quantitative research in geography during this time period, not just categorical methods. The spatial turn began in the mid 1990s when a number of geographers developed and introduced local alternatives to global research methodologies, such as regression. A traditional standard regression model captures the nature of the relationship between two or more variables generalized across space. In other words, the nature of the relationship is assumed to be globally applicable. It is generalized across space. The limitation of these global methods is their inability to capture the spatial variations in the nature of the relationships. These new methods focus on local indicators of spatial association.

One of the most important local modeling variations has been developed and disseminated by Brundson, Charlton, and Fotheringham, called geographically weighted regression. In essence, the method extends the global regression model by fitting a local regression model for each observation using a weighting distribution that represents spatial influence and proximity. Like many quantitative methods in the social sciences, these methods were not generally applied by researchers until the geographically weighted regression software became available. Binary logistic geographically weighted regression is an extension of the logit model described above, and enables the spatial variation in the log odds to be determined. The dependent variable is still binary and the explanatory variables can be either continuous or binary. The influence of nearby data points is included in the model by choosing an appropriate bandwith and kernel function. These are best determined by theoretical ideas of how space and distance operate on the processes under study. The specification for the geographically weighted logistic regression model is as follows:

$$y_i^* = \frac{e^{(\beta_0+\beta_1 x_{1i}+\cdots+\beta_k x_{ki})}}{1+e^{(\beta_0+\beta_1 x_{1i}+\cdots+\beta_k x_{ki})}} \qquad [10]$$

where y_i^* is the predicted value of y. Expressed in terms of the logit, this is:

$$\log\left[\frac{y_i^*}{(1 \quad y_i^*)}\right] = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} \qquad [11]$$

In these models the local parameter estimates are specific to each location. There is also a geographically weighted regression procedure available for Poisson (count) regression as well. The spatially specific parameter estimates can be mapped to indicate the variation and dynamics of associations across space.

These spatially explicit versions of categorical data analysis show enormous promise for research in human geography. Initially, categorical data methods were developed to extend quantitative methods to the particular challenges of social science. Now, these spatially explicit methods for categorical data address the particular challenges of studying human geography across space. Currently, these methods are being slowly adopted, and there is every expectation that the fields of human geography that have been greatly advanced by categorical data methods will soon blossom from the influence and findings of spatial categorical data analysis.

See also: Choice Modeling; Geographically Weighted Regression; Longitudinal Methods (Cohort Analysis, Life Tables); Regression, Linear and Nonlinear.

Further Reading

Agresti, A. (1984). *Analysis of Ordinal Categorical Data*. New York: Wiley.

Agresti, A. (1996). *Wiley Series in Probability and Statistics: An Introduction to Categorical Data Analysis*. New York: Wiley.

Allison, P. D. (2001). Binary logit analysis: Basic. In *Logistic Regression Using the SAS System*, ch, 2. New York: Wiley.

Allison, P. D. (2001). Binary logit analysis: Details and options. In *Logistic Regression Using the SAS System*, ch, 3. New York: Wiley.

Charlton, M., Fotheringham, S. and Brunsdon, C. (2005). *GWR 3, Software for Geographically Weighted Regression.* Newcastle upon Tyne: Spatial Analysis Research Group, Department of Geography, University of Newcastle upon Tyne.

Cox, D. R. (1970). *The Analysis of Binary Data*. London: Methuen.

Feinberg, S. E. (1980). *The Analysis of Cross Classified Categorical Data* (2nd edn.). Cambridge, MA: MIT Press.

Goodman, L. A. (1968). The analysis of cross classified data: Independence, quasi independence, and interactions in contingency tables with or without missing entries. *Journal of the American Statistical Association* 63, 1091 1131.

Knoke, D. and Burke, P. J. (1980). *Log Linear Models*. New York: Sage Publications.

Powers, D. A. and Xie, Y. (2000). *Statistical Methods for Categorical Data Analysis*. San Diego, CA: Academic Press.

Robinson, G. M. (1998). *Methods and Techniques in Human Geography*. Chichester: Wiley.

Wrigley, N. (1979). Developments in the statistical analysis of categorical data. *Progress in Human Geography* 3, 315 355.

Wrigley, N. (1982). Quantitative methods: Developments in discrete choice modelling. *Progress in Human Geography* 6, 547 559.

Wrigley, N. (1985). *Categorical Data Analysis for Geographers and Environmental Scientists*. London: Longman.

Yamaguchi, K. (1991). Log rate models for piecewise constant rates. In *Event HistoryAnalysis*, ch. 4, pp 71 100. New York: Sage Publications.

Cellular Automata

P. M. Torrens, Arizona State University, Tempe, AZ, USA

Glossary

Agent-based modeling Computer modeling in which synthetic computational entities are used to mimic agency in simulation.

Cellular automata Computable media, bound within a distinct cellular space, used as processors in models and simulation.

Multiagent systems Collectives of agent automata used to solve problems in computer models.

Introduction

Application of cellular automata to the study of geographical phenomena dates to the late 1960s and development of land use change models by Chapin and Weiss. While famous today for the invocation of Tobler's first law, a short paper on urban growth modeling by Waldo Tobler in 1970 is among the earliest examples of geographers' use of cellular automata. This work, while pioneering, was largely ignored by geographers for several decades, until interest was revived by Helen Couclelis in the mid 1980s, presaging a flurry of activity in the early 1990s and the emergence of automata modeling as a popular avenue of research inquiry thereafter.

The use of cellular automata in geographic research is illustrative of a broader paradigm shift in the social and life sciences, away from modeling using aggregated views of space and time and toward treatment of phenomena and systems as collectives of massive amounts of individual, independent, and heterogeneous entities – each represented at their own atomic spatial and temporal scale – connected and interacting dynamically in a complex adaptive fashion. In geography, this work draws inspiration from related research in sociology, economics, ecology, political science, physics, biology, chemistry, mathematics, and computer science. However, work by geographers is starting to have a reciprocal influence in these fields, infusing spatial thinking (and GI science in particular) into the social and physical sciences.

Cellular Automata: A Primer

Automata are, fundamentally, computing media. Their origins date to Turing's work on computable numbers in the 1930s. Examples from the computer sciences include finite state machines, Turing machines, and artificial neural networks. These basic computing automata may be characterized as follows:

$$A \sim (S, R);\ S = \{S^1, S^2, \ldots, S^k\};\quad R: (S_t, I_t) \rightarrow S_{t+1}$$

where A represents an automaton, characterized by states S and a transition rule R. The transition rule functions to manage changes in states S from time t to time $t+1$, given input of other state information from outside the automaton at time t. The transition rule becomes the medium of exchange and states become the raw material when many such automata are designed to co interact. Automata of this form resemble Markov models (which deal with serial autocorrelation in temporal transition) and raster models (which deal with layering of state space).

The cellular automaton (CA) adds an additional characteristic to Turing's automaton or finite state machines – the notion that automata should be considered as being housed discretely within the confines of a cellular unit. The idea for the CA comes from Stanislaw Ulam and John von Neumann, following their work during the early foundations of digital computing in the 1950s.

Cellular automata are characterized as follows. Cells dictate the discrete (spatial) confines of the automaton. A lattice is an arrangement of neighboring cells to make up a global geographic space. Neighborhoods are localized areas of spatially related cells around a given automaton, from which it draws input in the form of their state information. The neighborhood includes the target cell itself by convention. Transition rules are the processing engines for cellular automata, and may be designed in an almost limitless fashion. In fact, cellular automata are capable of supporting universal computation. Rules may be formulated as functions, operators, mappings, expressions, or any mechanism that describes how an automaton should react to input. Time is the final component of cellular automata and it is introduced as discrete packets of change in which an automaton receives input, consults its rules and its own state information, and changes states accordingly.

Adding cells to the basic automaton structure yields the following specification.

$$\mathrm{CA} \sim (S, R, N);\ S = \{S^1, S^2, \ldots, S^k\};\quad R: (S_t, N_t) \rightarrow S_{t+1}$$

This looks complicated, but it is really quite straightforward. Above, CA refers to a cellular automaton, characterized by states S, a transition rule (or vector of rules) R, and a neighborhood N.

Various properties of geographic phenomena and systems may be mapped to these characteristics. Cells can be used to represent an almost limitless range of geographic 'things,' from cars and land parcels to eco system ranges and microorganisms. Cells can take on any geographic form and this might be regular (squares, hexagons, triangles, and so forth) or irregular. Cells relate to the spatial boundary of an entity. The boundary can convey information of direct geographic significance, such as the edges of property ownership, the footprint of a vehicle, or outer walls of a microorganism. Neighbor hoods are used to represent the spatial (and temporal) boundary of processes that influence those entities: the community within which a property resides, the journey to work for a vehicle, or the range of chemotaxis for a microorganism. States are used to ascribe attributes to cellular entities: a parcel's land use, a car's velocity, or the preferred glyconutrient for a human cell. Moreover, the states can be tied directly to phenomena that act within and/or on the cell and the larger system that it exists within. This is achieved using transition rules.

Transition rules tie all of these components together; they are the glue that binds cells, states, and neighbor hoods. Cellular automata, like all automata, are universal computers. Given enough time, resources, and the right rule set, they should be capable of supporting any computable statement. This lends terrific power to the transition rule. Cellular automata should, in theory, be able to simulate virtually anything. This stands in stark contrast to more traditional modeling methodologies available to geographers, which often limit the range of questions that can be answered with the tool. A gravity model, for example, can only support interaction ex pressed in terms of aggregate flows. Derivation of tran sition rules is key to use of cellular automata in geographic research. Transition rules are inherently tied to space–time as well as pattern and process. The rules can, therefore, be used to represent fundamental pro cesses from GI science, spatial analysis, and quantitative geography: spatial interaction, space–time diffusion, dis tance decay and distance attenuation, centrifugal action, centripetal action, and so forth.

The Advantages of Cellular Automata Modeling in the Geographic Sciences

Cellular automata have many advantages for geographic modeling. They are capable of supporting very large parameter spaces for simulation. A one dimensional CA with a binary state set and 13 cells has 2^{13} possible configurations. A two dimensional version of the same CA has 2^{169} possible configurations, and a three dimen sional CA with the same specification has 2^{2197}! If we were to divide the world's land mass into 1 km^2 rasters and run a two dimensional CA over it with 50 states (this is not unlikely, for example, if we were to develop a simple global climate model), the space of possible state switches would be $50^{153,295,000}$. This is a large number.

The representation of space and time in cellular au tomata lends an inherent spatiality to the concept. Tra ditional modeling techniques in geography abstract from spatial detail. Cellular automata, on the other hand, make an implicit use of space and spatial complexity. Cells, neighborhoods, and lattices are inherently spatial. In addition, cellular automata are capable of supporting separate notions of space and time, as well as combined space–time relationships. The basic cellular concept also has a natural affinity with raster based data structures common to GI systems and image processing in remote sensing. The stratification of state variables in cellular automata is also synonymous with raster layering in GI systems. Similarly, the serial treatment of temporal re lationships allows for the introduction of a formal hier archy to dynamics that complements Markov like processes already popular in geographic modeling. Moreover, cellular automata are capable of representing form and function, pattern and process.

Cellular Automata and Complexity

Cellular automata have further advantages stemming from their relationship with complex adaptive systems. Complexity studies focus on the grassroots of system dynamics, emphasizing the local interaction among elements that give rise to phenomena at synoptic scales. In modeling, the complexity approach also focuses on issues such as the importance of historical (seed) con ditions, feedback between subsystems, interaction, dy namics, phase transition, noise and perturbations, and so forth. Cellular automata are among the best and most widely used tools for complexity modeling.

Human Geography Applications

Cellular automata have been employed in the study of a wide range of geographic phenomena. The plethora of applications is illustrative of the usefulness of the ap proach. Human geography applications are largely fo cused on issues relating to urban geography and behavioral geography.

Models of urbanization and urban growth feature prominently in the literature. Cells are generally used to represent areas of urbanization or land parcels, with cell boundaries matched to pixels in remotely sensed images. States can be relatively simple, with binaries representing urbanization, for example. State specification can, how ever, be much richer, formulated as layers of land use or activity (synonymous with attribute layering in a GI system), or as fuzzy membership. A variety of transition rules have been employed as representative of the forces of urbanization. These include urbanization based on agglomeration through diffusion limited aggregation; land use transition based on potential for development; physics driven mechanics; data driven rules from statistical analysis; data trained rule sets from artificial neural networks; and spatial heuristics. In some cases, different rule sets are employed as what if experiments to test the varying influence of urbanization processes. In this sense, cellular automata are used as artificial laboratories to test theory, as tools to think with.

Cellular automata have also been used to model other aspects of human geography. Work by Schelling and Sakoda with simple chessboard models is among the early antecedents of cellular automata simulations of sociospatial segregation. Cells are equipped with states that correspond to ethnicities, and rules based on preferences for co location in space are used to show how small biases can quickly lead to large scale segregation from random initial conditions, and to investigate the tipping point at which sociospatial segregation begins. More recent models of residential mobility follow similar schemes, with rules designed to introduce more elaborate mechanisms based on spatial choice and dissonance.

David O'Sullivan has also developed a model of gentrification dynamics based on cellular automata. Cells are used to represent gentrifiable real estate and rules are introduced to test rent gap hypotheses from the theoretical literature.

Automata based work on traffic simulation is largely agent based, but several applications have been developed using cellular automata. Cells are used to represent vehicles and pedestrians in these instances, with movement formulated by proxy using transition rules that pass the presence of these entities as state information between cells based on heuristics designed to mimic lane changing, collision avoidance, stopping, queuing, and junction navigation.

Physical Geography Applications

Cellular automata models are also popular in physical geography research. Global climate models (GCMs) and general circulation models in climate and meteorology studies function, at a global level, as cellular automata. At a macrolevel, GCMs are divided into large cellular grids composed of cells of around one degree of latitude and longitude in size. Various models simulate the internal climatic dynamics within these cells, but the results are generally exchanged through the larger grid on simple cellular automata schemes. Cellular automata are very efficient tools for representing processes of geomorphological transport and feature as the geographic engine in models of particle movement. Diffusion based rules have been used for these purposes in models of debris flow. Cellular automata are used in a similar fashion, as the mechanism of mobilization in models of wind transport for snow movement and dune formation, for example. Transition rules are naturally accommodating of calculus based equations in these applications.

Cellular automata are crucial to the design and application of fire models in physical geography work. Some of the most advanced computational work with cellular automata has been realized in development of wildfire simulations.

Geographic Applications in Other Sciences

Cellular automata have also been used in a variety of social science applications outside the geographic sciences, largely due to their value in representing geography in simulation. Cellular automata with fundamental space–time representations have been used in anthropology, to model the formation of societies and in political science and sociology to explore civil violence. Cellular automata have been particularly useful in infusing geography into work in economics. Models developed by urban economists, for example, rely heavily on cellular automata to represent processes of urban agglomeration. Cellular automata models that are fundamentally geographic in inspiration have also been widely used in ecology to simulate floral and faunal dynamics and in biology, where they are employed in the modeling of cellular dynamics and tumor formation.

The Future for Cellular Automata in Geographic Research

Research work in this area is focused on a broad range of interests. Development of new applications is still quite active, with new cellular automata models popping up for a variety of phenomena of geographic interest, with an associated growth in their use as test beds for theory, practice, and policy.

Researchers have also begun to focus on extending the basic idea of cellular automata from mathematics and computer science for geographic applications. Work in specifying cells, lattices, and neighborhoods through GI systems is particularly active, with recent advances in the use of graphs, Voronoi polygons, irregular cells, and 2.5 D lattices.

Cellular automata are naturally allied with remote sensing, GI systems, and the dataware associated with them. Various attempts have been made to develop cellular automata engines within GI systems and to build GI systems functionality into cellular automata. Recent work, however, has focused on the mutual links between cellular automata and GI science.

Research into the connections between cellular au tomata and agent automata is central to the current re search agenda. The two are popularly confused, even though the distinction on geographic grounds is reason ably straightforward: cells do not actually move within their lattices and engage in action by proximity, while agents can move freely in vector spaces and can engage in action at a distance. Tool kits and methodologies that support both approaches are just beginning to be built.

Issues surrounding calibration and validation of cel lular automata models are chief among challenges facing future research in this area. Existing work has an over arching focus on patterns and their role in benchmarking cellular automata models, but research into the role of processes in validation and calibration is comparatively underdeveloped.

See also: Artificial Intelligence and Expert Systems; Behavioral Geography; Chaos and Complexity; Geocomputation; GIScience and Systems; Quantitative Methodologies; Simulation; Spatial Science.

Further Reading

Batty, M., Couclelis, H. and Eichen, M. (1997). Special issue: Urban systems as cellular automata. *Environment and Planning B* 24(2).
Benenson, I. and Torrens, P. M. (2004). *Geosimulation: Automata Based Modeling of Urban Phenomena*. London: Wiley.
Clarke, C. and Gaydos, L. (1998). Loose coupling a cellular automaton model and GIS: Long term growth prediction for San Francisco and Washington/Baltimore. *International Journal of Geographical Information Science* 12(7), 699 714.
Couclelis, H. (1985). Cellular worlds: A framework for modeling micro macro dynamics. *Environment and Planning A* 17, 585 596.
D'Ambrosio, D., Di Gregorio, S. and Iovine, G. (2003). Simulating debris flows through a hexagonal Cellular Automata model: Sciddica S3 hex. *Natural Hazards and Earth System Sciences* 3, 545 559.
Deutsch, A. and Dormann, S. (2004). *Cellular Automaton Modeling of Biological Pattern Formation*. Boston: Birkhauser.
Guy, E., White, R., Uljee, I. and Drazan, P. (1995). Using cellular automata for integrated modelling of socio environmental systems. *Environmental Monitoring and Assessment* 30, 203 214.
Kronholm, K. and Birkeland, K. W. (2005). Integrating spatial patterns into a snow avalanche cellular automata model. *Geophysical Research Letters* 32(19).
Stuart, C. F. and Weiss, S. F. (1962). *Factors Influencing Land Development. An Urban Studies Research Monograph*. Chapel Hill: Center for Urban and Regional Studies, Institute for Research in Social Science, University of North Carolina.
Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46(2), 234 240.
Torrens, P. M. and Benenson, I. (2005). Geographic automata systems. *International Journal of Geographical Information Science* 19(4), 385 412.
Torrens, P. M. and O'Sullivan, D. (2001). Cellular automata and urban simulation: Where do we go from here? *Environment and Planning B* 28(2), 163 168.
Vandewalle, N. and Galam, S. (1999). Ripples versus giant dunes in a saltation avalanche model. *International Journal of Modern Physics C* 10(6), 1071 1076.
Wenzhong, S. and Cheung Pang, M. Y. (2000). Development of Voronoi based cellular automata an integrated dynamic model for geographical information systems. *International Journal of Geographical Information Science* 14(5), 455 474.
Wolfram, S. (1994). *Cellular Automata and Complexity*. Reading, MA: Addison Wesley.

Relevant Websites

http://theobio.mtbio.de
Andreas Deutsch's Computational Biology group at Technische Universitat Dresden.
http://ccl.northwestern.edu
Center for Connected Learning and Computer Based Modeling at Northwestern University.
http://eslab.tau.ac.il
Environmental Simulation Laboratory at Tel Aviv University.
http://www.geocomputation.org
Geocomputation Site.
http://www.geosimulation.org
Geosimulation.
http://www.casa.ucl.ac.uk
Michael Batty's group at University College London Centre for Advanced Spatial Analysis.
http://www.ncgia.ucsb.edu
The SLEUTH project at University of California, Santa Barbara.
http://transims.tsasa.lanl.gov
TRANSIMS project at Los Alamos National Laboratories.
http://www.wolfram.com
Wolfram Research.

Census Geography

W. Friesen, The University of Auckland, Auckland, New Zealand

Glossary

Census Error Errors which affect reliability of census data, often following predictable patterns.

Census Moment The moment at which the census is said to have taken place, even though enumeration probably takes place before and after this time.

***De Facto* Enumeration** The enumeration of people according to their location at census moment.

***De Jure* Enumeration** The enumeration of people according to their usual residence.

Geodemographics Analysis of demographic data using relatively small spatial units.

Politics of Participation Political aspects of participating in a census, including issues of representation, identity, and confidentiality.

Postenumeration Survey Sample survey usually taken soon after a census, to estimate the degree of undercount and its characteristics.

Undercount The tendency for a census to miss a proportion of the population.

Introduction

This article adopts a broader definition of census geography than is sometimes used. The more narrow definition, often used by statistical agencies, focuses on the spatiality of census taking and analysis, whereas this article also considers issues of census implementation and the techniques used by geographers and others in census analysis as well as the applications to which census data are put. It also moves beyond the descriptive to assess some political contexts of the census process.

A census is usually differentiated from a social or population survey by the fact that a census should enumerate (or attempt to enumerate) the total population of an area. Most population censuses are undertaken with a housing or household component as well. Since a census is normally taken at a single point in time, census data provide a cross sectional view of a population, in contrast to a longitudinal approach which collects data on individuals or households through time. There is some debate as to whether some of the mixed forms of registration and enumeration, which have replaced a full census in some countries, should be described as a census.

Census Implementation

History of Census Taking

It is not known when the first census was taken, but there is evidence of census taking nearly 6000 years ago, with occurrences recorded in Egypt, China, Babylonia, Palestine, and Rome. Early censuses were undertaken especially for the purpose of military conscription, but were also used for purposes of taxation and labor requirements of the state. Thus, in most cases, enumeration only involved adult male citizens.

A modern census is one which meets standards of completeness, accuracy, and simultaneity (discussed further below). These censuses also have broader purposes than the earlier censuses, and these include government planning for provision of services and often the determination of electoral boundaries. The first modern census is sometimes attributed to Sweden in 1749, but there were earlier censuses with modern attributes undertaken in Nouvelle France (Quebec), Acadia (Nova Scotia), and some of the Italian principalities. In the late eighteenth and through the nineteenth century, many other countries initiated full censuses: United States 1790, Spain 1798, England 1801, France 1801, Ireland 1811, Norway 1815, Australia 1828, Greece 1836, New Zealand 1851, Switzerland 1860, Italy 1861, and Canada 1871.

Through the twentieth century some colonial powers initiated censuses in their territories, although in some cases full censuses were not undertaken until independence. By the end of the century, national censuses had become the norm for nation states. In the 2000 round of census taking (1995–2004), about 200 nations undertook at least one census. Those that did not undertake a census tended to be nations with serious political problems such as civil war.

A new trend in recent years should be noted. Sweden has not undertaken a census after 1990, and in 2005 initiated a registration based census in which there was no comprehensive enumeration, but rather the analysis of other records of government combined with sample surveys. Denmark and Finland have population registers and these are combined with data from other administrative records. In 2004, France introduced a system in which data are gathered in successive annual collections rotating through municipalities every 5 years; these have replaced the comprehensive census. The American Community Survey (ACS) is administered annually to a sample of areas throughout the United States, with no household being sampled more than once in 5 years.

Although the 2010 census will still take place, the data collected in the ACS will replace the long questionnaire used in previous censuses. Such approaches are likely to be adopted in other countries where registration systems are comprehensive, but may not be possible in many other countries with less comprehensive systems and where certain privacy regulations are in place.

Objectives and Design of a Modern Census

In many countries, there is legislation which regulates the implementation and analysis of the national census. A primary legislative function of many censuses is to allow equitable construction of electoral boundaries, and these are often redrawn after each census. Most censuses also have objectives related to planning for the provision of services such as education and health, the development of a skilled labor force, and a range of other economic and social purposes.

In order to fulfill these objectives, most modern censuses are designed to be:

- universal: the total population should be enumerated;
- compulsory: all residents are obliged to take part in census enumeration;
- comprehensive: ideally a large number of variables should be collected, although in reality some censuses collect samples for some variables (e.g., US census in 2000 obtained full enumeration for six basic variables but a 15% sample for a larger number of variables); and
- simultaneous: the census should take place within a short period of time for the whole population to allow comparability; there is often a precise census moment (e.g., midnight on a particular date).

The periodicity (frequency) of censuses varies considerably between countries. The US and UK undertake censuses every 10 years (in years ending in 0 and 1, respectively), and these are an artifact of the timing of the first censuses two centuries earlier. The French census has been carried out every 7 to 9 years, sometimes, but not always, approximating the presidential electoral cycle. Censuses in many ex British territories (Canada, Australia, and New Zealand) have followed a 5 year cycle, although the electoral cycle in those countries is shorter than this. In the attempt to achieve universal and synchronous enumeration, the timing of a census is important. It should avoid periods of extreme weather (although Finland uses New Year's Eve), significant cultural and religious events, and periods of high levels of resident mobility.

Enumeration coverage (who should be enumerated) is also an issue. A *de facto* enumeration (e.g., UK) records people according to the place they are at census moment, while a *de jure* enumeration (e.g., US) records the usual residence of a person. The problem with the *de facto* approach is that it may incorporate many people who are not usually residents (e.g., tourists) and so may be less relevant to the planning and other purposes of the census. The problem with the *de jure* approach is that while it may reassign enumerated residents to their usual place of residence, this is not possible for usual residents who are outside of the country at the time of enumeration (unless they are intercepted at the border on their return). In some countries, both *de facto* and *de jure* data are collected and analyzed.

To encourage international comparability, the United Nations has proposed a standardized range of variables that should be collected in a census, including age, sex, birthplace, citizenship, language, education, economic status, family structure, fertility, and a number of others. In practice, however, a great range of different information is collected by different countries. Although most censuses aim to be as comprehensive as possible, the number and nature of variables collected depends on financial and human resources available for enumeration and analysis, as well as political, social, and cultural considerations. Also, in countries with large sectors of the population with limited levels of education, it is necessary to create questions which are readily comprehensible to both enumerated and enumerator. Political considerations include the desirability of convincing the population that questions are not designed to allow agencies of government to use census information for purposes of cross checking on aspects such as income, welfare benefit payment, and military service. Social and cultural considerations include sensitivities about questions perceived as private such as questions relating to personal behavior or characteristics such as fertility, smoking, disability, or religion. In some censuses, questions such as these are optional.

Role of Government and Politics of Participation

Even in those countries which have implemented neoliberal reforms in which some functions of government have been privatized, the central statistical agencies which implement censuses have remained under government control (although often with a legislated degree of autonomy). The role of government is critical to the implementation of a compulsory census, but this role also creates problems in relation to participation.

In the past, it was mainly small religious groups which objected to being enumerated in line with their aversion to participation in most aspects of government. However, the politics of participation has become more complex in recent times with objections to the census process also originating from political movements such as libertarianism and anarchism. Also, as marketers, market researchers, credit rating agencies, and spammers of all varieties have become increasingly intrusive using new technologies, census enumeration may be classed by some along with these other intrusions. On a general level, many governments have responded to these intrusions by strengthening privacy legislation, and more specifically by increasing the provisions for confidentiality within the census collection and dissemination processes. Most recent censuses have been preceded by publicity campaigns emphasizing the public good aspects of a census and the strict provisions for confidentiality. Nevertheless, in even the most successful of censuses there are individuals within the population who will be missed (and occasionally counted twice), and other errors related to individual and household characteristics will enter into the census process. Thus, it is necessary to consider the process of census enumeration and analysis in more detail.

Census Enumeration and Analysis

Although enumeration and analysis are generally considered as separate stages of the census process, technological changes have resulted in the boundaries between these processes becoming blurred since many of the same issues are found in both. The basic units of enumeration of most population censuses are the individual and the household. These are also the most common units of census analysis, but analysis may also be undertaken by aggregation within and beyond households to use other units including family and community.

In the past, spatial units of enumeration and analysis have largely been constructed to facilitate effective enumeration and to serve legislative and administrative purposes at analysis. Increasingly, with the use of GIS, small spatial units are seen as building blocks to devise aggregated spatial units custom made for almost any purpose. Of course, suburbs, wards, cities, districts, counties, regions, states, provinces, and other administrative and political units are still important in the process of census data analysis and dissemination.

Methods of Enumeration

A great deal of effort is put into the design and testing of questionnaires before enumeration begins, and the success of census enumeration and analysis depends on the quality of questionnaire design. Well designed census questions include those which are succinct, not open ended, and do not ask leading questions.

Although census enumeration in theory takes place at a specific point in time (census moment), in reality, enumeration takes place over days or weeks in order to gather information about the particular census moment. The two most common methods of enumeration in a comprehensive census are enumerator based distribution of questionnaires (using either direct interviewer or canvasser method, or the self enumeration or householder method) and postal distribution of questionnaires (which inevitably involves self enumeration but might include telephone follow up). In the United States, distribution and collection is by post while in Spain, Italy, Australia, and New Zealand both are done by enumerators, while in Canada and the United Kingdom distribution of forms is done by enumerator and return is by post. In many less developed countries there is much more reliance on direct interviewer methods. In the twenty first century, a number of census agencies have initiated online enumeration, and while relatively small proportions of these populations have been enumerated in this way, this form of enumeration is likely to increase rapidly in importance.

Population Coverage and Undercount

Although the ideal of census agencies is a 100% enumeration, most censuses fall short of that, and the rate of response is partly dependent on whether a census is compulsory or not. Although participation in the census in most countries is in theory compulsory, in reality it is not always possible to enforce. In the United States census of 2000, which relied more on promotion than compulsion, the final response rate was 67%, slightly higher than the 65% achieved ten years earlier. Even in countries where compulsory participation is more rigorously enforced, some individuals will be missed, and a few mobile individuals may be counted more than once, so that the net effect is a population undercount.

A post enumeration survey (PES) is the most common way of estimating the accuracy of census coverage not only in terms of population undercount, but also in relation to particular characteristics of the population. A PES is based on intensive sampling of particular populations and areas usually soon after the census itself to estimate the proportion of population missed by census enumeration. Target groups may include the homeless, transients, and people absent from their usual residence at census moment. Target areas may include areas with relatively lower educational levels, areas with higher numbers of undocumented migrants, and remote areas.

There is a significant politics surrounding population coverage and undercount. Since undercounts are more likely in areas of lower socioeconomic status, if these are not rectified by a PES or similar instrument, there may be a number of disadvantages for such areas. Since most electoral systems are based on proportional representation, such an area will be underrepresented in the political system. Further, in countries such as the United States where public funding is allocated according to population, health, education, and other social funding from federal and state sources will be reduced if the undercount is not considered.

Sources of Census Error

Data errors may occur when a respondent has provided incorrect or ambiguous information or when data are being coded and entered. Respondent errors fall into several categories:

- accidental inaccuracies, for example, visitors over looked, babies forgotten, and questions misunder stood;
- intentional misrepresentations, for example, hiding of undocumented migrants and misreporting of income; and
- variables with inherent ambiguity, for example, ethnicity (see below).

Errors related to coding and entering of data may be simply technical mistakes, but they also may relate to problems with classification of particular variables, and examples discussed below include classification of households and families, and classification of ethnicity and religion.

Geographies and Politics of Households and Family

Although households have served as one of the two basic units of census enumeration, they represent a complex series of debates within the social science literature. In *Progress in Human Geography* there has been an ongoing debate about the pivotal role of the household in shaping the geographies of gender, home, and everyday life. For example, studies of urban gentrification and its relationship to population dynamics have been enriched by the investigation of household level change. This debate incorporates the so called second demographic revolution which has been hypothesized around the deinstitutionalization of the nuclear (heterosexual) fam ily, and the increasing importance of a variety of other family types. Outcomes of this include more single parent and one person households, declining average household size, and a greater variety of individual and family transitions between different household structures.

Thus ways in which enumeration and classification of households and families are carried out in a census in fluences the nature of the social research which is pos sible. In many censuses, a household is synonymous with a dwelling, although more than one household may exist within a single dwelling. Also significant is the fact that some households may exist between more than one dwelling, even between countries as the concept of transnational households proposes. There are even greater difficulties in attempting to capture family rela tionships within a census. Censuses often attempt this through identifying the relationship of each individual in a household to a primary individual. In some cases there is an identifiable 'head', in the past the male parent of a nuclear family, but with the rise of feminist perspectives and the decline in the centrality of the nuclear family, many censuses have abandoned the use of this concept for some other proxy. The US census refers to this pri mary individual as person 1, but this does not always allow the reconstruction of family types within a household, so many households are classified as being comprised of 'mixed individuals' or something equivalent.

Politics of Classification: Ethnicity and Religion

Two of the most contentious of census variables are race/ ethnicity and religion. On one level, this relates to the perception that these are private and sensitive issues, and many censuses present a question on religion as optional. More fundamentally, especially in the case of ethnicity, the politics of classification are significant, since 'eth nicity' is a term incorporating aspects of language, heri tage, birthplace, nationality, cultural ritual and symbols, and religion.

The purposes of census questions on race or ethnicity are various. In colonial societies, the identification and classification of race was justified on the grounds of sci ence, welfare assessment, service provision, and in the case of South Africa, a mechanism to administer apart heid. Many modern censuses collect data on race/eth nicity, usually for purposes of services and related planning, but in some cases also for legislative reasons (e.g., the allocation of Maori electoral seats in New Zealand).

Earlier censuses presented classifications of race in a biologically deterministic way. Questions often asked what proportion of a person's 'blood' was derived from a particular 'race' with results sometimes presented in fractions. Most recent censuses have used self declared 'ethnicity' or 'ethnic background'. Nevertheless, during the enumeration and analysis stages, the classification of these responses is fraught with difficulties. During enu meration, a set number of categories may be presented, purportedly representing the predominant ethnicities of a country, usually with the option to specify one or more 'other' categories. Listed categories tend to be chosen over those not listed, and the self identified 'other' tends to produce some responses which are spurious. There are differences of opinion as to what constitutes an ethnicity. Those who analyze ethnic data maintain that Canadian or Australian are nationalities rather than ethnicities since they do not allow any differentiation between in digenous groups nor between various linguistic and cultural groups, but there are many who prefer those identities because of long term residence in the country in question or because of the complexities of multiple ethnic origins.

Data and Geography Encoding

The historical development of the post enumeration processes of a census may be conceived in stages as shown in **Table 1**. Although the US census used some computer processing in 1950 (UNIVAC 1), most census data before the 1960s were manually encoded. Census agencies tended to adopt new computer technologies soon after they became available, and in some cases were significant contributors to new technological develop ments. The first significant use of digital technology was for data encoding, and this was followed by the devel opment of spatial digitization for geography encoding. However, it is only recently that digital methods have been used for the actual geography design of censuses. Whereas the geographical census output was driven by the imperatives of the data collection process, georefer encing of census returns has allowed the output geog raphy to be created independent of the collection geography as illustrated in **Figure 1**.

Applications

A vast range of demographic, socioeconomic, and related analyses are based on census data, and those mentioned here are only indicative. In the past, census data were accessed through published volumes, but currently digital outputs predominate, ranging from online tabulations to customized data runs and the use of data labs. Data labs allow analysis down to the level of individual records, usually involving strict conditions of confidentiality. Two examples of these are the PUMS (Public Use Microdata Samples) in the US and the SARs (Samples of Anon ymised Records) in the UK.

Table 1 A four-stage model of census geography development

Stage	*Approximate date (in UK)*	*Data encoding*	*Geography encoding*	*Geography design*
1	Pre-1960	Manual	Manual	Manual
2	1960s	Digital	Manual	Manual
3	1980s	Digital	Digital	Manual
4	2000s	Digital	Digital	Digital

Source: Martin, D. (1998). Optimizing census geography: The separation of collection and output geographies. *International Journal of Geographical Information Science* 12, 673 685.

Demographic and Socioeconomic Analyses

The mainstays of demographic analysis are fertility and mortality. In countries which have birth registration systems, fertility rates are usually calculated from these data combined with census data on the cohorts of women of reproductive age. Birth registration systems usually collect a limited range of information related to the characteristics of the mother such as place of residence, age, and ethnicity but censuses which collect fertility information such as age at first birth, ever born children, etc. allow analysis using a much greater range of variables than is possible with the birth registration data. Also, in many less developed countries, vital registration systems are absent or unreliable, so census data are critical to the analysis of fertility. Likewise, basic rates of mortality can be derived from death registration data, but the popu lation on which these rates are based is almost always derived from a census. Other important demographic processes and characteristics which are usually derived from census data include population size and change, population distribution and movement, ethnic com position, age–sex structure, and household and family structure.

Much socioeconomic analysis is also derived from census data using variables such as income sources and levels, educational levels, occupation, industrial sector of employment, and so on. Further, multivariate indicators are constructed from these basic variables. For example, much effort has gone into the development of indicators of deprivation or well being such as the work on social indicators by the Canadian Council on Social Develop ment, which uses a range of census and non census sources.

Migration Analysis

While the statistical analysis of international migration relies on some data external to the census (arrival and departure data, residence permit data, etc.), census an alysis is important in analyzing immigration outcomes. Questions on birthplace and duration of residence within a country allow the identification of the migrant popu lation, and cross tabulations with various social and economic variables allow an assessment of levels of mi grant participation and well being. A limitation of the census method, however, is that longitudinal analysis of individuals or families is not possible except in some cases by inference between censuses.

The analysis of internal migration is highly dependent on census data in countries which do not have com prehensive systems of residence registration. A question on usual residence at a previous point in time (e.g., 1 or 5 years ago) allows the construction of matrices of move ment between different areas of a country. The degree of mobility measured by this method is dependent on the

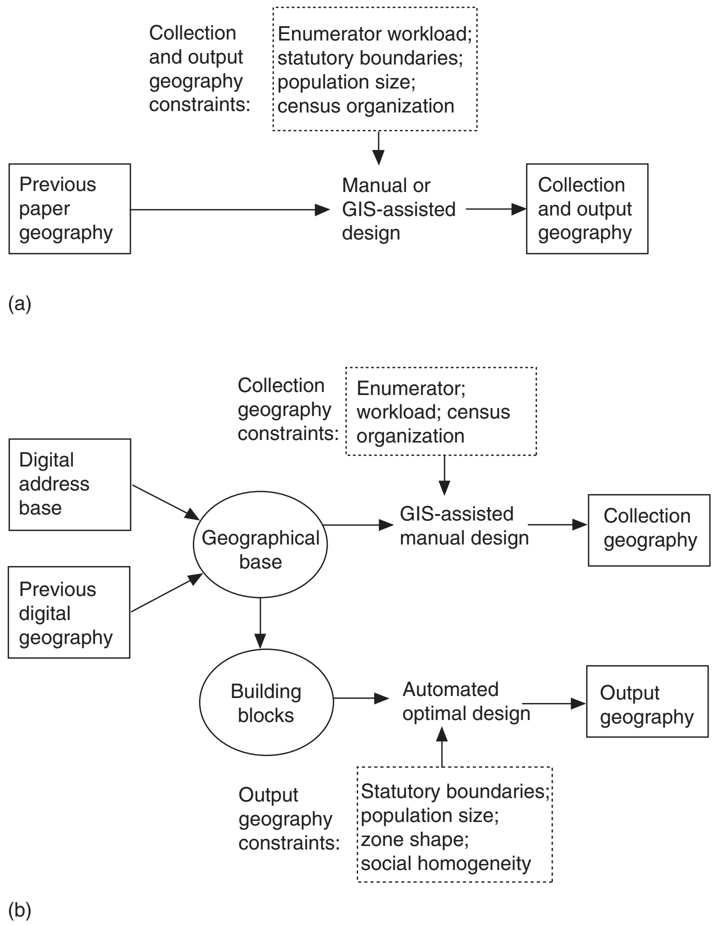

Figure 1 Census geography design procedures at (a) stage three, and (b) stage four. Source: Martin, D. (1998). Optimizing census geography: The separation of collection and output geographies. *International Journal of Geographical Information Science* 12, 673 685.

spatial resolution used, with the most detailed scale (any household movement) resulting in the highest levels of measured mobility. A common use of this method is to assess economic well being of different regions of a country in relation to population mobility, especially in relation to the educational and occupational character istics of the migrant population.

The residual method of migration analysis may be used in the absence of census data on movement into or out of an area. Age–sex characteristics of the population of an area are compared between censuses so that a survival rate is applied to age–sex cohort between census 1 and census 2 and the difference between the hy pothesized number of survivors at census 2 is compared to the actual cohort at census 2; the differences are taken to represent the net migration of each cohort between censuses.

Geodemographics

Although the term 'geodemographics' was first used by geographers more than a century ago to classify neighborhoods within cities, its contemporary usage can be attributed to its use within business demography to refer to the analysis of demographic data, usually using relatively small spatial units, for the purposes of market analysis and related purposes. The term and approach have more recently been used in other contexts, especially for social research and the provision of public services. One of the central approaches of geodemographics is the clustering of statistically similar neighborhoods or other areas. Census data are usually central to these approaches since geodemographics demands information at a detailed spatial scale and often involves a number of variables. Census data are then often supplemented by data from other sources such as market surveys, records of housing sales, electoral rolls, and so on.

Population Projections

Planning at national, regional, and local levels is usually dependent on the projection of future populations. Critical to projection methodologies is robust data to construct base populations and to develop assumptions.

Since undercount is likely in nearly all censuses, pro jection base populations are often constructed by in corporating the undercounted elements of the population, so these may be seen to be better represen tations of the usual population of a country than the census counts themselves. Fertility assumptions are usu ally derived from birth registration statistics but usually use census populations as their base, and similarly life tables used in projections are derived from both death registration data and census data.

International and internal migration assumptions for population projection may be the most problematic since there may be large and unexpected fluctuations in mi gration resulting from government immigration policy changes, economic cycles, or dramatic political or natural events. Nevertheless, census data on past migration trends may be useful in modeling possible and/or likely migration patterns over time.

Issues

The changing nature of census geography is being driven by several forces. The most obvious of these is techno logical change. The rapid acceleration of computer power has revolutionized all stages of the census process, ranging from enumeration to data entry to analysis to dissemination of results. Even more fundamentally, this acceleration and the development of GIS have altered the potential for spatial conceptualizations that were not possible earlier. Another force, partly related to techno logical change, is the ability of governments to gather a vast range of information on their citizens which has reduced the need for the comprehensive census taken at a fixed point in time. Thus some countries have abandoned the traditional census approach and it is likely that many more will follow. However, potentially at odds with this development is another significant force, namely the emergence of identity politics in many places which challenges the whole classificatory process that censuses or population registration systems represent. Parallel to this are concerns about the confidentiality of data col lected by governments and by private agencies and on going monitoring via registration systems may appear to be more intrusive than an occasional census was.

See also: Census Mapping; Demography; Electoral Geography; Geodemographics; Georeferencing, Geocoding; Population Geography.

Further Reading

Buzar, S., Ogden, P. E. and Hall, R. (2005). Households matter: The quiet demography of urban transformation. *Progess in Human Geography* 29, 413 436.

Christopher, A. J. (2005). Race and the census in the Commonwealth. *Population, Space and Place* 11, 103 118.

Howard, D. and Hopkins, P. E. (eds.) (2005). Editorial: Race, religion and the census. *Population, Space and Place* 11, 69 74.

Hugo, G. (2006). Population geography. *Progess in Human Geography* 30, 513 523.

Martin, D. (1998). Optimizing census geography: The separation of collection and output geographies. *International Journal of Geographical Information Science* 12, 673 685.

O'Sullivan, D. (2006). Geographical information science: Critical GIS. *Progress in Human Geography* 30, 783 791.

Siegel, J. S. and Swanson, D. A. (eds.) (2004). *The Methods and Materials of Demography*. San Diego: Elsevier Academic Press.

Subramanian, S. V., Duncan, C. and Jones, K. (2001). Multilevel perspectives on modeling census data. *Environment and Planning A* 33, 399 417.

Wilson, T. and Rees, P. (2005). Recent developments in population projection methodology: A review. *Population, Space and Place* 11, 337 360.

Relevant Websites

www.ccsd.ca
The Social Indicators Site of the Canadian Council on Social Development.

www.statistics.gov.uk
UK census homepage.

www.census.gov
US Census Bureau home page (contains census data but also links to other government statistical agencies and papers on census methodology, etc.).

Census Mapping

D. Martin, University of Southampton, Southampton, UK

Glossary

Census Complete enumeration of a population for statistical purposes.

Choropleth Map constructed from shaded areas in which shading represents the value of the attribute being mapped.

Ecological Fallacy The fallacy of making inferences about individuals from aggregated data.

Enumeration Method for collecting and recording characteristics on each member of the population, conducted by census officer called an 'enumerator'.

Geocoding Conversion of addresses, place names, or other geographic references to coordinates, such as latitude and longitude.

Geodemographic Classification Classification of areas based on their socioeconomic characteristics so as to place the most similar areas into the same groups.

Gerrymandering Deliberate manipulation of geographical boundaries so as to maximize the electoral advantage of a particular party.

Modifiable Areal Unit Problem The dependence of analysis of aggregate data on the precise reporting zones used for aggregation.

Introduction

Census mapping refers to the cartographic presentation of the results of censuses in order to reveal spatial patterns in the data. Despite difficulties in achieving complete enumeration, the high population coverage of censuses means that it is usually possible to obtain data for very small geographical areas which can form the basis for detailed maps. Censuses are usually administered by means of a questionnaire for each member of the population covering a range of demographic and socioeconomic topics. Postenumeration processing of these forms permits the production not only of simple counts and percentages such as numbers of households without a car or percentage persons unemployed, but also a wide range of cross tabulations and derived indicators such as social class, geodemographic classifications, and multivariate deprivation indicators. This article considers the overall role of censuses, the uses of census mapping, and a range of census mapping techniques.

The Place of Censuses

Censuses of population play an important part in the government of nations, particularly with regard to the spatial allocation of resources. The earliest modern censuses, from which results have been widely published, can be dated from around the start of the nineteenth century, with the first census of population in 1790 in the US and 1801 in England and Wales. Censuses represent one of a range of methods for counting population which also includes continuous administrative and registration based systems, particularly favored in Scandinavia, and indirect methods such as estimation from satellite remote sensing. A census is an attempt, typically decennial, to measure the location and characteristics of the entire population at a single point in time, in contrast to continuous approaches which rely on tracking population members, usually for administrative rather than statistical purposes. These approaches all contrast with sample survey methods which do not attempt to directly record the entire population. A successful census offers the single most powerful combination of population coverage, geographical and socioeconomic detail within a single integrated dataset, and can support a variety of data outputs, including, for example, microdata samples and interaction data which describe migration and commuting flows. The most traditional output has been aggregate responses for geographical areas, which form the basis for most census mapping.

Historically, there have been multiple motivations for conducting a census. An understanding of population resources may be required for purposes as diverse as raising an army, organizing political representation, levying taxation, or allocating welfare funding. The results of a census may be particularly politically sensitive when, for example, they demonstrate the relative strength of different population subgroups, such as the proportions and degree of mixing between Protestant and Catholic groups in Northern Ireland or the various ethnic groups in South Africa. Where there are conflicts within a population, subgroups may actively boycott census enumeration in protest about government policy, challenge the results of a census as not adequately representing their interests or lobby for the inclusion of specific questions or definitions. Some population subgroups such as unregistered migrants may actively seek to avoid detection by government, including a census. A longstanding criticism of censuses and their analysis is that they include only questions which are acceptable to the prevailing government and are therefore funda mentally limited in their ability to challenge established power relations in a society.

Modern censuses are generally conducted under specific statistical legislation and incorporate sophisti cated data protection techniques in an attempt to provide respondents with privacy assurances. These include the random swapping of individuals between areas, rounding, or modification of small counts and suppression of results for areas with small populations. The extent to which the census technical and legislative framework provides effective protection to the individual, rather than simply contributing to surveillance by the state, is a strongly contested issue and varies between national contexts. Any data adjustment process will have an impact on census mapping. In general, statistical disclosure control pro cedures have the biggest impact on data at the neigh borhood or village level, but are negligible at the scale of cities and regions.

Nevertheless, since the advent of computer based census operations the user base for census data has in creased dramatically. Computer readable census outputs are widely used by central and local government, healthcare planners, academic researchers, and the commercial sector for demographic analysis and business or service planning. Census mapping, in the 1970s the preserve of university computer laboratories, is now readily undertaken online by schoolchildren. There is thus a strong tension between increasing the content and outputs of censuses to meet contemporary demands versus a desire to maintain comparability over time and preserve response rates. Internationally, census response rates are falling due to both public concerns over the privacy of personal information and increasing difficulty in accessing and delivering census forms to every household. In the early 2000s nations are adopting dif ferent strategies to provide the statistical information required by government – either through increasing re liance on administrative data sources or by spreading census enumeration over time. France has adopted a rolling census, aiming to cover all areas of the country during a 5 year time period, while the US has moved to a short form only census accompanied by a continuously administered sample survey. In addition to these alter native strategies, increasing effort is required in order to deal with the impacts of census nonresponse, including enumerator follow up of households from which there is no return and statistical imputation of missing values in the census data.

The Role of Census Mapping

Modern census organization involves the coordination of a very large workforce and geographical information systems (GISs) are routinely used to manage the logistics of enumeration. This organizational mapping does not generally form part of published outputs but nevertheless underpins the operational success of the census. Suc cessful geocoding of individual census returns, linking them to the appropriate areas, is essential if counts are to be correctly aggregated. The management of address lists and geography lookup tables is a major element of census operations and often serves to underpin the geocoding of many noncensus data sources. Enumeration areas used are not necessarily the same as those used for publication of results, which usually need to be meaningful in terms of administrative and electoral geographies or are de signed for general purpose statistical reporting.

Mapping makes a unique contribution to the inter pretation of census results due to the ability of the human eye to identify pattern in graphical data. Thus, complex patterns of social geography may be readily conveyed by mapped census data which would be almost impossible to extract from tabular results. Ongoing fascination with these representations can be traced back to nineteenth century interests in the mapping of social conditions such as Charles Booth's poverty maps of London, through to the development of social area analysis and geodemo graphic classification whereby areas are grouped ac cording to their characteristics defined across a large pool of census variables. The production of a final map, for example, of an area classification scheme, may be the result of extensive unseen manipulation in GISs and statistical software. The dual independent map encoding (DIME) digital data structure developed to accompany the 1970 US census was influential in the emergence of GIS standards and one of the first examples of a statistical organization's digital boundary data to accompany a census. Digital boundaries allow users to map and analyze census data using their own software and provide the population base layer for many GIS applications. While geodemographic classification tends to find com mercial applications, such as direct marketing or site location for retail outlets, similar approaches also serve as the starting point for area based government policy, particularly the weighting of additional resources to neighborhoods with high levels of unemployment and multiple social deprivation.

All aggregated census data are susceptible to the modifiable areal unit problem (MAUP). This relates to the fact that observed patterns are a reflection of the number and placement of area boundaries as much as of the underlying population characteristics: by redesigning area boundaries it is frequently possible to change the apparent pattern in the map. Such aggregate represen tations are also subject to the ecological fallacy, whereby associations seen in the data at one level of aggregation will not necessarily be observed between individuals in the population, or at other levels of aggregation. These phenomena are particularly relevant to the use of census mapping in political districting. The general principle requires that small census areas be grouped into larger electoral districts of similar population size to achieve equality of representation, either by manual or automated means. The ability to concentrate supporters of particular parties by boundary redesign can also lead to the direct manipulation of boundaries for purposes of electoral gain, known as gerrymandering.

Techniques and Examples

By far the most widely used cartographic representation of census data is the choropleth, or shaded area, map. The example provided in **Figure 1** shows the 2001 census percentage of white ethnic groups in the city of Southampton, England. Values in the map are based on the aggregation of all the individual census results within each output area, clearly revealing how the outer suburbs are almost entirely white while some central areas are dominated by nonwhite ethnic groups. It is generally inappropriate to map counts in this way, as the values will often be related to the size of the area rather than prevalence in the population. A disadvantage of all choropleth mapping is that the visual dominance of an area is related to its geographical area rather than its population size. The largest areas in **Figure 1** comprise mostly docks and open water, while the most densely populated areas are so small as to be barely visible. This type of map is increasingly available online from the websites of national statistical agencies, such as that illustrated in **Figure 2**, from the Neighborhood Statistics service for England and Wales which again shows Southampton, this time the percentage of households without a car in 2001 for larger census areas called wards. In this case, additional interactivity is provided by dynamic linkage between the histogram and the map: pointing at an area on one will highlight its identity and position in the other. The figure also demonstrates the effects of generalization, compared to **Figure 1**, in which the map has been simplified to speed the web application, with the consequence that these area boundaries would be difficult to locate accurately on a street map of the city. The availability of some census data within online mapping and virtual globe software such as Google Earth has not to date produced new forms of mapping so much as increased options for access and interaction.

If area boundaries are not available, it is possible to produce simple maps by the generation of synthetic boundaries, known as Thiessen or Voroni polygons, or the placement of proportional symbols at centroid locations

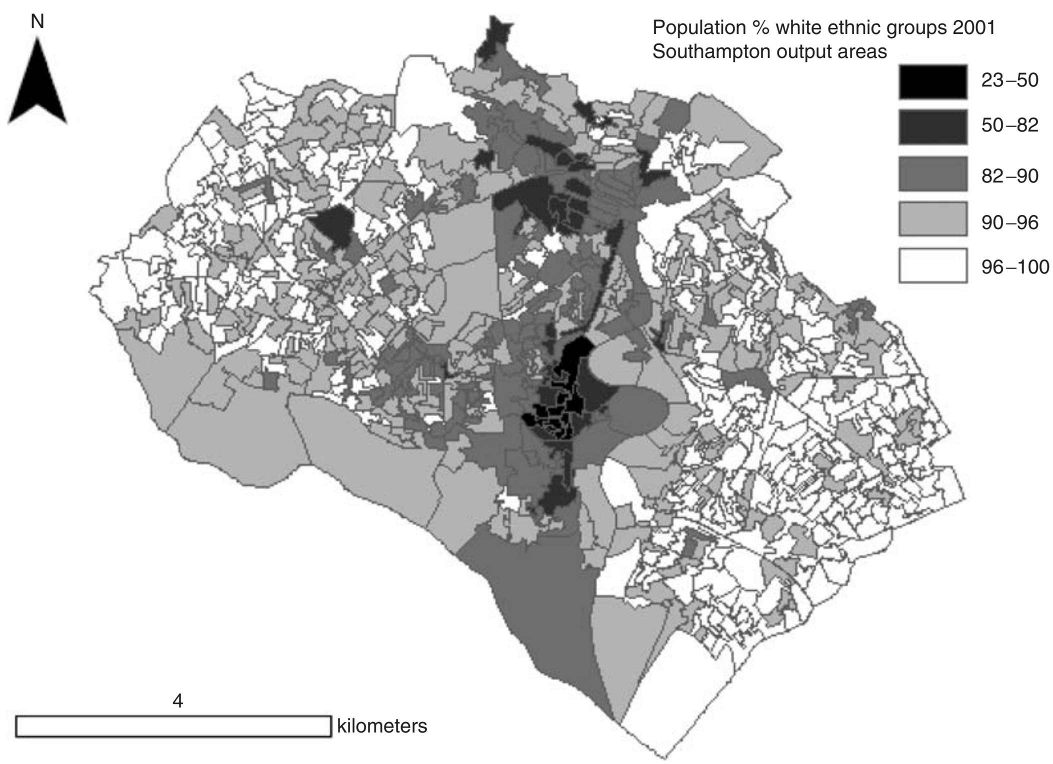

Figure 1 Choropleth map showing percentage population from white ethnic groups for output areas in Southampton, England. Source: Office for National Statistics, 2001 Census: Key Statistics (England and Wales). ESRC/JISC Census Programme, Census Dissemination Unit, MIMAS (University of Manchester); 2001 Census, Output Area Boundaries. Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

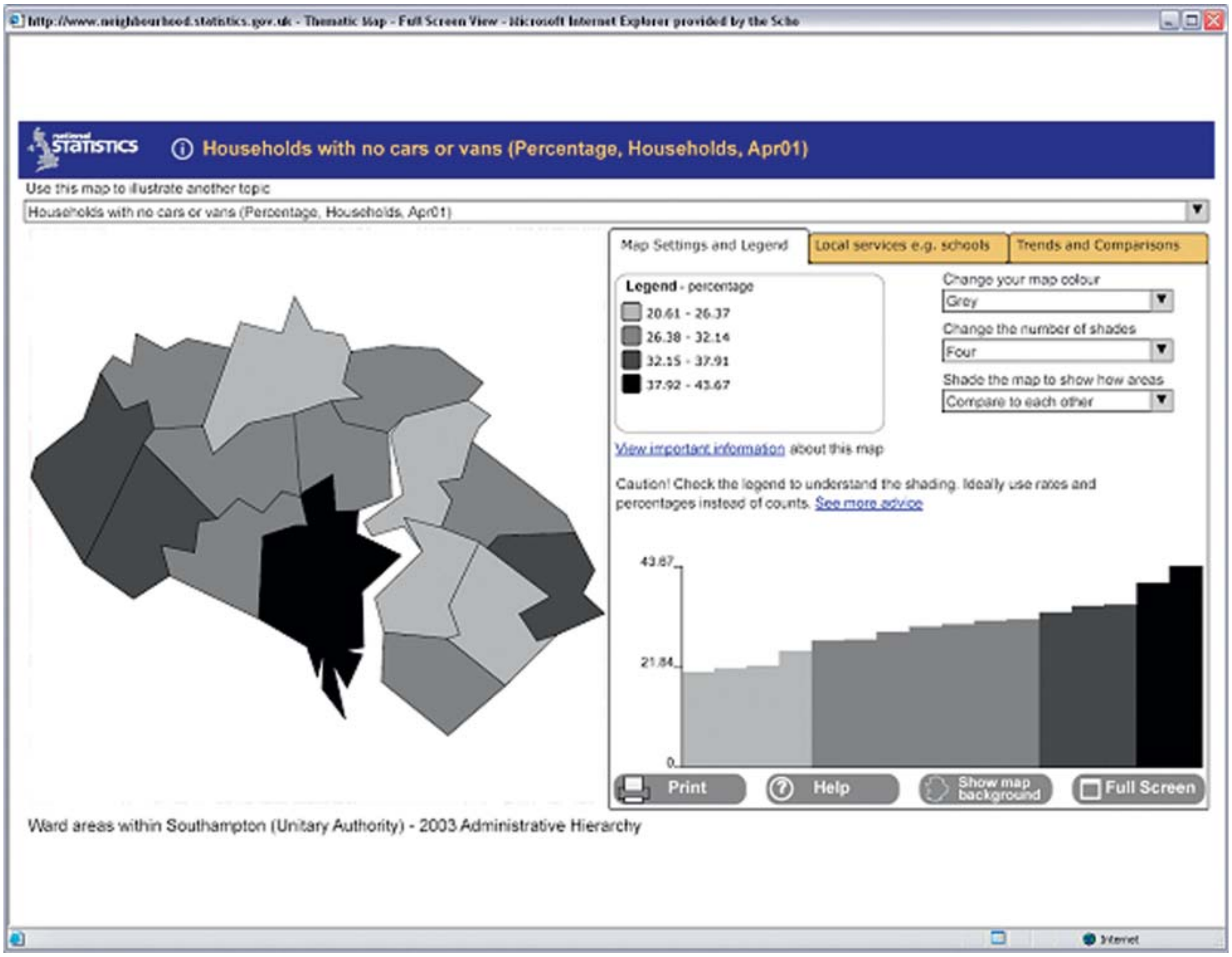

Figure 2 Online mapping of national statistical agency data: Percentage of households without a car for wards in Southampton, England. Source: National Statistics website: www.statistics.gov.uk Crown copyright material is reproduced with the permission of the Controller of HMSO.

representing each area, although these types of representations tend to introduce additional interpretational difficulties, for example, due to altered area geometry or overlapping symbols in areas of high population density. All census mapping is affected by general cartographic design considerations. Choices of map scale, number and placement of class intervals, selection and order of colors, and other layout considerations will affect how the map reader interprets the data. A further challenge facing census users is that of comparing data from successive censuses when there have been changes in the area boundaries. True comparisons can only be made for areas which encompass equivalent population groups. Where boundaries have changed it is necessary either to aggregate to the smallest possible comparable areas or to interpolate values from one time period to another. This challenge makes long term comparison of local populations especially difficult where census outputs are tied to administrative boundaries which themselves experience high levels of intercensal change.

A variety of alternative techniques are available which address some of the deficiencies of choropleth mapping. In dasymetric mapping, additional information is used to restrict the shading of mapped values. For example, a map of inhabited areas might be superimposed on census boundaries and shading displayed only within the inhabited areas, thus providing a much more accurate impression of the population distribution. Another approach designed to address this type of difficulty is population surface modeling, whereby census variables are redistributed onto the cells of a regular grid, as illustrated in **Figure 3**. This figure again shows the Southampton region, here showing population density in 200×200 m grid cells from the 1991 census. In this case, many of the cells remain unpopulated and the resulting map more accurately reflects the true distribution of population. These regular grid cells can aid comparison over time or with the output of other spatial models but the unconventional appearance of the map can deter users familiar with choropleth representations.

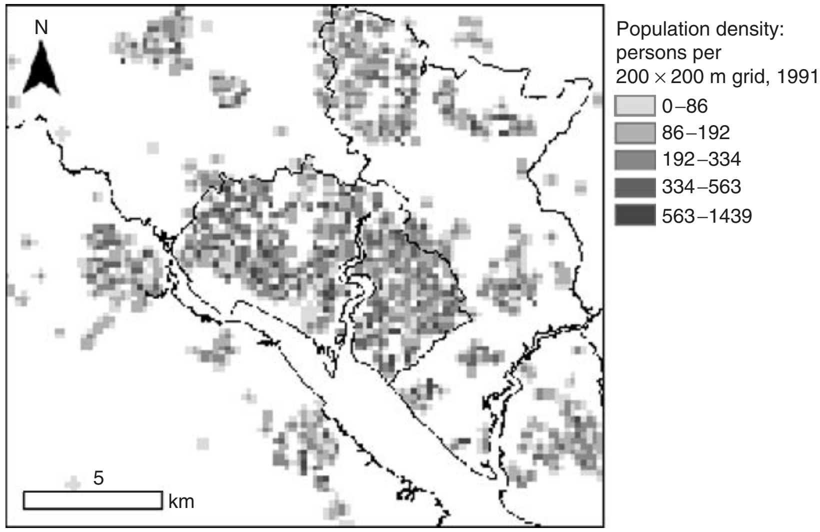

Figure 3 Population surface model showing population density for 200 × 200 m grid squares in Southampton region, England with coastline and local government boundaries overlaid. Source: 1991 Census: Small Area Statistics (England and Wales). Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland. Digitized Boundary Data (England and Wales) provided with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown and the ED-Line Consortium.

An entirely different data visualization is produced by the use of a population cartogram, in which locational accuracy is relaxed in order that each area can be represented proportional to its population size. Using this approach, areas with larger populations will take up a greater proportion of the map, but with the consequence that distance and direction cannot be accurately interpreted. An example of this approach is illustrated in **Figure 4**. This map shows change in the percentage of households without a car between 1991 and 2001. The cartogram illustrates the enormous population dominance of the London conurbation, represented by the dark area to the lower right of the map, where growth in car ownership has been much slower than in more remote areas with smaller populations. The figure embodies two important innovations over a conventional choropleth map: first, the sizes of the areas (here, counties and unitary authorities) are proportional to their population sizes and second, the use of stable geographical areas has permitted the creation of a map showing change over time.

Censuses continue to provide an enormously rich resource for those concerned with social geography. The large statistical data volumes make mapping an essential key to understanding geographical trends and processes in these data, although much spatial analysis is performed without the production of intermediate maps. The choropleth method continues to dominate census mapping, both in the popular media and academic publications, although its representational weaknesses are well documented and a variety of alternative methods are available which allow the user to manipulate and interrogate census data with greater geographical and statistical rigor.

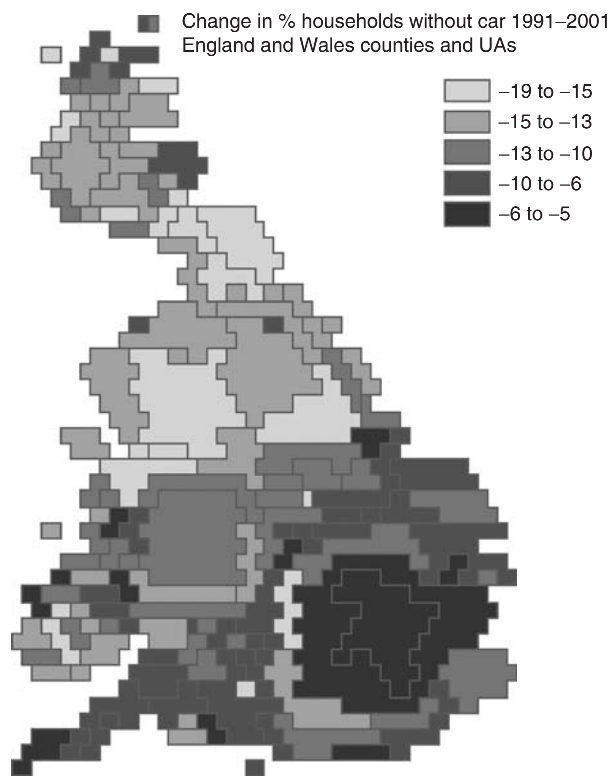

Figure 4 Population cartogram showing change in percentage of households without a car for counties and unitary authorities in Britain. Source: 1991 and 2001 Censuses: Small Area Statistics/Census Area Statistics (England and Wales, Scotland). Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland. Boundary data: JISC CHCC project www.ccg.leeds.ac.uk/teaching/chcc.

See also: Geodemographics; Georeferencing, Geocoding; Geovisualization; Gerrymandering; Internet/Web Mapping; Modifiable Areal Unit Problem.

Further Reading

Brewer, C. A. (2005). *Designing Better Maps: A Guide for GIS Users*. Redlands, CA: ESRI Press.

Brewer, C. A. and Suchan, T. A. (2001). *Mapping Census 2000: The Geography of US Diversity*. Washington, DC: US Census Bureau.

Cook, L. (2004). The quality and qualities of population statistics, and the place of the census. *Area* 36, 111 123.

Curry, M. R. (1998). *Digital Places*. Routledge: London.

Dewdney, J. C. (1983). Censuses past and present. In Rhind, D. W. (ed.) *A Census User's Handbook*, pp 1 16. London: Methuen.

Dorling, D. and Thomas, B. (2004). *People and Places: A 2001 Census Atlas of the UK*. Bristol: Policy Press.

Durham, H., Dorling, D. and Rees, P. (2006). An online census atlas for everyone. *Area* 38, 336 341.

Martin, D. (1996). An assessment of surface and zonal models of population. *International Journal of Geographical Information Systems* 10, 973 989.

Martin, D. (2006). Last of the censuses? The future of small area population data. *Transactions of the Institute of British Geographers* 31, 6 18.

Monmonier, M. (1996). *How to Lie with Maps* (2nd edn.). Chicago, IL: University of Chicago Press.

Openshaw, S. and Rao, L. (1995). Algorithms for reengineering 1991 census geography. *Environment and Planning A* 27, 425 446.

Singer, E. (2003). The eleventh Morris Hansen lecture: Public perceptions of confidentiality. *Journal of Official Statistics* 19, 333 341.

United Nations (2000). *Handbook on Census Management for Population and Housing Censuses*. New York: United Nations.

Wood, D. (1992). *The Power of Maps*. Routledge: London.

Relevant Websites

http://www.census.gov/main/www/stat int. html
List of national statistical organizations maintained by US Census Bureau.

http://www.neighborhood.statistics.gov.uk
Neighborhood Statistics Service for England and Wales census mapping tools.

http://www.personal.psu.edu/cab38/Color Brewer/Color Brewer.html
Online guidance on class intervals and color schemes for thematic mapping, Penn State Personal Web Server.

http://www.ccg.leeds.ac.uk
The CCG Online GIS atlas British census cartogram tools.

http://www.statistics.gov.uk
UK Statistics Authority.

http://factfinder.census.gov
US Census Bureau's American Factfinder census mapping tools.

Central Business District

G. A. Rice, University of Helsinki, Finland

Glossary

Edge Cities Refers to patterns of large-scale urban sprawl into the rural vastness of the US. Edge cities are quite different from other noncentral urban developments, like British New Towns, in scale, intensity, and accessibility. Also unlike New Towns, edge cities were not specifically designed to manage the overflow of post-war baby boomers but rather, they are a more explicit attempt to attract corporations and light industries through offering tax cuts and a local skilled labor force.

Technopoles Very similar to European industrial districts especially those found in Northeast Central Italy. While both technopoles and industrial districts aim to nurture a specific industrial atmosphere, in the former case the emphasis is on attracting established research-based companies and innovation centers which cater to smaller high-technology companies. Technopoles tend to follow a number of stages from proto-type to larger agglomeration that may eventually evolve into edge cities or large urban centers outside older established cities. Southern California is a case in point and thus is often referred to as a polycentric region.

Zone of Assimilation An area adjacent to the central business district (CBD) characterized by extensive redevelopment and the spread of new and typically upmarket shops and office developments (cf. zone of discard). This is the favored frontier and the physical direction that the CBD tends to extend along. As long as demand for space in the CBD continues to rise, the zone of assimilation will be constantly pushed and reshaped.

Zone of Discard An area at the edge of the CBD characterized by transitory business, boarded up premises, and unsuccessful commerce. The rent levels tend to be less expensive than in the main CBD which attracts smaller low-market retailers, warehouse activities, and wholesale trading premises. In many cities including Sacremento and Phoneix the zone of discard represents the old CBD that has become outstripped by newer developments and more general supply side changes in the retail and service sectors ;(cf. zone of assimilation).

Introduction

The central business district (CBD) is that part of the city which contains the principal commercial streets and main public buildings. Throughout history the CBD has been characterized by a number of land use changes that in clude industrial, residential, commercial, administration, and consumption. These spatial and temporal changes have been used to support and reject a number of theories and delimitations of the CBD. However, in the wake of post war decentralization and the rising popularity of polycentric regions in both Europe and the US much of this work has itself come under criticism. The main thrust of the criticisms suggest that CBDs are now more likely to experience atrophy due to the frenetic rise of suburban lifestyles, capital flight, and edge cities in regions like Southern California. Notwithstanding, parallel develop ments like inner city regeneration suggest that CBDs are now at the heart of a more widespread back to the city movement that (re)promotes the attractiveness of city center living through gentrification, enchanting cultural districts, and an *uber* cool social scene.

The following entry will chart the ways in which the CBD has been used to make a number of theoretical contributions to human geography. This is evident from the definitional nuances and the various ways in which CBDs have evolved and subsequently been delimited. While CBD evolution takes on local morphological characteristics, the case of Boston shows how CBDs evolve internally and expand physically. This is compli mented by a more general treatise of CBD evolution based on North American and Canadian cities. Various approaches to CBD delimitation are also discussed to demonstrate the importance of urban land rent and changes across the land use surface. The entry then fin ishes off by suggesting that retail decentralisation is one of the biggest threats to the future of the CBD. The case of the UK mirrors more general trends in Europe and the US and raises challenging issues for national planning and policy guidelines. The proliferation of regional malls is testament to the purported porosity of these guidelines that continue to galvanise various reactive urban renais sance initiatives encapsulated by the contemporary CBD.

Defining the Central Business District

The CBD typically refers to the commercial heart of the city. The term is often used interchangeably with downtown and city center to describe a distinct land use pattern that is delimited from the rest of the urban settlement (see below). In Canadian and US cities with more mixed downtown areas the terms midtown and uptown are also used to distinguish them from the geo graphical heart of the city in the CBD. In Waterloo, Ontario, uptown refers to the prime shopping area to distinguish it from downtown Kitchener. Other examples include the midtown area of New York City and uptown Minneapolis centered about the intersection of Lagoon Street and Hennepin Avenue. In world cities such as London the separation of business functions from the CBD and the shopping area in Oxford Street was under way by the end of the sixteenth century and by the nineteenth century the principal shopping area became detached from the main city (this refers specifically to the financial district). Smaller cities like Glasgow have fol lowed a similar geographical shift with the current CBD located to the west of where the city first began to grow at 'Mercat Cross'.

The definitional nuances and spatial fragmentation of the CBD have enabled a number of theories to emerge. A set of approaches that capture the general spirit of these theories are collectively known as core frame models. These models suggest that the core of the CBD is distinct from the wider frame (or fringe) due to differences in the types and intensities of land use (**Figure 1**).

This core frame spatial organization is also evident in non Western cities like Cape Town where the CBD is flanked by industrial parks in the southern suburbs of Claremont, Rondebosch, and Wynberg. Similarly, in UK cities other activities requiring less relative centrality, including court houses and libraries, are located toward the edge of the core while functional nodes make up the surrounding frame. Glasgow's CBD is a case in point with the Mitchell Library, Sheriff Court and warehouse clus ter located outside the main core area that is demarcated by primary retail streets. These frame functions tend to be focused around smaller nuclei but in cases like the

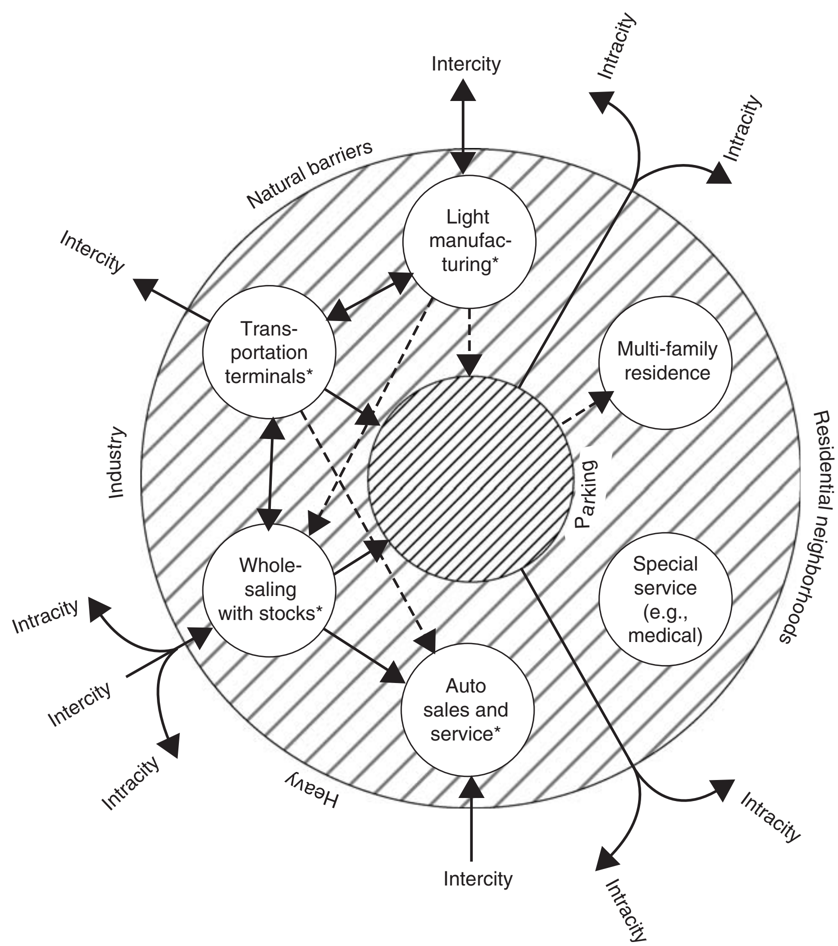

Figure 1 Core-frame model after Horwood, E. M. and Boyce, R. R. (1959). *Studies of the Central Business District and Urban Freeway Development*. Seattle: University of Washington Press, taken from Scargil, D. I. (1979). *The Form of Cities*, p 90. London: Bell and Hyman.

Tees Valley outside Newcastle in Northeast England the nuclei can grow into an extensive industrial complex. In Southern California the growth of light industrial tech nopoles in Palmdale are located some 60 miles to the north of the CBD in Los Angeles. In other cases, this decentralization proper is flanked by widespread resi dential sprawl and the rise of suburbia that in turn is stabilized by 'megamalls', including the Mall of America and the West Edmonton Mall. Consequently, core frame definitions of the CBD are now well overstretched.

Despite definitional nuances and morphological vari ations, it is generally accepted that the CBD is charac terized by a concentration of high order service activities that include retailing, office, legal, and other municipal functions. The literature suggests that there are three main reasons for the clustering of these activities with 'unique accessibility' being the first and most important. This is an 'all roads lead to Rome' type scenario and places the CBD at the hub of the primary transport network. Chicago's famous Loop is designed in this way with office complexes built above railway stations to fully exploit the local labor force. Second, the clustering of commercial activity can also bring economic advantages. While land use varies in the CBD, most of the functions are complimentary and independent. Firms can achieve external economies of scale by having access to what is, in effect, a proximate pool of common services. Numerous studies of London show the extent of spatial linkages between different types of office activities that include press, industry, insurance, banking, and commodities. Third, there are also behavioral reasons that help explain the centrality of high order service activities. In the business world a city center address arguably carries more prestige and the ability to afford such an expensive location is perceived as a sign of success. The 'skyscraper infatuation' that swept through 1920s America was characterized by the erection of prestigious tower buildings established as landmarks to illustrious entre preneurs. Furthermore, the social scene and leisure economy concentrated in and around the CBD further facilitates personal communication and tacit forms of knowledge favored by many businesses.

Evolution of the CBD

As the geographical concentration of commercial activity in the city, various CBD forms have been recognizable since at least the end of the eighteenth century. However, the manner in which the CBD functions is medieval as the linkages within the core area often need to be ac cessible by foot, which places an upper limit on the distance between functions that benefit from being linked together. This is particularly the case in smaller low rise urban centers like San Jose where city officials only need to walk one block to purchase spare car parts. However, in cities such as New York and London commercial pressure on space over time eventually led to the ex clusion of (most) other central business functions ex emplified by emergence of global financial districts in Lower Manhattan and Canary Wharf, respectively. Re search in smaller US cities including Philadelphia show that, as long as this pressure remains constant then the evolution of the CBD will always be in flux rather than some static element on the urban landscape. While scholars support this general thesis, there is less agree ment about the precise sequence of driving forces behind CBD evolution. In European cities like Utrecht the CBD evolved from the Industrial Revolution, whereas in Latin American and African cities rural to urban migration was the most important factor. To expose this lacuna the literature has focused on a number of trophy case studies to launch more general ideas about CBD evolution.

The Case of Boston

While scholars rightly continue to identify various local factors as driving forces behind CBD evolution, the case of Boston contributes something much more significant to our understanding of the process. It is the evolution of Boston's CBD from 1840 to 1920 that established a connection between the ways in which the CBD evolves internally 'and' expands physically. The evolution of Boston's CBD also fits with the various urban land rent models which suggest that those land uses least able to pay the highest rent will be relegated toward the edge of the city in zones of discard (discussed below).

The evolution of Boston's CBD from 1840 to 1920 is based around the emergence of 'modern' functional areas that sprouted from three specialized nuclei (**Figure 2**).

In 1840 food markets, warehouse, and financial func tions were largely restricted to central locations to gain access to Boston's large European immigrant population. By 1880 Boston's commercial activities had expanded to meet the demands of regional distribution that in turn led to the expansion of business premises. As a consequence many central residential areas suffered the social dis organization created by the problems of eviction and residential relocation. By 1920 Boston's CBD had split into more specialized functions than had initially spurted its growth. Finance and administration coalesced and expanded their accommodations into premium locations within the existing limits of the CBD, rather than at the expense of peripheral residential locations. By contrast the food industry reorganized their facilities by de centralizing into adjacent residential areas based upon the differing threshold populations required to support different types of retail function. This is because on the supply side different retail functions have different conditions of entry (thresholds), and thus demand

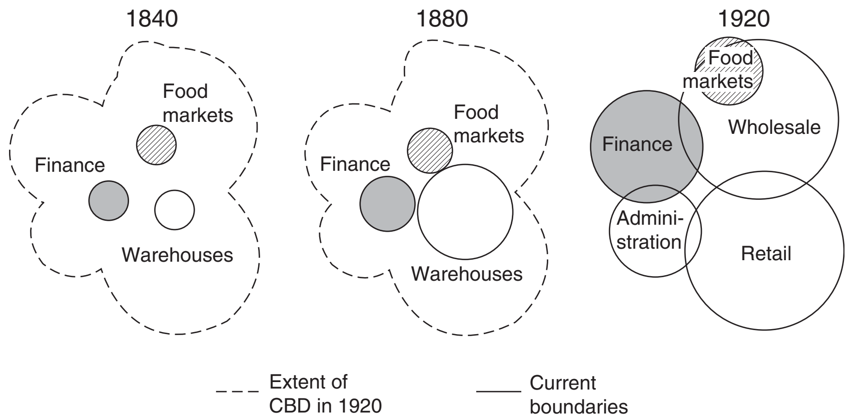

Figure 2 Evolution of Boston's CBD 1840 1920. From Ward, D. (1968). The emergence of central immigrant ghettoes in American cities: 1840 1920. *Annals of the Association of American Geographers* **58**, 343 359.

minimum trade areas for their support. On the demand side, consumers spend different proportions of their in come on different goods and services and purchase them with differing degrees of frequency. This produces the hierarchy of retail functions found in most American and European cities.

A CBD in Seven Stages

Other types of approaches are more schematic in ex plaining CBD evolution. A number of generic evo lutionary stages are proposed to account for the ways in which the CBD can be characterized by various moments of evolution ranging from 'inception' to the 'city of realms'. While there can be overlap between any of the individual stages, schematic approaches assume that there is a 'normal' trajectory that CBD evolution will follow. North American and Canadian cities are thought to be shaped (to different degrees of intensity) by the forces outlined below *ad seriatim*:

Inception

The original setting of a town or city contains pervasive influences on the future of the city and its citizens. Canadian settlement frontiers were linked to the building of railroads that nurtured the growth of prairie towns that grew through phases of urban redevelopment charac terized by various inceptions.

Exclusion

Once a town has been founded, growth occurs through the mechanism of the rent gradient (discussed below) that delimits the CBD from its surrounding area. This is evident in US cities such as Chicago where the CBD is the first of five zones radiating outward to form distinct concentric circles. The 'Chicago model' has greatly en hanced our understanding of how a CBD evolves by demonstrating that certain activities are 'pushed' outward due to the limited amount of central land coupled with the rent paying abilities of particular functions. Retailing is often used as a primary indicator of exclusion as higher order services including department stores can afford a central location while smaller shops and neigh borhood centers cannot. Exclusion also extends to non CBD uses such as residential, industrial, wholesale, and vacant land that characterize the zone of discard. The spatial patterns associated with exclusion are directly related to the physical direction of CBD growth along the zone of assimilation.

Segregation

While exclusion drives certain uses 'out' *of* the CBD, segregation drives certain land uses apart 'within' the CBD. As the clustering of activities grows in the CBD, the likelihood of functional segregation increases because greater amounts of one type of activity tend to ultimately exclude other CBD functions. New York's financial dis trict in lower Manhatttan and Vienna's Museum Quarter are cases in point. This functional segregation is a result of previous land uses becoming so self contained that they lack any significant external linkages or economies of scale.

Extension

As the CBD continues to evolve, distinctions are made between different types of extensions. In the nineteenth century the typical CBD was a compact unit at the focus of crosscutting routes and the typical residential area was a 'bead on a string' shaped by suburban railway lines. After World War I, CBDs generally began to grow out ward in thin alignments along arterial streets in the zone of assimilation, although cities like Chicago arguably extended into its hinterland as concentric circles. How ever, it is not the physical direction of the extension from the CBD but rather, the potential trade losses to outlying locations that have come to dominate the literature.

Replication and readjustment

Replication and the dynamic of adjustment are major concerns for the economy of the CBD. This is part of a more general trend of decentralization whereby the CBD experiences a trade deficit due to the rise of comparative services in suburbia and other outlying areas. The rise of regional shopping malls in the UK and technopoles in the US are indicators of this decentralization. The realization that business and trade can be successfully located out side the CBD is not just a consequence of limited central land supply, but also due to greater individual mobility and wider consumer choices. After World War II, in creasing decentralization and capital flight suggested that CBDs lacked the relative urban weight that they once had. The failure of various attempts to reshuffle the hierarchy back in favor of the CBD prompted a number of critical scholars to raise concerns about an 'urban crisis' or the 'crisis of contemporary urbanity'. This doom and gloom thesis became particularly influential in the 1960s and 1970s when the slowdown of manufacturing and heavy industry gave rise to a number of subsequent problems including high levels of unemployment and inflation.

Redevelopment

As a result of decades of urban decay in the CBD, re development and physical planning have taken on more teleological roles. At first these redevelopment initiatives attempted to stamp out urban decay only to realize that social problems cannot be solved by physical planning solutions alone. Since the 1980s redevelopment of the CBD has shifted to improving urban livability for con ditionally more prosperous residents. This is part of a widespread willingness of municipal authorities to accept change 'only if' it is 'for the better'. Thus, inner city slums are not replaced with similar low income housing but rather upmarket postmodern flats bankrolled by the private sector under the aegis of the neoliberal local state. The selectivity associated with the redevelopment agenda remains one of the most debated areas in human geography and right across the social sciences.

A city of realms

Despite continuing decentralization, the final stage in the evolution suggests that the CBD will remain the location of highly specialized functions. While there is general acceptance that some activities have moved out of the CBD, there is no universal agreement that this de centralization spells euthanasia for the CBD. This opti mism is partly related to the purported success of various forms of civic boosterism, the exploitations of location specific assets, and cultural regeneration programs col lectively referred to as 'selling places'. Appellations in cluding the European City of Culture and UK City of Architecture of Design and events like the Olympic Games reinforce the symbolic importance of the CBD as an integral component of urban civic consciousness. The ethos behind such initiatives is to (re)bolster the eco nomic and cultural values of the CBD as the prime lo cation for business and social life.

Delimiting the CBD

Just as the CBD has evolved through a number of dif ferent stages there have been various attempts to delimit its boundaries traditionally defined by a 'blighted zone' (of discard), characterized by obsolete and deteriorated property. At the most visible level the CBD can be delimited by the highest buildings, the heaviest concentration of traffic and pedestrian flows, and a large daytime working population and a low residential population. However, there is general agreement that these observations only provide crude indicators of the CBD's magnetism and are therefore not entirely satisfactory.

The main theoretical contributions for delimiting the CBD have come from approaches that employ land val ues and indices. These include floor space, business in tensity, and valuation of properties. The application of land rent is based upon earlier principles most associated with von Thunen's explanation of agricultural land uses. His central idea is that the economic gain from farming land decreases with an increase in distance from the market (CBD) due to more expensive transport costs. In urban environments the general assumption with land value delimitations is that individual land uses will compete for the most desirable central location in or as near to the CBD as possible. Potential benefits can in clude unique attributes such as strategic location, local market demand, integration with regional clusters, and human resources. However, these potential benefits come at a cost of having to pay the highest amount of rent to be closest to or at the peak land value intersection (PLVI). This is an area in the CBD where the main roads and other transport links intersect, but the actual position of the PLVI can shift depending on the direction of CBD growth along the zone of assimilation. A useful general ization based on North American cities shows how the CBD builds up around the PLVI first before extending out along the main transport corridors (**Figure 3**). Any further increases in distance away from the PLVI means a drop in land values which in turn produces changing physical urban land use patterns (retail, office, and residential).

The schematic representation in **Figure 3** is based on the land value surface that is, a reflection of accessibility within the urban area. Since accessibility differs across the urban area land use patterns are based on the ability to pay rent for the most accessible locations in the CBD.

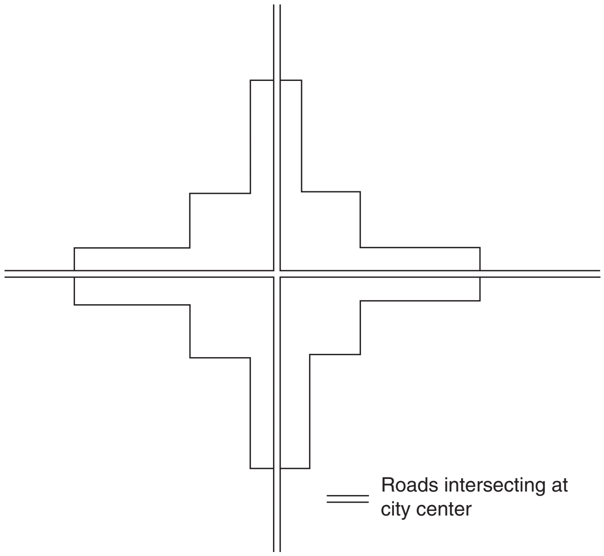

Figure 3 Idealized outline of the CBD according to Murphy and Vance 1954 taken from Scargil, D. I. (1979). *The Form of Cities*, p 87. London: Bell and Hyman.

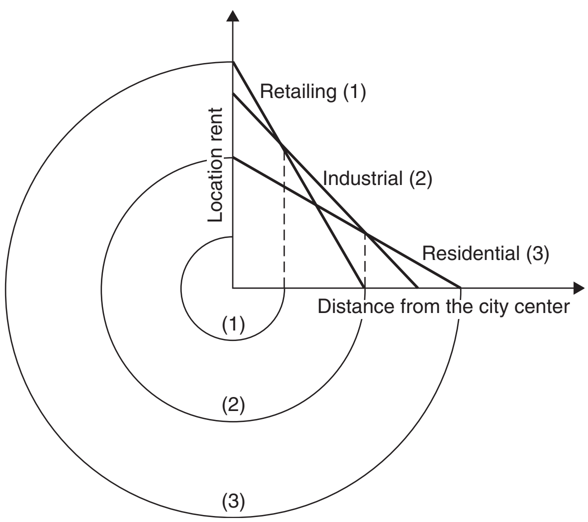

Figure 4 The Alonso Model after Alonso, W. (1946). *Location and Land Use: Toward a general theory of land rent*. Cambridge, MA: Harvard University Press taken from Johnston, R. J., Gregory, D., Pratt, G. & Watts, M. (2000) (eds.) *The Dictionary of Human Geography* (4th edn.), p 21 Malden, MA: Blackwell, (See also http://www.xreferplus.com/entry/733934)

This is supported by a number of other models which suggest that the land use most dependent upon accessi bility outbids other potential users. **Figure 4** shows that retailing (usually malls, department stores, and flagship stores) is most dependent upon accessibility followed by offices and then residential land uses. The resulting type of land use pattern is also associated with steeper declines in the offer price curve for those functions (retailing) that are most dependent upon a central location.

The CBD and the Threat of Decentralization

Many changes have occurred in Western society since 1945 that have implications for the twentieth first century CBD. These changes include fluctuating urban economic performance, the rise of global neoli beralism, and advances in process and product technol ogies. Notwithstanding, it is now near customary to claim that decentralization, especially in retailing, has had the biggest impact on the fate of the CBD. More recently the academic literature has been concerned with the con tinued growth of retail parks and regional malls in the UK and mega malls in North America that are purported to threaten the long term future of the CBD. This poses a problem because while CBDs are expected to compete with one another at the interurban scale, they must first and foremost be able to survive in their own regional economies. These trends have been complimented by a strong current of deindustrialization exemplified by the rise of technoploes in the US and growing support for polycentric regions in European planning circles. This restructuring of the space economy is particularly problematic because in relation to retailing and business parks at least, the CBD is unable to fully exploit the general shift toward major space units and the demand for adjacent car parking.

The Case of UK Retailing

As decentralization continues apace a number of CBDs have experienced propulsive growth at out of town lo cations. In the case of the UK this trend is nothing new, as retailing has existed at non CBD locations since before the end of the nineteenth century. It wasn't until 1947 that the Town and Country and Planning Act officially established a centralized system of planning. This meant a traditional approach to retailing as both central and local government aimed to preserve a hierarchy of shopping centers in accordance with central place theory. Consequently, proposals for peripheral sites were per ceived as a threat to the established retail hierarchy and the viability of the CBD as a whole. This stands in marked contrast to the post war situation in North America where out of town regional shopping centers and mega malls were commonplace.

However, since the 1960s the British retail environ ment has altered dramatically through decentralization that reflected both morphological and organizational changes. The opening of Brent Cross in 1976 was fol lowed by other out of town malls including the Metro Centre, Meadowhall, Merry Hill, Lakeside, Bluewater, and Braehead. These out of town sites offer good road access, large amounts of free parking, major retail chains acting as anchor tenants, and some form of nonretailing activity such as the ice rink at Braehaed and the concert venue at the Trafford Centre in Manchester.

One influential approach usefully captures this retail decentralization in three waves:

- Wave 1 – commencing in the late 1960s involved the decentralization of food retailing in the form of free standing superstores and hypermarkets.
- Wave 2 – beginning in the 1970s, was the decentral ization of stores selling household goods, including do it yourself goods, furniture, carpets, electrical goods, and motor accessories.
- Wave 3 – from the mid 1980s onward was associated with the exodus of comparison goods such as 'high streets out of town'.

With the election of Margaret Thatcher in 1979 the New Right aggressively pursued the politics of deregulation through 'opening up' the UK economy for international business. Shortly afterward, moves to make planning more 'business friendly' began with the Department of the En vironment's Circular 22/80. This asked local authorities to prioritize planning applications that contributed to na tional and local economic activity. In recognition of these evolving changes a final draft of the National Planning Guidelines in England and the National Planning and Policy Guidelines in Scotland was drawn up and planning authorities were asked to facilitate retail development, not only in existing centers but where appropriate on sites detached from such centers. These neoliberal conditions set the scene for the fiercely competitive 'store wars' that continue to pose a threat to British CBDs today.

While shopping malls pose a threat to the British CBD, they have also been at the forefront of the 'back to the city movement'. A number of cities up and down the UK including Birmingham, Leeds, and Glasgow have used the seductive trappings of retailing as a privileged tool to gain competitive advantage in their respective regional hinterlands. The new Bullring Mall has injected £1 billion into the center of Birmingham and the White City development in London is expected to generate more than £2 billion of fresh investment in the next 10 years after it opens in 2008. In Glasgow city center, the opening of Buchanan Galleries Shopping Mall in 1999 reflects another attempt where the CBD is attempting to become more competitive with out of town retail de velopments at Braehead and Glasgow Fort. Thus, the use of malls in the CBD is as much about restoring the downtown area as a center for economic activity as it is about capitalizing upon consumer demand for high order and salubrious retail facilities. This trend is compounded by more recent political shifts that will provide cities with the chance to elect their own mayors. Despite an initial lukewarm take up, the UK Government is pushing the idea of a mayoral model that if successful may pro vide stronger leadership for CBDs in the future.

Conclusions

Throughout much of the twentieth century, CBDs served as areas of supreme importance and all surrounding areas sorted themselves into ranks of decreasing eminence. This economic determinism meant that high order ser vices including retailing were located in the CBD based on their ability to pay rent. However, the use of the term CBD is now more debatable than ever before; many cities now have 'central business' all over the central area. This is a result of continuing pressure on finite CBD space that intensified in the post war period. Consequently, the emergence of more polycentric regions has become a threat to traditional central place type urban patterns. In both North America and Europe this threat has created a more competitive space economy consummated by various non CBD developments including shopping malls, suburbia, technopoles, and edge cities. However, since the 1980s these developments have galvanized a number of CBD initiatives encapsulated by a more general 'back to the city movement'. This ongoing re invention of the CBD reiterates the competitive advan tage of cities through taking stock of various tax breaks and financial incentives that thrive in a climate of de regulation. This approach has had some success in the CBDs of world cities like London, New York, and Tokyo. Whether or not the CBDs of smaller cities can regain their dominance in relation to proximate ex urban lo cations remains to be seen.

See also: Chicago School; Gentrification; Land Rent Theory; Retail Geographies.

Further Reading

Bourne, L. S. (ed.) (1971). *Internal Structure of the City: Readings on space and environment*. New York: Oxford University Press.

Frieden, B. J. and Saglyn, L. B. (1990). Downtown malls and the city agenda. *Society* 27, 42 49.

Gelder, K. and Thornton, S. (eds.) (1997). *The subcultures reader*. London: Routlege.

Hall, T. (2006). *Urban Geography* (3rd edn.). London: Routledge.

Knowles, R. and Wareing, J. (1976). *Economic and Social Geography made simple*. London: W. H. Allen.

Lawless, P. and Brown, F. (1986). *Urban Growth and Change in Britain*. London: Harper and Row.

Lefebvre, H. (1970). *The Urban Revolution*. Minneapolis, MN: University of Minnesota Press.

Mumford, L. (1986). What is a city? In Miller, D. L. (ed.) *The Lewis Mumford reader*. New York: Pantheon Books.

Park, R. E. (1997). The city: Suggestions for the investigation of human behaviour. In Gelder, K. & Thornton, S. (eds.) *The Subcultures Reader*. London: Routledge.

Scargill, D. I. (1979). *The form of cities*. London: Bell and Hyman.

Scott, A. J. (2002). Industrial urbanism in late twentieth century California. In Dear, M. J. (ed.) *From Chicago to L.A.: Making sense of urban theory*. Thousand Oaks, CA: Sage.

Ward, D. (1968). The emergence of central immigrant ghettoes in American cities: 1840 1920. *Annals of the Association of American Geographers* 58, 343 359.

Relevant Websites

http://www.brc.org.uk/
British Retail Consortium.

http://www.easternct.edu/depts/amerst/Malls.htm
Shopping Centre Studies.

http://www.icsc.org/index.php
ICSC.

Central Place Theory

J. Malczewski, University of Western Ontario, London, ON, Canada

Glossary

Central Function The goods and services provided at a central place.
Central Place Center (e.g., urban place) for distributing goods and services to its market area.
Centrality The relative importance of a central place with regard to the region surrounding it.
Market Area The territory surrounding a central place limited by the range of goods/services supplied.
Range The maximum distance that a consumer will travel to purchase a good or service.
Threshold The smallest market size for an establishment to be economically viable.
Urban System A network of urban places and their hinterlands.

Introduction

Central place theory is concerned with the size, number, functional characteristics, and spacing of settlements, which are nodal points for the distribution of goods and services to surrounding market areas. An interest in some aspects of the theory can be traced back to the eighteenth and nineteenth century, when German scholars attempted to identify the relations between settlements and their complementary regions. During the first two decades of the twentieth century, American sociologists proposed models dealing with the theoretical arrangement of central places. The first explicit statement of central place theory was made by German geographer Water Christaller and refined by German economist August Lösch. Although there are important differences between the models of Christaller and Lösch, they share a number of commonalities in terms of reasoning and underlying assumptions. They both concluded that the most efficient spatial arrangement of central places takes the form of a triangular lattice so that each central place has a hexagonal market area. In the post–World War II period, the classical models have been modified, extended, and tested empirically by researchers from various disciplines including geography, economics, regional sciences, and urban/regional planning. Together, the models of Christaller and Lösch as well as later modifications and extensions of those models form a subdivision of location theory that is called central place theory.

Classical Central Place Theory

A Single-Function Central Place System

The central place model is based on a series of simplifying assumptions. It assumes that economic activities are taking place on an isotropic surface – an idealized region which is completely flat and homogeneous in every respect. This implies that there are no barriers to movement and the transportation cost per unit of distance is the same in all directions. The theory also presupposes that population and income are evenly distributed over the region. Finally, it is assumed that both consumers and producers have perfect knowledge and act entirely rationally with respect to this knowledge.

Given these simplifying assumptions, central place theory introduces two economic parameters which control the potential locations of individual central places: the threshold (the minimum level of demand necessary to support business activity) and the range of a good (the maximum distance a consumer is willing to travel to purchase a good or service). The theory addresses the spatial interplay between the threshold and range. Using these two parameters the central place model demonstrates how an economic landscape evolves towards a spatial equilibrium in which a single function central place system consists of equal, regular hexagonal market areas, just large enough to contain the threshold demand, served by central places located at equal distance from each another.

Multiple-Function Hierarchical Systems

The hexagonal pattern of market areas generated by the single function model provides a starting point for a more complex multiple function hierarchical system of central places. A key concept underlying the multiple function framework is that of the order of a place which is derived from the orders of goods (and services) that are available in it. Some goods have a low threshold and a correspondingly low range (low order goods), whereas others need a much higher level of demand for their existence and are characterized by a more extensive range (high order goods). Since threshold demand and market range vary enormously for different goods and services it is impossible for all the central place functions to be carried on in all central places. Lowest order goods and services are available at a large number of locations, highest order goods at only a small number of locations. Accordingly, Christaller's urban system is arranged in a nested hierarchy of central places and market areas with lower order market areas completely contained within higher order areas (**Figure 1**). Central places on each level of the hierarchy are characterized by a specific mixture of functions in accordance with their threshold and range values. At the top of the hierarchy is the highest order center with many specialized functions (including other functions as well), at the bottom of the hierarchy there are a great many small central places with a few lower order functions serving small market areas. Each level supplies all the goods and services that the centers below it will provide plus some additional ones. This implies interdependency between levels of the hierarchy in that all centers except the smallest have other centers dependent on them for supply of certain goods and services.

Christaller suggested that the relationships between subsequent levels in the central place hierarchy are fixed and defined by a *k* value, which indicates the number of lower order centers (hexagonal market areas) served by a central place at the next highest order in the system. Using the *k* value, Christaller proposed three types of systems based on the marketing, transportation, and administrative principles (see **Table 1**). According to the marketing principle (or the $k=3$ network) the hierarchy and nesting pattern results in a system that serves a maximum number of evenly distributed consumers from a minimum number of central places. However, a trans portation system to serve such an arrangement is not efficient because the important links between larger central places do not pass through intermediate ones. Consequently, a k 4 network organized according to the traffic principle was proposed to account for situations in which the costs of transportation were significant. The transportation principle results in an arrangement that maximizes the number of central places located on the main transport routes. In the hierarchical arrangement according to the marketing and transportation principles, lower order places (or market areas) located between higher order centers are divided equally among three centers in the case of the marketing principle, and between two centers in the case of transportation principle. Such a division of smaller areas and larger ones is inefficient from the perspective of spatial arrangement for purposes of administration. Consequently, Christaller suggested that in regions with a highly developed system of central administration a $k=7$ hierarchy would tend to develop. In this arrangement, the hexagonal market

Table 1 Christaller's systems of central places

Principle	*k-Value*	*The goal of the system is:*	*Arrangement of central places*
Marketing	k 3	to serve a maximum number of consumers from a minimum number of central places.	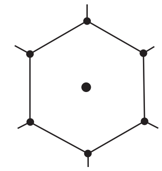
Transportation	k 4	to locate as many central places as possible along the main transport routes connecting the higher-order centers.	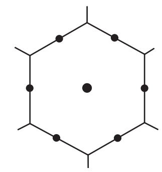
Administrative	k 7	to provide administrative control: lower-level administrative units are completely controlled by higher-level centers.	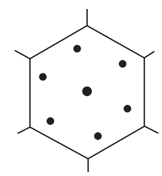

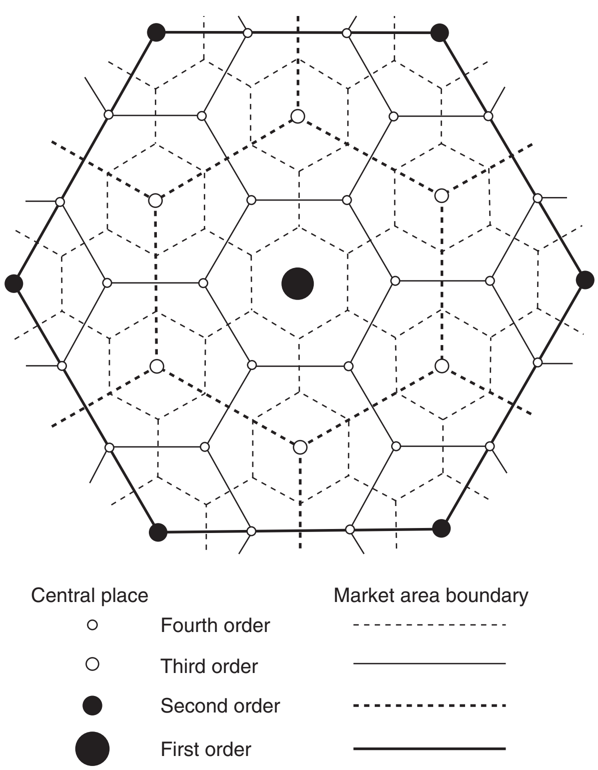

Figure 1 A hierarchical spatial agreement of central places.

areas of the smaller settlements are completely enclosed within the market area of the larger settlement. Efficient administration is the control principle in this hierarchy.

Lösch's Variable-*k* Hierarchies

Christaller's models are based on the assumption that the *k* values, once adopted, would be fixed; that is, they ap plied equally to the relationships between villages and towns, towns and cities, cities and metropolitan places, and so on, through all the tiers of the central place hierarchy. Lösch used a similar hexagonal unit for his central place model but he improved and extended Christaller's form. The major difference between the two approaches was that Lösch regarded the fixed *k* as sumption as a special limiting case. Lösch developed variable *k* hierarchies by allowing various hexagonal systems to coexist. This was achieved by superimposing the hexagonal network for one level of central places upon the network for another and then rotating the networks until the maximum degree of spatial association of central places was obtained. The reason for doing this was to minimize the route distances necessary to link the settlement pattern; as a result, it increased the efficiency of the spatial organization. In this manner, Lösch gen erated a pattern of sectors radiating from the largest urban center (the highest order central place) in a given region. The Löschian landscape is composed of city rich sectors alternating with sectors containing few urban places. This tendency of the urban system to form a sort of radial banding around the highest order central place is in reality reinforced by transportation patterns.

While in Christaller's model hexagonal market areas are arranged in a nested pattern with the market areas of lower order goods packed within those of higher order goods, in Lösch's model there is a much more complex web of interconnected market areas. Specifically, the Löschian spatial organization is characterized by a less distinct hierarchical structure than that proposed by Christaller because (1) central places of the same size do not necessarily perform the same range of func tions, (2) higher order central places need not have all of the functions of lower order centers, and (3) the city size distribution is more likely to follow the rank size rule.

Impacts on Urban Geography

There has been considerable support both for and against the overall effectiveness of central place theory in de scribing and explaining the locational relationships of central places and their market areas. On the one hand, the theory has been criticized for its emphasis on de scription rather than explanation. The classical models are extremely rigid and deterministic. They predict a partial equilibrium under a constant set of conditions. The models describe an essentially static set of locational relationships with no reference to how the spatial pattern might change over time. The social and cultural en vironment of the relationships is taken for granted making them independent of the political and social forces operating at various spatial scales.

In contrast to these limitations, central place theory has stimulated research in urban geography and related disciplines such as retailing/marketing geography by providing a framework for empirical and applied studies. The broader notions of the theory, particularly the hierarchy of functions and centers, the interdependence of settlements and market areas, threshold population, and spatial competition, have given extremely useful terms of reference and organizational concepts within which to examine systematic regularities in the size, number, and spacing of settlements, retailing centers, and public service facilities. A great deal of research has been built on these core ideas. The vast literature which has grown up around the classical central place theory may be broadly classified into two categories: (1) theoretical studies concerned primarily with evaluating and modi fying the various postulates of the theory and (2) em pirical and applied studies which have been more concerned with utilizing elements of the theory for de scriptive and planning purposes.

Modifications of the Classical Models

There is a considerable amount of research on theoretical refinements of the classical central place models. Ac cording to Berry *et al.*, this research can be subdivided into two groups of approaches aiming at: (1) extensions of the classical models, and (2) development of alternative approaches. The former studies attempt to modify the traditional models without changing their underlying assumptions, while the latter intend to make the classical models more realistic and to increase their theoretical generality.

Subsequent to Christaller and Lösch, considerable efforts have been made to modify and extend the classical central place models, particularly with regard to their nature and properties. These efforts have been directed toward developing more general hierarchical models, refining the spatial equilibrium (market area structures), examining the size and configuration of centers in a central place system. In addition, some attempts have been made to integrate the central place theory with fundamental concepts of urban geography such as the rank size rule. The rank size rule indicates that central places, at least in their population sizes, follow a kind of gradient rather than being arranged into a hierarchy. Beckmann resolved this inconsistency by suggesting that each level of the urban place hierarchy might be subject to random influences. If this provision for population variation is added to central place theory then the pat terns of center sizes would approximate the rank size distribution.

To make the classical models more realistic a number of attempts have been undertaken to scrutinize the central place system under various conditions including: a nonuniform population (or demand) density, spatial variation in the incomes of consumers, interdependencies between levels of urban hierarchy, price and transporta tion cost differentials, and variations in consumer be haviors. The alternative approaches to the classic central place models are mostly of a partial equilibrium nature, attempting to describe and explain some aspects of the central place system while disregarding others. One of the main problems with these approaches is that the economic rationality behind actions of individual con sumers and firms remains blurred. Researchers engaged in developing so called new economic geography have addressed this problem. These 'new' approaches propose a framework that adds dynamic factors to the classical models and incorporates the behaviors of individual consumers and firms into the central place framework. Krugman demonstrated that with some adjustments, the static central place models could be made sufficiently dynamic to explain the development of a hierarchical urban system. Fujita and his colleagues provided further insight into the nature of urban systems. They suggested that in a region with multiple industries that differ in terms of scale economies and/or transport costs, the urban system self organizes into a highly regular hier archical system along with a gradually increasing popu lation size of the region. This finding supports Christaller's central place hierarchy.

Empirical and Applied Studies

The theories outlined in the works of Christaller and Lösch have generated a large volume of literature re porting on empirical research into the spatial regularities suggested by the central place models. These empirical studies, however, have only partially been successful in demonstrating the existence of urban hierarchy. Although the theory has been verified by empirical evidence de rived from many regions of the world, the empirical studies have tended to focus on regions with specific characteristics. They have concentrated on rural regions with relatively homogeneous natural and cultural en vironments (such as the United States Midwest, the flat Gangetic Plain of India, Szechwan in China, and Sas katchewan in Canada) and therefore have failed to dis tinguish a complete range of upper order settlements such as national metropolitan centers and global level cities. Christaller's original hierarchy as established in southern Germany included seven orders of central places. There is some evidence to show that Christaller's hierarchy of urban places has changed in response to the process of globalization and development of tele communication technologies. Hall suggests that the two bottom levels of Christaller's hierarchy (i.e., market hamlets and township centers) have gradually dis appeared from the hierarchy. At the same time, a global and a subglobal level have been added at the top of the hierarchy of urban places.

Interestingly, most empirical studies have been con cerned with the distribution of consumer services and there has been little or no reference to the information dependent producer services. The economic landscape as prescribed by the central place models has changed due to the development of transport and telecommunication technologies. The technological advances have eroded the frictional effect of distance on consumer behavior, and consequently there has been a major shift in the organization of urban systems from spatial structures in which central places are graded with multilevel hierarchy (according to the models of Christaller and Lösch) toward systems shaped by the exchange of specialized information. Graham and Marvin suggest that a 'hub and spoke' urban network interconnected by fast transport infrastructures and telecommunication systems has gradually replaced the multi level functional hierarchy of central places. This claim echoes Pred's work, who argues that along with the structural changes of advanced economies the urban–hinterland relationships have be come more horizontal in terms of the hierarchy of central places, and they have become global in scale.

The classical central place models have had an im portant impact on spatial diffusion research. There have been a number of theoretical and empirical studies at tempting to link spatial diffusion and central place the ory. The results of these empirical studies suggest that the relations between central place hierarchy and spatial diffusions follow a two stage process: (1) a 'filtering down' of innovations through the urban hierarchy and (2) a contagious 'spreading out' from urban centers to other locations in their hinterland. Thus, the innovation po tential of a central place is related to its position in the urban hierarchy and the amount of interaction it has with centers that have already adopted the innovation.

The usefulness of central place models is supported by the successful applications of central place approaches for analyzing the locational patterns of the private and public sector services within urban areas. The most common and perhaps the most successful applications have been studies of the spatial patterns of retail activities. The classical studies are those by Berry and his colleagues. These studies confirm the applicability of the central place concept for analyzing the intra urban patterns of business centers. For example, the principles of central place theory have been applied to Chicago and its sub urbs, where a systematic hierarchy of centers and sub centers (albeit strongly influenced by the existence of major radial transportation lines) was identified. More recently, central place models have provided a foundation for a geographic information system (GIS) based analysis of the spatial patterns of marketing and retailing and for an empirical testing of the hierarchy of central places.

Central place models are considered to be useful descriptive tools, with some prescriptive value for re gional/urban planning schemes. The main stimulus for using central place theory in developing regional plans has been related to the efficiency of central place ar rangements that may have a positive effect upon the economic development of a given region. Central place theory has been successfully applied to regional plans in various regions and countries (e.g., USA, Canada, Europe, India, and Middle East). For example, the spatial pattern of new settlements on the drained polders of Holland was planned around the postulates of central place theory. Similarly, Israeli settelements on the Lakhish plains were developed according to a three level hierarchy.

See also: Christaller, W.; Diffusion; Hub Network Location; Location Theory; Networks, Urban; Quantitative Revolution; Retail Geographies.

Further Reading

Beaumont, J. R. (1987). Location allocation models and central place theory. In Ghosh, A. & Rushton, G. (eds.) *Spatial Analysis and Location Allocation Models*, pp 21 54. New York: Van Nostrand Reinhold.

Beavon, K. S. O. (1977). *Central Place Theory: A Reinterpretation*. London: Longman.

Beckmann, M. J. (1968). *Location Theory*. New York: Random House.

Berry, B. J. L. (1967). *Geography of Marketing Centers and Retail Distribution*. Englewood Cliffs, NJ: Prentice Hall.

Berry, B. J. L. (1971). Hierarchical diffusion: The basis of developmental filtering and spread in a system of growth centers. In Hansen, N. M. (ed.) *Growth Centers in Regional Economic Development*, pp 108 138. New York: The Free Press.

Berry, B. J. L., Parr, J. B., Epstein, B. J., Ghosh, A. and Smith, R. H. T. (1988). *Market Centers and Retail Location: Theory and Applications*. Englewood Cliffs, NJ: Prentice Hall.

Christaller, W. (1933). *Central Places in Southern Germany*, Baskin, C. (trans. 1966). Englewood Cliffs, NJ: Prentice Hall.

Davies, W. K. D. (1967). Centrality and the central place hierarchy. *Urban Studies* 4, 61 79.

Fujita, F., Krugman, P. and Mori, T. (1999). On the evolution of hierarchical urban systems. *European Economic Review* 43, 209 251.

Graham, S. and Marvin, S. (1996). *Telecommunications and the City: Electronic Spaces, Urban Places*. London: Routledge.

Greene, R. P. and Pick, J. B. (2006). *Exploring the Urban Community: A GIS Approach*. Upper Saddle River, NJ: Pearson Prentice Hall.

Hall, P. (1998). *Cities in Civilization: Culture, Technology and Urban Order*. London: Weidenfeld and Nicolson.

Hundson, J. C. (1969). Diffusion in a central place system. *Geographical Analysis* 1, 45 58.

King, L. J. (1984). *Central Place Theory*. Beverly Hills, CA: Sage Publications.

Krugman, P. R. (1995). *Development, Geography, and Economic Theory*. Cambridge, MA: MIT Press.

Losch, A. (1940). *The Economics of Location*, Woglom, W. H. and Stolper, W. F. (trans. 1954). New Haven, CT: Yale University Press.

Pred, A. R. (1977). *City System in Advanced Economies: Past Growth, Present Processes and Future Development Options*. London: Hutchinson.

Ullman, E. (1941). A theory of location for cities. *American Journal of Sociology* 46, 835 864.

Chaos and Complexity

G. Chapman, Lancaster University, Lancaster, UK

Glossary

Chaos The deterministic behavior of a dynamic system in which no system state is ever repeated. In terms of the phase space, this means that no point is ever revisited.

Complexity There is no agreed definition of complexity. Most definitions start with a minimum requirement that many parts are involved in nonlinear interactions, which create a whole with emergent properties. Other authors insist that the parts have a cognitive model of their role and position in the system: but since these cognitive models cannot claim to complete knowledge either of themselves or of the system, the partiality of knowledge is the reason that the characteristic behavior of the whole is seen to be emergent. For this reason, some authors insist that complexity is a property of biological and social systems rather than of physical or chemical systems.

Nonlinearity This occurs if change in one variable does not cause a proportionate constant change in a dependent variable.

Strange Attractor A region of phase space, described by a fractal; containing the path of a chaotic system.

Chaos and Complexity

There is a reasonable consensus on what the study of chaotic systems means, but there is very little consensus on what the study of complexity entails – it is a vaguely defined emerging field that is currently an unwieldy al liance of problems, aspirations, insights, and technologies. Most definitions start with a minimum requirement that many parts are involved in nonlinear interactions which create a whole with emergent properties. These prop erties are 'emergent' because they cannot be predicted from what is known about the parts and their relation ships. Thus far, such systems can be modeled on com puters, so that we can find out what actually does emerge from a known set of interactions, even if we cannot predict it. As soon as we introduce the idea that the parts should have agency, we enter a whole new minefield, because we cannot even know what are the relevant properties of the parts.

Without computers, systems theorists were able to make only small advances in understanding fluctuating system behavior. Volterra derived equations for the interaction of two species in an ecosystem, using non linear functions. The classic example is the relationship between basking sharks and shrimps. If the abundance of shrimps is low, the sharks starve, and there is a population crash. The shrimps can undergo a population explosion, since sharks reproduce slowly. But when the sharks do increase, shrimps decrease at an accelerating rate. The relationships are nonlinear, that is, the rate of change of one variable with respect to the other is not constant. The system results in 'repeated' fluctuations in the two species in time (**Figure 1**), which is indicated by the closed loop which shows the system in phase space (**Figure 2**). This behavior is often called dynamic equilibrium.

When small computers became available, systems theorists could run simulations of systems with more components. Although there had been mathematical in timations more than a century ago, it was not until these simulations were run that theorists finally understood there were classes of systems that never went to equilibrium, and never repeated a previous state. One of the first and a classic analysis of this was by the me teorologist Lorenz, who modeled the Earth's climate with three variables related to each other by nonlinear func tions. **Figure 3** shows how these three variables might change through time, and **Figure 4** shows the behavior in phase space. Since the system revisits the same neigh borhoods in space, we can say the path of the system has a kind of shape – here a shape with two lobes. This shape is known as a strange attractor. In this case, the system stays in one neighborhood for a while, then flips to the other ;neighborhood. The strange attractor actually exists in three dimensions, one for each variable. However, on two dimensional paper the diagram is also usually drawn in two dimensions, and it is important to realize

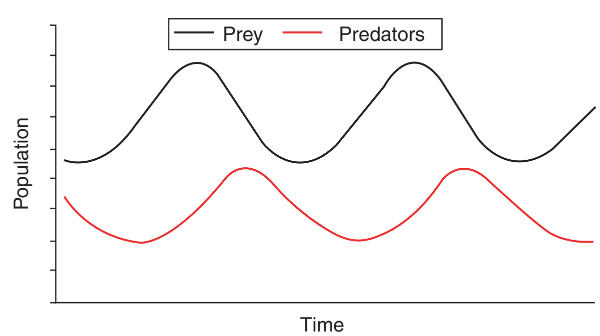

Figure 1 Lagged fluctuations in the populations of shrimps and sharks, Volterra Lotka equations.

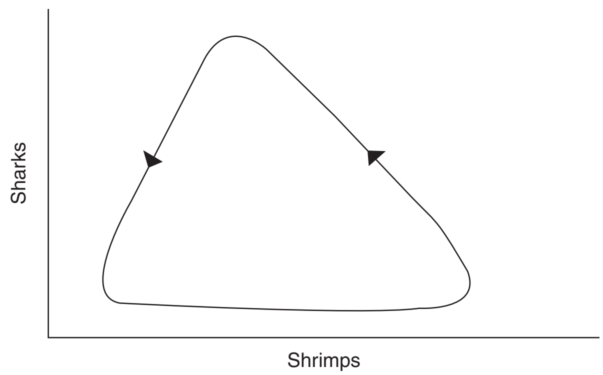

Figure 2 Figure 1 represented in a phase space.

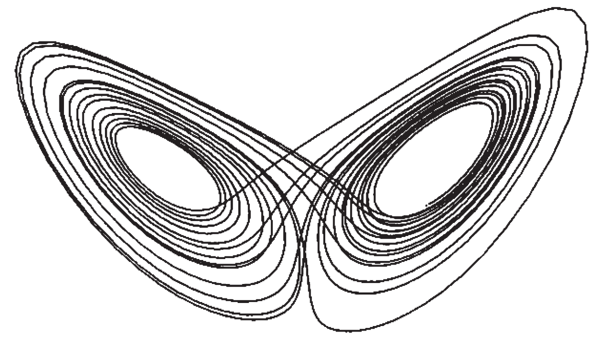

Figure 4 The Lorenz attractor.

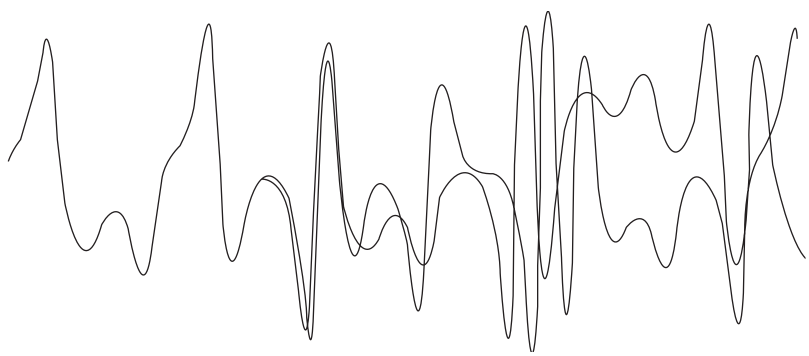

Figure 3 The behavior of a chaotic variable in time.

the implications of this. These systems are deterministic – they have no random component. For any given current state, there is one and only one next state. This means that from any point in the phase space there can be one and only one line segment going to the next point. Hence, wherever in **Figure 4** it appears that the lines touch and cross each other, they do not. They are separated in the third dimension, and pass either behind or in front of each other.

The trajectory of some systems can be modeled as a number, where each digit follows deterministically from the digit before. A system that goes to a stable end state might be something like this: 78265719746538434322222222222222. A system that reaches a dynamic equilibrium may be something like 7836591374652534141414141414141. With chaotic behavior the number is essentially random, in that there is no detectable pattern in it, and it never goes to a terminal state. This understanding leads to the use of the apparently paradoxical term – deterministic randomness. It highlights the fact that the systems are deterministic, but that there is no way we can predict the value of the *n*th digit, without calculating all preceding $n-1$ steps. The lack of predictability is a feature of randomness, but this kind of randomness does not come from flipping a coin. We will return to this feature in a discussion of measures of complexity, below.

So, we can summarize the main characteristics of chaotic behavior so far identified as: the system is deterministic, the system never repeats itself, and the system has a strange attractor. Now, consider the two lobes of **Figure 4**. Suppose one lobe represents ice age Earth, and the other, thermal period Earth. At some stage in the future, which state will the Earth be in? If the future we are thinking about is tomorrow (one iteration), then we have not much difficulty in saying tomorrow will be much like today. The only way we can find out what it will be like in 300 iterations, is to calculate all 300 steps. The future of the system is computationally irreducible. It is not that the system exhibits 'counterintuitive behavior'; the behavior has no intuitive characteristics. Further, if we change any of the initial conditions now by even the smallest amount, our prediction of tomorrow will change by an almost equally small amount, but the system in 300 iterations may be in a completely different state from the one calculated before. These systems are infinitely sensitive to initial conditions. This has been paraphrased as the 'butterfly effect': that a butterfly flapping its wings in West Africa may or may not set off dynamics that culminate in a hurricane in the Caribbean a month later. In history, the effect has been mythologized as: for want of a nail, a shoe was lost; for want of a shoe, a horse was lost; for want of a horse, a king was lost, and the kingdom too.

So, chaos theory is poorly named. It is not about randomness, it is about deterministic systems that are computationally irreducible. Further, it is about the im possibility of ever accurately mapping the real world onto such a system model. To do so, we would have to have an infinitely fine grained observation of the real world, and an infinitely large model. Increasing theore tical understanding of this problem has made us realize that we will never be able to predict the weather of northern temperate areas for more than a week, or two, at the most, by calculating forward from current system states. The present limitation of most weather forecasts to 5 days is because we are already nearing that limit. (This does not preclude some seasonal forecasting based on the historical understanding of repetitive behavior – as with El Nino or the North Atlantic Oscillation.)

Yet, although at first glance it might seem that we now 'know we cannot predict' the future, systems theorists have studied many different kinds of attractors, and to some extent can narrow down the range of possible futures.

Complexity

There may be a degree of agreement on the starting points for the study of complexity. But from then on, complexity becomes a circus big top for any number of performing intellectual animals, in an increasing number of circus rings.

The behavior of chaotic systems is a intuitive – that is, we cannot guess what they will do, we have to wait and see what they do. Complexity is one useful step beyond such a intuitive behavior. The dynamics of the system may well evoke in us some intuitive recognition of holistic behavior – as, for example, when a shoal of fish maintains a coherent pattern, but changes it to avoid a predator, or when a flock of birds seem to behave as a collective individual, forming and re forming a collective V shape as they fly on their migration routes. In these circumstances, we cannot pre dict exactly how the shoal or the flock will behave, but if we observe the same species, we begin to see the kinds of behavior that might occur. The same may be said of our understanding of riots in a football ground – we may know more about the pre conditions that might provoke or prevent a riot, but not be able to predict exactly when and in which part of the crowd it will start.

Increasing attention is now being paid to networks and emergent behavior. Networks may be richly connected, thinly connected, hierarchically connected, or of any combination. Increasingly we realize that the patterns of connectivity can have a significant effect on whether disruptions dissipate or propagate – as an example, it is clear that different patterns of road structure and con nectivity affect the way jams develop and propagate, as any motorist on the UK's motorway network will know. Another example is a power grid in a large city, which collapses unpredictably when a minor fault trips a se quence of more significant faults. Put simply, we are only just beginning to develop the tools to understand how complex dynamics develop on different network structures.

The examples of fishes, birds, and humans were bio logical (in that nervous systems were involved): the ex ample of the power system was physical. Some complexity theorists would claim that the two kinds of cases had the same basis, and that the distinction between biological and physical was 'old time disciplinarity' compared with the new interdisciplinarity. Others would claim that they are radically different, and cannot be fused. This dispute lies at the heart of our understanding of what is meant by 'science', and what is meant by 'being human'. In classical science the future is only an outcome of the present: the only questions that can be asked are 'how' questions, not 'why' questions. On the other hand, human beings believe they have agency, and can choose a future.

Both camps in complexity theory play with the idea of 'emergence'. This says that when some components are combined, there are properties of the whole that are not properties of the parts. At one level, this is trivially true (although still mysterious). Two gases, oxygen and hydrogen, combine together to make a compound, water, which has unique properties unlike those of either par ent: its solid phase (ice) is less dense than its liquid phase, and hence ice floats. At other levels, it is nontrivially true. Time and again one can read in the literature that the most complex entity in the known universe is a human brain, and human beings have self consciousness. Where does this self consciousness come from? The only answer so far (if one accepts the question as meaningful) is that when a sufficient number of neurons are connected together (and it is a number in billions) consciousness 'emerges' as an effect which is not attributable to the properties of the neurons and synapses of the brain. That is to say, given enough quantity, quality changes.

Those who claim that the shoal of fish and the power system can belong to the same class of complexity are basing their claim on this idea that complexity emerges as a result of quantity – of elements, and of connections between elements. The problem is that with conscious ness comes agency. For 'humanists' choosing a future implies that humans can imagine a model of a future, and that choice between them is purposive and not random. How do you explain behavior simultaneously as an 'outcome' and an 'intention'?

Cellular Automata

Cellular automata have been seized on by enthusiasts since computers have become ubiquitous to explore many as pects of what is claimed to be emergent behavior. In their simplest form, these are large 'boards' of many squares in a rectangular grid. The cells maybe either black or white. Any starting pattern of black and white may be used, and in any given simulation run, a set of transition rules is used to determine how a square may or may not change from black to white, or vice versa, in any one iteration.

If corner neighbors are counted, any square can have eight neighbors. For all unique combinations of black/white states for these nine cells, a transition rule specifies what the state of the central cell will be after the current iteration. There are therefore $2^9 = 512$ black and white combinations of the nine squares overall that will give an output state of either black or white for the target square. For many starting states and for many transition rules, patterns simply go to some stagnant and possibly monotone state. However, there are examples where dynamic patterns emerge which seem to behave 'organically'. An exemplar of this is Conway's famous Game of Life. The reader is urged to visit the website for this game, to see many of the 'life forms' which emerge. The neurophilosopher Daniel Dennett used this game as a starting point for his attempt to show that there could be a solution to the 'outcome/intention' paradox in a book which has a similarly contradictory title: *Freedom Evolves*.

Just as systems that go to equilibrium or dynamic equilibrium are not of interest to the chaos theorist, so also the complexity theorist is not interested in those simulations which reach a static terminal state. By contrast, the others have many significant and interesting properties. They are computationally irreducible – only by running it 1000 iterations can you know what state it will be in after 1000 iterations. Each cell is a 'self' with a boundary, and it has 'knowledge' of its own state and that of its immediate neighbors – so it is in an environment of which it has some knowledge, but not total knowledge. Its future states depend on how it reacts to its knowledge. Dynamic patterns self organize and higher scale (multicell) entities emerge, which have characteristic behavior patterns. Indeed this allure is so strong that there is an international club of people with their own vocabulary describing 'guns' and 'breeders' and 'gliders' that they have developed with their own sets of transition rules (**Figure 5**). 'Guns' are static shapes that fire 'projectiles' at regular intervals in straight directions; 'breeders' are shapes that produce new versions of themselves; 'gliders' are shapes that maintain their shape as they drift across the board.

Apparently this is a perfect metaphor for emergence, but actually it is not. The behavior of the whole is explicable by the behavior of the parts. Although it may appear that a 'breeder' is an organic whole, it has no self, no boundary to self, and no central knowledge processing. All of that still occurs at the level of the individual cells. The 'organisms' do not evolve – they might get eaten up or disintegrate, but the stable shapes that wander round the board do not randomly mutate and

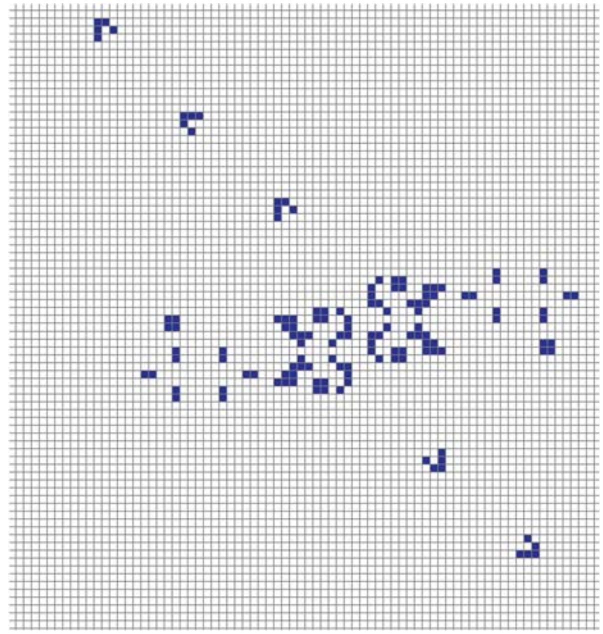

Figure 5 Breeders and gliders in Conway's Game of Life.

form new species. The simulations might however be good metaphors for the shoals of fish – there is no central processing there either, although an onlooker can sometimes get the sense that there could be. The behavior of the whole shoal may affect the survival of the species of fish, but this is an outcome, not an intention.

It is the human observer who imbues the fascinating patterns of Conway's 'game' with the dignity of 'life'. These 'creatures' lead frictionless lives in which they consume no negentropy, nor do they produce entropy. Notwithstanding this criticism, enthusiasts ardently pursue the field of artificial life (cyberlife), of which this simulation is a founding cornerstone.

Separately, there is no doubt considerable value in using these models to throw light on some kinds of complex dynamics. A favourite example is the modeling of traffic flows on motorways, where traffic can accelerate, then decelerate, and bunch up, often forming standing waves at points where there may have been an incident hours earlier. Running a model in which simple rules relate how cars adjust to each other in terms of spacing and speed can mimic these effects. In other complex networks like the power supply to the London Underground, if circumstances can be modeled which mimic a real world breakdown, then that does have a real world equivalent that we can construe as 'emergent', and one that we can usefully know more about.

The Big-Top of Complexity

Many of the topics which now vie for attention within the complexity circus had their origin somewhat outside of the new field, or before the new tent was erected. Ex amples include artificial intelligence, which itself comes in a variety of incommensurate forms. These include expert systems, which quarry human experts such as doctors diagnosing a patient, to elicit the rules under lying their decision making. Others may depend on ar tificial neural nets (ANNs). This is a development in computing which explicitly uses a simple model of data processing in brains. Neurons are connected in nets, and 'fire' each other off, or not, according to specified thresholds. They 'learn' to replicate the same relation ships between input data and the diagnoses made. The problem with the latter is that the relationships that link data with output cannot be isolated from the network: they are 'black boxes', in a sense truly emergent. It is then difficult to build confidence that the network will cor rectly identify novel cases. An apocryphal anecdote re lates that the US Army trained a neural net to distinguish between Soviet and US tanks by displaying thousands of photographs of each. When it came to using the net in the field to distinguish at high speed between friendly and enemy tanks, all tanks were found to be friendly. It turned out that the net had originally distinguished the two types of tanks by whether the photography was clear (a US tank) or grainy (a surreptitious distant shot smuggled out of the Eastern Bloc). All real time images were clear, therefore all tanks were friendly. Not with standing these limitations, ANNs have been used in geography in a variety of ways, though usually connected with physical and biogeography rather than human geography – for example, to classify ecosystems from remotely sensed satellite imagery.

A similar 'blind' and 'self organizing' tool is genetic programming, where thousands of arbitrary rules are assigned to solve a task, and those that are most suc cessful in any generation are selected for the next iter ation. The programs often involve random mating (swapping halves of code) and other variation as well. A program might be designed to learn how to play a card game like poker, or be applied to the prediction of energy use in a city.

Another popular development within the big top at the moment is the use of agent based modeling (ABMs). These are in principle rather more sophisticated versions of cellular automata. An application within geography is the modeling of cities, seeing how social segregation, or suburban sprawl, may result from interacting agents with differing behavioral rules. Studying the behavior that emerges, and testing new hypotheses and new rules, can help understand urban processes in a way which is rad ically different from old style classical spatial economics. These models can be linked to GIS systems so that the 'board' is a 'real' city. Fun, and useful, though these may be, complexity theory does little to guide the con struction of such ABMs.

Developments in robotics may come within the circus tent, not so much for the sake of applications within industry, as for a better understanding of biomechanical principles. The construction of a mechanical dog which can run with the same gait as a real dog reveals that only some of the simulation needs to be in information pro cessing. Much of the performance of the robot is better built into its own mechanical structures and material balances and tensions, so that the control is stimulating 'inherent response' harmonics of bouncing and stretching.

Perhaps the most far reaching of complexity's current buzz words is 'tipping point'. The implication that there are thresholds beyond which stability becomes instability, or leads to an irreversible, alternative, and undesirable state, has been given political legs, and is used almost regardless of whether there is a well founded model behind the statement or not. So in that sense, complexity has entered the popular imagination as a frightening metaphor. The climate will soon pass a tipping point, immigration will soon pass a tipping point, and the educational system will soon pass a tipping point.

Perhaps complexity's big top needs a little self or ganization and self regulation.

Measuring Complexity

Some things clearly are more complicated than others. By this account, it ought to be possible to provide a quantitative measure of complexity, so that the extent to which 'this' is more complicated than 'that' can be ex pressed accurately. Many authors believe that no measure of complexity has yet been discovered that has any useful meaning. They believe no such measure has ever been related to any other variable in any meaningful way, and that therefore it is useless in terms of theory building. However, there are some measures which attract intel lectual curiosity, and perhaps in the future some signifi cance will be derived from them. The most cited is the Solomonoff–Chaitin measure of algorithmic complexity. This says that the degree of complexity is defined by the shortest possible algorithm that can express the observed behavior. Formally this is defined in terms of the length of a program in a universal Turing machine (the most basic theoretical computer) required to print any given output. To go back to the numerical examples, a single decimal digit like 4 or 6 requires 3.322 bits of information to be expressed. Suppose the behavior of some system is represented by the string .33333333333333333333… one hundred decimal numbers long. To express 100 such numbers will take 332.2 bits. But we can also express the number as $n = 1/3$. This expression requires 2×3.332 bits, plus a small length of program for calculating the division. The number is clearly not very complex. By contrast, a string of 100 numbers from a chaotic system

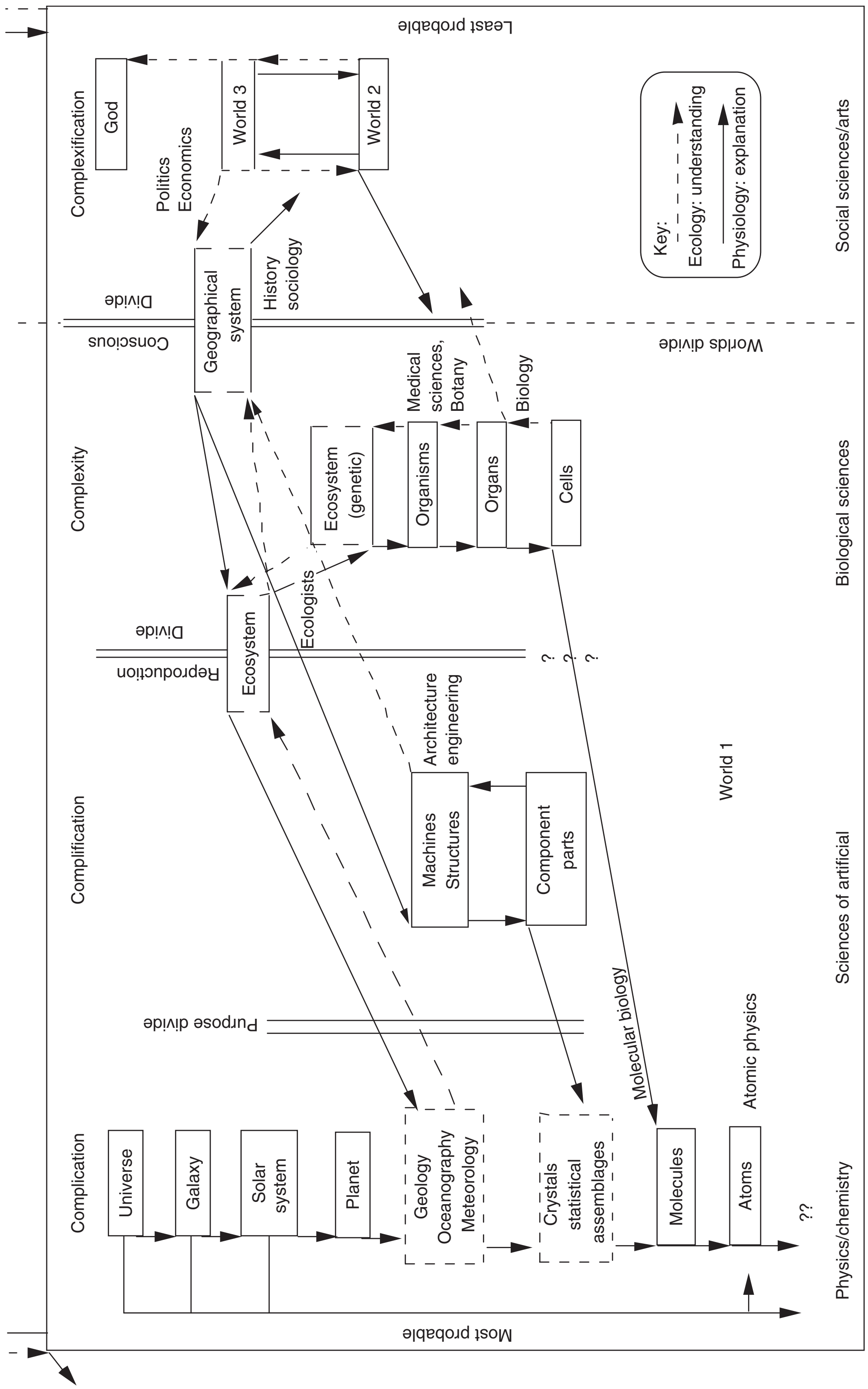

Figure 6 The four qua t es of comp ex ty.

Table 1 The four qualities of complexity[a]

	Complication	*Complification*	*Complexity*	*Complexification*
Understanding	Nil or cosmological	Social and historical context	Environmental	Consciousness (philosophy, ideology, values) Language
Explanation	Nonteleological Reductionist Universal (statistical)	Nonteleological Reductionist Deterministic Certainty Universal Ahistorical Bounded	Nonteleological Holistic (and reductionist in part) Emergent Historical Evolutionary	Teleological Holistic Emergent Reflexive Uniquist Historical Unbounded Uncertain Language
Future states	Calculable prediction	Replicable performance	Dependent on experimental trials	Contested imaginations Probabilistic correlation
Modes of analysis of two examples				
Irrigation	Kms. of channels # of outlets # of regulators Sedimentation Salinity	Integrated designed system	Ecosystem Plant growth Genetic diversity	Political mandate Crop localization/ protective/ Productive Communication Perception Corruption
People	Numbers Required food supply Demographics Malthusia	Robots	Patient of allopathic medicine	Language Personality Ambition Mood Aggression Altruism
Two examples of fields of practitioners				
Government and agriculture	Mouths to feed	Technical irrigation project	Agricultural research into HYV	Representation and redistributive policies
Disciplines	'Hard' sciences	Engineering	Biological sciences	Arts and humanities, reflexive social science

[a] Analytical possibilities of understanding and explanation from Figure 6.

cannot be summarized by an algorithm that is shorter than the 100 numbers, and it has therefore 'complete' complexity. Some fairly simple reasoning leads to the conclusion that the proportion of numbers that are not complex becomes vanishingly small, implying that stable solutions are extremely rare.

In statistical terms, if some sample of numbers are randomly drawn from a perfectly normal population, again we need not know all the numbers; we can use an algorithm to specify the distribution in terms of a few parameters such as mean and standard deviation. We need not know all the details of each number. Apparently complexity has been reduced. The use of statistical an alysis in social science has thus been based on the idea that it is the aggregates, not the individuals, which matter. However, no two nontrivially sized samples will contain the same numbers (hence the idea of the standard error of the sample mean). Now, suppose the behavior of the whole is not an aggregate, but the emergent effect of the individuals interacting with each other. Then clearly, the initial conditions in the two cases are different, even if the sample characteristics are the same. So if we are interested in emergent dynamics, the statistical approach will not work. As an example, suppose people have an individual and static propensity to riot. Then a statistical analysis of a sample of football fans will work out quite easily the likelihood of a riot of intensity x. However, if the propensity of each individual to riot is changed by the behavior of a neighbor (and worst of all it is a positive feedback loop), then what matters most is being able to predict a single trigger point somewhere. Such a trigger point may be the simple adjacency of two people who have poor self control and competing loyalties. Then there are further adjacencies to consider – whether those surrounding the flash point pair are likely to suppress or aggravate the situation. Even the same crowd of people may yield completely different behavior depending on who is adjacent to whom.

There is really only one possible conclusion from this about the state of our knowledge of our own societies. They are completely complex, and all models of them introduce simplifications that render the model predic tions unreliable.

The Complexity of Complexity

There is a determination by many theorists to show that complexity is the new unifying interdisciplinarity – even bridging the science–society gap. But the case is not yet proven, and it may be that there are different types of complexities in different realms. As suggested above al though computation can now produce lifelike simu lations, perhaps complexity is usually associated with living things – bacteria, brains, languages.

In **Figure 6**, this complexity is but one column. There are three other terms – two of a lower order complexity, and a fourth which is of higher order (**Table 1**).

The divisions between the four columns are import ant. On the left in 'complication' we stay firmly within the domain of physics and chemistry, and the realm of re ductionist explanation. The next column 'complification' is the world of artifice – the machines of mankind that obey the laws of physics and chemistry, but which have not been observed to occur spontaneously, and about which systems theory and operations research may have much to say. Although all these machines and buildings perform, in so far as they do, according to nonteleological explanation, their existence is understood in teleological and ecological terms – they are invented and built by man for man's own purposes. They are therefore split from the biological world by the 'purpose' divide. They have an exterior teleology. Mostly (to date) they have the property that they should not behave chaotically. Cars should not wobble erratically, nor should they behave differently according to their individual histories. For their operational success, fractal landscapes are replaced by smooth continuously curved roads.

The divisions between the inanimate world and the world of biology can be made on several grounds. The most immediate one is self reproduction perhaps one definition of life itself. Many people subscribe to the idea that sexual reproduction and genetic mixing is more complex than the dripping tap modeled by a theoretician of chaos. The next division is between the world of physical existence – world 1 – and the worlds of con scious and reflexive thought – worlds 2 (individual thought) and 3 (social epistemology). Note that the brain therefore simultaneously exists as an organ in world 1, and as thinking in world 2. The relationship between physical substance and the emergence of con sciousness was highlighted above. Complexity theory is the product of the human brain, and is part of how we 'choose' to describe some phenomena. Its categorization of nature may or may not prove to be better than dis ciplinary ones currently on offer.

In sum, the interdisciplinary unity of complexity is debatable.

Conclusion

Systems which are perfectly organized (ordered) do not evolve, unless by outside intervention. Systems composed of individuals which exhibit purely random behavior, like molecules of gas in a cylinder, do not evolve, and are best described in terms of their aggregate properties. Poised between randomness and order are other systems. Of biological ones we can state, we the observers, think they have the following properties: They have a large number of elements, because when the number of elements rise, formal description of relations between elements cease to give an understanding of the whole; there are dynamic interactions among this large number of elements, and many of these interactions are nonlinear; there is feed back within the system; they have a history; and every element of the system is ignorant of the behavior of the whole. With regard to even more complex (complex ification) reflexive social systems, none of the mappings between these conditions and the 'real world' is as yet unambiguous.

With reference to human geography, complexity has made an impact as a metaphor, the value of which Thrift (1999) has compared with actor network theory – where again there are many parts not fully aware of the way their behavior is mediated by the network. Harris (2007) has used insights drawn from complexity theory to grapple with the interface between society and planet in the quest for sustainable development. In more applied terms, agent based models are used to study system evolution, for example, with regard to cities by Allen (1997) and Batty (2007), and dynamics on networks, for example, with respect to traffic flows in cities by Tran sims (Los Alamos National Laboratory, USA). These models may provide us with lots of insights into how systems can behave, and how they might behave, but very little by way of predicting their exact behavior. The most certain prediction is that they will be unpredictable, because of sensitivity to initial conditions and bifurcation points.

See also: Cellular Automata; Complexity Theory, Nonlinear Dynamic Spatial Systems; Fractal Analysis; Neural Networks; Simulation.

Further Reading

Allen, P. M. (1997). *Cities and Regions as Self Organizing Systems: Models of Complexity*. Amsterdam: Gordon and Breach.

Allen, P. M. and Torrens, P. M. (eds.) (2005). *Special Edition on Complexity. Futures* 37.

Batty, M. (2007). *Cities and Complexity: Understanding Cities with Cellular Automata, Agent Based Models, and Fractals*. Cambridge, MA: MIT Press.

Casti, J. L. (1993). *Searching for Certainty: What Science Can Know about the Future*. London: Abacus.

Chapman, G. P. (2007). Evolutionary psychology, complexity theory and quantitative social epistemology. *Futures* 39, 1067 1083.

Chesters, G. and Welsh, I. (2006). *Complexity and Social Movements: Multitude at the Edge of Chaos*. London: Routledge.

Conway, G. The Game of Life. http:/psoup.math.wisc.edu/Life32.html

Harris, G. (2007). *Seeking Sustainability in an Age of Complexity*. Cambridge: Cambridge University Press.

Manson, S. M. (2001). Simplifying complexity: A review of complexity theory. *Geoforum* 32, 405 414.

Thrift (1999). 'The place of complexity' *Theory, Culture and Society* 16(3), 31 69.

Relevant Websites

http://transims.tsasa.lanl.gov
Transims, Los Alamos National Laboratory.

Chicago School

D. Sibley, University of Leeds, Leeds, UK

The Chicago School discussed in this article is that University's Department of Sociology, although Chicago has, or had, other celebrated schools, such as The School of Economics and The School of Pragmatism. The term 'school' may suggest a group of scholars working closely together and sharing philosophies and methods, as in Plato's academy. It may also suggest something rather special – a constellation of academics that is leading the way in an academic field, a group whose work inspires others in the discipline and shapes a paradigm for re search. In this sense, there have been a few schools in geography, notably the Berkeley School, centered on the key contributions of Carl Sauer to theorizing landscape, and many years later, the Los Angeles School of post modern urbanism associated with geography and plan ning at the University of California, Los Angeles, and the University of Southern California. The second case is more difficult to argue because individuals contributing to post modern theory are more widely dispersed – we would have to acknowledge contributions from Frederic Jameson at Santa Cruz and non Californian Homi Bhabha at the University of Sussex, for example. Gen erally, it could be suggested that we need to be skeptical about this Platonic and place specific vision of a school. Certainly, in the case of the Chicago School of Sociology, it would be necessary to talk about mythologies, fictions that reflect an external perception of the sociology de partment and have contributed to an uncritical view of the institution's contribution to knowledge. Any detailed examination of the work done in the department, par ticularly during the first 50 years of its existence, dem onstrates that it has accommodated radically different approaches to social research, often at the same time, and that the sociologists have had an uneasy relationship with academics in cognate fields, particularly academics else where in the University of Chicago.

The Chicago School began as a graduate department of sociology in 1892, in a period when the city was growing very rapidly. The department's first chair, Albion Small, had studied both in Germany and the United States and in Chicago he was concerned to promote a systematic, scientific sociology. He was a key figure in establishing sociology as an academic discipline in the United States as first editor of the *American Journal of Sociology*. Small appointed William Thomas, who had also studied philosophy and sociology in Germany and who, with Znaniecki, produced a major study of Polish peas ants in Europe and America between 1918 and 1920. Robert Park accepted an invitation from Thomas to teach in Chicago and, after 6 years in a part time position, he became a professor in 1914. Park's previous experience was diverse. His undergraduate degree at the University of Michigan was in philosophy, he worked as a journalist in New York City, did a master's degree in psychology and philosophy at Harvard, studied philosophy in Berlin under Simmel and in Strassburg under Windelband, and completed a PhD in philosophy at Heidelberg. He then lectured at Harvard, worked as press agent for the Congo Reform Society, and was personal assistant to Booker T. Washington at Tuskegee College, Alabama, until he moved to Chicago. These three – Small, Thomas, and Park – had some common ground, particularly their fa miliarity with Simmel's philosophy gained during their studies in Germany. Ellsworth Faris, appointed in 1919, had a doctorate in psychology and had worked as a missionary in Africa. Ernest Burgess, appointed in the same year as Faris, was the only member of the depart ment in the early period who was trained as a sociologist. Disciplinary boundaries in the humanities and social sciences in the early twentieth century were more fluid than later in the century but philosophical and meth odological diversity was a continuing theme of Chicago sociology. There were two other key figures in the early development of Chicago sociology. These were Louis Wirth, who became a professor in 1926 and was a major contributor to early studies of urban ethnicity, and Herbert Blumer, professor from 1931, who was concerned with relationships between urban ecology and social psychology.

A strong interest in philosophy and a concern with universal theory were fundamental to the development of scientific sociology at Chicago. The department is probably best known for its promotion of urban ecology. Park and Wirth, for example, saw the communities juxtaposed in the industrial city as an expression of an ecological order, a system of competition and temporary equilibrium based on spatial interdependencies. This competitive system, however, develops into a cultural order, whereby mutual understandings emerge from in creased interaction among members of different groups, and then a moral and political order, in which societal norms and goals are articulated, is realized. It is easy to see how this evolutionary system fitted American beliefs in assimilation and the adoption of 'American values' and how it fails to explain conflict, exploitation, and the persistence of groups who do not subscribe to main stream values. Robert Park's introductory essay in *The City*, published in 1925, demonstrates that he was well aware that socio spatial relations were more complex than this model of urban ecology suggests. In a long list of problems that he would include in an urban research agenda, Park identified social unrest, strikes, and vio lence, for example, but he did not provide any clues about how he would deal with them theoretically.

Although there were some differences in the theore tical approaches of members of the faculty – between Park and Burgess, for example – it has been suggested that there was a common philosophical underpinning to research in the School. This was pragmatism, a form of empiricism in which a fundamental proposition is that knowledge acquires meaning and validity on the basis of its usefulness. Theory, then, is inseparable from practice. Of the key figures in the pragmatist group in the late nineteenth and early twentieth century, John Dewey and George Herbert Mead were members of the University of Chicago and a third, William James, had taught Robert Park when he was an undergraduate at the University of Michigan. Mead was a social psychologist and professor in the sociology department from 1894 to 1931. Prag matism provided the philosophical basis for symbolic interactionism, a term coined by Herbert Blumer, who had been a student of Mead. We could think of it as an approach to understanding that involves engagement with others, accepting the pragmatist premise that the acquisition of knowledge requires involvement in the world and the observer's involvement affects outcomes and, therefore, what is known. Peter Jackson and Susan Smith argued that symbolic interactionism chimed with Chicago sociologists' strong interest in social reform but there are other views on their social commitment and involvement that will be turned to later.

The particular urban projects undertaken by the Chicago School during this period have been referred to as urban ethnographies although the level of involvement that is required in modern anthropological research was not evident in most of these studies and there was a common tendency to observe other cultures through an assimilationist lens. Projects included studies of vice (Reckless), the ghetto (Wirth), black families (Frazier), and mental illness (Faris). These were all focused on the city but one of the most innovative and outstanding studies was Thomas and Znaniecki's five volume *The Polish Peasant in Europe and America*, tracing the migration of Polish peasants from the villages of Poland to Chicago parishes, using a range of methods and materials, including oral histories, letters, and newspaper reports. A fundamental weakness of the urban community studies is demonstrated in Zorbaugh's writing on wealth and poverty – *The Gold Coast and the Slum.* Zorbaugh provided a detailed description of life in a relatively poor Persian district, close to the affluent Gold Coast, but interpreted institutions of everyday life, such as coffee shops, as evidence of 'social disorganisation'. Cultural difference was seen as a failure to adapt to progressive, modern American values. Despite such ethnocentricity, we could argue that Chicago sociologists in the early period laid the foundations for modern urban social geography. The way in which the work of the Chicago School entered geographical consciousness, however, was curious.

Human Geography and the Chicago School

The first commentaries on Chicago sociology in geog raphy are found in spatial science texts in the 1960s, with reference only to Burgess. Barry Garner, writing in 1967, referred to the 'classic' concentric zone model of Burgess that described the typical succession of land uses with distance from the city center. Walter Isard, in 1956, and Brian Berry, in 1959, had 'explained' this sequence in terms of the substitution of rents for transport costs. Spatial science in this period was primarily concerned with geometry, although neoclassical economic theory was often harnessed to provide explanations for spatial patterns. David Harvey, writing in 1969, only objected that Chicago was a sample of one so the land use suc cession described by Burgess could not stand as a gen eralization about urban spatial structure. Burgess did have a greater interest in cartography and demography than other Chicago sociologists so the connections be tween his writing and the interests of spatial scientists were closer than with other writers in the Chicago School but it is notable that the principal concerns of the sociologists, in the dynamics of social areas and the re lationship between the social and the spatial, were totally ignored by geographers at this time. It could be that the iconic concentric rings of Burgess's land use model, which became a standard feature of urban geography texts and university and high school curricula, hampered critical thinking about the capitalist city.

A reworking of the social aspects of urban research probably started in the 1970s with Ceri Peach in British geography. He was very enthusiastic about the Chicago School, describing it as 'the fountainhead from which all else flows' and he saw the basic propositions about social and spatial distance as key to an understanding of urban social and spatial segregation. However, there were some more critical reflections on Chicago sociology coming out of the urban ethnographies of Peter Jackson and others in the 1980s, including Jackson and Smith's text, *Exploring Social Geography*. Certainly, acknowledging the research of the early Chicago sociologists contributed to a change in emphasis in urban geography and helped to weaken boundaries between urban geography and similar work being done in sociology and anthropology but there are still questions about the centrality of the Chicago School in urban studies.

Other Chicago's

In the process of locating knowledge, in a particular discipline and a particular place, exclusions occur. This may not be a serious problem if histories and geographies of ideas identify only loosely bound places that are argued to be key nodes in the production of knowledge. However, in the case of the Chicago School, narratives that claim so much for the academics in the sociology department are largely silent about important urban research done elsewhere in the same university and the city. In 1936, the University of Chicago Press published *The Tenements of Chicago, 1908–1935*, edited by Edith Abbott. This collection of essays on housing questions had been preceded by papers on housing questions by Abbott and Sophonisba Breckinridge that appeared in the *American Journal of Sociology* in 1910 and 1911 and these two produced a book on housing problems in the city in 1912. The later work presents a strong contrast to the sociologists' monographs, with their emphasis on social disorganization and assimilation. In a chapter on landlordism, for example, Edith Abbott and Helen Rankin Jeter used Marxist concepts of use and exchange value (without referencing Marx) to explain how landlords contributed to inequalities in the housing market. Similarly, in a chapter on rents, Jeter noted how anticipated land value, taxation, and income from the site, pushed up subsequent valuations without regard for the needs of tenants or their ability to pay, or for the continuing decay of the building. This argument was well supported with evidence from Olcott's Land Values Blue Book of Chicago, which charted annual changes in the geography of land values in the city. The women's view of race and housing was also significantly different to that of the men in the sociology department. Following a paper by Breckinridge in 1913 and earlier research on race by W.E.B. DuBois in 1899, Edith Abbott and Mary Zahrobsky noted how housing submarkets were created by discriminatory practices. In particular, they argued that black families were paying significantly higher rents than white immigrants for comparable accommodation because discrimination against black people limited the supply of housing available to them. Scarcity meant that landlords could postpone repairs and raise rents knowing that an apartment could be let at any time. Another study by a woman, Mary Faith Adams, noted how Mexicans were similarly affected by racism in the housing market. We could compare these analyses of housing submarkets with Park's facile observation (in 1935) that because of the abandonment of middle class housing near the city center as a result of 'white flight', "by a singular turn of fortune, the southern Negro, lately from the 'sticks' – the man politically farthest down – now finds himself living in the centre of a great metropolitan city," making no reference to the subdivision of properties by landlords or their failure to invest.

In addition to what would now be considered incisive analyses of housing problems in Chicago, some of the women also made perceptive observations on the lives of recently arrived immigrants. These included an account of evictions of Italian families by a railroad company, noting the pain of separation from the rest of the community, and the adjustment problems of Croatians who still felt a strong attachment to their rural homeland. There is an awareness of injustice and exploitation that is missing from the urban ethnographies of the sociologists. The political implications of their research are spelled out in one of Edith Abbott's contributions to the book. Noting the problem of low pay and the lack of affordable housing provided by the private market, she makes a case for large scale public housing. It is clear from the essays in *The Tenements of Chicago* that this group of women were somewhere to the left of Robert Park.

Edith Abbott and Sophonisba Breckinridge were both on the faculty of the University of Chicago where Abbott was dean in the School of Social Service Administration so it is not as if the women's research was hidden from the sociologists. In fact, Burgess acknowledged Breckinridge's outstanding contribution to social work, while maintaining a distinction between 'scientific' sociology and 'practical' social work. The failure of the sociologists and Park in particular, to recognize their contribution to urban theory, appears to be a consequence of sexist attitudes and a distaste for the women's politics. One sociology student recalls being immersed in sociology as a science and hearing only occasional rumblings about the old maids downtown who were wet nursing social reformers. The gendering of academic knowledge was exacerbated by the editorial policies of the *American Journal of Sociology*. While Albion Small was editor, the proportion of papers by women fell from 15% to 5% suggesting that American sociology became a more masculine enterprise as the subject established a place in the academy. Park conflated gender and politics, once remarking that the Chicago women were more dangerous than the city's corrupt politicians or gangsters. This observation reflected conservative male anxieties about Hull House settlement and its most prominent member, Jane Addams. Hull House was one of a number of settlements established in large American cities (and in London, Manchester, and Liverpool) in the late nineteenth century to provide support for newly arrived immigrants and the poor more generally. Members of Hull House provided literacy classes and a crèche for working women, as well as undertaking research on housing conditions. Apart from her work in the settlement and elsewhere in the city, Addams was an important figure in the peace movement during World War I and she was prominent in the international women's movement. Edith Abbott and other contributors to *The Tenements of Chicago* had also been involved in Hull House and

Abbott acknowledged the help of Addams' co founder, Julia Lathrop, in the preface. The radical background of Abbott and other members of the Social Service Ad ministration faculty is clearly something that Robert Park found troubling. Until Martin Bulmer's 1984 history of The Chicago School of Sociology, there was neither ac knowledgment of the academic contribution of the women nor any sustained critical commentary on Park and his colleagues. Earlier accounts by geographers, in cluding Peach, Jackson and Smith, and Cater and Jones, all neglected to specify or mention the work of the School of Social Service Administration and its contri bution to urban theory.

Park, Science and Race

Most accounts of the Chicago School understandably present Robert Park as the central figure because he set research agendas and promoted urban ethnography. Al though others did most of the urban research, Park is associated with innovative methodologies and the iden tification of significant topics for research. There is a common assumption that the approach to research pro moted by Park involved 'getting your hands dirty' but there is evidence that Park himself insisted on keeping a distance from the research subject. This comes from his view of science. Showing the influence of Windelband's scientific philosophy during his education in Germany, and Auguste Comte, who distinguished between abstract and concrete science, Park maintained that sociology, including human ecology, was an abstract science, whereas geography and social work were concrete sci ences that provided the facts that sociologists used in their theorizations. According to his notional hierarchy of knowledge, abstract and theoretical sociology was 'higher' than concrete, fact gathering activities, like social work research. This had two implications. First, it allowed Park to devalue social work research and, second, it made it possible for him to argue that the sociologist should try to maintain objectivity by not getting involved with the subject. Interestingly, Burgess did not share this view and he called for the integration of social work and social theory. Park's position on the theoretical status of sociology was crucial in relation to work on race. In 1919, there was a 'race riot' in Chicago (there were similar but not so serious conflicts in Liverpool and Cardiff in the same year) and Park was appointed to the Chicago Commission on Race Relations to investigate the riot. A report on the riot was actually produced by the black sociologist, Charles Johnson, with some help and guid ance from Park. Martin Bulmer noted that it showed a high degree of scientific detachment from what he (Bulmer) called the 'emotional' subject matter. The poor state of race relations at this time encouraged sociology students to get involved but Park advised them that they should approach the problem with detachment, 'like a scientist dissecting a potato bug'. This clearly contradicts the view that Chicago sociologists promoted hands on research. Maybe they did but apparently not on questions that were politically charged. A survey of sociologists connected with Park at this time concluded that there was an overriding concern to distinguish sociology from socialism.

The assumption that the Chicago School in the early twentieth century initiated research on urban ethnicity and socio spatial relations has probably contributed to the neglect of earlier work on race and space by W.E.B. DuBois. In his research on the black population of Philadelphia at the end of the nineteenth century, DuBois, as a black (albeit middle class) academic, rec ognized not only the significance of racism in relation to employment and housing but also the collective strength black migrants gained through affiliation to black chur ches and (enforced) residential clustering. He also put the black experience into a comparative framework by looking briefly at the migration streams of Italian, Jewish, and Irish minorities. This was a landmark work. However, while Park acknowledged DuBois' contribution to an understanding of race in America in a general way, he took nothing in terms of theoretical insights from *The Philadelphia Negro*, published in 1899. Similarly, histories of urban social geography have neglected DuBois' con tribution. Thus, we could conclude that 'the Chicago School', as a collective endeavor in the production of knowledge, has, in part, been defined by omissions.

Conclusion

This article has focused on the early twentieth century because it was the academics in the Department of Soci ology during this period that have had most influence on human geography, initially in the realm of spatial science and later in broader urban social geography. Howard Becker, a member of the 'second Chicago School', has argued that the department was never a 'school' because there was no unifying theoretical or methodological per spective. Although most histories emphasize ethnography, there was also an early quantitative tradition established by William Ogburn in the 1920s and followed by Hauser, Stouffer, and Duncan. Similarly, Burgess favored ques tionnaire surveys and quantitative techniques rather than ethnography. Social geography's Chicago School has been built around Robert Park, who was a complex character. His positive role in developing urban ethnography must be acknowledged but his resistance to ideas coming from elsewhere, notably from the Chicago School of Social Service Administration, has contributed to partial and distorted histories of urban research in the early twen tieth century city.

See also: Berkeley School; Los Angeles School of Post-Modern Urbanism; Social Geography; Society-Space.

Further Reading

Abbott, A. (1999). *Department and Discipline: Chicago Sociology at One Hundred*. Chicago, IL: University of Chicago Press.
Becker, H. (1999). The Chicago School, so called. *Qualitative Sociology* 22(1), 3 12.
Deegan, M. (1981). Early women sociologists and the American Sociological Society: The patterns of exclusion and participation. *American Sociologist* 16, 14 24.
Jackson, P. and Smith, S. (1984). *Exploring Social Geography*. London: George Allen and Unwin.
Sibley, D. (1995). *Geographies of Exclusion*. London: Routledge.

Relevant Websites

http://userpages.umbc.edu/~lutters/1996 SWLNote96 1 Lutters, Ackerman.pdf
This is an introduction to the Chicago School of Sociology by Wayne Lutters and Mark Ackerman.

Child Labor

S. Punch, University of Stirling, Stirling, UK

Glossary

Division of Labor The widespread specialization of tasks and occupations within a society.

Majority World It refers to the developing world, which includes Africa, Asia, Latin America and the Caribbean. The term 'Majority world' reflects that the 'majority' of population, poverty, landmass, and lifestyles exist in the three economically poorest continents of the world.

Minority World It refers to the developed world, which includes the United States of America, Canada, Europe, Japan, New Zealand, and Australia.

Socialization The process whereby individuals come to learn the beliefs and behavior that are seen to be appropriate for a particular society.

Values Central ideals and beliefs that inform people's actions and can be common throughout a society, for example, the values of 'family life'.

Throughout the world, many children play a key role in both productive and reproductive household labor, and may contribute to their household maintenance in both paid and unpaid ways. However, many children's tasks are not recognized as work and can be underestimated or undervalued. For example, domestic labor is considered something that children should automatically carry out without pay. In this sense, child labor can be compared to the invisibility and undervaluation of women's work. It can also be invisible since children's work is often masked as 'training' or 'helping'. Partly, this may be because children's labor is often controlled by women, and is intended to help women with their reproductive tasks or to replace women releasing them for more productive labor. However, care must be taken not to assume that all children's work is within the stereotypical female sphere of domestic work and childcare.

Children carry out a wide range of unpaid and paid work in diverse contexts, such as at home, on farms, in shops, restaurants, factories, and in a variety of jobs in the tourist industry. Sometimes the work they do is similar to that of adults, and at other times it is adapted to their age and particular competencies. For example, children's division of labor within households may be divided according to their gender, age, birth order, and sibling composition.

Since the 1990s, an increasing number of studies have examined the nature of child labor from children's own perspectives. This is partly because of the growth of the new social studies of childhood which views children as social actors whose own perspectives should be sought when constructing knowledge about their daily lives. In relation to child labor, this has also led to a recognition of the importance of children's work during their childhoods in the present rather than only as having socialization value for their future lives as adults. Most of the literature on children and work can be placed within three broad categories: definitions of child work/labor; policy and legislative debates of protection versus abolition; and the reasons why children work, which include debates on the economic utility of children. Each of these will now be considered.

Definitions: Child Work versus Labor

There has been an ongoing debate about how children's labor is conceptualized: whether it is positive and a form of socialization, or negative and a form of exploitation. This led to a distinction between the definitions used to describe children's activities. 'Child work' was considered to be acceptable, a social good and a form of socialization useful for children's future lives as adults. Whereas 'child labor' was perceived as unacceptable, a social evil and a form of exploitation which could be detrimental for children's futures. A distinction was made between two categories of child work: unpaid family work and paid work outside the family circle. Recent thinking on child work abandons the work/labor dichotomy, and recognizes the complexity of the nature of children's work.

Certain activities are considered more appropriate, less harmful, and even beneficial for child workers. Children working in family enterprises or with close kin, especially in rural areas, tend not to be subject to the same risk of exploitation as those working in labor intensive industries or in an urban environment. For example, Boyden stated that in Peru, domestic service and agriculture are legally considered more appropriate for children because:

> … they involve non market relations of production and are both traditional and fundamental to Peruvian society. … The assumption is that young people involved in non waged activities, recruited into the labour market through kinship networks, or working in family enterprises, are in some way guaranteed protection. Exploitation is seen solely as a function of waged employment in large impersonal concerns. (Boyden, 1988: 199)

However, it is wrong to assume that child work is protected when it occurs within kinship relations.

Exploitation may be more concealed and difficult to accept in family enterprises but that does not render it nonexistent. It is often harder to intervene in cases of child exploitation when the child is dependent physically and emotionally on the exploiter. In Latin America, where fictive kinship relations exist, it can be common for rural children to work for their godparents as domestic servants in return for their upkeep or continuing their schooling. Domestic service is not only one of the main forms of child work but is also highly exploitative, as the child can be vulnerable to psychological, physical, verbal, and sexual abuse. Paradoxically, it is one of the activities considered most appropriate and least harmful for children, often not included in legislation or most lenient on the entry age of the child. Thus, exploitation can be concealed by ideas about reciprocity within kinship relations, which can facilitate a source of cheap labor for the wealthier of the fictive kin. It is important not to assume that kin relations are always based only on reciprocity and mutual support rather than exploitation and oppression.

Similarly, it should not be assumed that wage employment for children is necessarily exploitative. It can have positive benefits as it enables children to earn an income, however small, which can help to support themselves and/or their families. Paid work can stimulate children's personal growth, self esteem, and development, as well as encouraging a sense of family responsibility. By earning their own money, children have access to greater decision making power and autonomy. In contrast, unpaid family work can reinforce children's economic dependence on their parents. Some children prefer paid employment outside the home environment rather than engaging in unpaid family work. For example, children in Indonesia said they preferred to work for low wages in factories rather than for no wages for the family. Some children say they prefer paid work because it is less exploitative and has more future prospects.

Therefore, it is too simplistic to assume that child work in the urban world of commerce and industry is automatically exploitative, whereas child work in traditional occupations of farming and domestic service is a beneficial and essential form of socialization. Such a division became more accepted and accentuated because of the formation of child labor legislation. On the one hand, it tended to refer to the potential negative consequences and effects of child waged employment with non kin. On the other hand, it neglected, ignored, romanticized, or even encouraged nonremunerated work of children carried out within familial contexts. Such a simplistic division of children's work obscured the complexities involved.

It is difficult, therefore, to classify which types of child work are 'exploitative' child labor and which have positive benefits. White suggested that children's work should be viewed as on a continuum, rather than trying to categorize it as one of two extremes. It is argued that work is not just good or bad for children but on a continuum from best to worst, where at one end the experience of work is positive and beneficial, and at the other end, it is negative and destructive. Most child labor falls somewhere in between the two extremes. However, it can be difficult to decide where particular employment should be placed along the continuum, since the majority of work has both positive and negative effects simultaneously, both in the present and for children's futures. Work for children can be seen either positively or negatively:

> For large numbers of children work is an ordeal, a source of suffering and exploitation, and a fundamental abuse of human rights. Often, child labour results in educational deprivation, social disadvantage and poor health and physical development. Yet child work can be an important element in maturation, securing the transition from childhood to adulthood. It can also be essential for family survival. (Bequele and Boyden, 1988: v)

Ways in which children are considered to be 'exploited' vary and depend on one's definition of 'exploitation'. It can be seen in terms of present suffering or as having long term detrimental consequences. It is recognized that children can face many dangers at work, both physical and psychosocial. The physical hazards include the health and safety risks of dangerous working environments and the use of unsafe tools and equipment. The social dangers occur more frequently and include low pay; long hours; lack of legal protection; sexual, physical, or emotional abuse; slave like or socially isolating conditions, and work that is mundane and repetitive. Boyden *et al.* noted that there are major physiological, psychological, and social differences between children and adults which can increase children's susceptibility to such hazards. However, they discussed the difficulty of measuring the actual impact of work on children. This is because of the 'invisibility' or subtlety of some of the effects; differing cultural views of risk and benefit; difficulty of isolating work impact from other causes (such as poverty); and the difficulty of gaining an accurate diagnosis of children's health conditions. Consequently, they argued that:

> …discussion and information about the outcomes of work for children has been confined mainly to the risks rather than to actual impacts. Emphasis usually has been placed on physical and safety issues and on the adverse effects on schooling, while psychosocial effects have been largely ignored. The overall picture that has emerged suggests that in various ways and especially physically children probably are more susceptible to adversity than are adults, although conclusive evidence has yet to be provided. (Boyden *et al.*, 1998: 110)

Child work can also be beneficial both in the present and for the future. In the present, the child's contribution to the survival of the household is important in much of the Majority world where children are encouraged to contribute to household income and not be a financial burden. Work is also important for the children them selves as a source of pride, satisfaction, and self esteem. Work can have a moral value, giving children a sense of efficacy and responsibility. Many children like to feel that they are making useful contributions to their families as it can increase their status as family members, as well as building their confidence. Children's work can also give them access to a wider social network and to consumer goods. It can enhance their personal autonomy and give them a sense of independence and self reliance.

In addition, children develop useful skills which they will need for their future, and sometimes it may be crucial for them to build up their competencies with particular ways of working, such as learning to plough fields correctly. In family enterprises, the place of resi dence is also the place of work where children are pre pared for future economic responsibilities and acquire initial skills. The household and the family can play an important role in preparing children for their future by helping them acquire the necessary skills.

Therefore, children's work can have many benefits for children themselves and their families. These benefits and the subjective value given to work by children, their families, and communities mediate children's vulner ability and enable them to be more resilient. The ad vantages and disadvantages of children's work need to be weighed against each other, while also focusing on the perspectives of children and their family, taking into account their wishes and needs. For example, Woodhead showed that children in Bangladesh, Ethiopia, The Philippines, Guatemala, El Salvador, and Nicaragua demonstrated an ability to reason about which work was best for them. They considered a range of costs and benefits, including relative income, security, safety, haz ard, exploitation, independence, and autonomy. His study also showed that children's perceptions of the benefits of their work, such as enhanced self esteem and sense of responsibility, often outweighed the drawbacks, such as poor working conditions. Therefore, in order to reach an adequate understanding of the nature of children's work one must take into account the long term and short term outcomes, the specific social and cultural environment, the historical and economic context, as well as both adults' and children's perceptions of their own situation.

Legislation: Protection or Abolition?

The UN Convention on the Rights of the Child states that exploitative child labor should be eliminated, but the Convention's recommendations "may not be realistic for all countries, especially those whose economies and educational facilities are insufficiently developed" (Bequele and Myers, 1995: 93). The International Labour Organization (ILO) has been striving to ban child labor since 1919, and the ILO's Minimum Age Convention of 1973 (No. 138) sets out the current principles guiding international child labor. However, recent discussions of policymaking and legislation have shifted from a desire to ban all child labor to providing legislation to protect children from overexploitation. The illegal status of children's work forces it underground, to be denied by governments, employers, parents, and the children themselves, and where it is hidden it is not included in protective legislation. Otherwise, it may be disguised as an 'apprenticeship' or 'training', when in reality such a label is a convenient excuse to pay even lower wages.

However, illegal child labor tends to be confined to industrial or commercial jobs, especially factories, mines, and other hazardous employment. Labor legislation prohibits children under a certain age, usually 12–14 years old, from doing certain jobs which tend to exclude the agricultural and domestic sectors. Even where com prehensive legislation exists, it is rarely enforced in the Majority world. This is because there is a shortage of inspectors, who are badly paid and are often easily bri bed. Bequele and Myers outlined the preventative ap proaches which aim to tackle the underlying social and economic conditions, and rehabilitative approaches which deal with the symptoms of child labor. They highlighted the need to target the most at risk groups which include those working in hazardous situations and very young children. However, they also recognized that "protecting children against a particular hazard, and making their work more tolerable, may encourage them to stay in it." (Bequele and Myers, 1995:156.) There is no disputing that exploitative child labor conditions should be regulated, but the child's view, the context, and the consequences should also be taken into account.

Causes: Why Do Children Work?

Whether children work is not a question of choice for many children or their families. The causes of child labor usually stem from poverty and underdevelopment, but this is not the only reason why children work. They also work as a result of other structural constraints, including the failure of the education system, unemployment and underemployment, vested interests of employers, and rapid rural–urban migration. In addition, children may work because of a lack of parental awareness of the im plications for children's health and development, social and cultural attitudes, and lack of political will for effective action. Also, children may choose to work to help their families, to enhance their independence and competencies, to gain access to consumer luxuries, to gain useful skills, and as a means of self actualization.

On the demand side, employers prefer to employ children as they can do labor intensive tasks for lower wages than adults, they are a good source of casual labor, and tend to be a docile labor force as they are un protected by legislation and workers' rights. On the supply side, whether children work or not tends to de pend on the wealth of the household, the employment status and wage rates of employable adults within the household, the availability and cost of schooling, and the social and cultural environment. Family dynamics and household composition are also cited as influencing whether children work. For example, Boyden *et al.* noted that "in many places, a disproportionate number of working children appear to be from homes headed by a single woman" (1998:138). They observed that family emergencies, such as death or incapacitation of an adult earner, loss of a job, harvest failures, and severe weather may also increase the likelihood of children starting work.

The way in which society views childhood across time and space influences the nature of child work. Cultural expectations not only shape the types of work which children do but also impact upon the meaning and value of their work. Cunningham noted that changing concepts of childhood over time affect the growth and decline of child work. He observed that once childhood in the UK became viewed romantically as a special time of life to be protected, this helped to encourage the separation of children from the adult world of work.

In the Majority world, children are considered as economically valuable to their parents for two main reasons: in the short term because they contribute to the household with either unpaid work or financial contri butions from paid work. In the longer term because children are a source of financial security in old age. The debate on the economic utility of children considers to what extent the economic benefits of having children outweigh their economic rearing costs. Studies exploring the economic contribution of children to peasant households provide contradictory results. For example, on the one hand, some have concluded that children's economic costs outweigh their benefits as they consume more than they produce up until the age of 16. On the other hand, others have suggested that children make a significant economic contribution, and that boys in par ticular have produced more than they have consumed by the time they reach 16 years of age. It has also been argued that children are neither a net asset nor an eco nomic liability to the peasant household: their work makes some contribution to the agricultural household but each child is also an additional mouth to feed. Fur thermore, it has been suggested that the negative effect of nonproducing young children is compensated by the positive effect of the productivity of children when they are older.

Despite contradictory evidence regarding the eco nomic utility of children, it is widely recognized that children are valuable to their parents particularly in rural areas for their work on peasant farms. Although the children's contribution is considered important for family well being in poor households, it is debatable whether it is indispensable. The significance of the economic utility of rural children is that, to a certain extent, parents are dependent on their children both for their unpaid work and for security in old age, thereby reinforcing inter dependent household relations.

It is widely acknowledged that many children in the Majority world are expected to undertake unpaid household labor on a routine basis and may contribute to the survival of the household with paid work. However, it is worth bearing in mind that many children in the Mi nority world may also work, although some argue that their participation in work tends to be more symbolic: for socialization purposes rather than because of economic need. Recent research has shown that children in the Minority world work for a variety of reasons which can be related to opportunities and constraints. For some children, work provides opportunities for increasing fi nancial and social autonomy, developing new social networks, and accessing consumer and leisure cultures. For other children in the Minority world, poverty is the key reason why they work and their work experiences can be negative. They may begin work later than their peers in the Majority world, and they may work shorter hours, but children in the Minority world can also suffer from poor working conditions, such as unsociable hours and poor pay. Employment for children in the Minority world is part time rather than full time but, like children in the Majority world, their experiences of work are extremely diverse, encompassing both positive and negative features and consequences, and is often com bined with school. Despite the complexity and variety of the nature and causes of work for children, one thing is certain: work is a common part of many rural and urban childhoods throughout the world.

See also: Children/Childhood; Livelihoods; Poverty; Vulnerability.

Further Reading

Ansell, N. (2005). Work: Exploiting children, empowering youth? In Ansell, N. (ed.) *Children, Youth and Development*, pp 158 191. London: Routledge.

Bequele, A. and Boyden, J. (eds.) (1988). *Combating Child Labour*. Geneva: International Labour Office.

Bequele, A. and Myers, W. E. (1995). *First Things First in Child Labour: Eliminating Work Detrimental to Children*. Geneva: International Labour Office.

Boyden, J. (1988). National policies and programmes for child workers: Peru. In Bequele, A. & Boyden, J. (eds.) *Combating Child Labour*, pp 195 216. Geneva: International Labour Organisation.
Boyden, J., Ling, B. and Myers, W. (1998). *What Works for Working Children*. Stockholm: Radda Barnen and UNICEF.
Cunningham, H. (1995). *Children and Childhood in Western Society Since 1500*. London: Longman.
Fyfe, A. (1993). *Child Labour: A Guide to Project Design*. Geneva: International Labour Office.
Hobbs, S. and McKechnie, J. (1997). *Child Employment in Britain*. London: Stationary Office.
Miljeteig, P. (1999). *Special Issue on Child Labour Introduction: Understanding Child Labour. Childhood* 6(1).
Mizen, P., Pole, C. and Bolton, A. (eds.) (2001). *Hidden Hands: International Perspectives on Children's Work and Labour.* London: RoutledgeFalmer.
White, B. (1994). Children, work and 'child labour': Changing responses to the employment of children. *Development and Change* 25, 849 878.
White, B. (1996). Globalization and the child labour problem. *Journal of International Development* 8(6), 829 839.
Woodhead, M. (1998). *Children's Perspectives on Their Working Lives: A Participatory Study in Bangladesh, Ethiopia, The Philippines, Guatemala, El Salvador and Nicaragua*. Sweden: Radda Barnen.
Woodhead, M. (1999). Combating child labour: Listen to what the children say. *Childhood* 6(1), 27 49.

Relevant Websites

http://www.antislavery.org
Anti slavery International: Child Labour Programme.
http://www.globalmarch.org
Global March against Child Labour.
http://www.hrw.org
Human Rights Watch: Child Labor.
http://www.ilo.org
International Programme on the Elimination of Child Labour (IPEC).
http://www.unicef.org
The 1997 report of *The state of the World's Children* focuses on child labour and can be downloaded from the UNICEF website.
http://www.worldbank.org
Working paper from the World Bank's Human Capital Development and Operations, World Bank.

Children and Mapping

V. Filippakopoulou and B. Nakos, National Technical University of Athens, Athens, Greece

Introduction

The topic of children and mapping is a special subject within the broad area that examines the ways people see and interpret maps. Children are a special case of map users for two reasons. The first one pertains to the re lation between the development of children's conception of space and that of cartographic understanding. The theories of the children's spatial development, mainly deriving from psychological studies, have provided the theoretical basis for approaching the way children use maps. The second reason is the educational perspective of school maps and atlases. Most elementary or high school textbooks contain a large number of maps, mainly the textbooks related to geography and humanities courses.

Among the various kinds of maps that children are exposed to, school atlases are the most noticeable ex amples, being traditional educational tools that help children acquire spatial knowledge and mapping skills. Their origins are traced in 1697, as part of an atlas published by Louis Courcillon de Dangeau. In 1753, an atlas by the great mathematician Leonard Euler was one of the earliest German atlases explicitly made for use in schools. From the beginning of the nineteenth century, school atlases were systematically produced in Europe and North America. In our technological era, the atlas form has been changed and consists of packages, which are electronic atlases both software and spatial data, and characterized by various degrees of interactivity.

At the beginning of the 1980s, a discussion started on how to approach maps from the point of view of children. It was the beginning of a new perspective into cartographic research, approaching the meaning of maps from the user's viewpoint. The next three decades were productive as far as the theoretical and experi mental work done in this area is concerned. Children and mapping became a research topic in the fields of psych ology, geography, education, and cartography. Children's understanding of maps has been approached from dif ferent theoretical perspectives: the nativist views, Piaget's theory, Vygotsky's theory, and the cognitive perspective. Depending on the theory they are based on, the research and experimental studies in cartography can accordingly be distinguished as nativist, Piagetian, developmental (or neo Piagetian), culturalist, and cognitive. Many psycho logical studies have focused on spatial cognition and mental representation of space, and, as a consequence, maps have been used in these studies as means of ac cessing children's spatial thinking. Research on this topic from a strictly cartographic perspective has been less forthcoming and nonsystematic. So, there is a lot of re search evidence concerning preschool and early pri mary age children's spatial thinking with large scale maps, though the evidence is rather limited on issues such as: children's understanding of cartographic con cepts, the development of spatial thinking with small scale maps, how secondary school children deal with maps, the kind of maps that are more effective for children, the kind of maps children prefer, and the actual contribution of computer mapping to geographic education.

The following paragraphs summarize the theoretical perspectives of cartographic understanding, give some research evidence on the development of understanding spatial representations, refer to the basic characteristics of maps and the evidence as regards the children's de velopment of associated concepts, and, in closing, address the contribution of computer technology in children's dealing with maps.

Development of Cartographic Understanding

The development of cartographic understanding has been approached from four different theoretical per spectives.

The Nativist Approach

The first one, the nativist, posits that cartographic understanding is innate. Three kinds of argument have been advanced for this perspective. The first one comes from a particular view concerning evolutionary devel opment, according to which much of the structure of human mind is innately specified, evolved through nat ural selection during the evolutionary history of the species. The second argument in support of this view is based on conclusions from experimental studies, which claim that children develop mapping abilities at a very early age, before receiving any cartographic education, and also, the mapping abilities expressed by people in early stages of human history. The maps, made 5000 years ago, suggest that all people develop an under standing of the geographical space as well as the ability of representing it through available materials. A third argument comes from a 1986 study of the performance of a 4 year old girl, blind from birth, who could encode a tactile map and navigate between objects in a room. The girl demonstrated an understanding of the correspond ence between the map and real space, even though she had no experience either with the space or with maps. But, it has been argued that this is just one case, while other studies suggest that blind children show delays in spatial abilities.

Of support to this perspective are similar views that have been expressed about the knowledge of language. Darvine in 1871 was the first one to argue that language is an instinct. Almost 100 years later, in 1965, Noam Chomsky, based on linguistic analyses, argued that human language is a biological object, internal to the human mind/brain, and the knowledge of language is individualistic. Chomsky's ideas had a great influence on linguistics and cognitive science. Following Chomsky's views, Steven Pinker argued in 1994 that language is an instinct. Among other scholars, Chomsky's ideas have been also disputed from George Lakoff, who along with Mark Johnson in 1980 argued that human language is not entirely a genetic innovation, but rather central aspects of language arise through evolution from the neural systems that are present in nonhumans. A central point of Lak off's ideas is that the mind is inherently embodied, and thought is mainly unconscious. In 1987, Lakoff expressed his belief that humans' abilities to conceptualize the world are grounded in human–environment interaction. So, he claims, reason is not an essence that separates humans from nonhumans, but it is an essence that places all on a continuum.

There have been several counter arguments to the nativist approach to map understanding. One is that the theory cannot explain why there are many adults having difficulty in map reading. Another is that many re searches concluded that engagement with maps at an early age does not necessarily mean that children can understand cartographic concepts. On the contrary, there is evidence that map understanding progresses slowly and gradually from easy to difficult tasks. The productive result from this perspective is evidence that children can have some engagement with maps at an early age and can be introduced to cartographic concepts, starting from easy to more difficult ones. Neural studies in the first years of twenty first century give evidence for continuity between mental capacities found in humans, such as at tention, memory, and learning, and those of other species of primates, thus giving new potential to the nativist theory.

Piaget's Theory

The second perspective accepts that cartographic under standing follows the stages of intellectual development. Jean Piaget's theory on psychology of intelligence and Jean Piaget's and Barber Inhelder's theories on child's conception of space and child's conception of geometry determined the research on maps and children for almost three decades (from the 1970s and afterward).

Piaget and Inhelder proposed that children's environmental adaptation develops in a sequence of coherent and qualitatively different stages: the sensori motor, the preoperational, the concrete operational, and the formal operational. In the sensori motor stage, chil dren (from about birth to 2 years) do not have any mental function, and only at the end of this stage develop inner representations of the outside environment. At the pre operational stage, children (from 2 to 7 years) cannot focus their thinking on more than one thing. They per ceive space from an egocentric point of view and understand only topological spatial relations. At about age 3–4, they can recognize shape as the first topological spatial relation, also open and closed figures, but they cannot differentiate between closed figures (circle, square). They are not able to form a straight line. By about age 5–6, they begin to discriminate straight and curved lines, sizes of lines and angles (Euclidean prop erties), but, they cannot organize landmarks in an ob jective spatial whole, rotate a plane 180°, describe changes of position, and reconstruct a route in reverse direction. At this stage, there is nonconservation of overall distance, and only at the end of this stage children start to learn how to measure, qualitatively, without unit iteration. At the concrete operational stage, children (from about 7 to 11 years) can use operations (mental processes) which enable them to classify, arrange objects in series, and understand projective properties of space. However, children of about 7–9 years are unable either to coordinate a system as a whole, or to describe a route. At a mean age of 7.5, they show conservation of distance between two objects, and at a mean age of 8.5 they are capable of one dimensional measurements in an oper ational form and empirically discover the two dimen sional measurement. They also show conservation of area. Children of about 9–11 years can coordinate landmarks and changes of position, they can construct a topo graphical schema in a coordinate system, and they have a full appreciation of two dimensional and three dimen sional measurements. Finally, at the formal operational stage (above 11–13 years), children are capable of more abstract and logical thinking, they are able to understand Euclidean properties of space, they establish relations between lines and areas, and they understand volumes of objects in relation to the surrounding space.

At each stage, children's mental abilities are rather stable. The mental structures of knowledge and under standing are developed into new ones because of new experiences and adaptation to new information. In recent years, many psychologists reject the authenticity of

Piaget's theory and mainly the idea of discontinuous stages of development. According to them, many skills seem to develop more gradually and continuously. The distinction of spatial relations in topological, projective, and Euclidean is criticized as not being helpful to an analysis of spatial thinking. The theory is criticized for being focused on the level of the individual, ignoring the important role of the social processes involved in cognitive development. Another point of argument is the implication of Piaget's theory that adults are accurate in spatial understanding, which seems to be wrong.

Piaget's theory had a great influence on the research on children's mapping abilities. The Piagetian approach to cartographic understanding accepts cognition as the basic factor in the development of mapping competence. Children actively construct their knowledge about maps. This construction of knowledge is a continual interaction between children and related experience. This is a fundamental issue which determined both theoretical and experimental work on this topic. Children's perceptual abilities, at each stage, were a guide for the introduction of cartographic concepts gradually and according to the degree of difficulty involved. Piaget's experimental work using graphic representations and spatial models greatly helped predict how children conceptualize space through maps. The results of many experimental studies on what children dealing with maps can understand do agree with Piaget's theory, while other experimental studies suggest that children can do more with maps than what Piaget's theory suggests. In 1989, Lyn Liben and Roger Downs based on the Piagetian approach – although critically – proposed a developmental approach to children's understanding of maps. Based on many experiments in the 1990s, they concluded that map understanding is a complex procedure that develops gradually depending on the cognitive level and experience of the individual, and it has to be approached both from a developmental and a cartographic perspective. They also argue that children show competence in a few mapping activities at an early age, but cartographic competence improves slowly in tasks concerning more advanced spatial thinking.

Vygotsky's Theory

The third perspective considers the social factor as playing a central role in cartographic understanding. Among the psychologists that criticize the individualism of Piaget's theory is Lev Vygotsky. His theory on learning and cognitive development, as expressed in his work on thought and language and mind in society, has offered the theoretical basis in studies on cartographic understanding in recent years. The relation between the biological roots of behavior and the human activities in the social environment is a critical issue in every theory of development. Vygotsky used a key point in his approach of this relation, mainly, the functional system of learning, which differentiates the concepts of learning as defined in other theories. According to him, there are two kinds of concepts: the spontaneous results of everyday experience and the scientific results of systematic school learning. The first ones are concrete and unsystematic, while the second are systematic and hierarchically structured. There is a gap between spontaneous concepts (children's own learning) and scientific concepts (taught by teachers), which is wide in very young children and is gradually bridged after systematic learning. Scientific concepts are easier to learn, and it is better for children to formalize concepts in school with the help of skilled learners before experiencing them unsystematically. So, according to Vygotsky, school must play a central role in cognitive development.

Investigations in 1989 and 1991 demonstrated that guided participation may be important in the development of spatial thinking. In experiments in 1991, children were able to plan routes on a map more effectively when collaborating with adults. In experimental studies in 1998, children of 12–13 years of age working in groups showed a slightly better understanding of cartographic concepts than the ones working individually. In 2000, it was admitted that Vygotskian views have been prominent in research on spatial competence and also in a study in 2003, it was said that collaborative teaching in solving cartographic problems needs further development. But it has also been argued that Vygotsky's theory overemphasizes the role of the social environment in the development of mind.

The Cognitive Perspective

The fourth theoretical perspective to cartographic understanding is based on theories of cognition that developed the last decade of the twentieth century and is growing at an exponential rate. In the context of developmental cognitive neuroscience, studies from various disciplines give evidence about brain processes and resulting mental functions. The use of electrophysiological recording systems to record the human brain opened new horizons toward understanding cognitive development. These brain scanning techniques opened up the possibility to view the brain in action and localize the regions of the brain activated during different activities. Several regions appear to be devoted to spatial thinking, and it is concluded that these regions develop in very early childhood, contrary to what was believed. Based on these findings, suggestions have been expressed for starting cartographic literacy in kindergarten.

In the context of the cognitive perspective, there is another approach to cartographic understanding based on the theory of information processing. Information processing approaches (which first appeared in the 1960s) use models of the human cognitive system based on computer operations, where the hardware is the per ceptual and cognitive human system (eyes and brain), and the algorithms are the mental activities by means of which information is processed (perceived, encoded, stored, and used for problem solving). The major goal of this psychological approach is to describe the nature of thought, how the human mind represents and handles information, and how knowledge is processed and or ganized. The developmental changes are proposed to be gradual and continuous. What children know is not that important. Instead, of importance is how cognitive pro cesses change with age and experience. Children are seen as being active in interactions with the environment, because they perceive objects and events and then re member and draw inferences from them.

The cartographic perspective on the information processing theory is a new approach. The earlier cogni tive models related to cartographic concepts appeared around 1985. Since then, knowledge structures of map related information have been proposed, and in several studies models of cognitive structures have been used in experiments related to children's activities with maps. The strength of the approach is that it offers insight into children's thinking as to how they handle information while using maps. On the other hand, it approaches understanding from the individual point of view, ignoring the social parameters involved in this process.

Recent Views

The aforementioned theoretical approaches to carto graphic understanding differ mainly as to the point they consider as dominant in the development of spatial thinking. This consideration determines not only each theoretical position but also the experimental work on which it is established. As to what approach has to be followed in future research, the prevailing view is that there is no single scientific approach as to how children learn with maps. In order to face the educational map material from a critical point of view and maps as means that produce knowledge, the approach has to be holistic, taking into account the fact that children are individuals who perceive the geographical space and its represen tations through their senses, and at the same time, they are members of a society, inevitably influenced by their social environment as well as by the educational pro cedures. Such an approach to spatial development has been proposed by Nora Newcombe in 2000. She advo cates that her approach: encompasses nativism by con sidering early infancy as the starting point for spatial development but denying that the competencies of in fants are so fundamental to spatial development as na tivists argue, is interactionist without being Piagetian, and it encompasses interactions of the child with the skilled adults but denying their dominant role in devel opment. In 1995, McEahren had proposed a similar holistic approach to map understanding. He considers maps as spatial representations and contends that the concept of representation is fundamental to all ap proaches that can be taken to cartography. In his detailed analysis, he approaches maps at multiple levels: lexical, semiotic, cognitive, and social.

Understanding Spatial Representations

Using Models

According to Piaget, children of about 7 years of age start to appreciate the model as a representation. Experi mental studies in 1991 with children of ages between 2.5 and 3 years found that children appear to appreciate the correspondence between a room and a model at the age of 3 years, having more success when the room is familiar to them and when the model is a photograph or a drawing. Another study in the same year suggests that very young children recognize just unique objects and their representations in the models rather than fully appreciating the model as a representation. Other ex periments in 1997 indicate that children from about the age of 3 are able to use a model to find a location. In 2000, a review study concludes that children after 3 years possess the idea that models are tools of thought, sug gesting that the use of models by children needs initial support.

Using Aerial Photographs

Aerial photographs are considered means of introducing children to spatial representations. Without being ab stract and symbolic representations as maps are, large scale photographs remind one of geographical space. By photo interpretation procedures, children can be intro duced into the concept of looking at the Earth from above. The results of the rather few experimental studies with preschoolers using photographs are contradictory. In 1970, the results of the experiments with 5–7 year old children showed that all children were able to understand that the photograph was the view of a landscape from above, but they could recognize just a few features. In 1980, experiments with 3–5 year old children concluded that they were able to recognize many features on the photos. Contradictory results have been derived from studies with preschoolers in 1991. Very young children seemed to interpret successfully very large scale photo graphs of space familiar to them, as was found in two studies in 2002 and 2003. The results suggested that young children view aerial photographs as a collection of features and not as a representation of an area. In ex periments in 1971 with older children of ages 6–11 years, it was concluded that below age 9, children had problems in recognizing features, and only the older ones of ages 9–11 years seemed able to identify many characteristics. Even then, they were able to identify mainly the familiar ones and the ones whose image in the photograph looks like their view from the ground. However, they find it difficult to identify the area presented on the photo. Difficulties seem to appear when using small scale photographs. An experimental study in 1983 showed that many students aged 16 years could not successfully relate small scale photographs and corresponding maps. The same results were found in another study in 1979, in which high school students aged 15–16 years had difficulties in relating oblique aerial photograph with the map. It has been argued in 2006 that more systematic experimental research is required on this issue.

Using Maps

In the context of the developmental approach to cartographic understanding, the understanding that a map is a spatial representation can be distinguished as accomplished at two levels. At one level, the so called holistic, children understand the relationship between the map as a whole and the whole of the real world space to which it refers. At another level, the so called componential, children understand the symbol–referent relationship for each individual symbol. At the componential level, two things have to be understood: the first one is the geometric correspondence between the position of each individual symbol on the map and the location of the referent feature in the real space. The second is the representational correspondence, which refers to the symbolic representation of feature characteristics as recorded in the legend.

Experimental studies in 1987 and 1996 suggest that children from about the age of 4 understand that maps represent spatial information, but it seems that they do not have a full understanding of the representational correspondence between the map and the represented space. As was found in a study in 1996, children about 8 years old think that the map is only a small scale map used for finding ways and unknown places. Gradually, older children understand that the term map includes many different kinds of spatial representations. At the componential level, there is evidence from experiments in 1979 that children from the age of 3 years can use information from a simple map in order to identify a location. Other experiments in 1989, 1989, and 1996 concluded that kindergarten children can indicate their own location on the map but have difficulty in locating other locations. Moreover, they have significant difficulties in location and orientation tasks when the map is unaligned to the environment that it represents. The orientation task using unaligned maps is difficult, and only children about 10–11 years old succeeded in it. Representational correspondence appears to be achieved by about 6 years of age, as indicated in experiments in 1997. Other studies in 1979 and 1994 give evidence that 4 year olds appear to be able to understand representational correspondence in the case of pictorial symbols. The lack of systematic investigation of early understanding of symbols has been mentioned in many recent literature reviews.

Understanding the Basic Characteristics of Maps

The achievement of cartographic understanding has, as a prerequisite, the understanding of the basic characteristics of maps. Scale, map projections, generalization, and symbolization are common to every map and are considered as basic characteristics of maps. Since map literacy is not systematically included in the school curriculum, there are no definite and clear results referring to children's understanding of these basic map concepts. On the other hand, the results of experimental studies give some evidence as regards children's development of concepts associated with the map's basic characteristics.

Scale

Map scale is the ratio between the dimensions on the map and on those of reality. According to Piaget, understanding scale requires the understanding of proportionality, concepts achieved at the formal operation stage. Experimental studies in 1999 show that children around 3–4 years old are able to encode proportional distance. Other experiments in 1996 show that 7 year old children make substantial errors in map scale tasks. Also, in another study in 1995, 10 year old children perform worse than children of 11–13 years old in map scale tasks. In 2000, a review study of the above experiments suggests that the contradiction of the results leads to a reconsideration of Piaget's view that scale depends on acquisition of proportionality.

Scale is indicated on maps both in numerical and graphical representations. Graphic scale representations are more useful for young children in distance estimation tasks. In maps addressed to children, different graphic scale representations support different learning levels. In a study in 1971, difficulties and misconceptions were identified in tasks involving children's use of graphic scale such as: confusion by the use of different graphic scales in different maps, difficulty to measure long distances as easily as short ones, inability to understand that the outcome of the measurement is a true distance on the ground, and having problems in countries where the units of measurement are different from those on the map scale.

Map Projections

Through map projections, the spherical surface of the Earth is transformed into a plane. A map projection is a mathematical relationship of geographical coordinates (φ, λ) and plane coordinates (x, y). Whatever is the mathematical relationship applied to the transformation process, distortions are inevitable. The distortions of the geometric relationships on the sphere (distances, areas, angles, directions), when projected onto the plane, can be controlled by a suitable choice of projection.

Understanding map projections involves constructing the image of the graticule on the map. Children must be able to visualize the form of the graticule and the way it looks from different perspective views. Having achieved this knowledge, children are able to evaluate projections by comparing the geometry of the graticule as it is on the spherical Earth with the grid on the map. Gradually, they come to understand the effects different transformations have on the representation of landmasses as well as the importance of choosing the best projection for a particular map. Children of around 9–11 years of age (depending on different school curricula) are introduced to the concepts of the spherical Earth, the equator, the meridians and parallels. According to Piaget's views, children are able to use coordinates at about 8 years of age. But experiments in 1965 report that children of 10–14 years old have difficulty in using coordinates in mapping tasks. The use of dynamic representations of the globe and its perspectives from different points of view as well as its transformation to a plane through various kinds of map projections can be a strong educational tool for introducing children to the concepts of map projections.

Generalization

All maps are abstractions of the real world. As the scale gets smaller, the map content gets less. The information portrayed on maps has been chosen according to the scale and the purpose of the map. The chosen information has been subject to the generalization processes, which are: classification (order of features by their attributes), simplification (portrayal of important feature characteristics and elimination of details), exaggeration (enhancement of important characteristics), symbolization (graphical coding of information), and induction (inferences from the interrelations among the features on the map). All the processes of generalization are done by the cartographer, except that of induction for which it is the user who makes logical extensions of the portrayed data, so it depends a lot on the symbolization.

The extent to which the generalized information presented on maps affects the way children interpret maps has not been studied systematically. From the few experimental studies in 1972, 1980, and 1981, there is evidence that children of primary and also of secondary school make misinterpretations, thinking that what is presented on a map is all that exists in the real world. A review study in 1998 recognizes the need for children's introduction to the concept of generalization and to the processes involved in it. Generalization is a key step toward successful map interpretation.

Symbolization

Maps use symbols that stand for the features of the real world they portray. Since 1967, when Jacques Bertin introduced a semiotic approach to cartographic symbolization, cartographers have followed a systematic symbol design, developing typologies of symbol categories. The visual variables (shape, size, orientation, hue, value, chroma, pattern, and texture) were the basis of symbol design assigned to represent quantitative and qualitative variations of the data represented on maps. Although the use of visual variables in symbol design is practically a standard procedure, it has not been considered as a prerequisite knowledge for map use.

Relatively little is known about how young children interpret symbols in maps. According to Piaget, children are able to recognize shape at sometime after 3 years. In 2000, another scholar underlined the fact that children with appropriate guidance could appreciate the symbol–referent relation earlier. It is argued in many experiments that children have greater difficulty in understanding the geometric correspondence than the representational one. In 1996, experiments found that kindergarten children show great variation in symbol identification. Children are able to identify shape variable and show difficulty in color naming. They easily identify both pictorial and abstract line symbols, but they show difficulty with point and area symbols. In other experiments in 1996, very young children aged 5–7 years found pictorial symbols attractive and easier to interpret. Many of them were able to identify abstract symbols using a legend. Size is easily identified. The understanding of color as a variable is cognitively complex, since it requires matching color differences with object characteristics, and this ability improves by age. The level of knowledge (experience) and the level of development (verbal ability and amount of attention) are mentioned as responsible factors.

Elementary school students are gradually exposed to maps that apply abstract symbols and the use of legend. Matching symbols between the legend and the map involves holding the symbol characteristic in memory, so this task develops gradually. Experiments in 1984 show that children understand qualitative symbols first, and quantitative ones later. Size is easily identified as a variable, but it is difficult for children to compare the size of symbols whose area is proportional to a quantity.

School atlases addressed to children older than 11 years comprise thematic maps representing quantities that occur at points, lines, or areas. When quantities occur over an area, a statistical surface is created. The concepts of the statistical surface, of ratio, and pro portionality are very important in thematic mapping and a prerequisite knowledge for understanding thematic maps like dot maps, choropleth, and isarithmic common to all school atlases. Experiments in 2003 show that students even of third and fourth year of secondary school do not fully understand the concepts of ratio and proportionality.

The Relief

The relief can be represented in maps by various methods, such as symbols of stylized form (appear not only on early maps, but also today on maps addressed to very young children), hachuring (lines representing the greatest slope), hill shading, layer tinting (hypsometric coloring), and contouring. Relief can also be represented through perspective pictorial maps (block diagrams, ob lique views, and schematic maps). There are two aims of relief representation: first, the visualization of relief by the user when seeing the map as a whole and, second, the interpretation of elevation data. The methods of repre sentation that are effective in the visualization task have poor results in the interpretation tasks. Contouring gives measurable data, but is poor in visualizing. Hill shading gives a realistic visual representation but nonmeasurable data. Recently, topographic maps represent relief by using contours and hill shading, and the result is very effective.

According to Piaget, children are able to understand the relief representation not earlier than 9 years of age and the concept of contouring not earlier than 11 years. In experiments in 1979, it is argued that only at the age of 11 can children interpret simple landforms. But even younger secondary school students show dif ficulties in height estimation and relief interpretation in cases where contours are not closed. Their performance improves with maps in which contouring is combined with layer tinting. Experiments in 1983 argue that limi tations on language development seem to be a serious problem in understanding the relief on maps, since children do not know the geographical terms. The understanding of contours appears to be a difficult task as well. Slope estimation seems to be the most difficult task, even for 14 year olds, as reported in studies in 1979 and 1989. For younger children, three dimensional models seem to be helpful for an introduction to landscape surfaces, as reported in experiments in 1997. Three dimensional representations and pictorial maps are more effective for primary school students to visualize landforms.

Children as Map Users in the Information Technology Era

Cartography at the beginning of twenty first century is facing a technological revolution due to the widespread use of electronic media and especially of computers and information technology. In many countries, indi viduals have access to a computer usually connected to a worldwide network everyday. Such a technological advance not only affects the technical frame of cartog raphy, but it also changes decisively the relation between cartography and society. During the long history of cartography, the latter faced significant technological revolutions that changed the methods of map con struction dramatically. But in every case, the cartographic processes needed specially trained staff (cartographers) in order to be performed. As a result, the knowledge of how to construct a map was related to a very small section of the society, that is, the cartographers. The existence of such advanced technological tools is trans forming any member of the society, and especially chil dren, into a cartographer by offering electronic systems able to construct any kind of map.

Indeed, children have a more privileged place by being familiarized with the use of the computers. Children can easily convey the knowledge acquired from playing a computer game into an effective construction of a map. Two issues are important for the children's proper assimilation of the present technology. The first one is related to the dramatic change of the available edu cational tools provided by information technology. Soft ware packages provide the ability to simulate several spatial phenomena (construction of models, walk through, and fly through) and to construct virtual worlds and examples, while at the same time their function is characterized by interactivity. Thus, using electronic media, children can be exposed to map skills and to map concepts (i.e., generalization, map projections, and spherical Earth) by means of various and effective communicative ways. The second issue refers to children as mapmakers. In the past, children were constructing an analog map of a continent, for example, by tracing (from a published map) the coastline on a piece of paper. With a computer, they can construct a detailed map of the same continent by retrieving data from a spatial database and symbolizing various features in alternative colorful ways. A critical question is raised here as to whether these maps made by children are effective in developing cartographic knowledge, or are they just drawings. Specialists on children and mapping persist ently ask for map literacy.

See also: Atlases; Children/Childhood; Cognitive Geography; Internet/Web Mapping; Maps.

Further Reading

Carswell, R. B., Leeuw, G. A. and Waters, N. M. (eds.) (1987). Atlases for schools. *Cartographica*, 24/1, Monograph 36.

Castner, H. W. (1990). *Seeking New Horizons: A Perceptual Approach to Geographic Education*. Montreal and Kingston: McGill Queen's University Press.

Cohen, R. M. (ed.) (1985). *The Development of Spatial Cognition*. Hillsdale, NJ: Lawrence Erlbaum.

Downs, R. M. and Stea, D. (eds.) (1973). *Image and Environment: Cognitive Mapping and Spatial Behavior*. Chicago, IL: Aldine.

Gersmehl, P. J. (2005). *Teaching Geography*. New York: The Guilford Press.

Golledge, R. G. (ed.) (1999). *Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes*. Baltimore, MD: Johns Hopkins Press.

Kraak, M. J. and Brown, A. (2001). *Web Cartography*. London: Taylor and Francis.

Kraak, M. J. and Ormeling, F. J. (1996). *Cartography: Visualization of Spatial Data*. Harlow: Addison Wesley Longman.

Kosslyn, S. M. (1994). *Image and Brain: The Resolution of the Imagery Debate*. Cambridge, MA: MIT Press.

Liben, L. S. and Downs, R. M. (1989). Understanding maps as symbols: The development of map concepts in children. *Advances in Child Development and Behavior* 22, 145 201.

McEahren, A. M. (1995). *How Maps Work. Representation, Visualization, and Design*. New York: The Guilford Press.

Newcombe, N. S. (2000). *Making Space: The Development of Spatial Representation and Reasoning*. Cambridge, MA: MIT Press.

Wiegand, P. (2006). *Learning and Teaching with Maps*. London: Routledge.

Wood, C. H. and Keller, P. C. (eds.) (1996). *Cartographic Design: Theoretical and Practical Perspectives*. Chichester: Wiley.

Relevant Websites

http://lazarus.elte.hu
Commission on Cartography and Children, Eotvos University, Department of Cartography and Geoinformatics.

Children/Childhood

F. Smith and N. Ansell, Brunel University, Uxbridge, UK

Glossary

Cyberspace The abstract space which is created when computers across the globe communicate with each other.

Identity Identity refers to the way we understand ourselves, other people, and places. Geographers have drawn upon post-structuralist approaches to highlight the ways in which identities are multiple. Moreover, such approaches have also illustrated the dynamic nature of idenity and the importance of discourse and performance for its production and reproduction. Social and political processes play their part in the development of individual and group identities.

Neoliberal Neoliberal economic policy seeks to reassign control of the economy from the public to the private sector.

New Social Studies of Childhood This approach has its roots in sociology and developed following the recognition that childhood is a socially constructed phenomenon, differing across time and space. Concentrating on everyday life and the ways in which children act and interact in society, it focuses on the cultural performances and the social worlds children both construct and engage with. Theoretically and methodologically, the new social studies conceptualize children as social actors; active participants in society whose agency is central to the creation of their own social worlds.

Nonrepresentational Theory Nonrepresentational theory seeks to move away from geographies of representation and discursive interpretation of a socially and culturally constructed world. Instead, academics using nonrepresentational theory focus on the practices, actions, and interactions of everyday life.

Post-Structuralism This approach highlights the importance of language and discourse to the constitution of (multiple) identities and the practices by which we come to understand society and the world.

Poverty Poverty refers to the lack of resources enabling individuals to access commodities which meet basic human needs, including food, clean drinking water, and shelter. Political and social conditions are inherent to the persistence of poverty.

Social Construction The process through which the interactions of individuals over time become routinzed and eventually lead to the development of social meaning.

Definitions

Defining childhood is problematic due to the spatial and temporal fluidity of prevalent ideas and laws pertaining to children. On one level, the definition used by many human geographers and other social scientists reflects the 1989 UN Convention on the Rights of the Child, which defines childhood as the period up to 18 years of age. However, while this implies that children are defined by their chronological age, a collection of studies encompassed by the term 'the new social studies of childhood' has highlighted the complex and contested nature of the term child. This work has highlighted how childhood is socially constructed, and as such is historically and culturally specific rather than reflecting any biologically essentialist reality. Moreover, there is no universal category of child as children's lives and experiences are diverse, inextricably linked to other aspects of their social identity such as gender, ethnicity, class, (dis)ability and shaped by the places where they live, go to school, work, play, etc. As such it is important to remember the global differences and complexities inherent in children's lives and experiences. Furthermore, rather than conceptualizing children as immature, yet to reach the social maturity of adults, human geographers along with colleagues in other social science disciplines (re)conceptualize children as human beings, worthy of investigation in their own right. In addition to providing new theoretical and conceptual understandings of childhood, many human geographers seek to work with children in order to transform their lives at a variety of levels including the local, global, social, political, and economic.

Origins of Work on Children by Human Geographers

Children remained absent from most geographical work until William Bunge's detailed empirical studies of Detroit and Toronto in the early 1970s, in which he focused on the spatial oppression of children in the built environment. At a time when geography was heavily dominated by the quantitative revolution, Bunge used a variety of methods including observation of children at play to highlight the inequalities faced by children living in high rise environments and in poverty. At around the same time a fellow geographer, Jim Blaut, began conducting some empirical work on the mapping abilities of children, establishing an agenda for research into children's geographical knowledge which led to a focus in geography on child development, spatial cognition, and mental maps. Geographers like Hugh Matthews followed Blaut and successfully critiqued the work of developmental psychologists who conceptualized children as psychologically immature, in need of guidance to develop into normal adults, by demonstrating the mapping abilities of very young children. Others such as Roger Hart were more influenced by Bunge and produced detailed ethnographic studies of children's experience of place and began highlighting the relationship between the environment and socialization. Cindi Katz further developed this notion by exploring the role of the physical environment in rural Sudan and New York in relation to social and cultural reproduction, as well as the local consequences of global economic restructuring. Drawing on a number of the research methods used by Bunge and Hart, she produced an innovative account of the ways in which global processes have severe consequences for children living in very different local environments.

Despite these seminal studies, it was not until the publication of an article in the Journal '*Area*' in 1990 by Sara James entitled 'Is there a 'place' for children in geography?' that there was a public debate on the state of research on children in human geography. This debate marked the beginning of the growth in interest in children from geographers which has led to the development of the current subdiscipline of children's geographies. However, it is important to note that geographers have not remained isolated from other social scientists in this development. Children's geographers have both drawn upon and now contribute to the interdisciplinary new social studies of childhood. Such studies challenge the relative absence of children in social science research, highlighting that childhood is socially constructed, varying across time and space, and according to the way it interacts with other aspects of social identity. It is this assertion that different childhoods are socially constructed and contested in different ways which forms the basis of analysis. Thus children's geographers explore the ways in which children's identities are classed, racialized, and gendered in different places in different ways. Moreover, the new social studies of childhood have influenced human geographers by conceptualizing children as competent social actors. It is also important to note that the increasing focus on children in geography also reflects the broader growth in interest in sociospatial inequality and previously hidden and neglected geographies influenced by a feminist and post structuralist perspective.

Theoretical Concepts

The Social Constructedness of Childhood

The concept of a socially constructed child is relatively new in geography. It developed in part as a response to the crisis of representation in the social sciences which began in the mid 1970s. Focusing on academic texts, concern was expressed over the ability of academics to misrepresent others through their writing. Children were viewed as a particularly significant group in this debate given their lack of participation in creating academic texts which slowly began to change the ways in which social scientists conduct research with children and other relatively powerless groups. According to social constructivists, the experience of childhood is discursively produced. While modernists naturalize childhood as a time of development, essentializing normative notions of 'the universal child', social constructivists highlight the plurality of childhoods, contextualized by social, historical, political, economic, and geographic processes. Geographers utilizing this theoretical perspective explore the material and corporeal experiences that embody children's everyday lives and consider the ways children's social identities are gendered, classed, racialized, etc. within particular spatial contexts. Thus, notions of different axes of social identity as mutually constitutive provide an important theoretical basis for many geographers working with children.

Children as (Competent) Social Agents

Mirroring the development of the children's rights movement, including legislation such as the United Nations Convention on the Rights of the Child, children have been increasingly conceptualized within children's geographies and the new social studies of childhood as social actors. Such a conceptualization views children as human beings, not merely 'human becomings'. Rather than complying with contemporary interpretations of children as less competent than adults, unable to make key decisions about their lives, such a conceptualization views children as competent social agents, exercising agency to transform their own social worlds. However, this is not to say that a consideration of wider social structures that shape and constrain children's agency has been deemed unimportant. To the contrary, the ways in which children's lives are structured by wider global and local processes remain central to analysis.

Children's Geographies

The following sections will explore children's geographies by drawing on a range of empirically rich material from both the 'minority and majority world', to highlight the relationships between children and the spaces they inhabit. The first part will explore the socioeconomic structures, institutions, discourses, etc. – including globalization, 'development', and poverty that 'place' children in particular ways. The second part will explore the ways in which children experience place.

The Spatial Construction of Childhood

One of the most important ways in which human geography can enhance our understanding of childhood is to highlight the importance of space and place to children's lives. Geographers question essentialist notions of a universal child by illustrating through diverse empirical studies the ways in which children's experiences vary in different places and in different contexts. The notion of what constitutes a normal childhood in the global North, embodied by ideals of innocence, playfulness, and happiness, is not only challenged by a number of studies in the global South but also by research highlighting childhood as a time of work, responsibility, and hardship in the North.

At a broad international scale there are large variations in economic wealth that have an impact on the lives of children. There are also political, cultural, and ideological differences between the ways in which children are treated and catered for in different national settings. There are also distinctions between urban and rural environments that pattern childhoods at a more local level, as well as innumerable fine grained differences between places. However, not only does spatiality shape children's lives directly, but discourses of childhood are spatially constructed in relation to notions of, for example, the rural or of national identity, in turn contributing to the cultural constructs that give meaning to place.

The following sections examine some of the structural and discursive processes that operate to differentiate between childhoods at the local scale and globally.

Poverty

The economic resources available, at household level, within wider communities, and at national level, have a great impact on the lives of children. In any society, children are usually disproportionately represented among the poor. They are also generally more adversely affected by poverty: undernutrition in childhood can have effects lasting into adulthood, and famine exacts its highest toll among the very young. Poverty has implications for children's access to healthcare and education, as well as leisure activities. It is also the main reason why so many children worldwide participate in the workforce from a young age. It is also poverty that drives some young people to live on the streets or to engage in commercial sex work.

Economic Change and 'Development'

Children are affected not only by economic conditions but also by economic change. Historical studies have shown how the emergence of spaces specifically for children (particularly schools) in Western Europe was associated with the requirements of the Industrial Revolution for disciplined labor, as well as the wish to exclude children from workplaces. In the twentieth century, various sets of strategies were implemented in an attempt to improve the living conditions in the poorer parts of the world. These were generally undertaken without children specifically in mind, but nonetheless impacted on children's lives. 'Structural adjustment' policies, imposed by international financial institutions on governments in the global South as a condition of debt rescheduling, were notorious for their negative impacts on the children of the poor, particularly through the introduction of fees for use of health and education facilities.

International Child Rights, NGOs, and Cultural Change

Although children have seldom been a major focus of national politics, at the international level there is a long history of action in support of children and their rights. Save the Children was established in 1919, in the aftermath of World War I in Europe, and was involved in drafting the Declaration on the Rights of the Child, adopted by the League of Nations in 1924. The discourse of rights associated with children that is promulgated internationally has changed over the past eight decades, but remains firmly rooted in Western expectations of what is appropriate for children. The UN Convention on the Rights of the Child, agreed in 1989, was the most rapidly ratified international treaty ever. The convention inscribes rights to protection, provision, and participation, which inform many of the activities of international child focused nongovernmental organizations (NGOs), as well as governments. Both policies and discourse are contributing to changing cultural expectations concerning children.

Globalization

Globalization is a term that encompasses a number of related processes, all of which impinge on children's lives. Economic globalization has drawn some children into export oriented production. More notably, the fact that firms are very mobile means that they have little incentive to invest in social reproduction in particular settings: it is easier to move on and find a new workforce elsewhere. In cultural terms, children worldwide have (to varying degrees) access to similar commodities: music, films, clothing, food, drink. The fact that similar commodities are consumed does not, however, imply that their meanings remain constant: children in different local settings attach different meanings to particular items, and consume them in different ways.

Children's Experiences of Place

The following section will explore one of the key contributions geographers are making to the study of childhood, by focusing on the empirical work geographers are doing in the everyday spaces of childhood which shape (and are shaped by) their lives.

The Home

The home is an important site for the negotiation of child–adult power relations. The home is a bounded spatial setting in which adults generally determine the rules and regulations governing appropriate ways in which children are expected to behave. It is also a porous site, in many respects open to cultural influences through technologies such as television and the Internet, and the fact that people move in and out carrying ideas from beyond its boundaries. As such, the home provides a key setting in which children's gender, class, racial, and sexual identities are produced and reproduced. The work of geographers has been important in highlighting the diverse ways in which children experience their home environments. As well as being a space of refuge for children being bullied at school or in public space, the home can also have a darker side as it is also the site of most child abuse. Moreover while most children are cared for by adults within the home, others, including those with disabled parents, or parents suffering alcohol or drug dependency, may find themselves in the position of carer. This role is even more common in parts of the world characterized by higher levels of adult morbidity, and by orphanhood that leaves some homes without adult carers: a situation that is increasingly common in the parts of Africa that are worst affected by the AIDS pandemic.

Ideologically, in Western societies, children are considered to belong within the home. This notion has informed a great deal of child policy in Western countries and has been carried into other societies through instruments such as the UN Convention on the Rights of the Child, and through the policies and projects of many development agencies.

Institutional Settings (Childcare and School)

Worldwide, children are spending an increasing amount of time in institutionalized spaces, most notably school and daycare. Children are thus increasingly separated from adults and placed in specialized children's environments. Contemporary childhood, especially in the global North, can be conceptualized as a period of spatial separateness, in which children are excluded for a significant part of the day from the wider adult world. Moreover, geographic research highlights how such environments structure the ways children use and experience space and organize their time, enabling adults to practice surveillance and regulation of children, processes which children actively contest.

Schools are arguably the single most significant institution in most children's lives. Historical and geographical work has illustrated how schools were developed to socialize children into appropriately behaved citizens and productive workers. By creating a bounded environment in which groups of generally similarly classed children could be controlled and inculcated with appropriate notions of how to behave according to their class, gender, and ethnicity, schools have played and continue to play a fundamental role in the production and reproduction of children's individual identities, as well as society more generally.

Schools do not, however, play a uniform role. Different societies have different ideas concerning what children need to learn and the behaviors that are expected of them, as well as different levels of resource (made) available to commit to education. This has resulted in distinct differences in children's experiences of schooling in different parts of the world. In many parts of Africa and Asia, education systems mimic those of the former colonial rulers, with curricula and teaching methods geared around the need for children to pass public examinations.

The growth in institutionalized childcare environments in the global North is a response to a number of things. First, it reflects both the desire of increasing numbers of women to work outside the home, as well as changing social policy, most notably the growth in neoliberal workfare policy, which makes it harder for women living on welfare benefits to stay at home with their children. Second, it reflects changes to the family, in particular, the increase in lone parent households. Finally, formal childcare is increasing due to spatialized fears for children's safety in public space. Rather than being allowed to roam freely in what is increasingly considered to be dangerous public space, a growing number of children are placed in what are considered more appropriate adult controlled institutionalized environments in which they can be properly cared for and protected while their parents are in paid work. However, as a number of geographers have highlighted, this is not to say that children passively accept the regulation of such environments, but actively contest the adult centered organization of the space.

In the global South everyday childcare is generally arranged on a more informal basis, although under the influence of the World Bank, in particular, preschools are mushrooming in many countries. As in Western countries, these are intended both to better prepare children for school and to promote the integration of women into neoliberal workforces. Historically, impoverished societies had high rates of full time institutional care for children whose families could not look after them. Institutional residential care is, however, falling out of favor, particularly in light of the UN Convention on the Rights of the Child. Nonetheless, in societies affected by war or by AIDS, the demand for institutional care may increase.

Public Space

Geographers' main concern with public space has been to explore the ways in which children access and use it and the attachments they form to it. A number of studies have explored the opportunities children have to play in public space, highlighting uneven access to specialist play provision such as playgrounds and parks while also illustrating children's capacity to create play opportunities regardless of local formal amenities. Such work highlights the ways in which children utilize, experience, and appreciate space differently from adults. Children redefine everyday settings by creating opportunities to play and constructing secret places and dens which has led to the suggestion from some geographers that adults should stop providing formal playgrounds in favor of more open space, where children are enabled to create and recreate their own play environments. Moreover, historical work by geographers has highlighted the fact that the development of formal play provision in the global North has been a common response to the desire to contain children within a controlled and bounded spatial setting, rather than as a way of providing an empowering child centered environment. Work in the United States, for example, has illustrated how the growth of playgrounds in the early 1900s was a deliberate attempt to ensure immigrant children were socialized as appropriately gendered American citizens. Other research in the United Kingdom exploring the current growth in commercial play spaces also raises important questions about children's place in public space.

Recently, geographers have highlighted the growing concern for children's safety in public space in the global North. Children are conceptualized as being vulnerable to the danger inherent in public space from strangers and cars, as well as from other children. However, despite this, the street remains an important space for children, particularly for those from working class backgrounds. Work by geographers has been important in highlighting that while children are significant consumers of street culture and public space more generally, they are generally excluded from participating in the planning processes which impact upon the ways in which it is designed. A number of geographers are striving to combat this, by working in innovative ways with children and local planning authorities. However, there is some debate as to how effective this actually is and whether or not it is more tokenistic than participatory.

Research in the global South has tended to find that rural children have considerable spatial freedom, although their time may be subject to stricter management by adults. For most rural children, attendance at school entails a lengthy walk, in the absence of private and in many cases public transport. Children may also engage in (paid or unpaid) work in public space. In Bolivia and India, for instance, children are expected to undertake chores that take them away from the home for hours at a time: time which they use not only to complete the tasks expected of them, but also to take opportunities to play.

Urban environments are generally considered less suitable for children. In many cities, particularly in less affluent areas, children are at risk of violence when outside the home, and particularly if they stray from their own neighborhoods. Some choose to join gangs (sometimes as a means of self protection), which exposes them to even higher risks. While public space may be seen as unsafe for children, in many cities, substantial numbers of children live on the streets. These children are generally regarded as 'out of place', disrupting ideologies that place children indoors and within conventional families. That they are able to survive on the streets indicates their capacities as social actors, carving out niches for themselves within hostile urban environments.

Cyberspace

In the North cyberspace has become an increasingly important place for children to interact with others, both globally and locally. Cyberspace offers children the opportunity to network with peers who may live within their local environment or on the other side of the world, thus enabling children to (re)produce relationships across time and space. It is likely that children will increasingly form social relationships free from the constraints of spatial locality. Given the fears over children in public space discussed earlier, virtual space has become an important alternative to the material spaces of childhood. Indeed, virtual space may offer some children the opportunity to publicly interact in a world in which access to the public sphere is becoming increasingly restricted. Information technology thus offers children important new spaces for the development of children's online cyber communities. Within these communities, children have the capacity to contest and resist adult notions of acceptable behavior, which has led to a rise in moral panics around children accessing inappropriate material such as pornography or being subject to the predatory gaze of pedophiles. However, as a number of geographers have pointed out, this is not unique to virtual space, but reflects moral panics surrounding childhood more generally in public space. The increasing amount of time spent by some children in cyberspace is also leading to concern from adults over the amount of time children are now spending alone, inside, in front of a computer rather than interacting face to face with other children outside. Anxiety about children who are becoming socially isol ated loners spending all their free time in cyberspace has to some extent replaced concern over the amount of time children spend in front of a television. Such concern, however, is interesting to geographers as it reveals something about the way adults conceptualize the ap propriate place of childhood, one that is not socially isolated and indoors.

While virtual space shares some characteristics with material space, virtual environments are relatively unique in children's geographies in that they allow chil dren to (re)establish their own identities free from many constraints. Thus, while identity and individual charac teristics are limited by biology and society in the material world, within cyberspace children can choose, if they want, to change their age, gender, ethnicity, class, etc., in terms of the information they give to chat rooms, gaming systems, or individuals by e mail. However, although virtual space offers children the possibility to play with their identity, it is important not to overstate the ability of cyberspace to eradicate the realities of material ex istence for children.

Cyberspace remains a sphere of relative privilege. Although many children in the global North have access to the Internet, through school if not at home, worldwide its use remains highly uneven. In poor countries it is generally only the elite that can afford to allow their children to go online. There are, however, some in novative projects that allow school children to com municate ideas and compare lifestyles with their contemporaries in very different settings.

Critiques and Challenges

Outside of academia, the children's rights movement has, as yet, had only limited success. The global position of children remains poor, with significant numbers of chil dren lacking a basic education, clean water, or food. The UN Convention has been criticized for trying to impose a global vision of childhood which fails to adequately recognize cultural differences or varying needs. Moreover many individual countries have been criticized for failing to implement the UN Convention. Therefore, at both the local and global level, children still have little political or economic power and continue to form the world's largest minority, disenfranchised group.

Academically, children's geographies are challenged by recent critiques that call into question the utility of the main precepts of the new social studies of childhood. As children's geographers have become interested in emotions and embodiment, the focus on childhood as a social construct limits the possibilities for research. In creasingly, geographers need to understand childhood as 'both' a social 'and' a biological phenomenon. Likewise the emphasis on children as social actors, which tends to incorporate an understanding of children as individual human subjects, acting knowingly upon the world, is challenged by research that adopts a more relational view of children, and stresses the ways in which they are 'subjectified'. This is not to suggest that we need to return to a study of children as 'human becomings', as merely future adults, but to recognize that all human subjects are in a constant process of becoming, through their rela tionships with others and with the world. In line with these critiques, novel research into children's lives has begun to engage with nonrepresentational theory, and also to explore new ways of understanding children's lives that move away from a focus on enabling children to participate in research by describing the world from their own perspectives.

See also: Child Labor; Cyberspace/Cyberculture; Feminism/Feminist Geography; HIV/AIDS in Developing Countries; Home; Parenting/Motherhood/Fatherhood; Poverty; Education; Youth/Youth Cultures.

Further Reading

Aitken, S. (2001). *Geographies of young people: The morally contested spaces of identity*. London: Routledge.
Ansell, N. (2005). *Children, youth and development*. London: Routledge.
Blaut, J. and Stea, D. (1974). Mapping at the age of three. *Journal of Geography* 73, 5 9.
Bunge, W. (1977). The point of reproduction: A second front. *Antipode* 9, 60 76.
Hart, R. (1979). *Children's experience of place*. New York: Irvington.
Hart, R. (1997). *Children's participation. The theory and practice of involving young citizens in community development and environmental care*. London: UNICEF/Earthscan Publications Ltd.
Holloway, S. and Valentine, G. (eds.) (2000). *Children's geographies: Playing, living, learning*. London: Routledge.
Holt, L. (2004). The 'voices' of children: De centring empowering research relations. *Children's Geographies* 2(1), 13 28.
Katz, C. (1991). Sow what you know: the struggle for social reproduction in rural Sudan. *Annals of the Association of American Geographers* 81(3), 488 514.
Matthews, H. (1984). Environmental cognition of young children: Images of school and home area. *Transactions of the Institute of British Geographers* 9(1), 89 105.
Punch, S. (2002). *Youth transitions and interdependent adult child relations in rural Bolivia' Journal of Rural Studies* 18(2), 123 133.
Robson, E. (2004). Children at work in rural northern Nigeria: Patterns of age, space and gender. *Journal of Rural Studies* 20(2), 193 210.
Smith, F. and Barker, J. (2000). Contested spaces: Children's experiences of out of school care in England and Wales. *Childhood* 7(3), 315 333.
Wood, D. and Beck, R. (1990). Dos and don'ts: family rules, rooms and their relationships. *Children's Environments Quarterly* 7(1), 2 14.
Young, L. (2004). The 'place' of street children in Kampala, Uganda: marginalisation, resistance and acceptance in the urban environment. *Environment and Planning D: Society and Space* 21, 607 627.

Relevant Websites

http://www.gcyf.org.uk
Geographies of Children, Youth and Families Working Group, RGS IBG.

http://www.colorado.edu/journals/cye
University of Colorado at Boulder, Children, Youth, Environments.

Chinese Urbanism

L. J. C. Ma, University of Akron, Akron, OH, USA

Glossary

Central Plain The area in the middle reaches of the Yellow River, including northern Henan and southern Shanxi provinces, where the earliest cities emerged.

Danwei A work unit and basic spatial and social cell in China's socialist cities.

Hangtu A technique of city-wall construction whereby layers of earth are pounded down firmly in wooden frames until the desired height is reached.

Longshan Culture A late Neolithic culture (3000–2000 BC) during which walled towns and early cities first appeared.

Urbanism The most basic character and meanings of the fundamental and dominant aspects of the development of cities.

Introduction

Cities are complex spatial entities consisting of economic, social, political, and cultural dimensions that differ significantly in different cultures, where they tend to undergo varying patterns of transformation over time. Cities and states are dynamically related and mutually constituted, with the former often embodying the essence of the latter. This is especially true in the case of China. Relative to other major cultures of the world, China's urban experience has been quite unique in terms of the process of urbanization as well as the nature of urbanism. While urbanization is most frequently expressed as the increase in the number of cities, the expansion of urban space, and the growth of urban population, the concept of urbanism has been much more protean and scholars have used it more freely without a consensus of its definition.

Whereas studies on China's urbanization abound, few address Chinese urbanism, and the mutually reinforcing relationship between nation building and urban development in different historical periods has not been given the level of attention it deserves. In this article, cities are seen as the discursive and material sites where a nation's major political, cultural, and social systems are developed and specific projects for national development are concentrated, and urbanism is viewed as a historically dynamic phenomenon that undergoes significant changes in different eras. Conceived broadly, urbanism refers to the nature and meanings of the most fundamental, salient, and dominant aspects of city development in a particular nation or region in a given period of time. Operationally, this article adopts a macro historical stance to outline what major forces have affected city development and the resultant urbanism in different periods of Chinese history, paying particular attention to the mutually reinforcing relationship between the state and the city. As such, urbanism is at once a process and a spatial, cultural, and socioeconomic manifestation of urban change. The trajectories of Chinese urban development can be divided into five periods, each with a set of distinct characteristics of urbanism.

Neolithic Settlements and Incipient Urbanism (*c.* 5000–770 BC)

As in other parts of the world, early Chinese towns evolved from villages. During the Middle Neolithic era (*c.* 6500–5000 BC), large sedentary villages with a few hundred people based mainly on agriculture and animal domestication had emerged. Widely distributed in China, many of the villages were surrounded by circular ditches and walls constructed with the *hangtu* (tamped earth) technique, whereby layers of earth are pounded down firmly within wooden frames, a technique that continued to be used in all subsequent periods in China's city wall construction. City walls were the most enduring morphological feature of the Chinese city, serving as a spatial marker demarcating a privileged space for the security of the urban population.

China's first walled towns, some with moats, emerged during the Longshan period (3000–2000 BC) in the Central Plain in modern Henan Province. Other town sites were found in the mid Yangzi River valley, Shandong peninsula, and Inner Mongolia. Incipient urbanism appeared during the Longshan era. These walled settlements, most of them squarish or roughly rectangular in shape, while others roughly circular and oval, have been labeled *chengbao* (literally, walled fortresses) by Chinese archaeologists, a term that suggests that their main function was defense. Archaeological finds in these towns suggest that they were not yet cities, because territorial state and written language had not appeared. However, raised ceremonial platforms, cemeteries, and a large number of clay figurines of animals and humans, jade objects, and copper smelting slag were found, which suggest that the walled towns were religious and production centers with highly stratified societies.

The Longshan period overlaps in part with what Chinese archaeologists believe to be China's first dynasty and territorial state, the Xia (*c.* 2200–1750 BC), during which China's earliest city, Erlitou (1900–1500 BC), which is located near today's Yanshi City, Henan Prov ince, emerged. Occupying an area of 300 ha, the Erlitou site was most likely a capital of the Xia dynasty and possibly an early capital of the succeeding Shang dynasty (1600–1046 BC) as well. The Erlitou culture represented a highly significant phase at the very beginning of Chi nese civilization. At the Erlitou site, there are clear signs of state formation, urban planning, and the existence of complex societies with clear social stratification and production specialization. Whereas it is not clear which ancient capital mentioned in ancient Chinese documents corresponds with the Erlitou site, scholars believe Erlitou was a capital city with a large temple/palace complex as well as a large production center where bronze, bone, and ceramic objects were produced, many of them used for ritual and ceremonial purposes. The production of bronze weapons and ritual vessels was of particular im portance as ceremonial bronze vessels symbolized re ligious and political power. As such, the Erlitou site was clearly an early political, military, and religious center.

Abundant archaeological evidence has been found at three sites, all in modern Henan Province, that were unquestionably the largest urban centers of the Shang dynasty: (1) a site at Yanshi named by archaeologists as Yanshi Shang City (*Yanshi shangcheng*), (2) a site named Zhengzhou Shang City (*Zhengzhou shangcheng*) in the modern city of Zhengzhou, and (3) a large site at Xiaotun in modern Anyang long known as the Yinxu, which lit erally means 'the remaining sites of the Shang'. These three cities were unquestionably the major cities of the early, middle, and late Shang period, respectively. Scho lars agree that Zhengzhou Shang City and Yinxu served as capital cities of the middle and late Shang, but it is less clear if Erlitou and Yanshi Shang City were also, re spectively, capital cities of the late Xia and early Shang. It should be noted that no city walls have been found at Erlitou and Yinxu, which begs the question as to whether the presence of city walls should be considered as a ne cessary condition for defining early cities in China.

Archaeological data of the four earliest cities men tioned above suggest that early Chinese urbanism differs significantly from the key characteristics of the cities in ancient Mesopotamia, Egypt, and India. Aside from the fact that the population of these cities was unquestion ably larger than the nearby villages and that there was a clear division of labor in specialist handicraft production, archaeological assemblies at the four ancient Chinese cities do not display the other major characteristics of the ancient cities of the Old World as identified by the eminent British archaeologist V. Gordon Childe in his highly influential 1950 paper, where written language, long distance trade, monumental public buildings, and surplus accumulation were also listed as key features of the earliest cities. China was not even mentioned in this study. Early Chinese urbanism was based on the role of the cities as political, religious, and defense centers.

Administrative and Commercial Urbanism (770 BC–AD 1840)

Childe's key criteria defining early cities in the Old World show clearly the importance of economic pro duction and trade in the rise of its early cities, and eco nomic factors have remained dominant in subsequent Western urban history. In China, during the Zhou period (*c.* 1122–256 BC), a politically fragmentary era with many independent states contending for territory and su periority, cities grew as state capitals and centers of ad ministration, defense, and trade. With little interference from state authorities and significant improvement in agricultural productivity spurred by the use of iron tools and private land ownership, urban manufacturing and commerce flourished, which greatly facilitated social mobility and enhanced the economic functions of the cities. After the Qin (221–206 BC) unified the country and beginning with the Song dynasty (960–1279 AD), cities became increasingly secularized as aristocratic political authority gave way to bureaucratic state control in national polity. Gradually, cities lost their religious significance where early Zhou kings had worshipped heaven and Zhou noblemen kept their ancestral shrines. From the Qin to the mid nineteenth century, Chinese cities functioned essentially as seats of imperial ad ministration and centers of territorial control but with a significant level of economic activities. A well structured city hierarchy based on three or four levels of local territorial units was developed, with the county capitals that hovered around 1500 in number for much of the imperial period serving as the lowest level cities and key centers of imperial field administration. There were, however, numerous towns, mostly without walls, that flourished as grass roots market towns where the imperial government had no official presence.

The presence of the state in the cities was clearly visible and it strongly affected the morphology of the cities and the nature of urbanism in dynastic times, giving rise to what might be called 'administrative urbanism'. This term does not imply that the cities of this period had no other functions, as we know they also functioned as commercial, educational, and cultural centers for the local population. Here the term is used to emphasize the fact that Chinese cities grew mainly because they were cen ters of state administration, a fact that has never changed even in the twenty first century. As seats of administration and local political power, cities were walled for defense and local population control. Aside from city walls, the most salient spatial element of the cities was the gov ernment district consisting of a set of local government buildings (*yamen*) that served as the nerve center of a city. Other key spaces in the cities included official residences, military compounds, drum and bell towers, granaries, schools, temples, and commercial areas. When settlements appeared outside of the city walls, a second wall enclosure would be added to facilitate population control and urban administration. As centers of administration, cities could grow rapidly through the establishment of a local branch government and organized migration of population, but they could also be ruined rather suddenly as a consequence of frequent internal rebellions, invasions by foreign groups, and dynastic changes, which often involved bloody warfare.

Increasing secularization, bureaucratization, and economic growth after the Song dynasty led to highly significant development of urban commerce. Geared more and more to the needs of the general urban population instead of to the demands of the elite classes as in the past, China's vibrant urban economy during much of the millennium from the ninth to the nineteenth century was largely service oriented, centered around a variety of stores, numerous restaurants, hotels, entertainment places, and commodity storage buildings. Aside from merchants, there existed a huge number of government officials, their assistants, local staff and their families, public security force in the cities, and for the large cities – especially the national capital – there were numerous handicraftsmen working in state run workshops and large numbers of troops stationed in and near the cities. Their daily needs had to be met. Historical commercial urbanism was thus marked primarily by local consumption demanded by these urban dwellers and by those who were present in the cities to provide various services, but conspicuous consumption by the rich and powerful was also common. However, vibrant commercial urbanism had not always been present prior to the ninth century. With the exception of the Zhou period in which a laissez faire economy was developed in the cities, urban commercial activities, the locations of urban markets, and their trading hours were rigidly regulated by the state. After the ninth century, urban commerce was no longer restricted to specific urban wards but dispersed throughout urban space where businesses could be conducted with virtually no time limitations. However, significant commercial capitalism failed to appear in Chinese cities because the most profitable commodities such as salt and tea were monopolized by the state and the Confucian elites all preferred to pursue government service as their career goal. They showed little interest in commerce and disdained technology. However, consumption based commercial urbanism represented a significant dimension of China's traditional urbanism, which remained little changed until the mid nineteenth century when Western powers arrived in China demanding trade with the Middle Kingdom that until then had existed in 'splendid isolation' from the rest of the world.

The Arrival of the West and Modernist Urbanism (1840–1949)

Urbanism in China experienced a dramatic change in the mid nineteenth century when increasing numbers of Western traders arrived in China demanding expanded trade along the China coast, which precipitated the Opium War (1839–1842) between the British and Chinese empires. China's loss of the Opium War and other subsequent wars with foreign powers resulted in a series of unequal treaties that forced China to open an increasing number of 'treaty ports' where foreigners could reside and conduct business in 'concessions' or 'settlements' leased to foreign powers with extraterritorial privileges. Eventually, a total of 73 treaty ports and 25 'ports of call' were established for foreign traders, first along the coast and later in inland areas. In addition, China voluntarily opened up 32 trading ports that foreign steamships could visit.

In the treaty port cities, an entirely new form of urban landscape with Western style buildings and streets was created. Whereas the size of these foreign quarters varied in different cities, they were small relative to the indigenous sectors. However, the significance of foreign quarters goes far beyond their Western style urban landscape. These foreign enclaves, and by extension, the treaty ports themselves, while a symbol of foreign imperialism and China's humiliation, were also centers of dramatic change in almost every realm of China's development for more than a century. Foreign traders, with the aid of Chinese compradors in the beginning, brought with them not only new commodities that generated handsome profit for them, but also, more importantly, a wide range of new knowledge and ideas, including Western science, technology, and democracy that seriously challenged Chinese traditions and eventually precipitated profound changes in virtually every aspect of China's development. As such, the treaty ports, which by the 1930s had included all of China's major cities except for Beijing, can be considered spaces of Chinese enlightenment through which Western ideas, goods, and institutions flowed in and where the debates as well as the actual projects of China's early modernization in general and industrialization in particular were centered.

The impact of the West rendered Chinese cities socially and demographically much more heterogeneous than before as new classes of urban elite, including foreigners and Chinese entrepreneurs, emerged. Politically, the treaty ports introduced Western municipal administration and court systems to China. They also served as sanctuaries for many of China's revolutionaries who could get protection in foreign settlements. Spatially, the pattern of China's urban development was altered in favor of the coastal region, in sharp contrast to the earlier pattern that had been focused heavily on inland cities because the Chinese empire was essentially land based, and because, during the Ming (1368–1644) and the Qing (1644–1911) dynasties before 1840, the major function of the coastal cities was limited to military defense against pirates, a function that precluded meaningful coastal development. In 1943, the system of treaty ports ceased to exist as the unequal treaties were replaced by new and equal treaties.

However, China's modernization efforts continued well into the mid twentieth century. Despite the fact that before 1937 China had suffered from frequent internal conflicts among regional warlords and wars against Jap anese invasion, by the 1920s most large cities and many smaller ones in coastal and inland areas had initiated modernization projects that fundamentally altered tra ditional Chinese urbanism, with the reorganization of urban space and the introduction of new urban insti tutions forming the core of the nationwide modernist agenda. From the 1840s to the 1930s at least, Chinese urbanism was shaped essentially by modernization ef forts, and as such, it can be argued that it was a period of modernist urbanism, a term used here to emphasize the fact that the dominant concern of the national and urban leaders was modernization, which was city based.

Chinese cities during the first three decades of the twentieth century witnessed significant changes in urban design and infrastructure. Cities continued to be the breeding grounds of new ideas for national development and served as the platforms for the debate of national identity. More new ideas were introduced from abroad by, and debated hotly among, urban intellectuals, many of them trained in the West and Japan. Under the effects of a worldwide urban reform movement in the early twentieth century and embracing a common model of urban administration and city planning developed in the West and carried out effectively in Japan, Hong Kong, and the concession areas of the treaty ports, major Chi nese cities, coastal as well inland, initiated specific city planning projects. These cities included Beijing, Shang hai, Nanjing, Guangzhou, Nantong, Tianjin, Changchun, Wuhan, and Chengdu. A new urban polity centered on urban administration (*shizheng*) emerged, and city gov ernments were reorganized with city codes, appointed mayors, and assemblymen. For the first time in Chinese history, the city became a legal entity which represented a major departure from the lack of administrative status in traditional territorial organization. Prior to the early 1920s, there was no city government and all matters of a city, including national capitals, were under the ad ministrative purview of the magistrate(s) of a county (counties) in which the city was located.

The most visible changes in early twentieth century China occurred in the urban landscape. Cities underwent a major urban facelift through the removal of city walls, the widening and paving of streets, the promotion of personal hygiene and public sanitation, the installation of water supply and drainage systems, electricity and tramways, the construction of urban parks, department stores and hotels, hospitals, churches, museums, theaters and dance halls, modern industries, railway stations and Western style universities and schools, the introduction of a modern police system, chambers of commerce, banks, newspapers and magazines, and the initiation of 'local self governing' assemblies in urban administration. There was a uniform national consensus for major changes in all realms of national development and a strong national quest for modernity. Through the power of the press, scholarly journals, and popular magazines, a distinct urban culture emerged, characterized by the presence of avant garde writers and artists, a modernist mindset pervading among the educated urbanites, and the preference for modern lifestyle among all urban residents.

Under the strong current of modernization, there was a widespread concern for national cultural identity, es pecially in Beijing, Shanghai, and the tourist city of Hangzhou. Newly introduced ideological and material istic urban elements rendered the Chinese city a clearly recognizable legal, spatial, socioeconomic, and cultural entity that differed in major ways from traditional Chi nese urbanism. However, China in the first half of the twentieth century suffered from repeated wars and fre quent shifts of political regimes. Political fragmentation, war with Japan (1937–45) and an all out civil war be tween the Nationalists and Communists from 1945 to 1949 rendered the half century a period of major national disruptions and destructions. Whereas the national as pirations for modernization were high and the basic framework for modern urban administration was set up in China, constant political upheavals, incessant wars, and the lack of specialists in urban administration prevented the cities from successfully implementing urban reform projects. With the establishment of a socialist government in 1949, modernist urbanism which managed to bud but failed to bloom fully under extremely difficult national conditions was abruptly replaced by socialist urbanism which represented yet another major metamorphosis in the history of Chinese urbanism.

Socialist Urbanism (1949–1978)

Socialist urbanism in China was created very quickly during the first three decades of Communist rule. From 1949 to 1978, cities served two major functions. First, they continued to serve as centers of national and local administration, a function that directly affected their growth. Second, cities played a passive supporting role in China's centrally planned socialist industrialization, serving essentially as spatial containers instead of functioning as an active force, and socialist urbanism resulted directly from what is known as 'urbanization from above'. As central economic planning precluded cities from pursuing any independent courses of action and the budget of a city was allocated from the central state, cities themselves did not play an active role in China's socialist economy.

In terms of spatial structure, China's socialist urbanism was marked by several distinct new features. As seats of administration, cities were designed to glorify socialism. All major cities had a political center consisting typically of a large public square used for political rallies, one or more large government office buildings, and one or more broad avenues lined with tall buildings and trees radiating out from the center. In the middle of large traffic circles, one might find the statue of Mao with his right arm raised high in the air. Unlike in the West, commercial nodes in China's socialist cities were rarely impressive, visually or functionally, because urban consumption was severely limited.

The basic spatial and social cell of China's socialist city was the 'work unit' (*danwei*), which might be a factory, a state agency, or a university. Each work unit had a clearly demarcated space, often walled, in which the employees of the unit worked and lived in housing subsidized by the state. Large work units were self contained, complete with housing, schools, clinics, and other facilities. Work unit housing, which constituted the bulk of urban housing, was built with standardized criteria for per capita space and building styles varied little. The presence of a large number of fairly uniform five or six story walk up flats, together with low personal income and low level of urban consumption and the lack of service establishments, rendered urban life and city landscape rather drab and dull. Industrial work units tended to be found in different parts of a city, despite the claim of functional zoning in socialist city planning. As industrial production was carried out without any concern for environmental protection, air and noise pollution was common in the cities.

As in other socialist states, China's socialist urbanism was also characterized by low levels of urbanization accomplished mainly through strict control of rural–urban migration to reduce the cost of urbanization, including the cost for the construction of urban infrastructure, housing, and schools. Between 1950 and 1978, the levels of urbanization, defined as the proportion of total population living in cities and officially certified towns, hovered between 11.2% and 17.9% only. There were extremely low levels by any standards. Unlike in the capitalist world where industrialization and urbanization have gone hand in hand, the strategy of socialist development in China was industrialization with limited urbanization. To attain the goal of spatial equality in development, industries were built in many inland cities, especially during the First Five Year Plan (1952–57), including Baotou, Wuhan, Zhengzhou, and Xi'an which grew rapidly as result.

A unique phenomenon of China's urban experience during the 1950s to the 1970s was Mao's anti urban sentiment that resulted in a large number of urban cadres being sent down to the countryside for a short period of time to 'learn from the peasants'. During the Cultural Revolution (1966–76), millions of urban high school and college students were also sent to the countryside for re education and settlement. Such anti urban rustification campaigns were initiated by Mao to re energize people's revolutionary spirit and, during the Cultural Revolution, to root out Mao's opponents in the state and Party machinery in the cities. The campaigns ended with Mao's death in 1976. The initiation by Deng Xiaoping of the policy of "reform and open" to the outside world after 1978 has, once again, brought dramatic and profound changes to the Chinese cities and urbanism. In less than a century, Chinese urbanism has experienced three major metamorphoses, from the quest for modernity in the century after the mid nineteenth century to socialist urban transformation during the period 1949–78, and then to massive post Mao urban reconstruction that has occurred with unprecedented large scale, breath taking speed and great economic and social complexity.

Multiplex Urbanism in the Reform Era

China's dramatic urban transformation since the 1978 reforms is a consequence of the interplay between exogenous and endogenous forces at global, national, and local scales. As elsewhere, globalization and neoliberalism have strongly impacted the Chinese economy, but major endogenous policy shifts have been more important in effecting changes in the Chinese city, including the initiation of the policies of marketization, decentralization of political and fiscal powers to the local levels, including cities, and privatization of state housing and state owned enterprises. The introduction of urban land and housing markets and simultaneous de industrialization in the socialist sector and tertiarization in the new urban economy have been particularly important in reshaping China's postreform urbanism.

Since 1978, cities in China have been completely overhauled, affecting virtually every aspect of the city – spatial, social, cultural, economic, and political – resulting in what can be called 'multiplex urbanism'. Politically, reform era urbanism is characterized by the dominant role of the cities in local territorial administration. Promoted by the state as the engines of growth for national and regional economies, cities have been allowed to annex suburban counties and turn them into city districts so that a large amount of suburban land is now directly under the control of central cities.

Economically, reform era cities have attracted massive foreign capital from Hong Kong, Taiwan, and other countries and lured a greater amount of domestic capital from the state sector. Thus, new industries in new industrial spaces, new commercial establishments in new spaces of consumption, and various types of new services have appeared in the cities. With increasing exports of products from China, now the World Factory, a significant urban middle class has emerged who buys cars, homes, and other consumer goods. Commercialization of urban housing has served as a filter, sorting out the urban population in different areas with different housing quality, giving rise to new spaces of differentiation and marginalization exemplified by exclusive gated communities and dilapidated migrant enclaves. Socially, the urban population has become more heterogeneous than before. Rural migrant groups, foreigners, and businessmen from Taiwan and their families are found in significant numbers in many cities. Rising income gaps in urban China have led to significant social and economic polarization and inequality in housing ownership. Culturally, urban life has been greatly enriched, with a great deal more cultural activities available than before, including active night life, entertainment, high fashion, more reading materials, and a wide variety of Internet based games and personal interactions. The lively and bustling urban scene in the reform era contrasts sharply with the subdued and simple urban life during the socialist period.

A key component of China's multiplex urbanism is that the cities have become dynamic centers of production for the nation as well as the world, with money going to the coffers of municipal governments. Cities are no longer simply spatial containers of state industries. Their economic well being is now tied directly to the taxes and various fees that cities collect from the firms and facility users in the cities. With increasing per capita income in the cities, a subtle aura of optimism and rational exuberance about the future exists among the urbanites. A sense of pride among the urban residents can also be detected, flowing in partly from the new urban spaces that have been created, including new central business districts (CBDs), new financial districts, new housing estates, new urban expressways, and new malls and suburban shopping centers. On the other hand, the drive for modern architecture on the part of city leaders has given global architectural firms the opportunities to build ultramodern, exotic, and often highly controversial and even grotesque buildings in Chinese cities that seem out of place. As the tempo of urban construction is extremely fast and nobody seems to be concerned about building Chinese cities with some kind of Chineseness, in many respects, Chinese cities are evolving without a distinct personality. In some ways they are converging in form with cities elsewhere. However, one must not assume that the processes affecting China's urban form are similar to those in other parts of the country because the ideological, cultural, geographical, and political contexts in urban China are so different.

The Persistence of the State in Chinese Urbanism

It is clear that Chinese urbanism has evolved differently under different historical contexts. It has always contributed to state formation while reflecting the national concerns of the period in question. What sets Chinese urbanism apart from the urbanisms elsewhere is the persistently strong role of the state in shaping its character that began with China's first cities in the Shang and Zhou periods some 4000 years ago. Throughout Chinese history, the administrative function of the cities has always been extremely strong, and this has directly affected the growth of the size and number of cities and the nature of Chinese urbanism, especially the urban economy, the structure of urban space, and the composition of urban population. Traditionally, the Chinese state has always been stronger than either the market or the society, and there is no sign that this will change in the foreseeable future. The powerful role played by the Chinese state in shaping urbanism is particularly evident from the fact that the contemporary forces of urban restructuring in China, including globalization of production, marketization of the means of production, privatization of firms and housing, and internationalization of urban planning, have all been allowed to play a role with the blessing and support of the state.

See also: Neoliberalism; Urbanism; Urbanization.

Further Reading

Chan, K. W. (1994). *Cities with Invisible Walls: Reinterpreting Urbanization in Post 1949 China*. Hong Kong: Oxford University Press.

Chang, K. C. (1986). *The Archaeology of Ancient China*. New Haven, CT: Yale University Press.

Chang, S. D. (1963). The historical trend of Chinese urbanization. *Annals of the Association of American Geographers* 53, 109 143.

Childe, V. G. (1950). The urban revolution. *Town Planning Review* 21, 3 17.

Demattè, P. (1999). Longshan era urbanism: The role of cities in predynastic China. *Asian Perspectives* 38, 119 153.

Esherick, J. W. (ed.) (2000). *Remaking the Chinese City: Modernity and National Identity, 1900 1950*. Honolulu: University of Hawaii Press.

Hao, Y. P. (1986). *The Commercial Revolution in Nineteenth Century China: The Rise of Sino Western Mercantile Capitalism*. Berkeley, CA: University of California Press.

Knapp, R. G. (2000). *China's Walled Cities*. Oxford: Oxford University Press.

Li, L. and Chen, X. C. (2003). *State Formation in Early China*. London: Gerald Duckworth & Co.

Ma, L. J. C. (1971). *Commercial Development and Urban Change in Sung China (960 1279)*. Ann Arbor, MI: University of Michigan, Department of Geography, Michigan Geographical Publication No. 6.

Ma, L. J. C. and Wu, F. L. (2005). *Restructuring the Chinese City: Changing Society, Economy and Space*. London: Routledge.

Skinner, G. W. (ed.) (1977). *The City in Late Imperial China*. Stanford, CA: Stanford University Press.

Thorp, R. L. (1991). Erlitou and the search for the Xia. *Early China* 16, 1 38.

Wheatley, P. (1971). *The Pivot of the Four Quarters: A Preliminary Inquiry into the Origins and Character of the Ancient City*. Chicago, IL: Aldine Publishing Company.

Wu, F. L., Xu, J. and Yeh, A. G. O. (2007). *Urban Development in Post Reform China: State, Market, and Space*. London: Routledge.

Chinese-Language Geography

W. S. Tang, Hong Kong Baptist University, Kowloon Tong, Kowloon, Hong Kong, People's Republic of China

Glossary

Confucianism It is traditional Chinese society's most foundational body of thought, centered around the Confucian classics, which emphasizes the behavior of an individual within society.

Daoism It is a body of thought, attributable to Laozi, which focuses on the individual within the natural realm.

Fang Zhi It is a local gazetteer published by each county and province on a rich array of topics for the comprehensive description of a local setting.

Human–Heaven Harmony It refers to the Chinese conceptualization of the relation between human and nature, which emphasizes harmony at the expense of conflict.

Kejiao Xingguo It is a slogan coined by politicians and intellectuals and chanted by almost everyone since Western imperialism in the late 1800s, calling for scientific and technological development to revitalize the nation.

Wuxing In traditional Chinese philosophy, natural phenomena are classified into the five elements or five phrases to capture the interactions and relationships between them.

Yin* and *Yang It refers to, according to ancient Chinese philosophy and metaphysics, the two primal opposing but complementary forces found in everything in the universe.

A Long History of Chinese Human Geography

Recognition of Human–Land Relation during the Formative Stage

Human geography started to flourish in China more than 2500 years ago. The concept of 'geography' was first coined in the classic text of *Yi Jing* (or *I Ching, Book of Changes*), referring to the observation and understanding of the Earth's surface for the pursuance of human activities. This reflects an early recognition of the human–land relationship, in the language of '*Tian*' (the heaven) and '*Di*' (the Earth). The first regional geography text, *Yu Gong* (*Tribute of Yu*), was written around 500 BC. Ancient China was divided into nine provinces based on a survey of natural resources. For each province, the boundary, the mountains, the nature of the soil, the waterways, the kinds of products, field tax divisions, the names of tributes, and the minorities were described and the few waterways that provided routes of transporting the tributes were identified. The way China was regionalized implies some kind of administrative division. Besides, the concept of regional disparity, especially with respect to the nature of the soil, is also discernible in the text.

Constituting Part of the Official Historiography

Later, more systematic works in human geography can be found in more official historiographies. These include *Shi Ji* (*The Records of the Grand Historian*), written from 109 to 91 BC, by the *magnum opus* of Sima Qian (145–88 BC) and *Han Shu* (*History of Western Han*) by Ban Gu (AD 33–92). The former recounts Chinese history from the mythical Yellow Emperor to the contemporary Emperor Wu of Sima's time. Included in it is a biography, *Huozhi Liezhuan* (*Biography on Goods Multiplication*), the theme of which is to develop production, stimulate prosperity, and raise people's living standards. It represents an early attempt of economic geography and, together with the coverage of the location of cities, their conditions of development, etc., an urban geography text. *Han Shu*, in contrast, covers the Western Han period only, and includes in it a special appendix, *Dili Zhi* (*The Geographical Gazetteer*), which details accounts of the physical landscape, water conservancy, resource extraction, population census, changing territorial boundaries, and tourist resources. This inclusion initiates a new standard of practice in the official historiography of each dynasty. *Dili Zhi* started to be published independently since the Western Jin Dynasty (AD 265–316). *Fang Zhi* (local gazetteers) refers to the systematic account of the topography, history, and socioeconomic situation of any local administrative district. Originally restricted to districts of special concern from the point of view of any dynasty, *Fang Zhi* was later regarded as an essential documentation of each county, prefecture, and province since the tenth century. In the Qing Dynasty (AD 1644 –1911), a special office was even established to coordinate the tasks of writing *Fang Zhi*, which were also standardized with principles, methods, patterns, and contents. Since its publication, the numbers of *Fang Zhi* had accumulated to over 10 000 volumes by 1949.

Enriched by Foreign Explorations, Official, and Civil Alike

In addition, there had been many accounts of exploration beyond the Chinese territory undertaken by orders of an emperor or by missionaries and travelers since the early

Chinese history. Zhang Qian was ordered thrice (139–119 BC) by the emperor of his time to explore the Mediterraneans for geopolitical considerations. He described the land route across inner Asia and then to the Mediterranean shore, including the ethnic minority regions in the Yunnan–Guizhou plateau. Once the Silk Road was opened, Buddhism spread eastward to China, while many Buddhist monks from China went westward to India to learn more about it. So went, for example, Fa Xian (AD 399–413) and Xuan Zang (AD 627–644). Their publications detail the topography, city defense, transport, agricultural production, climate, people, cultural exchanges, etc., of many places. Besides, other Chinese explorers also went by sea. The most remarkable one is Admiral Zheng He's seven separate expeditions, between 1405 and 1433, to Southeast Asia, Ceylon, the west coast of India, and as far as the Persian Gulf, the Red Sea, and the east coast of Africa. Although his three books were not as detailed as local gazetteers, they represent valuable additions to the Chinese human geography literature. The last, but not the least, is Xu Xiake, an intellectual not interested in being incorporated by the examination system, who traveled around the country from 1613 to 1639 and systematically recorded all the physical and human geographical phenomena that he observed in his trips.

The Late Penetration of Western Environmental Determinism

The influences on human geography in China also came from the West. Dated back to the Italian Jesuit Matteo Ricci's first visit to China in 1528, there had been missionaries from the European Christian churches, bringing along Renaissance informed modern geographic thought. Another wave of interaction with the West came after imperialism from the mid nineteenth century to the beginning of the twentieth century. During this period, there were more than 100 geographical expeditions and surveys in China led by scholars from Germany, Britain, Sweden, Hungary, France, United States, Russia, and Japan. Some of their works have contributed to the understanding of human geography in China. Besides translations of Western geographical works into Chinese, a means of 'using the advantages of the "barbarians" to subdue them', China did send students abroad to study geography (partly under the Boxer Rebellion Indemnity Fund), growing out of the burning desire to employ the more advanced Western technologies to build a stronger nation. The most prominent figures include the Harvard geology doctorate Zhu Kezhen (1890–1974), Edinburgh–Cambridge–Glasgow–trained geologist Ding Wenjiang (1887–1936), and the Leuven geology doctorate Weng Wenhao (1889–1971). (One may add Zhang Shuangwen (1866–1933), who introduced Western geography to China through his command of the Japanese language and, then, translation of the Japanese geography literature.) Given the then dominance of Darwinism and, in turn, environmental determinism, the Chinese were introduced to the Western human–land relationship with Paul Vidal de la Blanche's humanism being a better version of it. In addition, due to their background, advances in human geography were weaker and fewer than those in its physical counterpart. Most of the works on human geography concentrated on the recording and general description of geographical objects. This is especially the case of settlement geography, which resembles *Fang Zhi* both in methodology and in contents. Seldom were there more theoretical understandings. Nevertheless, due to the strenuous efforts of these scholars, syllabi on geography education were enriched, the professional geography society – the Geographical Society of China – to promote the discipline was founded in 1934 (together with the launching of its journal *Acta Geographica Sinica*), and the Institute of Geography was established in 1941.

Reinforced by the Soviet Practical Orientation

The Chinese in Yan'an (Yan'an is located on the Loess Plateau in northern Shaanxi province and it was the seat of the Central Committee of the Communist Party from 1937 to 1947) and other liberation areas already had an early taste of the Marxist informed human geography, which approached geographical phenomena from the perspectives of colonialism and unequal development. Most people, however, did not really feel its impacts until the 1950s, when the Soviet influences loomed large in the newly established socialist nation. Unlike in the past, the Soviet influence had been more on the technical means of resolving practical issues than on theoretical and moral geographical knowledge. Before any industrialization program could proceed, knowledge about resources in the prospective region must be made available. There were then nationwide, integrative, regional surveys, spanning geography, geology, climatology, botany, zoology, soil, hydrology, agriculture, forestry, and economy, in which geography played an important role. Influenced by the development within the Soviet geography discipline itself, and propagated by the concept of territorial production systems in particular, human geography in China was dominated by economic geography which, in turn, concentrated on agricultural geography, industrial geography, and transport geography only. As the development strategy was transformed to 'walking on two legs' (industry was emphasized not at the total expense of agriculture) and, later, the Great Leap Forward and its aftermath, so did economic geography shift to focus primarily on agricultural geography in the 1960s and 1970s. The latter contributed to land use planning, agricultural resource assessment, and agricultural regionalization.

Yet, this development in human geography during the first thirty years of socialism should not be interpreted one sidedly; the influences of the Soviets were there for sure, but the Chinese traditional practice did make a difference too. Since the early 1950s, there were mass campaigns, one after another, to purge this or that category of citizens. The intellectuals were purged, while Western texts and ideas were adjudicated. Human geography was no exception to this rule. Between 1950 and 1953, *Acta Geographica Sinica* published only a few articles on human geography. They were all subject to open and severe criticism. In response, human geographers either converted themselves into physical geographers or merely stopped writing so as to avoid unnecessary scrutiny, harassment and, even worse, purge. This helps explain the comparatively sluggish development of human geography during this period.

Propelled by the Quest for Expansion in Productive Capacity

As the nation sought to recover from the brink of total collapse after the Cultural Revolution and started to recontemplate economic growth again, geography was called into service in the mid 1970s to take stock of resources within the territory. In particular, the Institute of Geography contributed to the large scale surveys of many regions in the country and, especially during the early 1980s, the formulation of master plans for many medium and small sized cities. In 1986, it was even officially announced to subordinate the Institute of Geography under the joint supervision of the China Academy of Sciences and the State Planning Commission. During these ten years, human geographers like Li Xudan emphasized understanding human geography in the context of human–land relations and economic geographer Wu Chuanjun highlighted the consideration of local specificity in territorial development and management. The publication of the specialized journal *Economic Geography* since 1981 and *Human Geography* from 1986 marked the diversification as well as resurgence of human geography. These attempts redressed the imbalance of the negligence of local details in public policy in the previous two decades. Concomitantly, besides economic geography, the quest for knowledge about population, natural, and tourist resources is responsible for the establishment as well as promotion of population geography, land resource geography, and tourism geography.

While cities had always played a recipient role in the nation's first thirty years of economic development, they are now considered active agents of growth in themselves as well as for their surrounding countryside. The rapid pace of urbanization alerted the national economic planners to questions like the relationship between the level of urbanization and the level of economic development, rural–urban migration, the urban size hierarchy, and the functional structure, catchment area, and spatial structure of cities. The complexity of the urban question required a more integrated approach to tackle it, leading to the birth of the multidisciplinary urban studies (including economics, sociology, anthropology, politics, history, architecture, and urban planning). The spatial, materialist perspective of geography, as persuasively put forward by Song Jiatai in his argument of the regional foundation of urban development, allowed geographers to dominate the urban discourse at least in the 1980s and 1990s, and so enhanced the subdiscipline urban geography. As urbanization grew even faster, urban geographers started to investigate a whole array of issues such as land assessment, commodity housing, urban redevelopment, suburbanization, the impact of foreign direct investment (FDI), urban governance, world city formation and, even, spatial inequalities. Lately, with the growth of the 'middle class' in the city, there are attempts to explore the issues of mass consumption, culture, and differences. Besides, as intellectual exchanges with the West broaden, urban geographers begin to employ quantitative techniques and other methodologies (such as behavioral and qualitative) in their analyses. (Since it embraced capitalism more warmly as well as being informed by the critique of colonization since the late 1970s, human geography in Taiwan has lately been introduced with some elements of critical geography.)

The above description of the development of human geography in China can be supplemented by another reference to the number of papers published in *Acta Geographica Sinica* for the 70 years between 1934 and 2004 (the journal was not published between 1967 and 1977). Although a total of 2239 papers were published, less than 24% of them could, according to one classification system, be categorized as human geography (**Table 1**). In other words, human geography had never occupied a more prominent position in the discipline during the contemporary period, especially under socialism. A breakdown of the papers on human geography between 1934 and 1988, as elaborated in **Table 2**, illustrates that urban geography recently grew faster than other subfields of human geography. (There are some inconsistencies in the total number of papers on human geography between the two tables due to slightly different classification systems.)

Characteristics of Chinese Human Geography

The Dominance of Human in the Human–Land Relation

During the dynastic past, the traditional Chinese culture had made three significant imprints on human geography

Table 1 The total number of articles published and its counterpart on human geography

	1934 48	*1950 66*	*1978 89*	*1990 2004*	*1990 2004 supplements*	*Total*
Total number of papers	146	363	435	1068	227	2239
Human geography	50	82	95	224	55	526
%	34.25	22.59	21.84	22.85	24.24	23.49

Source: Adapted from He, S. and Zhao, X. (2004). Analysis of the articles on human geography published in *Acta Geographica Sinica* in 1934 2004. *Acta Geographic Sinica* 59 (Supplement), 198 204.

Table 2 The changing distribution of articles published in different subfields of human geography

Subfields	*1934 48*	*1949 66*	*1978 88*
Economic geography and land reform	4	17	16
Agricultural geography	9	35	24
Industrial geography	1	2	9
Transport geography	5	3	6
Commercial geography	1	0	0
Human geography	4	4	3
Population geography	6	2	2
Urban geography	5	3	17
Rural settlement geography	4	1	1
Historical geography	0	11	3
Tourists geography	2	0	0
Ethnic geography	4	0	0
Political geography	0	1	0
Place names geography	2	1	1
Total	47	80	82

Source: Yeung, Y. M. and Zhou, Y. (1991). Human geography in China: Evolution, rejuvenation and prospect. *Progress in Human Geography* 15, 373 394.

in China. First, unlike the West, the discipline has a strong human dimension. This is best exemplified in the harmonious 'human–land relation'. According to the oldest Chinese thinking, the nature was the object of veneration. Being dominated by the nature, human ac tivities should follow it. This thinking was soon overtaken by another that emphasized human agency. Having the ability to empower nature just like nature has the ability to influence human activities, people can subdue nature for their own cause. However, informed by the 'human–heaven harmony' cosmology (constituted by both the Confucian and the Daoist components), people, being part of heaven, must realize that there is a limit to the exploitation of nature, beyond which people will suffer from their own deed. Nature and people merely create resonance in each other. That is why people should co exist peacefully with nature. This is a naïve 'human–land relation', which put humans in the priority, yet under scores the local conditions in implementation. An ex ample of this practice is the widely acclaimed dyke pond system in South China. While the latter undoubtedly exhibits a close harmony between human and nature, one should not neglect the fact that it was organized to en hance humans to produce more resources, including silk and aquatic produces. This is very different from the comparatively more land centered human geography of the West as well as its alleged separation between human beings and nature.

Conceptualization Is Based on Concrete Sense Impressions

The Chinese geographical conceptualization has favored the most concrete sense impressions at the expense of abstract speculation. Very early on in her history, China had already developed a naïve, but overarching, schema that explains the universe's dynamism in terms of basic forces in the nature, namely the complementary agents of *Yin* (dark, cold, female, negative) and *Yang* (light, hot, male, positive), and *Wuxing*, the five elements or the five phases (water, fire, wood, metal, and earth). Central to the latter are the two interactive cycles of generation (wood feeds fire; fire creates earth (ash); earth bears metal; metal collects water; and water nourishes wood) and restraining (wood parts earth; earth absorbs water; water quenches fire; fire melts metal; and metal chops wood). Most forces in the nature are seen as having the states of *Yin* and *Yang*, which are opposite, mutually nonexclusive, inter dependent, dynamic, and transformative. *Yin* and *Yang* so understood are not discrete duality, as one is inter penetrated into the other. It is the interaction of *Yin* and *Yang* that fuels the dynamics of *Wuxing*. The world, which is constantly changing and full of contradictions, must be understood relationally. To understand and appreciate one state of affairs, it is difficult to exclude its opposite, as what seems to be true now may be the opposite of what it seems to be. These basic forces of *Yin*, *Yang*, and *Wuxing* had in the past been applied to describe interactions and relationships between phenomena, be they seasonal changes, agricultural production, human relationships, or social institutions.

This methodology of understanding human activities favor concrete sense impressions. The custom of under standing nature and society holistically, in particularly from macro to micro and from big to small, builds on sense experience from the struggles in mostly agri cultural activities. Comprehending the pattern of any development by senses relies not so much on logic as on literary quotations with profound implicit meanings and lively metaphors to express and understand any object (e.g., the Chinese are fond of using 'the method of painting a dragon and dotting its eye' to convey the message of 'bringing out the salient point'). *Dao* ('the way') is something to be sensed after cumulative experiences, but not to be explained. Besides, the *Yin Yang* and *Wuxing* principle presumes no discrete boundary between an object and its external forces. Everything must be understood in complexity. In particular, the contradictions and their transcendence or integration as implied in *Yin* and *Yang*, a type of non Hegelian dialecticism, ensure that any object cannot be understood outside its context. Thus any attempt of de contextualization, such as analyzing the attributes of an object and categorizing it on the basis of its abstracted attributes with great precisions, was discouraged. Concomitantly, there was little discovery of laws that were capable of explaining classes of events. As will be argued in due course, the fact that the Chinese did not institutionally encourage debates also hindered the development of logics, the prerequisite for discovering truth. In fact, it was not necessary for the Chinese geographers to get their hands dirty, as they were equipped to resolve many epistemological issues by the overarching schema at a very early stage of development. In sum, the field of human geography was satisfied with vague, sensual understanding while showing a distaste for analytical explanation of individual phenomena.

Be Practical – For State Governance in Particular

Chinese human geography was practical in orientation and for the pursuance of state governance in particular. Undoubtedly, this was influenced by *Dao*, 'the way'. The aforementioned characteristics of human orientation and concrete sense impressions encourage Chinese geographers to assign top priority to phenomena that are either closely related to human activities or amenable to human control and reconstruction. This bias becomes comprehensible in the context of the state–society relations. Emperors of different dynasties, sons of heavens, practiced the emperor's way (*Huandao*). Since the latter intermediated the heaven's way (*Tiandao*) and human's way (*Rendao*), the emperors were justified to rule the people on behalf of the heaven. They, therefore, had an absolute authority over their subordinates, not only their right to live or die but also their right to cultural life. Whenever there was a debate among competing arguments or thinking, the power of arbitration, or even adjudication, also rested with the emperors. Since Confucian times, then, intellectuals in China had understood their limited roles in the society, not so much pursuing the truth as elaborating policies and devising techniques of implementation for state governance, and recognized the danger of claiming even a low intellectual autonomy. As the intellectuals put this mentality into practice, *Dao* that concerns less with the discovery of truth than with the finding of guidance to live in the world was reinforced. Obviously, the geographers could pay attention to nothing but practical issues.

The major concern of the Chinese, dated back to the ancient time, has been state building. To one's surprise, China in the dynastic past had been less unified than a casual historiography would have us believe. Geographically, China was basically divided into three big cultural regions: agriculture in the central, lower regions; animal grazing in the northwest, plateau regions; and fishing and hunting in the northeast, mountainous regions. It was the interaction of these three cultural groups over time, sometimes in war and the concomitant breakup, and sometimes in peace and then coexistence, that had written and rewritten the long Chinese history. During many periods, as mostly before the Song dynasty (i.e., before tenth century), there were boundaries rather than borders due to incessant fighting among competing kingdoms. The country was basically a class divided society in that the political elite was separated from the rest of society, the town from the countryside, and the center from the periphery. The emperor was interested in maintaining political stability by extending his control, usually with some degree of autonomy, to the periphery via the establishment of the local administrative system of counties. In other periods, as between Song and early Qing dynasties (i.e., from tenth to late nineteenth century), there were more or less clearly defined territories. One saw the increasing centralization of arm forces, but due to regionalization of the socioeconomy, localism grew and local resistance ensued. Therefore, it was still difficult for the state to carry out effective surveillance of the population in every corner of the country. Still in other periods, as from late Qing dynasty to the establishment of the People's Republic of China, the territory was identified and recognized by other nation states in the world. Besides mobilization of the means of production for industrialism, there was the centralization of not only means of military violence but also surveillance. The population started to be more vulnerable to state control. Yet, due to the incessant challenge from foreign powers and the fractionalization of internal political coherence, the development of the nation state had not really been completed. The above has shown that China, from the ancient time to the establishment of the socialist nation, had a long history of extending the state power over the territory and its population. At any time, it was imperative for the emperor or any group leader to make sense of one's situation and formulate strategies so as to unify the country. This reliance on practical knowledge for the secular end of governing is best captured in the oft cited slogan since the end of the nineteenth century: "To revitalize the nation by science and education" (*Kejiao Xingguo*).

This emphasis has directed the discipline of geog raphy to concerns that are one way or other related to state building. Practical knowledge about dynastic pol itics and the daily life of people was emphasized at the expense of the natural processes underlying the for mation of the earth's surface. Chinese geographers were long interested in the ownership and administrative division of land surface. The concerns of *Yu Gong* fall exactly into these domains. Cartographic production in the past dwelled on fine details of location for adminis trative control including mountain ranges, river drain ages, major towns, strategic defense posts, roads, and traffic. The contents of *Fang Zhi* consist of various issues of administrative control, from types and quantities of agricultural produce, the amount of arable land, eco nomic situation, designation of cities and towns, local taxes, to administrative division. Besides, the Chinese geographers were interested in the practical knowledge about people's daily life. The practical interest in agri cultural production has induced them to research into the relations between the climatic and hydrological conditions, on the one hand, and, on the other, sowing seeds, growth, and harvest of agricultural produces.

While the traditional Chinese culture was instru mental in the development of human geography in the dynastic past, the imperative of the nation state is de cisive under socialism (and, to some extent, during the first few decades of the twentieth century too). The so cialist state has managed to put its population under surveillance. This measure is implemented together with the centralization of capital and land for industrialization. On the one hand, the quest for order across the territory requires desperately practical knowledge not only about the availability of all means of production – the types, quantities, quality, and location – but also the possibility of combining them across space and over time. The magnitude and complexity of geographical knowledge so required differ from those of the dynastic past. On the other, the intimacy between the state and the society reduces further intellectual autonomy. Knowledge pro duction is under state surveillance in two related ways. The state and its local agencies 'order' the production of knowledge by researchers, whom they directly employ. As a matter of fact, as aforementioned, the Institute of Geography was regrouped under the auspice of the State Planning Commission in 1986. On top of this, there were occasional campaigns to purge the researchers, ensuring that they are under some kind of control. Un doubtedly, things have improved these days, but the quantitative changes have not led to any qualitative changes. As a result, the drive for truth is practically difficult. The fact that the slogan of calling researchers to contribute geographical knowledge to the promotion of a harmonious society by adopting the scientific develop ment worldview was chanted in one of the recent party committee meetings is a testimony to this argument. Examples of reducing the more academic subjects of various subdisciplines to corresponding practical knowledge abound: tourism geography becomes tourist development planning and management; population geography becomes population census and planning control; economic geography becomes economic distri bution and development planning; etc. In other words, the three characteristics of human geography in the dynastic past – human , sense , and practical oriented – could also be found under socialism, even attaining higher degree of maturity.

See also: Culture; Geographical Journals; Language; State.

Further Reading

Chiang, T. C. (2005). Historical geography in China. *Progress in Human Geography* 29, 148 164.

Teather, E. K. and Chow, C. S. (2000). The geographer and the fengshui practitioner: So close and yet so far apart? *Australian Geographer* 31, 309 332.

Whitehead, J. W. R. and Gu, K. (2006). Research on Chinese urban form: Retrospect and prospect. *Progress in Human Geography* 30, 337 355.

Wu, C. J. (1990). The progress of human geography in China: Its achievements and experiences. *Geojournal* 21, 7 12.

Yeung, Y. M. and Zhou, Y. (1991). Human geography in China: Evolution, rejuvenation and prospect. *Progress in Human Geography* 15, 373 394.

Choice Modeling

J.-C. Thill, University of North Carolina at Charlotte, Charlotte, NC, USA

Glossary

Choice Set A set of discrete choice alternatives available to an individual.
Markov Process A mathematical framework of discrete-time stochastic process characterized by a set of states.
Maximum Likelihood A statistical method to make inferences about utility parameters from a dataset.
Random Utility Utility or satisfaction of a choice alternative, comprised of a deterministic component (systematic utility) and a random term.
Rational Behavior A consistent and deliberate decision process driven by the maximization of an individual's utility.
Systematic Utility The deterministic portion of the utility of an alternative.

Introduction

A conventional premise of microeconomic theory is that demand and supply are traded off through continuous adjustments in quantities. Individual demand is determined by calculus methods that maximize consumer utility (i.e., satisfaction), subject to constraints. By the 1970s, it had become clear from the pioneering work undertaken in transportation, marketing, and economics that many consumer decisions do not conform to this postulate of marginalist consumer theory. Individual decisions actually often involve a finite number of options that are indivisible, mutually exclusive, and collectively exhaustive. The determination of a 4 year college one is to attend upon high school graduation; the car make to buy; whether to buy a car, lease it, or neither; the commitment to marry; and the choice of a travel mode for one's daily commute are all examples of discrete choices that cannot be comprehended through the lens of conventional decision theory. Choice modeling is the theory of individual decisions among discrete alternatives and its empirical derivatives in the form of measurement procedures and estimation methods. Although choice models are commonly based on the postulate of rational behavior (a normative stance), they purport to be a faithful description of the expected behavior of individuals.

From rather narrow beginnings rooted in the econometrics of qualitative response variables and in the mathematical psychology of rational behaviors, choice modeling first addressed simple choice situations (for instance, 'where shall I do my grocery shopping this week?'), but quickly gained in robustness to cope with complex, multilevel systems of intertwined and constrained decisions. Choice models continue to contribute to the social science understanding of how human and social systems take shape from the coalescence of decisions of multiple actors (consumers and producers) based on utilitarian principles. They are also valued as operational models in various areas of applied spatial sciences, such as retail analysis, transportation, business site selection, housing, and real estate.

Human geographers mistakenly refer to spatial interaction models as choice models on occasions. If indeed equivalence has been established between certain forms of choice models, such as the multinomial logit model (MNL) presented hereunder, and entropy maximizing models of spatial interaction, the analogy is nevertheless in the algebraic form of the two models and not in the implied behaviors. Spatial interaction modeling is the subject of another article.

Theoretical Foundations

A simple choice problem can be conceptualized as the selection of one alternative from a predefined set of available discrete choice alternatives that are mutually exclusive and collectively exhaustive, which is also known as the choice set. In keeping with the view on consumer economics that utility cannot be derived from commodities *per se*, but from their objectively measurable characteristics, the utility of each alternative is evaluated on the basis of their intrinsic attributes as well as on the personal characteristics of the decision maker. The latter dependency accounts for interpersonal differences in the subjective perception and evaluation of alternative attributes. Following Thurstone's random utility theory, it is typically posited that, while individuals deterministically choose the alternative with the greatest utility, the utility of an alternative cannot possibly be known with full certainty. Unobserved preference factors and measurement errors and approximations account for the random component of utility. Hence, the model cannot forecast what choice is made, but it can estimate the likelihood that a given option in the choice set be selected.

The actual calculation of choice probabilities is contingent upon the specification of the error term (random utility) of the model. In early studies on discrete choice modeling, it was assumed for convenience sake that errors are independently and identically distributed as a double exponential function. Choice probabilities then take an elegant closed form known as the MNL. More formally, if n denotes an individual and i is an alternative, the probability that individual n chooses i has the form:

$$P_n(i) = \frac{e^{\beta_i' X_{in}}}{\sum_j e^{\beta_j' X_{jn}}} \qquad [1]$$

where β is a vector of parameters to estimate empirically and X is the vector of measurable characteristics of the alternatives that are variables in the systematic utility. For empirical reasons, the MNL is attractive and has become the workhorse of choice modeling due to its tractability and the relative ease with which its parameters can be estimated. Models based on other distributional assumptions of errors remain, to this day, less tractable. Assuming that errors are normally distributed, the so called probit model has grown quite in popularity in recent years.

One of the noteworthy properties of the MNL that stems from the assumed distribution of errors is that it satisfies Luce's choice axiom and, therefore, exhibits independence from irrelevant alternatives (IIA). This entails that the ratio of choice probabilities of any two alternatives depends only on their systematic utility components and is totally unaffected by other alternatives that may be available to the decision maker. According to the choice axiom, no allowance whatsoever is made for different degrees of substitution or complementarity among choice alternatives. Consequently, any new alternative entering the choice set draws market share equally from each alternative already in the choice set. Whether the IIA property holds in the practice of specific decision situations has been the subject of considerable debate. If it does not hold, MNL predictions are biased and can be behaviorally counterintuitive. A famous illustration of this is Debreu's red bus/blue bus paradox whereby the introduction of a new bus service (red bus) competing against an existing bus service and private automobile would be forecasted by an MNL model to grant one third market share to each transport alternative, while intuition dictates that car would maintain a 50% market share and the balance would be split equally between the two bus services. This has led to alternative approaches to modeling discrete choices that impose less restriction on structures of substitution between alternatives.

Empirical Validation and Statistical Estimation

The implementation of discrete choice theory to real world decision situations involves significant measurement and statistical estimation issues. From the onset, two broad approaches have prevailed. In discrete choice models *sensu stricto*, actual choices are observed by survey and independent variables entering the utility function are either objective measures or subjective evaluations of the various choice alternatives of a prespecified choice set. The parameters of the utility function and the choice probabilities are estimated through data fitting procedures, usually a maximum likelihood method, so as to reveal preferences through calibration on overt behavior. This perspective has been very popular in transportation, housing, shopping, and other locational studies.

Alternatively, decompositional multiattribute preference models are estimated on the basis of an individual's stated preference ordering for a set of hypothetical choice alternatives. Choice set alternatives are composed according to principles of experimental design so as to insure statistical representativeness. Decompositional preference models first estimate the utility function of each individual on the separate contributions of each attribute level (using least squares regression analysis, analysis of variance, or linear programming), and apply a choice model to the resulting preference structure to simulate a choice outcome.

Each of the two approaches has different advantages and disadvantages. Because decompositional preference models are based on principles of experimental design, it is recognized that they can better account for the interpersonal heterogeneity in preferences. They can also handle choice alternatives that are beyond the realm of observation because they do not currently exist. This places them in a unique position to predict the success of new choice options, such as new forms of public transportation services, new electric automobile technologies, new supermarket concepts, or new concepts of housing developments, such as transit oriented developments and mixed use complexes as responses to the challenges of greater urban sustainability. It is undeniable, however, that the experimental nature of data collection in decompositional models renders them more complex, while the validity of their prediction of actual choices may be inferior to that of discrete choice models *sensu stricto*. Estimation procedures for combining revealed and stated preference data, along with psychometric data (i.e., attitudes and perceptions), have been available since the early 1990s and have been instrumental in enhancing the behavioral representation of the choice process.

Modeling Complex Choices

Admittedly, the choice model embodied by eqn [1] raises a number of issues regarding its behavioral underpinnings and its ability to capture more than just the simplest choice decisions made by human beings. These issues have driven the agenda of research on choice modeling for the past three decades.

Choice Set Formation and the Decision Process

The model of eqn [1] suggests that the choice set is known beforehand and independently of the choice process. In fact, the choice set of any single decision maker is a subset of the universe of all alternatives that are both feasible to the decision maker (living in the White House belongs to the realm of fiction for all practical purposes) and known at the time a decision is made (hence the power of advertising to raise people's awareness). Accordingly, individual behavior can be framed as a two step process wherein a choice set is specified first, and a selection among available alter natives is then made.

Using the universal choice set may be a poor surro gate for the decision maker's choice set because it fails to account for constraints precluding certain alternatives, such as their physical availability (e.g., absence of pas senger train service between many US cities), time availability (e.g., a grocery store that is too remote for biweekly food purchases), affordability, and others. When a decision maker has no or limited knowledge of some perfectly feasible alternatives, these alternatives are in effect removed from one's choice set. Hence, proceeding with a model based on the universal choice set is bound to bias estimation results (a nonzero probability is falsely assigned to alternatives outside the choice set) and to impute to the nonselection of this alternative to indi vidual preferences and free will instead of unawareness and infeasibility. Given these considerations and for pragmatic reasons of data availability and computational power, reduced choice sets are employed instead of universal choice sets in many decision situations of relevance to human geography. *Ad hoc* narrowing, sam pling, and aggregation strategies are routinely imple mented with reasonable success as far as the selection among geographically defined alternatives is concerned. *Ad hoc* narrowing of the choice set may be based on some distance cutoff or some other independent and well in formed knowledge of behavioral decisions. By reducing the choice set to a manageable number of alternatives, these strategies are also instrumental in enabling the modeling of spatial choices, which routinely involve a very large number of possible alternatives (hundreds of restaurants in a mid to large size metropolitan area, tens of thousands of housing units for lease or for sale, etc.).

Having decision makers specify their own choice sets is a very controversial matter due to the inaccuracy of the information in the light of the fact that the choice set is not explicitly kept in memory by the decision maker. A somewhat less arbitrary and more systematic approach to the two stage choice problem process consists in the implicit approximation of the decision maker's choice set. This approximation proceeds by assigning a weight to each alternative to reflect its degree or likelihood of membership to the decision maker's choice set according to some preset rules such as geographic proximity or accessibility. Finally, the choice set can be explicitly de termined, in which case the choice process represented by eqn [1] is replaced by

$$P_n(i) = \sum_{C \; 1}^{M} P_n(i|C)P_n(C), \quad \forall C \quad \text{where } i \in C \qquad [2]$$

Here, the second term models a given choice set C from the universal choice set M as a function of personal and household characteristics of the decision maker (in cluding one's locational profile) and of characteristics of the alternatives in choice set C; the first term involves the evaluation of alternative i conditional on choice set C. This framework has been successfully implemented in various spatial choice contexts, including the choice of shopping destinations.

Substitution among Alternatives

Because of its property of independence of irrelevant alternatives, the MNL (eqn [1]) imposes that all alter natives be uniformly substitutable (equal cross elas ticities). The red bus/blue bus paradox illustrates the senselessness of this assumption in many decision situ ations. This is proven to be particularly the case with spatial decision making because choice alternatives ex hibit well defined spatial structures resulting from ad justments to the preference of decision makers for minimizing effort in travel and search activities and from adjustments to the decision maker's processing of infor mation on spatial alternatives into clusters based on geography. In other words, choice alternatives that are located nearby are closer substitutes than distant alter natives. A variety of choice modeling approaches exists that afford a richer pattern of substitution among alternatives.

First, the handful of choice models mentioned in the previous section that implicitly or explicitly capture the extent of the decision maker's choice set are not affected by the IIA property. The same can be said of the multinomial probit model; this model encapsulates the degree of substitution between pairs of alternatives through a variance–covariance matrix estimated on em pirical choice data. In certain choice situations, patterns of substitution among alternatives can be figured out without great difficulty on the basis of existing cognitive and behavioral theories and built into the choice model. The nested multinomial logit model (NMNL) assumes that alternatives can be sorted hierarchically according to their degree of substitutability. A classical case is that of the housing choice problem. **Figure 1** illustrates such a problem, where housing units are clumped into types (single family vs. apartments) in first instance because of the low level of substitutability between housing types. Then, within each type, differentiation based on the neighborhood emerges (N_1 through N_3), and finally each unit of each type in each neighborhood is evaluated on its own merit (U_1 and so on). The NMNL exhibits the IIA property for choice alternatives belonging to the same cluster (or nest). This model is a member of a broader class of choice models that also allow for cross elasticities between alternatives. The primary drawback of the NMNL is that the particular structure of substitutability among alternatives must be defined by the analyst be forehand, while the hierarchical order is sometimes un clear or arbitrary. It should be noted that the hierarchical structure that underpins the NMNL creates a form of conditionality of choice subsets and by the same token solves the problem of the choice set definition.

Relationships among Choices

It is not unusual at all for decisions to be interrelated. An individual's choice of housing location is intertwined with the location of the place of work. These choices in turn are influenced by a number of the individual's nonspatial decisions, such as the decision to marry, raise a family, and adopt a certain lifestyle. Individual decisions may also need to be coordinated within the framework of a household. Finally, other decisions of more limited scope may be affected: decision to buy a car or not, choice of the commute travel mode, choice of a commute route, etc. The multiple facets of travel decisions have also been recognized as an area where an integrated modeling approach is beneficial. Simple origin–destin ation trips are subsumed by trip chaining and household daily activity patterns, where activity scheduling, stop frequency, timing and location, routing, and modal choices come together. Similarly, choices made by private business or government entities are seldom the results of isolated decisions.

The integration of complex sets of decisions within the framework of random utility theory has followed three main lines of research. In the first and simplest, complex decisions are modeled jointly, in a single step. The decision maker is seen as selecting from a super choice set formed of the Cartesian product of choice sets defined on each elemental choice dimension. For in stance, the integration of residential location choice (three neighborhoods to choose from) and of the choice of a commute travel mode (three modes available) creates a choice set made of nine composite housing location/ travel mode options. As a variant of this approach, trip chaining behavior can be modeled as a Markov process, with probabilities of transition between states rendered by a joint choice model.

Evidently, the joint modeling approach may not be well suited to the complexity of the multidimensional decisions due to the IIA property. With a decomposition of the choice set into hierarchical groups, a nested choice model captures the complexity through a series of simple choice models conditioning decisions along dimension to choices modeled at earlier stages. Finally, for decision situations where some choices are on a continuous scale (e.g., weekly commute time), mixed continuous/discrete choice models can be estimated.

Other Decision Rules

Random utility choice models make specific assumptions on the internal mechanisms used by decision makers to process information and derive a choice. The customary linear utility function supposes trade offs (or compen satory offsets) among attributes of choice alternatives, on the basis of which a deterministic decision is made that

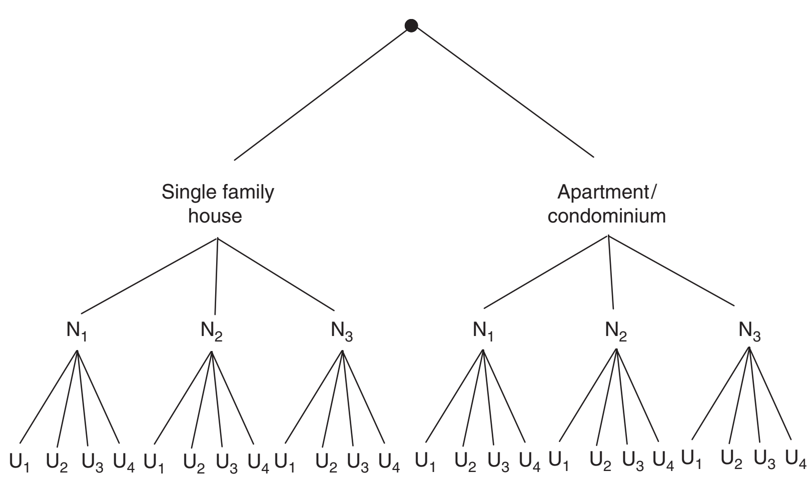

Figure 1 Nested structure of housing choice alternatives.

maximizes utility. In this view, poor performance of an alternative on one choice criterion can be compensated by a good performance on one or more other criteria. Various decision rules that do not espouse the principle of full compensation between attributes have been suggested. Some rules may require that an alternative must exceed specific values of each attribute to be chosen (e.g., a house should have a sale price under \$300 000 and be no more than 1 mile away from work). Alternatives that fail to reach a certain attribute level are rejected as unacceptable, even if they outperform other alternatives on some other attribute. Some other rule systems are less stringent and consider an alternative as worthwhile if it meets a preset threshold on one attribute only, regardless of its performance on others (a house should have a sale price under \$300 000 or be less than 1 mile away from work). Another popular rule system is the elimination by aspect. This rule proceeds in a stepwise fashion, considering each attribute in turn from the attribute deemed the most important to the least important. The selection process eliminates all alternatives that do not satisfy a threshold defined on the selected attribute and continues until a single alternative remains. Interestingly, these alternative decision rules can also serve as methods for defining choice sets.

Attribute trade offs are not the only property that sets apart rule based and random utility models. While the latter are parametric models of choice behavior, the former emphasize the algorithmic methods of information retrieval, representation, processing, and fusion that lead to a unique choice decision. Not surprisingly, diverse constructs of data mining and knowledge discovery have become prominent in implementing non compensatory choice models, such as decision trees and related methodologies. With decision trees, sequential sets of conditions on attributes are applied so as to identify the choice strategies that lead to alternatives being selected. Complex learning algorithms of rule induction developed in artificial intelligence have dramatically expanded the horizon of possibilities in modeling choices from a process oriented perspective. Some hybrid semicompensatory rule based systems even have the flexibility of blending random utility maximization with decision thresholds. Once calibrated on empirical choice data, rule based systems can be built into computational process models that simulate real world choice behavior over the full range of spatio temporal scales.

Decision Making and Modeling Choices

While this may not patently transpire from the discussion held thus far in this article, choices are the results of a decision making process involving multiple sets of considerations. Behavioral geographers and environmental psychologists have worked at crafting the conceptual framework that integrates the actual act of selecting an alternative with the mental and cognitive processes that operate under the influence and constraint of the ambient social and physical environment of the decision maker. **Figure 2** schematically represents such structure. Choice models have traditionally addressed the relationship between individual preferences on the one hand, and the evaluation of alternatives and the resulting choice of an alternative on the other, at the exclusion of all other considerations reported in the schematic framework. To this day, the theory of choice modeling still has a limited ability to incorporate these other considerations and, when it does, it is in a rather piecemeal fashion. The state of the art in choice modeling can positively be linked to so called hybrid choice models that integrate multiple models of choice under constraints and models of cognitive and psychological factors in a recursive structure. Owing to their flexibility and structural

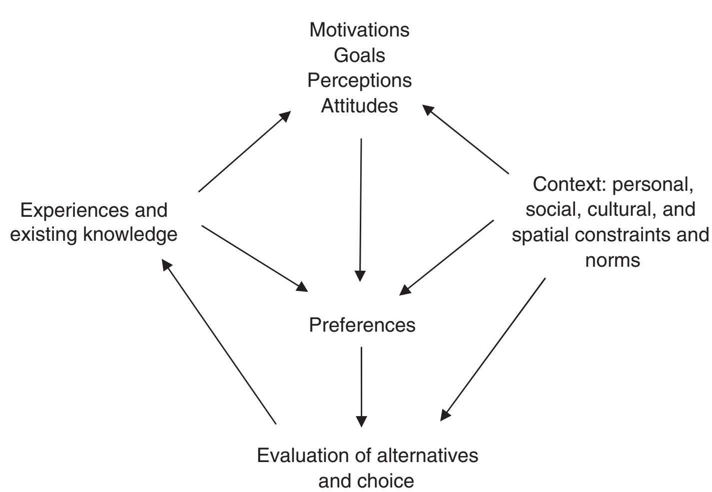

Figure 2 Conceptual framework of decision-making process.

similarities to the framework depicted in **Figure 2**, simulators and computationally intensive numerical procedures (e.g., artificial neural networks) can be expected to play a significant role in tomorrow's choice modeling.

See also: Behavioral Geography; Cognitive Geography; Neural Networks; Rational Choice Theory (and Rational Choice Marxism); Spatial Interaction Models; Time Geographic Analysis.

Further Reading

Ben Akiva, M. and Lerman, S. (1985). *Discrete Choice Analysis: Theory and Applications to Travel Demand*. Cambridge, MA: MIT Press.

Ben Akiva, M., Bradley, M., Morikawa, T. *et al*. (1994). Combining revealed and stated preferences data. *Marketing Letters* 5, 335 350.

Bhat, C. R. (2000). Flexible model structures for discrete choice analysis. In Hensher, D. A. & Button, K. J. (eds.) *Handbook of Transport Modelling*, pp 71 90. Elsevier: New York.

Domencich, T. and McFadden, D. (1975). *Urban Travel Demand. A Behavioural Analysis*. Amsterdam: North Holland.

Golledge, R. G. and Stimson, R. J. (1997). *Spatial Behavior. A Geographic Perspective*. New York: Guilford.

Hensher, D. A., Rose, J. M. and Greene, W. H. (2005). *Applied Choice Analysis: A Primer*. Cambridge: Cambridge University Press.

Louviere, J. J., Hensher, D. A., Swait, J. D. and Adamowicz, W. (2000). *Stated Choice Methods: Methods and Applications*. Cambridge: Cambridge University Press.

Luce, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology* 15, 215 233.

Manski, C. F. (1977). The structure of random utility models. *Theory and Decision* 8, 229 254.

McFadden, D. (1974). Conditional logit analysis of quantitative choice behavior. In Zarembka, P. (ed.) *Frontiers in Econometrics*, pp 105 142. New York: Academic Press.

McFadden, D. (1978). Modeling the choice of residential location. In Karquist, A., Lundquist, L., Snickars, F. & Weibull, J. L. (eds.) *Spatial Interaction Theory and Planning Models*, pp 75 96. Amsterdam: North Holland.

Pitfield, D. E. (ed.) (1984). *Discrete Choice Models in Regional Science*. London: Pion.

Recker, W. W. and Kostyniuk, L. P. (1978). Factors influencing destination choice for the urban grocery shopping trip. *Transportation* 7, 19 33.

Timmermans, H. J. P. and Golledge, R. G. (1990). Applications of behavioural research on spatial problems II: Preferences and choice. *Progress in Human Geography* 14, 311 354.

Relevant Websites

http://www.jocm.org.uk
Journal of Choice Modelling (a new open access journal covering theory and practice in the field of choice modelling).

Christaller, W.

R. Capello, Milan Polytechnic, Milan, Italy

Walter Christaller (**Figure 1**), a German geographer and one of the founding fathers of the 'central place theory', was born on 21 April 1893 in Berneck in the Black Forest, and died on 9 March 1969 in Königstein in Taunus (Hesse), at the age of 75. His mother belonged to a cultured middle class family of Darmstadt and in her mature years she was a successful novelist. His father, relatively poor, was a clergyman and theologian, but he became involved in doctrinal conflicts and as a result of his hostility to the church and of his partial deafness, he lost his office. During his childhood, Walter Christaller was fascinated by maps; he was spending hours looking at them, taking imaginary journeys, and marking lines between cities close to each other, forming networks and triangles. A nagging question emerged in his mind at that time: why is it that towns were so regularly spaced along the routes? Why did small towns follow large ones, in apparently regular progression? Years had to pass before he answered these questions to his own satisfaction.

His first studies – undertaken before World War I – were in philosophy and political economics at the universities of Heidelberg and Munich, rather than in geography. During the war he became a soldier and, wounded several times, later returned home. These experiences no doubt encouraged a strong interest in socialist ideas which remained with him throughout his life, although now and again he denied and sometimes modified them. During the 1920s, he experienced different jobs – miner, construction worker, newspaper article writer – but the fact that none of these jobs lasted for a long time created personal problems for him because, married in 1921, he was faced with raising three children. Christaller and his wife were divorced in 1928 with an amicable settlement, which allowed him to have good relationships with his three children; in fact, the older daughter recalled later on that during her childhood she drew hexagons for him. During those years Christaller did not pursue geography except for designing a list of code numbers of German towns. He called that activity a hobby, but it was indicative of his mathematical theoretical mind.

In the 1930s, Christaller continued his studies at the University of Erlangen (Bavaria), under the influence of Robert Gradmann, who had led the foundation of a location theory of cities. At that time, Christaller was able to reply to his childhood questions, by creating his so well known urban hierarchy model, presented for the first time in his doctoral degree in political economics earned in 1932. In that year, he left Germany for a short period because of Adolf Hitler's election and Christaller's linkage to the social democratic party. His PhD thesis *Die Zentralen Orte in Süddeutschland* was published in 1933 after his return to Germany. This work was destined to be a landmark in his carrier, a success throughout the world, in more recent time translated into English and into Italian. In 1937–38 he finished a dissertation thesis (*Habilitation*) on 'Rural settlement in Germany and their relation to community administration', which introduced him in the scientific world as a spatial planner.

Before his studies, no theory had ever tried (1) to interpret regularities in the structure of urban settlements and (2) to give a sound economic explanations to these regularities.

Walter Christaller was able to explain the existence of an 'urban hierarchy', and in particular the size and frequency of urban centers at every level in the hierarchy – and therefore the market area of each of them – and the distance between a particular city and those at the levels immediately below or above it – and therefore the geographical distribution of all the urban centers.

Together with the economist August Lösch, Walter Christaller was hence the first to formulate a model able

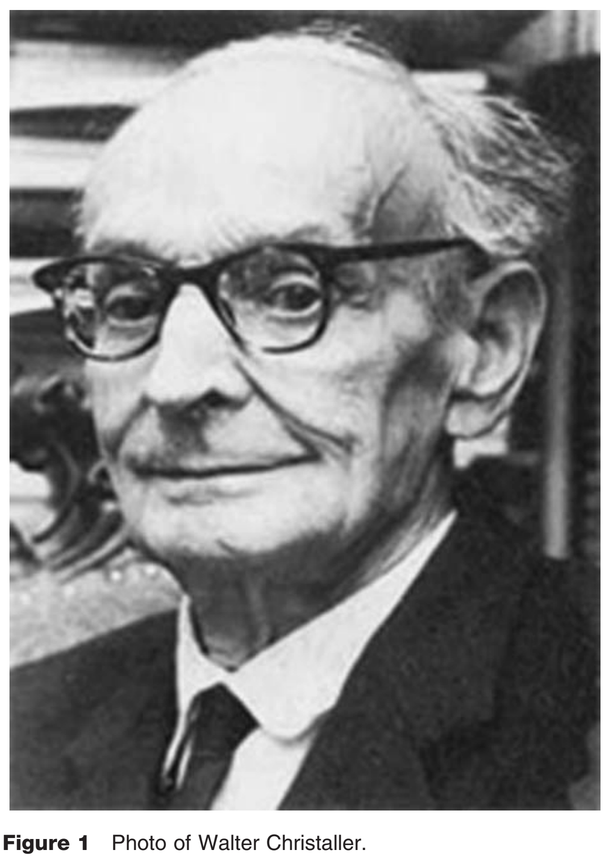

Figure 1 Photo of Walter Christaller.

to demonstrate the existence of an urban hierarchy in which each city of a certain size performs a specific function. Moreover, the model is able to furnish rules with which to identify the number of centers of a certain order, the size of each market area of each center, the distances among centers of the same order, and therefore their geographical distribution. Christaller's model is based on the assumption that an urban center exists where there are goods and services to be traded. This central 'place' (hence the name 'central place theory' given to the literature that Christaller's model inspired) must produce or supply goods or services to a population spatially dispersed across a uniform and isotropic surrounding territory. The aim of the model is to show how products and services (especially tertiary functions) come to be territorially organized in an urban hierarchy.

In 1940, Walter Christaller joined the Nazi party, eager to apply his central place ideas, and blind to the political misuse being made of them. He worked at that time for the *Reichskommisariat für die Festigung deutsches Volkstums*, created in 1939 under the direct control of Himmler. Christaller's job was to plan and classify East German and polish occupied cities of that time on the basis of his theory.

After the fall of the Third Reich, Walter Christaller joined the communist party with the hope that an authoritarian government "would have used its power to relocalize the cities destroyed during the war on the basis of the optimal framework suggested by the central place theory" (1949), and in 1951–52 became its deputy in the local council of Darmstadt, the city where he decided to live after the war, thanks to the availability of his mother's house. Because of cooperation in a communist study group on agricultural problems in both East and West Germany, he was charged with espionage, but after some years of legal proceedings all charges were dismissed.

Christaller worked in a field that was exploited politically to a high degree, nevertheless his theory became known very late, and appreciated first of all outside the geography's world. In 1941, thanks to Edward Ullman, Christaller's ideas were introduced in the US under the title 'A theory of location of cities'. In the course of time, Christaller's theory had made an impact throughout the world.

Only in his seventies, Christaller began to receive honors, first from abroad. He was given the outstanding Achievement Award of the Association of American Geographers in 1964, and in 1968 he received honorary doctorates from the University of Lund (Sweden) and of Bochum (Germany). In 2003 he received the posthumous title of 'The century's inventor' in the American journal *National Geographic*.

See also: Central Place Theory; Economic Geography; Location Theory; Spatial Analysis, Critical.

Further Reading

Beckmann, M. J. (1958). City hierarchies and the distribution of city sizes. *Economic Development and Cultural Change* 6, 243 248.
Beguin, H. (1984). The shape of city size distribution in a central place system. *Environment and Planning A* 16, 749 758.
Berry, B. and Garrison, W. (1958). Recent development of central place theory. *Papers and Proceeding of the Regional Science Association* 4, 107 120.
Christaller, W. (1933). *Die Zentralen Orte in Süddeutschland,* Darmstadt: Wissenschaftlische Buchgesellschaft. English edition (1966). *The Central Places in Southern Germany*. Englewood Cliffs, NJ: Prentice Hall. Italian edition (1980). *Le Località Centrali*. Milano: FrancoAngeli.
Eaton, C. and Lipsey, R. (1982). An economic theory of central places. *The Economic Journal* 92, 56 72.
Mulligan, G. (1984). Agglomeration and central place theory: A review of the literature. *International Regional Science Review* 1, 1 42.
Ullman, E. (1941). A theory of location for cities. *The American Journal of Sociology* 853 864.

Christian Geography

H. Aay, Calvin College, Grand Rapids, MI, USA

Glossary

Creation-Fall-Redemption Basic tenets of a (Reformed) Christian worldview: reality as creation; sin as misdirection of creation; redemption as restoration of creation.
Natural Theology The establishment of knowledge about God through study of the world.
Neo-Calvinism Revival of Calvinism led by Abraham Kuyper (1837–1920) beginning in the second half of the nineteenth century with a focus on relating Christianity to all areas of life.
Qualitative Pluralism Ontological irreducibility of distinct kinds of existence.
Reductionism Explanation in terms of one constitutive aspect or element.
Reformed Branch of Protestantism associated with the life and works of John Calvin (1509–1564) and his followers.
Religion Not an area of life but the whole of life. It is service of what is regarded as divine in every area of culture.
Worldview A vision of and for life; the comprehensive framework for one's basic beliefs about things.

Introduction

Many a different adjective and prepositional phrase added to the word 'geography' has become common currency to identify and parcel out geography's distinctive fields, topics, regions, paradigms, approaches, research groups, national traditions, and historical periods. Christian is almost never one of them. There is Berkeley geography, Classical geography, the geography of aging, Japanese geography, idealist geography, the geography of the Northern Netherlands, applied geography, sport geography, and the mountain geography specialty group. International and national synopses, dictionaries, and encyclopedias of geography survey reflect and police the current ontic and explanatory/interpretative organization of the discipline. Christian geography has not been an entry in such reference works. Yet there is a field or tradition or approach of Christian geography that can be identified and appraised, as in the 1998 survey edited by Henk Aay and Sander Griffioen and also in a special 2002 theme issue of the *Christian Scholar's Review*, 'Geography in Christian perspective', edited by Janel Curry.

In spite of later vigorous dismissals by those committed to scientific naturalism and positivism, geographic scholarship in the West from the beginning of the Middle Ages well into the eighteenth century may be classified as Christian geography because natural theology and Christian teleological frames were so prevalent. While a pervasive secularization of science and scholarship, including geography, prevailed until recently, nevertheless, a small Christian scholarly geographic thread remained visible, carried by some individual scholars at public universities together with geographers at a small number of Christian colleges and universities. More commonly, geographers who were Christians accepted a disconnect between personal piety and professional scholarship; except for Christian virtues such as honesty, charity, empathy, and diligence, Christian confessions did not enter the work floor of geographic research and teaching. When scientific rationalism came under fire from a renewed humanistic scholarly tradition in the early 1970s, Christian geographers such as David Ley and Ian Wallace, working from explicitly biblical views of humanity and world, were first taken seriously and regarded as contributing to the development of modern humanism in geography. In today's multi paradigmatic and postmodern climate, especially in human geography, there is much more ready acceptance of control beliefs and worldviews for informing geographic scholarship. Even against the background of postmodernism's pervasive relativism and skepticism, there is today more of an open window for Christian geography and a greater likelihood that it is taken seriously by academic communities and the general public. This is the case more in the United States than elsewhere. If there is a place for positivist, Marxist, and realist geographies, there is room for Christian geography, provided it is an open scholarly and not a closed, predetermined, and prescriptive enterprise. Moreover, Christian geography must be open to dialog with other approaches; to date, this has been lacking.

Faith Precedes Science, Faith Precedes Geography

Central to the project of Christian scholarship and, by extension, Christian geography, is the biblically based confession that there is no part of the world, no arena of life, and no human enterprise that is autonomous from the Creator and stands apart from the service of God. Like all other realms of life and world, science and scholarship in general, and the field of geography in particular, need always point to their origin and to their engagement with the purpose and meaning of human life. This strong position is certainly not shared equally by all traditions within Christendom; it is associated most clearly with a Reformed, neo Calvinist position, and somewhat less with, for example, Anabaptist, Lutheran, and Catholic positions. These circumscribe the relevance of the Christian faith more to certain arenas of human life and culture, such as evangelism, church, personal piety, and conduct and less, or not at all, for example, to science and scholarship, environmental issues, and globalization.

If the relevance of the Christian faith to world and culture is fully comprehensive and holistic, then it must include theoretical thought and with this the entire institutionalized and culture bound system of universities, science, disciplines, scholarship, and research is pulled into its orbit. It is one thing, however, to declare that geographic thought and research are not autonomous and not unrelated to religious starting points and calling, it is quite something else to give such a declaration a more case specific form. Fortunately, there is now a well established and accepted tradition in the philosophy of science that empirical research is always shaped by *a priori* beliefs and theories, by philosophical presuppositions and conceptual categories. Some of these are methodological and epistemological but other even more prior, fundamental control beliefs involve the nature of humanity, the origin of the cosmos, the unity, and coherence of reality as well as the essence of different aspects of the world studied by the sciences (by geography). Such beliefs and presuppositions are determined by or are part of faith; they cannot be given as answers by science itself.

Such presuppositions and basic beliefs are always elements in more comprehensive and integrated frameworks often designated as worldviews. Hinduism, consumerism, pantheism, humanism, and materialism are examples. Worldviews are prescientific systems of values about things; they are guides for living carried by all human beings; they provide a way of understanding all of reality. Some may be scientifically and formally organized and carried forward into philosophies and ethics, such as postmodernism, Marxism, and idealism and pragmatism. Worldviews serve as implicit guides and steering mechanisms for daily life in a culture; together with the philosophies based on them, they also guide and shape the scientific enterprise. A still excellent and telling example of a Christian analysis of presuppositions is geographer Rowland Moss' 1985 survey of the underpinnings of the metaphysical systems in environmental management, scholarship and practice. Evolutionary humanist, technological pragmatist, Marxist, mystical holist, magical, as well as several variants of a Christian worldview are here considered and interpreted.

Types of Christian Geography (and Types of Christian Geographers)

There are many geographers who intentionally or unintentionally keep their Christian faith and their geography profession very much apart. Faith is private or confined to church, so they might argue, and has no place in the university, scholarship, or scientific research. Science is an autonomous undertaking with its own *modus operandi.* Faith may shape the interaction with and treatment of fellow researchers, colleagues, and students; chapel and campus fellowship thus belong to its claims but not academic geography itself. Such a stance, itself grounded in belief, makes the phrase 'Christian geography' a contradiction. This position – Christian faith and geography have nothing to do with each other – represents one end of a spectrum along which a variety of ways exist to bring the two together. The five ways or types described below do overlap and the order in which they are presented does not necessarily represent an increasing integration of Christian faith and geography. It is also necessary to underline that the classes of scholarly approaches portrayed here are in no way the exclusive province of a Christian geography. The same subject matter may be investigated from diverse cosmological, ontological, and epistemological presuppositions, which together lead to different results and explanations.

'God's Fingerprints': Geography and Natural Theology

During the more than 2000 year history of geography in the West, natural theology has been the main type of Christian geography. Natural theology looks to the creation as a second source of revelation about the nature and attributes of God. From the Middle Ages to the eighteenth century, geographical writings were very commonly cast into a natural theology frame. The configuration of continents, the adaptability of human culture to environmental conditions, the alluvium deposited by rivers, the distribution of the human population, the commercial uses of particular rivers – all geographical patterns and human–environmental relationships – reveal divine design and specific attributes of God. Under attack from modern science, philosophy, and even theology, such teleological geographic interpretation receded in the nineteenth century and largely disappeared in the twentieth century. Yet, some professional academic geographers continued to refine the natural theology tradition in geography well into the twentieth century. A notable example is the neo Calvinist Dutch scholar Arie van Deursen (1891–1963), a pioneer Christian academic geographer for whom natural theology together with the scriptures provided the necessary additional Christian content of meaning for the observable geographic facts.

As the cultural force of Christianity waned, natural theology disappeared from geographic scholarship, replaced by a much more secular science. A formidable weakness of this approach to Christian geography throughout has been that theology or worldview had little or no impact on empirical research itself; this is left mostly untouched or unquestioned; only the results are read as evidence and proof for divine design and divine attributes. Such associations do not really belong to the research cycle and its published results. They are confessional exclamations. Just as a Christian astronomer comes away in awe of the Creator's design for a galaxy or a Christian ecologist for energy cycling in an ecosystem, so the Christian geographer's love for and knowledge of God is made more real, for instance, by the involved mechanics of snow avalanches. These are present day echoes of natural theology in geography and as confessional exclamations they have a place but not in formal geographic scholarship. It is not that Christian content is added to neutral research by attention to theological implications; these belong to theology not to geography. Christian metaphysics needs to help shape geographic study itself.

Christian Geography by Subject Matter

One very common, straightforward yet restrictive route to bring Christianity and geography together is for scholars very deliberately to choose research topics related to churches, worship, congregations, Christian organizations, and institutions, Christian education, evangelism, missions, neighborhood outreach, activities of clergy and laity, and so on. While others study the journey to work, Christian geographers study journey to church; while others study the spatial dynamics of firms, Christian geographers study the spatial dynamics of churches. The spatial, place, landscape, and human ecological dimensions of Christianity related subjects offer much scope for research, although only a few geographers such as Roger Stump follow this trail. Topics such as these are appropriate to the (social) sciences; the supernatural and otherworldly aspects of the Christian religion are not. In somewhat the same vein, in the context of applied geography, Christian geographers with specific skills, knowledge, and understanding would partner with churches other para church and Christian agencies to improve their effectiveness and ministry.

One can argue that because Christian scholars have a love for, dedication to, and special interest in subjects related to the Christian life, experience, and thought, they would want to direct their research energies to such topics. The research agendas at Christian colleges and universities are replete with such selectivity. The precise proportion of Christian to non Christian scholars pursuing Christian topics of all kinds is not known but it is safe to say that Christians are in the majority. One can also argue that without a cadre of Christian scholars many of these kinds of topics would lie fallow; others would show little inclination to undertake such research. It is therefore necessary and beneficial 'in house' academic work that records, interprets, and evaluates the Christian experience and affairs and holds up a mirror for thoughtful Christians everywhere.

The scholarly pursuit by Christians of such topics brings into especially sharp relief again the worldviews and presuppositions that undergird research. Some would maintain that a Christian worldview engaged in the scientific investigation of such Christianity related subjects would lead to a loss of critical distance and an inclination to concentrate on certain aspects and overlook others. Indeed, many scholars who are Christians, but who strictly separate religion and science, may for that reason follow a more positivist research paradigm for the study of such topics. Conversely, the possession of a Christian confessional vision and the knowledge and experience of a believer is regarded by others as necessary to fully and truly explain and interpret a religiously motivated and laden activity, event, or phenomenon. Some would still argue that only outsiders (non Christians) are able to conduct valid, balanced, and meaningful research on Christian institutions, behavior, and ministries, but such outsiders' own metaphysics and experiences would also shape the research and the interpretation of its results.

In this regard, it is helpful to remember that a Christian worldview based on the scriptures includes norms and rules for life and world; such general principles may be further differentiated and developed by Christian philosophy and ethics. These are then used to shape, interpret, and evaluate the empirical research conducted on topics related to Christian practice and observance. The love and respect that Christian geographers would naturally have for the empirical study of Christianity related subjects listed above would be restricted and policed by a prior and greater commitment to the norms and metric of an integral Christian worldview, ethics, and philosophy. These empirical states of affairs, like all others, may be found lacking, conforming to, falling short of, enhancing, distorting, and negating a Christian worldview. With such research safeguards in effect, Christian geographers in general as 'insiders' are arguably better placed to produce meaningful and discriminating scholarship on Christianity related geographic subjects. That is not at all to say that scholarship about such subjects by non Christians is without worth.

Christian geography by subject matter is much too restrictive. It creates a subfield rather than a legitimate scholarly passage into every topic of the entire discipline. Moreover, just choosing Christianity related subject material, of course, lacks any assurance of a Christian worldview interpretation. Marxist, postmodern, liberal capitalist, feminist, positivist, and Christian geographic research on, for example, the camp revival meetings on the American frontier during the Great Revival at the beginning of the nineteenth century would look quite different from each other. A topic, worldview and phil osophy must, it can be contended, be considered together.

Geography of the Bible and the Holy Land

The geography of the Bible and Holy Land is a distinctive subset of Christian geography by subject matter. Geog raphers who make the scriptures and ancient Palestine their specialization are not all Christians, but many are. To the question: "how can I be both a Christian and a geog rapher?" some have answered: with research on the geography of the Bible and ancient Palestine. If the Bible is God's authoritative word for salvation and a canon for human life and understanding our world, then what better way to integrate Christianity and geography than by deepening and enriching the power and meaning of scriptures through geographic inquiry. If there is a geo graphical side to every part of reality, then surely the geographical aspect of this God inspired book and this very special place deserve a very high research priority. In this, such scholars join a more than thousand year tradition of geographical writings about the Bible and Palestine by theologians, biblical scholars, archaeologists, geographers, historians, and Near East experts. Some looked upon the Bible itself as a source of geographical knowledge. Throughout the West, this work produced atlases, topo nymic surveys, physical and human geographies of Bible lands, as well as more detailed accounts of the environ mental and spatial settings of Bible stories and writings. These Bible geography writings were more synoptic and especially served the needs of Bible education in churches, schools, colleges, seminaries, and universities.

Today, religiously conservative and fundamentalist Protestant Christians scholars are especially busy with this kind of Christian geography, more than those from other traditions within Christianity and from Jewish and other non Christian backgrounds. The geography of the Bible research community is especially active in the United States where for more than 30 years academic meetings have been held to share and discuss research results. Jonathon Lu and William Dando are among the leaders of this research community.

This type of Christian geography welcomes research by scholars from every kind of religious or worldview persuasion. One does not need to be a Christian to en gage in geographic scholarship about the Bible or ancient Israel. All kinds of scholarship are welcome to contribute to this body of geographic knowledge. Every geographic topic, method, interpretive framework, and technique further develops the geography of the Bible and ancient Israel whether Biblical site identification, water conflicts in ancient Israel, satellite images of Bible lands, or the geography of the wilderness wanderings. While there are links to the larger subfields of cultural geography and the geography of religion, the geography of the Bible and Bible lands is a specialty rather closed off from the rest of the field. It has little to no impact on the philosophical and methodological directions of geography as a whole. It does not offer another way forward for geography. Inte grating Christian faith with scholarship through this specialty makes the project of Christian geography rather restrictive.

The Difference that Religion Makes to Human Geographies

The legacy and practice of scientific rationalism and positivism in geography and in science in general in cludes strong opposition to all religion viewed as ignor ant superstitions and myths from which science will deliver humankind. Against such a scientistic outlook, a Christian worldview stresses the centrality of religion in human life, points out that life is the living of religion (not just by the openly religious), and emphasizes that human beings are by nature religious creatures. Religion is belief that something or some things are divine and how human beings should relate to the divine. The divine may have as its focal point something out of this world (the God of the scriptures) or something in the world (reason). Religion is constituted by ultimate commit ments, ground motives, presuppositions, and worldviews. It shapes and directs all human undertakings, including theoretical thought and scholarship. With this, religion includes much more than a traditional theistic or poly theistic metaphysics of, for example, Christianity and Hinduism. Materialist metaphysics, including Marxism and consumerism, and idealist metaphysics such as ra tionalism and positivism are in every way religions as defined here, even though some may be more philo sophically developed and articulated than others.

With religion as the concentration point of human existence in a Christian worldview, it makes sense that geographers who are Christians would be interested to trace the effect and extent of the impact of religion so defined on human landscapes, spatial organization, land use, environmental management, environmental per ception, and other familiar subjects of geographic in vestigation. Such a research focus is part of the field of the geography of religion although underplayed because of its focus on formal worship, formal religious obser vances, and the supernatural.

Rather than assigning and compartmentalizing world views in the geography of religion and belief systems, it would be much more strategic to assign religion a more central and robust methodological place and role in human geography in general. A fine example of this ap proach is Christian geographer Janel Curry's research in the late 1990s on worldview and agriculture and rural communities. She constructed the worldviews of five different Iowa farming communities and found that these were responsible for creating different places, different perceptions of and relations to the natural world, dif ferent ethics, and different agricultural systems.

Of course, as with the geography of the Bible, a focus on worldviews and their geographical effects is not the special province of Christian geography, although it is an approach that it should champion. Rather, it is an ex planatory perspective and interpretive framework rec ommended for consideration for all topical specialties and for geographers of every stripe. It is significant that the overall conceptual organization and research output of geography and its history are commonly thought of and organized in terms of worldviews or metaphysical paradigms; in this case, this approach is accepted, valued, and effective. Evaluating such worldviews along with their geographical imprints and proposing alternate geographies, however, is very much dependent on the metaphysical frameworks of a research community or individual scholars.

Christian Ethics and Geography

Geographers who wish to make their Christian convic tions count in their profession may especially want to focus on ethics and activism. After all, God's call to love and serve the creation and 'neighbors' draws Christians in general and Christian geographers in particular to know, analyze, and confront conditions of injustice, oppression, homelessness, exploitation, hunger, poverty, violence, inequality, and other expressions of a broken world. The gospel is not something supplementary or contrary to culture; rather, it directs itself to healing and restoring all things in the world. The geographical dimensions of such pressing issues set the agenda for research and the pub lished results include remedies: economic, social, and political restructuring. Geography is just one piece; multidisciplinary, analytical, and policy frames prevail in this kind of scholarship. Geography itself fades somewhat into the background. Moreover, often there is a strong participant observation, a personal identification with the afflicted, and a social activist stance in such moral geographies. The focal points of a Christian worldview in this case are sin and redemption.

Of course, many other geographers with different ideological persuasions share such professional priorities and research interests, including Marxists, those with liberal political commitments and deep ecologists. Christian geographers for whom such issues come to the very front personally and professionally because of their Biblically based values and sensitivities may directly adopt and engage any one of a number of suitable theories for ethical evaluation and action. Or, more consistently, and more from the ground up, Christian geography would work with and from Christian schools of ethics, different from the foundations of other moral theories such as utilitarianism, virtue ethics, human rights, care theory, Kantian duty ethics. On such a basis, Christian geography would analyze and suggest ethically appropriate solutions for specific cases of locational decision making, environ mental conflict and place based and global inequalities and welfare issues.

Geographers who represent this kind of Christian geography, among others, include Paul Cloke, Johnathan Bascom, and Gerard Hoekveld. Cloke is a rural geog rapher who has devoted much of his career and life in research and engagement with issues of rural homeless ness, marginalization, deprivation, and poverty in Great Britain. He grounds his research in a (post structuralist) Christian ethic, and is prepared to name as ultimately 'evil' – and through committed practice he urges us to counter – the many banalities of human life that 'thoughtlessly' or 'carelessly' intersect with one another across space to create multiple miseries and oppressions for some in some places (but not for others in other places). Similarly, Bascom, a specialist in East Africa, has made the geography of refugees (their migration, pro tection, and assistance) a large part of his life's work; again, Christian convictions, ethics, and worldview an chor his scholarship and personal involvement. Hoek veld, as well, during the past decade, has written extensively about relationships between applied human geography and ethics as seen in locational decision making about infrastructure and urban renewal in the Netherlands. Christian ethics is a dynamic and rapidly developing scholarly field; as such, it can effectively and consistently bridge, translate, and mediate Christian convictions, values, and worldview with Christian geo graphic scholarship.

Christian Worldview and Christian Philosophy: A Paradigm for All of Academic Geography

There is another more comprehensive, complete, and perhaps fitting alternative for Christian geography. It takes its identity from the above mentioned premises about worldviews and geography. Accordingly, Christian geog raphy is not a pre scientific leftover from its long history, not a topical specialty, and not a branch of the geography of religion. Rather, it is a worthy and authentic scholarly and scientific approach to the entire field of geography based on a scripturally directed, comprehensive, and in tegral worldview. It rejects the view that geographic sci ence is autonomous, religiously neutral, or the only approach to valid geographic knowledge. David Ley hinted at such a Christian paradigm for human geography in his provocative 1974 paper, expressly offering a different understanding of the city to that provided by Marxists. David N. Livingstone, John L. Paterson, and Ab van Langevelde have at times also represented this approach; it constitutes the smallest wing of Christian geography but offers, the author believes, the greatest relevance, integrity and scope for bringing Christianity into geography. Where it is institutionalized, this approach is especially found in colleges and universities affiliated with the Reformed branch of Christianity.

Many would agree that the Reformed Christian tradition, in particular, has made particularly important contributions to advance the project of Christian scholarship. From within this tradition for the past hundred years the contours of a Christian worldview have been built and translated into Christian philosophy that is intensely relevant and serviceable for the special sciences such as geography. 'Creation fall redemption' is a well known shorthand description for this worldview. All of reality (humanity, society, ecology, the physical Earth, and universe) is God's creation and there is a good ordered creational structure for everything (laws and norms). A commitment to such a fundamental Christian creationism has nothing to do with other very different and programmatic types of creationist positions such as creation science, young Earth creationism, or intelligent design. The entire creation has been deformed and misdirected by the working of human sin, and cultural life consists of better or worse responses to creational norms. The gospel/Jesus Christ makes such norms visible and operative in culture again; and the entire creation may be redeemed and restored; the original goodness and intent of creation is regained.

The creational structure and order displays a diversity of irreducible kinds of functioning – physical, economic, aesthetic, spatial, to name just a few – and there exists a great pluriformity of things (glaciers, firms, congregations, regional economic development) that bundle and combine these functions in distinct ways. Such qualitative pluralism and multidimensionality are hallmarks of a Christian ontology that seeks to avoid reductionism. When any one aspect becomes an ultimate explanatory principle (spatialism, economic determinism, and environmentalism) or is consistently explained in terms of another, something in the creation is made absolute and idolized. This is a crucial ontological consideration for geography as for any other science. A geographic analysis and understanding of any issue, process, or entity is always partial and restricted and has to take its place alongside the scholarship of other sciences and, as has been repeatedly asserted, depends on a metaphysical paradigm to account for coherence, unity, origin, and destination. During the last half century the philosophy and canons for Christian scholarship have seen a great deal of discussion and development. Christian geographers therefore have the opportunity to seize upon that elaboration and refinement and make their world view count and be meaningful in their scholarship.

See also: Geography, History of; Marxism/Marxist Geography I; Philosophy and Human Geography; Religion/Spirituality/Faith.

Further Reading

Aay, H. and Griffioen, S. (eds.) (1998). *Geography and Worldview. A Christian Reconnaissance*. Lanham, NY: University Press of America.
Aay, H. and Van Langevelde, A. (2005). A Dooyeweerd based approach to regional economic development. *Journal of Social and Economic Geography* 96(2), 184 198.
Bascom, J. (1998). *Losing Place. Refugee Populations and Rural Transformations in East Africa*. New York: Berghahn Books.
Blamires, H. (1988). *Recovering the Christian Mind: Meeting the Challenge of Secularism*. Downers Grove, IL: Inter Varsity Press.
Cloke, P. (2002). Deliver us from evil? Prospects for living ethically and acting politically in human geography. *Progress in Human Geography* 26, 587 604.
Clouser, R. (2005). *The Myth of Religious Neutrality: An Essay on the Hidden Role of Religious Belief in Theories*. Notre Dame, IN: Notre Dame Press.
Curry, J. M. (2000). Community worldview and rural systems: A study of five communities in Iowa. *Annals of the Association of American Geographers* 90(4), 693 712.
Curry, J. M. (ed.) (2002). *Theme issue: Geography in Christian perspective*. *Christian Scholars Review* 30(4), 351 471.
Dando, W. A., Dando, C. Z. and Lu, J. J. (eds.) (2005). *Geography of the Holy Land: Perspectives*. Kaohsiung, Taiwan: Holy Light Theological Seminary Press.
Dooyeweerd, H. (1953 1958). *A New Critique of Theoretical Thought*, 4 vols. Philadelphia, PA: Presbyterian and Reformed Publishing Company.
Harrison, R. T. and Livingstone, D. N. (1980). Philosophy and problems in human geography: A presuppositional approach. *Area* 12, 25 31.
Hoekveld, G. A. (2005). *Het hoogste doel. Over de verhouding tussen toegepaste sociale geografie en ethiek (The Highest Goal. Concerning the Relations between Applied Human Geography and Ethics)*. Utrecht, The Netherlands: Faculteit Geowetenschappen.
Ley, D. (1974). The city beyond good and evil: Reflections on Christian and Marxist interpretations. *Antipode* 6(1), 66 73.
Marsden, G. M. (1997). *The Outrageous Idea of Christian Scholarship*. New York: Oxford University Press.
Moss, R. (1985). The ethical underpinnings of man's management of nature. *Faith and Thought, Journal of the Victoria Institute* 111, 23 56.
Park, C. C. (1994). *Sacred Worlds: An Introduction to Geography and Religion*. London: Routledge.
Paterson, J. L. (1998). *Geography and Religion, Agriculture and Stewardship: The Practice of Agricultural Stewardship in the Christian Farmers Federation of Canada*. PhD Thesis, University of British Columbia.
Wallace, I. (2006). Territory, typology, theology: Geopolitics and the Christian scriptures. *Geopolitics* 11, 209 230.
Wolters, A. M. (1985). *Creation Regained. Biblical Basis for a Reformational Worldview*. Grand Rapids, MI: Eerdmans.

Relevant Websites

http://www.freewebs.com/reformationalphilosophy
Resources for a Christian World View.

Chronic Disease

K. Robinson and S. J. Elliott, McMaster University, Hamilton, ON, Canada

Glossary

Comprehensive Intervention It focuses on multiple risk factors and approaches, addressing multiple segments of society, implemented in multiple settings/environments.

Determinants of Health The range of personal, social, economic, and environmental factors that determine the health status of individuals or populations.

Epidemiology The study of the health of populations.

Health Promotion The process of enabling individuals and communities to increase control over the determinants of health and hence improve health status.

Morbidity Illness.

Mortality Death.

Population Health Approach This approach, through a focus on the broad range of health determinants, aims to improve the health of entire population (as opposed to individuals) and to reduce health inequities among population groups.

Prevention It includes measures to prevent the occurrence of disease (primary prevention), as well as arrest its progress (secondary prevention) and reduce its consequences (tertiary prevention).

Introduction

Chronic diseases are ongoing/recurring, not caused by infection, or passed on by contact. They generally cannot be prevented by vaccines, do not resolve spontaneously, and are rarely cured by medication or other medical procedures. Chronic diseases are often caused by a combination of health damaging behaviors (smoking, drinking, high fat diet, and physical inactivity) and exposure to socioenvironmental conditions (stressful living circumstances, poverty, and automobile focused urban centers). The World Health Organization (WHO) identifies four main types of chronic diseases – cardiovascular disease (CVD; primarily heart disease and stroke), cancer, chronic respiratory diseases, and diabetes – responsible for the majority of related deaths. However, mental disorders, vision and hearing impairment, oral diseases, bone and joint disorders, and genetic disorders are also noted as chronic diseases that account for a substantial portion of the global burden of disease.

The Global Burden

Chronic diseases are the largest cause of mortality in the world, accounting for close to half of the world's deaths. The greatest contributor to this burden of illness is CVD, followed by cancer, chronic lung diseases, and finally diabetes mellitus. The picture of health has not always been this way. Many countries in the developed world have 'achieved' what is known as 'the epidemiologic transition', that is, the health status and disease profile of human societies have historically been inextricably linked to their level of economic development and social organization; when we lived as hunter gatherers, we had very high birth rates as well as high death rates. Major causes of death were starvation, warfare, interactions with fierce animals, high rates of maternal and child mortality, as well as frequent epidemics of infectious diseases and famines. Healthcare systems at this time represented indigenous systems, traditional healers, and herbal medicines. As we organized ourselves into agrarian societies, we had better access to food and therefore birth and death rates began to fall near the end of this period. High rates of fertility continued given persistent high rates of infant mortality and the need for farm labor, but both of these, as we saw, fell tremendously during the Industrial Revolution. Deaths from infectious diseases decreased dramatically, standards of living increased accordingly, and families therefore gave birth to fewer children. Farming practices became mechanized, more food was being produced for fewer people, and all in all people lived longer and healthier lives. While there is some suspicion that the primary reason for this shift in stage 3 was the development of, and increased access to, a modernizing healthcare system, others disagree. A historical epidemiologist named Thomas McKeown explored over 300 years of data to come to the conclusion that, indeed, healthcare and access to it had very little to do with the gains in health status the world saw during the 1800s and early 1900s. Rather, the primary contributors to enhanced health status were sanitation, improved nutrition, and family planning. Stage 4 in the epidemiologic transition represents our current state, where chronic diseases take the lead with respect to mortality and morbidity. The literature often refers to these as 'diseases of affluence' given that as a human society, the higher our standard of living and the more progressive our economic development, the more likely we are to die of a chronic (as opposed to infectious) disease. Now we start seeing deaths overtaking births in many parts of the world. The explanation for this is that women are more educated, take a full role in the work force, and are choosing to have fewer children. Fertility is, therefore, at below replacement levels (e.g., see the population trends in Greece, France, and Germany). As a result, some countries have to rely heavily on immi gration to fuel their economies. As we live longer, and make more money, rates of chronic diseases increase, due to the role of lifestyle (smoking, drinking, high fat diets, and lack of physical activity) and environmental (auto mobile based cities) factors. The final stage of the epi demiologic transition predicts effective chronic disease management that would see a substantial reduction in mortality from chronic disease. But we have to die sometime, of something. Researchers predict we reach our 'biological wall' at about 85 years. So, the true pattern of stage 5 is yet to be determined.

Regardless, the global prevalence of all the leading causes of chronic diseases is increasing, with the majority occurring in developing countries and projected to in crease substantially over the next 20 years. Chronic dis eases have not simply displaced acute infectious diseases as the major cause of death in most developing countries, but they have joined them. As a result, such countries now experience a double burden of disease; a polarized dis tribution of ill health, if you will, that makes any policy response that much more complicated. For example, in South Africa, infectious diseases account for 28% of years of lives lost while chronic diseases account for 25%. This distribution of disease can put a tremendous strain on any healthcare system. Further, this polarized distribution creates unhealthy, unintended consequences. For ex ample, antiretroviral regimens in HIV infected patients actually 'increase' the risk of CVD. Despite this increasing burden of disease, WHO spends only 50 cents per death per person on all noncommunicable diseases (NCDs) except for mental health; it spends a comparative \$7.50 on leading communicable diseases.

Determinants of Chronic Disease

Although chronic diseases are among the most common and costly health problems in the world, they are also among the most preventable. A small set of risk factors explains the majority of chronic disease occurrence at all ages and for both sexes. Four of the most prominent chronic diseases – CVDs, cancer, chronic lung disease, and diabetes – share common and preventable biological risk factors, specifically high blood pressure, high blood cholesterol, and overweight (**Figure 1**). These inter mediate, clinical risk factors are related to three modi fiable behavioral risk factors – unhealthy diet, physical inactivity, and tobacco use – as well as nonmodifiable risk factors (age and heredity).

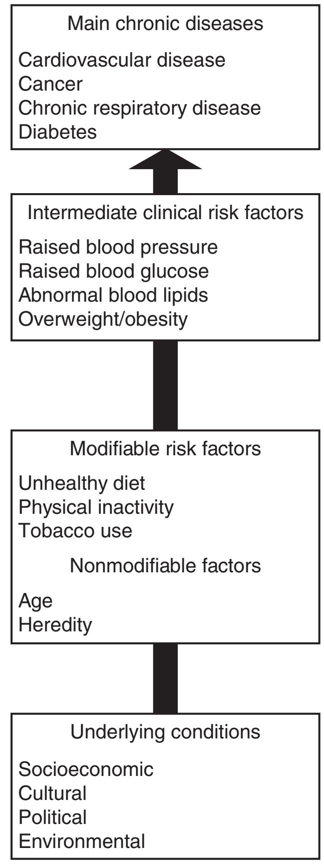

Figure 1 The determinants of chronic disease.

These modifiable risks are not solely determined by individual lifestyle choices, but are rather shaped by underlying socioeconomic and environmental conditions. The underlying determinants of chronic diseases are a reflection of larger societal and institutional forces driv ing social, economic, and cultural change such as glob alization, urbanization, population aging, and policy structures.

Greater exposure to physical risks, higher levels of psychosocial stress, and lack of access to healthcare make those living in poverty even more vulnerable to chronic disease; further, the experience of chronic illness can also push families into poverty.

A range of other factors contributes to a much smaller proportion of chronic disease such as excessive alcohol use and infectious agents (related to cervical and liver cancers). Physical environmental factors (such as air pollution) also cause several chronic respiratory diseases. Finally, genetics and psychosocial factors (such as stress) also play a role, although their interaction with other determinants is more complex.

A growing body of research, based on chronic disease epidemiology from a life course perspective, has dem onstrated the critical role of conditions in particular life stages in determining chronic disease. In particular, prenatal (maternal health, nutrient intake, and exposure to harm) and early childhood conditions (diet and socioeconomic conditions) are associated with rates of chronic disease in adulthood.

Geographies of Chronic Disease

Geographers have contributed to our understanding of chronic disease in a number of ways: distribution and diffusion through disease ecology approaches, the use of geographic information systems (GISs) and spatial analysis, contributing to our understanding of determinants, highlighting the inequalities in mortality and morbidity, contributing to our understanding of the impacts of chronic disease on individuals, as well as more recently contributing to our understanding of prevention.

Disease ecology is a long standing branch of medical geography that focuses on the distribution and diffusion of disease. Simply mapping the incidence of mortality and morbidity from chronic disease – at various spatial scales (local, regional, national, and international) – can provide clues as to its etiology and treatment. In a sense, therefore, these types of investigations are primarily hypothesis generating. When we pair incidence data with potential determinants, we can start to test whether or not hypotheses have some credibility. For example, we can map data on the hardness of water along with incidence of mortality and morbidity from CVD to assess whether or not these two things are associated in space. Or, we can map air pollution data along with CVD data to assess the same (potential) association. The introduction of sophisticated GISs and spatial analytic techniques allows us to explore these relationships even further. Take for example the simple mapping of CVD incidence data. We can map rates for, say, Canada and the US. From this, we can develop a powerful visual display of where rates are high and where they are low. While interesting enough, the application of spatial analytic techniques such as local indicators of spatial analysis (LISA) allows us to create a simple coefficient that tells us whether or not the patterns we see in the map are occurring due to some systematic distribution or simply because of chance. This process of identifying 'hot spots' of disease then focuses researchers and policymakers on areas deserving of greater attention.

Geographers have also contributed substantially to our understanding of the determinants of chronic diseases and their associated risk factors. The importance of space and place underlies much of this work. For example, geographers have contributed to our understanding of the different pattern and intensity of risk factors in remote northern communities in Canada. More recently, geographers have been undertaking a substantial amount of research on the role of the built environment as a determinant of chronic diseases and their related risk factors. Using GIS to objectively measure features of the built environment, for example, we can see the relationship between the built environment and physical activity, a major risk factor for overweight and obesity, CVD, diabetes, and some cancers. Some geographers have explored the assessment of a 'walkability' index using GIS in residential neighborhoods, where 'walkability' can be defined as the extent to which characteristics of the built environment and land use may or may not be conducive to residents in the area who walk for exercise, recreation, or leisure in order to access amenities or travel to work.

Recent developments in the area of multilevel modeling have increased the capacity of geographers to explore not only the spatial distributions of (chronic) disease, but also their determinants at the individual and ecological levels. For example, health geographers have always been curious – in terms of the determinants of disease more generally – about the relative contributions of the characteristics of the individual (e.g., sex, age, and income) versus the characteristics of the neighborhood in which the individual lives (e.g., average income, average education, and average dwelling value).

With respect to inequalities in the burden of chronic disease, geographers have focused on a range of characteristics: socioeconomic status in general (income in particular), race, ethnicity, education, and geographic location. For example, researchers have found a distinct social gradient in the incidence of chronic disease by socioeconomic status, indicating a stepped gradient in incidence with decreasing status. Geographers in Canada have deconstructed this further to indicate that while income is a major determinant of chronic disease inequality in the US, education plays a greater role in the Canadian context. This may be due to the lack of universal health insurance in the US. Rates of chronic disease, and their associated risk factors, are also higher among certain racial and ethnic groups. For example, many African Americans suffer disproportionately from hypertension, diabetes, and dyslipidemia (i.e., an abnormality related to fats or lipids in the blood), all of which are significant risk factors for CVD and some cancers. The primary factors contributing to this pattern could include chronic overweight and obesity, high salt intake, lack of adequate follow up, inadequate access to healthcare services, cost of treatment, etc. Further, geographic variation in mortality from chronic disease in African Americans has been the focus of several research studies, and substantial variation has been found. Similar results stem from studies of geographic variations in chronic disease rates among women, and particularly minority (i.e., African American and Hispanic) women. This translates to utilization of healthcare services for chronic diseases, with Hispanic women exhibiting very low rates of screening for breast cancer in the US.

While much of the above research in the geographies of chronic disease has employed quantitative methodologies, the use of alternative epistemologies by human geographers to explore the experiences of individuals diagnosed with a chronic disease has also contributed substantially to our understanding. These researchers focus on the meaning of chronic illness and the disruptions to life because of its onset. Major findings indicate a shrinking of geographical worlds as chronic illnesses progress and individuals find that the social and economic roles that they play(ed), as well as the spaces they inhabit(ed) begin to erode as working life and many aspects of social lives come slowly to an end as the chronic disease progresses.

Finally, health geographers have been moving into the area of chronic disease prevention and health promotion. Much of this work has focused on the capacity building efforts of both communities and public health systems in the context of chronic disease prevention. Health promotion specialists have begun to realize the importance of the environment (social, physical, and cultural) in the prevention of chronic disease and have begun to shift the emphasis from individual lifestyle type promotion activities (e.g., encouraging individuals to quit smoking) to community based interventions and policies (e.g., bylaws around the use of and advertising for tobacco). Prevention efforts are discussed in more detail in the next section.

Response to the Burden

Chronic diseases have generally been neglected in international health and development work. This is due, in part, to the commonly held notion that chronic diseases mainly affect high income countries, when the majority of global deaths due to chronic disease actually take place in low and middle income countries. Chronic diseases hinder economic growth in all countries. However, it is especially important that prevention efforts are undertaken in both countries experiencing rapid economic growth (e.g., China and India) and those that are the least developed countries.

Integrated Approach to Chronic Disease Prevention

A significant percentage of deaths from cancer, heart disease, and diabetes could be reduced or delayed through primary prevention. Based on the commonality of risk factors and approaches across many chronic diseases, there is a growing recognition of the strong rationale for pursuing an integrated approach to prevention efforts. Although an elevation of a single risk factor significantly predicts individual's ill health, the societal burden from NCD results from the high prevalence of multiple risk factors related to general lifestyles. Community based activities are required with an 'integrated' public health approach targeted to the population, in addition to those at high risk by linking prevention actions of various components of the health system, including health promotion, public health services, primary care, and hospital care. Integrated chronic disease prevention aims at intervention that addresses the common risk factors by the health system and other existing community structures, rather than an outside prevention program. It denotes a comprehensive approach, which combines varying strategies for implementation. These include policy development, capacity building, partnerships, and informational support at all levels. Integration also calls for intersectoral action to implement health policies. Finally, an integrated approach centers on strategic consensus building among different stakeholders, such as governmental, nongovernmental, and private sector organizations, in an effort to increase cooperation and responsiveness to population needs.

The concept of an integrated approach is not a new one. It was first introduced by WHO in 1981 and later endorsed by the World Health Assembly in 1985, which led the WHO program to test and evaluate comprehensive, integrated, community based approaches to chronic disease prevention and control (WHO's Countrywide Integrated and Non Communicable Disease Intervention (CINDI) Program).

Promotion of Health Equity

Traditionally, chronic diseases were considered diseases of affluence. Currently, there exists sufficient evidence to indicate that this relation has been inverted. Evidence clearly shows that the risk for some NCDs (such as CVD and certain forms of cancers) is now higher at lower socioeconomic levels.

Prevention strategies should consider underlying influences on health inequalities such as education, income distribution, public safety, housing, work environment, employment, social networks, and transportation among others. It is important that strategies are aimed at reducing overall population risk while simultaneously reducing the gap among different population groups. In many instances, this requires a redesign and an evaluation of interventions of well documented efficacy. It also entails the identification and special attention to key population groups, such as Indigenous Peoples, new urban migrants, and women.

Community-Based Interventions

Community interventions have demonstrated that they have a great effect on NCD prevention, since the interventions intend to act not only on the individual and its nearby social nucleus, but additionally on the social en vironment that determines behaviors. These inter ventions imply the active participation of families and communities, pooling and sharing resources to ensure integrated preventions, the identification of leaders, or ganized groups and institutions, and the development of strategic coalitions and alliances.

Looking Forward

Is the world close to achieving its demographic transi tion? Will infectious diseases be a thing of the past, with chronic disease rates putting an increasing strain on the world's economies and healthcare systems? Will the world end up with the bimodal health distributions seen in parts of Africa and other parts of the developing world? Will the stage of effective chronic disease man agement take root and bear fruit, altering the curve seen in stage 5 of the epidemiologic transition? These are only some of the questions that remain to be answered.

See also: Disease Diffusion; Epidemiological Transition; Health Geography; Health Inequalities; Medical Geography.

Further Reading

Dyck, I. and O'Brien, P. (2003). Thinking about environment: Incorporating geographies of disability into rehabilitation science. *The Canadian Geographer* 47(4), 400 413.
Kuh, D. and Ben Shlomo, Y. (2004). *A Life Course Approach to Chronic Disease Epidemiology*. London: Oxford University Press.
Leslie, E., Coffee, N., Frank, L., Owen, N., Bauman, A. and Hugo, G. (2007). Using geographic information systems to objectively assess relevant environmental attributes. *Health and Place* 13, 111 122.
McKeown, T. (1988). *The Origins of Human Disease*. London: Oxford University Press.
World Health Organization (2002) Reducing risks, promoting healthy life. *World Health Report 2002*. Geneva, Switzerland: WHO. http://www.who.int/whr/2002/en/whr02 en.pdf.
World Health Organization (2005). Preventing chronic diseases a vital investment. *WHO Global Report*. Geneva, Switzerland: WHO. http://www.who.int/chp/chronic disease report/en (accessed in April 2008).
Yach, D., Hawkes, C., Gould, C. L. and Hofman, K. (2004). The global burden of chronic diseases: Overcoming impediments to prevention and control. *Journal of the American Medical Association* 291, 2616 2622.

Relevant Websites

http://cbpp pcpe.phac aspc.gc.ca
Canadian Best Practices Portal for Health Promotion and Chronic Disease Prevention is designed to develop and disseminate best practices information for chronic disease prevention and control interventions.
http://www.cdc.gov
Centres for Disease Control and Prevention (US), National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP).
http://www.cdpac.ca
Chronic Disease Prevention Alliance of Canada (CDPAC) a networked community of organizations and individuals who share a common vision for an integrated system of chronic disease prevention in Canada.
http://www.phac aspc.gc.ca
Public Health Agency of Canada, Centre for Chronic Disease Prevention and Control.
http://www.who.dk
World Health Organization Countrywide Integrated Non Communicable Disease Intervention (CINDI).
http://www.who.int
World Health Organization, Department of Chronic Diseases and Health Promotion (CHP).

Citation Geography

M. Batty, University College London, London, UK
R. Carvalho, Queen Mary, University of London, London, UK

Glossary

Cartograms Maps of geographical activity that are distorted in such a way as to transform the spatial unit so that its area is proportional to some attribute, such as income. Such maps illustrate how city locations often dominate areas in their hinterlands while poor countries dwarf in size compared to rich countries.

Citation A reference to an article or book made by an author within the text and listed so that readers of the work can easily access the citation. Analysis of citations requires that these lists be catalogued and made available for other researchers to search them and identify patterns which usually indicate the productivity (and other attributes) of the named authors.

Citation Geography The generation of geographical patterns from citation databases, usually through data mining of digital datasets, such as those compiled by the ISI. The process of generating these data is one of compiling the geographical addresses of those authors who are cited, converting these to geo-coordinates, and then visualizing these locations through maps and rank plots.

Citation Networks Maps of relationships between authors and/or between papers and journals which are constructed to represent the frequency with which one author (or class of papers/journals) cites another. These are social networks in that they imply patterns of influence or mutual respect which reflect how authors cooperate. When rooted in terms of where authors are located, these networks can be geographical or planar, but usually, the network structure is purely topological.

Data Mining A generic process and set of digital methods for extracting pattern and order from extremely large data bases, often with millions of entries.

Institute of Scientific Information (ISI) The Institute of Scientific Information (ISI) which was founded in 1960 by Eugene Garfield and has remained the central archive for citation counts data in the arts, humanities, social science, and scientific worlds. Various aspects of these databases allow the calculation of measures of the impact of scientific research at the level of the researcher, the university, or institution, and the source of the output, such as the journal.

Power Laws and Scaling Frequency distributions of activity which are highly skewed to the left can be represented by power laws, usually relating the count of activity to its size in an inverse manner. Such distributions scale in that at whatever scale one examines the distribution, it has the same shape. Power laws can be seen as approximations to the fat or heavy tail of a lognormal distribution.

Rank-Size Distributions A method of transforming a frequency distribution into a form that rank-orders the objects from largest (rank 1) to smallest (rank *n*) which when plotted on logarithmic scales often yields a straight line.

Scientific Productivity The measurement of a scientist's output, usually in terms of papers and related printed material available to the wider scientific community, but also in terms of patents, and sometimes, products generated in a commercial context. Such outputs are increasingly digital and published in online resources and archives which enable their data to be mined.

Spatial Distributions Patterns of geographical activity that, in the socioeconomic world, are attributes or activities, such as populations, incomes, indices of deprivation etc., coded to Euclidean (map) space. In the case of citation data, these are volumes of activity associated with different geographical locations. Such distributions typically illustrate how geographic activity concentrates.

Distributions of Human Activity

When attributes and characteristics of human populations are ordered and counted, it is generally assumed that these are normally distributed with most of the population falling within some relatively narrow range of the average value of the characteristic in question. The classic example is adult male height where the current mean value in the UK is about 173 cm with 96% of the population falling within plus or minus 22 cm of this average. In statistical parlance, this is two standard deviations either side of the mean value. To all intents and purposes, most of the distribution lies within three such deviations on either side. Extreme values are very rare; for example, in 2006, the smallest, known, living, adult male was 72 cm, the largest 272 cm, which respectively are about nine standard deviations below and nine above the UK mean. When the frequency distribution of such heights is plotted, this follows the famous 'bell curve', shown in **Figure 1a**, that has been argued to apply to everything from the distribution of intelligence quotients

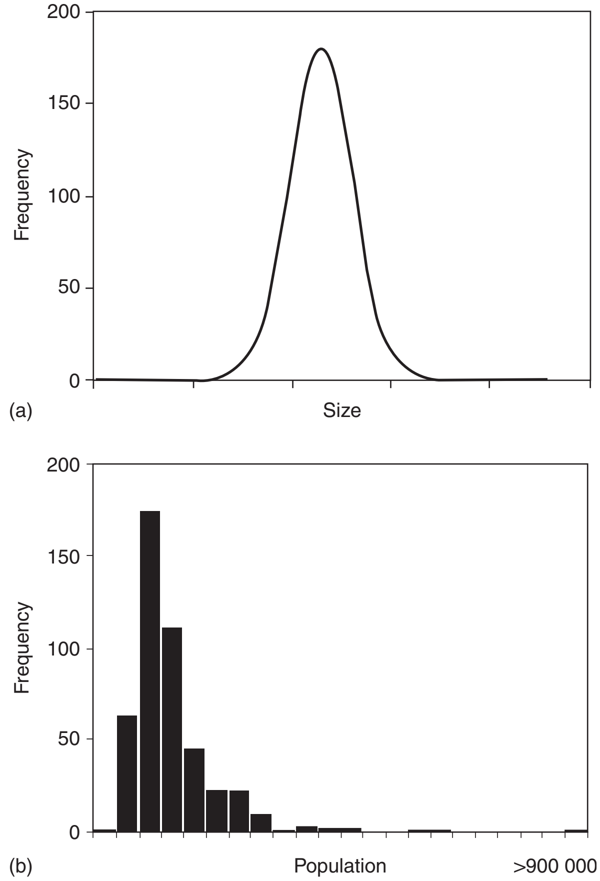

Figure 1 Normal and lognormal distributions: (a) a normal bell curve and (b) a 'typical' lognormal distribution of city sizes in Great Britain from the 2001 Population Census.

to the numbers of people buying 'lunch' at different times during the working day.

In fact, many distributions, particularly those that exist in geographical space, do not follow the normal curve. This may not be immediately apparent but a little casual reflection confirms that when it comes to size, competition ensures that not all events or objects can exist at the largest size and that it is more likely to find populations whose sizes are skewed toward the smaller end of the spectrum. Take incomes, for example. There is only one Bill Gates with an income in the multi billions but there are many John Smiths with modest incomes somewhere near the bottom of the pile. Geo graphy is also reflected in this kind of competition. For example, there is only one New York but there are many small town Americas, in terms of city size. In fact, when we plot these kinds of distribution, we no longer get a normal curve of size versus frequency but something that is called a lognormal – a uni modal, somewhat distorted, bell like curve which is highly skewed to the left, as shown in **Figure 1b**, for the top 458 urban areas in Great Britain from the 2001 population census. The reason why the curve is called the lognormal is that were we to transform the x or size axis logarithmically, then we would recover the normal distribution. Like the normal, this is an ideal type, but in fact, distributions that ap proximate this form are much more common than has been supposed, for they imply generative processes in society which reflect competition, one of the key features of human sorting that marks the distinct patterns that arise and emerge within geography.

Competition operating at the individual level is a key feature of human systems and this does not usually generate patterns where the majority of events and their attributes are ordered in the manner of the bell curve. It depends, to an extent, on what is being measured but if this is size of the events generated, then these are usually distributed lognormally. In terms of incomes, the num bers of the rich are dwarfed by the poor but at the very poorest levels, the numbers are also small because of lower limits on survival and various subsidies that are operated to ensure that these lower limits are met. In terms of the process of generating wealth, then the probability of getting richer gets progressively less as income rises; that is, if the process is random, which to an extent in society is the way things sort themselves out, then the probability of getting wealthier through a suc cession of large, random increments of wealth over time – in short, 'luck' – is increasingly unlikely. There are many simple models of such processes that illustrate how competition generates the kind of heterogeneity and extreme order that comes to pass.

It was Pareto in the late nineteenth century who first drew attention to this phenomenon, and many extreme or heavy tailed distributions which approximate the log normal in **Figure 1** are casually refereed to as 'Pareto' distributions. He is credited with coining the cliché (Pareto's law) that "20 percent of the population own 80 percent of the wealth" which is consistent with the dis tributions generated by such competition. In geographical terms, the best examples are cities which follow the same kind of lognormality. Their patterning and size can be explained by many models ranging from the simple sto chastic ones, just referred to above, to more elaborate conceptions based on central place theory that leads to the skewed order shown in **Figure 1b**: very few large towns, and very few small towns (because towns have to reach a certain threshold before they are classed as towns), with most towns being at the smaller end of the scale. These are the patterns that we find in citation. They exist not only for the productivity of scientists but for the productivity of places where this science is generated.

Working with Statistical Distributions

To work with distributions, such as the frequencies shown in **Figure 1**, we need to make some simple transformations that help us visualize these patterns in a more intuitively appealing way. What we first do is to form the cumulative distribution by adding the frequencies for all objects, up to the size in question, from the graphs shown in **Figure 1**, thus generating 'S' shaped curves, symmetrical for the normal distribution and somewhat asymmetrical for the lognormal. We then take these distributions, working backward from the largest sizes to the smallest, first noting the size for greatest cumulative frequency that is the size of the largest event, then the next cumulative frequency, and so on, plotting these sizes against the adjacent differences in cumulative frequencies taken backward. These values are the rank of the object in question. We thus form a 'rank size distribution' where the plot gives the size of the object on the vertical and the rank on the horizontal axis. Strictly speaking, this is the counter cumulative distribution and it is much easier to work with. The shape of this curve is downward sloping and as the greatest ranks are approached, the curve dips slightly. **Figure 2a** shows this pattern of the lognormal in **Figure 1b** and it is instructive to note that the largest sizes of object are the highest ranks where the distribution is now skewed to the left, but the smallest numbers are given on the curve to the right which is slowly declining: this rightmost curve is referred to as the 'long tail' whereas the upper tail is called the short, fat, or heavy tail. The long tail is the subject of much speculation in social systems for there are some who argue that it is not the short tail that is important for this is where the action has already happened – in big cities and among those with high incomes – but it is in the long tail that all the action resides because this is where growth will come in the future. To be a big city you need to have been small city first, to be a highly cited scientist, somewhere along the way you must have been somewhat less frequently cited scientist, and so on.

In many systems, researchers have taken the largest objects and tended to disregard the smallest for two reasons: first, size dominates our thinking for as the Pareto rule implies, if 20% of the people own 80% of the wealth, then we can go a long way by simply studying and measuring these top 20%. The second reason is that before computers and digital data, this bias toward the largest objects was reinforced as one could not handle the entire data using less automated methods. Thus, until quite recently, data and information might only have been collected for the largest objects in any set. We demonstrate this somewhat unwitting limit in our first foray into citation analysis in the next section. If we accept that our analysis is so structured beginning with the largest events, then what we are examining is the short or heavy tail, the left most portion shown in **Figure 2a**. This curve can be extremely well approximated by a simpler curve than the lognormal, by a power law that relates the size of the object $P(r)$ to its rank r through the

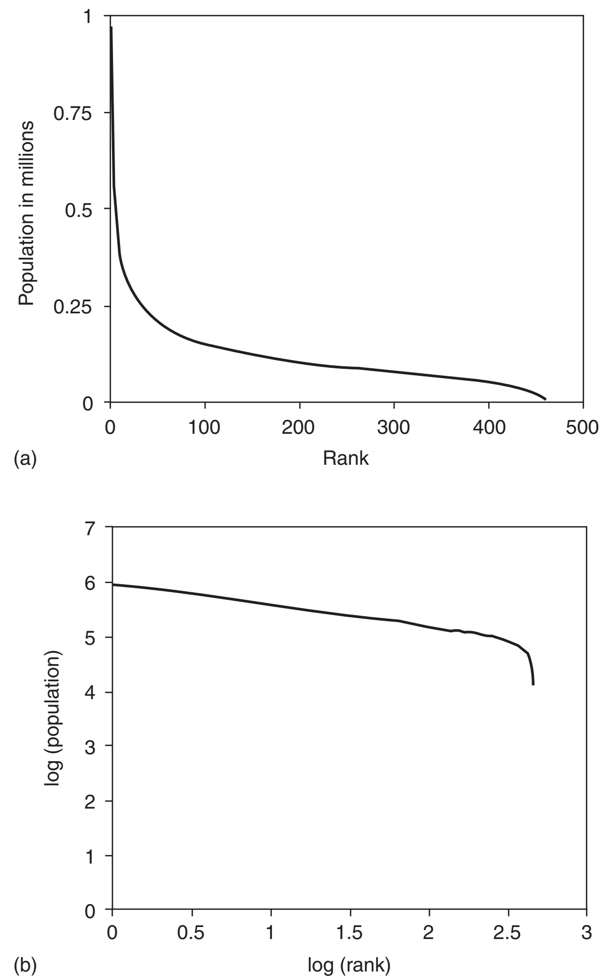

Figure 2 Rank-size plots: (a) The untransformed counter-cumulative distribution and (b) the logarithmically transformed distribution.

equation $P(r) \sim 1/r^{\alpha}$. This is an inverse power law with parameter α which controls the rate at which the size of the object declines with increasing rank r.

The rank size distribution is an example of a scaling relationship that means that at whatever scale (i.e., magnification of the graph) one examines the relation, it always looks the same. If this holds, and the evidence is that in the upper tail it does, then the same relation pertains to highly cited scientists as all scientists, indeed for any group of scientists taken within specific limits of productivity. Far from being unusual, these kinds of scaling laws actually appear ubiquitous, and as seen further, no matter how we carve up such a distribution in terms of the dimensions on which we measure it, we appear to get scaling laws which can be visualized as rank size plots. Before we conclude this brief foray into methods, we should note that it is particularly easy to fit a curve to a power law. If we logarithmically transform the size and frequency as log $P(r)$ and log r respectively, the curve in **Figure 2a** becomes a straight line as in

Figure 2b: the model is thus $\log P(r) = \log K + \alpha \log r$ where $\log K$ is the constant or intercept of the line and α is the slope which can both be determined using standard statistical methods.

It was George Kingsley Zipf, in 1949, in his book *Human Behavior and the Principle of Least Effort*, who first popularized the notion that scaling distributions could be ranked in this way. He demonstrated that for city size in the United States, the parameter value on the rank was equal to unity, meaning that if you took the largest city's population (New York), and divided it by the rank of the city in question, this would give the population of that city – try it! Or if you knew the population of the city, dividing its population into the largest city population would give its rank. In fact, this simple rule of thumb has been generally shown to be a special case but it still remains a benchmark against which other scaling relations are compared.

Ordering by Citation

A quick way to reveal how significant this kind of scaling is for citation analysis can be generated by taking a group of serious scholars and examining their citation rates. To show how generic this is, we have taken the 27 geographers who are Fellows of the British Academy (BA) in 2007, and simply keyed their names into Google Scholar, which is a widely used database pertaining to publications. It is well known that Google tends to undercount citations and thus this kind of analysis can only ever be preliminary to a full data mining of all citations associated with any particular scholar. Each scholar in this list publishes widely but they represent different fields of geography with different practices. What we have counted for each is their most heavily cited work and when we rank these counts (all 27) we generate the rank size plot which is shown in **Figure 3a** in logarithmic form. This is not a pure power law by any means. It seems that there are two distinct groups here, for if we were to fit the scaling model, we would identify a break into two regimes. This might be due to the differences within the field of human geography within the social sciences, with historical geography being less cited and different in publication focus than, say, more mainstream economic and urban geography. It might also relate to the way text books are cited which is different from research papers. But it does show that even from the most rudimentary data, these kinds of distribution are nowhere near being normal.

Our focus here is not on relative ordering of individuals *per se* but on the geographic distribution of citations which we can extract by suppressing individual authorship but count locations of where scientific papers and books are produced. In this way, we can show how patterns of concentration, with respect to authorship, are

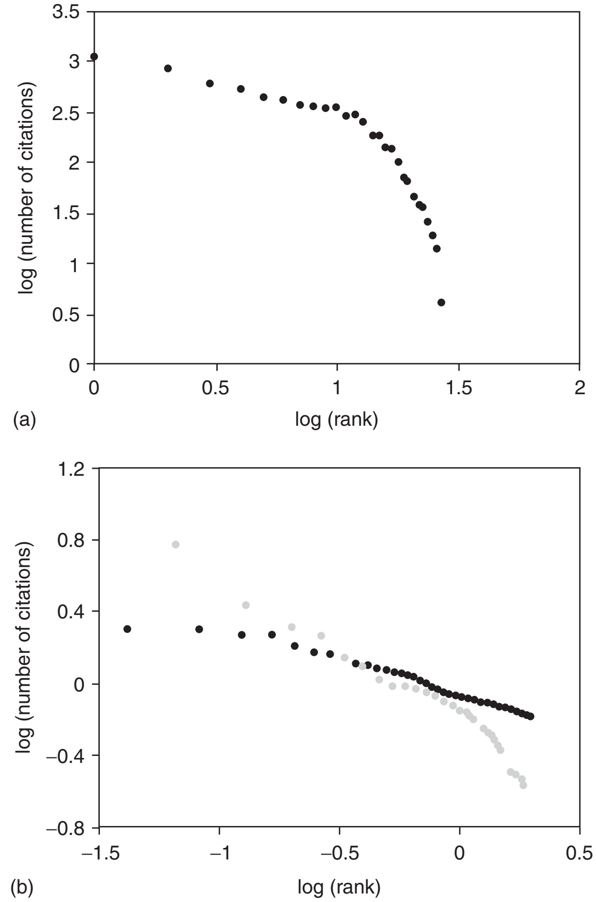

Figure 3 Zipf plots of scientific productivity: (a) The British Academy data and (b) The Economic Geography data; black points are article citations; grey points are book citations. From Foster, J., Muellerleile, C., Olds, K. and Peck, J. (2007). Circulating economic geographies: Citation patterns and citation behaviour in economic geography, 1982 2006. *Transactions of the Institute of British Geographers* 32, 295 312.

replicated when it comes to looking at where knowledge (as revealed through publications, of course) is actually produced. In this sense, citation analysis can be seen as part of the geography of scientific activity which, in turn, is indicative of the way the economy organizes itself spatially. Apart from the social interest in who produces what and their productivity and who cites who, this kind of analysis is central to the new economic geography which is deeply rooted in the economics of spatial agglomeration and growth.

A slightly more elaborate analysis by authors of works in economic geography has been developed recently by Foster, Muellerleile, Olds, and Peck, in 2007. They list the most highly cited authors in economic geography from 1982–2006 by their most highly cited article and text books. We show their data for the top 47 authors and top 29 books in **Figure 3b** as Zipf plots. There is much more evidence here of classic scaling than in the BA case and it is of interest to note that books are much more heavily cited than articles. This reflects many differences between writing and publishing a book versus an article. Both examples however, for the economic geographers and the BA Fellows, involve a politics of selection. Who we choose to select and how they are selected makes a difference which injects a level of caution into any such analysis. One last methodological note: In **Figure 3b** which compares authors' top cited articles against books, the plots are based on the rank size data which is normalized by dividing through by the average rank $\langle r \rangle$ and average size $\langle P(r) \rangle$ before the data is logged. This enables the graphs to be collapsed onto the same scale, and hence, easily compared.

Extracting the Geography of Citation

From the data shown in **Figure 3b**, Foster, Muellerliele, Olds, and Peck in 2007 also extract the location of the authors writing the highly cited articles and books and although they only list this data by country for various years, it is clear that the data decreases in the country citation counts with respect to rank in a manner similar to a power law. A slightly more considered approach to such analysis is to extract the geographical coordinates associated with the addresses of the authors producing these works, and our first demonstration is from the ISI's 'HighlyCited' Database. This is a list, of the most highly cited scientists in a number of different fields, that records not their citation counts but their institutional locations. Currently the database which is located as a free resource at http://www.isihighlycited.com contains 6611 scientists from 21 different fields. We have used past data from December 2002 where we selected 12 of these subject areas and examined the geographical locations of 1222 scientists listed. There are obvious limitations in such data sets but notwithstanding difficulties with respect to the 'institution' data relating to classifying where the researchers work, the data provides a good 'impressionistic picture' of this geography.

The pattern of concentration that this analysis reveals is remarkable. 1222 scientists work in 429 institutions which are located in 232 places in 27 countries. Almost half of these scientists are in 50 institutions in five countries, most being in the United States. The concentration increases as we aggregate the data from institution to place and thence to country. In **Figure 4**, we show these data sets as Zipf plots where we plot the log of the number of scientists at and in each institution, place, and country against the log of their ranks which as above, have been normalized by their means. Collapsing each data set in this way shows, quite clearly, how the concentration increases as we aggregate the data into places and countries. Somewhat serendipitously, we have fitted power laws to these plots, where for each of the three

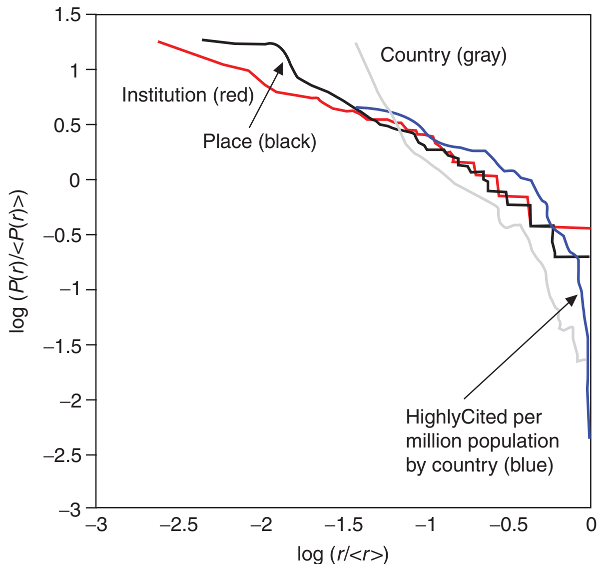

Figure 4 Rank-size distributions of highly cited scientists. The blue line is the plot for countries, normalized by population in millions, which illustrates a different pattern of concentration from the basic data. Linear plots have been fitted to the basic data using $\log(P(r)/\langle P(r)\rangle)\quad \log K\quad \alpha\log(r/\langle r\rangle)$. For institutions $\alpha\quad 1.049$ ($R^2\quad 0.962$), for places $\alpha\quad 0.816$ ($R^2\quad 0.938$), and for countries $\alpha\quad 1.997$ ($R^2\quad 0.949$). All these values are significantly different from zero at the 99% level.

respective aggregations, the value of the power is related to the degree of concentration (Redner, 1998).

Table 1 shows the top ten countries in terms of the number of scientists and places where they work. The ratio of scientists to places provides another measure of the concentration with the implication that the larger the number of highly cited in each country, the more likely they are to be highly concentrated in a small number of places. If we normalize the data by population, we get a slightly different picture; the top five are now Switzerland, US, Sweden, UK, and Israel with smaller countries becoming more significant. We might even consider normalizing by the square of population reflecting the potential interaction within a population although, for the scientific enterprise, this probably is not a good measure of where such interaction takes place. All this does is to sharpen the index even further with small countries dominating. Similar analyses for institutions and places are more complicated as the choice of a population for the normalization is uncertain. College towns begin to dominate, for example.

A particularly graphic indication of the basic pattern is illustrated in **Figure 5** where we have mapped the main locations of places by circles proportional to the number of cited scientists. Forty percent of the most highly cited scientists work in ten places of which nine are in the US. These locations bear out our perceptions of where the world's top institutions are most heavily concentrated: on the west coast of the United States, the

Table 1 Top ten ranking of highlycited scientists by country

Rank	*Country*	*No. of highlycited*	*No. of places*	*Concentration: scientists/places*	*HighlyCited per million population*
1	US	815	90	9.06	3.16
2	UK	100	24	4.17	1.72
3	Germany	62	21	2.95	0.78
4	Canada	42	15	2.80	1.53
5	Japan	34	14	2.43	0.27
6	France	29	11	2.64	0.50
7	Switzerland	26	5	5.20	3.78
8	Sweden	17	2	8.50	1.96
9	Italy	17	10	1.7	0.29
10	Australia	17	9	1.88	0.96

Boston Washington megalopolis on the east coast, central London and the golden triangle formed with Oxford and Cambridge, Chicago, and interestingly, in the cluster of towns around Research Triangle Park in North Carolina. We have not yet examined the local detail of where these institutions are located but casual knowledge suggests that these are even more highly clustered at ever finer scales. For example, the institutions in Boston are all within a 2 mile radius of the MIT Museum whereas in London they are within a 3 mile radius of the British Museum. At an even more local scale in central London, for example, the majority of the scientists cited are located within half a mile of Euston station in Bloomsbury. This is remarkable evidence of urban agglomeration and specialization.

Extracting Spatial Productivity from Massive Data Sets

It is also possible to extract the geography of scientific activity from data sets that actually contain articles themselves, without referring to citation, as many articles are now online. One of the most accessible archives was set up, in 1991, as a resource for preprints in theoretical physics. It is known as the 'arXiv' and by December 2004, it contained 303 614 accessible papers containing some reference to the geography of the authors. Currently the average submission rates to the 'arXiv' are about 3600 papers per month and it has come to reflect an essential resource for pre publication and peer review. Papers are often submitted in LaTeX but the source encodes author and address data in specific fields, which can be parsed using custom scripts (e.g., written in Perl), to extract the geographical location of the authors. There are many problems in doing this but we have extracted the geog raphy of paper production for 2326 papers produced in 2004 which had one or more US authors, but of these, some 517 or 6.5% were unusable in that we could not extract the geo coordinates. This data are plotted in **Figure 6** with the Zipf plot alongside where we have fitted this not with an inverse power law but with a stretched exponential. This gives an idea of the kind of resources that are now becoming available from which are likely to revolutionize the study of economic geog raphy in an online world.

A larger and more reliable data set we have worked with is the computer science archive known as 'Citeseer'. The Citeseer digital library locates computer science articles on the web in Postscript or portable document format (PDF) format and extracts citations from and to different documents. It has made its metadata available online and the inclusion of address and affiliation fields for each author allows a rigorous analysis of the geography of a very large bibliometric database. The 'Citeseer' metadata contains 716 772 records, some of which are repeated and some of which have authors with empty address fields. About 53% of these records identify all authors and their respective addresses. Out of these 379 111 unique papers, we analyzed the 128 348 (or 34% of these papers) that have one or more US authors. This is in reasonable ac cordance with Thomson ISI global indicators, which state that between 1997 and 2001, the United States produced some 36% of the world's highly cited publications. For each paper, we extracted the five digit Zip Code from each author's address field and geocoded this Zip into an *x,y* pair of coordinates. We identified Zip Codes from the address field by using regular expressions to match a five digit code (plus the optional four digit code, which we ignored), preceded or followed by a US state (or its ab breviation) or the acronym USA. We restricted the an alysis to the 48 conterminous US states plus the District of Columbia. This gave a total of 116 771 distinct authors with a US address. Out of these, 103 928 papers/authors list a single Zip Code in their address, 10 579 belong to institutions located in two Zip Codes, and 2264 are located in three or more institutions.

To investigate the concept of scaling in publication output of academic research centers, we computed the probability distribution of total paper output per Zip Code. If two research centers belonging to the same institution have addresses with distinct Zip Codes, we considered them as distinct centers. Our analysis identified 3393 different Zip Codes that matched the US

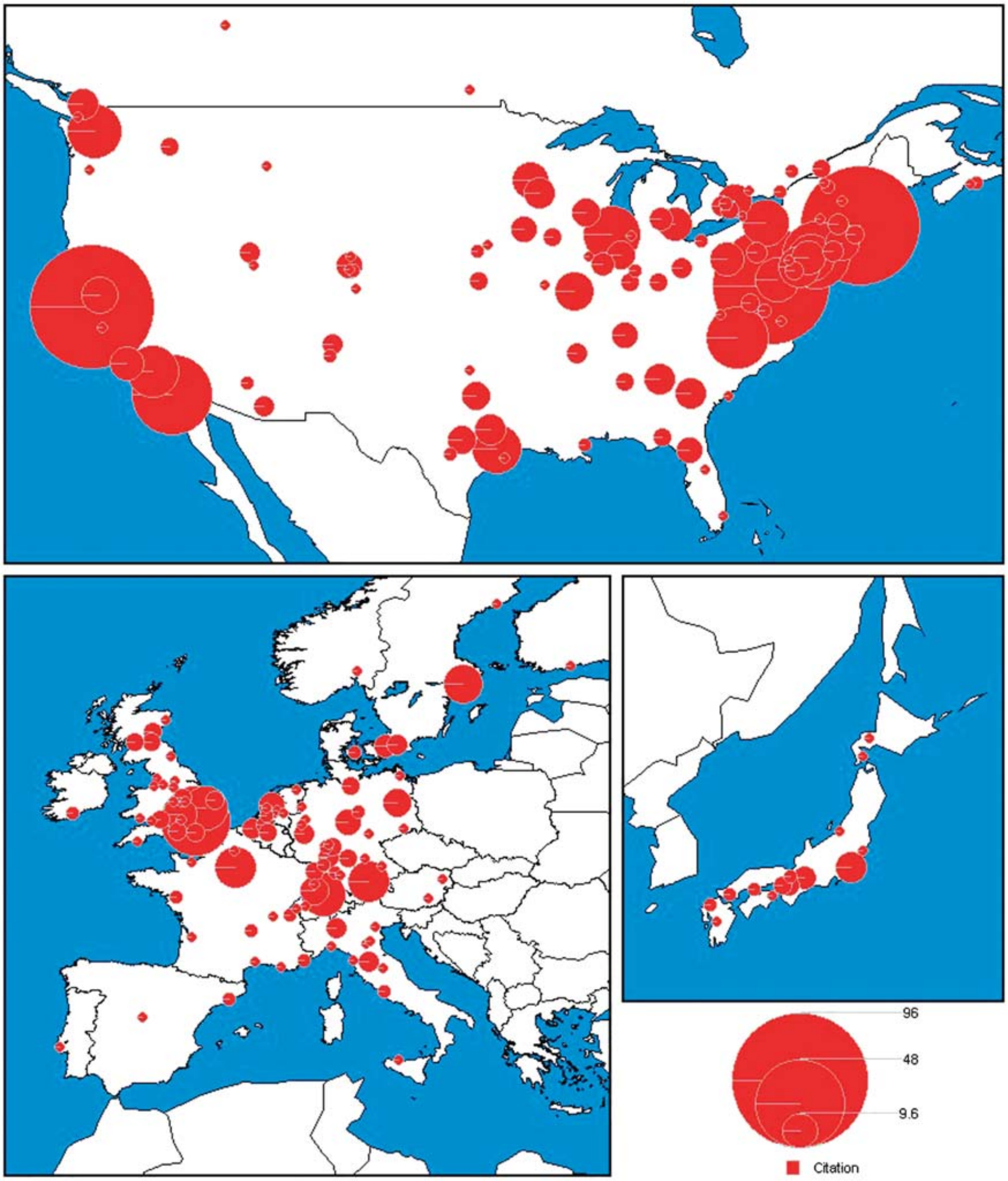

Figure 5 The geographical distribution of the HighlyCited: 12 subject areas with 1221 scientists (December 2002).

census bureau tables and we implemented fractional counting to compute the productivity of US research centers. For every paper, we parsed each author's address field and extracted the Zip Codes therein (there may be more than one Zip, if the author belongs to more than one US institution). From this data, we are able to plot the rank size scaling as well as maps of the concentration and/or dispersion of the productivity of centers by Zip Code in the continental USA. A Zipf plot of the 'Citeseer' data is shown in **Figure 7** and there is evidence here of two distinct regimes, not unlike the BA data, and this suggests that there are two kinds of universality at work here. We have also mined physics papers using the main 'Google' search engine, notwithstanding the caveats made earlier, and we show the Zipf plot for this alongside the 'Citeseer' data in **Figure 7** where it is clear that this gives a similar profile to the computer science papers. It is somewhat speculative to suggest that this is evidence of universality among all scientific papers but the notion of at least two regimes of competition at work in scientific productivity is appealing.

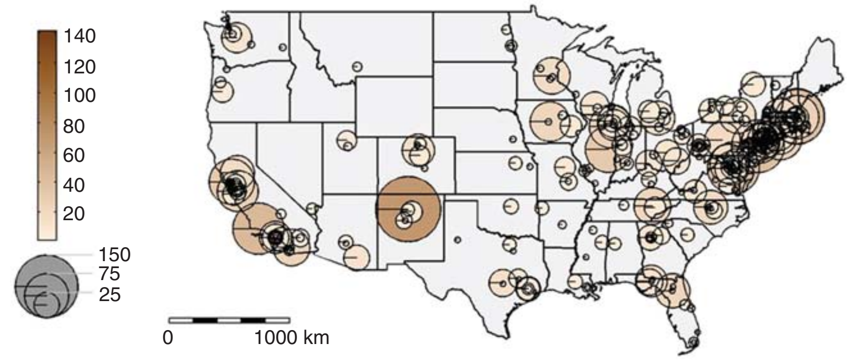

(a)

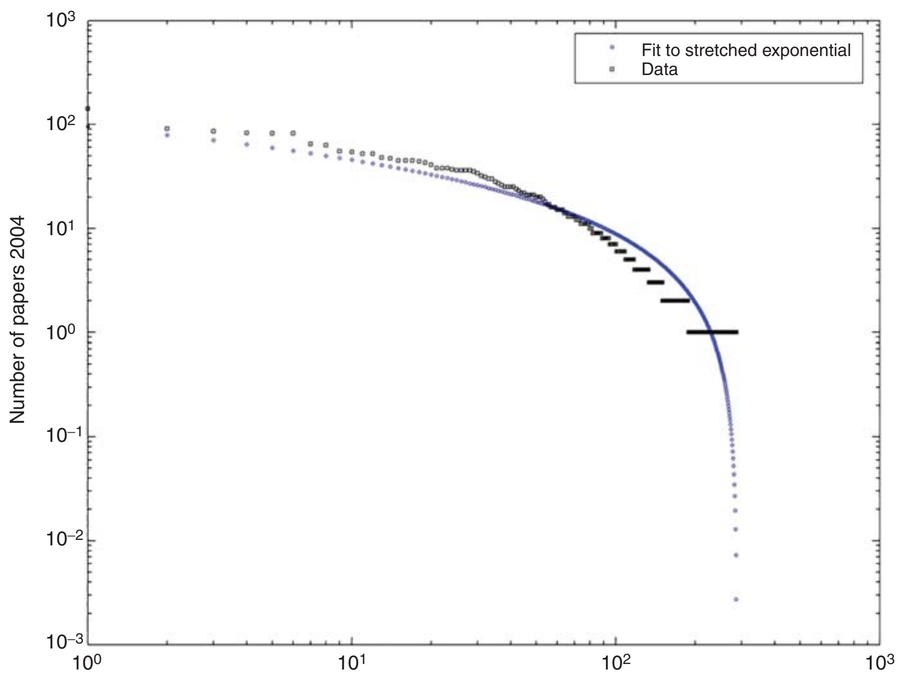

(b)

Figure 6 Physics articles produced in the US in 2004 from the arXiv: (a) Geographical locations (Top 287 Zip Codes) from the 2004 'arXiv' Data and (b) a Zipf plot of data ranked across Zip Codes.

We will not discuss the data and the scaling further, for we simply focus on the geographies shown in **Figure 8**. What these maps show is scientific productivity by center where each circle is proportional in area to the logarithm of the productivity count. **Figure 8a** shows this for equal map areas in conventional terms and this bears out all that we have been saying about the non normality and extreme nature of geographical distributions as well as the scaling inherent in such patterns. However, it is instructive to compare these patterns with other economic and demo graphic measures, such as the pattern of research and development (R&D) expenditures, and population size from year 2000 censuses. A technique for distorting the map shape to approximate the pattern of R&D or popu lation is based on redrawing the boundaries of the state units used to form the map to reflect the extent of R&D or population. In short, instead of actual land area, the area of each unit is now proportional to some other measure of location. This will make units that are larger, in terms of economic or demographic activity, larger than units with smaller activities, notwithstanding their actual land areas. Various techniques to generate such cartograms are available and we have used Gastner and Newman's dif fusion technique of 2004 which is also used in mapping poverty and equality in the 'WorldMapper project'. What **Figures 8b** and **8c** show is how the USA is distorted in area to reflect the concentration of highly productive re search centers – in computer science – which reflect the population and R&D expenditure in southern and mid California, the northeast seaboard and the Midwest, with the mountain states massively collapsed and forming a narrow bridge between the east/Midwest and the west

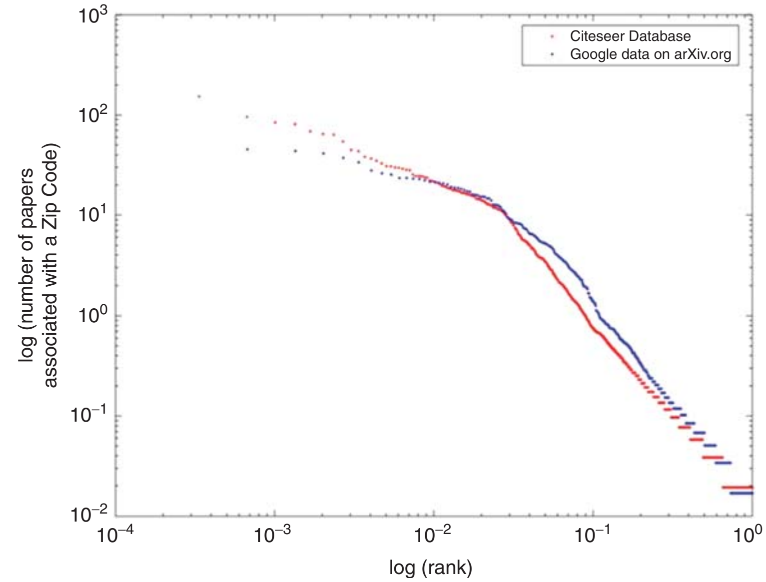

Figure 7 Zipf plots: Citeseer for computer science and Google data for physics. The plots show the actual computer science data from 'Citeseer' against an alternative analysis of physics papers taken from the 'Google' Search Engine.

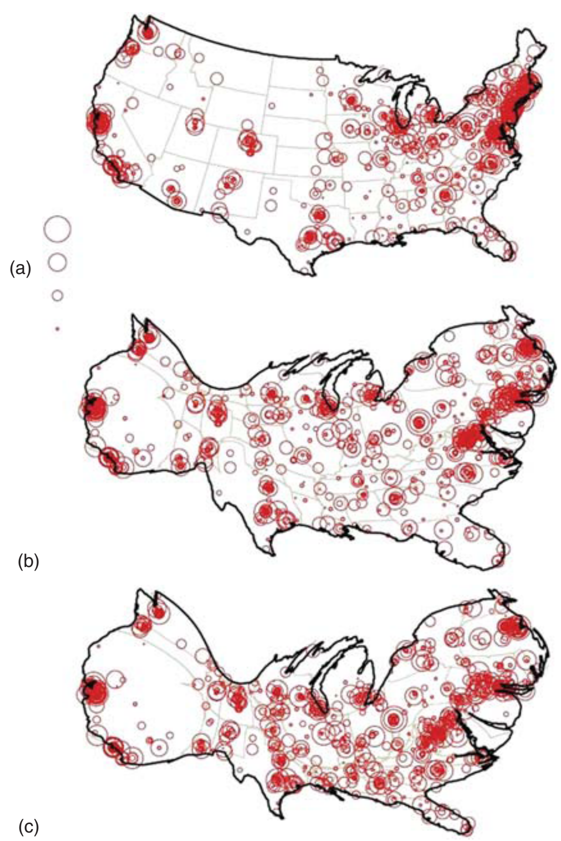

Figure 8 Equal area and cartograms displaying the number of articles produced by US computer scientists from the Citeseer database 1997 2004: (a) the data on an equal area map; (b) the data on a cartogram of US State areas reflecting R & D (2000); and (c) the data on a cartogram of US State areas reflecting population (2000).

coast: such is the economic geography of contemporary America.

Related Work

The development of citation analysis in the last decade has also taken a new twist with collaboration between scientists being studied, now. This maps easily onto work in social networks where the development of network science has become central to the development of new sciences that focus on relationships between entities. Citation networks have been studied for many years but only with the availability of online resources that can be mined digitally has this area become significant in that very large topological networks and their structure can now be analyzed. New measures of network structure as well as a classification of networks into different struc tures, such as random graphs, small worlds, and scale free networks (which have scaling and fractal properties similar to those introduced here), have become signifi cant with the notion of developing or growing networks becoming important. Mapping such topologies onto geographies as maps and cartograms has not yet occurred although this is likely before long as digital data mining comes of age in terms of its visualization. Katy Borner's *Places & Spaces: Mapping Science* shows how we might visualize new scientific data resources, products, and ideas in terms of the places that they originate from and are related to.

This article has not been simply about citation geography although this is a relatively new and novel way of illustrating geographical phenomena. It is as much about new methods for mining digital data and extracting geographic content, about understanding the patterns which are non normal and often embody extreme events, and about new ways of visualizing this data, first as Euclidean maps and their distortions as cartograms but later as Zipf plots and ultimately as social and physical networks. The challenge in this field is to link such techniques to distributions and processes in economic geography in particular and the wider panoply of geographic distributions in general.

See also: Geographical Journals.

Further Reading

Adamic, L. A. (1999). *Zipf, Power laws, and Pareto A Ranking Tutorial*. Palo Alto, CA: Information Dynamics Lab, HP Labs. http://www.hpl.hp.com/research/idl/papers/ranking/ranking.html (accessed Aug. 2008).

Anderson, C. (2006). *The Long Tail: Why the Future of Business Is Selling Less of More*. New York: Hyperion.

Bak, P. (1999). *How Nature Works: The Science of Self Organized Criticality*. New York: Springer.

Batty, M. (2003). The geography of scientific citation. *Environment and Planning A* 35, 761 765.

Carvalho, R. and Batty, M. (2006). The geography of scientific productivity: Scaling in US computer science. *Journal of Statistical Mechanics* 10(P10012), 1 11.

Foster, J., Muellerleile, C., Olds, K. and Peck, J. (2007). Circulating economic geographies: Citation patterns and citation behaviour in economic geography, 1982 2006. *Transactions of the Institute of British Geographers* 32, 295 312.

Garfield, E. (1983). *Citation Indexing, Its Theory and Application in Science, Technology, and Humanities*. Philadelphia, PA: ISI Press.

Gastner, M. T. and Newman, M. E. J. (2004). Diffusion based method for producing density equalizing maps. *Proceedings National Academy Science USA* 101(20), 7499 7504.

King, D. A. (2004). The scientific impact of nations. *Nature* 430, 311 316.

Leydesdorff, L. (2007). The interdisciplinarity of its environment and the citation impact. *Environment and Planning B* 34, 826 838.

Matthiessen, C. W. and Schwarz, A. W. (1999). Scientific centres in Europe. *Urban Studies* 36, 453 477.

Newman, M., Barabasi, A. L. and Watts, D. J. (2006). *The Structure and Dynamics of Networks*. Princeton, NJ: Princeton University Press.

Pareto, V. (1916). *The Mind and Society: A Treatise on General Sociology*, Borngiorno, A. and Livingston, A. (trans., 1935). New York: Harcourt, Brace, and Javanovich.

Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *European Physics Journal B* 4, 131 134.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison Wesley.

Relevant Websites

http://en.wikipedia.org
Citation Index; Citation, Wikipedia.
http://citeseer.ist.psu.edu
CiteSeer.IST, Scientific Literature Digital Library.
http://www.arXiv.org
Cornell University Library.
http://www.isihighlycited.com
ISIHighlyCited.com.
http://scimaps.org
Places and Spaces.
http://www.worldmapper.org
Worldmapper.

Citizenship

V. Chouinard, McMaster University, Hamilton, Canada

Glossary

Capitalist State The institutions, relations, and practices associated with the state in capitalist societies.

Citizenship The status of being a citizen, membership in a community, and how an individual responds to membership in a community.

Citizenship Formation The processes through which relations and practices of citizenship are formed in particular places and regions.

Civil Society The nonstate voluntary and civic organizations and institutions contributing to the functioning of society.

Differencing Processes through which people become more and less marginalized in society and space.

Globalization An economic, social, cultural, and technological process through which nation-states, communities, and households are increasingly integrated around the world.

Postcolonial Theories Theories concerned with the political and cultural position of persons formerly subjugated by colonial powers.

Introduction

Citizenship refers to particular ways of being situated within and responding to relations of power through which a community is governed or ruled. It involves claiming, exercising, and contesting rights, entitlements, and obligations (e.g., rights to vote and strike, the obligation to pay taxes) and diverse ways of engaging with the institutions (such as the state) and relations through which communities are constituted and governed. Citizenship has many geographical implications – determining where we can and cannot live, what rights we can claim in particular places and spaces of life, what obligations we are encouraged to fulfill, and the extent to which we can or cannot influence the regimes of power to which we are subjected in particular places and nations.

Geographers' interests in citizenship date back to the classical writings of thinkers such as Strabo and Ptolemy. However, it has only been since the 1970s, with the rise of radical and feminist geography, that geographers have given more sustained critical attention to citizenship as a sociocultural and political–economic phenomenon. Prior to this, political geographers tended to study citizenship tangentially, examining particular spatial manifestations of it such as voting patterns, voting districts, or national borders. Radical, feminist, and critical social geographers, in contrast, have sought to understand citizenship as complex, geographically uneven ways of engaging with relations of power through which particular communities are governed – engagements that occur both in and outside the state. Far from being universal, access to citizenship was increasingly seen as socially and geographically uneven; reflecting how diverse people were located as more and less entitled and marginalized members of real and imagined communities constituted at a range of geographic scales. This article discusses this more recent body of social geographic scholarship.

Radical Social Geographic Perspectives on Citizenship

The rise of radical political economic inquiry in human geography, during the 1970s and 1980s, was associated with growing attention to issues of social justice and to how different groups were situated within relations of power in capitalist society and space – particularly in terms of differences in class locations. Geographers were also increasingly concerned with the capitalist state and civil society as terrains of power contested by different groups of citizens. Fincher, for example, examined struggles over urban renewal schemes in Boston which would attract corporate capital but also displace low income residents.

As early issues of Antipode indicate, radical geographers were concerned with a wide range of topics related to citizenship. While relatively few articles explicitly discussed citizenship *per se*, topics relating to relations, experiences, and practices of citizenship included: advocacy in planning, neocolonial relations of power ghettoizing poor people of color in the slums and barrios of First and Third World cities, racist violence in response to school desegregation and busing in cities such as Boston, and 'socialist citizenship' in academic geography. In such various ways radical geographers helped to put issues of power, oppression, and marginalization in society and space associated with differences in income and class location, gender, race, and even ways of practicing human geography, on the geographic agenda. Further, by beginning to explore how geographic research could be used to promote social justice, radical geographers helped to encourage new visions of the roles that geographers could play in not only understanding the world but helping to change it. Feminist and critical social geographers continue to build upon these legacies to help advance our understanding of how and why so many citizens continue to be denied rights and entitlements in the neoliberal societies of the early twenty first century.

Feminist Contributions to Rethinking Geographies of Citizenship

From the 1970s onward, the second wave feminist movement, and the rise of 'New Left' scholarship and activism, helped to inspire feminist scholars to unsettle prevailing ways of understanding the state, relations of ruling, and citizenship. During the 1970s and 1980s, feminist scholars challenged 'gender neutral' conceptions of the state – helping to reveal how gender relations oppressive to women were reproduced through state rule. Catherine MacKinnon, for example, used a radical feminist conception of the state as perpetuating patriarchal forms of oppression to help explain why it was that female rape victims were 'revictimized' within Western judicial systems through court proceedings which closely scrutinized their past sexual histories. Carole Pateman challenged prevailing conceptions of citizenship as based on universal contractual relationships, arguing for greater attention to the sexual contract through which women's patriarchal subordination to men was secured (e.g., marriage). What feminist critiques of the state and relations of ruling did was to bring apparently 'private' matters, such as the sexual subordination of women by men, into the public sphere of the state and citizenship. Gender, class, and other differences such as ethnicity and race were, it was argued, critical in determining which citizens were more or less empowered through the state and its relations of ruling. Feminists were also concerned with critically assessing the extent to which two decades of feminist struggles in and against the state, around issues such as women's health, child care, and gender discrimination in employment, were translating into laws, policies, and implementation practices that made empowering differences in women's lives. For example, challenges in negotiating relationships between nongovernmental feminist organizations and 'femocrats' in the Australian state bureaucracy and in promoting affirmative action through the state were among the topics tackled in a book aptly entitled *Playing the state*. A key question in such state centered analyses was whether and if so how feminist activism was pushing the local state in directions which began to undermine gender relations of power in capitalist societies.

Feminist geographers began to rethink relations and practices of citizenship during the 1980s through studies examining the role of the capitalist state and classed and gendered relations of ruling in perpetuating women's subordinate position in society and space. Early work focused on the role of the state and planning processes in reproducing gender relations and sociospatial divisions of labor oppressive to women. Bowlby, for example, examined how post war planning practices in Britain gave rise to suburban environments designed to reproduce women's domestic roles in the home. As in the wider feminist literature, the state was recognized as a critical site of struggle over policies and practices shaping women's lives in areas such as employment equity, the provision of services such as daycare and public transportation, the availability of welfare supports, and planning the environments that women would live in. Much of this work was socialist–feminist in orientation, treating gender and class relations as pivotal, intertwined relations of power involved in reproducing capitalist societies and ways and places of life.

The 1990s saw feminist geographers engage in more explicit ways with questions about the relations and practices of citizenship in which women (and men) were engaged and the socially and spatially complex ways in which struggles over rights unfold in late capitalist societies. Feminist geographers made very important contributions to rethinking citizenship as a gendered set of relations and practices by challenging traditional conceptions of political acts as occurring exclusively in formal political spaces (such as state legislatures) pointing out not only that private spaces such as the home may serve as important spaces for political organizing amongst women, but also that women's political acts often blur the boundaries between the private and public and challenge assumptions about formal and informal politics. Feminist geographers helped to demonstrate that many women, denied full access to formal political spaces of life, have to be creative in making spaces for political action by, for example, making public what is usually private – through collective displays of individual grief at the disappearance and death of their children at the hands of an oppressive political regime in the case of the Madras (Mothers) of La Plaza de Mayo in Argentina for instance. Political actions may cross private and public boundaries, do not always involve direct engagement with the state, and may have effects that are private, public, or both. By thinking about gendered political acts in this broader, more spatially fluid way we can also appreciate that citizenship is much broader than formal political activities such as voting.

As the 1990s progressed, feminist geographers increasingly turned their attention toward what Kobayashi termed processes of differencing in society and space – which enabled and empowered some citizens and disempowered citizen and noncitizen 'others'. Feminist geographers began to explore how multiple other embodied differences, in addition to gender and class, empowered some and disempowered others. Growing concern with how a multiplicity of differences situated people in relation to rights of citizenship was associated, in and outside geography, with a shift away from state centered analyses of the forces shaping citizens' lives. Increasingly, feminist geographers understood citizenship as played out through multiple differences that situated diverse people in relation to regimes of power in civil society and the state, and in informal and private, as well as formal public spaces of political activity. Sexual citizenship, for example, was an important topic placed on the feminist geographic agenda during the 1990s. Valentine's early work examined how spaces such as urban streets are coded as for those with heterosexual identities and how this helped to marginalize those with non heterosexual identities and ways of taking up space. Building on such work, Bell examined what he termed the spaces of sexual citizenship: "… those locations where a citizenship is constituted through the citizen's sexuality…." He argued, using a case in which male same sex sadomasochism was actively disciplined through the British state and legal system, that such disciplinary regimes construct those with such dissident sexualities as citizen perverts. Geographers continue to be interested in sexually dissident forms of citizenship, as for example, Kitchin and Lysaght's work on sociospatial forces shaping sexual citizenship in Belfast, Northern Ireland, indicates.

Feminist geographers have also, in recent years, made important contributions to rethinking relations and practices of citizenship in transnational and postcolonial ways.

Fenster has examined how Israeli planning policies which have failed to respect differences between Ethopian immigrants and other Israeli citizens have led to the denial of citizenship rights for this group, for example, the right to a living space which accommodates extended family living arrangements. Kofman has examined the impacts of a multiplicity of political jurisdictions within the European Union on the status of groups such as migrant laborers, women immigrants, and refugees – showing how groups often constructed as 'outsiders' within civil society are marginalized in terms of abilities to claim full rights and entitlements of citizenship. More recently she has examined how Western European states have responded to migration pressures and concerns regarding social cohesion and national security by developing increasingly stratified systems of managing access to rights of citizenship and of promoting loyalty to nation states and prevailing ways of life. She shows how such post 9/11 managerial regimes place a growing emphasis upon the obligations of citizenship (e.g., respect for the rule of law, tolerance toward others, loyalty to the nation state).

As feminist geographers are helping to show, understanding how diverse people are situated in relation to rights, practices, and spaces of citizenship requires close attention to processes of inclusion and exclusion, differencing and oppression, which cross cut transnational, national, local, and personal scales. Pratt, for example, shows how the denial of immigrant status to female Filipino domestic workers in Vancouver, and their consequent dependence upon continued employment in order to maintain residence in Canada, leaves them vulnerable to employer practices which deny fundamental human rights such as the right to privacy. As authors such as Anderson stress, in a transnational, increasingly mobile world it is important that we look beyond processes of oppression and exclusion/inclusion such as racism to consider as well how discourses around 'belonging' (constructions of migrants for instance as 'outsiders') shape diverse people's capacities to claim and exercise citizenship.

Feminist geographers' efforts to advance our understanding of differential access to citizenship as the outcome of transnational processes such as managing migration have been complemented by research which draws on postcolonial theories. McEwan's work on gendered forms of citizenship and democratization in post apartheid South Africa stresses the importance of developing non Westernized conceptions of diverse women's citizenship. She discusses how black women in customary marriages in South Africa are denied fundamental rights of citizenship to an extent not found in Western nations – rights to inherit property or to engage in any political action in public spaces of life for instance – as a result of deeply entrenched, traditional structures of patriarchal authority. She calls for greater attention to how such women develop alternative forms and spaces of citizenship – for example, developing self help groups in the private sphere of the home in their struggle against domestic violence. Nelson has examined the local through global forces shaping how indigenous women in Cheran, Mexico, have struggled to redefine their roles as citizens and to break free of colonial, Catholic legacies which constructed women as not belonging in the public sphere. She details how women's struggles to establish markets for selling embroidery in other locales, struggles prompted by the declining viability of traditional subsistence practices with globalization, combined with men's and women's struggles in the late 1980s for more democratic local governance, to break down gendered barriers to women's political engagement and voice within the public sphere.

Feminist geographic research is clearly advancing our understanding of citizenship as a complex, contested sociospatial process. By exploring how relations and practices of citizenship develop within the context of a multiplicity of differences between and among women and men, and by insisting that ways of engaging in citizenship are far more complex than traditional conceptions of political action in formal political and public spaces allow, feminist geographers have shown how access to and engagement in citizenship is contested not only in relation to the state but in our homes and communities. Further they have helped us to understand differential access to and engagement in citizenship as the outcome of struggles that unfold across spatial scales that range from the intimately personal to the transnational.

Challenging Neoliberal Visions of Citizenship: Critical Social Geographic Perspectives

The rise of the New Right in nations such as the United States and Britain during the 1980s, with its neoliberal visions of citizenship as the freedom to pursue wealth through the unfettered operations of the market and its associated attacks on welfare state intervention, challenged New Left visions of citizenship as involving collective efforts to diminish structural inequalities in society. In geography, as in other disciplines, this challenge provided an important impetus to critical social and feminist ana lyses of citizenship formation as the socially and spatially uneven outcome of struggles over *de jure* and *de facto* rights and opportunities to participate in society. Influenced as well by the so called 'cultural turn' in geography and other disciplines, critical social geographies of citizenship emphasized not only the political–economic, but also the social and cultural relationships and practices of power in which citizens are enmeshed in particular places. As in feminist geographies of citizenship, this has encouraged closer attention to how differential access to and experi ences of citizenship in specific places are the outcome of processes which are simultaneously imagined and real – as much the product of discourses and imaginings about who belongs and who doesn't as of the material consequences of class, gender, and other divisions of power intrinsic to capitalist societies.

An important theme in critical social geographic ac counts of citizenship has been the critique of neoliberal and neoconservative policies and discourses aimed at promoting forms of social citizenship stressing obli gations and responsibilities as opposed to entitlements to rights and to full participation in society. Fyfe, for ex ample, has critically assessed the impacts of 'law and order' policies in Britain on citizenship practices, noting how an emphasis on active citizenship as individual ob ligations to assist in areas such as crime prevention and policing have eroded conceptions and practices of citi zenship as entitlement to rights. When combined with legislation such as the Public Order Act, an act which emphasized individual rights to the enjoyment and use of public space, as well as private property, over collective rights to freedom of assembly, the neoliberal state has actively promoted the devolution of obligations to indi vidual citizens and an erosion of collective entitlements of citizenship. Kearns has shown how the active citizen ship being promoted by the Conservative government from the late 1980s in Britain involved a shift away from more collectivist approaches to welfare provision toward an elitist mode of governance through which individual citizens were morally pressured to assume responsibility for assisting others through the voluntary sector.

In addition to critically assessing transformations in relations and practices of citizenship associated with New Right policies and legislation, critical social geographers have examined spaces of and possibilities for resistance to neoliberal individualized, consumerist forms of citizen ship. Pile has examined struggles to democratize local development processes in the London Docklands through a coalition of tenants' associations which insisted on people led forms of development. While not framing their struggles in the language of citizenship, this coali tion was able to empower local residents in the devel opment process thus challenging individualized and consumerist forms of participation promoted by the London Docklands Development Corporation. Brown has examined the multiple spaces of activism which developed in response to the AIDS crisis in Vancouver arguing that responses amongst activists within civil society both helped to open up spaces of radical demo cratic citizenship – for example, through a buddy system which challenged the alienation of those with AIDS from their biological families and the state – and closed other opportunities down when voluntary organizations pro viding services became increasingly subject to state regulation.

Critical social geographers have, like feminist geog raphers, stressed the complex spatiality of how rights and entitlements of citizenship are assigned and contested, showing that these processes unfold within multiple spaces of daily life and across multiple spatial scales. Blomley and Pratt have discussed the geographically uneven ways in which rights are assigned and contested in Canadian society and space. They note, for example, the importance of divisions between private and public space in determining the sorts of rights claims that can be made – how, for example, workers' rights to strike and picket have been deemed private rights and thus not protected constitutionally through the entitlements of citizenship laid out in Canadian Charter of Rights and Freedoms. They also point out how the bundle of rights recognized in particular spaces and places of life are subject to contestation; as for example, when residents in central city areas undergoing gentrification, such as Vancouver's Downtown Eastside, challenge individual ized conceptions of rights to property with more col lectivist visions of belonging to community and having a 'right to stay' in the face of displacement pressures. Struggles for entitlements to rights also unfold across geographic scales: when domestic workers use transna tional networks and discourses of human rights to press rights claims for instance.

Getting Global: Thinking Transnationally about Citizenship

In an increasingly transnational, global world, advancing our understanding of the political–economic and cultural forces shaping migrant rights has been an important aim of recent critical social geographies of citizenship. Leitner traces how Western European nation states have adjusted their responses to international migration pressures in accord with both changing economic needs (e.g., fluctu ating levels of need for foreign workers within domestic labor markets) and processes of cultural representation. She notes how migration pressures from Eastern Europe and the developing world have been portrayed as threatening national cultures and ways of life (e.g., as 'hordes' descending upon a nation state and society). Particular conceptions of what it means to belong to the nation are, she shows, at work at limiting who has access to citizenship and the degree to which certain rights such as voting can be claimed. Kofman has, similarly, stressed the ways in which citizenship in Western European cities is being extended to some and denied to migrant 'others'.

Focusing on the globalization of labor markets, specif ically on the export of labor from the Philippines and the import of labor to Japan, Ball and Piper have drawn at tention to how nation states at both ends of the migration chain are failing to protect the human rights of workers. On the one hand, the Philippine nation state has em braced neoliberal modes of regulation which devolve the responsibility for rights and welfare from the state to the individual workers and employers. On the other hand, and despite a structural dependence on migrant labor in some sectors of the labor market, the Japanese nation state has strictly enforced noncitizenship status and minimal labor rights for migrants working within its borders.

Mountz's analysis of state responses to human smug gling across Canadian borders provides another inter esting example of a critical social geographic account of the complex political–economic and cultural forces shaping the regulation of international migration and the rights afforded to those who enter a nation illegally. Focusing on a very controversial 1999 smuggled entry of migrants to the West coast of Canada, she argues for more embodied approaches to understanding how states respond to illegal migrants – which attend to how state power is enacted and experienced both by staff respon sible for assessing refugee claims and handling media reports (which in this case represented the state as 'losing control' over its borders) and by illegal migrants detained for assessment and possible deportation purposes. She notes the importance of examining cultural repre sentations of migrant bodies – in this case images of those bodies as threatening national borders, sovereignty, and existing immigration and refugee policies and pro cedures. Her work helps to remind us that state responses to challenges to existing ways of regulating access to citizenship and residency rights in particular places are more uneven, contradictory, and contested than unified views of state policies and procedures allow.

Looking Ahead: Toward New Geographies of Citizenship?

Over the past two decades, feminist and critical social geographers have increasingly turned their attention to processes of citizenship formation – processes that unfold in private and public spaces of life, and across personal, local, regional, transnational, and global scales. By at tending to complex processes of differencing, they have helped to keep questions of inclusion and exclusion, empowerment, and marginalization at the forefront of geographic analyses of citizenship. Further, by demon strating how socially and spatially uneven access to citizenship and human rights is they have challenged us to imagine futures in which all of our rights matter and are realized in practice.

What are some of the challenges that remain in ad vancing feminist and critical social geographic accounts of citizenship formation in society and space? One is expanding the range of processes of differencing con sidered in our explanations of the forces shaping relations and practices of citizenship in place. Processes of dis ablement, for example, play a critical role in disentitling people with mental and physical impairments to rights of citizenship in practice and yet disabling forms of citi zenship have received very little attention in the geo graphic literature. At a time of US led war in the Middle East and of racial profiling, surveillance and detention of suspected terrorists in the US and allied countries such as Canada and Britain, there is also a need to better understand the extent to which basic human rights are being extended to groups such as detainees. As Nagel and Staeheli's work on Arab American ways of asserting claims to citizenship in the United States suggests, there is also a need to better understand how particularly marginalized groups of citizens are struggling to assert the legitimacy of their citizenship claims.

There is also a pressing need to consider the kinds of citizenship practices that can promote a more just, sus tainable, and less imperialistic, militaristic world. On the latter, Mohanty challenges feminists, in particular, to grapple with how our teaching and scholarship may be complicit in the transnational US project of imperialistic, militaristic empire building. She argues:

> If the racialized, gendered, heterosexual figure of the citizen patriot, the risky immigrant, the sexualized and de masculinized external enemy and potential domestic terrorist are all narratives and state practices mobilized in the service of empire, an appropriate question to ask is whether and how the academy and the academic

> disciplines … are involved in contesting or buttressing these practices. (Mohanty, 2006)

She asks us, in short, to critically assess the projects of citizenship formation in which we are engaged and to develop radical practices of resistance which challenge oppressive relations and practices of empire.

Geographers have done much to transform our understanding of the processes through which the re lations, rights, and practices of citizenship are developed and contested in society and space. Will we also succeed in putting that knowledge into radical practices capable of forging a less divided, more inclusive, and just world? That is probably 'the' most important question facing all of us in the twenty first century.

See also: Activism; Belonging; Civil Society; Critical Geography; Democracy; Difference/Politics of Difference; Empowerment; Ethnicity and Resistance, Historical Geographies of; Governance; Governance, Good; Governance, Urban; Governmentality; Human Rights; Identity Politics; Imperialism, Cultural; Inequality; Liberalism; Nation; National Spatialities; Nationalism; Nationalism, Historical Geography of; Neocolonialism; Neoliberal Economic Strategies; Neoliberalism; Neoliberalism, Urban; Place, Politics of; Refugees and Displacement; Social Justice, Urban; Social Movements; State; Urban Policy; Welfare Reform.

Further Reading

Ball, R. and Nicola, P. (2002). Globalisation and regulation of citizenship Filipino migrant workers in Japan. *Political Geography* 21, 1013 1034.

Bell, D. (1995). Pleasure and danger: The paradoxical spaces of sexual citizenship. *Political Geography* 14(2), 139 153.

Blomley, N. and Geraldine, P. (2001). Canada and the political geographies of rights. *The Canadian Geographer* 45(1), 151 166.

Brown, M. (1997). *Replacing citizenship: AIDS activism and radical democracy*. New York and London: The Guilford Press.

Chouinard, V. (2001). Legal peripheries: Struggles over disabled Canadians places in society and space. *The Canadian Geographer* 45(1), 187 192.

Fenster, T. (1998). Ethnicity, citizenship, planning and gender: The case of ethiopian immigrant women in Israel. *Gender, Place and Culture* 5(2), 177 189.

Fyfe, N. R. (1995). Law and order policy and the spaces of citizenship in contemporary Britain. *Political Geography* 14(2), 177 189.

Kearns, A. (1995). Active citizenship and local governance: Political and geographical dimensions. *Political Geography* 14(2), 155 175.

Kofman, E. (2005). Citizenship, migration and the reassertion of national identity. *Citizenship Studies* 9(5), 439 451.

Leitner, H. (1995). International migration and the politics of admission and exclusion in postwar Europe. *Political Geography* 14(3), 259 278.

McEwan, C. (2005). New spaces of citizenship? Rethinking gendered participation and empowerment in South Africa. *Political Geography* 24, 969 991.

Mohanty, C. T. (2006). US empire and the project of women's studies: Stories of citizenship, complicity and dissent. *Gender, Place and Culture* 13(1), 7 20.

Mountz, A. (2004). Embodying the nation state: Canada's response to human smuggling. *Political Geography* 23, 323 345.

Nagel, C. R. and Lynn, A. S. (2005). 'We're just like the Irish': Narratives of assimilation, belonging and citizenship amongst Arab American activists. *Citizenship Studies* 9(5), 485 498.

Nelson, L. (2004). Topographies of citizenship: Purhepechan Mexican women claiming political subjectivities. *Gender, Place and Culture* 11(2), 163 187.

Relevant Websites

http://www.politicalgeography.org
Association of American Geographers, Political Geography Specialty Group.

http://www.emporia.edu/socsci/fembib/index.htm
Emporia State University.

http://econgeog.misc.hit u.ac.jp/icgg/
International Critical Geography Group.

Citizenship and Governmentality, Rural

L. Cheshire, The University of Queensland, Brisbane, QLD, Australia
M. Woods, Aberystwyth University, Aberystwyth, UK

Glossary

Active Citizens Individuals capable of regulating their own behavior according to prevailing standards of conduct.

Advanced Liberalism A political rationality that governs 'free' citizens through appeals to obligations to one's community.

Analytics of Government A study of government that shows how categories of state and society are brought into being through changing practices of rule.

Citizenship Commonly denotes membership to a national polity to which certain rights and responsibilities are attached.

Governmentality In a broad sense refers to the conduct of conduct. Applied more specifically to the governing activities of the state.

Rural Governance The study of the changing modes of governing in rural societies (usually in terms of the increased inclusion of rural citizens in the exercise of rule).

Introduction

In its most straightforward sense, citizenship is understood as bestowing upon individuals a legal status as members of a national polity, enacted by state apparatus and defended by legal and administrative rule and regulations. As members of that polity, individuals are entitled to certain political, civic, economic, and social rights, including individual freedom of speech, rights to property, the right to vote, and the right to live according to the prevailing standards of society – regardless of race, sex, or social class. At the same time, these rights are accompanied by a reciprocal set of obligations to the collectivity so that the rights of others can be secured.

It is more constructive, though, to consider citizenship not only in terms of legal status, but also as a device for understanding the relationship between the individual and the state in how society is governed. This relationship is not fixed and universal, but subject to change and contestation as particular definitions are enforced downwards by the state, and upwards by individuals and communities demanding an extension of their rights. Moreover, while in legal terms, citizenship is considered a universal entitlement for members of a nationality, experience shows us that citizenship rights have not been extended to all members of a society even when they hold formal citizenship status. For some social groups – women, indigenous people, and homosexuals – the granting of their citizenship rights has not occurred without significant struggle, but there are others too for whom the ability to participate as active and autonomous members of society is circumscribed by conditions of unemployment, poverty, and general social exclusion. Thus, we should see citizenship not as an automatic entitlement for all those bestowed with the legal status of belonging to a nation, but as something enacted, performed, and, increasingly, conditional upon the ability of individuals to conduct themselves in a way that is deemed morally and ethically responsible. As a concept, therefore, citizenship refers to a suite of rights and responsibilities that are negotiated between the state and the individual, but it is also a set of practices that constitute subject identities through material and institutional arrangements.

Such thinking is influenced by the governmentality writings of the French philosopher Michel Foucault, and his 'analytics of government'. From this perspective, entities such as state, society, and 'citizen' are social constructs developed under certain regimes of government that are historically specific. An analytics of government examines how these different regimes come into being, and the various govern mentalities that arise as those responsible for exercising rule pose questions about the appropriate domains of government (citizen, market, population, and community) and the means through which they might best be governed. What makes an analytics of government distinct from other approaches to the history of ideas, however, is that it moves beyond an examination of abstract political ideologies to examine the ways that certain categories of subject are constituted through the application of various techniques, instruments, and institutions of material intervention.

Since the mid 1990s, there has been considerable growth in the adoption of a governmentality perspective for exploring the changing form and function of the modern state and, by extension, the modern citizen. While there is nothing uniquely rural in these processes, there is a burgeoning body of literature on 'rural governance' that examines these changes in the context of rural societies of the advanced Western world. What follows is an overview of the construction of rural citizenship as an object of concern for political authorities, a brief outline of the key concepts and ideas that are common to governmentality analyses, and a discussion of the way this work has been applied to the field of rural studies.

Understanding Rural Citizenship

The concept of citizenship is not necessarily readily as sociated with rural societies. The term ‘citizen’ shares an etymological route with ‘city’ in the Latin word ‘civitas’ and, throughout history, citizenship has been primarily associated with urban societies, such as Greek and Roman city states and mediaeval boroughs. The coun tryside, in contrast, was associated with rigid social hierarchies supported by institutions such as the church and landed estates, and underpinned by particularistic relations between tenants and landowners. The appli cation of ideas of citizenship in the countryside can therefore arguably be considered as part of the urban ization of rural society.

The remaking of rural people as ‘citizens’ began with the rise of modern nation states from the late eighteenth century, and the promotion of ideas of national citizen ship. However, the rights and responsibilities of national citizens are not set, but open to political contestation and negotiation. In particular, the British sociologist T. H. Marshall suggested that states balance the rights of citi zens in varying degrees according to ideology and stage of historical development. Thus, in the twentieth century, broadly social democratic states in Europe, including Britain, emphasized social rights through the creation of welfare states. In a rural context, this meant that rural residents gained the same rights of access to public ser vices such as health and education as urban residents, despite the often higher costs of delivery; and the state took an interest in protecting farm populations from fluctuations in the agricultural economy through various price support schemes. Claims to similar social rights can be seen in rural politics and policies in North America and Australia, notably in campaigns for electrification and other infrastructure provision, and in some elements of agricultural support.

More recently, the ideological drive of state re structuring and the consequences of social and economic restructuring in rural areas have combined to reconfigure citizenship in rural society in a number of ways. First, as neoliberal states have retreated from their commitment to social rights, rural activists have mobilized to assert their ‘right’ to public services threatened by rational ization, or to demand state intervention to protect farms and rural communities from adverse economic pressures. At the same time, the inequalities of public service provision have been highlighted by in migrants who demand the same level of service provision in their new rural communities as previously experienced in towns and cities. Second, claims to ‘citizens’ rights have been used by external pressure groups to challenge some es tablished features of rural societies. These include the promotion of a right of public access to open countryside in Britain, which has challenged the historic privileges of landowners.

Third, some governments have sought to redefine the rights of citizens as akin to the rights of consumers, notably by introducing the right to choose into public service provision. Economies of scale have meant that this has been less successfully achieved in rural areas than in urban areas, creating a geographical disparity. The blurring of citizens’ rights and consumers’ rights has also been reflected in the assertion of consumer interests in food, agricultural, and environmental policy networks. Finally, the neoliberal state has also re imagined the balance between citizen rights and citizen responsi bilities. While the focus of citizenship has traditionally been on the rights it embodies, the last two decades have seen a renewed emphasis on the responsibilities that citizenship also implies. The concept of active citizenship has become engrained in rural policy across a range of domains, including rural development, conservation, re source management, public service provision and hous ing, and can be positioned as part of shift in the mode of governmentality, as discussed further in the remainder of this article.

Governmentality

Throughout the course of his work, Foucault came to use the term government in various ways. In his earlier writings, he defined it in broad terms as the conduct of conduct: a set of actions upon actions to shape, guide, or change the course of behavior. In later work, however, government came to refer more specifically to the exer cise of political power by the modern state, as outlined in his famous essay on ‘governmentality’. In this and sub sequent accounts of his work, two key themes on gov ernmentality can be discerned. The first relates to what he called a ‘problematics of government’ whereby those seeking to govern pose questions about how this might best be achieved – who should be governed, by whom, to what ends and by what means? Consistent with Foucault’s interest in government as practical activity, these ques tions were both discursive and technical in that they constituted reality in particular ways that made it amenable to intervention via the application of a broad range of practical techniques and procedures.

The second area of interest in the governmentality literature refers to the emergence, from the seventeenth century onwards, of a novel form of political power known as biopower. Biopower was distinct from sovereign forms of power in that it demarcated a new sphere of interest for political authorities – known as the population – and a new area of concern regarding the health and welfare of that population. This evolved in two basic forms. First, a disciplinary power, which targets itself at specific individuals, seeking to create mute and docile bodies through the application of certain disciplinary techniques of training and surveillance. The second, a biopolitics of the population, seeks to manage the population as a whole through techniques of normalization. In order for these new forms of governing to take place, the life of this population had to be made 'knowable' and measureable. Thus, an explosion of discourses about the problems of public health, housing, sexual activity and so forth occurred, accompanied by new techniques of calculation, audit, and inscription that brought the formerly hidden life of the population into the domain of state activity.

Liberalism

Contemporary governmentality theorists such as Nikolas Rose have extended Foucault's ideas on biopower to link its emergence with the rise of a new liberal regime of government, which sets limits on the extent to which the state can intervene in the workings of the market and civil society. As a 'practice' of government, liberalism involves the development of new techniques for governing which enable political authorities to govern people, events and places without destroying their freedom. It seeks to achieve this by modeling its own techniques of governing upon the self[-]regulating mechanisms that occur naturally within civil society. These techniques, which are readily apparent today, include exposing state administration and private individuals to the self[-]regulating principles of market competition; nurturing the norms and values that are believed to exist naturally in the sphere of civil society; and implicating private actors and nongovernmental agencies in the activity of governing. Central to all these techniques are practices of self[-]government, with liberal forms of rule relying upon the construction of a new kind of active citizen who is imbued with the skills and sensibilities of rational self[-]government, thereby rendering the direct imposition of state will unnecessarily.

Governing through the Social

During the course of the nineteenth and twentieth centuries, governments of advanced Western societies have experimented with various forms of liberalism in pursuing more effective means to achieve their ambitions for a well[-]ordered society. Two such models have been discerned over the last century, each conceiving the various domains of governmental concern in quite different terms, and each embodying a changed relationship and corresponding demarcation of responsibilities, between the state and citizen. The first is a 'managed liberalism', which manifested itself most clearly in the Keynesian welfare state of the mid[-]twentieth century. Where classic liberalism governed through the rational choices of autonomous individuals, welfarism demarcated a new object of government interest – known as the 'social' – which acted upon the well[-]being of individuals in the name of society. For social forms of government to work, it was necessary to nurture the collective ties of solidarity that linked individuals to wider society and to infuse those individuals with the norms of social obligation. In return, they were accorded equal rights to political representation and economic and social security via a national welfare state.

Advanced Liberalism

Empirical observation shows us that this form of government is no longer dominant. In its place is what has been referred to as an 'advanced liberal' rationality of rule, or a post[-]welfarist regime of the social, which contains elements of both classic liberal and social governmentalities. As with its earlier manifestation, advanced liberalism reconfigures the subjects of rule as autonomous and free actors, responsible for guiding their own conduct, but also as members of subnational communities and networks rather than as members of society as a whole. This suggests there has been a rescaling of citizenship, with allegiance to others enacted in the context of localized family and community obligations. Additionally, however, there is an increased emphasis on the performative aspect of citizenship, with rights increasingly contingent upon individuals demonstrating a willingness to accept personal responsibility for their own future. Here, it is possible to see a melding of the two poles of liberal and social citizenship. The first pole – what may be termed economic citizenship – relates to the required conduct of citizens in the sphere of the market as responsible and informed consumers, capable of managing their lives in an entrepreneurial and prudential manner. The second pole revolves around an ethic of citizenship that is more community based and invokes ties of obligations to others.

Governing Rural Citizens

The adoption of a governmentality perspective has done much to shed light on the changing practices of rural citizenship, albeit under the much broader heading of rural governance. Since the mid 1990s, governmentality analyses have become increasingly popular in rural studies as researchers have sought to understand the changing relationships between rural areas and the state, and the discursive practices through which the various domains of farmer, rural citizen, and community are demarcated as objects of government concern. In many parts of the advanced Western world, significant changes have occurred in the way rural areas are governed, underpinned by a broader shift from a welfarist to an advanced liberal rationality of rule. As evidence of this trend, researchers point to the dismantling of many of the social and economic forms of protection that were tra ditionally afforded to rural areas, and their replacement with new policy discourses centered around the promo tion of self help and self reliance in farm and natural resource conservation and management, rural service provision, and community renewal. For many observers, such moves are symptomatic of attempts by governments to divest themselves of their responsibility to rural areas, particularly at a time when so many are experiencing economic and social decline. Yet, we are reminded that advanced liberalism constitutes a new mode of rule, as opposed to an absence of rule, and that the enrolment of rural actors into the domain of governing has been ac companied by new governmental technologies that re configure rural citizenship into something that is active. This has also entailed the formation of a new category of citizen whose conduct does not accord neatly with pre vailing standards of responsible conduct.

In one of the earliest papers to examine state inter vention in rural matters using a governmentality per spective, Murdoch and Ward present an historical analysis of the practices through which British agri culture came to be established as a sector of interest to the state. Through much of the nineteenth and twentieth centuries, this interest in farm matters was enacted through the lens of a national rural policy framework organized around the domestic production of food and/ or minerals, or in the case of the Antipodes, for export back to the home country. The prevailing rationality of this social form of government was the importance of farming to the national economy, and a commitment by the state to ensure its prosperity was assured. In ac cordance with these goals, a whole regime of material and institutional farm support structures was established, which collectivized farm risk in the name of the social. This protectionist agenda also extended to rural areas more broadly, with rural citizens conceived as having equal rights to the same level of services and infra structure as their urban counterparts, which the state was prepared to meet. In the pioneering nations of Australia, the US, and Canada, governments were even more am bitious in their interventionist strategies as attested by the establishment of closer settlement and other regional development programs, such as the Tennesee Valley Authority in the US, and the new inland towns of Albury Wodondga in Australia.

In line with recent moves toward advanced liberalism, governments across the Western world have begun dis mantling the protectionist regimes of the last century amid criticisms that they have fostered a handout men tality in rural communities and perpetuated inefficiencies in agriculture. Compounded by economic rationalist ar guments that subsidized rural service provision is no longer cost effective, the result has been a withdrawal of state support in rural areas as rural people are en couraged to play a more active role in providing those services for themselves. To a large extent, this new agenda of rural active citizenship has been driven by political authorities and embodied in a range of new institutional and funding arrangements that reconstruct rural people as partners of the state rather than as passive recipients of welfare. Yet, at another level it has also been embraced by rural people who have responded to re duced state support by mobilizing into local action groups and formulating their own, sometimes novel, so lutions to the problems they face.

The reconstruction of citizen identity around the two poles of market and community is discernable within the rural sphere. With regard to the former, this is most apparent in the realm of farming where collective problems for the state have been reconstituted as prob lems for individual farmers. As Higgins has revealed in his study of Australian agriculture, citizenship can no longer be viewed in terms of social obligations, but be comes individualized and contingent upon the self gov erning capacity of farmers as manifested through their farm management skills. Moreover, in order for state agencies to assess these capacities, they must first be brought into the realm of the knowable through various techniques of calculation and farm audit and, for those considered lacking such qualities, enhanced through the application of farm business training and counselling schemes. If these strategies fail to reconstruct unviable farmers into rational business managers, the only kind of state assistance likely to be forthcoming are farm exit payments designed to facilitate their adjustment out of the industry.

Attempts to harness the self governing properties of individuals to economic imperatives can also be observed in governmental approaches to rural and regional de velopment, which are imbued with a similar discourse of increased self reliance and personal responsibility. For agencies such as the Organisation for Economic Co operation and Development (OECD), the problems fa cing declining rural areas arise from their lack of integration into the global economy and are rooted in the supposed deficiencies of rural people themselves: they are unskilled or have the wrong skills, they lack a self reliant mindset and, perhaps understandably, they have been rendered powerless and demoralized by their experiences of rural industry and community decline. With the causes of rural decline constructed in this way, the solution is seen to lie in the provision of 'capacity building' that directly works on the attitudes and values of rural people to build their self esteem, imbue them with a spirit of entrepreneurship and reconstitute them as self responsible and active citizens. While capacity building has been touted as a source of empowerment for rural people by providing them with the skills to articulate their own needs, it has also been seen to operate as a vehicle through which rural development initiatives can be governed according to state ambitions for a globally competitive rural economy. This is exemplified by Cheshire's study of the *Building Rural Leaders Program*: a training course in Australia designed to assist rural people in building their leadership and rural development skills. Drawing on Foucault's writings on disciplinary power, she suggests that Building Rural Leaders seeks to create economically productive but politically 'docile' rural leaders through its efforts to enhance their entrepreneurial disposition while simultaneously directing those efforts toward embracing, rather than resisting, the changes they face.

While market rationalities are often viewed as contradictory to the sentiments of community, the latter remains a central pole of action within advanced liberalism and often complements economic discourses of citizenship in unexpected ways. This has been referred to as 'governing through community', and plays out in rural areas in a number of ways. First, in the way community is deployed as a means of imbuing individuals with a collective, rather than an individualized conscience, thereby encouraging enterprising citizens to engage in rural development or natural resource management activities that benefit the rural community as a whole. Often referred to as 'social' or 'community' entrepreneurs, they are the consummate active citizen: imbued with all the enterprising qualities of rational self management but community minded enough that they voluntarily elect to channel their skills into community, rather than individualistic, endeavors. Second, governing through community discursively reconstructs rural issues as local community, rather than state concerns. In the realm of natural resource management, the advent of community based natural resource management schemes such as the Australian National Landcare Program demonstrates how responsibility for managing the problem of environmental degradation has been diffused throughout the rural community, with rural people now viewed as 'partners' of the state. Once again, though, research has suggested that such partnerships constitute a form of 'government at a distance', with state agencies continuing to exert considerable power in determining the type of action taken by local community groups. In the case of Landcare, it has been suggested by Lockie and others that the discourse and funding of Landcare works to deal with the negative environmental externalities of an increasingly intensive agricultural imperative, rather than encourage any profound reassessment of current farming practices to take place.

Third, in policy discourses, the economic and social well being of rural areas is frequently understood in terms of the presence or absence of a community culture. So called 'can do communities', which take active responsibility for their own well being, are commonly viewed as possessing a positive outlook in life and a consensual approach to solving problems. Conversely, struggling communities are viewed as an aberration to this and require the application of remedial or expert intervention to re establish the community culture that has been lost. In practice, this deployment of community frequently begins with the administration of a community audit or skills workshop in which local residents are encouraged to identify their own problems and collectively develop appropriate solutions using local skills and resources. While practices of community development and community capacity building are not unique to rural areas, the strong historical association between rurality and community serves to bolster the claim, not only that rural people should be expected to help themselves, but also that it is part of their culture to do so, and their right to expect this. Finally, it is argued that such discourses work to normalize particular ways for rural communities to think and behave, thereby creating a distinction between the civilized members of society who understand their obligations to self and others, and those who have failed to exercise their citizenship in a responsible manner. Those seen to lie beyond this normative framework are often considered beyond help and – in a policy climate of 'picking winners' for marketable rural development opportunities – government assistance often passes them by.

Finally, it is important to note that the reordering of rural citizenship has not been restricted solely to those who live in rural areas, but equally to the conduct of all citizens in how they relate to the countryside. Codes of conduct established specifically for nonrural dwellers visiting rural regions exhort a similar ethos of good (countryside) citizenship based on commonsense notions of responsible behavior and respect for other users. Recent work by Parker examines the various practices of self government that underpin the British Country code, which not only sets out a code of conduct for visitors to the British countryside but, importantly constitutes a new category of country citizenship: that of the nonrural dweller. With parts of the contemporary countryside increasingly becoming sites of consumption by nonrural citizens, the governance of these new populations, and the potential conflicts over land use and values, will become matters of increasing concern requiring new, unprecedented forms of intervention.

Summary

Studies of contemporary citizenship have advanced significantly in recent years as a result of the adoption of a governmentality lens through which the fundamental categories of state, citizen, and society, and the relationship between them, can be problematized. In the field of rural research, this approach has been adopted predominantly in the context of the changing governmentalities of the state and the emergence of an advanced liberal mode of rule which seeks to govern through the rational market orientated conduct of the enterprising farmer and/or the redefinition of rural dwellers as ethically informed community members. Empirically, this has led to a proliferation of studies examining the mentalities of rule as articulated in rural policy documents and – in fewer cases – exploring how they become embodied in material interventions such as rural leadership, community capacity building, and farm financial counselling programs. What has been missing, until recently, is a greater consideration of the way that rural people engage with these discourses of active citizenship – not merely as objects of political power, but as potential sources of resistance or contestation through which challenges to the new policy frameworks might occur. If the inclusion of rural people into the networks of the state has enabled political authorities to govern their conduct more directly, then it has also created a space for a much closer engagement of rural people with government policy. Indeed, more recent research has also suggested that the mobilization of rural people around protest activities can also be perceived through the lens of citizenship and understood as attempts by rural people to reassert themselves – as citizens – in response to the perceived erosion of their rights. What this new research reminds us is that citizenship remains a contested field – one that required a more detailed and ongoing analysis.

See also: Citizenship; Foucauldianism; Governance; Governance, Urban; Governmentality; Liberalism; Protest, Rural; State.

Further Reading

Cheshire, L. (2006). *Governing Rural Development: Discourses and Practices of Self Help in Australian Rural Policy*. Aldershot: Ashgate.
Dean, M. (1999). *Governmentality: Power and Rule in Modern Society*. London: Sage.
Dean, M. (2002). Liberal government and authoritarianism. *Economy and Society* 31, 37 61.
Foucault, M. (1991). Governmentality. In Burchell, G., Gordon, C. & Miller, P. (eds.) *The Foucault Effect: Studies in Governmentality*, pp 87 104. Hemel Hempstead: Harvester Wheatsheaf.
Higgins, V. (2002). Self reliant citizens and targeted populations: The case of Australian agriculture in the 1990s. *Arena* 19, 161 177.
Isin, E. and Wood, P. (1999). *Citizenship and Identity*. London: Sage.
Lockie, S. (2000). Environmental governance and legitimation: State community interactions and agricultural land degradation in Australia. *Capitalism, Nature, Socialism* 11(2), 41 58.
Murdoch, J. (1997). The shifting territory of government: Some insights from the Rural White Paper. *Area* 29, 109 118.
Murdoch, J. and Ward, N. (1997). Governmentality and territoriality: The statistical manufacture of Britain's 'national farm'. *Political Geography* 16, 307 324.
Parker, G. (2006). The country code and the ordering of countryside citizenship. *Journal of Rural Studies* 22(1), 1 16.
Rose, N. (1996). The death of the social? Refiguring the territory of government. *Economy and Society* 25, 327 356.
Rose, N. (2000). Community, citizenship and the third way. *American Behavioral Scientist* 43, 1395 1411.
Woods, M. (2006). Political articulation: The modalities of new critical politics of rural citizenship. In Cloke, P., Marsden, T. & Mooney, P. H. (eds.) *Handbook of Rural Studies*, pp 457 471. London: Sage.

Relevant Websites

http://www.countrysideaccess.gov.uk/things to know/countryside code
The Countryside Code.
http://www.daff.gov.au/natural resources/landcare/national landcare programme
The National Landcare Program.

City Marketing

E. J. McCann, Simon Fraser University, Burnaby, BC, Canada

Glossary

Broken Windows Strategy A popular yet controversial philosophy of policing that stresses the rapid response to minor crimes and blemishes on the built environment (e.g., vandalism and broken windows). It suggests that a quick response will discourage both minor and also major crimes.

Myths Narratives about the world that are not necessarily untrue. More importantly, they resonate with their audience and provide an agenda and legitimacy for action.

Revanchism A term generally referring to vengeful policies and used in the contemporary urban context to highlight the negative effects of zero tolerance policing and related policies on the poor.

Introduction

Classic city slogans like "I ♥ New York" and "Glasgow's Miles Better" are so embedded in the popular consciousness as to have long outlasted their official use by their respective municipalities, chambers of commerce, and convention and visitors bureaus. Cities and towns are increasingly associated with symbols, slogans, or taglines ("What Happens Here Stays Here. Las Vegas" or "Johannesburg. A World Class African City," for example). These slogans are the most widely recognizable elements of an interconnected group of contemporary strategies know by a number of names including, urban entrepreneurialism, civic boosterism, urban branding, and city marketing. They are intended to enhance a city's competitive advantage over others in the context of zero sum competition for resources, investment, jobs, and tourist revenue. While the ubiquity and persuasiveness of city marketing strategies often makes them appear as natural and normal activities of city governments and business organizations, geographers and others have shown that city marketing is a social process with a history, a geography, and a political economy. Through the empirical study of city marketing and through efforts to conceptualize the contemporary process in terms of its political, economic, and social contexts, geographers have identified how the character of city marketing has changed through time, how it draws on certain aspects of the built environment, and why and in whose interests city marketing has developed into its present form.

These themes highlight a number of myths, tactics, and intended audiences that characterize most city marketing strategies. Key myths that permeate city marketing schemes include: the city as welcoming and safe, vibrant and fun, tolerant and accepting of social and cultural difference, environmentally friendly, culturally rich, business friendly, and as strategically and conveniently located. Favorite tactics of city marketers, beyond sloganeering, include: the provision of packages of business incentives, the (re)building, policing, and cleaning of the urban built environment to keep its appearance in line with the city's marketing image, and continual efforts to maintain coherence in the city's marketing message by keeping disparate interest groups either 'on message' or out of the spotlight. Finally, many city marketing strategies are characterized by their attention to a range of audiences or niche markets: potential business investors, potential residents and workers, tourists and conventioneers, the promoters of mega events like the Olympics, grant giving agencies, and local residents. In this regard, city marketing is not only strategic; it is also powerful and political. It involves struggles by different groups seeking to shape cities in ways that satisfy their interests.

City Marketing in Historical–Geographical Context

City marketing did not emerge fully formed from the heart of the Big Apple, nor was it a child of Mr. Happy, Glasgow's cartoon symbol. Rather, the legendary New York and Glasgow campaigns of the 1970s and 1980s marked only the beginning of the contemporary phase of city marketing. While contemporary city marketing has important distinguishing contexts, features, and impacts, it is worth pointing briefly to earlier examples of the phenomenon. Place marketing has occurred in various forms since at least the middle of the nineteenth century. In the United States at that time, the development and sale of railroad towns on the Anglo frontier was facilitated by advertising in the East and elsewhere. In the first half of the twentieth century, resorts and newly built suburbs became major foci of place marketing efforts in Britain and the US as mass tourism burgeoned. In these and other contexts marketing was driven largely by private actors.

By the last third of the twentieth century, a series of political–economic restructurings set the context for the contemporary era of city marketing. Geographers generally regard the economic crises of the early 1970s as marking a series of political and economic shifts that fundamentally changed the character of cities and urban governance. This was a time, in the Western world, of de industrialization, high rates of unemployment, and a growing turn to fiscal austerity at all levels of the state. These conditions paralleled and precipitated the growing popularity of a neoliberal ideology that advocated the benefits of a market logic and for the 'rollback' of the state through, for example, the privatization of nation alized industries. In the many industrialized cities that had enjoyed economic prosperity in the post war period, local governments and chambers of commerce had largely managed the economy and social reproduction but had not engaged in the sort of boosterism charac teristic of private actors in earlier years. With the drying up of tax revenues and flows of funding from other levels of government, these actors found themselves compelled to act entrepreneurially in order to gain a competitive advantage as they sought to fix flows of resources, jobs, and capital in their city rather than see them stream into another.

Placing City Marketing: Geographical Approaches

Urban entrepreneurialism, then, sets a context for understanding contemporary city marketing, as do the activities of coalitions of political and economic elites.

Urban Entrepreneurialism, Growth Coalitions, and City Marketing

Urban geographers' conceptualization of contemporary city marketing is heavily influenced by David Harvey's classic article, 'From managerialism to entrepreneurial ism'. Harvey makes three related claims. First, that since the early 1970s, city governments have had to take an entrepreneurial stance in relation to capital as a result of the changes described above. Second, this shift has en tailed a relocation of control over, and risks emanating from urban economic development. Control of local governance is increasingly vested in private and un elected institutions. Local state agencies are, thus, left playing a facilitative role in urban policy making while assuming more of the risks of development. It is the re location of risk that Harvey sees as distinguishing con temporary entrepreneurial city marketing from earlier forms of civic boosterism. His third argument is that the entrepreneurial impulse has encouraged an initiatory or speculative ethos in entrepreneurial cities that leads coalitions of private and public actors to focus more re sources and attention on the construction of spectacular sites or built environments that will, it is hoped, prime the pump for further private investment. In this case, the geographical lens at work is one focusing on specific scales of action (such as the global, national, and local) and on the ways in which processes generally seen to be operating at one scale are, in fact, thoroughly imbricated with the others. This perspective provides a great deal of insight into the structural conditions for and practices that constitute city marketing.

Who are the actors involved in contemporary city marketing and what are their key interests? Entre preneurial urbanism has not sidelined the types of busi ness interests involved in nineteenth century marketing. In fact, these locally dependent place entrepreneurs, or rentiers, are crucial to the process because their eco nomic interests are based in drawing rent from land and buildings in a specific city. Therefore, it is generally in their interests to advocate for policies that will increase land values and intensify land uses in their locality. In order to achieve this goal, rentiers (i.e., developers, construction companies, property financiers, the real estate industry, etc.) must bring local politicians – a group that is as much, or more, locally dependent than they are – into a growth oriented coalition. The large scale political–economic conditions that Harvey identi fies in the post 1970 period make local policy makers increasingly likely to engage wholeheartedly in such a coalition, offering as it does an opportunity to increase tax revenues. These actors join with other institutions and groups who view their interests – economic or otherwise – as lying with local economic growth. The other institutions include local media, utilities, local arts and entertainment organizations, and universities. As members of a private–public growth coalition, locally dependent actors propagate the myths and engage in the tactics of city marketing in order to attract and fix mobile capital in their place.

The Role of the Urban Built Environment in City Marketing

Geographers have also employed a second, not unrelated lens to the study of city marketing. It involves a concern with the mutually constitutive relationship between the built environment and social action. This approach seeks to understand the 'rewriting' of the city through mar keting strategies in (1) textual and representational terms and (2) a physical sense.

City marketing frequently entails the writing and re writing of a particular 'script' or story about a city, par ticularly – but not only – for cities that are attempting to rebound from economic crisis and decline. There is an intent by the actors described above to present what they see as the most attractive essence of their city to the world via an assemblage of cut and pasted tangible and intangible elements of its landscape, economy, society, and culture. This brand is not only captured in a slogan and a logo but also involves the careful managing and presentation of specific images of the city through city websites, marketing brochures, and, where possible, via stage managed media events. The recent millennium celebration, for example, was used by cities across the globe to present their best and most spectacular face on live television by carefully staging their midnight fireworks celebrations against backdrops of their most picturesque or significant landscapes. Paris used the Eiffel Tower, Mumbai employed the Gateway of India, Moscow's fireworks exploded above St. Basil's Cathedral and the Kremlin, Rio de Janiero's lit up Copacabana Beach, and so on. Yet, the use of the urban built environment in urban image making is hardly a once in a millennium endeavor. Cities like Sydney, which have successfully competed for mega events like the Olympics, offer their iconic landscapes as stage sets for global media consumption during those events. More prosaic still, cities such as Barcelona use elements of their built environments, like unique styles of architecture, to convey specific meanings to particular audiences.

While the urban built environment can be used as part of a carefully constructed writing and rewriting of a city, geographers also note the ways in which the writing of cities for marketing purposes can provoke the reshaping, policing, and managing of its physical spaces. This form of physical reworking involves: (1) the building or renovation of elements of the built environment to enhance or correspond to a marketing goal and (2) the policing, cleaning, and managing of urban public spaces with the express intention of making them correspond to a carefully constructed brand and, thus, to make them feel safe and welcoming for those who the growth coalition sees as its most valuable clients. Examples of the first strategy abound. The most renowned ones in the literature are Baltimore's Harborplace development and the surrounding Inner Harbor and Boston's Faneuil Hall Marketplace but examples of these 'festival spaces' intended to contribute to revamped city images can be found across the world. A recent addition has been the Guggenheim Museum in Bilbao, which has been credited with markedly enhancing that city's global image and tourist economy. Indeed, the phenomenon of urban growth coalitions copying successful development models from elsewhere led Harvey, in his classic article, to identify a problematic 'serial reproduction' of the sameness across urban landscapes which, by definition, crowds the marketplace and dilutes the competitive advantage that the festive space, aquarium, or sports stadium provides to a city.

Every brand – its slogan, its logo, its iconic landscape elements – has a shelf life and the expiration dates seem to arrive all the more quickly when marketplace crowding compels other growth coalitions to innovate and find new marketing styles and strategies. The production of an urban brand is, then, an ongoing process, not a one off event. Not only does it entail careful monitoring of the representational strategy (including which elements of the built environment to foreground) but it also involves the constant policing and management of the city itself, so that its public spaces – and even its people – or at least those who are on public view in spaces likely to be traversed by tourists or business people – correspond to and enhance the brand. A number of related tactics are involved in this element of a branding strategy. At a general level, the city is frequently divided into 'frontstage' and 'backstage' areas. The former are tourist and/or business and shopping zones like Central Glasgow. The latter are areas seen as dangerous, unpalatable, or of little interest to highly valued visitors and residents. More or less subtle representational strategies can be used to channel valued populations in certain directions including, for example, the careful construction of tourist maps to focus the tourist gaze on specific sites/sights.

This channeling of people through the urban built environment is now frequently the purview of employees of Business Improvement Associations, Business Improvement Districts (BIDs), or their equivalents. These are private organizations, funded by special localized levies and taxes, that have, under conditions of neoliberal entrepreneurialism, assumed the formerly public functions of managing clearly delimited areas of the city (frequently the downtown). These private governance agencies not only offer tourists advice, but also frequently provide street sweeping services in their area and contract with private security firms to enforce bylaws and guard against so called 'quality of life' crimes (e.g., panhandling, public intoxication, graffiti tagging, and prostitution). BIDs, then, are inspired by and seek to reinforce some of the central myths of city marketing – the city as safe and friendly, fun, and without obstacles to commerce.

The Sharp Edge of the Hard Sell: Managing and Policing the Branded City

Many have noted that the attempts by tourist agencies, chambers of commerce, and BIDs to mange the urban environment in correspondence with a particular marketing strategy are often armored with a hard edge of high technology and 'zero tolerance' policing strategies. These strategies frequently involve the blanketing of city neighborhoods with closed circuit television (CCTV) surveillance, the reintroduction of neighborhood police stations, and the strict enforcement of laws for even the most minor infractions, under the 'broken windows' model of policing. While many applaud the 'cleaning up' of cities where these tactics are deployed, others argue that this sharp edge of image conscious urban governance marks a new high (or low) point in a mean spirited, 'revanchist' war against the poor, and particularly the homeless, in contemporary cities.

Many of these strategies and the politics around them are epitomized by Times Square in New York. In the 1990s, under the guidance of Mayor Giuliani and a pri vate–public development partnership, this neighborhood which was renowned as a 'red light' district but was also regarded by many as the vibrant heart of the city, was transformed into what proponents see as the welcoming face of the new tourist and business friendly Manhattan and skeptics regard as a bland, and literally Disneyfied space of consumption (featuring a Disney store, Disney theaters, and a studio of Disney owned ABC Television). The square has its own BID, whose street cleaners patrol constantly, a police substation, and is in Midtown Man hattan, a neighborhood where almost every square foot of public space is covered by at least one CCTV camera. Of course, the rewritten square is also the site of the annual, internationally televised street party to welcome the New Year – an event staged to convey the city's image as a friendly, safe, vibrant and fun, if sometimes chilly, place.

Know Your Audience: Targets of City Marketing

It is already evident from the discussion above that city marketers tend to have more than one audience in mind when they create a brand. Clearly tourists, potential residents, and business investors are in the sights of most urban growth coalitions when they charge marketing consultants with enhancing their city's image. It is worth elaborating on the range of audiences – and the audi ence specific tactics – that have been identified in studies of city marketing, however.

Potential business investors are enticed with talk of a generally 'business friendly' environment, tax incentives, a willing and well trained workforce, ongoing economic growth, an already existing cluster of similar businesses and suppliers, help with customized facilities, proximity or easy transportation to markets, and high quality of life for CEOs and valued workers. Potential residents, who in many but not all cases would be workers in the city, are enticed with the promise of jobs and abundant career opportunities, a low cost of living, and high quality of life, encompassing everything from proximity to attractive natural environments and recreational opportunities ad jacent to the city, to bike paths, high quality recreational facilities and spectator sports, good schools and colleges, abundant 'high' and 'street' culture, and so on. The combination of economic development and quality of life discourses has certainly been at the heart of many city marketing campaigns in the last 15 years. Marketers seek to present their city as an economic boomtown and, simultaneously, an ideal hometown. Among potential residents and workers who have become particularly coveted since the turn of the millennium in North America has been the so called creative class of young workers employed in the technology and design indus tries. Cities pursuing this market niche have sought to highlight their vibrant neighborhoods, coffee shops, ac tive (rather than spectator) recreational opportunities, diversity, cosmopolitanism, and 'cool'. Urban elites who market to tourists and conventioneers are, for their part, likely to touch on some of the same quality of life and 'quality of place' images outlined above but are more likely to foreground the more immediate, visceral, and ephemeral experiences available in the city – from art galleries to night life. Yet, the tourist market is highly segmented and city marketers are careful to tend to a variety of niches within it.

There are other important audiences for city mar keters beyond the 'big three'. For example, certain cities spend a great deal of resources on marketing themselves to the organizers of spectacular mega events, such as the Olympic Games, the soccer World Cup, or to those, like the officials of the European Commission, who can be stow titles like 'Capital of Culture'. These bids take a great deal of effort from a range of institutions and vol unteers and involve often quite detailed marketing strategies. The time, care, and money put into these bids indicates the substantial nature of expected returns and legacies from events like the Olympics, particularly as a result of the expected global media coverage. Clearly, various print and electronic media are used by city marketers, yet the marketers do not necessarily always control the ways their city will be portrayed. Thus, the media become not only an agent of city marketing, but also a target audience as city marketers seek to persuade them to present their city favorably in feature stories, travel articles, and the increasingly ubiquitous rankings of the best cites for business, quality of life, young people, etc.

Each of the audiences discussed here requires a dif ferent set of marketing tactics. Two more audiences, which also necessitate specific marketing styles, are worth noting. First, while cities' relationships with funding streams from other levels of government have changed since the 1970s, this is not to say that central government, for example, is no longer important as a source of rev enues. In some cases, like the UK, there is evidence of a focus by local elites on competitive bidding for grants from central government. Therefore, urban governments, private–public partnerships, etc. must present themselves in a particular way through the grant proposal process. Second, it is worth noting that city marketing is not al ways externally focused. Marketing often entails a focus on the population of the city itself. This is frequently seen as part of an effort to legitimize the policies and objectives of governing elites, to emphasize that entrepreneurial growth strategies are ‘value free’ (i.e., they benefit everyone in the city) or, in the case of the “We are LondONErs” campaign, to shape a particular identity among the population. London’s campaign was developed in response to the bombings of 7 July 2005 and is intended to portray the city as multicultural but united.

Staying on Message: The Politics of City Marketing

As the London campaign suggests, city marketing has a number of political motivations and implications. The importance marketers place on creating and managing a coherent, singular brand for a city necessitates political activity intended to persuade a wide variety of interests within a broad growth coalition to get on and stay on ‘the same page’. Growth coalitions comprise interests from a range of economic sectors, as well as state actors and representatives of other institutions, such as universities. Therefore, there can be different foci within a coalition. This ‘silo’ effect can be exacerbated in larger cities – or in urban regions where numerous cities have combined their resources to construct a single marketing strategy – since even interests from the same sector might find themselves at odds over development strategy due to their embeddedness in certain neighborhoods or municipalities. Thus, a continual politics of consensus building and persuasion within the coalition is necessary if a single brand is to be articulated and maintained. While this pluralistic political context does produce successful long term agreements on coherent marketing strategies, other examples such as that of the Ontario cities that brand themselves as ‘Canada’s Technology Triangle’ have highlighted the difficulties in maintaining coherence under a single marketing banner.

The politics of persuasion and consensus building operates not only within growth coalitions but also between the coalition and the larger population of a city. In order to present a unified and welcoming marketing image, marketers must find ways to dampen opposition to their particular branding and development vision. Thus, it is necessary to build what Cox and Mair call a ‘localist ideology’ where most residents of a city are persuaded that they are worthy, deserving, and unified in relationship to the ‘outside world’. Not surprisingly this sort of hegemonic discourse is never complete or entirely convincing. Not all residents will be persuaded by a marketing strategy or a particular development agenda. There will be protests, as there so often are before and during Olympic Games, as activists for groups whose interests and identities are elided by the marketing and development strategy coalesce around their own counter slogans such as “Healthcare Before Olympics,” and “Homes Before Games” (two oppositional slogans referring to Vancouver’s successful bid for the 2010 Winter Olympics). The failure of the politics of persuasion and the rise of dissent within or without a growth coalition can, in some cases, be quite damaging to a brand and can provoke elites to try even harder to put what they see as a city’s best face forward.

Future Directions

The study of city marketing has been a vibrant element of urban geographical scholarship since the beginning of the 1990s. Yet, there are aspects of the phenomena that have been less covered than others. Furthermore, there is a danger of the literature becoming rather uninspiring. In the latter regard, it seems that with a few exceptions, most approaches to city marketing come from the political economy perspective outlined above. While that is a powerful and valuable approach that is worth continuing to pursue, it is one that seems well worked out, conceptually. Therefore, there is a danger that new studies will, ironically, engage in a ‘serial reproduction’ of their own – one producing new case studies of slogans and strategies without furthering conceptual understandings. On the other hand, there are signs that the recent turn toward questions of affect in critical human geography and the relatively well established interest among geographers in psychoanalysis might lead to new insight into the ways in which brands, symbols, and slogans move their audiences on a variety of levels.

In the former regard, it seems that while some work is being done in the following three areas, more is necessary. First, the increasingly common way in which urban policy professionals, such as city planners, market their ideas, models, and achievements to their colleagues elsewhere through professional networks and organizations, international ‘best practice’ clearinghouses, and through the Internet is a very specialized type of city marketing, but one that has an effect on cities and one that can both inform and be informed by burgeoning geographical literatures on policy transfer and mobilities. Second, more detailed analyses and evaluations of the effects of marketing strategies are necessary. Can campaigns intended to attract new workers or investment be measured in order to evaluate their direct and indirect consequences? Third, as city marketing becomes more common beyond the Western industrialized world, it seems reasonable to assume that analyses conducted in a variety of different historical, cultural, and political–economic contexts would benefit from and lead to theorizations of postcoloniality and international development. Are standard theoretical perspectives tenable, for instance, if they are employed to study city marketing in socialist, post socialist, post apartheid, or developmentalist states? These are open and interesting questions.

See also: Governance, Urban; Neoliberalism, Urban; Tourism, Urban; Urban Growth Machine; Urban Policy; Urban Representation/Imagination; Waterfront Development.

Further Reading

Boyle, M. and Hughes, G. (1994). The politics of urban entrepreneurialism in Glasgow. *Geoforum* 25(4), 453 470.

Chang, T. C. (1997). From 'Instant Asia' to 'Multi faceted jewel': Urban imaging strategies and tourism development in Singapore. *Urban Geography* 18(6), 542 562.

Cochrane, A., Peck, J. and Tickell, A. (1996). Manchester plays games: Exploring the local politics of globalization. *Urban Studies* 33(8), 1319 1336.

Cox, K. R. and Mair, A. (1988). Locality and community in the politics of local economic development. *Annals of the Association of American Geographers* 78, 307 325.

Hall, T. and Hubbard, P. (eds.) (1998). *The entrepreneurial city: Geographies of politics, regime, and representation*. New York: Wiley.

Harvey, D. (1989). From managerialism to entrepreneurialism: The transformation of urban governance in late capitalism. *Geografiska Annaler* 71B, 3 17.

Kearns, G. and Philo, C. (eds.) (1993). *Selling places: The city as cultural capital*. Oxford: Pergamon Press.

Liebovitz, J. (2003). Institutional barriers to associative city regional governance: The politics of institution building and economic governance in 'Canada's Technology Triangle'. *Urban Studies* 40(13), 2613 2642.

MacLeod, G. (2002). From urban entrepreneurialism to a revanchist city? On the spatial injustices of Glasgow's renaissance. *Antipode* 34(3), 602 624.

McCann, E. J. (2004). 'Best places': Inter urban competition, quality of life, and popular media discourse. *Urban Studies* 41(10), 1909 1929.

Short, J. R., Benton, L. M., Luce, W. B. and Walton, J. (1993). Reconstructing the image of an industrial city. *Annals of the Association of American Geographers* 83, 207 224.

Ward, K. G. (2000). From rentiers to rantiers: 'Active entrepreneurs,' structural speculators,' and the 'politics of marketing the city. *Urban Studies* 37, 1093 1107.

Ward, S. (1998). *Selling places: The marketing and promotion of towns and cities, 1850 2000*. London: E & FN Spon.

Relevant Websites

http://www.glasgow.gov.uk/en/YourCouncil/PublicRelations/Campaigns/glasgowsmilebetter.htm
Glasgow City Council.

http://www.joburg.org.za/
City of Johannesburg.

http://www.techtriangle.com/
Canada's Technology Triangle.

http://www.downtownla.com/
Downtown USA.

http://www.london.gov.uk/onelondon/
We are Londoners.

City-Region

S. Davoudi, Newcastle University, Newcastle upon Tyne, UK

Glossary

City-Region City and its environs.
Inner Core Area Core, Center, node, city, Urban tract, the C zone.
Outer Surrounding Area Hinterland, umland, metropolitan community, region, field of association, catchment area, the S zone.

Origin of the City-Region

Since the 1990s, there has been a remarkable resurgence of the concept of the city region in both academic and policy communities. In broad terms, the concept articulates the relationships between the city and its environs. For almost 30 years the focus on the city region had been overshadowed by a shift of emphasis toward urban and particularly inner urban issues. Hence, its revival is not only a reincarnation of an analytical construct to understand the complex spatial relations, but also a manifestation of a political move toward new regionalism. This has been coupled with the struggle over the rescaling of the state intervention to intermediate levels such as the city region.

While the city region's rising popularity is recent, its origin is not. The idea that the city cannot be understood fully by reference only to its administrative boundaries has a long history. While the term itself is thought to have been coined by Robert Dickinson in 1947, the concept can be traced back to an even earlier time in both theoretical and applied studies as well as planning practices. An example of the latter is the 1909 Plan of Chicago which promoted a regional vision of the city which extended well beyond its administrative boundaries. Similarly, in 1915 Patrick Geddes, the Scottish planner, used the notion of conurbation to advocate the need for planning to take into account the resources of the region in which the historic but rapidly spreading cities are situated. In 1933, R. D. McKenzie, an American sociologist, pointed out that the metropolitan region was primarily a functional entity.

However, despite its long history the concept of the city region does not enjoy a common definition; neither in its use as an analytical term nor in its upsurge as a political one. Analytically, it represents different spatial entities depending on how it has been arrived at methodologically. Politically, it means different things depending on the type of policy agenda it serves. In practice, the concept is frequently used simply to refer to the areal extent of a metropolitan area. While there are clear linkages between the analytical and political usages of the concept, the main focus of this article is on the city region as an analytical construct. As such, a key contribution of the concept to spatial thinking is its departure from a preoccupation with the physical structure of the city and the urban form *per se* toward a focus on the relational dynamics of the social networks and urban functions which often transcend the bounded perceptions of space. Hence, the contemporary relevance and significance of the city region concept lies in its potential: first, to evoke a relational understanding of space and place in policy and practice; second, to encourage researchers to seek new methodologies for mapping the less tangible interconnections which define the virtual contours of what Manuel Castells calls the space of flows; and third, to contribute to the debate about metropolitan governance.

Conceptions of the City-Region

Since the 1950s, there have been a myriad of studies attempting to: define the components of the city region; identify, measure, and map its boundaries; and, understand the interactions between its constituent parts. Despite their variations, they share two common features: first, they portray an urban centric conception of the city region which puts the emphasis on the city, sometimes at the expense of neglecting the region. Second, they represent an economically driven approach to city region definition in which the dominant economic flows determine the extent of the city region. Exceptions to these do exist but are rare and more recent.

The supremacy of the urban centric view lies partly in the historical evolution of the spatial structure of the city region. This until the mid twentieth century was largely influenced by a continuing concentration of population and economic activities in the core city in most of the developed world. Although this trend began to reverse in the latter part of the twentieth century – when, mainly in the West, population and employment began to decentralize – it did not substantially change the urban centric perspective. Indeed, such a view has been further boosted in the twenty first century by the apparent re emergence of the cities as the magnets for attracting population and economic activities. It is argued that regions with successful core cities at their heart can develop a vibrant and continually evolving hierarchy of settlements of different sizes that carry out a wide range of functions. Embedded within this view of the city region is an implicit recourse to Christaller's 'central place' theory. The theory implies a nesting hierarchy of cities which serves populations drawn from within regular hexagonal catchment zones at successively larger scales.

The continuing dominance of the economically driven approach may be explained by the prevailing policy frameworks based on competitive city regionalism. In deed, what underpins the re emergence of the city region since the late 1990s is a growing conviction that participation in the service intensive global economy has led to greater spatial concentration of activities. As a result, cities are re emerging as production sites and as the powerhouses of modern, advanced economies, where the creation of unique knowledge is of greater signifi cance than exploitation of natural resources or mass production. It is argued that city location is as important for this new knowledge based economy as factory floor space was for the industrial economy. This has made city regions an insistent element of the geographic landscape. Will Hutton has called these post industrial cities the 'ideopolis', arguing that they represent a twenty first century metropolitan version of what we first saw in Italian renaissance city states. Their buoyant demand, intellectual capital, and business self confidence help to create and sustain new ideas. Among the common fea tures which underpin the successful performance of post industrial cities are a diverse economic base, internal and external connectivity, quality of life, skills and human capital, innovation in firms and institutions, leadership, and strategic decision taking capacity.

Built on these arguments, it is concluded that the competitiveness of major cities is the key determinant of regional performance in advanced knowledge based economies.

The City-Region Interactions

The economic approach to city region is a limited one because it fails to recognize a full spectrum of city regional relationships, which constitute a complex web of visible and invisible multidirectional flows of not only economic but also social, cultural, and environmental activities (see **Box 1**).

Through these interactions, cities and regions may share a mutually reinforcing relationship. Furthermore, each of these interactions creates its own spatial ima ginaries and functional boundaries which do not always overlap. The outcome is a complex web of interactions and a multiplicity of boundaries. Hence, city region boundaries do not constitute precise lines; instead they are blurred, if not indeterminate. However, despite a long standing recognition of the complexity and emer gence of the city region spatial structures, attempts to define the city region have remained entirely focused on its economic functions. Hence, it remains the case that the most widely recognized and deeply embedded in terpretation of the city region is its conceptualization as a functional economic space.

Box 1 Potential flows in the city-region

- People (daily commuting to work, shopping, leisure; nondaily commuting for cultural, entertainment, and recreational activities; and migration)
- Goods (manufacturing and semiprocessed materials between firms)
- Services (banking, educational, health, and business)
- Capital and assets (investment, taxes, land ownership, and property rights)
- Waste and pollution (solid waste, emission, and water pollution)
- Environmental resources (water and minerals)
- Knowledge (technical information, social ideas, and experiences)
- Social norms, values, lifestyles, and identities

City-Region as a Functional Economic Space

Conceptualizing the city region as a functional economic space – pioneered by Brain Berry in 1968 – has remained an enduring and powerful spatial imagination. Within this perspective, the city region is depicted as a largely self contained and integrated economic entity, com monly known as the functional urban region (FUR). The concept of FUR was first developed in America as a way of moving away from an earlier population based concept of metropolitan district which referred to urban ag glomerations with more than 200 000 populations. The shift took place when Gras used economic rather than spatial criteria to identify 14 metropolitan areas in North America. Later, the notion of metropolitan area was formally adopted by the United States Bureau of the Census in 1950. These were called the Standard Metro politan Statistical Area (SMSA) and referred to areal unit of over 50 000 population. The defining characteristic of the metropolis was considered to be the commuting pattern of the workforce which was assumed to be radial from the periphery to the center(s) where the jobs were located. In Britain, the notion of 'conurbation', which was coined by Patrick Geddes in 1915 and further developed by Fawcett in 1932, was spatially, rather than functionally, defined. Fawcett defined seven conurbations in Britain which more or less corresponded to those that were later delineated by the statisticians in the General Register

Office (GRO) using the 1951 census. So, at the time when the British statistical authorities were using the concept of conurbation defined largely on the basis of physical extent of the city, on the other side of Atlantic, SMSA put functional integration at the center of the definition of metropolitan areas.

Meanwhile, in other parts of Europe, researchers were drawing on criteria similar to the ones used to define SMSA in order to delineate city regions. A notable ex ample is Hans Carol's definition of Zurich city region in 1956. Given Christaller's strong influence in the contin ent, Carol combined the two approaches (one based on commuting data to define the city region and the other based on the range of services to define the hierarchy of service centers) to map the arrangement of central places in the city region of Zurich. He then argued that a planned decentralization of activities should be guided by this framework. By the 1970s, these were also introduced in the UK by Peter Hall and his study of Standard Metropolitan Labor Areas (SMLA) for England and Wales. Both SMSA and SMLA consist of the historic city and its commuting hinterland instead of being limited to the continuously built up area centered upon a particular city. The functional approach to city region definition has been further advanced by studies such as those undertaken by Mike Coombes and Paul Cheshire in Britain as well as through pan European research pro grams such as the European Spatial Planning Obser vation Network (ESPON).

Regarding its utility, the FUR is often seen as an or ganizing mechanism in the national space economies. However, evidence from the survey conducted by the Organization for Economic Cooperation and Develop ment (OECD) in 2002 of its members has revealed that this is not the case in countries such as Japan, Mexico, Korea, Spain, and Turkey. Furthermore, the official rec ognition of the FUR as a framework for policy imple mentation varies considerably from one country to another. In some countries, such as Austria, Denmark, and Switzerland, FURs are used as a framework for the implementation of policies relating mainly to labor market and transport. In other countries such as Finland, France, Germany, Italy, and the UK, FURs serve as a basis for identifying areas which qualify for national or European aid. In Norway, the concept of FUR has played a significant role in the discussions about regionalism and new regional policy. In a fourth category, countries such as Portugal, Sweden, Czech Republic, and the United States do not use FURs as their official units for policy implementation. Nevertheless, adopting a FUR approach (as a proxy for the city region) to policymaking has remained an attractive proposition particularly in the area of strategic planning where the need for coordin ation and integration of investments, policies, and programs is considered to be critical. Indeed, such arguments have underpinned the current debate over the co aligning of local government structures to that of the city regions.

Mapping the City-Region

While there have been numerous attempts to define the city region – as an FUR – the emphasis has remained on the city as the node of interactions, and on the economy (or indeed on work related commuting) as the deter mining function. This has largely shaped (and been shaped by) the methodologies that have been used to map the FURs. Despite their differences such method ologies fall into one of two broad categories: one is the top down and deductive approach and the other is the bottom up and inductive approach.

The Top-Down Approach

In the top down, deductive approach, analyses start from a predetermined set of cities as nodes or destinations for the economic flows and then move out to assign other areas to them. These cities (nodes) are selected on the basis of a specified set of functions, population size, economic performance (often measured by the gross domestic product), accessibility and connectivity, inten sity of financial and business services, and so on. De pending on the weight given to the criteria for the selection of nodes, the number of cities and hence the frequencies of the city regions reduce or increase. While a classic example of the top down definition is the official definition of Metropolitan Statistical Areas in the US, the approach has also been used in countries such as Canada, France (urban areas), Germany, and Portugal, as well as in pan European studies such as ESPON.

The Bottom-Up Approach

In the bottom up, inductive approach, the city regions emerge from the full set of commuting data through an algorithm which simultaneously optimizes the bound aries on the basis of a size of employment criterion and a minimum threshold of self containment of flows to workplaces. This approach has become feasible only re cently as a result of advances in data manipulation and computation. It still remains an exceptional feature of a landscape which is dominated by the traditional top down definition of the FUR. The most notable attempt to adopt the bottom up approach is the work undertaken by Coombes *et al.*, in 1986. Different variations of this ap proach are also used in countries such as Finland, France (for defining employment areas), and Italy.

In technical terms, the first approach is nodal and nonexhaustive (not covering the whole of the national territory), while the second one is non nodal (or multi nodal) and exhaustive, as illustrated in **Figures 1** and **2**.

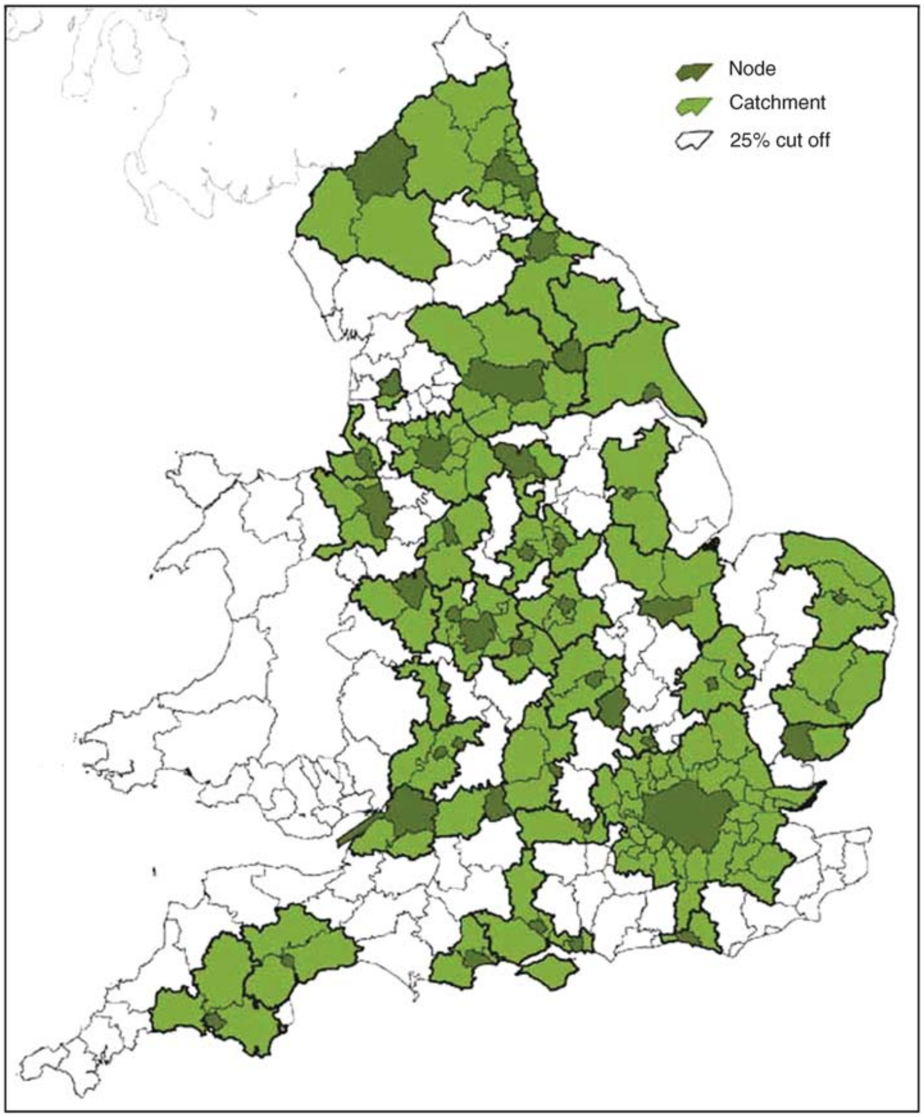

Figure 1 Example of a top-down approach to city-region definition in England, 2001. Source: Robson, B., Barr, R., Lymperopoulou, K, Rees, J., and Coombes, M. (2006). A *Framework for City-Regions. Working Paper 1: Mapping City Region*. London: Office of the Deputy Prime Minister, p 22, Figure 2.8: city region boundaries with a 25% cutoff.

The first one is underpinned by an urban centric conception of the city region, while the second one marks a departure from it, shifting the emphasis from the city to the region. However, while the bottom up approach provides a more inclusive way of understanding the complex flows of commuting across a wider territory, it remains an economically driven approach and largely reliant on travel to work analyses.

One advantage of the non nodal approach is that it can reveal potential polycentric patterns in the wider region. As Davoudi has shown, promoting the development of such polycentric urban regions – where a number of neighboring and nonhierarchical cities interact with each other economically and complement each other in service provision – has become a central objective of the European spatial policy. However, given the dominance of the top down approach and its upsurge in policy domain, the remaining sections of this article focus on providing an outline of its evolution and key features.

The Structure of the City-Region

The top down approach attributes two distinct but interrelated components to the city region: the inner core area (central urban tract) and the outer surrounding area which is associated with and sometimes dominated by the core. The two entities have been given various terminologies. The inner area has been called: the core,

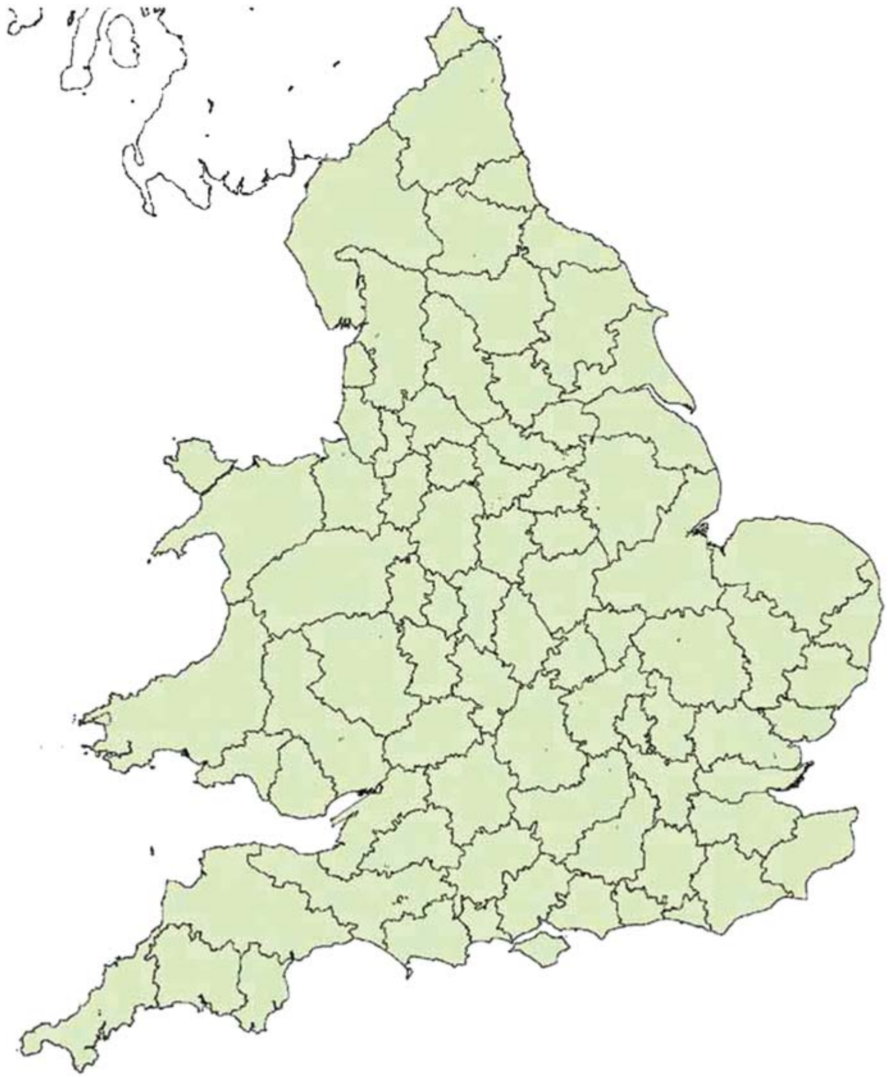

Figure 2 Example of a bottom-up approach to city-region definition in England and Wales, 2001. Source: Robson, B., Barr, R., Lymperopoulou, K, Rees, J. and Coombes, M. (2006). A *Framework for City-Regions. Working Paper 1: Mapping City Region*. London: Office of the Deputy Prime Minister, p 11, Figure 2.1: Bottom-up city regions: all workforce, self containment 70%.

the center, the node, the city, and the urban tract. The outer area has attracted terms such as: the 'hinterland' by Gras in 1922, *umland* (the land around) by the German scholar Schöller in 1957, 'metropolitan community' by Bogue in1949, and the 'region' by McKenzie in1933 and Dickinson in 1947. For Mumford, the city's region is its field of association or catchment area. In 2005, Parr co dified these two entities by calling them the 'C zone' and the 'S zone' to disassociate them from their historical baggage and to use categories that are exhaustive and mutually exclusive. For Dickinson, the outer area can be further divided into two categories: the city settlement area or the daily commuting area, and the city trade area or the catchment area for central services.

An early and striking example of a top down, urban centric view of the city region is the one portrayed by Donald Bogue in 1949 who, drawing on biological analogies, delineated 67 metropolitan regions in America based on an arbitrary system of decreasing values, be ginning with the metropolitan center as the dominant and followed by the hinterland city as subdominant, the rural nonfarm as influent, and the rural farm as subinfluent. Another example of identifying associational zones is the work undertaken by Boustedt in 1953. He defined the city regions (*stadtregion*) of Germany as consisting of a core (one or more contiguous adminis trative units and associated *gemeinden*), an urbanized zone (*verstädterte* zone), the fringe zone, and the wider area of *umland* within which the influence of the city predominates. Mapping the city region (or more pre cisely the FUR) has invariably involved identifying such zonal arrangements. In the top down approach, this involves first, defining the core and second, identi fying the extent of its influence by measuring the economic flows to and from the core. These are elaborated in turn.

Defining the City in the City-Region

A top down approach to map the city region starts with defining the core city (or urban tract). Tradition ally, this has referred to the compact and fairly continuous built up area. Such an area has also been a defining feature of conurbation as used by Fawcett in 1932. For him, a conurbation was an area occupied by a continuous series of dwellings which were not separated from each other by rural land. This approach and its follow up by British (GRO) and American stat isticians (SMSA) were designed partly to distinguish the predominantly urban areas from the predominantly rural ones. In that sense, their understanding of conurbation was not the same as the one used by Patrick Geddes. His was more consistent with the wider concept of the city region as a functional space and came closer to what Friedmann and Miller called the urban field.

Meanwhile, in France the notion of urban aggregates were defined – for the 1954 census – as those contiguous communes, which were linked together primarily in terms of family associations (*cadre de l' existence familiale*). This included the geographic framework in which, for example, schools and shops were located but excluded the workplaces. In Germany, while there were no official definitions of urban areas by census authorities at that time, there were several studies which attempted to define the city region. A notable example is Hesse's work on typologies of *gemeinden* in southwest Germany in the 1940s.

Using the compactness of the urban uses as a cri terion remains at the center of the statistical definition of the core city as well as the distinction between what is urban and what is rural. For example, in Finland (in common with other Nordic countries) a built up area, *taajama*, refers to a cluster of buildings containing at least 200 residents in which the distance between buildings does not exceed 200 m. However, given the diffusion of the urban–rural fringe, the definition of an urbanized core in terms of continuous urban land uses has become increasingly elusive and in many cases ir relevant. Definitions of the urban based on economic function have grown apart from definitions based upon physical development. For example, using the terms loosely, England is in physical terms predominantly rural but in socioeconomic terms overwhelmingly urban. Indeed, an important feature of the city region, par ticularly when it is based on the bottom up approach, is its potential to reveal the multifaceted interrelationships between urban and rural areas within a particular spatial context.

Delineating the Extent of the City-Region

There are two ways of defining the city region (or FUR) as a geographical entity: one is through statistical analysis of actual flows and the other is by an approximation of time–distance from the core. Both are briefly outlined below.

Measuring the flows

The focus of the statistically defined city region is on measuring the precise areal extent of the city's associ ations. While such associations are dynamic and include a wide range of economic, social, and environmental links, the hegemony of economic discourse and economic data has meant that the only city region interaction which has proved amenable to measurement and mapping is the flows of people or more precisely the flows of daily commuters to work. This is evident by the OECD's survey of its members which concludes that the most typical concept used in defining a functional region is that of labor markets.

Within this perspective, after the delineation of the core city, the extent of the FUR is determined by including each locality that has more than a given proportion of its active labor force working in the core city and hence commuting from such localities to the core on a daily basis. Clearly, the lower the pro portion the larger will be the extent of the FUR and vice versa.

In technical terms, a low percentage point will cause overbounding, while a high threshold will lead to underbounding. In the case of the former, areas whose labor market linkages with the core are not above a certain threshold will be excluded from the resulting city region, even if they may be tightly connected to it through, for example, environmental, cultural, or ad ministrative ties. Conversely, when the threshold is low, the resulting FUR will engulf those cities and towns that fall within that threshold even if they assume historical and political independence. The issue is not a mere technical one, and can have far reaching implications for spatial planning, investment programs, and the sense of place and identity.

Indeed, a major criticism of conceptualizing the city region as a FUR is its heavy reliance on not only one type of interaction (i.e., economic) but also on one eco nomic criterion (i.e., travel to work) for measuring that interaction. Such reductionism has been compounded by a lack of attention to the differences between labor market areas. Often, the average commuting distance of the higher order occupational groups (such as pro fessional and managerial) is much longer than the lower ones (such as unskilled manual), leading to more than one city region boundary for one type of commuting, as shown in **Figures 3** and **4**.

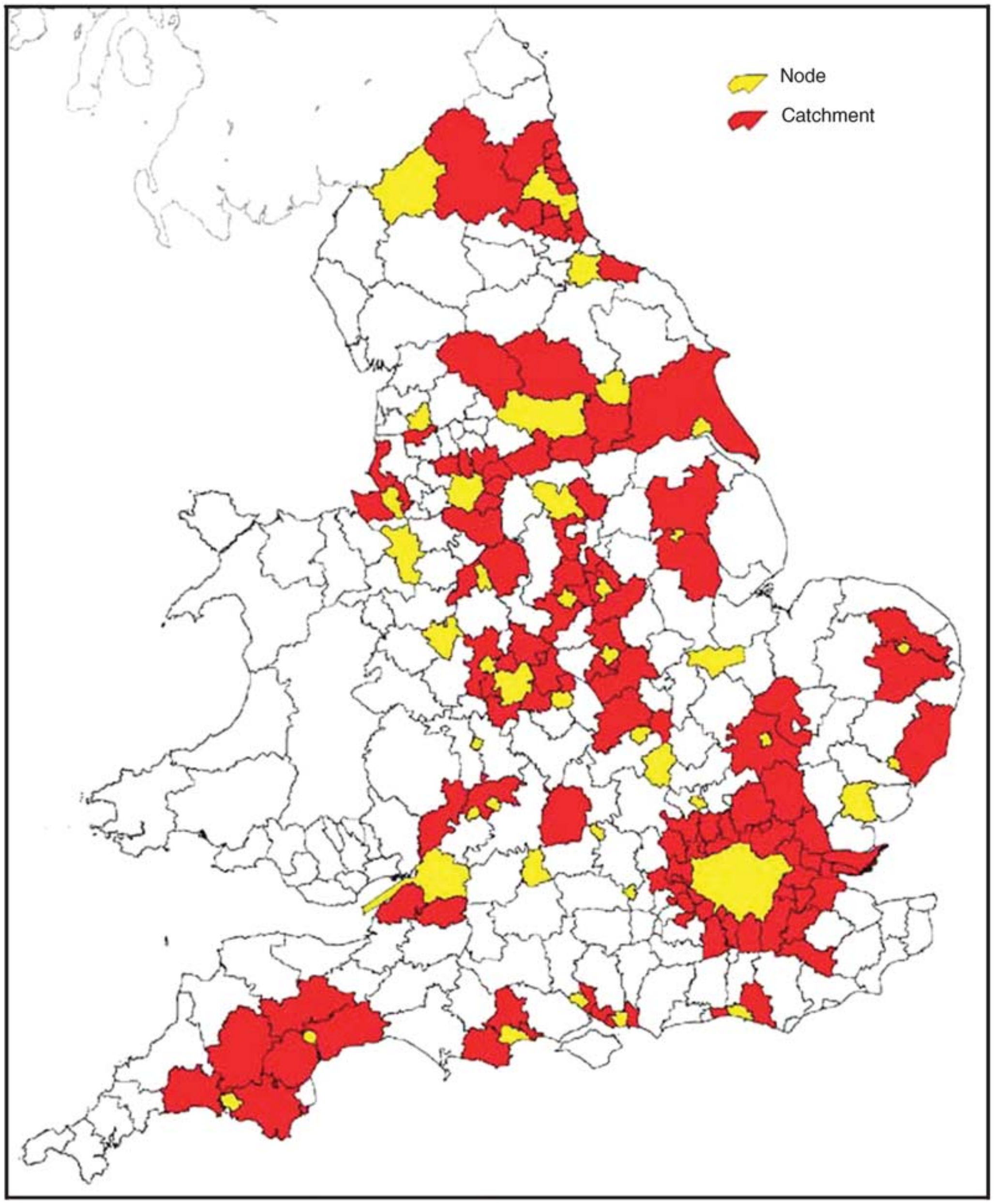

Figure 3 Example of city-region definition based on professional and managerial labor market area in England. Source: Robson, B., Barr, R., Lymperopoulou, K, Rees, J. and Coombes, M. (2006). A *Framework for City-Regions. Working Paper 1: Mapping City Region*. London: Office of the Deputy Prime Minister, p 25, Figure 2.10: Catchment for professional and managerial workers.

Furthermore, labor market analyses have a limited utility in capturing the economic relationships of the FUR. For a more sophisticated approach other nonwork trip generation activities, such as journey to shops, ser vices, and leisure facilities, need to be taken into ac count, despite the difficulties of measuring them. If that becomes feasible, the city region boundaries will further diverge. Yet another complexity arises when the housing market areas are taken into account. It appears that for households making residential location de cisions, the city region boundary is defined by their mental maps of areas with which they are familiar. Such areas do not necessarily match that of the labor market area, hence creating yet another boundary for the city region. In an ambitious attempt to move beyond the labor market areas and capture the economic relation ships of the mega city regions, a European project called Polynet has attempted to measure the interactions be tween the firms and particularly those which are key intermediaries in the knowledge economy such as, ad vanced business services, financial institutions, law, ac countancy, and management consultancies. Given that

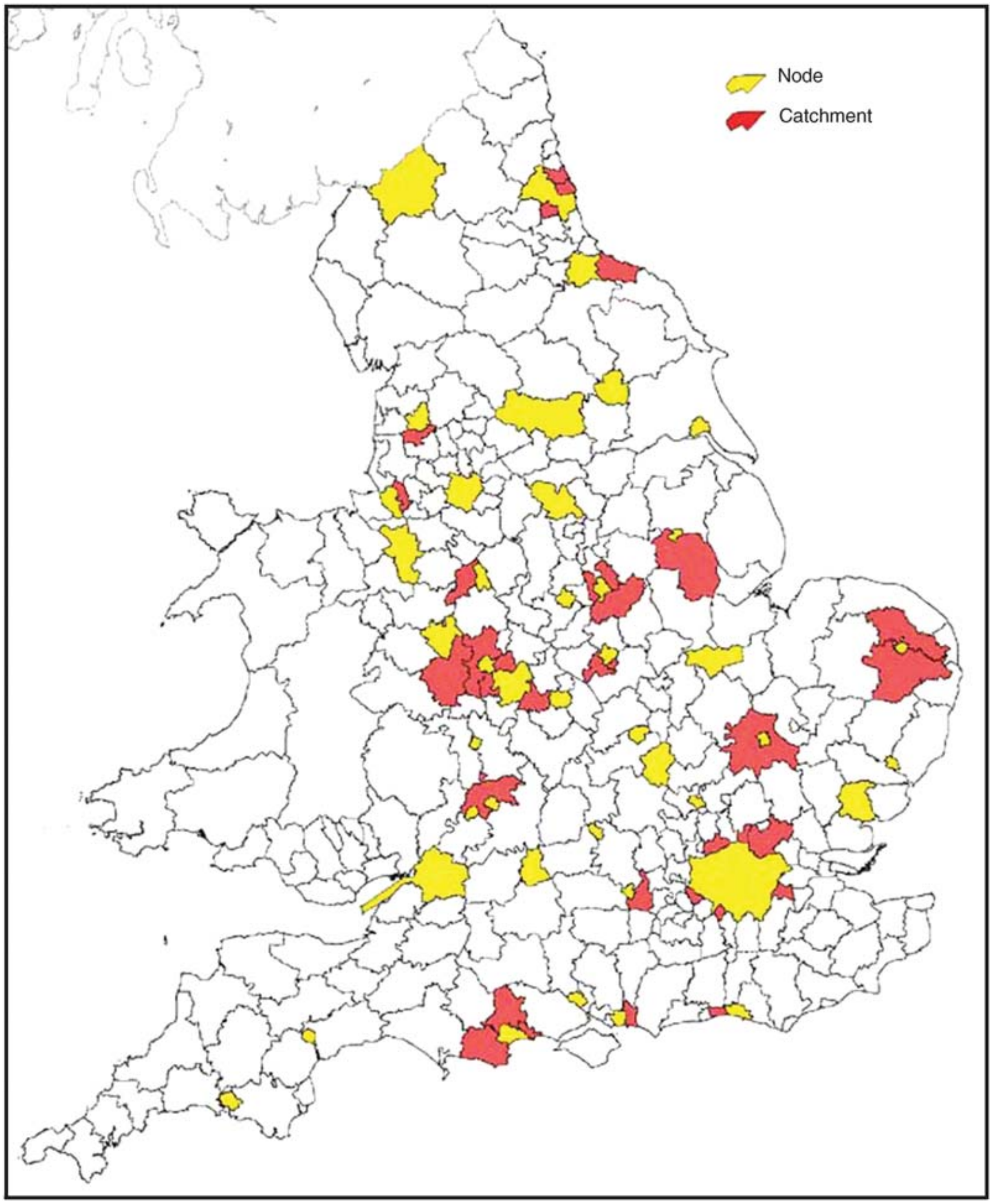

Figure 4 Example of city-region definition based on unskilled and routine labor market area in England. Source: Robson, B., Barr, R., Lymperopoulou, K, Rees, J. and Coombes, M. (2006). A *Framework for City-Regions. Working Paper 1: Mapping City Region.* London: Office of the Deputy Prime Minister, p 26, Figure 2.11: Catchment for semiskilled and routine workers.

much of such interactions take place in the cyberspace and via the Internet, measuring them has proved no toriously difficult. What is certain is that these interactions take place within a city regional boundary which is likely to be different from those mentioned above. The key message is that such complexities con firm the view that the city region geometry is best characterized by multiplicity, fuzziness, and overlaps, manifesting only an approximation of self containment which is likely to vary for different kinds of activities, flows, and functions.

Approximating the time–distance

In addition to the use of statistical analyses of the actual commuting flows, there have been other attempts to map the city region using an approximated time–distance considered to be reasonable or convenient for daily com muting. Prior to the introduction of public transport in the late nineteenth century, the 3 mile radius was considered as the limit of travel by foot and hoof. In modern metro polis, Blumenfeld argued that the reasonable traveling time from the outskirts to the center had to be no more than 40 min. For Geddes, the rule of thumb for convenient commuting distance was an hour. While most commen tators have used the latter as the maximum commuting distance, in 1995, Batten has argued for the lower limit of half an hour. More recently, ESPON research used a measure of 45 min drive time as a proxy for identifying functional urban areas across Europe. In all these, lines of equal time–distance to and from the city (isochrone) are used to define the extent of the city region.

Whatever considered as a comfortable commuting time, this methodology remains arbitrary and suffers from the same flawed assumption which was made by Bogue in 1949 when he mapped metropolitan regions in America. He assumed that accessibility between one center and another was a basic determinant of city regional relations. The assumption was that a metropolis can dominate all of the area which lies closer to it than to any other similar city, even if the other metropolis is larger. While proximity may be necessary for the creation of a functional economic space, it is certainly not sufficient. Having said that, it is true that as speed and convenience of travel increase, so will the isochrone and hence the extent of FUR, because people can, if other conditions are right, commute farther distances within the same time span. A notable example of the role of infrastructure in expanding the FUR is the development of Oresund Bridge between Copenhagen and Malmo which has led to the creation of a transnational city region.

City-Region as a Political-Administrative Space

Conceptualizing the city region as a political adminis trative space has been at the heart of debate about ap propriate spatial scale for strategic decision making and for coordination of public investments, policies, and programs. It is argued that the local government struc ture should be coaligned with that of the functional economic space. This argument has been strongly pur sued within the spatial planning communities. As early as 1915, Geddes was advocating that for sensible planning conurbation is the most appropriate spatial scale. Across the Atlantic, the 1909 Plan of Chicago provided one of the first comprehensive plans for a city region. However, by the end of World War II there were still only about ten city regional governments in place in the developed world, most of which were dependent on central governments.

In England, a major attempt to co align the ad ministrative boundaries with that of the functional urban regions was made in the 1960s by Derek Senior, a member of the Royal Commission on Local Government. His Minority Report advocated the establishment of 35 city regions for England to form the upper tier, of a two tier structure, of local government with responsibility for strategic planning. However, the subsequent legislation for reform failed to implement these aspirations. Hence, there remains in Britain – as in many other developed countries – the problem of overbounding when the mu nicipal boundary covers large tracts of rural land (see Leeds in **Figure 5**) and underbounding when the mu nicipal boundary does not cover even the whole of their built up area (see Birmingham in **Figure 5**) let alone their functional hinterland.

As regards the city region scale, the OECD survey (mentioned above) highlighted that in most of its member states there is no formal structure of governance operating at the level of functional urban regions. However, there has been a proliferation of voluntary intermunicipal coop erations at, loosely defined, city regional level. Given the variable geometry of the city regions, such voluntary collaborations provide a more flexible governance frame work, capable of adapting to the rapidly changing dy namics of the city region configurations. While such cooperative arrangements are motivated predominantly by the economic competitiveness agenda, they are also in creasingly justified by the sustainable development agenda, and in particular the need for managing the environmental footprints of the cities over their hinterlands. More im portantly, the environmental concerns as well as the rec ognition of the role of culture and identity are increasingly challenging the narrow, urban centric, and economically driven definitions of the city region. They are calling for a focus on the emerging relational dynamics of the city region space.

Conclusion

The city region has enjoyed a widespread resurgence in recent years as both an analytical and a political concepts. The focus of this article has been on the former, aiming to unpick the concept, trace its origin, and explore its evolution. Particular emphasis was placed on a critical review of the methodological approaches to the defin ition of the functional urban regions.

While in the last 60 years or so, many approaches have been adopted to define the city region and map its boundaries, this article has shown that with a few ex ceptions, the dominant conception of the city region has been urban centric and economically driven. While the dominance of this econo urban perspective has historical roots, it has been further boosted by the apparent re emergence of the cities as the magnets for attracting population and economic activities and a prevailing policy framework based on competitive city regionalism. It is argued that competitive cities create prosperous regions through a potential chain reaction which can follow from a high level commitment to strengthen a city's knowledge base as a launch pad for stronger economic performance

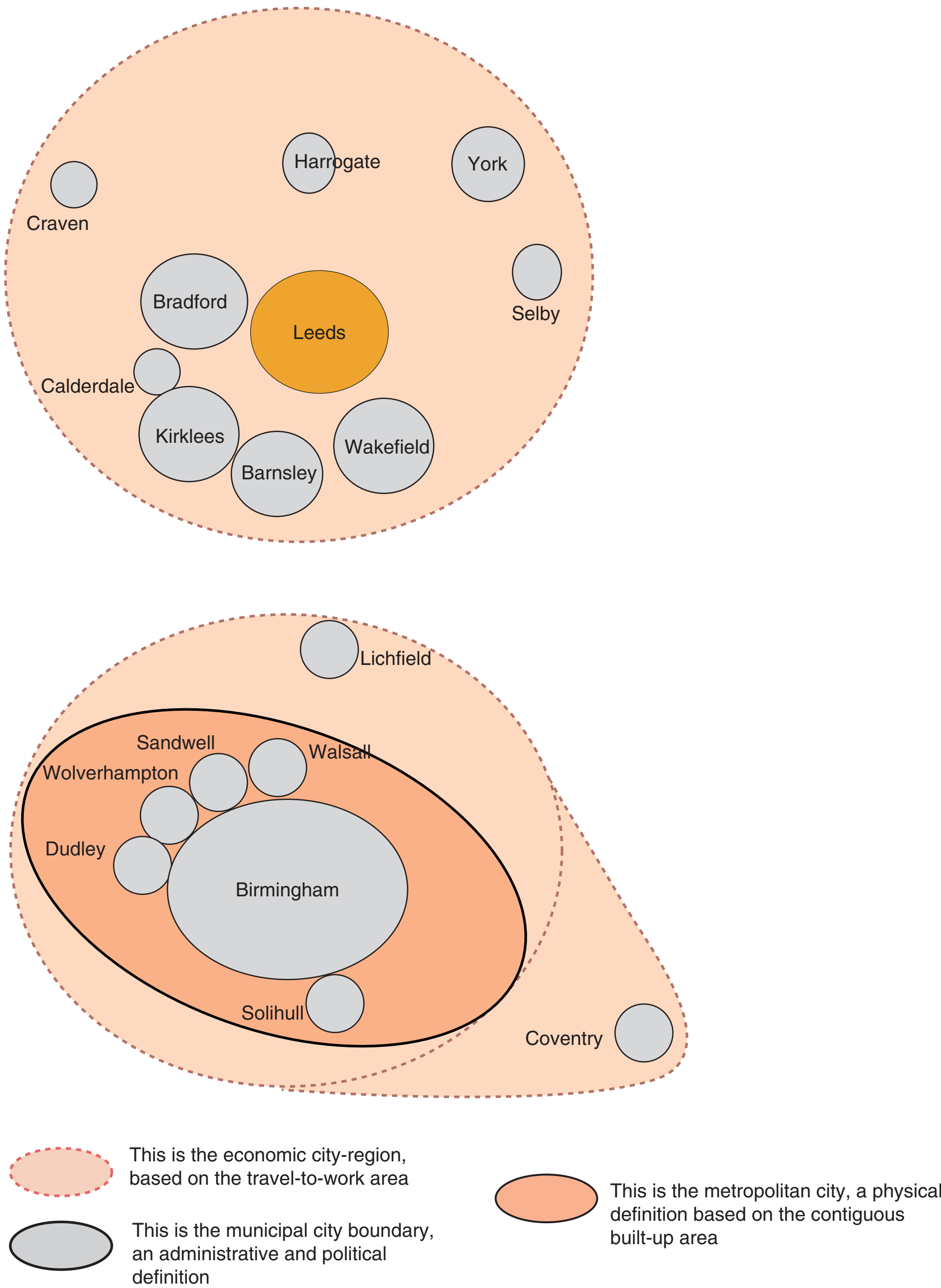

Figure 5 Examples of administrative overbound and underbound (Leeds and Birmingham in England). Source: HMSO (2006). *Devolving Decision Making: 3 - Meeting the Regional Economic Challenge: The Importance of Cities to Regional Growth*, London: HMSO, p 12, Figure 2.1: The coincidence between administrative cities and city-regions varies.

and a distinctive external profile. While the evidence for such trickle down or ripple out effects is yet to be documented, this argument continues to frame the debate about the city region in both analytical and political terms.

Analytically, it has fueled the dominance of the urban centric and economically defined conception of the city region, as clearly reflected in the interpretation of the city region as a functional urban region and the top down approach to define its boundaries. Politically, it has become the key justification for channeling policy attention, investments, and political powers toward the economic growth in major cities often at the expense of those parts of the regions which fall outside the cities' hinterlands.

Both the urban centric and the economically driven view of the city region have been challenged. The former is under increasing strain by the changing patterns of work, mobility, and lifestyle in rural areas, while the latter is questioned by a rising concern over the environmental footprint of the cities which affect their hinterlands and beyond. Despite this, there remains a key shortcoming in the current city region agenda; that is, the disconnection between the conceptions of the city region as an economic space from its conceptualization as an ecological space. Contemporary city regions are key sites of consumption of environmental resources and production of waste. As such, they are engaged in not only economic exchanges and distributions of benefits, but also in ecological exchanges and distribution of risks and disbenefits. An illustrative example of this is the city regional relationship based on the management of municipal waste which is based largely on generation of waste in the city and its disposal in its hinterland and beyond. Critical examination of how these changing ecological relationships may reshape our conceptions of the city region and deepen our understanding of the emerging relational dynamics of the city regional space will therefore make a valuable contribution to push the debate forward.

See also: Central Place Theory; New Regionalism; Regionalism.

Further Reading

Coombes, M. and Wymer, C. (2001). A new approach to identifying localities: Representing 'places' in Britain. In Madanipour, A., Hull, A. & Healey, P. (eds.) *The Governance of Place: Space and Planning Processes*, ch. 3, pp 51 69. Ashgate: Aldershot.

Dahmann, D. C. and Fitzsimmons, J. D. (eds.) (1995). *Metropolitan and Nonmetropolitan Areas: New Approaches to Geographical Definition*, Population Division working Paper 12, US Bureau of the Census, Washington DC.

Davoudi, S. (2003). Polycentricity in European spatial planning: From an analytical tool to a normative agenda. *European Planning Studies* 11(8), 979 999.

Dickinson, R. (1964). *City and Region*, Part III, *The City Region*, pp 227 435. London: Routledge and Kegan Paul Ltd.

Eurostat (1992). *Study on Employment Zones*, Eurostat Report E/LOC/20, Luxembourge.

Hall, P. and Pain, K. (2006). *The Polycentric Metropolis: Learning from Mega City Regions in Europe*. London: Earthscan.

Organization for Economic Co operation and Development (2002). *Redefining Territories, the Functional Urban Regions*. Paris: OECD.

Parr, J. (2005). Perspectives on the city region. *Regional Studies* 39(5), 555 566.

Ravetz, J. (2000). *City Region 2020: Integrated Planning for a Sustainable Environment*. London: Earthscan.

Robson, B., Barr, R., Lymperopoulou, K., Rees, J. and Coombes, M. (2006). *A Framework for City Regions. Working Paper 1: Mapping City Region*. London: Office of the Deputy Prime Minister.

Scott, A. (2001). Globalization and the rise of city regions. *European Planning Studies* 9, 813 826.

Relevant Websites

http://www.espon.eu

Civil Society

C. McIlwaine, Queen Mary, University of London, London, UK

Glossary

CBO Community-based organization functions at the community level, and is often based on membership and focuses on local issues such as neighborhood committees.

Good Governance Used by donor agencies to refer to the nature of public decision making and the exercise of public power. It is usually associated with institutions that support neoliberal development and is meant to encompass participation, consensus, transparency, and efficiency.

GRO Grassroots organization operates at the local level and usually around a specific issue such as securing a water supply or providing credit.

NGO Nongovernmental organization referring to a nonstate, nonprofit organization which can operate at a range of scales.

Social Capital Relations of trust and participation inherent within social networks and civic organizations.

Social Movement Refers broadly to groups of people that are protesting about an issue or challenging the *status quo* in some way. They are usually organized around a specific issue, such as land reform, gender, civil rights, the environment, and so on.

Introduction

Civil society is a contested term that broadly describes the arena of collective social interaction situated between the state, market, and household. Despite being used in a normative manner to encompass a host of values linked with participation and cooperation, it is more commonly used to describe a range of formal and informal nonstate organizations, groups, and associations that form part of the voluntary sector. These usually refer to community organizations, activist groups, trade unions, faith based groups, cooperatives, as well as more formal nongovernmental organizations (NGOs). While business, media, and professional organizations are usually included in definitions, this may vary, as does the inclusion of political groups. Civil society is often used synonymously with NGOs, especially in policy circles, even though, strictly speaking, NGOs represent only one element of civil society, albeit an important one.

Within the social sciences, civil society has become an especially important focus for research within international and post socialist development circles as part of debates linked with the collapse of communism in the former Soviet Union and Eastern Europe and democratization processes in the Global South. From a policy perspective, these debates have fed into the neoliberal concerns of the New Policy Agenda pursued by the bilateral and multilateral development organizations that have championed civil society as an antidote to the rollback of the responsibilities of the state. While arguably, geographers have come somewhat late to recognizing the importance of civil society, more recently, they have been crucial in calling for greater sophistication in the use of the term. This is particularly in terms of acknowledging profound variations in civil society activity, but also its inherently political nature. At the same time, civil society has been widely critiqued by those working from an international development perspective focusing particularly on the problems of using civil society as a policy tool rather than as an analytical construct.

Delineating Civil Society

Beyond these basic definitions of civil society are a host of variations influenced by disciplinary origin, political perspective, as well as the discord between theory and practice. Also important is that there are not only variations in definitions of civil society, but also diversity within the category. Indeed, there is often a tendency to assume that civil society is homogenous and monolithic. In turn, civil society organizations are often thought to share aims and work together in partnership, especially from a normative perspective. However, empirical studies throughout the world have illustrated that civil society is highly variegated across a range of scales from local to global, as well as by organizational type, purpose, function, and orientation. Especially important is the need to recognize geographical variations in civil society activity that is often strongest in certain areas such as capital cities. In addition, the contextualized local relationships of power, politics, and influence are also critical in shaping the nature of civil society. Thus, some note that a minimum distinction should be made between formal civil society organizations such as NGOs, and informal ones such as social movements. The former is associated with a view of civil society as a collective of organizations while the latter tends to encompass civil society as an arena for action, although these distinctions are not mutually exclusive. Others suggest that it is important to focus on the civil society that actually exists rather than what is presumed to or thought to exist. In turn, civil society must be recognized as dynamic at a temporal level as state–society relationships shift and allegiances within and between different organizations change. It is also important to view civil society as a sphere of social reproduction and organization that has always existed in some form in most societies, rather than an entity. In deed, definitional problems often arise when civil society is used as a catch all term that implies homogeneity, and for some even renders it useless and/or reified.

Ideological Roots and Theoretical Perspectives on Civil Society

Perhaps most significant in understanding the multiple meanings of civil society is to contextualize the emer gence of the concept through a brief examination of its ideologically complex origins. Indeed, only by doing this is it possible to comprehend the roles and functions ac corded to civil society.

As a concept, civil society was first coined in ancient jurisprudence and subsequently in Medieval political philosophy. While also mentioned in the seventeenth century, not until the eighteenth century did the term evolve into its current usage, largely through the work of Adam Ferguson. In the nineteenth century, civil society was widely debated by Hegel, Marx, and later Gramsci, together with others such as de Tocqueville. Some of the core discussions over the centuries centered around the ways in which the state and civil society were related or separate, how civil society encompassed economic interests, and how it functioned as a space for action, and as a buttress against the state. These ideas have fed into contemporary theoretical debates that can be somewhat crudely divided into two main perspectives.

The first is the liberal democratic approach that is closely associated with the writings of Alexis de Toc queville on nineteenth century North America and es pecially on the benefits of civic organizations for the creation and maintenance of democracy. Put simply, this view conceptualizes civil society as a largely autonomous sphere of freedom and liberty encompassing an organ izational culture that fosters democracy, both political and economic. Civil society is thus viewed as a positive force and potential conduit between the individual and the state. It ensures that the state does not become too powerful, and also instills a sense of participation, duty, and tolerance among citizens. There are several vari ations and interpretations of civil society within the broad liberal approach. An important one is the regime school that is circumspect about the ability of civil society to foster democracy and places emphasis on state reform. However, it is the neoconservative version that has been most influential in recent years, especially in influencing donor discourses and 'good governance' agendas (see below). Here, civil society is viewed mainly instrumentally as important in delivering services through NGOs in a context of a pared down state, as well as helping to prevent 'nondemocratic' forces from threatening state power.

Another variant encompassed within the liberal ap proach is the associational school that emphasizes the interrelationships between civil society and social capital. Heavily influenced by the work of Robert Putnam on social capital and civic participation in Italy and later applied to the US, there is thought to be a synergistic relationship between the emergence of a strong civil society and the formation of social capital. The creation of social capital, broadly defined as the relations of trust and participation inherent within social networks and civic organizations, is therefore thought to be one of the positive by products of strengthening civil society. Also Tocquevillian in inspiration, social capital is often viewed in a uniformly positive light, despite more recent cri tiques that highlight the lack of consideration of power imbalances and social differentiation which may ultim ately lead to the further homogenization of civil society.

The other main theoretical approach to understand ing civil society is the post Marxist or more specifically, Gramscian perspective. In contrast to the liberal demo cratic approach, this view conceptualizes civil society as internally divided along various lines of social cleavage. Organizations within this sphere may both challenge and uphold ideological hegemony. Most commonly, this ap proach is linked with the ways in which civil society can act as a site of resistance and struggle against authori tarianism. This was particularly relevant in understand ing resistance to such regimes in Eastern Europe and Latin America in the 1980s and 1990s. Furthermore, this perspective can also be associated with the emergence of social movements as important forces in challenging dominant thinking, whether about politics, poverty, gender, or the environment.

In part, the post Marxist or Gramscian approach in corporates the many criticisms that have emerged about the liberal democratic view from a theoretical per spective. Although much of this critique relates to the empirical reality of the role and function of civil society on the ground, and especially of NGOs, it still has im portant theoretical resonance. Perhaps one of the most common critiques relates to the separation of civil society from the state. Empirical reality has shown that the two are intricately interrelated and indeed, interdependent, and that the nature of the interactions is determined by the historical, economic, and political circumstances of particular countries and regions. Linked with these are critiques that challenge the notion of civil society as homogenous, as noted above, but also that it is inherently 'good' and automatically harbors and fosters democratic consolidation. It has been widely noted that it is much too simplistic to assume that civil society organizations, and NGOs in particular, will pluralize and democratize.

This is because, first, the mere existence of civil society organizations does not lead to a concomitant emergence of democracy, and second, that civil society comprises a huge diversity of organizations that are likely to have competing aims and political viewpoints. Indeed, civil society organizations are just as likely to be founded on authoritarian principles as on participatory and demo cratic ones. Furthermore, some civil society organizations may align themselves with the state in some circum stances depending on its political hue, while others may deliberately challenge the government of the day.

Another important dimension of the liberal and es pecially neo Tocquevillian interpretations is that as a concept, civil society is Western in origin, yet it is being hailed as a critical component of development efforts in the Global South. Moreover, while early theorization was based on Europe, more recent research has focused on the United States in particular, leading to what some call the 'Americanization' of the civil society debate. This refers not only to the normative, partial, and conservative forces behind the resurgence of the concept, but also the enthusiasm for it within US foundations as well as multilateral organizations. From a Southern perspective, there has been much consternation in some circles about the imposition of a European and North American inspired construct, especially in the African context. In particular, concerns have been voiced about the relevance of talking about fostering and strengthening civil society in countries where civil society has existed in various guises for centuries. Others compare the imposition of yet another concept as a form of neocolonialism and another way of imposing neoliberal values on African or other non Western societies. Having said this, most commentators have now come to terms with civil society. Most agree that as long as the concept is used in a way that does not homogenize or prescribe how civil society should function then it can be useful in thinking about how collective organization can be harnessed in order to promote democracy within countries.

In the empirical and especially the policy reality of using the concept of civil society, there is often no clear cut distinction between these various theoretical approaches. Indeed, what makes civil society such an important concept is that it appeals to both ends of the political spectrum. However, the liberal, and especially, the neo Tocquevillian approach has strongly influenced de velopment policy discourses led by donor agencies, with little consideration of Gramscian inspired interpretations.

Civil Society and Development Policy

Although the liberal approach has dominated policy discussions about civil society, theoretical wrangling is largely absent from donor discourse, with conceptual positions discerned only implicitly. Yet, this should not underestimate the importance of the concept of civil society for the development policy community. Few would contest the fact that civil society has formed an integral dimension of the 'good governance' agenda of organizations such as the World Bank, the UK's Department for International Development (DFID) and the US Agency for International Development (USAID). On one hand, this has highlighted the notion of civil society in checking the potentially authoritarian elements of the state, as well as promoting democratization through participation. On the other hand, it has also fed into neoliberal discourses of reducing the role of the state, enhancing the role of the market, and asserting the importance of the individual. Thus, since the 1990s, donor agencies have increasingly supported the idea of 'strengthening civil society' as a way of promoting sus tainable development through economic liberalization, state modernization, and democratization. This was fa cilitated through the establishment of a host of 'civil society units' within donor agencies, such as the Centre for Democracy and Governance established within USAID, and the State and Civil Society Unit within the Inter American Development Bank in the 1990s. Within the World Bank, a new structure was introduced in 2002 to facilitate their growing engagement with civil society. This entailed the appointment of 120 Civil Society Engagement Specialists working in country offices as Civil Society Country Staff, in a Civil Society Group in the Washington Headquarters; together they constitute the Civil Society Team.

One critically important aspect of donor engagement with civil society is not only the liberal and often neo conservative theoretical influence on understanding of the concept, but also the almost universal way in which donors tend to equate civil society organizations with and so reduce them to NGOs. Thus, donors have increasingly channeled funding through NGOs. This has involved first, an attempt to 'build civil society' and promote citizen engagement, mainly through institution and capacity building and partnership programs, and second, to finance the alternative delivery of services that have been eroded due to the paring down of the state through neoliberal structural adjustment policies.

Although NGOs have now become major players within development policy, their increasingly powerful roles have generated considerable criticism, some specific to NGOs and some linked with the wider use of the concept of civil society. Indeed, many of the problems with the enhanced role of NGOs in delivering develop ment relates back to the earlier critiques of the liberal democratic and normative theoretical perspective dis cussed above. One key issue is that donors often ignore the huge diversity of NGO activity and NGO types; just as civil society more generally is diverse, so are NGOs (see also below). This becomes a particular problem for donors who tend to assume that NGOs have a greater capacity to reach the poor, and to promote local participation. However, many NGOs do not necessarily work with the poor, nor do they involve citizens or those at the grassroots in their work. Indeed, it is really grassroots and community based organizations who work at this level, and who often have much more democratic internal management structures than larger NGOs. Linked with this is that the function of NGOs is also diverse yet often ignored by donors. One of the main distinctions, for example, is between those that provide services (sometimes referred to as ONGOs – operational NGOs), and those that are involved in policy advocacy around such issues as the environment or human rights (sometimes referred to as ANGOs – advocacy NGOs). Not surprisingly though, it is the former who receive most donor funding.

The other main area of criticism revolves around accountability. As NGOs become more dependent on donors for funding, so they may become more accountable to the donors rather than their beneficiaries. Among larger NGOs in particular, their relationship with donors can potentially undermine their effectiveness in terms of being able to claim or assert independence to the point that they are 'too close for comfort'. There are also issues regarding the conditions imposed on NGOs who receive funding. Often, monitoring of the performance of NGOs involves reporting how money is spent and whether the appropriate activities have been undertaken, but it rarely involves whether NGOs actually make a difference, especially in more intangible realms such as 'participation' and 'empowerment'. These issues are more important to some NGOs than others, for example, those addressing gender issues are more likely to view empowerment as a major measure of success (see **Figure 1**). All these concerns may ultimately serve to undermine the effectiveness of civil society and the neo Tocquevillian goals of promoting democracy and participation so valued by donors. At a broader level, donor attention to civil society through NGOs has also been criticized for failing to address structural issues underlying poverty and inequality by concentrating on the local, as well as to neglect the political dimensions of the functioning of civil society.

Despite these problems, civil society and NGOs still have an important political role in terms of mobilizing people to demand their rights, even if the sector does not command sufficient funds to transform the livelihoods of the poor on a large scale. In turn, while representation may be partial and uneven, at least there is now recognition of the importance of including the views of civil society at all stages of the project cycle, even though this may be implemented in an imperfect manner. One important example here is the role according to civil society in the World Bank's recent Poverty Reduction Strategies. An integral element of Poverty Reduction Strategy Papers (PRSPs) (which outline the formulation of this strategy) is that civil society is consulted and their views included as part of an effort to make them 'country driven' and 'stakeholder led'. Although the success of civil society participation has been criticized as being shallow and insufficient, efforts are being made to try and increase their participation. Another area where strengthening civil society has been important is in post conflict reconstruction. In so called 'post conflict countries' such as Guatemala, El Salvador, and Iraq the active construction of civil society organizations and their inclusion in

Figure 1 Meeting of a women's neighborhood committee, Guatemala City.

Box 1 Donor-supported civil society building: The Iraq Civil Society Program

Since 2004, the US Agency for International Development (USAID) through America's Development Foundation (ADF a nonprofit private voluntary organization dedicated to the international development of democracy) has provided US$59 billion to fund the Iraq Civil Society Program. This program aims to 'to promote an informed, sustainable, and active indigenous Iraqi civil society that effectively and responsibly participates within a democratic system of governance'. In line with a normative, liberal, neo-Tocquevillian approach to civil society, the program proposes to foster participatory governance in Iraq through four key activities:

1. Building civil society resource centers in regional hubs to provide assistance to Iraqi civil society organizations.
2. Targeted technical assistance and training to particular types of civil society organizations dealing with:
   a. Civic education
   b. Human rights
   c. Women's advocacy
   d. Anti-corruption
3. Small grants program to reinforce the work of the resource centers and targeted assistance. These will support activities that 'increase civic activism and promote the role of CSOs in advocacy and policymaking at the national, regional, and local levels'.
4. Support for the independent media through providing technical assistance to the commercial media sector focusing on news and public affairs. This aims to increase access to high quality information throughout the country through providing help with skills development for journalists, media business development, media law advocacy, building of professional associations and industry organizations.

From http://www.usaid.gov/iraq/updates/may06/iraq fs26 050506.pdf; see also http://www.adfusa.org/countries iraq.htm

national level decision making has been seen as central in peacebuilding initiatives (see also **Box 1**).

Civil Society across Space

As noted above, spatial variation is an important element in recognizing the diversity of civil society, prompted largely by the work of geographers. Not only are civil society organizations often distributed in spatially concentrated ways, they are located across the world at different spatial scales comprising a host of different sized organizations. Therefore, civil society organizations can refer to very small scale community based organizations at the grassroots level, to larger regional or nationally based organizations, to umbrella groups comprising several organizations, and to international NGOs. There are also complex spatial interactions between these different levels of civil society activity. In addition to donors, large international NGOs (sometimes referred to as INGOs or BINGOs – big international NGOs), are also increasingly channeling development funding through smaller civil society organizations. This is usually through national level NGOs, which tend to be single issue based (sometimes referred to as NNGOs). Therefore, transnational linkages have been important in influencing the nature of aid flows to NGOs through complex networks and geographies over a range of scales, with flows from larger northern NGOs to smaller southern NGOs being the most significant routes of such flows.

At the international level, there has also been considerable interest in 'global civil society'. Broadly conceived, this refers to global citizen action that operates beyond the limits of national boundaries, politics, and economic concerns. The emergence of global civil society is generally associated with the growth of international NGOs in the 1990s and their growing participation in global decision making in major international conferences (mainly the UN conferences such as the Rio Earth Summit in 1992, the Cairo Population conference in 1994, and the Beijing Women's conference in 1995). As with civil society, global civil society is a normative concept that tends to focus on the importance of dialogue and ethical principles through social change and democratization at the global level. It is also viewed as a space for 'anti politics' and is often associated with the anti globalization movement, and with groups linked with promoting a 'common good' rather than say global terrorist organizations (such as Al Qaeda). However, somewhat contradictorily, on the one hand, global civil society is linked with the propagation of Western and imperial hegemony, while on the other, a site for a radical rethinking of global governance that will provide an alternative to neoliberal globalization. As with civil society more broadly, these views depend on the theoretical viewpoint taken. What is agreed is that global civil society is strongly contested and cannot automatically be assumed to comprise a global democratic polity.

How Useful Is Civil Society?

In light of the highly contested nature of civil society at all scales, it is pertinent to ask whether civil society has any value as a concept and as a way of describing collective social groupings. It is undeniable that civil society has now entered the social science lexicon and that it is here to stay both theoretically and in terms of the policy environment. Indeed, as several commentators have suggested, civil society is more than a buzzword, not least because it facilitates the assessment of the relationship among the individual, society, and the state. However, for it to be useful, it is appropriate to distinguish between civil society as an analytical concept and as a policy tool as noted above. At the conceptual level, it has and will continue to generate very important debate about the nature of democracy and the role of nonstate actors locally and globally. In terms of policy, the integration of civil society into donor discourse has been critical in recognizing a sector of society that had hitherto been ignored. It has given civil society an intellectual and concrete space to have some say in how society is organized and how problems are solved. While certainly, civil society is ambiguous and often imprecise, this can also be a benefit. However, it is essential to recognize the varied intellectual dimensions of the concept and not assume it uniformly operates for the common good as the liberal approach suggests. In turn, civil society must be recognized as diverse, internally variegated, and comprising a host of different and often competing functions and aims. Civil society is dynamic, shifting, and varies across time and space reflecting the heterogeneity of the organizations that form it. Yet, it is also a vibrant arena that is essential in one form or another in all societies of the world, as long as its limitations are taken into account.

See also: Aid; Globalization, Cultural; Globalization, Economic; Governance, Good; Neoliberalism and Development; Nongovernmental Organizations; Participation; Resistance; Rio Summit; Social Capital.

Further Reading

Anheier, H., Glasius, M. and Kaldor, M. (eds.) (2004). *Global Civil Society 2004/5*. London: Sage.

Bebbington, A. (2004). NGOs and uneven development: Geographies of development intervention. *Progress in Human Geography* 28, 725 745.

Bebbington, A. and Bebbington, D. (2001). Development alternatives: Practice, dilemmas and theory. *Area* 33, 7 17.

Edwards, M. (2000). *NGO Rights and Responsibilities: A New Deal for Global Governance*. London: Foreign Policy Centre.

Edwards, M. (2004). *Civil Society*. Cambridge: Polity.

Edwards, M. and Hulme, D. (1996). Too close for comfort? The impact of official aid on non governmental organizations. *World Development* 24, 961 973.

Howell, J. and Pearce, J. (2001). *Civil Society and Development: A Critical Exploration*. London: Lynne Rienner.

Lewis, D. (2002). Civil society in African contexts: Reflections on the usefulness of a concept. *Development and Change* 33, 569 586.

McIlwaine, C. (1998). Contesting civil society: Reflections from El Salvador. *Third World Quarterly* 19, 651 672.

McIlwaine, C. (1998). Civil society and development geography. *Progress in Human Geography* 22, 415 424.

Mercer, C. (2002). NGOs, civil society and democratization: A critical review of the literature. *Progress in Development Studies* 2, 5 22.

Mohan, G. (2002). The disappointments of civil society: The politics of NGO intervention in northern Ghana. *Political Geography* 21, 125 154.

Munck, R. (2007). Global civil society: Royal road or slippery path? *Voluntas* 17, 325 332.

Van Rooy, A. (ed.) (1998). *Civil Society and the Aid Industry*. London: Earthscan.

Relevant Websites

http://www.adfusa.org
America's Development Foundation: Countries, Iraq.

http://www.civicus.org
CIVICUS (World Alliance for Citizen Participation).

http://www.civilsocietybuilding.net
Civil Society Building Network.

http://www.lse.ac.uk
London School of Economics and Political Science, London, UK: The Centre for Civil Society.

http://www.socialwatch.org
Social Watch Network.

http://www.jhu.edu
The Johns Hopkins University, Baltimore, Maryland, US: The Centre for Civil Society.

http://www.un.org
United Nations: Global Issue on the UN Agenda, Civil Society.

http://www.nu.ac.za
University of Kwa Zulu Natal, Durban, South Africa: Centre for Civil Society.

http://www.worldbank.org
World Bank: Civil Society.

Claval, P.

M. Bédard, Université du Québec à Montréal, Montréal, QC, Canada

Claval, Paul (1932–)

Paul Claval, born in Paris in 1932, is one of the most important geographers of his generation, both in terms of the extensive nature of his work and his originality, eclecticism, and international influence. His passion for geography developed during his teenage years in southwestern France. While accompanying his father on summer work trips, Claval very quickly became aware of the effects that distance and information have on the circulation of individuals, goods, and ideas within a given territory. This dual intuition has characterized all his reflections ever since, that is, during his university education in Toulouse, his teaching career in Bordeaux and Montpelier, his professorship at the Université de Besançon from the end of the 1950s to the beginning of the 1970s, his years at the Université de Paris IV (extending until the end of the 1990s), in his later role as professor emeritus.

The originality of his work lies in the fact that it systematically explores human activities and space by way of his chosen explanatory postulates: distance and information. These are the two cornerstones of his thought that he sees as major structuring elements of all geographical thought and yet are strangely absent from its formulations. A typical product of the 1960s, Claval is highly critical of the methodological and ideological errings of a geography that he deems overcautious or too comfortable given that it ignores or underestimates certain important problem areas such as justice, power, the weight of ideologies, identity, and values, not to mention the sociohistoric and paradigmatic ruptures arising from the globalization–internationalization–metropolization triptych and the passage from modernity to postmodernity. Claval is a critic whose work is innovative insofar as it suggests a number of alternative ways to more fully develop geography's cognitive richness and social relevance. This has motivated him not only to constantly explore the history and epistemology of the discipline, but also to open his thought to other disciplines and to foreign geographies, especially geographical currents from the English speaking world, as a means of renewing a geography that would seem to be slow in fully assuming its role as a provider of new frameworks and innovative explanations. In order to reach this objective, Claval criss crossed the fields of economic geography, social geography, political geography, urban geography, and then cultural geography, but he did not practice these disciplines so much as reinvigorate them by offering his own theoretical perspectives, epistemological positions, and reforming concepts as a means of offsetting the shortfalls that he uncovered in them.

Claval's body of work appears to be an ongoing schematic reflection upon *homo geographicus*, designed in the short term around case studies then adjusted over the medium term by the application of economical, social, political, urban, and cultural approaches that over the long term link, by way of feedback loops, his ongoing probing of the destiny and role of geography.

At the outset of his career, Claval observed that French human geography was unable to be anything other than a science of regional differentiation. His conviction that geography is a profoundly social discipline led to probings of an epistemological nature. Claval was well aware of the doubts assailing the geographers of the era and the difficulties they were experiencing in characterizing their discipline; he therefore suggested a historical and critical reflection on geographical knowledge as a possible solution. His goal, following the examples of Harsthorne, Hettner, and Kant, was to illustrate that geography is not a science subject but rather a science point of view, characterized by a unique way of grasping reality, whose ultimate objective is the reconstruction of the spatial aspects of our reality.

To reach his goal, Claval drew on Isard's work concerning central place theory, then on Ullman and Garrison on the city and the opposition between centrifugal and centripetal forces, in order to more fully understand how social life is conditioned by space. Claval's dissatisfaction with the era's essentially empirical and descriptive economic geography, seen as incapable of discovering regularities and explaining constants, led him to explore the roles of microeconomic equilibriums and macroeconomic mechanisms in the genesis and functioning of territorial wholes. For him, the function of transportation fees and information costs proved to be a key factor determining the farthest extent to which space is structured around poles of variable importance. The innovative nature of his approach to economics was demonstrated in his illustration of how territorial constructs are dependent upon their level of economic development and territorial control, that is, their capacity to adopt strategies of localization and modes of spatial organization that can explain both their growth problems and their successes.

Although useful, both in itself and in the service of reform, this approach, like the initiatives of Bunge and Harvey of the same era, proved to be too mechanical, too obsessed by the material aspect of observable phenomena, and by an urgent desire to model in order to assert its scientific nature. And while these colleagues from the English speaking world quickly realized the limits characterizing the positivism from which spatial economics arises and began to practice a more militant and politicized geography inspired by Marxism, Claval turned toward social concepts close to humanism, as well as toward other disciplines. Above all, Claval wanted to see economic geography stop being a geography of places and resources and become a geography of economic subjects and attitudes; this second conviction therefore prompted him to pursue his objective of renewing human geography in an entirely different way.

Disappointed by the inability of the era's geography to demonstrate the class structure characterizing each individual place, Claval explored the social dimensions of geography in conjunction with Gould and Harvey. Drawing upon the works of anthropologists and sociologists, he showed how the behaviour patterns of various players are conditioned by their partners and by the institutionalization of their relationships. In keeping with his explanatory postulates, he identified both the effect of distance on the emergence of class consciousness and the systematization of social relations as a means of illustrating the costs linked to the transfer of information. Once again his approach stood out for its probing of the spatial dimension of power reflected in these classes and relations, as well as the freedom of choice of the individuals that sustain them. Thus, as a complement to his social approach, Claval proceeded to deconstruct the tensions and forces arising with the insertion of humans into space and the relations they forge therein, thus laying the groundwork for a new political geography.

Tired of a born again French political geography that focused exclusively on the state, borders, and capitals, Claval also developed an original political approach through which to understand the facts of power as processes that model space. In fact, he postulated that any system of institutionalized relations serves to channel the effects of power, and he saw power as an essential element of any explanation of geography. Drawing upon the notion of feedback loops derived from cybernetics and systems theory, Claval identified, within an evolutionary spatial framework, the regulating relations linking civil society, the political sphere, and the administrative organization. He was especially concerned with the role of legitimacy in the functioning of any political system and he concentrated on analyzing area and distance as key factors of the freedom–domination dialectic. In addition, by focusing upon these spatial conditions and various forms of freedom in his analysis of power relations, Claval was an innovator, discussing political geography with an emphasis on the individual scale and without dwelling on territory or resorting to maps.

By mobilizing social, political, and institutional dimensions, Claval suggests an ingenious reading of society as well as of the myths of the social sciences and the ideologies that dominate therein, thus enabling a better understanding of the link between the facts of spatial configuration and those of the organization and structure of various societies. His reflections inevitably led him to realize that our social universe is increasingly typified by cities and by a mode of urban life that extends well beyond these particular concentrations. During the 1980s, this focus on the city enabled him to hone his understanding of the influence of space, especially metropolized space, on sociopolitical action.

Having grown weary of an urban geography that showed little interest in the causes and consequences of transportation and frustrated by a theory of central spaces perceived as incapable of explaining the emergence of central space, Claval turned toward various urban planning manuals as a way of better understanding the role of communications and the costs of commuting in the urban organization of space. In addition, inspired by trips to North America and Germany, he sketched a theory of urban networks and another of urban morphology structured around the fluidity and density of these same modes of communication and transportation. Proposing a conception of the city as the ultimate place of convergence, Claval once again made note of the erosion of urban identity within entities no longer belonging to the city or the country as a result of the spreading out and fracturing of their functions. This explains why his urban approach also focuses on the symbolic dimensions of the city, an approach which, while mainly characterizing certain places, enables him to unify urban, regional, and even national space. All things considered, these two theories and their symbolic extension make up an original urbanology treatise according to which the city is the fundamental form of spatial organization in the contemporary world. In the final analysis, destined as it is to facilitate all types of communication, exchange, and recognition, this triple vocation is triggered by metropolization more than it has ever been triggered before.

Wishing to construct an urban approach that is as comprehensive as possible, and observing the phenomenal growth of very large cities, Claval isolated a new dynamic of spatial polarization that he dubbed metropolization as early as 1986. Fleshing out his study of information circuits and communication networks as reflected in the rapid mutations that affect them, he underlines that, in parallel to the growth of central regions and to counterurbanization, development conserves a hierarchical dimension that is now divided among isolated, fragmented urban poles subject to regional or national constraints. Hence, like Giddens and Wallerstein, Claval seeks to interpret the emergence of this new urban scale that circumvents the hierarchies of central places in a world where metropolises, headquarters of global firms, and leaders of worldwide networks, are in direct communication and benefit from excellent accessibility to the economy as a whole. Claval contends that by applying this theory geographers can better grasp the underlying unity of urban concentrations and then more fully understand the processes that challenge the center/periphery dialectic, including those associated with the reemergence of the local and those that transform urban networks through the consecration of certain places where gestation, design, and control activities are concentrated.

This being said, according to Claval these metamorphoses and restructurings of urban settings and ways of life have greatly increased since the passage from modernity to postmodernity, a transition that began in earnest with the fall of the Berlin Wall and accelerated with the globalization of trade, the international division of work, and the increased mobility of individuals and ideas; this passage sounded the death knell for modern ideologies and typical ways of seeing the world and managing its evolution, as reflected in the political, identity, and environmental problems that have shaken the entire planet since the mid 1990s.

The fall of the Berlin Wall and the implosion of the pro Soviet regimes, followed by the extensive regional reconstruction caused by economic globalization, have sparked a questioning of a certain idea of the state, of growth, and of well being in accordance with the emergence of movements characterized by ethnocultural affirmation (nationalism, fundamentalism, etc.) and by sociolibertarian protest (antiglobalization, minority rights, etc.). These are all phenomena that Claval associates with the limits and errors of a way of thinking about our relations with others and the world, a thought process that may be in its death throes. Pursuing a process of reflection of a more clearly epistemological nature, Claval has come to focus on the social and environmental costs of cognitive, ideological, and political regimes that are incapable of providing satisfying solutions. For example, according to Claval, the passage from the popular culture/elite culture opposition to the mass culture/knowledge and technical culture dichotomy is responsible for the current shake up of identities. In fact, this substitution facilitates the eradication of the vernacular components of contemporary societies. And unable to refer to distinct landscapes, techniques, or customs that they may feel to be their own, our societies are completely adrift. This identity crisis is also caused by the collapse of philosophies of history that have dominated for several centuries and that are based on expectations of a better tomorrow, not rooted in a faith based belief in another world.

The problem for Claval, as for Jameson and Fukuyama, is that the privilege accorded to progress, and therefore to time, is no longer valid. And as we confront this void in our horizon of expectations, reflections on postmodernity go forth and multiply. From this postmodern perspective, it is no longer time but rather space that really matters, with individuals aspiring to live in the here and now within frameworks that are sensitive to their aspirations. Given that it is far from certain that this postmodern thinking, characterized by an extreme volatility in its refusal of all dogmatism, can solve anything at all, Claval takes advantage of the current strong social demand for a reform of our relations with our habitat to propose a geographical alternative. More specifically, he proposes a cultural approach that he believes is equipped to link individuals and societies and to give them a happier and more appropriate destiny.

For a long time, Claval came up against the fact that the relations institutionalized by human groups reflect beliefs that they share. As it happens, Claval has intuitively realized that a fundamental change of perspective is currently taking shape, as reflected in present probings and mutations, and even more in geographic research in the English speaking world concerning sense of place and French research on lived in space. This is why since the 1990s he has made every effort to bring out a cultural approach that, he claims, entirely underlies his previous approaches.

Claval addresses our relationships with the world and with other cultures from the perspective of systems of shared beliefs. In this way, he seeks to improve the geographical analysis of cultural realities, which traditionally studies the sole concrete manifestations of cultural dynamics by exploring more specifically that which is above and beyond and elsewhere (otherness, myths, utopias, etc.), elements which help us to better understand the basis of our ways of occupying space. Thus his cultural approach asks in what way individuals perceive the world, invest their passions, knowledge, and interests in it, and then live in it and represent it. And given that this approach focuses on the meaning that these same individuals give to their existence by way of their culture, and especially the way the latter is found in space and materializes it, Claval's cultural geography also insists on the territorial dimension of our ways of thinking about our habitat. By talking about territory rather than space, Claval is able to deepen his reflection on contemporary political, identity, and environmental problems by emphasizing that the places and especially the landscapes within which human existences unfold are constructed as much by actions as by the types of discourse applied to them. By adopting the idea that types of discourse not only articulate but also structure what is real and then reveal it, he argues that geography must consider representations and perceptions since the relationships that link individuals with their environment are as much material as symbolic.

This said, such an interest in values, symbols, and the meaning of place, as well as in territory, landscape, and discourse rather than space, no matter how productive and relevant, is not really innovative since many other geographers have already covered such territory, and for an even longer time (Buttimer, Cosgrove, Relph, and Tuan). What is innovative is not perceiving this in terms of the birth of a new form of cultural geography, but rather of another type of nonlinear causality which, by revisiting the epistemological bases of geography by proposing a dynamic, heuristic, and relational vision of our experience of place, could very well renew its geographicity. In fact, by insisting on the cultural dimensions of geography after its economic, social, political, and urban components, Claval, when all is said and done, advocates a broad and ambitious form of geography which, while simultaneously being concerned about material, social, and idea based realities, seems to him more equipped to prepare humans to be citizens of the world who are ethically committed and environmentally responsible. Such being the case, today he encourages his followers and practitioners to negotiate a paradigm shift in order to revive geography's social relevance and cognitive richness.

As innovative as it may be, Claval's thought would not be so noteworthy were it not for its extensive dissemination. By the end of 2005, Claval had nearly 550 publications to his name, including 40 books, not to mention their reprintings and translations!

For Claval, publishing is a privilege demonstrating respect for the reader; this explains the voluntary simplicity of his thought and the clarity of his writing, even when he addresses complex issues. Paradoxically, this absence of any obtuse convolution or terminological logorrhea may actually hinder the recognition of his scientific contribution. In addition, since out of a sense of rigor he quotes and makes many reference to other writers, some may believe that he has nothing original to say. Furthermore, this feeling is reinforced by the fact that he seems to be participating in what is today a broad based movement of which he was, however, a founder. However, if his thought really is so unimportant, why is he so often quoted, published, and invited to share his observations and chair various committees?

Sensitive to the prevailing winds and deeply convinced of the significant contribution that geography can make to a better understanding of the world, Claval serves first and foremost the discipline and not himself. This explains his insatiable curiosity and his very critical attitude toward geography, as well as his opening to other geographical cultures and disciplines, since he is very concerned with being open to anything that could enrich the geographical explanation of the world. Looking elsewhere whenever necessary for inspiration, Claval transmits and interprets the tendencies of the geography of the English speaking world to Francophone nations. One of his strong suits is, moreover, an ability to illustrate that the geography of the latter is far from being homogenous, whether in terms of ideologies or practices. Another strength is his ability to point out that the geography of the English speaking world has pushed its own transformation to such an extent that it sometimes neglects, for example, the restrictions that the environment or distances impose. As such, according to Claval, Anglophone geography has often become too attentive to what is said about the world and not enough to the influence people have on their environment.

Singificantly, over 125 of Claval's texts have been written in or translated into other languages. This is how he has contributed to raising the profile of French geography outside of France and even inspiring non Francophone researchers to read Francophone geography. Claval has also attempted to reach these same goals through other means, such as when he created the *Espace et Culture* (space and culture) laboratory in 1981, the activities and results of which have been made known since 1992 by way of the journal entitled *Géographie et Cultures* (geography and cultures) of which he was the editor for more than 10 years. Claval is also one of the main architects of the cultural geography and history of geography sections of the International Geographical Union. This far reaching organization, bringing together thousands of geographers with all types of backgrounds, annually organizes a variety of forums for exchange in many parts of the world.

Granted that only time will tell how important Claval really is, he is clearly more than a mere geographical tradesperson. In contrast to Brunet, Frémont, Gottman, Raffestin, and other French speaking geographers who are more easily seen as innovators, Claval is one of the main actors involved in the maturing of contemporary geography. In fact, if one broadly considers his un compromising vision of geography, society, and know ledge, the force of his beliefs and ideas, the foresight inherent in what he has demonstrated, as well as his conceptual and theoretical ideas as regards metropoliza tion mechanisms, the evolution of issues in geography, and the various approaches he has elaborated, etc., it becomes clear that Claval has greatly contributed to opening up geography in terms of ideology and meth odology. Breaking with the archaic nature of much geo graphic endeavor while holding out against doctrinaire abuses, he has been among those who have worked the hardest to point out that geographers trace the contours of space and explore complex territorial and relational matters that no one was aware of before.

Working in the field of geography is for Claval a way of dealing with the challenges that face humanity, exploring the changes in perspective that must take place, and im agining new ways of doing things and talking and thinking about our relationship with the world. This process in volves arming oneself with the necessary means to fully participate in a radical reform of the pact that unites individuals with the planet that sustains them. Also in volved is an attempt to understand what meaning should be given to life in a world that is becoming increasingly emancipated from its traditional perceptions and refer ents. In a sense, Claval's objective is to contribute to this by encouraging people to become aware that the world must be understood geographically and that human beings will only be able to fulfil their potential by ac cepting and pursuing their geographical condition.

See also: Cultural Geography; Francophone Geography; Lowenthal, D.; Tuan, Y.-F.; World/Global Cities; Wright, J. K.

Further Reading

Claval, P. (1963). *Géographie générale des marchés*. Paris: Les Belles Lettres.
Claval, P. (1964). *Essai sur l'évolution de la géographie humaine*. Paris: Les Belles Lettres.
Claval, P. (1968). *Régions, nations, grands espaces: Géographie générale des ensembles territoriaux*. Paris: Marie Thérèse Genin.
Claval, P. (1972). *La pensée géographique: Introduction à son histoire*. Paris: Société d'édition d'enseignement supérieur.
Claval, P. (1978). *Espace et pouvoir*. Paris: Presses universitaires de France.
Claval, P. (1980). *Les mythes fondateurs des sciences sociales*. Paris: Presses universitaires de France.
Claval, P. (1980). Epistemology and the history of geographical thought. *Progress in Human Geography* 4, 371 384.
Claval, P. (1981). *La logique des villes: Essai d'urbanologie*. Paris: Litec.
Claval, P. (1987). Communication theory as a basis for urban theory. In Yadav, C. S. (ed.) *New directions in urban geography*, pp 137 147. New Delhi: Concept Publishing Co.
Claval, P. (1990). A critical review of the centre periphery model as applied in a global context. In Shachar, A. & Oberg, S. (eds.) *The world economy and the spatial organization of power*, pp 13 27. Aldershot: Avebury Press.
Claval, P. (1993). *La géographie au temps de la chute des murs: Essais et études*. Paris: L'Harmattan.
Claval, P. (1996). From a cultural world to a political one. In Douglas, I., Huggett, R. & Robinson, M. (eds.) *Companion encyclopedia of geography. The environment and humankind*, pp 310 326. London: Routledge.
Claval, P. (2001). *Épistémologie de la géographie*. Paris: Nathan.
Claval, P. (2003). *La géographie du XXI[e] siècle*. Paris: L'Harmattan.
Claval, P. (2003). *Géographie culturelle: Une nouvelle approche des sociétés et des milieux*. Paris: A. Colin.

Relevant Websites

http://www.enec.paris sorbonne.fr/revue/num.html
Journal Géographie et Culture.

Climate Change

J. C. Sweeney, National University of Ireland, Maynooth, Ireland

Glossary

Anerobic Not in contact with the atmosphere, typically 'oxygen-free' conditions.

Axial Tilt A cyclical change in the tilt of the Earth's axis. This ranges from 22–25 degrees over a period of about 41 000 years and increases or decreases seasonal differences, especially at higher latitudes.

Carbon Isotope Dating A radiometric dating method based on the decay of the naturally occurring isotope carbon-14 to determine the age of carbonaceous materials.

Crossdating A technique used to match ring-width variations in a tree-ring record with corresponding ring patterns in another. Overlaps allow extension of record into the past.

Eccentricity A cyclical change in the Earth's orbit around the Sun whereby a more elliptical orbit gives way to a more circular orbit, meaning seasonal changes in radiation amounts received at particular latitudes occur.

Global Climate Models (GCMs) Three-dimensional mathematical simulations of the processes that regulate the global climate system. GCMs are used to provide future climate scenarios and to inform policy. A disadvantage is their inability to resolve small-scale processes and the coarse grid which they must use due to intensive computing requirements.

Precession of the Equinoxes A slow gyration of the Earth's axis occurs due to the pull of gravity from the Sun, Moon, and other planets. This produces changes in the date of closest (perihelion) and most distant (aphelion) approaches to the Sun, resulting in seasonal radiation receipt changes at particular latitudes.

Regional Climate Models (RCMs) To overcome the difficulties posed by coarse grid GCMs, RCMs with a higher resolution are constructed for smaller areas. These are driven at their boundaries by a parent GCM.

Younger Dryas A cold period, most pronounced in the North Atlantic, lasting from approximately 12 700 to 11 500 BP. Named after an alpine/tundra flower, this abrupt climate change is believed to have been caused by a shutdown of the North Atlantic thermohaline circulation.

Introduction

Communities which cope with climatic hazards by implicitly assuming a continuance of the *status quo* learn harsh lessons. Despite this, quite limited adaptation to known risks is often apparent in contemporary society. In the past, the catastrophe of harvest failure led to sometimes quite sophisticated coping responses and the shunning of areas of known high risk based on the collective folk memory of the communities concerned. In the present, however, the most typical response has been to develop a climatic fingerprint, usually of three decades or so of climate observations, as the basis for forward planning. Engineers, farmers, and planners extrapolate from this for guidance on the probable range of extreme events which will be experienced during the lifetime of a structure, or over a range of harvests. These are now demonstrably flawed approaches as the nonstationarity of climate time series has become apparent. An inherent characteristic of climate is change, and a period of rapid change is now underway which has the potential to threaten the fabric of society in many parts of the world during the present century. Indeed there are major initiatives emerging at both the highest political levels, as evidenced by international treaties such as Kyoto, as well as bottom up initiatives such as the 'transition towns' which are indicative of growing public sensitivity to the magnitude of the problem faced. A fundamental reassessment of human–environment relations is an integral part of coping with what is viewed as the most severe problem facing the world in the present century.

Evidence from the Past

Instrumental observations provide the optimum means of comparing past and present climates. Invention of key instruments, such as the thermometer (Galileo) and barometer (Torricelli) generally occurred in Renaissance Italy in the late sixteenth and early seventeenth centuries, though records from these times cannot be safely relied upon. The quality of the instruments, especially the glass used, varied widely and the multiplicity of scales and exposures posed additional problems of comparability from place to place and over time. It was the Industrial Revolution, with the need to construct public water schemes to alleviate the unhygienic conditions of the coalfield cities of Western Europe, which provided the main impetus for the development of a more elaborate meteorological network. The demands of the railways for a standard time system in the early years of the nineteenth century also required a more rigorous temporal regime within which comparable observations could be made. Gradually, the observational network was extended overseas by the European maritime colonial powers, albeit with a coastal bias which limited understanding of tropical climates for many decades. Outside of Europe and parts of North America, the length of the observational record is frequently quite short and provides only a limited insight into the climatic conditions before the middle of the twentieth century. Alternative ways of deducing past conditions have therefore been developed.

Documentary Sources

A long tradition of documenting climate extremes exists in some countries such as Iceland, China, and Ireland (**Figure 1**). Similarly, dates on which harvests occurred, or on which flowering of particular plants took place, or when other natural events such as bird migration occurred were often meticulously documented. Collective decisions about when in autumn to pick the grapes for wine making in some of the villages in northern France are recorded back to the fourteenth century, while in Japan the date in spring when the Emperor traveled to Kyoto to view the newly flowered cherry blossoms was similarly recorded. Comparison of when such events take place today with local instrumental observations enables calibration of such documentary sources to be achieved, providing rudimentary reconstruction of growing conditions for much of the past millennium for individual locations. Early newspapers, estate records, and ships' logs during the era of sail also provide useful documentary sources. For example, on ships of the British Navy the log was also maintained three times daily while in port, usually by the Master who, unlike the Captain, often stayed with the same ship for the duration of his

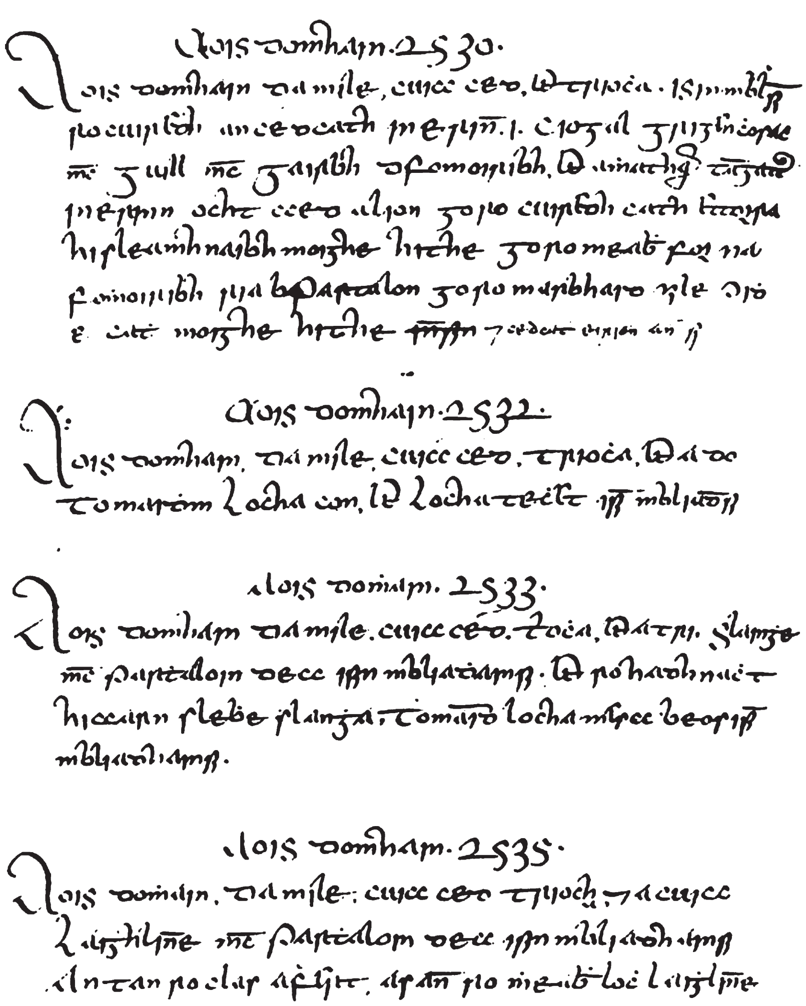

Figure 1 In this old Irish manuscript, reference is made to a storm on Lough Conn in Co. Mayo allegedly in 2668 BC. While the veracity of such documentary sources is questionable, they can provide glimpses of past climatic conditions. Source: Royal Irish Academy.

career. Compilation of such records enables rudimentary weather charts to be constructed and aspects such as storm frequency to be compared with the present day.

Proxy Sources

An indirect deduction of past conditions can be derived by examining relict phenomena sensitive to the climatic conditions which prevailed at the time they were laid down. Some tree species growing in climatically stressed environments produce annual growth rings symptomatic of the conditions prevailing during the season concerned. Calibration of the ring width or the cell density with proximal instrumental measurements of climate for the same period enables a transfer function to be established which can be applied to the rings formed before the observations were available. Moreover, by overlapping ring records from the present with older wood, a tech nique known as crossdating, the record can be pushed back several thousand years. In some cases, individual high magnitude, low frequency events, such as volcanic eruptions, can be detected.

Pollen grains provide a lower resolution record than tree rings, though they enable a better picture of overall environmental changes to be discerned. Pollen pro duction is prolific for some plants, and pollen grains resist decomposition very effectively if deposited in anerobic conditions, such as in a lake bed or peat bog. Since species tend to produce distinctively different pollen types, analysis of a particular layer enables the deduction of the compositional vegetation characteristics at the time the layer was deposited. From the vegetation as semblage, the prevailing climate may be deduced and its fluctuations charted over time by examining sequential layers of pollen. Although this generally only results in relative dating of events, a time calibration is often pos sible using carbon isotope dating of organic deposits such as bones, antlers, etc. also deposited in a particular layer. Closely related to the palynology methodology is cli matic reconstruction based on beetle remains. Again, the tolerance of the species to present day climate is used to infer palaeoclimatic conditions (**Figure 2**).

Deep cores have now been drilled into the icecaps on Greenland and Antarctica. The annual layers of ice here are delimited by slight melting during the short summer periods. While air bubbles trapped in the ice enable the composition of the atmosphere during past periods to be inferred, it is the isotopic composition of the oxygen making up the ice itself which provides the palaeocli matic information. Isotopes of an element occur because of an additional neutron or two in the atomic nucleus, rendering the nucleus slightly heavier than the normally occurring form of the element. In the case of oxygen, about 0.2039% of a typical sample will consist of ^{18}O atoms as opposed to the normal ^{16}O atoms which account

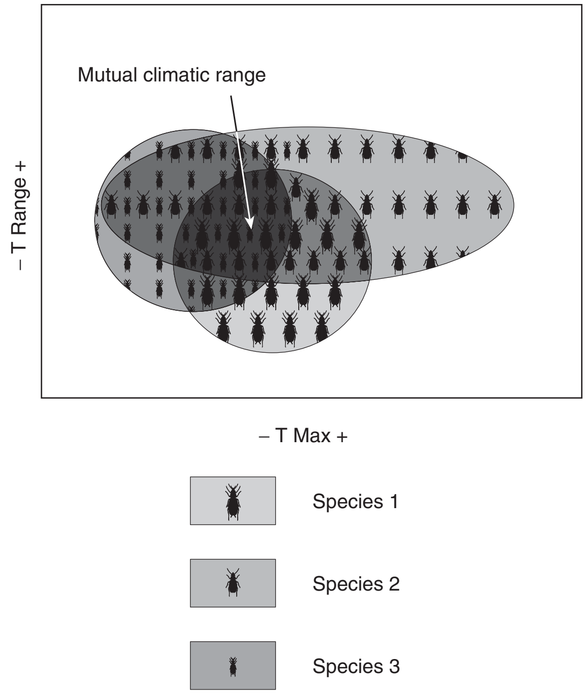

Figure 2 Reconstruction techniques based on known tolerance-of-climate characteristics of beetles.

for almost all of the remainder of the sample. During evaporation, ocean water made up of the lighter ^{16}O atoms enter the vapor phase more readily than their heavier counterparts. If this water vapor then condenses and falls as snow during cold periods it may not recycle back to the oceans for extended time periods resulting in the ocean waters becoming selectively impoverished in ^{16}O and enriched in ^{18}O. Measuring the ratio between these two isotopes in layers of glacial ice thus provides a geological thermometer enabling a snapshot of global temperature conditions at the time the snow was ob tained. Similar methodology can be used in analyzing the microscopic shells found by drilling into undisturbed layers of sediment on the ocean floor. These were built by creatures using the oxygen in the seawater in which they lived. This can yield data on the temperature of the ocean in which they swam as far back as 2 million years ago.

The Record of Climate Change

Climate Change in Prehistory

Reconstruction of climate over the past 2 million years confirms that the Earth then was a glacial planet. For much of that interval, a bimodal climate type prevailed with cold glacial periods lasting 90–100 000 years punc tuated by shorter, warmer interglacials of about 10 000 years. The peak of the last glacial extended from 25 to 18 000 BP and by 14 500 BP rapid warming was underway in the North Atlantic region. Abrupt termination of glacial events occurred with temperature rises of upto 2 °C per decade over large areas. As the Laurentide ice sheet melted, however, iceberg armadas released cold, low salinity and therefore relatively low density water into the Atlantic Ocean, diminishing the vigor of the North Atlantic Drift and triggering a dramatic return to cold conditions for much of Europe. This lasted from 12 700 BP to 11 500 BP, a period known as the Younger Dryas event. After this, abrupt warming resumed and the warmest times of the present interglacial (1–2 °C above mid twentieth century levels) were experienced around 6 000 BP, though the timing of this mid Holocene max imum varies by several millennia for different regions. Generally though, the warmth of the mid Holocene was associated with plant, animal, and human migration poleward. Agriculture became established in Europe and China and cave paintings from the Sahara show hippo potami hunts and cattle herding in areas that today would be considered hyperarid. After the peak of the mid Holocene warmth, drying introduced climatic stresses to many of these places and concentration around reliable water sources such as oases and river valleys seems to have occurred.

A strong correlation between CO_2 concentrations as deduced from analysis of air trapped in the ice layers and temperature as inferred from the isotope analysis is evident, implicating this greenhouse gas as a historic regulator of global temperature. However it is important to emphasize that it is not CO_2 which initiates the temperature rise at the end of a glaciation, and in fact a rise in CO_2 tends to lag several centuries behind the rise in temperature during this event. This is related to a delayed response from the oceanic CO_2 reservoir and, once this occurs, amplification of temperature rises takes place as a result of normal greenhouse gas processes.

Historical Climate Change

Cooler and moister conditions prevailed over much of the eastern Mediterranean in the centuries during which imperial Rome spread its influence across Europe. Gradually, increased frequency of drought conditions further east in Asia spread westward, bringing plagues and mass movements of population in its wake, ultimately to undermine Rome itself. Cooler, stormy conditions set in over much of Europe after 400 AD.

Toward the end of the first millennium AD, tem peratures in Northern Europe appear to have been about 1 °C warmer than the mid twentieth century. This Medieval Warm Period lasted for about three centuries, but was not a global warming event. Harvests were easier, especially in the northern lands where wheat could be grown in northern Norway, oats in Iceland, and barley and rye in Greenland. Viking voyages into the Medi terranean, up the rivers of northern Russia, to the edge of the polar ice, and into the western ocean as far as Greenland and Canada were facilitated by benign cli matic conditions. However, by 1200 AD a downturn was beginning. The ice was returning to the seas around the colonies in Iceland and Greenland and supply routes were becoming increasingly difficult. The Danish par liament debated whether Iceland should be evacuated after a major volcanic eruption in 1783–84 which caused further hardship. By 1540, the last inhabitants of the Greenland colony had died, their demise inevitable as a result of their failure to adapt to the changed climatic conditions as the Inuit had so successfully done.

The Little Ice Age (1450–1850) was also considered previously to have been a period of global cooling, though current views tend to suggest cooling varied greatly over time and space. However, for the Northern Hemisphere as a whole, the Little Ice Age most likely contained the coldest three century period of the entire postglacial record. The Mediterranean froze twice around Marseilles. The Thames froze on at least 20 occasions and slabs of sea ice 5 km in extent appeared in the English Channel. An Eskimo was observed paddling his kayak in the mouth of the River Don near Aberdeen in the early 1700s. Harvests failed and famines occurred in many parts of Europe. In North America, glaciers doubled in extent in the Rockies and moved downslope by 400 m. Recovery from the Little Ice Age has now taken place with a vengeance and furthermore, in contrast to the changes of the last two millennia, the changes of the twentieth century have an undisputed global warming signature.

Natural Causes of Climate Change

External Causes

The Sun delivers 1370 W m ^{2} on a level horizontal surface at the top of the atmosphere. Any variability in this could be expected to cause climate change and it seems probable that on geological timescales, this has been the case. On shorter timescales, small changes in solar irradiance have been detected by satellites over the past three decades. These, however, are too small to ex plain the temperature variations which have occurred over the same period. The best known causes of solar output changes are caused by disturbances on the surface of the Sun which manifest themselves as sunspots. These exhibit a cyclical behavior, typically of 11 years but embedded in longer cycles of 22, 80–90, and around 180 years. Attempts to link sunspots to climate trends have proven inconclusive, though the coldest decades of the Little Ice Age corresponded to a period when sunspot activity declined to almost zero.

Fluctuations in the receipt of solar energy also happen due to three cyclic variations in the Earth's orbit around the Sun. Eccentricity changes occur over a period of 96 000 years whereby the orbit changes from its most elliptical to a near circular mode. This results in upto 25% more radiation arriving at the Earth as a whole when it is at its point of closest approach to the Sun as opposed to 6 months later when it is most distant. The tilt of the Earth's axis also changes over a period of 41 000 years from a minimum of 21.8° to 24.4° and back. Sea sonal climate contrasts are sharpened by a greater tilt. Finally, the elliptical orbit itself slowly rotates changing the time of year when closest or furthest approach to the Sun occurs. This is known as the 'precession of the equinoxes' and occurs on a 22–23 000 year cycle. None of these changes affect the annual total of radiation receipt, though when combined they appear to have the potential to reduce seasonal receipt at high latitudes and have been implicated in causing the glacial/interglacial oscillations of the past 2 million years.

Internal Causes

Volcanic eruptions can eject both particulate matter and gases high into the atmosphere. Where aerosol particles, usually sulphates, get into the stratosphere they may linger in suspension for several years, gradually spreading out and forming a dust veil which has the effect of scattering solar energy and inducing cooling at the sur face. Typically, this cooling is of the order of 0.2–0.5 °C for a year or two after a major event. For maximum global climate impact, a number of factors are required:

- A low latitude location to enable global spread of dust veil
- Explosive power to inject material into the stratosphere
- High sulphur content to generate maximum aerosol sulphate load
- Clustering of major events

Human Causes of Climate Change: Global Warming

A Swedish chemist, Svante Arrhenius, suggested in 1896 that if atmospheric concentrations of the trace gas CO_2 were to double, global temperatures could increase by about 5°C. Though at that time thought of as a somewhat hypothetical possibility, the contention is proving in creasingly prescient. Preindustrial concentrations of about 280 ppm have now risen close to 380 ppm princi pally as a result of anthropogenic emissions. Annual fossil fuel related emissions are currently of the order of 7.2 Gt carbon with a further annual contribution from land use changes such as deforestation of 0.5–2.7 Gt carbon. In creases in other greenhouse gases are also occurring. Methane has already doubled from its preindustrial concentration. About 60% of emissions are related to anthropogenic activities. Nitrous oxide has increased by 13%, most likely related to increased fertilizer use in intensive agriculture. An important consideration for control policies is the residence time of greenhouse gases. While for some, such as CO_2, this is very dependent on which stores atmospheric carbon eventually ends up in, other greenhouse gases such as methane have relatively short residence times. Accordingly, targeting policies on such gases offers better short term reductions in overall greenhouse gas loadings.

Unequivocal evidence from a variety of sources con firms that a major global warming event is underway. Over the last 100 years, global average temperature has increased by 0.74 °C with an acceleration in warming evident over recent decades. The 11 warmest years in the instrumental record have occurred since 1995 (**Figure 3**). Associated with this warming have been changes in other parameters such as ocean temperatures and precipitation. Increases in precipitation have tended to be most pro nounced in the mid latitudes, with decreases more

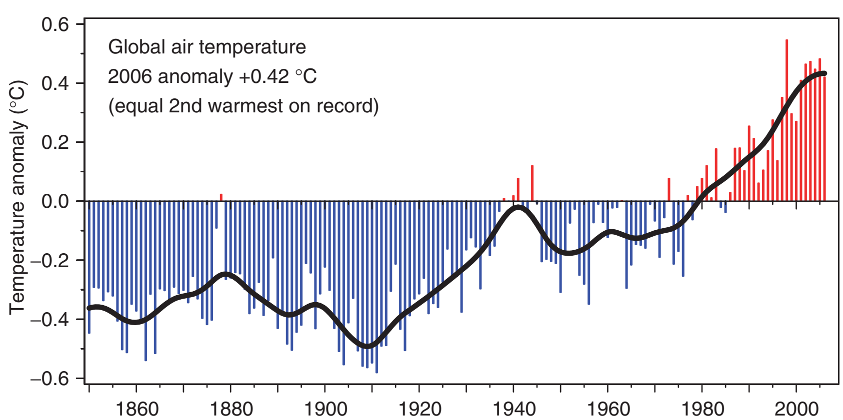

Figure 3 Global temperature trend. From Climatic Research Unit, University of Norwich. Brohan, P., Kennedy, J., Harris, I., Tett, S. F. B. and Jones, P. D. (2006). Uncertainty estimates in regional and global observed temperature changes: A new dataset from 1850. *Journal of Geophysical Research* 111, D12106, doi:10.1029/2005JD006548.

common in parts of the tropics where more intense and longer droughts have also been observed. Many extremes consistent with a warming trend are increasing, such as heat wave frequency, heavy rainfall events, and the in tensity (though not frequency) of tropical cyclones.

The recently published Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC), reflecting a consensus of 2500 climate scientists and endorsed by 130 governments, was the latest in a series of highly influential statements which reflects growing confidence that people are playing an increas ingly dominant role in changing global climate. The First Assessment Report (1990) was tentative: "…the observed increase could be largely due to this natural variability; alternatively this variability and other factors could have offset a still larger human induced greenhouse warming." The Second Assessment Report (1995) was less cautious: "The balance of evidence suggests a discernible human impact on global climate." The Third Assessment Report (2001) boldly stated that: "There is new and stronger evidence that most of the warming observed over the last 50 years is attributable to human activities." This con clusion is now strongly reinforced in the Fourth Assess ment Report (2007): "It is very likely that anthropogenic greenhouse gas increases caused most of the observed increase in globally averaged temperatures since the mid 20th century." In the carefully crafted language of the IPCC, 'very likely' equates to '90% sure'.

Modeling and Future Projections

Global Climate Models

Global climate models (GCMs) have developed rapidly over recent decades in response to major advances in computing power and better understanding of how the climate system works. GCMs work by solving primary equations describing energy, momentum, mass, and water vapor at thousands of grid points on, below, and above the surface. Each equation is solved at a fixed time interval (10–30 min) enabling a progression into the future to be derived from an initial starting set of conditions. Such is the enormous computational resources required that GCMs are only maintained at a small number of facilities around the world. Early generations also simplified many of the inputs. Coupling ocean and atmosphere models was a particularly difficult task because of the latter's much longer response time. Sulphate and nonsulphate aerosols, vegetation and atmospheric chemistry, and better representation of the carbon cycle have all recently been incorporated in most models and a range of input emission scenarios is now employed based on assump tions concerning the socioeconomic evolution of the global community over the next century. Nevertheless, GCMs have not completely overcome the difficulties associated with processes that occur on scales finer than their grid resolution, such as cloud processes. They also exhibit continuing difficulties in handling feedback pro cesses, such as ice albedo, cloud, and water vapor feed backs. Notwithstanding these difficulties, best estimates for global temperature increases for the end of the pre sent century relative to the end of the last, range from 1.8–4.0 °C, depending on the emission scenarios used.

Downscaling

The limitations imposed by computer processing cap acity means that GCM grid sizes of the order of 300 km are inappropriate for environmental managers and pol icymakers. Downscaling of GCM output to a finer mesh resolution has thus become a major research objective, and achievement, of climate scientists over the past decade.

Regional climate models (RCMs) are produced by nesting a secondary model within one or more of the grid spaces of the GCM. Outputs from the parent GCM, such as pressure, wind, temperature, and water vapor, at various altitudes for the area bounding a specified do main of interest, are used to drive the RCM. Within this domain, more spatially detailed output may be produced. Typically RCMs offer resolution of approximately 20–50 km, though even this may be too coarse for many users. In addition, the RCM suffers from any inherent deficiencies in the parent GCM since only a one way influence (GCM → RCM) is allowed. Multiple GCMs and ensemble based approaches are often used to address such uncertainties.

An alternative to regional climate modeling involves a range of statistical techniques designed to downscale from coarse GCM scales. Among these are regression based techniques, various scaling methods and the use of statistical weather generators. The most widespread is statistical downscaling (SD) which depends on strongly linking both the surface variables of interest and GCM output to a large scale atmospheric dataset from which transfer functions may be derived. Usually, upper air data is used to provide this link. Such a technique can provide downscaled output from multiple combinations of GCMs at a spatial resolution useful for guiding decisions which may be affected by future climate change. SD may re quire large amounts of good quality data for calibration and its effectiveness is ultimately dependent on the re liability of the parent GCMs employed.

Impacts

The impacts of climate change will depend not only on the magnitude of change experienced (**Figure 4**) but also on the socioeconomic development trajectories which the world may take over the coming decades. Adaptation

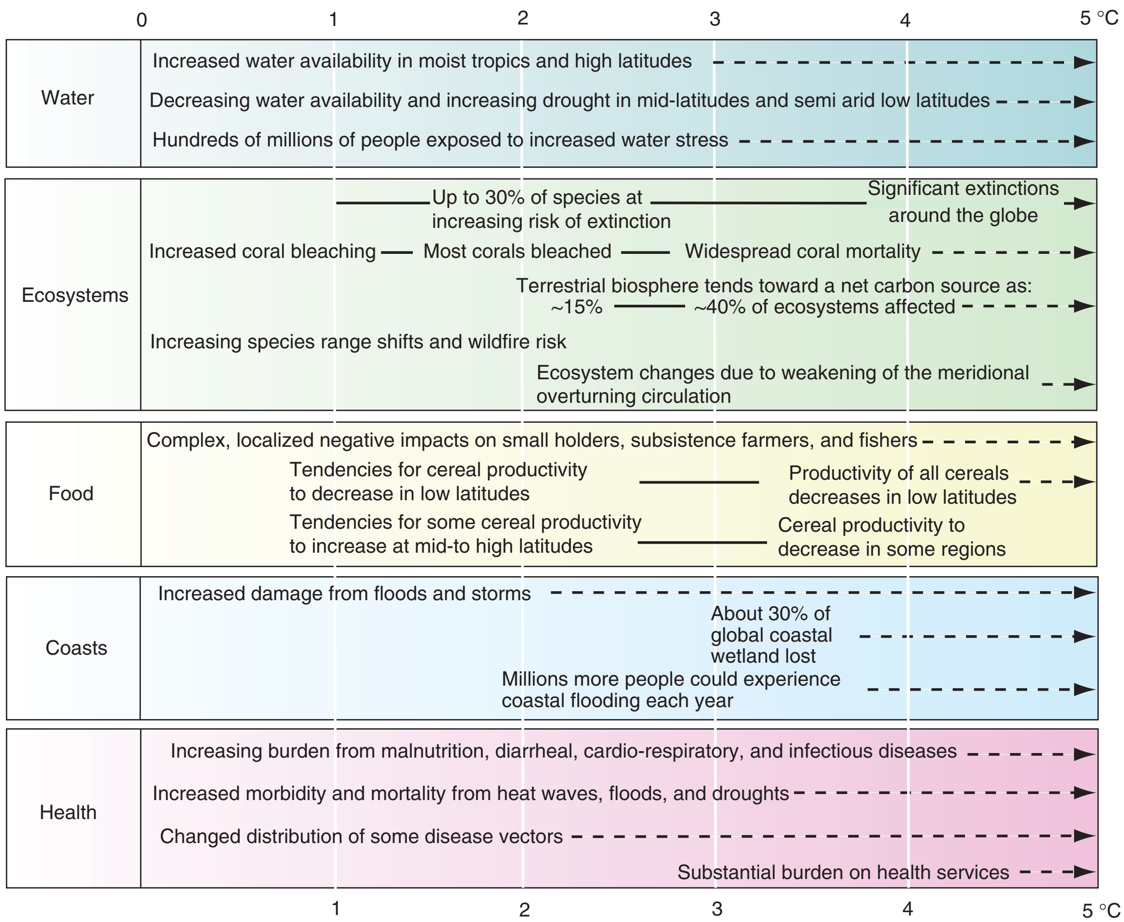

Figure 4 Global mean annual temperature change relative to 1980 99 C. Significant extinctions around the globe is defined here as more than 40%. About 30% of global, coastal wetlands lost is based on average rate of sea level rise of 4.2 mm yr ^{1} from 2000 to 2080. Source: IPCC, 2007.

measures which might mitigate impacts to some degree also require consideration. A cascade of uncertainty commencing with the socioeconomic assumptions and ending with impact assessment is evident (**Figure 5**). However, some confident projections are now emerging. Higher temperatures and reduced rainfall are likely to result in an increase in the extent of water stressed re gions in the years ahead. Mid latitude areas such as the Mediterranean and dry tropical regions such as the Sahel are particularly vulnerable. Notwithstanding this, the incidence of heavy precipitation events is likely to in crease significantly, altering flood frequency and severity in almost all parts of the world. The reductions in glacier fed river flows in the summer melting season are also likely to result in major reductions in water availability in parts of the world highly dependent on such resources.

Impacts on biodiversity are expected to be consider able with 20–30% of species facing extinction if tem perature rises exceed about 2 °C. Additional indirect stresses from fire, pests, and diseases are also likely to be experienced. Food potential may increase for a time at higher latitudes as additional CO_2 concentrations stimulate higher yields. In northern Europe, for example, reduced winter cold is likely to assist agriculture and hydroelectric power production. But for any increase above about 1–3 °C, and indeed for any increase at all on present levels for low latitudes, crop yield potential is likely to fall. Parts of Africa are particularly vulnerable to reductions in the area of land suitable for agriculture as well as reductions in the growing season.

Sea level rise estimates range from 18–38 cm for low emission scenarios to 26–59 cm for high scenarios. For densely populated deltas, particularly in Asia, and low lying islands this poses a major threat and hundreds of millions of people will become vulnerable to flooding, a situation exacerbated where risks are enhanced by tro pical storms or coastal subsidence. Some small islands will find their water resources and tourism compromised, for example, by erosion of beaches and coral bleaching. As with all climate hazards, poor communities will bear the brunt of these impacts.

While in some high middle and high latitude lo cations, warmer winters will reduce annual heat related mortality considerably, increased incidence of heat

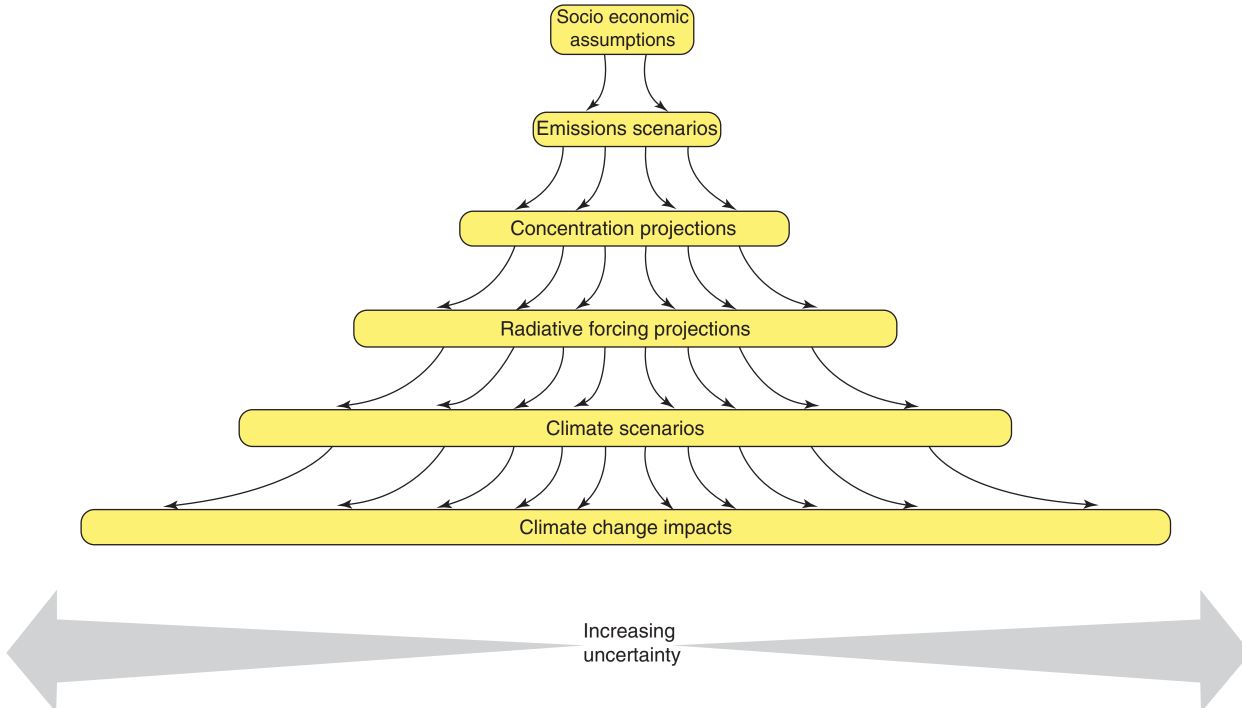

Figure 5 The uncertainty cascade in assessing climate impact.

waves, such as occurred in Europe in 2003, will be associated with increased mortality overall, and especially for the elderly. Changes in disease incidence, such as malaria and water related diseases, can also be expected.

Policy Responses and Adaptation

Vulnerability to climate change is greatest in developing countries with weaker economies and social institutions. People already living on the margins lack the resources necessary to cope adequately with the stresses of climate change. For some, such as the mega deltas of Asia or the low lying islands of the Indian and Pacific Oceans, the effects will be considerable; and yet the problem is not essentially of their making. This mismatch between the primary culprits and the primary victims has bedeviled international attempts to address the problem.

The United Nations Framework Convention on Climate Change (UNFCCC) was agreed upon at the Earth Summit in Rio de Janeiro in 1992. This had as its ultimate objective, the stabilization of greenhouse gases at a level which would prevent dangerous interference with the climate system. The convention recognized the difficulties faced by the developing countries and assigned the lead in combating climate change to the developed countries. More detailed legally binding commitments for industrialized countries emerged in new proposals in Kyoto, Japan in 1997. Under these proposals, the developed countries were required to provide financial resources to the developing countries and to facilitate the transfer of cleaner, less polluting technologies to them. Innovative flexible mechanisms, including trading of CO_2 emission rights, would be employed by the developed countries to reduce their overall greenhouse gas emissions by 5.2% by the period 2008–12 as compared with 1990. Individual targets for each country were set to take account of their particular circumstances. However, despite the Kyoto Protocol coming finally into force in 2005, progress has been slow. The USA withdrew its commitment to ratify the protocol in March 1991 and other nations such as Australia have also since refused to ratify. This, together with the fact that major developing countries such as China and India remain immune from any emission reduction requirements has limited the likely effectiveness of a treaty which was always going to be a small first step in tackling the problem. The 27 countries of the European Union (EU) have always been the main champions of emission reduction targets and have interpreted the ‘dangerous climate change’ clause of the UNFCCC to mean a global temperature increase of 2 °C. In 2007, the EU agreed a further unilateral set of cuts in their own emissions. These amount to 20% reductions on the 1990 levels by 2020. Thirty percent is on offer if the other developed countries currently not participating in the Kyoto Treaty commit to specific emission reductions as part of a new post Kyoto arrangement. However the effectiveness of any international agreement which does not involve some reductions from the most rapidly growing developing world economies is questionable. China, for example, is expected to surpass the USA as the highest emitter of greenhouse gases by as early as 2009.

See also: Environment; Environmental Policy; Environmental Regulation; Environmental Studies and Human Geography.

Further Reading

Brohan, P., Kennedy, J., Harris, I., Tett, S. F. B. and Jones, P. D. (2006). Uncertainty estimates in regional and global observed temperature changes: A new dataset from 1850. *Journal of Geophysical Research* 111, D12106. doi:10.1029/2005JD006548.

Intergovernmental Panel on Climate Change (2007). Climate change 2007: Impacts, adaptation and vulnerability. *Report of Working Group II of the Fourth Assessment Report of the IPCC.* Cambridge: Cambridge University Press.

Intergovernmental Panel on Climate Change (2007). Climate change 2007: Mitigation of climate change. *Report of Working Group III of the Fourth Assessment Report of the IPCC.* Cambridge: Cambridge University Press.

Intergovernmental Panel on Climate Change (2007). Climate change 2007: The physical science basis. *Report of Working Group 1 of the Fourth Assessment Report of the IPCC.* Cambridge: Cambridge University Press.

Relevant Websites

http://ec.europa.eu/environment/climat/eccp.htm
European Climate Change Programme.

http://www.ipcc.ch/index.htm
Intergovernmental Panel on Climate Change.

http://unfccc.int/2860.php
United Nations Framework Convention on Climate Change.

Cloke, P.

M. Phillips, University of Leicester, Leicester, UK

Glossary

Agape One of several Greek words for love. Often used in Christian writings to refer to selfless love, or a sense of love or care which extends beyond the self to encompass a sense of acting for others.

Applied Geography An approach to geography that values the application of geographical ideas, knowledge, and/or skills in social arenas beyond the academic.

Autoethnography An account of a group, situation, or place written by a member of this group, situation, or place, rather than by a person from outside who even while perhaps seeking to become a 'participant observer' will eventually exit the group, situation, or place. With autoethnography the researcher is not trying to become an insider in the research setting but is already so positioned.

Index of Rurality A measurement of how rural, and by implication how urban, a place is. While often portrayed as an objective indicator and constructed through statistical methods, in practice may be seen as an attempt to create formal representations which correspond with preestablished views of the character of particular spaces.

Key Settlements A term which originated in planning discourses in the early twentieth century to signify settlements which performed important functions not only for their own inhabitants but also for people from surrounding settlements. It was utilized both as a means of 'rationalizing' service provision in areas of rural depopulation and as means of restricting service and residential growth in rural settlements.

Logical Positivism A philosophy that emerged initially in the 1920s but was incorporated into many parts of geography as part of the so-called 'scientific' or 'quantitative' revolution.

Others and Otherness These terms came to prominence through human geography's engagement with ideas of postmodernism which emphasized, among other things, the degree to which people, events, practices, and so on, differ from one another. Attention was drawn to how geographers had tended to neglect such differences, often assuming that people were very much like, or the same, as they were.

Political Economy Approaches in which emphasis is placed on the political dimensions of knowledge production, with all knowledge both being value laden and exerting social effects, and on human societies being historically and spatially specific productions. Within geography there has been a particular focus on capitalist forms of production.

Cloke, Paul (1956 –)

Professor of Human Geography at the University of Exeter, England, having previously worked in geography departments at Lampeter, Wales, and Bristol, England; by the early 1990s Martin Bodman identified Paul Cloke as an emerging 'master weaver' in human geography, a phrase used to characterize people having a significant influence within the discipline. Bodman argued that Cloke was one of a group of human geographers emerging as major influences in the second half of the 1980s, linked in large part to the rise of political economy and social/cultural perspectives. Cloke has certainly acted as a spokesperson for such perspectives, although it should be noted that he was publishing highly influential studies from the late 1970s, the majority of which were, by his own admission, set in quite different traditions of human geographical research.

Paul Cloke has suggested, in his 'autoethnography; published in 1994' that his student experiences and early years as a lecturer in human geography were very much conditioned, "almost surreptitiously," by prevailing orthodoxies present in the period and places he inhabited. He argues, for instance, that in his undergraduate degree he was "introduced persuasively to the notion of progress through planning to a better, more efficient world." As he later reflected, this modernistic viewpoint was normalized within his geographical education, and indeed was of considerable significance in his PhD and related research on rural planning, as demonstrated in work such as *Key Settlements in Rural Areas* and *An Introduction to Rural Planning* which had clear applied and normative elements. These emphases were not only present with the work of his teachers and supervisors, but were also evident more widely in the 1970s and early 1980s when notions of 'applied geography' and 'applied rural studies' were frequently advanced. These approaches emphasized linking academic work with practical applications, commonly advancing these arguments in conjunction with 'scientistic' understandings of geography and logical positivist epistemologies and ontologies. These features can all be identified within Cloke's early writings, which he has argued were prompted and legitimatized by the then prevailing culture of social science which valued the employment of statistical methods. Cloke, who is now widely associated with the use of qualitative research methods, has expressed some reservations about his early research, although one of its products – the 'index of rurality' – remains widely cited and emulated.

During the course of 1980s, Paul Cloke can be seen to have become increasingly critical of the positivism and modernism in which he had been immersed. Books such as *Rural Planning: Policy into Action?* and *The Rural State?* reveal an increasing questioning of the effectiveness and impartiality of state activities such as planning, which was linked to adoption of a political economy perspective. A clear expression of this position was given in Cloke's 1989 contribution to New Models in Geography, where at the same time the positivism of rural studies is challenged and the benefits of applying a political economy per spective responsed. Among the themes identified as being of key significance to rural geographers was that of the relation between capital accumulation/restructuring and state/society relations. The late 1980s and 1990s saw Cloke pursue these issues through research on the iter ation of economic restructuring and social class recom position in rural areas, the role of planning within a capitalist state, the privatization/de regulation of the state, the commoditization of the countryside, and social marginalization and exclusion.

Paul Cloke has explained that although he was at this time establishing "forthright political economic ac counts," he actually had a number of concerns about them, including their apparent undermining of rurality as an analytical concept; economic and political re ductionism; and the conflation of political economy with Marxism**.** These concerns surfaced in research during the early to mid 1990s which, in various ways, sought to knit political economy together with recognition of the cul tural and experiential dimensions of rural spaces, a pro ject which he characterizes as an 'enculturing political economy'.

Cloke's endeavors can be seen as part of, and indeed contributing to, more general 'cultural turns' in rural and human geography, as well as a response to personal life/ research experiences. In respect to his research, his turn toward culture was undertaken in large part through continuation and transformation of work influenced by political economy: economic restructuring, for instance, became linked to and modified by concepts drawn from regulation theory, which in turn was employed in studies of social marginalization and poverty; this work, as well as the studies of social class and commodification, be came focused on symbolic representations, as indeed did research on social marginalization and exclusion. This last strand of work also evidenced a clear desire to give voice to experiences, concerns, and identities of rural people, particularly those experiencing some form of marginalization. In making such arguments, Cloke's work resonated quite directly with postmodern ideas, par ticularly as advanced in rural studies in relation to the study of 'neglected rural others'.

The movement from a political economy to a post modern perspective has not been universally welcomed and, indeed, Cloke has repeatedly withdrawn from complete embrace of postmodernism, expressing unease about its political and policy implications; the degree to which it adopts a moral understanding which views itself as being above social interest; the implicit privileging of certain lines of difference and identity over others; and superficial and/or inappropriate engagements with re search subjects. Such reservations were by no means unique in geography, although Paul Cloke has come to address these issues in a varied and distinctive manner.

A major strand of Cloke's work since the late 1990s has been further investigation of the experiential and cultural dimensions of social marginalization in rural areas. This has been pursued primarily through research on homelessness which has highlighted the presence yet apparent invisibility of rural homelessness. The research addresses many of the problematics identified in Cloke's enculturing of political economy, although a new concern with ethics emerged both in relation to the conduct of this particular research and with regard to research in human geography more generally. Cloke has differen tiated between human geographies '*of* the other' and '*for* the other', suggesting in an essay entitled 'Deliver us from evil?' that while it may be relatively easy to identify work which expresses an abstract intellectual fascination with others and otherness, "it is far more difficult to discover in contemporary human geography as a whole a sense for the other which is emotional, connected and committed." The former human geography exhibited the problematic features of postmodernism identified previously, while the latter, he suggested, "may necessitate a prompting of the moral imagination which includes political, ethical/ spiritual constellations of issues such as charity and agape, and evil."

Use of terms such as 'spirituality', 'agape', and 'evil' reveal a recasting of many concerns previously expressed in political economic terms into a language of ethics. In part, this can be seen as an acknowledgement of his long standing Christian faith, whose influence throughout his work is discussed in his autoethnography, although he states that his argument is not that of "the proselytiser" but rather that geographers from all manner of faiths and perspectives need to engage with such issues. He has highlighted similarities between secular social theory and Christian concepts such as charity and agape, as well as between the ethos of care for the homeless among Christian and non Christian charitable organizations. He has also conducted work on 'ethical consumption'. This work, as well as his most recent work on homelessness, has transcended his earlier focus on rural space. This

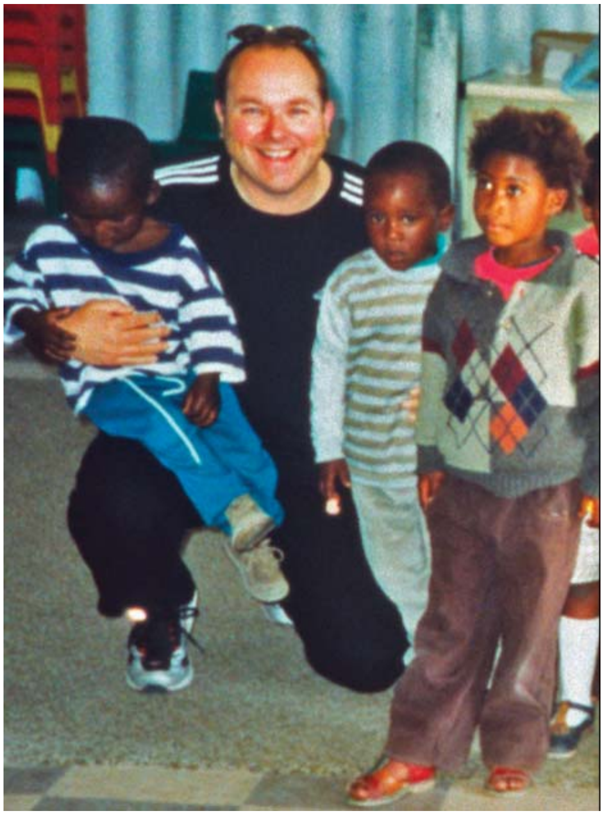

Figure 1 Paul Cloke.

feature is also evident in research on the agency of nature which has emerged since the mid 1990s. This research has not only made links to questions of ethics but also drawn extensively on post structuralist ideas of relational agency and performativity, as well as notions of socio natural hybridity and non human actants drawn from actor network theory.

As has been highlighted within many disciplinary histories, insight and influence do not necessarily proceed through research writings alone and can advance through a variety of means. As noted earlier, in his auto ethnography, Cloke articulated a variety of personal and contextual influences on his research, and these and new influences have certainly contributed to subsequent de velopments in his work. For instance, Paul Cloke has illustrated interconnections between his professional and personal life expressed through his engagement with Khayelitsha township, in South Africa. While connected closely with his academic work on ethics and otherness, and with his reading of philosophical and theological discussion of Christian concepts such as charity and agape, Cloke's account also reveals the significance of personal friendships and how his love of playing music and soccer informed his engagement with people in Khayelitsha (**Figure 1**).

In terms of his academic influence, it is important to recognize that since 1985 Paul Cloke has been the founding editor of the *Journal of Rural Studies*, which has been an extremely significant agent in the articulation of international rural research from a plurality of per spectives. Another important source of academic influ ence is undoubtedly through teaching, both in person and through the production of teaching texts. He has been influential through both routes, having produced a series of highly popular textbooks and, as a series of students would testify, a popular and inspiring teacher. His enthusiasm and passion is infectious and he also finds time to interact with people in a very personal way. Furthermore, just as Cloke himself has identified a range of influential collaborations and personal interactions that have influenced the development of his work, so many contemporary geographers would acknowledge that their work has been influenced by formal and in formal interaction with him. This is certainly the case for this author.

See also: Actor-Network Theory/Network Geographies; Applied Geography; Autoethnography; Christian Geography; Cultural Turn; Gentrification, Rural; Homelessness, Rural; Housing, Rural; Identity and Otherness, Rural; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Modernity; Other/ Otherness; Postmodernism/Postmodern Geography; Radical Political Economy; Rural Geography; Rural Populations.

Further Reading

Barnett, C., Cloke, P., Clarke, N. and Malpass, A. (2005). Consuming ethics: Articulating the subjects and spaces of ethical consumption. *Antipode* 37, 23 45.

Bodmin, A. (1991). Weavers of influence: The structure of contemporary geographic research. *Transactions of the Institute of British Geographers* 16, 21 37.

Cloke, P. (1977). An index of rurality for England and Wales. *Regional Studies* 11, 31 46.

Cloke, P. (1979). *Key Settlements in Rural Areas*. London: Methuen.

Cloke, P. (1983). *An Introduction to Rural Settlement Planning*. London: Methuen.

Cloke, P. (1987). *Rural Planning: Policy into Action?* London: Harper and Row.

Cloke, P. (1989). Rural geography and political economy. In Peet, R. & Thrift, N. (eds.) *New Models in Geography: The Political Economy Perspective*, vol. 1, pp 164 197. London: Unwin Hyman.

Cloke, P. (1994). (En)culturing political economy: A life in the day of a 'rural geographer'. In Cloke, P., Doel, M., Matless, D., Phillips, M. & Thrift, N. (eds.) *Writing the Rural: Five Cultural Geographies*, pp 149 190. London: Paul Chapman.

Cloke, P. (2004). Deliver us from evil? Prospects for living ethically and acting politically in human geography. In Cloke, P. (ed.) *Envisioning Human Geography*, pp 210 228. London: Edward Arnold.

Cloke, P. (2004). Exploring boundaries of professional/personal practice and action: Being and becoming in Khayelitsha Township, Cape Town. In Fuller, D. & Kitchin, R. (eds.) *Radical Theory/Critical Praxis: Making a Difference Beyond the Academy?* pp 92 102. Kelowna: Praxis (e)Press.

Cloke, P., Cook, I., Crang, P., Goodwin, M., Painter, J. and Philo, C. (2004). *Practising Human Geography*. London: Sage.

Cloke, P., Milbourne, P. and Thomas, C. (1997). Living lives in different ways? Deprivation, marginalization and changing lifestyles in rural England. *Transactions, Institute of British Geographers* 22, 210 230.
Cloke, P. and Thrift, N. (1987). Intra class conflict in rural areas. *Journal of Rural Studies* 3, 321 333.
Jones, O. and Cloke, P. (2002). *Tree Cultures: The Place of Trees and Trees in Their Place*. Oxford: Berg.

Relevant Websites

www.elsevier.com/wps/find/journaldescription.cws home/348/description#description
Journal of Rural Studies
http://sogaer.exeter.ac.uk/geography/people/staff/p cloke/main.shtml
Paul Cloke homepage

Cognitive Geography

D. R. Montello, University of California – Santa Barbara, Santa Barbara, CA, USA

Glossary

Affect Emotional component of psychological states including mood, stress, preference, and attitude.

Cognitive Map Spatial and nonspatial knowledge about places stored in the mind.

Cognitive Regions Regions in the mind, used by people informally to organize their understanding of the Earth surface.

Dead Reckoning Updating orientation by calculating or inferring a new location or heading based on knowledge about movement speed and direction from a known starting point, without recognition of specific features.

Hierarchical Spatial Reasoning Reasoning about spatial relationships on the basis of superordinate relations in a hierarchy of places and regions.

Landmarks Distinct objects or features in the world that can be noticed and remembered, and serve to organize experience and memory, guide action, and provide a basis for communicating about the world.

Route Knowledge Knowledge of linear sequences of landmarks connected by travel patterns, ordered but not necessarily metrically scaled.

Survey Knowledge Knowledge of two-dimensional layout from which spatial relations among places can be determined even if direct travel between them has never occurred.

What Is Cognitive Geography?

Cognitive geography is the study of cognition, primarily human cognition, about space, place, and environment. Cognition is knowledge and knowing by sentient entities, including humans, nonhuman animals, and artificially intelligent machines. Cognitive structures and processes include those of sensation, perception, thinking, learning, memory, attention, imagination, conceptualization, language, and reasoning and problem solving. Some of these structures and processes are consciously accessible, potentially available to awareness; others are nonconscious, outside of awareness. Cognition is functionally and experientially intertwined with affect, motivation, and behavior. Our beliefs and knowledge influence, and are influenced by, what we feel and what we do.

Cognitive geography emerged as an approach within human geography and as an interdisciplinary link with psychology and other fields during the 1960s but reflected strands of inquiry from at least as early as the beginning of the twentieth century. Those strands included an interest in understanding and improving spatial orientation and disorientation, geographic education, map design, urban planning and landscape design, and models of spatial behavior and interaction, including travel, communication, and economic activity. The study of cognition is thus a concern for geographers because it involves the fundamental geographic issues of space, place, and environment. Cognitive research not only holds the promise of improving noncognitive models of human activity but also includes problems that are part of the domain of geography in their own right.

Cognitive geography originated as a component of the behavioral approach in human geography; it thus shares much of the philosophical character of behavioral geography. The behavioral approach is the view that we can understand much about human geography by studying it at the disaggregate level of analysis – at the level of the individual person. As such, behavioral geographers examine data on the behavior of individuals, allowing for the likelihood that individuals vary from one another because of factors such as their intellectual abilities, gender, education, and culture. Behavioral geographers hold that models of human activity and interaction can be improved by incorporating into them more realistic assumptions about human behavior. As a case in point, the assumption of economic rationality holds that economic agents are motivated solely to maximize profit, and have complete and accurate knowledge of profit relevant information. This assumption is usually quite unrealistic, especially when studying economic agents that are individual people. Behavioral geographers argue that assumptions about human psychology like that of economic rationality can be replaced with assumptions that are more realistic. For example, gravity models hold that spatial interaction declines as a function of increasing distance, raised to the power of some exponent. Behavioral geographers assert that the distance term should properly refer to cognitive distance – distance as believed – rather than actual physical distance (or cognitive travel time rather than actual travel time, and so on).

Although signs of interest in cognition were present in the discipline of geography before the emergence of the behavioral approach (as suggested above), attempts to study human geography at the disaggregate level naturally led behavioral researchers to regard what the individual knows or believes about the world as playing an important role in explaining what the individual does or will do. That is, a key assumption of cognitive geography arising from its genesis within the behavioral approach is that people do what they do, at least to an important degree, because of what they think is true. People evaluate alternative decision options according to their beliefs, in order to make behavioral choices in space and place. What people think, in turn, arises from perceptual knowledge acquired via the senses, as organized and interpreted by existing beliefs and schematic knowledge structures and processes. These, in turn, are products of people's genetic and experiential histories.

Cognitive geography differs from other approaches to human geography, such as environmental geography, social physics, classical economic geography, and various forms of critical geography. It differs in its topical focus, its preferred methodologies and epistemological assumptions, and its basic conceptualization of the human–earth relationship – its version of a human geographic ontology. Four key tenets of cognitive geography express its topical focus, and its epistemological and ontological assumptions:

1. An appropriate level of analysis is disaggregate. The individual person is an informative unit of analysis (disaggregate or microscopic level) over and above the social or spatial group (aggregate or macroscopic level). Disaggregate analysis allows for – even insists on – possible individual variation based on the genetic and experiential variation of individuals. However, cognitive geography does not assume that this variation is completely random or idiographic; individual variation is assumed to exist within causally patterned networks of interrelatedness. That is, individual variation can systematically explain, and be explained by, other constructs to which it relates. Individual variation is typically studied in terms of constructs that characterize groups of individuals and are potentially causally relevant, including sex, gender, age, education, culture, residential environment, activity patterns, and social class. Thus, cognitive geographers believe it is valid to study individual variation scientifically, that is, according to systematic empirical observation and logically consistent theorizing that is potentially repeatable, accumulable, and generalizable. Of course, patterns of correlations between measures of group constructs and measures of geographic behavior (and cognition) are, by themselves, ambiguous with respect to causal explanation. The group construct could cause variations in the behavior, the behavior could cause variations in the group construct, or another construct or set of constructs that may not even have been measured could cause variations in behavior.
2. Behavior is based on subjective or perceived reality. Affect and behavior are based directly on subjective, or perceived, realities (plural because of individual variation). Behavior is based only indirectly and approximately on objective reality. Much of the indirect effects of objective reality are mediated by subjective reality, because subjective reality itself is directly based on one's experience of objective reality, which varies somewhat across individuals. Cognitive geographers recognize that subjective realities typically deviate from objective reality, sometimes greatly. The internal mental structures and processes believed to mediate observed behavior are not necessarily available to conscious awareness. Much behavior in fact depends on cognitive structures and processes that are not consciously accessible, such as when we respond favorably to a particular place for reasons we do not remember or understand. Furthermore, we sometimes delude ourselves about the reasons for our actions; we may provide explicit explanations for our actions that are actually personal theories or rationalizations, because we sometimes have no direct conscious access to our real motivations and do not feel comfortable accepting this lack of awareness.
3. Human–environment relations are dynamic and bidirectional. Neither environmental determinism nor its opposite, cultural determinism or autonomy, are viable theoretical frameworks for understanding human geography. Human–environment relations are bidirectional, insofar as the actions and cognitions of individuals both cause, and are caused by physical and social environments. Furthermore, these relations are dynamic, constantly emerging and subsiding, though remaining relatively stable over sufficient time periods to justify scientific study. Within the context of cognition, the notion of dynamic and bidirectional human–environment relations is reflected in the notion that humans are active gatherers and processors of information, not simply passive recipients.
4. Mind emerges from brain and nervous system, in a body that is in a physical and social world. Cognitive geographers, like psychologists and other cognitive scientists, recognize that mind depends on the complex organ of the brain and the rest of the nervous system. Following the lead of other cognitive scientists, cognitive geographers in the early twenty first century are beginning to show interest in cognitive neuroscience, the scientific discipline that studies mind–brain relations. Much recent interest in cognitive neuroscience on the part of geographers and others is fueled by excitement over the relatively new technologies of brain scanning, especially functional magnetic resonance imaging (fMRI). But cognitive geographers do not equate mind with brain; mind does emerge from brain, but it also emerges from the human body, existing in a social and physical world. That is, mind is embodied and situated. Attempting to

understand cognitive geography solely by reducing it to the brain would be like trying to understand glaciers solely by reducing them to hydrogen and oxygen. That is, cognitive geography is not reductionistic, although it accepts that reductionistic analysis contributes to a comprehensive understanding of geography. Although cognitive geographers could once be criticized for treating mind too individualistically, they increasingly recognize that mind emerges and functions within a social and cultural context. Similarly, the critique that cognitive theory leaves a person frozen in daydreams or pondering over alternatives has largely been supplanted by ecological and functional approaches to cognition that reconnect the knowing person to action and the physical world.

Research Topics in Cognitive Geography

As another approach to explaining cognitive geography, this section briefly reviews some of its research topics, including some specific concepts, theories, research questions, and empirical findings. Research in cognitive geography covers a diverse array of topics, as it includes all aspects of human cognition about spatial, temporal, and thematic aspects of the Earth and external representations of the Earth. The rest of this article focuses on research in four topical areas: geographic knowledge and reasoning, navigation and orientation, learning and development, and symbolic sources of geographic information. Other research topics in cognitive geography include spatial choice, artificial intelligence, environmental preference and esthetics, hazard perception, and individual and group differences in geographic cognition.

Geographic Knowledge and Reasoning

Central to cognitive geography is the idea that people acquire geographic knowledge about the spatial and nonspatial properties of the Earth surface, and the natural and human features and processes found there. This knowledge is encoded in the nervous system in the form of patterned internal, or mental, representations. These representations are variously referred to as cognitive (or mental) maps, mental models, memory images, schemas, concepts, and so on. Geographic knowledge also exists as symbolic artifacts – patterned external representations. Examples include cartographic maps, verbal descriptions, numerical equations, and so on. Whether internal or external, geographic knowledge includes knowledge of distances, directions, connectivity, and other spatial properties of varying geometric sophistication; it also includes nonspatial knowledge of temporal and thematic properties, such as storm frequency, soil fertility, city populations, and so on.

The term cognitive map suggests a metaphorical similarity between geographic knowledge and cartographic maps. This metaphor is apt insofar as both cognitive and cartographic maps are representations, contain spatial and nonspatial information, are selective, distort and schematize spatial and nonspatial properties, can encode information via different perspectives, represent on a continuum of abstractness, and have a variety of functions. The metaphor misleads insofar as cognitive maps are not unitary and uniform representations, as is suggested by the spurious idea that they are mental pictures in the mind. Cognitive maps are more like atlases or collages, in that they consist of multiple representations that are not necessarily mutually coordinated or consistent, and do not have constant, or constantly varying, scales. Furthermore, cognitive maps are like multimedia presentations in that they are derived from multiple sources of information and encoded in a variety of formats, not just graphical or pictorial. Cognitive maps do not necessarily encode Euclidean metric spatial relations or any other metric relations. Much research in cognitive geography and related fields has documented various patterns of distorted distances, directions, and other spatial relations that are evident in behavioral data based on cognitive maps.

Cognitive maps are thought to be composed of characteristic elements, including landmarks, paths, regions, and boundaries. Landmarks are distinct objects or features in the world that can be noticed and remembered. They serve to organize experience and memory, guide action such as navigation, and provide a basis for communicating about the world. Although they are often thought of as point like features, linear and areal features such as paths and regions can also serve as landmarks. Although one often thinks of landmarks as discrete features, such features often act as landmarks only within a context of their location or surrounding environment.

Regions are approximately two dimensional elements, spatial categories of Earth surface. Just as geographers use regions to organize the Earth surface, and institutions use regions to organize administration of the Earth surface, lay people use regions to organize their cognition of the Earth surface. Cognitive, or perceptual, regions are regions in the mind. People informally regionalize on the basis of physical boundaries and other discontinuities in the structure or appearance of the Earth surface, but they also regionalize on the basis of activities, residential and cultural patterns, past events, and other factors that may not be perceptibly present on the Earth surface. Geographers and others have further noted that people organize regions into full or partial hierarchies; hierarchical structures link regions at different levels of status (size, functional importance) according to relationships of containment or connectivity. For example, countries contain provinces. Hierarchical spatial reasoning occurs when people reason about spatial relationships on the basis of superordinate, rather than direct, relations among regions and places. Regional organization and reasoning, including that which is hierarchical, can be inferred from patterns of judgments people make in recalling places and their relations, and in estimating distances and directions (**Figure 1**); patterns of times people require to make such judgments are also telling about knowledge organization.

Navigation and Orientation

Navigation is coordinated and goal directed movement of one's self through the environment, where the goal is a place or location one strives to reach. It is an essential process involved in the temporary and permanent travel behavior that is of great interest to human geographers. A practical reason to study navigation is to understand cognitive aspects of accessibility, such as what information is needed by different classes of travelers, including older adults and those with sensory–motor disabilities.

Navigation may be understood to consist of the two broad components of locomotion and wayfinding. Loco motion is the coordinated movement part of navigation, wherein we walk or drive while maintaining a directional heading, avoiding barriers, and so on. Locomotion is coordinated to the local surrounds – the environment that is directly accessible to one's sensory and motor systems. If a person controls his or her own locomotion, as when driving a car, it is considered active locomotion; otherwise, as when riding in a car, it is considered passive. Wayfinding is the efficient goal directed planning and decision making part of navigation; it guides the loco moting person to a destination. It includes planning for travel and decision making during travel, including route choice and orientation to nonperceptible features. Way finding is coordinated to the distal environment – the environment that is not directly accessible to one's sensory and motor systems but is accessed via internal or external knowledge, such as that found in cognitive or cartographic maps.

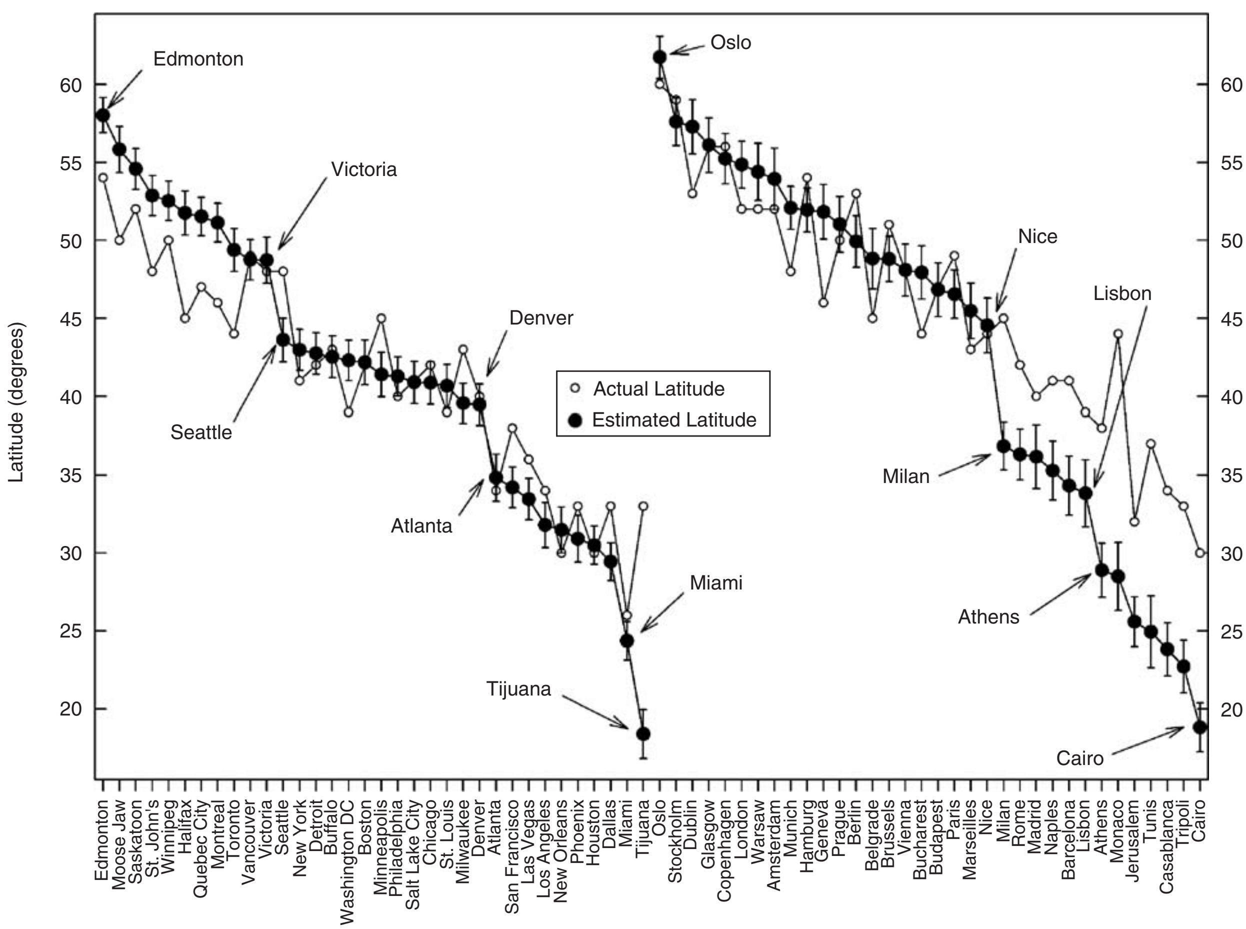

Figure 1 Regional organization and reasoning as reflected in patterns of latitude estimates for world cities. Cities (on the abscissa) are ordered according to estimated latitudes in the Western and Eastern Hemispheres. Adapted from Friedman, A. and Brown, N. R. (2000). Reasoning about geography. *Journal of Experimental Psychology: General* 129, 193 219.

Orientation is an essential component of navigation. Geographic orientation is knowing where you are on the Earth's surface and the heading to your destination. To the degree that people are uncertain about their location and/or heading, they are geographically disoriented. Orientation and disorientation are always relative to some reference system, whether based on one's body, on external landmarks, or on macroscopic properties of the Earth or the surrounding environment, such as magnetic north or the alignment of a visible mountain chain. As people navigate, they typically update their orientation. This can be done according to one of two broad classes of mechanisms, or a combination thereof. Landmark based updating (piloting, position fixing) is based on recognizing external features like landmarks. Dead reckoning updating (path integration, inertial navigation) involves updating by calculating or inferring a new location/heading based on knowledge about movement speed and direction from a known starting point, without recognition of specific features. Information about movement speed and direction can be based on internal signals, such as vestibular sensing of body acceleration, or external signals, such as optic flow in the visual field.

Learning and Development of Geographic Knowledge

People's geographic knowledge varies, in part, as a function of their age, their education, and their experience. That is, geographic knowledge depends on learning and maturation. Infants are born without specific geographic knowledge, although their nervous system is innately designed to acquire particular types of information (such as depth and directional relations) when exposed to the world through sensory–motor experience. As infants and children age, their geographic knowledge changes, because of changes to the nervous system, new experiences, and reorganizations of cognitive structures and processes that occur over time independently of specific experiences. Different theoretical approaches conceptualize development as occurring in relatively abrupt transitions between qualitatively distinct stages, or in relatively gradual and continuous transitions. These changes are both general, such as acquiring an understanding of spatial relations like hierarchical containment, and specific, such as acquiring an understanding of the layout of a particular neighborhood one has visited. Although these changes are usually thought of in terms of accumulating more knowledge and increasing the complexity or nuance of one's understanding, they also include forgetting and some loss in the ability to hold complex spatial relations in working memory.

Cognitive development over the lifespan of an individual includes changes over relatively short time spans as a function of experience – learning. A widespread theory of special relevance to geographers holds that people moving to a new place acquire an understanding of the layout of that place in a sequence of the three stages of landmark, route, and survey knowledge. Landmark knowledge is knowledge of the identity of landmarks, knowledge that is not intrinsically spatial. Route knowledge is knowledge of linear sequences of landmarks connected by travel patterns; routes are ordered but contain minimal metric scaling. Survey knowledge is knowledge of two dimensional layout from which spatial relations among places can be determined even if direct travel between them has never occurred; such knowledge is usually thought to be especially useful for formulating shortcuts and other instances of creative navigation. More recent theories question the sequential nature of this progression or even whether such qualitatively distinct stages of knowledge representation exist.

Geographic knowledge is acquired directly during sensory and motor activity in the world and indirectly from symbolic representations (more below). Sensory systems represent the first response of the nervous system to stimulation from the world. Sensory receptors transduce world energy patterned over space and time into patterned nervous system energy. For humans, relevant forms of energy in the world include electromagnetic, chemical, physical pressure and vibration, gravity, and heat; nervous system energy is electrical and chemical. Different types of sensory receptors that respond to different types of world energy lead to qualitatively distinct sensory modalities, including vision, hearing, smell, taste, pressure, texture, temperature, kinesthesis (body limb position and movement), and vestibular sensation (gravity and acceleration). Clearly, humans have more than the five senses commonly believed to exist. All senses contribute to the apprehension of geographic properties, but for humans, vision is especially powerful for apprehending detailed and precise spatial and thematic information at a distance. Knowledge about the world is acquired via motor activity, also. Movements such as turning one's head or walking around a corner provide us with information about the environment. That is, we do not move only to get to new places but to learn about them. Geographers are interested in how different modes of locomotion, whether mechanically aided or not, alter our experience of place as we travel and the information we pick up.

Symbolic Sources of Geographic Information

Indirect symbolic sources of geographic information include static pictorial representations such as maps, photographs, and remotely sensed imagery; static object representations such as physical models; dynamic pictorial representations such as movies and animations; and linguistic representations, including spoken or written natural languages, gestural languages, and formal computational or mathematical languages. The dynamic and actively controlled, first person computer graphic representations known as virtual environments (virtual reality) are indirect symbolic sources of knowledge, strictly speaking, but fully immersive systems simulate direct sensory–motor experience in real environments with increasing effectiveness.

From a cognitive perspective, maps and other representations of geographic information serve the purpose of communicating information about space, place, and environment to people, or otherwise inducing ideas or thoughts. Much cognitive research focuses on understanding the power of geographic information displays to support understanding and creative thought, or to support misconceptions and misleading conclusions. Cognitive geographers, including cartographers, study how information is perceived from maps and other information displays, how it is interpreted and stored in memory, how it is used to reason and solve problems, and how it is used to guide behavior. Maps and other displays represent features and other information with graphic symbols. The visibility, legibility, and semantic interpretation of these symbols are important, as is the sheer amount of information possible to display effectively at one time. Some symbols are relatively iconic, resembling or at least suggesting graphically what they mean; using blue to represent water may be easy to interpret but may also suggest that most water bodies are a particular shade of blue when they are actually green, brown, or black. Other symbols are relatively abstract and may therefore require more training for their interpretation; representing elevation by means of contour lines is an example. Computer technologies offer the possibility of tailoring information displays to the cognitive characteristics of individual users; research must determine the best way to measure these variations and help evaluate how displays should be modified to take account of them.

Cognitive geographers and others conduct research on the characteristics of spatial and environmental information as communicated via natural language, both spoken and written. This research will help us improve human interaction with digital geographic information systems, including navigation systems, digital libraries, and information systems for tourists and those with disabilities. Spatial information is to be found in most grammatical classes of words; nearly all prepositions refer to spatiality or temporality. Spatiality expressed in prepositions and other elements of language is almost always relatively imprecise and nonmetric, not expressing quantitative distances and directions. Many spatial terms depend on various aspects of context for their interpretation; ‘near the barn’ and ‘near Beijing’ surely refer to rather different proximities.

Linguistic descriptions of spatial layout designed to aid navigation are called verbal route directions. Cognitive geographers are interested in describing how people construct route directions and how people interpret them. For example, it has been observed that, as in other cases of linguistic communication, a speaker will attempt to assess what a listener knows in order to judge what level of complexity the listener can likely comprehend. Cognitive geographers are also interested in characterizing how people should deliver route directions so as to communicate optimally about navigation. Should directions focus on landmarks, or should they include survey layout information? How important is it to provide corrective or overshoot statements?

See also: Affect; Artificial Intelligence and Expert Systems; Behavioral Geography; Children and Mapping; Choice Modeling; Color, Mapping; Embodied Knowing; Experimental Design; Fuzzy Set and Fuzzy Logic; Geovisualization; Golledge, R.; Haptic or Touch-Based Knowledge; Landscape Perception; Language and Research; Lowenthal, D.; Map Perception and Cognition; Memory; Mental Maps; Neural Networks; Semiotics; Sense of Place; Subjectivity; Transport and Accessibility.

Further Reading

Downs, R. M., Liben, L. S. and Daggs, D. G. (1988). On education and geographers: The role of cognitive developmental theory in geographic education. *Annals of the Association of American Geographers* 78, 680 700.

Downs, R. M. and Stea, D. (1973). *Image and Environment*. Chicago, IL: Aldine Publishing.

Friedman, A. and Brown, N. R. (2000). Reasoning about geography. *Journal of Experimental Psychology: General* 129, 193 219.

Garling, T. and Golledge, R. G. (eds.) (1993). *Behavior and Environment: Psychological and Geographical Approaches*. Amsterdam: North Holland.

Golledge, R. G. (ed.) (1999). *Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes*. Baltimore, MD: Johns Hopkins Press.

Gould, P. and White, R. (1986). *Mental Maps* (2nd edn.). Winchester, MA: Allen and Unwin.

Kitchin, R. and Blades, M. (2002). *The Cognition of Geographic Space*. London: IB Tauris Publishers.

Lloyd, R. (1997). *Spatial Cognition: Geographic Environments*. Dordrecht: Kluwer.

Lynch, K. (1960). *The Image of the City*. Cambridge, MA: MIT Press.

Mark, D. M., Freksa, C., Hirtle, S. C., Lloyd, R. and Tversky, B. (1999). Cognitive models of geographical space. *International Journal of Geographical Information Science* 13, 747 774.

Montello, D. R. (2002). Cognitive map design research in the twentieth century: Theoretical and empirical approaches. *Cartography and Geographic Information Science* 29, 283 304.

Moore, G. T. and Golledge, R. G. (eds.) (1976). *Environmental Knowing*. Stroudsburg: Dowden, Hutchinson & Ross.

Portugali, J. (ed.) (1996). *The Construction of Cognitive Maps*. Dordrecht, The Netherlands: Kluwer Academic.

Tversky, B. (1992). Distortions in cognitive maps. *Geoforum* 23, 131 138.

Relevant Websites

http://cognitivesciencesociety.org/index.html
Cognitive Science Society
This professional society brings together researchers from many fields who hold the common goal of understanding the nature of the human mind.

http://www.edra.org/
EDRA The Environmental Design Research Association
This professional society is the home for researchers from several disciplines, including geography, psychology, and architecture, who study human behavior and environmental design, including cognitive research on relationships between people and their environments.

http://epbg.blogspot.com/
Environmental Perception and Behavioral Geography, Specialty Group of the AAG.
This specialty group is the home for cognitive and behavioral researchers in the Association of American Geographers.

http://cognet.mit.edu/
MIT CogNet
This site provides access to ten major reference works on cognitive science and neuroscience, including the "The MIT Encyclopedia of the Cognitive Sciences," edited by Wilson and Keil.

Cohen, S.

N. Kliot, University of Haifa, Haifa, Israel

Glossary

Arc of Instability South Asia with the Middle East form the Arc of Instability.

Geopolitical Region Meso-order tactical regions: North and Middle America or Maritime Europe and the Maghreb are geopolitical regions of the Trade Dependent Maritime Realm.

Geostrategic Region (or Realm) Macro-order strategic realms. The boundaries of such regions include areas that the major powers consider to be vital to their national interests.

Heartland Based on Holford Mackinder's (1861–1947) writing, this refers to the inner area of Eurasia, characterized by interior drainage and impenetrable by sea power. Later (1919) Eastern Europe was included in the Heartland.

Quarter-Sphere of Marginality The southern continents of South America and sub-Saharan Africa. These areas are characterized by poverty, illiteracy, diseases, and rebellions.

Shatterbelts Areas where competing geostrategic realms converge.

Cohen, Saul (1925–)

Born in Malden, Massachusetts, and was educated in Boston, Saul Cohen graduated in geography of Harvard (MA 1948, PhD 1954) and served in Europe in World War II. A liberal Democrat, he is married to Miriam, has two daughters and seven grandchildren.

In 50 prolific years of writing, teaching, and studying in geography, Saul Bernard Cohen has written 14 books and monographs, and over 100 articles and other publications (**Figure 1**). Cohen fulfilled many important academic and administrative positions, such as president of Queens College, New York City; Dean, Clark University Graduate School; president, the Association of American Geographers; and numerous positions and functions within the US educational system – the US Office of Education, the New York State Board of Regents and many more. His enormous contribution to education in geography is manifested by the hundreds of geographers who were educated by him in his rich teaching career at Boston University, Clark University, Queens College, and Hunter College. Furthermore, Cohen was a major force in several committees and organizations which focused on the advancement of geography in American Universities in his role in the American National Standards Committee and in his offices in the National Science Foundation, the American National Academy of Sciences, Office of the Geographer, US Department of State, and so forth. This genuine interest in education is expressed in *Problems and Trends in American Geography* (1967) in which the scientific knowledge of geography, its paradigmatic structure and perhaps its aspirations for the future, is presented in 19 chapters authored by prominent geographers of that era. An additional ten articles authored by Cohen are dedicated to education at large. Most conspicuous is 'Reflections on the elimination of geography at Harvard 1947–1951' which was published in the *Annals of the Association of American Geographers* in 1988.

It is most appropriate that a biography of Saul Cohen appears in the *International Encyclopedia of Human Geography*, as Cohen represents, through his writings, an interest and curiosity in a rich variety of topics: from *Store Location Research for the Food Industry* (1961); 'Israel's fishing industry'; 'Towards a geography of policy'; 'Geography and the Federal Government' (1968); 'Israel's place names as reflection of continuity and change in nation building'; 'The world geopolitical system – in retrospect and prospect'. Yet, Cohen, a combination of classical scholar with extensive research interests, is also a 'specialist' in its modern interpretation and is known as a world authority in modern geopolitics and political geography at large. His most important theoretical

Figure 1 Photo of Saul Bernard Cohen.

contributions are, however, in the area of political geography and geopolitics. Cohen is the geographical editor of the *Oxford World Atlas*, and editor in chief of the *Columbia Gazetteer of the World*.

Two of Cohen's research areas need to be discussed in depth as they embrace most of his research and writings: Israel and the Middle East, and political geography and geopolitics. His interests in the former were founded on frequent visits to Israel, which always involved joint re search projects. The research topics were mostly political and geopolitical: not only Jerusalem and Israel's borders but also Israeli society, ideology, politics, and the econ omy. Cohen, who is fluent in Hebrew, was able to conduct direct studies in Israel, which reflects another important quality: all were based on sound theoretical and empirical foundations. An example of such study is the Develop mental Approach to the Study of the Rural Sector in Israel, which was based on Cohen, Wapner, and Kaplan's 'An organismic developmental perspective for under standing transactions of men and environments'. This research, which was completed in the 1970s, was able to forecast processes which took place in Israeli rural soci ety 30 years later.

Cohen also wrote perhaps one of the best works on Jerusalem, in which he, with great skill, utilized his political geography craftsmanship to offer a govern mental/geopolitical integration structure for the city. His proposal for area wide integration is still relevant in the present time. In two other monographs, Cohen was able to analyze in a very eloquent manner all the aspects of the Israeli/Palestinian conflict, Israel's geopolitical stand ing in the Middle East, and its future. These works are still outstanding in their precise analysis of all the weak and strong points in Israel's geopolitical position. "To define secure borders in a narrow, military sense, is to adopt a too limited definition of geopolitical strategy." In addition to the abovementioned books, about one fifth of Cohen's scientific papers are dedicated to the study of Israel and the Middle East, and he brought to the area that original research perspective.

Yet, it is Cohen's broader contributions to political geography and geopolitics which is certainly his most significant contribution. The best expression of these original works is *Geography and Politics in a World Divided* and *Geopolitics of the World System*. In *Geography and Politics in a World Divided* Cohen presents, in the first two chapters, the essence of political geography and geo politics, and takes it very far from the German geo politics. Thus, Cohen guides political geography into six approaches to the study of the subject: the power analysis approach, the historical approach, the morphological approach, the functional approach, the behavioral ap proach, and the systemic approach. Cohen, together with Rosenthal, made a very important theoretical contri bution within the system approach. In the model proposed by those authors, a geographical system was advanced as the unit with which the political process interacts with geographical space. Political transactions, structures, and societal forces are the components of the process; place, area, and landscape are the components of geographical space. Process and space interact through the formation of political action areas, and various ideological attachments, organizations, and perceptions characterize these action areas. This approach charac terizes Cohen's unique perspective to geopolitics.

Following Gottman, MacKinder, and others, move ment and change are crucial elements in the geopolitical system as Cohen perceived it. He suggested two geo strategic regions: (1) The Trade dependent Maritime World and (2) the Eurasian Continental World. Pro jecting into the future, Cohen anticipated the eventual emergence of a third geostrategic region – the Indian Ocean Realm. The core of the Trade dependent Mari time World is the Maritime Ring of the USA and it is subdivided into four geopolitical regions: Anglo America and the Caribbean; Maritime Europe and the Maghreb; Offshore Asia and Oceania, and South America. The core of the Eurasian Continental World is the Russian In dustrial Triangle, with two geopolitical regions: the Russian Heartland and Eastern Europe and the East Asian mainland. Between these two geostrategic regions lie Shatterbelts – the Middle East and Southeast Asia.

In his 2003 *Geopolitics of the World System*, Cohen ad justed the global system to the various political changes. In his spatial hierarchy of the global structure, the highest level is the geostrategic realm. Cohen proposes three geostrategic realms: the Atlantic and Pacific Trade dependent Maritime Realm, the Eurasian Continental Russian Heartland, and the mixed Continental Maritime East Asia. Today's Trade dependent Maritime Realm, which embraces the Atlantic and Pacific Ocean basins and their interior seas, had been shaped by international exchange. The Eurasian Continental Realm, which is anchored today in Heartland Russia, is inner oriented and less influenced by outside economic forces or cul tural contacts. The boundaries of the Heartland Russian Realm have changed substantially. To its west, the East ern European states are no longer tightly within the grip of Moscow, but in the East, the former Soviet Republics of the Trans Caucasus and Central Asia are not free of Russia's strategic oversight. In the third Realm – the East Asia Realm – China emerged as a major trading nation, but has not become part of the maritime world, and is still continentally oriented. China extended its power southward into the Indochinese states of Vietnam, Cambodia, and Laos. South Asia stands apart from the three realms as an independent geopolitical region which, when linked with the Middle East, forms an Arc of Instability. The future of the Middle East Shatterbelt and Eastern Europe, as a gateway political region or as a

Shatterbelt, remains to be determined. The major feature of a Shatterbelt, as defined by Cohen, is that it presents an equal playing field to two or more competing global powers, operating from different geostrategic realms. The Shabberbelts of the 1970s – sub Saharan Africa, South east Asia, are no longer Shatterbelts, but the Middle East remains one. The future may bring additional Shatter belts onto the world scene: the Baltic through Eastern Europe and the Balkans are the most likely candidates. Cohen depicts in his book the evolution and develop ment of the global hierarchical structure and comple ments the jigsaw by two arcs of geostrategic instability. These are South Asia and the Middle East – and the southern continents – the Quarter Sphere of Marginality, as it is called.

Cohen treats the world as a general system. The de velopmental principle holds that systems evolve in pre dictably structured ways, that they are open to outside forces, that hierarchy, regulation, and entropy are im portant characteristics, and that they are self correcting. There is no doubting the theoretical crux of Saul Cohen's geopolitics, and the fact that it has endured well for more than 30 years of researching and writing about the up heavals of our world.

See also: Borderlands; Geopolitics; Geopolitics and Religion; Hegemony; Mackinder, H. J.; Political Boundaries; Political Geography; Superpower; World-System.

Further Reading

Cohen, S. B. (1957). Israel's fishing industry. *Geographical Review* 47(1), 66 85.
Cohen, S. B. (1961). *Store Location Research for the Food Industry*. New York: NAWGA.
Cohen, S. B. (1967). *Problems and Trends in American Geography*. New York: Basic Books.
Cohen, S. B. (1968). Geography and the Federal government in geographers in government. *American Geographical Society* 5, 1 4.
Cohen, S. B. (1973). *Geography and Politics in a World Divided*. New York: Oxford University Press.
Cohen, S. B. (1977). *Jerusalem Bridging the Four Walls*. New York: Herzl Press.
Cohen, S. B. (1986). *The Geopolitics of Israel's Border Question*. Tel Aviv: Tel Aviv University, Jaffee Center for Strategic Studies and Westview.
Cohen, S. B. (1988). Reflections on the elimination of geography at Harvard University 1947 1951. *Annals of the A.A.G.* 78(1), 148 151.
Cohen, S. B. (2003). *Geopolitics of the World System*. Lanham, MD: Rowman and Littlefield.
Cohen, S. B., Wapner, S. and Kaplan, B. (1973). An organismic developmental perspective for understanding transactions of men and environments. *Environment and Behavior* 7(3), 255 289.

Cold War

T. W. Luke, Virginia Polytechnic Institute and State University, USA

The Cold War was named in 1947 by Bernard Baruch in his advisory role to President Truman as the United States, Great Britain, France, and its other Western allies faced off against the Soviet Union, with mutually aggressive maneuvers and countermoves, especially in Europe. This label was publicized more broadly by Walter Lippmann in a 1947 book with same name, and it came to describe much of the world's twentieth century history. It was a multidimensional contest ranging from ideology, economics, and culture to geopolitics, diplomacy, and technology, which arguably began with the Bolshevik Revolution in 1917, but its tensions were diffuse as first liberal capitalist and then illiberal fascist regimes both courted Soviet diplomats and sought access to the USSR's markets. After years of intense competition, the Molotov Pact unified Nazi Germany and the Soviet Union in a joint invasion of Poland. Yet, within months, Berlin turned eastward in Operation Barbarossa, and Moscow allied itself with the capitalist West. Those struggles behind the Cold War were renewed near the close of World War II as the so called Big Five Allied powers began to disagree about how to organize the post 1945 world order with the defeat of the three major Axis powers – Fascist Italy, Nazi Germany, and Imperial Japan.

European Start of the Cold War

The Cold War's immediate origins lay in 1944 to 1946, but the conflict crystallized in 1947 as the United States and the Soviet Union began to engage in a very direct cultural, social, and technological competition as well as more indirect military confrontation. To conduct this war, in July 1947, only 2 years after President Truman, Prime Minister Churchill, and Joseph Stalin met at the Potsdam Conference between Great Britain, France, the United States, and the Soviet Union to create a peaceful post war world, President Truman reorganized the American military under a single Department of Defense, formed the Central Intelligence Agency (CIA), and created the National Security Council to wage the Cold War.

During World War II, the divergent visions of the Western Allies and the USSR over the nature of the post fascist governments in Europe were quite evident, but fairly moot as long as Nazi occupying forces were in place. By 1944, however, the conflicts became intense as Free French forces reestablished control in France, the German occupation of Italy ended south of the Piedmont, and the Soviet liberation of Poland brought Red Army units into Germany itself. Large numbers of communists fought against Hitler's forces in France and Italy, and a free Polish regime hoped to return to Warsaw. However, the USSR set up a pro Moscow provisional government in Poland, and the Allies worked to keep communists out of both France's and Italy's new governments. As Hitler's regime collapsed in May 1945, independent communist governments came to power on their own in Albania and Yugoslavia, while the USSR engineered the rise to power of communist regimes in Poland, Bulgaria, Romania, and Hungary.

Consequently, in his attempt to prevent the further spread of communism, President Harry S Truman promulgated the Truman Doctrine to aid anticommunist forces fighting in Greece and Turkey in March 1947. This move launched the official policy of 'containment', and it also triggered the Marshall Plan to fund Europe's industrial reconstruction. Greece might have gone the way of Yugoslavia, but British, and then American assistance helped stop the communist insurgency there during the bitter Greek Civil War. Even though Czechoslovakia initially installed an independent multiparty parliamentary government after the war, communist activists sparked a crisis in February 1948 that prompted Prague to ally itself with Moscow as well.

During 1947 and 1948 Britain, France, and the United States increasingly moved toward consolidating their respective zones of occupation in Germany into a single administrative unit, so the USSR imposed a land blockade in the summer of 1948 against the three Western powers' zone of occupation in Berlin. For 11 months, Berlin was sustained by a costly airlift, and this lifeline maintained Western control in the American, British, and French zones in the city. However, these developments led to the creation of West Germany in May 1949 and East Germany in October 1949 as separate states. The iconic center of the Cold War emerged here from a divided Berlin, a divided Germany, and a divided Europe. Similarly, the cooperative spirit of the Allied response during the Berlin Airlift continued with the creation of a formal alliance between the Western Powers around the North Atlantic Basin, or NATO (North Atlantic Treaty Organization). Even though minor progress was made after Stalin's death in 1953 with the State Treaty between Austria and the four wartime allies, ending its German style four power occupation and giving it independence in a formal state of neutrality during 1955, Europe remained very divided. And it stayed on the edge of war for the next 40 years after the creation of the two German republics. In

1955 as West Germany also attained full sovereignty, it was allowed to rearm itself, and then join NATO. East Ger many also began rearming in 1956, and then joined the East Bloc's military alliance.

Cold War Outside of Europe

Outside of Europe, the Cold War was a misnomer from the beginning as military conflict raged for decades. In China, the last revolutionary civil war between the Na tionalist, or Kuomintang, Party, and Communist Party led to the defeat of the Nationalists and their forced retreat to Taiwan by 1949, allowing the People's Republic of China to be declared in Beijing on 1 October 1949. The United States did not intervene decisively to assist the Kuomintang, but it guaranteed the Nationalist regime's security after the anticommunist forces fled to Taiwan. In less than a year, the partitioned Korean peninsula saw the Democratic People's Republic of Korea under Kim Il Sung invade its southern neighbor, the American backed Republic of South Korea created in August 1948 with UN support. The Korean War ensued from June 1950 through July 1953 as both Moscow and Beijing backed Kim's attempts to unify the Koreas militarily. An armis tice was accepted after 3 years of war between a United Nations coalition under American leadership with North Korea and its Chinese communist and Russian backers. The first of many full blown 'proxy wars' between the United States and the Soviet Union, the Korean War still has not ended in a formal peace treaty.

At the same time, French efforts to reestablish control over Indochina for Paris, after local communist and na tionalist rebels there fought with the Allies to defeat the Japanese, led to the First Indochina War from 1946 to 1954. The nationalist and communist Viet Minh forces under Ho Chi Minh divisively defeated the French in 1954, and the region was more formally partitioned into Cambodia and Laos, which remained under traditional royal rule, and two areas for a separate North and a South Vietnamese state. The Geneva Accords of 1954 promised national elections in 1956 to unify the nation, but the Southern regime in Saigon repudiated that pro vision, fearing a communist electoral victory. The Second Indochina War began in 1959, and it lasted until 1975. The Northern regime in Hanoi received arms and as sistance from the communist bloc, and the United States continued its support to anticommunist forces as it had done since 1946. In 1965, however, the United States sent large contingents of its own troops into Vietnam, but then withdrew in 1973 after 8 years of tactically successful but strategically indecisive battles. The conflict had been taken into Laos and Cambodia by both sides since the early 1960s. After the American invasion of Cambodia in 1970, the communist Khmer Rouge fought against American forces and the pro American traditional re gime until they too came to power in 1975 as the com munist forces of North Vietnam and southern Viet Cong partisans finally defeated the American backed South Vietnamese regime.

The 'hot war' nature of conflict outside of Europe largely was fueled by local nationalist struggles against weakened or collapsed old European empires. Some of these nationalist struggles began or grew, during World War II as anti Japanese, anti German, or anti Italian wars with Western backing, but communist parties often played a major role in many of anti fascist struggles. Others were more local conflicts, but they soon became sites of struggle between the United States and the Soviet Union because of the communist parties. The collapse of the Batista government in Cuba, the coming to power of Fidel Castro in 1959, and the support given to radical nationalist Arab powers after the emergence of Israel in 1947 in many subsequent conflicts are good cases in point. While more indirectly ideological in nature, these two sites were the most volatile flashpoints of the Cold War era.

The Kennedy administration's backing of the failed Bay of Pigs counterrevolutionary invasion of Cuba moved the Castro government closer to Moscow for military and economic support in 1961. By mid 1962, it was becoming apparent that the USSR was preparing to place a small number of intermediate range ballistic missiles in Cuba. While this deployment mirrored earlier American basing of comparable missiles in Western Europe and Turkey, the United States challenged Moscow in October 1962 with an 'embargo' of seaborne shipping into Cuba to prevent the completion of these Soviet missile sites and force their dismantlement. For nearly a week, the world teetered on a knife edge between this shaky peace and an all out nuclear war until Khrushchev backed away from confrontation. Such brinksmanship during the Khrush chev years, which had reached nearly the same levels of nuclear confrontation in Berlin and Taiwan, taught both sides to be more cooperative with each other in the Cold War's intense conflict in the coming years.

This more cooperative spirit was reflected in the more subtle support the Soviet Union provided to the frontline Arab states against Israel, which had strong American backing, in its 1956, 1967, 1970, 1973, 1980, and 1981 conflicts with its neighbors. Although Soviet aid backing for Egypt and Syria led to an American global nuclear alert in 1973 during the Yom Kippur War, Moscow acted more gingerly around the explosive Arab–Israeli conflicts even when Israel openly invaded Lebanon, fought air battles with Syria, or destroyed Iraq's developing nuclear weapons program by bombing the Osirak reactor com plex. Moscow backed the Palestinian Liberation Organ ization and radical regimes in Libya, Yemen, Syria, Iraq, and Somalia, but it carefully avoided a more direct face off with the United States in the region even when it became apparent that Washington was actively supporting the Islamic mujahaddin counter revolutionaries in communist led Afghanistan against the Soviet invasion there.

Most of the ideological contest between the United States and the Soviet Union evaporated as a battle of ideas in the 1960s and 1970s. Khrushchev's limited efforts to reimagine communism after Stalin's death in March 1953 in more flexible consumer oriented terms ended with his ouster in October 1964. The Warsaw Treaty Organization forcibly intervened in Czechoslovakia in 1968 to maintain Soviet style rule over and against more independent and innovative Czech ideological reforms, and a New Left all around the world repudiated the Brezhnev and Kosygin traditional party line as 'obsolete communism'. Likewise, Washington became bogged down in the quagmires of Southeast Asia fighting the communist led wars of national liberation in Cambodia and South Vietnam by backing corrupt military dictatorships, even as it accepted strategic parity with the USSR with regard to its nuclear strike forces.

The People's Republic of China erupted in a countrywide ideological civil war between radical Maoist 'red' activists and more technocratic 'capitalist road' economic reformers, which ended only with Mao's death in 1976 and then Deng Xiao Ping's consolidation of power in 1978. After these upheavals, Beijing slowly turned more toward the West just as a strange new era of *détente* unfolded between Washington and Moscow after West Germany made moves toward an 'Ostpolitik', or Eastern Politics, focusing on greater cultural, economic, and technical exchanges with the East Bloc. These geopolitical maneuvers and internal battles left Moscow's pretense to being the world's leading revolutionary center nearly a complete mockery, particularly in years between August 1968 after it invaded Czechoslovakia to impose a Moscow friendly communist regime after ousting a more maverick nationalist communist government and December 1979 when it charged into Afghanistan to prop up a putative Marxist–Leninist regime in Kabul. Rather than leading a global ideological movement, as it had tried successfully during the 1920s and 1930s, to convince others to embrace communism willingly, by the early 1980s Moscow was left embroiled in numerous interventions around the world in Latin America, Africa, the Mideast, and Asia seeking to impose or maintain its model of 'actually existing socialism', which fewer and fewer Soviet citizens even wanted anymore at home, with military assistance, economic subsidies, or actually armed interventions.

The neoliberal turn in Great Britain and the United States in 1979 and 1980 brought to power Margaret Thatcher and Ronald Reagan. Their revitalized faith in capitalist free enterprise and militant anticommunism created new tensions between the East and the West in the early 1980s as the Communist Party and Soviet Union saw four general secretaries and presidents come and go within 3 years. Recognizing that Leonid Brezhnev's 'actually existing socialism' was only masking the institutional failings of the socialist order, while it stumbled into more and more unsustainable international entanglements, Mikhail Gorbachev became President of the Soviet Union and General Secretary of the Communist Party in 1985. He tried returning to Khrushchev's reformist tendencies with his own three pronged program of greater *perestroika* (restructuring), *democratikizatsiya* (democratization), and *glasnost* (openness) in the USSR. These efforts were matched with new strategic arms accords with the West, a pullout from Afghanistan, and reductions in its military forces in Eastern Europe. By 1989, such rapid changes clearly were pulling apart the Soviet Union itself along with its satellite regimes in Eastern Europe. The nationalist and popular uprisings in Poland from 1980 triggered a real change as 1989 and 1990 saw a wave of peaceful popular uprisings replacing the Marxist–Leninist regimes of Eastern Europe. In November 1989, the Berlin Wall was breached, Germany reunited in 1990, and the Russian Republic essentially started refusing to collaborate in the Soviet Federation with the larger Soviet Union. By December 1991, the USSR itself dissolved, and the major republics within the Union became sovereign independent states under the rule of former communist authorities who generally repudiated their Communist Party memberships but not their authoritarian party leadership styles.

Legacy of the Cold War

The Cold War expressed at least two very different approaches to organizing and then living modern, urban, industrial life – one supposedly capitalist and one allegedly socialist. Despite each side's pretense of ideological faithfulness, neither liberal democratic capitalism nor totalitarian state socialism was true to either their original ideological designs or their actual ordinary expectations. The free market aspirations of nineteenth century liberalism always were constrained by state involvement, cultural inertia, or religious conviction just as the full communist utopia of nineteenth century socialism failed to be realized in conditions of economic underdevelopment, feudal governance, or permanent militarization.

Consequently, liberal democratic capitalism – even in the 1920s, 1960s, or 1980s – never proved to be entirely liberal, fully democratic, or truly capitalist as bureaucratic paternalism, constitutional limitations, and national protectionism frequently hobbled personal rights and social progress. Likewise, totalitarian state socialism – even in the

1930s, 1950s, or 1970s – always lacked total social control, rational state authority, and convincing socialized community that undercut the foundational ideas of collective economic, cultural, social, and individual freedom. Racism, nationalism, elitism, and militarism, which existed in both blocs prior to 1914, and then worsened into the 1940s, clearly deformed both the capitalist West and the communist East in ways that arose from each bloc's peculiar beginnings. Yet, their life or death struggle with fascism from the early 1930s to 1945 also created a system of permanent war that neither Washington nor Moscow, London or Beijing, Paris or Belgrade found easy to repudiate. What ideology could not sustain, geopolitics easily fueled for nearly 50 years.

In a truly Orwellian fashion, the peace that allegedly dawned in 1945 actually led to an era of continuous war; and, the victories and defeats incurred by both sides in supposedly 'low intensity' conflicts in Asia, Africa, and Latin America soon sparked new 'high intensity' confrontations between Cold War client states, parties, and movements and their former sponsors in Washington and Moscow. The spillover from this violence has plagued the daily lives of millions since the Cold War purportedly ended in 1991. Indeed, many of the seeds of today's post Cold War conflicts were sown in the 1970s as narcocapitalism developed out of socialist national liberation movements, radical jihadism came from fighting for Washington against Moscow's Mideastern client states, and maverick capitalist and socialist countries went off on their own irrational nationalist misadventures in Santiago, Buenos Aires, Pretoria, Taipei, Tel Aviv, Tehran or Havana, Pyongyang, Mogadishu, Baghdad, Hanoi, Belgrade, and Damascus. Anarchy in Africa, poverty in Latin America, jihadism in the Mideast, and genocidal nationalism everywhere all were fueled by the conduct of the Cold War.

Washington's inept diplomacy after 1991 in so many places around the world, and Moscow's diminished capacity to rule after 1989 led not 'the end of history' but a return of so much 'history repressed' since 1914 that the past two decades have been ones of geopolitical drift and destruction rather than years leading to a progressive equitable and developmental New World Order. In many respects, the Cold War has not truly ended. It has simply abated somewhat in its intensity or morphed significantly in its form. Two decades after Ronald Regan called for Mikhail Gorbachev to tear down the Berlin Wall and fully engage the communist bloc with global society, which have both occurred, the United States and Russia continue to maintain thousands of strategic nuclear weapons. Six thousand of these are deployed with long range strategic strike forces whose most probable use would be against each other – the two former Cold War rivals. Moscow continues to resist Western inroads into its former spheres of influence all across Central and Eastern Europe, while Washington still frets over the popularity of Fidel Castro's regime in Cuba and his close ties to populist anti American regimes at various times Venezuela, Ecuador, Bolivia, Peru, and Argentina.

Meanwhile the survival of ruling communist parties, and the rise of new authoritarian regimes with anti Western leanings, ranging from Cuba and North Korea to Belarus and Serbia, are persistent factors that keep an aggravatingly Cold War style of diplomacy alive and well in some corners of the world. Even though these smaller states no longer have the backing of powerful ideologically antagonistic patrons in Moscow or Beijing, these countries maintain repressive, and sometimes militant, governments that are anti Western, anti capitalist, and anti American in tone, which Russia and China occasionally have found useful to support. The USA, in turn, continues its policies of armed containment, economic sanctions, and ideological condemnation in many of its relations with these governments long after the demise of the Soviet Union. It also expects its allies to follow this lead in their respective regional ties, but this pattern varies. Most US allies, for example, maintain a Cold War stance vis à vis North Korea, but many others have developed more cordial relations with Cuba. Likewise, Moscow is threatened by the German, French, and the US advances into Central and Eastern Europe with the enlargement of the European Union and NATO.

The domestic effects of the Cold War in the United States and the Soviet Union, then, have been fundamental and long lasting. Whatever hope remained for those seeking to build a truly alternative modernity grounded in a revolutionized socialist society was lost in the permanent mobilization of a red garrison state. After the 17 June 1953 uprising in East Germany, Moscow had to forcibly impose, and then maintain, its socialist designs through military coercion for nearly four decades. Ultimately, the cultural, economic, political, and social costs of actually existing socialism during the Cold War led to the union's fragmentation, socialism's bankruptcy, the republics' secession, and the soviets' degradation. In December 1991, the USSR ended without a bang and with few whimpers. The remaining republican autonomy presumed to be held by the American states and citizenry has been continually beset by federal emergency measures in the USA. The young nation was instructed by its first president to avoid entangling foreign alliances and to stay true to its original revolutionary ideals. Instead, it created a vast network of treaties, an empire of bases, and an institutionalized military–industrial apparatus that has borrowed trillions, misspent billions, and disserved millions to attain its Cold War victories. In the meantime, the blowback from its decades of proxy wars, engineered *coups*, military interventions, and diplomatic misadventures has set the stage for much of today's twenty first century post Cold War global affairs. Just as there has not been an 'end to history', there is not yet truly an end to Cold War.

See also: Africa; Capitalism; Middle East and North Africa; Third World.

Further Reading

Booth, B. (1998). *Statecraft and Security: The Cold War and Beyond*. Cambridge: Cambridge University Press.

Carter, D. and Clifton, R. (2002). *War and Cold War in American Foreign Policy, 1942 62*. London: Palgrave.

Fischer, B. A. (2000). *The Reagan Reversal: Foreign Policy and the End of the Cold War*. Columbia, MO: University of Missouri Press.

Gaddis, J. L. (2005). *The Cold War: A New History*. New York: Penguin Press.

Garthoff, R. L. (1994). *The Great Transition: American Soviet Relations and the End of the Cold War*. Washington, DC: Brookings Institution Press.

Hanhimaki, J. M. and Westad, O. A. (2004). *The Cold War: A History in Documents and Eyewitness Accounts*. New York: Oxford University Press.

Lebow, R. N. and Stein, J. G. (1995). *We All Lost the Cold War*. Princeton, NJ: Princeton University Press.

Leffler, M. P. (1994). *The Specter of Communism: The United States and the Origins of the Cold War, 1917 1953*. New York: Hill and Wang.

Loth, W. (2002). *Overcoming the Cold War: A History of Détente, 1950 1991*. London: Palgrave.

Mastny, V. (1998). *The Cold War and Soviet Insecurity: The Stalin Years*. New York: Oxford University Press.

Sayigh, Y. and Shlaim, A. (1997). *The Cold War and the Middle East*. Oxford: Oxford University Press.

Taylor, P. J. (1990). *Britain and the Cold War: 1945 as Geopolitical Transition*. New York: Guilford.

Westad, O. A. (2007). *The Global Cold War: Third World Interventions and the Making of Our Times*. Cambridge, MA: Cambridge University Press.

Colonialism I

A. Lester, University of Sussex, Brighton, UK

Introduction

In recent years, there has been an enormous body of work which tries to understand modern European colonialism in all of its economic, social, and cultural dimensions, and to grapple with its legacies. The focus of this article is especially on British and, to a lesser extent, on French colonialism up to about 1870.

The terms 'colonialism' and 'imperialism' are closely interrelated. They are both concerned with the actions of people based in, or traveling out of, the world's most powerful polities located in metropolitan spaces, who attempt to dominate people elsewhere in the world. The main difference between the two terms is that 'colonialism' refers largely to the actions of those who migrate from metropolitan space to pursue various projects as colonists within subordinated territories. 'Imperialism' refers more often to a kind of 'remote control' exercised over other parts of the globe by powerful agents within metropolitan space, especially by the state. In practice, however, colonists tend to maintain dense networks of communication and exchange with metropolitan agents, and so it is easy to see why the two terms are often conflated. Both will be used in what follows.

A Brief Historical Geography of Modern European Colonialism

Empires, constructed as comprising metropolitan 'centers' and colonial 'peripheries', have been characteristic of social and political organization since historical records began. They are in part the result, and in part the cause, of uneven distributions of resources and they strongly influence the spatial transactions in such resources at scales from the local to the global. But they are also configured by, and help to reconfigure, movements of people, ideas, knowledge, and culture. The geographically extensive networks of communication and exchange that empires facilitate help determine the ways that collective social identities are re created in both metropolitan and colonial spaces. They do so by inscribing distinctions in power between colonists and colonized, governors and governed, and citizens and subjects. Although the precise terms of these distinctions are always contested within metropolitan cultures, since at least the eighteenth century and the rise of the modern European empires, they tend to have been strongly racialized, as well as gendered and sexualized.

The ancient and medieval empires founded for example in Greece, Rome, Peru, China, and Persia were based on a metropolitan polity's domination of contiguous land based territories. Distinctions of inherited status, religion, and gradations on a scale of civility operated as the axes of difference between colonizing and colonized populations in these empires, rather than the more modern concept of 'race'. From the fifteenth century, however, Western Europe became the locus of a number of new metropolitan polities that created empires in tandem with the dissolution of feudalism and the rise of mercantile capitalism, and which coexisted with continuing land based empires such as that of China and, later, Russia. Following Portuguese navigational innovation in the fifteenth century, these western European empires were largely maritime rather than land based, allowing them to span trans continental distance and connect more diverse peoples living in more dispersed places. These modern empires brought populations that had previously been entirely separated, into unequal relations with each other, forging new webs of long distance trade, fostering new movements of people and ideas, and bringing about new cultural formations.

The early modern Portuguese and Spanish empires were based on relations of tribute and plunder, especially in South America, where some historians have argued that the first genocides of indigenous peoples were carried out by the conquistadors in their pursuit of gold and other resources. From their very beginning these imperial ventures were the subject of moral debate and critique within the European metropoles, with Bartolome De Las Casas, for instance, campaigning against the atrocities that Spanish authorities were carrying out against indigenous peoples. Like the European humanitarians who later emulated him, De Las Casas had no objection to colonization *per se*, but rather to the irreligious manner in which it was being carried out. His injunction to the Spanish during the 1540s to use their influence in South and Central America in order to convert Indians to Christianity and thus bestow benefit upon them, rather than simply to exploit their land, labor, and resources, was an antecedent to both the British 'civilizing mission' and the French *mission civiisatrice* of the nineteenth century.

From the sixteenth century, the English state was licensing privateers to destabilize the Iberian powers' monopoly of South and Central American resources. By the seventeenth century, with new sea routes allowing Europeans to travel around Africa and into the Indian

Ocean, Dutch and English trading companies were able to vie for control of the lucrative trade in spices and silks with East and Southeast Asia as well. While Dutch, English, and French rivals established supply and trading depots in what are now South Africa, India, and Indonesia, French and British planters in the Caribbean and North America initiated what would become an enduring trade in enslaved people across the Atlantic from the coast of Africa. By the late eighteenth century, the material cultures of Britain and France were conditioned by well developed circuits of exchange connecting Western Europe, Africa, Asia, and South America. Within British and French ships, for example, Indian calicoes were transported to Africa to purchase slaves, Tahitian breadfruit was taken to the Caribbean to feed those slaves, Caribbean molasses was transported to New England where it was made into rum for trade with Native Americans, and tea, coffee, chocolate, tobacco, sugar, rice, and potatoes converged from sites distributed across the globe on the British and French cities and provinces.

The rivalry between the British and French states in the eighteenth and early nineteenth centuries brought geopolitical considerations into line with these extensive mercantile operations, especially in Asia and North America. Governmental intervention helped to transform small trading settlements into state supported colonies. The fiscal revenue to be exacted from their subjects' economic transactions with indigenous peoples (the fur trade in North America and tea, spices, and increasingly land rental in India) was vital to the maintenance of the military, and of the state itself, in both European countries. Control of trade routes was thus a motivation for governments to colonize, for example, when the British state seized the Cape Colony (and its vital port at Cape Town) from Republican France's ally, the Batavian Republic, in 1806. Similarly, the British state colonized New South Wales in Australia not just as a penal settlement but also in order to deter the French state from claiming it as a base for operations in the Pacific. In both places, British governments subsequently secured their strategic interests by promoting the emigration of British colonists, giving, renting, or selling them land regardless of prior indigenous occupation. Largely as a result of the wars with France, by 1820 the British Empire had already absorbed almost a quarter of the world's population.

However, modern colonization was by no means always the result of either trading opportunities or geostrategy. Although a trade in flax and concern about possible French pre emption were certainly considerations in Britain's colonization of New Zealand in 1840, for instance, its annexation was primarily the result of an evangelical humanitarian lobby wishing to minimize the supposed deleterious effect on Maori of Britons who had already emigrated there. Again, humanitarian concern over colonization manifested itself as a drive for morally charged, benevolent imperialism, rather than imperial withdrawal.

Colonial Projects and Outcomes

Imperial historians have persistently attempted to devise models of imperial expansion based on a single 'driving force', such as the deliberations of an official mindset, the imperatives of industrial capitalism, or the financial and cultural demands of 'gentlemanly capitalism'. But there was never one overriding reason why modern European colonialism extended in the ways that it did. Its emergence by the late nineteenth century was the product of geographical contingency, rival state deliberations, and warfare and a proliferation of different colonial agendas as Europeans realized the opportunities that diverse colonial spaces provided, for example, for material gain, enhanced family security, escape from the constraints of European gender and sexual mores, the acquisition of scientific knowledge, new careers, or Christian proselytization.

These different agendas or colonial projects often coalesced. For example, the geostrategic desire of the state to secure a particular colony from imperial rivals meant that settlers were given land and an opportunity for a new life. But colonial projects also diverged and could come into conflict, for instance when humanitarians agitated against settler brutality and the dispossession of indigenous peoples, and insisted on the moral purpose of colonialism outweighing its material value. The Europeans pursuing these multiple projects also constructed varying discourses of difference. Many settlers, for instance, represented 'racial' difference as immutable, developing biologically determinist notions of 'race' which justified the appropriation of indigenous peoples' land and on occasion even their 'extermination'. Humanitarians insisted upon a contrasting universalist view of humanity in which 'racial' differences were cultural rather than innate, and thus subject to erasure once Christianity had been successfully proselytised.

Colonial relations 'on the ground' were yet more complicated than colonial discourses of difference, because European colonial projects could never simply be imposed upon empty terrains. Colonists' projects were always either reinforced or more commonly deflected or undermined by the actions of those who already inhabited colonized spaces. In settings such as India, the selective cooperation of indigenous elites was necessary if a British administration was to be superimposed upon or coexisted with indigenous hierarchies. In South Africa and New Zealand, British authorities were able to forge alliances with some indigenous chiefdoms against others, enhancing the status of allies at the expense of those who resisted colonial expansion. Convergences between

European and indigenous structures, however, went deeper than just the deals struck by elites. To give just one example, systems of colonial labor extraction, which enabled businesses such as the mining industry in South Africa to grow, were themselves the product of a co incidence between dynamic European and African patriarchal structures. These meant that African men could be absent from their rural homesteads for extended periods to fulfill the mine owners' insatiable demand for cheap migrant labor, while African women were expected by both British and African men to take care of social reproduction in segregated 'reserves'. The existence of these reserves was itself in large part due to the relative success of African resistance to the appropriation of their lands during the numerous colonial wars of the nine teenth century. Colonial outcomes in any given imperial site, then, were the result of the uneven relationships forged between Britons and indigenous peoples along a spectrum from collaboration through negotiation, ap propriation, and resistance, rather than the fulfillment of metropolitan devised and imperially driven agendas.

Colonial Legacies: Economy, Governance, Education

Although decolonization led to the dissolution of the modern European Empires from the 1940s to the 1970s, it is widely recognized that these empires' reconfigura tion of relations between peoples across the globe over the preceding 300 years still has profound consequences. The benefits and costs of these consequences are the subject of fierce political and historiographical contestation.

Perhaps the most immediately obvious legacies of the modern European empires are the material inequalities between former metropoles and colonies. At the onset of European colonial involvement in any given place, there were not necessarily great disparities between the eco nomic resources available to colonizer and colonized. English East India Company entanglement in the gov ernment of India, for instance, was a piecemeal and gradual process, reliant as much on the fragmentation of an initially far more powerful Mughal empire as it was on the growing resources and governmental connections of the company itself. Historians argue over the economic effects of British rule thereafter. Niall Ferguson, for in stance, suggests that British investment and free trade policies helped to restructure the modern Indian econ omy (as well as the economies of other former colonies) broadly in ways that are favorable for post independence governments. By contrast, Marxist historians emphasize the exploitative nature of colonial economic relations, pointing, for example, to British taxation policies which discriminated against Indian textile producers and in favor of British exports, and to colonial agricultural regimes which skewed production away from the in digenous population's subsistence needs and toward a vulnerable dependence on metropolitan consumers' fickle tastes.

Even broadly pro empire historians have admitted that policies of free trade and the commercialization of agriculture in parts of Africa and Asia often militated against redistributionist measures that had formerly been in place, thus contributing to mass starvation in the famines of the nineteenth and twentieth centuries. It is also beyond doubt that in settler colonies (those where large numbers of Britons migrated and stayed), the most fertile land and the most valuable resources were in variably wrested from indigenous peoples and taken into the hands of settlers. It was overwhelmingly these set tlers, rather than the prior inhabitants, who benefited from the economic opportunities of colonialism. They were able to engage profitably in a trade with metro politan consumers of the resources that they extracted from or cultivated on 'their' new lands, in order to create the modern states, for instance, of Australia, New Zealand, Canada, and South Africa.

The effects of modern European colonialism on modes of governance and education are also contested. In each of the territories colonized by British interests, governmental structures and schooling systems were es tablished that bore resemblances to British models. Some historians see them as examples of the progressive 'modernization' that the British Empire brought to large swathes of the world, and evidence of the Empire's as similationist tendencies. However, there was considerable variation in the adaptations that colonial officials made to British models in colonial spaces, and 'British' ways of thinking about education and governance were affected in turn by colonial experiences.

In the settler colonies of the nineteenth and early twentieth centuries, a British model of education was often adopted for white settlers but no formal state education at all, or only a more restricted and utilitarian form of training, offered to indigenous peoples. Besides importing elements of the educational curriculum from Britain, settler educationalists were able to export their 'knowledge' of indigenous peoples back to Britain. Thus, in the late nineteenth and early twentieth centuries, the geography syllabus in British schools included biologic ally determinist 'explanations' of the limited intellectual capacities of indigenous peoples drawn from the exam ination of 'specimens' supplied, and often analyzed, by colonial scientists. A Department of Geography was founded at the University of Sydney by Griffith Taylor, who expounded his theory that 'racial evolution' had left the 'Negro far behind' in the main British Geography journals, and was influential in the British university curriculum during the 1920s. Aside from cementing ra cist discourses of difference between colonizer and colonized and enduring notions of civilised/developed and barbaric/undeveloped parts of the world, these formal lessons about 'other races' helped inscribe a hierarchy of knowledge in the consciousness of European people. White Europeans, largely men, were understood to be the producers of knowledge and indigenous peoples its subjects. This was despite the fact that exploratory, cartographic, and scientific understandings were invariably shaped in colonial spaces only with the collaboration of indigenous informants.

Outside of the governmental education systems in settler colonies, mission societies ran Christian schools which offered a more complete curriculum to indigenous converts, and took the lead in disseminating English as a language of indigenous subjects, as well as colonial citizens of Empire. From the late nineteenth century, it was in these establishments that indigenous leaders learned and appropriated the discourse of nationalism in order to found movements such as the African National Congress in South Africa and the Congress Party in India. Although they usually began by making claims for the extension of colonial privileges to indigenous elites, these movements ultimately developed into broad based alliances between extremely diverse indigenous groups. In the twentieth century they helped bring the British Empire to an end.

What spurred such nationalist movements to transform themselves from polite, elitist pressure groups to mass movements of colonized peoples was the fact that, in most political systems modeled on Britain's Westminster pattern, the franchise was actually limited overwhelmingly to white settlers. Generally only a small minority of indigenous subjects, if any at all, were qualified to vote by virtue of their exceptional education or property. With mission schools helping to produce increasing numbers of such potentially qualified black voters by the 1890s however, franchise criteria were progressively tightened in a number of settler colonies, so as to prevent indigenous voters ever outnumbering whites. Different colonial governments frequently shared ideas for procedures which would maintain white supremacy and limit indigenous peoples' capacity to participate in colonial citizenship, without the need for explicitly racist legislation. For instance, when the southern African colony of Natal devised an English language dictation test which would restrict further immigration from India in the 1890s, the idea was disseminated via a meeting of heads of colonial governments in London and adopted by Australian and other administrations. In Australia, it became a central plank of the subsequent 'White Australia policy'.

Even outside of the settler colonies, the Westminster political system tended to be applied selectively. In the British colonies of Africa, which included Crown colonies and Protectorates as well as the settler colonies to the south, limited forms of citizenship had been extended to urban African elites by the mid twentieth century. But indirect rule through the ossifying of 'traditional' chiefly authorities characterized the rural 'reserves'. These left a legacy of kinship and clientship based forms of authority that were inherited by post colonial African governments, and which have continued to facilitate corruption and anti democratic practices. When conservative modern historians proclaim the virtue of the Empire in leaving a legacy of British liberal notions of democracy and inclusive education, such exclusions, that were built into the very fabric of colonial life, and that helped to maintain its fundamental 'rule of difference', are often overlooked.

Colonial Cultures

From the late 1970s, the history of modern colonialism ceased to be the preserve of imperial historians concerned to locate the economic, political, or military 'root causes' of European expansion. The social, and above all, cultural relations and legacies of colonialism began to be tackled much more systematically within literary criticism, and a body of work known as postcolonial theory was developed. Of course, this work had its own antecedents, many of them drawn from the French colonial experience. There was Franz Fanon's analysis of the psychological results of colonialism in French Algeria, for instance, and C. L. R. James' classic account of the ways that French revolutionaries resorted to inscribing racial limits around the 'European' concepts of liberty, egality, and fraternity, when responding to the rebellion of slaves in Haiti. But it is Edward Said's book *Orientalism* (1978) which has galvanized the analysis of colonial identities and discourses from new cultural perspectives over the last 20 years.

Said's book was essentially an analysis of the ways that multiple European, and more recently American, writers had configured 'Oriental' spaces (the Middle and Far East in their world view) over a period of some 400 years. What was remarkable was the enduring and repetitive nature of the binary opposites between Occident and Orient that these texts established. In persistently defining the West as civilized and the Orient as savage, the West as enlightened, the Orient as ignorant, the West as democratic and the Orient as despotic, Western texts had brought into being a discourse that not only defined Western knowledge about the East, but also constituted the identity of the West itself, as the Orient's opposite. For Said, himself a Palestinian exile, colonial relations were not only about the diplomatic, military, and political mechanisms through which the British among other empires expanded, but more pertinently about the enduring tropes of cultural difference that, as Gregory has shown with reference to ongoing wars in Afghanistan, Iraq, and Palestine, still characterize and legitimate Western interventions overseas.

Other postcolonial critics have sympathetically critiqued Said's work, for instance, for paying insufficient attention to the intersections between 'race' and gender in colonial discourses, and for evading the issue of how inhabitants of the 'Orient' have constructed both themselves and the West as they return the West's 'gaze'. Homi Bhabha has complicated Said's notion of a discourse of binary differences between colonizing and colonized cultures, emphasizing instead those 'third spaces' in which cultural hybridities, psychological ambivalences, and subversive mimicry emerge as a result of colonial encounters. These proliferating postcolonial interventions tend to be preoccupied with the past in the present – the ways that modern European colonialism created not just enduring material imbalances across the globe, but also patterns of identity formation and discourse that continue to affect our understandings of what is civilized and what is not, for instance. These understandings continue to mean that some groups' representations of the world are empowered at the expense of others.

The 'cultural turn' that postcolonial scholars across a number of disciplines have brought to the field of colonial studies has been criticized by many imperial historians. They point to the sweeping theoretical generalizations that mark some of this work, its analysis of text devoid of temporal and spatial context. But a growing body of work, often referred to as 'new imperial history', has recently sought to bridge the divide between traditional imperial history and postcolonial scholarship. It does so by recognizing that colonialism was fundamentally constitutive of, as well as constituted by, culture, but also by acknowledging the need to ground such insight in contextual and critical archival enquiry. New imperial history also brings a more gendered analysis of colonial relations, and a more spatially sensitive mode of enquiry than either traditional imperial history (with its focus on white male elites and its metropolitan bias) or much of the foundational postcolonial scholarship (with its lack of gender differentiation and its aspatial theorization).

For example, Catherine Hall's work has focused attention on the ways that Victorian definitions of those social groups eligible for inclusion in the British body politic were both gendered and 'raced' with continual reference to colonial forms of exclusion and dominance rehearsed and practiced in Jamaica. Mrinalini Sinha and Antoinette Burton have focused on similar circuits of discussion over the definition of manliness and of feminism that connected India and Britain within what Sinha calls an 'imperial social formation'. Each of these authors has also been at pains to demonstrate the ways in which British culture has been fundamentally imbricated with imperial assumptions and colonial knowledges, and to point to the postcolonial dicscussions over 'racial' and cultural difference 'at home' in Britain that are one of the empire's most obvious legacies.

Colonial Geographies

Recent historical geographies of colonialism have intersected quite closely with these tendencies in new imperial history. There has been a similar determination to specify particular colonial projects, rather than derive general models of the driving forces of colonial expansion. There have been similar attempts to track these projects across both metropolitan and colonial spaces, showing how they helped, through their interaction with indigenous projects, to constitute colonial and metropolitan societies reciprocally. A few examples are given as follows.

Daniel Clayton has analyzed different discursive modes through which Vancouver Island was understood, both locally and in Britain. These modes of narration ranged from 'neutral', scientific accounts of exploration and contact with indigenous society in the late eighteenth century, through records of trading encounters founded on a degree of reciprocity, to metropolitan based geostrategic planning, which inclined to write indigenous agency out of the picture in the nineteenth century. Alan Lester has examined the ways in which the contested colonial projects of governmentality, humanitarianism, and settler capitalism were pursued through communicative networks that linked the Cape Colony and Britain during the early nineteenth century. Together with David Lambert, who has examined forms of whiteness that were constructed during the moral struggle over slavery in Barbados, he argues that the trans imperial contests between proponents of such different colonial projects helped to structure debates around Britishness and 'race' across the Empire as a whole. Moving from public and political (masculinized) debates to the more intimate relations that underpinned colonial cultures, Alison Blunt has been concerned with the ways that conceptions of 'home' are stretched across imperial terrain when colonists settle elsewhere, and when colonized peoples move to other imperial sites, and Richard Phillips has studied the ways in which debates about prostitution, homosexuality, and the age of consent circulated between different sites within the British Empire. Stephen Legg situates the discourses and practices of urban governmentality in colonial New Delhi in the context of broader bundles of relationships both between Britons and Indians, and between Britons in India, and those elsewhere in the Empire.

These networked historical–geographical approaches to empire also include studies of the trajectories that individual subjects, such as governors, missionaries, and consorts, took across imperial space, transforming both their own subjectivities and, to a greater or lesser extent, each colonial society in which they dwelt. In part, their purpose is to understand the colonial past on its own terms, but in part also they are intended to encourage more spatially informed accounts of the multiple and diverse legacies of modern European colonialism, and of the potential for anti colonial practice that colonial interconnections allow.

Colonialism in the Present

It is not just in relation to the current 'war on terror' that colonialism is now being debated widely again. In various sites of the former British Empire, the relations that settlers forged with indigenous peoples in the nineteenth century are being re examined in the light of current political projects. In Britain, Niall Ferguson's recently published and televised survey of the Empire establishes it as a model of progressive, orderly, modernization of the globe. Ferguson argues that the current US administration would be wise to emulate it. For him the problem with the USA's overseas interventions is that they do not demonstrate the level of imperialist commitment to the long term restructuring of other societies that Britain accomplished over a hundred years ago.

In Australia, the bicentenary celebrations of the British First Fleet's landing in 1988 prompted both historians and aboriginal peoples' representatives to argue for the inclusion of Aborigines' experiences of genocide in the nation's official histories. A right wing backlash ensued. Keith Windschuttle has been the most uncompromising participant in the ongoing 'history wars', accusing Australia's professional historians of 'fabricating' histories of colonial brutality against aboriginal people to fit in with their general left wing bias. Windschuttle seeks to minimize estimates of the violent aboriginal death toll during the early nineteenth century, to exonerate British settlers of blame for what deaths occurred, and to blame nineteenth century humanitarians instead, for inventing separatist policies that have failed aboriginal and white Australians ever since. In New Zealand, the judicial process administered by the Waitangi Tribunal has had recourse to the greatest triumph of the early nineteenth century British humanitarians, the 1840 Treaty of Waitangi. The Treaty, or rather the various versions of it, are now seen as the foundation of a pact between Maori and the Crown, representing Pakeha (British settlers, their descendants and other nonindigenous New Zealanders) to which contemporary Maori can appeal in their claims for the restoration of land and resources.

What debates in each of these postcolonial states, and in others, have in common is a discussion of the essential morality or justice of relations between colonists and indigenous peoples in the past, and their implications for the present. Ferguson's narrative of the British Empire reads like a moral balance sheet designed to aid contemporary Britons coming to terms with their imperial past. The slave trade, the tolerance of famine, and the Amritsar massacre in India, for instance, count on the debit side, while the abolition of slavery, the construction of educational and industrial infrastructures, and above all the facilitation of free trade are credits that in the end swing the moral balance in Britain's favor. Overall, Ferguson suggests, British colonialism was 'a good thing' for the world and for its indigenous subjects.

For Windschuttle and his supporters in Australia, Britons and their descendants need to acknowledge no moral responsibility for the destruction of aboriginal societies. Indeed, the liberal ideas and institutions that British settlers brought to the continent, suggests Windschuttle, actually saved many aboriginal people from the inherent vices of their own society. Windschuttle argues that where Australia has gone wrong is in listening to humanitarians and aboriginal representatives who encouraged the perpetuation of aboriginal cultural difference despite the settlers' invitation of assimilation. In New Zealand, the existence of the Treaty of Waitangi and the general acceptance that its provisions were frequently breached by British colonists have not prevented a much more restrained but nevertheless detectable Pakeha backlash against the Maori assertiveness associated with the Tribunal and its attempts to make redress and reparation.

Finally, it is worth noting that each of these struggles over the colonial past (and it would be possible to proliferate the examples), whether they adopt a local, national, or global frame of reference, takes place in a context in which those groups formerly placed at a lower point in the hierarchy of civilization within European colonial discourses are still placed on the lowest strata of various indices of welfare today.

See also: Colonialism II; Empire; Imperialistic Geographies; Postcolonialism/Postcolonial Geographies.

Further Reading

Blunt, A. (2005). *Domicile and Diaspora: Anglo Indian Women and the Spatial Politics of Home*. Oxford: Blackwell.

Burton, A. (1994). *Burdens of History: British Feminists, Indian Women and Imperial Culture*. Durham, NC: University of North Carolina Press.

Cain, P. J. and Hopkins, A. G. (2002). *British Imperialism 1688 2000*. London: Pearson.

Clayton, D. (2000). *Islands of Truth: The Imperial Fashioning of Vancouver Island*. Vancouver: University of British Columbia Press.

Cooper, F. and Stoler, A. L. (eds.) (1997). *Tensions of Empire: Colonial Cultures in a Bourgeois World*. Los Angeles, CA: University of California Press, Los Angeles.

Evans, J., Grimshaw, P., Philips, D. and Swain, S. (2003). *Equal Subjects, Unequal Rights: Indigenous Peoples in British Settler Colonies, 1830s 1910*. Manchester: Manchester University Press.
Fanon, F. (1968). *Les Damnés de la Terre*. Paris: Maspero.
Ferguson, N. (2003). *Empire: How Britain Made the Modern World*. London: Allen Lane.
Gregory, D. (2004). *The Colonial Present: Afghanistan Palestine Iraq*. Oxford: Blackwell.
Hall, C. (2002). *Civilising Subjects: Metropole and Colony in the English Imagination, 1830 1867*. Cambridge: Polity.
James, C. L. R. (1963). *The Black Jacobins: Toussaint l'Ouverture and the San Domingo Revolution*. New York: Vintage.
Lambert, D. and Lester, A. (eds.) (2006). *Colonial Lives across the British Empire: Imperial Careering in the Long Nineteenth Century*. Cambridge: Cambridge University Press.
Legg, S. (2007). *Spaces of Colonialism: Delhi's Urban Governmentalities*. Oxford: Blackwell.
Lester, A. (2001). *Imperial Networks: Creating Identities in Nineteenth Century South Africa and Britain*. London: Routledge.
Manne, R. (ed.) (2003). *Whitewash: On Keith Windschuttle's Fabrication of Aboriginal History*. Melbourne: Black Inc. Agenda.
Ogborn, M. (2008). *Global Lives: Britain and the World, 1550 1800*. Cambridge: Cambridge University Press.
Phillips, R. (2006). *Sex, Politics and Empire: A Postcolonial Geography*. Manchester: Manchester University Press.
Sinha, M. (1995). *Colonial Masculinity: The 'Manly' Englishman and the 'Effeminate Bengali' in the Late Nineteenth Century*. Manchester: Manchester University Press.
Stokes, E. (1992). The Treaty of Waitangi and the Waitangi Tribunal: Maori claims in New Zealand. *Applied Geography* 90, 176 191.
Windschuttle, K. (2002). *The Fabrication of Aboriginal History, Vol. 1: Van Diemen's Land 1803 1847*. Sydney: Macleay Press.

Relevant Websites

http://muse.jhu.edu/journals/journal of colonialism and colonial history/
Journal of Colonialism and Colonial Histroy.
http://www.empiremuseum.co.uk/
The British Empire and Commonwealth Museum.

Colonialism II

D. Atkinson, University of Hull, Hull, UK

New Imperialisms, New Colonialisms, and New Political Geographies

The later decades of the nineteenth century and the first years of the twentieth century witnessed European colonialism accelerate in pace and expand in scale. By 1914, much of the world was shaped by this process and its attendant, entwined webs of trade, migrants, political power, and knowledge. Historians usually trace this acceleration to the 1870s when free trade imperialism gave way to the more protectionist, competitive colonial order of the later nineteenth century. Certainly, global economic instabilities from 1873 to 1896 propelled the industrialized economies into an increasingly urgent search for new sources of raw materials and new markets for their manufactured goods. Further, when the leading imperial powers met at the Berlin Conference (1884–85) in an attempt to preserve free trade, their inability to agree a formula led to a more trenchant 'new imperialism' whereby the European powers carved up huge swathes of territory as secure resource bases and markets. The Scramble for Africa, for example, left 90% of the continent under colonial rule. At a global scale, from 1885 onward, an additional 20% of the world's territory and 10% of its population were brought under direct European control. For some commentators, this 'new imperialism' was the apogee of European expansionism; indeed, for some, 1870–1914 constitutes the 'Age of Empire'.

This second articles on colonialism outlines this process briefly. It focuses upon the ways that space and colonial territory became finite in this period, and consequently, how they became prizes to be claimed, then mapped, surveyed, and delimited, and thereafter to be disputed and redistributed by the Western imperial powers. In particular, it outlines the growth of the political geographical theories and practices that in some instances prompted these projects, and in many cases facilitated and legitimated them. It is noted that the production of geographical knowledge was not a unified process in this period, and that there was some ambivalence and debate about the ties between geography and empire. In general, however, geographical knowledge and practices provided increasing service that assisted, explained, and facilitated colonialism. This serious intellectual support helped to produce an increasingly variegated mosaic of colonial territories from the 1870s. It also resulted in an ever more politicized geography, and a new subfield of political geography, that addressed this colonial world.

Colonialism Amidst Global Closure

The colonial world was far more fluid and manifold than can be outlined here; it changed and evolved as the world it co constituted became more globalized, interconnected, and enmeshed. In addition, far from being the uniform, coherent impress sometimes envisaged in Western imperial capitals, colonialism was contingent and experienced differently in each context as colonial policies were negotiated locally by the many groups involved. These complex and shifting matrices of culture, power, diplomacy, technologies, consensus, and oppression therefore produced numerous different experiences of empire, and an account of this length can only be partial. That said, this article highlights two complications that particularly impacted upon the political geography of colonialism from the 1870s.

One complication of the late nineteenth century was that the reenergized imperialism of long established powers such as France and Britain was augmented by the nascent imperialism of more recent states. After unification in 1871, the German state established territories in Cameroon, Togoland, Namibia, and Tanzania in the mid 1880s. Meanwhile, the Italians, unified in 1861, also sought their 'place in the sun' with a stuttering colonial campaign in East Africa from 1885. Elsewhere, longer established states also began forging empires on the back of their expanding economies. The United States seized the Philippines from the dissolving Spanish empire in 1898, while the Japanese began expanding into parts of East Asia in the same decade. Meanwhile, the Russian Empire consolidated its possessions across Siberia and central Asia. Indeed, colonialism became so commonplace that the Belgian King Leopold II felt able to develop a private empire in the Congo from 1885. Imperial ambitions, therefore, were spreading in scope and scale, and direct colonial rule was visited upon still more world regions as a consequence.

A second complication arose in the form of the 'closing world'. Some commentators worried that the era of European expansionism since the fifteenth century was threatened by a looming shortage of (what was imagined to be) vacant space. Africa was almost fully partitioned by 1891, and even the North and South Poles would be 'conquered' by 1911. For anxious observers, this crisis of space portended a world where states would have to coexist in an interconnected, interdependent system. Some episodes in the mid nineteenth century had emphasized the increasing connections between European states. The cross border contagion of the continent wide crash of 1847–48, and the subsequent working class uprisings in France, Italy, and Germany demonstrated this growing relatedness. New communications and transport technologies also imploded the experience of space, distance, and time. The 1869 opening of the Suez Canal, for instance, collapsed the travel time between Asia and Europe. It brought the 'Orient' 'nearer' to Europe, and was further evidence of this space–time compression.

These changes therefore entailed a radical shift in the human experience of space. Geographical imaginaries had to be reconceptualized and geography was forced onto national agendas. The responses of European governments – as they struggled to reassert control – revolved around territorial concerns. First, governments increasingly erected tariff barriers in attempts to insulate flailing economies from the unstable world economy. Second, the European powers constructed a complex matrix of military alliances that, some argue, eventually precipitated World War I. Third, as noted above, states old and new engaged in colonial expansionism in search of raw materials and new markets, while hoping that empire building might also unite (or at least distract) metropolitan populations. The changing context also demanded new geographical theories to help explain this closing world. These will be expanded further after discussion of another aspect of geography that flourished in this period.

Making Colonial Space with Maps and Boundaries: Practical Political Geographies amidst Global Closure

One aspect of geographical 'science' fundamental to patterning the world with colonies was the practice of cartography. This was significant because it was through maps, surveys, and boundary making that colonizing powers were able to demarcate, represent, and, by extension, know their colonial territory. In addition to constructing these political units practically, maps also helped to constitute colonies symbolically for audiences in Europe. They were among the most significant tools of Western imperialism, and demonstrate clearly geography's central role in imperial service.

From the mid 1970s, critics exposed geography's service to imperial militarism and expansionism, and there is now consensus that rather than see geographical knowledge as objective, detached data, we should regard it as an assemblage of power–knowledge central to the construction and governance of territory. The role of maps in this process rests partially upon their frequent use in the direct application of political authority. Their collation of information about topography, settlements, population distribution, and other data makes them crucial to the effective military suppression of territory. Thereafter, this same encompassing grasp of data helps facilitate the direct policing, control, and governance of space. Yet, these explicit political applications are often masked by the widespread, popular faith in the apolitical objectivity of cartographic 'science', and this latent power of maps augments their explicit applications. Consequently, maps and survey routinely became an instrumentalized part of colonial government.

Perhaps inevitably then, the new imperialism changed the nature of colonial cartography. In Africa, for example, earlier European maps of the continent were constructed primarily to locate trading stations and ports for commercial purposes. After 1884, they focused increasingly upon defining and delimiting newly bounded colonial territories – for cartography was the way of cementing knowledge of, and control over, overseas regions. The lasting significance of this process is evidenced by the modern maps of Africa that are still patterned by many of these colonial boundaries. Indeed, we have now lost sight of how significant the practical business of boundary making was. In this era of colonial rivalry, borders needed to be defensible and sometimes militarized too; likewise, establishing firm divisions between Self and Others was also of symbolic importance for colonizers. In Britain, key politicians, such as George Curzon, and leading figures in geographical institutions, such as Thomas Holdich, each theorized effective boundary making at length. Meanwhile, those working in the field busied themselves creating robust and durable borders by fitting them to lines of latitude and longitude, or to natural features such as ridges, rivers, or lakes. Indeed, boundary making was seen as a significant, applied geographical practice, and record of this work can be found throughout the geographical journals of the period.

Yet the problems with grafting Western concepts of bounded nation states onto African landscapes were legion – particularly as the cartographers and boundary makers seldom paid attention to extant African territorialities or traditions of conceptualizing political space. As a result, nations were divided, and communities were subject to changes of imperial rule at a stroke. But again, the subsequent longevity of these alien territorial models is in part testament to the power of maps. For these colonial boundaries were established by an apparently accurate, irreproachable 'science', and once these newly surveyed lands were enshrined within cartographic documents as unified, European colonial territories, they would be reproduced as such increasingly, accruing ever more credibility over time. By contrast, the indigenous geographies and place names that these Western maps erased were increasingly forgotten. *In extremis*, imperial mapping flattened and buried local geographies, instead producing maps that portrayed desocialized, vacant lands ripe for colonization.

The disregard for African territorialities reached its nadir when colonial territory became a diplomatic currency traded by imperial powers after their conflicts elsewhere. For instance, the border between British Nigeria and the German Cameroons altered three times between 1886 and 1919. First, a geometric border was established as the British and Germans scrambled for African territory. Second, German expansionism from 1901 to 1906 established new 'natural frontiers' related to geographical features. Finally, after World War I, as the British and French squabbled over the partition of this German colony, a new border was: "…drawn in a casual way with a blue pencil… by a diplomat [in London] who knew nothing of the lands and peoples he was dividing" (Yearwood, 1994: 218). Again, for local communities, this arbitrary process entailed new rulers, new citizenship, and revised colonial burdens. Worse still, their creeping symbolic power meant that each iteration of these maps would further subsume local geographical knowledge. As many Europeans also saw the practice of cartography as another marker of their superiority over the colonized, map making visited further slights upon local communities.

The full complexity of the relationship between cartography and empire cannot be outlined here, but it should be clear that both the process of mapmaking and the maps themselves were significant elements in constructing the colonial world. Further, the belief that geographical experts – with maps and data to hand – were among the best placed to construct new states after conflict persisted through to the mid twentieth century.

Theorizing Political Geographies and Colonial Expansionism amidst Global Closure

Late nineteenth century Europe also saw geography established as an academic discipline in the main imperial states. By resonating with the prevailing Darwinian *zeitgeist*, geography positioned itself as the science that explained the differing connections between society and environments across space. It did this above all via theories of environmental determinism. Studies that mapped racial groups and related their supposed social development to differing geographical conditions became central to colonial discourse. But another legacy of this period was that many of these new professional geographers were keen to demonstrate both their practical reach and their political relevance to the states that funded them. One such public service was the attempt to conceptualize and explain the changing context of the closing world. A distinctive product of this response was the development of political geography – a perspective that focused upon questions of territory, state, and nation. Mindful of their obligations and despite their supposed objectivity, proponents of this new political geography often sought to justify colonial expansionism; likewise, others proposed ways to reorder the world after conflict.

For instance, the 1890s found Friedrich Ratzel, the figure who established German academic geography, worried about his country's future in a world of diminishing space. He proposed a way of responding to this context with his theories of the 'organic state' and *Lebensraum* ('living space') in his 1897 book *Politische Geographie*. Here, Ratzel moved beyond orthodox legalistic notions to theorize states as organisms that, given natural selection, had to expand and devour the space and resources of adjacent states or be consumed themselves. For Ratzel, because the Germans were a 'strong' people due to their long established roots in the German soil, they would inevitably expand to attain their appropriate *Lebensraum*. Indeed, this process could be read as the natural, biological law behind state growth, with colonialism revealing the Darwinian order of states. Yet Ratzel's theories were also inflected by his nationalism and his status as a colonial agitator keen to see Germany become a continental and colonial power. These interests ensured that Ratzel's theories were soon mobilized to provide sustained intellectual support for German colonialism in Europe and beyond.

In Britain, this changing world also agitated Halford MacKinder – the founding figure of British geography. He mirrored Ratzel's inseparable interests in geographical education and politics: he regarded 'thinking geographically' as essential to the continued health of the British Empire. He sought to educate Britain's elite about the looming threats of the closing world in his paper: 'The geographical pivot of history' in 1904. This article identified the potential of recent over land transport technologies (particularly railways) to integrate the vast resources of North, West, and Central Asia. As this extensive region was invulnerable to British maritime power, it had the potential to expand relentlessly, and unimpeded, throughout Eurasia and Africa, to eventually conquer the globe. Mackinder's theory was not only a warning for the British Empire, but it was also a way of rethinking the changing world by considering its intersections of space and politics, or better, its political geography.

Innovative though they were, these Eurocentric theories assumed that Western powers enjoyed a remit to intervene globally. They also saw colonial empires as a solution to the challenges of the closing world. A few dissenting voices objected to these cosy links between geography and imperialism. The Russian anarchist and geographer Petr Kropotkin complained that the discipline should be promoting cooperation between peoples rather than Darwinian struggles over territory. Other leftist geographers in France, Ireland, Italy, and the UK also worried about geography's state service. These concerns, however, were outweighed by the majority of geographers who put their scholarship at the service of expansionist states – a process that accelerated further in the twentieth century.

Reconfiguring the Colonial World: Redrawing Boundaries with a Politicized Geography

By 1914, one third of the world's population and a half of its territory were under colonial rule. Consequently, the European wars of the early twentieth century infected regions much further afield, and colonial troops died for 'mother countries' they had never seen. More specifically, World War I led to colonies being transferred – again, as a diplomatic currency – from defeated to victorious powers at the post war Versailles peace conferences. According to US president Wilson – the most powerful voice at Versailles – the world was to be reconfigured along lines of national self determination. This did not apply to those colonies deemed unready for self governance, however. They became League of Nations mandates overseen by established imperial powers. The Ottoman Empire's Middle Eastern colonies were divided between the British and the French, for example. Here was the political geography of colonialism writ large as borders were redrawn and colonies were mandated to new rulers overnight.

The modern state of Iraq finds its origins in this process. In 1919, the Versailles diplomats mandated the erstwhile Ottoman territory of Mesopotamia to the British, who had been developing their presence in the oil rich region since 1914. So alien was the initial administration imposed by the British, however, that both the Shia and Sunni communities united in armed resistance. The British quelled this opposition, and in 1921 proclaimed a new state that forced together the Sunni communities around Basra in the south, with the Shia regions around Baghdad in the center of the new territory. The Kurdish lands to the north were also incorporated, as was the region around Mosul in 1925, after a lengthy border dispute with Turkey. These different ethnicities were contained within die straight stretches of boundaries that betray their construction by British cartographers. In theory, the tensions between these groups would ease with time: 80 years on, they provide fault lines that appear to be unraveling – despite the West's continued belief in its right to intervene in the region.

The colonies were secondary to the main business of Versailles, however, which was the reconstruction of Europe. This process also involved territorial realignments as the shattered Austro Hungarian and German Empires were stripped of territories, and a series of new states were created. In part designed to contain any future German resurgence and to buttress Europe against the Soviet Union, these new states were also supposed to provide national self determination for the peoples of Central, Southern, and Eastern Europe. Once again, politicized geographers were involved in constructing new borders and states. Isaiah Bowman, director of the American Geographical Society, became the key player in the House Inquiry: the body that furnished the US diplomats with details, statistics, and maps as they sought to construct the world that Wilson's idealist agenda demanded.

At face value, the House Inquiry provided objective, scientific geographical knowledge: independent expertise that was entrusted with determining national destinies. In practice, this was not always so. When Bowman needed data to underpin the US plans for the former Austrian–Hungarian Balkan territories, for instance, he commissioned the respected Serbian geographer Johan Cvijic to provide it. Cvijic produced a series of maps representing the region's population density, and its religions and languages (which were considered markers of ethnicity). These data helped to underpin the creation of Yugoslavia: a new state proclaimed as a coherent ethnic mix, despite the Croat, Albanian, Bosnian, Montenegran, and Slovenian populations alongside the majority Serbs. Yet, Cvijic had been mapping these populations since 1906, and had gradually exaggerated the distribution of Serbs, while downplaying the scale of the Albanian and Macedonian populations to the south of Serbia. Nevertheless, Cvijic's skewed representations resulted in a Serbian dominated state. As we saw in the 1990s, after the demise of Cold War certainties and the death of long standing Yugoslav President Tito, this imbalance and the long subsumed ethnic rivalries of the former Yugoslavia prompted civil war.

These brief vignettes demonstrate how the Versailles conferences, often relying on geographical expertise, refashioned colonial territories. The diplomats hoped they were establishing a stable, peaceful world order. The perceived slights of the settlements, however, prompted revisionism from some of the totalitarian states that also emerged from the chaotic aftermath of World War I.

The Political Geographies and Accelerated Colonialisms of Inter-War Totalitarianism

The unparalleled slaughter of the Great War begot consequences beyond its stark demographic impact. So encompassing was the trauma that increasing numbers began questioning long established social and political models. The economic downturn from 1929 yielded additional damage and prompted further radical thinking. While liberal democracy struggled onward in most imperial powers, communism promised different solutions in the Soviet Union. Elsewhere – notably in Italy and Germany – fascist, totalitarian regimes emerged with alternative revolutionary programs. Both the Italian and German regimes were nationalistic, bellicose, and felt cheated of their territorial rights at Versailles. Subsequently, both developed clear foci upon territory and engaged in sustained, accelerated colonial expansionism. These two contexts also catalyzed Europe's largest geopolitical movements, which both supported colonial aggrandizement intellectually. This article continues by outlining the theories and practices of these totalitarian colonial geographies briefly.

The strictures imposed on Germany at Versailles were seen as vindictive by many. Karl Haushofer, a demobbed general who returned to civilian life as a Munich geography professor, was particularly irked by the Allies' use of geographical knowledge at the conferences. In response, he was determined to apply geographical knowledge to solve Germany's problems, and he thus developed a *geopolitical* perspective that melded the political and the spatial to explore the inter war flux of boundaries, territory, and resources. Articulated through his journal, *Zeitschift für Geopolitik*, this nationalist, applied perspective asserted Germany's right to reintegrate its territory and people, to develop the nation's geographical imagination, and to advise on its territorial reorganization. It also justified a renewed German empire overseas and in Eastern Europe by adapting Ratzel's 'organic state' and *Lebensraum* theories. With these theories pushed to their limits, *geopolitkers* proposed a vast German empire encompassing a North–South sector of the Earth that included all global environments. This empire would consist of a colonial hinterland in Africa and Scandinavia that served the industrial heart of Germany. Parallel pan regions in Asia and the Americas were also theorized as self sufficient territorial blocks. The plan, therefore, was for global scale colonial empires in the future.

Haushofer's geopolitics augmented other forms of applied science under Nazism that also called for renewed German imperialism overseas and in Eastern Europe. These ideas were realized from 1937 onward as Nazi expansionism ripped through Central and Eastern Europe, and into the Western Soviet Union. In the latter regions especially, the Nazis planned a distinctive accelerated colonialism whereby, after military conquest, these areas would be cleared of the peoples deemed alien to the Third Reich. The dark histories of this killing perhaps constitute the nadir of European expansionism. Eventually, Nazi ambitions envisaged the construction of German colonial landscapes. Although facilitated by the modern disciplines of area research and spatial planning, these were to be socially conservative landscapes of rural towns and settler colonies stocked by soldier farmers with large families. The aim of this 'total planning' was for 20 million ethnic Germans to be consolidated within the Reich by the late twentieth century. They would work the 'German' soil of Eastern Europe and constitute an 'East wall' against Bolshevism. Ironically, however, the maps and statistics produced by German geographers to support this vision were eventually used to partition Germany after 1945.

In Fascist Italy, geopolitical theories also developed in response to the Versailles conferences and their perceived failure to rectify Italy's paltry share of the colonial world. In 1930s' Trieste – itself a border city transferred from Austria in 1919 – political geographers Giorgio Roletto and Ernesto Massi developed an Italian version of geopolitics amidst the wider politicization of Italian geography under Fascism. Their monthly journal *Geopolitica* (1939–42) declared a unique, synthesizing perspective. It claimed to encompass all elements of the unstable modern world, while offering 'scientific' arguments for expansionism with this 'geographical doctrine of empire'. Such ambition resonated with Fascist Italy's wider colonial ambitions – not least because Italy was a relative latecomer to colonialism. Initial adventures in Eritrea and Somaliland from 1885 were inconclusive; later, an 1896 defeat at Adowa in Ethiopia was disastrous. Nevertheless, the desire to rank alongside other European powers prompted successive governments to nibble at the few remaining slivers of Northern and Eastern African territory. In 1911, Italy invaded Ottoman controlled Libya's coastal regions, and, after 1918, started to extend their rule southward into the Sahara. Under the belligerent Fascist regime from 1922, this process accelerated.

In Libya, for example, this entailed a threefold process of conquering the space and population militarily, then mapping and surveying the territory, and finally, settling a new colonial landscape with Italians. The military conquest involved using modern warfare against the indigenous resistance. In Cyrenaica and the interior, this entailed a lengthy campaign that cost of over 80 000 Bedouin lives. Once this 'pacification' was complete, the Italians began mapping and surveying the region. Between the later 1920s and 1942, the state cartographic service had mapped almost the entire 1.8 million km^2 of Libya. As elsewhere in the colonial world, the symbolism of mapping newly conquered territory and deploying this modern, 'scientific' practice was acknowledged explicitly. Also, as elsewhere, the cartographers were augmented by a series of interdisciplinary surveys of the Libyan interior from the early 1930s. The Società Geografica Italiana – with its close ties to the state and its tradition of supporting field sciences – organized this self proclaimed 'colonial science'. Seven expeditions were despatched into the Sahara, and returned enough data for the Italians to construct a European version of the territory that could be comprehended by them as a governable colonial domain. Indeed, so explicit was geography's service to this accelerated colonialism that one historian labeled Libya "an authentic laboratory" of "geographical sci ence" (Del Boca, 1988: 272).

Finally, the regime created settler colonies along Libya's coastal regions. These were to be populated by veterans and their families who were housed in the same explicitly modern, spatially ordered settlements that had been developed in the 'internal colonization' of the Pontine Marshes to the South of Rome. Save for those voices recovered by Ali Abdullatif Ahmida, we know little of the indigenous perspectives on this late European colonialism, and we need to explore these further. Similarly, we need to learn more about the lesser studied colonialisms, such as the Italian, Portuguese, or Dutch experiences, to broaden our appreciation of the whole imperial era.

Decolonization and Rethinking Colonialism

The Spanish Empire began to disintegrate in the nine teenth century, while, as noted above, the Ottoman, Austro Hungarian, and German empires were dissolved in 1919. Decolonization in the remaining European empires gathered pace in the inter war years as the growing opposition of the colonized surfaced amidst the period's wider turbulence. As the shattered economies of the Western powers reeled after World War II, the pro cess accelerated and was largely complete by the late 1960s.

There were some exceptions to this process although, in many respects, these demonstrate how imperialism changed and adopted more amorphous forms after 1945. Both superpowers of the Cold War period maintained relations with other states that can be considered colonial (or neocolonial) ties. The Russian Empire – as the Soviet Union until 1991 – retained its post 1945 Eastern European territories into the 1980s. It also developed a distinctive colonial socialism that extended into these 'satellite states'. Conversely, the United States of America entered 1946 with a growing economy and an interest in developing its overseas influence. The neocolonial hold that the US had maintained over South America since the nineteenth century was now extended to other world regions. Arguably this was US colonialism by stealth: a shift from discredited territorial possession toward the 'informal imperialism' that rendered the world as a market for the US goods. Indeed, Western development and aid policies aimed at the Global South attracted al legations of exploitative neocolonialism from depend ency theorists from the 1960s onward. The ideological, economic, and military contest between the superpowers also trailed a series of conflicts in some former colonial possessions and satellite states. In these respects, coloni alism did not fade away with the decline of formal im perialism, it simply changed its form and remained a significant actor in global affairs.

In the twenty first century, a growing debate explores the nature of contemporary empire. One position echoes the point above, and holds that formal imperialism has transmuted into new, flexible forms of capitalism that extend around the globalized world. But others insist that this deterritorialized world – with its flows of trade, fi nance, peoples, and cultures – still struggles with rem nant structures of territoriality, borders, and the regulation of bodies within these frames. Hardt and Negri highlight the tensions within this emerging global geopolitical framework as the focus of sovereignty shifts from state centric perspectives to more decentred, fluid, biopolitical foci. Yet others insist that neocolonialism retains the tendency to install social divisions between peoples and places in our contemporary world.

Postcolonial studies emerged in parallel with de colonization. This project sought to undermine the nat uralized assumptions of colonial discourse, to recover colonial voices and perspectives, and to complicate our appreciations of colonial experiences. It also sought to decenter Western narratives of the colonial era. Histor ically, for example, academic interpretations of imperi alism had often been resolutely Eurocentric. Influential early commentators such as Hobson and Lenin read colonialism as something visited upon the under developed world as a consequence of expanding Western capitalism. Many post 1945 Western models for the Third World echoed this ethnocentrism: designing states that mirrored their former colonial masters. Likewise, modernization theory – and its associated aid programs – assumed that the social, political, and economic devel opment of former colonies should also replicate prior European experience.

Rather, postcolonial perspectives and new imperial histories have been eager to emphasize how colonial processes impacted upon both colonial realms and Metropolitan lands. They have exposed the myriad hy brid cultures constituted relationally by empire and that remained hidden between the blunt binary divisions of European colonialism. They also address the variations of colonial experience in different places, and how colo nialism's influence crept into domestic spheres, impacted upon individual bodies, and how it stretched into the biopolitics of population management and control. In creasingly, and at their best, postcolonial approaches are also sensitive to how colonial experiences were inflected by race, gender, class, and other subjectivities. To this end, scholars have focused upon retrieving the voices and experiences of all corners of colonized subjectivity and, increasingly of late, exploring the experience of post coloniality – of being shaped by colonial cultures.

Geographers can serve this debate with discussion of the shifting spatialities of colonialisms past and present, of the changing natures of territory and boundaries, and of the roles of their discipline in representing, calcu lating, constructing, and reconfiguring the colonial world. This discussion of the political geographies of colonial ism is curtailed by space and this article is unable to stretch much beyond the usual categories of late Euro pean imperialism. Other colonialisms from beyond Western Europe remain under researched, and the study of colonial modernity in each different context requires further attention. We also need more reflection upon the countless ways that the cultures of colonialism echo into our postcolonial present – that is, recognizing that co lonialism did not simply happen in distant times and distinctive places, but produced much of our modern world.

See also: Colonialism I; Geopolitics; Political Geography.

Further Reading

Alhmida, A. A. (2005). *Forgotten Voices: Power and Agency in Colonial and Postcolonial Libya*. London: Routledge.

Atkinson, D. (2005). Creating colonial space with geographies and geopolitics. In Ben Ghiat, R. & Fuller, M. (eds.) *Italian Colonialism*, pp 15 26. New York: Palgrave MacMillan.

Del Boca, A. (1988). *Gli Italiani in Libia: Dal fascismo a Gheddafi*. Milan: Mondadori.

Duara, P. (ed.) (2004). *Decolonization: Perspectives from Now and Then*. London: Routledge.

Gregory, D. (2004). *The Colonial Present: Afghanistan Palestine Iraq*. Oxford: Blackwell.

Hardt, M. and Negri, A. (2000). *Empire*. Cambridge, MA: Harvard University Press.

Harley, J. B. (2001). *The New Nature of Maps: Essays in the History of Cartography*. Baltimore, MD: John Hopkins University Press.

Heffernan, M. (1998). *The Meaning of Europe: Geography and Geopolitics*. London: Arnold.

Jan van Pelt, R. and Dwork, D. (1996). *Auschwitz: 1270 to the Present*. New York: Norton.

Kearns, G. (1993). *Fin de siècle* geopolitics: Mackinder, Hobson and theories of global closure. In Taylor, P. (ed.) *The Political Geography of the Twentieth Century: A Global Analysis*, pp 9 30. London: Belhaven.

Lacoste, Y. (1975). *La Géographie, ça sert, d'abord, à faire la guerre*. Paris: Maspero.

Livingstone, D. N. (1992). *The Geographical Tradition: Episodes in the History of a Contested Enterprise*. Oxford: Blackwell.

Ó Tuathail, G. (1996). *Critical Geopolitics: The Politics of Writing Global Space*. London: Routledge.

Smith, N. (2003). *American Empire: Roosevelt's Geographer and the Prelude to Globalization*. Berkeley, CA: University of California Press.

Wolfe, P. (1997). Imperialism and history: A century of theory from Marx to postcolonialism. *American Historical Review* 102, 338 420.

Yearwood, P. (1993). 'In a casual way with a blue pencil': British policy and the partition of Kamerun, 1914 1919. *Canadian Journal of African Studies* 27, 218 244.

Colonialism, Internal

R. J. Das and S. Chilvers, York University, Toronto, ON, Canada

Glossary

Debt Peonage A process in which moneylenders force their debtors to work for them for wages which are not market determined. Debt peonage is a form of unfree labor.

Proletarianization It is the process in which those who perform direct labor are gradually separated from means of production such as land and eventually become property-less, and are therefore economically forced to work for capitalists for a wage.

Scheduled Castes These are communities in India who traditionally occupy the lowest stratum of the caste hierarchy (or even outside of the caste hierarchy). The Indian constitution has granted them some limited special rights.

Serfdom It is the condition of peasants who are tied to specific landlords and must work on their farms.

Surplus-Labor Appropriation It is the process in which those who control means of production, such as land, make workers/peasants perform excess labor above what would be required if these classes owned means of production.

Definition

Internal colonialism refers to a complex set of sociospatial relationships of exploitation and domination that characterize certain culturally distinct populations residing in sovereign, mainly Third World, societies. The two defining relationships of internal colonialism are that subjected populations are (1) exploited principally by mechanisms which may not be ‘capitalist proper’ and (2) institutionally dominated, both politically and culturally. In the first instance, internal colonies overwhelmingly produce primary goods for metropolitan markets, constitute a source of cheap labor for capital controlled from outside the colony, and/or constitute a market for the products and services from metropolitan areas. Subordinated populations part with their surplus labor through coercive social relationships, such as slavery, serfdom, debt peonage, and/or a general unequal exchange of commodities. Thus, surplus labor is extracted from the colonized population by mechanisms that largely differ from those found in advanced capitalist societies (where exploitation occurs through the employment of ‘free’ waged labor). The second trait of internal colonialism is that these ‘noncapitalist’ forms of surplus labor appropriation necessitate forms of political and cultural domination of a qualitatively different nature from that exercised by ruling classes in metropolitan capitalist society. The population of an internal colony as such faces severe discrimination in civil society, violence (as broadly defined), and restricted access to the state resources.

Conceptual Origins

The concept of internal colonialism has been applied to a variety of geopolitical contexts. In some instances, populations are said to be colonized by the entire state of which they are a constituent part and at times by dominant regions of the national state. Case studies have often been drawn from rural Third World locales, though arguments have been made for its applicability to spaces within the metropolitan heartlands of global capitalism. The theoretical geographical scope of the concept has thus been substantial. What is generally accepted is that the explication of internal colonization as a coherent concept – incorporating both the ‘economic’ dimensions of regional exploitation, and the ‘political’ aspects of social group domination – has its origin in radical Latin American scholarship. The concept was pioneered notably by Pablo González Casanova, Rodolfo Stavenhagen, and others as a part of a renewal of Marxism in Latin America during the 1960s, to explain the social problems of postcolonial societies at least partly in terms of their colonial heritage. Persistent poverty in the Third World societies could not, however, be wholly blamed on external mechanisms of colonialism, particularly after formal independence. Possibly the reason for internal colonialism cohering as a theory in Latin America, decolonized much earlier than other parts of the Third World, is that internal domination and exploitation of indigenous groups by other national groups was strikingly evident in the postcolonial period. Drawing upon histories then emerging of Latin America’s role in the development of capitalism in the West, particularly upon what became known as the dependency school, many writers began to draw parallels between the present and the colonial past. Theorists of internal colonialism rejected both the then influential dualist model of development (positing parallel ‘modern’ and ‘traditional’ economies within a national space), and bourgeois structural convergence models (convinced of the inevitability of the emergence of integrated national economic spaces). The internal domination and exploitation of ‘natives by natives’ (as González Casanova termed it) was highlighted instead, in particular the instrumental internal class relations that had emerged between different social groups and regions. It was also argued that internal colonialism was a specificity of postcolonial societies with large indigenous populations.

In this alternative scenario, the 'internally colonized' regions were envisaged as integral and necessary to the functioning of the national economy, and reference was specifically made as to how colonialism had wrought a sociospatial division of labor, imposing cultural distinctions and stratifications along class lines. In the specific case of Latin America, this division came between the Iberians, including Mestizos (populations of mixed European and indigenous descent) and indigenous peoples. However, in other contexts, such as in South Asia, the division was seen not so much as between indigenous and mixed descent populations, but between tribal and low caste populations and those who were upper caste, relatively Westernized, and English educated. Such divides, which had persisted since the days of colonialism, had earlier concerned a general category of writers, who were to have an antagonistic influence upon the development of internal colonialism theories. In the lead up to independence from the colonial powers, and in the postcolonial aftermath, various educational and philanthropic movements attempted to create 'authentic' national cultures in opposition to what was seen as the alien culture of the conquerors. In Latin America this was typified by the Indigenismo movement, composed largely of urban Mestizo intellectuals, who were fixated on recapturing the indigenous 'glories' of a pre Columbian era. The critique that emerged of this phenomenon, and which informed theses of internal colonialism, was of the perspective that the question of the fate of dominated populations could not be resolved by humanitarianism since the persisting exploitation was not a question of morals, culture, or ethics, but was a problem rooted in the materiality of the exploitative socioeconomic system itself.

Mechanisms of Internal Colonialism

Many scholars define the condition of internal colonialism as a situation whereby one distinct cultural/ethnic/racial group economically exploits and politically dominates another such group. In this view, cultural/ethnic/racial oppression is seen as irreducible to other social processes, such as class. The mutual relationship between 'economic' class exploitation and, for example, 'political' racial oppression is often, however, not clarified, and the term 'internal colonialism' is itself used loosely. Scholars broadly influenced by the Marxist thought (e.g., Wolpe, as well as Drakakis Smith, a geographer, and Hartwig, who are influenced by him) have attempted to directly theorize situations of internal colonialism to class relations, and in particular to the articulation between capitalist and non capitalist class relations. The argument is as follows: in certain conditions, within a country, capitalism may develop predominantly in relation to noncapitalist modes of production. Capital in such circumstances has a double tendency – to conserve as well as to dissolve noncapitalist social relations. This relationship with noncapitalist modes of production varies geographically. In terms of relations of dissolution, in some areas, capital may destroy precapitalist modes of production through primitive accumulation – that is, it may dispossess peasants of their means of production in order to create a property less class of wage laborers. In terms of relations of conservation, there are two possibilities: the extraction of the products of human labor or the extraction of human labor itself from noncapitalist areas. In one area, capital might extract goods produced under noncapitalist modes of production, for example, through sheer plunder, or the exchange of nonequivalents, or in a similarly forcible manner. The ethnically dominant groups, such as the Mestizos (mixed European and indigenous population) in Latin America or upper caste merchants from the plains in India, extract surplus from the internal colony by purchasing commodities cheap and selling dear and by charging usurious interest rates.

In other areas, capital may extract not the product of human labor but the capacity of humans to work, 'labor power' itself. This happens when laborers from non capitalist locales work in capitalist areas (e.g., as seasonal migrants). Since part of the cost of reproducing and sustaining this workforce is borne by noncapitalist modes of production (and not by capital), wages are lower than what they would otherwise be. Capital, simply, does not need to pay a wage sufficient to cover the reproduction of its living workforce. This is the case when, for example, land in the internal colony is owned and cultivated communally and/or where the produce is shared through ties of kinship or mutual obligation. This process thus contributes toward the reproduction of labor periodically working outside the spaces of the internal colony and keeps wages down for capitalist employers. It must be noted that whether the internal colony contributes goods directly, or the labor necessary to produce commodities, apart from economic mechanisms, the element of extra economic coercion is regularly used by owners of means of production. Additionally, in the realm of commodity and money circulation, various forms of unequal exchange between the internal colony and the advanced regions contribute to the stunting of the forces of production possessed by populations of the internal colony, and a general re capitalization of the region. With a mixture of different methods of production in the colony, the forces of production in the internal colony remain largely undeveloped. Since local or autonomous capitalist development is absent, the surplus extracted from the colonized population is rarely reinvested back into the territory, and investments of significance are generally unsustainable in terms of their long term economic, social, or ecological viability.

Spaces of internal colonialism thus come to be characterized by dependence on a restricted range of economic activities, mostly in the primary sector, and the region consequently suffers from a lack of diversification and, when technology does happen to be advanced, extreme degradation of the physical environment. Historically, with few livelihood alternatives, and as traditional occupations come under pressure, there is out migration and long term population declines occur (though interspersed with short term, localized boom periods, for example, as associated with prospecting for finite natural resources). With a lack of economies of scale, and due to political considerations, such as the weak enforceability of private property rights and sociospatial discrimination, the provision of various (private and state) services and investment is limited. This increases the pressure on smaller communities and creates uneven development as populations attempt to consolidate resources within the confines of the colony.

The human and physical geography of internal colonialism can thus be understood partly as the expression of underlying and contradictory social relations. However, insofar as the existence of an internal colony involves economic exploitation and social oppression, this requires political legitimization in the context of broader society – or else eventually becoming untenable over time. Exploitation, though, is legitimized not through state and ideological mechanisms based on private property, as characteristic of liberal capitalist democracies, but through political and ideological domination based firmly upon purported racial, ethnic, or national traits, that is, through a relationship that is primarily colonial in nature. This is the case, whether it is in the Third World, or in the First World settler colonial societies with substantial aboriginal populations. It is the essence of internal colonialism: a set of sociospatial practices of domination that 'justify' the ways in which capital benefits from and develops at the eventual expense of 'noncapitalist' modes of production. The fact that internal colony areas are less developed than the country as a whole requires justification, and here non class social relations are incessantly invoked. In other words, it is made to appear that the region is less developed because of, for example, racial or ethnic traits. Once these differences are used to justify the relations of exploitation, unequal exchange, and unequal geographical development between internal colony spaces and those pertaining to other areas, then discrimination reinforces these conditions. Purported differences, such as of race and ethnicity, thus obtain a material foundation for their sustenance.

To the extent that internal colonial relations are simultaneously about class and colonialism/ethnic oppression, dominated populations come to possess a contradictory identity, one that can be argued to have adverse implications for developing the collective agency necessary to counter internal colonization. On the one hand, populations see themselves as colonized, different from wider society, and, on the other hand, as members of an exploited class, in the sense that they are located in a typical exploited class situation. Since the 1960s, there has been a debate as to the influence of internal colonialism on the development of class consciousness versus other forms of consciousness. Stavenhagen proposed that the colonial relationships between indigenous communities and the larger society strengthened such marginalized communities and fomented their ethnic identity, while the development of capitalist relations in the internal colony tended to contribute to their disintegration as a community and integration into national society. In other words, ethnic consciousness makes the lines of social struggle more complex, as dominated populations may at times fight as antagonistic classes, and at times as oppressed cultures. Additionally, given that those classes and political groups ruling in the internal colony may not be dominant at the national scale (e.g., Mestizos in Latin American countries being subordinate to Iberians and landed interests in India being subordinate to the bourgeoisie) the population of the internal colony may struggle to integrate immediate tactical struggles with visions of temporally longer and spatially wider social change. Internal colonialism thus gives a distinctive character to class relations, techniques of production, and the class struggle in those countries where this phenomenon exists. Indeed, the phenomenon cannot be properly understood without reference to these dimensions.

In this respect, relations of internal colonialism differ from, for example, conventional urban–rural relations because they have a different historical origin and are based on active discrimination. They also differ from class relations as the sociospatial demarcation of the internal colony cuts across class lines. In this way, internal colonialism can be distinguished from general regional inequality under capitalism. The major difference between marginalized regions and the defined internal colony is found in the application of the different institutionalized practices of domination and methods of social control. The population of internal colonies is thus subject to discriminatory practices over and above those characteristic of relations between dominant classes and typical regional working classes. Furthermore, individuals from working class backgrounds in peripheral regions have opportunities for social mobility depending upon opportunity structures to the limited degree that they can shake off their class origins and become socialized into the skills, values, and attitudes of the mainstream society. By contrast, the individual from the internal colony is highly socially constrained, regardless of personal merit.

Whither Internal Colonialism?

Over time, relations and processes of internal colonialism can change and weaken. One fairly common line of argument is that dominant social–cultural groups always have a vested interest in maintaining relations of internal colonialism. However, this may not always be the case. Historically, as capitalist relations have developed and penetrated into remoter regions, class relations predicated on the accumulation of capital have entered into conflicts with colonial style relations based on mercantile relations or on exploitation through noneconomic forms of domination. To the extent that internal colonialism is based on the appropriation of unfree labor of a population, and to the extent that free wage labor may be more appropriate (i.e., more profitable) than unfree labor, then those specific conditions defining a situation of internal colonialism may weaken. In any case, even if unfree labor is used, it is not necessary that colonial political relations must prevail. In such scenarios, it is mainly those enterprises of the ethnically dominant whose technology has lagged behind and who continue to use cheap forced labor that may be interested in maintaining internal colonial relations, and, as such, there is ideological struggle within the dominant community over this matter.

Further, the weakening of internal colonialism may occur through class differentiation within the colonized community, consequent to the development of petty commodity production in rural areas. In particular, the rural petty bourgeoisie may forge an alliance with urban workers, socially marginalized groups in urban areas, and colonized groups in an attempt to break the patron–client relations in use by the dominant classes. Political struggles for land, for better working conditions, and for full political participation also can help contribute to the demise of internal colonialism, and need not always be at odds with the fuller development of capitalism (e.g., as historically associated with land reforms or with the establishment of parliamentary representation). Populations from the internal colony may, too, adopt those lifestyles associated with the ethnically dominant as a means of breaking out from their communities in a search for employment elsewhere. This process involves a combination of forced circumstances and choice, proletarianization, and rural–urban migration that contributes to the breakdown of internal colonialism. State led agrarian reforms, including public investment in internal colonies, can also lead to the closure of the inequality gulf between colonized populations and the national society.

Hence, on the one hand, the dynamics of internal colonial relations may become transmuted into those typifying other situations, such as of a capital/labor class relationship and a more conventional form of regional marginalization. Such developments may, though, mitigate against the demise of internal colonialism. For example, the national state by seeking to either redistribute land, manipulate guaranteed crop prices, divulge increased political autonomy, and/or invest in internal colony areas with public funds might rouse political opposition by nature of an alliance between rural land magnates in the colony and urban working classes (who might fear disinvestment in their neighborhoods and services or else in the rising price/insecurity of food supplies). The further penetration of capitalist–social relations might perhaps lead to lower sections of the ethnically dominant working class mobilizing to ensure social exclusion of colony populations from privileged openings in the social structure, such as in traditional commercial employment opportunities (consider the problem of racism in various trade union contexts) or positions in the bureaucracy (witness contemporary struggles over government quotas for scheduled castes and tribes in India).

Other developments can emerge to change the relations of internal colonialism in sometimes unanticipated ways. Attempts to impose forms of social control – such as the religion of the dominant community – upon the internal colony may, for example, act as a spur to cultural resistance. Similarly, extreme measures of social control in the internal colony, sanctioned by the prejudices of the national population, may result in a backlash and the radicalization of movements. In this case, colonized populations may challenge the hegemonic order in the name of a disadvantaged class, on the one hand, and in the name of a cultural minority, on the other. Social and cultural institutions that develop organically within the internal colony, and in competition with state supported institutions, may hereby become nodes for the crystallization of political opposition. This may either create a dual basis for a challenge against the hegemonic order in society or become a feature for co optation – the state making certain cultural and some (limited) political concessions, but without altering the fundamental basis for surplus labor transfers from the internal colony.

Critical Assessment

The theory of internal colonialism is not without deficiencies, and there are four main criticisms. First is the general understanding of the development of capitalism presented by the concept of internal colonialism. An implicit assumption of the internal colonialism theory is that peripheral regions in a given country will always be necessary to fuel the capitalist system, the absence of which would engender national economic crisis. What is disregarded is the possibility of the important role played by technological change in creating relative surplus value (in the advanced capitalist center of the economy), particularly after the historical phase of the dispossession of populations of their means of production has ended. There is therefore an undue amount of stress on absolute surplus value as it is produced by the native population and on its appropriation by those culturally and politically dominant groups. Additionally, to the extent that internal colonies in the Third World might be judged as historically and geographically unique, in comparison to the assumed origins of capitalism in the First World, it downplays the importance of the comparable overseas imperial relations that the emergence of Western industrial capitalism depended upon. In certain narratives of internal colonialism, there is, thus, a potential tendency toward economic determinism.

A second criticism is how class exploitation is understood in relation to ethnic oppression and the autonomy of these issues from one another. For example, the Marxist scholar Harold Wolpe alleges that much of the thinking on the topic of internal colonialism tends to treat ethnic relations as autonomous of class relations. This is problematic, since while workers belonging to the ethnically dominant group may benefit from surplus from other groups, they do not own/control the means of production or extra economic power critical to the successful exploitation and subjugation of other communities. Class distinctions are thus at risk of neglect in theories of internal colonialism as the tendency is to conceptualize an entire population ruling at the expense of another. The major division is hence not seen as between owners and nonowners of the means of producing everyday material life, but instead between groups whose differences are ultimately socially mutable. As another scholar, Cristóbal Kay, observes, "the key to the concept of internal colonialism deserves to be the 'relation' of domination and exploitation between two social groups." The actual racial, ethnic, cultural and even geographical elements are 'not' central to the mechanisms of internal colonialism – rather they 'merely' facilitate the exploitation of the internally colonized.

To the extent that internal colonial relation is an aspect of class relations, how then can the two be distinguished? Insofar as capitalism is the dominant mode of production in the country, and insofar as the forms of exploitation that are said to persist within internal colonies are common and apply to dominant social groups as well, then it is more appropriate to talk about class relations than colonial relations. When these forms of exploitation are accompanied by extra economic coercion and would not otherwise exist – and that it applies to particular groups only, or mainly – then it is valid to talk about colonial relations. The dimension of noneconomic political coercion is the key element in defining internal colonialism. Indeed, some scholars question whether internal colonialism has applicability to conditions where noncapitalist modes of production, if they still exist, are marginal. For, arguably, where they are marginal, dull economic compulsion will be sufficient to compel labor and products from noncapitalist areas to enter into fully capitalist production relations and circulation. Nevertheless, other writers continue to insist that the concept is analytically useful 'wherever' the relation of exploitation between capitalist areas and noncapitalist areas is expressed racially and/or in the garb of cultural/ethnic group identity, even in predominantly capitalist societies where there exists persistent and severe discrimination against aboriginal populations. Arguably, specific capitalists in certain places and at certain times do make use of extra economic coercion in the sphere of production (they do this, for example, when they deprive laborers of their right to freely bargain over their wages) and utilize mechanisms of racial–cultural–political oppression in order to obtain labor and commodities at below market wages and prices, respectively. Therefore, the general mechanism of internal colonialism – economic exploitation and racial oppression–political domination overlapping in a place – may not totally disappear with the coming of 'pure' capitalist relations.

A third criticism is the assumed homogeneity of internally colonized populations. On the one hand, there often – in the seeming rush to identify 'colonialism' – is an overlooking of 'emergent' class divisions within internal colony populations. On the other hand, there is often considerable local cultural and linguistic differences within oppressed populations and hierarchies of social power that operate within the wider population. These cultural distinctions may have divisive salience in domains of social life other than the workplace, and less traditionally theorized as being part of the class struggle – neighborhood politics, educational institutions, voluntary associations, and so on.

Finally, the concept of internal colonialism can be critiqued politically. According to some interpretations, to use the concept of internal colony automatically implies a strategy of 'national' liberation. To critics, this is the central weakness of the theory because it contains a hope for liberation from oppression and exploitation through an ethnic, cultural, and national autonomy. It, suggests Cristóbal Kay, can be used to raise political consciousness but cannot fulfill that stated goal. Structural conditions for establishing separate states do not exist in most situations, as most internal colonies are highly dependent on the workings of the integrated national capitalist system. There is, further, a pessimism to internal colonialism studies that, while often darkly illuminating, can at times hinder the recognition of substantial political gains (such as constitutional rights) as anything other than the attempt to co opt liberatory movements and aspirations within the colony. There is thus a deficient theorization of the capitalist state, as a political economic institution that emerges from, and has tensions reflective of, civil society at large.

Summary

The concept of internal colonialism emerged as, and remains, a historic attempt to understand the social and economic structures and inequities that persist in different spaces in the context of national capitalist development. In the effort to analyze often highly complex sociospatial formations, internal colonialism as a theory has both its strengths and weaknesses. Though a somewhat nebulous concept, due to its wide empirical application and popular theoretical appropriation, studies of 'internal colonialism' challenge many of the assumptions inherent in various international development theories, especially those emphasizing the inevitably revolutionary consequences of industrialization within territories, and those assuming 'pure' modes of production existing in isolation of capitalism. The key to the 'theory' of internal colonialism is to be found in an understanding of the particular sociospatial relations of domination and exploitation that characterize the relations between different ethnocultural social groups within a national space. The ethnic, social, cultural, and geographical elements in the conceptualization of internal colonialism are not central but do certainly facilitate the identification, establishment, and intensification of the phenomenon described as internal colonialism. Thus, the criticisms should focus on the theory's analysis of the relations of domination and exploitation. If these relations cannot be clearly defined, do not exist in reality, or are already part of a theory which has greater explanatory and predictive powers, then the concept of internal colonialism must necessarily be abandoned.

See also: Colonialism II; Dependency.

Further Reading

Drakakis Smith, D. and Williams, S. (eds.) (1983). *Internal Colonialism: Essays Around a Theme*. Edinburgh: Department of Geography, University of Edinburgh.

Gonzalez Casanova, P. (1965). Internal colonialism and national development. *Studies in Comparative International Development* 1(4), 27 37.

Hartwig, M. (1978). Capitalism and Aborigines: The theory of internal colonialism and its rivals. In Wheelwright, E. L. & Buckley, K. (eds.) *Essays in the Political Economy of Australian Capitalism*, vol. 3, pp 119 141. Brookvale, NSW: Australia and New Zealand Book Company.

Hechter, M. (1975). *Internal Colonialism: The Celtic Fringe in British National Development*. London: Routledge and Kegan Paul.

Hicks, J. (2004). On the Application of Theories of Internal Colonialism to Inuit Societies. http://www.cpsa acsp.ca/papers 2004/Hicks.pdf (Accessed in Aug. 2008).

Kay, C. (1989). *Latin American Theories of Development and Underdevelopment*. New York: Routledge.

Lovering, J. (1978). The theory of the 'internal colony' and the political economy of Wales. *Review of Radical Political Economics* 10(3), 55 67.

Robert, J. (1984). Facets of colonialism: The internal colonial concept. *Comparative Studies in Society and History* 26(3), 543 568.

Stavenhagen, R. (1965). Classes, colonialism and acculturation: Essay on a system of inter ethnic relations in Mesoamerica. *Studies in Comparative International Development* 1(6), 53 77.

Wolpe, H. (1975). The theory of internal colonialism: The South African case. In Oxaal, I., Barnett, T. & Booth, D. (eds.) *Beyond the Sociology of Development: Economy and Society in Latin America and Africa*, pp 229 252. London: Routledge and Kegan Paul.

Color, Mapping

A. L. Griffin, University of New South Wales-ADFA, Canberra, ACT, Australia

Glossary

Choropleth Map Thematic map where areas are shaded based on the value of a data variable.
Color Vision Impairment Inability to perceive differences in particular color hue pairs.
Electromagnetic Spectrum The entire range of wavelengths of electromagnetic radiation.
Simultaneous Contrast Apparent color shifts induced by surrounding colors.
Thematic Map A map displaying statistical data.
Visual Cortex The part of the cerebral cortex of the brain involved in processing visual information.

Introduction

The word color is often associated with sensations that are named, such as red, blue, purple, and brown. However, the phenomenon of color is actually multidimensional. Color can be differentiated along three different dimensions: hue, lightness, and saturation. Color hue is related to the wavelengths of light that are perceived by the map reader, and is typically described using names (e.g., 'the sky is blue' or 'the rose is red'). Color lightness is the perceived lightness or darkness of color, and is related to the quantity of light at a given wavelength. If an object reflects or emits a larger amount of light, it will be perceived as having a lighter color than an object that reflects or emits a smaller amount of light. Color saturation is a measure of color purity. Highly saturated colors are the purest, like lemon yellow or ruby red, while less saturated colors are muddy and grayish, like olive green. Highly saturated colors are created from narrow wavelengths of light, while less saturated colors contain light at a wider range of wavelengths (**Figure 1**).

Color Vision and Perception

Color vision, like all human vision, is the product of the human eye–brain system interpreting electromagnetic radiation. In other words, as a phenomenon, color does not exist independently of the image that the human brain creates.

The human eye–brain system is composed of three components:

- a source of visible light – the portion of the electromagnetic spectrum between the wavelengths of 370 and 770 nm (blue light and red light, respectively);
- the eye, whose lens focuses light that enters through the retina on a collection of specialized light sensitive cells, and whose optic nerve relays signals generated by these cells to the brain and;
- the brain, which constructs the sensation of color from the optic nerve signals after they reach the visual cortex at the rear of the brain.

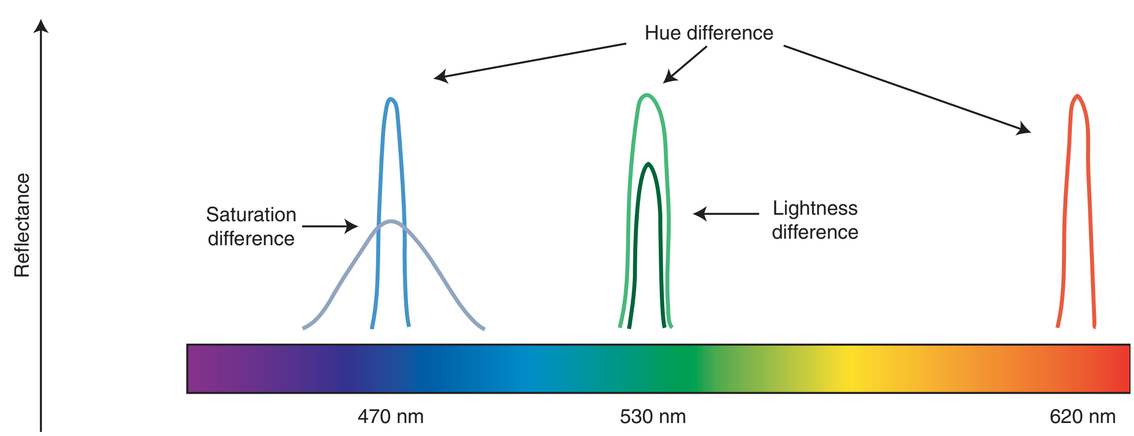

Figure 1 Differences in color hue (red, green, and blue peaks), color lightness (light and dark green peaks), and color saturation (pure blue and grayish blue peaks). ©1997 2008 The Pennsylvania State University. All Rights Reserved. Used by Permission.

Within the eye, there are two major types of light sensitive cells: rods and cones. Rod cells are sensitive to the intensity of light rather than its wavelength, and therefore do not play a large role in color vision. Cone cells, on the other hand, are sensitive to different areas of the visible portion of the electromagnetic spectrum. Human eyes have three types of cone cells, each sensitive to different wavelengths of light (i.e., color hues): L cones (long wavelengths – red light), M cones (medium wavelengths – green light), and S cones (short wavelengths – blue light). These different types of cone cells are differentially dis tributed in terms of both location and sensitivity across the retina. For example, while the S cones cover the most area, they are the least sensitive to light.

There are two main theories that explain how the brain constructs color from the signals generated by the eye's cone cells. The first, termed the trichromatic theory, holds that the color that map readers see is dependent upon the combination of the relative amounts of stimulation each of the three cone types receives. In other words, blue light will stimulate the S cones more than the M cones or the L cones. The second theory is the opponent process theory, which proposes that color perception depends on two opponent color channels (red–green and blue–yellow) and a lightness–darkness channel. The opponent color chan nels are based on the idea that there are no colors that are mixtures of red and green or that are mixtures of blue and yellow (as compared to the mixture of red and blue, which yields purple, or of red and yellow, which yields orange). Physiologically, the red–green channel is believed to result from comparing the signals from the L and M cones, while the yellow–blue channel results from a comparison of the S cones with the sum of the signal from the L and M cones.

Simultaneous Contrast

Simultaneous contrast is the phenomenon in which the eye perceives differences in color when two objects of the same color are surrounded by differently colored back grounds. This shift of color (i.e., the apparent differences) can be discussed in terms of the opponent process theory of color perception. In the black and white context, simultaneous contrast causes a shift toward the opposite end of the black–white color channel. In other words, if an object is surrounded by a white background, it will look darker than it will when surrounded by a dark background. In the color context, simultaneous contrast will induce color shifts toward the opponent color of that color channel. For example, a yellow background will make objects it surrounds look bluer (**Figure 2**).

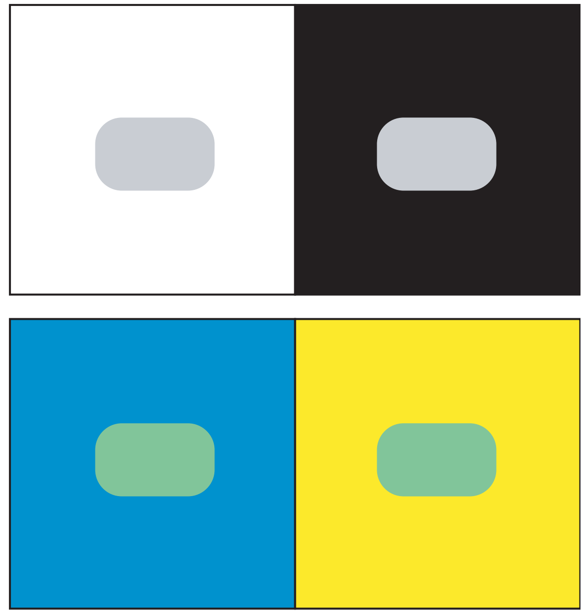

Figure 2 The color shifting effects of simultaneous contrast for both black and white and color surrounds.

Color Vision Impairment

The physiology of the human eye can also be used to explain individual differences in how map readers perceive color. Although its prevalence varies in different parts of the world, up to 4% of the population experi ences some form of color vision impairment. The con dition is more common in males than in females, with up to 8% of North American and European men experi encing color vision impairment.

The most common type of color vision impairment results in difficulty distinguishing between reddish and greenish colors, but the condition also affects hue pairs other than red and green, such as magenta and cyan. Scientists believe that this type of color vision impair ment may be related to changes in the L cones and M cones of impaired individuals. The effects of color vision impairment have been described using color confusion lines that identify color hues that may be confused by individuals with particular types of color vision impairment. Generally, hues that fall along one confusion line will be difficult for color vision impaired individuals to differentiate, while those that fall across the lines will be differentiable. Moreover, regardless of an individual's degree of color vision impairment, he or she will be able to see differences in lightness, so it is possible to design color combinations for maps that will accom modate map readers with color vision impairment (**Figure 3**).

Color Specification Systems

Color specification systems are scientific systems for identifying and precisely describing different

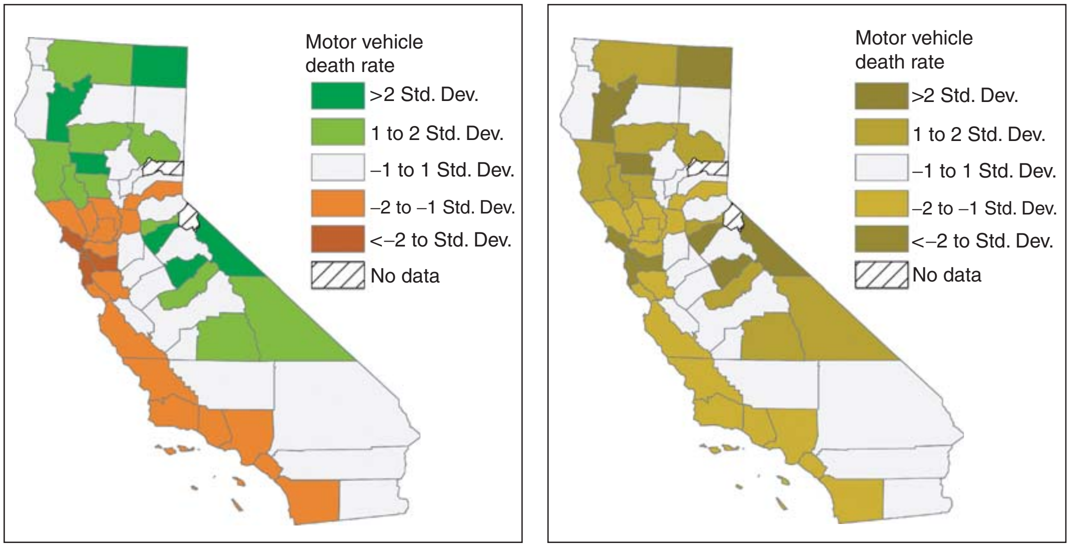

Figure 3 Example of a map that was not designed with color vision impaired individuals in mind (left) and a simulation created using VisCheck software of what a color vision impaired map reader would see (right).

colors. Color specification systems are needed because of the imprecision of the color vocabulary in everyday language. For example, the word 'blue' may cause some individuals to imagine the color of the sky on a sunny day, while others may imagine the color of blueberries or the color of blue jeans. They are also important because different methods of map pro duction require the use of different specification systems.

Color specification systems can be divided into two major classes: user oriented and production oriented systems. User oriented systems have the advantage that they explicitly consider human perceptual capabilities. However, user oriented systems eventually must be translated into a production oriented system to make the map. Production oriented systems, as the name implies, are based on the production technology used for con structing the map. The main disadvantage of using pro duction oriented systems is that they are specific to particular types of map production (e.g., print vs. elec tronic maps), and do not explicitly take color perception into account.

The primary user oriented color specification system is the Munsell system, which was derived from extensive user testing with color pigments. This system divides up the colors humans are capable of perceiving into equal perceptual divisions of color lightness and color saturation for each color hue. Working with this system, it quickly becomes apparent that humans do not perceive the same number of different shades of each color hue. A particular advantage of working with this system is that it is easy to create smooth transitions in color lightness or saturation when designing a color scheme (**Figure 4**).

Production oriented color specification systems can be further classed according to whether the production technology generates electronic or printed maps. The human eye senses two types of light: light that is emitted from a light source (such as a computer monitor or television) and reflected light (light that was generated by the Sun or another power source and has subsequently been reflected off the surface of an object). Additive color specification systems are used for describing emitted light, while subtractive color specification systems are used for describing reflected light.

The name 'additive' is derived from the fact that additional types of cones in the eye are stimulated when different wavelengths of light are mixed (e.g., adding red light to blue light to produce purple light will stimulate both L cones and S cones). The color specification system that is used for additive color is the RGB (red, green, blue) system commonly used for computer monitors. Computer monitors produce color by emitting light from three color guns (one each of red, green, and blue light). In this system, each color gun can emit light at 256 levels of intensity, producing a range of color combinations, with 255 being the highest intensity of light and 0 being the lowest (**Figure 5**).

Subtractive color specification systems are used when pigments (e.g., ink or paint) are used to produce maps. Subtractive colors are determined by the wavelength of light that is left after the pigments on the paper have absorbed other wavelengths. For example, we see the color blue when the ink on a page absorbs light in the red and green portions of the electromagnetic spectrum. The two main subtractive color specification systems are cyan, magenta, yellow, and black (CMYK) and spot

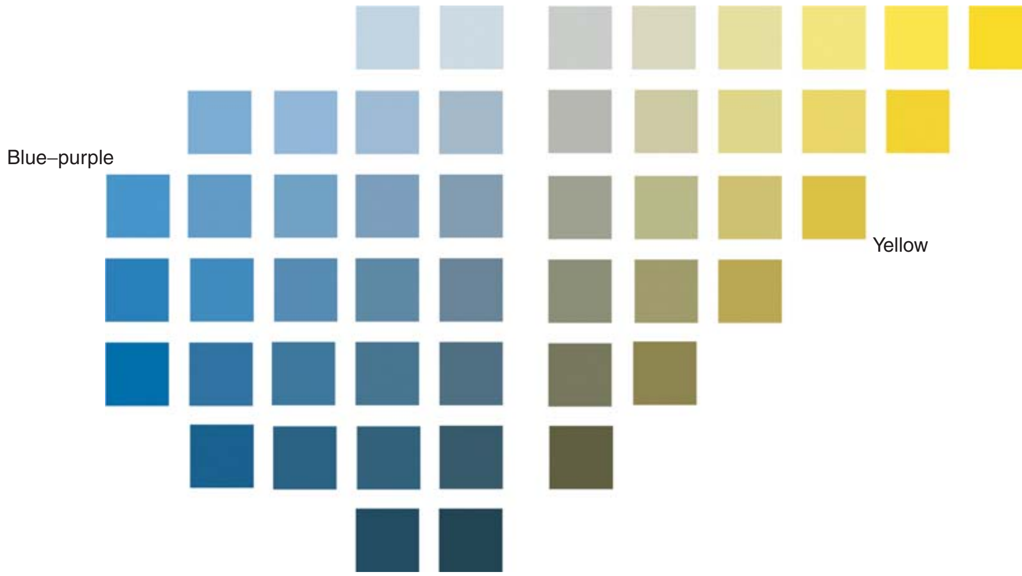

Figure 4 An example of a slice through the Munsell color space. Here, it is clear that humans perceive more shades of blue than yellow. This is a result of the fact that yellow is inherently lighter than blue, and therefore it is not possible to see as many dark yellows as dark blues. ©1997 2008 The Pennsylvania State University. All Rights Reserved. Used by Permission.

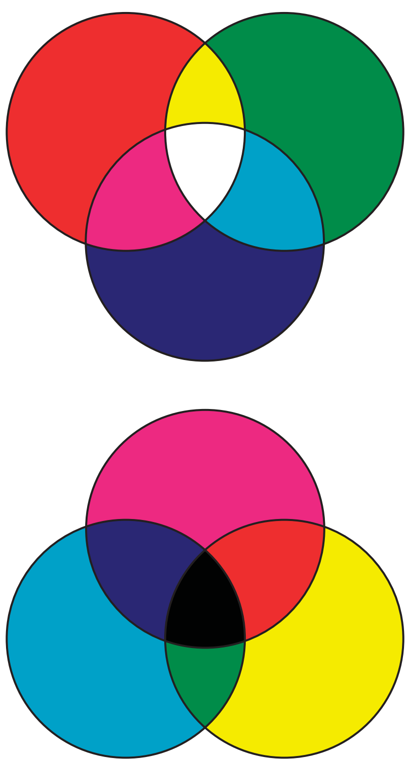

Figure 5 The additive (RGB; top) and subtractive (CMYK; bottom) color specification systems, used for electronic and print media, respectively.

color. In the CMYK system, which is commonly used in inkjet printers, laser printers, and printing presses, each of the four standard pigments is applied using a screen in amounts ranging from 0% to 100%, with 0% spe cifying that none of the pigment is placed on the page, and 100% indicating that the full strength of the pig ment is applied. Spot color is similar in that in this system the amount of pigment ranges from 0% to 100%, but different in that each spot color is a premixed pigment.

Choosing and Designing Color Schemes

There are three major types of color schemes that can be employed in thematic maps: sequential, diverging, and qualitative. Sequential color schemes use one or a small number of color hues, at a number of different lightness levels. Diverging color schemes are constructed by joining the lightest end of two different sequential schemes. The difference in color lightness between adjacent symbols in sequential and diverging color schemes should be similar so that map readers do not see a big jump between one class and the next. Quali tative color schemes are composed of symbols of similar lightness and saturation but different hues (**Figure 6**).

Cartographers choose the type of color scheme to be used based on characteristics of their data. The main distinction that they consider is whether their data are quantitative (showing differences in the amount of a phenomenon) or qualitative (showing differences in the kind of phenomenon). Sequential and diverging schemes are used for quantitative data. Most map readers associate light symbols with a smaller quantity and darker symbols

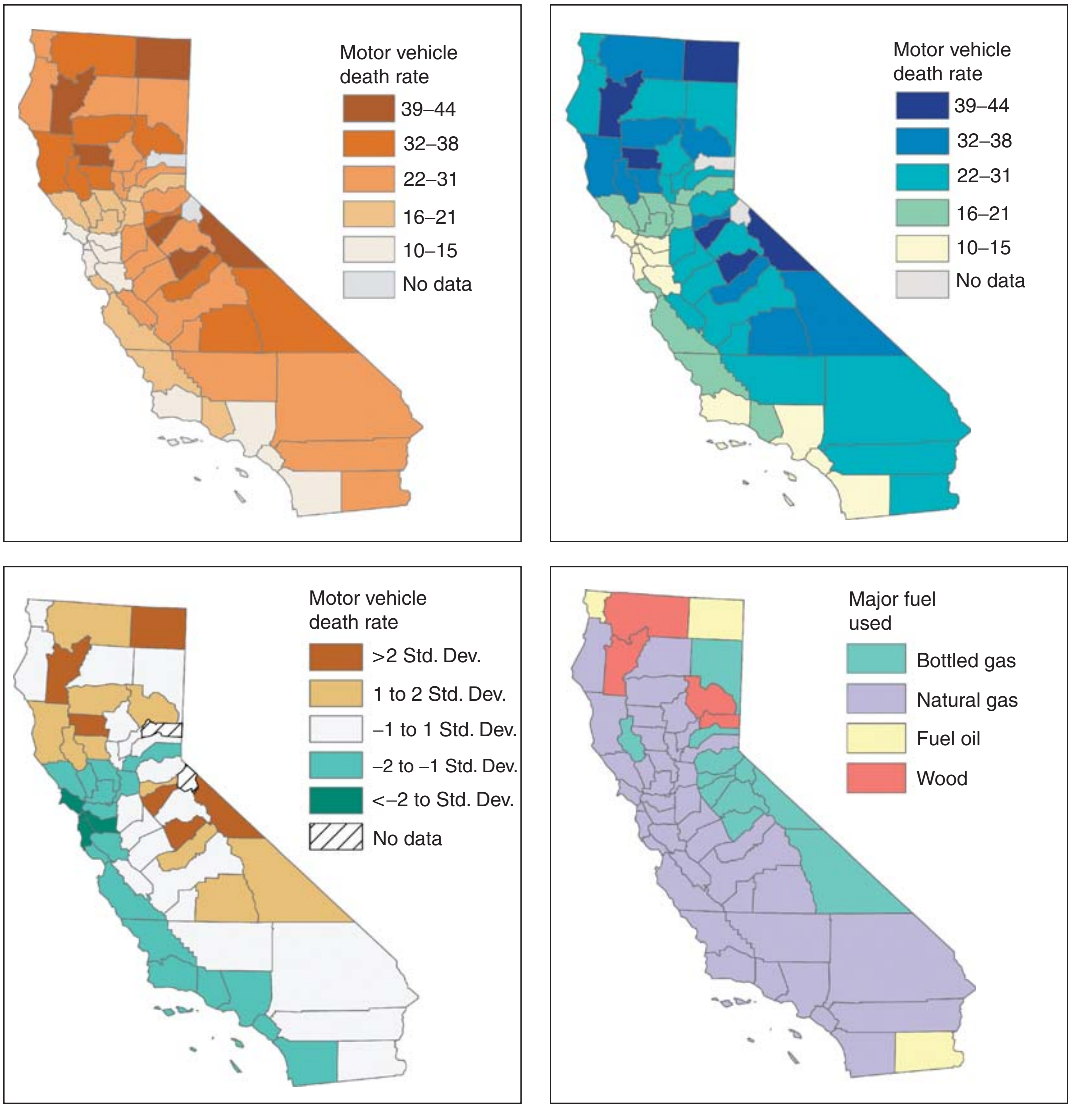

Figure 6 A one-hue sequential color scheme (top left) and two-hue sequential color scheme (top right) that use different lightness levels for representing quantitative information differences in the rate of motor vehicle deaths. A diverging color scheme (bottom left) differentiates areas that are above or below the mean motor vehicle death rate, which falls within the middle class. A qualitative color scheme (bottom right) uses symbols of similar color lightness and saturation, but different hues to show differences in the kind of fuel used to heat homes in California. ©1997 2008 The Pennsylvania State University. All Rights Reserved. Used by Permission.

with a larger quantity, which makes sequential and diverging schemes good choices for representing differences in amounts. In particular, cartographers choose diverging schemes when they want their map to allow map readers to make comparisons with critical values (e.g., zero population growth, mean precipitation, or the government defined poverty line). In this case, the critical value should fall in the central, light category. Qualitative schemes, on the other hand, contain hue differences rather than lightness differences. Because hues are not associated with quantities (e.g., red is not more than purple), they are effective for representing differences in kind.

Designing color schemes to accommodate individuals with color vision impairment requires additional thought. The effects of color vision impairment have been described using color confusion lines that identify color hues that may be confused by individuals with particular types of color vision impairment. Generally, hues that fall along one confusion line will be difficult for color vision impaired individuals to differentiate, while those that fall across the lines will be differentiable (**Figure 7**).

Change in the number of jobs by county

- Gain > 1000
- Gain 500–999
- Gain 1–499
- Loss 1–499
- Loss 500–999
- Loss > 1000
- No data

Confusing rendition

Color name	*x*	*y*	*Y*	R	G	B
Dark green	0.175	0.756	13	0	145	0
Green	0.268	0.658	30	115	185	0
Yellow-green	0.390	0.554	61	210	215	0
Yellow-orange	0.480	0.484	59	255	190	0
Orange	0.603	0.385	31	255	110	0
Red	0.646	0.353	13	190	60	0
Gray	0.328	0.411	19	150	150	150

Accommodating rendition

Color name	*x*	*y*	*Y*	R	G	B
Dark blue	0.153	0.064	9	50	50	250
Blue	0.205	0.164	29	150	150	250
Light blue	0.255	0.261	65	210	210	250
Light orange-red	0.413	0.360	47	255	180	180
Orange-red	0.519	0.380	28	230	130	130
Red	0.658	0.342	13	200	0	0
Dark gray	(Not measurable)			90	90	90

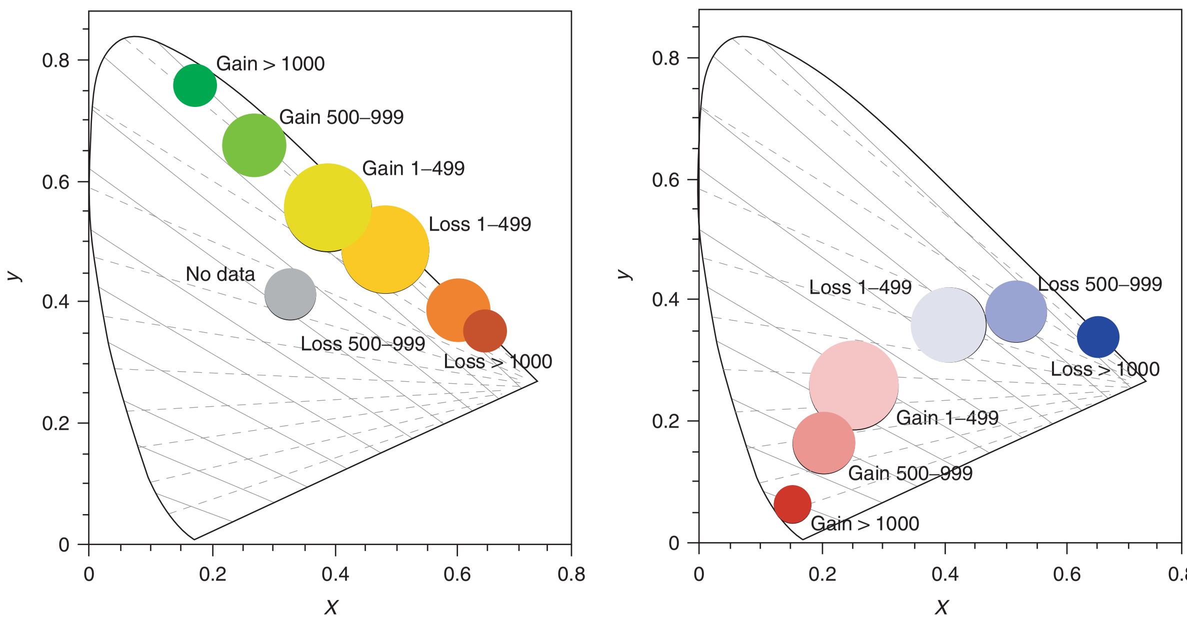

Figure 7 Examp e of a confus ng and an adjusted co or scheme p otted aga nst co or confus on nes for nd v dua s w th red–green co or v s on mpa rment. ©1997–2008 The Pennsy van a State Un vers ty. A R ghts Reserved. Used by Perm ss on. After O son, J. M. and Brewer, C. A. (1997). An eva uat on of co or se ect ons to accommodate map users w th co or-v s on mpa rments. *Annals of the Association of American Geographers* 87: 131. Courtesy of B ackwe Pub sh ng.

See also: Information Graphics; Map Perception and Cognition; Map Types.

Further Reading

Brewer, C. A. (1994). Color use guidelines for mapping and visualization. In MacEachren, A. M. & Taylor, D. R. F. (eds.) *Visualization in Modern Cartography*, pp 123 147. Tarrytown, NY: Elsevier Science.

Brewer, C. A. (1997). Evaluation of a model for predicting simultaneous contrast on color maps. *The Professional Geographer* 49, 280 294.

Brewer, C. A., Hatchard, G. W. and Harrower, M. A. (2003). ColorBrewer in print: A catalog of color schemes for maps. *Cartography and Geographic Information Science* 30, 5 32.

Brewer, C. A., MacEachren, A. M., Pickle, L. W. and Herrmann, D. (1997). Mapping mortality: Evaluating color schemes for choropleth maps. *Annals of the Association of American Geographers* 87, 411 438.

Brown, A. B. and Feringa, W. (2003). *Colour Basics for GIS Users*. Harlow: Prentice Hall.

Krygier, J. and Wood, D. (2005). *Making Maps. A Visual Guide to Map Design for GIS*. New York: Guilford Press.

MacEachren, A. M. (1995). *How Maps Work: Representation, Visualization and Design*. New York: Guilford Press.

Malacara, D. (2002). *Color Vision and Colorimetry: Theory and Applications*. Bellingham, WA: SPIE Press.

Olson, J. M. and Brewer, C. A. (1997). An evaluation of color selections to accommodate map users with color vision impairments. *Annals of the Association of American Geographers* 87, 103 134.

Slocum, T. A., McMaster, R. B., Kessler, F. C. and Howard, H. H. (2005). *Thematic Cartography and Geographic Visualization* (2nd edn.). Upper Saddle River, NJ: Prentice Hall.

Relevant Websites

http://www.colorbrewer.org
Colorbrewer: An interactive tool for selecting color schemes. This tool provides color specifications for CMYK, RGB, for a range of sequential, diverging, and qualitative color schemes.

http://www.vischeck.com
Vischeck: This website provides color vision impairment simulation capabilities. It allows the upload of image files in commonly used formats and produces a simulation of what map readers with three different types of color vision impairment would see.

Commodity Chains

M. Hassler, Philipps-University Marburg, Marburg, Germany

Glossary

Agro-Commodity Chain An agro-commodity-centered network, tracing the linkages and actor interrelationships from farm to household consumption.
Commodity Chain A network of labor and production processes whose end result is a finished commodity.
Governance Authority and power relationships between firms that determine how financial, material, and human resources are allocated and flow within a commodity chain.
Input–Output Structure A value-added chain of products, services, and resources linked together across relevant industries.
Node Specific process within a commodity chain and is embedded within larger networks of economic agents.
Production Network The nexus of interconnected functions and operations through which goods and services are produced and distributed.
Territoriality Geographic dispersion of production and marketing networks at the national, regional, and global levels, comprised of different sizes and types.

Commodity Chains: An Introduction

The production of any commodity, whether it be a manufactured product or a service, involves an intricate articulation of individual activities and transactions across space and time. Such production networks – the nexus of interconnected functions and operations through which goods and services are produced and distributed – have become not only organizationally more complex but also increasingly global in their geo graphical extent. They not only integrate firms (and parts of firms) into structures that blur traditional organiza tional boundaries, through the development of diverse forms of equity and nonequity relationships, but also integrate national economies (or parts of such econ omies) in ways that have enormous implications for na tional economic development. At the same time, the precise nature and articulation of such firm centered networks are deeply influenced by the concrete socio political, institutional, and cultural contexts within which they are embedded, produced, and reproduced. The process is especially complex because while the latter are essentially territorially specific (primarily, though not exclusively at the level of the nation state), the production networks themselves are not. They 'slice through' state boundaries in highly differentiated ways, influenced in part by regulatory and nonregulatory bar riers and local sociocultural conditions, to create struc tures that are 'discontinuously territorial'. At the same time, the geographical scale at which production net works are being configured continues to widen.

Central to any network concept is, therefore, the cross national segmentation and separation of all eco nomic actors and processes surrounding the production and consumption of a specific product. Within this con text, several approaches and traditions of theorizing and conceptualizing 'chains' in the production process are found in differing terminologies and disciplinary con texts. In particular, the debate on (global) commodity chains or, more recently, on (global) value chains has produced a comprehensive contribution to this con ceptual development. Much of the debate on commodity chains, the focus of this article, is derived initially from Wallersteinian world system theory. Both ap proaches apply a global perspective to conceptualize the integration of nations at different stages of economic development into the global economy. However, while world system theory is predominantly state centric, the commodity chain debate focuses on production as the main unit of analysis. A commodity chain has been broadly defined as "a network of labour and production processes whose end result is a finished commodity" (Hopkins and Wallerstein, 1986: 156). This article out lines the initial key arguments on which the subsequent commodity chain debate was based and on which a large body of literature has evolved. Furthermore, it provides a brief summary on how this concept has been applied to agro commodities.

Commodity Chains: A Conceptualization

The starting point of commodity chain analysis is to focus on the dynamics of contemporary changes within individual industrial sectors and their corporate strat egies. It aims to link structural changes at the global, macroeconomic level with contemporary firm specific, microeconomic organizations of production and distri bution and goes beyond the sole examination of the spatial extension of the commodity flow. The commodity chain analysis intends to answer questions of how the various economic agents of individual countries are linked to the global economy. This includes a perspective on the forces and systems of localization in the world economy in raising questions where the commodity chain geographically 'touches down' in the local economy and how it is connected to firms and households within. This also includes aspects of the realization and extraction of the economic surplus and the labor relations and com position. Commodity chains are:

> …sets of interorganizational networks clustered around one commodity or product linking households, enterprises and states to one another within the world economy. These networks are situationally specific, socially con structed, and locally integrated, underscoring the social embeddedness of economic organization. (Gereffi *et al.*, 1994: 2)

Initially, it was claimed that the novelty about com modity chains is not the theorizing of the fact that eco nomic activities cross national boundaries, but the inclusion of the role of strategic decisions within the internationalization process of production and trade. Economic actors tend to develop variations in the or ganization and coordination of economic activities based on the ownership of property rights, and the way these are obtained and exchanged. These variations in economic activities affect the governance of interrelationships and could be dominated by highly cooperative transactions or by a rather adversarial competition. Indeed, the com modity chain perspective stresses the variations and or ganizational scope of linkages between economic agents to unpack and understand their sources of stability and changes. This includes the extent to which economic actors are influenced by the institutional framework of the locational setting of their transactions, since production and trade have to coincide with the national and inter national regulations they are embedded within.

In traditional microeconomic terms, production is defined as a combination of inputs to produce an output as a finished product. However, chain concepts imple ment a more dynamic approach whereby production takes place in time and space and consists of subsequent and interconnected processes. Production is a series of functions, such as design and product development, raw material procurement, assembly and manufacture, and marketing and distribution. It is a chain of activities and exchanges in which inputs move forward as money moves back. Furthermore, the traditional definition of discrete industries in economics is becoming less useful in ana lyzing the contemporary interconnected relationships of firms. The issues of defining specifically what is an in dustry and which production processes belong to which industry leave many open questions and loose ends in traditional approaches. Indeed, contemporary inter industry input–output relations, in particular, demand a concept that is centered around production taking place within and between firms.

Instead of looking at a specific industry, a chain con cept offers the advantage of investigating the production processes, rather than focusing on manufacturing alone. This is based on the general definition of the capitalist firm, in which production is a result of adding value to materials/ inputs to produce a specific product/ output. Within this context, four distinct analytical dimensions to any commodity chain have been identified: (1) input–output structure, (2) territoriality, (3) governance struc ture, and (4) institutional framework. The input–output structure of the commodity chain approach emphasizes the embeddedness of economic agents such as raw ma terial suppliers, factories, traders, and retailers within wider sets of inter and intrasectoral networks. In other words, the economic activities of firms, by their very nature, add value to a specific product and are therefore generally embedded within larger networks of market participants defining their backward and forward link ages. They are involved in production processes that generate services and/or produce physical commodities, which locate them within specific commodity chains of extensive social and organizational relations.

The commodity chain literature refers to these pro duction processes as 'nodes', which identify specific processes of the chain and are embedded within larger networks of economic agents. Each node has a set of structural components, which include acquisition and/or organization of inputs, employment policies, transport, distribution, and finally, consumption. In other words, each node transforms supplied raw materials with a specific labor force and technical endowment into a product that is logistically submitted to a successive node of consumption. However, consumption does not neces sarily represent a consumer market and could also be a node of further processing. In consequence, the com modity chain concept allows an investigation of the backward and forward linkages of individual firms from the very early stages of raw material production to final consumption of the finished commodity. The nodal concept of commodity chains allows a division of the entire production process of a single product or com modity into its individual sequences, whereby each in dividual node's social and organizational construction can be analyzed. Every single node is clearly defined and can be consequently redefined. Technological innovations and/or changes in the organizational structures of busi ness corporations affect the (re)definition of these nodes.

A key argument of the commodity chain framework is that the nodes are the locus of profit and surplus rather than the national economies. In this context, the dualism between core and peripheral nodes is a development of the argument of core and peripheral states within world system theory. The analytical distinction between core and periphery in relation to the nodes evolved on the basis that commodity chains consist of a set and hierarchy of processes of various complexity and sophistication. The hierarchical affiliation of the node toward core or periphery is largely defined by the value added within, whereby core activities are those where the principal profit accrues.

It is argued that the higher the value added within a node, the higher is the competitiveness of the firm. Furthermore, firms operating in high value niches/core nodes aim to protect their competitiveness by creating barriers of entry for competitors. These barriers could be based on innovative process technologies, superior product qualities, or brand reputations as a result of cumulative marketing efforts. Competitive advantage is often created within a highly localized cluster of economic actors to guarantee spatial proximity to research centers, superior component suppliers, and competing firms. In contrast, low value processes/peripheral nodes are characterized by a large number of competing firms. Within peripheral nodes the involved processes show a lower complexity and are largely based on cost advantages generated by cheap labor resources, raw materials, and economies of scale. The geography of these nodes tends to be highly dispersed since the foundations of the competitive advantages can be easily replicated by other firms around the globe. Furthermore, semiperipheral nodes have been identified as an intermediary stage. They demand some form of organizational and industrial capabilities and integrate processes of relatively advanced manufacturing and low end services such as quality control and component sourcing. Semiperipheral nodes often represent the organizational linkages between core and peripheral processes.

Thus, the extent of geographical dispersion of these nodes, which comprise the pivotal point of production, defines the territoriality of the commodity chain. This includes the aspects of the geographical location of the nodes as well as the geographical structure of the firm's internal and external linkages. In consequence, the territoriality of chains has a spectrum of possibilities that range from geographical dispersion to high geographical concentration, from the global scale to the local scale. The geographical configuration of the chain is influenced to a large extent by the technologies of transport and communication. Thus, the innovation processes within these key technologies to integrate the spatial dimension of economic activities result in a continuous geographical extension of chains. These processes shape and reshape international economic activities into more complex and tightly connected production networks. However, in addition to the continuous spatial fragmentation and dispersion of global production networks, there has been an increasing geographical concentration and integration of economic activities at various regional scales. These regional network coordinations are largely a result of heterogeneous regional clusters of national economies. They operate on the basis of the same principles as the networks stretching over the global economy following the hierarchy of nodes.

Within this context the territoriality of production can also be conceptualized on the basis of the economic development process. The trajectory of world market integration sees economic development largely as a result of close interrelationships with transnational corporations (TNCs) from developed countries in which technology and knowledge transfers occur. At the starting point of world market integration, people and firms learn how to manufacture products to the specifications of local and foreign buyers in terms of quality, price, and delivery specifications. However, when initial exports are threatened by new competitors operating with lower costs or by protected and saturated export markets, local producers adjust their strategic behavior to the new conditions to maintain a position within these industries. This trajectory is structured into the stages of (1) primary commodity exports, (2) export processing assembly operations, (3) component supply subcontracting, (4) original equipment manufacturing (OEM), and (5) original brand name manufacturing (OBM), whereby each successive stage demands a higher degree of industrial capabilities. However, instead of defining the territoriality of production as a sum of nation states, commodity chain analysis claims to identify the fundamental structures of global capitalism, in which national boundaries are insignificant. This follows the argument that the world economy is organized through horizontal and vertical linkages of an international division of labor in which the modes of integration and geographical scopes vary over time.

To identify the form of integration in the global economy and the power relations between each economic actor, the commodity chain concept has placed a strong analytical emphasis on the governance of economic action. The governance structure of each commodity chain is influenced to a large extent through the various forms of intra- and interorganizational relationships. Distinct combinations of business relationships are created by firms of all kinds of organizational scales and scopes operating within an economy – transnational and domestic, large and small, and public and private. Despite this heterogeneous structure of the economic system, it is increasingly the TNC that shapes and coordinates the structure of the commodity chain. International capital flows are a basis of this global shift of economic activities, however, they do not lead necessarily to direct investments in production facilities. Looser forms of business relations, such as subcontracting and licensing, do not demand the relatively strong commitment of foreign direct investment while generating a certain power and control of production in external facilities. In consequence, the ownership of productive assets is not essential to empower the TNC with the control mechanisms over these assets. Furthermore, large retailers and brand name companies have the distinctive power to decide and control the products they buy, but without any direct involvement in the manufacturing processes.

Therefore, Gereffi argued that industrial and commercial capital has a central role in creating commodity chains of distinctive governance structures that are either 'producer driven' or 'buyer driven'. He bases his dual distinction on the debate emerging from Piore and Sabel's work on industrial organization. Producer driven models are concerned with mass production, while buyer driven models are production units concerned with flexible specialization, based on the growing importance of segmented demands in developed country markets. This is also based on the assumption that patterns of production shape the character of demand in producer driven commodity chains, whereas in buyer driven commodity chains it is consumption that determines where and how global manufacturing takes place.

According to Gereffi, producer driven commodity chains are vertical networks of large TNCs predominantly operating in capital and technology intensive industry sectors, such as heavy machinery, automobiles, aircrafts, and computers. TNCs in producer driven commodity chains are characterized by the ability to control their backward linkages to raw material and component suppliers, and their forward linkages into distribution. Producer driven commodity chains are controlled by firms at the point of production. However, the subcontracting of components and strategic alliances among rivals are common features of this type of commodity chains. They usually belong to global oligopolies, whereby the geographical scope of these networks varies according to the number of countries included in the commodity chain, as well as to their development stage.

In contrast, buyer driven commodity chains are trade based horizontal networks coordinated by retailers, trading houses, and brand name companies. Firms coordinating buyer driven chains deal predominantly with labour intensive, manufactured consumer goods, such as clothing, footwear, toys, housewear, and consumer electronics. The geographical separation of the production process from design, research and development (R&D), and marketing is the key characteristic of this type of commodity chain. Gereffi argues that core companies or buyers embedded within this type of commodity chains are 'merchandisers' rather than 'manufacturers'. These firms specialize in design and marketing activities without actually owning and operating internalized production facilities. This allows buyers to focus on final consumer markets to respond with strategic decisions on changes of consumption. They own brands and maintain these by designing appropriate products and creating extensive marketing campaigns in their target markets.

The highest barriers to entry in buyer driven commodity chains exist in marketing, design, and product innovation. Output of these nodes is directed at gaining competitive advantage in the marketplace. Companies located in core niches of this type of commodity chain are able to exert oligopolistic power because of their knowledge of markets, immense bargaining power, and retail outlets. In consequence, the accumulation of knowledge and power in high value research, design, sales, marketing, and financial services define the profitability of core companies. The exclusion of production as a core business obviates the need for economies of scale to achieve profitability, and instead allows a focus on economies of scope as a business strategy. This enables quick responses to market changes and the service of different market spaces to their own requirements with a unique branded product. Therefore, core profits are realized in design, retailing, and marketing, based on property rights such as brand names and trademarks, rather than manufacturing.

The distinction between producer driven and buyer driven commodity chains highlights the inter and intrasectoral variations of global capitalism. The firm has been described as a major force in developing these variations in the structural and organizational composition of the chain. Although the institutional framework is identified as the fourth dimension in this concept because "state policy plays a major role in GCCs" (Gereffi, 1994: 100), the driving forces within the commodity chain itself are identified as the main causes for global change and firm specific behavior. This follows largely from the argument of convergence in which the globalization process will result in rather homogeneous structures and strategies of TNCs. Gereffi argues that the institutional framework has an increasingly diminishing impact on the organizational scopes of firms.

This conceptualization of the commodity chain framework has also attracted some criticisms. One area of criticism, for example, was the fact that it oversimplifies the economic realities by a sole dual distinction between buyer and producer driven chain (this dualism has now been replaced by a more nuanced governance conceptualization within the value chain debate). Furthermore, it has been argued that the institutional framework plays a far more relevant role in defining globalization and economic processes than the initial commodity chain literature would stress.

Commodity Chains: An Application

Despite these points of criticism, the commodity chain concept has been widely accepted and used for empirical analysis of structures and dynamics within the global economy. For economic geographers, one of the main concerns within commodity chain analysis has been the examination of "the spaces through which consumers are connected to producers" (Hughes and Reimer, 2004: 1). In terms of sectors in which commodity chain research has been conducted, a rather strong focus on buyer driven sectors, such as clothing, has been identified. However, the concept has analytical potential for a wide range of economic/industrial sectors and the way con sumers and producers are interlinked (see **Figure 1**).

One area in which economic geographers have also widely applied the commodity chain concept is the agro food production system, from farm to final consumer. Indeed, it has a strong potential to investigate the interrelationship of consumption with production and vice versa.

As such, agro food systems have potentials for highly heterogeneous commodity chain coordinations. These coordinations are dependent not only on the type of the investigated produce but also on the way it is supplied to final consumers. The type of produce has an influential factor on the way production is organized, while the degree of processing has an influential factor on the actors included within the chain. It could be argued that the more processed the food is, the more industrialized becomes an agro food system and the more actors are included within a production system.

Therefore, the analysis of agro food commodity chains provides case studies of coordinations of highly localized spatial dimensions (e.g., farmers market) up to the very global scale (e.g., Ghana cocoa or Ethiopia coffee for consumption in Europe). In addition, they provide cases of very narrow producer–consumer organizational coordinations (e.g., again raw vegetables at a local farmers market traded directly to consumers) up to highly com plex organizational network coordinations (e.g., from cocoa farmer to a final Mars bar or the conventional meat production). In particular, coordinations that stretch across the globe and integrate developing countries into the global economy have further analytical dimensions. They provide case studies of trade dependencies of de veloping countries from developed countries.

As Gibbon (2001:60–68) states, "agro commodities re main a main strand linking the world's rural poor with global product markets." Cash crops, such as coffee, tobacco, cotton, sugar, rubber, tea, and cocoa, still repre sent a main source of income and foreign exchange for developing countries. In the late 1990s, these main tropical agro commodities still generated more than 60% of total agricultural exports of all LDCs. These agro commodities are still largely supplied to global markets as primary goods and not further processed. The secondary pro cessing of primary agro commodities, such as the roasting of coffee or the grinding of cocoa, is still usually based in developed countries. There are two main reasons why these processes are usually not conducted at locations of plantation. This is in particular (1) the lack of the neces sary knowledge of consumer tastes at target markets and (2) to obtain a rich flavor, processing has often to be conducted within short time before end consumption. Therefore, upgrading into secondary processing of agro commodities, such as coffee and cocoa, within developing countries is basically done only for consumption on the domestic market.

As the cases of coffee and cocoa demonstrate, the way developing country producers are integrated into the global economy depends to a large degree not only on the produce they grow, but also on the way it is con sumed. This is also the case for other agro commodities. In the case of horticulture commodity chains, con sumption also has a crucial role in the way farmers are integrated into the wider economic system. Although local or regional production systems have received

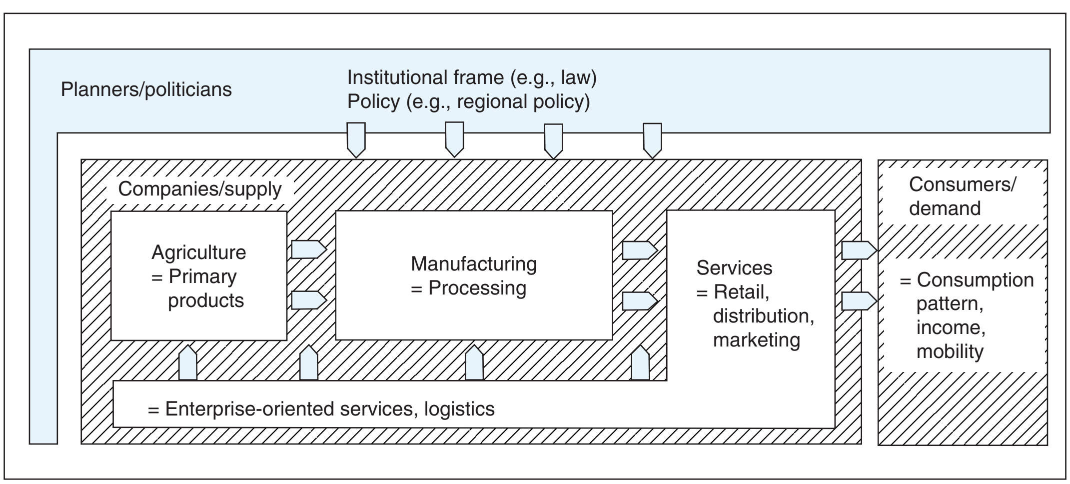

Figure 1 Sectors and actors in the commodity chain system. Reproduced from Kulke, E. (2007). The commodity chain approach in economic geography. *Die Erde* 138, 123.

increasing importance, in particular within the growth of demand of organic food, however, the bulk of con ventional food is traveling larger distances. This is, in particular, the case during the winter time of the large consumer markets, which are primarily located on the Northern Hemisphere. Farmers and growers on the Southern Hemisphere supply supermarkets in Europe and North America with fresh produce. As a result of the market consolidation within retailing, these supermarkets are also endowed with a strong market power in relation to their agro commodity suppliers.

> The horticultural … [commodity] chain linking UK consumers and supermarkets with export firms and farmers in Africa has been directly affected by this process of retail concentration. Whereas imported horticultural produce was previously channelled pri marily through wholesale markets, the largest UK re tailers now control 70 90% of fresh produce imports from Africa. There are signs of similar trends in other parts of Europe. (Dolan *et al*., 1999: 7)

In consequence, this results in some form of buyer driven chain, as in the case of manufactured industrial products, such as clothing. **Figure 2** illustrates the spe cific actors and nodes within a food commodity chain linking UK supermarkets with African producers. From

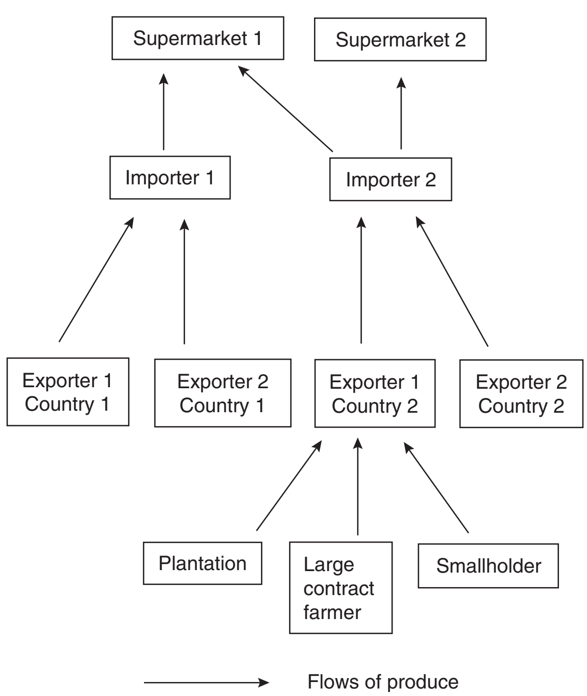

Figure 2 Flows of produce in African food (commodity) chains. Reproduced from Dolan, C., Humphrey, J. and Harris-Pascal, C. (1999). *Horiculture Commodity Chains: The Impact of the UK Market on the African Fresh Vegetable Industry*. IDS Working Paper 96, Brighton: Institute of Development Studies.

harvest to supermarket shelves, the produce is channeled through four stages. After the harvest on the fields of a plantation or the land of a smallholder, the produce is processed by an African exporter. This exporter is re sponsible for packaging, storage, and transport to the UK. Once the food is in the UK, it is taken over by an im porter, which has usually exclusive, country specific contracts with exporters. Again, the importer is respon sible for the quality control, storage, and transport of the goods to the warehouses of the supermarket chains. However, the organizational standards within these trade networks as well as the product prices are largely defined through the market power of the retailers.

The awareness of final consumers of these asym metrical power relationships and dependencies of de veloping country farmers from buyers in the developed world also resulted in initiatives and pressure groups for ethical trading. These campaigns aim to influence through consumer choice the code of practice of trans national manufacturing corporations and retailer organ izations. Environmental protection and improved social standards are the main goals of these initiatives. They aim to improve the circumstances of production under which, for example, agro commodities are grown and produced and the livelihood for the people involved. Ethical trade campaigns have been started for a number of commod ities, such as clothing and horticulture produce.

In fact, since research on environmental issues has a strong tradition within geography as a discipline, future commodity chain research could place an even stronger focus on environmental sustainability, in com bination with the social sustainability of agro production systems.

See also: Embeddedness; Global Commodity Chains; Global Production Networks; Globalization and Transnational Corporations.

Further Reading

Dicken, P. and Hassler, M. (2000). Organizing the Indonesian clothing industry in the global economy: The role of business networks. *Environment and Planning A* 32, 263 280.

Dicken, P., Kelly, P. F., Olds, K. and Yeung, H. W. C.h. (2001). Chains and networks, territories and scales: Towards a relational framework for analysing the global economy. *Global Networks* 1, 89 112.

Gereffi, G. (1994). The organization of buyer driven global commodity chains: How US retailers shape overseas production networks. In Gereffi, G. & Korzeniewicz, M. (eds.) *Commodity Chains and Global Capitalism*, pp 95 122. Praeger, CT: Westport.

Gereffi, G. (1999). International trade and industrial upgrading in the apparel commodity chain. *Journal of International Economics* 48, 37 70.

Gereffi, G. and Korzeniewicz, M. (eds.) (1994). *Commodity Chains and Global Capitalism*. Praeger, CT: Westport.

Gereffi, G., Humphrey, J. and Sturgeon, T. (2005). The governance of global value chains. *Review of International Political Economy* 12, 78 104.

Gereffi, G., Korzeniewicz, M. and Korzeniewicz, R. P. (1994). Introduction: Global commodity chains. In Gereffi, G. & Korzeniewicz, M. (eds.) *Commodity Chains and Global Capitalism*, pp 1 14. Praeger, CT: Westport.

Gibbon, P. (2001). Agro commodity chains: An introduction. *IDS Bulletin* 32, 60 68.

Hassler, M. (2003). Crisis, coincidences and strategic market behaviour: The internationalization of Indonesian clothing brand owners. *Area* 35, 241 250.

Hassler, M. (2003). The global clothing production system: Commodity chains and business networks. *Global Networks* 3, 513 531.

Hassler, M. (2004). Raw material procurement, industrial upgrading and labor recruitment: Intermediaries in Indonesia's clothing industry. *Geoforum* 35, 441 451.

Henderson, J., Dicken, P., Hess, M., Coe, N. and Yeung, H. W. C.h. (2002). Global production networks and the analysis of economic development. *Review of International Political Economy* 9, 436 464.

Hess, M. and Yeung, H. W. C.h. (2006). Whither global production networks in economic geography? Past, present, and future. *Environment and Planning A* 38, 1193 1204.

Hopkins, T. K. and Wallerstein, I. (1986). Commodity chains in the world economy prior to 1800. *Review* 10, 157 170.

Hopkins, T. K. and Wallerstein, I. (1994). Commodity chain: Construct and research. In Gereffi, G. & Korzeniewicz, M. (eds.) *Commodity Chains and Global Capitalism*, pp 17 20. Praeger, CT: Westport.

Hughes, A. (2001). Global commodity networks, ethical trade and governmentality: Organizing business responsibility in the Kenyan cut flower industry. *Transaction of the Institute of British Geographers* 26, 390 406.

Hughes, A. and Reimer, S. (2004). Introduction. In Hughes, A. & Reimer, S. (eds.) *Geographies of Commodity Chains*, pp 1 17. London: Routledge.

Kulke, E. (2007). The commodity chain approach in economic geography. *Die Erde* 138, 117 126.

Leslie, D. and Reimer, S. (1999). Spatializing commodity chains. *Progress in Human Geography* 23, 401 420.

Raikes, P., Jensen, M. F. and Ponte, S. (2000). Global commodity chain analysis and the French filiere approach: Comparison and critique. *Economy and Society* 29, 390 417.

Relevant Websites

http://www.globalvaluechains.org
Global Value Chains Initiative, Duke University.

Communicable Diseases, Globalization of

J. R. Oppong, University of North Texas, Denton, TX, USA

Introduction

Communicable diseases are the greatest disease threat facing humankind. Diseases previously considered eradicated are re emerging and new viruses are emer ging. Due to increased human population and accelerated global travel, local outbreaks become instant global threats. Despite technological advances in medicine, devastating communicable diseases such as Ebola, severe acute respiratory syndrome (SARS) or avian influenza continue to surface and ravage humankind. The 1918 influenza outbreak killed at least 20 million people in about 24 weeks, but HIV/AIDS has killed more than 25 million since 1981 and continues its march around the globe. Researchers are concerned that an avian influenza outbreak could kill many more people when it emerges because of the absence of immunity and human travel interaction patterns.

Diseases are not distributed uniformly in space. Throughout the world, certain locations have diseases that are not present in others. Even when a disease occurs in different locations, the rate of occurrence frequently varies between places. Some areas have high rates of a particular disease than others or may not have that dis ease at all. Generally, tropical regions with warm, humid climates tend to facilitate the proliferation of disease causing organisms, or insects and animals that carry such disease. Similarly, physical environmental characteristics, such as soil type and the mineral content of soils and rock, which vary from one location to another, may also inhibit or encourage disease. Swampy vegetation in low lying areas of the tropics is ideal for the proliferation of mosquitoes and thus, endemic malaria. Finally, cultural practices, including diet, behavior, and occupation, also contribute to geographic variations in health and disease.

Communicable diseases, diseases that are easily transmitted from one person to another, also vary spa tially. Geographic differences such as levels of crowding, sanitation, and health practices such as defecation, or even sexual behavior contribute to these differences. Medical geographers study the spatial differences in disease occurrence, how diseases spread across geo graphic space, and also the spatial distribution of health care resources. They seek to answer the question of who is getting what diseases/health care where and why? This article examines the geographic distribution of leading communicable or infectious causes of disease and death. We begin with one basic conceptual framework commonly used to account for spatial variations in disease.

Disease Ecology Model

The disease ecology model attributes spatial variations in disease to three major variables – environment, genetics and behavior. Environment is a broad context and in cludes not just the physical environment of livelihood, environmental pollution e.g. (including indoor pollution such as radon) and water quality, but also the socio economic environment, including relative deprivation, healthcare quality and access, land mines, and the po litical economic context. For example, unlike Canada, where universal access to care based on need is provided for all citizens, in the United States, access to care de pends on ability to pay. Thus, immunization is more easily available in Canada than in the US and this influ ences the infant mortality rate and life expectancy in the two countries. Genetic predisposition is important par ticularly for degenerative diseases. People with light skin pigmentation are more prone to skin cancer than others with a darker pigmentation. Similarly, recent research suggests that Ashkenazi Jewish women have a higher risk of breast cancer compared to other women. The human genome project is not only identifying genes responsible for several serious diseases such as Parkinson but also developing exciting new and effective therapies for them. Human behavior is the final component in the disease ecology model. Put simply, behavior facilitates or inhibits disease. Unhealthy behavior traits such as excessive al cohol consumption, overeating and failure to exercise, unsafe sexual practices and smoking are associated with various diseases. In contrast, regular exercise and other positive health behaviors, including proper diet, enhance health and prevent disease. Thus, spatial variations in Africa's environment, including widespread poverty that promotes extra marital and transactional sex, poor healthcare environment with limited access to anti retrovirals, cultural practices such as circumcision, widow inheritance, and preference for unsafe sexual practices, combine with other factors to make HIV/AIDS rates the highest in the world. Similarly, variations exist between and within countries. Besides vertical transmission (from mother to child), genetics is not as important in HIV transmission – all humans are genetically susceptible.

The disease ecology framework has been criticized for its positivist philosophy, dearth of social theory, neglect of social divisions such as race, gender, disability, and sexuality, and the trivialization of the role of place and locality. For example, Kearns called on medical geog raphers to incorporate social theory, but Mayer and

Meade and Paul chastised the supposed call for dis carding disease ecology in favor of narrow health geo graphy. Current work includes much broader frameworks such as vulnerability theory and political ecology.

Globalization and Communicable Diseases

Globalization, particularly the increased economic and cultural interdependence, and associated shrinking of time and space due to technological advancements, is transforming disease transmission patterns. Traditionally, immigrants have always provided a vital link in disease transmission between their home regions and their des tination. By establishing close contact between distant places, immigrant travels promote disease spread. For example, European travelers carried diseases that killed many people in the Americas. However, recent devel opments in technology and travel have escalated the role of travel related disease spread. In fact, the volume, speed, and reach of travel today have so accelerated the spread of communicable diseases that national bound aries no longer offer isolation or protection. The brief review that follows demonstrates that permanent mi grants are not the only means of disease spread. Short term visitors, including tourists, are also very important in the spatial diffusion of microbes and pathogens.

HIV/AIDS

Easily the world's worst communicable disease, HIV/AIDS has affected every country but sub Saharan Africa has experienced the worst devastation. The Caribbean and parts of Eastern Europe and Asia also have high rates. Even in the United States where rates are much lower, increasing numbers of people are living with HIV/AIDS, and prevalence is particularly high among subpopula tions such as Blacks, men who have sex with men, and intravenous drug users.

At the end of 2006, an estimated two thirds of the world's 40 million cases were in sub Saharan Africa, a region with 11% of the world's population, and 91% of the world's children living with the disease. HIV preva lence rates vary widely between African countries, ran ging from a low of less than 1% of the adult population in Senegal and Somalia to 15–20% in South Africa and Zambia. A few countries have even higher prevalence rates – Swaziland with 33.4% is the highest, Botswana has 24.1%, while Lesotho and Zimbabwe have 23.2% and 20.1%, respectively, at the end of 2006. Generally, West Africa has much lower rates than Southern Africa. Besides Cote d'Ivoire which has an adult rate of 7.1%, most West African countries have lower than 5%. In East Africa, prevalence exceeds 6% in Uganda, Kenya, and Tanzania.

Due to cost, antiretroviral medications reach only a small fraction of Africans who need them. Rwanda and Namibia serve about 70%, Kenya, Malawi, and Zambia range between 25 and 45%, but Ghana, Nigeria, and Mozambique serve fewer than 20%. As a result, life ex pectancy has fallen across the region and averaged 47 years in 2005.

Women are more vulnerable and thus more severely affected than men, and at much younger ages. The major driving forces of HIV/AIDS include poverty and eco nomic inequality which necessitate transactional sex, cultural practices such as widow inheritance, misguided national policies, and neglect such as in South Africa where the government, disputing the link between HIV and AIDS, refused to provide antiretrovirals to prevent vertical transmission of the disease. Global efforts to control the disease through improved testing and supply of inexpensive medications and other programs will help those who need care most in African countries.

Tuberculosis

Tuberculosis (TB) is an airborne respiratory disease caused by *Mycobacterium tuberculosis*. TB bacilli are pro pelled into the air when people who are sick with pul monary TB cough, sneeze, talk, or during normal breathing. Inhaling even a small number of these airborne bacilli is enough to cause infection. Most exposed per sons never become ill but develop latent tuberculosis. An important minority (5–10%) progress from latent to ac tive tuberculosis due, for example, to immunosuppression from co infection with the human immunodeficiency virus (HIV).

Despite treatment advances, tuberculosis remains an uncontrolled communicable disease worldwide. World wide about nine million persons are diagnosed and two million die yearly from tuberculosis particularly in de veloping countries. Currently, more than 90% of the global tuberculosis burden occurs within developing countries due to, among other things, widespread HIV/AIDS infection, malnutrition, crowding, medication shortages, and poor healthcare programs. In developed countries, newly diagnosed TB occurs most frequently among foreign born immigrants, and native born home less who typically live in extremely crowded and un sanitary conditions. For example, in Germany and France immigrants are 3 times and 6 times, respectively, more likely to be diagnosed with TB than are native born persons.

In Canada, while comprising only 18% of the Cana dian population, the foreign born accounted for 65% of all reported TB cases in 2000. In the US, 54.3% of TB cases were found in foreign born persons. Immigrants entering the US and Canada from countries with high tuberculosis incidence have a higher risk of latent or active tuberculosis when they arrive. Thus, states or provinces with high immigrant populations have high rates. For example, British Columbia, Ontario, and Quebec, Canada's leading immigrant destinations, ac counted for 87% of foreign born tuberculosis cases. Moreover, a disproportionate number of the cases were reported from large cities with significant ethnic and minority concentrations such as Toronto, Montreal, and Vancouver. In the US, five states – California, New York, Florida, Texas, and Illinois – account for 66% of the total foreign born tuberculosis cases reported. Similar to the Canadian pattern, major US cities, including New York City, Los Angeles, San Francisco, San Diego, Miami, Atlanta, and Houston, all report disproportionately high numbers of cases from immigrant neighborhoods.

Moreover, the origin of foreign born tuberculosis cases reported from each of these states was largely homogenous, revealing established migration. In 2003, New York reported approximately 64% of the US total of tuberculosis cases born in the Dominican Republic; Florida reported 60% of cases born in Cuba; Texas and California reported 60% of cases born in Mexico; and Minnesota reported 55% of Somali born cases.

Poliomyelitis

Polio is a viral disease that is usually spread through fecal contaminated water and food or oral human to human contact. Early symptoms of infection include fatigue, fever, vomiting, headache and pain in the neck, and ex tremities. The virus invades the nervous system and can cause paralysis in hours. The incubation takes between 3 and 30 days, but a person begins to shed the virus in their stool 7–10 days before, and 3–6 weeks after the onset of symptoms. A small minority of cases (0.5–1%) of cases result in paralysis, and while polio can affect people at any age, at least 50% of cases occur in children 3–5 years old. Most people infected with the polio virus have no symptoms at all but continue to spread the highly con tagious virus for weeks. Thus once introduced, polio spreads very quickly and usually infects many people before an outbreak is detected.

Once considered almost eradicated worldwide, polio has resurged not only in the six countries with endemic polio (Nigeria, India, Pakistan, Niger, Afghanistan, and Egypt) but transmission has been reestablished in an additional six countries (Burkina Faso, Central African Republic, Chad, Côte d'Ivoire, Mali, and Sudan). Despite massive efforts, total eradication continues to be an elu sive goal. The global campaign to eradicate polio by 2005 suffered a severe setback when parts of northern Nigeria rejected the polio vaccine because Muslim clerics claimed that the vaccine causes infertility and spreads HIV/AIDS and cancer. From those few states, polio spread through large parts of the country, neighboring countries, and eventually to Saudi Arabia through Nigerian Muslims on pilgrimage to Mecca. Indonesian Muslims, who were exposed during their pilgrimage in Mecca, carried the disease with them to Indonesia where it became reestablished. Polio reminds us that global cooperation is critical for controlling communicable diseases.

Sexually Transmitted Infections

Sexual behavior of travelers is contributing to spatial mixing of different strains of sexually transmitted pathogens worldwide. Consequently, sexually active tra velers risk being infected with new strains of known sexually transmitted infections. An often cited example is the current endemicity of quinolone resistant *Neisseria gonorrhoeae* (QRNG) in California and Hawaii. This pathogen was originally common only in Asia. According to the CDC, persons who engaged in sexual activity abroad first introduced these strains of QRNG in the US, but now incidence is no longer related to travel. Whereas less than 1% QRNG isolates were identified between 1999 and 2001, 9.4% were reported in 2005, and in people with no history of travel. In short, travel, whether in the form of short term recreational travel or per manent migration, facilitates the spatial spread of disease. Consequently, the spatial distribution of immigrant neighborhoods and ethnic islands may provide an im portant key to understanding the geography of disease diffusion.

Severe Acute Respiratory Syndrome

Severe acute respiratory syndrome (SARS) illustrates how migration facilitates the spatial spread of disease. From its source, Guandong Province in South China, SARS spread initially to Hong Kong from where it dif fused worldwide. When it was over, a total of 8439 probable cases and 812 deaths had been identified from 30 countries. The devastation of SARS was not uniformly distributed, however, but followed a path of spatially linked places and peoples facilitated by international travel.

Canadians returning from Asia imported SARS into Canada. The first cases involved a Canadian family of Hong Kong descent who live in Toronto. A 78 year old woman and her husband traveled to Hong Kong and stayed at a hotel where a cluster of 13 persons with suspected or probable SARS are known to have stayed. Two days after returning home, the woman developed what is now known as SARS and died. Several family members who had close contact with the index case then developed SARS symptoms, and one was later admitted to a hospital that subsequently became the epicenter for the Toronto outbreak. SARS killed 44 Canadians, caused illness in hundreds more, paralyzed a major segment of

Ontario's healthcare system for weeks, and put more than 25 000 residents of the Greater Toronto Area in quarantine.

H5N1 Avian Influenza

Some communicable diseases originate in animals and occasionally mutate to infect humans. An excellent ex ample is avian influenza. It occurs naturally among wild birds; they shed it in their saliva, nasal secretions, and feces, but usually do not get sick from it. Domesticated birds, including chickens, ducks, and turkeys, become infected through contact with contaminated surfaces, birds, or their secretions. Human contact with infected birds or contaminated surfaces produces infection, but human to human spread of avian influenza viruses is rare, typically limited, and unsustained. Symptoms of avian influenza in humans include fever, cough, sore throat, muscle aches, eye infections, severe respiratory diseases, and other life threatening complications. Human influ enza virus refers to those subtypes (currently H1N1, H1N2, and H3N2) that spread widely among humans. In the United States, human flu kills 36 000 people and hospitalizes more than 200 000 people every year.

The current communicable disease of global concern is caused by influenza A (H5N1) virus. Endemic in birds, H5N1 virus is contagious and often deadly. While it usually does not infect people, human infections have occurred, usually in people with direct close contact with infected poultry or contaminated surfaces. About 60% of the human cases have died, mostly as healthy children and young adults. Because humans have little pre existing natural immunity to H5N1 infection, it is feared that a pandemic will results if these H5N1 viruses gain the ability for efficient and sustained transmission among humans. While there has been some human to human spread of H5N1, it remains limited. The H5N1 virus responsible for recent human illness and death in Asia is resistant to two common antiviral flu medications and currently no commercially available vaccine to protect humans exists.

Research continues into new vaccines to counter a possible H5N1 pandemic but unfortunately, current vaccine production methods require an outbreak before an effective vaccine can be developed and even then it takes between 5 and 7 months. To avoid a global pan demic of H5N1 or other virus, new vaccine technologies are required, to respond quickly and effectively to new challenges of communicable disease outbreaks.

Mapping Communicable Diseases

Ever since the seminal work of Dr. John Snow, mapping has been a key tool in fighting communicable diseases. Thanks to recent technological advances, particularly in GIS, the role of mapping in disease tracking, surveillance, and control has burgeoned. Recent examples include West Nile Virus, and tuberculosis control. Cromley and McLafferty provide an excellent overview of GIS ap plications in public health generally.

One particularly fascinating recent development is mapping genotypes of communicable disease. Each disease causing organism has a specific phenotype, and the species members have a more unique genotype. Tools exist for separating and categorizing disease organisms with these more unique genotypes into groups based on their genetic structure (isolates). Two people with identical or matching isolates indicate recent transmission from a common source; unique isolates indicate remote transmission from a different source. Analyzing and mapping these genotypes allows researchers to distinguish between not just disease species (phenotype) but also locate areas with high inci dence and different strains of the disease, including recent versus remote transmission of disease. Areas with high rates of clustered genotypes indicate ongoing transmission (i.e., an outbreak), whereas multiple unique isolates indicate infections that were acquired elsewhere (are imported).

Moonan *et al.* based a recently successful TB inter vention on geographically targeted screening of clustered tuberculosis isolates. Because it assumed that persons with clustered strains of disease from geographically re lated areas represent ongoing community transmission that can be identified and interrupted through treatment, the study focused resources on zip codes with high in cidence and clustered isolates. During the 28 month intervention, the number of cases of tuberculosis de creased from 28.5 cases per 1000 screenings to 2.4 cases per 1000. The rate of developing latent tuberculosis in fection fell from 14.3 to 2.2 per 100 person years of ex posure. The intervention was successful because it distinguished between high incidence areas with high and low levels of clustered strains, and deployed re sources not just on the basis of incidence but on the basis of genotype. Since unique isolates represent remote transmission, location based screening in areas with high rates of unique isolates would be less likely to identify either persons with active tuberculosis or recent acquired latent tuberculosis infection and thus not as cost effective.

Conclusion

Communicable diseases remain a major threat to hu mankind. Effective surveillance and response remain our best protection against these deadly threats. Unfortu nately, surveillance remains poor and rudimentary in the most vulnerable countries where outbreaks are most likely. Strengthening the health and surveillance systems of poor developing countries is critically needed in a globalized world because communicable diseases do not respect political boundaries.

See also: Diffusion; Disease Diffusion; Disease Mapping; Health Geography; Medical Geography.

Further Reading

Booth, C. M., Matukas, L. M., Tomlinson, G. A. *et al*. (2003). Clinical features and short term outcomes of 144 patients with SARS in the greater Toronto area. *Journal of American Medical Association* 289, 1 9.

Brudey, K., Driscoll, J. R., Rigouts, L. *et al*. (2006). *Mycobacterium tuberculosis* complex genetic diversity: Mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiology* 6, 23 (doi:10.1186/1471 2180 6 23).

Carballo, M. and Nerukar, A. (2001). Migration, refugees and health risks. *Emerging Infectious Diseases* 7(supplement 3), 556 560.

CDC (2005). Controlling tuberculosis in the United States recommendations from the American Thoracic Society, CDC, and the Infectious Diseases Society of America. *Morbidity and Mortality Weekly Report* 54(RR12), 1 81.

CDC (2005). Gonorrhea STD surveillance 2005. http://www.cdc.gov/std/stats/gonorrhea.htm (accessed May 2008).

CDC (2006). The global HIV/AIDS pandemic, 2006. *MMWR* 55(31), 841 844.

CDC (2006). Trends in tuberculosis United States, 2005. *Morbidity and Mortality Weekly Report* 55(11), 305 308.

CDC (2007). Key facts about avian influenza (bird flu) and avian influenza A (H5N1) virus. Factsheet. http://www.cdc.gov/flu/avian/gen info/pdf/avian facts.pdf (accessed May 2008).

Cromley, E. K. and McLafferty, S. L. (2002). *GIS and Public Health*. New York: Guilford.

Denevan, W. M. (1992). The pristine myth: The landscape of the Americas in 1492. *Annals of the Association of American Geographers* 82(3), 369 385.

Etkind, P., Ratelle, S. and George, H. (2003). International travel and sexually transmitted disease. *Emerging Infectious Disease*. http://www.cdc.gov/ncidod/EID/vol9no12/03 0210.htm (accessed May 2008).

Gerbase, A. C., Rowley, J. T. and Mertens, T. E. (1998). Global epidemiology of sexually transmitted diseases. *Lancet* 351(supplement 3), 2 4.

Havlir, D. V. and Barnes, B. F. (1999). Tuberculosis in patients with human immunodeficiency virus infection. *New England Journal Medicine* 340(5), 367 373.

Health Canada (2003). Tuberculosis in Canada, 2000. http://www.phac aspc.gc.ca/publicat/tbcan00/pdf/tbcan 2000 e.pdf (accessed May 2008).

Health Protection Agency Centre for Infections, London (2005). Annual report on tuberculosis cases reported in England, Wales and Northern Ireland in 2003. http://www.hpa.org.uk/infections/topics az/tb/pdf/2003 Annual Report.pdf (accessed May 2008).

Kalipeni, E., Craddock, S., Oppong, J. R. and Ghosh, J. (eds.) (2004). *HIV and AIDS in Africa: Beyond Epidemiology*. Oxford: Blackwell Publishing.

Kearns, R. (1993). Place and health: Towards a reformed medical geography. *Professional Geographer* 45, 139 147.

Kearns, R. (1994). Putting health and health care into place: An invitation accepted and declined. *Professional Geographer* 46, 111 115.

Kearns, R. (1994). To reform is not discard: A reply to Paul. *Professional Geographer* 46, 505 507.

Long, R., Njoo, H. and Hershfield, E. A. (1998). Tuberculosis: Epidemiology of the disease in Canada. *Canadian Medical Association Journal* 160(8), 1185 1190.

Mayer, J. D. and Meade, M. S. (1994). A reformed medical geography reconsidered. *The Professional Geographer* 46(1), 103 106.

Moonan, P. K., Bayona, M., Quitugua, T. N. *et al*. (2004). Using GIS technology to identify areas of tuberculosis transmission and incidence. *International Journal of Health Geographics* 3(1), 23 33.

Moonan, P. K., Oppong, J. R., Sahbazian, B. *et al*. (2006). What is the outcome of location based targeted tuberculosis screening based on universal genotyping? *American Journal of Respiratory and Critical Care Medicine* 174, 599 604.

Oppong, J. R., Denton, C., Moonan, P. K. and Weis, S. E. (2007). Foreign born status and geographic patterns of tuberculosis in Tarrant County, Texas. *Professional Geographer* 59, 478 491.

Paul, B. K. (1994). Commentary on Kearns's 'place and health: Toward a reformed medical geography'. *The Professional Geographer* 46(4), 504 505.

Schneider, E., Moore, M. and Castro, K. G. (2005). Epidemiology of tuberculosis in the United States. *Clinic in Chest Medicine* 26(2), 183 195.

Snow, J. (1855). *On the Mode of Communication of Cholera*. London: John Churchill. http://www.ph.ucla.edu/EPI/snow/snowbook.html (accessed May 2008).

Struewing, J. P., Hartge, P., Wacholder, S. *et al*. (1997). The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. *New England Journal of Medicine* 336(20), 1401 1408.

Thompson, M. M. and Najera, R. (2001). Travel and the introduction of human immunodeficiency virus type 1 non B subtype genetic forms into western countries. *Clinical Infectious Diseases* 32, 1732 1737.

Versalovic, J. and Lupski, J. R. (2002). Molecular detection and genotyping of pathogens: More accurate and rapid answers. *Trends in Microbiology* 10, S15 S21.

Wallace, J. M., Rosen, M. J. and Reichman, L. B. (1997). Incidence of tuberculosis in the United States among HIV infected persons. The pulmonary complications of HIV infection study group. *Annals of Internal Medicine* 126, 123 132.

Watson, J. T., Jones, R. C., Gibbs, K. and Paul, W. (2004). Dead crow reports and location of human west nile virus cases, Chicago, 2002. *Emerging Infectious Diseases* 10(5), 938 940.

World Health Organization (2004). WHO guidelines for the global surveillance of severe acute respiratory syndrome (SARS). http://www.who.int/csr/resources/publications/WHO CDS CSR ARO 2004 1.pdf (accessed May 2008).

World Health Organization (2005). Global tuberculosis control: Surveillance, planning and financing. http://www.who.int/tb/publications/global report/2005/pdf/Full.pdf (accessed May 2008).

Relevant Websites

http://www.polioeradication.org/
Global Polio Eradication Initiative.

http://userpages.umbc.edu/~earickso/Index.html
Health and Medical Geography Specialty group of Association of American Geographers.

http://www.unaids.org/en/
Joint United Nations Program on HIV/AIDS.

http://www.nap.edu/catalog.php?record id=11588
The Impact of Globalization on Infectious Disease Emergence and Control: Exploring the Consequences and Opportunities, Workshop Summary Forum on Microbial Threats (2006).

http://www.cdc.gov/tb/
United States Centers for Disease Control and Prevention (CDC) Tuberculosis website.

http://www.who.int/csr/sars/en/
World Health Organization Severe Acute Respiratory Syndrome website.

Communist and Post-Communist Geographies

J. Timár, Center for Regional Studies, Hungarian Academy of Sciences, Békéscsaba, Hungary

Glossary

Forces of Production/Productive Forces The forces of production are the forms in which nature and raw materials are appropriated and used in the labor process. The productive forces are the unity of means of production and labor (with the mental and physical capabilities and experiences of human being exercised in labor process).

Means of Production The unity of the instruments and the subject of labor.

Mode of Production The method of producing the necessities of life. It is the unity of the productive forces and the relations of production. At a certain stage of their development, the productive forces come into conflict with the existing relations of production. This conflict is the economic basis of social revolution and leads to change in the entire society. Classic Marxist texts describe the historical sequence of modes of production such as capitalism which gives way to socialism.

Relations of Production The objective material relations formed between all people in the process of social production, exchange, and distribution of material wealth. The basis of these relations is ownership of the means of production. While capitalist property is a form of private property, social property appears in the form of public or state property in socialism and cooperative and community property in communism.

Efforts in geography to address the issues of socialism/communism failed to be seminal to the emergence of any powerful theory that could have been the result of putting the perspective of scientific communism on geography. It was the political economic approach of Marxism that imbued the geographies of state socialist countries the profoundly. Human geography was replaced with economic geography. Attention was focused on production.

In the Stalinist era and during the Cold War, ambitions to defeat 'bourgeois pseudo sciences' and building a socialist society were assigned as tasks to sciences. Geographical research was under state control and in support of the central planning. Physical and human geography were separated at the level of institutions for a long time.

In the countries of post socialist transition, geography is characterized by (1) the restoration of the 'pre socialist' relationship between economic and human geography; (2) a wider spectrum of researched areas, specialization; and (3) the emergence of pluralism in philosophy and social theories. The Western versions of post socialist geographies differ from their Eastern counterparts mainly in terms of their approaches, methodologies, and interpretation of space. They provide evidence of the fact that post socialism is both an Eastern and a Western phenomenon.

Definitions of Communist Geography

Geographies which (1) use the theoretical framework and concepts of scientific communism (socialism) for studying geographic issues; (2) provide support for a communist social and economic formation (mode of production) and social and political movements fighting for such a society; and/or (3) are institutionalized geographies cultivated in countries building state socialism (communism) can be regarded as 'communist'.

The main underlying reason for such a multiapproach definition is the numerous meanings of communism. One such meaning is what is called 'scientific communism', which is – in addition to dialectical and historical materialism and political economy – an important part of Marxism and Leninism; it is a theory according to which it is inevitable that communism should replace capitalism and which explores the possible directions of building a socialist communist society. One of the tenets of this theory is that history is the story of class struggles. Under capitalism, the bourgeoisie owning the means of production exploits the proletariat that only owns its labor through the surplus value generated by the latter. In the same society, however, conditions for eliminating social exploitation and establishing a new socialist social order, free from the contradictions of capitalism, were also created. Production, which also acquired a social feature through the multitude of workers concentrated in large factories, requires that means of production should be in public ownership. Balance between the forces and relations of production can be restored through the nationalization of the means of production, that is, through the exploitation, for the benefit of the society, of exploiters owning the means of production. The way to a new classless society is a proletarian revolution. Marx and Engels' theory of scientific communism in the mid nineteenth century was preceded by several other concepts and tenets – ranging from Plato's Utopia of communism, theories on socialism in the era of feudalism, the petty bourgeois and bourgeois version of socialism to critical utopistic socialism and communism – on the transformation of the society along socialist principles and the principle of a just and equitable society. According to Marx and Engels, however, as class struggle becomes more sophisticated so even critical utopistic socialism and communism becomes an obsolete, reactionary, or even conservative theory on socialism. Advocates of this theory intended to achieve their goals in a peaceful manner. The theory itself is not based on extensive knowledge of the laws and motives of social development and fails to recognize the historical inevitability of a socialist revolution and 'the vocation of the proletariat in world history'. By contrast, Marx considered scientific communism to be the generalized collective experience of the revolutionary movement of the working classes. Thus, communism also means a sociopolitical movement aimed at overthrowing capitalism and building a communist society.

Finally, communism is also used to refer to a socioeconomic formation which is based on the public ownership of the means of production and which can turn the principle of 'to each according to his needs, from each according to his ability' into reality. The first stage of the evolvement of this social formation is socialism, after it has been fully consolidated, followed by full communism with the complete 'withering away of the state'. Not all the Marxists of the nineteenth century shared the view on the necessity of a socialist stage characterized by the dictatorship of the proletariat or on the characteristics of this stage as were proposed by Marx. Conducted with anarchists, debates focused mainly on how the revolution, socialism, and the exploiting state were interrelated. On the eve of the revolution in Russia, Lenin worked out (and even 'fine tuned') both the strategy of the revolution and the theoretical foundations of how a socialist state should be built. In order for practical tasks that emerged during the building of a socialist state, for example, socialist industrialization, the reorganization of agriculture, and the cultivation of a 'socialist mindset', to be fulfilled, an increasingly high number of new theoretical issues had to be tackled. The Russian Communist Party took such tasks in its stride and regarded them to be the order of the day.

Considering the fact that the characteristics of communist geography, which later evolved into post communist (post socialist) geography, can best be studied in the former Soviet Union and in East Central European countries, which turned Marxism into a state ideology and which are today characterized by a post socialist (post communist) social order, what follows describes experience in this region.

Communist Geography as an Approach Linked to the Theory of Scientific Communism

We can hardly speak of communist geography as an approach that is linked to Marx's theory on scientific communism and one that distinctly differs from other approaches.

Radical Western geographies of the 1960s and 1970s proposed concepts such as 'integrated labor' (rather than a division of labor), self sufficient regions (which interchange both products and ideas with other regions), and community control, which used an alternative society based on anarchist principles as its starting point. In geography, the most powerful advocate of what was called 'anarchist communism' was Russian theoretician and geographer Kropotkin. Published at the turn of the twentieth century, his works argued against both political and economic centralization. They cherished the idea of communities operating on the principle of cooperation and mutual support, where the 'mess' of the people will organize itself by itself. Geographies based on the principle of anarchist communism were, however, diametrically opposed to the Marxist–Leninist interpretation of state level 'authoritarian communism'.

There is a school, however, which not only has brought Marxism into geography, but has also put geographical perspectives on Marxism. However, not even this school is labeled as communist; rather, it is referred to as Marxist geography, with the adjective 'Western' used for the sake of clarity and in order to set it apart from the geography cultivated in state socialist countries in the East. Followers of Marxist geography primarily use political economy and dialectical and historical materialism proposed by Marxism – for example, Harvey created a completely new 'historical geographical materialism'. Fundamentally, they were committed to studying capitalism. They did not dig deep into the theory of scientific communism or the workings of state socialist regimes, toward which they adopted a critical attitude.

In contrast, human geography did address the issues of socialism (communism) in state socialist countries. However, unlike radicalism that characterized the theory of scientific communism (and Western Marxist geography) at its (their) inception, geography in these countries was technocratic rather than radical. It failed to be seminal to any powerful theory that could have been the result of bringing scientific communism into geography. Paradoxically, the creation of such theories was foiled mainly by the very totalitarian system, in which the ruling communist party, reiterating the building of socialism (communism), turned to the theory of scientific communism and Marxism–Leninism as a whole in order to justify its own etatism. Justification, however, came at the price of distorting the original theory. In general, the end of Stalinism also marked the end of an era when social sciences had to meet strict criteria and were under close scrutiny. However, the picture was far more subtle. Brought about by the popular movements/revolutions of the mid 1950s in GDR, Poland, and Hungary, the Prague uprising in 1968 and the cultural revolution in China, etc., changes, which differed from one country to the next, also affected geography.

Especially in the Stalinist era, but several decades later as well, as attested to by the works of a few researchers, geography contributed to a state level corruption of Marxism and Leninism. One of the most frequent manifestations of such corruption was the permanent use of the terminology of scientific communism and, in a broader sense, that of Marxism and Leninism. Thus, for instance, the word 'population' in a demographic analysis was only too readily replaced with 'the forces of production'. Likewise, terms such as 'mode of production', 'relations of production', or 'territorial division of labor' were used left, right, and center in literature. Sentence starters like 'In compliance with the requirements of our socialist society ….' were unlikely to be the results of the actual study of social inevitabilities. Rather, they were prescribed by the communist party or used simply as propaganda sound bites or as declarations of a sense of belonging (the use of the various forms of the first person plural) and personal commitment.

It was perhaps the political economic approach of Marxism that imbued the geographies of state socialist countries the most. Human geography was replaced with economic geography from the outset. Industrial, agrarian, settlement, and population geographies were, as a rule, regarded to be its subdisciplines. Attention was focused – while in several countries social and cultural geographies were simply nonexistent – on production, in connection with a materialist *weltanschauung* and the political practice of forced industrialization. It was a common 'belief' that the development of the productive sectors would automatically lead to the development of the nonproductive ones, improve living conditions, and resolve the problems of social reproduction. Research focused on structures, especially on the macro scale. Locational analyses and the quantitative models of the optimal regional allocation of production with a command economy perspective put on them had the most far reaching influence in socialist countries. Theories of the 'territorial productive complex' (TPC) and economic regions (rayons), worked out in the Soviet Union, also 'reached' other countries, though their feasibility was the subject of heated debates.

In contrast to Marxist geography in Western countries, geography in state socialist countries did not participate in the conceptualization of the issues of class relations. Although descriptions of social structures were 'sneaked into' demographic analyses, changes in the ratio of the working classes to peasantry in agricultural cooperatives (and those in their living standards presented in academic literature from the 1970s and the 1980s) and how these changes were interconnected with industrialization and urbanization were the 'hottest' topics of the era. At best, geography refrained from going into details regarding the 'individualism of the peasantry' or the 'unprogressiveness of villages'. Marx did not 'promise' an equal standard of living for the socialist stage, since it was division according to the principle of labor rather than that of needs that prevailed. Nevertheless, official party ideology made egalitarianism into one of its high sounding slogans. This is very likely to be one of the reasons why the Marxist theory of uneven development was not further elaborated during state socialism. Whenever regional backwardness was identified, it was not its underlying reasons that were examined. The fact that certain elements of the capitalist mode of production prevailed for a while was considered to be a natural outcome of the laws of dialectics. Therefore, whenever backward regions/settlements were studied, it was development opportunities rather than the causes of backwardness on which attention was focused, thereby offering assistance for the state with central planning. There were countries where debates pointing out the negative role of the state were conducted, as was the case in Hungary in the 1980s in the case of sociologists Szelényi, Manchin, and Ferge. However, geographical experiences did not feature in such debates. The above debate was about whether market mechanisms offset inequalities generated by the state in the socialist era, or they add their own inequalities, which happen to benefit the same social strata. To give Marxist theoreticians their due, it should be noted that those (e.g., Lukács, Bloch, Kołakowski, and Gyilasz) whose oeuvres proved to be influential found themselves in confrontation with the ruling orthodoxy: they were deported, imprisoned, or forced to emigrate from their home countries.

Communist (Socialist) Geography Supporting the Building of State Socialism

Marxism strives to both understand and change the surrounding world. Since the overthrow of capitalism was no longer a topical issue in state socialism, during the Stalinist era and the Cold War struggle against 'bourgeois pseudo sciences' and the building of a socialist society were assigned as tasks to sciences. The latter task (i.e., the building of a socialist society) was in conformity with the tenet according to which Marxist–Leninist epistemology was based on practice. Geography, too, would openly identify the satisfaction of the needs of the society with the objective of 'serving working people'. However, apart from activities aimed at popularizing sciences among workers, such services were a far cry from today's people's geography project initiated by Western Marxist and critical geographies. In fact, they mostly meant applied research for and in support of the state. From then on, only within the boundaries of the prevailing *status quo* was it possible to make any change.

If there is communist (socialist) geography, direct or indirect assistance with central planning is very likely to be its universal characteristic. Economic (human) geography was most closely related to economic, regional, and settlement network development and planning in several different ways. Atlases popularizing the 5 year plans of the 'people's economy', empirical studies 'laying down the geographical foundations of economic development', and theories and models that made the delimitation of economic regions possible were the 'products' of the era. Staffs at research institutions had to contribute to planning at various scales and some researchers were retained as experts. As tight political control was easing so some criticism was also voiced. Several countries adopted Soviet experience in the participation of geography in the gigantic 'transformation of nature' or at least references were made to such utilization of research. It was sometimes the case that geographers, not always making it a point of honor to refer to their sources, imported Western theories such as Christaller's theory on the central place.

In the 1970s and 1980s, numerous studies (e.g., rural research and in Poland and Hungary social geographic research of German and British origin with an eye on practical application) were carried out in East Central Europe, which did put forward proposals for improving the living conditions of those in need. No comprehensive evaluation of such studies or applied geographic studies is available as yet. Some of the retrospective theoretical studies in the post socialist era assert that not only geographical expertise but also expertise in social sciences in general were used as a tool for the legitimation of the political will and power.

Geography in State Socialist Countries

The term 'communist' geography – better known as 'socialist' geography – was the original 'official' term used to set institutionalized geography in state socialist countries apart from other geographies. In this sense, all kinds of geographies cultivated in those countries were socialist (communist). The authoritarian regimes 'saw to it' that geography, and any discipline for that matter, was or at least looked ideologically consistent. The most universal characteristic of geographical knowledge produced in such a social environment was, therefore, its attitude toward the ruling powers, which mainly affected the network of geographical institutions.

This institutional system was controlled and, hence, centralized by the state (the communist party). Institutes of geographic research which were spun off from universities and which were parts of the national academies of sciences were established. These institutes, adjusting to central planning in people's economy, worked according to national research plans. Academies were responsible for using a Soviet type scientific rating system, which proved to be an efficient tool for state controlled cultural policy to establish a cadre system in the profession as a whole. During the Stalinist era, many geographers in several countries were severely criticized, stripped off their academic degrees, dismissed from work, or arrested. As actual cases attest to it, reasons included a World War II role, religious affiliation, a statement negating socialist development (e.g., that a given country is an agrarian rather than an industrial–agrarian country), the use of a 'bourgeois' approach (e.g., a morphologic rather than functional approach adopted in urban study), or the taking over of the locational theory from Western geographies. The scale of such 'cleansing' was completely different in the German Democratic Republic and Hungary (in the latter country even the Geographic Society, operational since 1875, was banned), due to their roles in the war, from that in Poland, where access to Western literature was easier and which had a wider network of international relations. Soviet hegemony was felt nearly everywhere; for example, guest tutors and researchers contributed to the teaching and the research carried out by their hosts and published in their journals, studies published in the Soviet Union were translated, etc. Generally, the grip of this hegemony only slowly eased, except in China, which broke away from the Soviet Union completely, Yugoslavia, which also kept its distance, and later in Romania and Albania. With dictatorship becoming softer, the number of bilateral Western relations grew and, in addition to the Council for Mutual Economic Assistance (COMECON) research, wider institutionalized cooperation in the International Geographic Union also played an important role.

The separation of physical geography from economic (human) geography was also reflected in institutional fragmentation. The story of such 'legal separation' and later their unification during socialism can be less well interpreted on a philosophical basis than in a political economic context. Since the spread of the dualist approach led to breaking away from the pre socialist tradition of unified geography in most countries, geographers turned to Marxism for explanation. Explanation was provided by the fundamentally different laws that underlay the development of the phenomena in the 'two geographies' and the suspected dangers posed by environmental determinism branded as 'bourgeois'. In the Khrushchev era, prompted by Anuchin, a major debate was sparked between 'monists' and 'dualists' in the Soviet Union. Finally, the party joined in the debate and declared that the rigid separation of nature from society was erroneous and of a 'Stalinist nature', expressing that such separation ran counter to (economic) planning interests. The fact that reasoning had identical roots clearly reflected that 'theoretical debates' had nearly nothing to do with philosophy proper. Rather, such debates were a reflection of the power struggle. Both parties sought justification in the tenets of Marxism–Leninism, the practical need for planning and the national geographies of the pre Stalinist era. At the time of a similar change in paradigms, Romania rediscovered its national roots, and Mihailescu's 1936 university textbook advocating the unified approach was reissued in 1969. For such, a change in orientation in the Soviet Union was necessary. Nevertheless, the explanation provided for the reissuance of the book was also in line with Romania's current regime's concern with nationalism. The fact that, for instance, in China the balance between physical and human (economic) geography could not be restored can be ascribed to its complete confrontation with the Soviet Union. It is another story that the dominance of physical geography – being politically less risky – prevailed in most state socialist countries even in the era of the new integrative 'constructive geography' and despite the spread of the anthropocentric approach of landscape and environmental research.

It is hardly a coincidence that integrated landscape geographic research adopting a new system approach became one of the areas that challenged descriptive static view. This, however, left the empirical or positivist nature of most human geographic research unaffected. The revamping and spread of quantitative methods also occurred in state socialist countries mainly in the 1960s and 1970s and due particularly to the weight of applied geography to which regional science quickly found its way. However, revamping failed to materialize in qualitative methods and social theories, except the Marxist–Leninist ones.

Characteristics of the Post-Communist/Post-Socialist Geographies

Unlike the term 'communist geography', 'post communist/post socialist geography' had a more unambiguous and more uniform meaning and gained more currency: it is used to refer to geography focusing on research on transition from state socialism (communism) into capitalism. Considering the fact that 1989–91 political changes in East Central Europe and the Soviet Union occurred before communism had been reached (only the Soviet Union declared officially the reaching of the stage of full communism), it stands to reason that the adjective 'post socialist' was more frequently used than 'post communist'.

Post-Socialist Geographies in Post-Socialist Countries

Naturally, post socialist geographies are cultivated in the countries concerned, that is, post socialist countries. Privatization, market economy, and multiparty democracy have changed the relationship of geographical knowledge production with political powers, its institutional frameworks and nature. The length of time for political, economic, and social transition to occur seems to be varying from one country to the next, which further differentiates, the national human geographies of the former Soviet block. These geographies are in the state of transition themselves: current trends are, in several countries, the organic continuation of the processes that started in the 1980s (or even earlier). In the early 1980s, Kuklinski deemed developments in Polish geography to be a change in paradigm in which the issues/problems of economic, social and ecological crisis, spatial differentiation, new local and regional communities, and the changing system of values must be addressed/resolved with new methods and by using interdisciplinary relations. At an international meeting of the directors of academy run geographic research institutions in September 1989, the representatives of several countries already put the tasks and situation of geography in the context of a 'political and economical crisis', 'social movements', and 'perestroika'. By contrast, at the same meeting, a participant from Cuba emphasized the importance of the second edition of a national atlas and a Mongolian attendee pointed out, among other things, the weaknesses of the supply of cadres. In the countries of post socialist transition, despite their differences, the transformation of human geography shares a few universal characteristics:

1. The relationship between economic and human geography has become what it used to be: they have 'changed places' in terms of their weight, hierarchical relationship, and institutional background. Research institutes, no longer controlled by party politics, found themselves to be subject to market conditions. University faculties have been reorganized – in general, there has been a marked change in direction, with everyone looking to the 'West'.
2. Human geography has become more diverse and further specialized. There has also been a renaissance of political geography, boosted by increasingly strong nationalism and as a revival of the traditions of 'bourgeois geography'. Numerous geographers have been rehabilitated and several areas of research that used to be off limits (e.g., geography of religions) are now again accepted. Charged with new political content, ethnic geography has achieved high visibility. New subdisciplines have emerged including, for example, election geography, which was inconceivable in the era of state socialism owing to the lack of free general elections. And geography of gender, though with low visibility, is already a presence. Obviously, further examples could be put on this 'list of changes'.

3. Democratization also paved the way for pluralism in philosophy and social theories. First, challenging Marxism and Leninism rather than the emergence of unmistakably new trends is more common. Currently, experience gained in the totalitarian regimes seems to be supporting 'value neutral sciences'. 'Situated knowledge' has not even been raised as an issue to be debated. Social and cultural theories as fomenters have hardly been able to penetrate into established geography. Critical geography has also hardly been able to gain in popularity in part for the above reasons and in part because of the negative connotation as sociated with left wing thinking in post socialist countries.

Some claim that characteristically post socialist dir ections of development underlie shifts in views that are relatively comprehensive and in keeping with Western trends. One example that supports this claim is a shift from the earlier dominance of structuralism toward human agents, individuals, their spatial behavior as well as the cognitive and decision making factors involved in such behavior. Underlying this trend is vigorous liberal ism represented by individualism. The latter was inspired under the new political circumstances by the need to 'liberate' the individual oppressed by the socialist state. There has also been a shift in the political economic focus in several countries. The simultaneous emergence/ strengthening of cultural geography has not, however, been a response to an increasingly self conscious cultural policy. Nor has it been inspired by identity based post colonial or environmental social movements, since such movements are rather weak. This is, in part, the reason why 'new cultural geographies' in post socialist countries are not linked to, for example, the homeless, the disabled, or the homosexuals whose spatiality is rather narrow, relative to dominant cultural realms. For the time being, it focuses on (mainly ethnic) groups that were 'off limits' in the socialist era.

Relations of Post-Socialist Geographies in the 'East' and in the 'West'

There have always been researchers in the West who have taken an interest in studying state socialist coun tries. By today, the Western versions of post socialist geographies have become institutionalized: in addition to a number of interdisciplinary institutes operational in several parts of the world, there is a Russian, Central Eurasian, and East European Speciality Group in the Association of American Geographers and a Post Socialist Geographies Research Group in the Royal Geographical Society and the Institute of British Geog raphers. These organizations help to bring together scholars with an interest in post socialist issues and organize East–West workshops. Similar to their Eastern counterparts, members of these organizations also con tribute to revealing the national and local characteristics of key social, political, and economic trends in post communist transition. However, while researchers in the West focus on for instance governance, power structure; socioeconomic and gender inequalities, social and cul tural identities, and marginalization, those in the East adopt a different approach to such issues, focusing mainly on regionalization, decentralization; regional, settlement inequalities; territorial identities, and underdeveloped areas.

Overall, though both 'parties' adopt a wide variety of approaches to research on post socialism, it seems that in post socialist countries, at least for the time being, few regard geographical relations to be the outcome of social processes (and when they do, it is to a much lesser ex tent). Rather, they treat such relations as spatial patterns. One reason is the dominance of the concept of absolute space – space as a container, space as separated from concrete social processes. Furthermore, the mapping of spatial patterns is of quantitative nature. The fact that the development of qualitative methods has been put on the back burner does not help strengthen 'social' in social geography. Attention is paid to the real material world rather than the symbolic space. There is also a strong social demand for sticking to reality and empiricism, the current challenge being the understanding of a new era where the importance of mapping and documenting is higher than average. The significance of applied geo graphy not only remains, but in East Central Europe it is also further strengthened because of regional European Union (EU) planning and because it is a source of income for many institutes. Applied geography, however, con tinues to be used in studies ordered mainly by the state. No studies as yet have been carried out in furtherance of social movements and in order to examine spaces of resistance.

One of the benefits of shared interest in post socialism and of joint East–West projects is, among other things, that they do examine such subject matters. Nevertheless, with sovietization completed, criticism has been voiced concerning the dangers of the new postcolonializing of geographical knowledge and a Western (and, within that, British and American) hegemony. Advocates of relational geographies and the reflexive approach are seeking to mitigate such risks. The key starting point of research adopting such an approach is that post socialist transition reshapes East and West alike and post socialism is both an Eastern and a Western phenomenon.

See also: Anarchism/Anarchist Geography; Applied Geography; Balkans; Capitalism; Capitalism and Division of Labor; Chinese-Language Geography; Cold War; Critical Geography; Cultural Geography; Determinism/ Environmental Determinism; East/West; Economic

Geography; Europe; German-Language Geography; Governance; Harvey, D.; Hegemony; Ideology; Kropotkin, P.; Liberalism; Marxism/Marxist Geography I; Marxism/Marxist Geography II; People's Geography; Physical Geography and Human Geography; Political Economy, Geographical; Political Geography; Positivism/Positivist Geography; Regional Science; Resistance; Russian-Language Geography; Situated Knowledge, Reflexivity; Social Class; Social Geography; Socialism; State; Structuralism/Structuralist Geography; Territorial Production Complexes; Uneven Development.

Further Reading

Demek, J. (1980). The geographical prognosis in present day Czech geography. *Sborník Československé Geografické Společnosti* 85(1), 3 8.

Domański, B. (2004). West and East in 'New Europe': The pitfalls of paternalism and a claimant attitude. *European Urban and Regional Studies* 11(4), 377 381.

Dovényi, Z. and Hajdú, Z. (2000). Evolution of human geographical thought in Hungary in the 20th century. In Kovács, Z. (ed.) *Hungary Towards the 21st Century: The Human Geography of Transition*. Studies in Geography in Hungary 31, pp 29 48. Budapest: Geographical Research Institute HAS.

Enyedi, Gy. and Kertész, A. (1984). South east Europe. In Johnston, R. J. & Claval, P. (eds.) *Geography Since the Second World War*, pp 64 78. Sydney: Croom Helm.

Hooson, D. J. M. (1984). The Soviet Union. In Johnston, R. J. & Claval, P. (eds.) *Geography Since the Second World War*, pp 79 106. Sydney: Croom Helm.

Jalowiecki, B. (1986). Social geography in Poland. In Eyles, J. (ed.) *Social Geography in International Perspective*, pp 172 184. London: Croom Helm.

Kuklinski, A. (1984). Poland. In Johnston, R. J. & Claval, P. (eds.) *Geography Since the Second World War*, pp 132 155. Sydney: Croom Helm.

Lichtenberger, E. (1984). The German speaking countries. In Johnston, R. J. & Claval, P. (eds.) *Geography Since the Second World War*, pp 156 184. Sydney: Croom Helm.

Marx, K. (1970). Critique of the Gotha Programme. In *Marx/Engels Selected Works*, vol. 3, pp 13 30. Moscow: Progress Publishers. Online Version: http://www.marxists.org/archive/marx/works/download/Marx Critque of the Gotha Programme.pdf

Marx, K. and Engels, F. (1969). Manifesto of the Communist Party. In *Marx/Engels Selected Works*, vol. 1, pp 98 137. Moscow: Progress Publishers. Online Version: http://www.marxists.org/archive/marx/works/download/manifest.pdf

Peil, T. (2006). Emerging, submerging and persisting ideas: Is there social and cultural geography in Estonia? *Social and Cultural Geography* 7(3), 463 492.

Stenning, A. (2005). Out there and in here: Studying 'eastern' Europe in the west. *Area* 37(4), 378 383.

Timár, J. (2007). Gender studies in the gender blind post socialist geographies of East Central Europe. *BELGEO* 3, 349 369.

Turnock, D. (1982). Romanian geography reunited the integrative approach demonstrated by the conservation movement. *GeoJournal* 6(5), 419 431.

Relevant Websites

http://www.marxists.org
Marxists Internet Archive.

www.geog.plym.ac.uk
Royal Geographical Society (with the Institute of British Geographers) Post Socialist Geographies Research Group.

http://www.rceeegeography.org
Russian, Central Asian and East European Specialty Group of the American Association of Geographers.

http://www.jiscmail.ac.uk
The 'Cities after Transition' CAT list.

Community

S. C. Aitken, San Diego State University, CA, USA and National Technical Norwegian University, Trondheim, Norway

Glossary

Anomie The breakdown of social norms in modern urban society and the concomitant rise of social pathologies.
Communitarianism Belief that values exist in public space and community is an important part of moral development.
Dreamscapes Affluent suburban and exurban escapist communities.
Individualism A moral, political, and social outlook that stresses human independence and the importance of self-reliance.
Liberalism A belief system that supports individual rights, and extends this to *laissez-faire* economic policies.
Neoliberalism Particular forms of economic and political relations that favor self-governance and a free-market ethic applied to all sectors of society.
Privatopia Evan McKenzie's term for the privatization of local communities esthetically, legally, and fiscally in gated communities and home-owner associations.

Community is one of those concepts about which social and cultural geographers worry. It is a difficult term to tie down, and it is used in a myriad of ways by all spectrums of society. In his influential *Keywords*, Raymond Williams points out that unlike other terms for social relations, such as state and nation, and even society, the term 'community' is almost never used unfavorably. Its ap probatory character and its abilities to prefigure meta phorical unity make the term valuable to all sides of political debate. Importantly from a geographic per spective, identities forged at other spatial scales, such as home or nation, are easily rolled into struggles over community. The problem for geographers is to decipher this struggle in terms of the ways community elaborates a politics of space.

This article begins with some common wisdom about what constitutes community as a form of social and spatial organization. Then it briefly traces some academic debates on place identity and community embeddedness, and counters these with contemporary problematization of the term as it is applied by neoliberal agendas. This is followed by a discussion of the ways some communities have been 'emptied out' and then 're placed' in the last half century.

Tying Down Communities

In some of its traditional configurations, community en genders notions of spatial contiguity as well as a sense of nested belonging. Not all contemporary communities arise within spatially contiguous areas: the terms 'com munity' and 'neighborhood' are not always synonymous. Although its modern and postmodern configurations may elaborate communities without propinquity on the Internet and through mobile phone conversations, there are nonetheless important relational and geographic practicalities to these configurations. There is, however, a certain allure associated with the desire for groups of people to live in local arrangements of personal ac quaintance, mutual aid, and governance. The continu ation of this allure raises important geographic questions, not least of which is the fomenting of an exclusionary sense of place that prohibits access to certain people based on income, ethnicity, age, and so forth.

In that it is often sought after and chosen, community garners significant commodity power. This is perhaps best seen with the rise of so called neotraditional and gated residential communities at the end of last century in a variety of locations around the globe. In many places, these communities by design evoke person centered liv ing and are advertised as immune from urban violence. A penchant for security guards and surveillance systems is hidden in a design that appropriates notions of village life around public space and recreational centers in settings that foster friendly and caring relations between families, and across social and economic classes.

Of further importance politically is the generally held conviction that a community can demand sacrifice of people's time, money, sweat, and blood. At its more in nocuousness, this may entail financial sacrifice to buy into the ideal family oriented community that fulfills desires for social wholeness and security, as well as good schools. Alternatively, undue self sacrifice may be demanded of residents of communities where spatial injustices such as the placement of noxious industries overwhelm the ability of local groups to address those problems. At the deepest level of commitment, political communities in war ravaged parts of the world may suggest an ultimate sacrifice from young suicide bombers.

The notion of individual self sacrifice raises important issues around community expectations and state inter vention. Neoliberalism is a term now commonly used to describe particular forms of economic and political re lations that favor self governance and a free market ethic applied to all sectors of society. It sets up individuals, communities, and cities – as well as corporations – as self seeking, responsible, economic agents that ensure their own means of survival and improvement. The notion of community in neoliberal parlance suggests increased mutual assistance and self sacrifice. In an essay entitled 'The trapdoor of community' in the *Annals of the Association of American Geographers*, Steve Herbert suggests that within contemporary neoliberal agendas local communities stand as recipients of devolved authority, and they may legitimate that devolution although, for the most part, and particularly in lower income areas, they cannot fulfill the obligations the state seeks to offload.

Imagining Community as Social and Spatial Organization

In his celebrated essay on 'Community and society', Ferdinand Tönnies set the stage for twentieth century thinking about community with his distinction between modern society (*gesellschaft*) and premodern communities (*gemeinschaft*). This dualism informed the twentieth century modernist debate on what constitutes community in terms of social and spatial relations, influencing the thinking of famous social theorists, such as Emile Durkheim, Max Weber, and Georg Simmel. The work of these theorists provides a foundation for contemporary philosophical debates between individualists, liberalists, and communitarianists as well as setting academic foundations for our understanding of urban social and spatial relations.

Gemeinschaft suggests that people are united by deep, horizontal social relations through familial and kinship networks. Place is valued as a primary loci for these networks and limited spatial mobility engenders intensive face to face relations. Place is seen as a circumscribed space within which people interact frequently in highly personalized ways. *Gesellschaft* does not include notions of place based community and, in spite of uniting factors such as spatial proximity and mutual needs, people live separate, individuated lives. Social relations are based upon efficiency and the contractual obligations of capitalist society. Individuals choose community in the form of special interest groups. Although people in *gesellschaft* society may live together in spatial proximate ways, these are limited liability communities wherein members share only a physical landscape and perhaps strive for similar material goals, such as the maintenance of property values.

Weber suggests that *gesellschaft* arises from so called rational capitalist forces, which creates communities for the market where none were previously known. Weber points out that these new communities are increasingly managed bureaucratically, which 'dehumanizes' the emotional context of their existence. More darkly, Durkheim's famous notion of *anomie* refers to breakdown of social norms in modern urban society to the extent that these norms (such as community based kinship ties) cannot be contextualized by bureaucracy. He argues that individuals cannot find their place in society without clear rules, and changing conditions through modernization as well as adjustment to urban life leads to conflict, deviance, and social pathology. The influence of Durkheim and Weber led Simmel, in his *Metropolis and Mental Life*, to postulate that a huge burden of urban society was its focus on organization and rationality over emotion and irrationality, and homogenization (through, among other things, bureaucracy), over differentiation and heterogeneity.

Gemeinschaft thinking translates lightly into modern notions of communitarianism, which began in the late twentieth century. Communitarians claim that values and beliefs exist in public space, and take form from debates that take place publicly. They argue that becoming an individual means taking a stance on the issues that circulate in public space. Alternatively, individualism – loosely arising out of *gesellschaft* – describes a moral, political, and social outlook that stresses human independence and the importance of self reliance. Classical liberalism also supports individual rights, and extends this to *laissez faire* economic policies. Many aspects of classical liberalism are reworked in contemporary neoliberalism, which joins with communitarianism to counter what is seen as individualism's and liberalism's emphasis on rights based philosophies over the importance of communal associations in creating morally responsible citizens (note the rise of governmental support for so called faith based social initiatives and ideological pronouncements such as 'it takes a village to make a child'). As Herbert points out "the neoliberal move to devolve state obligations resonates with long standing desires to make communities effective sites of moral development and political activity." As pointed out later, local places are increasingly burdened with this moral responsibility at a time when their power (and place) is emptying out.

Simmel's observations on the fleeting, intense, and diverse nature of modern social and spatial relations presages, to a degree, contemporary globalization, mass air travel, and Internet based virtual communities. Some commentators on the rise of virtual communities as evidenced through websites such as MySpace suggest that digital technologies give rise to a *gesellschaft* utopianism, precisely because 'no place' is 'every place'. In an important twist to a problematic dualism, Kevin Robbins probes the reasons for the development of these kinds of communities, and questions the nature of a desire that constantly seeks satisfaction through the 'other' of any place, which cannot be satisfied by a material place. He suggests that in a move from the modern to the postmodern there is an increased depthlessness to social and spatial relations, which sandwiches individual need for community between the development of war technologies and a growing consumerist needs for inventions like reality television. The latter, Robbins argues, develops as a consequence of need for material connection that is corroded by marketing 'presence' (here, now, and immediate). Material presence (not only kinship ties but also violence and terror) is denied by the presentation of images that can be engaged viscerally but do not require reason, analysis, or reflection. For Robbins, the postmodern moment is a precise tension between disengagement and withdrawal on the one hand, and a search for belonging and community on the other.

The Emptying Out of Place-Based Communities

Like Simmel, Durkheim developed Tönnies' concepts of community and society in terms of changes in social and spatial organization. In his hugely influential *The Division of Labor in Society*, Durkheim emphasized a progressive division of labor. His notion of organic solidarity as a form of social cohesion grows with increased division of labor in the workplace. Organic solidarity is based upon differences and is distinct from mechanical solidarity that is rooted in families and neighborhoods. Durkheim believed that the loss of mechanical solidarity with modernization is compensated with increases in organic solidarity. In later work, he traces a change from embedded community ties in premodern times to a spatial form of community characterized by what has been termed an 'emptying out' of places and their replacement with the formation of new political associations/communities. These new communities are an important move away from earlier romanticization of *gemeinschaft*, suggesting, for example, new forms of solidarity on the factory floor. Although problematically dualistic, the work of Durkheim suggests a breakdown of categories that presage Henri Lefebvre's project on *The Production of Space*. With modernization, Lefebvre argues, we lost the categories that not only comprise everyday ways of apprehending 'commonplaces' such as town and community but also common concepts such as history, paternity, and traditional morality. Durkheim's *anomie* describes the rootlessness in contemporary social life and then his later work suggests that the loss of embedded communities may be replaced with a new kind of workplace solidarity. For Lefebvre, 80 years later, all that is left with the disintegration of the material conditions of social life are floating, hollow, and abstract metaphors and myths. Robbins post structural perspective on community challenges the hollowness of these myths, but there is still a concern that although contemporary communities may not be hollow, they are placeless.

Placing Communities

Although the twentieth century witnessed the creation of modern communities by design, it was the less tangible ideas of planners and social activists such as Herbert Gans, Jane Jacobs, and Kevin Lynch that influenced debates on what constitutes community. Lynch argued that most important aspect of communities was how they were imagined by those who lived and worked within their borders. His work was primarily at the urban scale but influenced a generation of planners who were disenchanted with efficient, optimized, and people less planning practices. At a smaller scale, Jacobs recognized the virtues of locally based community ties as an important source of satisfaction with day to day life. In particular, Jacobs focused on the functioning of localities around key people such as store owners, priests, local politicians, as well as elderly residents who provide 'eyes on the street' because they are around so much.

Herbert Gans popularized the term 'urban villages' to celebrate the intense social cohesion he witnessed in inner city Boston. This cohesion, he averred, is based upon varied cultural pluralisms that revolve around residents' varied ethnicities, kinships, occupations, and lifestyles. He described tight territorial based local kinship networks in poor communities that moved across ethnic and lifestyle lines. When Gans extended his work to peripheral suburban communities with his study of Levittown, he encountered a much different style of community. Gans concludes that people in Levittown were unable to reconcile cultural pluralisms because they comprised young families that are inward looking and unable to establish a meaningful relationship between family and community. In an essay on 'The myth of community studies', British sociologist Margaret Stacey was one of the first to capitalize on what she saw as the universality of Gans' work. In *Tradition and Change*, using Banbury, England, as an example she showed how local patterns of mutual aid and community are disappearing as young families become more affluent and need no longer rely on each other. Countering the work of Simmel and Webber, she nonetheless charged that local community systems, where they existed, were viable cultural products. Her modernist detractors labeled Stacey a romantic with *gemeinschaft* utopianism for older and mythical community living.

Nonetheless, the dream of suburban community living is still pervasive in Western culture. In a paper in the *Annals of the Association of American Geographers*, Michael Dear and Steven Flusty coined the term 'dreamscapes' to describe the community based ramifications for middle class consumers of the postmodern esthetic. Described as places for concentrating 'like activities' and 'like people' in a society of increased interest and commodity specialization; peripheral 'dreamscapes'; and protected, gentrified, downtown communities create places for affluent groups who can afford to secure their investments against outside influences. They display characteristic forms of architecture, landscaping insurance, surveillance, and security. Part of comfortable middle class living are 'third places' that are neither home nor workplace, but somewhere in between that provide another context of community. Third places are places to meet, socialize, relax, work away from the office and places to eat and drink without pressure to consume or move on. The third place is epitomized by the modern coffee shop, with its sofas and newspapers – a revival of its eighteenth century role in community – or by the Internet cafe.

Community as an Impediment to Social and Spatial Justice

Iris Marion Young's *Justice and the Politics of Difference* brought to the attention of geographers the ways that community as a political category reifies difference. She pointed out that although contemporary society is discursively committed to equality, at the ideological level injustices to those categorized as 'other' are veiled in everyday habits and cultural meanings that are embedded in the standard practices of local politics and planning. Sidestepping patterns, Young focuses on variable vectors of power and the oppressions they produce. She takes explicit axes of oppression such as racism, sexism, ageism, and homophobia and reformulates them into five broader modalities of power that involve issues of justice beyond distribution; these she labels exploitation, marginalization, powerlessness, cultural imperialism, and violence. By so doing, Young is attempting to move beyond neat categories of difference and singular alignments between oppressed and oppressor. By pluralizing the category of oppression in this way, Young opens methodological possibilities that avoid the exclusive and oversimplifying effects of denying difference within groups or leaving out important ways that groups are oppressed. It addresses the problem, for example, that feminists and antiracist theorists have with Marxism's reduction of all oppressions to class oppression that leaves out the very specific oppressions of Blacks and women.

Young's argument is also about embodiment and community. By highlighting spaces of local activities as community, identity becomes conflated with a scale of the body that denies difference. The scaling of bodies (what is seen and represented) within community and local activism is criticized by Young because it expresses a desire for the fusion of subjects (such as in dreamscape space and third places) rather than honoring situated selves. The scaling of bodies in so called neotraditional and lifestyle communities reifies racism, sexism, ableism, and ageism. The problem with the search for locally based community is that it is often based on a desire for unity and wholeness, which, in turn, generates borders and exclusions. This desire denies difference in two ways: first, it does not allow a recognition of difference between and within subjects and, second, in privileging face to face relations, it establishes a model of social relations that are not mediated by space and time. Suggesting that the only authentic social relations are in face to face communities detemporalizes the process of social change into a static, before and after structure much like that proposed by Tönnies. A mythic and romanticized *gemeinschaft* with nurturing face to face relations is set in opposition to the present *gesellschaft* of alienation and *anomie*. This is wholly utopian because it fails to recognize that social relations in both *gemeinschaft* and *gesellschaft* are mediated in space and through time. A nurturing notion of community denies difference because it denies the contradictory and ambiguous nature of social and spatial life.

The communitarian model is potentially conservative because its basis in social constructivism collapses to radical relativism if all political positions (and all communities) have the equal status suggested by a neoliberal agenda. Some feminists argue that a communitarian elaboration of community and the so called unitary (or Habermasian) public are problematic in this sense because they suggest a monolithic public sphere to which all communities have access. Rather, they argue 'for various publics' that deny a homogenized, mechanistic, and institutional space for democracy and justice wherein communication and consensus may evolve, because such a space denies the practical implications of a social and hierarchical construction of scale that makes access from one scale to another or, alternatively, from a weak to a strong public, difficult. The intractable issue that is broached here highlights the ways political acumen pivots on how local community issues become politically charged at other scales so that they cannot be dismissed as something akin to public housekeeping. Two points are important here: scale as conceptualized here is political, not Cartesian; the local community may not be the most appropriate place for devolved politics.

Durkheim points out that the scale of larger cities means more diversity. This not only may produce *anomie* but it also means that diversity has more of a chance of being politically charged rather than being suppressed into communal norms or forced underground. Young believes with Gans and Jacobs that larger cities are not oppressive but, rather, they liberate people from conformist pressures. The uncontrived scale of the city celebrates difference because it is about strangers being together and it is about the creation, through the recognition of contradiction and ambiguity, of social justice through the recognition of diversity. The geographic form of Young's social justice challenges the democratic and neoliberal models of small, decentralized communities for urban decision making. Instead, she prefers as form of social justice at a larger scale, in regional units with mechanisms for incorporating and representing smaller communities. Her regional solution elaborates an important form of mobility through "…diverse and unequal neighborhoods, towns and cities, whose residents move in and out of one another's locales and interact in complex webs of exchange, only a sovereign authority whose jurisdiction includes them all can mediate their relations justly."

Reflexive Communities

Young's regional solution is important in the sense that she produces scale as a deliberate political act. In taking the scale issue out of geography, some have suggested the need for a 'flat ontology' that focuses on social and spatial relations rather than scalar hierarchies. In this sense, the notion of communities nested within cities or societies loses its power. Implicit within this redirection of a previously given geography is the loss of traditional divisions between neighborhoods, towns, cities, and regions. The importance of this to the geography of community is that it destabilizes current neoliberal rhetoric that assumes the fiscal and legal autonomy of 'community development block grants', 'neighborhood watch' programs, and associated measures of 'community policing', and so forth. Neoliberal support for autonomous local communities ties into their ability (1) to evaluate and act upon policies based upon day to day experience, and (2) to act collectively in a noncoerced way. Herbert counters this empirically, arguing that citizens for the most part do not see their community as a communal mechanisms for political action. Rather, residents across economic classes want to "locate themselves firmly in a known and supportive locale wherein they can understand their place and minimize their vulnerability to the unexpected and undesirable."

In his rendering of *Imagined Communities*, Benedict Anderson famously states that "communities are to be distinguished, not by their falsity/genuineness, but by the style in which they are imagined," and this imagination is sufficient to build nations and statehood. The chimera of how we imagine community politically is precisely its usefulness for geographers. However vague, the notion of community elaborates generally nurturing meanings, and is almost always something people desire. It is the tension between imagination, desire, and political manipulation that sets the most important stage for geographic inquiry.

See also: Belonging; Difference/Politics of Difference; Identity Politics; Sense of Place.

Further Reading

Anderson, B. (1991). *Imagined Communities*. London: Verso.
Aitken, S. C. (1998). *Family Fantasies and Community Space*. Princeton, NJ: Rutgers University Press.
Dear, M. and Flusty, S. (1998). Postmodern Urbanism. *Annals of the Association of American Geographers* 88(1), 50 72.
Fincher, R. and Jacobs, J. M. (1998). *Cities of Difference*. New York: The Guilford Press.
Gans, H. (1962). *The Urban Villagers*. New York: Free Press.
Gans, H. (1967). *The Levittowners*. New York: Vintage Books.
Herbert, S. (2005). The Trapdoor of Community. *Annals of the Association of American Geographers* 95(4), 850 865.
Jacobs, J. (1961). *The Death and Life of Great American Cities*. New York: Random House.
LeFebvre, H. (1991). *The Production of Space*. Oxford: Blackwell.
Lynch, K. (1970). *The Image of the City*. Boston, MA: MIT Press.
Robbins, K. (1996). *Into the Image*. London: Routledge.
Young, I. M. (1990). *Justice and the Politics of Difference*. Princeton, NJ: Princeton University Press.

Competitiveness

N. Lewis, The University of Auckland, Auckland, New Zealand

Glossary

Circults of Capital Marxian concept referring to the expansion of value through the generation of surplus value. It highlights the circulation inherent in the production of value, from its generation in the labor process, its transformation into commodities and their consumption, the reproduction of labor and realization of surplus value, and their reinvestment in the labor process as either labor power (mediated by social institutions) or the technical and social organization of work (surplus value mediated by capital markets). In geography, David Harvey has traced and theorized the spatialities of these circuits and their materialization as uneven development.

Global Shift The title of Peter Dicken's highly influential text examining the rise of transnational corporations and the remaking of the global economy. Reprinted in five editions, it traces the growth, extensification, increasing connectivity, and reorganization of economic exchanges related to the emergence of global production networks.

Neoclassical Economics Mainstream approach in economics to understanding decision-making and the operation of the economy. It focuses attention on the performance of rational, maximizing economic agents in markets that determine prices, outputs, and income distributions. The economics of neoliberalism, it presumes that markets will mobilize self-interest to deliver optimum social outcomes. On the supply side, cost-constrained, fully informed, profit-maximizing firms will compete equally. The most productive and efficient will win out.

Introduction

Although rarely confronted directly, competitiveness is a central concept in economic geography. In location and regional development theories, and in their work on industrial restructuring, uneven development, industrial district formation, learning regions, and neoliberalization, economic geographers have asked what gives particular advantages to enterprises located in particular places. Far from unified in their approaches to this question, different authors have derived many and varied answers. Sources of competitiveness have been identified in access to resources or markets, labor qualities, agglomeration economies, transactions costs, firm or commodity chain organization, social institutions, and government policy and spending. Geographers have studied the spatialities of social relations of production, production networks, path dependencies, circuits of capital, political processes, and investment cycles. Others have instead subverted the language of competitiveness to promote alternatives to a capitalist space economy.

Economic geographers, then, have in many different ways sought the geographical source of competitiveness. This article traces the history of this engagement through to the emergence in recent years of relational approaches that reject the quest for a spatial essence of competitiveness. Relational approaches deploy less rigid frames of analysis for identifying and studying the complex of agents, calculations, imaginaries, rationalities, and acts of investment that define geographies of competitiveness. They suggest we study the spatial organization of economic agency, the spatial strategies of agents, and the constitution of spatial advantage across the boundaries of firms and in constituted spatialities such as production networks, nation brands, communities of practice, diasporic strategies, or Argonaut networks. The article begins by highlighting several contradictions that have characterized geographers' engagements with competitiveness. These are used to sensitize a discussion organized into loosely periodized conjunctures of knowledge production, methodology, and policy interest. The article goes on to welcome the emergence of relational approaches and related concerns with diverse economies as opportunities to rework geographical insights about competitiveness.

Contradictory Geographies of Competitiveness

Geographers' engagements with 'competitiveness' have been characterized by many contradictions in addition to methodological and theoretical differences in approach. Here six such contradictions are identified that frame the way that competitiveness has been understood and examined in the literature.

First, geographers have achieved only limited success in their quest to specify the geographical source of competitiveness – precisely and ironically because geography does matter, in all its relationality and depth. Geography guarantees that competitiveness remains irreducible to a stable, universal, measurable, replicable, or calculable essence.

Second, partly as a result, geographers have confronted competitiveness indirectly. It rarely features in the indexes of geography texts. Even neoclassical economic analyses emphasizing the costs associated with spatial fixity and distance have generally failed to study the sites (balance sheets) and moments (board meetings) of decision making where competitive advantage is actually conceived/calculated and evaluated. Such forensics might unlock the calculations of productivity, profitability, politics, and risk that give meaning to competitiveness in any particular case. Instead, geographers have theorized spatial advantages from either surface level features such as distance from markets or from more deeply contoured topographies of institutional space.

Third, they have also commonly examined competitiveness after the fact, and as equivalent to success or growth. Businesses and regional or national economies that thrive and grow are seen *ex post* to be the most competitive, productive, and efficient. Collapsing success, competitiveness, efficiency, productivity, and profitability to survival are growth risks imputing the roots of competitiveness to spatial logics. The danger of paying insufficient attention to particular investment decisions (perhaps with complementary balance sheet forensics) is that more subtle geographies may be lost.

Fourth, geographers have often worked with politically predefined scales (city, industry, region, or nation) to imagine concepts like regional or national competitiveness and specify objects of analysis. These are often the spaces for which data are routinely collected, but may not be (or may have ceased to be) meaningful objects or subjects of competitiveness. Geographers working with more relational, constituted spatialities (communities of practice, industrial ecologies, clusters, and production networks) have responded to the challenge of identifying meaningful subjects and objects but must struggle to identify who is acting and in response to what impetuses.

Fifth, economic advantage is very often particular, situated, and contingent. Innovation, social capital, political networks, and strategies that configure value chains to exploit geographical difference, each the subject of increasingly sophisticated relational inquiry, might all give short term competitive advantages that have little to do with efficiency or productivity as understood in neoclassical accounts. Indeed, in practice, they may manifest in or emerge from a capacity to forestall or hold off competition. What is seen as competitiveness is often that which is, in neoclassical economics, its other (monopoly) or the result of its antithesis (cooperation).

Finally, treatments of competitiveness are neither built nor deployed in a political vacuum. Neo Marxists have long criticized the capitalist relations of production subsumed within notions of competitiveness. Erica Schoenberger adds that the term itself performs as a powerful moral economy – in economic geography and in wider politics. Collapsing success, efficiency, competitiveness, and productivity, disguise the exploitative practices of capital and naturalize the brutal spatial politics of accumulation. The super profits and sweatshops of transnational capital (TNC) and the redundancies of industrial restructuring are neither virtuous nor necessary. Borrowing Gibson Graham's terms, discourses of competitiveness 'other' cooperation and more diverse economic practices with different moral economies. They may foreclose on possibilities of thinking otherwise about the economy.

From Industrial Geography to Global Shifts

Early 'industrial geography' aimed to explain patterns in the landscape, as did much mid twentieth century geography. Generations of school children learned of the relationships between key resources such as coal and steel and the development of the industrial heartlands of the United States, Britain, and Europe. Geography was revealed to matter (location gave competitive advantages), even if space was largely assumed to vary only in terms of distance, transportation costs, the location of resources, and markets. Locational rents could be derived from either proximity to resources or markets and plotted out in different distance decay and bid rent curves for different commodities to map zonal patterns of land use. Geographers using quantitative techniques tested spatial patterns against the theories of Weber, Losch, Christaller, Von Thunen, and Alonso.

This spatial analytic phase of economic geography unfolded into greater interest in the actual location decisions of economic agents on the one hand, and a concern with structural conditions on the other. In mainstream approaches, hypothetical analyses gave way to empirical analysis of location. Although competitiveness was again studied indirectly via location, analysis of pattern became replaced with a greater emphasis on process. Contemporaneously, concern with path dependent or structural conditioning of process became registered in debates over whether development trickled down from growth poles or processes of circular and cumulative causation reproduced uneven development. These debates tied emerging concern with the fates of regions to the interests of geographers in Third World development.

Geographers came to address notions of regional and national competitiveness via analysis of uneven development and the emerging evidence of decline in key industrial regions. While those influenced by neoclassical economics continued to work on firm location, critical geographers turned to dependency and world systems accounts of the structural conditions of competitiveness, and their own emerging theories of uneven development. They asked questions about regional or national competitiveness, multinational corporations, industrial regions, and commodity chains. Armed also with Harvey's insights, their commitment to grounded investigation, and sensitivity to questions of 'where', geographers became pivotal critics of the emerging shape of what McMichael later called the 'globalization project'.

The 1980s were a high point in economic geography, or at least the critique of the spatialities of industrial restructuring, regional decline, and related questions of competitiveness. Harvey, Dicken, Schoenberger, Doreen Massey, Richard Walker, Michael Storper, Allen Scott, Neil Smith, and many others brought critical lenses to bear on the restructuring processes and the new spatial divisions of labor emerging from the collapse of Fordism and the spatial fix of post war Anglo American growth. With a Marxian sensitivity to capital–labor relations, Massey in particular deployed the language of spatial divisions of labor to place the labor relation and its geographies at the center of theories of competitiveness and social transformation. The move gave economic geography its critical mainstream.

With labor costs and conditions at the center of explanation, as Jamie Peck later argued, the (necessary) community based reproduction of labor ties places tightly into this equation. With labor tied to place, new technologies of production and organization were seen to free capital from it and to facilitate more flexible and dispersed accumulation. Spatial restructuring enhanced competitiveness and generated new spatial divisions of labor. Of course, as Harvey and Smith both demonstrated, capital is far from fully mobile, different capitals are differentially mobile, and labor markets are always differentially mediated by social, political, and cultural institutions. Logics of spatially uneven development inhere in capitalism. Investment continues to tie capital to place, at least temporarily, and place will always be a factor in the competitive advantages sought and experienced by capital. Decisions are made in complex and calculative matrices of competitiveness. Explanations assuming factor mobility and level playing fields had little hope of unlocking the secrets of geographies of competitiveness.

Armed with Harvey's theorization of the spatiality of circuits of capital and geographical uneven development, economic geographers came early to reframe concerns with competitiveness as a political economy critique not just of capitalism but of development and globalization. Industrial decline and social dislocation in the former industrial heartlands of the West highlighted the geographical responses of increasingly concentrated and TNC to crises of competitiveness. Facilitated by regulatory changes, new technologies of production, finance, and corporate organization were being worked into a spatial fix to the crises. In the work led most notably by Peter Dicken in the language of 'global shift', studies of competitiveness became subsumed within efforts to capture and account for the new globalized terrain of capital concentration, TNCs and their outsourcing networks. Competitive advantage became to be seen as a function of the economies of scale and spatial dynamism of these corporations/networks in relation to labor costs and regulation of production (particularly around the labor relation).

'Global shift' was a dominant narrative, but could not be a singular story. It began to morph at its center into concerns with flexible manufacturing, commodity chain analysis, and production networks that linked TNCs to multiple other firms and organizational forms. These cut across and remade established and hierarchical scales of political and economic agency and analysis. Although clearly linked to other economic domains 'global shift' was (is) a story of large firms and their contracted partners in predominantly manufacturing industries – a narrative of 'industrial geography'. Still dominant in geographical imaginations (appropriately, given the attention it directs to organization, power, and monopoly as sources of competitiveness), it is only one of several contemporary geographical narratives of competitiveness.

Industrial Districts and the New Regionalism

A chronologically parallel school of thought to 'global shift' built a new geography of competitiveness on the basis of ideas of flexibility and networked relations, but in relation to closely integrated small firm complexes and industrial district formation. Drawing on the experiences of the 'Third Italy' rather than 'global shift', other geographers focused on the achievement of flexible accumulation and competitiveness through social institutions.

By the early 1990s, a literature on industrial districts was theorizing away from the experiences of the Third Italy and Orange County. It connected directly with Michael Porter's highly influential ideas of competitive advantage and industrial clusters in business studies. In Porter's account, value chain linkages and complementarities tended to generate clusters of firms that provided competitive advantages for new entrants through access to immediately proximate interfirm relations. Technology and globalization had not overcome colocation which offered advantages associated with innovation, knowledge dissemination, and trust (replacing resource and market access). Industrial districts were explicable, virtuous, and could be encouraged through policy. Over the 1990s, theory began to merge with policy in a growing literature on clusters, their behavioral underpinnings, formation, durability, and policy relevance. Actual/emergent clusters became indistinguishable from imagined or planned clusters.

In a similar coupling with policy interest, institutional explanations of competitiveness were wedded to emerging theories of social capital. In the geographical literature this 'new regionalism' identified 'institutional thickness' as a localized source of competitive advantage. More and deeper social institutions produced greater trust, enhanced business connections, reduced transactions costs, promoted seamless flows of information, and stimulated collective learning environments. Although accused of failing to address wider production networks and fetishizing social institutions and place, the new regionalism contained critical elements. Its richer set of accounts of the relationships between place and competitiveness and their formation offered a critique of the Porter school, and provided a counter to the dominant 'global shift' narrative in scholarship. It accented co operative sources of firm advantage by contrast with competitiveness.

In policy spheres, the new regionalism gave regional authorities an alternative to policies seeking to attract investment via tax breaks, regulatory exemptions, and de unionized and cheap local labor pools – those classic strategies of the race to the bottom. Encouraged by notions of institutional thickness, policymakers began to promote the new panaceas of social capital and good governance as a basis for new industrial policies and fixes to the failures of neoliberal development. This conjuncture of policy and theory became enhanced and extended in the creative cities discourse of the early 2000s. Stimulated by Richard Florida's arguments about the creative classes, certain cities became held to have a competitive advantage in creative industries generated by their tolerance of difference, cosmopolitan vibrancy, civic amenities, and stimulating built forms. These ideas connect to new regionalist ideas of innovative *milieux*. Positively, they suggest alternative policy targets for competitive city strategies such as enhancing social institutions and tolerance. On the other hand, their foundational understandings of investment processes and competitiveness have been challenged by critics such as Peck.

Postindustrial Spaces

Other economic geographies have linked evidence of the competitive advantage of spatial clusters with the 'global shifts' narrative in their work on 'postindustrial' industries. Gordon Clark has highlighted both the advantages of spatial clusters and globalizing financial circuits of capital and associated concentration of capital in the financial services industry. Colocation in apparently footloose activities still generates competitive advantages. These advantages mirrored those found in manufacturing – the significance of proximity and social institutions in facilitating knowledge transfer, learning communities, and the building of communities of practice across firm boundaries; and the hard edged locational decision making and investment processes of spinningoff and positioning within specialized labor markets.

Geographers such as Allen Scott and Dominic Power also began to study cultural industries such as music, fashion, film, as well as other newer high tech creative industries. They again found distinct patterns of clustering in particular cities and locales within them related to social institutions, knowledge transfer, and interfirm connections. Geographers working with 'cultural economy', such as Louise Crewe and Wendy Larner, have pointed to the persistence of particular locales with value adding, culturally ingrained associations between place and activity embedded in historical trajectories and traditions. These associations give the areas and the firms within them particular cultural and political advantages and confirm that cultural features of place matter to economic subjects – consumers, investors, and workers (particularly, the highly skilled whose particular qualities, in the language of classical economics, generate economic rents). Where investors are also consumers and highly skilled workers become investors in spinoffs, these cultural features become instrumental in geographies of competitiveness. Although criticized for a narrow focus on 'sexy' industries and for decentering investment processes, these writers have directed attention to consumption and identity formation in circuits of capital.

In the policy language of the creative city inspired in part by Florida, these ideas are in turn linked to the notion of cultural vitality as a place based resource. New ideas are emerging about competitive cities and the 'creative industries'. Although understating the significance of particular urban fabric and wider economic structures in generating these cultural hot spots, the creative city thesis adds new complexity to the ideas of competitiveness.

Social Networks Facilitating 'Global Shifts'

Not all the stories of successful regions fall neatly into either the 'global shift' narrative or the largely European, new regionalism narrative. Regional success stories in the Asia Pacific have focused attention on cultural institutions and the way that these lubricate the transnational networks of exchange configuring commodity circuits. Anna Lee Saxenian's work on the new Argonauts and transnational business relations within social networks between Taiwan and the USA and Henry Yeung's later work on business networks within East Asia point to the way that cultural and social institutions become articulated to global production networks. In so doing, they give certain regions and economic agents competitive advantages, and/or facilitate arrangements not captured by established theories of competitiveness. This work has tracked the operation of circuits of capital (investment, trade, and commonly labor relations and flows) within diasporic communities framed by family and other culturally mediated relationships. Saxenian and Yeung are quick to distance their work from culturally essentialist accounts of 'guanxi' capitalism and simplistic notions of a cultural competitive advantage. Rather, they offer other bridges between 'global shift' narratives and the new regionalism that also connects with the work of historical economic geographers such as Gordon Winder who highlight global networks (often family centered) prior to the 'global shifts' of the mid late twentieth century.

While stopping short of understanding these networks as co constituted and emergent assemblages in post structural terms, this body of work takes the literature forward in at least three directions. It directs attention to business networks and to the interplay of cultural and economic relations in their establishment. There are different ways to generate competitive advantages through business organization that are shadowed in 'global shift'. Cultural knowledge, institutions, and experiences are differentially endowed assets (individual and collective), and agents extract competitive advantages by organizing their business to exploit them and their spatialities. The organizational charts of TNCs appear aspatial by contrast. Second, it reminds us that social institutions are always important, but need to be explored in articulation to circuits of capital that stretch beyond particular regions, product complexes, and cultural framings. With Indian diasporic communities now mobilized, the world is being influenced by another round of 'global shifts' that are built upon and seek to realize family and other cultural 'who you know' and 'how you know' advantages with geographical roots. Finally, the work confirms that old understandings of periphery and center have crumbled amid the emergence of Argonauts, TNCs, diaspora strategies, NICs, and the like.

Learning and the Sociality of Competitiveness

The new regionalism directed attention away from input costs (including labor) to the social organization (families and social networks) of circuits of production and the porous boundaries of the firm. Initial interest in cultural factors and the operation of social institutions in lubricating and organizing investment and production has given way to harder and more specific interests in innovation, learning, and knowledge transfer. The harder, commonly quantitative, edge of new regionalism sought to unpack in more mainstream economic language the competitive advantage afforded by social institutions, or what Gertler has called the 'undefinable tacitness of being (there)'. This literature asks how exactly institutional thickness translates into competitive advantages, and how these effects might be measured. Efforts to produce typologies of trust shifted to concerns with thickness of economic exchanges, knowledge dissemination, local labor market movements, and necessary conditions for start ups. More recently and more relationally, concerns have shifted to questions of learning, tacit knowledge, and the value of face to face meetings – increasingly theorized as 'buzz'.

Much of this work has been conducted in new, so called knowledge intensive industries in 'learning regions' or innovative *milieux*. The literature has placed knowledge, its production, and its exchange at the center of competitiveness. It has gone beyond identifying knowledge asymmetries as a source of competitive advantage, to emphasize that knowledge is experiential, cognitive, temporary, contingent, and contextual, and therefore to some extent tacit and borne of social context. If the interest is in the spatialty of that advantage, it must lie in how tacit knowledge becomes produced, appropriated, valued, and shared. The literature points out that this has much to do with social context, proximity (material, virtual, and social), social and localized professional institutions, and social learning processes. These processes exhibit a powerful spatiality, but one that need not be simply territorialized or scaled. Rather, it is co constituted in and with communities of practice which may have complex spatialities.

Political Economy and Political Agency

In all these narratives, competitiveness is at some level political. Geographies of competitiveness are constituted by and constitutive of political spaces. Economic activity takes place in political space and is in turn highly policied, while investment decisions have both political origins and consequences. Competitiveness owes much to struggles around the labor relation and increasingly the politics of environmental resourcing. States (national and local) alter competitive potential. Development agencies foster local, city, regional, and national competitiveness, while other agencies build social and physical infrastructure. Taxation, labor, welfare, education, cultural, and other policies mediate the constitution of social institutions and economic identities; while all manner of other agents and authorities from industry associations to trades unions, environmental groups, and indigenous peoples seek to advance all manner of political projects, either through state agencies or more directly.

Geographers studying economic activities at all scales routinely consider competitiveness within political economy approaches.

Much recent economic geography has centered on a critique of neoliberalizing globalization. In the initial or 'roll back' phase of neoliberalism, to use Peck and Tickell's terms, governments removed the subsidies, employment policies, price controls, welfare regimes, and cultural policies that configured the competitiveness regimes of Fordism and related social and spatial compacts. Marxian economic geographers traced the changing spatial divisions of labor of 'global shift' to the interplay of the political and the economic in the collapse of the Fordist space economy under the weight of its own contradictions and the ideology of neoliberalism. Neoliberal reformists saw competition as naturally and ideally perfect and space as naturally and ideally neutral. They claimed to be returning the game of capitalism to its natural state – a level playing field on which the most competitive win. In practice and in addition to dismantling welfare regimes, this meant marketizing flows of resources, money, labor, and goods, opening access to cheaper labor in less regulated conditions, and allowing certain fractions of capital greater control over the circuits of capital. Economic geographers, politically hostile toward, and intellectually critical of, such moves, dismissed notions of level playing fields and pointed instead to how neoliberal policy altered the conditions of competitiveness, and redistributed advantage across social groups and space.

Since the late 1990s, economic geographers have examined efforts to rebuild economic institutions at national, regional, and supranational scales, particularly in Europe where the European Union has refashioned regulatory space. Drawing on civic boosterism, the new regionalism, and competition state agendas, national and regional governments across the European Union, localities in the United States, cities in Australia and Pacific East Asia, and national states such as Finland, Singapore, India, and New Zealand, have launched programs to foster innovative *milieux*, learning regions, national systems of innovation, hubs, clusters, nation brands, diasporic investment communities, and creative cities. Tax breaks on research and development, the sponsorship of scientific research, innovation seminars and tours by celebrity academics, industry level market development assistance, immigration policies, and even income support schemes for workers have been deployed to build competitive knowledge economies. In this 'roll out' neoliberalism, states are deploying what Larner and colleagues have called 'after neoliberal' political projects to enhance competitiveness in a new race to facilitate run in parallel to the race to the bottom initiated by neoliberalism. These races frame Thrift's ideas about a defining double moment of capitalist exploitation in which, alongside primitive accumulation strategies, managers are seeking to extract new value from the psychophysical powers of certain privileged human bodies by speeding up the rate of innovation.

While connected by the materialities of 'global shifts', traveling rationalities, and what Thrift and Olds label 'cultural circuits of capital', neoliberal programs of both 'roll back' and 'roll out' varieties have definitive geographies. They have taken different forms at different times and in different places, harnessed to different political projects and experiences of restructuring. In an edited volume by geographers confronting this diversity, England and Ward highlight again the discursive significance of competitiveness. Circulating among business people, government agencies, and experts of various types, commonsense understandings can background questions of efficiency, productivity, and profitability. The multiple actors in such circulations make decisions targeting pools of value that do not necessarily derive from micro calculations of profitability. Despite the globalization of capital, they may have particular temporal horizons, emerge from particular historical trajectories, or be subject to myriad political, cultural, or economic contingencies. Richard Le Heron and Wendy Larner have in response called for a post structural political economy in which assemblages of economic activity are seen to be politically, culturally, and economically co constituted and to transcend prior territorial and analytical framings. Such approaches would signal a more open, partial, relational, and situated understanding of competitiveness, and a greater sensitivity to the temporality, contingencies, and social constitution of the frames through which we imagine and study it.

Diverse Economies

With this in mind, not all economic geographers have approached ideas of competitiveness from the direction of the conventional economy. Some have worked on various relocalizations of agro food chains such as farmers' markets, fair trade movements, and organic or slow foods. These works demonstrate particular, localized, and normatively constructed notions of competitiveness, which can look very different to both the 'global shift' and 'Third Italy' conceptions. Others, such as Andrew Leyshon, have identified alternative conceptions of competitiveness in localized systems of exchange and the persistence of face–face transactions in facilitating capitalist exchanges. Gibson Graham point to the emergence of alternative community economies in Australian settings and remind us of the long history of work on informal or shadowed activities in development contexts in the Global South.

This work focuses on exchanges in localized and marginalized settings, organized in conventions beyond dominant economic models and neoliberal ideologies – albeit in settings articulated to the global capitalist space economy. Authors use the term 'alternative' to make visible forms of exchange that are often marginalized by the dominant circuits of global production networks and supermarkets/chain stores. They also highlight the soci ality of exchange, and different ethics of economic practices, thereby challenging dominant capitalist prac tices and the regulatory order that argues there to be no alternative. They demonstrate that alternatives are pos sible and desirable. Further, as Gibson Graham argue, there are multiple capitalisms. Big 'C' capitalist practices are heterogeneous, ethically bound, and challengeable. By shifting attention away from workplace struggles around the employment relation to marginalized socio economic settings, the work on diverse economies high lights again the political performance of discourses about competitiveness. It challenges the hegemony of 'com petitiveness' as the assumed universal basis of decision making, and challenges economic geographers to think instead of viability and resilience.

Relational Geographies of Competitiveness

To conclude, three observations are made on the narratives outlined above. First, they suggest an uneven stumbling toward more relational understandings of competitiveness across economic geography. This movement is co constituted with a shift away from the study of manufacturing location in the industrial world to other industrial and postindustrial fields, and to more complex and open circuits of capital in a greater range of geographical settings. These shifts are in turn built on and from a collapse of positivist fundamentalism and a heightened sensitivity to the particular and 'messy' ac tualities of competitiveness and decisions made around investment. They also reflect an ongoing critique of the economic geographies written and actualized via policy or pivotal investment actors. Institutions, such as the World Economic Forum's *World Competitiveness Report* and International Institute for Management Development's (IMD) global benchmarking project, or Standard and Poor's ratings, continue not just to practice and facilitate in business studies an older style economic geography of competitiveness, but to actualize it by inscribing it into globalization policy and related circuits of capital. Eco nomic geographers must continue to engage with com petitiveness in many guises.

There is messiness in all of this, however, a promise of more methodologically diverse study of the constitution of each of the moments, sites, and agents of economy. The focus on buzz and learning, for example, is associated with geographers working in quantitative methodological traditions, but seems to me to encompass post structural aspirations and epistemology. It seems to necessitate analysis of the co constitutive relations among actors, institutional structures, material resources, spatio economic imaginaries, and processes of change. The 'buzz' literature takes up the challenge of identifying meaningful objects to study. It suggests an opening up of the political economy of competitiveness to both tighter case study analyses of actual investment decisions and more imaginative understandings of the relational spaces of competitiveness. At the same time, the application of the 'ecology' metaphor to particular configurations of actors and activities promises a new relational approach to competitiveness that recognizes more fully the chal lenges for the reproduction of competitiveness posed by economic exchanges with the environment.

Second, the term 'competitiveness' continues to be deployed politically, whether in 'roll back', 'race to the bottom' policies associated with free trade agreements or in 'roll out' or 'after neoliberal' projects seeking to build clusters of competitiveness at different scales. The spa tiality of competitiveness remains prominent in the for mation of national, regional, and city development in 'roll out' ism. This is particularly apparent in the economic nationalism of development states such as New Zealand, Ireland, Singapore, and Thailand, the learning region policies of larger, advanced industrial nations, the knowledge strategies of Singapore, Dubai, and Bangalore, and creative city projects across the industrial world.

Third, the opportunity to connect new relational understandings of competitiveness to new relational understandings of the production of space and scale may invigorate in Gibson Graham's terms a post capitalist politics of possibility. Such a politics may re imagine connections and help to establish new articulations be tween capitalist economies and the diverse alternative economies that are being imagined on the ground by communities in both the Global North and South. De stabilizing notions of competitiveness in this way might be accompanied by companion projects that examine the co constitution of wealth and consumption and labor and environment. Investigating such possibilities might re invigorate a politics of economic geography eviscerated by neoliberalism. It might allow economic geographers, for example, to relate their critiques of cluster, creative city, or social capital development programs to con ceptions of alternatives. That is, while the initial chal lenge is to recapture alternative understandings of competitiveness from the imagined singularities of global capital, the greater challenge is to identify conditions of possibility for reproducing small economic enterprises, families, and communities.

See also: Capitalism; Cultural Economy; Industrial Districts; Learning Regions; Location Theory;

Neoliberalism; New Regionalism; Social Capital; Uneven Development.

Further Reading

Christophers, B. (2006). Circuits of capital, genealogy, and television geographies. *Antipode* 38, 930 952.

Dicken, P. (2007). *Global Shift: Reshaping the Global Economic Map in the 21st Century* (5th edn.). London: Sage.

England, K. and Ward, K. (eds.) (2007). *Neoliberalization: States, Networks, People*. Malden: Blackwell.

Fagan, R. and Le Heron, R. (1994). Reinterpreting the geography of accumulation: The global shift and local restructuring. *Environment and Planning D: Society and Space* 12, 265 285.

Florida, R. (2002). *The Rise of the Creative Class, Cities and the Creative Class*. New York: Basic Books.

Gertler, M. (2003). Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). *Journal of Economic Geography* 3, 75 99.

Gibson Graham, J. K. (2006). *A Postcapitalist Politics*. Minneapolis, MN: University of Minnesota Press.

Harvey, D. (reissue) (1999). *Limits to Capital*. London: Verso.

International Institute for Management Development (IMD) (2007). *World Competitiveness Yearbook*. Lausanne: IMD Research Centre.

Larner, W. and Le Heron, R. (2002). The spaces and subjects of a globalising economy: A situated exploration of method. *Environment and Planning D: Society and Space* 20, 753 774.

Lewis, N., Larner, W. and Le Heron, R. (forthcoming). New Zealand designer fashion industry: Making industries and co constituting political projects. *Transactions of the Institute of British Geographers.*

Leyshon, A., Lee, R. and Williams, C. (eds.) (2003). *Alternative Economic Spaces*. London: Sage.

Malmberg, A. and Maskell, P. (2006). Localized learning revisited. *Growth and Change* 37(1), 1 18.

Massey, D. (1995). *Spatial Divisions of Labor: Social Structures and the Geography of Production*. New York: Routledge.

McMichael, P. (2004). *Development and Social Change: A Global Perspective* (3rd edn.). Thousand Oaks: Pine Forge Press.

Morgan, K., Marsden, T. and Murdoch, J. (2006). *Worlds of Food: Place, Power, and Provenance in the Food Chain*. Oxford: Oxford University Press.

O'Neill, P. and McGuirk, P. (2002). Prosperity along Australia's eastern seaboard: Sydney and the geopolitics of urban and economic change. *Australian Geographer* 33, 241 261.

Peck, J. (1996). *Work Place: The Social Regulation of Labor Markets*. New York: Guilford.

Peck, J. and Tickell, A. (2002). Neoliberalising space. *Antipode* 34, 381 404.

Power, D. and Scott, A. (2004). *The Cultural Industries and the Production of Culture*. London: Routledge.

Saxenian, A. L. (2006). *The New Argonauts: Regional Advantage in a Global Economy*. Harvard: Harvard University Press.

Schoenberger, E. (1998). Discourse and practice in human geography. *Progress in Human Geography* 22(1), 1 14.

Smith, N. (1990). *Uneven Development: Nature, Capital and the Production of Space* (2nd edn.). Oxford: Basil Blackwell.

Storper, M. and Walker, R. (1989). *The Capitalist Imperative (Territory, Technology and Industrial Growth)*. Oxford: Basil Blackwell.

Thrift, N. (2006). Re inventing invention: New tendencies in capitalist commodification. *Economy and Society* 35, 279 306.

Yeung, H. (2005). Rethinking relational economic geography. *Transactions of the Institute of British Geographers* 30, 37 51.

Yeung, H. (2005). The firm as social networks: An organizational perspective. *Growth and Change* 36, 307 328.

Complementary and Alternative Medicine

G. J. Andrews, McMaster University, Hamilton, ON, Canada

Glossary

Alternative Medicine Forms of medicine often used instead of conventional medicine.

Complementary Medicine Forms of medicine that exist, and are used, alongside conventional medicine.

Conventional Medicine Orthodox biomedicine, ranging from its scientific assumptions and approach to the character of its remedies and services.

Holistic Medicine Forms of medicine that deal with the person's mind and body and seek to investigate and address the wider causes and consequences of diseases as they occur throughout people's daily lives.

Therapist Public or private sector practitioners of complementary and alternative medicine.

Traditional Medicine Forms of medicine that originate in traditional societies and cultures, passed down orally through generations. They are often more commonly found in developing world countries and the Asia-Pacific region.

Introduction

The terms 'alternative medicine' and 'complementary medicine' refer to all therapies and treatments which are not currently part of mainstream conventional medicine. During the 1970s and 1980s, the term alternative medicine was the first universally accepted title. By the 1990s however, the term complementary medicine had gradually become more popular, primarily because it denotes working alongside conventional medicine rather than in opposition to it. This is a particularly important distinction for both people who work in the sector and researchers, many of who do not wish to be thought of as involved in a marginal or hostile field. Because the title alternative medicine has remained the most popular in North America, in order to avoid further confusion and un necessary debate, the all inclusive term 'complementary and alternative medicine' (often abbreviated to and pro nounced CAM) has become preferred in the international literature. Occasionally, the terms 'holistic medicine', 'in tegrative medicine', and 'nonconventional medicine' are also used as universal titles. One could argue that re searchers could select from these different titles to de scribe subtly different engagements between conventional and other medicines. However, though this is sometimes done, it is the search for an acceptable inclusive term for the entire sector that has dominated discussion.

CAM refers to an extremely diverse set of practices, grouped together as much because of their shared dis tinction from conventional medicine as because of their commonalities. Indeed, CAM is not a unified field, but is constituted of a wide range of materials and practices. Twelve core practice modalities of CAM exist: acu puncture, chiropractic, creative and sensory therapies, healing, herbalism, homeopathy, hypnotherapy, massage, reflexology, naturopathy, and osteopathy. When taking into account various subvarieties of these modalities, together with many other less known varieties, it is possible to list many hundreds of distinct forms of CAM. Moreover, the borders of CAM become blurred and its forms much more numerous, as it becomes incorporated into many materials and practices of everyday life. In this respect CAM increasingly permeates the daily lives of westerners ranging from aromatherapy shampoos to yoga workouts at workplace stations.

There are a number of ways of classifying and grouping different CAM modalities. A typical scenario is under four broad categories: 'mind body therapies', 'biological therapies', 'energy therapies', and 'physical therapies'. Such a classification serves the purposes of research where it is necessary to talk collectively yet provide some distinctions (e.g., with regard to education, policy, and regulation). In terms of regulation, certain CAM modalities are fully regulated often by national scale representative organizations. Other modalities are partially regulated and some continue largely un regulated. Much however depends on the particular modality and the national context.

Other important differences between particular mo dalities of CAM relate to their relationships with con ventional medicine. Some modalities are far more integrated with conventional medicine than others. In deed, those that draw closest to the medical model, in terms of their scientific basis and ability to provide an acceptable evidence base, typically have strongest rela tionships and often share spaces for practice (such as osteopathy and chiropractic). However, some researchers argue that debates about 'letting CAM into' conventional medicine and the categorization of CAM in terms of its difference and distance from conventional medicine (often for stated purposes of safety and regulation) serve as an orthodox medical control over CAM.

Research has shown that CAM has grown very rapidly during the last 20 years, and some commentators have called it the most significant health related consumer trend of the twentieth century. Studies frequently identify that over a quarter of the general populations of developed world countries use CAM regularly for a variety of con ditions – chronic, long term, musculoskeletal, and psy chological being the most common categories. A great deal of provision is by the small business private sector, the hundreds of thousands of private therapists worldwide – either working alone or in small group practices – con stituting an extensive global but cottage industry. Increasingly however, CAM is available in conventional healthcare settings, the most common scenario being its provision alongside general/family practice. Some ortho dox health professionals (such as nurses) with interests and expertise in CAM have developed therapies as part of their everyday work routines. This however is often in an unofficial capacity and therefore, on a service level, can be inconsistent. Less typical, yet becoming more common, is for CAM to be provided in hospitals and community settings through dedicated service initiatives. Meanwhile, a number of large companies selling herbal supplements, or with chains of outlets, represent the involvement of much bigger business in CAM. Nevertheless, much depends on the particular national context. North America, for example, is home to much larger CAM businesses than Europe.

The Organization of CAM Research

CAM research has developed into a discipline in its own right. Indeed, reflecting this are a number of dedicated academic journals and regular international conferences (see, e.g., the journals *Complementary Therapies in Clinical Practice* and *Complementary and Alternative Medicine*). Moreover, many scholars now research and teach CAM exclusively and regard themselves primarily as CAM re searchers. It is a discipline constituted of many different perspectives, including biomedical, economic, socio logical, psychological, policy, anthropological, and geographical. However, this diversity also reflects its dis parate institutional character. There are very few dedi cated university departments of CAM (see, e.g., the Department of Complementary Medicine at Exeter University, UK), and consequently the majority of re searchers are located in departments that reflect their 'parent' disciplines.

A small number of CAM research institutes exist and also places that disseminate research findings to the practice community. One example is IN CAM, the Canadian Interdisciplinary Network for Complementary and Alternative Medicine Research, created to foster excellence in CAM research. Specifically, its objectives are to build a sustainable network that facilitates and supports research; promote knowledge transfer among researchers, CAM and conventional practitioners, policymakers, and research funding agencies and – to avoid the duplication of research efforts – link with other CAM networks and institutions. Another example, also in Canada, but with a more regional orientation, is CAMera – Complementary and Alterative Medicine Education and Research Net work of Alberta.

Meanwhile, reflecting the emergence of online CAM resources and research evidence is CAMline, an evidence based website for both health professionals and the public. It offers access to an evidence base on natural health products, therapies, practitioners, and health conditions.

Social Scientific Fact Finding

Mirroring the relative infancy of CAM as a form of healthcare, research is still very much in an initial information accumulation phase. Much research has been in the form of clinical trials, designed to establish at least the beginnings of a scientific evidence base. Initial social science research on CAM during the 1980s, and throughout the 1990s, was also engaged in some basic 'fact finding' about the social nature of production and consumption. For example, rates of use were investigated for both CAM as a whole and between therapies, studies finding various manual therapies (such as massage and reflexology) to the most frequently provided and used. With regard to who uses CAM, studies have found fe males, younger adults, and the more affluent to be the largest groups of users, particularly due to the need to find out of pocket payments for many treatments. However, contrary to popular opinion, studies have also found that, with the exception of children, many sections of society, and all age groups, use CAM regularly. With regard to some trends in use, studies have found users to actively mix and match conventional medicine and CAM and feel empowered in creating their own treatment profiles. They have found people to be frequent users (typically one a week or every 2 weeks), and that they use CAM over long time periods, often for their general health maintenance.

It is also recognized that broader social trends and transitions have underpinned the recent CAM phenom enon. Commentators have found CAM use to be asso ciated with an emerging health consumer culture, and a greater number of people – often referred to as 'smart consumers' – who inform themselves about health and illness and exercise this knowledge when selecting from the treatment options available to them, or otherwise seek alternatives. Some commentators have linked CAM to spiritual cultures that have developed alternative philosophical visions toward health. However, in re sponse, others have been quick to stress that, while these cultures might have been common 20 years ago, CAM is now far more mainstream in terms of its users and their motivations. More practically, regarding the immediate reasons for using CAM, researchers have identified push factors from conventional medicine, including perceptions of its impersonal nature, a lack of caring, long waiting lists, and its inability of to cure long term chronic conditions. Meanwhile, identified pull factors of CAM include the longer time and dedicated attention that therapists are able to give to clients, and the individualized nature of their care.

Of research focused squarely on CAM provision, a great deal discusses technical issues relating to practice development. Recent work, however, is focused much more broadly on features and contexts of production. For example, the career pathways of therapists have been investigated – studies finding therapists to have entered the CAM sector from other caring professions (e.g., nursing, social work, and teaching). A range of ethical and moral issues in practice have also been discussed. These have included when and whether therapists refer clients to conventional practitioners, when and whether they discontinue treatments, and what they do if a client can no longer afford to pay for their services. Businesses, business planning, and career aspirations have been investigated – studies finding these to be typically short term, low level, and 'realistic'. Other research has considered regulation and professionalization, particularly in the context of current policy debates on how CAM should be provided in and integrated with conventional medicine. One of the main debates here has been agreeing on acceptable methods to collect evidence for practice. Some therapists claim that because their therapies do not work on scientific principles, they should not be subjected to clinical trials. Meanwhile others argue that the often subtle health gains obtained through using CAM are not adequately captured by clinical trials. However, most doctors, policymakers, and researchers strongly disagree and suggest that clinical trials are both appropriate and necessary in order for conventional healthcare systems to fund CAM, and for their clinicians to take CAM seriously.

The Origins of a Geographical Perspective

Mirroring the growing popularity of CAM during the past 20 years, and its connections to many aspects of social and cultural life, like other social scientists, human geographers are becoming increasingly interested in it as a subject of study. However, just as the roots of many forms of CAM are located in traditional medicine – often from the developing world and East Asia – so are the roots of geographical research on CAM. Traditional medicine typically refers to treatments associated with ancient societies, passed down through their generations, often in the form of oral knowledge. It is the methods, beliefs, and approaches that people employed to restore health before the arrival of conventional biomedicine. Examples include Pinyin (China), bone setters (worldwide), ayurvedic medicine (India), and Kampoh medicinal preparation (China).

The study of traditional medicine was an important interest of early medical geography. Charles Good's 1977 paper outlined an agenda that heavily influenced and directed research. Good pointed out the substantial health problems facing Africa, Asia, and Latin America, the lack of conventional services in many areas but, at the same time, the widespread availability and use of traditional medicine. He noted the vast manpower, resources, and access to local populations available to conventional medicine through integration or collaboration with traditional medicine. Although Good clearly assumed the superiority of conventional medicine, he did indicate the potential for collaboration from the national scale to the scale of individual facilities, and the need to investigate how integration relates to wider national and regional health goals and health planning. Mirroring the perspectives and concerns of the medical geography in the 1970s, the research agenda Good posed was focused on the accessibility and utilization of services, including investigating spatial arrangements of traditional medicine in urban and rural areas, factors in seeking traditional medicine, and various obstacles related to distance. Although these issues are still investigated to the present day by geographers, many now look for a deeper and often qualitative understanding of the cultures, values, and belief systems that relate to traditional medicine in particular places of the developing world. Meanwhile, other scholars have started to investigate the roles of traditional medicines used by indigenous people located in developed world countries (such as First Nation communities in Australia and Canada). Research here suggests that ideas about medicine and healing are often part of a broader cultural association with place and landscape. These can often conflict with dominant Western ideas. In contrast to these tensions, Maori notions of health and landscape have actively informed health policy in New Zealand.

Consumer and Provider Geographies

Not all research on CAM that has a geographical orientation has been conducted by geographers or is labeled as geography. For example, numerous studies by CAM researchers have considered regulatory issues within well defined areas of political jurisdiction, and profiles of CAM provision and use within and between places ranging in scale from countries to individual institutions. However, beyond sampling frames, in these studies, spatial issues are not often prioritized and given a central place in research questions and analysis.

The need for explicit geographical research on CAM was first highlighted by Anyinam in 1990. His argument was for wide ranging quantitative inquiry, including attention to local, regional, and national distributions of therapists, their relationships to biomedical services, their referral networks, and practice catchment areas. Only a modest number of studies answered this call and the ones that did mainly considered spatial trends in CAM provision at the regional level. In retrospect, given the vast range of CAM provided in so many places, responding effectively in this way was always going to be a very tall order. Moreover, another important disciplinary context was that very shortly after Anyinam's paper, medical geography was to undergo a significant reform that would witness the emergence of a qualitative 'postmedical' geography of health, focused far more on experiences of health and disease, than on distributive features of service provision as emphasized by Anyinam.

Since 2001, qualitative research on CAM, by Wiles, Andrews, Adams, Doel, Segrott, Williams, and others has been more varied and has had very different priorities. A consistent theme has been attention to cultural and consumer geographies of CAM. Studies are attempting to understand who uses CAM and why, how this differs between and across places, and how place itself shapes and is shaped by CAM. A variety of cultures are explored and articulated in research ranging from local spiritual and 'alternative' cultures in CAM, to 'high' consumer CAM cultures as progressed, for example, through lifestyle magazines. Research has also begun to consider how different cultures of CAM may be associated with versions of urbanicity and rurality. For example, a study in Devon, UK, articulated how certain forms of CAM are associated with spiritual ancient landscapes, idyllic landscapes, and certain rural cultures. Meanwhile other studies articulate how CAM is associated with urban living and city quarters. Studies have also considered design features of CAM settings and how they play an important part in the overall CAM consumer experience. For example, research has observed that some therapists' clinics are clearly decorated by them to resemble conventional medical places, though the majority of CAM clinics are purposefully designed and decorated to exert a subtle calming effect, and to provide a spiritual experience. Other research has focused on professional spaces where orthodox medicine and CAM are increasingly combined (such as GP surgeries).

Place and Presence

An emerging field of research considers the ways, beyond co presence, that place plays a part in the clinical practices and consumer experience of CAM. Williams termed this 'landscapes of the mind' in her discussion of imagery and visualization of place in psychotherapeutic practices, autogenic training, relaxation techniques, and meditation. Similarly, Bondi noted that many forms of counseling (often part of CAM) involve going 'with' someone – perhaps to the past, and often to a place, that may be real or may be metaphorical, and that places whether they are based on reality and fantasy, are used to explore clients' 'inner worlds'. Obvious connections can be made here to broader disciplinary theoretical interests such as psychoanalytic geographies. Indeed, Bondi suggested that some of the key theoretical texts that inform psychotherapeutic practices have geographical components such as Winnicott (potential space); Klein (splitting); Freud (time and unconscious); Wright (positions and viewpoints), and Kopp (journeying). These kinds of inquiry do offer critical insights into the nature of human perceptions and emotions and on the complex cognitive processes that link people and places. Notably, then, they also potentially extend interest in CAM to a wider audience of human geographers.

In 2004, Andrews followed up these observations on imagery and place with a qualitative empirical investigation of CAM practices. Imagined therapeutic places were found to be constructed and manipulated by both therapists and their clients. It was observed that, in some cases, mental imagery was subtle and supportive to other therapeutic techniques particularly manual techniques. In other instances, it was found to be a central feature of therapies and used more directly, particularly in dedicated imagework to construct mind bridges with past life events. Hypnotherapists, for example used their clients' unconscious minds in doing this. Connecting with nature and natural landscapes was found to play a significant part in mental imagery. These might be idyllic for purposes of relaxation or they might be strongly unpleasant and awesome places that clients paint as pictures and backdrops of how they feel. The issue of power relationships between therapists and their clients in mental imagery was also highlighted in Andrews research, particularly in terms of who moves imagery forward. Indeed, particular places being imagined by clients may be suggested or guided by the therapist, or may be directly chosen by the client, depending on the particular modality of CAM and the illness context. Certainly, a major observation of this study was that a degree of trust has to be necessarily given to those therapists who use imagework. Indeed, using mental imagery of places is not a fail safe method and a variety of problems are encountered, particularly with helping clients obtain or reach a specific state of mind.

Calls for Future Inquiry

Because CAM is a relatively new field of inquiry for geographers, naturally there are many opinions as to what future research should focus on and the per spectives it should take. CAM deconstructs traditional understandings of health and illness, and fuses a vast array of materials and practices. It therefore poses a serious challenge to human geography in terms of how to approach it. With its traditional concerns for services and health experiences, health geography alone might not be able to cover all the necessary ground. Indeed, as Doel and Segrott have noted, for health geographers terms like health, disease, illness medicine, and place need and use, are constant points of reference. However, as they also suggest, in CAM these are precisely the kinds of things that are subverted and redefined. Consequently, future geographical study on CAM will have to be able to deal with CAM much more broadly as a wide ranging con sumer phenomenon. This, as Doel and Segrott mention, should not be an overly difficult task as CAM is con sistent with human geography's current engagements with aspects of everyday life, human practices, and per formativity. Moreover, they suggest that CAM is con sistent with the principles of the emerging style of human geography with its emphasis on situated practice and nonrepresentational theory.

Others have argued that future research is needed to explore a greater variety of CAM places and practices. With regard to places, imagined places in practices (as outlined above), certainly provide an avenue for future inquiry. Another important issue is the transportation of traditional medicine from the places of their origin to the West. In this respect, postcolonial theory needs to be brought to bear in understanding of how traditional medicines might be morphed into CAM in new Western social and cultural contexts. With regard to perspectives, research on the political and economic geographies of CAM would certainly be welcomed. On the micro scale, this could involve, for example, researching aspects of integration with conventional medicine in shared places, and how different modalities of CAM may be integrated differently within different conventional places for healthcare. Also in terms of perspectives, there certainly needs to be a more thorough examination of gendered characteristics of CAM through a feminist lens, and commentators need to understand the ways in which gender may play out in various places for CAM. Also required is attention to how CAM is important to his tories of places, the historical identity of places, and how it might be associated with heritage sites and industries. There are certainly many other places and perspectives to be followed. However, the key message is that because CAM spans so many materials and practices of everyday social, cultural, and economic life, researching it must inevitably embrace multiple perspectives.

See also: Chronic Disease; Health Geography; Indigenous Health and Medicine; Medical Geography; Psychotherapy/Psychotherapeutic Geographies; Wellbeing.

Further Reading

Adams, J. (2004). Exploring the interface between complementary and alternative medicine (CAM) and rural general practice: A call for research. *Health and Place* 10(3), 285 287.

Andrews, G. J. (2003). Placing the consumption of private complementary medicine: Everyday geographies of older people's use. *Health and Place* 9, 337 349.

Andrews, G. J. (2004). (Re)thinking the dynamic between healthcare and place: Therapeutic geographies in treatment and care practices. *Area* 36(3), 307 318.

Anyinam, C. (1990). Alternative medicine in western industrial countries: An agenda for medical geography. *Canadian Geographer* 34(1), 69 76.

Clarke, D., Doel, M. and Segrott, J. (2004). No alternative? The regulation and professionalization of complementary and alternative medicine in the United Kingdom. *Health and Place* 10(4), 329 338.

Del Casino, V. J. (2004). (Re)placing health and health care: Mapping the competing discourses and practices of 'traditional' and 'modern' Thai medicine. *Health and Place* 10, 59 73.

Doel, M. and Segrott, J. (2003). Self, health and gender: Complementary and alternative medicine in the mass media. *Gender, Place and Culture* 10(2), 131 144.

Doel, M. and Segrott, J. (2003). Beyond belief: Complementary medicine, lifestyle magazines and the dis ease of everyday life. *Environment and Planning D: Society and Space* 21(6), 739 759.

Doel, M. and Segrott, J. (2004). Materializing complementary and alternative medicine: Aromatherapy, chiropractic and Chinese herbal medicine in the UK. *Geoforum* 35(6), 727 738.

Good, C. (1977). Traditional medicine: An agenda for medical geography. *Social Science and Medicine* 11, 705 713.

Stock, R. (1981). Traditional healers in rural Hausaland. *Geojournal* 5(4), 363 368.

Verheij, R., Bakker, D. and Groenewegen, P. (1999). Is there a geography of alternative medical treatment in the Netherlands? *Health and Place* 5, 83 97.

Williams, A. (1998). Therapeutic landscapes in holistic medicine. *Social Science and Medicine* 46(9), 1193 1203.

Williams, A. (2000). The diffusion of alternative health care: A Canadian case study of chiropractic and naturopathic practices. *Canadian Geographer* 44(2), 152 166.

Wiles, J. and Rosenberg, M. W. (2001). Gentle caring experience: Seeking alternative health care in Canada. *Health and Place* 7(3), 209 224.

Wilson, K. (2003). Therapeutic landscapes and First Nations Peoples: An exploration of culture, health and place. *Health and Place* 9(2), 83 93.

Wilson, K. and Rosenberg, M. (2002). *Exploring the determinants of health for First Nations peoples in Canada: Can existing frameworks accommodate traditional activities? Social Science and Medicine* 55, 2017 2031.

Relevant Websites

http://www.incamresearch.ca
Canadian Interdisciplinary Network for CAM Research (IN CAM).

http://www.camline.org
Evidence Based Complementary and Alternative Medicine.

Complexity Theory, Nonlinear Dynamic Spatial Systems

D. O'Sullivan, The University of Auckland, Auckland, New Zealand

Glossary

Adaptivity The ability of a system to alter itself in response to changes in its environment so as to preserve its own existence and operation.
Dynamics A general term for the ways in which a system changes its state, also a branch of physics devoted to the study of the dynamics of systems.
Graph Theory That branch of mathematics concerned with the analysis of network (i.e., graph) structures.
Linear (and Nonlinear) Linear relationships among a system's elements are those where small changes in the conditions reliably lead to small changes in the resultant effects; no such restriction applies to nonlinear interactions.
System A collection of entities (that may themselves be subsystems) that interact with one another so as to function as a coherent whole.

System complexity refers to structural and dynamic be havioral characteristics of systems that are considered qualitatively different from the systems studied in clas sical science. Classical systems consist of either small numbers of interacting elements, with interactions that are linear, so that small changes in the inputs lead to small changes in the outputs; or, very large numbers of elements (in the millions or more) where statistical ap proaches can account for system level behavior. These constraints on the systems studied in classical science result from the intractability of the mathematics required to represent systems of intermediate size with nonlinear interactions among elements. In recent decades, complex systems consisting of tens to millions of elements, with intricately structured, usually localized interactions among elements have been widely studied predomin antly using computational models. The resulting com plexity theory posits a number of structural and dynamic properties of complex systems, such as emergence, self organization, path dependence, and tipping points. These concepts have risen to prominence in many disciplines in recent decades, popularized not only by scientific inter est, but by the accompanying promise of a 'holistic' sci ence, with broad cross disciplinary appeal. It remains unclear whether the implicit promise of complexity theory to bring greater unity across physical and social sciences can be realized.

Origins

The roots of complexity theory are widespread. It can be regarded as a recasting of general systems theory as ad vocated by von Bertalanffy in the late 1960s. Other im portant jumping off points include artificial intelligence, where the work of Herbert Simon has been influential, and game theory in economics. Another source is non mainstream mathematical biology early in the twentieth century, particularly D'Arcy Thomson's *On Growth and Form*. More recently, various schools of thought in physics, such as Prigogine's nonlinear dynamical per spective on the second law of thermodynamics, and Haken's synergetics have also been widely taken up.

The origins of now dominant North American con ceptualizations of complexity lie in optimistic commen tary after World War II about the prospect of tackling social and economic problems by coordinated scientific effort. In the wake of the achievements of the scientific expertise brought to bear on wartime challenges (leading to the development of not only nuclear energy and the jet engine, but significantly in the current context, the first electronic computers), it was suggested that the social and economic problems facing the post war world might be tackled by similar scientific efforts. This view was most cogently expressed in a prescient piece entitled 'Science and complexity' by Warren Weaver, who dis tinguishes between the simple small systems of classical science, systems of 'disorganized complexity' such as gases, which are amenable to statistical explanation, and systems exhibiting 'organized complexity' consisting of many elements interacting in nonlinear ways. Weaver referred directly to the importance of computers in providing the capability to develop understanding of such complex systems, and suggested that operations research, systems theory, and cybernetics would be crit ical to the study of complex systems. Much of Weaver's vision has come to pass under the rubric of 'complexity science' particularly following the widespread availability of desktop computers accessible to most working scientists.

Structure and Dynamics

A complex system is a 'medium sized' collection of elements interacting with one another in more or less simple ways. In this context, 'medium sized' ranges widely from more than two to hundreds of thousands, even millions of elements. Elements are themselves simple, or are treated as such at the current level of analysis. Interactions between elements are not restricted to particular functional forms and may be governed by linear equations or by nonlinear interactions such that levels of interaction and their effects may in turn depend on other features of the elements interacting. In most accounts interactions among elements are localized, often but not necessarily spatially, but in highly structured ways, so that every element does not interact with every other at all times. This means that interactions among elements vary depending on location.

Systems considered 'complex' range widely from ant hills and brains, to cities, the Internet, ecosystems, and the global economy. Such systems have been studied using computational simulations, and the resulting knowledge about typical structural and dynamic characteristics may be loosely termed complexity theory. It is a central tenet of complexity theory that in spite of their diversity, complex systems share common characteristics and be haviors due to their similar structural features, and that these characteristics apply regardless of the particular details of the complex system in question. Thus, structural similarities between the Internet, the global economy, and the human brain in terms of interactions between the constituent elements of each – whether web servers, corporations, firms and governments, or neurons – matter as much, or even more, than obvious differences between those elements for understanding the systems and their behaviors.

Interaction Networks

An important theme in studies of complex systems is thus the structure of patterns of interactions among system elements. Structure may be measured and understood in terms of a network of relations among elements, which can be analyzed using graph theory. Some specific forms of structural organization are considered more inter esting due to their prevalence in complex systems.

Hierarchically organized systems consisting of one or more levels of nested subsystems may be considered complex. This perspective is prevalent in biological and engineering sciences where functional separation of elements into cooperating subsystems at a number of levels is a key feature. Biologically, interest focuses on how functional subsystems within an organism are organized to interact with one another. From an engin eering perspective, the interest is in what principles for the design of robust engineered systems can be learned from naturally occurring hierarchical systems. John Holland and Herbert Simon's work in particular has focused attention on hierarchically organized systems.

Other network structures have been the subject of considerable more recent interest. First is the apparently paradoxical 'small world', which combines strong local ized interactions with efficient communication from any element in the structure to any other. The term derives from the 'small world' phenomenon in social settings where, anecdotally, it is commonplace for apparently unrelated people to discover that they have acquaint ances in common. The operation of diffusion or similar processes is different on a small world network from a random or spatially structured network, and this has been a subject of considerable research.

Second is the scale free network where the numbers of neighbors associated with each node in a network conform to a power law distribution, meaning that there are few highly connected nodes and many peripheral or weakly connected nodes. Of particular interest is the nature of growth processes that lead to networks of this type, and the robustness of such systems when some links are lost. This research is closely related to work on city or firm size distributions.

Arguably, the structure of the interaction networks that might emerge between geographically embedded entities in a complex system is underexplored relative to these network structures. How geographically embedded networks (e.g., national or regional scale producers and distributors) and unembedded networks (e.g., global fi nance and multinational corporate networks) might interact is also a neglected area.

Emergence

Consideration of processes yielding different interaction structures leads naturally to the concepts of emergence and self organization. Emergence is the simple idea that complex aggregate outcomes may result from simple interactions among the elements of a complex system.

In spite of its apparent simplicity, emergence is problematic. Ontologically, it presents difficulties because it is not at all clear that any mystical notion of emergence is necessary if all that it entails is the existence of ag gregate system level properties based on the states of constituent elements. Emergence remains problematic where it implies more than this, particularly when it is taken to mean that aggregate level properties are not predictable based on an understanding of element level properties and interactions. In this case, emergence is an observer dependent concept, which is not robust, be cause, having once observed an 'emergent' aggregate outcome in a system it can no longer be unexpected. It may be more useful in such cases to think of emergence as a rhetorical device that enables claims about the ex planatory efficacy of the interactions proposed for low level elements, based on their ability to account for particular aggregate outcomes.

An alternative understanding of emergence is a broader one that defines a series of ontological levels of reality, such as subatomic particles, atoms, molecules, proteins, and so on, where elements observed at one level emerge from interactions among elements at the next level down. Each level can be related to a scientific discipline. This scheme recognizes different disciplines as a way of dealing with the different types of entities and interactions observed at each level.

While this approach is attractive for its neatness, at least three problems arise. First, this notion of emergence is rather different from more mundane notions expressed in terms of aggregate outcomes, so that care is required in using this framework to justify claims for the more workaday concept. Second, it is ironic that some would advocate such an understanding while at the same time asserting, as complexity theory does, the efficacy of abstract models of interacting elements as providing explanations of any transition from one level to another, independent of the disciplinary context. If the point is that entities and interactions different in kind emerge at each level, then surely different models are required for each level. Third, and most critically for human geography, the unambiguous series of levels proposed by such a scheme, if it is true of any phenomena (a moot point) appears untenable when we consider the 'levels' of existence in the social world. In the social realm, individual elements (human beings?) do not simply interact to produce a single higher level (society?) but instead act and interact in many diverse and interconnected settings (households, families, nations, ethnicities, pressure groups, workplaces, churches, unions, etc.) across a range of social and spatial scales.

Self-Organization

Self organization imposes on top of emergence the idea that the emergent outcomes a system exhibits are often structured so that the system can be considered to be organized into a number of interacting subsystems.

Self organization, partly because it demands more of the world than emergence, may be a more useful concept. The idea that it is possible for functionally distinct subsystems to develop in a complex system seems inherently geographic. Self organization presupposes spatial differentiation in a system and in a geographical context evokes concepts such as regions, or the functional areas of a city.

A further concept often overlaid on to self organization is adaptivity. A self organized system that is also considered adaptive is termed a complex adaptive system. Such systems are widely considered to be the paradigmatic objects of interest in complexity theory. How systems adapt to their environment in order to maintain their continuing function is a question with relevance in biology (where it is applied to organisms), but equally to social sciences where it may be applied to 'systems' such as language, society itself, or indeed any collective social structure.

Chaotic Dynamics and the Edge of Chaos

A chaotic system is one that, while appearing to behave randomly, is actually governed by simple deterministic rules, so that if its initial state were known precisely then all future states could be accurately predicted. However, because very small differences in the initial conditions of a system can lead to arbitrarily large differences in state at some subsequent time, it is impossible for the initial state of a system to be known precisely enough, and long term prediction is impossible. This is referred to as sensitive dependence on initial conditions. While many simple mathematical systems have been shown to exhibit chaos, it is unclear that such dynamics are a good model of any sociospatial system, nor are chaotic dynamics a necessary feature of complex systems. Indeed, from some perspectives, much of the interest in complex systems is in their avoidance of chaos: how does the relative order exhibited by complex systems emerge, in spite of the disordered environments in which they exist?

An interesting notion relating complexity and chaos to one another is the idea that complexity exists at the 'edge of chaos'. This idea suggests that the inherent dynamism of chaotic systems arising from their capacity for rapid change is a feature that makes for greater adaptivity in complex systems, and that for this reason many complex systems are poised on the 'edge of chaos'. Related, and highly ambitious, claims are made for 'self organized criticality', which suggests that many systems spontaneously evolve to a point where they are prone to rapid change at a very large range of scales. While both concepts have garnered considerable attention, particularly self organized criticality, it is not clear that either framework can be readily attached to sociospatial systems.

Positive Feedback and Path Dependence

A different type of sensitive dependence on initial conditions arises from the twin concepts of positive feedback and path dependence. Study of the dynamics of classical systems emphasizes negative feedback processes which tend to restore a system to equilibrium. Potentially disruptive positive feedback effects are of more interest in complexity science. A positive feedback is any effect that tends to reinforce a system's changes along whatever development path it is already pursuing. The most familiar example in geography is urban growth in capitalist economies where rapid growth tends to create the conditions for further rapid growth. Positive feedback effects create the potential for small initial changes to lead to rapid large scale changes in a system. Small initial competitive advantages for a region may, through positive feedback, lead to enormous differences in outcomes between regions. Silicon Valley's dominance of turn of the century information technologies provides a canonical example.

Positive feedback leads directly to the notion of path dependence, and in parallel, or equivalently in a spatially distributed system, place dependence. If small effects can be magnified via positive feedback into large differences, then the particular sequence of historical and geographical events across a system (or at localized places within it) is important in developing explanations. This perspective emphasizes the importance of historical and geographical accounts of complex systems, and draws attention to the importance of contingency in explanation.

Tipping Points and Critical Mass

Another frequently identified feature of complex dynamics is the existence of tipping points. A geographical example is urban residential change when slow shifts in a neighborhood's demographics may reach a tipping point as a result of a critical mass of (say) immigrants entering the neighborhood. Beyond the tipping point rapid neighborhood change occurs, leading quickly to complete demographic transition. The same concept might apply equally to incoming gentrifiers. Similar dynamics are posited for adoption of innovations such as mobile phones or fax machines. In most cases, such dynamics are driven by positive feedback. In a social context, if or when tipping points are perceived by the society or group in question is a key issue.

Issues Arising from Structure and Dynamics

Some general points can be drawn from the foregoing discussion:

1. The ambitious generality of claims made in complexity theory – the claim to generality, together with a degree of intuitive obviousness – insofar as concepts such as path dependence and emergence seem familiar – explains why complexity science has successfully established currency in so many disciplines. Complexity theory moves systems based science into a position where social phenomena may now be a viable object of study using these methods, because the models used are no longer restricted to idealized and unrealistic behaviors (such as the market equilibria of economics). The downside of course, is that the very generality of the concepts promises a great deal, but struggles to deliver much in the way of satisfying explanations of anything very specific. When brought into any particular context, the explanatory power of ideas from complexity theory often proves rather limited.
2. Often, in the application of complexity theory to social phenomena, emergence of aggregate behaviors from individual level interactions is emphasized over any downward causation effects operating in the opposite direction. While downward effects are implicitly considered insofar as individual behavior is affected by the localized characteristics of a system, in general, the overall aggregate state of the system is not a factor in the behavior of individuals. For sociospatial systems this seems a limitation.
3. A focus on the relationships between the (often spatial) structure of a system and its dynamics is a core theme in geography that finds strong support from complexity theory. A closely related observation is the inherently spatial nature of complex systems. The localized context of the actions of individual elements in a system is recognized as a key aspect of how system behavior unfolds, and spatial differentiation of a system is both expected and an outcome of interest. The self organization and path (or place) dependence concepts in particular, bring these aspects of complex systems to the fore.

The Role of Computer Models and Simulation

The foregoing discussion summarizes key phenomenological findings of complexity theory. At least, as interesting as these ideas are their epistemological underpinnings. As has been noted, systems theory, computer science, artificial intelligence, and cybernetics are important sources of complexity science, and in keeping with this ancestry, computer models are the dominant research practice in complexity science. This approach involves development of a computational model of the phenomenon (i.e., the complex system) of interest, and subsequent exploration of its behavior. Findings from the model are mapped back onto the real world as claims about possible behaviors of the original phenomenon.

This bald description hides a very wide range of actual research. Two particular model types have been of particular interest in explicitly spatial settings: cellular automata and agent based models. More generally, models vary in style from highly abstract and simplified thought experiments, to elaborate, realistic simulations. At the simple abstract end of the scale, models may simply be a way to sketch out broad characteristics of a system, often in terms of the features discussed above. At the realistic end of the scale, research may be more geared toward exploring scenarios in contexts closely related to policy formulation, with little concern for the broad system characteristics. Arguably, this approach is not complexity science so much as a pragmatic response to increasingly complicated decision making problems.

Many aspects of model based research are not scrutinized closely from a philosophical perspective, nor are they always explicitly reported in publications. Precisely how a researcher chooses to represent a system in a model is critical to all subsequent findings. Given the theoretical expectation of at least two 'levels' in a system, one the basic elements and their interactions, the other the system level itself, definition of elements and their interactions is crucial. Equally, which elements are omitted is critical. By their nature, these decisions are difficult to report adequately, when the focus is on the model that was developed, rather than on the infinite number of other models that might have been developed. Thus the representational adequacy of models is not often considered.

One strategy for sidestepping this problem is to argue that models are simply thought experiments or explorations of the implications of (usually) simplified assumptions about how the world is. While this position is intellectually defensible, in the absence of empirically grounded follow up, it is hard to justify such research as an end in itself. Such follow up remains rare, in part because of weak links between researchers using model based methods, and more mainstream social scientists. Without supportive empirically grounded research, assessing how well a model works as a representation of the phenomena it is purported to represent remains a matter for debate and argument.

In the absence of verification, many modelers turn to some form of validation. Validation can take two forms. Either a model is checked for internal consistency, or its outputs are validated against empirical data. While the former is necessary, consistency checking cannot establish the representational accuracy of a model. Validation against empirical data appears a more promising procedure, but also fails to establish the representational merits of a model: a failure to reproduce observed phenomena may suggest that a model is a poor representation, but the converse is not true, and in any case producing a good fit to observational data does not rule out the possibility that an entirely different model could produce a similarly good fit.

These methodological issues represent significant challenges to any research program substantially founded on computational modeling. Partly as a result, finding appropriate ways to report research based on computational models has become an important (if minor) theme in geography and elsewhere, with some research focusing on how model results are interpreted and deployed in large scale decision making (e.g., in the context of climate change), and others suggesting that it is important for model based efforts to be connected to other approaches more fully.

Impacts in Human Geography

Examples of self conscious complexity based research in human geography are unusual. Complexity science has risen to prominence during a period when human geography has been dominated by the humanistic, structuralist/Marxist, and post structural/cultural perspectives and the accompanying philosophical and methodological prescriptions are unsympathetic to the computational modeling approach.

The only field where complexity has had a direct impact is at the interface between urban planning and geography where researchers have developed model based approaches to understanding urban growth, sprawl, transportation, and other related phenomena. This area is fertile ground for the development of new variants of both the cellular automata and agent based model types. Another focus of attention is in studies of land use change where model based approaches, occasionally informed by complexity theory, have become commonplace. Another connection exists to the extent that world systems theory is influential in some areas of human geography, and is also consistent with complexity theory.

The impact of these strands of work on the discipline more widely has been limited, and other disciplines (particularly sociology, economics, planning, and architecture) have often been a more receptive audience. The geographic information science community has been a receptive audience within geography, and methodological issues relating to the integration of models with geospatial data and to the challenges of exploring the large data sets generated by models are of particular interest.

The evident isomorphism of many key themes in geography with those of complexity theory, together with the broad disciplinary reach of complexity, has more recently witnessed a greater interest from the geographical mainstream. For some, the isomorphism provides a new language to draw on in theorizing the global economy, a language that may prove more durable than earlier, rather loose analogies with chaos and fractals. Attention has also focused on the historical geography of complexity and on how its metaphors have successfully crossed disciplinary boundaries to permeate not only science but the culture more widely.

The overlaps and connections between post World War II technocratic science, complexity as a new scientific paradigm, new age perspectives on complexity as 'holistic' science, and the currency of complexity as a buzzword in business certainly suggest that complexity itself is a suitable topic for economic or cultural geographical study, given the popularity of actor network theory and other approaches from the sociology of scientific knowledge tradition. Whether or not complexity science simply represents a rebranding of social physics, its logical conclusion, or something more significant also remains an open question.

Such issues are critical to the long term potential of complexity as a bridge between quantitative and quali tative traditions in human geography. While complexity science brings to the fore a recognition of the intricately structured interconnectedness of the world in a way that previous quantitative perspectives arguably have not, properly understood, it also holds a place for concrete, empirically grounded research, both qualitative and quantitative. In addition, the role of computational models opens up avenues for dialog, at the same time that it makes for an explicit acknowledgement of the act of representation in quantitative work. These are aspects of complexity theory where human geography appears well placed to make significant contributions.

See also: Cellular Automata; Chaos and Complexity; Fractal Analysis; Network Analysis; Simulation; Social Studies of Scientific Knowledge; Systems; World-System.

Further Reading

Bak, P. (1997). *How Nature Works: The Science of Self Organized Criticality*. Oxford: Oxford University Press.

Batty, M. (2006). *Cities and Complexity: Understanding Cities with Cellular Automata, Agent Based Models, and Fractals*. Boston, MA: MIT Press.

Byrne, D. (1998). *Complexity Theory and the Social Sciences: An Introduction*. London: Routledge.

Cilliers, P. (1998). *Complexity and Postmodernism: Understanding Complex Systems*. London: Routledge.

Gleick, J. (1987). *Chaos: Making a New Science*. New York: Viking Penguin.

Harvey, D. L. and Reed, M. (1996). Social science as the study of complex systems. In Kiel, L. D. & Elliott, E. (eds.) *Chaos Theory in the Social Sciences: Foundations and Applications*, pp 295 323. Ann Arbor, MI: University of Michigan Press.

Holland, J. H. (1998). *Emergence: From Chaos to Order*. Cambridge, MA: Perseus.

Manson, S. M. (2001). Simplifying complexity: A review of complexity theory. *Geoforum* 32, 405 414.

Oreskes, N., Shrader Frechette, K. and Belitz, K. (1994). Verification, validation and confirmation of numerical models in the earth sciences. *Science* 263, 641 646.

O'Sullivan, D. (2004). Complexity science and human geography. *Transactions of the Institute of British Geographers* 29, 282 295.

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society* 106, 467 482.

Thrift, N. (1999). The place of complexity. *Theory, Culture and Society* 16, 31 69.

Urry, J. (2003). *Global Complexity*. Cambridge: Polity.

Waldrop, M. (1992). *Complexity: The Emerging Science at the Edge of Order and Chaos*. New York: Simon and Schuster.

Weaver, W. (1948). Science and complexity. *American Scientist* 36, 536 544.

Relevant Websites

http://www.cscs.umich.edu/
Center for the Study of Complex Systems at the University of Michigan.

http://necsi.org/
New England Complex Systems Institute.

http://cssociety.org
The Complex Systems Society.

http://www3.interscience.wiley.com/journal/38804/home
The journal 'Complexity'.

http://emergence.org/
The journal 'Emergence'.

http://www.santafe.edu/
The Santa Fe Institute.

Computational Human Geography

H. Couclelis, University of California, Santa Barbara, CA, USA

Glossary

Algorithm A set of precise step-by-step instructions to be followed by a computer in solving a problem or simulating a process.

Artificial Intelligence (AI) A computer science field aiming at the development of programs or devices exhibiting behavior that may be characterized as intelligent. Hard AI purports to replicate processes of human intelligence, whereas soft AI is only concerned with the outcomes of the behavior.

Augmented Environment A space equipped with sensors, transmitters, and receivers that allow suitably equipped users to obtain and/or post information about that space that would not normally be directly available to the unaided senses.

Complex System A system of many interacting components that is open to exchanges of energy, matter, or information with its environment. Complex systems have characteristic properties and behaviors (nonlinearities, unpredictability, self-organization, chaos, emergence, etc.) that differentiate them sharply from the simple, closed systems of classical science.

Deep Uncertainty Uncertainty about future states of the world and/or the outcomes of actions that is qualitative in nature, that is, cannot be expressed in probability terms.

Four-Color Theorem (Conjecture) The first major mathematical theorem to be proved with the help of a computer. The proof confirmed that four different colors are necessary and sufficient to color any arbitrarily complex map so that no two adjacent areas have the same color. This landmark proof has special significance for computational geography because the problem is explicitly spatial.

Heuristic A computational shortcut for obtaining approximate solutions to problems for which analytical solutions either do not exist or could not be derived within a reasonable amount of time.

Monte Carlo Techniques Computational procedures that use random sampling algorithms to derive the statistical properties of classes of patterns.

Numerical Solution A solution derived through a digital approximation method, as opposed to an exact analytic solution.

Optimization A class of techniques for finding minimum or maximum values of functions subject to a number of constraints.

Simulation Model A computational model that mimics the behavior of some dynamic empirical process of interest.

Introduction

Origins of the Notion

Computational human geography is part of a broad spectrum of fields in both the natural and the social sciences that have adopted the qualifier 'computational' to describe their methodological and epistemological approach to their subject matter. Thus we have, on the natural science side, computational physics, chemistry, biology, and so on, and even computational mathematics. Similarly, in the social sciences, the terms computational economics, political science, anthropology, linguistics, etc., have appeared in the literature in recent decades, sometimes also as titles of professional journals (e.g., *Computational Economics*, *Computational Linguistics*). All these relatively new fields are part of computational natural or social science, or computational science more generally. The term computational science is sometimes used interchangeably with scientific computing, though the latter may have a narrower, more applied meaning. The rapidly increasing number of journals with related titles (e.g., *Scientific Computing World*, *SIAM Journal on Scientific Computing*) and of computational science degree programs, from BS to PhD, at major universities world wide, attest to the growing interest in computational approaches across the sciences. In geography, the terms Geocomputation (the title of a successful international conference series), geoinformatics, or geomatics (the names of several engineering departments specializing in geospatial computing), and Computational Geography (the name of a research center at the University of Leeds) are closely related to computational human geography, though they are not restricted to the social science side of the discipline.

Computational Science and the Theory of Computation

Computational science has its origins in the first digital computers that became available for use by scientists in the mid to late 1940s. Even in their early, primitive incarnations these machines could tackle the massive calculations needed to solve complicated problems much more efficiently than professional humans could. But scientific computation proper began when the use of computers moved beyond mere number crunching to helping derive approximate numerical solutions to problems that were too large or difficult to solve ana lytically. In academia these computational approaches to problem solving were at first considered very much second best to traditional methods, especially in the formal sciences such as mathematics, theoretical physics, and economics. A turning point in the status of compu tation as a legitimate scientific approach was reached in the 1970s when the four color conjecture, one of the famous unsolved problems in mathematics, was proved; or rather, it was demonstrated to be true by two math ematicians who used a computer to perform an ex haustive search through some 1500 configurations representing all possible classes of four color maps. This caused quite a stir in the discipline. At first many mathematicians refused to accept the computer aided solution as a valid proof but eventually the work was verified independently, while other similarly derived proofs started appearing in prestigious mathematical journals. Yet, what boosted computation to the prominent position it occupies today is above all the rise of the complex systems paradigm in the natural and social sciences, which occurred around the same time. Research issues framed as complex systems problems are almost by definition not analytically tractable and require the de velopment and implementation of numerical algorithms for their solution. More importantly, complex systems are typically dynamic systems undergoing some process, and algorithms themselves describe a complex process. Thus, algorithms can be used not only to solve problems for which mathematical formulations already exist, but also to formulate complex system problems from scratch as complex processes.

The power of algorithms in representing all manner of process is established in the formal theory of computation, which is due in large part to the British mathematician Alan Turing. The theory of computation is a branch of modern (discrete) mathematics, which has its roots in set theory. A central idea in Turing's theory is that of an effective procedure, that is, a set of unambiguous step by step instructions for carrying out a task. Algorithms embody the notion of effective procedure. A still unproved conjecture known as the Church–Turing thesis states that any problem that can be represented and solved in any other way can also be represented as an effective pro cedure and solved by a corresponding algorithm. In other words, all of science – past, present, and future – may be recast as computational science. A well known recent ex ponent of this thesis is the mathematician Stephen Wolf ram, who claims that the computational structures known as cellular automata (CA) can provide models for all of science. These are part of automata theory, one of three branches of the theory of computation. Automata are formally specified classes of effective procedures with well defined properties. Their dynamic behavior is governed by sets of if–then rules that operate on the automaton's current and possible states. CA consist of one , two , or higher dimensional arrays of automata, each of which updates its current state by also taking into ac count the states of neighboring automata in the array.

Both automata and CA have become an integral part of computational human geography (see below), though their formal origins in the theory of computation – and thus their analytic power – are not always recognized. CA are of particular interest to computational human geo graphy because they are inherently spatial. They are used extensively to represent urban, regional, and environ mental dynamics and other phenomena of spatial change. Automata also underlie agent based models (ABM), which have become increasingly popular in recent years in geography and in the social sciences more generally. Both these very important classes of applications involve the development of simulation models, that is, algorithms embodying specific computational processes that are as sumed to mimic the empirical processes of interest. Be yond these process oriented applications there is a wide array of other methods and techniques that use compu tation to generate problem solving procedures for stat istical, graphical, and other forms of data processing. The next section discusses some of the very diverse uses of computation in human geography.

Computation in Human Geography

Historical Overview

Sometimes computation in (human) geography is thought to be practically synonymous with the use of geographic information systems (GIS), but in actual fact the history of the field's involvement with computation is both much longer and broader than its association with GIS. From the 1950s on, the rise of the quantitative revolution offered fertile ground for experimentation with computational techniques in the context of spatial analysis, optimization problems, and early models of spatial behavior and decision. Hägerstrand's pioneering work in spatial diffusion demonstrated two major notions related to computation: (1) the idea that spatial diffusion is a process that may be described as an effective pro cedure, and (2) the utility of a computational stochastic sampling technique from the class known as Monte Carlo simulations. Hägerstrand's work thus touched on both major aspects of scientific computation: as direct process representation, and as data handling methodology.

Human geography had another significant early en counter with computation in the form of the first gen eration of urban models that appeared in the 1960s. In practical terms these models turned out to be over ambitious, expensive, heavily criticized and ridiculed failures, especially in the USA, but the seeds for the computational solution of large empirical problems were planted in the discipline at that time. These early com puter models consisted of very large systems of simul taneous equations that would have been extremely difficult to solve manually. Approximate solutions were generated by iterative algorithms that assigned values to the different model parameters until an acceptable point of convergence toward equilibrium was reached. The failure of these models had less to do with computation *per se* as with certain immature conceptualizations of cities as static equilibrium systems.

Partly in reaction to such mechanistic representations of human phenomena as aggregate systems of statistical regularities, behavioral geography emerged as a movement emphasizing the perceiving, thinking individual who makes and carries out decisions in geographic space. Work in that area has been quite diverse, ranging from humanistic perspectives on values, attachments, and emotions relating to place, to theoretical models of spatial decision inspired from microeconomics or game theory, to empirical observations of individuals or groups making housing, transportation, shopping, land use, and other choices in urban or rural environments. In addition to introducing human geography to experimental psychology and cognitive science, this latter thrust in behavioral geography was particularly amenable to approaches from another major development to come out of Turing's the ory of computation: artificial intelligence (AI). Computa tional models of housing search, of orientation and route finding, of directional biases in spatial behavior, and so on, began appearing in the literature by the late 1970s. These AI inspired models emphasized the unobservable cogni tive processes thought to underlie spatial decision and behavior, including the generation of cognitive maps.

Beyond the modeling of spatial processes, and often in conjunction with it, is another major use of computation in human geography. This is the more standard use of the computer as a data handling and numerical problem solving device, providing solutions to mathematical and statistical problems such as those found in spatial analysis and spatial optimization. Such problems often can be solved only approximately in practice because an exact solution would take far too long to calculate, even on the fastest computers. In this case heuristics are used, that is, computational shortcuts that can produce good enough solutions. Heuristics are also used by default in the nu merous instances of problems in human geography that have no formal expression but are directly translated from the conceptual to the computational stage. These forms of computing as numerical problem solving also have a long and distinguished history in the field. Methods for searching large solution spaces, for model calibration, for function optimization, for spatial cluster detection, for dealing with data uncertainty, and many others were routinely used by quantitative geographers already in the 1970s. By the 1980s the term automated geography appeared in the literature, along with a number of sophisticated pattern recognition, classifi cation, optimization, and other heuristics derived from AI (e.g., neural network analysis, genetic algorithms, simu lated annealing, and self organizing maps). Interest in these techniques has grown along with the spread of computation in human geography.

While not directly part of computational human geography, computer cartography and GIS have affected the field profoundly. Computer aided cartography appeared as soon as computers were able to produce some primitive graphics, and around the same time the ancestor of GIS took shape in the guise of the computer based cadastre for Canada. Both fields quickly transcended their origins as the computerized versions of traditional ways of doing things to move in unanticipated, vibrant new directions. Computer cartography soon extended into animated cartography, digital terrain modeling, and three dimensional (3 D) visualization, the dynamic mapping of abstract flows (e.g., of migration or money flows), the in tegration of field and remotely sensed data (also from urban areas), and so on, and even took care of stylistic problems such as of the correct placement of lettering on maps. As for GIS, its explosive growth into a technology affecting a host of domains well beyond geography – human or other – has been amply documented. While being the most prominent piece of the computational revolution in (human) geography, GIS is clearly not the only one.

The Present – and beyond

All these early involvements of human geography with computation continue to flourish today. There are by now few efforts in human geography that are not touched by some development or other coming out of the com putational turn in the discipline – be it a computer generated map, the visualization of an environment, a simple GIS analysis, or a tool for spatial data exploration. Current work in computational human geography may be briefly discussed under three headings: computation as spatial process representation, computation as problem solving procedure and for visualization, and mobile and pervasive computation.

Models describing spatial processes have by now largely replaced static models of spatial organization in many areas of quantitative human geography, and in par ticular in the study of urban and regional growth and change. Considering the intuitive understanding of these phenomena as being essentially dynamic and complex, much of the work in these areas since the late 1980s has looked to the complex systems paradigm for method ologies as well as for conceptual inspiration. Of the com plex systems models available in the physico mathematical and computational literature, CA stood out as being dir ectly applicable to geographical problems. Indeed, CA have several desirable properties: they are inherently spa tial; because they generate complex macroscale patterns as the result of micro scale interactions, they may be seen as metaphors for urban and regional organization resulting from a myriad of interdependent decisions by individual actors; they occasionally produce surprising outcomes that stimulate new thinking about the real systems modeled; they are relatively easy to program and implement; and they produce striking dynamic visualizations. CA based models are used in human geography in two different ways: as simulations of actual urban, regional, and other land use change processes; and as instructive, intuition sharpening metaphors, providing opportunities for novel thought experiments illuminating the properties of spatial processes viewed as complex processes. In the first case, the rigid structure of formal CA must be substantially modi fied in order to provide credible empirical models of spatial change. While these generalized CA models can be calibrated to fit actual data very closely, many of their interesting dynamic properties established in theoretical computer science are thereby lost. In the second case, whereby simple CA models are developed as instructive metaphors, there is little attempt to make these look realistic, focusing instead on isolating and exploring spe cific properties of interest. One of the earliest and best known examples of this second kind was developed not by a human geographer but by a Nobel laureate economist. Schelling's model of spatial segregation demonstrates that even mild individual preferences for being close to one's 'kind' may lead to starkly segregated spatial patterns.

As for the processes of spatial decision and behavior, the AI inspired, custom built models of earlier be havioral geography have been largely superseded by more generic types of agent based models. Agents are classes of automata simulating decision making entities that can construct representations of their environment, interact with and act on their environment and other agents, and pursue goals. Simpler forms of agents have more limited sets of properties. Agents may be spatially mobile or not, and they may represent individuals, groups, institutional decision makers, etc. Interest in this class of models is growing rapidly in human geography, fueled by both the ability to represent decision and be havior with considerable realism, and by the availability of software that greatly facilitates implementation. In creasingly, CA and ABM are being integrated into more complex structures, with the former representing pro cesses in the spatial environment, and the latter the de cision making entities affecting, and being affected by these processes.

The use of computing for data processing and nu merical problem solving has also exploded in parts of human geography, in parallel with the trend for com putation to become ever more accessible and less tech nically demanding. Most of the techniques of earlier periods are still being used and refined today, with an emphasis on increasingly powerful and easy to use soft ware packages as well as good quality data, both of which are now often available free to researchers over the web. In conjunction with increasingly sophisticated possi bilities for data visualization there is also growing interest in exploratory data analysis and data mining, which are classes of techniques that facilitate the discovery of interesting patterns in large datasets. It is now standard for software to come with interactive graphic interfaces that blur the distinction between confirmatory and ex ploratory analysis. This fusion of computational data handling, analysis and visualization is a very powerful development because it capitalizes on both the delib erative and the intuitive capacities of the human mind, creating fruitful right brain–left brain synergies. Visual ization is also used extensively to communicate ideas among researchers and especially between researchers and the public, where the proverbial picture that is worth a thousand words can help break down the notorious barriers separating experts and nonexperts.

Next to these evolutionary computational develop ments there are also some fairly revolutionary ones that are beginning to affect parts of human geography. These derive from a number of new digital technologies that support distributed, wearable, and mobile computing. Part of this broader trend toward pervasive computing involves the convergence of existing technologies (e.g., cell phones and wireless Internet, or Global Positioning Systems (GPS) and GIS), while others are early appli cations of genuinely novel technologies, such as nano technology. Examples of new research areas in human geography that are beginning to capitalize on these de velopments are location based services (LBS) and aug mented environments. LBS allow subjects on the move to receive GIS generated maps on hand held devices such as cell phones that display the location of desired nearby services (commercial or other), or that show good routes from the subject's current position to a destination. Augmented environments are urban (usually) environ ments outfitted with built in sensors, wireless transmit ters, and receivers that can provide to suitably equipped users local information that would normally not be dir ectly accessible to them. Combinations of sonic, visual, or virtual reality (VR) interfaces may be used. Applications can range from narratives describing historical buildings or districts in a subject's immediate environment, to spontaneous commentaries by visitors in public places that subsequent visitors can access, to auditory signals allowing blind subjects to construct effective mental maps of their surroundings. These research areas are at the frontier of computational human geography, even tually requiring a radical rethinking of traditional notions of space and place and of the relationship between humans and their environment.

Issues and Critiques

Over the years computational human geography has had its strongly committed, sometimes strident champions as well as its highly skeptical, often dismissive critics. Critiques of computational human geography originate from both within the post positivist perspective of geography as spatial science as well as from the hu manistic and social theory perspectives. The internal critiques parallel those of all computational (social) science and tend to center around epistemological and methodological issues, whereas the external critiques are more likely to contest the ontological assumptions underlying many uses of computation in human geog raphy. Critiques of computational approaches in human geography should be distinguished from those, also found in the geographical literature, that examine the role of computers in society more generally. These focus on issues of empowerment, access, privacy, surveillance, digital divides, and other societal dilemmas, often in the context of discussions of GIS and its potential uses and misuses.

Internal Critiques

The most widely accepted definition of computational science as the use of computer based methods and techniques to solve scientific problems does not do justice to the extremely significant role of computation in not just solving but also in representing problems. This is an important distinction because there is little controversy surrounding the problem solving, numerical computing aspect of computational human geography. On occasion techniques may be used incorrectly or inappropriately but beyond that, no one misses the days of the slide rule. The representational aspect of computation, on the other hand, raises a number of important issues that have yet to be properly addressed, let alone resolved.

Computation has often been discussed as the third way of doing science, lying somewhere between theory development and experimentation. This implies a new approach to knowledge production and the need for a new kind of research methodology different from either the mostly deductive mode of theoretical work or the mostly inductive mode of experimental science. That third way centers on the construction of complex simu lated worlds, within which experiments may be run that would have been difficult or impossible to conduct in the real world. Examples from computational human geography would be process oriented models of people interacting with their environment, often expressed as coupled agent based and CA models. These may simu late farmer communities dynamically affecting land cover through adaptive land use decisions, households seeking suitable housing in changing urban neighbor hoods, or trip makers responding to various congestion pricing schemes on transportation networks. Such models are built to incorporate considerable empirical and in tuitive understanding of the complex processes of interest and when calibrated to actual data they are often pre sented as suitable for policy analysis.

The epistemological problem is that models of com plex open systems with deep uncertainties, as social systems nearly always are, cannot be used for prediction. Predictive models belong in the traditional scientific paradigm of theoretical closed system descriptions sup ported by experimental evidence, or at least of well established empirical generalizations such as human geography's spatial interaction models. Because this fact is not always appreciated, many computational modelers understand progress in the field to mean building simu lations that are increasingly more detailed and realistic, though increased detail can actually decrease any pre dictive value such models may have by introducing additional 'noise'. The meaning of model validation in this new world of computational process models thus remains open, and so does the question of how to derive valid insights that may be useful for both theory devel opment and for policy guidance.

Technically, the main reason why models of complex open systems cannot yield reliable predictions is that many (in some cases infinitely many) different models can provide acceptable fits to the data; in other words, any particular model is but one realization out of a large space of potential models, few or none of which may be correct by whatever definition of the term. This issue is sometimes addressed with Monte Carlo simulations that generate many versions of a particular model by sys tematically varying the parameters; model outcomes are then considered reliable to the extent that they are re produced by large numbers of different parameter sets. This methodology may take care of parametric un certainty but cannot address structural uncertainty, that is, the degree of confidence one may have in the structure of the model. Researchers in both the social and the natural sciences have suggested methods for generating large numbers of different model structures in a manner analogous to generating versions of the same model through Monte Carlo simulations. The idea is that ex ploring the properties of entire classes (ensembles) of models, even relatively simple ones, may yield more robust insights into complex processes than the study of even the more realistic looking individual model. There is fertile ground here for the more theoretically inclined computational human geographers to make contributions that will also benefit researchers in many other fields.

External Critiques

Computational human geography has evolved out of the discipline's quantitative revolution, and the external cri tique has mirrored that development. The basic onto logical objections to representing the human world in terms of (x, y) coordinates and measurable attributes have had little reason to change since the 1960s. For much of social science human and social phenomena are funda mentally irreducible; quantitative methods can only scratch the surface of what really matters. Thus, for many human geographers computational human geography is social physics that has evolved from the age of celestial mechanics to that of growing sand piles (an oft cited example of complex system behavior in physics). If anything, the picture has only gotten worse, considering that the theory of computation is also known as the theory of abstract machines. It is understandable that those who have objections of principle to the use of physicalist and mechanistic representations of human beings may feel even more threatened by today's clever and capable computational agents than by the point particles of the quantitative revolution. Indeed, in our technology obsessed age, computation as form of power and as alluring rhetoric may present new and troubling societal risks.

Beyond the evident inadequacy of computational models as representations of individual human con sciousness is their apparent total inability to represent central social theory constructs such as gender, race, class, and society, or more elaborate notions such as the social construction of space or social/spatial justice. There seems to be no language that can effectively translate between these two domains of human geography, the computational and that of social theory. Cross fence cri tiques have substantially waned by the early 2000s, the two sides appearing to have resigned themselves to co exist in mutual disregard. Glimmers of a rapprochement are evident however in parts of the field under the banner of mixed research methods (quantitative/computational and qualitative). A particularly noteworthy development is the adaptation of GIS tools to socially engaged appli cations, as spearheaded, for example, by the public par ticipation GIS (PPGIS) agenda. Persistent major differences in philosophy make it highly unlikely that computational human geography will be accepted by more than a fraction of the discipline in the foreseeable future. On the other hand, increasing numbers of scholars from other social sciences are already adopting ap proaches and methods from computational human geog raphy in their own research.

See also: Cellular Automata; Geocomputation; GIS, Mobile and Locational Based Services; GIS, Public Participation; Spatial Analysis, Critical.

Further Reading

Bankes, S., Lempert, R. and Popper, S. (2002). Making computational social science effective: Epistemology, methodology, and technology. *Social Science Computer Review* 20, 377 388.

Batty, M. (2005). *Cities and Complexity: Understanding Cities with Cellular Automata, Agent Based Models, and Fractals*. Cambridge, MA: The MIT Press.

Benenson, I. and Torrens, P. M. (2004). *Geosimulation: Automata Based Modeling of Urban Phenomena*. Chichester: Wiley.

Beven, K. (2002). Towards a coherent philosophy for environmental modelling. *Proceedings of the Royal Society of London A* 458, 2465 2485.

Couclelis, H. (1998). Geocomputation in context. In Longley, P., Brooks, S., McDonnell, R. & Macmillan, B. (eds.) *GeoComputation: A Primer*, pp 17 29. New York: Wiley.

Gilbert, N. and Troitzsch, K. G. (1999). *Simulation for the Social Scientist*. Buckingham: Open University Press.

Gimblett, R. H. (ed.) (2002). *Integrating Geographic Information Systems and Agent Based Modeling Techniques*. Oxford: Oxford University Press.

Longley, P. A., Brooks, S. M., McDonnell, R. and Macmillan, B. (1998). *Geocomputation: A Primer*. Chichester: Wiley.

Openshaw, S. (1998). Towards a more computationally minded scientific geography. *Environment and Planning A* 30, 317 332.

Portugali, J. (ed.) (2006). *Complex Artificial Environments: Simulation, Cognition and VR in the Study and Planning of Cities*. Berlin: Springer.

Wolfram, S. (2002). *A New Kind of Science*. Champaign, IL: Wolfram Media.

Relevant Websites

http://www.casa.ucl.ac.uk
CASA: the UCL Centre for Advanced Spatial Analysis. CASA is an initiative within University College London to develop emerging computer technologies in several disciplines which deal with geography, space, location, and the built environment. A majority of CASA's publications and research projects would qualify as 'computational human geography'.

http://ccl.northwestern.edu
Northwestern University's NetLogo Models Library: this is a rich, continuously growing collection of simple computational models, many of which represent familiar geographical processes. It includes a version of the well known Schelling model of spatial segregation.

http://www.geocomputation.org
The Geocomputation Conference Series: this is a series of international research conferences initiated in 1996. This site maintains papers and other information from previous conferences and links to upcoming conferences.

Concentrated Deconcentration

B. Lambregts, University of Amsterdam, Amsterdam, The Netherlands

Glossary

Concentrated Deconcentration The spatial process whereby people, jobs, and/or other resources move away from an urban center to 'reconcentrate' in smaller centers some distance away.

Introduction

Concentrated deconcentration refers to the spatial process whereby people, jobs, and/or other resources move away from an urban center – not to spread out uniformly over space (i.e., the process known as sprawl) but to 'recon centrate' in smaller centers at some distance away. This process gives rise to settlement patterns that are best conceived as a compromise between monocentric and fully dispersed patterns (**Figure 1**). Concentrated decon centration can be linked to various spatial scales. Theo retically, these may range from the local to the global. The concept, moreover, is essentially biscalar. It always involves deconcentration at one particular level and concentration at another, inevitably lower level. These levels usually, but not always, concern the regional and the local, respectively. Concentrated deconcentration may occur spontaneously or it may be pursued through spatial and/or economic planning strategies. The latter are deliberate attempts by (spatial) policymakers to set in motion the spatial process described above. Going synonyms for concen trated deconcentration are 'decentralized concentration' and 'concentrated decentralization'.

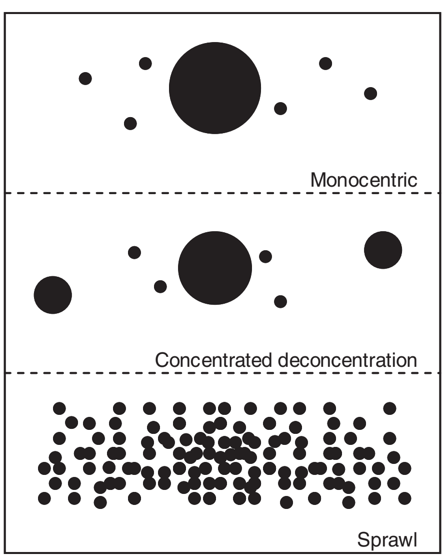

Figure 1 Alternative settlement patterns: monocentric, concentrated deconcentration, and sprawl.

Concentrated deconcentration is best known in its capacity as a planning concept. This dimension will therefore occupy center stage in this article. The next sections subsequently discuss the concept's background and origin, its various applications, its main achievements and flaws, and the ways it lives on today.

Background and Origins of the Concept

Processes of deconcentration and even concentrated deconcentration have been a feature of dynamic space economies right from the beginning. However, they be came the subject of deliberate action on a larger scale only in the course of the twentieth century, when the management of urban growth and regional development became a full grown concern of many a nation's societal and political elites. Before that, the planned movement of people and industry away from the often overcrowded nineteenth century industrial cities had already occupied the minds of some pioneering entrepreneurs and inter ested citizens. Among these was Ebenezer Howard (1850–1928), who in his famous book *Garden Cities of Tomorrow* (1898) specified in great detail how people, notably the urban working class, could be offered better living conditions by planning and developing concen trated but yet pleasantly spacious new settlements at some distance from the existing urban agglomerations. His 'garden cities' were not to be dormitory suburbs of which the residents depended for jobs on the nearby central city, but self supporting, economically viable communities that combined the key advantages of both 'the town' (i.e., accessibility to jobs and services) and 'the country' (i.e., healthy environments and space). Inter estingly and with foresight, Howard did not picture his garden cities as disconnected islands in an otherwise empty world but thought of them as constituents of larger (i.e., regional), multicentered systems of towns, cities, and villages set against a background of open space.

Although Howard's ideas were received with en thusiasm, it would take almost 50 years before the actual development of garden cities – then named 'new towns' – was taken up seriously. Basically, it was the British New Towns Act of 1946 that really set things in motion. This act paved the way for the development of no less than 20 new towns in the UK in the years following: a massive operation that would inspire planners around the world.

These included the Dutch and the French, but also the Soviets and their Eastern European 'allies', who, encouraged by strongly interventionist attitudes, did not shy away from initiating the development of hundreds of new towns, often in peripheral and/or raw material producing areas. Countries such as Japan, Brasil, Chile, Columbia, and Mexico were also among the early adaptors, and when international advisory bodies such as the World Bank and some UN divisions in the 1960s and 1970s also started to promote concentrated deconcentration as a regional development strategy, a variety of African, Middle Eastern, and (South)east Asian countries jumped on the bandwagon as well. The deliberate encouragement of regional (concentrated) deconcentration had quickly become an idea in good currency and was viewed by many spatial and economic policymakers as an effective answer to a variety of urban and regional problems.

A Multipurpose Concept

Howard had developed his garden cities model with basically one key objective in mind: to create better living conditions for the urban working class. Some 60 years later, the concentrated deconcentration strategies stemming from his model would serve other objectives as well. These can broadly be divided in two: those related to the management of urban growth; and those geared to the promotion of regional development. While these categories are not mutually exclusive (in practice, concentrated deconcentration would often be employed to serve urban growth management and regional development objectives at the same time), the distinctive features make separate treatment worthwhile.

Concentrated Deconcentration as a Growth Management Strategy

Concentrated deconcentration as a growth management tool typically relates to the scale of the city or metropolitan region. Prior to World War II, the toolboxes of urban planners were mainly filled with concentric growth models and some variations on the theme. Over time, however, the needs and desires of the (aspirant) urban population further diversified and mobility became an increasingly common good. Managing urban growth, in consequence, gradually became a more complicated task. During the first post war years, the 'normal' engines of urban population growth (i.e., births and immigration) in many parts of the world were running at full speed and produced a strong demand for housing and related services. This demand was further increased by the urban renewal programs that aimed to rehabilitate war damaged neighborhoods and address persistent inner city problems (e.g., high densities, poor services, and (consequent) substandard living conditions). These programs often involved the temporal and/or permanent displacement of inner city dwellers, which needed substitute housing in consequence. At the same time, improved transport and changing locational preferences among the more well to do urban households in particular further boosted the suburbanization process that had already started to manifest itself.

Against this background, London was perhaps the first city to become the subject of a comprehensive concentrated deconcentration policy used for growth management objectives. The now famous Abercrombie Plan for London of 1944 aimed to simultaneously reduce inner city densities, curtail metropolitan growth, and impose order to the deconcentration process that was already underway. Part of the solution was to direct London's population growth not to the city's fringes but to a number of existing country towns and almost a dozen new towns. The latter, which would eventually be realized under the 1946 New Towns Act, were to be developed from scratch in different places up to 50 km from the city's core (**Figure 2**).

Other countries were quick to adopt Great Britain's newly conceived growth management tool. The *villes nouvelles* developed around Paris in the 1960s and onward served largely similar goals as their UK examples and so did the satellite towns of, among others, Stockholm, Moscow, Hong Kong, Singapore, Tokyo, and Osaka. The concentrated deconcentration policies employed in these and other cases typically got flavored and fine tuned in accordance with local conditions and needs and the range of objectives served gradually broadened in result. Planners in the Netherlands, for example, used their combined concentrated deconcentration and growth center policy also to realize national (social) housing targets and to create favorable conditions for sustainable (i.e., public) transport solutions. The country then still faced a substantial housing shortage and its planners had already realized that uncontrolled sprawl and sustainable transport solutions do not go together very well. They figured that by concentrating the rather massive housing schemes on relatively cheap green field locations away from the cities, two birds could be killed with one stone. It was the time where spatial planning further matured as a policy discipline and optimism about the controllability of society ran high.

Concentrated Deconcentration in Regional Development Strategies

The use of concentrated deconcentration strategies in regional policies goes a long way back as well, and as in the case of growth management, the strategies over time have come to serve a variety of regional development

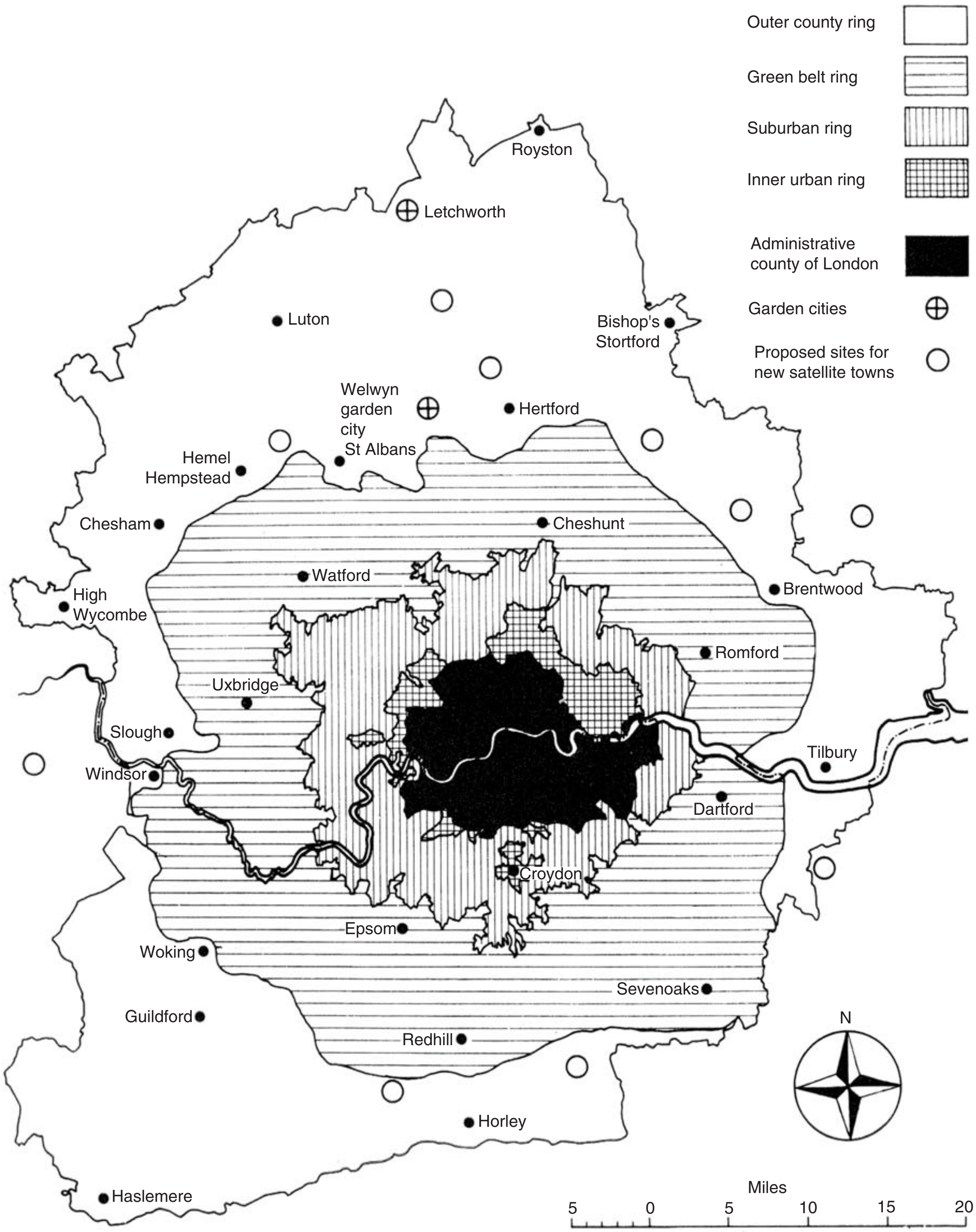

Figure 2 Concentrated deconcentration as a growth management tool: planned new towns in the Greater London area according to Abercrombie's 1944 Greater London Plan. Source: Schaffer, F. (1972). The new town movement. In Evans, H. (ed.) *New Towns: The British Experience*, pp 11 21. London: Charles Knight & Co.

objectives. These include the promotion of interregional balance and the revival of regions in decline.

Few countries are characterized by a spatially uniform distribution of resources and wealth. Core–periphery differentials are the rule rather than the exception. Throughout the twentieth century, governments across the globe for different reasons have tried to reduce such interregional disparities. They have used different

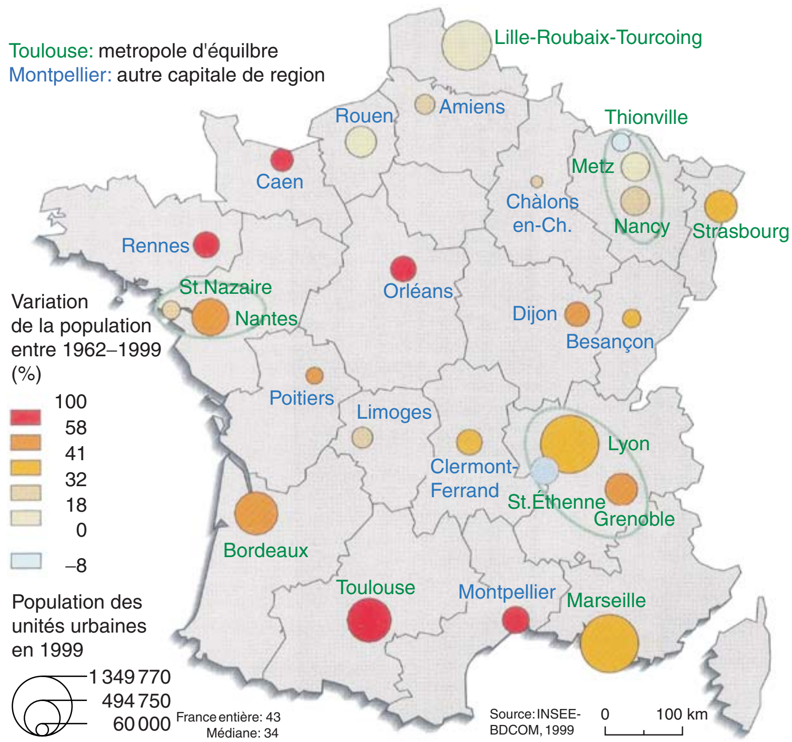

Figure 3 Concentrated deconcentration as a regional development principle: the French policy of the *metropoles d'équilibre*. Source: Reclus (2001). *Atlas de France*, *Territoire et aménagement*, Vol. 14, La Documentation française, p. 64. Taken from: http://www.senat.fr/rap/r02-241/r02-24127.html.

strategies, ranging from the provision of investment subsidies, via the orchestrated movement of public services, to the planned development of new towns or 'growth poles' in such regions. In general, such strategies sought to deconcentrate resources at the national or regional level while (re)concentrating them on a regional or local level.

For new towns or growth poles that are developed with the aim to uplift a region, economic viability is a crucial aspect. To secure economic viability spatial and economic planners often followed one of two approaches. They would either first create an attractive living environment in the assumption that economic functions would follow, or they would first facilitate the establishment of major job providing industries and only then develop the amenities and services that make a place a livable one. New centers developed according to these models have been built not only around basic industries and industrial parks as in the case of Ciudad Guayana in eastern Venezuela (metal works) but also on the basis of tertiary employment (e.g., government, education, research, and tourism). A fine example of the latter is the educational and scientific center Akademgorodok (Academy Town), which was established in the 1950s to boost Siberia's development and which involved the transfer of substantial scientific communities from Moscow and Leningrad to this formerly 'empty' quarter. Brasilia, Brazil's capital since 1960, is another good example of a new center developed for the purpose of opening up and elevating a formerly undeveloped region. In Europe, notably the French worried about the lack of interregional balance in their *territoire*. The 1947 report *Paris et le désert français* (Paris and the French desert) called attention to the rather asymmetrical spatial distribution of resources, with Paris dominating the country in virtually all respects. A more balanced development was advocated and as of the 1960s pursued through a policy which promoted the development of eight so called *métropoles d'équilibre* (**Figure 3**). Investors could expect substantial bonuses and exemptions if they chose to invest in these 'development poles'.

A special case of 'restoring' interregional balance concerns the revival of regions in decline. This became an issue for many governments especially during the second half of the twentieth century, when the ongoing automation of industrial processes and the emergence of a new global division of labor caused heavy job losses and income decline in many West European and North American regions notably. Some of these regions were able to absorb the shocks and adjust timely to the new conditions prevailing. Others, however, got caught in a downward spiral where high unemployment, low per capita income, poor educational attainments, and social deprivation reinforced each other and accumulated into a massive development challenge. Few national governments could afford to neglect these developments and extensive efforts were (and in some cases still are) made to get such regions back on track. Here again, concentrated deconcentration was sometimes used to create more competitive spatial structures for such regions. Public investments would be steered to one or more existing or newly planned centers that were thought to offer above average opportunities for some sort of economic development (e.g., tourism). It was then hoped that private investment would follow soon and provide jobs for the depressed region's workforce.

Whatever the exact motive, in (almost) all cases the application of concentrated deconcentration strategies in a regional development framework heavily depended on the capacity of notably national and regional governments to steer, amongst others by means of spatial planning, the investment flows of economic actors. More often than not, as we will see below, this proved to be quite a job.

Achievements and Flaws

Concentrated deconcentration strategies have been and still are employed in a wide variety of spatial, societal, and institutional settings and for many different purposes. Achievements and flaws are therefore not easily generalized. Yet, some general comments can be made.

Concentrated Deconcentration as a Growth Management Strategy

Did the use of concentrated deconcentration as a growth management tool relieve pressure on core cities, provide livable spaces for former and would be urban dwellers, and help to counteract uncontrolled suburban sprawl? The precise answers to these questions differ for every case, but a general outline can be given.

Ironically, the most successful feature of concentrated deconcentration as a growth management tool also caused its downfall from the late 1970s onward, especially in the Western economies. From that time it became increasingly clear that the pressure on many core cities had been taken away so drastically that the latter even started to suffer from it. The application of concentrated deconcentration as a growth management tool happened in times when deconcentration as an 'autonomous' spatial process was already underway. The two combined caused many Western cities to actually lose inhabitants. In the Netherlands, for example, Amsterdam and Rotterdam, both the subject of concentrated deconcentration policies, between 1965 and 1985 lost almost 25% of their inhabitants. The main problem, however, was that notably middle income households took the opportunity to swap their cramped city flats for the more spacious and comfortable terraced or detached houses built in the new centers outside the cities. Low income and otherwise vulnerable groups stayed behind. In their capacity as accumulations of purchasing power and taxable resources, cities rapidly became less prosperous places and when the general economic climate into the 1980s remained bleak, the concomitant problems in many cities reached serious proportions. Policymakers responded by shifting their attention toward large scale urban revitalization projects (e.g., in the UK) and 'compact' city development (e.g., in the Netherlands) and certainly in Europe the deliberate promotion of (concentrated) deconcentration in the context of urban growth management rapidly became less fashionable.

Concentrated deconcentration, in combination with some other factors, thus proved to be a threat to the vitality of many donor cities, but did it produce livable spaces for those who had left? The answer to this question again differs from case to case. For most of the newly planned centers ambitions ran high: they would offer attractive housing, good quality services, and sufficient employment for their inhabitants. In practice, however, many new centers started as bedroom towns with little employment opportunities. Over time, some of these developed into lively, versatile centers indeed. Examples include some of the French *villes nouvelles* and British new towns. Others, however, largely failed to diversify or are still in the process of doing so. Unsurprisingly, the economically successful centers are now also among the most livable ones. Yet, some predominantly dormitory towns have emerged as very popular places to live in as well, much depending on the type of houses built (owner occupied, low rise generally doing better than high rise public housing estates) and the presence of sufficient employment opportunities within commuting distance. However, there are less successful examples too. Some of the French *villes nouvelles* were among the places affected by the autumn 2005 riots (signaling that they have not evolved in too idyllic places over time) and according to a 2002 UK government report several of the English new towns suffer 'major' problems as well (e.g., high crime and unemployment levels, pockets of severe deprivation, and large amounts of the housing and infrastructure being in need of a thorough overhaul). The underlying causes include the fact that many of these new towns were built: (1) in a very short time span (meaning that many structures now simultaneously qualify for renovation); and (2) on the basis of (with hindsight) flawed planning and design principles. The fact that their speedy construction was followed by decades of relative neglect when policy attention shifted back to the core cities once again, did little good either.

A similarly mixed story, finally, goes for the question to what extent concentrated deconcentration has helped to counteract uncontrolled suburban sprawl. Sprawl has been criticized among others for causing excessive loss of open space and (transport related) energy use. Concen trated deconcentration was supposed to offer a more sustainable alternative. Where implemented, concen trated deconcentration strategies have without doubt helped to save open space, but the effects should not be overestimated. In many settings, in spite of concentrated deconcentration policies being in place, suburban sprawl stubbornly has refused to be contained. More rural and spacious living environments than those created in the new towns have continued to appeal to many middle and high income households. And local authorities, for obvious reasons, have remained willing to cater to the desires of these groups, even if this was at odds with the policies advocated by higher level planning authorities. Naturally, the rigidity of a country's planning system mattered a lot. Whereas, for example, in the Netherlands many people eventually moved to where the government said they should, the proportion of planned overspill from London accommodated by the city's new towns has been reported to be substantially less than anticipated.

Partly as the result of severe problems of measure ment, empirical research into the effects of concentrated deconcentration on transport behavior and transport related energy use has only recently gotten off the ground. The various reports are neither unanimous nor conclusive. Contextual aspects affect the actual outcomes to a considerable extent. Yet, most authors agree that concentrated deconcentration settlement patterns are more energy efficient than sprawl like patterns with the main controversy focusing on the question whether or not the former are also more efficient than monocentric settlement patterns.

Concentrated Deconcentration in Regional Development Strategies

With the use of concentrated deconcentration in regional development strategies policymakers intended to en courage economic activity and raise the level of welfare within regions. Systematic evaluations of the achieve ments in this matter are hard to come by and the high sensitivity to context specific factors again makes it hard to generalize.

In the early years of the application of concentrated deconcentration principles for regional development purposes, expectations about its effects ran high. However, already early in the 1970s it started to become clear that with some notable exceptions (e.g., some of the new towns in the UK and industrial poles in India) the actual results in many cases fell short. In the years following, many implemented concentrated deconcentration strategies were abandoned prematurely and by 1990 new regional development paradigms had taken over almost completely.

In many cases, policymakers turned to growth pole and concentrated deconcentration policies too eagerly and failed to properly analyze the exact nature of the regional problems at play. The intricacy and toughness of regional development problems on the one hand and the complicated nature of growth pole and concentrated deconcentration strategies on the other was often over looked. This resulted in many cases of failure and abandonment of the strategy. Implemented strategies could be judged unsuitable for that they did not address the actual problem(s) of the region; they could be in feasible due to the absence of enabling factors (e.g., fi nancing and effective planning powers); they could be unrealistic if they were based upon unjustified assump tions; or they could be invalidated in case they got compromised by too many, often conflicting secondary aims. The questions of how many (new) centers of what size should be planned at what distance from the existing centers and of what kind of economic activity should be promoted often posed a considerable challenge, if they were raised at all.

Most disappointing, however, was the fact that the anticipated positive external effects of the new economic growth centers in many cases failed to materialize or, worse even, were overshadowed by unforeseen negative spillover effects. The support for and the development of (new) centers of economic activity in peripheral and/or deprived regions was assumed – through spread and trickle down effects – to favor the entire region (or at least large parts thereof). The new industries would establish linkages with existing business in the region or pave the way for the establishment of such business, and the income generated by the new industries and their employees would also flow directly into the local econ omy. But in practice these anticipated benefits, if they occurred at all, were frequently neutralized or even overshadowed by negative side effects, notably by the so called backwash or polarization effects. The new, often heavily subsidized firms and industries could easily become responsible for driving out of business the firms in the growth pole's hinterland that had previously provided similar products or services to the regional market. And if the development of the growth poles went together with the construction of new extra regional transport connections (which was often the case), cheaply imported producer and consumer goods could do a lot of damage to the traditional local/regional production structure as well. Finally, if the new center(s) proved successful and created abundant job opportun ities, they could induce substantive labor migration from the hinterland to the new center. Such flows usually consisted of a region's most enterprising and prolific people and their leaving often meant a great drain on the hinterland's human capital resources. In the worst of cases, these negative spillover effects could lead to strong polarization at the regional level (i.e., great dis parities between the planned growth center and its hinterland), while planners had actually intended to mitigate such disparities (albeit at the interregional or national level).

Yet, it is not expedient to put away the use of con centrated deconcentration in regional policies as a complete failure. It has yielded results in some cases and caused troubles in others. Effects, moreover, could change over time. Positive spillover effects could dom inate during the first several years and give way to negative externalities in the years after, or the other way around. While today the majority of the professional and academic planning community has more or less dis missed the use of concentrated deconcentration and growth pole strategies for regional development object ives, the debate about its (theoretical) suitability will probably continue a bit longer.

Concentrated Deconcentration Today

Concentrated deconcentration strategies are an exponent of the conviction that people and capital can be steered and directed across space as deemed desirable. This conviction, however, is increasingly being questioned, especially in those societies where individuals and capital have a considerable say in where they wish to settle. At the same time, however, economies of scale and other externalities, in spite of the information and communi cation technology (ICT) revolution, remain important stimuli for economic agents to agglomerate in space. 'Concentration' and 'concentrated deconcentration' as spatial concepts have therefore not become fully obsolete yet, also because sustainability and the preservation of (increasingly scarce) open spaces are likely to remain on the agenda for yet some time.

These observations are mirrored by the fact that con centrated deconcentration as a spatial and/or economic planning concept is still in use today, sometimes in full view, sometimes in some sort of disguise. Concentrated deconcentration for urban growth management is still practiced notably in East and Southeast Asia, Latin America, and Africa, where cities expand at un precedented speed and careful planning is needed to channel growth in manageable and sustainable ways. China, for instance, aims to develop some 400 new towns still in the decades to come in order to accommodate the massive influx of people to the country's major cities. Some of these new developments (e.g., Dongtan) will be built on the basis of eco friendly principles. In North America and Australia, the more eco conscious and 'antisprawl' 'New Urbanism' movement may pave the way for a more sustained application of concentrated decon centration principles in those parts of the world. In Eur ope, finally, the concept secretly lives on in the 'new' concept of polycentric territorial development. This con cept is considered to offer a better fit to the increasingly complex spatial interdependencies produced by the evolving network society, but essentially still pursues a 'concentrated deconcentrated' spatial development pat tern. One of the concept's advocates is the European Union itself, which considers it a promising mean to strengthen its more peripherally located urban centers and regions and release pressure on the congested West European core area (growth management and regional development objectives combined!). But since polycentric territorial development, just like concentrated deconcen tration, is a multiscalar concept, it is possible to find ap plications at the national and regional levels as well. The áunderlying rationale in these cases is still that concen trating people and capital is to be preferred over sprawl, but having learned from past experiences, market forces are generally granted a bigger say in exactly where and to what extent concentration is to be encouraged and realized.

See also: Aménagement du Territoire: Territorial Development; Core-Periphery Models; Growth Poles, Growth Centers; New Towns; Polycentricity; Regional Development Models; Regional Development Theory; Regional Planning and Development Theories.

Further Reading

Alonso, W. (1968). Urban and regional imbalances in economic development. *Economic Development and Cultural Change* 17, 1 14.

Anderson, W. P. (1996). Urban form, energy and the environment: A review of issues, evidence and policy. *Urban Studies* 33, 7 36.

Bontje, M. (2001). Dealing with deconcentration: Population deconcentration and planning response in polynucleated urban regions in North west Europe. *Urban Studies* 38, 769 785.

Department for Communities and Local Government (2006). *Transferable Lessons from the New Towns*. London: DCLG Publications.

Faludi, A. and Valk, A. van der (1990). *Rule and Order: Dutch Planning Doctrine in the Twentieth Century*. Dordrecht: Kluwer Academic Publishers.

Gaile, G. L. (1980). The spread backwash concept. *Regional Studies* 14, 15 25.

Hall, P. (2002). *Urban and Regional Planning* (4th edn.). London: Routledge.

Hansen, N. M. (ed.) (1972). *Growth Centers in Regional Economic Development*. New York: The Free Press.

Higgins, B. and Savoie, D. (eds.) (1988). *Regional Economic Development: Essays in Honour of François Perroux*. Unwin Hyman: Boston, MA.

Moseley, M. J. (1974). *Growth Centres in Spatial Planning*. Oxford: Pergamon Press.

Parr, J. B. (1999). Growth pole strategies in regional economic planning: A retrospective view. Part 1. Origins and advocacy. *Urban Studies* 36, 1195 1215.

Parr, J. B. (1999). Growth pole strategies in regional economic planning: A retrospective view. Part 2. Implementation and outcome. *Urban Studies* 36, 1247 1268.

Reclus (2001). *Atlas de France*, *Territoire et Aménagement*, vol. 14, La Documentation française, p. 64. Taken from: http://www.senat.fr/rap/r02 241/r02 24127.html (last accessed in June 2008).

Richardson, H. W. (1976). Growth pole spillovers: The dynamics of backwash and spread. *Regional Studies* 10, 1 9.

Schaffer, F. (1972). The new town movement. In Evans, H. (ed.) *New Towns: The British Experience*, pp 11 21. London: Charles Knight & Co.

Schwanen, T., Dijst, M. and Dieleman, F. M. (2004). Policies for urban form and their impact on travel: The Netherlands experience. *Urban Studies* 41, 579 603.

Relevant Websites

http://www.englishpartnerships.co.uk
On British new towns.

http://www.villes nouvelles.equipement.gouv.fr
On French new towns.

http://www.senat.fr/rap/r02 241/r02 241/ mono.html
On French spatial development policy (in French).

http://www.newtowns.net
The European New Towns Platform.

Conservation and Ecology

M. Jay and M. Morad, University of Waikato, Hamilton, New Zealand, and London South Bank University, London, UK

Glossary

Biological Diversity or Biodiversity The diversity and variability of life at all levels from genes to populations, ecosystems, and landscapes.

Convention for Biological Diversity (CBD) An international agreement that came into force in 1993 following the United Nations conference on environment and development (the Earth Summit) in 1992.

Conservation Movement With respect to nature, a conservation movement is a social movement concerned with the protection or conservation of elements of the living environment through social or political change.

Ecology The discipline involving scientific study of the interactions of living organisms and their nonliving environment.

Ecosystem Means a system of living organisms and their nonliving environment.

Endemic As defined by ecologists and zoologists means the situation in which a species or other taxonomic group is restricted to a particular geographic region. Such a taxon is said to be endemic to that region.

Environmental Movement A social movement concerned with the protection or restoration of elements of the living and/or nonliving environment through social or political change. Environmental movements are frequently broader than conservation movements but may include nature conservation as one concern.

EU European Union.

IUCN World Conservation Union, formerly the International Union for the Conservation of Nature, an international organization combining government and nongovernment members, with its headquarters in Geneva, Switzerland.

NGO Nongovernmental organization. Examples include Greenpeace, Friends of the Earth, Conservation International, and World Wide Fund for Nature (WWF).

UNEP United Nations Environment Programme. The aims of the organization are: ''To provide leadership and encourage partnership in caring for the environment by inspiring, informing, and enabling nations and peoples to improve their quality of life without compromising that of future generations.''

UNESCO United Nations Educational Scientific and Cultural Organization. The aims of the organization are to promote international cooperation among its members in the fields of education, science, culture, and communication.

WWF An international nongovernmental organization with branches throughout the world.

Introduction

'Conservation' is a term that has been applied to multiple phenomena: conservation of nature, of matter, of energy, of art, of buildings, heritage sites, and places. Conservation of nature refers to norms, values, institutions, and social movements concerned with the protection or wise use of natural areas and living resources; and the term is largely a Western concept that has been associated with the creation and management of 'protected natural areas', including parks, game reserves, and wildlife sanctuaries. More recently, conservation has been applied to the notion of 'biological diversity' or 'biodiversity', a term referring to the variety of life in all forms and at all levels of organization.

Conservation may, but need not, involve preservation. Preservation involves maintaining an object or feature in a fixed state in perpetuity. Since most features of nature fluctuate and change, preservation is seldom an aim; rather, the aim is usually to ensure their continued existence at a socially determined level. Conservation of species for the purpose of fishing or hunting, for example, may involve the management of a target population such as grouse, trout, salmon, deer, etc. at a level where they are plentiful for hunters. Conservation of a landscape often means maintaining a particular relationship to land, for example, a relationship that preserves symbolic or iconic features in an otherwise changing landscape.

Conservation may involve structured social organizations such as government departments, nongovernmental organizations, or spatially discrete communities such as a village, as well as loosely structured, social movements or networks.

Ecology

The term ecology, coined in 1866 by the German biologist Ernst Haekel, is derived from the Greek word *oikos*, denoting home or habitat. Traditionally, ecology has often been considered a branch of biology, and historians have often argued that ecology was inspired by Darwin's theory of evolution. Life forms have been examined by biologists at various levels, from proteins and nucleic acids to cells and individual organisms. Ecologists generally engage in this type of study at the level of populations, communities, and ecosystems, or even the biosphere as a whole.

Ecology's central building blocks, ecosystems, are treated as integrated, complex units consisting of the living organisms as well as nonliving (physical forms such as energy and nutrients) objects in a given habitat. In turn, the complex relationships between ecosystems have led to the concept of the biome, denoting assemblages of interrelated ecosystems, such as desert biomes. The biosphere comprises all of the Earth's biomes where life forms can exist. The field of ecology may also be subdivided into the study of ecozones, which are generally demarcated by continental borders, and each ecozone can be further subdivided into ecoregions.

Ecological questions and scholarship have generated interest well beyond the life sciences, in fields such as epidemiology, urban planning, and human geography. Also, rising concerns for the future of the planet and natural habitats have (more recently) seen green issues at the center of wider social, economic, and political debates. In particular, the prospect of increased global warming has bestowed on these debates an unprecedented sense of urgency and mounting significance.

It is widely accepted that the birth of urban ecology in the 1920 (attributed to the 'Chicago School') has been responsible for the proliferation of human and social ecology into other disciplines, such as human geography. However, some geographers date back human geography's involvement with ecological questions to the beginning of the twentieth century, and more specifically to the publication in 1911 of *The Influences of Geographic Environment*, by Ellen Semple. The author famously claimed in her book, "Man is a product of the earth's surface."

The work of human geographers echoes and substantively overlaps with that of political ecologists, as both groups endeavor to bring together formerly unrelated areas of investigation. The catalyst to this growing body of scholarship may have been the field of cultural ecology, which examined how human cultures demonstrated high levels of adjustment as survival strategies toward their surrounding natural habitats.

Ecology and conservation have become linked in the area of biodiversity conservation and the management of valued landscapes and ecosystems.

Conservation: A Very Brief Overview

Although nature conservation has become a global phenomenon, the concept and practices are the cultural products of Western society in the late nineteenth and twentieth century. The institutions, values, and practices associated with nature conservation are almost universally influenced by a Western paradigm of scientific and rational management of dedicated spaces and political economies of power. The paradigm is one which perceives a clear distinction between the worlds of nature and humans, and which tends to value science and rationality as means of problem solving. The political economy of past and present colonial and neocolonial relations between western countries and countries of the Third World means that contemporary conservation organizations and programs, have been heavily influenced by western assumptions and values. This trend has been perpetuated through international organizations such as the United Nations Environment Programme (UNEP), the World Conservation Union (IUCN), and the World Wide Fund for Nature (WWF).

Conservation in the contemporary western world has evolved from two main roots, an elitist tradition, originating mostly in Britain and a populist tradition originating mostly in North America and Australasia. The elitist tradition stemmed from the English association of aristocracy with hunting and game management, influenced by 'enlightenment' respect for scientific approaches to knowledge and resource management. The populist tradition evolved from early romantic trends in American society, which linked wilderness to moral virtue and national identity.

In addition to the elitist and populist traditions, a distinction can be drawn between Old World and New World conservation. The colonists of North America, New Zealand, Australia, and South Africa perceived the land as largely empty. They set aside areas protected from human encroachment and thereby created landscapes that tended to be sharply divided between intensive human habitation and so called wilderness. Old World countries in contrast have been more concerned with conserving traditional landscapes against the encroachments of modern forms of development. Thus, the practices of conservation in Europe contrast significantly with those in North America, Australia, and New Zealand. Conservation in Old World countries was more gradual and mixed in terms of aims and processes, incorporating concerns for cultural as well as natural heritage. In the New World, conservation was more directly focused on the protection of native fauna and flora with little or no regard for the cultural heritage or possible rights of native people. The creation of parks and reserves was largely a matter for negotiation between different settler interest groups in the face of rapid environmental change and destruction of native ecosystems and species.

Conservation in the United States

Concern with the conservation of nature and wild areas began in the US during the nineteenth century. In 1836, Ralph Waldo Emerson published '*Nature*', in which he articulated the importance of wilderness as a source of moral and spiritual uplift. His ideas influenced Henry David Thoreau (1817–62), an eloquent and passionate advocate for nature and a key inspirational source of American environmentalism. The 1850s, 1860s, and 1870s, following the exploration and opening up of land west of the Mississippi River, brought a surge in the number of travel books and romantic literature by adventurers, explorers, and artists. They conveyed notions of romantic beauty and heroic grandeur linked to wild and empty spaces. A series of government sponsored survey reports of the west were published, and included artistic illustrations, bringing the scenic and natural wonders to the attention of eastern Americans.

In 1864, George Perkins Marsh published *Man and Nature: or Physical Geography as Modified by Human Action*, a work which strongly influenced intellectuals and policy makers of the day. Having traveled widely and noted the widespread ecological damage suffered in the Mediterranean, Marsh highlighted the ravages wrought on the natural world by development and warned of the need to protect natural resources. By the 1870s, public support for preserving areas of wilderness was firmly in place. Yellowstone National Park, the world's first, was created in 1872, and by 1892 John Muir had founded the Sierra Club, one of the oldest conservation organizations in the world. In 1897, the New York based Wildlife Conservation Society founded the New York Bronx Zoo.

American conservation was given a boost in 1901 when Theodore Roosevelt became President. Roosevelt believed strongly in the moral benefits of wilderness contact and recreation. He considered that it encouraged such manly qualities as courage, endurance, and self reliance. Furthermore, he believed that the land itself embodied the ideals of American independence, resourcefulness, and democratic freedom.

Contemporary responsibility for conservation in the US is distributed between numerous federal and state agencies. The Fish and Wildlife Service, as part of the Federal Department of the Interior, is responsible for federal wildlife laws, endangered species, migratory birds, nationally significant fisheries, and the management of national wildlife refuges. The Parks Service, a separate department within the Department of the Interior administers national parks, historic sites, recreation areas, and scenic rivers. States have their own departments and programs. For example, the State of New York has a Department of Environmental Conservation with a broad environmental mandate that includes conservation of open space and the creation of an open space network. The State of Washington, on the other hand has a Conservation Commission with no operational responsibilities for public land but with a mandate to provide leadership and advice.

Nongovernmental organizations within the US include agencies that are both large in terms of their American membership and massive on a world scale. American conservation/environmental organizations have had a major influence on conservation elsewhere in the world both as direct contributors to conservation programs and in shaping policies on conservation by international organizations such as the IUCN and the World Bank.

Conservation in the UK

Early conservation practices in Britain were largely related to the protection of habitat for game birds, deer, trout, and other species for hunting by the aristocracy. Populist concern for conservation began toward the close of the nineteenth century when groups such as the Commons Preservation Society, founded in 1865, were formed to preserve commons and countryside access for the public. In 1895, the National Trust for Places of Historic Interest or Natural Beauty ('The National Trust') was formed to promote the preservation of – and public access to – buildings of historic or architectural interest and land of natural beauty. A nongovernmental organization dependent on membership subscriptions and donations, it has grown to become one of the largest conservation organizations in Britain.

While populist concern for conservation was widespread at the beginning of the nineteenth century, the effort and concerns of voluntary organizations were resisted by entrenched interests such as the owners of large estates, competing government departments, and local county councils. Only in 1949 was a National Parks Commission formed and even then, the powers of the commission were feeble; they could designate national parks and provide advice, but administration was under the jurisdiction of county councils. Nevertheless, over the following decades the commission made a significant contribution to the development of protected areas in England and Wales. In addition to the commission, a nature conservancy was established with a mandate (and the funding) to identify and administer land as 'national nature reserves' or sites of special scientific interest (SSSIs) for protection of wildlife. Despite these measures, wildlife continued its steady decline.

Agricultural intensification during the 1950s, 1960s, and 1970s brought the destruction of many treasured elements of countryside landscape. The realization grew that it was not sufficient to preserve islands of habitat surrounded by ecological deserts of farmland; in order to protect species and biological diversity, it was necessary to preserve or restore ecological processes across whole landscapes. In 1968, the mandate of the National Parks Commission was broadened to advocate for conservation throughout the countryside (not just in national parks) and it became the Countryside Commission (and, after successive restructurings, Natural England under the Department of Food, the Environment and Rural Affairs). The Countryside Commission introduced the concept of Environmentally Sensitive Areas and the Countryside Stewardship Scheme (later the Environmental Steward ship scheme), to encourage farmers to conserve valued elements of the landscape and compensate them for loss of production.

Voluntary organizations have had major influence in terms of general public opinion and political pressure on local and central government officials. Such organizations include the Royal Society for the Protection of Birds, the Royal Society for Nature Conservation (formerly the Society for the Promotion of Nature Reserves), the National Trust, Fauna and Flora International, and na tional branches of Greenpeace, Friends of the Earth, and the WWF. Like the American organizations, they have had major influence on conservation in other parts of the world.

Conservation in Australasia

The first European settlers of Australia encountered an ecology that was entirely different from anything they had experienced before: a land of thin soils, drought, erratic rainfall, strange plants, and radically different fauna (e.g., marsupials and the strange duck billed platypus and spiny echidna). Royal societies and natural history societies sprang up almost in tandem with settlement. By the 1890s, every Australian colony sup ported a natural history society, royal society, or both.

Concerns for conservation began soon after settlement in response to mass slaughter of native mammals by commercial hunters and trappers and to the damaging impacts of European style development on native fauna and flora. In 1881, the New South Wales Zoological Society persuaded the state government to pass the Bird Protection Act and in 1883 the Queensland Royal Soci ety campaigned for the creation of national parks. The Royal Society of South Australia lobbied for the creation of national parks in the late 1880s and in Western Australia, the Natural History Society successfully campaigned for fauna and flora reserves.

From the 1920s to the 1950s, conservation in Australia was overshadowed by events such as the Great De pression, the two World Wars, the Korean War, and a prevailing development focus. By the beginning of the 1960s there was growing concern at the loss of native species, coupled with a greater appreciation of native fauna and flora as symbols of Australian identity. The Australian Conservation Foundation, as one of Australia's largest voluntary organizations, was founded in 1964 with the aim of protecting Australian rainforests, interior drylands, and the Great Barrier Reef. During the 1970s, conservation groups sprang up in opposition to logging of state forest, coastal sand mining in Queensland, uranium mining in northern Australia, and hydrodevelopment in Tasmania. As well as the activism of populist conser vation groups, conservation by Australian government agencies also increased from the 1970s onward.

A significant Australian initiative in the final decades of the twentieth century was the land care movement. While the primary aims of land care groups are land health and sustainability, many have discovered that long term sustainability under Australian conditions may require a return to native trees and deep rooted native grasses as an alternative to more productive but less ecologically resilient introduced pasture grasses. More recent official conservation concern, at both federal and state levels, has focused on implementation of a national biodiversity strategy. Published in 1996, the strategy has prompted numerous action plans and reports, but evi dence of effective change on the ground is still to come.

Conservation under the European Union

Conservation of nature in the European Union has de veloped from early efforts to limit transborder environ mental impacts, to a region wide concern with the protection of valued fauna, flora, and traditional rural landscapes. The 1972 Helsinki Convention on the Baltic Sea and 1976 Barcelona Convention on the Mediterra nean Sea, for example, were efforts to reduce the cross boundary effects of agricultural and industrial pollution by countries surrounding these water bodies.

A Europe wide perspective became more evident with the 1979 Bird Directive, which aimed at long term preservation of all bird species in the European Union. The directive identified endangered bird species and made member states responsible for designating special protected areas. Similarly, recognition of the importance of habitat for species conservation became evident with the 1992 Habitats Directive, which aims to protect wildlife species and their habitat. Member states were required to identify areas of special conservation interest and draw up management plans for their protection.

Over the 1990s, European agricultural policies began to address the damaging effects of agriculture on native wildlife and traditional agricultural practices. Changes were introduced to the Common Agricultural Policy (CAP) to enable member states to administer agri en vironmental policies that would encourage farmers to protect wildlife habitat and landscapes or features of natural or cultural value. In 2000 the program, Natura, was introduced as the focus of a common European conservation policy. Natura 2000 aims to create a net work of protected areas to conserve vulnerable species and habitats across their natural range irrespective of national or political boundaries. It links to other policy instruments, most particularly financial support through the EU Rural Development Regulation.

Unlike the North American commitments to 'wilder ness' conservation, European approaches are concerned with conserving a heritage of species and landscapes which has been constructed and managed over gener ations; the landscape is thus a manifestation of human history in a way that North American, Australian, or New Zealand wilderness is not. In these latter, parks and re serves are the creation of the colonial immigrants, not the indigenous people, and do not generally reflect the his tories of the settler society.

Conservation in Nonwestern Countries

Moving to countries formerly under British colonial rule in Africa East, Central and Southern Africa, conservation was strongly influenced by the British tradition of game protection for the elite, in this case, officers of the British civil service and their visitors and families. It focused on the protection of game animals and their habitat and was largely a response to uncontrolled slaughter by early explorers, traders, and settlers. By the beginning of the twentieth century, game reserves had been established in Sudan, Kenya, Uganda, and the British colonies of central Africa. The reserves were subject to strict regulations about hunting and in many cases indigenous people and their animals were barred from entering.

With the transfer of government away from Britain as countries became independent, conservation priorities changed and in many cases, reserves became protected areas in name only. The fate of many of the reserves created during the British colonial era continues to be uncertain, particularly given the population pressures that many of the former colonies face. Many of the re serves are subject to uncontrolled poaching or agri cultural encroachment.

Game tourism has become an important source of foreign exchange for some countries, however. Countries such as Kenya, Tanzania, and Uganda have significant tourism industries that depend in substantial part on the wildlife present in public and private game reserves.

India, as another country that experienced British colonial rule, was also influenced by British ideas and policy. At the beginning of independent rule in 1947, India inherited Corbett National Park, a popular pre independence hunting spot for the British (named after Jim Corbett, a tiger hunter turned conservationist). In addition, there were areas such as Keoladea National Park, a wetland sanctuary for birds created by the Maharaja of Bharatpur, Bandhavagarh, hunting grounds of the Maharaja of Rewa, and Bandipur, hunting grounds of the Maharaja of Mysore. Since Indian independence, many of the former hunting areas have been turned into national parks or wildlife sanctuaries, and new areas have been established. Problems with funding, population growth, conflict between people and wild animals, and development of land and water resources for agriculture have meant that many of India's parks and wildlife sanc tuaries are under pressure. Pressures include poaching, removal of trees for firewood, and lowering of water tables through irrigation of adjacent lands, primarily for agriculture.

New conservation partnerships are developing through out the Third World, however, between western based and international organizations and host countries. WWF, for example, runs a tiger conservation program that has involved partnership with governments and local organ izations in India, China, Bhutan, and Southeast Asia. Conservation International, in collaboration with govern ment and local farmers in Ghana, has encouraged farmers around Kakum National Park to grow cocoa without cut ting down forest canopy. The UNEP through its World Conservation Monitoring Centre together with the IUCN and the World Commission on Protected Areas have de veloped an important coordinating role for identifying protected areas and providing information that can assist with their management.

The Rise of Global Conservation Institutions

The rise of global conservation is largely a post World War II phenomenon and it is possible to discern two main stages in its philosophy and development. The first stage, from the early and mid 1960s to the early and mid 1980s could be characterized as an age of dawning realization of the global scale of environmental problems and the global interconnectedness of problems. The second stage, from the early and mid 1980s to the present has been gradual acceptance that conservation of nature cannot be separ ated from other aspects of human ecology; it is inextric ably linked to socially equitable sustainable development.

Global Conservation Stage 1: A Growing Awareness of Interdependence

The first truly international organization for conservation of nature was the International Union for the Conser vation of Nature, now known as the World Conservation Union (IUCN). It was formed in 1948 with support from UNESCO. Intended as a clearinghouse for government agencies and private national and international organiza tions, it was a blend of government and nongovernment organizations that became headquartered in Gland, Switzerland.

It was not until the 1960s, however, that widespread awareness of the global scale and interconnectedness of conservation and environmental issues began to be ap preciated. By 1960, certain types of habitat had already experienced worldwide depletion – for example wetlands – or damage as a consequence of industrial development – for example, the forests of Scandinavia and eastern Canada damaged by acid rain caused by industrial emissions from countries to the south. A key event included the 1962 publication of the book *Silent Spring*, by Rachel Carson, which clearly identified links between the decline of bird populations and application of pesticides. This insight brought home to many, an awareness that the chain of life is interconnected. Following growing environmental awareness in the 1960s, the 1970s experienced a number of important events, as indicated in **Table 1**.

Global Conservation Stage 2: Paradigm Change: Conservation Depends on Social Justice and 'Sustainable' Development

Through the 1990s and the new millennium, conservation organizations have increasingly come to accept that conservation of nature/biodiversity depends on the maintenance of ecosystem processes and critical interdependencies between plants, animals, and humans. In addition, they have come to realize that successful conservation requires community acceptance and involvement at local levels. If local people do not or cannot provide the conditions for other species to survive, national and international policies ring hollow. More and more, organizations such as Fauna and Flora International, WWF, and the Wildlife Conservation Society operate in association with national and local groups. There has been a paradigm shift, from a narrow focus on species (tiger, big game) 'by government decree', to programs that involve local communities in ways that will assist community members to sustain a livelihood.

The Contribution of Human Geography

Writings on conservation of nature/biodiversity by human geographers have been rich and various. Many have been flavored by an underlying normative concern for protection of nature, or, conversely, for social justice in the face of western style conservation solutions to nonwestern countries and cultures. Attempting to summarize the diversity of scholarly writing on issues of nature conservation and biodiversity is difficult, but a number of themes can be identified. They include empirical studies, critiques of conservation, examination of the epistemologies of conservation, the social construction of conservation places, the political ecologies of conservation, and the complexities of integrating conservation with social and cultural sustainabilities at local level.

Empirical and comparative studies of conservation organizations, policies, and programs have provided comparisons of conservation programs in different countries and the manner in which conservation policies and programs 'mesh' with other socioeconomic and political arrangements, including the values, concerns, and circumstances of local communities. They include analyses of the relationships of groups involved in conservation at local, national, and international levels and the role of international organizations and NGOs.

Critiques of conservation have attempted to address the questions: Who gains and who loses from conservation programs? How are the losers from conservation affected and how do they respond? How are national or international elites implicated in the creation of protected areas? Are the advantages of conservation equally shared? The studies have particularly focused on political ecologies of neocolonialism and globalization, which involve the imposition of conservation constraints on local communities by national or international elites. Such analyses have revealed the often unequal spread of costs and benefits from conservation.

Epistemologies and discourses of conservation have examined the ideas, assumptions, and knowledge of actors involved in conservation, including scientists, national policy makers, national and international conservationist organizations, and local communities. While many conservation organizations are prepared to acknowledge the value of local ecological knowledge, criteria for international funding frequently depends on scientific assessments of ecological value. Conservation management policies may depend on scientifically based proposals, which are problematic or inappropriate for local circumstances. Resistance to western scientific views of conservation may come from groups who perceive nature and their relationships to nature in ways that are at odds with those of the western scientific community. For example, while an organization such as the EU with its Natura 2000 program is clearly focused on creating a network of key sites for 'biodiversity' as identified by scientists, local communities may view the same landscapes as icons of national identity, as heritage, or as backward areas that need to be developed.

Social and cultural geographers have focused their attention on the social construction of conservation/biodiversity/protected areas; and these studies examine the ways in which conservation and/or biodiversity come to be construed by different groups and how protected areas are constructed from the collection of different (often divergent) interest groups involved. For example, in relation to biodiversity, groups vary on what they consider relevant elements of biodiversity (do worms, fungi, or bacteria count?). In addition, there is no consensus on the importance of different types of biodiversity (are key species such as tiger more important than ecosystems such as wetlands?), the kinds of places that deserve protection (should protected areas include privately owned land, or only public land?), and

Table 1 Important events: stage 1, 1970s and stage 2, 1980s and 1990s

Event	*Comment*
Stage 1	
1971 Ramsar convention	An international agreement, which provides the framework for conservation and wise use of wetlands.
1972 UN Conference on the Human Environment, Stockholm	The first international conference on environmental issues, organized by the UN Economic and Social Council. Representatives attended from 113 countries and 400 intergovernmental and nongovernmental organizations. Preparatory meetings before the conference identified environmental issues more clearly, including the differences between countries, and the overriding importance of development for poor countries.
1972 Paris Convention for the Protection of the World Cultural and Natural Heritage (under the auspices of UNESCO).	A convention that established the system of world heritage sites as cultural or natural features of outstanding universal value.
1972 publication of the Club of Rome report, *Limits to Growth* by D.H. Meadows, D.L. Meadows, J.Randers, and W.W. Behrens.	The book highlighted the implications of uncontrolled growth on renewable and nonrenewable resources, and the finite capacity of the earth to absorb wastes.
1973 United Nations Environment Program (UNEP)	Founded as an outcome of the Stockholm Conference on the Human Environment. The mandate of UNEP was to coordinate the development of environmental policy by keeping on monitoring the global environment and bringing emerging issues to the attention of governments and the international community.
1973 Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES)	Intergovernmental agreement to ensure that international trade in wild animals and plants does not threaten their survival.
1979 Formation of Greenpeace International	A voluntary environmental organization that was particularly effective in drawing media attention to the slaughter of whales.
1980 Publication of the World Conservation Strategy by IUCN and UNEP	One of the first publications by conservationists to recognize that conservation of nature is inseparable from social and economic sustainability and cannot be achieved without development to alleviate the poverty and misery of millions of people. It was the first major public document to introduce the term 'sustainable development'.
Stage 2	
1987 Publication of *Our Common Future*	Report of the United Nations World Commission on Environment and Development, otherwise known as the Brundtland report. The report was highly influential in its articulation that development should be 'sustainable' and that sustainability involves both intra- and intergenerational equity. With its unapologetic insistence on sustainability, no matter how vaguely defined, the report provided the basis for a changing world paradigm for conservation and environment and development.
1991 Global Environment Facility (GEF)	A fund established which helps developing countries to fund projects and programs that protect the global environment. In 2002, projects funded by the facility included biodiversity, climate change, land degradation, and the effects of pesticides.
1992 UN Conference on Environment and Development (Earth Summit), Rio de Janeiro	The Earth Summit was the largest gathering of heads of state, government officials, international organizations and nongovernment organizations held to that date. It provided the impetus for the Framework Convention on Climate Change, the Convention on Biological Diversity, and Agenda 21, a statement of an international program of action for the twenty-first century.
1993 the Biodiversity Convention came into force.	The convention aimed to ensure conservation of biological diversity by member states, sustainable use of biological resources and equitable sharing of benefits from genetic resources. As a consequence, a large number of countries have since produced strategies for conservation of their biological diversity and the term 'biodiversity' has become widely used.
1993 creation of the World Conservation Monitoring Centre (WCMC) by UNEP	An important facility for assessing the progress or otherwise of conservation efforts.
2005 The UN releases the report of the Millennium Ecosystem Assessment	The report by 1300 experts from 95 countries concluded that approximately 60% of the earth's ecosystems were being degraded or used unsustainably.

forms of management (how much should harvesting be allowed? should human habitation and production be included?).

One significant recent development prompted largely by the contributions of human geographers has related to the changing paradigms and a conservation of hope. These writings have noted the changing influence of concepts such as 'sustainability' on conservation object ives, and the ways in which the paradigms of western based conservation have been changing in response to global concerns such as the United Nations Millennium Development Project and its call to reduce absolute poverty and ensure environmental sustainability. The literature includes a normative concern to 'decolonize' conservation and to reduce the opposition between protected nature and spaces of human habitation that has frequently infused the language of both conservationists and scholars.

Some analysts have noted that protected areas do not always involve displacement or loss for local people; in some cases, wilderness can provide habitat both for in digenous people and for wild nature. A trend of the past two decades has been the intermixing of habitat pro tection with other forms of land use to create 'multi functional' rural landscapes. This is certainly clear in the EU, where agri environmental programs allow farmers to be compensated for costs involved in conservation of wildlife habitat or landscapes of cultural value. Concepts of 'ecological corridor' and habitat 'stepping stones', introduced from conservation biology and landscape ecology, have encouraged visions of landscapes which allow human and nonhuman species to coexist in shared space. The paradigm is of humans in nature, rather than humans at war with nature, or humans rescuing nature, or humans exploiting nature.

A problem for geographers interested in nature con servation is that the term 'nature' has been eclipsed or conflated with 'environment', and 'environment' has been appropriated by the hegemonic discourses of environ mental management and environmental science. Critical or reflective geographers concerned about conservation of nature or biodiversity are almost immediately trapped by the close historical association between protection of wildlife and habitat by elite groups and the dispossession of the poor, the indigenous, or the minority from tradi tional relations with their natural environment. Geo graphers thus have no vocabulary to describe or explain their interest without inviting the accusation that by doing so, they uphold colonial, neocolonial, or capitalist systems of oppression.

This tension is unfortunate because in the discourses about conservation of nature and biodiversity, human geographers have special insights to offer. Theirs is an awareness of whole system interactions; of the intimate co creation of landscapes by humans and nonhumans; of the importance of human subjectivities, and an understanding of the ways that power relations struc ture spatial organization and the allocation of resources.

In *Spaces of Hope*, Harvey (2000) explains that in our contemporary world, the rich are becoming more affluent and multinational corporations are gaining increasingly greater powers; but much of the world's natural resources and landscapes are being damaged and depleted. Harvey argues that there is nothing par ticularly new about this, but the absence of political will to do something about this constitutes the greater threat. In order to make unavoidable changes to redress the catastrophes we face, Harvey contends that hu manity needs to become the 'architects' of a different environment, spanning the meeting points between the personal and immediate scales (of the 'body') and the global political economy. This and other similar inputs from human geographers provide a compelling argu ment that humanity can draw inspiration and insight from utopian visions. If any vision of the environ ment (as 'spaces of hope') is to work, the argument goes, it must take account of our human qualities. Harvey's point is that human geographers should not shrink from the messy business of engaging with such major contemporary issues as environmental damage and biodiversity loss. 'Spaces of hope' require such engagement.

Examining the geographical literature of the past two decades on conservation, it is clear to see a rising interest in the political, social, and environmental implications of nature conservation – frequently one and the same as biodiversity conservation – and a rising effort to identify pathways of hope. Such pathways are offered as an al ternative to the bitter pessimism of environmental de struction that has characterized conservation thinking for so long. Much of the pessimism remains – the powerful social forces of environmental destruction continue apace. However, as the dynamic of environ mental destruction gains wider recognition, there is also the realization that humans are unavoidably in and of nature, and must find ways to reconcile human aspir ations with ecosystem limits. This realization is a source of hope.

See also: Biodiversity; Biopolitics; Colonialism, Internal; Culture/Natures; Deforestation; Empire; Environmental Justice; Environmental Studies and Human Geography; Environmentalism; Globalization, Cultural; Heritage and Culture; Heritage and Identity; Imperialism, Cultural; Indigenous Knowledges; Land Change Science; Nature, Historical Geographies of; Nature, Social; Nongovernmental Organizations; Political Ecology; Rio Summit; Sustainable Development; Third World; Wilderness.

Further Reading

Adams, W. M. (2003). *Against Extinction, the Story of Conservation*. London: Earthscan.

Adams, W. M. and Mulligan, M. (eds.) (2003). *Decolonizing Nature: Strategies for Conservation in a Post Colonial Era*. London: Earthscan.

Brechin, S. R., Wilshusen, P. R., Fortwangler, C. L. and West, P. C. (2003). *Contested Nature: Promoting International Biodiversity Conservation with Social Justice in the Twenty First Century*. New York: State University of New York Press.

Harvey, D. (2000). *California Studies in Critical Human Geography, 7: Spaces of Hope*. Berkeley, CA: University of California Press.

Hutton, D. and Connors, L. (1999). *A History of the Australian Environment Movement*. Cambridge: Cambridge University Press.

IUCN and WCMC (1998). *1997 United Nations List of National Parks and Protected Areas*. Gland: IUCN Publications.

IUCN UNEP WWF (1991). *Caring for the Earth, a Strategy for Sustainable Living*. Gland: The World Conservation Union, United Nations Environment Programme, and World Wide Fund for Nature.

Library of Congress *The Evolution of the Conservation Movement, 1850 1920.* http://memory.loc.gov/ammem/amrvhtml/conshome.html (accessed June 2008).

Millennium Ecosystem Assessment (2005). *Ecosystems and Human Well Being: Biodiversity Synthesis*. Washington, DC: World Resources Institute.

Millennium Ecosystem Assessment (2005). *Living Beyond Our Means, Natural Assets and Human Well Being*. Washington, DC: World Resources Institute.

Southworth, J., Nagendra, H. and Munroe, D. K. (2006). Editorial, introduction to the special issue: Are parks working? Exploring human environment tradeoffs in protected area conservation. *Applied Geography* 26, 87 95.

Zimmerer, K. S. (2006). Cultural ecology: At the interface with political ecology the new geographies of environmental conservation and globalization. *Progress in Human Geography* 30(1), 63078.

Relevant Websites

http://www.conservation.org
Conservation International.

http://www.foei.org
Friends of the Earth International.

http://www.greenpeace.org
Greenpeace International.

http://www.iucn.org
IUCN: International Union for the Conservation of Nature.

http://www.nationaltrust.org.uk
The National Trust.

http://www.nature.org
The Nature Conservancy.

http://www.unep.org
United Nations Environment Programme (UNEP).

http://www.wcs.org
Wildlife Conservation Society.

http://www.iucn.org
World Commission on Protected Areas.

http://www.unep wcmc.org
World Conservation Monitoring Centre.

http://www.iucn.org
World Conservation Union.

http://www.wwf.org
World Wide Fund for Nature.

Consumption

D. Leslie, University of Toronto, Toronto, ON, Canada

Glossary

Actor Network Theory (ANT) A body of theory associated with science studies and, in particular, the work of Bruno Latour, John Law, and Michel Callon. Actor network theory involves a material-semiotic method. The aim is to understand how people, ideas, technologies, and nature form networks. Relations within the network are not conceptualized as pre-existing, but rather are seen as precarious, emerging within the course of network interactions, and requiring constant repetition.

Commodity There are a variety of uses of the term commodity, including objects of value, things that are produced for the purpose of being exchanged, and objects that are exchanged for money. In a society driven by the production of commodities, an entrepreneur chooses both the commodity to be produced, and also the method by which it will be manufactured according to expectations about its salability on the market and ability to generate profits.

Commodity Chain A set of interrelated activities (such as production, consumption, design, retailing, marketing, and retailing) necessary for the production of a particular product or service. The concept is often associated with the global commodity chain (GCC) literature and in particular the work of Gary Gereffi. In this conceptualization, the economy is understood as an assemblage of individual commodity chains. The idea is that each commodity chain has a different form of organization, as well as different geographies and power relations.

Commodification The extension of the commodity form to goods and services which previously were not commodities as defined above. There are different meanings of commodification depending on how a commodity is defined.

Commodity Fetishism The notion of commodity fetishism emerges in the writing of Marx in 'Capital Volume I' and refers to the process whereby in the marketplace, commodities appear to have value in relation to each other (or in relation to money). The true source of their value – human labor – is not visible. Consumers perceive commodities as natural, as having a life of their own, rather than as socially produced.

Flexible Accumulation A term stemming from French Regulation Theory (associated with the work of Aglietta and Lipietz among others) which refers to the relationship between production and consumption which began to take shape after the early 1970s. Flexible accumulation is associated with flexible or just-in-time production systems, flexible labor markets, and more segmented and differentiated consumption patterns and is also referred to as 'post-Fordism' or 'reflexive accumulation'.

Governmentality The literature on governmentality stems from the work of Michel Foucault and is concerned with how government seeks to cultivate certain forms of behavior among individuals, populations, firms, and economies. Governmentality emphasizes the types of governing authority, the forms of knowledge, techniques, and the entity to be governed, as well as how it is conceived, and the consequences of governance. Power is not determined *a priori*, but is the result of a constantly changing assemblage of techniques, practices, and rationalities.

Material Culture A body of work which explores the relationship between physical objects and social relations. It draws upon a range of disciplines, including anthropology, archaeology, design studies, history, and geography. A key theme in the literature is that material objects are shaped by human actions, but individuals gain habits, ethics, and values through their interactions with particular material environments and objects.

Postmodern There are two main uses of the term postmodern. First, the concept refers to a 'postmodern' era characterized by the dominance of signs and symbols over use values. Some commentators read postmodernism as the cultural logic of late capitalism or flexible accumulation. A second use of the term is implied by the concept of postmodernism as an artistic or theoretical shift. Here, postmodernism refers to a set of aesthetic practices associated with irony, pastiche, and eclecticism and a blurring of the divide between high art and popular culture. Postmodernism also denotes a set of theoretical ideas demonstrating skepticism toward modernist universalisms, metanarratives, and grand theory.

Introduction

In recent years, there has been a growing interest in the dynamics of consumption in contemporary capitalism. This interest is reflected in the publication of a number of popular books on global branding and consumption such as Naomi Klein's 'No Logo', as well as an explosion of academic works on the topic. Geographers have turned their attention to consider spaces and scales of consumption, as well as issues of identity and ethics in consumption. Analyses have focused on a range of commodities, including food, clothing, furniture, and flowers.

Why this growing interest in consumption? In part, the attention to shopping is related to a broader 'cultural turn' in economic geography, which aims to bring together economy and culture, production and consumption, and the material and the symbolic. It is also tied to the shift toward a more flexible and reflexive regime of accumulation, according greater attention to the role of demand in innovation. An emphasis on consumption is a manifestation of growing concerns related to the safety, security, and sustainability of food and other commodities, as well as a generalized disillusionment with the growing expanse of consumer culture, including the diffusion of consumer subjectivities into political practice, education, the academy, and healthcare. For yet other more cynical analysts, the increasing interest in consumption relates to the exigencies of academic publishing and grant writing, which place an increased premium on 'newness', 'fashion', and 'marketability' in academic research.

This article reviews the geographical literature on consumption, paying special attention to the way in which consumption has been defined and theorized historically. Spaces of consumption will also be highlighted, along with the politics of consumption.

Defining Consumption

Consumption typically refers to the selection, purchase, use, maintenance, repair, and disposal of goods and services. Under capitalism, consumption is primarily oriented around the consumption of commodities. Commodities are objects or services which have a use value (or function) and an exchange value (in that they are exchangeable for money or for other goods through barter).

Capitalism is associated with the commodification of more and more aspects of social life. Virtually, every human activity in western, and increasingly nonwestern societies, relies on or has commodities associated with it, including marriage, birth, death, and war. Most recently, genes, embryos, and knowledge have all gained status as commodities. Yet, despite the growing trend toward commodification, consumption maintains a significant independence from capitalist relations; it takes place at a different time and place from commodity production, and it offers ways of decommodifying objects produced and exchanged as commodities in other sites. As Andrew Sayer (2003: 341–357) puts it, "Objects which begin their life as commodities do not necessarily remain commodities at all moments in their existence."

Consumption overlaps and is connected to other spheres of economic activity. Marx, for example, illustrates that the moment of production is simultaneously an act of consumption in that it involves the consumption of labor power and other commodity inputs. Similarly, the production of labor power entails the consumption of food and other commodities and is therefore also an act of consumption. At the same time, the work of consumption is an act of production. In consuming food, for example, the human being actively produces the physical body.

Commodities not only perform a useful function in the reproduction of the human being, but also hold particular associations or meanings for users. As such, commodities hold value within a system of other objects and have a sign value. Discussions of consumption such as those forwarded by Jean Baudrillard, stress the disconnection between use values and sign values in the contemporary or postmodern era, and the growing importance of sign values relative to use values. Suggests that the Baudrillard surface of the commodity has become detached from the commodity; this simulation of commodities has become its own reality or hyper reality. For Baudrillard consumption of signs of commodities has become more important than the consumption of commodities themselves. However, the primacy of sign values is widely contested in the literature on consumption, with many theorists maintaining that use values remain pivotally important in practices of consumption.

The consumption of goods and services is bound up in rituals and social relationships, including relations of inclusion and exclusion. Material goods are not only things, but rather they convey meanings of class, gender, and style. Consumption plays a role in the production of identities, and in the maintenance of individual narratives of the self. Most significant in this regard is the work of Pierre Bourdieu. In his book 'Distinction', Bourdieu illustrates how consumption choices are tied to status rivalry and class, and signify position in a social field.

However, consumption is not a wholly individualistic and self indulgent act. There are a variety of motives for consumption, ranging from selfish, egoistic ones to benevolent and altruistic aims. Consumption can be tied to external imperatives such as the pursuit of status, but it is also practiced for its own sake and in light of internal motivations. Recent studies demonstrate how consumption is often pursued with the interests, needs, and advice of others in mind, and in the context of relationships of caring and love. Anthropologist Daniel Miller has made a particularly important contribution to the field of consumption studies by highlighting the social and moral dimensions of shopping, much of which is directed toward family members and friends. Thus, rather than being a momentary and individual act of purchase, consumption is a social process whereby people relate to artifacts in complex ways, transforming the meanings of commodities dramatically as they incorporate them into their lives.

Theorizing Consumption

Given its significance in everyday life, it is perhaps surprising that until recently, geographers and social theorists have tended to ignore consumption, focusing instead on the realm of production. When consumption did enter the equation, as in the Frankfurt School, it was generally condemned as a realm of uniformity, falsity, and meaninglessness. Writing in the 1940s, authors such as Adorno and Horkheimer describe how consumption practices are increasingly dictated through the machinery of capitalist cultural industries, promoting a homogenized and materialistic American culture. For the most part, consumers are conceptualized as passive dupes, seduced by the schemes of retailers, marketers, and advertizers, and politically pacified through the consumption of useless objects and mass entertainment. For these thinkers, the dominance of consumption diminishes the possibility for freedom and individuality. The meanings and pleasures of consumption and the subjectivities they mobilize are seen as fundamentally comprised and inauthentic. These early discussions of consumption focused overwhelmingly on the negative dimensions of consumption, and are underpinned by an entrenched belief in the monolithic and determining nature of capitalist transformation. Ironically, thinkers on the right (such as F. R. Leavis) have historically been prone to a similar tone when writing of mass consumption. Concern is expressed, for example, regarding the loss of community and tradition, and the rise of hedonism.

Partly as a reaction to this bleak portrait of consumption, a new literature emerged in the 1970s and 1980s, which was much more optimistic in its outlook. This literature grew out of the work of a number of theorists based at the Birmingham Centre for Contemporary Cultural Studies (such as Stuart Hall). For these authors, moralistic pronouncements such as those of the Frankfurt School are based on elitist prejudices. Geographer Peter Jackson, for example, was quick to suggest that these critiques privilege academic knowledge over popular wisdom, showing little respect for the political judgment or moral integrity of ordinary consumers. Jackson claims that accounts often display a distaste for low (and female) culture and are blind to the multiple and complex meanings that shopping, commodities, and popular culture are capable of offering to individuals. In contrast to the conception of 'consumer as dupe', this new literature highlights the increasingly knowledgeable ways that buyers engage with consumer culture. Influenced by this broader shift, geographers began to produce a lively literature foregrounding the capacity of consumers to appropriate the schemes of commercial captors to their own advantage, creating diverse subjectivities. The emphasis is on the critical skills of consumers, the practices by which commodities are endowed with oppositional meanings, and on the ways in which consumption is utilized as a form of creativity and resistance.

Such interventions were not immune to criticism, however. For many social scientists, the pendulum had swung too far in the opposite direction, with a tendency to overcelebrate consumer agency and to ignore the control of producers. Still others argued that the celebration of everyday acts of subversion was obscuring the more genuinely radical potential of consumption. Today, a more balanced perspective has emerged in the discipline exploring the plurality of modern technologies of consumption. Influenced by Michel Foucault's theories of governmentality, this literature foregrounds how consumer conduct is mobilized, motivated, and distributed, and how consumer freedom and choice itself can operate as an effective form of power and governance. Modes of governance work through the capacities and freedoms of the governed in order to secure the ends of government. Foucault explores the way power enables as much as it constrains. Emphasis is placed on the way in which commercial institutions frame the subject of consumption in nonreducible ways, on how technologies of consumption mobilize consumer conduct by forging links between human motivations and the specific features of commodities. As Paul duGay has recently argued, the focus is not on producers or consumers but on the relation between the two, on the devices used to promote consumption, and on the dispositions of consumers. Consumers do not pre exist descriptions of them, but are enacted and performed through various discourses. In this way, regimes of governance do not determine forms of subjectivity, but rather they elicit, promote, and facilitate particular qualities in agents. They are successful to the extent that agents come to experience themselves through such capacities. The aim is not to establish a single theory of consumption but to trace the ways in which geographically and temporally specific modes of consumer conduct are mobilized. While consumers are governed, there can be moments of problematization where the techniques, knowledge, languages, and practices of governance are called into question by authorities (such as corporations) and the governed (in this case consumers). This outline of theoretical traditions is overly simplified. However, in one form or another, these various narratives continue to animate the stories we tell concerning consumption in the discipline.

Spatializing Consumption

Drawing on these theoretical insights, geographers have added a distinctly spatial dimension to the study of consumption. They have explored different scales of consumption – from the global to the local. Operating at the global scale, much work has been devoted toward an understanding of the globalization of commodity culture. Here, there has been an attempt to understand global flows of commodities such as McDonald's hamburgers, Coca Cola beverages, and GAP clothing. For many commentators, the diffusion of these brands leads to a process of cultural and economic homogenization, often referred to as a 'McDonaldization' of culture. However, geographers have been much more cautious and critical of such overarching claims. Dispelling many of the myths of globalization, they have been quick to cite the resilience of local cultures in a transnationalizing world, foregrounding the production of new forms of local culture and new meanings of the 'local'. Research has illustrated how goods assume different meanings in different places, and has illuminated the difficulties of forging global markets. As one example, global companies are often forced to customize their brands to appeal to local markets, repudiating generalized claims regarding the emergence of global brands.

Landscapes of consumption, and the power relations that characterize them, have also been an important focus of attention. This work has become all the more critical in the early twenty first century, as landscapes of consumption have become more ubiquitous and cities themselves have come to resemble theme parks and shopping malls, with their carefully designed streetscapes, and privately maintained security forces. More and more, urban economic development agendas focus on marketing the city as a site of entertainment, consumption, and culture, in order to attract tourists, talent, and investment. Geographical work traces the control exercised in the layout and design of shopping malls, department stores, high streets, and festival marketplaces. Much of this literature has explored the construction of gendered and class based identities in these spectacular sites, but also acknowledges the pleasure and autonomy that women, youths, and seniors derive in these quasi public spaces.

In recent years, there has been an increased focus on the embeddedness of commodities within more mundane spaces such as the second hand store or car boot sale. Geographers such as Nicky Gregson and Louise Crewe have drawn attention to the need for ethnographic research on what people actually do with their purchases, on how they transform commodities through repair, restoration, and alteration, and on the ways in which they use, reuse, and dispose of goods. The home has also assumed greater importance due in part to the expansion of e commerce and home shopping. While geographers have long been interested in the consumption of housing itself as an example of a positional good, they are beginning to explore the dynamics of consumption within the space of the home. The home affords an interesting opportunity to examine the relationship between people and goods because it is a space of both individual and collective consumption, where the goods purchased and the meanings assigned to them are negotiated between household members.

Finally, geographers have also explored how consumer discourses map the body including its size, shape, health, and sexual attractiveness. Discourses shape the body's spatiality, comportment, and motility and define its possibilities and limitations. Practices of discipline, surveillance, and self restraint form an important part of everyday life. Food consumption, in particular, illustrates the intimate connection between consumption and the body. In the case of food, the commodity is ingested and quite literally becomes part of the body. The corporeal nature of food consumption is illustrated in the growing concern over food safety and quality, and the heightened sense of public mistrust of scientific and government knowledge and regulations governing food. Food consumption patterns have long term consequences for health and mortality; inequalities in access to food based on class and geography have become an urgent concern for researchers.

Consumption incorporates a series of practices located in particular places, but it also serves to connect consumers with actors in other sites and locales. As Andrew Sayer notes, consumption encompasses a range of social and material relationships between the producer and consumer, between the purchaser and consumer, between a consumer and other consumers (joint consumers or third parties), and between consumers and other individuals who may have claims to the same resources. Consumption also entails a relationship between the consumer and the natural environment. A single act of purchase may have different impacts in terms of these various relationships and sites. It remains important to unravel the complex geographies of consumption.

Geographers have been at the forefront of a group of scholars calling for work which traces the entire trajectory of a product from its conception and design, through its production, retailing, and final consumption. This is referred to as a vertical approach to consumption because it highlights the uniqueness of individual commodities which are characterized by a distinct temporality and spatiality. This specificity has implications for chain dynamics and power relations as well as consumption practices, ethics, and politics. Such an approach differs from a horizontal analysis of consumption emphasizing the broader characteristics of consumption in a particular place or gendered and ethnic dimensions of consumption.

The introduction of a vertical heuristic underscores not only the need for multiple approaches to consumption, exploring the specificity of individual commodities, but also the overlap between commodities and commodity chains.

One useful contribution of the vertical approach has been to highlight the way individual commodity chains are organized and governed. A central theme in this literature relates to the different types of governance structures associated with commodity chains. According to Gary Gereffi, there are two distinct types of governance regulating the dynamics and power relations of commodity chains. The first encompasses producer driven commodity chains which are characteristic of large integrated and capital intensive industries such as aerospace and automobiles, where producers play a dominant role in design and in the production system. The second type of governance structure refers to buyer driven commodity chains, where large retailers and brand name marketers play a key role in the commodity networks. These power relationships are common in labor intensive, consumer goods industries such as garments, footwear, toys, and consumer electronics. Gereffi argues that, in recent years, we have seen a shift from producer driven to buyer driven commodity chains. Instead of having to put considerable time and money into finding markets for goods and persuading consumers to buy them, Gereffi suggests that retailers and marketers have become more demand led. While the binary between producer and buyer driven chains underestimates the interdependence between sites on a chain, and the ways in which different types of power may be exercised at different nodes, these discussions have advanced our understanding of the influence and control exercised by a range of intermediaries along a chain.

There are a number of different approaches to commodity chains, and it is not the author's aim to review them here. One important point, however, concerns the political value of tracing these connections. Much of the commodity chain literature is underpinned by a political economy perspective, emphasizing systematic inequalities between core and periphery in global commodity systems, and the flow of economic surplus across space. Often, the aim is to penetrate the fetishism of a particular commodity, to delve below the fantastic images associated with the product in the realms of advertising and retailing, and to trace the value of a commodity back to the hidden abode of production. For many analysts, commodity chain analysis offers the possibility of defetishizing commodities. This involves looking beyond the exchange of commodities, to the value created in producing the commodity, in order to fully understand and appreciate what a commodity is. This allows one to view commodities as thoroughly embedded in social and geographical processes, the idea being that increased education may lead consumers to question their role in perpetuating landscapes of exploitation.

Other analysts remain skeptical of the linear nature of such explanatory sequences, preferring to see chains as having no beginning or end. Inspired in part by actor network theory, this literature explores the interdependence between people, firms, states, and organizations in commodity networks. Agency is viewed as relational; nodes as such do not exist as concrete objects, but rather are constituted within the networks they are part of. Seen from this vantage point, commodity fetishes are not neatly woven veils masking the origins of consumer goods. Furthermore, there is no guarantee that simply providing additional knowledge to consumers about the production of goods will automatically lead to changes in consumer behavior. Relations between actors are prone to rupture and recombination in everyday consumption. The goal then is not to thicken surfaces, but to think constructively about how surfaces are productive of particular identities and relations.

A final point relates to the role of the commodity itself in these networks. Inspired by both actor network theory and work in anthropology on material culture, there is a growing interest in how the commodity itself is an important actor in commodity networks, along with other nonhuman actors such as technologies and nature. There has been increased attention to the role of objects and the ways in which they allow humans to order society in more durable and mobile ways. Material goods, buildings, and machines are not mere outcomes of human action: they actively participate in processes of interaction. In this vein, material objects and commodities can be considered as actors within commodity networks.

Politicizing Consumption

For Noel Castree, recent work on the geography of consumption displays a tendency to avoid normative judgment. Castree argues that there is a need for a more systematic consideration of how to weigh the 'goods' and 'bads' of consumption, along with a way of evaluating the values, practices, and impacts of consumption upon people within and between locales. He suggests that consumption researchers frequently adopt a situated and reflexive voice in order to challenge authoritarian standpoints such as those of the Frankfurt School. However, Castree worries that this 'modest' voice obscures the need for academics to take a stand, to offer principled arguments and suggestions in particular circumstances. While acknowledging the problematic nature of imposing normative standpoints *a priori* and the complexity of normative thinking, it is crucial to offer criticisms of consumption and commodification based on reasoned and case specific arguments. He raises issues surrounding the audience of consumption studies, suggesting that consumption researchers need to be more explicit about whom they are addressing in their work in order to reflect more explicitly on the political implications of their work.

All of this foregrounds the politics of consumption itself. In an influential 1995 polemic, Daniel Miller suggests that consumption is now the vanguard of history and the fulcrum of dialectical contradiction. He argues that on the one hand, it appears as one of the key social issues of our time, responsible for massive suffering and inequality. On the other hand, consumption is also at the heart of potential political movements. Miller casts the consuming housewife rather than the working class male as a progressive force of the twenty first century. Through the sheer authority of her collective decision making power, he argues, the consumer holds the potential to turn domination into empowerment. While not entirely optimistic that this power will be realized, Miller argues that if consumers could mobilize and channel this power, they could mount an effective challenge to the irresponsibility and amorality of the market.

While geographers studying the political economy of capitalism have typically not viewed it as a potential site of struggle, they are beginning to explore the political possibilities of this space. In recent years, there has been heightened attention to a range of social justice movements such as cooperatives, consumer boycotts, green consumption, the slow food movement, and organic and fair trade consumption. These forms of activism cannot simply be dismissed as liberal or individualistic; for many encompass radical and collectivist forms of struggle. Furthermore, these movements speak to groups historically marginalized in production based struggles such as women and youth. Gender bias and a moral distaste for shopping have obscured this radical potential. The global mobility of capital makes struggles at the point of production increasingly difficult to wage. There is, therefore, a need to consider new sites of organizing outside the workplace.

There are of course many limitations associated with consumer based movements – few represent a genuine resistance to capitalism. While there have been selective improvements in working conditions in certain industries, ultimately systematic inequalities are not addressed in many ethical consumption initiatives. In the case of organic food consumption, the impact on the environment may be positive, but geographers have shown that working conditions can be just as exploitative in organic agriculture as in the dominant food economy. Others have suggested that many ethical campaigns are merely marketing initiatives designed to reassure concerned consumers. These limitations, however, should not blind us to the potential of these movements, and to the need to establish a politics that can challenge global capitalism from a variety of vantage points and spaces, including consumption.

Conclusion

From a marginalized field of study, consumption has emerged as one of the most important terrains of enquiry and political practice. However, consumption still needs to be taken seriously in many branches of the discipline, particularly in economic geography where there is still a tendency to prioritize production. It is really only in the last few years that economic geographers have begun to emphasize the role of demand in innovation systems. This is particularly important since a growing proportion of profits are derived from the stages nearest to the customer, and innovation in customer service is now critical to firm success. The customer is increasingly the focus of innovative and entrepreneurial activity. Sophisticated users are viewed as being actively involved in evaluating, conceiving, and modifying products, particularly in specialized product categories such as sporting goods and computer games. The focus on the customer shifts the emphasis beyond the region or industrial cluster in the attempt to understand innovation systems, since customer relationships often occur at a distance. The increased focus on users in innovation is a much needed corrective to the earlier tendency to ignore the role of consumption.

In sum, consumption studies are coming of age, leaving behind earlier simplifications regarding agency and determination, and examining the geographical complexities in the way consuming subjects are constructed and governed. Geographers have drawn attention to the diverse motivations that underpin consumption and the multiple linkages that exist between economic spheres, as well as between consumers and other actors, objects, and nature. While processes of commodification have become increasingly prevalent and insidious, geographers have come to recognize the political possibilities of mobilizing at the site of consumption.

See also: Commodity Chains; Corporate Responsibilities; Food Networks; Global Commodity Chains; Material Culture.

Further Reading

Adorno, T. and Horkheimer, M. (2002). The culture industry: Enlightenment as mass deception. In *Dialectic of enlightenment: Philosophical fragments*, Cumming, J. (trans.), pp 94 136. New York, NY: Stanford University Press.

Baudrillard, J. (1998). *The consumer society: Myths and structures*. London: Sage.

Bourdieu, P. (1984). *Distinction: A social critique of the judgment of taste*. London: Routledge.

Castree, N. (2004). The geographical lives of commodities: Problems of analysis and critique. *Social and Cultural Geography* 5(1), 21 35.

Cook, I., Crang, P. and Thorpe, M. (2004). Tropics of consumption: 'Getting with the fetish' of 'exotic' fruit? In Hughes, A. & Reimer, S. (eds.) *Geographies of commodity chains*, pp 171 192. London: Routledge.

duGay, P. (2004). Devices and dispositions: Promoting consumption. *Consumption, Markets and Culture* 7(2), 99 105.

Fine, B. and Leopold, E. (1993). *The world of consumption*. London: Routledge.

Gereffi, G., Korzeniewicz, M. and Korzeniewicz, R. (eds.) (1994). *Commodity chains and global capitalism*, pp 247 265. Westport, CN: Greenwood Press.

Glennie, P. and Thrift, N. (1992). Modernity, urban and modern consumption. *Environment and Planning D: Society and Space* 10(4), 423 443.

Gregson, N. and Crewe, L. (2003). *Second hand cultures*. Oxford: Berg.

Jackson, P. (1999). Commodity cultures: The traffic in things. *Transactions of the Institute of British Geographers* 24, 95 108.

Klein, N. (2000). *No logo. Taking on the brand bullies*. Toronto: Vintage.

Maskell, A. and Power, D. (2005). On the role of global demand in local innovation processes. In Fuchs, G. & Shapira, P. (eds.) *Rethinking regional innovation and change*, pp 273 290. New York: Springer.

Miller, D. (ed.) (1995). *Acknowledging consumption: A review of new studies*, pp 164 203. London: Routledge.

Miller, D. (1998). *A theory of shopping*. Cambridge: Polity.

Nava, M. (1992). *Changing cultures: Feminism, youth and consumerism*. London: Sage.

Sayer, A. (2003). (De)commodification, consumer culture, and moral economy. *Environment and Planning D: Society and Space* 21(3), 341 357.

Wrigley, N. and Lowe, M. (eds.) (1996). *Retailing, consumption and capital: Towards the new retail geography*, pp 48 67. Harlow: Longmans.

Relevant Websites

http://www.behindthelabel.org
Behind the Label.

http://www.cleanclothes.org
Clean Clothes Campaign.

http://www.fairtraderesource.org
Fair Trade Resource Network.

http://www.globalvaluechains.org
Global Commodity Chains, Duke University.

http://www.nologo.org
Naomi Klein.

http://www.nosweat.org.uk
No Sweat.

Content Analysis

J. Baxter, University of Western Ontario, London, ON, Canada

Glossary

Code (noun) A theme or category used to group text units by similar meaning; (verb) the process of ascribing text units to themes or categories.

Codebook A list of codes and their definitions for categorizing communication messages.

Inter-Rater/Coder Reliability A measure of the degree to which different coders apply codes in the same way to a range of text units. Typically, a correlation coefficient above 0.8 is considered good.

Message Any communication between humans – for example, a newspaper article, interview, or speech.

Text Unit A subunit of a larger text – for example, paragraphs, lines, words.

Unit of Analysis The unit actually measured, typically the text unit.

Unit of Observation The unit about which study conclusions are drawn – typically, it is the same as the unit of analysis, but sometimes it is something the unit of analysis represents – for example, newspapers, political parties.

Introduction – Content Analysis in the Social Sciences

Content analysis is a social scientific methodology for making sense of recorded human communication including news media, policy documents, letters, and even video or novels – particularly, written texts. In practice, content analysis involves breaking down texts into smaller units: paragraphs, sentences, phrases, or single words. These units are then grouped according to common meaning – codes. This methodology is typically used to make at least three types of inferences; those about: the producers of the texts; the messages in the texts themselves; and the audience for those texts.

Though this methodology was formalized around the 1920s, it grew rapidly during World War II in communications research to study propaganda in the media. This era of popularity continued through the following decades as the quantitative revolution took hold across the social sciences. Though the methodology was heavily developed as a quantitative method during that era, in the decades since the quantitative revolution the term 'content analysis' has been used in much broader terms. The main distinction is between a primarily quantitative and qualitative version of content analysis as 'methodology' – with version specific ontological and epistemological underpinnings – and not more simply as a technique or a method. Thus, not only is content analysis unique in this regard, it is paramount to understand how the term is being used in practice/context.

Early definitions of content analysis are still heavily cited in the social sciences and these emphasize positivistic underpinnings, including quantification, replication, and objectivity. Some of the more recent definitions acknowledge the inherent ambiguity of objectivity, in particular that this implies essentialist notions of meaning in text – that meanings are singular and fixed over time and space – something most geographers likewise no longer accept. Thus, these more modern definitions emphasize the 'validity' of inference rather than the repeatability of inference.

Despite such advances, most books on content analysis almost exclusively discuss content analysis as a positivistic, quantitative endeavor that involves counting the occurrences of predefined, though admittedly, theoretically informed codes. Yet, some of the most cited authors acknowledge the notion that content analysis can also be qualitative. Though the books by those authors still mainly concern quantitative implementations of content analysis, they have left the door open for the development of content analysis as a qualitative practice. For the qualitative content analyst, the presence or absence of a phenomenon, however arrived at, is what is important, not the number of times it occurs in a text or group of texts. What follows is a more detailed discussion of these two very different forms of content analysis.

Quantitative Content Analysis

Many maintain that, in the strictest sense of the term, content analysis refers to its quantitative form. Some key characteristics that comprise this form include adherence to the scientific method, the message as the unit of analysis and/or observation; phenomena must be counted; and explanations are nomothetic. Among the characteristics of the scientific method that many content analyses sourcebooks cite are that a study must be: objective, based on an *a priori* research design, reliable, replicable, valid, generalizable, and deductive. For the quantitative content analyst, the messages, not the audiences for or producers of those messages, are supposed to be the unit of analysis and/or observation. That is, there must be some 'communication' between individuals or groups in a content analysis study, even if the conclusions concern people/groups. Thus, not just texts are open to study by the content analyst, any form of communication is available for study. While the requirement that the phenomena must be counted is rather straightforward, the nomothetic mode of research is not. The latter refers to research that seeks out broad, generalizable, parsi monious, explanations of a wide array cases, not just a specific case or a few cases.

The typical process for conducting a quantitative content analysis should be familiar to any social scientist who understands the hypothetico deductive process in cluding the following ordered steps: theoretical rationale, conceptualization, operationalization, *a priori* develop ment of code definitions, sampling, application of code definitions to communication messages, tabulation and statistics, reporting, and conclusions. An example from geography concerning the relationship between distance and bias in newspaper reporting on spotted owl habitat protection will be used to demonstrate each step.

Theoretical Rationale

As with all deductive research, the process starts with a theoretical rationale, specifically a relationship between concepts that is to be tested empirically. In the case study, one general class of relationships tested is the variation over space in the coverage of an environmental issue. Specifically, two hypotheses (among others) are tested: (1) newspaper coverage concerning logging and the protection of spotted owl habitat decreases with physical distance from the habitat forests in question – a distance decay effect of 'newsworthiness' and; (2) reporting bias likewise varies systematically with physical distance.

Conceptualilzation and Operationalization

The conceptualization and operationalization process helps link theory with the available data. In our example study, 'physical distance' is measured as the number of kilometers between the city which publishes a news item on spotted owls and the nearest city which is contained within the owl's USA habitat range (i.e., the northwest). The 'amount of newspaper coverage' is measured in two ways: the 'number of stories' and the 'average story length in words'; whereas the 'bias in coverage' concept (i.e., toward logging forests within the range of the spotted owl habitat) is measured in two ways: as pro cut and pro save sources, and pro cut and pro save themes. Thus, the study tests whether the independent variable 'physical distance' predicts each of these four dependent variables.

A Priori Development of Code Definitions

This stage requires that a list of code definitions – a codebook or dictionary – is developed prior to coding. That is, the strictest practitioners will not read their data prior to creating their codebook, rather such researchers will refer primarily to extant theory. There are pre existing dictionaries, such as the Harvard and Lasswell general purpose dictionaries which, because they are standardized, facilitate cross study comparisons. Two key aspects of quantitative coding are that there can only be one code per text unit and that each code must be unique from all other codes – mutual exclusivity. This seems a rather strict requirement and would frustrate most qualitative analysts, but it is considered necessary by purists in order to meet the basic assumptions for stat istical analysis. In our example, if a passage (message text unit) concerns mainly a 'source', it may only be coded as pro cut source; pro save source; or neutral/unknown source, and nothing else. Thus, very strict decision rules must be put into place ahead of time. At this stage initial tests of inter rater or inter coder reliability of human coding may be calculated to ensure that definitions can be consistently applied by different people. This is es sentially the correlation across each of the test coders' applications of the codebook to the same data. Most authors agree that a correlation coefficient of 0.80 or higher is indicative of a reliable coding scheme.

Application of Code Definitions to Communication Messages

There are two ways to apply codes – using humans or using a computer. A potential source of confusion is that both may actually be facilitated by computer. Typically, codebooks are used by humans, whereas dictionaries are used by computer programs and both have the goal of ensuring consistent application of the codes. Though this seems like a purely mechanical process, there is fre quently some code refinement during this phase whereby preliminary reliability checks are made and the coding scheme refined. Unlike hermeneutic qualitative content analysis which may require special knowledge of social theory in the area being studied, quantitative content anlaysts advocate that ideally, once the definitions are established, human coding requires no special training as a social scientist, merely training about the specific codebook. There are numerous software packages which facilitate computer dictionary based coding (e.g., Ver batim Blaster) as well as human coding (e.g., NVivo). While it is typically beyond the scope of most empirical publications to include the codebook or dictionary, even as an appendix, a good research paper will provide at least one example of how a key coding choice was made.

Sampling

Since a census of all communication messages is often impractical (e.g., all newspaper articles on spotted owls), typically a theoretically relevant sample of messages is selected. The sample may be narrowed through random selection of elements on such dimensions as time period, place, and source. The example study is narrowed to ten daily newspapers from the USA during a 5 year period from 1990 to 1994. The rationale is that the issue peaked in this period and the daily papers they picked had cir culations that were between 300 000 and 600 000 to control for known bias effects from very large circulation newspapers.

Tabulation and Reporting

Many of the texts concerning content analysis involve detailed discussion of the various ways the data may be presented quantitatively and these mirror procedures for secondary data analysis in general methods texts. Data may be presented simply, as frequencies or in more so phisticated terms as relationships between variables in bivarate tests and multivariate models. In the case ex ample, histograms show frequencies of stories, average words per story, and bias ('themebalance' and 'source balance') with the category axis representing the news papers ordered by distance from the spotted owl habitat (**Figure 1**). Though the predicted distance decay pattern does appear for the 'number of stories' and somewhat for 'average length of stories', no support is found for a spatial dimension to 'source' or 'theme bias'. This con clusion is reinforced with a multivariate regression an alysis (not included here) showing that physical distance is a statistically significant predictor of 'story frequency' and 'average story length', but not of 'source' or 'theme bias'. Thus, the conclusion is that distance does not predict reporting bias, but it does predict the amount of newspaper coverage.

Qualitative (Hermeneutic) Content Analysis

Though there are plenty of standardized procedures for quantitative content analysis formalized in several text books outside of geography, far less is written about qualitative content analysis – anywhere. This may come as a surprise to many geographers, since indeed this raises questions about the appropriateness of using the term 'content analysis' in reference to qualitative re search. Several researchers claiming to conduct content analysis might instead use terms such as interpretive

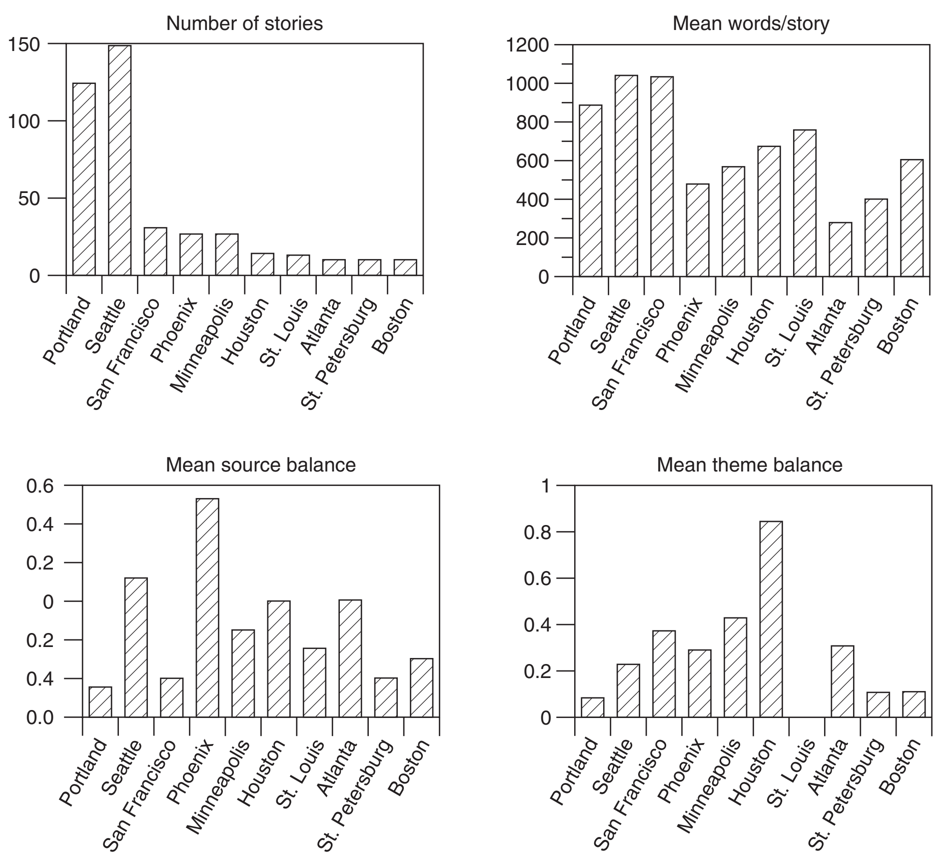

Figure 1 Example summary data from a quantitative content analysis study. Reproduced Bendix, J. and Leiebler, C. (1999). Place, distance, and environmental news: Geographic variation in newspaper coverage of the spotted owl conflict. *Annals of the Association of American Geographers* **89**(4), 658 676 figure 3, with permission from Blackwell.

analysis, hermeneutic analysis, ethnographic analysis, discourse analysis, grounded theory, critical analysis, or semiotic analysis. Part of the problem may stem from a simple misunderstanding of terminology, but equally likely is that qualitative content analysis is a legitimate and unique methodology which simply shares epi stemological and ontological elements of its qualitative cousins just listed. It is difficult to tell which is the case though, since journal articles rarely allow for an adequate elaboration of methodology.

Many of the books on (quantitative) content analysis discuss qualitative content analysis in rather pejorative terms. One of their critiques against qualitative message analysts is that the latter tend to overemphasize the creators and audiences of messages (texts) rather than the messages themselves. Indeed, this is viewed by the qualitative practitioners to be one of the major strengths of their approach – that the research lens is appropriately focused on the people involved in message creation and con sumption since these involve power relations (**Table 1**). Some argue that the differences between quantitative and qualitative content analysis are at least partially a matter of degree – for example, that qualitative content analysts focus on deep (latent) meaning, whereas quantitative analysts focus more on surface (manifest) meaning. It is more often the case though, that these practitioners advocate funda mentally different approaches to social science. The following is a list of characteristics that, though they may be present in quantitative content analysis they are, nevertheless, 'most commonly' found in qualitatively oriented content analysis:

- focus on theory development (induction) not just theory testing (deduction);
- researcher is the instrument, not the codebook;
- skilled social scientists are required to do the work;
- latent content (between the lines; 'hidden' meaning) is more important than manifest content;
- more flexible coding rules;
- attention is paid to how a message is produced in the first place (e.g., interviews);
- researcher is often a co creator of the messages (e.g., interviews);
- inferences often concern producers of messages and/ or the intended audiences;
- inductive development of concepts (codes);
- use of quasi quantitative language (e.g., 'repeatedly', 'often', 'rarely') rather than precise numbers (e.g., '80% of the time');
- identification of the presence/absence of a concept rather than precise frequencies;
- multiple content types (e.g., newspapers, policy docu ments, archival material) considered simultaneously, each with potentially different text unit sizes;
- heavy use of example quotations from the data; and
- attention to valid inference at the expense of large sample size and strict *a priori* coding rules.

By way of summary, **Table 1** highlights some of the strengths and limitations of both quantitative and quali tative content analysis.

Many researchers today now advocate for the com bination of methods where appropriate, and increasingly we see content analysis studies that use both quantitative and qualitative methodologies in the same study.

Table 1 Relative strengths of quantitative and qualitative content analysis

Strengths	*Quantitative*	*Qualitative*
Reliable	More[a]	Less
Transparent	More	Less
Generalizable	More	Less
Nonreactive/unobtrusive	More	Less
Inexpensive	More	Less
Wide spatial and temporal coverage	More	Less
Valid	Less	More
Methodologically flexible/ adaptable to research situation	Less	More
Findings relevant to resolving a specific problem	Less	More
Connection to/engagement with those studied	Less	More
Conditions of document creation are clarified	Less	More
Useful for answering 'why' questions	Less	More
Takes advantage of researcher-as-instrument	Less	More

[a] Each strength is considered to be a continuum whereby 'more' and 'less' concern characteristics of each content analysis methodology in relation to the other.

Content Analysis in Human Geography

How human geographers use the term 'content analysis' varies. Further, a survey of geographic literature on the topic leaves the distinct impression that content analysis has fallen between the cracks of quantitative and quali tative geographic research. Current texts on geography methods and some dictionaries of geography have either no entry whatsoever concerning content analysis or only brief mention. In addition, entries for the term 'text' likewise make no/scant mention of content analysis. In stead, entries for 'text' are more likely to mention, and discuss at length other, explicitly qualitative, interpretive analytical techniques such as semiotic analysis, discourse analysis, landscape analysis or cultural analysis. This may signal at least two possibilities. One is that geographers eschew content analysis as a worthwhile form of human/ social research. Another possibility is that nobody is really sure what it is that content analysis is meant to do for geographers. The answer seems to fall somewhere between these two explanations.

For example, as further indication that this method ology is marginalized in geography, when the term 'content analysis' does appear in geography methods texts, less than a page and little elaboration are typically provided. Further, the passages in those texts tightly align with the notion, and somewhat pejoratively so, that content analysis is solely a quantitative endeavor.

Since content analysis is typically discussed in texts on geography methods in the sections on qualitative meth ods rather than the quantitative sections, it is perhaps unsurprising that this methodology receives such short shrift. Nevertheless, this underscores the unique position of this methodology as one that can be implemented quantitatively or qualitatively.

Despite the dearth of material in geography methods texts, geographers do actually use the term 'content analysis' frequently in their research and writing. For example, a search of the Geobase geography journal database (between 1986 and June 2006) for 'content analysis' anywhere in the title or abstract produces over 750 results. Once references from physical science/ physical geography journals are eliminated, about half of those references relate to what is discussed here and the majority of those are empirical studies. A more precise search of three influential geography journals – the *Annals of the Association of American Geographers* (AAAG), the *Professional Geographer* (PG), and the *Transactions of the Institute of British Geographers* (TIBG) – over the shorter period from 1996 to June 2006 produces a considerable total of 16 empirical articles five, nine, and two, respectively. It seems content analysis is alive and well in geography. What seems to separate geographers from the early, history of content analysis in the social sciences is that the communications being analyzed are often 'created' by the research process itself. That is, the 'texts' in much of the empirical work by human geog raphers come from interviews and focus groups that were led by the researchers themselves – they are par ticipants in the data creation. While this would certainly offend the purist quantitative content analyst on the grounds that it violates objectivity, such an approach to data 'creation' receives ample epistemological and ontological support from qualitative researchers across the discipline.

Debates

There are at least four important debates concerning content analysis that are worth highlighting. These reflect general methodological debate in social science and human geography.

What Constitutes Objectivity?

At least two related issues are worth considering. First, while some quantitative content analysts endorse that a hallmark of objectivity is defining codes prior to even looking at the message data; other quantitative content analysts advocate that an inductive reading of the mes sages to develop codes is entirely acceptable. The qualitative content analyst typically prefers the latter approach since theoretical development is typically their goal anyway. Second, most current discussions of ob jectivity center on the notion of intersubjective inter pretations that can be agreed upon by a wider group; this is why there is heavy reliance on inter rater reliability in content analysis. Yet, in practice the 'raters' are often supervisors and their students or otherwise like minded individuals. So we must be cautious about considering that such results represent 'reality', even an inter subjectively interpreted one.

What Is the Relative Importance of Manifest Content as Opposed to Latent Content?

In theory, manifest content concerns the most 'com monly' understood intersubjective interpretation of a message while latent content concerns more holistic and/ or deeper meanings and even the relationships between messages – which are also intersubjective. Some consider this difference equivalent to the difference between 'surface' and 'deep' meaning and many equate manifest content with quantitative content analysis and latent content with qualitative content analysis. Such dichoto mies are in practice false, whereby for example the content sought, whatever it may be, is a selective process. Ultimately, all content will fall somewhere along what may be considered a continuum from manifest to latent meaning. What is equally important is to recognize that where the researcher looks for meaning, whether deeply or not, is often determined by their own preexisting methodological and theoretical preferences. This reflects ongoing debates concerning the crisis of representation marked by the cultural turn in geography and the social sciences generally.

Is Frequency Related to Importance?

Some assume that knowing the precise frequency of some phenomenon is more important than simply knowing whether it exists in a text. In the example above it may be important in some instances to know how many different newspapers are biased one way or the other, while in others it may be enough to know that the Houston Chronicle is heavily biased toward the 'pro cut' frame. The simple presence or absence of a phenomenon may be enough to address the problem at hand.

Do Computers Make It Too Easy To Rapidly Produce Irrelevant Results?

This is a common criticism against computerized con tent analysis software (CCAS), and even computer assisted qualitative data analysis software (CAQDAS) generally. This goes hand in hand with concerns about an overemphasis on manifest content – a reason many qualitatively oriented geographers dismiss content an alysis. There are over 35 computerized CCAS packages excluding all purpose CAQDAS packages like NVivo or Altas.ti. Indeed, programs like General Inquirer can, in seconds, produce output using preloaded dictionaries. There are at least three sources of concern. First, such programs are accused of being little more than keyword counters – consistency at the expense of relevance. Sec ond, is the concern that computers are a poor replace ment for humans as creative, adaptive, interpretive thinkers – manifest content of the worst kind. Though computers may consistently apply rules, they may not always accurately identify the appropriate, contextual ized use of a message. This inaccuracy increases with the size of the text unit. Third, many lament the potential loss of meaning from simply applying preexisting dic tionaries or coding schemes – comparison at the expense of new understanding. The best use of CCAS and CAQDAS is to help facilitate the task of interpretation; the worst use is to simply allow it to produce frequencies without any theoretically informed reflection on the how those frequencies came to be.

When used appropriately, content analysis can provide extremely useful insights about 'texts'. However, since the term 'content analysis' can be used in a variety of contexts, it is important to understand 'how' the term is being used in each specific case. Foremost, consumers of empirical research must understand whether it refers to quantitative or qualitative content analysis since each has very different epistemological and ontological foundations and standards for rigor. Likewise, practitioners should be careful to define the precise way they are using the term. Since the roots of content analysis are decidedly quantitative, those adopting a qualitative approach may opt for a less ambiguous label like interpretive analysis, hermeneutic analysis, ethnographic analysis, or discourse analysis.

See also: Discourse Analysis; Grounded Theory; Positivism/Positivist Geography; Quantitative Methodologies; Semiotics.

Further Reading

Babbie, E. (2003). *The Practice of Social Research* (10th edn.). Belmont, CA: Wadsworth/Thompson.

Bendix, J. and Leiebler, C. (1999). Place, distance, and environmental news: Geographic variation in newspaper coverage of the spotted owl conflict. *Annals of the Association of American Geographers* 89(4), 658 676.

Berelson, B. (1952). *Content Analysis in Communication Research*. New York: Hafner.

Crang, M. (2005). Qualitative methods: There is nothing outside the text? *Progress in Human Geography* 29(2), 225 233.

Holsti, O. (1969). *Content Analysis for the Social Sciences and Humanities*. Don Mills, ON: Addison Wesley.

Krippendorff, K. (2003). *Content Analysis: An Introduction to Its Methodology*. Thousand Oaks, CA: Sage.

Larsen, S. (2004). Place identity in a resource dependent area of northern British Columbia. *Annals of the Association of American Geographers* 94(4), 944 960.

Lowe, N. (2002). Software for content analysis: A review. Report for the Identity Project. http://people.iq.harvard.edu/~wlowe/Publications/rev.pdf (accessed on February 2008).

Neuendorf, K. (2002). *The Content Analysis Guidebook*. Thousand Oaks, CA: Sage.

Weber, R. P. (1990). *Basic Content Analysis*. Newbury Park, CA: Sage.

Relevant Websites

http://academic.csuohio.edu
Companion to Neuendorf's book, The Content Analysis Guidebook Online, The Academic Server.

http://www.courses.washington.edu/socw580/contentsoftware.shtml
General content analysis, links to software.

http://www.content analysis.de
General resource, German language version of site.

http://www.webuse.umd.edu:9090
Online version of General Inquirer software, Scientific Research on the Internet, University of Maryland.

Corbridge, S.

M. S. Kumar, Queen's University Belfast, Belfast, UK

Glossary

Anti-Essentialist Dogmatism and essentialism have been associated with the Marxian analysis of underdevelopment. Essentialism arises because of the disjuncture between theory and the 'real' world. The obsession with the transcendental qualities of capital and of structures which bind the poor in the developing world are the key attributes of essentialism. This leads to a rigid form of dogmatism which colors all worldview, theory, and ideology. It provides a highly generalized perspective of the real world and ignores the potential of those in the periphery to undertake their own development.

Anti-Intentionalists Like anti-essentialists, anti-intentionalists too seem to believe in the power of the market to provide solutions to the endemic problems of poverty, poor health, and lack of livelihood faced by the poor in the developing world.

Corbridge, Stuart E. (1957–)

Corbridge states quite emphatically that "I as a middle class white Englishman, cannot or should not make an effort to engage in an 'imaginative' conversation with the Detroit auto worker or with an adivasi mine worker in Bihar, India" (**Figure 1**). Corbridge shuns the light heartedness of the postmodernist and the hyper universalist at the same time. He challenges the continuing normalizing discourses of development across regions and communities which reinforce the impossibility of development studies to transform human lives. He thus poses the question as to how we might put ourselves in the shoes of the 'other'? Indeed, according to Corbridge, "any adventure in the tropics would require engagement not just with trade and fiscal policy, but also of good governance, public service reforms, assymetrical information base and sustainability, not to mention the politics of the region."

Stuart Corbridge was born in West Midlands and studied at Sidney Sussex College, University of Cambridge. He is a leading geographer with a determined public face who has championed the cause of the developing world in general, and India in particular, for more than two and a half decades. Having graduated with first class honors in geographical tripos, he obtained a PhD under the supervision of Professor Ben Farmer from the University of Cambridge in 1978. His PhD was titled: '*State, Tribe and Region: Policy and Politics in Jharkhand, India, 1880–1980*'. He started his career at Huddersfield Polytechnic where he worked from 1981 to 1985, and moved to Royal Holloway and Bedford New College, (London University) from 1987 to 1988, he was appointed as Associate Professor of Geography at the Maxwell School of Public Affairs, Syracuse University, in the US. He then came back to the UK, Cambridge and worked as lecturer in South Asian geography from 1988 to 1999. In 1999, he was appointed as Professor of International Studies at the University of Miami, US. He has since come back to the UK in 2001 as Professor of Development Studies, London School of Economics, UK. In 1993, he was nominated as Visiting Professor, CSRD, Jawaharlal Nehru University, New Delhi. He was also nominated for the Benjamin Meaker Visiting Professorship in Politics, Bristol University, UK; Distinguished Visiting Professor of Development Studies, University of Oslo, Norway; and in 2006, as Visiting Professor of Asian Studies, University of Texas Austin. His main research interest lies within the Indian subcontinent and he has been a major contributor to the debates on the transitions in the subcontinent. His teaching has invigorated generations of students in geography who imbibed critical thinking on a wide range of issues from nationalism, democracy, and development; to governance, participation, and empowerment among the poor; the politics of forest policies which infringe upon the livelihoods of the tribal adivasis; the politics of transformation of traumatic spaces in Ayodhya and Bhuj in India; geopolitics of development; of the burden of debt; and of critical debates in development theory.

Through his journey, he has emerged as a singularly powerful and reasoning voice in the development studies

Figure 1 Stuart E Corbridge.

circuit for his commitment to and insights about the understandings on development in general and India in particular. The main intellectual influence on his thinking and writing can be effectively traced to Benny Farmer's area studies' tradition and Derek Gregory's challenges to social theorization. Like Farmer, Stuart was never inter ested in 'geography' with a capital 'G' and his writings and seminars attest to this fact. He draws widely from 'the social sciences', of politics, economics, history, and an thropology in a rigorously multidisciplinary manner. The eminent economist, Professor K N Raj, at the Centre for Development Studies in Trivandrum (Kerala) was over whelmed with Stuart's presentation and pronounced it as one of the best in social science tradition. Thus while many pay lip service to multidisciplinarity, Stuart has doggedly persisted in breaking through the straitjacket of disciplinary rigidity, which has been recognized in his taking up the challenging task of editorship of the pres tigious multidisciplinary journal, *Journal of Development Studies*. Besides, he is also a member of the editorial board of *Review of International Political Economy*; *Singapore Journal of Tropical Geography*; *Ethics*, *Place and Environment*; *Progress in Development Studies*; *Economy and Society;* and *Journal of South Asian Development*. He acknowledges the intellectual debt of gratitude he owes to two books in his evolution as a scholar namely, Ben Farmer's *Introduction to South Asia* and David Sopher's *An Exploration of India*. He believed that knowledge produced at the boundaries of differing intellectual traditions, of theoretical, empirical, and pol itical enquiry makes the world of scholarship exciting and engaging. He shunned the idea of looking for a theoretical certainty in a largely disparate and uncertain world.

He adopted one of the most underdeveloped regions of eastern India – Jharkhand – in order to understand the dilemmas of development. Here, he investigated the disjuncture in the allocation of resources between the rural and the urban. The state, in India, played a major role in the redistribution of wealth across population subgroups and he recognized early on that the state, more than the free market, was critical in advancing or re stricting developmental initiatives in the eastern region. Corbridge started publishing research based on fieldwork in India from 1982, and his avid engagement with eastern India politics and development issues was sustained over the last 25 years, highlighting policy and questions of agrarian change, development policy, and governance. He has made significant contributions to the geopolitics of development in India. In 1986, starting with Capitalist World Development, Stuart went on to chart the issue of Global Debt and Development; with John Agnew, he then jointly authored the volumes, *Mastering Space: He gemony* and *Territory and International Political Economy*. His jointly authored volume on India with John Harriss, *Re inventing India: Liberalisation, Hindu Nationalism and Popular Democracy*, was widely appreciated for the breadth of scholarship. This was followed by his edited volumes, *Jharkhand: Environment*, *Development, and Ethnicity* with S Jewitt and S Kumar; *Seeing the State: Governance and Gov ernmentality in India* (G Williams, M Srivastava, and Rene Veron); and finally *Colonial and Postcolonial Geographies of India* jointly edited with S Raju, and M S Kumar. Overall, he has contributed extensively to political and economic debates on India in major international journals and has been the special adviser to the Select Committee for International Development, House of Commons, 2004–05 (UK development assistance to India). The graduate students he supervised are all well placed in key positions and continue the legacy of his scholarship in the UK, India, and the US.

Corbridge scholarship over the years has attempted to ground the key issues of development and under development. He highlights the diverse challenges to mainstream development studies as presented by both the New Right and the post Left (1998). Corbridge's empathy with the developing world was reflected in his 1993 classic *Marxisms*, *Modernities and Moralities: Devel opment Praxis and the Claims of Distant Strangers*, where he argued that it was important "to care about the lives of distant strangers, people whom we do not know but without whom our lives would be impossible."

Corbridge has provided a sustained analysis of de velopment and its engagement with people in diverse spaces. His ideas on the ethics of development, as well as the geopolitics of money, space, and power have provided new departures in the field. His commentaries on the nature of development in *Reinventing India* along with his recent research output of *Seeing the State* extends the debates on governmentality, corruption, and the civil society in India. These commentaries provided lessons for the developing world in engaging with public policy and spelt out the choices available to them in the newly reorientated Washington Consensus of the neoliberal world order.

His first book – *Capitalist World Development* (1986) – marked a fresh departure when critique of established radical and Marxist modes of analyses was not com monplace in geography or even in general social sciences. With the collapse of the Washington Consensus, and the post Suez crisis, Corbridge, like many other scholars, was grappling with the nature of the 'impasse' in the Left's analysis of development and underdevelopment. He as serted that the neo Marxist model was incapable of re solving developmental problems in the poorer regions of the world. For this, he suffered a vicious backlash from some senior scholars for daring to call the bluff of 'methodological exclusivism', and was accused of a dog matism which was based on excessive dependence on economistic explanations. His analysis contributed to a first round of critique questioning such dogmatism and essentialisms within Marxist and radical development theories. Yet, it must be said in his defense that Corbridge never took an extreme position insofar as the scholarship or the academia was concerned. He was truly grounded in the Indian 'middle way' and always accepted the critical role of Marxism in the politics of transformation and progress and highlighted the plight of those dis enfranchised by the power of capital. He emphasized the need to engage with the nature of exploitation. Being embedded with the realities in the field in the developing world, he was always skeptical about the disjuncture between theory and empiricism. Thus for him E P Thompson and Raymond Williams were just as import ant as Rosa Luxemburg and Karl Marx.

He sought to move beyond the limitations of Marxist and neo Marxist approaches including structuralism, without necessarily condemning the essential scope of the transformatory character of radical Marxian per spective. Corbridge accordingly made a claim for a rad icalized modernity, without necessarily staking a claim to Marxism and its attendant discourses. He acknowledged the sensitivity of Marxism to the politics of transforma tion and progress and explicitly recognized that the un even development of the modern world was founded on 'the exploitative and anarchic logic of expanded com modity production'. He allied himself with those dis enfranchised by the power of capital who were subjects to exploitation.

He hence questioned the idea of blaming 'capitalism' for the underdevelopment in the developing world and instead focused on the geopolitics of underdevelopment and its role in creating the persistent inequalities that continue to plague the world at large.

In challenging the neutrality of social sciences, Cor bridge provided a strong critique to neo Marxists' ap parent lack of appreciation of the role of government, and of the developmental state in the Third World in assisting capitalist development. He thus made a coherent and important contribution to the critical debates of world capitalist development, which is contingent and reflective of real world issues in the developing world. He em phasized that nation states are intrinsically empowered to determine the course of development within their boundaries. Throughout, Corbridge allied himself with those disenfranchised by the power of capital and 'who' thereby succumbed to exploitation. This was evidenced in his exposition about the de territorialized adivasi tri bals in eastern India. His extensive synthesis made use of concepts and ideas and provided an incisive critique of the dogmatic positions that continue to plague and foster an 'impasse' in development studies.

In his writings, he has retained his strong belief in rad icalism to foster dialog for critical thinking in geography and development studies in particular, while outlining the position of post Marxism, postdevelopmentalism, and pro letarian essentialism. Indeed, Stuart challenged all forms of essentialism. He has been a keen student of the emerging new agrarianism in rural India leading to a rise of active rural politicization resulting in a change of political balance. As one who is firmly rooted in the fieldwork tradition, Stuart's insights are clearly demonstrated in his provocative piece titled, 'The merchants drink our blood' written in 1997. He asserts that keeping one's feet firmly on the ground is far more critical than engaging in the polemics of theory and ideology of development.

He makes a persuasive case for mainstream develop ment studies against certain moral claims which were being promoted by both the New Right and a diverse post Left. While seeking the morality and ethics of this discourse, he contends that citizens and states in the advanced industrial world have a responsibility 'to attend to the claims of distant strangers' in an interdependent economy. Thus, for him development ethics is not just about questions of transnational justice and positionality, but also about the construction of plausible alternative worlds, that is, a 'middle way' between growth, redistri bution, and governance involving practical development policies.

Corbridge goes beyond dogmatic attachments to sci entific Marxism and has little truck with those who make a career out of condemning the very idea of develop ment. He has highlighted the fact that, in recent years, development studies faced a major onslaught on its hu manistic traditions. Development for him is therefore both intentional and immanent. He is conscious about the need for and the value of intervention in the poorer regions of the world: the idea being that there is a positive relationship between economic growth and human development, which goes beyond simple quanti tative measures of development. There is an implicit belief that developmental states have "some responsi bility and capacity to stimulate growth and an implicit belief that rich countries have a duty towards meetings the needs of poor nation states." He decried the lack of "moral (normative) bases of international solidarity" and highlighted the negative role played by anti intention alists of the New Right and anti essentialists of the post Left schools of thought. He was acutely aware of the exclusionary tendencies of the New Right and the post Left in development studies. In other words, he advo cated the importance of strategic intervention for the betterment of the poor. Such interventions included trade versus aid, debt relief, technical cooperation, and exchange of expertise. Here, ethics and transnational justice were critical for strategic interventions.

According to Corbridge, the rights and voices of dis tant strangers are crucial for grounding the ethical and moral agenda of development. Therefore, lack of en titlement and choices available to the poor present challenges that go beyond the immediate gratification of needs. For him, "a minimalist universalism in respect of basic human needs is more persuasive than arguments which maintain we need to de link the lives of We and our Others." Thus, he maintains that a morality in de velopment studies must contain a morality of its critique too. Engagement with the moral and ethical foundations of mainstream development polices are critical for the survival of the agenda for development. Can we return once more to the ethical and moral foundations of de velopment? He denounces 'untrammeled individualism' and free market without regulation.

He clearly advocates a change in the relationship of care and duty that is being mapped out in the wake of globalization with respect to the developing world. He also highlights the key role that activists and scholars may play in providing alternative patterns of development agenda. The needs and rights of distant strangers are critical in this new configuration. He definitely does not subscribe to the caricatures of developmentalism and cosmopolitanism and maintains that development in the 1990s is vastly different from that in the early 1950s or the 1960s. He is against essentializing and universalizing crude versions of developmentalism. Thus, the lives of distant strangers are deeply embedded in all of us in the form of largely unequal social relations. In a way, the politics of development play a primary role in defining the end points of development. He therefore reiterates the need to adhere to less exclusionary forms of devel opment ethics. This means a moral commitment, to discharge our responsibility toward ameliorating the distress of distant strangers, which is cultural and region specific.

He calls for us to be aware of the privileges that we enjoy in this unequal world. Having a heightened sensi tivity to the problems of thinking, as an 'other' is what Stuart strives to address beyond the rhetoric of devel opment. He highlights the danger of subscribing to the post Left's antidevelopment ideas, which raises the specter of indifference within the local. As he states, "equality of opportunity cannot be squared with hunger, poverty, slavery and state organised torture of innocent victims." He states that romantic populism rarely en hances the lives of the poor on a daily basis and urges us to look more critically at the "real dilemmas of devel opment…." He has contantly raised critical issues of positionality, ethics, and of discursive power in his en gagement with the poor. As he notes, basic human needs for development have not necessarily been provided by the agents of the free market: an ethical critique of de velopment policies is accordingly far more important than simple derision of policies.

On questioning whether there is a space beyond de velopment, Corbridge emphasises that the emergent trends "both from a theoretical and a policy relevant perspective concerns the ways in which states and markets can be considered as institutions and can be combined in a more or less synergistic fashion." He acknowledges that key, exciting intellectual developments have been occur ring in the political anthropology of development and that the state may play an important role in regulating the process of both capital accumulation and development. He is convinced that the politics of the future must be a 'people centered' one, or one which contests the control ling ambitions of both the state and the market. He highlights the role of the Chipko movement (*andolan*) in the Uttarakhand region of northern India. Thus he calls for "empowerment as a substitute for development, where self worth or *izzat* is reclaimed by the victims of state sponsored development."

He cautions that capitalism is constantly evolving and changing to suit the imperatives of world development. As far as is discernable, he calls for a realistic revision of the rhetoric in the light of the evidence from the field.

Written with lucidity and rigor, Corbridge and Harriss, in *Reinventing India*, note that democratic institutions have been conspicuously absent on the Indian political land scape. They dismiss the idea that Indian history can only be read in terms of essentialist categories such as mod ernity. They are indeed comfortable with the idea that India and Indians straddle different moments in time, "living more comfortably on the borders between both the rural and urban worlds." Indeed the state continues to occupy a central place in the imaginations of the deprived communities in India. This idea is further explored by Corbridge and helps to destabilize the politics of major itarianism that almost found its rationale in Indian pol itics. A fascinating paper by Simpson and Corbridge emerged, which further engaged with the "geography of things that may become memories," and highlighted the contested politics of reconstruction and of place making in the aftermath of a major earthquake in western India. Indeed, Corbridge cannot be faulted for being re ductionistic and dogmatic. As he earlier noted, "a radical perspective need not be so pessimistic."

His co authored book *Seeing the State* was based on a detailed fieldwork extending the debates by Arturo Escobar, James Ferguson, James Scott, and Partha Chatterjee. It engaged with the de politicized practices of development which tried to steer clear of excessive state directed development. Corbridge, along with his co authors attempted to identify the key aspects of repetitive participation and good governance in the eastern Indian states of Bihar and West Bengal. This book breaks new grounds on engaging with the technologies of rule, and of governance. The general conclusions drawn are that a working civil society in India is able to interact with the state and this is largely the result of deepening dem ocracy. There is greater evidence of a decentralized, participatory form of democracy which is fostering active citizenship among the Indian subalterns. This work breaks new grounds in inviting readers to think politically about the diverse forms of governance, of governmentality based on evidence from a suite of pro jects such as the Panchayati Raj, Joint Forest Manage ment, and the Employment Assurance Scheme.

As he notes, "Globalisation is not only a synonym of disempowerment: it creates certain conditions for dem ocratization, de centralization and empowerment as well as for centralization and standardization. Globalization therefore opens as many doors as it shuts." He thus presents a sustained analysis of India's fragmentary na ture of economic development by integrating evidence from the margins of Jharkhand to the subcontinent as a whole in a well articulated synthesis with John Harriss. He has recently completed a jointly edited volume on *Colonial and Postcolonial Geographies of India*, which has received commendation from across the international press for showcasing the key issues from the Indian subcontinent.

In conclusion, Corbridge's work transcends disciplin ary boundaries and his contributions in the field of de velopment studies are well acknowledged. He provides a critical edge to debates on capitalism as a sustainable model for development. He constantly urges the discip lines to engage with a realistic revision of the diverse rhetoric, that seems to have swamped current scholarship, based on clear evidence from the field. He calls for em pirical data to substitute empty rhetoric in development studies. It is important to acknowledge that in all his scholarship, he has never privileged space, or geography as being the center of his argument. He constantly ac knowledges the power of space, (both rural and urban) in enhancing critical explanations. He asks the question as to whether geography matters more than institutions to help explain the convergence or divergence in the state of development? Why do some states 'fail' and others do not? How far can globalization be linked to deepening poverty and inequality at different spatial scales? What does the increasing liberalization of the agricultural sector mean for food security? Corbridge constantly challenges the illusion of a transcendental market in all his writings. He calls for a re politicization of develop ment studies by focusing on the politics of policymaking. As he states, "Development involves concerted struggles against the powers of vested interests, at all spatial scales." Thus development geography has to be attentive to the needs of 'distant strangers' whose choices and in dividual freedom remain circumscribed and structurally constrained by the stranglehold of globalization. Indeed, development studies cannot be reduced to a singular ideology of development, and transcendental claims of the state and market need to be reevaulated.

See also: Capitalism; Development I; Development II; Postdevelopment; Uneven Development.

Further Reading

Agnew, J. and Corbridge, S. E. (1995). *Mastering Space: Hegemony, Territory and International Political Economy*. London: Routledge.

Corbridge, S. E. (1986). *Capitalist World Development: A Critique of Radical Development Geography*. London: Macmillan.

Corbridge, S. E. (1993). Marxisms, modernities and moralities: Development praxis and the claims of distant strangers. *Environment & Planning D: Society and Space* 11, 449 472.

Corbridge, S. E. (1997). The merchants drink our blood: Peasant politics and farmers' movements in post Green Revolution India. *Political Geography* 16(5), 423 434.

Corbridge, S. E. (1998). Development ethics: Distance, difference, plausibility. *Ethics, Place and Environment* 1, 35 54.

Corbridge, S. E. (2001). Beyond developmentalism: The turn to cultural anthropology. *New Political Economy* 6(1), 81 88.

Corbridge, S. E. (2002). Development as freedom: The spaces of Amaryta Sen. *Progress in Development Studies* 2, 183 217.

Corbridge, S. E. (2007). The (im)possibility of development studies. *Economy and Society* 36(2), 179 211.

Corbridge, S. E. and Harriss, J. (2000). *Reinventing India: Liberalization, Hindu Nationalism and Popular Democracy.* Cambridge: Polity, and Delhi: Oxford University Press.

Corbridge, S. E., Jewitt, S. and Kumar, S. (2004). *Jharkhand: Environment, Development, Politics*. Delhi: Oxford University Press.

Corbridge, S. E., Williams, G., Srivastava, M. and Veron, R. (2005). *Seeing the State: How Poor People Experience Governance and Democracy in Rural India*. Cambridge: Cambridge University Press.

Raju, S., Kumar, M. S. and Corbridge, S. E. (2006). *Colonial and Post Colonial Geographies of India*. New Delhi: Sage.

Core-Periphery Models

B. Ramírez, Universidad Autónoma Metropolitana, Mexico City, Mexico

The core–periphery categories have explained the relation between regions that have experienced capitalist development and concentration, and the ones set apart. These terms have been used by different theoretic currents and in different ways since the end of the nineteenth century. They have two different aims, depending on the model: to analyze change and transformation of different societies or modes of production, and to propose different ways in which change and transformation could affect countries, regions, or cities.

These models have been developed by different areas of knowledge in an attempt to generate links and connections within them. Urban or economic geographies have been crossed by economic, political, sociological, or planning perspectives generating different ways to go beyond the economic or the geographical view. Not in all discourses are the above categories used as part of the construction of the model. Nevertheless, others such as developed–underdeveloped, metropolis–satellite, modern–traditional, and North–South, implicitly refer to the relation between one central zone and another that surrounds it in the so called periphery.

The main perspective in this discussion assumes a change of societies and territories as the central approach. One can recognize three different perspectives of change: transition, organization, or integration. In the first, transition adopts a macro scale dimension. At the end of the nineteenth century, this discussion was led by Marx, Durhein, and Weber, among others, and related to the modern aspects of capitalism but, specifically, to an attempt to understand the nature of the transition from feudalism to capitalism. This comprehension was different for each of the authors and their positions were sometimes even opposed.

On another scale, the second processes are related to the urban–rural scenario, which means they emphasize the regional perspective of the geographical situation of space. Part of this view consists of the ways in which cities reorganize the surrounding countryside, integrate production to the requirements of urban industry, concentrate services, and organize central places. Three aspects are discussed here: that of integration of rural to urban resulting in regions, organization of the surrounding centers, and transitions from rural to urban.

Lastly, integration of developed and underdeveloped countries and the way in which links and relations are generated in the world market, or at the regional scale as well, are part of these arguments. It is important to mention that the adopted frameworks are also a part, sometimes explicitly but others implicitly, of the explanations given to understand intranational processes between different regions.

One of the main characteristics of the discussion is that the leading vision of the problem always juxtaposed two different geographical territories in a confrontation of hierarchical spaces: the center and the periphery. This juxtaposition assumes not only the possibility, but also the necessity to generate change in the periphery in order to achieve goals and richness in the center. But at the same time, it assumes either the possibility that all spaces in some sense need to be like the center, or within an historical interpretation, that all non European spaces are underdeveloped. In that sense, some authors have characterized the discussion as Eurocentric. The search for equilibrium between spaces is at the core of the discussion.

In general terms, we can recognize two types of models: those emphasizing the national approach, where the assumption is that the process of social and territorial change is essentially internal to nation states; emphasis is placed on the intranational situation in order to solve regional differences. On the other hand, the international models, with an economic emphasis, assume that the world markets and trade are the main aspects producing change and evolution in countries and regions. In both cases, arguments regarding causes of development and underdevelopment generate different theories on the subject.

Intranational Models

Within the national perspective we can recognize three different positions. First, that of economic geography at the end of the nineteenth century, developed by Vön Thünen; later, between World Wars I and II, we have the central place theory developed by Lösch and Christaller. Both focused their interest on studying centrality as a way of organizing a city and its surrounding rural areas. The former, in a circular way, organizes agriculture, cattle production, etc., in the periphery, leaving the urban center as the main point to be developed according to urban industrial necessities. The latter is a geometric and hexagonal model, first organizing the main center and its hinterland, depending on the services provided and the cost of transportation defining the limits. But at the same time, different and less important centers are organized in a hierarchical structure of less important units around the main urban central place. Since these models have been widespread among geographers and economists, they are still important today. We will return to this perspective when we explain Krugman's contemporary proposal.

Within this perspective, linking economy with geography we can include Perraux and Boudeville's proposal, known as 'poles of development' or 'growth poles'. It came out in France after World War II and assumes that growth and development do not exist in all regions at the intranational scale, leading to the recognition of two different types of regions: the homogeneous and the polarized, which could introduce change through planned activities. The former are those where continuity and concentration are the main features making the region coherent and interdependent; the polarized regions are backward but have possibilities for industrialization, allowing for transformations in production and services. Lastly, regional plans are those spaces where investment could generate the established economic goal of change and growth. Even if this proposal does not assume the center–periphery discourse, the main features of the model are within this perspective: the objective is to transform the polarized regions into developed ones by planning activity imposed at the regional level as well.

As a second perspective, we include Rostow's transition model – an important position in the 1960s. He does not concentrate on the dichotomy between central and peripheral countries, but points out the possibility that those who are backward can achieve development through the application of the industrial modernization of the West, focusing on the ability of the entrepreneurial elite to stimulate growth and change. The model assumes a linear view of growth and not the polarization perspective, as his implementation has to be done in five stages, each of them representing a transition from the traditional to the modern society. We could argue that different stages of backwardness might achieve the growth of the center if they follow the path of development and change defined by the model. It has a double target: on the one hand, it allows for evolution and change, but on the other, it assumes an integrative possibility for the less modern or even backward countries, at any stage of their transition, to achieve the goals and principles of the developed ones located at the center of the process. In that sense, we can affirm that it is a model that tries to find equilibrium and homogeneity within economies and territories, with a strong influence in models like Friedmann's, as will be developed later.

The third national perspectives, mainly economic, which are widely known among geographers, include those of Myrdal, Hirshman, and Friedmann. Developed in different contexts and even times, they have been very widespread among academics and politicians in order to promote change and development. The Swedish politician and economist, Gunnar Myrdal, pushed forward his vision at the end of the 1950s in an attempt to generate a world antipoverty program. His proposal, called the Circular and Accumulative Causation Model, emphasized the vicious and circular cause of poverty. Change, he assumed, could be not only economic but of any kind, since all movements could generate others in an interwoven and circular manner. In that sense, his arguments support a strong interdependence between economic and political phenomena and between the institutions supporting these processes. Thus, state intervention is important in the economic and political life of any country.

Changes in the economic and social system do not generate the compensations required for equilibrium. On the contrary, there are other processes moving the system in the same direction as the original change. As a result, there is an increase in movement, giving an impulse to the accumulative transformation of social processes, resulting in a disequilibrium and sometimes accelerating rhythm and evolution.

Causes for change are cumulative in a circular system and they could generate an ascendant movement. The more we know about the different factors and the way in which they interrelate, the more one can cause a maximum increase in policy efforts, changing the social system. Industry is considered to be the main agent of change, producing, in a centrifugal manner, new impulses to other localities. Inequalities are seen only as a problem of the rhythm of progress between regions within a country. In that sense, there is a need to generate compensatory changes in external diseconomies in order to compensate impulses that evade changes. The role of public intervention is important in order to generate movement and to provide harmony for social change; to prevent poverty and inequalities between regions, industries, or social groups.

During the 1960s, when the welfare state was still one of the main agents supporting development, Albert Hirshman, a German economist who lived in the USA for many years, generated a possibilistic escape route from narrow economic structures. His main interest was how to make a bridge between politics and economy, endogenous and exogenous factors for development. In this case, his model is a conjunction of political and economical forces, organizing, in an appropriate place, the political dimensions of economic phenomena, with both forces treated as endogenous variables. The model assumes that economic growth will reduce political tensions and can produce more cohesion within economic and political agents.

In this view, the market is the only possibility to create changes through growth and industrialization. The role of the State in development via the model of import substitution is important. Maturity of a country is achieved when it is industrialized; the infrastructure required is provided by the State as well as subsidies for new industry. In that sense, hidden resources have to be distributed through State intervention.

Lastly, with strong influence from Myrdal, Hirchmann, and Rostow, and other positions like those of Prebisch and Isard, in the late 1960s John Friedmann developed a simple and normative model of spatial development in backward countries. Despite the fact that he later moved his view to other assumptions, it is an important model to analyze as it incorporates many of the dominant ideas of the moment.

The model assumes, in a linear perspective, the evolution of a spatial system in four stages, which traces the transformation from an underdeveloped to a developed situation in a national or regional context. The first stage assumes the existence of independent local centers with scarce population, mainly rural or devoted to primary sector, without any exploitation or hierarchy among them. In the second stage, industrialization marks the beginning of the transformation, assuming the concentration of investment in one or two cities, creating the emergence of a dualistic spatial structure due to flows and exchanges, with a center that is experiencing rapid and intensive development, and a periphery whose economy imperfectly relates to this center and is either stagnant or declining.

Maturity in industry and concentration of wealth in the center bring opposition to political consciousness in the periphery. These are some of the main features of the third stage. The dynamic of regional differentiation is observed as national governmental actions and the growing market brings about more regional disparities. Poverty is confined to limited areas of the periphery, with strong possibilities of being eradicated through state interventions.

The final stage creates a functional, independent system of cities at the national level with a balanced, urban sized distribution emerging with regional centers competing effectively for resources using national capital. The assumption of national integration, efficiency in the location of firms, and growth at maximum and minimum interregional imbalances are part of the main features generating the model. Apart from the author's movement from these ideas to others in subsequent works, it was considered weak in two senses: as a descriptive view of the process and as a guide for public intervention. More arguments against this view are that it did not take into account differences in colonization within different Third World countries, natural resources, and economic activities resulting in the 'new activities' of the colonies. He undermined the role of foreign influence and emphasized the possibility of internal development without external forces constraining or controlling changes. In other perspectives, the model could be characterized as apolitical, considering that regional policy is a natural interest of governments simply due to public interest, and that poverty is treated as a regional and not a social problem, assuming that it could be easy to solve by following a spatial model.

International Models

These visions support trade and international relations as the main origin of core and periphery relations. Three groups of positions can be found in: the neoclassical model of Ricardo and Samuelson, the Marxist tendency, wherein we include the 'imperialist position' developed by Lenin, the 'world system' of Wallerstein, 'unequal exchange' introduced by Argyri Emmanuel and adapted by Samir Amin and others, and lastly, the 'dependency theories' in Latin America. International models are recognized by some authors as exogenous and always connect one space with another, mainly through trade and the world market.

Ricardo developed the 'law of comparative advantages', in which international exchange will lead countries to production specialization according to the relatively lower prices of the labor force, producing revenues for all countries. Neoclassical economies added to Ricardo's proposal that natural resources and capital were important elements in order to define the comparative advantages among different countries.

On the other hand, Samuelson admits that international exchange solves economic inequalities among nations; therefore, underdevelopment could be reduced through world specialization of production. The above arguments are supported by authors who argue for the comparative advantages between countries and free exchange as a means to correct inequalities between production factors and the distribution of resources among countries. Those positions are a pure and free defense of international exchange, as well as an assumption of homogeneity in production factors, level of knowledge and technology, and the natural and social conditions existing in the countries.

Contrary to the above concepts, Marxist theory created an important framework to understand capitalist relations at the international scale. In the first place, theories of imperialism seek to explain the continued growth of capitalism by focusing on the way in which the European expansion and the acquisition of colonies allowed for the exportation of capital, thus avoiding crises in the centers. There are two main positions sustaining imperialism. First, that of Hobson, who argued that in the absence of effective demand in the metropolis, the need arises to find new markets for commodities overseas. On the contrary, Lenin's argument sustained that to eliminate the declining rate of profit in the core, the opening of the colonies was an opportunity to find more profitable investment opportunities abroad, mainly through the flow of investment to the periphery. Lenin conceived imperi alism as a developed stage of capitalism and not as the relation between two territories and economies. This stage was characterized by four features: dominance of monopolies and financial capital; export of capital more than commodities; creation of world monopolies; and partition of the world between various imperialist powers.

Years later, the 'world system' model generated an advanced discussion of Lenin's imperialist perspective. Wallerstein, his main supporter, argued for the necessity to understand world history within the context of the capital accumulation process on the world scale, implying a draining of capital from the periphery to the center of the world economy. This position, shared with Emmanuel, Gunder Frank, Cox, and Samir Amin, ac cepts that theory and history are not only related but are the same: theory is history. The model has two main characteristics: on the one hand, it sees class as a group in the world economy, and on the other, it does not analyze class exploitation but efficiency along the various scales of the world system.

The name of the world system is capitalism, and more precisely, the capitalist world economy in process with a core and a periphery formed since the beginning. The core produces high valued goods and products, the periphery those of low value. There could be an external arena waiting to be incorporated into the periphery (empires or countries without the capitalist mode of production) that are sending products to the center as well. The contradictions of the system will allow us to understand its nature, characteristic structures and pro cess, as well as to mark tendencies, contradictions, anti nomies, and paradoxes. It will not last forever, since it cannot be programmed to do so.

Within this framework and with a strong criticism of the ordinary theory of underdevelopment, the unequal exchange theory of Samir Amin analyzes relations of accumulation at the world scale as well, with an emphasis on the flows and transfers of value generated by the re lations between the developed center and those of the underdeveloped periphery. When precapitalist countries interact, they relate with the center through open or hidden transfers of value from the latter to the periphery. Some critics of the unequal exchange theory have argued that Marxist theory is mistaken in that exploitation takes place in the sphere of circulation and not in that of production. This criticism is shared with that of the de pendency school as well.

During the 1950s, strongly focused on Latin America and undoubtedly influenced by the discussions going on at that time, the United Nations Economic Commission for Latin America and the Caribbean (ECLAC) model was of great importance for the continent because it supported the welfare state, generating industrialization via the substitution of imports. This had an impact on the transformation of the continent during the second half of the twentieth century, going thereafter into positions that created the Latin American dependence model.

According to authors such as Prebisch, Cardoso, and Falleto, dependency refers directly to those conditions of existence and functioning of economic and political sys tems, and to the internal and external foundations within Latin American countries. It is a model of relations sharing both national and international characteristics generating development. They argue that conditions of underdevelopment are the result of the relation between central and periphery societies. Through a transfer of resources from the latter to the center, it generates dis tortions in the economies of the periphery and limits the allocation of resources which create economic growth. This transfer adopts many faces and forms.

The underdeveloped situation is historically defined, therefore it is not a model of organization of economic data, where it is important to differentiate processes of relations between the colonial or national peripheral situation. It was generated with commercial and indus trial capitalist expansion when, despite different degrees of productivity, each country adopted diverse positions in the global capitalist system. Countries then have different positions within an international economic structure, as well as different positions of domination.

Dependency theory argues against the possibility of autonomous national development of Latin American economies and had a strong influence, managing to characterize processes during the 1960s and 1970s in the continent. Despite all the critics this model had, authors such as Gunder Frank and Castells used this framework to explain underdevelopment in the first place or, second, to characterize the urbanization of Latin America as dependent.

What are the links between the three forms of Marxist exogenous models? In Roxborough's opinion, if imperi alism focuses on the gains of the center, dependency focuses on the disadvantages of the periphery and un equal exchange in the forms and conditions of the movement of commodities and value.

Despite the importance the above theories had in explaining differences at a world scale, arguments dir ected at analyzing regional unequal development at the national level were used in order to define transfers of value between developed and underdeveloped regions within nations. They were part of the geographical dis cussions regarding regional transformations or relations among countries as well.

Contemporary Discussions

The discussion on core–periphery models converges today on five visions, with different positions. The first includes some authors assuming that, due to the centrality and concentration character of models like those of urban economics and industrial location, these become important in order to understand 'the new economic geography'. Krugman explores the possibility of generating two different kinds of core–periphery disequilibrium models: those related to the endogenous situation of the national scale and those of the international scale within the neoclassical economic perspective.

Concentration is at the core of Krugman's arguments, as well as the creation of centers of population and industry. He debates on the possibility of having a center–periphery model that depends negatively on the costs of transportation, and positively on demand and scale economy. The above elements are part of an accumulative process that generates and increases regional differences. The same perspective could be used to understand the organization of metropolitan areas, as well as interactions between countries at an international scale, and between regions in an intranational manner. The way in which those factors interact as part of a system generates the core–periphery model; however, the impossibility of finding equilibrium is part of Krugman's arguments.

The second position is mainly political and relates to the way in which the world perceives integration and reorganizations under the present economic and political conditions, which are moving toward globalization. Integration has three visions – the European, where the search for autonomy from Washington has generated new centers in Berlin and Paris which, without any doubt, reorganize the European periphery according to the relations of each area with the United States. The way in which the European community has been gaining strength in the last few years offers arguments that fortify this view of the world situation.

On the other hand and within Latin America, the same globalized context allows us to conceive of the integration of the continent's regional development through the generation of blocs and alternative postures along with the ones from Washington, or through intranational concepts, wherein the Economic Commission for Latin America, in its modified form, is still important today. The United States' perspective included a free market with Mexico and Canada through North American Free Trade Agreement (NAFTA), despite the fact that mobility and free circulation of the labor force is not a part of this proposal. The center, of course, is the USA, and the other allied countries are the periphery. The reorganization of agreements among countries of the world and, specifically, those of Latin America, has generated changes in the way in which the organization of commerce and trade is taking place in the world market. These visions change the concept of only one center and many peripheries to that of multiple centers, possibly hierarchized. This is part of the contemporary view and, of course, could be applied as well to other regions, such as the Asian part of the globalized triad of the world.

Third, at the intranational scale, this view of multiple centers with corresponding peripheries is important in order to understand regional imbalances and differences as well. From only one center and one periphery, competition for primacy among regions and cities is part of the contemporary tendencies in countries' reorganization. At the same time, sustainable development requires the development of urban areas, specifically metropolitan cities, needed to open the perspective of urban periphery for the conservation and preservation of the natural environment, more than development or urban growth. Another position within this scale is the perspective of the global cities, implying a concept of center developed through the concentration of services more than production. The tertiary sector is now more important than ever, especially that which is related to the financial and information sectors as activities generating concentration more than industry. Peripheries are the areas where those activities do not exist, as nowadays these areas are more spread out than usual.

Fourth, there are some views which do not integrate space through center and periphery but use flows to generate relations and links in a space of discontinuity more than a homogenized territory. Fragmentation is an important part of this concept, wherein some views of the regulationist theory are important, with authors such as Vetz leading this proposal. The possibility of explaining processes at both scales – the international or intranational – are included in this position, wherein centers are elements for the concentration of different kinds of flows within the national or international arena.

Finally, the critic's vision of the dualistic concept of the core–periphery vision has two aspects: one developed within the Marxist perspective against the concept of the ECLA in 1960, and the contemporary vision that came out with the discussion regarding space. The former has been made within the Brazilian context by Francisco de Oliveira; he comments that the dualist model of the ECLA in the first instance, and the dependency theory afterward, were not correct. He argues that rent concentration and accumulation in the periphery are part of the reproduction of history and capital, and therefore also reproduce the history and construction of capitalism in the central countries. Both histories are then interrelated in the production of wealth, in rent extraction, and in the differential impact of trade. In a sense, its conjunction is part of the assumptions adopted by the world system and by the unequal development perspective as well.

In a different manner and with arguments influenced indirectly by Marxist theories, but including the critics of modernization, Eurocentrism, and others, nowadays

Doreen Massey has set forth another opinion on the dualistic concepts of world development and transfor mation. Given the general reflection on the concept of space and the way in which it relates to time in a rhythmical and joint movement, changes and relations among countries, cities, and regions could be understood in a different manner. Encounters of different societies with particular stages of development and diverse histories could cause multiple and heterogeneous trajectories, directions, and velocities of change and movements.

The story of the West, its modernism and capitalism, is not the only possible history for the future. Each space, in any scale it might be considered, could contend with particular space–time movements, together with different trajectories and directions, in a situation in which en counters between different stages of development could generate new possibilities of change and relations.

Coexistence and multiplicity in networks of trajec tories and relations are the results of different relations not working as opposites but just as different parts and stages of the same process. In this perspective, 'back wardness' is not such, but is only one form of transfor mation with a direction and movement that is different from the 'established' history. For this perspective, there is not only one center but multiple centers, as well as different and also multiple peripheries moving in co existence as the result of a coherent and multiple movement of places with particular space–times.

This perspective is dramatically changing the way in which spaces are considered. Homogeneity and equi librium are not the images we need to achieve, as it was in the previous vision of modernism. On the contrary, simultaneous heterogeneity of different images of space–times in coexistence could achieve more democratic and politically active places where the images of people working for their own transformation could take place in simultaneous times.

This is indeed a new and different way to analyze space. The dualistic perspective is nowadays, in turn, hoping to find more open and democratic ways of visu alizing transformation and change for the future.

See also: Cumulative Causation; Dependency; Industrialization; Modernization Theory; Regional Inequalities; Uneven Development; Uneven Regional Development.

Further Reading

Amin, S. (1973). *Le Development Inegal*. Paris: Minuit.
Amin, S. (1974). *Accumulation on the World Scale: A Critique of the Theory of Underdevelopment*. New York: Monthly Review Press.
Cardoso, F. and Faletto, E. (1969). *Dependencia y Desarrollo en América Latina* (1st edn.). México: Siglo XXI.
Freedman, M. (1980). *Free to Choose. Towards a New Economic Liberalism*. New York: Harcourt Brace Javanovich, Inc.
Hirschman, A. (1971). *A Bias for Hope: Essays on Development and Latin American*. New Haven, CT: Yale University Press.
Krugman, P. (1995). *Development, Geography and Economic Theory.* Cambridge, MA: MIT Press.
Lenin, V. I. (1969). *Imperialism: The Highest Stage of Capitalism*. London: International Publishers.
Massey, D. (2005). *For Space*. London: Sage.
Myrdal, G. (1957). *Economic Theory and Underdeveloped Regions*. London: Gerald DucKworth.
Rostow, W. W. (1960). *Stages of Economic Growth: A Non Communist Manifesto*. Cambridge: Cambridge University Press.
Roxborough, I. (1979). *Theories of Underdevelopment*. London: MacMillan.
Wallerstein, I. (1976). *The Modern World System: Capitalist Agriculture and the Origins of European World Economy*. Cambridge: Cambridge University Press.

Relevant Websites

www.cepal.org
Latin American and Caribbean demographic center.
www.redtercermundo.org.uy
Red Del Tercer Mundo: Third World network.
www.eclac.org
United Nations Economic Commission for Latin America and the Caribbean.

Corporate Responsibilities

A. Hughes and F. Wray, Newcastle University, Newcastle upon Tyne, UK

Glossary

Corporate Citizenship The participation of a corporation in environmental and societal concerns beyond the financial bottom line.

FTSE4Good Index Series A performance measurement of global corporate responsibility standards, designed to inform investment decisions.

Multistakeholder Organization Institution comprising more than one kind of representative public or private caucus.

Nongovernmental Organizations Institutions operating independently of national governments, including civil society organizations and grassroots social movements.

Stakeholder Individual or group affected by, and affecting, a corporation's activities.

Transnational Corporation A company with operations, through ownership, in more than one country.

Overview of Corporate Responsibilities

Up until the early 1990s the central, and arguably only, responsibility of corporations was doing business, ensuring profitability, and maximizing returns to shareholders. This view was epitomized by the arguments of the free market economist, Milton Friedman, and the extensive rolling out of neoliberalism in the 1980s. However, over the past two decades, corporations have been viewed increasingly as having sets of obligations extending beyond core capitalist objectives associated with the financial bottom line. It is now a mainstream view, both within and outside the corporation, that business is also responsible for aspects of society and the environment that are affected by its activities. Corporations are therefore now seen as having some degree of moral agency with respect to various groups and environments influenced by their business strategies and practices. Terms such as corporate citizenship, corporate social responsibility, and sustainable business are now commonplace in the lexicon of what can, overall, be recognized as a corporate responsibility movement. And while there are many competing definitions of corporate responsibility, the groups and spaces most prominently regarded as being responsibilities of corporations are customers, employees, communities, and natural environments, as well as shareholders.

While key commentators such as Mick Blowfield and Jedrzej Frynas explain that the ethics of business has been a subject addressed for centuries, dating as far back as the first century BC, the current trends in corporate responsibility are argued by these authors and others to have their roots in late nineteenth century regulation to check the development of the modern corporation, as well as food boycotts during the same time period. The influence of late nineteenth and twentieth century corporate philanthropy associated with family run businesses, such as Cadbury and Rowntree in the UK, is also acknowledged. However, corporate responsibility in its current form appeared to develop from the early 1990s, with a distinct incorporation of issues associated with development, human rights, and environmental protection on a more global scale. Corporate responsibilities until this time tended to refer either to legal obligations to society or to philanthropic acts including charitable giving and strategies of caring for employees and local communities. It now involves a greater integration of responsibility toward workers, customers, communities, and natural environments into core business strategies involving, for example, corporate compliance with labor and environmental codes of practice.

A range of factors has fuelled the emergence of current developments in voluntary forms of corporate responsibility, including: (1) the globalization of corporations and their supply networks, and the associated context of trade liberalization and deregulation since the 1980s; (2) anticorporate activism; (3) the rise in the power of corporate brands and reputation, which makes large companies vulnerable to negative publicity; (4) an increase in public awareness of corporate activity and its impacts via improvements in global communications; (5) the growing importance to the investment community of ethical performance on the part of public companies; (6) the emergence of a strong business case for corporate responsibility; and (7) corporate scandals.

The context of globalization and the liberalization of trade and investment from the 1980s set the scene for voluntary corporate initiatives in the field of social and environmental responsibility because it engendered a decline in the regulation of corporate activity by traditional governmental institutions. The power of corporate capital encouraged by this climate of deregulation was increasingly called into question by campaigners, the critical media, and consumer groups during the 1990s. The protests by antiglobalization campaigners at the 1999 meeting of the World Trade Organization in Seattle have come to symbolize a much wider set of movements questioning corporate power. Campaigns concerning the contribution of corporate activity to environmental degradation and poor working conditions, for example, have been conducted by a host of campaigning groups including nongovernmental organizations (NGOs) and trade unions, as well as college students and critical journalists. Many campaigns have involved alliances between these different groups, frequently originating in the global North but sometimes extending to the global South. Most commonly, the object of anticorporate campaigning have been high profile, often transnational, brand name companies such as Shell (in the extractive industries), McDonalds (in the fast food retail sector), and Nike and Gap, Inc. (in the apparel sector). Such corporate names have become synonymous with global capitalism, and the adverse impacts they have sometimes had on the environment and worker welfare – their corporate irresponsibilities – have been widely criticized. However, such prominent corporations, as a result of campaigning, have since the turn of the century become leaders in the field of corporate responsibility with codes of conduct and extensive corporate monitoring programs.

Added to the influence of anticorporate campaigning and critical journalism, institutional investors such as pension funds are evermore concerned with corporate social and environmental performance, with indicators of ethical responsibility increasingly influencing their investment decisions. Such indicators include the FTSE4Good Index Series. Moreover, the business case for corporate social and environmental responsibility, shaped by the aforementioned actors, is holding sway with many corporate directors and managers, incorporating views about the management of risk through the avoidance of negative publicity, the retention of staff by ensuring responsibility to employees, the benefits to profitability of being branded as an ethical corporation and, more controversially, the possible avoidance of certain kinds of labor and environmental legislation by demonstrating existing responsible business through voluntary codes of conduct. The longstanding recognition of responsibility to shareholders has also tightened since the 1990s as corporations have been called to account through accounting scandals like those involving Enron and WorldCom in 2002.

The response to these political–economic forces of the majority of corporations, in particular transnational corporations (TNCs) and those at the leading edges of their industries, has been to introduce corporate responsibility programs, which address the impacts of their business on social groups (often referred to as stakeholders) and the environment. Many corporations have corporate responsibility representatives at director, as well as managerial, level and specific aspects of responsibility are integrated into a whole range of corporate functions – human resources management, supply chain management, quality assurance, customer relations, and more. Business schools and consultants in corporate responsibility and sustainability are also highly influential in guiding responsible corporate practice. There are many cases of improved relations with, and positive impacts on, stakeholders, though critiques of corporate responsibility abound (addressed in the final section of this article titled 'Corporate responsibilities: an economic–geographical research agenda'). Geographical studies include research into the relationship between anticorporate campaigning and corporate responsibility, the influence of institutional investors, the organization of ethical supply chain management and its impacts on labor standards in economically less developed countries, and the adoption of environmental standards by companies in different geographical regions. However, much research on corporate responsibilities is interdisciplinary in nature, incorporating contributions from economics, management, politics, and development studies.

Concepts and Tools of Corporate Responsibility

At the heart of developments in corporate responsibility and corporate citizenship is the notion of the triple bottom line. Coined by John Elkington in 1994, the term refers to the principle that business should respond not only to the financial bottom line concerned with profitability, but also to the protection of social needs and the natural environment. The concept of stakeholder engagement is closely aligned to the recognition of a triple bottom line. A company's stakeholders are the individuals and groups affected by its corporate activities, including employees, consumers, workers employed by suppliers, and communities. Stakeholder groups can also include organizations such as NGOs and trade unions who work to protect environmental and social interests affected by the company. Engagement with these stakeholders, involving consultation, monitoring, reporting, and collaboration, is a central part of how corporate responsibility is practiced.

The values of corporate responsibility are commonly crystallized into specific standards. These can be set by institutions working at the local, national, regional, or global scales. Examples of standards for corporate responsibility set by international organizations are the ISO 14000 environmental management standard set by the International Organization for Standardization, the SA8000 workplace standard set by the US multistakeholder organization, Social Accountability International, and international standards for the production of fair trade goods set by the Fairtrade Labelling Organizations International (FLO). The participation of companies is voluntary. Some businesses choose to work with such formalized standards and labeling initiatives, while others work with their own, or other organizational, codes of conduct. Voluntary codes are guidelines drawn up by companies, sometimes aided by stakeholder groups, and serve to specify issues of business responsibility by constructing a set of minimum standards and procedures for companies to follow in their routine business. Some codes incorporate conventions of international organizations such as the United Nations (UN) and the International Labour Organization (ILO), discussed in the following section. Such codes can apply to procedures covering all kinds of aspects of corporate activity, for example, customer care, human resources, waste management, and the management of supply chains. Codes have therefore become an accepted tool used in the management of corporate responsibility.

Accountability is another key concept in the field of corporate responsibility, referring to the notion that companies are answerable to the groups and environments their actions affect. Such accountability tends to refer not only to the general understanding that companies are responsible for their actions, but also that responsibilities toward stakeholders are enforceable and reported. In other words, the standards and codes in themselves are not enough to ensure that corporate responsibility is applied. Rather, procedures for monitoring, reporting, and verifying the implementation of these codes and standards are equally important in order to ensure transparency and accountability. Tools and techniques of social and environmental auditing have therefore developed, influenced by traditional procedures of financial auditing. What auditing provides is a process of checking, monitoring, and reporting evidence of corporate performance against the standards encapsulated in codes. Such voluntary codes and auditing procedures represent private interest forms of social and environmental regulation that appear to replace traditional, state led forms of regulation.

Globalization, Development, and Corporate Responsibility

There is widespread acknowledgment amongst corporate responsibility practitioners and academic commentators that globalization, trade liberalization, and deregulation have set the scene for current trends in transnational forms of corporate responsibility since the early 1990s, as outlined previously. However, Mick Blowfield, an influential author in the field, also insightfully points out that the practice of corporate responsibility itself also reflects some key dimensions of globalization, including rationalist standard setting, the dominance of business values and systems, a tendency toward measurement, and progress through alliances between private and public organizations. Moreover, the globalization of corporate responsibilities to incorporate overseas stakeholders affected by corporate activity means that corporate social and environmental responsibility increasingly involves economically less developed countries. Recognition that corporate activity has significant impacts on the development process means that corporate responsibility agendas are increasingly incorporating development targets. Examples include the UN Millennium Development Goals set in 2000, including the eradication of poverty, universal primary education, and environmental sustainability. Many business and development practitioners believe that corporate social and environmental responsibility, through the implementation of codes and standards, can serve to benefit workers and communities in developing countries. However, the key to achieving this in practice is seen to be effective stakeholder engagement, and there are significant cultural and language based challenges associated with this.

Corporate responsibility takes on global dimensions when it incorporates transnational business activity in two main forms. The first concerns corporate responsibility associated with TNCs, in other words where corporations operate in multiple countries through their foreign direct investment (FDI) strategies. The second form concerns corporate responsibility in firms' transnational networks, including supply chains involving overseas producing companies not owned by the purchasing corporation. In terms of the context for corporate responsibility applied by TNCs, first, Rhys Jenkins, Ruth Pearson, and Gill Seyfang explain that, although initial attempts in the 1970s to regulate the activities of TNCs were made by regulatory institutions at the international and national levels, neoliberal policies and increasing deregulation from the 1980s onward meant that the governance of TNCs took a voluntaristic turn. Trade liberalization, export promotion, and the encouragement of FDI fuelled a move away from state regulation of TNCs and toward corporate self regulation. Within the private sector and development agencies, the dominant view is that TNCs operating in developing countries are beneficial to the development process. However, civil society organizations and critical academics question this view in light of, for example, the low wages often paid to workers and the environmental degradation frequently caused by production. Codes and standards for business responsibility are increasingly used by TNCs to govern their operations in developing countries and improve their social and environmental performance in areas where the achievement of development goals is so important. Effective engagement of the TNCs with local level stakeholders is critical to the success of TNC corporate responsibility strategies.

With respect to corporate supply networks involving relations between purchasing (retail/brand name) corporations and their overseas suppliers, corporate responsibility has been implemented widely through labor and environmental codes of conduct involving minimum standards for workers and environmental protection at sites of production. Initiated by campaigning on the part of civil society organizations and trade unions often in the global North, along with media exposure of poor working conditions in export producing countries, programs of corporate responsibility involving labor and environmental codes are applied to ensure that minimum standards are met with respect to wages, hours of work, health and safety, the right to belong to trade unions, worker welfare, and environmental impacts.

Multistakeholder organizations involving institutional collaboration between companies, NGOs, and sometimes trade unions are particularly influential in guiding corporate approaches to labor standards in global supply chains. The UK's Ethical Trading Initiative (ETI), for example, influences the codes and procedures used by many high profile retailers and brand manufacturers in the governance of ethical supply chain conduct. The ETI was formally established in 1997 and now represents one of the largest and most strategically significant civil initiatives of its kind in Europe. It consists of corporate, NGO, and trade union members and is supported by the UK government's Department for International Development (DfID). Lying at the heart of the ETI is the base code of conduct, which the organization encourages all corporate members to use in the implementation of responsible business standards in the context of their own brand supply chains. This base code consists of nine provisions that build directly on core ILO conventions. These provisions, exclusively concerning labor standards, are: employment is freely chosen; freedom of association and the right to collective bargaining are respected; working conditions are safe and hygienic; child labor should not be used; living wages are paid; working hours are not excessive; no discrimination is practiced; regular employment is provided; and no harsh or inhumane treatment is allowed. The Fair Labor Association (FLA) in the USA and the Fair Wear Foundation in the Netherlands perform similar roles to the ETI. Significantly, each of these organizations has core ILO conventions at the center of its labor code (see **Figure 1** for the FLA Workplace Code of Conduct).

International development agencies are also playing a growing role in encouraging and guiding corporate responsibility. The World Bank encourages it through its Corporate Social Responsibility Practice, for example. In addition, the United Nations launched its Global Compact in 2000 – an international initiative to develop corporate responsibility. The Global Compact involves collaboration between UN agencies, labor, and thousands of companies worldwide to meet social and environmental objectives. At the center of the initiative are ten universal principles concerning human rights, labor standards, the environment, and anticorruption, which are drawn from the UN Declaration of Human Rights, the ILO's Fundamental Principles on Rights at Work, and the Rio Declaration on Environment and Development. Companies are encouraged both to incorporate the principles into their core business operations and to submit examples of their corporate responsibility practice. While participation in such initiatives is voluntary, they are nonetheless helping to shape progress in the application of corporate responsibility to the field of development on a global scale.

Critiques of the Corporate Responsibility Movement

There are many and varied critiques of corporate responsibility, despite the progress made by the aforementioned developments and initiatives. On the political right, some commentators continue to maintain that the only responsibility of corporations should be ensuring profitability and maximizing shareholder return. According to this view, corporate social and environmental responsibility runs the risk of compromising profitability. Of those commentators and practitioners who generally support the broad level objectives of corporate social and environmental responsibility, there are still points of contention. One concern is that the business case for corporate responsibility – suggesting that social and environmental responsibility is key to profitability – dominates the agenda, along with the tools and language of management associated with codes, auditing, and reporting. Criticism of managerial approaches to responsibility suggests the need for a closer engagement and dialog with noncorporate actors and their lifeworlds. More critical readings of the business case go so far as to suggest that it is all associated with dubious ethical branding and misleading public relations exercises.

Amongst critical academics and sections of civil society, serious concerns are raised about the replacement by private sector corporate responsibility initiatives of traditional governmental legislation. It is suggested that private sector initiatives do not always cover many important development indicators. Moreover, the voluntary nature of corporate responsibility, in contrast to enforced legislation, is called into question. Voluntary corporate initiatives mean that there is only a partial coverage of industrial sectors and corporate activities with respect to the implementation of social and environmental standards, as corporations can opt into or out of initiatives at will. The focus of media and NGO campaigns on certain brand name corporations and particular sectors, for example, frequently determines the pattern and success of corporate responsibility initiatives. Hence, criticized companies such as Nike and Gap, Inc.

> Forced Labor**:** There shall not be any use of forced labor, whether in the form of prison labor, indentured labor, bonded labor, or otherwise.
>
> Child Labor**:** No person shall be employed at an age younger than 15 (or 14 where the law of the country of manufacture allows) or younger than the age for completing compulsory education in the country of manufacture where such age is higher than 15.
>
> Harassment or Abuse**:** Every employee shall be treated with respect and dignity. No employee shall be subject to any physical, sexual, psychological, or verbal harassment or abuse.
>
> Nondiscrimination**:** No person shall be subject to any discrimination in employment, including hiring, salary, benefits, advancement, discipline, termination, or retirement, on the basis of gender, race, religion, age, disability, sexual orientation, nationality, political opinion, or social or ethnic origin.
>
> Health and Safety**:** Employers shall provide a safe and healthy working environment to prevent accidents and injury to health arising out of, linked with, or occurring in the course of work or as a result of the operation of employer facilities.
>
> Freedom of Association and Collective Bargaining**:** Employers shall recognize and respect the right of employees to freedom of association and collective bargaining.
>
> Wages and Benefits**:** Employers recognize that wages are essential to meeting employees' basic needs. Employers shall pay employees, as a floor, at least the minimum wage required by local law or the prevailing industry wage, whichever is higher, and shall provide legally mandated benefits.
>
> Hours of Work**:** Except in extraordinary business circumstances, employees shall (1) not be required to work more than the lesser of (a) 48 h per week and 12 h overtime or (b) the limits on regular and overtime hours allowed by the law of the country of manufacture or, where the laws of such country do not limit the hours of work, the regular work week in such country plus 12 h overtime and (2) be entitled to at least one day off in every 7-day period.
>
> Overtime Compensation**:** In addition to their compensation for regular hours of work, employees shall be compensated for overtime hours at such premium rate as is legally required in the country of manufacture or, in those countries where such laws do not exist, at a rate at least equal to their regular hourly compensation rate.

Figure 1 The Fair Labor Association (FLA) Workplace Code of Conduct. From http://www.fairlabor.org/conduct (accessed 12 April 2008).

tend to have more developed social and environmental compliance programs.

An important geographical issue regarding the critique of the corporate responsibility movement concerns its Western centrism and the assumption that codes, standards, and procedures of implementation developed in the West are universally applicable. It is suggested instead that more contextualized and place sensitive approaches to corporate responsibility should be practiced in order to ensure greater inclusivity and to avoid the marginalization of some stakeholder groups. This view about the Western centric nature of much corporate responsibility dovetails with the argument that the assumptions of corporate social and environmental responsibility are neither sufficiently acknowledged nor challenged. In a particularly incisive and infuential critique, Mick Blowfield argues that the fundamental values of capitalism concerning profitability, free trade, the freedom of firms to disinvest, private property, commodification, and market power are unaffected by the vast majority of corporate responsibility initiatives, with programs only addressing what he refers to as the more negotiable aspects of corporate activity. This point is exemplified by the limitations of labor codes applied by retailers and brand manufacturers in their global supply chains. Some aspects of the codes are measurable and relatively practical to monitor and report, such as ensuring no child or slave labor. Other clauses concerning the payment of a living wage and the prevention of discrimination are more difficult to implement due to debate over definitions and challenges associated with monitoring, respectively. Moreover, in line with Blowfield's critique, the codes fail to address both the right of a retailer to terminate contracts with suppliers and the power of retailers' purchasing practices driven by ever more intense demands for greater volumes of product at ever lower prices. There is a concern that the failure of corporate responsibility initiatives to address some of the core tenets of capitalism may leave gaps in the governance of labor rights and environmental protection that, as a result of corporate self regulation displacing state legislation, are simply not addressed at all. Overall, the corporate responsibility movement has become a key part of economic governance on a global scale since the early 1990s. There are problems, however, associated with the movement that are bound up in prevailing processes linked to globalization, capitalism, and governance.

Corporate Responsibilities: An Economic–Geographical Research Agenda

The task of engaging with and understanding ongoing developments in the corporate responsibility movement is clearly an interdisciplinary one. However, it is also one in which geographers can play and are playing a sig nificant role. In terms of future directions for geo graphical research on this subject, there are three key areas that might be considered as priorities. First, there is room for a deeper understanding of the ways in which aspects of environmental responsibility are incorporated into business strategies and management systems. Sitting at the interface of the physical and social sciences, geography as a discipline can potentially offer insight into these processes, particularly at a time when en vironmental issues are rising so high up political and economic agendas. Second, there is scope for geographers to increase their contribution to knowledge on the pro cesses that link new trends in ethical consumption with developments in corporate responsibility. Research in the two areas has tended to be conducted separately, but understanding the networked connections between sociocultural spaces of ethical consumerism and the economic spaces of the business world is important given the power and authority of consumers and consumption in advanced capitalist economies. Third, given the cri tique concerning the Western centrism of the corporate responsibility movement, more critical studies of trans national forms of corporate responsibility are required in which geographers can play a role. Acknowledging that such corporate responsibility initiatives need to engage more deeply with noncorporate actors in the global South and their lifeworlds, methodologies associated with participatory research and theoretical frameworks linking insights from postcolonial theory with concepts of the capitalist space economy offer potential for advancing critical debate on this theme. Geographers have been recently active in these methodological and theoretical developments. In sum, addressing the spatiality of the concepts and practices that drive corporate responsibility represents an important contribution to understanding and engaging with what has become a very powerful political–economic movement.

See also: Commodity Chains; Environmental Regulation; Global Commodity Chains; Transnational Corporations in Developing Countries.

Further Reading

Blowfield, M. (2005). Corporate social responsibility: Reinventing the meaning of development. *International Affairs* 81, 515 534.
Blowfield, M. and Frynas, J. G. (2005). Setting new agendas: Critical perspectives on corporate social responsibility in the developing world. *International Affairs* 81, 499 513.
Clark, G. L. and Hebb, T. (2005). Why should they care? The role of institutional investors in the market for corporate global responsibility. *Environment and Planning A* 37, 2015 2031.
Friedman, M. (1962). *Capitalism and Freedom*. Chicago, IL: University of Chicago Press.
Gouldson, A. (2006). Do firms adopt lower standards in poorer areas? Corporate social responsibility and environmental justices in the EU and US. *Area* 38, 402 412.
Hughes, A. (2001). Global commodity networks, ethical trade and governmentality: Organizing business responsibility in the Kenyan cut flower industry. *Transactions of the Institute of British Geographers* 26, 390 406.
Jenkins, R. (2005). Globalization, corporate social responsibility and poverty. *International Affairs* 81, 525 540.
Jenkins, R., Pearson, R. and Seyfang, G. (eds.) (2002). *Corporate Responsibility and Labour Rights: Codes of Conduct in the Global Economy*. London: Earthscan.
Matten, D. and Crane, A. (2005). Corporate citizenship: Toward an extended theoretical conceptualization. *Academy of Management Review* 30, 166 179.
Oglesby, E. (2004). Corporate citizenship? Elites, labor and the geographies of work in Guatemala. *Environment and Planning D: Society and Space* 22, 553 572.
Prieto Carrón, M., Lund Thomsen, P., Chan, A., Muro, A. and Bhushan, C. (2006). Critical perspectives on CSR and development: What we know, what we don't know, and what we need to know. *International Affairs* 82, 977 987.
Sadler, D. (2004). Anti corporate campaigning and corporate 'social' responsibility: Towards alternative spaces of citizenship? *Antipode* 36, 851 870.
Thompson, G. F. (2005). Global corporate citizenship: What does it mean? *Competition and Change* 9, 131 152.
Welford, R. (2000). *Corporate Environmental Management 3: Towards Sustainable Development*. London: Earthscan.
Zadek, S. (2001). *The Civil Corporation: The New Economy of Corporate Citizenship*. London: Earthscan.

Relevant Websites

http://www.fairlabor.org
Fair Labor Association.
http://www.fairtrade.net
Fairtrade Labelling Organizations International.
http://www.ilo.org
International Labour Organization.
http://www.iso.org
International Organization for Standardization: ISO System of Standards and the ISO 14000 Environment Management Standard.
http://www.ethicaltrade.org
UK's Ethical Trading Initiative.
http://www.unglobalcompact.org
UN Global Compact, an International Initiative to Encourage Corporate Responsibility.

Corridor and Axis Development

L. Albrechts, Catholic University of Leuven, Leuven, Belgium
T. Tasan-Kok, Delft University of Technology, Jaffalaan, The Netherlands

Glossary

Functional Axis Similar economic activities that form linear spatial clusters.
Linear Development Linear (spatial) development.
Mega-Structures Large-scale spatial development patterns.
Polycentric Development Multicentered spatial developments developing as functional spatial activities ranging from transnational to intra-urban scales.
Ribbon Development A built-up area along a road running outward from the city center.

Introduction

The terms 'corridor' and 'axis development' carry a variety of meanings. The differences relate to the function (urbanization corridors, ecological corridors, transportation corridors, and economic development corridors), views (geographers, ecologists, spatial policy planners, transport engineers, economists, etc.), and scale (from local to global). Corridor and axis refer to linearity and, in some cases, to continuity. The idea of linearity can be traced back to the urban utopias of the late nineteenth century. The early references (by architects, engineers, policymakers, geographers, and planners such as Soria y Mata, Milyutin, Leonidov, Ginzburg and Baršc, and Benoit Lévy) introduced the corridor as a physical linear (spatial) development pattern. While these early references introduced corridors and axes implicitly as continuous and linear urban forms, with the passing of time the concept of corridors and axes took on other dimensions (from spatial pattern to development tendency), scales (micro to macro), and meanings (virtual, analytical, policy, etc.). The meaning of corridors and axes shifted in the twenty first century from a reference to the spatial development pattern as a continuous and linear urban form to a variety of meanings such as developments linked to the spatial concentration of economic activities, polycentric urban development, multimodal transport bundles, linkage between habitat patches to allow the migration of species, etc. In this article, we organize the different meanings under two main categories, namely, analytical and policy approaches, where scale plays an important role. The scale includes micro to macro scale developments (from the inner city to interregional or cross border transnational developments).

Although corridor and axis seem to refer to similar types of development and come from the same family of linear development, there is a slight difference between the two terms. Generally, corridor development refers to the development of macro scale settlement structures whereas axis development refers to micro scale linear continuity. In other words, axes are concentrations of urban functions along a linear pattern whereas corridors refer to patterns of spatial development.

The sections below focus on three main topics for the purpose of putting corridors and axes in context. The opening section introduces the historical development of the concept with an introduction to the birth of the idea of linearity, while the next section focuses on criticism of corridor and axis development. In the final section, types and different interpretations of corridor and axis development are organized under the headings of analytical and policy categories.

Brief History of the Development of the Concepts of Corridor and Axis: Birth of the Idea of Linearity in Urban Utopias

Although conceptually defining an urban form, linearity has become associated with corridor development, a term which is used to define large scale development patterns (either physically or virtually). Their linear structure is in fact the most distinctive characteristic of corridors. Going back in history, the birth of 'linearity' was seen as an ideal solution for development and transport issues. The idea of a linear city may have earlier roots than the beginning of the nineteenth century, but the first one to mention linear development as an urban development model is the Spanish urbanist Arturo Soria y Mata (1844–1920). Soria y Mata's 'linear city' was intended to integrate urban and rural development through a linear development stretch, projecting an extended urban belt parallel to the transport lines. Although his linear city–region dream, structured along road and streetcar infrastructure, was never realized, a linear suburban development around Madrid was realized as a small scale example (**Figure 1**). However, Soria y Mata did influence many architects internationally at the beginning of the twentieth century, including the Soviet de urbanists and utopians who used the concept of linearity (**Figure 2**). Nikolai Alexander Milyutin, for instance, adopted the idea of linearity and justified placing production enterprises and schools in the same band in the 1920s, while Ivan Leonidov aimed at an overall linear structure with bands of transportation for industrial zones and for other technical and cultural infrastructure to meet the needs of the new socialist republic with his Magnitogorsk project of the 1930s. In 1930, the architects Ginzburg and Baršc designed a 'green city', with the principle of settling the city's working people along the main arteries connecting Moscow and the nearby towns. Although these plans were never fully realized, linearity has become an important element in the theory of urban systems (**Figure 2**).

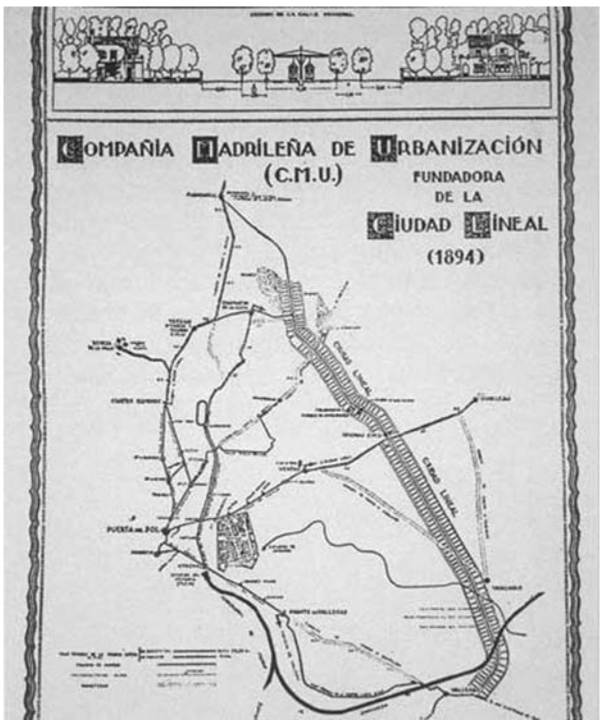

Figure 1 Historic Madrid. Source: http:// www.madridhistorico.com/seccion7 enciclopedia/ index enciclopedia.php?id C&idinformacion 175&pag 8

Criticism of Corridor and Axis Development: Linearity versus Compactness

The idea of linearity collides with the idea of the compact urban form. Industrial cities of the nineteenth century displayed a relatively compact form, which was easy to distinguish from the rural areas and other cities. However, the dynamics of twentieth century cities have resulted in a tendency toward decentralization, economic growth, and outward population growth of the urban areas since the 1960s. Many (planners) believe that the compact city with an optimal density should replace suburban sprawl as the dominant development pattern in the future. From this point of view, corridor and axis development are criticized as they are associated with the decentralization of certain functions.

In Central Europe the decentralization of the population started in the 1960s and 1970s, sometimes in the form of new satellite towns, high rise suburbs, or leapfrogging urban sprawl. The decentralization of economic activities started in the late 1970s, first for retail activities, and later also for manufacturing. The process of decentralization was supported/made possible by changes in the economy, an explosion of personal mobility, and the emergence of new lifestyles. Peter Hall summarizes some of the criticism of these developments: the waste of land, the profligate use of natural resources, the pollution, and cost in lives resulting from the dependence on the private car, and the homogeneity of standards brought on by suburbanization. Many planners of the twentieth century defined the 'concentration of urbanization' as a basic principle in efforts to counteract the development patterns (urban sprawl, ribbon development along traditional road infrastructures, etc.) resulting from decentralization. Different concepts were proposed for a more sustainable and livable urban form (not only compact cities, but also 'pearls on a string' along transport

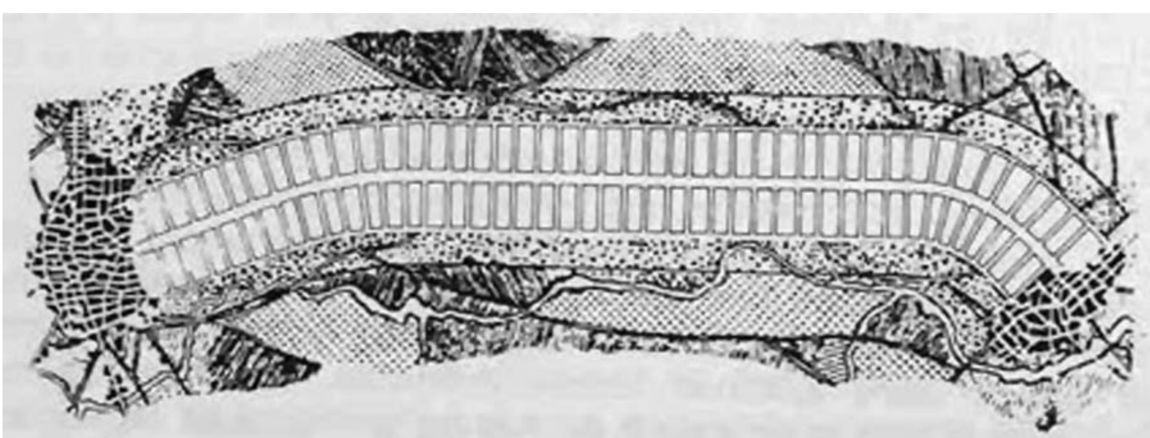

Figure 2 Illustrations from Soria y Mata's Ciudad Lineal (Linear city) in Madrid. Source: www.wikipedia.com

axes). The primary goals were to save the landscape and to reduce car mobility.

In North America – where land is abundant, trans portation costs are low, and the economy has become dominated by service and technology industries – dis persed urban land development patterns connected by major transportation routes have been dominant over the last 50 years. Decentralization of activities resulted in two opposite effects: first, the commuting time has remained relatively constant and second, the commuting distance tends to be longer and to be covered using the auto mobile rather than public transport as the dominant mode. As a result of this permanent outward moving process, suburban highways are often as congested as urban highways, defining corridors of massive commuting traffic between cities and suburbs and cities.

Types and Different Interpretations of Corridor and Axis development

Corridor and axis development have different meanings that are interrelated but cannot be distinguished from one another very easily. We can differentiate the approaches to corridors in terms of two main categories: (1) corridor and axis are approached as analytical concepts and (2) corridor and axis are approached as policy concepts.

The analytical approaches may lead to policy dis courses in which the multiple interpretations at various spatial scales are used. Scale plays an important role in both categories. Developments occurring either on the macro scale or on the micro scale display different characteristics in terms of the appearance or functionality of the corridor and axis developments as illustrated below. Micro scale here means the intra urban scale, whereas macro scale refers to a range of developments from inter urban to interregional and transnational.

Analytical Approaches

In this category, the corridor and axis development concept is generally used to explain spatial development patterns in a given space, where a range of activities take place with a varying geographical reach involving a variety of actors. Accordingly, three analytical categories can be defined:

1. corridors as structuring elements for large scale de velopment patterns (mega structures);
2. corridors as linear urban forms; and
3. corridors as functional axes.

Corridors as structuring elements for large-scale development patterns (mega-structures)

In this category, the term 'corridors' refers to structuring elements of large scale development patterns, as we may call the mega structures that are shaped by the processes of the growth of cities, the division of labor, massive subur banization and sprawl, and technological development. These processes have given rise to conglomerations of urban and suburban places spread over a large area. Cor ridors can be defined here as lines, as begin–end or trunk feeder networks, or as physical forms that resemble networks such as backbone or ladder networks, string of pearls, etc. The number and functions of the nodes on these networks may differ with the scale level of devel opment. For instance, on the highest, transnational scale it could be an ecumenopolis (such as Beseto Ecumenopolis) or at interregional scale a mega corridor, such as the so called 'blue banana' (a European core area from London to Milan), which can be considered a (virtual) line network or a string of pearls. Moreover, at the transnational scale another example would be the triangle between Lille, Amsterdam, and the Ruhr area with an extension south eastward to Frankfurt, the Northwest European Delta Metropolis, or at the interregional scale there is a large, hybrid urban pattern of urban concentration areas (Rand stad and Flemish Diamond), with suburban zones and open spaces. Some other examples can be categorized as de scribed below.

Ecumenopolis

This analytical term refers to developments at the transnational or interregional scale. According to this utopian idea of Doxiadis, a continuous settlement structure would lead to a universal city, where metrop olises and megalopolises would become interconnected to form a worldwide city. Such cities would connect within a single continuous network. This concept will be demonstrated in East Asia if the urban corridor between Beijing, Seoul, and Tokyo (Beseto Ecumenopolis) con tinues to grow. This enormous region includes five mega cities (Beijing, Tianjon, Seoul, Tokyo, Yokohama, and Osaka Kobe) within the 1500 km strip of four densely populated countries. Doxiadis also suggested that tradi tions and trends of urban development may lead to the first European *eperopolis* (continent city), which would cover the London–Paris–Amsterdam area.

Mega-corridors

In this sense, corridors are physical and socioeconomic links between distant large urban agglomerations (inter regional as well as transnational in scale) with comple mentary strengths as a means to achieving interregional cooperation. These are usually referred to as mega cor ridors, which is an analytical concept developed to examine to what extent networks are spatially linked. Large urban regions in Northwest Europe – the Rand stad, the Flemish Diamond, the RheinRuhr area, Lille, Paris, London, and the West Midlands – are examples of this type of corridor. Some developments in Southeast Asia such as Kuala Lumpur–Klang Valley, or the corridor between Batam, Indonesia–Johore, and Malaysia–Singa pore are examples of such mega corridor developments.

Megalopolis

A continuous and densely populated development was described as a megalopolis by Jean Gottman, a French geographer, to define the belt on the east coast of the United States in the beginning of the 1960s. The term refers to interregional developments where metropolitan regions or large urban corridors are structured by transport infrastructures and have developed when urban areas grew together.

Polycentric development

Although the concept dates back to early twentieth century work of the Chicago School, in the beginning of the twenty first century Simin Davoudi's work has shown that polycentricity is still lacking a clear definition. It refers to different analytical terms such as a planning tool, a spatial structure, or a policy goal for achieving a balanced regional development or a lifestyle. Concerning large scale development patterns (mega structures), polycentricity can be defined as multicentered spatial developments as functional spatial entities that refer to a wide range of activities from transnational to inter regional, intercity, and even intra urban scales.

Polycentricity can be within a region (inter urban scale), that is, the organization of the urban systems and networks within a functional region; it can be at the interregional scale, that is, the organization of urban systems and networks within a wider region going be yond national borders; or it can be at the transnational scale such as European polycentricity, that is, urban systems and networks at European level. Although early references to the concept of polycentricity at the inter urban scale date back to the 1930s, Larris and Ullman's study in 1945 contained the first reference to the term 'polycentricity'. They claimed that cities grow around several economic centers and evolve as multi nuclei. In fact, in the 1970s the urban structures shifted from the monocentric to the polycentric model, a fact which was mainly related to the rapid decentralization of economic activities and the increasing mobility of people. This development at the intra urban scale caused a fragmented spatial distribution of activities, which were no longer concentrated around a single core (or a single center). Changing economic rela tionships between firms and the distribution of em ployment also created the complex commuting patterns of people between job and home. Moreover, due to the increasing popularity of post Fordist (flexible) pro duction systems, economic activities began to cluster in several centers. The post Fordist production model began to be a widespread phenomenon from the late 1970s onward. In this system, the production process became fragmented as individual firms specialized in their areas of expertise and began to be concentrated around industrial districts or clusters of integrated firms – like Silicon Valley (USA), Jutland (Denmark–Germany), Småland (Sweden), and several parts of Italy. Within this framework, urban systems became more complex and multinodal.

Corridors as linear urban forms

This approach highlights the developments mainly at the intra or inter urban scales. The relationship between transport and movement within cities and their regions has been important for the development of cities and the urban form throughout history but this relationship began to change at the end of the nineteenth century. Ribbon development at the intra urban scale, for instance, gen erated great concern in the UK during the 1920s and 1930s, as well as in numerous other countries. Ribbon development refers to a built up area along a road running outward from the city center. This type of development is characteristic of many Midwestern settlements in the USA. In the mid twentieth century we realized that the analysis of urban form, structure, and pattern had become an important area of interest among planners (e.g., Lynch, Webber, etc). Corridor and axis development was not yet directly referred to at the beginning, but initial steps were made toward defining corridors as linear urban forms in those years. Harris and Ullman brought the element of lineation into the discussion of urban patterns with their well known article on 'The nature of cities in the USA around 1945'. Linear concepts also became fashionable in urban planning during the post World War II era in the UK, although implementation was very limited.

In the late 1960s, Doxiadis provided a systematic analysis of urban form and structure and argued that certain forces (natural and man made) shape urban sys tems. He emphasized that linear forces tend to form linear settlements as long as these settlements are small and do not need any common services. Doxiadis was perhaps also among the first to analyze and categorize the forms of settlements on the basis of their essential structures, characterized in terms of circular and (geo metric and nongeometric) linear forms. In those years, 'corridor' meant an inter urban scale development with the linear patterns of the highly developed bundles of transport routes joining the major towns. Although lim ited, ideas surrounding linearity were also reflected in the actual plans in the UK in those years. Caersws new town plan in Wales, for instance, suggested series of villages along a new major road, which would be paral leled by local roads joining the main road at junctions.

Corridors as functional axes

Corridor also refers to discontinuous zones of focal points at the interregional scale, largely corresponding to the exchange points of transport flows, rather than to 'corridor' as a continuous ribbon of industrial sites. Here, the spatial clustering of certain functions is gaining in importance. Silicon Valley in California (USA) is a good example of this type of functional concentration axis. Historically dating back to the late 1930s, Silicon Valley is an axis of high tech firms that are located in the Santa Clara Valley at the southern end of San Francisco Bay. Silicon Valley originally offered a high concentration of semiconductor industries in the corridor, which have shifted in recent years toward software innovation and Internet and computer related service activities. Another similar case is the M4 corridor adjacent to the M4 motorway, to denote the far eastern part of the motorway from Swindon to London. The M4 corridor is home to a large number of high tech businesses such as Microsoft, Cisco, and Hewlett Packard.

On a very different scale, ecological corridors can also be looked upon as functional axes. They refer to linear green patterns/areas. Ecological corridors can be defined as linear green axes where animals can travel, plants can grow and spread continuously, and threatened species can be replenished from other areas. In this sense, an eco logical corridor may define a transnational scale linear green axis, which can relate to a cross border develop ment, where small scale green patches can be connected to sustain the continuity of green layers. Ecological corridors can be considered as linear landscape features. Ecological stepping stones are similarly considered to be a linear arrangement of habitat patches. Ecological cor ridors cover broad areas of habitat, either continuous or discontinuous, connecting relatively closely spaced pat ches of green. Examples of ecological corridors can in clude river basins or valleys; forests at the base of large mountains; the banks of streams, canals, rivers, and small ponds; and the strips running along the sides of roads. It is important to distinguish ecological corridors as phys ical features that provide connectedness in green land scape by physically connecting core habitats, and the biological attributes of landscape connectivity. Ecological corridors may be defined differently in each country context. Australian and South African ecological cor ridors, for instance, are defined on the basis of strips of remnant forest along roadsides and river banks. However, in the European context ecological corridors are defined for the purpose of promoting substantial habitat restor ation, and these need to be distinguished as restored corridors.

Policy Approaches: Corridors and Axes in Spatial Planning and Urban and Regional Policy Documents

Corridors are recognized in spatial planning and regional policy documents (drawn up by individual countries or by groups of countries), in which they are seen as an element of balanced and integrated regional (and trans national) development within different spatial policy contexts. Accordingly, four normative categories can be defined:

1. the corridor as an instrument of economic develop ment policy;
2. the corridor as an instrument of (transnational) spatial development policy;
3. the corridor as an instrument of transportation policy; and
4. the corridor as an instrument of ecological (or en vironmental) policy.

Although policy documents quite often refer to corridor and axis development, it seems to be a big step to move from the analytical framework to the implementation of policy. Moreover, the cross border developments – such as transnational transportation corridors, mega corridors, polycentric developments, ecological corridors, etc.– that require cross border cooperation require new governance approaches and regulations to deal with these types of development. In Europe, recent attempts to establish a more loosely organized structure in the Delta Metropolis provide an initial example of the establishment of such governance institutions. The spatial development policy of the European Union (EU) gives special attention to a variety of corridor developments that are defined ana lytically above.

The corridor as an instrument of economic development policy

Individual countries define specific spatial strategies at different scales, ranging from the micro to the macro (urban, regional, and national) scales, where interregional transportation networks, regional–spatial development corridors, or inner city functional corridors make sense. The Multimedia Super Corridor (MSC) of Malaysia and Guangzhou China are good examples of these.

The MSC (Malaysia), a 50 km long high tech indus try zone stretching southward from Kuala Lumpur, offers a 15 km wide concentration of global information and communication technology (ICT) industries. Started in 1996, the MSC Malaysia has become a dynamic ICT hub, hosting more than 900 multinationals and foreign and domestic companies focused on multimedia and com munications products, solutions, services, and research and development.

South China's Guangdong Province, one of the most developed areas in the country, built an information in dustrial corridor in the Pearl River Delta during the period 2001–05. The corridor focuses on information technology, which was selected as a pace setting and priority industry, with Shenzhen and Guangzhou cities as leaders. The provincial capital of Guangzhou will stress software development; Shenzhen, a special economic zone, will prioritize high technology; and Dongguan will endeavor to build itself into the world's largest computer production center.

The corridor as an instrument of (transnational) spatial development policy

Policy documents on cross border cooperation also de fine corridor development at the macro level. In the EU, for instance, the tendency is to implement an indicative spatial development perspective, where individual countries are invited to contribute to the organization and development of transnational (spatial, infrastructural, etc.) networks. For instance, European Spatial Develop ment Policy (ESDP) is an organized and integrated spatial policy at a transnational level. In contrast, the United States has almost no strategy to anticipate and manage comparable concerns throughout the country. Recent academic research (carried out by the Lincoln Institute, New York), however, aims to define conditions for an American spatial development policy. This policy would define the megalopolis as a model for cooperation among the cities and regions in the US that are currently growing together and creating diseconomies within a matrix of congested transportation networks, thus cre ating a new urban form. However, at the policy level this project has not yet been realized.

The development of a balanced and polycentric urban system and a new urban–rural relationship is one of the targets of the territorial development strategy of the ESDP, which defines the corridor concept as a tool to reconcile economic growth, competitiveness, and sus tainable development. ESDP offers a geographical image of the European economic space – a polycentric urban system, linked by integrated communication corridors, and separated by protected environmental and water catchments. Pan European or Euro corridors, for in stance, are defined by the ESDP.

The corridor as an instrument of transport policy

The transport connections between important economic regions (mega urban regions) define the corridors as transport axes. The trans European networks, for in stance, are part of a wider system, including a tele communications network and a proposed energy network. In addition to the various trans European net works, there are ten pan European corridors, which are paths between major urban centers and ports. In order to create a dense network of feasible transport infrastructure throughout Europe, the European Commission contrib utes to the organization and development of Trans European Networks (TENs) in the areas of transport, telecommunications, and energy supply infrastructure. TEN is a European scale program that will interconnect the national transport infrastructures to create a genuine trans European network.

In this context, railway networks offer important po tential for easing traffic on major trans European roads and for connecting urban regions. The five railway axes within the EU are intended to define the transport pri orities: the Paris–Bratislava axis, which will improve connections with the new member states; the Lyon–Budapest axis, also called the East–West corridor south of the Alps; the high speed southwest Europe axis, con necting France to the Iberian Peninsula; the Berlin–Pa lermo axis, also called the North–South corridor of Central Europe; and Rail Baltica, which will connect Helsinki to Warsaw via the Baltic states.

The corridor as an instrument of ecological (or environmental) policy

Ecological corridors are usually seen as important elements of planning policy at several scale levels. In Europe, for instance, several organizational bodies and institutions have been established for protecting and managing ecological corridors. For instance, the Pan European Ecological Network aims to ensure appropriate interconnectivity between the selected core (green) areas that provide the optimum achievable quantity and quality of environmental space at the transborder scale. Here, the corridors are seen as instruments to ensure that species populations have adequate opportunities for dispersal, migration, and genetic exchange. Adopted by the Council for the Pan European Biological and Landscape Diversity Strategy in 1999, the guidelines for the development of this ecological network take the form of a reference docu ment for use by actors, including policymakers, parlia mentarians, natural resource managers, spatial planners, researchers, the academic community, representative organizations, private enterprises, and members of non governmental organizations. Within this framework, several transborder initiatives have been taken and pilot projects have been implemented such as the Transborder Cooperation between Estonia and Latvia on Trans border Ecological Network Development and Water Resource Management; the Development of the Trans boundary East European Model for the Polesie: the Belarus–Ukraine–Poland Regional Ecological Network; and the cross border ecological network of the Bialo wieza Forest Region (between Poland and Belarus). In these initiatives, not only have ecological corridors been defined as instruments for protecting certain green areas, but cross border policy exercises between various countries have been implemented.

See also: City-Region; Networks; Planning, Urban; Polycentricity; Urban Growth Machine.

Further Reading

Albrechts, L. and Coppens, T. (2003). Megacorridors: Striking a balance between the space of flows and the space of places. *Journal of Transport Geography* 11, 215 224.

Chapman, D., Pratt, D., Larkham, P. and Dickins, P. (2003). Concepts and definitions of corridors: Evidence from England's Midlands. *Journal of Transport Geography* 11, 179 191.

Davoudi, S. (2003). Polycentricity in European spatial planning: From an analytical tool to a normative agenda. *European Planning Studies* 11(8), 979 999.

Doxiadis, C. A. (1968). *Ekistics: An Introduction to the Science of Human Settlements*. London: Hutchinson.

Gottmann, J. (1961). *Megalopolis: The Urbanized Northeastern Seaboard of the United States, xi, 810 p. illus., maps, tables, 25 cm*. New York: Twentieth Century Fund.

Harris, C. D. and Ullman, E. L. (1945). The nature of cities. *Annals of the American Academy of Political and Social Science* 242, 7 17.

Losch, A. (1954). *The Economics of Location*, translated by Woglom, W. H. New Haven, CT: Yale University Press.

Lynch, K. (1960). *The Image of the City*. Cambridge, MA: MIT Press.

Whebell, C. F. J. (1969). Corridors: A theory of urban systems. *Annals of the Association of American Geographers* 59(1), 1 26.

Relevant Websites

http://www.madridhistorico.com
Historic Madrid: Encyclopedia.

http://w3.cnice.mec.es
The Linearity: In Spain, the architect and town planner Arturo Soria y Mata, inspired by the idea of Howard of the garden city, designed model linear city.

Cosgrove, D.

K. D. Lilley, Queen's University Belfast, Belfast, UK

Denis Cosgrove (1948–2008)

Denis Cosgrove (**Figure 1**) was born in 1948 and educated in Liverpool at St. Francis Xavier School and then St. Catherine's College, Oxford, subsequently taking up posts at Loughborough University and then Royal Holloway and Bedford New College (University of London). In 2000 he became the first Alexander von Humboldt Professor of Geography at University of California, Los Angeles (UCLA). Cosgrove was one of the foremost cultural geographers of his generation, putting 'landscape' firmly back on geography's disciplinary map in the 1980s and 1990s with a series of theoretically informed and challenging books and articles, particularly *The Iconography of Landscape* (edited jointly with Stephen Daniels in 1988). His conceptualization of landscape proved very influential, in particular, the idea that landscape is a 'way of seeing' rooted in Western culture and in Renaissance Europe, as well as exploring the idea of landscape through empirically and theoretically informed studies. Cosgrove also contributed to debates on the 'visual' in geographical thought and representation, using landscapes, architecture, art, maps, and textual and pictorial imagery as a means of questioning and examining human relations with constructed

Figure 1 Denis Cosgrove.

environments. To this end, he worked within a broadly defined cultural materialism forged within an arts and humanities tradition, as was evident in his work in the 1980s, such as *Social Formation and Symbolic Landscape*, a book that followed cultural theorist John Berger in regarding landscape as 'an ideological concept'.

Some measure of Cosgrove's impact on the way that geographers have come to think of landscape as a cultural construct can be seen in the textbooks of the 1990s and 2000s, for in these, landscape has become conceived of as 'a composition and structuring of the world', not neutral or inert, but a map of complex social and cultural meanings. This differs from the landscape geography of a previous generation, a landscape 'history' in the tradition of those following W. G. Hoskins and C. O. Sauer. As a result, 'landscape' has itself become a contested terrain among geographers (and others concerned with landscape), with some arguing that too much emphasis is placed on represented or imagined landscapes, rather than the physical and material, and others questioning this very distinction. Certainly, much of Cosgrove's interest was on issues of representation, and through them he has shown how author and authority need to be taken seriously when it comes to trying to understand texts not only in a literal sense, but also 'text' in a metaphorical sense (which includes landscapes as representations as well as representations of landscapes). The concern for metaphor in text and image related to Cosgrove's continued interest in art, architecture, and landscape. He undertook innovative collaborative work with artists, photographers, and landscape architects, closing the gap between those involved in landscape praxis and those in academe. This took the form of work on mapping, evident, for example, in his edited *Mappings* as well as exhibitions and installations, and also through the contributions Cosgrove made, particularly in the early 1990s, to geographical debates on textuality and subjectivity that shaped the discipline's 'cultural turn' – heralding the emergence of what was to be termed a 'new cultural geography'.

Throughout his academic career Cosgrove was influenced by disciplines beyond geography, such as literary studies and architectural history, which informed his thinking on the relationships existing between the humanities and geography. This was given particular intellectual momentum in 1993 with the foundation of a new journal, *Ecumene, A Geographical Journal of Environment, Culture and Meaning* (now *Cultural Geographies*), established under his editorship, fostering a certain brand of cultural geography that was both historically and geographically rooted and placed, yet at the same time theoretically informed and critical. This was very much the essence of Cosgrove's own work, characterized by a blend of both the empirical and the theoretical. It is to be seen, for example, in his work on Palladian landscapes of six teenth century Italy, as well as later dealings with car tography and imagined geographies of the Renaissance and Enlightenment. This work, itself emerging from a humanistic Europe, was still concerned with those issues that informed the 'new' cultural geography of the 1980s and 1990s, namely, the idea that the world is human centered, subjectively constructed, contextual, and con tingent – yet also structured along axes of power. What marked Cosgrove's work out as especially important was his ability to delve deeply into particular periods and subjects through archival study of primary and secondary historical sources, yet retained a broad perspective that – even if the reader was not too concerned with the em pirical details – drew out a wider significance that con nected to philosophical and theoretical questions, both ontological and epistemological. At the close of the 1990s, at a time when so much of human geography had become either overtly theoretical in approach or ahis torical in content, Cosgrove's balance of the particular and the general, the empirical and the theoretical, the historical and the contemporary, provided a salutary re minder of how geographers can contribute not only to their own discipline's development and interests but also to a much greater project, of knowledge creation and human understanding in which geography is central. His death in 2008 after a sustained battle with cancer robbed cultural geography of one of its most admired and well respected figures.

See also: Cultural Geography; Historical Geography; Humanism/Humanistic Geography; Landscape.

Further Reading

Cosgrove, D. E. (1984/1998). *Social Formation and Symbolic Landscape*, London: Croom Helm (2nd edn.) with additional introductory chapter. Madison, WI: Wisconsin University Press.
Cosgrove, D. E. (1993). *The Palladian Landscape: Geographical Change and Its Cultural Representations in Sixteenth Century Italy.* London: Leicester University Press.
Cosgrove, D. E. (ed.) (1999). *Mappings*. London: Reaktion Books.
Cosgrove, D. E. (2001). *Apollo's Eye: A Cartographic Genealogy of the Earth in the Western Imagination*. Baltimore, MD: Johns Hopkins University Press.
Cosgrove, D. E. and Daniels, S. (eds.) (1988). *The Iconography of Landscape: Essays on the Symbolic Representation, Design and Use of Past Environments*. Cambridge: Cambridge University Press.
Freitag, T. and Jons, H. (2005). Vision and the cultural in geography: A biographical interview with Denis Cosgrove. *Die Erde* 136, 205 216.
Lilley, K. D. (2004). Denis Cosgrove. In Hubbard, P. J., Kitchin, R. & Valentine, G. (eds.) *Key Thinkers on Space and Place*, pp 84 89. London: Sage.

Relevant Websites

http://www.geog.ucla.edu/cosgrove.php
http://www.gg.rhul.ac.uk/cosgrove/

Cosmopolitanism

J. Binnie, J. J. Holloway, S. Millington, and C. Young, Manchester Metropolitan University, Manchester, UK

Glossary

Commodification To turn something into, or treat something as, a commodity.

Enlightenment An eighteenth-century philosophical movement stressing the role of 'scientific' thinking and a belief in human reason as the basis of modernity.

Nationalism Devotion to the interests or culture of one's country.

Normative A set of sociocultural values that seek to prescribe a certain way of doing things.

Patriotism Love of, and devotion to, one's country.

Transnationalism Various processes through which people, principally migrants, produce and sustain complex relationships across state boundaries.

Introduction

Cosmopolitanism is a broad ranging term in socio political and moral philosophy which has been the subject of debate in human geography and the social sciences. It is a highly contested term which has been conceptualized and defined in different ways. Cosmopolitanism is generally associated with the concept that all of humanity could (and should) belong to a single community. Someone who adheres to such cosmopolitan principles or exhibits cosmopolitan characteristics (however they are defined) is referred to as a 'cosmopolite' or 'cosmopolitan'. However, the exact nature of this community, and how it could be constituted and created, has been the subject of much debate. Should it be based, for example, on new global political and institutional forms and new market relations which are open to all, or particular moral norms or new inclusive relationships with others, or the production of cultural expressions appreciated by all citizens? A dominant perspective in many, but not all, cosmopolitan stances is an emphasis on getting beyond forms of social interaction which limit the possibilities of achieving this single community. Thus, cosmopolitanism is often (but not necessarily correctly) conceived of as an alternative to patriotism and nationalism and as critical of parochial attachments to place. Instead, emphasis is placed on the search for more inclusive forms of social relations between 'global citizens' and a global outlook and openness to difference. Cosmopolitanism thus has relevance to a wide range of geographical issues including: geopolitics, governance, and territoriality; globalization and new transnational social movements; international migration, multiculturalism, and living with difference; the branding and marketing of cosmopolitan space and cities; and moral and ethical geographies. While debates have often focused on abstract philosophical discussions, more recent research, particularly within human geography, has examined how notions of cosmopolitanism are constructed and deployed in particular sociospatial contexts and their social and spatial implications. This research both questions abstract notions of cosmopolitanism and also offers direction for how it may work as a normative project.

Definitions of Cosmopolitanism

The Historical Development of Cosmopolitanism

Definitions of cosmopolitanism are wide ranging and highly contested. Its origins are attributed to classical Greek philosophers who debated notions of political culture and citizenship. These debates predominantly saw citizens' allegiances as tied to the polis or city. However, they also included the principles of relationships that citizens should perform with regard to 'others', particularly foreigners residing within their cities. Thus, there was a tension in classical Greek philosophy between ideas about citizens' ties to (local) politics and about citizens' ties to humanity more generally. This was reputedly stated more explicitly by Diogenes (*c.*412 BC), who, in response to the question 'where do you come from', was said to have replied "I am a citizen of the world." Diogenes' exact meaning is arguable, but this notion was developed in contested ways by classical Greek and Roman philosophers, who debated forms of 'world citizenship' which implied that individuals should do the most they could to help other human beings, regardless of their location.

The Enlightenment saw a revitalization of these varied notions of cosmopolitanism. An increase in global trade, travel, and the 'discovery' of the world, linked to growing philosophical interest in human rights, and the debates surrounding the American (1775–83) and French (1789–99) Revolutions led Enlightenment philosophers to propose a variety of cosmopolitanisms. In everyday usage, the term developed a meaning of an attitude of open mindedness, an interest in travel, and encountering difference as part of a particular lifestyle, as opposed to identification with a particular religious, political, or national affiliation. The philosophical development of the term, however, opposed this individualism. Philosophically, the term came to be based on the acceptance of a moral ideal of a universal human community based on apparently universally shared human characteristics. In turn, various philosophers linked this to the development of political theories of international relations. For some, cosmopolitanism implied the establishment of a single world state and the abolition of nation states. For others, notably Immanuel Kant, the cosmopolitan world political order implied a series of free sovereign states acting voluntarily within a wider world federal structure, with ‘world citizens’ exhibiting an ethical stance of responsibility toward others around the globe, including respect of their human rights. Kant also proposed the concept of ‘cosmopolitan law’ which encompassed the notion that humans have rights and responsibilities as ‘citizens of the world’ and not just as citizens of sovereign states.

In the nineteenth and twentieth centuries, competing conceptions of cosmopolitanism continued to be developed – particularly in the context of the increasing globalization of economic activity. Marxist thought advanced notions of the withering away of the state under Communism and the universal characteristics of labor around the world, potentially leading to international workers’ movements. Economic forms of cosmopolitanism were developed by advocates of global free trade. The need to find ways of maintaining peace in the aftermath of the two world wars, and the growing appreciation of a global interconnectedness in areas such as famine and disaster relief also fostered thinking about global social responsibility. Indeed, authors point to the establishment of international organizations, such as the League of Nations and later the United Nations, the International Red Cross and Red Crescent, and the role of the International Criminal Court in developing the rights and responsibilities of individuals as international citizens, and international responses to major humanitarian disasters as evidence of the emergence of new international institutional forms representing cosmopolitan ideals.

In the twenty first century, some theorists (particularly Ulrich Beck) argue that we are entering a period of global transformation of modernity which can be characterized as the ‘age of cosmopolitanism’. This argument rests on the assertion that the key issues in today’s increasingly globalized world (e.g., a variety of ‘risks’ associated with international trade and labor issues, human rights, migration, the environment and climate change, the ‘war on terror’, and so on) do not lead to the end of the nation state but nevertheless demand other forms of international organization at a range of spatial scales. As a result, it is argued, we must move beyond sociological understandings of the world based on a nation state centered definition of society and politics to a cosmopolitan outlook. This is necessary both to change the nature of social scientific understanding of the world around us and to underpin new forms of social and political action.

Contemporary Cosmopolitanisms

The last 20 years have witnessed a rapid growth in social science literatures drawing on this history of contested notions of cosmopolitanism. The result has been a further round of debates about definitions of the concept. Cosmopolitanism thus remains a highly contested notion which is difficult to summarize, but its main contemporary meanings can be outlined as below, though these definitions cannot easily be divided into discrete categories and this list is not exhaustive.

Cosmopolitanism as a New Sociocultural Condition

For some observers, the late twentieth and early twenty first centuries have been characterized by a qualitative shift in the nature of society, culture, and politics related to globalization. It is important not to suggest that globalization is a homogeneous phenomenon which has already been achieved and is uncontested. However, developments in transportation, communication, travel and tourism, and consumption, among others, are argued to be producing a world which is more cosmopolitan in nature, in the sense that time–space compression is exposing people to cultural diversity on an everyday basis and at a scale not previously seen. What is new is not the intermingling of cultures but global awareness of that intermingling, its associated risks, and potential benefits. Some argue that this is producing a new sociocultural condition, an ‘age of cosmopolitanism’. A more nuanced view suggests that this new era is characterized by a process of cosmopolitanization, in which the increasing interdependence of social actors across national boundaries accompanying globalization is inadvertently producing forms of lived cosmopolitanism, though not necessarily as the outcome of a normative cosmopolitan project (e.g., the engagement of a number of national scientific communities around the issue of climate change).

Some argue that this view is further evidenced by the fact that cosmopolitanism, in various contested ways, is increasingly penetrating everyday discourses in a range of settings, including consumption and tourism, city planning and city marketing, and education and academia. For example, it is now commonplace for people (ranging from city authorities and developers to tourists and consumers) to refer to particular cities or areas within cities as ‘cosmopolitan’, in the sense of possessing a vibrant cultural diversity. Others argue for the importance of education in developing a cosmopolitan ethos, for example, through exposing children to cultural difference, languages, and so on. For some, the existence of this ‘age of cosmopolitanism’ is held to require a shift in the practice of social science itself from a methodological nationalism (in which the social categories of analysis are primarily conceived of in relation to the framework of the nation state) to a methodological cosmopolitanism. In short, what is required is a cosmopolitan turn in the social sciences to adjust to the 'real world' sociocultural condition of the cosmopolitan age.

Cosmopolitanism as Political Project

For some, cosmopolitan thought underpins political projects which seek to improve the lot of humanity. This viewpoint is also related to the increasingly global, or at least international, nature of the various risks and challenges that spill over national borders. This standpoint is moral, ethical, and normative, with the goal of creating a more progressive and just world and a new global democratic order. Thus, some schools of cosmopolitan thought debate the possibilities of creating norms and institutions 'above' the nation state. Commonly cited examples of institutions which already exist and which demonstrate some characteristics of this approach are the European Union, the United Nations, and international aid, human rights, and charitable organizations. While this is sometimes referred to as achieving cosmopolitanism 'from above', other theorists stress the development of cosmopolitanism 'from below', particularly with reference to the emergence of transnational social movements promoting the rights of marginalized social groups. Again, neither of these positions necessarily advocates the end of the nation state but instead they recognize the possibility of citizens having multiple affiliations 'above', 'below', and congruent with states.

Cosmopolitanism as Global Citizenship

This understanding of cosmopolitanism concerns a political geography and philosophy of reframing global citizenship. Again, cosmopolitanism is seen as a normative project underpinned by ideas of universal standards of morality, political action, and law. In this view, cosmopolitanism is closely related to the complex set of processes making up globalization which makes us more aware of the world as a single place and one in which there is an increasing potential for encounters with difference. A variety of risks (terrorism, human rights issues, the environment, and so on) are held to have transformed the spatial scale of political agency, justice, and citizenship toward the global. Thus, citizenship, law, and politics must now be thought of in global terms. Such a position also implies the development of new forms of cosmopolitan public consciousness. Competing interpretations of cosmopolitanism hold that this either means a reduction in the significance of the nation state as these issues become more global, or that national and local scale allegiances can still be significant in this context. This vision of cosmopolitanism as a philosophy of world citizenship has been articulated through the notion of cosmopolitan law which emphasizes individuals and groups as legal entities. Cosmopolitanism thus requires legal mechanisms that act in the interest of global citizens (though this does not mean the end of the nation state). Although often closely related to transnationalism in the literature the two terms are not synonymous, as transnational social or political movements may not be cosmopolitan in their goals or nature.

Cosmopolitanism as an Attitude and Competence

A further commonly held understanding of cosmopolitanism relates to the supposed cosmopolitan characteristic of possessing a particular attitude and disposition toward difference (an interpretation particularly associated with Hannerz). In this view, cosmopolitanism is associated with an openness to, even a desire for and an appreciation of, sociocultural difference. This search for difference by the cosmopolite is both an aesthetic and an intellectual stance. The cosmopolitan is thus held to actively seek out difference or the 'Other', and as a result apparently embraces a global outlook which can involve rejecting their own cultural background and the apparently constricting confines of local communities. The cosmopolite seeks out new forms of identification based on globality and diversity. From this viewpoint cosmopolitanism also encompasses a set of dispositions and forms of practice within and toward diversity. Cosmopolitan practice thus comprises a set of skills which are applied to these encounters with difference. The cosmopolite is held to possess a degree of reflexive ability which allows them to negotiate the risks and pleasures of encountering difference. The skills, knowledge, and transnational connections possessed by the cosmopolitan can also form the basis of their cultural capital. In this way cosmopolitanism has often, but incorrectly, been narrowly equated with particular middle class consumption practices and the lifestyles of global elites.

Critiques of Cosmopolitanism

Given this diversity of contemporary conceptions of cosmopolitanism it is not surprising that it remains difficult to define and open to a range of critiques. Indeed, some authors suggest that the term may lose any real meaning and fail to achieve its progressive potential. These critiques center on the degree to which there really is an ongoing cosmopolitanization process, and how and to what extent the complex and uneven processes that make up globalization are actually linked to the creation of cosmopolitanism. A range of related arguments focus on the question of whether those cosmopolitanization processes that are occurring displace locally embedded sources of allegiance and other sources of sociocultural identity, such as the nation, class, race, gender, and sexuality. Furthermore, in a world of global inequalities, are current manifestations of cosmopolitanism anything more than the products of the decisions of certain empowered global elites? Finally, these explorations of actually existing cosmopolitanism, of multiple cosmopolitanisms shaped in different sociospatial contexts, question the extent to which we can ever realistically expect cosmopolitanism to function as a normative political project.

Globalization and Cosmopolitanism – Is a Global Cosmopolitan Condition Possible?

A first set of critiques of cosmopolitanism focus around the extent to which the existence of a normative political philosophy in the context of globalization can actually produce a sense of cosmopolitanism. Critics of the globalization literature (such as Kennedy), while acknowledging the increasing incidence of the global interrelatedness of people, suggest that this is overemphasized. The proportion of the global population who possess the capital, time, incentive, knowledge, and reflexive skill to think through their global interconnectedness may actually be quite small. Although more and more people are brought into sets of interrelations that transcend national boundaries, this does not mean that they will necessarily develop a sense of global consciousness or empathy with their fellow 'global citizens', although it is important not to rule out the possibility of unreflexive forms of cosmopolitanism. In fact, such knowledge may produce indifference or even hostility toward global 'Others' as a reaction to the fate of those others within the inequalities of globalization or the perception of the possible imposition of other identities and values. Furthermore, most people still live their lives in and through locally embedded networks of social relations and affiliations. This is not to deny the possibility of cosmopolitanism but to suggest caution about naïve visions of the intertwining of a normative political philosophy, globalization, and transnationalism in producing cosmopolitanism. Certain sets of circumstances may lead to the development of particular sets of global interconnections into actual examples of global public spheres or discussions, but this remains relatively rare.

However, other perspectives (particularly Beck and Sznaider) interpret the nature of global interconnectedness more optimistically and argue that the production of global norms does not necessarily rely on the positive promotion of a norm by certain groups. Rather, it is argued that the current climate of increasing awareness of, and concern about, global risks will over time produce an unavoidable need to cooperate at the global scale. The evaluation of global crises and threats to humanity will potentially 'negatively' fuel a shared sense of global responsibility which will produce cosmopolitan norms and agreements and, eventually, political projects in the form of institutional cosmopolitanism, though the difficulties involved in this progression are acknowledged.

Other literatures point to the actual emergence of new global social movements as examples of the production of the specific instances of a globalized sense of responsibility for 'Others', though these movements are often focused around a shared concern, such as the rights of particular (often marginalized and less powerful) social groups. Feminist critiques of cosmopolitanism point to transnational networks based around the struggles of particular groups against forms of exclusion. These networks are underpinned by concerns with ethics, empathy, and notions of care for 'Others'. Importantly, these forms of cosmopolitanism have not risen from attempts to develop a normative, universalistic version of cosmopolitanism which insist on a common political culture ('cosmopolitanism from above'), but instead have emerged through practices on behalf of marginalized groups in specific sociospatial contexts which retain cultural diversity ('cosmopolitanism from below'). To an extent, such examples demonstrate that people can see themselves as simultaneously embedded in local concerns and allegiances and develop a personal moral sense of global concern, even if this is not thought of in terms of cosmopolitanism. This point raises further critiques of cosmopolitanism that it is internally fractured along different axes of power and local allegiances.

Cosmopolitanism, the Nation-State and Nationalism

Although cosmopolitanism in the majority of its forms does not necessarily foresee, or even require, the end of the nation state, it is emphasized as an ideal form of organization and condition that transcends the state. Criticisms of cosmopolitanism note that even though more and more people are experiencing global interrelatedness, they still draw on more locally rooted and embedded forms of allegiance in identity construction and senses of the self. One of these forms is nationalism. Sentiments toward the nation remain an important force in the global economy, polity, and society despite ongoing cosmopolitanization. Thus, there has been a debate over whether cosmopolitanism and nationalism can coexist. From certain views this is a possibility. Citizens, for example, may express pride in the contribution of their nation to the lot of broader humanity. Individuals can celebrate different ways of being in the world, including a sense of global connectedness, at the same time as being committed to a single nation state.

However, cosmopolitanism is further endangered by the possibility of it becoming part of the wider political, social, and economic projects of particular globally dominant nation states. In such cases, notions of cosmopolitanism are defined according to the cultural assumptions and national prejudices of states, but are then mobilized as representing certain 'universal' and 'naturalized' truths which become difficult to challenge. Cosmopolitanism is thus advanced as a means of legit imating the political projects of particular states on the global stage.

Cosmopolitanism, Sociocultural Identity, and Space

This association of cosmopolitanism with the projects of nation states links to critiques based on attacking cosmopolitanism as an elite construct. From some per spectives, cosmopolitanism has been conceptualized as a disposition of a Western elite of educated, transnational professionals who travel globally and feel at home in, and can negotiate, a multitude of global cultures and spaces. By contrast, locally bounded working class neighborhoods have been represented as noncosmopolitan, in the sense that they apparently lack a global outlook and connect edness. This perspective has been soundly critiqued as an elite conceptualization of cosmopolitanism on a number of grounds. The spaces of elite global travelers are, in fact, often tightly regulated and homogeneous (executive air port lounges, multinational hotel chains, and the spaces of 'ex pat' work and leisure) and can serve to isolate this elite from the world of difference around them. Research also indicates that such transnational professional elites may encounter difference through business and work, but in their home lives they associate much more closely with people similar to themselves and live lifestyles in neigh borhoods in which everyday contact with difference is all but eliminated. Again, knowledge of, and exposure to, difference will not necessarily produce cosmopolitanism. By contrast, working class neighborhoods may exhibit a great deal of sociocultural diversity (through, for example, immigration) and their inhabitants may have to develop skills in negotiating everyday encounters with difference in their locality, though some of these encounters may be the ones that they do not wish to have or cannot avoid.

A similar tension is also found at the heart of par ticular forms of cosmopolitan urbanism, where urban elites design, designate, and regulate certain areas of cities as 'cosmopolitan', but where this designation typically refers to a form of cosmopolitanism associated with certain middle class consumption practices. The designation of certain quarters or areas of cities as cosmopolitan implicitly labels others as non cosmopolitan, leading to a differential valuing of parts of cities and the people who live there. In these processes, the commodification of cosmopolitanism as part of urban regeneration and gentrification may paradoxically lead to the exclusion of certain forms of difference (e.g., gen dered, racial, ethnic, (dis)ability, age, class, or sexual difference) which are deemed unacceptable by elites. Traditional working class neighborhoods may be por trayed as noncosmopolitan and assumed to be more likely to be associated with reactionary forms of nationalism, or the suburbs may be represented as banal compared to the vibrant cultural diversity of 'cosmopolitan' city centers and consumption districts. Even where difference is ap parently celebrated in the city center, such as in com modified 'China Towns' or 'Gay Villages', there are still limits to what is deemed acceptable difference (e.g., certain nonheterosexual sexual practices and identities). Again such simplistic dichotomies do little to reveal the potential of actually existing cosmopolitan relations. Some studies of major cities reveal how in certain neighborhoods consumption practices and uses of public space lead to cosmopolitanism in everyday exchanges without the intervention of elite notions of the concept or attempts to plan it into urban space and communities.

These examples of the elite control over the desig nation of what counts as cosmopolitan raise issues of power in relation to cosmopolitanism in the context of global inequalities. They highlight that cosmopolitanism is not always the result of a conscious and voluntary choice and that different social groups have different opportunities to make such choices. This has been well illustrated by studies which have examined the everyday experiences of transnational un and semi skilled labor to uncover various forms of working class cosmopolitanism (though these arguments could be extended to include other forms of sociocultural differences, such as gender, age, and (dis)ability). Far from being the archetypal globe trotting elites, such migrant workers have to ne gotiate the difficulties and dangers of the global economy and the realities of un and semi skilled labor markets from a much less powerful position. For many people placed in this situation, not entirely through their own choice, cosmopolitanism is developed as a part of survival strategies. Migrant street sellers from the developing world who work in the developed world, for example, have to negotiate complex sets of relationship with fellow street sellers, with suppliers of goods, with local au thorities, laws and police, with customers who may be locals or tourists, and with social actors who control transnational international flows of currency (legal and illegal) back to their countries of origin. Transnational un and semi skilled labor may have to negotiate com plex intercultural relations on an everyday basis in workspaces such as the building site, in the street, and in shared rented accommodation.

Such practices are typically seen as cosmopolitan in the sense of being an aptitude and skill in negotiating difference. However, in other contexts, such as the family, in migrant residential spaces, while socializing or eating with fellow nationals, or in religious institutions, at other times of the day this cosmopolitan attitude may be forgotten in favor of cultural, national, or religious preferences. Such studies demonstrate that cosmo politanism has a temporality related to the negotiation of everyday spaces (and others demonstrate the fluidity of cosmopolitanism throughout different stages of peoples' lives). People in such contexts may possess cosmopolitan attributes, but these do not arise through their own choices and do not necessarily become a fixed part of their being.

However, it would be wrong in this evaluation to neglect other situations in which encounters with sociocultural difference have had positive outcomes and led to fulfilling cosmopolitan experiences. These include intercultural personal relationships, such as binational marriage partnerships or same sex partnerships, and personal biographies, for example, of migration and ac ceptance into new communities (e.g., the arguments of Nava on encounters with racial and gender difference). While historically, various forms of non normative sexualities have been seen as a threat to the state, certain figures such as gay men and women have been viewed as positive agents of cosmopolitanism. In addition, it is ne cessary to acknowledge again the more positive experi ences of transnational networks based around care, empathy, and the struggles of social groups against exclusion.

Cosmopolitan Futures?

All these points raise considerable questions over cosmopolitanism as a normative political project. As demonstrated, cosmopolitanism is characterized by multiple and highly contested definitions. What emerges are multiple cosmopolitanisms produced in different sociospatial contexts and which can also exhibit tem porality and fluidity and overlap with other allegiances at a range of scales. Cosmopolitanism exists as a normative political philosophy which is centered on universalistic notions of global citizenship. However, this view is very different from the multiple contested forms of actually existing cosmopolitanism which emerge through the 'grounding' of cosmopolitanism in the spatial and its exploration in a range of sociospatial contexts. Different senses of cosmopolitanism are performed through the embodied social experiences of individuals and groups in different contexts within a world of inequalities.

This presence of multiple cosmopolitanisms does not indicate a clear future for it as a concept or a normative project. It may increasingly become the tool of neoliberal states which will use it as means of legitimating the political projects of the nation state through an appeal to apparently universal notions. It may well continue for some time as a marker of difference in space and society mobilized in particular by certain elite groups as part of their cultural capital. Associated with this, it may have a (limited) future in a commodified form to indicate some apparently superior condition of cities and certain urban spaces which is used to help cities compete in the global economy. As global flows increase, and more people are drawn into interdependencies that transcend state boundaries, it may be deployed as a particular aptitude or skill which certain people can deploy to negotiate life in an increasingly globalized world, but remain superficial and fluid as people dip in and out of contexts that require cosmopolitan competence.

More optimistically, perhaps education could help to contribute to a more intensified cosmopolitan sense of living in the globalized world. Political and ethical action combined with education may underpin the bottom up development of new social movements and the 'scaling up' from the local to the global of a revolutionary cosmopolitanism. Finally, the continuing intensification of transnational interdependencies in the face of global threats may lead inadvertently to the production of cosmopolitanism as an unconscious side effect of dealing with global concerns, or cosmopolitanism may continue to arise from everyday encounters in mundane 'micro publics' such as schools, workplaces, and leisure spaces. Some kind of 'accidental cosmopolitanism' may thus become a more prevalent form of the production of a cosmopolitan consciousness.

See also: Difference/Politics of Difference; Globalization, Cultural; Identity Politics; Multicultural City; Other/ Otherness; Transnationality.

Further Reading

Appiah, K. A. (1998). Cosmopolitan patriots. In Cheah, P. & Robbins, B. (eds.) *Cosmopolitics: Thinking and Feeling beyond the Nation*, pp 91 114. Minneapolis, MN: University of Minnesota Press.

Beck, U. (2006). *The Cosmopolitan Vision*. Cambridge: Polity Press.

Beck, U. and Sznaider, N. (2006). A literature on cosmopolitanism: An overview. *The British Journal of Sociology* 57, 153 164.

Beck, U. and Sznaider, N. (2006). Unpacking cosmopolitanism for the social sciences: A research agenda. *The British Journal of Sociology* 57, 1 23.

Beck, U. and Sznaider, N. (eds.) (2006). *Special Issue: Cosmopolitan Sociology. The British Journal of Sociology* 57, 1 164.

Binnie, J. and Skeggs, B. (2004). Cosmopolitan knowledge and the production and consumption of sexualized space: Manchester's gay village. *Sociological Review* 52, 39 61.

Binnie, J., Holloway, J., Millington, S. and Young, C. (eds.) (2006). *Cosmopolitan Urbanism*. London: Routledge.

Cheah, P. and Robbins, B. (eds.) (1998). *Cosmopolitics: Thinking and Feeling beyond the Nation*. Minneapolis, MN: University of Minnesota Press.

Datta, A. (2009). Places of everyday cosmopolitanisms: East European construction workers in London. *Environment and Planning A*, 41, 353 370.

Hannerz, U. (1996). *Transnational Connections: Culture, People, Places*. London: Routledge.

Kennedy, P. (2007). Global transformations but local, 'bubble' lives: Taking a reality check on some globalization concepts. *Globalizations* 4, 267 282.

Kurasawa, F. (2004). A cosmopolitanism from below: Alternative globalisation and the creation of solidarity without bounds. *European Journal of Sociology* 45, 233 255.

Lamont, M. and Aksartova, S. (2002). Ordinary cosmopolitanisms: Strategies for bridging racial boundaries among working class men. *Theory, Culture and Society* 19, 1 25.

Mitchell, K. (2007). Geographies of identity: The intimate cosmopolitan. *Progress in Human Geography* 31, 706 720.

Nava, M. (2007). *Visceral Cosmopolitanism. Gender, Culture and the Normalisation of Difference*. London: Berg.

Popke, J. (2007). Geography and ethics: Spaces of cosmopolitan responsibility. *Progress in Human Geography* 31, 509 518.

Sandercock, L. (2003). *Cosmopolis II. Mongrel Cities of the 21st Century*. London: Continuum.

Vertovec, S. and Cohen, R. (eds.) (2002). *Conceiving Cosmopolitanism. Theory, Context and Practice*. Oxford: Oxford University Press.

Werbner, P. (1999). Global pathways. Working class cosmopolitans and the creation of transnational ethnic worlds. *Social Anthropology* 7, 17 35.

Young, C., Diep, M. and Drabble, S. (2006). Living with difference? The 'cosmopolitan city' and urban reimaging in Manchester, UK. *Urban Studies* 43, 1687 1714.

Counter-Mapping

R. Rundstrom, University of Oklahoma, Norman, OK, USA

Glossary

Hegemonic The capacity of one group to dominate and oppress another primarily through a combination of coercion and consent of those dominated, and also occasionally by force.

Nongovernmental Organization (NGO) A nonstate, noncorporate organization or institution, usually privately owned, whose primary goal may be to promote environmental conservation, human rights, economic development, or technological change.

Reification A distortion in which something abstract or nonexistent is transformed into material reality, or when something is treated as real when it is not.

Introduction

In the early 1990s, the end of the US–Soviet Cold War and the concomitant rise of geographic information technologies produced opportunities for marginalized peoples to renew their effort to gain greater control over their lands and lives in a manner that could influence their government and its corporate and NGO allies. Long despised and stereotyped by government leaders, bureaucrats, and social elites in most countries as backward savages impeding national economic development, indigenous peoples and their long history of grievances resurfaced into public consciousness as the all consuming Cold War discourse began to subside. Many of these complaints – such as land dispossession, resource abuse, social disruption, and population decimation – were not new, although some of the complainants had been introduced to the negative effects associated with economic globalization only recently. What were new, however, were some of the technologies for representing the geographical facts of these grievances, such as geographic information systems (GIS), satellite images, global positioning systems, and the Internet. Nonetheless, a lot of early counter mapping used manual techniques successfully, and many still use them. But few projects of any kind would have been undertaken without these additional new developments: an abundance of sympathetic NGOs, the globalization of mass media attention to problems of ecological and cultural sustainability, and the emergence of media savvy indigenous leaders linked together in vast international networks.

Counter mappers are openly political about their task. Indeed, they seek to represent places in a manner that not only counters the maps made by state and corporate authorities but also reveals the hegemonic politics inherent in those maps they are countering. Counter mapping is widely understood as a part of postcolonial politics, a task taken up primarily by indigenous peoples who often find themselves in a perpetual state of emergency.

Origin, Current Usage, and Similar Terms

'Counter mapping' was coined by Prof. Nancy Peluso in 1995. She examined the origins and implementation of two kinds of forest mapping in Kalimantan, Indonesia. One set of maps belonged to state forest managers and the large, international financial institutions like the World Bank that supported them. The other set was more recently created by a network of local activists to represent – sometimes in sketch maps, and sometimes through digital imposition on existing state maps – village claims to forests that village residents had long managed themselves. The goal of this second set of maps was to co opt the mapping techniques and conventions of the colonizer, the Indonesian state in this case, to legitimize the claims to the forests and rights to their use emanating from the colonized, the Dayak people indigenous to Kalimantan, and to counter the abuses committed by the state against their lands and lives. Here, counter mapping was initiated in two different ways that now characterize such exercises wherever they occur: professional geographers and anthropologists working through and frequently funded by large multinational NGOs; or local residents often working through a local NGO organize the resistance, often enlisting the help of cartographic, GIS, and other experts from more developed countries or who may already be living in the region. The first may be termed 'top down', the second 'bottom up'.

Now, 12 years later, counter mapping still refers primarily to actions taken by indigenous peoples. But it has also begun to be applied more widely to nonindigenous peoples similarly dispossessed in rich, economically developed countries. For example, farmworkers in the United States (US) have used counter mapping to raise awareness of and provide impetus for later collective actions aimed at improving facilities for protecting workers against contamination from pesticide and other agricultural pollutants. Working class residents of The Bronx in New York City used participatory GIS to help combat environmental injustice and to counter police portrayals of their neighborhoods as crime ridden, and small scale fishing communities in the northeastern US are working on an atlas of their fisheries to combat the rights assumed by multinational industrial trawlers to control offshore fishing grounds. Even efforts by a weak state, East Timor, to control oil and gas deposits via a seabed boundary dispute with a stronger state, Australia, have been termed 'counter mapping'. Some are already seeing counter mapping by indigenous peoples as just one subtype of community or participatory mapping.

Readers should be aware of a number of closely related expressions enjoying varying degrees of use today: ethnocartography, community mapping, alternative cartography, mapping back, counter hegemonic mapping, and (public) participatory mapping. The following are sometimes used synonymously but encompass much more: bioregional mapping, critical cartography, subversive cartography, and remapping. In almost every case, GIS may be used in place of mapping or cartography. Two observations may be made of this abundance of similar terminology: (1) hundreds of similar projects in as many places have been undertaken involving hundreds of different academic consultants and (2) embedded in all of these terms is the widely recognized condition that until recently influential maps were produced solely under the aegis of governments, corporations, or other large institutions.

Theoretical Status

With only a few exceptions, experienced scholars have considered counter mapping a pragmatic exercise. They have tended to focus on establishing an empirical record of successes and failures, the reasons for the outcomes, and what practices do and do not work in the regional contexts in which they have worked. Theorizing their work in a broader context has been of secondary or no concern, perhaps understandably so. Suggested contexts include development theory, postcolonialism, and perhaps most importantly, theories of culture change. Yet, there are some notable exceptions to this general trend. Lefebvre's *The Production of Space* has been used to explain the possible conceptualizations of space underlying the motives of participants in these projects. The concepts of access and territorialization at work in counter mapping have also been explored by another experienced researcher. Yet another has suggested that the binary distinction between indigenous and expert scientific–governmental concepts of resources is false. The argument is that discourses constructed this way pit local communities of producers against expert managers in ways that do not accurately represent the blurred quality of such a boundary between supposed adversaries, a highly permeable line that may be crossed repeatedly by those from either side. Indeed, it has been argued that all 'players' – the professional, the bureaucratic, and the local indigenous or artisanal – bring highly partial interests to bear on issues of resource definition and control. Some of these interests diverge, while others coincide. What our general understanding of the role of individual agents in cultural and environmental change can say about it all, and how the hundreds of empirical examples may be used to inform theories of technology and change are matters that have yet to receive much attention.

The Diffusion of Counter-Mapping

Singular, small mapping exercises conducted sporadically in the arctic and subarctic regions of Alaska and Canada began in the 1950s, and gathered momentum in the region in the late 1960s, particularly in Canada. There, land use and occupancy projects (LUOPs) became normalized by the 1970s. Vast regions of untapped resources existed in the north and west, in the province of British Columbia and the territory then known as the Northwest Territories. These too were areas where aboriginal title had not yet been extinguished according to the needs of English common law. Indian and Inuit leaders were alert to the approaching resource grab for oil and gas and precious minerals just as they were concerned about cultural persistence. The impetus for native peoples to produce hundreds of map biographies – the stories of their lives on the land recorded on maps – thus came directly from indigenous political organizations themselves, an indication of their ability to leverage a degree of political power in a liberal parliamentary democracy. The three volume *Report of the Inuit Land Use and Occupancy Project* is the most comprehensive example. Conducted under the auspices of Inuit Tapirisat of Canada and published in 1976, the project documented Inuit land use ideas and practices across 1.5 million square miles of land and sea that comprise their domain, thus laying the groundwork not only for contesting the state's presumption of unrestricted access but also for the emergence of the Inuit governed territory now known as Nunavut.

Indigenous peoples of the circumpolar Arctic and subarctic are connected especially well. It was in 1975 that a Saami activist and author began making maps of Saami lands and lives following his participation in the first International Conference of Indigenous Peoples in British Columbia (BC), Canada, which was organized by George Manuel, a Secwepemc (Shuswap) from Canada, who, later, as president of the Union of British Columbia Indian Chiefs (UBCIC) brought the UBCIC's Council of Nations to approve a LUOP for northern BC in the early 1980s. As president of the World Council of Indigenous Peoples from 1975 to 1981, Manuel traveled to Argentina, Chile, Guatemala, Nicaragua, Peru, and Sweden to meet with indigenous political leaders in part to promote conventional of indigenous LUOPs as a means of fighting land takings by states.

Links among the indigenous peoples living in rich and poor countries were strengthened at the supranational scale in 1982 with the first meeting of the United Nations (UN) Working Group on Indigenous Populations, later superseded by the UN Permanent Forum on Indigenous Issues formed in 2000. UN sponsorship continues to in crease. In 2004, the UN Educational, Scientific, and Cul tural Organization (UNESCO) held cultural resource mapping seminars in Bangkok. Since 2007, UNESCO has been working with the Indigenous Peoples of Africa Co ordination Committee to foster their capacity to protect their biological and cultural resources through partici patory mapping projects. Presently, both top down and bottom up counter mapping now occur on every contin ent and in every major country in which indigenous peoples are found, with the likely exception of China.

Benefits: Rights, Possession, and Sustainability

From indigenous perspectives, counter mapping has the potential to assist their efforts to gain recognition of land use rights, to pool and protect community ecological and land use knowledge, to clarify the meaning of geographic features, and ultimately, to protect themselves against land dispossession and other forces associated with the unwanted aspects of economic globalization. As such, counter mapping has become an important part of the cultural triage that minority societies must constantly undergo when faced with an oppressive state authority. Aside from providing a means to fight dispossession, counter mapping may also present an occasion for other social activities crucial to cultural survival, including: strengthening kinship and other social networks through self education, providing an impetus for future collective efforts, building threat awareness and mobilizing activist behavior in the pursuit of conflict resolution, and uni fying communities by building consensus and a sense of group empowerment and self determination. In addition, these human potentials may all be achieved as ecological diversity and stabilization are promoted around the world.

Some indigenous organizations have gone so far as to publish or endorse publication of their own atlases. The Inuit, Wet'suwet'en, Gitxsan, Coast Salish, Zuni, Saami, and Maya have pursued these massive undertakings that intensify all the benefits listed above. To date, atlas projects have been most popular in North America, but their production is spreading.

These goals have been abetted by the production of manuals of carefully outlined procedures and issues associated with the pursuit of counter mapping in Southeast Asia, Latin America, and North America. For example, both top down and bottom up projects in Indonesia got a boost from a 1996 manual jointly pub lished by the Ford Foundation, World Wildlife Fund, and several Indonesian NGOs that included summaries of seven pioneering case studies of community mapping projects in Kalimantan, Sulawesi, and Sumatra. The manual promotes mapping as a means of mediating re source conflict in the hope of producing equal, effective partnerships among government agencies and local communities. Subsequently, thousands of Indonesian villages have engaged in counter mapping projects. Other detailed manuals are available for those working in Latin America, Canada, and the US.

Pitfalls: Transformation, Reification, and Illegitimacy

Enthusiastic proclamations of the value of counter map ping in the 1980s and early 1990s have now been replaced with a more tempered appreciation of its potential, in cluding a healthy skepticism of its universal value. Indeed, wherever counter mapping is undertaken, prickly issues emerge about the transformative and reifying power of maps and GIS, the transparency of the political interests of those involved, and the overall legitimacy of projects. Counter mapping typically involves a multicultural team working together to map truths to power – conditions that raise sensitivity to the above issues still higher. So, members of a community may be unexcited by the pro spect of counter mapping because the project may not succeed in producing change, it may exacerbate existing sociopolitical frictions within the community, and it may further worsen relations with opposing institutions of an already oppressive government.

Its transformative and reifying power lies in the fact that counter mapping is a contradictory pursuit. The purpose is to counterpose indigenous ideas of spatial organization and land use against those of the state. But because of the power imbalance, the language and tools of the dominant society must be used by those under its control. That process of using another's language and tools can change the ideas so represented. The ultimate effects of mapping have an unpredictable quality as a result.

The best example involves the problem of bounding territories on maps. For indigenous peoples, the spatial organization of what the Western world calls 'resources' cannot be mapped independent of the social relations that produce it. That is, rights to resources on specific lands are defined by customary and dynamic social practices more than by firm ownership within solid boundary lines marking mutually exclusive territories. Kinship networks and obligations, marriage practices, rights to procuring food through hunting and agriculture, customs in disposing of the dead, and other factors impose a dynamic social atmosphere on land use rights, one that is frequently at odds with the Western concept of fixed resources on fixed territories. Indeed, the techniques used in counter mapping are usually silent about the histories and precedents involved in customary land uses, effectively cancelling them and creating new realities based on alien ideas. In short, land use can be a nonterritorial issue, one in which community members struggle more with what boundaries mean from time to time than with establishing, moving, or eliminating them permanently. Yet, mapping back at the state usually requires fixing boundaries and imposing stasis on the land which, in turn, overthrows the old ways and reinvents land use as a territorial claim that can provoke internal conflict. For example, counter mapping has transformed property rights discourse in some parts of Indonesia and thereby contributed to new kinds of ethnically based violence a few years ago. In southwestern Africa, counter mapping projects led to unexpected land grabs by large conservation NGOs based in Europe and North America that were detrimental to the mapmakers themselves.

Reification of territories and boundaries is only one part of the problem. Counter mapping can racialize land areas, uses, and residents too. This occurs as people get categorized by language or culture and mapped onto lands falsely ascribed to be theirs. In the most diverse areas of Africa, Asia, and Latin America, for example, the process of mapping may arbitrarily assign lands or districts to an ethnic majority in a particular area based on statistics, a rule by majority process that abrogates customary practices for dealing with social differences. Counter mapping may also reify villages or communities as the primary unit of geographical analysis. Yet, such places may not actually reflect the local organization of a society on the land so much as the categories imposed on them by earlier colonial administrations, or even external NGOs.

Finally, the project's transparency influences whether the counter mapping effort will have legitimacy for the people involved. The relevant concerns can be grouped into six categories: the goals of everyone involved, especially those of large external NGOs and their donors; the accessibility of appropriate social groups to the project, notably women and youth; authorship, or the related question of who gets to prepare the maps and how credit is assigned; accountability of mapmakers to non mapmaking members of the community; the openness with which the project is pursued; and the clarity of its overall goals and expected effects to community members.

As important as these concerns are to mapping activists, in the map or be mapped binary thinking common today, the immediacy of the threat to lands and lives takes precedence. Yet, people in indigenous communities must still calculate the social cost. Not so long ago, in the 1990s to be more precise, a Penan possessing a map could get arrested in Sarawak. So, the decision to make and use maps, to provide moral and political support to the effort, to use them in seeking redress of grievances in a court of law ruled by adversaries who also control an army, and many other decisions are all calculations that must be carefully made by community leaders who frequently work under severe constraints not understood by associates of external NGOs. Such calculations require careful social networking and a long, communal deliberation process. The fact that counter maps must often be prepared and deployed quickly puts stress on this system of decision making.

Conclusion

Counter mapping existed long before the term was coined in 1995. Since then, the openly political and anticolonial practice has diffused worldwide, providing significant benefit to indigenous peoples seeking greater control of and securer connections to the places where they live. Counter mapping projects also increase their capacity for galvanizing other social actions. But such mapmaking is not without pitfalls, including the necessary transformation of one society's geographical ideas into the terms and rubrics of the one to which they address their grievances. In seeking change, the colonized must use the language and tools of the colonizer – a risky situation at best. For this reason, the relationship between cartography, GIS, and other digital spatial technologies and the lives and lands of indigenous peoples will remain an uneasy one for a long time.

To this point, research on counter mapping has focused on recording the positive and negative aspects of its practice in countless empirical examples. This research has largely been written from the standpoint of non indigenous academic participant observers. Of great value in its own right, this literature might be expanded by experienced anthropologists and geographers to include the following: empirical descriptions written by indigenous participants of what counter mapping projects have meant to them; longitudinal studies of the long term effects of counter mapping in specific regions (to date, there appear to be only two of these); analyses of the maps themselves and other results of counter mapping projects; and more theoretical work on the place of counter mapping in development theory, postcolonial studies, and theories of cultural and technological change.

See also: Biodiversity Mapping; Critical Cartography; Critical GIS; Cross-Cultural Research; Deforestation; Extended Metropolitan Region; GIS and Society; GIS,

Public Participation; Hegemony; Human Rights; Imperialism, Cultural; Indigeneity; Indigenous Geographies; Indigenous Knowledges; Land Rights; Literature; Mapping, Race and Ethnicity; Maps and Governance; Maps and Protest; Maps and the State; Neocolonialism; Neoliberalism; Neoliberalism and Development; Nongovernmental Organizations; Postcolonialism/Postcolonial Geographies; Resource and Environmental Economics; Science and Scientism, Cartography; Spatial Ontologies; State; Territory and Territoriality.

Further Reading

Avataq Cultural Institute (1990). *Inuit Place Name Map Series of Nunavik*. Montreal: Avataq Cultural Institute.

Brody, H. (1998). *Maps and Dreams: Indians and the British Columbia Frontier* (first publ. 1981). Prospect Heights, IL: Waveland Press.

Chapin, M., Lamb, Z. and Threlkeld, B. (2005). Mapping indigenous lands. *Annual Review of Anthropology* 34, 619 638.

Chapin, M. and Threlkeld, B. (2001). *Indigenous Landscapes: A Study in Ethnocartography*. Arlington, VA: Center for the Support of Native Lands.

Cooke, F. (2003). Maps and counter maps: Globalized imaginings and local realities of Sarawak's plantation agriculture. *Journal of Southeast Asian Studies* 34, 265 284.

Cravey, A., Arcury, T. and Quandt, S. (2000). Mapping as a means of farmworker education and empowerment. *Journal of Geography* 99, 229 237.

Ferguson, T. and Hart, E. (1985). *A Zuni Atlas*. Norman, OK: University of Oklahoma Press.

Freeman, M. M. R. (ed.) (1976). *Inuit Land Use and Occupancy Project*. Ottawa: Ministry of Supply and Services Canada.

Herlihy, P. and Knapp, G. (2003). Maps of, by, and for the peoples of Latin America. *Human Organization* 62, 303 314 (special issue on participatory mapping and indigenous peoples).

Hodgson, D. and Schroeder, R. (2002). Dilemmas of counter mapping community resources in Tanzania. *Development and Change* 33, 79 100.

Johnson, J. T., Louis, R. P. and Pramono, A. H. (2006). Facing the Future: Encouraging Critical Cartographic Literacies in Indigenous Communities. http://www.acme journal.org/vol4/JJRLAP.pdf (accessed in June 2008).

Maantay, J. (2002). Mapping environmental injustices: Pitfalls and potential of geographic information systems in assessing environmental health and equity. *Environmental Health Perspectives* 110(supplement 2), 161 171.

Nevins, J. (2004). Contesting the boundaries of international justice: State countermapping and offshore resource struggles between East Timor and Austrlia. *Economic Geography* 80, 1 22.

Orlove, B. (1993). The ethnography of maps: The cultural and social contexts of cartographic representation in Peru. *Cartographica* 30, 29 46.

Parker, B. (2006). Constructing community through maps? Power and praxis in community mapping. *The Professional Geographer* 58, 470 484.

Peluso, N. (1995). Whose woods are these? Counter mapping forest territories in Kalimantan, Indonesia. *Antipode* 27, 383 406.

Peluso, N. (2005). Seeing property in land use: Local territorializations in West Kalimantan, Indonesia. *Geografisk Tidsskrift* 105, 1 15.

Phadke, R. (2005). People's science in action: The politics of protest and knowledge brokering in India. *Society and Natural Resources* 18, 363 375.

Poole, P. (ed.) (1995). Geomatics: Who needs it? *Cultural Survival Quarterly* 18(4), 1 77 (special issue on counter mapping and indigenous peoples).

Poole, P. (2003). Cultural Mapping and Indigenous Peoples. A report for UNESCO. http://www.iapad.org/publications/ppgis/cultural mapping.pdf (accessed in June 2008).

Ribot, J. and Peluso, N. (2003). A theory of access. *Rural Sociology* 68, 153 181.

Riewe, R. (1992). *Nunavut Atlas*. Edmonton: Canadian Circumpolar Institute and the Tungavik Federation of Nunavut.

Robbins, P. (2003). Beyond ground truth: GIS and the environmental knowledge of herders, professional foresters, and other traditional communities. *Human Ecology* 31, 233 253.

Sletto, B. (2002). Producing space(s), representing landscapes: Maps and resource conflicts in Trinidad. *Cultural Geographies* 9, 389 420.

St. Martin, K. (2005). Mapping economic diversity in the First World: The case of fisheries. *Environment and Planning A* 37, 959 979.

Warner, C. (2005). Community mapping, local planning, and alternative land use strategies in Bali. *Geografisk Tidsskrift* 105, 29 41.

Warren, C. (2005). Mapping common futures: Customary communities, NGOs and the state in Indonesia's Reform Era. *Development and Change* 36, 49 73.

Relevant Websites

http://geography.berkeley.edu
Maya Atlas The Struggle to Preserve Maya Land in Southern Belize, Department of Geography, University of California at Berkeley.

http://www.ubcic.bc.ca
Traditional Land Use Studies, Union of BC Indian Chiefs.

Counterurbanization

C. J. A. Mitchell, University of Waterloo, Waterloo, ON, Canada
C. R. Bryant, Université de Montréal, Montréal, QC, Canada

Glossary

Amenity Things beyond life's necessities that contribute to psychological, rather than physical, well-being.

Hoover Index A measure of population concentration ranging from 0 to 100, or from a reading of a population that is distributed perfectly uniformly to one resembling perfect concentration.

Jeffersonian A pastoral existence that rejects city life.

Kondratiev Wave A regular S-shaped cycle consisting of alternating periods of high and low economic growth.

Postproductivist Countryside Rural areas that no longer rely solely on economic activity related to primary-sector production.

Regional City An expanded urban form that includes both an urban core and functionally integrated periphery.

Conceptualization

Counterurbanizing describes the process whereby a regional or national settlement system undergoes a transformation from a state of population concentration (i.e., an urban state) toward a state of population deconcentration (i.e., a counterurban state) (**Figure 1**). In this state, a larger proportion of the population resides in less concentrated settlement areas, than in those with higher populations. A system is counterurbanizing when growth rates in smaller settlements exceed those recorded in larger locales. This scenario emerges as a consequence of higher rates of net natural population increase and/or net migration in the former, when contrasted with the latter (**Figure 2**).

Levels of net natural population increase are a function of change in two variables: the number of live births and the number of deaths occurring within a system. Historically, population discrepancies between urban and rural areas could be explained largely by disparities in these two components. Over the course of the twentieth century, spatial variations in these rates have diminished across much of the developed world. Consequently, contemporary population change in these countries is usually a function of variations in net migration (and, in some cases, regional or local variations in rates of natural increase).

Net migration levels are derived from the number of people entering (in migration) or leaving (out migration) a settlement system from within (internal migration) or beyond (external migration) the area under study (a regional or national system). The geographic destination, and origin, of these migrants causes a system to either urbanize or counterurbanize.

For example, if net migration in larger centers exceeds that occurring in smaller destinations, then an urbanizing process may be underway. If net migration in smaller locales is greatest, this may be indicative that the system is counterurbanizing. Furthermore, if residents who originate in small settlements choose to relocate to more concentrated destinations (a movement defined as urbanization), then this may contribute to an urbanizing settlement system. Alternatively, if migration is greatest from larger to smaller settlement areas (a movement defined as counterurbanization), then a system may be

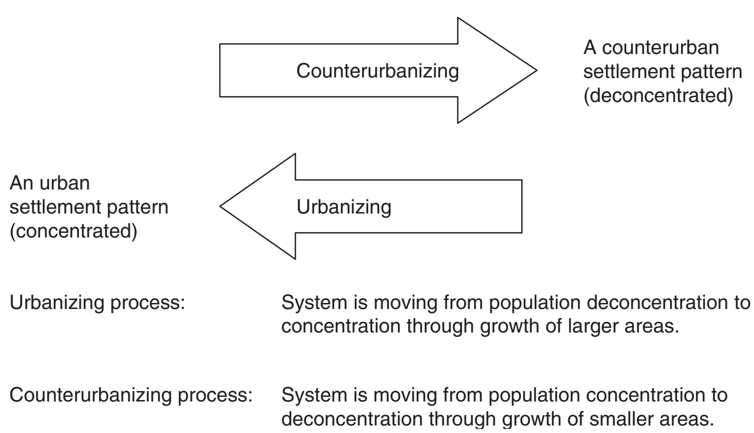

Figure 1 Population redistribution.

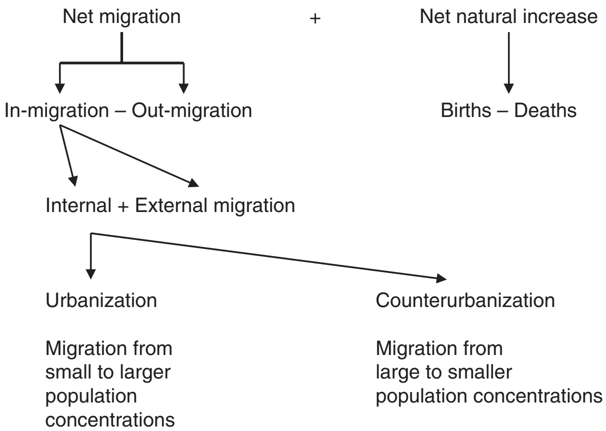

Figure 2 Demographic influences on population redistribution. Adapted from Mitchell, C. J. A. (2004). Making sense of counterurbanization. *Journal of Rural Studies* 20, 15 34.

counterurbanizing. These trends are confirmed only by comparing the resulting levels of growth occurring in the system's settlement areas.

Evidence

Since the onset of the Industrial Revolution, the dom inant trend has been for settlement systems to urbanize. Historically, this situation was driven by the migration of large numbers of rural residents to concentrated settle ments in search of employment in burgeoning manu facturing sectors. Subsequent centuries witnessed a continuation of this trend, culminating in the creation of largely urban settlement patterns across much of the developed world.

Migration tendencies began to shift in the latter half of the twentieth century (**Table 1**). In many developed (e.g., the United States and Japan) and select developing nations (e.g., Estonia and Hungary), rural depopulation slowed while rural in migration levels began to rise (e.g., in Belgium, Denmark, Norway, Switzerland, the United Kingdom (UK), France, West Germany, and Russia). In some cases, these trends contributed to a temporary halting of the urbanizing process. The specific extent and timing of this occurrence varied globally. Deconcentra tion in some regions was reflected in either a sporadic or ongoing trend of higher levels of growth in intermediate sized urban center (called polarization reversal or re directed urbanization). In other areas, comparable tendencies were found to be occurring in a system's smallest settlements (defined as 'rural', 'nonmetropolitan', or simply 'small').

North American researchers, including Brian Berry and Calvin Beale, were the first to document that a rural renaissance was underway. Analysis of American census data for the period 1970–74 demonstrated that popu lation growth levels were higher in nonmetropolitan than metropolitan counties, although later research revealed that growth levels had actually been retained in smaller metropolitan areas, with only the largest concentrations reporting heavy loss. Growth was not ubiquitous, but favored nonmetropolitan counties of the northeast, west, and southern states, in areas with high incomes, high population densities, a state college or university, a large military population or retirement amenities. Calculations of national indices of population concentration (i.e., the Hoover index) for the 1960s and 1970s provided con firmation of these national turnaround trends.

Population deconcentration is restricted neither spa tially nor temporally. It has occurred in many countries, including Canada, Britain, West Germany, France, Spain, Sweden, Australia, and Japan, and at various times. Like the United States (US), Canada and Australia underwent a period of deconcentration during the 1970s. In the UK, this trend continued into the 1980s, although at a re duced level. Spain and West Germany saw deconcen tration prevail into the 1990s. Russia also experienced its

Table 1 National trends in urbanization and counterurbanization in the 1970s and 1980s for selected European countries

Country	*1970s*	*1980s*	*Shift*	*1980 84*	*1984 88*	*Shift*
Austria	+0.38	+0.01		0.25	+0.47	+
Belgium	0.36	0.44		0.49	0.33	+
Denmark	0.79	0.01	+	0.04	0.16	
Finland		+0.69		+0.51	+0.80	
France	0.26	0.36		0.33	0.31	NC
Germany	0.29			0.63	0.08	+
Ireland	+0.43	0.35				
Italy	+0.12	0.21		0.16	0.33	
Netherlands	0.83	+0.12	+	0.24	+0.46	
Norway	+0.21	+0.69	+			
Portugal	+0.36	+0.52	+	+0.39	+0.53	+
Sweden	0.26	+0.35	+	+0.14	+0.53	+

Data are correlation coefficients of relationship between net migration rate and population density. + one indicating urbanization; one indicating counterurbanization; NC: no significant change. Adapted from Champion, A. B. (1992). Urban and regional demographic trends in the developed world. *Urban Studies* 29, 461 482.

first counterurbanizing process during this decade, while, at the same time, the US saw a brief return to higher levels of deconcentration (1990–94).

Since its widespread recognition in the 1970s, the counterurbanizing process has become increasingly spatially selective. In many countries, growth is now restricted to rural areas, and small towns either falling within an extended sphere of metropolitan influence, or in more remote regions with attractive physical or constructed amenity. Furthermore, unlike the earlier decades, contemporary growth contributes to the counterurbanizing of regional, rather than national, settlement systems (e.g., in Canada, Spain, Australia, Denmark, Norway, and Sweden).

Explanation

Social Factors

The trend of nonmetropolitan population growth is attributed to a variety of factors, whose influence is uneven over both space and time. Most widely recognized are those of a social nature that attribute the increase in rural population to the escalating importance of environmentalism over productionism; from a lifestyle devoted to improving one's job, education, and income, to one focused on improving the condition of one's lived space. In some cases, this transition has driven individuals into revitalized city centers where a more cosmopolitan lifestyle awaits. In others, it is the lure of an Arcadian ideal that entices an adventitious population into the city's countryside, and sometimes, even further beyond.

Real, or imagined, images of rurality encompass myriad attributes that are believed to be lacking in large, urban concentrations. Open space, a safe environment for children, peacefulness, ethnic homogeneity, desirable single family housing, small town living, an attractive natural or constructed landscape, abundant recreational opportunities, and preexisting kinship or friendship affiliations are among the characteristics that create a village in the mind of many would be migrants. This image took on new meaning during the latter half of the twentieth century, as time–space compression intensified. Relocation to a bucolic setting was seen as a means to escape the undesirable attributes emerging from modernity (i.e., the negative side effects of urban concentration). This movement was made even more enticing, and viable, by the diffusion of desirable urban amenities (infrastructure, communications, transportation) into the countryside. While some rural in migrants chose to retain a partial tie to the urban core through their commuting behavior, others sought to relinquish this bond altogether.

One of the earliest types of counterurbanites, the American 'ex urbanite', sought to combine the benefits of a rural, and 'Jeffersonian' existence, with employment opportunities available in adjacent urban cores (e.g., New York City). These moves later were enhanced by rapid expansion of the intercity freeway system, which extended the city's influence farther into the fringe, giving rise to new urban forms, such as the regional city. More recently, the availability of telecommunications technology enables the maintenance of this tie over increasingly larger distances (e.g., from metropolitan centers of California to its high mountain areas). Populations of ex urbanites are not only restricted to peri urban areas of America, but are also found surrounding major centers in Canada (e.g., Toronto, Montreal, Vancouver, Edmonton, Calgary, Winnipeg, Halifax, Quebec City), Australia (e.g., Adelaide, Sydney), Spain (e.g., Madrid, Barcelona), the Netherlands (e.g., Amsterdam), Finland (Helsinki), Ireland (Dublin, Belfast), and most metropolises of the developed world.

Ex urbanites are accompanied by other migrants who choose rural locales as an attractive space to combine both workplace and residence. In these cases, employment is the facilitator that enables these 'anti urban' residents to realize their desire of returning to, or taking up, residence in a less concentrated setting. In the Andalusian countryside of Spain, for example, an inward flow of older, married migrants has secured employment in their cultural hearth. In Australia, the attainment of job transfers and promotions by the managerial and professional classes has facilitated a flow of lifestyle motivated migration to the North coast. The Highlands and Islands of Northern Scotland have received migrants from England and Wales who voluntarily fill positions vacated by local residents or who engage in entrepreneurial activity, such as craft production or service provision. In the high mountain areas of British Columbia (Canada), California, and Nevada (US), young, and often single, migrants seek out (or create) conventional employment in their new place of residence. In Britain (and on a smaller scale in the Appalachian region of the US), some migrants choose to adopt a radically new and often self sufficient lifestyle (i.e., back to the landers). This variant of green wave, or countercultural, migration has been a vibrant part of many rural economies of the UK for more than 40 years.

For some urban residents, rural locales are embraced during retirement years. These destinations inevitably offer the benefit of small town living in a climatically favorable and scenic (often coastal) landscape. Research in North America, in particular, highlights the importance of early retirement as providing the opportunity for many people to follow their dream and move to amenity rich areas. The towns and villages bordering Georgian Bay, Ontario, Canada, for example, have experienced a graying over the past 30 years. This trend is widespread in many coastal regions including those of the US (e.g., the upper Great Lakes, Florida, Texas, Arizona), Western

Europe (e.g., the Mediterranean coasts of Spain and France and the southwest peninsula of Great Britain), and Australia (in New South Wales).

Economic Factors

Social motives are responsible for voluntary shifts from urban to rural locales. Structural realities related to economic conditions give rise to nonvoluntary moves. For some, the decision to relocate to a nonmetropolitan set ting is a reflection of a productionist lifestyle and simply the need to make ends meet. For these 'displaced ur banites', a rural setting is sought not for its amenity characteristics, but to satisfy a need for affordable hous ing or lower costs of living. At least part of the growth occurring outside Adelaide, Australia, and New York City, is attributed to this type of migration. In many transitional countries undergoing a post socialist trans formational crisis (e.g., Russia, Estonia, Hungary), urban out migration is a necessity brought about by economic hardship.

The availability of employment opportunities outside metropolitan cores also prompts urban out migration, or reduces levels of migration from rural locales. This scenario arose in the US during the late 1970s when the rate of increase in employment was greater in counties supported by less than 2500 residents than in counties with larger population concentrations. In Britain, sig nificant growth also occurred during the same period when economic activity in small towns and rural areas coincided with a decline in employment levels in the London conurbation.

Nonmetropolitan employment opportunities in all economic sectors (primary, secondary, tertiary/quater nary) emerged during the 1970s. In some areas, growth was spurred by a renewed interest in domestic energy sources. In the US, the energy crisis of the early 1970s promoted development of coal mines in Appalachia and the northern Great Plains. Similarly, in the UK, the North Sea oil developments of the same decade provided employment opportunities, leading to growth of these energy rich regions.

The post 1970 period also saw some manufacturing relocate to nonmetropolitan sites. The less urbanized and rural regions of the EU, for example, reported stronger industrial performance during this decade than did their urban counterparts. Similarly, during the same period, manufacturing employment levels were higher in America's nonmetropolitan counties. Here, as in other regions, deconcentration occurred as a consequence of economic restructuring or, in some areas, as an integral part of the rise of an information economy.

According to the neo Marxist school, the 1970s marked the beginning of economic restructuring, char acterized by the replacement of regional sectoral specializations with the spatial division of labor. This shift saw managerial functions retained in urban cores (e.g., Frankfurt, New York, Tokyo London, Paris, Amsterdam), with more routine functions relocating to small and medium sized towns where the diseconomies of metro politan location could be avoided. This spatial reorgan ization of capitalistic enterprise has been at least partially responsible for precipitating the counterurbanizing process.

In other locales, nonmetropolitan centers have been the beneficiaries of the information age – what some believe to be the arrival of a fifth Kondratiev wave. The diffusion of telecommunications technology into the nonmetropolitan realm has granted employers consider able latitude in their location decisions. As distance constraints continue to disappear, the amenities of small town living have become increasingly more attractive to industries that were formerly urban based. In the US, for example, significant nonmetropolitan growth occurred in the high technology sector between 1975 and 1982; a trend also apparent in parts of nonmetropolitan Canada and Spain.

In some regions (e.g., southern Italy, West Germany, Canada), metropolitan deindustrialization has been ac companied by a proliferation of tertiary sector employ ment in nonmetropolitan centers. This may be viewed as part of the general tertiarization that is taking place within many national economies of the developed world. The escalating growth of both producer and household services in scenic mountain regions of Europe and the US is indicative of this trend, one that contributes both to the counterurbanizing phenomenon and, also, to the creation of a multifunctional, postproductivist landscape.

Political Factors

Since World War II, some countries have enacted pro grams to encourage growth outside the metropolitan orbit. In Britain, planners actively promoted the growth of country towns during the 1960s to reduce the devel opment pressures on the London conurbation. Similar policies were adopted in Northern Ireland a decade later, to curtail the areal expansion of Belfast. In Spain, growth was refocused on more peripheral regions as a con sequence of major administrative restructuring during the same decade. In Sweden, the decentralization of government employees also during the 1970s was initi ated to curb population growth in Greater Stockholm.

Other policies have either directly, or indirectly, en hanced the appeal of rural residence. In the southern US, extensive dam building programs were implemented in the 1960s and 1970s to promote development of these regional economies. During the same period, investments in sports, hospitals, schools, and community facilities were initiated in various parts of the UK to raise the socioeconomic attractiveness of its rural areas. West Germany also promoted the decentralization of higher education to reduce spatial inequity between its settlements. In the transition economies of Eastern Europe, urban to rural migration has been enhanced by government programs, which, in this case, restore land to precommunist owners. In Russia, for example, land restitution policies contributed to the counterurbanization movement that prevailed during the 1990s. Since the mid 1980s, Canada too has attempted to induce community economic development in rural and resource based regions. In Quebec, for example, programs promoting network development have been implemented both to attract urbanites and former ruralities, and to stem the outflow of current residents. These actions have increased in intensity and interest as more rural areas have experienced a positive net in migration.

Other initiatives provide direct, or indirect, incentives for relocation. In the US, an increase in federally funded retirement benefits and mortgage guarantees during the 1970s provided residents with greater financial security and the means to move to a desirable nonmetropolitan location. In the UK, government improvement grants for the repair of older or substandard housing also became available during this time, making it possible for urban residents to relocate to remote parts of Scotland. Business, too, has benefited from government programs. Labor subsidies and capital investment grants made available during the 1960s and 1970s were instrumental in enticing manufacturing plants to relocate to more rural regions of the UK (e.g., to Devon). In Quebec, Canada, programs have been developed to enable young migrants in several peripheral regions to create employment in their new place of residence. For a number of these, as in other countries, opportunities have been created through new information communication technologies via the Internet.

Statistical Factors

Statistical underbounding is a final factor that is used to explain population growth in rural areas. As cities grow outward, adjacent rural locales may be drawn into the metropolitan orbit. This phenomenon of metropolitan overspill, or decentralization, provides a boost to census areas defined as rural. Until these areas are appropriately designated as part of the urban or megalopolitan core, their growth rates distort the magnitude of population change that is occurring within a rural settlement system. Thus, what may appear to be a deconcentration trend may, in fact, simply reflect further concentration and a continuation of the urbanizing process as cities spread outward in a wave like fashion.

Statistical underbounding, and associated boundary changes, has been responsible for at least part of the growth reported in Japan and South Africa during the 1960s, Canada and the US throughout the 1970s, Britain during the 1980s, and Sweden during the 1990s. Although accounting for part of the growth occurring in rural areas, this factor cannot explain the higher rates of growth that have been experienced in areas well beyond metropolitan influence. This explanation only comes from considering other political, economic, and social factors that, when combined, contribute to an understanding of the counterurbanizing process.

Implications

There are numerous implications of population deconcentration for rural communities located both within and beyond the urban fields of metropolitan regions. These focus first, on how to handle and prepare for the impacts on settlements that are experiencing a renewal in their population, and second, on how to effectively integrate newcomers into the life of the community. Addressing impacts and integration does not occur ‘naturally’, as both require a concerted effort by local actors and citizenry. Without intervention, new migrants may feel uncomfortable, and even unwanted, in their new surroundings.

Implications related to impacts on settlements are many, including additional costs to the local tax base, and the capacity of local systems (education, infrastructure, housing, healthcare) to cater to new needs. In the absence of an adequate supply of appropriate housing, for example, new migrants may bid up local prices, making it more difficult for young, local residents to enter the housing market. At the same time, higher property values may also decrease the likelihood of attracting additional migrants. This has been observed in several rural regions, including the Haliburton area of Ontario, Canada. In other areas, these problems are caused by those who purchase properties for a summer residence, and sometimes for investment purposes, as in the case of the Isle of Skye in western Scotland.

The counterurbanizing trend also has implications for local healthcare systems. Growing numbers of retirees may place a strain on existing facilities. This situation is compounded by the general thinning out of healthcare professionals that has been documented in myriad countries of the developed world. The attraction of new medical professionals has become an important agenda item for many rural locales. In North America, for example, some communities have embarked on aggressive advertising campaigns that focus on the merits of rural living. Others have lobbied for modern medical facilities, at least on a regional basis, while some offer financial incentives and/or accessible housing to young, medical professionals willing to relocate to their municipality.

A significant implication arising from population growth is the challenge of integrating diverse newcomers into community life. New migrants, located in both the city's countryside and more remote rural areas, may ar rive with perspectives that differ radically from those held by long term residents. In the rarer case where new migrants are recent immigrants, such differences can be even more pronounced. Providing a means by which these interests can be integrated into community decision making is an ongoing challenge for communities benefiting from migration induced population growth.

In some cases (e.g., the Haliburton area of Ontario, Canada), early newcomers have been excluded from engaging in community governance. The failure to rec ognize new voices ensures the continued hegemony of long term residents, and ultimately leads to the creation of contested rural space. In many regions, the situation is only improved when significant changes are made to the orchestration of development planning, as is the case in the Haliburton area. Constructing new forms of gov ernance to better reflect the evolving composition of expectations and place identities necessitates a shift in the attitudes and actions of longer term residents. This is no simple matter, and how communities respond con tributes even further to the differentiation of community development trajectories among rural and small town settlements.

Population increase in rural settlements is far from uniform, and significant spatial variation may occur at relatively small scales of geographic analysis. This means that neighboring settlements may experience very dif ferent demographic profiles and trajectories, which, in some cases, may reflect the scale of geographic filter that has been used. In other cases, this mosaic of trajectories is related to the heterogeneity of rural amenities and en vironments that exist within a local area. Furthermore, in a cumulative process of change, the very actions of new migrants, and longer term residents (e.g., elected officials or other community leaders) can enhance, or reduce, the probability of others arriving. This reveals a portrait of rural areas as being heterogeneous, even at the very local scale.

Counterurbanizing tendencies that are affecting nonmetropolitan space are not necessarily permanent. Despite the development of programs to foster infra structure development and employment opportunities in many areas, governments also continue to mount major programs and interventions that favor large urban and metropolitan centers. The continuation of these initia tives may have a long term impact on population deconcentration in the future. On the other hand, it is clear that how local rural and small town communities prepare for in migrants and actively work to facilitate integration will contribute much to making newcomers a permanent part of their growth and development.

See also: Berry, B.; De-Industrialization; Demography; Fordism, Post-Fordism and Flexible Specialization; Immigration I; Industrial Restructuring; Knowledge Economy; Migration; Modernity; Rural Populations; Rurality and Post-Rurality; Telecommunications; Transitional Economies; Urbanization.

Further Reading

Beale, C. L. (1977). The recent shift of United States population to nonmetropolitan areas, 1970 1975. *International Regional Science Review* 2, 113 122.

Berry, B. J. (1976). *Urbanization and Counterurbanization*. Beverly Hills, CA: Sage Publications.

Bryant, C. R. and Joseph, A. (2001). Canada's rural population: Trends in space and implications in place. *The Canadian Geographer* (Anniversary Issue) 45(1), 132 137.

Champion, A. B. (1992). Urban and regional demographic trends in the developed world. *Urban Studies* 29, 461 482.

Champion, A. G. (1990). *Counterurbanization: The Changing Pace and Nature of Population Deconcentration*. London: Edward Arnold.

Cloke, P. (1985). Counter urbanization: A rural perspective. *Geography* 70, 13 23.

Geyer, H. and Kontuly, T. (1993). A theoretical foundation for the concept of differential urbanization. *International Regional Science Review* 15, 157 177.

Keith, H. (2006). From dropping out to leading on? British counter cultural back to the land in a changing rurality. *Progress in Human Geography* 30, 309 336.

Marsden, T., Lowe, P. and Whatmore, S. (eds.) (1990). *Rural Re structuring: Global Processes and Their Responses*. London: Fulton.

Mitchell, C. J. A. (2004). Making sense of counterurbanization. *Journal of Rural Studies* 20, 15 34.

Sant, M. and Simons, P. (1993). The conceptual basis of counterurbanization: Critique and development. *Australian Geographical Studies* 31, 113 126.

Vartiainen, P. (1989). Counterurbanization: A challenge for socio theoretical geography. *Journal of Rural Studies* 5, 217 225.

Relevant Websites

http://esa.un.org/unup
Further Information on Global Urban/Rural Population Shifts, UN.

http://www.worldbank.org
Further Information on Trends in Global Population Growth.

Cox, K.

H. van der Wusten, Universiteit van Amsterdam, Amsterdam, The Netherlands

Glossary

Gerrymandering The practice of altering political constituencies to acquire electoral advantage, typically by shifting the boundaries of a district or electoral tract between elections.

Growth Coalition A partnership or tacit understanding between local politicians, businesspeople, and institutions in which there are shared local visions, typically pro-economic growth.

Local Economic Development An area of urban policymaking in which the focus is on increasing inward investment, indigenous growth, and property values.

Cox, Kevin R. (1939–)

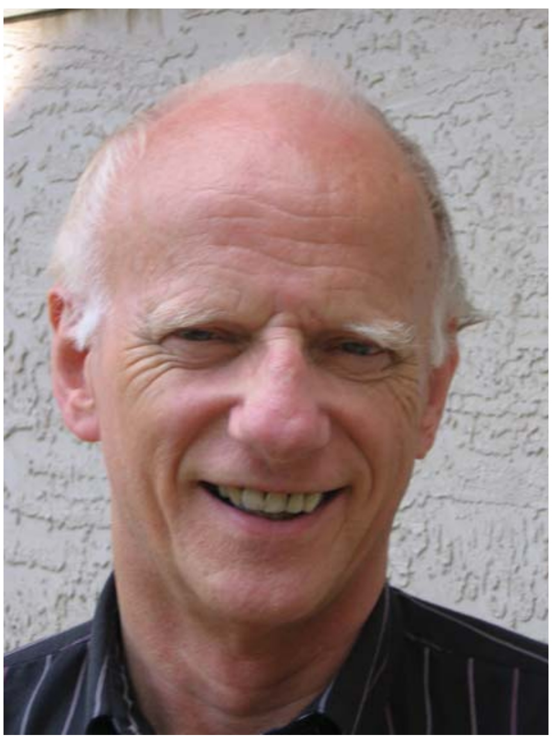

Born in Warwick, England in 1939 in a politically con servative family of lower public sector professionals, Kevin R. Cox got his secondary education in a public school that harbored a number of well read teachers. Apart from successfully preparing his entry in a good university they made him acquainted with a general outline of socialist thought. He did his undergraduate studies in geography in Cambridge, where he met Peter Haggett as a teacher and David Harvey as an older stu dent. During those years he participated in a Marxist debating club and a socialist association.

After his BA in 1961 he went to the US, initially for a year, and then stayed. He took his MA (1963) and PhD (1966) at the University of Illinois in Urbana Champaign. His PhD dealt with voting behavior. In 1965 he started teaching at the University of Ohio in Columbus. In 1971 he became full professor, and has been a distinguished university professor in the same university since 2003. He has lectured widely, in Europe and elsewhere, won an Honours award of the Association of American Geog raphers (AAG) and a Guggenheim Fellowship. He loves and plays jazz music, which – though rarely – shows up in his work (in an illustration of American stamps fea turing African American musicians). He is not known for much applied work or a great deal of recognized service outside academia. He has been an adviser to 17 suc cessfully defended PhDs.

Initially Cox was an ardent participant in what was then known as quantitative geography (i.e., the effort to put geography on a more strictly nomothetical basis and to apply matching research methods). He has always been keen on systematizing, careful argument, and theoretical grounding. From early on he was familiar with the con struction and analysis of mental maps and social net works and he cooperated in that field with Reg Golledge. The experience helped him to quickly produce his own general human geography textbook *Man, Location and Behavior* in 1972.

Cox has always considered himself a political geog rapher. In graduate school and shortly thereafter he de veloped a model and did widely regarded empirical work with British electoral data. This was published not only in major geography journals but also in highly respected political sociology publications in the circle of Stein Rokkan. As a sequence he organized – together with Reynolds – a collection in the overlapping area of geography and political science. Rokkan wrote, in his foreword to that book, that it had been a revelation to read Kevin Cox's thesis on the implications of the recent developments in mathematical geography for the study of politics. But Cox had by that time moved on from an interest in voting and elections to a more widely shared concern within geography: the distributional outcomes of politics, welfare geography, particularly in the urban context. This showed in the book and it did not sit well with the focus of political sociology. It would, however, guide Cox's further work.

His studies of urban issues later refocused from problems of distribution to problems of local economic development. His emphasis on local economic develop ment has implied an interest in the formation, mainten ance, and results of politically active local growth coalitions and questions of administrative territorial fragmentation. His concentration on the local level – the sites where relations of production get shaped – was somehow challenged by the globalization literatures. In the 1990s, he extensively wrote on scale questions in geography, undermining all simple notions of predeter mined levels in geographical analysis and also chal lenging the complete predominance of something that could be called 'the global level'. Like others he has struggled with the significance of territoriality – which he has come to see as the very core concept of political geography – versus the apparently growing importance of network connections. Most of his work on these topics is concerned with the US, though in the 2000s he added long distance labor migration to South African cities to his list of research interests.

His approach gradually became Marxist. From the early 1970s onward, he returned to Marx's own texts and came to see Marxism as "the major force of my intel lectual life" as he wrote, looking back, in 2005. One major inspiration was Harvey's *Social justice and the city* (re viewed by Cox in a piece that appeared in 1976). He has extensively criticized those who, in current debates within Marxism, deviate from the central tenets of re lations of production–class–exploitation to consider the much more generally accepted notions of relations of exchange – inequality. This also fits in with his primary interest in sites of production as central anchors for the spaces of dependence and engagement in which social life unfolds. The connections of Marxism and geography that he pursues have been inspired by Harvey and he also calls his ultimate aim 'historical geographical material ism'. This is not to say that he has lost his sensitivity to his early interests in more cultural topics, but he insists that they should be studied in a materialist framework for which Marx has proposed the crucial basis.

Cox has always remained active in the field of political geography, but has within that field moved far beyond mere electoral concerns and wants to shift the emphasis of traditional political geography (less attention to lo cation of capital cities, gerrymandering, and irredentism as he stated already in 1979). He has latterly written a textbook – *Political Geography* – that appeared in 2002, and published a reader on critical concepts in political geography (2005), both reflecting his research agenda, and his continuous ability for principled argument and systematic presentation.

See also: Electoral Geography; Local Economic Development; Marxism/Marxist Geography I; Marxism/ Marxist Geography II; NIMBY; Political Geography; Rational Choice Theory (and Rational Choice Marxism); Structural Marxism.

Further Reading

Cox, K. R. (1968). Suburbia and voting behavior in the London metropolitan area. *Annals of the Association of American Geographers* 58, 111 127.

Cox, K. R. (1972). *Man, Location and Behavior: An Introduction to Human Geography*. New York: Wiley.

Cox, K. R. (1973). *Conflict, Power and Politics in the City: A Geographic Approach*. New York: McGraw Hill.

Cox, K. R. (ed.) (1997). *Spaces of Globalization: Reasserting the Power of the Local*. New York: Guilford Press.

Cox, K. R. (1998). Spaces of dependence, spaces of engagement and the politics of scale, or: Looking for local politics. *Political Geography* 17, 1 24.

Cox, K. R. (2002). *Political Geography: State, Territory and Politics*. Oxford: Blackwell.

Cox, K. R. (2005). The politics of local and regional development in comparative perspective. *Space and Polity* 9, 191 200.

Cox, K. R. and Johnston, R. J. (eds.) (1982). *Conflict, Politics and the Urban Scene: Case Studies in Urban Political Geography*. London: Longman.

Cox, K. R. and Mair, A. (1988). Locality and community in the politics of local economic development. *Annals of the Association of American Geographers* 78, 307 325.

Relevant Websites

http://geog www.sbs.ohio state.edu
Cox's own website, Ohio State University.

Creativity

A. Currah, Oxford University Centre for the Environment, Oxford, UK

Glossary

Agglomeration The sustained clustering of activities, people, institutions, and resources around a specific location or set of locations, which eventually leads to the formation of cities, industrial regions, and other centers of specialization.

Creativity The mobilization, appropriation, and utilization of knowledge to create ideas, interpretations, forms, goods, methods, or technologies that are distinctly original and imaginative in their application and long-term impact.

Intellectual Property A historically and geographically specific set of rights (comprising copyrights, patents, trademarks) devised and enforced by Western states, which were introduced to protect and encourage acts of creativity.

Knowledge The awareness and understanding of facts, principles, and truths (acquired and refined over time) that broadly relate to the origin, form, function, evolution, and interaction of the natural and human geographical environment.

Transnational Corporation A public or privately owned economic entity, with locations in multiple countries, which has the scale and resources to produce, market, and distribute goods and services around the world.

Human creativity is arguably one of the principal foundations of modern civilization. It has been fueled by a range of basic human characteristics: for example, the constant search for meaning, the desire to invent, the power of our imagination, both individual and collective, and the pursuit of economic rewards, as well as social status. Together, these traits have served to advance the scope and function of civilization over the centuries, as indicated by the evolving nature and culture of cities. In effect, creativity represents the mobilization of resources, both physical and intangible, in pursuit of a vision that then finds its expression, to varying degrees, in the human and natural geographical environment.

An appreciation of creativity is therefore a vital component of human geography, and economic geography in particular. As we will see, creativity acts as a fulcrum around which many other forces crosscut and interact, both in geography and in other related disciplines. A wide range of changes in the human geographical environment can, to varying degrees, be related to corresponding changes in the nature and direction of human creativity. While this article will hopefully introduce the reader to themes that are specifically related to the economic and geographical significance of creativity, it will also touch upon general themes that are covered elsewhere in the encyclopedia.

Throughout our relatively brief history, the creation, appropriation, and application of ideas – for example, designs, writings, technologies – have transformed the economic landscape, as well as broader social, cultural, and political structures. In a very general sense, it is 'creativity' that has provided the fuel for Schumpeter's gales of 'creative destruction', or Kondratieff's 'long waves' of economic development. To be sure, a substantial portion of creativity takes place beyond the realm of the market, where economic considerations of profit, efficiency, and cost prevail. Indeed, for many individuals, the very act of creating is a distinctly 'noneconomic' activity, driven by desires and incentives other than the accumulation of profit. Nonetheless, creativity does foster the application of novel and economically valuable knowledge, such as new goods or services. In turn, this reconfigures the spatial and organizational contours of capital and its commodity chains – for example, patterns of production and resource allocation, social and spatial divisions of labor, or the sites of distribution and consumption.

For all these reasons, creativity has been a pivotal force in the evolution of the economic landscape. Curiously, however, the topic of creativity has only recently entered the currents of mainstream debate and research in economic geography. It is only really since the mid 1990s that geographers have begun to seriously and explicitly interrogate the economic and spatial dimensions of creativity. Before that, creativity was largely treated as background 'noise', which was examined, albeit indirectly, through detailed studies of entrepreneurship and technological innovation.

In the past decade or so, geographers have sought to examine the growing importance of creativity in advanced economies, principally in North America and Europe, where there has been a shift toward 'knowledge based' economic activities, and where, as a result, skills such as organizational learning and adaptation are increasingly vital to competitive success. Studies along these lines have contributed to a thorough reconceptualization of the firm in economic geography. Thus, according to current perspectives, the firm is not simply a nexus of economic imperatives; it is also a site of creative practices, which are played out across relational networks at different spatial scales.

The 'creative turn', as it might perhaps be termed, has also been characterized by a more specific set of questions. In particular, geographers and others have been drawn to explore an emergent class of 'creative industries', which are believed to lie at the heart of advanced economies such as the United Kingdom or the United States. In general, these industries can be said to produce goods or services that comprise a significant amount of knowledge in their design, function, and appeal. That is to say, the economic output of firms in these industries depends explicitly on the creation, appropriation, and application of new ideas – indeed, frequently in a form that is entirely 'nonrivalrous' and 'nonexcludable', to borrow the language of economists (see below). In the UK, for example, the Department of Culture, Media and Sport (DCMS) has incorporated sectors such as advertising, architecture, crafts, design, fashion, film, publishing, software, and radio into its official classification of the creative industries.

The conceptual boundaries of the creative industries are blurred, however, as a result of overlaps with another domain of research in economic geography, which is concerned with the cultural underpinnings of the economy. In the spirit of this approach, geographers have studied industries that produce goods with relatively higher levels of 'sign' versus 'utility' value. The esthetic appeal of goods such as fashion clothing or film is exemplary. The key point is that creative industries may be cultural in their orientation, as the DCMS classification attests; but equally, they may also be oriented around the production of goods and services that have relatively higher levels of 'utility' versus 'sign' value – for example, sectors such as biotechnology, engineering, law, or applied scientific research. These, too, are inherently creative domains of activity and innovation.

The multidimensional nature of creativity, and its expression in an expanding array of industries, means that we still lack a clear, rigorous, and unambiguous definition of the 'creative industries', not only in economic geography but also more broadly across the social sciences, as well as in policy circles (see below). The identification of a group of industries, united by their creative foundation, is made challenging by the fluid and fuzzy meaning of 'creativity' in contemporary debates. For example, creativity is often used interchangeably with related terms such as culture, innovation, knowledge, learning, and so on. This complicates the terms of the debate, as well as the clarity and policy relevance of empirical research.

Nonetheless, the study of creativity is still at a very early stage in economic geography. The topic is now at the forefront of debate. It will naturally take time to refine concepts and empirically test the strength of different theoretical viewpoints. There is also a pressing need for comparative case studies to examine the different meaning, form, and evolution of the creative industries in different contexts. Sitting at the interface between the 'cultural' and the 'economic', creativity has much to contribute to contemporary economic geography, which itself remains in a state of flux. The creative industries, however they are defined, are clearly at the forefront of modern capitalism and continue to raise a plethora of questions. The purpose of this article, then, is to highlight some of the key themes of conceptual debate and research that presently encompass the study of creativity in economic geography.

The discussion is structured around three general issues. The first concerns the geographically bounded nature of creativity – in particular, the continued agglomeration of the creative industries in a globally networked hierarchy of city regions. The second examines the debate about the relationship between creativity and regional economic development. And finally, the third considers the challenges inherent to the governance of creativity, with specific reference to recent developments in the area of intellectual property rights.

The Geographically Bounded Nature of Creativity

One of the core strengths of economic geography is its capacity to show how economic actions and processes are 'embedded' in specific contexts, and in turn how they change over time. The same is true of creativity. Here, economic geographers have examined the geographical dynamics of creative practices, which also take form in specific contexts.

An important finding of this research is that, despite the inexorable march toward globalization, creative industries continue to be located in zones of economic agglomeration, which in turn are tightly integrated with each other and global markets. It might be more accurate, therefore, to say that these industries exhibit a 'dual' geography. That is, the conception and creation of creative goods and services are increasingly concentrated in particular places, but at the same time these very goods and services are also being channeled and sold to consumers through increasingly extensive chains of distribution. As we will see below, there is a distinctive 'political economy' to geographies of creative production and distribution. In most sectors, a handful of 'transnational corporations (TNCs)' account for the bulk of revenues. Crucially, these corporate entities have the scale and resources to funnel their goods and services around the world.

This pattern is especially evident in those industries that exhibit a high 'cultural' or 'symbolic' content – for example, fashion, film, music, or television. Such industries typically generate their goods in a tightly defined 'creative field', which comprises the resources, individuals, firms, and institutions necessary to initiate and sustain creativity. In the case of film, a prominent manifestation of a creative field would be the area popularly known as Hollywood, where the 'major' film studios are a dominant force. Or, in the case of fashion, an equivalent creative field would be Paris, London, or New York, where, likewise, major fashion houses also occupy a dominant position. The lesson also holds, to varying degrees, for other creative industries such as biotechnology or information technology, where the manipulation of strategic bodies of knowledge by skilled workers is also critical to success. Agglomerations such as Route 128 in Boston, Massachusetts, or Silicon Valley in California, underscore the significance of the creative field, albeit in a different context.

The trend toward spatial agglomeration has been reinforced by the growing prevalence of vertically disintegrated, 'project based' modes of production in many of the creative industries. That is, the tendency for goods and services to be produced through temporary coalitions of firms and individuals, as well as related suppliers and specialized service providers. The short term, dynamic, and frequently unstable nature of project based work demands that participants are in close proximity to one another. The vertical disintegration of the film industry is a classic illustration of this trend, as evidenced by Christopherson and Storper. Here, the rise of projects has steadily given rise to a 'transactions intensive' creative field in the Los Angeles region. Today, independent producers and specialized suppliers in this region interact with clients across the media and entertainment industries.

What theoretical characteristics does the creative field actually exhibit? Although there are both subtle and complex differences between zones of economic agglomeration, it is possible to outline some of the basic characteristics of the regions that typically underpin creative industries.

First and foremost, the creative field is characterized by a dense collection of firms, often small to medium in size, which are involved in the exchange of traded goods and services, as well as untraded 'interdependencies', such as cultural conventions, norms, and trust, often via decentralized project based relational networks. Second and related, the creative field represents a 'vortex' of sociocultural reproduction, where particular ways of doing and knowing the business of creativity are forged and reinforced over time. This is sustained through face to face contact, the development of social networks, and the movement of individuals between different firms and organizations. A vital extension of this is the presence of a diverse labor market, which is able to rapidly mobilize resources to serve the myriad needs of different projects. Third, and finally, the creative field is also shaped, enabled, and (perhaps, occasionally) constrained by what is known as the 'regional milieux'; that is, the presence of local institutions (such as colleges, schools, training centers, trade bodies, and public–private partnerships), as well as the architectural, physical, and technological infrastructures (such as road networks, telecommunications, housing, and amenities) that enable the movement of goods, services, people, and information. Combined, these myriad characteristics serve to create distinctive place based identities over time.

The picture is incomplete, however, without consideration of the connections that bind together the matrix of creative fields around the world. As recent work in economic geography has revealed, there are also important processes of territorial competition, decentralization, and specialization at work in the area of creativity. The creative field must be understood within an increasingly global economic context, where there are different centers of creativity, as well as flows (of people, information, and goods) between those centers. Advances in transportation and telecommunications mean that creative projects can now be modularized and outsourced to specific teams, in different locations, with far greater ease and at lower cost. A typical film project, for example, now comprises a wide range of locations, stretching from the capture of footage (often in several production sites) through the postproduction editing of music and video to product marketing. What this means is that a far greater range of centers (current and aspiring) can participate in the amorphous industry of creativity, whether it is the design of a new product, the rendering of video and animation, or the analysis of commercial data. This is also conducive to heightened levels of territorial specialization; for example, with certain centers developing competencies in specific areas of creative activity. To be sure, barriers to entry do exist. There are powerful increasing return effects that reinforce the strengths of existing clusters, while excluding others. Nonetheless, there is scope for increased territorial competition between centers both in the area of productive investments (as locations fight to attract investment) and in the area of product distribution and marketing (as different firms and industries compete to attain exposure in the global market).

It is insufficient, therefore, to examine the geography of the creative field purely at a local scale. Rather, it is vital that we situate the geographical dynamics of creativity within the wider, extralocal context of networks and global flows of goods, services, and people. On that front, it is also vital that we take into account the political economy of commodity chains in the creative industries – that is, the respective power of different economic entities and zones of agglomeration, and the manner in which that power is enacted and expressed. The scope for heightened territorial specialization and competition in this domain has spurred efforts, principally by institutional and governmental elites, to nurture the development of 'creative clusters' so as to enhance the economic prospects of particular regions, as well as the competitiveness of the home country in global markets. There is growing pressure, in other words, for governments to attract and retain a new class of 'creative capital' and 'creative talent'. It is to this issue that we now turn.

Creativity and Economic Development

Since the 1980s, there has been a profound shift in the focus and direction of urban and regional governance in the advanced economies of North America and Europe. In particular, the emergence of neoliberalism has led to an increased emphasis by policy makers on mechanisms of economic development, as opposed to mechanisms of economic redistribution and equity. Broadly speaking, policy makers have sought to devise initiatives that both nurture local firms and attract external investment. Set against this background there has also been growing awareness in policy circles of the economic importance of human capital and skills.

Given this context, it is of little surprise that the creative industries have captured the imagination of policy makers and government leaders. For in recent years, the discourse of creativity has quickly suffused the language of urban and regional economic policy. The interest surrounding creativity both in geography and related social science disciplines has helped to fuel this trend. Prominent books such as *The creative city* or *The creative class* have transformed creativity from an academic concept into a compelling brand, which is especially appealing for investment hungry political elites. The essential premise of these and other related books is that creativity is now a decisive source of economic and competitive advantage for cities and regions worldwide. Florida, for example, claims that over 30% of the US workforce is employed in its classification of the 'creative industries'. And so, the locational decisions of these creative individuals are believed to have a significant impact on the economic geographies of growth. The implication is that cities and regions should therefore adjust their economic development policies to attract and retain individuals that are employed in creative occupations.

In practice, the underlying argument calls for policies that are specifically tailored to the lifestyle interests of the so called 'creative class'. In brief, Florida has argued that individuals from the creative class are likely to prefer to live and work in places that are socially diverse, tolerant, and open to new forms of expression and economic innovation. As such, policy makers have been advised to focus on developing 'lifestyle assets' (such as a vibrant art and music scene, with ample recreational opportunities), creating spaces for social interaction (such as 'cultural quarters' and 'bohemian districts'), or fostering diversity and openness (along ethnic, racial, demographic, and sexual metrics). The success of academic celebrity consultants such as Landry and Florida has, indeed, begun to make an impact on the kinds of policies favored by urban and regional political elites. However, there are growing concerns about the prevalence of this version of 'creativity' in the policy making process. A powerful critique has emerged, offering a range of alternative perspectives. The two themes below capture the essence of the debate.

First, branded concepts such as 'the creative city' or 'the creative class' advance an overarching and misleadingly simple view of creativity and its relationship to economic development. Like the related concept of economic clusters, the concept of creative cities or classes still lacks clarity and rigor. The geographical measurement of creativity is also problematic. Metrics such as Florida's Bohemian or Gay Index are alluring because they create 'league tables' in which policy makers can vie for promotion; but they also imply a causal correlation between certain characteristics and economic development. It is true that open and tolerant cities are likely to foster greater amounts of interaction and innovation. Dense and diverse labor markets may indeed provide some of the original sparks for creativity. The agglomeration of creative firms and individuals may also provide the basis for local 'spill over' effects. However, it is also true that creativity may be a symptom – rather than a cause – of economic development. In addition, the implicit assumption that a class of highly skilled and mobile workers is responsible for the bulk of economic development means that other kinds of creative individuals are likely to be neglected by policy.

Second, the creativity script dovetails perfectly with the prevailing market led focus of urban and regional policy. Worryingly, it provides a revamped guise for neoliberal elites who are interested in packaging and marketing urban space to a global market of investors. There are potentially negative ramifications to a creativity led policy of development, as recently recognized by Florida, Landry, and others. In particular, the emergence of urban quarters, districts, and neighborhoods that are designed to appeal to the 'creative class' may accentuate sociospatial exclusion. The rent inflation that often follows such urban redevelopment threatens to displace lower skill workers, potentially eroding the diversity that made the area attractive. A market led policy may unduly constrain or threaten those aspects of urban life that are more collective, communal, and nonmarket in nature. Indeed, there are myriad forms of creativity lying outside the market that add to the diversity and appeal of cities. The prevailing policy environment tends to favor only those forms of creativity that are able to promote profit and growth. And in the long run, the incessant pursuit of lucrative creative led policies may also, paradoxically, lead to increased homogeneity and standardization in form and function between cities. There is clear scope, therefore, for cities to embrace a much wider and more inclusive definition of creativity in their policies.

Overall, then, creativity has become an important if contentious topic in relation to studies of economic de velopment. The specific mechanisms that connect cre ativity to economic development, at a range of spatial scales, require further investigation. In particular, there is a pressing need for comparative case studies that help to elucidate these issues in different locations. It is clear, for example, that creativity is culturally and geographically specific in its growth and long term impact. This leads us to the final theme of the discussion, which concerns the economic governance of creativity – that is, the political and regulatory structures that shape the reproduction of information rich creative goods and services over time. As we will see, the globalization of the creative industries has exposed tensions inherent to the creation, appropri ation, and commercial application of ideas.

The Economic Governance of Creativity

The goods and services that emanate from creativity are distinctive in economic terms because they typically exhibit the characteristics of 'public goods'. That is to say, creative goods and services are essentially built on the packaging and exploitation of information. As such, they can be reproduced at virtually zero marginal cost after the initial act of conception, design, and production. The information that comprises creative expressions is therefore both 'nonrivalrous' (it can be appropriated by multiple users) and 'nonexcludable' (it is difficult to ex clude nonpaying users). These characteristics are a po tential deterrent to acts of creation and investment by individuals and firms. The freely shareable nature of digital creative works such as music, movies, and software is illustrative. It is for these reasons that creativity, at least in a commercial form, has relied upon a specific form of economic governance provided by the state – namely, intellectual property rights (IPRs). Dating back to the fifteenth century, IPRs come in the form of copyrights, patents, and trademarks.

IPRs are significant from an economic and geo graphical perspective because they govern the movement and use of particular goods and services. Their under lying objective is to realize a balance: on the one hand, to enable creators to extract a reasonable return from their investments; and on the other hand, to ensure that society has access to the latest creative works, both cultural and scientific, which will eventually pass into the public do main where they can be appropriated and reused in fu ture acts of creativity. IPRs, then, connect local fields of creation with global markets of consumption. The Hol lywood studios, for example, rely on copyright legislation to extract revenues from their films, which are then funneled back into the Los Angeles region in the form of profit and new creative investments. In recent decades, however, the core function of IPRs has been distorted by the growing size and market power of TNCs, which control extensive portfolios of copyrights, patents, and/or trademarks. Such TNCs typically sit at the heart of vertically disintegrated project based networks, pro viding the finances necessary to create, market, and dis tribute a wide range of goods and services.

The economic value of intellectual property has meant that TNCs, from the media, entertainment, soft ware, and pharmaceutical industries, have been able to successfully reshape the regulatory environment in their favor. In the US, for example, the broad spectrum of intellectual property now accounts for over $1.2 trillion, or 12% of gross domestic product. In many advanced economies, the creation and defense of IPRs is now seen as a critical requisite to national competitiveness, not least because of the steady transition to a globalized knowledge based economy. The result has been a pro gressive strengthening and extension of IPRs, in the area of copyright, patents, and trademarks. This has occurred not only in North America and Europe, but also at a global scale. Increasingly, TNCs are using their power to incorporate IPRs into trade bargaining. For example, US trade negotiators successfully included new IPR obli gations in recent bilateral free trade agreements with Australia, Chile, Morocco, and Singapore. The case of IPRs therefore highlights the socially constructed and dynamic nature of the governance structures that shape the evolution of the economic landscape. Indeed, IPRs are illustrative of what Clark has termed 'real regulation'; a view that state created regulations are shaped and performed through the interactions between firms, lobbyists, and policy makers.

IPRs must also be viewed as a geographically and historically specific mode of economic governance. The concept of owning creativity is arguably ethnocentric: IPRs are embedded in a distinctly Euro American matrix of thinking. The principal bastion of that approach, the World Intellectual Property Organization (WIPO), ad vances a particular narrative of creativity that celebrates the individual creator and normalizes the defense of property rights. This approach, however, runs into con flict with alternative creative practices, some that are well established, others that are emerging. In non Western cultures, for example, the creation of new ideas – both cultural and scientific – often takes place in a col laborative and nonproprietary fashion. In such cases, the mode of governance is premised upon social norms and conventions, as opposed to prices and individual property rights. Indeed, the literature is replete with case studies of 'gift economies', where the emphasis is on the sharing rather than the hoarding of resources. This mode of creation is well established, especially in the field of music where the process of industrialization is arguably a historical anomaly. Nonetheless, IPRs are not programmed to recognize the diffuse network of social relations from which creativity emerges.

The situation has been complicated, in recent years, by the rapid emergence of a digitally networked economic and social environment across advanced economies. The growth of the Internet, and a related ecology of devices and computers, has made it possible for millions of computer users to share their resources (time, hardware, and information) with each other for free. In particular, the Internet has given rise to new patterns of creative production and distribution, which still await detailed investigation and theorization. There are myriad examples. Consider the sharing of code among open source software developers; or the peer production of information based resources such as encyclopedias (Wikipedia), or photo (Flickr) and video (YouTube) archives. These spaces of creativity defy conventional economic logic. Operating in the absence of prices or managerial commands, millions of anonymous users have proved adept at sustaining efficient channels for the creation and dissemination of digital creative works. However, although the Internet has radically enhanced our access to the cultural and historical record, the scope for creativity in this new environment is severely constrained by the scope of IPRs. The restrictive nature of copyright, for example, effectively criminalizes a vast amount of derivative and transformative creativity that has been made possible by the digital age.

The Internet has brought to the fore a basic economic tension that affects the art, commerce, and governance of creativity – specifically, how to balance market based modes of exchange (oriented around the commodification of resources) with nonmarket based modes of exchange (oriented around the social production and sharing of resources). For centuries, the creation, appropriation, and application of new ideas have flourished without the market. To be sure, the industrialization of creativity – with the support of IPRs – has generated a tremendous amount of innovation and progress. Copyrights have assured authors, filmmakers, and songwriters of a return on their time and investments. Patents have enabled the development of new drugs and technologies. Trademarks have enabled individuals and firms to protect the value inherent to their brands and commercial identities. However, a substantial portion of creativity continues to flourish outside the formal structure of the market. And moreover, nonproprietary forms of sharing can also generate economic value over time, as the commercial success of open source software attests.

The economic geography of creativity must therefore take into account a wide array of alternative spaces and practices, which are oriented around desires and incentives other than the pursuit of profit. Finding ways to balance the market with nonmarket activity will be critical to the future governance of creativity. By shedding light on the embedded and contingent nature of creativity, economic geographers can play a vital role in this broader conversation and debate.

See also: Commodity Chains; Cultural Economy; Economies, Alternative; Entrepreneurship; Firms; Globalization and Transnational Corporations; Innovation; Knowledge Economy; Relational Economic Geography; Technological Change.

Further Reading

Bassett, K. (1993). Urban cultural strategies and urban regeneration: A case study and critique. *Geoforum* 25, 1773 1788.

Bassett, K., Griffiths, R. and Smith, I. (2002). Cultural industries, cultural clusters and the city: The example of natural history film making in Bristol. *Geoforum* 33, 165 177.

Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. New Haven, CT: Yale University Press.

Bollier, D. (2002). *Silent theft: The private plunder of our common wealth*. London: Routledge.

Bollier, D. (2005). *Brand name bullies: The quest to own and control culture*. Chichester: John Wiley.

Caves, R. E. (2000). *Creative industries: Contracts between art and commerce*. Cambridge, MA: Harvard University Press.

Christopherson, S. (2002). Project work in context: Regulatory change and the new geography of media. *Environment and Planning A* 34, 2003 2015.

Christopherson, S. and Storper, M. (1986). The city as studio; the world as back lot: The impact of vertical disintegration on the location of the motion picture industry. *Environment and Planning D: Society and Space* 4, 305 320.

Clark, G. L. (1992). 'Real' regulation: The administrative state. *Environment and Planning A* 24, 615 627.

Coe, N. M. (2000). On location: American capital and the local labour market in the Vancouver film industry. *International Journal of Urban and Regional Research* 24(1), 79 94.

Coe, N. M. (2001). A hybrid agglomeration? The development of a satellite marshallian industrial district in Vancouver's film industry. *Urban Studies* 38(10), 1753 1775.

Coe, N. M. and Johns, J. (2004). Beyond production clusters: Towards a critical political economy of networks in the film and television industries. In Power, D. & Scott, A. J. (eds.) *The cultural industries and the production of culture*, pp 188 204. London: Routledge.

Currah, A. (2003). Digital effects in the spatial economy of film: Towards a research agenda. *Area* 35(1), 64 73.

Currah, A. (2006). Hollywood versus the internet: The media and entertainment industries in a digital and networked economy. *Journal of Economic Geography* 6(4), 439 468.

Currah, A. (2007). Managing creativity: The tensions between commodities and gifts in a digital networked environment. *Economy & Society* 36(3), 467 494.

Department of Culture, Media and Sport (2006). Creative economy programme, http://www.culture.gov.uk/what we do/ Creative industries/creative economy programme.htm (accessed Oct. 2007).

Florida, R. (2002). Bohemia and economic geography. *Journal of Economic Geography* 2, 55 71.

Florida, R. (2002). The economic geography of talent. *Annals of the Association of American Geographers* 92, 743 755.

Florida, R. (2002). *The rise of the creative class*. New York: Basic Books.

Ghosh, R. A. (2006). *Code: Collaborative ownership and the digital economy*. Cambridge: MIT Press.

Gibson, C. and Kong, L. (2005). Cultural economy: A critical review. *Progress in Human Geography* 29, 541 561.

Grabher, G. (2001). Ecologies of creativity: The village, the group and the heterarchic organization of the British advertising industry. *Environment and Planning A* 33, 351 374.

Grabher, G. (2002). Cool projects, boring institutions: Temporary collaborations in social context. *Regional Studies* 36, 205 214.

Hall, P. (1998). *Cities in civilization*. New York: Pantheon.

Harvey, D. (1989). From managerialism to entrepreneurialism: The transformation in urban governance in late capitalism. *Geografiska Annaler, Series B Human Geography* 71, 3 17.

Hess, C. and Ostrom, E. (2007). *Understanding knowledge as a commons: From theory to practice*. Cambridge, MA: MIT Press.

Hyde, L. (2006). *The gift: How the creative spirit transforms the world*. London: Canongate.

International Intellectual Property Alliance (2004). *The Copyright Industries in the US Economy*. http://www.iipa.com/pdf/ 2004 SIWEK FULL.pdf (accessed Oct. 2007).

Kearns, G. and Philo, C. (eds.) (1993). *Selling places: The city as cultural capital, past and present*. Oxford: Pergamon.

Landry, C. and Bianchini, F. (1995). *The creative city*. London: Demos/ Comedia.

Lessig, L. (2002). *The future of ideas: The fate of the commons in a connected world*. New York: Random House.

Lessig, L. (2004). *Free culture: How big media uses technology and the law to lock down culture and control creativity*. New York: Penguin.

Leyshon, A. (2001). Time space (and digital) compression: Software formats, musical networks, and the reorganization of the music industry. *Environment and Planning A* 33(1), 49 77.

Leyshon, A. (2003). Scary monsters? Free software, peer to peer networks and the spectre of the gift. *Environment and Planning D: Society and Space* 21, 533 558.

Markusen, A. (1999). Fuzzy concepts, scanty evidence and policy distance: The case for rigour and policy relevance in critical regional studies. *Regional Studies* 37(6/7), 701 717.

Mokyr, J. (2002). *The gifts of Athena: Historical origins of the knowledge economy*. Princeton, NJ: Princeton University Press.

Molotch, H. (1996). LA as design product: How art works in a regional economy. In Scott, A. & Soja, E. (eds.) *The city: Los Angeles and urban theory at the end of the twentieth century*, pp 225 275. Berkeley: University of California Press.

Peck, J. (2003). Fuzzy old world: A response to Markusen. *Regional Studies* 37(6/7), 729 740.

Peck, J. (2005). Struggling with the creative class. *International Journal of Urban and Regional Research* 29, 740 770.

Porter, M. E. (2004). The economic performance of regions. *Regional Studies* 37(6/7), 549 578.

Power, D. and Hallencreutz, D. (2002). Profiting from creativity? The music industry in Stockholm, Sweden and Kingston, Jamaica. *Environment and Planning A* 34, 1833 1854.

Power, D. and Scott, A. (eds.) (2004). *Cultural industries and the production of culture*. London: Routledge.

Pratt, A. (1997). The cultural industries production system: A case study of employment change in Britain, 1984 1991. *Environment and Planning A* 29, 1953 1974.

Pratt, A. (2002). Hot jobs in cool places: The material cultures of new media product spaces: The case of South of Market, San Francisco. *Information, Communication and Society* 5(1), 27 50.

Rantisi, N., Leslie, D. and Christopherson, S. (2006). Placing the creative economy: Scale, politics and the material. *Environment and Planning A* 38, 1789 1797.

Saxenian, A. (1994). *Regional advantage: Culture and competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.

Saxenian, A. (2006). *The new argonauts: Regional advantage in a global economy*. Cambridge, MA: Harvard University Press.

Scott, A. J. (1996). The craft, fashion and cultural products industries of Los Angeles: Competitive dynamics and policy dilemmas in a multisectoral image producing complex. *Annals of the Association of American Geographers* 86, 306 323.

Scott, A. J. (1999). The cultural economy: Geography and the creative field. *Media, Culture and Society* 21, 807 817.

Scott, A. J. (2000). *The cultural economy of cities: Essays on the geography of image producing industries*. London: Sage.

Scott, A. J. (2004). Cultural products industries and urban economic development: Prospects for growth and market contestation in global context. *Urban Affairs Review* 39(4), 461 490.

Scott, A. J. (2005). *On Hollywood: The place, the industry*. Princeton: Princeton University Press.

Scott, A. J. and Storper, M. (2004). Regions, globalization, development. *Regional Studies* 37(6/7), 579 593.

Sell, S. K. (2003). *Private power, public law: The globalization of intellectual property rights*. Cambridge: Cambridge University Press.

Stolarick, K. and Florida, R. (2006). Creativity, connections and innovation: A study of linkages in the Montreal Region. *Environment and Planning A* 38, 1799 1817.

Storper, M. (1995). The resurgence of regional economies ten years later: The region as a nexus of untraded interdependencies. *Journal of European Urban and Regional Studies* 2, 191 221.

Storper, M. (1997). *The regional world: Territorial development in a global economy*. New York: The Guilford Press.

Storper, M. and Venables, A. J. (2004). Buzz: Face to face contact and the urban economy. *Journal of Economic Geography* 4, 351 370.

Sunstein, C. R. (2006). *Infotopia: How many minds produce knowledge*. Oxford: Oxford University Press.

Weber, S. (2004). *The success of open source*. Cambridge, MA: Harvard University Press.

Zukin, S. (1995). *The cultures of cities*. Oxford: Blackwell.

Relevant Websites

http://www.creativeclass.com
Creative Class.

http://www.creativecommons.org
Creative Commons.

http://www.culture.gov.uk
Department of Culture, Media and Sport.

http://www.onthecommons.org
On The Commons.

Crime/Fear of Crime

H. Koskela, University of Helsinki, Helsinki, Finland

Glossary

Commodifying Crime A process where people participate in the production of crime by consuming its representations as well as security services and products.

Crime A violation of a legal code which, nevertheless, is a culturally, geographically, and historically specific social construct.

Critical Criminology A subfield of criminology in which crime is understood as a part of complex social structure reflecting the power relations and structural inequalities in a society.

Culture of Fear A state of society in which fear is pervasive and perceived as omnipresent that overcomes normal encounters and, thus, has negative effects on everyday relations.

Fear of Crime A feeling which is partly unrelated to actual victimization, contributes to lowering the quality of people's lives, and reflects race, age, class and gender positions.

Globalization of Fear A post-9/11 state of society in which fear is amplified, is perceived as an all-pervading condition of living, and in which global tensions and local insecurities are intertwined.

Othering A process in which certain groups are labelled as dangerous or unwanted without a clear (crime related) reason.

Social Production of Fear The construction of fear that takes place in everyday warnings and discussions and in the cultural reproduction of ideologies about particular groups of people.

Urban Fortress Model A model of urban space where safety is guaranteed by segregation, protective and exclusionary design, and technological surveillance characterized by defence mentality.

Zero Tolerance A crime prevention strategy which aims to strictly deter minor crimes perceived as signals of a more serious criminal activity.

Crime Is a Social Construct

Crime is fascinating. It is present in news and novels, in horrendous fears, and thrilling imaginations. It has all the elements of both nightmares and daydreams. The media contributes strongly in sustaining the everydayness of crime and in bringing it close. Yet, simultaneously, true crime remains remote to most people. While crime seems to be ubiquitous, placing the issue under a more serious and academic consideration makes it quite complex. There is no simple, uniform definition of crime. What is considered to be criminal varies across cultures, geographical locations, and the periods of history. Crime can be described as violation of a legal code but this does not undermine the fact that legal codes themselves are culturally, geographically, and historically specific. Criminal law can also be considered as reflecting the interests of the powerful, or repeating structural inequalities in society. In others words, both crime and criminal law are social constructs.

Crime as a social problem has also long been a field of academic interest. Research addressing crime from various viewpoints has been conducted within the fields of criminology, sociology, anthropology, and geography. Geographers have been specifically interested in the connections between crime and space. The spatiality of the patterns and effects of crime has been a focus of such prominent books as *Crime, Space and Society* and *The Geography of Crime.* The geography of crime can be considered as simply examining the spatial distribution of criminal activity. From a wider perspective, there is more to it: crime has multiple and variable relations to space, and different types of crime have different spatialities.

Different spatial relations include, for example:

- global flows of crime (such as drug traffic or human traffic);
- crime which typically takes place in particular settings defined as hot spots (such as assaults or street violence);
- crime which challenges the boundaries of private and public space (such as violence against women);
- crime which aims to restrict other's right to space (such as hate crime);
- crime with a territorial function and a battle over space (such as urban graffiti); and
- more or less a spatial crime (such as fraud).

First studies concerned with the geography of crime had their roots in ecological and environmental understandings of crime patterns, diffusions, and criminal individuals. Research relied strongly on official definitions of crime and deviance. The old fashioned geography of crime and social ecology approach has been quite influential. Nevertheless, their conceptualization of space has been somewhat naive: space is understood as a relatively static arena within which crime takes places and distributes, often unevenly. Lately, geographical information systems (GIS) have become popular in mapping crime and in developing technologically based crime prevention strategies. In critical criminology, on the contrary, crime and criminals have been understood as a part of complex social structure within which relations of power and processes of criminalization must be grasped.

There are many ways of understanding the spatiality of crime. The development of the geography of crime includes various discourses on several subjects; some traditions have developed harmoniously, others have had polemic relations. There is no generally agreed account of the issue. The research on this subject can be outlined in relation to the development of the conceptualization of space: the development of the geography of crime echoes the more general development of geography from a science of surface towards the interpretation of images and meanings of places, and further, towards theoretically informed social science. Geography in the latter form is interested in interrogating space as both materiality and as produced in social practices. Differences in the notions of space are parallel to differences in the examination of crime.

Perhaps the best example, which crystallizes different dimensions of the social and spatial production of crime, is graffiti. It is clearly historically and geographically specific: it has not been criminal always, and at present it is criminal in some cultures and legal or tolerated in others. While the graffiti movement is global, the local practices vary. Accurately described as crimes of style, the spray painting practices are inherently paradoxical. Whether being criminal or not, graffiti is always socially contested: some people regard it as vandalism, others as street art. Furthermore, graffiti is a spatial crime in many senses. It clearly has a territorial function: it is all about marking spaces, either for political, subcultural, or aesthetic reasons. Graffiti is about owning space, and fighting against graffiti is a battle over urban space, questioning who has the right to be in what space. However, it is rather the symbolic conquering of space than literal occupation that matters. Graffiti can also be considered as an indicator of social conditions. It has long been used to identify ethnic or other tension zones related to social change in restless neighborhoods.

Fear Is a Social Construct (with Spatial Consequences)

Fear of crime has been identified as a problem in its own right. The relationship between fear and crime is more complicated than it seems to be at first sight: fear is partly unrelated to actual victimization. Fear of crime can contribute to lowering the quality of people's lives as much as crime does, and fear can thus be as great a problem as crime itself. Furthermore, as much as crime, fear of crime also has spatial dimensions.

Fear of crime is constantly modifying spatial realities. It leads people to take spatial precautions, such as avoiding certain places, using different routes in different hours, or not going out at all after dark. For some people, fear means fleeting moments of uneasiness, for others, constant and serious restrictions that profoundly shape their everyday lives. Temporality, as well as spatiality, is an important factor in fear: daytime is hardly ever thought to be as frightening as nighttime. Even the darkness of the night itself, a natural element of environmental difference frequently implicated in the fear of crime, is socially mediated. What makes people cautious is the social nature of night: what is going on and how others behave in particular places.

Fear cannot be regarded merely as a response to crime, but must be understood in a wider context of life. Fear has a socially produced reality of its own. The social construction of fear takes place in parental or spousal warnings, discussions among friends, warnings that are faced in discussions with anybody, and further, in the cultural reproduction and transmission of ideologies about women and men, children and adults, and the concept of family. All in all caution, mistrust, and tense social relations form a condition which has been called the culture of fear.

From the point of a geographer, the relations and events that affect people's lives begin to be of especial interest at the point where social experiences turn out to have spatial consequences. Three, nevertheless partly overlapping, traditions of the geography of fear can be identified. The first may be called a macro scale approach. In this tradition, research aims to identify differences in fear of crime by region. Individual perceptions of fear are mapped in order to get a picture of the most frightening areas of a particular city, for example. The research methods used have been either cognitive mapping or survey questionnaires. This tradition has its origins in the well known moral maps of the city developed by the Chicago School sociologists in the 1920s and 1930s. Information about the perception of fear, together with maps based on crime statistics, is used to identify areas at risk and vulnerable regions. In contemporary geography, research based on this tradition is still common. Computerization of geographical research has meant that this field is entering a new era; an increasing amount of research and development is conducted through the use of geographic information systems (GIS) to collect and store information on crime and to create GIS applications for crime prevention. A characteristic of this tradition is that it recognizes space as a surface.

Another research tradition examining the geography of fear may be called a micro scale approach. Instead of – or sometimes in addition to – regions, this type of research has focused on more specific places of fear. This tradition investigates experienced fear as socially differentiated and aims to identify both the physical forms and the social situations that arouse feelings of fear. The methods used have varied from questionnaire surveys to interviews. This approach was typically used between the 1970s and early 1990s, within humanistic and feminist geography. Research examined perceptions of either a whole city or a smaller area, a particular suburb or a park, for example. It was found out that fear may be attached to spaces which appear to contrast: both empty and crowded spaces, and both open and closed spaces. This tradition attempts to lead the conceptualization away from treating space as a surface and understands space as a combination of physical and social dimensions.

The third approach may be called a production of space approach. This expression applies to research which does not focus on fear in relation to the scale in question but rather is based on a particular understanding of space. It is acknowledged that the geography of fear cannot be presented just as a combination of frightening physical structures and social situations but that the physical and social dimensions are fundamentally intertwined. This approach is less interested in mapping the frightening regions or places, and refuses to see fear of crime as a private problem of an individual. Instead it endeavors to identify the everyday practices which produce fear or undermine it, and to understand the social power relations behind fear. Fear of crime is to be conceptualized through the emotion related and power related production of space. Space and social characteristics are seen as mutually modifying, interacting dimensions which deeply affect the nature of fear and shape it. Research aims to take account of the symbolic connotations embedded in space and the social relations which change the nature of space. The situations where fear is constructed or deconstructed, and the ostensibly private and personal but simultaneously power related decisions to go or not to go, demonstrate how space is produced in social practices.

Fear is deeply embedded in a sense of vulnerability – both physical and social vulnerability – and powerlessness. Fear has been shown to be most acute where communities feel a sense of powerlessness and a lack of local democratic control. The groups that suffer the most oppression – who most often feel marginalized and helpless – are the ones that are most afraid. For those who have few resources and who are at the periphery of society, basic existential insecurity is easily transformed into fear. There is a connection between isolation and fear of crime: the feeling of being left out in society and a weak position in social networks amplifies fear of crime. Fear is deeply related to race, age, and class positions – and especially to gender.

Gender Makes a Difference

Female victimization is of especial interest since gender is one of the most crucial factors affecting fear of violence and geographical mobility. Spatial constraints are intertwined with gendered power relations. However, it is important to understand that women do not form a uniform category but there are differences, that it is not an inborn quality of women to be fearful, and that many women are spatially very competent. The origins of fear are multiple ranging form specific events and changes to structural vulnerability and power relations. Women experience several kinds of offensive behavior which they may perceive as offensive but which would not necessarily be considered criminal. Minor offences can cause fear of a more serious attack by reminding women of their vulnerability and the possibility of victimization. The everyday experiences which contribute to women's fears – such as sexual harassment – are often trivialized. However, the incidents of sexual harassment in public space serve to remind women of sexual danger. Sexual harassment arouses fears of more severe sexual attacks by creating an atmosphere of insecurity and unease. Women are also indirectly reminded of their vulnerability by how the structures of power are represented in the construction of contemporary urban forms, as well as in the social landscapes of gender roles and sexualities.

Women are more likely than men to restrict their access to and activity within public space because of fear. Women's fear is crucially different from men's fear: it differs in its extent, its nature as well as its effects on women's lives. One reason for this is that they perceive a unique and serious threat barely felt by men: sexual violence. Violent attacks and sexual harassment remind women every day that they are not meant to be in certain spaces. When women adopt strategies to cope with fear they commonly distance themselves – in space and time – from potential attackers. Thus, public space can be considered as one territory to which men hold greater rights than women: a territory from which women are often excluded by fear of male violence. Fear produces a sense of space as something that needs to be constantly negotiated.

Space is produced by gender relations, and reproduced in those everyday practices where women do not – or dare not to – have a choice over their own spatial behavior. Experienced violence, threat of violence, sexual harassment, and other events that increase women's sense of vulnerability reinforce masculine domination over space. Fear can be argued to be a consequence of women's unequal status, but simultaneously it is contributing in perpetuating the gendered inequalities. The spatial constraints in women's lives are a reflection of gendered power relations.

A Vicious Circle of Defence

Fear reaches beyond individual relations – it has consequences to politics, economy, and urban structures. On a more conceptual level than the everyday experience, the social production of fear affects a whole range of settings and relations in contemporary societies. The widening gap between different social groups is associated with rising fear of crime and concerns about safety. The political and normative dynamics within the sociospatial order raises questions concerning the processes relating to disobedience and exclusion in urban space. Fear affects the urban form: both the building of housing areas and the development of public urban space. In the name of fear, urban space is becoming regulated and segregated on a moral basis which often defines social order in a notably narrow way.

What is characteristic to contemporary social relations is that experiences previously considered ordinary are constantly reformulated as dangerous. In this culture of fear carefulness, caution, and suspicion have grown to cover almost every aspect of life. When crime is seen as normal – an unpleasant but unavoidable part of life – an individual's strategy is not to endeavor to reduce crime but rather to protect oneself and one's family. People concentrate their energies in avoiding trouble and danger, not in understanding them. This leads to a condition where people believe that they should defend themselves against dangerous enemies rather than be concerned of their welfare and the prospects for rehabilitation. Social integration and solidarity are replaced with segregation and zero tolerance. Instead of being devoted to supporting the welfare state and taking care of others, people are committed to sustaining social control and spatial exclusion.

The social tensions are followed by spatial ones. Political and economic developments lead first to the segregation of social groups, and subsequently to spatial development which contributes to that segregation. Spatial constraints are increasingly important in understanding how social relations and power relations take shape in present societies. Some changes in cities predict rather pessimistic images of the future. The urban fortress model where safety is guaranteed by exclusionary design and technological surveillance has faced strong criticism. The writings of critical scholars have aimed to show where the defence mentality in its extreme forms can lead – Los Angeles being the most famous example of a city with protective design. Social polarization, inequality, and segregation are clearly legible in urban form. Spaces have been divided and polarized. Some parts have been transformed into controlled, guarded fortresses which are privately owned and maintained, whereas others have been neglected and left to deteriorate. Increased fear has caused social groups to flee from each other into isolated homogeneous enclaves. Design and urban policy are characterized by paranoia, protectionism, and lack of common interest. Various forms of surveillance are used to maintain the existing order.

Access control and spatial restrictions can be used as a tool for reinforcing the purification and homogenization processes of urban space. While increasing control may make some feel safer, it also generates cumulative fear, racism, and distrust. Restricted access and the increased control are creating an urban space which is more difficult to approach and stay in and is less lively and less spontaneous. Open and democratic public spaces are replaced by highly regulated spaces to which certain groups labelled as deviants are not welcome. The perceived probability of someone being dangerous, however, can be associated with completely legal behavior. The regulation of public space is also a process of 'othering'. City bylaws, public order acts, and other regulations contribute in defining not only what but also who is acceptable – who is respectable and who is excluded, who has the right to (be in) what space. Discussions about crime and fear are turned into discussions about right to urban space and geographical exclusion.

The massive expansion of protection is argued to lead to a vicious circle of defence. In housing areas, gated communities are the most extreme reaction to crime – and fear of crime. The development of gated communities can be understood as part of a more general shift towards security oriented architecture and planning in which the major concern is to manage space and to segregate different types of people from each other. If a city is designed in such a way that it looks like the environment is dangerous, the danger will be embedded in people's imaginations although there originally would have been no objective reason for that. What makes people cautious in these kinds of environments is not the actual experience of victimization, but instead, the ubiquitous probability of victimization. Fear is continuously reproduced by the design of urban space.

The Globalization of Fear

The terrorist attack on New York's World Trade Center (WTC) on 11 September 2001 has fundamentally changed the landscapes of fear. Fear has amplified, spread all over, and gained multiple roles it did not have before. Rather than being afraid of a stranger lurking in a street corner, people tend to be more generally worried: fear has become an omnipresent, all pervasive condition of living. Fear is simultaneously ubiquitous and enigmatic; danger is everywhere, yet difficult to identify and, thus, avoid. Global tensions and local insecurities are linked in a new way. This can be described as the globalization of fear.

The changing role of fear means, for example, that it is extensively used as a political tool – both for the Western regimes and for the 'others'. Fear can be spoken to, exaggerated, and articulated for particular political ends. In the name of personal safety, many people are ready to accept the kind of rules and constraints that would be perceived as intrusive or moralistic if the rea soning behind them would be based on any other notion. The mobilization of fear for political purposes is clearer in the post 9/11 era than ever before. Fear is a political tool which tames political life, not merely subduing political subjects but altering their guiding principles under the weight of its intense coercion. Fear has also increased tensions in racial and religion related relations. While gender relations are still meaningful in everyday encounters, they are often overpowered by a more in definite fear of the 'other', defined in terms of race, colour of skin, religion, culture, or country of origin.

For the sake of security, societies face widespread surveillance and control. The surveillance society spreads from urban space to national borders and con sumer behavior, being not only more intense, but in creasingly computerized, integrated, and globalized. The use of power is almost instinctive; people are controlled, categorized, segregated, disciplined, and normalized without any particular reason. Surveillance cameras and other monitoring systems follow people literally every where. When these devices are combined with visitors' registers, people finding tools, and biometric devices, the visible surveillance is linked to other forms of infor mation, creating a net of technological control. Further more, increased surveillance has strengthened the link between fear and the security business. This is apparent both in the marketing of security technologies – not only to governments and private companies but also to indi vidual households – and in the increase of private po licing. The blooming security business, however, will always remain unethical because the ones who run it will gain the more the more people are afraid.

The post 9/11 era has also meant that the meaning of the mass media has increased tremendously. The media are the main symbolic battleground of our times. The WTC attack literally showed how to be seen – not to hide – is what makes criminal acts meaningful and thrilling for the criminal. This type of dramatized rep resentations in the media will help fear reproduce itself. Furthermore, in and through these images, crime be comes a commodity. Both crime itself and fighting crime can make an exciting, fascinating show. Crime prevention programs present material from surveillance camera tapes creating an impression of the viewer being able to both verify and participate in what is happening in the world of crime. With a pretext of crime control, people are able to both moralize and peep into the field of criminal action. Crime becomes tied to the consumer society and people become tied to the production of it. By consuming the filming of real and fictional crime, people end up becoming part of the process of pro duction of crime.

While previously crime and fear of crime were rea soned as (partly) separate, in contemporary culture they seem to interrelate. As the dominating crime prevention strategy is to separate the dangerous 'others' – either locally or globally – fear has got a bigger role in all types of crime control strategies than ever before. Crime and fear of crime melt into each other.

See also: Agoraphobia; Chicago School; Defensible Space; Embodied Knowing; Emotional Geographies; Emotional Knowing; Feminism/Feminist Geography; Foucauldianism; Gated Communities/Privatopias; Gender in the City; Law and Law Enforcement; Moral Landscapes; People's Geography; Policing; Public Space; Public Spaces, Urban; Society-Space; Surveillance; Youth/Youth Cultures.

Further Reading

Bauman, Z. (2000). Social issues of law and order. *British Journal of Criminology* 40, 205 221.

Craglia, M., Haining, R. and Wiles, P. (2000). A comparative evaluation of approaches to urban crime pattern analysis. *Urban Studies* 37(4), 711 729.

Evans, D. J. and Herbert, D. T. (eds.) (1989). *The Geography of Crime*. London: Routledge.

Ferrell, J. (1996). *Crimes of Style. Urban Graffiti and the Politics of Criminality*. Boston: Northeastern University Press.

Furedi, F. (2002). *Culture of Fear. Risk Taking and the Morality of Low Expectation*. London: Continuum.

Garland, D. (2001). *The Culture of Control. Crime and Social Order in Contemporary Society*. Oxford: Oxford University Press.

Koskela, H. (1997). 'Bold walk and breakings': Women's spatial confidence versus fear of violence. *Gender, Place and Culture* 4(3), 301 319.

Mitchell, D. (2003). *The Right to the City. Social Justice and the Fight for Public Space*. New York: Guilford Press.

Pain, R. (2001). Gender, race, age and fear in the city. *Urban Studies* 38(5 6), 899 913.

Presdee, M. (2000). *Cultural Criminology and the Carnival of Crime*. London: Routledge.

Robin, C. (2004). *Fear: The History of a Political Idea*. Oxford: Oxford University Press.

Smith, S. J. (1986). *Crime, Space and Society*. Cambridge: Cambridge University Press.

Spencer, C. and Ratcliffe, J. (2005). *GIS and Crime Mapping*. Chichester: Wiley.

Valentine, G. (1989). The geography of women's fear. *Area* 21(4), 385 390.

Young, J. (1999). *The Exclusive Society. Social Exclusion, Crime and Difference in Late Modernity*. London: Sage.

Relevant Websites

http://www.homeoffice.gov.uk
British Crime Survey Data: Long Term Trends, Home Office.

http://www.ojp.usdoj.gov
Bureau of Justice Statistics, Office of Justice Programs.

http://www.be safe.org
Community Safety Series.
http://www.crime.org
Crime.org.
http://www.crimereduction.gov.uk
Crime Reduction.
http://www.csas.org.uk
CSAS (Community Safety Advisory Service).
http://www.renewal.net
Renewal.
http://www.surveillance and society.org
Surveillance and Society.
http://www.scotcrim.u net.com
The Scottish Centre for Criminology: Fear of Crime Research.

Critical Cartography

D. Wood, Raleigh, NC, USA
J. Krygier, Ohio Wesleyan University, Delaware, OH, USA

Glossary

Art Maps Maps created by artists.
Cartography Mapmaking, including allied theoretical and practical aspects.
Critique Assessment.
Eco-Mapping Mapping for ecological purposes, usually at the grassroots by environmental activists. Similar to green mapping.
Ethnocartography Study of non-Western mapping practices. Similar to indigenous mapping.
Green Mapping Mapping for ecological purposes, usually at the grassroots by environmental activists. Similar to eco-mapping.
Hegemony Domination of one group or person or idea over others.
Indigenous Mapping Study of non-Western mapping practices. Similar to Ethnocartography.
Mapmaking General practice of making maps, usually thought of as separate from more theoretical or philosophical issues.
Parish Mapping Large-scale, local mapping of areas usually for activist purposes or the collection and mapping of local knowledges.
Participatory Mapping and GIS Practices for gathering and mapping information whereby participation of nonexperts is encouraged.

The standard story about critical cartography is that it developed during the late 1980s and early 1990s in opposition to the hegemonic tradition of mapmaking as a progressive and value free transcription of the environment.

Often cited in this account is the deconstruction of a North Carolina state highway map by Denis Wood and John Fels in 1986; the publication of the first volume of Brian Harley and David Woodward's monumental reconstruction of the history of cartography in 1987; a series of polemics by Harley, especially 'Maps, knowledge, and power' in 1988, 'Deconstructing the map' in 1989 and 'Can there be a cartographic ethics' in 1991; the first of a number of papers by Robert Rundstrom reassessing the place of mapping among First Nations peoples in 1990; the publication of a paper critical of GIS by John Pickles in 1991; *The Power of Maps*, an exhibition curated by Wood for the Cooper Hewitt National Museum of Design in 1992, and the Smithsonian in 1994, and Wood's best selling *The Power of Maps* in 1992; David Turnbull's critique from the perspective of the sociology of scientific knowledge, *Maps Are Territories* in 1993; Doug Aberley's *Boundaries of Home: Mapping for Local Empowerment* in 1993; and Pickles' *Ground Truth* in 1995.

Binding these disparate efforts together was the nature of their critique. This was less about ferreting out bad maps or making better ones than about trying to lay bare, understand, and call into question the presumptions of professional cartography, 'professional cartography' referring at once to official mapmaking, the dominant map houses, and academic cartography.

As a story about the origins of a now pervasive critique of the assumptions and practices of professional cartography, the story has undoubted merit. The period did see a self conscious engagement with the fundamentals of cartographic thinking and behavior. But construing critical cartography in the narrow beam of contemporary critical theory forecloses an awareness of both a precedent history of critique within professional cartography and a much longer history of critical practice in mapmaking at large.

Mitigating against this vision of a critical past has been the conflation of cartography – a comparatively recent professionalization of mapmaking dating to the first third of the nineteenth century – with the whole of mapmaking, much of whose history preceded the development of cartography, and the balance of whose history has run parallel with it. A strategic move in this conflation was the effort on the part of Max Eckert and others to articulate mapmaking around a hegemonic vision of timeless principles (what Arthur Robinson called 'the essential cartographic process') in an effort to recast mapmaking as a science, as a *Wissenschaft*, rather than a form of human communication closer to writing. As a science, cartography was understood to progress from the solution of one problem to another – as other sciences were construed – and its past was recast as a seamless accumulation of knowledge and technique.

Early Critique in the History of Mapmaking

In fact, mapmaking has been perpetually transformed, all but dialectically, by successive critiques. These may not have been critiques in the sense inaugurated by Immanuel Kant, but they were emphatically critiques, embedded, often as not, in novel ways of making maps, novel map subjects, or both. A classic example is the world map published by Gerard Mercator in 1569 and the projection implicit in it. This was not, as it is so often presented, the solution to an urgent problem (as is made amply plain by the fact that it took two centuries for the projection to be widely adopted), but neither was it merely a novelty. It was, however, deeply critical, of both the conical Ptolemaic projections popularized by Renaissance scholars and the plane charts (*portolanos*) long used by mariners.

This is not something we have to infer. Mercator made his critique explicit in the 15 texts scattered over his map. About the Ptolemaic maps, for example, Mercator wrote that, "Indeed the forms of the meridians as used till now by geographers, on account of their curvature and their convergence to each other, are not utilizable for navigation; besides, at the extremities, they distort the forms and positions of regions so much, on account of the oblique incidence of the meridians to the parallels, that these cannot be recognized nor can the relation of distances be maintained." As for the charts of the mariners, "the shapes of regions are necessarily very seriously stretched and either the longitudes and latitudes or the directions and distances are incorrect; thereby are great errors introduced." Nor was it solely with respect to form that Mercator was critical. Among other things, Mercator abandoned the Ptolemaic prime meridian for another, and adduced a north polar land mass, a second Greenland, and a huge protuberance in southwest South America.

Traditionally staged as "a paradox of advances and retrogressions" in the drama, *The Progress of Cartography*, Mercator's map is praised for its ingenuity and condemned for the 'cartographical mistakes' it disseminated. In fact, both the map's form and its content are more usefully approached as embodiments of Mercator's critical engagement with his sources. Not only did nautical charts disagree with each other, as did the maps of the scholars, but also the two sets of maps were difficult to reconcile, a compelling problem for Mercator whose life work consisted in compiling maps from the maps of others. "I had to wonder," Mercator once wrote a friend, "how it could be that ship courses, when the distances of the places were exactly measured, at times show their differences of latitude greater than it really is, and at other times on the contrary, smaller … [T]he matter caused me anxiety for a long time, because I saw that all nautical charts, by which I was hoping especially to correct geographical errors [that is, errors in the maps of the geographers], would not serve their purpose."

Confrontation with conflicting reports inescapably brings the problem of knowledge to the foreground, raising its contingent status before even unwilling eyes. Mercator's critique of the *portolanos* and the Ptolemaic conics famously took the form of his eponymous projection, a spatial frame which was no sooner published than it too became the subject of a critique that continues into the present. Among those objecting to it first were the navigators for whom it was expressly designed. They found it an oddity on which it was hard to measure distances as a consequence of the projection's increasing scale with distance from the equator. It is this characteristic that has sustained the most extended critique. J. H. Lambert's 1772 critique, which took the form of a veritable counter projection, shrank what Mercator had stretched to preserve area instead of shape. Lambert's cylindrical equal area projection became the first of a family of such projections driven by resistance to the world brought into being by Mercator. Among them were those of James Gall in 1855; Walter Behrmann in 1910; Trystan Edwards in 1953; and Arno Peters in 1967.

Gall critiqued Mercator's commitment to navigators, writing, in 1855, that "Mercator's projection sacrifices form, polar distance, and proportionate area, to obtain accurate orientation for the navigator; whereas to the geographer, form, polar distance, and proportion of area are more important than orientation," which, while reversing it, recalls Mercator's critique of the Ptolemaic conics popularized by the scholastic geographers. Peters, on the other hand, critiqued Mercator for being "the embodiment of Europe's geographical conception of the world in an age of colonialism." Though like Lambert and Gall, Peters was not a cartographer (Lambert was a physicist and mathematician, Gall a clergyman, Peters historian), Peters had no hesitation about critiquing the profession for clinging to a "closed body of cartographic teaching which has developed into a myth." In response to Peters, an embattled cartography united to condemn all rectilinear world projections, hysterical reaction that underscored the depth of the wound sustained from Peters' critique of the profession's epistemological foundations.

Critique within the Profession of Cartography

Not all critique has come from outside the profession. Though some internal critique has echoed that of outsiders, more has been directed toward transforming cartography into a science. Max Eckert, for example, published the first volume of his 1500 page *Kartenwissenschaft* in 1921, the second in 1925. Although vehement in its opposition to the use of non equal area projections in geography (especially the Mercator), the *Kartenwissenschaft* was most influential in shaping cartography's reconstruction as a science. Though generally traditional in his goals – who could ever have disagreed with Eckert's demand that maps be "correct, complete, appropriate, clear and distinct, readable, and handsome"? – the route Eckert sketched for reaching these goals was distinctive, largely the application of psychology to map design. "The question," he wrote, "whether an economical map should demonstrate the distribution of only a single phenomenon or of a lot of them will not bring anyone to confusion if his thinking is logically based and if the designer has paid regard not only to the scale and to the purpose of the map, but also to the visual capability of the human eye and to the receptivity of the human brain," adding that, "It would be an extraordinary progress if a scientifical cartographer and a psychologist could jointly proceed to empirical tests clearing up by which map charge the human eye and the human brain will be overcharged."

Since it may be doubted that Eckert's question had ever brought many into confusion (who, making a map, or any communication, would want to 'overcharge' the human brain?), Eckert's program may be more appropriately appreciated as a critique of the practices of his peers, a critique arising, in fact, from what one of Eckert's memorialists has referred to as "Eckert's rage against overcharging maps with signs." Eckert's desire to 'scientifically' validate his Apollonian preferences resonated with enough others that from the 1950s onward, academic cartography was heavily invested in the psychological testing of map readers' abilities. Arthur Robinson's textbook, *Elements of Cartography*, encouraged this point of view, especially its later editions. While acknowledging that cartography was not a science, Robinson's third edition (1969) stressed that cartography "employs the scientific method in the form of reason and logic in constructing its products … It has its foundations in the sciences of geodesy, geography, and psychology." Eric Arnberger's text, *Handbuch der Thematischen Kartographie* in 1966 also followed Eckert in attempting to impose order on the "wild branch that has grown untended and unpruned on the trunk of the topographic map" by formulating a theoretical framework for the establishment of cartography as a *Wissenschaft*.

It was the unrelenting focus on how undergraduate students read various arrangements of graduated circles, line widths, and color schemes (undergraduate students were invariably the subjects of the psychological tests) that ultimately led to the countervailing internal critique of the profession, its construction of mapmaking as a science, and its construction of its history as a value free and progressive transcription of the environment. Wood, Fels, Harley, Woodward, Rundstrom, and Pickles were all professionally involved with cartography. Their critique aimed at overturning the paradigm of Eckert and Robinson by shifting attention from the 'form' of the map, with which the profession was obsessed, to its 'meaning for behavior'. Instead of asking whether the brain was overcharged by the density of symbols, these critics asked how the body of the subject was constructed by the map, that is, how the map oppressed, subjugated, or otherwise impinged on people.

This shift in commitments first surfaced in differences over Peters' projection, and Peters' explicit outsider's critique of cartography's political, indeed colonialist, even racist, dimension. Most professionals, and the official professional organs, pretended to either outrage or bemusement, wondering how a projection (only a mathematical formula after all!) could be political in the first place, but nonetheless took time to swipe at Peters' projection as ugly, not Peters' (it's identical to Gall's), or otherwise inappropriate (world projections should never be rectilinear). Internal critics, on the other hand, understood and in their various ways empathized with Peters' project. By 1994 Jeremy Crampton had characterized the affair as "cartography's defining moment." In 2003 Wood claimed that the affair had been, "in its way, the death knell of the profession."

The Outside Critique

Paralleling the internal critique was an external one. Doug Aberley, a bioregional planner, published *Boundaries of Home* in 1993, but the book reflected mapmaking that had been going on for a while, in the case of First Nations mapping, for quite a while. First Nations, or Indigenous, mapping offers a critique of official mapmaking with respect to its prerogatives, form, and content. The origins of the contemporary movement may be traced to the early 1970s, especially to the publication in 1976 of the three volume *Inuit Land Use and Occupancy Project,* which pioneered the use of individual map biographies. In these, "hunters, trappers, fishermen, and berry pickers mapped out all the land they had ever used in their lifetimes, encircling hunting areas species by species, marking gathering locations and camping sites – everything their life on the land had entailed that could be marked on a map."

Neither the mapmaking of their Inuit ancestors nor that of contemporary cartography, these 'map biographies' inaugurated a new trajectory in the history of mapmaking. If the maps were scientific, it was not in the vein of geodesy, geography, and psychology as Robinson had dreamed, but in that of ethnography (the practice has been called a kind of ethnocartography), though more significant by far was their use in land claims negotiations. Thanks to a Royal Proclamation of 1763 which asserted an Aboriginal title – a form of property right arising from long and continuous use and occupancy of land prior to the arrival of colonial peoples – their maps enabled the Inuit to assert an Aboriginal title to 2 000 000 km^2 of Canada. A land claims settlement was proclaimed in 1993, and in 1999 the new Territory of Nunavut was created. When its 19 member assembly assumes all governing power in 2009, the Inuit of the former Northwest Territories will be among the first indigenous peoples in the Americas to achieve self government in recent times.

The role of indigenous mapping in this process was lost on no one. Similar mapping projects were immediately initiated among the Inuit, Settlers, and Naskapi Montagnais of Labrador, the Dene of the Mackenzie River basin, the Indians of the Yukon, and the Inuit and Cree of northern Quebec, among others, and the idea rapidly spread around the world. An important benchmark was the publication of *The Nunavut Atlas* in 1992, and the identification, also in 1992, at the United Nations Rio Summit, of community based mapping, using both local and computer technologies, as a key research, community building, and planning method. Publication of the special issue of *Cultural Survival Quarterly: Geomatics: Who Needs It?* in 1995 with its examples of ethnocartography from around the world, Nancy Peluso's articulation of counter mapping, also in 1995, and *Maya Atlas: The Struggle to Preserve Maya Land in Southern Belize* in 1997, were other signal moments. The 1998 establishment of the Aboriginal Mapping Network marked a coming of age for the movement. In 2003, some 120 Aboriginal mapmakers from across North America and as far away as Panama, Taiwan, and Malaysia met for the third international GIS mapping conference hosted by the Aboriginal Mapping Network.

This wholesale assault on the presumptions of professional cartography extends to its most fundamental categories. "Mapping, and cartographic technologies have progressed immensely over the past decades," write the initiators of a project for the International Polar Year 2007–08. "And yet, the representation of landscapes, topology, toponymy, and landforms remains focused on just that – land. North American topographic maps continue to represent landscapes as interpreted, described, and named over a history of European, American, and Canadian exploration. In Canada and Alaska, efforts have been made, and are currently underway, to begin 're mapping' the North according to the rich diversity of Inuit knowledge (e.g., place names, oral history, and land use and occupancy projects) that is generally overlooked in conventional mapping initiatives. However, the large expanses of blue that delineate the Arctic Ocean and Hudson Bay, among other major water bodies, are left relatively empty in most maps. These 'blank' areas are actually ice covered white expanses for three quarters of the northern year." The project proposes to map each Inuit community's past and present sea–ice use patterns using methodologies and philosophies patterned after those of the 1976 Inuit Land Use and Occupancy Project.

As suggested by this example, First Nations mapping has ties to other forms of what has become known as counter mapping, counter, that is, to the mapping of professional cartography. These linkages, proliferating today – to bioregional mapping, community mapping, public participation GIS (PPGIS), participatory rural appraisal, green mapping, and still others – had already been anticipated in Aberley's book which, while pointing to the achievements of the Inuit in the Northwest Territories and Labrador as exemplars, as also to Common Ground's Parish Map Project and others, was essentially a bioregional primer. What united most of Aberley's projects, of course, was an emphasis on the local, but most of them also shared an indifference to, if not a disdain for, professional cartography and its obsessions. Invariably the focus was on significance at the expense of formal values as when, in a complete rejection of Eckert's overriding concern, Aberley advised, "Alternately, you can either purchase multiple copies of your base map, or chart all your data directly on the base map. The latter technique will create an extremely busy image which will be useful in conveying the richness of natural and human activity in your neighborhood area, but can be short on clarity."

More difficult to deal with was professional cartography's spatial epistemology, including its reliance on discrete boundaries, especially when these came bundled in computer mapping technologies, like GIS and GPS, exported to support indigenous mapping activities. Indeed, counter mapping called forth critiques from the very practitioners who first advocated its adoption, aware as they were of the conceptions of property implicit in cartographic conceptions of space, conceptions wildly at odds with those of many of the peoples attempting to exploit computer mapping's potential. As Peluso wrote as long ago as 1995, "The key theoretical questions about the impacts of counter mapping on resource control are to what degree new notions of territoriality reflect older ones; how the reinvention of these traditions benefits or works to the detriment of customary practice, law, and resource distribution; and how the intervention of NGOs … affect the villagers' access to and control over … resources." In contradistinction to Bernard Nietschmann's 1995 insistence that an indigenous map made with computer technology, "… will have transcendental powers because it can easily be translated by everyone everywhere; it transcends literacy; [and] it is visually comprehensible," comes Peter Walker and Pauline Peters' 2001 caution that, "The job of mapping should not end with the drawing of boundaries; where social scientists assist social groups to draw maps, it is crucial that they also document and communicate what these boundaries mean for local people," that is, it has to be an ethnocartography in fact.

Ethnocartography, eco mapping, PPGIS, anticipatory rural appraisal, green mapping, Parish mapping, all these and others have fed the stream of maps, growing in volume ever since the late 1950s, that have been made by artists. From the hands of surrealists, situations, pop artists, conceptual artists, Earth artists, eco artists, in stallation artists, and others have come a flood of maps challenging not simply Western capitalist society, but the authority of Western capitalist cartography to map the world. Art maps contest not only the authority of pro fessional mapmaking institutions – government, business, and science – to reliably map the world, but they also reject the world such institutions bring into being. Art maps are always pointing toward worlds other than those mapped by professional mapmakers. In doing so, art maps draw attention to the world making power of professional mapmaking. What is at stake, art maps insist, is the nature of the world we want to live in. In pointing toward the existence of other worlds, real or imagined, map artists are claiming the power of the map to achieve ends other than the social reproduction of the status quo. Map artists do not reject maps. They reject the authority claimed by professional cartography uniquely to portray reality as it is. In place of such professional values as accuracy and precision, art maps assert values of imagination, social justice, dreams, and myths; and in the maps they make hurl these values as critiques of the maps made by pro fessionals and the world professional maps have brought into being. Artists insist that their maps chart social and cultural worlds every bit as 'real' as those mapped by professional cartographers. Some, Guy Debord among them, have explicitly called for a 'renovated cartography' as a form of intervention. The project of art mapping is nothing less than the remaking of the world.

Narrowly construed, critical cartography was a late phase in the history of cartography, a professional mo ment in the broader history of mapmaking; but criticism more broadly understood has been an aspect of map making from its earliest days. In some sense it is inherent in the practice, in which even the most fundamental forms of field mapping involve the self conscious reso lution of conflicting reality claims. In more common forms of mapmaking, compilation surfaces the problem of knowledge in an inescapable fashion (as it did for Mercator), as do symbolization, generalization, and classification. Undoubtedly professionalization, with its guidelines and texts, handbooks and standards of practice, constituted a form of anxiety reduction during a period of rapid growth in mapmaking, pushing aside any need for an overwhelmed class of mapmakers to directly confront the problem of knowledge. But in the end, even for professionals, the problem would not go away, and hence the rise of the internal critique. Outside the pro fession, conflicting reality claims lay at the heart of the practice. It may be anticipated that the critical claims making of these nonprofessionals will increase ex ponentially as the technologies developed over the past century and a half by the professionals are increasingly embedded in increasingly accessible online mapping and mapmaking tools. It may not be anticipating too much to look forward to the day when mapmaking remakes itself into a completely critical practice.

See also: Cartography, History of; Counter-Mapping; Critical GIS; Mapping, Philosophy; Mapping, Race and Ethnicity.

Further Reading

Crampton, J. W. (2003). *The Political Mapping of Cyberspace*. Chicago: University of Chicago Press.

Crampton, J. and Krygier, J. (2006). An introduction to critical cartography. *ACME: An International e Journal for Critical Geographies* 4(1), 11 33. http://www.acme journal.org/vol4/JWCJK.pdf (accessed in June 2008).

Harley, J. B. (2001). *The New Nature of Maps: Essays in the History of Cartography.* Baltimore, MD: Johns Hopkins University Press.

Pickles, J. (ed.) (1995). *Ground Truth: The Social Implications of Geographic Information Systems*. New York: Guilford Press.

Pickles, J. (2004). *A History of Spaces: Cartographic Reason, Mapping, and the Geo Coded World*. London: Routledge.

Wood, D. (1992). *The Power of Maps*. New York: The Guilford Press.

Critical Geography

U. Best, Technische Universität Chemnitz, Chemnitz, Germany

Introduction

Critical geography is an increasingly international movement that interprets geography as critical social practice. The term 'critical geography' came into use in the 1970s in Anglophone geography. At that time, it was not used as a label for a single approach or subdiscipline of geography, but as an attribute – 'a critical geography', not 'critical geography'. The term was preceded by a number of different terms connected with other moments in the development of critical geography. The idea of a social geography (again: not 'social geography') was first used by Reclus and others in France in the 1870s and 1880s. Later, a materialist, Marxist geography was aimed at by Wittfogel and Horrabin in the 1920s and the French Marxists in the 1940s and 1950s, as well as German student groups in the 1970s. The French Marxist geographers also introduced the idea of a committed or active geography. In the 1970s in North America and the UK, 'radical geography' was the term chosen by the movement. This term is also used today to denote this specific phase and situation of critical geography.

The term 'critical geography' is sometimes understood to convey a distancing from this specific radical geography – as less radical or less Marxist. In general, critical geography is used as an umbrella term for all these different approaches and movements, (among others) radical, Marxist, feminist, left wing, or activist geographies and also as a term that bridges the differences between the different usages and approaches developed from these earlier movements. More specifically, it is also used for the developments since the 1990s.

The article first introduces different definitions of critical geography, then portrays the development in France, Germany, and in the English language, and finally discusses the International Critical Geography Group and the debates in international critical geography.

What Does 'Critical' in Critical Geography Mean?

Initially, the meaning of 'critical' was not derived from a tradition of geography itself, but drawing on a wider social theory. For example, it could refer to a Marxist approach, as with Wittfogel, who does not elaborate the theoretical aspects of critique but starts off criticizing mainstream geography from a Marxist standpoint. The first possibility for a critical geography is therefore to develop the spatial aspects of critical social theories.

This way, Marxist geography, for example, deals with the role of space in Marxist theory; feminist geography deals with spatial or material aspects of gender relations; postcolonialist geography deals with the role of space in postcolonialist theory, and so on. This perspective however assigns a minor role to geography – that of applying the theories developed in other disciplines, filling the gaps. In traditional geography, concepts of landscape and region served as the theory of geography. Critical geography in this sense does not have a theory 'of its own', but draws on social theory and criticises geographical concepts. With the further development of critical geography, however, different meanings of 'critical' emerged.

The Critique of Geography

The Critique of Geographical Mystification

Critique of geography can refer to a critique of geographical arguments in politics and the society, which are often used to mystify power relations. One well known example is the critique of ghetto formation. In a naturalizing discourse, the 'problems' in segregated neighborhoods can be blamed on the 'area' and its inhabitants. An argument of critical geography would be that the problem is better approached through an analysis of class relations, capitalist urban development and racism, and its expression in a specific spatial constellation. Another example of a critique of geography is the critique of geopolitics, as first exercised by Wittfogel and later in critical geopolitics. The traditional formulation of geopolitics would posit a state with its territory, and then outline specific problems emerging from the location of the state, its size, and the position or shape of its borders. Geopolitics as practiced by the state would picture an issue as a problem for 'the country' – for example, the location of a pipeline, or the significance of a certain (foreign) territory for the state. A critique would first of all criticize the idea of the state as an acting entity or a community of interest and then analyze the use of geographical ideas in the interest of the state elites.

The Critique of the Discipline

The second aspect of the critique of geography is the critique of the discipline and its role in this mystification or naturalization of power structures, or of its direct participation in oppression. The critique of the discipline can take the shape of a critique of contemporary practices of mainstream geography, or aspects of it, or of the past of the discipline and its legacies. An example for the latter is the criticism of the involvement of British geography in imperialism, of German geography in national socialism, or of French geography in colonialism. Further critiques of the practice of geography attack it as Eurocentric or as 'white'. Also, more recent and contemporary engagements of mainstream geography (or more specific geographies) are the objects of a critique. Graham, for example, describes his participation at (and criticizes) a conference on urban warfare and antiguerilla tactics, which was geographical if not explicitly in its discipline, so certainly in its content.

The Critique of the Internal Structures of Academic Geography

A further field of practical commitment of critical geography is the critique of the situation that determines the practice of (critical) geography – the structures of academia, research, and communication. These can take the form of the critique of academic institutions and their rules (e.g., patriarchal systems). Currently, in the international critical geography movement, the critique of the power structures of international geography plays a main role (see below).

Revolutionary and Counter-Revolutionary Theory

Writing in the early 1970s, David Harvey considered the question of how geography can be critical in more detail. In his 1972–73 article, 'Revolutionary and counter revolutionary theory', Harvey criticizes Kuhn's model of the development of scientific paradigms which pictures science progressing to an increasing understanding of the world, by a constant improvement of theories and their correspondence to the observed reality. Harvey criticizes Kuhn's model as idealistic due to its 'abstraction of scientific knowledge from its materialistic basis'. He proposes a more society conscious model for the evolvement of theories, pointing out that science usually serves the interest of middle class academics, of the state or business elites that fund research and teaching. On these grounds, he criticizes the 'quantitative revolution' as determined by three factors: a genuine interest in new explanations, an interest in academic distinction, and a demand by the state for better means of planning and control.

His idea of revolutionary theory goes further, however. With the wider acceptance of a revolutionary theory, social changes would follow. Harvey bases this view of the role of the intellectuals on Gramsci, where they would play the avant garde in a struggle for hegemony. A revolutionary theory does – in Harvey's concept – not attempt to contribute to the development of capitalism, but aims at revolution, and offers a radical new formulation of theoretical understanding. A counter revolutionary theory, on the other hand, would be meant to prevent social changes from happening by confusing or distorting the revolutionary theory. Counter revolutionary theories emerge from similar positions as revolutionary theories but only aim at empathy or temporary solutions. In his initial article, Harvey considers Gunnar Olsson's linguistic approach as potentially counter revolutionary. In addition, Harvey defines '*status quo* theory' to be a theory that only upholds the societal *status quo* because it is limited to the questions posed by those that profit from the *status quo*. A *status quo* theory for him is exemplified by the Chicago school model of urban structure – it does not propose any change to social inequalities apart from the individual moving up the ladder.

Harvey proposes a model that considers science to be always connected to the state and determined by the class relations in a state. A given phase in the capitalist mode of production requires a certain set of theories which produce and legitimize its ideology and practice. This was the case with, for example, traditional geography and geopolitics which were required by a nationalist, imperialist state. In a changed social situation with a new mode of production – which is also characterized by a different production of space – a different geography is required to legitimize the power structures. For example, while the nation states of the nineteenth century required national (but also imperial) geographies, the globalized world economy requires geographies with a global, transnational perspective. Multiculturalism can therefore be interpreted as an ideology that aims at the production of a flexible subject that is most useful for current forms of production and government.

Geography as Critical Social Science

Derek Gregory, another geographer who explored geography as a critical science, bases his concept partly on the critical theory of the Frankfurt school. Gregory's argument in *Ideology, Science, and Human Geography* is initially similar to Harvey's. Social science is determined through the mode of production, Gregory argues with Althusser. The role of critical theory in this situation is much more complicated, however. Traditional theory produces a division between the scientist and the political citizen – the scientist is not meant to have a political interest and mainly fulfills his/her role as a specialized worker for the state or the economy. Critical theory, in the words of Horkheimer as taken up by Gregory, questions this division and aims at a politically conscious role of science, one that denies its function for the elites. Critical theory then does not try to improve the functioning of the structures of power. It is not, in other words, a 'constructive' critique that comes up with suggestions of how to better solve a problem defined by the elites or the ruling class of a society (for Harvey, this would be a *status quo* theory, for Horkheimer, it is traditional theory). Critical theory is a critique that aims at the transformation of the society. This definition includes two further aspects of critical theory. If it does not support the structures of power, it must first become aware of these power structures – critical theory therefore requires an analysis of power. Second, it requires an analysis of the role of science in these structures of power ('traditional theory'). Finally, it needs to develop a position critical of these power structures and of the role of science. Similar to this, Gregory formulates three propositions for a critical geography: it must provide a structural explanation of a condition, a reflexive explanation of the frame of reference of the scientist, and it must be a committed explanation that expresses its emancipatory interest.

Critical Geography as a Standpoint Epistemology

This leads to further contributions to the question of what makes critical geography critical. Where Gregory (with Horkheimer) argues for a self critical scientific practice, this is in some ways reflected in the standpoint epistemology developed by feminist philosophers of science, social scientists, and anthropologists and taken up by geographers. While in classical Marxist theory, the standpoint to be taken was that of the working class (and mainstream theory was denounced as bourgeois – that is, attributed to a different standpoint), and Horkheimer denied the firm connection of critical theory to any class or societal group, the issue of the role of the own position in a society gained much importance in feminist theory. Standpoint theory means that science is made explicitly from the perspective of a gendered subject, more concretely, the very specific subject that is the author or researcher. Science can, therefore, not be unbiased but is always biased – and in order to reveal the male bias of most of social science, standpoint epistemologies stress their different take on the topics. This also includes an analysis of one's own involvement with the research process and the 'objects' of research, and finally a questioning of the outcome and validity of the research. Knowledge is understood to be always situated in the specific social context of its production. This standpoint epistemology also leads to the development of the 'positionality' approach, which similarly stresses the reflection of the own role in the research process and the specific problems (or chances) emerging from it.

Geography and Critical Practice

Critical geography also analyzes social practices considered as critical, social movements or radical politics. Different terms have been used to characterize the geographies. Routledge chose the term 'geographies of resistance' for his analysis of the spatial relations of social movements like the Chiapas movement, an alliance of indigenous and 'other' excluded people in Mexico who have formed their own structures of subsistence and self governance. He also termed the places of these struggles terrains of resistance. Blunt and Wills analyzed the spatial relations of anarchism, Marxism, feminism, queer theory, and postcolonialism, as well as the geographical concepts building on these theories, and termed these 'dissident geographies'. Others analyzed the 'autonomous geographies' of local social movements or the 'politics of place' in the struggles in the London docklands. In economic geography, the idea of alternative economic spaces has gained some attention. These geographies of critical practice refer to the contestation of the structures of power and argue that power is nonmonolithic, that change can take place through human agency, both locally and (possibly) on a wider scale. In *Spaces of Hope*, Harvey goes so far to propose a utopian geography.

These geographies also express complicity or support of the movements they analyze, and with their analyses and in their scientific practice also attempt to make a contribution to them. They point toward a commitment beyond academia, to struggles that geographers are involved in or are sympathetic of. In feminist geography, for example, researchers often take sides with the groups they write on and are involved in their struggles. This issue has also been discussed under the question of activism and academia.

National/International Critical Traditions

Critical geography in almost all of the aspects considered above is not 'new', but there have been critical perspectives and critical thinkers (and practitioners) in geography from the start of the academic discipline in the nineteenth century. Since the late 1960s, for example, radical geography in North America and the UK has tried to draw attention to Pyotr Alexeyevich Kropotkin (1842–1921) as a radical political figure in the history of geography, highlighting his political significance as well as some of his concepts for a better society. Kropotkin was a highly international figure, a Russian anarchist involved in uprisings in Central Europe, in struggles in Germany and in Russia. Many of the early critical thinkers in geography were highly international, but their work was contained in situations that were also determined by language, national academic structures, and disciplines as well as the political conditions. While the following sections will be dedicated to the development of three main national/language contexts of critical geography, there are of course a number of other such settings for critical geography. Japan, for example, has a relatively long tradition of critical geography, originally developing a materialist/Marxist approach to economic geography. Other strong trends of radical geography could be located in Scandinavia or in Brazil.

Critical Geography in France

An acquaintance of Kropotkin's was the French geographer Élisée Reclus (1830–1905). Reclus had studied in France and Germany, had traveled widely, and was a prominent figure in the anarchist movement. He published widely on anarchism and was involved in the struggles of the Paris commune in 1871. He was also an important figure in French geography. His main work is the 32 volume *Nouvelle Géographie Universelle*. His interest in his geography lay in the conditions of human beings and how to improve them. Reclus – in contrast to the second founder of French geography Vidal de la Blache – included political and social aspects in his geography. He is one of the inventors of the term 'social geography', which he defined as drawing on the social facts of class struggle, of balance and change, and of the sovereign decision of the individual (**Figure 1**).

However important he was at his time, Reclus did not create a school of followers and remained at the margin of the 'classic French school of geography' dominated by Vidal de la Blache. He did not enter the canon and was – until recently – regularly omitted from the history of French geography.

The legacy of Reclus and of critical geography in France is therefore fragmented. Two generations later, a different kind of critical geography emerged. In the interwar years, the questions of a 'géographie engagée' had been raised by Blanchard, who was not radical in the sense of today, but who inspired a group of younger geographers who all became interested in Marxism. Jean Tricart, Pierre George, and Jean Dresch were members of the French communist party and after 1945 were all based in Paris. They also published geographical articles in the communist journal *La pensée*. Dresch analyzed French colonialism mainly in North Africa, and Tricart, who later became a supporter of applied geography, even proposed a difference between Marxist and bourgeois geomorphology. George introduced a Marxist economic geography. They remained important figures in French geography and set the trend for French economic geography to become strongly influenced by Marxist concepts. This was particularly important as an opposition to the traditionalist Vidalian school, in order to counteract the power of geomorphology in the discipline and to establish a clear perspective of human geography. Dresch and George were also influential as academic sponsors of some of the next generation of critical geography in France. Dresch supervised Yves Lacoste and André Prénant; George supervised Raymond Guglielmo and Bernard Kayser. There was a circle of communist students of geography in the late 1940s and early 1950s which held regular meetings and discussions, and some of whom started a journal called *Géographie Nouvelle.*

The directions these geographers took were strongly orientated toward the analysis of French industry, of regional inequalities and urban rural relations, and of urban development. Research in the Third World also played a strong role, and its anticolonialist stance was later fuelled by the Algerian war of independence.

In the 1970s, several new journals were founded. The early issues of *Espaces et Societés* (founded in 1970 by Henri Lefebvre and Anatole Kopp) give evidence of the strong trend in Marxist spatial analysis and urban theory (although with few contributions from the discipline of geography). Some of the articles from the journal were translated in *Antipode.* In 1976, Yves Lacoste founded *Hérodote*, initially subtitled *Stratégies Géographies Idéologies* (later *Revue de Géographie et de Géopolitique*) (**Figure 1**). The early issues of *Hérodote* bring together all the force of critical geography – a critique of mainstream geography, a critique of geography as a tool for the military, as an ideological tool of the state, and a critique of the institution of geography and the 'professors' geography'. Issues were themed on the military, on struggles in South America, on Africa or on a critique of the traditional regional approach (the 'paysage'/landscape approach). Main conceptual debates in early *Hérodote* were on the question of the 'militant' geographer, and on whether critical geography should rather follow a Marxist political economy approach or an analysis of the state and the military. The contributors to the journal in general followed the second line of investigation. The journal was radical in its style. Some articles were printed in two columns, with the main column presenting the body of the article, and the side column adding comments (and criticism) by other members of the editorial group. *Hérodote* also documented interferences of government institutions with another new journal, *Espaces/Temps*. This journal, founded by Jaques Lévy, was not radical throughout but regularly carried articles connected to critical geography, discussing Marxism in geography or critiques of planning.

The 1970s not only saw new developments in critical geography. The new branch of critical geography – as evidenced in *Hérodote* – criticized Marxian political economy and established a perspective analyzing the discipline, the state, and the military. However, Marxism had been so strong in French geography since the 1950s

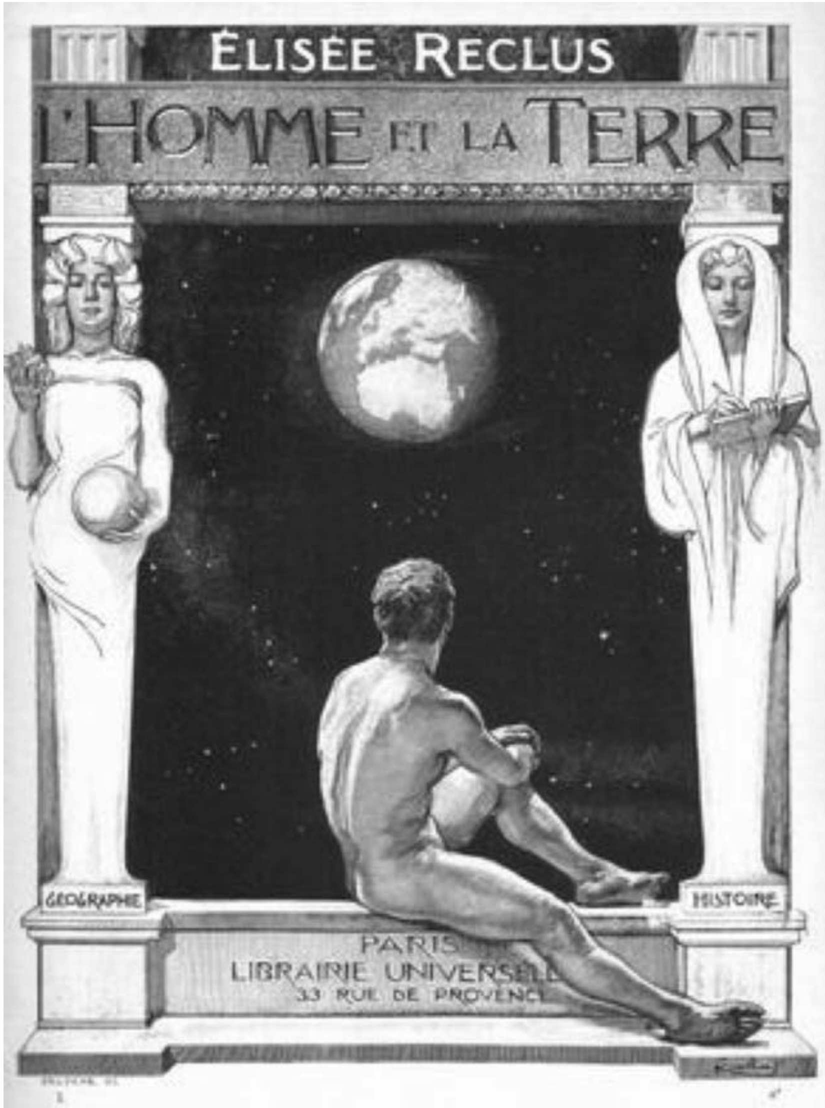

Figure 1 Title of Reclus' L'homme *et al.* Terre (vol. 1, 1905). Modern feminist geography would note that a male figure is contemplating about the Earth, but the female figures also represent geography and history.

that mainstream geography also now openly debated Marxism (and criticized that it was 'not geographical enough'). Debates in nonradical journals give evidence of the threat that Marxism was perceived to be, especially since it had also become an international movement in geography. *L'Espace Gèographique* dedicated a long discussion to David Harvey's 'Social justice and the city'. Another issue of the journal took on the issue of geography and 'pouvoir' (power). At this time, radical geography had started to define the issues and mainstream geography in many cases only reacted (**Figure 2**).

Since the 1970s, there has been a reconsideration of Reclus. Before, some books had been published looking mainly at his politics, but in 1981, *Hérodote* dedicated a whole issue to Reclus as a precursor of radical geography (a second theme issue on Reclus came out in 2005). His main work 'L'Homme et la Terre' was reissued in an edited version in 1982. Roger Brunet, also emerging from the Paris circle of communist geographers of the 1950s, started a research center called gip RECLUS with the aim of compiling a new *Géographie Universelle* in the spirit of Reclus (funding for the center ceased in 1997 and it was continued as an association).

Generally, however, the 1980s were marked by a distinct feeling of disillusionment among the left. In 1981, the socialist candidate Mitterrand was elected president. He appointed several members of the communist party as ministers. However, Mitterrand abandoned his social project and the communists left the government. The wider dissatisfaction and the effects of the breakdown of the Soviet Union and its socialist project may have contributed to a decline of critical, especially Marxist approaches (although of course very few post war critical geographers advocated a Soviet model of development).

Hérodote increasingly turned to straightforward geopolitical analyses. Yves Lacoste came to express his support for the idea of the nation, and Béatrice Giblin, co editor of *Hérodote* , tried to distance their project from the antinationalist Reclus.

The beginning of a fourth generation of critical geography can be located in the late 1990s. It was fuelled by international developments and the high relevance of critical approaches in the 'new' debates that shaped international geography. Georges Benko started the journal *Géographie Economie Société*, which has a strong international focus. It featured, for example, reevaluations of the regulation school, of territorial governance, contributions from Richard Peet, and discussions of Neil Smith's work on gentrification. More recently, the diffusion of spatial theory and the new gained importance of 'space' in critical theory was also evident in more general leftist journals. The journal *Actuel Marx*, dedicated to the critical discussion of Marxist theory with contributions mainly from philosophers, had contributions from geographers – notably several from David Harvey on the 'new imperialism'.

French critical geography was always highly international. In the formative years of the discipline, Reclus traveled widely, was an international anarchist activist, and published a geography of the world. The second wave of critical geography after World War II drew much political motivation from its involvement in the communist party and its critique of French colonialism Europe. Due to the long tradition, Marxist debates in geography were at a high theoretical level. This generation was powerful in the discipline and highly influential in Spanish and Italian geography, a trend that was continued in the third wave after 1970. In Italy, for example, an Italian version of *Hérodote* was circulated for some time. Connections with Quebec were also always of

Hérodote **janvier-mars 1976**

revue trimestrielle

1

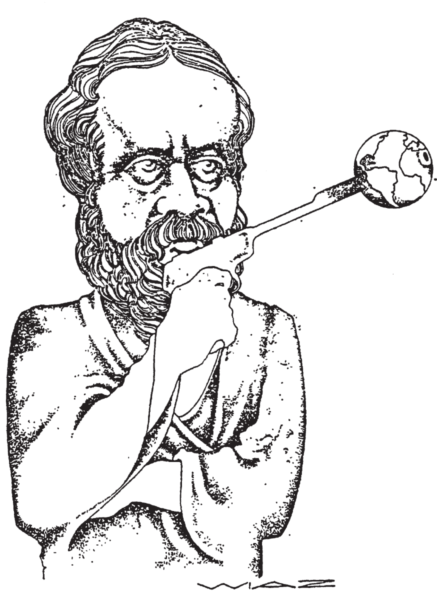

SOMMAIRE

PHOTO DE COUVERTURE. Rizières du Nord-Vietnam soumises à un bombardement en tapis. La couronne qui entoure chaque cratère est formée par des terrains du sous-sol qui recouvrent la terre arable.

Figure 2 Inside title of the first issue of Hérodote (1976), Hérodote and the globe-pistol. The table of contents show (among others) articles about the bombing of the Red River dams in Vietnam and an interview with Michel Foucault.

relevance in French critical geography. There was con siderable interaction with Spanish critical geography, which also started drawing on English language radical geography in the years of post Franco democratic tran sition. The current phase of critical geography in France also draws on the now established Anglophone inter national critical geography.

Critical Geography in Germany

Critical geography in the twentieth century often built on the work of two Germans, who could therefore be con sidered as important figures in critical geography, despite not being geographers: Karl Marx, who is often reported to have attended Carl Ritter's geography course at the uni versity of Berlin, and Friedrich Engels, whose analysis of the conditions of the working class in Manchester is often referenced as an early critical urban geography. Still, they were no geographers. The first explicitly critical geog rapher in Germany did write in a Marxist tradition. Karl August Wittfogel, an expert on China and a member of the Frankfurt school of critical social theory (and of the com munist party), in the 1920s developed an approach to geography based on a consideration of the society nature relationship the resource base of a society, and the control of these resources by the ruling class. The article was published first in Russian, then in German in a journal of the communist international. Writing after German geog raphy had been established as a mainly conservative dis cipline, Wittfogel posited himself as an explicit critic of the growing influence of right wing conservative German geopolitics, which he criticized as the ideology of German imperialism. Wittfogel was also the translator of the eco nomic geography of James Francis Horrabin, which he considered a good example of a materialist (yet not thor oughly Marxist) approach to geography. At around the same time, Sandor Rado, a Hungarian geographer and cartographer, also studied and worked in Germany. He was a communist by conviction and a dedicated cartographer who produced maps for newspapers as well as map guides. Rado was in close contact with the Soviet government: in 1925, he published a guide book to the Soviet Union. In 1930, he published the atlas *Der Imperialismus*, as the first volume of a projected series. The atlas pictured the de velopment of the colonial empires, global economic re lations and political constellations. In the preface, Theodor Bernstein, director of the Institute for the World Economy in Moscow, described the atlas as a project opposed to the 'new trend' of German geopolitics, showing the real basis of global inequalities and thus empowering the workers. Later, Rado fled to Paris and Switzerland, where he was also involved in an espionage ring working against Nazi Ger many. Rado and Wittfogel however had little influence on German geography, although Rado was an important geographer in Hungary after World War II (and after he came back from containment in the Soviet Union). Witt fogel emigrated to the US and later testified in the McCarthy hearing against Owen Lattimore, also an expert on China and critical academic.

With the rise of national socialism, these critical ap proaches in geography disappeared in Germany. After 1945, in West Germany, there was a fairly strong anti communist climate, and many earlier Nazi geographers continued teaching at universities. In Eastern Germany, Schmidt Renner developed an approach to economic geography that was based on historical materialism but that was only rarely a point of reference for West German geographers. Mostly, however, East German geography was applied geography that made contributions to the regional planning or economic organization of East Germany. In general, geography in East Germany has had very little effect on West German and post 1989 German geography.

In West Germany, the first new attempts at a critical geography developed in the late 1960s, in the wake of student radicalization. Geography students were involved in the struggles for a reform of the universities toward more democratic institutions. They also saw the devel opment of critical approaches in the social sciences and tried to apply them in geography. A number of short lived journals resulted from this involvement. The jour nal *Geografiker* (1968–1972) was a publication by the geography student association of the Free University of Berlin. It served to document their efforts at university and curriculum reform and exchange their suggestions with other student associations in West Germany. Many of the suggestions were calls for a more 'useful' know ledge. In its later issues, *Geografiker* also included articles that criticized not only the form of geography, but also the theories of traditional German geography (the *Länderkunde* model). *Länderkunde* was denounced as an ideology, but *Geografiker* did not offer a clear alternative model for geography, although it contained some articles of a broadly Marxist approach as well as those in the spirit of positivist 'science'. The Geographentag con vention of 1969 (and the following convention in 1971) featured a session where some students voiced a more radical, Marxist critique of geography. The main trend of the revolution in geography, however, was quantification. The radical critique stayed marginalized. Students setting up stalls at geography conventions displaying radical journals were in a number of cases evicted by the police. Another student journal, *Roter Globus* (red globe) was published in Marburg from 1971 to 1973 (**Figure 3**). The contributions were more radical critiques of traditional geography and the authors were generally located in the discussion of the radical left (and communist) groups of the time. It attempted a Marxist foundation of geography, but was not very much influenced by the international

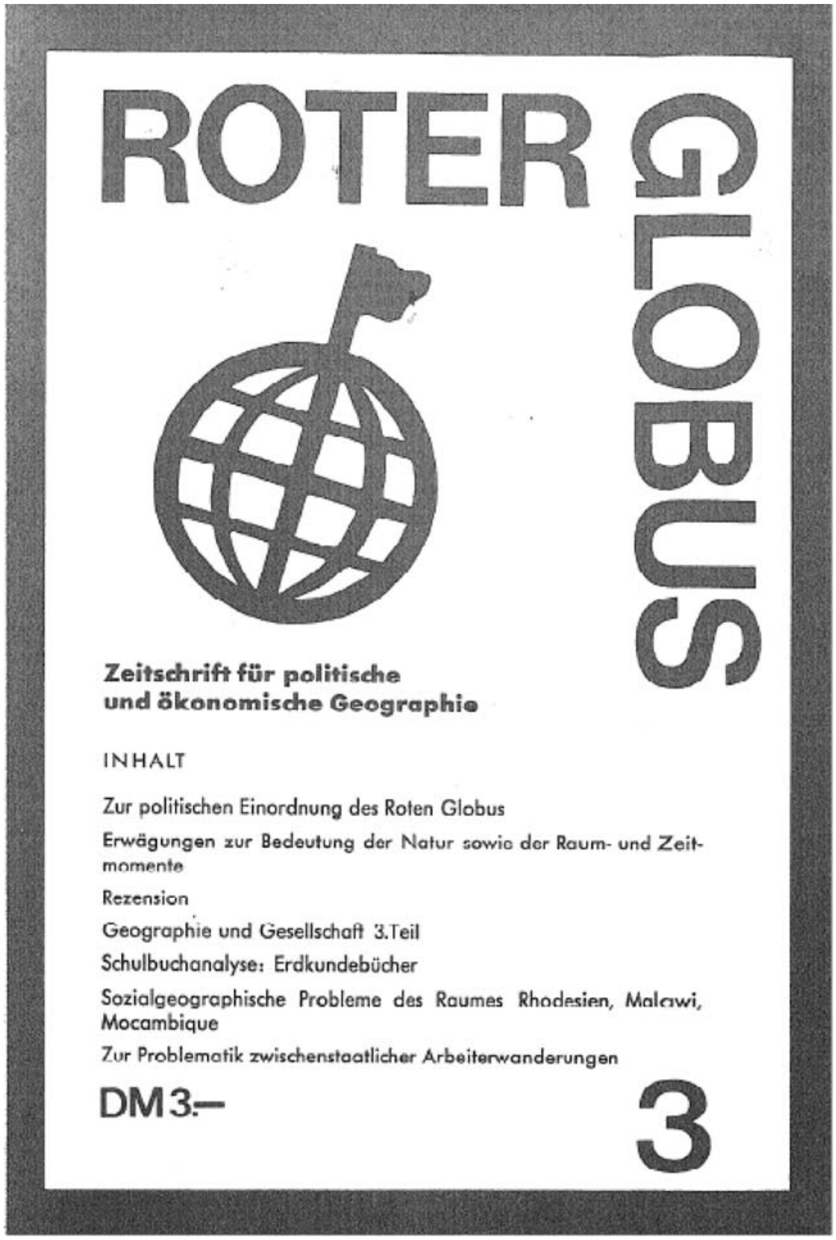

ROTER GLOBUS

Zeitschrift für politische und ökonomische Geographie

INHALT

Zur politischen Einordnung des Roten Globus

Erwägungen zur Bedeutung der Natur sowie der Raum- und Zeitmomente

Rezension

Geographie und Gesellschaft 3.Teil

Schulbuchanalyse: Erdkundebücher

Sozialgeographische Probleme des Raumes Rhodesien, Malawi, Mocambique

Zur Problematik zwischenstaatlicher Arbeiterwanderungen

DM 3.–

3

Figure 3 Title page of Roter Globus (issue 3, 1972). Here, the subtitle of the journal was still *Journal for Political and Economic Geography.* With issue 4, this was changed to *Journal for the Critique of Bourgeois Geography.* This issue contained (among others) articles on the political position of the journal, nature society relations, a school-book analysis, the histroy of geography, and migration. Reproduced with permission from Wolfgang Plucinski.

developments in French and English language critical geography, which were at a considerably higher level. One reason was that these were mainly student journals with a very small discussion base. A third short lived journal, *Geographie in Ausbildung und Planung* ('Geography in teaching and planning', 1973–1976) even entered an exchange with *Antipode.* The Marxist margin of geography in Germany in the 1970s was also manifest in a book series (*Geographische Hochschulmanuskripte*, 1973–1985) which featured contributions in critical geography, among others the translation of a Marxist geography course book by the Danish geographer Buch Hansen. Parallel developments also took place in Switzerland and Austria. In Austria, a critical geography association was established in 1982, which has established several book series, and in Switzerland, a group of critical geographers was established in the 1980s.

In the late 1980s, these attempts at a Marxist geography almost disappeared. English language radical geography was ignored by the German mainstream. A new critique emerged in feminist geography, which was highly international in Germany almost from the outset and even founded a subgroup of the German geographical association. Further, the critique of the history of the discipline gained new momentum, primarily targeting the role of geopolitics, geography, and geographers in national socialism. Student groups were again important in this critical analysis of the discipline. It was connected to a new interest in the critique of geopolitics on the international scale.

The 1990s saw a continuation of these new trends in critical geography. Since the late 1990s, due to the increased international role of critical geography and the growing internationalization of German geography, critical geography has gained ground in Germany. The new relevance of space in the critical social sciences was a further factor that led to increased possibilities for critical geography in Germany. The Hettner lectures (1997–2006), for example, a seminar series in Heidelberg sponsored by a software billionaire and in itself by no means 'radical', were an evidence of this changed role of international critical geography: many of the invited speakers were influential critical geographers, from Doreen Massey to Michael Watts. The seminars also sparked off a number of meetings by feminist students. Geographers were involved in the critical association for urban research and activism INURA. Since 2001, there have also been *Antipode* sponsored lectures at the German geography conventions, the most recent ones looking at the changing conditions for critical geography in Germany and in international comparison. Student groups have also organized further lecture series on critical geography. The interdisciplinary journal *Anarchitektur* (since 2002) is dedicated to a critique of planning, architecture, and spatial politics. The new journal *Geographische Revue* published an issue dedicated to Marxism in geography, and German geographers are actively involved in the international critical geography conferences.

Critical geography in Germany presents a case of a relatively isolated development with no uninterrupted tradition and a history of marginalization and exclusion from the main discipline. No tradition was established until the 1970s when in student circles critical geography was relatively strong, but excluded from the curriculum. Only very recently, with the growing internationalization of German geography and the weight of critical geography internationally, some space has been established for critical approaches.

Critical Geography in North America and the United Kingdom

Anglophone geography has become the branch of critical geography with the strongest presence globally at the moment, but these approaches set in a little later than in the other two cases portrayed here. British and North American geography – and particularly critical geography – were very connected, with a high exchange not only of ideas but also people. As pointed out above, *An Outline of Economic Geography* by the later MP James Francis Horrabin, published by the Plebs league for working class education, was written as a critical materialist approach and can probably be considered the first such attempt in the English language. It was translated into German in 1926 by Wittfogel and into Russian as an 'economic geography from a Marxist perspective' in 1928 (further translations into other languages followed). Horrabin illustrated numerous books by left wing authors and provided the maps. He also published an atlas of the British Empire, and Rado's abovementioned atlas of imperialism was published in the UK in 1938 with a much extended text section but without the political introduction. These were isolated attempts – there did not emerge a critical tradition in geography until the 1960s. Apart from individual radicals like Keith Buchanan, an early leftist critic of British traditional geography, it was only in the course of the opposition against the Vietnam War, black power, and student movements, as well as the paradigm shifts of the 1960s, that geographers started turning to critical approaches. In the USA, the idea of a 'social geography' that recognized problems like poverty and racism was an early focus for critical geography. William Bunge's efforts to use geography against the exclusion of black and poor people were a radical version of this social geography. The foundation of the journal *Antipode* in 1969, created after a radical geography caucus, signaled the beginning of a broader radical movement (**Figure 4**). The term 'radical geography' came into use, along with the terms 'critical geography' and the older 'social geography', all aiming at 'socially relevant research'. The debate on relevance was not limited to radical circles – it could either mean relevant for the application by the state, or relevant for the oppressed in concrete struggles. *Antipode* was fairly clearly on the side of 'radical' relevance. *Antipode* printed, for example, a syllabus of a course on poverty and an issue on radical research methods like participatory action research. Richard Peet wrote in an editorial in 1970: "We want to print alternatives – the people's geography, survival geography, humanistic (lovin') geography, radical geography." Early issues were also illustrated with cartoons commenting on the articles. David Harvey's article on revolutionary and counter revolutionary theory (see above) sparked a debate on the role of theory in social struggles and the relation between theory and practice. Other articles analyzed the institutional structures of 'establishment geography' or introduced feminist/gender issues. A further focus was to contribute to an anarchist geography, building on the legacy of Reclus and

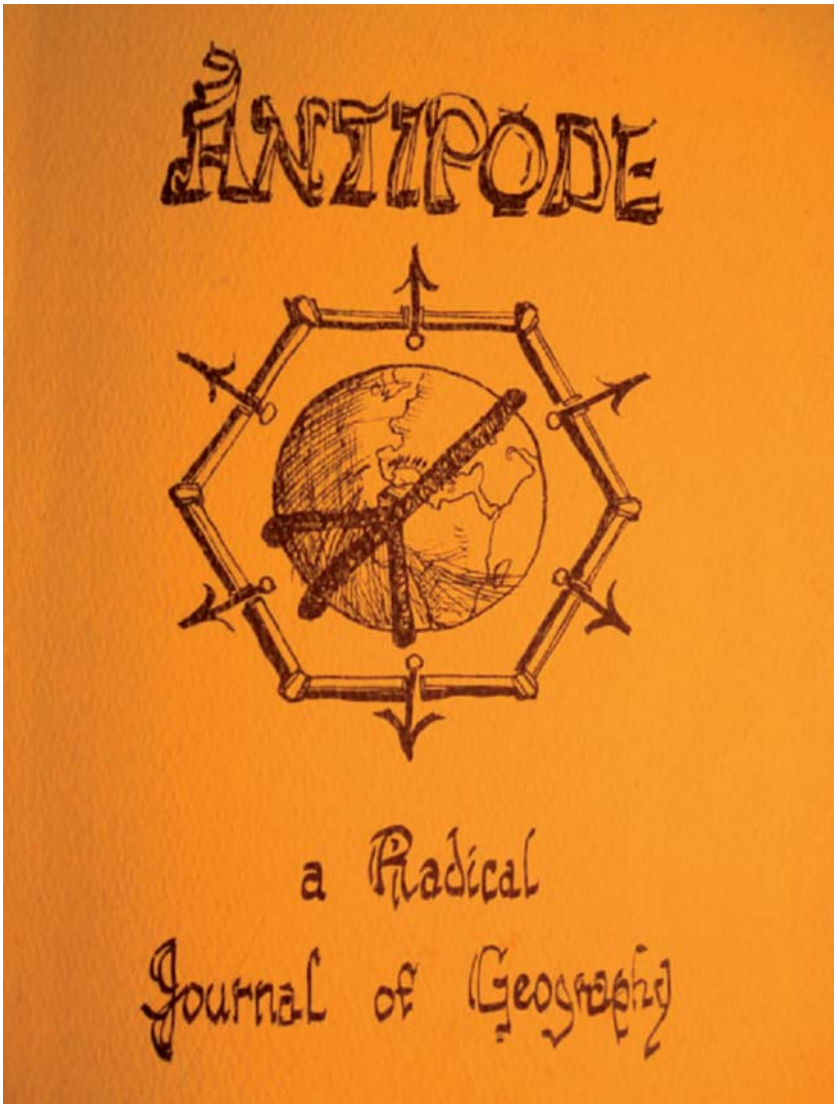

Figure 4 Cover of Antipode 1, 1969. Reproduced with permission of Blackwell publishers.

Kropotkin. *Antipode* was also relatively international in the composition of its authors – it featured regular contributions by British geographers, Scandinavian geographers, as well as translations from the French journal *Espaces et Sociétés*. Milton Santos, in an *Antipode* article from 1974, also used French geography as an ex ample of a developed Marxist and activist approach.

In terms of organization, North American geog raphers founded the Union of Socialist Geographers (USG, Toronto 1974) which also had a British wing and lasted into the 1980s. Later, the Socialist Geography Specialty Group of the (Association of American Geographers) was established (now Socialist and Critical Geography Specialty Group). The radical movement in North American geography was not uncontested, however. There were strong institutional resistances, and the establishment of *Antipode* and of the USG has to be understood as a reaction to these attempts of exclusion.

A number of developments coincided in the 1980s. First, there was a differentiation of Marxist approaches that came along with the increasing specialist knowledge of Marxist theory within geography, so that different positions could emerge.

Second, and connected to this, an increased attention was dedicated to the state. New approaches drew on Wallerstein's or Althusser's reformulations of Marxism. In the late 1980s, critical geopolitics developed a special perspective on the state and geopolitics. English language critical geography thus followed the path that French critical geography had taken a few years earlier, when *Hérodote* turned to an analysis of the state (although *Hérodote's* approach was less Marxist inspired).

Third, postmodern approaches became important, and the relation between postmodern and Marxist approaches was discussed at great detail. The journal *Society and Space* (*Environment and Planning D*, launched in 1983) included approaches that broadly fit in with the established critical tradition, but was later also at the forefront of a turn to Foucault and postmodern/post structuralist approaches.

Fourth, starting in the 1980s, 'critical geography' became a label to encompass a diversity of approaches – and also a selling point. A number of book series have since used the label, starting with the Macmillan book series *Critical Human Geography* in 1982, to the series *California Studies in Critical Human Geography* (since 1997) and the Routledge series *Critical Geographies* (since 1999), which are very broad in their scope, from disability and bodies to nature and racism.

A further new focus of the 1980s was the rise of new cultural geography. Its radical branches included a renewed critique of geography and the analysis of the relevance of place, identity constructions and discourses in social struggles.

In the 1990s, these trends toward a more diverse critical geography with a number of established approaches continued – including feminism, varieties of Marxism, Foucaultian analyses, and post structuralist perspectives. The late 1980s had also witnessed an increasing attention to 'space' in the wider social sciences, and this trend also continued in the 1990s. The 1990s saw the institutionalization of feminist and gender approaches in geography (like in the journal *Gender, Place and Culture*, since 1994). Queer theory, however, entered geography only in the 2000s.

Institutionally, in the UK there emerged another point of contention. Since 1933, there were three main geographical societies: the Geographical Association, focused on geography in schools, the Royal Geographical Society (RGS), which was at that time strongly dedicated to exploration and expeditions, and the Institute of British Geographers (IBG), which had initially been founded by academic geographers who did not find themselves represented in the RGS. When a merger with the RGS was discussed in the IBG in 1994, a majority voted for it, but a minority were unhappy with the decision, due to the more traditional outlook of the RGS and its colonialist past. The critique intensified in the debate about Shell as a sponsor of the RGS and its role in the violence against the Ogoni people in Nigeria and the execution of Ken Saro Wiwa and other Ogoni activists. The RGS was called on to renounce the sponsorship from Shell. This conflict was one of the key issues in the foundation of the critical geography forum, initially an association of critical geographers inside (and outside) the merged RGS/IBG and now a mailing list.

After strong contention in the discipline in the 1970s, critical or formerly critical geographers have reached the highest positions of academia in the UK in the 1990s, to a greater extent than in North America and much more than in continental Europe. Since the 1980s, critical geography papers are regularly being accepted for publication in traditional journals, and explicitly critical journals have reached top positions on citation indices. This has also led to some conflicts, when individuals and their schools are in a position to strongly influence the development of not only critical geography, but also of the whole discipline, and act as new gatekeepers. In this situation, critical geography has been criticized as a fashionable label, as uncritical critical geography or neocritical geography which has established its internal power structures, its rules of writing and theorizing, and has become a selling point in the academic labor market.

International Critical Geography

The international critical geography movement was built on different regional or language based associations. There had been a Scandiavian (Nordic) critical geography conference in 1996; there had been critical geography meetings at the British and US geography conventions; and there had been numerous conferences organized on a smaller regional scale. In 1996, the idea of an international critical geography conference (ICGC) was raised, as an alternative to the national association conventions. The conference took place in Vancouver, Canada in August 1997 and attracted about 300 geographers from 30 countries. It resulted in the formation of the International Critical Geography Group and has been followed by a number of regional conferences (like the East Asian Regional Conferences in Alternative Geography/EARCAG). In 2000, the second ICGC took place in Taegu, South Korea. The third conference (2002 in Békéscsaba, Hungary), the fourth (2005 in Mexico City), and the fifth (2007 in Mumbai) continued the series. These conferences contributed to a more international discussion, but also allowed some tensions in international critical geography to come to the surface.

One top issue was the question of language. While in many ways a pragmatic issue, at most sessions, English was the language that was spoken. This obviously privileged native speakers of English over others that needed to translate their ideas into English. It also meant that concepts were generally understood to be used as in the English language literature. Those from outside the UK, North America, and the 'Antipodes' were supposed to know the concepts used in these countries, in order to be able to communicate their perspectives. To tackle this problem, pragmatically, translation, and conceptually, the relativity of the own position became a key issue at the conferences.

The issue is closely connected to the second issue that became the 'Anglo hegemony' debate. It was – and is – centered on the inequalities of international critical geography. The 'geography of critical geography' is a scalar one, in which, as Simonsen argues, the 'Anglos' dispose of knowledge "'unlimited', 'universal' or at least 'transferable', while knowledge produced in almost all other places through the 'case study from another place' syndrome is constructed as 'limited', 'local' and deeply 'embedded' in its local/national context." Vaiou offers a different perspective: in Greece, 'other' critical geog raphies (in this case, French critical theories and later English language critical geography) have been an im portant factor to contribute to a critical social science. It was the international connection of Greek academics that furthered critical approaches. In Germany, the internationalization of German geography – and the streamlining according to international models like cita tion indices and 'high impact journals' – are currently privileging the import of critical geography.

This links in with the third issue that has recently become important in international critical geography: the different working conditions for critical geographers in different national contexts. In the UK, a critique of as sessment procedures has been voiced (with an inter national focus). Harvey went so far to suggest that due to the British assessment system, few good books were being written in the UK anymore. In the US, the working conditions as well as the streamlining of academic geographers in the course of their professional training have been criticized.

The topic of working conditions also concerns pub lications. As a reaction to the increasing corporate role in the publishing business, the high prices of journals (and the profit logic that drives their publication) have come under scrutiny. One reaction has been the establishment of the free online critical geography journal *Acme* in 2002, which has also tried to become more international in its scope than the traditional English language journals. Another reaction was the boycott of the publisher Reed Elsevier when it emerged that it was also organizing arms fairs. The boycott was started in the medical sciences but taken up in geography, and it led to the withdrawal of Reed Elsevier from the arms business.

Fourth, activism beyond academia has not been a key topic in the international critical geography movement, although there have been efforts to include perspectives on activism and activists at the critical geography con ferences. This highlights a deficit: it has not been easy to connect different understandings of academic work and activism. The relationship of activism and academia is determined by the political systems of the respective countries, by the construction of intellectual and political work, and by the specific struggles taking place. To tackle this issue in theory and practice will be a task for the future of international critical geography.

See also: Critical Theory (After Habermas); Economies, Alternative; Feminism/Feminist Geography; Marxism/ Marxist Geography I; Marxism/Marxist Geography II; People's Geography; Place, Politics of; Postcolonialism/ Postcolonial Geographies; Radical Geography; Resistance; Socialism.

Further Reading

Basssda (2006). A kind of queer geography/Raume Durchqueeren: The Doreen Massey reading weekends. In: gender, place and culture. *A Journal of Feminist Geography* 13(2), 173 186.

Bataillon, C. (2006). Six géographes en quête d'engagement: Du communisme à l'aménagement du territoire. Essai sur une génération. Cybergeo : Revue européenne de géographie, No. 341, 27 June 2006. http://www.cybergeo.eu/docannexe/file/1739/341.pdf.

Berg, L. D. (2004). Scaling knowledge: Towards a critical geography of critical geographies. *Geoforum* 35, 553 558.

Blomley, N. (2006). Uncritical critical geography? *Progress in Human Geography* 30, 87 94.

Blunt, A. and Wills, J. (2000). *Dissident Geographies: An Introduction to Radical Ideas and Practice*. London: Prentice Hall.

Cox, K. R. (ed.) (2005). *Political Geography. Critical Concepts in the Social Sciences*. London: Routledge.

Fahlbusch, M., Rossler, M. and Siegrist, D. (1989). Conservatism, ideology and geography in Germany 1920 1950. *Political Geography Quarterly* 8, 353 367.

Folke, S. (1985). The development of radical geography in Scandinavia. *Antipode* 17(2 3), 13 18.

Fuller, D. and Kitchin, R. (eds.) (2004). *Radical Theory/Critical Praxis. Making a Difference Beyond the Academy?* Okanagan: Praxis e press.

Graham, S. (2005). Remember Fallujah: Demonising place, constructing atrocity. *Environment and Planning D: Society and Space* 23, 1 10.

Gregory, D. (1978). *Ideology, Social Science, and Human Geography*. London: Hutchinson.

Harvey, D. (1973). *Social Justice and the City*. London: Arnold.

Harvey, D. (2000). *Spaces of Hope*. Edinburgh: Edinburgh University Press.

Harvey, D. (2006). The geographies of critical geography. *Transactions of the Institute of British Geographers* 31(4), 409 412.

Mizuoka, F., Mizuuchi, T., Hisatake, T., Tsutsumi, K. and Fujita, T. (2005). The critical heritage of Japanese geography: Its tortured trajectory for eight decades. *Environment and Planning D: Society and Space* 23(3), 453 473.

Oßenbrugge, J. (1983). Recent developments in political geography in West Germany: An ahistorical reflection of the past, an apolitical analysis of the present, and possibilities for the future. *Political Geography Quarterly* 2, 71 80.

Peet, R. (ed.) (1977). *Radical Geography. Alternative Viewpoints on Contemporary Social Issues*. Chicago: Maaroufa Press.

Reclus, E. (1982). *L'homme et laTterre*. Paris: La Découverte. First published in 1905 1908 in six volumes (reissue under the direction of B Giblin).

Santos, M. (1974). Geography, marxism and underdevelopment. *Antipode* 6(3), 1 9.

Sharp, J., Routledge, P., Philo, C. and Paddison, R. (2000). *The Entanglements of Power: Geographies of Domination/Resistance*. Routledge: London.

Simonsen, K. (2004). Differential spaces of critical geography. *Geoforum* 35, 525 528.

Smith, N. (2005). Neo critical geography, or, the flat pluralist world of business class. *Antipode* 37, 887 897.
Vaiou, D. (2004). The contested and negotiated dominance of Anglophone geography in Greece. *Geoforum* 35, 529 531.
Wittfogel, K. A. (1985). Geopolitics, geographical materialism and marxism. *Antipode* 17(1), 21 71. First published in 1929.

Relevant Websites

http://www.acme journal.org
ACME: An international E journal for critical geographies.
http://econgeog.misc.hit u.ac.jp
The International Critical Geography Group.

Critical Geopolitics

J. P. Sharp, University of Glasgow, Glasgow, UK

Glossary

Discourse A discourse can be thought of as a temporally and spatially specific lens through which the world is interpreted, thus defining the boundaries of what can be known.

Geopolitics Geopolitics is an aspect of the practice and analysis of statecraft which considers geography and spatial relations to play a significant role in the constitution of international politics.

Realism A theory of international relations which regards the state to be the only genuine source of agency in the international realm.

Introduction

The term 'geopolitics' was first used by the Swedish political scientist Rudolf Kjellen in 1899, but did not become popular until used in the early twentieth century by British geographer and strategist Halford Mackinder. Mackinder promoted the study of geography as an 'aid to statecraft', insisting that geographers had a key role in the practices of international relations. The study of geo politics focused upon the ways in which geographical factors – such as proximity, land mass, geographies of resources, and so on – shaped the historical unfolding of international politics. He argued that certain spaces would be more vulnerable to attack because of the effects of distance on politics and that certain topographical features could promote security. In geopolitical argu ments, the effect of geography on politics is based upon 'commonsense', rather than ideology: the 'facts' of geography are seen to have predictable influences upon political processes.

The concept of security is fundamental to the study of geopolitics. This refers to the protection of the state from the threat of external powers. Geopoliticians argue that they can aid national security by analyzing the effects of a country's geography, and that of potential conquerors on future power–political relations. Geopolitics claims to be able to predict which areas could strengthen a state, helping it to rise to prominence, and which might leave it vulnerable. As a result, geopolitics has traditionally been considered to be a very practical and objective study: the actual 'practice' of international relations has been seen to be quite separate from political theory.

Critical geopolitics refers to a range of approaches that arose through a revival of interest in political geography to challenge dominant geopolitical arguments in the 1980s. This involved a number of intellectual positions from European and Anglo American geography which arose to challenge the state centrism and factual accuracy of conventional geopolitical theories. For in stance, critical analysis of American exceptionalism and Cold War theories of the domino effect revealed the inherently ideological nature of the use of these geo graphical arguments. Such critical perspectives came from intellectual strands of Marxism and world systems approaches but the conceptual approach that has come to be most closely associated with the term 'critical geo politics' has its origins in post structuralism.

Post-Structural Critical Geopolitics

Founded initially by political geographers Simon Dalby and Gearoid Ó Tuathail (writing together and separately) in the late 1980s, critical geopolitics influenced by post structuralist concerns with the politics of representation considers the ways in which the use of particular dis courses shape political practices. Using the apparently oxymoronic term 'critical geopolitics', Dalby and Ó Tuathail sought to create a critical approach to subvert the taken for granted meanings of geographical dis course in explanations of international politics. The rise of critical geopolitics in political geography can be seen in a wider context. Post structural approaches in inter national relations theory deconstructed the foundational concepts upon which the discipline had been built, in cluding realism and sovereignty, in a way mirroring the theoretical agenda of critical geopolitics. However, the rise of both critical geopolitics and post structural ver sions of international relations theory have created a very geographically specific tradition with few practitioners beyond North America and the UK (a notable exception would be Fin Anssi Paasi).

Critical geopolitical approaches seek to examine how it is that international politics is imagined spatially or geographically and in so doing to uncover the politics involved in writing the geography of global space. Rather than arguing over the true effects of geography on international relations critical geopolitics asks whose models of international geography are used, and whose interests these models serve. This approach owes much to Michel Foucault's insistence that power and knowledge are inseparable. For geopoliticians, there is great power available to those whose maps and explanations of world politics are accepted as accurate because of the influence that these have on the way the world and its workings are understood, and therefore the effects that this has on future political practice.

Dalby and Ó Tuathail likened critical geopolitics to writing over the accepted political maps of the world to expose what these omitted and to highlight the power/knowledge relations upon which they depended. Unlike conventional geopolitics, for critical geopolitics geography is not a collection of incontrovertible facts and relationships 'out there' in the world awaiting description but is a discourse. Geographical orders are created by key individuals and institutions and then imposed upon the world. Ó Tuathail has called this process 'geo graphing' – earth writing – to emphasize the creativity inherent in the process of using geographical reasoning in the practical service of power. Geographical discourse used in representations of international politics then is the product of cultural context and political motivation. Those adopting critical geopolitical approaches grant a range of power to discourse, from those who see geopolitical discourse becoming out of synch with the geopolitical reality it seeks to describe and so causing inappropriate state practice, to a figure like French philosopher Jean Baudrillard for whom language and representation are everything (he famously illustrated this point by arguing that the Gulf War did not exist, it only occurred on television).

Conventional geopolitics draws on a discourse of objectivity. In part, this comes from the use of the apparently unchanging facts and laws of geography. However, the idea of geopolitics as an objective form of knowledge (and therefore an unbiased form of political practice) also emerges through the privileging of sight (especially with the use of maps and diagrams) over other senses in geopolitical reasoning. This allows the geopoliticians to write as if from afar, as if somehow unconnected to the world being surveyed reinforcing the idea of an objective account rather than one written from a position embedded within the events being discussed. It hides the fact that the geopolitician has his or her own point of view and loyalties. Although it is generally accepted that Nazi *Geopolitik* had a political agenda, this is considered to be an aberration of the 'science' of geopolitics. Yet the Foucauldian approach to discourse makes it clear that no geopolitician can be innocent of interest. For example, naval historian Mahan's geopolitics were influenced by his desire to build up the US Navy at a time when other technologies seemed to make naval power less important while George W. Bush's geopolitical construction of an 'axis of evil' stretching from Iraq to North Korea provides Americans with a named and territorialized enemy 'out there' in its 'war on terror'.

In addition to recognizing the situatedness of geopolitics, critical geopolitical approaches have focused on ways in which geopolitical arguments have sought to create emotional response on the part of its subjects encouraging them not only to believe in certain representations of the world, but also, in some cases, to create active citizens remaking their world.

Critical Geopolitics and Identity

Conventional geopolitics is based around realist political thought. For realism, world politics is made up of an anarchic, unregulated space of the international and the ordered space of the sovereign state. States are understood to be the only (or main, in the case of neorealism) actor in international politics. The main role of the state is to provide security, protecting the domestic space and its citizens from the threat of the chaotic international. Realism is based around a binary of inside (domestic, state) and outside (chaotic, international). By privileging one side, a hierarchy is formed which naturalizes the unequal relationship between self and some defined other. For realism, ordered international space represents a peaceful system, whereas anarchy signifies global political space which has yet to be domesticated by a sovereign state. This discursive structure is necessary to the maintenance of the naturalization of the realist position as the explanation of international politics. Thus, legitimacy is provided to the states wishing to extend the order of their sovereignty to this chaotic space.

Critical geopolitical analysis of security discourse demonstrates that it is difference that threatens the states so that the essential moment of geopolitics is the division of space into 'our' place and 'their' place; its political function being to incorporate and regulate 'us' or 'the same' by distinguishing 'us' from 'them', 'the same' from 'the other'. In arguing this, critical geopolitics suggests that geopolitics is not something simply linked to describing or predicting the shape of international politics, but is central to the ways in which identity is formed and maintained in modern societies. National identity is not simply defined by what binds the members of the nation together but also – perhaps even more importantly – by defining those who exist outside as different from members of the nation. Drawing borders around territory to produce 'us' and 'them' of the nation and those who are different, does not simply reflect the divisions inherent in the world but also helps to create differences. While some critical geopolitics has sought to challenge the state centrism of conventional accounts by highlighting the role of actors 'above' and 'below' the state, and those actors and issues which transcend the boundaries of states (such as environmental security), it has also sought to highlight this role of realism and the exclusion of difference in the construction of identity.

It has been argued that the first instance of geopolitics was Orientalism, the imaginative geography of the West versus the rest so eloquently elucidated by Edward Said. Drawing upon the Western philosophical tradition of ordering the world through conceptual binaries, Said argued that Europe's representation of the Orient as other, was used to reflect back a positive image of the self. The West is everything that the rest is not. In other words, he is suggesting that the concept of difference between East and West is a geopolitical difference which is inscribed throughout the texts of Western culture whether travel writings, political texts, paintings, or in academic discussions. These texts 'preceded' experience, so empirical evidence was included but fitted into the categories that were already constructed. Travelers saw what they expected to see. For Said, this is particularly important because of the link between this imaginative geography and European power. This imaginative geog raphy was made manifest over space as it was built into colonial policy, into the institutions of governance, and, more recently, into practices of aid and development. The imaginative geography of Orientalism shaped the real geographies practiced in the space of the Orient.

While it is easy to see the continuation of Orientalism in contemporary geopolitics of the 'war on terror' (and the identities this has inscribed on Americans and their others), critical geopolitics argues that we can see this logic playing out between other states too. For example, when the USSR was constructed as being completely unlike the USA, any description of the USSR as evil, aggressive, and unreasonable implies goodness, tolerance, and reason on the part of Americans. Again, geopolitics does not simply reflect the facts of geography but also in dividing the world into a state and the international realm helps to form geographical orders and geo graphical relationships.

The construction of 'otherness', and particularly the sense of danger that this presents, have implications for the practice of domestic affairs in addition to foreign policy. Critical geopolitical analysis has shown how the presence of external threats has been used within geo political discourse to disciplining domestic dissent. The construction of otherness simultaneously presents a normative image of identity to be imposed on national citizens.

The Critical Geopolitics of the Cold War

A good deal of critical geopolitical analysis has been turned to Cold War geopolitical discourse. Here it is very clear how geopolitical discourse was presented as factual and objective, but that it tied clearly into developing a normative sense of US identity through the creation of the USSR as entirely different and other to America.

One of the formative documents of the Cold War was sent as a telegram from Moscow, by George Kennan – 'Mr X' – a US official in the Soviet Union at the end of World War II. Kennan argued that the Soviet Union was absolutely different from the USA so there could be no possibility of compromise between the two. This image of two distinct and incompatible territorial blocks was re inforced by political discourse whether Churchill's image of an Iron Curtain dividing Europe, or Ronald Reagan's depiction of the Soviet Union as an 'evil empire'. A number of interrelated geopolitical concepts reiterated this binary geography in political discourse. These were, most importantly, containment, domino effects, and dis ease metaphors.

Containment, first outlined by Kennan, referred to the military and economic sequestration of the Soviet Union. Russia's historical geography, and not simply its political and cultural difference, was invoked to give this argu ment scientific respectability: the USSR was seen as an inherently expansionary force which had to be kept in check.

The inevitability of Soviet expansion was also ex pressed in metaphors of dominoes or disease. Such metaphors saw the spread of communism or socialism as a result of proximity to territory ruled by Soviets. The domino theory assumed that Soviets, communists, and socialists everywhere were unqualifiedly evil. An Ameri can nuclear strike on Vietnam was essential in order to halt a Viet Cong victory which would set off a chain reaction of countries falling to the communists, like a row of falling dominoes. The domino effect can actually be seen to underlie the Vietnam war more generally. The argument insisted on the necessity for the United States to fight and win in Vietnam, for if South Vietnam fell to the communists and turned communist, then this would set in motion the tumbling of other states in the region – like falling dominoes; as Cambodia, Laos, Thailand, and Burma would turn communist as well. Let unchecked, this process would continue to spread around the world, and would not stop until it reached the last standing domino, the USA. This discourse made future political action appear inevitable, unless proactive action – such as con tainment or preemptive strike – were enacted here and now.

The domino metaphor embodied a zero sum game, a political system within which only two powers existed (the USSR and the USA), where only force could oppose force and where the unfolding of the process was in evitable – once started, the continuing fall of states was as unavoidable as stopping a line of dominoes from toppling once the first had been pushed. Disease metaphors were structurally very similar, relying upon notions of con tagion or the malign spread of infection, again depending upon a simple notion of geographical proximity as the basis for social and political change. Even more so than with dominoes, disease metaphors illustrated the neces sity for immediate action in order to prevent the further spread of the malady. Both were based on 'natural' pro cesses which implied that there was no politics or ideology involved only a 'commonsense' decision that action must be taken to stop them. The complexities of regime change were reduced to a simplistic mapping of location – proximity to a communist state made another vulnerable regardless of the political, social, and cultural characteristics of that state.

Although apparently a geographical theory, the domino theory eliminated any geographical analysis. Rather than being concerned with understanding geo graphical process, geopolitics reduces spaces and places to concepts or ideology. Space is reduced to units which singularly display evidence of the characteristics which are used to define the spaces in the first place (Asia 'is' exoticism, the USSR 'is' communism, Iran 'is' funda mentalism, the USA 'is' freedom and democracy, and so on). Geopolitical discourse is clearly linked to identity formation. All those on the inside have particular char acter traits (or should have, in this normative vision) which are completely at odds with those of outside. Any description of the other as negative would automatically reflect back a positive image onto the self. Every de scription of the Soviets as totalitarian or repressive, im plied the opposite characteristics in the binary opposite space of the USA.

Popular Critical Geopolitics

Like conventional geopolitics, many early accounts of critical geopolitics maintained a clear distinction be tween high and low culture, seeing formal, theoretical, and practical forms of geopolitics lying with the 'intel lectuals of statecraft'. There is a sense that international relations are beyond the realm of ordinary people, and beyond the effects of the politics of everyday life. Ó Tuathail insists that it is important to maintain a certain specificity to the term 'geopolitics' to keep it as a term closely associated with the writing of global space by those involved in international relations. However, it has been argued that this may overlook important practices of international relations which are not located within the formal practices of statecraft. Simply to describe a foreign policy is to engage in geopolitics and so normalize par ticular world views. Any statement concerning inter national relations involves an implicit understanding of geographical relationships or a world view.

It is unrealistic to assume that the intellectuals of statecraft are somehow beyond or outside of hegemonic national culture, or that their pronouncements are somehow unaffected by the circulation of ideas and be liefs therein. Key social believes are not simply imposed from above but emerge through hegemonic assemblages (government but also education and media) which may not all present a seamless image, and through which 'intellectuals of statecraft' themselves are socialized into the world. Moreover, in order to make their arguments believable to their audiences (in many cases, ordinary people, and not just other intellectuals of statecraft), they must refer to concepts and values that have consonance for the population at large, if their support is to be as sured. The media – both high and low culture alike – is intimately bound up with geo graphing the world, as are a range of activities normally described as occurring outwith the sphere of international politics. There are numerous examples of the permeability of boundaries between media geopolitics (factual and fictional) and formal geopolitical practice, from the role of CNN as favored US narrator of the two most recent Gulf Wars to President Reagan's enthusiastic suggestion after seeing *Rambo: First Blood Part II* (1985) that the next time American hostages were taken in the Middle East he would know what to do about it, and recent reports that the American government intelligence specialists have held secret meetings with Hollywood filmmakers and writers to map out possible future scenarios of terrorist threat.

As a result of the influence of cultural context, dif ferent countries' geopolitical traditions draw upon spe cific metaphors to create images of international geography. Political elites must use stories and images that are central to their citizens' daily lives and experi ences. By reducing complex processes to simple images with which their audiences would be familiar, geopolit icians could render political decisions to be quite natural, or could make the result of the process appear pre determined (as the domino example has demonstrated). For example, sport metaphors have been particularly prominent in the USA. Such language points to the 'es sential' differences between national potentials for world class performance and naturalizes a global arena in which the rules of the game are understood, and within which there are clear (unequivocal) winners and losers. The political difficulties and violence of conflict are reduced to technicalities in game play.

Embodying Geopolitics

Because of its relentlessly critical nature, the work of critical geopolitics is never done. There is a need to constantly be critical of any statement regarding the geography of world politics. As any statement must al ways be situated and partial, critical geopolitics can al ways seek to reveal the interests lodged within it. The post structural variety of critical geopolitics has been critiqued for this relentlessly negative critical attitude, from those who regard this lack of ontological commit ment as a limitation of the approach. There is critique but no positive alternative. The dominant form of textual analysis has also faced critique for leaving critical geo politics detached from the real practice of politics, and the visceral implications of these practices on peoples around the world.

Feminist political geographers, while sharing critical geopolitics' critique of dominant representations of international politics, have nevertheless argued that the textual focus of critical geopolitics restricts the nature of what alternative forms of politics can be imagined, and instead insisting on a form of geopolitics that is attentive to the embodiment of the people through whose lives it is articulated. This perspective insists that critical geo politics thinks of bodies as sites of performance rather than narrowly defined as simple surfaces for discursive inscription. Discourses do not simply write themselves directly onto the surface of bodies as if these bodies offered blank surfaces of equal topography. Instead, these concepts and ways of being are taken up and used by people who make meaning of them in the different global contexts in which they operate. Various theorists have tried to work in notions of embodiment to produce a more materially grounded critical geopolitics, incorpor ating ethnographic approaches alongside textual analysis, and in terms of the bodies of those caught up in the representational practices analyzed by critical geopolitics.

Feminist critics have noted a more general failure to acknowledge feminist heritage, so that the widening of 'the political' is seen to come from post structuralist sensibilities rather than feminist arguments about the personal being the political, reproducing the idea of geopolitics as a masculinist tradition. While women have not generally been part of the hallowed intellectual so cieties or elite 'intellectuals of statecraft' critiqued by critical geopolitics, their bodies most certainly have al ways been caught up in the resulting political geog raphies: flows of migrants and representations of vulnerable women and children to be protected by the state. So, although women have not been found in the formal spheres of geopolitics, they are very much caught up in the day to day practices upon which international relations depend. As feminists have highlighted, while men can stand in for their country in battle, women tend to literally embody the nation as its mothers and daughters, their bodies are often clearly inscribed with geopolitical meaning as the territory to be defended by men, potent symbols of national purity to be protected from foreign violation. For women in such conditions, of course, geopolitics is never simply a textual device.

Others have analyzed the production of radical al ternatives to dominant geopolitics, not in academic cir cles, but in the grassroots of collectives existing outside of the state. Such 'antigeopolitics' also seeks to open up space for forms of resistance beyond academic textual analysis.

See also: Anti-Geopolitics; Cold War; Critical Geography; Environmental Security; Foucauldianism; Geopolitics; Military and Geography; Political Geography; Poststructuralism/Poststructuralist Geographies; Terrorism; War; World-System.

Further Reading

Ashley, R. (1987). The geopolitics of geopolitical space: Towards a critical social theory of international politics. *Alternatives* 14, 403 434.

Campbell, D. (1992). *Writing Security*. Minneapolis, MN: Minnesota University Press.

Dalby, S. (1990). *Creating the Second Cold War: The Discourse of Politics*. New York: Guilford.

Dalby, S. (2003). Environmental geopolitics: Nature, culture, urbanity. In Anderson, K., Domosh, M., Pile, S. & Thrift, N. (eds.) *Handbook of Cultural Geography*, pp 498 509. London: Sage.

Dalby, S. and Ó Tuathail, G. (eds.) (1998). *Rethinking Geopolitics*. New York: Routledge.

DerDerian, J. and Shapiro, M. (eds.) (1989). *International/Intertextual Relations: Postmodern Readings of World Politics*. Lexington, MA: Lexington Books.

Dowler, L. and Sharp, J. (2001). A feminist geopolitics? *Space and Polity* 5, 165 176.

Flint, C. (2006). *Introduction to Geopolitics*. New York: Routledge.

Hyndman, J. and de Alwis, M. (2004). Bodies, shrines and roads: Violence, (im)mobility and displacement in Sri Lanka. *Gender, Place and Culture* 11, 535 557.

Ó Tuathail, G. (1996). An anti geopolitical eye: Maggie O'Kane in Bosnia, 1992 93. *Gender, Place and Culture* 3, 171 185.

Ó Tuathail, G. (1996). *Critical Geopolitics*. Minneapolis, MN: Minnesota University Press.

Ó Tuathail, G. and Agnew, J. (1992). Geopolitics and discourse: Practical geopolitical reasoning in American foreign policy. *Political Geography Quarterly* 11, 190 204.

Said, E. (1978). *Orientalism*. New York: Vintage.

Sharp, J. (2000). *Condensing the Cold War: Reader's Digest and American identity 1922 1994*. Minneapolis, MN: University of Minnesota Press.

Sparke, M. (2005). *In the Space of Theory: Postfoundational Geographies of the Nation State*. Minneapolis, MN: Minnesota University Press.

Critical GIS

N. Schuurman, Simon Fraser University, Burnaby, BC, Canada

Definition: Critical GIS

Critical GIS is a unique amalgamation of social theory and geographic information science (GIS) that is characterized by its emphasis on extending the functionality, as well as democratization of the technology. This dual social/technical emphasis gives rise to an agenda that is concerned with socioeconomic, feminist, epistemological, ontological, and participatory elements of GIS. While critical GIS emerged from a series of critiques from human geographers, it has been redefined as a creative blend of human and technical geography that has the potential to uniquely shape GIS and obliquely influence other information sciences. There are challenges, however, that range from a need to gain disciplinary approbation to the practical problem of transforming technology to respond to social concerns. There are a number of issues that contribute to an understanding of critical GIS. They include its history, relationship to science and technology studies (STS), feminism and GIS, ontology research, and participatory GIS (PGIS).

Critical cartography far preceded critical GIS with roots in late 1970s when David Woodward paired with Brian Harley to develop a *History of Cartography*. Harley is perhaps better known for revealing the critical dimensions of cartography. He claimed, for instance, that maps have always been a means of depicting and producing social relations. He used the famous example of the king's castle that is frequently illustrated in large size on feudal maps while whole villages comprising cottages are missing. Indeed, Harley famously concluded that the map is not the territory but a representation of social relations. It follows that maps not only represent but enable certain relations of power. This view corresponded to sociologist John Law's conceptualization of maps as a mechanism for ordering societies. Law noted that between 1400 and 1800, Europe underwent transformations in social organization during which maps were critical for the representation and imposition of the 'truth'. Harley's legacy was in drawing attention to the unity between maps and power. Raper and Livingstone applied this structure to GIS arguing that data models are theories of the world. In this view, measurements are not taken without being informed by theory. In addition, representation is closely linked to ideology – a key tenet of critical GIS.

Critical GIS began to converge as an identifiable subdiscipline in the early 1990s as human geographers became aware of the increasing importance of GIS within geography. While some human geographers were suspicious of GIS' claims, others were concerned about the direction that GIS was leading the discipline. Critiques of GIS that focused on apprehensions of the effects of GIS on geographical knowledge and disciplinary epistemologies were first published in 1990. The next 5 years witnessed many publications by human geographers who were often skeptical about the role that GIS could productively play in the discipline.

Critiques of GIS by human geographers can be characterized as three distinct periods. The first, between 1990 and 1994, was characterized by rather fierce criticisms of GIS and spatial analysis by human geographers. Though many of these analyses were warranted – especially as GIS had been oversold in its early days – the tone of the papers was frequently vituperative. The arguments were, however, relevant as authors appealed for a more thoughtful and reflective GIS that might concern itself with epistemological and ontological implications of spatial analysis and representation. The flavor of early critiques of GIS was, however, unwittingly antagonistic.

The second period of critique took place between 1995 and 1998. Over 40 papers were published in 1995 alone by a mix of human geographers worried about the implications of a widely disseminated GIS. Many of these publications were included in two key 1995 anthologies: *Ground Truth* edited by John Pickles, and *GIS and Society*, a special issue of *Cartography and GIS* edited by Eric Sheppard. Concern was expressed that GIS best served large corporations and public agencies rather than the disenfranchised. Critics were especially skeptical about links between the quantitative revolution and GIS. Human geographers largely regarded the quantitative revolution as a mechanistic approach to geography and worried that GIS propagated many of the underlying principles. Critics were also suspicious that GIS was overly masculinist and an invasion of spatial privacy. The period of critical GIS from 1995 to 1998 was thus characterized by a mixture of substantive issues and theoretical concerns that predominated in human geography.

During this period, however, a number of critics began to build liaisons with GIS researchers in the field. The groundwork for this closer collaboration was laid by the US based National Center for Geographic Information Analysis (NCGIA) Initiative 19 *GIS and Society*. Initiative 19 provided the basis for close contact between critics of GIS and academic researchers, and led to a more relevant set of arguments about the role and potential of GIS in society and within the discipline of geography. Thus by 1998, critics of GIS were more engaged with the technology. There was also increasing acceptance that GIS occupied a permanent place in the discipline of geography. More importantly, by the late 1990s, critical GIS had moved from an antagonistic activity to a means of positively affecting a technology that was being widely adopted in other disciplines and in the commercial sphere.

This trend toward a potentially positive, creative critical GIS can be attributed to the large number of collaborative efforts between GIS researchers who were receptive to critiques from human geographers and other critics. Moreover, early critical GIS spawned recognition among GIS scholars that critical geographers could identify social and even technical implications of GIS technology. The cooperation depended, however, on GIS researchers being willing to extend the scope of legitimate research questions in the discipline.

The beginnings of this dynamic cross fertilization were spawned during a meeting at Friday Harbor in 1993. The Friday Harbor meeting was unusual by any standards given its goal of reconciling disciplinary antagonisms. That initial conference was spearheaded by Tom Poiker who was perturbed by rumblings of discontent about GIS from human geographers; the organizing committee included Eric Sheppard and John Pickles both of whom had recently published critiques of GIS in the *Professional Geographer*. The Friday Harbor meeting created a precedent of bringing together diverse group of scholars from within the discipline of geography, and began what is arguably the unique discourse of critical GIS.

Dissemination of collaborative work between critical theorists and GIS scholars has aided the cross fertilization that characterizes critical GIS today. A meeting, organized by Francis Harvey in the spring of 1999 and held in Kentucky brought together critics of GIS, as well as well known classification theorists Geoff Bowker and Susan Leigh Star. Selected papers from this meeting were subsequently published in *International Journal of Geographical Information Science*, thus increasing the opportunity for dialog between the technical and critical communities. More recently, Mei Po Kwan, Francis Harvey, and Marianna Pavlovskaya organized a conference on critical GIS at Hunter College in 2001. Papers from this conference were published as a special issue of the journal *Cartographica* in 2005. Likewise, David Unwin sponsored a meeting on Spatial and Temporal Representation in GIS in the UK in August of 2001. The meeting was designed to bring together critical theorists and GIS analysts to develop an agenda for diversifying representation in GIS, and incorporating time based analyses. Papers from that meeting were published in 2005. Through efforts such as these, critical GIS was incrementally disseminated to a broader community of human geographers and information scientists.

Critical GIS, Science Wars, and Social Constructions of Technology

The 1990s witnessed a phenomena known as the science wars in which critics of science pointed to social influence on scientific constructions of knowledge. The subsequent emergence of critical GIS has roots in these debates though it cannot be subsumed under the rubric of the science wars. To do so undermines the considerable efforts made by the GIS community to accommodate criticism as evidenced by NCGIA I19 and the subsequent Varenius Project. Like earlier NCGIA initiatives, the Varenius Project was funded by the National Science Foundation under the NCGIA rubric. Varenius is named for the seventeenth century Dutch geographer, Bernhard Varenius, who initiated mathematical bases for geographical inquiries. The project was directed toward research in three areas of geographic information science: (1) cognitive models of space; (2) computational aspects; and (3) 'geographies of the information society'. For the critics, it trivializes their criticism by assuming that they are repeating axioms from STS. It also fails to recognize that many STS researchers focus on fields other than their own. In STS, sociologists study and critique physics or biology. In geography, however, critics have turned their attention to their own field, an area in which they have considerably more stake. It also discounts the willingness of critics to work with GIS scholars.

At the core of the science wars was the epistemological privilege enjoyed by science and the degree to which science is culturally influenced. Social scientists were increasingly critical of science 'proper'. Proponents of science defended themselves with the publication of books and articles intended to undermine the politics and motivation of its critics. The science wars were made truly contentious by the 1996 publication of an article called 'Transgressing the boundaries' by physicist Alan Sokal. Sokal putatively discussed parallels between twentieth century mathematics and physics and their relation to ongoing work in STS. His was a dense, meandering text that should not have evoked much comment nestled in the company of excellent companion pieces on the cultural construction of science. On the day of its publication in *Social Text*, however, Sokal published an exposition in *Lingua Franca* in which he claimed that his paper was a hoax, a parody of attempts by social scientists to link science and culture. *The New York Times* considered this trick worthy of an editorial and Sokal, himself, was eager to talk to the media. The science wars went public. In the aftermath, there were many rounds of debate.

Critics of GIS from human geography neither exclusively emphasized the social construction of GIS nor did they, for the most part, engage with the theoretical critiques emerging from STS. Staples of STS such as actor network theory (ANT) and the symmetry of true and false claims were eschewed in favor of epistemological arguments, and specific attention to the 'effects' of GIS, as well as the direction of the discipline. Debates over GIS in geography were motivated by a genuine desire on both sides to steer the discipline in an appropriate and responsible direction. The history and content of those debates is as distinguished by its differences from the science wars as its similarities. In each instance, social scientists initiated a critique of their counterparts in science and technology. It is the nuances of the debate 'within geography' that mark critical GIS as it is emerging in the early twenty first century.

Critical GIS is clearly related to STS though the two remain distinct. Critical GIS is, first of all, situated within geography. It involves a very different set of negotiations to study one's own discipline. STS tends to focus on disciplines outside of the researcher's domain. John Law a leading sociologist of science, for instance, spent considerable time following scientists 'in action' in Daresbury Nuclear Laboratory. Critical GIS researchers focus on our own discipline, seeking to uncover contingent and social practices that contribute to the development of GIScience.

As outsiders, STS scholars can afford to be more aggressively critical of science. Donna Haraway has discussed the use of tactics by social theorists to undermine science. She expresses caution about attempts by social constructionists to demonstrate that science is just another text, like history, produced for the power and profit of dominant players. Schuurman and Pratt argued in 2002 that there is an inside/outside to critique. External critique has little invested in the outcome while internal critique is necessarily cautious and careful, as it has a stake in the future of the technology. To be constructive, critique must care for the subject. While critical GIS began as an external critique with an attendant harshness, it has decidedly coalesced as an influence within GIS that is wedded to the goal of a better technology. Critical GIS shoulders the burden of responsibility to a discipline while simultaneously deconstructing it bases.

In addition, critical GIS has been differentiated historically from STS by its methodologies. Critical human geographers employed few STS tools until recently. In the latter half of the 1990s, a few critical GIS researchers began to use ANT to describe the development and dissemination of GIS technology. ANT is a systematic method of tracing the development of sciences and technologies through social processes. Frequently, ethnographic techniques such as interviews are used to track the development of ideas and technologies. Such studies are aided by the analysis of 'inscriptions' related to the technology such as publications, press releases, or grant proposals that authenticate the science in an iterative fashion. Though ANT has been instrumental in illustrating that science and technology depend on social processes to develop, it is limited in that it tends to emphasize the science as consequence of sociology with few links to materiality. This leads very easily to a rejection of STS claims by scientists who frequently rely on a realist/empiricist interpretation of the world.

If critical GIS is to remain a system of constructive internal critique, then it must find ways to communicate social processes to scientists in a manner that is relevant to their own work, and acknowledge a fundamental materiality. A framework that satisfies these demands was described by David Demeritt in a 2001 article in the *Annals of the Association of American Geographers*. Heterogeneous constructivism acknowledges that natural systems including geographical phenomena are influenced by a broad range of social practices, but, are nevertheless linked to a fundamental reality. It follows that we cannot ascertain the degree to which science reflects phenomena 'as they truly are', but accept that they are a hybrid reflection of material and social processes. This iteration of STS bestows an epistemological freedom on critical GIS research while emphasizing that social influences on science and technology modify but do not substitute for material reality. By integrating heterogeneous constructivism – with its implicit recognition of the politics of knowledge – critical GIS is better positioned to successfully communicate insights to developers and users of GIS.

Heterogeneous constructivism is the rubric for a recent trend in critical GIS that points to social influences in the development of technology. This technique can point to specific instances where GIS is affected by, or constrains social processes. For instance, generalization research in GIS was delayed by a cartographic tradition that stressed map simplification rather than database, or model, generalization. Mei Po Kwan has found that conventional spatial methods for evaluating accessibility suffer from gender bias, and are inadequate for demonstrating the space–time constraints in women's lives. In both instances, points of social influence were identified in GIS construction. Heterogeneous constructivism is a tool for studying the algorithmic basis for GIS. It implies a more GIS savvy researcher who is interested in questions that extend beyond social construction to reconstruction.

Feminism and GIS

External critiques from human geographers tend to be concerned with epistemological assumptions and social repercussions, while internal critiques have focused on the technical. However, there is a further difference. Internal critiques have a stake in the future of the technology while external ones tend not to. To be constructive, critique must care for the subject. A feminist approach urges a change of attitude in critiques of GIS, from morally and intellectually superior outsider to a critic who is more open about his/her complicities in shared circuits of power and intellectual traditions. This change in dir ection must be accompanied by a refocusing of the ob jectives of criticism, from that of exposing error to the task of uncovering the production of truth.

Understanding the effectiveness of GIS and the pro duction of truth within and through GIS, requires a specialist's knowledge of it. Feminists may see tactical ways of using existing GIS technologies to further goals for social justice. Mei Po Kwan argued, for instance, that critical GIS in a feminist context needs to be more re flexive in order to produce truths otherwise concealed. Likewise, Kwan has demonstrated that it is possible to develop feminist discourses in GIS using visualization. The technology can, in fact, be used to enrich feminist geography and practices. Despite these promising de velopments, feminist geography and GIS – two very ro bust elements of the discipline – remain ironically separate.

Critical GIS and Ontology Research

Early critiques of GIS focused on the need for prac titioners to examine the ontologies they worked with and the epistemologies of their practices. Ironically, main stream GIScientists took up the call to address onto logical and epistemological issues. In the process, they have integrated many of the concerns expressed in early critiques but within the technological framework or dis cursive regime that underlies GIScience. There is a powerful divide between the abstract concepts that are used to articulate critiques of GIS and the formal lan guage – ultimately code – that is required to implement changes to the technology. Concepts such as subject– object dualism, or epistemological intransigence do not translate neatly into machine language. As a result, it has been easier for GIScientists to develop computable mechanisms for inclusion of multiple ontologies in a GIS that can represent multiple points of view.

Ontology research in GIScience has increasingly fo cused on problems with categories and their failure to travel well across the boundary between conceptual ization and computation. For example, the European Union (EU) has developed a vegetation classification system based on biotypes. Some conservationists have noted, however, that UK classifications do not match Russian vegetation types well. Also, some Irish vegetation types are unclassifiable using the system. The classifi cation systems reflect not only different frames of refer ence, but were developed under different vegetative and climatic regimes. Neither reflects a universal reality but local settings and their institutional cultures. Ontology research in GIS is charged with developing a mapping from multiple spatial ontologies to database models with minimum loss of meaning.

Multiple models for achieving this goal have been developed under the rubric of ontology research. One method uses the analogy of node arc structures from graph theory. In this scenario, nodes are equated with conceptual elements (intrinsic properties) and arcs with relationships (extrinsic properties). The goal of this ap proach is to identify close or near concepts and tag database elements so that proximate concepts and en tities can be linked. Methods of addressing this problem also include restructuring databases to illustrate links between semantic objects based on context. Another method for including multiple ontologies is the use of inclusion rules for determining which elements belong together. Each of these methods is linked by emphasis on including context in the final database as a means of ensuring that the terms of classifications can be recog nized in different settings. Through ontology research, GIScientists have contributed to the body of research called critical GIS.

Participatory GIS

A key contribution of critical GIS has been to include communities in the discourse of GIS. PGIS is envisioned as a flexible system, comprising a suite of integrated methods and technologies that, through the incorporation of multiple perspectives and a diversity of alternative information forms, facilitates collaborative planning ef forts, supporting inclusive public participation in de cision making processes. It is both a computer based system employing a collection of software tools and an interactive human process. The public participation pro cess involves an exploration and description of the problem, exploration and evaluation of solutions, creation and evaluation of alternatives, feedback, and finally pre scription of an acceptable approach. It aspires to increase public involvement in defining, analyzing, and resolving issues related to space and location, especially for groups traditionally excluded from such processes. It also rep resents the expansion of GIS research to the sociopolitical implications of public use and the role of technology in democracy.

Since the development, design, and implementation of GIS is also social process, the technological components cannot be isolated from the sociopolitical context of its application. GIS has been perceived both as an em powering and marginalizing force upon community groups, shaping and being shaped by associated politics and power relationships. Harris and Weiner argued in 1998 that empowering and disenfranchising processes both occur simultaneously depending on perspective. Accordingly, societal consequences have become a priority within PGIS research. PGIS research studies the value GIS offers marginalized groups, the nature of their application, and further develops innovative ap proaches, technologies, and institutions to support their empowerment.

A strong interest in promoting public participation and understanding the underlying mechanisms and methods began to develop in the late 1960s and early 1970s. By the late 1970s and 1980s, substantial criticism had developed toward the prevailing scientific approach to planning, which failed to recognize the link between the act of planning and the implications for the society being planned while defining spatial components in isolation from their contextual situation. This encouraged the new perspective that planning must be integrated within the framework of diverse interest groups, rather than existing independently from its sociopolitical context.

The principle goal of the PGIS movement is to pro mote the empowerment of citizens by facilitating col laborative planning such that community groups, who were traditionally excluded from planning processes by the technocratic elite, have the ability to participate in policy related decision making. This is embodied in a system of tools, techniques, and approaches, both tech nological and social. Another aim is to maximize acces sibility, both in physical access to the system and comprehension of its functionality. The incorporation of local knowledge is a crucial advantage not only for community members but for planners who recognize that local knowledge is often superior to conventional 'expert' knowledge of a place and may contribute to stronger analyses.

Dissemination of information is critical for the strength, progress, and success of community groups. Community groups may use geographic information for specific administrative, strategic, tactical, and organiza tional purposes but this underlies the broader theme that better information allows community stakeholders to ask better questions about the issues about which they are most concerned. PGIS is capable of providing citizens with information and dissemination options not only by bringing together multiple perspectives but also through the capacity to act as portal to government and pro fessional data. The power to assimilate data from mul tiple sources and formats increases the depth and diversity information available, enables communities to better represent their conceptualization of space, and aids stakeholders in gaining better understandings of the complex interrelationships that underscore public policy decisions. Technological components of this system must be designed such that the ease of use and comprehension supports effective use by all sides of public debates. The implementation of PGIS is also greatly influenced by the local social and political context of its use. Greater depth of involvement also gives every project the potential to improve the community–government dynamic.

PGIS has perhaps been the most effective form of critical GIS. It has transformed the area of criticism into a constructive exercise with the participation of millions of people worldwide. Its successes have furthered the goals of critical GIS because: (1) PGIS has been forced to develop user friendly interfaces in order to increase ac cessibility of GIS; and, as a result, (2) GIS has been disseminated to a broader circle of users thus democra tizing the technology.

Critical GIS in the New Millenium

Critical GIS has been a part of geography's landscape for long enough to have a history and respected role – ar ticulated by Eric Sheppard in 2005. Sheppard examined the history of critical GIS, drawing attention to its key moments. He concluded with a caution that critical GIS scholars must continue to pay attention to the suppos itions that animate the technology and their interpret ations of it. David O'Sullivan also summarized the progress of critical GIS in 2006. O'Sullivan characterized critical GIS as a means of theorizing GIS through understanding its origins and development. This is reminiscent of Harley's attention to the historical rela tionship between power and mapping and consistent with many of the goals of critical GIS thinkers. O'Sullivan argues that critical GIS will remain relevant so long as it hosts an intellectual fusion between technologists and theorists. Marianna Pavlovskaya has articulated means for combining qualitative approaches in GIS with the quantitative and illustrates that the two approaches to research have never been separate. Certainly that is a tenet of critical GIS.

The goal of critical GIS to extend the functionality of GIS while democratizing it has been enhanced by in fluences from STS and critical geography, feminist geographers, ontology research, and the successes of PGIS. The future of critical GIS hangs, however, on the willingness of geographers and users to engage both theoretically and technically with the systems and sci ence of GIS.

See also: Actor-Network Theory/Network Geographies; Critical Cartography; Feminism, Maps and GIS; Geodemographics.

Further Reading

Curry, M. (1997). The digital individual and the private realm. *Annals of the Association of American Geographers* 87(4), 681 699.

Demeritt, D. (2001). The construction of global warming and the politics of science. *Annals of the Association of American Geographers* 91(2), 307 337.

Elwood, S. (2004). Partnerships and participation: Reconfiguring urban governance in different state contexts. *Urban Geography* 25(8), 755 770.
Fisher, P. and Unwin, D. (eds.) (2005). *Re presenting GIS*. London: John Wiley & Sons, Inc.
Ghose, R. (2005). The complexities of citizen participation through collaborative governance. *Space and Polity* 9(1), 61 75.
Goodchild, M. F. (1992). Geographical information science. *International Journal of Geographical Information Systems* 6(1), 31 45.
Harris, T. and Weiner, D. (1998). Empowerment, marginalization and 'community integrated' GIS. *Cartography and Geographic Information Systems* 25, 67 76.
Harley, B. (1989). Deconstructing the map. *Cartographica* 26, 1 20.
Harley, B. (1990). Cartography, ethics and social theory. *Cartographica* 27, 1 23.
Harvey, F. and Chrisman, N. R. (1998). Boundary objects and the social construction of GIS technology. *Environment and Planning A* 30, 1683 1694.
Jankowski, P. and Nyerges, T. (2003). Toward a framework for research on geographic information supported participatory decision making. *URISA Journal* 15, 1 APA I.
Kwan, M. (2002a). Is GIS for women? Reflections on the critical discourse in the 1990s. *Gender, Place and Culture* 9(3), 271 279.
Kwan, M. P. (2002b). Feminist visualization: Re envisioning GIS as a method in feminist geographic research. *Annals of the Association of American Geographers* 92(4), 645 661.
Law, J. (1994). *Organizing modernity*. Oxford: Blackwell.
McLafferty, S. (2002). Mapping women's worlds: Knowledge, power and the bounds of GIS. *Gender, Place and Culture* 9(3), 263 269.
O'Sullivan, D. (2006). Geographical information science: Critical GIS. *Progress in Human Geography* 30(6), 783 791.
Pavlovskaya, M. (2006). Theorizing with GIS: A tool for critical geographies? *Environment and Planning A* 38(11), 2003 2020.
Pickles, J. (ed.) (1995). *Ground truth*. New York: Guilford Press.
Raper, J. and Livingstone, D. (1995). Development of a geomorphological spatial model using object oriented design. *International Journal of Geographical Information Systems* 9(4), 359 383.
Raper, J. and David, L. (2001). Let's get real: Spatio temporal identity and geographic entities. *Transactions of the Institute of British Geographers* 26, 237 242.
Schuurman, N. (1999). Critical GIS: Theorizing an emerging discipline. *Cartographica* 36(4), 1 107.
Schuurman, N. (2000). Trouble in the heartland: GIS and its critics in the 1990s. *Progress in Human Geography* 24(4), 569 590.
Schuurman, N. (2002). Reconciling social constructivism and realism in GIS. *ACME: An International E Journal for Critical Geographies* 1(1), 75 90.
Schuurman, N. and Geraldine, P. (2002). Care of the subject: Feminism and critiques of GIS. *Gender, Place and Culture* 9(3), 291 299.
Sheppard, E. (1995). Sleeping with the enemy, or keeping the conversation going? *Environment and Planning A* 27, 1026 1028.
Sheppard, E. (2005). Knowledge production through critical GIS: Review and assessment. *Cartographica* 40, 5 22.
Sieber, R. E. (2004). Rewiring for a GIS/2. *Cartographica* 39(1), 25 40.
Weiner, D. and Harris, T. (2003). Community integrated GIS for land reform in South Africa. *URISA Journal* 15, 1 APA II.
Wright, D. J., Goodchild, M. F. and Proctor, J. D. (1997). GIS: Tool or science? *Annals of the Association of American Geographers* 87, 346 362.

Relevant Websites

http://www.geog www.sbs.ohio state.edu/faculty/mkwan/cgis/CGIS Home.html
The Critical GIS Site and Listserve.

Critical Rationalism (After Popper)

D. Bennett, Carleton University, Ottawa, ON, Canada

Glossary

Coherence Theory of Truth A theory of truth in which a statement is true if it coheres with (fits with) other statements, false if it does not fit.

Correspondence Theory of Truth A theory of truth in which propositions are true if, and only if, they correspond in some way with facts, or the world, or states of affairs.

Deduction Also called 'nonampliative inference' is inference from the general to the particular. In deductive argument, for a given set of premises, the conclusion must follow.

Empiricism The philosophical doctrine that our knowledge comes only from experiences which arrive in the mind which is otherwise blank, a tabula rasa, through the five senses.

Epistemology The philosophical theory of knowledge, it is concerned with criteria for distinguishing between our beliefs and our knowledge claims.

Induction Also called 'ampliative inference' it is inference from the particular to the general. In inductive reasoning, the conclusion only follows 'more or less'.

Metaphysics That part of philosophy which ruminates on questions of being, of what there is, and why there is something rather than nothing. It is opposed by logical positivists, by analytical philosophers, and by postmodernists and deconstructionists, separately and on very different grounds.

Popper

Popper contributed to epistemology in natural and social science, to the philosophy of history, the philosophy of language, to logical theory, and to political theory. He belonged to a generation of outstanding scientists, mathematicians, and philosophers who emerged from Central Europe, from Vienna in particular, in the last two decades of the nineteenth century and the first half of the twentieth century. They were German speaking, but their influence was negligible in German culture. At the out break of World War II, a few opted to stay in Axis Europe but most had already fled to Britain or America. As their impact on Anglo American science and phil osophy grew, the gap between the British analytical tra dition and the French and German styles of existential and speculative philosophy widened.

Karl Raimund Popper (1902–94) was born into a middle class family in Vienna. His parents were Jewish but had converted to Lutheranism: Karl was baptized a Protestant. His mother, Jenny Schiff Popper, was an ac complished pianist, and eminent musicians were family friends. Young Karl was a chorister with the avant gar dists Schoenberg and Webern, and in his youth he con sidered but decided against a career in music. His father, Simon, was a successful liberal barrister and Doctor of Law at the University of Vienna, where Karl studied mathematics, music, psychology, physics, and philosophy. Popper received a primary school teaching diploma in 1925 and his PhD in philosophy in 1928. In 1929, he qualified to teach mathematics and physics in secondary schools and between 1930 and 1936, the years of the logical positivists at their height, he was a school teacher in Vienna.

Reacting to growing apprehension about the future course of Nazism, which had already destroyed the lo gical positivists' Wiener Kreis, Popper took up an ap pointment as a lecturer in philosophy at Canterbury University College, Christchurch, New Zealand, in 1937. In 1945, he accepted the post of Reader at the London School of Economics (LSE). He became a naturalized British citizen in the same year. Popper frequently complained that he was socially ostracized in his adopted country, but he was actually a workaholic who eschewed a social life, seeking seclusion and preferring to stay home and work. From 1949 until his retirement in 1969, he was Professor of Logic and Scientific Method at LSE. After retirement, he continued to live quietly in and around London and to work indefatigably at home until his death from kidney failure and pneumonia, compli cations of cancer, in a Croydon hospital in 1994. He had been knighted in 1964. His wife had died in 1985. He had no children. Popper was a Fellow of the British Academy and of the Royal Society, quite a rarity.

Popper and the Positivists

Critical rationalism addresses the problem of induction, the inability of any finite set of cases to lend logical security to an unrestrictedly general statement. Recog nition of the problem goes back at least as far as Hume's *Enquiries Concerning Human Understanding* (1748). Over shadowed by the material successes of applied Newton ian science in the Industrial Revolution, the problem of induction came to be seen as no more than a logic puzzle for philosophers until Popper took seriously its implications for epistemology and science. Popper said that Hume is correct, we cannot move logically from particular instances, no matter how many, to a universal generalization. This is a leap we cannot make in logic, but it is a leap we all in fact make in life because we must, and in this, science is no different from anything else that humans do. We must accept the problem of induction as a limit and seek to understand how we cope with it and its consequences in some other way, because we obviously do. For science, he developed just such another way by replacing the logical positivists' propositional knowledge criterion of verifiability with the criterion of falsifiability. In Popper's hands, this seemingly minor modification ramified into an effective logical demolition of the positivist project, and it entailed replacing the classical conception of scientific knowledge as the indubitable outcome of a disembodied and logically secure process of discovery with the view that scientific knowledge is a human creation which is inherently theory laden, and which necessarily involves elements of metaphysics and myth.

Popper was never a member of the Vienna Circle. He was not invited to sit in on even one of their meetings despite maintaining friendships with Waismann, Feigl, and Kraft, a fact which irked Popper for the rest of his life. Otto Neurath called Popper 'the official opposition', which suggests friendly respect for a fellow parliamentarian who is, for now, on the other side of the argument, but in the long term is committed to the legitimacy of the process and the institution. Popper himself described the relationship as 'friendly/hostile', hostile particularly to Wittgenstein's influence on the Vienna Circle, blaming him for pushing the positivists' antipathy to metaphysics even further into a broadly antiphilosophical attitude. Moritz Schlick, the leader of the positivists, may have kept Popper out of the Circle in retaliation for his being difficult over this issue. Schlick agreed with Wittgenstein that philosophy would inevitably fade away because its audience would come to see that philosophy has nothing meaningful to say. On the other hand, Carnap, the most productive member of the Circle, expressed favorable interest in many of Popper's ideas. But on balance, the positivists thought that Popper's position differed only slightly from their own. In *Confessions of a Philosopher*, Magee suggests that the positivists were incapable of grasping that what Popper was saying was radically counter to their position, and, thus blinkered, they recast critical rationalism into their own categories. They thought that Popper was disputing their answer to their question "what is it that renders some statements about the world meaningful, even if false, and others not?" Their answer was that it resided in empirical confirmation, but as profound difficulties emerged around the logic of verifiability, many positivists simply replaced the verifiability principle with a confirmability principle or with the somewhat stronger idea of a testability principle, and those interested more in the actual structure of science than they were in the integrity of positivism *per se,* took up Popper's notion of the falsifiability principle. They thought Popper had provided rather a clever solution to a technical problem, but that was all. Popper protested that he was not putting forward falsifiability as a criterion of meaningfulness, but rather as a criterion which demarcates science from pseudoscience and from nonscience. The positivists responded that it amounted to the same thing! A. J. Ayer certainly saw it this way, and because the English language world's first acquaintance with the ideas of positivism lay in his 1935 book *Language, Truth, and Logic*, the conflation of Popper and positivism took an early and tenacious hold. This was aided by British, especially Oxford, philosophy's parochialism which embraced Ayer's introduction of positivist ideas despite the fact that Popper had published an effective logical demolition of them before Ayer even wrote his book. Popper's critical rationalist ideas had to wait a quarter of a century for an English translation to make them more widely available.

Critical Rationalism

In a very general sense, Popper and the Vienna Circle were interested in the same questions: where does scientific knowledge come from; what gives it its authority; how is it added to; how does it grow; is it true? But they differed radically in their approaches to these questions, with the result that they differed radically in their answers. Far from being a fellow believer, Popper was profoundly opposed to positivism, which he regarded as fundamentally misconceived.

Another name for logical positivism is logical empiricism. Empiricism is the doctrine that everything that a mind knows comes from experiences of things outside the mind. According to Magee, the members of the Vienna Circle "believed – as had virtually everyone since Newton – that science was an edifice of certain and secure knowledge that was built up by logically deductive argument on the basis of scientific laws that had been conclusively verified." The Baconian–Newtonian view held that science appropriated the world in a totally unconceptualized way – the view from nowhere – reading off the laws of nature from an accumulation of disinterested observations of an unproblematic reality. There were patterns embedded in the observations; the observer had only to find them. Scientific laws, and in turn theories, were built out of such discovered empirical regularities. Thus was constructed a body of knowledge which was universal, correct, and certain. It grew as discovered truth was piled on discovered truth, revealed fact added to revealed fact.

The Vienna Circle set out to protect science as the sole possible way of arriving at this version of truth, by establishing a way of cleaving scientific statements from all other forms of utterance. The keystone in the positivists' conceptual arch was the notion that the meaning of a proposition is the method of its verification. To be verified is to be true: to be true is to be verified. Propositions could be either analytically true or empirically true. Analytical truths, such as those of mathematics, are tautologies, true because they are different ways of saying the same thing. Tautologies can never say anything new. For the positivists, that left empirical truths as the only truths. The positivists tried to work entirely within the correspondence theory of truth, with limited success. In the correspondence theory, empirical truths are statements which are validated by somehow expressing external states of affairs, which is to say that they are propositions which depend on there being some way, in principle at least, of making observations which confirm the proposition. Any proposition which could not, under any circumstances, be translated into an empirically testable form was dismissed as a metaphysical proposition, and all metaphysical statements were nonsense. That is not to say that they were random utterances, nor to deny that they could have moral or rhetorical impact, nor that they could cohere with other metaphysical statements. But they must remain as beliefs or imprecations which we could deny at no cost: they simply did not have the potential to become knowledge because there was no conceivable way of testing them empirically.

In positivism, for a proposition to be true, it had to be empirically verified by induction from a body of observations or test cases, all of which must support the proposition. Popper pointed out that the best we can really say is that all cases thus far fail to falsify the proposition. This may create a psychological habit of expectation that all future cases will do the same thing, but in logic we do not have, and we cannot have, conclusive grounds for claiming that the proposition has been verified. The positivist project was beset by many problems of articulation but loss of the verification principle was, in logic, fatal.

Critical rationalism was presented in Popper's first book, *Logik der Forschung,* which was published in the Vienna Circle's Schriften zur wissenschaftlichen Weltauffassung series in 1935. He later claimed that in creating critical rationalism, he had destroyed positivism. In a logical sense, he had, but in practice and in history *Logik der Forschung* had little impact until it was published in English as *The Logic of Scientific Discovery* in 1959. Popper himself did the translation. Popper is quite unusual in philosophy in that he wrote copiously in advocacy and defense of his own work; so, he took the opportunity to make quite substantial changes to the book under the guise of translation. The English version became more of a revision than a simple translation.

Rather than conclude that appeals to empirical evidence enjoyed no greater status and security than other knowledge claims, Popper argued that what set scientific statements apart was that they could be conclusively shown to be 'false' in logic by the existence of 'one' empirical case to the contrary. He removed the problem of verifiability – or proof – by saying that there can be no proof: indeed, he found the idea of proof totally useless. He substituted the idea of refutation, of falsifiability. And he made the remarkable claim that this solves the problem of induction – or, at least he made the problem go away – by denying that it plays any part in scientific reasoning!

Popper argued that science does not come by its hypotheses, laws, and theories through uncontaminated, pretheoretical observations. Science always sees reality (whatever that is) through some lenses, made up of hypotheses and theories. Apprehension of reality is post theoretical. For Popper, Newton's laws are not laws of nature that Newton discovered, they are laws of Newton, laws that he made up.

If theories are brought to nature rather than vice versa and thus determine the kinds of observations that can be experienced, then where do they come from? They come, said Popper, from imagination, and from our propensity to guess. They may be intuitively arrived at, they may be metaphysical, and they may be what he called myths. Popper held that pseudoscience or 'myth' could inspire proper science, and could, in principle at least, eventually develop into science. Indeed, he considered the general progress of human knowledge to be the process whereby myth is converted into scientific knowledge by its subjection to critical examination. And he was willing to say that, in certain cases, metaphysical statements could fulfill a valuable heuristic function, which set him at loggerheads with the more dogmatic of the positivists.

He held that the proper way of proceeding is to be relentlessly critical, and he favored a deductive process for arriving at empirically examinable statements. Thus Popper combined an empiricist ontology with a rationalist epistemology. Popper called his philosophy of science 'critical rationalism'.

According to Popper, scientific knowledge consists only of those statements which are amenable to empirical interrogation and which have not yet been falsified. Scientific theories live by not dying. So, science must be critical, open ended, situated, incomplete, risk taking, and contingent. Far from seeking ways of creating a canon of empirically supported beliefs, science should seek out disconfirmations. In logic, scientific knowledge emerges from criticism, an endless cycle of conjectures and refutations. In logic, one reliable contrary case can falsify, but an infinity of confirmatory cases cannot prove.

In practice, as Kuhn and others have argued, scientists do not behave in this way, always seeking to falsify; rather, they try to save what they believe is true for as long as possible. But Popper was not concerned with the actual behavior of scientists, or with scientific practice. He was concerned with the logic of confirmation/falsification, and his argument is compelling in its domain. While falsifiability is a simple logical notion in principle, in practice it is less straightforward – no single observation is reliable enough to conclusively falsify a theory, and Popper conceded that, in reality, a single conflicting instance is not enough.

Nonetheless, Popper says that no theory can ever be conclusively verified, proven to be an unqualified and universal truth. It is not a case of needing to try harder, or getting more data, or clarifying the language. Science as the pursuit of certainty is a misguided enterprise. So is philosophy. Certainty is an unachievable state, although it can be expressed as an ideal. We must settle for less, and the western, centuries old, search for certainty has to be abandoned. The real question is not 'must we accept uncertainty', but 'how much uncertainty must we accept'? In everyday life, we use theories that yield practical results in that they solve puzzles that confront us. We humans continue to use provisionally successful theories until they fail to solve important puzzles, then we replace them with something better, or we don't and possibly we die. So, according to Popper, we are problem solvers, not knowledge builders. Science is not a robust edifice of knowledge brick placed on knowledge brick, an accumulation of empirically confirmed verities. Rather, science is a perpetual work in progress, which is useful to us in that it has given us as a species some successes in dealing with important, including life threatening and life sustaining, problems. And in this respect, science is not markedly different from everyday life: it does not have a unique method – it codifies, clarifies, and culturally transmits practices that humans have developed to deal with survival problems.

Criticism is the crucial engine driving this forward, although if we have given up – as we must – the possibility of completed certain knowledge, we cannot say where we are being driven to. A falsified theory is replaced by another, better theory, better in the sense that it has not yet been falsified. That seems a pretty straightforward substitution. But how are we to choose among several theories that are all ostensibly the same in that they are currently alive, that have not yet been falsified? Popper says that living theories are to be preferred according to how falsifiable they are, how risk taking they are. The more a theory is open to refutation but has not yet been refuted, the more it is to be preferred. It has more empirical content. Popper goes on to claim that surviving an attempted refutation 'corroborates' a theory, and to the extent that the theory is more readily falsifiable, a failed refutation affords greater corroboration. This seems to readmit the inductive method that he rejected at the outset, and this remains problematic. In this same vein, Popper avoided, for a long time, saying 'truth'. It was not a concept with which he was comfortable. But in comparing and choosing between theories, he found himself talking about which was better inasmuch as it approached the truth – which he called verisimilitude, itself something of an obfuscation.

Popper's Political and Social Philosophies

During his time in New Zealand, Popper turned his attention to social and political issues which he had first encountered in his early teens. In 1919, he had become heavily involved in left wing politics in Vienna. He was, for a short while, a Marxist, but became disillusioned by its doctrinaire nature which he once saw played out in a clash between unarmed Communist youths and police, in which several of the youths were shot dead. The local communists accepted this as a tactically useful outcome. For the true believers, regret was a useless emotion: events such as this were to be considered inevitable on the road to revolution. Popper found this form of fatalist capitulation of individuality to be morally repellant. Later, during the 1930s, Popper was appalled by the failure of Austrian democratic parties to confront and repel the rising tide of fascism, and by the welcome extended to it by the Marxists, because it fitted with their theories of historical, scientifically underpinned, inevitability.

But before then, he had joined the Democratic Socialist party, and continued to work with the disadvantaged in Vienna, where he encountered the application of the psychoanalytic theories of Freud when he worked – under the direction of Adler – with handicapped children. Popper was struck by what he saw as the contrast between Einstein's openness to critical inspection and the total absence of any critical spirit in the theories of Freud, Adler, and Marx. Also, he noted, Einstein's theories could be falsified empirically, whereas the others set up their theories in such a way that they were amenable only to confirmation. Popper realized that critical rationalism gave him weapons with which to attack Marxism and Freudianism by portraying them as pseudosciences. He was not particularly concerned about traditional pseudosciences, such as phrenology, astrology, or alchemy. These had been around for a long time, but they had shown little capability to compete seriously with science. But Marxism and Freudian psychoanalysis, the new and 'imposing' worldviews which had captivated many intellectuals, especially in Vienna, at the turn of the century, worried Popper. These theories claimed to be scientific, to speak with the authority of science, and they certainly appeared to have some of the language and characteristics of science. Popper argued that neither was a science, and never could become scientific because they could explain everything – which is what many find attractive about them – which meant that they could never fail, so they could never be tested. Science and pseudoscience were separated not by the capability of the former to demon strate that its statements 'were' so, but by the latter's in ability, in principle, to be able to show any instance in which they were 'not' so. No imaginable empirical cir cumstance can falsify a nonscientific statement, because a nonscientific statement cannot be put into a form where it can be put to an empirical test. Marxist and Freudian theories were, and would remain, ideologies, preepistemic belief systems: they were – and are – faiths.

Beginning in the 1920s, Popper started to develop a critique of Marxism. He first discussed it publicly in 1935 in a talk entitled 'The Poverty of Historicism'. He sub sequently extended the ideas of critical rationalism into the social sciences and in *The Poverty of Historicism* (1945 and 1957) he attacked the Marxian notion that there are laws of development in history which human beings can discover, and which they are fated to obey. Popper dedicated the book to all victims of the communist and fascist belief in laws of historical destiny. Popper has not been popular with the dominant academic left in the social sciences and humanities which tend to dismiss him as a 'positivist', an all purpose demonization, a needs no explanation 'othering'. By the time *Historicism* became widely accessible in English in 1957, Popper was already well established as a public intellectual in his adopted country, having been propelled from academic obscurity to full blown celebrity by *The Open Society and Its Enemies*, written in English, published in 1945.

Popper's basic argument was that history is influenced by the growth of knowledge, which he had shown to be both unpredictable and intrinsically incomplete. And as Popper regarded history as a chronicle of knowledge, inexorable laws of human history are, he argued, un attainable. Popper's target was not the philosophy of politics but the philosophy of history, specifically the doctrine of historicism, which holds that the goal of social and historical inquiry is the formulation and sub stantiation of general laws of historical development. Such laws permit the prediction or prophesizing of future stages with certainty, and knowing what the future state must be, provide the sole rational means of forming political policy and action now. In other words, present effects can have future causes.

In the two volumes of *The Open Society and Its Enemies* Popper applied the ideas of critical rationalism to polit ical and social theory. He argued that certainty is not available in science, nor anywhere else, so the imposition of a single point of view is never justified. The worst forms of modern society are totalitarian regimes which tolerate no dissent, stifle all forms of criticism, and im pose policies as part of a top down, centralized planning process within a paternalistic state. The vital antidote is criticism, endless conjecture, and refutation in political and social actions. The goal should be the replacement of ideas and policies with better ideas and policies. They are better not in that they lead asymptotically to an ideal state, they are better in that they prevent, seek out, and remove the worst cases of social evils. As we don't know what makes people happy, but we do know what makes them miserable, policies should be aimed at removing avoidable suffering and reducing the effects of handicaps. The society most likely to make this happen exped itiously is the open society, in which critical discussion and opposition is not just tolerated but is the lifeblood of the society. Openness and tolerance are essential.

Popper believed that all totalitarians are ideologues, people with a simple, clear, easily grasped message about what is wrong, who provide one equally simple, clear, easily grasped answer as to how to put it right. Popper argued that they knowingly co opt and exploit ideas – perhaps considerably modified or disguised along the way – from the works of one great philosopher or an other. Popper singles out Plato, Hegel, and Marx as in herently open to totalitarian exploitation.

Popper's claim that Hegel and Marx are historicists is unarguable. However, the case for branding them as in herently totalitarian is not. The connection of Hegel to fascism can only indirectly be established via Hegel's influential assault on the primacy of the individual, a precursor to fascism's belief in the moral supremacy of the collective over the individual. Marx is clearly the founding theorist of what became communist totalitar ianism, but his personal values were libertarian: his his toricism was utopian, yet violent, but not inherently totalitarian. It was as operationalized by Lenin and per fected by Stalin that the Soviet instantiation of Marxism became totalitarian.

Popper adopted Weber's concept of 'methodological individualism' to stand against historicism's structures which respond only to inchoate social forces. In the early 1970s, reacting to an emerging Marxist geography, hu manist geographers would ask "where are the individuals with (some) free will?" Popper's original formulation of methodological individualism seems to have been an elaborate way of saying (as Margaret Thatcher did indeed say) 'There is no such thing as society'. But he later conceded that individuals are what they are, to some extent, because of the institutions in the sway of which their lives are lived. For Popper, institutions and their actions can only be explained as the net aggregation of actions by individuals acting rationally within their social situations as they perceive them to be, as best they can. Resonances with social theory, agency/structure, and structuration theory abound.

Scholarly inaccuracies in *The Open Society* are plentiful and easily found, and Popper readily concedes as much. He attributes them to lack of access to library resources during his wartime isolation in New Zealand. Poor resources notwithstanding, Popper dismisses his critics as, what Berkley called, the 'minute philosophers' who indulged in hairsplitting criticism, scholasticism, burying great ideas in floods of words, then arguing about the meaning of the words. This, in Popper's opinion, is bound to be fatal to proper philosophy. And he was quite sure his work was proper philosophy.

In the late 1940s, and early 1950s, one third of humanity lived in nominally Marxist states. The largest was a Stalinist despotism operating a very corrupted version of Marxism as modified by Lenin. In the West of the early Cold War, Popper's ideas were widely well received, and he was propelled from academic obscurity to lionized public intellectual. Later, the adoption of many of his ideas by Thatcherites allowed the left to cast him as a reactionary, rather than as an ultralibertarian, and the collapse of the Soviet Union robbed him of his main target.

Popper's theory of knowledge (critical rationalism) and his social theory are linked by fallibilism: progress in science depends on theories being endlessly exposed to critical scrutiny – in society, progress depends on policies being exposed to public discussion, reformulation, and criticism as the sole valid process of legitimation in an open society. Political discourse, as much if not more than scientific discourse, required the unrestricted play of ideas and criticism. Dogmatism led not to utopia but to the dystopia of totalitarian oppression.

Popper's Theory of Objective Knowledge: Three Worlds

Popper employed Darwinist metaphors throughout his writings on critical rationalism, and in his political and social philosophy. He changed his mind about Darwinism several times, sometimes regarding it as tautological and therefore unfalsifiable, at other times seeing it as a historical hypothesis, open to refutation and therefore open to change. Sometimes it was a science, sometimes not. Eventually, he settled on the view that knowledge is an evolutionary product. It grows through an ongoing cycle of change (conjecture, testing, and refutation) although not toward some final state of completed truth. It progresses, but not toward a perfect endpoint. It exists as stored knowledge even if no one consults it. In ch. 2 of *Objective Knowledge* (1972), he developed the notion of a theory of knowledge without a knowing subject, challenging monist (physical materialism) and dualist (physical world plus the world of mind) views of what there is in the universe, suggesting instead that a pluralism (at least three worlds) is more convincing. He readily conceded that he could not conclusively refute theories of worlds one and two, falling back instead on mounting a strong challenge: refutation, he said, is conclusive and unavailable, whereas persuasion is not conclusive but is the best one can do.

Popper asserted that (1) the problems of the theory of knowledge (epistemology) are at the heart of philosophy, popular and academic, (2) that common sense is our only possible starting point: it is not sufficient, but it is necessary, and we can improve on it, which science does, and (3) there exist three worlds: the world of matter (world one), the world of mind and coming to know (world two), and the products of mind, objectively existing creations/constructions (world three).

Russell, Moore, and the analytical philosophers who followed them took knowledge to be instantiated in the minds/mental states of individuals. Popper, however, held that although knowledge is the product of mental activities, it is the thing produced, not the psychological or neurophysiological processes that produce it. And once produced (in an individual mind, which then expresses it publicly), it has existence and is apprehendible to others. It enters world three which consists in the publicly accessible embodiments of knowledge, man made yet autonomous. Objective knowledge resides in the reality of world three. Here can be found the products of the human mind, languages, myths, stories, scientific theories, music, pictorial art, architecture, feats of engineering, and so on. World three objects belong at the same time to world one where they have physical, concrete expression as artifacts and objects.

Popper argues that world three objects are real, when what is real or what exists is whatever has a causal effect on physical things. World three objects have causal impacts on humans' world two by being available to inform subjective knowledge, and thereby can causally impact world one. Popper says that world three is very much the same thing as what anthropologists call culture.

Popper and (Human) Geography

In geography, Popper suffers the double indignity of being misunderstood on his own, and having his work mistakenly attached to a misunderstood enterprise. In geography, Popper is typically portrayed as some sort of positivist fellow traveler. Most geographers subscribe to what Bennett in his 2005 paper, 'Replacing positivism in medical geography', has called the 'standard version' of positivism in which Comtean positivism, logical positivism, science as practiced and as theorized, empiricism, quantification, science politics, scientism, and so on have been conflated into a blanket pejorative 'positivism' which is considered to be shallow, inherently reactionary, and to underpin multiple oppressions. At best, critical rationalism is framed as merely a patch applied to positivism to staunch a technical leak, a perception traceable back to Harvey, Johnston, and Gregory, who set the basic parameters of the debates over positivism and its successors, and gave Popper little credit for being distinctive or epistemologically fruitful.

In his ground breaking and hugely influential *Explanation in Geography* (1969) Harvey seriously misrepresents Popper as philosopher of science by attributing to him the belief that science is characterized by certainty, and that any retreat from insisting on certainty is a retreat from science. As we have seen, this is actually the opposite of Popper's arguments. *Explanation in Geography* is only implicitly positivistic. It treats philosophy as a synonym for metaphysics, and methodology as a synonym for epistemology. Harvey was concerned in part with the validity of the covering law model of explanation as it could apply to history, and in this context Popper's ideas about the poverty of historicism are given brief mention.

Johnston's *Geography and Geographers*: *Anglo American Human Geography since 1945* is now in a sixth edition. In the almost 30 years since the first edition, several generations of geographers have grown up getting their disciplinary history from Johnston. They have been told in every edition that Popper is noteworthy mainly for his disputes with Thomas Kuhn, whose *Structure of Scientific Revolutions* developed the notion of paradigm, paradigm shift, and the development of science not as one continuous upward flow but rather as a series of steps in which periods of intellectual tranquility are punctuated by conceptual upheavals. Other than that, Popper is credited with creating the falsificationist modification to the positivist criterion of verification. In the sixth edition, Johnston is still saying that the goal of positivism is the same as the goal of critical rationalism, the development of comprehensive theories with high predictive capabilities at high levels of certainty. Indeed, Johnston says that positivism and critical rationalism are cut from the same cloth: "The two differ on means, not ends." The misrepresentation of Popper as a minor figure, an aftermarket tinkerer who improved one of the poorly designed parts of the positivist vehicle is duplicated in Johnston's *Philosophy and Human Geography*, another standard source for geography students and professionals through 20 conceptually tumultuous years.

For the most part, Gregory (*Ideology*, *Science*, *and Human Geography*) discusses Popper in the conventional context of the logic by which theories can be (dis) confirmed, but he goes on to examine more thoroughly, the theory observation dialectic and its implications for critical sciences. Gregory also recognizes that the differences between critical rationalism and positivism are not as slight as stated elsewhere, but still says that they are not fundamental. Gregory is concerned to critique the entire edifice of deductive nomological geography as spatial science by developing structural, hermeneutic, reflexive, and committed explanations, heading in the direction of social theory. He doubts the applicability of either confirmation or falsification as workable criteria in social science, a view echoed by Sayer, which leaves us to ask 'if we can neither confirm nor deny, what is it that we can do other than endlessly and unconstrainedly speculate'?

James Bird engaged more comprehensively with Popper in a series of papers, culminating in a (strangely neglected) book length meditation on a wide range of conceptual and theoretical issues. In his first foray ('Methodological implications for geography from the philosophy of K. R. Popper', which can be found in *The Scottish Geographical Magazine*), Bird sees Popper providing serious competition for Kuhn as the most important single source for geographers concerned with the philosophy of science. He goes on to state that, contra Johnston, the implications for geography of taking on board some of Popper, are quite different from those arising from encounters with Kuhn. Bird is unusual in that he talks about the origins of theories, and admits issues of mind and neurobiology into the otherwise recursive or infinitely regressing observation/theory/observation sequence. Bird outlines Popper's testability/falsifiability methodology, and explicitly considers the notion of verisimilitude as part of an assessment of its applicability to the status of theories in human 'and' physical geography. He argues that although Popper has shown that comprehensive and exact theories which fit all cases are impossible to attain, theories may be arranged along a scaled continuum of the degree to which unattainability prevails. This depends largely on the data involved in the theorization. Social data, he says, are less tractable than physical. He also develops a brief overview of Popper's three worlds, and suggests that this is a useful conceptualization inasmuch as it helps to dissolve such dichotomies as physical/human geography, science/social science, and field/armchair. Bird also draws in Popper's political philosophy to inform analyses of planning processes and policies of social engineering. In his initial explorations of the potential benefits of encountering Popper, Bird is eclectic and brief. In *The Changing Worlds of Geography*, Bird explicitly adopts the Popperian doctrine of conjecture and refutation as it bears on the central activity of problem solving. He portrays geography developing as a science in a mode of constant revision, in which he argues there has only ever been one revolution and it concerns not science *per se*, nor quantification, nor spatial theorization, but the recognition that humans are generalizers – because to survive they must be – and that uniqueness and singularity are not the same thing. According to Bird, the conceptual revolution which hit geography in the 1950s and 1960s was fundamentally about the recognition that unique cases/people/places/events/objects are members of classes based on qualities and attributes, while retaining their aggregate combinatorial uniqueness. This is the logic of classes which Russell worked out long ago.

Bird elaborates a wide ranging and sophisticated analysis of geographic methods and concepts, built largely around Popper's three worlds, from which he constructs a schema, a conjecture, PAME, as a possible way forward. PAME stands for pragmatic analytical methodology epistemology. Perhaps, Bird was a little too idiosyncratic? Geography has paid little attention to his ideas, which is geography's loss.

In the 1980s, a few geographers (e.g., Alan Hay, R. J. Bennett, John Marshall) tried to unravel the by then entrenched muddling of science, quantification, spatial theory, and so on which is widely known as positivist geography. Marshall, in his 1985 paper, 'Geography as a scientific enterprise', gave a full, knowledgeable, accurate, and nuanced account of Popper's critical rationalism. He sorted out some of the Sherlock Holmesian confusion of induction with deduction, succinctly explained the problem of induction, Popper's solution for it, and went on, uniquely in human geography, to spell out the implications of Popperianism for the atavistic conception of science as the pursuit of completed, certain, truth. He clearly embraced the idea that science proceeds, much as Popper says, with theorization/speculation preceding observation, explicitly recognizing the value laden nature of empirical investigations. And he chastised those who, putatively familiar with Popper's work, still ignored or misrepresented the differences between critical rationalism and positivism, exclaiming that in some cases Popper is actually described as a positivist. Unfortunately 20 years later, this point of view is still being uncritically regurgitated. Hoggart, Lees, and Davies in *Researching Human Geography* (2002) set a new standard for misinformation by declaring that Popper reformulated "…positivism into logical positivism (sometimes also called critical rationalism)."

Haines Young and Petch (*Physical Geography: Its Nature and Methods*) comprehensively explored the relationships between science and physical geography, moving from logic, to the classical view of science, before developing the Popperian, Kuhnian, Lakatosian, and Feyerabendian critiques, extending them into impacts on methods, measurements, and modeling, eventually calling for the formulation of a critical tradition in physical geography. Whether or not this has cross over relevance to human geography is very much contested. In her 1997 review of 'Philosophies underlying human geography', Elspeth Graham shows that the question of a shared epistemology for social and natural sciences is framed by the larger debates over naturalism/antinaturalism, realism/antirealism, and the hermeneutic turn in philosophy.

This all leads to questions of social theory, humanistic conceptions of free will, reflexivity, rationality, intentionality, agency, and structure, and so on, where a revisited and properly understood Popper could bring some clarity. As it is, Werlen (*Society, Action and Space: An Alternative Human Geography*) alone seems to have tried to see what comes of Popperian notions such as the three worlds, and objective knowledge without a knowing subject, when bent to the task of developing geographical theory in which action takes primacy over spatiality. Werlen began by critiquing Popper but came to realize that those who criticized him had not read him, and that Popper was often unfairly pilloried. Yet, Werlen too is disrespectful toward Popper, explicitly acknowledging that Popper would not agree with his work being used as Werlen intends to use it! Nevertheless, Werlen produced a very densely argued and minutely constructed exploration of Popper's three worlds, and their relevance for social theory. Werlen claims that Popper's philosophy of science starts with an objective theory of knowledge, whereas in fact Popper ended up there, perhaps a little reluctantly. The arguments are – as they always are in social theory – highly abstract, obtuse, and self referential. At the end, they depend on a scholasticist appreciation of fine points of internal coherence, as there are no discernible empirical consequences.

Legacies: Philosophical, Scientific, and Geographical

The lasting impact of a philosophical system depends on the net balance between problems solved and problems created. No philosophical system is impregnable and complete: all have shortcomings and all are susceptible to criticisms, broad or picayune. Larger and more ambitious systems are more open to criticism – according to Popper, that makes them more corroborated when they cannot be conclusively rejected. How, then, does Popper fare?

Scientists rarely take notice of what philosophers of science say. In fact, as John Casti (1989) shows in *Paradigms Lost: Images of Man in the Mirror of Science*, scientists know other scientists and their work, know the work of historians of science, but have never heard of most philosophers of science. But if they have heard of one, that one is Karl Popper. Popper was and continues to be popular among scientists. Many outstanding scientists (e.g., Medawar in biology, Hermann Bondi in physics, Eccles in neurophysiology) thought that Popper gets it more right than anyone else, and they like critical rationalism because it embraces uncertainty, thus providing a place for imagination in science. And Popper guarantees that science can never end: there would always be work to be done.

Medawar believed that Popper was the best philosopher of science there has ever been. Fellow philosophers have been less kind: they point out that his philosophies are rife with contradictions. That is, they fail to cohere perfectly. But this can be said of any and all philosophies. However, professional philosophers had to deal personally with Popper, and this was no easy task. Professional philosophers, as with professional scientists, are the product of the post World War II expansion of universities and the role of the state in higher education. Being a professional philosopher turns rumination into an academic job, typically one concerned with the minutiae of language. Philosopher and professor of philosophy are not necessarily the same thing, in the same way that playwright and critic stand in relationship to one another. Popper asserted that, for philosophy (or for a philosopher) to be primarily concerned with the meanings of words was an enormous error, diverting philosophical activity into an endless self referential cycle, in which triviality and scholasticism eventually prevail. Popper dismissed professional philosophers' concerns with language, with the meaning of words, as an utter waste of time. The only thing that the study of the meaning of words reveals, he said, is the meaning of words.

Professional philosophers' disdain for Popper returned his contempt for them. His personal behavior was so at variance with his proclaimed beliefs as to undermine his own philosophies. His critics were always wrong: he never was. Magee says that pluralism is the essence of liberalism: once disputants repeat themselves they must agree to disagree and mutually respect the right of others to hold different opinions. Popper understood none of this. He thought his way through to a rationally formulated conclusion, an intellectually liberal and right minded outcome, which was correct, so he had the right to impose it by will, never letting up until he got his own way. Popper inveighed against dogmatism, but was pathologically dogmatic. He argued that there was no single correct method in science or elsewhere, that there could be no single correct view, yet he hounded to exhaustion those who would not explicitly concede that he was right. He had to win. He was intolerant. At LSE he was nicknamed 'the illiberal liberal', and it was said by his students that *The Open Society and Its Enemies* should have been called 'The Open Society by One of Its Enemies'.

Popper, and Lakatos, and Feyerabend, both of whom he taught at the LSE, have been called the irrationalist philosophers of science. Popper opened up several Pandora's boxes: once released, his ideas were developed in ways that dismayed him. For example, Feyerabend took his ideas to an extreme anarchist endpoint, arguing that 'anything goes' as far as generating hypotheses is concerned. Feyerabend has often been misunderstood, mischievously perhaps, by those in the constructionist wing of the humanities who wish to reduce science to merely a form of storytelling, as saying that 'anything goes' in 'all' of science, rather than with respect to only hypothesis generation. Popper never accepted that his ideas necessarily must lead to Feyerabend's unbridled liberalism about the origin of theory.

Although Popper eventually conceded that if we want to know what it is that makes a theory scientific then we must look at how it is handled by people, rather than consider 'only' its logical structure, he never believed that knowledge was inherently and unavoidably socially constructed. Popper abhorred those who claimed (1) that science is incapable of establishing truth at all, and (2) that scientists adhere to philosophies (indeed, that they construct philosophies) for social, cultural, and political reasons, rather than rational ones. In contemporary terms, he had no time for post everything extreme relativists. Popper said that such critics resent being regarded as inferior to scientists, so they adopt strategies to reduce science to their own level: they are trying to change their standing in the pecking order of those who have things to say, and if they can't do it within the rules, then they will change the rules, or say that there are no rules at all.

Overall, Popper has not done well in geography, which is really to say that geographers have not allowed ignorance of his work to inhibit their commenting on it. Most research methods texts for students – under and postgraduate – repeat the mantra that 'Popper = falsifiability and only falsifiability', and in so doing, perpetuate a superficial and ill informed portrayal which stops well short of encountering Popper's most radical and most interesting ideas.

Popper is well worth (re)visiting. There is an argument waiting to be made that logical positivism is the last gasp of the philosophy of Newtonian science, whereas critical rationalism is the first deep breath of the philosophy of post Einsteinian science. Popper may just be the first postmodern philosopher of science. Popper would probably have found this outrageous, but if the cap fits....

See also: Philosophy and Human Geography; Positivism/Positivist Geography; Scientific Method; Spatial Science.

Further Reading

Bennett, D. (2005). Replacing positivism in medical geography. *Social Science and Medicine* 60, 2685 2695.

Bird, J. (1993). *The Changing Worlds of Geography: A Critical Guide to Concepts and Methods* (2nd edn.). Oxford: Clarendon Press.

Graham, E. (1997). Philosophies underlying human geography research. In Flowerdew, R. & Martin, D. (eds.) *Methods in Human*

Geography: A Guide for Students Doing a Research Project, pp 6 30. Harlow: Addison Wesley Longman Limited.

Hay, A. (1985). Scientific method in geography. In Johnston, R. J. (ed.) *The Future of Geography*, pp 129 142. London: Methuen.

Hoggart, K., Lees, L. and Davies, A. (2002). *Researching Human Geography*. London: Arnold.

Johnston, R. J. and Sidaway, D. (2004). *Geography and Geographers: Anglo American Human Geography Since 1945*. London: Arnold.

Magee, B. (1973). *Popper*. London: Fontana.

Magee, B. (1997). *Confessions of a Philosopher: A Journey through Western Philosophy* (ch. 3: Logical positivism and its refutation; ch. 11: Getting to know Popper). New York: Random House.

Marshall, J. U. (1985). Geography as a scientific enterprise. In Johnston, R. J. (ed.) *The Future of Geography*, pp 113 128. London: Methuen.

Popper, K. R. (1978). *Three Worlds*. The Tanner Lecture on Human Values delivered at The University of Michigan, April 7, 1978. http://www.tannerlectures.utah.edu/lectures/documents/popper80.pdf (accessed Jul. 2008).

Popper, K. R. (1986). How I see philosophy. In Shanker, S. G. (ed.) *Philosophy in Britain Today*, pp 198 211. London: Croom Helm.

Quinton, A. (1967). Popper, Karl Raimund. In Edwards, P. (ed.) *The Encyclopedia of Philosophy*, pp 398 401. New York: Macmillan Publishing Co., and The Free Press.

Quinton, A. (1982). *Thoughts and Thinkers* (ch. 27: Popper 1: Objective knowledge; ch. 28: Popper 2: Political Theory). London: Duckworth.

Werlen, B. (1993). *Society, Action, and Space: An Alternative Human Geography*. London: Routledge.

Critical Realism/Critical Realist Geographies

Andy C. Pratt, London School of Economics, London, UK

Glossary

Critical Realism A philosophical position that develops sophisticated claims about what is and should be taken as 'real' by the social sciences, advancing arguments about an ontology of the (social) world that regards the 'real' as embracing more than just the 'surface' features of everyday human existence.

Critical Realist Geographies The translation of critical realist ideas into human geography, possessing both a methodological dimension (after Andrew Sayer) and the profound ontological challenges of Roy Bhaskar.

Realism A term covering a range of philosophical positions, but when used without a qualifier it tends to mean a so-called 'naive' belief that what is 'real' is what is immediately present to, and open to the apprehension of, the basic human senses (especially sight and touch).

Realism: A Conceptual Introduction

The aim of this article is to outline the key aspects of the philosophy of realism and its application in the social sciences more generally, and human geography in particular. Realist philosophies go under a number of names: naive, transcendental, and critical. The focus here is on critical realism, one of several versions of realism, and, reflecting geographical practice, it is rooted in Roy Bhaskar's interpretation of realism outlined in both his 1978 book, *A Realist Theory of Science*, and his 1989 book, *The Possibility of Naturalism*. While the concern here is with geographical practice, the aim is to situate this debate in the usage of realism in the wider social sciences. While it may seem that the social sciences are driven by philosophical fads, it is suggested that it might be better to view the social sciences as responding to an ongoing debate about what constitutes knowledge. The first part of this article addresses the three 'cornerstones' of the philosophy of the social sciences that have been challenged by realist scholars: ontology, causality, and abstraction. Bhaskar describes critical realism as "ontologically bold, and epistemologically timid;" in this sense we will deal with epistemology separately, and in so doing open up the broader interpretation that Andrew Sayer's reflections on Bhaskar's work, *Method in Social Science*, written in 1984, did in geography. After a discussion of cognate social science appropriation of critical realism, we return to a central issue: how to 'do' realist research. Finally, we outline some of the points of criticism to which critical realists are open, and why.

Part of the problem with philosophy for the general social scientist, or geographical, reader is that it has developed a very precise language. Where this precise language has been popularized – as it has through (positivist) science – terms are apt to be misused, or their philosophical meanings obscured. Hence, a number of terms in the realist lexicon – above and beyond 'the real' – require careful use; for example, cause, mechanism, and model. Perhaps the best place to begin an elaboration of realism is with Rom Harré's 1986 book, *Varieties of Realism*: an accessible text that admirably achieves what it says in the title, as well as outlining the philosophical assumptions and positions associated with the terms. Put simply, the notion of realism in philosophy ranges on an continuum from a (naive) realism (or empiricism), a sort of 'what you see is what you get', unmediated world that has no 'depth', all the way to the strong version of (transcendental) realism where what is seen (or perceived) is regarded as a result of some other structures of mechanisms underlying (the perhaps obscuring facets of) surface 'reality'. The important point to stress here is that it is not a 'perceptual' difference of opinion being debated, but rather the nature of the 'stuff' of reality. Hence, this is a fundamental point of division: either there is, or there is not, 'more' to that which we 'see' (even after accounting for different perceptions, or means of sensing things: more powerful microscopes, X rays, etc. In short is there anything more even if we can't see it?). Thus, Harré, in his 1972 book, *The Philosophies of Science*, stresses the importance of recognizing the notion, and degree, of ontological depth: ontology being the nature of things, and ontological depth being the assertion that there is more to things than is immediately apparent in observation. The notion of transcendentalism is that the thing remains the same from whichever perspective one apprehends it, whether one closes ones eyes, or even when 'we' are not present. Transcendental realism is thus a way of saying that 'thing ness' is substantial and independent of observation (without denying that observers may apprehend a thing differently; it is they, not it, that are changing).

Ontology

Ontology is the first key dividing line for realists, a dualism upon which all philosophical investigations must be founded and thus will be divided. In this sense, you are either a realist or not. In a slightly different dimension, while Bhaskar initially aligned himself with transcendental realism, in his pioneering book written in 1978, *A Realist Theory of Science*; he subsequently sought to open up a further distinction, in his 1989 book, *The Possibility of Nat uralism*. It is often suggested that rocks and people are different: in terms of their capability to be reflexive, or in terms of their humanity. This distinction is commonly used as a critique of positivist science and to legitimate the addition of an evaluative character, or method, to the social sciences: because the things under investigation are dif ferent. Bhaskar's argument is to turn this argument on its head. He asserts that physical and human objects/subjects are to be considered the same – philosophically speaking – but that the positivist model is flawed. Bhaskar argues that naturalism does hold, 'but only' from a realist perspective. We will return to this issue in the discussion of naturalism, critical realism, and epistemology below. It is this com bination that Bhaskar labels 'critical realism'. It is useful to note that radical naturalism is also found in actor network theory (as an attempt to rebalance the point of reference between people and inanimate objects). However, actor network theorists tend to a 'thin' ontological point of view.

Causality

A second line of distinction, one that is important to critical realists, concerns causality and how the different objects and 'levels' of reality are joined. The dualism drawn here is between successionist notions of causality (if A then B) popularized by Hume: ones that underpin popular notions of positivism and empiricism. Thus, in positivist analyses causality can only be inferred (as there is no 'depth' below the surface, as causality cannot be directly seen it can only be inferred). By extension, if causality cannot be seen it cannot be empirically sensed or measured, thus it does not exist: this is a problem for empiricists. The realist strategy is to substitute instead the notion of generative causality, where real objects interact and under particular conditions react together to produce, or generate, an effect. Realists refer to 'neces sary' conditions (to make a reaction work) and 'contin gent' factors (that may or may not be present: enabling factors). Thus, cause is identifiable and a result of par ticular mechanisms: 'real' mechanisms. However, the simple presence of mechanisms alone does not generate action. This is where context (or space and time) res onates with geographers. As we will see below, this is a very helpful defense against determinism that is leveled at other realist philosophies, in particular structuralism; as well as at empiricism/positivism. We recognize the realist pattern of explanation from experience: the match does not strike due to the introduction of wet or damp; the bread does not rise due to the lack of yeast. The contrast is with successionist causal models, such as that employed in positivism, where these models are simply used as explanatory devices, like algebraic formulas, and do not seek to represent how the world is/things are. This raises another common distinction that critical realists like to make between the concepts of 'models' in positivism and realism: realist models suggest that reality is thus, whereas positivist models are only interested in predicting the right output from a given input, essentially a 'black box' approach. In both senses realist models are more 'real' than positivist ones.

Abstraction

A third leg to the critical realist position is the process of abstraction, that is, what is the nature of the rela tionship between the different 'levels' or concepts of 'reality'. Methodologically, for realists, it is because reality is 'structured' or 'deep' that the selection of re search objects – as represented as taxonomies, natural orders – is a critical starting point. Nonrealist taxo nomies may be legitimately selected on the basis of similarity of color, shape, pattern, etc. For the realist, a coherent research object should be the sum of a com plete causal mechanism and its effects. Some realists have referred to this process as 'rational' abstraction, and counterpoised it to 'chaotic' conceptions. This is perhaps unfortunate, as it risks adding a level of judgmentalism into the issue, as well as to invite confusion over the notion of 'rational' action (or worse still, rational choice). So, abstraction identifies the appropriate mechanisms and the conditions under which a cause may produce an effect. Rational abstractions include only 'necessary re lationships'. Realists have stressed the importance of necessary and contingent relationships to make such distinctions. Once identified, such mechanisms will dis close the relevant set of objects to be investigated. It might look like deducing the world from a model, but realists argue that the whole process involves what is termed 'retroduction' (as opposed to induction or de duction): that is, the movement between the real object, the necessary and contingent conditions, a rational ab straction, and the specification of causal mechanisms. For realists, this is an iterative and reflexive methodological process. Finally, the notion of emergence is used to ar ticulate the unique outcomes of processes thus identi fied, under specific conditions.

Epistemology

We have now introduced the three main tenets of the ontologically 'bold' aspect of critical realism. Next we come to Bhaskar's so called 'timid' epistemology.

Epistemology is the logic by which we decide how knowledge may be added to, or rejected as 'false'. One popular approach is that of positivism, which, rather arrogantly, has adopted the mantle of 'the' scientific method. Critical realism represents a full frontal rebuttal of positivism, while maintaining that the project of science 'is' possible. As Bhaskar notes in his 1989 book, *The Possibility of Naturalism*, a common method to investigate both humans and objects is possible and desirable (as opposed to creating different types of science for physical and human processes). In this way Bhaskar seeks to 'wrong foot' positivism and traditional science by arguing for the position of realism as the appropriate ontology for the social sciences, one that may be replicated in the physical sciences. Bhaskar also claims that positivists are 'really' realists. He expresses this through an account of the 'diurnal' notion of scientific endeavor; that in practice, scientists (by day) do, and think, 'realism', although they report their experimental practice (at night) as positivism. Clearly, this is contradictory and logically invalid. In this sense Bhaskar's (critical) realism can be articulated to the social studies of science tradition, and to hermeneutics. Moreover, critical realists are scathing of positivists who champion the deductive purity of their logic, by pointing out the deceit of inference involved in accounts of cause.

Bhaskar introduces his notion of epistemology, a theory of knowledge of how we narrate change, confrontation, and the 'development' of ideas and knowledge, as the transformational model of social activity (the TMSA). This step, seldom referred to by geographers, is a logical continuation of naturalism: namely, the model of science is one based on the 'social'; and therefore it needs to incorporate a model of social change as well. This is his version of a resolution of the structure–agency model, but one rooted in a realist ontology. So, the TMSA seeks to account for both the vertical and horizontal logics: the sequential, if A then B; and also the possibility that Z is the surface effect of A, that deeper cause. Other versions of the structure–agency resolution, notably Anthony Giddens's 1984 summary of his structuration theory, *The Constitution of Society*, are philosophically noncommittal on their ontological position. A subsequent development of this question, one that Bhaskar does not explore in huge detail, is carried forward in Margaret Archer's work such as her 1988 book, *Culture and Agency*, and her 1995 book, *Realist Social Theory* . Archer deals with these epistemological and ontological debates in more depth, and she uses the term 'morphogenesis' to signal her resolution of structure and agency, and ontology. Bhaskar's promotion of the TMSA is 'timid' because it is meant to be critical (reflexive). From Bhaskar's point of view, it should be possible to be critical of the epistemological stance taken, as well as the findings, and hence to develop more satisfactory accounts of the world.

Critical Realism in Human Geography and Cognate Social Sciences

Critical realism has had a significant impact in the discipline of geography, and in particular it offered a resolution of the disenchantment with positivism, especially – although not exclusively – within human geography, heralding an approach that, it is claimed, avoids the perceived shortcomings of subjectivism (as in the humanistic or phenomenological geography advance from positivism). However, as we have already noted, realism has a prehistory linked to older philosophical notions of realism, and this prehistory can be partially mapped back into geography. The position of realism vis à vis positivism is fairly clear, as is its position against subjectivism, although in the latter case there have been claims for realist phenomenology. Again, phenomenology has often been used as a possible solution to the subjective–objective/structure–agency debate, notably by Derek Gregory writing in 1978. However, common usages of phenomenology are not philosophically realist and, strictly speaking, also stray from Edmund Husserl's original interpretations of phenomenology itself.

Realism (re-)entered the geographical lexicon first via Gregory's pioneering 1978 book, *Ideology, Science and Human Geography*, and in his later discussions, from the mid 1980s, of the nature of practices, spaces and places, and critical hermeneutics. Later, Sayer's writings, leading from his crucial 1984 book, *Method in Social Science*, became a common point of departure: especially in economic geography, perhaps because they were offered as 'method'. By contrast, Gregory's contribution emerges from ongoing debates within historical geography where philosophical and theoretical discussion was more nuanced. However, the two threads came together in the 'localities' work of the late 1980s. Arguably, in the process, the empirical focus of work on localities became quickly detached from philosophical debates and – arguably – this productive thread was broken.

In recent years interest in realism has flowed from another source of disenchantment, this time with the structuralism articulated in much Marxist thought from the 1970s and 1980s. Advocates sought to develop a rigorous method for historical–materialist approaches (as well as tempering the determinism of structuralism, particularly Althusserian stucturalism: a previous attempt to develop a 'scientific' basis of Marxism), and without a reduction to historical materialism (it is claimed by critical realists that realism was historical materialism's method; however, it is worth stressing that, logically, realists are not necessarily historical materialists if in practice many lean in this direction). Critical realism's impact was felt most directly in economic geography, through the work of Sayer *et al.*, as Andy Pratt notes in his 2004 essay on Sayer; although it was also deployed in urban and social geography as well. In part, its popularity was due to the increasing wariness of economic geographers at what were perceived as the excesses of Marxist approaches; most keenly felt in geography as they often denied the importance of space. Thus, arguably, realist notions of space as 'effect', coupled with contingency, provided a useful template for geographers. In fact, the argument seemed so persuasive that (geographical) commentators such as Kevin Archer, writing in 1987, suggested that "we're all realist now." As with many philosophies, the core protagonists saw such superficial acceptance of realism as a call for business as usual rather than a real change (this was perhaps facilitated by the superficial similarity and familiarity of terms and language deployed in critical realism).

Critical realism was also influential in sociology, initially through the work of Russell Keat and John Urry in their 1975 book, *Social Theory as Science*, and later via their colleagues at Lancaster University, and most notably through the Economic and Social Research Council (ESRC) funded 'localities' projects: research projects that created both a strong common interest between sociologists and geographers, and a significant use of realism in practical research. In many ways this accomplishment built upon a theoretical body of realist sociological work that had been developed previously at Leicester University, notably by Derek Layder in his 1990 book, *The Realist Image in Social Science*, and his 1993 book, *New Strategies in Social Research*, although this work is seldom referred to by geographers. On one hand, realist research was empirical (and theoretical): what Sayer termed 'extensive' and 'intensive'; however, often it was read as an excuse for empiricism *tout court*. On the other hand, in their localities, work at least realists were labeled as postmodernist and post structuralist. Realists arguably 'are' post structuralists; however, most versions of post structuralism have been driven by a relativism which is avowedly anti ontological. Some have argued that realist post structuralism and constructivism are perfectly compatible: that 'truth' can be known, but is always known under contingent conditions.

Realism perhaps retains its strongest foothold in economics where, especially through Tony Lawson's 2003 book, *Reorientating Economics*, the Cambridge economics group have developed a strong critique of mainstream neoclassical economics. Their critique is an avowedly philosophical one. They criticize economics, and its obsession with mathematical modeling, as misplaced; in a rather pointed judgmental critique, they describe the contemporary discipline of economics as 'autistic'. The characterization is intended, problematically perhaps, as a derogative term, as if economists see the world as do sufferers of Asperger's syndrome with both inordinate attention to detail (in this case specifically to the mathematical abstractions or models) and poor cognizance of the 'bigger picture', indeed of social relations in general. It is claimed that a critical realist informed economics can provide a methodological foundation for a 'Post Autustic Economics' (PAE); one that attends to 'real' practices rather than idealized and abstracted models expressed in the apparent precision of a language of algebra.

Critical Realist Method

Given critical realism's immanent critique of positivism, and its attention to the real, it is unsurprising that expectations may be raised about 'realist methods'. Some geographers have sought to reflect upon the application of realism. What realists reject are 'cookbook' approaches implying that methods are intrinsically realist or not, and supposing that the same methods can be applied in all circumstances; rather, the argument is that appropriate methods be used to investigate particular causal mechanisms. Perhaps the simplest statement is from Sayer who argues for the combination of 'intensive' and 'extensive' methods. 'Extensive' methods are usually glossed as sample data collections from whole populations: these allegedly show generalized outcomes. 'Intensive' methods, using fine grained qualitative approaches, are usually advocated for examining causal processes. The argument is that intensive and extensive approaches should be used to complement one another. In practice, 'extensive' has been taken to mean quantitative methods, and 'intensive', qualitative methods. In so doing, realism risks being stripped of its philosophical baggage and simply presented as the advocacy of 'common sense'. Moreover, the analysis of secondary data with descriptive statistics potentially contradicts the realist claim to examine 'real structures or mechanisms'. By definition, most secondary statistics are derived from positivism's models and are – according to realists – based upon chaotic conceptions. Accordingly, such data, let alone the accompanying analyses, are unlikely to be helpful.

However, it should be noted that the selection of what constitute the objects of analysis are, for realists, rooted in the process of abstraction, and most definitely not defined *a priori* or based on simple 'surface' regularities. Thus, it is conceptualization and abstraction that mark a key differentiator rather than the method *per se*. Both Andy Pratt and Henry Yeung, in their 1995 and 1997 contributions, offer nuanced reflections on the use of methods, as well as on the realist mode of research. They point to the iterative and heuristic process, and to the challenge of 'proof' (which realists refer to as 'practical adequacy'), as opposed to the normative linear exploration in the vein of hypothesis, data, and proof. In a decidedly 'post structuralist' mode, some realists point to the importance of narrative in scientific explanation that reflects the process of investigation. As we have noted above, realists claim that positivists may 'do realism' (i.e., they assume causal relations), but report in the guise of positivism by discussing correlation. In this way, positivists commit the error of eliding correlation and causation.

Critical Realism's Challenges

Initial attempts to develop critical realism came from feminist writers who sought to develop a nuanced critique of patriarchy by distinguising necessary and contingent relations, notably in the 1986 paper by Jo Foord and Nicky Gregson. The fruits of this work, and the legacy of the 'localities' debate, are to be found in the work of Simon Duncan and his coauthors from the late 1980s to the late 1990s. The significant challenges to realism have seemed to come from post structuralism; a philosophical debate that raged on the heels of realism (e.g., actor network theory or nonrepresentational theory). At first sight, realism is the opposite of post structuralism, given, in particular, the disavowal of the notion of a single unitary Truth in the latter. It can be argued that the picture is not so black and white, however, in that some realists argue that, while Truth may exist, the problem is how we know it or represent it. In practice, we cannot see Truth or real objects outside of representations. Post structuralists and constructivists begin at this point, arguing that we can never know the Truth (although this is often elided into the claim that there is 'no' Truth). Generally, post structuralists are silent on ontology, suggesting either that we do not need to answer such a question or that there is nothing but 'surface'.

Lest such distinctions seem too obscure, we may take the problem of dealing with texts. We need to ask if the text is 'real' or it is itself a 'representation' or 'effect' of some process. The former pushes one to view words how we commonly view (but philosophically erroneously) numbers, as if they were absolute and objective. The latter moves one to recall Humpty Dumpty who says (apparently in a conventional positivist mode, underlying the sovereignty of the 'author') in Lewis Carroll's *Through the Looking Glass*: "When *I* use a word … it means just what I choose it to mean, neither more nor less." To which Alice replies "The question is … whether you *can* make words mean so many different things." Humpty Dumpty's (seemingly, post structuralist) retort is "The question is … which is to be master that's all." In the current terminology, we would say that language is performative and context dependent. The question of meaning and truth is thus slippery: whom and how do we determine who is to be master: who's interpretation is final? The nub of the point – if this is seen as a dualism – is the ontological assumptions and the method that acts upon them. Of course, the point where this article opened was that these questions are the object of ongoing philosophical debate and social–theoretic critique. It may be difficult, and messy: on one hand, there are many overlaps between post structuralism and critical realism; on the other hand, one can say the same for post structualism and empiricism. The twist is: what role is assigned to ontology? In this sense, Bhaskar's account of critical realism as "ontologically bold, but epistemologically timid" should not be read as apologetic, but rather as a principled defence of a reflexive and critical social science: in this sense, a 'critical' realism worthy of its name.

See also: Actor-Network Theory/Network Geographies; Humanism/Humanistic Geography; Intensive/Extensive Research; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Phenomenology/Phenomenological Geography; Positivism/Positivist Geography; Poststructuralism/Poststructuralist Geographies; Structuration Theory.

Further Reading

Archer, K. (1987). Mythology and the problem of reading in urban and regional research. *Environment & Planning D: Society and Space* 5, 384 393.

Archer, M. S. (1988). *Culture and Agency: The Place of Culture in Social Theory*. Cambridge: Cambridge University Press.

Archer, M. S. (1995). *Realist Social Theory: The Morphogenetic Approach*. Cambridge: Cambridge University Press.

Bhaskar, R. (1978). *A Realist Theory of Science*. Hassocks: Harvester Press.

Bhaskar, R. (1989). *The Possibility of Naturalism: A Philosophical Critique of the Contemporary Human Sciences*. New York: Harvester Wheatsheaf.

Cloke, P., Philo, C. and Sadler, D. (1991). *Approaching Human Geography: An Introduction to Contemporary Theoretical Debates*. London: Paul Chapman.

Duncan, S. and Edwards, R. (1999). *Lone Mothers, Paid Work, and Gendered Moral Rationalities*. Basingstoke: Macmillan.

Duncan, S. and Goodwin, M. (1988). *The Local State and Uneven Development: Behind the Local Government Crisis*. Cambridge: Polity.

Foord, J. and Gregson, N. (1986). Patriarchy: Towards a reconceptualisation. *Antipode* 18, 186 211.

Gibson, E. M. W. (1981). Realism. In Harvey, M. E. & Holly, B. P. (eds.) *Themes in Geographic Thought*, pp 148 162. London: Croom Helm.

Giddens, A. (1984). *The Constitution of Society: Outline of the Theory of Structuration*. Cambridge: Polity Press.

Gregory, D. (1978). *Ideology, Science and Human Geography*. London: Hutchinson.

Gregory, D. (1985). People, places and practices: The future of human geography. In King, R. (ed.) *Geographical Futures*, pp 56 76. Sheffield: The Geographical Association.

Gregory, D. (1985). Suspended animation: The stasis of diffusion theory. In Gregory, D. & Urry, J. (eds.) *Social Relations and Spatial Structures*, pp 296 336. London: Macmillan.

Harré, R. (1972). *The Philosophies of Science: An Introductory Survey*. London: Oxford University Press.

Harré, R. (1986). *Varieties of Realism: A Rationale for the Natural Sciences*. Oxford: Basil Blackwell.

Keat, R. and Urry, J. (1975). *Social Theory as Science*. London: Routledge and Kegan Paul.

Lawson, T. (2003). *Reorienting Economics*. London: Routledge.
Layder, D. (1990). *The Realist Image in Social Science*. Basingstoke: Macmillan.
Layder, D. (1993). *New Strategies in Social Research: An Introduction and Guide*. Cambridge: Polity.
Pratt, A. C. (1995). Putting critical realism to work the practical implications for geographical research. *Progress in Human Geography* 19, 61 74.
Pratt, A. C. (2004). Andrew Sayer. In Hubbard, P., Kitchin, R. & Valentine, G. (eds.) *Key Thinkers on Space and Place*, pp 245 250. London: Sage.
Relph, E. (1989). A curiously unbalanced condition of the powers of the mind: Realism and the ecology of environmental experience. In Boal, F. & Livingstone, D. (eds.) *The Behavioural Environment*, pp 277 288. London: Routledge.
Sayer, R. A. (1984). *Method in Social Science: A Realist Approach*. London: Hutchinson.
Sayer, R. A. (2000). *Realism and Social Science*. London: Sage.
Yeung, H. W. C. (1997). Critical realism and realist research in human geography: A method or a philosophy in search of a method? *Progress in Human Geography* 21, 51 74.

Relevant Websites

http://www.econ.cam.ac.uk/seminars/realist/index.htm
Cambridge Realist Workshop.
http://www.criticalrealism.demon.co.uk/
Centre for Critical Realism (CCR).
http://www.raggedclaws.com/criticalrealism/
Critical Realism.
http://criticalrealism.wikispaces.com/
International association for critical realism (IACR).
http://www.equinoxjournals.com/ojs/index.php/JCR
Journal of Critical Realism.

Critical Theory (After Habermas)

M. Phillips, University of Leicester, Leicester, UK

Glossary

Communicative Action A term central to Habermas' theory of action, but used in at least two different ways. First, and its most restrictive sense, Habermas uses it to refer to actions that incorporate consensual motivations related to the intersubjective recognition of validity claims. Second, and more generally, Habermas uses the term, along with social interaction, to refer to intersubjective interaction.

Critical Theory An attempt to understand the world in such a manner that enables people to become more emancipated.

Lifeworld A concept coined originally by phenomenologists such as Edmund Husserl and Alfred Schütz to refer to the world as understood by participants in that world. For Habermas the lifeworld is intersubjectively established, forms an arena of communicative action, and involves cultural transmission, social integration, and personal identity formation.

Public Intellectual A term which has been used to refer to academics who write for a public as opposed to an academic or policy-related audience.

Public Sphere A term that is often used in two ways. First, to refer to arenas of institutionalized political power associated with a sovereign state. Second, the public sphere is used to refer to where things are accessible or visible to others. Habermas used both senses in his discussion of the public sphere.

Purposive–Rational Action Action focused on the means to pursue specific ends. Elements of the world are encountered as objects that can, in principle, be manipulated and controlled. In Habermas' work, this form of action is often differentiated into instrumental and strategic action and is contrasted with communicative action in which there is a focus on the formulation of ends and where elements of the world are encountered as speaking and acting subjects.

Reconstructive Analysis A form of analysis which aims to make explicit practically mastered but untheorized knowledges. Differs from empirical–analytical analyses in that it does not see itself producing new or better knowledge but rather as simply clarifying what is already practically mastered. These reconstructions are also seen to differ from hermeneutic understanding in that their goal is not to paraphrase or translate meanings, but rather to explicate what is known, albeit un- or semi-consciously, by people and used in everyday life.

Steering Media A phrase used by Habermas to refer to mechanisms for integrating action between people which are detached from processes of mutual understanding and communicative action and thereby involve purposive–rational action. Habermas identifies money and power as forms of key steering media.

Universal/Formal Pragmatics These terms were coined by Habermas to describe his analysis of language and specifically his claim that reconstructive analysis of speech can reveal certain conditions which are universally present or implied in the act of speaking. In particular Habermas argues that speech relies of the positing, and potential discursive redemption, of one or more validity claims.

Introducing Habermas, Critical Theory, and the Frankfurt School

Jürgen Habermas is a German philosopher and social theorist/analyst who, in part because he worked at and with members of the Institute for Social Research in the Johann Wolfgang Goethe University in Frankfurt, is seen to be a member of the so called Frankfurt School of critical theory established in the 1930s. As is the case with many designated schools of thought, there is considerable debate as to the value of title, the names of the people to be associated with it, and the criteria for inclusion and exclusion. This is despite Max Horkheimer, the second director of the Institute, providing formal outlines of the nature of critical theory and his hopes for the Institute. As outlined in Held's *Introduction to Critical Theory*, Hor kheimer highlighted three key concerns for members of the Institute: (1) the development of research which examined fundamental philosophical questions through sophisticated research methods; (2) a reconstruction of Marxism so as to enable it more fully to achieve its ob jectives; and (3) the production of social theory capable of explicating the mediation of economic, psychological, and cultural facets of social life. Horkheimer's de scription, while not characterizing the work of everyone subsequently associated with the Frankfurt School of critical theory, does illustrate something of its profoundly ambitious nature whereby many of its members sought to combine philosophy, social theory, and in depth empir ical research in programs of work which drew together elements from Marxist political economy, Freudian psychoanalysis, contemporary and historical cultural analysis of both high and mass cultural forms, all with the practical intent of criticizing and subverting social domination and exploitation. This constellation of con cerns and interests show clear parallels with many of those pursued in contemporary critical human geog raphy, although often pursued in less synthesized man ners to that envisaged by many of those associated with the Frankfurt School.

The focus of this article will be primarily on the work of Habermas, detailing how his ideas came to be drawn on by geographers, as well as outlining elements of his work which have been rather ignored within geography although being arguably of clear relevance to issues and debates pursued within contemporary human geography. Given the parallels between Horkeimer's ambitions for critical theory and motifs of contemporary human geography, many similar and other lines of argument could be advanced concerning the geographical value of other work associated with Frankfurt School. In addition to Habermas and Horkheimer, other people associated with the Institute of Social Research who might be re garded as key participants in the Frankfurt School of critical theory include Theodor Adorno, Herbert Mar cuse, and more recently, Klaus Offer, Albrecht Wellmer, and Axel Honneth.

As demonstrated in **Figure 1**, there has been con siderable variation in the duration of these peoples' in volvement with the Institute, which itself migrated away from Frankfurt for a period as a consequence of the rise of Nazism in Germany, spending time based in Geneva (1933–35) before moving to the USA, where it was for mally attached to the Columbia University, New York (1935–46), as well as being informally run from locations in California (1941–50) before formally being reconsti tuted in Frankfurt in 1950. While Horkeimer and Adorno were the leading figures associated with the Institute to return to Frankfurt, Marcuse among others remained in the USA. Subsequent to its return to Frankfurt, a series of new members of the Institute were recruited, including Habermas, and somewhat later, Albrecht Wellmer and Klaus Offe, who are sometimes characterized as a 'second generation' of the Frankfurt School, although much of the work associated with this group was actually con ducted while Habermas was working at the Max Planck Institute near Munich. Habermas did, however, return to Frankfurt in 1983 where he appointed Axel Honneth, who has been a student of his in Munich, to the position of assistant professor. Alex Honneth was subsequently to take on Directorship of the Institute of Social Research, and has been seen as a leading figure of a third generation of the Frankfurt School of critical theory.

In addition to people directly associated with the In stitute of Social Research, albeit in some cases for only a short period, many other prominent social theorists and philosophers – including Walter Benjamin, Hannah Arendt, Richard Bernstein, Nancy Fraser, Thomas McCarthy, and Selya Benhabib – have been influenced by, and worked with, members of the Institute, and hence might be seen to be part of the Frankfurt School and its various generational forms, although never formally part of the Institute. The influence and significance of the Frankfurt School is hence wide ranging and diffuse, and clearly extends beyond the principal focus of this article, the work of Habermas.

Habermas himself originally joined the Institute in 1956, although remaining there for only 3 years during which time he principally worked with Adorno on a study of student protest. His work was, however, seen by Horkheimer as being potentially politically damaging to the Institute and Habermas left for the University of Marburg to write what was to become the first of many major studies (see **Table 1**) which led to him being seen as the leading figure of the second generation of the Frankfurt School of critical theory. In 1964 Habermas returned to Frankfurt as Professor of Philosophy and Sociology, a position previously held by Horkheimer, and he quickly sought to connect his work with the vision of critical theory outlined by his predecessor and critic. From the 1970s many of Habermas' works started to be translated into English and were soon cited in Anglo phone geography. Since this time Habermas has con tinued to produce a series of major works addressing a wide range of social, political, and moral issues while continuing to champion the concept of theory with critical intent. Although references to Habermas have continued to circulate within geography, mentions of his notion of critical theory have become much less preva lent in recent years. This is not to say that his writings have little contemporary relevance to geography: as will be shown, across a range of contemporary studies in human geography there is still clear recognition of at least some of the concepts and concerns which exercised Habermas from his most early work to his most recent. There is also, it will be suggested, some connections between geography and Habermas' more recent studies which have as yet to be charted.

Habermas and Geography: Initial Engagements

It was in the late 1970s that a series of Anglophone geographers began to cite the work of Habermas. This was a period of philosophical awakening and turmoil within human geography, with geographers both ques tioning adherence to scientistic philosophies such as lo gical positivism and advancing a range of alternative positions. Many of the earliest references to Habermas were linked to these dual endeavors. Brief references to Habermas, and indeed also to Horkheimer and Marcuse, for instance, appear scatted across Ley and Samuels'

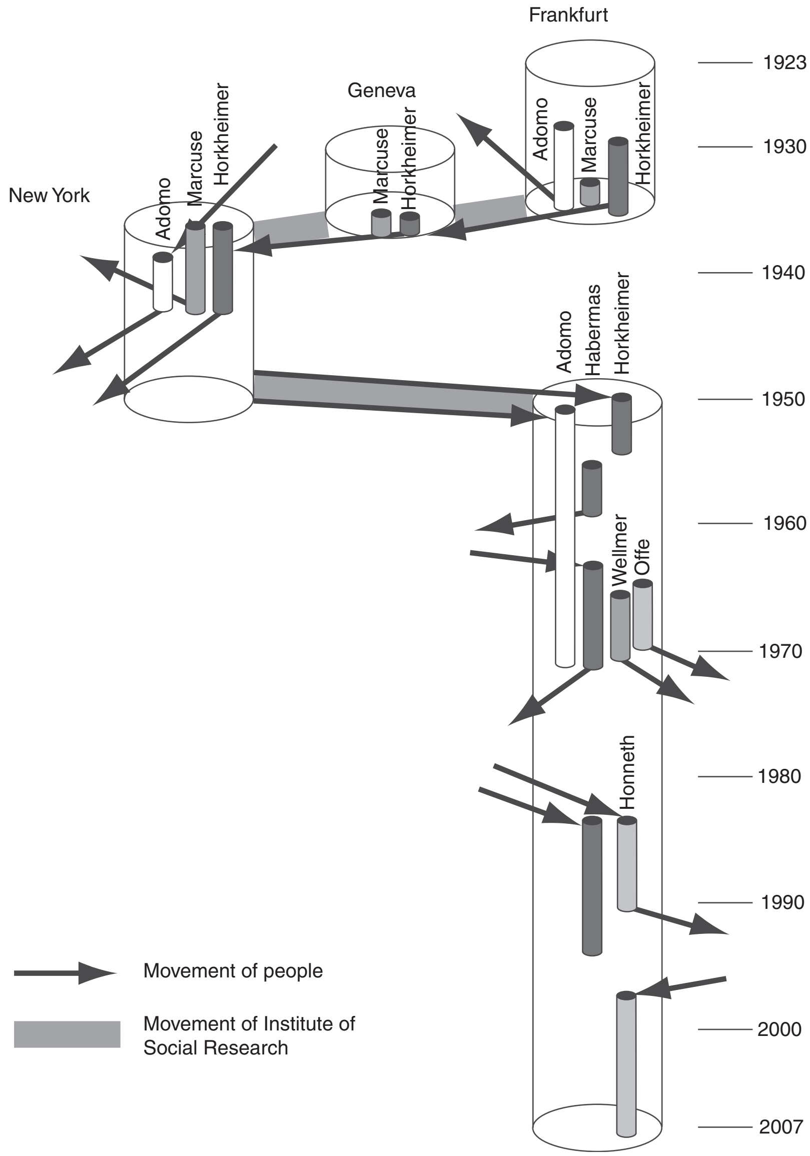

Figure 1 Time space paths of the Institute of Social Research and its advocates of critical theory.

Humanistic Geography: Prospects and Problems, principally directed at criticizing the self understanding of positivistic spatial science, a use to which Habermas' work is also put in Dietrich Bartels' discussion of theory in geography. Bartels actually draws quite extensively on Habermas' *Knowledge and Human Interest* to identify different rationalities involved in both positivist spatial science and earlier regional/landscape studies, suggesting that the former employed an instrumental rationality focused on technical manipulation while the latter employed an interpretative rationality focused on correspondence to the values, beliefs, and opinions present in a situation under examination. Bartels suggests that both rationalities are partial and need to be superseded by second and third 'waves of rationality' which, respectively, identify the underlining assumptions of the first wave rationalities and then account for their formation. Bartels argues that geography was indeed in the process of moving through these waves of rationality, although there is no argument presented as to the dynamics of such movement.

Gregory's *Ideology, Science and Human Geography* provided a more sustained exploration of Habermas' work within geography, drawing on this not simply to critique positivistic–spatial science but also as a means to justify and develop a different form of human geography, namely critical human geography. In advancing these arguments, Gregory both more fully elaborated on Habermas' arguments concerning the limitations of scientistic approaches and highlighted the relevance of

Table 1 The major works of Jürgen Habermas translated into English by date of original publication (date of first translated edition in brackets)

1962	The structural transformation of the public sphere (1989)
1963	Theory and practice (1974)
1967	On the logic of the social sciences (1988)
	Towards a rational society (1971)
1968	Knowledge and human interests (1971)
1973	Legitimation crisis (1976)
1975	On the pragmatics of social interaction (2001)
1976	Communication and the evolution of society (1979)
1981	Theory of communicative action, volume 1 (1984)
	Theory of communicative action, volume 2 (1987)
1983	Philosophical-political profiles (1983)
	Moral consciousness and communicative action (1990)
1985	The philosophical discourse of modernity (1987)
	The new conservatism (1999)
1988	Postmetaphysical thinking (1992)
1991	Justification and application (1994)
1992	Between facts and norms (1996)
	On the pragmatics of communication (2000)
1996	The inclusion of the other (1998)
1997	The liberating power of symbols
1998	The postnational constellation (2001)
1998	Religion and rationality (2002)
2001	The future of human nature (2003)
	Time of transitions (2006)
2004	The divided West (2006)

further aspects of his work in understanding the value and limitations of other philosophical and methodological approaches being debated within geography in the 1970s, such as humanism, structuralism, and various self styled radical approaches such as Marxism.

One line of argument advanced by Gregory was for the development of a more radical approach to geography which viewed the concept of detached objectivity so valued by positivistic and scientistic approaches as both undesirable and unattainable. Knowledge, it was suggested, was both value laden and had important social consequences, and as a result was clearly connected to the exercise of power and should be seen as inherently political. Radical geography was, hence, concerned to politicize the epistemology of geography, to make it as Gregory argued, 'a committed enterprise'. It was in pursuit of this task that Gregory in particular turned to the work of Habermas and, to a lesser extent, some other members of the Frankfurt School of critical theory.

Habermas was a key contributor, along with Marcuse, in the so called 'positivist dispute' within German sociology instigated by Adorno in 1961. Geographers remained largely immunized from this debate, which was effectively an argument between advocates of the critical rationalism espoused by people such as Karl Popper, and the ideas of critical theory being developed by members of the Frankfurt School. Gregory, however, highlighted the relevance of this debate to attempts to critique positivist approaches in geography, as well as looking to arguments that Habermas developed later in works such as *Towards a Rational Society, Theory and Practice*, and *On the Logic of the Social Sciences*. These studies suggested that the supposed value neutrality of positivistic approaches concealed a commitment to instrumental – means and effort focused – control, a point clearly, and to geographers relevantly, evidenced by systems theory, an issue that Gregory explored in more depth in an article entitled 'The ideology of control'. Habermas further argued that this instrumentalization was transforming politics, which had changed from being intimately connected to notions of what is desirable and right to issues of what is merely possible. Habermas sought to resist this form of rationalization, arguing for recognition of the need to link theory and practical knowledge into political processes of public will formation orientated toward achievement of justice and a good life.

As well as using these ideas to critique the supposed value neutrality of knowledge, Gregory, and also Lewis and Melville, used the work of Habermas, and particularly the concept of knowledge constitutive interests propounded in *Knowledge and Human Interests*, to suggest that a particular value orientation, namely emancipation understood as freedom from avoidable constraints, should be an explicitly acknowledged value orientation in geographical studies. This was based on an argument that knowledge involves three quasi transcendental cognitive interests: the technical, the practical, and the emancipatory (see **Table 2**). Habermas argued that knowledge acquisition was inherently linked to one or more of these interests which were rooted in three distinct but universal dimensions of human existence: work, interaction, and power. He further suggested that these interests served as the basis for three different types of knowledge, each with its own distinctive object domain, methodology, and purpose or aim.

This theory of knowledge not only emphasized its social rootedness, thereby debunking any notion of social detachment which characterized the scientism of positivist approaches, but also resonated well with the lines of philosophical debate raging in geography in the 1970s. So, not only could positivist spatial science be equated with empirical–analytical science orientated to a technical interest in expanding social control over nature and people, but the counter reactions of humanistic and radical geography could be mapped onto the historical–hermeneutic/practical and critical/emancipatory as well. Habermas seemed to provide both an interpretation of the knowledge terrain of geography, and also a justification for the adoption of a radical/committed perspective in that critical knowledge was seen to both incorporate the other forms of knowledge and transcend their problematic features. The phrase critical could, and indeed sometimes was, viewed as purely epistemological, involving identification of the failings of other

Table 2 Habermas' concept of knowledge constitutive human interests

Quasi-transcendental human interest	*Type of knowledge*	*Object domain*	*Methodology*	*Aim*	*Dimension of social existence*
Technical	Empirical analytical	Object world	Experiment	Successful explanation	Work
Practical	Historical hermeneutic	Subject world	Empathy	Successful interpretation	Symbolic interaction
Power	Emancipation	Ideological world	Reflection	Successful emancipation	Power

approaches, although this was far from the implications of Habermas' own writings which repeatedly emphasized that critical theory was theory with practical social intent: it was theory orientated toward emancipating people from constraints.

This conception of critical theory chimes closely with Marxist claims that the purpose of theory is not simply to interpret the world but to change it, and indeed the critical theory of the Frankfurt School and Habermas explicitly acknowledge connections to Marxism. Habermas' notion of critical theory shows clear parallels with the concept of radical theory as outlined in David Harvey's Marx inspired, *Social Justice and the City*, although this book contains no reference to Habermas or the work of other members of the Frankfurt School, save for a few critical references to Marcuse's *One Dimensional Man*. The writings of Habermas and other members of the Frankfurt School on the other hand repeatedly demonstrate an association with Marxism, although they draw upon a wide range of other sources and also develop a series of criticisms and reconstructions of Marxism. Critical theory can hence be seen to have an uneasy relationship with Marxism, a feature that meant it defied easy emplacement in the stark lines of philosophical and theoretical debate that characterized geography in the 1970s and 1980s.

On the one hand, Habermas' work not only seemed to provide epistemological justification for radical geography but also showed strong parallels with the ontological arguments being developed by Marxist geographers. In particular Habermas engaged with the system theories of Luhmann and Parsons, developing conceptions of social systems that showed clear parallels with the structural Marxist formulations associated with Althusser, Castells, and Poulantzas. A key parallel was the desire to resist the reduction of social life to an economic base which was prevalent within many strands of Marxism and a source of criticism both within and beyond geography. While much of the geographical criticism in the 1970s and 1980s was undertaken from positions constructed as being antithetical to Marxism, such as humanistic geography, the criticisms of Habermas, and indeed those seen as structural Marxists, were conducted from positions of explicit allegiance to Marxism. Habermas' turn to systems theory and the elaboration of economic, political, and socio cultural systems and crisis tendencies as outlined, for instance, in *Legitimation Crisis*, can be seen as quite in tune with the vision of critical theory outlined by Horkheimer, involving a reconstruction of Marxism to enable consideration of the psychological and cultural, as well as political–economic, facets of social life.

On the other hand, Habermas' embrace of Marxism and systems theory was neither full nor uncritical. He, for example, clearly identified the failures of scientism in the epistemologies of both Marxism and systems theory, and also suggested that both their ontologies needed reformulation. In the former case this was to avoid overemphasis on the instrumental action of labor and the transformative agency of the proletariat, and in the latter case to incorporate recognition that notions of structure should be melded with those of agency. From his early works, Habermas was concerned to highlight that social life was irreducible to systemic demands but always required consideration of the self understanding of people that were constituted, he argued, through communicative interaction. As will be outlined in the next section, Habermas came to develop these ideas into a series of quite specific conceptual arguments, although they can also be seen to connect into concepts and debates routinely employed within human geography in the late 1970s and 1980s, albeit often with few direct references to the work of Habermas. So, for instance, throughout much of his work Habermas has drawn extensively on the concept of lifeworld, offering critiques of its conceptualization by phenomenologists such as Edmund Husserl and Alfred Schütz, while also continuing to maintain its significance. Proponents of humanistic geography also emphasized the significance of the concept, although in these cases associations were drawn primarily to its phenomenological conceptualizations. Phenomenology and the concept of lifeworld themselves became subjects of geographical debate in the late 1970s and early 1980s, often centered on the significance of human agency and social structure. Some reference in this context was indeed made to the work of Habermas and his conceptualizations of system and lifeworld, although these were often situated alongside, and given less emphasis, than ideas derived from the work of social theorists such as Giddens, Bhaskar, and Bourdieu, despite the former making frequent reference to the work of Habermas and his concept of critical theory.

Studies that drew more directly and extensively on the work of Habermas often involved theorizations of the state and drew on Habermas' *Legitimation Crisis*, as with Clark and Dear's *State Apparatus*: Structures and Language of Legitimacy, Short's *The Urban Arena*, King's studies of 'Capital switching and the role of ground rent', and in a very different spatial context, Pile's *The Private Farmer: Transformation and Legitimation in Advanced Capitalist Agriculture*. Miller has also addressed issues of state power and contestation using Habermas, although the focus in the *Geography and Social Movements: Comparing Antinuclear Activism in the Boston Area* is on the system/lifeworld distinction as elaborated in Habermas' *Theory of Communicative Action*. In a quite different vein of work, Gunnar Olsson also drew on the work of Habermas, as well as that of Adorno, in his individual transition from proponent to critique of positivistic spatial science which focused on the power and limits of language, concerns and sources which were also utilized in Pile's elaboration of interpretative approaches to human geography. Habermas' work was also drawn into discussions of time geography and the historical geographies of the Industrial Revolution and modernity, which in turn were often framed through the contours of the agency–structure debate of the 1980s.

From the early 1990s, however, geographers also began to pay increasing attention to Habermas' early study of the public sphere, which he defined as an arena of social life whereby people who were formally unrelated to each other could come together in order to enact procedures of public will formation through unconstrained argumentation.

Initially published in 1962, *The Structural Transformation of the Public Sphere* was only translated into English in 1989, although it had been the subject of some Anglophone commentary and critique before then. In the 1990s, however, this discussion intensified, not least within geography. Howell, for instance, considered Habermas' analysis of the public sphere as an illustration of theorizations which have normative political, as well as socially descriptive dimensions, suggesting that the key feature of Habermas' study of the public sphere was its demonstration that this was both an historically and spatially specific phenomena and simultaneously a normative–political ideal. Howell argues, however, that Habermas' theorization of the public sphere is insufficiently localized; a feature that he suggests is, by contrast, clearly evident in its conceptualization by Hannah Arendt, although subsequently Howell, drawing on the work of Cohen and Arato, as well as later work by Habermas, comes to promote Habermas' analysis of the public sphere over and above the emphasis on localism fostered by postmodern, communitarian, and feminist theorists.

The work of Arendt is also discussed briefly by Phillips, whose discussion has other parallels in its suggestion that Habermas' conception of the public sphere has utopian and relational elements, which can be seen to broadly equate Howell's notions of political and social theories. Phillips also highlights questions of spatiality, suggesting that Habermas linked the public sphere to a range of territorialized spaces. Phillips explores the degree to which the concept might be applied to the countryside – the public sphere appearing in Habermas' work and in many subsequent studies as a highly urban phenomena – as well as highlighting how the concept of the public sphere connects into a range of themes in Habermas' later work and into notions of critical theory and critical geography.

Unsurprisingly, the spatiality of the public sphere has been the motif for several other geographical studies. Laurier and Philo, for instance, highlight studies which question the degree to which the public sphere is constituted within clearly bounded spaces, before considering the degree of variability of practice within one of the sites of emergence of the public sphere identified by Habermas, namely the urban coffee house. Laurier and Philo argue that these institutions were spaces of quite diverse sociability in which reasoned debate was often far from predominant. Withers adopts a rather broader focus, suggesting through an examination of the production, circulation, and audiences of geographical knowledge in eighteenth and nineteenth century Scotland, that public awareness was constituted across and between a series of locations, including some which might be seen as private rather than public. The call for recognition of wider locations is also evident in the work of Ogborn and Bridge. The former explores the ways in which the streets of eighteenth century London were constituted within philosophical, political, and literary discourses as spaces in which private individuals come together as a public, albeit in often quite contrasting ways. Reference to the street as a public sphere is also evident in the work of Bridge, who seeks to explore what he characterizes as spaces of rationality within the city, including within the public spaces of the street and spaces of civic sociability such as parks and public buildings. In developing his arguments Bridge draws primarily on Habermas' elaboration of a theory of communicative action rather than his early analysis of the public sphere, although there are strong lines of connection between these two strands of work. Bridge suggests that Habermas' conception of communicative actions needs broadening to encompass a wider range of activity spaces than implied within

Habermas' conceptualization of the public sphere, as well as seeking to deepen it to encompass non discursive and embodied activity. In his pursuit of these ends, Bridge suggests that recent reformulations of pragmatism within communication theory and feminism provide significant benefits over the theorizations of Habermas, not least because of the emphasis given to particularity and difference. This is an argument that has clear resonances with issues raised within several of the aforementioned geographical studies of the public sphere, and also with a series of more recent studies by Habermas and associated debates that have hitherto received little attention within geography.

Geography and Habermas: Signs of Disengagement

With the exception of work on the public sphere, engagement with the work of Habermas has waned within geography since the 1980s and early 1990s. In part this can be seen to reflect the success of critical geography, which has become a central theme, and perhaps even prevailing orthodoxy, in Anglo American human geography, with many of the targets of critique, such as notions of value free science, now being widely rejected. It is also the case that a series of other critical perspectives have moved into geography subsequent to its initial engagement with the critical theory of Habermas and the Frankfurt School, and as a consequence geographers now express similar concerns but in a different, if commensurable way. However, it is also clear that the work of Habermas has itself become the subject of criticism within these new perspectives.

One area of criticism relates to Habermas' concept of the public sphere. Habermas argued that the concept of a public sphere whereby unrelated people would come together to discuss issues and establish common viewpoints through unconstrained argumentation emerged in seventeenth and eighteenth century Europe, and although never fully enacted was highly significant in the emergence of capitalism. He claims that the public sphere provided an arena where mercantile and industrial capitalists excluded from political institutions of the state could nevertheless direct the activities of the state via the elaboration of policies identified as rational, and also a form of action which transcended capitalist relations in that it established the principle that reason and emancipation from social constraints could follow from open and unconstrained debate between people. Habermas then seeks to trace the expansion and yet dissolution of the public sphere in association with organizational change in capitalist production and state power, and the emergence of mass communication media during the course of the nineteenth and twentieth centuries, which he suggests together led to fragmentation of the public sphere into publicity, public relations work, and public opinion research.

While initially quite slow to pick up on this aspect of Habermas' work, geographers have, as noted above, recently come to pay it some attention. At least in part this is probably because the concept of the public sphere appears to be quite readily spatialized: public sphere hence becomes read as public space or spaces of the public. However, while such readings have clear appeal to some geographers, it is also evident that many geographers with an interest in public space are less impressed with its emancipatory potential than Habermas. Feminist geographers, for instance, have argued that public/private distinctions often disempower women, who may find themselves enclosed in private, domestic spaces and denied access to public spaces. Habermas certainly recognized that bourgeois public will formation of seventeenth/eighteenth century Europe was far from open to all, it being constructed through restrictions both of gender and property. It has been suggested, however, that not only are there other lines of exclusion from the public sphere, but that the public sphere might be a space where such differences are in part constituted and where people relate to each other in ways other than through processes of rational debate and argument.

Habermas' more general concept of emancipatory theory was also subject to criticism by feminists, it being suggested that the notion that acquisition of knowledge brings about freedom was a highly masculinist conception. Postcolonial theorists, including the former promoter of Habermasian critical theory, Derek Gregory, criticized the Eurocentricism of Habermas, suggesting he failed to sufficiently recognize the degree to which his arguments privilege Western constructions of knowledge. Habermas was also high on the target lists of postmodern critics who identified him as an advocate of a modernist project of enlightenment that was not simply as yet unrealized but was inherently flawed in that its actions, while conducted in the name of emancipation, brought about further relations of domination. Habermas acknowledged his allegiance to notions of enlightenment and to what to him was still the unfinished project of modernity, although many of the criticisms made by postmodernists and others were well known to him because they had been identified by earlier members of the Frankfurt School, not least by Adorno and Horkeimer in *The Dialectic of Enlightenment*. Indeed, Habermas' study *The Structural Transformation of the Public Sphere* was criticized by Adorno and Horkheimer for being insufficiently critical of enlightenment concepts. Yet in this study, and in Habermas' subsequent work on knowledge constitutive interests, there is clearly recognition that knowledge is linked to the exercise of power, although for

Habermas this power can be potentially emancipatory, as well as often a means of oppression and domination.

Within geography, association of knowledge with en lightenment and emancipation was viewed with increas ing skepticism through the 1990s, and while work with an explicit commitment to effecting social change became increasing prominent, this tended to be expressed very much as a commitment to furthering the interests of particular social groups rather than as an expression of some generic human interest. Habermas has come to change many of his views in the period since geographers began to engage with them, in part as a response to critics. However, rather than turn away from his notion of a critical theory as a rational justified orientation toward emancipation, Habermas has continued to espouse this view, albeit through four distinct, but often overlapping, lines of arguments. As will be demonstrated below there may well be considerable scope for further engagement by geography in these lines of argument.

Habermas Post 'Human Interests'

Speech Acts and Universal/Formal Pragmatics

While Habermas' theory of human interests was picked up by some geographers as providing an epistemological framework for understanding geography and promoting the adoption of critical perspective, beyond geography it was widely criticized for being metaphysical, setting up a series of *a priori* foundational concepts, criticisms which were recognized but not elaborated on in Gregory's *Ideology, Science and Human Geography*. Habermas himself came to accept the arguments of his critics, recognizing that in doing so he risked losing any rationale for the adoption of a critical approach to knowledge production, in which case the orientation to emancipation would simply become a matter of preference rather than ra tional choice. Habermas was not prepared to endorse such a relativist position, and sought to see if he could establish an alternative theorization of why people should adopt a critical position.

To develop this argument Habermas undertook a linguistic turn, looking to see if an analysis of language use could provide a rationale for a concern for emanci pation. He came to argue, in his theory of universal or formal pragmatics, that the orientation to emancipation was not simply one among a variety of other perspectives from which people could choose, but rather was a value orientation that people, albeit frequently unwittingly, enacted in their daily lives. Specifically, he suggested that normal communication or dialog assumes a background consensus that involves the implicit recognition of four validity claims: intelligibility, truthfulness, rightness, and sincerity. This background and associated validity claims must, he suggested, be maintained if communication is to be achieved, although they may be routinely questioned and recovered through discussion (see **Table 3**).

As **Table 3** illustrates, Habermas' theory of universal/ formal pragmatics suggests that speech not only involves the use of symbolic expressions but also makes con nections to the extra linguistic worlds of objects/events, social expectations/obligations, and personal experi ences/emotions. As a consequence, within language people necessarily, although frequently not explicitly, make truth or validity claims about their connection with these three worlds.

The notion that language has connections to any ex ternal referent has been widely questioned within the so called crisis of representation, it being extensively argued that language functions, at best, as such an opaque me dium of representation that notions of a correspondence with any world outside language is an illusion. Habermas likewise rejects so called correspondence theories of truth, but also argues that language inherently makes claims about worlds beyond itself. For Habermas the problems of representation, although important, do not negate the raising of truth claims. The problem for Habermas is not with raising such claims, indeed for him this is an unavoidable aspect of language use. The problem rather lies in suggesting that truth claims are infallible, that is, in suggesting that they take the form of statements that are seen as immutably correct. For Habermas validity claims are potentially revisable even when widely accepted, and the question that exercises Habermas is how some truth claims are accepted and others rejected. He attempts to explain this by suggesting

Table 3 The raising and recovery of validity claims in speech acts

Validity claim	*Problematizing question*	*Maintenance response*	*Type of validity*	*Relational connection*
Intelligibility	What do you mean?	Clarification	Condition of discourse	Linguistic worlds
Truthfulness	Did it happen like that?	Explanation	Discursive validity claim	External world of objects and events
Rightness	You shouldn't say that?	Justification	Discursive validity claim	Social world of expectations and obligations
Sincerity	I don't believe you?	Sincerity	Non-discursive validity claim	Internal world of personal experience

that the use of language in everyday conversation pre supposes an ability to continue and re establish lost consensus through argumentative processes in which reasons are advanced, debated, and evaluated.

For Habermas this argumentative procedure provides the answer to how truth or validity claims are accepted and rejected. He suggests that deciding on their accept ability or nonacceptability rests on argumentative rea soning: that is, the putting forward of justifications in support for a validity statement so that other people understand the statement and are in a position to accept or reject it. For Habermas problematic validity claims are accepted or rejected through consensual agreement fol lowing a reasoned argument of all those implicated in a particular truth claim. Truth claims are hence inter subjectively established.

Habermas uses this argument about language in a whole variety of ways, including within his theory of social rationalization, his notion of a philosophy of intersubjectivity, and his discussion of ethics, universal ism, and pluralism. It has, however, been objected that he neglects many aspects of language use in his focus on reaching understanding, and also that there are many forms of communicative activity other than through speaking, arguments which are very clearly evidenced within the criticism of Habermas in Bridge's *Reason in the City of Difference: Pragmatism, Communicative Action and Contemporary Urbanism.*

Action and Rationalization

Although much of Habermas' work is philosophical in character, he has sought to continually integrate this with social analysis, a feature that is evident in the work of most members of the Frankfurt School and has frequently been seen to constitute one of its defining features. Habermas' *Structural Transformation of the Public Sphere* constitutes an early example of this, it representing both a study of change in European society from the seven teenth century and also an attempt to recover the voice of reason championed by Enlightenment philosophers such as Kant. Habermas' conception of the public sphere shows clear parallels with his subsequent elaboration of speech acts and the discursive recovery of validity claims.

Knowledge and Human Interests was very much focused on epistemological arguments, but subsequent to this Habermas produced a series of studies that integrated philosophical arguments with social analysis. In *Legitim ation Crisis*; *Communication and the Evolution of Society*; and particularly his *Theory of Communicative Action*, *Volumes 1 and 2*; Habermas established analyses of societal devel opment and social action orientated both to re constructing the ontology established by Marxist historical materialism and to developing new epistemo logical support for the concept of critical theory.

A key feature of these analyses was the claim that social action needed to be differentiated into labor, or purposive–rational action, and social interaction, or communicative action. Marxist historical materialism had focused on the former, effectively arguing that social history has seen societies expand their control over ex ternal nature through the elaboration of systems of purposive–rational action: Marx's concept of modes of production. Heightened productive capacity has not re sulted in emancipation as implied by Marx's theoretical schemas, but rather, and here Habermas draws exten sively on Weberian analysis of the protestant work ethic and bureaucratization as well as earlier Frankfurt School analyses, has seen the emergence of new forms of dom ination centered on communicative interaction although still often interpreted through the lens of purposive–ra tional action.

Habermas argues that there is, hence, a need to sup plement analysis of the growth of systems of purposive– rational action (which he often characterized as steering media of system integration) with examination of mechanisms of communicative interaction (which he characterized within his *Theory of Communicative Action* as involving cultural reproduction, social integration, and socialization). He further argues that there has been a tendency for systems of purposive–rational action to become uncoupled from, and predominate over, com municative interaction, an argument which has clear echoes of his earlier analysis of the dissolution of the public sphere but which he comes to describe in terms of the mediatization and colonization of the lifeworld. He suggests that systems of purposive–rational integration, such as the money economy and bureaucracies, both become disconnected from processes of communicative will formation and come to regulate arenas of social life such as cultural transmission, socialization, and subject formation (or the lifeworld) which are more appropri ately the domain of communicative action, a develop ment which as **Table 4** illustrates, he sees as productive of a series of pathological results such as loss of cultural meaning, anomie, and mental illness. **Table 4** also illus trates how these arguments link back to Habermas' an alysis of the relational connections enacted within speech acts and thereby to his broader arguments about critical theory. As with his earlier work on the public sphere, philosophical argument is linked to social analysis.

While there are clear parallels between Habermas' *Theory of Communicative Action* and his earlier study of the public sphere, and also with Horkheimer's early call for critical theory to encompass the mediation of economic, psychological, and cultural facets of social life, his new work has excited the attention of far fewer geographers. Reference to it has been made in discussions of mod ernism by the likes of Harris and Gregory, with parallels and contrasts being drawn with the work of people such

Table 4 Habermas' theorization of social action

Form of action	*Action system*	*Focus of action*	*Pathological tendency*	*Relational connection*
Purposive rational	System integration	Achievement of aims	Alienation	External world of objects and events
	Cultural transmission	Meaningful interpretation	Loss of meaning	Linguistic worlds
Communicative action	Social integration	Social legitimacy/ acceptance	Anomie	Social world of expectations and obligations
	Socialization	Personal identity	Mental illness	Internal world of personal experience

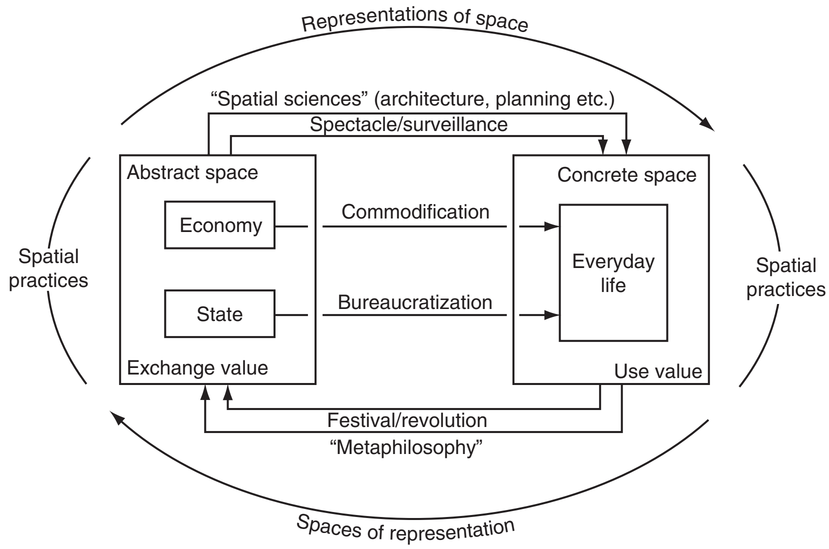

Figure 2 A geographical configuration of modernization processes mimicking Habermas's theory of communicative action. From Gregory, D. (1994). *Geographical Imaginations*, Figure 30. Oxford: Blackwell.

as Foucault, Giddens, Lefebvre, and Harvey (see **Figure 2**). Such studies, however, remain at present largely schematic and often the emphasis in the contrasts being drawn is to imply that Habermas pays insufficient attention to the geographical/spatial dimensions of social life. *Theory of Communicative Action* is, however, quite central to Miller's *Geography of Social Movements: Comparing Antinuclear Activism in the Boston Area* and to Bridge's *Reason in the City of Difference: Pragmatism, Communicative Action and Contemporary Urbanism*, with the former study drawing on notions of system and the colonization of the lifeworld, while the latter focuses primarily on the differentiation of instrumental and communicative action. The former more readily embraces Habermas' arguments than the latter, with Bridge being highly critical of the dualistic character of Habermas' theorization of actions, arguing that instrumental actions are more communicative in character than Habermas supposes, and also that communicative actions are more strategically self interested and non consensual, as well as non linguistically mediated, than conceived by Habermas. The overall outcome is arguably to undermine the whole logic of Habermas' theorization of action, although Sayer presents an alternative response to the fuzziness of Habermas' conceptual distinctions, claiming for instance that the distinction between system and lifeworld highlights important dimensions of social life even though they never exist in a pure form. Sayer further adds that many of the social pathologies which might be identified as products of the systemic colonization of the lifeworld are both readily apparent in contemporary analysis of the cultural economy and the politics of identity and distribution, and also in earlier theorizations by both classical and Marxist political economy. Indeed, while geographers may be hesitant to draw directly on the theories of action and rationalization outlined by Habermas, many of his concerns and claims may be seen to have flowed quite surreptitiously into geography, both because

Habermas himself drew on literatures such as Marxism, phenomenology, and pragmatism which exerted an influence on human geography discourse and because his work impacted on others whose work in turn has been drawn on by geographers. In the case of Sayer, for instance, he connects his discussion of system and lifeworld into the work of Nancy Fraser who, as noted earlier, was herself very much influenced by Habermas.

The Philosophy of Intersubjectivity

Habermas' emphasis on the integrative aspects of language is a major point of difference from, and disagreement with, the emphasis on the problems of representation within postmodern/post structural lines of thought. In books such as *The Philosophical Discourse of Modernity* Habermas addresses such criticisms, focusing particularly on the critique of language advanced by postmodern thinkers such as Derrida and Foucault. Habermas argues that such work expresses the desire to deny or suspend the presentation of truth claims, but that these constitute acts of performative contradiction in that truth claims are contained in the very utterances of those who seek to deny or suspend truth claim. As outlined in his theory of universal pragmatics, for Habermas the very practice of uttering implies that people, even postmodernists and post structuralists, implicate themselves in an argumentative procedure in which they seek to justify their own position and criticize the position of others. Postmodernists, and indeed earlier generation Frankfurt School critical theorists such as Horkheimer and Adorno, are, Habermas suggests, tied to a philosophy of conscious, or philosophy of the subject, which centers on whether or not individuals can come to some understanding or representation of the world around them. Habermas argues that whatever the answer provided, such perspectives neglect the role of intersubjective relations. Theoretical reflections, he argues, always have to be seen as being accountable to others and hence are justified not through the intellectual prowess of a particular person but rather via the prospect of public argumentation orientated to reaching mutual understanding.

Habermas' criticism of postmodernists and post structuralists has been taken on board by people such as Nancy Fraser and Thomas McCarthy who seek to identify performance contradictions within the work of Foucault and Derrida, respectively. Such ideas have also been employed within geography, although often without direct reference to the work of Habermas, save for Phillips' use of them in his discussion of rural studies and philosophy in human geography. The repeated surfacing of expressions of concern about relativism and the political dimensions of postmodernism/post structuralism, along with calls for more participatory and dialogical modes of research and more relational ontologies of the self, may be seen to relate to issues pertinent to Habermas' concept of the philosophy of the intersubjectivity, although as yet there has been no substantive attempt to import this concept into geography.

Communicative Ethics

Habermas' discussion in *The Philosophical Discourse of Modernity* can be said to reveal the normative or evaluative dimension of writing, even among those that would seek to deny this dimension any practical significance. In a series of subsequent books, Habermas seeks to explore in more detail how one should deal with a plurality of moral positions. Specifically in books such as *Moral Consciousness and Communicative Action*, Habermas seeks to develop a theory of practical discourse, that is a theory concerning arguments about morals and ethics. The purpose of this theory is to demonstrate that normative statements, that is, remarks about whether a person is right to do something, are open to reasoned arguments as much as statements about whether something exists or not. Habermas claims, in line with his universal pragmatic theory of language, that norms can be considered valid or right if they would be consented to in free discussion of all concerned. Discourse or communicative ethics seeks to suggest that social norms of equality and mutual recognition of others could be sustained by such argumentative procedures.

Habermas terms his work as a minimalist ethics in that it is concerned only with what he has termed questions of morality, rather than with ethics. While both morals and ethics imply normative statements about how people ought to live and behave, morality *per se* is, Habermas suggests, concerned with formal, general, statements of social justice and individual autonomy. These issues are, Habermas argues, inescapable presuppositions of all moral and ethical discussions and are also, he claims, based on the intersubjectivity of social life through which people are 'individuated' (made into individuals) through socialization. Morals are therefore seen by Habermas to be the universal core of normative statements, while ethics involve much more personal, contextual, judgments.

The suggestion that there might be a universal core to at least some moral positions is clearly a contentious claim. As might be expected, Habermas seeks to sustain or ground this claim with reference to his reconstructive analysis of language. Specifically he suggests that the moral precepts of justice and solidarity reflect inescapable facets of the argumentative procedures that form the deep rules of language use. Simplifying greatly, Habermas develops this argument in three stages. First, he suggests that valid moral precepts are ones that have the quality of fairness or impartiality, which implies that all those implicated in a particular discourse would agree on them. Second, he suggests that the qualities of fairness and impartiality can only be expressed by use of a principle of universalization. By this, Habermas means the ability to generalize modes of action and maxims, and the interests these actions and maxims entail. This is necessary so that people can decide whether or not they are prepared to accept the consequences and side effects of other people acting in accordance with a particular moral principle. Moral norms require that all those implicated by it can accept the consequences and side affects of all others following it also. Third, Habermas suggests that the universalization principle can itself be rationally justified in that it enacts the procedures of argumentation that are inherent in language use.

The desire to establish a universal core to moral positions, even after the farming out of much of that which is usually taken to constitute morality and moral questions, is clearly open to criticism that it is a totalizing project, seeking to reduce difference and individual choice. Habermas, and sympathetic critics such as Seyla Benhabib, Richard Bernstein, and Thomas McCarthy have sought to demonstrate that the implications of the universalization principle does not involve exclusion of the other, but rather should be seen as an attempt to disclose the basis for respecting difference at the same time as considering and regulating its effects on the lives of others.

There has been little direct reference to Habermas' discourse ethics within geography. There has been some passing reference to it within work on the geographies of morality and care: Habermas' universalization principle resonates clearly with debates over care and responsibility for distant others. Benhabib has shown that Habermas' discourse ethics connects with issues of multiculturalism, citizenship, and governance, the first of which is an important element of Bridge's *Reason in the City of Difference: Pragmatism, Communicative Action and Contemporary Urbanism*. Bridge, however, is skeptical about the universalization he detects in Habermas, an issue that also concerns Howell, although he comes to argue, in line with likes of Benhabib, that Habermas' minimal ethics provides a means of respecting difference.

It is important to note that Habermas, as well as developing his highly abstract notion of discourse or communicative ethics, has also sought to engage practically in political and ethical debate, undertaking a series of commentaries on contemporary cultural and ethical issues, including globalization, religion, terrorism, and the implications of genetic engineering. Heffernan is one geographer who has paid attention to one such commentary and given the subject matter of the others, it could be argued that Habermas has come to act as a public intellectual in geography, albeit one whose interventions have been largely ignored within academic human geography.

Concluding Remarks

As clearly evident in **Table 1**, the work of Habermas is very extensive, and this article has had to be highly selective. Attention has focused on outlining how his work was initially drawn into Anglo American geography and some of the ways in which his work has changed since the writings that attracted the attention of human geographers. Some reference has been made to Habermas' connection to the Institute of Social Research and the so called Frankfurt School, but clearly there is much more to be said about the work and influence of other members of the Institute/School in geography. The work of Klaus Eder and Klaus Offe has, for instance, been quite widely utilized within geography in relation to studies of economic geography, social movements, class identities, and environmental issues, while Walter Benjamin's work has received considerable attention in studies of urban cultural geography (e.g., see Derek Gregory's *Geographical Imaginations* and Steve Pile's *Real Cities*). Benjamin was only ever loosely connected to the Institute of Social Research in Frankfurt, and there are clear points of difference between his work and that of Habermas, and indeed between each of the people identified as part of the Frankfurt School. The reason for focusing on Habermas is that his work, particularly his book *Knowledge and Human Interests*, was highly influential in the emergence of notions of a critical geography in the late 1970s. The arguments outlined in this book and their application within human geography has been discussed, as has subsequent points of divergence between the discourses of critical human geography and later work by Habermas. This should not be taken to imply that Habermas' works are of no relevance to human geography. Several of his ontological concepts have attracted the interest of geographers, most notably the notion of the public sphere and also social and system integration and the colonization of the lifeworld. Furthermore Habermas' criticisms of postmodernists and post structuralists have clear parallels within contemporary debates in human geography. It may also be that geographers could usefully pay more attention to Habermas' attempts to justify critical perspectives and to his discussions of the philosophy of intersubjectivity and discourse ethics, as well as to the work of a series of social theorists and philosophers who have been influenced by Habermas and who perhaps constitute a new generation of Frankfurt School critical theorists.

See also: Critical Geography; Critical Rationalism (After Popper); Functionalism (Including Structural Functionalism); Gregory, D.; Humanism/Humanistic Geography; Phenomenology/Phenomenological Geography; Positivism/Positivist Geography; Postmodernism/Postmodern Geography;

Poststructuralism/Poststructuralist Geographies; Pragmatism/Pragmatist Geographies; Private/Public Divide; Public Space; Radical Geography; Structuralism/Structuralist Geography; Structuration Theory; Structurationist Geography; Systems Theory.

Further Reading

Benhabib, S. (1992). *Situating the Self*. Cambridge: Polity.

Benhabib, S. (2002). *The Claims of Culture: Equality and Diversity in the Global Era*. Princeton, NJ: Princeton University Press.

Bridge, G. (2005). *Reason in the City of Difference: Pragmatism, Communicative Action and Contemporary Urbanism*. London: Routledge.

Clark, G. and Dear, M. (1984). *State Apparatus: Structures and Language of Legitimacy*. London: Allen and Unwin.

Cohen, J. and Araro, A. (1992). *Civil Society and Political Theory*. Cambridge, MA: MIT Press.

Fraser, N. (1989). *Unruly Practices: Power, Discourse and Gender in Contemporary Social Theory*. Cambridge: Polity.

Gregory, D. (1978). *Ideology, Science and Human Geography*. London: Hutchinson.

Gregory, D. (1980). The ideology of control: System theory and geography. *Tidjschrift voor Economische en Sociale Geografie* 71, 327 342.

Gregory, D. (1994). *Geographical Imaginations*. Oxford: Blackwell.

Heffernan, M. (2005). The end of Atlanticism: Habermas, Derrida and the meaning of Europe in the twenty first century. *Geopolitics* 10, 570 575.

Held, D. (1980). *Introduction to Critical Theory: Horkheimer to Habermas*. London: Hutchinson.

Honneth, A. (1991). *The Critique of Power: Reflective Stages in a Critical Social Theory*. Cambridge, MA: MIT Press.

Horkheimer, M. (1972). *Critical Theory: Selected Essays*. New York: Seabury Press.

Horkheimer, M. and Adorno, T. (1972). *Dialectic of Enlightenment*. New York: Herder and Herder.

Howell, P. (1993). Public space and the public sphere: Political theory and the historical geography of modernity. *Environment and Planning D: Society and Space* 11, 303 322.

Howell, P. (1994). The aspiration towards universality in political theory and political geography. *Geoforum* 25, 413 427.

Jay, M. (1996). *The Dialectical Imagination: A History of the Frankfurt School and the Institute for Social Research 1923 1950* (2nd edn.). Berkeley, CA: University of California Press.

King, R. J. (1989). Capital switching and the role of ground rent: 1, 2 and 3. *Environment and Planning A* 21, 445 462, 711 738, 853 880.

Laurier, E. and Philo, C. (2007). 'A parcel of muddling muckworms': Revisiting Habermas and the English coffee houses. *Social and Cultural Geography* 8, 259 281.

Lewis, J. and Melville, B. (1978). The politics of epistemology in regional science. In Batey, P. (ed.) *Theory and Method in Urban and Regional Analysis*, pp 82 101. London: Pion.

Ley, D. and Samuels, M. (1978). *Humanistic Geography: Prospects and Problems*. London: Croom Helm.

Marcuse, H. (1964). *One Dimensional Man: Studies in the Ideology of Advanced Industrial Society*. London: Routledge and Kegan Paul.

McCarthy, T. (1978). *The Critical Theory of Jürgen Habermas*. Cambridge: Polity.

Miller, B. (2000). *Geography and Social Movements: Comparing Antinuclear Activism in the Boston Area*. Minneapolis, MN: University of Minnesota Press.

Ogborn, M. (1998). *Spaces of Modernity*. New York: Guilford Press.

Phillips, M. (1994). Habermas, rural studies and critical social theory. In Cloke, P., Doel, M., Matless, D., Phillips, M. & Thrift, N. (eds.) *Writing the Rural: Five Cultural Geographies*, pp 89 126. London: Paul Chapman.

Phillips, M. (2005). Philosophical arguments in human geography. In Phillips, M. (ed.) *Contested Worlds: An Introduction to Human Geography*, pp 13 83. London: Ashgate.

Pile, S. (1990). *The Private Farmer: Transformation and Legitimation in Advanced Capitalist Agriculture*. Aldershot: Dartmouth Press.

Pile, S. (1991). Practising interpretative geography. *Transactions of the Institute of British Geographers* 16, 458 469.

Pile, S. (2005). *Real Cities: Modernity, Space and the Phantasmagorias of City Life*. London: Sage.

Sayer, A. (2001). For a critical cultural political economy. *Antipode* 33, 687 708.

Short, J. R. (1984). *The Urban Arena*. London: Macmillan.

Unwin, T. (1992). *The Place of Geography*. Harlow: Longman.

Wellmer, A. (1991). *The Persistence of Modernity: Essays in Aesthetics, Ethics and Postmodernism*. Cambridge: Polity.

Wiggershaus, R. (1994). *The Frankfurt School*. Cambridge: Polity.

Withers, C. (1999). Towards a history of geography in the public sphere. *History of Science* XXXVII, 45 78.

Relevant Websites

http://www.habermasforum.dk
The Habermas Forum Online Collection of Articles, Reports, and News Relating to Habermas.

http://www.ub.uni frankfurt.de
The Horkheimer Pollock Archive.

http://www.ifs.uni frankfurt.de
The Institute of Social Research, Johann Wolfgang Goethe University, Frankfurt am Main.

Cross-Cultural Research

T. Skelton, National University of Singapore, Singapore

Glossary

Transculturation This term was coined by Fernando Ortiz, a Cuban anthropologist, in 1947 to describe the phenomenon of merging and converging cultures, predominantly within the context of colonialism, invasion or slavery. He proposed that there were phases to the process of transculturation. The merging and converging are established when conquered people choose and select what parts of the dominant culture they will accept. This stands in opposition to 'acculturation' because it recognises the fact that subordinate groups have power and agency to create their own forms of the dominant culture.

Introduction

Space, place, and spatiality are central themes for many human geographers. Cultural identities and practices are located within spaces and places. Consequently, many geographers conducting research in 'other' places and spaces are often engaging cross culturally. However, not everyone describes or defines their research processes and practices as 'cross cultural'. This means that there is often more cross cultural research taking place within human geography than is evident when searching for publications on the subject. In most contemporary books on methodologies within human geography, there is a chapter on cross cultural methods but the focus of these chapters can be quite different. Hence, trying to definitively define cross cultural research in geography proves to be difficult. Nevertheless, an overview of published materials that either explicitly name cross cultural approaches, or imply their use, provides an insight into specific complexities and features that have to be taken into account when embarking on cross cultural research.

These complexities and features are evident at 'all' stages of the research process (designing the research questions, conducting the research, analyzing the data, and writing up the findings) and of research practice (methodological approaches used, role of research assistants and interpreters, and engagements between researchers and participants). It is important to note that defining elements and identified features of cross cultural research are not exclusive to a cross cultural approach. Similar traits can be identified within feminist geography, postcolonial, and research ethics discourses about doing research (to name the three main areas). Indeed, publications that engage with debates about cross cultural research often link with postcolonialist, feminist, and/or ethical approaches. However, many human geographers argue that extra special care and attention is usually demanded within cross cultural research because of the evident (or presumed) differences between and across researchers and participants, a range of cultural sensitivities (and insensitivities), language, interpretation, and cultural practices and meanings.

What Is Cross-Cultural Research?

Generally, this is understood as research conducted within cultural contexts 'different' to those of the researcher or the research team. Often this 'difference' is manifest through different spaces, places, or spatialities to those the researcher would usually identify themselves with (while at the same time acknowledging that identities with place and space are not constant and fixed). However, because of the complexity and diversity of cultural aspects then a researcher may remain in their same 'usual' space or place and yet be able to conduct cross cultural research. The problem with this generic understanding of cross cultural research is that almost all human geographical research could be construed as cross cultural – which renders the concept meaningless. The other problem is that such definitions can construct the researcher as a fixed entity who enters a different place/space and/or culture and the cultural difference is always defined in relation to the researcher, thus always placing the researcher as central. In reality, as human geography repeatedly demonstrates, identities are not fixed or immutable. Consequently, researchers may cross cultures and spatialities and experience changes in their own identities and positionalities. The researcher does not remain culturally unchanged by these encounters across spatial and cultural borders.

Within the literature, however, there are often two central presumptions made about what defines cross cultural research. One is that cross cultural research occurs in a place different from the daily milieu of the researcher, usually overseas or across national borders. The second is that the research takes place within the same spatiality that the researcher is familiar with (their own country, for example) but engages with people of a different culture (such as white Australians working with aboriginal or Torres Straits Islander peoples). In the second definition, the researcher crosses certain cultural (rather than spatial) boundaries. Such cultural boundaries may be based upon socially constructed differences such as class (interviewing elites, for example), race, ethnicity, gender, religion, sexuality, language, or other cultural practices, or upon any combination of these bases of difference (ethnicity and gender, for example). For some human geographers, cross cultural research is said to take place when these 'two' elements are present – the crossing of spatial and cultural boundaries. For others, their research is defined as cross cultural because of the need for interpreters or translators who play a significant role in the actual research practice. This often involves a further layer of cross cultural relations – that of the cross cultural connections or disconnections between the researcher and the interpreter or translator through the research process. In such a situation, there are at least three actors working across cultures and spaces: re searcher, research participant, and interpreter/translator.

Doing Cross-Cultural Research

In human geographers' discussions about the process and practice of conducting cross cultural research, there are a range of themes that are addressed. These often connect with wider debates within research methods and are not exclusive to cross cultural debates. These involve more everyday issues such as differential conceptualizations of time, reciprocity, gift giving, hospitality, social manners, and so forth. These can all be part of the frustrations and fascinations of doing cross cultural research. Almost all cross cultural researchers discuss the ways in which they had to learn different ways of being, of conducting re search, and of engaging with people. Overarching these more mundane (but extremely important) experiences, it is possible to identify specific thematics that are con sidered highly pertinent and essential to cross cultural research.

The Role of Ethics

Many of the ethical considerations made by those writing about cross cultural research are relevant to all research with human subjects. However, many authors suggest that in a cross cultural context ethical practices are even more important to think about in research planning, design, and implementation. This is because cultural complexities can confound particular ethical approaches which are invariably constructed within the cultural and research context of the researcher, and not necessarily of the participants. What appears ethical research practice in one culture may not be so in another. For example, the ethical insistence of gaining written, informed consent demanded by most academic ethical review boards/ committees can be highly problematic in cultures with limited literacy skills or a negative history of form signing whereby the signatories did not know what they were signing for. Hence, alternative, culturally acceptable and ethically valid means of negotiating informed consent have to be utilized. Interviewing in a way that maintains Euro American senses of privacy (on a one to one basis, away from the household or community) may be seen as rude and suspicious in a culture with different senses of personal privacy. In a cross cultural setting (particularly in ones that are postcolonial or still colonial), the stereotype and construction of the researcher as all knowing, potentially dangerous, and powerful are still very pervasive and a great deal of work and energy is required to counter this presumption on the part of the researcher. It is important to bear in mind that ethical principles are social and cultural constructions. Con sequently, the cross cultural researcher has to be open to the possibility and necessity of reframing the ethical underpinnings of their research in line with the group of people who are the participants in the project. The ethical research practice of the researchers can some times conflict with those of locally recruited research assistants, translators, or interpreters and hence com promise and adaptation will be essential. The research questions that are asked and how they are asked can be distinctively different in cross cultural settings; some will breach the ethical or moral sensibilities of those being asked. Understanding the cultural, ethical meanings and practices in the site of the research is an essential con sideration in cross cultural research.

Insider/Outsider

This binary is often discussed by those doing cross cul tural research. It is assumed to be an indication of a researcher's connection or disconnection with the cul tural group they are working with. The presentation of these two terms as a binary is problematic as cross cul tural researchers tend to find themselves in a fluid and slippery space 'between' the two. Indeed, the term 'betweeness' is often used to indicate that it is possible in certain contexts to be both or neither insider and out sider; to both connect and disconnect in relation to a cultural community. It is also possible that different re search participants (even from the same cultural com munity) can read researchers in alternative ways and define them as insider or outsider. The researcher may connect with something that one participant sees as a significant part of their social or cultural identity (their gay sexuality, for example), but may not be able to connect with another aspect of a participant's identity, such as being a mother.

Both sides of the binary identities can bring advan tages and disadvantages to cross cultural research. The insider/outsider status can also be complicated in re search teams as 'lead' researchers may be perceived as 'outsiders', but their research assistants, translators, or interpreters recognized as 'insiders'. In some communities, people will not agree to interviews, and not participate in discussion groups nor complete questionnaires (of course, there are more research methods that are used cross culturally than this, but these are the most common types reported) because the researcher is an 'outsider'. In different communities, this 'outsider' status may be seen as a real strength because the researcher is considered to be external to, and separate from, the internal sociopolitical and economic–cultural relations of the research participants. In other locations or communities, 'insiders' will be considered as more empathetic, trusted, or reliable than outsiders. If cross cultural research is conducted longitudinally in the same cultural place, then an 'outsider' researcher may become an 'insider' over time – but that is something that the research participants decide upon, it is not something researchers can claim.

Cross cultural researchers often record the ways in which their insider/outsider status helped them gather 'different' kinds and levels of information. 'Insiders' might be privy to more intimate information that requires good linguistic and cultural understanding; 'outsiders' might be given more detail and explanation as research participants want to make them fully understand the cultural context rather than presume that the researcher already knows their stories and experiences.

Sameness/Difference

This is another binary that researchers discuss as a key element of doing cross cultural research. It is linked to the above insider/outsider binary but is distinctive. This binary tends to be based upon the identity or positionality of the researcher and research participants rather than on the ascribed degree of cultural association or disassociation that a researcher may/may not have with a cultural group. This is an aspect of cross cultural methodology, whereby researchers consciously acknowledge difference between themselves and research participants and integrate this difference and the possible consequences of it into their research practice. This might entail acknowledging the differences from the outset and choosing methods or approaches that might counter this or working with an interpreter or locally based researcher, for example, to bring in a degree of sameness. It might also be an element that cannot be countered but is used to the advantage of the researcher, and they play on their difference and their separateness in order to elicit more explanation or detail from informants. Such an approach can often be useful when trying to work with elite, cosmopolitan, or transnational cultural groups.

One of the most common areas of difference discussed in relation to cross cultural research is that of race and/or ethnicity. This is not a difference, if it exists, that can be easily transformed or overcome and so the research methods have to be utilized in ways that take this particular difference into account. Generally, however, the published material tends to focus much more on the ways in which researchers aim to find a basis of sameness and commonality with the cultural communities they are working in. The aim is to find a connectivity that may transcend difference, and so allow the creation of an affinity or alliance which encourages the cultural and research exchange. Feminist geographers working cross culturally have reported using gender, motherhood, or sexuality as a connection. Such a construction of sameness at one level may counter balance differences at another, gender may be a connection where race is not. Inevitably such constructions of sameness and difference may have various consequences depending on who researchers are working with; they are not fixed. What is evident, however, is that there is often a reciprocal exchange between researchers and research participants as they all search for possible commonalities to create some shared spaces of similarity from which meaning and other understandings can emerge.

Translators and Interpreters

Some human geographers define cross cultural research as that which requires the use of translation and/or translators or interpreters. Others might argue that all forms of cross cultural research require some degree of interpretation in order to be able to talk across difference.

Considering written translation first, cross cultural research requiring multiple translations of questionnaires, for example, has to have the same research meaning within very different cultural contexts. For example, in a large international and cross cultural research project on the health and welfare of children, what does the concept of 'physical punishment of children' actually mean in different cultural contexts? Does it mean slapping, punching, hitting with an object, or beating with a cane or stick? How long and how hard is the punishment administered? How can these elements be translated in a meaningful way across cultures to allow cross cultural, cross national comparisons of the ones required by organizations within the United Nations, for example? Meaningful and effective translation of the research questions then becomes a central feature of such a project and creates considerable problems for comparative cross cultural research. Much of the research project time is taken up, with the team trying to reach agreement and negotiation of terms, definitions, and linguistic interpretations.

In the case of translators or interpreters, the role of these third parties within the research process can be extremely advantageous as well as problematic. Indeed, the insider/outsider binary can get more raveled when translators and interpreters come into the frame, because they can have different aspects of being insiders or out siders. Often translators and interpreters are recruited from within the cultural community in which the re search is taking place, and they are often presumed to 'understand' the community. However, their social class or educational experience may set them apart from many in the community, their familial connections may place them differently to some participants, and they may perpetuate stereotypical interpretations of some group members. The role of these 'third parties' therefore has to be handled extremely carefully and selection and training of a translator/interpreter can be an essential and yet very difficult part of cross cultural research. The com plexities of these research relationships are often an important part of discussion in the publications from cross cultural projects.

Translation is what takes place where cultural and linguistic differences between the researchers and par ticipants are not so great, for example, a north Indian researcher working in southern India, where the cultural and linguistic differences are very wide, then a process of interpretation takes place. What is required is transcul turation whereby the translation or interpretation in cludes the cultural context alongside recognition from the researcher that they are always working with hybrid texts that have been created through a process of inter pretation, translation, transculturation, and transcription. Cross cultural researchers talk of the ways in which this process of translation/interpretation creates a new space of exchange and learning where researcher, translator/ interpreter, and participant have to move into a cross cultural relationship created by the subject area of re search. There might be a distortion of the language, but through this cross cultural process new meanings are created as 'in between' forms of understanding. However, it is equally essential to recognize that there will always be degrees of incomprehensibility; knowledges kept hidden and not told. Honesty about this is an important part of the research process.

The Politics of Cross-Cultural Research and Methods

A consideration of the politics of this type of research leads to an extremely tangled web of interlocking pro cesses and potentially problematic areas, both ethically and politically. This can render the doing and writing of cross cultural work extremely difficult and complex. However, despite these problems, cross cultural research practiced effectively can provide rich, new, and constructive knowledges that may counter perceived prejudices, stereotypes, and power–knowledge relations. Human geographers have conceptualized political issues of cross cultural research in several ways, which include not only examination of the power relations, inequalities, and injustices that allow and enable research to take place but also exploration of possible ways to challenge these in order to carry out the research. This article draws upon the three most commonly discussed approaches and demonstrates the interaction between cross cultural ap proaches and discourses of feminism and postcolonialism.

Questions of Postcolonial/Northern/Western 'Authority'

This political approach forces researchers to question why research is being done within the South or a post colonial (or continued colonial) context, and what are the consequences of such research for the people and places involved. Similar questions should be asked when re searching cross culturally within the place the researcher is located, particularly when working with groups with less or little social power. Many native and indigenous people continue to connect research with colonialism and imperialism, hence the importance of treading carefully. In order to question a particularly located powerful authority (in relation to knowledge con struction), there are a series of questions that have to be asked at all stages of the research process. Why is the research being done in these places and with these people? Who benefits from the research? What are the power relations at play both in the contemporary context and historically? How will the past connect with, and impact upon the present? As a researcher where do we fit (or where will we be placed) within these power relations and what can be done to counter them? What status will the postcolonial context create for research participants? Will they be rendered subalterns by the cross cultural research process? Can they speak, in whose language, and with which vocabulary? Which politics of postcoloniality will be used to frame and interpret participants as people, as knowledge holders, and as agents? As researchers, how can we counter Northern (Eurocentric, Western) theo rizations and representations of the 'other'? How do we avoid constructing ourselves as 'transcultural knowers' and research participants as 'native informants'?

Cross cultural research has been envisioned as a process in which connections and dialogs with others can be enabled in ways that do not perpetuate postcolonial/ Northern/Western authority, but rather create spaces and discourses of progressive politics. Such a politics can challenge academic authority and open up, make space, and cross over cultural and spatial differences to establish alternative interpretations and practices of power. This can only be done effectively, however, if many of the above questions are answered meaningfully, thoughtfully, and sensitively.

(Mis)Representation and the Production of Knowledge

Thinking through the politics of representation and knowledge production is essential within cross cultural research, because much is at stake when a group or a place is researched, spoken, and written about. Researchers are implicated in the production of people and places, as knowledge is produced, and they can be responsible for misrepresentation. Equally, cross cultural researchers learn from others but have to guard against the appropriation of knowledge. There needs to be a continual questioning within this methodological approach of why knowledge is produced, who is it produced for, and how can it be best (in the context of social justice) represented. Many cross cultural researchers also practice a critical questioning of, and resistance against, the Euro American forms of knowledge and systems of knowledge production that constitute academic geography and render people and places 'elsewhere' as case studies. Within cross cultural research, there is a need to continually ask how particular kinds of literacy and engagement that are genuinely sensitive to the spaces 'between' disparate cultures of knowledge production can have an impact on the formulation, implementation, and communication of geographical research?

Cross cultural researchers argue strongly that sensitivity about representation is even more acute in this context. In order to avoid and actively counter the reinforcement and perpetuation of certain cultural stereotypes (misrepresentation), it is essential that the processes of representation are considered at all stages of the research. The research questions asked will contribute to a particular production of knowledge, which will then go through a process of representation. The onus of getting those research questions 'right' falls upon the shoulders of the researcher and some geographers practice the production of research questions through dialog and negotiation, learning all the time from the people they work with. Many cross cultural researchers face dilemmas of how to write and produce the research and may have to abandon or exclude certain findings and things learned in order to follow a particular political and/or ethical commitment to the people they have worked with.

The end product of cross cultural geographical research is the representation of the research participants, and so there is an inevitable power relation; the researcher is perceived to be the knowledge producer. However, it has been argued that a process of auto ethnography can counter this representational dilemma. It allows for the recognition of, and sensitivity to, the research participants as knowing subjects, proactive, and self confident in their engagement with the researcher and project as they represent themselves on their own terms and remain faithful to their own understandings of themselves.

Through transculturation, processes of engagement across difference and at times of nonunderstanding, in between spaces are created (as discussed above). These spaces can bring forth instances of enlightenment, a chance to see the world through someone else's reality and the experience of a loss of privilege. While being very challenging elements in the research, they are also important sites and times for the production of knowledge. The locatedness and cultural context of such knowledges should always be articulated and theorized. The South and postcolonial spaces are not just empirical sites of research but they are also arenas for the production of theory. Genuinely, cross cultural research, learning, dialog, and knowledge production recognizes this and respects the analytical insights offered and created.

Positionality, Power, and Reflectivity

Cross cultural or transcultural interactions are the discursive and material connections that occur when cultural groups are engaged in extended contact. The cross cultural researcher is clearly embedded in these interactions and hence must continually assess one's positionality and the power relations at play through a process of reflectivity (also termed reflexivity). These elements are at play in all research with human subjects, but cross cultural researchers argue for greater vigilance because of the complexities of power and injustice indicated above. This is also about recognizing the shifting nature of power over time and space.

The role and positionality of the researcher or the research team and of research participants have to be scrutinized and interrogated because of the impact they can have on the research process and the learning that takes place. Positionality is not always fixed but there are elements which are relatively secure and are read in particular ways – such as being white in an African postcolonial context; being from the colonial power in the context of a still colonial Caribbean island. Honesty and negotiation about what such aspects of a researcher's positionality mean to the research process have to be acknowledged in the production/writing/dissemination of the research. This is known as a process of reflectivity. Good practice in cross cultural research is to reflect back and forward on the process, the role of the researcher, the intricacies of power relations played out through the research and the most effective and fair way of producing/writing the research. Reflectivity happens continually throughout and beyond the time of the research project.

See also: Culture; Cultural Geography; Feminism/Feminist Geography; Postcolonialism/Postcolonial Geographies.

Further Reading

Butz, D. and Besio, K. (2004). The value of autoethnography for field research in transcultural settings. *The Professional Geographer* 56(3), 350 260.

Dyck, I., Lynam, J. M. and Anderson, J. M. (1995). Women talking: Creating knowledge through difference in cross cultural research. *Women's Studies International Forum* 18(5 6), 611 626.

Herbert, J. (2007). Negotiating boundaries and the cross cultural oral history interview. In Rodger, R. & Herbert, J. (eds.) *Testimonies of the City: Identity, Community and Change in a Contemporary Urban World*, pp 251 267. Aldershot: Ashgate.

Howitt, R. (2005). The importance of process in social impact assessment: Ethics, methods and process for cross cultural engagement. *Ethics, Place and Environment* 8(2), 209 221.

Jazeel, T. and McFarlane, C. (2007). Responsible learning: Cultures of knowledge production and the North South divide. *Antipode* 39(5), 781 789.

Malam, L. (2004). Embodiment and sexuality in cross cultural research. *Australian Geographer* 35(2), 177 183.

Martin, P. M. and Glesne, C. (2002). From the global village to the pluriverse? 'Other' ethics for cross cultural qualitative research. *Ethics, Place and Environment* 5(3), 205 221.

Mohammad, R. (2001). 'Insider' and/or 'outsiders': Positionality, theory and praxis. In Limb, M. & Dwyer, C. (eds.) *Qualitative Methods for Geographers*, pp 101 117. London: Arnold.

Mullings, B. (1999). Insider or outsider, both or neither: Some dilemmas of interviewing in a cross cultural setting. *Geoforum* 30, 337 350.

Raghuram, P. and Madge, C. (2006). Towards a method for post colonial development geography? Possibilities and challenges. *Singapore Journal of Tropical Geography* 27, 270 288.

Raju, S. (2002). We are different, but can we talk? *Gender, Place and Culture* 9(2), 173 177.

Skelton, T. (2001). Cross cultural research: Issues of power, positionality and 'race'. In Limb, M. & Dwyer, C. (eds.) *Qualitative Methods for Geographers*, pp 87 100. London: Arnold.

Smith, F. (1996). Problematizing language: Limitations and possibilities in 'foreign language' research. *Area* 28, 160 166.

Staeheli, L. and Nagar, R. (2002). Feminists talking across worlds. *Gender, Place and Culture* 9(2), 167 172.

Twyman, C., Morrison, J. and Sporton, D. (1999). The final fifth: Autobiography, reflexivity and interpretation in cross cultural research. *Area* 31(4), 313 325.

Relevant Websites

http://www.anu.edu.au/culture/
Australian National University, Centre for Cross Cultural Research.

http://www.victoria.ac.nz/cacr/
University of Wellington Victoria, Centre for Applied Cross Cultural Research.

Cultural Capital

J. L. Waters, University of Liverpool, Liverpool, UK

Glossary

Credentialism In society, it refers to reliance on academic qualifications and credentials to confer jobs, status, and other rewards.
Gentrification A process by which buildings and neighborhoods are transformed and given value through physical renovation. It is usually associated with the practices of the middle class.
Meritocracy A social system that rewards merit (natural talent and ability) over unearned titles and privilege.
Neighborhood Effects Where the local social and geographical contexts shape the experiences and aspirations of the people living there.
Social Exclusion It refers to an inability to access resources considered normal for a particular society or social context. Exclusion usually occurs as a result of perceived social group membership.

Introduction

Cultural capital is most commonly associated with the work of French sociologist Pierre Bourdieu, who has written extensively on the relationship between capital in its various forms (cultural, social, symbolic, and economic) and the reproduction of class status in contemporary society. According to Bourdieu, cultural capital exists in three principal states: embodied in an individual person, objectified through cultural artifacts, and institutionalized in academic qualifications and credentials. It can be both intentionally accumulated through active self improvement and unconsciously acquired through socialization within the family.

Uses of the Concept of Cultural Capital

Research in the social sciences has made extensive use of Bourdieu's notion of cultural capital, referring to its three distinctive states. To begin with embodied cultural capital, this describes aspects of a person's culture or social standing as revealed through the attributes of the body such as a muscular physique, dress sense, posture, or accent – qualities that are difficult to objectively quantify. This type of cultural capital necessarily takes time and a degree of effort to acquire and cannot simply be passed on (as in a gift or purchase) from one person to another. Research has shown, for example, that immigrants will often deliberately seek to attain cultural capital that is valued by the host society, aiding the process of integration and, eventually, full citizenship. Aihwa Ong has observed that government and nongovernment institutions in the US are often directly involved in the process of turning refugees into new citizens through attention to their bodies. In these examples, embodied cultural capital can include language and 'know how' as well as bodily comportment, food and medical preferences, and dress sense.

Objectified cultural capital describes the possession and coveting of valued material artifacts such as paintings, a type of car, or a piece of rare furniture by an individual or family. Unlike its embodied counterpart, this type of cultural capital can be instantaneously transmitted from one person to another, is far easier to measure, and appears to have an objective and enduring market value unrelated to current societal trends. This appearance, Bourdieu suggests, is misleading: the value of objectified forms of cultural capital is constantly reinforced by and through the actions of individuals and groups who have a stake in policing the boundaries of acceptable taste in order to maintain the value of their own cultural capital.

Institutionalized cultural capital – the third state in which cultural capital can exist – describes the value ascribed to the possession of formal academic qualifications, which provides the bearer with an institutionally recognized guarantee of their cultural competences. This conception of cultural capital has had the biggest impact on research in the social sciences to date. Institutionalized (like objectified) cultural capital appears to possess an unambiguous value, providing an objective measure of attainment and enabling meritocratic forms of social organization. In fact, Bourdieu argues, meritocracy is a myth; formal educational qualifications serve to reinforce existing class boundaries in two important ways. First, schools tend to reward the cultural capital already possessed by the children of middle class backgrounds; formal examinations are not objective assessments of ability or intelligence but, in fact, evaluate a child's knowledge and understanding of what are predominately middle class cultural values and concerns (such as music, art, literature, and history), thereby inadvertently neglecting and devaluing alternative (e.g., working class) forms of culture. Second, the value of qualifications is not independent of society but rather depends upon a continual process of recognition enacted by employers, school teachers, school and university alumni (such as 'old boy networks'), and professional associations. These actors police the boundaries of privilege. The concept of cultural capital has, therefore, been primarily used to explain the process by which class status is reproduced through formal education and the rise of credentialism in advanced industrial societies.

Spatial Dimensions of Cultural Capital

According to Bourdieu, the structure of society (and classes within society) reflects the distribution of the different types and subtypes of capital at any given mo ment in time. This claim is implicitly geographical – access to and availability of capital varies spatially as well as between social groups.

The Spaces and Places of Cultural Capital

While there is a dearth of scholarship specifically examining the spatial dimensions of cultural capital, contemporary work by geographers on gentrification and social exclusion, as well as emerging scholarship on education and local labor markets, have emphasized the important relationship between cultural capital and place. Research has examined different social and eco nomic dimensions of spatial variations in cultural capital – some places are particularly rich in cultural capital while others experience a notable deficit. This geography has important implications for developing a better understanding of enduring inequalities within society.

Bourdieu's concept of cultural capital has significantly informed recent work attempting to illuminate the mechanisms underlying the gentrification process. The increasing presence of cultural capital can help explain why particular 'less desirable' neighborhoods become desirable (first 'culturally', before becoming economically profitable). According to David Ley, the process of gen trification is partly attributable to the commodification of the 'esthetic sensibilities' of artists, who are rich in cul tural capital but relatively poor in economic capital. In choosing a residence, artists often deliberately seek out unusual, less conventionally elegant places. The presence of artists in an area (artists are often viewed as the har bingers of fashionable taste) has the effect of increasing its cultural capital, which over time becomes converted into economic capital as more and more people buy (literally) into the culture of the area, purchasing, through housing, an objectified form of cultural capital.

While places can confer cultural capital on individuals, they can also, conversely, deny it. Residing in areas devoid of (what is commonly considered to be) valuable cultural capital can have a detrimental impact upon life chances. This has been shown by work examining the importance of neighborhood effects on educational attainment, implying that a lack of existing cultural capital within an area impedes its further accumulation. Within these neighborhoods, children generally have low educational attainment while parents are more likely to be un employed or working in low skilled, low paid jobs. Low levels of cultural capital in an area (which can include well staffed, well resourced schools, leisure facilities, ac cess to transportation, quiet places to study, resources such as books and computers, friendships that valorize studiousness and parental support with schoolwork) en sure that children are unable to break free from a cycle of underachievement, perpetuating high levels of place specific social deprivation. The presence or absence of cultural capital is seen to be instrumental in persistent forms of social exclusion.

Closely related to this, research in geography has examined the construction and reproduction of so called stigmatized places – places widely perceived to be sig nificantly lacking in cultural capital. These places contain and reproduce alternative, unvalued, and ostensibly dysfunctional forms of culture. This is explored in the work of Harald Bauder, who has examined the relation ship between deprived inner city neighborhoods, stig matization, and cultural capital. Frequently associated with an urban underclass, these neighborhoods experi ence multiple forms of deprivation in spheres such as housing, schools, employment, and public amenities. However, Bauder is critical of much existing research on neighborhood effects and cultures of poverty, which identifies what he describes as a 'neighborhood based cultural pathology' in explanation. According to this in terpretation, the absence of cultural capital in an area is self perpetuating and inevitable; the problem lies within (and not external to) the neighborhood itself. As Bauder suggests, what is important is not the presence or absence of cultural capital *per se*, but how that presence or absence is actively constructed and ascribed by individuals and groups with the power to affect the life chances of those who live within these neighborhoods, such as employers, schools, and service delivery institutions. This underlines the fact that cultural capital exists as a potent force only insofar as it is given meaning and value by wider society.

Spatial Strategies in the Accumulation of Cultural Capital

As capital is unevenly distributed over space, so mobility affects the capacity of individuals and groups to acquire cultural capital. In general, research has shown that members of the middle class demonstrate greater spatial mobility than their working class counterparts.

Middle class families can be often seen to employ mobility in a strategic fashion when accessing different forms of capital. Recent work has highlighted the importance of spatial strategies when examining the accumulation of cultural capital through education. Re search by Stephen Ball, Richard Bowe, and Sharon Gewirtz has examined the interplay between class, cul tural capital, and educational choice in Greater London. Parents are divided into working class locals and middle class cosmopolitans: the former group typically choose the local school for their child, reflecting the practicalities of day to day living, while the latter group will frequently travel outside the locality, engaging with the education market in a more strategic fashion and maximizing, through spatial mobility, the cultural capital available to their child. Tim Butler's work, also on London, demon strates similar findings regarding the tendency of middle class parents to send children outside their neighborhood for secondary schooling. This serves to isolate children from the local community while at the same time re inforcing existing forms of middle class cultural capital through socialization with 'people like themselves'. Travel and mobility are key strategies in the deliberate accu mulation of cultural capital.

These findings are paralleled in research by Waters that considers the transnational accumulation of cultural capital through international forms of education. Waters has shown how middle class immigrant families from Hong Kong use international migration to accumulate cultural capital (in both its embodied and institutional ized states) and this is achieved through the children's acquisition of a 'Western' education. When taken back to Hong Kong, the Western university degree confers sig nificant cultural capital (and economic capital in the labor market), securing and strengthening families' middle class status on their return. These ideas have been developed in recent work examining the geo graphical dimensions of an emergent international edu cation industry, predicated on spatial variations in the value of institutionalized cultural capital.

These examples highlight the importance of a geo graphical perspective on cultural capital. The value of cultural capital is not fixed but varies spatially, according to the scale of analysis, and is embedded within particular societal contexts.

See also: Gentrification; Mobility; Neighborhood Effects; Transport and Social Exclusion.

Further Reading

Bauder, H. (2001). Culture in the labor market: Segmentation theory and perspectives of place. *Progress in Human Geography* 25, 37 52.

Bauder, H. (2002). Neighbourhood effects and cultural exclusion. *Urban Studies* 39, 85 93.

Bell, D. and Valentine, G. (1997). *Consuming Geographies: We Are Where We Eat*. London: Routledge.

Bondi, L. and Matthews, M. H. (eds.) (1988). *Education and Society; Studies in the Politics, Sociology and Geography of Education*. London: Routledge.

Bourdieu, P. (1984). *Distinction: A Social Critique of the Judgement of Taste*. Cambridge, MA: Harvard University Press.

Bourdieu, P. (1986). The forms of capital. In Richardson, J. G. (ed.) *Handbook of Theory and Research for the Sociology of Education*, pp 241 258. New York: Greenwood Press.

Bourdieu, P. and Passeron, J. (1977). *Reproduction in Education, Society and Culture*. London: Sage Publications.

Bridge, G. (2001). Estate agents as interpreters of economic and cultural capital: The gentrification premium in the Sydney housing market. *International Journal of Urban and Regional Research* 25, 87 101.

Butler, T. (2003). Living in the bubble: Gentrification and its 'others' in North London. *Urban Studies* 40, 2469 2486.

Dyck, I. (2006). Travelling tales and migratory meanings: South Asian migrant women talk of place, health and healing. *Social and Cultural Geography* 7, 1 18.

Ley, D. (2003). Artists, aestheticisation and the field of gentrification. *Urban Studies* 40, 2527 2544.

Ong, A. (1999). *Flexible Citizenship; the Cultural Logics of Transnationality*. Durham: Duke University Press.

Ong, A. (2003). *Buddha is Hiding: Refugees, Citizenship, The New America*. London: University of California Press.

Warrington, M. (2005). Mirage in the desert? Access to educational opportunities in an area of social exclusion. *Antipode* 37, 766 816.

Waters, J. L. (2005). Transnational family strategies and education in the contemporary Chinese diaspora. *Global Networks* 5, 359 378.

Waters, J. L. (2006). Geographies of cultural capital: Education, international migration and family strategies between Hong Kong and Canada. *Transactions of the Institute of British Geographers* 31, 179 192.

Cultural Economy

A. C. Pratt, London School of Economics, London, UK

Glossary

Cultural Industry A term coined by Adorno to negatively characterize the mass production (and consumption) of culture.

Cultural Industries A term developed by French scholars, and developed by British writers, that stresses the variety of cultural production, and underlies a positive role in economic development.

Creative industries A term used by the UK government to promote its intellectual property/content industries. The term has been adopted internationally. The adjective 'creative' is not as controversial as 'cultural', and creativity seems to align it to the information economy.

Economization The notion that cultural has been drawn into economic production (and separated from the arts).

Culturalization The idea that the economy as a whole has become more 'cultural' in the sense that values and design rather than simple price are market signals.

Cultural Industries Production System A term that expresses the span of material cultural production processes across production and consumption, and all points in between.

The Cultural Mode of Economic Action: The Cultural Economy

Arguably the longest running debate about cultural economy can be tracked more precisely to its status as a 'mode of analysis' and to debates about the role of eco nomic sociology as a 'subdiscipline' and its position vis à vis economics. In the early variants of classical Eco nomics what we now consider the social, cultural, and political were an internal part of Economic discourse. Later, neoclassical analyses sought to concentrate on atomistic, uniquely economic activities described in terms of rational calculation. In effect, this created a 'boundary dispute' with respect to the social. In soci ologist Talcott Parsons' influential formulation of Eco nomic Sociology, this divide was formalized (and remained so for 40 years or more). In Parsons' view, neoclassical Economics should be separate from the so cial: Sociology was allocated the task of exploring the 'social setting' for economic action, economics was left with the 'pure Economics'. Thus the social (and cultural) dimensions of the economy were allocated to 'context'. For many years Economic Sociology, and Geography (which drew on the same formulations), upheld this distinction. It is only relatively recently that writers have sought to challenge this dualism (the economic and the social); in so doing they have opened up the field for a reconsideration of the explanation of economic action proper and thus challenged neoclassical Economics on its own grounds. This is the key to understanding the cur rent re engagement with the cultural economy.

This is a complex and emerging field. In brief we can identity at least six lines of current debate about the cultural economy. There is no space to elaborate on these here. First, researchers have sought to show how eco nomics is one rhetoric amongst others; it is a mode of argumentation rather than revealed or undisputed truth. Second, others have sought to demonstrate how truth claims are sustained in economics (and science) via a web of correspondences. Third, in an anthropological turn some have sought to examine the practices by which economic life are carried on. In other words to see how economic life 'is done'. We can note the contrast with economists who argue about how economic life 'should be' done. Fourth, we can point to anthropologists of material culture who have rendered the economic field into the familiar language of kinship. Fifth, alongside these has been a Deleuzian influenced debate about performativity and affect which has on one hand highlighted the role of the 'passions' in economic life, and on the other hand, sought to recast the economic field within a post representational mode. While these lines of theory are in essence Economic Sociology, they have also shaded into Economic Anthropology, and more generally come to encompass the cultural turn in the social sciences; hence cultural analyses of the economy. In so doing, the cultural turn has been strongly articulated as a methodological critique against quantitative and in favor of qualitative approaches. Sixth, we can identify a dimension of debate that has sought to explore the moral dimension of social and economic life; for some writers, this has been aligned to a critical take on the 'cultural turn'.

Much of the theoretical debate about the 'cultural turn' in the discipline of geography has been positioned against Marxist structuralism and production, while theoretical debates about postmodernism and post structuralism highlighted the play of images. In the practice of advertising, filmmaking, and similar forms, it was used to frame the consumption experience. Although not a necessary consequence of such an approach, more often than not the object of interest has been consumption. There are a lack of new Cultural Studies, Economic Geographies, or Cultural Geographies that apply their attention to production, let alone production 'and' consumption. In Economic Geography the 'cultural turn' was in part figured as a reaction to the economically reductivist, or productivist, focus. Such a shift can be evidenced by the renewed interest in business organiza tion; some researchers have examined the unusual forms of economic organization, namely 'project working', that are commonly found in the cultural economy. In parallel, there has been a growth in management literatures concerning the cultures of management, and the par ticular innovative practices that constitute the creative/ cultural industries. The instrumental focus of this knowledge seeks to achieve competitive advantages through innovation or product design.

A common dualistic couplet deployed in the analysis of the cultural economy is that of culturalization–econ omiztation. As a dualism this is, on the one hand, pre sented as economic activities being increasingly inflected with cultural topes and performance, etc., rather than simple differentiation of value. This has been taken as an account for the deployment of culture to increase market segmentation and develop niche markets in a circum stance of oversupply, or market saturation. Some writers make a link with flexible specialization and what they term a reflexive turn, that is indexed to growing demand from consumers for an ever differentiated product mar ket (by design) and the organizational and productive means to satisfy it (flexible specialization). Hence, it is argued, culturalization is endemic in late capitalism.

Economization is commonly overlain with a moral sentiment. This is the classic formulation of Adorno writing in the late 1930s that rejected commodification as separating art from its aura. It is from this perspective that Adorno labeled (and defined) the Culture Industry in such negative terms. At one and the same time he sought to legislate what was culture, and to equate this with the market, thus creating an unhelpful market/ nonmarket equation with art and commerce. Later work by political economists – common In Media and Com munications studies – has carried on Adorno's moral baggage and added its own critique of power and control.

Returning to the foundational provision of cultural goods, cultural policy has been based upon market failure associated by inadequate price signals. Thus, 'creeping economization' is feared to morally 'devalue' culture; hence, the necessity of public intervention to counter these 'incorrect' market signals. Other binaries such as high–low culture, or public–private, for profit–not for profit have served to reinforce the notion that commercial/mass cul ture/private/for profit are aesthetically inferior, despite the fact that the arguments (on one hand Adorno's 'aura', and on the other, market failure) are different.

The cultural studies tradition has also argued that the high–low culture distinction is false, but from a different discursive point of view. It is argued that there is a case, albeit not in such clear terms, to promote forms of cul ture that are not commercially orientated in order to promote diversity or representation. There is a long history of the analytical elision of high culture with higher social classes. Against this the value of the cultural studies tradition has been to recognize that culture is 'classed' (and 'spaced'); that is, it is associated with, and the product of, particular class fractions. Thus, the insight that in favoring one particular esthetic, one serves to support and legitimize that particular class (or regional) position and worldview.

Increasingly culture is being used in instrumental ways to promote cities – for example, by creating dis tinctive modern art galleries. The argument here is that cultural differentiation is a way of marking place rooted uniqueness (inherent or created). A line of argument in continental Europe has been the discussion of local food cultures, or food products that differentiate place and markets, and are supported by 'regional branding'. A different, commercial, articulation is the increasing use of the 'experience economy' to promote a 'feel good ex perience' that helps to sell (otherwise similar) goods and places: from people dressing up and performing service, to street theater and animateurs. Ethnographic research has highlighted the 'hidden' or emotional labor in such practices. The latest version of this is the use of cities to attract what has been termed the 'creative class' because this group of workers who work in hi tech are attractive to new hi tech companies who are drawn to labor pools. The key point that detracts from this work, for the purposes of this article, is that it is focused on cultural consumption, not production; and, that culture is being deployed in an instrumental manner.

The Economy of Cultural Products and Services: The Cultural Economy

If we turn our attention to the part of the economy dealing with the production of cultural goods and ser vices, the 'cultural economy', then we are immediately faced with two problems. First, what are its boundaries and definitions? and second, why has it grown? Em pirically, these two problems are interlinked as the taxonomies used for empirical data collection on eco nomic activities are based upon historically determined conceptions of the economy which are themselves con structed around concepts of a normal (old) industrial economy and are intrinsically ill prepared to identify 'newness'. This is a problem associated not only with culture, but one with new technology based industries, and the service sector more generally.

The section titled 'The cultural mode of economic action: the cultural economy' has already differentiated the 'cultural economy' from the 'cultural' economy, the latter would include the cultural dimensions of the whole economy. Adorno, when he coined the term 'culture industry', was seeking to point to those parts of the economy that were mass producing what he identified as inferior culture. A radical transposition of the notion was developed by French Media and Communications scholars in the 1980s; they sought to take this economy seriously, and to recognize that the 'Culture Industry' was a legitimate expression of culture and that its production was 'plural' and various: they renamed it as 'the Cultural Industries'. This notion has been discussed in the UK (and much copied elsewhere), and underpins the policies of the 1980s for cultural regeneration pursued by UK metropolitan authorities. Conceptually, two lines of emphasis have emerged from this root: first, those that stress the textuality of the cultural industries and have a closer theoretical lineage to Media and Communications studies; and second, those that stress the production system of cultural products and services that have a closer alliance with Economic Geography. Yet another term has recently become popular with policy makers: the creative industries, although there are critical differences in the interpretation of this term in practice usage overlaps.

The cultural industries production system approach has been proven important for policymakers; it seeks to join both questions of the 'breadth' of culture (activities that should be included: film, television, books, computer games, theater, music, etc.), and the 'depth' of cultural production (activities that are required to produce cultural outputs: manufacturing, distribution, and consumption). Some analyses have concentrated on the former and operationalized this through counting numbers involved in artistic occupations, other analyses have explored cultural trade. The cultural industries production system approach argues that occupation – as used by Florida – offers partial analyses, as well as failing to capture the social reproduction of labor and knowledge in the cultural industries. Andy Pratt has developed a statistical framework that seeks to both define and operationalize both the breadth and depth definitions which focus on 'industry' rather than 'occupation' as a fundamental unit of analysis; it is an approach that has been subsequently taken up by national governments and United Nations Educational, Scientific and Cultural Organization (UNESCO) in the production of so called 'mapping documents'. One outcome of such studies has been the fact that the creative/cultural industries/economy constitute around 10% of some economies; indeed, they are the third major industry in London. Overlain on this political driven discourse has been a further confusion concerning the terms the creative industries, or the cultural industries: moreover, the cultural economy and creative economy.

Alongside enumerating and defining the cultural economy there has been a debate concerning its spatiality. The cultural economy has a distinctive geography, one that is strongly, although not exclusively, articulated to urban areas in the developed world. A particular policy driven debate has concerned the agglomeration of cultural industries, so called 'clusters'. While policy makers have, as with other industries, sought to promote clusters, there is currently a poor understanding of the process of interaction at the local scale, empirical work on the cultural economy points to the role of knowledge, reputation, and untraded dependencies; moreover, these are both localized and face to face, and internationalized through in such nodes. This articulation of place and globalization in the cultural economy is taken up in the debate about cultural industry commodity chains.

Once we have accepted that the cultural economy does exist, and that it has grown very rapidly from a relatively insignificant economic and cultural low point, we must consider the question 'why?' This is an area of ongoing research. There are two lines of debate. First, without doubt, arguments concerning the growth of income, leisure time, and the proportion of disposable income spent on culture account for one part of the 'demand'. However, to this we need to add the creation of demand through the development of various means of persuasion (advertising) and the social transformation wrought by the extension of full time education, and increased participation in tertiary education. This latter process is linked to the creation of a cohort of young people with few responsibilities and a considerable amount of money and time. Added to this economic potential is the creation, from the mid 1950s onward, of the social identity of 'teenager' that has been heavily inflected with advertising. Finally, the cultural explosion and liberal forms of artistic expression that developed in the 1960s. There is no space to explore these linkages, although it goes without saying that their interrelationships need care in interpretation.

A second theme of accounts of the emergence and growth of the cultural economy is via an extension of the work on the postindustrial modernization of society. Researchers have proposed that the information, knowledge, and creative sectors of the economy constitute the 'quaternary sector', succeeding the agricultural, manufacturing and service sectors as the 'highest stage' of growth. This line of argument has been criticized arguing that the linkage to production is empirically evident (and important) rather than absent, as implied in the postindustrial conceptualization.

Beyond these disputes about the long term development of economies and the role of the cultural economy in such a transformation are a number of important empirical changes that research is highlighting. First, as noted above, the realization that the cultural economy is becoming an important aspect of some developed economies, in terms of turnover, employment, exports, as well as meaning. Second, as also noted above, is the fact that the cultural economy does seem to have some particular patterns and processes of location, practice, and organization. Third, that the nature of work and employment in these industries is distinctive. There has been much discussion about the distinctive rise of flexible work and self employment, the concentration of freelancing, and casualization of labor, plus the degree of serial contracting and project working, found in the cultural economy. Researchers are increasingly exploring these processes, and their variance across both the creative economy, plus the extent to which they are found in the rest of the economy. Finally, work on production and commodity chains has highlighted the importance of the recursive linkages between production and consumption processes and places.

If we add this increasing fragmentation of production processes with the other characteristic highlighted by traditional media and communications scholars, namely the degree of media (structural) concentration, and mix in the diverse and unique spatial clusters explored by geographers, it is clear that the cultural economy will keep researchers busy for many years to come.

See also: Creativity; Cultural Geography; Economic Geography; Innovation.

Further Reading

Adorno, T. and Horkheimer, M. (1977). The culture industry: Enlightenment as mass deception. In Curran, J., Gurevitch, M. & Woollacott, J. (eds.) *Mass Communications and Society*, pp 349 383. London: Arnold.

Amin, A. and Thrift, N. (eds.) (2004). *The Blackwell Cultural Economy Reader*. Oxford: Blackwell.

Bathelt, H., Malmberg, A. and Maskell, P. (2004). Clusters and knowledge: Local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography* 28, 31 56.

Bell, D. (1973). *The Coming of Post Industrial Society*. New York: Basic Books.

Blair, H. and Rainne, A. (2000). Flexible firms. *Media Culture and Society* 22, 187 204.

Callon, M. (ed.) (1998). *The Laws of the Markets*. Oxford: Blackwell Publishers.

Coe, N. M. (2001). A hybrid agglomeration? The development of a satellite Marshallian industrial district in Vancouver's film industry. *Urban Studies* 38, 1753 1775.

Crang, P. (1994). It's showtime: On the workplace grographies of display in a restuarant in Southeast England. *Environment and Planning D: Society and Space* 12, 675 704.

Crewe, L. (1996). Material culture: Embedded firms, organizational networks and the local economic development of a fashion quarter. *Regional Studies* 30, 257 272.

Crewe, L. and Beaverstock, J. (1998). Fashioning the city: Cultures of consumption in contemporary urban spaces. *Geoforum* 29, 287 308.

DCMS (1998). Creative industries mapping document. *Report of DCMC*. London: Department of Culture, Media and Sport.

Ferguson, M. and Golding, P. (1997). *Cultural Studies in Question*. London: SAGE.

Florida, R. L. (2002). *The Rise of the Creative Class: And How It's Transforming Work, Leisure, Community and Everyday Life*. New York: Basic Books.

Garnham, N. (1987). Concepts of culture public policy and the cultural industries. *Cultural Studies* 1, 23 37.

Garnham, N. (2005). From cultural to creative industries: An analysis of the implications of the 'creative industries' approach to arts and media policy making in the United Kingdom. *International Journal of Cultural Policy* 11, 15 30.

GLA Economics (2004). Measuring creativity: 2004 update of the GLA's creative industry economic data. London: Greater London Authority.

Grabher, G. (2002). The project ecology of advertising: Tasks, talents and teams. *Regional Studies* 36, 245 262.

Hannigan, J. (1998). *Fantasy City: Pleasure and Profit in the Postmodern Metropolis*. London: Routledge.

Hesmondhalgh, D. (2002). *The Cultural Industries*. London: SAGE.

Hesmondhalgh, D. and Pratt, A. C. (2005). Cultural industries and cultural policy. *International Journal of Cultural Policy* 11, 1 14.

Hochschild, A. R. (1983). *The Managed Heart: Commercialization of Human Feeling*. Berkeley, CA: University of California Press.

Law, J. (1991). *A Sociology of Monsters: Essays on Power, Technology and Domination*. London: Routledge.

Markusen, A. (2006). Urban development and the politics of a creative class: Evidence from the study of artist. *Environment and Planning A* 38, 1921 1940.

McCloskey, D. N. (1986). *The Rhetoric of Economics*. Brighton: Wheatsheaf.

Miege, B. (1987). The logics at work in the new cultural industries. *Media Culture Society* 9, 273 289.

Miller, D. (1997). *Capitalism: An Ethnographic Approach*. Oxford: Berg.

Molotch, H. L. (1996). LA as design product: How art works in a regional economy. In Scott, A. J. & Soja, E. (eds.) *The City: Los Angeles and Urban Theory at the End of the Twentieth Century*, pp 225 275. Berkeley, CA: Unversity of California Press.

Palladino, G. (1996). *Teenagers: An American History*. New York: Basic Books.

Parsons, T. (1961). *The Structure of Social Action: A Study in Social Theory with Special Reference to a Group of Recent European Writers*. New York: Free Press of Glencoe.

Pine, J. P. II. and Gilmore, J. H. (1999). *The Experience Economy: Work Is Theatre and Every Business a Stage*. Boston, MA: Harvard Business School.

Pratt, A. C. (1997). The cultural industries production system: A case study of employment change in Britain, 1984 91. *Environment and Planning A* 29, 1953 1974.

Pratt, A. C. (1997). Production values: From cultural industries to the governance of culture. *Environment and Planning A* 29, 1911 1917.

Pratt, A. C. (2000). New media, the new economy and new spaces. *Geoforum* 31, 425 436.

Pratt, A. C. (2004). The cultural economy: A call for spatialized 'production of culture' perspectives. *International Journal of Cultural Studies* 7, 117 128.

Pratt, A. C. (2005). Cultural industries and public policy: An oxymoron? *International Journal of Cultural Policy* 11, 31 44.

Sayer, A. (2003). *Valuing Culture and Economy*. Lancester: Department of Sociology, Lancaster University.

Scott, A. J. (2000). *The Cultural Economy of Cities: Essays on the Geography of Image Producing Industries*. London: Sage.

Thrift, N. J. (2005). *Knowing Capitalism*. Thousand Oaks, CA: Sage Publications.

Throsby, D. (2001). *Economics and Culture*. Cambridge: Cambridge University Press.

Warhurst, C., Nickson, D., Witz, A. and Cullen, A. M. (2000). Aesthetic labour in interactive service work: Some case study evidence from the 'new' Glasgow. *Service Industries Journal* 20, 1 18.

Cultural Geography

C. Gibson and G. Waitt, University of Wollongong, Wollongong, NSW, Australia

Glossary

Affect In verb form, to influence someone or something; to produce a mental or emotional effect. In noun form, the conscious subjective dimension of emotion or feeling.

Agency A capacity to control actions or exert power.

Binary A system of ideas or theories resting on two parts or components, often diametrically opposed antonyms.

Discourse Literally, a conversation or exchange of verbal expressions. In cultural analysis, sets of statements that through language and social circulation make people, plants, places, and things knowable and understandable.

Representational An epistemological perspective in the humanities which emphasizes that all things, people, and landscapes have symbolic or semiotic meanings separate from consideration of their material or literal existence. These meanings can be 'read' or 'interpreted' in cultural research using methods of deconstruction and discourse analysis.

An Introductory Note on the Meaning of 'Culture'

'Cultural geography' is now accepted as a major subfield of human geography. However, its update in teaching and research is relatively recent. What is cultural geography shifts with context, philosophical traditions, and most importantly, rests on one's understanding of 'culture' and 'the cultural'. There are many variations on the manner in which 'culture' is used to describe a type of geographical knowledge. A literal reading of the term in a verb form links 'culture' to practices of cultivation – to grow, extend, feed, or rear – as in 'agriculture', or in the production of material objects and artifacts in particular places and times. In the early twentieth century, another popular meaning was a modern, 'humanist' one that defined 'culture' as something possessed by a society having progressed or 'civilized' to develop a set of shared norms, behaviors, and rationalities (hence the related term, 'cultured'). In this vein, culture has often been used as a synonym to describe the creative expressions, and intellectual and esthetic achievements of talented people within particular societies. People who understood and valued these intellectual and creative expressions were often positioned as critics, agents, and managers – essentially arbiters of taste responsible for differentiating between what was considered 'high art' and 'common', popular culture, such as popular songs. However, perhaps the most common definition of 'culture' is an identifiable 'way of life'. In this definition, culture is understood as a set of shared values and beliefs based on language, religion, customs, and ethnicity. These cultural traits are then expressed in material cultures, including creative expressions, clothing, foods, buildings, and occupations.

In geography, this latter definition – of culture as a way of life – remained uncontested for over six decades from the 1920s. In geographical research, the idea of culture as a way of life was often deployed to make generalized claims about all people within a particular geographically bounded area, termed a cultural realm or region. So, for Rubenstein, culture is:

> the body of customary beliefs, material traits and social forms that together constitute the distinct tradition of a group of people. Geographers distinguish groups of people according to important cultural characteristics, describe where particular cultural groups are distributed, and offer reasons to explain the observed distribution. (Rubenstein, 2005: 24)

Culture as a way of life is approached as a unitary variable that is able to explain a spatial pattern. Following this approach, 'culture' is seemingly distinct from politics, economics, or demographics. However, many contemporary cultural geographers often resist any attempt to approach culture as a cause, or indeed define culture as a 'thing', 'type', or even 'subject field'. Drawing on a number of geographical traditions, such as feminism and postcolonialism, they challenge the unquestioned assumption of the 'realness' of culture in the 'way of life' interpretation, arguing that it is impossible to analyze culture to reveal intrinsic properties. Indeed, if Don Mitchell is correct when he argued that "there is no such thing as culture," by inference, 'cultural geography' also ought to be considered a fabricated category of academic knowledge, without any real material existence.

Such arguments are informed by a less everyday understanding of 'the cultural': as knowledge production. Culture as a 'way of thinking and knowing' requires researchers to address two fundamental philosophical questions: first, what exists in the world, and, second, how do we know, what we know about the world? This way of understanding culture – as a way of thinking and knowing – has become particularly germane with the advent of post structuralist influences in geography. Post structuralist theories encourage researchers to identify links between knowledge production, power, practices, behaviors, and possibilities for change and transformation. Thinking about culture as knowledge production is particularly relevant when attempting to understand how people make sense of their lives and the spaces and places they live in and move through. Culture as knowledge production is informed by a concern for how we make sense of ourselves and the world through the ongoing, unfolding, and emergent relationships – be they fleeting or more permanent – between animals, plants, material things, and people. Culture as a way of thinking and knowing is informed by a concern for examining how particular ideas are established, circulated, maintained, and challenged.

Clearly, given these conflicting definitions, understanding what is meant by the word 'culture' is not straightforward. Equally, understanding what is meant by 'cultural geography' is not a simple task, given that these competing definitions of culture have seeped into academic geography in various times and places, and have enabled particular ways of doing cultural geography. Attempting to decide what defines a singular 'essence' or meaning for cultural geography is therefore impossible. Instead, our intention is to acknowledge multiple and particular narratives about 'cultural geography'. We have enacted this in the structure we use in the remainder of this article. Rather than a singular narrative, the account of cultural geography offered here is written as if we were directors of the opening scenes of a motion picture – *Cultural Geography: The Movie* – structured into a series of three parallel 'takes', each of which could replace the others depending on personal preferences. To some, this experimental structure may seem a little too clever, or may be outright annoying if all one wants is a quick reference guide and an explanation without fuss. Nevertheless, this film making approach serves a number of purposes. First, rather than offering one grand narrative, our three 'takes' provide a more open ended way to understand cultural geography. Each 'take' provides a particular angle on the intellectual imperatives of cultural geography. While each take builds up the picture, even when read together they only provide a particular interpretation of the stories of cultural geography. Second, while paying attention to origins, transformations, and successions, our 'takes' allow us to point toward themes, relationships with other disciplines, and address political considerations. Third, our three takes highlight how this article on cultural geography is 'directed' from a particular perspective. In other words, whatever working explanation is given is always situated in relationship to the authors' own geopolitical and institutional position. Critical analysis of geographical knowledge has emphasized its partiality and situatedness. Our writing perspective is informed not only by our many axes of difference – including our gender, age, language, ethnicity, and training within Anglo American geographic traditions – but also betrays our institutional position in Australia, where geography continues to be fashioned by Anglo American powers and interests. Our knowledge of what we think as cultural geography is therefore partial, situated, and anchored in our own context.

Finally, and related to our third point, is that 'cultural geography' is a heuristic device that now means certain things in English speaking geographical traditions. Embedded in contemporary understandings of cultural geography are meanings that dominate over those others emanating from some non English speaking worlds. In Australia, we are in a place where it is acutely possible to reflect upon this. Our position allows us to be immersed in and make to contributions knowledges that are labeled as cultural geography in Anglophone academic geography, but we are also aware of the manner in which they do not reflect the cultural geographical knowledge production undertaken in nonacademic, everyday circumstances by other social groups. Such groups are immediately in our midst, for instance, Australian aboriginal groups, such as the Anangu, Bunjalung, or Wadi Wadi. These groups have particular ways of interpreting the relationship between 'culture' and geographical space, and they create knowledges about this relationship (for instance, through visual artistic practices and oral histories). Definitive statements about subdisciplines thus invariably obscure the productive sites of cultural geographical knowledge in non Anglophone and nonacademic worlds. Yet – and herein lies a crucial paradox – we do not (and cannot) seek the role of benevolent spokesperson for these non Anglophone or nonacademic geographies in some universal manner. Although it is possible to identify epistemological plurality, we can never assume to be speaking 'authentically' across the myriad of cultural differences that exist in the world.

For these reasons, we have adopted the format below of three 'takes' on cultural geography. Rather than giving the reader a sense of comfort that the boundaries of cultural geography are knowable and fixed, instead our aim is to challenge readers to recognize the contributions we have identified as belonging to a subfield of cultural geography, and yet also continually question them. We suggest that retaining a sense of discomfort and ambiguity around the subfield of cultural geography is important, and more 'true' to the intentions and imperatives of cultural geography itself.

Cultural Geography, Take One: In the Beginning

Our first take follows a conventional narrative plot that begins with 'origins' and a 'classical period', then unfolds in a linear narrative of ongoing progress of 'new', 'newer', and 'newest' cultural geography. This will give the reader a sense of comfort typical with linear, 'progressive' stories, and it will suggest that the boundaries of cultural geography are knowable, periodic, and fixed. This will be deliberately challenged in the takes that follow.

Classical cultural geography is conventionally traced back to origins in the 1920s, with the work of Carl Sauer and his colleagues at the University of California, Berkeley, United States of America. The 'Berkeley School', as it would become known, embedded an understanding of culture as both 'cultivation' – to grow or rear – and as 'way of life'. Carl Sauer coined the term 'cultural landscape' to describe the manner in which place was 'fashioned from a natural landscape by a cultural group'. For Sauer,

> culture [as a way of life] was the agent, the natural area the medium, the cultural landscape… the result [of cultivation]. Under the influence of a given culture [as way of life], itself changing through time, the landscape undergoes development, passing through phases, and probably reaching ultimately the end of its cycle of development. With the introduction of a different that is alien culture [or way of life], a rejuvenation of the cultural landscape sets in, or a new landscape is superimposed on remnants of an older one. (Sauer, 1925)

Hence cultivation and way of life were intimately linked through the concepts of cultural and natural landscapes. Groups of humans with discrete population sizes, densities, mobilities, housing styles, agricultural styles, and social customs – in short, cultures with particular ways of life – would literally transform the prehuman natural landscape by cultivating a new cultural landscape. Consistent in much Sauerian cultural geography, even into the 1970s, was a 'superorganic' or 'cultural determinist' approach. Culture was a 'whole', rather than an amalgam of the actions of individuals:

> We are describing a culture, not the individuals who participate in it. Obviously, a culture cannot exist without bodies and minds to flesh it out; but culture is also something both of and beyond the participating members. Its totality is palpably greater than the sum of its parts. (Zelinsky, 1973: 40)

In Rowntree's words, Sauerian cultural geographers "depicted the personality of geographical space in historical perspective." This approach – especially followed in North America in the decades following Sauer – tended to examine the geography of the material cultural landscape, organized, patterned, and located usually in a rural context at the regional scale. Common topics included the study of the diffusion of rural farming practices, modes of agrarian life, distributions and patterns of material cultural products (from vernacular architectural styles to musical instruments), and culturally specific land use practices.

There is further historical context also worthy of brief explanation: in the 1920s Sauer was reacting against a particularly mechanistic approach to understanding relations between humans and nature – environmental determinism – which had dominated geography until that time. Environmental determinists sought to identify causal links between ecological and terrestrial variations and cultural appearances, traits, and behaviors across the Earth's human population distribution. Environmental determinists were prominent in Europe (e.g., Mackinder and Ratzel) and their disciples brought it back to America (e.g., William Morris Davis and Ellen C. Semple) and Australia (e.g., Griffith Taylor), under the banner of 'anthropogeography' or sometimes, more simply, 'human geography'.

Environmental determinists sought not only to describe culture as way of life, but also heavily emphasized a sense of civilization or progress – cultural difference was judged through the lens of environmental determinists as moral and intellectual superiority based on a scale of perceived development. Humans were not all considered equal. While humans may have 'risen' from 'nature', according to environmental determinists, some were less human than others depending upon where they were 'located' along a course of ascension 'above' nature. Ascendance above the nonhuman world was understood by environmental determinists as a process of becoming civilized, and becoming cultured. Humans were differentiated by being classified into 'races'. These classifications were regularly disputed, and relied upon crude techniques, such as anthropometrics (body measurement), or drew upon the now discredited scientific ideas of the 1930s, including eugenics and social Darwinism. Certain 'races' were regarded to have achieved 'higher' levels of civilization – literally, acquiring cultured traits (like reason, rationality, technology, etc.) – as they evolved away from nature. McClintock demonstrates how in the nineteenth century Europe, these ideas of racial superiority were naturalized through depictions of the human 'family tree' – which placed the white races securely on the top branches. Fanciful assumptions were made that the environment somehow determined cultural differences, including morality and intellect. In other words, climate, remoteness, topography, and available ecological resources were responsible for variations in ways of life, and enabled (or limited) peoples to become cultured.

Such theories are dubious not only because of the inherent racism and lack of cross cultural understanding typical of that time. They are also logically inconsistent because environmental determinists mistook material cultural evidence – the extent of cultivation (literally, in the case of the sophistication of farming practices) and assembly of material objects and edifices (like industrial technologies, buildings, cities, etc.) – as proof of ascent (or otherwise) along hierarchical scales of civility and cultural advancement. Enormous presumptions were made about what evidence constituted culture as a way of life, which was itself poorly theorized. For instance, environmental determinists were too quick to view the absence of grand buildings in some indigenous cultures as evidence of lack of advancement. At the same time, the depth and complexity of indigenous cultural practices and traditions was rarely recognized, or could rarely be imagined outside the dominant Western hierarchical worldview of the time which posited such peoples as 'lower' or 'less cultured'. Ideas of culture as a way of life, when partially and selectively deployed, operated to justify a bounding of certain human worlds as cultured, as separate from others; the remainder categorized as less civilized, primitive, or as belonging to the natural world. Such a conception of culture – a 'thing' possessed by certain humans to various extents, in opposition to nature (as 'without culture') – became perhaps the most pervasive and influential example of binary thinking in geography, sustaining imagined boundaries between the civilizations of Europe and the savagery of 'new' worlds. Further, in this (European) human centered moral universe, rights were allocated to only those certain people who sat above animals, plants, and minerals. Indigenous rights to land and resources in settler societies were not recognized or were traded away in treaties – acts which set in train conflicts that remained the subject of political struggle for centuries. Geographical knowledges thus enabled European colonial dispossession to be seen as the 'survival of fittest' cultures and states over others, while missionary evangelism and the appointment of Aboriginal 'protectors' could be justified as the benevolent ushering of indigenous and 'lower' races along the civilizing spectrum – diffusing civilization and 'culture' through Christianization.

Although contemporary cultural geographers might, with understandable moral outrage, recoil at the idea that such ideas were a foundation for their subdiscipline, it is important to note that environmental determinists were, in effect, writing cultural geography prior to the name 'cultural geography' coming into widespread use with the Berkeley School. Environmental determinists conjectured on the qualities of culture, cultural differences, and geographical distributions. The logic of environmental determinist thinking in turn had its own historical context – it did not appear from a vacuum either. It had been influenced by Western philosophy as far back as Aristotle and Plato, and later Locke, Darwin, Montesquieu, and Lamarck. It is thus possible to argue that the production of cultural geographical knowledges has been a mainstay of Western intellectual efforts throughout many hundreds of years. In regular use, however, the term 'cultural geography' only became prominent after Carl Sauer and the Berkeley School rejected environmental determinism, introduced the concept of cultural landscape, and injected into geographical theory the capacity of humans to transform their surroundings through a particular way of life.

For essentially half a century, the superorganic, Sauerian understanding of cultural landscape dominated cultural geography, especially in North America, until the emergence of humanistic geography in the 1970s, and the so called 'cultural turn' of the late 1980s, which transformed the subdiscipline and stretched what was meant by culture. Throughout the 1960s, geography had been engaged in an excursion into mathematical modeling and positivist exploration of spatial processes – the so called quantitative revolution. In the 1970s, geographers reacted against this, drawing on Marxist theories of uneven development, class conflict, and the structural contradictions of the capitalist system, to enliven a new radical geographical perspective. Throughout these decades, cultural geography – still very much seen as in the Sauerian tradition as the study of cultural landscape, region, ecology, and diffusion – was a persistent, albeit sidelined presence. Cultural geography did contribute to the growing, interdisciplinary fields of cultural and political ecology, but by the 1970s it had become less popular and less visible, a specialty considered by many to be arcane or inconsequential.

By the late 1980s, however, Lester Rowntree, summarizing in *Progress in Human Geography* the advances made by 'new' cultural geographers such as Derek Gregory, Peter Jackson, James Duncan, and Dennis Cosgrove, was led to the following observation:

> For geographers accustomed to the low, yet enduring profile shown by cultural/humanistic geography over the decades, a silhouette that sometimes engendered a certain defensiveness by its practitioners, this last year has been characterized instead by highly visible activity: a well known, committed and productive cultural geographer as AAG president, recognition of cultural geography as a speciality group within the association, a multitude of panels and special sessions on 'new directions' and 'emergent themes' in cultural geography, even multiple edition textbooks that attest to strong undergraduate enrolments in the area. Has a phoenix arisen? (Rowntree, 1988: 575)

Rowntree was describing the postmodern 'cultural turn' (as it would become known) that in the late 1980s and early 1990s networked its way through Anglophone geography, and to some extent further afield. The timing of the 'cultural turn' can be linked to a broader dissatisfaction across the social sciences and humanities –

including geography – with existing conceptual tools and their ability to help understand the complexity and vola tility of contemporary social change. The cultural turn was influenced by the writings of theorists outside geography, such as Pierre Bourdieu, Raymond Williams, and Clifford Geertz, and captured in a series of important books on meaning, power, and the symbolic landscape. According to Cook *et al.*, the foundational narratives and initial energy for the turn in geography came mainly from geographers based in the United Kingdom. They credited the col lection by Chris Philo – *New Words*, *New Worlds* – for putting the 'new' into 'cultural geography', although manifesto like statements about the need for 'new' cultural geography had appeared earlier, in particular the papers organized for Cosgrove and Jackson's session at the 1987 Institute of British Geographers (IBG) conference, on 'new directions in cultural geography'. During the 1990s, momentum gathered for 'new cultural geography' under a series of conferences organized with the support of the Social and Cultural Geography Research Group of the Royal Geographical Society and the IBG.

The curiosities of 'new' cultural geographers in the 1980s and 1990s can be interpreted as a series of broad intentions. First, although postmodernism was the catchcry, much of cultural geography after the cultural turn was politically post Marxian, in the sense of either seeking to advance from, or reacting to, the Marxist political economy that dominated human geography from the 1970s. Humanist geographers writing in the late 1970s and early 1980s were keen not only to reflect more theoretically on the nature of the tensions between socioeconomic structure and human agency, acknow ledging Marxist insights into the macro scale processes and conditions that create social divisions and determine life chances, but also to recognize how human agency is enacted within the bounded and structured confines of particular places and times. Influential at the time were perspectives from phenomenology and structuration theory. Even though Marxism emphasized structures of capitalism, it did enable cultural geographers to move further from superorganicism by recognizing the manner in which esthetic and moral values were contested, and "configured in such a way that they reinforce [society's] economic and political structures" (Shurmer Smith, 2002: 29).

Post Marxist cultural geographers were also heavily influenced by feminist thought and philosophy, and in particular by the realization that socioeconomic class was not the only axis of oppression. Whereas Marxist his torical materialism provided a useful theoretical per spective for radical geographers in the 1970s who sought explanations for the way in which capitalism was re sponsible for socio economic forms of oppression, those seeking explanations for racism, sexism, and homophobia required different kinds of theoretical tools and empirical approaches. By this time, racial conflict was widespread and the Civil Rights Movement had reversed segregation in the United States, the sexual revolution had con fronted conservative norms about gender roles and had empowered women, and decades of international migration and the growth of tourism had produced more heterogeneous cities. The idea of culture as a stable, superorganic 'way of life' held collectively by populations needed to be improved upon. Culture became under stood more relativistically as identities and behaviors, held by some in a cultural geographical group (and not by others), and deployed by individual people at different times and in differing ways depending on context. This theoretical shift was necessary for researchers interested in confronting oppression, to understand human cultural difference, challenge the idea of 'race', uncover the gendered nature of social institutions, and unsettle con servative ideas of 'normal' sexuality and family.

For example, the concept of 'queer' (understood both as an adjective and as a verb) became crucial to ques tioning and contesting normative assumptions about sexuality, gender, and space, encouraging researchers to replace assumptions considered 'fixed' and 'natural' with more fluid and unbounded perspectives. Bell *et al.* dem onstrated how space is often taken for granted as het erosexual by discussing the hostility experienced by those who acted outside the codes and norms of heterosexuality, for example, same sex kissing in the street. More recently, the challenges posed by the gay friendly marketing of nations, cities, and festivals has been discussed, in par ticular how such efforts operate to assimilate particular understandings of gayness into mainstream life. Other related debates have included the practicalities of doing and writing queer geographies, as well as potential pol itical interventions that encapsulate the philosophical commitment to ideas of slippage, in between ness and liminality.

A second and related intention of the cultural turn was to uncover how ideas, knowledges, and social practices are produced, maintained, and circulated, es pecially in the realm of everyday life. Whereas Marxist geographers, with their intent to explain socioeconomic oppression, sought to understand the structure and politics of the world capitalist system, cultural geog raphers interested in sexism, racism, homophobia, and other axes of oppression needed to move beyond superorganic ideas of 'systems' and 'structures' and grasp with more subtlety the manner in which ideas and at titudes about people and places infused social life, and were responsible for the ways that repression and cruelty materialized. Influences from post structuralist literary theory permeated geography: meanings for culture were no longer taken as fixed or stable; instead, depictions and representations of places and peoples became subject to analysis. Foucault's idea of knowledge as power and the related concept of 'discourse' (understood as sets of statements that make people, plants, places and things understandable) were particularly influential. Represen tations and discourses could be captured as 'data' in formal documents, such as government policies and planning approvals, and in 'everyday' sources such as newspapers, films, television shows, and songs. Analysis of this could reveal the origins and contours of dis cursive formations – ideas, knowledges, beliefs, attitudes, depictions, and 'common sense' notions that permeate society and shape the contemporary world's cultural geography. For instance, racism toward 'Asians' in Britain, or Muslims in the United States, could be re vealed through understanding how both groups have been depicted (often in a demonizing fashion) in tele vision and newspapers. Methodological advances in cluded the literary technique of deconstruction, and the development of latent and manifest content analysis – a more numerical, coding based approach to represen tational analysis using the language and pictorial ma terial in everyday media as evidence.

Thus, overtly drawing on post structuralist semiotics, geographers could 'read' from everyday discourses the signs and symbols that embody meaning. What these meanings were – and thus how researchers interpreted them – was argued to be open to political and ideological processes, as different groups sought to maintain or contest dominant meanings, or replace them with alter natives or pluralistic interpretations. Cultural represen tations in the everyday were outcomes of relations of power, from contestations between hegemonic interests (who install dominant meanings) and subordinate groups, who to various extents resist these dominant meanings and ideologies, and express their own interpretations.

Concurrent with this shift toward the representational and the everyday was the reclamation in analysis of 'popular' forms of culture. Inspired by the manner in which cultural studies emerged as a new interdisciplinary field seeking to challenge the stuffy orthodoxies of lit erary criticism, classics, and musicology, geographers embraced popular culture – once considered fanciful, escapist, or common – as a new area of research to be taken seriously. The meaning of culture 'as art' was ex posed as elitist and deeply tied to imperial notions of European civilization as more 'cultured' than other so cieties. Instead, popular culture in all its forms, from hip hop to sit coms, and magazines and comic books, became possible sources of representational material for cultural geographical analysis.

Despite the exciting possibilities offered by working outside conventional paradigms, advances in 'new' cul tural geography were not without their critics. The of fences supposedly committed can be condensed down to at least five. Cultural geographers were charged with neglecting the immediately political – of drifting away from a concern with oppression. At best, 'new' cultural geography was all hype and no action. Second, cultural geography was charged with ignoring questions con cerning rigor, morality, and truth. Cultural geography lacked methodological rigor and had become an 'any thing goes' subdiscipline. Third, cultural geography was charged with speaking an exclusionary language of post structuralist 'jargon' filled with its own self importance. Fourth, driven by theory, the cultural turn had trans formed the word into the world. Scant empirical data became a veneer, allowing theory as fashion to run rampant. Conversely, a final critique suggested that the cultural turn had discarded the possibility of integrative or holistic theory, transforming the world relativistically into a series of case studies, with a soft theoretical veneer. At best, the cultural turn resulted in a number of highly reflexive case studies. Thrift alerted us that such charges are of benefit. Crucially, he pointed to the importance of the application of analysis of everyday geographies into government policies through initiatives in both teaching and training. Others argued that cultural geographers have continued to work politically (on forms of op pression beyond capitalist exploitation), that methodo logical experimentation was precisely what was required to push the barriers of knowledge beyond problematic assumptions and staid conventions. Futher, cultural geog raphy's contemporary terminology was appropriate, and no different to the technical language of the physical sciences – having its own theoretical origins, and specific intended purposes and meanings.

Yet, throughout the 1990s, and on into the 2000s, cultural geographers themselves would voice dissatis faction with the dominance of the now mainstream representational strand of cultural geography. The ar gument was that cultural geography had become too reliant on textual analysis and cultural discourse, without requisite ethnographic work required to understand how these representations impacted on people, social policy, and the material landscape. Instead, it was recommended that geographers promote efforts to 'rematerialize' geography, through a 'newest' cultural geography, in tended to supersede the 'new' cultural geography of the 1980s and 1990s.

One response was through the importation of yet an other set of outside theoretical influences, this time from history and philosophy of science and the work of authors such as Bruno Latour: the so called 'actor network the ory' with its focus not on representations or discourse, but on the relations forged in an ongoing manner between people, objects, plants, and animals. The core of this theoretical perspective was recognition that humans did not have a monopoly over culture, nor over agency; in stead, nonhuman objects, animals, and plants were theo rized as agents with equal capacity to exist and enact agency in networked sets of relationships with humans and other beings. These sets of relationships – often described as 'assemblages', 'actor networks', or 'hybrid geographies' – move cultural geography away from a purely discursive focus, and advance an understanding of the world in which dualistic ideas about humanity and nature as separate spheres are no longer assumed.

While actor network theory provided an excellent tool for challenging nature–human dualisms, concerns were raised about how understandings of place rested within this conceptual framework. Cloke and Jones extended the concept of networks by turning to the concept of dwelling. It offered deeper insights into how (non)human actors are relationally co constituted in landscapes and places, as well as networks. Examples would be 'the city', 'the orchard', or 'backyard', conceptualized not as bounded geographical entities, but as a set of continually revised relationships between people, the material objects (such as cars, roads, and ports, in the case of the city), and ecological systems containing plants, birds, insects, etc. Thrift also pointed out the failure of actor network theory to conceptualize place, using the term 'ecology' to signal that thinking about relational places involves understanding interactions between a wide spectrum of entities, some human, some physical, some biological, and some human made. Further, Thrift argued that actor network theory gave conceptual priority to the technical over the human body – that is, its perceptual mechanisms, memory, and various bodily skills. Hence, Thrift extends relational thinking about the spatial by directing attention to Judith Butler's concept of performativity. In this view, identities are unstable, and not innate; instead, they are performed repetitively by subjects interacting (whether consciously, or in an embodied, unconscious level) with historically embedded discourses, norms, and ideals. Gender is not a given biological fact; rather it is performed by subjects in relation to social norms and ideals. This has enabled the rethinking of the relationships between scale, subjectivity, the body, and mobility. For example, Knopp rethinks the role of mobility in the lives of nonheterosexual people. Rather than explain the mobility of people with same sex desires purely through attributes of the urban or the rural (as destinations and/or places of origin), the embodied motivations of individual people are also seen as crucial. On the one hand, particular sexual desires may be played through differences nonheterosexual people imagine between the city and the country. On the other hand, identities are created and performed through the experiences and acts of physically moving through space. A focus on embodied (dis)placement is a constant reminder that personal identity formation is spatially co constituted, progressive, and fluid, and never complete or fixed.

Cultural Geography, Take Two: …The More Things Stay the Same

Take one, which rests on an origins narrative, suggests a logical, linear progression through which modern cultural geography has developed. 'Classical', 'new', and 'newest' cultural geography can be read off as sequential paradigmatic episodes: one orthodoxy is challenged, and replaced by an alternative, which itself becomes orthodox. However, in take two, we want to complicate this overly linear narrative by illustrating the ways in which some themes endure, albeit refashioned through an alternative conceptual lens, while other paradigmatic 'revolutions' can be alternatively viewed as more gradual and uncertain transitions. We argue one way cultural geography defies the origins narrative is through how some objects of analysis continue, like landscape, despite the changing definitions of the cultural. In other words, through a discussion of different definitions of cultural geography we hope to show that modifications and variations notwithstanding, there is also stability.

One particular stability in cultural geography is in the concept of 'culture' adopted – in particular the idea of culture as 'way of life'. Even though humanists and 'new' cultural geographers sought to move beyond superorganic conceptions of culture, enabling consideration of individuals and subjectivity, the idea of culture as shared traits linked to bounded geographical entities has persisted – perhaps in part because this view of culture is presumed common sense, and is thus so pervasive in everyday life.

Cultural geographers influenced by the 'cultural turn' in the 1980s were quick to criticize the idea of culture as 'way of life'. They revealed how, in the idea of culture as 'way of life', important assumptions based on learned behavior were made about where to draw lines separating allegedly distinctive cultures. The assumption that everyone within a geographical bounded area had similar behavior patterns, values, and attributes was deeply problematic. Important differences between people by age, ethnicity, sexuality, class, and gender are masked by the idea that people share a definable way of life in a given culture region. At best, culture as a way of life noted the importance of human agency in the transformation of place; at worst, it supports highly essentialist, sexist, and racist arguments.

Yet, despite critiques, the Sauerian tradition continues to inform some introductory cultural geography textbooks, particularly those written for the United States market. According to Jordan, Domosh, and Rowntree culture is the:

> learned collective human behaviour, as opposed to instinctive, or inborn, behaviour. These learned traits form

> a way of life held in common by a group of people. (Jordan *et al.*, 1997: 5)

Despite cautionary statements, within the United States, in particular, the Sauerian concepts of cultural landscape and culture as a way of life have ongoing appeal. The understanding of cultural geography as concerned with regional variations – in cultural practices and norms, landscape, sport, popular culture, and so on – is still practiced.

In some respects, the culture as a 'way of life' approach has been put to important use, even after critiques, as a way of replacing dominant Western understandings of the world with more pluralized geographical imaginations. Moving beyond its problematic colonial past, the culture as a 'way of life' approach persists in studies of the geography of language which have sought to reveal the extent to which smaller languages are under threat of domination from English, or from government policies seeking to encourage indigenous populations to abandon traditions and adopt national languages (as in Indonesia and Australia). Although such studies are critical, and informed by an intention to resist dominant exercises of power over subordinate groups, the conception of culture adopted remains similar to that which characterized early twentieth century geography. Much of the contemporary research in indigenous geographies (including analysis of kinship systems, ceremonial performances, political struggles, and relationships to land) still assumes some level of collective shared culture and defined territory. Similarly, many ethnographic studies of international migration, diaspora, and cultural practice in multicultural societies retain some sense of the 'way of life' approach – for instance, in research examining transnationalism, cultural maintenance, community festivals, music, marriage practices, sport, and so on.

Culture as way of life also persisted in other 'classic' ongoing debates, such as the cultural homogenization and cultural hybridization debates. Despite first being voiced in the 1980s, arguments still abound about cultural homogenization and the influence of American popular culture, fast foods, television, and the English language on global cultural diversity. American culture is positioned as a hegemonic force, with detrimental impacts on the diversity of ways of life in other parts of the world – homogenizing and standardizing the world into an oligopoly of pastimes, consumer behaviors, and brands. The related debate about 'cultural imperialism' concerns the manner in which norms and morals of dominant cultural groups are rendered 'natural' within or across nations at the expense of subordinate groups, whose ways of life are denigrated and whose members are often subject to racist attacks. Processes of cultural homogenization and cultural imperialism both rely on interpretations of culture as the way of life, albeit viewed through a more thoroughly critical, political lens than in the superorganic era of cultural geography.

Others have argued that diversity is growing in the midst of Anglo American cultural imperialism (through, for instance, widespread international migration and tourism), as people in their everyday lives adopt, mutate, and reclaim as their own the products, stories, and meanings of popular cultural flows – in other words, hybridization. For example, a transnational fast food corporation founded in America, like McDonalds, may not be thought of by many of its customers in Australia, China, or Russia as 'foreign', but merely as part of everyday life. For Appadurai, the cultural landscapes imagined as stable in previous eras have been replaced by new 'mediascapes', 'ethnoscapes', and 'technoscapes', with more complex sets of relationships between global flows and everyday lived realities. Examples include television watching practices outside the United States, and the formation of musical subcultures that blend American styles with local instruments and styles.

Another example of stability in cultural geography is provided by a particular theme – landscape – that is persistent, albeit reinterpreted through new theoretical lenses. 'New' cultural geographers certainly differentiated their interpretation of landscape from the classical work of Carl Sauer. First, the idea that landscapes as a signifying system had inherent meanings became relevant, emphasized by a humanistic approach. Landscape was understood as socially constituted and the product of an ideologically laden 'way of seeing'. Although 'new' cultural geographers rejected Sauerian superorganicism, landscape could be read as text, just like films and newspapers; and just like those other texts, it embodied ideology and meaning. Second, drawing on the work of Marxist cultural geographers, the analysis of landscape was explicitly political, particularly through how different social groups discursively layered meanings over a particular landscape. For Daniels, an advocate of the concept of landscape within new cultural geography, this last point was crucial. He emphasized understanding of landscape as site of contest, through which oppositional social worlds were made meaningful through historically situated visual codes and conventions. Following this, from the early 1980s, new cultural geographers began to investigate landscape as a "discursive terrain across which the struggle between different, often hostile, codes of meaning construction has been engaged" (Daniels and Cosgrove, 1993: 59). Landscape came to be understood as one of many texts, including those on canvas and on paper to be interpreted, decoded, or deconstructed. Some critics suggested that all cultural geography became swept away by critical techniques of deconstruction. Yet, this is an exaggeration. For example, Don Mitchell is a supporter of ongoing critical engagement with the material dimensions of landscape. While Mitchell acknowledges the importance of paying attention to 'symbol and metaphor', he underscores a commitment to understanding landscape as always more than represen tational. He suggested that "landscape may demand a theory of landscape, but it also demands that theories of capital circulation and crisis, of race and gender, and of geopolitics and power be built right into it" (Mitchell, 2003: 790). In other words, cultural geographers should not simply be producing accounts of ideologically loaded landscapes, but working toward producing landscapes of social justice through an account that situates landscape within the relations of production that underpin the global economy.

In other ways, the idea of a 'cultural turn' in geog raphy, responsible for a paradigmatic revolution, has been exaggerated – either by those wanting to proclaim the 'new' (and their place in it), or by those seeking to generalize about new research prior to criticizing it. Cook *et al.*, for instance, warned that in geography there was much less a revolutionary 'cultural turn' than a far more modest series of engagements with post structuralist writings and debates, borne out of curiosity for their implications when thinking spatially and doing geography. Indeed, in a theoretical sense, geography's 'cultural turn' was less a revolution and more the culmination of a steady transition already underway, stemming from the reinvigoration of the humanist tradition in geography. The same impulses that drove Sauerian cultural geographers in North America and the 'possibilists' in France to react against the disempowering determinism of their predecessors had reemerged by the late 1970s. Social theory – and dom inant theoretical perspectives in geography in particular – had become once again too doctrinaire, too dogged in pursuit of all encompassing explanation; the subtlety, variation, and capacity of humans to influence their living conditions appeared to have been thoroughly erased. By the mid 1980s, attempts have already been made to survey the resulting eclecticism that the 'humanist' and 'behaviorist' umbrellas captured and to more thoroughly question the manner in which philosophical assumptions were imported to geography in previous geographical paradigms from neighboring disciplines, such as economics and sociology. Terms like 'agency', 'sense of place', 'perception', and 'mental maps' had already fully entered the geographical vocabulary before the 'cultural turn', as geographers sought to understand more fully the subjective links between individuals and their surrounding environments, providing "an expansive view of what the human person is and can do" (Tuan, 1976). The entry of more humanist concepts into geography in the 1970s and early 1980s took place in a series of transitional moments – without fervor or tumult – which set the scene for the 'new' cultural geography that followed.

Transitions were not only limited to cultural geog raphy either. The dissatisfaction with existing conceptual tools was manifest in a more widespread suspicion of the ability of universalizing explanations to accurately and adequately capture the complexity of contemporary economic and political processes. To illustrate, in the very same issue of *Progress in Human Geography* that Lester Rowntree remarked on the rise of new cultural geography, Trevor Barnes wrote an extensive and com prehensive critique of universalizing theories in eco nomics and economic geography – in particular the notion of *homo economicus* (a theory that humans acted rationally in ways that always maximized an individual's ends given the limited means available), and Felix Driver wrote a critical review of universalist concepts of history in geography, arguing instead for context in theory, and a breaking down of binaries between 'historical geography' (the study of the past) and other types of geography (only dealing with the present). In other words, although the 'cultural turn' is now usually perceived as a radical change in geography's direction stemming from cultural geography, and in turn then influencing other sub bran ches of the discipline, the changing theoretical directions were actually more widespread, were perhaps not even 'cultural' at all. Rather than working to 'zone off' a subdisciplinary field of cultural geography, themes such as identity, power, and discourse common in cultural geography have ebbed and flowed in a parallel fashion throughout research in political, economic, development, social, and population geography. The 'cultural turn' was thus more accurately part of a discipline wide transition from the desire to develop complete, fool proof theories about human societies to a desire to develop theories that were supple, less bombastic, and more capable of being sensitive to contingencies, variations, and contradictions.

Cultural Geography, Take 3: Re-Placing 'the Human'

The third, and perhaps most unsettling take on cultural geography, concerns the very manner in which human beings have been positioned in research and thinking. Although cultural geographers' engagements in the 1980s and 1990s with themes such as landscape and identity opened up the discipline to ideas of power, discourse, and ideology, they left largely intact the idea that humans, possessing culture, were somehow separate to nature. For instance, Mitchell assumed that the cultural landscape was an exclusively human attainment. Humans were the sole actors in generating 'cultural landscapes'. If land scape is conceptualized as a product of human culture then there is no possibility to recognize nonhuman agency in the land. Research that conceptualizes culture as humans 'bearing traces', 'influencing', or 'creating' landscapes refuses the possibility of rocks, water, sun, soil, and nonhuman animals having agency. 'Nature' becomes an inert medium, a passive field, rather than a creative agent. Critical questions have emerged about how such assumptions have been made, by whom, and to what effect. Insights from studies of the cultural geography of human/nature relations in non Western societies have illuminated just how limiting European and North American perspectives on the agency of nature have been, in contrast to indigenous societies, where the agency and cultural significance of landforms, rivers, glaciers, and other natural features is more fully realized. More deeply, geographers have questioned the very ontology of 'human' and 'nature' as meaningful categories, inviting difficult questions about our existence and relationships with various parts of the nonhuman world. Dialogs have emerged particularly with science and technological studies, and feminist philosophers such as Val Plumwood. Rethinking agency in a way that involves collaboration between human and nonhuman 'actors' has necessitated conceptual tools that do not separate culture (human presence) from nature (human absence). The conceptual frameworks of researchers such as David Crouch, Nigel Thrift, and Sarah Whatmore incorporate the concept that landscapes are always more than human. They offer narratives that prioritize the active role of nonhuman worlds through concepts of reciprocity, networks, assemblages, multiple interactions, and collaborating agencies, instead of treating humans as superior and apart from nature. One challenging research agenda adopting such a collaborative model is to decenter the human in policies designed to manage environmental futures.

As well as paying attention to the agency of actors in nonhuman worlds, cultural geographers are also interested in multiplying the sensory pathways through which analysis takes place. Anglophone academic geography, deeply embedded in the Western philosophical tradition since the Enlightenment, has become heavily dependent on the visual sense to provide it with suitable evidence and analysis. Other senses such as smell, touch, and sound rarely featured in research, and were considered less 'reliable' than sight when forming the basis of geographical explanation. Only sight appeared to provide evidence worthy enough of rationalist argument. Recently, cultural geographers have sought to decenter the importance placed on sight, and suggest instead that our webs of relationships with other human and nonhuman actors needs to be understood as mediated by all the senses. Moreover, these relationships take place in ways that involve humans as physical bodies interacting with other actors and environments before and beyond cognitive processing through language and contemplation. The focus within 'new' cultural geography in the 1980s to 1990s on the representational was supplanted with other insights that emphasized the more embodied ways of knowing space. While representational or textual understandings of the spatial remain important, and are not jettisoned by those interested in pursuing embodied approaches, an overemphasis of the cognitive and the contemplative means geographers miss most of what is going on in everyday experiences – including smell, sound, emotion, memory, and touch. These sensory pathways do not rely upon the internal processing of representations of an external world. In other words, nonrepresentational experiences are central to our making sense of space. As Noel Castree explains, such approaches:

> focus on a world in which we are dwellers not observers, multi sensual participants not detached spectators. We come to know by doing, and we do because of what we already know in an iterative process where the material world affects us and we affect it… we engage with the material world… using all our senses: we are practical beings not just intellectual ones. Much of our understanding of, and action upon, this world is thus, in Thrift's view, never formally represented or representable at all because it is tacit, sensuous, habitual and precognitive. (Noel Castree, 2005: 229 230)

Cultural geographers have thus pursued 'soundscapes', 'smellscapes', or 'olfactory geographies'. Space has acoustic, rhythmic, and sonorous qualities, and thus sound, noise, and music can be seen as integral to understanding, experiencing, and making everyday geographies, for example, in the use of personal stereos or in dance clubs or street parades. These sorts of studies provide insights into how people do things, and how people constitute their worlds through sensing and experiencing. Music and sound elicit emotions that help comprise the spatial through the ways in which people individually respond.

By focusing on experiential responses, including those aroused by smell, touch, music, and sound, cultural geographers have started to think about the emotional and affective dimensions of the spatial. Deborah Thien attributed the registration of geographies of emotion to what she terms an "affective turn in social and critical thought," meaning a shift toward understanding how agents influence each other not just cognitively, but also emotionally and bodily. The concept of affect has philosophical origins with Spinoza and Deleuze and Guattari, and is widely used in psychology to describe the emotions, as opposed to cognitive, rationalist behavior. In more recent cultural geography, its deployment is particularly meant as a way to open up theorizing on the way in which humans relate with surrounding environments and nonhuman actors, and less based on clear distinctions between affect and rationality (distinctions which are blamed for denigrating emotion as ‘less reliable’ or ‘valid’ in a research sense). Emotions are a social projection of an individual feeling, and while feelings are understood as personal and biographical because they rely upon interpretation and categorization from previous experiences, expressions of emotion are sometimes refracted when experienced in particular contexts. Affect is pre personal; it is a nonconscious experience of intensity that is incapable of absolute explanation through language because it exists prior to and outside of consciousness. Understanding how the world is mediated by affect is therefore commonly understood as challenging legacies of rationalist and masculinist social science that kept feelings out of academic research and practice. What bodies may be and do is now subject to debate because of theoretical advances on affect, and how affect shapes the relationship between subjectivity and space.

A third, voiced intention of researchers attempting to reposition ‘the human’ in geographical research is to move beyond methodological orthodoxies and now standard practices of ‘reading’ cultural discourse. Further, questions about sensory pathways, affective relationships, and the agency of nonhuman actors demand a much broader range of skills and methodologies than in the ‘tool bag’ of the cultural geographer trained from the 1930s to the 1980s. Typically, cultural geographers were devoted to in depth interviews, alongside some ethnographic techniques, and later on, deconstruction and content analysis. While these tools remain important, the emphasis on understanding the affective making of everyday lives necessitates what Thrift terms a ‘methodological break out’. Such a break has inspired a plethora of experimental methodological strategies ranging from performative writing, interactive websites, photo and film journalism, and 3D perception mapping using geographical information systems, attempts to co produce the world, for instance, through street theater, and various methods which work with bodies, such as music therapy.

Also infusing cultural geography in recent years has been a concern with researchers themselves: how they come to do research from particular perspectives, how their identities shift and are challenged in research, and how the work they do is conducted and interpreted in relation to the institutions of universities, research facilities, and disciplinary traditions in different countries and cultural contexts. Indeed, although ‘cultural geography’ is now accepted as a major component of human geography (even if impossible to bound or define), its meaning and use is relatively recent, is geographically contingent, and is interpreted varyingly by researchers coming at their projects and written work from the particular positions they occupy. How cultural geography is done varies enormously in different linguistic and national traditions, and conversely, researchers are increasingly realizing that they cannot pretend to be researching from a placeless, genderless, ‘unethnic’ zone of impartiality.

We have already encountered some of this variation in positionality – when in the above discussion of the manner in which Sauerian cultural geography persisted in North America more than elsewhere; and how in Britain, in particular, ‘new’ cultural geography found traction. Far more complex again is the extent to which perspectives and approaches in cultural geography have diffused and been hybridized in other parts of the world, and how researchers have negotiated their own positions, identities, and limitations when undertaking cultural geographical research. ‘Spaces of geographical knowledge’ is a theme avidly explored in the country reports of the journal *Social and Cultural Geography*, where differences in national practice are discussed. For instance, Bunnell, Kong, and Law report on how in Southeast Asia, research on national culture, religion, and multiculturalism dominate (linked in part to domestic political priorities, as well as the relatively recent histories of nation states there). Whereas in places like Greece, New Zealand, the Caribbean, and Australia, there is a combining and more ambiguous position prescribed for social and cultural geography where geography departments are small and lecturers cover more than one subject area. Concerns have also been expressed about how Anglo American geography is producing geographies that are assumed to be universal, creating discourses of what ‘international’ research means, even though such research is thoroughly positioned from a particular perspective, time, and place. In other outlets, cultural geographers have debated the related issue of what acknowledgement of positionality in research means practically, in terms of writing and the conduct of using new methods in cross cultural circumstances, and more philosophically, when coming to assess self reflexivity in research and the value of case study research vis à vis links to theoretical developments.

So, cultural geography has come to be a diverse, enigmatic subdiscipline – illustrated by the various entries of Anderson *et al.*’s (2003), *Handbook of Cultural Geography.* As cultural geography rose in popularity in the 1980s and 1990s, singular definitions of ‘culture’ were cast aside and replaced instead with epistemological uncertainty and a desire to promote plurality, self reflexivity and transgression of inherited intellectual binaries. At the same time, older and simpler traditions survived and retained some meaning and relevance, particularly for debates about cultural change and continuity in an ever globalizing context. All the while, transitions occurring in cultural geography were also taking place in allied subdisciplines, such as economic and political geography. At its most experimental – even nihilistic – edge, cultural geographers continue to work with elementary problems about human existence, and our position in the world, in such ways as to call into question the very utility of categories such as 'human', 'nature', and 'culture'. It may turn out to be that cultural geographers retheorize the world in such a way as to render the identity of the subdiscipline itself redundant. If so, there could probably be no more defiantly radical final act.

See also: Actor-Network Theory/Network Geographies; Affect; Berkeley School; Content Analysis; Cultural Turn; Culture; Deconstruction; Discourse; Dwelling; Ecology; Historical-Geographical Materialism; Human Geography; Humanism/Humanistic Geography; Hybridity; Identity Politics; Imperialism, Cultural; Landscape Perception; Landscape; Marxism/Marxist Geography I; Mental Maps; Non-Representational Theory/Non-Representational Geographies; Performativity; Phenomenology/ Phenomenological Geography; Popular Culture; Possibilism; Poststructuralism/Poststructuralist Geographies; Quantitative Revolution; Queer Theory/ Queer Geographies; Sauer, C.; Sense of Place; Sensorium; Situated Knowledge, Reflexivity; Structuration Theory; Transnationalism.

Further Reading

Ang, I. (1985). *Watching Dallas: Soap Opera and the Melodramatic Imagination*. London: Routledge.

Agnew, J. A. and Duncan, J. S. (1981). The transfer of ideas into Anglo American geography. *Progress in Human Geography* 5(1), 42 57.

Amin, A. and Thrift, N. (2002). *Cities: Reimagining the Urban*. Cambridge: Polity Press.

Anderson, K. (2007). *Race and the Crisis of Humanism*. New York: Routledge.

Anderson, K., Domosh, M., Pile, S. and Thrift, N. (2003). *Handbook of Cultural Geography*. London: Sage.

Anderson, B. and Harrison, P. (2006). Questioning affect and emotion. *Area* 38(3), 333 335.

Appadurai, A. (1996). *Modernity at Large: Cultural Dimensions of Globalization*. Minneapolis, MN: University of Minnesota Press.

Barnes, T. (1988). Rationality and relativism in economic geography: An interpretative review of the homo economicus assumption. *Progress in Human Geography* 12(4), 473 496.

Barnett, C. (1998). The cultural turn: Fashion or progress in human geography? *Antipode* 30, 379 394.

Basch, L., Schiller, N. G. and Blanc, C. S. (eds.) (1993). *Nations Unbound: Transnational Projects, Postcolonial Predicaments, and Deterritorialized Nation States*. New York: Routledge.

Bell, D., Binnie, J., Cream, J. and Valentine, G. (1994). All hyped up and no place to go. *Gender, Place and Culture* 1(1), 5 27.

Berg, L. and Kearns, R. (1998). America unlimited. *Environment and Planning D: Society and Space* 16, 128 132.

Binnie, J. (2004). Quartering sexualities: Gay villages and sexual citizenship. In Bell, D. & Jaynes, M. (eds.) *City of Quarters: Urban Villages in the Contemporary City*, pp 163 172. Aldershot: Ashgate.

Bourdieu, P. (1984). *Distinction: A Social Critique of the Judgement of Taste*. London: Routledge and Kegan Paul.

Browne, K. (2006). Challenging queer geographies. *Antipode* 38(4), 885 893.

Bull, M. (2000). *Sounding Out the City: Personal Stereos and the Management of Everyday Life*. New York: Berg.

Bunnell, T., Kong, L. and Law, L. (2005). Country reports: Social and cultural geographies of South East Asia. *Social and Cultural Geography* 6(1), 134 149.

Butler, J. (1990). *Gender Trouble: Feminism and the Subversion of Identity*. London: Routledge.

Castree, N. (2004). Differential geographies: Place, indigenous rights and 'local' resources. *Political Geography* 23, 133 167.

Castree, N. (2005). *Nature*. London and New York: Routledge.

Castree, N. (2005). The epistemology of particulars. *Geoforum* 36(5), 541 544.

Cloke, P. and Jones, O. (2001). Dwelling, place, and landscape: An orchard in Somerset. *Environment and Planning A* 33, 649 666.

Connell, J. and Gibson, C. (2003). *Sound Tracks: Popular Music, Identity and Place*. London: Routledge.

Cook, I., Crouch, D., Naylor, S. and Ryan, J. (eds.) (2000). *Cultural Turns, Geographical Turns*. Harlow: Prentice Hall.

Cosgrove, D. (1984). *Social Formation and Symbolic Landscape*. London: Croom Helm.

Cosgrove, D. (2000). Cultural geography. In Johnston, R. J., Gregory, D., Pratt, G. & Watts, M. (eds.) *The Dictionary of Human Geography*, pp 134 138. Blackwell: Oxford.

Cosgrove, D. and Daniels, S. J. (eds.) (1988). *The Iconography of Landscape: Essays on the Symbolic Representation, Design and Use of Past Environments*. Cambridge: Cambridge University Press.

Cosgrove, D. and Jackson, P. (1987). New directions in cultural geography. *Area* 19, 95 101.

Crouch, D. (2001). Spatialities of the feeling of doing. *Social and Cultural Geography* 2(1), 61 75.

Cruikshank, J. (2005). *Do Glaciers Listen? Local Knowledge, Colonial Encounters, and Social Imagination*. Seattle, WA: University of Washington Press.

Davidon, J. and Bondi, L. (2004). Spatialising affect; affecting space: An introduction. *Gender, Place and Culture* 11(3), 373 374.

Desbiens, C. and Ruddick, S. (2006). Speaking of geography: Language, power, and the spaces of Anglo Saxon 'hegemony'. *Environment and Planning D: Society and Space* 24, 1 8.

Dodman, D. (2007). A place or a people? Social and cultural geographies of the Anglophone Caribbean. *Social and Cultural Geography* 8, 143 150.

Dowling, R. (2005). Country report: Social and cultural geographies of Australia. *Social and Cultural Geography* 6(5), 767 775.

Driver, F. (1988). The historicity of human geography. *Progress in Human Geography* 12(4), 497 506.

Duffy, M., Waitt, G. and Gibson, C. (2007). 'Get into the groove': The role of sound in creating a sense of belonging in street parades. *Altitude* 8, 1 22. http://www.altitude21c.com (accessed in June 2008).

Duncan, J. S. (1987). Review of urban imagery: Urban semiotics. *Urban Geography* 8, 473 483.

Dunn, K. M., Gandhi, V., Burnley, I. and Forrest, J. (2003). Racism in Australia: Cultural imperialism, disempowerment and violence. In Gao, J., Le Heron, R. & Logie, J. (eds.) *Proceedings of the 22nd New Zealand Geographical Society Conference*, pp 175 179. Conference series 22, School of Geography and Environmental Science, University of Auckland, 6 11 July.

Ekinsmyth, C. (2002). Feminist methodology. In Shurmer Smith, P. (ed.) *Doing Cultural Geography*, pp 177 185. London: Sage.

Elder, G. (2002). Response to 'Queer patriarchies queer racisms, international'. *Antipode* 34(5), 988 991.

Forrest, J. and Dunn, K. M. (2006). 'Core' culture hegemony and multiculturalism. *Ethnicities* 6(2), 203 230.

Foucault, M. (1980). *Power/Knowledge: Selected Interviews and Other Writings, 1972 1977*. Brighton: Harvester Press.

Gibson, C., Costello, L. and Hughes, R. (2004). Doing cultural geography. *Australian Geographer* 35(2), 123 130.

Giddens, A. (1984). *The Constitution of Society*. Cambridge: Polity Press.

Gregson, N. and Rose, G. (2000). Taking Butler elsewhere: Performativities, spatialities and subjectivities. *Environment and Planning D: Society and Space* 18(4), 433 452.

Gutiérrez, J. and López Nieva, P. (2001). Are international journals of human geography really international? *Progress in Human Geography* 25(1), 53 69.

Halberstam, J. (2005). *In a Queer Time and Place: Transgender Bodies, Subcultural Lives*. New York: New York University Press.

Hamm, B. and Smandych, R. C. (eds.) (2005). *Cultural Imperialism: Essays on the Political Economy of Cultural Domination*. Peterborough, ON: Broadview Press.
Haraway, D. (1991). *Simians, Cyborgs and Women: The Reinvention of Nature*. London: Routledge.
Head, L. (2000). *Cultural Landscapes and Environmental Change*. New York: Arnold.
Head, L. and Muir, P. (2007). *Backyard: Nature and Culture in Suburban Australia*. Wollongong, NSW: University of Wollongong Press with Halstead Press.
Jackson, P. (1989). *Maps of Meaning*. London: Routledge.
Jackson, P. (2000). Rematerializing social and cultural geography. *Social and Cultural Geography* 1(1), 9 14.
Jordan, T. G., Domosh, M. and Rowntree, L. (1997). *The Human Geography Mosaic: A Thematic Introduction to Cultural Geography* (7th edn.). New York: Harper and Row.
Kearns, R. and Panelli, R. (2006). Country report: Directions to enlarge our worlds? Social and cultural geography in New Zealand. *Social and Cultural Geography* 7(2), 321 330.
Knopp, L. (2004). Ontologies of place, placelessness, and movement: Queer quests for identity and their impacts on contemporary geographic thought. *Gender, Place, and Culture* 11, 121 134.
Kobayashi, A. (2003). GPC ten years on: Is self reflexivity enough? *Gender, Place and Culture* 10(4), 345 349.
Kofman, E. (1988). Is there a cultural geography beyond the fragments? *Area* 20, 85 87.
Kong, L. (1995). Popular music in geographical analyses. *Progress in Human Geography* 19, 183 198.
Latour, B. (1993). *We Have Never Been Modern*. London: Harvester Wheatsheaf.
Ley, D. and Samuels, M. S. (1978). Contexts of modern humanism in geography. *Humanistic Geography*. Chicago, IL: Maaroufa Press.
Lorimer, H. (2005). Cultural geography: The busyness of being ‘more than representational. *Progress in Human Geography* 29(1), 83 94.
Malbon, B. (1999). *Clubbing. Dancing, Ecstasy, Vitality*. London: Routledge.
McClintock, A. (1995). *Imperial Leather: Race, Gender and Sexuality in the Colonial Contest*. New York: Routledge.
McGregor, A. (2004). Sustainable development and ‘warm fuzzy feelings’: Discourse and nature within Australian environmental imaginaries. *Geoforum* 35, 593 606.
Mitchell, D. (1995). There’s no such thing as culture: Towards a reconceptualization of the idea of culture in geography. *Transactions of the Institute of British Geographers* 20, 102 116.
Mitchell, D. (2000). *Cultural Geography: A Critical Introduction*. Malden, MA: Blackwell.
Mitchell, D. (2003). Cultural landscapes: Just landscapes or landscapes of justice? *Progress in Human Geography* 27, 787 796.
Nettle, D. and Romaine, S. (2000). *Vanishing Voices: The Extinction of the World’s Languages*. Oxford: Oxford University Press.
Norton, W. (2000). *Cultural Geography: Themes Concepts Analyses*. Oxford: Oxford University Press.
Olwig, K. (2002). *Landscape, Nature, and the Body Politic: From Britain’s Renaissance to America’s New World*. Madison, WI: University of Wisconsin Press.
Peach, C. (1968). *West Indian Migration to Britain: A Social Geography*. Oxford: Oxford University Press.
Peet, R. (1986). The destruction of regional cultures. In Johnston, R. J. & Taylor, P. J. (eds.) *A World in Crisis? Geographical Perspectives*, pp 150 172. Oxford: Blackwell.
Philo, C. (ed.) (1991). *New Words, New Worlds: Reconceptualising Social and Cultural Geography*. Lampeter: Social and Cultural Geography Study Group.
Plumwood, V. (2002). *Environmental Culture: The Ecological Crisis of Reason*. London: Routledge.
Rodaway, P. (1994). *Sensuous Geographies: Body, Sense and Place*. London: Routledge.
Rowntree, L. B. (1987). Cultural/humanistic geography. *Progress in Human Geography* 11(4), 558 564.
Rowntree, L. B. (1988). Orthodoxy and new directions: Cultural/ humanistic geography. *Progress in Human Geography* 12(4), 575 586.
Rubenstein, J. M. (2005). *An Introduction to Human Geography: The Cultural Landscape,* (8th edn.). Upper Saddle River, NJ: Pearson Prentice Hall.
Said, E. (1993). *Culture and Imperialism*. New York: Vintage.
Sauer, C. (1925). The morphology of landscape. *University of California Publications in Geography* 2, 19 54.
Shouse, E. (2005). Feeling, emotion, affect. *M/C Journal: A Journal of Media and Culture* 8, 6. http://journal.media culture.org.au/0512/ 03 shouse.php (accessed in June 2008)
Shurmer Smith, P. (2002). Marx and after. In Shurmer Smith, P. (ed.) *Doing Cultural Geography*, pp 29 39. London: Sage.
Smith, S. J. (1994). Soundscape. *Area* 26, 232 240.
Smith, S. J. (2000). Perfoming the (sound)world. *Environment and Planning D: Society and Space* 18(5), 615 637.
Sutherland, E. (2004). In the field politics: Multiple voices of outworkers and researcher. *Australian Geographer* 35(2), 161 168.
Thien, D. (2005). After or beyond feeling? A consideration of affect and emotion in geography. *Area* 37, 450 456.
Thrift, N. (1996). *Spatial Formations*. London: Sage.
Thrift, N. (1999). Entanglements of power: Shadows? In Sharp, J., Routledge, P., Philo, C. & Paddison, R. (eds.) *Geographies of Domination/Resistance*, pp 265 277. London: Routledge.
Thrift, N. (2000). Afterwords. *Environment and Planning D: Society and Space* 18, 213 255.
Thrift, N. and Whatmore, S. (2004). Introduction. In Thrift, N. & Whatmore, S. (eds.) *Cultural Geography*, pp 1 18. London: Routledge.
Tuan, Y. F. (1976). Humanistic geography. *Annals of the Association of American Geographers* 74, 353 374.
Vaiou, D. (2005). Country Report. The ambiguities of social and cultural geography in Greece. *Social and Cultural Geography* 6(3), 455 460.
Waitt, G. (2005). Doing discourse analysis. In Hay, I. (ed.) *Qualitative Research Methods in Human Geography*, pp 133 146. Melbourne, VIC: Oxford University Press.
Walmsley, D. J. and Lewis, G. J. (1984). *Human Geography: Behavioural Approaches*. London: Longman.
Warf, B. (1986). Ideology, everyday life and emancipatory phenomenology. *Antipode* 18, 268 283.
Watson, J. (2000). China’s Big Mac attack. *Foreign Affairs* 79(3), 120 130.
Whatmore, S. (2002). *Hybrid Geographies: Natures, Cultures, Spaces*. London: Sage.
Williams, R. (1988). *Keywords: A Vocabulary of Culture and Society*. Glasgow: Fontana Press.
Women and Geography Study Group of the Institute of British Geographers (WGSG) (1984). *Geography and Gender: An Introduction to Feminist Geography*. London: Hutchinson.
Young, I. M. (1990). *Justice and the Politics of Difference*. Princeton, NJ: Princeton University Press.
Zelinsky, W. (1973). *The Cultural Geography of the United States*. Princeton, NJ: Prentice Hall.

Relevant Websites

http://cgj.sagepub.com/
Cultural Geographies academic journal homepage.
http://cultural.missouri.edu/
Cultural geography specialty group of the Association of American Geographers.
http://www.iag.org.au/study groups/cultural geography study group/
Cultural geography study group of the Institute of Australian Geographers.
http://www.elsevier.com/wps/find/journaldescription.cws home/713880/description#description
Emotion, Space and Society academic journal homepage.
http://www.tandf.co.uk/journals/titles/0966369X.asp
Gender, Place and Culture academic journal homepage.
http://www.tandf.co.uk/journals/titles/08873631.asp
Journal of Cultural Geography academic journal homepage.

http://www.tandf.co.uk/journals/titles/14649365.asp
Social and Cultural Geography academic journal homepage.

http://scgrg.blogspot.com/2008/02/rgs ibg scgrg sessions 2008.html
Social and cultural geography research group of the RGS IBG.

http://www.carleton.ca/space/
Space and Culture academic journal homepage.

http://www.cultural geography.net/
The cultural approach in geography Commission of the International Geographical Union.

Cultural Materialism

G. Helms, University of Glasgow, Glasgow, Scotland, UK

Glossary

British New Left Emerging in the mid-1950s in dissent from communist party politics in the wake of Stalinism, the New Left political organization centered around the Campaign for Disarmament and global human rights/social justice movements.

Critical Marxism A strand of Western Marxism which highlighted concerns over agency, culture, and alienation.

Cultural Materialism A historical materialist theory of culture as social and material productive process and as the use of material means of production.

Hegemony The maintenance of control of one social group over others, which crucially relies on consensus and not merely coercion.

Structures of Feeling Lived, affective experiences which are shared intersubjectively among a social group. Empirically specific in historical and geographical terms, these structures inform emergent cultures.

Introduction

Cultural materialism, as the theoretical framework of the British Marxist and literary theorist Raymond Williams, seems a somewhat curious entry to the encyclopedia. It is curious in two aspects: first, since its common frame of reference is firmly situated within literary criticism or cultural studies rather than human geography; second, while a focus on culture has received sustained interest and scholarship, not at least through the Anglo American new social and cultural geography, its coupling with (historical) materialism has been viewed rather skeptically. Nonetheless, Williams' influence as one of the founding members of the British New Left and the wider politics of post World War II Western Marxism on what would be come from the 1970s onwards known as the study of culture, has been considerable.

While much of Williams' work and the theoretical context of cultural materialism has been appropriated by literary criticism and cultural theory, his arguments around emergent cultures, ways of life, and structures of feeling were for him an indispensable part of a Marxist social and political theory and as such broader than ad vanced literary criticism. In fact, it is the relevance of culture, a concept whose complexity remains virtually unequalled in the English language, as Williams argued succinctly in his *Keywords*, which has led geographers to engage with Williams' essays and theoretical scholarship as well as his range of novels.

For Williams, cultural materialism:

> … is a theory of culture as a (social and material) pro ductive process and of specific practices, of 'arts', as social uses of material means of production (from language as material 'practical consciousness' to the specific tech nologies of writing and forms of writing, through to mechanical and electronic communications systems). (Williams, 2005: 243)

In order to understand Williams' insistence on the centrality of culture and the contributions cultural ma terialism would be able to make, it is important to get a sense of the key concepts and debates within which he embedded his proposal on how the production of culture, and with it arts and literature, can be integrated within historical materialism. These starting points were for once the developments in literary criticism on how cul ture would not only include elite cultural practices and institutions such as fine arts and literature, but also to regard cultural practices as fully embedded and inte grated within all aspects of social life (that culture was in fact ordinary). The second point of entry was his con tinued reception of, engagement with and critique of culture within historical materialism, notably in relation to a dominance of the base/superstructure model of so ciety, as well as debates around hegemony, agency, and (practical) consciousness. These all led on to the key elements in Williams' work: dominant, residual, and emergent cultures; and structures of feeling.

The relevance of Williams' work for human geog raphers is firstly and most widely visible in the early engagement from within historical geography with his arguments on an ordinary and actively produced work ing class culture. Relatedly, a more specific uptake of key cultural materialist concepts has occurred through Williams' work on the country and the city in English culture; and his writings on community politics and a militant particularism that fueled such local struggles. An interest in Williams has more broadly taken place along two further lines: first, a more indirect route by which a study of culture became the object of Anglo American social and cultural geography through an engagement with the growing field of cultural studies; and second, cultural materialism has in turn featured prominently as the frame of reference for a critical interrogation of the cultural turn.

Historical Materialism and the Problem of Culture

In order to situate cultural materialism theoretically, let us first retrace the struggles over culture within historical materialism, which marked a key reference point for Williams. The societal model which had become widely accepted within official Marxism as party politics in the Soviet Union and of European communist parties since the beginning of the twentieth century was that of a rather mechanistic understanding of how an economic base determined an ideological superstructure. Within the latter, politics as well as culture would be located. In such a model, culture and cultural practices have often been narrowed down to the concept of a mirror, and a (more or less consciously applied) false consciousness. Such truncation of the problem of culture to one of reading off culture from the economy had been frustrating for a wide range of Marxist activists and intellectuals – Antonio Gramsci, George Lukács, Theodor Adorno, Max Horkheimer, just to name a few.

As part of a wider political movement on the Left, the British Marxism of the New Left was firmly part of the Western European tradition of a critical Marxism that developed across Western Europe, notably Germany, France, and Italy, in opposition to official communist Marxism. Often in close relation to political struggles, such as the uprisings in Hungary and East Germany, or the student unrests in Western Europe, a central focus for these developments rested on questions over agency and practical consciousness, ideology, and alienation. Thus positioned, cultural materialism is characterized by a socialist humanism which sets it apart not only from Louis Althusser's structuralist Marxism but also from the antihumanism of post structuralism.

Within these critiques of both official and structuralist Marxism, materialism became to lie primarily within the actual life processes – simply in being alive – and not in confining it to material (as physical) entities, and even less to economic activities. Consequently, for Williams, any concept of culture as a reflection of the material was in fact not material enough since any cultural practice was material practice insofar as it involved human beings working (in the sense of engaging, changing, modifying) their social and material environment.

Such a refocusing towards materiality in the context of social praxis was undertaken initially by Lukács in relation to a social totality that encompasses all social practices. This meant that the base/superstructure duality would no longer be the only, and indeed stifling, backdrop for a Marxist problem of culture. Williams warned of a too general use of such totality as merely encompassing all social practices and indeed urged for a consideration of the intentions that provide an organizing frame for such entirety of social practices. For him, such intentions could not be bracketed off as (class) interest but instead point toward the need for incorporating concepts of power – and with this he clearly points toward Gramsci's hegemony – into an analysis of social ontology. In so doing, Williams (1977: 87) proposes an understanding of determination which "is never only the setting of limits; it is also the exertion of pressures". Yet, aware of the dangers of a voluntarism which affords all explanatory power to individual agency, Williams argued that rather than individuals being able to exert such pressures, it was specific social formations within which such pressures can be built up.

The importance of struggles over culture and cultural practices as 'political' struggles rather than simply matters of taste, distinction, or esthetics marks a key contribution of cultural materialism. It is this contribution which has reverberated throughout the social sciences from historical geographies of the industrial revolution to studies of social movements. With this political understanding of culture as everyday experience, cultural materialism engages with Gramsci's contributions on popular culture and the significance of hegemony in the struggles over what constitutes such culture. Gramsci's understanding of power does not solely focus on oppression but integrates cultural hegemony as a persuasive consensus alongside forceful coercion. As Don Mitchell (2000: 51) puts it "the problem of cultural hegemony is one of determining how and why subordinate people seem to consent to their own domination, why they seem to so frequently go against what 'objectively' looks to be in their own best interest." The conjunction of practices, ideology, and interest becomes obvious: practices are imbued with cultural significations and as such embedded in particular social relations. Hence they always carry particular interests. However, these interests are often neither clearly visible nor identifiable. The concept of ideology which is used to describe these processes of concealment becomes wider than a mere misleading of other (subordinate) people: through processes of alienation, social products (both physical and social) become reified as natural objects.

Politics, Pedagogies, and Cultural Materialism

Alongside other British Marxists such as Edward P. Thompson, Williams' cultural theory rested on the central tenet that culture could be separated neither from the politics nor from the economy. Moreover, it was the struggles and negotiations over culture which informed cultural materialism as agency and classed relations received particular attention – be it in Thompson's seminal *Making of the English Working Class*, or Williams' *Long Revolution.*

The emergence of new social movements across Western Europe, notably student, antinuclear/ecological and feminist ones, constituted a political platform that expanded on some of the earlier popular education projects of the British Left such as the Mass Observation Project with newly established local forums and clubs in the late 1950s. These provided a stimulating momentum across the country for reading and study groups, and a lively mix of contemporary youth subculture as much as older traditions of workers' education. Moreover, exploring the changing cultural institutions and new technologies also provided opportunities for interventions in the politics of education, mass media, and their potential for democratic involvement.

Dominant, Residual and Emergent Cultures

The Long Revolution as one of Williams' earliest studies put forward a specific exploration of social change, calling into question previously held socialist ideas of how politics and culture were interrelated. At its center stood the long term cultural revolution of British social life that was unfolding along three interrelated social processes: a democratic revolution, an industrial revolution, and a cultural revolution. Cutting through all aspects of British social life, this ongoing long revolution was taking place neither in predetermined nor linear ways.

By rejecting that elements of culture such as fine arts or literature could be deduced from the economy, and by instead focusing on the ways in which specific cultural values became dominant in particular societies, Williams offered an anthropological and holistic reading of the social world. In this social world, culture as a way of life was afforded central status.

On the basis of social change, tensions within specific social configurations were played out in such a way that dominant and subordinate cultures existed simultaneously and in the same places. It was, moreover,

> … impossible for a dominant culture to contain subordinate groups and ideas in a thorough way, resulting in cultural practices generating values which were contradictory and contestable. Any social or cultural formation could be divided on a tripartite basis into emergent, dominant and residual elements. Because of the uneasy co existence of these parts, a subversive potential characterised the dominant order. (Kenny, 1995: 91)

It is such an understanding of power relations and struggle over dominance that clearly bears reference to Gramsci's concept of hegemony. It is through such notion of hegemony that cultural materialism regards dominance as an active process – and one which extends far beyond traditional notions of politics into the fabric of society, its traditions, institutions, and cultural formations. Through such active process, Williams (similar to Gramsci himself) placed particular importance on emergent cultures, which correspond with the latter's counter hegemonic forces. Such understanding has attracted two criticisms: firstly, that such emergent cultures at the same time are in danger of harking back to romanticized notions of community; and, secondly, that the extent to which such community is based on the white, male, English working class, and thus exclusive of others had been insufficiently reflected by Williams.

Structures of Feeling

More so than cultural materialism itself, structures of feelings have been most closely attributed to Williams. The vagueness of the concept accounts for some of the continuing engagement with as much as the critique of it. Nonetheless, recent explorations in relation to Foucault's episteme, Bourdieu's habitus and Gramsci's hegemony place the salience of the concept centrally to the study of culture. The concept, not at least by Williams himself, has been applied in a variety of meanings, yet, two key elements can be identified within such structures of feeling:

1. An empirical historical and geographical specificity: not concerned with abstracted human nature, it is actual structures of feeling that are at stake. Neither sociological nor biological, these structures give phenomenological evidence to a consciousness that is intersubjective. Formed by a range of interactive structural processes, these structures become part of emergent cultures.
2. It is in the fields of arts and literature that actual structures of feelings are most readily accessible. However, they may also be found in social and cultural practices of institutions that are not part of dominant or hegemonic cultures.

Filmer outlines that the "relationship between these two constant features implies that structures of feeling are generated through the imaginative interactional social and cultural practices of initiation and response – quintessentially social practices of reflexive communication of experience which are at the root of stability of and changes in human societies." Drawing on the lived experience of people as embodied and practical consciousness emphasizes the social, or rather societal, relations in which individual emotions are enframed.

Such focus on emergent cultures has been central to the explication of structures of feeling as those articulations of something that cannot be yet articulated, and thus defies objectification. In this sense, structures of feeling directly point to the future possibilities of current experiences of the otherness of emergent cultures (cf. Lefebvre's work on articulating such notions within a related frame of Western Marxism).

Conceptually, a number of criticisms have been leveled against cultural materialism (as indeed the wider critical Marxism): most famously, this has been the charge of voluntarism as class struggle and agency was given too much importance over structural determinations (in particular if we think about the argument between Althusser and Thompson). Relatedly, the insistence on a social totality as based on internal relations of its different moments has been criticized for its functionalist and organicist tendencies. Similarly, an all encompassing concept of culture can become too vague to contribute to knowledge production or indeed political commitment.

The Sociology of Culture: Applying Cultural Materialism

Cultural materialism embeddedness within literary criticism was not to a small extent due to Williams' own disciplinary belongings as Professor of Drama and Fellow of Jesus College, Cambridge. However, it was part of a much wider societal process in post war British socialist politics by which literature and previous upper class 'high' culture was seen to become appropriated by the social democratic politics of mass education, social mobility, and mass civilization. Yet, it is in this confluence of politics and academia that Williams and the British New Left went beyond the traditions of literary humanism and embraced a more active socialist humanism as political project.

For Williams this was justified by the extent to which cultural practices and cultural production itself were central to the constitution of society. Here, the sociological sense of culture as a whole way of life means that culture as a signifying system shapes considerably a whole range of social processes. It is with this understanding in mind that Williams alongside others was critical of the pejorative use of the term mass culture (which posited popular mass culture as low culture vis à vis an elitist high culture) and instead preferred to focus on ordinary and everyday cultural practices.

Reflecting the far reaching social changes in the middle of the twentieth century, and in particular the optimism of the British post war social democratic settlement, societal change rested right at the centre of Williams' first two major studies, *Culture and Society* and *The Long Revolution*. Similar to Thompson's *Making of the English Working Class*, these studies were principally concerned with the experiences of social and cultural changes in English social relations. A focus on culture in these studies rested firmly on the concept of community. Challenged through advancing capitalism and the rise of mass production, such community became the focal point for political organization and solidarity – a focal point that Williams maintained throughout his novels also. Yet, and here some of the criticisms leveled at Williams, as much as others of the first New Left, for a romanticized and populist notion of community, are somewhat misplaced. Rather than invoking organicist, nostalgic communitarian values, Williams in *The Long Revolution* investigated the processes by which modernity impacted on cultures and values, and thereby proposed a concept of community that had to be able to act as a source for social subjectivity as much as diversity.

The second key element of a sociology of culture was the more accepted use of culture as artistic and intellectual activities – in literature or fine art. Here, Williams stressed the relevance of the respective signifying practices – including language, philosophy but also media and emerging cultural industries. It is in this second field, that his work on cultural institutions and formations has been developed further in relation to the production of culture.

Cultural Materialism and Human Geography

Culture understood as everyday experiences extends beyond individual lifestyles to social relations, formations, and processes. In fact, such understanding has become one of the most common uses of culture within human geography. For cultural materialism, culture needs to be set in relation to ways of life as much as structures of feeling. These reference points have been significant throughout human geography's engagement with cultural studies more generally. In particular, the work of historical geographers such as Dennis Cosgrove, Derek Gregory, Mark Billinge, and Stephen Daniels engaged with Williams as much as with Thompson, in terms of ordinary cultures, agency, and production of culture, with reference to landscape as much as to the making of class, hegemony, and power. Beyond such general uptake of cultural materialist work, there have notably been three influences of Williams' work on human geography: first, his study into the relationship between the city and the countryside in English literature from the sixteenth to the twentieth century; second, a militant particularism as political stance that was rooted in common lived experiences and provided a concept to explore political activism; and third, a more critical (and reflective) engagement with cultural studies and culturalism within human geography itself.

The Country and the City

Understanding culture as thoroughly implicated in everyday social relations was arrived at through literary criticism which merely remained within the confines of the academic discipline but extended beyond to studies of social history as much as political commitments. Here, *The Country and the City* provided the most explicit en gagement with place and space: while neither of these was one of Williams' keywords, community as a concrete set of social relations was for him also bound by place. Yet, narrowing community down to both a long lost past and a bounded place seems to be too limiting for the uses that Williams put the concept to: instead, the particular ways in which community was constructed in English literature was employed powerfully to investigate and critique the social background of this literature. The central concerns in Williams' study lay in the relational dimensions of community as much as the ways in which the city and the countryside were pitched as opposite poles within social life and culture. Here, literary criticism excavated a social and intellectual history which resonated with much of modernity's classical social theory. Williams found that most novels operate with the concept of a community that is knowable, meaning that people, and their relationships could be easily communicated to the readership. More importantly, however, it was a knowable community that was highly socially selective, as, for instance, in relation to Jane Austen's novels:

> It is outstanding face to face; its crises, physically and spiritually, are in just these terms: a look, a gesture, a stare, a confrontation; and behind these, all the time the novelist is watching, observing, physically recording and reflecting. That is the whole stance the grammar of her [Jane Austen's] morality. Yet while it is a community wholly known, within the essential terms of the novel, it is an actual community very precisely selective. Neigh bours in Jane Austen are not the people actually living nearby; they are the people living a little less nearby who, in social recognition, can be visited. (Williams, 1975: 203)

In contradistinction, the city in English literature became the place where such knowable communities were hidden from view and easy visibility – but had, as in Dickens' work, to be revealed and teased out. The strength of the analysis lies in its careful empirical tracing of cultural practices, cultural formations, and the construction of social relations within these. While its source material, English novels, is of course the source material of literary criticism, the analysis went beyond literary criticism making frequent use of cultural materialism as a theo retical toolbox. It was these sets of poignant observations which filtered into historical geography studies on landscape and the cultural production of these. Here, Daniels' work on the duplicity of landscape and its careful production provides a good example of how geographers drew on cultural materialism.

Militant Particularism and Internationalism

The second strand through which place frequently sur faced in Williams' work is his notion of a militant particularism – standing by one's loyalties where actual lived experiences formed the basis for (working class) solidarity and could not be abstracted from. Harvey in particular had paid close attention to how environment, place, and space provide a referencing framework for cultural materialism. Such attention embeds lived experiences within the daily lives, its social practices, and relationships. That place, and community for that matter, is made and changed by processes beyond the local, was clearly understood by Williams with his own biographical experience of local attachment and internationalism, growing up in a working class family in the Welsh border town of Pandy; similarly so in his views on community (albeit problematically). Relatedly, Thompson's study of the making of working class identities was critically engaged within Gregory's work who emphasized the importance of local studies which would be in a stronger position to trace nuanced versions of working class formation and bounded agency.

For Harvey, Williams' novels offer a far richer en gagement than his more academic works with these concerns over loyalties to rather different social contexts and related structures of feelings. In novels such as *Loy alties*, *Border Country*, and *Second Generation* the tensions between lived experiences and abstract convictions are ever present and unresolved. As Harvey argues:

> Loyalties contracted at one scale, in one place, and in terms of a particular structure of feeling, cannot easily be carried over without transformation or translation into the kinds of loyalties required to make socialism a viable movement either elsewhere or in general. But in the act of translation something important gets lost leaving be hind a bitter residue of always unresolved tension. (Harvey, 1996: 39)

Yet, again in relation to earlier points made about the localness of such experience of solidarity and a militant particularism, Featherstone critiques the extent to which Harvey sees it as bounded and thereby restricted to lo cality and community. Instead, militant particularism is able to provide a suitable framework for exploring recent resistance to neoliberalism. Exploring the ways in which such resistance is formed and networked in eighteenth century social struggles in England, Featherstone points to how networks of social relations operate across dif ferent scales and can indeed provide points of bringing together different types of militant particularisms. At the same time, he remains attentive to the existing tensions within Williams' work over 'authentic' experience (structures of feeling) and a lack of exploration in how far different experiences of internationalism/localism were indeed connected up.

Beyond 2000 and Towards the Future?

Surveying the relevance of cultural materialism within human geography points finally to its role in providing an explicit frame to engage critically with cultural studies and cultural geography more specifically. Much of this relevance originates from within Williams' own enduring critique of post structuralism and the role it affords to representation. Within human geography itself, Harvey's exploration of militant particularisms is an important reference point. The second one is Mitchell's work on and within cultural geography. The latter's provocation 'that there was no such thing as culture' took on the postmodernist 'new' cultural geography in particular for its inability to hold on to the 'indissolubility' of culture from political economy, and in turn for its reification of culture, and in particular material culture.

One of the interests for employing Williams' work, as much as, for instance, Henri Lefèbre's or other Marxists who were on the sidelines (or in active opposition to) orthodox, official Marxism while at the same time not taken in by structuralist readings of Marx (such as those by Althusser), lies in the attentiveness not only to agency but also to everyday experiences, while, as Mitchell stated above, positioning this within a Marxist political economy. The firm rootedness of Williams' work within English traditions of literary humanism, the political experiences of the New Left and British Marxism are evident throughout his writings, yet, at the same time, lines of communication, exchange and political concerns extend well beyond the UK, most visibly in relation to Western Marxism, but also in his continuing critical engagement with post structuralism.

Williams, as Mulhern (2000: 72) argues, set out to develop "an integrally socialist politics of culture" which would maintain a political commitment without falling into either elitist cultural criticism or populist cultural studies. It is with such political project in mind that Williams' writings cannot be subsumed under either literary criticism or cultural studies. In collections such as *Resources of Hope*, posthumously published in 1989, or the earlier *Towards 2000* the political analysis of the presence and the past is employed decisively to examine what he calls "Resource for a journey of hope" (Williams, 1983). The strength of his call lies not a romanticized view of communities gone by but instead in trying to understand how contemporary political crisis management as the management of actual social relations and actual experienced pressures can be employed for the analysis of dominant social orders, how they are maintained and challenged so "… that this kind of general humanist analysis can significantly contribute to thinking about the future" (Williams 1983: 15).

See also: Culture; Determinism/Environmental Determinism; Hegemony; Historical-Geographical Materialism; Humanism/Humanistic Geography; Ideology; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Social Class; Structural Marxism.

Further Reading

Althusser, L. (1969). *For Marx*. London: Allen Lane.

Billinge, M. (1984). Hegemony, class and power in late Georgian and early Victorian England: Towards a cultural geography. In Barker, A. & Gregory, D. (eds.) *Explorations in Historical Geography*, pp 28 67. Cambridge: Cambridge University Press.

Daniels, S. (1989). Marxism, culture, and the duplicity of landscape. In Peet, R. & Thrift, N. (eds.) *New Models in Geography*, vol. II. pp 196 220. London: Unwin.

Eagleton, T. (1989). *Raymond Williams: Critical Perspectives*. Cambridge: Polity Press.

Featherstone, D. (2005). Towards the relational construction of militant particularisms: Or why the geographies of past struggles matter for resistance to neoliberal globalisation. *Antipode* 37, 250 271.

Filmer, P. (2003). Structures of feeling and socio cultural formations: The significance of literature and experience to Raymond Williams's sociology of culture. *British Journal of Sociology* 54, 199 219.

Gregory, D. (1982). *Regional Transformation and the Industrial Revolution*. London: Macmillan.

Harvey, D. (1996). *Justice, Nature and the Geography of Difference*. Oxford: Blackwell.

Jackson, P. (1989). *Maps of Meaning. An Introduction to Cultural Geography*. London: Routledge.

Jones, P. (2004). *Raymond Williams's Sociology of Culture: A Critical Reconstruction*. Houndmills: Palgrave.

Kenny, M. (1995). *The First New Left: British Intellectuals after Stalin*. London: Lawrence and Wishart.

Mitchell, D. (1995). There's no such things as culture: Towards a reconceptualization of the idea of culture in geography. *Transactions of the Institute of British Geographers* 20, 102 116.

Mitchell, D. (2000). *Cultural Geography. A Critical Introduction*. Oxford: Blackwell.

Mulhern, F. (1989). Towards 2000, or news from you know where. In Eagleton, T. (ed.) *Raymond Williams: Critical Perspectives*, pp 67 94. Cambridge: Polity Press.

Mulhern, F. (2000). *Culture/Metaculture*. London: Routledge.

Simmel, G. (1993). The metropolis and mental life. In Frisby, D. & Featherstone, M. (eds.) *Simmel on Culture*, pp 174 185, (first pub. 1903). London: Sage.

Thompson, E. P. (1963). *The Making of the English Working Class*. London: G. R. Elton.

Williams, R. (1961). *The Long Revolution*. London: Chatto and Windus.

Williams, R. (1975). *The Country and the City*. St Albans: Paladin.

Williams, R. (1976). *Communications*. Harmondsworth: Penguin.

Williams, R. (1977). *Marxism and Literature*. Oxford: Oxford University Press.

Williams, R. (1983). *Keywords: A Vocabulary of Culture and Society*. London: Fontana.

Williams, R. (1983). *Towards 2000*. London: Chatto & Windus, the Hogarth Press.

Williams, R. (2005). *Culture and Materialism: Selected essays*. London: Verso.

Relevant Websites

http://www.massobs.org.uk/index.htm
The Mass Observation Project.

Cultural Politics

S. Rycroft, University of Sussex, Brighton, UK

Glossary

Appropriation The process whereby elements of hegemonic culture are taken and made to mean something new and exclusive to counter-hegemonic social groups.

Counterculture A broad social group, originally conceived as being of middle-class origin, which uses a range of political, cultural, and direct-action tactics to resist hegemonic culture, and which often attempts to establish permanent alternative ways of life.

Counter-Hegemonic Relating to social groups that in a variety of ways challenge the dominant order of a society.

Hegemonic Culture A faction in a society who represents dominant values and mores.

Incorporation The process whereby a counter-hegemonic group is absorbed into hegemonic culture, usually through commercialization of their styles.

Subculture A social group, originally conceived as being of working-class origin, which coalesces around a specific set of styles and values and thereby resists hegemonic culture.

Introduction

Cultural politics is a term that has a 40 year history but only entered the field of geographical studies in the mid 1980s. At that time, when agendas were being set for a 'new' cultural geography, one approach suggested was to utilize a notion of cultural politics that had been developed by sociologists over the previous 20 years. Geographers were asked to attend to culture as a politically contested social construction and to do so they needed to adopt a broad conception of culture that moved beyond a focus on elite cultural forms and explored more popular and resistant forms of cultural practice. Cultural politics has become a concept used in helping explain the dynamics of cultural processes and is central to the work of new cultural geographers. Indeed, it is difficult to find a piece of work in cultural or social geography that does not deal with cultural politics in some measure. This is because cultural politics refers to the tactics and strategies used by social groups, both powerful and oppressed, to express their views of the world.

In the mid 1980s, so called 'new' cultural geographers in the UK and the US adopted cultural politics as a broad theoretical framework with which to understand the effect of culture on everyday lives and geographies. Initially, there were slight divergences between the US and the UK approaches. While geographers on both sides of the Atlantic were trying to develop more complex and nuanced understandings of culture and its role in political processes, in the US there was already a long and established history of cultural geography. There, the move to a cultural politics approach revolved around the rejection of an established approach to cultural geography (sometimes called the Sauerian/Berkeley School approach) which viewed culture as a pregiven thing rather than a process. In the rejection of these established approaches, American cultural geographers initially emphasized the role of the individual in affecting and changing culture in a kind of humanist Marxism. In the UK, without that disciplinary history to address, the turn to cultural politics was more directly informed by a theoretical approached form outside of the discipline through the pioneering work carried out by the Centre for Contemporary Cultural Studies (CCCS) at Birmingham University. Within a few years, however, as the two approaches informed one another through publications and conferences, there was little difference between the works of 'new' cultural geographers on either side of the Atlantic in terms of their conception of cultural politics.

This article has two main sections. The first analyzes the ways in which cultural politics has been theorized. In this section, the ideas of the CCCS take center stage. It must be noted from the outset, however, that the CCCS offered a pure and perhaps overly systematic theory of cultural politics that geographers rarely take on board wholesale. However, it is in the work of the CCCS that concepts central to the notion of cultural politics as adopted by geographers, such as hegemony, counter-hegemony, resistance, and the connection between socioeconomic and cultural processes, were most clearly articulated. The next section works through some of the ways in which geographers have employed these theories especially with respect to studies of resistance, and also emphasizes how the theories have been changed by geographers.

Hegemony, Culture, and the Political Economy

Hegemony

An influential concept of cultural politics emerged from a series of studies of resistant cultural practices carried out by the CCCS throughout the 1970s. The CCCS concerned themselves with the manifestation of cultures of resistance in terms of the ways in which they reflected a position in relation to the forces of production. Broadly speaking, this means that they adopted a Marxist perspective on culture primarily influenced by the works of Antonio Gramsci and Raymond Williams.

A new understanding of hegemony worked through by Gramsci and Williams was central to the CCCS' early work. Prior to Gramsci's reworking of the idea, hegemony was conceived as a condition of uncontested political supremacy: one dominant social group dictated the social, cultural, and political norms of subordinate orders. Gramsci's notion of hegemony, however, was based upon an understanding that people are not coerced by dominant culture but, rather, accept the dominant ideas of society, while they are also empowered to challenge them from below. Hegemony, in this conception then, was always strived for, but never fully achieved by the dominant or hegemonic culture.

Raymond Williams was concerned with theorizing cultural processes and finding a more active role for culture than had previously been recognized by Marxist scholars. For Williams, culture was not simply a function of the political economy as most Marxists felt, but in fact crucial to creating the circumstances for political and economic change: cultural change and changes in the political economy were entwined and inseparable. Using Gramsci's ideas, Williams suggested that hegemony was a process and not a defined system or structure and that the reality of cultural processes must always include the efforts and contributions of those who are outside of a hegemony. While hegemonic order was still understood to set the limits through which conflicts move and were resolved, it would not prescribe ideas because groups or classes which were not powerful could express their subordinate position and experiences through their culture. In other words, we must attend to the experiences and cultures of subordinate groups to understand not only cultural change, but also change in the political economy.

The Cultural Is Political

Armed with these theoretical perspectives the CCCS set about evolving an understanding of cultural politics through theoretical explorations derived from empirical studies. The CCCS was established at a time characterized by significant social upheaval. The post war period had witnessed the emergence of numerous subcultures and countercultures, each characterized by a particular style, lifestyle, taste, and outlook. The project of the CCCS was to explain the emergence and function of these cultures of resistance, and to assess their role as agents of political and economic change. They did so mostly through ethnographic studies of various groups, including Mods, Bikers, and Hippies, and wider studies of cultural political themes such as racism and feminism.

Cultural politics, as formulated through these studies, built upon the broad definition of culture and its processes as developed by Williams. In this conceptualization, culture has a direct connection to the material base because culture is the medium through which people use everyday resources and transform them into objects and practices that are invested with symbolic meaning and value. Culture is political too because it is also the medium through which different versions of the world, as expressed through these meanings and values, are contested.

In their studies of post war cultures of resistance, focused almost exclusively upon youth cultures, the CCCS first distinguished between subcultures and countercultures – a distinction which is rarely made today. These two forms of resistant cultures differ in their 'parent cultures'. Subcultures like Teddy Boys, Mods, and Rockers come from working class roots, whereas members of countercultures, like the hippies, come from middle class backgrounds. Members of a subculture have a more immediate connection to the material (economic) base because they are required to work in order to facilitate their participation in the group. A classic example of this was the stereotypical Mod who worked in menial, low paid work from Monday to Friday and used their wages to finance their weekend leisure time, often maximizing that time through the use of amphetamines. Those direct connections with the world of work means that incorporation back into the mainstream is more or less assured as members of the subculture mature and acquire the responsibilities of adulthood.

Middle class countercultures, on the other hand, have a less immediate, although no less significant, connection to the material base. Middle class youth are afforded the material and nonmaterial resources that allow some of them to 'drop out' for extended periods. Because middle class youth have a safety net in terms of financial security and sets of skills afforded by greater educational opportunity, it is possible for them to take the kinds of risks that cannot be entertained by working class youth. They can drop out and do something unconventional. The kinds of activities that such people often engage in are entrepreneurial – underground media production, music performance and production, craft based small scale manufacture, organic farming, and so on – and so quite easily incorporated back into hegemonic culture and economy.

It is that process of incorporation that is central to the revised idea of hegemony as well as the possibilities and the limits of cultural politics. Cultures of resistance are founded upon a series of appropriations of everyday resources and the transformation of those appropriated resources into something that is meaningful for them. In most cases, it is a style that is appropriated and made symbolic: in the case of the Teddy Boys it was Edwardian suits; for Mods, Italian modernist style; for skinheads, Crombie overcoats normally favored by city stock brokers; and for the hippies a wider range of styles, in cluding musical motifs, Asian designs, native American artifacts, and so on. These objects are invested with ex clusive, implicitly oppositional meanings specific to the culture of resistance. For the CCCS it is these restyled objects that form the glue that hold cultures of resistance together, allows them to function, and defines them to those outside of the group. The challenge to hegemony in this way is never a direct articulation, rather, it is ex pressed obliquely in style, and it is that style that is eventually incorporated back into hegemonic culture. What begins by offering symbolic challenges to authority, ends by establishing new hegemonic conventions, in corporated economically as a commodity, or incorpor ated ideologically through the media.

But hegemonic culture itself does not remain un changed by this process. In fact, change in the hegemonic order that is brought about by the incorporation of subordinate cultures is central to explaining changes in the broader political economy.

Cultural Politics versus Politics Proper

This process can be illustrated with reference to the 1960s counterculture. The kinds of hegemonic values that they were resisting were precisely the values that have shifted in the West since the 1960s: attitudes toward the institutions of marriage and the family for instance, the role of women in the workforce and public life generally, the more general acceptance of alternative ways of being, and so on. One could argue that, in a sense, this change in the hegemonic order is a kind of victory for the counterculture. However, a different argument is made by the theories developed by the CCCS, namely, that there was a crisis of hegemony during the post war period characterized by the breakdown of class relations and of control over the youth of the dominant class. This crisis allowed the space for the middle class counter culture and countless working class subcultures to es tablish themselves. These cultures of resistance actually helped provide the social and cultural circumstances that facilitated a shift in the political economy. Western economies transformed in the post war period into what some have called a postindustrial state, where con sumption rather than production became far more sig nificant in driving economic growth and more generally, our everyday lives. In order to facilitate this change, at titudes toward consumption and commodities in general had to change. Because a crisis in hegemonic order allowed the proliferation of many alternative lifestyles, each of which, significantly, based upon different and shifting consumption practices, gradually, hegemonic attitudes toward consumption changed.

Because style, the appropriation and transformation of style, and the hegemonic incorporation of style is central to this theory of cultural politics, style was the focus of research for scholars at the CCCS. The nature of a subcultural or countercultural ideology they felt could be read through the encoded signs of dress, lifestyle, and music. Political points are made by resistant groups through the transformation in and interpretation of everyday objects as style. This is where they saw the ultimate weakness of cultural politics: ultimately, the resistant group may be ignorant of its higher revo lutionary project precisely because attacks on authority are generally expressed obliquely through style. The root causes of oppression and alienation are never directly articulated, nor are the fundamental causes or structures which affect existence. It was in the arena of politics proper that cultural politics met its limits because the struggles it represented are waged at the level of lifestyle.

The Geographies of Cultural Politics

It is with the logical conclusions of this systematic ren dering of cultural politics that most contemporary geographers take issue. It would be a mistake therefore to assume that whenever one encounters the phrase in the work of a scholar, or indeed in the rhetoric of a social or political movement, that it is meant in the pure theore tical form set out above. Many scholars, for instance, have serious doubts about the connections between appropri ation of a counter hegemonic style and the kinds of wide ranging effects on the political economy predicted by the CCCS. Many too find the theory inflexible and monolithic, unable to account for the complexities of involvement in particular forms of resistance. It is only if one considers cultural politics as defined by consumption practices that this aspect of the CCCS theory works. However, when geographers use the term cultural pol itics they are often referring to aspects of culture and society that are not easily and simplistically reduced to style and consumption.

The cultural politics associated with the geographies of class, gender, race, ethnicity, and sexuality, for instance, is more than stylistic politics; it involves very real pol itical struggles and tactics that do indeed demonstrate a conscious articulation of the root causes of alienation and oppression. Such identity politics, while there might be some associated consumption practices, are far more than just consumption. They frequently emerge not just as a stylistic choice but also as a reaction to marginalization. For example, the cultural politics associated with the hip hop and gangsta subculture of Southcentral Los Angeles, not only combines a recognizable, appropriated, and refigured style, but also articulates, through powerful lyrics and performance, conditions of oppression, violence, racism, de industrialization, and economic marginalization. The fear and moral panic that this subculture continues to engender is itself testament to an effective political strategy.

Geographers working on cultural politics have been at the forefront of recent challenges to systematic theories of cultural politics. Grounded geographical studies of cultural politics have led to a revision of the concept, questioning the role of agency and the binary nature of the theory. Recent work has suggested that it is rather difficult to sustain the binary opposition of resistance and domination explicit to these theories. Geographers have begun to employ 'post structuralist' ideas – that is, ideas that move away from the restrictive structured and sys tematic explanations (like those of the CCCS) – to help explain the messy reality of situations on the ground. Cultural politics is now beginning to be understood 'where' it takes place and not through theories of dom ination and resistance. Grounded geographical studies of resistance then suggest that there are many flows and connections between dominant and dominated that cannot easily be mapped onto the theory.

Cultural Politics and Spatial Politics

The geographies of cultural politics are not simply a question of taking the concept outlined in the first section and mapping it. Cultural politics are also, at the same time, geographical politics. Cultural politics and spatial politics are mutually constitutive and spatial strategies are either intentionally or unintentionally integral to the formation of cultural politics. Cultural ideologies, hege monic or counter hegemonic, are spatial and affect spa tial practices. Indeed, dominant cultural ideologies are often difficult to map and it is only when those ideologies are challenged through transgression that they manifest themselves. When modes and norms of behavior are challenged by transgression, hegemonic discourses tend to come to the fore, identifying and attacking values and behaviors that seem out of place. In other words, when a grounded geographical approach is used, the nuances and complexities of cultural politics are revealed – they are not everywhere the same.

It might be that a particular movement is not con sciously employing spatial strategies in its tactics, but either the spaces in which its activities occur shapes its cultural politics, or its cultural politics actually creates spaces out of existing spaces. The same space has mul tiple versions, hegemonic and counter hegemonic. Cer tain spaces have normative hegemonic values attached to them and it is often the tactic of subordinate groups to subvert or transgress these hegemonic values. Sometimes this can be spectacular such as in the case of mass demonstrations or occupations, and sometimes it can be subtle, like willful disobedience in the classroom. Geographers have worked on role of space in the cultural politics of sexuality, for instance. In these studies, they have mapped out the ways in which everyday spaces, which are normatively valued as heterosexual, can be transformed into homosexual spaces through a variety of cultural political practices, including listening to par ticular genres of music, couples openly displaying their affection for one another, and so on. Through such practices everyday spaces are transformed allowing for acts that would otherwise be deemed inappropriate in those spaces. Moreover, through this combination of spatial and cultural politics subordinate identities can be expressed and affirmed.

Cultural politics then can also inscribe space with identity, literally appropriate and redefine a particular space in exactly the same way that styles are appropri ated. The most obvious example of this is graffiti, where spaces and territories are marked for a variety of reasons, from gang activity to political statement. But there are other forms too. Certain territories become integral to the performance of alternative identities. The disused warehouses that remained in the northwest of England after de industrialization, for instance, became spaces crucial to the development of cultural politics associated with the rave scene during the late 1980s and early 1990s. Many subcultures have become associated with certain territories. Often, these associations are the scenes of spectacular events such as the battles on the beaches of Brighton and Clacton between the Mods and the Rockers during the early 1960s. It is not that these spaces provide a 'home' to these groups, but rather that they function as symbolic spaces central to the functioning of the subculture.

Cultural political groups can use space for their own ends and in doing so also create new spaces. The con temporary global anticapitalist movement, for instance, has been characterized as a 'convergence space': a global space in which numerous and often politically divergent local and global resistant groups converge. The space created by this alliance is not a physical space but a networked space with a variety of nodes and connections. These convergence spaces are not so much expressing a common identity as a common cause and in doing so, use already existing spaces both real and virtual (such as the Internet) to develop their politics.

The Place of Cultural Politics

Cultural politics is also situated politics, that is, it can take place in specific locales. In some cases, the associ ation between territories and particular forms of cultural politics become more permanent and long standing and situated cultural politics develops in tandem with mor phological changes in the territories to which it has be come associated. Many of the territories associated with countercultural living during the 1960s such as the Haight Ashbury district of San Francisco or Greenwich Village in New York are good examples of this process. Another, Venice Beach in Los Angeles, California, serves as a useful illustration: Venice, sometimes alternatively known as Venice West, was constructed in 1905 as a fantasy version of the real Venice, incorporating canals complete with gondoliers, lined with colonnaded build ings, designed as a resort town for the elite of downtown Los Angeles. By the late 1940s the resort had fallen into disrepair and cheap rents allowed it to be colonized by a community of Beat poets and artists. They tended to occupy and perform in the mock Venetian infrastructure of the district and their cultural politics became closely associated with that infrastructure. Later, in the 1960s, as Beat cultural politics evolved into countercultural re bellion in Venice, the same buildings became the bases of operations. This association between cultural politics and physical infrastructure was so strong that any inter vention by the city authorities was seen as part of a wider campaign to rid the city of undesirable elements. There were a series of city ordinances that targeted counter cultural activities such as impromptu performance, and a redevelopment initiative in the form of a Master Plan that was proposed by the City of Los Angeles in the late 1960s. The attack on the Venetian architectural esthetics engendered in the plan was perceived and resisted in the same way as concurrent ordinances directed toward as pects of countercultural lifestyle.

The symbiotic connection between cultural politics and the physical place of Venice West remains so strong be cause rather than appropriating a style, the cultural politics of the area is based upon the material appropriation of a distinctive physical space, a place. In this way then cultural politics can sometimes find a strategic geopolitical base. This need not imply that the place in which particular forms of cultural politics are rooted is not contested, that there are not any alternative, hegemonic versions.

Elite Cultural Politics

So far we have exemplified ideas about cultural politics and its geographies by using counter hegemonic groups. However, cultural politics is not the preserve of the op pressed or the resistant because within hegemonic culture there is more than one cultural configuration in operation. Advanced industrial societies are characterized by a mass of competing forces – social, political, economic, and cultural – which challenge and shape dominant forces. These challenges need not only arise from the oppressed. It is worth looking at a case of cultural politics from middle class hegemonic culture, not just to redress this imbalance, but also to help understand the operation of cultural politics in an everyday contemporary environment.

Numerous middle class communities are currently being established on the periphery of large cities in places like Orange County in Southern California. Sometimes called 'edge cities' or 'exurban communities', these developments have become a significant aspect of urban development over the past two decades not only in the West, especially the United States, but also in more prosperous pockets of the developing world. These ex urban communities are often political entities with city status and thus the power to raise taxes and spend public money. Many academics have written about the ways in which these communities are exclusive and a response to urban failure in the large cities that they orbit. Indeed, often, the residents of edge cities are fleeing the effects of cultural politicking that we have outlined above, espe cially the cultural politics associated with race and eth nicity. But it is also interesting to note too the operation of cultural politics within their boundaries.

What strikes one about exurban communities is the extent to which residents are prepared to restrict their freedoms in order to live there. Specifically, there are a range of regulations related to the maintenance of en vironmental esthetics: residents have a very limited pal ette from which to choose a color for their homes; planting schemes for gardens are prescribed, and gardens must be maintained to a high standard; garbage can only appear on certain days; and certain activities usually associated with street life such as ball games and car maintenance are also often forbidden. In the non residential environments of these communities, be havioral restrictions also apply, including ordinances on inappropriate apparel and accepted codes of conduct. At first glance this might appear to be a heavy handed hegemonic strategy, easy to identify, and thereby resist (and indeed there are a growing number of youth sub cultures emerging in these communities). It is, however, no different from cultural politics operated by sub ordinate groups. These restrictions are all directed toward maintaining the image of the community through environmental management practices. A defined set of environmental values are being enacted in middle class cultural politics in these cities, just as a differently de fined set of environmental values form the cultural pol itics of more 'radical' ecological groups. Cultural politics is not the preserve of the young and dispossessed, they form the core set of values which inform actions in the world and make everyday geographies.

Recent Developments

We end with a brief note on recent trends in the geog raphies of cultural politics. As noted above, geographers have begun to use poststructuralist approaches to cultural politics, effectively moving away from the systematic understanding of culture and cultural politics outlined in the first section. This trend is something that has been identifiable over the last 10 years. Geographers especially have been at the forefront of this move and they are beginning to think differently about culture. They are abandoning binary models and thinking about culture less as the product of a system and more as a set of social practices and performances. It is only abstract theories that suggest that at some stage those social practices and performances add up to a 'culture'. The systems that we discuss when we talk about cultural politics and its geographies – hegemony, normative space, environmental values, etc. – are simply collections of creative social practice, and those creative social practices are increasingly the focus. In practice then this has meant that studies of cultural politics by geographers now look at situations of resistance and domination on the ground less as a set of symbolic acts that are easily decoded for those with a knowledge of the broader culture, and more as a set of repeated practices and performances.

See also: Cultural Geography; Culture; Identity Politics; Resistance; Youth/Youth Cultures.

Further Reading

Cresswell, T. (1996). *In Place/Out of Place: Geography, Ideology, and Transgression*. Minneapolis, MN: University of Minnesota Press.

Hall, S. and Jefferson, T. (eds.) (1976). *Resistance through Rituals: Youth Subcultures in Post War Britain*. London: Hutchinson in association with the CCCS, University of Birmingham.

Ingham, J., Purvis, M. and Clarke, D. B. (1999). Hearing places, making spaces: Sonorous geographics, ephemeral rhythms, and the Blackburn warehouse parties. *Environment and Planning D: Society and Space* 17, 283 305.

Nandrea, L. (1999). Graffiti taught me everything I know about space: Urban fronts and borders. *Antipode* 31, 110 116.

Pile, S. and Keith, M. (eds.) (1997). *Geographies of Resistance*. London: Routledge.

Rose, M. (2002). The seductions of resistance: Power, politics, and a performative style of systems. *Environment and Planning D: Society and Space* 20, 383 400.

Routledge, P. (2003). Convergence space: Process geographies of grassroots globalization networks. *Transactions of the Institute of British Geographers* 28, 333 349.

Rycroft, S. (2003). Mapping underground London: The cultural politics of nature, technology and humanity. *Cultural Geographies* 10, 84 111.

Sharp, J. P., Routledge, P., Philo, C. and Paddisone, R. (eds.) (2000). *Entanglements of Power: Geographies of Domination/Resistance*. London: Routledge.

Skelton, T. and Valentine, G. (eds.) (1997). *Cool Places: Geographies of Youth Cultures*. London: Routledge.

Cultural Studies and Human Geography

D. Bell, University of Leeds, Leeds, UK

Glossary

Hegemony A key concept in cultural studies, drawn mainly from the work of Antonio Gramsci, who used it to theorize how powerful groups maintain their power not simply through the threat of force, but by gaining the consent of the subordinate, thereby naturalizing power structures in society.

Introduction

Human geography has long had a reputation as a magpie discipline – scavenging the shiny ideas of other subjects, and making new uses of them. For many people working in human geography, this is a real strength; the openness or porosity of the subject means it is always transforming, morphing, moving in new directions. For critics, conversely, this signifies everything that is bad about human geography: it is a rag bag – a messy mix of bits and pieces – so it lacks coherence and is too faddish. The debate between fans and critics of this peculiar characteristic of human geography can become very heated, especially when it seems that new ideas might be transforming human geography so as to make it almost unrecognizable. One such debate occurred around the concept of culture, and the traffic in ideas between human geography and cultural studies. This interchange is sometimes referred to as the 'cultural turn' in human geography and the 'spatial turn' in the humanities and social sciences (which includes cultural studies). This, then, is a story of interactions between bodies of knowledge, ways of thinking and doing, and academic subject areas. As will become clear, these 'turns' to culture in space and to space in culture have been incredibly productive, and recognized as such, while also generating heated debate. This, too, seems to be something positive about human geography: the scrutinizing of the changes in the discipline, the debate, and dissent. It is certainly a sign of life.

Cultural Studies

This section concentrates on British cultural studies as a distinct and important segment of a broader intellectual project with different trajectories in different places and times, but with a common aim to explore and understand 'culture' – though it should be noted at the outset that the interplay between cultural studies and human geography is now thoroughly internationalized. Cultural studies began to form in the UK as an academic project in the 1960s, although its origins reach even further back than that. The story of British cultural studies shows it to be similar to human geography in its openness to different ideas and perspectives. It has to begin with the often repeated problem: how to define what 'culture' is. Cultural studies textbooks always start with this core question, and usually work it away from a narrow definition of culture as the arts or as analogous to civilization, as when someone is described as 'cultured' (these ideas are sometimes tagged in cultural studies using terms like high culture), and toward a broader conceptualization of culture as 'ways of life'. This reflects early debates that influenced cultural studies, and reverberate through it to this day. These debates are important because they also reveal the political dimension to cultural studies: it is not just the objective analysis of culture, but is an intervention into those ways of life. For some commentators, a clear political commitment is what defines cultural studies – more so than any particular theories, methods, or thinkers define it. As it became consolidated as an academic subject, cultural studies retained this political edge, and built up a toolkit to ask important questions about ways of life as they are, and as they might be. In Britain, cultural studies began to ask and try to answer these questions, drawing on a range of perspectives drawn from politics, philosophy, linguistics, social theory, and psychoanalysis. At the heart of these enquiries was the question of power: how does culture work to maintain relations of dominance and subordination? How do powerful groups form, how do they retain power, and how do they convince the rest of us that it is right that they are powerful and we are not? The late 1960s was a period when these kinds of questions were being asked outside of academia, of course, in protests and strikes, in the ideas of the hippies and the emerging counterculture, by the workers' and students' movements, and so on. Cultural studies was, therefore, a product of its times.

In the UK, the institution most synonymous with cultural studies in this period was the Centre for Contemporary Cultural Studies (CCCS) at the University of Birmingham. In particular, with Stuart Hall as its director, the CCCS developed a perspective on culture inspired by an assortment of theories and thinkers, combining Marxist and post Marxist ideas with structuralist and poststructuralist linguistic theory and Freudianism, to produce new insights into culture as politics. Among the key interests of the CCCS were the workings of the powerful, to which they applied ideas drawn from thinkers like Antonio Gramsci (his concept of hegemony, in particular) – ideas they later applied to analyze the phenomenon of Thatcherism in the UK. But they were also interested in resistance and opposition, and in particular, how the cultural sphere could become the stage for contesting power. Perhaps most famously, they developed an interest in so called 'subcultures', exploring how and why young people, in particular, formed into 'tribes' – such as 'mods', 'teddy boys', or 'punks' – and how the kinds of cultural products and practices of these groups could be understood as a symbolic working through of their experiences of oppression, marginalization, or powerlessness. The innovativeness and the notoriety of cultural studies led to a diffusion of the subject through higher education, particular in the late 1980s and 1990s. Cultural studies expanded rapidly, and its ideas blossomed and multiplied. Its influence was also seeping into other subjects by this time, including human geography.

The Cultural Turn

The story of the impact of cultural studies on human geography is often referred to as the cultural turn, and this is now a story that has been told many times. In the UK, a series of events and conferences organized by the Social Geography Study Group (SGSG) of the Institute of British Geographers (IBG) was an early landmark, during which the question of culture was hotly debated. This turn to culture, as we shall see, marks something of a paradigm shift in human geography – especially in overturning its previous reliance on ideas of positivism and spatial science. It should be noted, of course, that there had been something called cultural geography for decades prior to this; however, what came to be known as the 'old' cultural geography was untouched by cultural studies, and was a specialized subdiscipline concerned with mapping 'culture regions' or 'culture areas' – geographical locations where particular cultures existed. It therefore looked for the material traces of a culture, such as distinct building types, ways of speaking, customs or artifacts, and then sought to describe their distribution, without recourse to theories of culture. These so called 'superorganic' views of culture were more widespread in the USA, where they are especially associated with Carl Sauer and the subsequent 'Sauerian tradition' and 'Berkeley School'. In the UK, cultural geography was something of a marginal pursuit; nevertheless, by the late 1980s the concept of culture, and the ideas of cultural studies were being noticed by some branches of human geography. The SGSG's interest in rethinking how culture and geography might be brought together led to its re branding as the Social and Cultural Geography Study Group (SCGSG), and to the publication of *New Words, New Worlds: Reconceptualising Social and Cultural Geography* in 1991. This small, desktop published book produced by the Study Group is an understated benchmark in the cultural turn, and a record of the liveliness of debates and the enthusiasm among a generation of social geographers to think seriously about culture and about cultural studies, and to the birthing of the 'new cultural geography'. This prompted considerable debate about relations between 'old' and 'new' cultural geographies, with considerable name calling on both sides. The heat generated by this initiative, therefore, also stimulated a much broader turn to culture in human geography as more and more scholars picked up on the ideas and writings pouring out of cultural studies. While today there are some debates about the true extent of this turn, and of the appropriateness of even calling it a turn, there is little doubt that something happened to the shape and form of many aspects of human geography around this time.

This cultural turn coincided and overlapped with the impact of assorted ' isms' spreading across academic subjects. Feminism and postcolonialism were challenging the 'centricities' of sites of knowledge production and bodies of knowledge, but it was postmodernism that arguably became (rightly or wrongly) most associated with the cultural turn. The postmodern disavowal of grand narratives, its collapse of the distinction between image and reality, and its restless experimentation made a lasting impact on human geography, too. But at the same time, postmodernism became quite an easy target for those critics who worried about the excesses of the cultural turn – whether these were in the form of incomprehensible theoretical writings, or a perceived floating free from reality and from 'real issues', or a ridiculous chase to find the next trendy thinker to shoehorn into geography. While there may be some truth in all those charges, the backlash against this 'new' cultural geography, paralleled in a backlash against cultural studies, applied a frustratingly conservative brake to new developments. For many in the discipline, the capriciousness of all this 'newness' could be written off as whacky pretension, its key thinkers dismissed as dilettantes and wannabes.

Yet, at the same time, the 'new' cultural geography and cultural studies were blooming; publishing houses pumped out books and journals, conferences were filled with sessions on cultural themes, cultural geography became a mainstream part of most curricula, and cultural studies witnessed growth in student numbers, in reputation, and in impact. In geography, new journals such as *Ecumene* (later renamed *Cultural Geographies*) and *Social and Cultural Geography* swelled with articles from this bloom. Both subjects arguably suffered a similar set of fates: their trendy faddishness allowed their critics to refuse to take them seriously, feeding broader debates about the 'dumbing down' and 'trendification' of higher education, and of the humanities and social sciences, in particular (media studies was often welded to cultural studies in this regard). At the same time, having made, for many people, a convincing case for culture, the ideas and methods of cultural studies and the 'new' cultural geography began to be adopted by other disciplines and subdisciplines, until (almost) everyone was getting interested in culture. Cultural history, cultural sociology, and other similar subjects laid claim to culture, and moreover, flexed their longer disciplinary histories to say that they had always really been doing culture – so there was no longer any need for a specialist area called cultural studies. Similarly, in human geography, while the cultural turn had its origins in social geography, soon it too was influencing and impacting on economic geography, political geography, historical geography, and so on. Suddenly, everyone was doing culture. Paradoxically, around the same time, it seemed everyone was doing space.

The Spatial Turn

Around the time of the turn to culture, something else was afoot in the social sciences (and the humanities): the 'rediscovery' of space. Space had been disappeared from the mainstream social science agenda, arguably because it had become the province of human geography, so it was out of bounds for sociologists and social theorists. If they were interested in a variable, it was time, not space. But then something changed, and space became 'sexy'. The interest in postmodernism helps account for this spatial turn, especially since time was seen as collapsing into space, and into instantaneity, as a result of the speeding up of the world. Key thinkers writing about postmodernism started talking about space, for example, via discussions about architecture, and the spaces of a postmodern mall or hotel lobby. Space became an interesting theoretical device, and also provided a rich set of metaphors for understanding the postmodern world: maps of this, cartographies of that, topographies, landscapes, sites, locations, etc. Moreover, there was a convergence with one key branch of the 'new' cultural geography, which was concerned with understanding landscapes. This branch approached landscape like a 'text', full of meaning (both explicit and implicit), and this 'textual' approach chimed with the use of spatial metaphors. The connection with postmodern semiotics – the analysis of 'signs' and their meanings – made for fertile ground.

So there was two way traffic, of sorts, between human geography and cultural studies, and some geographers, especially those writing on postmodernism, found their work being read and talked about across the social sciences and beyond. But note this is a 'spatial turn', rather than strictly speaking a 'geographical' turn. The flow of traffic between cultural studies and human geography was not equal in both directions. Cultural studies was (and still is) very selective in its engagements with human geography, often preferring to draw on the work of sociologists, anthropologists, and social theorists with an interest in space, such as Georg Simmel or Michel de Certeau, instead of 'mainstream' geographers. In a chapter on 'Topographies of culture: geography, power and representation', in the textbook *Introducing Cultural Studies* (which gives an unusual prominence to cultural geography, perhaps unsurprising given that one of its authors is a geographer!), only one of the three key thinkers singled out for special treatment is from a human geography background (David Harvey; the two others are Edward Said and Paul Gilroy). And a recent cultural studies textbook focused on 'cultural readings in the material and symbolic city', by Ben Highmore, rarely cites the work of geographers, even as it discusses shopping spaces or city streets. Clearly, even when space is 'sexy', geography perhaps has still to cast off its 'Anorak' status. So the interest in cultural studies around issues of space – for example, about globalization, geographical mobility, or the cultures of cities (and, of course, countrysides, too) – has too often been 'discovered' as if anew, while cultural geography continues a parallel exploration of the very same themes and ideas.

Doing Cultural Geography and Cultural Studies

The influence of cultural studies on human geography (and vice versa) extends not just to theories, but also to methods. Here the traffic may be more multidirectional, as cultural studies has moved toward a more empirically informed mode of inquiry while cultural geography has a longer lineage of methodological experimentation. Debates about methods in both subjects have been long and heated, while the developing toolkits have both been open to influence from other disciplines. Perhaps most notably, there are two major 'methodological turns' that we can detect in both subjects – a 'textual turn' and an 'ethnographic turn'. The textual turn, as noted, includes understanding all kinds of things, including places and landscapes, as cultural texts, written by someone and read (or readable) by someone. These texts are full of meanings, often about power and resistance, and these meanings are contested (or contestable) – an insight that linguistics has taught. Recently, however, there has been a turn away from seeing everything as texts, everything as meaning – so called nonrepresentational theory (NRT) is one important way in which this is being worked through. But a more prominent countertrend to the textual turn, which was seen as problematic in that it made assumptions about how these texts were read, was a renewed interest in empirical investigation, and particularly in the anthropological techniques of ethnography. While the cultural studies and 'new' cultural geography take on ethnography has been criticized by anthropologists for bastardizing an established methodological approach, there can be no doubt that this 'ethnographic turn' opened up new ways of doing cultural studies and human geography, while also helping to 'reground' some of their more theoretical flights of fancy. Huge debates about how to do this kind of research ensued, inevitably, and these connected back to the textual turn through arguments over how to represent, in academic writing, the 'messy texts' produced by ethnographic fieldwork, and how to locate the researcher herself in those accounts. A new way of researching culture was, therefore, birthed in the productive interchange between these disciplines, and in their parallel willingness to try something new, to borrow, and to bend.

Other Cultural Turns

As already noted, it is important to remember that the 'cultural turn' did not just affect cultural geography. While more and more bits of human geography came to resemble the 'new' cultural geography, either theoretically or methodologically (or both), the impact of cultural studies was more widely, if unevenly, felt. Political geography, for example, opened up to consider new forms of politics, such as cultural politics and identity politics, new ways of understanding politics, and new ways of thinking about the politics of geography itself. Rural geography experienced a noticeable cultural turn, asking questions about the construction of ideas about the rural, about representations of country life, and about the inclusions and exclusions of rurality and rural studies. Yet some areas of human geography seem to have remained more immune to the turn to culture, even if the objects of their interest have become incorporated into cultural analysis. Transport geography, for instance, has arguably yet to undergo a cultural turn, despite the current fascination, in the humanities and social sciences, with travel, movement, and mobility. Of course, this raises a key question and a key critique of the cultural turn: Where are its limits? What is beyond culture? There has been a long running debate about this point. Is nature the opposite of culture? For some thinkers, nature is revealed to be a thoroughly, inescapably cultural thing: how we understand nature, the values we attach to it, and so on. Recently then, there has been talk of 'natureculture' or, symptomatically pluralized, 'naturecultures', to reinforce this entanglement. So there is room for a cultural turn in physical and environmental geography, too. Obviously, to some doubters and critics, this is pure nonsense, and stretching culture so far makes the term meaningless – which brings us full circle, back to the issue of definition. If culture is anything and everything, all geography is cultural geography. Or is the study of anything and everything cultural studies?

There is a further problem to the all encompassing cultural turn identified by some commentators, and this concerns the uses of culture in politics. This problem has shifted in form from an attack on intellectuals around the issue of 'political correctness' to a hijacking of the language of cultural difference to argue the incompatibility of different cultural traditions and practices. Some thinkers previously comfortable with talking about culture have declared themselves 'postculturalist' in an attempt to avoid the pitfalls of politically dubious or dangerous formulations of cultural difference. Other critics warn that an overemphasis on the cultural ignores 'real' issues, such as social injustice, poverty, and violence. So the cultural turn in social geography has been cautioned not to go too far, and has been slowed by a call to 'resocialize' geography. That call was twinned with a note to 'rematerialize' geography, when the cultural turn was seen to be making everything into representations and symbols, ignoring the 'thingness' of things as things. A turn to material culture in cultural studies, and theoretical developments such as NRT, suggests that this rematerialization is well underway.

Last Thoughts

As someone who has an educational training in human geography and more than a decade of teaching experience in cultural studies, the author has been witness firsthand to the traffic between disciplines. Many of the discussions about these cultural and spatial turns seem quite abstract, based on flows of ideas from one subject to another. Clearly, at this level, there has been a lot of shuttling back and forth, though arguably the volume channeled from cultural studies toward human geography is greater than the counter flow. But much of the day to day business of thinking and doing these two subjects has a resilience, or a recalcitrance, that suggests less has changed than the 'revolutionary' accounts of the turns suggest. The nuts and bolts, bread and butter work of teaching and learning are still recognizably different from one another. The 'habitus', to use a sociological term for the ingrained, embodied, habituated ways of working, just feels different. The material artifacts of the two disciplines, the dispositions that they afford to help us think, these too are more remarkable for their differences than for their similarities. To use one last key cultural concept, it seems, to the author at least, that the 'structures of feeling' in the two academic disciplines are at times utterly unalike.

See also: Cultural Geography; Cultural Politics; Culture; Popular Culture; Postmodernism/Postmodern Geography.

Further Reading

Anderson, K., Domosh, M., Pile, S. and Thrift, N. (eds.) (2003). *Handbook of Cultural Geography*. London: Sage.

Baldwin, E., Longhurst, B., McCracken, S., Ogborn, M. and Smith, G. (1999). *Introducing Cultural Studies*. London: Prentice Hall.

Barnett, C. (1998). The cultural turn: Fashion or progress in human geography? *Antipode* 30, 379 394.

Cook, I., Crouch, D., Naylor, S. and Ryan, J. R. (eds.) (2000). *Cultural Turns/Geographical Turns*. Harlow: Prentice Hall.

Duncan, J. S. and Duncan, N. G. (2004). Culture unbound. *Environment and Planning D: Society and Space* 36, 391 403.

Highmore, B. (2005). *Cityscapes: Cultural Readings in the Material and Symbolic City*. Basingstoke: Palgrave.

Jackson, P. (1993). Berkeley and beyond: Broadening the horizons of cultural geography. *Annals of the Association of American Geographers* 83, 519 520.

Lorimer, H. (2005). Cultural geography: The busyness of being 'more than representational. *Progress in Human Geography* 29, 83 94.

McDowell, L. (1994). The transformation of cultural geography. In Gregory, D., Martin, R. & Smith, G. (eds.) *Human Geography: Society, Space and Social Science*, pp 146 173. Basingstoke: Macmillan.

Mitchell, D. (2000). *Cultural Geography: A Critical Introduction*. Oxford: Blackwell.

Philo, C. (comp.) (1991). *New Words, New Worlds: Reconceptualising Social and Cultural Geography*. Lampeter: SCGSG.

Turner, G. (2003). *British Cultural Studies: An Introduction* (3rd edn.). London: Routledge.

Relevant Websites

http://www.meccsa.org.uk/
MeCCSA, the UK's Media, Communications & Cultural Studies Association.

http://cultural.missouri.edu/
The Cultural Geography Specialty Group of the AAG.

http://scgrg.blogspot.com/
The Social & Cultural Geography Research Group of the RGS IBG.

Cultural Turn

C. Philo, University of Glasgow, Glasgow, UK

Glossary

Cultural Turn A vague and contested term, but, at root, indicating the enlargement of the purview of human geography after the late 1980s through introducing numerous new 'cultural' subject-matters and a range of new conceptual resources for doing justice to these subject-matters.

Introduction: A 'Cultural Turn'?

> … it is already obvious that there has been less a *cultural turn* than a series of cultural insights, turns, multiple circuits. Geography has become, perhaps, embodied by cultural discourse. (Cook *et al.*, 2000: xi)

A thumbnail definition of the 'cultural turn' within contemporary human geography would stress how the focus of disciplinary attention has been considerably enlarged, since around the late 1980s, by attending to a much greater array of entities, processes, and relations that arguably warrant the designation of 'cultural' than had occurred previously. This definition might also explain that bringing such cultural matter into the heart of disciplinary inquiry has demanded shifts, some subtle and others far reaching, in our underlying epistemological and ontological commitments regarding 'what' exists in the world inhabited by human (and other) beings and 'how' knowledge can be generated both about this world and the ongoing 'building' of it. Such a definition quickly becomes formidably abstract, in part, precisely because geographers have been asking the label of cultural turn to cover so much and in effect to do so much referential, conceptual, and historiographic work. Nonetheless, these twin claims, about subject matters and certain theoretical commitments, can signpost key ways into appreciating what the cultural turn in human geography might be said to have entailed. In what follows, an attempt will be made to elaborate these twin claims, and in so doing to split them into a threefold characterization of the cultural turn prompted by recognizing two distinguishable, if not wholly distinct, moves occurring in the more theoretical registers of the so called turn. The article will then proceed with a more grounded nod to the history of the 'cultural turn' in British human geography since the late 1980s.

To commence, though, it is worth amplifying the somewhat cautious tone of the preceding paragraph. To be sure, the notion or attribution of a cultural turn within contemporary human geography is remarkably commonplace, and can be found in countless books and articles, as well as peppering numerous other articles in this encyclopedia. The impression conveyed is hence of something quite certain, a clearly known, understood, and delimited (and delimit able) approach to human geography that has emerged in recent years, about which there is a measure of consensus and around which there is a readily fathomable set of debates. Yet, it can be convincingly responded that there is misplaced concreteness in the term, that writers have rather different things in mind when deploying it, and that any sustained examination tends toward its deconstruction – its decomposition into a number of only tangentially related features – and to accepting that it is really little more than a convenient shorthand for developments whose complexity, heterogeneity, and frequent contradictoriness outrun what the stability of the term promises. If anything, therefore, the message of this article is that the notion or attribution of the term cultural turn *can* be made, that there are good reasons for doing so, which can be laid out as below, but that it really does need, however irritating stylistically, to remain in scare quotes or, to use a deconstructionist mode, to be constantly placed under erasure as, its name allowed to remain in the discussion but always seen as problematic, unstable, claiming more than it can or should.

Three Senses of the Cultural Turn in Human Geography

Putting aside what might be said about a cultural turn coursing through other realms of the academy, as well as more specifically in a discipline like anthropology, it is possible to discern three senses of the term which 'do' display some limited coherence, and can be advanced with some measure of justification, when discussing recent developments within human geography. These three senses can be examined in turn.

Cultural Turn I: Finding New Subject-Matters

The narrowest sense of the cultural turn refers to a progressive opening up of academic human geography, from at least the early 1980s, to a whole range of what might be regarded as 'new subject matters' as studied by other disciplinary fields. There is a long history within academic geography of speaking about culture, but the key claim is that what was usually meant by culture tended to be restricted to quite a narrow and – the significance of which will become more apparent further on – resolutely tangible realm of human achievement. Culture talk in the discipline goes back to at least the nineteenth century, when it was coincident with the emergence of geography as an institutionalized academic discipline, particularly in how founding figures of French and German geography such as Paul Vidal de la Blache and Freidrich Ratzel sought to carve out a role for a distinctive *human* or *anthropo* geography. Of special relevance here, though, are North American traditions arising from the 1920s that tackled what were explicitly badged as cultural geography, cultural landscapes, and cultural ecology. Culture here was commonly treated as the visible or 'material' manifestations of human activity in the physical environment, occurring as people rework the Earth's surface through diverse forms of agriculture or societies pockmark this surface with all manners of productions such as buildings, settlements, roads, industries, and the like. In the hands of influential scholars such as Carl Sauer and Fred B. Kniffen, the focus was how the 'physical landscape' became overprinted by a 'cultural landscape' – one might say how physical geography was topped off by a human geography. Attention was paid to the practices of cultivating, chopping, building, hauling, and so on, as thoroughly embodied engagements between humans and nature, but what might be deemed the more invisible or immaterial aspects of culture – meaning people's perceptions and understandings or a given population's worldview, mythology or religion, as linked to customs, rituals, and the like – remained largely obscure.

With some honorable exceptions, geographers prior to the early 1980s, whether environmental determinists, regional geographers, or spatial scientists, paid virtually no attention to how culture has long been understood in the broad church of the arts and humanities. Culture in the latter, on one level, has meant the storehouses of human meaning making, through intellectual, spiritual, and indeed all manner of spoken, written, or otherwise performed endeavors enabling people to create, share, and spread beliefs, interpretations, and evaluations – in sum, one might say their 'truth claims' – of many diverse kinds: scientific, moral, esthetic, whatever. Some arts and humanities studies might adopt a 'global' perspective, addressing the overall culture (and cultural evolution) of human society in the round, but other accounts break culture as meaning making down into smaller components indexed to particular instances or types of society, different strata, groupings or organizations within particular societies, or, relatedly, different regions, places, and spaces of social life. Culture here, on another level, has meant a dizzying diversity of artifacts and performances of various kinds – artworks, sculpture, literature, poetry, music, theater, dance, and much more – that in one way or another, and in different ways with different effects for different societies, times, and places, express and engage the emotional and intellectual capacities of human beings. Similarly, geographers prior to the early 1980s also largely failed to engage with culture as increasingly understood in the wider social sciences, maybe as the comprehensions, codes, and censures establishing the norms of everyday social systems and lifeworlds, or maybe as the many accoutrements of living in modern society, meaning the items of consumption, products of all kinds, with which peoples of the post World War II West have increasingly been encouraged to fill their existences.

By the early 1980s, however, there was a growing interest on the part of many geographers in studying precisely these many 'other' subject matters clustering under the rubric of what for other academics, as well as more popularly, gets taken as culture. On the one hand, this meant opening their studies to the many products of culture, both stereotypically 'high' and 'low' culture, seeking to establish how such products – and to an extent the underlying assumptions and dynamics of which the paintings, records, clothing, and the like were surface expressions – hold a variety of geographies, from place based differences in their character, to their spatial diffusion and circulation, to their impacts upon environments inhabited and landscapes witnessed. On the other hand, it meant looking again at, in the Kariels' terms, the more 'nonmaterial elements of culture', thereby commencing the task of unpacking the inner workings of cultural phenomena such as dwelling, religious belief, conversation, learning, and so on, and borrowing, where necessary, from those traditions of inquiry within the arts and humanities or wider social sciences outlined in the previous paragraph. Thus, a vast new span of subject matters, a smorgasboard of objects that might be deemed cultural, crept into the human geographer's purview. Whereas geography journals might previously have been full of articles talking about field systems, barns and fences, about houses, roads, and settlement layouts, or about mines, steelyards, and seaports, now they gradually started to include articles about artworks and artists (say, J. M. W. Turner), about literature and writers (say, John Steinbeck), poetry and poets (say, John Ruskin), about music and musicians (say, John Lennon), or about cars, chapels, cuisine, cradles, or cable TV.

New cultural geographers such as Denis Cosgrove and Stephen Daniels began to examine how certain cultural products, particularly but not exclusively more 'elite' ones, came both to represent the spaces and places of the world, comprising so called 'symbolic landscapes', and to embody often unequal power relations forged in the crucibles of class, status, and 'national' identity, the axes of so called 'social formations'. Such cultural products might also extend to the manipulating, the manicuring even, of built and 'natural' landscapes, from the buildings of central business districts to the grounds of country houses, and in this respect mention can be made of ongoing work by James and Nancy Duncan on the admixing of esthetics (‘tastes’) and social standing in the shaping of ‘exclusive’ suburban landscapes. Another strain of ‘new’ cultural geography has led the likes of Peter Jackson and Linda McDowell to consider how other kinds of cultural products, particularly the more popular ones entrained in what people like to consume, in every sense of that word, are intimately bound up in their senses of personal and collective identity, their ‘place’ literally and figuratively in the world. Whereas Cosgrove *et al.* have tended to draw inspiration from the arts and humanities, notably the methodologies of iconography or critical image interpretation, Jackson *et al.* have looked instead to sociology, or more precisely to versions of cultural and media studies with their loosely Marxian, or at least cultural materialist, approaches to critical social analysis. Whereas Cosgrove *et al.* have tended to think about broader epochs and trajectories of cultural life, the rise and fall of different ‘ways of seeing’ keyed into processes of ‘civilization’, modernization, industrialization, and nation state building, Jackson *et al.* have tended to emphasize the more grounded cultural politics of identity, the power relations of domination and resistance played out between hegemonic forces and ‘subcultural’ groupings wherein struggles over meaning – over who labels who, when, where, for what reasons, with what consequences – come to the fore. In highly condensed fashion, then, these remarks serve to depict the new cultural subject matters ushered into contemporary human geography, ones amounting to the first sense of a disciplinary cultural turn.

Cultural Turn II: Discovering the Immaterial

A second, more expansive sense of the cultural turn lies in how many human geographers writing from the late 1980s, but with earlier precedents, have started to centralize what might be termed ‘immaterial’ facets of human being and social life. In part, this claim has already been substantiated above, when indicating how the more interior, invisible, and immaterial aspects of culture moved centerstage in what a great many human geographers elected to study after the early 1980s. That said, there is arguably a wider set of claims to be registered about this shift from the material to the immaterial – insofar as these simple designations are adequate for the purposes of this discussion – since it was not merely a question of replacing (or supplementing) one set of subject matters (or ‘objects’) with another, but also a matter of adopting a range of new conceptual lenses, complete with fresh epistemological and ontological provocations, capable of appropriately addressing the many different mutations of immaterial psychosocial human worlds now brought into disciplinary view. For the author of the present article, it is here, if anywhere, that the real kernel or import of the cultural turn may be said to lie.

As already indicated, what passed for academic geography until the mid twentieth century – progressing in the veins of environmental determinism or regional geography – tended to close its optic around the exterior, visible, and material components of the cultural landscape. This was not entirely the case, and a small sample of exceptions would be North American geographers like J. K. Wright addressing the geographical ideas (or ‘geosophies’) of everyday peoples, British geographers like H.J. Fleure considering the ‘folklores’ of rural peasantries, or French geographers like Jean Bruhnes identifying ‘the psychological factor’ intervening between physical environments and human activity. Even so, it was, by and large, to anthropologists and sociologists that such more intangible dimensions of culture were left during the first half of the twentieth century. Moreover, in the fires of Anglophone geography’s pervasive experiment with spatial science during the late 1950s through into the early 1970s, questions of culture rarely featured except insofar as some spatial scientists hunted the ‘laws’ of spatial organization supposedly governing the distribution of tangible cultural phenomena. As one instance, Herbert Kariel and Patricia Kariel’s 1972 textbook *Explorations in Social Geography* explicitly offered a ‘spatial analysis’ of many phenomena, such as architecture and housing, religion, language, and education that might be couched as cultural. Intriguingly, these authors speculated that “cultural geography would deal with … the spatial aspects of both material and non material elements of culture,” (Kariel and Kariel, 1972: v) but in the chapters that followed – as demanded by the need for ‘things’ that could be readily observed, counted, and mapped – the material(izable) elements repeatedly evicted, or stood in for, the immaterial. Distribution maps of houses duly won out over the meaning of dwelling; churches and congregations over the convictions of faith; language and dialect speakers over the cadences of speaking; schools, teachers, and curricula over the philosophies of pedagogy.

To offer a grand if not entirely indefensible generalization, the twin routes ‘away from’ spatial science post circa 1970 signposted by, respectively, radical/Marxist geography (with its focus on social structure) and behavioral/humanistic geography (with its focus on human agency) ‘both’ offered startling new moves from an obsession with the material into the complications of the immaterial. Some might object that the first mentioned route, into radical/Marxist geography, was avowedly ‘materialist’ and as such did not entail the brush with the immaterial just claimed. Given that we are here referencing a Marxist ‘historical materialism’, subsequently reframed as a ‘historical–geographical materialism’ by many authors, and an orientation proclaiming its analytical foundations in the elemental operations of humans on nature, then such an objection is justified. It might nonetheless be replied that the whole architecture of Marxist thought – conducting theoretical work to unpack the hidden structures of social relations as bound up with control over the means of production – is all about bringing into the consciousness of the academic (and maybe of the downtrodden too) the reality of structures that are themselves immaterial (in the sense of not being immediately available to human sensory apprehension) but which have dramatic material effects in the well-being or otherwise of everyday people. For behavioral and (even more obviously) humanistic geographers, the encounter with the immaterial was of course much more obvious, since these scholars readily began detailed investigations into the shadowy realms of human perception, cognition, interpretation, emotion, meanings, values, and so on, creating a rich vein of inquiry for which terms such as 'mental maps' and 'senses of place' stood as helpful, if inadequate, shorthands. The details as such do not matter, however, for the crucial thing is that the seeming fear of the immaterial was breached: in all sort of ways, from a multiplicity of perspectives, the immaterial began to be released into the studies of human geographers.

The next claim is that these maneuvers paved the way for the cultural turn *per se*. There is still a danger of oversimplifying, but it seems clear that these wider horizons of human geography from the late 1980s have included a bewildering variety of developments extending the Marxist and humanistic revolutions of the previous decade, more or less dramatically recasting their coordinates, even entirely renaming them, in moments of deconstructive hostility, reconstructive empathy, and all points in between. The many returns to Marxist geography – the spatializing of Marxist concepts; the restructuring, spatial divisions of labor, and locality studies debates; the advent of regulation and regime theories; the critical realist encounter; the agitations around post-Marxism – have heightened a receptiveness to the possible impacts of immaterial cultural processes in material political–economic spaces (querying the very discriminations between immaterial and material in the process). While many commentators may not see the cultural turn as having anything to do with Marxist geography, and indeed only see here an implacable opposition, much of what has transpired within the cultural turn does, and arguably should, have reference points back in ongoing dialogs between Marxism and geography. More self-evidently perhaps, the parallel recalibrations of humanistic geography – the reappraisal of the landscape tradition in addressing how 'we' see, inhabit, and dwell within environing nature; the attention to inter subjective meaning systems spun in semiotic webs with sociospatial fix points; the alertness to discursive formations cementing together diverse texts, archives, and effects of power/knowledge; the fragmenting of the singular, highly problematic figure of 'man' in recognizing the sheer diversity of peoples occupying this planet; the cartography of the human subject as it discloses a fragile networked topology of subject positions and related identity politics; the openness to the psychodynamics and psychotherapeutics, the emotions, passions, joys, dreads, and consolations, of gender, race, disability, and sexuality – have fostered a sensitivity to the many and varied dimensions of immaterial culture that inhere in the making of all human spaces and places (as themselves always intense condensations of the immaterial and the material). This recasting of humanistic geography, albeit rarely named as such, fueled by inputs from feminism, postcolonial theory, psychoanalysis, social psychology, cultural anthropology, and many other fellow traveling theoretical edifices, comprises the most obvious manifestation of the cultural turn within and across the discipline.

There is evidently a close relationship between the developments sketched out in the above paragraph and the interest in new subject matters introduced in the previous subsection, and in many ways it is impossible, misleading even, to hold them apart. 'All' of this stuff, all of these 'cultural insights, turns, multiple circuits', to echo Ian Cook *et al.*, must nonetheless be seen as stirred together to produce *the* cultural turn, even as we deny the singularity, stability, and containment implied by the definite article here, and certainly to furnish the discipline's 'embodi(ment) by cultural discourse', as these authors also put it. One only needs to consult the magisterial edited collection, Kay Anderson *et al.*'s 2003 *Handbook of Cultural Geography*, to gain a feel for how the *entirety* of human geography is now, or potentially could be, infused by the kaleidoscope of subject matters and theoretical innovations introduced thus far in this article. The editors make clear that, for them, the *Handbook* is precisely *not* a report back from a particular subdisciplinary field, that it is not really about cultural geography *per se* at all, but rather a statement for a much more broadly and deeply 'enculturated' human geography:

> … if there is one thing about cultural geography that we know for sure, it is that it is not a field! As we debated this 'border' problem, it became clear to us that the field of cultural geography was better marked both by its disruption and by its insatiable enthusiasm for engaging new issues and ideas – whatever their source. (Anderson *et al.*, 2003: xiii)

Framing cultural geography as 'a style of thought' thereby shifts it from being a subdiscipline – the editors insert a picture of a grave entitled 'Here lies buried cultural geography … died 2002. In loving memory' – to being more the full span of provocations inserted by the cultural turn (and its descendants) 'throughout' human geography (and seeping into cognate disciplines):

> Cultural geography as a style of thought is not, to be sure, a singular worldview (not one way of thinking about the world' not a fixed question and answer session, as if in the latest game show), but a place from which to ask valid and urgent questions of the world; one in which the geographical is seen as constitutive of how the world is 'made up'. … The cultural has modified the geo graphical, making it possible to study more and more 'things', but also to bring more and more 'things' under critical scrutiny. … It is therefore no accident that this book has sections that seem to belong to other books on the economy, maybe, or on the social (and we even begin the book with these). (Anderson *et al.*, 2003: xix)

Should the 'cultural turn' in human geography need a manifesto, then maybe this is it! Although there is now a backlash against the rise of the immaterial in human geography, as will be noted shortly, it is readily apparent that the Anderson *et al.* volume – and countless other texts in contemporary human geography – simply could not be what they are, cover the ground, nor operate with the theories that they do, without the sustained engage ment with the immaterial in all of its multiple, contra dictory, compelling, but frustrating guises as mooted here.

Cultural Turn III: Conflations with the 'Posts'

A third, most extended but arguably most problematic, even inappropriate, sense of the cultural turn resides in how it has occasionally been associated or even conflated with various meta level theoretical – one might say less social–theoretical and ultimately more philosophical – stances recently taken in human geography, most notably the 'posts' as in postmodernism and post structuralism. To clarify, what has happened in some quarters is that the cultural turn becomes the umbrella term applied to a basket of theories whose invocation is less to do with any particular connection to the substance of cultural subject matters – that is, what is at issue is not their status as cultural theories or theories of the cultural – and more to do with a prevailing sense of their newness, strangeness, and challenge within the higher conceptual orbits of the discipline.

One logical step behind such thinking is that post modernism is often first identified as a 'movement' within certain cultural domains – specifically architecture, arts, literature, and music, but also spiraling out into the planning and management of public spaces and even whole urban districts. In David Harvey's famous 1989 text *The Condition of Postmodernity*, subtitled 'An enquiry into the origins of cultural change', the charge hence becomes how to bend historical–geographical material ism to explain the emergence of postmodernism as, in Frederick Jameson's phrase, 'the cultural logic of late capitalism'. Much confusion has then attended the rela tionship between this inquiry *into* postmodernism as a landscape of cultural objects and the adoption of a postmodernist perspective, one skeptical of 'meta narra tives' and seeking more humble, nontotalizing forms of 'minor' theoretical engagement with fractured local worlds, by some human geographers from the late 1980s. The latter formulation of postmodern geographies, awkwardly comprising a meta theoretical appeal pre cisely to dismantle 'grand theory', chopping it down to a more 'human scale', is not inherently *about* culture, and yet a too easy conflation ends up whereby postmodern ism as a new strain of theorizing is then installed as part of – even the leading edge in – the discipline's cultural turn. Such a conflation occurs in Don Mitchell's other wise superb 2000 *Cultural Geography* textbook, while Ron Johnston and James Sidaway, in their 2004 *Geography and Geographers* (6th edn.) textbook, preface their chapter 7 on 'postmodern geographies' with comments about the cultural turn. In fact, the latter authors are well aware of the conflations being made here when they write:

> … the cultural turn runs through the themes that follow but these are not contained by it … . So this chapter [their chapter 7] will be somewhat more disjointed than its predecessors as we work through a series of interwoven, postmodern, poststructuralist and postcolonial strands of contemporary human geography. (Johnston and Sidaway, 2004: 268)

It would surely be preferable *not* to run together these theoretical stances, but rather to inspect them carefully, critically, but sympathetically on their own (really quite different and distinct) terms, and even more importantly they should not be homogenized as somehow consti tutive, indicative, or derivative of a more overarching cultural turn. To do so risks losing conceptual precision, and thereby evacuating the cultural turn of whatever focus, specificity, and coherence *can* be detected in the first two senses of the term outlined above.

A qualification is perhaps in order, however, in that there are a few contributions to the literature that *do* reference the cultural turn as a meta theoretical move ment, but in a manner that retains rigor and is productive of insights. A prime example in this respect is Trevor Barnes' 2001 paper on how economic geography has changed from the quantitative revolution, or when it was prosecuted as a spatial science, through to the cultural turn, the latter being understood by Barnes as carrying the 'hermeneutic theorizing' espoused by the likes of Richard Rorty (pragmatist philosopher) and Donna Haraway (feminist science studies writer). More broadly, Barnes is here underlining the importance of what might be deemed post structuralist preferences for styles (and politics) of theory that are nonreductive, anti foundationalist, always alert to the social locations of knowledge production, and operating with quite different 'topologies' of space to those favored by the spatial sci entists. In so doing, he convincingly demonstrates how economic geography, and in effect much of human geography, has now been reconfigured through these novel theoretical sensibilities; and, insofar as the term cultural turn is deployed in the process, it is done so consistently, precisely, and sympathetically. What must also be acknowledged, lest there is any confusion, is that many of the new conceptual resources discussed in the previous subsection, all of those enlarged Marxist and humanistic–geographical positions opening up to the multiple immaterialities of human social worlds, *can* be cast as attuned to or informed by post structuralism, even if the connections are not always made explicit. It might be noted that the same could be claimed about an alle giance to postmodernism as a 'stance' before the world, but, for whatever exact reasons, the latter has largely disappeared from the most recent argots of geographical theorists. The basic point, nonetheless, is that, if it *is* thought useful to trace (or parallel) the lineages of the diverse culturally inflected considerations from the pre vious subsection back into (or with) post structuralism – by implication what occurs in Barnes' paper – then such a task is not without intellectual merit.

That said, such a task is not essential. The new con ceptual resources at issue are hardly exhausted, nor their import secured, or their shortcomings exposed by being dubbed post structural, while the more ontological and epistemological mutations of the post structuralist cor pus potentially take us in directions scantly reflected in said conceptual resources. This is why a separation is being envisaged here in two somewhat different senses of cultural turn, the first referring to theorizing that is grounded, inflected toward the world, while the second is ultimately more abstracted, more philosophical. More over, to pull out further the reservations implied so far about the third sense of the cultural turn, it might be objected that its deployment in this guise tends to be effected by commentators 'external' to it (whatever exactly 'it' might be) and by critics – 'scientific, Marxist, humanist, whatever – who move swiftly into negative, detractive remarks. To be harsh, therefore, the use of the term cultural turn in this guise, not arising all that often in print but all too regularly in the more informal settings of the seminar or common room, reflects an intellectual sloppiness: a readiness to lump together what are actually diverse meta theories with differing ontological, epi stemological, and methodological credentials, lending them that spurious coherence indicated at the outset of this article. This is a tactic of enemies, it might be added, utilized as a prelude to the making of critical, destructive, and often unhelpful comments smearing accusations of various hues – often collapsing into attacks on supposed formless relativism, intellectual fecklessness, and associ ated ethico political irresponsibility – across whatever gets caught in the crosshairs. There *are* telling critiques to be raised against elements of what cluster under the heading of cultural turn, to be sure, but less satisfying, to put it mildly, is the tactic of bundling together many things about which one might feel some nagging unease, albeit for reasons that are analytically unclear, and then dismissing them all at once through the device of tarring them with one brush (one overall side sweep at *the* (chronically underspecified) cultural turn in human geography).

The Cultural Turn in British Human Geography

To put a little historiographic flesh on the bones of the article as composed so far, it is proposed very briefly to review two moments in the history of the cultural turn in British human geography. In practice, the cultural turn, as in the various senses and elements discussed above, can be found coursing throughout Anglophone human geography and indeed beyond (e.g., the collection edited by José Sarmento *et al.*), but it is matter of record that the most explicit agitations and reflections in this regard have emerged from the British academy.

New Words, New Worlds

Toward the close of the 1980s, and tied up with the decision to rename the 'Social Geography Study Group' (of the Institute of British Geographers) as the 'Social and *Cultural* Geography Study Group' (SCGSG), an 'initiative' was launched that sought to reflect upon what might be at stake in 'reconceptualizing' human geog raphy through new 'cultural' – and, by implication too, 'social' – lenses. The initiative hinged around the com mittee of the SCGRG writing quite a lengthy discussion document, entitled 'De limiting human geography: New social and cultural perspectives', buttressed by shorter 'position papers' prepared by pairs of committee mem bers, as a collective provocation to a larger constituency of British human geographers who attended and pre sented at a conference held in the University of Edin burgh (in September 1991). These written pieces, together with versions of all of the presentations given at the conference, and a few additional contributions and reflections, were compiled by Chris Philo into the desktop published volume *New Words*, *New Worlds* – the title for the initiative, as suggested by Jacquelin Burgess – that appeared in November 1991. In a preface to the volume, Philo talked circumspectly about a cultural turn, and maybe here is to be found the first explicit framing of matters in such a vocabulary. He borrowed from Denis Cosgrove's earlier comments on the renaming of the study group:

> The change in title … is an entirely welcome event for someone like myself who has always believed that human geography should celebrate the cultural diversity of our world and pay attention to the ways in which human beliefs, values and ideals continuously shape its land scapes. It is a change which signals a profound and, in some respects, an overdue change in geographical phil osophy and methodology. (Cosgrove, 1988: 3)

As Cosgrove had also declared, this change inserted "a sense of history and of human agency as the collective expressions of real people with imagination, ideas, de sires, beliefs and values – all those characteristics that are captured within that powerful but thankfully elusive term, 'culture'." (Cosgrov, 1988:3) Above all, for Philo, the call was now to situate the social realm "amongst the complex webs of cultural signification through which people understand themselves, their shared identities (their moralities and politics) and their geographies (their positionings in space and attachments to place)." (Philo, 1991:11)

These statements hardly do justice to the richness of the possibilities that were tumbling out of, and around, the SCGSG initiative. On one level, the SCGSG was clear that at least four distinctive fields of inquiry sug gested themselves – 'processes of social differentiation', 'constructions and boundaries of the self', 'globality and territoriality', and 'society, culture, and the natural en vironment' – and, in consequence, contributors ranged across such a vast array of social, cultural, economic, political, and environmental literatures that any sim plistic rendition of *the* supposed conceptual 'core' to what was being advocated would be misleading. Marxist and humanistic angles jostled alongside feminist, psycho analytic, and (from many avenues) social and cultural constructionist arguments, and those rather differently orientated representatives of the so called 'new cultural geography' as discussed earlier – Denis Cosgrove (as well as Stephen Daniels) and Peter Jackson (as well as Susan Smith) – all made key contributions (revealing in this context at least some commonality of purpose). Christian geography, the ideas of Norbert Elias on self and mod ernity, the post structuralist account of consumer society given by Jean Baudrillard, the postcolonial (imaginative) geographies of Edward Said, and even an early caution by Nigel Thrift about the dangers of attending to 'over wordy worlds' (see below): all of these ingredients of the *New Words*, *New Worlds* volume pointed toward different volleys of theory and research that have since been fired more loudly and lengthily across the bows of human geography (in Britain and beyond). What might be added is that academics partaking in the initiative and volume would likely, then and now, position themselves variously in relation to the subdisciplinary landscape of human geography – that is, as cultural, social, economic, polit ical, historical, urban, or rural geographers – and, in all kinds of ways, with alternating effects and outcomes, both positive and on occasion negative, there has, since 1991, been a proposed, and in some cases fully attained, 'enculturing' of the subdisciplines.

Cultural Turns/Geographical Turns

Keeping tabs on these many developments from around 1991 onward, making sense of them, and ascertaining gaps to be filled and, more importantly, critiques to be met, has been a formidable undertaking. A conference at the University of Oxford in September 1997 organized by the SCGRG (now renamed a 'research group', hence SCGRG) brought together many of the participants from the 1991 meeting with a new generation of geographers who were growing up in the orbit of the so called cul tural turn, even coming to see it as itself an orthodoxy to be challenged. An output from this event was another volume, this time a commercial publication, edited by Ian Cook, David Crouch, Simon Naylor, and James Ryan, and appearing in 2000 under the instructive main title of *Cultural Turns/Geographical Turns*: instructive in part because the pluralizing of 'turns' spoke to the feeling that there was too much diversity to be adequately subsumed by both the definite article and singularity of the term, as in *the* cultural turn, but also instructive because a message of this volume was the extent to which a parallel, com plementary alertness to geography, a 'geographical turn', or what Don Mitchell and others have called a 'spatial turn', was then arising in the wider academic landscape of cultural studies and social science. As the editors explained:

> This book stretches the idea of 'the cultural' in geog raphy to embrace large areas across the discipline. Moreover, it breaks the boundaries of geography itself and overlaps with work in anthropology, cultural studies, sociology and the humanities. Yet it holds all of these aspects in a relationship with geography and with what geography does and what geographers contribute. The book is divided into five parts and the part titles dem onstrate something of this range: 'Popular culture and cultural texts', 'Culture and political economy', 'Nature and society', 'Spaces and subjectivities'. Within these parts lie further categories concerning, for instance, na tions and nationalism, cities and urban spaces, regions, ethnicity, sexuality and gender … . (Cook *et al.*, 2000: xi xii)

It is easy to discern continuities with the *New Words*, *New Worlds* volume, not least in terms of the dazzling spectrum of new subject matters coming into the geographer's purview, but also in terms of the varying conceptual engagements with a host of differing material *and*, crucially, immaterial phenomena, forces, and affects. It should not be difficult to map across from a reading of *Cultural Turns/Geographical Turns* to the first two senses of cultural turn outlined above.

What permeates the *Cultural Turns/Geographical Turns* collection, though, is arguably a hesitation about what had been achieved, a loss of confidence, possibly even of excitement, as compared to the heady days of the *New Words, New Worlds* initiative (which, it should be underlined, is in no way a criticism of the editors). The problem, insofar as it can be conceived as such, is that it is possible to hear a number of criticisms, more or less incipient or articulated, running through many of the pages, particularly in worrying about the cultural turn having been *too* successful – having too quickly insinuated itself into every pore of human geographical inquiry – with the upshot of pushing to the margins (what might be construed as) *non* cultural facets of human social life (most obviously the economic but also some lineaments of the social and also those features of the political *not* immediately cast as 'cultural politics'). Such had never been the intention of the *New Words, New Worlds* participants, to be sure, for whom the intellectual 'imperialism' of imposing a cultural turn on the whole discipline had never been the ambition; but it could be agreed that an unintended consequence was for the hare set running in 1991 to dash around too widely and erratically. Some of the criticisms were undoubtedly unfair, melding with that line of unthinking denunciation of cultural turn as post structuralism identified above, but there was (and continues to be) salience in the complaint that the cultural turn, in many of its faces, has voyaged too far away from a concern with the material.

One line of critique has objected to the cultural turn's seeming abandonment of material social relations, meaning the ways in which human beings enter into complex, often fractious and unequal, social (power) relations between one another in the course of securing the material conditions of their own reproduction (at the bottom, the base realities of eating and drinking, being clothed and sheltered, and kept sufficiently physically and mentally healthy to continue living and functioning). Such a critique of course owes much to the historical (occasionally reworked as cultural) materialism of scholars inspired by Karl Marx, and Chris Philo's essay in the *Cultural Turns/Geographical Turns* collection – itself an echo of claims advanced by geographers such as Don Mitchell and Neil Smith writing from North America – debated whether the time had come to retrieve such materialisms *within* (not necessarily against) the discipline's cultural turn. A second line of critique has objected to the cultural turn's apparent loss of interest in the material phenomena of the world, the many 'things' of varying sizes and compositions with which humans can not but dwell in the world – humanity's *own* physical embodiment included – and which are always implicated in 'our' (human) ongoing being in and acting toward the (environing) world. Philo hinted at paying heed to the 'bump into able' things of the world, although in the process, and in conjunction with his appeal to historical materialism, he clearly conflated different theoretical understandings of the material (and materialism) as well as neglecting others.

Nigel Thrift's brief introductory piece in the Cook *et al.* collection, subtitled 'Dead or alive?', more fully anticipated, and to an extent delivered, if not so much a unitary materialist critique of the cultural turn, then undoubtedly the rudiments of a thorough going assault on it for neglecting a whole raft of 'materialities' emergent in – and constituting the 'spacings' and 'timings' of – how humans co mingle, vitally (in a lively, sensation full, generative manner) and emotionally (in an affective, passionate, or careworn manner), with all of the many things, real and virtual, promiscuously tugging at them from every side in every place (and hence Thrift's much expanded 'ecology of place'). Thrift's words here alluded to a host of (post structuralist and posthumanist) conceptual possibilities then starting to circle in human geography, and also in cognate fields, including the nonhumanism of actor network theory (ANT) and the 'world making' vitalism informing the intellectual genealogy that can be caricatured as Spinoza to Bergson to Deleuze and Guattari. He also adopted a stance shaped by what has been dubbed nonrepresentational theory (NRT), noting that "[o]n the whole, cultural geographers' use of theory remains resolutely … representational," leading the discipline's cultural turn to remain "dominated by tired constructionist themes." (Thrift, 2000:2 and end note 2, p.5) The charge was that human geographers continue to work with a model of the human being as much like themselves – as a 'mini' academic, inadvertently privileging an intellectualist, in some senses elitist (in 1991 Thrift said 'middle class') mode of being in the world that involves detachment, evaluation, and even irony – whereas so much of human being in the world is *simply* acting in that world, always already being entirely wrapped up in the ongoing conduct and performance of living, doing, sensing, emoting, responding, experimenting (because there are no or only sketchy prior 'mental maps' for how to proceed), and moving on (to the next moment (timing) and place (spacing)). Another term sometimes wielded in this critique has been 'culturalist' – objecting that 'we' tend to position our research subjects as if they are knowing cultural critics – and it is in response to the full force of these Thriftian objections that large swathes of British human geography, for good or ill, have now embraced the twin propulsions of both an 'emotional turn' and a 'practice turn', encompassing too a constant search for the new materialities of human social life, and have stepped back from (association with) the cultural turn. The extent to which this development amounts to the death of the cultural turn, however exactly delimited, or, from another vantage point, just another set of moves within its overall empire, will be clarified over the coming decade.

See also: Actor-Network Theory/Network Geographies; Cultural Geography; Cultural Materialism; Feminism/Feminist Geography; Geography, History of; Historical-Geographical Materialism; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Nature-Culture; Non-Representational Theory/Non-Representational Geographies; Postcolonialism/Postcolonial Geographies; Posthumanism/Posthumanistic Geographies; Poststructuralism/Poststructuralist Geographies; Psychoanalytic Theory/Psychoanalytic Geographies.

Further Reading

Anderson, K., Domosh, M., Pile, S. and Thrift, N. (eds.) (2003). *Handbook of Cultural Geography*. London: Sage. (esp. 'Preface' by Anderson, K. *et al*., pp xiii xix.)

Barnes, T. J. (2001). Retheorising economic geography: from the quantitative revolution to the 'cultural turn'. *Annals of the Association of American Geographers* 91, 546 565.

Barnes, T. J. and Duncan, J. S. (eds.) (1992). *Writing Worlds: Discourse, Text and Metaphor in the Representation of Landscape*. London: Routledge.

Barnett, C. (1998). The cultural worm turns: fashion or progress in human geography? *Antipode* 30, 379 394.

Cook, I., Crouch, D., Naylor, S. and Ryan, J. T. (eds.) (2000). *Cultural Turns/Geographical Turns: Perspectives on Cultural Geography*. Harlow, Essex.UK: Pearson Educational. (esp. 'Foreword' by Cook, I. *et al*., pp xi xii.)

Cosgrove, D. (1984). *Social Formation and Symbolic Landscape*. London: Croom Helm.

Cosgrove, D. (1988). The cultural in human geography. *Newsletter of the Social and Cultural Geography Study Group Spring,* 2 3.

Crang, P. (1997). Cultural turns and the (re)constition of economic geography. In Lee, R. & Wills, J. (eds.) *Geographies of Economies*, pp 3 15. London: Arnold.

Daniels, S. (1993). *Fields of Vision: Landscape Imagery and National Identity in England and the United States*. Cambridge: Polity Press.

Duncan, J. and Ley, D. (eds.) (1993). *Place/Culture/Representation*. London: Routledge.

Duncan, J. S. (1990). *The City as Text: The Politics of Landscape Interpretation in the Kandyan Kingdom*. Cambridge, UK: Cambridge University Press.

Gregson, N. (1993). 'The initiative': delimiting or deconstructing social geography. *Progress in Human Geography* 17, 525 530.

Gregson, N. (2003). Reclaiming 'the social' in social and cultural geography. In Anderson, K., Domosh, M., Pile, S. & Thrift, N. (eds.) *Handbook of Cultural Geography*, pp 43 57. London: Sage.

Harvey, D. (1989). *The Condition of Postmodernity: An Enquiry into the Origins of Cultural Change*. Oxford: Blackwell.

Jackson, P. (1989). *Maps of Meaning: An Introduction to Cultural Geography*. London: Unwin Hyman.

Jackson, P. (2000). Rematerialising social and cultural geography. *Social and Cultural Geography* 1, 9 14.

Johnston, R. J. and Sidaway, J. D. (2004). *Geography and Geographers: Anglo American Human Geography since 1945* (6th edn.). London: Arnold.

Kariel, H. G. and Kariel, P. E. (1972). *Explorations in Social Geography*. Reading, Mass., US: Addison Wesley.

McDowell, L. (1994). The transformation of cultural geography. In Gregory, D., Martin, R. & Smith, G. (eds.) *Human Geography: Society, Space and Social Science*, pp 146 173. London: Macmillan.

Mitchell, D. (1995). There's no such thing as culture: Towards a reconceptualisation of the idea of culture in geography. *Transactions of the Institute of British Geographers* 20, 102 116.

Mitchell, D. (2000). *Cultural Geography: A Critical Introduction*. London: Blackwell.

Philo, C. (2000). More words, more worlds: Reflections on the 'cultural turn' and human geography. In Cook, I., Crouch, D., Naylor, S. & Ryan, J. T. (eds.) *Cultural Turns/Geographical Turns: Perspectives on Cultural Geography*, pp 26 53. Harlow, Essex. UK: Pearson Educational.

Philo, C. (compiler) (1991). *New Words*, *New Worlds: Reconceptualising Social and Cultural Geography*. Lampeter: Social and Cultural Geography Study Group (of the Institute of British Geographers). (esp. 'Preface' by Philo, C., pp 1 13, and 'Over wordy worlds' by Thrift N., pp 144 148.)

Sarmento, J., deAzevedo, A. F. and Pimenta, J. R. (eds.) (2006). *Ensaois de Geografia Cultural*. Porto: Figueirinhas.

Smith, N. (2000). Socialising culture, radicalizing the social. *Social and Cultural Geography* 1, 25 28.

Thrift, N. (2000). Introduction: Dead or alive? In Cook, I., Crouch, D., Naylor, S. & Ryan, J. T. (eds.) *Cultural Turns/Geographical Turns: Perspectives on Cultural Geography*, pp 1 6. Harlow, Essex.UK: Pearson Educational.

Culture

K. J. Mee, University of Newcastle, Callaghan, NSW, Australia

Glossary

Hybridity Something new created by the merging of two things previously thought to be qualitatively different.
Material Culture The relationship between things and people.
Nonrepresentational Theory Concerned with the performance of everyday life. Investigates the ways in which people become subjects through practices that involve embodied interactions with objects and nonhumans.
Popular Culture Cultural practices of the majority of people.

Introduction

Culture is a key concept in contemporary human geography that is remarkably difficult to define. The way that culture is used shifts, depending on the context in which it is used and also varies between groups of human geographers. Research into issues that can be broadly described as 'cultural' has increased substantially in human geography since the end of the twentieth century. This is due in part to a resurgence of cultural geography in the discipline. It also reflects the cultural turn in geography, which has led to an increased study of culture more generally.

How human geographers conceptualize culture has been the subject of considerable debate in cultural geography. In 1995 Don Mitchell famously argued that "there's no such thing as culture." Mitchell later elaborated on this theme thus:

> There is no culture in the world, only differing arrays of power that organize society in this way, and not that. Hence there is only a powerful *idea* of culture, an idea that has developed…. under specific historical conditions and was later broadened as a means of explaining material differences, social order, and relations of power. (Mitchell, 2000: 74 75)

Despite Mitchell's concern about whether culture is a robust intellectual concept, geographers have been actively pursuing the study of culture and mobilizing the idea of culture in their work. Mitchell argues that "cultural geographers should be engaged in the task of determining not what culture is – since it is nothing – but rather how the *idea* of culture works in society" (p. 77). In the spirit of this, this article seeks to outline how the idea of culture is currently used by the human geographers. To begin with, it outlines the major ideas associated with culture that relate to work in human geography, and then outlines more specifically how human geographers use culture.

Ideas Associated with Culture in Contemporary Human Geography

The use of culture in human geography has been heavily influenced by the work of the cultural studies theorist Raymond Williams. Williams notes the complexity of culture as a concept, ranking culture as "one of the two or three most complicated words in the English language" (Williams, 1988: 87). The word 'culture' does different work in different languages. Williams traces the complexity of the meaning of culture in English to the different ways the word is used in English, French, and German, for example. This article focuses on the use of culture in contemporary English speaking human geography. The term may well be employed in a different way, or be used with a different inflection in other geographical traditions.

There are ten key ideas associated with the use of culture in contemporary English speaking human geography, as can be seen in **Table 1**. The first three of these ideas stem directly from the work of Williams, and relate to basic definitions of culture used by many human geographers. The following seven ideas of culture relate to more specific claims about culture, and the importance of culture in human geography.

The first key idea associated with culture then is that culture is associated with intellectual, spiritual, and esthetic development. In this usage, culture is strongly associated with ideas of civilization and progress. The development of culture is seen to distinguish people from animals, and different groups of people from each other. Cultural geographers have examined the work done by the idea that some people are more 'cultured' than others. For example, postcolonial geographers have examined how the idea of cultural superiority informed (and informs) colonizing practices.

Second, according to Williams, culture is connected to "a particular way of life, whether of a people, a period, a group or humanity in general" (Williams, 1988: 89). This is the most common use of culture in geography and the social sciences. This idea of culture is examined through a range of elements of cultures, including beliefs, ideas,

Table 1 Key ideas associated with culture in human geography

Ideas associated with culture	*Importance to the idea of culture*
Intellectual, spiritual, and esthetic development	Connected to the idea of civilization and progress Culture as distinguishing humans from animals Culture as distinguishing groups of humans
Way of life of a people, period, or group	Culture is shared by a group Culture is a marker of difference
Elements include beliefs, ideas, values, symbols, things, languages, rules, laws, and practices	Culture is actively practiced; beliefs, ideas, etc., are generated by deliberate and unintended actions
Works of intellectual or artistic activity, for example, music, literature, painting, sculpture, theatre, and film	These types of activities are understood as qualitatively different to other types of activity People make judgments about the value of these types of activities, and may rank them hierarchically
Culture as a process	Culture is not passively received and passed on Culture is actively experienced and shaped by individuals and groups
Culture forms through flows and relations	Culture or cultures are created through relations with other cultures. They are not fixed and unchanging; nor are they separate
Culture is used to mark boundaries, for example, culture/nature, culture/economy, culture/society	The meaning of culture shifts, depending on what binary culture is mobilized within. Culture has a much broader meaning when opposed to nature than it does when opposed to economy or society
Material and symbolic aspects of culture	Culture is associated with things (material artifacts) and symbols (which provide meaning for things and experiences)
Culture and power	There is unequal access to creating meaning; therefore, culture is linked to power
Plurality of cultures	There is no single culture; but rather multiple intersecting cultures
Culture as a powerful idea	There is no such thing as culture it is a powerful idea mobilized to serve particular ends

values, symbols, things, languages, rules, laws, and practices. Important to this notion of culture is the idea that culture is shared by a group. Culture is a marker of group difference. Moreover, geographers emphasize that culture is actively practiced. People reproduce the elements of culture through participating in activities. Culture is therefore not simply passed down and inherited, but rather deliberately created and recreated through practice.

Third, culture is associated with artistic activities, for example, music, literature, painting, sculpture, theater, and film. Investigations of culture in these areas take a number of forms. Geographers have examined the meanings found in artistic activities, for example, investigating the meanings of song lyrics, paintings, or films. Other geographers have examined the so called cultural economy, the business of producing cultural products. Geographers have made connections between perceptions of cultural activities as fulfilling and the poor working conditions found in some cultural industries. Geographers have also examined the reception of cultural activities by audiences. Moreover, geographers have considered how cultural activities are valued, arguing that some types of cultural activities are positioned at a higher status than others. All the types of culture discussed thus far come together through the notion of 'elite culture'. Thus the designation of some types of activities as more 'cultured' is related to ideas of human civilization and is linked to the participation and appreciation of particular groups. Engaging in these types of cultural activity is a marker of status through cultural capital.

Fourth, cultural geographers have argued that culture is a process. Culture is not passed unchanged between generations. Culture is actively produced by individuals and groups. This relates closely to the fifth point that cultures form through interactions. They are not fixed, but rather come about through the relations of groups of people with different ways of life. The interactions between cultures may lead to a mixing of cultures, and the formation of new ways of life. Geographers have investigated such interactions through the notion of hybridity.

Sixth, the word 'culture' is used to mark boundaries between qualitatively different things. Culture is juxtaposed variously to nature, economy, and society. This is discussed in more detail ahead in the section on how cultural geographers use the idea of culture.

Seventh, geographers have delineated material and symbolic elements of cultures. This theorization of culture has done significant work in the discipline. Considerable energy has also been expended in recent years arguing for more research into the material aspects of culture. The study of material culture in cultural geography has increased as a result. Some geographers argue that more consideration is required of the mutual constitution of material and symbolic elements of culture.

Eighth, contemporary cultural geography is focused on the links between culture and power. Cultural geographers focus on the work that culture, or aspects of culture, do to advantage some groups and disadvantage others.

The ninth point mentions, as will be clear from the previous discussion, that much work in cultural geography draws on a notion that there are many cultures, what is termed a plurality of cultures. This is reflected in how some cultural geographers use the word 'culture', mobilizing the word 'cultures' (in the plural) rather than 'culture' (in the singular) to point to the multiple and intersecting ways of life that they are investigating. The use of the singular form of 'culture' for these geographers denotes a focus on a very specific, but nonetheless interconnected culture. The use of the plural 'cultures' is more common among cultural geographers than other human geographers; however, not all cultural geographers use 'cultures' (Mitchell, for example, only uses the word 'culture'). The next section examines the tenth point, that culture is a powerful idea. It does this through an examination of how geographers use the word 'culture', what ideas of culture these uses involve, and the ends served by these uses of culture.

How Geographers Use Culture

In light of the understanding of culture in human geography presented previously, we can only answer a question about how culture is used in a historically specific fashion, through the lens of a particular entry point. This section is based on analysis of articles published in five selected human geography journals from 2002–06. All the journals chosen are English language journals, and thus the focus is largely on understandings of culture by English speaking geography academics. All of the journals are international journals in the sense that they publish articles by authors based in a number of countries, but the vast majority of articles are by and about English speaking countries. Two of the journals chosen are important specialist journals of cultural geography, *Cultural Geographies* and *Social and Cultural Geography*. Within the pages of these journals we can see traces of how cultural geographers in particular make use of the idea of culture. As culture is increasingly important in human geography, more general journals were also examined. The three journals chosen, *Progress in Human Geography*, *Environment and Planning D: Society and Space*, and *Geoforum* are all highly cited and regarded international human geography journals that publish work from a range of fields in human geography. A different picture of how culture works may have emerged from the inclusion of specialist journals from other areas of human geography. In addition, this approach glosses over differences between the ways that geographers from different traditions of cultural geography, or within different countries use the word 'culture'.

'Culture' is a word that is very frequently used in the pages of the journals examined. The analysis of this entry was based on the appearance of the words 'culture' or 'cultures'. 'Cultural', the adjective, was also used frequently in the articles but is not the focus of the discussion ahead. While there are many different ways that culture is used by human geographers, this entry focuses on the three most common ways the word in used (see **Table 2**). First, culture is used as a boundary word opposed to another word (e.g., culture nature, culture economy, culture society). Second, culture is used to point to particular types of process through the expression 'culture(s) of… '. Third, culture is used to point to the practices of a particular group through the use of '… culture(s)'. Within each of these categories there are significant variations in how culture is used.

Before turning to a discussion of the way that culture is used in human geography it is important to note the difference between the use of the singular culture and the plural cultures. The plurality of cultures is an important idea associated with culture as was noted earlier, and this can be seen in the use of the term 'cultures' rather than 'culture'. Generally speaking, those publishing in cultural geography journals, or who are identified as cultural geographers, are more likely to use 'cultures' rather than 'culture'. In **Table 2**, I have included 'culture' whenever the word use or combination was only found in the singular, 'cultures' whenever the word was only used in the plural and 'culture(s)' when the word usage was sometimes singular and sometimes plural. Given the importance of the plurality of cultures this provides insights into the idea of culture being mobilized in an article.

Culture in a Binary with Another Key Term

'Culture' is frequently used as a binary term with nature and economy, less frequently with society. In these categories, culture is much more commonly used as plural in relation to nature or natures than it is with economy or society. Geographers are more likely to mobilize the 'cultural and the social' as terms rather than 'culture society'.

So how does 'culture' work as an idea when mobilized as part of a binary? The meaning of 'culture' shifts, depending on its binary context. When used in the culture nature dualism, it tends to mean the practices of people quite broadly. When used as cultures natures, it also brings in traces of the idea of plurality of cultures, and

Table 2 Contemporary uses of 'culture' or 'cultures' in human geography[a]

How culture is used	*Idea(s) of culture associated with this usage*	*Examples of this usage of culture*
Culture in a binary with another key term	Culture as boundary	
	Culture as the practices of humans	Culture/nature
	Culture as the practices of specific groups	Cultures/natures
	Culture as a particular set of practices, often referring to artistic activities or the symbolic aspects of culture	Culture/economy
	Culture is more narrowly framed and relates to the symbolic aspects of culture	Culture/society
Culture of…	Culture as a process	
Cultures of…	The practices of particular groups	Cultures of mothering Cultures of poverty
	There are important symbolic elements of practices often associated with economy	Culture(s) of consumption
	An attempt to move away from binary conceptions of culture and economy	Culture of economy Culture of risk Cultures of alientation
	Links between culture and power, leading to conflict or contestation	Culture of resistance
…culture	Culture as a way of life associated with particular practices	
…culture(s)	Ways of life characterized by a particular time period	Modern culture, primordial cultures
	Ways of life of broadly defined groups associated with ethnicity, nationality, etc.	Aboriginal culture, African culture, Anglo Celtic culture, Celtic culture(s), Canadian culture, diasporic culture(s), English culture, European culture, Indian culture, Scottish culture, Welsh culture, white culture(s)
	More specifically associated with oppression	Racist culture
	Ways of life associated with groups defined by their religion	Mormon culture, Lutheran culture, Protestant culture
	Ways of life associated with gender	Female culture, feminist culture, lad culture, male culture, masculinist culture, women's culture, yobbish culture
	Ways of life associated with sexuality	Gay culture, lesbian culture, queer culture(s), sexual culture
	Ways of life associated with class	Poor culture, working class culture(s), welfare culture
	Ways of life associated with age	Youth culture(s)
	Ways of life associated with physical disability	Ableist culture
	Ideas, beliefs, values, and practices of particular types of work, workplaces, or industries	Audit culture, bureaucratic culture, business culture, entrepreneurial culture, workplace culture(s)
	Ideas, beliefs, values, and practices associated with other types of 'popular culture' activities	Car culture, culinary culture(s), ceremonial culture, drinking culture, food culture(s)
	Practices associated with specific types of places	Bar culture(s), childcare cultures, island culture, metropolitan culture, school culture, rural culture(s), urban culture
	Practices or ways of life associated with specific types of places at a particular scale	Body culture(s), global culture(s), home culture, local culture(s), neighbourhood culture, national culture, regional culture, street culture
	Ways of life of a specific place	Karelian culture, Highland culture
	Linked to the artistic activity idea of culture	Elite culture, high culture, mass culture, popular culture, vernacular culture
	Also linked to the culture as civilization related to an implied hierarchy of culture	
	Culture and power	Dominant culture(s), hegemonic culture, minority culture
	Elements of ways of life and their impact on the practices of geographers	Disciplinary culture(s), field culture, visual culture

[a] Derived from the use of the words culture or cultures in articles published in *Cultural Geographies, Environment and Planning D: Society and Space, Geoforum, Progress in Human Geography, and Social and Cultural Geography* between 2001 and 2006 where applicable.

their multiple interactions with a plurality of natures. When used as a binary with economy, culture implies a narrower spectrum of human activity, either connected to artistic activities or the symbolic elements of human practice. The binary of culture society delineates a more specific meaning for culture, primarily concerned with the symbolic aspects of culture.

Recent research has been targeted at examining how culture and nature are intertwined. This research picks away at the culture nature binary in an attempt to show the work done by separating people from nature. Other geographers have been keen to unpack the binary be tween culture and economy, including an examination of 'cultural economy'. The binaries of culture nature and culture economy invoke to foreground new under standings of people in places, in order to open new possibilities for understanding and intervention. Culture is clearly a powerful name in these literatures. While cautioning against the simplistic use of an idea of culture, and certainly pointing to the dangers of dualisms in the use of culture, the geographers drawing on culture re main alert of the communicative power of the idea.

Culture of, Cultures of

The second way that culture is used is in the phrases 'culture of… ' or 'cultures of'. The use of these phrases implies the idea of culture as a process, with people ac tively reproducing and reproducing shared sets of meanings. There are three distinct ways that 'culture of' is used. First, it is used to relate to the practices of particular groups. For example, in her 2000 work on cultures of mothering, Dowling examines the relationships between ideas that women hold about mothering and how tasks associated with mothering should be performed and the daily practices that result from them, and reproduce these ideas of mothering. The use of the plural (cultures of mothering) is important here; there are many overlapping cultures of mothering in a given context.

Second, 'culture of' or 'cultures of' is also found in work about cultural economy. It is particularly prevalent in work concerning cultures of consumption. A key re search focus of this work is the intertwining of the material and the symbolic elements of commodity transactions. Such work intersects with work unpacking the culture economy binary.

Third, 'culture of' is used to highlight the relationship between culture and power; highlighted by phrases such as 'cultures of alienation' and 'culture of resistance'. Examinations of a culture of resistance mobilize the idea of culture as an active process to investigate how power is negotiated and contested. In using 'culture of resistance' (as opposed to resistance culture) authors draw upon a broader theoretical tradition (the work of Edward Said, for example).

…Culture or …Cultures

By far the most common way that 'culture' is used is preceded by a descriptive word which specifies what as pect of the idea of culture is being discussed or examined. Most of the uses of the term in this way center around different examples of culture as a way of life, as can be seen in **Table 2**. They inform the reader of the con textually specific nature of the culture under investi gation. **Table 2** shows us that geographers continue to use culture to specify the ways of life of a people, a period, or a group. The many different types of cultures identified in the Table provide insights into the types of ways of life that concern human geographers at the be ginning of the twenty first century. Geographers are concerned with different ways of life associated with time periods; thus modern culture specifies the practices of people in the modern era. The term is used with little explanation, implying that what modern culture is, will be self evident to the reader. 'Culture' preceded by a descriptive adjective is more commonly used to specify a particular group or groups defined by ethnicity, nation ality (or even the continent on which they live), religion, gender, sexuality, class, age, or physical disabilities. The theoretical concepts associated with the terms used are important to the implied culture being expressed. Thus for example geographers use gay culture and lesbian culture (in the singular) but are far more likely to use queer cultures in the plural (although this term is also used in the singular). The use of queer cultures relates to the fluidity of the concept of queer, and implies plural interconnected queer cultures.

'…culture(s)' is also used as an idea in connection to particular types of work practices. Thus when geographers use the term 'audit culture', they mean a set of beliefs and ideas about the value of auditing that have become one of the practices of workplaces, particularly in the public sector. There is a difference in this use of culture com pared to the broader notion of culture used when applied to a group. This type of understanding of culture is similar to that applied to other types of activities (that might be broadly understood as popular culture activities, here re ferring to activities connected to but outside of work places), for example, car culture or food culture(s).

Unsurprisingly geographers also use culture in ways that are specific to particular types of places. Thus geographers discuss specific cultures at a range of scales. Some work aims specifically to consider the way that scale is culturally produced, and how scaled practices of culture impact upon ways of life. These types of words are used with different levels of specificity; thus geog raphers frequently use 'local cultures', but tend to use 'neighborhood culture' rather than 'neighbourhood cul tures'. The idea of a plurality of cultures is therefore unevenly reproduced in geographic writing.

Interestingly geographers rarely use a particular place name smaller than a country to describe the culture of a particular place. Thus while it is relatively common to see terms such as Canadian culture or Indian culture, and there are numerous examples of the use of terms such as Scottish culture or Welsh culture, it is relatively rare to find regional cultures specified at a smaller scale than this (although it occurs occasionally, e.g., the use of Highland culture for the specific culture of the Scottish Highlands). This might reflect the set of journals chosen for analysis here; the idea of naming a specific regional culture might be found elsewhere. It also reflects the practice of geography in the contemporary era, where identifying a specific subnational area's culture is rare.

Terms such as elite culture, high culture, mass culture, and popular culture continue to have currency in geog raphy. Often these terms are used in a broadly descriptive way, as the background to the specific aspect of culture (e.g., popular culture) under investigation. Similarly con nections between culture and power appear through the use of terms such as hegemonic culture and minority culture. These terms are often used as the background to the investigation of a more specific set of practices, to provide the context for a case study, for example.

Finally geographers use culture(s) to communicate aspects of the doing of geography. The disciplinary cultures of geography and geographers thus become something to investigate and a potential space of intervention. The ways in which geographers do the work of geography have been named through the idea of 'field culture'. In these usages geographers turn their attention to culture as actively practiced, examining the links between how ideas and practices (both past and present) shape how geographers go about doing geography and imagining their objects and subjects of analysis.

How then do geographers go about investigating culture or cultures? Because culture is used in so many ways by geographers, numerous different types of methodologies are used to examine culture. Useful sources which detail approaches to studying culture in clude the 2004 special issue of *Australian Geographer*, and the edited collections of Blunt *et al.* in 2003 and Shurmer Smith in 2002. Particularly important in the study of culture or cultures are qualitative methodologies (al though some quantitative analysis is found in work that deals with culture). Qualitative methodologies used to investigate culture include the collection and analysis of existing texts, including archival information, letters and diaries, newspaper articles, novels, tourist brochures, and postcards. Other forms of qualitative research involve collecting the stories of participants involved in a par ticular process or event, through methodologies such as interviews or focus groups, and even on occasion through questionnaires. Many geographers combine these meth odologies; for example, Allison Blunt in 2003 combines the analysis of letters and diaries with the analysis of interview texts in investigating the life stories of Anglo Indian women. Drawing on personal experiences, either through techniques of participant observation or through autobiography, is also an important means of analyzing culture. Analysis of visual texts, such as photographs and films, has also been important in recent years. In the early twentieth century, geographers were interested in ana lyzing culture by examining material artefacts and landscapes. The 1980s saw a resurgence of the study of culture in landscape. Since the 1990s there has been an increasing interest in the study of objects and their role in and relationship to culture. The rise of nonrepresenta tional theory in geography has led to attempts to analyze practices such as dance and walking to provide new in sights into the ways that culture is actively produced through embodied encounters.

Culture and cultures are clearly powerful names in geography at present. Culture(s) are used in such a wide array of ways that it is tempting to think of culture as everything and therefore nothing (to paraphrase Mitchell in 1995). However, 'culture' and 'cultures' clearly does important work in geography at present. Culture needs to be understood in a contextually specific fashion. Despite the array of meanings associated with culture, what the author means by culture in the context of a given piece of academic work is generally clear. There is a shared set of meanings (shown in **Table 1**) within geography about ideas associated with culture. Culture remains an im portant lens for understanding the world in con temporary human geography.

Postscript

This entry has focused on the idea of and use of culture in contemporary English speaking human geography. Culture clearly operates as a powerful name more broadly in the social sciences and humanities, and in societies more generally. Considering the differences between the work done by culture in human geography and the work done by culture elsewhere (and the inter actions between the two) is a pressing task.

See also: Cultural Capital; Cultural Economy; Cultural Geography; Cultural Turn; Hybridity; Landscape; Material Culture; Material, The; Nature-Culture; Popular Culture; Queer Theory/Queer Geographies; Youth/Youth Cultures.

Further Reading

Anderson, K. and Gale, F. (eds.) (1999). *Cultural Geographies* (2nd edn.). Hawthorne: Longman.

Anderson, K., Domosh, M., Pile, S. and Thrift, N. (eds.) (2003). *Handbook of Cultural Geography*. London: Sage.
Barnett, C. (2001). Culture, geography and the arts of government. *Environment and Planning D: Society and Space* 19, 7 24.
Bennett, T., Grossberg, L. and Morris, M. (eds.) (2005). *New Keywords: A Revised Vocabulary of Culture and Society*. Malden, MA: Blackwell.
Blunt, A. (2003). Home and identity: Life stories in text and in person. In Blunt, A., Gruffudd, P., May, J., Ogborn, M. & Pinder, D. (eds.) *Cultural Geography in Practice*, pp 71 90. New York: Edward Arnold.
Blunt, A., Gruffudd, P., May, J., Ogborn, M. and Pinder, D. (eds.) (2003). *Cultural Geography in Practice*. New York: Edward Arnold.
Cosgrove, D. (2000). Culture. In Johnston, R. J., Gregory, D., Pratt, G. & Watts, M. (eds.) *The Dictionary of Human Geography* (4th edn.). pp 143 145. Malden, MA: Blackwell.
Dowling, R. (2000). Cultures of mothering and car use in suburban Sydney: A preliminary investigation. *Geoforum* 31, 345 353.
Duncan, J., Johnson, N. and Schein, R. (eds.) (2004). *A Companion to Cultural Geography*. Malden, MA: Blackwell.
Gesler, W. M. and Kearns, R. A. (2002). *Culture/Place/Health*. London: Routledge.
Gibson, C. (2003). Cultures at work: Why 'culture' matters in research on the 'cultural' industries. *Social and Cultural Geography* 4, 201 216.
Gibson, C., Costello, L., Hughes, R. *et al*. (2004). Special Issue: Doing Cultural Geography. *Australian Geographer* 35.
Jackson, P. (1989). *Maps of Meaning*. London: Unwin Hyman.
Jackson, P., Cosgrove, D., Duncan, J., Duncan, N. and Mitchell, D. (1996). Exchange about Don Mitchell's 'there's no such thing as culture'. *Transactions of the Institute of British Geographers NS* 21, 572 582.
Mitchell, D. (1995). There's no such thing as culture: Towards a reconceptualization of the idea of culture in geography. *Transactions of the Institute of British Geographers NS* 20, 102 116.
Mitchell, D. (2000). *Cultural Geography: A Critical Introduction*. Malden, MA: Blackwell.
Radcliffe, S. and Laurie, N. (2006). Culture and development: Taking culture seriously in development for Andean indigenous people. *Environment and Planning D: Society and Space* 24, 231 248.
Rankin, K. (2003). Anthropologies and geographies of globalization. *Progress in Human Geography* 27, 708 734.
Rose, G. (2001). *Visual Methodologies: An Introduction to the Interpretation of Visual Materials*. London: Sage.
Shurmer Smith, P. (ed.) (2002). *Doing Cultural Geography*. London: Sage.
Williams, R. (1988). *Keywords: A Vocabulary of Culture and Society*. Glasgow: Fontana Press.

Relevant Websites

http://cgj.sagepub.com/
Cultural Geographies.
http://www.ctheory.net/home.aspx
Ctheory.
http://www.tandf.co.uk/journals/titles/14649365.asp
Social and Cultural Geography.

Culture/Natures

J. Sundberg and J. Dempsey, University of British Columbia, Vancouver, BC, Canada

Glossary

Corporeal Refers to a material, physical body or bodies.

Culture The material and symbolic productions of a given social group and/or their way of life. Culture is used to draw lines between humans and nonhumans, as well as social groups as evidenced in binaries such as culture/nature and civilized/primitive.

Environmental Determinism The belief that human activities and/or characteristics are ultimately defined and controlled by the physical environment (i.e., climatic variations produce different types of cultures and work habits).

Epistemology A theory about what constitutes valid or legitimate knowledge.

Hybrid A word used to stress the co-evolution and co-production of all things.

Ontology The study of being or the nature of knowable things.

Race Describes a historically contingent social construction that serves to establish an order of difference and hierarchy between particular social groupings as if grounded in nature.

Introduction

The last two decades have witnessed a monumental shift in how cultural geographers engage with nature. Whereas once, nature was treated as an ontologically separate domain and geographers worked to understand how societies interacted with it, what is known commonly as nature is now often understood as a social formation with multiple meanings. As Noel Castree elaborates, nature may refer to the material world of bodies, birds, and rocks, the essence of something, or the order of things. As concept and material reality, nature is enlisted to ground sociopolitical projects.

In this article, we examine how geographers have contributed to understandings of nature: how culturally specific formations of nature come into being and how they are put to work to achieve particular political ends. We begin by outlining some origins of dualistic understandings of nature and culture in Western thought and practice, including geography. We then shift to recent geographic work on nature as a historically contingent cultural formation. Specifically, we focus on research detailing how nature is enrolled in the production of interior topographies and exterior terrains, subjects, and territories.

Before we begin, it is important to note that while ontological distinctions between nature and culture are generally considered axiomatic in Western societies, this assumption is by no means universal in time or space. Indeed, as we illustrate, European colonial power/knowledge regimes have played a key role in universalizing the divide. Therefore, to write about nature/culture constructions is to presume the existence of a divide and thus constitutes a Eurocentric and Euro centered endeavor.

Origins and Legacy of the Nature/Culture Divide

Derek Gregory and others have noted that the nature/culture dualism is central to modernity, a social formation and worldview originating in sixteenth century colonial encounters and the emergence of an imperial global economy controlled by Europeans. Here we outline four features of Western thought and practice central to modernity that contribute to on going boundary making between nature and culture: patriarchy, colonialism, scientific epistemologies, and capitalist–political–economic relations. We also include the discipline of geography; as the embodiment of these power/knowledge regimes, the discipline has played a critical role in institutionalizing nature/culture dualisms.

A key commonality of all these modalities is that boundaries between nature and culture are used as a way to mark off 'difference', or the 'self' from an 'other'. While the nature/culture dualism is not the same in time and space, and is flexible and malleable depending on specific sociopolitical configurations, it is always connected to making and sustaining hierarchies, and thus is a technology of power.

Patriarchal Social Relations

Feminist environmental philosopher, Val Plumwood, argues that a fault line running through Western thought constitutes nature and culture as ontologically separate and separable realms. For Plumwood and other feminist scholars like Carolyn Merchant and Donna Haraway, the two realms – nature and culture – are not equal, but rather defined by historically contingent relations of privilege and oppression, including those that mark gender differences. For example, nature is often feminized and made passive as compared to culture, which is masculinized and active. Ancient appeals to care for Mother Earth, however, were superceded in the seventeenth century by what Merchant calls rape scripts, in which the denudation and degradation of the natural environment was justified in gendered and hetero normative terms. As such, the dualistic conception of nature/culture is produced by, and contributes to the continuation of phallocentrism and patriarchy in con temporary society.

Colonialism

In colonial encounters from the sixteenth century on, nature became a site in which Europeans delineated and secured boundaries between themselves and colonial others. Europeans distinguished themselves as civilized and modern by recalling their rise above the state of nature, while those subject to violent colonization were framed as living in nature and, consequently were placed outside of modernity and civilization. To this day, for instance, Indigenous people in parts of Latin America are referred to as *naturales*, a term from the colonial era meaning both natural and native.

At the same time that nature was degraded, Derek Gregory argues, it also became a site of normalization. Europeans treated the environments of their homelands as the norm against which they measured unfamiliar natures. Europe's temperate climate and environment were said to be responsible for the perceived rationality and industriousness of European civilization. In contrast, the centers of European colonial enterprises – the tro pical zones – were framed as regions where nature was an overwhelming force; while wondrous and bountiful, the tropics were said to be incompatible with civilization. In colonial geographical imaginaries, the temperate and tropical zones constituted realms of radical difference. For instance, Alexander Von Humboldt suggested that soil and plant fertility in the South American tropics only served to retard the development of civilization therein. Humboldt's prolific writings played an important role in inventing the tropics as a geographic field of enquiry and making it available for imaginative and material appro priation by Europeans.

Colonial science and exploration also played an im portant role in producing and mapping difference through nature. Geographer, Bruce Braun, documents how nineteenth century geologist, George Dawson, ex plored northern British Columbia and in the process mapped physical and cultural landscapes into separate domains. In doing so, Dawson detached resident In digenous groups from their land and produced the physical space as empty and awaiting European capital and labor. Braun further argues that colonial framings inform contemporary environmentalist materials, which place First Nations people in nature and outside history, with the effect of erasing their contemporary claims for land and justice. In both the historical and contemporary case, Braun shows how the discursive construction of nature as an abstract entity, separate from humans, helps erase Indigenous human–land relations, making the land available for Western colonial and national uses.

Scientific Epistemologies

Critical to the emergence of modernity, the scientific revolution of the sixteenth and seventeen centuries is a source of the nature/culture dualism. Revolutionary changes in scientific knowledge and practices, such as the turn to empiricism, emptied nature and hence the body of spiritual significance and made them available for in terrogation, measurement, and analysis by human sub jects, who were positioned in an instrumental relationship to both. The dualistic relationship forged between humans/nonhumans was codified in epistemologies built on René Descartes' vision of a mind/body split, which frames human reason as external to the body and the life world. Only knowledge produced by an unmarked ra tional gaze – envisioned as de corporealized, de con textualized, and distanced – is legitimated as objective and universal. Such framings of scientific objectivity work to sustain nature/culture dualisms.

Moreover, scientific claims to objectivity mask and protect very particular perspectives and interests in formed by cultural assumptions and social hierarchies. As Haraway illustrates in *Primate Visions*, early twentieth century primate studies claimed that female primates exist in a near constant state of sexual receptivity. This belief was used to defend authoritative constructions of heterosexual marriage and the nuclear family as the natural order of things. Any deviance from this ideal was labeled as a source of sexual and social chaos.

Science studies scholar, Bruno Latour, argues that much contemporary science turns on a process of on going purification that attempts to create nature and culture as two distinct ontological zones. Such purifi cation, he argues, obscures the world we live in, which is composed not of humans and nonhumans interacting, but rather hybrid entities impossible to separate into distinct ontological categories. Indeed, Latour proposes that epistemologies of purification contribute to the explosion of environmental problems, and close off the possibility of democracy for all worldly entities – human and otherwise.

Capitalist Ideologies of Nature

Marxist geographers, Neil Smith and David Harvey, argue that the boundaries between nature and culture are ideological formations connected to the emergence of capitalist–political economies during the seventeenth and eighteenth centuries. In *Uneven Development*, Smith argues that industrial capitalism, more than anything else, is responsible for setting contemporary (albeit changing) visions of nature. While capitalist processes increasingly remake nature to suit specific needs, ideological visions of nature as external prevail. For Smith, these ideological constructs stem from bourgeois imaginaries and serve to conceal the voracious destruction of capitalism.

Smith also argues that notions of human nature and universal laws are ideological formations. In his detailed reading of Thomas Malthus, David Harvey shows how concepts like scarcity and overpopulation do not reflect nature's inherent limits, but rather class interests. Mal thus argued that lower class people are unable to check their sexual behavior; their resulting poverty and famine act as a natural check on their population growth. For this reason, Malthus claimed, social assistance only under mines the natural order of things. At the same time, as Harvey cogently points out, Malthus suggested that upper classes continue to consume at appropriate rates to maintain a functioning economy. Together, the concepts of overpopulation and natural limits work to naturalize poverty, famine, and class inequity.

Political ecologists have built on the critique of so called natural laws to demonstrate that predominant understandings of desertification and deforestation reflect state and elite interests rather than an external reality and ultimately, work to dispossess already marginalized peoples of land and livelihood by framing them as in capable of managing soils or forests.

Nature/Culture in Geography

As an instrument of empire and capitalist expansion, the emergent discipline of geography provided ways of knowing and seeing the world that engaged with and furthered dualistic ontologies. For this reason, Gillian Rose argues in *Feminism and Geography*, geography is a masculinist discipline steeped in the hierarchical dual isms of nature/culture, body/mind, and object/subject. Thus, despite claims to be an integrative, unifying dis cipline, much geographical thought and practice con tribute to the nature/culture dualism. To illustrate how, we briefly outline important moments in the discipline.

Nature shapes culture

The environmental determinism of eighteenth century explorers, such as Humboldt, informed the work of nine teenth century geographers, like Harford Mackinder, who ascribed causal power to the natural environment. Writing in the early twentieth century, Anglo American geog rapher, Helen Churchill Semple, argued that nature – as an immutable and unchanging force – was the basis of human societies. Although always contested, deterministic arguments dominated the field for a century or more, and were a way of accounting for and hierarchically ranking cultural, political, and economic differences between people and places. Although the explanatory powers of environmental determinism have been officially dis credited, scholarly and popular discourses continue to ground arguments about racial hierarchies in nature. For example, geographical textbooks on Latin America often stress the limitations and obstacles posed by climatic conditions and geographic features. Mountains are said to retard communication and economic development, while the tropical climes handicap mental agility and ambition.

Culture shapes nature

In the early twentieth century, Carl Sauer shifted the intellectual course of geography away from deterministic claims. Arguing that cultures work on and with nature, Sauer viewed landscapes as the material manifestation of cultural processes. The task he set for geographers was to analyze the processes by which cultural landscapes came into being. Even so, his focus on how social groups transform and dominate the environment left the nature/culture divide intact. As an analytical concept, the cul tural landscape came to dominate human geography for decades. Building on Sauer's methodological foundation, scholars in cultural geography, as well as the related subfields of cultural ecology produced rich, empirical analyses of how cultures interact with and transform natural landscapes.

Studying culture, not nature

However, by the late 1960s, interest in nature/culture relations waned as the discipline became increasingly divided into human and physical geography. This splin tering, as Castree suggests in *Nature*, deepens as human geographers begin to develop and experiment with other approaches (i.e., behaviorist, humanist, Marxist, feminist), which give little consideration to the nonhuman.

De-naturalizing nature

In the 1990s, some human geographers were drawn to nature as a conceptual field of significant cultural en gagement. This shift was prompted in no small part by the work of influential feminist theorists, such as Donna Haraway, Judith Butler, and Emily Martin; environmental historians, William Cronon and Richard White; as well as sociology of science scholars, Michel Callon, Bruno Latour, and John Law. Drawing on post structural theory and anti essentialist conceptions of social categories, these scholars challenge received views of nature as a self evident realm that exists outside of and separate from social relations. Instead, they argue, nature is thoroughly social, the effect of historically contingent social and material practices.

This approach is at the forefront of Haraway's work, wherein what counts as nature and matters of fact are not self evident or neutral, but power laden. From this perspective, the culture/nature dualism is a construct in which social groups invest in order to draw boundaries between themselves and others, between in groups and out groups. For Haraway, these boundaries mark off social territories like the family, gender, race, class, citizen, and nation. Claims about nature, then, are always about ordering worlds, worlds that are often sexist, racist, homophobic, and environmentally destructive. In other words, worlds that work well for some, but not for others.

Interior Topographies and Exterior Terrains: Nature/Culture Formations

Within geography, anti essentialist approaches ushered in a challenge to the nature/culture divide, much entrenched in the discipline. As scholars braid together the work of diverse thinkers, geography has exploded with groundbreaking work on nature. In this section, we engage with recent geographical research showing how nature is enrolled to legitimize social hierarchies and cultural norms, and divide (stolen) territories. This work demonstrates that a nature/culture split is not something that emerged from the Garden of Eden or was solidified in seventeenth century laboratories, but must be and is continuously produced through many diverse mediums: academia, popular culture, media, governments, and so on. As Haraway suggests, immense resources are expended to produce this divide and police the boundaries. Here, we focus specifically on the production of subjects and territories, although the division is superficial as more often than not, these two cultural formations are overlapping and mutually constituting. Nonetheless, they are useful markers of how nature is employed to make categories and distinctions. While a wealth of such research exists in cultural geography, our examples are necessarily limited and selective due to space constraints.

Nature/Culture Subjects

Feminists have long argued that dualisms such as male/female, culture/nature, mind/body, and civilized/primitive are interrelated and mutually reinforcing. Although these dualisms call forth seemingly established and clearly demarcated ontological domains, the boundaries between them are always in the process of being redrawn. Such negotiations are nowhere more evident than in debates about who or what counts as human. The human/nonhuman dualism draws conceptual strength from the debasement of those aspects of social life associated with nature, such as bodily functions, physical necessities, and emotions. Indeed, controlling and bounding bodily flows have been a requirement of rationality, civility, and inclusion in the body politic. In what follows, we present work in geography analyzing how nature becomes a medium through which subjects are constituted.

Gendering subjects

Feminists have long argued that the mind/body split is built upon a male/female binary. Historically, women have been associated with the body and so called natural processes like menstruation or reproduction and, on these grounds excluded from embodying the human as a legal and political category. Along these lines, Robyn Longhurst's research with pregnant women found that they are associated with bodily leakages and as such, are cast outside the boundaries of public spaces.

While white masculinity is measured in terms of distance traveled 'away' from nature, Sherene Razack argues that in white settler societies like Canada, subject formation often involves journeying from civilized spaces into those constituted as liminal – as in wilderness – and back out again to civilized space. In *Mapping Men and Empire*, for example, Richard Phillips shows how the male heroes of nineteenth century Canadian adventure stories come into manhood and whiteness through encounters with the rugged Canadian wilderness and its so called wild Indians.

Naturalizing racialized subjects

The notion of bodily leakages also is constitutive of a racialized binary between civilized versus primitive bodies. Given that all bodies leak, as Longhurst points out, the civilized body comes into being through disciplined management of leakages. Geraldine Pratt's work with Filipina live in domestic workers in Canada demonstrates the extents to which the women go to ensure that they leave no bodily traces in their employer's homes. Filipina workers try to erase the smells of Filipino food, while ensuring that they do not leave fingerprints on their employer's valuables. These efforts to contain corporeal traces, to discipline their bodies are attempts to work against racialized stereotypes that keep Filipina nannies outside the space of rights.

The boundary between civilized and primitive is delineated through land management as well. In sixteenth century colonial New Spain, Andrew Sluyter argues, Europeans did not recognize indigenous modes of land management as 'cultivation' and on these grounds, cast native peoples as undeserving of rights to their lands. John Locke's theories of labor and property subsequently justified policies of dispossession in English New World colonies. For Locke, labor was key to improving land and establishing rights to property. As Kay Anderson notes in her recent book, *Race and the Crisis of Humanism*, Locke's theory established plant and animal cultivation as the only forms of labor requiring the exercise of reason, seen as a crucial capacity in Western civilizational definitions of the human as an ontological category. Seen as merely plucking fruits from trees or animals from forests, native peoples were said to be lacking in reason.

Nature/Culture Territories

Nature also is called upon to demarcate territorial boundaries, thereby constituting bounded spaces of difference. Within this section, we outline how nature/culture dualisms are enlisted to create particular under standings of the nation and demarcate belonging within nations. Following this, we examine how territories are divided into spaces designated as natural or cultural.

Nature/culture nation

Building on environmental–deterministic philosophies, the nation has tended to be conceptualized as an organic product of unique relations between people and nature. In the United States, contact with the wilderness has been linked to the production of a distinct Anglo Saxon American character, set of values, and democratic insti tutions. In lamenting the loss of a wilderness frontier in the late nineteenth century, elites in the United States were worried about the loss of sites in which rugged masculinity and whiteness had been produced. Efforts to preserve wilderness as markers of national (read white, masculinist, and bourgeois) identity resulted in the dis possession of Indigenous inhabitants and the creation of national parks, which have become key sites wherein the nation's heritage is narrated and belonging delimited.

The biophysical environment is also a site wherein national identity may be renegotiated. Allaine Cerwon ka's research in Australia suggests that many white Aus tralians are reconfiguring their relations to the colonial past and the multicultural present by literally uprooting exotic species like the willow trees and rose gardens privileged by British settlers and replanting native spe cies like eucalyptus. Through these practices, white Australians differentiate themselves from British colon ists of the past and reposition themselves as environ mental stewards with rights to national territory.

Nature/culture natural spaces

Western nations have a long history of marking natural spaces from cultural ones, but the meanings assigned to those spaces are incredibly flexible. Here we examine the demarcation of resource territories from conservation ter ritories; the former spaces are abstracted from their con texts for exploitation, the latter for protection from humans.

Making national resources

Much Western resource management is based upon the abstraction of a particular, most valued resource – say forests – from its socionatural environment (i.e., other species, ecosystems, humans). This construction of forests is not simply a reflection of how the forest works or is, but rather has often been part of a larger project of bringing land and labor into capitalist markets and under cen tralized control.

For example, David Demeritt chronicles a shift in the late nineteenth and early twentieth century United States, when quantification and graphic picturing were first used to produce US forests as a single national entity abstracted from historical, ecological, and social relation ships. These representational practices went hand in hand with the creation of a particular kind of nature: the normal forest, with scientifically predictable, regular harvest times and growth patterns for national production and con sumption. Along with these picturings and productions of the forest, argues Demeritt, came the consolidation of forestland ownership, the centralization of management, and increased disciplinary power to scientific experts and managers. Marking the space for national, scientific re source production in this case meant obliterating, as much as possible, anything that did not fit such as older (i.e., old growth) forests and the lay knowledge of loggers.

Conservation natures

Since the mid twentieth century, a global conservation movement concerned with the disappearance of natural landscapes and species has prompted many a nation state to carve out parts of national territory as spaces of and for nature. Conservationist efforts to separate nature from culture have led to acts of violent dispossession the world over. Roderick Neumann's research in Tanzania dem onstrates that so called natural areas actually have long histories of occupancy and use. The creation of protected areas in such places, he found, leads to violence and long term conflict, as peasant communities are deprived of access to land and natural resources.

Conservation also is a site in which exclusionary practices may be reconfigured. Juanita Sundberg's re search in Guatemala suggests that while transnational conservation projects often reproduce preexisting social hierarchies, they also create opportunities for change. For instance, administrators of the Maya Biosphere Reserve in Guatemala followed conventional practices of ex cluding Indigenous peoples – in this case the Itza' – from decision making circles. At the same time, however, Itza' men and women saw new value placed upon their en vironmental knowledge and therefore sought out spaces wherein race and gender hierarchies could be renegotiated.

Conclusion

In this article, we suggest that nature and culture are not neutral categories that merely describe the world; rather, they have been and are constituted dualistically to mark difference and hierarchy. As our examples illustrate, what comes to count as nature and culture has profound pol itical consequences. For this reason, de naturalizing these taken for granted categories is an incredibly important area of geographical research and theorizing. More generally, this research suggests that contemporary so cieties desperately need to challenge and ultimately move beyond dualistic framings of nature and culture, to learn how to live with difference differently.

New modes of relationality, of living with difference, is the motivating force behind Haraway's *Companion Species Manifesto*. This work implodes existing conceptualizations of nature/culture through rich descriptions and analyses of Haraway and others' relationship with dogs. Such re lations, she suggests, are not interactions between pre existing entities, but rather involve processes of co evolution and co constitution. Recognizing mutuality is part of Haraway's ethic of significant otherness: living with intimate others must be based on an openness toward and profound interest in the other's ways of being in the world, whether that other is human, animal, or inanimate.

Haraway's insistence on studying the ways dogs and other nonhumans act in and make a difference to the world avoids the critique often made of research showing nature to be thoroughly social: that de naturalizing de naturalizes in ways that risk erasing the materiality, so ciality, and sacredness of other species. While Haraway's new work has yet to be taken up in geography in any significant way, research on animal geographies, re lational ontologies, and posthumanism promises to offer new paths toward conceptualizing difference differently while also re enchanting nature. The challenge ahead involves translating such concepts into political practice.

See also: Animal Geographies; Cultural Geography; Feminism/Feminist Geography; Geography, History of; Nation; Nature; Nature, History of; Natures, Gendered; Posthumanism/Posthumanistic Geographies; Race.

Further Reading

Anderson, K. (2007). *Race and the Crisis of Humanism*. New York: Routledge.

Braun, B. (2002). *The Intemperate Rainforest: Nature, Culture, and Power on Canada's West Coast*. Minnesota, MN: University of Minnesota Press.

Castree, N. (2005). *Nature*. New York: Routledge.

Cerwonka, A. (2004). *Native to the Nation: Disciplining Landscapes and Bodies in Australia*. Minneapolis, MN: University of Minnesota Press.

Demeritt, D. (2001). Scientific forest conservation and the statistical picturing of nature's limits in the progressive era United States. *Environment & Planning D: Society and Space* 19, 431 459.

Gregory, D. (2001). (Post)colonialism and the production of nature. In Castree, N. & Braun, B. (eds.) *Social Nature: Theory, Practice, and Politics*, pp 84 111. Cambridge, MA: Blackwell Publishers.

Haraway, D. (1989). *Primate Visions*. New York: Routledge.

Haraway, D. (2003). *Companion Species Manifesto*. Chicago, IL: Prickly Paradigm Press.

Harvey, D. (1996). *Justice, Nature, and the Geography of Difference*. Cambridge, MA: Blackwell Publishers.

Latour, B. (2004). *The Politics of Nature*. Cambridge, MA: Harvard University Press.

Longhurst, R. (2001). *Bodies: Exploring Fluid Boundaries*. New York: Routledge.

Neumann, R. (1998). *Imposing Wilderness: Struggles over Livelihood and Nature Preservation in Africa*. Berkeley, CA: University of California Press.

Plumwood, V. (1993). *Feminism and the Mastery of Nature*. New York: Routledge.

Sluyter, A. (2002). *Colonialism and Landscape: Postcolonial Theory and Applications*. Lanham, MD: Rowman & Littlefield Publishers.

Smith, N. (1980). *Uneven Development*. Oxford: Basil Blackwell.

Sundberg, J. (2006). Conservation encounters and transculturation in the 'contact zones' of empire. *Cultural Geographies* 13, 239 265.

Relevant Websites

http://www.bruno latour.fr/
Bruno Latour's website.

http://www.lancs.ac.uk/fass/sociology/profiles/John Law/
John Law's website.

Cumulative Causation

E. Delaney, US Deparment of the Interior, National Park Service, Washington, DC, USA
J. W. Harrington, Jr., University of Washington, Seattle, WA, USA

Glossary

Disequilibrium Economics Any aspect of economic theory or modeling that recognizes increasing returns from the application of capital, highly skilled labor, or technology. Increasing returns can help explain the development of large firms and the sustained growth of wealthy, capital-rich regions or countries.

Elastic Demand Aggregate demand for a good or service that increases when its price declines, as is the normal expectation in economics.

Keynesian Economics Macroeconomic theory or policy which emphasizes the roles of the state in influencing aggregate demand for the production of goods and services.

Cumulative causation (CC) is a label for a set of mutually reinforcing processes leading to differential growth and decline among interacting regional economies. Scholars differ in the set of processes included in explanations of CC. Originally and most generally, CC is an expression of disequilibrium economics, in which regional economies respond to internal and external factors and experience different rates of economic growth and development.

CC is evidenced through differences in growth rate and social development between related or interdependent countries or regions. A body of literature has grown up around the work of a few key individual researchers, including Allyn Young, Gunnar Myrdal, A. O. Hirschman, and Nicholas Kaldor. The supply–demand relationship between regions may influence their relative development. Regions that consume raw materials sourced from supplier regions may experience economic growth and development, at the expense of supplier regions. If trade between regions consists primarily of factors of production being transported to regions where more advanced products are produced and are sold back to the supplier region, the more developed region tends to continue to develop while the supplier region tends to remain static.

Microeconomic Origins

The basis for CC was initially proposed in 1928 by Allyn Young as a response to, and extension of, Alfred Marshall's modification of orthodox economic theory to include increasing returns to firms through increases in the scale of production in manufacturing and the effects of external economies on production.

Young recognized the interdependencies between firms, markets, industries, and the overall economy. He observed that changes in firms' technology and the means or organization of production would have effects beyond the firm itself. In particular, as the market for a firm's production increased and the scale of production consequently increased, Young argued that firms would make capital investments in machinery to increase the specialization and efficiency of production. When these investments took the form of standard products of other firms (e.g., machine tools), rather than creating their own specialized and unique tools, the firm was both achieving savings through the use of standardized tools and influencing other industries.

Under conditions of elastic demand for the products of the firm, and scale related reductions in the cost of production passed on to the consumer in the form of lower prices, the competitive market leads to feedback between increasing sales, reduction in costs, and purchases by the firm of capital goods from other industries. Therefore, as the market for one product increases, the market for other products will increase, although not necessarily at the same rate. These relationships provide entrepreneurs with incentive to expand sales to increase profits by taking advantage of reductions in their costs.

Though Young did not use the phrase CC, his initial contribution established several tenets that underlie subsequent discussions and extensions of CC theory. First, CC is a process of disequilibrium. Second, as the scale of a market increases, increasing specialization of production leads to interdependencies between industries becoming established, with growth in one industry leading to growth in other industries. Third, the firm learns by doing, and fourth the entrepreneur plays a role by taking advantage of the economics of production and seeking new markets.

Myrdal: Social and Macroeconomic Development

In the second half of the 1940s and into the 1950s, much of the industrialized world was in economic ruins because of the effects of the Great Depression followed by a nearly global war. In the few years after the end of the war, many colonies of European countries gained their independence, leaving formerly dependent colonies in need of economic growth. The understanding of the processes by which poverty stricken, post war, and 'underdeveloped' countries and regions 'develop' was far more than an academic undertaking; it was a humani tarian necessity for regions to become self sustaining.

Myrdal was both a professor at the Stockholm School of Economics and a member of the Swedish Parliament from the Great Depression era into the post World War II era (1933–47). He rejected the theoretical simplifi cations of static economic equilibrium in analyzing re gional economies. Instead, he sought to explain why some regions grow and others either do not or regress. To do so, his theoretical framework diverged from neoclassical economic explanations where economic actors or firms are in control of their environment (e.g., the assumption of perfect knowledge). He did not completely abstain from using any simplifying concepts, however. Myrdal drew upon other fields, including the study of physics, for example, accepting that in a frictionless environment a moving object will continue to move until acted upon by an external force. This provided a framework where ac celeration and external forces contributed to his explanations.

Myrdal proposed the concept of circular causation to explain the relationships between more developed re gions experiencing growth and less developed regions supporting that growth. This concept, embraced by geographers because it expresses dynamic social and economic relationships between regions, was initially developed to explain the set of forces that kept many American Blacks in a state of poverty. Myrdal noted that a large number of interdependent factors were at work, including "employment, wages, housing, nutrition, clothing, health, education, stability in family relations, law observance, cleanliness, orderliness, trustworthiness, loyalty to society at large, etc. – which are all interrelated in circular causation." Increasing any one of these factors would result in an increase in all of the related factors.

When later applied to regional growth and develop ment, circular causation takes "into account the entire social system, including everything else of importance for what comes to happen in the economic field." When forces external to existing economic and social activities are applied to a region, it is possible to influence the state of socioeconomic factors within the region. Myrdal, a Keynesian economist, saw the possibility of government policies stimulating the interrelated social conditions within the region by inducing growth in one of the fac tors, typically by inducing economic activity.

In Myrdal's circular causation, development is ex pressed as an increase in measures of social well being. When one of the interdependent factors is encouraged to grow, others follow. Economic growth allows and en courages improvements in nutrition, health, education, law observance, etc., in the same direction. Because the circular causation leads to improvements in many interrelated social conditions, Myrdal noted that "cir cular causation normally will have cumulative effects." Certain conditions (e.g., cultural, social, or other inertia) in a region may lead to slow growth in some factors relative to others, making the 'coefficients of inter relations' a function of regional characteristics.

Myrdal described interregional dynamics that occur between growing developing regions and dependent, more static and less developing regions. Spread effects are the diffusion of positive aspects of growing regions to dependent, more peripheral regions through trade or through the effect of policy. Backwash effects strengthen the dominant/dependent relationship between de veloping and dependent, less developed regions. This is manifest through the movement of resources (e.g., skilled labor, capital that could be used to stimulate local growth, natural resources) from less developed to developed re gion. This can be an important source of dynamism in a developing region, but occurs at the expense of the less developed region because its potential for future devel opment is lessened. The classic colonial relationship ex emplifies this, as the colony ships raw materials to the colonial power in trade for manufactured goods.

Hirschman: Inter-Industry Linkages

A. O. Hirschman attempted to explain differentials in growth rates between interdependent regions after World War II. Hirschman explored the linkages between regions experiencing unbalanced growth. A significant contri bution was his exploration of how the flows between developed and less developed regions changed over time as the difference in development status became more equal.

Initially, Hirschman described a relationship where the developed region shipped finished goods to the less developed region for final processing, for example, packaging, warehousing, and shipping to consumer. The developed region enjoys an industrial environment where established linkages between producers and suppliers already exist. Raw and semifinished products and pro ducer services are available to the producer, allowing the efficient production of finished goods with economies of scale. The less developed region only has access to fin ished goods through trade. The absence of inter industry linkages, which can be expressed in an input–output table, presents a barrier to development.

Hirschman's conceptual framework, like his pre decessors', was of a manufacturing economy. Significant to Hirschman were the use of the input–output structure of the economy, and the presence of business acumen. Even if the material inputs for manufacturing were present in a region, a characteristic of development was the decision making and management capability required for successful industrial activity. The potential for development does not in itself lead to change; it may be latent in a region until an entrepreneur provides organization. Once inter industry linkages are established, it becomes possible for development to occur.

As domestic industries develop in the less developed region, some eventually achieve a state where they both return profits to investors and are able to compete with imports from the more developed region. As firms grow, investors will see opportunities in the creation of local suppliers to growing firms. These relationships are known as backward linkages, and growing supplier firms may stimulate the development of further backward linkages to support them. As industries grow, there is also the potential for the presence of semifinished goods to stimulate new industries that take advantage of them. These new markets represent forward linkages.

For Hirschman, an underdeveloped region was one where the potential for interindustry linkages were not realized; a developed region had both resources and entrepreneurial activity to establish interdependent industries. His definition of development and underdevelopment focus strictly on economic linkages, where Myrdal's definitions reflect interdependent economic activity and social conditions. Hirshman's use of the input–output structure to identify relationships and to explain how growth in one sector leads to growth in other sectors provides geographers with important theoretical and methodological frameworks for explaining how regions grow and change.

Kaldor: Income Elasticity

Nicholas Kaldor's work on CC rounds out the main theoretical framework for economic growth and development by considering the effects of demand, technological change, and agglomeration economies. Kaldor recognized that personal income can have a strong influence on the types of products in demand, and that different income classes tend to consume different types of products. He noted that the poor consume necessities like food and shelter. More affluent consumers purchase manufactured goods, and the rich are larger purchasers of services. In this way, income elasticity plays a significant role in how regions and countries relate to each other. Developed regions tend to experience trade with other developed regions, which can supply sophisticated manufactured goods and services. Poor regions focus on trade with regions that supply foodstuffs and basic goods, and only achieve developed status with difficulty.

As demand for products in one sector increases, related sectors experience increases in demand for their products. Technological change can influence demand through the introduction of new methods and products which lead to decreasing prices to producers. Agglomeration economies can provide firms with advantages due to reduced costs, but can also allow for industry specific skilled workers and an increase in the availability of information related to specialization.

Similar Concepts in Marxist Theory and Modern Economics

Marxist theory recognizes a tension between the demands of capitalist accumulation and the requirements of social reproduction. Accumulation of profit requires some displacement of costs outside the economic accounting of capitalism: to parts of the natural world that have not been valorized, to future generations, and to particular places. In geographic terms, this tension is manifested in capital's search for increased profit through (1) spatial divisions of labor that make use of low cost locations for low value added activities and (2) continuous redeployment to more profitable (typically lower cost) locations. The resultant spatial and temporal unevenness of economic development is problematic for people and social institutions that are attached to particular places.

Modern economic models place an emphasis on imperfect competition. Most relevant to CC is the presence of increasing returns to production scale and economic agglomeration. Paul Krugman modeled economic location and trade based on increasing returns and path dependence, and concluded that as transportation costs fall, local specialization increases in centers that have developed requisite institutions, infrastructure, and supporting services. Places lacking these external economies have lower returns to investment, despite their lower costs. This pattern of CC can be disturbed only by major technological (and perhaps other environmental) changes that reduce the value of scale in existing centers.

Summary

In sum, CC is a term that encapsulates complex and dynamic relationships among countries and among constituent regions. The CC framework emphasizes individual economies in disequilibrium, and differential growth among regional economies. While the effects of CC can be observed in comparisons of regions' or countries' annual flows of goods, investment, and people, the term itself refers to processes that occur over time. Once established, patterns of growth and development tend to be perpetuated, but changing economic and institutional conditions can interrupt the process, even with continued interaction among regions.

See also: Agglomeration; Development I; Input-Output Analysis.

Further Reading

Appelqvist, Ő. and Andersson, S. (eds.) (2005). *The Essential Gunnar Myrdal*. New York: The New Press.

Kaldor, N. (1989). *The Essential Kaldor*. New York: Holmes & Meier Publishers.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy* 99, 483 499.

Toner, P. (1999). *Main Currents in Cumulative Causation*. New York: St. Martin's Press, Inc.

Cyberspace/Cyberculture

D. Bell, University of Leeds, Leeds, UK

Glossary

Cybernetics A theory of the control and communication of regulatory feedback in biological, technical, or social systems. Changes to the external environment are looped back to the system, which makes adjustments to maintain a steady state. The term 'cybernetics' stems from the Greek word *kubernites*, meaning steersman, governor, pilot, or rudder. Cybernetics is the science of communication and control in living beings or machines.

Digital Divide A key term used to describe new inequalities that result from the distribution of access to information and communication technologies (ICTs) at different spatial scales. New maps of 'haves' and 'have nots', many of which reproduce older forms of inequality, can be seen from the global scale down to the household.

Futurology The 'science' of predicting future scenarios; in a rapidly changing environment, futurology has become very important, but is also prone to misprediction given the uncertainty of paths of development in, among other things, new technologies.

Me Media Against arguments that cyberspace is a new domain of corporate control, users are constantly finding innovative new ways to make their own uses of new technologies. Weblogs (blogs) have become very important manifestations of this user-controlled desire to produce, disseminate, and consume new media content, sometimes tagged Me Media to reflect its personalized formats.

Introduction

Cyberspace has been a hot topic across a range of academic disciplines, including geography, for close to two decades now. It has also been subject to intense popular media attention, both factual and fictional, littering news stories, movies, urban myths, and everyday talk. Credit for coining this neologism routinely rests with the science fiction writer William Gibson, who conjured a future image of vast 'datascapes', landscapes made from computerized information, which users accessed by connecting their consciousness directly to global computer networks. This process, of 'jacking in' to 'the matrix', remains a potent image of the experience of entering cyberspace – the 'space between the screens' or the 'world of the wires' that we enter when using the Internet, the World Wide Web, or any of the other proliferating forms of computer mediated or digital communications, information, or entertainment resources around us.

A keen geographical eye will immediately notice the suffix ' space' that makes up half this new word. While the 'cyber ' prefix is taken from the science of cybernetics – the study of self regulating feedback systems – this suffix alerts us to the spatiality, or geography, of cyberspace. But what kind of space is it? As Gibson foresaw it, it is an imaginary, immaterial space: a cityscape made of data, standing like skyscrapers or flashing by like car headlights. It is a way of describing the mass of information shared across all computers, databases, and other containers, and the connections that weave them into a web. Does this mean that its geography is also imaginary, a 'Gibsonian' mental map that helps us understand what we're doing when we go online? Or are there other geographies of cyberspace to consider?

I am writing this sitting in front of a computer, which is running a word processing program but which is also online, allowing me to periodically 'jack in' to cyberspace, to check my e mails, search the web for a key fact, keep an eye on the day's news, and so on. My computer is a node in cyberspace, then, connecting through wires (and wireless connectors) outward, into the ether. However, it is also grounded, localized: on a desk, in a home office, in a house, in a city, and so on. It is 'here'. And so am I, typing away, looking at the screen (or screens, as I click between different programs and files). So, there are at once two geographies that I am experiencing: one of global connectivity, of immaterial, aspatial, 'everywhereness'; the other, perhaps more mundane, of absolute 'hereness', of locality and place. While this 'hereness' is also a reflection of profound changes brought about by cyberspace – such as the changing work patterns enabled by cyberspace that have brought academic work so fully into my home, for better or for worse, not to mention the manifold transformations of work practices, the deskilling and reskilling, brought about by computerization – there is, without doubt, a 'here' here. Early discussions of cyberspace, by contrast, tended to stress its dematerializing, disembodied, aspatial character, and grew from that point a series of powerful ideas about this new world some of us were moving in to: 'leaving the meat behind', playing with identity, building new 'virtual' communities, altering the very fabric of society. Some of these things have taken place, but they have not necessarily taken place away. Perhaps it is useful, in order to illustrate this key point, to break things down a little, by looking at particular geographies of cyberspace. For reasons of brevity, the entry will focus here on two: economic geographies and cultural geographies.

Economic Geographies of Cyberspace

The economic geographies of cyberspace are arguably the most important and problematic spatializations to understand. Gibson prefigured this, in a way, with his fiction's emphasis on corporate control of information – his heroes are often hackers trying to access data held in corporate hands, which they can then sell on into the informal informational economy. So the first economic geography concerns the value of information, and the industries of production, storage and retrieval, circulation and consumption of data. Critics argue that this new industrial sector reduces all of human activity to codable, sellable data, mined through the many electronic actions, interactions, and transactions many people participate in daily, from web surfing behavior to credit card transac tions. However, the information economy is more com plex and multilayered than that; it is a global system of trade in data (or information capital), a new set of eco nomic activities and a new form of economic com petitiveness, founded not on 'traditional' economic activities but on these new sources of postindustrial wealth. And while this new economy might be imagined to be 'footloose', not tied to particular locations, what has instead emerged is a new set of uneven economic geographies, a new distribution of core and peripheral economies, new riches and new forms of poverty – all of them mapped onto space (and cyberspace). At various spatial scales, there are those who are information rich and those who are information poor; there are vast in equalities in access to cyberspace's hardwares (phone lines) and softwares (not just programs, but also computing literacy and skill). The information economy, moreover, produces a new division between the 'high end' services and the 'low end' marginal activities that support and enable them, such as component manufacture or data processing. Not everyone with a job in cyberspace is a highly paid programmer or a dot.com entrepreneur. While even 'low end' activities may represent new eco nomic opportunities in marginal or peripheral econ omies, this complex new economic geography reinforces the importance of exploring how locational factors con tinue to shape economic activity and opportunity. For consumers, too (or would be consumers), there are huge digital divides, separating the cyberspace 'haves' and 'have nots' at global, national, regional, and local scales.

This new economic geography demands we ask some quite old fashioned questions; political economy questions that were eclipsed in earlier optimistic readings of the promises of cyberspace. Questions like: who can afford to go online? Who has the capacity to use data, rather than to be merely made into data? How has cyberspace impacted on and in different places, pro ducing not only new opportunities but also new in equalities? Has the Internet produced greater equality or greater polarization? If, as the sociologist Manuel Castells argues, there are now two worlds, the 'space of flows' and the 'space of places', how do these interact, what are the power relations, who's exploiting whom? These are vital questions for geographers to ask and try to answer, be cause they are vitally 'geographical' questions.

One key site for examining these new economic geographies has been the city. Early futurological ac counts of the impacts of information and communication technologies (ICTs) were that their networked capacities would finally obliterate the necessities of place, making us all footloose or nomadic. As a result, it was suggested that old geographical formations, such as cities, would dissemble. Cities were seen as problematic legacies of an earlier age – congested, polluted, atomized, and anomic, and a brighter future was predicted, with teleworkers relocating away from urban blight to live networked lives in idyllic isolation. A lot of the technical innovations prefigured in these accounts – not just working at home, but shopping, sourcing entertainment, learning, social izing, and so on, virtually – have come about in some form or another. Yet cities still exist, even dominate; they have, in fact, become more important, more concen trated, even as they have sometimes become dis embedded from their immediate environs. A global network of densely connected cities has grown up, housing the command centers of the new economy, and new spaces of spectacular consumption. To understand why the futurologists were so off target means looking not just at the economic impacts of cyberspace, but also key cultural factors.

Cultural Geographies of Cyberspace

While economic analyses of cyberspace were initially neglected, perhaps for being too old fashioned and 'un sexy', discussions of the cultural transformations of new ICTs have been much more vocal, and also heavily po larized between those who see in these changes the dawning of a bright, liberated new age, and those who see either an intensification of all society's current ills, or new problems ahead. To get an angle on these two positions, it might be wise to begin with one key area of debate: the impact of cyberspace on 'community'. This has arguably been one of the longest running and bitterly contested debates, which shows little sign of resolution. Enthusiastic proponents of new online communities see them offering countless advantages to members – ad vantages often attributed to the transcending of 'real' space or geography. Communities can now form around interests, practices, or commonalities, rather than being the result of proximity or propinquity. People can go online and build their own communities, selectively and electively, making bonds as strong or weak as they wish, joining as many different communities as they choose, and staying a member only for so long as they consider it worthwhile. These positive attributes of online com munity are quickly turned into problems by critics, however. For those pessimistic about or hostile to the idea of online or virtual community, selectivity and electivity are weaknesses, meaning that members have only narrow, instrumental reasons for belonging. 'Real' communities, they argue, take the rough with the smooth, grow over time, and are multifaceted. Exchanging e mails is too ephemeral to constitute a community.

These debates have been circling for more than a decade, as 'virtual communitarians' extol the virtues of new forms of belonging, and critics argue the in compatibility of 'virtual' and 'community' – although often with exceptions, as in discussions of diasporic communities, where online co presence is discussed more positively (sometimes at least). This last example opens up a second key issue: the question of identity in cyberspace. Again, there has been a polar debate about the pros and cons of 'online identity play' – about the possibilities of entering cyberspace freed from 'real' identity, and the ability to remake oneself online by se lecting an identity. Especially in earlier, text based online forums or communities such as chat rooms or multiuser domains (MUDs), this option generated a lot of dis cussion (both inside and outside participants themselves). At opposite ends of the scale were those who rejoiced in the freedom to endlessly make and remake self identity, to morph into elaborate, fantastic new beings, and those who saw in these practices only deceit, schizophrenia, paranoid lack of trust, and the unleashing of demons.

Identity and community debates connect to a third important topic area in cyberspace that has attracted a lot of attention and that is very amenable to a geographical analysis: debates about what kind of political space cyberspace affords. Are we seeing new democratic arenas being created, a new public sphere or agora, where people can freely exchange ideas and critique? Or is cyberspace so thoroughly colonized by the powerful, by corporate capitalism, that it can only ever be a space of domination and exploitation? The early history of cyberspace developments certainly foresaw democratic and countercultural potential; indeed the broader history of personal computing is tied to the 1960s hippie culture. Yet the Internet is part of another history, too – the Cold War military history of 'mutually assured destruction' and total war. For some critics, the origin of the Internet in military research makes it always a tool of oppression, despite attempts to disembed it from this past. And as the Internet and the Web in particular have developed, it is undeniable that they have been thoroughly corporatized, as the computing and media giants battle for domination.

Yet even under such conditions, there is something unruly and ungovernable about cyberspace, as the recent explosion in 'Me Media' shows. Me Media describes a new type of content available online, produced and consumed by 'ordinary users', such as blogs. The rise in this kind of 'amateur' content production has certainly questioned the criticism that most of us have become merely passive consumers of cyberspace, surfing content made by corporate providers. Me Media suggests the inventiveness of users to work with, around and against the wholesale corporate colonization of cyberspace. Control of cyberspace by corporations or nation states is also contested by hackers, who have a longer lineage of countercultural activity (though this has largely been presented as a form of malicious 'cybercrime' rather than a legitimate challenge to colonization).

These debates about community, identity, and dem ocracy have been advanced, and maybe even moderated, by the growing number of academic studies venturing online. In place of polemic, scholars have been advancing theo retical and empirical agendas, in order to better understand the new worlds that are emerging. Identity, community, and democracy have long been hot topics in this growing interdisciplinary enterprise, which some people have tag ged with another neologism, cyberculture studies.

Cyberculture Studies

Cyberculture studies represents the coming together of diverse strands of academic work across a range of subject areas. As such, it is work carried out in a diversity of (sometimes conflicting) intellectual and institutional lo cations, often embedded within more 'traditional' subjects and departments, such as cultural studies, sociology, or human geography. This gives cyberculture studies a heterodox richness, and an anticanonical stance: it is both theoretically and methodologically promiscuous. How ever, this can also make it seem like a fringe activity, not a 'proper' subject; it is too 'faddish' and lacks the patina of older disciplines. Its proponents would argue that these are advantages, signaling an openness that many 'tradi tional' subjects lack. To sketch its parameters rather than attempting a strict definition (which is often changing, sometimes elusive), cyberculture studies includes the following:

- Cultural studies approaches to understanding the material, symbolic, and experiential dimensions of cyberspace and to cybercultural forms, practices, politics, and identities.

- Work by geographers on the manifold spaces of cyberspace, and work in urban studies on ‘cybercities’ (see below).
- Work in computer science and other related ‘cyber technosciences’, including hardware and software de velopment and user modeling, robotics, artificial intelligence (AI) and a life, nanotechnoscience, and so on.
- Insights from the history of science and technology, for example, on the histories of computing, and from science and technology studies.
- Sociological studies of the uses and impacts of new technologies.
- Literary theories and studies, for example, those concerned with science fiction and cyberpunk.
- Media studies work on new media, multimedia, or digital media, and film studies work on sci fi cinema, digital filmmaking, new modes of film production, distribution, and consumption.
- Philosophy of science and technology, and philo sophical theories used to think about cyberspace and cyberculture.
- Economics and organization studies of changing work patterns in the information and knowledge economies.
- Feminist studies of science and technology, including ‘cyberfeminisms’ and ‘cyborg feminisms’.
- Social, political, and cultural ‘cybertheory’ in its di verse forms.
- Research in the biomedical and biotech sciences on the interfaces between bodies and technologies, such as gene therapies or xenotransplantation.
- Policy oriented studies, whether in social, welfare, communications, cultural, or new, hybrid ‘cyberpolicy’ contexts.
- Studies of the creative and applied arts intersecting with new technologies, and studies of the esthetics of new technologies.
- Work on ‘cyberpsychology’, on the psychological im pacts of cyberspace.
- Research in linguistics into the languages of new technologies and their users.
- Cross disciplinary futurology that predicts ways of living with machines yet to come.

As such, cyberculture studies is a complex field which not only makes use of diverse academic traditions and theo retical perspectives, but also deploys a diversity of re search methods and approaches. In addition, each of these domains has its own geographies, its own spatial mani festations. One of these strands will be picked out here for illustration, and to connect back up to earlier key points.

Cybercities

This has been a key focus of geographical work: to ex plore the intersections between digital technologies and cultures, and urban life. There has been a blossoming of theoretical, empirical, and applied research over the last decade, leading at least one writer, Stephen Graham, to call for a distinct subdiscipline called ‘urban ICT studies’. As noted already, it was not that long ago that futur ologists were predicting the end of city life as we know it, making a term like ‘urban ICT studies’ apparently oxy moronic. The fact that this has become such a fertile area of research tells us that things often don’t turn out as predicted: ICTs have led to intensified urban concen tration, even as they have radically transformed the patterns of city life. The rapid growth and diversification of the new information economy has led to the devel opment of specialized urban centers in this new ‘network society’; while some writers see these command centers as disembedded from their regional and national locale, they remain definitively urban. In some cases, such cities have fostered innovative mileux, attracting high tech workers and firms and providing an environment con ducive to innovation. And, of course, cities have become the major markets for ICTs, where a culture of ‘techno logicalness’ is fostered and where new technologies are ‘beta tested’, adopted, and domesticated. Yet cities are changed by cybertechnologies: their built form, their infrastructure, their uses, their cultures are all trans formed, reconfigured, or repurposed. Of course, this cybercities discourse needs moderating by asking im portant questions about specificity: not all cities are being transformed in the same ways. There are winners and losers as the urban hierarchy shifts to accommodate cybercities, not every city successfully morphs into a ‘city of bits’. However, it is clear that ‘urban ICT studies’ will continue to be an important thread in studies of cyber space and cyberculture, and one to which human geog raphers have much to contribute.

There is one important final point to make: there has been a lot of debate about the words ‘cyberspace’ and ‘cyberculture’, a feeling that they may be old fashioned and imprecise terms, especially given all the unremitting convergence and ‘remediation’ that makes the landscape of ICTs so complex and ever changing. Cyberspace is, for some critics, a term too closely associated with either earlier stages in the development of both the technolo gies and ways of thinking about them; it is connected to ideas about virtual reality, or text based MUDs or cy berpunk. For some, this earlier phase has been super ceded, leading some scholars to prefer to talk of ‘digital culture’ or ‘new media’ – one exasperated critic even wrote of ‘fill in the blank’ studies. The name change has been mirrored in a shift in the key concerns of scholars in the field, too, for example, in a turn toward more ethnographic work and in a growing focus on the ‘ma terial culture’ of cyberspace. While the names we give to these topics are as yet unsettled, this must surely be taken as a sign of the vibrancy and pace of change in the field;

there can be no doubt that human geographers (or 'cybergeographers') have a key role to play in its on going analysis.

See also: Informational City; Knowledge Economy; Mapping, Cyberspace; Public Space.

Further Reading

Abbate, J. (1999). *Inventing the internet*. Cambridge, MA: MIT Press.
Bell, D. (2001). *An introduction to cybercultures*. London: Routledge.
Bell, D. and Kennedy, B. M. (eds.) (2007). *The cybercultures reader* (2nd edn). London: Routledge.
Crang, M., Crang, P. and May, J. (eds.) (1999). *Virtual geographies: Bodies, spaces and relations*. London: Routledge.
Dodge, M. and Kitchin, R. (2001). *Mapping cyberspace*. London: Routledge.
Graham, S. (ed.) (2004). *The cybercities reader*. London: Routledge.
Kitchin, R. (1998). *Cyberspace: The world in the wires*. London: John Wiley.
Kitchin, R. (1998). Towards geographies of cyberspace. *Progress in Human Geography* 22, 385 406.
Loader, B. D. (ed.) (1998). *Cyberspace divide: Equality, agency and policy in the information society*. London: Routledge.
Marshall, P. D. (2004). *New media cultures*. London: Arnold.
Zook, M. (2005). *The geography of the Internet industry: Venture capital, dot coms, and local knowledge*. Oxford: Blackwell.

Relevant Websites

http://www.mappingcyberspace.com
Maps to Accompany Dodge and Kitchin's Book.
http://www.cybergeography.org/atlas
More Maps and Visual Depictions of Cyberspace.
http://www.rccs.usfca.edu
Resource Center for Cyberculture Studies.

D

Darby, H. C.

R. A. Butlin, University of Leeds, Leeds, UK

Darby, Henry Clifford (1909–1992)

Henry Clifford Darby (see **Figure 1**) was one of the most influential figures in the field of historical geography in the twentieth century, pioneering the mapping of medieval English data (notably the land measurements and other material contained in the Domesday survey of 1086), the description and analysis of regional landscapes, notably the Fenland of Cambridgeshire, the represen tations of the historical geographies of England, and some of the related methodological and conceptual problems of describing and accounting for changes in the past. While based at Cambridge during World War II, he played a major role in the production of 30 Admiralty Handbooks: books produced to provide information for the British Navy on the history and geography of states of Europe and other countries involved in the war. He built up groups of researchers in historical geography in the university departments where he was Professor of Geography: at Liverpool (1945–49), at University College, London (1949–66), and at Cambridge (1966–1976). He was recognized as a polished lecturer and writer, and an experienced and effective administrator.

Figure 1 Henry Clifford Darby (1909 1992).

Darby was born in Resolven in the county of Glamorgan, South Wales, on 7 February 1909, was edu cated at Neath County School, and in 1925 became a student at St Catharine's College, Cambridge, where he graduated with first class honors in geography in 1928. He then became a research student at the same college, spending part of his first research year working in Northern and Southern Rhodesia and Nyasaland on pioneer settlement as part of a project managed by Isaiah Bowman, the director of the American Geographical Society. He obtained his PhD in 1931 (the first geography PhD awarded at Cambridge) with a dissertation on the role of the Fenland in English history, supervised by Bernard Manning (a lecturer in geography with strong interests in medieval history). He was appointed to a university lectureship in geography at Cambridge in 1931, and elected to a fellowship at King's College, Cambridge, in 1932.

Darby's early publications in the late 1920s and early 1930s included papers on the architectural geography of Great Britain, the tinplate industry in the Vale of Neath, settlement geography in Northern Rhodesia, and a series of papers on the Fenland. He published two books on the Fenland in 1940: *The Medieval Fenland* and *The Draining of the Fens* (the latter being republished in 1956). Many years later, in 1983, he published *The Changing Fenland*, comprising revisions of five chapters of *The Draining of the Fens* plus an introductory chapter on the Middle Ages. As was the case for much of Darby's regional work, these volumes evidence the most thorough interrogation of manuscript and published sources, detailed knowledge based on personal field experience, and meticulous presentation.

However, H. C. Darby is perhaps best known for his collaborative studies of the geographies of the material contained in the volumes of the *Domesday Book*, using a range of mapping techniques to portray a wide range of phenomena that help to identify and characterize essential features of the administrative, settlement, and economic geography of medieval England at a particular point in time. This was a considerable feat, accomplished in the days well before computers and geographical information systems. Seven volumes of the *Domesday Geographies* were published by Cambridge University Press between 1952 and 1975, five of them dealing with regions and the remaining two a gazetteer and a review volume. These studies also paved the way for similar analyses and mappings of historical data, such as tithe and census data, which contributed considerably to the understanding of historical geography.

Darby also edited and contributed to two important sets of essays on the changing historical geography of England: *An Historical Geography of England before A.D. 800*, published in 1936, and a *New Historical Geography of England*, published in 1973. The bases of these books were chapters written by distinguished scholars, not all geog raphers, dealing with England and some of its regions, either at specific periods in time or longer periods of transition and change. The first of these books was made up of (largely thick) cross sections of history, such as 'The human geography of Roman Britain' (Chapter 2), 'Leland's England' (Chapter 9), 'The draining of the Fens' (Chapter 12), and 'The growth of London, AD 1660–1800'. This volume was seen by Darby as a methodo logical compromise in the management and presentation of place through time, and hence the second book adopted a slightly different approach, using both time slices but also narratives of change through time. The titles of chapters 5–8 were: 'Changes in the later Middle Ages', 'England *circa* 1600', 'The age of the improver: 1600–1800', and 'England *circa* 1800'. They were illus trated by meticulously drawn maps and bore the signs of very careful editing. The first of these two books was in every way a pioneering product, the second a way of attempting to address some methodological and practical problems of the first (though partly also a vehicle for new research findings). Darby's interest in mapping was also reflected in his joint editorship of *The University Atlas*, published first in 1936 and running into 22 editions, and the *New Cambridge Modern History Atlas*, published in 1970.

Darby addressed some of the basic characteristics and problems of historical and human geography in his published essays and addresses to learned societies, notably his 1962 paper on 'The problem of geographical description', but theory was not high on his list of intellectual priorities. Nonetheless, he published a wide range of papers on methodology that revealed what Alan Baker has described as his 'mission to establish a new, catholic historical geography and his passion for methodological challenges and experiments'. He pre ferred on the whole to work on empirical data and their analysis and presentation, and research that led to better understanding of the historical geography of landscape. A seminal paper of 1951 on 'The changing English landscape' was a timely summary of such processes of change as woodland clearance, marshland drainage, heathland reclamation, changes in arable land, landscape gardening, and changes in towns and industry. Although, as Michael Williams has suggested, "It caused few flutters in the geographical world…Few realised that it was, in fact, a radical departure," that "sought to integrate people and the world they inhabited." Darby was always interested in and connected with the historical geography of countries and regions outside Britain, notably France, Europe more generally, and the US, and used material from readings and visits in some of his publications and academic seminars, though a number of manuscripts on these subjects were not published until after his death.

Clifford Darby's outstanding work was recognized with a knighthood in 1988 for 'services to historical geography', by the awards of seven honorary doctorates, and by his election to fellowship and membership of a range of august institutions, including the British Academy and the Royal Commission on Historical Monuments. His work was directly and substantially influential, along with that of others, in establishing historical geography as a major subject of research and teaching in universities in Britain and elsewhere, and of encouraging younger scholars, usually on his own terms, to commit to this fascinating and challenging subject. Some of his own students would later produce critiques of his work, which itself had earlier been a critique of the practices of historical geography and geographical history of the 1920s and 1930s.

See also: Archives; Bowman, I.; Historical Geography; Regional Geography II.

Further Reading

Baker, A. R. H. (1992). Obituary. Henry Clifford Darby 1909 1992. *Transactions of the Institute of British Geographers, New Series* 17, 495 501.

Baker, A. R. H. (2003). *Geography and History. Bridging the Divide*. Cambridge: Cambridge University Press.

Clout, H. D. (2007). Henry Clifford Darby 1909 1992. Geographers. *Bibliographic Studies* 26, 79 97.

Darby, H. C. (1989). The published writings of Sir Clifford Darby, C.B.E., F.B.A. *Journal of Historical Geography* 15(1), 5 13.

Darby, H. C., Williams, M., Clout, H., Coppock, T. and Prince, H. (2002). *The Relations of History and Geography. Studies in England, France and the United States*. Exeter: Exeter University Press.

Harley, J. B. (1989). Historical geography and the cartographic illusion. *Journal of Historical Geography* 15(1), 80 91.

Lawton, R. and Butlin, R. (1989). Clifford Darby: An appreciation. *Journal of Historical Geography* 15(1) 1, 14 19.

Mead, W. R. (1992). Professor Sir Clifford Darby 1909 1992. *The Geographical Journal* 158(3), 352 355.

Meinig, D. W. (1989). Clifford Darby: An American appreciation. *Journal of Historical Geography* 15(1), 20 23.

Prince, H. (1992). H. C. Darby, 1909 1992. *Journal of Historical Geography* 18(4), 456 460.

Williams, M. (1989). Historical geography and the concept of landscape. *Journal of Historical Geography* 15(1), 92 104.

Williams, M. W. (1995). Henry Clifford Darby 1909 1992. *Proceedings of the British Academy* 87, 289 306.

Darwinism (and Social Darwinism)

H. Winlow, Bath Spa University, Bath, UK

Glossary

Descent All animal species are linked through time and space and have a common ancestor.
Gradualism A series of small-scale changes which over time result in large-scale change, for example, in the evolution of geological forms or plant and animal species.
Monogenism The idea that all human races are united as one species.
Natural Selection The individual traits within each species are selected by environmental conditions which, over time, will result in the emergence of new species types.
Organism Analogy The comparison of larger groups in society to the development of an individual organism at a variety of scales, for example, a city, a state, or the world.
Polygenism The idea that different human races are different species and that some of these species are biologically and socially inferior to others.

Darwinism: Definition

The term Darwinism stems from Charles Darwin's (1809–82; **Figure 1**) biological evolutionary theories, including his development of the theory of natural se lection, as outlined in his 1859 text *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. Following the publication of *The Origin* the term was initially coined by Darwin's supporters, but over time it became associated with a wider suite of biological evolutionary theories (despite Darwin's attempts to narrow the term to incorporate only natural selection). This resulted in its application to forms of evolutionary theory which were quite distinct from Darwin's natural selection and his stress on haphazard development. In the late nineteenth and early twentieth centuries many scientists, including geographers, described themselves as Darwinian, al though their work clearly displayed the influence of a number of biological evolutionary theories, includ ing neo Lamarck(ian)ism which emphasized linear pro gressionism. Conventionally evolutionary theory has been widely regarded as synonymous with Darwinism, but historians of science note an overuse of Darwin in accounting for the widespread influence of evolutionary theories on emerging scientific disciplines, including geography and anthropology, in the late nineteenth century.

Origins of *The Origin*

Charles Darwin has been popularized as the champion of nineteenth century biological science, and natural selection has been portrayed as a turning point in social and scientific thought – a misconception that remains within mainstream representations of evolutionary sci ence. In attempts to dispel this myth, recent historians of science have argued that natural selection represented a continuation of ideas, rather than a turning point. In contrast to Kuhn's paradigmatic framework recent accounts emphasize a slow revolution, where Darwin's work is situated in the context of a longer standing tradition of biological and geological investigation and within the context of the work of his contemporaries.

The construction of the geological column and the establishment of the antiquity of the Earth in the early nineteenth century represented a divergence from the creationist view of world history. Charles Lyell's 1830–33 text *Principles of Geology* suggested a gradual, rather than a catastrophic, evolution of landscapes. This was partly evidenced by research undertaken on the Jurassic

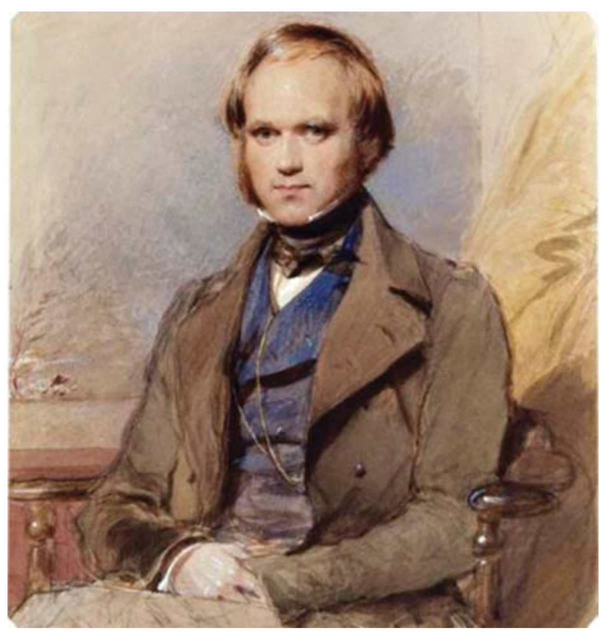

Figure 1 Charles Darwin in 1840, by George Richmond.

coastline in Dorset, England, and his emphasis on gradual change reflected a British conservative political outlook, which contrasted with revolutionary changes in Europe. Lyell's geological system meant that fossils could no longer be seen simply as markers of strata but as evidence of former life. The acceptance of his timescale marked a cornerstone in the development of geology, and influ enced other fields of research, including developmental biology. While Lyell and Darwin shared a gradualist approach to science, Lyell rejected the Darwinian tree of life and associated concepts of the ape ancestry of the human species – links that can be traced back to the Linnaean taxonomy of the eighteenth century. In *The Origin* Darwin used Lyell's stress on gradual temporal changes in support of the theory of natural selection:

> I am well aware that this doctrine of natural selection … is open to the same objections which were at first urged against Sir Charles Lyell's noble views on "the modern changes of the earth, as illustrative of geology;" but we now very seldom hear the action, for instance, of the coast waves, called a trifling and insignificant cause, when applied to the excavation of gigantic valleys or to the formation of the longest lines of inland cliffs. Natural selection can act only by the preservation and accumu lation of infinitesimally small inherited modifications, each profitable to the preserved being; and as modern geology has almost banished such views as the excavation of a great valley by a single diluvial wave, so will natural selection, if it be a true principle, banish the belief of the continued creation of new organic beings, or of any great and sudden modification in their structure. (Darwin, 1859: 95 96)

While the idea of the antiquity of the Earth gradually became incorporated into theological perspectives, the origin of plant, animal, and human species remained controversial. In the first half of the nineteenth century, the creationist view of development dominated society and science – although earlier evolutionary theories, including Lamarckian and Geoffroyan doctrines of comparative evolution, had gained support among London's working class evolutionists. Realizing that his ideas would not gain support from the scientific elite in the 1830s, Darwin delayed publication for at least 20 years, during which time he consolidated his work and gathered further evidence to support his theories. When *The Origin* was published in 1859, evolutionary ideas were becoming less controversial among the middle classes. Darwin, however, still received criticism from the clergy and the popular press, as evidenced in a number of cartoons (**Figure 2**). Darwin published his theories after receiving a letter in 1858, from fellow naturalist Alfred Russell Wallace, which revealed that he had come to similar conclusions about species evolution.

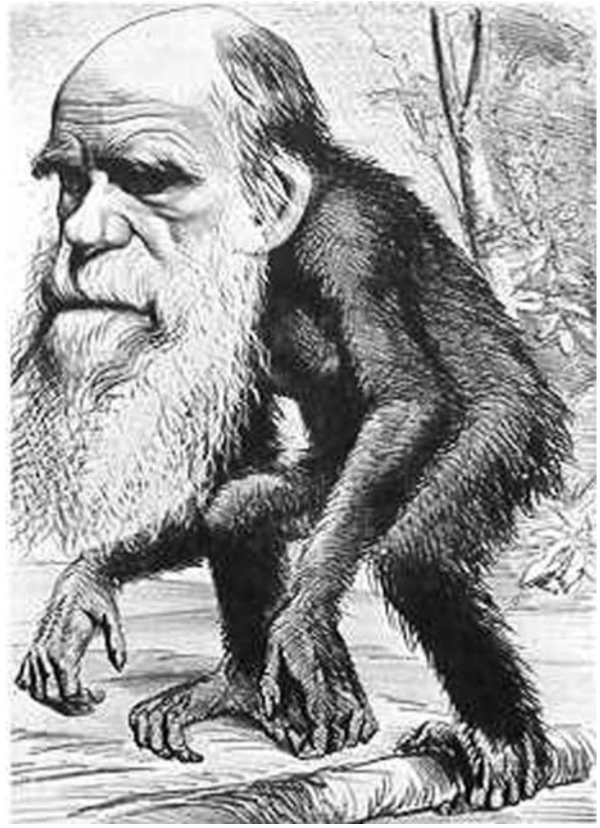

Figure 2 Darwin as an Orangutang. Source: *The Hornet* (1871).

Several major influences on Darwin's development of natural selection have been discussed by geographers and historians. These influences include: first, discoveries made while on the Beagle voyage of 1831–36; second research into pigeon breeding; and third, Reverend Thomas Malthus' 1798 *Essay on the Principle of Population*. The 5 year Beagle voyage made a strong impression on Darwin's life and intellectual development. He traveled extensively and observed a wide variety of plant, animal, and human species in their native environments. One reason that *The Origin* received rapid recognition from contemporaries was because Darwin was known as the author of *The Voyage of the Beagle* (originally published as the third volume of the *Journal of the Researches into the Geology and Natural History of the Various Countries Visited by HMS Beagle* in 1839), an ethnographic account which included descriptions of indigenous peoples and their struggle for life, particularly in relation to colonization. The geographical distribution and diversity of species encountered on the voyage led Darwin to question Paleyian natural theology. It was clear that geographic environment, be it continental or island based, had played a role in the development of species variety, as evidenced, for example, by the Galapagos turtles and the unique finch varieties found on each separate island.

It was not until the late 1830s, after his return to England, that Darwin developed the idea of natural selection.

Following his return Darwin undertook investigations into pigeon breeding and observed that, over time, decisions by breeders to select specific characteristics had resulted in many different breeds. He speculated that nature may behave in similar ways and developed the metaphor of nature as breeder. If nature was viewed in the same way as generations of human breeders it may be possible to discover the principles of selection in nature.

Commentators have stressed, to varying degrees, the influence of Malthusian concepts on Darwin's work. Darwin read Malthus' *Essay* in September 1838 and the ideas expressed there helped shape his ideas on natural selection. Malthus observed that the human population increased exponentially, while food production only increased arithmetically, and stressed the need for positive and preventative checks for population control. Positive checks included starvation, famine, pestilence, and war, while preventative (or social) checks included moral restraints, late marriage, and cultural practices, such as abortion, infanticide, and birth control. Darwin viewed the struggle for existence identified by Malthus as applicable to all animal and plant species:

> A struggle for existence inevitably follows from the high rate at which all organic beings tend to increase. … Hence, as more individuals are produced than can possibly survive, there must in every case be a struggle for existence, either one individual with another of the same species, or with the individuals of distinct species, or with the physical conditions of life. It is the doctrine of Malthus applied with manifold force to the whole animal and vegetable kingdoms; for in this case there can be no artificial increase of food, and no prudential restraint from marriage (Darwin, 1859: 63 64).

Natural Selection

The theory of natural selection was the main tenet of Darwin's biological evolutionary scheme, although other important aspects included descent, gradualism, the multiplication of species (including geographic speciation), and sexual selection. While incorporating Malthusian aspects of the multiplication of species and population checks, Darwin's insight was that the stronger members of each species would survive and the weakest would be eliminated (become extinct). These concepts formed the basis of natural selection, which was Darwin's 'mechanism' of evolutionary change. Darwin argued for a natural process of variation, struggle, and the selection of traits favorable for survival:

> Owing to this struggle for life, any variation, however slight and from whatever cause proceeding, if it be in any degree profitable to an individual of any species, in its infinitely complex relations to other organic beings and to external nature, will tend to the preservation of that individual, and will generally be inherited by its offspring. … I have called this principle, by which each slight variation, if useful, is preserved, by the term of Natural Selection, in order to mark its relation to man's power of selection (Darwin, 1859: 61).

Darwin's theory of natural selection incorporated the idea that all species were linked to one another in time and space by descent. This was expressed through the metaphor of a branching tree of life, where the trunk symbolized the common ancestor from which all animals had originated. Central to this model of evolutionary development was the idea of a haphazard process of evolution, which rejected progressionist notions of linear development. This haphazard process occurred as each species became better adapted to its surrounding environment. Any environmental changes that occurred would affect the development of the species, and could cause species to diverge onto different branches of evolution, and in some cases to become extinct. The branching tree metaphor was considered early in the development of the theory of natural selection, as shown by a sketch in Darwin's *Notebook B* (**Figure 3**) and a detailed diagram was included in *The Origin* (**Figure 4**). Here the letters A–L represent parent species which resemble each other to varying degrees (thus spaced unevenly). The dotted lines diverging from each original species represent offspring from these species. Through the process of natural selection over many generations, some original species will develop new varieties and others will become extinct. New species varieties are represented at the top of the diagram (a14, q14, and so on). Darwin meticulously explained this diagram in chapter IV of *The Origin*, and the strength of his account is reinforced through the written description with which he concluded the chapter:

> … As buds give rise by growth to fresh buds, and these, if vigorous, branch out and overtop on all sides many a feebler branch, so by generation I believe it has been with the great Tree of Life, which fills with its dead and broken branches the crust of the earth, and covers the surface with its ever branching and beautiful ramifications (Darwin, 1859: 129 130).

The theory of natural selection did not occur in an ideological void but consolidated earlier debates on the development of species. However, despite Darwin's already well formulated ideas on human evolution – as reflected in his notebooks from the late 1830s – he largely avoided this issue in *The Origin.* Darwin eventually addressed human evolution in his 1871 text *The Descent of*

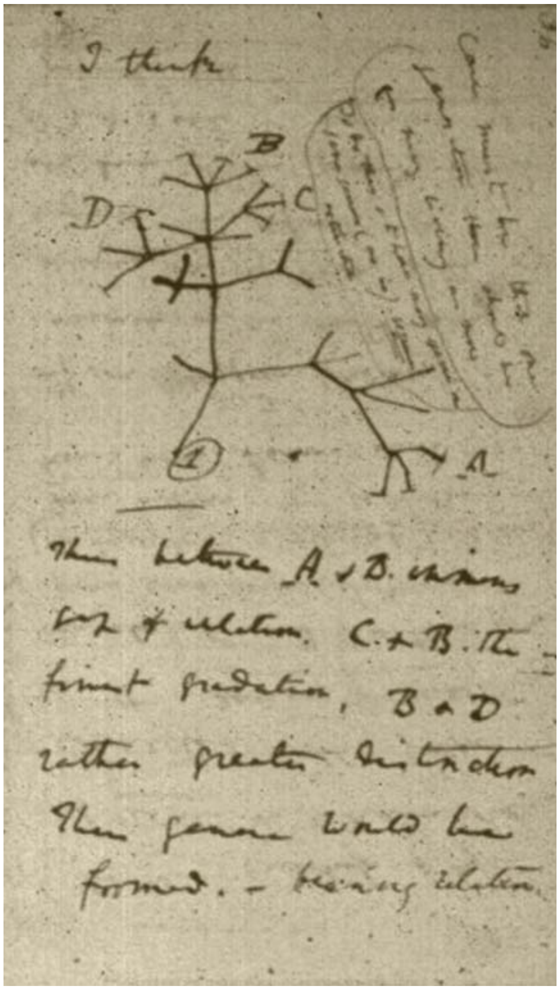

Figure 3 A page from Darwin's *Notebook B* showing his first evolutionary tree diagram, London 1837. From http://darwin-online.org.uk, The Complete Works of Charles Darwin Online.

Man and Selection in Relation to Sex by which time debates over human ancestry were well established.

Social Darwinism: Definition

Social Darwinism is commonly understood to mean the application of Darwin's biological theories to the social and cultural realm. In reality, sociocultural evolutionary theories developed in parallel to biological theories, rather than emerging from them. Social Darwinism is an umbrella term which was applied indiscriminately to a variety of social theories that emerged in the late nineteenth century, with often little resemblance to Darwin's original theories. Since science itself cannot easily be separated from influences within society, there is no clear set of concepts which can be defined as Social Darwinism. There are grounds to question whether the label named any actual social movement, or whether it was an abstraction created by historians and read into the past. The term was not commonly used until the turn of the century and has negative associations. Over the course of the twentieth century, Social Darwinism has been equated with racism, Nazism, and the eugenics movement, perhaps explaining why more recent engagement with evolutionism has been conspicuous by its absence within human geography.

Forms of Social Darwinism

Classical theories of social evolution coincided with the rise of Darwinian bioevolutionary theory and the two areas of inquiry influenced one another. The application of biological evolutionary theories to society naturalized the social order, and was, in many instances, used to legitimize class and racial hierarchies. Nineteenth century social evolutionary theories tended to incorporate ideas of progressionism carried over from the Enlightenment period, as shown in the work of British philosopher and social theorist Herbert Spencer (1820–1903) and of British anthropologist Edward Burnett Tylor (1832–1917).

Spencerianism is the set of theories most commonly equated with Social Darwinism, despite the fact that Spencer's work was heavily influenced by Lamarckian, and not Darwinian, evolutionary concepts. Spencer published his first book *Social Statistics* in 1851 in which he integrated aspects of Lamarckian progress with *laissez faire* economics, and developed the metaphor of the social organism. In his promotion of universal progressionism, Spencer linked strands of biological, psychological, and social evolution. These theories could be used both to describe the origin of racial differences, to account for deviations from the unilineal sequence of development, and to explain the evolution of higher mental faculties.

Spencer was responsible for coining the term 'survival of the fittest' which over time became linked to Darwinism. This was in part due to Darwin's incorporation of the term in the fifth edition of *The Origin*, published in 1869, where he changed the title of chapter IV from 'Natural selection' to 'Natural selection; or the survival of the fittest'. This move proved to be unfortunate and cleared the way for cruder forms of Social Darwinism which emphasized the exclusion of the weaker (or inferior) members of society in a competition for profit.

Edward Burnett Tylor's cultural evolutionary theory stressed a linear model of cultural development. Tylor argued that similarities between cultures in different areas of the world could be explained by independent invention and that cultures were forced into parallel channels of development by the need to follow a hierarchy of cultural stages. The existence of persistent relics (or survivals) from previous stages of advancement which

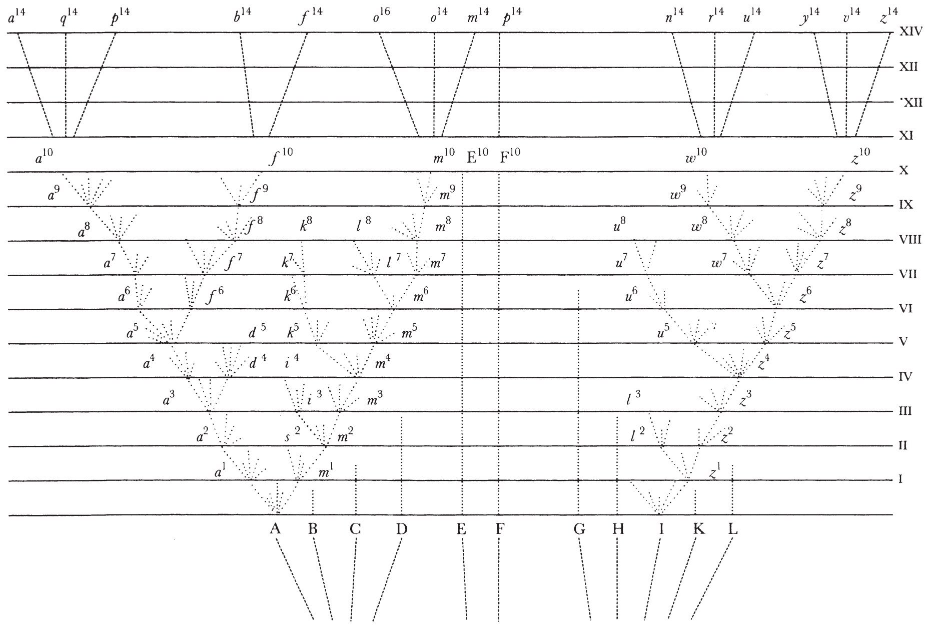

Figure 4 Branching tree diagram of species descent. Source: Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (1st edn.), 1st issue. London: John Murray.

no longer served a purpose was cited as evidence. Tylor was a monogenist, and for him the acceptance of linear progressionism meant that the idea of unity need not be abandoned. The ladder metaphor could be used to explain the existence of apparently inferior groups in biological and sociocultural terms. Tylor, like Darwin, originally denied links between the cultural progress and biological origins of the human race for fear of implying intellectual inferiority, but eventually both accepted these links, becoming more racialist in approach.

Geographies of Darwinism

The reception of Darwin's ideas varied between different places within the wider context of societal values, ideals, and practices. The reception of Darwin's theories in different geographical locations is explored in a 2005 paper by Livingstone, who has developed the idea of a geography of reading. Livingstone (2005:391) argues that "*where* scientific texts are read has an important bearing on *how* they are read." The reception of Darwin's bioevolutionary theories in different places is explored, focusing on Charleston Museum of Natural History (USA), Wellington Philosophical Society (New Zealand), and the St. Petersburg Society of Naturalists (Russia).

Scientists based at Charlestown Museum of Natural History developed a style of Southern science based on political ideology which incorporated polygenist principles and legitimated slavery. In this political environment Darwin's theories of species transmutation were regarded as a threat to Southern racial and religious culture. In Wellington, Darwin's theories were also read from a racialist perspective, but here they were interpreted as supporting the success of white colonialism. A set of public lectures in 1868–69, held at the Colonial Museum in Wellington, introduced Darwin's ideas to the public. This included an address by politician William Travers, who brought to Darwin's text the long standing colonial conviction that the Maori were fated for extinction by nature. In St. Petersburg, an 1879 essay by Karl Kessler titled 'On the law of mutual aid' condemned Darwin's law of the struggle for existence, and stressed the importance of mutual aid. Kessler's stress on mutual aid was later vigorously promoted by Peter Kropotkin. These examples illustrate how the politics of place shaped the ways in which Darwinian theory was read and understood by nineteenth century academics.

Darwinism in Geography

Geography, Evolutionary Theory, and Imperialism

Evolutionary perspectives coupled with the imperialism of the age had a lasting effect on the emerging discipline of geography. Evolutionary theories were translated into geography in large part through environmental determinism, which can be perceived as a version of Social Darwinism. Geographical perspectives promoted overseas expansion and included assumptions about the moral, intellectual, and biological superiority of white Europeans in comparison to other racial groups.

Most geographers accepted the organic metaphor and utilized it in different ways. The organism analogy was one thread of Darwinian theory but was also important in neo Lamarckian biology. This concept played a formative role in the work of geographers Freidrich Ratzel (1844–1904) in Germany, William Morris Davis (1850–1934) in the USA, and Halford Mackinder (1861–1947) in Britain.

The relationship between geography, imperialism, and evolutionary theories is evident in the work of Ratzel who had the twin concerns of scientizing geography and putting it into the service of imperialism. Ratzel's 1881 text *Anthropogeographie* illustrated his version of human geography through the main themes of migration, isolation, space, and environmental determinism (themes that had previously been identified in the work of Moritz Wagner, who stressed the importance of migration in species variation). In *Anthropogeographie*, Ratzel introduced the concepts of *Lebensraum* (living space) and *Grossraum* (large space). Building on Spencerian notions of society as organism and Malthusian notions of population expansion, Ratzel's stress on *Lebensraum* developed the idea of the state as organism. Ratzel argued that population enlargement would reach the point where territorial expansion was inevitable. The state would eventually aspire to *Grossraum*. This was seen as a natural evolutionary process, incorporating concepts of competition and survival (of the fittest state). Thus, the late nineteenth century European expansion into Africa was legitimated as a natural process.

Davis applied evolutionary theories to landscapes through his development of the cycle of erosion, first presented in 1889. The cycle began with an uplift of the land, which would gradually wear back to a peneplain, after going through a number of sequential forms. The influence of evolutionary theory was reflected in his use of biological terms, including birth, adolescence, maturity, and death. Davis' work incorporated the Darwinian stress on the importance of time in evolutionary development, although his work also contained neo Lamarckian elements.

Mackinder conceptualized the discipline of geography as a means of encouraging British people from a young age to think imperially. The influences of imperialism and evolutionary theory were apparent in his 'Heartland Thesis', a concept first outlined in a paper on 'The geographical pivot of history', presented at the Royal Geographical Society (RGS) in January 1904. This theory stressed the advancement of civilization and reflected Ratzel's concept of *Lebensraum*. The heartland was conceived as a land locked area in Eurasia, where a power base could be created. In Mackinder's imperial approach, the exercise of world power was vital if the world organism was to grow democratically. The discourse of race was a central element of his theorizing – if the British race failed in its domination of the globe then another race would become the world power. His belief in environmental determinism led him to view the natural environment as the stage for the struggle of the races and, as in Ratzel's geography, imperial expansion was naturalized.

Racial Classification and Mapping

Modern understandings of race have their roots in the Enlightenment period. Kant, for example, stressed the links between distance from the equator and skin color, while the concept of 'The Great Chain of Being' influenced later ideas of racial hierarchy. From the 1850s in Britain, evolutionary theories and debates over the origin and unity of the human species shaped scientific thinking. Allied to these developments was the idea that a discrete number of racial types could be delineated, and the measurement of racial traits dominated physical anthropology. Previous assumptions about racial superiority and inferiority influenced racial classification systems. For example, since the Enlightenment it had been assumed that Australian aborigines were a doomed race that had not progressed far up the ladder of civilization and would soon become extinct. Late nineteenth and early twentieth century racial taxonomies generally placed nonwhite groups further down the ladder of evolution, with white Europeans perceived as superior and fitter in biological, social, and intellectual terms.

Working at the intersection of anthropology and geography a number of late nineteenth and early twentieth century academics classified and mapped human racial groups. These anthropogeographers were strongly influenced by evolutionary theories (both Darwinian and neo Lamarckian), environmental determinism, and systems of measurement derived from anthropology.

In his 1899 volume *The Races of Europe*, American economist and anthropogeographer William Zebina Ripley (1867–1941) suggested a tripartite racial scheme for Europe (Teuton, Alpine, and Mediterranean), which combined a reliance on anthropometric measurement with the assumption that racial characteristics had been molded by environmental conditions (for example, he mapped the increase in pigmentation moving South through Europe). Ripley's work was later utilized by anti immigration supporters in the US in attempts to prevent migration from Southern and Eastern Europe (Ripley's Mediterranean category), groups who became perceived as inferior to Northern Europeans. The prominent Australian geographer, Griffith Taylor (1880–1963) attempted to delineate discrete racial groups through his zones and strata theory (developed in 1919), where the migration and settlement of racial groups was crudely equated to geological formations at a world scale.

His model combined an emphasis on migration, on en vironmental determinism, and on heredity of particular racial characteristics, especially cephalic index. Taylor divided the world into six racial categories and in corporated preexisting assumptions on development. Australian aborigines and black Africans were ranked in the lower layers of strata, while Europeans and Asians were ranked more favorably. A strongly deterministic approach to the evolution of race was clear in the work of Ellsworth Huntington, who, in 1915, produced world maps purporting to illustrate the relationship between climate and civilization. Huntington argued that climate had enervated those in the Northern latitudes, while the tropics had debilitating effects on development. Not surprisingly, given his original assumptions, the areas he defined as having the best climate overlapped with those defined as the most civilized and occurred in Europe and other areas of successful white settlement.

Other Applications of Evolutionary Metaphors

Urban ecology

Evolutionary metaphors were applied to studies of urban development as evidenced by the work of the Chicago School which embarked on a major program of research on the American city in the early to mid twentieth century. Urban ecology was central to the work of Robert Park and stressed the idea of society as organism, with each part being symbiotically related to all others in a web of competitive and cooperative behavior (incorpor ating Darwin's stress on these two elements). Central to Park's analysis of the city was the idea that humans would naturally be channeled into different groups which would find residential and occupational niches in society. Burgess' 1920s concentric zone model of urban devel opment incorporated the idea of invasion and succession based on the assumption that cities, as organisms, must grow, and that individuals within the economy would be upwardly mobile. Urban ecology has been criticized for emphasizing the supposedly natural development of cities and ignoring social factors, such as discrimination and social exclusion. The urban models were also seen as endorsing the modern capitalist economic system.

Capitalism

Organic metaphors have been associated with the de velopment of the capitalist economy since the late eighteenth century (predating Darwin). Commentators, including Marx, have observed that Darwin must have been affected by the era of competitive capitalism in which he lived, as reflected in his development of con cepts such as competition and the struggle for existence. In the nineteenth century ideas of biological organisms fighting for survival became linked to ideas of natural dominance and in economic terms this translated as the dominance of capitalism and the capitalist class.

Gibson Graham, in her 1996 book, argues that the ladder metaphor of development (stemming from the Enlightenment and reflected in Lamarckian theory) has been applied to the transition from pre Fordism to Fordism to post Fordism, regarded as an improvement in the capitalist organism. This model, however, fails to recognize the diverse forms of capitalism that may have developed or the ways that economic development occurs unevenly over economic landscapes. Gibson Graham argues that the idea of an evolutionary bush is a more informative way of regarding changes in both capitalist and noncapitalist (domestic) forms of economy (**Figure 5**), reflecting Darwin's branching tree analogy.

Darwinism in the Current Landscape of Discipline

In the biological sciences the modern synthesis emerged between 1920 and 1950 and involved the combination of

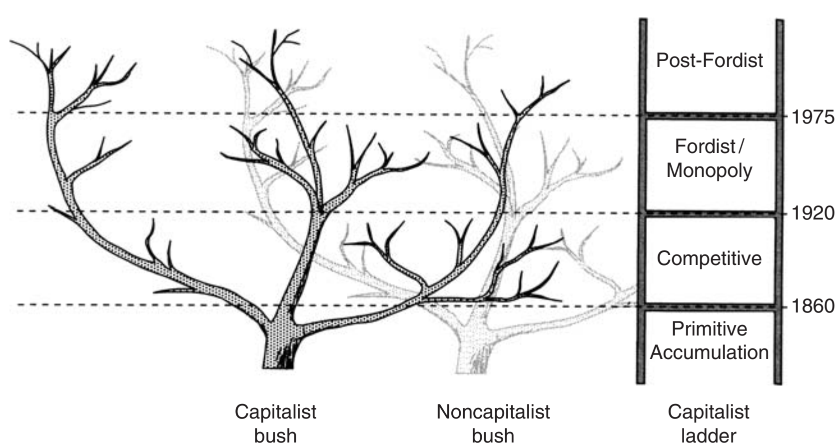

Figure 5 Branching trees and ladders of capitalism. From Gibson-Graham, J. K. (1996). *The End of Capitalism (As We Knew It): A Feminist Critique of Political Economy*. Oxford: Blackwell.

Darwinian biology and Mendelian genetics. The discovery of the structure of DNA in 1953 further strengthened the modern synthesis and validated key components of Darwin's theory. An engagement with evolutionary theory is less developed in the social sciences, especially within human geography. There remains an unease to engage with evolutionary theories due to the previous uncritical application of biological theories to the social realm, which resulted in a number of problems, including links between Social Darwinism and Nazism, social antagonisms created by sociobiology, and the eugenics movement.

Geography as a discipline has historically been conceived as incorporating both studies of nature and of humanity. Rather than being fully integrated, a divide emerged between the human geographical focus on society and physical geographical studies of the natural environment. Recent posthumanist perspectives in human geography are beginning to address the wider connections between evolution, adaptation, and environment–organism relationships. Geographers such as Harvey, Whatmore, and Thrift have noticed an artificial break between human society and nature. Thrift considers the relationships between humans, animals, and technological developments (such as tool use and the development of the human hand, and the idea of writing as technology) and argues that vital spaces exist where a network (rather than a tree) of biological and technological interaction occurs. In his 1996 text *Justice, Nature and the Geography of Difference*, Harvey notes the importance of developing a language which would incorporate evolutionary ideas into societal perspectives within a positive ecological framework. He draws attention to four aspects of interactions between living organisms, which integrate aspects of Darwinian theory. These four relational categories are: first, the competition and struggle for existence; second, the adaptation and diversification into environmental niches; third, collaboration, cooperation, and mutual aid; and fourth, environmental transformation. Constructive reengagement with biological and ecological theory by human geographers may lead to a closer understanding of the interrelationships between humanity and nature and a greater respect for, and understanding of, our environment.

See also: Anthropogeography (After Ratzel); Determinism/Environmental Determinism; Indigenous Geographies; Lamarck(ian)ism; Mackinder, H. J.; Mapping, Race and Ethnicity; Taylor, G.

Further Reading

Beer, G. (1988). Introduction. In Darwin, C. (ed.) (1859) The *Origin of Species,* (2nd edn.) pp vii xxviii. Oxford: Oxford University Press.
Bowler, P. J. (1986). *Theories of Human Evolution*. Oxford: Blackwell.
Browne, J. (1995). *Charles Darwin: Voyaging*, vol. 1. London: Jonathan Cape.
Burrow, J. W. (1966). *Evolution and Society; a Study in Victorian Social Theory*. Cambridge: Cambridge University Press.
Carroll, J. (2000). Introduction. In Darwin, C. (ed.) (1859) *The Origin of Species* (1st edn.), reprinted, pp 9 72. Peterborough: Broadview Texts.
Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, (1st edn.) 1st issue. London: John Murray.
Desmond, A. and Moore, J. (1991). *Darwin*. London: Michael Joseph.
Fitzsimmons, M. (2004). Engaging ecologies. In Cloke, P., Crang, P. & Goodwin, M. (eds.) *Envisioning Human Geographies*, pp 30 47. London: Arnold.
Gibson Graham, J. K. (1996). *The End of Capitalism (As We Knew It): A Feminist Critique of Political Economy*. Oxford: Blackwell.
Harvey, D. (1996). *Justice, Nature and the Geography of Difference*. Oxford: Blackwell.
Livingstone, D. N. (1992). *The Geographical Tradition: Episodes in the History of a Contested Enterprise*. Oxford: Blackwell.
Livingstone, D. N. (1994). Science and religion: Foreword to the historical geography of an encounter. *Journal of Historical Geography* 20(4), 367 383.
Livingstone, D. N. (2005). Science, text and space: Thoughts on the geography of reading. *Transactions of the Institute of British Geographers, NS* 30(4), 391 401.
Mayr, E. (1985). Darwin's five theories of evolution. In Kohn, D. (ed.) *The Darwinian Heritage*, pp 755 772. Princeton, NJ: Princeton University Press.
Stoddart, D. R. (1966). Darwin's impact on geography. *Annals of the Association of American Geographers* 56, 683 698.
Thrift, N. (2005). From born to made: Technology, biology and space. *Transactions of the Institute of British Geographers, NS* 30(4), 463 476.
Whatmore, S. (2002). *Hybrid Geographies: Natures Cultures Spaces*. London: Sage.
Winlow, H. (2006). Mapping moral geographies: W.Z. Ripley's races of Europe and the United States. *Annals of the Association of American Geographers* 96(1), 119 141.

Relevant Websites

http://www.lamarck.cnrs.fr
Much of Lamarck's original writing is free online.
http://darwin online.org.uk
The Complete Works of Charles Darwin Online.

Dear, M. J.

A. Latham, University College London, London, UK

Micheal J. Dear (**Figure 1**) has, at various times, been an urban geographer, homelessness researcher, postmodern theorist, and central member of the Los Angeles School of urban studies. Co author of two seminal monographs on mental healthcare provision and urban homelessness in North America, Dear is nonetheless most well known for his more recent work on the urban dynamics of Los Angeles and his attempts to develop a postmodern style of geographical analysis. His central argument is that Los Angeles – long thought of as a city defined by its singularity – has come to define the paradoxes and pos sibilities of the contemporary metropolis. Drawing on continental European philosophy, neo Marxism, and mainstream social science, in combination with techni ques from the humanities and creative arts, Dear's often experimental, and highly distinctive, essays on postmodern Los Angeles have – along with those of Ed Soja – mapped out the critical edge of postmodern human geography.

If Dear's recent – and most influential – work is characterized by its experimental and *avant garde* feel, his early academic trajectory hardly pointed in this direction. Born in Treorchy, South Wales, in 1944, Dear read for his undergraduate degree in geography at the University of Birmingham, before completing an MPhil in town and country planning at University College London in 1969. He then moved to the University of Pennsylvania to pursue doctoral work. He was awarded his doctorate in 1974 for a dissertation 'Locational analysis for public mental health facilities', at which time he took up a teaching appointment at McMaster University in On tario, Canada. Dear's early publications focused – like his PhD dissertation – on the geography of mental health care provision. And, like his dissertation, they had a strong regional science feel (not surprising given the University of Pennsylvania's strength in this discipline at the time).

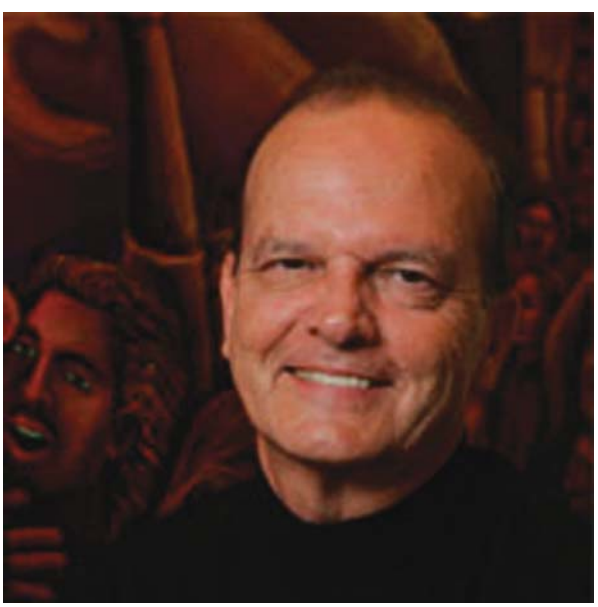

Figure 1 Micheal J. Dear.

Notwithstanding his methods, however, Dear's work on the provision of mental healthcare facilities did have a critical edge. His publications highlighted the many inadequacies of then current mental health policies, in cluding the remarkably uneven geographic distribution of community healthcare facilities in American and Canadian cities, the overwhelming tendency of such community facilities to be clustered in the poorest and most socially disorganized parts of cities, and the lack of objective locational criteria used by policymakers in their decisions as to where to locate mental healthcare facilities. Dear's early research also highlighted the emergence of mentally disabled 'ghettos' in the inner city of many North American cities, as mentally ill indi viduals who had formerly been warehoused in residential asylums were released into 'community care' in 1960s and 1970s as part of a general policy shift toward the deinstitutionalization of mental healthcare provision. This "ghettoization" (Dear 1977, page 588) of service dependent mental health patients was in part simply a product of the low cost and accessibility of the housing in these areas. But as more and more community mental health facilities came to be located in inner city neighborhoods – because of low real estate costs, the availability of large, formerly upper middle class houses easily converted into hostels or treatment centers, the presence of restrictive zoning restrictions in other neighborhoods, the lack of organized community opposition – such neighborhoods also began to attract deinstitutionalized mental health patients simply because of the density and extent of care provision available. In effect, then, and as an unintended result of the general policy shift toward the deinstitutionalization of mental healthcare provision, areas of North American cities were "becoming [in effect] the location of an asylum without walls" (Dear 1977, page 588).

As Dear extended his analysis of mental healthcare provision in the late 1970s and 1980s, his work developed an increasingly critical and political inflection. For one, his work began to explore the links between the inner city 'ghettoization' of mental health patients, and the precipitous rise of homelessness and rough sleeping that was simultaneously occurring in Canadian and American cities. For another, to understand the social, economic, and political dynamics that were driving the patterns of mental healthcare provision (or indeed lack of it) his work was documenting, Dear began to develop an increasingly sophisticated interest in political theory, critical social theory, and political economy. The patterns of welfare neglect, self interested neighborhood pro tection, and local and national policy development, that were driving many of the trends in mental health pro vision required an in depth consideration of the insti tutional structures that brought these trends into being if they were to be explained with any intellectual rigor. This in turn required careful consideration of what exactly the state is and how it – and its various agents – defines and governs territory. Thus, in *State Apparatus: Structures of Languages of Legitimacy*, co authored with Gordon Clark, Dear turned to examining "how the state and the state apparatus structure the sociospatial organization of capitalist society" (Clark and Dear 1984, page 1). From the start Clark and Dear highlighted the problems they saw with liberal and pluralist accounts of the modern state – their inability to recognize the inequalities that capitalism is based upon, nor to fully account for the extent to which the state apparatus reaches into evermore dimensions of economic and social life. Nonetheless, while Clark and Dear's analysis was heavily influenced by the radical critique of mainstream social science (and especially the mainstream liberal politics that Marxist geographers like David Harvey, Richard Peet, and Neil Smith had been developing since the early 1970s), they stopped short of endorsing a Marxist theory of the capitalist state. Instead, they con cluded (not entirely convincingly) that the primary goal of the state – and the only one that could ensure its legitimacy – should be ensuring "social justice" (Clark and Dear, page 191).

State Apparatus was a densely argued and largely theoretically oriented text. However, many of its themes are reprised in more empirically grounded form in Dear's two following monographs on homelessness and mental healthcare: *Landscapes of Despair: From Deinstitutionalisation to Homelessness* and *Malign Neglect: Homelessness in an American City*, both co written with Jennifer Wolch. *Landscapes of Despair* is an in depth analysis of the health and welfare policies that had led to the mental health 'ghettoization' described in Dear's earlier paper. But as well as mapping and describing the emergence of areas dominated by mental healthcare users as his earlier work had done, *Landscapes of Despair* also presents a detailed historical geography of the political and economic dynamics which had shaped mental healthcare policies. As the title suggests, Dear and Wolch also explicitly linked the deinstitutionalization of mental healthcare provision with the dramatic rise in urban homelessness that many North American cities experienced in the 1980s.

As with the emergence of the mental healthcare 'ghetto', which Dear and Wolch now called the zone of dependence, the rise of homelessness was not simply a straightforward product of deinstitutionalization. Rather, it was produced through the interaction of a range of often unexpected forces – cuts in welfare provision, social housing policies, the decline in hostels offering single rented rooms, economic restructuring which destroyed many low paid, low skill jobs, to name just a few of the more prominent causes. Moving beyond these causes, *Landscapes of Despair* also explored how the service dependent ghetto functioned; despite its haphazard constriction it often offered a genuinely supportive environment for those housed within it, as well as gen erating a kind of shadow state of charity and voluntary organizations funded by state money but not publicly accountable to any electorate. This was a set of themes reprised in more detail in *Malign Neglect* which, unlike *Landscapes of Despair*, focused on a single city, Los Angeles (described in the 1980s and early 1990s as the 'homeless capital of America'). However, *Malign Neglect* also examined the emergence of a 'proto homeless' popu lation, a stratum of people that stretched well beyond the mentally ill and disabled who because of the decline in welfare provision, a lack of social housing provision, and the precariousness of their economic position, were on the edge of falling out of the housing market.

Landscapes of Despair and *Malign Neglect* both were – and indeed remain – widely admired, widely read, and frequently cited texts. In many ways they still define the central terrain of contemporary geographical debates about homelessness, vagrancy, and mental health pro vision. With these two books Dear established himself not only as a preeminent figure in health and welfare geography, but also as one of small number of high profile scholars defining the frontier of possibility for a newly theoretically literate human geography. Given the as suredness of Dear's arguments, and the degree which they are quite clearly rooted in the broader shift within Anglo American human geography toward the social sciences, it is perhaps surprising that just months after the publication of *Landscapes of Despair*, Dear was declaring in an article in *The Transactions of the Institute of British Geographers* that "this is a time of intellectual crisis" (Dear 1988, pages 262), and that human geography was caught up within this crisis. "[Contemporary] geography," he wrote, "is an eclectic mish mash of old and new, per tinent and irrelevant, the quirky and incomplete. […] There seems to be no consensus on what human geog raphy is, or what are its objectives in analysis."

This seems a strange argument for someone who has just shown the benefits of human geography's engagement with the social sciences in general, and social theory in particular. But Dear's intervention is not nearly as odd as it first appears. Dear worried that human geography's engagement with the social sciences had not gone far enough. While much of Anglo American human geography had moved toward the social sciences, the discipline remained riven by a labyrinth of subdisciplinary rivalries and divisions. Each subdiscipline defined its own engagement with wider social scientific and social theoretical debates, and there was little ground for debate within the discipline on how these different perspectives might relate to each other. Thus, while a range of cognate social science disciplines such as anthropology, sociology, and history were experiencing a so called 'spatial turn' – discovering that space did indeed matter – geography remained trapped by its own parochialism, unable to reap the benefits of this sudden interest in the discipline's core focus, spatiality. Dear argued that to escape this parochialism geography needed to complete the reinvention of the discipline that the initial discovery of social theory in the 1960s and 1970s had begun.

How to bring about this reinvention? In Dear's view, what was required was not the reassertion of a singular vision of human geography – for example, a return to positivism, or a total embrace of Marxism, or political economy, or post structuralism. What was instead needed was some shared sense of what human geography's core concerns were, some common basis for discussion across differing approaches to undertaking geographical research. Turning to a series of ongoing debates in philosophy, the creative arts, and some parts of the social sciences, Dear argued that what geography needed to do was engage with its diversity through rethinking itself as a postmodern discipline. Postmodernism is defined through its assertion that there is no foundational position from which knowledge claims could be judged. The point (at least in Dear's reading) is not that all knowledge claims are equal, but that there is no one theory that can claim a privileged position. Thus, "the postmodern challenge is to face up to the relativism of human knowledge, and to proceed form this position to a better understanding" (Dear 1988, page 271).

Following on from this argument, and drawing inspiration from his move in 1986 from McMaster University to a chair in geography at the University of Southern California, where in 1995 he became the founding director of The Southern California Studies Center, Dear's writing began to focus almost exclusively on articulating what a postmodern geography would involve. In particular, in concert with a number of other Californian based urban scholars such as Ed Soja, Micheal Storper, Allen Scott, Mike Davis, and Jennifer Wolch, Dear connected his call for a postmodern geography with the claim that the Los Angeles metropolitan region offered a 'prototypical' example of an emergent form of postmodern urbanism. Dear's conviction was that LA with its sprawling, decentralized form, its hyperautomobility, its ethnically diverse population, its rich mix of cutting edge technology and media industries, classic Fordist manufacturing industries, and all sorts of low technology, low wage, industrial sectors, and its complex international connections, in some sense defined the postmodern moment. So, where "Los Angeles [had] almost always been viewed as an exception to the mainstream of American urban culture" (Hise *et al.* 1996, page 1), for Dear and his fellow Los Angeles School theorists LA had became the place where things 'all came together'.

This was – and remains – a rather grand claim. In a series of edited collections and a single authored book, *The Postmodern Urban Condition*, Dear has tried to demonstrate the productiveness of such an approach. A key dimension to this is Dear's experimentation with narrative, and exploration of novel forms of argumentation and representation. Thus, like his fellow Los Angeles School colleagues, Ed Soja, and Mike Davis, Dear's analysis is driven by great flocks of neologisms: 'flexism' (capitalism organized through ultra responsive supply chains); 'keno capitalism' (seemingly random patterns of urban land use and development); 'cyburbia' (suburbs that highly integrated into national and international electronic communication networks); 'commudies' (communities based on a pre packed lifestyle); 'cyberia' (areas without access to electronic communications like the Internet); 'cybergeogeoisie' (the privileged elites who run the global economic system); 'protosurps' (the unskilled, marginalized, and disposable labor that allows flexism to function); 'holsteinization' (the 'monoculturing of consumers' to make them better targets for marketing); and mimetic contagion (the spontaneous crossover of cultural practices from one group to another). Indeed, in places Dear's writing reads more like a novel by Phillip K. Dick or William Gibson than it does serious academic prose.

But while critics such as Richard Symanski, Peter Jackson, Robert Lake, Robert Beauregard, and Donald McNeill have found Dear's approach exaggerated, self indulgent, or simply pointless, there is a method to all these superlatives and hyperbole. They are meant to shake the reader (and indeed the writer) out of established patterns of thought. The hyperbolic, exaggerated, prose employed throughout much of Dear's writing on postmodern urbanism – most notoriously in his *The Postmodern Urban Condition* – is meant to "encourage new ways of seeing" (Dear 2003, page 202). Dear's writing does do this. Even if readers are often unconvinced by Dear's more outlandish claims, they are forced to consider why they disagree with what he is saying. This, perhaps, is the most significant achievement of Dear's postmodern urbanism. It is not so much that it has pre sented a new, definitive, way of doing urban geography. Rather his postmodern urbanism has opened up an exciting terrain through which novel and challenging forms of urban analysis might be legitimately explored and experimented with.

See also: Health Geography; Homelessness; Los Angeles School of Post-Modern Urbanism; Postmodernism/ Postmodern Geography; Poststructuralism/ Poststructuralist Geographies; Soja, E.

Further Reading

Clark, G. and Dear, M. (1984). *State Apparatus: Structures of Languages of Legitimacy*. Boston, MA: Allen Unwin.

Dear, M. (1977). Psychiatric patients and the inner city. *Annals of the Association of American Geographers* 67(4), 588 594.

Dear, M. (1987). The postmodern challenge: Re constructing human geography. *Transactions of the Institute of British Geographers* 13, 262 274.

Dear, M. (2000). *The Postmodern Urban Condition*. Oxford: Blackwell.

Dear, M. (ed.) (2002). *From Chicago to L.A.: Making Sense of Urban Theory*. Thousand Oakes, CA: Sage.

Dear, M. (2003). Superlative urbanisms: The necessity of rhetoric in social theory. *City and Community* 2(3), 201 204.

Dear, M. and Flusty, S. (1998). Postmodern urbanism. *Annals of the Association of American Geographers* 88(1), 50 72.

Dear, M. and Leclerc, G. (eds.) (2003). *Postborder City: Cultural Spaces of Bajalta California*. New York: Routledge.

Dear, M. and Wolch, J. (1987). *Landscapes of Despair: From Deinstitutionalisation to Homelessness*. Princeton, NJ: Princeton University Press.

Leclerc, G., Villa, R. and Dear, M. (eds.) (1999). *La Vida Latina en L.A.: Urban Latino Cultures*. Thousand Oakes, CA: Sage.

Wolch, J. and Dear, M. (1993). *Malign Neglect: Homelessness in an American City*. San Francisco, CA: Jossey Bass.

Debt

David Simon, University of London, Egham, UK

Glossary

Club of London The group of creditor commercial banks and other financial institutions that negotiates collective policies in respect of indebted countries and repayments.

Club of Paris The group of official (governmental and intergovernmental) creditor institutions (e.g., government aid agencies, the regional development banks, World Bank, and IMF) that negotiates collective policies in respect of indebted countries and repayments.

Countertrade A more sophisticated form of barter involving bridging finance to enable the traded commodities to be exchanged at different times.

Debt–Equity Swaps A debt reduction mechanism whereby commercial creditors accept equity stakes in debtor state bonds or enterprises, usually at a discount, in lieu of actual debt repayments.

Debt for Nature Swaps Similar to debt–equity swaps except that the debtor invests in environmental conservation, for example, national parks, biodiversity reserves, pro rata for the value of debt written off.

Debt–Export Ratio The total external debt as a proportion of the value of exports of goods and services.

Debt–Interest Ratio Debt interest payments as a proportion of the value of exports.

Debt Rescheduling A renegotiation of the original terms and conditions of loan repayment to a creditor, usually involving an extension of the grace and repayment periods in return for imposition of a higher interest rate.

Debt–Service Ratio Total debt service burden as a proportion of the value of exports.

Long-Term Debt Debt with a maturity of more than one year.

Poverty Reduction Strategy Papers (PRSPs) Successor to structural adjustment and economic recovery programs introduced in 2000–01 by the IMF and World Bank to refocus debtor governments' efforts and donor budgetary support on poverty reduction in line with the Millennium Development Goals and to ensure civil society participation in their production.

Total External Debt Debt owed to non-residents and payable in foreign currency, goods, or services.

Introduction

Few episodes have so profoundly affected the global political economy over recent decades as the onset in 1980–81 of the so called 'debt crisis' which destabilized, and for a time threatened the very survival of, the entire global trading and financial system. Its impact still dominates key national and international policy arenas, affecting both governmentality and – especially in the more heavily indebted countries – processes of govern ance. While the crisis erupted into the news headlines when Mexico in 1982, and then Brazil and Argentina, announced their inability to service their accumulated debt, it is important to understand the origins of the crisis in structural and historical terms. Such an explanation also provides an appreciation of the limitations of most solutions proposed over the intervening years. Ulti mately, tackling the problem will require more funda mental measures to address the underlying inequalities that produced it.

Although inequalities in the global trade regime were longstanding – with origins dating back to the European imperial and colonial eras but especially grandiose modernization strategies after independence – the more immediate roots of the crisis lie in Middle Eastern geo politics. During the Arab–Israeli war of late 1973, the key oil producing Arab states retaliated against Western support for Israel, and sought to raise additional revenues for their war effort, by means of a dramatic increase in the hitherto low producer price of crude oil. At the time, international commodity trade occurred mainly in terms of fixed prices set by producers. The success of this stratagem resulted in a second price increase in 1979, the combined effect of which raised prevailing prices more than sixfold over early 1973.

The shockwaves of so dramatic a price increase of this essential commodity, on which most modern eco nomic activity depends, rapidly ended the period of unprecedented international economic growth and opti mism that had prevailed since the mid 1960s and trig gered a severe global recession. Efforts to reduce oil consumption impacted upon industrial output, which in turn depressed the hitherto buoyant prices of other pri mary commodities that formed the bulk of poor country exports. Although the impact everywhere was substantial, among the worst hit were non oil producing poor countries, which suffered the triple shock of falling ex port prices, rapidly increasing import prices for oil and industrial commodities, and a rapid decline in official development assistance (ODA) as donor governments cut back as part of their overall adjustment to domestic recession.

Conversely, the large oil producers were unable to absorb the vast flows of 'petrodollars', as oil revenues came to be known, into their often small modern economic sectors. Hence many billions of surplus dollars were invested abroad, in European and North American real estate and in the international banking system. Unable to offload all these funds in recession hit markets of the global North, the banks found a ready market with governments and parastatal corporations in the newly industrializing and other lower middle income countries eager to replace their dwindling ODA inflows in order to complete the many flagship large scale, capital intensive industrial and infrastructural projects being undertaken as part of their economic modern ization drives. Indeed, many saw the excess international financial liquidity as an opportunity to obtain funds to accelerate their ambitious plans. This process greatly accelerated the historical shift away from grant based ODA toward loans, increasingly from commercial sour ces. Although 36% of total Latin American debt had been with official sources and 64% with private sources in 1970, by 1981 the respective percentages were 12 and 88.

Crucially, real interest rates (i.e., rates that have been discounted for inflation) were at historically low levels when these new commercial loans – mostly with a maturity period of five or more years as a reflection of the still optimistic expectations in borrowing countries – were negotiated. However, despite the prevailing practice of arranging fixed interest loans, the lenders insisted on using variable interest rates in order to compensate for higher perceived risks. Apart from their perceptions of borrowers' ability to afford the higher interest rates, banks favored countries with the best growth records or prospects, and perceived political stability. As the global recession deepened rapidly, interest rates climbed steeply, reaching historically high levels within the life of these large loans. This imposed a rapidly increasing debt service burden on the borrowers, and, when the loans matured in large numbers toward the end of the 1970s, repayment of the principal became impossible. Even where the projects had been completed successfully, the 'triple whammy' referred to above meant that public sector balances of payments were in a parlous state. Increased borrowing became necessary to offset balance of payments deficits and to repay earlier loans. When the three largest debtors, Mexico, Brazil, and Argentina, declared their inability to continue servicing their debt in the early 1980s, the 'debt crisis' broke.

Anatomy of Debt

The essence of the crisis has not been the absolute value of the external debt as such, but its scale relative to other economic indicators which determine the ability of a country to service that debt. The crisis deepened and broadened rapidly during the 1980s, with even more countries defaulting or threatening to default on one or both components of debt service, namely interest pay ments and repayment of the 'principal', that is, the capital loan. Poland was the first European country to reach crisis point, in 1981 – even before Mexico – but as an apparently isolated case it did not have the same global impact, despite the severity of the problem domestically. By the end of 1985, Mexico and Brazil each owed over US$100 billion and Argentina some US$50 billion. Latin American debt was fairly evenly split between official and commercial sources, whereas in sub Saharan Africa the debt was substantially to official sources. By early 1986, total Third World debt reached US$1 trillion (1000 billion) and continued to rise. However, not all countries and regions fared equally badly. The harsh counter measures enforced and adopted in Latin America, for example, led to a reduction in overall indebtedness by 1990 on two of the three conventional indicators (debt–export ratio, debt service ratio, and debt interest ratio), whereas sub Saharan Africa, South Asia, and the Middle East and North Africa saw their aggregate indebtedness rise sharply during the 1980s, despite lower dollar values of debt. The precise extent of these trend changes depends on which of the different, but related, variables is used (see **Table 1**). The crisis in Latin America was so severe in terms of its overall developmental and social impacts that the 1980s became known as the Lost Decade. Debt–export ratios of over 150% and debt–service ratios above 15–20% are generally regarded as unsustainable in the medium or long term because they displace productive investment and expenditure.

Over the period 1990–2003, the total external in debtedness of poorer countries continued to rise, reaching over US$2.5 trillion, an increase of over 91% (**Table 2**). However, the increase for low income countries was a modest 24.3% at current prices (representing a decrease in real terms), whereas, for middle income countries as a whole, and for both sub categories, debt more than doubled at current prices and in 2003 accounted for 84% of the total for both categories. Geographically, the increases ranged from a mere 13.8% in the Middle East and North Africa, and 30.8% in sub Saharan Africa, to 124.5% in East Asia and the Pacific and a massive 211.2% in Europe and Central Asia (**Figure 1**).

The largest individual debtor countries in 2003 were (all in billions of US dollars) Brazil (235.4; 330%), China (193.6; 48%), Russian Federation (175.3; 135%), Argentina (166.2; 533%), Turkey (145.7; 251%), Mexico (140; 88%), Indonesia (134.4; 204%), and India (113.5; 123%), where the percentages represent the respective debt–export ratios. These included the world's most populous and among the economically fastest growing

Table 1 External debt burden, 1990 (and 1980)

Region/Country	*Total external debt as percentage of exports of goods and services*	*Total debt service as percentage of exports*	*Interest payments as percentage of exports*
Sub-Saharan Africa	324.3 (96.8)	19.3 (10.9)	8.9 (5.7)
Low-income countries	218.5 (105.1)	20.1 (10.3)	9.3 (5.1)
Ethiopia	480.3 (136.2)	33.0 (7.6)	8.1 (4.7)
Somalia	2576.2 (252.0)	11.7 (4.9)	5.8 (0.9)
Lower middle-income countries	179.0 (115.2)	20.3 (18.8)	8.4 (9.1)
Côte d'Ivoire	487.4 (160.7)	38.6 (28.3)	13.3 (13.0)
Upper middle-income countries	132.1 (159.6)	17.9 (31.0)	8.2 (16.6)
East Asia and Pacific	91.1 (88.8)	14.6 (13.5)	5.8 (7.7)
Korea, Republic	130.6 (44.0)	19.7 (10.7)	12.7 (3.5)
South Asia	281.5 (162.9)	25.9 (12.2)	13.1 (5.2)
Bangladesh	448.2 (345.6)	25.4 (23.2)	7.7 (6.4)
India	282.4 (136.0)	28.8 (9.3)	15.9 (4.2)
Europe	125.7 (90.6)	16.9 (15.9)	6.8 (7.1)
Middle East and North Africa	180.3 (114.9)	24.4 (16.4)	8.1 (7.4)
Latin America and Caribbean	257.4 (196.8)	25.0 (37.3)	13.3 (19.7)
Mexico	222.0 (259.2)	27.8 (49.5)	16.7 (27.4)
Brazil	326.8 (304.9)	20.8 (63.1)	8.2 (33.8)

Source: Simon, D. (1995). Debt, democracy and development: Sub Saharan Africa in the 1990s. In Simon, D., Van Spengen, W., Dixon, C. & Narman, A. (eds.) *Structurally adjusted Africa: Poverty, debt and basic needs*, p. 25. London: Pluto.

Table 2 External debt and debt burden, 1990 and 2003

	Total of external debt ($/mn)			*Short term debt as % of total debt*		*Total public and publicity guaranteed debt ($/mn)*	
	1990	*2003*	*% change*	*1990*	*2003*	*1990*	*2003*
Low income	341 580	424 472	+24.26	11.0	7.9	7692	23 821
Middle income[a]	995 453	2 129 666	+113.9	16.6	17.3	52 315	571 305
Lower middle income	653 470	1 392 764	+113.1	16.7	17.1	34 132	338 332
Upper middle income[a]	341 983	736 902	+115.5	16.4	17.7	18 182	232 973
Low and Middle Income[a]	1 337 033	2 554 138	+91.03	15.2	15.7	60 007	595 126
East Asia and Pacific	234 081	525 535	+124.5	16.0	23.9	21 635	120 503
Europe and Central Asia	217 224	675 998	+211.2	18.2	19.9	4921	206 601
Latin America and Caribbean	444 900	779 632	+75.2	16.5	10.7	25 029	227 856
Middle East and North Africa	139 555	158 827	+13.8	14.1	12.8	1418	6747
South Asia	124 396	182 785	+46.9	9.9	3.9	1727	17 796
Sub-Saharan Africa	176 878	231 360	+30.8	11.6	13.4	5276	15 623

[a] includes data for Gibraltar.
Source: World Bank (2005). *World Development Indicators 2005*. Washington, DC: Tables 4.16 and 4.17.

states, the largest transitional economy, a geostrategically important and pro Western Islamic state, and the three original big debtors from 1980. Together, these eight biggest debtors accounted for US$1304.1 billion, approximately 51% of all low and middle income countries' debt. However, their respective debt–export ratios vary considerably, representing different debt burdens, from a very manageable level in China and Mexico to unsustainable burdens in Argentina and Brazil. The heaviest debt–export ratios in the world in 2003 were 1952% in Liberia, 799% in Sudan, 697% in Sierra Leone, and 632% in Rwanda, all amongst the poorest and most war ravaged economies. Liberia would require almost 20 years of current export revenues to pay off its external debt.

Such aggregate and snapshot data conceal both sub national variations in terms of geography, class, ethnicity, and agro ecological system, and also trend changes within the overall period. For instance, despite the in creased absolute indebtedness in 2003, Latin America's fortunes are generally held to have improved sub stantially during much of the 1990s, sometimes dubbed the Miracle Decade in contrast to the Lost Decade of the 1980s. However, the late 1990s and early 2000s witnessed renewed financial turmoil and short term crises in the major debtor states.

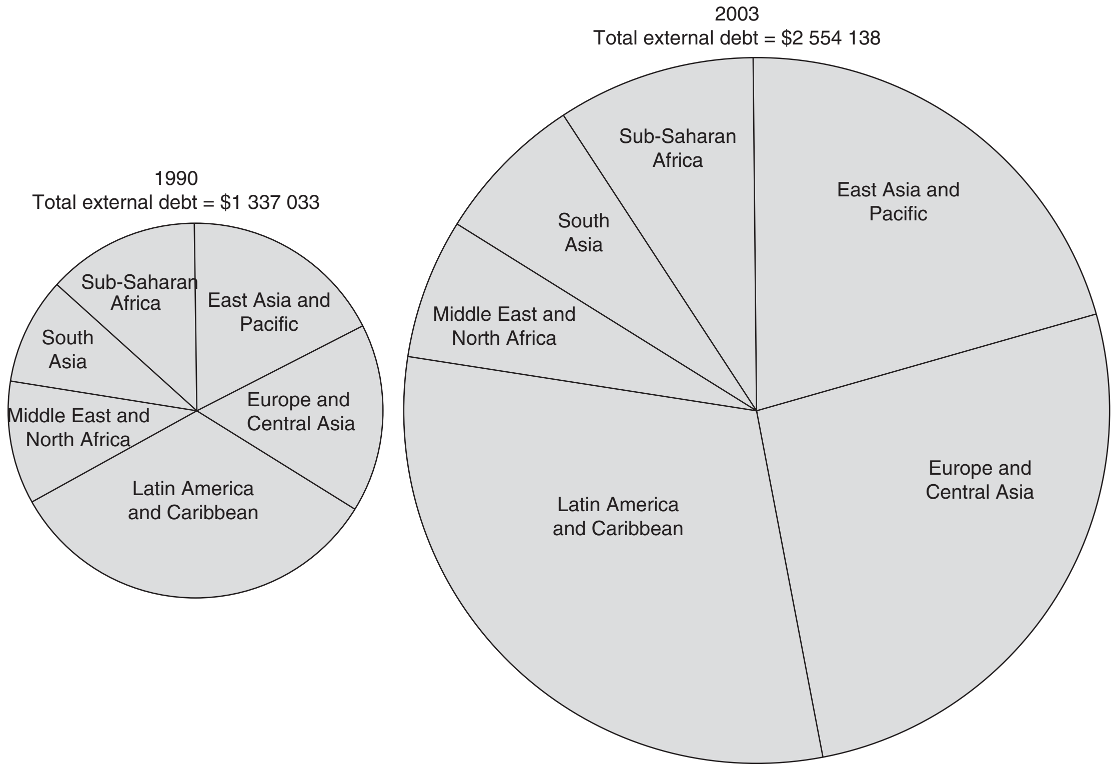

Figure 1 Geographical distribution of external debt, low- and middle-income countries, 1990 and 2003 ($ millions).

Response of the Global Regulators

The likely implications of the debt crisis were immedi ately appreciated by the international financial insti tutions (IFIs) like the International Monetary Fund (IMF) and World Bank (WB) that regulate the global economy, along with the major donor governments and bilateral aid agencies of the Organization for Economic Cooperation and Development (OECD), and the banks themselves. Some of the most heavily exposed banks were threatened with bankruptcy, while any collective default – such as was seriously considered by the Car tagena Group of Latin American states in 1985–86 – would have threatened the entire global economic and financial system. In defense of their own interests, the banks sought and usually received assistance (from tax write offs to direct bale outs) from their home govern ments while pursuing debt repayment with vigor.

The IMF, WB, and main creditor agencies adopted an inflexible position that blamed debtor countries alone for their predicament and ignored the role of the creditor banks and agencies in having been the knowing partners to loan agreements with duly appraised risks. Accord ingly, the IFIs' strategy – that became their holy grail – was to enforce a package of harsh remedial measures (known as Structural Adjustment Programs (SAPs)) on debtor countries as a precondition for any further financial assistance, including rescheduling of public debt. They also strenuously opposed any attempt by debtor countries to unite into a grouping as a counterpart to the London and Paris Clubs of creditors, insisting on dealing with defaulters individually. Their ability to enforce this has both reflected and reinforced the struc tural imbalances of the global political economy and the relative weakness of debtors (neocolonialism).

Until the late 1980s, private and public creditors were reluctant to accept that even the most severely indebted countries were unlikely ever to be able to repay much of their accumulated debt. Accordingly, they opposed write offs in principle and because they feared a domino effect. Instead, they undertook debt reschedulings, with terms reflecting their perceptions of likely repayment and sought, as a minimum, to ensure continuing interest repayments.

Tackling the Debt Crisis

SAPs were designed as short term strategies to stabilize heavily indebted economies, and were followed by long term Economic Recovery Programs (ERPs) designed to restructure and revitalize the economies through extensive liberalization and privatization, diversification of exports, and the like. The first generation programs were applied blueprint like to very different contexts and were unduly harsh, especially on social expenditure.

These shortcomings were eventually acknowledged and addressed, to some extent, after the late 1980s but not before many of the education, health, and welfare gains of previous decades had been undermined. By 1997, the importance of effective state capacity was also recognized in the World Development Report, heralding a partial U turn that sought to reinvigorate government insti tutions selectively, alongside attention to social capital.

Although SAPs and ERPs were intended as bitter medicine to restore the sick debtor economies to health so that they would be able to repay their debts, it became clear that debt reduction required more direct action. Already by the mid 1980s, Latin America and sub Saharan Africa became net capital exporters, that is, their debt repayments exceeded capital inflows through ODA, new loans, and FDI. Still reluctant to write off debt unilaterally, creditors devised a series of other mech anisms to reduce the debt overhang (a situation where new loans are taken out entirely to service existing ones and where the debt service burden displaces productive economic activity). These are discussed next.

Debt–equity swaps were usually undertaken at a dis count, the rate of which reflected the likelihood of the debt repayment. A secondary debt market then emerged in which creditors traded discounted debt certificates according to each party's perception of the eventual re demption prospects. Considerable debt reductions did occur; for example, Mexico reduced its burden by almost US$1.5 billion in 1987. However, the practice declined in the 1990s as a result of increasing domestic criticism in debtor countries that this represented economic neo colonialism, whereby economic sovereignty was being eroded in favor of foreign governments and institutions.

Debt–for–nature swaps represented an effort to over come this criticism while also promoting biodiversity conservation and sustainable development. Costa Rica was the country most deeply involved in this mechanism, recording some demonstrable benefits but ultimately suffering similar concerns about economic and environ mental sovereignty.

Debt auctions involved debtor countries offering their debt certificates to the highest bidder – at discounts representing the purchaser's perception of the prospects of repayment – for conversion into local equity. Auctions were held every month or two in the late 1980s, often involving amounts of over US$100 million. The equity could be for new investment, not just shareholdings in existing companies, hence potentially benefiting the debtor country. By the mid 1990s, however, debt auctions had been discontinued for similar reasons to debt–equity swaps.

Countertrade represented a strategy by debtor coun tries to increase trade among themselves by circumventing the need for hard currency (in which conventional trade is undertaken), and thereby also promoting South–South cooperation. Governments, trade institutions, or donors like oil exporting countries provided bridging funds to enable the commodities being bartered to be consigned at different times. For instance, if Oman and Zimbabwe desired to trade oil for beef but the oil was needed in order to process and transport the beef, the oil–beef exchange rate would be negotiated at the outset, then the oil shipped to Zimbabwe and, after an agreed period of time, the beef would be sent to Oman. The bridging funds would cover transport, demurrage and interest payments. Ultimately, the time consuming negotiations required for each deal, as well as limited countertrade possibilities, undermined the mechanism.

Poverty reduction strategies (PRSs), more commonly known by the title of their principal documents, poverty reduction strategy papers (PRSPs), replaced SAPs and ERPs as the principal IFI adjustment vehicles in 2000–01. PRSPs are required to be compiled through a process of wide ranging consultations between governments and civil society, but the extent of such engagement varies widely and has often been token. In accordance with the Millennium Development Goals (MDGs), the central focus is required to be on the reduction of poverty. Adoption and then IFI approval of a PRSP is – like its predecessors – a prerequisite for adjustment lending or debt relief under the Highly Indebted Poor Country (HIPC) initiative.

The HIPC initiative is a program through which OECD member states provide debt relief to the world's poorest countries with the heaviest debt burden by reducing it to sustainable levels. Launched in 1996 by the IMF and WB, it has involved negotiations with each of the identified debtor states to agree the donors' terms, namely that resources saved by debt forgiveness must be used to tackle poverty and could not be spent on military, security, or other nonpoverty activities. The target is to reduce countries' debt–export ratios to a supposedly sustainable 150%. Signing up the debtors has proved very time consuming and the original target of 20 countries by the end of 2000, to coincide with the Jubilee 2000 campaign by international civil society for large scale debt relief, was not met. The target has since been expanded to 25 and now 38 (of which 32 are in sub Saharan Africa), although not all have yet qualified to participate.

A report in May 2006 by the Independent Evaluation Group (IEG), the WB's self auditing arm, showed that 29 countries had reached the first stage of debt relief, called 'decision point'. This means that they had qualified to receive interim debt relief, ceased to make debt service payments, and had begun to implement the donor policy requirements. The Republic of Congo, which qualified in early 2006, is expected to have debts of US$2.9 billion written off. Eighteen countries had reached the second stage, known as 'completion point', that is, the countries had implemented the donor requirements – principally implementation of a PRS (see above) – and received final debt relief totaling US$19 billion. However, these countries had made only modest progress toward achieving the MDGs. Furthermore, in 11 countries where debts had already been canceled, total external debts had actually risen on account of conditions attached to new loans, exchange rate movements, difficulties in diversifying exports, and so forth. Eight of them (Rwanda, Ethiopia, Uganda, Tanzania, Mauritania, Burkina Faso, Ghana, and Mali) still had moderate debt distress risk, remained vulnerable to external shocks, and would continue to require highly concessionary financing. When the 29 countries receive all eligible relief after reaching their completion points, the total debt forgiven will be approximately $73 billion.

These new data underscore some of the HIPC's limitations, including the need for complementary policy actions beyond the scope of HIPC itself, and the ability of the IFIs to introduce new preconditions for qualification even once the process is well advanced. In contrast to the IEG report, bilateral agencies like the UK's Department for International Development (DFID) report major progress through increased expenditure on health, education, and other antipoverty measures in beneficiary countries as a result of HIPC debt forgiveness.

The HIPC has become the OECD members' principal multilateral debt relief vehicle, although members are also encouraged to forgive bilateral debt in parallel. Like some other donors, the UK writes off 100% of such bilateral debt, compared to about 50% by multilateral bodies. In addition, debt service payments made by a country after reaching decision point are held in trust by the UK and then returned when the completion point is reached. By May 2006, the combined value of the UK's HIPC and bilateral debt forgiveness totaled £1225 million (c. US$2450 million at late 2006 exchange rates). Prime Minister Blair's Commission for Africa report in 2005 saw a positive ongoing role for the HIPC initiative and associated debt relief. This, in turn, fed into the commitment by the G8 summit at Gleneagles in July 2005 to provide enhanced debt relief by canceling 100% of the remaining debt owed to the concessionary lending arms of the WB, African Development Bank (AfDB), and IMF. If agreed by other countries and the Boards of the WB, AfDB, and IMF, the deal could be worth up to US$55 billion for 38 HIPCs.

Conclusions

The debt crisis has been with us for over a quarter of a century and, despite many initiatives by creditors and indebted countries, it is unlikely to disappear in the foreseeable future. The largest absolute external debtors are middle income or, like China and India, economically fast growing countries. However, not all of these are experiencing the heaviest debt burden; this depends on the ability of a country to service its debt. Although the US government is by far the world's largest debtor, being unable to cover its fiscal deficit (which stood at $8.348 trillion in late May 2006), this debt is predominantly domestic rather than external, and there is little concern about possible economic collapse because the debts are generally (if not universally) felt to be sustainable. Nevertheless, it is salutary to note that this debt was more than three times higher than the total external debt of all low and middle income countries in 2005.

While some donors have shown increasing urgency and commitment, such attitudes are far from universally held. However, for impoverished people in different debtor countries – not just the HIPCs – meaningful action to tackle the debt burden is literally a matter of life or death. Poor people live in rural, periurban, and urban areas, all of which must be addressed by appropriate policies. In other words, nationally and locally specific conditions can be crucial; simplistic blueprint type policies implemented across the board stand little prospect of success.

Schemes like debt–equity swaps and countertrade have been supplanted by the combination of donor driven PRSs and the HIPC initiative as the principal current mechanisms for economic restructuring, debt relief, and poverty reduction. Some real benefits have been demonstrated in countries like Zambia, where primary school fees have been abolished, teacher recruitment increased, and school rehabilitation schemes undertaken with the resources freed from debt service. However, these initiatives alone are inadequate. Concerted civil society protests and activism, through international campaigns like Jubilee 2000 and the Trade Justice Movement, have undoubtedly influenced official bilateral and multilateral policies positively in recent years. Ultimately, only more equitable terms of trade will remove the foundation of the debt crisis. As demonstrated here, debt write offs can help but will not, on their own, preclude further debt accrual or solve wider problems.

See also: Aid; Foreign Direct Investment; Geopolitics; Modernization Theory; Neocolonialism; Poverty; Social Capital; Structural Adjustment; Sustainable Development.

Further Reading

Booth, D. (ed.) (2003). *Fighting poverty in Africa: Are PRSPs making a difference?* London: Overseas Development Institute.

Craig, D. and Porter, D. (2003). Poverty reduction strategy papers: A new convergence. *World Development* 31(1), 53 69.

Fraser, A. (2005). Poverty reduction strategy papers: Now who calls the shots? *Review of African Political Economy* 32(104 105), 317 340.
George, S. (1988). *A fate worse than debt*. London: Penguin.
George, S. (1992). *The debt boomerang: How third world debt harms us all*. London: Pluto.
Green, D. (2003). *Silent revolution: The rise and crisis of market economics in Latin America* (2nd edn). London and New York: Latin America Bureau and Monthly Review Press.
Helleiner, G. (1992). The IMF, the World Bank and Africa's adjustment and external debt problems: An unofficial view. *World Development* 20(6), 779 792.
Neumayer, E. (2002). Is good governance rewarded? A cross national analysis of debt forgiveness. *World Development* 30(6), 913 930.
Pettifor, A. (2000). Debt cancellation, lender responsibility and poor country empowerment. *Review of African Political Economy* 83, 138 144.
Schatan, J. (1987). *World debt: Who is to pay?* London: Zed Books.
Simon, D. (1995). Debt, democracy and development: Sub Saharan Africa in the 1990s. In Simon, D., Van Spengen, W., Dixon, C. & Narman, A. (eds) *Structurally adjusted Africa: Poverty, debt and basic needs*, pp 17 44. London: Pluto.
World Bank (2005). *World development indicators 2005.* Washington, DC: The World Bank.

Relevant Websites

http://www.wdm.org.uk
World Development Movement.
http://www.oneworld.net
NGO network, debt.
http://www.brettonwoodsproject.org
Bretton Woods Project, a forum focusing on critiques of the Bretton Woods approach to debt and development.
http://www.dfid.gov.uk/mdg/debt hipc faqs.asp
DFID UK Department for International Development, The Heavily Indebted Poor Countries (HIPC) Initiative Frequently Asked Questions.
http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/0,menuPK:336998~pagePK:149018~piPK:149093~the SitePK:3369992,00.html
PovertyNet, the WB's central focus on poverty, including Poverty Reduction Strategies.

Deconstruction

C. Barnett, The Open University, Milton Keynes, UK

Glossary

Différance A neologism that refers to both a sense of spatial dispersal and temporal deferral characteristic of textuality.

Logocentrism The belief in the unmediated relationship between thought and linguistic expression.

Metaphysics of Presence The privileging of the present tense in the determination of Being in Western metaphysics.

Performative A linguistic act that brings into existence or enacts what it seems to describe, for example, "I name this ship… ."

Textuality A conceptual figure for the generalization of those attributes normally ascribed to writing (e.g., absence, chance, spatial dissemination, temporal delay, impurity) to practices of signification in general.

Introduction: What's in a Name?

Deconstruction is a distinctive tradition of philosophical analysis and textual criticism. It is indelibly associated with the writings of Jacques Derrida (1930–2004). Deconstruction first came to prominence in the 1960s and 1970s as a strand of a broad wave of structuralist and poststructuralist thought emanating from the French academy. Derrida is often thought of as the quintessentially 'French' theorist, but deconstruction has been most influential in the English speaking academy. By the 1990s, Derrida's writings were translated into English almost without delay. Deconstruction is a tradition that exists in and through the medium, institutions, and conventions of translation.

Deconstruction has a distinctive presence in contemporary popular culture. Derrida has had a pop song written in his honor, and a documentary film in which he 'starred'. The academic reception of deconstruction is difficult to disentangle from a succession of very public controversies in which Derrida himself, or deconstruction as an intellectual movement, became more or less willingly involved. This includes the 'Cambridge Affair' in 1992, when a few disgruntled academics formally objected to the University of Cambridge conferring an honorary degree on Derrida, and the invocation of deconstruction in the 'Sokal Affair' in the late 1990s, a controversy surrounding the status of social constructionism and scientific authority. When Derrida died in 2004, comments extended well beyond the usual academic memorials and broadsheet obituaries to the letters pages and op ed pages of major national newspapers in North America and Europe. Though less controversial than in the 1980s, Derrida's legacy nonetheless remained a matter of more than academic interest.

This presence of deconstruction in popular culture provides an interesting counterpoint to what is often a rather self righteous academic insistence, by Derrida and others, that deconstruction is peculiarly resistant to definition. This is normally how entries on deconstruction, such as this one, would start. Derrida once commented that any proposition of the form 'Deconstruction is X' was missing the point. Beneath this claim is a serious line of reflection on the philosophical construal of the mediums through which things present themselves to thought and action. But the popular presence of deconstruction should not be dismissed too rapidly. Deconstruction has become, whatever Derrida's intentions, a popular verb referring to the dismantling and taking apart of authoritative narratives, opinions, or identities, exposing their contingency to view. It is, in this usage, a popular synonym for the more austere sounding concept of critique. Deconstruction is also associated, more positively, with particular styles of artistic practice that take apart and then reassemble the materials and conventions of established genres. The model here is so called deconstructivist architecture, with which Derrida himself was closely associated through his collaborations with internationally renowned architects, such as Peter Eisenman and Bernard Tschumi.

If this populist reference seems out of place in an academic entry on deconstruction, then it is worth acknowledging that this mixing of genres and disrupting of boundaries is a defining feature of Derrida's academic work. Derrida's writing ranges over a huge range of topics and fields. This includes the canon of Western philosophy from Plato onward, including Rousseau, Kant, and Hegel, and Nietzche, Husserl, and Heidegger. But Derrida's reading of philosophy is explicitly inflected by an engagement with the canon of modernist literature, including Mallarmé, Celan, Blanchot, and Joyce. One recurrent theme of Derrida's work is the degree to which a clear cut distinction between philosophical reason and literary expression is not possible. Derrida also engages with the work of key modern social and political thinkers, including Marx, Freud, Saussure, and Levi Strauss. Here, the recurrent theme is the degree to which modern social science entrains a series of philosophical tropes just as it breaks from classical metaphysical modes of reflection. In his engagement with the work of Emmanuel Levinas, Walter Benjamin, and Søren Kierkegaard, Derrida simi larly raises the question of whether the secularization associated with post Enlightenment reason can or should break wholly with intuitions of religious faith. In his consistent engagement with nonwritten arts, including painting and architecture, Derrida likewise raises ques tions about the hierarchies of different sensory modalities that elevate philosophical reflection to the highest order.

Deconstruction is best approached as a way of reading. As a method of analysis, deconstruction exposes un acknowledged implications in existing traditions. If this is understood solely as a critical impulse, however, then deconstruction is hardly distinct from a 'hermeneutics of suspicion'. But Derrida was less a critic than he was an advocate of the writers and texts that he subjected to such forensic attention. Derrida leaves us with a whole new genre of theory, in which thinking is nothing more than the working through of an inheritance and a taking of responsibility for traditions of thought.

The general significance of Derrida's work cannot be detached from the distinctive style of his writing. All of Derrida's significant works are commentaries on texts by other writers. Deconstruction thus works through the elaboration of particular texts, rather than creating con cepts or general systems. The concepts most often asso ciated with deconstruction – such as 'supplement', 'pharmakon', 'trace', and 'chora' – are found by Derrida in the texts he reads. He shows that they turn out to have greater potential meaning than they are usually ascribed by their host traditions. This systematically parasitical dependence on other texts for the elaboration of a deconstructive conceptual vocabulary implies that the generalizing application of deconstructive motifs in other contexts should be undertaken with great care.

Given the size, range, and depth of Derrida's *oeuvre*, it is useful to think of his career falling into two phases. In the first phase, deconstruction was an avowedly critical enterprise, and became closely associated with a rigorous mode of literary–theoretical textual analysis. In the second phase, deconstruction takes on a more affirmative tenor.

Deconstruction: The Basics

Derrida's most important work is *Of Grammatology*, first published in France in 1967, and translated into English in 1976. *Of Grammatology* lays out the basic lineaments of a deconstructive method of philosophical analysis, which refuses to ignore the textual medium in which universal philosophical propositions are presented. It is in this book that Derrida makes the famous claim that "There is nothing outside the text." This remark has become one of the most widely cited, and most widely misunderstood, thoughts in twentieth century philosophy. It is best understood as an interpretative rule, according to which reading follows the immanent patterns of texts to lay bare inconsistencies and contradictions, rather than apply external criteria of interpretation in order to impose unity on them. This protocol of literal reading draws on an older and deeper tradition of heterodox hermeneutics that stretches back to Spinoza.

Of Grammatology consists of close readings of canonical texts by the founder of modern structural linguistics, Saussure; the anthropologist, Levi Strauss, who pion eered structuralist anthropology; and one of the pivotal figures of modern political thought, Rousseau. This is, then, a critical engagement with some of the founding texts of radical egalitarian thought. In the case of each author, Derrida identifies a tendency to denigrate writing as a secondary, contingent medium for the articulation of pure thoughts properly expressed in direct speech. Derrida calls this privileging of expressive speech over the dangers and impurities of materially mediated com munication 'logocentrism'. Logocentrism embodies a deep prejudice in Western thought for the ideal of a disembodied, isolated subject free from dependence on others.

In the first, critical phase of deconstruction, this in herently normative evaluation of the relationship between speech and writing, and orality and literacy, is subjected to a sustained analysis that leads to apparently perverse conclusions. Derrida argues that if writing is able to act as a material support to the purity of expressive speech which always threatens to supplant the original, then this logically implies that something essential must be absent from the pure form. It turns out that far from being a mere supplement, writing is a necessary supplement to the supposedly pure form of speech.

This analysis of speech and writing in *Of Grammatology* exemplifies a more general theme in deconstruction, whereby what is secondary, accidental, or contingent turns out to be fundamental to the working of identities, meanings, and systems. In 1988, Derrida characterized deconstruction as "a practical analysis of the parasite and of the axiomatics upon which its interpretation is based." In invoking the parasite as the term with which to question the normativity of received understandings of belonging and property, Derrida made recourse to a resolutely spatial figure.

Derrida calls the assumption that phenomena, such as meaning or identity, must have essentialist forms, "the metaphysics of presence." This term indicates the rele vance of deconstruction to geography's concern with spatiality and temporality. Deconstruction is indebted to Heidegger's argument that Western thought has consist ently privileged the present tense when trying to ap prehend the nature of Being. By affirming the irreducible role of writing in the expression of thought, Derrida argues that all those qualities for which writing (or tex tuality) stands – absence, spatial, and temporal extension, the threat of intention going astray or of destinations not reached – are constitutive of apparently free standing entities such as the unified self identical subject of philosophical reason. This theme is articulated by one of Derrida's most important neologisms, the notion of *différance.* This term refers to the movement of spatial dispersal and temporal deferral that Derrida holds to be the condition of possibility for any and all identity, punctuality, or unity.

In geography, as in any other fields, Derrida's work is most often thought of as a poststructuralist radical ization of structuralist accounts of the relational quality of meaning (and by extension, of identity). On this understanding, Derrida is usually interpreted as affirming an endless, interminable slippage of meaning that escapes any attempt at closure. But the concern in his work with issues of presence and absence also indicates the degree to which deconstruction is indebted to phe nomenology. This is a much more important context for understanding deconstruction than is usually acknow ledged in the interpretation of Derrida as a post structuralist thinker. Derrida's earliest work was on Husserl, the founder of modern phenomenology. His analysis of the metaphysics of presence, as already suggested, is developed in a tense relation of attraction and repulsion to the work of Martin Heidegger, the other pivotal thinker of the phenomenological tradition. And one of his closest mentors, with whom his work continues on an oblique dialog from the 1960s onward, was Emmanuel Levinas, whose phenomenological writ ings challenge Heidegger's privileging of fundamental ontology and reassert the priority of ethical orientations to others as the constitutive mode of human relationality. By bringing his analysis of logocentrism to bear on this tradition, Derrida points up the limitations of the persistent privilege accorded to unmediated subjective experience in phenomenology.

In relation to both structuralism and phenomenology, Derrida's itinerary remains firmly entrenched in what has become known as 'continental philosophy'. But Derrida's own work engages in important ways with what is often considered a distinctively Anglophone tradition of 'analytical philosophy'. In Derrida's case, this is most clearly evident in his long time engagement with the writings of J. L. Austin, the most famous exponent of ordinary language philosophy. Derrida's treatment of Austin's idiosyncratic 'linguistic phenomenology' is the most important single source for the popularization of the notion of performativity across the social sciences and humanities. In the 1970s, Austin became the focus of an ill tempered exchange between Derrida and the American philosopher John Searle. This exchange has become a model for the mutual hostility and incompre hension between analytical and continental philosophy. But Derrida's ongoing appreciation of Austin, as well as the degree to which Austin's vocabulary becomes an essential aspect of how deconstruction is made intelli gible by its supporters from the 1980s onward, is indicative of a much more messy, even parasitical rela tionship between deconstruction and ordinary language philosophy than is often admitted by either 'side' of this division.

One lesson to take from this relationship between the quintessentially French philosopher Derrida, and the quintessential exponent of 'Oxford philosophy' is that deconstruction is less a tradition concerned with lan guage, textuality, or meaning in narrow ways, but is part of a broader revival of interest in the arts of rhetoric, in language in use. Rhetoric is classically defined as the art of persuasion and eloquence. The history of Western thought can be understood as a long quarrel between serious reason and rhetoric. From one perspective, sci ence, rationality, and enlightenment overcome religion, superstition, and magic. A counternarrative sees in this process only the subordination of visceral, creative pluralism to soulless reason. The quarrel turns on a shared set of opposed pairs: reason versus passion; fact versus opinion; neutral versus partisan; and reason versus rhetoric. This opposition underplays the extent to which rhetoric, as a classical discipline, was concerned with the ways in which audiences could be swayed through a combination of both rationality and emotion. Derrida's consistent concern with the medium in which philo sophical truth is communicated – its textuality – and with the parasitic relationship between the value of truth and the forms of fiction and imagination, sits squarely within a tradition of thought that endeavors to reorder these oppositional pairs, rather than assert the superiority of one over the other.

There is another important feature of Derrida's debt to Austin. Austin's account of ordinary language focused on the 'force' which secured the authority of acts that had no final foundation in representational truth. Derrida uses this analysis of the conventional qualities of public acts of language use to open up a line of questioning concerning the foundational conditions for the success of various sorts of 'performative' acts. This includes an analysis of the peculiar force of the constitutional foundations of modern democracy. The reference to Austin is also evident in Derrida's most influential political essay, the *Force of Law,* which is an extended reflection on the relationships between justice, law, and violence. It is Austin's resolutely English, almost paro chial philosophical vocabulary, that serves as the privil eged register in which Derrida, the embodiment of 'French theory', sought to articulate the politics of deconstruction from the late 1980s onward.

Deconstruction and Politics

Derrida's work has been consistently subjected to a demand for politics. The political relevance of deconstruction in its critical phase was somewhat unclear. The elusive quality of Derrida's writing generated a great deal of commentary, in which deconstruction was interpreted as a variety of ideology critique, which can help to expose the contingent foundations of naturalized ways of thinking about identity, meaning, or subjectivity. In this interpretation, deconstruction's demonstration of ambivalence, contradiction, or silence in canonical traditions of thought is assumed to be automatically disobliging: it is assumed that showing the contingent conditions of beliefs or values weakens the level of confidence which should be invested in them. This interpretation, while widespread, mistakes the ethicopolitical force of Derrida's work. Any political interpretation of deconstruction must give due weight to how Derrida's writing, in its very method and style, reorders the assumed relationships between what is given and what is possible that underwrites standard models of academic critique. Derrida does not present what is given as in need of negation or transcendence, or as something to be overcome, transformed, or sloughed off. Rather, the given is treated as an inheritance, one that must be approached responsibly in order to judge what should be subjected to criticism, what should be jettisoned, and what should be affirmed. This means that any exercise in deconstruction is just as likely to be vindicatory as it is disobliging.

This affirmative ethos of deconstruction became much more explicit in the last decade and a half of Derrida's career. Deconstruction had reached its institutional zenith in the 1980s, primarily in literary studies in the USA. Up to this time, deconstruction had been less than warmly received in mainstream English language philosophy. The shift in Derrida's work from taking Western philosophy to task for its blind spots toward developing a vocabulary of ethical concern and political responsibility coincides with a series of controversies that propelled deconstruction onto the front pages of national newspapers in the USA, UK, and Europe. One of these controversies surrounded renewed interest in the depth and profundity of Heidegger's Nazi affiliations in the 1930s and 1940s, in view of deconstruction's intellectual proximity to Heidegger's project. A much more influential controversy followed the revelation in 1987 that Paul de Man, Derrida's close friend and the leading deconstructionist critic in the USA until his death in 1983, had penned anti Semitic literary journalism in wartime Belgium.

There is little doubt that in the wake of these controversies Derrida's writing undergoes an explicit ethical and political turn. This leads, paradoxically perhaps, to a much more receptive engagement with deconstruction across the social sciences than had previously been the case, when it had been so closely identified as literary theory. The ethical dimensions of deconstruction became increasingly foregrounded as the importance of Levinas as a reference point for Derrida's project became more widely appreciated. Derrida's own writing also begins to address itself much more explicitly to ethical and political topics in various traditions of political theory, moral philosophy, and theology.

Derrida's later work focused on a family of virtues and values of relating: hospitality, forgiveness, confession, tolerance, testifying, bearing witness, gift relations, mourning, justice, and friendship. All of these refer to a set of attentive, generous, and responsive ways of relating to others. Derrida's ethical writings have thus contributed to the idea that ethics is primarily concerned with the relationship between responsibility and otherness, a theme developed in recent work in geography on post structuralist ethics and nonrepresentational theory.

However, Derrida's philosophical revival of this family of topics should not be too quickly consigned to the ethereal arena of ethics. What is at stake is an effort to think about politics without conforming to established models of political relationality, which depend either on idealized models of consensus and reciprocity or darkly pessimistic views about violence and antagonism, which evacuate any and all space for normative judgment from the world. Derrida's discussion of this family of topics is, furthermore, explicitly presented in a political register in which geographical themes play an important part. For example, his discussions of hospitality and cosmopolitanism are addressed to contemporary debates about immigration, asylum, and multicultural in European nation states. His evocation of the rhetoric of ghosts and haunting in his long awaited reading of Marx rethinks the possibility of international solidarity in an age of economic globalization. And his treatment of the theme of friendship rethinks traditional conceptions of sovereignty, law, and statehood. At the same time as Derrida's reflections on ethics and politics address key geographical assumptions of modern thought, they also focus on a set of understandings of historicity and temporality. For example, Derrida's concern with 'democracy to come' explores the relationship between the promise of political ideals and their institutionalization. Any concrete manifestation of democratic norms in legal form will always be exceeded by demands for justice that cannot be anticipated.

Derrida's more explicit treatment of ethical and political issues has meant that deconstruction has been reconciled with traditions of Marxist and post Marxist critical theory that were previously doubtful of its political credentials. This has coincided with a degree of rapprochement between deconstruction and analytical traditions of philosophy. Where once it was claimed by

Derrida's detractors that 'proper' philosophers did not bother with deconstruction, there is now a flourishing literature concerning the overlaps between deconstruction and analytical approaches. One important reference point for this movement is the idiosyncratic philosophy of Stanley Cavell, which displays a guarded appreciation of Derrida's philosophical project as a complement to his own. But above all, perhaps, it has been provoked by the renewed interest among Anglo American philosophers in philosophical pragmatism triggered by Richard Rorty in the late 1970s. It is through this pragmatist interpretation of Derrida that deconstruction has been most influentially translated into the disciplinary debates shaping human geography.

Deconstruction in Geography

Deconstruction has had a variable reception history in geography. The ways in which deconstruction has been interpreted disclose some of the conventions of theory construction in the discipline. Derrida is rarely cited as a key thinker on issues of space and place. Nor does his work conform easily to the preference for materialist analysis that emphasizes relations of power which has shaped the discipline's engagement with post structuralism. Deconstruction has been thought to over emphasize textuality, without much care being taken to consider how Derrida's re inscription of the texture of textuality might carry some important insights for how to conceptualize spatiality and temporality. Nonetheless, Derrida is a background presence in a number of intellectual developments in the discipline in the last decade and a half. For example, many of the feminist theorists with whom geographers have most closely engaged are strongly indebted to Derrida and deconstruction. This is true, for example, of Judith Butler's politics of performativity; Elizabeth Grosz's work on embodiment; Gayatri Chakravorty Spivak's Marxist deconstructionist postcolonialism; and Iris Marion Young's retheorization of justice and the politics of difference.

Deconstruction first came to prominence in geography as part of debates about postmodernism in the late 1980s and early 1990s. It was invoked as an authoritative reference point for critiques of essentialism and foundationalism in epistemology. This epistemological reading of deconstruction saw Derrida's situated readings of philosophical traditions externally applied to support generalized arguments about the contingency of all knowledge claims and the constructedness of social phenomena. The construal of deconstruction as anti essentialist and antifoundationalist owes a great deal to the ways in which geographers' discussions of postmodernism were shaped by Richard Rorty's distinctive revival of philosophical pragmatism. Derrida has become the authoritative reference point for a set of general claims about binary oppositions, dichotomies, and dualisms that shape the terms of theoretical criticism in poststructuralist geography. His work is meant to prove beyond doubt that all meaning depends on the temporary stabilization of the differential movement of signification; that the validity of all foundations is dependent on the operation of hierarchical binary oppositions; and that all identities are constituted through arbitrary acts of exclusion and closure. Deconstruction is used as the source of external criteria that can be applied to particular objects of critical analysis. In this usage, geographers have applied deconstruction as a variety of ideology critique to debunk claims of objectivity and naturalness.

The epistemological framing of deconstruction has been supplemented more recently by a more sophisticated focus on the spatial and temporal vocabularies through which Derrida's writing questions philosophical reason. The single most sustained engagement with Derrida's ideas in geography is Marcus Doel's treatment of the spatial and temporal metaphysics of deconstruction in *Poststructuralist Geographies*, which lays out an alternative spatial grammar of mobility, relations, and foldings. This treatment underscores two features about theory construction in geography. First, Derrida continues to be interpreted as distinctively 'French' and/or 'Continental', covering over the conditions of translation and controversy through which deconstruction has developed as a hybrid tradition of thought. Second, Derrida increasingly comes into focus for geographers because of a creeping ontologization of theory in the discipline. This is one index of a more fundamental problem with the use of deconstruction in geography. There is a tendency to apply deconstructive concepts as if they were social–theoretical concepts. Derrida's engagement with traditions of philosophical, literary, and aesthetic reflection are almost completely devoid of any mediation by sociological or historical conceptualization. When deconstruction is mistaken for social theory *per se*, this ends up producing highly abstract, and somewhat, simplistic formulae.

In geography, deconstruction is primarily treated as a stock of ideas to be applied, rather than a practical, methodological orientation to research. There are at least three examples of such 'application' of deconstruction to social theory that have been given an airing in geography. First, it is often assumed that the deconstruction of the grammatical subject of philosophical reasoning maps directly onto an account of the contingency of personal or social identities. In this application, deconstruction tends to be run together with other poststructuralist accounts of subject formation, for example, from psychoanalysis or from discourse theory, to produce a standard account of 'the social construction' of identity. What distinguishes this variety of social constructionism from older ones is a strong emphasis on language as the medium in and through which subjectivity is constructed. To borrow a distinction from Nigel Thrift, this style of poststructuralism ends up overemphasizing the degree to which subjectivity is socially 'constructed' at the expense of exploring how people are always busy socially 'constructing' subjectivity in interaction.

Second, this style of argument about the social construction of the subject is closely associated with a model of power that appeals to the authority of deconstruction. Here, Derrida's demonstration that the intelligibility of meaning depends on a play of what is present and what is absent is reinterpreted into a claim that all identities, institutions, or meanings depend for their coherence on the exclusion of some element that, in turn, returns to potentially undermine this coherence. Understanding power in terms of exclusion and closure has an obvious appeal to geographers. It has become influential in various strands of relational thinking in geography. However, like the argument about subjectivity with which it is closely associated, it tends to subsume deconstruction under a broader psychoanalytical vocabulary (and specifically, it insists on reading Derrida through the lens of Lacan's interpretation of structural linguistics).

In both cases, in the account of the subject and of power, deconstruction is made applicable to social science by adhering to a fairly standard model of language as a system of meaning. Deconstruction is invoked to claim that meaning is inherently fluid, only temporarily stabilized through arbitrary acts of closure. These sorts of arguments, by investing so much theoretical authority into theories of language, miss the point of Derrida's critique of logocentrism: the assumption that the subject is located wholly on one side of a division between thought and the world is precisely the target of Derrida's work. It is this focus that brings his work into more or less unlikely proximity with other critics of philosophical skepticism since Wittgenstein, such as Stanley Cavell or John McDowell. And as already suggested, these arguments also underplay the extent to which deconstruction is really concerned with the arts of rhetoric rather than developing a theory of language, focusing on textualized modalities of action rather than on concepts of meaning and signification.

This brings us to a third application of deconstruction in social science, in which Derrida's work is invoked to argue that all social analysis is really just a form of 'writing'. This sort of argument has to invent a straw figure of naturalist social science as naïve empiricism or positivism in order to claim that deconstruction demonstrates that all description or explanation of the real world is dependent on conventions of narrative and storytelling. The real, it turns out, is just an effect of fictional representations. This sort of argument presents deconstruction as the latest in a longer line of broadly interpretative approaches to social science. It radicalizes these approaches by claiming that all writing is a variant of fictional storytelling. In so doing, this application of deconstruction ends up inadvertently relying on subjectivist, expressive conceptions of literary writing that deconstruction, in an earlier incarnation as literary theory, had systematically criticized. It also tends to reduce deconstruction's analysis of textuality to a simplistic understanding of writing as a system of conventional meaning making.

Each of these three applications of deconstruction to social science is visible in geography. In each case, deconstruction is positioned as a variant of interpretative social science, made to bolster this broad perspective against the imagined threats of naturalist, objectivist, quantifying explanatory social science. What seems to get lost in this application of deconstruction is the extent to which Derrida's work, in its critical phase, actually calls into question some of the basic assumptions about the relationships between meaning, experience, and subjectivity upon which interpretative approaches to social science themselves rely. The targets of Derrida's deconstructions are, remember, the key thinkers in the structuralist and phenomenological canon, the two most important sources for interpretative social science in the twentieth century. What emerges from this phase of deconstruction is a persistent focus on the way in which these avowedly radical approaches tend to install hidden normative criteria of harmony, originality, purity, and unity into the interpretative methods of analysis, evaluation, judgment, selection, and reading, which they bequeath the social sciences. It is here that deconstruction's relevance to social science lies. If deconstruction is critical, then it is first and foremost critical of the elitism, moralism, and piety of radical intepretativism in the social sciences as well as the humanities. Deconstruction encourages us to reconsider the normative assumptions built into theory formation, empirical investigation, and concept use. It is unwise to assume that the force of deconstruction rests on a debunking of claims to objectivity or rational argument in the social sciences. To do so is to reproduce a long standing romanticism, not least in human geography, wherein the value of authentic experience of places, of direct contact with the field, and of unmediated and expressive forms of communication continue to be held in the highest regard.

The affirmative phase of Derrida's writing abounds with figures of attentive, generous, and responsive relations of openness, such as arrival, hospitality, and forgiveness. This work is easily read as providing a kind of ethical vocabulary, which can serve as the regulative ideals with which to oppose the grubby realities of worldly exclusions, closures, and resentments. But this is probably the wrong way to interpret this work, not least because it ends up overemphasizing the spatial aspects of these figures at the cost of taking seriously the distinctive temporality of relating that they stand for. What Derrida seems to have been striving for in this work is not an 'ethics', but rather a way of bringing into view those aspects of togetherness that 'conventional' understand ings of sociability, whether based on models of interest or meaning, cause or motive, squeeze from view: the emergent dimensions of being with others which are contingent in the sense that they can and do adjust to unanticipated events, rely on tacit understandings yet to be articulated, and may take forms that cannot yet be imagined.

There are certainly areas of geographical research in which these aspects of deconstruction have been ac knowledged and productively developed. For example, one substantive area of research in geography which is heavily inflected by deconstruction is postcolonialism. Derrida's concern with issues of reading, interpretation, and context are intimately related to a wider critique of Western historicism. The deconstruction of the speech/ writing distinction in *Of Grammatology*, for example, was presented as a critique of a deep seated ethnocentrism that inhered in even the most radical aspects of Western social and political thought. The degree to which deconstruction in translation might have been always already postcolonial is indicated by the fact that this book was translated into English by Spivak, a leading post colonial theorist and maverick deconstructionist critic at the same time. This postcolonial dimension to decon struction is also evident in Derrida's treatment of dem ocracy to come. The idea of the West as the unique bearer of principles of Enlightenment (e.g., liberty, democracy, or justice) which are diffused outward to non Western cultures is challenged by his affirmation of the radically open ended quality of forms of democracy yet to be anticipated.

Another related area where deconstruction has made its mark in geography is in discussions of radical dem ocracy. Derrida's discussions of sovereignty, hospitality, friendship, and cosmopolitanism challenge some of the basic assumptions about the spaces and times in which democracy is enacted. Deconstruction is thus closely related to projects aimed at rethinking 'the political'. However, there is a tendency in this style of political deconstruction to ontologize relations of hostility and antagonism as the primary generative force of human affairs. Derrida's account of the performative force of relations of law and justice is easily interpreted as an account of how all sovereign power is necessarily underwritten by relations of violent exclusion. This sort of account makes it very difficult to account for observable relationships of compromise, negotiation, and cohabitation. Yet these were precisely the focus of Derrida's readings of various attentive, generous, and responsive modes of relating in his later work.

An alternative interpretation of deconstruction as bearing on issues of democratic theory would start from the observation that these modes of relating all enact some form of public space of address and encounter. Another of the unexpected rapprochements in Derrida's later career was a convergence with the critical theory of Jürgen Habermas. These two thinkers, so often taken to stand for fundamentally opposed perspectives on the legacy of the Enlightenment tradition, turn out to have a number of shared interests. For example, both owe a debt of gratitude to Austin's ordinary language philosophy, and its intriguing implications about how actions bring into existence different scenes of public engagement. Both also share a concern to redeem what is of lasting value from Kant's treatment of universality and cosmopolitanism in an age in which issues of global responsibility are acutely felt. In both respects, there is considerable potential for connecting Derrida's work to the task of rethinking the terms in which geographers have conceptualized practices and regulative ideas of public space.

Conclusion

The controversies that characterized Derrida's academic and public career, and out of which the styles and topics of deconstruction emerge, reveal the expectations that academics bring to their engagement with complex conceptual and philosophical ideas. Deconstruction has often been found unsettling because it does not culmin ate in clear criteria of judgment or practical reason. Even many of Derrida's fans tend to suggest that he just does not know when to stop with the relentless ques tioning: that deconstruction needs to be followed by reconstruction. This sort of claim only reproduces the idea that deconstruction is a negative, disobliging pro cedure of demystification. But deconstruction is better thought of as a practice of reasoning governed by the imperative of working through inherited traditions in critical, inventive, and responsible ways.

See also: Language and Research; Other/Otherness; Performativity; Phenomenology/Phenomenological Geography; Philosophy and Human Geography; Postmodernism/Postmodern Geography; Poststructuralism/Poststructuralist Geographies; Pragmatism/Pragmatist Geographies; Semiotics; Structuralism/Structuralist Geography; Text and Textuality.

Further Reading

Barnett, C. (1999). Deconstructing context: Exposing Derrida. *Transactions of the Institute of British Geographers (New Series)* 24, 277 293.

Barnett, C. (2004). Deconstructing radical democracy: Articulation, representation and being with others. *Political Geography* 23, 503 528.
Barnett, C. (2005). Ways of relating: Hospitality and the acknowledgement of otherness. *Progress in Human Geography* 29, 1 17.
Critchley, S. (1999). *The Ethics of Deconstruction* (2nd edn.). Edinburgh: Edinburgh University Press.
Derrida, J. (1976). *Of Grammatology*. Baltimore, MD: Johns Hopkins University Press.
Derrida, J. (1982). *Margins of Philosophy*. Chicago, IL: University of Chicago Press.
Derrida, J. (2002). *Acts of Religion*. London: Routledge.
Derrida, J. (2002). *Negotiations: Interventions and Interviews 1971 2001*. Stanford, CA: Stanford University Press.
Derrida, J. (2005). *Paper Machine*. Stanford, CA: Stanford University Press.
Doel, M. (1999). *Poststructuralist Geographies*. Edinburgh: Edinburgh University Press.
Glendinning, S. (ed.) (2001). *Arguing with Derrida*. Oxford: Blackwell Publishing.
Harpham, G. G. (2002). *Language Alone: The Critical Fetish of Modernity*. London: Routledge.
Mouffe, C. (ed.) (1996). *Deconstruction and Pragmatism*. London: Routledge.
Norris, C. and Benjamin, A. (1988). *What is Deconstruction?* London: Academy Editions.
Royle, N. (ed.) (1999). *Deconstructions: A User's Guide*. Basingstoke: Palgrave.
Thomas, M. (2006). *The Reception of Derrida: Translation and Transformation*. Basingstoke: Palgrave.
Thomassen, L. (2006). *The Derrida Habermas Reader*. Edinburgh: Edinburgh University Press.
Wheeler, S. (2000). *Deconstruction as Analytic Philosophy*. Stanford, CA: Stanford University Press.
Wigley, M. (1993). *The Architecture of Deconstruction: Derrida's Haunt*. Cambridge, MA: MIT Press.
Young, R. (1990). *White Mythologies*. London: Routledge.

Defensible Space

E. Warwick, King's College London, London, UK

Glossary

Defensible Space A design concept identifying spaces that can be controlled and supervised by the occupants of the surrounding buildings.

Environmental/Architectural/Social Determinism The extent to which the environment, architecture, or social factors dictate human behavior.

Image The impact of a physical design's form and materials to influence the occupants' perception of it, particularly negative associations with class, the isolation, and stigma of poor quality housing.

Milieu The positive influence of adjacent activities, for example, the juxtaposition of active, well-used streets, on the security of adjacent areas.

Surveillance The opportunities from the physical design of buildings and streets to provide surveillance to residents. Defensible space principles emphasize natural surveillance (location of overlooking windows, visible pathways, and spaces) over technical means such as CCTV, or external activities such as police or security guards.

Territoriality The feeling of ownership of a space by its occupants, with an associated sense that strangers are unwelcome. The concept of defensible space identifies territoriality as the ability of the physical environment to create identifiable zones of influence.

The Concept of Defensible Space

At its simplest, defensible space is a design concept defining spaces that can be controlled and supervised by the occupants of the surrounding buildings. Oscar Newman, its best known advocate, conceptualized it as the way that residential environments can be designed so that their physical characteristics (the layout of the site, the buildings, and the architectural design of the homes themselves) allow the residents to directly participate in ensuring their security.

Lack of defensible space is perceived to have a negative effect, resulting in a range of undesirable behaviors; from increased vandalism to weakening of casual social interaction and good neighborliness. The strength of design in countering or preventing antisocial behavior is contested. But despite the uncertainty surrounding its validity, defensible space's potential as an explanation for, and a possible solution to, many of the failures of late twentieth century mass housing, has warranted its continued interest to both academics and practitioners. The application of its principles has been extensive and is commonly understood by a variety of disciplines and areas of study (e.g., urban geography, urban sociology, architecture, planning and housing policy, and even spawned subfields in criminology such as environmental criminology). While it has primarily been used to explain the interaction between the physical design of spaces and crime, the concept is far richer and more complex than simply the degree that the design of space facilitates or inhibits criminal activity, and has influenced interpretations of place, meaning, and behavior.

The Dimensions of Defensible Space

The criteria required to achieve defensible space are a mixture of social and physical circumstances. There is a cluster of criteria influencing perceptions of safety and control; "visibility to witnesses, community spirit and being prepared to guard neutral territory, stream of potential witnesses, and demarcation of private territory physically and symbolically" (Mawby, 1977: 171), with a second cluster related to building form, appearance, and location. Newman named these four dimensions territoriality, surveillance, image, and milieu.

Territoriality

Territoriality has an extremely broad interpretation, from the concept of place attachment or social commitment to location, to animal like defending of 'turf' through symbolic or physical acts. Similarly, territoriality can be applied at a range of scales, relating to specific sites or wider neighborhoods.

The positive and negative attributes of territoriality are explored across the social science disciplines; a contradictory interpretation of territoriality arises from interconnected mechanisms that establish control over space. This control can be communicated in a multitude of ways: physical (a person's presence reducing and inhibiting criminal activity or a physical barrier such as fences or walls); symbolic (plants and or others symbolizing ownership of the space); visual (graffiti tags); or physiological (encouraging a sense of safety through well lit, well looked after places). Control mechanisms applied to the same space that communicate possession, can result in alternative readings from a range of users.

To Newman, territoriality consisted of several inter connected effects:

- the process of establishing a sense of ownership for legitimate users of the space;
- which in turn provides a clear definition of areas controlled and influenced by inhabitants; and
- is encouraged through familiarity with neighbors or passers by.

Territoriality is often characterized as 'the absence of anonymity' for legitimate inhabitants, according to Ley, and he emphasizes opportunities for casual contact, leading to joint ownership of space and to the exclusion of noninhabitants. For Newman, territoriality is directed at strangers by controlling their access and movement through an estate. Yet, territoriality might influence factors that jeopardize as well as improve security. The number or frequency of passers by has contradictory interpretations, providing a spectrum of explanations; many strangers might provide greater anonymity for criminals, but the same number of familiar passers by might act as potential witnesses or provide positive opportunities for social interaction.

Surveillance

Surveillance is especially emphasized in the concept of defensible space, which favors the use of natural or in formal surveillance (windows overlooking public spaces, or active street frontages) over mechanical (CCTV) or formal forms of supervision (security guards). However, even with an as apparently simple approach as natural surveillance, there may be contradictory consequences to design decisions; ground floor windows, which may improve visibility over adjacent spaces, might also be advertising opportunities for burglary.

Image

Image is the least well defined, but in some ways the most tangible of Newman's dimensions; he includes the use of distinctive built forms, materials, finishes, or aesthetics, which are identified as having associations with a par ticular social class or lifestyle. Image, and its resultant associations, suffuse housing design from the urban scale to the detail, from the size and scale of the blocks themselves to the selection of particularly institutional or vulnerable materials. Image is, of course, affected by the physical condition of the buildings and spaces, by signs of decay, neglect, or poor maintenance. The dominance of materials in the reading of spaces is widespread; cracked fluorescent lights cruelly exposing the flaws in cold damp concrete ceilings speak of tight budgets, lack of care, and neglected maintenance, and have become synonymous with public housing stigmatized by its appearance and design.

Milieu

Milieu is the positive influence of activities perceived as safe (such as police stations or well used streets) adjacent to areas suffering from crime. Greater levels of activity contribute to a positive milieu through a potentially benevolent increase in passers by. It encompasses the effects of neighborhood context and reputation, and identifies urban design decisions for the location of amenities as an explanation for spatial concentrations of crime or the presence of unpopular areas. Certain activities or amenities (shops, clubs, or bars) act as attractors for crime. Here, Newman touched on the Chicago School ideas of Burgess, Shaw, and McKay of crime hot spots through mapping high and low risk neighborhoods. He ascribes a simplistic juxtaposition of high crime and low crime areas and proposed pepper potting potentially difficult residents within quiet well behaved neighborhoods to reduce crime levels.

The Emergence of Defensible Space from Various Disciplines

Defensible space unites reading of urban space of varied types and scales, across the disciplines influencing ideas of urban design, geographies of crime, and ethnographies of crime.

Initially, research into defensible space focused on low income, high density urban social housing, first in America and then later in the UK. Two key proponents of defensible space were Oscar Newman, an American architect and planner (1935–2004), and the British geographer Alice Coleman (1935).

While Newman and Coleman can be credited with popularizing the term defensible space in the 1970s and 1980s, particularly its application to spaces around housing estates, it was clearly not a new concept. The theoretical foundations of the idea can be traced back to Lewis Mumford's positive examples of humane urban environments described in *The Culture of Cities*, and the concepts of encounter and the interconnectedness of common space that were being explored by urban re searchers. From this developed an understanding of the physical impact of the built form that incorporated image, meaning, and legibility in the formulation of perceptual concepts of place. These strands are united in *The Death and Life of Great American Cities*, Jane Jacobs' positive numerous social interactions were encouraged and iden tified as indicators for successful well designed places.

In his influential paper on the Pruitt Igoe housing project, the sociologist William Yancey found that resi dents clearly recognized the difference between 'private space' in apartments and 'public space' such as shared grounds. In his critique of the estate's design and layout, he identifies a lack of 'semipublic space', which he feels is referred to by housing management professionals as 'wasted space'. But far from being wasteful, he argues that the lack of provision of this type of semipublic space results in residents retreating into internal private spaces. Yancey termed these semiprivate areas that might act as foci for informal networks to develop as defensible space exploring, several years before Newman used the term. A decade later, Wilson and Kelling also identified a retreat into domestic space as a response to, and an escape from, neglected spaces. This sparks a cycle of decay where residents' retreat from public spaces further undermines the social control of communal areas, leading to increased vandalism and disorder.

There is a similarly extensive spatial perspective on crime, and David Herbert's *Geography of Crime* succinctly records the parallel development of criminological and geographical theories. Historically, both disciplines focused on problem areas of the city, looking for patterns of crime and concentrations of criminals (e.g., the nineteenth century studies of British rookeries or Burgess' pre World War I concentric crime model of Chicago). This neutral locational analysis of neighborhood mapped purely spatial patterns of crime, proposing an environment without behavioral differentiation. The areal approach was expanded through an ecological analysis of the impact of poor housing, poverty, transient or foreign populations on crime during the 1950s. Yet, these methods still maintained that a specific behavioral setting had the power to elicit similar responses from diverse occupants and remained at an aggregate interurban scale of investigation.

The evolution of a humanistic geography required a more sensitive approach to reading the mutual interaction of place and people. David Ley's exploration of gang territory in Philadelphia during the 1970s provides a useful explanation of the defensible space concept. His ethnographic approach combined environmental psychology and behavioral science to show how the identity of spaces reinforces the social identity of the group that inhabits them. The meaning of these spaces contains clues that encourage appropriate forms of behavior. Ley's gangs displayed a sophisticated mapping of their turf as defensible space and a complex reading of the physical and symbolic markings of territory and boundary. Similar sensitivities to spatial meaning are displayed in groups seen as victims, the elderly, women, or ethnic minorities.

Herbert proposed a research agenda to be applied to understanding the geography of crime, and particularly defensible space: the classification and understanding of vulnerable environments related to the potential for different offences; the balance between the social and physical factors that expose these vulnerabilities; and finally, most significant to human geographers, whether these social and physical qualities can be classified subjectively by the occupiers of the spaces. Applying this approach to his study of burgled homes in Swansea, Herbert argues that the interpretation of design features (even apparently simple placement of entrance doors or windows) needs to be related to how they are perceived by occupiers and offenders.

It was not until 20 years later that research interests moved from mapping the location of offenders or offences, to explore the idea of the local environment as a framework for individual behavior. Reviewing the literature on the fear of crime, Susan M Smith identifies how vandalism and urban decay are primary sources of residents' anxiety. Anxiety is then easily translated into fear of crime. The higher levels of anxiety and fear found in deprived urban neighborhoods, particularly rundown poor council estates, can explain the anomalies found between the objective distribution of offenses to the perception of the risk of crime. The evolving geographies of crime, deviance, and fear provide multiple readings of the same physical urban environment, requiring differentiation between the responses of diverse demographic and social groups. The shift toward multiple readings; to the experience of the victim; to inhabitants as agents of control; to prevention rather than containment of crime; and to a blurring of boundaries between the social science disciplines, has led toward a more holistic understanding of environment and social influences of crime.

The long term ability of defensible space design principles to influence crime rates is still to be confirmed; its power and appeal to current human geographers clearly lies in the potential of environmental improvements to encourage a greater sense of safety to a broad community. Defensible space theorists' concentration on the communal spaces surrounding mass housing as a vulnerable environment, establishes it as a particular subset of the discussion of crime in the wider urban realm. The restrained social expectations for the threshold spaces to our homes, which vary from the perceptions of acceptable behavior for a busy public street, intensify rather than simplify this complex interactive issue.

In and out of Favor: The Investigation and Practical Application of Defensible Space in the US and the UK over the Last 30 Years

Newman's pioneering work on the concept started in St. Louis where he taught architecture and city planning at Washington University from 1965. He writes how watching the decline of the infamous Pruitt Igoe public housing project in St. Louis at first hand, already a dangerous, unpopular place to live only a few years after its completion in 1956 and uninhabitable long before its demolition in 1972, acted as a stimulus for developing the idea of defensible space. He explored the effects of defensible space in a research study in which he com pared the crime rates of two housing estates in Yonkers and the Bronx, New York. Newman found that one (with high rise blocks) had 50% more criminal offenses than the other which consisted of six story blocks designed to conform to the characteristics of defensible space. Newman publicized his findings in his populist, and publicly well received, 1973 book *Defensible Space: People and Design in the Violent City*.

Over the next 20 years, there were relatively few practical applications of the defensible space approach in the USA (mainly due to Reagan era financial constraints restricting the construction or remodeling of public housing) until the 1990s. Newman remodeled a few decaying and unpopular estates in Chicago in the late 1980s, and most recently adapted a medium size neigh borhood in Dayton, Ohio, which had fallen into decline during the early 1990s. Newman continued to write design guidance, to campaign for improvements in the design of housing and residential neighborhoods, and to proselytize for defensible space until his death.

Publicity surrounding Newman's book was extensive in both the USA and the UK, and the ideas it contained were perceived as contentious but attractive to a range of professionals. It reinforced contemporary research in Britain where Mawby's study of high rise housing in Sheffield explored the relationship of the built environ ment to crime in 1969. During the 1970s, the defensible space concept infiltrated several major housing renewal projects such as the Liverpool Inner Area Research Team's improvements to unpopular 1950s estates and the Department of the Environment's (DoE) research into remediation of difficult to let estates. But little British 'evidence' existed and it was not until the end of the following decade that Alice Coleman's research gathered any substantial data on the impact of defensible space in the UK context. Coleman's research began in 1979 with an extensive study of over 100 000 homes on council estates in south and east London, undertaken by the Land Use Research Unit at King's College, London. Coleman's research was published in the sensationally titled *Utopia on Trial* in 1985, and the presentation of her findings as 'suspects' and 'prosecution evidence' established an adversarial courtroom tone to her argument.

To test her ideas in practice, funding was provided in 1988 by the DoE to remodel five estates in north London, Northampton, Lancashire, and Manchester through the 'design improvement controlled experiment' (DICE) initiative. A £43M program of work included dividing large blocks of flats into smaller self contained sections, providing individual entrances to homes, enclosing private gardens, removing walkways, constructing new roads to improve access, and building infill housing to occupy the underused open spaces on the estates. The DICE initiative continued until the mid 1990s. A thorough evaluation of DICE in 1997 by consultants Price Waterhouse made unfavorable comparison to other renewal programs such as the 'estates action scheme', or the 'priority estates program' (PEP). This report for the DoE suggested that while the design improvements had made some impact, DICE was not a more successful regeneration initiative than others such as PEP, just more expensive, failing to achieve the expected savings in maintenance costs, or higher rental income from tenants. Slight improvements in residents' perceptions of the estates and feelings of safety were noticed, but these effects were not sustained. Again, the extreme claims of its originators were not demonstrated conclusively, and the evidence to support defensible space still proved inconclusive.

From Theoretical Concept to Operationalized Approach

The translation of the research on defensible space into actual design guidelines required a simplification of the concepts to facilitate their application to discrete examples of architecture and urban design. The advice concentrated on two aspects: the classification of space, and establishing criteria for the design of specific aspects of buildings or spaces.

The Practical Elaboration of a Public/Private Space Taxonomy

The division of human activities into public and private spheres has long been the foundation of political and social theories. Translating those social concepts into theories of urban experiences requires a spatialization of these disparate spheres.

Jacobs first provided a succinct argument for the necessary delineation of urban public and private space. Acknowledging that space in a suburban setting differs in function to that of an inner city housing project, Jacobs criticized urban spaces that ooze into each other, iden tifying the lack of definition and demarcation as sending out confusing signals for responsibility and ownership over those spaces.

Newman's concept of defensible space elaborated these two existing definitions of space as either discrete public or private space, into a further four categories: private, semiprivate, semipublic, and public space. Interior spaces within flats were obviously classified as private and streets as public. Lobbies, stairs, or shared internal spaces would be considered semiprivate spaces, and external grounds or gardens accessed by a wider number of residents were classified as semipublic. However, depending on the design and layout, lifts and other internal circulation areas such as walkways or access decks can be classified as public.

Coleman later expanded these categories further to include ‘confused’ space, to describe the grassy patches of indeterminate function often between post war housing blocks, corresponding to the planning term SLOPE – space left over from planning. Later studies recognized that there may be more subtle hierarchies within public space rather than such distinct categorizations, for example, a spectrum of public accessibility between highly public spaces and less public space.

While terms such as semiprivate space were commonly used across the disciplines of sociology and urban design; by the early 1970s, there was a growing emphasis placed by geographers on the classification of space and a desire to codify and empirically measure the effects of space.

Design Elements Affected by Defensible Space

Newman identified eight specific design variables that contributed to poor defensible space:

- size and scale of the block or estate;
- number of dwellings using the same entrance;
- number of storeys in a block;
- interconnected vertical routes (either stairs or lifts);
- presence of multiple alternative escape routes;
- degree that grounds and common areas are shared by different families;
- presence of internal corridors rather than open visible access ways; or
- location and form of the entrance (flush entrances being preferable to set back front doors or entrances facing away from the street).

To these, Coleman added a further seven design variables:

- storeys per dwelling (flats being preferable to maisonettes which Coleman perceived as a proxy for family accommodation and were thus associated with the problems of children living above ground level);
- overhead walkways;
- type of entrance;
- blocks raised up on pilotti or above garages;
- number of blocks on the site;
- number of access points onto the site; and
- presence of play areas.

These cluster into those variables directly limiting surveillance and increasing opportunities for crime (poor visibility in the design of entrances) and those variables that more generally affect positive territoriality and ownership of common areas. The types of negative effects resulting in these design characteristics were overused entrances, isolated and deserted internal corridors or access decks, and unused vulnerable unsupervised places such as pram stores or garages subject to vandalism. Talking to residents, Coleman found it was not just the visible signs of destruction that shaped residents’ concepts of these spaces but the noise disruption from poor soundproofing between flats or from children in inappropriately located playspaces.

The powerful effect of Newman’s first three variables arises from the large number of neighbors sharing entrances or blocks, rather than the height of the dwellings, and Newman identified a high number of dwellings sharing the same entrance as the most influential factor. Emphasizing the impact of density of occupation, Newman placed an almost moral emphasis on the overlarge scale of the estates resulting in a greater number of residents affected by the poor design. Coleman also mentioned the related issue of high child density.

But the interaction of these design variables is complex and possibly contradictory. While an individual front door or garden might facilitate pride and responsibility for its upkeep, unless a front garden is designed as a place to linger, it introduces a buffer zone reducing contact between neighbors. Similarly, breaking estates into smaller parts, limiting access points, or removing walkways, may reduce potential escape routes for criminals but it may also increase residents’ sense of isolation and disconnection with the wider urban fabric.

Linking Design Variables to Social Data

Using these design variables, Newman tried to establish a connection between design characteristics and police crime data on robbery and burglary as well as visible signs of disorder such as broken lights in communal corridors. He made no distinction between the seriousness of the criminal activity, whereas Coleman’s research relied on a purely visual inspection of the disordered environment, and measured visible signs of litter, graffiti, vandalism, or urine and feces. She used this visual evidence as a proxy for broader social malaise and her only social measure was the number of children taken into care. She argued that the sense of unease arising from a chaotic disordered environment is a more immediate gauge of community control and satisfaction than crime events or fear of crime.

From these 15 variables, the concept of defensible space was codified into several indicators, each with acceptable thresholds: Coleman devising the ‘design disadvantage score’ and Newman developing a similar 100 point scoring system in the *Defensible Space Guidelines for Yonkers RFT.* It was these processes of being able to score and assess a design and thus make a predictive assessment of proposed physical changes which appealed to housing policymakers.

Both Newman and Coleman argued that by following these indicators, the redesigned housing and its surroundings would be transformed, converting dangerous crime ridden estates into safe, well integrated neighborhoods. They suggested practical modifications through minor interventions to the typology (dividing open spaces into private gardens) and more radical changes to the urban grain; replacing high rise towers with terraced row houses, and implementing resident controlled streets and cul de sacs.

As well as reducing ongoing management and maintenance costs caused by vandalism, Coleman claimed that designing for defensible space solved many other trying symptoms of 'social malaise' from antisocial behavior, delinquency, and even drug dealing. Here was a way of salvaging existing mass housing and making good the mistakes of the past.

Suspicion and Criticism of the Idea

There have been specific criticisms of the concept of defensible space from a range of professional, as well as academic, viewpoints. In the US, the architecture and planning profession initially reacted positively to Newman's publication. Yet, in the UK, designers appeared more suspicious of the concept, exhibiting a sense of British superiority over what were perceived as extreme problems of American public housing.

Newman's original studies were heavily criticized on methodological grounds. Subsequent researchers have identified statistical flaws, questioning both the statistical analysis and the selective use and underplaying of statistics. Coleman's lack of statistical sophistication and failure to control for background variables was similarly attacked. Several critics implied that Newman made careful selection of the projects to study, and he is accused of omitting potentially conflicting or explanatory data; for example, income, or data on social class.

A more fundamental and deep rooted criticism is Newman and Coleman's misunderstanding of demographic or socioeconomic factors, and their failure to take account of the broader social context. Coleman's frame of analysis is myopically locked into a dislike of a specific historical aspect of architectural style. This leads to some ludicrous claims of causality, linking the growth in crime rates in Britain since World War II to the introduction of 'modern movement' architectural esthetics as the commonest style for public mass housing. Her assertion, that the increase of illegal drug usage on the poorest designed estates is due to design and layout, completely ignores the increased accessibility, affordability, and growth in the popularity of particular drugs that can have little to do with the shape or form of the built environment.

Defensible Space Has Been Blamed for Many Things

In practice, defensible space has proved less successful than predicted by its proponents. Applying defensible space principles has been found to contradict construction and fire regulations or to inhibit residents' privacy. The principles also challenge urban design ideals such as connectivity and permeable space, potentially resulting in fortress like housing estates and gated communities, isolated from their surroundings. Evaluation of remodeled housing estates where the principles have been applied show the claims that the capital cost of the physical improvements will pay for themselves through reduced repair costs, or in nonmonetary improvements in residents' health, have proved unfounded.

The debate about defensible space hinges on the relative significance of physical space or social factors on an individual's behavior. Environmental determinism, the extent to which surroundings dictate human actions, has been an ongoing anthropological and geographical tradition since Darwin. The strength of this effect has been fiercely debated but continually recurs in explanations of spatial behavior.

During the 1980s and 1990s, while Coleman was promoting her work in Britain, architectural determinism was an acutely potent accusation. As a writer on architectural and urban space, Hillier is clear about the potential difficulties in attributing causality to buildings and their architecture. "There is a problem of method in establishing any kind of link between architecture and social outcomes… Housing is invariably a social process as well as a physical product." He points out that it is impossible to separate either architecture or built form as an independent variable in a process of social disadvantage (both markets and bureaucracies assign poor people to poor housing). He also identifies a second difficulty around the strength and scope of defensible space as a sufficiently strong effect to influence both individual and group behavior. If it is hard to establish credible models of causality for an individual's behavior, then extending a robust causality to whole communities is even more problematic.

However, Hillier argues against the converse position that, as environments are behaviorally neutral, their design is irrelevant. Clearly, the reading and understanding of space is significant. But the investigation of defensible space needs to be reformulated to separate out low level effects of space from more complex social systems.

Both Newman and Coleman presented the design of defensible space as a panacea to overcome social differences, and improve behavior, as well as to reduce crime, and as such were typecast as determinists. However, while extensively asserting the dominance of design's influence, both were careful to provide disclaimers. Newman was initially circumspect in his claims for architecture's positive behavioral influence:

> Architecture operates more in the area of influence than control. It can create a setting conducive to realising the potential of mutual concern. It does not, and can not, manipulate people towards these feelings but rather

> allows mutually beneficial attitudes to surface (Newman, 1972: 201).

Newman's later publications became even less definite on the primacy of physical design; acknowledging that there were requirements for both behavioral and physical effects to achieve a successful defensible space and describing it as a sociophysical phenomenon. Good design could facilitate defensive behavior, but only under certain social conditions.

Despite these disclaimers, it has been suggested that Newman's position popularizing environmental determinism forced a reaction from behavioral scientists reestablishing social determinism as the dominant urban science approach. And this cycle had parallels among policy circles; as physical interventions became unfashionable, their perpetrators were berated as environmental determinists.

The resulting dominance of social policy initiatives is equally open to accusations of social engineering. The artificial dichotomy between physical or social influences appears to have been fueled by the need to distinguish varied disciplines of practice (differentiating urban designers, planners, and social scientists). And beneath the blanket accusations of determinism (both environmental and social), the subtleties of the interaction between people and places, and the nuances of how we read, understand, and use spaces were often lost.

Designing Safe Spaces Where People Want to Live

As concerns increase over the growing seriousness of the crimes and levels of violence experienced on housing estates, the desire to design safe spaces to live in has rekindled interest in defensible space. In 1994, residents surveyed by the US Federal Housing and Urban Development unit (HUD) recorded growing concerns about gun or drug crime. There is an element of wistful nostalgia that during the same period, in the UK, Coleman's researchers could have been measuring symptoms of social malaise in terms of the number of discarded crisp packets, amount of dog poo, or defacement of council property. Yet, despite being at opposing ends of a spectrum of seriousness, the solutions to these concerns continue to privilege the offender through focusing on minority behavior. Defensible space principles focus on design interventions to limit the actions of a criminal minority rather than facilitate the improved quality of life for a respectable, legitimate majority.

Herbert (1982: 46) usefully sets out a tripartite definition of defensible space summarizing the range of its effects:

- as a model for residential environments which inhibit crime by creating the physical expression of a social fabric that defends itself;
- as a surrogate term for the range of mechanisms – real and symbolic barriers, strongly defined areas of influence, and improved opportunities for surveillance – that combine to bring an environment under the control of its residents;
- as a living environment which can be employed by its inhabitants for the enhancement of their own lives, while providing security for their neighbors, families, and friends.

It is this third way that attempts to redress the imbalance between the concept of defensible space as an exploration of how spatial structure underpins social life and interactions, rather than simply as an explanation for the location of crime.

Both Newman and Coleman argued that crime can be built into housing form, and that applying defensible space principles provided far reaching solutions to a range of social ills. Newman believed that remaking poorly designed housing would reduce criminal acts and foster a sense of community and place; Coleman, that the design principles would provide a more universal panacea – even with the potential to improve educational attainment and the concentration of school children. In this, they display an unacknowledged but equally ambitious belief in the power and value of design. Defensible space as a concept to salvage problem mass housing estates displays as lofty an idealism as the utopian dreams of the architects who originally built them.

See also: Gated Communities/Privatopias; Neighborhoods and Community; New Urbanism; Public Spaces, Urban; Urban Design; Urban Policy.

Further Reading

Coleman, A. (1985). *Utopia on Trial: Vision and Reality in Planned Housing*. Oxford: Hilary Shipman.

Herbert, D. (1982). *Geography of Urban Crime*. New York: Longman.

Hillier, B. (1996). *Space Is the Machine: A Configurational Theory of Architecture*. Cambridge: Cambridge University Press.

Jacobs, J. (1961). *The Death and Life of Great American Cities*. New York: Random House (Vintage edition 1992).

Ley, D. (1974). *Black Inner City as Frontier Outpost: Images and Behaviour of a Philadelphia Neighbourhood*. Washington, DC: Association of American Geographers.

Mawby, R. I. (1977). Defensible space: A theoretical and empirical appraisal. *Urban Studies* 14, 169 179.

Mayhew, P. (1979). Defensible space: The current status of crime prevention theory. *Howard Journal of Penology and Crime Prevention* 18, 150 159.

Merry, S. E. (1981). Defensible space undefended: Social factors in crime control through environmental design. *Urban Affairs Quarterly* 16(4), 397 422.

Mumford, L. (1938). *The Culture of Cities*. New York: Harcourt Brace and Co.

Newman, O. (1973). *Defensible Space: People and Design in the Violent City*. London: Architectural Press.

Newman, O. (1996). *Creating Defensible Space*. Washington, DC: US Department of Housing and Urban Development Office of Policy Development and Research.

Pain, R. (2000). Place, social relations and the fear of crime: A review. *Progress in Human Geography* 24(3), 365 387.

Smith, S. J. (1987). Fear of crime: Beyond a geography of deviance. *Progress in Human Geography* 11, 1 23.

Wilson, J. W. and Kelling, G. L. (1982). Broken windows. *The Atlantic Monthly* 3, 29 38.

Yancey, W. L. (1971). Architecture, interaction, and social control: The case of a large scale public housing project. *Environment and Behaviour* 3(1), 3 21.

Deforestation

J. J. Metz, Northern Kentucky University, Highland Heights, KY, USA

Glossary

Kuznets Curve The U-shaped graphical relationship between a dependent variable (e.g., income inequality, environmental quality, forest cover) registered on the *y*-axis and an independent variable measuring economic growth, located on the *x*-axis. Originally proposed in a 1955 article by economist, S. Kuznets.

Monocrop A single species, and often single variety, to which an agroecosystem is planted.

Shifting Cultivation The farming system which operates within a forest by cutting, drying, and burning a small patch of forest, then planting multiple crops in the ashes for one to several years. When the plot is abandoned, the forest reoccupies the site and is left uncut for an extended period before renewing the cycle.

Introduction

Concern over deforestation of the tropics has increased steadily over the last decades, as increasing numbers of people have come to believe that deforestation and the related processes of global warming and biodiversity loss threaten our future. This concern is mirrored and reinforced by the growing numbers of scientific studies to specify these processes and the potential losses they are creating. Yet, 'deforestation' is an ambiguous term, not only because it encompasses a variety of processes and results, but also because deforestation has been, and continues to be, judged both a positive and a negative process. The soybean farmer in the Amazon says he is only doing what farmers of the developed world have done in the past: converting forests to agricultural fields in order to gain personal wealth and feed the world. His point reminds us that deforestation has a long history, one which must be considered as the context for contemporary processes.

This article, therefore, will proceed by examining the definitions and data by which we can evaluate deforestation in its first section titled 'What is a forest? What is deforestation?' The second section titled 'Patterns of forest conversion and degradation' will sketch the history of deforestation leading to and summarizing contemporary processes. The third section titled 'Environmental and social costs of deforestation' will sketch the social and environmental impacts of deforestation. The fourth section titled 'Causes of forest conversion and degradation' will explore the causes of deforestation, suggesting a nested hierarchy of forces operating at local, national and international scales which interact in specific geographic landscapes and through specific histories to produce deforestation. A fifth section titled 'Policies to preserve forests and forest services' will review policies to improve forest management and some of the potential pitfalls of such policies. Finally, the article will conclude with a summary of major points.

What Is a Forest? What Is Deforestation?

Deforestation is ambiguous because it is used for the 'conversion' of forests to other land uses, for the 'degradation' of existing forests via logging, forest product extraction, or environmental modification, and for all the combinations of these processes. Moreover, how we define a forest will influence how much of it we find: some estimates define 'forest' as areas with only 10% of the land covered with trees. Finally, people have judged deforestation as both 'good' or 'bad', which reveals how the term and concept are embedded in the broader narratives or systems of interpretation which we use to interact with our world and give meaning to our lives.

Concern over the environmental and social impacts of deforestation has grown during the same years that satellite imagery has allowed humans to observe the Earth and record changes. Yet, obtaining data on the extent of forest change is more difficult than one would expect. Analyzing the entire world is a formidable task, so there is a trade off between breadth and depth of coverage. Early satellite imagery was unable to assess the quality of the resource, so degraded or regenerating forests could not be unambiguously identified, and this meant that time series data had to be based on aggregated data which might not show a change in forest area even though new forests had been cleared while abandoned cleared forests had regrown. More recently, as satellite imagery and geographic information system (GIS) technologies have improved, spatially explicit pixel by pixel comparisons over time have become more common, but in tropical areas only India and Brazil are doing such comprehensive monitoring, though other countries have used good data from different time periods to estimate change. Other sophisticated analyses have applied high resolution (30 m pixel) imagery to 10% samples of tropical forests to identify forest conversion and degradation. Yet, recent low resolution imagery continues to have difficulty identifying degradation, especially in dry forests. These remotely sensed data are seldom country specific and since the forest policies and economic opportunities that influence forest conversion and deg radation 'are' country specific, country based data are crucial to identifying and understanding the trends.

The best country based global data on forests comes from the United Nations Food and Agricultural Organ ization's (FAO) Forestry Resource Assessments (FRA) which were made at 5 year intervals between 1948 and 1963, and then resumed with increasing sophistication in 1980, 1990, 2000, and 2005. These rely on country reports by national forest experts of the reporting countries (138 for the 2005 FRA), but because resources vary, not all countries have comparably detailed and reliable information. In the discussion below, the FAO data from the 2005 FRA have been used primarily for contemporary estimates of forest cover and change. The FAO illustrates the problems resulting from how we define 'forest'. Until the 2000 survey the definition of forest varied; it was defined as greater than 20% ground cover of 5 m tall woody plants in the developed countries, but only greater than 10% ground cover in the de veloping world. The 2000 survey extended the 'greater than 10% cover' definition to the entire world, thereby expanding the 'forests' of Russia and Australia.

Patterns of Forest Conversion and Degradation

Contemporary concerns center on forest conversion and degradation in the tropics, and this concern is warranted because over the last quarter century 4–5% of tropical forests have been disappearing each decade. Yet, mature temperate and boreal forests continue to be logged around the world, especially in Russia. Although tropical forest conversion and degradation are estimated to be contributing 20% of the CO_2 being added to the at mosphere and to be threatening mass extinctions at rates not seen since the demise of the dinosaurs, anthropogenic deforestation has been occurring at accelerating rates since humans first gained control of fire.

Forest conversion and degradation rates increased dramatically as human societies shifted from hunting and gathering to fixed field farming and again from feudal agriculture to industrial capitalism. The invention of agriculture and production of agricultural surpluses freed some members of society to specialize in handicraft production, soldiering, and ruling. With population growth and technological advancement, these feudal/ tributary societies converted more and more forests and grasslands to agriculture and agricultural surpluses to luxury items for trade with similarly organized societies. The pattern of interaction accelerated again in the 1490s when one of the world's civilizations, Europe, gained access to all other civilizations, concentrated mercantile surpluses in their own hands, and began to integrate the world into one economic system, increasingly based on traded commodities.

During the eighteenth century a new nickpoint in the upward trajectory of change emerged as a few of the merchants stopped trading handicraft goods and began producing commodities in factories with machines operated by wage labor and fueled by fossil energy. This revolution in production led to a total reorganization of society and of personal experience, as ever expanding numbers of people organized their lives around factory and office work rather than agricultural production, and as an unending stream of standardized, increasingly cheap commodities poured out of factories and into workers' hands. A key change in this new industrial system was the competition driven need to reinvest surpluses into improving the production system and developing new products. This system can only remain healthy with endless growth in consumption.

All these changes led to expanding populations of increasingly voracious consumers, and this led to expanding demands on forests, grasslands, and all the Earth's ecosystems. Timber and naval stores for ship building stripped forests from the lands surrounding the Baltic Sea, and then from eastern North America, Cen tral America, and the Caribbean in the sixteenth through seventeenth centuries. Sugar produced by slaves con verted the forests of the coast of Northeast Brazil and then of the Caribbean islands. Tea expansion in south eastern China, opium in India, cotton in the US south, grains in the central plains of the US and South American pampas, coffee and bananas in Latin America – all were parts of the global transformation of the world's ecosystems. Contemporary deforestation of the tropics is only the latest phase of that long process.

From an ecological perspective these human activities have had two general trends: ecological simplification and cashing in of natural capital. In forests and other ecosystems little modified by human activities, solar energy is captured by plants and flows through myriad pathways to support many species of herbivores, carni vores, and detritivores (organisms that eat and decompose dead plants and animals). Some of the energy is stored in the structure of the ecosystem, like tree trunks, soil organic matter, plant nutrients. Human agroecosystems simplify the systems so that the maximum amount of solar energy flows into the food or raw materials for industrial production. Breeding of domestic species, with its latest genetic engineering phase, produces large areas of genetically identical monocrops, narrowing the diversity further. Moreover, the wood, nutrients, and humus stored in the system are depleted by the con version to and operation of agroecosystem. By 1900, in dustrial processes using fossil energy began to provide energy subsidies, like machinery, fertilizers, pesticides, which produced huge increases in yields, but masked the loss of order and diversity.

From a social perspective, we see ever increasing numbers and percentages of the world's people trading relatively self sufficient production systems for minor supporting roles in the expanding world economy. As the system unfolds, surpluses and wealth continue to concentrate under the control of fewer and fewer entities.

These processes are statistically manifested in shrinking forests and expanding agricultural fields and pastures (**Table 1**). From 1700 to 1980 world croplands expanded almost fivefold from a small base to 1235 million ha, while the world's large forest estate has decreased by 1015 million ha, or 19% (**Table 1**). Some of the agricultural expansion has been into grasslands and some forests converted to farmlands have reverted to forest, especially in Europe and North America. This reforestation trend has continued since 1980 and has been supplemented by forest plantations in some countries, but in most of the world the net result is smaller areas of forest (**Tables 2–4**).

The 2005 Forest Resource Assessment of the FAO shows Europe with the largest area of forest because all of Russia, the country with the world's largest area of forest, is classified as Europe (**Table 2**). Much of Russia's forest is boreal forest, dominated by three, two, or even one species of relatively low stature conifer, but Russia's southeastern regions are covered with more diverse forests of larger broadleaf deciduous trees. Since the end

Table 1 Changes in land cover/land use, 1700 1980

Region	*Vegetation*	*Area in millions of ha (% change from previous time)*					*% Change*
		1700	*1850*	*1920*	*1950*	*1980*	*1790 1980*
Tropical Africa	Forests/woodlands	1358	1336 (1.6)	1275 (4.6)	1188 (6.8)	1074 (9.6)	20.9%
	Grassland/pasture	1052	1061 (0.9)	1091 (2.8)	1130 (3.6)	1158 (2.5)	10.1%
	Croplands	44	57 (29.5)	88 (54.4)	136 (54.5)	222 (63.2)	404.5%
N. Africa Mid East	Forests/woodlands	38	34 (10.5)	27 (20.6)	18 (33.3)	14 (22.2)	63.2%
	Grassland/pasture	1123	1119 (0.4)	1112 (0.6)	1097 (1.3)	1060 (3.4)	5.6%
	Croplands	20	27 (35.0)	43 (59.3)	66 (53.5)	107 (62.1)	435.0%
North America	Forests/woodlands	1016	971 (4.4)	944 (2.8)	939 (0.5)	942 (+0.3)	7.3%
	Grassland/pasture	915	914 (0.1)	811 (11.3)	789 (2.7)	790 (+0.1)	13.7%
	Croplands	3	50 (1666.7)	179 (258)	206 (15.1)	2039 (1.5)	6666.7%
Latin America	Forests/woodlands	1445	1420 (1.7)	1369 (3.6)	1273 (7.0)	1151 (9.6)	20.3%
	Grassland/pasture	608	621 (2.1)	646 (4.0)	700 (8.4)	767 (9.6)	26.2%
	Croplands	7	18 (157.1)	45 (150.0)	87 (93.3)	142 (63.2)	1928.6%
China	Forests/woodlands	135	96 (28.9)	79 (17.7)	69 (12.7)	58 (15.9)	57.0%
	Grassland/pasture	951	944 (0.7)	941 (0.3)	938 (0.3)	923 (1.6)	2.9%
	Croplands	29	75 (158.6)	95 (26.7)	108 (13.7)	134 (24.1)	362.1%
S. Asia	Forests/woodlands	335	317 (5.5)	289 (8.8)	251 (13.1)	180 (28.3)	46.3%
	Grassland/pasture	189	189 (0.0)	190 (0.5)	190 (0.0)	187 (1.6)	1.1%
	Croplands	53	71 (34.0)	98 (38.0)	136 (38.8)	210 (54.4)	296.2%
SE. Asia	Forests/woodlands	253	252 (0.4)	247 (2.0)	242 (2.0)	235 (2.9)	7.1%
	Grassland/pasture	125	123 (1.6)	114 (7.3)	105 (7.9)	92 (12.4)	26.4%
	Croplands	4	7 (75.0)	21 (200.0)	35 (66.7)	55 (57.1)	1275.0%
Europe	Forests/woodlands	230	205 (10.9)	200 (2.4)	199 (0.5)	212 (+6.5)	7.8%
	Grassland/pasture	190	150 (21.1)	139 (7.3)	136 (2.3)	138 (+1.5)	27.4%
	Croplands	67	132 (97.0)	147 (11.4)	152 (3.4)	137 (9.9)	104.5%
Russia/USSR	Forests/woodlands	1138	1067 (6.2)	987 (7.5)	952 (3.5)	941 (1.2)	17.3%
	Grassland/pasture	1068	1078 (+0.9)	1074 (0.4)	1070 (0.4)	1065 (0.5)	0.3%
	Croplands	33	93 (184.8)	178 (89.4)	216 (21.3)	233 (7.9)	606.1%
Pacific developed countries	Forests/woodlands	267	267 (0.0)	261 (2.2)	258 (1.1)	246 (4.7)	7.9%
	Grassland/pasture	639	638 (0.2)	630 (1.3)	625 (0.8)	608 (2.7)	4.9%
	Croplands	5	6 (20.0)	9 (216.7)	28 (47.4)	58 (107.1)	1060.0%
Total	Forests/woodlands	6215	5965 (4.0)	5678 (4.8)	5389 (5.1)	5053 (6.2)	18.7%
	Grassland/pasture	6860	6837 (0.3)	6748 (1.3)	6780 (+0.5)	6788 (+0.1)	1.0%
	Croplands	265	537 (102.6)	913 (70.0)	1170 (28.1)	1501 (28.3)	466.4%

Source: Adapted from Richards, J. F. (1990). Land transformation. In Turner, B. L., II., Clark, W. C., Kates, R. C. *et al*. (eds.) *The Earth As Transformed by Human Action*, pp 179 201. New York: Cambridge University Press.

Table 2 Forest area by world region in 2005 (Ten largest forest-holding countries listed)

Region/subregion	*Forest area (1000 ha)*	*% Global forest area*	*% Land area*
Total Africa	635 412	16.1	21.4
Eastern/Southern Africa	226 534	5.7	27.8
Northern Africa	131 048	3.3	8.6
Western and Central Africa	277 829	7.0	44.1
Democratic Republic of Congo	134 000 (seventh largest)		
Total Asia	571 577	14.5	18.5
East Asia	244 862	6.2	21.3
China	197 000 (fifth largest)		
South and Southeast Asia	283 127	7.2	33.4
Indonesia	88 000 (eigth largest)		
India	68 000 (tenth largest)		
Western and Central Asia	43 588	1.1	4.0
Total Europe	1 001 394	25.3	44.3
Russia	809 000 (largest forest holdings)		
Total North and Central America	705 849	17.9	32.9
Caribbean	5974	0.2	26.1
Central America	22 411	0.6	43.9
North America	677 464	17.1	32.7
Canada	310 000 (third largest)		
United States	303 000 (fourth largest)		
Total Oceania	206 254	17.9	24.3
Australia	164 000 (sixth largest)		
Total South America	831 540	21	47.7
Brazil	478 000 (second largest)		
Peru	69 000 (ninth largest)		
World	3 952 025	100	30.3

Adapted from FAO (2006). *Forest resource assessment 2005*. Tables 2.1, 2.3, Figure 2.3 (pp 16 18). Food and Agricultural Organization. Viewed July 2008 at www.fao.org/forestry/site/fra2005.

of the Soviet era, Russian forests are being logged aggressively and often illegally in eastern Asia, part of a global pattern of unsustainable logging.

South America has the second largest area of forests because it contains the Amazon, three quarters of which remains intact. Brazil has the world's second largest area of forest, and Peru's Amazon forests make it the ninth largest forest holder (**Table 2**). Nevertheless, conversion and degradation of the Amazon forests allow South America's loss (8 153 000 ha) to almost equal the area lost in Africa (8ha) between 1990 and 2005 and makes Brazil the country with the world's largest forest loss, about three eights of the world total (**Tables 3** and **4**). Africa's smaller area of forests have a higher rate of forest loss (–0.63% per year) than South America (–0.47% per year). Asia has had less aggregate forest loss, and even a gain from 2000 to 2005 because China has embarked on a massive program of establishing plantations, over 2 million ha added from 1900 to 2005 (**Table 5**). China's plantation expansion masks Southeast Asia's forest loss, which is the highest rate (– 0.91% per year) of any major subregion. Central America has a higher rate, but only a small area of residual forest (**Tables 2, 3,** and **5**).

The cases of China and of Europe reveal expanding forests and plantations. Indeed, the analysis of 227 countries reporting forest area in the 2005 FRA found only 88 (39%) reporting a decline in forest area, 60 (26%) reporting expanding forests, and 79 (35%) reporting no change. This trend comes either from 'reforestation' of former forest lands through spontaneous regeneration and nurtured protection or from 'afforest ation', the planting of trees and creation of forests on lands that were not recently forested. Forest expansion, called 'forest transitions' by scholars, has been occurring in developed countries for 50 years at least as people leave the poverty of rural areas for urban opportunities. Because of infertile soils, cleared lands in the tropics are often abandoned and sometimes revert to forest. Plantation expansion is different in that it requires greater investment. Plantations are prevalent in countries with large forest product demands and strong govern ments which guarantee secure tenure and create tax or other incentives to reward planting (**Tables 4** and **5**).

Forest transitions are an example of environmental Kuznets Curve phenomena: the U shaped pattern in which measures of environmental quality (the *y* axis) at

Table 3 Net annual change in forest by world region or subregion from 1990 to 2000 and from 2000 to 2005

Region or subrebion	*Change 1990 2000 1000 ha/yr*	*Change 1990 2000%/yr*	*Change 2000 2005 1000 ha/yr*	*Change 2000 2005%/yr*
Total Africa	4375	0.64	4040	0.62
Eastern and Southern Africa	1731	0.71	1702	0.74
Northern Africa	1013	0.72	982	0.73
Western and Central Africa	1631	0.56	1356	0.48
Total Asia	792	0.14	+1003	+ 0.18
East Asia	1751	+ 0.81	+3840	+1.65
South and Southeast Asia	2578	0.83	2851	0.98
Western and Central Asia	+34	+0.08	+14	+0.03
Total Europe	+877	+0.09	+661	+0.07
Total North and Central America	328	0.05	333	0.05
Caribbean	+36	+0.65	+54	+0.92
Central America	380	1.45	285	1.23
North America	+17	Not significant	101	0.01
Total Oceania	448	0.21	356	0.17
Total South America	3902	0.44	4251	0.50
World	8868	0.22	7317	0.18

Adapted from FAO (2006). *Forest resource assessment 2005*. Table 2.4 (p 20). Food and Agricultural Organization. Viewed May 2007 at www.fao.org/forestry/site/fra2005.

Table 4 Ten countries with largest annual net loss or net gain of forest area, 2000 05

Countries with loss	*Annual change loss (1000 ha/yr)*	*Countries with gain*	*Annual change gain (1000 ha/yr)*
Brazil	3103	China	+4058
Indonesia	1871	Spain	+296
Sudan	589	Vietnam	+241
Myanmar	466	United States	+159
Zambia	445	Italy	+106
Tanzania	412	Chile	+57
Nigeria	410	Cuba	+56
Democratic Republic of the Congo	319	Bulgaria	+50
Zimbabwe	313	France	+41
Venezuela	288	Portugal	+40
Total	8216	Total	+5104

Adapted from FAO (2006). *Forest resource assessment 2005.* Tables 2.5, 2.6 (p 21). Food and Agricultural Organization. Viewed May 2007 at www.fao.org/forestry/site/fra2005.

Table 5 Recent increases in plantation forests

Country	*Added 1990 2000 (1000 ha)*	*Added 2000 2005 (1000 ha)*
China	545.8	1489.0
United States	596.9	157.0
Russia	270.9	320.4
Vietnam	108.3	129.0
India	85.1	84.2
Indonesia	79.3	79.4
Chile	61.3	61.4
Australia	46.2	56.2
Turkey	46.5	46.6
Korea	44.0	35.2
New Zealand	50.8	16.6
Uruguay	46.8	19.4
Argentina	30.9	30.2
Myanmar	30.2	30.6
Peru	45.2	7.8

Adapted from FAO (2006). *Forest resource assessment 2005*. Annex 3, Table 10. Food and Agricultural Organization. Viewed May 2007 at www.fao.org/forestry/site/fra2005.

first decrease as a society's average income (the *x* axis) increases, but then level off and begin to improve as the society's wealth continues to increase beyond the inflection point. Forest area and quality are seen to be following this pattern because as societies gain wealth and opportunity, rural populations move to cities reducing subsistence pressures on forests and because those controlling forests apply scientific management to them. Some ecologists have suggested that the pro jected problems of global warming, mass extinctions, and environmental decline will be mitigated by these trends, but others argue that forest transitions only occur when either labor becomes so scarce and expensive that land is abandoned or when the forest resources are so severely depleted that governments encourage individuals and/or communities to initiate and protect forest plantations or natural regeneration. In areas with many underemployed workers and/or large forest resources and/or governments unable/unwilling to maintain order, forest decline will continue. Moreover, other critics note that severely fragmented primeval forests will not regenerate sufficiently to maintain the intact ecosystems needed to preserve biodiversity. What is clear is that plantations will in no way maintain biodiversity. Other concerns about planting trees to offset carbon emissions arise from attempts by the Kyoto Protocol of the UN Convention on Climate Change to use market forces to promote greenhouse gas reductions. The plan is to set up an international market through which carbon emitters can reduce their carbon footprint by paying for projects to reestablish forests in the developing world. Other proposals call for greenhouse gas emitters to pay governments and logging concession grantees not to log, convert, or degrade forests. One well known example in the 1990s was a US electric utility paying the government of Bolivia to establish a 149ha national park in its portion of the Amazon basin. Problems with these schemes are primarily technical but are highly significant: how can an independent agency determine the value of the preserved forest, both in terms of its carbon content and its 'environmental services', like biodiversity preservation? How can an independent agency ensure that the preservation of an area of forest does not induce excessive degradation of nonpreserved or reforested areas nearby? Will governments have the resources and will to monitor these projects effectively? Can governments be assured that participating in these plans will not threaten their sovereignty? Theoretically, these types of projects offer great promise, but they are extremely difficult to implement.

Although forest area continues to decline in only 40% of the world's countries, most of these are in the biologically rich tropics. We have not considered how forest quality is declining. The FRA provides estimates of the area of forests considered 'primary forest' (native species, no evidence of human disturbance), 'modified natural forest' (native species, some evidence of humans), 'semi natural forests' (native species established by planting or assisted natural regeneration), and 'plantations' (see **Tables 6** and **7**). **Table 6** provides the percentages of these categories for the largest 15 countries, and Table 7 shows recent changes in primary forest for the same countries. Much of the Amazon remains in the primary category, while developed countries have most forests in the modified natural category.

Environmental and Social Costs of Deforestation

The local, national, and global impacts of converting and degrading forests and other intact ecosystems are forcing people to recognize the essential services these ecosystems provide. First, intact forests can provide a steady supply of products: foods, construction materials, medicines, fuel, industrial raw materials, fresh water, and

Table 6 Percentages of forests in FAO forest quality classes: 15 largest forest-owning countries

Country (Area in forest million ha)	*Percent primary forest*	*Percent modified natural forest*	*Percent semi-natural forests*	*Percent protective plantation*	*Percent productive plantation*
World	36.4	52.7	7.1	3.0	0.8
Russia (808.8)	31.6	66.3	0	1.5	0.6
Brazil (477.7)	87.1	11.8	0	1.1	0
Canada (310.1)	55.3	46.7	0	0	0
United States (303.1)	34.4	57.9	2.1	5.6	0
China (197.3)	5.9	58.0	20.3	14.5	1.4
Australia (163.7)	3.2	95.7	0	1.0	1.5
Democratic Republic Congo (133.6)	No data	No data	No data	No data	No data
Indonesia (88.5)	55.0	0	41.1	3.8	0
Peru (68.7)	88.8	10.1	0	1.0	0
India (67.7)	0	48.7	46.6	1.6	3.2
Sudan (67.5)	20.0	70.0	2.0	7.0	1.0
Mexico (64.2)	51.1	47.2	0	0.1	1.5
Colombia (60.7)	87.4	12.1	0	0.5	0
Angola (59.1)	0	99.8	0	0.2	0
Bolivia (58.7)	50.0	50.0	0	0	0

Adapted from FAO (2006). *Forest resource assessment 2005*. Annex 3, Table 8 (pp 220 231). Food and Agricultural Organization. Viewed May 2007 at www.fao.org/forestry/site/fra2005.

Table 7 Changes in primary forest in 15 largest forest states, 1990 2005

Country	*Area in primary forest (millions ha)*			*Changes in primary forest (1000 ha/yr)*	
	1990	*2000*	*2005*	*1990 2000*	*2000 2005*
Russia	241.7	258.1	255.5	+1640	532.2
Brazil	460.5	433.2	415.9	2729.3	3466.0
Canada	165.4	165.4	165.4	0	0
United States	105.3	105.3	104.2	1.0	215.2
China	11.7	11.7	11.7	0	0
Australia	No data	5.2	5.2	0	0
Democratic Republic of Congo	No data	No data	No data	No data	No data
Indonesia	70.4	55.9	48.7	1447.8	1447.8
Peru	62.9	62.2	61.1	72.2	224.6
India	No data	No data	No data	No data	No data
Sudan	15.3	14.1	13.5	117.8	117.8
Mexico	38.8	34.8	32.8	395.0	395.0
Colombia	53.8	53.3	53.1	51.1	56.2
Angola	0	0	0	0	0
Bolivia	460.5	433.2	415.9	2729.3	3466.0

Adapted from FAO (2006). *Forest resource assessment 2005*. Annex 3, Table 9 (pp 232 237). Food and Agricultural Organization. Viewed May 2007 at www.fao.org/forestry/site/fra2005.

DNA. Second, forests and other ecosystems also provide environmental services by regulating environmental processes. They can clean the air and water of pollutants. They regulate local, regional, global climates by damp ening daily cycles of temperature and humidity, by transpiring water so that moisture penetrates the in teriors of continents (half of the Amazon's rainfall is thought to come from transpired water), by absorbing and sequestering greenhouse gasses, especially CO_2. They provide habitat for insect, bird, and mammal pollinators of human crops. They regulate pests, buffer the impacts of natural hazards (e.g., mangroves and wetlands), reduce soil erosion, and minimize runoff of nutrients and rain fall. Forests also provide cultural and social services by being sites of spiritual, religious, and esthetic experience, by providing people with habitats within which they develop their cultures, by offering sites for recreation and ecotourism.

Hitherto, people have seen the costs resulting from forest loss as local or at worst regional, costs which can be remedied by reforms at home or ignored when they only affect people in far off lands. However, the growing awareness that global warming and biodiversity loss can personally affect even the affluent has focused attention on these issues. About 20% of the CO_2 gasses entering the atmosphere come from forest burning and de composition, most of which occurs in the tropics. Con versely, expanding forest can absorb and sequester CO_2 and is proposed as one of the major paths to reducing global warming. Expanding plantations is an easy way toward that goal, but it does virtually nothing to preserve biological diversity.

Biologists define three levels of biodiversity: at the genetic, species, and ecosystem levels. Ultimately, all this diversity can be preserved for future human use and for its own sake only by preserving the ecosystems which sustain it. There are attempts to store the diversity of life. As technological agriculture and genetic engineering are eliminating the genetic variability of key crops and livestock, breeding institutions have stored the genetic material of many tens of thousands of crop and domestic animal varieties in refrigerated storehouses, placing the 10,000 year heritage of human breeding experience into a few vulnerable places. Zoos and other storehouses of nondomesticated organisms can also preserve some of the diversity, but it will only be a tiny fraction. Without maintaining intact ecosystems, the projected extinctions will accelerate. Although the world is now experiencing extinction rates between 50 and 500 times higher than those recorded in the fossil record, the continuation of current trends is estimated to raise the rate another 10 times higher.

Causes of Forest Conversion and Degradation

Although no widely accepted, comprehensive expla nation of deforestation has emerged from the hundreds of papers analyzing the processes, the best summaries argue (1) that multiple factors interact to produce forest loss; (2) that both direct actions (like agricultural expansions or logging) and indirect or underlying factors (like gov ernment colonization programs or export crops) operate; and (3) each case is geographically and historically spe cific. This last makes generalizations difficult. We can advance understanding by basing analysis on a nested hierarchy of local, national, and global scales at which

Global Scale: how have forests been affected by:
- Development institutions: how have bilateral and multilateral institutions changed with regard to their roles, importance, and use of grants vs. loans?
- Development theories: how have modernization, basic needs, neoliberal theories affected forest uses?

National–Subnational Scale: how have forests been affected by:
- Development policies like: colonization schemes; promotion of exports; subsidies – to whom and for what?
- Tenure arrangements: Nationalized forests, recognition of communal rights, grants of large blocks of forest to elites for development projects?
- Governance practices: Level of control over state territory; unity of citizenry– level of differences based on wealth, race, ethnicity; relative democratic vs authoritarian political structures; susceptibility to outside pressures; amount of debt?

Local Scale: how have forests been affected by:
- Local environment: climate - length and timing of dry season; soil quality; topography?
- Agricultural expansion:
  - Small holders: traditional shifting cultivators; immigrant slash and burn farmers; permanent fixed field farmers?
  - Large holders: plantation agriculture; cattle ranching; tree plantation?
- Logging: road construction; clear-cutting; selective cutting; other extraction methods?
- Mineral extraction: mining, forest clearing, road building?

Figure 1 Nested hierarchy of causes of deforestation.

the forces influencing forest use operate (**Figure 1**). The actors at these various scales are all influenced by the specifics of geography and history.

Local Scale

At the local scale in the tropics, agricultural expansion is often identified as a major cause of forest conversion and degradation, but who are the actors: small scale subsistence farmers or large scale commodity producers? If subsistence farmers, are they traditional shifting cultivators, immigrant slash and burn cultivators, or permanent field farmers? 'Slash and burn' is a label used without distinguishing between traditional shifting cultivators and recent immigrants who are often unfamiliar with local conditions.

This leads to considering the local environment. Most upland tropical soils are infertile because many of their nutrients have been leached by the heavy rainfall. Tropical forests, unlike mid latitude forests and soils, have most nutrients held within the living organisms, which efficiently capture and recycle them. Clearing and burning the vegetation releases the nutrients, but the closed nutrient cycles are broken and nutrients disappear within a few years. Traditional shifting cultivators clear and burn small patches of forest, plant them to many crops, and abandon the plots so that the surrounding forest can reoccupy the site. They leave the abandoned plot to the regrowing forest for long periods before returning. They have been using these forests in this way for thousands of years, so they are not the cause of tropical forest loss. Outside immigrants, in contrast, settle in areas adjacent to roads, but abandon their fields to clear new forest when fertility declines and weeds take over. They often continue to burn their abandoned fields which seldom revert to forest, so they create the fishbone pattern of deforestation revealed in satellite images. Some tropical soils retain fertility because they form from recent volcanic deposits or on floodplains whose nutrients are replenished annually. Such sites can support permanent agriculture, though the favorable climate allows pests to flourish and induces heavy pesticide applications.

Soil quality is one example of how local conditions influence potential uses of tropical forests. The Amazon basin has few volcanic soils, but parts of Central America and of Southeast Asia do, so permanent agriculture is much more successful in these regions. Climate and topography are other locally specific conditions which influence the productive potential of a site. Is there a dry season? During what time of the year does it fall? If the country is mountainous or hilly, higher elevations have cooler temperatures, which reduces water demand, and higher or lower precipitation depending on location relative to prevailing winds.

Another local and regional factor affecting the likelihood of disturbance is the percentage of trees that are commercially valuable timber. For example, Southeast Asian forests had large numbers of dipterocarp species like teak, while the Amazon and Africa had considerably fewer valuable trees, so logging causes more forest loss in Asia than elsewhere.

Logging is often done when a government sells a 'concession' to a private individual or corporation. These usually specify the species and size of the trees that can be cut; and sometimes they specify the way the timber must be removed. In very few cases are these regulations observed. A World Bank study in 2006 estimated that timber worth $10 billion is taken illegally each year, and this costs governments another $5 billion in lost royalties and taxes. The report listed 17 countries where more than 50% of the timber is illegally logged. Over the last 15 years China has become the center of furniture, flooring, and plywood production, relying on imported wood. As long as the timber has paperwork claiming it was logged legally, they process it. Logging exemplifies well the way different causes of deforestation interact: logging roads open forests to other exploiters: squatters, ranchers, miners. Brazil's Amazon reportedly had 105 000 miles of such roads in 2006, more than 10 times the total of government roads. Roads, mines, pipelines, and other types of development also destroy forests directly and open areas to others.

That wealthy individuals and corporations are able to buy logging rights and ignore tree harvesting regulations reveals the role power and wealth play in deforestation. Land development schemes in the Amazon granted rich individuals several thousand hectare blocks of forest to convert to ranches or soybean farms. The total impact of these few wealthy exploiters is often greater than the thousands of poor seeking to survive on the forest fron tier. This discussion leads to the role of national governments.

National- and Subnational-Scale Forces

Government policies are major causes of deforestation in multiple ways. During colonial times, many governments promoted forest conversion to agriculture in order to increase tax revenue. However, by the mid nineteenth century, as industrial processes spread into the colonies, forest departments began to oppose revenue ministries over whether forests should be converted to agriculture or preserved for industrial uses. This contest developed first in India, as timber for railway construction con sumed huge volumes of wood. The forest department induced the government to nationalize and restrict access to many forests, thereby outlawing essential subsistence uses and alienating subsistence users. Colonial foresters first developed the narratives of deforestation induced environmental destruction that later were used in African colonies and continue to justify excluding traditional forest users around the world.

Forest policies of postcolonial states vary, depending on how much forest they control. States with large holdings (Brazil, Indonesia) promote forest conversion and exploitation, but countries whose forests are depleted (India) have, until the mid 1980s, followed colonial policies outlawing subsistence use. Governments pro moting forest exploitation do so through their general development policies, through land tenure arrangements, and through how effectively they govern their territory.

Development policies can directly support forest exploitation by granting logging concessions, initiating forest colonization projects for the landless poor (Brazil, Peru, Indonesia have done this, usually to avoid re distributing land owned by the elite), or by giving large blocks of forest to wealthy citizens able to develop it (e.g., cattle or soybean estates in Brazil). More ubiquitous and subtle are the indirect ways governments have supported forest use. These include development policies (1) that build roads and other infrastructure into forested regions, (2) that promote exports crops, and (3) that provide tax and other subsidies to forest users.

Land tenure arrangements also affect forest use. Colonial and independent governments claimed all for ests as state property early on ignoring the customary claims of traditional users. As the forest becomes devel opment frontier, three groups often contend for control of the forest: indigenous users, who often had managed communal forests sustainably; immigrant small holders seeking new farm land; and wealthy outsiders seeking to log or establish large scale agricultural operations. Governments have almost always supported the last group by sanctioning their tenure and ignoring their legal transgressions. Since the mid 1980s pressure from international donors and nongovernmental organizations (NGOs) have forced many governments to recognize the tenure claims of indigenous groups.

The quality of state governance also influences forest preservation or conversion. The first consideration is whether the state is capable of controlling its territory. Does it have the financial and personnel resources to govern? Does it have a citizenry unified by a common vision? Most developing countries were created by co lonial occupiers with no regard for territorial viability or integrity. When the resulting country contains fragments of many ethnic groups and few valuable resources, it faces great challenges.

A second factor is how evenly the land and wealth of the country are distributed. Does a small minority control most wealth and the state apparatus? This is a common situation which will inevitably lead to the government ignoring the needs of the majority either directly, via state support to the already rich, or indirectly by failing to enforce the law and allowing 'corruption' to flourish. A third factor is the extent to which the state is demo cratic; this means more than having sham elections: to what extent are the leaders accountable to those affected by their choices? State policies lead to global influences.

Global-Scale Influences

The impact of global actors is already apparent from the discussion above, but explicit discussion will clarify its importance and changing history. During colonial times colonies were affected by the changing demands of the occupier: agricultural commodities and tax revenues before 1850 but thereafter timber and industrial wood. In postcolonial times direct control of production has given way to indirect control via development policies and financing, yet the overall demand for export commodities has remained. Three periods are commonly recognized. In the first period, from 1945 to about 1970, modernization theories and bilateral grants dominated. Modernization encouraged industrialization, often guided by government planning. Agricultural aid sought to develop large scale crops for export. Land redistribution policies were promoted, but implemented in ways that did not threaten elite control. Economies grew, but the wealth did not trickle down, so relative and even absolute poverty expanded. The second period which dominated through the 1970s sought to meet the basic needs of the poor by continuing growth oriented policies while directing the benefits to the poor. During this basic needs period loans replaced grants and multilateral agencies assumed larger roles. Debt burdens of many countries grew rapidly and forced governments to emphasize exports even more strongly. Also during this period environmental concerns intensified in the developed world, and the first concerns about biodiversity loss and global warming were articulated.

By the 1980s debt crises swept the developing world and the third period of neoliberal development theory took hold and continues to guide policies to this day. Neoliberalism holds that the market rather than government planning and intervention could create sustainable development. As debt burdened countries refinanced, the International Monetary Fund (IMF) forced them to adopt structural adjustment policies of the neoliberal model, that is, to offer unrestricted access to foreign investors, to reduce and end tariffs and to terminate government run enterprises. During the neoliberal era, however, concerns over deforestation, biodiversity loss, and global warming have become so powerful that the multilateral and bilateral donors have forced recipient countries to structure their development projects to minimize these problems. Hence, Brazil now requires that for each hectare of forest converted, 2–4 ha must be preserved.

Environmental concerns have interacted with neoliberal theory in interesting ways. Environmentalists and human rights groups have complained that state policies have destroyed forests, while indigenous forest users have managed forests much better than the state; hence, indigenous groups must be given tenure rights to their forests. Donors have forced many states to do this. This devolution of power to local users fits well with the neoliberal goal of privatizing resources, though communal property rights differ from the preferred private property ideal. Claims that they support sustainable development and pressure from citizens and environmental groups have forced donors to recognize that environmental services provided by intact ecosystems (biodiversity preservation, pollination, climate regulation) are externalities the market fails to include in its magic. The World Bank's 2004 forest strategy proposes developing market values for these services and establishing markets to compensate states and individuals for these services when they forego forest conversion. Whether the international donors will implement and/or force governments to follow these requirements remains to be seen, but these seem to be the only hope for preserving forests and the services they provide.

Table 8 Carbon in forest ecosystems of regions of world in 2005 (tonnes/ha)

Region or subregion	*Carbon in living biomass*	*Carbon in dead wood*	*Carbon in litter*	*Carbon in soil*	*Total carbon*
Total Africa	95.8	7.6	2.1	55.3	160.8
Eastern and Southern Africa	63.5	7.5	2.1	No data	73+
Northern Africa	26.0	3.3	2.1	33.5	64.9
Western and Central Africa	155.0	9.8	2.1	56.0	222.9
Total Asia	57.0	6.9	2.9	66.1	132.9
East Asia	37.0	5.0	No data	No data	42+
South and Southeast Asia	77.0	9.0	2.7	68.4	157.1
Western and Central Asia	39.7	3.6	11.4	41.0	95.8
Total Europe	43.9	14.0	6.1	112.9	176.9
Total North and Central America	60.1	9.0	14.8	36.6	120.6
Caribbean	99.7	8.8	2.2	70.5	181.2
Central America	119.4	14.4	2.1	43.3	179.2
North America	57.8	8.8	15.4	35.8	117.8
Total Oceania	55.0	7.4	9.5	101.2	173.1
Total South America	110.0	9.2	4.2	71.1	194.6
World	71.5	9.7	6.3	73.5	161.1

Adapted from FAO (2006). *Forest resource assessment 2005*. Annex 3, Table 2.8 (p 34). Food and Agricultural Organization. Viewed May 2007 at www.fao.org/forestry/site/fra2005.

Policies to Preserve Forests and Forest Services

The previous section presented the best hopes for pre serving forests and their services. If current trends con tinue, there is little doubt that the developing world will follow the example of the developed countries and convert their forests to short term profit generating uses. The hope of establishing preserves to maintain diversity and plantations to sequester carbon may compensate for CO_2 emissions, but will only preserve a small percentage of the diversity (see **Table 8**). The world's 30 000 pro tected areas cover only about 7% of the 250 million ha of forest and if they are well maintained will only save 50–70% of the existing species. However, preservation is a continuing expense which is a burden on developing countries and will be ignored in times of crisis. Inter national financing of biological preserves is also un reliable: it fell by 50% from 1995 to 2001. Ecotourism can provide income to support bioreserves, but is not likely to provide sufficient income to sustain them or the broader array of ecosystems needed. Establishing a system to value and compensate forest owners for not destroying forests is a challenge that remains to be implemented.

Some scholars argue that only indigenous com munities and settled agroforestry farmers have the in centives and skills to preserve forests, but they must be granted secure tenure and be assured that they can gain fair value for the forest products they harvest. Over the last 20 years land tenure reform has given indigenous communities or small agroforestry farmers legal owner ship or management rights over 25% of the developing world's forests. Current markets, which favor large enterprises, must be modified to make capital, infor mation, market access, and appropriate prices available to small holders. Payments for maintaining environmental services must be delivered to forest users.

These changes are being considered by donors and grudgingly accepted by governments. They fly in the face of trends promoting large plantations for carbon sequestration and bioenergy production. Whether they are implemented or not will likely determine the future of forests.

Conclusions

People have been degrading forests and converting them to other land uses since humans first evolved, but the areas converted and degraded have accelerated con tinuously as agriculture was invented, populations grew, and the global economy expanded to incorporate even the most remote communities into its processes. Current forest losses are caused by multiple factors operating at local, national, and global scales, as these processes are influenced by local geographies and histories. National policies, supported by global actors, drive deforestation by encouraging elites to harvest forest wealth through logging, through large scale forest clearing for ranching and plantations, and through maintaining highly con centrated land ownership patterns, thereby driving poor farmers into forests for survival. National governments, again supported by global organizations, establish forest preserves, but they are usually too small and too poorly protected to prevent massive extinctions. Only by en suring that poor forest dwellers have legal management rights to forests and are fairly compensated for sustain ably harvesting forest products will it be possible to block catastrophic loss of forests.

See also: Development I; Development II; Environment; Industrialization; Modernization Theory; Neoliberalism and Development.

Further Reading

FAO (2006). Forest Resource Assessment 2005. Food and Agricultural Organization. http://www.fao.org/forestry/site/fra2005/en (accessed April 2008).
IPCC (Intergovernmental Panel on Climate Change) (2007). Summary for policy makers. *Working Group II Report*: *Impacts, Adaptation, Vulnerability.* http://www.ipcc.ch (accessed April 2008).
Kohm, K. A. and Franklin, J. F. (eds.) (1997). *Creating a Forestry for the 21st Century*. Covelo, CA: Island Press.
Leach, M. and Mearns, R. (eds.) (1996). *The Lie of the Land. Challenging Received Wisdom on the African Environment*. Oxford: International African Institute.
Molnar, A., Scherr, S. J. and Khare, A. (2004). *Who Conserves the World's Forests? A New Assessment of Conservation and Investment Trends*. Washington, DC: Forest Trends. http://www.forest trends.org/documents/publications/Who%20 Conserves long final%202 14 05.pdf (accessed April 2008).
Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. and Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature* 403, 853 858.
Orme, C. D. L., Davies, R. G., Burgess, M. *et al*. (2005). Global hotspots of species richness are not congruent with endemism or threat. *Nature* 436, 1016 1019.
Richards, J. F. (1990). Land transformation. In Turner, B. L., Clark, W. C., Kates, R. C. *et al*. (eds.) *The Earth As Transformed by Human Action*, pp 179 201. New York: Cambridge University Press.
Richards, J. F. and Tucker, R. P. (eds.) (1988). *World Deforestation in the Twentieth Century*. Durham, NC: Duke University Press.
Rudel, T. K., Coomes, O. T., Moran, E. *et al*. (2005). Forest transitions: Towards a global understanding of land use change. *Global Environmental Change* 15, 23 31.
Sayer, J. (ed.) (2005). *The Earthscan Reader in Forestry and Development*. London: Earthscan.
Sharma, N. P. (ed.) (1994). *Managing the World's Forests. Looking for Balance between Conservation and Development*. Dubuque, IA: Kendall Hunt Publishing.
Tucker, R. P. and Richards, J. F. (eds.) (1983). *Global Deforestation in the Nineteenth Century World Economy.* Durham, NC: Duke University Press.
Wertz Kanounnikoff, S., DeLuca, G., Thomas, T., Buys, P. and Chomitz, K. M. (eds.) (2006). *At Loggerheads? Agricultural Expansion, Poverty Reduction and Environment in the Tropical Forests*. Washington, DC: World Bank Publications. http://www.worldbank.org/tropicalforestreport (accessed April 2008).

WRI (World Resources Institute) (2005). *Millennium Ecosystem Assessment*. Covelo, CA: Island Press. http://www.maweb.org/en/index.aspx (accessed April 2008).
Wright, S. J. and Muller Landau, H. C. (2006). The future of tropical forest species. *Biotropica* 38(3), 287 301.

Relevant Websites

http://www.cifor.cgiar.org
Center for International Forestry Research.
http://www.fao.org
Food and Agricultural Organization of the United Nations: Forest Resource Assessment 2005.
http://www.forests.org
Forest Conservation Portal. A Large Inventory of Information and Links.
http://www.forest trends.org
Forest Trends. A Large Inventory of Information and Links.
http://www.ipcc.ch
Intergovernmental Panel on Climate Change.
http://www.ITTO.or.jp
International Tropical Timber Organization.
http://www.millenniumassessment.org
Millenium Ecosystem Assessment.
http://www.rightsandresources.org
Rights and Resources Initiative. Collaborative Site Focused on Community Control of Natural Resources, Especially Forests.

De-Industrialization

A. Pike, Newcastle University, Newcastle-upon-Tyne, UK

Introduction

The contraction and rationalization of manufacturing industry – 'de industrialization' – occupies a central role in the uneven geographical development of capitalism. De industrialization has been experienced unevenly in different countries, regions and localities, and between different industries, firms, social groups, and individuals. Building upon the pioneering work of economists in understanding the phenomenon and its narrowly eco nomic dimensions from the late 1970s, geographers were the first to explain the geographical differentiation of de industrialization over time, across space, and in place. Significant theoretical and empirical advances were provided by revealing the geographical extent of de industrialization through mapping where and when it was happening, and deepening our understanding of its precise nature through conceptualizing the different forms of contraction and varying kinds of economic activities, firms, and jobs being lost. As the geographical reach of de industrialization has extended, new insights are being uncovered into its forms, degrees, and character by geographers working internationally.

De industrialization is an ongoing historical process and its causes and implications are deeply geographical. Adaptation to manufacturing decline in localities and regions is often highly uneven and differentiated. Some places are better able to adapt to economic change than others. Economic inefficiencies emerge, however, as the result of particular forms of adjustment. Productivity in the economy as a whole falls when a skilled manufacturer who loses his or her industrial job ends up in lower value added, lower wage, and lower productivity service employment. Output and relative living standards fall as a consequence. De industrialization imposes often huge social costs on people, households, and families formerly dependent upon manufacturing employment for their livelihoods and well being. Deep seated dislocations caused by concentrated and prolonged local and regional manufacturing decline include: the economic, social, and psychological costs of job loss and unemployment; the risks and anxiety felt by 'reluctant entrepreneurs' forced into self employment or establishing new busi nesses by the absence of alternative local employment opportunities; economic, social, and environmental costs generated by longer distance commuting flows as people are forced to work further away from home; and, com munities being fragmented as people move and migrate in search of work in the face of negligible local job prospects.

Defining and Explaining De-industrialization

What do we mean by de industrialization? At its most basic level, de industrialization refers to the contraction and decline of the weight of manufacturing industry within an economy. This is usually measured in terms of output (the total value produced by the sector) and employment (the number of people working in the sector). A distinction is drawn between absolute declines in output and employment and relative declines in the proportion or share of manufacturing in total output and employment. This is important because in a growing economy, manufacturing may be declining in relative terms but maintaining its absolute size and importance. Another key measure of de industrialization is trade competitiveness. This focuses on the relative share of international markets held by manufacturing exports from particular countries and regions. Where such shares are falling, this reveals declining competitiveness in manufacturing against other producers. Falling demand for output and labor can generate de industrialization. When considering de industrialization as a particular process of change the scale is important – how much has output, employment, or exports fallen – as well as the rate at which any declines have occurred. The magnitude and speed at which de industrialization is experienced shapes its impact and the potential for adaptation in localities and regions.

Explanation of de industrialization has focused on four main sets of arguments. First, the 'maturity thesis' emphasizes the 'natural' evolution of the economy through primary (agriculture, mining) to secondary (manufacturing) to tertiary (services) and then quater nary (knowledge based) stages over time. **Figure 1** depicts this theory in terms of changes in employment and per capita income as part of a transition between sectors through time. De industrialization – falling proportions of employment in industry – is seen to occur alongside expansion of employment in the service sector. De industrialization, then, is interpreted as a direct consequence of the evolution and stage of maturity of the economy. As economies develop through this model, they evolve into more advanced forms of economic activity.

'Smokestack' manufacturing industries – typically char acterized as heavy engineering, shipbuilding, steel, and textiles – are left behind as the economy develops. New service activities expand and absorb the surplus labor ejected from manufacturing. For some, what is at issue is not whether de industrialization is occurring but whether it is occurring fast enough to shift economic resources from outdated and uncompetitive industrial sectors into new and more advanced economic activities – a process Joseph Schumpeter described as 'creative destruction'. The maturity thesis has strong roots in the Fisher–Clark theory of economic evolution and Daniel Bell's 'postindustrialism' thesis. **Table 1** demonstrates how the relative contribution of manu facturing to output has declined to varying degrees in all areas internationally between 1990 and 2003, except East Asia and the Pacific where it has grown from 30% to 36% of GDP and the Middle East and North Africa where it is stable at 13%. Such geographically variable experience supports critics who argue that economic change does not follow the inevitable, linear, and singular route suggested by the maturity thesis. The agency of institutions and people in localities and regions means the outcomes are more contingent with potentially multiple pathways of adaptation, including possible reversals and stasis.

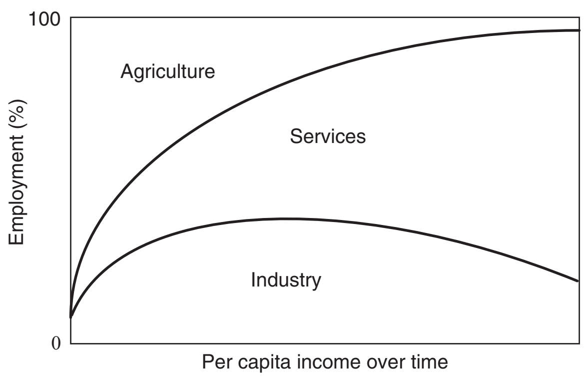

Figure 1 Changes in employment (%) and per capita income over time by sector.

Second, drawing upon the orthodox economic theory of comparative economic specialization and the benefits of trade, the 'trade specialization thesis' emphasizes how places specialize in economic activities where they hold a comparative advantage against other places due to their economic assets and capabilities. Following this logic, for example, relatively low wage economies specialize in labor intensive activities in which this characteristic offers them a comparable economic advantage in com petition with other producers. Economic specialization is shaped by a complex mix of factors, including natural resource endowments and socially and insti tutionally produced capabilities, that can be accumulated and lost over time through shifts in market and industry structure, technologies, competition, and regulation. As depicted in the growth of output in East Asia and the Pacific (**Table 1**), the geographical locus of manufacturing specialization has shifted internationally from West to East in recent decades as part of the new international division of labor. Former regional and local centers of manufacturing have consequently suf fered de industrialization in its wake.

Third, de industrialization has been explained in terms of the 'competitive failure of manufacturing' in its international context. Producers in a specific locality or region are interpreted as uncompetitive against other producers located elsewhere within international mar kets. **Figure 2** illustrates the dramatic growth in the value of merchandise exports from the early 1980s in the 'newly industrializing countries' (NICs) in Central and Latin America and Asia, especially China. This reveals the degree of specialization in export led manufacturing in these countries and intensified competition in export

Table 1 Changes in structure of output (% of GDP) by country type and sector, 1990 2003

	Agriculture		*Industry*[a]		*Manufacturing*		*Services*	
	1990	*2003*	*1990*	*2003*	*1990*	*2003*	*1990*	*2003*
High income	3	2	33	27	22	18	65	71
Europe EMU	*4*	*2*	*34*	*28*	*24*	*20*	*62*	*70*
Middle income	15	10	39	36	24	22	46	54
East Asia and Pacific	*25*	*14*	*40*	*49*	*30*	*36*	*35*	*36*
Europe and Central Asia	*16*	*8*	*43*	*31*		*18*	*41*	*61*
LA and Caribbean	*9*	*7*	*36*	*27*	*22*	*16*	*55*	*66*
Middle East and N. Africa	*14*	11	38	43	13	13	48	47
South Asia	*31*	*22*	*27*	*26*	*17*	*16*	*43*	*51*
Sub-Saharan Africa	*19*	*17*	*34*	*31*	*17*	*14*	*47*	*52*
Low income	32	24	26	27	15	14	41	49
World	5	4	34	28	22	18	61	68

[a] Industry includes mining, manufacturing (columns 6 7), construction, electricity, water, and gas.
Source: Adapted from The World Bank (2005). *World Development Indicators 2005*. Washington, DC: The World Bank.

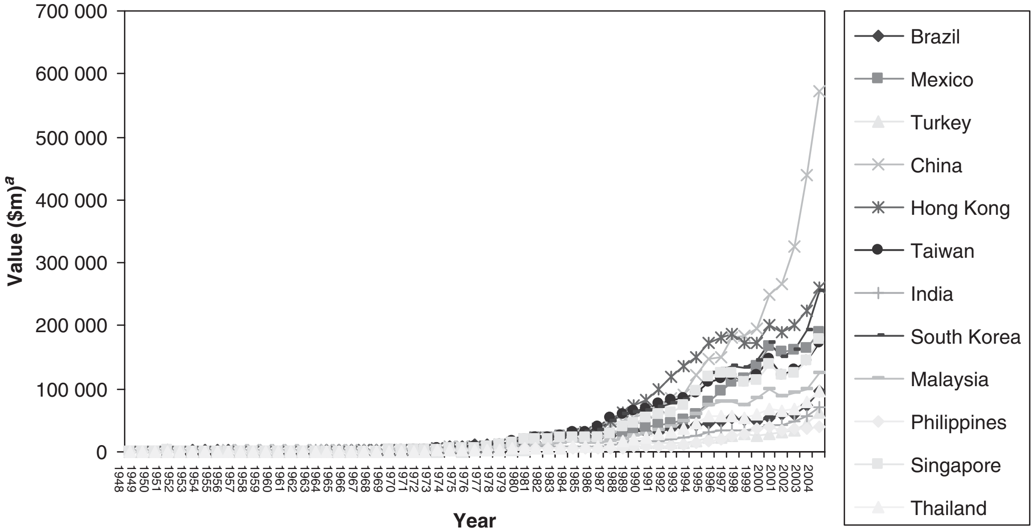

Figure 2 Merchandise exports by value ($m) and country, 1948 2004. [a]$m unadjusted values. Source: Calculated from UNCTAD Statistical Database.

and domestic markets for manufacturers in the historically industrialized countries. Loss of market share, sales, and revenues challenges and potentially undermines manufacturers' profitability and viability. In response, firms attempt to adapt their economic activities, not always successfully, to changes in markets and among competitors. Declines in output and/or employment – de industrialization – can result. Britain's early, large scale, and intense experience of de industrialization in the late 1970s and early 1980s, for example, has been explained by some in terms of a particular 'British disease' comprising weak management, short term investment, poor quality, low productivity, inability to compete with exports from the NICs, and militant industrial relations. For orthodox neoclassical economic geographers, failure to grow or even sustain market share is interpreted as a clear market signal that demand is not present or sufficient for the goods that manufacturers in a particular place are producing. Rational economic factors responding to signals through the market adjustment mechanism will therefore shift investment resources, output, and jobs into activities where demand exists. What is left of manufacturing will be the most technologically innovative and productive that can compete internationally. Manufacturing will be 'leaner' and 'fitter' having shed its inefficient parts.

Last, the 'disinvestment thesis' explains de industrialization as the active disinvestment in manufacturing by firms and, in the case of nationalized industries, states. For firms with increasingly internationalized production networks expanding geographically to take advantage of differential production costs, 'hollowing out' their manufacturing activities can make economic sense. This typically involves a shift to lower value added assembly rather than higher value added manufacture – turning manufacturing plants into 'warehouses' rather than 'factories'. In this way, firms can still access high value and lucrative advanced country markets and bypass any tariff or nontariff barriers to trade on imported, fully assembled goods by importing components and subassemblies manufactured more competitively elsewhere.

Where manufacturing remains relatively capital intensive, as compared to services, and only weakly profitable, it struggles internally to generate sufficient investment capital for its needs and increasingly comes to rely upon the external capital market. In the context of the increasing dominance of the international financial system in shaping access to investment capital, financial institutions have grown in sophistication and in their ability to assess potential returns on capital among parcels of economic activities competing for investment funds. Ostensibly manufacturing firms can therefore be encouraged through their owners, investors, and/or managers to search and switch into activities potentially offering higher relative returns and generating greater 'shareholder value'. Crucially, it is often not that manufacturing is not profitable. It is that manufacturing is not profitable enough in comparison with other economic activities. During the 1980s, Barry Bluestone and Bennett Harrison termed this 'conglomeratization' as manufacturing firms diversified, sometimes unsuccessfully, into often unrelated sectors such as property development in search of relatively higher investment returns to satisfy increasingly demanding capital markets. More recently, geographers have begun to focus upon the impact of new financial vehicles such as hedge funds and private equity in promoting and facilitating such disinvestment in manufacturing. A flight from manufacturing and de industrialization, then, accompanies such disinvest ment and restructuring.

Where manufacturing industry has been state owned, de industrialization can follow the reduction of state ownership – either directly through reduced financial stakes or wholesale privatization or indirectly through the reduction of levels of subsidy or other forms of state aid. Restructuring formerly state owned manufacturing industry has generated de industrialization as firms are rationalized and sometimes closed in a bid to strip out the profitable activities in the context of increased exposure to competitive external markets. Privatization of manufacturing has been an important cause of de industrialization especially in strategic industrial sectors such as defence, including aerospace and shipbuilding, and steel. The starkest recent example is in the formerly centrally planned economies in Central and Eastern Europe. Large scale privatization has occurred as part of the transition to mixed economies of market and state, resulting in huge scale de industrialization and the loss of thousands of jobs.

Table 2 Industrial employment in European Union countries, 1965 95

	Industrial labour force (%)			
	1965	*1975*	*1985*	*1995*
Austria	45	na	38.1	35.4
Belgium	46	39.9	31.8	28.3
Denmark	37	31.5	27.9	27.1
Finland	36	na	31.9	27.9
France	39	38.7	32.4	26.9
F.R. Germany	48	46	41	36
Greece	24	29.2	25.7	23.2
Ireland	28	30.5	30	31.4
Italy	42	39.1	33.5	32.1
Luxembourg		46.3	32	25.5
Netherlands	41	34.6	28.2	22.8
Portugal	31	33.8	33.9	32.2
Spain	35	38.3	31.8	30.2
Sweden	43		29.9	26.6
United Kingdom	47	40.7	34.6	27.4
EU15				30.3

Source: Adapted from Cumbers, A. and Birch, K. (2006). *Divergent Pathways in Europe's Old Industrial Regions?* Draft Paper for SPIF Meeting, 27 October, Glasgow, UK: Centre for Public Policy for Regions (CPPR), University of Glasgow.

Uneven Histories and Geographies of De-industrialization

De industrialization is characterized by highly uneven histories and geographies. The ways in which the process has unfolded is geographically differentiated in time and space. Different localities and regions have experienced de industrialization in different ways and to varying degrees, shaped by their particular histories and insti tutional legacies of industrialization, labor, culture, and politics. De industrialization has become a more widespread geographical phenomenon, extending inter nationally in recent years. De industrialization is mani fest at different geographical scales. National experiences differ and are mediated by the different institutional formations and variegated forms of capitalism in different national political economies. Inter and intraregional transitions at the subregional, local, and urban scales reveal further differentiation, diversity, and variety at finer levels of geographical resolution.

In the midst of the economic dislocations in the late 1970s and early 1980s, the Western European hearths of industrialization during the Industrial Revolution of the mid nineteenth century were among the first to experi ence de industrialization, particularly the UK's north, Scotland's central belt and south Wales, as well as France particularly Lorraine and Nord Pas de Calais, and Ger many especially the Ruhrgebeit and Saarland. **Table 2** reveals the geographically differentiated experience of de industrialization in Europe over time. The older industrialized countries and regions, such as Belgium especially Wallonia, Denmark particularly Jutland, and the Netherlands especially Limburg, began their de in dustrialization earlier in the 1960s. Indeed, between 1965 and 1975, several peripheral European countries ex perienced industrialization, including Greece, Ireland, Portugal, and Spain. Between 1975 and 1995, de indus trialization spread out unevenly and to varying degrees from the older heartlands of industrialism, affecting places such as País Vasco in Spain and Setúbal in Por tugal. By 1995, only Ireland and Portugal had higher proportions of industrial employment than in 1965. Greece too has held its share, despite falling from a peak in 1975.

Anglo American geographers were first to grapple with the uneven geographies of de industrialization. In the UK from the late 1970s, a 'North–South divide' emerged at the regional level as waves of de industrial ization hit, first, the old industrial regions of the north west, north, Yorkshire and Humberside, Scotland and Wales and, second, the post war growth areas in the West Midlands and Greater London. Pervasive processes of change reached deep into the industrial fabric of manufacturing regions negatively to affect networks of small and medium sized subcontractors. At the subregional scale, an urban–rural shift occurred with de industrialization concentrated in London and the provincial cities of Birmingham, Liverpool, Manchester, and Newcastle while re industrialization occurred in the smaller urban centers, new towns, and rural areas, especially in the 'Golden Crescent' from East Anglia around Cambridge across to the 'M4 Corridor' to the west of London.

Such intense experiences prompted debate among economic geographers about the causes and interpret ation of de industrialization. Inspired by neoclassical economics, the 'location factors' approach explained this geography of de industrialization in terms of industrial, firm size, ownership and urban structures, and regional policy. De industrializing places overspecialized in de clining sectors, large and foreign owned or nationalized firms and were concentrated in cities. Neo Marxian analysts situated the changes in structural shifts in Britain's changing post imperial role in the international economy, firms and industry's changing positions in the spatial division of labor and consequent restructuring, urban labor militancy and the potential flexibility of new 'greenfield' plants, especially for feminization of the workforce. Both approaches agreed that regional policy played a role in encouraging de industrialization through modernization and spatial relocation. Specifically, per ipheral regions that benefited from the relocation of nationalized industries as part of regional policy sub sequently bore the brunt of rationalization following privatization.

In North America in the late 1970s and early 1980s, the geography of de industrialization comprised a shift from the 'snowbelt' of the former industrial heartlands of the north and east such as Michigan and Pennsylvania to the 'sunbelt' of the southern states such as Arizona and New Mexico. Older vintages of manufacturing in cities, for example, Los Angeles in California, were affected too. The relocation to the southern states was caused by the relatively lower wage costs for labor of comparable skills and productivity, lower business taxes, 'right to work' legislation preventing strikes and industrial militancy, and 'green', unorganized labor that promised enhanced flexibility for manufacturers. The older industrial heart lands remained important but shrunk in their relative size and weight as new rounds of investment went south. More recently, de industrialization in the northern, eastern, as well as some southern states has occurred through manufacturing relocation to Mexico as activity grew among the maquilladores operations that emerged along the border to serve the US market in the context of the North American Free Trade Agreement (NAFTA). Geographies of de industrialization continued to be highly differentiated locally and regionally in response to such broader pressures for change.

De industrialization has not remained confined to the advanced Western European and American heartlands of industrialization. As an inevitably geographical phenom enon, de industrialization has extended its spatial reach to the former peripheries of the new international division of labor established as assembly and production outposts from the core manufacturing nations and regions, as well as components of national industrialization and develop ment strategies during the 1970s. Indeed, some argue that because of an increasingly 'globalized' economy the 'industrialization–de industrialization' cycle has speeded up due to heightened levels of interconnection and interdependency facilitated by new information and communication technologies, intensified competition, and the geographically extended coordination opportun ities afforded by 'global production networks'.

In Southern Europe, for example, Greece, southern Italy, Spain, and Portugal, work by geographers has re inforced the understanding that de industrialization is a profoundly geographically uneven phenomenon that can occur simultaneously with industrialization – as **Table 2** reveals – and is economically uneven because it does not just affect large scale Fordist or even post Fordist high volume manufacturing. Geographers have demonstrated how manufacturing small and medium sized enterprises (SMEs) in the particular industrial structures of Southern European states experienced de industrialization albeit in differentiated ways. Accession to the EU and the es tablishment of the Single European Market and Eco nomic and Monetary Union in the late 1980s and 1990s has shaped particular experiences of de industrialization through the heightened mobility of capital and labor, intensified competition, corporate concentration and centralization, rationalization, 'Europeanization' and internationalization of supply networks, liberalized pub lic procurement, and reduced state subsidies. Even the paradigmatic local and regional development model of agglomerated industrialization in the industrial districts of the 'Third Italy' around Emilia Romagna is being de industrialized and reconfigured by 'offshoring' and the international decentralization of production.

Following the 'Velvet Revolution' in 1989, the for merly centrally planned economies of Central and Eastern Europe have begun a highly geographically un even transition to mixed economies of market and state. De industrialization in post socialist countries has oc curred on a massive and unprecedented scale and in a highly compressed time period. Historically, the centrally planned economies maintained relatively high levels of industrialization to promote economic growth and pro vide employment and military capacity. Following the collapse of communism, a process of transition has been unfolding through the introduction of market reforms and privatization. A dramatic shake out has occurred as manufacturers and jobs formerly serving protected markets within the Soviet trading system Council for Mutual Economic Assistance (COMECON) were newly exposed to international market competition. The mag nitude of de industrialization is unprecedented, espe cially in employment terms, and highly geographically uneven with differentiated transition paths evident at the national and regional levels. **Table 3** reveals the uneven geography of industrial employment collapse, especially (outside areas of conflict in the Balkans in the 1990s) in

Table 3 Employment in industry by country, Eastern Europe, Baltic States and CIS, 1990 1999 (indices 1989 100)

Country	*1990*	*1999*
Eastern Europe[a]	*95.6*	*60.9*
Albania		
Bosnia and Herzogovina	98.3	28.9
Bulgaria	91.0	46.7
Croatia	102.4	58.3
Czech Republic	95.8	73.4
Hungary	97.0	66.4
Poland	93.7	69.5
Romania[b]	96.5	49.5
Slovak Republic	95.7	65.4
Slovenia	95.1	66.7
Macedonia	95.3	55.4
F.R. Yugoslavia	100.9	65.3
Baltic States	*96.8*	*54.0*
Estonia[c]	96.9	63.6
Latvia	97.0	45.7
Lithuania	96.7	55.7
CIS	*98.2*	
Armenia	102.6	40.5
Azerbaijan	97.1	53.6
Belarus	98.6	76.2
Georgia	104.2	
Republic of Moldova[c]	102.4	35.9
Russian Federation	97.7	62.6
Ukraine	98.1	54.4
Total above	*97.4*	
Former Soviet Union	98.2	

[a] Excluding Albania.
[b] End of year.
[c] Excluding Transdniestria since 1993.
Source: Adapted from UN/ECE Common Database data in Philipov, D. and Dorbritz, J. (2003). *Demographic Consequences of Economic Transition in Countries of Central and Eastern Europe*. Population Studies No. 39, DGIII Social Cohesion, Strasbourg, France: Council of Europe Publishing.

the Republic of Moldova, Armenia, Latvia, Romania, Bulgaria, and Ukraine, for example, in its industrial centers in Donetsk and Dnepropetrovsk. Those that have experienced de industrialization to a lesser degree in clude Belarus, Czech Republic, Poland, Slovenia, the Slovak Republic, and Hungary, buoyed by simultaneous industrialization through the wholesale or partial acqui sition of indigenous industry and relocation of typically foreign owned manufacturing capacity from Western Europe. The post Soviet bloc is highly heterogenous, however, with relatively 'successful' reformers Czech Republic, Hungary, and Poland experiencing de indus trialization alongside a rapid growth of services. Whereas Moldova, Romania, and Ukraine, as well as to a lesser degree Bulgaria and Russia have seen an increased employment share in agriculture following the decline in industrial employment but only sluggish service em ployment growth. In these places, a return to subsistence forms of economic activity is a rational economic response to the calamitous collapse in local industrial output, income, and job opportunities but may be per ceived as a backward step in terms of the linear stages model of economic evolution and modernization de scribed by the maturity thesis.

Beyond Europe, the fast growth NICs from the 1970s have since experienced uneven de industrialization in distinct waves. In Asia, pioneering work by geographers demonstrated how the rapid appreciation of the Japanese currency, the yen, during the 1980s – the 'endaka' phe nomenon – undermined the export competitiveness of Japanese manufacturers and, by cheapening imports, triggered decentralization of their production networks more broadly within East Asia. This accounts for part of the growth in industrial employment in East Asia and the Pacific in **Table 1**. Crucially, this decline in manu facturing output and employment affected not just the large original equipment manufacturers but also their hierarchically organized tiers of SME subcontractors. **Table 4** illustrates how Japan has continued to experi ence de industrialization in terms of output and employment through the 1990s, especially in the indus trial belts of Kanagawa, Osaka, Tokyo, and Saitama. In later waves, de industrialization has hit both South Korea especially in employment terms and in places such as Busan and Ulsan in the south, and older industrial centers in Taiwan within and beyond Taipei. De industrialization has created unprecedented challenges for these developmental states and their export led industrialization strategies that they are only now struggling to come to terms with. Prior to its incorpor ation into China in the late 1990s, Hong Kong had experienced de industrialization from the 1980s through the 1990s, especially the small locally owned manu facturers, where manufacturing collapsed from 17% to 5% of GDP and from 37% to 27% of men's jobs and from 27% to 10% of women's jobs in the context of an export dependent economic structure, rapid tertiariza tion, and a noninterventionist state (**Table 4**). Australia too experienced contraction of its share of industrial employment, especially in steel and motor vehicles, following its deeper integration into the international manufacturing division of labor and evolution into a service dominated economy. De industrialization is evi dent in Asia Pacific as at least labor intensive production shifts geographically to emerging centers of manu facturing specialization and growth in China – whose share of manufacturing output rose from 33% to 39% of GDP in a fast growing economy between 1990 and 2003 (**Table 4**) – and India, as well as Cambodia, Indonesia, Laos, Thailand, and Vietnam.

In parallel from the 1980s, Central and Latin America, including Mexico, Argentina, and Colombi a and especially Brazil and Uruguay, have experienced

Table 4 Manufacturing output (% of GDP) and industrial employment (% of total male and female employment) by selected country, 1990/92 2002/03

Country	*Manufacturing output (% of GDP)*		*Industrial employment[a] (% of total male and female employment)*			
	1990	*2003[b]*	*1990 92*		*2000 02*	
China	33	39				
Singapore		28	36	32	31	18
Argentina	27	24	40	18	30	12
South Korea		23	40	28	34	19
Japan	27	21	40	27	37	21
Uruguay	28	19	36	21	32	14
World	*22*	*18*				
India	17	16				
Colombia	21	14	35	25	19	17
Australia	14	12	32	12	30	10
Brazil	25	11	27	10	27	10
Hong Kong	17	5	37	27	27	10

[a] Includes mining and quarrying (including oil production), manufacturing, construction and utilities (electricity, gas, and water).
[b] Entries are listed in decreasing order of manufacturing output for 2003.
Source: Adapted from The World Bank (2005). *World Development Indicators 2005*. Washington, DC: The World Bank.

de industrialization to differing degrees. Brazil, particu larly, has seen manufacturing's share of GDP and em ployment fall from 25% to 11% and from 37% to 10%, respectively, between 1990 and 2003 (**Table 4**), especially in its industrial heartland in São Paulo. Deepened eco nomic integration, including the establishment of the MERCOSUR trading bloc, international market and trade liberalization, and the reduction of tariff and non tariff barriers have all underpinned a reduction in the relative weight of manufacturing in total output and employment. The shifting geographical locus of trade specialization in labor intensive manufacturing to the Asian nations of the Pacific Rim has further reinforced such changes in Central and Latin America.

Recent evidence suggests de industrialization is becoming even more international, even global, in its geographical reach. Localities and regions industrialized as part of earlier waves of development in the emergent economies China and India, as well as even some parts of Africa are beginning to experience de industrialization. Sukti Dasgupta and Ajit Singh call this 'premature de industrialization' in the developing countries where declines in the proportion of manufacturing jobs in total employment are happening at much lower levels of per capita income than occurred in today's historically industrialized developed countries.

The Politics of De-industrialization

For those localities and regions affected, de industrial ization raises fundamental questions about politics, pol icy, and the roles of localities and regions within the broader spatial division of labor. Contesting and resisting de industrialization has a geographically uneven and differentiated history, shaped by political cultures, tra ditions, and institutions. Geographers identified the fundamental tension between class and territory as bases of social organization in responding and seeking to shape industrial change. This work was highly innovative in situating campaigns to save a specific plant in a particular place in the context of placing other plants in other places at risk of rationalization. As social relations were stretched across space through the spatial division of labor, defending manufacturing in one place could in advertently mean betraying class by deflecting rational ization elsewhere. During the 1970s and 1980s, laborism and trade unionism were relatively stronger, especially in the industrial cores of the developed world, and political activism in concert with local states and communities against de industrialization was more evident and wide spread. In more recent years, such social agency in localities and regions facing de industrialization has often been replaced by greater degrees of fatalism and acknowledgment of their fate in the context of the growing volatility of investment flows and overwhelming economic and political constructions of 'globalization'. Colored by their experiences in failing to prevent and perhaps only slowing down acute industrial adaptation, localities and regions have developed strategies of coping with de industrialization. Policy and institutional innov ation and experimentation have occurred. Task forces have become a standardized response to manufacturing rationalization to coordinate economic development activities, including the sale of businesses and sites and addressing labor market and supply chain adaptation through retraining and market diversification.

The politics of de industrialization raise normative questions about what direction and form the local and regional economy should be taking and the role of state and other institutions in hastening or managing manu facturing's orderly decline or nurturing and sustaining its continued growth. Geographers have innovatively drawn on evolutionary and institutional economics in seeking to interpret whether and how traditional in dustrial regions adapt to new economic development trajectories. Older industrial regions internationally are seen to suffer from different varieties of 'lock in' – economic, cognitive, political – that impart strong path dependencies upon their future development tra jectories. These may constrain or prevent places from collectively renewing and/or modernizing their existing industrial structures or diversifying and moving into new activities, potentially reorienting existing, as well as developing new knowledges, skills, and assets to shift localities and regions formerly dependent upon manufacturing onto development trajectories based upon new economic activities. Preventing or delaying such adjustment is seen by some to be slowing down inevitable evolution in economic structures. Localities and regions experiencing de industrialization face tough challenges about the nature and course of their future direction. Challenging the conventional wisdom of the maturity thesis that de industrialization is inevitable or desirable are arguments that underline manufacturing's distinctive contribution to local, regional, and national economies.

Coda: Manufacturing (Still) Matters

In the wake of its absolute and/or relative decline, policy debate has focused upon the importance of manu facturing to economies, particularly when de industrial ization appears to show few signs of abating even among those first gripped by its effects. Some argue that the transition toward services is desirable, inevitable, and evidence of the development and maturity of economies. Ushering in and addressing the adjustment costs of a transition to a post industrial economy are seen as the appropriate aims of policy. Making things is consigned to an earlier, less advanced stage of economic evolution. Production is best undertaken by those localities and regions specialized and cost competitive in manu facturing internationally. Services, it is argued, are potentially more knowledge intensive, specialized and sophisticated, higher value added and less vulnerable to imitation and cost based competition in a globalizing economy.

Others claim that manufacturing remains vital to local, regional, and national economic growth. Their arguments cite Nicholas Kaldor's famous axiom that manufacturing is the 'flywheel of growth' because the rate of growth of manufacturing output tends to drive the rate of productivity growth in manufacturing and in services. This is because manufacturing has a much greater potential for technological progress and prod uctivity growth, especially relative to services, and has the potential for dynamic increasing returns to scale and the ability to absorb surplus labor from other sectors. The growth of tradeable services – so called 'invisibles' such as insurance – has grown substantially especially in advanced economies in recent years. However, the export and trade of manufacturing goods remain relatively more important to the national econ omy's ability to pay its way in the world through earnings from the overseas sales of its exports, maintaining its balance of payments between income and expenditure in its national accounts, managing inflation, and sus taining the value of the exchange rate of its currency internationally.

The post industrial arguments ignore another two dimensions fostered by the ways in which manufacturing has been changing in recent years. First, some parts of manufacturing have become more knowledge intensive and sophisticated. Even so called 'smokestack' sectors, such as shipbuilding, involve highly complex microelec tronic systems and software and, for example, textiles are increasingly overlapping with nanotechnologies in bio medical applications. Where comparative advantage is based upon innovativeness, technology, and organization, manufacturing can retain a significant high value added role within the economy requiring highly specialized and skilled labor. Manufacturing can, then, be much more productive and employ many fewer people in more skilled jobs than hitherto. Second, the traditional dis tinction between 'manufacturing' and 'services' has become somewhat blurred. Think about the complex bundle of services wrapped around the manufactured commodity of the mobile phone, digital music player, or laptop computer. Manufacturing entities are increasingly looking to more sophisticated and higher value added services – for example, branding, subcontracted on site support, logistics, web based networks, after sales, recycling – in which to embed their products as a means of providing differentiation, sustaining competitiveness, enriching profitability, and locking customers into longer term relationships and repeat business. Such ser vices, crucially, are dependent upon the manufactures at their heart. Despite the ongoing waves of geographically differentiated de industrialization being experienced globally, these commodities still have to be made by someone somewhere.

Acknowledgments

Thanks to Costis Hadjimichalis, Stuart Dawley, and John Tomaney for constructive and incisive advice on the content and Andy Cumbers, Danny MacKinnon, and

Kean Birch for discussions about adaptation in old in dustrial regions. The usual disclaimers, as always, apply.

Further Reading

Blackaby, F. (ed.) (1978). *De Industrialisation*. London: Heinemann.

Bluestone, B. and Harrison, B. (1982). *The Deindustrialization of America: Plant Closings, Community Abandonment and the Dismantling of Basic Industry*. New York: Basic Books.

Bosch, G. (1992). *Retraining Not Redundancy: Innovative Approaches to Industrial Restructuring in Germany and France*. Geneva: ILO.

Cohen, S. S. and Zysman, J. (1987). *Manufacturing Matters: The Myth of the Post Industrial Economy*. New York: Basic Books.

Cumbers, A. and Birch, K. (2006). *Divergent Pathways in Europe's Old Industrial Regions?* Draft Paper for SPIF Meeting, 27 October, Glasgow, UK: Centre for Public Policy for Regions (CPPR), University of Glasgow.

Daniels, P. (2006). On services and economic geography. In Bagchi Sen, S. & Lawton Smith, H. (eds.) *Economic Geography: Past, Present and Future*, pp 112 125. London: Routledge.

Dasgupta, S. and Singh, A. (2006). *Manufacturing, Services and Premature De Industrialisation in Developing Countries: A Kaldorian Empirical Analysis, Working Paper 327*. Cambridge: Centre for Business Research, University of Cambridge.

Dawley, S. (2007). Making labour market geographies: Volatile 'flagship' inward investment and peripheral regions. *Environment and Planning A* 39(6), 1403 1419.

Dicken, P. (2006). *Global Shift: Reshaping the Global Economic Map in the 21st Century* (4th edn.). New York: The Guilford Press.

Edgington, D. (1994). The geography of Endaka: Industrial transformation and regional employment changes in Japan, 1986 1991. *Regional Studies* 28(5), 521 535.

Evans, Y. and Smith, A. (2006). Surviving at the margins? De industrialisation, the creative industries and upgrading in London's garment sector. *Environment and Planning A* 38(12), 2253 2269.

Grabher, G. (1993). The weakness of strong ties: The lock in of regional development in the Ruhr Area. In Grabher, G. (ed.) *The Embedded Firm: On the Socio Economics of Industrial Networks*, pp 255 277. London: Routledge.

Hadjimichalis, C. (2006). The end of the third Italy as we knew it. *Antipode* 38(1), 82 106.

Harvey, D. (1996). *Justice, Nature and the Geography of Difference*. Oxford: Blackwell.

Haslam, C., Williams, J. and Williams, K. (1990). The hollowing out of British manufacturing. *Economy and Society* 19(4), 456 490.

Hassink, R. and Shin, D. H. (2005). Guest editorial: The restructuring of old industrial areas in Europe and Asia. *Environment and Planning A* 37, 571 580.

Hudson, R. and Sadler, D. (1986). Contesting works closures in western Europe's industrial regions: Defending place or betraying class? In Scott, A. & Storper, M. (eds.) *Production, Work, Territory*, pp 172 193. Winchester, MA: Allen and Unwin.

Lever, W. F. (1991). Deindustrialization and the reality of the postindustrial city. *Urban Studies* 28(6), 983 999.

Liu, T. L. and Chiu, S. (1993). Industrial restructuring and labour market adjustment under positive non interventionism: The case of Hong Kong. *Environment and Planning A* 25(1), 63 79.

Lloyd, R. and Sazanami, H. (eds.) (1991). *Industrial Change and Regional Economic Transformation: The Experience of Western Europe*. London: Harper Collins.

Lloyd, R. and Sazanami, H. (eds.) (1989). *Deindustrialization and Regional Economic Transformation: The Experience of the United States*. Boston: Unwin Hyman.

Martin, R. and Rowthorn, B. (eds.) (1986). *The Geography of De Industrialisation*. Basingstoke: Macmillan.

Massey, D. and Meegan, R. (1982). *The Anatomy of Job Loss*. London: Methuen.

Massey, D. and Meegan, R. (1985). *Politics and Method*. London: Methuen.

Mickiewicz, T. and Zalewska, A. (2002). *De industrialisation. Lessons from the structural outcomes of post communist transition, William Davidson working paper No. 463, January*. Ann Arbor, MI: William Davidson Institute, University of Michigan.

Morris, A. S. (1996). Regional industrial promotion in Argentine Patagonia. *Tijdschrift voor Economische en Sociale Geografie* 87(5), 399 406.

Philipov, D. and Dorbritz, J. (2003). *Demographic Consequences of Economic Transition in Countries of Central and Eastern Europe*. Population Studies No. 39, DGIII Social Cohesion, Strasbourg, France: Council of Europe Publishing.

Pieper, U. (2000). De industrialisation and the social and economic sustainability nexus in developing countries: Cross country evidence on productivity and employment. *Journal of Development Studies* 36(4), 66 99.

Pike, A. (2002). Task forces and the organisation of economic development: The case of the north east region of England. *Environment and Planning C* 20, 717 739.

Pike, A. (2005). Building a geographical political economy of closure: The closure of R&D Co in north east England. *Antipode* 37(1), 93 115.

Pike, A. (2006). Shareholder value versus the regions: The closure of the Vaux Brewery in Sunderland. *Journal of Economic Geography* 6, 201 222.

Pike, A., Rodríguez Pose, A. and Tomaney, J. (2006). *Local and Regional Development*. London: Routledge.

Pitelis, C. and Antonakis, N. (1998). *Manufacturing and De Industrialisation: Theory, Evidence and the Case of Greece*. Cambridge: Judge Institute of Management Studies.

Rodríguez Pose, A. and Tomaney, J. (1999). Industrial crisis in the centre of the periphery: Stabilisation, economic restructuring and policy responses in the Sao Paulo metropolitan region. *Urban Studies* 36(3), 479 498.

Sawers, L. and Tabb, W. K. (eds.) (1984). *Sunbelt/snowbelt: Urban Development and Regional Restructuring*. Oxford: Oxford University Press.

The World Bank (2005). *World Development Indicators 2005*. Washington, DC: The World Bank.

De-Localization

A. Kalogeressis and L. Labrianidis, University of Macedonia, Greece

Glossary

Foreign Direct Investment An investment involving a long-term relationship and reflecting a lasting interest and control by a resident entity in one economy (foreign direct investor or parent enterprise) in an enterprise resident in an economy other than that of the foreign direct investor (FDI enterprise or affiliate enterprise or foreign affiliate).

Outsourcing Outsourcing is the delegation of tasks or jobs from internal production to an external entity (such as a subcontractor). Most recently, it has come to mean the elimination of native staff to staff overseas, where salaries are markedly lower. This is despite the fact that the majority of outsourcing that occurs today still occurs within country boundaries.

Subcontracting Subcontracting is the manufacture of goods by one firm (the subcontractor) for another (the lead firm) based on the specifications of the latter. Often there can be several layers of firms or intermediaries mediating the relationship between the actual production workers and the end product market. The lead firms normally exercise considerable control over their subcontractors in terms of price, quality and timing of the products they supply.

Introduction

Originally delocalisation was conceived as another variant of the long list of terms referring to the splitting of a production process, including but not limited to disintegration, internationalization, intra mediate trade, intra product specialization, kaleidoscope comparative advantage, multistage production, outsourcing, slicing up the value chain, splintering, subcontracting and vertical specialization. Some of these terms, such as subcontracting and internationalization were, of course, already in use with slightly or completely different meanings. Others, such as kaleidoscope comparative advantage and slicing up the value chain were too idiosyncratic to gain wide acceptance. Finally, others – delocalisation being one of them – gradually took on a quite different meaning.

Therefore, delocalisation is a term referring to the spatial restructuring of industry at a national, regional or global scale. Its primary elements are Foreign Direct Investment (FDI) and outsourcing, although it also refers to all other types of cross border business interactions. The direction of the movement is usually from the more developed to the less developed countries. In other words, delocalisation is a term that is 'wider' than the other terms, which, at least in mainstream international economics, are identified with and limited to trade in intermediate products. To give a simple example, any industry that is moving, either directly or indirectly, its most labor intensive sections abroad is delocalizing. Broadly speaking FDI could be considered the direct movement, while subcontracting or outsourcing account for the indirect movement.

Before analyzing delocalisation, let us first try to find our way through the 'forest' of terms and try to produce a taxonomy of firm organization (**Table 1**). Let us suppose that we have a domestic firm producing a single product using two intermediate goods (1 and 2 respectively, the former being labor and the latter knowledge intensive), domestic capital and domestic labor.

Let us start with a vertically integrated domestic firm. In this case both intermediate goods are produced in house using domestic capital and labor. Under this configuration, the firm can only be linked with the world market through its final good exports, or raw materials imports.

The first possible deviation occurs when the firm realizes that, for a number of reasons a foreign market can be better served by producing the product there rather than exporting it. This implies a duplication of the production process as additional plants are established to supply different locations: a horizontal FDI.

On the other hand, when the firm realizes that, for example intermediate good 1 can be produced more efficiently in a less developed (and lower labor cost)

Table 1 A Taxonomy of delocalization

Location of production	*Production*	
	Internalized	*Externalized*
Home country	Production kept in-house at home (vertically integrated domestic firm)	Outsourcing (production outsourced to third-party firm, e.g., a subcontractor)
Foreign country (delocalization)	Vertical or horizontal FDI	International outsourcing
For service industries Offshoring	Intra-firm (captive) offshoring	Outsourced offshoring

country, it may decide to set up a plant there to produce it. Although such movements have traditionally been described as vertical FDI, quite recently the alternative term intra firm or 'captive' offshoring has entered the International Business (IB) vocabulary. The differences between the two terms are almost inexistent, although the latter tends to apply more to service industries, while the former to manufacturing ones.

Alternatively, the firm may decide that it is in its best interest to focus on its core competences, which we may assume are better employed in the production of inter mediate good 2. This could imply that the firm wishes to stop producing intermediate good 1. Assuming that the latter is still essential to the firm, it would probably outsource it to either a local firm in a foreign country, or to an affiliate of another Trans national Corporation (TNC).

The inherent difficulties in defining the terms are now more or less evident. One question that arises is whether firms that traditionally bought the intermediate product (i.e. never produced it in house and therefore never stopped producing it) are outsourcing. Another issue is related with subcontracting: is it merely a special case of outsourcing or a completely distinct category? Finally, one last issue ignored by this relatively simple taxonomy – however very important in our context has to do with the governance; are relationships based on formal, contractual agreements, the outcome of specific social environments, trust and embeddedness?

The changing Geographies of Production

In the 60 years between the end of the 2nd WW and today, the global map of production of goods and services has changed significantly. Although three fourths of the global manufacturing still takes place in Developed countries, the share of Developing countries has risen considerably; from 5% in 1953 to almost 24% in 2001. Another significant feature of the world production map is that a very small number of countries produce a sig nificant part of the global output. In 2003 the 15 most significant producers contributed 78% of the global manufacturing value added (**Figure 1**).

A simple illustration of the extent of the inequalities is the fact that the manufacturing value added of the Russian Federation (the last country in our top 15 list) is more than the sum of the 86 countries at the bottom of the list!

The situation within the various groups of countries also changed significantly. Concerning the Developing countries, almost all the change in the region's per formance was due to the rapid increases in a relatively small number of countries in S.E. Asia. Led by the four Asian 'tiger' economies of S. Korea, Taiwan, Singapore and Hong Kong, which were later followed by a number of other countries, the share of the region in world manufacturing value added (see **Figure 2**) more than tripled during the last two decades. Among the countries of the region, it was China that displayed the most spectacular performance, recently overtaking Germany as the third most significant manufacturing producer in the world.

In contrast, the performance of the other subgroups of the Developing countries group was rather disappointing. Specifically, the Developing countries of America lost ground, producing approximately one third of S. & E. Asia in 2001, while only two decades ago the situation was reverse. On the other hand, Africa and the Developing countries of W. Asia and Europe were and still are marginal in the global map of production.

Within the Developed countries there have also been considerable changes. North America emerged as the dominant region with more than 30% of the world manufacturing value added in 2001, outperforming Western Europe during the last decade. The performance

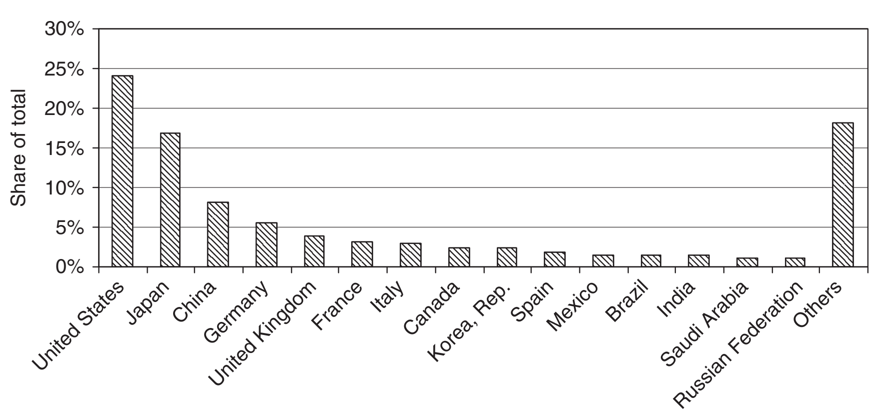

Figure 1 The 15 most significant producers (industry, value-added constant 2000 US$). Data from World Bank WDI database (http://devdata.worldbank.org/dataonline/).

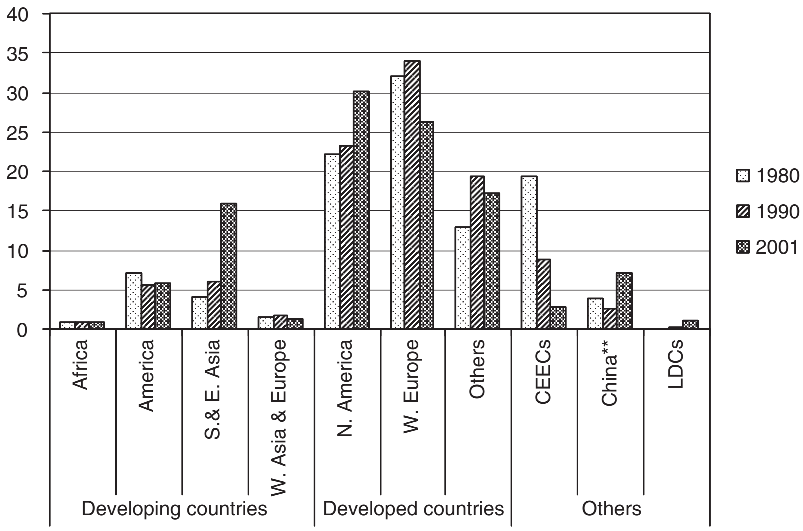

Figure 2 Evolution of world manufacturing value added. Source: Data from UNCTAD Globstat (http://globstat.unctad.org/html/index.html).

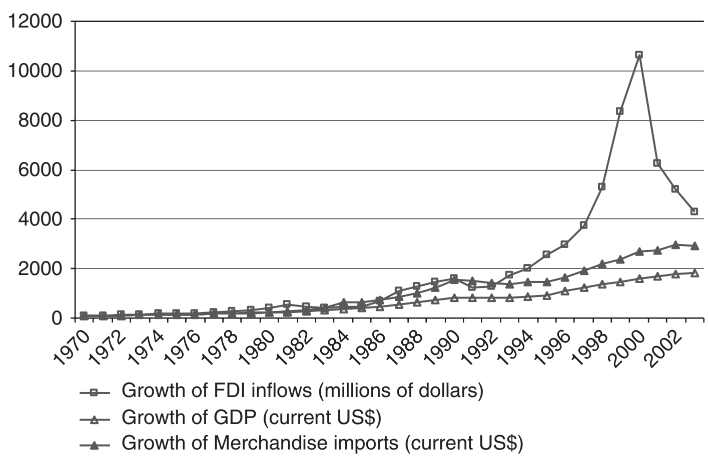

Figure 3 Evolution of FDI, GDP and imports (1970 100). Sources: Data from World Bank WDI database (http://devdata.worldbank.org/dataonline/) for imports and output, and UNCATAD FDI Database (http://stats.unctad.org) for FDI inflows.

of the 'other' Developed countries seems to depict the slump of the Japanese economy during the 1990s.

However, the regions where the most impressive (although very different in direction) changes took place are the CEECs and China. The former appear to have experienced a dramatic crisis, currently accounting for less than 3% of the world manufacturing value added (down from almost 20% in 1980). China, on the other hand, is currently the fastest growing economy.

FDI Growth and Distribution

Although with significant fluctuations, FDI has grown enormously during the last 35 years, and especially, since 1985. More specifically, during the period 1970–2003 the average rate of change of FDI inflows was 14.7%, growing significantly faster than exports (11.2%), and even faster than output (9.3%).

Until 1982 world output and trade grew at about the same rate (**Figure 3**). Since then, trade started growing significantly faster than GDP. In 2003 trade was 29 times its 1970 level, while output had only grown 18 times.

More importantly, since 1993 (almost a decade after trade started growing faster than output), trade growth was in its turn superseded by FDI growth (**Figure 3**). FDI growth was in fact so spectacular, that after three decades of growth, in 2000 (its peak year) it was more than 100 times higher than its 1970 level, with more than 80% of that growth occurring during the last decade.

TNCs and accordingly FDI (outward, **Figure 4**) as well as inward) are still concentrated in Developed countries. Low labor costs alone are not sufficient for a

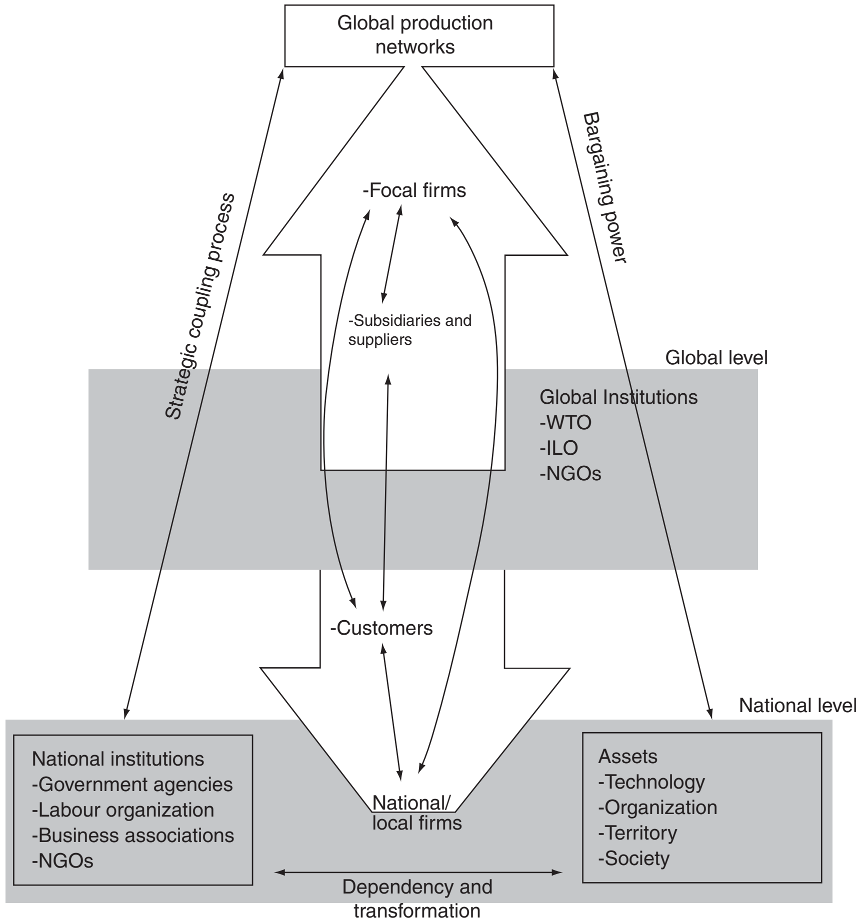

Figure 4 (A conceptual framework for analyzing the impacts of delocalization. Source: Coe, N. M. *et al.* (2004). Globalizing regional development:A global production networks perspective. *Transactions of the Institute of British Geographers*, 29(4), 468 484.

country to attract FDI. There are other more important factors, including for example physical and non material infrastructure, socio economic stability and human capital. In 2004, the inward FDI stock of Developed countries amounted to $6 766 per capita, while the respective figure for Developing countries was only $438. Furthermore, the stock of outward FDI of Developed countries was $9 005 per capita, while the equivalent figure for Developing countries was only $223.

Another important aspect regarding FDI is its distribution between developing countries. **Table 2** provides a number of quite interesting hints about the distribution of FDI within developing countries:

- The 20 most important FDI recipients absorb an extremely large share of the total inward FDI (84.1% in 2004).
- although the concentration is very significant, it nevertheless is gradually being reduced. However, this does not imply that more countries are becoming significant players (in 1990 the countries receiving more than 1% of the total inward FDI stock were 20, while in 2004 they were 21). It rather appears as a consolidation of the place of established countries.
- East and South east Asian countries along with a small number of Latin American countries appear to have benefited the most. On the contrary, African countries are clearly underrepresented.

The Growing Importance of Outsourcing

The problems related to the definition of outsourcing, briefly analysed above, are reflected in the difficulties to measure it. In all types of efforts, outsourcing is measured in terms of trade flows or as shares of total trade flows. Intra firm trade (trade between parents and affiliates in a TNC network) is one of the measures often used to capture outsourcing. As in the case of FDI, most of the empirical evidence comes from the USA, Japan and Sweden. In these three countries, intra firm trade accounts for 38–45% (in the case of the USA), 30% (Japan) and 50% (Sweden) of the total trade of the parent companies.

Table 2 The 20 most significant developing FDI recipients

Country	1980 Share (%)	Rank	1990 Share (%)	Rank	2004 Share (%)	Rank
China, Hong Kong SAR	15.4	1	12.4	1	20.5	1
China	0.9	25	5.7	5	11.0	2
Mexico	0.7	31	6.2	4	8.2	3
Singapore	5.3	5	8.4	3	7.2	4
Brazil	12.7	2	10.2	2	6.8	5
Russian Federation					4.4	6
Bermuda	3.0	8	3.8	7	3.5	7
Korea, Republic of	1.0	23	1.4	19	2.5	8
Chile	0.9	30	2.8	11	2.4	9
Argentina	4.0	6	2.4	15	2.4	10
Thailand	0.8	27	2.3	17	2.2	11
Malaysia	3.5	7	2.8	10	2.1	12
South Africa	9.8	3	2.5	13	2.1	13
Venezuela	1.2	20	1.1	20	2.0	14
China, Taiwan Province of	1.7	16	2.7	12	1.7	15
India	0.4	45	0.5	34	1.7	16
Cayman Islands	0.2	60	0.5	31	1.6	17
Turkey	5.9	4	3.1	8	1.6	18
Nigeria	2.0	15	2.3	16	1.4	19
Viet Nam	0.9	22	0.5	35	1.3	20
Total	85.9		86.5		84.1	

Source: UNCTAD FDI Database (http://stats.unctad.org).

The second, and much more popular, measure is the various approaches of intermediate inputs. One such approach has been the effort to estimate the level of 'Vertical Specialisation' (VS) as a measure of the imported inputs in products that are then exported. This measure is a subset of all intermediate inputs, as some of these are not embodied into export products. In studies of OECD and developing countries increasing shares of VS trade in the majority of countries have been observed. Furthermore, it has been argued that about 30% of the growth of exports of these countries is attributed to VS.

The globalisation of services

While the delocalisation processes in services and manufacturing are similar, there are important differences too. First, although the services sector is much larger than the manufacturing sector, only some 10% of its output enters international trade, compared with over 50% for manufacturing. Second, the pace of Globalisation of services affected by the tradability revolution is faster than manufacturing. Third, whereas the relocation of goods production has involved, overwhelmingly, firms in manufacturing, service functions are located abroad by companies in all sectors. Fourth, the skill intensity is generally higher for tradable services than for manufacturing located abroad, thus affecting white collar jobs in particular. And finally, services located abroad may be more footloose than relocated manufacturing activities because of lower capital intensity and sunk costs, especially services that do not require high skills.

The Globalisation of services reflects nothing less than a revolution in the tradability of services. Traditionally, most services have been non tradable in that they require buyers and sellers to be in the same place at the same time. Unlike physical products they could not be traded between parties located in different countries; a haircut, for instance, is impossible to deliver across a distance. Many services however, do not require physical proximity, but have usually taken place face to face because of technical constraints, habits or customs. These services center on exchange, storage, processing and retrieval of information broadly defined.

New ICTs are dramatically changing the tradability of the information centered set of services, in several ways. For example, all kinds of information can be stored by digitization. And much cheaper and faster transportation allows the instantaneous exchange of digitized information and voice communication between people anywhere around the globe. The use of ICT allows knowledge to be codified, standardized and digitized, which in turn allows the production of more services to be split up, or "fragmented," into smaller components that can be located elsewhere to take advantage of cost, quality, and economies of scale or other factors. This makes it possible to produce certain services in one location and consume them (or use them in further production) in another – either simultaneously (e.g. information provided via call centers) or at different time (e.g. data entry, software development).

Services delocalisation is not simply a North – South issue. *A significant share of relocation takes place among*

Developed countries (i.e. Canada, Ireland and various W. European countries remain among the most attractive locations). Lower wages are not the only driver of services relocation. The phenomenon is essentially a manifestation of a shift in production in response to comparative advantage. It offers all the advantages and costs of such a shift. Services relocation is not a "zero sum game," in which one party (the countries receiving service work) gains at the expense of the other party.

Interesting parallels can be drawn with the relocation of jobs in ICT related manufacturing, when assembly operations shifted to East Asia. Together with technical progress, the Globalisation of hardware cut the prices of ICT products, increased investment in ICT hardware and contributed to higher productivity and growth. Many new jobs emerged in the US to integrate ICT equipment into the workplace and such jobs grew twice as fast as overall employment.

A Conceptual Framework

In discussing a conceptual framework for delocalization, there are two distinct questions to be answered. Both have been adequately addressed by quite distinct schools of thought, however, there have been precious little efforts to bring them together. The first is 'why', while the second is 'how' are firms delocalizing.

The first question has been the main subject of a number of theories falling into the wider International Business (IB) discipline, the most influential of which have been the internalization theory and Dunning's eclectic (or otherwise known as Ownership, Locations, Internalisation – OLI) approach.

Concerning the second question, one needs to start looking at how the economy is organised from production to consumption. Although there have been numerous efforts it is Gereffi's Global Commodity Chains (GCCs) and the Global Production Networks (GPNs) that have been the most influential paradigms (particularly the former).

The conceptual framework proposed for analysing delocalisation is based on the GPN approach and has *four main analytical dimensions* that are intersected by wider categories i.e.

1. The firm, with its own unique set of resources and competitive advantages
2. The sector, with its given technologies and markets
3. The "environment" (local/ regional/national) with its unique institutions, civil society, history and policies
4. The global "environment" with its unique institutions, governance and power relations

Finally, the wider categories intersecting the four analytical dimensions are value, embeddedness, and power.

Furthermore, in order to understand how delocalisation affects development (which is identified with value creation, enhancement and capture) one could look at the interplay between three large groups of variables: assets, institutions and GPNs.

Each country or region is endowed with (or has created) a set of assets, some of which are ubiquitous, while others are more or less exclusive to the country.

Firms lie at the other apex of the triangle. These may be the focal firms, their subsidiaries or suppliers or their customers. However, this can also be expanded to firms that are not part of GPNs or GCCs. Firms operating within bi national or regional networks, which are most usually considered as quite irrelevant to not only the more mainstream IB or neoclassical approaches, but also to the more 'socialised' network approaches, need to be seriously considered, as they often support or outright substitute GPNs in their role of growth enhancement. Specifically, a region's (and therefore a nation's) assets (in fact the economies of scale and/or scope implied by the specific assets) can bring growth inasmuch as they complement the strategic needs of trans local actors situated within GPNs. This strategic coupling (of the firms' of GPNs' strategic goals and the regions' economies) is therefore of primary importance to development. Nevertheless, we should try not to ignore the possibilities of a decoupling process that can be equivalently detrimental to development. This possibility points to the importance of embeddedness to the sustainability of development. Here the term embeddedness is used in a very precise way, in order to capture the extent to which an activity becomes fixed in a particular place and, as such, contributes to local/national economic development through both direct and indirect spinoffs (including backward linkages with local suppliers)'. It is clear then that even local firms can become dissembedded.

Finally, we have the institutions at all levels of governance (national, regional or local). During the last decades, there has been a growing devolution of political and economic power from the central state to local and regional institutions. This has resulted to a multitude of configurations of institutions, affecting both the regional or national sets of assets and the bargaining power vis à vis the TNCs.

The result of this interplay between countries or regions with their assets and institutions and firms trying to take advantage of the assets is development, which is conceptualised as value that is created, enhanced and captured. Assuming that value is created in the country, the big question then is how to enhance this value and, more importantly, how to capture it.

Finally the two levels of analysis (the global and the national/local), as well as the three tangible dimensions parallel to the two levels, i.e. the firms, the assets and the institutions, get transcended by the value created in the production and consumption of goods. The territorial embeddedness of the firms, as well as the power they possess within their respective networks, more or less defines how much of this value will be captured by the regions.

It is clear that these factors combine in a rather individualistic way, consequently rendering prescriptions extremely difficult.

A Case Study: The Delocalisation of the Garment Industry in the Southern European Periphery

This section is a case study of the delocalisation of the garment industry through N. Greece, and particularly the city of Thessaloniki and its neighbouring prefectures.

Figures 5–8 describe the various stages in the transformation of the sector. In order to better comprehend the various movements, we have divided the whole process into four quite distinct stages, which, although highly stylised, reflect the reality of the vast majority of the sector. During the first stage (1970–80 – **Figure 5**) orders to Greek subcontractors started coming from W. European (mainly German) and to a lesser extent American lead firms. This happened for a number of reasons, such as the country's relatively low labour cost, as well as its privileged status as a trade partner of the EU. These orders were normally executed within the company, although in some cases, there were also 2nd layer subcontractors within the same city (e.g. Thessaloniki), so as to lower labour costs.

The change that takes place during the second stage (1980–1990/91; **Figure 6**) comes from the Greek subcontractors, since they gradually develop quite extensive subcontracting networks of 2nd layer subcontractors within the city and mainly in the surrounding villages. Although a considerable part of the less labour intensive tasks are still performed in house by the first level subcontractors, the simpler tasks are assigned to 2nd layer subcontractors, who, in their turn often develop their

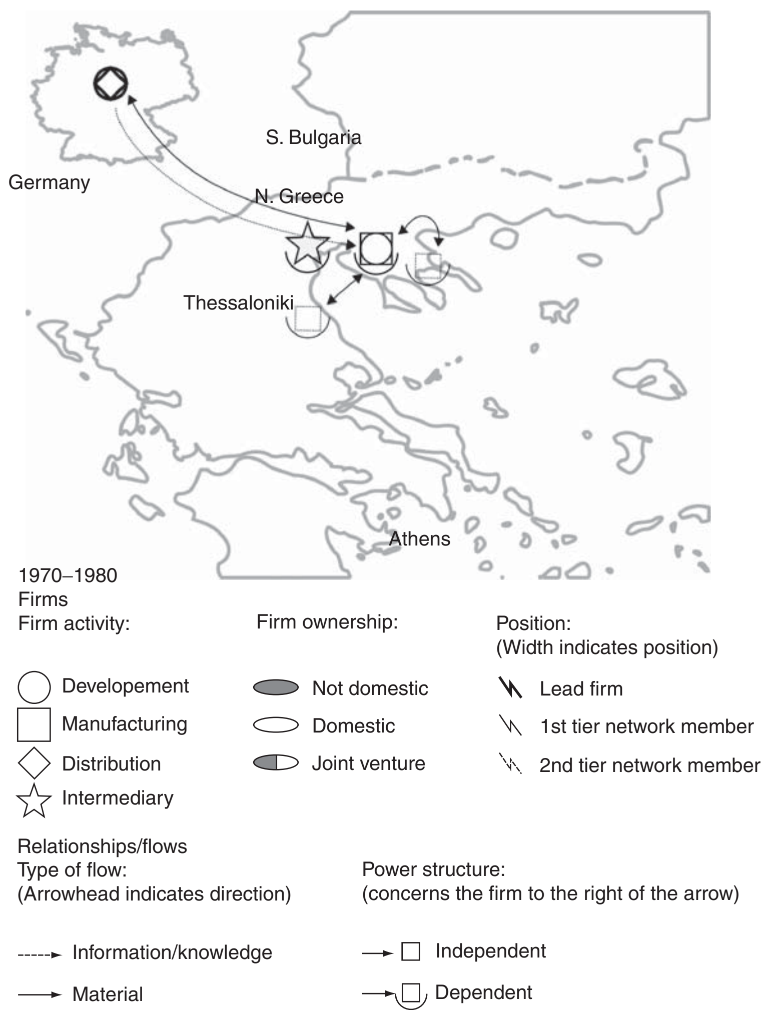

Figure 5 Stage 1 of the delocalization of the garment industry through Greece.

Figure 6 Stage 2 of the delocalization of the garment industry through Greece.

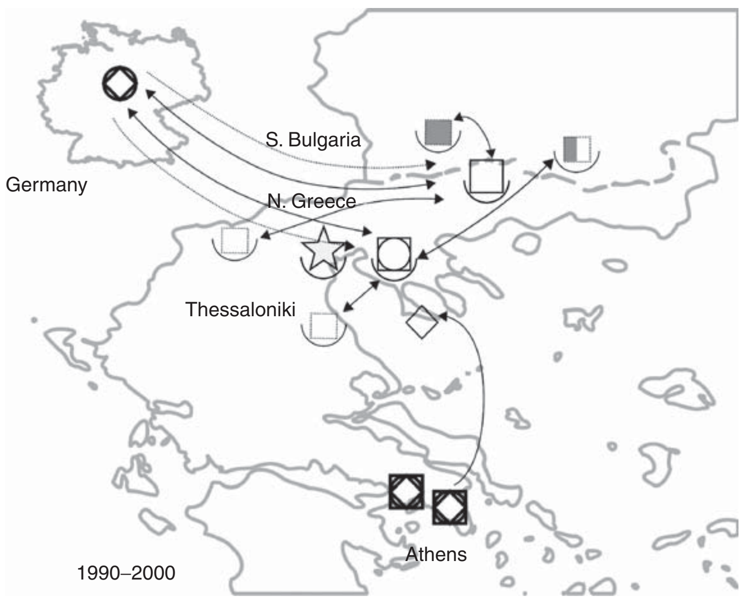

Figure 7 Stage 3 of the delocalization of the garment industry through Greece.

own – smaller – networks of 3rd level subcontractors, often homeworkers.

The third stage (1990/1–2000 – **Figure 7**) is the ‘triangular’ manufacturing stage. Lead firms in e.g. Germany or the US, continue placing their orders to their established Greek partners. However, soon after the collapse of the state socialist regimes, the latter are beginning to transfer parts of their activities to 2nd layer subcontractors in the Southern parts of the neighbouring countries, particularly Bulgaria, and FYROM. By the end of this period, most Greek garment firms are mainly responsible for the organisation of the network. It is interesting to note that for variable periods (ranging from several months to few years) the lead firms were un informed of this development. This gave first movers a considerable advantage, since they were paid Greek prices for Eastern European labour costs. Finally, during the final stage (2000 – today), more direct links are gradually developed between Germany and Bulgaria.

The Greek – German – Bulgarian ‘triangle’ highlight a number of issues which are of relevance to the understanding of delocalisation. Firstly, extreme mobility

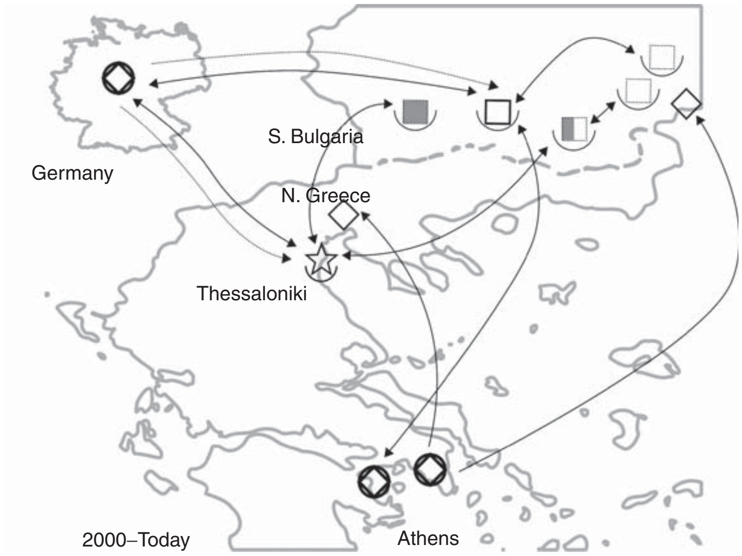

Figure 8 Stage 4 of the delocalization of the garment industry through Greece.

often co exists with various degrees of locational sticki ness. In our case, it seems to have taken many of the German lead firms more than 30 years to abandon their Greek partners. Developing a relationship with a new contractor usually takes a long time; thus lead firms may prefer to follow their established patterns wherever they decide to shift to a new country and keep on ne gotiating the conditions of their relations rather than looking for new suppliers. Moreover, dependence and asymmetrical power relations may co exist with mutual confidence. These two factors can explain, at least in part, why Greek firms continue to form one of the apexes of the German Greek – Bulgarian garment manufacturing triangle. Of course, the role of proximity in the creation of the Greek Bulgarian cluster has been immense. In fact, the Greek Bulgarian borders highlight the case that in tense delocalisation in the border areas can provide a fertile ground for companies that otherwise would not be able to go international. On the other hand, the very fact that even very small companies can easily go inter national can create a shock, at least in the short term, to the local economy of the home country.

However, when firms or sectors fail to functionally upgrade or diversify, there seem to be limits to the degree of stickiness, since, sooner or later, the costs of main taining established relations exceed those of establishing new ones. So, the second point is related to the devel opment implications. In **Figure 8** it becomes evident that the garment sector is in a sharp recession phase in Thessaloniki and the surrounding area. Not only are German firms increasing establishing direct links with Bulgaria, but local producers never really managed to gain a foothold in the local, not to mention national market, which during the last few years is increasingly dominated by firms originating from Athens. In other words, functional upgrading for the cluster's firms was a real rarity. How did this come to happen? This is cer tainly not a case of obstacles imposed by the buyers. What seems to be the answer to this question is the fact that Greek subcontractors became locked in into what was, at least initially, a very rewarding arrangement. Almost overnight these firms broke into foreign markets, eschewing the considerable sunk cost necessary in order to become exporters. In fact, most of these firms never even searched for new clients. In a sense, the obstacles to upgrading were imposed by the firms themselves and their unfortunate short termist view of success.

This, of course, emphatically highlights the extreme difficulty to generalise both about the course of delocalisation, as well as its implications.

Implications of Delocalisation

The issue of the implications of delocalisation is very interesting in the sense that the literature appears to largely ignore the fact that, to a significant extent, similar questions have been asked and answered in the past. So, in both the academic and non academic literature the threat of jobs relocation to lower labor cost countries often emerges as an unprecedented phenomenon. The rise of global outsourcing has triggered waves of con sternation in advanced economies about job loss and the degradation of capabilities that could spell the dis appearance of entire national industries. It is not the first time the international activities of national firms is causing such concerns, since almost 40 years ago (1968) the first report concerning the impact of FDI on the source country, prepared by a Commission headed by W. B. Reddaway, was published. It was followed by a very large number of papers studying almost every possible aspect of the impacts of FDI including the impact of FDI on exports, employment, industry structure and domestic investment.

Delocalization has both advantages and disadvantages for host and home countries. In h*ost countries* in particular: FDI in export oriented services offers a number of economic development benefits such as increased export earnings, job creation, higher wages and upgrading of skills.

Concerns have been expressed that the growth of not only blue, but also white collar jobs in host countries signals employment losses in the home countries (i.e. Developed countries) and potential harm to their economies. However, somewhat paradoxically, the latter are expected to reap a considerable part of the gains from relocating. Benefits include lower prices to consumers, expanding markets for exports and higher corporate profits. A study has found that for every dollar spent on services relocation to India, firms in the US reaped $1.12 $1.14 in benefits. Hence, there are good reasons for Developed countries to welcome the trend. This does not mean that this process will necessarily be smooth; there are bound to be short term challenges for policy makers, especially in terms of adjusting to the restructuring taking place in response to shifting comparative advan tage. What is not clear here is who uses this money and for what purposes. As when production remains domestic then this money is passed to the employees and then in the domestic economy, whereas in the latter case the final recipient of this money may be the US company, but it is not at all clear where the money is spent, invested or taxed. So it is vital to distinguish between the micro and the macroeconomic levels.

Not surprisingly, because of the important role of out sourcing and subcontracting, most of the concerns about delocalisation center on jobs, wages, and skills in the countries involved. In fact, the concerns are so grave that one could easily get the impression that this is the end of the world as we know it. This is certainly an optical illusion.

There is an over exaggeration of the number of jobs that have been lost by Developed countries. For example, in the US, only 2.5% of all job losses during the first quarter of 2004 were the result of relocation. 850 000 jobs may move from US. But even though this number sounds alarming it represents a relatively small slice of the 131 million people employed in the US. Technical change is a far more important cause of job losses. Bank tellers, answering services and secretaries are replaced by ATMs, voice answering technologies and word processing soft ware. Those are two parallel phenomena, both of vital importance for labor intensive industries. However, technological change has been around for several dec ades, whereas delocalisation is a new trend – at least in its current form.

On the other hand, there is an over exaggeration again of the benefits in employment created in the countries hosting such activities. For example, India's high tech boom, even with an estimated 700 000 computer software professionals, still involves less than 0.2% of the nation's more than 400 million of economically active population.

In turn, what is most concerning about delocalisation is that it's diverting the attention of business leaders from building successful, innovative companies to integrating the firm's specific assets that they control to simply finding cheaper inputs. Of course there are innovators who are using delocalisation to get cheaper inputs in a way that is really speeding up their innovation cycle as well.

See also: Foreign Direct Investment; Globalization, Economic; Networks; Globalization and Transnational Corporations; Transnationalism and Labor Geography.

Further Reading

Blomstrom, M. and Kokko, A. (1996). How foreign investment affects host countries. *World Bank Policy Research Working Paper No. 1745*.

Coe, N. M. *et al*. (2004). Globalizing regional development: A global production networks perspective. *Transactions of the Institute of British Geographers* 29(4), 468 484.

Dicken, P. (2003). *Global Shift: Reshaping the Global Economic Map in the 21st Century* (4th edn.). London/New York: Sage Publications/ Guilford Press.

Dicken, P., Kelly, P. F., Olds, K. and Yeung, H. W. C. (2001). Chains and networks, territories and scales: Towards a relational framework for analysing the global economy. *Global Networks* 1, 99 123.

Dunning, J. H. (1993). *Multinational Enterprises and the Global Economy*. Workingham: Addison Wesley.

Feenstra, R. C. (1998). Integration of trade and disintegration of production in the global economy. *Journal of Economic Perspectives* 12(4), 37 50.

Gereffi, G. and Korzeniewicz, M. (eds.) (1994). *Commodity Chains and Global Capitalism*. Westport, CT: Praeger Publishers.

Hirst, P. and Thompson, G. (1999). *Globalization in Question: The International Economy and the Possibilities of Governance* (2nd edn.). Cambridge: Polity Press.

United Nations Conference on Trade and Development (UNCTAD) (1993). *World Investment Report: Transnational Corporations and Integrated International Production*. New York: United Nations.

United Nations Conference on Trade and Development (UNCTAD) (2004). *World Investment Report: The Shift towards Services*. New York and Geneva: United Nations Publications.

Relevant Websites

http://www personal.umich.edu/~alandear/glossary/
The Terms of Trade and Other Wonders: Deardorff's Glossary of International Economics.

http://www.unctad.org
United Nations Conference on Trade and Development.

http://www.unctad.org/Templates/Page.asp?intItemID=1888&lang=1
United Nations Conference on Trade and Development Statistical Databases on line.

http://worldbank.org/
The World Bank.

Democracy

C. Barnett, The Open University, Milton Keynes, UK
M. Low, London School of Economics and Political Science, London, UK

Introduction

Democracy has a simple meaning: 'Rule by the people'. However, the meaning of 'the people' and of 'rule' is far from straightforward. The historical geography of democracy involves the ongoing process of finding answers to various practical problems: Who belongs to the demos? How should rule be organized? What is the scope of the activities over which democratic rule should be extended? Disagreement over basic definitions, concepts, and institutions is constitutive of democracy's dynamism, inclusiveness, and widespread appeal.

These questions of the scope of rule and the delimitation of the people raise important geographical questions. In large territorial states, the question of rule raises issues of the spatial reach and territorial organization of the mechanisms of popular rule. Is democratic rule limited only to electoral practices and representative mechanisms, or should it extend to more participatory practices such as citizens' juries or participatory budgeting? Likewise, the question of the people raises issues of how to define membership of a polity. Should cultural criteria of shared belonging to a community prevail? Or should criteria of place of residence be used? Or should a principle of affected interest be used to define the identity of those who should have a say in shaping decisions?

Democracy has not been an explicit object of analysis in geography until very recently. The heightened interest in this topic is in part a reflection of real world processes of democratization, and awareness of the worldly relevance of electoral geographies to the exercise of power across the globe. It is also, however, a reflection of shifts in the normative paradigms which underwrite self consciously critical human geography. Until recently, there has been little explicit focus in human geography on the normative questions that are at the core of debates about the relationship between democracy and spatiality. The revival of political philosophy since the 1970s was more or less systematically marginalized by more radical perspectives, often dismissed as too 'liberal'. However, the 1990s saw a shift in various subdisciplines toward investigating the way normative issues central to democratic theory played out in practice. This is reflected, for example, in discussions of issues of citizenship, deliberation, and participation in areas such as development geography, urban studies, planning studies, and environmental studies. Democracy now provides the rallying call of even the most radical of geographical analyses of neoliberal capitalism. The absence of democratic politics is now recognized as a key factor in the reproduction of social injustice and inequality, and the exposure of vulnerable or marginalized groups to serious harm. At the same time, the terms of democratic politics have been reconfigured by innovative mobilizations by social movements. At a more abstract theoretical level, the styles of spatialized ontology that geographers have pioneered have opened up new understandings of 'the political'. This has drawn the discipline into debates informed by a distinctive strand of contemporary political theory that focuses on the agonistic, dissensual aspects of democracy. Aspects of democratic theory are now present across human geography. This work can be divided into two broad areas of concern: research into and evaluations of liberal democracy; and explorations of a variety of alternative forms of democracy.

Geographies of Liberal Democracy

Liberal democracy is usually taken to refer to forms of institutionalized popular representation, involving periodic mass election of representatives to authoritative legislatures (and in many cases of executive heads of state), under conditions of free speech and association. This is a model unevenly developed in the post Enlightenment West, and widely emulated and/or imposed in other parts of the world. Liberal representative democracy is usually legitimated by appealing to the basis of politics in the preferences and liberties of private individuals recognized as equal citizens under the rule of law. There is an enormous field of literature, largely in political science, devoted to examining the conditions under which such liberal democratic institutions develop, as well as to particular aspects of these systems, including political parties, elections, voting behavior, citizenship, and the media. There is also a substantial body of liberal political theory associated with the justification of institutions of liberal democracy, a body of work that often provides key critical insights into the operations of actually existing liberal democracies. Both these fields of research – political science and political theory – tend to assume a given spatial formation of democracy, consisting of territorialized national states, internally divided in some way or another.

Modern democratic theory has developed around a specific geographical imagination. This assumes that democracy is framed by bounded territories, involving a nested hierarchy of scales contained within the nation state. Thinkers in this tradition have focused considerable attention on the geographical organization of democratic politics in complex, spatially extensive territories. Research on liberal representative democracy in geography has focused on two areas: electoral geographies and geographies of democratization.

The subfield of electoral geography has consistently focused on aspects of how geography and democracy are related in liberal polities. Much of this research has been focused on explaining and mapping the spatial distribution of votes. Geographers have demonstrated how the spatial organization of electoral systems effects how votes are translated into representative majorities in liberal democracies. Recent attention has focused on developing more sophisticated, contextually sensitive explanations for voting behavior. Research on electoral processes has broadened out to include the geographies of campaigning, party formation, and political communication. The spatial organization of formal democracy therefore has consequences for democratic outcomes in terms of basic normative criteria of equality and representativeness. Accordingly, recent research in electoral geography has more explicitly addressed the normative issues raised by practices of gerrymandering, political redistricting, and the design of representative devices.

Research on the geographies of democratization is part of a broader tradition of work examining the conditions for the formation and maintenance of liberal democratic polities. Since the late 1980s, the diffusion of democracy as a global form of governance has followed in the wake of the collapse of communism in Eastern Europe, political transitions away from authoritarianism in Latin America, Africa, and Asia, and the application of norms of democratic governance in the geopolitics of Western international financial policy, trade negotiations, and military engagements. Geographers have investigated the degree to which the adoption of democratic forms of governance can be accounted for by specifically geographical factors. This research contributes to renewed debates over whether democracy can only be established and sustained after various socioeconomic and cultural prerequisites have been met.

Geographers have also critically assessed the theoretical assumptions and the practical devices through which liberal forms of electoral democracy have been circulated as the global norm. Debates about democratization raise fundamental questions regarding the degree to which the norms of Western, liberal, representative democracy can and should be practically applied in non Western contexts and deployed as normative benchmarks of critical analysis. The geographical mobility of democratic practices suggests that the devices through which different imperatives of democratic rule are enacted can be combined, adapted, and reordered in different geographical contexts. As a consequence, the values enacted through these devices – the ways in which interests are represented, conflicts resolved, participation practised, accountability enforced – might be highly variable between different geographical contexts.

Alternative Democratic Spaces

Criticisms of theories of liberal representative democracy that assume the nation state as the natural container of democratic politics have encouraged geographers to give increasing attention to various alternative models of democratic politics. There are two aspects to this focus: a concern with scales of democratic politics above and below that of the nation state; and a concern with the potential for radicalizing democratic politics in established and new arenas.

The first aspect of the attention on alternative models of democratic politics is the focus on the spaces of democratic politics that exist above and below the level of the nation state. While electoral geography tends to focus on the translation of subnational electoral outcomes into aggregate national election results, a burgeoning literature focuses on the city or the region as a privileged scene of democratic politics in its own right. This focus is related to claims about place as the key spatial modality in which different identities and interests are experienced and negotiated. At the same time, in political science and political theory, the territorial framing of democracy is increasingly subjected to critical investigation by confronting democratic theory with the facts of contemporary transnationalism and globalization. This marks one area of overlap with research in geography. Settled assumptions about inclusion and participation in democratic politics are disturbed by transnational social relations and identifications, which point up the limitations of territorialized models of representation. Settled assumptions about the objects of democratic decision making are likewise increasingly under strain with the proliferation of non territorial concerns, such as global climate change, pandemic disease, and integrated transnational financial markets.

This shift of attention to the complex spaces and scales through which democratic politics operates is related to a second aspect of research on alternative democratic spaces in geography. It is concerned with searching out and affirming various fractures in liberal representative regimes. These explorations are committed to thinking of democracy as more than simply a procedure for legitimizing the decisions of centralized bureaucracies and holding elected representatives accountable. This reflects a broader turn to thinking of democracy as a contestatory regime. This is indicated by the prevalence of theories of democracy in geography which focus on redefining 'the political' as a realm in which new identities are formed and new agendas are generated, and through which the stabilized procedures, institutions, and identifications of official politics are contested and potentially transformed. This is indicative of an alternative sense of 'democratization', understood not so much as the geographical diffusion of established norms of democracy, but the deepening of democratic impulses and their extension of new arenas of everyday life. In contrast to the focus of electoral geography and mainstream political science on the formal democratic procedures of elections, voting, and parties, attention is paid to the diverse practices and sites where questions of accountability, citizenship, justice, and participation are contested.

These broad schools of democratic thought share an emphasis on the instrumental and intrinsic value of 'participation' as the central normative feature of democratic politics. They belong to a broader family of twentieth century egalitarian thought about the relationships between democracy, rights, and social justice, including post Rawslian political philosophy; post Habermasian theories of democracy, including their feminist variants; various postcolonial liberalisms; agonistic liberalisms and post structuralist theories of radical democracy; and the revival of republican theories of democracy, freedom, and justice. These traditions share a conviction that democratic politics amounts to more than formal procedures for the aggregation of individualized voter preferences, and taken together amount to a broad tradition of theories of 'radical democracy'. However, radical democracy in this expansive sense is not best thought of as a rejection of liberalism 'tout court'. One of the lessons drawn by diverse strands of radical political theory from the experience of twentieth century history is that struggles for social justice can create new forms of domination and inequality. This leads to a grudging appreciation of liberalism as a potential source for insight into the politics of pluralistic associational life. The problems of coordination, institutional design, and justification which liberal political theory dwells at length on are unavoidable in any normatively persuasive and empirically grounded critical theory of democracy.

In geography, 'radical democracy' has often been taken to be synonymous with a specifically post structuralist approach developed by Ernesto Laclau and Chantal Mouffe. However, the term is better used to describe a broad variety of approaches that emphasize the inclusive, participatory, and contestatory qualities of democratic politics. Perhaps the most influential of these approaches is the broad family of theories of deliberative democracy, often sharing a critical interest in the ideas of Jürgen Habermas. Approaches that develop various strategies for informed, structured participatory discussion and decision making are now central to debates over democratic theory. These generally imply a much more active role for citizens in all facets of decision making, as well as, in many cases, the extension of democratic norms to a far wider array of activities. They also involve thinking of a range of activities that are not directly political, in the sense of being tied directly to decision making processes, as nonetheless important facets of a broader democratic culture by virtue of their contribution to the quality of the public sphere. Post structuralist theories of radical democracy have also had significant influence in geography. These emphasize the idea of democracy as a diverse set of practices of challenging the boundaries of 'the political' in two senses: by establishing new issues as counting as political; and by focussing on the active formation of subjectivities and identifications in and through political practices of agonistic contestation and collective mobilization. Post structuralist theories of radical democracy tend to lay considerable emphasis on space, as a scene of multiplicity and encounter with difference, as the very scene of democratic politics. These theories are also related to a distinctive understanding of 'democracy to come', derived from the work of Jacques Derrida, which emphasizes the idea that democracy is an inherently contested, evaluative, incomplete formation which is always open to further perfection, without hope of finally culminating in a pure form.

Theories of radical democracy, whether of a Habermasian or post struturalist inflection, focus on forms of broadly noninstrumental cultural practice and performance, understood as key mediums of democratic opinion formation or democratic subject formation. In turn, both broad approaches share in the problem of whether the forms of 'influence' that can be generated in the public sphere or in cultivating an agonistic democratic ethos can or should be articulated with the institutionalized 'exercise' of power. Post Habermasian approaches finesse this issue through concepts of weak and strong public spheres. Post structuralist theorists, who address the communicative conditions of pluralistic democracy through figures of democratic ethos, are even more wary of drawing too close to sites of decision making and rule. For them, the force of democratic politics is reserved for the disruptive energies of contestatory practices.

In practical terms, theories of radical democracy, broadly defined, are often associated with calls for the 'decentralization' of decision making and political participation to subnational scales of regions and cities. This 'localizing' impulse easily falls into the trap of assuming that subnational scales of governance are somehow more democratic by virtue of being closer to people's everyday concerns. At the same time, there is increasing attention given to emergent forms of transnational democracy, focusing on the problem of whether systems of globalized economic and political governance can be subordinated to democratic oversight. Research on this dimension of contemporary democratic politics has important practical and theoretical lessons for thinking about the degree to which radical practices of deliberation, participation, and contestation might always have to contend with the problem of representation. It is in work on issues of global justice and transnational democracy that one finds sustained attention to the diverse agents of justice through which democratic justice can be pursued and secured, moving beyond an exclusive focus on states as the privileged containers of democratic politics. This emphasis on the global dimensions of democratic politics has two important implications for geographical research on democracy. First, it indicates that, rather than op posing representative to participatory forms of dem ocracy, any viable form of democratic polity is likely to combine aspects of these practices in different ways. For both practical and normative reasons, representation seems an irreducible aspect of any viable, pluralistic model of democracy. Not only do representative pro cedures enable the 'time space distanciation' of demo cratic politics, but they also embody important principles of difference and nonidentity within the 'demos'. Rep resentation is also an unavoidable mechanism for the integration of so called mute interests, such as future generations or nonhuman actors. The second reason why the globalization of democracy is significant is that it suggests a move beyond the predominant territorial framing of the spatiality of democracy. Rather than thinking simply in terms of the need to articulate sub national and national scales with global scales, dis cussions of these topics increasingly focus on the diverse spatialities of democracy, ones that articulate territorial and nonterritorial practices, and scalar and nonscalar conceptualizations of space.

Conclusion

It is significant that in those fields of social science where issues of democratic theory and practice are primarily considered, in political science and political theory, for example, contemporary debates often hinge on geo graphical themes and geographical problems. Geog raphers' contributions to further spatializing the analysis of democratic practices such as elections, and to im agining alternative democratic geographies, synthesize quite well with emerging concerns about how democracy can be understood in a world where the sites of politics are multiplying and crucial decisions are made in ways that cut across territorially organized liberal regimes. In this sense, geography is at stake in a new way across the social sciences: understanding contemporary democracy requires geographical contributions by everyone, and democracy as an object of analysis offers considerable potential for new forms of cross disciplinary engagement. However, further engagement with these debates by geographers presents a challenge for prevalent understandings of critical vocation in the discipline. The predominant understanding of critical analysis in human geography has been developed by engaging with the defects of liberal theory, primarily by reference to eco nomics. But as we have already suggested, it is difficult to appreciate the dialogic contexts of contemporary theories of radical democracy without acknowledging the degree to which these theories draw on key themes from tradi tions of political liberalism, such as pluralism, citizenship, rights, or the separation of powers. It is tempting to see in the energy of theories of radical democracy a promise that everything can and should be democratized and that all interests can and should be included in democratic practices. However, an abiding aspect of liberal theories of democratic politics is a concern with the justifiable limits of collective decision making, a concern informed by an acknowledgement of the potential harms, injustice, and violence that can be generated by the putatively democratic exercise of coercive power. Democracy, it should be remembered, is a way of doing politics, and not a substitute for politics. Tallying with the ways in which nondemocratic relationships, such as those based on faith, or those mediated through bureaucratic institutions, or those underwritten by the force of violence can and should be democratized is a pressing problem of empir ical, normative, and theoretical analysis.

See also: Citizenship; Civil Society; Electoral Geography; Governance, Good; Participation; Political Representation; Social Movements.

Further Reading

Agnew, J. (1996). Mapping politics: How context counts in electoral geography. *Political Geography* 15, 129 146.

Amin, A., Thrift, N., Baker, H. and Massey, D. (2005). Centres don't have to be points: Politics beyond state boundaries. In Latour, B. & Weibel, P. (eds.) *Making Things Public: Atmospheres of Democracy*. Cambridge, MA: MIT Press.

Anderson, J. (ed.) (2002). *Transnational Democracy: Political Spaces and Border Crossings*. London: Routledge.

Barnett, C. (2003). *Culture and Democracy: Media, Space and Representation*. Edinburgh: Edinburgh University Press.

Barnett, C. (2004). Deconstructing radical democracy: Articulation, representation and being with others. *Political Geography* 23, 503 528.

Barnett, C. (2005). Temporality and the paradoxes of democracy. *Political Geography* 24, 641 647.

Barnett, C. and Low, M. (eds.) (2004). *Spaces of Democracy: Geographical Perspectives on Citizenship, Participation and Representation*. London: Sage.

Bell, J. E. and Staeheli, L. (1998). Discourses of diffusion and democratization. *Political Geography* 20, 175 195.
Benhabib, S. (ed.) (1996). *Democracy and Difference*. Princeton, NJ: Princeton University Press.
Braun, B. and Disch, L. (2002). Radical democracy's 'modern constitution'. *Environment and Planning D: Society and Space* 20, 505 511.
Brown, M. (1997). *Replacing Citizenship: AIDS Activism and Radical Democracy*. Minneapolis, MN: University of Minnesota Press.
Burgess, J., Stirling, A., Clark, J., Davies, G., Eames, M., Staley, K. and Williamson, S. (2007). Deliberative mapping: A novel analytic deliberative methodology to support contested science policy decisions. *Public Understanding of Science* 16, 299 322.
Cohen, J. and Fung, A. (2004). The radical democratic project. *Swiss Political Science Review* 10(4), 23 34.
Connolly, W. (1993). *The Terms of Political Discourse* (3rd edn.). Oxford: Blackwell Publishers.
Connolly, W. (1995). *The Ethos of Pluralization*. Minneapolis, MN: University of Minnesota Press.
Connolly, W. (1999). *Why I am Not a Secularist*. Minneapolis, MN: University of Minnesota Press.
Dahl, R. (1970). *After the Revolution*. New Haven, CT: Yale University Press.
Dahl, R. (1989). *Democracy and Its Critics*. New Haven, CT: Yale University Press.
Dahl, R. (2000). *On Democracy*. New Haven, CT: Yale University Press.
Derrida, J. (2005). *Rogues: Two Essays on Reason*. Stanford: Stanford University Press.
Dryzek, J. S. (2002). *Deliberative Democracy and Beyond*. Oxford: Oxford University Press.
Dryzek, J. S. (2005). Deliberative democracy in divided societies: Alternatives to agonism and analgesia. *Political Theory* 33, 218 242.
Dunn, J. (2005). *Setting the People Free: The Story of Democracy*. New Delhi: Atlantic Books.
Entrikin, N. (2002). Democratic place making and multiculturalism. *Geografiska Annaler* 84B, 19 25.
Ettlinger, N. (2007). Bringing democracy home: Post Katrina New Orleans. *Antipode* 39, 8 16.
Fainstein, S. (2000). New directions in planning theory. *Urban Affairs Review* 35, 451 478.
Goodin, R. (2003). *Reflective Democracy*. Oxford: Oxford University Press.
Habermas, J. (1996). *Between Facts and Norms*. Cambridge: Polity Press.
Hannah, M. (2001). Sampling and the politics of representation in US Census 2000. *Environment and Planning D: Society and Space* 19, 515 534.
Held, D. (1995). *Democracy and the Global Order*. Cambridge: Polity Press.
Held, D. (2006). *Models of Democracy* (3rd edn.). Cambridge: Polity Press.
Hickey, S. and Mohan, G. (2005). Relocating participation within a radical politics of development. *Development and Change* 36, 237 262.
Howard, D. (2002). *The Specter of Democracy*. New York: Columbia University Press.
Johnston, R. J. (1999). Geography, fairness and liberal democracy. In Proctor, J. & Smith, D. M. (eds.) *Geography and Ethics*, pp 44 58. London: Routledge.
Johnston, R. J., Pattie, C., Dorling, D. and Rossiter, D. (2001). *From Votes to Seats: The Operation of the UK Electoral System Since 1945*. Manchester: Manchester University Press.
Katznelson, I. (1996). Social justice, liberalism and the city. In Merrifield, A. & Swyngedouw, E. (eds.) *The Urbanization of Injustice*, pp 45 64. London: Lawrence and Wishart.
Keane, J. (2004). *Violence and Democracy*. Cambridge: Cambridge University Press.
Laclau, E. and Mouffe, E. (1985). *Hegemony and Socialist Strategy*. London: Verso.
Latour, B. (2004). *The Politics of Nature: How to Bring the Sciences into Democracy*. Cambridge, MA: Harvard University Press.
Low, M. (1997). Representation unbound. In Cox, K. (ed.) *Spaces of Globalization*, pp 240 280. New York: Guilford.
Low, M. (2003). Political geography in question. *Political Geography* 22, 625 631.
Low, M. (2004). Cities as spaces of democracy. In Barnett, C. & Low, M. (eds.) *Spaces of Democracy*, pp 128 146. London: Sage.
Mann, M. (2005). *The Dark Side of Democracy: Explaining Ethnic Cleansing*. Cambridge: Cambridge University Press.
Massey, D. (1995). Thinking radical democracy spatially. *Environment and Planning D: Society and Space* 13, 283 288.
Mouffe, C. (1995). Post Marxism: Democracy and identity. *Environment and Planning D: Society and Space* 13, 259 265.
O'Laughlin, J., Ward, M. D., Lofdahl, C. L. *et al*. (1998). The diffusion of democracy, 1946 1994. *Annals of the Association of American Geographers* 88, 545 574.
O'Neill, O. (2001). Agents of justice. *Metaphilosophy* 32(1/2), 180 195.
Pattie, C. J. and Johnston, R. J. (2002). Political talk and voting: Does it matter to whom one talks? *Environment and Planning A* 34(6), 1113 1135.
Pattie, C. J. and Johnston, R. J. (2003). Local battles in a national landslide: Constituency campaigning at the 2001 British General Election. *Political Geography* 22(4), 381 414.
Pateman, C. (1970). *Participation and Democratic Theory*. Cambridge: Cambridge University Press.
Pettit, P. (1999). Republican freedom and contestatory democracy. In Shapiro, I. & Hacker Cordón, C. (eds.) *Democracy's Values*, pp 163 190. Cambridge: Cambridge University Press.
Przeworski, A. (1995). *Sustainable Democracy*. Cambridge: Cambridge University Press.
Pugh, J., Featherstone, D., Marres, N. and Thien, D. (2007). On the political. *Area* 39, 130 135.
Purcell, M. (2007). *Recapturing Democracy: Neoliberalization and the Struggle for Alternative Urban Futures*. London: Routledge.
Rydin, Y. (2003). *Conflict, Consensus and Rationality in Environmental Planning*. Oxford: Oxford University Press.
Saward, M. (2003). Enacting democracy. *Political Studies* 51, 161 179.
Sen, A. (1999). Democracy as a universal value. *Journal of Democracy* 10(3), 3 17.
Scheuerman, W. (2006). Critical theory beyond Habermas. In Dryzek, J., Honig, B. & Phillips, A. (eds.) *The Oxford Handbook of Political Theory*, pp 85 105. Oxford: Oxford University Press.
Shapiro, I. (1999). *Democratic Justice*. New Haven, CT: Yale University Press.
Shapiro, I. and Hacker Cordón, C. (eds.) (1999). *Democracys Edges*. Cambridge: Cambridge University Press.
Slater, D. (2002). Other domains of democratic theory: Space, power, and the politics of democratization. *Environment and Planning D; Society and Space* 20, 255 276.
Warf, B. (2006). Voting technologies and residual ballots in the 2000 and 2004 presidential elections. *Political Geography* 25, 530 556.
Young, I. M. (2000). *Democracy and Inclusion*. Oxford: Oxford University Press.
Young, I. M. (2001). Activist challenges to democracy. *Political Theory* 29, 670 690.

Relevant Websites

http://www.democraticaudit.com/index.php
Democracy Audit.
http://www.freedomhouse.org/
Freedom House.
http://www.opendemocracy.net/
Open Democracy.

Demography

P. Rees, University of Leeds, Leeds, UK

Glossary

Age–Time Graph The graph used to represent the age and time properties of demographic data and model variables (also known as Lexis diagram, after one of its first developers).

Cohort-Component Projection Model A model for projecting population using all the components of change for age groups in a period-cohort framework.

Components of Change The population flows that change the population: births, deaths, and migration.

First Demographic Transition The shift from a regime of high fertility and mortality rates to a regime of low fertility and mortality rates.

International Migration Migration from one country to another. Sometimes referred to as external migration.

Intranational Migration Migration from one location to another within the same country. Sometimes referred to as internal migration.

Life Expectancy The average (expected) length of life, free of the bias of the current distribution of deaths by age, which depends in part on the age distribution of the population at risk. The period life expectancy model assumes that current age-specific mortality rates will persist through the life of the current population. In the cohort life expectancy model, the age-specific rates are projected into the future and so life expectancies computed for each birth cohort.

Life Table A table of rates of mortality, survival probabilities, and life expectancies based on a stationary population model.

Long-Term Migrants Migrants who have stayed in a destination for 12 months or more (retrospective measure) or who intend to reside at a destination for 12 months or more (prospective measure).

Migrants People who have migrated from an origin to a destination and survive there. Also referred to as transitions.

Migration Distance The average of distances between origins and destinations weighted by the size of the migration flow.

Migration Expectancy The expected number of migrations that a person would experience over a lifetime, allowing for mortality.

Migrations The number of relocations of usual residence that occur in a time interval. For any migration flow, the number of migrations (events) will exceed the number of migrants (transitions/survivors). Also referred to as movements.

Multistate Model A model for computing the life table or for projecting the population in many states (e.g., regions and educational grades) which incorporates the movements or transitions between states.

Net Migration The difference between in-migration to and out-migration from a geographical unit.

Population Accounts Tables for organizing population stocks and components of change for consistency checks.

Population Ageing The process through which the population becomes older over time. The mean age of the population rises and the share of the population that is old rises. The same process occurs within the older population. The process is a consequence of the first demographic transition and becomes even more pronounced with the second demographic transition. It may be mitigated to a partial extent by the third demographic transition.

Probabilistic Projections A large set of projections of the population in which the uncertainty in future trajectories of the components of change has been handled by sampling from error distributions around those trajectories. Using the set of projections, statements can then be made about the probability that the future population will lie between an upper and a lower limit.

Proximate Determinants of Fertility A set of multiplicative factors which explain the level of fertility in a population. The factors involve marriage levels, use of contraception, use of abortion, the degree of natural infertility, and the maximum potential number of children.

Radix The assumed stationary number of births and deaths used in constructing life tables. Typically, 100 000 births and deaths per year are used.

Replacement Fertility The level of total fertility rate (TFR) needed to ensure that the population is replaced in the long run. It depends on female mortality rates in combination with age-specific fertility rates. In the lowest-mortality countries, replacement TFR has now fallen to 2.07–2.08 children per woman, close to the minimum of *c*. 2.05 when there are no female deaths before the end of female reproductive careers and when boys make up 51% of births.

Scenario Projection A population projection which uses conditional assumptions (e.g., what if fertility rates were at replacement level) to explore special futures for the population.

Second Demographic Transition The shift to a demographic regime of below-replacement fertility rates.

Short-Term Migrants Migrants who have stayed in a destination between 3 and up to 12 months (retrospective measure) or who intend to reside at a destination for 3–12 months (prospective measure).
Standardized Mortality Ratio (SMR) The ratio of observed mortality in a population to the mortality that would have occurred if it had the same age-sex distribution as a standard population (e.g., European standard population).
Third Demographic Transition The shift to a regime of low fertility combined with positive net migration from other countries.
Total Fertility Rate (TFR) The total number of children a woman would have if she experienced the age-specific fertility rates of either a current period (period TFR) or of an estimation/projection of age-specific fertility rates for a birth cohort (cohort TFR).
Variant Projection A population projection which uses assumptions about one or more of the components of change that is different from the assumptions used in a central or principal projection.
Visitors Persons who have stayed in a destination for less than 3 months (retrospective measure) or who intend to reside at a destination for up to 3 months (prospective measure).

Definitions

People are normally counted at their usual place of residence. By usual residence is meant the place where they spend most of their nonwork time, and from which they make journeys to work, to shop, to school, to leisure, and to holiday locations. The concept of usual residence is unproblematic for most of the population, but a significant minority have several places of residence, some temporary, some permanent. In many censuses, a separate record is made of people who are temporarily resident on census night, either because they are visiting or because they are working away from home.

In societies where permission was needed to take up permanent residence in new places (under the Hukou registration system in China, for example), counts of the population of cities depended on whether people with temporary residence were included as well as those with permanent registrations. Many Western societies attract short term labor migrants (e.g., United Kingdom with its large intake of labor migrants from the new European Union member states in the period 2004–07), but official population estimates only include long term international migrants and not those who are short term. Temporary migration results in the population present fluctuating during the year, and so resident population estimates are used. As population counts and estimates play a crucial role in the computation of key demographic indicators, the user of those indicators needs to bear in mind that they are subject to a margin of error based on the uncertainty of the population stock estimate.

Population Characteristics

Demography studies the groups within the population that differ in their attributes and behavior. We consider the following dimensions that differentiate the population: location, sex, age, time; marital status, living arrangements, and household membership; social and economic statuses, including educational attainment, occupation, and income class; and race and ethnicity.

Location, Sex, Age, and Time

All populations are distinguished by location in geographical space. They are usually studied at country level where national laws and boundary controls make people's locations reasonably unambiguous. Within countries, demographic analyses are carried out for regions, local districts, and small areas. There are difficulties in comparing territorial units across countries (particularly when migration is considered) and across time because changes to administrative units are made. The European Union has made efforts, for example, to standardize territorial units across Europe in the Nomenclature of Units for Territorial Statistics (NUTS).

Virtually, all demographic analysis distinguishes between the sexes both because of differences in biology and because of differences in behavior and welfare. For example, women live longer than men in virtually all countries of the world, but these differences are narrowing in the most developed nations.

A crucial variable in demographic analysis is age, which is measured as the integer number of years elapsed since birth. Mortality, fertility, and migration processes vary profoundly and in different ways by age. Age interacts with the passage of time to give rise to a number of key concepts which are displayed in **Figure 1**, which consists of a set of age–time graphs. The top row of graphs illustrates the concepts of age group (persons with the same ages), period (an interval of time), and cohort (persons with the same birth year who proceed together through age–time space). We can place within the age–time graph both demographic events (births, deaths, and migrations), which occur at a point in time to persons of an exact age, and demographic transitions (changes of status measured at two time or two age points, such as survival or being a migrant). There are four schemata for capturing demographic events/transitions (shown in the bottom block of four graphs): period–age (e.g., deaths in a calendar year to persons aged 80 at death), age–cohort (e.g., survivors from one birthday in 1 year to the next

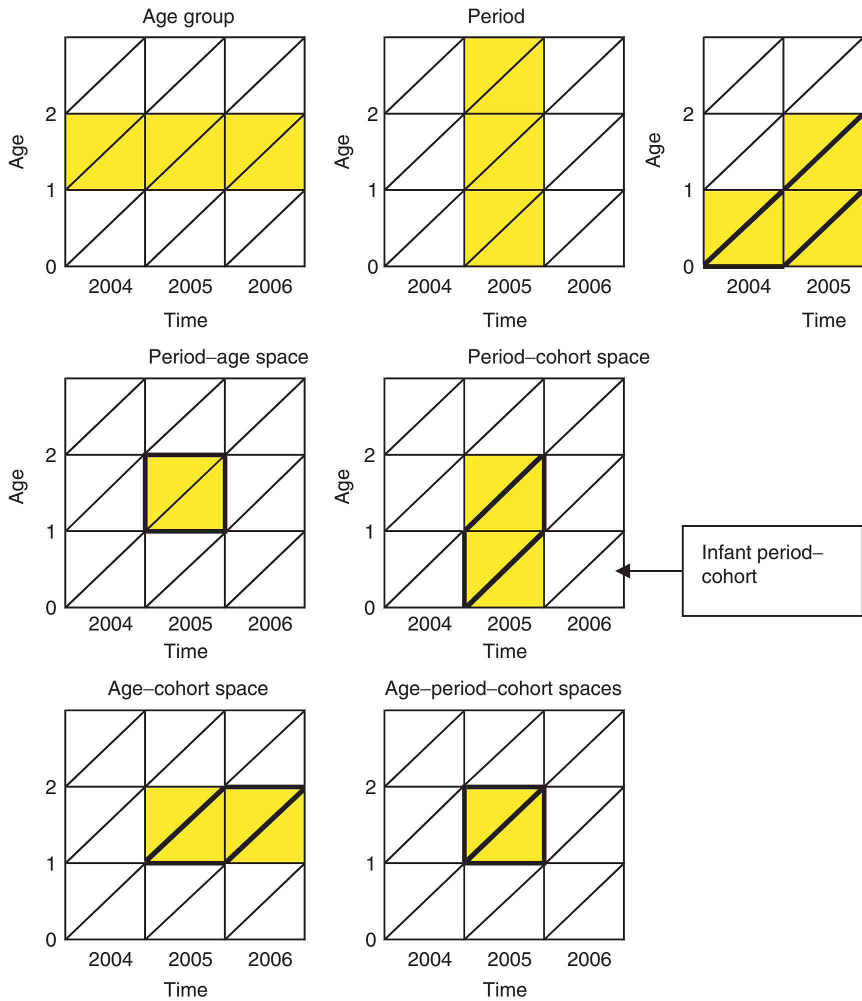

Figure 1 The age time framework at the heart of demographic analysis.

birthday in the following year), period–cohort (survivors of one age group at the start of the year to the next age group at the end of the year), and age–period–cohort (e.g., births to mothers born in year by age at maternity and year of birth). The age–cohort schema is used in the life table model. The period–cohort schema is used in the cohort component population projection model. The age–period–cohort schema is used to record demo graphic events where year of birth and age of event are both recorded. Data from this schema can be reassembled for use in either life tables or projection models. The period–age schema is used when year of birth has not been recorded for the demographic event and some as sumptions are needed in order to use these data in demographic models.

Marital Status, Living Arrangements, and Household Structure

A second set of population attributes which affect demographic processes concern small groups: couples, families (living together), and households (person sharing the same living accommodation). These units take eco nomic (e.g., consumption) and social (e.g., childrearing) decisions together (at least in part). Marital status in cludes the states of being single, married (first married, remarried), separated (married but living apart), and di vorced. In many societies, cohabitation by couples out side of marriage has become important, though the living arrangement may lead to marriage. In the future, social and demographic surveys may need also to count same sex unions where these are recognized.

Social and Economic Status: Education, Occupation, and Income Class

Although social and economic attributes are the focus of social geography rather than demography, demographic analysis has been usefully applied to the growth of populations by educational grade. Projections of the world's population by educational attainment under varying scenarios demonstrate the enormous future importance of the educated labor force of China and India, for example.

Race and Ethnicity

Through past and present international migration, national populations, particularly in the developed world, are becoming more diverse in their racial and ethnic composition. The main focus of social research has been on integration/segregation of new groups within national societies, but demographers are now tracking the changing composition of national populations using a variety of projection techniques. Europe's population will become less 'European' and more 'foreign' and mixed in ethnicity over the next 50 years.

Demographic Models

To understand how and why population sizes, compositions, and distributions change, we use models of the processes of population change. Although simplified from complex reality, demographic models are powerful tools for understanding population change.

Models of the Components of Population Change

Populations of the same territory at two points in time are connected by the following 'components of change' relationships:

$$P^{t+n} = P^t + N^t + M^t \qquad [1]$$

where P^{t+n} is the population at time $t+n$ years, P^t is the population at time t, N^t is the natural increase of the population in the time interval t, $t+n$, and M^t is the net balance of in and out migration. Natural increase is the surplus of births over deaths in a population, and the net migration balance is the difference between in and out migration streams. So it is useful to break down the change components further into:

$$P^{t+n} = P^t + B^t \quad D^t + M^t_{\mathrm{in}} \quad M^t_{\mathrm{out}} \qquad [2]$$

where B^t, D^t, M^t_{in}, and M^t_{out} are, respectively, births, deaths, in migrations, and out migrations in the time interval starting at time t. However, it is normally important to distinguish the migration streams into those internal to a country from those crossing international boundaries, because the data sources may be different and because international migration is influenced by explicit policy measures, whereas internal migration is only indirectly and weakly affected. So the third version of the 'components of change' equation is

$$P^{t+n} = P^t + B^t \quad D^t + M^t_{\mathrm{in}} \quad M^t_{\mathrm{out}} + I^t \quad E^t \qquad [3]$$

where M_{in} and M_{out} refer only to internal migration and I to immigration and E to emigration. **Table 1** shows a components table for one ethnic population, White males, in the UK region of Yorkshire and the Humber in the 2001–02 (mid year to mid year interval), decomposed by a single year of age. Each row of the table records what happens in a period–cohort, labeled by the start and finish ages in the table. The first row refers to the infant period–cohort, with a start stock of births and a finish stock of population aged 0. For a single population, the components of change account completely for population flows into and out of the population.

Population Accounting Models

When many populations that interact (e.g., a set of regions within a country) are considered together we can expand the components into a population accounts table. **Table 2** shows how this is done for the components

Table 1 An example of the components of population change for period cohorts, Yorkshire and the Humber region, UK, 2001 02, White males[a]

Start age	*Population 2001*	*Deaths*	*Out-migration*	*Emigration*	*In-migration*	*Immigration*	*Population 2002*	*Finish age*
Births	24 814	147	805	106	864	105	24 724	0
0	24 338	23	691	92	742	91	24 365	1
1	25 175	11	609	94	653	93	25 208	2
2	26 358	8	559	86	600	85	26 391	3
3	26 829	6	518	85	556	84	26 860	4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
98	123	42	2	0	2	0	81	99
99	107	41	1	0	3	0	68	100
100 +	142	142	0	0	0	0	0	101 +
Sum	2 274 520	27 056	38 451	7786	41 256	7707	2 250 190	
Population	2 249 706							

[a] Each row contains frin row contains the stocks and flows for one period cohort. The sum of the Population 2001 column is the combination of the 2001 population and births in 2001 02.

Source: Adapted from WHO (2007).

defined as demographic events by explicating repre senting the rest of a country and the rest of the world as system states. Very similar accounts can be defined using transition information, which is convenient when mi gration data derive from retrospective questions in a census or national survey. Such population accounts implicitly underpin multistate population models, widely used in population projection. The multistate framework has the key advantage of linking the population changes in one region or country to those in other regions or countries from which it receives its migrants. The ac counting framework has the key advantage of forcing the user to achieve consistency between start and finish population stocks and component flows.

Life Table Models/Stationary Population Models

One of the most widely used demographic models is that of the life table used not only by national and inter national statistical agencies to monitor the health of populations but also by the actuarial profession and in surance industry to compute the funds needed to support population groups (firm employees and members of life assurance funds). Life table models are designed to re move from mortality analysis the effects of varying births to a population and the effect of the current age structure on cruder measures. So the life table model adopts a hypothetical population in which deaths and births are equal. **Table 3** presents an abridged life table for Aus tralia using 2005 mortality data for persons. To the hypothetical stationary population are added 100 000 babies each year (row 1, column 4) and the total popu lation is just over 8.1 millions (row 1, column 7). Most demographic texts give accounts of how the variables set out in **Table 3** columns are computed. Mortality rates, ${}_nM_x$, are computed by dividing observed death counts by period–age by a mid interval or average population in the age group. Mortality probabilities, ${}_nq_x$, are computed from mortality rates using simple assumptions to convert to age cohorts. The complement of these probabilities, the survival probabilities, are applied to the life table radix, 100 000, to compute survivors, ${}_nl_x$, and to compute nonsurvivors, ${}_nd_x$. Life years lived in an interval, ${}_nL_x$, are a sum of the survivors at the end of the interval, l_{x+n}, plus the nonsurvivors, ${}_nd_x$, weighted by the length of the age interval, n multiplied by the fraction spent alive, ${}_na_x$. The total life years beyond age x, T_x, are a sum of the life years lived from the oldest to age x. Life expectancy, e_x, is the ratio of T_x to the survivors to age x, l_x.

Full accounts of the life table model equations are provided in the main demographic texts. Here we com ment on features of the life table which are neglected in conventional descriptions.

Although most life table models list six equations connecting the life table variables as laid out in **Table 3**, life expectancies beyond age x can be expressed in one sum of products equation as a function of the age intervals, n, life fractions in age intervals, ${}_na_x$, and mor tality rates, ${}_nM_x$:

$$e_x = \sum_{k\ x}^{k\ z}\left\{ n_n a_x \prod_{y\ 0}^{y\ k}[1 \quad ({}_nM_x/(1 \quad {}_n a_{xn}M_x))]\right\} \qquad [4]$$

where z is the last age beyond which all survivors are assumed to die; k, x, and y are age indices. It is thus possible to compute life expectancies or any other life

Table 2 Population accounts for a two-region system, White males, ages 1 2

		End state in interval t, t+n				
		Region (1)	*Rest of the country (2)*	*Rest of the world*	*Deaths*	*Totals*
Start state in interval *t*, *t+n*	*Region (1)*	R^1	M^{12}	E^1	D^1	P^{1t}
	Rest of country (2)	M^{21}	R^2	E^2	D^2	P^{2t}
	Rest of world	I^1	I^2	0	0	I^*
	Totals	P^{1t+n}	P^{2t+n}	E^*	D^*	T
	Period cohort: age 1 to age 2	*End state in 2000 01*				
		Yorkshire and the Humber	*Rest of UK*	*Rest of world*	*Deaths*	*Totals*
Start state in 2000 01	Yorkshire and the Humber	24 461	609	94	11	25 175
	Rest of UK	653				
	Rest of world	93		0	0	
	Totals	25 208				

Notes: R^1 = accounting residual for region 1 = P^{1t} M^{12} E^1 D^1 etc. The asterisk (in D*, I*, E*) indicates summation over the superscript replaced and thus represents totals for the variables.

Table 3 Abr dged fe tab e for Austra a, 2005, persons

Ages (x)	*Age interval (n)*	*Years alive ($_na_x$)*	*Mortality rate*	*Mortality probability*	*Cohort numbers surviving*	*Cohort numbers dying*	*Life years*	*Total life years*	*Life expectancy*	*Survivorship probability*	*Ages at start*	*Ages at end*
			Period–age ($_nM_x$)	*Age–cohort ($_nq_x$)*	*Exact age (l_x)*	*Age–cohort ($_nd_x$)*	*Age–cohort ($_nL_x$)*	*Sum of age cohorts (T_x)*	*Exact age (e_x)*	*Period–cohort ($_nS_x$)*		
0	1	0.1	0.00 511	0.00 509	100 000	509	99 542	8 141 182	81.4	0.99 455	B rth	0–4
1	4	0.4	0.00 024	0.00 096	99 491	96	397 735	8 041 640	80.8	0.99 912	0–4	5–9
5	5	0.5	0.00 011	0.00 055	99 395	54	496 841	7 643 905	76.9	0.99 947	5–9	10–14
10	5	0.5	0.00 010	0.00 051	99 341	50	496 579	7 147 064	71.9	0.99 885	10–14	15–19
15	5	0.5	0.00 036	0.00 180	99 291	178	496 008	6 650 485	67.0	0.99 768	15–19	20–24
20	5	0.5	0.00 057	0.00 285	99 112	282	494 856	6 154 477	62.1	0.99 700	20–24	25–29
25	5	0.5	0.00 063	0.00 315	98 830	311	493 371	5 659 621	57.3	0.99 655	25–29	30–34
30	5	0.5	0.00 075	0.00 375	98 519	369	491 670	5 166 249	52.4	0.99 583	30–34	35–39
35	5	0.5	0.00 092	0.00 459	98 149	450	489 621	4 674 580	47.6	0.99 439	35–39	40–44
40	5	0.5	0.00 133	0.00 663	97 699	648	486 876	4 184 958	42.8	0.99 192	40–44	45–49
45	5	0.5	0.00 192	0.00 954	97 051	926	482 941	3 698 082	38.1	0.98 821	45–49	50–54
50	5	0.5	0.00 283	0.01 407	96 125	1352	477 245	3 215 141	33.4	0.98 238	50–54	55–59
55	5	0.5	0.00 429	0.02 121	94 773	2011	468 838	2 737 896	28.9	0.97 217	55–59	60–64
60	5	0.5	0.00 704	0.03 458	92 762	3208	455 791	2 269 058	24.5	0.95 512	60–64	65–69
65	5	0.5	0.01 143	0.05 555	89 554	4975	435 335	1 813 267	20.2	0.92 775	65–69	70–74
70	5	0.5	0.01 883	0.08 993	84 580	7607	403 882	1 377 932	16.3	0.88 094	70–74	75–79
75	5	0.5	0.03 268	0.15 107	76 973	11 628	355 795	974 050	12.7	0.80 189	75–79	80–84
80	5	0.5	0.05 806	0.25 352	65 345	16 566	285 308	618 255	9.5	0.68 341	80–84	85–89
85	5	0.5	0.10 034	0.40 107	48 779	19 564	194 983	332 947	6.8	0.49 715	85–89	90–94
90	5	0.4	0.16 897	0.56 065	29 215	16 379	96 936	137 964	4.7	0.33 547	90–94	95–99
95	5	0.3	0.27 815	0.70 471	12 835	9045	32 519	41 028	3.2	0.20 739	95+	100+
100	2.25	0.45	0.44 543	1.00 000	3790	3790	8509	8509	2.2			

Source Adapted from WHO (2007)

table variable without necessarily constructing the full life table, though presenting the intermediate compu tations in life table form is advisable.

Life tables involve explicit assumptions about the fraction of an age interval spent alive by persons who die. This variable is labeled ${}_na_x$ and is reported in the third column of the table. For most age intervals, an assump tion of 0.5 is fine, but for the first and second intervals it is better to substitute fractions based on a study of the detailed distribution of deaths by smaller age intervals. This leads to an adoption of 0.1 for the interval birth to age 0, 0.4 for the interval age 1 to age 5, 0.4 for the interval age 90–95, and 0.3 for the interval 95–100. The life spent beyond age 100 is computed as the reciprocal of the mortality rate and the fraction is computed for a 5 year interval. More precise estimates based on empirical information about the distributions of deaths within age intervals can be made, but do not make much difference to the computed expectations of life.

It is important when computing life tables for most countries to adopt at least 100 as the highest age, as more and more people survive to that advanced age. In many countries, national statistics are already prepared for much higher ages (e.g., to 120). Under 2005 mortality conditions shown in **Table 3**, 50% of Australians will survive to 84.6 years. On only mildly optimistic as sumptions about continuing declines in mortality rates (2% per year), it is likely that more than half the 2006 birth cohort in the UK will still be alive at age 100. The implications of this trend for society will be profound.

The column titled 'survivorship probability' is not usually included in published life tables. Period–cohort survival probabilities are computed as L_{x+n} divided by L_x, where L is interpreted as the number in the stationary population of the life table. These probabilities can be used directly in cohort survival calculations in population projection models.

Australia's life table represents one of the most ad vanced in the world: better life expectancies are found only in Japan and Iceland. It should be stressed that these life expectancies are period specific. Cohort expectancies are likely to be considerably higher, although they will depend on forecasts of mortality into the future.

The Cohort-Component Model for Projecting the Population

One of demography's main contributions to societal planning is to provide projections of the future popu lation. A population projection is a future trajectory of the population and its constituent groups based on par ticular assumptions about the drivers of change, mortal ity, fertility, and migration. There are thus as many projections as there are sets of plausible assumptions.

The main device used for population projection is the cohort component model, an expansion of the com ponent model, adding age in period–cohort form. This is conveniently organized in the following sequence of equations in single region form.

For all period–cohorts from $x=0$ to $x=z$ (the last) and recognizing populations by gender, g, the com ponents of change equation is

$$P_{xg}^{t+n} = P_{xg}^{t} \quad D_{xg}^{t} + M_{\text{in},xg}^{t} \quad M_{\text{out},xg}^{t} + I_{xg}^{t} \quad E_{xg}^{t} \qquad [5]$$

with a parallel equation for the infant period–cohort (from birth to age 0)

$$P_{-1g}^{t+n} = B_{-1g}^{t} \quad D_{-1g}^{t} + M_{\text{in},-1g}^{t} \quad M_{\text{out},-1g}^{t} + I_{-1g}^{t} \quad E_{-1g}^{t} \qquad [6]$$

This formulation assumes that people remain in the same period–cohort in a time interval, so we need to add an equation to transfer them to the next period–cohort at the start of the next time interval:

$$P_{x+ng}^{t+n}(\text{next}) = P_{xg}^{t+n}(\text{current}) \qquad [7]$$

with the necessary addition of the end populations for the last, open ended age group

$$P_{xg}^{t+n}(\text{next}) = P_{x-ng}^{t+n}(\text{current}) + P_{xg}^{t+n}(\text{current}) \qquad [8]$$

The cohort component model also requires that the age and time intervals be equal in length (and preferably single years of age and annual time intervals).

The equations above are simply accounting identities, which must be turned into model equations for use in projection. Occurrence–exposure rates are computed as the ratio of demographic events to the population at risk for a unit time interval. A simple assumption, used fre quently, for the population at risk is that it is a linear average of the start and end populations in an interval. Using d to represent mortality rates, e emigration rates, o out migrations rates (to the rest of a country), M_{in} to represent in migration flows (from the rest of a country), and n the age/time interval, the projection equation for the end of interval population of gender g in the period–cohort that starts in age x is projected as

$$\begin{aligned} P_{xg}^{t+n} = {} & P_{xg}^{t}\left[1 \quad n\tfrac{1}{2}\left(d_{xg}^{t} + e_{xg}^{t} + o_{xg}^{t}\right)\right] / \left[1 + n\tfrac{1}{2}\left(d_{xg}^{t} + e_{xg}^{t} + o_{xg}^{t}\right)\right] \\ & + M_{\text{in},xg}^{t} / \left[1 + n\tfrac{1}{2}\left(d_{xg}^{t} + e_{xg}^{t} + o_{xg}^{t}\right)\right] \\ & + I_{\text{in},xg}^{t} / \left[1 + n\tfrac{1}{2}\left(d_{xg}^{t} + e_{xg}^{t} + o_{xg}^{t}\right)\right] \end{aligned} \qquad [9]$$

Once the population that exists at the start of the interval has been survived forward to end of the interval, new borns can be projected and assigned a gender:

$$B_{x}^{t} = u_g \times n \times f_{x}^{t} \times \tfrac{1}{2}\left(P_{xf}^{t} + P_{xf}^{t+n}\right) \qquad [10]$$

where u_g is the proportion of births that are of gender g (male, female) and f_x are the period–age fertility rates. Equation [9] is then used for the infant period–cohort with births rather than start population to survive and migrate newborns in the interval.

This schema for computing the cohort component model avoids the need for iteration and can be easily generalized from a single region to many regions, con verting the scalar equations into matrix equations.

The Multistate Demographic Model

American demographer Andrei Rogers and colleagues developed the life table and cohort component models to handle many populations simultaneously. Populations of particular 'states' (e.g., regions) are the result of the flows from and to all the other states in the system being studied. The populations of states are therefore dependent on the populations of all other states: out migrations from a region are in migrations to other re gions. It took some time for full agreement to be reached on the exact specification of the multiregional demo graphic model (in its life table or projection forms), be cause different operational realizations were needed, depending on the nature of the migration flow data available in any applied context. If migration data derive from retrospective questions in censuses, surveys, or linked registers, then a transition form of the multistate model is appropriate; if migration data derive from population or administrative registers which record flows as events, then a movement form of the model is needed. But the only real difference between these approaches is the way the input multistate transition probabilities are computed. Once computed, the models are identical.

The single region cohort component model has been generalized thus:

$$p_{xg}^{t+n} = \left[1 \quad n\tfrac{1}{2}\left(M_{xg}^{t}\right)\right]\left[1 + n\tfrac{1}{2}\left(M_{xg}^{t}\right)\right]^{-1} p_{xg}^{t} + \left[1 + n\tfrac{1}{2}\left(M_{xg}^{t}\right)\right]^{-1} i_{xg}^{t} \qquad [11]$$

where p_{xg}^{t+n} is the column vector of regional populations for gender g and period–cohort x at the end of the time interval, p_{xg}^{t} is the equivalent for start of time interval populations, i_{xg}^{t} is the column vector of immigration to regions in the time interval and M_{xg}^{t} is a matrix con taining the negative of mortality and emigration rates in the principal diagonal and region to region migration rates in the off diagonal cells. The other parts of the single region model can be similarly generalized to many regions in matrix form. Of course, this brief account of the multistate demographic model cannot provide read ers with more than the basic idea, and they will need to pursue a sustained reading course to build a full under standing. Multistate methods are widely employed by national statistical offices in their regional and local populations. They are also used by regional authorities to project their local populations and by Lutz and col leagues to project sets of world regions and countries. Ways of reducing the number of variables in the multi state model have been explored through substituting functions for the component rate age schedules or through representing the full origin–destination–age–sex array of migration rates with simpler models. Software (LIPRO) for implementing the multistate projection model is available from the Netherlands Interdisciplinary Demographic Institute (NIDI).

Other Models for Projecting the Population

The models discussed so far are based on the average behavior or experiences of population groups and the number of states through which the population flows is limited to the main demographic variables, location, sex, and age in combination with marital status, ethnic status, labor force status, or education status. We may wish to examine the full interplay of these variables on popu lation dynamics, but face the problem of rapidly ex panding numbers of variables and the difficulties of estimating transitions between states.

So researchers have adopted a different model structure in which the lives of individuals are represented and then assigned events or transitions by sampling from a distribution of event rates or transition probabilities. Multistate and micro simulation projection models are closely related, and have advantages and disadvantages which suit each to particular applications. Many im portant cross sectional micro simulation models have been constructed in order to study the impact of eco nomic policy changes on households, but relatively fewer dynamic models have been built, because of paucity of good transition data.

We now briefly discuss the way projection models have been used, examining, first, variant projections and, second, probabilistic projections, both attempts to cap ture and report the uncertainty inherent in projections.

Variant Projections

The projection considered most likely is termed the 'principal' and alternatives are labeled 'variants'. A common scheme for preparing variant projections in volves deciding on high, medium, and low assumptions for each of the three components (or four if migration is split into its internal and international streams). This then generates a schema of some 27 variants (three as sumptions for each of three components). **Figure 2** dis plays the principal 2004 based projection for the United Kingdom plus eight variant projections. It is also common to carry out particular 'what if' or scenario projections to achieve additional understanding of the future

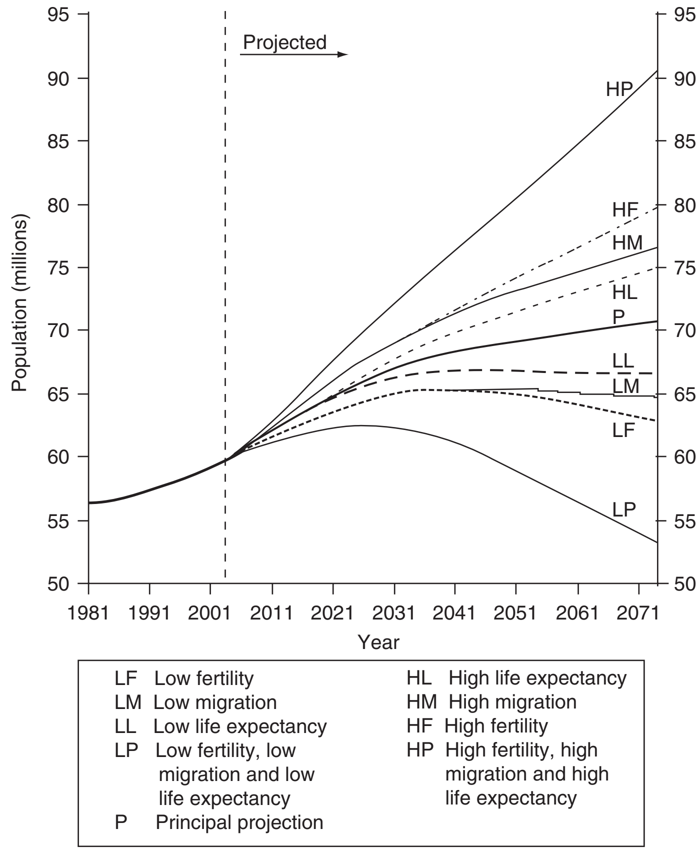

Figure 2 Population of the United Kingdom according to principal and variant 2004-based projections, 1981 2074. Source: Crown Copyright, ONS.

population. For example, a projection that assumed no international migration would be compared with the principal projection to assess the impact of the inter national migration assumptions on the future population.

Probabilistic Projections

In the past 15 years several demographers have argued that users want to know what probability can be placed on a particular range of future populations or what the probability that a population will reach a particular level. A group led by Lutz at the International Institute of Applied Systems Analysis has pioneered techniques for producing probabilistic projections. The essence of the method is to estimate the error distribution in key component parameters (e.g., total fertility rate) and to sample from these distributions year by year in the fu ture. Each set of parameters is used to project the population into the future, and from the full set of hundreds of projections run a probability distribution of outcomes can be constructed. **Figure 3** shows such an outcome for Australia. The median projection for Aus tralia sees the population growing from 20.0 millions in 2001 to 28.0 millions in 2051. Around this median are the uncertainty distributions of population: around two thirds of outcomes will lie between 24.4 and 31.8 millions, while there is a 95% probability that the 2051 Australian population will lie between 21.1 and 36.0 millions.

Main Components of Change and Associated Indicators

Much effort has been expended by demographers to describe, explain, and forecast the main drivers of population change. We consider some features of this effort, with some discussion of sources of data and of the important relationship of each component with age.

Mortality Measures and Trends

To measure the mortality level experienced by a popu lation, good statistics on deaths and population are nee ded. In most developed countries, deaths must be legally registered with cause(s) of death recorded by a physician. In countries with less developed vital registers, survey methods are used that ask households about deaths of members. This was the method used to estimate the very high death rate that occurred in Iraq in the 2003–06 period after the invasion of the country by American and coalition forces, though the method of sampling house holds has been challenged.

Register or survey deaths combined with population estimates are used to compute age specific mortality rates. **Table 3** displays such a set of rates for Australia in 2005. The relationship between mortality and age has the general form shown in **Figure 4** (for the United Kingdom in 1998). Rates for the very young are high (because of infectious disease and malnutrition in poor countries, and because of malformations at birth in richer countries) and then decline quickly to a base level between ages 5 and 10. From age 10, mortality rates rise steadily in a straight line on the semilog graph, indicating exponential increase with age. The smoothness of the rise is inter rupted between ages 15 and 25 by above trend mortality rates, mainly associated with accidents and violence. Populations differ in the level of this rate–age curve. In **Figure 4** we can see that female mortality rates are lower than male, except before age 5 and after age 95 when they are nearly the same.

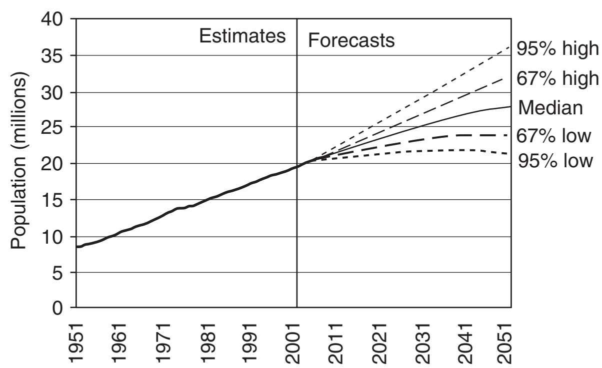

Figure 3 The probability distribution of Australia's future population. Used with kind permission of Wilson, T. and Bell, M. (2004). Australia's uncertain demographic future. *Demographic Research* 11(8), 195 234. Online at: http://www.demographic-research.org/Volumes/Vol11/8/, Figure 12.

Using age specific mortality rates, it is usual to com pute standardized mortality rates which filter out the effect of the age structure of a population on the level of mortality. For comparison of mortality rates across sub national areas within a country, standardized mortality ratios (SMRs) are used in which the mortality experience of a local population is compared with that of the country which are given the value of 100. Areas with worse than the national average mortality SMRs are above 100; areas with better than average mortality experience have SMRs below 100.

Life tables, discussed earlier in this article, provide the best means of comparing the mortality of different populations. Where it is difficult to make good estimates of the age specific mortality rates that are the key input to the life table model, then recourse can be had to collections of model life tables. These use common pat terns of the mortality–age relationship and generate life tables for a set of given life expectancies (20, 25, …, 80,

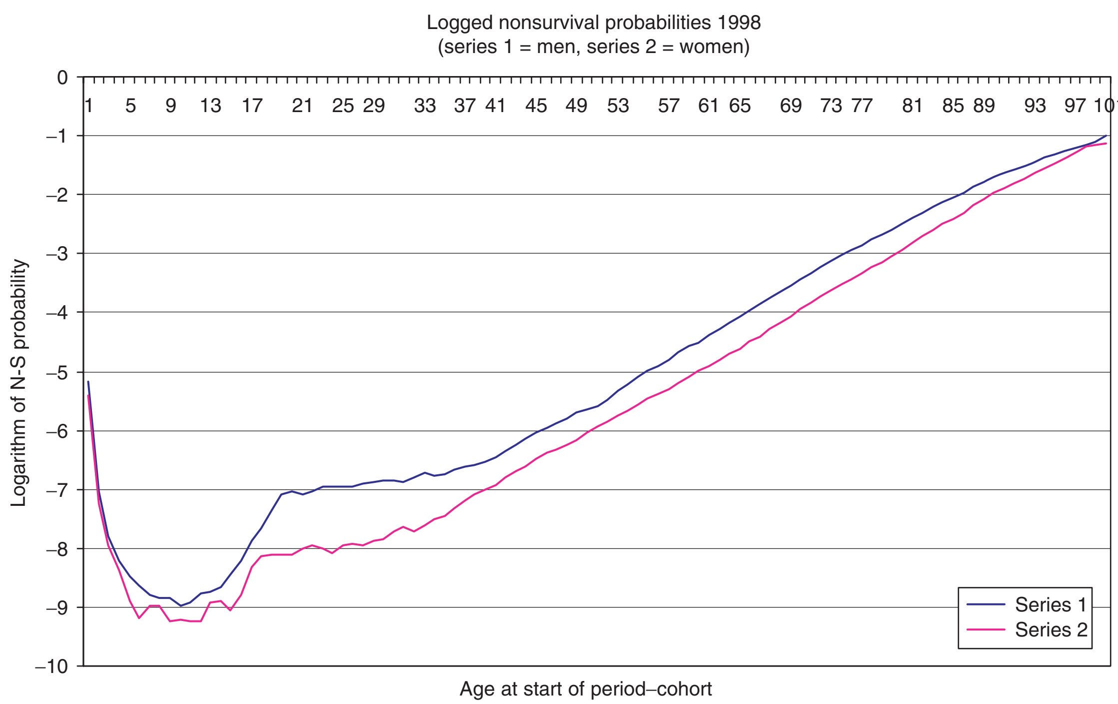

Figure 4 Mortality probabilities by age, United Kingdom, 1998. Computed by the author from data on deaths and population published by the Office for National Statistics and other UK national statistical offices (Crown Copyright).

85). However, the World Health Organization has de veloped an extensive library of country life tables, and there is now less need to use model life tables.

Mortality trends have in the past six decades, in general, moved steadily downward, resulting in steady and sustained rises in life expectancies in most countries. For developed countries, the rate of improvement in life expectancy has been about 2 years per decade with higher rates of change in fast developing countries in Asia. However, there are notable exceptions such as the post Soviet transition countries which experienced a decline in life expectancy as a result of economic hard ship. Most Central and East European countries are now experiencing improving life chances. The other countries which have experienced declines in the last two decades have been sub Saharan countries, such as Botswana or Zimbabwe, which have been affected by the HIV/AIDS epidemic or severe economic crises.

There is a vigorous debate about the future of life expectancy, which most international and national pro jections assume will continue to improve, at least in the short run. On the one hand, 160 years of historical re cords analyzed by demographers Vaupel and Oeppen tell us that the best national life expectancies have improved at 0.20–0.25 of a year per year, with the implication that projection assumptions should, in any one country, adopt a similar linked trajectory. On the other hand, Olshansky and colleagues emphasize the threat of increasing obesity to this record of continuous improvement. This debate about longevity has considerable relevance for economic and social policies, particularly for the provision of pensions, health, and social care for the elderly.

Fertility Measures and Trends

To measure the fertility level experienced by a population, good statistics on births and population are needed. In most developed countries, births must be legally registered by the parents. In countries with less developed vital registers, survey methods are used that ask households about past births of mothers. From these data are com puted crude birth rates (births/population), general fer tility rates (births/females age 15–44), and age specific fertility rates or ASFRs (births to mothers by age/females by age). ASFRs are multiplied by the relevant age interval (usually single years) and summed to yield the total fer tility rate (TFR) for a population. The TFR is a synthetic indicator of women's fertility: it is a prediction of the number of children that a woman would have if she ex perienced the current period ASFRs. Cohort TFRs sub stitute actual (past) or predicted (future) fertility rates to arrive at a better indicator of what might happen but re quire a long time series of ASFRs or a projection of them.

A version of the TFR which allows for the effect of female mortality between age 0 and 45 is the net reproduction rate (NRR), which is the number of daughters a woman would have by the end of her re productive life. When NRR is 1, then the population replaces itself, generation by generation. Because com puting the NRR means that the female life table must be computed first, researchers interested in the level of fertility in relation to replacement use the concept of the replacement TFR, commonly specified as 2.1 children. The replacement TFR varies between world regions and countries because female mortality varies. In the least developed regions in 1995–2000, 2.75 children were needed to guarantee that one daughter is alive at the end of the reproductive lifetime. The replacement TFR for England and Wales has been recently computed to be 2.07 in 2001, close to the minimum of about 2.05.

Trends in fertility in recent decades have been pre dominantly downward or, if low levels have already been achieved, fluctuating around those low levels. In de veloping countries, the fertility transition is part of the first demographic transition to smaller families con sequent on industrialization, urbanization, the intro duction of universal education, and a shift in the balance of childrearing costs/benefits. The downward shift has been promoted by family planning programs and the development of new effective methods of contraception (the contraceptive pill, the intrauterine device, etc.). An influential framework called proximate determinants of fertility has been developed by Bongaarts and colleagues for understanding the level of fertility in countries in transition.

Developed countries saw low fertility in the Great Depression years of the 1930s and the war years of the early 1940s, followed by fertility booms in the quarter century after 1945. Since 1970, most European, North American, and East Asian populations have seen fertility decline to below replacement, as women's independence has grown through labor market entry, educational ad vancement, and female control over conception, com bined with the demise of the traditional child centered family and the rise of individualism. This fertility shift has been termed the second demographic transition. One of the factors underpinning continuing below replace ment fertility in developed countries has been the post ponement of births to older ages: mean age at first birth is now in the 25–29 age group, and in many countries the peak fertility rate has shifted into the 30–34 age group. But if postponement finishes, then fertility decline stops. In 2000–06, half of European countries experienced small increases in TFR, while half experienced small declines. These data have been interpreted as showing two fertility regimes in Europe: northern countries from Finland to France where TFRs are around 1.8 and have risen a little since 2000, and Southern, Central, and Eastern Europe where TFRs are below 1.5 where there are both small increases and decreases. A suggested determinant of this pattern is the degree of support for early childcare and reentry into the labor market after childbirth provided by the state, institutions, and the market to mothers.

International Migration Measures and Trends

International migration is the movement of people across international borders for the purpose of settlement. International migrants change their usual place of residence from one country to another. The United Nations suggests that the degree of permanence of the migration should be measured over a 12 month period, so that shorter stays in another country are not classified as permanent international migration. When passengers arrive in a country, they are asked whether they intend to stay for less than 3 months, classifying them as visitors; for between 3 months and 12 months, classifying them as short term migrants; or for 12 months or more, classifying them as long term migrants. This is a prospective measurement of migration. Alternatively, people can be surveyed at their current place of residence and asked where they were living 12 months ago. If the answer is another country, then they are classified as international migrants. This is a retrospective measurement of migration. It is also possible to measure international migration by asking migrants leaving a country to register their departure and to ask a question about migration of those new registering in countries which maintain a comprehensive population register.

All measurements of international migration are fraught with error and disagreement. Those expressing an intention to stay for more than 12 months may return home before the year has ended, while visitors may decide to apply for permanent residence after arrival. Short term migrants may become long term and vice versa. People leaving a country frequently fail to register their departures, and in many cases they are not departing from one permanent residence but moving between one residence and another on a seasonal or periodic basis. Poulain and colleagues have documented the statistical systems used to record international migration between European states, and present a matrix of migration flows between EU member states, from both receiving and sending country statistics. Most cells have at least one entry, though there are a few empty cells. The discrepancies between the two views can be very large: in 2002 Germany reported 78 739 people migrating to Poland, while Poland reported only 2335 arriving! In general, immigration statistics are considered more trustworthy as immigrants normally have to fill in a form on entry or to obtain work, education, or healthcare. To arrive at a reasonable estimate of the migration flows between European countries, Raymer had to design an ingenious and elaborate algorithm to make best use of the available data. There is also error introduced as a result of failure to count illegal migrants.

Flows within Europe, though large, are not the largest international migrations in Europe, which involve migrations from and to the world outside Europe. In recent years, both Spain and Italy have received very large flows of migrants from abroad, 2004 of 610 000 and 558 000 net migrants, respectively, compared with 204 000 into the United Kingdom, 105 000 into France, and only 82 000 into Germany, according to statistics assembled by the Vienna Institute of Demography.

A team led by Kupiszewski at the Central European Forum for Migration and Population Research has recently carried out explorations of the future population of 27 European countries under different international migration scenarios. They show that under no plausible scenario of net gains from international migration will there be much alleviation of the effects of population aging on old age dependency. The projection model interestingly divides the international migration flows into two sets, those between the 27 countries which are modeled as part of a multistate population projection, and the net immigration flows from outside Europe which will make the biggest impact on the future population. In effect, what was formerly international migration within Europe has become very similar to domestic or intranational migration.

When we move from a European to a world canvas, it becomes even more difficult to paint a reliable picture of international migration flows. One attempt to do so is by Dorling and colleagues in their Worldmapper project. They have assembled statistics on the populations in each country born abroad which yields a lifetime immigration estimate, and then used that information to reconstruct the distribution of those foreign populations by origin country to yield a lifetime emigration estimate. Subtraction of emigration from immigration produces a net international migration estimate based on foreign population stocks. When we compute those net balances for 200 countries, expressed as percentage of the resident population, and plot them against median male earnings, we find a moderate positive correlation ($r = 0.28$) between balance and the income variable, indicating that the higher a country's per capita income the more likely it is that the balance of migration will be inward.

Because of difficulties in measuring international migration in a comparable way, most work has focused on immigrant populations, and the ways in which they interact with and integrate with host populations and into sources of conflict and lack of cohesion.

Intranational (Internal) Migration Measures and Trends

Intranational migration is defined as changes of usual residence within countries. Some researchers have distinguished between local mobility and longer distance

Table 4 Comparable measures of internal migration, Australia and Britain

Measure	*Australia*			*Britain*		
	1980 81	*1995 96*	*1995 96*	*1980 81*	*1990 91*	*1990 91*
Migration expectancy (moves per lifetime)						
Males	10.7	12.8		6.4	6.2	
Females	11.1	13.6		6.7	6.5	
Median distance moved (km)		Inter-zone	All moves		Inter-zone	All moves
38 city regions		248	41			
65 Time Consistent Statistical Districts		178	41			
67 counties					94	28
459 districts					34	6

Source: Adapted from Bell, M., Blake, M., Boyle, P. *et al.* (2002). Cross national comparison of internal migration: Issues and measures. *Journal of the Royal Statistical Society A* 165(3), 435 464.

migration because of the different mix of motivations at the two scales, but, in practice, is difficult to compartmentalize what is a continuously varying phenomenon.

Migration levels and intensities vary with the spatial scale of measurement: the more territorial units there are the greater the level of migration, when migration is defined as a border crossing. This makes it difficult to construct comparable measures of migration between countries, except for migration at the smallest scales of all residential moves. Long, an American demographer, has established that the new countries of North America and Australasia have much higher internal migration levels than European countries. Courgeau, a French demographer, proposed a simple linear model of the relationship between migration intensity and the logarithm of the number of spatial units used in migration measurement be used to compare internal migration between countries, but the idea has not found favor. A full audit of measures that can be used to compare internal migration across countries has been undertaken by Bell and colleagues. They recommended adoption of measures capturing four dimensions of internal migration: migration intensities, migration distances, migration connectivity, and migration effects or impacts. Their discussion is illustrated by a rigorous comparison of migration in Britain and Australia using comparable measures. **Table 4** reports on migration expectancy and on median distance moved in Australia and Britain, showing Australians to be twice as migratory as the British and migrating over twice as far. However, this measurement agenda has yet to be adopted by any international statistical agency.

Most applied work that compares internal migration between countries has focused on the structure and patterns using net migration, one of the impact measures. For example, a study of internal migration in ten European countries in the 1980s and 1990s identified whether internal migration was contributing to urbanization or to counter urbanization.

The relationship of migration intensity and age has been studied and modeled extensively by Rogers and collaborators over several decades. They have developed a family of models, based on combining exponential functions, which describe the schedule of migration intensities plotted against age. **Figure 5** illustrates the ways that these can be assembled. All four versions contain the single negative exponential decline during the childhood ages and the double exponential that represents rising and then falling mobility of labor market entry, career moves, followed by a settled occupation. Some migration intensity/age schedules exhibit an elderly retirement peak (another double exponential), when couples relocate after exit from the labor force. This type of schedule is found in migration flows from metropolitan centers to regions of retirement living. More common are schedules with a single positive exponential after retirement, reflecting migrations triggered by loss of spouse/partner or increasing infirmity.

Because migration consists of a move from an origin to a destination over intervening space, a large number of spatial interaction models of migration have been developed that seek to explain the intensity of migration in terms of a set of origin variables, a set of destination variables, and measures of the cost of crossing the space between origin and destination. There have been a huge number of the determinants of migration flows between sending and receiving regions within countries, using a variety of statistical models of migration. A recent analysis of UK migration constructed models for 98 origins/destinations and 14 age–sex groups and used more than 40 potential explanatory variables.

Processes of Population Change and Their Implications

This article concludes with the study of population through the lens of demography by putting the different components of change together to see what major changes have occurred or are occurring to populations in different parts of the world.

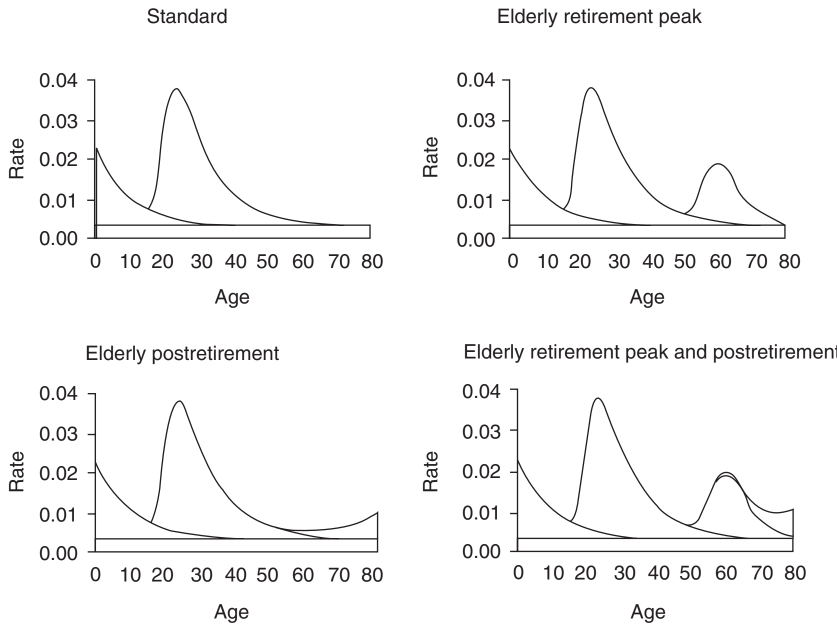

Figure 5 The four main families of multi-exponential model migration schedules. Reproduced with the kind permission of Prof. Andrei Rogers.

The Demographic Transitions

Over the past century and a half, national populations have experienced a shift in their demographic regimes from high fertility, high mortality, and low and uncertain growth to low fertility and low mortality, and uncertain growth through an intermediate phase (transition) of fertility decline and mortality decrease. This is the first demographic transition. Because mortality decline preceded fertility decline and often proceeded more quickly, the result was high natural increase and rapid population growth. The demographic transition, its course in different continents and epochs, and its drivers and causes are extensively discussed in demographic texts. What is less explored is the way the demographic transition has favorable economic impacts because of the increasing concentration of the population in the working ages (fewer children, relatively, and more surviving into later working ages). Bloom and Williamson have argued convincingly that an important factor explaining the extraordinary growth performance in East and Southeast Asia in the past 30 years has been this concentration of the population in the productive ages. In the current and next two decades, East Asia will lose this advantage as its population ages beyond the working years.

Europe's fertility decline did not stop at replacement fertility level but moved into much lower fertility territory after 1970. This is the second demographic transition. By 1985, virtually all European countries had fertility levels below 2.1 children per woman and low fertility has persisted since then. A range of explanations have been advanced to explain this new demographic regime: these include shifts to new philosophies of individualism and careers over family for women, women's freedom to control their own fertility through the contraceptive pill, the increase in female labor force participation (and opportunity cost of children), increased female incomes and therefore independence, the decline of marriage and the rise of cohabitation, the renewed rise in childlessness, women's increased participation in higher education, and the failure of the state and market to provide enough early childcare support (to make possible both work and children for women). Interestingly, fertility intentions (in an ideal world) seem not to have shifted in surveys from a modal choice of two children.

Persistence of low fertility over several decades leads to reduced entries to the labor force and so to shortages of workers. In many European countries, this has led to renewed immigration by foreign workers and their families, which over time will begin to change the composition of the population, particularly if the immigrants have, at least initially, higher fertility than the native population. This is the third demographic transition. Coleman has documented this process for a number of European countries, and gathered together evidence on projection of European country populations distinguished by foreigner/citizen or ethnicity. This shows substantial changes in nationality or ethnicity composition of country populations taking place, with more anticipated in the next half century, although the shifts are dependent on the immigration policies in place. This population replacement process is also occurring in North America and Australasia.

The Ageing of the World's Population

The most important consequence of the demographic transitions is population aging. Population aging is the process that results in higher average ages and higher shares of the population being old. The threshold age used by the United Nations for defining the old age population is 60 years. Population aging occurs because smaller new birth cohorts are born into populations as a result of fertility decline and because more people survive to older ages with lower mortality rates. Migration redistributes both younger and older populations, and so affects the degree of aging of countries and subnational regions and localities. The United Nations report on World Population Aging makes several critical observations: (1) that population aging has been a new phenomenon of the twentieth century and will be even more pronounced in the twenty first century; (2) that aging affects virtually all countries in all continents, whether rich and poor; (3) that the pace of aging for many countries in East and Southeast Asia will be more rapid than that experienced by countries in Europe and North America, because their fertility decline in the demographic transition has been so much faster in the past 40 years; (4) that it is highly unlikely that the process will be reversed, given that it occurs under virtually all plausible future scenarios for demographic change; and (5) that aging will have profound social and economic implications for human societies.

Migration and Ageing

National policy makers and employers now often ask whether the effects of aging on the labor supply and on the support available for retired workers can be mitigated by accepting immigrants from other countries. This is called the replacement migration question, and was explored in an influential report by the United Nations. The migration assumptions in population projections for eight developed countries far into the aging process were adjusted to achieve particular outcomes in terms of age structure. These scenario projections showed that to achieve, for example, a long run, constant old age dependency ratio, would require immigration at completely unprecedented and unrealistic levels. The same exercise was carried out for 27 European countries by the Central European Forum for Migration Research using a projection model that handled migration in a more sophisticated fashion. Again the results were that international migration into European countries could make a small impact in slowing aging, but to achieve a stable level of the old age ratio or the old age share would require huge migration flows, which were demographically, let alone politically, infeasible.

Aging poses huge challenges for societies affecting pension and social security systems, demand for healthcare and the needs of the infirm elderly for family, social, and state support. Decisions need to be taken about ages of entitlement to state and occupational pensions and the length of working life, the relative responsibilities of individuals, families, and younger generations for support for older people. New intergenerational contracts will be needed. But population aging should be seen as an enormous success for the societies experiencing it and one that can be coped with through new fair arrangements and through continued growth in economic productivity. The knowledge of demographic trends past and future, such as outlined in this article, provide policymakers with long lead times for taking appropriate action.

See also: Census Geography; Emigration; Immigration I; Immigration II; Migration; Population Geography.

Further Reading

Bijak, J., Kupiszewska, D., Kupiszewski, M., Saczuk, K. and Kicinger, A. (2007). Population and labour force projections for 27 European countries, 2002 2052: Impact of international migration on population ageing. *European Journal of Population* 23(1), 1 31.
Bloom, D. E. and Williamson, J. G. (1998). Demographic transitions and economic miracles in emerging Asia. *World Bank Economic Review* 12(3), 419 455.
Bongaarts, J. (1978). A framework for analyzing the proximate determinants of fertility. *Population and Development Review* 4(1), 105 132.
Coleman, D. (2006). Immigration and ethnic change in low fertility countries: A third demographic transition. *Population and Development Review* 32(3), 401 446.
Lutz, W., Sanderson, W. and Scherbov, S. (eds.) (2004). *The End of World Population Growth in the 21st Century*. London: Earthscan.
Oeppen, J. and Vaupel, J. (2002). Broken limits to life expectancy. *Science* 296(10 May 2002), 1029 1031.
Olshansky, S. J., Passaro, D. J., Hershow *et al*. (2005). A potential decline in life expectancy in the United States in the 21st Century. *The New England Journal of Medicine* 352(11), 1138 1145.
Rogers, A. (1995). *Multiregional Demography: Principles, Methods and Extensions*. Chichester: Wiley.
Rowland, D. T. (2003). *Demographic Methods and Concepts*. Oxford: Oxford University Press.
United Nations (2002). *World Population Ageing: 1950 2050*. Department of Economic and Social Affairs, Population Division. New York: United Nations. http://www.un.org/esa/population/publications/worldageing19502050/index htm (accessed Jun. 2008).
van de Kaa, D. (1987). Europe's second demographic transition. *Population Bulletin* 42(1), p64.
Wilson, T. and Rees, P. (2005). Recent developments in population projection methodology: A review. *Population, Space and Place* 11, 337 360.

Relevant Websites

http://www.who.int
Data and Statistics, World Health Organisation.
http://www.demographic research.org
Demographic Research.
http://www.coe.int
Demographic Yearbook, Recent Demographic Developments in Europe, Council of Europe.
http://www.popassoc.org
Demography, Population Association of America.

http://epp.eurostat.ec.europa.eu
Eurostat.
http://www.springer.com
European Journal of Population, Springer.
http://www.ined.fr/en
Institut national d'études démographiques (INED), Paris.
http://www.census.gov
International Programs Center, US Census Bureau.
http://www.jpr.org.au
Journal of Population Research.
http://www.demogr.mpg.de
Max Planck Institute for Demographic Research, Rostock.
http://www.nidi.knaw.nl/en
Netherlands Interdisciplinary Demographic Institute (NIDI), the Hague.
http://www.popcouncil.org
Population and Development Review, Population Council.
http://www.jstor.org
Population Index.
http://popindex.princeton.edu
Population Index.
http://www.prb.org
Population Reference Bureau.
http://www.ined.fr/en
Population, Resources and Documentation, INED Publications.
http://www3.interscience.wiley.com
Population, Space and Place, Wiley.
http://www.tandf.co.uk
Population Studies, Taylor and Francis.
http://www.un.org
United Nations Population Division.
http://www.oeaw.ac.at
Vienna Institute of Demography (VID), Vienna, Osterreichische Akademie der Wissenschaften.
http://www.worldbank.org
World Bank.
http://www.worldmapper.org
Worldmapper.
http://www.iiasa.ac.at
World Population Program, International Institute for Applied Systems Analysis (IIASA).

Dependency

E. Hartwick, Framingham State College, Framingham, MA, USA

Glossary

Center Point in space where power accumulates.
First World Rich industrialized countries of the world.
Import Substitution Industrialization Replacing imported manufactured goods with locally produced goods.
Periphery Zones from which power and resources are drained.
Terms of Trade Ratios at which goods are exchanged during trading.
Third World Poor underdeveloped countries of the world.
Uneven Development Tendency for development to occur differentially in different places.

Introduction

Dependency is a basic characteristic of the operations of power in space. Under most kinds of class societies, especially capitalism, power accumulates at points in space, be these significant individuals or global cities. These points, often designated as 'centers', exercise power over domains, or 'peripheries' that vary from personal space to regional and global zones. The power of centers is derived from draining the peripheries of people and wealth. Power is exercised in a number of forms: as knowledge and discourse, as capital investment and financial expertise, or as militaristic intervention and the occupation of countries. The peoples, institutions, activities, and states of the peripheries have a dependent relation to the centers – dependent meaning that their choices are constrained and their possibilities limited by their power relations with centers. While the geography of power relations may change over time as, occasionally, new centers arise from what previously were peripheries, and as the form taken by power relations shifts from one kind of dominance to another, critical theorists claim that dependency continues as a fundamental characteristic, a basic inequality in human, social existence.

Maps of the World

When critical human geographers gaze at global space, they see not stretches of even development, as with the homogeneous plains assumed by industrial and agricultural location theory, but uneven development, regional inequality, centers dominating peripheries and inventing sophisticated legitimating ideologies that excuse their doing so. Perhaps the main tradition in critical human geography expressing this general perspective looks at relations between the global center known as the 'First World' and the range of dependent regions collectively known as the 'Third World'. The discipline of human geography has long had an interest in the Third World. This interest was revived by US and British involvement in the Vietnam War, which many radical academics interpreted as a symptom of First World imperialism. A series of articles drew on the existing stock of sophisticated radical ideas dealing with questions of underdevelopment, center–periphery relations, and imperialism to provide theoretical insight for oppositional political practice.

In the best of this work, James Blaut drew on the notion of dependency to argue for a biased European and North American, First World view of the world that made it impossible for them to understand erupting global events. Conventional Western science, Blaut argued, is so closely interwoven with imperialism (defined as white exploitation of the non white world) that it can only describe and justify, but not explain or control, major events like wars and invasions. Imperialism, he said, is underpinned by Western 'ethnoscience', that is the cognitive system characteristic of a given cultural–linguistic universe – in this case the white, European West – that provides the system of scientific ideas with which people think and work. European ethnoscience contains a set of historical beliefs and social–scientific generalizations about the world biased in favor of whites and congruent with the interests of Western imperialism. For Blaut the European model of the world has a unicentric spatial form with a distinctive geometry, an inner European space originally closed from an outer non European space. Inside this space, 'The West' exercises some kind of unique historical advantage (race, environment, culture, mind, spirit) giving it superiority over all other peoples. European civilization is supposedly generated by inner processes – Europe makes history – while non Europeans play little or no crucial role in epochal events – 'the rest of the world' has history made for it by the progressive white people in the classic style of dependency. Non Europeans are characterized as primitive and unprogressive, barbarous and heathen, less intelligent and less virtuous than Europeans. The expansion of the Europeans is thought of as self generated, a 'striving outward', with progress diffusing from points of origin, and becoming less evident with distance from the center. Whenever non Europeans show evidence of pro gress, this is proportional to the European impact on their society – this amounts to a global diffusion model and the belief generalized is 'diffusionism' (i.e., cultural processes flowing from the European center to the non European peripheries). Blaut listed 30 reasons why Europeans were supposed to be better than everyone else. These included: factors of environment, as with climate, soil fertility, land form structure, coastline shape, vegetation, the distribution of natural disasters, and the incidence of disease; factors of culture, as with inventiveness, rationality, scientific thinking, progressiveness, religion, and institutions like markets; social factors, like class structure, family, and cities; and political characteristics as with the form of the state and the rise of democracy. In Blaut's mind, environmentalism continued to form the underlying causal emphasis behind most theories of national and regional supremacy.

For radical geographers like Blaut, Eurocentrism had to be criticized because of its intellectual corruption and its arrogant dismissiveness. However, there are many ideas widely considered to be great themes of history that are utterly mistaken. Thus, additionally, Eurocentrism needed criticizing out of existence because it was dangerous to the world's peoples. For, in its name, white people went out to conquer, control, and irrevocably change a world they fundamentally misunderstood. Time after time, the Euro peans committed genocide on peoples they never knew. Eurocentrism had to go because it was a fatal ignorance.

Blaut compared this with a Third World ethnoscientific model of the world as multicentric, with centers of incipient or proto capitalism springing up at strategic points throughout the Old World. In a Third Worldist under standing, the multicentered pattern of relatively equal levels of development was disrupted by the European plunder of the New World (the European 'discovery' being due solely to the fact that the Iberian centers of expansion were closest to America), the flood of bullion into Europe, and the re sulting commercial, industrial, and scientific–technological development. After that, the dialectic of development and underdevelopment intensified, and the world economy fixed itself in place. From this perspective, Blaut argued against the notion of the 'European miracle' by undermining its more concrete theory of the autonomous rise of Europe. Thus: (1) Europe was not superior to other regions prior to 1492; (2) as in dependency theory, colonialism and the wealth plundered from Third World societies were the basic processes leading to the rise of Europe and the fall of Third World civil izations; and (3) Europe's advantage lay solely in the mun dane realities of location – that is, nearness to the Americas.

Dependency Theory

Blaut's map of the world provides a good context for understanding the main line of theoretical elaboration of the concept of dependency. The notion of dependency in development is best expressed by a school of thought called dependency theory. The basic idea of the de pendency school is that the development of Europe and North America was predicated on the active 'under development' of the non European world – that is, making Third World societies less developed than they had previously been. As a result, Third World countries can achieve at best forms of development that make them dependent on the First World for capital, technology, markets and, especially at times of emergency, aid, loans, and grants. Dependent forms of inequality are built into global space. These inequalities are inherent in the un even development of capitalism. They are caused by the exploitation of the peoples and societies of peripheral regions by the capitalist class of the center. Global in equality may be ameliorated by charity, but it can be eradicated only through revolutionary transformation to a new global order.

Dependency theory began with criticisms by Latin American economists of the claim to universal applic ability made by classical and neoclassical economics – that is, the same economic laws applied to all countries and regions. To the contrary, the United Nations Economic Commission for Latin America (ECLA), working during the 1950s and 1960s, found that conventional comparative advantage and trade theories expressed the experience of the industrial countries of the center, rather than the agricultural countries at the periphery of the world, and were inadequate for understanding the historical experi ence of the underdeveloped world. Instead an appreciation of the different historical contexts and natural situations of these countries, their different social structures, types of behavior, economies, and external relations required a new structuralist perspective that took basic structural differ ences among economies into account. In particular, Raul Prebisch, the most distinguished theorist working with ECLA, saw the world not in mono economic terms, as one homogenous system, but as two distinct areas, a center of economic power in Europe and the United States, and a periphery of weak countries in Latin America, Africa, and Asia. Whereas comparative advantage theory argued that the exchange of central industrial goods for peripheral primary goods was to the periphery's advantage, Prebisch argued that Latin America's peripheral position and pri mary exports were exactly the causes of its lack of pro gress, specifically because a long term decline in the periphery's terms of trade (the ratio between the value of exports and the value of imports) made international ex change work almost exclusively in favor of industrial producers. For Prebisch, the solution for Latin America lay in structural change and specifically the industrial ization of peripheral countries, using an import substi tution strategy (i.e., replacing industrial imports with domestic production, under the cover of tariff protection).

However, the ideas of ECLA and Prebisch were later criticized by the Latin American left for ignoring class relations: state intervention in the economy, such as the protection of infant industries from foreign competition, could end up subsidizing the profits of the local bour geoisie, with consumers paying higher prices – at one time the tariff on refrigerators imported into Mexico was 800%. A more radical dependendista position was pieced toge ther by writers such as Galeano, Sunkel, Furtado, Cardoso, Falleto, and Dos Santos, and popularized in the English speaking world through the writings of Andre Gunder Frank.

Chains of Exploitation

Frank was a leading critic of modernization theory – the notion that progressive change began in the center and spread to a periphery that was always relatively back ward. This entailed Frank's criticism of the 'dual society' thesis, that underdeveloped societies had a dual structure of modern and traditional sectors, each with its own dynamic, and poorly fitted together. For Frank attributing underdevelopment to traditionalism (or feudalism), ra ther than capitalism, was a historical and political mis take. World capitalism had destroyed or transformed earlier social systems even as it came into existence, converting them into sources of its own further devel opment. For Frank, the economic, political, social, and cultural institutions of the underdeveloped countries resulted instead from the penetration of capitalism, ra ther than being original or traditional. Frank focused on the metropole–satellite (or center–periphery) relations he found typical of Latin America. The underdevelop ment of peripheral capitalist regions and people, he said, was characterized by three contradictions: (1) the con tradiction of the monopolistic expropriation of economic surplus; (2) the contradiction of metropolis–satellite po larization; and (3) the contradiction of continuity in change. Drawing on Marxist analyses of the class ex propriation of surplus value, especially Baran's version which emphasized the potential surplus which could be made available for investment under noncapitalist cir cumstances, Frank argued that external monopoly re sulted in the expropriation (and thus local unavailability) of a significant part of the economic surplus produced in Latin America. The region was actively 'underdeveloped' (made less developed) by the expropriation of its surplus product (source of investment capital in Marxist theory). Using a case study of Chile, Frank described the pattern of surplus movement as a massive, geographical ex propriation system reaching the most remote corners of the region, with a surplus extraction system running through it. This chain like set of links connects the capitalist world and national metropolises to the re gional centers (part of whose surplus they appropriate), local centers, and so on, down to large landowners or merchants who expropriate surplus from small peasants or tenants, and sometimes even from these latter to landless laborers exploited by them. At each step a relatively few capitalists above exercise monopoly power over many below, expropriating their economic surplus. Thus at each point, the international, national, and local capitalist system generates economic development for the few and underdevelopment for the many. This idea of surplus transfer over space was further developed by Frank as a system whereby center and periphery become increasingly polarized as capitalism developed the one, and underdeveloped the other, in a single historical process. In this perspective, only a weaker or lesser de gree of metropole–satellite relations allowed the possi bility of local development. These two systems suggested a third to Frank – the continuity and ubiquity of struc tural underdevelopment throughout the expansion of the capitalist system.

From this perspective on underdevelopment, Frank generated more specific hypotheses that could be used in guiding development theory and policy. In contrast to the world metropolis, which was satellite to no other region, the development of national and regional metropolises was limited by their satellite status – for example, local metropoles such as São Paulo, Brazil, or Buenos Aires, Argentina, could only achieve a dependent form of in dustrialization. Real development thus entailed separ ation and autonomy from the global capitalist system. Similarly, in a hypothesis directly opposed to the finding of modernization geography, that development was spread through contract with the metropolis, Frank hy pothesized that the satellites experienced their greatest development when ties to the metropolis were weakest – historically during wars, geographically in terms of spa tial isolation. By extension, regions which had the closest ties to the metropole in the past were the most under developed in the present – Frank found this confirmed by the 'ultra underdevelopment' of the sugar exporting re gion of northeastern Brazil and the mining regions of Bolivia. In summary, underdevelopment in Frank's theory was not an original condition, nor the result of archaic institutions surviving in isolated regions, nor even did it come from Third World irrationalism. Underdevelop ment was generated by the same processes which de veloped the center; in particular, underdevelopment in the periphery resulted from the loss of surplus which was expropriated for investment in the center's development.

An immediately noticeable weakness in Frank's theory lay in his failure to specify the economic mechanisms of surplus extraction. In some cases the mechanisms of surplus extraction were obvious – for example, when European, North American, or (later) Japanese or Chi nese corporations owning land, factories, and resources in Latin American countries and withdrawing surplus as rent or profit, or when the international banks, the International Monetary Fund (IMF) or World Bank, loaned capital to peripheral states and enterprises and withdrew surplus as interest payments. But what of peasant producers, owning their own land, and producing cash crops for export to center markets, a situation typical of much of the peripheral agriculture in the nineteenth and twentieth centuries?

Here the beginning of an answer was provided by Arghiri Emmanuel in the theory of unequal exchange. Like the ECLA economists, Emmanuel argued against neoclassical trade theory, which said that the inter national division of labor and system of trade had ad vantages for all participants. Emmanuel argued instead that trade made poor countries poorer, and rich countries richer – so that arguments for free trade were arguments for maintaining the advantage of developed countries. Emmanuel assumed the perfect international mobility of capital, but the immobility of labor between countries – hence wage rates persistently differed greatly between the two sets. Peripheral countries exported agricultural products which embodied large quantities of cheap labor and imported industrial products embodying small amounts of expensive labor. This led to terms of trade favoring the higher cost products of the center, while devaluing the lower cost exports of the periphery. Per ipheral countries were prevented from achieving devel opment because they sold their goods at prices below values (the socially necessary labor embodied in the products), while rich countries sold goods at prices above values. For Emmanuel, unequal exchange (through trade) was a hidden mechanism of surplus extraction and a major cause of the economic stagnation in the periphery. Samir Amin estimated the amount of surplus transferred from poor to rich countries via unequal exchange to be 1.5% of the product of the rich countries, but 15% of that of the poor countries, an amount, he thought, that was "sufficient to account for the blocking of the growth of the periphery." From the perspective of dependency theory the peripheral countries borrowed back their own surplus to finance 'development'. The geopolitical im plications of this are enormous – debt forgiveness is justified as reparations for injuries done in the past and continuing into the present, rather than the conventional justification, as charity for the poor.

For dependency theorists, Europe's development was based on external destruction – brutal conquest, colonial control, and the stripping of non Western societies of their peoples, resources, and surpluses. From historical processes like these came a new global geography of European First World center and non European Third World periphery. The relationship between center and periphery assumed, for the Brazilian geographer Teontonio Dos Santos, the spatial form of dependence, in which some countries (the dominant) achieved self sustaining economic growth, while others (the dominated and dependent) grew only as a reflection of changes in the dominant countries. The incorporation of Latin America into the capitalist world economy, directly through (Spanish and Portuguese) colonial adminis tration, but more subtly through foreign trade, geared the region's economies toward demands from the center. Dependence skewed the region's social structure so that local power was held by a small ruling class which used the gains derived from exporting for luxury consumption rather than investment; real power was exercised from external centers of command in dominant ('metro politan') countries. Dependence continues into the pre sent through international ownership of the region's most dynamic sectors, multinational corporate control over technology, and payments of royalties, interest, and profits to external corporations, banks, and development agencies.

Yet, there were other, more serious, criticisms of Frank. The Brazilian economist Fernando Henrique Cardoso found the notion of the development of underdevelop ment a neat play on words, but not helpful in concrete terms. In Latin America, he said, multinational corpor ations invested in modern industrialization, while sup posedly traditional sectors (agriculture, mining) operated in technically and organizationally sophisticated ways, and both were parts of an advanced yet still dependent cap italist development. However, he added that in countries like Argentina, Brazil, and Mexico spatial and sectoral dualisms had emerged, with space divided between ad vanced regional economies tied to the international cap italist system, and backward sectors, or internal colonies, isolated from progress. Multinational corporations were interested in at least some development for dependent countries because of the markets this prosperity provided. However, the Latin American countries remained dependent for technology on the United States. In contrast to Frank's universalism, Cardoso wanted to look at specific situations in particular parts of the Third World, where development and dependence could be found in tandem. We might note that, on becoming president of Brazil in the 1990s, Cardoso adopted a neoliberal policy position in line with his notion of dependent development.

Exceptions to Dependency

The most serious criticism of dependency in develop ment comes from the apparent ability of previously Third World countries to industrialize and achieve nondependent forms of development. Balassa argued for a stage theory emphasizing the choice of economic policy in setting a course and pushing a country through his torical phases of industrialization. Industrial development was said to begin in Third World countries as a response to the needs of the primary sector (processing raw ma terials, providing simple inputs, etc.). Subsequently a first or 'easy' stage of import substitution, of the kind sug gested by dependency theory, entailed the local manu facture of previously imported nondurable goods, and a second stage involved the replacement of imports of intermediate goods (petrochemicals, steel), producer durables (machinery), and consumer durables (auto mobiles) by domestic production – this was undertaken in the post World War II period by the Latin American countries, some South and East Asian countries, and the East European countries. Problems such as the small size of the market (when industries are characterized by considerable economies of scale) led to the need for considerable state protection.

Export oriented policies were adopted by Japan in the mid 1950s, Korea, Singapore, and Taiwan in the early 1960s, and the Latin American countries in the mid to late 1960s. Korea, Singapore, and Taiwan which imple mented 'free trade regimes' (exporters could choose to use local or imported inputs in their manufacturing) increased industrial exports rapidly in the early 1960s, and in the late 1960s and early 1970s various incentives for exporting led to further growth in manufacturing employment and higher incomes. Countries that retained inward looking strategies (India, Chile, Uruguay) re mained at the bottom of the industrial growth league. For Balassa the evidence was conclusive: countries applying outward oriented development strategies had a superior performance in terms of exports, economic growth, and employment, whereas countries which continued inward orientation encountered increasing economic difficulties. This view was given credence by the continuing success of the New Industrial Countries (NICs) in the 1970s and 1980s. The developed countries had economic growth rates averaging 4.8% a year in the period 1964–73 and 2.1% a year in 1973–83; the NICs (Brazil, Mexico, Hong Kong, Singapore, Taiwan) collectively had growth rates of 8.4% and 5.3% in the same periods, with the East Asian countries sustaining growth rates in the order of 10% a year often for a decade or more. In the mid 1980s a

new wave of industrialization in Indonesia, Malaysia, Thailand, and (later) China seemed to confirm the view that dependency theory had been killed by develop mental reality.

Dependency Reasserted

The development model of the NICs retains many aspects of dependency. The NICs usually rely on foreign direct investment in their early stages. They import technology continuously from existing centers of accumulated knowledge, for which licensing fees have to be paid. In addition, they serve markets in the First World where competition is severe from countries like China, where labor costs are even lower. The period of 1980–2006 saw recurrent debt and other kinds of financial crises exactly in countries like Mexico, Brazil, Argentina, South Korea, and Thailand that supposedly provided the 'exceptions' to dependency. In such times of crisis, the NICs applied to the IMF for short term loans to see them through, and to the World Bank for longer term restructuring loans. The international financial institutions (IFIs) are domin ated by the biggest economies in the world, as with the US, Japan, Germany, and the UK. The IFIs placed 'con ditionalities' on the loans, by which countries undergoing crisis had to restructure their economies in the direction of neoliberal models of development. By this means de pendency is deepened to include control over the eco nomic structure and the trajectory of development. Thus, in counterargument, the rumors of the death of depend ency have been exaggerated. Dependency is not dead. It has just changed its institutional form, and widened its application.

See also: Capitalism; Colonialism I; Colonialism II; Debt; Development I; Development II; East Asian Miracle; Empire; Eurocentrism; First World; Latin American Structuralist School; Modernization Theory; Neocolonialism; Social Class;Third World.

Further Reading

Baran, P. (1960). *The political economy of growth*. New York: Monthly Review Press.

Balassa, B. (1981). *The newly industrializing countries in the world economy*. New York: Pergamon Press.

Blaut, J. M. (1976). Where was capitalism born? *Antipode* 8(2), 1 11.

Blaut, J. M. (1994). *Colonizer's model of the world*. New York: Guilford.

Blaut, J. M. (2000). *Eight Eurocentric Historians*. New York: The Guilford Press.

Cardoso, F. and Faletto, R. (1979). *Dependency and development*. Berkeley: University of California Press.

Dos Santos, T. (1970). The structure of dependence. *American Economic Review* 60, 231 236.

de Janvry, A. (1981). *The agrarian question and reformism in Latin America*. Baltimore: Johns Hopkins University Press.

Emmanuel, A. (1972). *Unequal exchange: A study of the imperialism of trade*. New York: Monthly Review Press.

Frank, A. G. (1969). *Latin America: Underdevelopment or revolution?* New York: Monthly Review Press.

Frank, A. G. (1969). *Capitalism and underdevelopment in Latin America*. New York: Monthly Review Press.

Frank, A. G. (1979). *Dependent accumulation and underdevelopment*. New York: Monthly Review Press.

Galeano, E. (1973). *Open veins of Latin America: Five centuries of the pillage of a continent*. New York: Monthly Review Press.

Peet, R. and Hartwick, E. (1998). *Theories of development*. New York: Guilford.

Prebisch, R. (1972). *International economics and development*. New York: Academic Press.

Desertification

C. J. Barrow, Swansea University, Swansea, UK

Glossary

Biodiversity Biological diversity in terms of species and variation within each species.

Desiccation Drying conditions (aridification) resulting from a dry phase of climate, lasting a decade or more.

Drought A period of more than 2 years with below-average precipitation.

Dust Bowl An area stretching across several of the Midwest states of USA, where poor land management and drought led to severe soil degradation and dust storms between 1933 and 1936. This region of desertification stimulated interest in soil and water conservation and improvement of land husbandry. Dust Bowl is sometimes applied to other regions showing similar degradation.

Global Environmental Facility (GEF) Established in 1991 as a corporate venture between industrialized and developing nations it is jointly managed by the World Bank, United Nations Development Programme (UNDP), and United Nations Environment Programme (UNEP). It is intended to assist poorer countries to develop policies and programs that offer global environmental benefit; for example, desertification control; biodiversity protection; management of international waters; mitigation of global warming; reduction of stratospheric ozone depletion, and so on.

Land Degradation Loss of utility or potential utility or the reduction, loss, or change of features or organisms, which cannot be replaced. This can be due to natural causes, but is more often wholly or in large part caused by human mismanagement. It can be viewed as the opposite of sustainable land use. Desertification implies degradation toward 'desert' conditions, some danger of irreversibility, and more prevalent in periodically dry regions.

Retrogression A regressive process whereby climax vegetation (that in 'balance' with local conditions) and soil reverts to a less-developed stage, which might even be bare ground.

Sahel The belt of semiarid and subhumid land with variable seasonal rainfall lying south of the Sahara, extending from the Atlantic Cape Verde Islands and West Africa through the Sudan to the Indian Ocean.

Salinization Land degradation and desertification due to the accumulation of salts and sodic compounds through natural or human causes.

Introduction

Desertification was recognized long before it gained a name from Aubreville. Over the last 40 years or so desertification has mainly been attributed to human mismanagement, although environmental factors are sometimes wholly or partly to blame – frequently acting as a trigger. Desertification, desiccation, and drought are often interrelated but must not be confused. As much as one fifth of the global population lives in regions at serious risk from desertification and includes some of the poorest peoples. Possibly 35% of the world's land surface is affected (**Figure 1**).

The extent has often been misreported, and the pro cesses causing it are not adequately understood. 'Desert spreading' is sometimes used to describe desertification, which is misleading, because, while deserts may expand, much desertified land develops *in situ* (sometimes far from existing deserts). Poor definitions mean that re searchers may make differing measurements, giving false pictures and making it easier for agencies and policy makers to argue whether to fund countermeasures. It has been argued that desertification should be recognized only when degradation is irreversible. However, there are regions where it has been bad and then ameliorated naturally and others where remediation has been suc cessful or may be technically possible. Although in many cases amelioration is theoretically possible, it is not af fordable or it is not practical to organize it. Desert ification occurs in rich and poor countries, where there is high or low population, in communist or capitalist states, and it can be found in quite wet and cool environments.

There is no universally accepted definition; however, most recognize that vegetation and soil undergoes re gression to something more like desert. It is a process of 'degradation of' land mainly 'in arid, semi arid, and dry subhumid areas resulting from various factors including climatic variations, but predominantly human activities'. However, some feel 'desertification' is a poor term and favor 'dryland degradation'.

The Process of Desertification

Palaeo dune fields ('fossil dunes'), desert varnish (rock encrustation indicative of past desert conditions), and fluctuating lake levels show deserts fluctuated before humans had any influence. Historical, archaeological, and palaeoecological evidence indicates humans started to cause desertification in some regions many thousands of

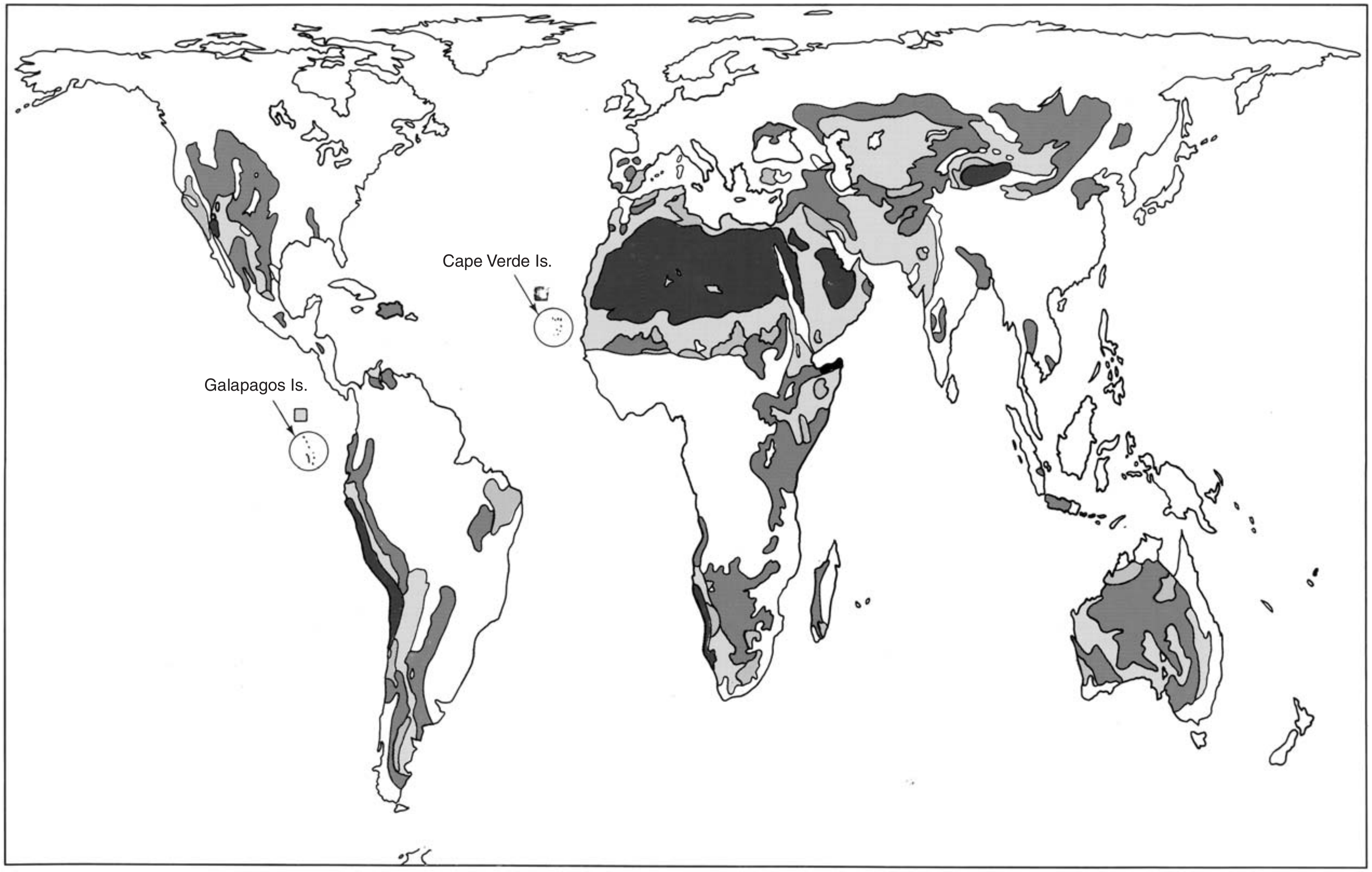

Key:

- Slight desertification
  Little or no degradation of plant and soil cover or salinization.
- Moderate desertification
  26 to 50% of plant community consists of climax species, *or* 25 to 75% or original topsoil lost, *or* soil salinization has reduced crop yield by 10 to 50%.
- Severe desertification
  10 to 25% of plant community consists of climax species, *or* erosion has removed practically all topsoil, *or* salinization has reduced crop yield by more than 50%.
- Less than 10% of plant community are climax species, *or* there are many sand dunes or gulleys, *or* salt crusts have developed.

Figure 1 Deserts and areas subject to desert ficat on.

years ago through burning, hunting, and deforestation. In the last few centuries human causation has overtaken natural, and there are fears desertification will spread and get worse as global warming occurs.

Vulnerable soils, delicate vegetation, and harsh cli mate (especially drought) make land vulnerable to desertification and it is human mismanagement which catalyses the process. Desertification is often described as a 'vicious spiral' – people degrade the land, struggle to maintain livelihoods, fail to invest in good husbandry, and cause further deterioration. However, there are situations where such a Malthusian decline is not the pattern. Once initiated there may be positive feedbacks, which reinforce the process and make recovery less easy. For example: soils lose organic matter and develop poorly permeable layers making moisture recharge less easy and increasing runoff; plant recolonization may be difficult once grazing or fire destroy sources of seeds; altered albedo (solar reflection) accelerates de vegetation and soil damage. There have been studies, which suggest desertification can lead to more arid regional conditions; and there may be wider climate impacts through airborne dust, and reduced carbon sequestration in damaged soils.

It may be difficult to distinguish desertification symp toms from causes, which can lead to misguided remedial action. Careful study should precede any attempt at management. The desertification process may operate so slowly that people fail to note it until too late. Or, sudden drought, fire, or locust attack makes it appear so suddenly that counteraction is difficult. In practice poverty and civil unrest may prevent mitigation, and governments may be unaware or unwilling to acknowledge it is underway. Often the process is unfairly blamed on unforeseeable natural disaster or careless peasantry ('poor locals').

Where the process is gradual (due to grazing, acid deposition, climate change, or whatever) it may only be apparent when a plot of land has been enclosed, also re mote sensing or aerial photography can reveal a sharp boundary between protected vegetated and damaged land where there is a fence or other divide. Desertification, can be an episodic or gradual process (or both). It can be a treacherous process: where land use is within the assumed carrying capacity and unexpected events still trigger desertification – a bushfire on adequately managed land at just the wrong moment may destroy seeds of annual plants, yet at any other time the fire would be unproblematic (a mistimed bout of overgrazing could have similar effect).

Causation

Desertification is often the result of a complex chain of causation, perhaps rooted in local land mismanagement or resulting from far distant policy making and governance. For example, abandonment of Roman grain production in North Africa was once attributed to reduced rainfall, in direct nonenvironmental causes such as the loss of cheap slave labor and free trade impacts are more likely. Factors, which alone or in combination, may trigger desertification, include: grazing (overgrazing); precipitation changes; altered intensity of land use – such as reduction of fal lowing; vegetation changes; salinization (salt/soda buildup in soils); pest activity; bushfire; and fuelwood collection. Desertification vulnerability is partly a function of the harshness of the environment and the ease with which vegetation cover and soil are damaged.

Desertification has increased since the 1950s and over this time many of the vulnerable lands have experienced marked population growth, mainly through in migration and improved health measures. Livestock herds have often multiplied in those areas (thanks to economic causes and improved veterinary care), over the same period there has often been breakdown of traditional drought coping strategies because of development. Some desertification has happened where there is lighter population which practices intensive ranching or cultivation which degrades the land in time of low rainfall – for example, Australia. There are fears that some modern developments, such as fertilizer use, are temporarily concealing ongoing desert ification (a desertification 'time bomb').

Regions vulnerable to desertification may tempt ex ploitation in wet periods; settlers are then reluctant or unable to adapt in drier times, so damage sets in. Another cause may be innovations, such as new crops or farming techniques, or tax changes and new market opportunities which can prompt unsustainable land use. Drought is fre quently seen as a trigger for desertification; for example, in Northeast Brazil and the sub Saharan Sahel and it can di vert attention from socioeconomic causation (see **Figure 2**). Desertification can be regarded as a 'litmus' (indicator) of land management that fails to adapt to environment.

Modern management is not good at coping with 'boom and bust' conditions; traditional practices may be well adapted, but have often broken down through developments such as population increase or spread of cash cropping or ranching. Causes may be complex, indirect, and cumulative, so each region should be carefully researched. National or international economic forces may trigger land degradation: structural adjustment may cause cutbacks in subsidies, or the market price for produce falls – so farmers resort to damaging land use. Poverty or unavailability of affordable alternative fuel can result in catastrophic removal of vege tation (and even livestock dung) causing desertification.

The Malthusian concept that population increase overstresses environment and resources and leads to impoverishment and degradation is sometimes wrong. While such a chain of causation may operate in some situations, population increase can also drive innovation to more intensive land use that is less damaging and able to feed more. Much depends on population growth not

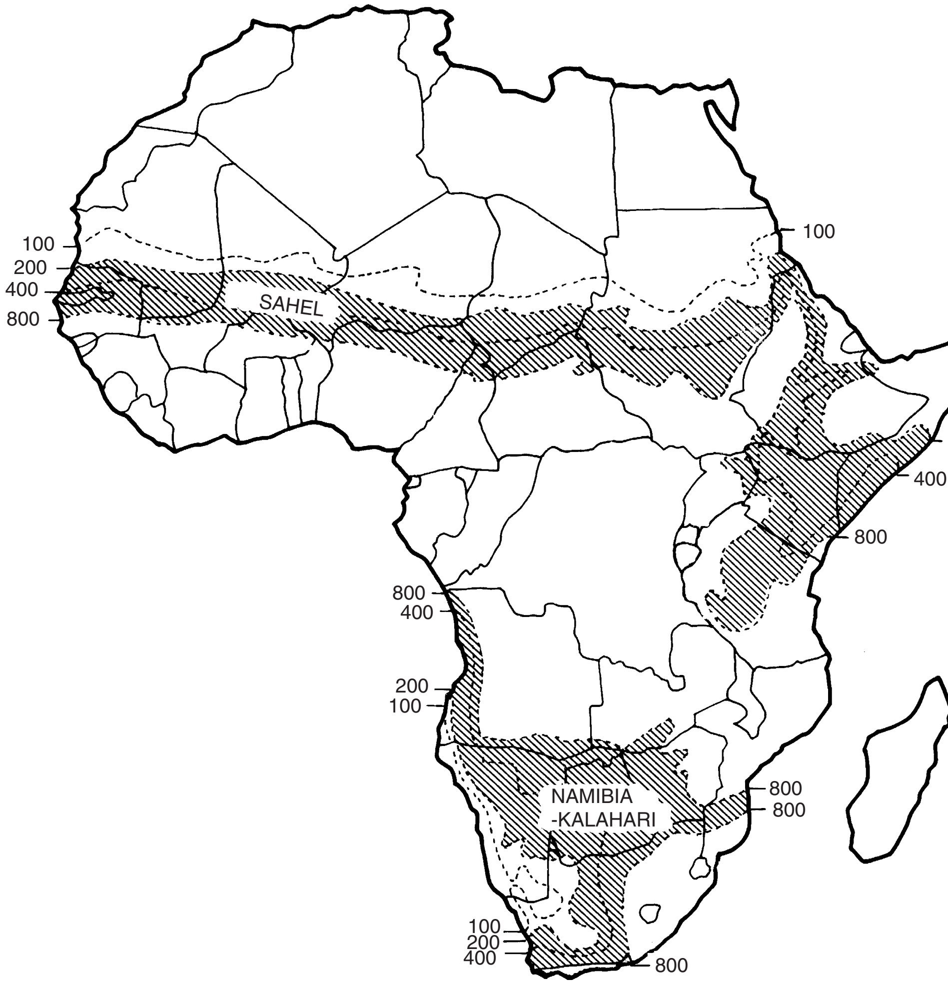

Figure 2 Africa area suffering or at risk of desertification. Shaded area lying between the 200 and 800 mm y [1] isohyets (90% probability) and the Sahel Zone south of the Sahara. Areas of Malagasy are at risk but have not been shaded.

being so rapid that it outstrips opportunities to inno vate and human willingness to change. Inadequate management of common land (not owned by individuals, but used by a number) has been seen as a cause of desertification; each user has an interest in maximizing their gains and without ownership or adequate controls is unlikely to act with restraint or invest in sustaining production. As a group the users of common land may cause damage if not controlled by laws or the forces of the market, and such forces may be lacking.

The in migration of people to desertification risk areas is frequently a process of marginalization caused by com petition for land or warfare elsewhere. However, sometimes migrants are arriving because they perceive opportunities.

Impacts

Ultimately, desertification means lost opportunities for food and commodity production and biodiversity conservation, and it is likely to mean those affected exist in less secure, as well as more impoverished conditions. Huge sums of money are wasted in countering desertification and in the associated opportunity costs (funds might have been use fully spent on other things).

Desertification also has an impact beyond the areas actually suffering it: degraded soils sequester less carbon helping to cause global warming; silt running from damaged land may damage rivers, irrigation schemes, and lakes hundreds of kilometers away; rivers draining af fected regions may become more prone to flash floods and periodic drying up. Desertification may cause dust storms, which sweep into surrounding regions to cause damage and disruption well outside the desertified area – currently large parts of China experience these in spring. Eco refugees fleeing desertification are a growing prob lem, and one already challenging a number of countries where they congregate. Destitute relocatees cannot easily reestablish adequate livelihoods and refugee camps may cause fresh land degradation as occupants scavenge for food and fuel wood. Some eco refugees move further afield seeking asylum, sometimes causing problems in distant countries. Loss of productive land makes food and commodity production less secure at the world scale.

Desertification destroys biodiversity and the plant and animal species lost may have great potential for crop breeding, pharmaceutical, or other uses.

Extent

Boundaries of desertified regions may not be distinct and can fluctuate a lot from year to year and season to season. Estimates of extent and severity were generally inaccurate before the 1990s, and still vary wildly, being exaggerated by some governments and agencies and downplayed by others. Recent assessments suggest that (globally) losses of cultivable land to desertification are around 12 million hectares each year (see **Figure 1**). The Global Assessment of Human Induced Soil Degradation (GLASOD) esti mates around 15% of the world's total land area was seriously affected by desertification in the early 1990s.

Very dry regions attract few people so suffer little desertification (deserts are already degraded), desert ification is most prevalent in regions with something like 600 mm average annual rainfall (average is not very meaningful where rains fluctuate a lot). Africa is widely held to be worst affected – in 2002 the United Nations Environment Programme (UNEP) claimed 45% of the continent was seriously desertified. While there was a trend toward desertification in the 1970s and 1980s in the Sahel (**Figure 2**), recent studies suggest marked wide spread improvement. Sahelian farmers have replanted acacia trees and are renewing productive land use over wide areas and rains seem to have returned. In northwest China desertification is now worsening, apparently through population increase, larger numbers of livestock, and herd improvements leading to greater fodder de mands. Northwest China and Inner Mongolia now have dust bowl conditions similar to that of 1930s USA.

Desertification can be widespread or a patchy local problem; for example, where livestock congregate and overgraze because a borehole has been provided there tends to be a teardrop shaped zone of damage facing into the prevailing wind; where a railway or road allows excessive exploitation of a region, there may be local 'hot spots'.

Monitoring and Predicting Desertification

Caution is needed:

- to ensure local scale desertification ('hot spots') is not misinterpreted as indicative of widespread degradation;
- to check if the problem is short term or long term;
- to establish whether the threat is decreasing or growing;
- to determine if the problem could in practice be remedied;
- to provide adequate and timely warning if things are getting worse;
- to conduct wherever possible, more than one type of test to confirm trends;
- to monitor causation in case it alters;
- measurement may give a 'snap shot' view, limited spatially and temporally. Real time continuous moni toring is preferable; and
- remote sensing must be backed up with adequate ground truth studies.

Empirical formulas have been developed which give some prediction of soil degradation for a given location and land use; the most widely used being the Universal Soil Loss Equation (USLE) and its adaptations for use in nontemperate environments, such as the Revised Uni versal Soil Loss Equation (RUSLE). Computer or analog simulation models may give an indication of how vul nerable a region is to various degrading processes. Re mote sensing and on the ground vigilance can also help warn of desertification. However, it is vital to have widely agreed reliable and easily measured indicators that can be monitored. Not all of these are physical, there may be useful socioeconomic indicators that show when people are being disrupted by desertification: sale patterns of livestock, pawning of jewelry, out migration, and so on. In addition, social studies may reveal when social capital declines and may make people more vulnerable to desertification. In the past monitoring vegetation change has often given misleading results because it has not been carefully conducted. It is important to establish, map, and monitor clear categories of degree of desertification from slight to severe. Uniform desertification is unlikely; pat ches of severe and irreversible damage likely lie in areas of less severely damaged land.

Some parts of the world are subject to variable pre cipitation, which may reflect quasi periodic phenomenon like the El Niño Southern Oscillation (ENSO). If this is the case, a developing ENSO event may enable fore casting of drought in desertification prone regions more than 12 months ahead. Locust swarm prediction and control are well developed, but in recent years have been neglected through warfare and lack of funding.

Combating Desertification and Rehabilitating Desertified Land

Between 1951 and 1962 the United Nations Educational, Scientific and Cultural Organization (UNESCO) under took a Major Project on Scientific Research on Arid Lands which helped stimulate interest in desertification. The 1968–73 drought in the Sahel caused the death of as many as 250 000 people and prompted the UN Conference on

Desertification (UNCOD) in 1977. UNCOD published a (global) Plan of Action to Combat desertification (PACD) and called upon each affected country to implement a national PACD in the hope that within 20 years the problem would be solved. Unfortunately, UNCOD was not very successful. The 1992 UN Conference on Environment and Development at Rio established an Intergovernmental Negotiating Committee on Desertification, augmented with a UN Consultative Group on Desertification Control (DESCON), an Intergovernmental Working Group on Desertification (IWGD), and a UNEP Desertification Board. Many nations signed a 1994 Treaty on Desertification and a UN Convention to Combat Desertification (UNCCD) and the UN declared 2006 International Year of Deserts and Desertification. However, overall success has been limited.

Rehabilitation may be possible if there are sufficient funds, suitable techniques, and people are supportive. Locals need to be motivated or desertification control measures will be superficial, neglected, or even damaged. What appears to an outsider to be a good approach may not be attractive to locals – for example, some terracing projects take up scarce land and hinder traditional tillage and so will be resisted. Acceptable means must be established to protect soil from erosion by wind or water – using windbreaks, stone lines, terraces, check dams, surface sealant sprays, as appropriate. Ways of improving soil organic matter content are also desirable, and it may be necessary to restrict grazing and control pests. China is currently planting a 5700 km long Green Wall, a belt of trees and shrubs to combat the desertification in its northeast. Revegetation hopefully leads to better rainfall infiltration, more soil moisture, and improved soil organic carbon. It requires costly inputs of fertilizer, irrigation, and labor to get such improvements. Countering desertification is as much about motivating people as developing techniques and providing the material inputs. There are examples of land users who cooperated as long as they were paid or fed, but then abandoned maintenance. Poor people in harsh, fragile, often remote, desertified environments are unlikely to be able to do much to improve things without aid. Also, they must be consulted to see what they need and will support. Sometimes 'anti-desertification' measures are inappropriate and do little to help local people or counter desertification; for example, export crop irrigation schemes profiting private business or government.

Where people lack security because there is unrest, warfare, or an inadequate legal right to land they are unlikely to invest in sustaining land use or remedying degradation. They reason that someone other than themselves or their relatives would benefit from any investment. Where there is more than one group using the same land and they fail to cooperate desertification can arise – for example, where nomadic herders or wild game graze unfenced lands left fallow to recover or which have been otherwise improved by sedentary agriculturalists. Land reform and enclosure may be necessary before attempting remediation.

Grazing controls, livestock reduction, alternative feeding strategies (such as stall feeding), or rangeland improvement may be helpful. China has recently controlled goats in Mongolia and western China because flocks have expanded to supply the fashion for cashmere and are damaging the land. Where fuel collection is a problem alternative energy may offer a solution, or the establishment of plantations or hedgerows may supply needs. Where inappropriate tillage is to blame, soil and water conservation techniques, irrigation, and other innovations may help.

See also: Climate Change; Deforestation; Intermediate Technology; Rio Summit; Sustainable Development; Vulnerability.

Further Reading

Aubreville, A. (1949). *Climats, forêts et désertification de l'Afrique tropicale*. Paris: Société d'éditions géographiques, et maritimes.

Batterbury, S. P. J. and Warren, A. (2001). Desertification (US Geological Survey). http://pubs.usgs.gov/gip/deserts/desertification (accessed on 25 August 2007).

Fairhead, J. and Leach, M. (1996). *Misreading the African Landscape: Society and Ecology in a Forest–Savanna Mosaic*. Cambridge: Cambridge University Press.

Geist, H. (2005). *The Causes and Progression of Desertification*. Abingdon: Ashgate.

Glantz, M. H. (ed.) (1994). *Drought Follows the Plough*. Cambridge: Cambridge University Press.

Goudie, A. S. (1990). *Techniques for Desert Reclamation*. Chichester: Wiley.

Grainger, A. (1990). *The Threatening Desert: Controlling Desertification*. London: Earthscan.

Mainguet, M. (1994). *Desertification: Natural Background and Human Mismanagement* (2nd edn.). Berlin: Springer Verlag.

Middleton, N. and Thomas, D. (eds.) (1997). *World Atlas of Desertification* (2nd edn.), (1st edn. 1992). London Arnold/Nairobi: UNEP.

Oldman, L. R., Hakkeling, R. T. A. and Soembroek, W. G. (1991). *World Map of the Status of Human Induced Soil Degradation* (revised edn.), (1st edn. 1990). Wageningen: ISRIC/Nairobi: UNEP.

Stebbing, E. P. (1938). The advance of the desert. *The Geographical Journal* 91, 356–359.

Thomas, D. S. G. and Middleton, N. J. (1994). *Desertification: Exploding the Myth*. Chichester: Wiley.

Relevant Websites

http://www.fao.org
Desertification, Food and Agriculture Organization of the United Nations.

http://www.fao.org
FAO/AGL National Soil Degradation Maps.

http://www.unep.org
Global Deserts Outlook, United Nations Environment Programme.

http://www.unccd.int
United Nations, Convention to Combat Desertification.

Determinism/Environmental Determinism

H. Ernste, Radboud University Nijmegen, Nijmegan, The Netherlands
C. Philo, University of Glasgow, Glasgow, UK

Glossary

Determinism A way of thinking that assumes one 'thing' (situation) is determined by logically and chronologically prior causes, an approach to explanation with deep theological roots but one that has become integral to most conventional forms of scientific inquiry.

Environmental Determinism A way of thinking common in academic geography, if now somewhat discredited, that supposes human existence and society, arguably including everything from settlement to language, to be determined by prior and external natural environmental conditions.

Possibilism A way of thinking in academic geography that 'softens' the outlines of a strict environmental determinism, supposing instead that the natural environment throws up possibilities for shaping human existence and society (not iron determinations).

Probabilism A claimed 'third way' between environmental determinism and possibilism, one stressing that the natural environment makes certain human outcomes more probable than others.

Voluntarism A way of thinking about human existence and society that supposes that human beings to have some independent capacity for influencing their own worlds, rather than being fully determined by controls from without.

Introduction

Etymologically, 'determinism' stems from Latin *determinare*, meaning 'to specify', 'to determine'. Determinism as a way of thinking has a very deep ancestry, certainly in Western thought, not least through entanglement with theological arguments about the primacy of an all powerful, hence all determining deity, in the face of whose 'commandments' it becomes an issue as to whether the 'mere' things of creation, humans included, possess any capacity for being and acting in a nondetermined way. From the ancient scholars, and then continually recast over centuries in the natural sciences, it has also become an epistemological strategy at the heart of conventional approaches to scientific explanation. Here it is assumed that one situation or configuration of phenomena, when subjected to knowable changes such as the application of particular forces, will automatically and by necessity lead to another situation, with the phenomena now differently configured. This assumption is the basis of cause and effect thinking, positing that the phenomena of the world consistently 'obey' principles or laws of causal determinism that the researcher believes either to have been proven or potentially to be provable. As a corollary, a deterministic epistemology also tends to assume the ready identification, delimitation, measurability, and computability of the variables comprising the situations in question. This determinism inevitably stands in contrast to a statistical or probabilistic position that allows chance elements to play upon outcomes, a stance that accepts a measure of nondetermination in the 'systems' under study, although for some scientists the chance element is conceived less as chance *per se* and more what 'we' do not as yet understand. It is telling to realize that in physics itself, with shifts from classical or Newtonian mechanics to relativity theory, quantum physics, and accepting 'uncertainty' in the calibration and prediction of certain phenomena, there has now been some retreat from, or reconceptualizing of, determinism. Hence, this article concludes with an addendum briefly reviewing this changing sense of determinism in physics.

Unsurprisingly, the problem of determinism has arisen most sharply for geographers at the interface, where they closely relate to the natural sciences of the external environment, at which point the issue becomes the extent to which human beings are determined in their constitution, organization, and behavior by the (kinds of) causal principles reckoned to govern all environmental phenomena. At various points in the history of geographical inquiry, efforts have been made to specify deterministic relations controlling whatever might be regarded as 'geographical phenomena', and a recurring theme ever since the labors of the ancient geographers have been in dispute over whether or not it is appropriate to conceive of overarching 'geographical laws' supposedly governing the disposition and character of these phenomena. More specifically though, when geographers think about 'determinism', it is usually in the context of debating the merits and demerits of 'environmental determinism', the name given to a form of geographical study emphasizing how the natural environment (in the shape of climates, landforms, rivers, forests, etc.) influences, controls, indeed determines, how humans organize their lives, potentially including everything from their agricultural practices to their religious beliefs. This approach to geography, called by some 'the geographical philosophy', was dominant in the late nineteenth and early twentieth century, coincident with the institutionalization of geography as an academic discipline in European universities. It does have deep historical roots, though, with many branches and variations spreading through into the twentieth century, and arguably feeding through into more recent attempts as forging a new 'environmental geography' that (re)centralizes human–environment relations. This article outlines such roots and branches.

Human geography has been further impacted by deterministic reasoning, however, in that, by analogy with the natural sciences, the social sciences, including 'human' geography, often quest for a determinism describing necessary and sufficient causal relations between different social situations or events. The core question at stake becomes how far human behavior is determined in the same manner as is seemingly true of 'matter', dead or alive, and how much 'space' is leftover for human free will, voluntary choices, and independent agency. In this respect, it is possible to distinguish between different interpretations of determinism. On the one hand, there are the 'incompatibilists' who view determinism and free will as mutually exclusive, and they can be further subdivided into those who reject determinism but accept free will, called 'libertarians', and those who believe that free will is an illusion, called 'hard determinists'. These different opinions reflect different interpretation of free will: sometimes free will is assumed to be a metaphysical truth, while sometimes it is assumed to be only a 'feeling of agency' that humans experience when they act. On the other hand, there are the 'compatibilists' or 'soft determinists' who believe that the two ideas can be coherently reconciled.

However framed, though, this social scientific perspective on determinism focuses directly on the problem of free will, with determinism at one extreme of the continuum, and 'voluntarism' at the other. The latter term, introduced by Ferdinand Tönnies, the nineteenth century German sociologist, describes the assumption that it is, in the first place, the workings of free will that determine every theoretical and practical human action. A prime representative of voluntarism before Tönnies was the German philosopher Arthur Schopenhauer, who compared the will with Immanuel Kant's *Ding an sich* ('thing in itself'). It is the will that causes things to appear to the human individual as they do, seemingly dictating the person's course of action, and as such will is identical to, in effect a version of, the 'things in themselves' (the *noumena* of the world, the things in the world, logically prior to the 'phenomena', mere appearances of things as perceived by the human). Will is hence a necessity, and exactly identical to what we otherwise call causality, and yet in human consciousness 'we' (humans) can also know and reflect upon, and even resist, our own will, thereby creating our 'own' determinancy. By this maneuver, the distinctions between determinism and voluntarism arguably break down, and yet, in practice, social scientists, including human geographers, continue to grapple with the tensions between what appear to be 'determinations' pressing upon human beings from without – pressures, demands, exhortations, etc., impelling them to act in a certain (predetermined) way – and what seem to be instances of humans acting otherwise, resisting external determinations, responding instead to inner impulses, convictions, and deliberations. In the long run of geographical inquiry, the external determinations have largely been regarded as environmental, and their degree of control debated, but other, broadly equivalent determinations have been posited as well. The issues raised in the last two paragraphs stand in the background of what is discussed in this article, and is briefly revisited more explicitly in the final section.

(Environmental) Determinism in Geography: Roots and Branches

Thinking along the lines of environmental determinism can easily be traced to the ancient world; it is said that Strabo, one of the earliest published 'geographers', was 'fond' of showing how the physical features of a land clearly influenced the 'character and history' of its human inhabitants. In his epic work from 1967, *Traces on the Rhodian Shore*, Clarence Glacken charts the many twists and turns of environmental thinking from Ancient Greece through to the likes of the sixteenth century French writer Jean Bodin, for whom "religion, culture, and morality were all subject to environmental control and therefore in some degree relative to the particular geographical circumstances in which they developed" (Livingston, 1992: 97).

It is nonetheless in the writings of those celebrated (European) Enlightenment and the early nineteenth century German 'geographers' Alexander von Humboldt (1769–1856) and Carl Ritter (1779–1859), notably the latter in the introduction to his *Die Erdkunde*, that the outlines of an intellectual environmental determinism began to be sketched out. Theirs was a complex, holistic science predicated on seeing in Nature – meaning 'all' Nature, the whole panorama of creation – an essential unity, whereby 'inorganic' phenomena (rocks, soil, and water) were causally linked to 'organic' phenomena (plant, animals, and human beings). For instance, Humboldt observed that:

> My attention will ever be directed to observing the harmony among the forces of nature, to remarking the influence exerted by inanimate creation upon the animal and vegetable kingdom. … [T]he geography of organic beings of plants and animals is connected with the

> delineation of the inorganic phenomena of our terrestrial globe. (Humboldt, 1951: 53 54)

It is vital to appreciate that this was primarily a vision of the influences running from the inorganic to the organic, a more expansive vision that one is only concerned with environmental controls on human beings. Intriguingly, it was one echoed by the great American geomorphologist William Morris Davis (1850–1934), who spoke of geography as the 'study of relations' whereby the inorganic shapes the organic: "any statement is of geographical quality" if it elaborates "some relation between an element of inorganic control and one of organic response" (in Johnson, 1909: 10).

This being said, the ingredients of a cruder environmental determinism, basically of positing stark environmental controls on human existence and action, still surfaced in the writings of Humboldt and Ritter. Again using Humboldt as an example, he observed that:

> Dependent, although in a lesser degree that plants and animals, on the soil and on the meterological processes of the atmosphere with which he is surrounded escaping more readily from the control of natural forces by activity of mind and the advances of intellectual cultivation … man [*sic*] everywhere becomes most essentially associated with terrestrial life. (Humboldt (in Tatham), 1951: 54)

Despite possessing qualities and powers different from plants and animals, then, Humboldt did envisage human beings as shaped in large measure by the properties of 'terrestrial life', inorganic, 'and' organic, which hence served to propel humans to the fore as the 'objects' to be explained (seen as determined not just by the inorganic world but also by its nonhuman organic elements). Arguably, this was a narrowing of the lens, producing what might be termed a cruder form of environmental determinism that began to hold sway over much of the nineteenth and early twentieth century thinking (in the discipline of geography and beyond).

This blunter form of environmental determinism affected (one might say afflicted) many writers, but an immediate sense of what was involved can be borrowed from Immanuel Kant, the German philosopher of the late eighteenth and early nineteenth centuries. A genius, pondering over such tricky subject matters as 'pure reason', 'sense perceptions', and the nature of time and space, he also gave lectures on geography which could include the following:

> Inhabitants on the coast of New Holland have half closed eyes and cannot see to any distance without bending their heads back until they touch their backs. This is due to the innumerable flies which are always flying into their eyes. All inhabitants of hot lands are exceptionally lazy; they are also timid, and the same two traits characterize also folk living in the far north. … Laps, Greenlanders & c. resemble people of hot lands in their timidity, laziness, superstition and desire for strong drink, but lack the jealousy characteristic of the latter since their climate does not stimiate their passions so greatly. (Kant (in Tatham), 1951: 130)

Extremely simplistic cause and effect determinisms were thus said to run between features of the natural environment, such as hot or cold climates, as associated with quite specific attributes, such as the presence of 'innumerable flies', and dimensions of human life within the environments so depicted. As is clear from the above quote, these determinisms readily teetered on the brink of deeply troubling negative stereotyping of certain peoples in certain parts of the world: in a word, the specters of racism and even a white supremicism were never far behind. In order to give a further taste of this really quite 'crass' environmental determinism, four different exponents can briefly be mentioned.

A rare woman academic geographer from this period, and someone who had studied in Germany with Ratzel (see below), Ellen Churchill Semple (1863–1932) became known as a 'hard core' determinist who exerted an enormous influence in the American and British human geography. At the outset of her book from 1911, *Influences of Geographic Environment*, she made her position unmistakably clear:

> Man [*sic*] is a product of the earth's surface. This means not merely that he is a child of the earth, dust of her dust; but that the earth has mothered him, fed him, set him tasks, directed his thoughts, confronted him with difficulties that have strengthened his body and sharpened his wits, given him his problems of navigation or irrigation, and at the same time whispered hints for their solution. She has entered into his bone and tissue, into his mind and soul. (Semple, 1911: 1)

Another American geographer, Ellsworth Huntington (1876–1947), had a similar impact with his theses regarding the dependency of civilization on climate. He traveled widely, and in his 1907 book *The Pulse of Asia* – revealingly subtitled 'The geographic basis of history', and dedicated to W. M. Davis – he studded a record of his journeying in Central Asia with a sustained environmental determinist refrain, concluding that 'geography,', as in physical–environmental conditions and changes, "especially through its influence upon [human] character, is the basis of history in a way that is not generally recognized" (Huntington, 1907: 402).

Even some years into the twentieth century, similar tones could be heard. Thomas Griffith Taylor (1880–1963), a British geographer who spent much of his career in Australia and Canada, developed "theories of determinism easier to understand but less subtle than those of earlier writers" (Freeman, 1961: 325). He referred to his approach as 'geocratic' rather than 'theocratic', thereby positioning geography (or the natural environment) as the driving force behind human existence in a manner analogous to how religious writers position God. In another guise, he produced maps of 'future' white settlement across the globe, identifying regions where climatic conditions were suitable for white colonization (irrespective of who might actually be living there presently). A less (in)famous geographer advocating a strict environmental determinism, meanwhile, was another American, Roderick Peattie (1891–1955), whose 1940 text *Geography in Human Destiny* assertively restated an 'environmentalist' perspective in the discipline – he called it a 'geographic philosophy' – and pictured human action as constrained along three axes (**Figure 1**) the environmental (*y*), the genetic (*x*), and human choice (*z*). Peattie then delighted in locating his explanation for just about everything in the human realm as near to the *y* axis, and as distant from the *x* and *z* axes, as he could. He also talked about what he nicely (and one might say appropriately!) called 'catastrophic geography', discussing events such as earthquakes and hurricanes, and demonstrating the clear 'geographic determinism' that would result for the humans caught up in such disasters (see **Figure 2**).

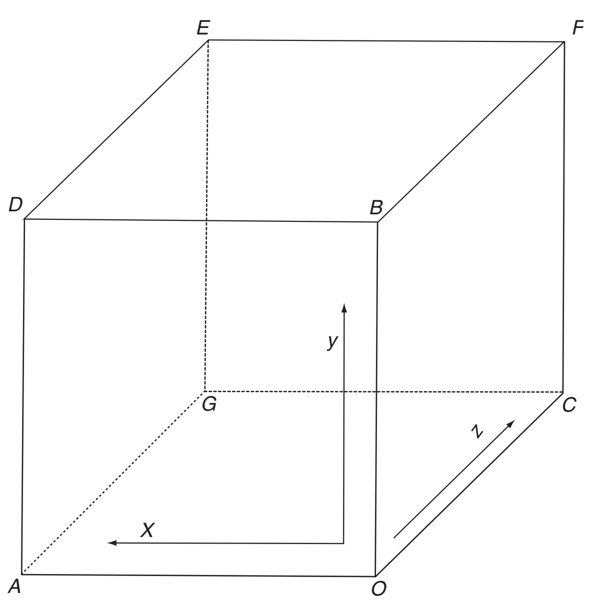

Figure 1 "If *x* is increasing control by inheritance, *y* the environmental factor in one's life, and *z* the increase of human choice in any event, any fact relating to man's [*sic*] action may be plotted within the cube." From Peattie, R. (1940). *Geography in Human Destiny*. Boston, MA: George W. Stewart.

Science, Evolutionary Theory, and an Environmentally Determinist Human Geography

Weaving through the developments noted above, in part fueling them but on occasion checking their excesses, was the effort, from the late nineteenth century, to create a more convincingly 'scientific' version of academic geography. At this time, it can be claimed, science became the new 'religion', and the enormous successes of the natural sciences rendered it a model for all fields of inquiry striving to be taken seriously as 'scientific'. At one level, this meant that there was indeed an imperative to be deterministic, to be able to excavate causal relations and underlying deterministic principles, as (supposedly) was both the goal and the accomplishment of all 'proper' sciences (one's cleaving to a Comtean positivism). At another level, this meant attempts to translate ideas, models, and methods from 'successful' natural scientific fields into as yet less 'successful' fields, in which regard the work of Charles Darwin, the brilliant natural scientist, natural historian, and sometime 'biogeographer', found such a huge resonance. With evolutionary theory, as codified in Darwin's 1859 text *The Origin of the Species by Means of Natural Selection*, for the first time it seemed feasible to explain all observable 'facts' about the living world, including human thinking and human behavior, without having to resort to some supranatural power. Human being was not the center of the world, nor the crown on God's creation anymore, but rather the product of random effects or of sufficient causes, and the natural environment was seen as the forcing house of this process of evolutionary development. Only the environmentally best adapted organisms within a species would survive, and it was this struggle to adapt, and hence to survive, that was taken to explain the different features of life as distributed across both the many species and ecological niches spread across the planet. Putting aside the many complexities about Darwinism, including its misrepresentation in many quarters, it is entirely unsurprising that many emerging disciplines – including the likes of sociology and philosophy, with Herbert Spencer's attempts to foster what was to become known as a 'social Darwinism' – sought to acquire, and to secure legitimacy from, a Darwinian cast.

Given its substantive interest in human–nature relations, academic geography was more plausibly set than some other disciplines to receive, and to rework, Darwinian materials. Environmental determinism was duly given a productive shot in the arm, a tool in its own armory, although a more systematic and rigorous borrowing from Darwin might have enabled the discipline to 'avoid' some of the worst excesses detailed above (not least by recognizing, with Darwin, the complicated 'two way' relationship between organism and environment). A

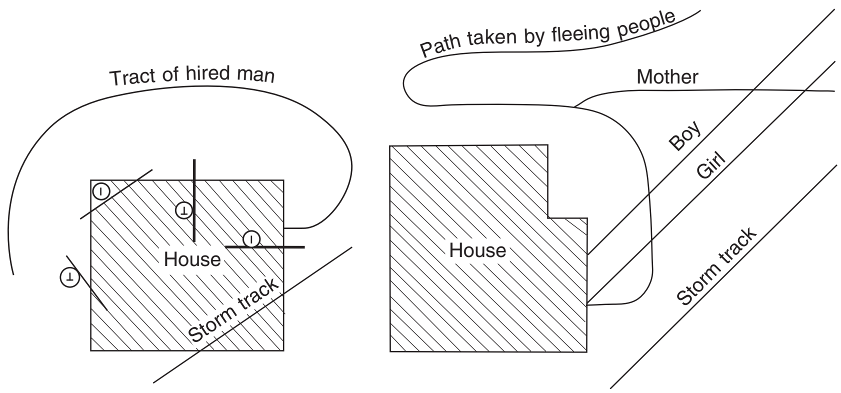

Figure 2 "Flight diagram of victims of geographic determinism during two tornadoes, showing actual paths through the air of the victims of the storms." From Peattie, R. (1940). *Geography in Human Destiny*. Boston, MA: George W. Stewart.

name often mentioned as a translator of Darwin into geography is that of the German scholar Friedrich Ratzel (1844–1904), who trained as a zoologist and early started to recognize the challenge of Darwin's evolutionary theory. The historian of the discipline George Tatham thus suggests that:

> Ratzel saw man [*sic*] as the end product of evolution, an evolution in which the mainspring was the natural se lection of types according to their capacity to adjust themselves to the physical environment. … Ratzel ten ded to see man as the product of his environment, moulded by the physical forces that surrounded him, and succeeding only insofar as he made the correct adjust ment to their demands …. (Tatham, 1951: 64)

This being said, Ratzel was conscious that a simple environmental determinism, which derives the contours of human life 'directly' from the compulsions of physical–environmental spaces, would be inadequate in itself. In stead, following the lead set by Moritz Wagner's 1868 text *Die Darwinsche Theorie und das Migrationsgesetz der Orga nismen*, he proposed a less direct relationship. In this view, 'natural space' does not directly produce human com munities and their cultures, but rather produces a natural setting for migratory and cultural movements across space, which then will lead to a spatially differentiated distri bution of human activities. Ratzel applied this thinking, mainly in the field of political geography, where he tried to explain the emergence of states as an evolutionary ne cessity. In analogy with biology, he interpreted the state as an organism which undergoes (population) growth until the locally available resources are exhausted, at which point territorial expansion becomes a necessity for the state to survive. States need *Lebensraum* (living space or 'room') to exist, which would explain their 'natural' ten dency toward territorial expansion, a notion that of course could be, and likely was, mobilized to dubious ends. It is nonetheless possible to observe a discrepancy between the strong theoretical propositions that were made by Ratzel and the careful, indeed quite nuanced, conclusions that he drew from them himself. In this context, his followers seemed to be much more dogmatic than the master himself, contributing, as with the cruder claims of Semple *et al.* as retold above, to the disrepute which began to befall environmental determinism in the second half of the twentieth century.

Softening Environmental Determinism

Human geography, in this tradition, tried to show that the natural environment was the causal determinant of not only the spatial distribution but also the entire 'nature' of human endeavor: "Places vary in their natural conditions and these variations determine what people do there," as Ron Johnston summed up this tradition. Put like this, environmental determinism appears as a very plausible, legitimate, and indeed incontrovertible enterprise, and it might be argued that in the hands of most geographers and others (e.g., human ecologists) who have worked within an environmentalist orbit – laboring hard, if often quite an onymously, unlike the standout but controversial figures mentioned above (i.e., Semple *et al.*) – a 'softened' version of environmental determinism was what ended up actually practiced. Indeed, what was sustained of determinism in human geography was mainly a free and 'mild' form of determinism, more a metaphor and an *ad hoc* hypothesis than a rigorous principle, as Gerhard Hard proposed in his 1973 text *Die Geographie. Eine wissenschaftstheoretische Ein führung*. In this rather diffuse way environmental deter minism lived on in landscape geography and regional geography as well as in many school textbooks.

To elaborate, in continental Europe, determinism early attracted serious criticisms, in both its general and its specific (i.e., environmental) forms. In Germany, the philosophy of idealism, with its convictions about both the freedom of human will and the historicity of personality, went through a revival, and evidently exerted an influence upon the country's geographers. Alfred Hettner (1859–1941), for instance, proposed that, instead of talking about necessary and certain (deterministic) influences of nature on humanity, a more sophisticated conception might be one stressing how nature merely restricts the number of opportunities available for how these lives might be lived. Modifying Kant's concept of chorography for the description of all phenomena that are found in the same place, Hettner coined the term 'chorology' as the scientific explanation of 'all' causal relationships between these phenomena. These causal relationships, in his view, were not unidirectional or hierarchical, and so nature was not seen as the sole de termining or causal power in the making of 'real' human geographies on the ground. The further implication was that people dwelling in any one environmental context should be conceived as having a substantial range of options from which to choose, when seeking to survive and maybe flourish in that context, suggesting that the human beings here possess causal powers.

In France, Bergsonian thinking, after Henri Bergson, gained influence with its focus on the unpredictable *élan vital* and the unforeseeable creative activity of human beings. Paul Vidal de la Blache (1845–1918) probably picked up on this current, and his posthumously pub lished 1926 text *Principles of Human Geography* duly identified the variety of possible 'local solutions' which people could devise to the 'problem of existence' posed to them by their host natural environments. His approach came to be known as 'possibilism', perhaps most obvi ously in the work of the French historian Lucien Febvre. As summarized in the 'Foreword' to Febvre's 1925 text *A Geographical Introduction to History*:

> Against the geographical determinism of Ratzel, [Febvre] sets the possibilism of Vidal de la Blache. 'There is no rigid and uniform influence of four or five great geo graphical fatalities weighing on historical individualities'. The true and only geographical problem is that of the utilization of possibilities. 'There are no necessities, but everywhere possibilities. (Barr, in Febvre, 1925: xi).

According to both de la Blache and Febvre, it was unreasonable to draw boundaries between natural and cultural phenomena; rather, they should be regarded as united and inseparable. Each human community adjusts to prevailing natural conditions in its own way, and the result of the adjustment is not one sidedly due to the natural conditions but the result of a long process of mutual adaptation.

In practice, then, the terms of a debate were estab lished between determinism and possibilism as perhaps the most deep seated of controversies running through the early to mid twentieth century academic geography. The landmark edited collection from 1951, *Geography in the Twentieth Century*, was framed by this debate, and in the introduction it was stated that "the authors will take opposite sides in certain problems, as for instance in the broad question of Possibilism versus Determinism" (Taylor, 1951: 3). Moreover, in this introduction, Taylor contrasted the deterministic line taken by both a 'theo cratic' position and his own 'geocratic' position (see above) with the so called 'we ocratic' position of the possibilists, one centralizing the role played by 'we' humans in shaping our own worlds and destinies. He also expended much energy here and elsewhere in the vol ume to disputing possibilism, and to propounding what he called a 'scientific' or 'stop and go' determinism, the basic thrust of which is that the natural environment in a given region charts out a 'main road' along which hu manity will inevitably develop in that region, but that the people involved will be like 'police officers' controlling the 'traffic' along this road: that is, allowing it through, maybe accelerating the overall movement, or stopping it, decelerating the movement.

In practice, however, enthusiasm for the strict en vironmental determinism favored by Taylor was definitely on the wane by the 1950s, and at the same time the de terminism–possibilism debate was itself fading from view, at least in the terms detailed above. Insofar as geographers through the later years of the twentieth century and up to the present have considered where they stand on the de terminism–possibilism axis, they would likely admit to following the softer, modified stance on environmental determinism indexed by the possibilist positions. One further contribution of note occurred in contributions from 1952 and 1957 when O. H. K. Spate (1911–2000) proposed a middle way between environmental determinism and possibilism, something that he called 'probabilism', em phasizing that the natural environment makes certain 'possibilities' (of how humans in a given environment might develop through time) more 'probable' than others. More recently, John E. Chappell has sought to recast the debate in terms of 'environmental causation', seeking thereby to avoid the excesses of an older environmental determinism, while various geographers – a prominent example is David Stoddart – have sought to reclaim a stronger sense of the natural–environmental dimension central to almost all facets of human existence, not least in a contemporary era of dramatic global climatic–environmental change (posing all manner of intractable questions for not just human flourishing but also maybe the very survival of humanity on this planet). The name of 'environmental geography' has sometimes been attached to this latter development, not least in the context of undergraduate textbooks and teaching, but there is potential for such a rubric to en compass a variety of conceptualizations and methodologies, older and newer, around the axes of humanity and en vironment, or nature and culture.

Determinism, Spatial Science, and Agency-Structure

As noted, the determinism–possibilism debate, as conducted around issues of 'environmental' influence, waned after c. 1950, not least because a very new set of concerns began, quite dramatically, to reconfigure all aspects of the discipline in Anglo Saxon speaking lands (and then spreading globally). These concerns arose with the advent of geography as spatial science, a quantitatively literate attempt to forge a properly 'scientific' geography able to specify foundational spatial (or morphological) laws of how phenomena (physical or human) are distributed across the Earth's surface. Other articles in this encyclopedia provide more detail on this far reaching sea change in geographical thought and application, but of relevance here is how the search for spatial laws, commonly conceived through a lens of cause and effect, clearly committed the discipline to a fresh era of determinism. Revealingly, in one of the founding statements of spatial science, Ian Burton's 1963 paper on the 'quantitative revolution', environmental determinism is treated favorably, not for its specific claims, which Burton basically rejects, but for the 'ambition' of identifying deterministic, causal, law like relations when examining geographical phenomena. There has been sustained argument over the validity of seeking to convert geography, particularly 'human' geography, into a 'science' akin to other conventional sciences, complete with a strong deterministic cast. Such arguments have also continued to rage over other versions of human geography that have emerged in opposition to spatial science, notably the more 'economistic' versions of Marxist geography, wherein deterministic laws of capitalist accumulation are advanced in the explanation of sociospatial inequalities. Realist, post structuralist, and postmodernist geographies have all questioned conventional deterministic reasoning, in whatever exact guise it has been prosecuted (environmental, spatial, or economistic), albeit mounting their criticisms from rather different bases, with differing levels of hostility, but not necessarily then endorsing voluntarism. Other articles in the encyclopedia ably cover the many and varied problematics involved, and so little more need be said here.

What can be noted is that several commentators on disciplinary history detect the same parallels as just spelled out. Indeed, the likes of Derek Gregory posit an essential continuity between the environmental determinism of Semple *et al.*, the 'geometric determinism' of the spatial scientists and the 'economic determinism' of certain Marxist geographers, particularly in how they all end up allowing only scant room for human beings to be and do, to think and act, in anything other than a fashion determined for them by external compulsions. Thus, for the spatial scientists, spatial forms (e.g., a 'central place' hierarchy of settlements) are reckoned to be just as controlling of human behavior as are natural–environmental contexts (e.g., a low lying, ill drained valley) for an environmental determinist. Or, for certain Marxist geographers, capital logics (e.g., the geographical transfer of 'surplus value') are reckoned to be equivalently controlling of human behavior. Unsurprisingly, alternative perspectives have been advocated by human geographers unhappy about the implied negation of human freedom, free will, or, in a more recent vocabulary, agency. The result has been, as hinted earlier, the appearance within human geography of voluntarist conceptions standing in opposition to, or at least endeavoring to qualify the excesses of extreme, determinist thinking.

Behavioral and humanistic geographers made the initial moves, demanding that people 'as' people, not as objects akin to rocks or rivers, nor as unthinking automata responding soundlessly to the determinations of environment, space, or economy, be placed center stage in inquiry. In turn, such moves were themselves criticized for being too voluntaristic, and for thereby positioning human beings as simply 'too' powerful, too able to shape their own destinies, when a host of constraining, if not strictly determining, influences surely did need to be taken more seriously. Here, therefore, was the fraught terrain of debate over what increasingly became known as 'agency' and 'structure', where agency equated with 'human' agency (the capacity for experiencing, decision making, acting, etc., of human beings) and structure with 'social' structure (potentially encompassing all manner of broader 'ecological', economic, and political imperatives comprising a structural 'level' above, below, or behind (the backs of) human being individually and collectively). A diversity of theoretical maneuvers were advanced in order to overcome this so called dualism, most obviously drawing from the 'structurationist' social theory of Anthony Giddens, an eminent sociologist, but commonly reworked to incorporate an enhanced alertness to how the 'resources' of time and space unavoidably enter into the recursive mediations of agency and structure (whereby agency 'makes' structure 'makes' agency, etc.). Once again, it is enough to reference other articles in this encyclopedia that chart in more detail the contours of the agency–structure debate, and how it was supposedly resolved in structuration theory and structurationist geographies. Nonetheless, the implications for thinking about determinism and voluntarism, a theme inevitably swimming throughout 'all' versions of human geography, past and present, whether or not named as such, should be obvious. In practice, perhaps akin to how most human geographers now effectively subscribe to an environmental possibilist (or maybe probabilist) line, even if rarely expressing their preference thus, it can be argued that the structurationist line on the balance between agency and structure, or between determinism and voluntarism, has now become a kind of 'unspoken' ground zero for many of us.

Determinism in Physics: An Addendum

In classical mechanics, the situation of a system S of n mass points at time point t_0 is fully known if, for each of the mass points i with $1 \leq i \leq n$, their locations (or co ordinates) $x_i(t_0)$ and their impulses (or momentums) $p_i(t_0)$ at time t_0 are known. Every future situation S at time $t(t_0 + j)$ as well as every situation in the past $(t_0 - j)$ can then be exactly calculated from the situation at t_0 with the help of the Hamiltonian differential equation

$$\dot{x}_i = \frac{\partial H}{\partial p_i}\ \dot{p}_i = \quad \frac{\partial H}{\partial x_i}$$

in which, the dot denotes the ordinary derivative with re spect to time of the functions $p_i = p_i(t)$ and $x_i = x_i(t)$ (called coordinates), taking values in some vector space, and $H = H(p_i, x_i, t)$ is the so called Hamiltonian or (scalar valued) Hamiltonian function. Therefore, given the situation at t_0, with the help of the above Hamiltonian nomological hy pothesis, one can derive situation S through a deductive nomological explanation. If one interprets the initial situ ation as cause of sufficient ground and every future situ ation as effect, then one accepts the theorem of Leibnitz, according to which every mechanical natural situation is uniquely determined by sufficient grounds, or that the same causes will always result in the same effects. This property, however, is only true under the assumption that we can measure the situation with infinite exactness, and of course, no human being has this ability, which can only reside in some supranatural creature (Laplace's 'demon' or 'omniscient calculator'). Nevertheless, it is this conception of mechanical causation, which underpins (virtually) all conventional scientific understandings of determinism.

What is true for mechanical systems is also true for electromagnetic fields in combination with the Maxwell–Lorentz' equations, while the differential equations of relativity theory describe a deterministic system. In the latter, though, the speed of light c is a limiting factor; and in the Minkowsky space of special relativity theory, two events (x_1, t_1) and (x_0, t_0) can only follow each other in time if $c^2(t_1 - t_0)^2 - (x_1 - x_0)^2 \geq 0$. Therefore, we can also distinguish events which cannot be qualified as past or future events and thus cannot be related with other events along the time axis required by standard 'cause' and 'effect' reasoning. Otherwise, however, in relativity theory we are still dealing with a fully deterministic system.

In contrast, in quantum mechanics, it has been shown, according to the Heisenberg 'uncertainty principle', that for a single electron 'not' all of the variables needed to describe its situation (location) can ever be measured at the same time. As a consequence, in the wave mechanics of Schrö dinger, the classical description of the situation is replaced by a ψ function, with which the situation of the system is not described exactly but is expressed in the form of a probability distribution of location (x_i) and impulse (p_i) at time t. With this mode of description and with the help of the Schrödinger equations, the ψ situation at t can be fully determined from the ψ situation at t_0. The physical systems of quantum mechanics are therefore not deterministic but only probabilistic. One big implication is that the so called 'new' physics actually departs from conventional scientific understandings of determinism, acknowledging rather than denying the role of uncertainty, chance, and the often un expected juxtapositional effects of space–time. As Doreen Massey has explained, this realization is crucial for at least two reasons: (1) it suggests a more sophisticated version of natural science than that to which academic geography, physical and human, has all too often been deferential; (2) it thereby raises the possibility of geographers being confident to operate with, and to contribute to formulating, alternative frameworks for 'scientific' inquiry and explanation; and (3) it also holds open the promise of 'nondeterministic' con ceptions and methods derived from academic geography and related (natural and social) sciences, complete with challenging insights into the character of time–space, be coming more widely influential.

See also: Behavioral Geography; Darwinism (and Social Darwinism); Ecology; Foucauldianism; Humanism/ Humanistic Geography; Positivism/Positivist Geography; Possibilism; Probabilism; Ritter, C.; Spatial Science; Structuration Theory; Structurationist Geography.

Further Reading

Beck, H. (1953). *Moritz Wagner als geograph. Erdkunde* 7, 125 128.

Burton, I. (1963). The quantitative revolution and theoretical geography. *The Canadian Geographer* 7, 151 162.

Chappell, J. E. Jr. (1981). Environmental causation. In Harvey, M. E. & Holly, B. P. (eds.) *Themes in Geographic Thought*, pp 163 186. London: Croom Helm.

Claval, P. (ed.) (1993). *Autour de Vidal de la Blache: La Formation de l'école Française de Géographie*. Paris: Centre National de la Recherche Scientifique (CNRS).

Febvre, L. (1925). A Geographical Introduction to History. London (esp. Foreword by Berr, H, pp v xx): Routledge & Kegan Paul.

Freeman, T. W. (1961). *A Hundred Years of Geography*. London: Gerald Duckworth.

Glacken, C. J. (1967). *Traces on the Rhodian Shore. Nature and culture in Western thought from ancient times to the end of the eighteenth century*. Berkeley, CA: University of California Press.

Gregory, D. (1981). Human agency and human geography. *Transactions of the Institute of British Geographers* 6, 1 18.

Hard, G. (1973). *Die Geographie. Eine Wissenschaftstheoretische Einführung*. Berlin: de Gruyter.

Hassinger, H. (1930). *Über Beziehungen Zwischen der Geographie und den Kulturwissenschaften*. Freiburger Universitats Reden. no. 3. Universtat Freiburg, Freiburg.

Huntington, E. (1907). *The Pulse of Asia: A Journey in Central Asia Illustrating the Geographic Base of History*. Boston, MA: Houghton, Mifflin.

Johnson, D. W. (ed.) (1907). *Geographical Essays by William Morris Davis*. New York: Dover Publications.

Lewthwaite, G. (1966). Environmentalism and determinism: A search for clarification. *Annals of the Association of American Geographers* 56, 1 23.

Livingstone, D. N. (1992). *The Geographical Tradition: Episodes in the History of a Contested Enterprise*. Oxford: Blackwell.

Marsh, W. M. (2005). *Environmental Geography: Science, Land Use and Earth Systems* (3rd edn.). London: Wiley Higher Education.

Martin, A. F. (1951). The necessity for determinism. *Transactions of the Institute of British Geographers* 17, 1 12.

Massey, D. (1999). Space time, 'science' and the relationship between physical geography and human geography. *Transactions of the Institute of British Geographers* 24, 261 276.

Peattie, R. (1940). *Geography in Human Destiny*. Boston, MA: George W. Stewart.

Platt, R. S. (1948). Determinism in geography. *Annals of the Association of American Geographers* 38, 126 132.

Ritter, C. (1822 1859). *Die Erdkunde im Verhältnis zur Natur und Geschichte des Menschen, oder Allgemeine Vergleichende Geographie als sichere Grundlage des Studiums und Unterrichts in Physikalischen und Historischen Wissenschaften*. Berlin: Reimer.

Schopenhauer, A. (1819). *Die Welt als Wille und Vorstellung*. Leipzig: Brockhaus.

Schulz, H. D. (1980). *Abhandlung des Geographischen Instituts Anthropogeographie, Vol. 29: Die Deutschsprachige Geographie von 1800 bis 1970*. Berlin: Geografisches Institut der Freien Universitat.

Semple, E. C. (1911). *Influences of Geographic Environment on the Basis of Ratzel's System of Anthopo geography*. New York: Holt.

Spate, O. H. K. (1957). How determined is possibilism? *Geographical Studies* 4, 3 12.

Stoddart, D. R. (1987). To claim the higher ground: Geography for the end of the century. *Transactions of the Institute of British Geographers* 12, 327 336.

Tatham, G. (1951). Environmentalism and possibilism. In Taylor, G. (ed.) (1953) *Geography in the 20th Century*, pp128 162. New York: Philosophical Library.

Tatham, G. (1951). Geography in the nineteenth century. In Taylor, G. (ed.) *Geography in the 20th Century*, pp 28 69. New York: Philosophical Library.

Taylor, G. (1951). Introduction. In Taylor, G. (ed.) *Geography in the 20th Century*, pp 3 27. New York: Philosophical Library.

Wagner, M. (1868). *Die Darwinsche Theorie und das Migrationgesetz der Organismen*. Leipzig: Duncker and Humblot.

Relevant Websites

http://plato.stanford.edu/entries/determinism causal/
Stanford Encyclopedia of Philosophy: Causal Determinism.

Development I

K. D. Willis, Royal Holloway, University of London, Egham, UK
M. S. Kumar, Queen's University Belfast, Belfast, UK

Glossary

Eurocentrism A way of thinking which assumes that the European experience is the only or the best way of achieving something and that Europe should be the model for other parts of the world to follow.

Global North A term derived from the Brandt Commission's reports in the 1980s which divided the world into two parts: North and South. The Global North encompassed the economically richer countries of Europe, the Soviet Union, Japan, Australia, New Zealand, the United States of America, and Canada.

Global South The collective term coined by the Brandt Commission in the 1980s to refer to the economically poorer countries of Latin America and the Caribbean, Africa, and Asia.

Gross Domestic Product (GDP) The value of all goods and services produced within a national boundary.

Gross National Product (GNP) The value of all goods and services claimed by the residents of a particular country, regardless of where the goods and services were produced.

Post-Structuralism A theoretical approach which recognizes that knowledge is always partial and situated, rather than adopting singular interpretations of the world.

Introduction

Conceptions of development and underdevelopment vary over time and space and the explanations for differences in development are similarly diverse. Geographers have considered 'development' in two different ways: first as a term encompassing ideas of economic, political, and social progress and second, as referring specifically to intentional interventions aimed at achieving such progress. The latter became much more common in the post World War II period when international attention turned to the situation in Africa, Asia, Latin America, and the Caribbean, especially in the context of decolonization and the Cold War. Thus, 'development' is often considered as a set of processes which relate purely to the parts of the world collectively entitled the 'Third World' or 'Global South'.

'Development' has received attention from a range of social sciences, including, of course, 'Development Studies' Geographical approaches to development have focused on issues of spatialities and scale, with particular attention being paid to flows of ideas, capital, and people, and the exercise of power at a range of scales. Three main themes come out of geographers' work on development: first, how and why definitions of development vary across space and how this reflects, reinforces, and challenges prevailing power relations; second, a consideration of how and why development varies across space; and third, how development interventions are experienced and generated at different spatial scales.

Development is a grounded process, carried out by different actors attempting to achieve a range of goals. Geographers and others have sought to examine the actions of key development actors, most notably, governments (in both the Global North and Global South), individuals, and communities in the Global South, private sector companies, and civil society organizations. The role of these different actors has varied over time as a range of development approaches and theories have emerged, and emphasis has been placed on particular sectors as the key routes to development.

In this article, we present a largely chronological discussion of how geographers have considered development. However, key questions cut across the different time periods and while we have to present a rather simplified view for the purposes of this discussion, it is important to remember that approaches and theories coexist and overlap at times.

Development as Modernization

Ideas of progress and civilization were key to the colonial projects embarked upon by European powers from the fifteenth century onward. However, this was rarely framed in terms of development as a concept. The idea of development really emerged in the post World War II era when it was believed that poverty was not the inevitable destiny of humankind. This awareness was spurred by the emergent affluence witnessed in the developed Global North. After the horrors of war and the earlier period of global economic insecurity in the 1930s, and within a context of increased decolonization as the countries of Africa and Asia moved toward independence, debates around the correct development path became more common. The other major influence for the emergence of development as an 'ideal' was the Cold War, whereby both the United States and the Soviet Union were competing to attract newly independent Third World countries to join their respective camps – capitalist or communist/socialist. They did this partly by providing development assistance and support.

Development from the point of view of the 'First World' or capitalist world was based on the experiences of the Global North and placed a particular form of modernity as the goal. This included urbanization, in dustrialization, nation state building, and the replace ment of traditional patterns of thoughts and belief with a notion of scientific economic rationality. Development was therefore seen as a process of catching up with the Global North by adopting the policies that had been viewed as successful in North America and Western Europe. These policies were to be implemented by governments in the Global South, often with assistance or aid from multilateral institutions (such as the World Bank) and Northern governments.

Economic growth was a key indicator of modern ization or development, and a range of economic indi cators were used to measure development. These included gross national product (GNP) or gross domestic product (GDP), GNP or GDP per capita, and annual changes in these measures. Other indicators used assessed the contribution of agriculture, industry, and services to the national economy or the make up of the national labor force. The assumption was that national level economic progress would be associated with improved standards of living as health and education levels im proved due to rising levels of wealth. There was also a belief that the benefits of modernization would eventu ally spread throughout the entire population through processes of 'trickle down'.

Geographers became involved in mapping different levels of development (based on economic measures) at both national and subnational levels. Differences between countries were understood to represent different stages in a linear progression from 'traditional' or 'backward' so cieties to 'modern' or 'developed' ones. The solution was to follow the examples of the countries of the Global North to ensure that internal conditions were appropri ate for development.

Within countries, development or modernization was assumed to spread outward from a number of core areas. Geographers represented such diffusion through map ping the spread of infrastructure, such as roads and railways, health centers, and schools. The main actors of such modernizing tendencies were national governments, external donors, and private sector companies. Policy debates flowing from such analyses focused on whether the concentration of investment in key locations would exacerbate or alleviate spatial inequalities over time. Based on his research in Venezuela, John Friedmann argued that the use of growth poles as sites of investment would eventually lead to the spread of economic devel opment out from these key sites. In contrast, Gunnar Myrdal claimed that without state intervention, growth poles would suck in investment and people, leaving more peripheral areas behind.

Emphasis on economic progress came to occupy the centerstage of development in the post war period. Here economic growth as a mythically imbibed concept be came a proxy for development and economists main tained the primacy of growth of per capita income as a standard index for development. Increased production and consumption were seen as twin processes to achieve the goals of development. Indeed, distribution of wealth and assets were rarely considered appropriate in a growth obsessed society. However this focus has been criticized as being Eurocentric both in the definition of development used and the routes by which development is to be achieved.

Marxist and Structuralist Approaches to Development

In the 1960s and 1970s, geographers, among others, con tinued to focus on processes of uneven development, both globally and within countries. However, rather than viewing spatial differences in development as reflecting variations in stages along a development path, many geographers adop ted a more radical interpretation, arguing that despite all the optimism of development there were structural factors which both explained existing patterns of development and which prevented countries in the Global South achieving development in the future. Processes of capitalism, coloni alism, and imperialism were used to explain global devel opment inequalities. As with the modernization ideas discussed earlier, development was again defined largely as economic progress, which, it was assumed, would result in welfare improvements.

Many geographers drew on ideas and theories coming from the Latin American experience. These included the Latin American structuralist ideas derived from the work of Raúl Prebisch and colleagues at the United Nations Economic Comission for Latin America (ECLA or CEPAL in its Spanish acronym). In the post World War II period, Prebisch argued that Latin American countries would not be able to achieve the levels of economic development found in the countries of the Global North. This was because the global economic system was very different and unequal in the middle of the twentieth century, compared with that found in the eighteenth and nineteenth centuries when many European countries and the United States of America had made great economic progress. The economic power of Global North countries in the twentieth century meant that it was almost impossible for poorer countries to progress in the same way. For Prebisch, a key policy solution was for Latin American nations to protect their national economies, particularly infant industries, through the use of tariff barriers and other forms of protectionism.

Other scholars drew on the theories of the so called 'dependency school' which adopted Marxist approaches to explanations of development and underdevelopment in the Global South. For example, Andre Gunder Frank's work on economic change in Brazil and Chile since the start of European colonialism in Latin America concluded that capitalism and colonialism created conditions of underdevelopment and exacerbated economic inequalities at a range of scales. Thus, according to dependency theorists, global development inequalities were not just a reflection of different stages of capitalist development, but rather were a creation of capitalist development in the Global North which was partly based on the exploitation of the Global South.

Geographers who interpreted capitalist development processes as the causes of global inequalities between the North and South were part of a broader field of radical geography which developed in the 1960s drawing on Marxist theory. The inevitable outcome of such analyses was to advocate an end to capitalism.

Neoliberalism as a Route to Development

Under post war modernization interpretations of development and those of Marxists and structuralists, the state at national and local levels was viewed as an important actor in guiding and shaping development, although, of course, theorists differed in the extent to which they believed the state should be in overall control. By the 1970s, however, a growing body of theory, particularly in political science and economics, was emerging which was drawing on ideas from the nineteenth century and earlier about the role of the market in economic progress and development. This set of theories was termed 'neoliberalism'. Rather than being a positive or productive force in development, neoliberal theorists argued that the state acted as a brake on growth as it was inefficient and open to corruption and mismanagement. Instead, decisions about what to produce, where to produce, and how much to charge should be determined by the invisible hand of market. The central priorities for essential development were structural reform, financial deregulation, liberalization, and privatization. Once again the core meaning of development was economic growth, but now the agency of development shifted from the state to the market.

Global economic events in the 1970s provided the conditions whereby neoliberal policies could be implemented. Sharp rises in oil prices followed by a slowdown in the global economy and an increase in interest rates meant that many countries found it increasingly difficult to meet their external debt payments. During the post war period, many governments in both the Global North and Global South had borrowed significant amounts of money to fund large scale modernization and development projects. With the economic changes of the 1970s, such a development approach came under threat. Following Mexico in 1982, many countries defaulted on their debt repayments. In order to gain access to foreign currency, the only option available was to approach the international financial institutions of the World Bank and the International Monetary Fund (IMF).

These organizations agreed to provide financial support on condition that countries followed certain policies encompassed in 'structural adjustment programs' (SAPs). These were based on key neoliberal principles of reducing state involvement in production decisions and promoting the role of the free market. The main dimensions of SAPs were privatizations, reductions in tariff barriers and import quotas, currency devaluation, and an opening up to foreign investment. Cutting back on government spending also involved the laying off of large numbers of staff and major reductions in welfare expenditure.

Geographers have considered neoliberalism in the context of development in two major ways. First, they have examined how neoliberalism as an ideology and set of practices has moved around the world and been implemented in particular places through processes of neoliberalization. While the World Bank and IMF set the conditions for SAPs and governments have few, if any, alternative options for continued financing, there are variations in how governments and national elites are engaged in implementing the policies. The importance of certain places in the Global South in both shaping neoliberal thought and being used as exemplars of successful economic development under a neoliberal model should also be recognized. Key examples are Chile after the 1973 coup when General Augusto Pinochet implemented neoliberal economic policies based on the teachings of Milton Friedman at the University of Chicago, and the East Asian economies which were held up in the early 1990s as neoliberal development successes having achieved high levels of economic growth without increased inequality. In the case of East Asia, critics argued that the role of the state was far more important in economic growth success than neoliberals acknowledged and that the East Asian financial crisis of 1997 was in large part a result of liberalization of the financial sector. In Chile, Pinochet's economic successes were achieved in the context of authoritarian rule.

The second way in which geographers have examined neoliberalism within a development context is through investigating the impacts of SAP's and associated policies on people, particularly marginal groups such as the poor and women. Such research has often concluded that while macroeconomic indicators may have stabilized and improved as a result of SAP's, social and economic inequalities have increased both in the long and short term as a result of the drastic policies. Reduced workers' rights due to changing labor legislation, the rise of informal sector working, and increases in basic food costs as government subsidies are withdrawn have often combined to increase the precariousness of everyday existence for millions of people. This is not to deny the role of neoliberal policies in job creation and improved services in some contexts, but to highlight their potential devastating effects.

SAPs and the implementation of neoliberal policies in the 1980s and early 1990s were criticized for their insenstivity to cultural and social differences, presenting a set of technical solutions which could be applied regardless of location. These critiques led to a shift toward nationally specific 'poverty reduction strategies' which have to include consultation with appropriate stakeholders throughout the country. While this move is welcome, it is still based on an attempt to implement technical solutions rather than considering how power is exercised and experienced, that is, an understanding that development has a political dimension.

Participation and Development at the Grassroots

The definitions and approaches to development which have been discussed so far have been based on ideas about development which stress the economic and lead to policies which have been imposed from the top down. This does not consider other definitions and how people at the grassroots envisage development. So called alternative forms of development have often stressed the importance of focusing at a local scale, allowing individuals and communities to create and agree on their own ideas of development and how it should be achieved. Such an approach is not only more effective, it is argued, because it is appropriate to a particular context, but it is also more democratic and has the potential to transform existing power relations by encouraging participation.

The focus on community and social sector development, particularly since the 1980s, has promoted alternative forms of development, for example, those informed by ideas of entitlements and capabilities as outlined by Amartya Sen. Sen's approach to development focuses on seeing individuals and communities as having agency and abilities, rather than presenting poor or marginal communities purely through the lens of 'problems' which outsiders need to solve. He also recognizes the limits which are present due to economic, social, political, or cultural marginalization resulting in inequalities in entitlements. Sen sees development as encompassing ideas of freedom whereby individuals and communities have the choices and abilities to shape their lives as they see fit.

The role of nongovernmental organizations (NGOs) has expanded with the growing focus on grassroots development. NGOs are often seen as ideal institutions to support local people because they are separate from the state and they are not driven by profit motivations. They are also viewed as being sensitive to local contexts and more efficient as a result.

While grassroots development has been presented as a contrast to other forms of top down development, they are in fact strongly linked. The growth of NGOs and a focus on community based development is not purely a reflection of alternative development strategies; the rise of neoliberal development thinking and policies has had a significant effect. This is for two main reasons; first, as governments have implemented neoliberal policies involving cuts in government spending, services which were previously supplied by the state are either provided by the market at a cost, or are not provided. This has meant the rise of NGOs in filling the service gap for poorer communities in sectors such as housing, health, and education. The notion of 'self help' has often been used to describe this form of activity as governments and private sector companies are not involved.

Second, a key dimension of neoliberalism has been decentralization and a discourse of greater participation. NGOs are seen as ideal institutions to facilitate such processes given their perceived effectiveness at a grassroots level and their ability to engage with a range of social and cultural groups. For both these reasons, international financial institutions and Northern government donors have increasingly directed aid through NGOs, rather than through Southern governments. Thus, the idea that NGO spaces of operation are inherently autonomous of, or alternative to, formal neoliberal development spaces is not reflected in reality.

The role of grassroots development and NGOs in particular in the context of participation has also been critiqued. Grassroots activities can be empowering as people who are normally left out of decision making processes are able to participate, thereby creating new active modes of citizenship and reducing dependency on the state. However, participation can vary greatly from mere attendance at a meeting where 'participants' are informed of development policies and projects, to meaningful participation involving active engagement in the initial development and design of projects. Additionally, discussions of participation are sometimes based on the false assumption that 'communities' are internally homogenous. In reality, power relations within communities can affect who participates and how.

Sustainable Development

In the field of development, geographers have been particularly keen to examine the relationship between development processes and the physical environment. In the post war period and the focus on economic growth under a modernization model, the natural environment was viewed as a resource which could be exploited for human advancement. This failed to recognize the potential for environmental destruction, especially when techniques used in the mid latitude temperate environments of Europe and North America were transferred to tropical zones with very different soil and vegetation conditions.

Environmental awareness in Northern development circles rose greatly during the 1960s and 1970s, but it was in the 1980s, with the report of the Brundtland Commission in 1987 and then the Rio Conference on Environment and Development in 1992, that environmental concerns came into mainstream government and donor thinking, particularly around the concept of 'sustainable development'. The Brundtland Commission (also known as the World Commission on Environment and Development) saw sustainable development as encompassing practices which met the needs of today's population, but did not undermine the availability of resources for future generations to meet their needs. Of course, the concept of 'needs' is very contested and varies over time and space and there have been highly contrasting debates around how to achieve this goal.

Two main approaches to sustainable development can be identified; the first is most widespread and can be termed a 'technocentric' approach, while the second is 'ecocentric' and involves transformations in the current forms of development. Technocentric approaches see environmental problems as something which requires a technical solution, such as a new way of generating power or traveling long distances. Under such approaches, economic growth is still viewed as important and desirable as it is market pressures which will help drive innovation. Putting a price on environmental goods, such as carbon trading, is also part of this approach. Such approaches have been criticized for not addressing the massive inequalities between the rich and the poor on the planet and doing too little to prevent future environmental disasters.

In contrast, ecocentric approaches place the environment at the heart of the discussion and see current forms of development as being unsustainable. According to ecocentric theorists and practitioners, sustainable development can only be achieved if present patterns of consumption and levels of economic growth are dramatically changed. Such approaches usually advocate small scale, community based development as this will allow projects which take into account local environmental conditions. It may also embrace indigenous knowledges about local environments and appropriate environmental management. However, ecocentric approaches require major shifts in current consumption patterns among the richer people of the world, and deny poorer populations the chance to enjoy standards of living which have so far been denied to them.

Postdevelopment and Postcolonialism

In this article, we have highlighted the ways in which development has been defined and implemented. A recognition of the ideological nature of development and the way in which definitions and policies reflect and reinforce the exercise of power has been a theme throughout. Developments in postcolonial theory and the rise of what has been termed 'postdevelopment' has contributed to understandings of development as a contested and non neutral process.

Postcolonial interpretations of development are that it is a largely Eurocentric product based on an unawareness of, refusal to engage with, or explicit strategy to destroy, non European interpretations of the world and what progress might be. Postcolonial approaches involve contesting ideas which are seen as 'natural' from a 'European' viewpoint, such as progress meaning economic growth and the focus on individual rights. Such approaches also challenge the Eurocentric representations of the peoples and places of the Global South as poor, uneducated, and passive.

This latter point has been a major part of the set of ideas termed 'postdevelopment'. These ideas came to the fore from the 1990s onward and involved an attack on development as an idea, arguing that it was Eurocentric and involved the mobilization of particular discourses about progress and modernity. For example, Arturo Escobar (1995) used the case of Colombia to argue that in the 1950s and 1960s the USA and the World Bank had constructed Colombia as poor, populated largely by uneducated peasants and passive women who needed development. This development was to be provided in the form of large scale top down projects, the benefits of which would trickle down to the people at the grassroots. These discourses were often adopted and internalized by government officials and the population more generally such that development became something desirable.

For postdevelopment theorists, grassroots community based development is a much more appropriate path forward. This would allow much greater involvement of local people in determining how they want to live their lives. However, as critics of postdevelopment have highlighted, this focus on community development is something which has been occurring in more 'mainstream' development for a number of decades. The top down, modernization driven approach critiqued by post developmentists has been discredited in most situations. In addition, as with advocates of grassroots development, more generally, there is a need to consider how far 'lo cally based' projects can achieve their goals within the context of broader national and global structures.

As this article has demonstrated, geographers have long been involved in challenging hegemonic ideas of development and investigating its material outcomes. However, postcolonial and postdevelopment research, as well as post structural ideas have become increasingly important since the late 1980s. This has led to a massive growth in development research which is aware of how gender, ethnicity, religion, and other forms of identity both shape and are shaped by development processes.

Summary

This article has outlined the key ways in which devel opment as a concept and set of policies has been dealt with within geography and more broadly in policy en vironments mainly since World War II. There has been a clear recognition of the contested nature of the concept and how development as an ideology has been used in particular places and times with certain outcomes. Geographers' engagement with postcolonial and post structural accounts of development and approaches to analysis have been important in this process. In addition, the general trend away from a purely economic focus has been obvious, as has been the tendency to consider grassroots and community based scales of analysis and policy intervention. This focus on the local is welcome in that it can allow for greater understanding of diverse opinions, but it should not hide the inherently political nature of development and how power relations and in equalities are embedded in the very concept. Develop ment despite its feel good factor, implies struggle and conflict to acquire resources and freedom of choice for the betterment of life for the majority.

See also: Brandt Commission; Debt; Dependency; Development II; East Asian Miracle; East/West; Eurocentrism; First World; Globalization, Economic; Human Geography; Latin American Structuralist School; Neocolonialism; Neoliberalism and Development; Nongovernmental Organizations; North-South; Participation; Postcolonialism/Postcolonial Geographies; Postdevelopment; Poverty; Rio Summit; Second World; Sustainable Development; Third World.

Further Reading

Baran, P. A. (1957). *The Political Economy of Growth*. New York: Monthly Review Press.
Corbridge, S. (1986). *Capitalist World Development: A Critique of Radical Development Geography*. Basingstoke: Macmillan.
Corbridge, S. (2002). Development as freedom: The spaces of Amartya Sen. *Progress in Development Studies* 2(3), 183 217.
Cornwall, A. and Brock, K. (2005). What do buzzwords do for development policy? A critical look at participation, empowerment and poverty reduction. *Third World Quarterly* 26(7), 1043 1060.
Crush, J. (ed.) (1995). *Power of Development*. London: Routledge.
Escobar, A. (1995). *Encountering Development: The Making and Unmaking of the Third World*. Princeton, NJ: Princeton University Press.
Friedmann, J. (1966). *Regional Development Policy: A Case Study of Venezuela*. Cambridge, MA: MIT Press.
Kambhampati, U. S. (2004). *Development and the Developing World*. Cambridge: Polity.
Lawson, V. (2007). *Making Development Geography*. London: Hodder Arnold.
Leftwich, A. (2000). *States of Development: On the Primacy of Politics in Development*. Cambridge: Polity.
Potter, R. B., Binns, T., Elliott, J. A. and Smith, D. (2008). *Geographies of Development: An Introduction to Development Studies* (3rd edn.). Harlow: Pearson Prentice Hall.
Power, M. (2003). *Rethinking Development Geographies*. London: Routledge.
Rahnema, M. and Bawtree, V. (eds.) (1997). *The Post Development Reader*. London: Zed Books.
Sachs, W. (ed.) (1992). *The Development Dictionary*. London: Zed Books.
Sen, A. (1999). *Development as Freedom*. Oxford: Oxford University Press.
Williams, G., Meth, P. and Willis, K. (2009). *Geographies of Developing Areas: The Global South in a Changing World*. Abingdon: Routledge.

Relevant Websites

www.imf.org
International Monetary Fund.
www.worldbank.org
World Bank.
www.wdm.org.uk
World Development Movement.

Development II

B. Korf, University of Zürich – Irchel, Zürich, Switzerland

Glossary

Complex Emergency Humanitarian crisis situation that is complex, combining political and natural causes that go beyond unidirectional explanations.

Developmental State State as an agent that is actively promoting development in a symbiosis of the political and economic elements of statehood.

Neopatrimonialism System of statehood governed through personal loyalties and reciprocal relationships of an essentially personal kind between leaders and their followers within the formal hierarchy of the state.

Shadow State Profits that can be made through contraband trade and smuggling have attracted state officials who operate a parallel shadow state, while keeping control of formal institutions of statehood.

Sobel Soldiers by day, rebels by night.

State Collapse The crumbling of institutions, where the state's structure, authority, law, and political order have fallen apart.

State Failure It is defined by the nonperformance or meager performance of specific state functions.

War Economy Bush economies – with a link to global contraband trade – organized around the power realms of roving bandits, who utilize violence and suppression of local civilians as a conscious strategy of economic reward.

Washington Consensus Emerged in the 1980s to describe the agreement of international financial institutions to pursue neoliberal policies whereby the function of the state was reduced to being an arbiter of market forces.

Introduction

Development and state building in postcolonial societies are twin processes. Social and economic development cannot be thought of without a state providing basic functions of welfare and security. It may be debatable what the functions of a state are, but development policy discourses commonly assume that development requires a functioning state. However, in many postcolonial societies, the construction of a state has been a difficult and contested political process. The challenges of building a postcolonial state – often based on arbitrary territorial boundaries – have been experienced by a large number of societies that have undergone violent conflict, civil war, and social anomie in the last few decades.

With the end of the Cold War, the phenomenon of societies experiencing partial or complete state collapse appears to have become more pertinent as exemplified in Somalia's slide into social anomie after 1990, but there are many more countries that have followed since then, such as Sierra Leone, Liberia, Congo, and Afghanistan. Public debate has attributed this to a coming anarchy or has linked those stateless territories with the breeding of global terrorism. Some scholars have started talking of a new type of war – the new wars – that are anarchic, fought by militias and warlords and rather intrastate than interstate. In some cases, state collapse and the new wars were mainly attributed to outside interventions (e.g., Iraq), but in other situations, it was argued that the lack of international intervention had resulted in an internally driven process of social deterioration (e.g., Somalia).

Although some of the implicit assumptions of the 'new wars' discourse are problematic, the phenomenon of state failure and collapse has been linked with many of the most appalling instances of human suffering in the modern world. It therefore poses challenging questions for development theorists and practitioners alike: What makes states fail – and what does failure mean in this context? How can the outbreak, persistence, and spatial extent of social anomie and violence – the complete collapse of state (and social) order – be explained? How can the international community deliver humanitarian assistance under such conditions? And is there a way out of protracted violent conflict to rebuild the state in violence striven societies?

State, Violence, and Development

A number of different words are used for describing phenomena that link social anomie and political violence to the state and its existence, namely state collapse, state failure, state fragility, or complex (political, protracted, humanitarian) emergencies. All those notions carry in them an implicit normative ideal, that of the Western liberal state as a rational organization with extensive powers and massive bureaucracies against which the current status of those borderland societies are measured. With the failure of the USSR as an alternative to liberal democracy, the only accepted global model of legitimate statehood is based on principles of multiparty democracy, respect for human rights, and good governance.

Two conceptions of the state in development have been popular. The concept of a developmental state was based on the premise of a supportive two way relationship between the political and economic elements of statehood. The developmental state was an active enabler and creator of development. South Korea is the classic example where a developmental state in the 1960s and 1970s was successful in steering a process of economic development. The second view of the role of the state is the Washington Consensus that emerged in the 1980s and reduced the role and responsibility of the state to that of an arbiter of market forces, placing the state in the role of servant of the market. In this conception, the state is the essential manager of a more or less regulated market economy and responsible for legal order, currency, infrastructure, and the provision of basic social services – or the tripartite functions of order, welfare, and security. Although both concepts differ with regard to their conceptions of the role of the state, they share the presumption that development is not possible without a state.

Where states do not work, it is presumed that they have failed to adopt the formula of success of the liberal democratic model. The phenomenon of state collapse as the most radical form of state failure is often represented as the 'other' of development, modernity, and the Western liberal state, as the irrational, barbaric instincts of the human specie coming to the fore. These metaphors contrast the opposing characteristics of order (us) and chaos (them) by juxtaposing traits of barbarity, excess, and irrationality on borderland or peripheral territories experiencing state collapse and characteristics of civility, restraint, and rationality on metropolitan, developed societies and territories.

Fragile States – Collapse, Failure

It is useful to distinguish between state collapse, which is concerned with the crumbling of institutions, where the state's organizational structure, authority, law, and political order have fallen apart, and state failure, which is defined by the nonperformance or meager performance of specific state functions. State failure may be a matter of degree, not all failing states collapse, but states that fail to deliver essential functions of statehood (as defined by the liberal model) without complete collapse may be termed weak states or disintegrating states, although their façade of statehood is often still intact. For example, Sri Lanka as a state fails to provide protection and security to a considerable number of its citizens, but the state structures are still visible and functioning even in most of the civil war affected areas in the northeast, which are contested by Tamil insurgents. In the case of Rwanda, the Hutu dominated state was effectively organizing militias that conducted the killing, so the state was failing to provide security to some of its citizens, mainly Tutsi and oppositional Hutu, but it had not collapsed.

There are several explanations about the mechanisms that weaken state structures or bring about state failure or state collapse. Broadly speaking, some scholars emphasize internal factors, others external factors, but depending on the specific context, the causal links between internal and external factors that make states collapse or fail may differ considerably. Processes of failure are often explained with the inability of a state to deliver its social, political, and economic functions. What the exact functions are and how the nondelivery of such functions undermines the coherence and strength of statehood is subject to debate, but it is clear that the provision of these functions requires a costly social organization, which may exceed what poorer societies with scattered populations in inaccessible territories may be able to afford.

Representation, identity, and loyalty

In many postcolonial societies, authoritarian rulers took power after decolonialization. In some cases, such as South Korea and Taiwan, such authoritarianism was successful in creating a common national identity, but in many other postcolonial states, this identity formation toward a nation state failed to be successful. Authoritarian regimes were often supported by the West as guarantors of political stability and as benevolent dictators that would be better able to construct a nation than a representative system of government that may lead to political fragmentation. These authoritarian nation building projects were embedded in a mass of symbols and rituals pretending a functioning state where there was often only a shadow of it.

In many countries where identity and loyalty was traditionally accorded to kinship, public institutions of statehood were subverted into private mechanisms designed to promote wealth and power for a specific kin group. Statehood became governed through personal loyalties – reciprocal relationships of an essentially personal kind between leaders and their followers within the formal hierarchy of the state. Neopatrimonialism is based on short term and individual reciprocities and often builds on kinship bonds. Neopatrimonial rulers have to pay off their clients in these chains of reciprocity, which diverts resources away from public use. This can undermine the very basis of neopatrimonialism in the longer term when it leads to a shrinkage of the resource base upon which the state can draw in the future. Neopatrimonialism contradicts the pertinent liberal model of the modern state, which is based on a rational, impartial bureaucracy eager to protect the common good based on generalized reciprocity of all citizens.

Neopatrimonial rulers utilized the anticolonial ideology of nationalism to cover up the cracks in the façades of common identity and statehood. This, however, created a dilemma. While the nation building narratives of postcolonial rulers triggered the expectations and demands on what the state should deliver, the neopatrimonial structures of the state undermined the basis upon which the state could have delivered these demands to a broader populace. This rising gap between promise and performance of postcolonial rulers undermined their credibility and legitimacy among the socially excluded parts of the postcolonial nation. Neopatrimonialism effectively hollowed out the capabilities of statehood in the long term and encouraged insurgent movements, ethnic conflict, and military *coups d'état*.

Wealth and welfare

The concept of a developmental state was built on the supportive two way relationship between the political and economic elements of statehood. Economic globalization and structural adjustment programs as part of the Washington Consensus, such as privatization of parastatal corporations, the reduction of tariffs, and the opening of markets, reduced the rent collecting capacities of state elites. The planning and developmental state was dissolved. This encouraged the expansion of spaces of the nonformal economy side by side with the global market economy. Territorial based and nonterritorial shadow economies involve innovative extralegal local–global networks linking local resources (e.g., drugs, diamonds, gold, and minerals) and international markets.

This process of privatization that was encouraged by the metropolitan powers of the West opened new economic opportunities for rulers in fragile states. Access to natural resources and their marketing through transborder networks became a new basis of political power: The profits that can be made through contraband trade and smuggling have attracted state officials who operate a parallel shadow state, while keeping control of formal institutions of statehood (i.e., keeping their official job as state employees). In other cases, smuggling and transborder trade has deprived state bureaucracies of revenues, while providing a revenue base for state destroying insurgencies. These shadow, bush, and war economies have often reversed the economic logic of statehood: the economy of the bush is based on exploitation of extralegal mining of mineral resources or cultivation of drugs and is located in rural border regions which have become revitalized as economic centers of transborder trade and exchange, whereas urban centers – the core of a nation's statehood – are in economic decay.

As those shadow economies depend on the appropriation of easily exploitable natural, mostly mineral, resources, some scholars argue that such resource abundance can become a resource curse, when it supports clientelist states and triggers political disintegration. When states collapse due to insurgent violence, ethnic conflict, and political disintegration, such economic networks are often reshaped into an economy of violence: physical force and anomic violence become strategies of so called warlords to monopolize resource rents. These war economies are not an expression of irrational madness, but are organized around the power realms of roving bandits, who utilize intimidation, violence, and suppression of local civilians as a conscious strategy of economic reward. Humanitarian aid that is distributed to the suffering civilian population has often been incorporated as resource rent in the war economies. Aid then triggers the shadow economies and structures of violence rather than supporting a 'developmental' state.

Order and security

The territorial boundaries of postcolonial states have been largely given, defined, and often arbitrarily drawn by colonial powers. The global doctrine of state sovereignty that bans interventions into internal affairs of a state made any ruler by definition the legitimate ruler of a specific country. Whenever security was threatened by external attack or internal uprising, incumbent rulers as recognized governments could often gain more external military support than their insurgent opponents and allowed the state to survive and its elites to stay in power. This external support to sovereign rulers could become counterproductive by increasing the alienation of the population or parts of the population from their rulers. It allowed quasi states to survive whose existence was guaranteed by global conventions, while their elites and rulers did not have to mobilize their own populace and to build up trusted and commonly accepted institutions of statehood.

Internally contested rulers need to rely on the loyalty of their security forces. Such loyalty is often built on kinship bonds rather than generalized social and political ties between security forces and the population (in its social and ethnic diversity). In societies where loyalty to the state is secondary to kin loyalty as in many multiethnic societies, internal insurgency tends to reinforce the commitments of security forces and their soldiers to their own kin. Similarly, insurgency groups often primarily represent and feel loyal to their specific kin and revolutionary political ideologies are upheld for mobilizing reasons, but become secondary in the further pursuit of military contestation.

In disintegrating societies, when law and order collapse or are close to collapse, feeble governments often recruit militias to shore up their power against insurgent groups. Such militia recruitment often further corrodes the functioning of statehood, because the boundaries between government soldiers and rebels become increasingly blurred with both terrorizing civilians for private gains, while the political course of fighting gets more and more out of sight. This has created a new type of militia: the sobels – soldiers by day, rebels by night. Sobels become essentially private operators, lacking discipline or – at least at first sight – any kind of public agenda. Child soldiers often guided by spiritual beliefs and under the influence of drugs, are an essential element of such disintegrating military combat.

While there has been a tendency in the literature to primarily analyze the dynamics of violence and to de scribe the privatization of warfare, it needs to be kept in mind that many of those combatants have legitimate grievances and pursue related aspirations, such as se curing their livelihoods. Privatization of violence in the form of multiple armed gangs, militias, soldiers, and 'rebels' have undermined the state's monopoly of vio lence. But the grievances that nurture these gangs of violence have often emerged because state elites have abandoned the public pursuit in favor of private enrichment.

Spatialities of State Failure

The explanations of state failure and collapse thus far have primarily been functionalist – they have explained the processes that made functions of statehood disinte grate. Human geographers have emphasized the specific spatialities inherent in such disintegrating processes, which often arise from uneven development. Ethno graphic studies of organized violence have brought out this multileveled and ambivalent character that is rep resented in different governable spaces, those of kinship, ethnic community, and nation state, which are often not compatible with each other and coexist in a complex and contradictory political economy of violence and resource appropriation. This contested political and economic space produces powerful centrifugal forces that are fragmenting the state, creating multiple unstable, un governable spaces culminating in a form of civic vigilantism.

The political economy of resource violence is not confined to local processes, but crosses different spatial scales. The particular territoriality of many fragile states is shaped by loosely populated spaces, which are difficult to control spatially, especially their remote borderlands. Although many insurgencies are fought as internal con testations within a state, they involve transborder eco nomic and military flows. Insurgents often operate across boundaries, the bush, shadow, and war economies operate beyond the local and across state borders and are em bedded in global commodity chains. Therefore, a dif ferentiation between internal and external triggers of state failure is rather artificial. The new wars, the failure of the state and the global economy are two sides of the same coin – a global network economy. Political boundaries are porous – they allow the flow of economic goods and persons (as refugees) across political borders.

Some authors assert that a new geography of war has emerged where resource concentrations become more important than political boundaries. It has been argued that abundance of resources provides incentives for privatization of violence and thus the disintegration of statehood. Whether or not resource appropriation trig gers the disintegration of statehood and how such dis integration occurs often depends on the relative location of such resources vis à vis the power centers, ethnic minorities, and border areas and the spatial distribution of resources (diffuse or concentrated) that determines the technologies of exploitation and whether or not such exploitation requires large scale investment and man agement. However, such a link between resource location and attributes should not be understood as being causally deterministic.

Development and Security

With the end of the Cold War and the global military domination by the US, the global ideology of state sov ereignty upheld throughout the Cold War has been weakened with large scale interventions occurring in several failing states. A new cartography has emerged that considers underdevelopment to be dangerous, for the people involved, but also for the world more broadly. This view has gained ground in the aftermath of 11 September 2001 with the global terrorism discourse. While the US led interventions in Afghanistan and Iraq are prominent expressions of this new global concern, the changing attitudes in aid policy emerged prior to these developments as a response to the so called new wars in Africa and Central Asia.

Complex Emergencies

In the 1980s and 1990s, in view of large scale famines linked with civil wars and disintegrating statehood in several African countries and after the Rwandan ethni cide, humanitarianism and policy discourses of aid agencies largely represented those fragile, failing, and collapsing states as complex emergencies where natural disasters often coincided with protracted political con flict, thereby creating devastating humanitarian disasters for a large populace. The term complex (political or protracted) emergency highlighted that the situation was one of an emergency that had political causes and showed a social complexity that went beyond simple and unidirectional explanations and solutions. The complex emergency discourse hinted at the fact that these emer gencies probably were beyond the reach of resolution and aid was confined primarily to provide basic humanitarian needs to a suffering civilian population.

Aid and humanitarian agencies realized that they had, in some instances, contributed to the prolongation of emergencies by providing resources for combatants or political support to divisive political elites. Development and humanitarian aid had, it was noticed, not always been the means of alleviating underdevelopment, but could in some circumstances unintentionally trigger social conflict or allow the protracted duration of state collapse, social anomie, and political violence. Out of this recognition, donor agencies claimed to have stopped working around conflict (by ignoring their own role in its dynamics). Instead, they developed new policies to work in conflict (by avoiding to do harm when delivering aid) and on conflict (by actively trying to contribute to conflict trans formation through aid). Those discourses on complex emergencies and working in and on conflict, however, retained a managerial understanding of development that was limited to fixing the most severe consequences of state collapse.

Rebuilding the State

Since the end of the 1990s and especially after 11 Sep tember 2001, this managerial understanding of aid in complex emergencies is paralleled by a much more am bitious project of state building and social reconstruction in war torn societies. The new emphasis in these policy discourses is that underdevelopment and state collapse are not only bad for the borderlands, the underdeveloped periphery, but also threaten the stability of the metro politan areas as well through terrorism and global mi gration. Development is considered as a strategy to ensure global security.

In these conceptions, development aid becomes an instrument of international donors for their will to gov ern (globally) – to reconstruct complete societies after their social disintegration as part of a global project of democracy building. This new explicit linking of security with development has encouraged new collaborative engagements of development agencies with limited or full scale military interventions. Two different models can be observed. In the fully interventionist model, core metropolitan states are prepared to deploy military for ces and to engage in long term commitment and to govern these states for an interim period in the form of international trusteeship before they are handed over to newly capacitated local authorities. Examples for such type of intervention are East Timor, Afghanistan, and Iraq. In the weakly interventionist model, indigenous political groupings are induced, encouraged, and forced to attempt a peaceful settlement with some external supervision and assistance, possibly the deployment of a UN peacekeeping force, but the overall commitment of core metropolitan states is much weaker and no inter national trusteeship for the core institutions of statehood is taken on by metropolitan states. The weaker model applies to many settlements of African or Central Asian civil wars.

State building, especially in the fully interventionist model, has often focused on constructing a democratic state with the Western liberal model as a template. Therefore, free elections for parliaments have often been considered as the cornerstone of such state building processes. This is linked with donor discourses of good governance as a precondition and trigger of social and economic development. Such conceptions disregard the different logic of reciprocity that underpins many soci eties with kin based loyalty systems. It constructs con ceptually the dichotomy of modernity, civility (the Western model) as a desirable template on the one hand and tradition, archaic customs, and kin based clientelist societies as the source of state collapse. These conceptions have few suggestions of how to integrate and combine the different logics of good governance and kin loyalty.

Thus far, the experiences with both models have been disappointing and have underlined the difficulties in re building societies after they have disintegrated into social anomie. The case of East Timor underlines that even large scale assistance in a small country and under stable security conditions may not be sufficient to build trusted institutions of statehood and heal the wounds of past conflict. In Afghanistan, where the US led military intervention in 2001 was followed by a large scale influx of aid monies, the success of state building remains un clear. In insecure territories, the military took over a twin role as security forces and deliverers of development to local populations, thereby blurring the boundaries be tween development and security. At the same time, the weakly interventionist model has not produced very promising results either as can be seen in many African civil wars, although the successful transition process in Mozambique is a notable exception.

Issues

State collapse as represented in the new wars discourse is an imagined geographical space of breakdown, excess, and barbarity. Viewed from the Western liberal model of the state, state collapse is the 'other' of development, civility, and progress. The representation of state collapse and new wars as a failure of modernity and as a space of suffering provides justification and legitimacy for the urge to govern – to expand the metropolitan project of development from more modest good governance models to the complete reconstruction of societies as modern, democratic states. With the new discourse of global ter rorism, this project of state building has come to dom inate the international agenda, but its success is still to be proven. The challenge will be how to reconcile different governable spaces, different logics of loyalty, and to allow alternative models of statehood to gain international acceptability.

See also: Borderlands; Critical Geopolitics; Development I; Ethnic Conflict; Ethnicity; Geopolitics; Global Production Networks; Governance, Good; Governmentality;

Hegemony; Imperialistic Geographies; Neocolonialism; Political Ecology; Postconflict Geographies; Sovereignty; State Theory; State; Terrorism; Third World; Uneven Development; Vulnerability; War; Watts, M. J.

Further Reading

Anderson, M. B. (1999). *Do No Harm: How Aid Can Support Peace or War*. Boulder, CO: Lynne Rienner.

Bayart, J. F., Ellis, S. and Hibou, B. (1999). *The Criminalization of the State in Africa*. Bloomington, IN: Indiana University Press.

Berger, M. T. (2006). From nation building to state building: The geopolitics of development, the nation state system and the changing global order. *Third World Quarterly* 27, 5 25.

Clapham, C. (2002). The challenge to the state in a globalized world. *Development and Change* 33, 775 795.

Cramer, C. (2006). *Civil War Is not Stupid Thing*. London: Hurst.

Duffield, M. (2007). *Development, Security and Unending War*. London: Polity.

Goodhand, J. (2006). *Aiding Peace? The Role of NGOs in Armed Conflict*. Boudler, CO: Lynne Rienner.

Keen, D. (2007). *Complex Emergencies*. London: Polity.

Le Billon, P. (2001). The political ecology of war: Natural resources and armed conflict. *Political Geography* 10, 561 584.

Luke, T. W. and Toal, G. (1998). The fraying modern map: Failed states and contraband capitalism. *Geopolitics and International Boundaries* 3(3), 14 33.

Ottaway, M. (2002). Rebuilding state institutions in collapsed states. *Development and Change* 33, 195 211.

Reno, W. (1995). *Warlord Politics and African States*. Boulder, CO: Lynne Rienner.

Rotberg, R. (ed.) (2004). *When States Fail: Causes and Consequences*. Princeton, NJ: Princeton University Press.

Watts, M. J. (2004). Antinomies of community: Some thoughts on geography, resources and empire. *Transactions of the Institute of British Geographers* 29, 195 216.

Wimmer, A. (1997). Who owns the state? Understanding ethnic conflict in post colonial societies. *Nations and Nationalism* 2, 631 665.

Zartman, I. W. (ed.) (1995). *Collapsed States: The Disintegration and Restoration of Legitimate Authority*. Boulder, CO: Lynne Rienner.

Relevent Websites

http://belfercenter.ksg.harvard.edu/
Belfer Center for Science and International Affairs Harvard University.

http://www.css.ethz.ch
Center for Security Studies Swiss Federal Institute of Technology, Zurich.

http://www.csis.org
Center for Strategic and International Studies Washington DC.

http://www.crisisstates.com
Crisis States Research Centre London School of Economics.

http://www.unu.edu/pg
Peace and Governance Programme United Nations University.

http://www.prio.no
PRIO The International Peace Research Institute Oslo.

http://www.sfb governance.de
Research Center 700 (SFB) Free University of Berlin.

http://statebuilding.org
Sustainable Peacebuilding Network.

Developmentalism

M. Watts, UC Berkeley, Berkeley, CA, USA

Broadly construed, developmentalism is a wide ranging and complex philosophical position – underwritten by a panoply of development institutions, discourses, and practices – that sustains the idea of 'development' as a normative goal. It is a distinctively modern philosophy (a philosophy at once theoretical and applied) – guided by the desire to transform (to improve, to develop, to change through planned action) those parts of the world characterized by conditions of mass poverty – variously called the developing world, the Global South, the Third World. For much of the post 1945 period, the canonical vehicle for developmentalism is the 'developmental state' – the idea that developmentalism is best pursued through an interventionist, centralized, bureaucratic, and authoritarian 'state' apparatus that self consciously governs the market and shapes the process of capitalist accumulation, and by extension of class formation and class power. The so called 'neoliberal' counter revolution unleashed in the 1980s against the dominant Keynesian thinking has, however, directly challenged this notion of a *dirigiste* state in the pursuit of development. Whether developmentalism is in practice underwritten by a developmental state – as opposed to free market or civil institutions – is a historical, political, and ideological question. Developmentalism in this sense refers to a wider field of ideas (encompassing socialist and free market strategies and ideologies) and meanings, from human freedom to economic growth. At the current moment, the hegemonic form assumed by developmentalism as a philosophy is neoliberalism – itself a complex field of ideas that purports to offer a sort of unified field theory of the economy, the state, and political rights. Its most recent grand theoretician is Friedrich von Hayek, but the philosophical lineage is much deeper. At present, the struggle over neoliberal ideas and practices in development is seen vividly in the debate over foreign aid and planning; on the one side, a former World Bank economist for whom all forms of aid and planning are disastrous failures, and on the other, Columbia economist Jeffrey Sachs (and rock star Bono) for whom the meeting of the Millennium Development Goals for impoverished parts of Africa or South Asia will require a massive expansion of foreign assistance.

Meanings and Origins

To see developmentalism as a field of related philosophical positions is to immediately confront the word 'development', which is, after all, its root. Development is one of the most complex words in the English language. In his book *Keywords* published in 1976, Raymond Williams notes that the historically complex genealogy of development in Western thinking can "limit and confuse virtually any generalising description of the current world order," rather it is in the analysis of the "real practices subsumed by development that more specific recognitions are necessary and possible." The history of these real practices is, however, long and their meanings unstable and labile. Development came into the English language in the eighteenth century with its root sense of unfolding and was granted a new lease on life by the evolutionary ideas of the nineteenth century. As a consequence, development has rarely broken from organicist notions of growth or from a close affinity with teleological views of history, science, and progress in the West. By the end of the nineteenth century, for example, it was possible to talk of societies in a state of 'frozen development'.

There is another aspect to the genealogy, however, traced by Mike Cowen and Robert Shenton in their 1996 book *Doctrines of Development* to the eighteenth and nineteenth century notions of 'progress', and specifically to development as a sort of theological discourse set against the disorder and disjunctures of capitalist growth. Classical political economy, including the work of Adam Smith, David Ricardo, and Thomas Malthus, is suffused with the tensions between the desire for unfettered accumulation on the one hand, and unregulated desire as the origin of misery and vice on the other. Development in Victorian England emerged in part, then, as a cultural and theological response to progress. Christopher Lasch in his important 1990 text *The True and Only Heaven*, for example, has described a late nineteenth century obsessed by cultural instability and cataclysm. Saint Simon devoted himself in his last years to a new creed of Christianity to accompany his industrial and scientific vision of capitalist progress. Accordingly, trusteeship, mission, and faith were the nineteenth century touchstones of development.

There is a body of development theory (and of developmentalism), of much more recent provenance, dominated by the profession of economics, however, which began to emerge between the two world wars. This is a complex story to tell since this theory is not simply an invention of Western economics or political economy, but has been shaped and profoundly contributed to by various Third World intellectual traditions (e.g., Maoism from China, dependency theory from Latin America, and

Tier Mondism from Africa). To simplify an enormously complex field, one can identify four broad conventional streams of developmentalism understood as lineages of theory:

- that is broadly 'Keynesian or neo Keynesian' which emphasizes planning and the fundamental role of state interventions in market operations;
- that draws upon 'neoclassical economics' and the centrality of free markets and robust property rights;
- that is 'Marxist' and turns on the critical analysis of the processes of capitalist accumulation, competitions, and class power; and
- that is 'institutional' which encompasses a number of theoretical positions (rational choice and discursive practice) but takes as a starting point the political economy of development institutions, whether the World Bank, the Grameen Bank, corporate social re sponsibility, or trade unions.

Each of these four has complex genealogies and histories which privilege the state, the market, and civil society in different ways; each is, in addition, shaped historically by the conditions to which they are made to speak. The broad outlines of these 'conventional' development the ories have been reviewed in a number of texts. In all of these theoretical traditions there is no common ground as regards developmental as a goal, which may mean eco nomic growth measured through gross domestic product or human development; or freedom measured through income, life expectancy, and education; or of devel opmentalism as a set of philosophical beliefs (liberal, social democratic, or socialist).

Third World Developmentalism: From Interwar to Postwar Constructions

There is a sense in which Third World developmentalism – in its specific forms of state and multilateral policy harnessed to the tasks of championing economic growth, 'catching up', 'improvement', and producing governable subjects – is of quite a recent origin. It has been argued that development was invented or discovered by President Truman in his famous 1949 speech on 'fair dealing' and 'undeveloped areas'. The key conjuncture is the decade after the end of World War II, in spite of the fact that the Third World was seemingly of marginal concern in relation to the reconstruction of Europe, and the revivification of the world financial system. The Third World – the very origins of the term lie in the 1950s with the growing clamor of newly independent states to not to be drowned by the Cold War struggles – is forged in the crucible of in dependence movements and the collapsing of European empires, a global war (1939–45), and the hardening of geostrategic and ideological tensions between communism and capitalism. In short, the origins of development theory and practice as an academic and governmental enterprise, and of development economics as its 'hegemonic' ex pression, are inseparable from the process by which the 'colonial world' was reconfigured into a 'developing world' beginning in the 1930s, but especially in the aftermath of World War II, Africa, for example, became a serious object of planned development after the Great Depression of the 1930s.

The British Colonial Development and Welfare Act in 1940 and the French Investment Fund for Economic and Social Development in 1946 both represented responses to the crises and challenges which imperial powers confronted in Africa, providing a means by which they could negotiate the perils of independence movements on the one hand and a perpetuation of the colonial mission on the other. But the process by which the co lonial crisis produces "development as a historically singular process" (Escobar, 1995, p. 9) requires a sensi tivity to regions, political economy, and politics: global discourse and local power, and local discourse and global power, intersect and reproduce particular regional de velopment formations. Development was necessarily Eurocentric because its origins lay in the European ef forts to deal with the essential fact of capitalism, as Schumpeter put it, and the tragedy of underdevelopment and the paradoxical unity of modernity. In the same way that Karl Polanyi in *The Great Transformation* published in 1944 saw the later eighteenth century welfare debates in England as the discovery of the 'social', so the invention of development in the nineteenth and twentieth cen turies was about the 'failure' of free market capitalism. Within the belly of free market capitalist modernization, and its destructive effects, there lay the seed of devel opment, the notion that the destructive effects of the market would produce a reaction, a planned response to the costs of the unfettered operations of the market. Development and developmentalism, one might say, began as a form of protection – planning, self conscious directed improvement – against a liberal order.

In his important book *Encountering Development* pub lished in 1995, Arturo Escobar begins from the premise of development as a particular sort of social imaginary. Development, in his view, "relies on setting up the world as a *picture*, so that the whole system can be grasped in some orderly fashion as forming a structure or system" [emphasis added]. His book purports to show how this image came to be held as a sort of blueprint for a panoply of institutions, a diverse community of planners, polit icians, and bureaucrats, and for a battery of technologies harnessed for social engineering on a global scale. In short, Escobar portrays how the hegemonic vision of postwar development was institutionalized and with what consequences. Escobar necessarily starts with the fun damental position occupied by economics in the knowledge and practice of development, but posits the economy as an institution composed of production, power, and signification. The economic sphere in this rendering is as much cultural as material, and hence as much about the production of social order, truth, and forms of subjectivity as child mortality, gross national product (GNP), and high yielding varieties of rice and wheat.

Arturo Escobar finds modern development discourse to be the latest insidious chapter of the larger history of the expansion of Western reason. In his diagnosis, de velopment was an invention – more properly, a "histor ically produced discourse" – of the post 1945 era. This discourse is governed by the same principles as colonial discourse, but has its own regimes of truth and forms of representation. Development is about forms of know ledge, the power that regulates its practices, and the forms of subjectivity fostered by its impulses. Hegemonic development discourse appropriates societal practices and meanings into the modern realm of explicit calcu lation, thereby subjecting them to Western forms of power knowledge. It ensures the conformity of peoples to First World economic and cultural practices. Develop ment has in short penetrated, integrated, managed, and controlled large parts of the globe in increasingly per nicious and intractable ways. It has produced under development, a condition politically and economically manageable through normalization, the regulation of knowledges, and the moralization and technification of poverty and exploitation as political and material prob lems. The new space of the Third World, carved out of the vast surface of global societies, is a new field of power dominated by development sciences replete with their own truth claims and constructed subjectivity. The pol itical technologies of development practice which sought to erase underdevelopment from the face of the Earth have, to employ Escobar's language, converted a "dream into a nightmare." Development in this account plays a similar role with respect to the making of the Global South as does Orientalism in the production of the Middle East.

In keeping with a number of other Southern activists and hybrid intellectuals – perhaps most notable are those associated with the Delhi Center for Developing Societies, including Rajni Kothari, Ashish Nandy, and Shiv Vishva nathan – Escobar sees poverty as invented and globalized with the creation of a battery of transnational 'welfare' institutions at Bretton Woods, and in San Francisco fol lowing the signing of the United Nations charter. The discourse of national and international planning and de velopment agencies was able to constitute a reality by the way it was able to form systematically the objects of which it spoke, to group and arrange them in certain ways, and to give them a unity. Patriarchy, race, gender, and nationalism were all embraced in this discourse, at the same time that economists were privileged within its ranks. This rule governed system has remained unchanged at the level of practice, although the discursive formations have been unstable. In all of this modernity's objectifying regime of visuality turned people of the South into spectacles, and the panoptic gaze of development became an apparatus of social control. Institutions such as the World Bank thus embody what has been called the 'God trick' of seeing everything from nowhere. Development is constructed in large part through keywords – 'toxic words' – which really mean something else: 'planning' normalizes people; re sources desacralize nature; poverty is an invention; science is violence; basic needs are forms of cyborg; and so forth. In all of this, the space that came to be the Third World was converted into a population of normalized develop mental subjects. Developmentalism in this view is a part of a global project of governmentality.

Alternatives to Developmentalism ('Postdevelopmentalism')

There is a complex and differentiated body of work gradually emerging since the 1970s which represents an effort to imagine alternatives to developmentalism (a 'postdevelopmentalism'), which sees development (as practice) and developmentalism (as philosophy) as a ser iously flawed, many would say a disastrously failed, 'modernist' project. The intellectual and theoretical origins of this critique of what one might call conventional or mainstream development theory and practice – of both neoclassical economics, the new institutionalism and Marxist political economy – is complex, but sheds much light on both the imaginative geography of development, and its imaginary alternatives. In general, it needs to be said that the entire field of post or antidevelopment stands uneasily with respect to modernity, in some cases as a reactionary antimodernism, in some cases as antimodernist cultural relativism, and in others as a celebration of 'crit ical traditionalism'. But it is also curious that this flowering of imaginative geographies of development attached to antimodernist sentiment – of thinking about alternatives, of alterity, and of otherness – occurs precisely at the apogee of two variants of what might be called ultra or hypermodernism (or fundamentalist developmentalism perhaps). One is the forging of the Washington Consensus, the triumph of the neoclassical counter revolution: that is to say, the very idea in the wake of 1989 that, to employ the language of the World Bank, there are no alternatives to capitalism and free markets. The other is the Euro American vision of modernity, the object of the post development community's wrath, which is, in some quarters, seen to be unstable, differentiated, and internally reconstituting itself. The notions of 'reflexive modern ization' (of new and variegated modernities driven by industrial capitalism's ability to undercut itself, to generate life threatening risks) and 'post traditionalism' and 'post scarcity' (globalized forms of decentered autonomy, emotional democracy, and debureaucratization) posits another sort of postdevelopment at the capitalist core. In both of these 'alternatives' within the West, some of the attributes which so enrage the likes of Kothari, Alvares, or Escobar – developmental linearity, faith in progress, functional autonomization, instrumental rationality, and mass politics – are themselves seen to be under threat by the very dynamism and achievements of modernization itself.

The Genealogy of Postdevelopmentalism

What then is the particular genealogy of postdevelop mentalism thinking? The short answer is that it emerged from within the contradictions and failures of devel opmentalism itself.

The crisis of needs in the 1970s. One thread surfaces from the 1970s' disillusion with the world economic order and with the fact that the poorest 40% had somehow missed the boat. It is from this pessimism, disenchantment with growth, and trickle down that the Dag Hammarskjöld Foundation called for 'another development' geared to need satisfaction, self reliance, and endogeneity Bjorne Hettne's important book *Three Worlds of Development*, published in 1990, is the pivotal text in the reaction to failed modernization, and offers a rejection of the 'Eurocentric model' and a plea for egalitarian, partici patory ecological, self reliant, and ethnodevelopment studies.

The 'impasse' of Marxian political economy. Another thread emerges from the debates within Marxism, and the sense that Marxian political economy had entered an epistemological, practical, and theoretical cul de sac. In order to jettison its economism, reductionism, totalizing history, and class determinism – as it were the worst sins associated with the work of French structural Marxism (Louis Althusser in particular) – political economy had to encompass diversity, agency, local initiatives, and het erogeneity. Much of this work turns to networks, local institutions, and community institutions – that is to say, to the subaltern communities and knowledges invoked by Escobar – as a way of rethinking rural development.

Postcolonialism and the 'ur' history of development. Devel opment theory was driven in radically new directions by the post structural turn, and specifically by the twin themes of postcoloniality and discourse. In the latter, the 'ur' history of the doctrines of development are exca vated, and development texts examined as particular forms of text and storytelling. In the latter, post colonialism signals the proliferation of histories, tem poralities, and spatialities, and, to quote Stuart Hall, "the intrusion of difference and specificity into the general ising and Euro centric post enlightenment grand narra tives….[to emphasize] the multiplicity of lateral and decentered cultural connections, movements and migra tions which make up the world today" (1996: 248).

Rediscovering civil society and the new social movements. However, subject to hyberbole and exaggeration, the gradual 'democratization' of the post socialist bloc and the bureaucratic authoritarian states of the South un questionably provided a new breathing space for civil society – what has been called the "resurrection, re emergence and rebirth of civil society" (Cohen and Arato, 1992: 29). Indeed, the contraction of the state under (imposed) economic austerity created opportun ities for all manner of nongovernmental organizations (NGOs), social movements, and associational networks to flourish as never before. It is this renewed concern with civil society, with communities, popular movements, and social networks, which represents in some quarters the prospect of a new way of doing politics, and the possibility of alternative ('grass roots', 'participatory', 'subaltern') visions of development outside of the conventional horizon of both state and market.

Indigenous knowledge and global threats. Ecological con cerns pushed the antidevelopment agenda in two senses. First, the environmental toll of rapid industrialization in the newly industrializing countries (NICs) and within the socialist bloc, coupled with potentially catastrophic forms of 'global' degradation (ozone depletion and global warming), fueled new concerns over the sustainability of mainstream development. Second, the growing theore tical interest within political ecology and the repositories of local knowledge (e.g., forms of 'peasant science' as alternatives to 'Western expertise' documented by Paul Richards' foundation text *Indigenous Agricultural Revolution* published in 1985) reaffirmed the salience of subaltern knowledge and indigenous local practice, typically em bodied in the proliferation of all manner of Southern green movements, as a counterweight to big science and multilateral greening of the World Bank sort.

Globalization and its malcontents. Finally, the rapidity with which the neoliberal orthodoxy was adopted, through a mixture of coercion and consent, throughout the Second and Third Worlds (keywords which themselves became increasingly inadequate as ways of classifying geographical space), inevitably triggered a counter reaction against the disruptive social, economic, and political consequences of 'globalization' (the dominance of market forces, the inte gration of the global economy, the genesis of worldwide consumerism and media integration, and the transforma tion of production and labor markets). To take an example, the United Nations Research Institute for Social Devel opment (UNRISD) refer to the 'states of disarray' pro duced by the 'painful adjustment' of market integration and globalization. The devastating costs in Africa of the structural adjustment programs of the 1980s, for example – and the global proliferation of a vast slum world of poverty and immiseration – provided a strong antidote to the tri umphalist thinking of free market globalization.

Postdevelopmentalism can be understood in part, then, as a reaction to and a product of a particular congeries of theoretical development (knowledge production) and pol itical economic transformations. The *évenements* of 1989, the unprecedented dislocative effects of globalization and market integration, a deepening sense of global ecological 'crisis', the flowering of civil society and social movements intersected with what one might call in its panoply of guises. There is of course a polyphony of voices within this postdevelopmentalist community: Vandana Shiva, Wolf gang Sachs, Arturo Escobar, Gustavo Esteva, and Ashish Nandy, for example, occupy quite different intellectual and political locations. But it is striking how intellectuals, ac tivists, practitioners, and academics within this diverse community participated in a global debate. The journal *Alternatives* is in some sense its institutionalized mouth piece, and some sites (in India and Mexico, for example) have emerged as its intellectual flagships (the Delhi Center has even been referred to as the 'new Frankfurt School').

Debating Alternatives

Debates over postdevelopmentalism have opened up critical avenues for the critical exploration of devel opmentalism as a philosophy and set of practices. On the one hand, it is important to examine the *longue duree* of development and to take seriously the ways in which what passes as development knowledge and practice are insti tutionalized and with what effects. The language of development is not neutral, and nor are its institutions. To deconstruct sustainability as one powerful aspect of developmentalism – to interrogate the assumptions of particular institutionalized visions of green development – is useful on many counts. To see how and with what consequences 'ecological economics' has been employed by the Global Environment Facility might shed much light on the sorts of projects funded by the World Bank and the consequences of specific types of large multilateral programs. Conducting ethnographies of development institutions, as Ferguson effectively does in *Anti Politics Machine* published in 1990, his account of a rural devel opment project in Botswana is singularly helpful in understanding how particular places and problems are constructed and legitimated by experts and managers.

Some of these new critical approaches to develop ment ironically have similarities with earlier forms of radical development thinking, notably 1960s' 'dependency' theory , a perspective largely discredited on the grounds of its simplistic theory of power and a crude sense of political economy. Like the Latin American *dependentistas*, it confers enormous power on an external world system, privileges local autonomy and cultural identity, and sees sovereignty as a central plank of its own vision of development. Like dependency theory, the nature of external power is often crudely articulated in bold outline – which in Escobar's case is ironic insofar as he employs Foucault to express an alternative theory of capillary power. Escobar's account of the World Bank as an instrument of modernity's power, for example, reads like Third World nationalism of old. No attempt is made to lay out the complex internal divisions within the bank or the important reversals made by the bank around some environmental and dam projects, or (on the basis of data he himself provides) the relative in significance of World Bank resources in relation to other capital flows to the South. Indeed, the antidevelopment communities' focus on keywords and the representation of the Third World could be pulled from a cursory reading of Che Guevara's meditation on the condition of under development (the Third World as a 'stunted child') penned at least 30 years ago.

Developmentalism and the Rise of Neoliberalism

The reality is, of course, that postdevelopmentalism as critique has arisen against the backdrop of a powerful resurgence – a counter revolution – of a particular form of developmentalism, namely, neoliberalism and free market capitalism as development. David Harvey in *A Brief History of Neoliberalism* published in 2005 attempts to map the dismantling of the social democratic world of embedded liberalism. Harvey aims to describe and ex plain how it happened that capital's compromise was supplanted by the regime of neoliberalism, which has "not only restored power to a narrowly defined capitalist class…[but] also produced immense concentrations of corporate power in energy, the media, pharmarceuticals, transportation, and even retailing." The results, tabulated in the book, have been strikingly disadvantageous, and even disastrous, for a large majority of the world's population, not to mention planetary ecology, and this despite the promises made by the salesforce of neoli beralism. Harvey returns again and again to the notion of a 'contradiction' between neoliberalism (as a form of developmentalism) and the actual practice of neoliber alization on the ground, which always seems to fall short of market utopian, textbook, and ideals (as in 'rising tide' or 'trickle down').

Confronted with neoliberalism's atrocious record, Harvey understandably asks: How is it, then, that the rest of us have so easily acquiesced in this form of devel opmentalism? This suggests that neoliberalism should be approached as an ideology. Any adequate account of its rise to hegemony should not assume a pure body of neoliberal theory 'lurking in the wings' of history but from the contested discursive field of developmentalism itself, among whose keywords are 'freedom', 'market', and 'private property'. The rise of neoliberal developmentalism is the rise of an unplanned form of development – meaning capitalism. To this extent its great arc is the historic victory of the idea that Friedrich von Hayek laid out in *The Road to Serfdom* published in 1944. Hayek, a native of Vienna who trained as an economist at the feet of Ludwig von Mises but is forever associated with a largely noneconomic corpus produced at the London School of Economics (LSE) and the universities of Chicago and Freiburg between 1940 and 1980, is widely recognized as the intellectual architect of the neoliberal counter revolution. It was Margaret Thatcher after all who pronounced, at a Tory Cabinet meeting, "This is what we believe," slamming a copy of Hayek's *The Constitution of Liberty* onto the table at 10 Downing Street. His critique of collectivism – that it destroys morals, personal freedom, and responsibility, impedes the production of wealth, and sooner or later leads to totalitarianism – is the 'ur' text for market utopians. Collectivism was by definition a made rather than a grown order, that is, a 'taxis' rather than a 'cosmos'. Collectivism was, Hayek said, constructivist rather than evolutionary, organized not spontaneous, an economy rather than a 'catallaxy', coerced and concrete rather than free and abstract. Its fatal conceit was that socialism (and social democracy for that matter) admitted the "reckless trespass of *taxis* onto the proper ground of *cosmos*" (Anderson, 2005: 16).

The other half of Hayek's project was an apologia for Western civilization, conceived of as liberty, science, and the spontaneous orders that co evolved to form modern society ('great society' as he termed it). It is a defense of the liberal (unplanned) market order from which the preconditions of civilization – competition and experimentation – had emerged. Hayek, like Max Weber, saw this modern world as an iron cage constituted by impersonality, a loss of community, individualism, and personal responsibility. But contra Weber, these structures, properly understood, were the very expressions of liberty. From the vantage point of the 1940s, this (classical) liberal project was, as Hayek saw it, under threat. Indeed, what passed for 'liberalism' was a travesty, a distorted body of ideas warped by constructivist rationalism, as opposed to what he called 'evolutionary rationalism'. The distance between 'actually existing' liberalism and Keynesianism was, on the Hayekian account, disastrously slight. What was necessary, as he made clear at the foundation of the Mont Pelerin Society in 1947, was a grand purging, a restoration of true liberalism by way of the removal of 'accretions'. There was to be no compromise with collectivism; all the ground ceded to creeping socialism had to be regained.

In his writing and his promotion of think tanks, such as the Institute of Economic Affairs in Britain – a brains trust for the likes of Keith Joseph and Margaret Thatcher – Hayek aggressively launched a cold war of ideas. He was one of the quartet of European theorists (Carl Schmitt, Leo Strauss, and Michael Oakeshott were the others) whose ideas, while in a tense relationship to one another, have come to shape a large swath of the intellectual landscape of the early twenty first century. Hayek was not in any simple sense a conservative or libertarian; nor a voice for *laissez faire* ('false rationalism' as he saw it). He identified himself in the individualist tradition of Hume, Smith, Burke, and Menger and provided a bridge which linked his short term allies (conservatives and libertarians) to classical liberals in order to make common cause against collectivism. To roll back the incursions of *taxis* required a redesign of the state. A powerful chamber was to serve as guardian of the rule of law (striking all under the age of 45 off the voting rolls, for example) to protect liberty against popular sovereignty.

The Long March from Mont Pelerin to the collapse of the Berlin Wall took about 40 years, by way of the 'Chicago boys' (a cohort of monetarist economists trained at or associated with the University of Chicago) in Chile, the World Bank–International Monetary Fund complex, and the Reagan–Thatcher dispensations. As Antonio Gramsci might have put it, there has been a Hayekian 'passive revolution' from above. We have witnessed what the Left's great pessimist Perry Anderson has dubbed a 'neoliberal grand slam'. The 'fluent vision' of the Right, he notes, has no equivalent on the Left; embedded liberalism (let alone something called 'socialism') is now, we are told, as remote as 'Arian bishops'. In this account, neoliberalism rules undivided across the globe and is the most successful ideology in world history.

The process by which neoliberal hegemony was established, and its relation to forms and modes and sites of resistance, remains a story for which, even with Harvey's synoptic survey at hand, we still have no full genealogy. The cast of characters may be lined up – from the school of Austrian economics to the Reagan–Thatcher–Kohl troika – but this explains very little, or rather it poses more questions than answers. Neoliberalism was a class reaction to the crisis of the 1970s (Harvey (2005) talks of a "restoration of class power"): on that much Milton Freidman and David Harvey might agree. 'Transnational capital' and the Wall Street–Treasury nexus imposed brutal forms of economic discipline –'structural adjustment', in the argot of neoclassical economics – to eradicate forever any residue of collectivism in the Third World. But beyond such descriptions, which Harvey lays out clearly enough, we are still left with many paradoxes and puzzles.

Why, for example, did the LSE and Chicago – once the respective centers of Fabianism and a certain version of (American) liberalism under Robert Hutchins – become the forcing houses of neoliberalism? What were the facilitating conditions that fostered the arrival of the maverick Ronald Coase in Chicago, marking a neoliberal turning point? How did the Chicago boys come to occupy the commanding heights in post Allende Chile and how did they live down the fact that their effects were, to use the language of the World Bank, 'disastrous'? How did the World Bank – a bastion of postwar development economics and, it must be said, of statism – become the voice of *laissez faire*? Harry Johnson, who held chairs at the LSE and Chicago, is a spectral figure in the liberalization of the World Bank, but how can we explain the capture of key sectors of the bank (often by second rate economists) against a backdrop of robust Keynesianism? How can we grasp the fact that 'shock therapy' in Eastern Europe was more the product of the enthusiastic Hungarian reformers than of the more reticent American neoliberal apparatchiks?

It is sometimes noted that the 1991 World Development Report (shaped by former US Secretary of the Treasury Lawrence Summers) marked a neoliberal watershed in its refiguring of the role of the state. But was not the foundational moment a decade earlier with the 1980 Berg report on Africa, named after Elliot Berg, a Michigan economist to whom it seems nobody paid much attention for 20 years? It was Africa (not Latin America or Eastern Europe) that proved to be the first testing ground of neoliberalism's assault on the overextended public sector, on physical capital formation, and on the proliferation of 'market' distortions by government. Understanding the rise of neoliberal developmentalism – of a market utopian vision – is one of the most important projects confronting contemporary development.

Developmentalism and Development Geography

In its approach to the relations between space, environment, and development, development geography as a subfield has inevitably been shaped by the broader theoretical and discursive turns and trends within development thinking over the past half century. On the larger canvas of the rise of free market utopian thinking and the so called TINA worldview ('there is no alternative'), much of critical development geography of late has been animated by the need to both think about alternatives – for example, the new peasants movements such as MST in Brazil or the contributions of indigenous movements to citizenship debates – and to investigate the ways in which neoliberal practice (contemporary developmentalism in its various forms) have been formulated, circulated, normalized, rendered persuasive, and installed in particular institutions (say the World Bank) or discourses (say 'ecological economics' or micro finance). In general, these sorts of approaches have enervated the field rather than signaled its 'death by developmentalism'. In particular, both the search for alternatives and the explorations of neoliberal developmentalism as a form of rule (implying forms of expertise, forms of subjectivity, and forms of identity) have compelled geographers to deepen their ethnographic understandings of how institutions produce forms of knowledge, how such knowledge does or does not become hegemonic, and how counter hegemonic movements and practices can emerge from the complex articulations of the local and the global. Looming in the background, of course, is the rise of two new developmental giants – China and India – characterized by enormously different trajectories and political economics but both pursuing outwardly oriented neoliberal strategies against a historical backdrop of differing forms of commitment to Third World socialism, colored in different ways by 'Asian characteristics'. In the future, these two mega developmental states will draw much geographical scrutiny.

See also: Aid; Civil Society; Imperialism, Cultural; Green Revolution; Indigenous Knowledges; Neoliberalism; Postcolonialism/Postcolonial Geographies.

Further Reading

Alvarez, C. (1992). *Science, Development and Violence*. New Delhi: Oxford University Press.

Anderson, P. (2002). Internationalism: A breviary. *New Left Review* 14, 5 25.

Anderson, P. (2005). *Spectrum*. London: Verso.

Beck, U. (1994). The reinvention of politics. In Beck, U., Giddens, A. & Lash, S. (eds.) *Reflexive Modernization*, pp 1 55. Cambridge: Polity.

Berman, M. (1982). *All That Is Solid Melts into Air: The Experience of Modernity*. New York: Penguin.

Booth, D. (ed.) (1994). *Rethinking Social Development*. London: Methuen.

Cardoso, H. and Faletto, E. (1979). *Dependency and Development in Latin America*. Berkeley, CA: University of California Press.

Chari, S. (2005). *Fraternal Capital*. Stanford, CA: Stanford University Press.

Cohen, J. and Arato, A. (1992). *Civil Society and Political Theory*. Cambridge, MA: MIT Press.

Cooper, F. and Packard, R. (eds.) (1998). *International Development and Social Sciences*. Berkeley, CA: University of California Press.

Corbridge, S. and Harriss, J. (2000). *Reinventing India*. London: Polity.

Corbridge, S., Williams, G., Srivastava, M. and Véron, R. (2006). *Seeing the State*. Cambridge: Cambridge University Press.

Cowen, M. and Shenton, R. (1996). *Doctrine of Development*. London: Routledge.

Crush, J. (ed.) (1995). *The Power of Development*. London: Routledge.

Cypher, J. and Dietz, J. (1997). *The Process of Economic Development*. London: Routledge.

Dallmayr, F. (1996). Global development? *Alternatives* 21, 259 282.

Davis, M. (2005). *Planet of Slums*. London: Verso.

Easterly, W. (2007). *White Man's Burden*. New York: Penguin.

Escobar, A. (1995). *Encountering Development*. Princeton, NJ: Princeton University Press.

Esteva, G. (1992). Development. In Sachs, W. (ed.) *The Development Dictionary: A Guide to Knowledge as Power*, pp 6 25. London: Zed Books.

Ferguson, J. (1990). *The Anti Politics Machine*. Cambridge: Cambridge University Press.
Fox, J. (1995). Governance and development in Mexico. *Journal of Development Studies* 34, 610 644.
Giddens, A. (1994). Living in a post traditional society. In Beck, U., Giddens, A. & Lash, S. (eds.) *Reflexive Modernization*, pp 56 109. Cambridge: Polity.
Goldmann, M. (2005). *Imperial Nature*. New Haven, CT: Yale University Press.
Gupta, A. (1998). *Postcolonial Development*. Princeton, NJ: Princeton University Press.
Hall, S. (1996). When was the post colonial? In Chambers, I. & Curti, L. (eds.) *The Post Colonial Question*, pp 242 260. London: Routledge.
Harraway, D. (1991). *Simians, Cyborgs and Women*. London: Routledge.
Harvey, D. (2005). *A Brief History of Neoliberalism*. London: Clarendon.
Hayek, F. (1944). *The Road to Serfdom*. Chicago, IL: University of Chicago Press.
Herbert, C. (1990). *Culture and Anomie*. Chicago, IL: University of Chicago Press.
Hettne, B. (1985). *Three Worlds of Development*. London: Methuen.
Hoover, K. (2003). *Economics as Ideology*. Boulder, CO: Rowman and Littlefield.
Hsing, Y. T. (1998). *Making Chinese Capitalism*. Singapore: Oxford University Press.
Kohli, A. (2005). *State Directed Development*. Cambridge: Cambridge University Press.
Kothari, R. (1989). *Rethinking Development*. New York: Horizons.
Lasch, C. (1991). *The True and Only Heaven*. London: Norton.
Leach, M. and Fairhead, J. (eds.) (2003). *Science, Society, Power*. Cambridge: Cambridge University Press.
Marglin, F. and Marglin, S. (eds.) (1990). *Dominating Knowledge*. New Delhi: Oxford University Press.
Mitchell, T. (2005). *The Rule of Experts*. Berkeley, CA: University of California Press.
Mosse, D. (2005). *Cultivating Development*. London: Pluto.
Nandy, A. (1987). Cultural frames for social transformation. *Alternatives* 12(1), 101 123.
Nandy, A. (1998). Culture, state and the rediscovery of Irdion politics. *Interculture* 21, 2 17.
Parajuli, P. (1991). Power and knowledge in development discourse. *International Social Science Journal* 127, 173 190.
Peet, R. and Watts, M. (1995). Introduction. In Peet, R. & Watts, M. (eds.) *Liberation Ecologies*. London: Routledge.
Pieterse, J. (1996). *My Paradigm or Yours?* Working Paper #229. The Hague: Institute of Social Studies.
Pottier, J. B. A. and Sillitoe, P. (eds.) (2003). *Negotiating Local Knowledge*. London: Routledge.
Polyani, K. (1944). *The Great Transformation*. Boston, MA: Beacon.
Prasad, M. (2006). *The Politics of Markets*. Chicago, IL: University of Chicago Press.
Preston, D. (1996). *Development Theory*. Oxford: Blackwell.
Radcliffe, S. (ed.) (2005). *Culture and Development in a Globalizing World*. London: Routledge.
Rahnema, M. and Bawtree, V. (eds.) (1997). *The Post Development Reader*. London: Zed Press.
Richards, P. (1985). *African Agricultural Revolution*. London: Methuen.
Roe, E. (1991). Development narratives. *World Development* 19, 287 300.
Routledge, P. (1994). *Resisting and Shaping the Modern*. London: Routledge.
Sachs, J. (2006). *The End of Poverty*. New York: Penguin.
Sachs, W. (ed.) (1992). *The Development Dictionary*. London: Zed Books.
Said, E. (1979). *Orientalism*. New York: Harper.
Sen, A. (2002). *Development as Freedom*. New York: Knopf.
Shiva, V. (1989). *Staying Alive*. London: Zed Books.
Shiva, V. (1993). The greening of the global reach. In Sachs, W. (ed.) *Global Ecology: A New Arena of Political Conflict*, pp 53 60. London: Zed Books.
UNDP (1996). *Human Development Report*. London: Oxford University Press.
UNDP (2006). *Human Development Report*. London: Oxford University Press.
UNRISD (1995). *States of Disarray*. Geneva: United Nations Research Institute for Social Development.
Wade, R. (1994). *Governing the Market*. Princeton, NJ: Princeton University Press.
Watts, M. (1995). A new deal for the emotions. In Crush, J. (ed.) *The Power of Development*, pp 44 62. London: Routledge.
Williams, R. (1976). *Keywords*. Oxford: Oxford University Press.
Wolford, W. (2005). Edible ideology? *Geographical Review* 86, 475 491.
World Bank (2006). *Development Report*. Washington, DC: The World Bank.
Wright, A. and Wolford, W. Q. (2003). *To Inherit the Earth*. Oakland, CA: Food First.

Relevant Websites

http://www.cfda.ac.in/
Center for Development Alternatives.
http://www.cgdev.org/
Center for Global Development.
http://www.globalexchange.org/
Global Exchange.
http://www.worldbank.org/
The World Bank.
http://www.nadir.org/nadir/initiativ/agp/free/wsf/
World Social Forum.

Devolution

M. Keating, European University Institute, Florence, Italy

Glossary

Competences The policy areas in which a given level of government can act.
Corporatism Government by concertation between the state and organized interests.
Devolution The delegation of power to local and regional levels of government.
Federalism A system of government with two territorial levels, each of which has constitutionally guaranteed powers.
Meso-Level The level of government and administration between the central state and municipal government.
Regional Policies State policies to reduce interregional inequalities.

What Is Devolution?

Devolution is defined in the *Oxford English Reference Dictionary* as "the delegation of power, especially by a central government to local or regional administration." This encapsulates both the meaning and the difficulty of the phenomenon, since it implies both the transfer of power downwards and the retention of ultimate control at the center. It is for this reason that it has become a means for unitary states to recognize their own territorial diversity and to relieve pressure on the central state, while avoiding federalization, that is, the formal division of sovereign authority, with the limits that this would impose on the center. Devolution has a long history in the United Kingdom, where it was canvassed as a way of reconciling unity with diver sity from the late nineteenth century, although, apart from the special case of Northern Ireland between 1922 and 1972, it was not implemented until 1997–98 when autonomous assemblies were established for Scot land, Wales, and Northern Ireland. Even then, the program ran into the sands in 2004 when the voters in Northeast England rejected proposals for a regional assembly. From the late 1990s, the term appeared in Italy, much favored by the Lega Nord and was incorporated into the program of the 2001–06 Berlusconi government, albeit in a set ;of rather incoherent proposals designed to satisfy different wings of the governing coalition.

Why Devolution?

The concept of devolution, if not the term, however, has wider application and we can usefully apply it to the rise of regional or 'meso' level governments since the 1960s in all the large European countries and some of the smaller ones. There are several reasons for the growth of this intermediate level of administration, policymaking, and regulation in the modern European state. One set is functional, the rise of the region as a key level for the understanding and management of economic and social change. From the 1960s, European states engaged in large scale planning and modernization projects requiring a level above the municipalities, provinces, and counties. Regional policies, the spatial complement to national Keynesianism, aimed to reduce spatial dis parities in development and bring into use idle resources in the lagging regions. Welfare state development and planning often also had a regional focus. Such policies were pursued by national governments, usually by national instruments such as the French *Délégation à l'aménagement du territoire et à l'action régionale* (DATAR) or the Italian *Cassa per il Mezzogiorno* and should not in themselves be confused with devolution, although, as we will see, they did have devolutionary implications. From the 1980s, the European Commission started to take over the regional development policy role through the ex panded and reorganized structural funds, incorporating the older European Regional Development Fund. Again, we must guard against the common tendency to confuse these regional policy initiatives with pressures for decentralization, but there were political implications.

A second influence is that of state modernization and rationalization. Governments from the 1970s sought to increase their steering capacity while reducing their need for direct intervention and administration. Good examples are France and Italy, where the central state had become overburdened with local and regional demands and sought to recover its own freedom of action by offloading detailed management and administrative decisions from the center. In France, there was an effort to separate local and regional politics, to reduce the accumulation of mandates whereby national politicians simultaneously held local office, and to cut down on the reference of decisions to Paris. In Italy, it was hoped to reduce clientelism and corruption, although the old patterns soon reestablished themselves at the regional level.

A third influence is democratization and demands from new social forces. In states like France and Italy, the 'Jacobin' tradition associated democracy and equality with centralization and uniformity. By the 1970s, this was coming under question as new social movements re engaged with the local; the social democratic left adopted many of their themes and democracy was linked again with decentralization and local autonomy. Protests against centralized regional policies and their neglect of local and regional concerns crystallized into demands for constitutional change and more sensitive development programs. In Spain, the transition from a centralized dictatorship was closely linked to decentralization as had been the case in Italy earlier. In Poland, decentralization and regional government were part of the democratic agenda for some, but were contested by Jacobin forces. Here, as elsewhere, support for devolution often depended on whether a party was in government or opposition. There was also concern in many countries that the growth of a regional state in the form of agencies and boards was creating a democratic deficit.

Fourth, there is the question of the management of national diversity. Immediately after World War II, Italy established four (later five) special status regions in the northern border areas and the islands. From the 1960s, Belgium moved in steps from being a unitary and centralized state, managed by consociational accommodation at the center, to a regionalized state, and then a federation. Spain had experimented with autonomy for its historic nationalities under the Second Republic in the 1930s and recognition of their rights was, by the 1970s, an integral part of the democratic reform agenda. The United Kingdom had struggled over how to manage its national diversity since the nineteenth century and the issue has recurred constantly since. Central and East European countries have been very reluctant to go down this road because of the political sensitivity of the issues.

Devolution Models

Devolution is not a single model of government that is then applied to individual states. It does not, unlike federalism, have a set of core features and a well-developed doctrine. Indeed the term itself is used inconsistently and sporadically. It should be seen, rather, as a practice, a set of instruments used by governments, but in which we can detect some underlying patterns.

The weakest form of devolution is 'administrative devolution', as it was known in the United Kingdom, which bears a resemblance to the French concept of deconcentration. In this case, regional branches of central government are given a degree of discretion in the implementation of policy and in coordinating its various strands. Ireland was long governed by a chief secretary who was a member of the British government, although not usually an Irish MP. In 1885, Scotland gained a secretary (of state) who was by convention always a Scottish MP as well as a member of the British cabinet. The arrangement was extended to Wales in 1965 although later Conservative governments did not have enough Welsh MPs to staff it and the secretary of state often sat for an English seat. In France, regional prefects were appointed in 1963 to coordinate national policy at the regional level, including the work of the long-established prefects in the *départements*. Successive governments repeatedly sought to enhance the role of prefects by strengthening their control over the field administration of government departments and giving them more power to make decisions on their own; but local interests constantly defeated this and made their way back to Paris. More recently, the French prefectoral model has been extended to Poland, where the *voivod* is a central official operating at the regional level.

A second model is that of 'functional devolution', that is, the transfer of specific tasks to regional agencies, with the remit of complementing and implementing national policy. An early example of this was in France, where the *Commissions de développement économique régional* (CODERs) were set up in 1963 to assist in regional planning. These were corporatist bodies, comprising business, agriculture, trade union, and independent members (50%), central government nominees (25%), and local councillors (25%) and intended to communicate with local civil society, notably the *forces vives*, the dynamic and modernizing elements in the regions. In 1972, following de Gaulle's failed referendum on regional and senate reform, these were replaced by indirectly elected regional councils drawn from national and local politicians. Alongside these regional councils were economic and social councils representing social and economic groups. Similar arrangements existed briefly in Italy and Belgium, and in 1964, Great Britain introduced Regional Economic Planning Councils consisting of local councillors, central government nominees, and representatives of social and economic interests. These were complemented by Regional Economic Planning Boards of representatives from central government departments in the regions. In the late 1990s, a similar system of functional devolution was reintroduced for England, with government offices for the regions, regional development agencies, and regional chambers (or assemblies) drawn from local government and socioeconomic actors. Hungary also adopted a model of development agencies alongside indirectly elected councils to manage development in the 1990s.

Such corporatist models of functional devolution have proved unstable and have either failed, with a lapse back to centralization, or evolved into elected regional government. They have to combine a 'top down' role, of carrying out central policy in the regions, with a 'bottom up' one, of providing a local input and direction. Their legitimacy, given their composite membership, is questionable, making it difficult to take hard decisions. They were born of an age in which development policy could be presented as a technical matter, a search for the right solution, and thus effectively depoliticized. In practice, any development strategy will produce winners and losers and rapidly be repoliticized. Indeed, it seems that regional planning and regional policies, while very different in principle from regional government, do reframe political issues in regional shape and encourage mobilization around regional themes. Yet the corporatist model does not provide any obvious ways for the resulting politics to be managed. Members of the corporatist bodies in consequence rapidly become disillusioned and often resign. In Italy and Belgium, indirectly chosen regional bodies rapidly gave way to elected regional government. In the former case, this had been foreseen in the post war constitution but postponed for over 20 years for political reasons. In Belgium elected regional governments, alongside devolution to language communities, was a means of managing the twin problems of economic discontent and language conflict. French regional government took a slightly different turn in the 1970s. President de Gaulle's regional government proposals in 1969 had aimed, among other things, to undermine the power of the old elites (*notables*) in the local political system by bringing in an elected regional level and reforming the Senate, which was elected from the localities. Failure in the referendum led to de Gaulle's resignation and his successor, Georges Pompidou, instead entrusted the regional level to precisely these *notables*, by having it indirectly elected from local government. The move to direct elections came only in the 1980s with the arrival in power of the Socialists, who had built up their local and regional power bases. Even they, however, postponed direct election until the very end of their term, in 1986. Hungary has proposed to move from corporatist to elected regional government, but has long postponed the transition. Britain's experiment in functional/corporatist regionalism ended with the failure of the National Plan in 1967 although the Regional Economic Planning Councils persisted until 1979 when they were abolished in the English regions by the incoming Thatcher government. The functional need for a regional tier of administration, however, persisted and in the 1990s the Conservative government established government offices for the regions. After 1997 the Labour government set up regional development agencies and regional chambers (which now call themselves assemblies).

The final model of devolution is the elected multifunctional government at the meso level. In principle, Italian regions are a tier of general purpose government but in practice they have been confined to rather narrow functional roles, notably in the administration of the health service. French regional councils, while directly elected, are still functionally restricted, their main tasks being planning and investment programming. Belgian regional governments were stronger and had taken on more tasks in successive rounds of constitutional reform, until Belgium was converted officially into a federation in 1993. In Flanders, where the regional and community institutions have been largely merged, this created a strong tier of general purpose government, but in Brussels and Wallonia there are weaker institutions. The Spanish autonomous communities are strong general purpose governments, which have increased their responsibilities over time, to the extent that many observers see Spain as a federation in the making. Scotland and Northern Ireland have strong elected governments with powers over everything that it is not specifically reserved to the center, while Wales has an assembly with administrative control over a defined range of functions. Poland has a directly elected regional tier of government, although its powers remain relatively weak.

One advantage that devolution possesses over classic federalism is that it is much easier to design asymmetrical devolution, given that, at least in principle, the central state remains in charge of the whole. It thus enables governments to respond in cases where there are autonomist demands in one part of the state, but where other parts prefer centralization. This is borne out by these cases. The United Kingdom has a highly asymmetrical constitution, with different autonomist settlements for Scotland, Northern Ireland, and Wales, while England, accounting for some 85% of the population, is governed from the center. Italy has special status for its five peripheral regions, without the need to generalize the model. Belgium was able to make rather different arrangements for Flanders, Brussels, Wallonia, and the German community, although its conversion into a federation has complicated matters. Even France has been prepared to make special concessions to Corsica, although not to extend these to regions of the mainland. In Spain, there is a constant tension between the historic nationalities, who insist on their distinct status, and the center, which has sought to homogenize the statutes of autonomy.

Powers and Competences

The allocation of powers in devolved systems varies greatly. In Scotland and Northern Ireland, everything is devolved except that which is specifically reserved. In principle, this is supposed to be the case in Belgium but the need to indicate whether it is the regions or the language communities that get the residual powers means that these need to be specified. Spain has a rather confusing system in which the constitution lists competences that must be devolved, those that must be kept at the center, and those that may be devolved by individual statutes of autonomy. There is also provision, however, for devolving reserved powers. French, Italian, and Polish regions have those powers that are specifically devolved to them. In principle, governments try to devolve clear blocks of competences and avoid overlapping central and regional responsibilities, but in practice the nature of modern government makes this impossible. Even should a clear separation of tasks be agreed upon at one point, the shifting nature of government will soon make it redundant. In any case, governments have rarely been willing to devolve whole areas of public policy. This has been the cause of considerable confusion, when politicians in Italy and Spain have talked of giving regions 'exclusive responsibility' for various functions. In practice this usually means the whole management of the field, but within legislative parameters set by the central state. Only in Scotland and Northern Ireland does exclusive responsibility mean that devolved governments can repeal existing state laws in the relevant fields. Scotland, Northern Ireland, and Wales are also unusual in that central ministers and departments cannot act or spend money within the devolved territories except in pursuance of a reserved competence.

The process of devolution and the entrenchment of devolved powers also varies. In Belgium, competences can be transferred by act of the central parliament, but cannot then be taken back. In Spain, statutes of autonomy were drawn up in the late 1970s and early 1980s for the 17 regions, within the parameters of the 1978 constitution but each statute is slightly different. Initially there was a fast track, aimed at accommodating the three historic nationalities (Catalonia, the Basque Country, and Galicia) but also used by Andalucia, giving large competences from the outset. Other autonomous communities acquired their powers more slowly, although the intention of the central government was that eventually they would all reach the same level. Autonomous communities can propose amendments to their own statutes and after the change in the Spanish government in 2004 a number started the process of framing new statutes. Once devolved, competences cannot be taken back by the center. French regions were not even recognized by the constitution until 2004 and their powers are transferred by ordinary law. Italian regions have powers bestowed on them by the law and the constitution and since the 1990s, there has been a long debate on giving them stronger and entrenched powers, but political divisions have meant that little has been produced. The United Kingdom represents both ends of the devolution spectrum, in that the powers devolved are very extensive, but are subject to ordinary law and parliamentary supremacy so that, in theory, Westminster can override their laws and even abolish their institutions. In practice, a convention has been established for Scotland (the Sewel Convention) that Westminster will legislate on devolved matters only at the specific request of the Scottish Parliament. This device is used frequently for technical reasons (to do with the overlap of functions) and sometimes to enable the Scottish government to offload difficult decisions.

Of course, formal constitutional guarantees to devolved governments are only one part of the story. Equally important is their political entrenchment. The Scottish parliament is, according to the Westminster doctrine, the creation of the UK parliament, which retains full sovereignty and could abolish it at will. Scottish political and public opinion, however, sees it as the product of an exercise in self determination and the representative of the Scottish nation. Any government trying to abolish it would be faced with the prospect of Scotland seceding from the United Kingdom altogether. The theoretical possibility for the Spanish parliament to override autonomous institutions, as some central politicians have threatened in relation to the Basque Country, is subject to more legal control, requiring an absolute majority in the senate and grave conditions (which would presumably be open to review by the constitutional court). More importantly, as in the UK, it would produce a political crisis and must be considered highly unlikely. Recentralizing competences in France, Italy, or Poland, on the other hand, would be much less difficult.

Finance has been one of the most difficult areas in devolution. True to the original devolution concept, governments have tended to try and decentralize power while retaining overall control. Initially, this usually means that they have retained taxing power at the center and transferred funds to the regions for specific purposes. This restricts the discretion of devolved governments and may make them little more than administrative agents of the center, as was the case for a long time in Italy and France. Over time, however, governments have seen the benefits of converting transfers into block grants, so giving regions freedom to allocate resources in accordance with needs and (an important political consideration) forcing them to take responsibility for rationing. More recently, governments have been willing to decentralize tax raising powers. Again the arguments are partly to do with efficiency but more perhaps with politics, the consideration that the regions have enjoyed the privilege of spending money while the center has had the pain of raising it. So the Spanish autonomous communities, the Italian regions, the French regions, and the Belgian regions and communities have gradually been given more fiscal autonomy. This has not yet happened in the United Kingdom. Indeed Northern Irish politicians insisted on removing a provision allowing for fiscal autonomy from a first draft of their devolution bill, fearing that at some time the British might leave them to their own devices.

It is difficult to generalize about the powers that have been devolved to meso level governments, but there are some trends. The main macroeconomic and stabilization functions have been retained at state level, unlike in some federal systems. Regulatory and market integration functions have also been retained. Major redistributive services, notably cash benefits, have been reserved to the center in the name of national solidarity and to maintain the widest risk pooling, although there are pressures in Flanders to decentralize the social security system. Promotional activities, including economic development and planning, have been widely devolved. Indeed, one of the functional reasons behind devolution since the 1990s has been the move away from centralized regional development policies – in which regions are integrated into national economies, to a decentralized mode – in which regions compete in national, European, and global markets. Public service provision is widely devolved, including health, where there is evidence of regional divergence in priorities and instruments. Education is often devolved, although in France and Italy this refers to the maintenance of the infrastructure and planning for provision, rather than the hiring of teachers or the curriculum, seen as matters of vital national interest. In Scotland and Northern Ireland, education is entirely devolved, with no input from London, while in Spain the division between the center and the regions has caused tensions in the historic nationalities. Social services are usually devolved, while policing is devolved in Scotland and some of the Spanish autonomous communities, and in principle, can be devolved in Northern Ireland. One area in which devolved governments have become active lies in the interface between education, training, and economic development, especially in active labor market policies, which now have a strong regional dimension almost everywhere.

Relations between central and devolved governments are, in principle, regulated by intergovernmental institutions and procedures. These include joint ministerial committees and memoranda of understanding in the UK, sectoral conferences in Spain, and planning mechanisms in France. In practice, however, issues tend to be resolved at the political level, through partisan or personal contacts. This is another contrast with federal systems, where institutional forms are more important.

It is not surprising that foreign relations are a reserved matter everywhere, but even here there are important differences. Most devolved systems make provision for devolved governments to represent their interests in the European Union. Belgium, the United Kingdom, and Spain, in different ways, allow devolved governments to participate alongside the state in the council of ministers. Nearly all devolved governments have offices in Brussels; they are in the Committee of the Regions, and are active in European networks. More broadly, they are increasingly engaged in 'paradiplomacy', seeking to promote their interests in world markets and seeking out allies in other regions. State governments have at times sought to restrict this as an infringement on their prerogatives, but at other times have been quite supportive, not seeing this as a challenge to classical diplomacy; it largely depends on the political relationships between the two levels. Belgium has the most far reaching arrangements here, with regions and communities possessing external competences equivalent to their domestic ones, subject only to their not infringing Belgian foreign policy interests.

Devolution and Beyond

Devolution is a concept developed in particular circumstances, the unitary European state, and even there it comes in many varieties. Its very openness may make it attractive at a time when the role of the state is being transformed and power is moving upwards to European and global levels, downwards to regions and localities, and outwards to the market and civil society. Devolution has been subject to two recurring criticisms. The first is that power devolved is power retained, in that the center can always take it back. In practice, however, this depends on the political conditions. The second is that it is inherently centrifugal in that regions will constantly be competing with each other to gain more powers. There is evidence for this in Belgium, where the process has taken on a dynamic of its own, rather than responding to public opinion and, to a lesser extent, from Spain. On the other hand, the United Kingdom shows a very different picture, with the population accepting that different parts of the state should have a different constitutional order, and voters in England rejecting devolution for themselves. A more radical version of this critique, that devolution is a slippery slope to secession, finds even less support from the evidence. Devolving systems are in permanent evolution, some heading perhaps to become stable federations, while others are asymmetrical and multinational unions. In other cases, again, devolution may fail and states revert to centralization, as happened in Italy after the first regional reforms, or in England on two occasions. Similar processes may be seen in other parts of the world and the term devolution has been used also to refer to local government or even to decentralization within organizations. To stretch the term too far, however, would reduce its utility and risk confusion with the broader concept of decentralization.

The Research Agenda

Devolution has stimulated a new wave of academic research. At a very general level, it has reminded social scientists of the historical contingency of the unitary nation state and its constant restructuring. Geographers have drawn attention to the spatial rescaling, as functions, institutions, and policymaking have migrated to new levels. There is a lot of work on the political economy of regionalism, focusing on the changing patterns of eco nomic management, promotion, and regulation. More recently, attention has been given to welfare state re structuring as competences are redistributed and regions have to manage policies of social inclusion while main taining their competitive position and attractiveness to inward investment. There is also a large amount of research on the place of devolved regions in the decision making process of the European Union, as competences devolved to regions overlap with those taken up to the European level. Early visions of a Europe of the Regions alongside nation states have faded, but the role of regions continues to feature in discussions about constitutional reform in the EU.

See also: Nation; Political Geography; Political Representation; Regionalism; Sovereignty; State; State Theory.

Further Reading

Balme, R. (ed.) (1996). *Les politiques du néo régionalisme*. Paris: Economica.

Caciagli, M. (2003). *Regioni d'Europa*. Bologna, Italy: Il Mulino.

Keating, M. (1996). Regional devolution. The West European experience. *Public Money and Management* 16(4), 35 42.

Keating, M. (1998). *The New Regionalism in Western Europe*. Cheltenham, UK: Edward Elgar.

Keating, M. (ed.) (2004). *Regions and Regionalism in Europe*. London: Edward Elgar.

Loughlin, J. (ed.) (2001). *Subnational Democracy in the European Union*. Oxford: Oxford University Press.

Sharpe, L. J. (ed.) (1993). *The Rise of Meso Government in Europe*. London: Sage.

Swenden, W. (2006). *Federalism and Regionalism in Western Europe*. Basingstoke, UK: Palgrave.

Relevant Websites

http://www.a e r.org
Assembly of European Regions.

http://www.devolution.ac.uk
Economic and Social Research Council, Programme on Devolution.

http://www.essex.ac.uk
European Consortium for Political Research, Standing Group on Regionalism.

http://www.cor.europa.eu
European Union, Committee of the Regions.

Dialectical Reasoning and Dialectical Materialism

E. Swyngedouw, Manchester University, Manchester, UK

Glossary

Dialectics It is the science of the general and abstract laws of the development of nature, society, and thought. Its principal features are: 1) the universe is not a disconnected mix of things isolated from each other, but an integral whole, with the result that things are relationally interdependent; 2) nature - the natural world or cosmos - is in a state of constant motion; 3) development is a process whereby insignificant and imperceptible quantitative changes lead to fundamental, qualitative changes. The latter occur not gradually, but rapidly and abruptly, in the form of a leap from one state to another; and 4) all things contain within themselves internal dialectical contradictions, which are the primary cause of motion, change, and development in the world.

Doctrine of Internal Relations It is the philosophical view that things are constituted out of internalized relations. These relations are essential to them and the bearers would not be what they are without them. It implies that everything has some relation, however distant, to everything else.

Exchange Value It is a quantitative relationship that underpins and permits the exchange of different use-values in an economy based on generalized exchange (such as capitalism). The (exchange) value of a commodity is related to the socially necessary labour time required to produce that commodity.

Georg Wilhelm Friedrich Hegel Hegel (1770–1831) was a German philosopher and one of the founders of German Idealism.

Socially Necessary Labour Time It is the labor time required to produce a commodity under average conditions of production and performed with the average degree of skill and intensity of labour prevalent in that society.

Teleology It is the view that the end state of affairs is already contained (in a deterministic) manner in the original conditions. It holds that all things are designed for or directed toward a final result, that there is an inherent purpose or final cause for all that exists.

Use Value It is the utility of a thing and relates to the properties of a commodity. Use values are realized by use or consumption.

> The river
> Where you set
> Your foot just now
> Is gone
> Those waters giving way to this,
> Now this. (Heraclitus)

> But dialectical materialism insists on the approximate and relative character of every scientific theory of the structure of matter and its properties; it insists on the absence of absolute boundaries in nature, on the trans formation of moving matter from one state into another. (Lenin, 1972: 258)

Introducing Dialectics

> Motion is the mode of existence of matter. Never any where has there been matter without motion, or motion without matter, nor can there be. (Engels, 1940: 74)

Although academic writing on dialectical thought and practice can be complex, over abstract, and arcane at times, its basic premises are rather simple and straight forward. "You can never step into the same river twice," the ancient Greek philosopher Heraclitus maintained and this dictum has become a standard reference to ex press one of the key features of dialectics. For Heraclitus, everything flowed and changed. Behind the world of things and events lie a whirlpool of change, flux, trans formation, and process. This process, whereby the order of 'being' is turned into the process of 'becoming', holds both for the realm of material things (whether given by nature or made by human beings) and of ideas, dreams, or events. In other words, the world should be seen, so dialecticians maintain, simultaneously as a collection of things and a maelstrom of processes, whereby the things embody and express the dynamic relationships through which these things came into being. In sum, dialectics prioritize relation, change, openness, movement, and process over closure, stasis, condition, and fixity.

Such dialectical views are in fact widely held across the social and physical sciences, and often practiced in everyday life activities. Consider, for example, how bio logical evolution theory sees a particular species as the result of a historical process shaped by a combination of random events and dynamic survival of the fittest within particular environmental relations. New species relate to their environment in new ways, and the dynamics of their interaction will produce new ecological configurations which, in their turn, will influence the future trajectories of adaptation, change, and new species generation. Levins and Lewontin, for example, very much understand biological and geoecological evolution as a dialectical process. In physics, process based approaches to explain the nature of things are now also quite common. Consider, for example, how subatomic particles are understood simultaneously as waves and things or, how in mechanics, the action/reaction dynamic remains one of the foundational principles to explain change and/or movement. Physicists such as A. N. Whitehead argue that "[n]ature is a structure of evolving processes. The reality is the process" (Whitehead, 1985: 90), or Bohm (1983: 48), who asserts that "[n]ot only is everything changing but all is flux. That is to say, what is is the process of becoming itself, while all objects, events, entities, conditions, structures, etc. are forms that can be abstracted from this process." In geology too, Sir Charles Lyell's foundational work on the Earth's geophysical arrangement as historical process was welcomed as a prime example of dialectical thought by Friedrich Engels in *The Dialectics of Nature*.

Dialectical principles are often at work in the domain of intellectual reasoning too. For the Greek philosopher Plato, in *Republic*, dialectics were equated with a particular and superior form of reasoning that combines a deconstructionist process (to attain basic principles) with a subsequent argumentative reconstruction by means of dialogical encounter, something commonly referred to as the Socratic method. This stands for the process through which understanding is achieved by means of the unfolding of argument and counter argument. Many teaching and learning environments are very much based on such dialectical methods (as are many coffee shop discussions).

For Wilhelm Friedrich Hegel, the first modern philosopher to embrace dialectical thought, dialectical movement is the process through which the absolute idea could be reached and absolute freedom achieved. For him, the historical process is a dialectical one in which the mind or spirit gradually liberates itself in the direction of the absolute. This end state of affairs would bring about an ideal state and an ideal society. This Hegelian idealist dialectical perspective opens a set of practices that many of us are familiar with on an everyday basis. When, for example, we see all manner of injustices in the world (e.g., mass hunger or poverty), our minds and creative thought can easily come up with a set of principles, ideas, and even concrete proposals that would not only remedy the situation but make it structurally better for all involved. Of course, these ideas, although they may seem eminently reasonable and straightforward, are often very difficult, if not impossible, to implement as they hit the constraints imposed by the material world. This is a classic procedure in which an idealist dialectic is at work. It is exactly this problem of idealism (i.e., it discounts the 'constraints' of the 'really existing' world) that the materialist dialectic, mobilized in historical materialism by Marx and many Marxists, tries to deal with. For Hegel, the dialectic operates between mind and world, while for Marx the dialectic is constitutive of the world (and includes the process of thinking).

Karl Marx's materialist dialectic 'turns Hegel on his head' by insisting that the starting point for accurate historical–geographical understanding should not be the mind (thought in itself), but rather the social and material relations through which individuals and their societies maintain coherence and organize change. For Andy Merrifield, "[d]ialectics is both a statement about how the world is and a method of organizing the world for the purposes of study and research" (Merrifeld, 1993: 517). Although a materialist dialectic insists on the dialectical nature of the world and the power of dialectical reasoning, these two are not identical to each other. Dialectical reasoning is of course something that works exclusively in our minds, while many (but certainly not all) of the dialectical processes in the material world unfold independently of thinking. This is one of the radical differences between a Hegelian and a Marxist dialectic. There is a disjuncture (or a gap) between the real of the world and the real of thought. While the world moves dialectically, that is, in a process of continuous change and transformation, dialectical thought aims at re presenting this process in thought. The latter of course can never be identical to the former. As Noel Castree (1996: 347) remarks: "theory is an abstraction in thought of a real object and this is not identical to the real." This, of course, means, as David Harvey maintains, that theory (i.e., thought) "cannot hope to explain everything there is, nor even procure a full understanding of singular events" (Harvey, 1982: 451). There is always an excess or remainder that escapes representation. It is of course from here that the possibility of the new, of the unexpected, and of a politics emerges.

In what follows, we shall chart, first, the basic contours of materialist dialectical reasoning. Second, we shall explore how dialectical reasoning works more specifically in the context of a historical–geographical materialist perspective. We shall then turn to interrogating the place of dialectics in geographical enquiry and provide some examples of dialectical geographical analysis.

The Dance of the Dialectic

> For dialectical philosophy nothing is final, absolute, sacred. It reveals the transitory character of everything and in everything; nothing can endure before it except the uninterrupted process of becoming and passing away. (Engels, 1976: 8)

A materialist dialectic insists that the human and nonhuman world is inseparable and transforms through myriad of relationships that are internally contradictory. Consider, for example, how all living organisms have to transform the material world they inhabit (like inhaling oxygen and in the process producing carbon dioxide and other chemicals). This is a process of creative destruction in which everything is continuously destroyed while new things are produced. Similarly, the computer on which these lines have been written is the end product of a series of sociometabolic processes (with socionatural properties as Marx insisted in *Capital*) in which parts of nature have been destroyed to produce a new bit of 'nature', a process that implied, among others, the mobilization of resources (nature) and machinery (socially transformed nature) procured by capitalists and the harnessing and controlling of human labor power by capital to perform a set series of procedures. While the worker was executing them, he or she could be making love or holidaying: work negates leisure; leisure negates work; the two define each other, yet they are contra dictory. In human societies, this sociometabolic process of the necessary transformation of nature in order to assure generational and intergenerational sustainability is always a social–geographical as well as a physical–geographical process. This means than the contradiction between humans and their conditions of existence is contained (or managed) through a particular set of social relations that organize the metabolic process. Put simply, the sustainability of society is predicated upon a trans formation of nature (through which existing forms of nature are destroyed while new natures are produced in the process). This transformation is always socially organized through a set of social relations which define the control over, access to, or distribution of the products of nature. These social relations can take the form of, for example, slavery, feudal, or capitalist relations. In the latter society, for example, the capitalist owns the means of production (i.e., transformed nature) while workers only possess their own nature (of which the capacity to work is the important one from the perspective of the capitalist). In order to sustain their nature, workers have to sell their capacity to labor to capitalists in exchange for a salary. The latter, in turn, permits them to get sufficient (albeit this is not always guaranteed) access to the sort of natures required for their sustenance (food, shelter, etc.). Over time, these social relations change as well and new forms of social organization emerge from the debris of older ones.

The materialist dialectic, as a mode of representation, starts from the real or material world as given. It accepts that a real world exists outside of the thought process that tries to make sense of that world. Yet, it also insists that the real world (as a collage of things, events, thoughts, and texts) is not transparent in itself as the dialectical relations and processes through which this collage comes into being are not directly visible. They have to be 'reconstructed' through the process of thinking. It is very much a way of approaching the world that is comparable to that of the archaeologist or geomorphologist. He or she peels away layer after layer to reveal clues about the processes that produced the things or conditions that he or she seeks to understand. These clues, in turn, are taken as starting points to reconstruct in thought the many interrelated processes through which they came into being. Knowledge, then, is a process of re presenting the realities of worldly practices and processes, whereby knowledge is never self identical to the real. The medi ation between the real (presentation) and re presentation operates by means of concrete abstractions that permit to construct knowledges about the world. These know ledges, in turn, of course matter in important ways in co shaping future processes. Consider, for example, the Marxist notion of value. Under capitalism, a commodity expresses both use value (i.e., it possesses a series of qualities some people like to enjoy) and exchange value (i.e., a monetary expression that makes the commodity interchangeable for all other commodities on a market). Although both use value and exchange value are abstract concepts, they simultaneously express two essential, yet contradictory, qualities without which the 'thing' could not function as a commodity. While value (as socially necessary labor time) cannot be seen, using this concrete abstraction permits us to represent a commodity as simultaneously possessing exchange and use value. Value contains and expresses the dialectical relation between use value and exchange value whose contradictory existence permits things to function as commodities. It brings out the contradictory unity of the value embodied by the commodity. Indeed, both exchange and use value cannot be simultaneously realized. More importantly, it permits showing how the simultaneous but contradictory presence of both types of value is predicated upon the existence of capitalist social relations. As Ollman (1993: 11) puts it:

> Dialectics restructures our thinking about reality by replacing the common sense notion of 'thing', as some thing that has a history and has external connections to other things, with a notion of 'process', which contains its history and possible futures, and 'relation', which con tains as part of what it is its ties with other relations.

According to Louis Althusser, a Marxist philosopher who took his cue from Antonio Gramsci and Mao Zedong, the materialist dialectic differs radically from the Hegelian or idealist dialectic in three main respects. First, the materialist dialectic is nonteleological. This means that there is no automatic inbuilt final social state of affairs, an absolute and total (ideal) world. Although

Marx has often been accused that he saw communism as the final (and ideal) stage of human society, it is abundantly clear from his writing that such stage does not appear automatically. It would come into being only through determined political struggle from the working class, a clear political strategy, and a continuous fidelity to the efforts of producing a more genuine human society, something that is by no means automatic or guaranteed, and once established, would not necessarily endure. On the contrary, the materialist dialectic is radically open ended and the future is always up for grabs. The future is the domain of myriad possibilities that are open for realization (depending on the dialectics of social struggle). Second, the materialist dialectic is not singular. This means that historical processes cannot be reduced to a singular conflict or contradiction that moves or powers everything else. While, of course, Marx insisted that all history is the history of class struggle (and this suggests the presence of a singular contradiction as the only operative force), later Marxists contended that any given nature–society complex expresses a multitude of tensions and contradictions in which the class relationship might be the most important one, but it is certainly not the only and, depending on the historical conditions, not necessarily the most important one. A materialist dialectic insists simply that, for human societies at least, every process of theorizing has to begin from the dialectical relationships through which nature is socially transformed in ways that permits the existing social order to sustain and (occasionally) transform itself. The third characteristic of the materialist dialectic is that, although it is about totalities in which everything is seen and thought to be related to everything else, it is not totalitarian. The latter refers to a condition in which one domain (say the economic) determines and shapes everything else (like the form of the state, hegemonic ideologies, even lifestyles and culture). In fact, all manner of tensions, conflicts, and contradictions (e.g., gender or ethnic relations) exist in any given society that renders the total ensemble of relations extraordinarily complex and variegated. In other words, the totality is multiple, it is internally heterogeneous. In fact, many Marxists insist, the dialectical movement of history produces ever more complex relations and contradictions. However, the relationships between the human and the nonhuman world and how these relationships are socially organized remain the necessary starting point.

We are now in a position to outline the key principles of materialist dialectical reasoning. First, dialectics emphasize the study of flux, process, and change over that of things, structures, or ensembles. Dialectics are a strictly relational perspective. As Harvey contends (1996: 49), "dialecticians in effect hold that elements, things, structures, and systems do not exist outside of or prior to the processes, flows, and relations that create, sustain, or undermine them." The world, consequently, consists simultaneously of things and processes. A personal computer (PC) is the product of a wide range of sociophysical relations that made it possible to exist and to be sustained as a PC. This thing, in turn, enters in a new set of relationships through which this text becomes constituted as both an intellectual achievement (at least the hope/desire is it does) and an economic commodity (the author gets paid for it and readers have paid for reading it – or they may have subverted this by making an unauthorized copy!).

Relation and contradiction are mobilized in order to perceive change. Contradiction is the principal feature of dialectics; it may be understood as some kind of incompatible development or movement of different elements within the whole whereby each part within a relationship simultaneously supports and undermines the other. Consider, for example, how the capital–labor relationship simultaneously defines workers and capitalists, while pitting their material interests against each other. Second, all contradictions are part of a relational assemblage that relate everything to everything else. This also means, as Ollman (1990: 38) puts it, that "each part is viewed as incorporating in what it is all its relations with other parts up to and including everything that comes into the whole." If one is interested in dams, for example, materialist dialectics would interrogate the multiple contradictory relationships through which a dam becomes constituted and subsequently sustained (or not as the case may be). This would include ecological processes, social, economic and political relational dynamics, and so forth. Third, it follows from the above that "parts and wholes are mutually constitutive of each other" (Harvey, 1997: 53). For example, the integrity of our body is constituted through the relationships that are forged with parts of the 'external (to the body)' environment while the latter is co shaped by our bodily processes (e.g., all manner of excrements). This leads to an extremely important fourth principle, namely the interchangeability of subject and object, of cause and effect. From a materialist dialectical perspective, there are no one way determinations or cause–effect relationships but rather mutually interactive dynamics that shape the coming into being of new conditions, assemblages, and possibilities, while others whither away. Fifth, as change is central to dialectics, the foundational question of dialectical thinking is about how, when, and into what ensembles change. This, then, leads us to the most formidable contribution of dialectics, namely the force and importance of imagination. Given that the future is undecided *a priori*, exploring possible other worlds is an integral part of dialectical reasoning. Since, by virtue of the multiple contradictory relations that constitute a given situation or ensemble, a wide variety of possible avenues open up in the interstices of such contradictory processes. It is in these spaces that the real possibilities for social transformation reside. Authors as diverse as Slavoj Zizek, Henri Lefebvre, David Harvey, Neil Smith, Vladimir Lenin, Guy Debord, or Andy Merrifield have explored the formidable possibilities of creative materi alist imaginations for changing the world toward a more humane condition. And of course, myriad workers' and other social movements around the globe have and continue to fight for a better life and for more just and equitable sociospatial relations.

The Dialectics of Place and Space: A Historical–Geographical Materialist Perspective

> Nothing lasts long under the same form. I have seen what once was solid earth now changed into sea, and lands created out of what was ocean. Ancient anchors have been found on mountaintops. (Pythagoras of Samos)

We have explored above the theoretical and epistemological thinking on dialectics. Historical–geographical materialism mobilizes a dialectical per spective to understand the production of nature and space and to propose possible emancipatory geographical transformations. Merrifield (1993: 7) summarizes a dialectical perspective on place and space as follows:

> Social space must be posited as material *process*. This process represents the rootless, fluid reality of material flows of commodities, money, capital and information which can be transferred and shifted across the globe. … From this standpoint, social space … represents simul taneously a network of exchange and a flow of com modities, communication, energy and resources.

Similarly, space as a thing acquires meaning, signifi cance, resonance, and even a particular geographical form in and through the multiple relations with which it is infused and through which it becomes produced. For example, the making of the built environment or the cutting down of the Amazon rain forest is realized in and through sociospatial processes of appropriation, of capital accumulation, of the imagineering and scripting of peo ples, places, and natures. It has been the staple of Marxist geography to disentangle the sociospatial processes through which, particularly – but not exclusively – under capitalism, spatial configurations are produced (in their economic, political, social, cultural, and ideological in stances) and transformed. Places or concrete geographies become then a moment in the perpetual dialectical dynamics of sociospatial processes. In other words, historical–geographical materialism turns all geography into a historical socio environmental geography (not just of the past, but also of the present and the future).

Geography, then, can be understood from the vantage point of the above dialectical and historical–geographical materialist perspective. Production, defined as the social metabolization of nature, is the starting point. The dynamics of the particular social relations under which this takes place in a capitalist market economy are associated with a particular temporal and spatial organ ization of society. This territorial structure of capitalism is contradictory (that is full of tensions and conflict), and the sociospatial dynamics of capitalism are inherently unstable. These socioeconomic geographical dynamics are always inscribed in a historically produced insti tutional, political, ecological, ethnic, and cultural landscape. This produces territorial configurations or coherences that are highly differentiated and give the geographical landscape its sweeping diversity, hetero geneity, and difference. Therefore, from a materialist dialectical perspective, social processes are both space dependent and space forming. Through contradictory sociospatial dynamics, actual places and concrete geog raphies (social and physical) become constructed/ produced. These historical geographical dynamics are shaped by continuous and diverse social struggles, in which the class struggle for control over and ownership of (produced) nature is the central one. All of these struggles are always infused by a myriad of non class based cleavages and conflicts such as ethnic, gender, or territorial conflicts or by conflicts outside the realm of production, and take outspoken geographical forms.

Doing Dialectics in Geography

> Dialectic, in fact, is the only procedure which proceeds by the destruction of assumptions to the very first principle …. When the eye of the mind gets really bogged down in a morass of ignorance, dialectic gently pulls it out and leads it up. (Plato, *The Republic*)

The 'dialectical turn' in geography has taken a great number of twists and turns over the past four decades or so. William Bunge was certainly an early pioneer of a more radical geography. The pages of the journal *Antipode* were also largely devoted, at least during the 1970s, to alternative and often highly debated readings of spatial ized dialectics. Authors as diverse as David Harvey, Gunnar Olsson, Doreen Massey, Richard Peet, Edward Soja, Costis Hadjimichalis, and Alain Lipietz contributed to the exploration and formulation of a sociospatial dialectical approach, one that would be able not only to transform the discipline of geography, also but more importantly, to transform the world in an emancipatory and more just manner. While Lipietz, Massey, and others were inspired by a spatialized structural analysis that drew on Althusserian interpretations of Marxism, Harvey and Ollson pursued a dialectical investigation more akin to Henri Lefebvre's or Bertel Ollman's view of dialectics as constituted through internal relations. From the 1980s onward, these perspectives would spiral out in a variety of directions, embracing more nuanced relational perspectives. In particular, David Harvey would articulate a fully developed dialectical perspective in a series of pathbreaking books, starting with *Limits to Capital*. In *Justice, Nature and the Geography of Difference*, he would propose a more systematized dialectical approach to sociospatial processes.

During the 1970s, dialectical analysis in geography tended to focus on spatialized political–economic questions and problems, such as urban and regional change or processes of global uneven development. More recent work has embraced socioecological transformations, variable sociocultural formations, ethnicity and gender, and the thorny issue of the dialectic of language, representation, and sociomaterial change.

While Marxism and dialectics have received considerable and often critical attention in geography, it is surprising that relatively few empirical geographical studies have been published by geographers that explicitly or implicitly explore the world from a dialectical historical–geographical materialist perspective. Doing dialectical materialist research requires a painstaking and theoretically informed reconstruction of the contested making and remaking of human and nonhuman geographical landscapes. While not pretending to be exhaustive in any way, there are a number of monographs that explore geographical conditions and configurations from a decidedly dialectical perspective. They may give some pointers as to how a historical–geographical materialist and dialectical geography might be practiced. Great dialectical urban monographs would include Mike Davis' *City of Quartz*, David Harvey's majestic study of nineteenth century Paris (*Paris, Capital of Modernity*), Andy Merrifield's *Dialectical Urbanism*, Matthew Gandy's *Concrete and Clay*, and Maria Kaika's *City of Flows*. Exciting socioecological studies would include Richard Walker's *The Conquest of Bread*, Michael Watts's *Silent Violence*, or Judith Carbey's *Black Rice*. The author would also single out George Henderson's *California and the Fictions of Captial*, Don Mitchell's *The Lie of the Land* and the author's own *Social Power and the Urbanisaton of Water* as dialectical historical–geographical materialist studies. Great Marxist studies of globalization by geographers would include Neil Smith's work on globalization and David Harvey's inspiring analyses of imperialism and neoliberalism.

While the above books all explore the difficult pathways of dialectical reasoning in geographical studies, they also share a passion and commitment to exploring possible *Spaces of Hope* as David Harvey likes to call them, and to thinking about a politics that would turn the actually existing perverse geographies of exploitation, domination, hunger, and war into a genuinely humanizing spatial practice.

See also: Critical Theory (After Habermas); Geopolitics; Harvey, D.; Historical-Geographical Materialism; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Radical Geography; Smith, N.; Social Movements; Structural Marxism; Structuralism/Structuralist Geography.

Further Reading

Althusser, L. (1965). *For Marx*, Brewster, B. (trans.). [Reprinted 1969, London: New Left Books.]

Balibar, E. (1993). *The Philosophy of Marx,* Turner, C. (trans.). [Reprinted 1995, London: Verso.]

Berman, M. (1983). *All That Is Solid Melts into Air: The Experience of Modernity*. London: Verso Books.

Bhaskar, R. (1993). *Dialectic*. London: Verso Books.

Castree, N. (1996). Birds, mice and geography: Marxisms and dialectics. *Transactions Institute of British Geographers NS* 21, 342 362.

Debord, G. (1967). *The Society of the Spectacle*. London: Rebel Press.

Engels, F. (1940). *Dialectics of Nature*. London: Lawrence and Wishart.

Engels, F. (1976). *Ludwig Feuerback and the End of Classical German Philosophy*. Peking: Foreign Languages Press.

Harvey, D. (1982). *The Limits to Capital*. Oxford: Blackwell.

Harvey, D. (1984). On the history and present condition of geography an historical materialist manifesto. *The Professional Geographer* 36, 1 11.

Harvey, D. (1985). The geo politics of capitalism. In Gregory, D. & Urry, J. (eds.) *Social Relations and Spatial Structures*, pp 128 163. London: Macmillan.

Harvey, D. (1996). *Justice, Nature and the Geography of Difference*. Oxford: Blackwell.

Harvey, D. (2000). *Spaces of Hope*. Edinburgh: Edinburgh University Press.

Lefebvre, H. (1974). *The Production of Space*. [Reprinted 1989, Oxford: Blackwell.]

Lenin, V. I. (1972). *Materialism and Empirio Criticism Critical Comments on a Reactionary Philosophy. Lenin Collected Works, Vol. 14*. Moscow: Progress Publishers.

Levins, R. and Lewontin, R. (1985). *The Dialectical Biologist*. Cambridge, MA: Harvard University Press.

Marx, K. (1990). *Capital Volume 1*. Harlow: Penguin Classics.

Marx, K. (1993). *Grundrisse: Foundations of the Critique of Political Economy*. London: Penguin Classics.

Merrifield, A. (1993). Place and space: A Lefebvrian reconciliation. *Transactions, Institute of British Geographers N.S.* 18, 516 531.

Merrifield, A. (2002). *Dialectical Urbanism*. New York: Monthly Review Press.

Ollman, B. (1993). *Dialectical Investigations*. London: Routledge.

Ollman, B. (2003). *The Dance of the Dialectic: Steps in Marx's Method*. Champaign, IL: University of Illinois Press.

Smith, N. (1984). *Uneven Development*. Oxford: Blackwell.

Swyngedouw, E. (1999). Marxism and historical geographical materialism: A spectre is haunting geography. *Scottish Geographical Magazine* 115, 91 102.

Swyngedouw, E. (2006). Circulations and metabolisms: (Hybrid) natures and (cyborg) cities. *Science as Culture* 15, 105 122.

Tse Tung, M. (1957). *On Contradiction*. Peking: Foreign Languages Press.

žižek, S. (2006). *The Parallax View*. Cambridge, MA: The MIT Press.

Dialogism (After Bakhtin)

J. Holloway, Manchester Metropolitan University, Manchester, UK
J. Kneale, University College London, London, UK

Glossary

Carnivalesque Popular cultural practices and representations that are critical of or subvert official culture. Ambivalent images and rituals that emphasize renewal and the cyclical nature of time are central to the carnivalesque (e.g., images of birth accompanied by those of death).

Chronotope The temporal and spatial matrix through which narratives and notions of subjectivity are organized – the inseparability of time and space in life and representation.

Dialogism A central Bakhtinian term suggesting that all language and existence are relational and responsive. Each speech or communicative act is dialogical through its responsiveness to other past, present, and future speech acts.

Grotesque Realism Central to the carnivalesque, grotesque realism principally involved practices and images with an exaggerated emphasis on the lower bodily regions, thus inverting the official legitimacy of the idealized body.

Heteroglossia A term used to describe the multiplicity and diversity of discourses in the social world – literally, many languagedness. Heteroglossia refers to heterogeneous and centrifugal forces which exist in constant ideological tension to the centripetal or monological forces which seek to unify the linguistic sphere.

Monologism An ideological force in culture and society which seeks to unify the diversity of heteroglossia. Any attempt to order or homogenize linguistic diversity can be described as monological.

Novelness A term used to describe the constant renewal and change characteristic of heteroglossia. Bakhtin identified the novel as the literary form which best reflected the unceasing change and diversity inherent to language and speech.

Polyphony Literally, many voices, polyphony refers to the dialogical encounter where each voice has equal value and standing.

Speech Genre A relatively stable form of language incorporating shared style, intonation, and content. Certain speech genres seek to ideological ascendancy and thus attempt to monologize heteroglossia.

Utterance The basic unit of communication in dialogism, utterances vary in length and are separated by a change in the speaking subject. The utterance is formed in a dialogical encounter through responding to past and anticipating future utterances.

Introduction

The work of Mikhail Mikhailovich Bakhtin (1895–1975) continues to inspire and influence in a variety of ways across a diverse range of disciplines, including literary theory, history, cultural studies, anthropology, and human geography. Bakhtin's writings, from the early philosophical works of *Towards a Philosophy of the Act* and *Art and Answerability* to the studies of folklore and medieval popular culture in *Rabelais and his World*, contain both varied and recurring themes. For some, especially Tzvetan Todorov, an early proponent of his work, the inherent unsystematic nature of Bakhtin's work is a problem that needs to be addressed. Yet it seems apposite that Bakhtin's work shows signs of both consistency and diversity. For, as we shall see, Bakhtin argued that the social world is simultaneously composed of heterogeneous/centrifugal forces *and* centripetal forces that seek homogeneity. Furthermore, when one tries to categorize Bakhtin's ideas, politically or theoretically, one is faced with a diversity of labels that never seem to coalesce: Is Bakhtin a Marxist? A post Marxist? Is he a liberal thinker or an early post structuralist? Once again, it seems appropriate that a thinker, who, as we shall see, repeatedly interrogates ideas of becoming and unfinalizability, should never be compartmentalized. Thus, any attempts to systematize or complete Bakhtin's ideas disavow the very philosophy and politics he valued.

After studying classics and philology at the University of Petrograd (1913–18), Bakhtin moved to Nevel in western Russia to work as a schoolteacher for 2 years. During this time and after relocating to Vitebsk in 1920, Bakhtin engaged with a group of intellectuals that included (among others) Valetin Voloshinov and Pavel Medvedev. Voloshinov's *Marxism and the Philosophy of Language* and *Freudianism: A Marxist Critique*, as well as Medvedev's *Formal Method in Literary Scholarship* are often attributed to Bakhtin himself. This 'authorship dispute' has never been and is unlikely to be settled. Instead, it is common to use the shorthand the 'Bakhtin Circle' to register the mutual influence and interaction among these writers in these 'disputed texts'. Less in dispute is

Bakhtin's arrest in 1929, under one of the numerous Stalinist purges of unorthodox intellectuals, and his exile in Kazakhstan. During this period, he wrote his important essays on the novel and Rabelais. Bakhtin defended the latter as a doctoral dissertation at the Gorky institute in 1941. However, this award was delayed and downgraded by both World War II and its controversial content. After exile, Bakhtin moved to Saransk University and became the head of the department of Russian and world literature. It was during the 1950s that Bakhtin was 'rediscovered' by a number of graduates from Moscow, who managed to persuade him to update his book *Problems of Dostoevsky's Poetics* and make him known to a wider and, in the 1960s and 1970s, Western world. Mikhail Bakhtin suffered all his life with the crippling bone disease, osteomyelitis (he had a leg amputated in 1938), and due to his ill health he moved back to Moscow in 1969, where he died in 1975.

Inspecting Bakhtin's Key Concepts

Dialogism and the Utterance

One of the key aspects of Bakhtin's theory of dialogism is its opposition to structuralist theories of language. For structuralists, especially those influenced by Ferdinand Saussure, language can be divided in two ways: langue, which is a set of normative rules which govern language, and parole, the implementation of these rules in any given speech or linguistic act. Parole therefore comprises the infinitely multiple and everyday ways we talk and communicate. Due to the infinite diversity of everyday speech, Saussure argued that parole should be ignored in favor of langue. Bakhtin and Voloshinov took the opposite view and argued that rather than seeing the multiplicity and particularities of speech in its everyday performance as something accessory to the rules which govern and structure language, we should seek to interrogate the situated and historical utterances of everyday life. Bakhtin and Voloshinov described the arguments of structuralists as abstract objectivism and monologism (which we will return to). They argue that this abstract objectivism sucks the life out of language as it is used and evolves in everyday life. Moreover, the unfinished and alive language of everyday life eludes all systems and laws through its incessant change. Yet, Bakhtin and Voloshinov also dismissed what they call theories of individual subjectivism where language is purely the product of individual consciousness, abstracted and isolated from the wider social world in which we make meaning and sense.

In distinction to these two theoretical trends, Bakhtin and Voloshinov advanced their theory of dialogism. With its focus on language in its everyday and social context, the key unit of communication in dialogism is the utterance. This varies in size and length – it can be a single word or a novel – and is delimited by the change to another speaking subject. In fact, the role of this other, whether this is another speaking person or a written response, is crucial to understanding the nature of the utterance. This is because the utterance simultaneously looks forward to the response it partly shapes and backward to the utterances it responds to. The utterance is double voiced and is responsive to past, present, and future utterances that mold its development. Meaning is therefore always constructed between utterances and speaking subjects. Shared understandings and cultural meanings are created at the boundaries and borders between utterances. The emphasis here again is on the lived and concrete construction of meaning, rather than abstracted rules and governing systems of language. Moreover, the individual utterance is part of a long chain of other utterances. Consequently, meaning can never be finalized or its origins found: language for Bakhtin and Voloshinov is a constantly generative and open ended activity, full of potentiality for words and utterances to take on new meanings through time. The dialogical utterance is thus formulated both synchronically and diachronically, and always has the potential, among and in conjunction with multiple other (past, present, and future) speech acts, to generate new and unforeseen meanings.

Bakhtin's theory of dialogism is thoroughly ethical. He argued we do not choose whether or not to enter into dialogical relations with others, their utterances, or the world, because we are continually and unavoidably responding to the world and others around us. Each of us occupies a unique and particular position in the world, which we must seek to actualize, and thus we are always already bound to respond to the others and the world. As such, in responding to the world, we have a responsibility to the world and other people: we must address the world and the world (must) address us. Bakhtin argued that this state of responsibility or addressivity is inescapable. This ethical stance is once again derived from the actual, concrete, and lived event, rather than through systems and structures. In *Towards a Philosophy of the Act*, Bakhtin disputed Kant's ethical imperative (the need for ethical principles that govern our responses and actions) and argued that ethics can only be located in particular acts as we respond and address to those events as they happen. If it were possible to abstract and represent what those reactions should or would be, then we could know what existence would be in advance. Likewise, ethical relativism, which leads to a state of nonresponsiveness to the world, is impossible in Bakhtin's schema. As we cannot know life in advance of its happening nor can we escape ethical responsibility, once again Bakhtin stressed the lived actuality of events and argues that ethical actions are unavoidably a result of the unfinalizability of practice in the world and toward others.

Speech Genres, Heteroglossia, and Polyphony

Bakhtin and the 'Bakhtin Circle' sought to trace a middle path between structure and system on the one hand and individualism and relativism on the other. Another way this was achieved was through Bakhtin's notion of the speech genre. Speech genres are deployed in utterances through typical and common forms of intonation, style, content, and composition. For example, we might talk in a different style and about different things in the pub than we do in the workplace. Or we might give different intonation to our speech and discuss different subjects when speaking to a close relative or lover instead of a casual acquaintance. Therefore, the utterance is achieved through the deployment of different conventional and generic ways of talking, which allow it to become double voiced: that is, to register past and future speech acts, speech genres act as a force of ordering and form on the utterance (without being finalizable). Importantly, speech genres reveal the social context of talk and speech acts in that utterances are always enabled by and reproduce socially shared ways of understanding. Bakhtin's theory of dialogism makes it clear that all ways of talking and communicating are social, value laden, and consequently ideological. Language is thus saturated and stratified by power. The deployment of different speech genres registers social hierarchies and differences, and each utterance is ideologically accentuated. The early work of the British Centre for Contemporary Cultural Studies found Voloshinov's nonstructuralist theory of language attractive for this reason.

Shaped in anticipation and response to past and future speech, enabled by, and reproducing speech genres, the utterance is a living and concrete speech act that is thoroughly social and ideologically permeated. Dialo gism then recognizes the performance of language and discourse as everyday, open to change, and subject to wider socio ideological forces. The multiplicity of socio ideological discourses, languages, and speech genres that penetrate and are performed through the utterance Bakhtin named heteroglossia. Each utterance is formed in and through heteroglossia – the multilanguagedness of the social world. Referring to the juxtaposition of different languages, speech genres, and dialects, hetero glossia is a term Bakhtin used to describe both the multiplicity of different ways of speaking and the nature of the relations between different languages. Reflecting once more upon the ideological nature of language, the heteroglossic context of different discourses and speech genres, diachronically and synchronically forms the double voiced utterance. Therefore, the context of the utterance is always dialogized heteroglossia.

The relation between the different discourses is often one of discord and struggle, and therefore relations of power infuse heteroglossia. Bakhtin argued that certain speech genres seek dominance and ascendancy, and thus seek to homogenize the heterogeneity of heteroglossia. These centripetal tendencies Bakhtin called the forces of monologism or monoglossia. In opposition to the centrifugal and diverse forces of dialogized heteroglossia, monological forces seek to unify the socio ideological linguistic sphere. Indeed, Bakhtin saw structuralist the ories of language, which make accessory the living world of everyday speech, as monological through their attempt to discover the unifying rules of language. The hetero glossic forces of the living utterance always relativize monologism: the diversity of heteroglossia undermines monoglossia. This does not mean that heteroglossia represents some form of linguistic anarchism: Bakhtin argued that there must be some form of order to the heteroglottic sphere (principally through the relative unity of different speech genres), or else meaningful communication would be impossible. However, this order is always open to change and mutation.

Besides reserving criticism for the abstract objectivism of structuralism, Bakhtin and other members of the Bakhtin Circle most particularly Medvedev, heavily criticized Russian formalism. This movement in literary theory rejected the everyday linguistic sphere and fo cused on the rules governing language in poetry (which they saw as the privileged site for meaning generation and true creativity). Bakhtin and Medvedev rejected this as another form of monologism abstracted from the di versity and creativity of heteroglossia. Indeed, Bakhtin championed the novel as the literary form that reflects or has the potential to reflect the diverse heteroglot lan guages of dialogism. In particular, Bakhtin argued that Dostoevsky's novels (such as *Poor Folk*) were important as they reflected the different ideological discourses of dialogized heteroglossia, and thus presented a multi plicity of different voices in dialogical relation. Moreover, Bakhtin argued that in Dostoevsky, we see not only the interaction and interpenetration of diverse ideologies and discourses, but also that these voices remain on an equal footing, with no one voice being privileged over another. This Bakhtin termed polyphony: an artistic orchestration of multiple and diverse discourses where none become dominant or seek to monologize another.

The novel is given privileged status in Bakhtin's the ories: he argued that in the novel the centripetal forces of heteroglossia are artistically fully realized. These diverse and multiple languages are, as we have seen, never finalized and are always open to metamorphosis. Indeed, the incessant becoming of heteroglossia and the forces which pull apart monologism represent what Bakhtin deemed novelness – a constant force of renewal in the world. As such novelness, which characterizes the centrifugal power of dialogized heteroglossia, is seen as a normative project – through novelness, all centripetal tendencies and monological utterances are pulled apart and turned upside down. Yet, here we see a slight contradiction in Bakhtin's work. For if dialogism is a universal of all communication – the utterance is unavoidably double voiced and unfinalized – then how can dialogism and novelness be something we must work toward? Ken Hirschkop argues that this tension (dialogism as simultaneously universal and normative) is central to Bakhtin's work.

Carnival and the Carnivalesque

One of Bakhtin's most famous contributions to the social sciences is his study of carnival and the carnivalesque. Bakhtin explored carnival images and carnival practices in *Problems of Dostoevsky's Poetics* and in his essay on the chronotope, but their most lengthy and detailed discussion is presented in *Rabelais and his World*. This major work explores medieval folk practices and humor as represented by the sixteenth century French humanist scholar Francois Rabelais (*c.* 1494–1553). Rabelais' series of five books, *Gargantua and Pantagruel*, recorded and praised medieval folk culture at the juncture of the Renaissance and the beginnings of modern humanist thought. Bakhtin read these texts as revealing a persistent force of subversion and parody of all that is held as 'high' or 'sacred' in official or, more specifically, monological culture. Many critics have read the *Rabelais* book as a veiled attack on the Stalinist context in which it was written, although this hardly exhausts its meanings.

The carnivalesque represents the practices of unofficial popular folk culture that subverts the practices and decrees of officialdom. The world of labor and feudal relations, as well as the Church and its proclamations of 'cosmic terror', are mocked and subverted. The carnivalesque is both a series of representations and rituals that turn the world upside down: a critical inversion of all high culture, normalized behavior, and accepted categories, everything that is held to be refined, ordered, serious, or principled is laughed at and ridiculed through the carnivalesque. As such, the carnival was a time and space that existed in a state of constant tension with the official. More specifically, the carnivalesque is performed in a dialogical relation to official culture through its heteroglossic and centrifugal dismantling of the centripetal forces of monologism. All those forces which seek to homogenize and bring into stasis the heterogeneous and unfinished sociocultural world are scorned, derided, and reversed.

The carnivalesque is a force of novelness and thus renewal, through what Bakhtin called the grotesque realism of carnival images and practices. Here, exaggerated images of the body become a means of defacing the intellectualist and spiritual transcendent forms of the official. Instead of the closed, smooth, or idealized body, the grotesque body values the lower bodily regions, and, in so doing, subverts their denial in official culture. In particular, Bakhtin argued that the focus on orifices in carnival culture revealed the grotesque body as open to the world, or more specifically, as existing in an unfinished and dialogical relation with others and the world around it. Once again, the border zones of openness and contact with the social world, the social world of dialogical relations, are given preference over closed and monological tendencies. The carnival was a thoroughly social space and time where no distinction was made between audiences and performers. Moreover, the grotesque body is thoroughly social in that the carnivalesque image of corpulence, excess, and protuberances is the body of the people rather than the individualized body. As the body of the people, the carnivalesque is a social phenomenon – through images and practices of birth and renewal, the social world dies and is born again. This devouring/devoured, open/closed, grotesque body points to the endless reproduction of the world. The time of carnival is thus a cyclical time, where popular folk culture values the continuous reproduction of life through a focus on birth and death, destruction and renewal, instead of teleological temporalities, visions of the end times or the certainties of the final word. The grotesque body was accompanied by modes of carnivalized language or speech. Again, Bakhtin focused on those speech genres which performed an open and dialogical relation to the social world – in particular 'Billingsgate', the language of the marketplace, the street, and the public square. This profaning and cursing speech genre is simultaneously full of ironic praise and excessive celebration – it is Janus faced. Indeed, the carnival and the carnivalesque are characterized by this deep ambivalence: praise is always accompanied by ridicule; images of birth are always associated with death.

For some commentators and critics, Bakhtin's emphasis on the social rather than the individualized body denies the historical realities of carnivalesque spatialities and temporalities. For example, it was often specifically the female body which was the object of ridicule and abuse. Similarly, images of women appear throughout *Rabelais and His World*, but their voices remain unheard and thus Bakhtin could himself be accused of monologism. In addition, there is some debate between commentators as to whether carnival survives today. This is partly due to Bakhtin's own rather ambiguous discussion; throughout the *Rabelais* book he argues that the eighteenth and nineteenth centuries witnessed the weakening of the rejuvenating powers of folk laughter, yet he also suggested that carnival is indestructible. In parallel, some have argued the carnivalesque was both episodic and licensed by officialdom, and consequently its revolutionary potential was or is negated. In this interpretation, the fact that carnival proper was (and is) restricted in a historical time and place, taking place at set times of the ritual calendar and in the public squares and streets of the city, marks it out as a ritual staging of symbolic protest, a fundamentally empty gesture but one which allows the expression of a culture's social and psychic tensions (and thus a legitimated 'safety valve' in culture and society). Yet, one need not see carnival as somehow fixed in time and space, but instead an ever present force of potentiality and renewal in the everyday. This we would argue is complementary to some of the common themes we have identified in Bakhtin's work: the social and cultural force of the carnivalesque lies in its disorientating, regenerative, and unfinalizable becoming in, and not sequestrated from, the dialogical practices of the everyday.

The Chronotope

The final idea of Bakhtin's discussed here is the chronotope, which literally means the inseparability of time and space. In particular, Bakhtin used the notion of the chronotope to explore the relation between time, space, and subjectivity or the self as it is aesthetically expressed in literature. Thus, the carnivalesque can be seen as form of chronotope in literature: it expresses a relation between subjectivity as part of a communal social body, an artistic expression of time through the notion of cyclical and regenerative time, and a spatiality of the public square, street, or tavern. Moreover, the chronotope of the carnival is such that we cannot separate time from space or vice versa – both are intrinsically and inseparably fused in its artistic and literary expression.

Bakhtin explored how chronotopes, represented in literature, have changed over time. For example, he explores how the Greek Romances of the second to sixth centuries AD differ in their representation of time–space and self, to those of the nineteenth century novel. In discussing these sources, we see how chronotopes anchor the narratives of different forms of literature. In the Greek literature, which Bakhtin collectively deems adventure novels, it is the chronotope of the road which allows the tying together of a narrative, especially in the form of encounters (random or otherwise) between characters. However, chronotopes are not only of interest as formal properties of text and narrative: Bakhtin argued that texts can also be read chronotopically to reveal ideological values regarding the relationship between space, time, and self. As such, different chronotopes in literature reveal the social, cultural, and political context of the time (and place) in which they were written. This move allowed Bakhtin to explore the decline of particular types of chronotope – for instance, the development of ideas of the self in ancient autobiography, or the 'sticky' time–space of Flaubert's provincial town or Tolstoy's noble estates. In other words, the chronotope allows us to trace the construction of time and space in different time periods and in different spaces.

Dialogism and Human Geography

Despite the value and potential application of Bakhtin's ideas for human geography, there has disappointingly been little uptake of his work in the discipline. Yet, this final section briefly reviews how Bakhtin has been utilized in human geography and applied to human geographical concerns. Broadly these applications fall into four categories. First, dialogical relations of meaning creation have been utilized by human geographers. For example, in a paper on the role of women in NGO led development, O' Reilly explores how women are placed in contradictory positions (as both central to NGO development *and* excluded from total participation) through contested dialogical relations in different spatial contexts. Here, dialogism is explicitly spatialized to show how different sites, locations, and places affect the meanings and ways of speaking available to differently gendered actors. The focus on the everyday and contested nature of meaning, and dialogism's key notion that language is a form of practice, is also a central to Thrift's focus on Bakhtin. As such, language is not concerned only with representation, but is also part of how space is continually ordered and transformed through practice. In a similar move, Painter focuses on the everyday and practical constitution of the state – what he deems stateness. Painter is keen to show how the state is not an abstract structure or reified concept by drawing upon Morson and Emerson's idea of Bakhtinian prosaics. In this way, for the state to have effects on its citizens (his example being the implementation of the so called 'Anti Social Behaviour Orders' in the UK) requires a whole host of actors and quotidian actions to be assembled, maintained, and performed. Painter draws upon Bakhtin to reveal how the everyday world (and the state's action in it) is unfinalizable, instable, and uncertain, as well as analyzing government documents to show how their meanings are constructed through the multiple overlapping and interacting discourses of heteroglossia. The unending forces of heteroglossic dialogical relations is also drawn upon by Folch Serra, whose paper remains one of the most sustained inspection of Bakhtin's work in human geography. She argues that dialogism is a counter to centripetal and monological forces previously at work in human geography as a discipline (such as the mathematical refinements of the quantitative revolution) in that it pulls apart grand narratives in favor of multiplicity and heterogeneity. In addition, Folch Serra values the shifting relations to Otherness found in dialogism and explores this process in terms of the ethnographic encounter. The utility of dialogism in relation to how human geographers research and write about the human world is further explored by Sutherland in her discussion of the politics of fieldwork, by Chatterton in his exam ination of encounters between activists and oil refinery workers, and by Crang in his analysis of polyphony and the construction of human geographic texts. In a different manner, alterity and the unavoidable responsibility to Otherness, detailed in Bakhtin's early work, is taken up by Holloway in a study of embodiment and the sacred rural landscapes in Glastonbury, UK.

Second, those interested in textual and literary geographies have drawn upon Bakhtinian theory; Marc Brosseau explores the representation of the city in John Dos Passos' novel *Manhattan Transfer*. Brosseau shows how, in content *and* form, the novel expresses urban life, such that the text becomes a city text. In particular, *Manhattan Transfer* emulates the experience of city life through the temporality (e.g., fragmented chronologies, a sense of simultaneity, and the 'here and now') and spa tiality (e.g., the contingency of place and urban diversity) of its composition. Here, Brosseau explicitly explores the chronotope of modern urbanity as it is expressed in this particular novel. This chronotopic analysis is also de veloped in Kneale's study of the horror fiction of H. P. Lovecraft, particularly the chronotope of the threshold, and specifically the graveyard and grotesque body as thresholds (sites of connection between outside/inside, or life/death) in Lovecraft's *The Unnamable*. Folch Serra again provides another utilization of the idea of chronotope, by combining Bakhtinian concepts to understand landscapes, regions, and place as produced and reproduced through dialogical forces. Here the idea of chronotope is understood as constructed through the relative weight and impact of both centripetal monological and centrifugal dialogical forces. As such, landscapes, places, and regions are constructed through socio ideological conflict.

Third, Bakhtin has been evoked in geographer's explorations of the cultural politics of carnivals everyday life and language. For example, Peter Jackson has written about Bakhtin and cultural politics in relation to London's Notting Hill carnival and Gavin Brown has compared the emancipatory potential of the carnival esque, antidogmatic, and antiauthoritarian spaces created by both the anticapitalist movement 'Reclaim the Streets' and through sites of public (homo)sex and sexual pleas ure. A fourth way in which Bakhtinian ideas have been used to consider human geographical concerns involves notions of movement, exile, and hybridity – especially in postcolonial theory. For example, Gilroy's enquiry into the 'double consciousness' of the Black Atlantic looks at how the chronotope of the ship, traveling between sites in the colonial world, brought forth different dialogical, asymmetrical, and hybridized encounters between iden tities and spaces. This application of Bakhtinian ideas to movement and exile is further adopted in DeSantis' study of migrants to the US. Through highlighting some of the struggles over identity these exiles face, DeSantis argues that their stories are understood as dialogically generated and thus migrant's experiences are endlessly produced through different centrifugal and centripetal tensions.

See also: Becoming; Body, The; Cultural Politics; Discourse; Festival and Spectacle; Ideology; Jackson, P.; Language; Literature; Medieval Historical Geographies; Popular Culture; Poststructuralism/Poststructuralist Geographies; Public Space; Self-Other; Semiotics; Space-Time; Structuralism/Structuralist Geography; Text and Textuality; Text, Textual Analysis; Thrift, N.

Further Reading

Bakhtin, M. M. (1981). *The Dialogic Imagination*. Austin, TX: University of Texas Press.

Bakhtin, M. M. (1984). *Problems of Dostoevsky's Poetics*. Minneapolis, MN: University of Minnesota Press.

Bakhtin, M. M. (1984). *Rabelais and His World*. Bloomington, IN: Indiana University Press.

Bakhtin, M. M. (1986). *Speech Genres and Other Late Essays*. Austin, TX: University of Texas Press.

Bakhtin, M. M. (1990). *Art and Answerability: Early Philosophical Essays*. Austin, TX: University of Texas Press.

Bakhtin, M. M. (1993). *Toward a Philosophy of the Act*. Austin, TX: University of Texas Press.

Burke, P. (1994). *Popular Culture in Early Modern Europe*. Aldershot: Scolar Press.

Brosseau, M. (1995). The city in textual form: *Manhattan Transfer's* New York. *Ecumene* 2, 89 114.

Brown, G. (2004). Sites of public (homo)sex and the carnivalesque spaces of reclaim the streets. In Lees, L. (ed.) *The Emancipatory City? Paradoxes and Possibilities*, pp 91 107. London: Sage.

Chatterton, P. (2006). 'Give up activism' and change the world in unknown ways: Or, learning to walk with others on uncommon ground. *Antipode: A Radical Journal of Geography* 38, 259 281.

Crang, P. (1992). The politics of polyphony: Reconfiguration in geographical authority. *Environment and Planning D: Society and Space* 10, 527 549.

DeSantis, A. D. (2001). Caught between two worlds: Bakhtin's dialogism in the exile experience. *Journal of Refugee Studies* 14, 1 19.

Folch Serra, M. (1990). Place, voice, space: Mikhail Bakhtin's dialogical landscape. *Environment and Planning D: Society and Space* 8, 255 274.

Gilroy, P. (1993). *The Black Atlantic: Modernity and Double Consciousness*. London: Verso.

Hirschkop, K. (1999). *Mikhail Bakhtin: An Aesthetic for Democracy*. Oxford: Oxford University Press.

Hirschkop, K. and Shepherd, D. (eds.) (1989). *Bakhtin and Cultural Theory*. Manchester: Manchester University Press.

Holloway, J. (2003). Spiritual embodiment and sacred rural landscapes. In Cloke, P. (ed.) *Country Visions*, pp 158 175. Harlow: Prentice Hall.

Holquist, M. (2002). *Dialogism: Bakhtin and His World* (2nd edn.). London: Routledge.

Jackson, P. (1988). Street life: The politics of carnival. *Environment and Planning D: Society and Space* 6, 213 227.

Jackson, P. (1989). *Maps of Meaning*. London: Routledge.

Kneale, J. (2006). From beyond: H. P. Lovecraft and the place of horror. *Cultural Geographies* 13, 106 126.

Medvedev, P. N. (1985). *The Formal Method in Literary Scholarship: A Critical Introduction to Sociological Poetics*. Cambridge, MA: Harvard University Press.

Morris, P. (ed.) (1994). *The Bakhtin Reader: Selected Writings of Bakhtin, Medvedev and Voloshinov.* London: Edward Arnold.

Morson, G. S. and Emerson, C. (1990). *Mikhail Bakhtin: Creation of a Prosaics*. Stanford, CA: Stanford University Press.

O'Reilly, K. (2004). Developing contradictions: Women's participation as a site of struggle within an Indian NGO. *The Professional Geographer* 56, 174 184.

Painter, J. (2006). Prosaic geographies of stateness. *Political Geography* 25, 752 774.

Sutherland, E. (2004). In the field politics: Multiple voices of outworkers and researcher. *Australian Geographer* 35, 161 168.

Thrift, N. (1996). *Spatial Formations*. London: Sage.

Todorov, T. (1984). *Mikhail Bakhtin: The Dialogical Principle*. Minneapolis, MN: University of Minnesota Press.

Vice, S. (1997). *Introducing Bakhtin*. Manchester: Manchester University Press.

Voloshinov, V. N. (1973). *Marxism and the Philosophy of Language*. Cambridge, MA: Harvard University Press.

Voloshinov, V. N. (1976). *Freudianism: A Marxist Critique*. New York: Academic Press.

Relevant Websites

http://www.geocities.com//CollegePark/Campus/8297/bakhtin.html/
An online Bakhtin bibliography, Geocities.

http://www.shef.ac.uk
The Bakhtin Centre at the University of Sheffield (include useful analytical database for searching works on or about Bakhtin).

http://www.utm.edu
The Internet Encyclopedia of Philosophy's entry on the Bakhtin Circle, The University of Tennessee at Martin.

Diaries (Video, Audio or Written)

P. Meth, University of Sheffield, Sheffield, UK

Glossary

Diary A product of recording (either in written, video, audio, photographic, or illustrative format) the views or experiences of an individual usually in chronological order.

Solicited An account or recording produced by a research participant at the request of a researcher.

Unsolicited An account or recording produced for personal reasons and not at the request of a researcher or other party.

Introduction

The term ‘diaries’ covers a broad range of methodological choices and outcomes for geographers; indeed, the term can be quite deceptive in itself, and a range of approaches are often described generically as ‘diary methods’. This article will help readers to distinguish between solicited and unsolicited diaries, as well as diary methods which are explicitly qualitative in approach versus those that are explicitly quantitative in focus – and all the approaches in between!

A common characteristic and advantage of using the diary method is its longitudinal rather than singular snapshot view. Written or recorded over a period of time diary keeping is discontinuous and thus allows for variation and flexibility in the entries that are provided. Changes or consistencies in participants’ experiences and records can be observed allowing a researcher to understand the complexities of life more fully. Diary keeping also takes advantage of the currency of recording, thus proving to be more accurate than retrospective methods which rely on memory. Diary keeping is usually chronological in form – although assumptions about chronological ordering may prove false with cross cultural research and it is important to be aware of this possibility. Given these features and despite their recent rise in popularity, the use of diaries (written, video, audio or blogs, etc.) as a methodological technique within human geography is a relatively poorly discussed subject. Much of the discussion of diary usage refers to the use of historical and often unsolicited diaries where the personal diaries of particular individuals (usually deceased) are used to explore and reflect on socio cultural and political trends of their time. These diaries are largely handwritten and we see very little evidence of geographers making use of historical audio or video diaries in their work – although the use of historical film, art, and music (nondiary in form) is growing in popularity. Less commonly discussed within geography is the use of solicited diaries for research purposes, although this is rapidly proving an interesting way of gaining significant and current data. The following considers in more detail some of these issues, and aims to offer geographers advice when selecting this particular method.

The Dominance of Written Contributions and the Rise of ‘Alternative Techniques’

Within the social sciences more generally there is a strong reliance on written material rather than audio or video recordings, both as a source of data as well as a preferred academic output. With respect to diaries historically, this reflects the technologically determined trend of handwritten personal diary writing, producing discrete paper based ‘books’ by particular authors, and the current associations linked to diary writing remain tied to this technique. Current written diaries do not have to conform to this handwritten approach and can make use of word processors instead – if suitable for the diary writers. The author used the latter approach with international PhD students in a qualitative methods course and it certainly gets around the issue of transcription very nicely! The use of alternative diary techniques has recently grown within the social sciences more generally. These alternative techniques refer to video, audio, photography, illustration, and blogs and are discussed in turn in the following sections.

Video Diaries

Video can be used in different ways but primarily the distinction is between how and where the equipment is utilized and the purpose of the method. One approach is to provide participants with a self managed or supervised video recorder which they take home with them (or take elsewhere) and record diary entries in their own time. Participants are provided with tape, or with digital recorders and are expected to return the equipment and data to the researcher. This method may be suitable with a range of participants but particularly with children who might find this a user friendly method that allows their voices to be heard. A second approach is to establish and designate a ‘diary room’ in a fixed location whereby participants visit the room and are filmed within the confines of that space. Again, although appropriate for a variety of participants, the TV show 'Big Brother' style of this approach may chime with children and teenagers' sense of purpose. This approach has the advantage of providing a pre given and 'private' location, but would be less useful for research which required data on other places. Geographers should consider the purpose of using video technology for diary entries and reflect on the advantages of using a video camera in both public and other private spaces, such as at home, or on the bus, or at work. Given our disciplinary intrigue with the complex human environment, using video diaries outside of a prescribed research setting may be very fulfilling – this all depends on the nature of your research aims.

Audio Diaries

Like video, audio diaries can be used in various ways to suit the aims of the research. Here participants are pro vided with a digital or tape based recording device which they can either take with them, and record their verbal thoughts over a period of time, or they can attend audio recording sessions in a fixed location (e.g., an interview room) and record diary based entries in this manner. This technique can be used to explore participants' perceptions of events, people, or places and offers them the flexibility of using a fairly discreet device at any time they choose. This approach also offers relative anonymity for the participant, although some may feel uncomfort able verbalizing their thoughts. It does however resolve issues associated with literacy which occurs with written diaries.

Photography

The use of photography within geographical research is rising. Photograph taking does not necessarily perform the role of 'diary making' in itself, although it could do, but rather photography can be used in conjunction with written or other diary making. Photography is particu larly useful when one is researching the nature or ma teriality of place or the significance of what a place, object, or person looks like for the participant. In add ition, photography can provide the researcher with valuable visual data and insight into areas inaccessible to researchers, possibly because of privacy issues (like a home) or dangers associated with the setting (such as violent informal settlements). Key to the success of this technique is providing the participants (photograph takers) with the opportunity to explain and describe their photographs after they have been printed. These verbal or written explanations can then be used to further il lustrate the meanings of the photographs. Whether or not photography in this sense can strictly be described as a diary technique is open to debate and really depends on what one understands a diary to be. But certainly photographs can be used to further stimulate or replace written diary accounts and may provide participants with opportunities for creative expression previously denied to them. Participants may also find this technique more enjoyable than other diary methods.

Illustration-Based Diaries

Like photography, the use of drawing within geography has growing respectability, particularly within social, cultural, and development geography. Drawings can be used as pictoral diaries in themselves, or as illustrations informing and elaborating on written or other diary ac counts. Drawings are commonly used with children (but definitely not exclusively) as a way of encouraging them to communicate their stories and interpretations about events, people, and places. The issue of analysis of il lustrations is key here and it is important, as with pho tography, to encourage participants to explain and describe their interpretation of their illustration. In addition, it is crucial for the researcher to have a strong awareness of the context within which the illustrations were made. A lack of awareness of context may mean that a researcher over analyzes the significance of particular illustrations and, in doing so, neglects other interpret ations. This is an important reminder indeed for all diary methods, not simply illustrations.

Blog-Based Diaries

A very recent form of diary keeping is that of blog writing. Blogs are websites which are generated by the user (i.e., anyone who chooses to do so) and entries are written or placed in a journal style and normally dis played in reverse chronological order. A brief history of blogs on Wikipedia reveals that since their first use in the mid 1990s their popularity as online diaries has dra matically escalated since this time with Technorati (a search engine) tracking approximately 60 million users by the end of 2006. The commentary of many blogs is highly geographical in nature with many focusing on recent political concerns and local news. Although most blogs are largely written, many users also contribute blogs in other formats such as video, audio, and pho tography. Evidence of the use of blogs for geographical diary research is not widely available in published jour nals, but given its popularity it is likely to form a very useful methodological tool in the future. Students who are interested in using blogs in their work should run a search on the internet viewing the range and possibilities of blogging. A key advantage of using blog based diaries is that users can be scattered across the globe, they can maintain their anonymity, and the resultant text usually will not need transcription as it will already be typed. Relying on computers is however highly exclusionary.

Issues Arising Out of Using Different Techniques

It is all too easy to be seduced by the benefits of tech nology without thinking through the implications care fully. All diary keeping techniques have pros and cons attached to them and geographers must consider very carefully the consequences of the choices they make in terms of sampling, the experience of the diary keeper, and the form of data produced.

As explained above substantial numbers of potential participants will be excluded from techniques such as video, blogging, and word processed diary keeping be cause of issues of cost, access to electronic equipment, and technological skill competencies. All of these issues are particularly relevant when researching or working with participants on low incomes or in areas of high poverty and low technical skills or access such as the global South. In these contexts using handwritten diaries with participants is a very cheap and accessible method as it simply requires the purchase of a cheap paper booklet and a pen or pencil on behalf of the researcher. However, written diaries may prove more costly for the researcher during analysis and transcription, as typing up diary entries is very time consuming. Barriers of cost, access, and ability can be overcome and participants can be taught to use different technologies. It is up to the researcher to factor in and invest time in the training of participants to use the equipment. Equipment can also be purchased by the researcher to cover the costs and hence broaden the scope of individuals who could be involved. The cost of disposable cameras for example is not that significant and training participants in their use is fairly straightforward.

Alternative technologies hold advantages too. Blog ging may allow a researcher access to diary keeping over a potentially global scale, broadening ones understanding of processes dramatically (although possibly narrowing the depth of understanding). Video, audio, illustrations, and photography may provide participants with literacy concerns an alternative means of expression which they may find very empowering. Literacy as an issue should be taken seriously and not just in the global South. Re searchers must assess their participants' abilities carefully and be sure to support and offer advice on ways of 'getting around' literacy concerns without humiliating their participants. Illiterate or semi literate participants could be encouraged to use alternative technologies (such as audio diaries) or they could make use of a friend or family member to complete their writing for them – this obviously requires a level of trust between participant and scribe, and may compromise the privacy and thus openness of diary entries.

Solicited or Unsolicited? A Crucial Distinction

All diaries whether written or other can be separated into unsolicited or solicited diaries. They are quite distinct from each other and each form presents a host of ad vantages and concerns.

Unsolicited Diaries

Within geographical research, unsolicited diaries, or diaries which were not written at the request of a re searcher or other party are often historical in nature. These diaries were often written for private purposes and have been made available (often long after the writer is deceased) for academic interest or general consumption through publication. There is a strong tradition with geography of using historical diaries in research and students interested in this form of research should con sider the work of historical geographers for inspiration and guidance. Historical unsolicited diaries can provide exceptional insight into the daily lives, experiences, and thoughts of a range of participants allowing the re searcher exclusive insight into life at that point in time. This has been very popular with feminist geographers who have broadened and challenged the discipline's understanding of its own history by bringing the accounts of female travelers and explorers into the disciplinary domain.

With unsolicited diaries, similar concerns are raised as when using archival material. Researchers may require permission to use historical diaries and access may be restricted, costly or delayed. Ethical issues also arise when using material specifically intended for private use. However, historical diaries are intensely rich sources of material, offering a unique insight into lives and contexts where little is known. In general, unsolicited diaries are usually highly qualitative in nature with diary writers using a handwritten format to express often rather poetic prose.

Solicited Diaries

These types of diaries require a researcher to specifically request the production of a diary (written, video, audio, illustration, photography, or blog) from a participant or group of participants. In other words, the person writing or making the diary account is doing so in full knowledge that the finished product is for external consumption. The production of the diary is thus negotiated between the researcher and the diary maker. Furthermore, be cause the diary has been requested by a researcher it is very likely that its content will reflect the concerns that the researcher wishes to read, for example, a researcher might ask participants to keep diaries of their daily travel patterns or of their weekly leisure activities, or of their experiences of violence. Because of this it is important to remember that the diary is not necessarily evidence of a participant's total point of view or of their interpretation of what is and what is not most significant to them. Rather the diary reflects the focus that the researcher has re quested. If left up to the participants they might have written about their experiences of eating or their child hood memories rather than their experiences of violence. Because of this the diary that is produced for the re searcher is not a private document. It may contain some private thoughts but it has been written for a researcher to read and this will explicitly shape what the participant chooses to write. As a result of this process, solicited diaries embody the subjectivities of both the producer and the researcher.

Forms of Solicited Diaries

Solicited diaries are growing in popularity within geo graphical research; however, their forms are highly var ied. If one views solicited diary keeping as a continuum, the forms of diaries vary from 'strictly chronological brief entries' (a more quantitative form) to 'loose prose verging on personal story writing' (a more qualitative form). At the quantitative end of the spectrum, diaries are often used to research people's travel patterns or their daily or weekly activities such as medicine taking, social habits, or use of particular places or facilities. Researchers using this approach would provide high levels of information and 'instructions' to participants to ensure that the cor rect and required information was recorded. These diaries are usually recording factual information, al though they may provide some space for reflective re cording too. At the opposite end of the spectrum, more qualitative diaries are used for recording participants' thoughts, feelings, views, and experiences of a range of concerns, such as violence, health, sense of identity, sense of place, etc. These diaries can be highly unstructured or semi structured in style and may include a range of narrative strategies (or ways or writing or recording). For example, participants may write in a story format, or even use poetry to articulate their thoughts. Geographers should select the form of diary most appropriate to their respective research aims.

Researcher Diary Keeping

A third type of diary is that of researcher's keeping their own personal diaries during or throughout their research projects. Again this is a poorly documented activity. Personal diary keeping for researchers may be motivated by a range of reasons such as keeping records of research activities, or recording personal feelings and emotions experienced as part of the research process. The former is really covered by what we term 'field note taking', whereas the latter is often an activity undertaken for personal and cathartic reasons. Researchers who adopt a strongly ethnographic and reflexive approach to their project may find personal diary keeping an essential part of their research project, and indeed may make use of their diary entries in their analyses and publications. The use of this material in publication or academic output may include the diaries of both researchers and research assistants. These diary accounts often provide important insights into the methodological practices shaping the overall argument of the research, for example, why par ticular methodological choices were made and not others.

The Politics of Diary Keeping – The Potential for Empowerment?

Diaries, particularly unsolicited personal diaries, are typically assumed to provide empowering and cathartic relief to their writers, as so many of those who have kept a personal diary would understand. It is also possible for solicited diaries to provide their writers/diary makers with therapeutic benefits as well as forms of empower ment. The likelihood of this happening is more common when the production of a diary allows the participant to experience one of two things. First, it may allow the participants to share and thereby relieve themselves of the burden of memories associated with events which may have been painful for them and which they have not shared with any or many others. Examples here may include experiences of trauma, such as violence, un employment, eviction, public shame, or events over which they feel deep guilt or embarrassment, such as theft, failure to follow advice, etc. A second experience which may facilitate empowerment is when the diary production reinforces or illustrates for the participant the fact that they are achieving or managing an event or issue more successfully than they had realized. For example, if a participant is asked to record in a diary their patterns of sociability, and they begin the project feeling reclusive and socially deskilled, a consistent record of interaction and efforts to make interaction may prove that indeed they are more socially able than realized. This may have a positive impact on their sense of self.

Diary keeping then can be employed by geographers who have an explicit agenda for empowering their par ticipants by offering them the opportunity to express and explore often deeply hidden thoughts and feelings. As celebrated by feminists, qualitative methodology which both gives participants a voice as well as empowers them offers a unique method for practicing progressive re search. However, empowerment is not a simple quest in research and should be thought through very carefully.

Geographers should make use of feminist texts which outline the principles of feminist research and what it means to empower before embarking on this approach. In addition, the complexities associated with 'progressive research' must be carefully understood well before research begins. These relate to the researcher's skills and the potentially destabilizing implications of enacting research which can have impact on participants socio-political lives.

Geographers should also not assume that using diaries is necessarily an empowering technique. If research methods can be emotionally progressive then so too can they be emotionally regressive. Diary keeping may indeed prove to be highly destabilizing and upsetting for participants, and researchers should reflect carefully on the ethics of causing unnecessary distress. Distress may also be experienced upon receiving, reading, viewing, or hearing individuals' diary accounts. Essentially, one is gaining a window into someone's life and it is thus not surprising that this can reveal highly emotive material.

Combining Diaries with Other Methodological Approaches

Geographers and other social scientists often use diaries in conjunction with other methods and there is good evidence that combining methods is intellectually productive. Various researchers combine diary keeping with interviewing and there is evidence that combining solicited diaries with focus groups, individual interviews, and photography can prove insightful. Diary keeping can also be used in conjunction with more quantitative techniques such as GIS or statistical analysis so that information recorded in diaries can be correlated with either spatial or numerical references.

Diaries can be used to explore in far more detail issues raised in focus groups for example, or alternative techniques such as photography can be used to illustrate or deepen diary entries. Different methods can be used to explore related but different research questions within a single project, or the different methods can be used to support each other and further validate the findings. Indeed, combining methods in this way is not exclusive to diary use, and the benefits for triangulation of using a mixed methods approach are well documented. From a participants point of view, different methods often provide them with more or less comfortable (or interesting) forms of engagement between which their responses can vary quite widely. Researchers should bear in mind however that diary keeping can be very time consuming for participants and this effort must be considered carefully. In some projects, payment for diary keeping may prove necessary.

Interpretation and Analysis

Analysis of data is often overlooked in methods discussions. There can be no single prescription for analyzing diaries because the form of analysis entirely depends on the form of diary used, its original purpose within the broader research project, and the theoretical approach of the research. Written diaries which fall under a more 'qualitative' label are more likely to be analyzed (once transcribed) using a textual analysis approach whereby themes and narrative trends are coded (by hand or using packages such as Nudist or NVivo). After coding, interpretation occurs and coherent stories are derived from the material. An appreciation of the context within which the diary was produced is fundamental to the analysis and the researcher recognizes their subjective role in the meanings derived from the material. Qualitative diary keeping – particularly written – raises some very real issues around truth. How does one assess the 'truth' of what the participants have written? This is a very difficult issue to gauge within social science in general and the advice would be to aim to triangulate research wherever possible so that stories and events can be checked. Using a combination of techniques is often helpful here. This is however not simply a problem for diary analysis, but all methods in general – how can one ever know if the participant is telling the truth. In some situations what participants do not say or write, can be equally revealing.

Diary material which is visual in form, such as video, photography, or illustrations, requires the use of carefully thought through techniques of analysis. Fortunately, there is very useful advice available to geographers about how to proceed with such analysis and researchers should seek the work of visual geographers to take this further. The interpretation and analysis of more quantitative diary forms normally make use of basic descriptive statistical techniques to facilitate understanding. Diaries can be subjected to an auditing procedure where particular events, places, or people can be counted (such as forms of public transport used in a week, or visits to a hospital). Unlike more qualitative diaries, researchers using quantitative diaries need to carefully explore their sampling procedures and techniques of analysis before drawing conclusions. However, all researchers using diaries must be critically aware of the ways in which they are analyzing their data, what (and who) is omitted or excluded from their data, and recognize the limitations of their arguments as a result. Indeed, feminist geographers would stress that the very choice of what sections of a person's diary a researcher chooses to include in their writings and what is excluded is a political decision in itself and this is a clear example of how diaries are constructed both by the writer and, through research and analysis, by the soliciting researcher.

See also: Archives; Autobiography; Auto-Photography; Ethical Issues in Research; Feminist Methodologies; Longitudinal Methods (Cohort Analysis, Life Tables); Mental Health; Mixed and Multiple Methods; Participatory Video; Representation and Re-presentation; Subjectivity; Time-Space Diaries; Transcripts (Coding and Analysis).

Further Reading

Bell, L. (1998). Public and private meanings in diaries: Researching family and childcare. In Ribbens, J. & Edwards, R. (eds.) *Feminist Dilemmas in Qualitative Research: Public Knowledge and Private Lives*, pp 72 86. London: Sage Publications.
Blunt, A. (2000). Spatial stories under siege: British women writing from Lucknow in 1857. *Gender, Place and Culture* 7(3), 229 246.
Dodman, D. (2003). Shooting in the city: An autophotographic exploration of the urban environment in Kingston, Jamaica. *Area* 35(3), 293 304.
Elliott, H. (1997). The use of diaries in sociological research on health experience. *Sociological Research Online* 2(2). http://www.socresonline.org.uk/2/2/7.html (accessed on February 2008).
Kwan, M. P. (1999). Gender, the home work link, and space time patterns of non employment activities. *Economic Geography* 75(4), 370 394.
McGregor, J. (2006). Diaries and case studies. In Desai, V. & Potter, R. (eds.) *Doing Development Research*, pp 200 206. London: Sage.
Meth, P. (2003). Entries and omissions: Using solicited diaries in geographical research. *Area* 35(2), 195 205.
Meth, P. and Malaza, K. (2003). Violent research: The ethics and emotions of doing research with women in South Africa. *Ethics, Place and Environment* 6(2), 143 159.
Noyes, A. (2004). Video diary: A method for exploring learning dispositions. *Cambridge Journal of Education* 34(2), 193 209.
Punch, S. (2001). Multiple methods and research relations with children in rural Bolivia. In Limb, M. & Dwyer, C. (eds.) *Qualitative Methodologies for Geographers: Issues and Debates*, pp 165 180. London: Arnold.
Rose, G. (2006). *Visual Methodologies: An Introduction to the Interpretation of Visual Materials* (2nd edn.). London: Sage Publications.
Thomas, F. (2007). Eliciting emotions in HIV/AIDS research: A diary based approach. *Area* 39, 74 82.
van Blerk, L. and Ansell, N. (2006). Imagining migration: Placing children's understanding of 'moving house' in Malawi and Lesotho. *Geoforum* 37, 256 272.

Relevant Websites

http://www.shef.ac.uk
Clare Rishbeth at Walking Voices, ESRC Project, University of Sheffield.

Diaspora

K. McKittrick, Queen's University, Kingston, ON, Canada

Glossary

Creolization Creolization describes the coming together of diverse cultural traits or languages. Often the consequence of contact, violence, and colonialism, creolization identifies a new or different (rather than subordinate/dominant) cultural practice that encompasses the social and political 'relations' of two or more ethnic groups. Unlike hybridity (see below), creolization identifies the renewed sociospatial consequences of a violent past.

Displacement The process of being removed, exiled, or marginalized. Geographically, displacement often refers to expulsion from a region, nation, or community. In social theory, displacement can occur due to racism, sexism, homophobia, or other acts of discrimination and violence that subordinate particular communities, even though these communities may not have the means to escape the discrimination.

Homeland A territory to which one belongs was born in, or is tied to ancestrally. Also understood as a 'country of origin', a homeland couples ethnic groupings to a particular nation that has hosted their community in the past or the present.

Hybridity Used in social theory to challenge essentialism, hybridity denotes the ways in which different cultures, ethnic groups, and geographic regions enmesh to produce a 'mixed' or 'fused' identity, place, culture, or area.

Migration, Migratory Subjects Any movement of humans from one location to another. Migratory subjects points to the experiential tenets of migration that, in diaspora theory, can refer to forced expulsions, colonial displacements, or economically induced movements – all of which tend to be long migrations.

Nation Often used to demarcate the geographic location of a particular ethnic group, nation refers to a social, political, economic, and historic territory that a government manages and protects.

Introduction to Diaspora: Exiles and Displacements

Derived from Greek verb *diasperien*, 'to sow over' or a 'scattering or sowing of seeds', diaspora is used to refer to the scattering of people across spaces, borders, and na tions. Historically, ethnic and racial traumas, as well as geographic losses, have underwritten the forced or induced movement of people. Specifically, diaspora in dicates that particular groupings of people have been violently forced to leave their homeland because they are racially, ethnically, and/or religiously different from those in power; forced expulsion might be coupled with ethnic cleansing. Jewish exiles, for example, have been regarded as forced scatterings instigated by anti Semit ism, genocide, and war. Indeed, the Jewish diaspora spans several centuries due to ongoing, and global, anti Semitic intolerances and ethnic struggles, including but not limited to: expulsion from Palestine in roughly the third century BC; the Crusades; the Spanish Inquisition; pogroms inside and outside of Eastern Europe; the holocaust and Shoah; and historic and contemporary conflicts over Israel and its attendant regions.

In addition to physical violence that targets and expels a particular ethnic group, diaspora points to the question of home, nation, and location. Lack of a stable nation space and geopolitical independence, and transnational dispersals, can shape the possible desire to establish and secure a location that can replace former geographic losses. In the twentieth and twenty first centuries, par ticularly after World War II, Shoah, and the creation of Israel in 1948, both Jewish and Middle Eastern com munities experienced violence and dispersal, which also instigated assertive claims to home, nation, and ethnicity. Importantly, the establishment of an Israeli homeland and the subsequent conflicts have also initiated a Middle Eastern diaspora, with communities from regions such as Palestine and Lebanon enduring violence, war, and dis placement. At the same time, forced exile and the loss of homeland must be understood alongside the region wherein the diasporic subject relocates. In the case of the Jewish diaspora, for example, formal and informal links with Israel, or former European communities occupied and lost during Shoah, are frequently developed within new spaces of settlement in North America, Africa, Australia, South America, and Asia. Holocaust/Shoah memorials, which have been constructed in various sites internationally, are evidence of how the violent dis placement and genocide that occurred during World War II are remembered vis à vis multiple diasporic locations.

Questions of home, nation, ethnicity and violence have also been used to clarify the African or black diaspora. Between the fifteenth and nineteenth centuries, African peoples were violently exiled during the trans atlantic slave trade. Expulsion and bondage had black subjects sold and transplanted into the Americas and Europe, and geographically disconnected from their former African homelands. Indeed, under transatlantic slavery, forced geographic displacement was coupled with sociocultural ruptures: languages, family genealogies and lineages, religious practices, and cultural, political, and economic norms were intended to be, and in many cases were, eroded and expunged from the lives of African slaves. Like the Jewish diaspora, the history of black diaspora is entangled with centuries of violence wherein the question of homeland and geographic loss shapes how belonging and nationhood are experienced within the new region of relocation. More specifically, diaspora carries with it the experience of human removal, death, and displacement; settling after such exile carries with it, rather than completely erases, the history of removal, death, and displacement. For African or black subjects, relocation in the Americas is, at times, understood through their diasporic relationship to the middle passage, the continent of Africa and/or transatlantic slavery. While debates over this history and the experience of enslavement have changed over time – some details forgotten, some memories too unbearable to return to, some generations rethinking the past and their relation to it – it is a history, some black subjects argue, that should not be forgotten.

While forced removal, exile, and resettlement structure the aforementioned histories, additional communities have also taken up the term diaspora to better understand displacement and international migrations. The Armenian community has experienced a long history of displacement and exile. Centuries of ethnic and religious subordination, culminating in the massacres of 1915–16, reveal that the Armenian diaspora must be understood alongside mass genocide and displacement. Discrimination, war, poverty, colonialism, ethnic cleansing, and violence have removed and moved Irish, Asian, South Asian, Caribbean, Eastern European, and Middle Eastern groups from their countries of birth; these movements are often, but not always, to regions such as North America, Australia, and Central Europe. Indeed, the violence of colonialism in many nations triggered diasporic exiles as the homeland, reimagined and managed as a colonial space, was rendered unfamiliar to its indigenous communities. These processes – war, ethnic cleansing, colonialism, and various genocides – paved the way for thinking about diaspora more broadly. Furthermore, recent investigations into gay, lesbian, transgendered, and bisexual communities have considered the possibility of queer diasporas: those displaced from a homeland, as well as spaces of resettlement, due to homophobia and heterosexual gender conventions. As queers are always already outside the heterosexual nation, queer diasporas explore the ways in which ethnic migratory struggles intersect with sexual tenets of displacement, home, and nation.

Additionally, economic migrations between over developed regions – Australians to Central Europe, to give just one example – have also been labeled diasporic cultures, albeit with much debate regarding whether or not the economic privilege and the individual (rather than collective) travels of the migrants constitutes a ‘real’ diaspora scattering. Regardless, recent discussions of diaspora, it has been argued, do not always replicate the histories and experiences of Jewish and black communities in that mass genocide and/or collective exile are not necessarily instigating travels. Often coinciding with post World War II and postcolonial global migrations, these dispersals have shown that diaspora can be voluntary, involuntary, or both. The context of these migrations underscores the ways in which globalization and transnationalism inform and complement – but do not twin – diaspora experiences. Unlike transnationalism and globalization, which encompass broad economic global flows, such as the cross national and international movement of technologies, economies, ideologies, consumer products, as well as people, diaspora centers on the condition of those humans who are displaced. Put simply, diaspora is a ‘human phenomenon’ whereas transnationalism and globalization bring into focus multiscalar economic patterns. This is not to suggest that diaspora does not have an economy, but rather that diaspora migrations are folded into, but do not replicate, processes of transnational globalization.

Implicit to the human emphasis on diaspora migrations are returns. The process of returning, whether imaginary, real, filmic, or fictional, distinguishes diaspora activities from other migratory processes. Thus, in what ways do diaspora communities return to their former homes and nations and in what ways are returns possible? Do diaspora cultures return through travel, remembrance, imagination, remittances, yearnings, stories, or songs? Are past cultural practices retained upon relocation and do these cultural practices carry in them the history of violence, dispersal, and memories of home? As noted above, the creation of Israel in 1948 geographically marked a homeland that some Jewish subjects could, and do, return to. In the African or black diaspora, homelands in Liberia and Ethiopia, as well as the ‘door of no return’ on Goree Island (the site where African subjects were kept in bondage before crossing the middle passage), have been imagined and returned to through travels, writings, music, and films. Political ideologies such as Afrocentrism, Negritude, and Pan Africanism produce links between and among global black cultures. Many diasporic communities return through economic remittances: from the nation of relocation, funds are sent ‘home’ to assist family and friends or to secure future departures. Additionally, many returns arise through cultural production: films that take up departures or violence; novels that remember slavery; poetry that explores forced human migrations; Internet sites that map holocaust memorials and record the lives of survivors; songs that sample the music of Bob Marley, and are then played on local radio stations, globally; live performances of Ghidda in the UK or Canada. These processes illustrate an expanding understanding of diaspora as complex exchanges between former colonies, migratory subjects, and new areas of settlement are maintained. Diasporic returns can therefore be real, imaginary, political, economic, and creative. These texts, memories, politics, and economic activities foster what has been called 'imagined communities': vacations, travels, and permanent returns can be understood alongside novels, letters, media, memories, films, songs, economic remittances, fashions, and foods, that cross borders and contribute to the meaning of homeland for those living outside its borders. All of these processes – migrations, memories, textual invocations of home and nation – also reveal that diasporic subjects make implicit or explicit connections among those who have been dispersed.

If diaspora fundamentally centers on the movement of people, are not all traveling cultures diasporic? If departure is voluntary, or individual rather than collective, is this a diasporic process? Do all diasporic subjects return to their homeland? While all diasporas are historically contingent to the specificities of time and space, and therefore differentially experienced, four geographic themes arise through which diaspora can be conceptualized and expanded upon which distinguishes it from simply describing a multitude of traveling cultures and individuals:

1. Exile of a substantial population based on ethnicity, race, religion, or difference that is often coupled with violence and/or genocide.
2. Loss of or displacement from homeland.
3. The mutual construction of identity and place as they are understood by migratory communities.
4. Returns.

These four themes, together, impart what is referred to as the 'routes and roots' of diaspora experiences. Routes explicate the forced exile, formal and informal connections generated among those dispersed, and imaginary and real links to homeland; roots identify former homelands, collective histories, and the relationship between identity, nation, and ethnicity.

Diasporic cultures are not simply traveling and migrating with ease. Rather a particular event or events – poverty, violence, war – has initiated flight and this movement is difficult psychically because it is wrapped up in loss and different types of return. These experiences are collective, often creating explicit or unspoken ties between and among the group that has been exiled. This is not to suggest that all diasporic subjects can or do return to their country of birth, or that they cling to the emotional moment of past violent exile; nor does it suggest that all diasporic cultural practices are embedded with nostalgia for homelands and psychically relying on the lives of others exiled to understand their history and selfhood. Rather, diaspora can be a powerful migratory process through which exiled communities formulate their collective identities not as victims, but rather as subjects who negotiate, transgress, and fracture the boundaries of and between nation, territory, and culture.

Diaspora movements and cultural practices give rise to very difficult human migrations and border crossings that inform how exile and territorialization (the jurisdiction and regulation of land and its citizens by a particular government) underscore alternative understandings of modernity, nation, identity, and place. Herein lie the complexities of diaspora theory, migrations, and cultures: they delineate exiles and travels but do not assume a clear or stable 'beginning' and 'end'; they site and cite human violence, but do not explicitly locate a region of liberation or freedom at the end of diasporic travels; they are concerned with nation, but are critical of borders, nationalism, and the limits of the nation state citizenship; they imagine homelands and returns, but do not assume that 'home' is either returnable or innocent; they take seriously the links between identity and place but, because diasporic subjects are migratory, put forward that identity and place are neither stable nor naturally entwined.

Modernity

While different theorists have conceptualized modernity in different ways, in terms of the diaspora, this epoch signals transnational flows beginning in the fifteenth century (colonial expansion, exploration, transatlantic slavery, indenture) which were crystallized in the seventeenth century during European Enlightenment when multiscalar social hierarchies began to shape economic, political, cultural, and social systems inside and outside imperial nations. European imperial expansion, the shift away from feudal political systems, the development of the modern nation state, and new international economic profits illustrate some of the ways in which systems of knowledge became increasingly based on categorization in the name of rationality, progress, and democratic liberalism. These systemic changes entrenched a corporeal hierarchy based on sexuality, race, ethnicity, phenotype, and gender wherein the white man symbolically came to represent and embody democracy and rationality. Concurrently, empirical analyses and inventories of plants, insects, animals, women, the poor, and the insane were conducted as were positivist mappings of European, indigenous, and colonial physical and human geographies. It follows that the underpinnings of transatlantic slavery, and the various colonial projects in the Americas, Asia, South Asia, and Africa were, in part, either classification missions designed to 'discover' and ultimately transform 'primitive' or 'nonrational' cultures or 'enlightened' justifications for racial–sexual subordination. This kind of classification, of bodies and cultures, stabilized racial differentiation; white and nonwhite ways of knowing bifurcated while encoding the world according to Eurocentric ideologies.

Modernity starkly draws attention to colonialism, genocide, and transatlantic slavery and the racial, sexual, and economic exploitation of subaltern subjects. The movement of African subjects to large and small plantations in the US, Canada, and the Caribbean; the technological underpinnings of the Holocaust/Shoah during WWII (trains, gas chambers, work camps); the colonial displacement and disappearance of Aboriginal cultures; imperial educational requirements in the colonies. These are just some events that suggest that modernity is a violent project. Thus, some of the central arguments brought forward by diasporic (and Aboriginal) communities are regarding the violence of modernity. What is at stake, they ask, when bondage, genocide, violence, and colonialism are equated with progress, enlightened reason, democracy, and egalitarianism? In what ways do diasporic subjects negotiate modernity if, in its various implementations, it negates their worldview and figures, even prefigures, them as irrational, uncivilized, and worthy of expulsion and bondage? And finally, if modernity often instigates diasporic flight and exile, what do these migratory communities bring to bear on the formation of the nation state? These critiques and understandings of modernity accentuate the tensions between identity and place, in particular the complex understandings of nation, identity, and location that diasporic subjects impart.

Nation

While modernity is often equated with democratic citizenship, technological progress, and new ways of organizing the world, these geographic, political, economic, and intellectual shifts also validated the expulsion of particular groups from their country of origin.

As noted above, the question of violent displacement reveals that diaspora is entwined with the expulsion from homelands or home nation. The nation is central to theorizing diasporas in that it clarifies the ties between migrations and the production of space. That is, forced dispersal carries with it the idea of 'displacement from' somewhere and 'displacement to' somewhere, and these travels necessarily effect how diasporic communities understand themselves in relation to the nation. While the formation of modern nation states often coincides with the production of ethnic communities and racial identities within a prescribed country or region (Vietnamese live in Vietnam, Germans live in Germany, Algerians live in Algeria), diaspora experiences illustrate the ways in which the nation, as a political bounded entity that safely houses and supports its citizens, is constantly being breached. This occurs in two ways. First, the homeland or nation through which expulsion takes place (displacement from) discloses that some citizens are not, in fact, welcome or at ease within the boundaries of their country of birth. Second, arrival in the nation of relocation (displacement to), as a refugee, economic migrant, or exiled subject, is often fraught with ambivalence due to barriers such as immigrant/refugee status, noncitizenship, and racism or other acts of discrimination.

What these two points suggest to diaspora theorists is that the nation and nationalism are limited geographies, ideologies, and processes. More clearly, the nation that expels particular communities because they do not 'fit in', and the nation that scrutinizes new immigrants and refugees, reveals that nationalism and the production of the nation state are determining the meaning of citizenship, culture, sexuality, and race in exclusionary ways. It follows that diasporic subjects who exit and enter different nations on a global scale, and whose travels are underwritten by discrimination, poverty, and/or violence, have a sense of nation and homeland that is complex and shifting rather than stable and permanent. To put it another way, diasporic subjects are not attached to the nation in any straightforward way. Indeed, while nation and homeland are crucial to conceptualizing the diaspora, the ambivalent relationship to these particular geographic regions reveals that diaspora subjects do not necessarily embody an essential attachment to their country of birth or nation of relocation; rather diasporic subjects, are constantly challenging and transforming ideologies that assume an uninterrupted connection between nation, ethnicity, and identity precisely because they come to express their shared histories and exiles through heterogeneity and difference. This does not position diasporic subjects outside of modernity. Rather, these migratory and displaced subjects are demonstrating the contours of the modern project and their intellectual contributions to this epoch. Diaspora theorists insist that the experiences of removal, travel, and returns are indicative of how roots and routes – indeed diaspora spaces – intersect with, and therefore are indicative of, modern geographies. A concrete example of this intersection can be illustrated through the black diaspora: people of African descent, living in, say, Britain, arrived in this nation due to transatlantic slavery, colonialism, and/or induced (economic) migration; they maintain ties with other global black communities, experience discrimination and success within Britain, and also call into question the naturalness of a white/British nation space. Together, these processes of displacement, outer national ties, and settlement demonstrate the workings of modernity and the modern nation not as bounded or unchanging, but as a territory that is inflected with difference.

Identity

Identity refers to one's self affiliations, specifically one's culture, nation, gender, sexuality, class, and so forth. Diasporas bring into focus the ways in which identity is produced in relation to collective histories, migrations, modernity, and nation. These themes and affiliations are developed in three ways. The first conception of identity centralizes collective histories. As noted above, discrimination and violence push some groups outside their country of birth. These diaspora migrations are indicative of a common history, wherein particular groups understand themselves and their community in relation to a traumatic past. In this case, identity is secured to the social and cultural codes that initiated expulsion, as well as the region through which the displacement occurred. Identity brings into focus both difference and continuity. Understood in this way, a two way process that is bound up in modernity arises: categorization and expulsion of subaltern communities; reunifying diaspora cultures through exploring their commonalities and collective experiences of expulsion. Identity in this formulation attempts to recuperate or restore a broken past. During anticolonial struggles, for example, theorists, artists, and novelists in or from South Asia and the Caribbean brought forth representations of difference that explored violent displacement (transatlantic slavery, indentureship, colonialism) and also considered how masses of people might, together, relate to this history vis à vis their common identities and contemporaneous political struggles against oppression. In short, diaspora identities were and are soldered as a counternarrative to Eurocentric modernity and the nation state that regard some cultures as absolutely 'other' or wholly different from the hegemonic class.

The second conceptualization of identity is a much less stable or unifying theorization and considers the ways in which diaspora identities change from moment to moment and place to place. It is precisely the historic and geographic break that refuses an easy relationship between identity and place for diaspora cultures. The original geographic loss, what some refer to as the moment of traumatic rupture, cannot be fully recovered because the implementation of modern Eurocentric social systems work to deny and dislodge, rather than recover, the past. At the same time, diaspora cultures are developing their identities according to the exclusionary citizenship policies within the nation of relocation, sometimes using counternarratives to develop a unifying cultural critique of modernity. Both of these processes suggest that the production of a diaspora identity can only be understood as fluid and changing: there is no satisfactory relationship with historic regions of displacement; it is difficult to fully belong in a nation of relocation. It is useful, then, to consider that diasporic cultures are geographically migratory and that this shifting spatialization produces complex identities that are unsettled. Even exiles that occurred during transatlantic slavery can be understood as migratory in that this brings into focus the ways in which black diaspora cultures continue to develop their identities simultaneously through and away from the continent of Africa and/or moment of rupture even as they continue to experience discrimination within the nation of relocation (racial profiling, segregations, incarcerations, poverty).

Understanding diaspora identities as fluid and shifting, but also as developing imaginary coherences across time and space, allows for analyses of how migratory cultures impact upon the meaning of modernity and nation. Neither outside modernity nor the nation, diaspora identities are often conceptualized as hybrid or creolized. Both of these terms are attempts to clarify cultural identities that are, through their displacement, negotiating the surroundings of their region of relocation and their ethnic ties in complex ways. The third conceptualization of identity thus considers the ways in which collective histories, migrations, and ethnic differences are embedded in, and evidence of, modernity. Through diaspora identity, it is not possible to separate European and non European versions of modernity. Hybrid or creolized diaspora identities enmesh different cultural practices, intellectual histories, social norms, that cannot be traced to a unitary ethnic homeland or nation or a pure ethnicity. Instead, diaspora identities combine and creolize multiple points of identification, thus embodying and asserting how diversity and difference are lived and experienced as a process of entanglement. To give a few examples: Indo Caribbean men and women educated at Oxford University; African American intellectuals theorizing Sigmund Freud and Frederick Neitzche alongside Frantz Fanon, Ralph Ellison, and Toni Morrison; dragon boat races in Toronto, Canada. Importantly, this particular production of diaspora identity is not a one way process, with subaltern subjects borrowing and absorbing the dominant culture. Rather this formulation also points to the ways in which diaspora cultures impact upon their surroundings and other diasporic cultures: white, Aboriginal, and Asian youths wearing hip hop fashions; black feminists embracing Latina feminism; queer dragon boat races.

Diaspora and Geography

While discussions of diaspora have been fostered in cultural studies, literary criticism, and critical race studies, human geographers have, in the last two decades, begun to integrate this field into their analyses. Diaspora is, many human geographers have noted, a fundamentally geographic process as it is underwritten by exiles, displacements, and regional specificities. Many geographic investigations use the term diaspora as a descriptor for movement. Rather than exploring questions of identity, modernity, violent exile, collective histories, and ambivalent nation making, diaspora is simply synonymous with migration and immigration. However, those geographers interested in the ways in which critical race theory, trauma, displacement, and identity shape and are shaping the production of space, have engaged the work of diaspora theorists outside the discipline of human geography in order to advance a unique conceptualization of what is often called 'diaspora spaces'.

The major critique of diaspora and diaspora studies is focused on the highly metaphoric understanding of this migratory experience. Because diaspora is primarily theorized through the production of identity, rather than a materiality of displacement, many geographers have noted that questions of hybridity, imagined communities and returns, and collective histories lack geographic specificity and political economy. Indeed, some geographers are frustrated with the unrepresentability of the psychic themes in diaspora studies. Thus, these human geographers have worked on spatializing the theoretical paradigms that are debated in diaspora studies. Close attention is therefore paid to the material production of home and nation in order to clarify how diaspora spaces might be produced within and beyond the nation of relocation; details, for example, regarding how diaspora subjects talk about home and nation – in their three dimensionality and geographic specificity – are coupled with emotional and psychic processes. Some human geographers also investigate the materiality of modernity, specifically positivist mappings, and consider how this process initiates difficult spatial conditions of which diasporic exile is a part. Homeland, therefore, is theorized as simultaneously material and metaphoric in order to bring into focus how diaspora spaces are produced through violence. Additionally, human geographers have paid close attention to the ways in which some global migrations are also diasporic: housing and immigration patterns, school curricula changes that include subaltern histories, racial profiling, 'multicultural' food distribution, ethnic enclaves, are just some examples of the ways in which diaspora cultures impact upon the materiality of local life. It has been argued, however, by diaspora scholars that this critique of the unrepresentability of diaspora experiences is hasty. Some diaspora scholars insist that questions of memory, psychic attachment, past exiles, and collective histories always carry in them the materiality of displacement. That is, through the language of diaspora, one must always imagine the three dimensionality of expulsion in order to fully grasp the assertion that diaspora is a human phenomenon.

See also: Ethnic Conflict; Memory; Migration, Historical Geographies of; Race.

Further Reading

Anderson, B. R. O. G. (1983). *Imagined communities : Reflections on the origin and spread of nationalism*. London, England: Verso Editions/NLB.

Anthias, F. (1998). Evaluating 'Diaspora': Beyond ethnicity. *Sociology* 32(3), 557 580.

Brah, A. (1996). *Cartographies of diaspora: Contesting identities*. London: Routledge.

Blunt, A. (2003). Geographies of diaspora and mixed descent: Anglo Indians in India and Britain. *International Journal of Population Geography* 9, 281 294.

Carter, S. (2005). The geopolitics of diaspora. *Area* 37(1), 54 63.

Clifford, J. (1994). Diasporas. *Cultural Anthropology* 9, 302 338.

Cohen, R. (1997). *Global diasporas: An introduction*. Seattle: University of Washington Press.

Davies, C. B. (1994). *Black women, writing, and identity: Migrations of the subject*. London: Routledge.

Gilroy, P. (1993). *The black Atlantic: Modernity and double consciousness*. Cambridge, Mass.: A Harvard University Press.

Gopinath, G. (2005). *Impossible desires: Queer diasporas and South Asian public cultures, perverse modernities*. Durham: Duke University Press.

Hall, S. (1990). Cultural identity and diaspora. In Rutherford, J. (ed.) *Identity: Community, culture, difference*, pp 222 237. London: Lawrence and Wishart.

Habib, J. (2004). *Israel, diaspora, and the routes of national belonging*. Toronto: University of Toronto Press.

Lowe, L. (1996). *Immigrant acts: On Asian American cultural politics*. Durham: Duke University Press.

Patton, C. and Sánchez Eppler, B. (2000). *Queer diasporas*. Durham: Duke University Press.

Shilav, Y. (1993). Ethnicity and geography in jewish perspectives. *GeoJournal* 30(3), 273 277.

Walcott, R. (2000). Pedagogy and trauma: The middle passage, slavery, and the problem of creolization. In Simon, R. I., Rosenberg, S. & Eppert, C. (eds.) *Between hope and despair: Pedagogy and the remembrance of historical trauma*, pp 135 151. Rowman and Littlefield: Lanham.

Yeoh, B. S. A. and Willis, K. (1999). 'Heart' and 'wing,' nation and diaspora: Gendered discourses in Singapore's regionalisation process. *Gender, Place and Culture* 6(4), 355 372.

Dicken, P.

J. V. Beaverstock, The University of Nottingham, Nottingham, UK

Dicken, Peter (1947–)

One of the world's leading economic geographers, Peter Dicken spent nearly four decades teaching and researching the geographies of industrial restructuring and regional change from a global perspective, based in the University of Manchester's School of Environment and Development. In 1988, he was awarded the personal title of Professor, followed by several distinguished awards, including a Fellowship of the Swedish Collegium for Advance Study in the Social Sciences (1999), the Victoria Medal of the Royal Geographical Society with the Institute of British Geographers (2001), and an honorary doctorate of philosophy from the University of Uppsala, Sweden (2002). During his long reign at Manchester, he has been an academic visitor at universities in Australia, Canada, Hong Kong, Mexico, Singapore, Sweden, and the US. From the mid 1980s onwards, with the publication of *Global Shift: Industrial Change in a Turbulent World*, Dicken can be credited for putting economic globalization firmly on the map in geographical thought. His influence has subsequently become apparent in studies of economic and employment change in major global industries (such as automobiles, electronics, textiles, pharmaceuticals, and logistics); industrial restructuring and organizational behavior in transnational corporations through the analysis of foreign direct investment flows and business strategy; and global production networks and territorial redevelopment. In short, Dicken's research and writings have brought a 'global perspective' to studying the geography of industrial restructuring and territorial redevelopment at the regional and local scale.

Summarizing, Yeung and Peck suggest that there are four very distinctive contributions that Dicken has made to the study of human and economic geography, and globalization. The first of these concerns firm spatial decision making – externalization of control and regional development. In this respect, Dicken was highly influenced by a 'behaviorist turn' in economics and management sciences in the early 1970s. Dicken began to put the action of the firm and its externalities of control under the microscope to explain the geography of location and regional development. But, perhaps the most profound step in Dicken's reputation was the analysis of multinational enterprises (nee corporations) decision making activities in location and industrial restructuring. Dicken's work became an exemplary for studying the behavior of multinational and transnational corporations in the context of the world economy recession of the post 1973 era, where externalities of control and the branch plant phenomenon were dominant structures of economics in the rapidly changing industrial world.

During the 1980s, Dicken's second contribution – and probably most significant contribution to human geography – was the first edition of *Global Shift: Industrial Change in a Turbulent World*. Since 1986, there have been a further four editions of this book published (1992, 1998, 2003, and 2007). *Global Shift* remains one of the key textbooks in economic geography. The 1986 edition introduces and discusses the economic geographies of industrial change in the world since the end of World War II from the global scale perspective. Dicken argued that the global shifts that occurred in many industrial sectors of the world economy were produced by three significant intertwining processes, namely: the rapid growth, proliferation, internationalization, and geoeconomic power of transnational corporations worldwide; the changing role of national government through patterns of international trade, investment, regional development policies, and macroeconomics; and the impact which informational and communication technologies have had on transport (e.g., containerization), the organization and geography of production, employment practice, and firm locational strategy.

Global Shift as a textbook has been of paramount importance for students of economic geography. It explained the processes of international restructuring through the actor of the transnational corporation in several key sectors of the world economy: textiles and clothing; iron and steel; motor vehicles; and electronics. The four subsequent editions of *Global Shift* have added new sectoral chapters on contemporary topics, such as the internationalization of services, distributive industries, and agro business. But moreover, *Global Shift* was viewed by the wider academy as 'one of the pioneering globalization texts'. Dicken skillfully advocated the global–local dialectic through an exploration of 'global' interconnectedness in economic processes and resultant uneven development in different geographical time–spaces.

Dicken's third major contribution to geography concerns his analysis of firms, nation states, and global networks. Indeed, through the 1980s and beyond, Dicken continued to research and publish widely on the interrelationship between the transnational corporation and the nation state in making the economic geography of regions and the locale. Dicken continued to argue that national governments did advocate agency in global ization in their interrelationships with firms, regulators, and other actors. This actor focused approach to explain the processes and patterns of global economic re structuring and change intersected very much with the economic sociologists notions of embeddedness and so cial action. Dicken was very quick to highlight the 'webs of enterprise' that made up embedded, relational inter and intrafirm organizational networks in contemporary transnational corporations. Dicken, therefore, was highly influential in bringing the study of transnational cor poration's organizational networks onto the agendas of economic geography from the early 1990s.

Fourth, and into the new millennium, Dicken with colleagues in his own department, at Manchester Busi ness School and the National University of Singapore, brought to the attention of geographers, economists, and sociologists, the theoretical framework associated with relationality and global production networks to explain territorial redevelopment brought about by transnational corporations' inter and intrafirm networks of relation ships. The notion of the firm–place relationship is central to Dicken's thesis on global production networks, where such relationships are essentially wrapped around four processes of relationality: intrafirm; interfirm; firm–place; and place–place. Concepts of social action and economic embeddedness are very important facets of these rela tionships especially when theorized in the scalars of globalization, from the local to the global.

A final and very important point to make in summing up Dicken's contribution to the academy is that his work is fundamentally interdisciplinary in scope and scale. Dicken's seminal work, *Global Shift*, makes interventions in the worlds of organizational studies, strategy, and political economy. But, moreover, *Global Shift*, in com bination with his latter work on *Global Production Networks*, makes a significant contribution to understanding the complexities and unevenness of globalization *per se*. Dicken is not only a significant economic geographer of our time, but is also a respected commentator on globalization.

See also: Economic Geography; Foreign Direct Investment; Global Production Networks; Globalization and Transnational Corporations; Globalization, Economic; Industrial Location.

Further Reading

Coe, N., Hess, M., Yeung, H. W. C., Dicken, P. and Henderson, J. (2004). Globalizing regional development: A global production networks perspective. *Transactions of the Institute of British Geographers* 29, 468 484.

Dicken, P. (1971). Some aspects of the decision making behaviour of business organizations. *Economic Geography* 47, 426 437.

Dicken, P. (1976). The multiplant business enterprise and geographical space. *Regional Studies* 10, 401 412.

Dicken, P. (1980). Foreign direct investment in European manufacturing industry. *Geoforum* 11, 289 313.

Dicken, P. (1986). *Global Shift: Industrial Change in a Turbulent World* (1st edn.). London: Harper Row.

Dicken, P. (2004). Geographers and globalization: (Yet) another missed boat. *Transactions of the Institute of British Geographers* 29, 5 26.

Dicken, P., Kelly, P. F., Olds, K. and Yeung, H. W. C. (2001). Chains and networks, territories and scales: Towards a relational framework for analysing the global economy. *Global Networks* 1, 99 123.

Dicken, P. and Miyamachie, Y. (1998). From noodles to satellites: The changing geography of Japanese sogo shosha. *Transactions of the Institute of British Geographers* 23, 55 78.

Dicken, P. and Thrift, N. (1992). The organization of production and the production of organization: Why business enterprise matter in the study of geographical internationalisation. *Transactions of the Institute of British Geographers* 17, 101 128.

Lloyd, P. E. and Dicken, P. (1972 and 1976). *Location in Space* (1st and 2nd edn.). New York: Harper Row.

Yeung, W. C. H. and Peck, J. (2003). Making global connections: A geographer's perspective. In Peck, J. & Yeung, W. C. H. (eds.) *Remaking the Global Economy*, pp 2 23. London: Sage.

Relevant Websites

http://www.sagepub.co.uk
SAGE publications: *Global Shift.*

http://www.sed.manchester.ac.uk
School of Environment and Development, The University of Manchester: Peter Dicken, geography staff.

Difference/Politics of Difference

T. Jazeel, University of Sheffield, Sheffield, UK

Glossary

Assimilationist Policies or political strategies that promote the merging and absorption of traits from previously distinct groups into a dominant civil group.

Citizenship Commonly conceived as a condition bestowed upon individuals considered as a member of a particular society or nation-state, wherein social rights, privileges, and duties are bestowed upon that individual.

Différance Concept derived by the French post-structural theorist, Jacques Derrida, that describes how the meaning of a (linguistic) sign is always deferred through other signs in relationships of irreducible relationality and intertexuality.

Identity Commonly regarded as a sense of self or community, providing sameness, continuity, and collective membership over time and at different geographical scales.

Liberalism A modern social and political theory founded on beliefs in the natural goodness of humans, autonomy of the individual, civil and political liberty, and freedom from tradition.

Subaltern Commonly used to refer to a person or proposition in a position subordinate to a dominant. In social theory used in conjunction with a body of work that outlines non-Euro-American modes of political resistance and knowledge.

Introduction

Difference and sameness are most commonly evoked in Human Geography in relation to identity theory. Over the last three decades or so, post structural theory has challenged Marxist analyses that read the production of difference through notions of the class and class consciousness alone that Marxists believe fall out of the different stages of society's economic development. In stead, post structural theory has influenced social the orists to explore a broader range of registers through which human difference is produced, and through which politics are lived and mobilized, for example, gender, race, sexuality, and disability. Post structuralism has also reasoned that the politics of identity and difference exist in conditions of relationality to one another. For there to be something that people/communities at any scale can identify with, there must be a constitutive outside. Sameness requires difference. Geographers' particular contributions have been to explore the particular spatialities and spatial languages that difference and the politics of difference not only occur in, but also are central to their very production.

That difference relies on a geography of sorts emerged strongly in Edward Said's work on Orientalism in the late 1970s. For Said, the two geographical sectors known as the 'Orient' and the 'Occident', or 'West' and 'East', are not only human and discursive (imperial) productions, they are geographical imaginations that support and to an extent reflect each other. The rela tionship of the 'Orient' to the 'Occident' is constitutive, that is, the 'Occident' is produced through the difference posed by that designated and simultaneously produced as 'Orient'. Difference then is a fundamentally spatial re lationship. Geographers working in the wake of Said have helped to destabilize any sense of a straightforward re lationship between place and identity by pointing to the differences within as well as between spaces and places. The development of a vocabulary of ideas around dias pora, transnationality, borderland, hybridity, centers and margins, and third spaces have helped to variously tease out just some of the multiple spatialities through which difference is produced at different scales, and through which a politics of difference is mobilized.

Liberalism, Difference, and the Politics of Difference

Post structural understandings of difference commonly work from particular understandings of the liberal pol itical subject and its relationship to collectivity. The neutral citizen of liberal theory Iris Marion Young argues is coded white, male, bourgeois, able bodied, and het erosexual. It is this pervasive coding of the liberal citizen that marks modes of non white, female, disabled, or homosexual subjectivity as different within liberal dem ocracies. But as stressed above, these modes of difference themselves help (re)produce the liberal templates of normal citizenship. The one depends on the other as a condition of dynamic relationality. In his early theories of post structural deconstructivism, Jacques Derrida fam ously referred to this relationship as 'différance'. The word, derived from the French verb 'différer' which can mean to either differ or to defer, is used to covey a sense that meaning is always merely the product of the dif ference between signs. Meaning is always deferred, elsewhere, like in a dictionary, or indeed this encyclo paedia. In identity theory then, difference can only be understood in relation to the normalizing codes of liberal citizenship and vice versa, but importantly 'différance' posits this as a relationship without end, never complete. Subject positions are always being made and remade relationally.

Within liberal democracy the play between the normalcy of liberal citizenship and its constitutive outside has produced powerful narratives and politics of the emancipation or assimilation of minorities. Assimilation to the liberal norms of European citizenship emerged as a conduit for rendering different modes of subjectivity transparent, legible, and safe. Within colonial encounters, for example, this kind of assimilatory politics of difference emerged as colonialism's emancipatory pretension to 'civilize the native'. In contemporary identity politics, state authority and the principles of reasonable governance can be seen to promulgate emancipation or assimilation through debates over the 'tolerant' or 'good' society, or even through culturalist understandings of state sponsored and ordered cosmopolitanism. However, both these models of emancipation and assimilation have never meant the dissolution of forms of difference in liberal culture. Far from it.

Crucially, in terms of theorizing the politics of difference, for many minority groups the effacement of difference through 'emancipation' has never been a political objective, and for many social theorists it has never been a realistic (let alone ethically desirable) political aim. Instead, for minority groups the politics of difference evokes the struggle to mobilize forms of resistance to the social exclusions suffered as a result of the group's difference. In relation to colonial encounter specifically, Homi Bhabha's influential body of work has exposed the contradictions inherent in colonial discourse itself. For Bhabha, the assimilationist or emancipatory effort within colonial encounters produces spaces and subjectivities in which borders are straddled, edges and horizons are blurred, and where differences bleed into one another, ultimately producing subversive and hybrid newness. The traces of irredeemable and recalcitrant difference pose immanent threats to colonialism's (or more generally, liberal political culture's) capacity to contain and naturalize difference as minorities appropriate the normalizations of 'civility', 'liberty', 'citizenship', 'progress', and 'development'. Difference always remains as a space of political resistance and/or trace within liberal political subjectivities.

That difference does not exist *a priori* has led scholars to explore a wide and wavering net of interlocking yet specific historical and cultural processes through which it is (re)produced. For human geographers the spaces and spatial interconnectedness through which difference is produced have become the subject of a rich seam of scholarship over the last few decades. Perhaps most obviously, difference has been explored in relation to nationalism, broadly conceived as feelings of collective identity which may be associated with the nation state or other regional scales in connection with 'national' and territorial formations. But geographers inspired by the post structural manoeuvres sketched above and by feminist theory have also moved to reject forms of territorialized essentialism in their spatial explorations of difference. In addition to the (usually) territorial spaces of national belonging, they have also explored the 'in between', interstitial and metaphorical spaces through which identity and difference are produced, and through which the contours of a politics of difference are discernible. Geographical work on these marginal spaces and subjectivities has also emerged as an effective way of understanding the forms of power that emerge from the heart of liberal social and political cultures and systems. The spatial production of race, gender, sexuality, and disability are perhaps the most common and effective registers through which human geographers have explored the spatial politics of difference, and it is to a discussion of these that this entry now turns. But the effort of doing so should not be taken as an inscription of race, gender, sexuality, and disability as discrete or essential categorizations through which identity politics and subjectivity can be understood. Rather contemporary geographical research has focused on understanding how each category of difference emerges through particular social and spatial relationships, and how they intersect with one another through lived experiences that resist essential categorization.

Race

The value of race as a scientific concept has long been repudiated, but it continues to be deployed as popular marker of human difference based on bodily characteristics: skin color, hair type, nose shape, and muscular definition. It is also used commonly to refer to people who share cultural or historical backgrounds. Despite these popular usages of the concept, critical social research has more or less come to the conclusion that the division of human beings into discrete races is not in any sense natural, but rather a social, political, and cultural construction. Equally, attempts to implant racial hierarchies within and across societies have been regarded as inseparable from cognate political projects (like colonialism or nationalism) rather than any kind of decoding of the natural world. Indeed, racial theorizations and instantiations of difference have underpinned some of the most repressive political regimes and violent geopolitical contestations in recent history (Nazism, Apartheid, civil war in Sri Lanka, genocide in Rwanda, for example).

These developments in ways of critically engaging the concept of race have seen geographers over the last 30 years or so turn from mapping the distribution and segregation of raced groups across space, to critical explorations of the processes involved in the racialization of particular people or groups. In other words, the mainstay of contemporary critical geographical engage ments with race has focused on the different spatialities through which the logic of race comes to be lodged in the body. These kinds of explorations have treated race as the lingering aftereffect of racist gazes; racism, as a violent set of representational practices, precedes race in this an alysis because racism in its various forms constitutes ways of mapping the body. By mapping the body, geographers have referred to the ways that racist acts, representations and policies (Social Darwinism, anti semitism, apartheid) inscribe the biological illusion of race onto the body.

The range of spatialities that geographers have iden tified with the mapping of raced bodies varies immensely, from landscapes, parks, and other spaces of nature; through diasporas, border regions, and transnational spaces; news media, entertainment, and cinematic spaces; to the institutional spaces of genealogical research, DNA code identification, and territories and flows regulated by the increasing prevalence of biopolitical technology. These kinds of explorations are likely to increase in scope and diversity as theorizing racialization demands that geographers confront the plethora of everyday and socially embedded spaces through which the materialities and politics of racial difference percolate and structure our world. Such research has shown that while race's value as a scientific concept has been widely rejected, it continues to wield significant and constitutive power as a structural force in the world today. Importantly, it is not just the non white, non European or non North American body/community that has been theorized in such work. Human geographers have also paid attention in recent years to 'white ness' as a specific racialized and relational category that is itself produced through social, political, and communitarian discursive practices that have their own spatial formations.

Recent critical geographical research since the de struction of the World Trade Center on 11 September 2001 and the subsequent 'Global War on Terror' has also noted a politics of difference increasingly divided po lemically and ideologically between a (neo)liberal 'West' coded white, secular and civilized, and a global Islamic nation coded non white, sacred and – according to the recurrence of pervasive enlightenment logic – 'uncivil ized' and violent. In addition to the tendencies to fold difference into distance in ways remarkably consistent with Said's Orientalizing discursive practices, geog raphers and other social theorists drawing lessons from Auschwitz during World War II have noted the existence of 'exceptional' spaces within the 'West' (Guantanamo Bay, for example) or 'Western' enclaves (Abu Graib) in which bodies are routinely interned without due legal process on the very basis of their religious and perceived racial difference, and in which all human and legal rights to the equal and humane treatment of human beings are suspended. Here, amidst the 'Global War on Terror', the performative symbolism of this type of spatiality of ex ception maps the Muslim body as potential terrorist, in turn marking the Muslim body as different, as always outside the codes of liberal, Western political subjectivity and collectivity. The essential logic of these kinds of practices cannot be operationalized without enabling forms of neo imperial and neo liberal, geopolitical power relationships, which reminds us that racialized practices of marking out human difference always walk hand in hand with power.

Gender

Despite a rich history of feminist activism and scholar ship across the world, gender roles and relations as markers of human difference grounded in biology con tinue to be commonplace ways of conceptualizing re lations within and across societies. In Western European philosophy, the gendered distinction between masculine culture and feminine nature that has been mapped onto man/woman predates Cartesian dualism, and can be found in Pagan and Christian beliefs in women's attu nement with the natural world through their child bearing capabilities or menstrual cycles. But feminist scholars point to Descartes' sixteenth century declar ation, "I think, therefore I am", as the birth of scientific, rational and cultural man whose mastery over nature (matter) also entailed controlling and subverting women. The social and political effects of objectifying women along with nature have subjugated women in society more generally, be it through pornography, women's subordination in the workplace, or the relative inequality of political opportunities available to women and men.

Contemporary feminist identity politics then seeks to articulate understandings of the social and patriarchal production of gendered relations of human difference by teasing out the discursive and performative instantiation of dominant types of femininity and masculinity. Femi nist geographers have accordingly tended to move away from explorations of gender roles for fear of re essen tializing the difference between feminized and mascu linized categories of reproductive and productive labor, respectively. Instead, they have moved toward explor ations of how relations between men and women are constructed through the spheres, and moreover, spatial ities of everyday life. Over the last 20 years or so, there has been a rich seam of feminist geographical scholarship that has explored the problems that inhere in the di chotomous and gendered relationship between a mas culine public sphere of economic labor, most often equated with formal notions of Politics, and a feminized private sphere of cultural reproduction implicitly coded as apolitical. Space – particularly those distinctions between notionally public and private spaces – therefore emerges as a central register through which the gendered production of human difference must be understood. In addition to problematizing the authority of this gendered and sexist binary geography, this work has also reclaimed the political dimensions of unpaid, 'feminized' labor in the private sphere.

More broadly, however, this effort to remove women from nature and place them within culture has often entailed the strategic conceptual separation of sex and gender. In other words, if gender is a cultural production, then logic dictates that sexually humans can be biologically differentiated into male and female. The problem with this maneuver is what to do with sexual difference. Treating gender and sex as separate in this way repeats the problem of a nature/culture binary by re essentializing sexual difference just as gender is emancipated through theories of its cultural production. Here, Judith Butler's early work has been particularly influential among post structural and feminist geographers. She argues that regarding heterosexuality as a regulatory regime is key to understanding how the body, sexed categorically as either male or female, is brought into being. That in between male and female sexes cannot exist in Western society without significant psychological trauma is not so much a biological fact as it is an effect of the power of the discursive regulatory regime of heterosexuality within society. Neither gender nor sex then are ontological pre givens according to Butler. Our designations as male or female emerge from discursive categorizations, and gender becomes an effect of one's stable (discursive) sexual identity as either male or female that is iteratively reinscribed on the body through performances that themselves are gendered as either masculine or feminine behavior. For geographers these insights have facilitated the exploration of a whole host of activities that occur in public and private spaces respectively that (re)inscribe gendered and sexualized understandings of difference.

Sexuality

The theoretical links between gender and sex mean that the two are commonly discussed together. However, geographical research on sexual identity and difference, while not exclusive of heterosexuality, has developed to pursue sophisticated explorations of the relationships between lesbian, gay, bisexual, and transgendered space and subjectivity.

Early gay and lesbian activists emphasized the authenticity of their stable, core sexual identities, therefore rooting their modes of sexual difference within biology. This intuitive hostility to constructionist accounts of identity and difference emerged not just from feelings that many individuals could not have been anything but gay, but also from distinctly strategic concerns around how to mobilize an effective homosexual politics of difference within hegemonically straight society and space. Biological explanations of sexuality for some still offer the best form of defense against forms of homophobia that suggest gays can voluntarily change their sexually 'immoral' behaviors.

The emergence of queer theory, and its incorporation as queer geography within the discipline, represents a paradigm shift and reaction against this uncritical early work on stable lesbian and gay identities. Challenging notions of biologically fixed hetero or homosexual identities, queer theory and queer geography adopts Butlerian ideas on the iterative and performative, sexualized subjectivities and gendered relations that normalize and stabilize sexual identities. Particularly important to the queering of understandings of sexual as opposed to gendered difference, however, is Michel Foucault's work on the *History of Sexuality*, which famously points to the normalization of heterosexuality as a regulatory regime in Western society since the seventeenth century. Instead of the repression of talk about sex and sexuality within an ostensibly conservative society, Foucault's rigorous historical research demonstrates the proliferation of talk and scientific investigation into the right type of (hetero normative) sex. Out of this discursive proliferation and production of scientific inquiry into and about the right types of reproductive and hetero orthodox sex, (homo)sexual minoritization emerges as stigma, abnormality, immorality and confession. Crucially, homosexuality emerges as difference only through the normalization of heterosexuality.

The application of these ideas through spatial registers has allowed geographers to explore how sexual difference is produced through particular spatializations of hegemonic discourses of sexual virtue, respectability, and reproductive coupling. Geographers have explored how the geographical theaters of everyday life constitute crucial technologies that effectively discipline the sexualized body, regulate sexual desire, and safeguard hetero normative maculinities and feminities. Queer emerges as a social positioning against these muted backdrops of the proper. Similarly, queer spatialities and spatial metaphors, like 'the closet' for example, emerge in irreducible connection to these dominant hetero normative spatialities. In terms of theorizing a sexual politics of difference geographers have also explored how gay, lesbian, bisexual, and transgendered people's subversive treatment and negotiation of forms of hetero normative spatiality constitute forms of resistance, or attempts to erode the terms of 'proper' space that produce heterosexuality and queerness.

Disability

Despite the predominance of medical models of disability within the medical profession, in social theory the trope of social construction has become increasingly common place in theorizations of human difference rooted in physical and/or mental impairment. Within human geography specifically, scholars have emphasized the importance of understanding disabled people's geo graphical experiences as a way of understanding how disablement is produced through the impaired body's negotiation of social and spatial norms. These social and geographical models of disability have theorized disability as the societal, attitudinal, or spatial restrictions placed on physically or mentally impaired people, to the point where they become marked as different because of their inability to fully integrate themselves in liberal societies.

Access to, and mobility through, built and natural environments emerges as central to the geographical study of disability. For example, Rob Imrie's work on the relationships between disability and the city has effec tively shown how building and home design continue to oppress physically and mentally impaired peoples' rights to the city. And others have shown that it is not just building design that produces disability as human dif ference. There are a dense network of institutions and agents within liberal socio political structures that con dition disabled people's experiences of their immediate environments. Geography emerges as a constitutive part of disabled people's oppression and marginalization in society.

What emerges from much of this contemporary geo graphical work on disability is an implicit sense of the pervasive presence of da Vinci's 'Vitruvian Man' as the structuring motif, the building block, within the cultural and political organization of liberal, Western society. It is unsurprising then that the body has emerged as central to this field of enquiry, and in this sense explorations of human difference marked as disability have enabled productive connections with feminist scholarship. One of the effects of this conversation has been the broadening out of what modes of difference can conceptually be referred to through the rubric of disability if social ex clusion results from degrees of difference from an un marked liberal and masculine, bodily norm. The liberalization of culture and space in public places like the shopping mall or cinema, for example, has been shown to produce pregnant or 'fat' bodies as a form of disablement because of the relative difficulty with which those bodies are able to negotiate such environments.

Given the predominance of social theories of dis ability within human geography, it would be wrong to suggest that medical theories of, and treatments for, various forms of physical and mental impairment have become irrelevant. Yet, in terms of the politics of dif ference disabled people have been just as involved in efforts to change the social norms that pervade their built and natural environments, as they have in seeking med ical treatment and assistance for impairment. The overriding concern has been that disabled people's needs are catered for and they can enter society fully no longer marked as different. Philosophically, however, in terms of the politics of difference, there is a tension between charitable approaches to disability that attempt to ob literate, or emancipate, disability as difference, and dis abled rights conceived as access to society. The struggle for disabled people to enter fully into society as citizens, it is argued, should not be confused with any need to eradicate disability as a marker as difference. The former is a human, social, and political right, whereas the latter reflects a liberal society's inability to deal with any form of human difference. Crucial here then is the physical agency required for disabled people to bring themselves into representation.

Difference and (Geographical) Knowledge

Any discussion of difference and the politics of difference as it has been worked through the (Western) academy must rub up against the academy's own encounters with difference and the important epistemological and polit ical questions that these encounters have generated. The main body of this entry has discussed markers of differ ence that have emerged through geography's theoriza tions of the production of identity. If, in this respect, post structuralist theory has challenged the singularity of the politics of class difference that emerges with Marxist analysis, then post structuralism and other forms of critical thinking have also posed important challenges to a rigid and formulaic Marxism's encounters with differ ent ways of being, knowing and doing beyond its epi stemological and geographical comfort zone.

In the 1980s, scholars within the field of South Asian historiography, now known as the Subaltern Studies collective, became vocally dissatisfied with orthodox European Marxist theory's economic determinism as it was thought through South Asia and elsewhere beyond the West. Although these intellectual battles were not fought within the discipline of Human Geography, they were extremely geographical because they involved thinking and working the politics of difference across a planet marked with what the Subaltern Studies scholars collective insisted are irreducibly 'different' histories, literacies and formations. In short, the Subaltern Studies movement contended that difference theorized through orthodox Marxism not only marks groups apart within a particular society on the basis of class alone, but its in tellectual premise could not help but theorize global difference through developmentalist, progressivist nar ratives that ground history in the unity of capital, and subsequently unrefined Marxism is wholly unable to account for the dynamic and improvisational modes of peasant political agency within South Asia.

In response, from 1982 onwards Subaltern Studies scholarship sought out self conscious breaks from the economic determinism of much Marxist scholarship by attempting to write histories from 'below' where sub altern classes are conceived as subjects in the making of their own history. But the Subaltern Studies movement never sought wholesale rejections of the purchase of an anticolonial, Marxist politics. Instead, by developing so phisticated theories of hegemony and power it sought revisionist perspectives that brought 'native' knowledge in South Asia into focus as a form of difference with its own distinct politics of resistance. In doing so, it also drew attention to the subjugating effects of Eurocentric and universalizing Marxist theory culture that folds dif ference into historicism. What emerged across the social sciences and the humanities then was a fecundity of otherness not just in terms of difference conceived as the différance of multiple subject positionings vis à vis identity explored in the first sections of this entry, but also the politics of difference conceived as the epi stemological challenge posed by the philosophical cer tainty of irreducible alterity elsewhere. Within geography these maneuvers have had profound effects on explor ations and theorizations of, for example, regional mod ernities, subaltern cosmopolitanisms, or ordinary cities.

Subaltern Studies itself, however, by attempting to allow the subaltern to speak, remains the subject of an important critique from within its own conceptual ter rain. Specifically, and in terms of political activism, Gayatri Spivak has questioned whether the subaltern figure can speak, so ventriloquised and caught in trans lation are they by the postcolonial (migrant) investigator. As she poses, the moot decipherment by another in an academic institution (or as she puts it, "a knowledge production factory") must not be too quickly identified with the 'speaking' subaltern. For Spivak, once the native informant comes into representation through the first world intellectual, once difference is rendered transparent as knowledge, subalternity moves elsewhere; the subaltern subject is irretrievably heterogenous. Philosophically, in terms of a politics of difference this poses the enormous challenge that there is always difference be yond that which our (or any) discipline can grasp. Per haps grasping that ungraspability is key to securing progressive and humble future research into the politics of difference.

See also: Body, The; Citizenship; Cosmopolitanism; Diaspora; Disability and Chronic Illness; Gay Geographies; Gender, Historical Geographies of; Hybridity; Identity Politics; Poststructuralism/ Poststructuralist Geographies; Queer Theory/Queer Geographies; Racism and Antiracism; Self-Other; Subaltern.

Further Reading

Bell, D. and Valentine, G. (1995). *Mapping Desire: Geographies of Sexualities*. London: Routledge.
Bhabha, H. (1994). *The Location of Culture*. London: Routledge.
Brown, M. (2000). *Closet Space: Geographies of Metaphor from the Body to the Globe*. London: Routledge.
Butler, J. (1993). *Bodies that Matter: On the Discursive Limits of 'Sex'*. London: Routledge.
Butler, R. and Bowlby, S. (1997). Bodies and spaces: An exploration of disabled people's use of public space. *Environment and Planning D: Society and Space* 15, 411 433.
Derrida, J. (1967). *Writing and Difference* (trans. Bass, A. 1978). London: Routledge & Kegan Paul.
Foucault, M. (1976). *The History of Sexuality I: An Introduction* (trans. Hurley, R. 1979). London: Allen Lane.
Guha, R. (ed.) (1982). *Subaltern Studies 1: Writing on South Asian History and Society*. Delhi: Oxford University Press.
Imrie, R. (1996). *Disability and the City: International Perspectives*. London: Paul Chapman Publishing.
Jackson, P. and Penrose, J. (eds.) (1993). *Constructions of Race, Place and Nation*. London: UCL Press.
Longhurst, R. (1998). (Re)presenting shopping centres and bodies: Questions of pregnancy. In Ainley, R. (ed.) *New Frontiers of Space, Bodies and Gender*, pp 20 34. London: Routledge.
McDowell, L. (1999). *Gender, Identity and Place: Understanding Feminist Geographies*. Cambridge, MA: Polity Press.
Said, E. (1979). *Orientalism*. New York: Vintage.
Spivak, G. (1999). *A Critique of Postcolonial Reason: Toward a History of the Vanishing Present* ch. 3. Cambridge, MA: Harvard University Press.
Young, I. M. (1990). *Justice and the Politics of Difference*. Princeton, NJ: University of Princeton Press.

Diffusion

L. A. Brown, Ohio State University, Columbus, OH, USA

Glossary

Culture Trait A single phenomenon seen as representative of a particular, typically ancient, culture which, as data, is used to trace the extent or influence of the culture. Examples include pottery as represented by pot shards, Bronze Age artifacts, cowry shell, counting systems, and tillage tools.

Diffusion Diffusion involves the propagation of a phenomenon in a manner such that it spreads from place to place, leading to maps of the phenomenon that differ from one time to another.

Diffusion or Simultaneous Invention in the Formation of Cultural Landscapes A point of contention for cultural geography, simultaneous invention posits that societies evolve through a similar set of stages, not unlike human beings going through childhood, adolescence, middle age, and old age whereas the diffusion approach argues that societies evolve by intermixing, cross-fertilization, and often, dominance or empire building.

Distance Decay or Friction of Distance Effect An effect which recognizes that interactions over space are generally attenuated according to the distance separating two bodies such that there is more interaction at closer distances and less at farther distances.

Empirical Regularities of Diffusion These include an S-shaped curve for diffusion over time, a neighborhood- or contagion-effect for diffusion across undifferentiated space, and a hierarchy effect for diffusion through an urban system.

Enabling Innovations An innovation that facilitates the diffusion of another innovation, or is required by another innovation. The gasoline station, for example, enabled adoption and diffusion of the automobile, and the adoption of battery-powered vehicles is inhibited by the lack of charging stations.

Innovation A new product, new idea, new technology, new organizational structure, new way of doing things, or a new phenomenon. In this context, new means new to a particular place by way of adoption, not necessarily that the innovation was created in the near past.

Innovation Diffusion, Adoption versus Market and Infrastructure Perspectives Adoption focuses on demand for the innovation, emphasizes a communications process, and assumes that all individuals have an equal opportunity to adopt. Market and infrastructure perspective emphasizes marketing and infrastructure creation as the process by which innovations are differentially made available to various population segments, focus shifts from the adopter to the structure within which individuals operate, and adoption is seen as a choice within a constraint set.

Mean Information Field (MIF) A component in Hagerstrand's Monte Carlo simulation model which is comprised of probabilities that represent the likelihood of information traveling from the center cell of a 5×5 grid to surrounding cells, and a representation of a local network of social communications.

Monte Carlo Simulation A computational algorithm that, having established conditions under which certain outcomes occur, relies on random numbers to determine how those conditions manifest themselves and the resultant outcomes. Relying on the law of large numbers, the exercise is repeated a multitude of times in order to better calculate expected outcomes. Applications date at least to the 1930s by physicists.

Resistance to Adoption It reflects an individual's hesitancy to embrace an innovation; it may stem from values that are inconsistent with adoption, that is, 'social' resistance, or it may stem from practicalities which make adoption difficult or impossible, that is, 'economic resistance'.

In human geography, diffusion involves the propagation of a phenomenon in a manner such that it spreads from place to place, leading to maps that differ from one time to another. Generically, the phenomenon is considered to be an innovation – a new product, a new idea, a new technology, a new organizational structure, a new way of doing things, or a new phenomenon. In this context, new means new to a particular place by way of adoption, not necessarily that the innovation was created in the near past. Hence, introduction of clean drinking water or aspirin to a remote Latin American village is considered to be an innovation in that context, even though these items have long been commonplace in Europe and North America. Further, phenomena studied under the rubric of diffusion cover an exceedingly wide range that includes the automobile and other transportation modes, broadcast and cable television, credit cards, cultural practices, epidemics or disease, family planning, farming techniques, frontier development, Internet and broadband availability, modernization in Third World settings, political movements, production practices such as the assembly line or just in time inventorying, shopping centers, urban ghettoes, and urban areas themselves.

While geographers have studied diffusion processes throughout most of this century, the focus of that effort has shifted over time, mirroring the broader concerns of the discipline as a whole. Many of these shifts have entailed dramatic departures from the prevailing norm of diffusion research and a change in the course of study. As a consequence, diffusion studies of the cultural geographer early in this century are hardly recognizable as being of the same genre as more recent studies which focus on the role of entrepreneurial agents in the diffusion of innovation.

This article provides a historical perspective on our study of diffusion, giving particular attention to developments set in motion by the work of Torsten Hagerstrand in the early 1950s. Six major sections frame the discussion. The first deals with the cultural geography tradition of diffusion study, an antecedent of the article's central focus. Attention then turns to the work of Torsten Hagerstrand, wherein both his operational and conceptual contributions are summarized. The next three sections trace research trends in the 25 plus years since Hagerstrand, focusing initially on efforts related to operational models of diffusion; then on Third World development as a diffusion process; followed by efforts related to the conceptual model of diffusion. The sixth section takes a broader view of spatial diffusion research in terms of a philosophy of science perspective and the overall trends of the geography discipline. A skeleton sketch of the ensuing discussion is presented in **Figure 1**, which summarizes the progress of diffusion research in geography.

The Cultural Geography Tradition of Diffusion Research

Diffusion has been the subject of geographic enquiry for most of the twentieth century, if not earlier. The spread of a phenomenon, idea, or technique throughout a population or region incorporates basic geographic elements of distance, direction, and spatial variation, and thus forms a valid field of geographic interest. The most immediate approach from which present day work descends was generated within the context of cultural geography, whose research goal was to identify environmental features characteristic of a given culture and if possible to discover what role human action plays or has played in creating and maintaining given geographic features.

'Culture' and the human or cultural 'landscape' are the main concerns, and diffusion studies are relevant insofar as they clarify questions concerning cultural origins, the spatial configuration of cultural areas, and culture related or culture influenced features on the Earth's surface. Carl Sauer is generally recognized as the central figure in this approach, which tended to be closely aligned with and complementary to concerns of anthropology.

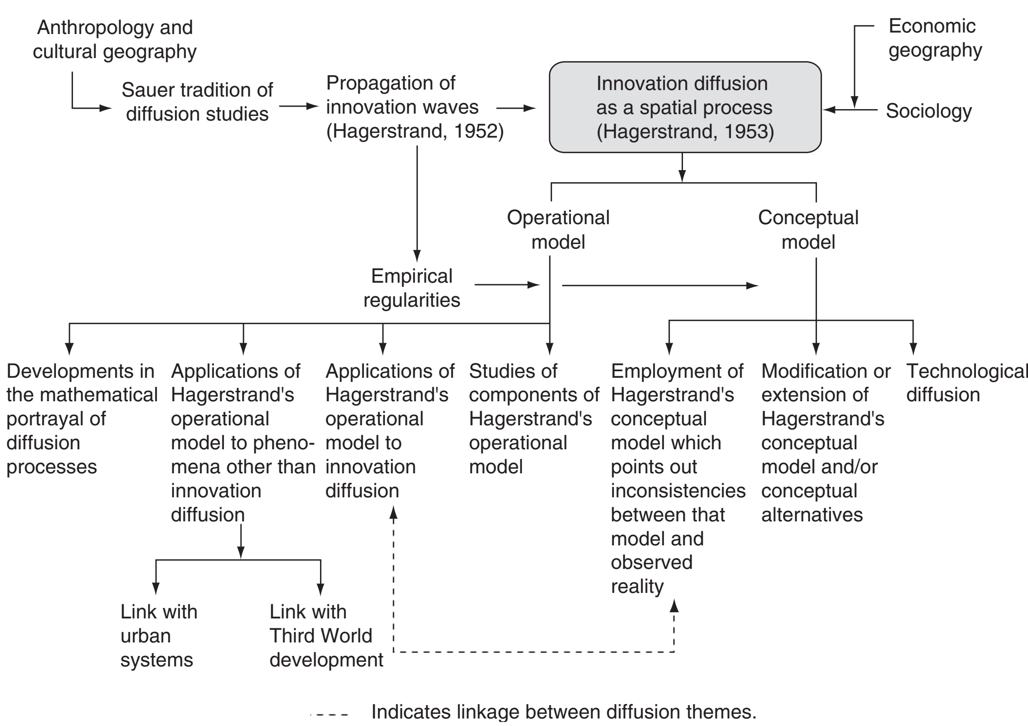

Figure 1 Diffusion research in geography: Themes and their interrelationships.

Many studies within this framework were formulated within the context of the controversy regarding the relative importance of 'diffusion' or 'simultaneous invention' in the development of the cultural landscape. They demonstrate a strong bias toward an empirical, inductive approach with emphasis upon landscape development in prehistory and over a considerable period of time. Often, the investigation considered a single phenomenon, which is seen as a 'cultural trait', and attempted to explain its observed spatial distribution as the outcome of a process unfolding over space and time. A typical procedure was to examine a phenomenon's spatial distribution at different times or its frequency of occurrence at different locations at a single time. From these observations, inferences are drawn regarding origins, means, and routes of dispersal of both the observed phenomenon and the culture traits, ideas, or techniques for which it is a surrogate.

To illustrate the point of contention here, simultaneous invention posits that societies evolve through a similar set of stages, not unlike human beings going through childhood, adolescence, middle age, and old age. Concerning society overall, for example, Hans Bobek posited (1) a food gathering stage; (2) specialized collectors, hunters, fishermen; (3) clan peasantry, pastoral nomadism; (4) organized agrarian society; (5) early urbanism and rent capitalism; and (6) productive capitalism, industrial society, and modern urbanism.

By contrast, the diffusion approach argues that societies evolve by intermixing, cross fertilization, and often, dominance or empire building. The culture trait provides evidence of this occurrence. It might be a type of pottery that directly links with one culture and not another, as evidenced by pot shards; a Bronze Age pin with a double spiral head found in Europe and Western Asia but traced to Troy and the Aegean Islands; cowry shells used for currency in West Africa which trace to Egyptian influences, counting systems, or tillage tools.

In stressing the role of diffusion in the development of the human landscape, many studies within the cultural geography tradition proceeded further to consider the indirect effects of the culture trait innovation on the physical and cultural environment. A notion basic to their discussion is that the diffusing item is both a stimulus to new innovation and itself subject to modification as it spreads from its point of origin. The relation between diffusion, the item being diffused, and the human landscape is therefore complex and subject to continual change. One result of this view was an emphasis on describing these ever changing relationships, rather than upon understanding the specific processes by which the item moves from one location to another.

Later in the twentieth century, this tradition came to focus on more contemporary society, but cultural artifacts remained in focus. Examples include studies of place names, house or barn types, language differences, and religion. Differences in each category were related to the dominance, or passing through of one societal group versus another. Considering language differences in the Northeast United States, for example, there are place to place differences in whether a running narrow body of water is referred to as a creek, brook, run, kill, lick, or stream; similarly, house types in Louisiana included built in porch, attached porch, open passage double log pen, single log pen, Midwestern, shotgun, bungalow, trapper, and oysterman types.

Hagerstrand on the Diffusion of Innovation

Sauer's counterpart in the research of primary concern to this article is Torsten Hagerstrand, whose work of the early 1950s initiated a shift in focus toward the generative processes underlying diffusion. Since this focus is not inconsistent with Sauer's conceptualization of diffusion, it is interesting to speculate why it was not a major topic of research at an earlier date. One reason may be that the identification and definition of specific processes controlling locational change are less feasible in the sparsely documented situations studied by cultural geographers. Second, and perhaps more important, the main impact of Hagerstrand's work was on researchers trained in economic and urban geography who were primarily concerned with location and locational processes rather than landscape. Whatever the reason, the shift was extensive. Illustrative of this is Peter Haggett's classification of research under the Sauer School as predominantly 'landscapist' or 'ecologist' and research under the Hagerstrand School as 'locationalist'.

Hagerstrand's research focused on the diffusion of manufactured innovations, unlike work under the Sauer tradition. His earliest English language publication, *The Propagation of Innovation Waves* (1952), is an evident continuation of the earlier cultural geography tradition (**Figure 1**) by its use of a primarily inductive approach and its concern with diffusion patterns rather than diffusion processes. However, empirical study at a somewhat later date led Hagerstrand to observe that "The spatial order in the adoption of innovations is very often so striking that it is tempting to try to create theoretical models which simulate the process and eventually make certain predictions achievable."

This statement is indicative of a change of emphasis in Hagerstrand's work from description and inductive generalization to a more deductive approach focusing on generative processes. Concomitant with this we observe a willingness to draw broad, general conclusions from one or a few empirical examples that did not characterize the cultural geographers. The basic work in this shift is

Hagerstrand's PhD dissertation, completed in 1953 and translated into English under the title *Innovation Diffusion as a Spatial Process* (**Figure 1**).

Hagerstrand made three contributions that played a role in ongoing research activities for three decades and contributed significantly to shaping the discipline as we know it today. These are (1) a conceptualization of the diffusion of innovation process, (2) development of a technique for operationalizing that conceptualization, and (3) the identification of empirical regularities in diffusion. Brief consideration is given to each of these contributions.

Hagerstrand's Conceptualization of the Diffusion Process

The basic tenet of Hagerstrand's conceptualization of the spread of innovation across the landscape is that the adoption of an innovation is primarily the outcome of a communications process by which the potential adopter learns of the innovation. This implies that factors related to the effective flow of information are most critical and, therefore, that a fundamental step in understanding the process is to identify the spatial characteristics of infor mation flows and resistances to adoption.

In Hagerstrand's conceptualization, (**Figure 2**), in formation originates as messages from the existing adopters or from an external source such as the mass media. He posits that the destination of the personal messages depends on the sender's network of inter personal contacts or his 'network of social communi cations', and the configuration of this network is primarily dependent on the presence of various social and terrestrial barriers which impede, divert, and/or channel communications. Particular attention is given to terrestrial barriers such as lakes, forests, difficult terrain, and the geographic distance separating two potential communicants.

The conceptualization further recognizes that resis tance levels differ from one individual to another, as a function of both personal and group characteristics, and that higher levels of resistance require more information for adoption to occur. Hagerstrand's discussion leads to a dichotomy of resistance either as stemming from values that are inconsistent with adoption of the innovation ('social' resistance) or stemming from practicalities which make adoption difficult or impossible ('economic' resistance).

Finally, to address the problem of diffusion at differ ent geographical scales, a hierarchy of networks of social communication is posited. As an example, one network may operate locally and another regionally. The first of these would control diffusion among farmers in a single locale, whereas diffusion at the central place level of aggregation might utilize a regional network comprised of persons in different central places that communicate with one another. The superimposition of these two levels would provide a comprehensive picture of dif fusion in the area represented (**Figure 3**).

In summary, Hagerstrand's conceptualization pro poses a transformation of the population from one with a low proportion of adopters to one with a high proportion of adopters by means of information dissemination through interpersonal contact or from an external source

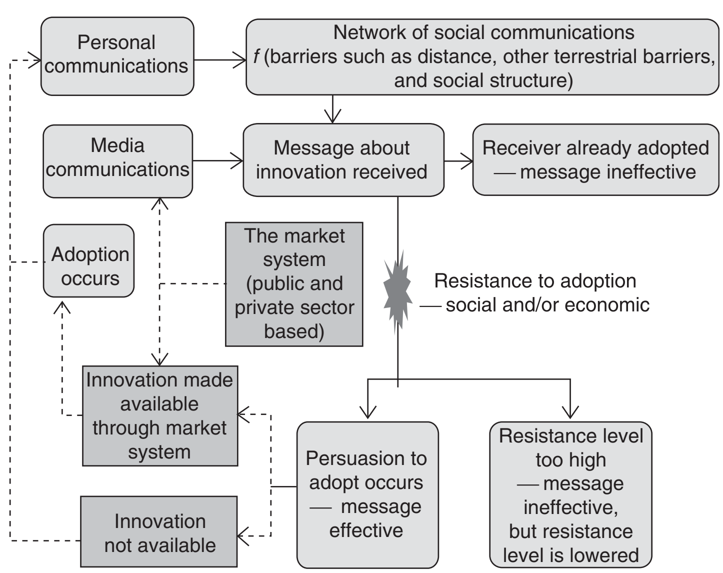

Figure 2 A flow diagram of the adoption perspective on innovation diffusion. From Hagerstrand with modifications reflecting more recent work on the market and infrastructure perspective. Dotted lines and shaded boxes indicate aspects of the matter and infrastructure perspective.

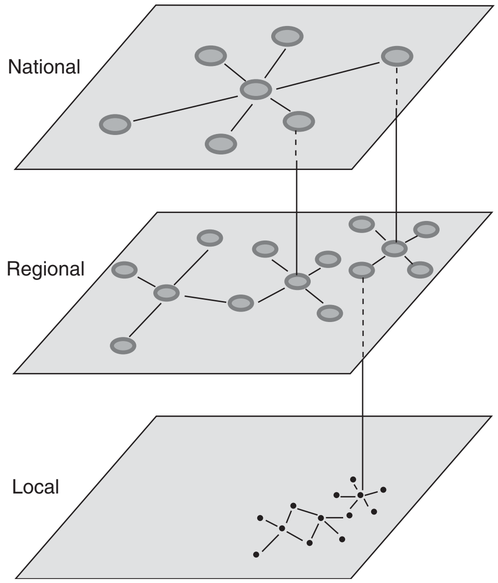

Figure 3 Schematic portrayal of diffusion at different spatial scales and their interconnection.

such as the media. In a spatial frame of reference, the principal mechanisms of this transformation are networks of social communication which are characterized by biases and distortions that are reflected in the spatial patterns of innovation diffusion.

Hagerstrand's Methodological Framework

To implement his conceptualization, Hagerstrand used Monte Carlo simulation. Several models of varying complexity were developed. The one he ultimately settled on, and which received the widest attention, focuses on interpersonal communications. Adoption takes place after a specified number of messages are received, the specified number varying according to an individual's resistance to adoption. Time is treated in discrete units of equal interval, and each past adopter sends a message in every time interval. The destination of each message is dependent on the probability of contact between teller and potential receiver. This probability is a function of the distance between them unless water or forest barriers intervene. For the case in which distance only is considered, a 'distance decay function' is employed to compute the probabilities of interaction between each of 25 cells in a 5×5 grid and the center cell of that grid, using the centers of each cell to compute the intervening distance and specifying probabilities such that all within the grid sum up to one. The resulting grid of probabilities is termed the 'mean information field' (MIF). It is employed with a map that is gridded in a similar manner, but with many more than 25 cells. The 5×5 MIF is centered over a map cell containing an adopter of the innovation at time t, and a random number is drawn to determine which of the 25 surrounding map cells receives a message to adopt in time $t+1$. If the message passes over a water or forest barrier, it is either nullified or a new random number drawn, depending upon the nature of the barrier. This procedure is repeated for every adopter and for every time interval. If a previous adopter receives a message, the message is lost; that is, it has no effect and is not repeated. If a message is received by a nonadopter who, owing to his/her resistance level, requires more than one message before adoption, the message is stored until sufficient messages for adoption are accumulated (**Figure 2**).

Empirical Regularities in Diffusion

Three patterns generally have been associated with diffusion processes (**Figures 4a–4c**). Over time, a graph of the cumulative level of adoption is expected to approximate an S shape. In an urban system, the diffusion is expected to proceed from larger to small centers, a regularity termed the 'hierarchy effect'. Within the hinterland of a single urban center, diffusion is expected to proceed in a wave like fashion outward from the urban center, first hitting nearby rather than farther away locations, and a similar pattern is expected for diffusion in a rural area. This third regularity is termed the 'neighborhood' or 'contagion effect'.

While Hagerstrand was not the first social scientist to identify these regularities, he is credited with introducing the concepts to the geography discipline. He also provides a framework which makes evident the interrelationships among these regularities. For diffusion in a central place or urban system context, this posits three phases: a 'primary' stage during which the initial diffusion centers are established; a 'diffusion' stage during which neighborhood type diffusion occurs in areas near diffusion centers and secondary diffusion centers are established in lower order urban places; and a 'condensing' or 'saturation' stage during which filling in occurs and diffusion eventually ceases (**Figure 4d**).

The framework thus illustrates the complementarity of the hierarchy and neighborhood effects and their temporal positioning within the diffusion process. It also implies that, if the locales of adopters were tabulated, higher order urban places would characterize the left tail of the S curve of adoption; lower order urban places would characterize the middle, sharply rising portion of the S curve, along with rural hinterlands in proximity to urban centers; and remote rural hinterlands would characterize the right tail of the S curve (**Figures 4e** and **3**).

Diffusion Research since Hagerstrand: Efforts Related to the Operational Model

Hagerstrand's operational model coincided with the 'quantitative revolution' in geography, but it also

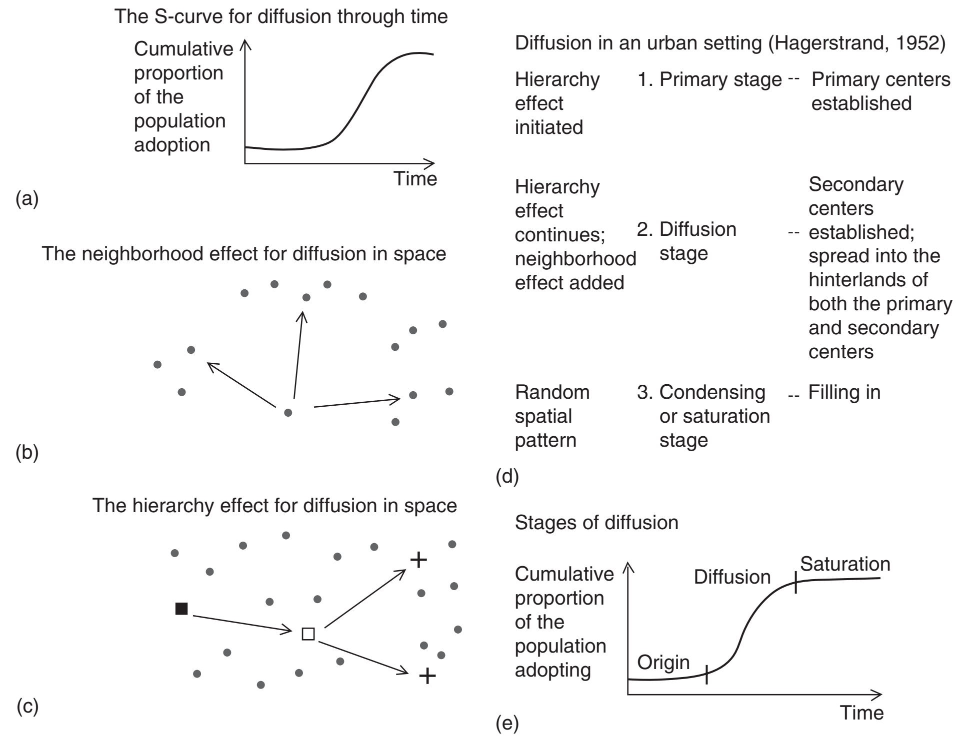

Figure 4 Empirical regularities in diffusion. Note in (**e**) the congruence of the schema with that of Hagerstrand in (**d**).

provided a unique contribution in going beyond conventional social science statistical procedures. Further, owing to its empirical regularities (**Figure 4**), diffusion is readily modeled by mathematical procedures. Accordingly, the topic became a centerpiece of geographers' efforts at the then new frontier of quantification, primarily focused on operational models of diffusion. Most attention was given to the Hagerstrand model itself, but new models for different situations also were proposed. Four types of research are evident (**Figure 1**): Studies of components of Hagerstrand's operational model, applications of Hagerstrand's operational model to innovation diffusion, applications of Hagerstrand's operational model to phenomena other than innovation diffusion, and developments in the mathematical portrayal of diffusion processes. Each of these is now considered.

Studies of Components of Hagerstrand's Operational Model

The MIF is the component of Hagerstrand's operational model that received most attention. Perhaps this is because the MIF embodies the spatial dynamics of the model, being the mechanism by which diffusion impulses are relocated from a point of origin to a point of destination. One set of concerns was mechanistic, focused on the procedures for creating an MIF from raw data, on the best mathematical function to represent the attenuation of spatial interaction with distance or the 'distance decay effect', and on estimating the goodness of fit between a simulation produced by the MIF and its real world counterpart. Configuring an MIF was done through data representing the distance between interactions, drawing on the generality that fewer interactions occur at greater distances and more interactions at closer distances. There is a greater probability, for example, that the origin and destination of migration, the stores at which one shops, or the pre marriage home locations of husband and wife (marriage distances) will be closer rather than farther apart; and expressing this empirical regularity might be done with a negative exponential function. Other surrogates explored include telephone calls, acquaintances, commuting, and shopping distances.

Computer programs for running the Hagerstrand model also received attention. These programs made the Hagerstrand model readily accessible and account in part for its widespread use. Also enhanced by programming developments were efforts to examine in an experimental setting the various diffusion patterns that result from different initial or ongoing conditions. A set of these efforts was concerned with geographic barriers, other than distance, which shape the MIF and distort its symmetry. Two properties of barriers deemed relevant were their shape (for example, parallel barriers, bar barriers with end gaps, bar barriers with a center gap) and their permeability (for example, absorbing, reflective, permeable). Directional distortions also were investigated, particularly for urban contexts where a bias would be introduced by the location of an urban place or the central business district.

Application of Hagerstrand's Operational Model to Innovation Diffusion

The extent of attention given to particular aspects of Hagerstrand's model is reflective of its importance and utility in the geography discipline of the 1960s. Not surprisingly, then, there are many examples of its appli cation, both to the diffusion of innovation and to other diffusion or diffusion like situations.

Examples of application directly to innovations in clude the diffusion of a community healthcare plan within Detroit, Harvestore storage systems in northeast Iowa, strip cropping in southwestern Wisconsin, irri gation wells in the Colorado High plains, and agricultural innovation in Mysore, India.

To illustrate the approach of this line of research consider the irrigation wells study. This employed both a local communications field, based on distances traveled to attend a free barbeque in Yuma, Colorado, and a re gional communications field, based on long distance telephone calls from Yuma. These were combined to form a single MIF, of circular rather than square shape, with possible contacts ranging from 1 to 40 miles. In addition, there was a ceiling on adoption, or a saturation level, of 16 wells per township. Expect for these minor variations, the model used in this study is virtually identical to that of Hagerstrand. Further, fit of the simulated diffusion pattern with the actual one was quite good, enabling the researcher to project the future adoption of irrigation wells and estimate the rate of de pletion of groundwater resources – a demonstration of the practical utility of the Hagerstrand model.

Applications of Hagerstrand's Operational Model to Phenomena Other than Innovation Diffusion

In applying the Hagerstrand operational model to phe nomena other than innovation diffusion, there was rela tively little modification, certainly not enough as to constitute a significant alteration. Phenomena examined in this way include migration in Australia, Ohio, and south central Sweden; spread of the Negro ghetto and growth at the urban fringe, both in Seattle Washington; spread and growth of urban centers in southern Sweden, a settled area; spread of urban centers in frontier areas of northern Sweden and Israel; and intra urban travel and transportation network growth in North America.

As a critique of these efforts, most applied the struc ture of Hagerstrand's operational model directly to their own problem, without significant modification to reflect the peculiarities of the process being studied. As a result, relatively little was learned about the particular structure of each situation or processes governing it. The models contained sufficient parameters to allow a reasonable fit between the simulated and real world diffusion patterns, and allow the conclusion that the suggested structure provides a plausible explanation of the actual process, albeit in retrospect, not a persuasive one. Indeed, the research findings would be of greater interest today were attempts made to determine if alternative structures are also plausible in the sense that they can be made to ap proximate the real world patterns.

In essence, then, applications of the Hagerstrand model, either to innovation or to other diffusion phe nomena, were accomplished without raising the question of whether the substantive assertions of the model – for example, that the process involved information flow or spatial interaction – were in conformance with the sub stantive reality of the processes being modeled.

Developments in the Mathematical Portrayal of Diffusion Processes

In the mathematical portrayal of diffusion processes some efforts were concerned with the operational model of Hagerstrand. Looking toward a more mathematically formal statement of the model, these generally involved the borrowing or restatement of models developed in the physical or biological sciences. The use of neural net models from mathematical biophysics is an example. The majority of efforts, however, focused on diffusion in volving urban places, putting forth models that differ in operational structure from that of Hagerstrand, while leaving unaltered the tenets of his conceptual model.

More generally, modeling efforts were guided by an approach that emphasized pattern over substantive pro cess and had a strong attachment with the precepts of 'general systems theory'. The latter perspective is illus trated by the observation that many different phenomena involving movement – for example, innovation, migra tion, travel, trade, industrial movements – can be exam ined within a diffusion framework. This would be based on the 'isomorphism' that all such situations consist of six basic elements an area or environment within which movement occurs; a temporal dimension; an item being diffused; places in which the item is located at the start of a time interval, or nodes of origin; places in which the item is newly located at the end of a time interval, or nodes of destination; and paths of movement, influence, or relationship between origin and destination places, referred to abstractly as edges.

The last four elements are represented as a dynamic graph set in a temporal geographic space (the first two elements). This is illustrated in **Figure 5** for 'relocation type diffusion', in which some members of the population change their locations from time t to $t+1$, and for 'ex pansion' or 'contagion type diffusion', in which new

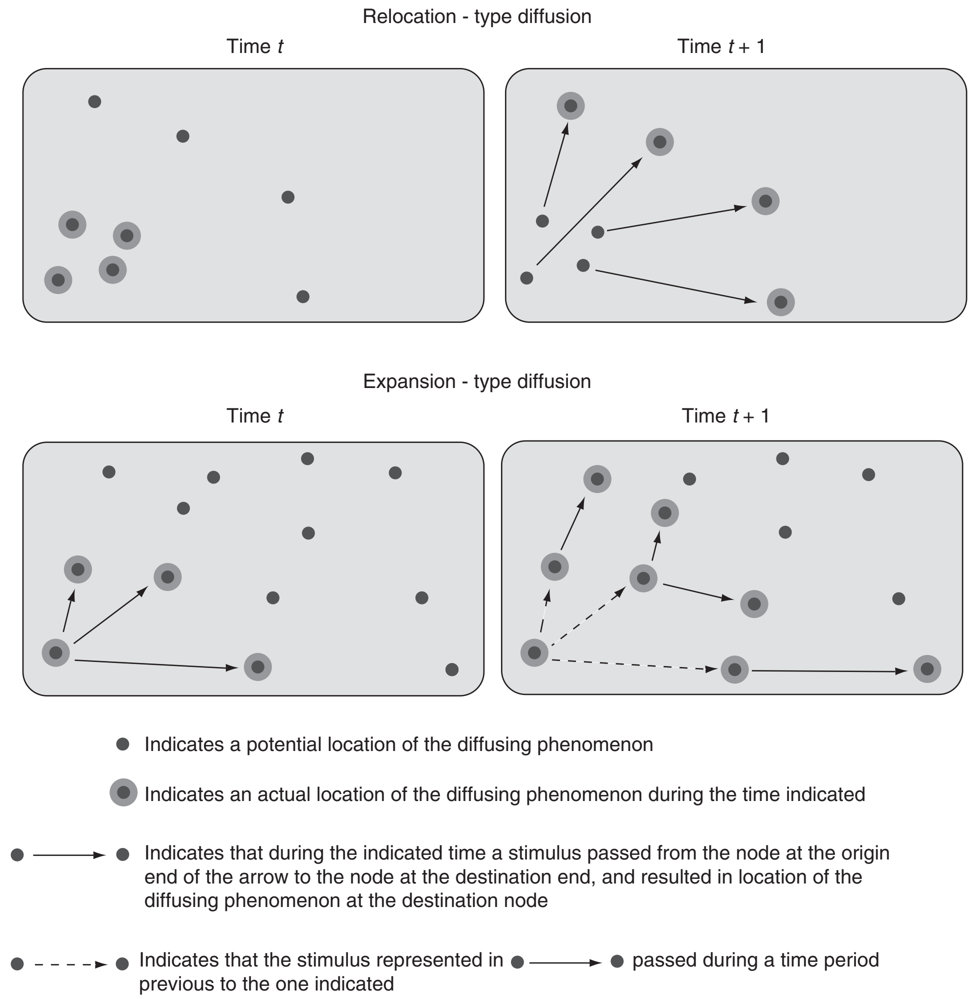

Figure 5 Diffusion portrayed as a dynamic graph.

members are added to the population between times t and $t+1$ and located so as to alter the population's lo cational pattern. The point here is to link disparate phenomena by a set of common structural characteristics that then give rise to general principles, thus providing a basis for a general theory of spatial diffusion or movement.

In a more concrete approach, different phenomena were seen to share some common and some distinct elements, leading to another set of models. The usual procedure was to take one or more of the empirical re gularities as given and then create a model which, as at least one outcome, would reproduce that regularity. Such regularities include an outcome that replicated the S shaped curve, produced the neighborhood effect in a wave model of diffusion outward from an urban center, and, for diffusion in an urban system, utilized the hier archy effect regularity or combination of hierarchy effect and neighborhood effect (**Figure 4**).

Taken together, these efforts moved the study of dif fusion toward a much greater integration with work on urban systems more generally. Thus, the newer models incorporated concepts of central place hierarchies, the rank size rule, spatial interaction, and the gravity model, as well as the friction of distance and barrier concepts associated with Hagerstrand's work. In the context of central place theory, for example, the 'order of good' concept indicates that higher order innovations would be limited to a particular portion of the urban system, unlike lower order innovations which could be found through out the urban system. Similarly, the variance in possible outcomes is considerably increased if interurban contact in not assumed to follow a gravity model formulation, a modification well supported by communication devel opments over the last quarter of the twentieth century. And models in this genre link space and time by, for example, fitting a logistic function to the S curve of diffusion, and then employing the latter to generate the number of adoptions for any given real time interval such as a single year.

In summary, Hagerstrand's model began as an offshoot of the cultural geography tradition that was dominant pre World War II, but by the 1980s it had morphed into a far more complex undertaking that both brought it into the broader realm of urban systems and regional devel opment and also fueled developments in that realm which led to its preeminence through the end of the twentieth century.

Diffusion Research since Hagerstrand: Third World Development as a Diffusion Process

With its strong regional dimension, geography also car ried the diffusion paradigm into African, Asian, and Latin American settings. This took several different forms.

Formal models of development can be divided into three phases that differ in their perspective: (1) stages of growth, modernization, and dual economy models which are aspatial, even though those dimensions are implicitly there; (2) core periphery and growth center frameworks which are explicitly spatial; and (3) dependency and Marxist frameworks which give more emphasis to forces that inhibit or block transformation. Central to these frameworks, albeit less so for (3), is the notion that Third World settings are comprised of traditional and modern sectors, and that development consists of, or necessitates, transforming traditional to modern – whether the focus is individuals, families, groups, economic sectors, infra structures, landscapes, regions, etc. Taken together, models of development divide Third World landscapes into dynamic, propulsive, modern cores, and traditional peripheries; development is represented by transforma tion and expansion of the modern; and this occurs through mechanisms such as information and capital flows, human capital improvement, migration, learned behaviors or imitation effects, infrastructure expansion and upgrading, and the like. Spatially, these transforma tions are seen in terms of two categories – backwash effects, whereby the core is strengthened at the expense of the periphery or hinterland, and trickle down effects, whereby core growth expands outward and stimulates periphery or hinterland development.

Geographers made relatively little contribution to the modernization, dual economy, and stages of growth frameworks but were highly involved in formulating core periphery, growth center, dependency, and Marxist approaches. Despite this difference, in that innovation diffusion was seen as integral to the social and economic change needed for transformation, geographers and many other social scientists became involved in this Third World project.

One approach involved choosing an index of devel opment and tracing its spread across the landscape – in some sense an unwitting restoration of culture trait studies from the Sauer tradition. Such indices included agricultural cooperatives, banking, communications in frastructure (telegraph, telephone), educational facilities, electricity, employment patterns across economic sectors, governance structures, medical services, police organi zations, postal agencies, transportation modes (roads, rail), urbanization, and other artifacts of a market economy.

In this enterprise, then, diffusion becomes a segue way into tracking development, its extent and paths of spread from one place to another, and the spatial integration of a region or nation. While these studies discuss the factors underlying diffusion, they generally do not interrogate them in a manner that moves forward our understanding of diffusion as a process or, perhaps, even development. One example is a study of the expansion of the trans portation system in Ghana as an index of development. While excellent at the time, and highly cited to this day, in hindsight we see that the transportation network really represents the spread of the colonial system and, de pending on one's perspective, may or may not represent real development.

In large part, the aforementioned blind spot was a product of the 1960s and 1970s era, one that was shared by most social scientists, not simply geographers, and one wherein a Eurocentric, often mechanistic, view of de velopment was the norm, as noted by James Blaut. In this regard, it is relevant that a later study uses the co operative movement in Sierra Leone as an index, but on behalf of a conceptual move forward that better linked development and diffusion frameworks, not simply to trace the spatial characteristics of development. Similarly, later studies of frontier development in the United States, Ecuador, and Brazil led to significant increases in our substantive understanding of the processes underlying this particular diffusion – ones that revolve around capital generation and control; migration, travel, and trade interactions; the role of particular economic sectors such as banking and wholesaling; and transportation linkages.

Diffusion Research since Hagerstrand: Efforts Related to the Conceptual Model

Prior to the 1980s, geography research on diffusion and related topics, such as urban dynamics and Third World development, tended to focus on models and modeling techniques. While the contributions of these efforts were considerable, and certainly an important step forward, the increase in our substantive knowledge was, at best, modest. Hence, in 1975 the accepted conceptual model of innovation diffusion was still that of Hagerstrand, proposed 25 years earlier. Since then, however, there has been a dramatic shift toward substantive and conceptual concerns (**Figure 1**), a shift that is documented in this section. The first topic is empirical research which pointed out a number of inconsistencies between observed reality and the Hagerstrand model. Attention then turns to work that addressed those inconsistencies and proposed conceptual alternatives and/or modifications to the diffusion paradigm. Finally, a brief look is given to technological diffusion which has become a major concern since the mid 1980s.

Inconsistencies between Observed Reality and the Hagerstrand Model

Efforts related to the operational model of Hagerstrand, described earlier, led many to expect coalescence in the work of geographers studying diffusion and those studying the dynamic aspects of other phenomena. This occurred to some degree, but an unanticipated outcome was to point out inconsistencies between reality and the model(s).

One strand of the coalescence is comprised of work that employs diffusion theory in another, usually a broader context such as growth pole and regional development processes, spatial decision making, and urban system development. To illustrate, the diffusion of technical innovations and of entrepreneurial or other organizational entities was seen as an important element underlying differential growth rates among urban areas. This led to accounting for, and modeling, evolution of the urban landscape by drawing on the conclusion that hierarchy and neighborhood effects characterize the diffusion of entrepreneurial and other organizational entities; that information transfer is a critical element underlying these diffusions; and that models incorporating population and distance components, such as the gravity model, suffice to account for the patterns of spread. But in weighing such a formulation against reality, it was concluded that the simplistic population and distance notions of traditional diffusion models were not adequate to account for the diffusion underpinnings of landscape evolution. Instead, looking toward the organizational decision making process, which often involves turning to the unique institutional setting in which such decisions are made, was deemed to be necessary.

A second strand of the coalescence in the work of diffusion researchers and that of geographers studying the dynamic aspects of other phenomena stems from the 'time geography' of Hagerstrand and research associates of the 1970s such as Tommy Carlstein. In that innovations affect the use of resources in space and over time, some innovations serve to increase capacity and thus release a particular resource(s), while others require additional inputs of human time, settlement space, water, energy, and so on – resources which are available only in certain time–space locations. Accordingly, an 'activity systems' perspective was advocated wherein adoption is viewed in the context of time budgets and the social, economic, and personal resources characterizing the milieu or locale of the potential adopter(s) – and the innovation's impact on, or congruence with, these resources is a critical element in diffusion. From the time geography perspective, then, that a potential adopter knows about the innovation is a necessary but not sufficient condition; and diffusion research of the past was overly, or inordinately, narrow in its focus solely on communication processes related to adoption and single, rather than a composite of, innovations.

Examples that illustrate the preceding point include the following. The introduction of formal schooling to an agrarian society is a 'time demanding' innovation that takes children away from their tasks in the village production process, so that acceptance should be inhibited. However, if electric light, which is 'time expanding' in that it extends the day, also were introduced, the school would compete less with the scarce production time resource, and acceptance would be more likely. Alternatively, grinding mills, which replace manual grinding of grain into flour, also are time expanding in that they allow completion of other tasks by freeing up daytime light. Also important is both the 'synchronization' and 'synchorization' of resource inputs related to an innovation, that is, resource inputs must be available both 'when' and 'where' they are needed. Finally, some innovations, such as the motorcar, create 'flexibility' in time and space, which increases their likely impact on time–space systems.

A third strand of coalescence in the work of diffusion researchers and that of geographers studying dynamic aspects of other phenomena stemmed from empirical research on diffusion itself. This finds a number of situations for which Hagerstrand's conceptual model is, at best, a partial explanation, and ultimately leads to considering factors other than information flows. One question was whether interpersonal communications or transportation is more important in accounting for innovation diffusion at the regional scale in India; later, a Mexican example concluded that infrastructure is more important. Others, drawing on empirical examples from Canada, Britain, North America, Southern Asia, and Sweden, noted that the communications process in diffusion is not spatially patterned in the way suggested by Hagerstrand's conceptualization; also that the shopping trip behavior of potential adopters and the locational behavior and marketing strategies of propagators of the innovation are more important determinants of diffusion patterns at the regional scale. We even find interrogation of spatial regularities in diffusion. It was noted, for example, that 'hierarchical filtering' or 'trickle down' could result from a market searching process in which an expanding industry exploits market opportunities in a larger to smaller sequence, or when an activity faced with rising wage rates in larger cities moves to smaller ones in search of cheaper labor; and that 'imitation' effects could result from entrepreneurs in smaller centers mimicking the actions of those in larger cities.

Together, these raised voices arrive at the conclusion that diffusion patterns are, in large part, better explained by entrepreneurial behavior rather than social inter action. This point is embodied in the market and infra structure perspective articulated by Lawrence Brown.

Market and Infrastructure Perspective: Extending and Changing the Conceptual Model

Hagerstrand's formulation, and those of the many other social scientists studying diffusion from the 1950s on ward, took an atomistic view of society as individuals with free will, any of whom could adopt an innovation. Thus, the onus of adoption was on individuals, who were classified on a scale ranging from innovator to laggard, with the social meaning, or significance, that each term implies. This is termed the adoption perspective, focus ing primarily on the demand side of innovation diffusion.

The market and infrastructure perspective focuses on supply or availability of the innovation. Rather than assuming all have an equal opportunity to adopt, this is seen as purposely and egregiously unequal, so that adoption represents a choice within a constraint set; government, nonprofit, and entrepreneurial institutions establish these constraints; and attention is given to the process by which innovations and conditions for adoption are made available to individuals, households, or firms. In general, then, focus is on the access to innovations or the way in which innovations are differentially made avail able to various population segments, and focus shifts from the individual to the structure within which individuals operate, paralleling realignments in social science overall. Nevertheless, the market and infrastructure perspective is put forth as complementary, not contradictory, to the adoption perspective. Broad outlines of diffusion are shaped by where (and to whom) innovations are (made) available, but within those outlines, more detailed vari ation in diffusion patterns reflects adoption behavior (**Figure 2**).

Central elements of this perspective include the propagator of the innovation, the entity behind its dif fusion; diffusion agencies (outlets) through which in novation is distributed to the population at large; and diffusion strategy, which is designed to encourage adoption and includes decisions on agency location, pricing, advertising, population segments targeted by promotional efforts, etc. Hence, the locations and tem poral sequencing of diffusion agencies determine where and when an innovation will be available and provide the general outline of the diffusion pattern. Diffusion strat egy further defines the pattern by creating different levels of access depending on a potential adopter's economic, locational, demographic, and social characteristics. The final pattern occurs via adoption behavior, the last stage of the process. Interacting with these factors are infra structures such as transportation and utility networks that may be an artifact of diffusion strategy, but more generally are independent of the innovation propagator and diffusion agencies. Further, in so far as such infra structures facilitate diffusion, they are generically seen in their own right as enabling innovations.

Hence, the market and infrastructure perspective emphasizes the marketing aspects of diffusion, especially strategies for market penetration which have a distinct spatial dimension. Whether the propagator is profit or nonprofit motivated, broad acceptance is a common goal, and a strategy will be designed to accomplish this, that is, to best penetrate the market. This also is true whether propagation involves centralized decision making, de centralized decision making with a coordinating prop agator, or decentralized decision making without a coordinating propagator, although such distinctions do affect procedures by which diffusion occurs. Similarly, diffusion agencies may be represented by salespeople or agricultural extension agents, but either should be guided by the marketing strategy and, accordingly, selective decisions should be made regarding which potential adopters will be contacted or cultivated, which not, and to what degree. In this context, then, that Iowa was the first state to have widespread adoption of hybrid corn in the 1930s is not necessarily because Iowans were more innovative; rather, they were targeted by a combination of agricultural extension agents(cies), seed companies, im plement dealers, and other entities comprising the agri cultural industrial complex of the time.

Considering the role of infrastructure, broadcast tele vision stations were needed before television sets dif fused. Today, cable, satellite, and broadband services vary enormously from place to place, affecting innovation availability, an occurrence acknowledged by terms such as 'digital divide'. Enhanced accessibility through com munication or transportation improvements increases the market areas of diffusion agencies, and renders innov ations more attainable by potential adopters – an extreme example is e shopping in the twenty first century. Con cerning enabling innovations, irrigation infrastructure is generally critical to the success of high yielding variety seeds in agriculture; expansion of the transportation system in Ghana, discussed earlier, can be seen in the same light.

Finally, the market and infrastructure framework calls into question empirical regularities such as the hierarchy and neighborhood effects because, from a marketing perspective, relevant factors are often not arranged in that manner. The diffusion of cable television in Ohio, for example, occurred initially in small, more remote places where broadcast signals were weak or nonexistent, because cable was a means of enhancing television reception and, thus, sales by appliance stores.

Technological Diffusion

Agricultural technology has been a major focus of diffusion studies. Beginning in the 1930s, and accelerating thereafter, rural sociologists and agricultural economists, working through agricultural experiment stations and extension agents, carried out an enormous number of studies focused on farm innovations such as hybrid corn, fertilizers, silage methods, high yielding varieties, and the like. Until recent industrial agriculture, the adopting unit was the farm household, thus joining this effort with diffusion research concerned with consumer innovations. However, manufacturing firms that are well beyond household size, also adopt technological innovations, as do service and retail firms. This has been studied extensively, but rarely joined with the body of diffusion research described previously.

Rather than emphasizing communications, studies of firm adoption gave more attention to characteristics of the firm and innovation itself. Also significant is the fact that innovation replaces an older technology with a newer one, highlighting comparative costs, benefits, and risks; and factors such as relative profitability, investment, sunk costs, and the like. There also is greater emphasis on innovation as a continuous process; one that might be tailormade for a particular firm or firm setting, rather than as a singular, unchanging phenomenon; and one that involves more intimate and ongoing relationships between the innovation supplier and user. Issues such as competition among firms and relative factor prices, favoring one technology over another, also play a role.

As these factors indicate, economic perspectives have been important to investigating the diffusion of technological innovations among firms. But spatial considerations enter in two ways. First, there was a tendency to assume that larger firms in larger urban agglomerations are earlier adopters, giving rise to a self reinforcing, 'cumulative and circular causation', hierarchy effect pattern of diffusion. Alternatively, other research pointed out that earlier adopters are just as likely to be in less than the largest size urban areas, which might correlate with aggressiveness and innovativeness of management. It also was noted that there might be a spatial order to innovative change, such that newer economic regions are favored, leading to regional rotation or spatial succession; Silicon Valley provides an example. The second way spatial considerations come into play is that the study of technological innovation has become closely intertwined with understanding regional development and regional economic change, and this shift emphasizes an entirely different set of factors such as regional production complexes, entrepreneurship, research and development infrastructures, government policy, agglomeration economies, and the like. Hence, the questions may not be inherently geographic in terms of process, but very geographic in outcome.

Diffusion Research in Geography as a Mirror of the Discipline

That the shifting focus in diffusion research coincides with broader trends in the discipline has been noted several times in this entry. For example, the Hagerstrand model was an important medium of the quantitative revolution, and the subsequent focus on the conceptual model of diffusion coincided with a disciplinary shift toward a concern with substantive reality, relevance, and societal utility. These shifts and their manifestation in day to day research are the concern of this section.

To elaborate the point, it is useful to employ Kuhn's view of paradigms and their importance to normal science. These foster a coherent tradition of scientific research by suggesting which experiments or intellectual exercises are worth performing, by directing research toward given goals, and by providing both a set of rules or standards for scientific practice and a consensus that eliminates the unproductive work of dealing with controversy about the legitimacy of a paradigm. But of course, the existence of a paradigm also implies barriers to the development of alternative paradigms which may be seen as illegitimate, irrelevant, or metaphysical; as the concern of another discipline; or as too problematic to be worth one's time. A paradigm might also insulate the community from socially important problems that are not reducible to the puzzle form because they cannot be stated in terms of the usual concepts and tools. How, then, does change occur in science? The paradigm incorporates a mechanism for its own replacement or alteration by providing expectations. Challenge occurs when these are not borne out by research, and as the anomalous findings, or anomalies, multiply, a new paradigm is created, often one in which the anomalous becomes the expected.

In this context, the quantitative revolution, embracing the tenets of logical positivisms, may be seen to represent a paradigm about research procedures which dictated matters pertaining to research tools, the questions asked, and substantiation required. On a lesser scale, the Hagerstrand model may be seen to represent a paradigm pertaining to a segment of substantive reality. The Sauer School also embraced a paradigm concerning the questions asked, methods used, and conceptualization being addressed.

To illustrate such shifts, **Table 1** compares research epitomizing the Sauer tradition, the transition from that

Table 1 A comparison of the major paradigms of diffusion research in geography

	Sauer tradition of diffusion	*Transition (Hagerstrand 1952)*	*Hagerstrand tradition of diffusion research (Hagerstrand 1953, 1967)*	*Market and infrastructure perspective on innovation diffusion and adoption (Brown 1981)*
Type of situation studied	Ancient; e.g., the grid-pattern town, domestication and spread of crops and animals, pottery styles, language, architecture	Contemporary; e.g., the automobile, radio, technical and social innovations	Contemporary; e.g., state-subsidized pastures, bovine tuberculosis vaccine, soil mapping, postal-checking services, automobile, telephone	Contemporary; e.g., ice cream retailing, credit cards, cable television, family planning, agricultural practices in the US, agricultural products in Mexico and Kenya, Third World development
Research methods employed	Qualitative analysis of archaeological data, archival documents, cartography, field work central to the endeavor	Cartography, graphs	Cartography, graphs, statistical analysis, mathematical modeling, Monte Carlo simulation	Qualitative analysis via case studies, cartography, statistical analysis, economic analysis, mathematical modeling
Type and generality of findings	Findings pertain to both diffusion pattern and diffusion process, but applied only to the phenomena being considered	Findings pertain to diffusion pattern only, but seen to apply in general to innovation diffusion in urban systems	Findings pertain to both diffusion pattern and process; seen to apply in general to innovation diffusion at all spatial scales	Findings pertain to both diffusion pattern and process; seen to apply in general to innovation diffusion at all spatial scales
Dominant process posited	None, but motivated by process of cultural change in society as either simultaneous invention or stages of evolution vs. diffusion, cross-fertilization, and intermixing	None	Communications	Market and infrastructure development, expansion, penetration

to the Hagerstrand tradition, the Hagerstrand tradition itself, and the market and infrastructure perspective. The critical change, illustrative of paradigm shift on the substantive level, is in the dominant process posited, discussed in earlier sections of this article. But changes representative of the methodological shift of the discipline also have occurred, and are manifest in the research methods employed and the type and generality of findings. There also are shifts in the type of situation studied.

Without Kuhn's view of normal science, and with the benefit of hindsight, the extensive work embracing the Hagerstrand model might be deemed myopic and ill directed. But the effort seems eminently reasonable when viewed as an elaboration of the dominant paradigm, which required the discovery of an anomaly, between that model and substantive reality, before the research direction could be altered.

One example of recognizing the anomaly is trying to apply Hagerstrand's conceptual model to diffusion of television sets at a regional scale. In that context, it was recognized that there needed to be a consideration of the shopping trip behavior of potential adopters and distri bution policies of propagators since it was necessary to know whether the market utilized by households was one chosen as a distribution point by the propagator – elements not a part of Hagerstrand's model. Sometime later, in examining spread of the artificial insemination of cattle, it was observed that the speed of information transfer, the central element of Hagerstrand's model, did not coincide with the spatial extent of the diffusion wave.

Nevertheless, it was at least 6 years after the obser vation of anomaly that the market and infrastructure perspective came to be accepted. In Kuhn's terms, this period would represent the extended exploration of anomaly step of the paradigm shift process, and over coming resistance to change. However, this period also corresponded with a broader shift in disciplinary focus from the quantitative revolution to a concern with rele vance, societal utility, and substantive reality.

Thus, the market and infrastructure model can be seen to play a role in the focus of the discipline in the late 1970s/early 1980s, just as the Hagerstrand model was a medium of the quantitative revolution. Arguably in fact, such congruence was a prerequisite to the wide spread acceptance of either model; that is, that the ex tensive elaboration of the market and infrastructure paradigm would not have occurred without the con comitant shift in disciplinary focus.

Reflecting a further shift in disciplinary focus, Derek Gregory provides a highly nuanced essay that begins with components of Hagerstrand's diffusion model, re phrases the usual interpretation, and uses it as a launch pad to advocate a broader and deeper understanding of social processes, including the role of inequality, class, and the social milieu within which adoption/diffusion occurs. This leads to extensive elaboration on Hager strand's time geography, described above, and provides a basis, in Gregory's view, for contemporary social theory. More recently, under another shift in disciplinary focus, Mei Po Kwan and Harvey Miller have taken the time geography concept into the realm of contemporary techniques by developing methods for GIS based geo computation and 3D geovisualizations based on Hager strand's time–space prism idea.

Another consideration is illustrated by the shift from the cultural geography paradigm of diffusion to that of Hagerstrand. As noted, this had less to do with the study of innovation adoption *per se* than with advent of the quantitative revolution, but interestingly, serendipity also played a role. In March through June 1959, Hagerstrand spent an academic term at the University of Washington with William Garrison and his students, William Bunge, Michael Dacey, Arthur Getis, Duane Marble, Richard Morrill, John Nystuen, and Waldo Tobler, among others. These were economic and urban geographers who would later be largely responsible for the quantitative revo lution. In this context, one might speculate what would have happened to diffusion research, and the revolution, had Hagerstrand instead visited Berkeley, the seat of cultural geography and largely substantively oriented diffusion research.

Finally, in this section and throughout this article there has been a tendency to refer to various models or perspectives as though they were discontinuous and mutually exclusive of one another, the earlier being re placed by the latter. While this is convenient for ex pository purposes, in reality the process of change is more one wherein there is considerable overlap between the new and the current.

See also: Anthropology and Human Geography; Berkeley School; Berry, B.; Bobek, H.; Chicago School; Core-Periphery Models; Dependency; Development I; Development II; Disease Diffusion; Distance; Economic Geography, Quantitative; Economics and Human Geography; Epidemiological Transition; Garrison, W.; Hagerstrand, T.; Industrialization; Innovation; Local Economic Development; Modernization Theory; Networks; Quantitative Revolution; Regional Development Theory; Regional Planning and Development Theories; Retail Geographies; Sauer, C.; Simulation; Spatial Expansion Method; Systems Theory; Technological Change; Technology and Regional Development; Telecommunications; Third World.

Further Reading

Blaut, J. M. (1993). *The Colonizer's Model of the World: Geographical Diffusionism and Eurocentric History*. New York: Guilford.

Brown, L. A. (1981). *Innovation Diffusion: A New Perspective*. New York: Methuen.

Brown, L. A. (1991). *Place, Migration, and Development in the Third World*, ch. 2. London: Routledge.

Brown, L. A. (1999). Change, continuity, and the pursuit of geographic understanding, Presidential Address. *Annals of the Association of American Geographers* 89, 1 25.

Carlstein, T., Parkes, D. and Thrift, N. (eds.) (1978). *Timing Space and Spacing Time, volume 1: Making Sense of Time, volume 2: Human Activity and Time Geography, volume 3: Time and Regional Dynamics*. London: Edward Arnold.

Gregory, D. (1985). Suspended animation of the stasis of diffusion theory. In Gregory, D. & Urry, J. (eds.) *Social Relations and Spatial Structures*, pp 296 336. London: Macmillan.

Hagerstrand, T. (1952). *Lund Studies in Geography; The Propagation of Innovation Waves*. Lund: Gleerup.

Hagerstrand, T. (1967). *Innovation Diffusion as a Spatial Process*. Chicago, IL: University of Chicago Press. (translation into English of *Innovationsforloppet ur Korologisk Synpunkt*. Lund: Gleerup).

Haggett, P. (1965). *Locational Analysis in Human Geography*. London: Edward Arnold.

Haggett, P., Cliff, A. D. and Frey, A. (1977). *Locational Analysis in Human Geography* (2nd edn.). London: Edward Arnold.

Kuhn, T. S. (1996). *The Structure of Scientific Revolutions* (3rd edn.). Chicago, IL: University of Chicago Press.

Kwan, M. P. (2002). Methods, models, and GIS: Feminist visualization: Re envisioning GIS as a method in feminist geographic research.

Annals of the Association of American Geographers 92, 645 661.

Malecki, E. J. (1997). *Technology and Economic Development: The Dynamics of Local, Regional, and National Competitiveness* (2nd edn.). Essex: Longman Scientific.

Miller, H. J. (2005). A measurement theory for time geography. *Geographical Analysis* 37, 17 45.

Rogers, E. M. (2003). *Diffusion of Innovations* (5th edn.). New York: Free Press.

Sauer, C. O. (1952). *Agricultural Origins and Dispersals*. New York: National Geographic Society.

Sauer, C. O. (1969). *Agricultural Origins and Dispersals; The Domestication of Animals and Foodstuffs*. Cambridge, MA: MIT Press.

Digital Data, Historical Geography and

H. Southall, University of Portsmouth, Portsmouth, UK

Glossary

Digital Imaging It is the creation of a digital surrogate via scanning or digital photography.

Dublin Core It is the best-known metadata standard in the cultural heritage field, but limited because it is built around a minimal subset of content, linking cataloging standards used in the libraries, museums, and archives sectors.

Extensible Markup Language (XML) It is a text format for defining formal representations of information, such as Dublin Core or TEI.

Gazetteer Service A system that receives place name enquiries over the Internet and sends back locations and other information.

Metadata It is data about data. Traditionally, it is a library catalog and other finding aids, today usually computerized and represented as XML.

Ontology It is a formal representation of a set of concepts and their relationships. It can be used to define some aspects of localities without coordinate data.

Optical Character Recognition It is the processing of a digital image of text to create editable and searchable text.

Text Encoding Initiative (TEI) It is an XML-based standard for capturing the meaning as well as the structure of text.

Thesaurus It is a hierarchic list of words identifying preferred and alternate terms, that is, synonyms. A gazetteer can be represented as a thesaurus.

Vectorizing It is the creation of a mathematical representation of lines on a map, from either a digital image or, using a digitization table, directly from a paper map.

Digitization

Before discussing these potentials, three distinct meanings of digitization of historical sources must be understood: (1) To most librarians and archivists, and probably to the generality of historians, digitization means digital imaging, that is, making pictures of documents. (2) In geographical information systems (GIS), especially as applied in human geography, digitization usually means recording the contents of maps as sequences of coordinates representing points, lines, and polygons. This work is more precisely described as vectorization. (3) Transcription re creates sequences of characters and numbers inside a computer, either by manual re typing or using optical character recognition technology.

Digital imaging is a much cheaper process than transcribing text or vectorizing maps, but the resulting files are much larger, and substantial collections only became practical in the 1990s as computer storage prices fell: maps should usually be scanned in full color at a resolution of 400 dots per inch (d.p.i.), or 300 d.p.i. as a minimum, resulting in files of 250 Mb upward per sheet. Digital photography is far less likely to damage historical documents than scanning, as it involves neither intense light nor physical contact between the document and the imaging equipment. However, it is much harder to avoid distortion, and even the largest digital cameras struggle to achieve the necessary resolution for large maps. In practice, there may have to be a trade off between preserving the information in a document, via digitization, and preserving the document itself.

Digitization projects led by librarians and archivists generally create collections of digital images organized like the original collection of documents or books, held in a content management system which is generally a nonhierarchic database holding the images and associated metadata. 'Preservation metadata' records the type of image, the size, the color depth, etc.; the GIF and JPEG formats are most widely used for dissemination, but compression leads to some loss of information and the TIFF format is generally preferred for archiving. 'Descriptive metadata' enables users to discover and identify digital resources. The best known descriptive standard, the Dublin Core Metadata Element Set, was designed to allow integration of catalogs from libraries, museums, and archives. Its 15 elements include 'coverage', defined as "the spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant." However, most Dublin Core implementations are limited to the standard elements found in library catalogs: author, title, date, and place of publication, etc.

Most digital collections of historical maps therefore consist of images of the separate sheets, the only clearly identified item of geographical information being the place of publication, with the actual area covered possibly being identified somehow within the title. This is still far more detailed than the cataloging of books simply because there is a catalog entry for each 'page', that is, map sheet. Edinburgh University's Geo X Walk project showed that it was possible to find maps of, for example, 'Kent', by accessing standard online library catalogs using the Z39.50 protocol but employing a central knowledgebase which 'knew' that, say, 'Dover' and 'Maidstone' were the names of settlements within 'Kent'.

Gazetteers

Few historical documents contain lists of geographical coordinates; even historical gazetteers more usually give relative locations: "Colwall, par. and vil. with ry. sta., SE. Herefordshire, $3\frac{1}{2}$ miles NE. of Ledbury"; "Ledbury, market town and par. …, Herefordshire, … 120 miles NW. of London" (examples from Bartholomew's *Gazetteer of the British Isles*, 1887). The first priority in making more effective geographical use of digitized historical documents is not adding coordinate data but exploiting the abundant geographical names they and associated metadata contain. The Dublin Core standard comments on the 'coverage' element: "Spatial topic may be a named place or a location specified by its geographic coordinates. … A jurisdiction may be a named administrative entity or a geographic place to which the resource applies. Recommended best practice is to use a controlled vocabulary such as the *Thesaurus of Geographic Names*. Where appropriate, named places … can be used in preference to numeric identifiers such as sets of coordinates."

The most basic controlled place name vocabulary would be simply and literally a list of names, defining which strings of characters are allowed in the coverage element of a metadata record, or allowing text analysis programs to identify geographical references within text. In practice, all such lists are also gazetteers and provide some form of locational information. This may be absolute locations such as latitude and longitude; relative locations like those above; or hierarchical locations such as 'in Herefordshire'. This is because geographical names are constantly re used across the Earth's surface, so some additional information is needed to disambiguate particular places: Worcester, England, versus Worcester, Massachusetts. In this sense, some of the earliest documents of interest to geographers can serve as gazetteers: medieval taxation lists generally order parishes within counties, while a major source for the China Historical GIS are the administrative gazetteers from the Han dynasty (206 BCE–CE 220) onward.

Recent years have seen extensive research into architectures for digital gazetteers, although involving few geographers. The *Thesaurus of Geographic Names* (TGN), developed by the Getty Information Institute and referenced by the Dublin Core standard, is organized primarily as a thesaurus: Stoke 'on' Trent should be used instead of Stoke 'upon' Trent; 'Stoke on Trent' is within the 'United Kingdom' which is within 'Europe'. The TGN includes a single point coordinate using latitude and longitude for each entry – even the United Kingdom – but does no mapping.

The Encoded Archival Context (EAC) initiative, sponsored by the Research Libraries Group, has created a standard which is even more purely textual, making no provision for coordinates at all. It is mentioned here because it is the emerging standard for 'archival name authorities', including geographical names. As previously discussed, most cataloging systems in libraries and museums do not systematically record locations, but some archives do index by place. More specifically, British local record offices generally index documents by the 'ancient parish' of origin. This requires a source which lists what parishes existed, each with an agreed spelling of its name. 'Ancient parishes' are not loosely defined 'places' but legally defined administrative units, and standards such as EAC treat such units as corporate bodies, like companies; but geographers would note that unlike companies, their boundaries are a fundamental part of their legal definition and without boundary information, documents such as the deeds of a house cannot easily be assigned to the appropriate parish.

Two other standards for digital gazetteers are more geographical in their approach. The Open Geospatial Consortium's 'gazetteer service profile' is an extension to their 'web feature service' standard, enabling a collection of geographical features to be searched by name, and hold information on variant names. The Alexandria Digital Library's (ADL) family of gazetteer standards defines the content of digital gazetteers, the types of feature they can contain, and a protocol by which networks of gazetteer servers can be queried over the Internet. Each entry in a minimal ADL compliant gazetteer has four elements: a unique identifier; a name, which is not necessarily unique; a 'feature type', such as 'fortifications' or 'mountains'; and a locational 'footprint', which can be a point coordinate, a bounding box, or a full boundary polygon. However, optional extensions to the ADL gazetteer content standard allow an entry to include detailed information on changing names, identifying names in different languages, and providing citations; information on complex hierarchies of relationships between geographical entities; multiple polygons tracking changes over time, with dates of varying precision.

Even a minimal ADL entry may demand too much of historical sources. A source may simply tell us that some geographical feature was 'in Herefordshire', although one approach is simply to associate a bounding box for Herefordshire with the feature. Other 'geographical features' referenced in historical sources may not necessarily exist; for example, and in a British context, should we include 'Camelot' and 'Avalon' in a gazetteer? The ADL standard also requires that every geographical feature be assigned a feature type, but this is to deny the concept of a 'place'. The particular feature type thesaurus developed by the ADL project is heavily influenced by the set of features identified on United States topographic maps, and terms like 'correctional facilities' and 'badlands' are not widely used elsewhere.

The audience for place name based online resources is at least as great as that for the statistical mapping which has preoccupied most historical geographers focused on digital technology. Even researchers studying nineteenth century sources will encounter many place names, notably for parts of towns, which have since gone out of use; and researchers studying earlier periods have also to deal with large numbers of names which differ substantially from their modern form; for example, Celia Fiennes writing in the early eighteenth century refers to both Norwich in Norfolk and Northwich in Cheshire as 'Norwitch'. Further, archivists and other catalogers need reliable name authorities if records are to be searchable by locality, and very large numbers of genealogists have specific requirements of gazetteers.

Geo-Referencing

Some would describe the products of digitizing historical documents as 'digital surrogates', but the digital version of the document should include all the information in the original while opening up new potentials. For the geog rapher, the most important of these is geo referencing: directly associating parts of the document with the real world locations they refer to. The most obvious appli cation is to maps, but the term has rather different meanings with statistical information and text.

The starting point is a digitized map, most obviously an image created by scanning but it could be the result of tracing a paper map using vector graphics software such as Adobe Illustrator, which understands only the 'local coordinates' of a piece of paper. Conversely, geo referencing is only possible within software which under stands real world coordinates like latitude and longitude. The precise method depends on the software, but some literally involve overlaying the map to be geo referenced, rendered semitransparent, over a map that is already geo referenced, aligning matching features in the two maps. If the map that is already geo referenced is a modern map, its publisher may well be able to claim a copyright in the geo referenced historical map.

Even if the historical map is published by the same national mapping agency as the modern map, it should not be assumed that they use the same map projection. For example, Britain's Ordnance Survey originally used the Cassini projection, and established a different prime meridian for each county. Only in 1931 did they begin to use the Transverse Mercator projection used in modern British maps. However, experiments with geo referencing First Series Ordnance Survey one inch maps from the mid nineteenth century showed that basic inaccuracies in the surveying were a larger problem than technical differences between map projections. Historical maps from before the nineteenth century are generally so topographically inaccurate that geo referencing them has little meaning, and researchers need to manually transfer specific features from them onto a more recent map, and geo reference that. This approach was used, for example, to construct an atlas of London in the 1690s, combining contemporary street maps with the earliest large scale Ordnance Survey maps from the nineteenth century.

The majority of historical documents are not maps, and 'geo referencing' them means identifying and giving meaning to the geographical names they contain. Stat istical reports are considered separately because they have been the focus of intensive study, especially by historical demographers working with censuses and vital registration data, and because such statistics almost always relate not to a point or vague 'place', but to a clearly delineated area and usually to an administrative area, such as a parish or a county. Census reports consist mainly of tables with se quences of geographical names down the left hand side; but while these names may have some immediate mean ing, as the names of modern towns or localities, these may well mislead. For example, English administrative geog raphy before 1974 featured units called 'rural districts' which were usually named after market towns but did not contain those towns, only the surrounding rural area. An extreme example is Portage County, Wisconsin, which was originally created in 1836 around Portage City in the south of the state, in the mid 1840s stretched all the way north to the Michigan border, and is now in the middle of the state based on Stevens Point.

In practice, geo referencing historical statistics means constructing a historical GIS, as discussed below, or at very least linking the data to a systematic record of what units existed, that is, a kind of authority list. What needs to be noted here is that establishing the boundaries of the reporting units used in a particular set of statistics may be a major project, even without any attempt to construct a GIS. For example, the first British census to clearly document its reporting geography was that for 1971. The records of early American county boundaries take the form of textual descriptions of boundary lines held in county courthouses, while our knowledge of the ancient parish boundaries of England has had to be pieced together mainly from tithe and enclosure maps of indi vidual parishes.

Geo referencing text means identifying geographical names within digitized texts, and if possible marking them up so as to associate them with other information about the same places. The first problem has been addressed as a branch of computational linguistics, the main conclusion being that holding a large body of information on what geographical names exist is more important than de veloping especially sophisticated algorithms. Even then, the complexities of human use of language may defeat the smartest computer systems. Consider this sentence from Boswell's account of his tour of the Hebrides with Dr. Johnson: "Simpson said, he would willingly go for Col, if young Col or his servant would undertake to pilot us to a harbour"; the first reference to 'Col' refers to an island, and the second to a person, the laird of that island.

The best developed system for marking up geographical names within texts has been developed by the Text Encoding Initiative, although most projects which have implemented the TEI guidelines have not implemented the section relating to geographical names; the most striking exception is the Perseus Project. The guidelines specify a place name tag, the simplest form of which is:

<placeName > Norwitch</placeName>

However, additional attributes can be added to the tag to standardize spelling, and to disambiguate:

<placeName reg = Tidenham cnty = Gloucestershire>
Tudenam<placeName >

Research in this area is poorly developed, and in particular the current TEI guidelines make no attempt to capture the 'role' of geographical names within text. This example actually appears in the guidelines, but the fact that 'Babylon' appears as a metaphor is ignored. In this case, it refers to New York but other writers from the same period use exactly the same soubriquet for London:

After spending some time in our <placeName key = NY1> modern.
<placeName key = BA1> Babylon </placeName>
<placeName key = NY1> New York </placeName>, I have proceeded to the
<placeName key = PH1> City of Brotherly Love </placeName>

Historical GIS: Principles

This article has explained how historical documents including maps can be digitized, and how such digitized documents can be enhanced to link them to the real world and to other digital information. GIS technology enables a wide range of historical information to be brought together into a single system, and creates a potential for broad reconstructions of historical environments. However, GIS systems are not the only way of combining historical sources, and existing commercial GIS solutions impose undesirable limitations on how we work with those sources.

It has already been noted that most library led digitization projects hold their images in content management systems which are essentially flat file databases, but designed to hold extensive descriptive and preservation metadata. Here there is no attempt to reconstruct past landscapes, and digital map libraries held in this form tend to emphasize the history of cartography, not the territory being mapped. However, much geographical research is concerned as much with people as with places, and prosopographical databases, usually constructed using relational database software, may be more useful to many historical geographers than GIS.

Mainstream commercial GIS software, and especially ArcGIS and MapInfo, appear well suited to mapping historical census data and similar statistics, and the problems finding suitable base maps to be digitized have already been discussed. The most obvious problem is that GIS packages provide much weaker support for recording dates than do conventional database programs. One approach is to construct separate coverages for each date, which works well for ten yearly censuses but less well with vital registration or data in economic distress: the data may well be available quarterly or monthly, so if boundaries change constantly a very large set of very slightly different coverages need to be created and managed. For this reason, the original Great Britain Historical GIS, discussed below, instead constructed a database of boundary line segments, or 'arcs', and used a large body of custom code to manipulate dates and assemble boundary polygons valid for specific days.

An alternative view is that the largest problem facing historical GIS specifically is not including a time dimension, but the large uncertainties surrounding historical information, and the consequent need to document sources in great detail. In particular, the whole approach of mainstream GIS assumes that locations are known with at least as much certainty as all other information, so that a set of geographical objects – points, lines, and polygons – become the central framework to which all other knowledge is added as attributes. Conversely, with historical sources we often know geographical names with far greater certainty than we 'know' the associated locations: locations may be conceptually problematic, as with Camelot and Avalon; real but unknowable, as with now defunct villages appearing in a medieval tax listing; known only approximately because the only maps are inaccurate; or simply known only approximately because the very expensive task of researching and computerizing accurate historical boundaries is yet to be funded.

The cost of historical GIS projects is rarely discussed other than in the context of the failure of particular funding bodies to support particular projects. However, projects such as those discussed in the final section generally cost more than academic projects working with modern GIS data, because the latter generally work with digital data created elsewhere, by national mapping or space agencies. Further funding can possibly be obtained if projects aim to serve wider nonacademic audiences, but they impose new requirements. The needs of archivists and the generality of local and family historians certainly require much greater attention to geographical names, and it is notable that the websites of some major histor ical GIS projects simply cannot be searched by place name.

The cost becomes still greater if a wider range of geographical information is included. One obvious ex tension beyond administrative boundaries is the inclusion of transport routes, but if the system is designed to trace changing geography over time it needs to record the detailed process of route construction and, sometimes, closure. This is possible for railway systems, although very extensive archival research may be required, and is arguably impossible for roads. The need for such extensions should be very carefully assessed, as building complex historical GIS systems often becomes a goal in itself, like assembling an evermore complex model railway. This is visibly a problem with urban historical GIS projects, which generally take one town and trace its development over time: ideally, such a system needs to record each building, in three dimensions, as the built form evolves, together with roads, railways, and so on.

An obvious problem with a GIS which records only one type of feature, be it administrative boundaries or railway lines, is that it is hard for users to orient them selves: the map lacks context. Adding additional layers of features is almost always possible but very expensive. However, an alternative approach is much cheaper and at least as effective. Methods for geo referencing scanned images of historical maps have already been discussed. The Great Britain Historical GIS has created multiple sets of such one inch to one mile maps of Great Britain, for different dates, plus broadly comparable less detailed maps for the same dates and for the rest of Europe. The resulting collection of scanned maps is effectively a his torical GIS in its own right, allowing users to zoom in and out to see different amounts of detail, and to change date. It is then used to provide underlays beneath stat istical mapping, which are sufficiently transparent for town names, railway lines, and so on from the scanned maps to show through.

The GIS field has generally developed around pro prietary software, notably ArcGIS and MapInfo, and the file formats they define. This tends to make GIS a closed environment, not interacting with other digital content. However, the Open Geospatial Consortium has de veloped a suite of standards allowing web based col laboration. The scanned map server developed by the GBH GIS, described above, supports the OGC Web Map Server standard, enabling other websites to use the server as a source of base maps. More advanced OGC standards, especially those of web feature servers, are yet to be implemented by historical projects, but would provide polygon base maps and support place name enquiries.

Historical GIS: National Systems

Even excluding archaeological GIS, there are many dif ferent kinds of GIS that have been constructed to support historical research, or to aid in the presentation of past environments. However, technical developments have been led by systems tracing the development of whole countries, generally focused on presenting historical census data.

The first such projects were in the United States. The *Atlas of Historical County Boundaries* began in the late 1970s as a collaboration between the University of Wisconsin's Automated Cartography Laboratory and Chicago's Newberry Library, but technical difficulties led to the project at the Newberry Library reverting to traditional methods. The Great American History Machine, despite its name very much a GIS, was developed independently by David Miller and colleagues at Carnegie Mellon University. Again pre dating commercial GIS software, it was built on top of the Andrew Toolkit, a pioneering windowing system. This severely limited the hardware it would run on, and a Microsoft Windows version was developed but never released. Much more recently, the US National Historical GIS project at the University of Minnesota has constructed sets of GIS coverages at census tract level and assembled and documented all existing computerized transcriptions of historical aggre gate census data.

European projects have developed largely independ ently. In the late 1980s and early 1990s, national GIS systems were constructed for Belgium, the Netherlands, Norway, Portugal, and Sweden. They were generally based on specially programmed systems, including the time dimension by various methods. More recently, they have been converted to work with commercial GIS software. Conversely, the Great Britain Historical GIS began somewhat later and worked with Arc GIS, but has moved to a different architecture based on ontologies implemented within an object relational database. De velopments outside North America and Europe have been generally more limited because of cost and more limited sources. However, the China Historical GIS covers a much longer time span than any European system.

See also: Digital Earth; GIS and Cartography; GIS and Society; GIScience and Systems; Spatial Ontologies.

Further Reading

Berman, M. L. (2005). Boundaries or networks in historical GIS: Concepts of measuring space and administrative geography in Chinese history. *Historical Geography* 33, 118 133.

Besser, H. (2003). *Introduction to Imaging* (revised edn.). Los Angeles, CA: Getty Research Institute.

Fitch, C. A. and Ruggles, S. (2003). Building the national historical geographic information system. *Historical Methods* 36, 41 51.

Goerke, M. (ed.) (1994). *Coordinates for Historical Maps*. Gottingen: Max Planck Institut fur Geschichte.
Gregory, I. N. (2003). *A Place in History: A Guide to Using GIS in Historical Research*. Oxford: Oxbow Books.
Gregory, I. N. and Southall, H. R. (1998). Putting the past in its place: The Great Britain Historical GIS. In Carver, S. (ed.) *Innovations in GIS 5*, pp 210 221. London: Taylor and Francis.
Hill, L. L. and Janee, G. (2004). The Alexandria Digital Library Project: Metadata development and use. In Hillmann, D. I. & Westbrooks, E. L. (eds.) *Metadata in Practice*, pp 117 138. Chicago: American Library Association.
Knowles, A. K. (ed.) (2002). *Past Time, Past Place: GIS for History*. Redlands: ESRI Press.
Knowles, A. K. (ed.) (2008). *Planing Histroy: How Maps, Spatial Data, and GIS and Changing Historical Scholarship*. Redlands: ESRI Press.
Miller, D. and Modell, J. (1988). Teaching United States history with the Great American History Machine. *Historical Methods* 21, 121 134.
Smith, D. M., Crane, G. and Rydberg Cox, J. (2000). The Perseus Project: A digital library for the humanities. *Literary and Linguistic Computing* 15, 15 25.
Southall, H. R. (2008). Visualization, data sharing and metadata. In Dodge, M., McDerby, M. & Turner, M. (eds.) *Geographical Visualization: Concepts, Tools and Applications*, pp 259 275. Chichester: Wiley.

Relevant Websites

http://www.visionofbritain.org.uk
A Vision of Britain through Time (GB Historical GIS).
http://www.alexandria.ucsb.edu/gazetteer
Alexandria Digital Library Gazetteer Development.
http://www.newberry.org
Atlas of Historical County Boundaries (US).
http://www.fas.harvard.edu
China Historical GIS.
http://www.davidrumsey.com
David Rumsey Map Collection.
http://dublincore.org
Dublin Core Metadata Initiative.
http://www.ecai.org
Electronic Cultural Atlas Initiative.
http://www.hgis germany.de
HGIS Germany.
http://www.hgis.org.uk
Historical GIS Research Network.
http://www.opengeospatial.org
Open Geospatial Consortium.
http://www.perseus.tufts.edu
Perseus Project.
http://www.tei c.org
Text Encoding Initiative.
www.getty.edu/research/conducting research/vocabularies/tgn
Thesaurus of Geographical Names.
http://www.nhgis.org
US National Historical GIS.

Digital Divide

T. Unwin, University of London, Egham, Surrey, UK
G. de Bastion, Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ) GmbH, Eschborn, Germany

Glossary

Broadband Broadband is a high data-transmission rate Internet connection. DSL and cable modem or wireless local area network (WLAN) services are typically capable of transmitting 256 Kbps or more, starting at approximately four times the speed of a modem using a standard digital telephone line. Generally, it is defined as a transmission capacity with sufficient bandwidth to permit combined provision of voice, data, and video.

DSL Digital subscriber line. DSL is a technology, which brings high-bandwidth information to homes and small businesses over ordinary fixed telephone lines. Typically, the download speed of DSL ranges from 128 to 24 000 Kbps depending on DSL technology and service level implemented.

ITU International Telecommunication Union, headquartered in Geneva, Switzerland, is an international organization within the United Nations system where governments and the private sector co-ordinate global telecom networks and services.

Internet Backbones The Internet backbone refers to the main connections of the Internet. It is made up of a large collection of interconnected commercial, government, academic, and other high-capacity data routes that carry data across countries, continents, and oceans.

Wireless Generic term for mobile communication services which do not use fixed line networks for direct access to the subscriber.

WSIS The World Summit on the Information Society, the first phase of which took place in Geneva in December 2003, with the second phase in Tunis in November 2005.

Introduction

The notion of a digital divide emerged during the mid 1990s, initially to refer to the differences between communities in their access to computers and the Internet. Since then, it has been used more widely to refer primarily to differences in the global distribution of digital technologies and content, not only computers and the Internet, but also mobile telephony, radio, television, and software. Moreover, the term refers both to absolute presence or absence, and also to the quality of access that people have to hardware and software.

The digital divide can usefully be examined at a variety of scales, from the global to the regional, national, and indeed local. Furthermore, attention to the transformatory concept of 'bridging the digital divide' forces us to confront moral agendas about the rights and wrongs of digital differences, and the potential benefits that poor and marginalized communities can gain from digital empowerment. The increasingly hegemonic notion of 'development' that has come to dominate international agendas during the first decade of the twenty first century has drawn its legitimacy from the twin rhetorics of the free market and liberal democracy. Information and communication are central to these agendas. Thus, there has been widespread interest in the role of information and communication technologies (ICTs) in 'development', and whether or not digital divides can indeed be 'bridged'. Central to these concerns has been the commitment made by the United Nations in 2000 in the form of the Millennium Development Goals, the eighth of which includes the aspiration: "In cooperation with the private sector, make available the benefits of new technologies – especially information and communications technologies."

Spatial Representations of Digital Inequality

Interpreting Digital Inequality

Digital inequality refers not only to spatial differences in the distribution of particular digital phenomena, but also to issues of social differences, access, political control, and cultural differences. Moreover, it is important to consider precisely what is meant by the word 'digital' itself. Often, it is used primarily to refer to hardware and infrastructure, but a more comprehensive definition includes the information and communication flows that are enabled by such technologies. Thus, the notion of digital inequality is concerned just as much with the ways in which people communicate by mobile/cell phone, as it is with digital content accessible via television, and access to information about local or national government practices through the Internet. The social and political dimensions of the 'digital divide' are particularly apparent through the varied access that people from different backgrounds and with different skills have to such digital content.

Global Digital Divides

There are marked inequalities between the ICT infra structures and services available across the world. Inter national Telecommunication Union (ITU) statistics indicate that in 2004 Africa averaged 1.74 PCs per 100 inhabitants, whereas Oceania averaged some 50.84; 14% of the world's population live in the G8 countries, but these account for 34% of the world's total mobile phone users; and on average there are 11.50 telephone sub scribers per 100 inhabitants in Africa, compared with 32.98 in Asia, 76.66 in the Americas and 111.72 in Europe (see **Figure 1**). In addition to these spatial differences, global inequalities are also reflected socially. Female Internet users thus ranged from 51% of the total in the USA and Canada in 2001–02, compared with 43% in Argentina, Norway, and the UK, but only 37% in Germany and Italy, and 35% in Malaysia. Figures for Africa are even lower, with only 19% of Internet users in South Africa and 12% in Senegal estimated to be women. There is some evidence that at a global scale the 'digital divide' shrank between the mid 1990s and the mid 2000s. ITU statistics again thus indicate that in 1994 the 'de veloped world' had 73 times as many Internet users, 11 times as many telephone lines per 100 inhabitants, and 27 times
as many mobile telephone subscribers per 100 inhabitants as had 'developing countries', but by 2004 these differ ences had fallen to only 8 times as many Internet users and 4 times as many telephone lines and mobile sub scribers per 100 population.

Regional Differences

The above statistics emphasize that Africa is by far the least well served part of the world in terms of ICTs. But even within Africa there are huge differences, both between countries and also between rural and urban parts of the continent. All ICT indicators across Africa indicate that a few places, notably the island territories of Mauritius, Réunion, and the Seychelles have vastly higher connectivity than the remainder. In 2003, Réunion had 198 times the number of telephone lines per 100 inhabitants as did countries such as Uganda or Niger. Spatially these differences are clearly revealed in **Figure 2**, which indicates that countries such as Burundi, the Democratic Republic of Congo, Ethiopia, Liberia, and Niger, all only had about one Internet user per 1000 inhabitants in 2003. Such regional differences are equally encountered across Asia and Latin America.

Local Divides

Digital divides are not only evident between rich and poor countries, but also within the developing world. Even in the world's richest countries there remain deep social and spatial divides in terms of digital access to information and communication. In 2005, the UN e Government Global Readiness Report noted that the UK was the highest ranking country in the world in terms of e participation. However, as **Figure 3** highlights, broadband access to the Internet varied enormously across the country in 2005. Dense middle class suburbs, where both cable and DSL are available had the highest proportion of households with broadband. More than 40% of households, though, did not even have access to the Internet. This is widely seen as accentuating socio economic differences, particularly for its effects on educational usage by young people. Such inequalities can be even starker in poorer countries, where access to ICTs, and the Internet in particular, are highly differentiated. In the third quarter of 2005 alone, China added some 3.6 million new broadband lines (far more than the

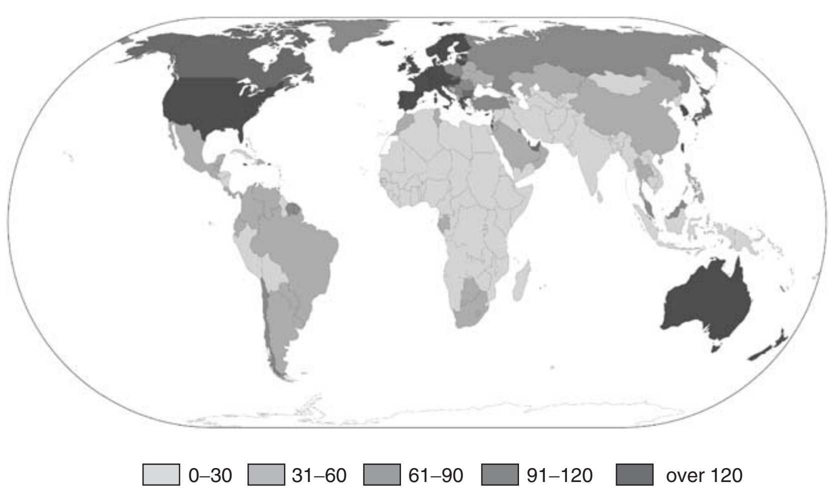

Figure 1 Telephone subscribers per 100 inhabitants in 2004. From ITU (2005) *World Telecommunication/ICT Development Report 2006: Measuring ICT for Social and Economic Development.*

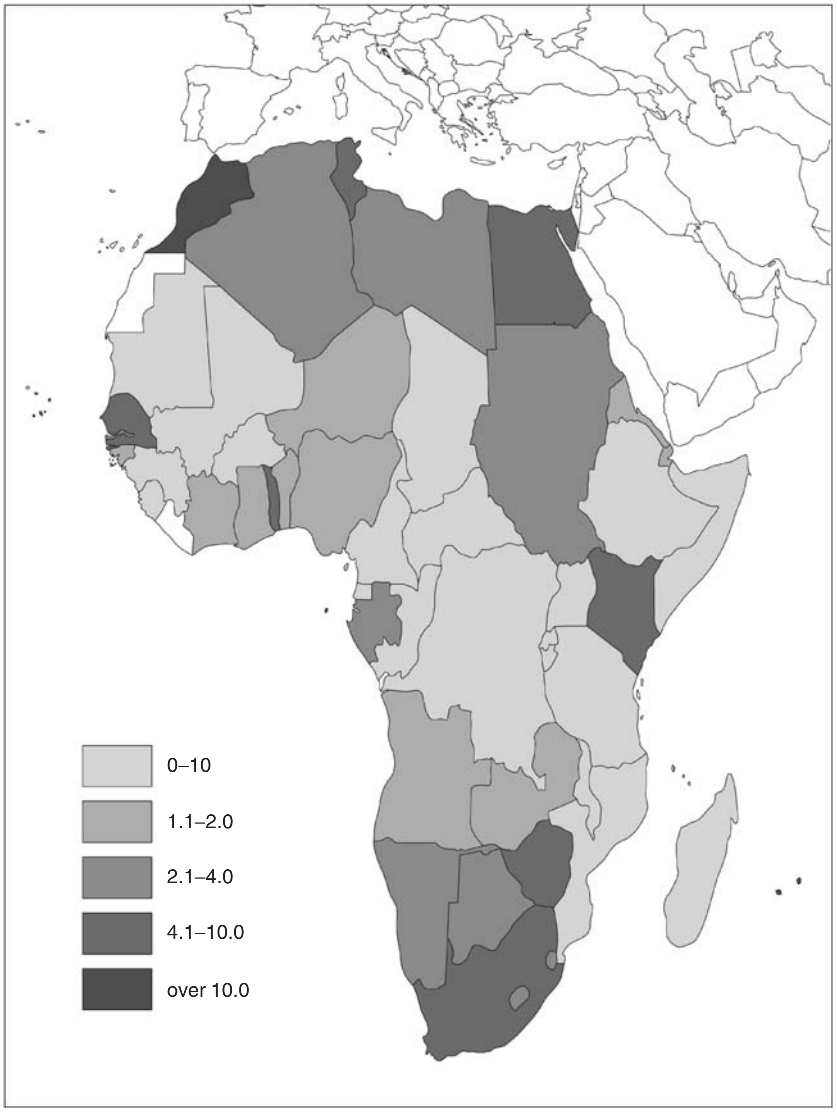

Figure 2 Internet users per 100 inhabitants in Africa, 2004. ITU Africa ICT Indicators.

2.7 million new lines in the USA, or the 1.2 million new lines in the UK during the first two quarters of 2005), but these are heavily concentrated in Beijing, Shanghai, and Guangdong, with rural areas, especially in the west, remaining much less connected.

Shaping the Digital Divide

The above section has emphasized the ways in which it is possible to visualize the digital divide a range of scales. Here we highlight the main structural factors that shape such divides.

Connectivity

The digital divide is above all shaped by the presence or absence of appropriate infrastructure to enable the use of digital technologies. The availability and quality of ICT infrastructure, including both IT hardware and telecommunication networks, particularly the Internet, is referred to as connectivity. The term subsumes issues ranging from the global distribution of Internet backbones to the problem of the 'last mile' – actually connecting the individual to the Internet. The numbers of fixed lines, mobile telephones, or PC owners are just some of the ways in which disparities in connectivity can be measured. The advance of mobile technologies and wireless Internet has facilitated affordable connectivity in remote and rural areas, especially through the spread of cellular/mobile telephones and prepaid phone cards. In 1999, Africa had a total of 2.2 fixed line and 1.0 mobile subscribers per 100 inhabitants. By 2004, the number of mobile subscribers had increased to 8.8%, whereas the fixed line subscribers only rose to 3.2%.

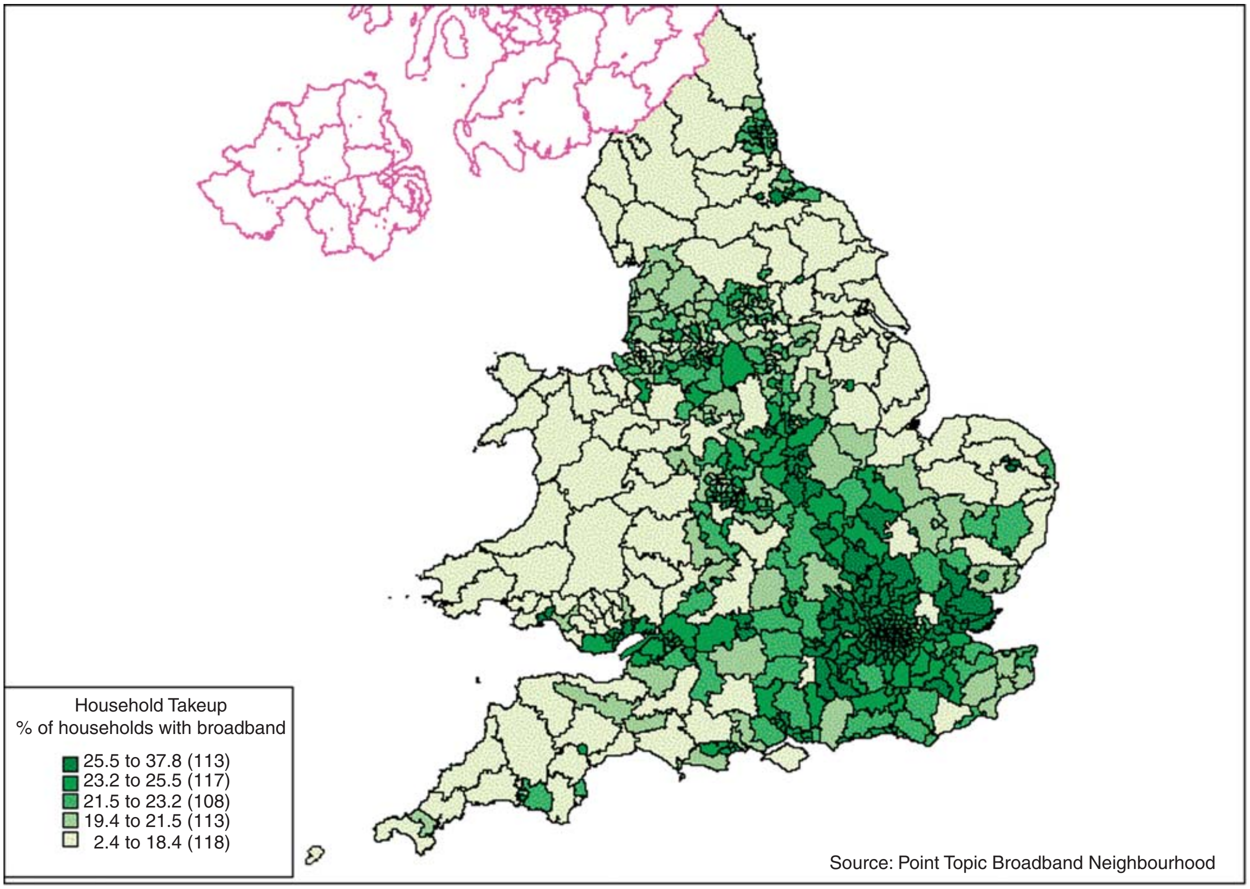

Figure 3 Broadband use in England and Wales, 2005. Reproduced from Point Topic (2005) *Broadband, the Internet and the Election*, with permission.

Accessibility

Although the physical access to ICTs remains unevenly distributed, the preoccupation with connectivity has lessened in recent years. The definition of the digital divide has been expanded to include the gap between those who have access to, and can effectively use ICTs, and those who cannot. In this context, the digital divide can be seen as a reflection of broader socioeconomic inequalities. Accessibility encompasses social, cultural, political, and economic factors that enable or prevent people from using digital technologies. Disparity in resource distribution is crucial in determining accessibility; income, occupation, and education all influence the extent to which people can use digital technologies. The significance of economic factors, particularly cost, can be seen in the correlation between levels of income and Internet usage, as well as the cost of access and Internet penetration rates. Likewise, the level of education, gender, age, and disability all determine the levels of access to ICTs.

Literacies

The emergence of ICTs has transformed the way in which we think about literacy. As ICTs become increasingly important for communication in the broadest sense, the inability to use them can disadvantage entirely new groups of people who previously thought of themselves as 'literate'. The elderly often, for example, find themselves marginalized by their inability to understand and use new communication technologies. However, ICTs also have the potential to enable previously 'illiterate' people to communicate effectively. Voice recognition software can thus help people to send written documents to others, even if they are unable to write themselves. Thus, the digital divide is fundamentally about the ways in which technologies are used to serve the needs of different people and communities.

Content and Information

The access to digital technologies can be meaningful only with the availability of relevant and affordable content and hence can become an integral part of people's lives. One important aspect of relevance is language. In general, digital technologies were first introduced and disseminated as mainly monolingual media, utilizing the language of the producers and not the consumers. Most were dominated by European languages, primarily

English. Due to the emergence of a multilingual character code standard during the 1990s, the number of languages being supported by major ICT desktop platforms has now increased. Although efforts to provide both open source and proprietary localized software are increasing, few Asian and African languages are yet catered for, with of course the major exception of Chinese. The use of local knowledge and adaptation of software to meet local demands is a prerequisite in utilizing ICTs effectively for educational, economic, or political purposes. With continued commodification of digital resources, affordability has also become an increasingly relevant issue, especially in education. There is a fear that the gap will widen between those who can afford to pay for, or have access to, such content and those who cannot.

Networks and Communication

Digital technologies, such as the Internet, have not only opened new channels for the distribution of information, but also have created new means of communication and transaction that facilitate different levels of networking, both interpersonal and through mass communication, irrespective of geographical location. Economic globalization processes and technological developments enabling and reducing the costs of information sharing and communication have resulted in what Manuel Castells calls ‘networked societies’. Organizational structures have been created that no longer conform to spatial centrality and conventional decision making hierarchies. Whilst in theory global participation in these networks is possible, in reality their participants are mainly found in the urban centers of the economically dominant nations. Enabling the usage of digital technologies is therefore not just about passive reception but includes active participation in the exchange and creation of knowledge and the ability to network and communicate freely using digital technologies, be it for educational, business, or political purposes.

Making a Difference

Having summarized the key aspects of the digital divide, we now focus on ways in which different types of organization have sought to reduce the inequalities that exist in access to, and use of, ICTs.

Global Agendas

The potential of ICTs to make a significant difference to the lives of poor people has been widely recognized at the international level. Initiatives such as the G8’s Digital Opportunity Task (DOT) Force (established in 2000), the UN ICT Task Force and most recently the World Summit on the Information Society (WSIS) held at Geneva in 2003 and at Tunis in 2005, have all sought to help create a global information society that will enable all people to benefit from the potential offered by ICTs. Whether they will achieve these ambitious objectives remains to be seen, especially at a time when many of the richer countries of the world are also seeking to remain ‘competitive’ knowledge economies. Nevertheless, UNESCO’s work in the field of cultural diversity and social aspects of the digital divide, as well as initiatives such as the UN’s Digital Diaspora Network, designed to promote development in Africa, through the mobilization of the technological, entrepreneurial, and professional expertise and resources of the African diasporas in the field of ICT, do have the potential to help bridge the digital divide. One of the key issues of contemporary global discussion is Internet Governance, which gained particular prominence at the 2005 phase of WSIS. This has led to an agreement to develop a more participatory approach to control the Internet and the creation of a forum, which will address issues such as connectivity, Internet backbones, and multilingualism.

Role of National Governments

Above all, the role of national governments is to provide regulatory and policy frameworks that enable the provision of affordable and accessible infrastructure, services, and content. This requires addressing issues ranging from trade and commerce, to security and informational freedom. Governments are expected to reduce digital inequalities, for instance, by providing ICT access to underprivileged groups within their territories. When telecommunication markets were liberalized in Europe, The German government guaranteed universal service to reduce the risk of people with low incomes losing access, by ensuring telecommunication services are provided to all households at affordable prices, including those in rural and remote locations. In developing countries, this concept is rarely adopted, due to lack of financial resources. Village computers and telecenters are examples of initiatives taken to provide connectivity alternatives in remote and rural areas in developing countries. Furthermore, national governments carry the responsibility of employing ICTs in administration and service provision, in order to increase efficiency, transparency, accountability, and opportunities for participation. Donor country governments are funding a number of such initiatives as their engagement has shifted from a technology driven approach to a focus on mainstreaming and accessibility issues in developing countries. Engagement to promote capacity building has increased, as has the integration of ICTs into specific sectors, such as education, healthcare, trade, and commerce. Beyond sector support, the assistance in policy formulation is also considered a central area of devel opment cooperation.

Private Sector

The private sector has a significant role to play in both bridging, and yet also in shaping, digital divides. It is a key partner in the provision of ICT infrastructure, ser vices, and applications, and through its ability to create competitive markets, lower prices, and increase the range of available options for citizens and organizations it can indeed help to bridge the divide. In total, tele communication projects accounted for 6% of foreign direct investments in low income countries in 2003. Latin America and the Caribbean attracted one half of those investments, whilst Eastern Europe and central Asia received a quarter. However, due to lack of profit ability, market driven private sector initiatives have nevertheless tended to neglect rural or remote areas in developing countries in preference for urban centers, and this often increases local digital inequalities. Multi stakeholder and public–private partnership initiatives have sought to encourage private sector engagement in addressing such underserved communities. As well as the generation of additional finances, one of the rationales behind private sector involvement in development ini tiatives is to increase economic growth and sustainability. Their aim is to open new markets with programs targeted at lower income groups. The Grameen Village Phone program in Bangladesh and Uganda, for example, pro vides mobile phones, financed by loans from the Gram een Bank, to individuals who in turn, sell their use on a per call basis. This business model aims to make mobile phone usage affordable and profitable in rural areas.

Civil Society and the Academy

The role of civil society is also important in delivering ICT access at the community level, in addressing the needs of citizens in policy formulation processes and in monitoring the rollout of such policies. This was recog nized by the UN in both phases of the WSIS and civil society organizations were formally involved in the pre paratory meetings as well as in the proceedings of the summit. The establishment of the international Civil Society Bureau for the second Preparatory Committee meeting represented a new approach in UN policy, in sofar as, for the first time, civil society was granted structural mechanisms to ensure effective participation at the decision making level of an UN summit. Issues concerning human rights, people centered sustainable development, and access to global knowledge were high on the agenda of civil society representatives partici pating in the WSIS and other international forums. Universities and research institutions also need to be involved in developing policy frameworks and undertaking rigorous monitoring and evaluation studies of ICT4D if the digital divide is indeed to be bridged.

Partnerships in ICT4D Practice

These illustrations have all emphasized that the digital divide cannot be 'bridged' by one community or sector alone. The ubiquity and diversity of ICTs means that complex partnerships that combine different interests and expertises need to be shaped, bringing together marginalized communities, governments, civil society, the private sector, and academic institutions, to help poor people use ICTs to help transform their life experiences.

Solutions in Search of a Problem

Many ICT projects designed to use innovative methods to 'bridge the digital divide' have failed, often because they have been designed in a top down and supply driven manner; often primarily serving the external interests of donor organizations or the private sector. There is a critical need to shift the balance toward a needs based approach, so that poor and marginalized communities are themselves involved in designing and shaping the implementation of such initiatives. However, two further issues must also be addressed if the un doubted benefits that ICTs can bring are to be realized.

Knowledge Societies in a Competitive World

The concept of a knowledge based society and economy suggests a focus on the development of an individual's capacity as a carrier of professional knowledge and as a member of institutions and society, in which knowledge needs to be effectively acquired and shared. ICTs are the medium through which individual and institutional knowledge can be stored and retrieved, transferred, and communicated. Whilst the attainment of equity in the distribution of digital technologies and resources is a goal on the agenda of the international development com munity, leading producers of hardware and software, as well as content distributors also carry an interest in maintaining their competitive advantage. Bridging digital divides can also be a way of opening up new markets and creating new consumers. Disputes in the scientific pub lishing sector are just one example of how the interests of commercial publishers and content producers conflict with the ideal of an open networked society in which relevant knowledge can be shared and accessed freely, especially in developing countries.

Digital Commons

From its inception as a medium through which knowledge could be exchanged 'freely', the Internet has now evolved to incorporate market forces with commercial providers charging for content, particularly through licensing fees. This is part of a wider trend in which knowledge is be coming increasingly commodified. The growing Free and Open Source Software community, the Open Access movement, and Creative Commons are all expressions of rising tensions between the protectors of 'proprietary' and supporters of 'free' digital content. The relevance of such alternative approaches to resource acquisition in developing countries has now become of particular pertinence. In comparison with the acquisition of propri etary content, the promotion of Open Source software communities and its uses in the public and private sector, civil society, and academia are increasingly considered to provide more sustainable development opportunities. The academic world has also reacted to the commercial monopolization of intellectual property with the initiation of the Open Access and Open Educational Resource movements, as well as by pursuing the creation of free to end user, and global access to scientific literature.

Conclusions

There are many digital divides – social, economic, political, spatial, and even ideological. However, it is not the technology that has created these, but rather the ways in which people have chosen to use ICTs to gain economic advantage, propagate a particular political message, or reinforce social divisions, that have forged the unequal digital world that we experience today. These technolo gies, though, have immense potential to help transform the lives of poor and marginalized communities. They can open up a whole new world to people with visual and hearing impairments. They can transform the lives of people with physical disabilities, and can enable young people who are unable to attend school to gain learning experiences of relevance to their lives. Whether or not they are indeed used to transcend existing digital divides, and to create a truly shared world of information and communication opportunities depends on the political will and the economic capacity of global partnerships to help poor and marginalized communities to deliver sustainable solutions through which they can be empowered.

See also: Civil Society; Empowerment, Foreign Direct Investment; Knowledge Economy.

Further Reading

Castells, M. (2000). *The information age: Economy, society and culture* (2nd edn.). Oxford: Blackwell.
Gelbstein, E. and Kurbalija, J. (2005). *Internet governance Issues, actors and divides.* Malta and Kuala Lumpur: DiploFoundation and Global Knowledge Partnership (http://textus.diplomacy.edu/textusbin/env/scripts/Pool/GetBin.asp?IDPool = 641).
Grameen Technology Center (2005). *Village phone replication manual Creating sustainable access to affordable telecommunications for the rural poor.* http://www.infodev.org/files/2868 file VillagePhoneReplicationManual.pdf
Greenberg, A. (2005). *ICTs for poverty alleviation: Basic tool and enabling sector*. Stockholm: Sida.
InfoDev (no date). *Open source software Perspectives for development.* Washington DC: InfoDev (http://www.infodev.org/files/837 file Open Source Software.pdf).
ITU (2005). *World telecommunication/ICT development Report 2006: Measuring ICT for social and economic development*. http://www.itu.int/ITU D/ict/
Norris, P. (2001). *Digital divide: Civic engagement, information poverty, and the internet worldwide*. New York and Cambridge: Cambridge University Press.
OECD (2005). *Financing ICTs for development Efforts of DAC members, review of recent trends of ODA and its contribution Report to the UN Task Force on Financial Mechanisms for ICT for Development (TFFM)*. Paris: OECD, Development Assistance Committee (DAC). http://www.oecd.org/dataoecd/41/45/34410597.pdf
Paolillo, J., Pimienta, D., Prado, D. *et al*. (2005). *Measuring linguistic diversity on the internet*. Paris: United Nations Educational, Scientific and Cultural Organisation, UNESCO.
Point Topic (2005). *Broadband, the internet and the election*. http://www.point topic.com/content/dslanalysis/UK%20election%20guide.htm
Schware, R. (ed.) (2005). *e Development: From excitement to effectiveness*. Washington DC: The World Bank Group.
Unwin, T. (2005). *Partnerships in development practice: Evidence from multi stakeholder ICT4D partnership practice in Africa*. Paris: UNESCO for the World Summit on the Information Society.
Weigel, G. and Waldburger, D. (2004). *ICT4D Connecting people for a better world*. Berne and Kuala Lumpur: Swiss Agency for Development and Cooperation and Global Knowledge Partnership.
World Bank (2006). *2006 Information and communications for development (IC4D): Global trends and policies*. Washington DC: World Bank.
WSIS Declaration of Principles (2003). *Building the information society: A global challenge in the new millennium*. Geneva: UN Summit of the Information Society (Paragraph 28, Paragraph 22 (C7, b). http://www.itu.int/dms pub/itu s/md/03/wsis/doc/S03 WSIS DOC 0004!!PDF E.pdf).

Relevant Websites

http://creativecommons.org
Creative Commons Developing Nations License.
http://www.itu.int
ITU statistics.
http://www.itu.int
WSIS statistics.

Digital Earth

G. A. Elmes and D. Weiner, West Virginia University, Morgantown, WV, USA
C. D'Alessandro-Scarpari, Université Lumière Lyon 2, Lyon, France

Glossary

DERM The Digital Earth Reference Model, templates for the standards and specifications to enable Digital Earth interoperability.
Digital Earth A virtual, multiresolution, three-dimensional (3-D) representation of planet Earth that enables a person to explore and interact with natural and social information at many scales of resolution.
Discrete Global Grid A set of areal cells that form a partition of the Earth's surface, usually represented by some form of topologically equivalent approximation such as a sphere, a spheroid, or a geoid.
Framework Data A set of interoperable spatial data themes forming the basic content for a geographic information system (GIS). Usually includes geodetic control, cadastral, elevation, hydrology, administrative units, transportation networks, and orthoimagery.
Geobrowser A program or tool allowing visual, spatial, and other searches for precise geographical data.
Geographic Information System (GIS) An organized collection of computer hardware, software, geographic data that allows spatial or geographically referenced data to be captured, stored, updated, manipulated, and displayed.
Geolibrary A digital collection of text, maps, images, and other materials, cataloged and searchable on a geographical basis, that is, by location.
Georeference To establish a relation between coordinates on a map and real-world coordinates.
Interoperability The ability of two or more systems or components to exchange information and to use the information that has been exchanged.
Metadata Information relating to another set of data ('data about data').
Microsensor A miniature electronic device that detects information about a specific variable.
Open Systems Open systems entail hardware and software in a system that meet standards enabling them to be replaced with that of another vendor. Open source code is available to the public, it does not necessarily meet open standards.
Remote Sensing The science of acquiring information about the Earth's surface by sensing and recording reflected or emitted energy and processing, analyzing, and applying that information.
Spatial Data Infrastructure The technology, policies, standards, and people necessary to permit geospatial data sharing among users.
Spatial Multimedia The combined use of georeferenced media, such as movies, music, speech, CD-ROMs, and the Internet, as for education or entertainment.

Introduction

As initially envisioned, Digital Earth (DE) is a virtual, multiresolution, 3 D representation of our planet that enables a person to explore and interact with the vast amounts of natural and social information gathered about the Earth. Both a metaphor for the representation of the planet and an intellectual driving force for the endeavor to map the planet comprehensively and seamlessly, Digital Earth is an integrated storehouse and source for many forms of geographically referenced information. Eventually Digital Earth is predicted to be for geospatial data what html and Transmission Control Protocol/ Internet Protocol (TCP/IP) are for the World Wide Web. The various efforts for realization of Digital Earth have profound technological, social, political, environmental, and economic implications.

At one level, Digital Earth is a 3 D graphic user interface or geobrowser to a distributed geographic in formation system (GIS). High speed, broadband networks linking software and hardware bring together and process georeferenced data from multiple sources. An interactive, immersive interface grants users access to a represen tation of the Earth, which assimilates numerical data, text, images, and maps, along with the relationships be tween them and models of their underlying processes. From another perspective, Digital Earth is a less formal amalgam of multiple innovations, some cooperative, some competitive, which bring geospatial and other infor mation to broad audiences. Yet, as the product of the shifting convergence of rapidly changing technology and computational capabilities, Digital Earth is mutable, and while its current manifestations range from giant globes to Internet browsers, it is difficult to predict exactly what new versions will emerge in future.

Digital Earth has an intellectual history that draws on deep seated conceptions of space, scale, and place. Themes of technological significance include standards, visualization, and spatial analysis of global data.

The implications of universally available global data on human scale relationships and society at large are diverse and distributed unequally among the human population. Foremost among these human scale questions are issues relating to the ethics of surveillance, privacy, power, and consent.

An Intellectual History of Digital Earth

The first use of the term Digital Earth is generally attributed to US Vice President Al Gore, although the history of the idea is much older. In a much cited speech published in 1998, Gore defined Digital Earth as a multiresolution, 3 D representation of the planet, into which we can embed vast quantities of georeferenced data. One antecedent to the concept is common to cartography in many cultures, an endeavor to represent the Earth as a geometric solid, usually a sphere, taking the Earth's surface as an organizing metaphor for information about places. A second antecedent is the goal to make information ubiquitously available as an essential component of democracy, and also in attempts to manage the global environment, and create sustainable economies. A third contribution is the aspiration to create tools which are able to represent and analyze data across multiple scales from the global to the local (see **Figure 1**). Inputs to this goal stem from radical transformations in computing that have resulted in massive capabilities for data sharing and processing across networks; the ability to combine data of many different types from multiple, distributed sources; and search engines (web browsers) that are able to find and link information about any particular location on the planet.

The nature of spatial data as a dyad of location and attributes is a defining feature of GISs. The concept of a Digital Earth emerged within cartography, and from political, environmental, educational, and technological initiatives. The globe is one locational metaphor for recording information about places. Terrestrial and celestial globes have been known since antiquity. Celestial globes serve as an aid to astronomy whereas terrestrial globes serve as a demonstration of mathematical and geographical hypotheses. Martin Waldseemüller was probably the first to use printed gores for recording cartographic information in 1507. Both aspects of organizing framework and content are present in the globe as a model for representing the Earth's surface.

The innovations of the industrial age enabled the size and scale of representations of the Earth to increase dramatically. Constructed in Paris in 1826, Charles Delanglard's Georama was a 10 m diameter globe that could be entered as an immersive environment. A similar Great Globe, 20 m in diameter, was designed and built in Leicester Square in London by James Wyld, and stood as a tourist attraction from 1851 to 1862. Accurate geographical information, showing physical and cultural features, was imprinted on the concave interior surface. In the quest for ultimate accuracy, the French geographer Elisée Reclus proposed a globe, to the scale of 1:100 000, which would have necessitated the construction of a sphere nearly 120 m in diameter.

The step from increasingly accurate, large scale physical renditions of the Earth to electronic ones was proposed by Buckminster Fuller as early as 1947. He envisioned the Geoscope, a 200 ft diameter geodesic sphere or hemisphere, suspended 100 ft above the ground. A computer was to control the intensity of

Figure 1 Early NASA Vision of Digital Earth. Courtesy of NASA.

electric light bulbs that entirely covered the exterior and interior surfaces. The number and arrangement of the bulbs was such as to produce the visual resolution of a fine screen halftone print, or that of a cathode ray tube picture. Had it been built, Geoscope would have provided an immersive media in which to discover vast amounts of geospatial information. As a tool for Earth stewardship revealing formerly invisible patterns, it was designed for forecasting and planning.

Images of Earth's sphere were first made available by extraterrestrial flight in the 1960s, followed by the widespread dissemination of the analogy of Earth as a blue marble, notably in a historic series of NASA photographs. The launch of a sequence of LANDSAT remote sensing satellites was instrumental in transforming the visual image of Earth to limitless quantities of renewable geospatial data. Along with the source of digital imagery, the Eames' 1977 film, *Powers of Ten*, was an intellectual milestone in the conceptualization of Digital Earth. *Powers of Ten* depicted the relative scale of the universe from 10^{-18} to 10^{25} m. The film zooms out ten times for every 10 seconds from the image of a sleeping man at 1 m resolution (10^{0}) to the scale of the universe (10^{25}) and back to the resolution of an atom (10^{-18}) illustrating a continual seamless transition across scales. In 1992 David Gelernter's *Mirror Worlds* extended the analogy of the microscope and telescope from film to massive computerized information systems and models of complex geographical realities, such as cities and other social phenomena. *Mirror Worlds* put forward computer based viewing tools focused not on the hugely large or small, but on the human scale, social world of organizations, institutions, and machines; promising the same microscopic and telescopic increase in depth, sharpness, and clarity illustrated in *Powers of Ten.*

Advancing these data visualization tools required the acquisition of multiple types of data from many remote locations, integrating them and making them accessible to the public through networks. Global representation, visualization, transformation, and resolution across scales, increasing demand for accurate large scale data, and computerization form the intellectual threads that lead to the Digital Earth initiative as articulated in 1998.

Digital Earth and Technology

The development of the Digital Earth has proceeded from 'vision' to several demonstration prototypes and practical implementations in less than a decade. When originally proposed, Digital Earth had two elements: (1) a satellite based remote sensing platform directed at the daytime face of the Earth, and (2) an immersive environment for exploring Earth's societies and environment. The specific satellite component has disappeared, although the US Earth Observation System and satellites from many nations perform similar, if not identical functions. The 3 D immersive interactive environment element has proved robust (see **Figure 2**). As in a conventional GIS, Digital Earth provides for the integration of geospatial and statistical data, with the additional ability to integrate, simulate, and analyze knowledge of human and physical systems in an interactive interface. Google Earth and Microsoft Virtual Earth are contemporary representatives of several prototype 3 D geobrowsers that have been developed and implemented on the Internet.

Technical Foundations

The Interagency Digital Earth Working Group (IDEWG) identified areas of knowledge and technology as foundations for the Digital Earth project and was subsequently adopted by international teams. The fields are computational science, mass storage, satellite imagery, and other georeferenced data, broadband networks, interoperability, and metadata.

Computational Science

Today's computers integrate and analyze huge quantities of raw geospatial data and simulate complex phenomena and processes that cannot otherwise be measured directly. High performance machines integrate data from numerous sources to create forecasts, scientific understanding, and outcomes of nonlinear, dynamic, even chaotic processes. The continued operation of Moore's Law (the doubling of the number of transistors on integrated circuits every 18 months) is a rough measure of the exponential increase in computer processing power which is expected to continue as nanotechnology is implemented. For example, high resolution spatial and temporal modeling of geophysical and earth science

Figure 2 Desktop 3D interface.

processes at a global scale has been the ostensible reason for building the Earth Simulation Model in Japan, one of the fastest computers on Earth in the first years of the twenty first century.

Mass Storage and Imagery

Mass storage capacity is increasing rapidly and simul taneously becoming cheaper, but a fully implemented Digital Earth will require the storage of petabytes (10^{15} bytes). Goodchild estimated that one petabyte is required to store a seamless satellite image of the Earth at 1 m resolution. NASA's Mission to Planet Earth alone gen erates a terabyte of information each day and 1 m reso lution imagery or better has become routinely available from a number of commercial satellites. On the basis of regular updates from high resolution satellite and aerial imagery, cartographic framework data, such as land cover classifications; road and river networks, are derived and kept current. Satellite imagery and other environmental data, acquired by large number of organizations and in creasingly sampled at very high temporal and spatial resolution by distributed microsensors, require broad band networks to connect data producers to servers and thence to users.

Networks and Interoperability

These data do not reside in one huge database. Servers, connected by high speed broadband networks, permit the transmission of data objects and visualizations. At the leading edge of technology 10 gigabit per second net works have been tested, carrying data streams up to 1000 times larger than is currently available using the World Wide Web. Public access to the network for transmission, interaction, and collaboration is proliferating. Ease of access to the World Wide Web has succeeded because of the emergence of interoperability through widely agreed upon, simple practices, such as the IP. GIS researchers and developers address many of the issues surrounding spatial data transmission and sharing through the Open GIS Consortium and the development of formal inter national standards. For example, the technical committee of the International Standards Organization (ISO) TC211 establishes a set of standards for information concerning objects or phenomena that are directly or indirectly as sociated with a location relative to the Earth. TC211 standards specify, for geographic information, methods, tools, and services for acquiring, processing, analyzing, accessing, presenting, and transferring such data in digital/electronic form between different users, systems, and locations. One reality and drawback of the generation of formal standards is their slow pace of development and implementation. In the fast paced highly competitive environment that characterizes the computer industry, less formal, *de facto* mechanisms for creating the spatial data infrastructure are based on the exchange of innov ations in best practice at forums such as the Where 2.0 Conference.

Metadata

Data about data, or metadata, are vital for georeferenced information to be usable from one computer, application, or jurisdiction to another. The name, location, source, date, data format, resolution, and accuracy are categories of information necessary for the meaningful and reliable combination and use of geospatial data. Technology is also advancing for the fusion of different data types from multiple sources, automated interpretation of imagery, and the deployment of intelligent computer programs (agents) to seek, gather, and integrate information based on a locational footprint. Valid outcomes of these func tions depend on the availability of complete metadata. At the same time, developments in location based mobile and ubiquitous computing, in spatially enabled radio frequency identification (RFID), and other sensors are creating real time sampling over large areas; creating a digital microgeography, which can be seen as the equivalent of magnetic resonance imaging (MRI) and computed axial tomography (CAT) scans in medicine. These developments place ever greater requirements on technological advances in storage, transmission, inter operability, and metadata.

Development and Realization of Digital Earth

Following Vice President Gore's statement, the IDEWG organized the initiative in the US. Representatives from 17 federal agencies, together with industrial and aca demic advisors, investigated topics related to visual ization, data access and distribution, standards, education, and others. Political change in the US, however, has re sulted in change in leadership for Digital Earth. Since 1999, the Chinese Academy of Sciences' (CAS) Institute of Remote Sensing Applications (IRSA) has led inter national collaboration on the Digital Earth concept and the *1st International Symposium on Digital Earth* was held in Beijing. In cooperation with a growing community of international agencies and NGOs, the Chinese govern ment continues to sponsor international symposia and projects. Today, the International Society for Digital Earth (ISDE) fosters the international Digital Earth community. Following the meetings in Beijing, there have been three further international symposia on Digital Earth; in Canada, the Czech Republic, and Japan. The 5th International Symposium on Digital Earth was held in Berkeley, California in 2007.

Three major projects involving global data and standards are in progress: The Global Map, Global

Spatial Data Infrastructure (GSDI), and Digital Earth. Global Map is a product of the International Steering Committee for Global Mapping (ISCGM). Initiated by the Geographical Survey Institute of Japan, the devel opment of Global Map resulted from Agenda 21, the document calling for global data that was an outcome of the 1992 United Nations Conference on Environment and Development (UNCED). A global cartographic framework database is planned, consisting of elevation, vegetation, land use, drainage systems, transportation, and administrative boundaries layers. The Global Spatial Data Infrastructure (GSDI) is defined as the policies, organizational remits, data, technologies, standards, de livery mechanisms, and financial and human resources necessary to ensure that those working at the global and regional scale are not impeded in meeting their object ives. The work of GSDI is accomplished by represen tatives of national mapping agencies, NGOs, and standards organizations. The success of the Digital Earth project is directly related to Global Map, and other global database and infrastructure programs as outlined in the GSDI Cookbook. Digital Earth provides a vision that is tied to the maturation of the Global Map, GSDI, and associated programs.

A useful scheme for defining the major development communities and concentrations necessary for the real ization of Digital Earth has been identified by Foresman and others as follows:

1. Visualization and exploration: the hardware, software, data, and methods for viewing and exploring Digital Earth.
2. Advance prototypes: the projects, test beds, and fa cilities through which the Digital Earth is tested and promoted.
3. Science and applications: the community that de velops and validates Digital Earth content and outcomes.
4. Standards and architecture: the community that en sures the infrastructure and interoperability protocols for a sustainable Digital Earth.
5. Data access and distribution: the gatherers, producers, and distributors of georeferenced data.
6. Education and outreach: the users of Digital Earth.

Digital Earth is dependent upon the cross referencing and coordination of technical innovations in the six areas of concentration. The Digital Earth Reference Model (DERM) was created to provide operational standards to enable these areas to work together. The DERM was the result of a public and private partnership to define the standards and architecture guidelines of Digital Earth. As an efficient approach to a rapidly changing, technology driven field, the successors of DERM, for example, the Geospatial Interoperability Reference Model (GIRM), remain significant reference tools for rapid prototyping. By 2002, a number of 3 D visualization prototypes be came available, and although the formal role of NASA as lead agency had ended, commercial and international interests took up the initiative. For example, Keyhole Inc. was a private venture, based on DERM and supported under the auspices of IDEWG that became the first widely available 3 D geobrowser. It is now available as Google Earth.

For political, technological, and economic reasons, the concept of Digital Earth has evolved in different forms than originally proposed. Arguably the strongest research initiatives have been associated with distributed geoli braries: including the NCGIA's Alexandria Project, (NCGIA – National Center for Geographic Information and Analysis), NASA's Land Processes Distributed Active Archive Center, which includes the Earth Observing System Data and Information System, and the discrete global grid initiatives, such as the 2nd International Conference on Discrete Global Grids. Other practical developments have been realized by the Federal Geo graphic Data Committee (FGDC) in the National Spatial Data Infrastructure (NSDI). In the private sector, SRI International's, Defense Advanced Research Projects Agency (DARPA) sponsored Digital Earth project in corporates Gore's vision of a connected suite of tech nologies with the vision of enabling a massive, scalable, and open model of the planet where millions of users can interact with vast quantities of geographically referenced data over the Web. SRI's stated goal is to create an open infrastructure that allows anyone around the globe to publish or to search for data based upon a specific lo cation, however, SRI is a major military contractor, so the real motivation for the effort may well be hidden. Most recently in the face of competition from web based computing, notably from Google, Microsoft has launched Virtual Earth. Currently various examples of Digital Earths are flourishing. The vision is indeed reality, albeit in as yet partial and immature form.

Space and Scale in the Context of Digital Spatial Information

One approach in human geography encourages the critical examination of the roles and effects of technology within society. As with all landscapes, the landscapes of Digital Earth are socially produced and contested; their meanings are layered, variously hidden and revealed, apparently neutral and objective, but in fact subjective, ambiguous, and complex. One direction of critical inquiry is the consideration of representation and its limits. Language is also key to the production and control of, and access to, spatial information. The politics of space and scale are expressed in control and censorship; in what is chosen to be represented in Digital Earth, and by what is omitted. Scale reproduces, reinforces, and reflects power; con versely power enforces scale. Many of these concerns have emerged from the debate on GIS and society.

Digital Earth and Society

Since 1995, a lively GIS and society debate on the im plications of geographical technologies has lead to inquiry into the relationships between science, technology, and human beings in the broad context of Digital Earth. As with all technology, the creation and effects of Digital Earth on society are seen to be diverse, contested, am biguous, and contradictory. Within the GIS and society perspective, an institutional theme focuses primarily on the effectiveness, efficiency, and equity of the implemen tation and use of geospatial information in public settings. The institutional focus emphasizes the positive economic, educational, and political contributions of the knowledge economy, sharing geospatial information, and the dem ocratizing power of increasing accessibility to information of all types. Digital Earth viewed from this aspect will be a major contributor to economic development; a technology that flattens the Earth and provides a more even playing field for global competition.

A second GIS and society theme, grounded in the various concerns of critical social theory, approaches Digital Earth from a more reflective and circumspective point of view. Critical social theory imparts a political economic focus on the production and control of digital spatial information. Understanding the many dimensions of political and institutional frameworks by and through which data are produced and distributed is a key to this method of analysis, including the appreciation of the ability to control, manipulate, and possibly subvert, Digital Earth data and subsequently generated information. To date, one common aspect of the various initiatives has been to make global data available through a single portal. There is vigorous competition to become the primary Digital Earth portal which may be interpreted as one effort to exercise control. Numerous private firms (Glo beXplorer, Google, Informix, and Microsoft corporations, among them) and some government (NASA) and non government organizations (United Nations Environment Programme, UNEP) are positioning themselves to be come principal distributors of Digital Earth data. Inter nationally, China, Japan and other forces seek to influence the direction of development.

A third theme in GIS and society examines how access to digital spatial information at multiple scales influences the relationships of power between creators and users and how, in turn, power relations influence access and participation. Individuals and groups may become spatial actors through the mediation of Digital Earth technology, or may be excluded, inadvertently or otherwise. Case studies of the Digital Earth project encompass social and environmental processes at multiple scales of represen tation and analysis. This global and globalizing technol ogy presents influential and apparently official and authoritative information to users in local contexts. The interface of global and local raises potential conflicts between global and locally inscribed data and knowledge. The inscription and inclusion of local knowledge is a key element in participatory and community GIS, issues that must be resolved in Digital Earth, if it is to make a contribution to the economies and societies of developing nations and is not to aggravate marginalization.

Much of the core spatial data for Digital Earth is produced by national mapping organizations (NMOs). In this guise cartography, has long been employed by the state for territorial control and private ownership. The decision to create or not to create data, its accuracy and attribution, and to disseminate it, is a political and eco nomic one. For example, data created for military pur poses may find its way on to Digital Earth only in censored, generalized forms, or long after it was acquired. While digital spatial data are produced by states, regional organizations, counties, municipalities, and grassroot or ganizations, the reasons and purposes for its creation may not be apparent or well documented.

A view currently presented on Google Earth provides little information on the lineage of the data. Many people consider the images and information to be in real or near real time. Others are surprised at the great variations in spatial resolution available from one location to another. If people do not understand the origins of the data and their subsequent lineage, the value of the information pre sented is substantially reduced, if not totally undermined.

Scientific visualization in Digital Earth privileges tangible information that is easily rendered over in tangible data of phenomena, processes, and relationships. Terrain is easily depicted in four dimensions; human beliefs, feelings, experiences, and interactions are not. The emphasis on the visual serves to strengthen positivist and objectified conceptualizations of the world and human activity at the expense of the daily practices of life. In assessing digital places, Curry and others have spoken eloquently about the limits of representation and of language in geographic information technologies. The power of the visual, coupled with the potential of spatial analysis and modeling for public and private decision making, reinforces the capacity of Digital Earth for in corporating unseen, deliberately introduced inaccuracy, and the opportunity for misuse of analytical results and predictions. An occidental view and interpretation of the world is an inevitable result of current Western scientific hegemony and the predominance of the English lan guage. As these globally available data are mediated through locally contingent applications, they produce, confirm, or challenge particular spatial representations and imaginary landscapes. There are a multitude of di verse participatory GISs in use today, but few of them readily admit local understandings of environment and place on an equal footing with traditional spatial data derived from 'expert' sources. With the emergence of mashups in the Google Earth environment (a mashup is a website or web application that seamlessly combines content from more than one source into an integrated practice), and the Open Source standards of NASA's World Wind, the incorporation of diverse types of local information within Digital Earth is becoming easier and, superficially, more accessible to the population at large. However, the necessary access to high performance personal computers and broadband access to the Internet is denied to the vast majority of people, and is likely to remain so for the foreseeable future.

Although access to the Internet is considered ubi quitous in the industrialized world, this is a patent oversimplification of the reality. Gender, class, ethnicity, education, all mediate access to computers; hence access to Digital Earth, even in the developed world, is similarly constrained. A growing digital divide determines access, and lack of access continually accelerates and exacerbates the division of knowledge and information. In this regard, Digital Earth is simultaneously part of the problem and of the solution.

See also: Critical GIS; Geodesy; Imperialistic Geographies; Local-Global; Mapping, Cyberspace; Remote Sensing; Scale.

Further Reading

Cosgrove, D. (2001). *Apollo's Eye: A Cartographic Genealogy of the Earth in the Western Imagination*. Baltimore: The Johns Hopkins University Press.
Curry, M. R. (1998). *Digital Places: Living with Geographic Information Technologies*. London: Routledge.
Elmes, G., Epstein, E., Sheppard, E. and Tulloch, D. (2004). GIS and society. In McMaster, R. & Usery, L. (eds.) *Research challenges in GIS*, pp 287 312. Boca Raton, FL: CRC Press.
Foresman, T. W. (2003). Digital Earth: The status and the challenge. In *Proceedings of the Global Mapping Forum*. Ginowan, Okinawa: International Steering Committee for Global Mapping, Japan Geographic Survey Institute.
Foresman, T. W. (2004). Digital Earth visualization and web interface capabilities utilizing 3 D geobrowser technology. In *Proceedings of the 20th ISPRS Congress, Commission IV*, pp 885 889. Istanbul: IRPRS.
Foresman, T., Walker, S. and Zimble, D. (2003). Mandate for Digital Earth Geobrowsers: Status and Recommendations. *First International Digital Earth Workshop*, July 2003, http://www.ncgia.ucsb.edu/digitalearth/3DGeobrowser/files/mandate2003.pdf (accessed on 18 July 2007).
Fuller, R. B. (1962). *Education Automation: Freeing the Scholar to Return to his Studies*. Carbondale: Southern Illinois University Press.
Gelernter, D. (1992). *Mirror Worlds: The Day Software Puts the Universe in a Shoebox…How it will Happen and What it will Mean*. Oxford: Oxford University Press.
Gore, A. (1998). The digital earth: Understanding our planet in the 21st century. California Science Center, Los Angeles, California, January 31, 1998. http://www.isde5.org/al gore speech.htm
National Research Council (1999). *Distributed Geolibraries: Spatial Information Resources, Summary of a Workshop, Mapping Science Committee, Commission on Geosciences, Environment and Resources*. Washington DC: National Academies Press.
Nebert, D. D. (2004). *Developing Spatial Data Infrastructures: The SDI Cookbook*, Version 2.0. Reston, Virginia: Global Spatial Data Infrastructure Association.
Seigfried T. (2000). *Machines and Metaphors*. In *The bit and the pendulum: from quantum computing to M Theory The new physics of information*. New York: John Wiley and Sons, Inc.

Relevant Websites

http://www.isde5.org
5th International Symposium on Digital Earth.
http://rsd.gsfc.nasa.gov
Blue Marble, Visualization of Remote Sensing Data.
http://www.ncgia.ucsb.ed
Digital Earth 3D Geobrowser.
http://www.discreteglobalgrids.org
Discrete Global Grids.
http://www.earthscope.com
EARTHscope.
http://www.gsdi.org
Global Spatial Data Infrastructure Association.
http://earth.google.com
Google Earth.
http://local.live.com
Microsoft Virtual Earth.
http://worldwind.arc.nasa.gov
NASA, World Wind.
http://www.opengeospatial.org
OGC, Open Geospatial Consortium, Inc.
http://www.powersof10.com
Powers of 10 CD and online multimedia site for Charles and Ray Eames' 1977 film "Powers of Ten".
http://www.ai.sri.com
Digital Earth, SRI.
http://www.geosphere.com
The GeoSphere Project, The Earth Imagery.
http://edcdaac.usgs.gov
The Land Processes Distributed Active Archive Center (LP DAAC), Earth Observing System (EOSDIS).

Disability and Chronic Illness

R. Wilton and J. Evans, McMaster University, Hamilton, ON, Canada

Glossary

Ableism It is a term used to name the systematic oppression faced by disabled persons as a minority population in a society where the 'able-body' occupies a position of privilege.

Embodiment The notion that disability as a lived phenomenon is, by its very nature, experienced through the body. This concept is used to move beyond the impairment/disability binary of early social model formulations.

International Classification of Functioning, Disability, and Health It is a conceptual scheme used by the World Health Organization to describe and compare people's ability to function physically and socially while living with an array of different health conditions and impairments.

Participatory Action Research The research that questions traditional researcher/participant roles and requires researchers to involve participants in the design, implementation, and analysis stages of a project. Emphasis is placed on conducting research of immediate relevance to the participant population, and using findings to facilitate social change.

Social Model of Disability A conceptual approach to disability that emphasizes the oppressive character of the social and built environments as a primary cause of disabled people's inability to participate in mainstream, everyday life.

Human geographers have made numerous theoretical and applied contributions to current understandings of disability and chronic illness. These contributions have engaged with many different impairments and disability experiences, employing a number of spatial theories and techniques in the process. In what follows, we review and reflect on disability and chronic illness research in human geography. We begin with some basic definitions and key concepts encountered within social scientific discussions of disability and chronic illness. We then move on to summarize the diverse geographical approaches taken by human geographers in recent years. Finally, we reflect on the multiple challenges encountered when carrying out disability research and the research gaps that remain to be examined.

Definitions and Key Concepts

The way we define disability and chronic illness has fundamental implications for how we make sense of the human geographies of people living with a diversity of conditions, including hearing impairment, intellectual disabilities, and illnesses (such as arthritis and fibromyalgia, spinal cord injuries, and epilepsy). Broadly speaking, we can distinguish between individualized medical definitions of disability and chronic illness, and those that emphasize the social, cultural, and/or political character of these conditions. Medical conceptions of disability and chronic illness tend to focus on the specifics of the 'problem' confronting an individual's body or mind. The following excerpt is from a neuroscientist describing the nature of a spinal cord injury:

> Take, for example, a person who has had a burst fracture of the C5 vertebral body… Due to edema (swelling of the spinal cord), the biceps (C5) may be initially weak but should recover. The wrist extensors (C6), however, should remain weak and sensation at and below C6 should be severely compromised. (Young, 2002)

Here, the 'burst fracture of the C5 vertebral body' produces a series of outcomes including the loss of the use of the individual's legs. Appropriate responses may include surgical interventions to maximize existing movement, efforts to deal with changes in the body's capacity for bladder and bowel control, and a process of rehabilitation to assist the individual in learning to live with a spinal cord injury and trying to return as much as possible to a 'normal' life. Conceptually, this individualized conception of disability begins with the impact of an injury – or a congenital defect or chronic illness – on the functional capacity of an individual, with particular attention to the ways in which medical specialties might repair and/or rehabilitate the body. To the extent that medical interventions are possible, individuals may be able to return to 'normal'. Where unsuccessful, individuals are disabled by their diminished functional capacity.

Given the status of biomedicine in Western society, this conception of disability and chronic illness as stemming exclusively from the individual body has had the status of commonsense in the twentieth century. However, several other ways of conceptualizing disability have emerged over the past 40 years that challenge the medical conception. While distinct, they share three things in common. First, they are explicitly collective in orientation, making sense of disability not as something localized in the body of a single individual but as something common to, and experienced by, a group of people in society. Second, to varying degrees they challenge the centrality of medicine in making sense of the meaning of disability. Third, to varying degrees, these conceptions emerge from activists and scholars within the disabled population. It is useful to comment on three of these conceptual approaches here.

Minority Group Model

The first of these conceptual approaches is a 'minority group' model of disability. Here, disabled persons are conceived collectively as an oppressed group within society. This conception does not necessarily question or critique the underlying assumptions about the nature of disability as defined in the medical model. Rather, it uses the experience of disability as the basis for a collective political identity, arguing that recognition of the rights of disabled people is essential for their fair treatment in a society where the able body is a privileged norm. An important variation of the minority group model is the development of 'Deaf culture'. Proponents reject disability as an identity, arguing instead that deaf people constitute a minority with their own language (sign language), culture, and traditions, and should be recognized as such by the broader society. Proponents call for the development of separate institutions to sustain and enrich the traditions of this cultural minority.

Social Model

A second model, and arguably the most influential in recent decades, has been the 'social model' of disability with origins in both the Disability Movement of the 1970s and in scholarship by writers such as Vic Finkelstein and Michael Oliver. Unlike the minority model, the social model explicitly challenges the fact of functional limitation as the basis of disability by distinguishing conceptually between impairment and disability. The following definitions, from the British Council of Organizations of Disabled People, make this clear:

> Impairment is the functional limitation within the individual caused by physical, mental or sensory impairment. Disability is the loss or limitation of opportunities to take part in the normal life of the community on an equal level with others due to physical and social barriers. (Barnes, 1991)

This approach separates the impairment experienced by an individual from 'problems' the individual may confront in society. To take the example of paid employment, a medicalized conception of disability would explain high rates of unemployment and low levels of labor market participation in terms of the limited functional capacities of individuals. By contrast, the social model posits that exclusion from paid employment is principally a function of barriers in the social environment (e.g., physical barriers at work and in the public transit system, marginalization of disabled people in the educational system, the negative attitudes of employers, and so on). Within the social model tradition, Finkelstein and Oliver linked the oppression of disabled persons in Western societies to the rise of industrial capitalism. Oliver paid particular attention to the development of the factory system that demanded a specific nonimpaired norm, the introduction of individual wage labor, and the rise of medicine and associated professions to deal with 'problem' groups excluded from the system of production. The association of 'independence' with participation in paid work and the resulting stigmatization of dependency held fundamental implications for disabled people.

Cultural Model

A third model emerged partly in response to the social model. Informed by post structural and feminist theories, proponents of what might be broadly termed a 'cultural model' of disability took the issue with the materialist emphasis of social model scholars. The social model had little to say about discourse or cultural representations, and the crucial role that such representations might play in the construction of people as disabled or chronically ill. Take, for example, the social model's opposition of impairment and disability outlined above. While politically potent in its capacity to identify the 'disabling' character of the social environment, this formulation leaves the category of impairment as a largely unquestioned and medicalized fact that always already exists beyond its social interpretation. As Shelley Tremain has argued, the social model fails to question how 'people with impairments' are produced as subjects by specific forms of power/knowledge. By contrast, a Foucauldian analysis directs attention to the practices through which 'impairment' comes to stand as a prediscursive and supposedly natural category, and the ways in which this category is itself productive of governmental practices.

An equally important contribution of this approach has been to insist on the importance of the body to conceptions of disability and chronic illness. Social model scholars challenged medicine's preoccupation with the pathologies of the individual body but, in so doing, they risked removing the body from their theoretical formulations. This move was critiqued for several reasons. While the distinction worked well for some conditions (e.g., people with paraplegia using wheelchairs who maintained good health following their injury), it was much less appropriate for others. Scholars discussing chronic physical illnesses argued that the social model erased the realities of pain and fatigue that played a central role in shaping daily geographies. In response, writers have argued for a return to the body, not just as an object of medical analysis but as a site of cultural inscription and contestation. For this reason, qualitative accounts of life with disability and chronic illness have come to the fore, both to validate individuals' embodied experience and as a means to challenge dominant narratives of disability as personal tragedy.

Collectively, efforts to move beyond an individualized medical conception of disability have had important implications beyond the academy. In 2002, the World Health Organization replaced its International Classification of Impairment, Disease and Handicap with the International Classification of Functioning, Disability and Health, making explicit reference to the social model in recognizing the capacity of the environment to disable and/or enable people. Over the past decade and a half, many Western countries have enacted comprehensive disability rights legislation calling, at least on paper, for the removal of barriers to full participation in social life for persons living with a multiplicity of impairments.

Disability/Chronic Illness Research in Human Geography

Several geographic approaches to research on disability and chronic illness can be discerned. To some extent, these approaches both reflect and contribute to the broader developments in social science scholarship on disability and chronic illness outlined above.

The earliest geographical engagements with disability were positivistic in approach and utilized quantitative spatial analysis methods. Quantitative approaches provided objective descriptions of the spatial distribution of impairment and chronic illness in society, often in relation to existing patterns of service provision. Studies were primarily concerned with mapping the locations, movement, and demographic attributes of persons categorized as physically or intellectually impaired, or diagnosed with chronic illness. These studies relied on prevailing biomedical categorizations of impairment and disease. Today, quantitative approaches to disability and chronic illness are less common, leaving a gap in knowledge with regard to current locational characteristics of impairment. Exceptions include Robert Gant's work on the mobility and accessibility needs of disabled people in public spaces, as well as research by Howard Foster on the spatial ecology of illnesses such as Multiple Sclerosis and Alzheimer's.

Another way that geographers have engaged with disability is by using behavioral approaches to identify the 'internal' processes that persons with impairments use in the course of everyday interactions with their environments. Similar to quantitative approaches, behavioral approaches adopt an individualized conception of disability as a form of personal constraint. A majority of behavioral research has involved blind and vision impaired persons, and has taken a more practical focus emphasizing cognitive mapping and spatial mobility. Through exploring the spatial abilities of individuals with vision impairments, behavioral geographers have been at the forefront of developing an array of assistive technologies for accessing geographical information and negotiating the built environment. Work by Reginald Golledge has been foundational for the development of technologies like tactile maps, electronic 'haptic' tools, and GPS navigation devices, all of which have been used to assist vision impaired persons to travel independently. Similarly, Mark Blades and colleagues have explored the production and use of tactile maps by children and adults, demonstrating that vision impaired people are able to develop cognitive maps of an unknown area and navigate routes between places by using hard copy maps that are interpretable through touch.

A third approach encompasses geographic scholarship engaging with the social model of disability. Broadly materialist in orientation, this scholarship has sought to elucidate the spatial dimension of disability oppression, with particular attention to the process of 'disablement' specific to the capitalist mode of production. A key conceptual contribution to this literature has been provided by Gleeson's *Geographies of Disability*, which advances an analytical framework based on historical–geographical materialism. Gleeson uses empirical data to trace how evolving processes of labor commodification and urbanization in modern capitalist societies produced exclusionary physical and social environments that devalued workers with impairments and facilitated the marginalization of impaired people. Gleeson's work illustrates how standardized forms of industrial production in the space of the factory were organized around, and presupposed the normality of, the able bodied worker, producing a 'disabling division of labor'. This division of labor has had an enduring influence on labor market experiences of people with impairments and chronic illnesses. For example, the growth of flexible work practices in postindustrial economies has been suggested as a possible source of employment opportunity for people who are unable to fit a standard nine to five workday. However, recent research by Robert Wilton has shown that flexible work environments continue to be organized around a nondisabled norm. In these contexts, the benefits of flexibility accrue primarily to employers.

Focusing on disablement and the urban environment, Rob Imrie has offered an analysis of the socioinstitutional and political practices that perpetuate what he calls 'ableist' spaces: material environments that propagate powerful physical and social barriers to participation in 'normal' everyday activities. In particular, Imrie has shown how embedded socioinstitutional practices involved in the production of the built environment systematically overlook bodily differences and thus deny accessibility and mobility to those with impairments, a state he calls 'architectural apartheid'. Inaccessible public and private spaces are produced through the intersection of economic forces, architectural design traditions, and urban planning practices, resulting in the exclusion of disabled people from 'mainstream' community activities. Broken and uneven streets and sidewalks, building architecture with inaccessible stairs or doorways, and housing designs that presume an able body combine to produce an exclusionary physical environment, restricting the ability of disabled people to carry out daily activities such as shopping or socializing with friends. Rob Kitchin and Robin Law offer a striking example of the exclusionary nature of urban space in their study of access to public toilets. They found that poor design and lack of provision constrained the spatial behavior of people with physical impairments. These everyday experiences with essential services speak to the spatialized and ableist 'power geometries' that reproduce the exclusion of disabled people within social spaces.

Alongside geographers' interest in spatializing the social model, there has also been recognition that other theoretical perspectives are essential for understanding the complex geographies of disability and chronic illness. In recent years, strands of post structuralism and feminist theory have been the most influential. Geographers, for example, have used post structural theory to consider disability in terms of the sociospatial production of 'difference'. Work has examined the power of discourse to divide or partition 'normal', 'healthy', and 'able' bodies from 'abnormal', 'unhealthy', and 'disabled' bodies. From this perspective, discursive processes of boundary making shape social reactions to impaired persons, contributing to landscapes of (in)tolerance and exclusion, a process Mike Dorn has called 'politico moral territorialization'. In this way, social space is disabling not only in a physical sense, but has also come to symbolize the deviation of impaired bodies from able bodied norms. Rob Kitchin, for example, has argued that unaccommodating built environments are social texts which inscribe certain bodies as 'out of place'. Such environments reflect and reproduce the privilege of an able bodied norm, while at the same time perpetuating the 'Otherness' of disabled bodies.

Elsewhere, Robert Wilton has used psychoanalytic theory to argue that the construction of impaired bodies as 'Other' triggers psychosocial anxieties that in turn motivate efforts to exclude visibly disabled people from social space. This work offers a cultural complement to earlier studies that emphasized the material bases of exclusion (e.g., that NIMBY reactions to housing and services for disabled persons are sourced in concerns about property values). In a context where the fact of human mortality is disavowed and where forms of 'body maintenance' are readily available to those able to pay, visible impairment challenges taken for granted assumptions about our ability to control the body. Also in the post structural tradition, Louise Holt has used the work of Judith Butler to explore the ways in which learning (dis)ability among children is constituted through performance in school classrooms. She draws attention to the ways that 'disability' can be enacted in quite different ways within the context of different classrooms, reflecting the discourses and spatial practices adopted by teachers and students.

Others geographers have used a feminist materialist perspective to examine the spatiality of chronic illness and impairment in the everyday lives of women. The work of Isabel Dyck and Pamela Moss is well known here. Focusing on sites such as the workplace and the home, Dyck and Moss have investigated the intersection of the material and the discursive in the formation of identity. Their work has carried forward a concern with notions of difference, by examining the ways in which dominant biomedical discourses are applied to distinguish between healthy and unhealthy bodies. These inscriptions are contested by women as they renegotiate place specific identities and corporeal changes associated with illness. For example, Dyck has shown how, as women experience chronic illness, the workplace can become a place of risk where taken for granted identities are fragmented and destabilized. A key issue for many women is disclosure, a moment of intense risk and anxiety. With disclosure, the body is 'marked' as disabled, an identity that is devalued within the workplace. As mentioned earlier, the workplace is normatively understood to be a place for 'healthy', able bodied workers. Thus, it is often the case that women manage disability, and issues around disclosure, through 'concealment strategies', that minimize symptoms and visibility. A major accomplishment in the work of Dyck, Moss, and others using feminist materialism has been to demonstrate that the body is simultaneously a material and discursive product.

Ethics and Empowerment in Disability Research

As geographers have engaged in research on the geographies of disability and chronic illness in recent years, they have also had to confront difficult questions about the ethics and politics of this academic inquiry. As part of asserting the rights of disabled people to speak for themselves, disability movements have come to regard academic experts (especially nondisabled academics) with some suspicion. Activists have questioned the motivations of academics, pointing to the ways in which existing scholarship has been used to classify disabled people and, in some instances, to legitimate their marginalization. Expressions like 'Nothing about us without us' signaled resistance to more traditional research relationships where nondisabled academics decide what is an appropriate topic, and retain the authority to speak for study populations. For some activists, research by nondisabled academics has little or no place in the political project of the disability movement. Others argue not for a complete exclusion of researchers, but for a restructuring of the relationship between researchers and participants.

A key development in this vein has been the call for participatory or 'emancipatory' research, where people with disabilities and chronic illnesses figure not just as respondents but as active stakeholders with influence over the research questions, the kinds of knowledge produced, and the uses to which study results are put. Geographers such as Brendan Gleeson and Vera Chouinard have advocated for the adoption of this approach, with explicit attention to the ways that research does or does not help to challenge forms of oppression faced by disabled people. In the late 1990s, Rob Kitchin outlined two approaches to emancipatory research. On the one hand, projects can remain guided by a professional model of research, where an academic is responsible for conducting the research but where participants monitor the process of research and offer feedback and criticism throughout. On the other hand, there can be an explicit effort to move toward a participatory model where people are involved in the entire process of research.

While these approaches have the capacity to challenge inequalities characteristic of more traditional research, there is a danger of romanticizing participatory projects. For example, a move toward increased participation assumes that people have the time and the energy to be involved in the entire research process. Finding innovative ways to both train and remunerate disabled people as research workers may help to avoid these pitfalls, and foster greater inclusion.

At the same time, scholars have cautioned that a participatory approach may effectively exclude some groups. Ed Hall and Robin Kearns argue that for people with learning (intellectual) impairments, a lack of education, social isolation, control by caregivers, and as well as the individual's intellectual capacities may make participatory research difficult. This draws attention to the need for a range of innovative approaches to research that maximizes opportunities for meaningful input from people while recognizing both embodied constraints and social barriers to participation. In practice, much recent geographical research has not been fully participatory in nature. However, recent debates about power relations within research have meant that geographers are now more willing to question how their studies address the interests and opinions of disabled persons in the design, implementation, and outcomes of studies.

Beyond the design of research projects, questions have also been asked about the extent to which the academy itself constitutes an inclusive environment for people with disabilities and chronic illnesses. While physical access on college and university campuses has improved, progress in other areas is mixed. Blind and visually impaired students, for example, continue to face challenges in accessing alternative format texts, while students and faculty with impairments confront enduring prejudices in their search for accommodation.

Future Directions

Recent years have seen a significant growth of work on, and interest in, the economic, political, social, and cultural geographies of disability and chronic illness. On the one hand, this is cause for optimism. Geographers have produced a wide, perhaps somewhat disparate, range of research on disability and chronic illness, forging connections with a larger population of disability scholars in the social sciences and humanities on the one hand, and with communities of disabled and chronically ill persons on the other. Despite this progress there is still cause for concern. For one thing, it is not yet clear whether disability and chronic illness will be taken seriously in the discipline and the curriculum beyond the confines of medical/health geography. Notwithstanding the work of the geographers discussed above, disability – as a dimension of oppression and a subject for inquiry – does not have the status of race, class, gender, and sexuality in the broader subfields of social, cultural, urban, economic, or political geography. This is despite the fact that disabled persons constitute a significant – and growing – minority of the population in Western countries and they continue to confront and struggle against manifold barriers to full participation in, and access to, housing, paid employment, political participation, public space, social lives, and cultural representation.

At the same time, existing studies have painted only a partial picture of the geographies of disability. The majority of work is from the developed world, while impairment, disability, and chronic illness are global realities. More work is needed to examine the complexities of these experiences in non Western contexts. For example, globalization has shifted economic production to poorer nations, with implications for the social and economic geographies of disabled and nondisabled persons alike. At the same time, structural adjustment programs often require developing nations to cut or forestall social spending to receive international loan assistance. Such cuts have implications for the incidence of impairment as well as the ability of disabled persons to access even basic services in these contexts. Finally, many questions surround the applicability of theories such as the social model of disability beyond the Western in dustrialized contexts in which they originated.

See also: Body, The; Ethical Issues in Research; Feminism/Feminist Geography; Golledge, R.; Health Geography; Historical-Geographical Materialism; Medical Geography; Mental Health; Participatory Action Research; Psychoanalysis; Social Geography; Social Justice, Urban.

Further Reading

Barnes, L. (1991). *Disabled People in Britain ande Discrimination: A Case for Anti discrimination Legislation*. London: Hurst, in association with the British Council of Disabled People.
Blades, M., Ungar, S. and Spencer, C. (1999). Map use by adults with visual impairments. *Professional Geographer* 51(4), 539 553.
Butler, R. and Parr, H. (1998). *Mind and Body Spaces: Geographies of Illness, Impairment and Disability*. New York: Routledge.
Chouinard, V. (1997). Making space for disabling differences: Challenging ableist geographies. *Environment and Planning D: Society and Space* 15(4), 379 387.
Dyck, I. and Moss, P. (2002). *Women, Body, Illness: Space and Identity in the Everyday Lives of Women with Chronic Illness*. New York: Rowman & Littlefield.
Gleeson, B. (1999). *Geographies of Disability*. London: Routledge.
Golledge, R. (1993). Geography and the disabled: A survey with special reference to vision impaired and blind populations. *Transactions, Institute of British Geographers* 21(2), 63 85.
Hall, E. and Kearns, R. (2001). Making space for the 'intellectual' in geographies of disability. *Health & Place* 7, 237 246.
Holt, L. (2005). Children with mind body differences: Performing disability in primary school classrooms. *Children's Geographies* 2(2), 219 236.
Imrie, R. (1996). *Disability and the City: International Perspectives*. London: Paul Chapman.
Kitchin, R. (1998). 'Out of place,' 'knowing one's place': Space, power and the exclusion of disabled people. *Disability and Society* 13(3), 343 356.
Kitchin, R. and Law, R. (2001). The socio spatial construction of (in)accessible public toilets. *Urban Studies* 38(2), 287 298.
Oliver, M. (1990). *The Politics of Disablement*. London: Macmillan.
Tremain, S. (ed.) (1998). *Foucault and the Government of Disability*. Ann Arbor, MI: University of Michigan Press.
Wilton, R. (2004). From flexibility to accommodation? Disabled people and the reinvention of paid work. *Transactions, Institute of British Geographers* 29(4), 420 432.
Young, W. (2002). Spinal cord injury levels and classification . Sci Info OPages http://www.sci info Pages, http://www.sci info pages. com/levels. html (accessed 11 September 2008).

Relevant Websites

http://www.adapt.org
ADAPT (US disability rights organization).
http://www.disabilitystudies.ca
Canadian on Disability Studies.
http://isc.temple.edu
Disability and Geography International Network (DAGIN), Temple University.
http://www.dawncanada.net
Disabled Women's Network (DAWN) Canada.
http://www.rehab international.org
Rehabilitation International.
http://soundscapes.geog.ucsb.edu
Research Unit on Spatial Cognition and Choice (University of California Santa Barbara).
http://www.bcodp.org.uk
United Kingdom Disabled People's Council (formerly the British Council of Disabled People).
http://www.who.int
World Health Organization's 'International Classification of Functioning, Disability & Health.
http://www.wid.org
World Institute on Disability.

Discourse

T. Cresswell, Royal Holloway, University of London, Egham, UK

Glossary

Epistemological Concerning how we know the world.
Ideology Meanings that tend to support existing relations of domination.
Representation A likeness or image of something that can be textual, visual, or even performed. In a representation something is 're-presented' or presented again.

Defining Discourse

Discourse is a term used with increasing frequency since the late 1980s in social, cultural, and political geography. Its use, however, is far from consistent. Often it is used to vaguely refer to spoken and written language while at other times it includes all forms of representation and refers quite specifically to theories derived from the work of post structuralism in general and the work of Foucault in particular. Foucault's theory of discourse approaches epistemological questions of how we know what we know, where such knowledge comes from, who is authorized to profess it and how, ultimately, 'truth' is established.

In everyday speech discourse usually refers to either spoken or written communication. More specifically it is often used to refer to an organized spoken or written text about a particular subject. An example of this would be Jean Jacques Rousseau's *Discourse on political economy* (1755). These general meanings for the word discourse are often used in geographical literature to loosely refer to a more or less organized set of words about something – 'the discourse of devolution' or 'rural discourse' perhaps. These uses, however, are frequently confused with a more precise theoretical use of the word discourse that originates, predominantly, from the work of Michel Foucault. In his theoretical and substantive work Foucault sought to show how linguistic and other elements are combined to shape and delineate particular kinds of objects of knowledge (madness, illness, the pathological, etc.), new sets of concepts and categories, new claims about what can be said and what must remain unsaid, and a new set of subject positions (the mad, the criminal, the deviant as well as the expert, the psychologist, etc.). In this sense discourse is no longer simply about words (written or spoken) but about a whole set of words, actions, institutions, and infrastructures that are more or less logically coherent and produce new regimes of truth. Foucault's use of the term discourse was far from consistent. He admitted that sometimes he used it to refer generally to all statements, often quite small and individual in conception. At other times he used it to refer to an overarching and organizing set of regulating principles that helps to account for the production of a multitude of statements at a smaller scale. He also used it to refer to statements in between these two poles. An example of the later would be the discourse of perversion or the discourse of criminality.

Foucault's most famous book (at least to geographers) is *Discipline and punish*. In this book he shows how the criminally deviant individual was transformed in the late eighteenth century through the growth of a disciplinary society which had at its center a set of legal and penal discourses which produced a new kind of criminal subject. This new kind of person had its own sets of practices, heredity, medical diagnoses, and other characteristics. New forms of punishment and surveillance arose hand in hand with these new subject positions as did new kinds of practice (in science, in law, in politics, etc.). These discourses were not simply sets of spoken or written words (as suggested by general definitions of the meaning of discourse) but constellations of words, practices, institutions (such as the prison), and things. Neither are they clearly 'about something'. Whereas common uses of the word discourse assume that words refer to something that preexists the words, the more precise formulation originating in the work of Foucault suggests rather the opposite – that new 'truths', new possibilities of being, emerge out of particular discursive formations. Discourses can thus be seen as establishing new networks of meaning and practice which delineate, produce, and reinforce relations between what it is possible to think, say, and do. Objects and subjects are not external to this process but are thoroughly constituted through it.

The more technical and Foucauldian use of the term discourse thus has two primary differences from its more everyday use. First, discourse is considerably more than spoken and written words (it includes all forms of representation and, in some readings, institutions, objects, and practices) and, second, discourse is not simply 'about' something (politics, illness, sexuality, etc.) but brings these things into being. In the sense that discourses produce reality they are performative. Discourse does not simply describe something that preexists it, rather it brings it into being. In this sense it is more than a reflection of 'reality' that can be true or false but is in fact performing 'reality' and producing 'truth effects'.

One way to specify the Foucauldian meaning of dis course is to compare it to the concept of ideology. Dis course is, in some ways, close to the term 'ideology'. There are, however, important differences. Ideology, like discourse, considers the relations between meaning, power, and truth. Most theorists of ideology define it as sets of meanings in the service of the relatively powerful – meanings which reinforce already existing power relations. Discourse differs from this in important ways. While ideology relies on some notion of truth and objectivity ('real interests', for instance), that is external to it, discourse produces truth. Foucauldian discourse analysis is not concerned with separating out the truthful and the objective from the rest but with showing how historically 'effects of truth' are produced discursively. In addition, ideology (at least in its Marxist formulations) is always part of a superstructure which is caused by some other realm of human interaction – usually the world of political economy. Discourse has no such 'base' to appeal to or that can be used to explain it. Truth, in other words, is not in some other place but is within discourse itself. Finally the idea of ideology rests on the notion of a knowing and intentional subject – it is fundamentally humanist in its conception. Foucault's notion of disourse rejects humanism and, in particular, the humanist notion of the subject at the center of the universe. The subject in Foucault is not unitary and knowing but fragmented and decentered – produced discursively through the process of subjectivization. An example would be the ways vari ous medical texts, beliefs, and practices, as well as the construction of the physical environment, produce the 'disabled' subject who is both marginalized and endowed with particular forms of competence.

Discourse analysis is thus very different from the analysis of ideology. While the analysis of ideology (and 'critical theory' in general) seeks to peel away obfuscation to reveal the 'truth' of a situation, the analysis of dis course tries to define not what is concealed by a discourse but what is produced by it – the rules of the discourse itself and the new regimes of truth that it enables.

Discourse and Sexuality

Perhaps the clearest delineation of discourse provided by Foucault is in the introduction to his *History of sexuality*. Here he describes a predominant approach to sexuality he calls the repressive hypothesis. According to this, sexuality has been repressed since the seventeenth cen tury when sexual activity was free of hang ups. Between the seventeenth century and Victorian times, the hy pothesis goes, sexuality was gradually repressed, made invisible, and submerged. It became embarrassing. This repression coincided more or less with the advent of industrial capitalism and the rise of a bourgeois order. In fact capitalism and its attendant work imperative became an explanation for the repression of sexuality. This, however, is not Foucault's approach. Rather than looking for economic explanations he turns to discourse. In this time of supposed repression, he points out, there has been a surfeit of discourse on sex. The question for Foucault is not why are we repressed but why do we say (so often and so loudly) that we are repressed?

> Why has sexuality been so widely discussed, and what has been said about it? What were the effects of power generated by what was said? What are the links between these discourses, these effects of power, and the pleasures that were invested by them? What knowledge (savoir) was formed as a result of this linkage? . . . The central issue then (at least in the first instance), is not to determine whether one says yes or no to sex, whether one formu lates prohibitions or permissions, whether one asserts its importance or denies its effects, or whether one refines the words one uses to designate it; but to account for the fact that it is spoken about, to discover who does the speaking, the positions and viewpoints from which they speak, the institutions which prompt people to speak about it and which store and distribute the things that are said. What is at issue, briefly, is the over all 'discursive fact', the way in which sex is 'put into discourse'. (Foucault, 1990: 11)

So far from being repressed, Foucault suggests, sex and sexuality is being discursively produced through an excess of things said. The point, he underlines, is not to find the truth behind the discourse (as an ideological analysis might do) but to find the truth produced within the discourse. Power here is not repressive but productive.

The Example of the Tramp

Take the example of the tramp in the United States in the period following the Civil War. A careful observer would have noticed increasingly frequent references to tramps in the pages of the New York Times from the 1870s onward. In fact they would not have seen any references to tramps before February 1875. It would have been as though they did not exist. A year later the first 'tramp law' came into being in New Jersey. A picture was beginning to emerge both in the press and in the defin itions provided by law of what a tramp was. Apparently idle, without work, and frighteningly mobile – able to be in New York one day and in California two weeks later thanks to the recently completed transcontinental rail road. In the early 1890s the New England reformer, John McCook, was writing sympathetic columns about tramp life which resulted from extensive questionnaires and interviews with tramps. By 1899 all kinds of books were appearing either about or by these tramps. Indeed a new academic discipline, sociology, was dedicating a large amount of time to recording the lives of these people and further classifying them into different kinds of tramps, as well as distinguishing them from hobos and bums. One such 'sociologist', Josiah Flynt, spent a year dressed as a tramp and traveling with them. He reported his adven tures in the popular magazine *Atlantic Monthly*. All kinds of theories about these disorganized people began to emerge from the new Chicago School of Sociology. In the early twentieth century the sociologists were joined by the medical profession and, particularly, by eugenicists, who suggested that the tramps suffered from an inherited predisposition to wander endlessly.

Tramps also became central to other new forms of knowledge. Cartoonists frequently featured tramps in cartoon strips in the daily paper as figures of fun. Vaudeville performers made fun of the tramp's hygiene and work aversion. The new silent movies were to put the tramp at center stage as Charlie Chaplin used his 'little tramp' to poke fun at conventional society. The practice of documentary photography was producing picturesque portraits of tramps in New York's back alleys. Given that our careful observer would never have seen the use of the word tramp as a noun until 1875 it is surely remarkable that such an array of words, images, charts, movies, diagnoses, treatment, and punishments should have been brought to bear on this previously invisible and un named character. It should be noted that sociology, newspaper cartoons, documentary photography, vaude ville, silent movies, and eugenics were also new to this period. It was as if they had arisen hand in hand with the tramp.

There are clearly at least two ways to think about this. The conventional way would be to say that a new phe nomenon has arisen – a new thing in the world – the tramp – and that, once noticed, knowledge had to ac count for it. The second way would suggest that with the construction of new discourses (sociology of deviance, vaudeville humor, etc.) it became possible to be a tramp. In this second sense the tramp was 'made up' discursively. Importantly this is not simply to say that new ideas and images of the tramp were produced culturally (this would hardly be controversial) but that tramps were literally brought into being – it became possible to be a tramp. And with the arrival of the tramp came new possibilities for practice, new punishments, new medical diagnoses, new spaces and places.

Geographies of Discourse

So what is geographical about discourse? There are four principal ways in which geographers either have or po tentially could engage with discourse.

First, the idea of discourse places a great deal of importance on context. Knowledge, texts, truths, prac tices, and realities are all products of particular times and spaces. They are not universal but contextual. Clearly, then, discourses have a distinctive spatial character. The work of Foucault, for instance, is often grounded in particular times and places and the knowledge that arises there is specific to that context. His studies of madness and discipline, for instance, were clearly set in France and the setting impacted upon the discourse. Discursive knowledge is not universal but particular. Its truths are not general but specific.

Second, and more important, discourses have very specific sites where they arise. Laboratories, clinics, uni versities, and asylums are all micro geographies which are both the context for and part of discourse. Most clearly, perhaps, modern discipline is unthinkable without the modern prison as famously represented by Bentham's panopticon. However, all kinds of discourse require au thority for someone to be able to use it and produce truth effects. Part of this authority comes from the place from which discourse was produced by the author.

Third, geographical knowledge itself can be subjected to discourse analysis. Geography departments and soci eties are sites from which discourse is produced, as well as material parts of discourses themselves. Somewhere like the Royal Geographical Society, for instance, is a site where geographical discourse constituted particular kinds of colonial and imperial subjects and produced particular kinds of colonial and imperial truths. Geography has its own discourses and some of them have been historically powerful as, for instance, Felix Driver has shown.

Fourth, discourse is implicated in the production of places and, in particular, in the judgment of people's practices within places. What counts as acceptable, ap propriate behavior, for instance, if often determined by a nexus of place and discourse. Subjects are not simply constituted anywhere but on a particular terrain. The discourse of urban planning, for instance, is implicated in the production of the 'ghetto'. The ghetto then histor ically became a site associated with particular people (often racially coded) and particular kinds of practice.

These different aspects of geographies of discourse have, in one form or another, circulated through geo graphy's subdisciplines since the mid 1980s. Cultural geographers, with their interests in textuality and rep resentation, began to invoke the notion of discourse quite loosely alongside literary concepts such as metaphor and rhetoric. The general 'cultural turn' in human geography saw the migration of discourse analysis, and particularly the Foucauldian version of it, into other subdisciplines such as political geography and economic geography. Critical geopolitics, for instance, often accounts for pol itical power on the international stage through analysis of a range of discourses from the popular media to official governmental policy. Rather than considering geopolitics as an objective form of knowledge about the world, critical geopolitics considers them as discursive con structions of the political world constructed from par ticular places by particular people with the authority to speak. Critical geopolitics thus considers all manner of representational forms from official documents to news paper cartoons as elements in the formation of the pol itical landscape.

The historical geographer, Matthew Hannah, con ducted a particularly intensive discursive analysis of the development of the census in the United States alongside a new statistical discourse which brought the new cat egory of the 'population' into being. This analysis involved an understanding of the places from which statistics emerge, as well as how particular people are given the authority to speak the discourse of statistics about popu lations. Census statistics, it turns out, are not simply about populations – they create them. Once populations are invented all kinds of statistical practice (and other kinds of practice) can be exercised on and through these popu lations. All of this is intensely geographical.

Lines of Critique

The Foucauldian notion of discourse is not without its critics. A number of important critiques have been made of discourse and discursive analysis, both by left wing commentators (mainly upset by the challenge to theories of ideology for instance), and by right wing critics who despair at what they see as the relativism and antirealism of Foucault's thought. An important ontological question, for instance, is 'what lies beyond discourse?' A strong reading of Foucault would seem to suggest that nothing lies beyond discourse. A sceptic might point to the hardness of rocks or the laws of gravity and say that such things are clearly not created discursively. Indeed, a supporter of Foucault, Ian Hacking, has gone to great pains to suggest that there is something importantly different about a 'tramp', for instance, and a 'waterfall'. While it is true that a waterfall would exist with or without our apprehending it, a tramp is literally made up through our knowledge about it and our representations of it. However, even a waterfall has to be apprehended and, a Foucauldian might argue, our understanding of waterfalls only happens through discourse. So here we have a weaker discourse theory that acknowledges things beyond discourse but says there is no way to know such things except discursively.

Among the more thought provoking critiques of dis course theory is the observation that if objects are internal to discourses (i.e., are constituted by them) then how are we to judge that discourse has constructed its object ap propriately, validly, or justly? Or to put it more straightforwardly, how is it possible to be right or wrong? This question is both an ontological one and a moral–political one. Unlike the theory of ideology, for instance, there does not seem to be a place from which to judge a discourse and the truths it constructs. How, for instance, could we say that Nazi discourse, or racist discourse, or homophobic discourse is wrong? On what grounds could we seek to change it? Related to this is the suggestion that Foucault's discursive approach leaves little room for re sistance to domination except through yet more discourse.

Finally theories of discourse are subject to the general critique of representation articulated by nonrepresenta tional theory. Nonrepresentational theory suggests that a focus on representation (and discourse is heavily repre sentational) ignores the world of emotion and affect that forms a central part of existence (human and otherwise). The suggestion that there is nothing (that can be ap prehended) beyond discourse suppresses and marginal izes these important aspects of life on Earth.

See also: Foucauldianism; Ideology; Non-Representational Theory/Non-Representational Geographies; Performativity; Representation and Re-presentation; Sexuality.

Further Reading

Barnes, T. and Duncan, J. (eds.) (1991). *Writing worlds: Discourse, text and metaphor in the representation of landscape*. London: Routledge.
Cresswell, T. (2001). *The tramp in America*. London: Reaktion.
Dalby, S. (1991). Critical geopolitics: Discourse, difference, and dissent. *Environment and Planning D: Society and Space* 9(3), 261 283.
Driver, F. (1992). Geography's empire: Histories of geographical knowledge. *Environment and Planning D: Society and Space* 10, 23 40.
Foucault, M. (1972). *The archaeology of knowledge*. London: Routledge.
Foucault, M. (1981). The order of discourse. In Young, R. (ed.) *Untying the text: A post structuralist reader*, pp 48 79. London: Routledge.
Foucault, M. (1990). *The history of sexuality, vol. 1: An introduction*. New York: Vintage Books.
Foucault, M. (1991). *Discipline and punish: The birth of the prison*. Harmondsworth: Penguin.
Hacking, I. (2002). *Historical ontology*. Cambridge, MA: Harvard University Press.
Hannah, M. (2000). *Governmentality and the mastery of territory in nineteenth century America*. New York: Cambridge University Press.
Mele, C. (2000). The materiality of urban discourse: Rational planning in the restructuring of the early twentieth century ghetto. *Urban Affairs Review* 35(5), 628 648.
Mills, S. (1997). *Discourse*. London: Routledge.

Relevant Websites

http://extra.shu.ac.uk/daol/about
Discourse Analysis Online.
www.eng.fju.edu.tw/crit.97/Foucault/Foucault.htm#power
Discourse, Power, Subjectivity.

Discourse Analysis

L. D. Berg, University of British Columbia, Kelowna, BC, Canada

Glossary

Discourse This term is often confusing because it is used in two ways. Linguists tend to use it for particular passages of speech and writing. Drawing on the work of Michel Foucault, social theorists use it in a broader fashion to refer to the taken-for-granted, and most often, hidden, frameworks of ideas that structure both knowledge and social practice. Discourses are thus sets of ideas 'and' practices that give statements, texts, rhetorics, and narratives particular kinds of meanings. Somewhat enigmatically, Foucault argued that "discourses are practices that systematically form the objects of which they speak." Many complex and often contradictory discourses can be in circulation at any one time, but one discourse is usually hegemonic, and thus it tends to constitute the general conditions under which dominant members of a society 'know' their world.

Episteme The historical formations of knowledge that sets the conditions of possibility for how we understand our social life. An episteme can be seen to share many characteristics with 'zeitgeist', in that it provides a structure of feeling about a specific era or epoch. Similarly, and as the term suggests, it shares certain characteristics with the term epistemology, in that it relates closely to how we come to understand things. However, both zeitgeist and epistemology tend to ignore the power relations that structurally contain and enable such processes, whereas the idea of the episteme is designed to directly confront the circulation of power in societies, and the 'productive' character of such power (i.e., its ability to create both subjects and objects of power).

Power/Knowledge In order to better understand the relationship between power and systems of knowledge, Michel Foucault coined the term power/knowledge. He argued that power and knowledge could not be separated analytically, as "there is no power relationship without the correlative constitution of a field of knowledge, nor any knowledge that does not presuppose and constitute at the same time, power relations." Importantly, Foucault does not see power as operating top-down, but, instead, he sees power as infusing all social relations and operating to produce subjects and objects (rather than operating between pre-given subjects and objects). Power/knowledge has become an important concept for geographers interested in understanding the complex relationship between knowledge, space, and power.

Subjectivity/Subject Position All of us work with some kind of theory of the subject, that is, what it is to be a sentient being and to live, act, and have relationships in the world. For many years, geographers drew on a relatively fixed notion of the subject exemplified by the term 'identity', and involving identification as, for example, a male, a female, a heterosexual, or a disabled person. Key to the concept of identity is attachment to the existence of some essential, natural, or central core qualities. More recently, post-structuralist and feminist geographers have drawn on the concepts of 'subjectivity' and 'subject positions' to understand what it is to 'be' a sentient being and engage in relations with other subjects and objects. Unlike the notions of relative fixity associated with identities, subjectivities are seen as always in-the-making, they are often contested, and in many cases they incorporate contradictory elements. Geographers have become key theorists of subjectivity because of the important role that space and place play in subject formation and thus subject 'positions' in social and spatial relations.

Introduction

In his inaugural lecture of 2 December 1970 at the Collège de France, – a unique institution of higher education in that its classes are both free and open to anyone who wishes to attend, and its professors are elected to their posts by an electorate of France's leading scholars – Michel Foucault set out to explain one of his most important concepts, namely, that of 'discourse':

> … in every society the production of discourse is at once controlled, selected, organized and redistributed by a certain number of procedures whose role is to ward off its powers and dangers, to gain mastery over its chance events, to evade its ponderous, formidable materiality. (Michel Foucault, 1984: 109)

Since that time, Foucault's notion of discourse has in formed much theorizing about the relationship between power and knowledge in the structuration of society and space. Critical geographers, particularly feminist, anti racist, and postcolonial geographers, have begun to focus on the role of language and discourse in the maintenance of political arrangements of domination and oppression. Much of this work is informed, implicitly if not explicitly, by a post structuralist sensibility. This sensibility is based on the argument that language does not mirror some preexisting reality, but instead language constitutes the conditions under which we 'know' reality. Indeed, much of the post structuralist argument concerning the status of truth centers on language because of "the realization/belief that language lies at the heart of all knowledge." As Simon Dalby suggests, this is a shift from positivist ap proaches and epistemological concerns that:

> … involves conceptualizing social existence as human practice. Social life is active creation, albeit within cre ated frameworks of custom, economy, power and lan guage. Social life is understood in and through language, and hence the structures of language reflect and create social life. (Dalby, 1990: 5)

The centrality of language in meaning production has meant that discourse analysis tended to focus on 'text' as an object of analysis. Following Roland Barthes, however, text is conceived broadly, and thus discourse analysis has focused on all sorts of social and cultural productions from printed words, paintings, maps, landscapes to social, economic, and political institutions. Such texts at once are created by, impose, and maintain particular discursive formations that involve specific epistemological claims, circumscribe legitimacy, and "provide the intellectual conditions of possibility of particular institutional and political arrangements."

This article attempts to link Foucauldian notions of discourse, with a more 'practical' guide to the practice of discourse analysis. The first half of the article begins with an introduction to Michel Foucault's conceptualization of discourse, then moves to a discussion of its relationship to power, continues with a discussion of some of what Foucault termed the 'rules' of discourse, and then con cludes with a brief introduction to Foucault's 'methods' of discourse analysis. The second part of the article at tempts to link Foucault's somewhat abstract approach to method to a more grounded understanding of 'doing' discourse analysis in human geography. It is important to note at the outset, however, that geographers have rarely explained in detail their discourse analytic method, so the article relies on a few key sources to elucidate some 'methods' of discourse analysis.

Foucauldian Conceptualizations of Discourse

Discourse as a Concept

Discourse is a difficult concept, primarily because there exist a large number of sometimes conflicting and often overlapping definitions formulated from various theo retical, methodological, and disciplinary standpoints. Nonetheless, two broad 'understandings' of discourse can be identified. Linguists tend to use discourse to refer to extended examples of either spoken or written language. On the other hand, while recognizing the centrality of language and text to discursive analysis, social theorists tend to view discourse in a much broader fashion. In much social theory, and particularly that based on the work of Michel Foucault, discourse refers to ways of structuring knowledge and social practice.

Following this latter approach, discourses can be characterized as "frameworks that embrace particular combinations of narratives, concepts, ideologies and sig nifying practices, each relevant to a particular realm of social action." Discourse is somewhat like grammar, in that it is not a thing to be seen and touched, but instead an implicit "set of capabilities, an ensemble of rules by which readers/listeners and speakers/audiences are able to take what they hear and read and construct it into an organized meaningful whole." Discourse is not merely reflective of dominant social values, but is also consti tutive: "different discourses constitute key entities (be they 'mental illness', 'citizenship' or 'literacy') in different ways, and position people in different ways as social subjects." Accordingly, discourse can be seen more as a set of unspoken rules which govern, control, and produce knowledge in a culture.

Discourse and Power

One key aspect to Foucault's notion of discursive power is that such power is not just 'repressive' in the sense that it prevents action or controls existing agents and outcomes, but that it is also 'productive' in that it creates new ac tions, events, agents, and outcomes. A number of feminist theorists have explored the productive role of power, and two metaphorical concepts are key here: the panopticon and the confessional. The panopticon is a metaphor for people's internalization of the perspective of a usually hegemonic Other that produces a self monitoring subject. Force is rarely required to ensure conformity, for the (subordinate) subject ensures its own conformity to dominant norms. For example, business people may conform to a certain 'norm' of dress and personal grooming, despite their own personal beliefs about such matters. The confessional is a metaphor for the "social urge to explain, justify and seek forgiveness for intimate feelings, decisions and actions." Played out in numerous settings – the church, the doctor's office, therapy sessions, the office – the confessional is "the liberational antithesis of power." In these ways, both knowledge and cultures are 'constructed' by the power inscribed in discursive practices. As social constructions, discursive formations have institutional origins and commitments; they are embodied in technical processes, institutions, and pat terns of general behavior. Thus, discourses are "both reinforced and renewed by whole strata of practices, such as pedagogy … and the system of books, publishing, libraries; learned societies in the past and laboratories now." Their knowledges are political products, bound up with power; they are thus both the product and medium of power.

A second important concept associated with discursive power is the way that its relation to the subject is conceptualized. In this regard, Foucault conceptualizes power as operating everywhere, and thus he does not see a need to theorize an agent of power. Perhaps more importantly, subjects are constituted in and through the power of discourse. This is distinctly different from other critical theories, such as Marxism, which tends to view power as operating top down, from a pre given subject (the ruling classes, who hold power) to another pre given subject (the working classes, who are relatively 'powerless'). This is an important theoretical move that helps to understand power when it is at its most effective – when power is exercised by everyone, yet by no one in particular. Power works best when it is anonymous, and in such cases, it would be misguided attempting to identify a specific agent of power. Instead, discourse analysts are much more interested in coming to understand not only which social groups benefit from discursive power, but also, how group and individual identities (subjectivities) are constituted in power relations.

Rules of Discourse

Foucault suggests that discourse may be identified by the existence of certain 'rules' which he calls 'practices'. Such discursive practices systematically form for the discourse "groups of objects, enunciations, concepts, or theoretical choices." For example, as Foucault suggests with reference to medicinal practices, such rules govern modes of speaking or 'enunciations' by defining the answers to questions such as "who, among the totality of speaking individuals, is accorded the right to use this sort of language? Who is qualified to do so?" Similarly, they govern the institutional 'sites' from which the medical practitioner makes such enunciations, and from which this discourse derives its legitimate source and point of application.

Discourses or discursive formations are complex systems of signification governing the production and use of knowledge and meaning in cultures. Cultures themselves are established through the complex interdependent relationships formed between, in, and through discursive practices. Such practices are governed by rules that circumscribe the objects of discourse, the speakers of discourse, and the forms of discussion that may take place in particular discourses. Discourses are governed externally, internally, and through control of access. Discourses are governed 'externally' through prohibitions of speech, divisions between reason and madness, and oppositions of truth and falsehood. They are also 'internally' policed through the application of commentary and ritualized discussion. Finally, 'access' to discourse is controlled via speech rituals, societies of discourse, doctrine, and social appropriation. These 'rules' for discourse interlock in intricate ways that are constantly shifting and regrouping, such that discursive practices themselves are also constantly transforming and changing in complex ways that are not easily traceable because they cannot be characterized merely as a paradigm shift or change in *zeitgeist*, collective attitudes or state of mind.

> The transformation of a discursive practice is linked to a whole range of usually complex modifications that can occur outside of its domain (in the forms of production, in social relationships, in political institutions), inside it (in its techniques for determining its objects, in the adjustment and refinement of its concepts, in its accumulation of facts), or to the side of it (in other discursive practices). (Foucault, 1977: 200)

Nevertheless, it is possible to trace the effects of these discursive relationships, and it is to Foucault's 'method' for doing so that I now turn.

Foucault has utilized his concept of discourse to analyze, among other things, the discursive practices of psychiatry, medicine, penal institutions, and regimes of sexuality. He has illustrated the ways in which madness is created as the antithesis to reason, diseased to healthy, delinquent to reformed, and deviant to normal. In so doing, Foucault has illustrated the way that discursive practices delimit their objects of study (or surveillance, discipline, etc.) and then define the legitimate perspective for the knowing subject. Accordingly, he has focused his analysis on the way a dominant identity that is self defined as 'normal' is usually structured in opposition to some external Other defined (by members of the dominant group) in contradistinction to the dominant identity as 'abnormal'. This division between the normal and the abnormal is a key theme in Foucault's analysis of society.

> When a judgement cannot be framed in terms of good and evil, it is stated in terms of normal and abnormal. And when it is necessary to justify this last distinction, it is done in terms of what is good or bad for the individual. These are expressions that signal the fundamental duality of Western consciousness. (Foucault, 1977b: 230)

Foucault's analysis of the normal/abnormal dualism is founded in the work of Nietzsche who analyzed the 'fundamental duality' of Western society in terms of three types of questions that have been described as 'objective', 'subjective', and 'functionary'. In objective terms, Nietzsche asked under what conditions were the value judgments, good and evil, constructed. Foucault, translating Nietzsche's questions into a secular society, focused not on good and evil but on normal and ab normal. This examination of the political economy of normality and abnormality is evident in all of Foucault's substantive works, which are devoted to examining the ways that judgments of normality and abnormality have been constructed. In subjective terms, Foucault has ex tended Nietzsche's notion of religious discourses causing an 'inner split' in the subject. Foucault suggests that

> [w]here religions once demanded the sacrifice of bodies, knowledge now calls for experimentation on ourselves, calls us to the sacrifice of the subject of knowledge. The desire for knowledge has been transformed among us into a passion which fears no sacrifice, which fears nothing but its own extinction. (Foucault, 1977c: 163)

The will to know about oneself and to study one's psyche produces a split subject/object, whereby the subject's own subjectivity becomes an object for the will to truth. Foucault asks how this inwardly divided self has come about, and who has benefited from its construction. His (partial) answer is that the oppositional discourse of truth and falsehood is somehow implicated in this process, and much of his substantive work attempts to illustrate the way this occurs in specific places and times. Finally, in functionary terms, Foucault extends Nietzsche's questions about the role of the 'ascetic priest' in promoting the divisions and dualisms that have come to dominate human thinking in Western societies. Writ ing in a secular era, Foucault asked what role the 'intel lectual' (the educated elite, not necessarily academics) played in the construction and elaboration of the social divisions between true and false, normal and abnormal. How were social divisions created? What power interests were served by these divisions? Having discussed in some detail Foucault's conceptualization of 'discourse', and his somewhat vague methods for analyzing discourses, it is now possible to focus attention toward how geographers might be 'doing' discourse analysis.

'Doing' Discourse Analysis

Discourse Analysis as a Difficult-to-Define Method

Social scientists (and others) have inherited from Foucault some reticence for explaining discourse analytic 'methods'. Clearly some of this reticence comes from Foucault's own distaste for outlining his methods, but it also comes from a post positivist sensibility that is itself wary of the idea of a simple list of methods that can be applied to discourse analysis. As Jonathan Potter and Margaret Wetherell observe, "one of the difficulties in writing about the process of discourse analysis is that the very category 'analysis' comes from a discourse developed for quantitative positivist methodologies." The result is that few social scientists write about methods in discourse analysis. In a recent review of discourse ana lysis as method, Gordon Waitt notes that:

> discourse analysis is positioned as a 'craft skill' (Potter 1996, p. 140), the researcher's ability 'to customise' (Phillips and Hardy 2002, p. 78), 'rigorous scholarship' (Gill 1996, p. 144), or 'human intellect' (Duncan 1987, p. 473). The maxim is 'learning by doing'. Through practice, discourse analysis is typically held to become intuitive. Scholarly passion seemingly underpins critical textual analysis. The methodology is often left implicit rather than made explicit. Undermining the very basis of discourse analysis is research that is too systematic, mechanical, and formulaic (Burman and Parker 1993) (emphases and citations in original). (Gordon Waitt, 2005: 179)

Human geographers have perhaps been even less forthcoming about their discourse analytic methods. In a recent discussion of a series of special sessions focused on discourse analysis in urban geography held at the Asso ciation of American Geographers 2003 conference, Loretta Lees observed that "out of sixteen papers, not one outlined in any detail how they had gone about their particular discourse analysis." In a similar review of discourse analysis in policy research Yvonne Rydin notes "the lack of methodological detail in many accounts of discourse analysis." Two review essays are emblematic of the tendencies in human geography. Henry Yeung undertook a methodological review of 'new economic geography', and in spite of the fact that part of the focus of the review was on the analysis of economic (and noneconomic) discourses, he never mentioned any of the actual methods of discourse analysis. Similarly, in his recent review essay, Keith Jacobs sets out, in part, to provide "an evaluation of its [discourse analysis] strengths and weaknesses as a method", but then does not actually discuss the specificities of the 'methods' of dis course analysis, focusing instead on the epistemological concerns of the papers that use discourse analysis. The point here is not to single these two authors out for such failures, but instead to point out how they represent the standard practice in human geography.

Clearly, there is more at work here than straight forward opposition to the discussion of methods, and in part, the nature of discourse analysis itself poses prob lems for those who attempt to describe it. Indeed, it is clear that many human geographers have taken their cues from writers such as Margeret Wetherell and Jonathan Potter , who argue that "much of the work of discourse analysis is a craft skill, something like bicycle riding or chicken sexing, that is not easy to render or describe in an explicit or codified manner." The result is that human geographers have taken an eclectic and often 'intuitive' approach to analyzing discourse, and that few of their specific methods have been recorded or discussed in a systematic manner.

Doing Discourse Analysis

Recognizing the shortcomings inherent in an approach with no written canon of methods, Gillian Rose set out to explicitly define seven key methodological components to discourse analysis of visual materials. Gordon Waitt has adapted these to the wider analysis of discourse in other geographical objects, and the following discussion attempts to summarize the seven key methodological issues discussed by these authors.

Suspending preexisting categories

Given that one of the key components of any discourse analysis is to attempt to identify hegemonic discourses in action, it is especially important for a discourse analyst to step outside hegemonic discourses. This can be seen as part of an attempt to examine a text "with fresh eyes and ears." It is also part of the process of engaging in reflexive analysis, where the analyst examines their own position in the discourse, and how that position helps to constitute particular understandings of the issues under analysis. Interestingly, immersing oneself in readings in critical social theory is perhaps the best way to gain the skills to step outside hegemonic discourses such as liberalism.

Absorbing oneself in the texts

It is especially important for any discourse analyst to absorb themselves in their objects of analysis. This involves becoming fully familiar with the texts that one is studying, and beginning, through the process of 'reading' and 're reading', to identify any particular themes that arise in the reading of texts.

Coding Themes

Once familiar with the objects of analysis, it is time to start to 'code' for particular themes. Coding can take many forms, but it usually involves some form of categorization of particular aspects of the objects of analysis. It is important at this juncture to pay particular attention to the ways that the producer and consumer of a text is positioned by the text. It is also important to come to understand how objects discussed in the text (people, places, and things) are also positioned textually. Some questions one might ask include does the producer use third person narrative, distancing themselves from their text? How is the author of a text explicitly or implicitly gendered? Is the text written from a 'masculine' or 'feminine' perspective? Is the reader assumed to occupy a particular social class? Are people in the text racialized? Are there particular stereotypes drawn upon or reinforced in the text? What role does space play in the constitution of subjectivities and subject positions? These are all difficult questions to answer, and they often require specific theoretical knowledge in order to do so.

Identifying 'regimes of truth'

All discourses rely on particular knowledge that defines the legitimacy of any particular statement of narrative in terms of truth/falsehood, normal/abnormal, moral/immoral, etc. It is useful to try to understand the mechanisms by which a particular discourse is seen to have both validity and worth. What patterns of 'truth' can be discerned? What experts are called upon to pronounce truth? In a pathbreaking analysis of news media discourse, Stuart Hall refers to a repetitious construction of "a world at one with itself" in news stories. This process involved a cyclical story starting with a crisis, a response to the crisis, and then a recourse to government reassurances that the crisis would never happen again. Time and again, Hall noted, government experts or senior ministers were called upon to let the reader know that all was again well in the world. Similarly, Bruce Braun identified a series of 'buried epistemologies' that underpin discussions of forest management in British Columbia and act to exclude any discussion of aboriginal rights and title to lands in question.

Identifying inconsistencies

Hegemonic discourses are always contested by subordinate discourses and all discourses are characterized by inconsistencies, contradictions, and paradoxes. It is theoretically important to identify these inconsistencies, and it is equally important to identify how such inconsistencies might contest or reinforce the dominant meanings constituted in a given discourse. Perhaps more importantly, these inconsistencies and disruptions allow for the creation of new subject positions and identities in discourse. It is important for the discourse analyst to be aware of these kinds of processes, especially when undertaking a genealogical or historical analysis of discursive changes over time.

Identifying absent presences

Discourses often rely on silences for their power and it is important for discourse analysts to identify these silences and to theorize how they operate to create particular subjects (and how they might erase others). Nationalist discourse provides a particularly telling example of the work that silence does. In the case of Canada, for example, there exists a dialectic of remembering and forgetting that:

> is expressed through efforts to tell stories of the 'long history' of white settlers and their place in the nation,

> along with two concomitant silences: the first silence elides the brutality of colonial settlement, especially the confiscation of land and subsequent placement of the first inhabitants … on reserves (Harris, 2002); the second si lence hides the relative recentness of colonial settlement. In this way, white Canadians construct a story of "European settlers as bearers of civilization, while si multaneously trapping Aboriginal people in the pre modern" (Razack 2002: 2), thus excluding them from participating in modern life. (Berg *et al*., 2007)

In this way, aboriginal people occupy discussions of the Canadian nation as an absent presence that reinforces white Canadians' sense of superiority and self.

Identifying social contexts

Finally, it is especially important to identify the social context within which discourses arise, or that may be constituted in and by discourses. While this may seem like a simple task, it is important to remember that the key goal here is to attempt to link the production of discourse with the production of key subject positions (audience, reader, writer, producer, etc.), and to think about how power operates in these contexts to (re)produce social relations and subject positions for social actors.

Summary

While human geographers have drawn upon a range of influences, it would be fair to argue that the predominant approach to discourse analysis today takes its cues from the work of Michel Foucault. Foucault's reticence to explain his methods in any thoroughgoing fashion has resonated with subsequent generations of discourse analysts, and human geographers have rarely provided any explicit explanations about how they go about 'doing' discourse analysis. In part, this is because discourse analysis itself operates outside hegemonic discourses of 'research method'. Accordingly, and while there are now a few discussions of methods, including helpful works by Rose and Waitt, it is clear that discourse analysis remains a difficult 'craft skill', one that takes some time to learn and one that is always fraught with the very problems it seeks to define.

See also: Discourse; Foucauldianism; Poststructuralism/ Poststructuralist Geographies; Text and Textuality; Text, Textual Analysis.

Further Reading

Bannet, E. T. (1989). *Structuralism and the Logic of Dissent. Barthes, Derrida, Foucault, Lacan*. London: Macmillan.

Barnes, T. J. and Duncan, J. S. (eds.) (1992). *Writing Worlds. Discourse, Text and Metaphor in the Representation of Landscape*. London: Routledge.

Barthes, R. (1973). *Mythologies* (trans. Lavers, A.). St. Albans: Paladin.

Berg, L. D., Evans, M., Fuller, M. D. and the Okanagan Urban Aboriginal Health Research Collective (2007). Special issue of ACME: Ethics, hegemonic whiteness, and the contested imagination of aboriginal community in social science research in Canada. In Pain, R. & Kindon, S. (guest eds.) *Participatory Research Ethics*. ACME: An international e journal for critical geographies. http:// www.acme journal.org/vol6/LDBetal.pdf (accessed Jun. 2008).

Cooper, D. (1994). Productive, relational and everywhere? Conceptualising power and resistance within Foucauldian feminism. *Sociology* 28, 435 454.

Dalby, S. (1990). *Creating the Second Cold War*. London: Pinter.

Dear, M. (1988). The postmodern challenge. Reconstructing human geography. *Transactions, Institute of British Geographers NS* 13, 262 274.

Deleuze, G. (1983). *Nietzsche and Philosophy* (trans. Tomlinson, H.). London: Athlone.

Derrida, J. (1976). *Of Grammatology* (trans. Spivak, G. C.). Baltimore, MD: Johns Hopkins University Press.

Fabian, J. (1983). *Time and the Other: How Anthropology Makes its Object*. New York: Columbia University Press.

Fairclough, N. (1992). *Discourse and Social Change*. Cambridge: Polity Press.

Fairclough, N. (2003). *Analyzing Discourse: Textual Analysis for Social Research*. New York: Routledge.

Foucault, M. (1965). *Madness and Civilization: A History of Insanity in the Age of Reason*. New York: Pantheon.

Foucault, M. (1970). *The Order of Things: An Archaeology of the Human Sciences*. London: Tavistock.

Foucault, M. (1972). *The Archaeology of Knowledge*. London: Routledge.

Foucault, M. (1973). *The Birth of the Clinic: An Archaeology of Medical Perception*. London: Tavistock.

Foucault, M. (1977). *Discipline and Punish. The Birth of the Prison*. New York: Vintage Books.

Foucault, M. (1977). History of systems of thought: Summary of a course given at Collège de France 1970 1971. In Bouchard, D. (ed.) *Language, Counter Memory, Practice: Selected Essays and Interviews*, pp 199 204. Ithaca, NY: Cornell University Press.

Foucault, M. (1977). Nietzsche, genealogy, history. In Bouchard, D. (ed.) *Language, Counter Memory, Practice: Selected Essays and Interviews* (Reprinted from *Hommage à Jean Hyppolite*, Paris: Presses Universitaires de France, 1971), pp 139 164. Ithaca, NY: Cornell University Press.

Foucault, M. (1977). Revolutionary action: 'Until now'. In Bouchard, D. (ed.) *Language, Counter Memory, Practice: Selected Essays and Interviews* (Reprinted from *Actuel*, vol. 14, Nov. 1971), pp 218 233. Ithaca, NY: Cornell University Press.

Foucault, M. (1978). *The History of Sexuality, Vol. I: An Introduction*. London: Penguin.

Foucault, M. (1980). Truth and power. In Gordon, C. (ed.) *Power/ Knowledge. Selected Interviews and Other Writings, 1972 1977*, pp 109 133. New York: Pantheon.

Foucault, M. (1984). The order of discourse. In Shapiro, M. J. (ed.) *Language and Politics* (Reprinted from Robert, Y. (ed.) (1981). *Untying the Text*. Boston, MA: Routledge), pp 108 138. Oxford: Blackwell.

Hall, S. (1973). A world at one with itself. In Cohen, S. & Young, J. (eds.) *The Manufacture of News: Social Problems, Deviance and the Mass Media*, pp 85 94. London: Constable.

Jacobs, K. (2006). Discourse analysis and its utility for urban policy research. *Urban Policy and Research* 24(1), 39 52.

Johnson, B. (1981). Translator's introduction. In Derrida, J. (ed.) *Dissemination*, pp vii xxxiii. Chicago, IL: University of Chicago Press.

Mills, S. (1997). *Discourse*. New York: Routledge.

Nietzsche, F. (1967). *The Will to Power* (trans. Kaufmann, W. and Hollingdale, R. J.). New York: Random House.

Nietzsche, F. (1969). *On the Genealogy of Morals* (trans. Kaufmann, W. and Hollingdale, R. J.). New York: Vintage.

Ó Tuathail, G. and Agnew, J. (1992). Geopolitics and discourse: Practical geopolitical reasoning and American Foreign Policy. *Political Geography* 11, 155 175.

Patton, P. (1994). Foucault's subject of power. *Political Theory Newsletter* 6, 60 71.

Potter, J. and Wetherell, M. (1994). Analyzing discourse. In Bryman, A. & Burgess, R. G. (eds.) *Analyzing Qualitative Data*, pp 47 66. London: Routledge.

Rose, G. (2001). *Visual Methodologies: An Introduction to the Interpretation of Visual Materials*. London: Sage.

Rydin, Y. (2005). Geographical knowledge and policy: The positive contribution of discourse studies. *Area* 37(1), 73 78.

Said, E. W. (1978). *Orientalism: Western Conceptions of the Orient*. London: Penguin.

Soja, E. (1989). *Postmodern Geographies. The Reassertion of Space in Critical Social Theory*. London: Verso.

Waitt, G. (2005). Doing discourse analysis. In Hay, I. (ed.) *Qualitative Research Methods in Human Geography* (2nd edn.). pp 163 191. Melbourne, VIC: Oxford University Press.

Wetherell, M. and Potter, J. (1992). *Mapping the Language of Racism: Discourse and the Legitimation of Exploitation*. New York: Columbia University Press.

Wetherell, M., Taylor, S. and Yates, S. J. (2001). *Discourse Theory and Practice: A Reader*. Thousand Oaks, CA: Sage.

Willems Braun, B. (1997). Buried epistemologies: The politics of nature in (post)colonial British Columbia. *Annals of the Association of American Geographers* 87(1), 3 31.

Yeung, H. W. C. (2003). Practicing new economic geographies: A methodological examination. *Annals of the Association of American Geographers* 93(2), 442 462.

Relevant Websites

http://users.utu.fi/bredelli/cda.html
Brett Dellinger's Critical Discourse Analysis'.

http://www.aber.ac.uk/media/Documents/S4B/semiotic.html
Daniel Chandler's "Semiotics for Beginners".

http://extra.shu.ac.uk/daol/editorial/
Discourse Analysis Online.

http://www.lboro.ac.uk/research/mmethods/resources/links/da.html
Loughborough University Department of Social Sciences Discourse Analysis.

http://www.ischool.utexas.edu/~palmquis/courses/discourse.htm
University of Texas Discourse Analysis.

Disease Diffusion

A. Schærström, Kometvägen, Täby, Sweden

Glossary

Carrier An infected but symptomless individual.
Case Fatality The percentage of a group affected by a certain disease ending with death.
Contagious 1. (about diseases) Infections which are directly transmissible between humans or between animals and humans. 2. (spatial diffusion) Diffusion by proximity.
Determinant A factor or circumstance that does not directly cause a certain kind of ill health but triggers or promotes the development of disease.
Disease A particular, medically defined condition of bad health, usually defined in terms of causes, signs, and symptoms, which are described in the ICD.
Disinsection Spraying aircrafts against insects.
Endemic A disease being continuously present in a certain place is said to be endemic.
Epidemic Outbreak and extensive spread of an infectious disease.
Epidemiological Transition The shift from a stage of predominant, infectious, parasitic, and nutritional diseases to a stage characterized by 'man-made' diseases. It is also an essential component of the demographic transition.
Epidemiology The scientific study of the extension, causes, and processes of diseases.
Epizootic The rapid and extensive diffusion of an animal disease, corresponding to an epidemic.
Etiology The cause or causal circumstances of a disease.
Habitat The natural living environment of a species.
Host The organism that is infected by a pathogenic agent.
ICD International Classification of Diseases issued and regularly revised by the World Health Organization.
Illness Often understood as the subjective perception of being in poor health cf. disease and sickness.
Incidence The number of individuals who fall ill (i.e., new cases) during a period of time compared with the population at risk (usually expressed as cases per 1000 or 10 000), cf. prevalence.
Infection A disease caused by a microorganism.
Intermediate Host An organism that is necessary for a pathogenic organism at some stage of its life cycle, cf. host.
Lentivirus Virus with delayed effects causing symptoms a long time after infection.
Nosocomial Infection originating in a hospital environment.
Osteoporosis Bone fragility.
Pandemic Epidemic spreading globally.
Pathogen A factor which may actually cause a disease, for example, bacteria, viruses, carcinogenic substances, toxic compounds.
Pathogenesis The process that leads to pathological changes in a living being.
Prevalence The frequency of a disease at a certain point in time, usually expressed as a ratio between the number of cases and the population at risk (such as cases per 1000 or 10 000), cf. incidence.
Reservoirs Animal populations in which pathogenic microorganisms can survive and circulate between outbreaks among humans.
Sickness Often the socially accepted perception of being ill, the concept used for administrative purposes, like sick-listing, cf. illness and disease.
STD Sexually transmitted diseases.
Time Geography A theoretical and conceptual framework, first conceived by Torsten Hägerstrand, based on the axiom that everything has a location and an extension in time as well as in space, supported by a particular visualization technique.
Vector 1. An animal that transmits the actual pathogen from one individual to another, also between animals and humans, such as fleas, mosquitoes, snails, or ticks. 2. Arrow indicating movements, for example disease diffusion routes, in maps.
Zoonose A disease that can be transmitted from animals to humans under natural conditions.

Introduction

It is obvious with many diseases – and with health status in a wider sense – that their geographical distribution tends to change, that is, the spatial occurrence or extension of diseases varies through time. So does the intensity of a certain disease, calculated as incidence, prevalence, mortality, or case fatality. As diseases – and health status in general – are largely functions of individual properties, behavior, and environmental circumstances, their spatial distribution will tend to change over time as individuals come and go and as the physical and social environment changes. Infections will spread with people, animals, and things moving from place to place. The distributions of noninfectious diseases will change too, and ostensibly these redistributions will appear as spatial diffusion and waning, but the character of the process will depend on the type of the disease and a multitude of geographically different physical and social circumstances.

With increasing mobility and environmental change, disease diffusion is an urgent public health issue on local, regional, national, and global levels. Recently several infectious diseases, which have been more or less under control or even eradicated from large parts of the earth, have reemerged, while 'new' diseases have appeared and noninfectious ones have expanded in numbers and spatial distribution.

Along with epidemiology, concepts and methods from geography and cartography can be useful for understanding, describing, and analyzing the processes leading to new spatial patterns of ill health, but current epidemiological trends call for methodological and theoretical development.

Disease Occurrence – Disease Distribution

Basically, disease distribution or occurrence can be perceived in two different ways. On the one hand, it can be thought of as the places where cases (i.e., affected individuals) are found. On the other hand, it could be perceived as the locations where the necessary circumstances for causing illness are prevalent, for example, the microbes or living conditions that constitute health hazards.

In the first view, disease is seen as an outcome which can be observed, like any state of health, in an individual. In other words, information about disease occurrence is linked to the individuals. This view is a basis for epidemiology, which measures incidence and prevalence in search of correlations with circumstances that might clarify the etiologies of diseases. The occurrence of disease could thus be considered equivalent with the individuals themselves and from this perspective geographical distribution can be understood as a number of more or less mobile points with varying lifetimes, sometimes appearing in clusters or within more or less clearly distinct regions or zones. This notion has led to important discoveries and progress in epidemiology and medical geography, most notably John Snow's mapping of cholera in Soho (**Figure 1**). The alternative view focuses on the causes, which means that disease occurrence is associated with an area where some potential health hazard prevails. Pathogenic factors may be around, but cases actually occur only when susceptible individuals are exposed to them. This perspective is the basis for identifying risk areas often visualized on maps in vaccination clinics.

Consequently, there are also two main ways to understand disease diffusion – either as the diffusion of cases or as the diffusion of the causes and hazards.

Since every spatial observation is valid at a certain point or period in time the notion of disease diffusion is a spatial–temporal concern.

In a formal sense disease occurs when a case is discovered, diagnosed, and registered, making it available for research and statistics. In a more real and personal sense disease occurs once an individual feels ill, which often but not necessarily happens before the diagnosis is made. It is also possible that a disease (e.g., cancer in its early stage) is discovered at a routine screening before the patient has noticed any symptoms. In a wider perspective, a pathogenic process may begin with exposure to circumstances that may eventually lead to a disease. Between exposure or the start of a pathogenic process and the onset of symptoms lies a period of incubation or latency. With influenza, the incubation time is a matter of days, whereas AIDS, which is caused by a lentivirus, has an incubation period of months or years and hepatitis C might cause only mild or no symptoms and therefore remain undetected for decades. With noninfectious diseases, such as cancer and neurological disorders, the latency period may be not only years, but even decades. In the meantime, the affected individuals may have moved, which means that the exposure or pathogenic onset may have happened in a different place and a different environment than the place where the case is eventually recorded. This time–space lag complicates the notions of disease occurrence and diffusion.

Mapping disease diffusion must account for the spatial changes in the time dimension. Thus, cartographical methods have included vectors, isopleths (isochrones), map sequences, and time geographical visualization.

Diffusion Process

While medicine focuses on causes, exposure, transmission of pathogens, and the ensuing pathological outcome, geographical issues concern the locations and numbers of cases plus the directions and pathways that diseases apparently follow to reach new victims as well as the means and geographical circumstances that favor or inhibit the spatial spread of ill health.

Spatial diffusion theory makes a main distinction between the dispersal and the structural dimension of a diffusion process. Whereas the dispersal aspect of the diffusion concerns the routes of the movements, the structural dimension refers to the reciprocal relationships between the locations along the route.

The dispersal dimension is designated either contagious (contact) or hierarchical while the structural dimension is characterized as either expansion (expansive) or relocation. The word contagious is used as a metaphor in diffusion theory. When it is used in connection with disease diffusion it does not necessarily refer to infections

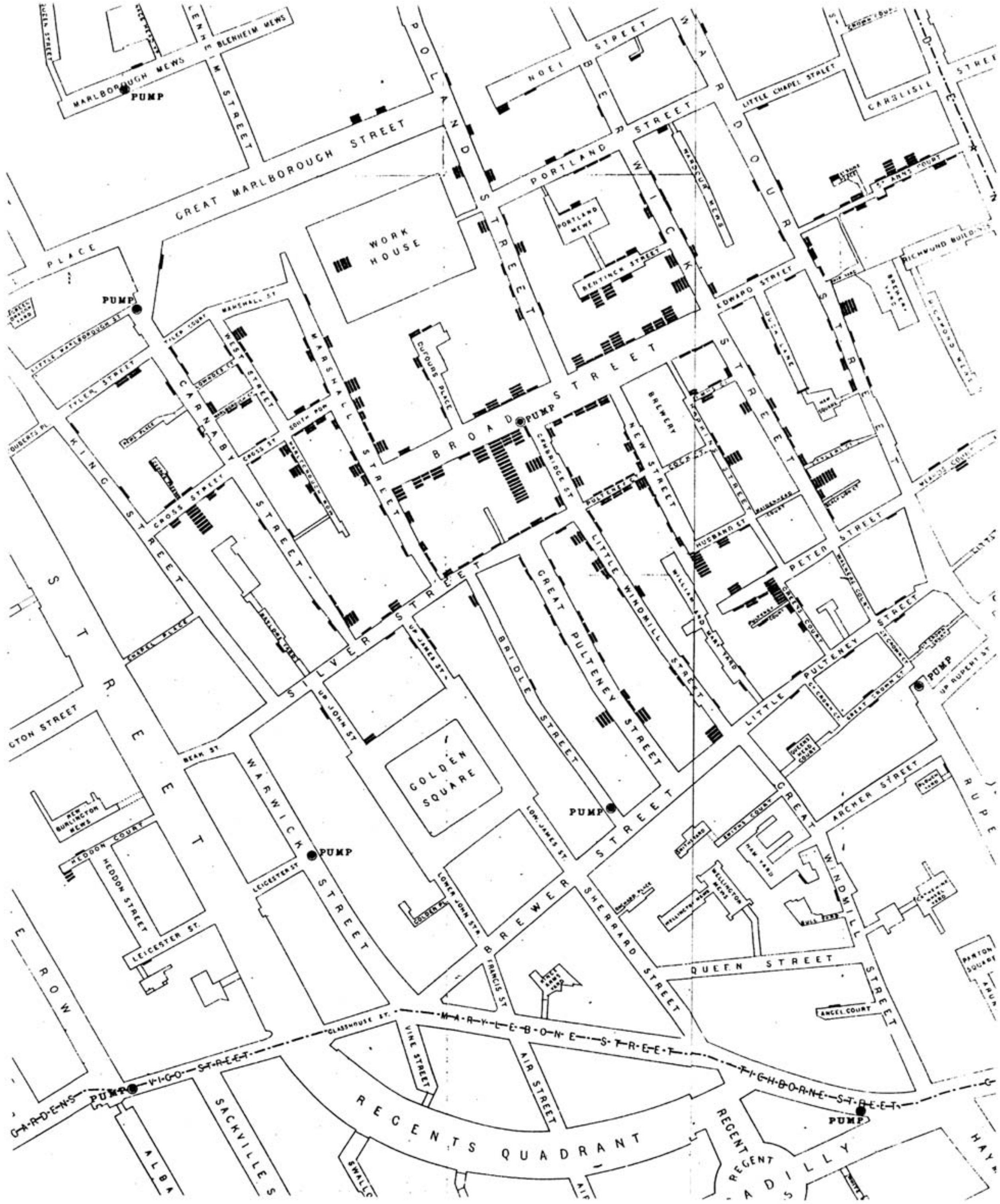

Figure 1 During a cholera epidemic in London 1853 54, Dr. John Snow mapped every death on a map and found that they clustered around one of the public pumps. This observation supported the hypothesis that cholera was spread by water.

that are literally transmitted from individual to indi vidual. Thus, drug abuse and its health consequences might spread from a metropolitan area to neighboring suburbs because of their proximity before spreading to regional centers and towns further down the urban hierarchy. In this context, contagious means that some thing tends to spread by proximity – to nearby places, whereas hierarchical diffusion takes place in a stepwise manner – from one, usually more central, location to the other, not necessarily nearby, locations from which the diffusion may proceed.

Considering the different ways of defining disease occurrence, this typology is not unconditionally applic able to disease diffusion. Medical geography has largely concentrated on the diffusion of infections, but the oc currence of noninfectious disorders displays similar trends. Epidemics/epizootics and pandemics spread as microorganisms and vectors move over distances and through time – just like goods, travelers, innovations, information, or habits. As to noninfectious diseases, the actual diffusion process may be more complex and con stitutes a methodological challenge.

Increasing numbers of cases, as such, reflect a diffusion process, but the spatial process may take very different shapes. First, diffusion might imply that more individuals are taken ill within unchanged outer spatial limits, that is, a condensation but no spatial expansion of cases takes place. Second, diffusion could imply that more individuals are affected within a wider area (i.e., the outer limits of a core area are widened and the process is expansive), for example, as drug abuse spreads from the inner city to the suburbs. Third, diffusion could mean that individuals in other places fall ill. In the latter case, different types of dif fusion may be more or less distinguishable. If an epi demic affects one country or continent after another, the extension of the disease is enlarged as more individuals in new places fall ill. Eventually the disease might subside in the locations where it struck first, which would appear as a relocation of the disease, but the individual cases are not relocated. Relocation of cases would imply that af fected individuals have moved to other places. This, however, is conceivable with genetically conditioned disorders.

A crucial issue in epidemiological geography is to what extent the observed cases of disease in a certain place are caused there as a result of selective migration, that is, individuals with a certain predisposition have moved to that place. For example, it has been suggested that persons with poor mental health or risky behavior (e.g., abuse) would tend to move to cities (drift), which would imply a relocation of health problems. Alter natively, physical and psychosocial stimuli in the urban environment would cause (breed) ill health. Sometimes, this issue is referred to as the breeder and drift hypotheses.

The diffusion process involves numerous events on all spatial levels from intercontinental transport and climatic variations down to households, workplaces, and hospitals. Geographically speaking, besides what is being spread, a diffusion process involves a place of origin (a source), one or more receiving places, the pathways of spread, and a force driving the diffusion.

1. The potential for spatial diffusion depends on the character of the causative agents, such as their trans mission modes, mobility, reach, and viability, as well as various determinants. Exposure (i.e., the contact be tween causes and susceptible individuals) is crucial for transmission and the start of a pathological change in a human or other living being.

   Contagious diseases (e.g., influenza or STDs (sexually transmitted diseases)) may be transmitted either directly to a susceptible person through in timate contacts with infected humans and animals or indirectly by a vector (e.g., ticks – tick borne en cephalitis (TBE) and Lyme disease; mosquitoes – malaria). The pathogen may also enter a human body through the air as droplets, with blood, food, and water as well as by surface contacts with objects. The viability of pathogenic agents and their vectors is a crucial factor with infectious diseases. Unless they find a suitable host soon enough they will perish, no transmission will take place, and the diffusion will stop.

   Noncontagious diseases may be caused or triggered by some physical, chemical, or even social exposure. Some diseases may be the outcome of genetic pre dispositions, either inevitably caused by genes or triggered by external circumstances. The reach of noninfectious pathogens (such as radioactivity or pollutants) is closely related to the medium which carries them, for example, winds and waters.
2. Like other geographical objects, sources of disease diffusion may assume point, line, area, or diffuse shapes. Point sources may be infected wells or pol luting anthropogenic plants, such as incinerators. Line shaped potential sources could be polluted riv ers, heavily trafficked roads, or high voltage cables. Area sources of disease diffusion may be endemic regions and natural reservoirs or places with high radon content in bedrock or buildings. Finally, diffuse sources could be certain behaviors, habits, and practices.
3. Next, no infectious diffusion can become apparent (in terms of cases) unless the pathogenic influence reaches a place with a population including suscep tible or vulnerable individuals.
4. The courses of diffusion are guided and shaped by available routes and nodes as well as barriers. Throughout history, contagious diseases have fol lowed people and animals along rivers, roads, rail roads, and caravan trails. Eventually railways, steamships, trucks, buses and, not least, airlines have altered the modes, velocity, and routes of disease diffusion. Winds and water also play a role for the transportation of organisms and vectors. Thus, foot and mouth disease has probably been spread by winds from farm to farm without any direct contact between

infected and susceptible animals. Cholera diffusion may have been influenced by wind directions.

Diffusion of contagious diseases may also follow social routes without obvious physical manifestations. Such diffusion has been mapped, for example, HIV in personal networks.

Noninfectious pathogens, such as radiation, chem ical emissions, and toxic compounds from volcanic eruptions may be carried by winds and waters.

Nodes of diffusion are places where people fre quently meet. Typical nodes for the transmission of contagious diseases are urban centers, ports, refugee camps, marketplaces, trading towns, and border towns. Due to poor sanitary and crowded conditions refugee camps are high risk areas for infections. Thus, in African refugee camps diseases like AIDS and other STDs, cholera, dysentery, plague, diphtheria, Ebola and Lassa fever have been reported. Other nodes are large workplaces, shelters for homeless, schools, and prisons. Likewise, not only permanent localities, but certain occasions such as local festivities and similar social events may increase transmission risks. For ex ample, haymaking in the summer has probably been an occasion for spreading measles in Iceland, con tributing to characteristic seasonal waves.

On the contrary, disease diffusion may be stopped or deflected by permanent or temporary barriers. Physical barriers, such as vast spaces of water, mountain chains, or large ice sheets, can be effective against 'contact' diffusion (to nearby places). Diffusion may be stopped or slowed down by biological barriers such as widespread immunity and habitats unsuitable for agents or vectors. Further, social barriers involve cultural practices, habits, and knowledge. Thus, dis ease diffusion may be precluded by public health interventions and precautions, such as quarantine procedures, disinfection, disinsection, and vaccination campaigns, or by widespread knowledge about per sonal protection.

5. Finally, diffusion depends on micro scale events that bring individuals in contact with health hazards. Humans, animals, transport networks, vehicles, and various objects may be involved in the process. Oc cupational conditions, daily personal routines and attitudes may be decisive for exposure. Sexual and other cultural practices play an important role for either spreading or barring the spread of STDs – for example, promiscuity versus fidelity, polygamy, cir cumcision and other rites.

   Key groups for infectious transmission are highly mobile people like migrants (such as migrant la borers), refugee populations, truck drivers, and com mercial sex workers and their clients. In situations of armed conflicts, troops often appear as another key group.

In reality, it is not unusual to see different epidemic waves of a certain disease behaving differently or to find both 'contagious' and hierarchical phases simultaneously or sequentially as parts of a larger diffusion process. Also, new geographical conditions affect the diffusion routes. For example, cholera epidemics swept across the United States and Russia several times in the nineteenth century and it has been ascertained that in the later part of the century both diffusion routes and velocity were affected by the growing railway networks. Instead of advancing slowly to the most nearby places the disease would leap from place to place along the lines (**Figure 2**).

Political circumstances sometimes affect, indirectly, disease diffusion in different ways. Thus, the South African apartheid regime seems to have impacted the spread of HIV in the region in two ways. As the regime imposed restrictions on nonwhites' movement within the country, the diffusion was inadvertently constrained. Then large numbers of opponents went into exile in neighboring countries where HIV was more prevalent, thus becoming exposed and later bringing the disease to South Africa when they returned after the change of regimes.

In all, the conditions for disease diffusion are rooted in complex ecological circumstances and chains of causes and determinants. Proximal determinants (sometimes called downstream factors) are involved with the actual transmission, such as commuting and sexual behavior, whereas distal determinants (or upstream factors) cover micro and macro scale, socioeconomic, and cultural conditions, like migration, poverty, urbanization or religion.

Epidemiological Landscapes

The geographical context of health and disease can be thought of as an epidemiological landscape or a landscape of exposure. The pathogenic complex with a similar concept distinguishes natural bio or eco pathogenic complexes from two types of man made complexes. Whereas socio pathogenic complexes refer to potentially harmful social circumstances, such as standard of living, occupational environment, food habits, and other cultural phenomena, techno pathogenic complexes are associated with technical products, by products, or waste.

Conversely, the term therapeutic landscape has been suggested for places that exert a healthy influence.

Several concepts are used to designate areas with certain disease ecological characteristics. Natural focus refers to a landscape associated with biotic and abiotic factors which are vital for the pathogens and other factors involved in the ecology of a particular disease. Natural *nidus* (plural. *nidi*) is defined as a micro scale region constituted of a living community where a disease agent

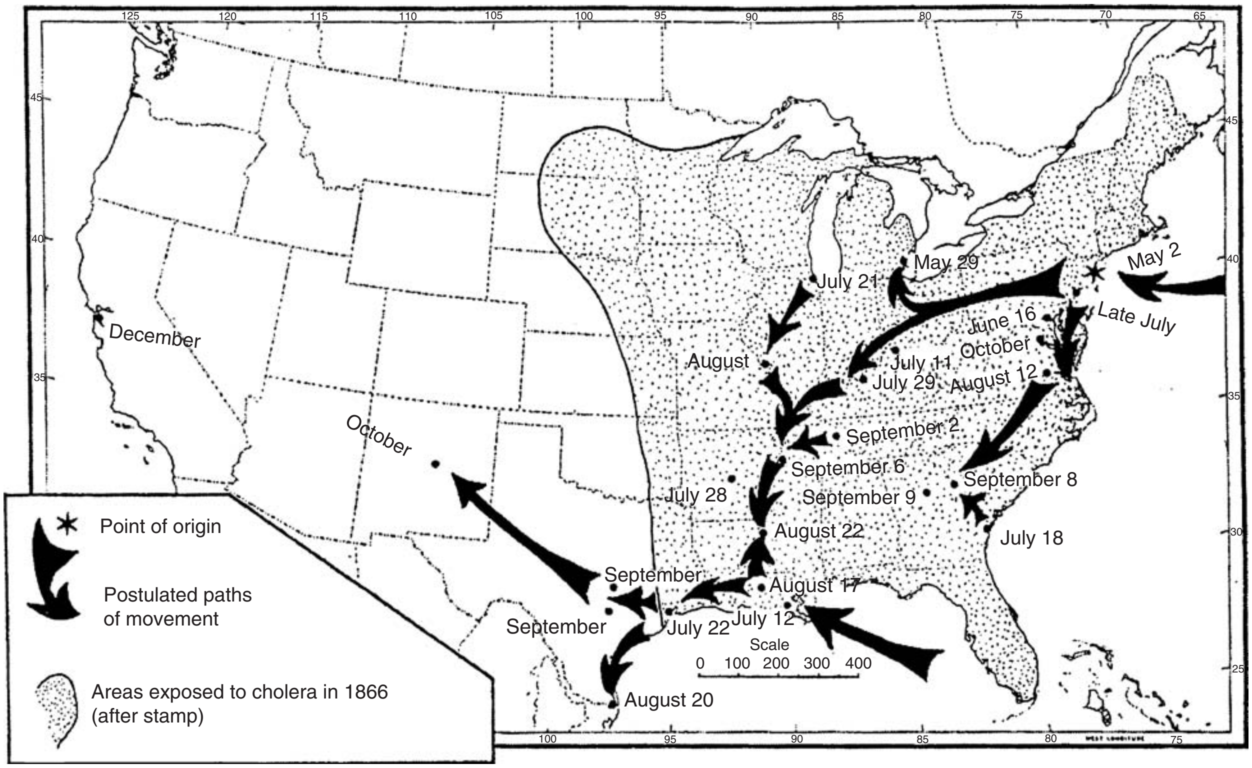

Figure 2 The movement of cholera epidemic (1866) through the United States. The first case was recorded in New York, from where the disease spread in a largely hierarchical way. The railroad net which had been developed in the eastern part of the country facilitated the disease diffusion.

continually circulates among the individual members and the habitat conditions are necessary to maintain the disease system. Nosochoros and urban risk cells are other concepts with similar meanings.

Landscapes are dynamic, that is, their contents and extensions will change over time with consequences for health and disease. This change is closely connected with the process known as epidemiological transition.

Epidemiological Transition and Future Prospects

Throughout history, epidemics and pandemics of in fections have repeatedly swept across the earth from their endemic areas to vulnerable populations and places. The era of European discoveries brought people and in fectious diseases from the Old World and other contin ents in contact with each other.

However, disease ecology has gone through an epi demiological transition which implies that diseases caused by infections, parasites, and poor nutrition eventually have been mastered, more or less, but superseded by others, usually called chronic, degenerative, or man made diseases. This transition, which started in the nineteenth century with a protracted process in England and some other European countries, has eventually reached other parts of the world, where the process has tended to be faster. This transition might appear as a spatial diffusion of diseases, but is rather the outcome of changing bio logical as well as geographical preconditions for disease diffusion, including diffusion of determinants.

First, social and structural changes have altered the ecological landscapes. Large scale urbanization and new lifestyles provide new opportunities for transmission and exposure. Further, as demographic structures are chan ging, susceptibility and vulnerability assume new spatial patterns.

Second, the conditions for diffusion have been rad ically altered by increasing spatial connectivity – modern transportation systems as well as increasing and diversi fied mobility. Migration and other types of mobility ap pear in many shapes with complex effects on the diffusion and distribution of health and disease. The notion of a 'healthy migrant effect' is based on the as sumption that persons in good health are more likely than others to migrate. However, 'forced migration' is a common effect of natural disasters, enduring violence, and adverse living conditions, which might induce people in poor health to move. Also, elderly or persons already suffering from poor health may move to healthier places or closer to healthcare. Thus, migration entails large scale relocation of health related factors. Commuting entails regular exposure to different environments, while tourism and other long distance traveling bring people in occasional contact with unfamiliar risks.

Third, global and local environmental changes alter the health prospects. In the future, the living conditions for agents and vectors involved in disease ecology might change in the wake of climatic change, just like prevalent winds and weather conditions might favor or disfavor the diffusion of a particular epidemic.

In all, the present global panorama of health and disease includes new, emerging, and reemerging diseases as well as increased frequencies of chronic disorders. Infectious epidemics continue to change the distribution of health in short and long time perspectives, but non infectious health issues are also sometimes said to assume epidemic proportions, for example, smoking, overweight, and obesity as well as drug abuse and its health

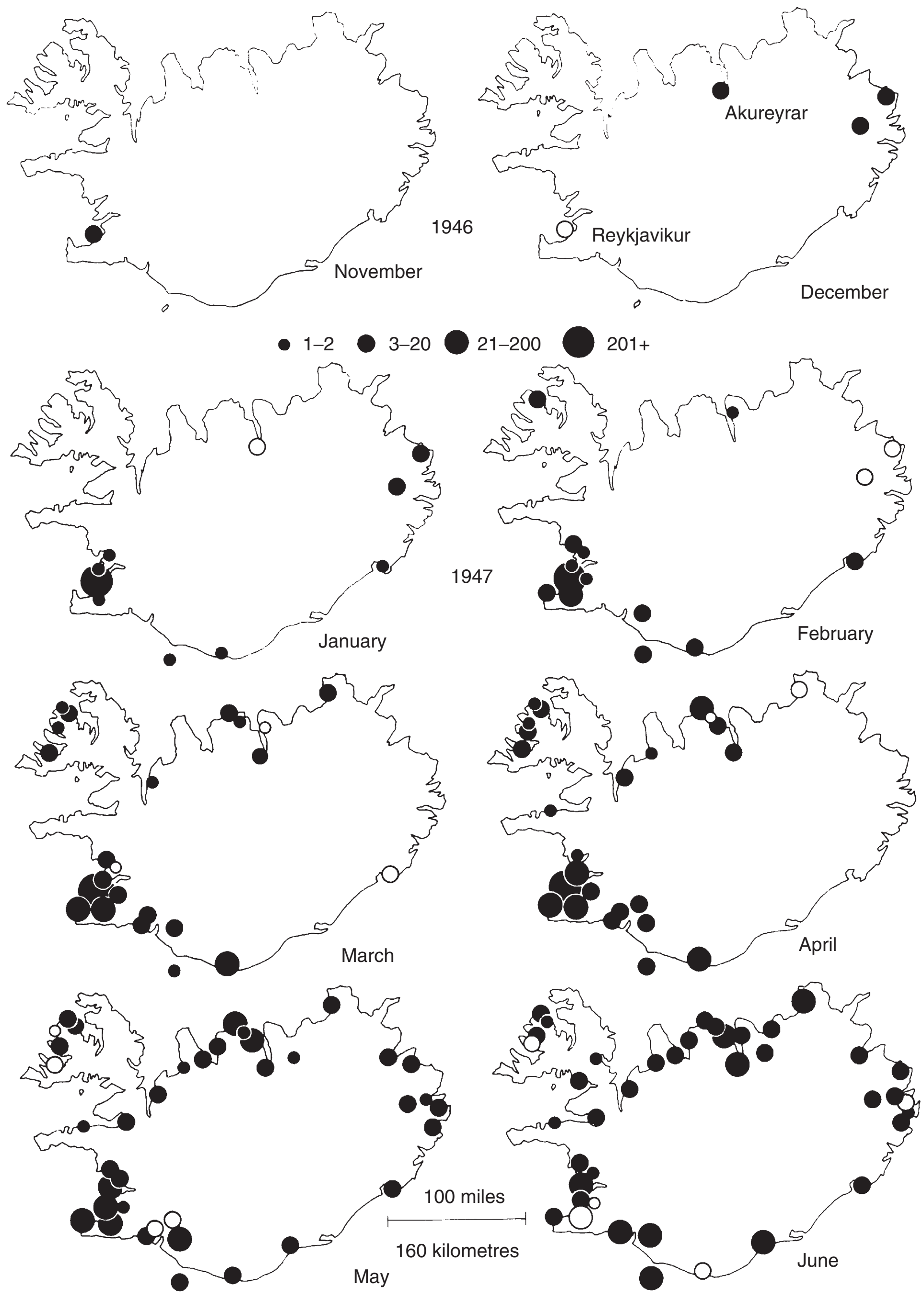

Figure 3 The spread of a measles epidemic across Iceland month by month (1946 47). From Cliff, A. D., Haggett, P., Ord, J. K. and Versey, G. R. (1981). *Spatial Diffusion. An Historical Analysis of Epidemics in an Island Community*. Cambridge: Cambridge University Press.

consequences. However, smoking, obesity, etc. are actually not diseases *per se* but rather health hazards.

Several infectious diseases had been perceived as new when they had been discovered in human population but had probably been circulating in animal populations (reservoirs). Thus, AIDS, now a major public health concern, was perceived as a new disease in the 1980s. Later examples include the 'severe acute respiratory syndrome' (SARS), the West Nile virus, and Ebola.

While infectious and parasitic diseases are often successfully combated, degenerative – often 'man made' – disorders are gaining ground as a consequence of altered physical environments, pollution, and occupational hazards as well as new lifestyles and disruption of social networks. It has been suggested that the epidemiological transition leads to a situation where patterns of mortality by causes differ significantly between countries that have reached the same stage in terms of general morality. Such geographical differences might reflect underlying national and cultural influences but remain to be explored.

Changing Conditions for Old Diseases

The distribution and diffusion of infectious diseases is affected not only by (micro) biological changes, but also by social and structural changes in societies.

Since some organisms remain unchanged, infection and survival will yield future immunity, whereas mutations will prevent immunity. For example, influenza virus regularly appears in new strains, preventing lifelong immunity, while measles virus remains stable. Current health threats are mutations giving rise to new varieties and drug resistant forms of causal organisms, for example, methicillin resistant staphylococcus aureus (MRSA) and tuberculosis (TB).

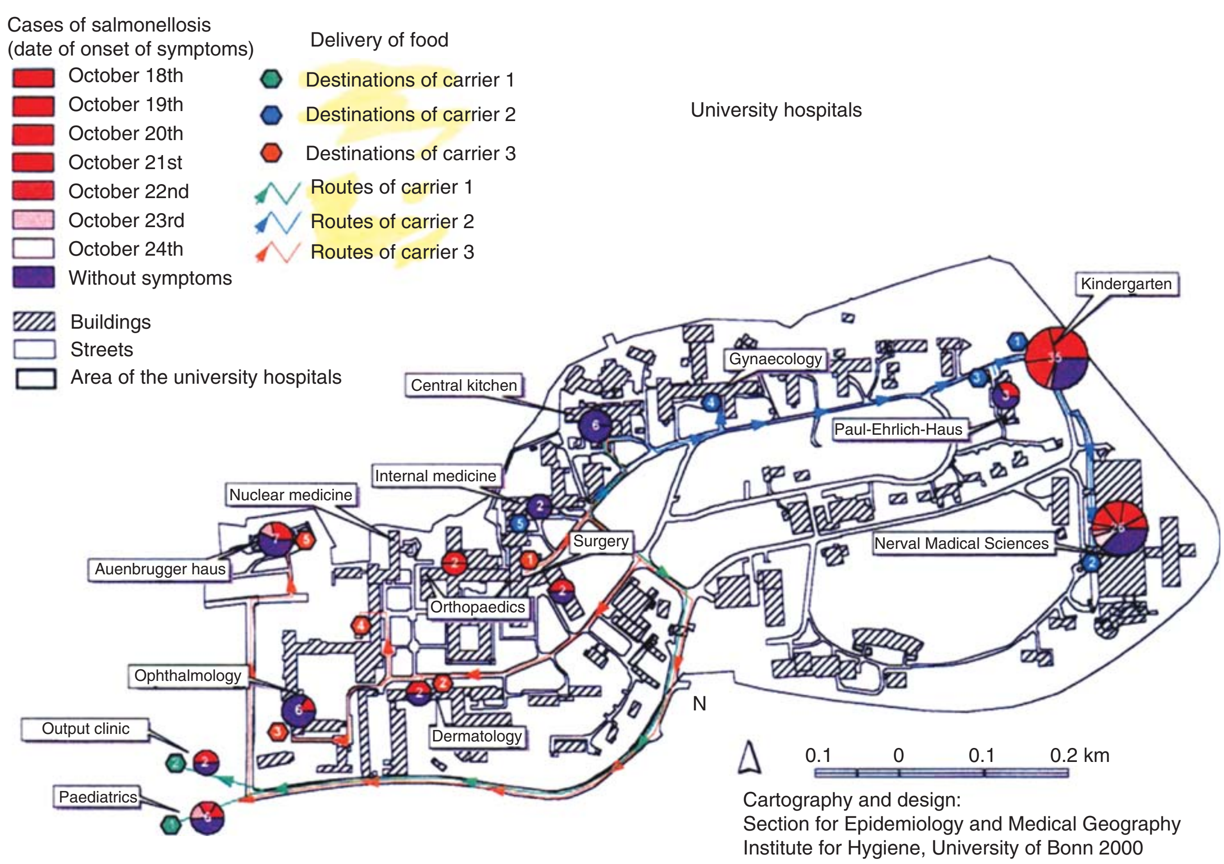

Figure 4 Salmonella diffusion in a local context. The diffusion of salmonella is a complex chain (network) of farming, food processing, and transportation, retail, and central kitchens. In October 1995 an outbreak of nosocomial salmonella was discovered at a university hospital, affecting nine buildings, including several clinics, a kindergarten and the central kitchen. Data about the infected staff, patients, and children, the internal logistics, and a retrospective analysis of meal preparation on the day of the outbreak were georeferenced and analyzed in a high-resolution GIS. The source of the salmonella diffusion was found in the kitchen, where vanilla pudding was prepared in the spatial vicinity of defrosting poultry and the pudding was probably contaminated by thawing water. From Kistemann, T. *et al.* (2000). GIS-supported investigation of a nosocomial salmonella outbreak. *International Journal for Hygiene and Environmental Health* 203, 117 126.

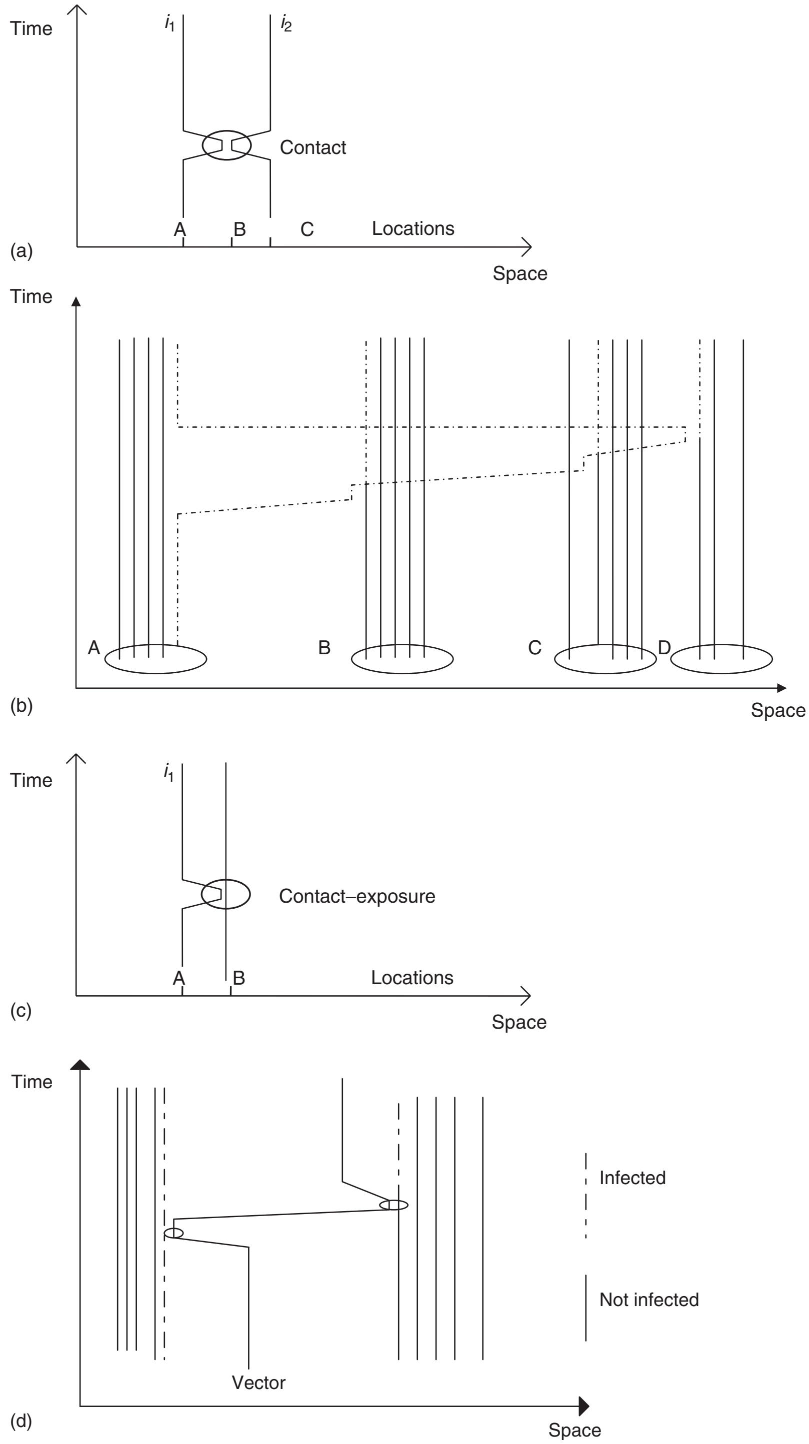

Figure 5 Transmission exposure diffusion: time geographic outlines. Transmission and diffusion happens on any scale from the individual (micro) to the global and often in complex chains. The figures are based on time-geographical visualization principles. The vertical lines in the figures represent life-paths (trajectories) of individuals. (a) In the most obvious kind of disease diffusion, pathogens are transmitted from one individual to another susceptible individual. One contact may be enough. (b) Sometimes, only one infected person is sufficient to introduce a communicable disease to another area with a virgin population and to start an epidemic. In fact, highly mobile individuals carrying a communicable disease may be in a position to spread the pathogen widely to one person after another and even one place after another. Healthy individuals are represented by full-drawn lines while persons carrying diseases are represented by dashed lines. (c) Short-term relocation and occasional exposure, say during a holiday, when susceptible individuals expose themselves to risks, like excessive sunlight, unfamiliar food, unfamiliar microorganisms, temporary sexual contacts, etc. (d) Diffusion by relocation of vector or artifact. Another type of transmission and diffusion occurs when pathogens are relocated from place to place, between spatially separate populations, without any direct human involvement but with a vector, for example, malaria transmitted by a mosquito. In a similar way, hepatitis and HIV can be transmitted if a syringe is first used by a carrier and then by a noninfected person. (e) The

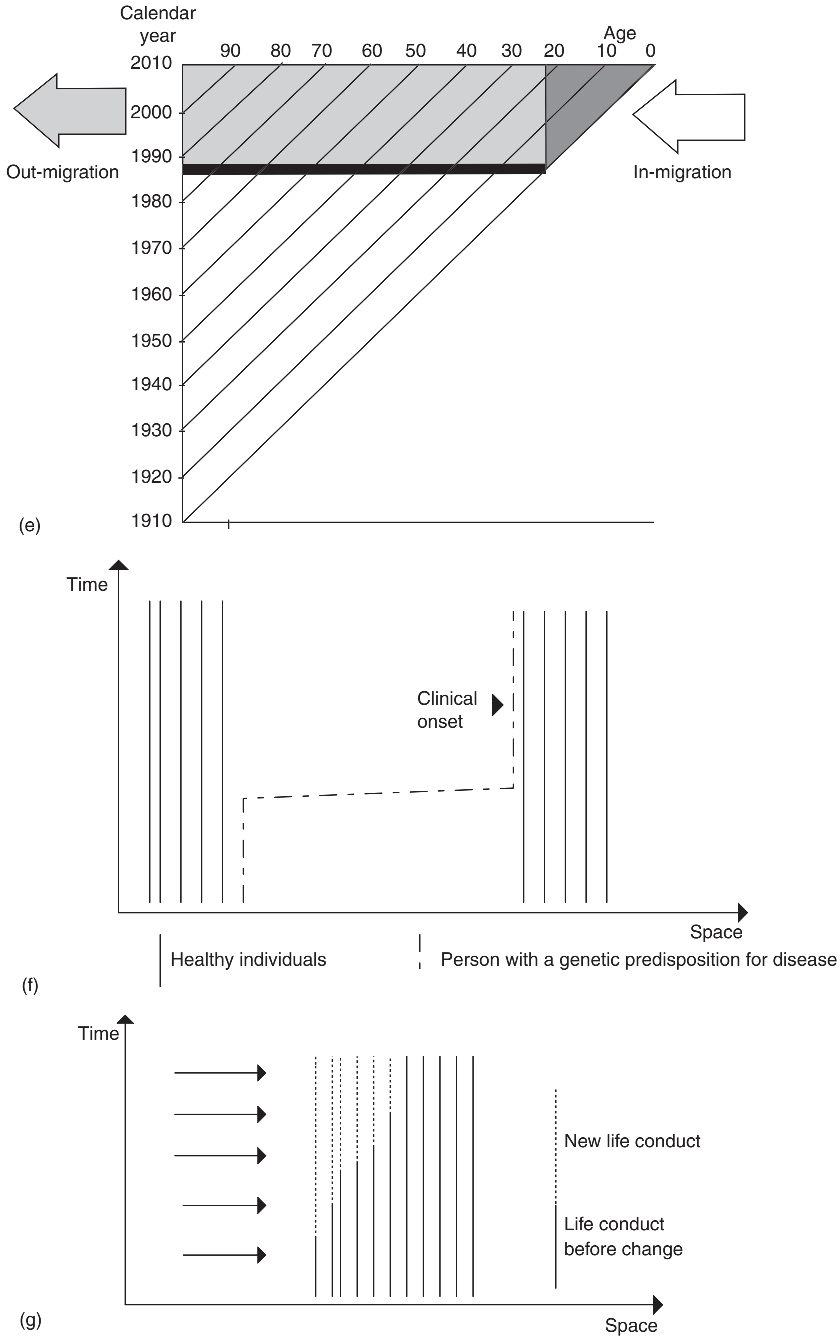

population in a certain place may be exposed to a sudden and temporary insult, such as radioactive fallout after a nuclear accident (thick black horizontal line). The cohorts living in the place at the time will probably carry the effects of the incident during the following years (dashed area). Children born in the place might be affected by lingering effects or genetic damages (shaded area). Out-migrants will bring relocate their health effects, while in-migrants from other places will be unaffected. Similarly, external factors may exert an insult on a whole population for any length of time or just a certain group, for example, schoolchildren or infants. Likewise the population or certain groups can be affected by health-promotion intervention. (f) Diffusion by relocation of predisposed individuals. Further, people may carry conditional 'predispositions' for diseases when they move. Predispositions are not communicable but congenital genetically inherited or individually inborn. This situation in turn opens two alternatives. The predisposed individual will develop the disease either in any case or only under a certain environmental influence. In diffusion terms, a relocation of the disease takes place (cf. **Figure 5d**). The diffusion may proceed further as a genetically inherited disorder eventually becomes predominant in an isolated population. Generally, large groups of migrants are needed for major changes of the gene frequency in a population but one migrant can be sufficient to spread a dominant, potentially pathogenic gene to another area and another population. (g) Pathogenic diffusion by innovation spread. Finally, just like innovations of any kind are diffused, lifestyle stimuli may reach a place and a population from outside. Subsequent health effects may produce the impression of disease diffusion or the reverse, that is, the disappearance of ailments. In fact, the epidemiological transition has largely been due to such changes.

Measles epidemics in Iceland illustrate the effects of changing conditions for diffusion. Measles is not endemic in Iceland, since the country is topographically isolated and its population is too small for the virus to remain in circulation. Thus, each epidemic is introduced from abroad, usually from Scandinavia or Britain. Mapping all epidemics of measles in Iceland from the 1870s has re vealed several changes of diffusion, probably reflecting structural modernization. Since 1945 measles epidemics have become more frequent and always started in the capital region, close to the international airport, whereas some previous epidemics started in harbor villages along the coast. The spread within the country became faster during the century, probably as a result of increasing urbanization, improved transportation, and a shift from an ambulatory school system in rural areas to fixed schools (**Figure 3**).

In the wake of increasing intercontinental mobility, infectious and vector borne diseases are returning to Europe, North America, and other parts of the world where they have been absent or confined for decades. TB and malaria provide examples of 'reemerging' diseases with different geographical characteristics. Globally, TB has never been overcome. According to World Health Organization (WHO), the current prevalence is some 14 million worldwide and some 1.6 million deaths per year can be ascribed to TB. Incidence and mortality are highest in the African region, while prevalence is highest in Southeast Asia. The 'return' of TB in Europe and the American continents is a consequence of mi gration, microbiological changes, and social conditions. TB now appears in drug resistant forms such as multi resistant (MDR TB) and extensively resistant (XDR TB) forms.

As for malaria, not only do tourists visiting malaria endemic areas return infected, but persons who have never been there have been diagnosed with so called 'airport malaria' after being bitten by odd mosquitoes that have been relocated by international flights and survived for some time at the destination on higher latitudes. Certain intercontinental flight routes have been identi fied as particularly risky and the majority of cases have been found in France, Belgium, and United Kingdom. This development can be put down to international mobility, but even nosocomial cases have been found, when patients have received transfusions of infected blood. Although the phenomenon is rare – since the first confirmed case was diagnosed in 1977, on average two cases annually have been recorded – it is a potential future risk. If global warming induces ecological changes, new habitats for malaria might emerge. An even more imminent hazard associated with intercontinental air transportation is the risk that malaria is re introduced in regions which are currently free but provide favorable conditions for the disease.

Tools for Understanding Disease Diffusion

Understanding the geographical diffusion of diseases is highly important for long term prediction and contain ment of diseases. Adequate surveillance is necessary for public health intervention in emergencies.

Many conventional maps of disease distribution are crude and inaccurate, but tools have been developed for monitoring, surveillance, simulation and prediction, based on better understanding of the spatial and tem poral behavior of diseases. Geographic information systems (GIS) can be very useful for surveillance, simulation, and analysis, provided that cases and risk factors can be registered with high spatial and temporal accuracy. GIS has also proved efficient for tracing the sources of infection in a micro scale context, such as salmonella diffusion in a hospital area (**Figure 4**). Re mote sensing can be used for surveillance of ecological conditions and habitats for disease agents and vectors. GIS supported simulation has been used to forecast the diffusion of HIV/AIDS and TB as well as the spread and health outcome of air pollution and radioactivity. With sufficiently accurate data, simulation can be applied on any level down to blocks and workplaces.

Early warning is important for national security in case of serious infectious epidemics, whether natural or deliberate, terrorist induced attacks. RODS (Real time outbreak surveillance) is a type of software which can be used for early detection of epidemics. If cases are re ported immediately to the system it can discover signs of epidemics which individual physicians are unable to see from their limited perspective.

Growing understanding that exposure in early life may have health effects in adult or advanced age has inspired the insight that similar exposures in different places and environments might lead to similar health outcome, in different locations. Therefore, a life course approach to epidemiology has been suggested to take into account such early exposures. Time geography, which offers a similar approach with a more explicit spatial dimension, has also been suggested as a method for detecting similar life paths and exposures of individuals who do not belong to spa tially contiguous populations. The time geography ap proach can also be used to visualize and clarify various epidemiological processes (**Figures 5(a)–5(g)**).

In spite of accumulated knowledge about disease diffusion, current epidemiological trends still call for methodological and theoretical development.

See also: Diffusion; Mental Health; Migration; Psychotherapy/Psychotherapeutic Geographies; Radical Geography; Simulation; Time and Historical Geography; Time Geographic Analysis.

Further Reading

Bentham, G. (1988). Migration and morbidity: Implications for geographical studies of disease. *Social Science and Medicine* 26(1), 49 54.

Cliff, A. D. and Haggett, P. (1988). *Atlas of Disease Distributions. Analytic Approaches to Epidemiological Data*. Oxford, UK and Cambridge, MA, USA: Blackwell.

Cliff, A. D., Haggett, P., Ord, J. K. and Versey, G. R. (1981). *Spatial Diffusion. An Historical Analysis of Epidemics in an Island Community*. Cambridge: Cambridge University Press.

Gatrell, A. (2002). *Geographies of Health*. Oxford, UK and Malden, MA, USA: Blackwell.

Gatrell, A. and Loytonen, M. (eds.) (1998). *GIS and Health. GISDATA 6*. Taylor & Francis.

Haggett, P. (2001). The global burden of disease. In Haggett, P. (ed.) *Geography A Global Synthesis*, ch. 20. Harlow: Prentice Hall.

Kawachi, I. and Wamala, S. (2007). *Globalization and Health*. Oxford, UK: Oxford University Press.

Kistemann, T., Dangendorf, F., Krizek, L., Sahl, H G., Engelhart, S. and Exner, M. (2000). GIS supported investigation of a nosocomial salmonella outbreak. *International Journal for Hygiene and Environmental Health* 203, 117 126.

Krieger, N. (2001). Theories for social epidemiology in the 21st century: An ecosocial perspective. *International Journal of Epidemiology* 30, 668 677.

Meade, M. and Earickson, R. J. (2000). *Medical Geography* (2nd edn.). New York and London: Guilford.

Selinus, O., Alloway, B., Centeno, J .A. *et al.* (eds.) (2005). *Essentials of Medical Geology*. Burlington, MA; San Diego, CA; London: Elsevier.

Thomas, R. (1992). *Geomedical Systems. Intervention and Control*. London and New York: Routledge.

Relevant Websites

http://www.cdc.gov
Centers for Disease Control and Prevention.

http://www.ecdc.europa.eu
European Centre for Disease Prevention and Control.

http://www.who.int
World Health Organization.

Disease Mapping

T. Koch, University of British Columbia, Vancouver, BC, Canada

Maps as a Medium for Medical Science

Nobody maps health. We have no atlas of healthy persons; we make few maps of nonsick, mundane folk. But since the seventeenth century, there have been many thousands of maps (and map collections) of disease incidence in local, regional, national, and international populations. Just as marine cartographers have always charted shoals and rocks that impeded harbors, students of disease have used the map to identify the location of this or that disease in order to study its possible cause or to devise better treatment plans. It is not that wellness lacks a geography that can be located in the two dimensional (2 D) plane of the map but that, compared to chronic illness and epidemic disease, the healthy subject lacks urgency, and thus interest. To map health is to focus on the absence of illness or disease, to focus on a state that, most believe, needs little attention.

Mapping the single, rare case of an unusual disease serves no purpose. Mapping an afflicted cohort in a community defines a local health problem and, ideally, contributes to its resolution. Mapping a group of cases over time often reveals environmental characteristics that promote this or that disease in place, resulting in not only simply knowledge of a specific condition but also suggesting a potential for its interdiction. In the map, lists of mortality and morbidity are transformed into classes of incidence that argue a density of occurrence whose location provides at once an argument for care and a site for research.

This is what maps do: they distill data (e.g., mortality reports) into event classes to argue a proposition about the nature of a disease in place and within a population. There can be monsters there, the old marine charts said. Look out. In medicine, the monsters 'are' disease symptoms, either the frightful, physical manifestations of epidemics that decimate populations or the chronic conditions that limit the lives of individuals. Central to the enterprise of understanding this or that individual disease, and disease processes in general, has been the map whose technologies of production and distribution both defined and reflected the increasing arsenal of technologies that have permitted the development of ever more precise approaches to disease through progressive attempts to formulate a medical science that can understand both their cause and effect.

While medical mapping includes a wide array of projects, from the location of hospital service areas to the creation of catchments for disease reportage, these administrative approaches are inherently responsive, resulting from a theory of disease and how best to develop responses to disease incidence. Here, the focus is on the construction of disease theories through maps and the method in which those theories have been first inscribed and then tested in disease maps.

Disease studies are a fundamentally ecological enterprise, an attempt to understand the nature of the monsters that affect our communities through a study of the environments that promote monster growth or inhibit monster presence. To slay the dragon, or at least hold it at bay, demands we craft zones of safety where danger, and death, are previously held dominion. This in turn requires that we understand disease states as a shared domain in which the microbial agent and the host are conjoined in place, the environmental elements promoting their union identified. From the start, scientific interest reflected in cartographic studies is not the rare case but rather in the disease agent that affects us generally, in groups.

Yellow Fever: Valentine Seaman

An early example of this disease mapping is Valentine Seaman's late eighteenth century mapping of a yellow fever (see **Figure 1**) outbreak in the New Slip area of New York City. Among the most feared diseases of that century, in the 1890s yellow fever decimated the burgeoning coastal American cities of Philadelphia, New York, and Boston. Some believed the disease was imported in trading ships from tropical areas where yellow fever was known to exist. Were this true, the logical response would be quarantines that constrained shipping and trade. Others, however, believed yellow fever arose from the foul odors of the human and animal wastes that accumulated in the unsanitary, nineteenth century city. Were this the case, no restraint of trade would be necessary and the best response would be in developing urban sanitation.

To test the later theory Seaman first mapped the location of cases of yellow fever occurring in the New Slip area. Numbering the cases, in the map, Seaman distinguished between fatal and nonfatal cases with different symbols. In his writing, he lamented the paucity of cases (how did he show 'mild' as opposed to 'severe' cases in the map?) and the difficulty of including all cases in his map.

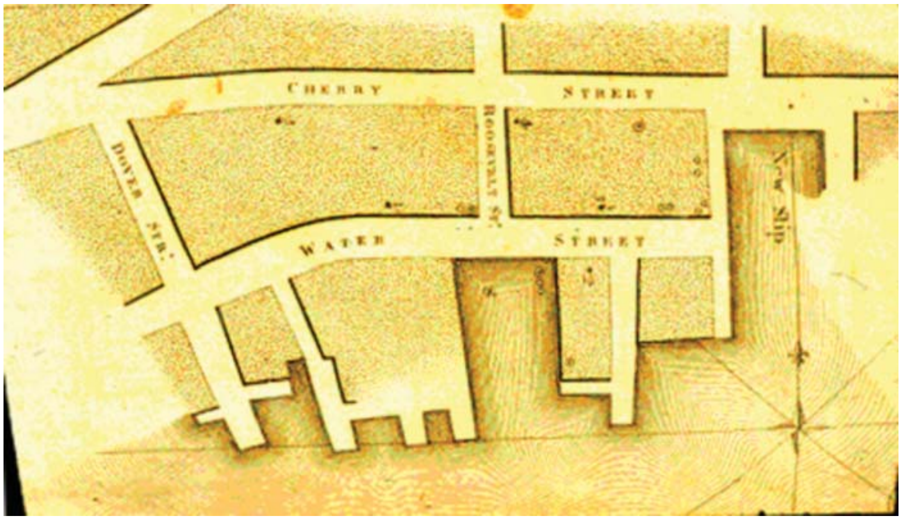

Figure 1 Valentine Seaman's map of yellow fever in New York City, 1989 sought to argue the disease was generated in urban waste sites. Image courtesy: the National Library of Medicine.

Still, the result was a map that showed a concentration of cases in the Water Street area. In a second map, Seaman then located local odiferous waste sites that he believed gave birth to the disease. Comparing the two maps Seaman saw congruence between the greatest concentration of deaths and the location of the malodorous sites. Proximity, he reasoned, implied causality.

Seaman instantiated in the map what Wood and Fels would describe as a logical, 'if–then' argument that is the conceptual heart of all mapping. First, he created from mortality reports a class of events, 'yellow fever' in which the individual cases together created a range of disease. "This," the map said, "is here, on the mapped streets. We know it." In a separate map Seaman created a second event class composed of local waste sites whose odor he perceived and located in the copperplate map. Considering both maps together he then argued that 'if' yellow fever originated in the foul odors of urban wastes 'then' this should result in congruence between the area of yellow fever cases and of odiferous sites. The congruity that resulted proved, Seaman believed, the yellow fever outbreak originated in the waste sites of the neighborhood.

Seaman was right. Proximity 'did' suggest causality although the process, we now know, was more complex than he assumed. Seaman also reported that the mosquito population was extremely heavy in the area of waste sites where cases of yellow fever were strongest. These were, for him, extraneous nuisances and not critical vectors of disease diffusion. It was only at the end of the nineteenth century that disease researchers identified mosquitoes as the yellow fever vector. It was through their bite, rather than the odor of the areas where they swarmed, that the disease spread.

Mapped arguments in medicine typically have this 'if–then' form to propose one of three distinct but related forms: taxonomic, topographic, and statistical. Taxonomies define diseases by location, fusing areas of incidence with either geographic or political boundaries to create a catalog of 'disease states' based on location. The logic of the map that resulted permitted areas of unusual disease occurrence to be identified at typically coarse scales of analysis in a manner permitting a more detailed, topographic analysis *in situ*. Finke's 1792 *Practical Medical Geography*, with its now lost 'Nosological map of the world', for example, attempted to argue a taxonomy of then known diseases in Western Europe as well as the native plants and resources that might serve to treat them.

In so doing Finke transposed the evolving political geography of Europe into a medical geography in which the boundaries of states became the borders of disease areas. Inherent in Fink's approach was the proposition, argued graphically, that disease is an outcome of local environmental conditions for which remedies might be found within areas of greatest incidence. Where diseases arouse, then, nature would provide a natural remedy that might serve to protect local populations.

Taxonomies and Topographies

Mapped taxonomies were precursors to nineteenth century topographies that sought to establish a relationship between the incidences of disease and suspected causal factors that promoted the occurrence of disease in place. It was through mapping that the idea of environmental disease agents within areas of incidence was first and best explored. A. K. Johnston's 1856 map of 'Health and disease in connection with chiefly natural phenomena', published in his *The Physical Atlas of Natural Phenomena*, is a good example of the transformation of

Figure 2 Johnson's world map of diseases also argued the relation between areas of similar disease and their potential transmission along trade routes. Here, yellow fever is shown to occur in the Caribbean and southern United States along general trade routes, marked in red. From Koch, T. (2005). *Cartographies of Disease: Maps, Mapping, and Medicine*. Redlands, CA: ESRI Press. Courtesy: National Library of Medicine.

taxonomic to topographic mapping. In it the incidence of disease is lodged in nations or national regions where different diseases were reported by travelers and local authorities. At one level, the map explored a traditional disease taxonomy that perceived 'tropical' versus 'temperate' diseases, conditions common to lowland as opposed to Alpine airs. The result was a global landscape of disease districts (the province of goiter, the state of catarrh, and the city of leprosy) whose effect was summarized in 'Value of life' charts on the map's lower right side comparing life quality and disease incidence for specific populations.

The ideas embedded in the mapping were twofold. First, that a disease taxonomy could be spatially presented in a manner permitting the identification of specific geographies where one or another condition dominated the landscape. Second, that the effect of a disease in place could be quantified precisely in terms of life quality and life years. The greater the local incidence of a disease, the less likely it is that life quality will be maintained.

Johnston's map also presented a 'topographic' argument proposing a potentially causal relationship between disease incidences among different regions. Lines along navigation routes joining the Caribbean to the southeastern United States show the apparent progress of yellow fever from south to north (**Figure 2**). Lines from Europe to the Americas tracked the apparent progress of cholera from England and Europe to North America. The result made the argument, then debated, that cholera traveled from 'here' to 'there', along the land and sea routes that defined the emerging globalization of commerce and migration.

Topographies of Disease

To investigate the nature of an individual disease and its causes typically required a finer resolution and a more focused scale of address. Such studies in the nineteenth century proposed a relatively simple topographic relationship between intensity of disease in a metropolitan region or city district and local factors that were assumed to promote its occurrence. These studies, relying on increasingly comprehensive and broadly available public health statistics, related disease incidence in an area to climatic, geographic, and socioeconomic. In the first decades of the nineteenth century, such mapped propositions increasingly employed a broad, urban scale of address.

In 1849, for example, William Grainger proposed a geographic relationship in which proximity to the low lying, central Thames River correlated positively with greater cholera incidence (**Figure 3**). In his density map cholera mortality reported in local registration districts was aggregated to argue an inverse relationship between cholera mortality and altitude above sea level. The result defined cholera as an area of lowland, river areas where bad airs were typically found. At higher altitudes, the air was believed to be healthier, and thus to serve as a barrier to disease incidence. Where small outbreaks occurred at higher altitudes Granger mapped potentially complicit wells, cesspools, or swamps whose odor polluted the air and promoted disease.

At a very different scale, John Snow famously created a map of a cholera outbreak (see **Figure 4**) in his St. James neighborhood in 1855 to argue that cholera was in fact waterborne rather than airborne. In mapping what he called 'the most terrible outbreak of cholera which ever occurred in this kingdom', Snow identified 578 cases of cholera occurring in August and September of 1854. These deaths, each symbolized by a bar, were mapped on the streetscape with the principal water sources that served the neighborhood district. "It might be noticed," Snow wrote, "that the deaths are more numerous near to the pump in Broad Street." The density of cases proximate to Broad Street, he argued, insisted upon that pump as a source of the outbreak.

Like Seaman and Grainger before him, Snow created from mortality reports two classes of events – one of mortality and the other of suspected disease agents. Like them, he argued, the proximity of disease incidence to mapped environmental elements, in this case a water source, demonstrates a causal theory of a disease. In making this argument Snow and his predecessors relied on what today is known as Tobler's law, or the 'first law of geography': "everything is related to everything else, but near things are more related than distant things." In applying this principle Snow had to account both for areas near the pump that seemed cholera free, for example, at local workhouse, and those areas where cholera occurred in areas nearer to other water sources. He did this through interviews with local residents, pursing a 'shoe leather epidemiology' of local interviews still employed today.

Modern Dot Maps

Some argue that all this is old hat and that nineteenth century methodologies do not serve the twenty first century needs. Maps may serve as illustrations, representations of statistical analytics, but not as primary tools of disease exploration. And yet, in local, regional, and national health offices, hundreds of diseases are mapped

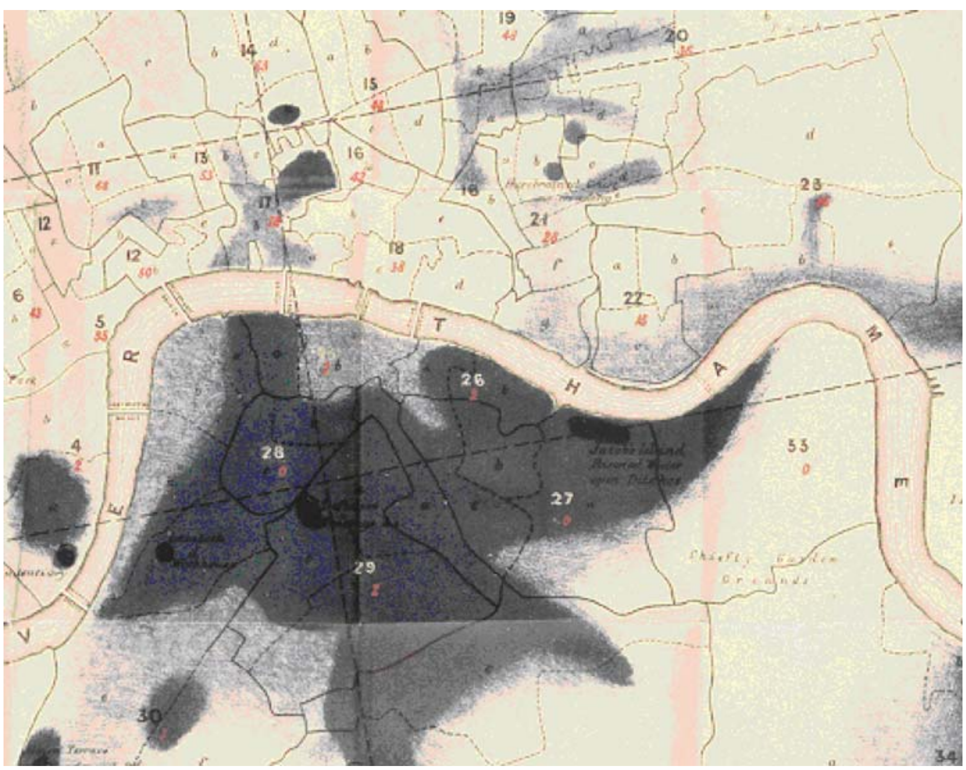

Figure 3 William Grainger mapped the relative intensity of cholera across London registration districts. Red numbers represent altitude above sea level while black numbers and letters identify individual districts. From Koch, T. (2005). *Cartographies of Disease: Maps, Mapping, and Medicine*. Redlands, CA: ESRI Press.

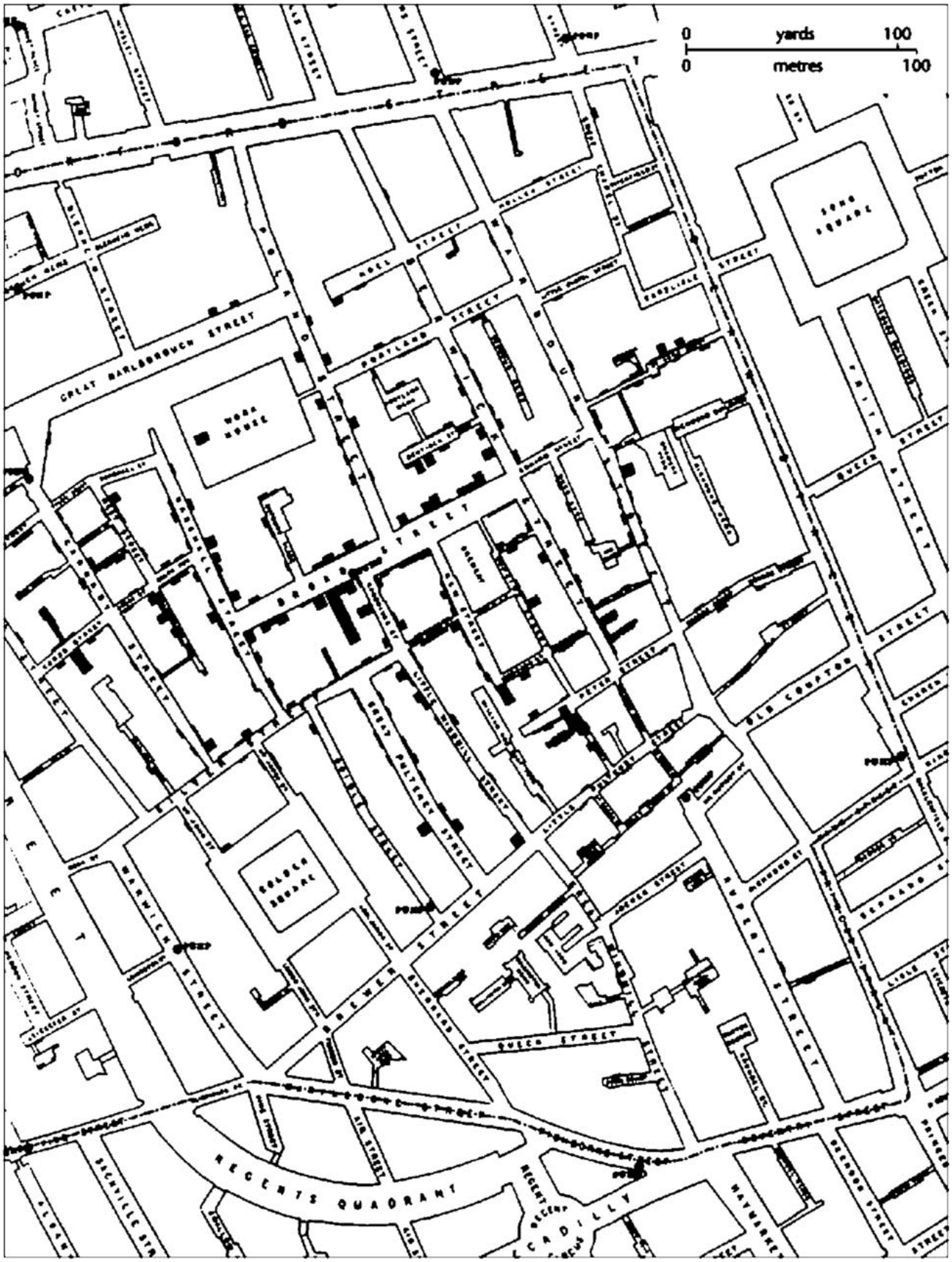

Figure 4 This section of John Snow's map of the St. James, Westminster cholera outbreak of 1854 in London demonstrates his central argument of a relation between cholera density and the water pump at the apparent center of the outbreak. From Koch, T. (2005). *Cartographies of Disease: Maps, Mapping, and Medicine*. Redlands, CA: ESRI Press. Courtesy: National Library of Medicine.

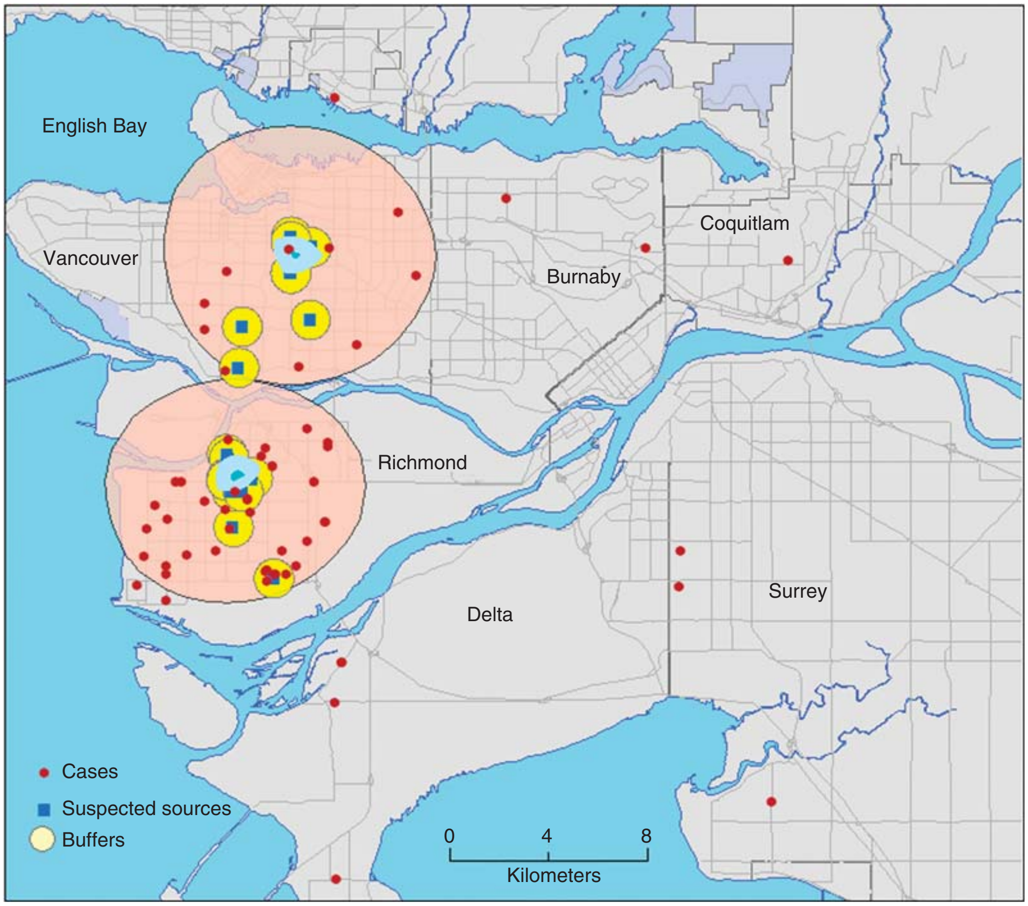

Figure 5 A series of interlocking buffers centered on suspected food sources was used to identify likely sources of a salmonella outbreak in Greater Vancouver in 2001. For the sake of clarity the number of suspected sources (blue squares) and cases (red dots) are reduced in this map. Source: Author.

to promote an understanding of their occurrence and, in some cases, the pattern of their diffusion. The methods employed have refined the nineteenth century approach but retain the basic spatial argument inherent in disease maps by his time. For example, in 2001, a brief outbreak of food poisoning affected persons in Greater Vancouver, Canada (**Figure 5**). The assumption was that the out break was a form of food poisoning but there were concerns that work on local sewers might somehow have contaminated the area's water supply. More than 60 persons were diagnosed during the outbreak and, epi demiologists learned, those persons had purchased food at any of 24 stores and restaurants.

To demonstrate contemporary disease mapping pro cedures the location of the reported cases and, separately, the food outlets where patients had eaten prior to the onset of their symptoms were mapped. A series of hier archical buffers were then used to develop a weighted center to the outbreak, narrowing the likely sources down to two or three food outlets located at the epicenter of the outbreak. A major highway serving the Richmond–Vancouver corridor bisected both buffers, suggesting a possibly portable source of the outbreak. Of the three food outlets at the center of the southern buffer was one company providing a product sold on site and to other restaurants in the area. Later testing by bacteriologists showed this was indeed the source of the outbreak.

Disease as Mapped Statistic

With the migration of mapping and statistics from paper to computer media a number of changes have occurred in both the analysis of disease and the distribution of dis ease related studies. With off the shelf programs, those who previously would have relied on professional map makers may now themselves use the mapped environ ment to store data on disease incidence and location, on the one hand, and manipulate that data electronically, on the other. Simultaneously, computer advances have en abled a series of progressively sophisticated statistical programs that permit increasingly robust, and certainly more rigorous, data manipulation than was possible even 30 years ago. Graphic and statistical media are not distinct, but related, with maps more and more fre quently serving as a medium for data manipulation, storage, and the presentation of results.

Computerized mapping technologies and the digita lization of data have enabled a vastly more complex approach to disease studies at every scale. Cancer inci dence, for example, appears to be causally related to a complex of behavioral, environmental, and genetic fac tors. Where behavioral and especially environmental conditions exist, these may be related to socioeconomic factors determining, for example, housing location. What is required is a methodology that identifies areas of greater disease incidence and then considers potential disease contributes *in situ*.

One difference between twentieth century and early nineteenth century studies is that contemporary mapped studies typically consider not simply incidence counts but incidence as a function of population. By the mid nineteenth century, it was common to express epidemic disease as a function of cases (and more specifically deaths) per 10 000 or 100 000 persons. In exploring dis eases in modern populations this approach has become the norm. In considering multiple risk factors it has also become common to use increasingly sophisticated stat istical measures in the mapping to 'smooth' data across a broadly mapped, geographic region and to argue probabilities of relations between suspected variables and identified clusters on the basis of the result.

Mapping of breast cancer is an example of the com plexity of the resulting work (**Figure 6**). Several ideas are inscribed in the accompanying map of breast cancer in cidence in the USA. At one scale the incidence of disease is proposed as a national issue in which variable disease rates in US county populations argue for the focus of investigations in those county areas with disease inci dence. Second, the map distinguishes breast cancer from all other cancers as a unique disease rather than one of a complex of potentially related cancer types. Variations in incidence across the map argue for specific studies of environmental and social factors in areas of greatest relative incidence. It was in part this type of study, and an analysis of the results, that led to the now famous Long Island Breast Cancer study, an intense investigation of a single cancer concentration. While the analysis was statistical, seeking positive correlations at a significant level, it was also cartographic: the work was carried out in the medium of the map and its electronic database in which data were stored and results presented.

While the enabling technologies permit a graphic and statistical rigor unimagined by Seaman, Grainger, or Snow, the basic thinking remains the same. By identifying areas where disease occurrence is greatest, one can focus

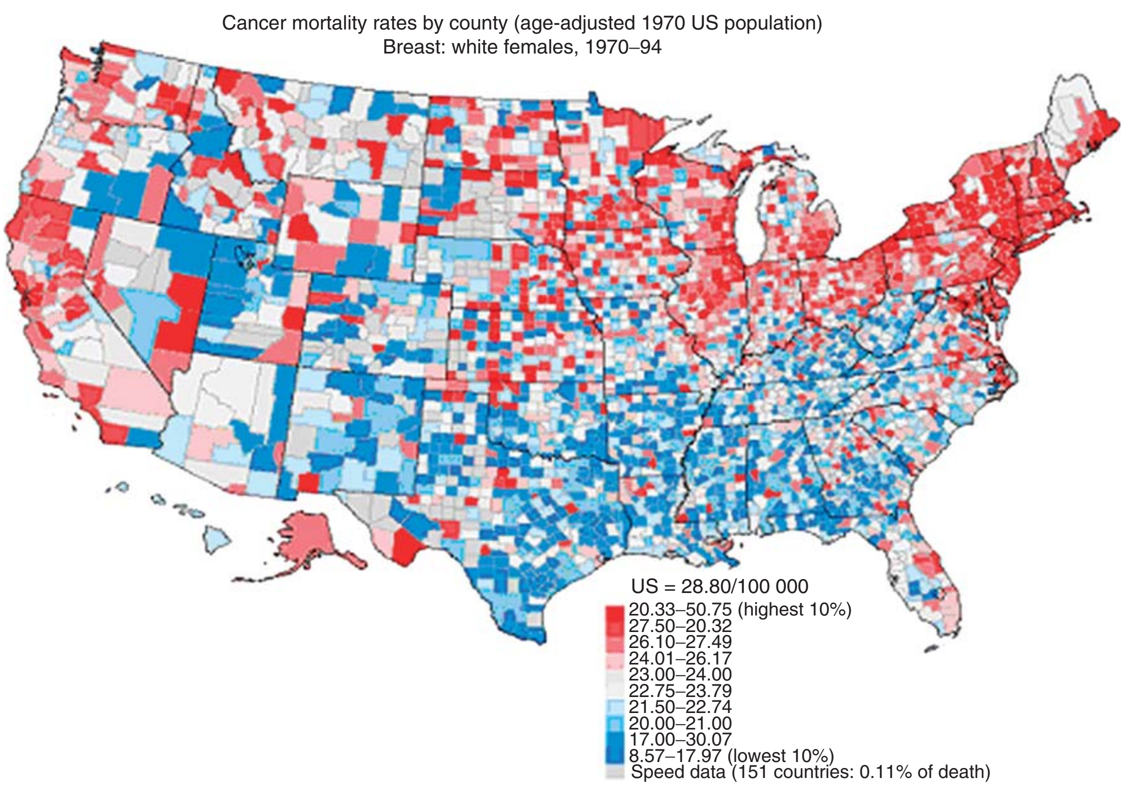

Figure 6 Variations in rates of breast cancer within US county populations argued for intense investigations of potentially causative factors at a finer scale. From National Cancer Institute, National Cancer Atlas.

on environmental elements that may be generative or, at least, may contribute to disease incidence. Having mapped those areas one then identifies their proximity to the potential contributors and seeks to confirm or reject the argument of a causal relation.

Conclusion

To the extent that the mapped argument becomes generally accepted, it becomes knowledge that could be applied by those in society whose responsibility is the social weal. From its earliest beginnings medical mapping has served both a scientific and a public function. For those involved in the study of specific diseases, maps of incidence and of possible causal factors have served to test propositions about the nature of a disease and the reasons for its incidence. Where casual factors are identified they are used to argue for the public containment or elimination of those factors. 'If' we can demonstrate a concentration of breast cancer cases in a neighborhood or region, 'then' we seek to assure that both treatment and diagnostic centers are available in those areas. We locate the area of best location in the map. 'If' we can demonstrate a community at risk from, say, influenza, 'then' we locate flu vaccination centers at the locations that are most likely to serve the largest possible number of persons at risk. 'Where' we find the greatest number of automobile crashes on a city street, we then may put in a traffic light or add an ambulance to the local district to assure faster medical treatment.

The nature of such mapped propositions is of the same type as those used for the study of disease incidence itself. These spatial propositions are the heart of the science of disease studies and of administrative responses to accident and illness. From both perspectives, the map is a critical medium in which the experience of disease is transformed into a study that develops spatially grounded data in an applicable fashion.

See also: Cartography, History of; Chronic Disease; Disease Diffusion; Mapping, Philosophy; Maps; Medical Geography.

Further Reading

Adamski, R. K. and Brunton, H. (2003). Developing and maintaining an enterprise GIS: An in depth look at the LI GIS. International Medical Geographers Symposium July 7.
Johnston, A. K. (1856). *The Physical Atlas of Natural Phenomena*. Edinburgh: W. Blackwood.
Koch, T. (2005). *Cartographies of Disease: Maps, Mapping, and Medicine*. Redlands, CA: ESRI Press.
Koch, T. and Denike, K. (2004). Medical mapping: The revolution in teaching and using maps for the analysis of medical issues. *Journal of Geography* 103(2), 76 85.
Koch, T. and Denike, K. (2007). Aaron's solution, instructor's problem: Teaching surface analysis using GIS. *Journal of Geography* 106(2), 69 77.
Pickle, L., Mungiole, M., Jones, G. K. and White, A. A. (1999). *Atlas of United States Mortality*. Hyattsville, MD: Centers for Disease Control and Prevention, National Center for Health Statistics.
Seaman, V. (1798). Inquiry into the cause of the prevalence of yellow fever in New York. *Medical Repository* 1(3), 314 332.
Snow, J. (1855). *On the Mode of Transmission of Cholera* (2nd edn.). New York: The Commonwealth Fund.
Wood, D. (1992). *The Power of Maps*. New York: Guilford Press.
Wood, D. and Fels, J. (2008). *The Natures of Maps*. Chicago, IL: University of Chicago Press.
Wood, D., Kaiser, W. L. and Abramms, B. (2005). *Seeing through Maps: Many Ways to See the World* (2nd edn.). Amherst, MA: ODT.

Distance

G. H. Pirie, University of the Western Cape, Cape Town, South Africa

Glossary

Distanciation The process of distancing or separating places, people, or things from one another.

Introduction

Distance is one of the key elements of the world. Dis tances on our planet are in some senses definitive: they are the way we get to know it, speak about it, associate with it, and navigate our way around it. The Earth's size is measured not by mass but by circumference, con ventionally at the equator. Referring to distances across continents and oceans, and between cities, is a popular way of registering where we live and work, how close or how far we are from other places, and how big an undertaking some journey may be. Place and distance are mutually constitutive. The distance we are from events elsewhere is often a measure of how seriously we take them; places that are out of sight are often out of mind. The distance that people traverse is an indicator of their mobility; the distance that places are from us is a clue to their accessibility.

The (f)actuality of distance is undeniable, yet as this chapter argues, distance is also an idea. Personal ex periences of distance are varied. The diverse ways in which distance has been used, worked with, and thought about in human geography also indicate that it is not a simple, one dimensional phenomenon. Like other terms that we take for granted and seldom interrogate, distance refers to an elusive phenomenon. It is more or less visible and it has a concrete identity, but it is also an abstract notion whose nature, dimensions, and meaning are dif ficult to pin down.

Arguably, the concept of distance is one of the key building blocks of human geography. Yet, whereas many people would readily associate geography with contin ents, countries, and hemispheres, with ocean currents, national capitals, and mountain ranges, few would quickly identify distance as geography's stock in trade. The observation suggests that distance does indeed have an abstract quality. Professional geographers are more likely than members of the public to nominate distance as a core concept: geographers would regard distance as complementing other fundamentals such as scale, place, region, and boundary.

Starting a half century ago, several geographers have argued that distance is one of geography's basic elements. In his inaugural address as Professor of Geography at the University of Edinburgh in 1954, J Wreford Watson went so far as to assert that geography was a discipline about distance. His presentation actually made a better case for the notions of extent, reach, spread, and distribution being at the heart of the discipline. Distance, Watson suggested, was a useful surrogate indicator of the extent to which objects have adapted to or dominated the en vironment. In geography, he noted, the emphasis on distance was an emphasis on extent. In Watson's view, geography mapped humanity's success or failure to ex tend ideas, ways, numbers, and control. In the final an alysis, he wrote, geographical distances were the distances of the human spirit.

In the age old conception of geography as a discipline that deals with regional diversity and dissimilarity (even uniqueness), distance presented itself as something that could account for difference. The argument that distance was a central plank in human geography received a boost in the 1960s when a new scientific geography was being formulated and championed. At a time when young geographers were eagerly injecting scientific rigor into their work, the standardized and mathematically meas urable properties of distance were very attractive. In the hunt for law like patterns and principles of spatial or ganization, quantifiable distance was an entity that re search scientists could use to explain similarity and difference. Several American geographers working in the new spatial paradigm stressed the importance of distance to their subject. William Bunge's 1962 book on theoretical geography claimed that nearness was the cardinal prob lem in geography. In 1968 John Nystuen reasoned that distance and separation were fundamental geographical properties. The following year Waldo Tobler argued that the first law in geography was that nearby places and phenomena were more similar than those far apart.

Emphasis on the centrality of distance to human geography faded after the 1960s. However, as if con firming that geographically informed conversation, an alysis, and speculation is impossible without some notion of distance, the d word continues to thread through all geographical discourse. This does not amount to a field of study. The term distance seldom appears in any keyword or subject search among geographical research papers. Even dictionaries of human geography take the view that the definition of distance is self evident.

The subject areas in which distance remains a defining topic are at a scale quite different to that at which geography operates. Astronomy and space science are at the one extreme. These disciplines deal with the greatest conceivable physical distances. Estimating, measuring, and assessing the significance of the vast spaces between planets, stars, and even galaxies remain a major scientific challenge. At the other extreme are the miniscule intervals that fascinate and bewilder engineers and physicists. At a rather larger scale are the culturally variable interpersonal distances tolerated during personal interaction such as greeting, sitting, standing, and talking. These psychosocial distances are the subject of proxemics rather than human geography.

This chapter is about distances that are part of everyday human public experience – these distances define and occur in the public rather than the private spaces that human beings occupy, confront, and move across. In human geography the distances of interest are those between places on Earth, and those that involve more than a few steps of travel in outdoor spaces. Put another way, geographers' distance currencies are miles and kilometers rather than stellar parsecs, distances to do with eye contact, speech audibility or human intimacy, or micro distances ranging from hair's breadth microns to yoktometers. Geographical distance is also generally horizontal as opposed to the centimeters and meters (or inches and feet) that measure the height, width, depth, length, and breadth of visible objects.

Discourses of Distance

The many literal and figurative references to distance in spoken and written language confirm the hold that it has had on everyday life. Among the explicit references to distance in English are phrases to do with stamina and additional effort. The coinage 'go the distance' refers to completing a metaphorical journey: boxers who fail to go the distance have been knocked out, counted out, or thrown in the towel. To 'go the extra distance' means doing more than is necessary or expected. Here the reference is often to a particular unit measure of distance, as in 'go the extra mile'. A place said to be miles away means it simply feels far. To say that a miss is as good as a mile indicates that the magnitude of shortfalls or gaps is irrelevant. Judging something miles better than another is a colloquialism that has nothing to do with actual distance. Similarly, running a mile means taking steps to avoid a personal pest. To speak of a proposal not having much mileage in it refers to the absence or presence of advantage; conversely, getting a lot of mileage out of something means putting something to good use. The persistence of these mile based phrases reflects the tenacious link between the English language and imperial distance measures.

The admonition 'keep your distance' is a recommendation or warning not to get unduly or inappropriately close to some person or event. Close parallels are 'stay away', 'know your place', and 'respectful distance' – phrases loaded with geographical innuendo. No precise distance is specified; it is assumed that the 'right' distance is known intuitively or (can be learned) experimentally. Animal growls and hisses are warnings about getting too close; humans have their own repertoire of gestures about crossing an imperceptible line and invading personal space.

Indirect references to distance in everyday conversation are possibly more plentiful than direct references. At least in English, notions of distance are wound into a rich spatial vocabulary. At its very simplest, distance is implied in some unspecified distinction between 'here' and 'there'. In its idiomatic use, the phrase 'neither here nor there' is corrupted to mean indifference or that something (often a choice) does not matter. Taken literally the phrase implies that distance does not differentiate and is of no consequence; locating at one point rather than another is of no moment.

Some notion of distance is also contained in a range of English nouns such as close, near, far, remote, and horizon – and in modern colloquialisms such as 'close in on'. To zoom in and zoom out are contemporary ways of speaking about telescoping, or artificially bringing close what is actually far, and vice versa. Metaphorical phrasings to do with nearness and remoteness obviously trade on notions of distance. 'So near but so far' is an ironic saying that parallels (and universalizes) 'a miss is as good as a mile'. The reference to missing a target – being wide of the mark – works on encouraging people to finalize tasks and work with precision. To stop short is to fail. These expressions resonate with geographical references to journeys, destinations, and distances. In these constructions of success and fulfillment, getting there is the thing rather than the journey itself. Similarly, to judge that a presentation or performance is good as far as it goes is to condemn with feint praise. The implication is that the activity has not succeeded as well as it might have; it goes some way, but not a long way, and certainly not far enough. Here, distance is again invoked as accomplished journey. A similar linguistic operation is at work in saying that a performer is far and away the best in a group. To outdistance the competition is to win.

Terms that are less self evidently related to distance than are 'far' and 'near' can still be difficult to conceive or comprehend without some prior notion of distance. Examples are proximity, range, reach, centrality, neighbor, core, and periphery. Concepts of distance are also embedded in language about spatial processes such as movement and diffusion. The clinical neutrality of these references to distance contrasts with more profoundly valued notions of distance embraced by the terms exile, banishment, deportation, marginalization, and alienation. Here distance has negative connotations. Enemies who are not far enough away, and lovers who are too far apart, also attest to distance being a thing to which value can be attached. Children, swimmers, hikers, and strangers in a town are warned not to venture too far. There is not just a vocabulary of distance; there is also a rhetoric of distance.

The references to distance in everyday English occupy a spectrum from bland to emotionally charged. The variety suggests that even outside professional geographical circles distance can be matter of fact (as in a street map or atlas), but can also be indeterminate and individualistic. Up to a point, distance is what people make of it; it may be as complex and nuanced as one wishes.

Physical distance is perhaps its simplest expression. The artifacts of physical distance include the stage posts that were positioned adjacent to animal drawn carting and coaching routes in the days of premotorized land transport. These stone or wooden distance markers measured the passing of defined stages of a journey, such as halfway. They also might have doubled as indicators of the stages at which teams of horses needed resting or changing. Watering points, stables, and inns that were established at such places replaced the knee high roadside markers.

Milestones were an equivalent but more formal measure of location on a route, and a pointer to both the distance traversed and remaining. Physical markers are now part of what might be termed distance heritage. Yet the importance and fascination of marking distance traveled remains: today's children still want to know with irritating frequency how far it is to a holiday destination. Airline passengers knocked senseless by stress and boredom find comfort gazing at representations of their flight progress on screen displays in passenger cabins: real time maps show the approximate location of the airliner and digital records measure the distance (and time) flown and the distance (and time) remaining to destination. More poignantly, one imagines that displaced people and refugees fleeing famine and war register acutely their distance from origin and their distance to safety.

The notion of a milestone marking progress is captured by its adaptation as a figure of speech that relates to people's careers. Long after travel distance could be 'clocked' independently by a milometer inside a motor vehicle, and long after road signs erected at junctions would indicate distances in fractional units to people traveling in multiple directions, people still benchmark personal advancement and critical steps in a life and a career by referring to milestones. Attaining a university degree is one such. Becoming a parent is another. Successfully completing a first 2 m high jump would be a milestone in an athlete's career. Facing up to an awkward telephone call rather than avoiding it would be a milestone in a shy person's personal development. A parallel speech form would involve talking about a person having come a long way.

Linguistic instances of distance give some insight into the history of dealing with the phenomenon, and also into its complexity. A good example is contained in the word speed. Although the reference to distance is hidden, it is one of two crucial constituents – the other is time. Speed is a measure of the time taken to travel a specified distance. An element of distance is intrinsic to the conception and measure of speed, and to its derivatives, namely, acceleration, deceleration, and momentum. In human geography, the idea and metaphor of a shrinking world is inconceivable without some preconception of distance and speed.

The link between distance and time is articulated in other common speech forms such as phrasings that gauge distance in time units: a friend may be said to live a 10 min drive away. The statement may reveal ignorance of actual distance. It also presumes that a listener will understand the re codification by virtue of having been a private vehicle driver or passenger. Timing distance is an effective way of combating the difficulty of sensing distance on journeys that involve disorienting twists and turns. It is also an effective way of gauging the length of walking journeys made without access to a vehicle milometer or a clear map scale. Referring to travel time instead of physical distance is perhaps the most common translation of distance; questions about how far places are apart are often answered in time units. Another instance of distance translation occurs in the context of family trees: a distant relative may be genetically remote, but may live close by. Distance can indeed be ambiguous, both figurative and real.

Dimensions of Distance

For most of human history distances have been apprehended, sketched, written, and spoken about in vague terms. There is little research on pre literate representations of distance in cave wall paintings and hieroglyphs, but it appears that distance was a practical concern, a matter for travelers and not a common reference point or a form of general knowledge. Being mobile, or immanently mobile, was the reason to discover or inquire about distance. Measures of distance might be gauged by the reported possibility of reaching a destination before sunset, or by the number of times one slept on a journey. The crudity of these measures resides in the vagueness and individuality of journey speed, and the difficulty of accounting for short nights of sleep that lengthen daylight travel. As modern travelers know, discomfort may indeed index length of journey, but it is not necessarily a reliable or shared measure. Short journeys can be worse than long; traveling the identical distance in different fare class accommodation can be a very different experience.

Other primitive measures of (travel) distance also revert to time as a base. Historically, distance was also recorded in terms of the number of days spent on horseback or the number of weeks spent at sea. In these cases imprecise measurement lies in the variability of equine performance and in the vagaries of wind, currents and sailing technique. Measures of distance were relative rather than absolute, and were imprecise but useful re flections of orders of magnitude.

Impressionistic, individualized, and subjective meas ures of distance have dominated much of human history, notably in pre literate societies. In the West, efforts to attain a standardized, mathematized, and more or less precise measure of distance date at least from Roman times. A Roman unit of distance (called a mile) was de fined as the equivalent of 1000 paces. Variable length of stride would have made the measure an imperfect and unstable standard. Fractions of a Roman mile were simple to calculate but not always expressed simply: the unit of distance named a stadium was the eighth part of a Roman mile. This was the equivalent of 600 Roman feet (of in determinate size). In time, the Roman mile became an English mile; a stadium was then made the equivalent of 220 yards; conversions could be approximations and quite arbitrary. The adoption of the statute mile addressed the need to standardize discrepant mile measures in various countries: it was defined as 1760 yards, longer than the approximately 1618 yards that constituted the English and Roman miles.

Not only have units of length baring the same name varied according to period and locality, but human his tory is also littered with locally unique and purpose specific distance measures. A furlong (still used on horse racing tracks) was originally equivalent to the length of a furrow in a common field, a site equivalent to a square of 10 acres. The persistent problem of internal cross ref erencing is clear: the length of one thing depends on another length that also requires definition. Similarly, a furlong was denominated in poles, being equivalent to 40, laid end to end. Like the exact length of a pole, the length of each of the 20 measuring rods that constituted a chain could be quite variable. Benchmarks could be quirky, yardsticks less so. Minor differences do not always matter: the base line precision sought by surveyors, machine engineers, rocket scientists, and surgeons is certainly not required by human geographers. For them it is academic that, since 1983, the international standard measure for a meter has been the distance traveled by light in a sealed vacuum in a specified fraction of a second.

Despite unitization and standardization, there is still no single universal measure of geographical distance. Metric measures (meters and kilometers) first adopted in late eighteenth century France had the important feature that they were part of a scale of easily divisible numbers. The elegant system has since been adopted in many countries, but geographic distance is still measured in the clumsier imperial system in countries such as the UK and the USA. Other anomalies exist. The universal measure of distance at sea is the nautical mile. Equivalent to 1852 m (6087 feet), it is sometimes called a geographical mile and is precisely the length of one degree of arc at the equator. Conventionally, across the globe, ship speeds are measured in knots, and flying and cloud altitudes are recorded and represented in feet.

Effective Distance

Despite the curious and amusing historical geographical variations in measuring distance, and despite persistent anomalies and inconsistencies, the measurement of geo graphical distance is simple. There is also a high degree of certainty about the accuracy and comparability of distance measures. Public faith in the measures is high, and they are seldom if ever controversial. But simple distance measures can be misleading. The notion of effective distance adds a layer of complexity that is ne cessary in some contexts. Effective distance distinguishes between actual geographical line distance (whether straight line or on a circuitous route) and the distance that is effectively embedded in a transport network or in a series of journeys. Effective distance is a combined measure: examples are passenger miles and ton miles. These compound measures should not be confused with the trademark airline passenger loyalty reward scheme branded as air miles.

The need to combine distance with a measure of quantity arises most obviously when measuring prod uctivity in the transport industry, especially comparative performances across different networks and organiza tions. A transport operator, A, whose 100 vehicles tra verse 1 million km in a year may appear to be making more efficient use of them than operator B whose fleet of 100 vehicles cover less distance. But B, who operates an urban transport service, carries more freight and more passengers: multiplying distance by load will show that B may outperform A (who operates long distance intercity services) in terms of passenger miles and freight ton miles. Effectively, B delivers commodities a greater distance than A; the effective distance covered by A is less.

Whereas they are intended to present a faithful pic ture of haulage actually accomplished, combinatorial measures involving distance are also open to misunder standing and manipulation. Corporate reporting can show impressive ton mile (ton km) or passenger mile (passenger km) statistics whose meaning is difficult to grasp intuitively and will always be opaque without any corresponding raw data on distances actually traveled or absolute loads. An impressive ton mile figure may reflect low loads carried long distances or large loads carried short distances.

Route miles are another common measure of effective distance in the transport industry. In the case of fixed line transport by rail and tram, for instance, it is often useful to distinguish between track length and route length: the length of single track infrastructure is half that of a parallel double track installation but the route length is identical. Comparing only the track lengths of train or tram operators can give a misleading impression of the geographical extent of their services. Track configuration and density is also germane: 1000 km of densely packed double track may service a smaller area than 500 km of single track.

Relative Distance

Indexes of effective distance emphasize that although line distance is easy to measure and comprehend, it is a simplistic measure. The notion of effective distance also indicates that distance needs to be and can be measured in relation to other elements. Physical distance – some times called absolute distance or engineering distance – most certainly does exist, but not all distance is a mere attribute or property of the physical world itself or of its mappings.

Relative distance is similar to effective distance, al though it is more intuitive and more widely adopted. The idea that distance is relational does not contradict the sheer physicality of distance. The two can coexist. The direct or crow fly distance between two points is a measure of uninterrupted, straight line distance. The distance by bus or underground train between the same two points is likely to be longer because roads and track wind round or under buildings. In this case, length of route and length of journey exceed the direct distance. The cyclist who cuts across fields and down narrow lanes creates a journey of even shorter distance between the identical two points. These various distances are medi ated by the mode of transport; journey length is relative to the type of transport used. In a deltaic or wetland area, the river journey between two villages might be 5 km whereas the overland journey is 20 km. Whether walking, riding on horseback, or motoring, people who choose to use a road between two points are traveling a distance that is route mediated. This denaturalized distance is an attribute of the built environment not the natural environment.

Route distance has been the norm for so long in human affairs that it is seldom referred to. The habit contrasts with the unspecified rule for measuring distance as the shortest straight line interval between two points. But crow fly distance is never used in tables of distances in road atlases; instead they measure (the most direct) route distance. If route distance has to all intents and purposes become our basic human register of distance, it cannot be said that this is an absolute distance. The point about anything absolute is that it is stable. Route dis tances change, reflecting the construction of new route infrastructure such as a motorway.

Route distance is only one instance of relative dis tance. Other types of relational distance that fall under the umbrella of technologically mediated or machine distances are better known. In 1954, Watson referred to the human factor in distance rather than to any techno logical consideration, but the distinction is slight. The two best known and most commonly used instances of relative distance are cost distance and time distance. The two indexes are not combinatorial like effective distance. Rather, they are a measure of distance expressed in cost or ticket price on the one hand, and in elapsed journey time on the other.

Time Distance

The time taken to cover distance is an extremely com mon measure of separation or journey length. Time may even be a more common register of geographical distance than physical length. The innumerable occasions on which people refer to a place being a few minutes away, or within an hour's drive, or about a 20 min walk, or an overnight flight away, reveals a more or less universally understood gauge of distance. Precision is not required or expected in such conversations – estimates suffice and orders of magnitude matter most. Units and fragments of time, it appears, are better understood than units of distance; time is visible (on a wristwatch or cell phone), whereas distance can be hard to see and its passing harder to monitor. Distance is easily denominated in minutes and hours.

The re registration of distances as time is a little surprising given that they are not homologous. Except under controlled conditions, it will seldom if ever be true that some fixed unit of distance traveled invariably equates exactly to a fixed unit of time traveled. Public transport schedules show different travel times for identical journeys: some are nonstop express services, others are all stops. Variable speeds are also due to variable degrees of traffic congestion and bad weather. Even under the most favorable conditions, topography can mean that the evening bike ride from town takes longer than the morning bike ride into town along the identical route. There is no single time distance between two places, whereas there is a single straight line physical distance.

Like other relational distances, time distance is tech nologically mediated. Generally speaking, advances in transport and telecommunications technology have ef fectively reduced the geographical distances between places. Over the centuries successively faster forms of transport have moved people and cargos more quickly between places. In the sailing ship era, the trans Atlantic journey was longer than in the succeeding steamship era. Steaming across the Atlantic made it seem less wide, and New York and London appeared closer. Edinburgh and London became less far apart in terms of time when railway transport became an alternative to overland coaching or coastal sailing. On long range, one stop jet service over the North Pole, contemporary flights from Britain to Australia track an entirely different great circle course and take fewer hours than the number of days consumed when banished prisoners were transported by sailing ship round Africa. Similarly, improvements in each distinctive mode of transport have lessened the time taken to reach outlying towns from a city, and vice versa. Transcontinental flights by jet brought Chicago closer to Los Angeles in time space than when they were served by propeller engine service on the same route. Distances, it appears, are plastic.

It is relatively easy to verbalize and graph the effect of faster transport on time distance, but it is less easy to represent on one map the time distance convergence between all places. Graphing time space compression between Paris and one other French town over a century is not difficult. Nor is it hard drawing an isoline map showing how many more places in France were access ible within a 2 h journey from Paris in 1990 than in 1950. But these illustrations cannot simultaneously also show time space compression between Lyon and Marseilles. Irrespective of scale, the visualization and depiction of multiple time space compressions is difficult.

The sophisticated cartographic transformation re quired is not unlike that involved in redrawing the world map of countries according to their relative population sizes or wealth, for instance. The difficulty in compiling such cartograms is keeping the shape, borders, and relative location of countries more or less congruent and recognizable. Distance transformation is equally un wieldy and problematic, and was not resolved until the 1970s. Since then, a powerful computer based statistical technique called multidimensional scaling has enabled representation of relative distances between all places in a set. The technique analyzes and methodically re sorts data in a way that expresses the strength of their interrelationships.

The output from multidimensional scaling of time distances is a new mapping of the location of places relative to one another in time space. Places 1 h apart are remapped as closer than places that are close together on the ground but 2 h journey apart. Two towns 50 km apart but connected only by irregular air service via a remote third town would appear further apart on a time space map than on a normal map. Map bearings are absent from this newly configured space and the absolute (real world) locations of places are unrecognizably distorted. Al though the transpositions are striking, they are not easily understood, so production of time space maps has been slight and sporadic. No world atlases publish such radical reviews of the world.

Cost Distance

Sometimes called economic distance, this measure of relative distance is not as common as time distance. There are certainly no linguistic expressions in English which convey distance in terms of cost. People are not heard conversing about places being 2 dollars distant, 5 pounds closer than another place, or 3 euros further on. One good reason may be the (unpredictable) asymmetry between the price of travel and the distance covered. Ticket pricing on public transport is structured in (in creasingly) opaque ways that are not one to one or in any other stable ratio to actual journey length. Variable terminal charges, flat zone charging, backhaul charging, and clever marketing are partly responsible. Budget air lines, for example, charge different fares for outbound and inbound flights, and some promotional fares on one leg of a long haul flight are pitched at less than for a few stops on urban transit.

The extent to which cost distance calculations enter into locational decision making, commodity delivery, and personal travel behavior is not well researched. Com mercial travelers, and businesses and corporate divisions that specialize in logistics, presumably have regard for transport distance configurations that are costly and those that are less so. And it is not inconceivable that private individuals who have detailed knowledge about public transit fares may decide where to rent accommodation or buy a home by costing annual and decade long savings by living one side of a fare zone boundary rather than at the next station down the line in another fare zone. The introduction of innercity road user charging schemes in cities such as Singapore and London will have made motor vehicle users more conscious of cost distances.

Effort Distance

Effort distance is a little used construction that may have referred initially only to the physical energy required to move some distance. Over time, the amount of physical energy needed to carry, push, pull, paddle, and cycle has been diminished by motorized transport. Nevertheless, there remain many places in the world where consider ation of the bodily energy required to move might re main a consideration – where choice of destination is made according to which journey taxes the human body or animal carriers least. The idea of effortless travel punted by airlines and railway and shipping companies clearly takes no account of the irretrievable energy ex pended in making the vessels and vehicles, or the energy sunk in transport infrastructure.

The criterion of minimizing effort distance was elevated into a locational principle long before the contemporary alarm and anxiety about depleting energy stocks and the need to conserve energy. The principle might have even more traction now when companies and private individuals seek to minimize their use of fossil fuels and their ecological footprint by traveling less often or less far. In an age of bureaucratized and congested hyper mobility the notion of effort distance might also be broadened to encompass the bother and emotion expended on moving between two points. Effort distance re calibrates real distance against the fuss and strain involved in crossing it. However popular it may be to measure frustration distance, the multiple subjective and experiential variations involved in measuring and comparing inconvenience render standardization and intelligibility almost impossible. By contrast, expenditure of time and money is more objective.

Cognitive Distance

Whereas distance measured relative to time, cost, and irritation emphasizes the negative aspects of crossing the known physical distance, the notion of cognitive distance brings into question even the salience of true or absolute distance. In keeping with the behavioral revolution that swept through their subject in the 1970s, human geographers queried the assumption that real physical distance (or perfectly known linear distance) was a major consideration in personal or corporate location and travel decision making. As opposed to omniscient and rational actors in some theoretical universe, real people – so it was observed and suggested – took decisions in relation to their perception of distance to a visible destination, or their cognition of distance to an unseen destination. Exaggerated or underestimated, if either such distance stood in a constant ratio to actual distance, the concern would be immaterial. But interpersonal variations are highly likely; they would probably express socio economic variations. People with access to private transport may sense distance quite differently to people who are captive users of bus transport and whose urban travel routes meander through suburbs. Both young and elderly people are likely to sense distance from world views unique to their generation: small children from innocence of distances longer than those at home, old people from memories of transport and travel in a bygone era.

Differences between cognitive and actual distance can emerge through directed interviewing or incidentally through conversation and sketch mapping. The mental maps on which people choreograph their daily activities may lack fidelity, but they do show how accumulated knowledge and limited activity fields distort scale. Under and overestimates of distance occur on least traveled routes. On the fringes of behavioral human geography and psychology there is interest in how people acquire and respond to the mind maps that they have of places and of spatial relationships. Research into cognitive mapping provides some insight into how people apprehend and estimate distance, and how mental yard sticks form and change. Imagined distance cannot emerge from nothing.

Affective Distance

Felt distance is a close relation of cognitive distance. The difference is that this measure carries an emotional charge. The affectation is mostly the creature of a past and wider world in which the distances that separated people were more significant than in an age of jet travel and voice and image connectivity via telephony and the Internet in real time over long distances. Distance mattered more in the past; separations at railway stations and harbors were often permanent and were emotionally taut events. Writers knew the weight of distance, and in classical literature there are frequent references to distancing being associated with lovers missing each other and pining. The first syllable of the word longing is redolent of distance. The cliché about absence making the heart grow fonder approximates some of the classical allusions. The condition of not being present has qualities of time distance; unusually, however, distance becomes ambiguous, at once despised and welcomed.

Social distance is another kind of felt distance: it measures the gap between social classes. The social divide between blue collar and white collar workers cannot be measured with a ruler. Similarly, the differences between the lives of poor, middle income, and rich people are only notionally related to distance. The width and intensity of the barrier between immigrants and locals is gauged better by expressions of xenophobia than by sheer physical distance between the places where they congregate. Yet geographical distance can index social distance. In apartheid South Africa, for example, the distance that the state put between different races by forced residential separation was intended to reinforce and flag social distance. The distances between spontaneously formed ethnic and religious clusters in all cities reflect something of the degree of social division. Distance differentiates.

The Significance of Distance

The experience of distance as an expression of absence and social difference gives it considerable pungency. These personal registers of distance in public spaces are not ones that have roused many human geographers; their dominant treatment of distance is as an obstacle and a deterrent.

Celebrations of distance as a positive phenomenon are rare. The first syllable of the word even has a negative connotation and, like dislocation and disease, distance is generally unwelcome. Yet there are instances of inten tional distancing to attain freedom, progress, and peace. Voluntary relocation can be positive, as in putting dis tance between an aggressor or oppressor and a victim. Many young people choose to escape the conservative and constricting influences in a small town. Isolation is actively sought by some writers and artists and retirees. Self imposed exile is another instance of regarding and embracing distance as a protective shield.

Instances of positive distanciation aside, the ascription of negative effects to distance is overwhelming. Distance is almost entirely regarded and presented as some thing that enervates action and that attenuates spatial interaction: people lose contact with distant friends and traders focus on nearby markets. The undesirable, en ergy sapping effect of distance is captured in the classic but unattributed reference to the friction of distance. The nod to physics suggests that distance has some natural, inherent properties. But physics also teaches that friction can be overcome by judicious lubrication and design engineering. Attributing frictional properties to distance invites solutions to overcoming something unwanted. Viewing distance in negative terms aligns with bombastic terminology about smashing it.

On the human stage, where personification of nature is common, the negativity of distance has been captured most strikingly by allusions to the tyranny of distance. The historian Geoffrey Blainey adopted the phrase as the title of his 1966 book about a key condition of and in fluence on Australia's past. Five years previously a set of public lectures had reflected on the meaning of and public consciousness about the remoteness and isolation of New Zealand. The notion of tyranny resonates with oppressive power and arbitrary force. The implication is clear that remoteness generates adversity, disadvantage, and myopia, and that the requisite travail (the struggle of overcoming distance) is financially costly and can take its toll on human health and life.

Arguing that the destiny of a people and a continent was wrapped up in its antipodean location, Blainey felt able to track Australian development only to the point where distance had been tamed rather than conquered. Efficient transport (which he termed the enemy of dis tance) had altered much in Australia, but not everything. Distance had been tamed less in people's minds than on maps, he suggested. Some attitudes still belonged to a more isolated era. Blainey's remark that Australians still lived in one of the world's billabongs could cut two ways. Either they continued to behave as if they were still settlers in a land far from British influence, or Australians were in denial of distance and conscientiously repro duced British ways of life in their diasporic community.

By the time he wrote the preface for a revised edition of his book in 1982, Blainey noted that the tyranny of distance had become a catchphrase in Australian political and economic debate. During the last quarter century of globalization, improvements in transport and communi cations will have shifted feelings and attitudes of re moteness. Reorientation of Australian affairs and affinities into the Pacific world will have lessened its geographical isolation if not the feeling of isolation. There may be some pride, after all, in accepting and approving the distance that is definitive of a country and a culture.

Distance Decay

Possibly the most familiar profiling of distance in human geography has been in studies of its attenuating effects. The term distance decay has long been used to capture diminution due to distance. An example of distance decay is the declining level of aircraft noise away from an airport. In the absence of some other remote source, the occurrence of lead poisoning decreases away from a motorway. Fewer migrating birds reach their destination than started out. Species diversity intensifies away from the species core; exoticism is about distance.

As in nature, so too in social affairs the effect of distance is to reduce the intensity of phenomena. More commuters travel short distances than long distances. Attendance at cinemas falls off the further away people live. The degree to which Celtic is spoken and understood diminishes as distance from Ireland, Scot land, and Wales increases. The empirical observation of distance decay is nothing new. Mapping the incidence of cholera in Victorian London showed clustering round one contaminated water source. In the mid nineteenth century an American psychiatrist and statistician noted that use of a medical facility declined with the distance it was away. In the 1930s research into the selection of marriage partners showed that most people met and married people from nearby. A wag once reported moral decay among British expatriates living far from the restraining influences of family, peers, and church. As a final example, a shrinking and reconfigured world has always featured in writing about the effects of aviation.

Observation and analysis of numerous instances of distance decay became something of a cottage industry in human geography in the last quarter of the twentieth century. Numerous studies were conducted of the rates of spread and adoption of things such as agricultural practices, of innovations such as television, and of the spatial diffusion of disease by contagion or other means.

Instances of distance decay are no longer clear cut in a globalizing world being changed by cheap travel, tel ephony, and the Internet. In the past it would have been easier to claim that radio station listener and newspaper reader numbers taper off from the source of broadcast and publication, that fewer long distance telephone calls are made than short distance calls, and that more ham burgers are eaten in the USA than on other continents. Nearby things are no longer always more similar than things far apart. The number of spectators at Wimbledon still falls off rapidly as distance from London increases, but the ratio of Londoners to outsiders and foreigners has probably altered significantly. Certainly, the number of people round the world who watch the tournament live on television exceeds the number of Londoners who do so. Time of day of live broadcast in far off countries (time zone distance rather than time distance) affects viewer numbers more than distance from the event.

Scientific study of the way diffusion declines with distance from a source, or how distance lessens spatial interaction between two points, uses mathematical no tation and modeling to express attenuation. Starting in the 1960s human geographers drew on Newtonian physics to relate the intensity of trade or movement between an origin and destination to the inverse of their distance apart. In gravity models, distance appears as the denominator in the equation; it is a variable that reduces the volume of interaction between two points whose at traction is measured by the numerator. In keeping with molecular physics, when calibrating a gravity model against observed trip volumes, the distance denominator is generally adjusted by raising it by a power. Typically, physical distance is squared so as to capture the geo metric rather than the additive effect that incremental units of distance have on deterrence: people are more than twice as unlikely to travel 40 miles as they are 20 miles. How people regard and weigh distance may be at least as important as sheer physical distance itself. Powering distance mathematically honors the power of (cognitive) distance.

Like most arguments that work with distance as an explanatory variable, gravity models imply that people know distances accurately and are infinitely sensitive to their small variations. Tying behavior to estimated dis tance – even to distances ranked on a scale of closest to furthest – may capture more accurately how people consider distance. A new grocery store built 100 m closer to a housing estate than an existing store will not necessarily deflect customers. Loyalty and anger over the new commercial development may outweigh slight savings in time, shoe leather, and petrol. However ir rational it appears from an economic point of view (especially if prices at the new store are lower), emotion, habit, and reluctance to investigate a new option may mean consumers do not attempt to minimize effort dis tance, cost distance, or time distance. The marginal benefits, tested or guessed, real or perceived, might just be too small.

Decaying Distance

The term distance decay refers to decay by distance and not to the decay of distance. Yet claims do erupt peri odically about the demise of distance. Cognizant of the shrinking world, it is reasonable to argue that distance has for some time had diminishing significance in human affairs.

In ancient history, when technologies of transport and communication were rudimentary, distance had extra ordinary significance. In nonindustrial societies, distance was an almost insurmountable barrier and, in con sequence, mythical qualities were ascribed to distance. To travel beyond the visible horizon was not just to journey a long distance, it was to reach the conjunction of earth and sky. There, elite politico religious adventurers were believed to have either encountered the past (be ginnings and ancestors) or to have seen and experienced the future (utopias, afterlife). Observers and travelers alike vested distance with symbolic power. Cosmological distance was associated with mystery and the super natural, and long distance travel with the acquisition and exercise of power and knowledge. Distant lands and people were sources of esoteric knowledge and exotic material wealth. The ability to overcome distance was a sign of power, ability, wisdom, and work; to conquer distances was to obtain religious grace.

The contemporary world retains little faith in such meaningful distances. The demystification of distance has been followed by its romanticization. Tourism discourse, for example, advertises that the grass is greener else where, that the beaches are whiter, the sun warmer, and time less pressured. In a virtual world, references to distance are figurative rather than literal, and the debate now is not about the significance of physical distance but about its insignificance. In a globalizing world that ap pears to be shrinking, distance cannot retain the same force it did 10 years ago, let alone a century ago or 1000 years ago. Technologies of transportation and com munication have created a global village where every thing seems close and nothing far. Notices about the death of distance echo populist claims about the end of geography, the annihilation of space, and the flattening of the world.

These catchy assertions work with notions of relative distance rather than absolute distance and do not claim that the Earth itself is shrinking. There is no contention that absolute distance does not exist, but it is argued that distance is not entirely elemental. The claims about the shrinking world point to the passing of distance as a material consideration in human affairs. The claims move away from the presumption that distance is merely a property of and an object embedded in the physical world. The objectification and reification of distance in the nat ural world is countered by a different paradigmatic view in human geography which stresses that distance (and proximity) is a relational entity. But geographical distance is not just a technical determination either, fabricated by and in thrall to the speed and price of (virtual) mobility. Post positivist geographers regard distance as being a human construction, endlessly fashioned and shaped by human agency and not just by machines. As the pre industrial case shows, and as cognitive mapping demon strates, distance is also an idea. And, in the new mobilities paradigm, distance is coming to be problematized as a residue of codifications, activities, experience, imagin ation, and memory. Any universalizing and standardizing of distance strips away its rich cultural, personal, and institutional associations. Reductive and essentializing representations of distance may strip away its abstraction and baffling complexity, but they denude it of meanings that are socially constructed, negotiated, and contested.

In addition to concealing layers of distance making, catchy claims about the demise of distance fail to deal with the nuances of contemporary spatial reconfigur ation: the world is not shrinking uniformly, and de distanciation is accompanied by re distanciation of places in plastic spaces. Not all communities feel the taming of distance let alone its destruction. The aptness of the shrinkage metaphor is open to question when it is so uneven; instead, the world map might be regarded as bending and twisting. A contorting world is moulding new global peripheries. Marginalization occurs at places that are eclipsed despite being relatively near the world's economic hubs. Forty years after being written, the final sentence in economist Wilfred Owen's book on 'distance' and 'development' may be truer than ever. Resorting to a common expression that uses figurative distance to con vey degree, he concluded that distance and isolation went a long way to explaining the poverty and stagnation of 80% of humanity.

See also: Maps; Mobility; Scale; Space I; Space II.

Further Reading

Blainey, G. (1966). *The Tyranny of Distance*. Melbourne: MacMillan.

Cairncross, F. (1997). *The Death of Distance: How the Communications Revolution Will Change Our Lives*. Cambridge, MA: Harvard Business School Press.

Ewing, G. (1974). Multidimensional scaling and time space maps. *Canadian Geographer* 18, 161 167.

Falk, T. and Abler, R. (1980). Intercommunications, distance and geographical theory. *Geografiska Annaler (Ser. B)*, 62, 59 67.

Friedman, T. L. (2006). *The World is Flat: The Globalised World in the Twenty First Century*. London: Penguin.

Gattrell, A. C. (1983). *Distance and Space: A Geographical Perspective*. Oxford: Clarendon.

Helms, M. W. (1988). *Ulysses' Sail: An Ethnographic Odyssey of Power, Knowledge, and Geographical Distance*. Princeton, NJ: Princeton University Press.

Janelle, D. G. (1973). Measuring human extensibility in a shrinking world. *Journal of Geography* 72, 8 14.

Kitchin, R. and Freundschuh, S. (2000). *Cognitive Mapping: Past, Present and Future*. London: Routledge.

Leyshon, A. (1995). Annihilating space?: The speed up of communi cations. In Allen, J. & Hammnet, C. (eds.) *A Shrinking World? Global Unevenness and Inequality*, pp 11 54. Oxford: Oxford University Press.

May, J. and Thrift, N. J. (2001). Timespace: Geographies of Temporality. London: Routledge.

Miller, H. J. (2004). Tobler's first law and spatial analysis. *Annals of the Association of American Geographers* 94, 284 289.

Olsson, G. (1965). *Distance and Human Interaction: A Review and Bibliography*. Philadelphia, PA: Regional Science Research Institute.

Owen, W. (1968). *Distance and Development*. Washington, DC: Brookings.

Sinclair, K. (ed.) (1961). *Distance Looks Our Way: The Effects of Remoteness on New Zealand*. Auckland: University of Auckland Press.

Ward, R. G. (1995). The shape of tele cost worlds in the Pacific island case. In Cliff, A. D., Gould, P., Hoare, A. & Thrift, N. (eds.) *Diffusing Geography*, pp 221 240. New York: Blackwell.

Watson, J. W. (1955). Geography: A discipline in distance. *Scottish Geographical Magazine* 71, 1 13.

Dudley Stamp, L.

D. Linehan, University College Cork, Cork, Republic of Ireland

Dudley Stamp, Lawrence (1898–1966)

Lawrence Dudley Stamp was born in London in 1898 and admitted to King's College at the age of 15 to study botany and geology. Following graduation he joined the Royal Engineers and served briefly in France during World War I. Illustrating his exuberance for the subject and his phenomenal work rate, he gained his doctorate in geology in 1921 while simultaneously taking the London BA in geography. He was then employed as a geologist in Burma (Myanmar). While stationed there, he was ap pointed as the first professor of geography in Rangoon in 1923, where he taught until 1926. In Burma, Stamp first specialized in tropical geography. In one of his first books – *The Vegetation of Burma from an Ecological Standpoint* – he enrolled researchers from of the Indian Forest Service to collect data and adopted a classic chorographic method to describe the country. Drawing upon the ecological theories of A. G. Tansely and a model for the systematic classification of natural regions developed by the geographer A. J. Herberson, his tech nique examined the causes and the interrelations of phenomena in the landscape, primarily in terms of cli mate and geology. While he would continue to innovate – he was among the first geographers, for example, to promote the use of aerial photography – this approach to the analysis of land use and resource management would characterize his career. His land use survey technique and representation first developed in Burma, was later scaled up to encompass Britain in the 1930s, and underwrote his ideas for a world land use survey spon sored by United Nations Educational, Scientific and Cultural Organization (UNESCO) he directed in the 1960s.

A postcolonial reading of Stamp's work in tropical geography can align his interests with the demands of the colonial state. His paper on Burma published in 1933 in *The Geographical Review* concluded that the country's 'wasteland' – which he depicted as unoccupied – was open for further colonization: "Upcountry one is struck by the immense areas of cultivable land which remain waste for want of cultivators" (Stamp, 1933: 87). He could also indulge his reader with a dash of adventure and hint of the exotic: "The Salween still indulges on the quiet in the time honoured practice of head hunting and their territory has been placed out of bounds to Europeans" (Stamp, 1933: 83). Most of Stamp's work, however, was very applied in orientation. This was well reflected in his stewardship of *Commercial Geography* which he inherited from the economic geographer G. G. Chisholm. This is not so much a book but a database of commodities and the ever expanding resource frontier required by a co lonial empire. At this time, it is arguable that Stamp – notably with his work on the 'colonial soil' – participated in a school of Western intellectuals who mobilized sci ence in an attempt to guarantee the effective manage ment of imperial sources.

In 1926, Stamp returned to England to take up a post at the London School of Economics (LSE). The LSE positioned him in a metropolitan network that supported a career at home and abroad. During this period, Stamp consolidated his activities as a successful author of popular geography textbooks. For example, he collabor ated with his wife Elsa Rea to produce an innovative *New Age* series for junior schools, based around photographs taken during his world travels. *The World*, first prepared for the school market in India in the mid 1920s, was adapted widely throughout the British Empire and, over 40 years, sold over a million copies, running to some 17 editions. His work on school textbooks for both the na tional and colonial marketplace enhanced his association with the Geographical Association which he used to build support for the first Land Utilisation Survey of Great Britain. The idea of a land use survey of Britain had been an objective of the Royal Geography Society from the late 1880s, when the survey "…was regarded as a project of Imperial importance" (Stamp, 1931: 42). Drawing upon his work in the tropics, Stamp mobilized a chorographical technique to devise six categories of land use for the purpose of the survey: meadow and per manent grass, arable land, heath land, and moor land or rough hill pasture forests and woodland, gardens, and unproductive agricultural land. This survey – which began in 1930 – was a huge undertaking but by framing it in the context of both geographic and civic education, Stamp was able to enrol school children and their tea chers to collect primary data on land use, and transfer this data by hand onto 6″ to 1 mile paper maps. The results are summarized in *The Land of Britain: Its Use and Misuse*. H. C. Darby concluded it was the greatest achievement of British geography. The survey success and its immediate value for agricultural and physical planning during World War II and after was Stamp's *entrée* into public service in Britain.

As questions around physical planning and issues of environment and habitat management have intensified in

Britain, Stamp's work on the Land Utilisation Survey remains highly regarded. But as one of the stalwarts of regional geography, the paradigm shifts in the 20 fol lowing this death – first toward quantification and then to radical geography – have often undermined his repu tation among contemporary geographers. In 1974, when David Harvey characterized geography's history as a shift from a concern with empire into "…the technics and mechanics of urban, regional and environmental man agement guided by a principles of benevolence and reformism…" Stamp could well have been his model (Harvey, 1974: 20). Ron Johnston argued that Stamp's promotion of applied geography defined geographers "…as information gathers and synthesisers who stood outside the political process within which planning goals were formulated and pursued" (Johnston 1979: 30). Even one of his warmest observers H. C. Darby conceded that his approach to geography was firmly pragmatic: "…it was said that he had a tidy mind and a passion for facts" (Darby 1983: 20).

These summaries are more problematic than first as sumed. In its various forms, versions of 'applied geog raphy' had always been central to the activities of the state. They came into their ascendancy in post war Britain, when the increased state intervention in the economy and environment presented geographers with innumerable opportunities to identify practical appli cations for geographical techniques and expertise. In this context, Stamp saw applied geography as the application of geography for the public good. Any claim that Stamp saw geographers as 'outside' the political process cannot be sustained. Stamp was part of a group of intellectuals in the 1930s that included Max Nicholson, Vaughan Cornish, and Julian Huxley who foresaw the role science would play in the planning and reconstruction of Britain. Geography for Stamp was a matter for action and intervention. During World War II, he served as vice chairperson of the Scott Committee, which set out the framework for the establishment of the town and country planning policy. He was also employed as Chief Adviser in Rural Land Utilisation at the Ministry of Agriculture. In the post war period Stamp served a member or chairperson of four major official committees – Land Use in Rural Areas, Nature Conservancy, Common Land, and National Resources. Undoubtedly he played a key role in shaping the post war spatial settlement in Britain. It is also clear in this period that he revised his perspective on the Global South. His work in tropical geography turned from a concern with colonial interests to issues related to the 'developing world'. His collaboration with the International Geographical Union (IGU) and UNESCO in particular presented him with new challenges and in their own way are reflected in clear pleas for global re sponsibility in *Geographies of Life and Death* and criticism of neocolonialism in his book *The Developing World*.

Toward the end of his life, Stamp reacted strongly against quantification which he saw a major challenge to applied and regional geography. As president of the In stitute of British Geographers (IBG), he controversially compared Peter Haggett to a theosophist and concluded that "Quantification has many points in common with Communism: it has become a religion to its devotees, its golden calf in the computer" (Stamp 1965: 18). Rather than this path, he argued for public engagement over academic specialization. He also saw geography as a holistic discipline, the unification of the physical and human worlds. Despite the growing challenge then emerging from spatial science, Stamp's profile as Britain's foremost public geographer was assured. In 1965 he was knighted by Queen Elizabeth II. This honor extended to an invitation on to the British Broadcasting Corporation's (BBC) flagship radio program 'Desert Island Discs' where guests are asked to imagine which music and texts they would take with them as if cast away on a tropical island. As he had done with audiences all his life, Stamp charmed the public with a discussion on his work with the world land use survey, the demographic pressure on land resources, the use and abuse of toxic chemicals in intensive farming, the possibilities of land reclamation, and marveled at the possible future cultivation of the sea's harvests. The radio show confirmed his vivid interests in environmental issues and his wide geo graphical imagination. It seems that throughout his life Stamp epitomized the motto of his old regiment, the Royal Engineers – *Ubique* (everywhere).

In spite of the turn toward continental social theory in Anglo American geography, Stamp's contribution to geography remains more central to the discipline than often realized. Stamp exemplified a philosophy about geographical knowledge which located it at the interface of pubic service, education, and civic responsibility. He understood the central role that geography played in the organization of social, environmental, and economic life and was able to articulate those ideas to broad and mixed constituency: school children, the general public, and policymakers. He helped create a geographical discourse which continues to frame many debates about the role of the discipline, particularly those concerning the appli cation of geographic information system (GIS), remote sensing, and land use studies to matters of public policy.

See also: Colonialism I; Colonialism II; Regional Geography I; Tropical Geography.

Further Reading

Anker, P. (2001). *Imperial Ecology: Environmental Order in the British Empire, 1895 1945*. Cambridge, MA: Harvard University Press.

Darby, H. C. (1983). Academic geography in Britain: 1918 1946. *Transactions of the Institute of British Geographers* 8(1), 14 26.

Hoskins, W. G. and Stamp, L. D. (1963). *The Common Lands of England and Wales*. London: Collins.
Stamp, E. C. and Stamp, L. D. (1930). *The New Age Geographies*. London: Longmans, Green.
Stamp, L. D. (1925). The aerial survey of the Irrawaddy delta forests (Burma). *The Journal of Ecology* 13(2), 262 276.
Stamp, L. D. (1925). *The Vegetation of Burma from an Ecological Standpoint*. Kolkata: Thacker, Spink and Company.
Stamp, L. D. (1931). The Land Utilization Survey of Britain. *The Geographical Journal* 78(1), 40 47.
Stamp, L. D. (1933). Burma: An undeveloped monsoon country. *Geographical Review* 20, 86 109.
Stamp, L. D. (1938). Land utilization and soil erosion in Nigeria. *Geographical Review* 28(1), 32 45.
Stamp, L. D. (1940). Southern margin of the Sahara: Comments on some recent studies on the question of dessication in west Africa. *The Geographical Review* 30, 297 300.
Stamp, L. D. (1948). *The Land of Britain: Its Use and Misuse*. London: Longmans, Green.
Stamp, L. D. (1960). *Applied Geography*. London: Penguin Books.
Stamp, L. D. (1965). *The Geography of Life and Death*. Ithaca, NY: Cornell University Press.
Stamp, L. D. (1966). Ten years on. *Transactions of the Institute of British Geographers* 40(2), 11 20.
Tansely, A. G. (1923). *Practical Plant Ecology*. London: George Allen and Unwin.

Dutch Human Geography

H. Ernste and L. Smith, Radboud University Nijmegen, Nijmegen, The Netherlands

Glossary

Action Theory An approach within the social sciences which was introduced in human geography by the Swiss geographer Benno Werlen. This approach systematizes the until then rather eclectic humanist geography to a full- fledged action theoretic approach within human geography. Later on this was developed further in a language pragmatic way by Wolfgang Zierhofer and introduced by Huib Ernste in Dutch human geography.

Ecological Complex A term introduced by the Dutch geographer Christian van Paassen, with which he tried to build on the ideas of the American Geographer Philbrick to designate the complex relationships between human activities and values, resources, built environment, institutions, and their external relationships. This idea was translated in terms of system theory by the Dutch geographer Jan Lambooy and made popular by Gerard Hoekveld through his famous 'Cube of spheres'.

Sociography A school of thought within Human Geography, which tries to combine Sociology with Human Geography. The highly descriptive sociography approach within human geography, focused on comparative and differential analysis formed the basis of the Amsterdam School of Sociography established by Sebald Rudolf Steinmetz. in the first half of the 20th century. Later on this approach phased out by the increased use of quantitative methods in human geographical research.

Introduction

Human geography as a discipline focuses on the contextuality and spatiality of human activities; as such it is evident that the activities conducted in the framework of this discipline can be related to the specific context in which they take 'place'. The Dutch geographer Max Heslinga therefore concluded that it would be impossible to write a 'single' history of 'the' geography. Rather one had to differentiate specific 'national' geographies. This view is also reinforced by David N. Livingstone in 1992, when he states that: "[…] geography has meant different things to different people in different places" and by David Harvey in his *Historical Materialist Manifesto* with respect to the history and condition of geography, where he writes: "The history of our discipline cannot be understood independently of the history of the society in which the practices of geography are embedded," nor from the history and geography of the academic institutions in which human geography developed. On the one hand, this supports the tradition of describing the history of the human geographic discipline from the specific national context as defined by a nation state or language area. On the other hand, one can criticize this position, because it already implicitly assumes a specific geographical structure and neglects possible other aspects and relationships. Here we focus on 'Dutch' human geography as it was established and developed further in Dutch universities. 'Dutch' in this sense is therefore understood as a specific (national) institutional context. The development of Dutch human geography is, however, also closely related to the development of geography as a discipline in secondary education and to the emergence of a strong geographical professional praxis in the Netherlands. Societal 'needs' for human geographic knowledge therefore played a crucial role. The development of human geography in the Netherlands is, at the same time, also linked to human geographic practices in other parts of the world. As we see, it is especially in this respect that Dutch human geography indeed takes a special position in continental European human geography. In this article, a short outlook on human geography in other Dutch speaking areas is also presented, although the links with 'Dutch' human geography are very limited.

Development of Dutch Human Geography

The Establishment of Dutch Human Geography in Academia

For the establishment of Dutch human geography, the Royal Dutch Geographical Society (KNAG), founded in 1873 in Amsterdam, played a crucial role. At that time, the Netherlands was still a strong colonial power and was exploring the unknown parts and resources of the Dutch colonial empire. Motivated by the promising returns of these colonial endeavors, the Royal Dutch Geographical Society, with the support of the Dutch government, initiated a number of expeditions to the Dutch colonies in which not just valuable knowledge of unknown areas was collected, but also the sphere of influence of Dutch government in the colonies was strengthened. At that time, geography was not represented at Dutch universities, instead flourishing outside universities in the form of applied knowledge of the colonies and applied geographical knowledge of the Netherlands as concerning its regional diversity and continuous struggle against flooding. The origins of Dutch human geography therefore have to be sought in different fields of the nineteenth century professional practice, ranging from (colonial) trade, commerce, shipping, land registry, civil and hydraulic engineering, and law to (colonial) administration.

Geography was first introduced as a subject in high school in 1863. As a logical next step, the Royal Dutch Geographical Society lobbied in 1874 for establishing independent chairs in geography at universities, mainly for the certification of geography teachers for high schools. In 1877 they were successful, as the city council of Amsterdam – the city with the strongest colonial interests – decided to transform the local Atheneum into a Communal University of Amsterdam (UvA) and appointed Cornelius Marius Kan (1837–1919) as the first professor in human and physical geography, including the geography and cultural anthropology of the Dutch East and West Indies. Kan mainly focused on educating geography teachers. As he had the conviction that physical and human geography were too distinct for them to be dealt with by one and the same person, notably because they both required close cooperative ties with neighboring disciplines, Kan, at his retirement, suggested that he would be succeeded by two professors, one for the field of physical geography and one for human geography. This suggestion had long lasting effects, providing the basis for what later was to become one of the specific features of Dutch geography, namely, the disciplinary separation of physical and human geography. Thus, upon the retirement of Kan, Eugène Dubois (1858–1940) and Sebald Rudolf Steinmetz (1862–1940) were appointed as professor for physical geography and human geography, respectively.

The Amsterdam School of Sociography

Steinmetz energetically pushed forward the development of human geography. He perceived human geography as a science with a very strong social–scientific orientation and founded what later came to be known as the Amsterdam School of Sociography. At that time, human geography and economics were the only sciences at the university to deal with society in which it was also possible to do a major, as sociology and anthropology only became major subjects for study at universities in the Netherlands after World War II. Sociography, defined by Steinmetz as comparative and differential sociology, focused on nations, peoples, or other human groups, and not so much on countries or regions. Steinmetz also sought to avoid reducing scientific explanations of human activities to those aspects which could be directly related to the physical environment. In his view, other factors, such as institutions, personality, cultural heritage, etc., also needed to be taken into account. At the same time, Steinmetz emphasized the need for detailed and in depth empirical knowledge. For him, there was a division of labor between the more concrete and descriptive, inductive sociography and the more abstract and theoretical and deductive sociology.

After the retirement of Steinmetz, his chair was divided up into a professorship in ethnology, for which Johannes J. Fahrenfort, a lector who continued Steinmetz's anthropological legacy, was appointed, and a professorship in sociography, for which Steinmetz scholar Henri Nicolaas ter Veen (1883–1949) was appointed. Under ter Veen's aegis, a series of descriptions of villages, towns, and regions were produced. These descriptive studies were devoid of generalizations, abstraction, and theory, and they incited a number of sociographers, with more theoretical ambitions, to defect to sociology after World War II. They moved away from the inductive approach of ter Veen, pressing instead for a more nomothetic approach with a consistent theoretical framework. They found inspiration for this in the structural functionalism of Talcott Parsons and Robert K. Merton.

The Emergence of Applied Human Geography

Around the 1920s, the position of human geography in Dutch universities also changed. In addition to the traditional fields of application of human geography knowledge in the Netherlands, namely, the colonies and high school teaching, a third field developed in the form of socioeconomic policy and spatial planning. Thus, ter Veen's PhD dissertation provided a sociographic description of the reclamation and colonization of the Haarlemmermeer, now the location of the Amsterdam Schiphol Airport, and he treated social facts as explanations for the success or failure of colonization processes. In 1918, the Dutch government had just decided to start the reclamation of the Zuiderzee, making ter Veen's study highly relevant for steering this colonization process. This also chimed well with the political current of those days, in which the liberal political philosophy was gradually replaced by a welfare state based on Christian and social democratic principles. The world economic crisis of the early 1930s further stimulated this tendency toward an intervening government. The Dutch government was very willing to develop policies for public welfare and for spatial planning, partly also inspired by the Tennessee Valley Authority, as part of Roosevelt's New Deal. In consequence, many of the Amsterdam sociographers became involved with government policies and civil services, founding a number of government related institutions for practical sociospatial research in the process. During the same period, various sociographic agencies were established by municipalities. These had the objective of monitoring local sociospatial developments.

In 1901, with the implementation of the first law for public housing, a start was made with systematic urban planning at the local level, soon followed by general spatial planning at the regional level in the 1920s and at the national level in the 1930s. In 1941, under German occupation, a central agency for national planning was established. This created many new job opportunities for sociographers and human geographers and provided a firm foundation for the applied character of human geography for which Dutch human geography continues to be known to this day. This propensity toward applied human geographic research found its parallel in the development of the Utrecht School in human geography.

The Utrecht School of Human Geography

In 1908, the Dutch government felt that the time was ripe for setting up a full fledged, well equipped geography department and of the three state universities – Leyden, Groningen, and Utrecht – the last was chosen to host this department. The applied geographer, generalist, and schoolteacher, Jan Frederik Niermeyer (1866–1923) was appointed as chair of 'political, economic, and general geography', while the German Karl Oestreich (1873–1947) was appointed chair of physical geography at this department. Niermeyer strongly opposed the 'at that time dominant' German influence on Dutch geography. He argued against the dogmatism of the physical determinism of Friedrich Ratzel (1848–1904) and conducted a fundamental discussion with Oestreich's main source of inspiration, Albrecht Penck (1848–1945). In the framework of this debate he argued for a human geography in which physical determinism was replaced by the possibilism of the French school. The *genre de vie* and the pursuit of a livelihood were most important in his view. Niermeyer also strongly focused on the economic aspects of the man–land relationship, which also led to the call for a part time professorship in economic geography at the School of Economics (now Erasmus University) at Rotterdam in 1913. This emphasis placed Niermeyer in a clear counterposition to the Amsterdam Sociographic School, although he shared their preference for inductive empirical descriptive research.

Niermeyers' pronounced positioning had far reaching consequences for human geography at the University of Utrecht, which continued to remain strongly influenced by the French school well into the 1960s and which, to date, continues to emphasize economic geography. When Niermeyer died in 1923, he was succeeded, initially by Willem Boerman (1888–1965), thereafter, from 1927 onward, by Louis van Vuuren (1873–1951). van Vuuren was a colonial civil servant who had not enjoyed tertiary training. Perhaps this also explains his definition of human geography as a 'concrete science'. The relationship between human beings and their physical environment, in his view, was realized through the production process, which he considered to be culturally determined. The cultural landscape therefore formed the starting point for his description of the human pursuit of affluence. Under the leadership of van Vuuren, human geography at the University of Utrecht expanded its commitment to applied research. In a paper published under the title of 'Practical geography' he formulated his geographical creed: "Nothing inspires love for a subject as much as it practical application. Even the most dedicated tuition fails to do so." In consequence, he supervised, during the 1930s and early 1940s, the publication of a number of regional monographs and collected cases (polygraphs). van Vuuren was convinced that applied regional geography should give a descriptive overview of characteristics, problems, and potentialities of an urban or rural area, and that these inventories should provide a basis for the spatial planning of that area. These applied studies clearly showed the influence of the French school and also had typical Dutch traits. Thus, by way of example, van Ginkel wrote with regard to the social conditions and open economy of the Netherlands, how these studies contrasted with the containment of the physical historical regional entities that was so characteristic for the French concept of 'pays'. In addition, stimulated by the general economic crisis of the 1930s, the studies conducted in the framework of the Utrecht School also had a clear focus on welfare issues. These studies were often contracted and financed by the government and were intended to serve policies focusing on distress relief or spatial planning toward developing the economic potentialities of regions.

Scientifically, the monographs of both the Amsterdam Sociographic School and the Utrecht School did not amount too much. In both cases, these were purely local descriptions that were without theoretical ambitions or reflections. Some of the polygraphs, however, were more thematic and contributed more to the Dutch scientific debate about the conceptualization of 'welfare' long before 'welfare' became an issue in international geographical literature. Thereby, some of van Vuuren's scholars stressed the economic component of welfare (e.g., H. J. Keuning), while others (e.g., G. J. van der Berg, Chr. van Paassen, and W. Steiginga) advocated a broader social scientific interpretation of welfare as 'social justice'.

After World War II, van Vuuren withdrew from his position as rector of the university, as he was discredited for his attitude toward the German occupants during World War II. Initially he was succeeded by Jan Otto Marius Broek (1904–74), who was strongly influenced by Carl Sauer following his stay in Berkeley for some time. This also led him to introduce American literature in the human geography curriculum – a major turn from the dominant French and German streams in Dutch human geography. Under his aegis, applied geography lost its dominance for some time. However, when Broek left the Netherlands to take up a position at the University of Minnesota, the tradition of regional studies, established by Niermeyer and van Vuuren, was reinvigorated by his successor, Adriaan de Vooys (1907–93).

1945–70

In the period following World War II, Adriaan de Vooys dominated human geography, together with Christiaan van Paassen (1917–96). Until 1973, when de Vooys retired from Utrecht University, the view remained dominant that human geographers should study the ways in which socioeconomic groups made a living in their local geographical setting. In maintaining this perspective, de Vooys did not refrain from going into a heated debate with Hans Dirk de Vries Reilingh (1908–2001), who had succeeded ter Veen in Amsterdam in 1950, with regard to an article which de Vries Reilingh had published on 'human geography as discipline of the humanities'. In this article, de Vries Reilingh accused the Utrecht human geographers of focusing too much on economic aspects, thereby largely neglecting cultural, political, and social aspects. However, de Vooys obstinately stuck to his view that economic geography formed the core of human geography. In the same way, he also defended the descriptive approach to human geography taken at Utrecht against alleged competitors from the side of sociology. Sociology, as it grew out of the sociographic tradition, contrasted the Utrecht approach by focusing much more on theory and model building, thereby position itself as a 'real science'. Only upon the retirement of de Vooys, in 1973, did a younger generation, led by Marc de Smidt, appointed as professor for economic geography in Utrecht (1977 onward), self consciously head into new directions.

Christiaan van Paassen was the other dominant human geographer scholar. Appointed in 1965 in Utrecht for the chair of human geography and spatial planning, van Paassen inspired his scholars with his personal involvement, theoretical inclination, and visionary views. Therein he was especially inspired by the philosophy of science which the German philosopher Theodor Litt (1880–1962) had developed and the phenomenology and existentialist schools. van Paassen argued that, in essence, human geography is a cultural science that needs to maintain a close relationship with the historical sciences. Hence, in contrast to de Vooys, he emphasized the cultural aspects of human geographic phenomena. In his well known article on 'geographical structuration and ecological complex', which was published in 1962, he provided a firm theoretical basis for his views. Nevertheless, his approach did not find much resonance in Dutch human geography, as his theoretical work was difficult to understand and diverted substantially from the mainstream approach in the Netherlands at that time. Especially with respect to his ideas of embodied spaces, he was far ahead of his time and his views were only picked up again by Huib Ernste (1957–) when expressed in his 2001 inaugural lecture at the Radboud University of Nijmegen.

Competition in the field of human geography increasingly also came from within the discipline, when new geography departments were established at various other universities across the Netherlands. This was a direct response to not only a growing demand for human geographers in the field of urban and regional planning, but also a growing demand for geography teachers. Furthermore, the subdiscipline of spatial planning took on a decisive own course. Most of the first chairs in spatial planning in the Netherlands were taken by scholars of van Vuuren, such as Willem Steigenga (1962, Amsterdam), Christiaan van Paassen (1965, Utrecht), Gerrit Jan van den Berg (1968, Groningen), Gerrit Wissink (1962, Radboud University, Nijmegen), and Jaap Buit (1972, Free University of Amsterdam), continuing the focus on applied and integral approaches to planning.

In 1949, Hendrik Jacob Keuning (1904–85) was appointed as professor in economic geography at the new Faculty for Economics at the University of Groningen. One year after his appointment, he had already established a new curriculum for human geography. In the beginning, this curriculum mainly served to provide education to geography teachers, and only in the 1970s did human geographic research fully unfold in Groningen. As a scholar of van Vuuren, Keuning 'exported' the orientation on the French school and the focus on economic issues to Groningen. In his research he applied a functionalistic view, whereby only those phenomena were taken into account which he considered to be human geographically relevant, that is, which played a functional role in the human pursuit of making a living. These functional relationships were initially described separately by sector. Later, during the 1950s, Keuning increasingly also explored the horizontal relationships between different sectors. Regions continued to be the main focus for human geographic research in Groningen, but Keuning defined regions much more as dynamic human constructs and as the product of human technologies and functional structures. Thereby, he diverged from the conceptualization of regions as pregiven natural entities, which was still rather common in both the Amsterdam Sociographic School and the Utrecht School. In 1964, a second chair was created for human geography at the University of Groningen. This was first taken by Johan Winsemius (1910–64), who may be considered exemplary for the transition of Dutch human geography from a colonial science to a science serving the welfare state. He was soon succeeded by Rob Tamsma (1922–2005). Tamsma mainly concentrated on educating geography teachers. In 1973, Piet Lukkes, formerly a researcher at a real estate developing company, was appointed on the new chair for methods and techniques of geographical research, at the University of Groningen. He expanded the focus on economic issues, such as spatial organization of firms, project development, marketing, and investment and also formulated a concept for regional analysis ('*regiologie*') which was directed more toward applied nomothetic analyses of regions, inclining toward generalizable knowledge and reliable predictions. This strong focus on commercial aspects of geography of space is still a prominent characteristic of current human geography at the University of Groningen.

In 1958, Ruud Cools (1908–87) was appointed as the first professor for human geography at the (Catholic) Radboud University in Nijmegen. He was also a scholar of van Vuuren and was strongly influenced by the French school. In his PhD thesis, he analyzed the contribution of one of the most remarkable representatives of the French school, Jean Brunhes. He set up the Nijmegen curriculum along the lines of the French human geography. In 1970, the interest for methodologically more sophisticated human geographic research approaches resulted in the appointment of Piet Kouwe (1928–97) at the Radboud University, as professor for methods and techniques of human geographic research. Together with Christiaan van Paassen and Gerrit Wissink, he had conducted a seminal research project on the formation of different residential milieus in the Haarlemmermeer. In this study, they deliberately avoided the usual method of regional monography, instead focusing on processes of integration and differentiation. In this way, the road was paved for a more process oriented human geography. In 1971 following the resurgence of interest for the relationship of the Netherlands with former colonies around the world, a professorship was created in the field of 'geography of development', to which Jan Kleinpenning was appointed.

The establishment of a department of human geography at the Catholic Radboud University in Nijmegen led the Protestant Free University of Amsterdam to establish a fifth academic educational program in human geography in 1961, not wanting to lag behind. The de Vooys scholar, Max Heslinga (1922–) was appointed for this chair. He specialized in historical geography and the history of the geographic discipline. He was soon assisted by Gerard Hoekveld, who would develop urban geography and ensured that human geographical knowledge was made more accessible to geography teachers. Especially significant was his translation of the theoretical concepts introduced by Christiaan van Paassen into a 'framework of action' ('*handelingsverband*'), schematized in his 'Bollenkubus'. This displayed the relationship between the framework of action and sociospatial structures and artificial morphological patterns. In doing so, he opened a window of exchange between Anglo Saxon and Dutch human geography. In addition to Helsinga and Hoekveld, Jan Lambooy was appointed for the chair of economic geography and Ad de Bruyne for the geography of underdeveloped areas. Jan Lambooy contributed strongly to the development of economic geography with his popular standard textbook of 1981 and his active discourse with regional economics. However, in terms of student numbers, the department at the Free University of Amsterdam remained one of the smallest. This made it vulnerable to budget cuts at the Ministry of Education. Indeed, in 1984 this program had to close down due to such government measures, and most of the staff was transferred to the University of Amsterdam and the University of Utrecht.

At the University of Amsterdam, the establishment of sociology and anthropology as separate educational programs meant that human geography was, more or less, squeezed in between sociology and the more economically oriented human geography in Utrecht and Groningen. de Vries Reilingh, therefore, initiated a reorientation and chose an approach that focused on the spatiality of human activities, without thereby losing the emphasis on social and cultural factors. Furthermore, a large contract research project for the municipality of Amsterdam on the development and functions of the inner city of Amsterdam enforced a theoretical reorientation on urban geographic research. During this phase, Willem Heinemeyer (1922–99), who was appointed as professor for human geography at the University of Amsterdam in 1969, also in close cooperation with the successor of de Vries Reilingh, Christiaan van Paasen, introduced the socioecological approach to spatial problems, which came to dominate Dutch urban geography for a long time. The Amsterdam urban geography chair has continued to maintain its focus on Amsterdam and its metropolitan region, thereby specializing in the labor market, the housing market, and the position of ethnic minorities. The observation by Dietvorst *et al.*, that the complexity of the big city formed a lasting research theme for human geographers at the University of Amsterdam, continues to apply today.

1970–Present

The 1960s and 1970s can be characterized by a growing number of students and, concomitantly, a growing number of chairs in human geography and, thereby, a growing range of specializations. More specializations and more staff also implied more differentiation and less clear cut local 'schools of thought'. New generations of human geographers were clearly also much more open to new developments in the international field of human geography. Especially developments in Anglo Saxon and Swedish human geography were readily absorbed, although Christiaan van Paassen observed that the appreciation of new geographical ideas was rather slight, not least in Utrecht, with the biggest human geography department. This related to a number of causes. First, most departments in the Netherlands still have a strong tendency to concentrate on local environments. This is partly not only due to the strong regional recruitment of students, but also due to the strong tendency toward applied contract research, for which commissioners often stem from the direct surroundings. Finally, appointment policies, even at present, show a strong inward looking tendency. Especially in a small academic community like that of the Netherlands, this results in a situation in which most new appointments involve persons who are also intellectually socialized in the same community, and consequently often do not have strong innovative inclinations for Dutch human geography.

Thus, traditional local specializations are much more rigid than one would expect from, otherwise fairly dynamic, academic institutions in the Netherlands. In 1984, this led Christiaan van Paassen to write how: "'the new professional guard' did not change that [the bipolar structure based on Amsterdam and Utrecht] [...] despite the challenges to academic traditions from changes within society and competition from other rapidly expanding social sciences." Thus, Amsterdam has maintained its close proximity to sociology, with specializations in fields such as political and cultural geography, while Utrecht has continued to prefer the study of economic phenomena, favoring a geographical orthodoxy. Groningen, as an offspring from the Utrecht school, can also still be typified by its strong economic focus on spatial investment and real estate policies. Nijmegen has developed along the same lines, although recently various radical changes have taken place (see below). These strong local characteristics can also still be observed in the latest self studies put together by the respective departments for the accreditation of the human geography curriculum in 2007. Notwithstanding these rigidities, from the late 1960s onward, Dutch geography has become increasingly influenced by Anglo Saxon ideas.

Along with international developments the 1970s also featured fierce, ideologically tinted scientific debates, mainly in relation to theoretical issues. Christiaan van Paassen thus reports how: "at Nijmegen, involved in the spiritual and political turmoil within the Roman Catholic world, a conference on 'critical geography' was organised in 1973; the 'Positivismusstreit' was discussed, and Habermas and Popper compared." In 1976, the KNAG organized a conference discussing theoretical reflections on the course of geography. At this event, Jansen in 1976 introduced the concept of policy oriented geography and van Paassen outlined his existentialist anthropological viewpoint. It was also Jansen who criticized work on centrality and hierarchies by Buursink, and Buit and Nozeman and advocated a phenomenological approach.

Following the development of a spatial analysis approach in human geography, the methodological sophistication of human geographic research increased substantially. The appointment of Piet Kouwe in Nijmegen and Piet Lukkes in Groningen for methods and techniques of human geographic research had already been mentioned. In Utrecht, Joost Hauer was appointed as the professor for human geographic research methodology in 1975, while Frans Marinus Dieleman (1942–2005) was appointed as professor of methods and techniques of geographical research at the Free University of Amsterdam in 1978.

Not only methodological sophistication, but also the productivity was raised substantially and specializations, such as economic geography (Utrecht, Nijmegen, Groningen, and Free University of Amsterdam), geography of development (Utrecht, Nijmegen, Free University of Amsterdam, and University of Amsterdam), and political geography (University of Amsterdam), began to flourish. The increasing numbers of chairs and staff, and the growing differentiation, also brought an end to the long standing Amsterdam–Utrecht rivalry, although they continued to maintain their own orientation.

A number of developments that have taken place since the 1970s have led to new changes in the Dutch human geographic praxis. First of all, the welfare state, which was so important for emergence of applied human geography, is on its retreat since the 1980s. The direct implication was a cutback on government spending on policy related research. At the same time, in many fields where the government reduced its activities, private companies, such as consultancy firms, commercial research agencies, developers, institutional investors, and media, stepped in, providing new work for geographers. Second, in the 1980s, the number of students enrolled in the human geography programs almost halved. As a direct consequence, several newly created human geographic specializations were closed or merged, some majors were turned into minors, and in 1984 the department of human geography at the Free University of Amsterdam was disbanded altogether. From 2000 onward, the number of students seems to have stabilized. However, fading disciplinary borders, increased dynamics in the landscape of human geography, and related programs in other disciplines following the introduction of the bachelor–master system, imply that it is difficult to predict trends in the number of students. Latest data show a slight increase in student numbers, especially in newly created master programs at the interface between human geography and other disciplines, such as development studies in Amsterdam and Utrecht and conflict management in Nijmegen.

Under the pressure to increase output ('publish or perish') Dutch human geographers increasingly resort to less problematic and easier to publish mainstream research and also, notably in comparison with other continental European countries, publish extensively in English language periodicals. Notwithstanding the high presence of Dutch geographers in English language journals, their impact has remained modest. The tendency toward the mainstream also makes it increasingly difficult to identify specific trends. Hence, de Pater and de Smidt conclude how: "Sharply honed polemics are hardly heard anymore in this more businesslike than lofty discipline." Thus, the broad mainstream is best typified as comprising geographers who combine modest forms of spatial analysis with behavioral geography.

Starting in the 1980s and culminating at the end of the 1990s, various counter currents, inclining especially toward post positivist approaches, seem to be gaining support. Thus, at the University of Amsterdam, critical geopolitics has emerged as a field in which the social construction of political spaces through discourse is studied. Partly this is due to the invitation of prominent foreign geographers and social scientists to lecture and write on the Dutch capital (e.g., Manuel Castells, John Loughlin, Edwars Soja, Lyn Lofland, Paul Claval, Ulf Hannerz, Donald Olsen, and John Mollenkopf).

It is also through new appointments that such changes can anchor. In the case of the Radboud University of Nijmegen, the appointment of Huib Ernste, who had been abroad for almost 17 years, as the successor to Jan Buursink, provided opportunity to reinvigorate an interest in theory building and fundamental research in Nijmegen, in divergence to the long Dutch tradition of applied human geographic research. Thus, Ernste introduced the phenomenologically based action theoretical approach, elaborated by Benno Werlen in German speaking human geography. This approach was then developed further in a language pragmatic fashion and contrasted with post structural approaches in a new 'multiparadigmatic' curriculum for human geography. In his inaugural lecture Ernste also tried to bridge the gap between these different paradigms by making use of the philosophical–anthropological approach of Helmuth Plessner, which resembles the anthropological approach of Christiaan van Paassen formulated 26 years earlier. To substantiate this theoretical orientation also the state of the art Alexander von Humboldt lecture series were initiated, whereby prominent geographers and social scientists are invited to Nijmegen. His interests also helped to establish the theory driven Centre for Border Research in Nijmegen, one of the main innovations of Dutch human geography, in the last decade. In the case of the University of Utrecht, the reshuffling of the economic geography section partly in relation to the creation of a new school of economics and some new appointments such as Ron Boschma and Oedzge Atzema on the one hand initiated a research program in the field of evolutionary approaches in economic geography which strengthened its status in the international academic community, and on the other hand continued the line of more applied economic geographic research. Otherwise new emphases in the field of human geography in the Netherlands are rather modest, notwithstanding the fact that the Dutch university landscape is characterized by high institutional dynamics. These institutional dynamics also did not result in a substantial increase of the number of women appointed on chairs for human geography in the Netherlands. In this sense, Dutch human geography clearly lags behind.

One of the other major driving forces of changes in human geographic curricula and research programs in the Netherlands are the regular external evaluations by an independent agency, the Quality Assurance Netherlands Universities (QANU). The research is assessed according to a protocol developed by the Royal Academy of Arts and Sciences (KNAW), the Netherlands Organization for Scientific Research (NWO), also known as the Science Foundation, and the Association of Universities in the Netherlands (VSNU). The results of these evaluations are not directly related to the funding of research or education, but indirectly, through the process of forced self reflection in the course of providing self studies as the basis for the external evaluations and through voluntary self reflection on the basis of the assessment results; they do result in increased quality and quantity of research output and they do lead to high standards in teaching. Independent of these assessments also the changing structure of Dutch research funding leads to some shifts in emphasis. As Jan van Weesep reported, research funds are increasingly being made available through the National Science Foundation on the basis of increasingly large scale programs. Another typical Dutch feature in human geographic research, which fits the traditional strong focus on applied research, is the big share of large scale programs with mixed research funding, where research funds are partly provided by private parties such as private companies and government agencies as well as by the National Science Foundation and Universities themselves. Together this has led to a tendency toward less academic freedom and more mainstream and less curiosity driven innovative research.

Development of Human Geography in Flanders

Dutch human geography is also linked to human geography in Flanders, the Dutch speaking area of Belgium. However, the development of the Belgian universities clearly also followed a different path. Human geography is taught at three Dutch speaking Belgian universities, namely, the University of Gent, the Catholic University of Louvain, and the Protestant free university of Brussels. At all three universities, however, in contrast to the

Dutch universities, geography is still taught as an inte grated discipline, comprising both physical and human geography. In general, geography at the Belgian uni versities in Flanders is less specialized than is the case in the Anglo Saxon countries, in the Northern European countries, or in the Netherlands. Partly as a result of a highly and centrally regulated curriculum until the be ginning of the 1990s, these three Belgian universities are very similar in their approach. They all use the natural sciences and mathematics as the backbone of the cur riculum which is then gradually enhanced with social scientific subjects. Human geographic research therefore is also very often focused on issues at the interface be tween physical and human geography. It therefore is also understandable that some of the best known Flemish contributions to the field of human geography stem from researchers who are active in other disciplines such as urban sociology or urban history and that there are not many sources to be found on specific schools of thought or on the historic development of human geography in Flanders (the main source being the Belgian geographic journal *Belgeo*). Flemish human geographers traditionally play an important role for Dutch human geographers in the role of peer assessors in the Dutch evaluation system. Finally, as Flemish human geographers also cooperate substantially with the French speaking human geog raphers, the Flemish human geographers embody a window between Dutch and French schools of thought in human geography.

Addendum

Introduction

South Africa's human geography features a number of specific historic developments which relate strongly to changes in the political context in which it has often found itself to be strongly embedded. The most salient change, obviously, was the gradual manifestation of Apartheid over the course of the twentieth century, and its formal demise in the early 1990s. South Africa's human geography field follows this trajectory in its manifestations to some extent. In relation to Dutch geography the immediate connection is that of language; with Afrikaans having its roots in the Dutch language while, at present, it still shares many similarities.

In the past Afrikaans was a medium used for teaching and research in human geography at various universities across the country. Therein a differentiation needs to be made between the formerly 'black' and 'white' uni versities (or, alternatively: 'formerly advantaged' versus 'disadvantaged' universities, respectively), with the for mer based in the former homelands while the latter were located in those parts of the country designated as being 'for whites only'. In a very general sense it might be argued that in the past geography departments at 'black' universities maintained an applied nature in their teaching and research, in line with the general objectives of these universities to train a black administrative elite for the governance of the homelands, while geography departments at former 'white' universities maintained a more heterogeneous curriculum and research agenda.

The second differentiation concerns the difference between the former 'white' English and Afrikaans speaking universities. At present four universities still maintain Afrikaans as a primary language of communi cation, albeit that there is variation in the extent to which this policy is implemented. These four universities are: Stellenbosch University, Potchefstroom University for Christian Higher Education, University of Pretoria, and the University of the North West (Potchefstroom). At other universities Afrikaans is not given such priority, and there English is usually the preferred mode of communication.

Hence, although South Africa does note a volume of academic work in Afrikaans within the field of human geography, there is little to argue that it is fundamentally different enough for it to be described as a separate geographical school from other practices of human geography in South Africa, notably in relation to more global practices. Thus, in the text that follows, a general, nonlingual differentiation is made of the most salient developments for human geography in South Africa.

The Origins of Human Geography in South Africa

Regarding the origins of the discipline in South Africa, Crush, Reitsma, and Rogerson note that, in essence, human geography constituted a 'colonized' discipline with a strong Euro American bias in its theoretical out look. This origin, in their view, is of significance for two main reasons.

The first concern relates to the origins of human geography and the very applicability of its 'expatriate' Euro American theoretical framework for explaining processes in South Africa. In essence this implied that social formations in South Africa were adapted to this framework, rather than the framework emerging from social formations. Yet it took until the 1980s for sufficient momentum to be built up to culminate in a new school of critical, postcolonial, post Ford human geographers.

The second issue raised by Crush *et al.* regards the essentially colonial origins of human geography for providing a scientific rationale for economic, social, but also geographic, differentiation between races in South Africa. This was the case during the nineteenth century under British colonial administration and thereafter under Apartheid policies of various white governments. Segregation in geography studies was not only manifested in the focus of research but also achieved through a differentiation between White and Black Universities. Not only did this institutionalize funda mentally different orientations, leading to a steadily in creasing rift in the field, it also marginalized the role of non white geographers. In reaction to these develop ments South African geographers found themselves in creasingly excluded from international geography exchanges.

Post-Apartheid: Toward a New Human Geography Approach

In the wake of the political turmoil of the 1970s and 1980s, human geography began to take on an increasingly critical political, even activist, stance, committing itself increasingly to the concerns of the indigenous, that is, black, majority of South Africa, doing so through its 'geographies of protest'.

The 1990s sees the rise of a number of parallel de velopments. Thus, one substantial group of human geographers take on positions of an applied nature, using multidisciplinary, multifaceted approaches in their work to study changes effectuated by governments, but also those brought about by social movements such as labor unions, civic and other community based organizations, urban and rural. Other geographers, notably historical geographers, have taken on a postcolonial research agenda, following structuralist Marxist paradigms to take critical distance from their predecessor geographers' liberal historical stance.

One of the most prominent liberal historical geog raphers of the past was A. J. Christopher. While generally praised for his pioneering work of the 1970s, mapping racial segregation and exclusion throughout South Africa, Christopher was also heavily criticized. This mainly re lated to his neutral, liberal stance toward the extent to which South Africa's government had succeeded in seg regating its population along racial lines, privileging its white population while marginalizing and oppressing others. Thus, Crush argues that Christopher, and other historical geographers of the same period, failed to en gage with the radical changes brought about in South Africa, instead producing: "antiseptic landscapes of much existing historical geographic writing."

Future Prospects for South African Human Geography

Concerning the road ahead for human geography in South Africa, Rogerson, an eminent human geographer at Witwatersrand University, calls on his fellow geographers to engage in a concerted effort to critically explore contemporary social and economic reconstruction pro cesses and the developments these produce, notably the reconfiguration of economic and social spaces in relation to the rural–urban divide. Thereby human geography will also need to engage with other disciplines, notably development studies, building on their efforts. The focus advocated by Rogerson is presently finding increasing impetus and is, for instance, reflected in the work of Brij Maharaj, who studies entitlement conflicts in the muni cipality of Durban.

In response to the need for a critical South African human geography and a truly post Apartheid situation black human geographers will need to be stimulated to take on a stronger role. At present they are strongly underrepresented in scholarly output. Thus, Sumner finds that of all scholarly output from South African geographers only 22% features black authors (with 69% of these publications having a black author as first or single author). In addition, there is also general consensus that South African geographers need to also expose their work more at a global level. This, for instance, re quires them to more present at global forums and international conferences, and to publish more outside local journals, notably the *South African Geographical Journal (SAGJ)*.

See also: Francophone Geography; German-Language Geography.

Further Reading

Beavon, K. S. O. and Rogerson, C. M. (1981). Trekking on: Recent trends in the human geography of southern Africa. *Progress in Human Geography* 5, 159 189.

Boerman, W. E. (1960). *Vijftig jaren economische geografie. Tijdschrift voor Economische en Sociale Geografie (TESG)* 51(4/5), 85 89.

Bosman, J. and de Pater, B. (1998). The international position of Dutch human geography: A quantitative exploration. *GeoJournal* 45(1 2), 85 96.

Bremen W.J. van den De Nederlandse geografie van 1850 tot 1950. In Dietvorst A.G.J., Ginkel, J. A. Van, Kouwenhoven, A. O., Pater, B.C. (eds.) *Algemene sociale geografie: ontwikkelingslijnen en standpunten*, pp 160 174. 1984 Weesp: Romen.

Brink, P. P. W. J. and Moree, P. (1996). *Met kapmes en kompas. Vier eeuwen Nederlandse ontdekkingsreisen en reisverslagen*. The Hague: Koninklijke Bibliotheek.

Buit, J. and Nozeman, E. F. (1976). *De Gewenste Spatiering van het Voorzieningenapparaat*. Amsterdam: Geografisch en Planologisch Instituut, Vrije Universiteit.

Buursink, J. (1971). *Centraliteit en Hierarchie*. Assen: van Gorkum.

Buursink, J. (1977). *De Hierarchie van Winkelcentra*. Groningen: Geografisch Instituut, Universiteit Groningen.

Buursink, J. (1998). *Nederland in geografische handen. Honderd jaar regionale geografie van Nederland*. Utecht: Koninklijk Nederlands Aardrijkskundig Genootschap (KNAG).

Crush, J. (1986). Towards a people's historical geography for South Africa. *Journal of Historical Geography* 12(1), 3.

Crush, J. (1991). The discourse of progressive human geography. *Progress in Human Geography* 15, 395 414.

Crush, J. (1992). Beyond the frontier: The new South African historical geography. In Rogerson, C. & McCarthy, J. (eds.) *Geography in a Changing South Africa: Progress and Prospects*, pp 10 37. Cape Town: Oxford University Press.

Crush, J., Reitsma, H. and Rogerson, C. (eds.) (1983). *Special Issue: Decolonizing the Human Geography of Southern Africa. Tijdschrift voor economische en sociale geografie* Lxxiii, 4.

de Pater, B. (2001). Geography and geographers in the Netherlands since 1870s: Serving colonialism, education, and the welfare state. In Dunbar, G. S. (ed.) *Geography: Discipline Profession and Subject since 1870. An International Survey*, pp 153 190. Dordrecht: Kluwer.

Dieleman, F. (1998). Research progress Dutch style. *Environment and Planning A* 31, 191 195.

Dietvorst, A. G. J., van Ginkel, J. A., Kouwenhoven, A. D., de Pater, B. C. and van den Bremen, W. J. (1984). *Algemene sociale geografie: Ontwikkelingslijnen en standpunten*. Weesp: Romen.

Ernste, H. (2002). Transgressing borders with human geography. *Inaugural Speech*. Radboud University, Nijmegen, 27 June.

Ernste, H. (2004). The pragmatism of life in poststructuralist times. *Environment and planning A* 36(3), 437 450.

Ernste, H. (2007). The international network university of the future and their local and regional impacts. In Scott, A., Harding, A., Laske, S. & Burtscher, C. (eds.) *Bright Satanic Mills: Universities, Regional Development and the Knowledge Economy*, pp 69 92. Aldershot: Ashgate.

Fairhurst, U. J., Davies, R. J., Fox, R. C. *et al*. (2003). Geography: The state of the discipline in South Africa: A Survey, 2000/2001. *Report Compiled for the Society of South African Geographers*.

Ginkel, H. van (1996). Professional practice and the development of modern geography in the Netherlands. In Berdoulay, V. & van Ginkel, J. A. (eds.) *Geography and Professional Practice. Nederlandse Geografische Studies*, vol. 206, pp 31 45. Utrecht: KNAG.

Ginkel, H. van and Smidt, M. de (eds.) (1983). *75 years of human geography at Utrecht*. *Tijdschrift voor Economische en Sociale Geografie (TESG)* 74(5), 313 406.

Giraut, F. and Maharaj, B. (2002). Contested terrains: Cities and hinterlands in post apartheid boundary limitations. *GeoJournal* 57, 39 51.

Harvey, D. W. (1984). On the history and present condition of geography: An historical materialist manifesto. *The Professional Geographer* 36, 1 11.

Heinemeyer, W. F. (1971). Reilingh en de sociale geografie. *Geografisch Tijdschrift* 5(4), 291 303.

Heslinga, M. W. (1978). *Geography and Nationality. The Estyn Evans Lecture 2*. Belfast: Department of Geography, The Queen's University.

Heslinga, M. W. (1983). Between German and French geography. In search of the origins of the Utrecht School. *Tijdschrift voor Economische en Sociale Geografie (TESG)* 74(3), 317 334.

Heslinga, M. W. (1999). Grondleggers. In de Pater, B. (ed.) *Een tempel der kaarten. Negentig jaar geografiebeoefening aan de universiteit Utrecht*, pp 34 73. Utrecht: Faculteit der Ruimtelijke Wetenschappen, Universiteit Utrecht.

Jansen, A. C. M. (1977). *Over Armoede van Distributie Planologisch Onderzoek*. E.G.I. paper No. 14. Amsterdam: Economisch Geografisch Instituut, Universiteit van Amsterdam.

Jansen, A. C. M. (1980). Een fenomenologische orientatie in de ruimtelijke wetenschappen. *Geografisch Tijdschrift* 14, 99 111 and 324 329.

Keuning, H. J. (1955). *Mozaïek der functies: Proeve van een regionale landbeschrijving van Nederland op historisch en economisch geografische grondslag*. The Hague: Leopold.

Knippenberg, H. and van Schendelen, M. (eds.) (2002). *Alles heeft zijn plaats. 125 jaar Geografie en Planologie aan de Universiteit van Amsterdam*. Amsterdam: Aksant.

Lambooy, J. (1973). Economic and geonomic space. *Papers of the Regional Science Association* 31, 145 158.

Lambooy, J. (ed.) (1978). *Beleidsgeorienteerd Onderzoek ter Diskussie*. E.G.I. Paper No. 17, Amsterdam: Economisch Geografisch Instituut, Universiteit van Amsterdam.

Lambooy, J. (1981). *Ekonomie en ruimte: Inleiding in de ekonomische geografie en de reginale ekonomie*. Assen: van Gorkum.

Lester, A. (2003). Space, place, landscape and identity: Historical geographies of southern Africa. *Journal of Southern African Studies* 29(3), 595 613.

Livingstone, D. N. (1992). *The Geographical Tradition. Episodes in the History of a Contested Enterprise*. Oxford: Blackwell.

Musterd, S. and de Pater, B. (2003). Eclectic and pragmatic: The colours of Dutch social and cultural geography. *Social and Cultural Geography* 4(4), 549 563.

Ormeling, F. J. (1974). In memoriam Prof. Dr. Jan O. M. Broek. *Geografisch Tijdschrift* 8(5), 403 406.

Pater, B. de (2002). *R. H. A. Cools vertolker van de Franse geografie. Geografie* 11(1), 16 17.

Pater, B. de and de Smidt, M. (1989). Dutch human geography. *Progress in Human Geography* 13(3), 348 373.

Paassen, C. van (1962). Geografische structurering en oecologisch complex, een bijdrage tot sociaal geografische theorievorming. *Tijdschrift van het Koninklijk Nederlands Aardrijkskundig Genootschap* 79(4), 215 233.

Paassen, C. van (1976). Human geography in terms of existential anthropology. *Tijdschrift voor Economische en Sociale Geografie (TESG)* 67, 324 341.

Paassen, C. van (1984). Human geography in the Netherlands. In Johnston, R. J. & Claval, P. (eds.) *Geography since the Second World War. An International Survey*, pp 214 234. London: Croom Helm.

Pellenbarg, P. (1999). Vijftig jaar geografie in Groningen. *Geografie* 8(1), 43 44.

Pirie, G. H. (1985). Geography in South Africa. *Professional Geographer* 37(4), 479 482.

Rogerson, C. M. (1990). A new South Africa   a new South African geography? *GeoJournal* 22(3), 227 229.

Rogerson, C. M. (1993). South African geography and post apartheid construction. *GeoJournal* 30(1), 5 7.

Schrader, R. (1974). Honderd jaar Koninklijk Nederlands Aardrijkskundig Genootschap 1973 1973. *Geografisch Tijdschrift. Nieuwe Reeks* 8(4), 1 164.

Smidt, M. de (1983). Dutch economic geography in retrospect. *Tijdschrift voor Economische en Sociale Geografie (TESG)* 74, 344 357.

Steinmetz, S. R. (1900). *Classification des types sociaux et catalogue des peuples. L'année sociologique*, 43 144.

Steinmetz, S. R. (1913). Die Stellung der Sociographie in den Geisteswissenschaften. *Archiv für Rechts und Wirtschaftsphilosophi: mit besonderer Berücksichtigung der Gesetzgebungsfragen* 6(3), 492 501.

Sumner, P. D. (2006). On the nature of southern African geographical publications, 2000 to 2004. *South African Geographical Journal* 88(22), 213 218.

Vlaamse Interuniversitaire Raad (ed.) (2002). *De Onderwijsvisitatie Geografie. Een onderzoek naar de kwaliteit van de opleidingen Geografie aan de Vlaamse universiteiten.* Brussels: Vlaamse Interuniversitaire Road.

Vries Reiling, H. D. de (1961). De sociale aardrijkskunde als geesteswetenschap. *Tijdschrift van het Koninklijk Nederlands Aardrijkskundig Genootschap* 78, 112 122.

Vuuren, L. van (1924). Practische geografie. In van Balen, C. L. (ed.) *Geografische bijzonderheden. Anniversary Volume for the 70th Birthday of R. Schuiling*, pp 346 353. Groningen: Noordhof.

Weesep, J. van (1987). Academic urban geography in the Netherlands. *Urban Geography* 8(1), 65 76.

Weesep, J. van (2006). Research assessment in the Netherlands. In Castree, N. (ed.) Research assessment and the production of geographical knowledge. *Progress in Human Geography* 30(6), 747 782.

Wellings, P. (1983). Review of 'decolonizing the human geography of southern Africa'. *Journal of Modern African Studies* 21(2), 360 364.

Werlen, B. (1988). *Society, Action, and Space. An Alternative Human Geography*. London: Routledge.

Werlen, B. (1995). *Sozialgeographie alltäglicher Regionalisierungen. Band 1: Zur Ontologie von Gesellschaft und Raum*. (Social Geography of Daily Regionalisations. Vol. 1: Ontology of Society and Space.) Stuttgart: Steiner.

Werlen, B. (1997). *Sozialgeographie alltäglicher Regionalisierungen. Band 2: Globalisierung, Region, und Regionalisierung*. (Social Geography of Daily Regionalisations. Vol. 2: Globalisation, region, and regionalisation.) Stuttgart: Steiner.

Werlen, B. (ed.) (2007). *Sozialgeographie alltäglicher Regionalisierungen. Band 3: Ausgangspunkte und Befunde empirischer Forschung*. (Social Geography of Daily Regionalisations. Vol. 3: Assumptions and results of empirical research.) Stuttgart: Steiner.

Wusten, H. van der (2005). Sebald Rudolf Steinmetz 1862 1940. In Armstrong, P. H. & Martin, G. J. (eds.) *Geographers. Biobliographical Studies*, vol. 24, pp 109 123. London: Thoemmes Continuum.

Zierhofer, W. (2002). Speech acts and space(s): Language pragmatics and the discursive constitution of the social. *Environment and Planning A* 34, 1355 1372.

Relevant Websites

http://fmg.uva.nl/home.cfm
University of Amsterdam, Faculty of Social and Behavioral Sciences

http://www.rug.nl/
University of Groningen

http://www.uu.nl/
University of Utrecht

Dwelling

O. Jones, CCRI, Cheltenham, UK

Glossary

Affect The vital systems of life beyond or pre rational, reflexive thought and language use.
Agency The capacity to act creatively.
Becoming A view of life as always in creation and thus always in motion and only possessing immanent and emergent form rather than static, predetermined form.
Embodiment An approach which takes bodies seriously as against the mind–body dualism.
Hybridity The understanding that much of life is only possible through routine interconnections of differing types of things and the idea that bodies themselves are not singular and pure but multifaceted in form, origin, function, and identity.
Taskscape Landscapes and places made by the repeating actions of organisms over time.

Introduction

Dwelling approaches life as a process of being in the world which is open to the world. Human and nonhuman life is read as an immediate, yet also enduring, relational process of bodies in environment (space and place) which are (with variation) mobile, sensing, engaging, responding, exchanging, making, using, remembering, and knowing. This process based, vitalist view of life is closely linked to phenomenology, and stands in opposition to dualized, rational Cartesian based approaches.

As Tim Ingold suggests when setting out Heidegger's foundational work on this concept, dwelling encompasses the whole manner in which one lives one's life on the Earth; thus "I dwell, you dwell" is identical to "I am, you are." For Heidegger, to be a human means to be on the Earth as a mortal. It means to dwell. This offers an alternative to Descartes' *cogito* – 'I think, therefore I am' – the foundation of rationalism which helped set up/reinforce a series of persistent, related dualisms including those of mind–body, subject–object, society–nature, and human–nonhuman. These, some argue, are impediments to understanding the relational nature of life, not least in terms of important questions of humans in nature/environment/place, and environmental sustainability.

Dwelling is thus offered as an alternative ground from which human and nonhuman life can be rethought. Given its focus on the body, which is inevitably in place and at the center of its own space, it has been an important theme within geography for those thinking about space, place, and landscape, and within related disciplines such as sociology, anthropology, archaeology, and landscape studies, which also address questions of society, nature, and environment. In recent times, there has been an upturn of interest in dwelling as various disciplines seek to develop richer, relational, process based notions of life in terms of ecological relations, hybridity, affect, posthumanism, performativity, and nonrepresentation.

Dwelling differs from social construction approaches to life/nature through its stressing of the physical, the relational, the senses, the orchestration of body and space/environment. It has more affinity with other relational process views such as actor network theory (ANT). It differs in that it retains a topographical interest in space as place and the space that moves with and around bodies, whereas ANT takes a more technogrammatic, topological approach to unfolding life.

Dwelling takes an ecological view of how humans and other animals make and inhabit 'life worlds' through registers of specific bodily practice. It can be used to consider places and landscapes as temporal extensions and entanglements where all manner of beings, things, and processes come into specific relations and settle out into the varied patterning that marks the world.

Foundations

The genesis of the concept, as discussed here, can be precisely located in Heidegger's (**Figure 1**) essay *Building*

Figure 1 Martin Heidegger (1889 1976).

Dwelling Thinking. This was one of the seven essays in the work *Poetry, Language, Thought*, but the notion of dwelling captures much of the ambitions in Heidegger's later work. It was thus part of the so called 'turn' (*Krhre*) in Heidegger's thought through which he moved away from some aspects of the position of *Dasein* set out in his earlier work. *Dasein*, as is dwelling, is essentially a conception of being in the world and being inseparable from the world. It attempts to create a new ontology of being which more or less reverses and reconnects the way in which the self engages with the world as presented by Cartesian ontology.

In the later work the lived, temporal, mortal nature of being continues to be a central theme but the focus is more strongly on how being, and its spaces as places, are always 'emergent, relational processes' and thus continuous or chestrations of body, environment, culture, and time.

The turn to the 'dwelling perspective' is often ex plained in comparison to the 'building perspective'. The latter stands accused of seeing human thought and action as somehow isolated from the world and thus in a pos ition to impose on the world, to, in some cases, literally build upon it ideally generated blueprints. The dwelling perspective points out that any form of life emerges from the world and there is never a gap, or break, in which thought and practice can completely free itself. Any building, including rational thought and knowledge, emerges from dwelt life. The subject–object dualism is scrambled and objective 'views from nowhere' impossible.

The continuities of dwelt life can be seen in Heidegger's most famous illustration of the concept. This describes a 200 year old peasant farmhouse in the German Black Forest. It is built on the lee of a hill, facing South, near meadows and a spring which supply sus tenance. The form of the roof (pitch and overhanging eves) is shaped to shelter the house from winter snow and storms. The interior, with its rooms, artefacts, beds, tables, cots, and shrines, is expressive of the lived lives. Dwelt life is articulated and practiced though continu ities of elements, bodies, processes, materials, artifacts, and actions.

Nigel Thrift suggests that turn to dwelling can be understood as a turn from thinking about space as an abstract, *a priori* ground to be filled by life, to thinking about place, in which relational transactions bring spaces, always specifically articulated, into being. These alter native views of space as spacing are of obvious interest to geography and reflect divisions that have, and still do exist within it.

Things are critical to dwelling, and in a sense have their own being. They 'gather or unite'. To illustrate this, Heidegger uses another oft cited example of a bridge, which, he suggests, does more than simply connect two banks of a river. Rather it creates a place, and rear ticulates the spaces and places, and the lives therein, around it. This quite literal example should not obscure the idea that all things are connectors and assemblers, Heidegger reminds us that the old German word for 'thing' meant as such. The plethora of things which fill the world are tools by which selves create dwelt lives.

Other key elements are the decentering of the modern human subject and a concern for nature and the en vironment as an active field. Heidegger states that the Earth is the serving bearer, blossoming and fruiting, spreading out in rock and water, rising up into plant and animal and should be 'spared'. This sparing of the Earth, or of the things of the Earth, is letting something be the way it is in itself. Such a letting be is sharply contrasted to the modern domination of things/nature. Not by controlling or ruling, but by respectful sparing, can we be open to the Earth appearing in its gifts. Here, we see a proto version of radical ecology, and indeed, some en vironmental philosophy does make specific connection to Heidegger's work.

Heidegger's notion of dwelling has poetic and spiritual dimensions, seeing continuities of dwelt life as preserving 'the foursome' – the unity of Earth, sky, gods, and mortals. Heidegger's desired authentic life requires these mystical elements which are best expressed and judged by poetic and spiritual vision. Dwelling is to be on the Earth as a mortal, cultivating, constructing, and cherishing, but this needs to be framed, judged, and given meaning by and within the vastnesses of time, Earth, sky, and gods.

Heidegger considered that modernity, in the form of industrialized capitalism, consumerism, and the spec tacular society, forcibly detached the self from of the world and thus stifled true dwelling. He was highly critical of the technological age and its conception of, and use of, things. Think of the contrast between craft, say a carpenter – working with well honed tools and skills to make things out of an intimately known material, in a process where skilled movement, material quality, and technology are harmoniously orchestrated and expressed in the process and the outcome – with industrial mass production where workers might do repetitive, possibly unskilled, isolated (from the wider process/outcome) tasks. The poetry, soul, and (harmoniously) dwelt life is missing from the latter. As we will see, later adaptations of his ideas rejected this split between authentic (dwelt) life and inauthentic (detached) life yet retained the central idea of being in the world.

Post-Heideggerian Dwelling. Temporalities, Landscapes, Taskscapes, and Politics

Heidegger's ideas have been hugely important in (re) shaping ontologies of life and in understandings of how life is lived in space, place, and landscapes as they have been carried forward in various disciplines. Significant developments in anthropology, archaeology, and soci ology have been underpinned by the work of Tim Ingold, Barbra Bender, and others. Ingold's influential work on dwelling is set out in two main articles 'The temporality of the landscape' and 'Building, dwelling, and living', and four other related articles, all now collected together under the heading 'Dwelling' in a collection of his pub lished works.

Ingold takes Heidegger's notion of dwelling and combines it with the related phenomenology of Merleau Ponty, elements of ecological physiology, and Bordieu's theory of practice. For Ingold, dwelling is a perspective that treats "the immersion of the organism person in an environment or life world as an inescapable condition of existence." The world continually comes into being with and around the inhabitant, and its constituents take on significance through their incorporation into a regular pattern of life activity. There is an overt continuation of Heidegger's dwelling, but there is also an important difference. In Heidegger, dwelt life was considered to be eroded to the point of extinction by modernity. For Ingold, "dwelling is the inevitable form of life. All living things dwell, thus it becomes universally applicable to human and nonhuman life."

The qualities of each dwelt life might be markedly different, but it is not a case of dwelt or undwelt life. All of life and its skills and practices stem from an ecology of action which is emergent, relational, and extended through 'hybrid' articulations of bodies/materials/prac tices. Life and its spacings are always unfolding. We grow up in the world and inevitably it enters us. Ingold describes how human children, like the young of many other species, grow up in environments furnished by the work of previous generations, and as they do so they come to carry the forms of their dwellings in their bodies – in specific skills, sensibilities, and dispositions.

Ingold's primary concern has been to move under standings in archaeology and sociocultural anthropology from the 'building perspective' to the 'dwelling per spective'. To do this, Ingold considers the buildings/ architecture of humans and nonhumans as extensions of Von Uexkűll's life worlds, which suggest that each life form dwells in a distinctive way and from this, practices and forms (of whatever kind) stem. In doing so, the gulf between human and nonhuman life, for example in terms of architectures, is erased. Both emerge out of lived life and are ongoing processes (albeit articulated in solid material form).

As it is centered on ongoing emergence and its resi dues, dwelling inevitably brings temporality and process to the center of considerations of places and landscapes which are most obviously understood as spatial for mations. Only by engaging with the temporal dimension of practice can one begin to understand the natures of current landscapes and how they are immanent to their emergence rather than expressions of some fixed pre formation. Specifically situated practice is all there, and there is no generality apart from that.

Ingold illustrates his arguments with an analysis of Pieter Bruegel the Elder's painting *The Harvesters* (**Figure 2**).

In the foreground of the painting is a pear tree. Around it some people are harvesting a field of grain, while others eat and rest in its shade, or are moving to and fro along a path. A church stands half concealed in the background. A vista of landscape stretches away from

Figure 2 *The Harvesters;* Pieter Bruegel the Elder.

the high on a hill vantage point. Other distant churches (markers of villages), fields, and woods can be seen.

Ingold considers the painting from the perspectives of time, actions (repeated), and the specificity and the relational agency of the actors (human and nonhuman) depicted. In effect, Ingold treats the landscape in the same way that Heidegger considers the peasant house in the Black Forest, thinking about how (dwelt) life flows in and through the elements in the scene.

Agency is distributed between the humans and nonhumans depicted. There is an ongoing interplay of relations between elements, bodies, and time. The people make the landscape through their bodies, treading the path by repeated walking. The landscape marks them and their practices, such as their walk being shaped by the pull or push of the hill as they move to and fro, or the actions of swinging a scythe. The shape of one moment of the social that the painting depicts – the seated group eating – is defined by the pool of shade made by the relationship of the earth, the tree, and the sun.

Although life is emergent, a becoming in the ever ongoing moment of the present, this takes place within the (material, emotional, cultural) accretions of past moments thickened and made enduring in various ways. The emergent life of the tree is mapped in the precise configuration of its branches. Like the churches in the background, the tree not only embodies and marks the temporal depth of the landscapes, it makes or gathers the place/space around it. Dwelt life is thus both in the now and in deep(er) time where past and present are blurred together, not least through materiality, memory, and (habitual) embodied practice.

Heidegger's example of a peasant house and then Ingold's use of this picture has seemingly rendered dwelling as an expression of authentic rustic (past) life. This has obscured its relevance to other forms of life which, on the face of it, seem markedly different (e.g., contemporary urban life which is addressed later in this article).

To capture the relational, emergent, creation of landscape, Ingold coins the term 'taskscape', where the spatiotemporal patterns of the environment take settled physical form. But this is to make the point that all landscapes (and places) are taskscapes (both human and nonhuman) – the distinction between them (landscape and taskscape) is ultimately dissolved. Cities, countrysides, wildernesses, etc., are all 'hybrid' taskscapes where human and nonhuman are bound into specific relational, dwelt, spatiotemporal patterns.

Recent eminent sociological approaches to nature and landscape have made this post Heideggerian form of dwelling the theoretical hub of understanding of nature–society relations as articulated in contested landscapes; Macnaghten and Urry suggest that dwelling helps overcome the conflicts between 'realist' and 'idealist' approaches to nature and environment. They emphasize that it is specific, dwelt, social practices which produce and transform different natures and different values. These are spatially and temporally distributed tasks; the past is continually redefined in terms of the present and landscapes are never completed but are always subject to contestation and renegotiation.

This highlights the important point that although Heidegger and Ingold illustrate ideas of dwelling in apparently harmonious, authentic, rural settings, this does not mean that power, contestation, economy are necessarily excluded. A political economy analysis is also possible of the Bruegel painting, and obvious questions of power, ownership, unjust labor conditions might well arise, but these do not negate the relationships that Ingold highlights; however, these questions might make them more complex, multiple, and contested. Ingold has recently acknowledged that his development of dwelling lacked overt political analysis and seeks to address this in more recent articles.

Drawing upon Macnaghten and Urry's work, Adrian Franklin places dwelling at the heart of a new anthropology of nature which breaks through obstacles in structuralist approaches to nature which ignored practice, practical knowledge, things, technologies, and embodied sensual experience. Franklin seeks to push dwelling toward a more animated, turbulent vision of the world unfolding in a burgeoning, far reaching (in time and space) interfolding of processes. He stresses the material and the technological as woven into 'social' fabrics and concludes that the world looks different this way and to see it and sense it requires a more active study of engagement with the world.

The basic building blocks of this new anthropology are unmediated perceptual knowledge, practical experience and knowledge of the world, the technologies that link humans and nonhumans, the aesthetic and sensual composition of experience, and the cultural choices that are made in reference to these. These kinds of interpretations and uses of dwelling link into current interests in geography in affect, performativity, and nonrepresentation which are considered below.

So, through the deployment of dwelling, nature and the social (understood in an embodied, affective sense) are refolded into relational co constituting patternings where the emergent territorializations become places which are physically manifest topographical patterns, but which remain open to the surrounding world and to culture, politics, and economy.

As we have seen, ideas of Heideggerian and post Heideggerian dwelling have often been discussed and illustrated through and in relation to ideas of rural landscapes, home, nature, localness, and rootedness. This sits uneasily with the apparently mobile, speeded up, stretched out urbanized nature of much of contemporary life. But in such life, embodied emersion in an environment, and time/space deepened experience re mains inevitable. New forms of repeated practices and encounters still abound, such as commuting to work, which may be seen as urban taskscapes. Examples of life without place(s) (home, work, cities), rhythm, and repe tition are rare or non existent in large part because of the very nature of our nature. Ideas of human and nonhuman dwelling which center on the life worlds made by body in environment seem to be applicable to contemporary urban life as anywhere else.

Dwelling and Geography

Heideggerian and post Heideggerian dwelling has influ enced various strands of human geography over the last half century or so. The former has been deployed in ongoing streams of phenomenological and humanist geographies; both in more recent geographies seeking to develop posthuman geographies informed by post structuralism, ANT, relationality, nonrepresentation emo tion, and affect. Both these streams are seeking to establish alternative ontologies and epistemologies of being on which geographical understandings of life are based. In addition to this, and more specifically, Heideggerian and post Heideggerian dwelling have offered very different approaches to landscape to those within mid–late twen tieth century geography, but now, in conjunction with the above, these are coming together in exciting new accounts of landscape as described below.

Humanistic Geography and Phenomenology

From the 1970s onwards, humanistic geography sought to develop understandings of social life embedded in sub jectivity, space, place, and everyday life. In part, this was to counter the prevailing geometric paradigm with its concerns for abstract spatial relations which largely ig nored subjective human experience and the emergent nature of knowledge/practice. Some geographers, for example John Pickles, turned to phenomenology to underpin the approach, drawing particularly on the work of Husserl, while others turned to Merleau Ponty and elements of existentialism. This shared much with Heidegger's approach centered on being in the world and focused very much upon how the world makes itself known, and how it can be made known through study. There are differences between Husserl's and Heidegger's positions and some geographers sought to base their phenomenological approaches more directly on Heidegger, suspicious of the desire to reveal a universal and *a priori* nature of the life world, in the former.

Place became a central focus of some key expressions of phenomenological and humanist geographies. Seamon and Mugerauer's *Dwelling, place and environment* published in 1989, refers back to Heidegger, Bachelard, the phenomenologist architecture of Christian Norberg Schulz, and pioneering geographers in this field – Yi Fu Tuan, Anne Buttimer, and Edward Relph for whom nature, place, and architecture were key experiential milieu of human life. Within these broad aims, various foci were to be found: Seamon sought to understand the experiential and contingent character of place through movement (anticipating the movement in Ingold's account of tasks cape); Relph deployed Heidegger's dwelling to escape from crude notions of place as location; Tuan's highly influential work explored distinctions between space and place and notions of 'topophilia' – the affective bond between people and place which shared much with Heidegger's idea of dwelt life. These, and other, approaches of place have been usefully summarized by Tim Creswell.

Geography and Landscape

Heideggerian and post Heideggerian dwelling shares some traits with the Sauerian approaches to landscape which were prominent in mid twentieth century geog raphy in that both are interested in interrelations be tween humans and the environment. However, whereas Sauer developed an approach aimed at scientific mapping of the transformation of natural landscapes into cultural landscapes, dwelling sees the production of landscape as more thoroughly relational, and not on some uni directional trajectory of nature to culture. This is a dualism in which dwelling does not even acknowledge.

Later cultural geographies of landscape, which 'read' landscapes as constructed texts and expressions of power (colonial, gender, economic, ideological), differ from dwelling in that they do not focus on the emergent coming into beingness of landscape in which the prop erties, vitalities, agencies, trajectories, and velocities of humans and nonhumans (rocks, weather patterns, plants, animals, people, technologies) combine to produce a unique chemistry out of which particular patterns of landscape, place, and life form.

In recent works, new geographical approaches to landscape and place have merged with wider ontological and epistemological questions which have adopted dwelling as part of their theoretical armory. This is be cause the latter are interested in bodies in place/space and bodies in ongoing relational action 'as central elements of being' and thus have often been discussed in the context of rethinking landscape and place.

Ecologies of Places and Landscapes

Dwelling is being deployed within poststructuralist and posthumanist geographies to counter, or develop, cultural approaches which were deemed too constructionist and too centered on the narrowly defined social. In some cases, these deployments of dwelling have harked directly back to the work of Heidegger, but they have also drawn upon post Heideggerian dwelling, notably that of Ingold. There has been a good deal of exchange between geog raphy, sociology, and landscape studies as these new configurations have developed.

Nigel Thrift and others frequently return to dwelling to remind us of the basic principle of life being an em bodied, embedded, relational unfolding process. This manoeuvre is included in recent attempts to rejuvenate ideas of place in geography in which a move from the building perspective to the dwelling perspective is cen tral. Thrift discusses ecologies of place, where places are performed inter being/material processes which coalesce into known/knowable spatial forms but remain fully wired into the flux of the wider world. Dwelling also brings the past, memory, 'hauntings', and the contin gencies of shared life into understandings of everyday life/space, which more topological approaches, such as ANT, find hard to pick up. For David Harvey the im portance of Heidegger's notion of dwelling is that it re minds us of how places are constructed in our memories and affections through repeated encounters and complex associations, and how place experiences are necessarily time deepened and memory qualified. These new dir ections are important because the very idea of place has come under critical scrutiny in some poststructuralist/ ANT based geographical thinking and under pressure from ideas of globalized flows, and yet, they remain a critical aspect of everyday life, not least because of the way in which life is dwelt life.

Geographies based upon Heideggerian dwelling (e.g., David Seamon and Robert Mugerauer) tried to show that dwelling can deal with interconnection and flows in and out of 'place' as well as connections within place, and the newer deployments of dwelling continue in this vein. In work on places and landscapes where trees are present, Cloke and Jones address a number of questions within the dwelling perspective. First, they problematize the dwelt/ authentic–undwelt inauthentic life dualism which is present in Heidegger and, to some extent, subsequent accounts of dwelling, by showing that the taskscape of an orchard is a highly complex, hybrid assemblage of old and new practices/technologies, material, and imagina tive presences. Second, they problematize the dualism between bounded space and network, by showing how the orchard is simultaneously articulated in both senses and could only be understood as such. Third, they challenge the idea of landscape as a fixed/framed view by placing the embodied (moving) witness within the ever changing landscape of movement, sound, light, and smell. Dwelling implies being in the landscape and once that is allowed, then there is all the business of the mo ment to consider; for example, the dynamics of weather and light, as well as the inevitable momentum of the past, and this goes for all who experience the landscape – visitors, workers, owners, and researchers.

Performativity and Nonrepresentation

Dwelling has much affinity with artistic, poetic, and lit erary mappings of being in place/landscape, as the wri ter Jonathan Bate shows, and thus will be of interest in relation to the increasingly evident intersections of geographical and artistic interests. Recently some geog raphers have developed a more performative account of dwelling. David Crouch, referring to Ingold's practical ontology in the world, feels that understandings of the production of nature (in the form of gardening) need to focus more strongly on the performative practice of doing. More emphasis is placed on the moment of emergence rather than on the field of emergence.

John Wylie suggests that Ingold's approach offers a potentially fruitful means of reconfiguring cultural geographies of landscape within the ambit of embodied practice and performativity but that it needs adjusting in three ways.

First, the temporal aspects of dwelling need to take more account of fleeting and mobile experiences, as well as stressing known, habitual landscapes. He suggests that dwelling as a form of phenomenological ontology should be the milieu for material cultures and ways of being that are productive of multiple spatialities and temporalities, longstanding and momentary, rural and urban, fixed and mobile, coherent and imaginary. In particular, it must enable the register of the transient and the fleeting as well as the enduring.

Second, attempts to move away from the visual as the way of engaging with/reading the landscape need to be tempered with the idea that the visual – the gaze – is itself an embodied practice.

Third, geography needs to develop methods appropri ate to a scripting and interrogation of embodied experi ences. These developments, along with Harrison's recent exploration of Levinas's view of dwelling with a concern for its ethical resonances, mean that ideas of dwelling within geography have again a renewed momentum.

Nonhuman Dwelling

One very interesting and important development in no tions of dwelling is the use of the concept to consider the lives of nonhumans. These often draw upon Von Uexküll's idea of life world to stress the 'intelligences' of dwelt animal life, how they are markedly various, and how they intersect; all this with the view of under standing human dwelling. Sarah Whatmore and Steve Hinchcliffe in their work on cities which are 'habitable' (for humans and nonhumans alike) consider animals and their urban dwellings as more than human agents exer cising unbidden, improvised, and sometimes, disruptive energies.

A recognition of dwelt animal life worlds has signifi cant implications for understanding the being of animals.

As the philosopher Raimond Gaita points out, an animal's intelligence is entirely active, its understanding entirely practical. In other words, animals practice embodied, practical, enacted knowledges – as self, being, and 'voice'. This challenges anthropocentrism and anthropomorphism. We all dwell, and there are commonalities which can be considered. It also links directly to questions of animal welfare. It is of note that organizations such as Compassion in World Farming see the Cartesian architecture of modern knowledge as culpable for denying animal rights. The appeal for domesticated, zoo, and companion animals to be able to express their natural behavior as much as possible is, in effect, an appeal to let them dwell in their environment, and thus create their life worlds and practice their enacted being. Such arguments have been developed by Jones as a way of countering ANT ideas of symmetry and purely relational identity in which the dwelt becomings of animals might get lost.

Final Points

It is the recognition in dwelling of the rich, intimate, creative ongoing togetherness of beings and things, the recognition of agency other than human, the recognition of time deepened processes, embodied experience, the experience of rootedness, the richness of things together over time, and the recognition of specificity, which makes it such a potent approach in posthumanist geography.

Dwelling is one conceptualization amongst others at work in human geography that together are trying to build a fuller account of the complex processes of life, which does not divide off and privilege the 'social' in narrow terms, and see the social just in reflexive, rational, language based terms. Doreen Massey is critical of Heidegger's notion of place as being too rooted but, she adds that in principle his reformulation of space as place seems to point in the right direction.

See also: Actor-Network Theory/Network Geographies; Affect; Becoming; Ecology; Emotional Geographies; Humanism/Humanistic Geography; Hybridity; Landscape; Non-Representational Theory/Non-Representational Geographies; Performativity; Phenomenology/Phenomenological Geography; Place; Posthumanism/Posthumanistic Geographies.

Further Reading

Bate, J. (2001). *The Song of the Earth*. London: Picador.

Buttimer, A. (1976). Grasping the dynamism of the life worlds. *Annals of the Association of American Geographers* 66, 277 292.

Casey, E. S. (1998). *The Fate of Place: A Philosophical History*. Berkeley: University of California Press.

Cloke, P. and Jones, O. (2001). Dwelling, place, and landscape: An orchard in Somerset. *Environment and Planning A* 33, 649 666.

Creswell, T. (2004). *Place: A Short Introduction*. Oxford: Blackwell.

Franklin, A. (2002). *Social Nature*. London: Sage.

Gaita, R. (2003). *The Philosopher's Dog*. London: Routledge.

Harrison, P. (2007). The space between us. Opening remarks on the concept of dwelling. *Environment and Planning D* 25, 625 647.

Harrison, S., Pile, S. and Thrift, N. (2004). *Patterned Ground: Entanglements of nature and culture*. London: Reaktion Books.

Harvey, D. (1996). *Justice, Nature, and the Geography of Difference*. Oxford: Blackwell.

Heidegger, M. (2001). *Poetry, Language, Thought*. New York: Perennial Classics.

Hofstadter, A. (2001). Introduction. In Heidegger, M. (ed.) *Poetry, language, thought*, pp x xxv. New York: Perennial Classics.

Ingold, T. (1993). The temporality of landscape. *World Archaeology* 25, 152 174.

Ingold, T. (1995). Building, dwelling, living: How people and animals make themselves at home in the world. In Strathern, M. (ed.) *Shifting contexts: Transformations in anthropological knowledge*, pp 57 80. London: Routledge.

Ingold, T. (2000). *The Perception of the Environment. Essays in Livelihood, Dwelling and Skill*. London: Routledge.

Ingold, T. (2005). Epilogue: Towards a politics of dwelling. *Conservation and Society* 3(2), 501 508.

Jones, O. (2003). The restraint of beasts: Rurality, animality, actor network theory and dwelling. In Cloke, P. (ed.) *Country visions*, pp 450 487. London: Pearson Education.

Macnaghten, P. and Urry, J. (1998). *Contested Natures*. London: Sage Publications.

Massey, D. (2005). *For Space*. London: Sage.

Massey, D. and Thrift, N. (2003). The passion of place. In Johnston, R. & Williams, M. (eds.) *A century of British geography*, pp 275 299. Oxford: Oxford University Press/British Academy.

Pickles, J. (1985). *Phenomenology, Science and Geography: Spatiality and the Human Sciences*. Cambridge: Cambridge University Press.

Relph, E. (1976). *Place and Placelessness*. London: Pion.

Seamon, D. (1980). Body subject, time space routines, and place ballets. In Buttimer, A. & Seamon, D. (eds.) *The human experience of space and place*, pp 148 165. London: Croom Helm.

Seamon, D. and Mugerauer, D. (1989). *Dwelling, Place and Environment: Towards a Phenomenology of Person and World*. Haven: Yale University Press.

Thrift, N. (1996). *Spatial Formations*. London: Sage.

Thrift, N. (1999). Steps to an ecology of place. In Massey, D., Sarre, P. & Allen, J. (eds.) *Human geography today*, pp 295 352. Oxford: Polity.

Thrift, N. (2005). From born to made: Technology, biology and space. *Transactions of the Institute of British Geographers* 30, 463 476.

Tuan, Y. F. (1971). Geography, phenomenology, and the study of human nature. *Canadian Geographer* 15, 181 192.

Tuan, Y. F. (1974). *Topophilia: A Study of Environmental Perception, Attitudes and Values*. Englewood Cliffs, NJ: Prentice Hall.

Villela Petit, M. (1996). Heidegger's conception of space. In Macann, C. (ed.) *Critical Heiddegger*, pp 134 158. London: Routledge.

Vycinas, V. (1969). *Earth and Gods: An Introduction the Philosophy of Martin Heidegger*. The Hague: Martinus Nijhoff.

Wylie, J. (2003). Landscape, performance and dwelling: A Glastonbury case study. In Cloke, P. (ed.) *Country visions*, pp 136 157. Harlow: Pearson.

Relevant Website

http://pegasus.cc.ucf.edu

Place and Space in Geography, Environment, Environmental Ethics, Ecology, Research on Place and Space; Pegasus, University of Central Florida.

E

East Asian Miracle

C. Dixon, London Metropolitan University, London, UK

Glossary

Conditionalities The conditions attached to loans, particularly from such agencies as the IMF and World Bank, but bilateral donors also make conditional loans. Policy conditionalities have been particularly associated with 'structural adjustment programs', financial rescue packages, and the neoliberal agenda of liberalization, privatization, and reduced state activity.

East Asia A group of countries to the east and southeast of the Eurasian landmass. Sometimes referred to as Pacific Asia or Asia-Pacific or separated into two regions, East Asia and Southeast Asia.

GATT (General Agreement on Tariffs and Trade) From 1947 until the establishment of the World Trade Organization (WTO), GATT provided the institutional focus for world trade.

The East Asian Miracle The World Bank is usually attributed with coining this term in the report of that name published in 1993. However the term 'miracle' was applied much earlier to individual Asian economies, for example, by Chalmers Johnson in *MIT and the Japanese miracle* and Hafiz Mirza's description of Singapore as 'a miracle by design' in *Multinational corporations and the growth of Singapore*.

Introduction

The label 'miracle economies' has been applied to a group of East Asian economies that since the early 1960s experienced high rates of economic growth, associated with low levels of income inequality and improved living standards. These economies have been frequently presented as comprising a series of waves or generations, each characterized by high rates of growth of GDP (**Table 1**) and exports of (initially) low tech/low skill manufactured goods. The lead was taken by Japan in the 1950s, followed by Hong Kong, Singapore, Taiwan, and South Korea (the Newly Industrializing Economies, NIEs) in the 1960s and 1970s, Indonesia, Malaysia, and Thailand (sometimes termed the new NIEs) in the 1980s, and China and Vietnam (the Asian Transitional Economies, ATEs) starting in the 1990s. These developments have been depicted by Japanese economists as a flight of geese, with growth spreading from Japan, driven, at least in part, by progressive relocation of lower tech activities in the face of cost pressure and the upgrading of technology and skills. Such processes were accompanied by Japanese and later NIE based companies going transnational on a regional basis. This form of organization was also followed by Western companies looking for low cost export platforms in East Asia. This disaggragating of production at the regional scale, taking advantage of cost, technological and skills levels had by the mid 1990s, produced a complex pattern of production networks and associated trade and investment flows, such that these countries were increasingly trading with and investing in one another. A very dynamic process of informal regional integration was operating behind which lay a series of nested regional divisions of labor (**Figure 1**).

By the early 1990s it was generally expected that the miracle would spread and deepen, with projections of extraordinary growth of wealth and markets, but it was also considered that the strength and increasing regionalization of growth would insulate East Asia from events in the rest of the world. However, some commentators were beginning to have doubts over the sustainability of the miracle. Thailand was showing signs of overloaded infrastructure, rising costs, and an inability to move to more skill and technology based production. The US based economist Paul Krugman questioned the extent to which the miracle was based on the additions of labor and capital rather than increased productivity. Such views were very much in the minority and refuted from an International Monetary Fund (IMF) perspective by, for example, Michael Sarel (an economist with the South East Asian and Pacific Department of the IMF). As late as March 1996 the IMF was endorsing the economic health and prospects for the sustained growth of such economies as Thailand and South Korea. However, in July 1997 a

Table 1 Annua average percentage rate of growth of GDP, 1960–2006

	1960–69	*1970–79*	*1980–89*	*1991*	*1992*	*1993*	*1994*	*1995*	*1996*	*1997*	*1998*	*1999*	*2000*	*2001*	*2002*	*2003*	*2004*	*2005*	*2006**
ndones a	3.5	7.7	4.9	7.0	6.3	6.6	7.0	7.2	7.8	4.6	–15.0	0.8	4.9	3.5	4.4	4.7	5.1	5.6	5.4
Ma ays a	6.5	7.6	4.7	8.6	7.8	8.5	8.6	9.2	8.6	7.8	–7.0	6.9	8.0	0.3	4.1	5.5	7.2	5.2	5.9
Ph pp nes	5.1	6.3	2.0	0.0	0.1	1.7	4.5	5.6	5.7	5.2	0.3	3.3	4.5	4.5	4.4	4.9	6.2	5.0	5.5
Tha and	8.2	7.2	6.8	7.9	7.2	7.9	8.8	8.6	6.7	–0.4	–7.0	4.4	4.8	2.1	5.3	7.1	6.3	4.5	4.6
S ngapore	8.8	8.8	6.8	6.7	6.0	9.8	6.8	8.3	6.9	7.8	1.3	6.8	9.7	–1.0	4.0	2.9	8.7	6.4	8.4
V etnam	–	–	4.6	6.0	8.6	8.1	8.8	9.5	9.3	8.6	4.0	4.1	6.8	6.9	7.1	7.3	7.8	8.4	7.7
Hong Kong	10.0	9.3	6.6	3.9	5.0	5.9	4.9	5.0	5.0	5.3	–5.5	3.4	10.2	0.5	1.8	3.2	8.6	7.5	6.8
S. Korea	8.5	9.5	6.8	8.4	4.8	4.5	6.3	9.2	7.1	5.5	–5.9	9.4	8.5	3.8	7.0	3.1	4.7	4.0	5.1
Ta wan	9.2	8.0	7.8	7.0	7.0	6.0	6.2	6.6	5.7	6.8	4.9	5.4	5.9	–2.2	4.2	3.4	6.1	4.0	4.3
PRC	–	–	9.2	7.2	12.0	10.2	11.0	9.9	9.6	8.8	7.8	7.1	8.0	7.5	9.1	10.0	10.2	10.2	10.5

*Provisional

Source World Bank

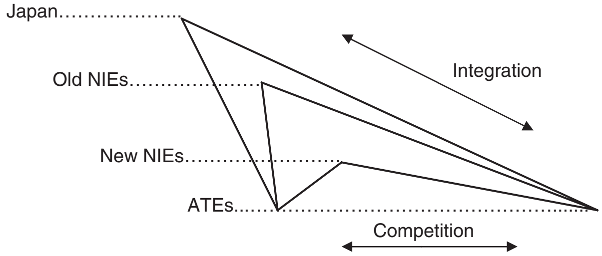

There is very considerable complementarity between the groups but significant competition within them for markets and investment.

Figure 1 Nested regional divisions of labor.

Pre-crisis
Growth economies
Miracle economies
Highly performing Asian economies
Open economies
Correct policies
Market friendly policies
Low levels of price distortion
Models for the rest of the Third World

Post-crisis
Government interference
Closed economies
Restrictions on trade, finance, and foreign ownership
Inadequate regulation
Corruption
Casino capitalism
Incorrect policies
Warnings to the rest of the Third World

Figure 2 Changing views of the East Asian economies.

major financial crisis emerged in Thailand, spreading rapidly through the region, reflecting the close trade, investment, and currency linkages that had developed. The most seriously affected economies, Indonesia, South Korea, and Thailand (**Table 1**), had to seek the assistance of the IMF, which provided rescue packages of US$23 billion, US$21 billion, and US$3 billion, respectively. The crisis produced a most extraordinary about face on the part of the majority of commentators and international agencies (**Figure 2**). Subsequently, major debates have ensued over the nature of the East Asian economies, their success, and the causes and impact of the 1997 crisis, adding to the long term disputes between what might be broadly called the neoliberal and developmental state positions outlined below.

The World Bank and the Selling of the East Asian Miracle

From *c.* 1980, what is sometimes termed the neoliberal counter revolution in development orthodoxy, focused attention on national policy and the manner in which economies dealt with the international system. Successful development was depicted as the result of open economies, free markets, and limited state intervention. Increasingly, the East Asian economies were presented as the product of such policies. These were at the root of long term growth of these economies and the very limited extent to which they were affected by the unstable international trading conditions of the 1970s, the Third World debt crisis, and the recession that characterized the early 1980s.

There are two key elements in the neoliberal packaging of East Asian development. First, prior to the rise of neoliberalism, there were Western geopolitical and ideological Cold War imperatives for successful economic reconstruction and growth in such economies as South Korea and Taiwan to be presented as the product of Western free market capitalism acting on previously disrupted and backward economies. Hence, the tendency to ignore both the importance of pre 1940s development and the role of US aid. Post 1945 both Taiwan and South Korea became enmeshed in a system of American hegemony which brought economic and military aid on an unheard of scale. Between 1945 and 1980 US nonmilitary aid amounted to US$5.6 billion for Taiwan and US$13.0 billion for South Korea, compared to US$6.9 billion for the whole of Africa and US$14.8 billion for Latin America. Indeed, it can be argued that in terms of economic and military assistance and access to Western markets, all the pro Western East Asian economies received highly preferential treatment. Second, concerns over retaliations through General Agreement on Tariffs and Trade (GATT) procedures made the East Asian economies wary of making subsidies, protection, and other distortion to the market too obvious – this was strikingly the case in South Korea. However, the highly protected nature of the East Asian economies and their subsidizing of key sectors and exports did become an issue for many Western trade associations who lobbied their governments for action. Resultant negotiations and compromise trade deals suggest an element of doublethink among the advocates of East Asia as the showpiece of neoliberal policies.

As the East Asia success cases moved central to the neoliberal blueprint for the Third World as whole, they become associated with a series of slogans: get prices right; get the basics right; and sound macroeconomic policies. That is, limited price distortions while ensuring low inflation, high levels of domestic saving and investment, low or nonexistent budget deficits, and low levels of international debt. It is certainly the case that by and large the East Asian economies did follow such policies, with the notable exception of price distortions. Indeed, in 1994 Alice Amsden suggested that the success of the East Asian economies reflected the extent to which (in neoliberal terms) they got prices wrong.

In 1993 The World Bank brought the various neoliberal views of the East Asian economies together in *The East Asian miracle*. This study focused on the seemingly miraculous growth of Japan, Hong Kong, South Korea, Singapore, Taiwan, Indonesia, Malaysia, and Thailand, terming them the 'Highly Performing Asian Economies' (HPAEs). This study had its origins in a project financed by the Japanese government, aimed at promoting the Japanese form of development as an alternative to the World Bank's neoliberal model. However, in the event it appears that the World Bank was not prepared to countenance the promotion of an alternative to its established position. In addition, the malaise that characterized the Japanese economy from *c.* 1990 seriously undermined the Japanese case. Thus, the final version of the study highlighted the ways in which the HPAEs had achieved their success through sound neoliberal policies, with limited state intervention and significant investment in education and training. Where the state had intervened this was described as 'tailored' to deal with specific market and sectoral weaknesses, with the state operating a system of rewards and penalties that replicated the market. However, the details and contribution of state intervention do not emerge clearly. The authors suggest that it is impossible to test the importance of state activity, dismissing this aspect of the HPAEs' development with the speculation that if there had been less state intervention then perhaps growth would have been even more miraculous. Other countries were warned that while state intervention was a feature in some of the successful East Asian economies, this did not mean that it should be attempted elsewhere, nor used as an excuse to resist market oriented reform. Not surprisingly there has been considerable debate over this aspect of the study, particularly from those that had long been critical of the neoliberal position and viewed the East Asian miracle as a product of high levels of state intervention (see below). The later position, referred to in the World Bank study as the 'revisionist position', is dismissed in less than two pages (out of 389).

The study has also been heavily criticized for ignoring historical, geopolitical, and cultural factors, glossing over continuing poverty, high levels of environmental damage, abuse of human rights, and the limited development of civil society, while lumping eight extremely diverse economies together under the HPAE label. For many critics the attempt to reduce East Asian development to a single model with wide application is the study's most basic flaw.

The Developmental State View

From the early 1980s a series of studies of East Asia emphasize the central role of the state. Among the most influential were Johnson's 1982 study of Ministry of International Trade and Industry (MITI) and the growth of Japan, Alice Amsden's 1989 account of South Korea, and Robert Wade's 1990 exposition of market governance and the developmental state. For Johnson, Japan was a developmental state which prioritized industrial policy and managed domestic industry so as to improve international competitiveness. Latter writers have broadened the scope of the developmental state concept to include the management of social transformation. For such writers as Gordon White and Richard Wade this involved pervasive political controls, ideological mobilization, and social engineering.

To be effective a developmental state needs a high level of capacity to formulate and implement long term goals. This in turn necessitates a degree of autonomy, that is, a state able to resist interest groups and prioritize long term goals over short term advantage. A critical element in this is an effective bureaucracy that can implement policies. However, there are very considerable debates over the origins of developmentalism, its operation, and the nature of the concept as a whole.

It is sometimes argued that developmentalism comes out of periods of major domestic upheaval and external threats, with authoritarian rule been legitimized through economic growth, improved living standards, and defending the state. At the general level, this could be seen as the background to the development of a number of the East Asian states, notably Japan, Taiwan, and South Korea. However, many states have faced such disruption and threats and become highly authoritarian without becoming developmental. Thus, the question of developmentalism in East Asia (and its absence elsewhere) cannot be understood outside of the particular (and perhaps unique) series of interactions between the historical, cultural, and geopolitical contexts of a very varied group of East Asian economies.

If a successful developmental state has to be highly autonomous, this implies a separation from society, which poses interesting questions over effective information gathering, policy formulation, and implementation. Robert Wade, however, sees the critical relationship being between business and the state, which comprises various authoritarian–corporatist forms. Harder authoritarianism in Taiwan and South Korea and a softer form in Japan are providing the basis for the effective operation of the centralized bureaucracy in order to gather information and implement policies.

Other writers, for example, Joel Migdal, have stressed the importance of a strong state which has the capacity to penetrate and regulate society, and to extract and appropriate resources, but is firmly embedded in society. Under this view, there is constant and intense interaction between the state and the society. This is an argument that Peter Evans takes this further. While stressing the embedded autonomy of such organizations as MITI, in terms of establishing policy, he gives equal weight to the networks that link the bureaucracy to society and the market. This interaction enhances the effectiveness of the bureaucracy. In this context lobbying by various interest groups can be a positive part of the process. There is here a question of balance between autonomy and capacity that has perhaps not been sufficiently researched. Linda Weiss has also seen these linkages as critical to a state's capacity and adaptability. However, it is apparent that there is considerable variation in the level of interaction, its impact, the level of capacity, both within and between states, how it is exerted, and the degree of authoritarianism.

A further variation on the developmental state appears to be emerging in China and Vietnam. These have been termed developmental party states and have developed in the wake of economic reform and re engagement with international capital. In both cases intensive business state linkages are developing, particularly at the local level, which are taking various corporate forms. In China the local state has become a critical element in growth, policy formulation, and implementation. Significant conflicts between the local and the central state are occurring – a situation that is also emerging in Vietnam – as the state attempts to accommodate new social forces and establish effective information gathering and regulatory structures.

The extensive literature and related debates over developmentalism concentrate on Japan and the NIEs, with some recent work on China and Vietnam. In comparison, little discussion has taken place over Indonesia, Malaysia, and Thailand. The World Bank labeled them HPAEs and until 1997 they were seen as the next NIEs, but to what extent could they be considered developmental states? Certainly some key elements of developmentalism are present and they do contrast markedly with the Philippines, where the failure to participate in the Asian miracle is generally seen as a result of state failure. However, in terms of historical experience, social and economic structures, and developmental problems they have little in common with Taiwan and South Korea.

In Thailand a critical role was played by the technocratic core of the Thai bureaucracy, particularly in the National Economic and Social Development Board (NESDB), the close state business linkages, and the informal linkages with the banking system. However, this has to be set against the lack of coordination of policy among the rest of the bureaucracy, lack of strategy and capacity in education, training, and infrastructure. There is here a partial developmental form, though one that in growth terms has shown considerable consistency and resilience. Malaysia could be considered a more effective and adaptable form and Indonesia rather less so.

While a significant level of state intervention appears, in different ways, to have been significant in all the cases of successful development in East Asia, generalization is difficult, even misleading. Indeed, in that sense any attempt to produce a unified developmental state model is at least as unsound as any general neoliberal model. While the state led view is much closer to the reality of East Asia, it is important to see success in the context of particularly (and probably unique) internal and external circumstances. While the East Asian miracle may constitute a major challenge to neoliberalism, it does not provide a replacement blueprint for the rest of the Third World.

The Impact of the 1997 Crisis

Most commentators have followed the neoliberal view that the causes of the 1997 crisis were internal. Blame was directed at national governments, elements of domestic capital, and their interrelationships. The IMF called for the dismantling of an economic system based on collusion between state, banks, and business and in the conditionalities attached to the rescue loans, stipulated rapid liberalization. Such positions imply that either the pre 1997 neoliberal view of the East Asian economies was wrong, or the economies had changed or had failed to adapt to changed global circumstances. In addition, such internalizing of the causes diverts attention from the operation of increasingly liberalized global financial markets and the lending policies of the international banks.

Overall, for the most seriously affected economies, there was a lack of the state capacity necessary to deal with the liberalized financial sector and the changed circumstances (see **Figure 3**). To a degree, this reflected political change, particularly in Thailand and South Korea. In the former, progressive democratization, combined with the weaknesses of the electoral system in which vote buying was rife, produced a series of coalition governments. These were weak, unduly influenced by business groups, and unable to address the problems of declining competitiveness, overloaded infrastructure and skill shortages, or the regulatory issues that had emerged following the liberalization of the financial sector. In addition, there was a politicization of key technocratic posts within the bureaucracy – notably in the Bank of Thailand and the Ministry of Finance – that further limited the effectiveness of regulation and policy implementation. When the crisis broke, there was a measure of policy paralysis.

While the crisis had some impact through East Asia, many economies emerged relatively unscathed (**Table 1**). Generally this reflected the capacity to take effective action to deal with currency values, current account

- A decline in export growth which reflected:
  - Loss of international competitiveness, resulting from currency adjustments, notably the appreciation of the US dollar to which most East Asian currencies were linked
  - Large inflows of FDI which tended to push up wages and other costs, further reducing competitiveness
  - Competition from China and Vietnam – in 1996 hourly labor costs in the textile sector were reported as US$1.49 in Thailand and US$0.49 in China
  - Reduced markets following the establishment of NAFTA and the EU single market
  - Glut in the international electronics sector – in 1996 50% of South Korea's export earnings were from electronics
- Related current account deficits which drained foreign reserves and put serious downward pressure on currency values
- Liberalization of financial regimes which gave the private sector easy access to large volume of comparatively cheap international funds
- Poor investment decisions by many domestic banks
- The emergence of high levels of private sector short-term debts which had fuelled stock markets and property asset bubbles

Figure 3 Factors behind the 1997 financial crisis.

deficits, and levels of corporate debt and counter recession with tax cuts and public expenditure. Most striking was the rapid, decisive, and effective response in Singapore, one of the world's most internationalized economies, though perhaps most attention focused on Malaysia's reimposition of capital controls. Such action was reinforced by lower levels of exposure (Malaysia), more limited integration with the international system (China and Vietnam), and less financial liberalization (Taiwan).

There has been considerable criticism of the IMF for not detecting the warning signs, particularly the dramatic decline in export growth and competitiveness during 1995–96, the stock and property asset bubbles, and the levels of short term debt and financial movements. To be fair, no one else noted the dangers inherent in these factors, least of all the international banks and other investors. However, there is considerable evidence that the IMF made the situation worse by misreading the nature of the crisis, applying uniform structural adjustment programs (SAP) like prescriptions aimed at restoring macroeconomic balance, and insisting on the closure of financial institutions. Indeed, subsequently the IMF admitted that mistakes had been made.

The IMF conditionalities stressed: the opening up of economies to foreign ownership and competition; privatization; corporate reform; reform and regulation of the financial sector. Particular attention focused on the banking sector and the high levels of debt–equity ratios – frequently 3:1 or more, when in the West an approach to 1:1 is considered a cause for concern. This practice combined with the close links between corporations and banks to facilitate the rapid raising of capital without recourse to the equity market. For the IMF and the Western financial regulators this form of funding is one of the root causes of the financial crisis. However, while there clearly was some unsound lending – as there was on the part of the international banks – it is also the case that the banking corporate linkages were central to the flexible and rapid responses to market opportunities that characterized the successful East Asian economies.

The basic aim of reform is to make the Asian corporate structures, governance, and regulation more like those of the West. From a neoliberal perspective, this was deemed necessary for sustained recovery and limiting the possibility of further crises. However, in so far as reforms are effective, they may undermine those features that were key to East Asia's long term dynamism. On the other hand, it may be that a combination of liberalization, globalization, and political change has made some of the key elements outmoded.

Rethinking the Miracle

The rapid recovery that most East Asian economies made from the 1997 crisis confounded most commentators. To considerable extent, the recovery must be seen as the result of the efforts of the states themselves, rather than just general improvements in global trading conditions. Given the dynamism and adaptability that the successful East Asian economies have exhibited over the last 40 years perhaps the recovery should not be so surprising. For all their faults these economies have achieved remarkable results and it might be more fruitful to ask why there had been so few major crisis. Similarly, rather that stressing the downsides of state business relations, corruption, cronyism, lack of transparency, and banking business linkages, it might be more useful to focus on the positive elements of such interaction.

The 1997 crisis revealed the very different geopolitical position that the East Asian economies have come to occupy since the end of the Cold War. In the initial stages of the crisis there was an expectation of US assistance and the sort of comparatively benevolent treatment from the IMF and World Bank that was experienced during the early 1980s. In the event, there was no US assistance, a blocking of attempts to put in place a regional financial institutional structure and rescue operation, and some harsh conditions were attached to the IMF loans. The East Asian economies ceased to have any role as showpieces for Western capitalism in the Third World and had become directly at odds with, in particular, American interests and ideology. However, the rapid recovery of East Asia (**Table 1**), the generally limited impact of reforms, and the rise of a new state led form in China and Vietnam, suggest that the region continues to provide a major challenge to neoliberalism.

See also: Developmentalism; Economic Crises; Economic Geography; Export Processing Zones; Foreign Direct Investment; Global Production Networks; Globalization and Transnational Corporations; Human Rights; Investment Promotion; Neoliberal Economic Strategies; Neoliberalism; Neoliberalism and Development; Pacific Rim; Spatial Division of Labor; Structural Adjustment; Trade Blocs; Transitional Economies; Transnational Corporations in Developing Countries.

Further Reading

Amsden, A. (1989). *Asia's next giant: South Korea and late industrialization*. New York: Oxford University Press.

Amsden, A. H. (1994). Why isn't the whole world experimenting with the East Asian model of development: Review of the East Asian miracle. *World Development* 22, 627 633.

Berger, M. T. and Beeson, M. (1998). Lineages of liberalism and miracles: The World Bank, the East Asian trajectory and the international development debate. *Third World Quarterly* 19, 487 504.

Brodsgaard, K. E. and Young, S. (eds.) (2000). *State capacity in East Asia*. Oxford: Oxford University Press.

Evans, P. (1995). *Embedded autonomy: States and industrial transformation*. Princeton, NJ: Princeton University Press.

Johnson, C. (1982). *MITI and the Japanese miracle: The growth of industrial policy 1925 1975*. Stanford, CA: Stanford University Press.

Krugman, P. (1994). The myth of Asia's miracle. *Foreign Affairs* 73, 62 78.

Migdal, J. (1988). *Strong states and weak states: State society relations and state capacity in the Third World*. Princeton, NJ: Princeton University Press.

Mirza, H. (1986). *Multinationals and the growth of the Singapore economy*. London: Croom Helm.

Page, J. M. (1994). East Asian miracle: An introduction. *World Development* 22, 615 625.

Sarel, M. (1996). *Growth in East Asia: What we can and what we cannot infer from It. Economic Issues. Number 1*. Washington: International Monetary Fund.

Stiglitz, J. and Yusuf, S. (eds.) (2001). *Rethinking the East Asian miracle*. Washington: World Bank.

Wade, R. (1990). *Governing the market: Economic theory and the role of government in East Asian industrialization*. Princeton, NJ: Princeton University Press.

Wade, R. and Veneroso, F. (1998). The Asian debt crisis: The high debt model versus the Wall Street Treasury IMF complex. *New Left Review* 228, 3 23.

Weiss, L. (1998). *The myth of the powerless state*. Ithaca, NY: Cornell University Press.

White, G. (1988). *The developmental state in East Asia*. London: Macmillan.

World Bank (1993). *The East Asian miracle: Economic growth and public policy*. New York: Oxford University Press.

World Bank (2000). *East Asia: Recovery and beyond*. Washington: World Bank Publications.

East/West

A. Bonnett, Newcastle University, Newcastle upon Tyne, UK

Glossary

Asian Values The notion, developed in political and cultural circles in East Asia from the 1970s onwards, that the indigenous cultures of (East) Asia encourage hard work, thrift, strong family ties, and social solidarity.
Bolshevism A variety of Marxist political activity associated with the ideas and practice of Lenin.
Occidentalism Stereotypes of the West; a set of interconnected representations that act to construct a vision or imaginary geography of the West.
Orientalism Stereotypes of the East; a set of interconnected representations that act to construct a vision or imaginary geography of the East.
Western Supremacism An ideology that claims the superiority and inevitable triumph of the West.

Historical Context

'East' and 'West' are contemporary geopolitical and geocultural designations. Yet they have come to us through many centuries and must be approached historically before their modern significance can be appreciated.

Our ability to locate ourselves in relation to the physical and human world demands 'orientation'. *Oriens* is Latin for east, derived from the verb, *oriri* 'to rise'. The word contains a clue to one of the universal material realities that shapes the geographical imagination, the arc of the Sun, its rising in the east and setting in the west. The Sun's daily journey provides us with our most basic sense of orientation.

Literally speaking, 'orientation' means to face the east. Medieval maps of Christendom were drawn with Jerusalem at the center and/or top. In this way maps were made to face east, or orientated. For the same reason, some early maps of the island of Britain had Wales at the bottom and East Anglia at the top (as seen on Richard Gough's map, from 1360). There is a particular Christian reason why nations should turn toward Jerusalem. But Christian orientation drew from a much wider and older tradition of veneration for the rising Sun. The positioning of buildings toward the east and south can be traced back to pagan structures, such as Stonehenge. Conversely, the fact that the west is the place where the Sun goes down has ensured its enduring association with death. The west has been seen as a site of life's ending and, by extension, of completion, maturity, and mystery. In the late fourth century AD, the Christian theologian Bishop Severian of Gabala (in Syria) explained why God had placed the Garden of Eden in the east, "in order to cause [man] to understand that, just as the light of heaven moves towards the West, so the human race hastens towards death."

These symbolic associations may also be traced from Egyptian deities of death and the west, through to notions of the west as a site of imperial culmination and rebirth in sixteenth century Europe. In fact, the connection between death and the west has numerous lines of ascent, many of which have led toward feelings of suspicion and dread of the west. The fact that white is the traditional color of death in some East Asian and African societies, when conjoined with the spectacle of 'white' conquerors coming from the west, the direction of death, produced a fearful connotative concoction.

Unsurprisingly, it is in China, a country with a vast amount of contiguous land to its west, that we find the oldest heritage of discovering and interpreting western horizons. Accounts of 'traditions regarding Western countries' became a regular part of dynastic histories from the fifth century AD. From early accounts of explorations as far west as Syria and beyond to 'the western terminus of the world', through to depictions of the modern West, one can identify a continuous tradition of Chinese commentary on the exotic lands of the setting Sun.

The notion of light and of life rising in the east and setting in the west appears to be a universal form of human orientation. However, even such basic ideas are subject to change as new geopolitical realities arise. The historical lineage of the modern geopolitical West is traditionally chartered from the breakup of the Roman empire, into western and eastern parts. The disintegrating western empire gradually came to be drawn together into what was later termed 'Western Christendom', a diverse community centered on Rome and Papal authority. Following schisms with the Eastern Church in the eleventh century and the mutual excommunication of the leaders of the two faiths in 1054, 'Western Christendom' achieved a clearer identity. The sacking by 'western' crusaders of Constantinople, the capital of the Eastern Church, in 1204, reinforced a sense of suspicion between the two Christian traditions.

The early history of 'Western Christendom' provides us with vague geographical parameters for the West. But it contains precious few indications of the emergence of the kind of wider, political and cultural, understanding of the West that we are familiar with today. If we want to find early associations that contain echoes of contemporary usage, we must look first to the rapidly changing world of sixteenth century Europe. For it was then that the unity of 'Western Christendom' was challenged by the Protestant Reformation. Protestantism created a new schism in the Christian world, a split that provoked the use of the secular term, 'West', as a noncontentious category for what was once called 'Christendom'. It was a word that papered over the cracks in Western Christianity. This secular usage was overlaid and complicated by the geography of the Reformation. For Protestantism was firmly associated with Northern and Western European countries and Catholicism with the South. Thus 'the West' began to be adopted within Protestant lands as a self designation, one that carried specific religious connotations.

However, a third connotation was also emerging. For a westward looking, colonial project was developing among those European powers facing the Atlantic. The 'Westernization of Europe' followed as the New World came to shape Western Europe's geopolitical horizons. Combined with its Protestant connotations, the notion of an imperial and colonial West might be imagined to have secured the status of 'the West' as the key geopolitical term both of the ages of discovery and Enlightenment. Yet it remained surprisingly peripheral: its usage largely prosaic and ideologically unelaborated. The West was, at best, a supplementary expression, used when talking about Europe. As this implies, it was 'Europe', not 'the West', that replaced 'Christendom'. However, the ground had been laid for the emergence in the late nineteenth century of 'the West' as the key geopolitical category of the modern world.

The Rise of the West

The West is an old word, but is also a modern idea. The assumption that being Western means being law governed and socially and technologically advanced is relatively recent. It was only in the late nineteenth and early twentieth century that the idea of the West acquired the role and range of meanings that are familiar to Westerners today. Moreover, the West's West has been changing ever since.

The rise to power of 'the West' is signified in a number of ways. One of the least obvious, because so taken for granted, is that the western rump of the Asian landmass has been awarded the status of being a continent, namely Europe. Another is that the West has come to signify the finished point of civilization.

Hegel gave an influential fillip to this notion when, in the early nineteenth century, he outlined his Occidentalist vision of Enlightenment. In a famous passage from *The Philosophy of History* (first delivered as lectures in 1822) Hegel explained that "[t]he History of the World travels from East to West, for Europe is absolutely the end of History." The notion that human freedom cannot exceed its modern, Western incarnation emerged from Hegel's association of Protestantism with rationalism. The Reformation was 'the all enlightening Sun', ensuring that "[t]he German Spirit is the Spirit of the new World." When Hegel talked about History traveling to the West, it was toward a Protestant horizon that he saw it going. However, despite elaborating at length on the Oriental world, Hegel had little to say about the West as a unity. When he turned to the Occident in *The Philosophy of History* he promptly fragmented it into three distinct civilizations (Greek, Roman, and German). Hegel had little use for any explicit or overarching sense of Western identity.

Similarly, in Britain, throughout the early to mid nineteenth century there appears to have been little serious interest in elaborating the idea of the West or imagining something called 'Western civilization'. Nevertheless, the association between the West and social dynamism and progress can be witnessed developing into a stereotype. The phrase 'the Western world' is used by Marx in 1853 when commenting upon British colonial policy in Asia. The same context, if different political conclusions, occasioned the use of the term by Macaulay 18 years earlier, in 1835. These were not merely descriptive usages. Clichés of Eastern social rigidity and conservatism and Western dynamism had already become firmly entrenched. It is a stereotype to the fore in Marx's depiction, in *Capital*, of Asiatic 'unchangeableness'. The radical British member of parliament, Joseph Cowen summed up these stereotypes pithily in a speech to the House of Commons (4 September 1880) when he depicted 'the conflicting civilizations of East and the West – the one iconoclastic and progressive, the other traditional and conservative'.

By the mid nineteenth century 'the West' is a word and an idea of growing but still unclear power and potential. It carries traditional religious and political connotations and these were, fitfully, being put to use to interpret Europe's rise to global power. But why, from the late nineteenth century, and particularly from the first two decades of the twentieth, did the West, became a 'central' idea, a ubiquitous category in the articulation of the modern world? The notion that Western society is a unity, with its own discrete history; that it is, moreover, a 'perspective', an ethno cultural repertoire, is a creation of little more than the last 100 years. The modern West is a relatively recent idea that absorbs and exceeds the term's older meanings. The development of this contemporary idea of the West can be explained in terms of the impact of specific events. The Bolshevik revolution, the rise of US hegemony, and the loss of colonial power are the most important of these; each acting to make the modern idea of the West seem more important, more necessary. A little of the complexity of this process by reference to the first of these events, the Bolshevik revolution, shall be indicated here.

Communism between East and West

Bolshevik communism in the USSR/Russia began as a project of Westernization. For Lenin and Trotsky Russia was ripe for Westernization; it was something that needed to happen not merely for the revolution to succeed but for it to be thinkable. In his typically imperious style Lenin decreed in 1918 that,

> it is our task … not to spare dictatorial methods in order to hasten the copying of Westernism by barbarous Russia even more than did Peter [the Great], not shrinking from barbarous methods of struggle against barbarism. (Lenin, 1965: 340)

The Bolshevik's association of barbarism and slavery with the East was, in part, a reflection of their faithful reading of Marx, who expressed similar views. However, it also represented the continuation of the views of the Russian elite. The conviction that progress was Western and conservatism Eastern was mapped onto an existing imaginative geography of the West being modern, enlightened, and civilized and the East being the antithesis of all three. 'Asianism' (Aziatchina) represented everything that was old and rotten, everything that needed to be ripped out of both Russia and her colonies. Mixing images of political reaction, with those of decay and infestation, Trotsky looked forward to the development of a clean, new Western civilization. "The revolution" he wrote in 1923, "means the final break of the people with Asianism, with the 17th century, with holy Russia, with ikons and cockroaches … an assimilation of the whole people of civilisation."

However, the 'moral support' for the Soviet state offered by some Western workers did not compensate for the brute fact that the revolution did not spread to the West. The Bolshevik project was premised on internationalism; more specifically, on revolution in the West. It was thought impossible for a communist state to survive in the midst of hostile capitalist powers. The absence of revolution, or even genuinely popular mass revolutionary movements, to materialize in the West, changed the nature of Soviet politics. It allowed Stalin to turn his back on the West and develop a form of nationalist socialism in which Russia was proclaimed to be the home of world revolution. "The whole world now admits," Stalin declared in 1930, "that the centre of the revolutionary movement has shifted from Western Europe to Russia." This attitude encouraged an almost phobic withdrawal from the West. In *The Captive Mind* (first published 1953), Czeslaw Milosz explained that,

> The official order is to evince the greatest horror of the West. Everything is evil there: trains are late, stores are empty, no one has money, people are poorly dressed, the highly praised technology is worthless. If you hear the name of a Western writer, painter, or composer, you must scoff sarcastically, for to fight against 'cosmopolitanism' is one of the duties of a citizen. (Czeslaw Milosz, 1985: 43)

With the collapse of communism in the late 1980s and early 1990s, Russia and its former colonial possession sought once again to redefine themselves in relation to the West. The outcome of this process is still not clear. Contemporary Russian governments like to deploy both pro Western and anti Western rhetoric. The latter is, in part, a response to the ability of the West, notably the USA, to proclaim ideological victory in the global battle for ideas.

Western Supremacism

Over the past two decades a muscular self confidence has taken hold of the Western media. This self confidence has led to a distinctly partial way of judging non Western representations of the West. Open contempt for 'their' atavistic refusal to acknowledge 'our' superiority is widespread. The collapse of communism is the most important cause of this new mood. However, the terrorist attacks on the USA of 11 September 2001 consolidated the trend and have come to shape a new popular idea that Western civilization is pitted and defined against Islamic extremism. Thus Victor Davis Hanson, who completed his book *Why the West has Won* in 2000, was embellishing an established narrative when, in 2002, he wrote his article, 'Occidentalism: The false West', for *The National Review.* In this piece Hanson argues that 'the East continues to stereotype the West, with not a clue about its intrinsic nature'. He mockingly portrays non Westerners as baffled by a 'mysterious Western paradigm – the freedom to speak freely'. Turning up the rhetorical heat to full blast, Hanson tells us that the most unaware and ignorant American knows more about 'what his own culture is and is not, about; what the Arab world stands, and does not stand, for' than the 'most sophisticated' intellectual of Middle Eastern origin.

In January of the same year another article titled 'Occidentalism' appeared making a similar point. For Margalit and Buruma, writing in *The New York Times Review of Books*, Occidentalism "is a cluster of images and ideas of the West in the minds of its haters." They add that what is really hated about the West, by those who only know it from afar, is its secularity and rationalism: "Occidentalists extol soul or spirit but despise intellectuals and intellectual life." In a recent book (*The Idea of the West*, 2004) the author has challenged these representations of Occidentalism. For the historical and international evidence suggests that the idea of West is used in multiple ways. Across the world stereotypes of 'the West' are employed to construct both pro Western and oppositional identities and pathways through mod ernity. Western supremacism makes it difficult to ap preciate this diversity.

Light from the East

Edwin Arnold's hugely popular epic poem *The Light of Asia*, a panegyric on the life of Buddha, was published in 1879. This work, along with Arnold's version of the Hindu epic the *Bhagavad Gita*, impressed and inspired Gandhi during his time spent studying law in London in 1888–91. On a less elevated plane, the assertion that Asia constituted a racial and political unity, 'a comity of Asia', was central to Meredith Townsend's *Asia and Europe*, which Nehru was later to cite as one of the influences upon his political imagination.

Over the past 50 years, the way Westerners have rep resented and stereotyped the East has been an object of fascination for Western critics. As we have just seen, the notion that the West played a part in the 'invention' or 'construction' of the East offers a partial truth. It is a partial truth that appears to attract Western intellectuals who cast themselves as critics of Western power. Hence, the idea of 'Asia' and the 'East' is persistently represented as something supplementary: as an idea dependent on and derivative of the all powerful (and all corrupting) ur continent of Eur ope. Yet this formula overlooks the ambiguous, transcultural origins of notions of Orient and Occident, Asia and Europe, both in premodern and modern Asia.

It is increasingly clear that an obsession with Western representations of 'others' is symptomatic of an Euro centric and increasingly dated insistence that the West is the defining center of world history. Two examples are offered here to develop this point. The first refers to Tagore's and Okakura's 'invention' of the East and West within their mission to imagine a nonmaterialist Asian modernity. The second example turns Tagore's and Okakura's hopes on their head for it introduces the contemporary attempt to construct the West as lazy and decadent compared to the hardworking, productive cul tures of Asia.

The Ideals of the East

It is difficult for us to sense the excitement and political provocation contained in the opening words of Kakuzo Okakura's *The Ideals of the East* (which first appeared in 1904), that 'Asia is one'. Yet, what a stupendous and im probable idea it is. The notion that there is some shared essence between all the different cultures of this vast space – or, at a more particular level, between, for ex ample, 'Indian culture' and 'Japanese culture' – is an imaginative leap of considerable proportions.

The Japanese art critic Kakuzo Okakura (1862–1913), along with his friend the Indian poet, Rabindranath Tagore (1861–1941), claimed to detect such an essence in spiritual traditions that stretch across a greater portion of the con tinent, most notably in Buddhism. Yet to even begin naming commonalties immediately opens the implausible nature of attempts to see Asia as a single entity. Buddhism, after all, is a minority religious current in Asia. Moreover, its ab sorption into regional cultures, from India to Japan, has taken very different paths. Thus any claim of contemporary solidarity based upon Buddhism becomes an act of histor ical retrieval, a reclamation, and reinvention of a distant past. More general affinities, which might seem plausible to many Westerners, such as the idea that Asia, as a whole, is more religiously inclined than the 'secular West', are also highly dubious. Formal religious observance in the USA remains today, as it was when Tagore was alive, a far bigger part of everyday life than it is in East Asia. Indeed, the deeply rooted nonreligious philosophical traditions of East Asia might make us wonder how Westerners ever managed to claim secularism as one of their defining characteristics.

Tagore wanted to break the association between mod ernization and Westernization. 'Modernism is not in the dress of the Europeans; or in the hideous structures, where their children are interned to take lessons' he argued: "These are not modern but merely European" (Tagore cited by Hay, 1970: p70). "True modernism," he continued, "is freedom of mind, not slavery of taste. It is independence of thought and action, not tutelage under European school masters. It is science, but not its wrong application in life."

However, although Tagore claimed to find a 'creative unity' between East and West, his dialectical logic was constantly interrupted by the stereotypes of both East and the West he had worked so hard to develop. What this means is that, because Tagore's West was a place of instrumentalism and soulless anomie, it was also a place quite unsuitable for 'creative unity'. It was a civilization that did not want real contact with others and that was, at root, destructive: "The dominant collective idea of the Western countries is not creative." Tagore asserted, "It is wholly wanting in spiritual power to balance and har monise; it lacks the sense of the great personality of man."

The Economic Ascendancy of the East

In the late twentieth century, Tagore's concern with soulless and mechanical Westernization was turned on its head. Eastern leaders began to accuse the West of in efficiency, decadence, and laziness. The East, in East Asia and, increasingly, in India, came to be associated with entrepreneurial endeavor and hard work. The West's 'undoing' is not merely a favorite topic among advocates of Eastern or 'Asian values' but a 'defining' theme: with out 'Western decadence' the notion that real or good Asians are devoted to wealth creation would be incomprehensible and easily rendered as a shallow and materialistic subversion of Asian identity. In 1998 the Singaporean ambassador to the United Nations, Kishore Mahbubani, offered a typical account of Western decline:

> Only hubris can explain why so many Western societies are trying to defy the economic laws of gravity. Budgetary discipline is disappearing. Expensive social programs and pork barrel projects multiply with little heed to costs. The West's low savings and investment rates lead to declining competitiveness vis à vis East Asia. The work ethic is eroding while politicians delude workers into believing that they can retain high wages despite becoming internationally uncompetitive. Lead ership is lacking. Any politician who states hard truths is immediately voted out. (1998: 97)

The twin themes of the work shy West and the self indulgent, 'out of control' West can be found throughout Asian values narratives. It is an interpretation of Western decay that carries a historical judgment. For it shows, as Singapore Prime Minster, Goh Chok Tong, put it during his National Day speech on 21 August 1994, that 'soci eties can go wrong quickly'.

> US and British societies have changed profoundly in the last 30 years. Up to the early 60s they were disciplined, conservative, with the family very much the pillar of their societies. Since then both the US and Britain have seen a sharp rise in broken families, teenage mothers, illegitimate children, juvenile delinquency, vandalism and violent crime. (cited by Sheridan, 1999: 72)

Traditionally, both Asian Westernizers and Eastern critics of Western soullessness saw themselves as under the thumb of the West, as reacting against its crushing force. In this context one can read a certain pleasure in between Goh and Mahbubani's lines: "they are recording that the tables have been turned, that it is now 'we' who can pat ronise and lecture 'them'."

Yet however sweet and, given the historical record, temperately expressed, this revenge must seem, a number of persistent ironies sour its enjoyment. The over zealous way that antisocial behavior is presented as Western offers a glib despatching of the sins of modernity. Goh's and Mahbubani's attitude toward the West represents a geographical 'othering' of internal problems. It is a pro cess of purification of the nation and of the East that sanctions and demands strict protection and self discip line as well as the perpetuation of an image of the decadent Westerner as a spatially displaced 'folk devil'.

See also: Asia; Europe; Modernity; Orientalism.

Further Reading

Arnold, E. (1879). *The Light of Asia*. New York: A.L. Burt.
Bonnett, A. (2004). *The Idea of the West*. Basingstoke: Palgrave.
Carrier, J. (ed.) (1995). *Occidentalism: Images of the West*. Oxford: Oxford University Press.
Hanson, V. (2001). *Why the West Has Won: Carnage and Culture from Salamis to Vietnam*. London: Faber and Faber.
Hanson, V. (2002). Occidentalism: The false west. *National Review* May 10.
Hay, S. (1970). *Asian Ideas of East and West: Tagore and His Critics in Japan, China, and India*. Cambridge: Harvard University Press.
Hegel, G. (1991). *The Philosophy of History*. Amherst: Prometheus Books.
Huntingdon, S. (1997). *The Clash of Civilisations and the Remaking of the World Order*. London: Simon and Schuster.
Lenin, V. (1965). *Collected Works, Volume 27, February July 1918*. Moscow: Progress Publications.
Mahbubani, K. (1998). *Can Asians Think? Understanding the Divide between East and West*. South Royalton, VT: Steerforth Press.
Margalit, A. and Buruma, I. (2002). Occidentalism. *The New York Times Review of Books* January 17.
Milosz, C. (1985). *The Captive Mind*. Harmondsworth: Penguin Books.
Okakura, K. (2000). *The Ideals of the East, with Special Reference to the Art of Japan*. New York: ICG Muse.
Said, E. (1978). *Orientalism: Western Representations of the Orient*. Harmondsworth: Penguin.
Sheridan, G. (1999). *Asian Values, Western Dreams: Understanding the New Asia*. St Leonards, NSW: Allen and Unwin.
Townsend, M. (1905). *Asia and Europe*. London: Archibald Constable.

e-Business and e-Commerce

K. E. Corey and M. I. Wilson, Michigan State University, East Lansing, MI, USA

Glossary

Business-to-Business (B2B) Activities utilize information and communications technologies to facilitate transactions involving intermediate goods and services.

Business-to-Consumer (B2C) Transactions use information and communication technologies (ICT)-enabled systems linking businesses and their customers for delivery and payment of final goods and services.

Consumer-to-Consumer (C2C) Interaction is ICT-enabled links between consumers for the purchase or exchange of goods and services.

e-Business Is the use of ICT to facilitate the operations of an organization through the purchase or exchange of goods and services, customer service, or provision of information.

e-Commerce Uses ICT for transactions online, and is a subset of e-business.

e-Government Is the use of e-business approaches to the delivery of public services.

Government-to-Consumer (G2C) Transactions are a form of e-business using ICT to link government and its citizens.

Introduction

Electronic business, or e business, is a wide ranging col lection of functions and activities that are facilitated by digital and electronic means. The ultimate goal of e business is to establish an online presence for an organi zation and to provide as much interactivity and func tionality as possible. e Business relies on information and communication technologies (ICT) to improve and ex pand the operation and functions of an organization, such as a business, government agency, or nonprofit organi zation. e Business is a recent phenomenon, lagging several years behind the initial popularity of the Internet in the early 1990s, and today with less than a decade of common presence in the most advanced economies. While often used as a synonym for e business, e commerce is a more specific concept associated with ICT enabled financial transactions. Other concepts associated with e business and e commerce include online interactions by business, government, and between individuals.

This article frames and discusses the concepts of e business and e commerce and many of the interdependent relationalities among these economic and societal functions. In addition to the characteristics of ICT enabled business, these activities also have different geographies or spatial and locational patterns. From the fundamentals and knowledge introduced in this article, the reader may be motivated to consider engaging in e business, thereby taking advantage of the new oppor tunities that are offered by digital and electronic technologies.

The past 50 years have witnessed the development and merger of computers and communications technol ogies. The power of these combined technologies created the Internet, which offers new possibilities for efficient and low cost electronic interaction for organizations and individuals. The existence of ICT and the Internet, however, does not guarantee the presence of e business functions. Effective and successful e business requires three elements. First, access to and use of the Internet, which includes the ability to go online for a reasonable cost by local standards, and the knowledge and skills needed to operate a computer and to navigate cyber space. Second, a developed financial services sector that allows electronic payments for goods and services, through banks, credit cards, invoicing, or third party banking equivalents. The third element is an efficient delivery system through postal or transportation services to complete the transaction. The absence of one or more of these elements makes the conduct of e business dif ficult if not impossible. Countries with well established e business sectors have sophisticated Internet, financial services, and delivery systems to support online business.

There are many reasons for adopting e business practices, with several of the leading factors listed next. Use of ICT often allows organizations to reduce costs through labor savings (automated systems, users enter data, offshoring), reduced transaction costs, more efficient delivery (electronic vs. ground/air/mail), and better tracking of operations. Organizations may provide im proved services, such as recent and greater information to customers by firms or citizens by government. New markets may be served, as the global reach of the Internet offers a far larger market than possible previously, al though the larger market may also carry greater com petition. Consumers may prefer online sources as they have 24 h access, reduced or no travel times, and the ability to rapidly survey the market for lowest prices or best conditions. The many advantages of an e business strategy also needs to be considered in terms of initial greater costs, not just in terms of establishing a web presence, but the possible need to re engineer organ ization or back office practices to function electronically.

Scope and Scale of e-Business

With over 1 billion people online globally, the potential now and in the future for e business activities is sub stantial. Estimates of the scale of e business activities are limited, as measurement lags behind this phenomenon, and the tendency for organizations to see data as a trade secret to be denied to competitors. In the 25 member European Union in 2005, the European Commission (2006) reported 91% of firms using the Internet, with 61% of firms having a web presence and almost a quarter of firms engaging in business to business (B2B) transac tions online. At the national level, many countries cal culate the level of online business. For example, in 2004, the United States Census Bureau (2006) cited over US$130 billion in online retail trade, with B2B e com merce exceeding US$1.9 trillion.

e Business is often characterized by the parties in volved in an online transaction, between and among business, government, and consumers. This section will briefly summarize common forms of e business. The most easily recognized form of online activity is business to consumer (B2C) transactions, which link businesses and their customers for information, purchase, delivery, and payment of final goods and services. Success with B2C occurs for firms that benefit from cost savings through electronic products and content that do not in volve physical movement (airline and hotel reservations, music and video downloads, etc.). Also successful are businesses that avoid local legal constraints (online gambling and adult content) and activities that can pro vide a wide selection at competitive prices (books, music, electronic goods, etc.). Despite the popularity of B2C commerce, it is only a small part of total online business. In the United States, B2C is only 7% of total e commerce.

The largest component of e business, and most powerful economically, is B2B. B2B utilizes ICT to fa cilitate transactions involving intermediate goods and services. B2B activities include online purchasing and distribution, supply chain management, and automated trading. The challenge for analysts of B2B e commerce is that B2B systems are an important competitive element for a firm, and system design and data are often con sidered proprietary and not publicized. A subset of B2B includes B2G when firms sell to governments.

e Government has several elements of e business, such as government to consumers or citizens (G2C), as well as government services to business (G2B). Govern ment is an important source of information, often at no charge, and the Internet is a powerful tool for the dissemination of public data and reports. Many organ izations rely upon government for business and demo graphic data, legal decisions, information about public policy, weather and climate, and access to forms. Gov ernment also transacts business online, such as for taxa tion, licenses and permits, and payment of fines. One developing function is electronic voting, which repre sents a complex intersection between technology, privacy, equity, and accuracy. As firms often found a need to re engineer for e business, government also found that the structure of government, with defined roles in specific departments, did not easily translate into cyberspace. The best government websites are those that address the needs of the user, designing the site to feature needs (taxation, travel, children, etc.) rather than force visitors to search across many departments to identify the in formation they need.

A final, and rapidly growing, dimension of e business is transactions among consumers (C2C). The most common formats are auction websites that allow indi viduals and businesses to buy and sell products and services online, such as eBay. These are third party websites that link consumers in a form of electronic flea market that also is emerging as a force for commercial sales and the auctioning of intermediate goods and ser vices. Somewhat related to the C2C format is peer to peer (P2P) online interaction, although this form tends not to use third parties to avoid legal implications associated with downloading of copyrighted material, such as music and video.

Functions and Content of e-Business

As a force that has had significant influence on the ways we live, work, learn, and entertain ourselves, e business is barely a human generation old, from electronic financial transactions 40 years ago to Internet enabled commerce only a decade old. However, during this relatively short life span e business has evolved from infancy to maturity, and e business functions are continuing to develop. These functions have evolved from consisting principally of simple commercial transactions on the Internet to include also informational exchanges that draw on myr iad technologies that transform the ways that we go about the business of living and of doing business. These transformations impact the economy, society, culture, polity, environment, geography, and time. The deep penetration of and growing reliance on e business and e commerce have resulted in the blurring of the lines between the physical worlds and the virtual online worlds. For the purposes of the conceptual framework being introduced here, it means that the contemporary concept of e business is a blend of material and elec tronic functions. Increasingly, it means that today, e business is a complementary mixture and partial integration of electronic technology supported business and physical business functions. Over time, the proportion of this mix is shifting increasingly to the electronic enabled end of this electronic to material spectrum.

Given the increasing complexity and diversity of e business functions, it is useful to frame them for clearer understanding and for use in business strategizing. Given the contemporary forces of information technology and globalization, these various e business functions should be conceptualized within the context of the global knowledge economy and network society. In this context three major e business activities may be identified and framed in a spectrum of business and economic functions and factors. These include: production functions; consumption functions; and quality of life and amenity factors. These three general sets of functions are linked and networked by means of such electronic technologies as the Internet and other telecommunications with multimedia content delivered via cable, telephone line, wireless, and satellite broadcasting.

Production Functions

Production may involve the generation of an innovation, such as an idea, the creation of a service, or the making of a good. These outputs in turn may or may not be developed further into a commercially marketable item. These outcomes may be electronic or material, but they were produced as a result of being electronically enabled. Examples include: basic research conducted in a university laboratory that produces an innovative scientific idea that has preliminary attractiveness for venture capital investment; a research and development department within a company tests and further demonstrates the commercial viability of life science process to be used in medical care, and it buys majority equity in its further development; production functions also include business and producer services whereby one firm provides special services, such as legal or accounting or advertising services to other firms; and public institutions such as governments provide producer services, as in the case of weather forecasts, taxes, and regulations.

Consumption Functions (or e-Commerce)

Consumption involves the use of a process, service, or good. Consumption may or may not require the exchange of money such as in the cash purchase of an automobile; this is an e commerce example. Further, e commerce may involve the exchange of information in order to gain access to services or to be more informed as in the example of searching many e government functions. As in the cases of the various production functions, in consumption functions there is not necessarily a sharp distinction between the electronic world and the material world. As e commerce deepens and matures, there often is better integration between online consumption and traditional physical retailing. For example, it is more common today to be able to make a purchase over the Internet and retrieve it or return it at a physical location such as the firm's store.

Quality of Life and Amenity Factors

In a global economy, where in many advanced national and regional economies, commercially value added knowledge and innovative services are creating more jobs than manufacturing industries. In such places, the needs of knowledge workers must be addressed, if such workers are to be grown locally and other such workers are to be attracted to the locality. Consequently, human capital development, a long recognized factor of production, may be more strategically framed as a factor of quality and amenity so that it is instrumental in nurturing and retaining a knowledge economy workforce for a particular region. This is an e business factor that can confer competitive advantage on a locality. In addition to human capital, quality of life and amenity factors such as a nearby attractive natural environment and creative and cultural assets are elements that enrich and reinforce the knowledge economy e business advantage of a place. Further, in a global economy labor market where high demand knowledge workers have virtually unlimited mobility, a local climate of tolerance, and a creative culture are critical factors that complement all the other e business functions noted throughout this article.

e-Business Spectrum

In sum, e business may be framed by a conceptual framework encompassing the principal economic, social, cultural, environmental, and spatial organizational factors that shaped development in a global knowledge economy. These functions are connected and linked by flows of information and knowledge that result in geographies of flows and content. The principal challenge empirically in engaging locally the study and/or the initiation of e business and e commerce is overcoming data and information availability – or lack of it. Frequently, marketing data may be available, but without the locational specificity and geographical breakdown needed for maximizing business decision making.

Spatial Organization of e-Business

Geographies of e-Business

The geography of e business has many elements and dimensions depending on the aspects addressed. At one level is the spatial pattern of ICT infrastructure, decided by the location of fiber optic and telecommunications networks that carry telecommunications and the Internet, and the nodes that direct and focus flows of electronic information. Patterns of infrastructure are evident at a number of scales, with leading and lagging cities, regions, countries, and continents. e Readiness indices based on ICT data show that the leading continents are North America, Europe, and Asia, and leading countries are Denmark, the United States, Sweden, Switzerland, and the United Kingdom. Least ready to engage in e business are Africa and Latin America. More concentrated than Internet infrastructure is the location of Internet hosts for e business, which tend to be found in the most advanced economies and leading cities. At its most advanced, the pattern of global e business consists of flows among the world's leading urban economies, with cities such as New York, London, and Tokyo playing core roles in this phenomenon.

Among the spatial factors that shape e business are language and culture, local and national laws, regional tastes and preferences, currencies, and privacy regu lations. Scaling up from the local to regional or national to international level means e business operations face a far more complex and nuanced market than previously. The three elements of e business noted at the start of this article – access, transaction, and delivery – all have spatial dimensions that need to be considered by pro ducers and consumers.

New Geographies in an Era of Digital and Intelligent Development

The reader should not be misled into thinking that business and commerce, enabled by ICT has eliminated distance or geography. Location and proximities still matter a great deal and they affect the ways that we behave and conduct business. These technologies, and our growing reliance on them and our expectations for their continuous improvement, have influenced and are changing the ways that we think about and perceive distance and the spatial organization of our new in creasingly networked worlds.

While there are patterns of convergence of electronic technologies and interconnectedness, and there is grow ing ubiquity of these technologies globally and locally, there remain many areas of the world, peoples, busi nesses, and institutions that are not networked or con nected or if connected, are poorly so. Therefore, even in this widely discussed networked society and global economy, there is a great deal of unevenness in the dis tribution of information and digital infrastructure and/or access to such technologies.

Cities and city regions function as principal locations where trade and e business converge. In the global knowledge economy city regions serve as hubs where networks and e business flows of transactions come together. These highest order urban places, in the con text of the global knowledge economy, host globally dominant financial services and complementary func tions such as world class cultural institutions in the arts, higher education, and corporate headquarters, among other business related and business supported assets.

The overarching business functional context of the global knowledge economy and network society is de velopment. In this context, economic development and economic growth alone are not sufficient for being competitive in an era of technology induced global ization. Development in the global knowledge economy, while driven by economic development and progress, recognizes the importance also of civil societal ad vancement, including material and environmental ad vancement and equity promotion for a majority of the populations of city regions. The economic development component of this comprehensive development frame work in the global knowledge economy is driven by the e business spectrum and its production, consumption, and quality of life and amenity factors. Because of spatial unevenness, there frequently needs to be a parallel public policy priority for equity to counter the inexorable mo mentum and movement toward development inequities and disparities.

In the current era of ICT facilitated globalization, two additional types of development require explication: digital development and intelligent development. Digital development involves ensuring that high speed broad band Internet infrastructure is distributed equitably and provided throughout a city, region, or country's settle ment space. Given the continuous improvement and pervasive competitive environment of the global econ omy, ensuring an initial distribution of first generation digital infrastructure may not be sufficient to ensure ongoing comparative advantage; consequently, constant upgrading and modernization of ICT infrastructure is required. In addition, the business and government interests of an area must recognize and operate with complementarity, usually in some form of alliances or partnerships, and from a position of intelligent develop ment. Intelligent development is digital development that is informed by effective market data and empirical re search, combined with the application of cutting edge and tested theory and conceptual frameworks that reflect best practices and benchmarking for monitoring business innovations and evaluation of the implementation of the selected business model.

Intelligent development in the context of effective e business practice involves attention to the geography of digital development and intelligent development pat terns. Areas that are underserved by digital infrastructure have less business potential and opportunity. These are places that represent gaps and disparities in the spatial organization of development; such areas can retard the competitiveness and intelligent development trajectory of a city region, thereby slowing the entire region's growth and advancement. Policies of equitable development and private–public partnerships are means that can be em ployed to address the need for widespread intelligent development.

What Does the Future Hold for e-Business?

The bursting of the dot com bubble in the year 2000 demonstrated that doing business and conducting com merce electronically did not exempt us from the need to adhere to the fundamentals of sound investment prin ciples and the best practices of planning and doing business. The relatively short life span of e business has been characterized by experimentation, learning, and change. No longer are these increasingly pervasive electronic facilitated economic and societal functions new to our individual and collective experience. These functions have evolved and can be expected to continue to be dynamic and continuously changing well into the foreseeable future.

As computers and Internet access move from being optional to being a necessity in most advanced econ omies, we confront the significance and implications of ubiquitous computing. Once access is assumed rather than an option, it means that lacking access, as an indi vidual, organization, community, region, or country, carries with it exclusion from the leading form of eco nomic interaction. Also, the ability to shift production to low cost areas suggests a world of new centers and per ipheries, where the centers are the leading urban nodes in the network society and the peripheries are those lo cations that cannot or will not engage with the new technologies. The equity implications and spatial con trasts such outcomes suggest may reshape the nature of the knowledge economy and network society.

As e business and e commerce evolve we can expect further melding of the electronic and material worlds by means of such activities as social networking, for ex ample, Second Life. This site reveals a virtual world that is three dimensional; it consists of a digital continent that includes a wide spectrum of economic and social activ ities where participants can create and construct mar ketplace and other e realities that involve trading and the use of currency. Increasingly, there are spillovers from and to the virtual and physical worlds. Second Life has attracted nearly 3 million inhabitants from around the world. Social networking has been attributed to stimu lating new interest in e commerce.

The concepts and the conceptual frameworks intro duced in this article should enable the reader to earn an initial understanding of the complexities and evolutionary changes that are inherent in e business and e commerce. Given the individual and personal scal ability of these two functions, this article may serve as a stimulus for readers to consider initiating and engaging in e business and/or e commerce.

See also: Business Services; Capital and Space; Capitalism; Civil Society; Competitiveness; Digital Divide; Entrepreneurship; Global Production Networks; Globalization and Transnational Corporations; Information Technology; Knowledge Economy; Knowledge Intensive Business Services; Local Economic Development; Uneven Development; Venture Capital.

Further Reading

Acs, Z. J. and Armington, C. (2006). *Entrepreneurship, geography, and American economic growth*. Cambridge: Cambridge University Press.
Albrechts, L. and Mandelbaum, S. J. (eds.) (2005). *The network society: A new context for planning?* New York: Routledge.
Beinhocker, E. D. (2006). *Origin of wealth: Evolution, complexity, and the radical remaking of economics*. Boston: Harvard Business School.
Clarke, S. E. and Gaile, G. L. (1998). *The work of cities*. Minneapolis: University of Minnesota.
Corey, K. E. and Wilson, M. I. (2006). *Urban and regional technology planning: Planning practice in the global knowledge economy*. New York: Routledge.
European Commission, Enterprise and Industry Directorate General (2006). The 2006 European E Business Readiness Index. http://ec.europa.eu/enterprise/ict/policy/ebi/ebi 2006 sh prel data rep.pdf (accessed Sep. 2007).
Fletcher, R. G., Moscove, B. J. and Corey, K. E. (2001). Electronic commerce: Planning for successful urban and regional development. In Williams, J. F. & Stimson, R. J. (eds.) *International urban planning settings: Lessons of success*, pp 431 467. Amsterdam: JAI.
Friedman, T. L. (2006). *The world is flat: A brief history of the twenty first century*. New York: Farrar, Straus and Giroux.
Gottmann, J. (1983). *The coming of the transactional city*. College Park, MD: University of Maryland Institute for Urban Studies.
Huggins, R., Izushi, H. and Davies, W. (2005). *World knowledge competitive index 2005*. Pontypridd: Robert Huggins Associates Ltd.
Kellerman, A. (2002). *The internet on earth: A geography of information*. Chichester: Wiley.
Leinbach, T. R. and Brunn, S. D. (2001). *Worlds of e commerce: Economic, geographical and social dimensions*. Chichester: Wiley.
Murdoch, J. (2006). *Post structuralist geography: A guide to relational space*. London: Sage Publications.
Rogers, E. M. (1976). Communication and development: The passing of the dominant paradigm. *Communication Research* 3, 213 240.
United States, Census Bureau (2006). E Stats. http://www.census.gov/eos/www/papers/2004/2004reportfinal.pdf (accessed Sep. 2007).
Yeung, H. W. C. (2005). Rethinking relational economic geography. *Transactions of the Institute of British Geographers* 30, 37 51.

Relevant Websites

http://www.americasnetwork.com
America's Network.
http://www.topics.developmentgateway.org
dgCommunities, Knowledge Sharing and Collaboration Worldwide.
http://www.forrester.com
Forrester Research.

http://www.lboro.ac.uk
GaWC Globalization and World Cities, Loughborough University.
http://www.globaltechforum.eiu.com
Global Technology Forum.
http://www.InsidePolitics.org
Inside Politics, Darrell West of Brown University.
http://www.internetworldstats.com
Internet Usage World Stats Internet and Population Statistics.
http://www.jupiterresearch.com
JupiterResearch Market Research, Analysis, and Advice.
http://www.itu.int
Market Information and Statistics (STAT), International Telecommunication Union.
http://www.oecd.org
Organisation for Economic Co operation and Development (OCED).
http://www.pewinternet.org
Pew Internet and American Life Project.
http://www.secondlife.com
Second Life: Your World. Your Imagination.
http://www.census.gov
The 2007 Statistical Abstract, US Census Bureau.
http://forum.maplecroft.com
World Economic Forum Global Health Initiative, Maplecroft.

Ecological Fallacy

D. Sui, Texas A&M University, College Station, TX, USA

Glossary

Ecological Fallacy The fallacy of making inferences about individuals using aggregate data.

Ecological Inference The process of using aggregate data to infer relationships at disaggregate/individual levels.

Modifiable Areal Unit Problem (MAUP) Analytical results often change when the scale or the areal boundary used in the analysis changes.

Multilevel/Cross-Level Analysis Analyses that conceptualize and analyze associations at different levels of aggregation, ranging from individual to area-based data. These analyses typically entail the use of variance component models to partition the variance, and to examine the contribution of factors measured at these different levels to the overall variation in the outcome.

The so called ecological fallacy refers to the inferential mistake that relationships observed for groups also necessarily hold for individuals. It is a logical fallacy in making inferences from group, aggregated data about individual behavior. Ecological fallacy is closely related to ecological inference (EI), which refers to the process of using aggregate (i.e., 'ecological') data to infer discrete individual level relationships of interest when individual level data are not available. Ecological fallacy is a manifestation of the so called modifiable areal unit problem (MAUP), which refers to the fact that statistical results are often contingent upon how data are aggregated using different areal units or scales. Analysis based upon a single set of areal unit or scale does not always warrant a reliable conclusion.

For example, a geographer may be interested in the level of development for countries in the world. One indicator one may consider using is per capita income at the country level (it could also be regional or municipal or local level). The per capita income for country A is \$45 000. To state that the average income for residents in country A is \$45 000 is true and accurate. The ecological fallacy occurs when one infers, based on the per capita data, that citizens in country A earn about \$45 000. In reality, earnings by each individual in the country can vary a great deal, from a few thousands to a few millions. As a general rule, one should keep in mind that assumptions made about individuals based on aggregate data are always vulnerable to ecological fallacy. In other words, the relationship between two variables at an aggregated level can be purely due to aggregation effects rather than to any real association. Another reason for the ecological fallacy is that there often exists a series of confounding factors that may cause the relationship between two variables to change drastically when moving from individual to aggregated analysis.

Geographers have studied this issue quite extensively, especially under the banner of the MAUP, but the problems related to EI and ecological fallacy can be traced back to Durkheim's study on differential suicide rates in Europe. Durkheim noticed that suicide rates were higher in countries where there were a large percentage of Protestants, and he speculated that suicide was promoted by the social conditions of Protestantism. This is often regarded as one of the earliest examples of EI as Durkheim used aggregated data but tried to make inferences about the relationship between suicide and religious faith.

The issue of EI was more formally discussed by Ogburn and Goltra in 1919, who sought to know how women, who were about to have the vote nationwide, would decide to cast their ballots. The secret ballot and the EI problem prevented analysts from distinguishing the votes of women from the remaining (male) votes in the same electoral precincts. In 1950, Robinson further raised the scholarly community's awareness on the issues related to ecological fallacy by demonstrating that relationships existing at the individual level can in fact be the inverse of aggregate level relationships – an observation that has been consistently confirmed across a variety of disciplines.

Traditionally, the ecological regression technique (techniques for estimating regression parameters using aggregated data when disaggregated data is not available) has been the most frequently used method for making EIs. Another deterministic method of EI is the method of bounds, which uses the limited information that is known with certainty from the cross tabulated aggregated data. However, the method of bounds only produces a range of possible values for the individual level relationships; thus, it does not really provide a solution to the EI problem. Following these earlier pioneering works, statisticians have developed many other approaches to handle EI. Among the methods developed so far are: the maximum entropy method, the aggregated compound multinomial (ACM) model, the ecological logit model, the Lupia–McCue partitioning approach, the neighborhood model, double equation regression, and quadratic context effect models.

Widely claimed as one of the major methodological advances in social science during past 25 years, Gary King's book *A Solution to the Ecological Problem: Re constructing Individual Behavior from Aggregate Data* has re ceived intense interdisciplinary scrutiny and is perhaps one of the most important progresses in addressing the problems related to ecological fallacy. By combining the method of bounds with the least squares regression, Gary King developed a new solution to the EI problem. In his book, he not only proved that various versions of the aggregation problems are mathematically equivalent, but also developed several software packages that implement the statistical procedure (via random coefficient model ing), diagnostics, and graphics so that researchers can apply his method in practice. King also empirically tested his statistical method with datasets of groups for which individual behaviors were known. He made more than 16 000 comparisons from five datasets between his esti mates and individual's behavior.

Despite the current efforts in data mining to discover patterns using data collected at the individual/disaggre gate level (the opposite problem of EI), EI is still required in social science research when individual level surveys are unavailable (as in local electoral politics because of the secret ballot), unreliable (as in racial politics, respondents do not always report opinions), inaccurate/insufficient (as in a vast political geography), unfeasible (such as questions relating to historical times), inaccessible (due to privacy concerns), or simply too expensive. In addition, due to legislations regarding individual privacy in many coun tries, aggregation across spatial units is the only way to ensure confidentiality for much of the socioeconomic and public health data collected by the government.

The process of aggregating data often conceals the variations within the dataset that are visible at the dis aggregate level. However, being vigilant about ecological fallacy does not entail that identifying associations be tween aggregate data is necessarily erroneous. It also does not necessarily mean that any inferences drawn about associations between the characteristics of an aggregate population and the characteristics of subunits within the population are absolutely wrong either. Otherwise, we could commit what some scholars call the ecological fallacy, when in fact individual level models are not ne cessarily better specified than those at the aggregate level. Schwartz further elaborated on the validity approach, which examines all studies in terms of internal and external (construct) validity problems for cross level inferences.

Because aggregate data are more easier to obtain than data on individuals and many decisions in the real world are based upon administrative units (based upon which various social, economic, and public health data are often collected), ecological analysis is unavoidable and more often than not is a necessity. In addition, ecological analysis can offer help clues about individual behavior. Under many circumstances, EI is often the only alter native in the absence of individual data. In the context of public health and epidemiology, population level studies often play a very important role in defining the most important public health problems to be tackled, and in generating hypotheses as to their potential causes. It also has been increasingly recognized that some risk factors for disease genuinely operate at the population level. Thus, only through ecological studies can these types of risk factors be detected.

Due to the low likelihood of statistically resolving confounding effects and aggregation bias, multilevel an alysis techniques (also known as hierarchical linear modeling) have been widely used and tested for nested datasets. Based on regression analysis, multilevel methods enable researchers to model the associations between a dependent variable and independent variable measures at different hierarchical levels, such as the individual, household, and neighborhood levels. Multilevel analysis allows researchers not only to examine how variability changes among levels, but also to model directly the effects at specific levels – that is, effect of state level Medicaid policy on access to health, or effect of school district expenditures. Multilevel analysis, when used in conjunction with these EI techniques, can help us better understand the complex relationships among many dif ferent variables operating at different levels.

At a broader theoretical level, ecological fallacy and EI are closely related to the debate regarding structure and agency in social sciences or the contextual and compositional factors in epidemiology. The structure–agency debate has been revolving around the complex and subtle relationship between knowledgeable and capable human agents, and the wider social systems and environmental structures in which they are implicated. The theoretical insights gained from the structure–agency framework have often been translated into com positional (involving human agents at the micro level) versus contextual (involving social/environmental structures at the macro level) factors in empirical studies. Deeper insights can be gained if empirical studies using EIs can be closely linked to these theoretical contexts. Only then can we possibly see the spider in the complex web of causation among the phenomena we are studying using a combination of individual and aggregate data.

See also: Modifiable Areal Unit Problem; Regression, Linear and Nonlinear.

Further Reading

Achen, C. H. and Shively, W. P. (1995). *Cross Level Inference*. Chicago, IL: University of Chicago Press.

Amrhein, C. G. (1996). Searching for the elusive aggregation effect: Evidence from statistical simulations. *Environment and Planning A* 27, 105 119.

Anselin, L. (2000). The alchemy of statistics, or creating data where no data exist. *Annals of the Association of American Geographers* 90(3), 586 592.

Bickel, R. (2007). *Multilevel Analysis for Applied Research: It's Just Regression! (Methodology in the Social Sciences)*. New York: Guilford.

Blakeley, T. A. and Woodward, A. J. (2000). Ecological effects in multi level studies. *Journal of Epidemiology and Community Health* 54, 367 374.

Brown, P. J. and Payne, C. D. (1986). Aggregate data, ecological regression, and voting transitions. *Journal of the American Statistical Association* 81, 452 460.

Cho, W. (1998). If the assumption fits: A comment on the king ecological inference solution. *Political Analysis* 7, 143 163.

Cleave, N., Brown, P. J. and Payne, C. D. (1995). Evaluation of methods for ecological inference. *Journal of the Royal Statistical Society, Series A, General* 158, 55 72.

Diez Roux, A. V. (1998). Bringing context back into epidemiology: Variables and fallacies in multilevel analysis. *American Journal of Public Health* 88, 216 222.

Duncan, C., Jones, K. and Moon, G. (1998). Context, composition and heterogeneity: Using multilevel models in health research. *Social Science and Medicine* 46, 97 117.

Duncan, O. D. and Davis, B. (1953). An alternative to ecological correlation. *American Sociological Review* 18, 665 666.

Durkheim, E. (1897). *Le Suicide*. (English transl. Spalding, J. A., 1951. Toronto: Free Press/Collier MacMillan). Paris: F. Alcan.

Freedman, D. A. (2001). Ecological inference and the ecological fallacy. *International Encyclopedia for the Social and Behavioral Sciences* 6, 4027 4030.

Freedman, D., Klein, S., Ostland, M. and Roberts, M. (1998). A solution to the ecological inference problem. *Book Review. Journal of the American Statistical Association* 93, 1518 1522.

Freedman, D. A., Klein, S. P., Sacks, J., Smyth, C. A. and Everett, C. G. (1991). Ecological regression and voting rights. *Evaluation Review* 15, 673 711.

Gelman, A. and Hill, J. (2002). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

Goodman, L. (1953). Ecological regression and the behavior of individuals. *American Sociological Review* 18, 663 664.

Gould, P. R. (1970). Is 'statistix inferens' still the geographical name for a wild goose? *Economic Geography* 46(supplement), 439 448.

Greenland, S. and Robins, J. (1994). Invited commentary: Ecologic studies biases, misconceptions, and counter examples. *American Journal of Epidemiology* 139, 747 760.

Grofman, B. (1995). New methods for valid ecological inference. In Eagles, M. (ed.) *Spatial and Contextual Models in Political Research*, pp 127 149. London: Taylor and Francis.

Johnston, R. J., Hay, A. M. and Taylor, P. J. (1982). Estimating the sources of spatial change in election results. *Environment and Planning A* 14, 951 961.

Johnston, R. J. and Pattie, C. (2001). On geographers and ecological inference. *Annals of the Association of American Geographers* 91(2), 281.

King, G. (1997). *A Solution to the Ecological Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton, NJ: Princeton University Press.

Krieger, N. (1994). Epidemiology and the web of causation: Has anyone seen the spider? *Social Science and Medicine* 39(7), 887 903.

Lohmoller, J. B., Falter, J., Link, A. and de Rijke, J. (1985). Unemployment and the rise of national socialism: Contradicting results from different regional aggregations. In Nijkamp, P. (ed.) *Measuring the Unmeasurable*, pp 357 370. Martinus Nijhoof: Den Haag.

Lupia, A. and McCue, K. (1990). Why the 1980s measures of racially polarized voting are inadequate for the 1990s. *Law and Policy* 12, 355.

Morgenstern, H. (1995). Ecologic studies in epidemiology: Concepts, principles, and methods. *Annual Review of Public Health* 16, 61 81.

Ogburn, W. F. and Goltra, I. (1919). How women vote: A study of an election in Portland, Oregon. *Political Science Quarterly* 34, 413 433.

Openshaw, S. (1984). Ecological fallacies and the analysis of areal census data. *Environment and Planning A* 16, 17 31.

Owen, G. and Grofman, B. (1997). Estimating the likelihood of fallacious ecological inference: Linear ecological regression in the presence of context effects. *Political Geography* 16, 675 690.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review* 15, 351 357.

Schwartz, S. (1994). The fallacy of the ecological fallacy: The potential misuse of a concept and the consequences. *American Journal of Public Health* 84(5), 819 824.

Sui, D. Z. (1999). GIS, environmental equity, and the modifiable areal unit problem (MAUP). In Craglia, M. & Onsrud, H. (eds.) *Geographic Information Research: Trans Atlantic Perspectives*, pp 41 54. London: Taylor & Francis.

Sui, D. Z. (2000). New directions in ecological inference. *Annals of the Association of American Geographers* 90(3), 579 606.

Thomsen, S. R. (1987). *Danish Elections 1920 1979: A Logit Approach to Ecological Analysis and Inference*. Arhus: Politica.

Williams, G. H. (2003). The determinants of health: Structure, context and agency. *Sociology of Health and Illness* 25(1), 131 154.

Withers, S. D. (2001). Quantitative methods: Advances in ecological inference. *Progress in Human Geography* 25(1), 87 96.

Wrigley, N., Holt, D., Steel, D. and Tranmer, M. (1996). Analyzing, modelling, and resolving the ecological fallacy. In Longley, P. & Batty, M. (eds.) *Spatial Analysis: Modelling in a GIS Environment*, pp 25 40. Cambridge: GeoInformation International.

Relevant Websites

http://www.Gking.harvard.edu
Garry King.

Ecology

P. McManus, The University of Sydney, Sydney, NSW, Australia

Glossary

Abiota The nonliving parts of an ecosystem, including soil, water, and air.

Adaptation Specific changes that enable a species to survive and/or thrive in a specific setting.

Balance of Nature The concept that all life is interrelated and inherently in equilibrium. This idea is rejected by scientific ecologists who recognize that Nature is dynamic, but not in balance. The idea forms the basis for a number of radical environmental philosophies that critique human society and its impacts because of its destruction of the balance of Nature.

Biota The plant and animal life of an ecosystem, popularly known as flora and fauna. Biotic refers to the living parts of an ecosystem.

Ecosystem Any unit that includes all of the organisms in a given area interacting with the physical environment so that a flow of energy leads to an exchange of materials between living and nonliving parts of the system.

Environment The surroundings of an organism, both living and nonliving, that potentially affect the organism.

Organisms All living things, including people, plants, animals, bacteria, and fungi.

Radical Ecology The term used to cover a range of environmental philosophies that reject incremental changes in political systems and cultural values, but favor more fundamental changes to save or improve the environment. This is sometimes known as radical environmentalism.

Restoration Ecology The establishment on disturbed and/or degraded sites of former plant and animal communities which existed there. This may involve establishing appropriate conditions, thereby creating a trajectory for Nature to continue the restoration.

Succession A sequential change in vegetation, often prompted by changes in other abiotic and sometimes biotic conditions. A concept applied to sequential changes in human populations in an area.

Introduction

Ecology is a branch of science devoted to the study of living organisms and the relationships between organisms and their environment. Other definitions of ecology have concentrated on the patterns in Nature, how they originated, how and why they change in space and time, and the vulnerability of those patterns.

The term ecology was coined in 1866 as *oekologie* by Ernst Haeckel (1834–1919), a Prussian biologist and zoologist. It is derived from the Greek terms *oikos*, meaning household, and *logos*, which translates into study. The Greek word *oikonomia* is the root of both economy and ecology. The former term means management of the household, and the term ecology means the science of the household.

Ecology is a plastic term. It is used in a scientific, or professional, sense but it is also used in other ways. In the scientific sense, ecology is a discipline that is practiced by ecologists. Applied ecology is concerned with the management and conservation of habitats, including potentially degrading factors such as pollution. Ecology sometimes constitutes a subdiscipline of biology, although, even in scientific ecology literature, there is an argument that ecology is the culmination of all the sciences. This is because it integrates the various sciences that are concerned with the physical and chemical conditions of what are constructed as components of ecology. In other uses, the term 'ecology' is intentionally meant to be more holistic. Within scientific ecology there are subareas, including marine ecology and population ecology. In other fields there are important debates about the role of science versus overtly value-driven approaches, in ecology. This debate exists within restoration ecology, and is captured in the tension as to whether restoration ecology is a subarea of ecology, or whether it should be termed ecological restoration, which is more inclusive and to which scientific ecological knowledge is only one, albeit an important, contributor.

Ecology has been called the broadest of the life science disciplines, and therefore, the most indistinct. This is the case both within scientific ecology, and the use of the term in other contexts to mean an environmental philosophy or as a pseudonym for environment. It has been claimed that no other discipline seeks to achieve what ecology does, that is, to explain a variety of phenomena at a number of levels. This broad coverage within scientific ecology has probably contributed, perhaps indirectly, to its use in general conversation, further broadening its use but making it even more difficult to define. Ecology may be used in general conversation to mean how any organism relates to its environment. In everyday language, it is sometimes a synonym for the environment. In some approaches, 'ecology' is used in a nonscientific sense to highlight the importance of connections, in contrast to environment, which are understood as surrounds. It may also be used in place of the term environmental philosophy to describe different political or philosophical movements concerned with the environment. Sometimes ecology is constructed as a more radical concept, as in approaches such as deep ecology, political ecology, or radical ecology. In other approaches it is used as a way of linking Nature with other systems, for example, in the concept of industrial ecology, human ecology, and social ecology. The contemporary politicization of ecology can be traced to the late 1960s, when ecology was reinterpreted by some authors as no longer being a scientific discipline but rather an attitude of mind or the philosophy of a social movement. It is, however, part of a longer tradition of ecology as a natural science engaging with, and being engaged by, social and political thought. In this article these approaches are identified collectively as normative ecology and their similarities and differences are highlighted.

The use of the term 'ecology' in ways that may be considered nonscientific may arise for a number of reasons, but could indicate a particular critique about scientific ecology. This critique centers on the supposed apolitical nature of scientific knowledge and/or the limitations inherent in the application of scientific processes to Nature. 'Ecology' is not being employed loosely, as some scientists may argue, but critically to identify the shortcomings of the scientific world view.

In summary, ecology may simultaneously become the attitude of mind (i.e., reverence for the Earth, ecological processes, and communities), the normative project (i.e., protection for the Earth, ecological processes, and communities), and the scientific justification that is considered necessary for the normative project.

History

The Nineteenth-Century Foundations of Ecology

Prior to Haeckel coining the term 'ecology', a number of prominent scientists contributed to the development of knowledge that would form the foundations of the scientific discipline. In the early nineteenth century, Alexander von Humboldt identified relationships between organisms and their environment, including relationships between plant species and climate, in what he considered would be a future plant geography. In 1807, he recommended that in the great chain of causes and effects nothing and no activity should be regarded in isolation. Unlike other scientific disciplines that attempted to isolate particular phenomenon from their context, ecologists attempted to understand phenomenon through their relationships. This connectivity was captured by Arthur Tansley, who in 1934 coined the term 'ecosystem' to link ecology and system. Well before this step was taken, there were similarities between ecologists and nineteenth century naturalists, but the development of ecology as a science, and the rapid growth of scientific ecology knowledge, meant that a comprehensive general knowledge moved beyond the scope of an educated person. Ecology developed as an organized, scientific activity that classified knowledge about Nature in particular ways.

There was substantial debate among scientists in the nineteenth century about Nature, God, evolution, how to study these phenomena and who should be studying these phenomena. The work of Charles Darwin is crucial in the mid nineteenth century, because he developed the notion that the interaction of various organisms with their environment was essentially the struggle for existence. This was the process of natural selection, a term which became modified in social and economic languages to become the survival of the fittest. While Darwin did not use the term 'ecology', his research on the relationships between living things and their environment is generally recognized as being at the core of ecology. His work influenced Haeckel, whose definition of ecology initially appeared in a book that was developing Darwin's notion of competition in Nature. While German in origin, Haeckel's works were widely translated into English and were influential in the English speaking world.

The roots of ecology may be found in physiology. According to Haeckel, ecology was the practice of taking physiology from the laboratory and into the field. The focus of this research was largely on the abiotic environment. Ecologists, however, tended to emphasize the biotic environment. This meant a split between ecology and physiology, and a split between laboratory based research (which included zoos) and field research where species were studied in their natural surroundings. This raised important methodological questions, particularly the importance of a species' environment in influencing its behavior. Was it possible to study animal behavior and derive meaningful conclusions if the animal was isolated in a zoo or laboratory and deprived of the stimulation and community in its natural surroundings?

The emphasis on the biotic environment within ecology has been challenged in recent years, largely due to significant changes in the abiotic environment. Models, such as the interactions between biotic and abiotic factors shown in **Figure 1**, are important in developing an understanding of how ecosystems work, and how they are likely to respond to changes in parts of the system. Importantly, many of these changes have been caused, if not entirely then certainly in large part, by humans. Ecology cannot continue to exist as a

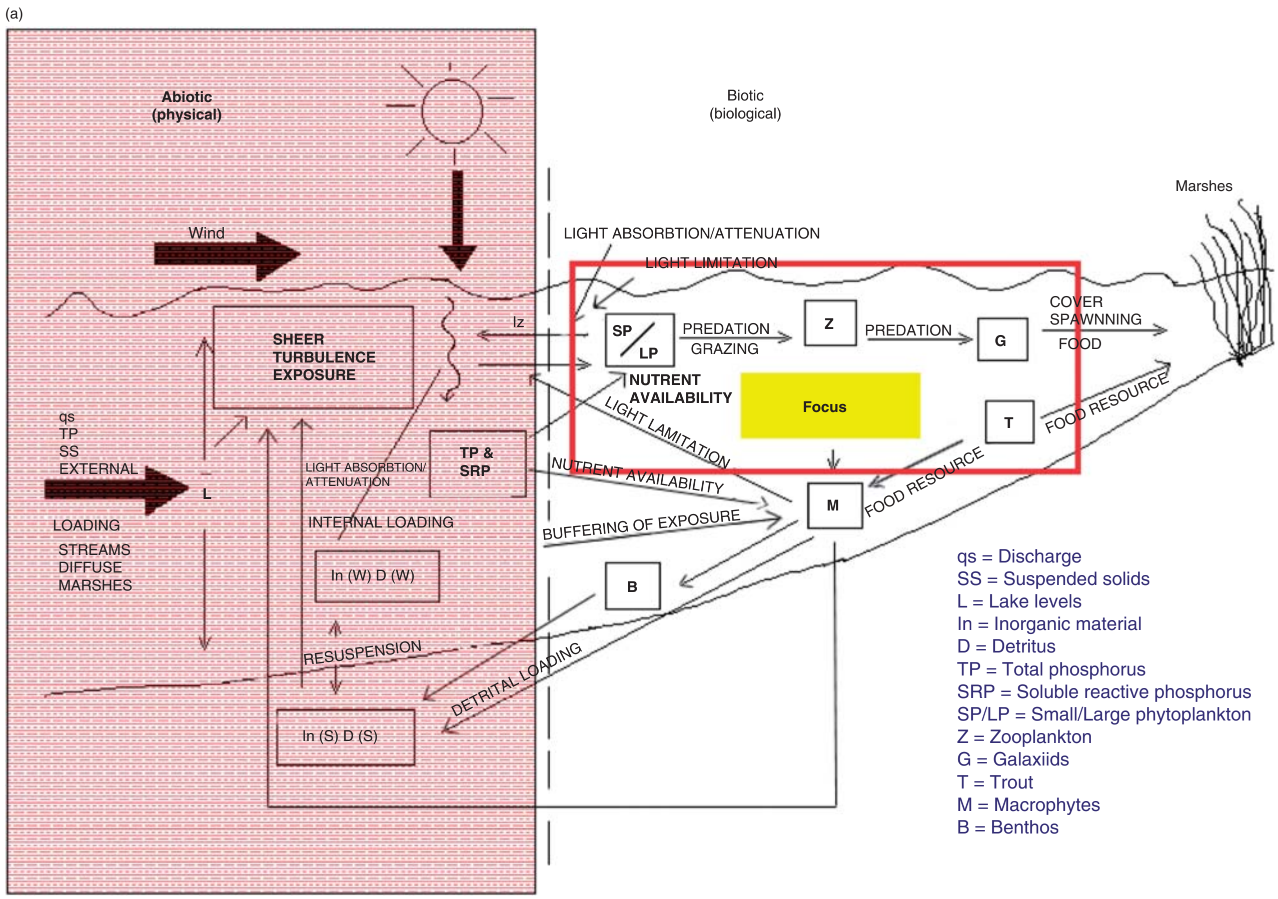

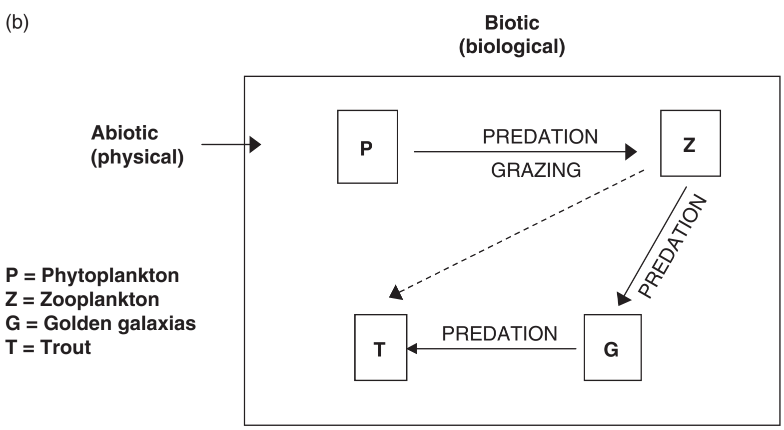

Figure 1 (a) A schematic diagram outlining the abiotic and biotic pathways present in a shallow lake ecosystem. (b) A simplification of (a), depicting the dominant biotic interactions between higher trophic levels in lakes Sorell and Crescent.

discipline that studies organisms in relation to each other and their surroundings without taking into ac count the modification of environmental conditions by humans. This includes changing the biotic environment (e.g., by disrupting food chains, or by intentionally or unintentionally introducing new species into an eco system) and the abiotic environments (human induced climate change).

Tensions in Ecology

Scientific ecology claims to be objective. It follows tra ditional scientific approaches to knowledge, such as quantification and experimentation. Normative ecology is an attitude of mind that focuses on reverence for the Earth. This reverence comes from intimate contact with the Earth as subject. Ontologically, scientific ecology perceives both competition and cooperation, but often implicitly emphasizes competition for food, habitats, and light. Nature is generally perceived to be dynamic, with no sense of balance. Ecologists in the scientific ecology approach often present the findings of scientific studies to political decision makers.

In contrast, normative ecologists often identify a balance of Nature, although it is recognized that Nature is dynamic. Normative ecologists generally critique political institutions, and are more likely to take direct action against existing institutions to achieve their normative project. This approach stems largely from an ontological position that believes Nature is based largely on cooperation, with competition occurring within limited domains. The cooperation in Nature includes humans cooperating in political action.

These tensions partly reflect changes in ecology over time. The first generation of white American ecologists, such as Frederic Clements, understood natural systems as progressing towards a climax. Once achieved, this was the balance of Nature that humans should not disturb. This approach to ecology underpins some normative ecology positions, but it has been thoroughly critiqued from a scientific ecology basis by A. G. Tansley, who coined the term ‘ecosystem’, and more recently by Daniel Botkin and the nonequilibrium thermodynamic ecologists who stress dynamism and change in ecosystems. The quest for order in early scientific ecology, and in some normative ecologies, is predicated on the notion of balance, and the concept of thresholds in a system that is out of balance would attempt to revert to a balanced state. This is apparent in many human ecology approaches to the issue of population, whereby studies on overshoot among animal species are used as a warning for what is likely to occur if the human population exceeds the capacity of the Earth to support this population.

These tensions between normative and scientific ecology, and within various normative ecologies, are deep divides based on ontological and epistemological differences, perceptions of Nature, and normative projects. Within these tensions is the question of the similarities between processes and structures in human society and those of animals, plants, and the Earth. Geographers potentially have an important role in moving debates beyond environmental determinism, and dualisms of human and Nature, to focus on the ongoing processes that constitute humans as part of Nature.

There is another tension that derives from ecology’s perceived lower positioning within the hierarchy of scientific disciplines. While this entire concept is open to challenge, there is some degree of consensus among scientists that certain branches of science, notably physics and chemistry, are more worthy of the title science than other disciplines, notably ecology and biology. This has been attributed to the normative uses of the term ‘ecology’, with other science disciplines keen to distance themselves from particular political agendas. It is also due to the debate about whether ecology has laws, similar to physics and chemistry. If ecology cannot produce such laws, is it really a science? At best it can be seen as a soft science, in relation to the hard sciences such as physics and chemistry, or as some ecologists would argue, in the process of refining hypothesis through scientific research in order to develop laws of ecology. Debates have centered on the definition of a law in science, and whether it allows exceptions, is explanatory, is predictive, and whether it invokes an idealized situation. In this regard, some writers within ecology argue that ecology should be seen as similar to other scientific disciplines such as physics.

Tensions around ecology as a science, and its engagement with nonscientific endeavors, are highlighted in the debate about the use of either ecological restoration or restoration ecology. Ecological restoration has been defined as an ensemble of practices that constitute the entire field of restoration. These include restoration ecology plus the participating human and natural sciences, politics, technologies, economic factors, and cultural dimensions. By way of contrast, restoration ecology means the scientific practices that, together, constitute a subdiscipline of ecology. In order to simplify the definition, ecological restoration has been defined as the process of assisting the recovery of an ecosystem that has been degraded, damaged, or destroyed. This raises questions about so called natural ecosystems and human involvement in the degradation and repair of ecosystems. Some authors challenge the explicit lack of value recognition in ecological restoration, accusing it of being an overly scientifically oriented activity. They highlight that values are crucial in deciding which ecological processes or attributes of a landscape should be restored. This work is similar to many of the tensions between human geographers and scientific ecologists, and while these tensions may not come to a head in debates (although in the past they have done so), this means that certain ecological approaches are not incorporated into human geography and knowledge which is generally recognized in human geography can be absent in ecological analysis.

Normative Ecology – Sub-areas

Within what may be termed the normative ecology approach, there are a number of approaches that range from left to right on the traditional ideological spectrum, although many proponents would reject this ideological spectrum and place ecology above, beyond, or in front of it. The range may also be understood as being from strongly anthropocentric to ecocentric, from romantic to technocratic, and from anarchist to authoritarian. The following approaches are indicative of some uses of the term 'ecology' in ways that differ from scientific ecology. Some geography literature is claimed as being representative of more than one tradition, hence highlighting the fluidity of the categories and the search for articulations between some of the approaches discussed below.

Cultural Ecology

This approach to integrating ecological concerns with human geography is gaining influence, particularly, in relation to the issues of biodiversity and the establishment and maintenance of protected areas. The approach is a critique of environmental management approaches that do not adequately consider the impact of environmental decisions on the livelihoods of people, nor do they consider the importance of local peoples' knowledge and participation in contributing to positive environmental outcomes. It is a critique of environmental decision making at a distance. The concept of scale employed in this approach is a relational concept of scale, unlike some scientific ecology approaches that predetermine the scale and then attempt to work within this limitation. Having said this, the approach favors the scale of the household, and is noticeable in that it differs from political ecology which has been understood as concentrating on the political, economic, and social impacts of resource related developments but not necessarily at the micro level of the household. The cultural ecology approach shares similarities with many other environmental approaches that favor decentralization of decision making and with feminist ecological approaches that focus on the micro scale of the household.

Deep Ecology

This is a radical environmental philosophy which argues that Nature has inherent rights to exist, that humans are part of Nature, and that our ecological awareness comes from experiencing ourselves within Nature. The term emerged in 1973 when the Danish philosopher, Arne Naess, distinguished between shallow ecology (technocentric, anthropocentric, and reformist) and deep ecology. Deep ecology was popularized in North America in the 1980s, and led to the emergence of activist groups, including Earth First! Deep ecology values all life as being intrinsically important, not simply as being of use to human beings. It is ecocentric. Deep ecology advocates dismantling the dominant socioeconomic systems through which humans appropriate Nature and argues for biocentric equality so the desire by humans to dominate Nature is eliminated in favor of humans, as one species, living in Nature. Humans, like other species, must respect Nature's limits and thresholds in order for all forms of life to live on a sustainable basis. The quality of life, rather than the quantity of life, is emphasized. Deep ecology has been critiqued as romanticism, fundamentally flawed in its conception of nonhuman parts of Nature having intrinsic value, and antihumanist in its approach, for its inability to distinguish between different parts of humanity (men and women, rich and poor, different cultural beliefs and practices) and its celebration of outdoor experiences in Nature. This approach has been influential in the environmental movement, but has had little impact on human geography to date.

Feminist Political Ecology

This approach draws on both feminist ecology and political ecology. It has been influential in human geography, perhaps more so than many other feminist ecological approaches. Ecofeminism is the lens through which many feminist geographies of Nature have been produced. While recognized as a broad church, this approach does not necessarily draw on ecology. Feminist political ecology questions the construction of identity, particularly as a basis for situating the researcher in relation to the research being undertaken. It recognizes multiple subjectivities and seeks to combine traditional geographical research techniques with feminist approaches of participatory mapping and oral histories. The approach also recognizes the gendering of environments. Recent work within this tradition has also noted the rural bias within political ecology, and attempted to address this issue.

Global Ecology

This approach is the radical political critique of global and international institutions and events that are perceived as reformist environmentalism. Scale is crucial here because although these authors adopt a global focus, they are very concerned with the links between local and global representations and impacts. Other uses of this term tend to assume a more physical, and less overtly political, approach, and include the study of the interactions among the Earth's ecosystems, land, atmosphere, and oceans to understand how these interactions shape the behavior of the earth system. It is global in the sense that the Earth is seen as a single unit. The approach is not favored in human geography because of the emphasis on multiple scales and the recognition of heterogeneity rather than a single system.

Human Ecology

Human ecology studies the relationships between people and their social and physical environments. It is an extension of concepts drawn from ecology to the social realm. The human ecology approach developed in the second decade of the twentieth century, but was made famous in the 1920s by the Chicago School of sociologists, including Park, Burgess, Thomas, and Wirth.

These sociologists imported concepts and analogies from ecology (such as community, competition, disturbance, climax equilibrium and invasion, and succession) to de velop theories and models on human society. They were not explicitly concerned with Nature, but they saw similarities in the processes of conflict and adjustment undertaken by waves of immigrants to the mid Western cities of the USA to the struggle for existence that Charles Darwin had seen in Nature. The term came to geography via sociology, when a University of Chicago geographer, Harlan Barrows in 1923, defined geography as the science of human ecology. This definition gave the term a different sense than that used by the sociologists. Barrow's focus was on human adjustment to physical environments. It was seen as deterministic, and was not favored by geographers and sociologists.

The term was used less in geography until the early 1970s, following the United Nations Conference on the Human Environment (Stockholm 72), when it became more influential as geographers responded to research agendas about the human environment. In the latter versions of human ecology, humans were not perceived as being different from other species because of their ability to adapt to new environments and environmental change through the use of culture; instead, humans were one species among many that interact with bounded en vironmental conditions. The re emergence of human ecology was critiqued within geography. As early as the 1970s, human ecology was seen by some geographers as being a futile attempt to regain something natural. With the rise of Marxist geography in the early 1970s, geo graphers tended to look more towards political and economic structures, rather than natural processes, to explain changes in a range of environments.

Contemporary human ecology, in contrast to the work of geographers nearly 90 years ago, understands the Earth as being finite but not a closed system, that people must use Nature to survive and that humans, unlike other species, regulate their environment not just through physical measures such as temperature control, but also through symbolic constructions and understandings. The recent use of the term in geography has been somewhat limited, possibly resulting from various interacting fac tors, including the vagueness of the term through use in many disciplines and its connotations with ideas that were popular earlier in the twentieth century.

Industrial Ecology

Industrial ecology is an approach that applies the eco system metaphor and model to suggest that industrial systems should be restructured in order to make them compatible with the way natural ecosystems function. It is promoted as an approach to close industrial production loops and reduce waste, thereby making better use of resources and preventing the overuse of raw materials.

In this approach, corporations should optimize the consumption of energy and materials, minimize waste generation, and use the effluents of one process as the raw material for another process. The existing industrial systems are seen as being antithetical to Nature in that they have linear throughput, whereas in Nature there is roundput as loops are closed. Terms such as 'roundput', 'closing the loop', and 'industrial symbiosis' are used to distinguish industrial ecology from existing throughput systems of industry that are materially open and rely on nonrenewables. This approach has been used by some leading environmental thinkers to offer critiques of capitalist and socialist modes of economic production as being expansionist and linear and therefore non sustainable.

The industrial (eco)system is seen as analogous to a natural ecosystem, where there is no waste. This ap proach has been somewhat influential in engineering, chemistry, and other scientific and technological discip lines, but has been both lauded and critiqued by urban planners and geographers for its selective use of eco logical metaphors and its inadequate treatment of social and political–economic processes. The tensions between firms in competition in the marketplace, and firms co operating to develop industrial ecology processes, is under theorized in the industrial ecology literature. Work by human geographers on trust has added an im portant perspective beyond the technical literature or unexamined assumptions about the nature of competition and cooperation.

Political Ecology

This is a term that has been used by both anarchist and Marxist inspired authors in international development. The latter use of political ecology has been more in fluential within geography, but has moved somewhat from its Marxist origins and is often the subject of tension be tween the politics and the ecology. Political ecology is seen as being at the intersection of a number of disciplines (see **Figure 2**), including human geography. Within pol itical ecology there is emphasis on understanding power relationships, and how these relationships are constituted, perpetuated, and changed through engagements with na ture. The use of political ecology approaches to the understanding of issues such as water allocations, land management, deforestation, and so on, has increased in recent years. Calls for a Third World political ecology have emphasized the importance of environments in de veloping countries – their discursive construction being different than that of environmental discourses in de veloped countries – and the political–economic systems that connect developed and developing countries that

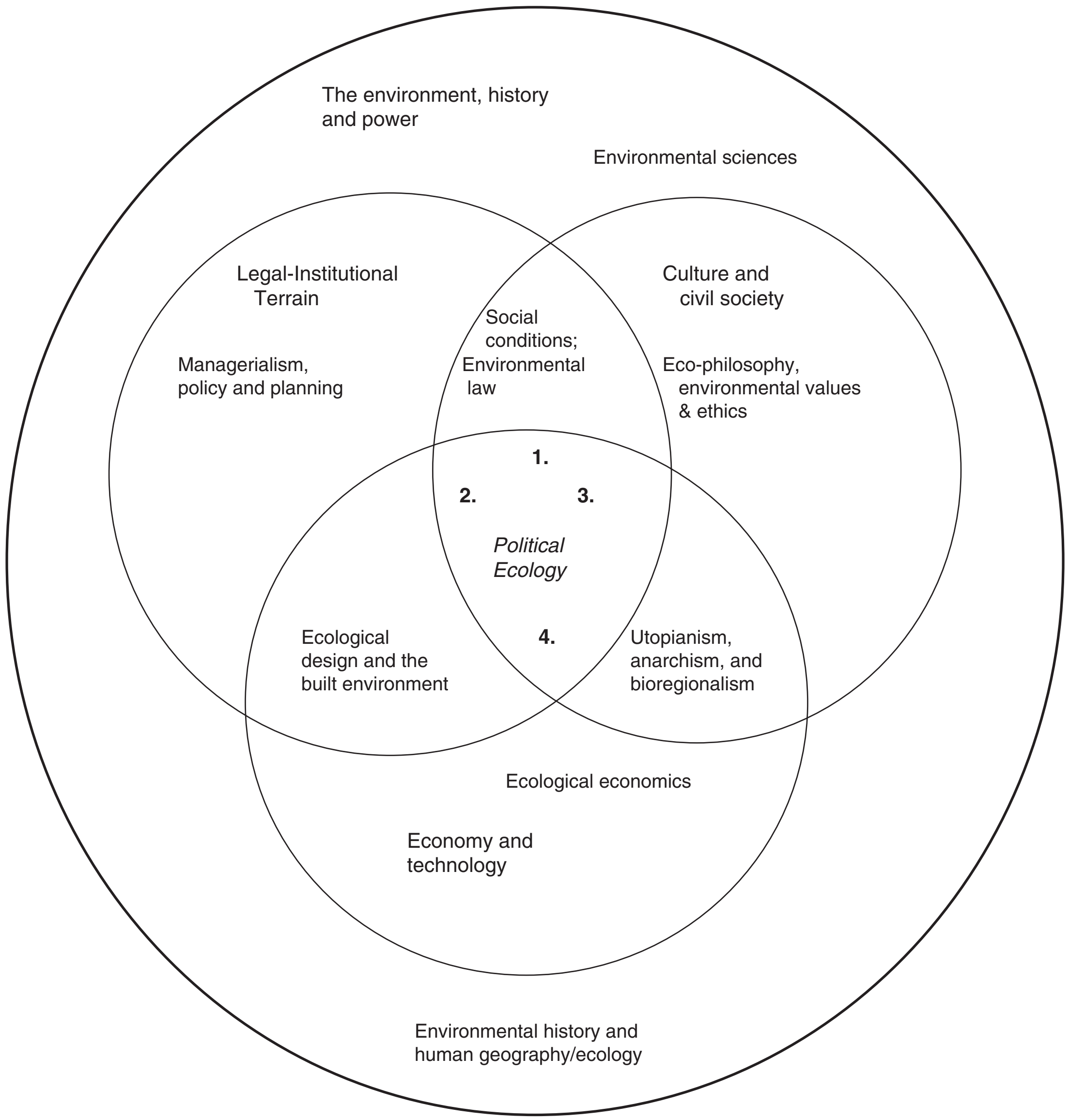

Figure 2 The sustainable development literature located within its four intersecting spheres of concern. The numbers locate the four intersecting challenges of political ecology: (1) holism, (2) empowerment and community-building, (3) social justice and equity, and (4) sustainable production and reproduction.

contribute to environmental impacts. This has been in fluential in human geography, particularly with the rec ognition that meso and macro scale political and economic forces establish a context for local environ mental engagements. These engagements concern both humans and the ecology of nonhuman systems, thereby bringing in nonhuman actors.

Social Ecology

This approach is most closely associated with the eco anarchists. In this approach, there are no hierarchies in Nature that serve as a natural foundation and justification for the existing hierarchies in human society. Instead, Nature is a web of life. It is believed that human dom ination of Nature will persist as long as hierarchies within human society are maintained. The term 'social ecology' has been used in human geography, largely to denote an approach that builds on the work of the Chicago School but looks at the processes of formation and succession through an understanding of political economy, the div ision of labor, structuration theory, and housing markets.

Urban Ecology

Urban ecology applied many of the concepts of human ecology to the urban realm. The city was understood as an ecosystem wherein areas within the city experienced processes such as invasion and succession. This applied to the movement of migrant populations, particularly, in regard to a portion of the city being the location of successive waves of groups from different backgrounds moving through an area.

The study of urban ecology grew in importance following the Habitat Conference in Stockholm in 1972. Some uses of the term 'urban ecology' are premised on the balance concept, as in the ecological city being a city that is in balance with Nature. Urban ecology, as it has developed in the human geographical literature, goes beyond the traditional human ecology in that it recognizes populations, technology, organization, and environment as important variables to be employed in the study of people in cities. Importantly, the concept of environment drawn from ecology is premised on the city being an open system rather than a closed system and therefore is not moving towards a state of equilibrium. This means that there is not a balance of Nature in the social aspects of the city. Urban ecology in the 1970s, influenced by Marxist geography, recognized the important role of ideology and values in shaping ecological processes. Ecology as scientific, implying freedom from values, was rejected in favor of an ecology that was about influencing ecological systems, but being aware of the values promoting, and being promoted by, changes made by humans.

Urban Political Ecology

Urban political ecology combines elements of urban ecology and political ecology. It seeks to expose the complex socioecological relationships that shape urban environments. It can be understood as a political economy of urban systems, but focusing on the environmental aspects of the city. In this sense, it builds on a long tradition of academic research into human–Nature relationships that have led to the construction of urban environments. Within urban political economy there is a focus on power and how it is exercised in shaping the city and the lives of people, and other species, in the city. This involves both structures and individual actions. The political ecology background is predominantly Marxist, or at least Marxist inspired, and builds on earlier Marxist urban geography. There are many different and sometimes overlapping approaches within urban political ecology, including those that emphasize urban metabolisms, those that challenge neoliberalism as a system of governance, and those that focus on a particular issue, such as water, poverty, or transport.

The Influence of Ecology in Human Geography

Ecology has been influential in many other disciplines, including anthropology, sociology, and geography. It has influenced both human and physical geography in different ways, and at differing times in the history of the disciplines. Some of this influence is identified above, in relation to specific applications of ecological thought within particular normative ecological approaches. Physical geography drew upon the work of geologists and of evolutionary biologists as it emerged as geomorphology and biogeography. In relation to human geography, ecology has been identified as a link between the natural and the social scientists. This is because of its content, its language, because of issues of scale, and because the methodologies used in the study of ecology are more closely aligned with social sciences than are other scientific disciplines. This is partly due to the origins of ecology in philosophical debate, but also due to the increasing recognition that humans are part of ecosystems and that it is impossible to study ecology devoid of human impact, and that environmental management (even from a scientific basis) cannot be separated from human values and beliefs.

Scale is important for both geographers and ecologists, although there is no unit of scale that is paramount within each discipline. At various times in the history of each discipline various scales have been privileged (e.g., regional geography, and more recently the corporeal geography that has been strongly influenced by feminist theory). Ecology may be concerned with individual organisms (including behavioral ecology, feeding strategies), populations (including numbers and distribution of a species), entire communities (e.g., competition between species or predator–prey relationships), or at larger landscape scales (including how spatial variation in the landscape affects the distribution and flow of energy, materials, and individuals).

Geographers have also used work from within the scientific ecology tradition. There have been numerous instances of ecological concepts and content being borrowed from geographers, economists, and anthropologists by ecologists, and vice versa. The fields of human ecology, urban ecology, industrial ecology, and landscape ecology are attempts to link the separate disciplines and/or to overcome the binary between human and Nature. For instance, urban ecology challenges the dualism between natural and built environments, but still tends to emphasize biophysical structures and processes relative to social structures and processes. Industrial ecology is increasingly being seen as a way to move towards sustainable development by seeing industries not as isolated entities, but as part of an industrial (eco)system. Outputs from one industry become inputs into another industry, thereby closing waste loops, and conserving resources and energy.

Human ecology drew on scientific ecology in a number of ways. These included the application of concepts and principles from scientific ecology, particularly concepts such as competition, succession, and their consequences. Some earlier work within human geography took ecological concepts such as competition and succession, and saw within the city a natural order. More recent critiques of this approach have highlighted the importance of political–economic systems and structures in creating areas within cities, and have noted that studies of these areas are like snapshots of a point in time. Human ecology research also included area based studies that combined the physical features of an environment, described in ecological terms, with the socioeconomic characteristics of an area. There were also studies undertaken that plotted observable social phenomena such as crime rates and mental disorders, and sought to identify the ecological correlation between the social and environmental mapping of an area.

Arguably the most important influence of ecology within human geography has been the provision of methodology. This was particularly the case when ecology was understood as the ordering of natural science knowledge in a rational, comprehensible, and comprehensive way. When systems geography was in vogue in the 1960s (which was consistent with the popularity of systems theory in a number of disciplines), the transference of methodologies applied in the study of ecosystems to the study of land use planning, geomorphology, and cities was considered unproblematic. Recent developments in geography that question the basis of classification systems, emphasize the importance of discourse and highlight that situated knowledge means that the transference of ecological methodologies to human geography is curtailed. The benefit of linking geography closely with ecology in the 1960s was that it drew geography closer to mainstream scientific thought, thus giving geography greater prestige and access to resources. This is still pursued in many universities, largely due to the importance of resources, such that geography has become geographical science. Ironically, the change in name has occurred when the ways of doing geography have moved away from the methodological approaches that are considered rigorous in contemporary scientific ecology.

In the 1960s, it was possible for some geographers to say that the discipline stood in much the same position in relation to ecology as it did 100 years previously. This can be summarized as geographers borrowing concepts from ecology, and applying them in geography. Since the zenith of systems geography, ecology has contributed less to the methodological approaches within human geography and, arguably, more to the content of human geography when geographers draw on ecological knowledge to research complex environments and the perception and use of these environments by people.

Conclusion

Ecology is a complex term that is a scientific discipline, used to refer to various environmental philosophies, and is used in popular vernacular to refer to the environment. Ecology engages with many issues that are familiar to human geographers, with ideas and concepts moving between the two disciplines. This is likely to continue, albeit in a critical manner, as popular awareness and concern about the importance of the environment is likely to increase in forthcoming years. Within human geography, the multitude of theoretical approaches that draw on, or intersect with, differing notions of ecology, is likely to be expanded in future years as a result of the theoretical developments in both ecology and geography. It could also be anticipated that links between the various forms of normative ecology within human geography will continue to develop, as researchers identify the strengths and weaknesses of various approaches, the potential for articulation, and the need to establish new research agendas to respond to changes in the world.

See also: Anarchism/Anarchist Geography; Biodiversity; Chicago School; Conservation and Ecology; Darwinism (and Social Darwinism); Determinism/Environmental Determinism; Environmental Studies and Human Geography; Environmentalism; Political Ecology; Radical Environmentalism; Systems Theory.

Further Reading

Barrows, H. (1923). Geography as human ecology. *Annals of the Association of American Geographers* 13, 1 14.

Berry, B. and Kasarda, J. (1977). *Contemporary Urban Ecology*. New York: Macmillan.

Botkin, D. (1990). *Discordant harmonies: A New Ecology for the Twenty First Century*. Oxford: Oxford University Press.

Bramwell, A. (1989). *Ecology in the 20th Century*. New Haven and London: Yale University Press.

Chorley, R. (1973). Geography as human ecology. In Chorley, R. (ed.) *Directions in geography*, pp 155 169. London: Methuen & Co.

Colyan, M. and Ginzburg, L. R. (2003). Laws of nature and laws of ecology. *Oikos* 101, 649 653.

Devall, B. and Sessions, G. (1985). *Deep Ecology: Living as if Nature Mattered*. Salt Lake City, UT: Peregrine Smith.

Keil, R. (2005). Progress report Urban political ecology. *Urban Geography* 26, 640 651.

Lawrence, R. (2003). Human ecology and its applications. *Landscape and Urban Planning* 65, 31 40.

Luke, T. (2002). Deep ecology: Living as if nature mattered: Devall and Sessions on defending the earth. *Organization and Environment* 15, 178 186.

Merchant, C. (2005). *Radical Ecology: The Search for a Livable World* (2nd edn.). New York: Routledge.

Naess, A. (1973). The shallow and the deep, long range ecology movements: A summary. *Inquiry* 16, 95 100.

Worster, D. (1977). *Nature's Economy: A History of Ecological Ideas*. Cambridge: Cambridge University Press.

Zimmerer, K. (2006). Cultural ecology: At the interface with political ecology The new geographies of environmental conservation and globalization. *Progress in Human Geography* 30, 63 78.

Relevant Websites

http://www.stetson.edu
Association of American Geographers: Cultural and political ecology specialty group.

http://www.che.ac.uk
Centre for human ecology.

http://www.esa.org
Ecological society of America.

http://www.is4ie.org
International society for industrial ecology.

http://www.ser.org
Society for ecological restoration international.

Economic Crises

M. Webber, The University of Melbourne, Melbourne, VIC, Australia

Glossary

Accumulation The accumulation of capital is the increase of wealth. In orthodox economics, the equivalent term is capital formation. Capital can be accumulated in tangible means of production (real investment) or as finance (in paper assets or such nonproductive assets as residences). Orthodox economists also identify human and social capital formation, but these terms confuse investment with its effect.

Asset Appreciation/Depreciation Productive assets are used to produce goods or services, rather than held principally for resale. For example, machines, buildings, patents, or licenses can be fixed assets of a business. Financial assets include financial instruments (such as stocks) and real estate, the price of which depends on supply and demand rather than their costs of production. Appreciation is the increase in value of an asset, and depreciation the decrease. An asset-based economy is one in which growth is based largely on the appreciation of financial assets.

Capital Money invested in assets for profit.

Development State/State-Led Development It refers to countries in which the state has, through macroeconomic planning, led private capital to develop specific industries that are thought to serve the national interest. The term is especially applied to countries in East Asia in the late twentieth century.

Equilibrium A state in which economic actors have no incentive to alter their decisions. For example, in a competitive market, there is a price at which the quantity demanded by consumers is equal to the quantity supplied by producers. At this price, consumers and producers are satisfied with their decisions. According to trade theory, a trade imbalance (country A consistently exports more than it imports) means that A buys more foreign currencies than foreigners buy of its currency; therefore, the price of A's currency tends to rise, making its exports relatively more expensive. Such price changes in A's currency continue until the trade imbalance is corrected.

Governance As a process, it uses social institutions and structures to develop and manage consistent policies and decision-rights for a given area of responsibility (a corporation or a state, for example). Government is a mode of governance of the state.

Rate of Profit The amount of profit in relation to the size of the investment. In Marxist political economy, the rate of profit is surplus/capital invested, where the surplus includes profits, interest, and rent.

What Is an Economic Crisis? The Case of East Asia

In May 1997, global investors began to sell East Asian currencies. Despite rapidly rising interest rates, currencies fell sharply in Thailand, the Philippines, Malaysia, Indonesia, and Singapore. The Hong Kong dollar also came under attack, but was rigorously defended by the Chinese government. By November 1997, as it became clear that eight of Korea's 30 largest chaebol were either bankrupt or financially strained, the Korean won was being forced downward. By early 1998, the currencies of Australia and New Zealand had been drawn into the vortex. The Taiwanese dollar ($NT) fell steadily too.

The only country in this region that largely escaped the turmoil was China: its currency was not freely traded internationally, and it supported the Hong Kong dollar. The Japanese yen also escaped devaluation. Between 1998 and 2002, the Singapore, Taiwan, and Vietnamese currencies had been devalued by an average of 15–20% against the US dollar, while the currencies of Australia, Cambodia, Korea, Malaysia, New Zealand, and Thailand all traded at about 30% below 1997 levels. The Philippines peso and Indonesian rupiah fell even more sharply. Ten years after the start of the crisis, the currencies of Australia and New Zealand have recovered their lost values, and those of Korea, Singapore, and Taiwan have largely recovered. But the Cambodian, Malaysian, Thai, and Vietnamese currencies are still trading at 30% below the 1997 levels, the Philippines' currency has fallen to almost a half of its precrisis value, and the Indonesian rupiah is still worth only a third of its precrisis value.

The crisis also affected rates of growth of output. The most obviously affected have been Thailand, Korea, and Indonesia, all of which were forced to seek the International Monetary Fund (IMF)'s support. However, other countries in the region have suffered too, particularly Malaysia, Singapore, and Taiwan, which are closely connected to the three principal patients. The impact of the crisis on levels of output is startling: in Korea and Taiwan by 2005 levels of output were 20–25% below what they would have been without the crisis, and in the other four countries, 30–35% (see **Figure 1**).

This was the 'Asian financial crisis', one of the most dramatic of a series of financial crises that have affected developed and developing countries since the late 1970s. Similar events in the developed world include the attack on the sterling and lira in 1992 and on the Australian

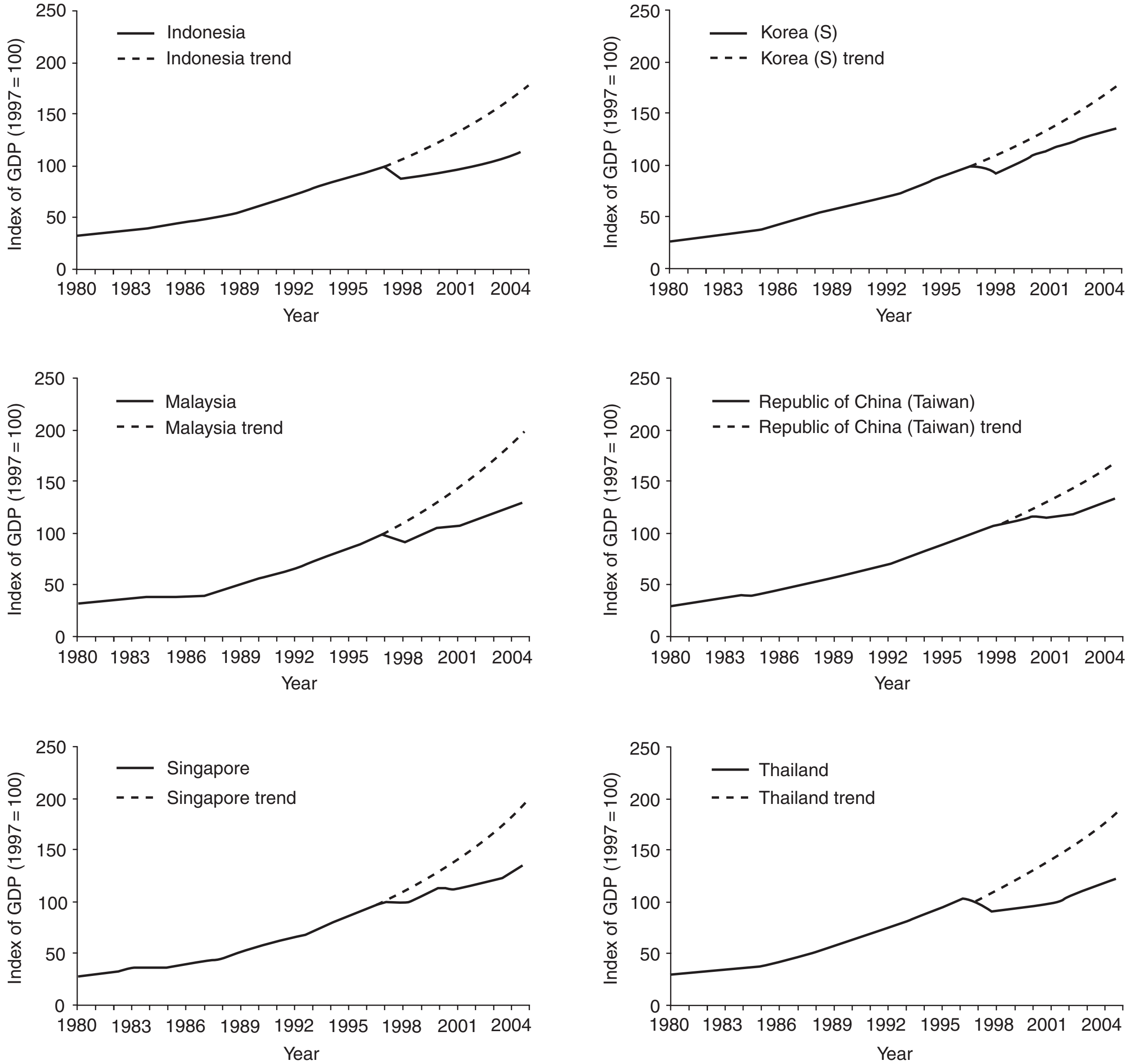

Figure 1 Index values (1997 1) of gross domestic product in constant US dollars of 2000, 1980 2006 in six East and Southeast Asian countries. Note: solid lines track actual levels of GDP; the dotted lines post-1997 indicate what GDP would have been had it continued to grow at the average rate for 1988 97. The difference is an indication of the effect of the 1997 Asian financial crisis. From World Bank, World development indicators; for Taiwan, Director General of Statistics, Executive Yuan, Republic of China.

dollar in 1985, and the banking crises of Finland (1991–93), Norway (1987–89), Spain (1977–85), Sweden (1991), and the USA (1984–91). This crisis is important for three main reasons.

First, for many people in East and Southeast Asia, the crisis is likely to be a defining event of their lives. Throughout the region, the incidence of poverty has increased; divorce and suicide rates in Thailand sky rocketed after the crisis; the number of infants abandoned after birth or placed in orphanages increased; and there was a rise in the number of arrests for drug related criminal activity and property crimes. Within Indonesia, people were killed in riots triggered by the economic collapse.

Second, governments have been ejected and political power within the region has been substantially realigned as the crisis has forced changes to the political economy of the Asia Pacific region. The IMF has become a dominant force in world politics: the US$110 billion in financial assistance to Indonesia, Korea, and Thailand that the IMF brokered has enabled it to demand a variety of reforms in those countries. The Asian Development

Bank (ADB), by contrast, was virtually ignored, and the World Bank's concern about poverty and sustainable development carried little weight against the demands for market reform. Similarly, despite the magnitude of its assistance to the region, Japan has lost authority in comparison to the USA and China.

Finally, ideologies have gained and lost too. By the argument that the cause of the crisis is local, the entire idea of a miracle system has been dismantled. By Japan's own inept management, the whole idea of state led de velopment is ridiculed. To the IMF, to most of the Western world's economists and press, the crash in East Asia was not simply a change in rates of growth, but was equally the signifier of the failure of a development model. The localization of the social and political insti tutions that sustain markets by regulating, stabilizing, and legitimizing them was ignored under the rush to impose American styles of economic governance.

This article explains some of the theories that seek to interpret such events. It briefly outlines the explanations typically offered for the crisis in East Asia, arguing that these pay insufficient attention to the fact that similar crises have affected a great variety of countries – not simply developing countries with supposedly poor standards of governance. Following that argument, a theory of crisis that is in the Marxist tradition is outlined.

Explaining the East Asian Crisis

Most explanations of the financial crisis in Asia are real and specific, in the sense that they identify characteristics of East Asian economies and societies to which investors reacted. Of course, crises are overreactions to events as investors panic, in a chaotic follow the leader process. A real, specific explanation, therefore, means an explan ation that starts with events or conditions in the real economy of a specific place: something wrong in East Asia causes investors to change their behavior (and per haps panic). The chain of causation runs: East Asia > finance system.

What economic events in the year or two preceding the crisis may have triggered the change in expectations and the panic? Perhaps the most common explanation of what went wrong in the short run is that current account deficits widened, causing exchange pressures to mount. Once doubts about the solvency of the borrowers arose, investors began to withdraw financing and exchange rates fell. As the economies came under pressure, property and equity investments lost value and the currencies became the targets of speculation, facilitated by financial liber alization, which made it easy for speculators to take advantage of falling currencies.

But this argument leads to another question: what were the underlying social conditions of economic governance that themselves permitted (or encouraged) those events? Some of the most politicized interpretations of the crisis have focused on this second question.

One of the most influential politicized interpretations is 'failure of governance'. This thesis takes several forms. The first is a story about homegrown vulnerabilities: lack of transparency, insufficient bank regulation, and personal influences on credit allocation. The second is a story about moral hazard – the belief by Western banks that Asian banks and corporations were in effect guaranteed by their governments, so the risks were perceived to be at gov ernment rather than private levels. However, the East Asian financial institutions were effectively unregulated, making loans that were risky and that drove up the prices of assets in East Asia. These two forms of argument identify a failure of governance, either state or corporate: lack of adequate controls to ensure safe lending.

A second long run explanation begins with changes in the nature of capital flows into developing countries. Private capital flows to developing countries increased sixfold during the 1990s, providing finance of a kind and magnitude that such countries had previously not learned to manage. These huge inflows of capital entered what were more or less development states, in which high rates of household savings were deposited in banks, which 'on lent' them to firms. Firms had high debt–equity ratios, but grew fast. This bank based, high debt model became the development state in Japan, Korea, and Taiwan, steering market forces in line with national goals (**Table 1**).

Because the economies of East Asia were successful, they attracted high levels of foreign capital. That capital, leveraged through the domestic banking systems, drove very rapid growth in supply. The argument goes: with financial liberalization, with faster and easier flows of short term capital, and, in some countries, with dem ocratization, firm–bank–government collaboration became more fixated on short term issues rather than long term development objectives. Particularly, the management of high levels of indebtedness proved problematic when capital could more easily flow across borders and burst property bubbles. Those problems led in the end to a classic property and equity speculative bubble, which eventually began to burst in Thailand in 1995.

However, if the crisis was primarily the result of deep rooted problems ('naughty' East Asians or vast inflows of unregulated foreign capital), it should not have taken the world by surprise. Until the third quarter of 1997, there was widespread optimism about the region among international bankers (the spreads between Asian and comparable US interest rates narrowed during the 1990s through the first half of 1997), credit rating agencies (ratings remained unchanged throughout 1996 and the first half of 1997), securities firms (published forecasts), and the IMF itself (in its published country assessments).

Table 1 Capital flows into developing countries, 1990 97

	1990	*1992*	*1994*	*1996*	*1997*
Private financial flows[a]	29.8	57.1	107.1	170.6	156.5
Official financial flows	56.3	55.5	45.7	40.8	100.0
Foreign direct investment	32.8	49.6	90.4	128.7	125.0

[a] units are billions of US$.

From: Howell M. (1998). Asia's 'Victorian' financial crisis, paper pre sented at the East Asia Crisis workshop at the Institute of Development Studies, http://www.ids.ac.uk/ids/research/howell.pdf.

Real analyses of the causes of the crisis assume that events in the real economy drove the financial system. Thus, they pay far too little respect to the relations be tween the financial and the real economy that are re vealed by this and earlier banking crisis. Similar crises occur in even the best regulated of financial systems. Perhaps, then, the central theoretical implication of this crisis is that the real cause is irrelevant? Perhaps, that is, rather than looking for the cause of the financial crisis in the conditions of the East Asian economies, we should recognize that the financial system is now driving itself, independent of conditions in the real economy? How does such a state of affairs come to be?

Theory of Crisis

In order to understand the emergence of such a financial system, let us go back to the real economy of production and exchange. The history of capitalist accumulation suggests that growth has continued now for several centuries, but that periods of rapid expansion are limited. Many Keynesians and Marxists argue that economic expansion is checked by forces internal to the capitalist mode of production.

Marx's theory of the falling rate of profit provides the foundation on which the account rests. Versions of the theory motivate different understandings of the historical economic geography of the post war world. Theories of the rate of profit and of fluctuations in growth are hotly con tested, but the variety of positions in that debate reflect different and partial models of the growth process, all of which insist that growth follows equilibrium characteristics.

Accumulation is a defining feature of the capitalist mode of production. Capitalists – owners of the means of production and controllers of the social surplus – have constantly to readvance capital in search of profit. How ever, techniques of production change and, like Smith and Ricardo before him, Marx argued that the process of capital accumulation was subject to a series of checks and contradictions – internal barriers to the self expansion of capital. Three barriers to the accumulation of capital were identified: crises arising from underconsumption (after prolonged growth, the capacity of the economy to produce exceeds the demand for commodities), from a squeeze on profits (after prolonged growth, increasing demand for labor leads to an increase in wages relative to profits), and from the replacement of labor by plant, machinery, and equipment in production (which means that exploitable labor – the source of profits – comprises an ever smaller proportion of the cost of production).

Early writers, including Marx, identified these bar riers to the accumulation of capital, but their analyses were qualitative and could not specify the relative mag nitudes of the forces tending to increase and the forces tending to decrease the rate of profit. The technical ap paratus of theory was not able to characterize this problem until the 1960s, when a stream of attempts were made to specify how all these variables are related. These formal mathematical models of the economy generally relied on the assumptions that prices are at equilibrium, that real wages are constant, that all that is produced is sold, and that all capital is productively employed; with these assumptions, the models demonstrated that cost reducing technical changes increase the equilibrium rate of profit; that is, there is no real internal barrier to the self expansion of capital.

Actually, though, economies are never in equilibrium. Instances of a commodity do not all command the same price; the demand and supply of commodities do not match; wages depend not only on the demand and supply of labor but also on historical and geographical con ditioning; capital is under or oversupplied. Productivity change is a probabilistic process in which marginal ad justments and learning reflect imperfect attempts to find better technical conditions rather than optimal states. Location is dominated by history and by local attempts to overcome history: the East Asian dragons and other newly industrialising economies (NIEs) have demon strated just how powerful local struggle can be in es caping structural traps. So macroeconomic dynamics turn on the relations between the demand for labor, wages, the demand for commodities, and the supply and demand for capital – in an economy that is not evolving along an equilibrium growth path. In such a general economic model, the history of profitability depends on the re lations that have been assumed away in the equilibrium models: between accumulation and the rate of growth of the labor force; between wages and the demand for commodities; between investment policies and supply; and between capacity and demand.

In the short run, the rates of change of unit costs, net exports, output, and capital can take on unrelated values. So, the rate of profit can rise or fall: trade, the rates of growth of demand, supply, and capital stock, and the rates of technical and wage change all interact to produce an actual history of profitability.

In the longer run, if the rate of profit is rising and production is expanding, capacity constraints begin to bind. The critical constraints are that demand cannot exceed supply; the demand for labor cannot exceed its supply; and output cannot exceed the capacity to produce. These constraints mean that: either the rates of investment, growth of output, change in net exports, or change in unit costs of production cause the rate of profit to fall, or the rate of profit is rising and the economy is driven up against the constraints. In turn, the constraints mean that increases in the rate of profit depend on reducing unit costs of production or on increasing the level of net exports as a proportion of domestic demand.

But in the longest run, other nations do not permit an economy to continually increase its net exports, for one country's export success is another country's failure. In the extreme longest run, therefore, net exports are a constant proportion of domestic demand (either exactly or on average over a period of years). In that case, if the constraints are met as equalities (either exactly or on average over a period of years), the rate of profit is constant (or fluctuates about a constant mean) and is limited by the rate of growth of the labor force; if the constraints are met as inequalities, the rate of profit falls. The history of capitalism so far indicates that the constraints are met as equalities over the very long run (of hundreds of years), but that over shorter periods of 50 years or so the constraints are met as inequalities and profit rates fall.

Finance

There is good evidence that the profitability of production in capitalist economies fluctuates in this manner (see, for example, **Figure 2**). In the USA, there were two periods of relatively high profits, in the 1870s/early 1880s and in the period after World War II, and two periods of lower profit rates, between 1880 and the 1930s and after 1970. (The timing of such fluctuations is different in different countries.)

How does investment respond to such fluctuations in profitability? As the value of investment in production falls, so the relative benefits of financial speculation rise. Histories of Western capitalism suggest that periods of profitability and growth in the real economy are succeeded and separated by periods of dominance by financial capital and the emergence of a new leader in the international economy. So it was in the Italian city states in fourteenth to mid seventeenth centuries; in the Netherlands between late seventeenth and late eighteenth centuries; in Britain between late eighteenth and early twentieth centuries; and in the USA between late nineteenth century and the present day.

Finance capital is comprised of money drawn out of production. Firms disinvest; thus the low rate of profit is translated into low rates of growth of output (and falling

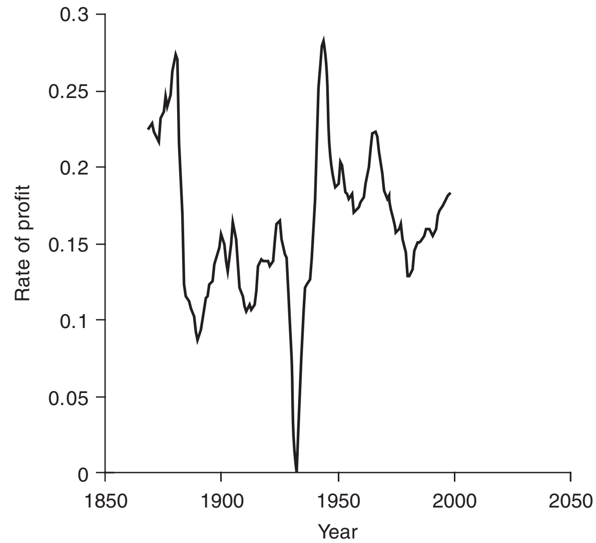

Figure 2 The rate of profit in the US nonresidential economy, 1870 1999. Date from Dumenil, G. and Levy, D. (2001). Periodizing capitalism: Technology, institutions and relations of production. In Albritton, R., Itoh, M., Westra, R. & Zuege, A. (eds.) *Phases of Capitalist Development: Booms, Crises and Globalizations*, pp 141 162. London: Palgrave.

working conditions). Rather than extracting surplus from workers, finance capital earns profits: (1) from other capitalists (by buying low, selling high) and (2) by reallocating itself into new spheres of production (new commodities and new locales for capitalist production). The post 1970 emergence of 'global capital' is in this interpretation merely the latest in a series of phases of alternation between dominance by commodity and finance capital.

To understand the financial economy, let us start by thinking about the way in which the value of money is set in trade. Suppose that there are two countries, A and B, that trade. As B exports to A, people in A must use their $A to buy $B in order to pay the producers of B for their exports. The demand for $B relative to $A rises, which tends to raise the relative value of $B compared to $A. A contrary motion occurs when A exports to B. If the balance of trade is such that A exports more than it imports, the demand for $A is greater than the demand for $B, and so the value of $A rises relative to the value of $B: A's exports tend to become more expensive. Two features of this story are important: first, the real movement of commodities drives the contrary movement of money and so the relative prices of currencies, and second, the movements are compensating – a trade imbalance calls forth currency changes that tend to remedy the trade imbalance.

But the financial flows (like those into East Asia) were not of this kind. Under conditions of global liquidity and easy transfers of money, the strength of the East Asian economies attracted capital. Those inflows did not compensate for the movement of commodities: the in flows of money were themselves the independent vari able. However, as money flows from B to A, three effects occur. First, a flow of money from B to A implies an increase in the demand for $A and so an increase in its price. Second, since much of this capital was used to buy fixed assets rather than productive facilities, the inflow stimulated rising asset prices. Third, the increased availability of money in relation to local production within A (and the lack of money in B relative to pro duction) tended to increase the trade deficit of A. This system is not self compensating: the revaluation of the $A and the inflation of asset prices both tend to encourage increased inflows of capital by raising their profitability.

The relationship between the flows of commodities, the flows of money, and changes in the relative values of money thus depends on which flow is driving the system. If the values of currencies are set mainly by trade, then a flow of money from B to A implies that the $A is re valued, the trade imbalance falls, and money flows di minish. If the values of currencies are set mainly by financial considerations and asset prices, then the re valuation encourages additional inflows of money and growing trade deficits. Increasingly, the latter consider ation became dominant in East Asia: capital inflows stimulate asset appreciation, trade deficits, and currency revaluations (or at least relative currency stability in the face of trade deficits). Conversely, in the case of Japan: capital outflows have stimulated asset depreciation, trade surpluses, and currency devaluations (or at least currency stability in the face of trade surpluses). In other words, the exchange rates are shifting in nonorthodox directions, dancing to the tune of capital flows rather than to trade flows. Under such a regime, capital inflows created the twin problems of East Asia: an asset bubble and in creasing competition within producers' export markets.

Conclusions

There may have been failures of governance, degener ation of the development state model in East Asia. But these conditions merely identified the target for the speculation (and perhaps exacerbated its effects). How ever, there would have been a financial crisis somewhere, just as there have been currency and banking crises since the 1980s in the developed as well as in the developing world. In other words, explanations of the long run or underlying causes of the crisis must focus not on conditions in East Asia but on the mode of operation of global financial markets. That in turn reflected the de clining profitability of production within the Western world since the 1960s and therefore the emergence of incentives for firms to reallocate money from production to finance. With increasing sums of money withdrawn from production in the West came both the crises of their manufacturing sectors in the 1970s and 1980s and the new prominence of international finance capital.

In the past, such episodes of international financial capital have ended with the recovery of profitable pro duction (as, for example, in the emergence of the US manufacturing in the late nineteenth century and the reconstruction after World War II). Perhaps the emer gence of China and India is signaling the start of a new spurt of production and the relative decline in the he gemony of finance? In the past, too, for reasons explored here, such spurts have also ended in falling profits and the search for new outlets for unproductive capital.

See also: Capitalism; East Asian Miracle; Finance, Historical Geographies of; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Regulation.

Further Reading

Braudel, F. (1981 1984). *Civilisation and Capitalism*, 3 vols. New York: Harper and Row.
Dumenil, G. and Levy, D. (2001). Periodizing capitalism: Technology, institutions and relations of production. In Albritton, R., Itoh, M., Westra, R. & Zuege, A. (eds.) *Phases of Capitalist Development: Booms, Crises and Globalizations*, pp 141 162. London: Palgrave.
Harvey, D. (1982). *The Limits to Capital*. Chicago, IL: Chicago University Press.
Lipietz, A. (1986). Behind the crisis: The exhaustion of a regime of accumulation. A 'regulation school' perspective on some French empirical work. *Review of Radical Political Economics* 18, 13 32.
Mandel, E. (1975). *Late Capitalism*. London: New Left Books.
Webber, M. J. and Rigby, D. L. (1999). Accumulation and the rate of profit: Regulating the macroeconomy. *Environment and Planning A* 31, 141 164.
Webber, M. (2001). Finance and the real economy: Theoretical implications of the financial crisis in Asia. *Geoforum* 32, 1 13.

Relevant Websites

http://www.marxists.org
Marxists Internet Archives.
http://www.rgemonitor.com
RGE Monitor.
http://www.worldbank.org
The World Bank.

Economic Development, Rural

D. Storey, University of Worcester, Worcester, UK

Rural Socioeconomic Change

Throughout most of the so called 'developed' world, agriculture is of declining importance in economic terms. What are commonly referred to as processes of mod ernization have contributed to an expansion in average farm size, increased mechanization, and concomitant declines in the numbers employed in agriculture. In the United Kingdom in the early 1950s, agriculture em ployed over one million people; this number has since halved. Amalgamation into larger farm units has seen the number of holdings decline from over 520 000 in 1939 to a current level of just over 200 000 while France's agri cultural population has more than halved since 1980. Simultaneously, there has been a reduction in the relative contribution agriculture makes to national economies. In Spain, agriculture now accounts for around 5% of em ployment and some 3% of GDP while in the USA and the UK less than 2% of employment is in agriculture. (It should of course be noted that agriculture continues to be of major significance in other parts of the world contributing well over half of GDP and employment in some countries of Asia and sub Saharan Africa.) Paral leling this decline has been a growing diversity of em ployment in many rural localities reflecting the growth of other economic activities in the countryside. Rural areas now have a more diverse economy with a growth of (nonagricultural) industry and, more particularly, ser vices. The footloose nature of some industries has meant they may find it advantageous to locate in rural areas in order to avoid some of the diseconomies that may be associated with congested urban centers. The growth of modern information and communications technology, most notably the internet, means an increasing number of companies and individuals can, in theory at least, work almost anywhere thereby (it is argued) rendering place unimportant.

The increasing role of the service sector globally is one significant element. Tourism in particular has been seen to be of major importance with rural places increasingly endeavoring to repackage themselves in order to tap into the tourist market. The growing shift away from what have been referred to as landscapes of production to landscapes of consumption and the apparent commodification of the countryside means we are in an era where rural places are increasingly endeavoring to reimagine themselves in order to deal with the broader processes of rural change and restructuring which they have experienced. This attempt to promote some form of local distinctiveness is part of a broader strategy of place promotion which is seen as necessary in a competitive visitor market. These changes indicate that rural areas have become more globalized and, rather than being linked primarily into the local area, they are now increasingly enmeshed in national and international linkages.

One obvious consequence of these structural shifts is the changing social composition of the countryside. There has been in migration into many rural areas from urban centers, a phenomenon generally referred to as counter urbanization. In England, for example, the population of rural areas increased substantially in the latter part of the twentieth century. While these changes are spatially and temporally uneven, the increasing economic and social heterogeneity of many rural areas means that rural and agriculture are no longer synonymous and this in turn has led many governments to recognize the need to devise rural development policies which embrace a broad spectrum of economic activities and more socially het erogeneous populations.

The traditional approach to development and plan ning centered on planners and development specialists – a professional elite – advocating specific developmental pathways for cities or for rural areas. Attempts to entice industry to locate in more peripheral rural locations such as the west of Ireland or the Scottish Highlands, through mechanisms such as preferential grant rates and the construction of 'advance' factories in rural areas, were among the elements in such policies. This 'top down' development philosophy was much criticized for a var iety of reasons. Chief among these was a concern that such an approach failed to consider the views of those most directly affected – local residents. Such criticisms have in recent decades led to the promotion of more 'bottom up' approaches centered on the involvement of local people and the empowerment of those people so that they actively participate in devising strategies for their own areas.

Integration, Participation, and Partnership

Changing economic and social relations have given added urgency to the need for more imaginative and integrated approaches to tackling the problems faced by rural areas. These approaches are characterized by an espousal of the importance of integration and participation. Integration can be interpreted in both a sectoral sense, whereby the various elements of the rural economy have a role to play in the development of rural places, and in a political and administrative sense, whereby various interested partners, organizations, and individuals are engaged within the development process. In this latter sense, the current development philosophy emphasizes the importance of participation and is underpinned by an espousal of the centrality of community involvement, empowerment, and capacity building. Four broad reasons might be suggested for the promotion of this strategy. First, the 'bottom up' approach could be said to reflect a democratization of the development process. Second, it has the potential to provide economic benefits to increasingly diverse rural areas where there is a need for more locally attuned ap proaches rather than generic blueprints. Third, there is perceived to be a need for local responses to those global processes which have tended to open up localities to outside competition. Finally, the involvement of local individuals and agents appears logical in that it utilizes local knowledge and skills and it may ensure more success given it enjoys a degree of local support. The approach is seen then as not just socially and politically desirable but also more effective and economically beneficial. All of these point away from cross spatial solutions imposed from outside and toward the need to be flexible in the approaches taken to rural development in order to take account of local circumstances. This more flexible ap proach could be seen to tap into the nature of people's identification with locality and community – what might broadly be seen as their sense of place.

This version of rural development is delivered through partnership arrangements bringing together interested groups and individuals from the statutory and voluntary sectors, including representatives of relevant agencies as well as local actors. This would appear to be a logical extension of a bottom up approach. In the European Union (EU), the Cork Declaration in 1996 gave more substance to a partnership approach, and integration, participation, and empowerment have now become buzzwords within the contemporary development dis course in much of Europe. The idea that developmental solutions to local problems can be dealt with through mechanisms that draw together representatives of local authorities, the private sector, and voluntary and com munity groups is now the accepted orthodoxy throughout EU countries.

The emphasis on partnership could be said to signify a shift from government to governance, a perspective which extends beyond the formal structures of central and local government to embrace the interactions be tween a range of actors at various levels in the spatial hierarchy and having varying degrees of influence. In general terms, the move to a more participative and partnership based approach might be seen in a positive light. On the surface at least, it appears to encourage decentralized decision making and recognizes the role of local people in the development process. This apparent local sensitivity might be seen as more desirable than externally imposed strategies. The encouragement of indigenous potential may also serve to break down the division between 'professional' and 'amateur', between internal 'resident' and external 'expert'. In broader terms, this shift might be seen as a reflection of the replacement of a modernist vision of planning with a more post modernist approach in which rural diversity and local differences are highlighted and celebrated. Underlying much of this appears to be an assumption that partner ship and participation are intrinsically good things to the extent that the supposed merits of the approach have become acts of faith not to be questioned.

Contemporary Examples

The EU's LEADER program provides a good example of this partnership approach to rural development. A fourth phase of this program is underway aimed at fostering economic development in rural areas utilizing a part nership approach and emphasizing innovation in the types of projects to be encouraged. LEADER has recognized that in addition to the need to demonstrate some form of sustainable economic output such as job creation, social, cultural, and environmental outcomes are of importance. LEADER has operated on the basis of providing part funding to local projects which could be seen as innova tive within that particular area. The program operates through a series of geographically based 'local action groups' consisting of representatives of the appropriate local authorities, other development agencies, and local community groups. In this way, the statutory, private, and voluntary sectors are brought together under one um brella. The types of project eligible have included craft type activities, local heritage initiatives, tourist related activities, support for the marketing of local produce, and promotion of local events. Another of LEADER's under lying aims has been the encouragement of local action. As a consequence, funding has also been used to encourage local animation and capacity building thereby em phasizing the community development aspect.

While agriculture may be seen to be of diminishing importance, recent years have seen a growing emphasis on local food production driven by a number of factors, including health concerns, perceived higher food quality, the desire to support local producers, resistance to supermarkets, and the reduction of 'food miles'. As a consequence, the promotion and branding of local and regional foods and drinks has become a noteworthy phenomenon. An example of successful regional branding is the Fuchsia brand in West Cork in Ireland under which much local food produce is marketed. Elsewhere in Europe, North America, and Australia an increasing emphasis on regional, niche and heritage foods is discernible. Products such as wine and olive oil are increasingly incorporated into tourism promotional strategies in southern France, Italy, and elsewhere in the Mediterranean. Alongside this there has been a resur gence of local farmers' markets in the UK and the USA. More broadly, these developments can be seen as symptomatic of increasing pluriactivity as farms diversify into other arenas leading some to characterize this as a shift from a productivist to a postproductivist era.

An extension of this 'theming' idea is to market a locality under a 'brand name' in order, among other things, to increase visitor numbers. The idea of place branding has a long history reflected in 'designated' landscapes such as the national parks of England, their extensive equivalents in the USA and Canada, or *parcs naturels regionaux* in France. However, its usage as a promotional tactic is evidenced through the use of lit erary figures with the likes of Thomas Hardy's Wessex and Catherine Cookson Country in the Northeast of England as examples of local 'literary branding'. In the United States many small settlements engage in local branding: the former silver mining town of Wallace, Idaho, styles itself 'the silver capital of the world'. Others associate themselves with a nearby attraction so the small town of Columbia Falls in Montana is the 'Gateway to Glacier National Park'. Others endeavor to find a niche for themselves, an example being the promotion of bookshops in Bécherel, in Brittany, in northern France. Besides being a *cite du livre*, Bécherel like other small towns in the region is designated *un petite cité de caractére*, a devise used to boost visitor numbers. Other examples of local level initiatives aimed at encouraging regeneration of rural localities include government sponsored at tempts to regenerate market towns in England (and similar attempts to rejuvenate small towns in Australia and the USA) and the promotion of walking, cycling, and driving routes in rural localities.

Much has been written by geographers and others on these changes in emphasis and two key strands appear to emerge. One of these is a focus on what might be seen as technical considerations surrounding how best for part nership working to occur and to be successful. Within this, attention has been paid to evaluating the effective ness of programs and the devising of toolkits to improve the 'output' of these. However, such apolitical assump tions elide important questions of power. Consequently, a more critical perspective has emerged directed at the underlying ethos of contemporary development prac tices. From this perspective, it is argued there is a need to consider the political dimensions of rural develop ment because (like any other form of development) it is not a politically neutral process as measures imple mented are likely to benefit some while simultaneously disadvantaging others. Within a 'Third World' context, the replacement of a 'top down' orthodoxy by another participative one to which all must subscribe is seen by some as merely a new form of tyranny rather than her alding a new and radical approach to development. In essence, what this suggests is that more attention needs to be paid to issues of power and the variable power re lations at play.

The Contested Nature of 'Community'

Contemporary development rhetoric highlights the im portance of community development. 'Community' is a term with potent meaning and one which tends to induce warm and comforting feelings. It is also something felt by many people to be inherent within rural areas and is associated with a more general lauding of a rural way of life as distinct from a supposedly more atomized urban lifestyle. By virtue of this, the promotion of a com munity based approach to economic development is something which is useful in its appeal to people's in built regard for community. However, community needs to be seen as a social construction rather than a concrete entity – a construction which can be both celebrated and abused. This is not to suggest that community does not exist but it is to warn of the implications of assuming it to be an essentialized and largely unproblematic entity.

In very broad terms, community might be defined either geographically or on the basis of shared interests. A problem arises when these place based and interest based interpretations are seen to be co terminus and where a commonality is assumed between all those resident in a geographically defined area. Territorially based partnership approaches tend to elevate residential location above social stratification and, as a consequence, various fault lines such as class, status, or gender are pushed to the background. While local residents them selves may be well aware of tensions which affect their understandings of what might be meant by community, the lauding of the local may result in the elision of ser ious differences which themselves may be linked to the problems manifest within a locality. While attention has been paid to devising mechanisms to draw out local potential and build the capacity among local actors to effect change, there is a risk of assuming that a consensus can be reached at a community level as to what path to follow. What this means in practice is that developmental initiatives endeavor to find common ground. While this may be possible on some issues, it is unlikely to be the case with regard to everything. As a consequence, agreement may well be reached on lowest common de nominator issues with more contentious matters being either ignored or downplayed. The sidelining of potentially divisive issues in favor of harmony may result in the implementation of a standard set of interventions. That which can be achieved will be pursued, while those things which might be desirable, but are more contentious, will be kept off the agenda.

Partnerships and Power

The arguments surrounding the meaning and nature of community give rise, in turn, to questions over the setting of the development agenda at a local level. It is not just in the tension between statutory and nonstatutory or between national, regional, and local that power relations are important. While participation may appear as an intrinsically good thing, this assumption ignores the fact that some actors are stronger than others. In other words, it plays down the ways in which power diffuses geographically. The assumption that power resides in specific sites, rather than being more diffuse, leads to the erroneous assumption that devolving to the local solves the problem. This assumes that power as a reified entity can be simply transferred. Furthermore, it ignores the reality of local power relations and the relative power (or powerlessness) of some individuals and groups. Some actors and groups may wield considerable power, others may be involved (or feel they are involved), while still others are excluded (or choose to exclude themselves). This means that some actors may be more influential than others. Partnership arrangements may not necessarily equate to widespread participation with some voices more likely to be heard while others are ignored or remain silent. Research in Scandinavia, Australia, North America, and elsewhere has raised questions about the ways in which, for example, women or young people may be marginalized within these structures. Discourse may be dominated by the 'usual suspects' who are regularly heard while a 'silent majority' sits outside the consultative structures.

This is not to deny that the use of a territorial framework may serve as a mechanism which induces a real sense of community and which may also facilitate practical outcomes. The imposition of a territorial frame on a 'community' may also serve to mobilize residents and give rise to a sense of local solidarity in the face of unwanted developments. However, it is generally held that those with the time and the energy are more likely to participate in community initiatives and organizations resulting in those less articulate and the more disadvantaged citizens being largely excluded. All this raises serious questions over the ways in which we measure community vibrancy and mobilization.

Although central and local government may appear to champion participative governance, the existence of informal networks means that some power resides in 'elite spaces' beyond direct democratic or community control suggesting a need to take account of the political and social relations present within a locality. Community development in its broadest form has become professionalized through the activities of individuals employed by government agencies or local authorities and those who represent community groups. The balance between statutory and nonstatutory partners may vary quite considerably and in some instances statutory agencies within local areas may continue to exert considerable control – a sort of top down imposition of a bottom up strategy. Furthermore, those who wield more power are in a better position to control, or at least influence, the development discourse thereby steering groups toward particular outcomes. In this way, it is meaningful to talk of discursive power whereby the terminology is utilized and controlled in particular ways which enhance the prospects of certain outcomes. This means that the practice of partnership working may be considerably at odds with the rhetoric.

To a considerable extent, this emphasis on community, partnership, and participation appears to be underpinned by the implicit idea that the local can be unambiguously defined and that somehow it is always best. While it may well be appropriate to be skeptical of external 'experts', it is dangerous to assume a misty eyed romanticism with regard to local groups. It does not necessarily follow that local solutions are somehow better. It can reasonably be argued that local knowledge, no more than expert knowledge, does not itself possess an intrinsic value. It is far from certain that a knee jerk lauding of the local, insofar as one distinct local can be identified, is in any way intrinsically 'better' than an externally imposed view. Local knowledge is not fixed and is shaped by various interests. The more romanticized justifications for local involvement tend to gloss over the possibility that some local perspectives may reflect insular and conservative values that may in turn translate into exclusionary practices directed against specific groups or individuals within a locality such as those not deemed to belong. In this way, an emphasis on place based communities runs the risk that it could legitimate forms of social exclusion whereby some people are made to feel very much 'out of place'. None of this is to impugn the motivations of local actors and those who play a role in rural development. Rather, it is to caution against simplistic assumptions concerning the value of local involvement.

Incorporation, Co-Option, and De-Radicalization

The argument to this point suggests there are internal contradictions within contemporary rural development philosophy. Taking a step back from this, it might be suggested there is a broader set of political processes operating here. Whether intentional or not, the partnership approach may involve incorporation and deradicalization through the co opting of local agents into the development structures. This runs the risk that those who might otherwise espouse more radical positions find themselves inside the structures of governance. Indeed, the importance attaching to the achievement of consensus is such that those who are seen as less cooperative might be branded as negative influences resulting in their becoming sidelined. One effect of this is the neutralizing of certain arguments and the evasion of conflict. It might be argued that the incorporation of voluntary groups and local activists may well serve to blunt criticism. Through co opting possible sources of dissent, more powerful partners, whether central government, local authorities, or other statutory bodies, may find their position enhanced. Less powerful partners may consent to be co opted by those who are more powerful in order not to be totally excluded. In effect, the discourse of participation, community, empowerment, and so forth disguises the various power relations in operation. Local people may have limited, if any, real control and the involvement of local groups may become a mechanism through which places are governed rather than reflecting a genuine shift in democratic structures.

The importance of variable arrangements enhances the possibility of 'local fit'. However, although local differences are emphasized, a partnership approach has nevertheless become a new orthodoxy and therefore one to be implemented in all contexts. The designing and operationalizing of toolkits along with the spread of things such as rural appraisal and parish surveys in Britain, while perhaps being of benefit, might also be seen as buying into a particular way of doing things which is ultimately disseminated downward rather than rising from below. It may also be the case that partnerships behave in a very functional sense, in that they operate so as to gain access to funding. Indeed their very existence in some cases may be a consequence of funding considerations. Rather than arising organically, many partnerships may come into being and may behave in particular ways for purely pragmatic or instrumental purposes. Further, the emphasis on territorially based partnerships means there is intense competition between places in the search for funding and an increasing need for local organizations to become adept at writing funding bids.

Conclusions

Undoubtedly there are many positive outcomes, tangible and intangible, associated with contemporary rural development initiatives. However, these may be difficult to determine for a wide range of reasons. While useful attempts at devising evaluative frameworks have been tried, an evaluation needs to do more than simply view these in terms of stated outcomes. The actual aims and objectives of various partners and agents may be somewhat different to those stated. An investigation of the appropriateness of current initiatives to address problems of marginalization and social exclusion needs to address fundamental issues of power. Are major shifts occurring or are we simply seeing old wine in new bottles? Is participation simply obscuring power relations and thereby ensuring a continuance of deep seated structural problems? Rather than witnessing a radically new developmental ethos some contemporary actions may be strong on 'spectacle' but weak on substance.

In some instances there appears to be a simplistic assumption that 'doing something' is sufficient in and of itself or that all actions are inherently good and devoid of self interest. The various power relations operating across spatial scales all need to be explored rather than simplistically assumed to operate in certain ways. While the 'production' of strong rural communities may be worthwhile in some respects, there is no guarantee that radical alterations in underlying structures are going to flow from this.

See also: Governance; Rural Communities; Rural Populations; Tourism, Rural.

Further Reading

Cooke, B. and Kothari, U. (eds.) (2001). *Participation: The New Tyranny?* London: Zed Books.

Commission of the European Communities (1996). *Cork Declaration A Living Countryside*. Brussels: Commission of the European Communities.

Edwards, B., Goodwin, M., Pemberton, S. and Woods, M. (2000). *Partnership Working in Rural Regeneration*. Bristol: Policy Press.

Edwards, B. and Woods, M. (2004). Mobilizing the local: Community, participation and governance. In Holloway, L. & Kneafsey, M. (eds.) *Geographies of Rural Cultures and Societies*, pp 173 196. Ashgate: Aldershot.

Goodwin, M. (1998). The governance of rural areas: Some emerging research issues and agendas. *Journal of Rural Studies* 14(1), 5 12.

Liepins, R. (2000). New energies for an old idea: Reworking approaches to 'community' in contemporary rural studies. *Journal of Rural Studies* 16(1), 23 36.

Moseley, M. (2002). *Rural Development. Principles and Practice*. London: Sage.

Murdoch, J. and Abram, S. (1998). Defining the limits of community governance. *Journal of Rural Studies* 14(1), 41 50.

Osborne, S. P., Williamson, A. and Beattie, R. (2004). Community involvement in rural regeneration partnerships: Exploring the rural dimension. *Local Government Studies* 30(2), 156 181.

Storey, D. (1999). Issues of integration, participation and empowerment: The case of LEADER in the Republic of Ireland. *Journal of Rural Studies* 15(3), 307 315.

Westholm, E., Moseley, M. and Stenlås, N. (eds.) (1999). *Local Partnerships and Rural Development in Europe. A Literature Review of Practice and Theory*. Falun: Dalarnas Forskningsråd.

Woods, M. (2005). *Contesting Rurality. Politics in the British Countryside*. Aldershot: Ashgate.

Economic Geography

T. J. Barnes, University of British Columbia, Vancouver, BC, Canada

Glossary

Commodity Chain Analysis A technique developed during the 1970s concerned with representing the entire trajectory of the life of a commodity and its associated geographies: from conception and design to consumption.

De-Industrialization A process beginning in many Western high-income, industrialized nations during the 1970s and involving the large-scale decline of manufacturing employment especially in old industrial regions as manufacturing firms moved offshore, went bankrupt, or were radically restructured.

Fordism/Post-Fordism Fordism was a particular form of industrial production originating in the early twentieth century, and characterized by the mass production of a standardized good, assembly-line techniques, use of dedicated machines, a large, de-skilled, and often unionized (male) labor force, and a centralized corporate industrial structure. Post-Fordism arose during the late 1970s as a replacement for Fordism, and characterized by batch production of specialized commodities, the use of flexible (computer driven) machines, flexible skilled labor, and a decentralized industrial organization.

Globalization The increasingly interconnected relation among places in the world brought about by cheapening transportation and improved communications, and reflected in increasing international investment, trade, and the movement of people and ideas.

Industrial District First discussed by the English economist Alfred Marshall in the early twentieth century, an industrial district is a tight-knit interlinkage of many highly specialized firms concerned with collectively producing a single product type such as clothing or furniture.

Location Theory A body of interlinked analysis that began in the early nineteenth century with writings by Johannes von Thünen and continues to the present, and concerned with the logically and empirically rigorous explanation of the spatial arrangements of economic phenomena and related flow patterns.

Multinational Corporation A large firm that owns facilities or carries out operations in at least two different countries.

Neoliberalism An idea developed during the inter-war period of the twentieth century averring that the most efficient form for organizing economic activities is the free market, and if markets do not exist they require creation even under the auspices of the state.

Regional Science A hybrid discipline originating in 1954 as a result of the energetic efforts of its founder Walter Isard, and concerned with twinning formal economic theory and quantitative methods to analyze spatial issues in economics, geography, and planning.

Spatial Divisions of Labor An idea popularized by Doreen Massey in a book of the same name and concerned with the pattern and consequence of local geographical economic specialization.

Introduction

Economic geography is a subfield of human geography concerned with describing and explaining the varied places and spaces in which economic activities are carried out and circulate. The discipline was institutionalized in the late nineteenth century in both Western Europe and the US, and remains one of the core subfields within Anglo American geography. There have been repeated attempts to forge links between economic geography and its seeming intellectual soul mate, economics, but none have held. Economic geography from the beginning was more empirically grounded, concerned with context, and conceptually open minded, and at the same time less abstract and formally theoretical than economics. There were moments when the two disciplines became close, but mostly they held themselves apart at a distance (as they generally do now). Further, unlike economists, economic geographers never settled on a single methodology, set of techniques, list of venerated luminaries, disciplinary problematic, or definitive definition.

Change has been incessant, the field continually reinventing itself. It makes for an exciting, dynamic, open subject, one that never looks back, and a frequent conduit for new ideas into the rest of human geography, but there is a nagging sense that before old promises are realized new ones are made. Economic geography sometimes seems like Penelope's shroud: spun during the day, and unraveled the same night.

Perhaps one of the reasons that economic geography has been subject to so much change is because of how closely the discipline is tethered to its empirical object of inquiry, and which over the subject's lifetime has undergone profound alteration, with concomitant shifts in the discipline's analysis of space, place, and geo graphical circulation. This is perhaps no better illustrated than at present. The end of the late twentieth and early twenty first centuries are witness to periods of rapid economic geographical change, and propelled by a combination of technological revolutions around trans portation and communications, and a social revolution around the introduction of the market in places that had either not been exposed to it (such as in communist China and the Soviet Union), or experienced constrained versions of it (such as in India). The consequence has been speedy, and in some cases monumental transfor mation on the ground. Contemporary economic geog raphers are trying to represent and analyze the emerging geographical constellation, emphasizing the centrality of flows and spatial interconnection, as well as the con tinuing importance of place. As they do so, the discipline itself changes. Economic geography lives in exciting times, but it sometimes leaves the discipline breathless.

This article is divided into two main sections. The first presents the main phases of the discipline's historical development from the late nineteenth century when it was institutionalized within universities in Western Europe and North America to its present incarnation. Five main periods and associated approaches are recog nized: commercial geography (late nineteenth/early twentieth century); regional economic geography (the inter war period); spatial science (from the mid 1950s and continuing in attenuated form to the present); Marxism and radical economic geography (from the mid 1970s to the present); and the 'cultural turn' (from the mid 1990s to the present). The second section outlines the contemporary character of the discipline by re viewing some of the primary clusters of research and writing within the field. Eight areas are discussed: theory and methods; globalization and neoliberalism; firms, in dustry, agglomerations, and networks; innovation and high tech; labor, bodies, and work; retailing and con sumption; producer services and finance; and nature and resources.

History and Approaches to Economic Geography

Commercial Geography

Existing in embryonic form as commercial geography, economic geography was formally defined as a discipline in 1882 by the German geographer Götz. Commercial geography had been less an academic discipline than a branch of the European imperial project. Commercial geography's charge was providing practical geographical knowledge to the military, the business class, and the colonial bureaucracy, including information about where places were located, the kinds of goods they produced, and the networks of available transportation. In contrast, under Götz's conception the purpose of economic geography was to fulfill a scientific and not a geopolitical end. It was to address the causes and reasons of economic geographical specialization, and determining the pro duction of commodities and the movement of goods.

By the 1890s, economic geography courses were ap pearing in US university calendars. In 1903, Lionel W Lyde was appointed professor of Economic Geography at University College London, and in 1908, the University of Edinburgh created a lectureship for the economic geog rapher, George Chisholm, author of the earliest English language economic geography text, *Handbook of Commercial Geography*. Economic geography was up and running.

Chisholm's case is an interesting one. Born in Scotland in 1850, he later worked in London making a living from writing and editing geographical textbooks, gazetteers, and atlases (he previously worked for the Edinburgh publisher W. G. Blaikie & Son on such projects as the *Imperial Dictionary*). From 1896, Chisholm supplemented his income by lecturing on commercial geography at the University of London's Birkbeck College using the *Handbook* as his textbook. In 1908, he came home, ap pointed as lecturer at the newly formed Department of Geography, University of Edinburgh.

Chisholm's *Handbook* was written mainly in the com mercial geography style, jammed with maps, tables, and economic geographical facts. It was a book designed to give especially the English business class a practical education, to make them more competitive against Western European rivals, particularly the Germans. But Chisholm also wanted to practice an economic as op posed to a commercial geography by scientifically ex plaining rather than merely describing. He did so by positing 'nature' as the central causative agent, arguing that the environment made each place uniquely fitted to undertaking a particular type of economic activity.

This position became more extreme, one with clear racist overtones, in the hands of a number of environ mental determinists later working in economic geog raphy at the beginning of the twentieth century. The Harvard geographer Ellsworth Huntington was the most notorious arguing that the level of economic develop ment in a region was a consequence of the climatic re gime that held. Climate, he contended, determined physical and mental efficiency, which in turn established the level of economic development. In Huntington's famous phrase, climate was the "mainspring of civil ization." Temperate climates found in Western and Northern Europe and parts of North America were conducive to high efficiency. While tropical and sub tropical climates found in Africa, large parts of Asia, and Central and Southern America produced low energy, hampering economic development, ensuring such places remained 'backward'.

Regional Economic Geography

The intellectual backlash against environmental determinism, as well as a changed global economic context in which imperialism and empire were no longer drivers, generated a different economic geography during the inter war period, a regional one. The focus was not as it was in Chisholm's work on global commodity production, trade, and transportation, but rather on local economic interconnections that produced unique and singular regions.

As a form of enquiry the regional approach was readily seen in the various economic geographical textbooks published in North America from the mid 1920s onward. It involved characterizing regions by a common typological scheme, one, for example, broken down by leading industries, natural resources, modes of transportation, and so on. Once all regions were so described, their differences, and thus their individual uniqueness, were immediately evident by reading across the typological grid.

For example, Vernor Finch's and Ray Whitbeck's *Economic Geography* is typical. The diverse body of detailed economic geographical facts they provide are ordered under an identical fourfold typology for each of the regions investigated: agriculture; minerals; manufacture; and commercial trade, transportation, and communications. Or, another example is Clarence Jones' textbook *Economic Geography* that deploys an eightfold typology. While more finely variegated than Whitbeck's and Finch's, Jones' typology performs the same role: classifying observations that are mapped, tabulated, photographed, and listed under the appropriate classificatory heading. By typologically comparing the facts of different regions, economic geographical differences are immediately seen, and regional uniqueness shines by its own light.

It was not until just before World War II that an intellectual justification was provided from the American geographer Richard Hartshorne. Arguing against any separation between economic and regional geography, Hartshorne asserted that unique regions were the natural units in which the world expressed itself. That uniqueness stymied any hope of producing scientific explanation and generalization in economic geography, however. Science was successful, he thought, because it was able to identify common properties in the phenomena it studied, enabling it to generalize, and at the limit to posit explanatory laws. Economic geographical regions, however, were not the same kind of natural phenomena because each region was a collection of exceptional features, describable only in its own terms. No generalization was possible. The consequence was, as Hartshorne wrote in *The Nature of Geography* that set out his larger thesis: "Regional geography, we conclude, is literally what its title expresses. … [I]t is essentially a descriptive science concerned with the description and interpretation of unique cases.…" (p. 44).

Spatial Science

Hartshorne's book was published at exactly the wrong time. The regional descriptivism it proposed for geography, and economic geography in particular, quickly became out of step with a set of larger changes that pointed in exactly the opposite direction. In particular, from the beginning of World War II, a number of social sciences, and even some humanities, were transformed from a descriptive approach of the kind championed by Hartshorne to one that instead emphasized scientific generalization and explanation. In part because of Hartshorne's influence, economic geography initially resisted that impulse, but by the mid 1950s it joined in too. The resulting shift to spatial science, and represented by the 'quantitative revolution', profoundly altered economic geography, sweeping away talk of regional uniqueness in favor of the systematic application of scientific forms of general theorizing and rigorous statistical techniques of description and analysis.

Spatial science within economic geography was defined by five main features. First, there was the use of formal theory and mathematical models, many of which were begged, borrowed, and stolen from neoclassical economics (in particular, from rational choice theory, general and partial equilibrium, and German location theory), and, a more unlikely source, physics (spatial interaction theory and later entropy maximization models). Second, there was the use of an increasingly sophisticated arsenal of quantitative methods. Initially used were off the peg inferential statistical techniques, but later specialized statistical measures and methods were designed in house to meet the peculiar features of economic geographical data (e.g., techniques of spatial autocorrelation). Third, there was the use of computers. At first they were very crude and limited, but within a decade they performed hitherto unimaginable calculations, for example, the inversion of large urban and regional economic input–output matrices that would otherwise have taken a lifetime to calculate by hand. Fourth, there was a philosophical justification made for spatial science based on positivism, the idea that only scientific knowledge is authentic knowledge. Fred K. Schaefer provided an early influential justification based on logical positivism, but it was broadened and deepened by David Harvey in his *Explanation in geography*. Finally, there was a focus on abstract spatialities and geometrically defined location. Regions remained part of the economic geographical lexicon, but conceived utterly differently than under the previous regional geographical regime. Regions were now explanatory, theoretical, and instrumental, a spatial unit to achieve functional objectives (and brilliantly realized in the parallel movement of regional science). The consequence was that people like Finch, Whitbeck, and Fielden Jones (and their respective books) were no longer recognizable as part of the discipline. They were not in the same field, not on the same planet.

Spatial science like most intellectual revolutions began at only a few sites, later diffusing more widely. In Europe it was associated with Cambridge University, Bristol University, and Lund University, and within North America, the Universities of Iowa, Washington, Chicago, and Toronto, as well as Ohio State. In each of these places, groups of young, bright, ambitious, com petitive, and almost exclusively male students gathered to participate in the revolution.

The University of Washington at Seattle provides a good case study. The graduate students who gathered there from 1955, and given the moniker the 'space cadets', included Brian Berry who was to make a series of seminal contributions to spatial science. Berry and the others were drawn by two faculty members interested in establishing a more theoretical and quantitative economic geography: Edward Ullman and William Garrison. Gar rison was especially important acting as a mentor, pro tector, critic, muse, and paymaster. In 1955, Garrison offered the first advanced course in statistics ever in a US geography department, and the same year he gave a course on spatial economic theory using Walter Isard's regional science textbook, *Location and space economy* (1955). Such training held the 'cadets' in good stead for their next project, and funded through Garrison, which was to evaluate a proposal for federal highway develop ment around Seattle. It was the perfect assignment, allowing the students to hone their analytical and theo retical skills, to bootstrap learn from other disciplines such as engineering and economics which were also re cruited, and to display their newfound knowledge of abstract theorizing, quantitative methods, and computer programming. The project produced a revolutionary book, *Studies of Highway Development and Geographic Change*. Packed with calculations, data matrices, statistical tech niques, cost curves, and demand schedules, and con ventional maps overlaid with numbers, arrows, starburst lines, and balancing equations, it was a volume like none other previously published in the name of economic geography. But it became the norm as spatial science took over the discipline in the 1960s.

Marxism, Political Economy, and Radical Economic Geography

The first signs of trouble for spatial science appeared in the early 1970s when David Harvey disavowed his earlier support, declaring that the quantitative revolution had run its course, telling us less and less of any import. Harvey was the first of a series of high profile defections. In retrospect, the problem was the abstract, closed, and narrow conception of economic geography that spatial science proffered. It was not true to economic geography's own variegated disciplinary history; not true to the his torical moment of the early 1970s that was increasingly politicized and drawn to issues of social relevance; and not true to its own scientific logic as assorted logical contra dictions, inconsistencies, and aporias revealed.

An alternative was political economy, colored espe cially by Marxism. Certainly, this was Harvey's choice after his spatial science renunciation. Subsequently, he provided a set of what have proved to be valuable and enduring concepts, for example, 'space–time com pression', drawn from Marx for understanding capital ism's geography of accumulation (best found in Harvey's magisterial *The Limits to Capital*).

In Britain, political economy surfaced in the late 1970s, emerging within industrial geography. Because of de industrialization Britain was hemorrhaging manu facturing jobs (and occurring also in other Western countries). Political economy with its focus on economic crisis and instability offered an ideal explanatory frame. Moreover, it was usefully supplemented by a methodo logical approach, critical realism, developed also in the late 1970s by the economic geographer Andew Sayer that linked political economy with specific on the ground strategies for empirical research and explanation. To explain economic geographical events, argued Sayer, re quired an intensive research strategy to separate out necessary structures that bore on their occurrence from merely contingent ones that happened to be present. Doreen Massey and Richard Meegan relied explicitly on political economy and critical realism in their important dissection of the *Anatomy of Job Loss*. And 2 years later, Massey brought everything together – political economy, critical realism, and the larger context of de industrial ization and industrial restructuring – in her watershed book, *Spatial Divisions of Labour*. Of the volume's many accomplishments, reestablishing place and region on the economic geographical agenda was central. This wasn't back to Hartshorne, however. Place and region were to be understood theoretically, as an evolving recursive re lation between rounds of capitalist accumulation and the socially constructed character of geography. Massey's work sparked the UK locality project, a large multisite, multiyear venture charged with explaining the changing fortunes of British regions.

While British economic geographers during the 1980s slogged in the grim trenches of industrial decline, a number of American economic geographers instead idled in the brilliance of industrial regeneration and growth. This was the flip side of capitalism: its innovativeness, creativity, and powers of recuperation. Drawing on a combination of political economy and institutional eco nomics, the 'California school', including AnnaLee

Saxenian Allen Scott, Edward Soja, Michael Storper, and Richard Walker, carried out during the 1980s a series of impressive theoretical and empirical studies based mainly in the Los Angeles and San Francisco Bay regions on high tech, the motion picture industry, and garment manufacture. They showed industrial capitalism was capable of rewriting its economic geography, producing new industrial spaces, and requiring a new theoretical lexicon of representation and explanation.

While initially these British and American versions of political economy were like ships passing in the night, they came together in the 1990s through regulation theory that later provided the template for economic geographical work during the rest of that decade. Initially developed by the French economists Allain Lipietz and Robert Boyer, regulation theory was an attempt to explain why capitalism survived in spite of Marx's best prediction of its demise. The regulationists argued that capitalism endured because historically an appropriate conjunction formed between the regime of accumulation, defined as the macroeconomic relationship between consumption and investment, and the mode of regulation, defined as a set of institutional rules, norms, and laws. For example, the post war Fordist regime of accumulation defined by mass production and mass consumption was sustained because of the attendant mode of regulation, the Keynesian welfare state, which promoted private investment and consumer spending necessary for its maintenance. Moreover, regulationism explained how British and American economic geographical experiences were so different. While British economic geographers were documenting the disintegration of an older Fordist regime and mode, the California school was looking at the formation of a brand new one labeled variously post or neo Fordism, or flexible production. They were different sides of the same capitalist Janus face.

The 'Cultural Turn'

By the 1990s another approach was emerging, the 'cultural turn'. In part it reflected a larger move across English language social sciences and humanities in taking culture more seriously (and linked to the rise of cultural studies and interest in post structural theory). But it also tapped existing disciplinary tendencies and inclinations especially since Massey's *Spatial Divisions* that invoked a set of cultural understandings for explaining economic geographical events, including gender, patriarchy, religion, and cultural politics. It took almost a decade more, but by the mid 1990s the 'cultural turn' was turning.

The agenda continues to be developed, but two main effects are apparent: rethinking the discipline's object of inquiry, theory, and methods, and using case studies of particular substantive topics to work through a cultural approach. Proponents aver that the emphasis on culture is not mere intellectual fashion, but reflects fundamental changes in capitalism as it moves to a 'soft', 'reflexive', or 'symbolic' form where the line between culture and economy is not just hard to see but is no longer there.

The traditional conception of economic geography presumed: (1) a clear object of study, the economy; (2) rationalist forms of theorizing; and (3) empiricist methods. The 'cultural turn' took aim at them all, finding each wanting.

(1) J. K. Gibson Graham in *The End of Capitalism* (*As We Knew It*) attacked the idea of a monolithic, purified economy. That idea possessed traction, Gibson Graham argued, only because of the grip of Western dualist metaphysical thinking. Once shed, however, the economy will be seen as it is: hybridized, heterogeneous, multiple, nonmarket, bleeding into other spheres such as the cultural, the social, and even the environmental. Consequently, it is not sufficient to do only economic geography, but one must engage other kinds of geography too.

(2) Rationalist, top down theorizing, the second characteristic, is discredited by post structuralism, and implying one must do theory in different ways as well. Following post structuralism, new theorizing should be reflexive, open ended, and catholic in its sources. Theory should be conceived as a vocabulary to achieve new ends, rather than mirroring the object of investigation. As a result, what counts as theory dramatically expands. It might include established frameworks such as globalization modeled as a process of spatial diffusion, but it could also include globalization modeled as rape (as in Gibson Graham's *The End of Capitalism* (*As We Knew It*)).

(3) Because of the increasing influence of feminist scholarship associated with the 'cultural turn', there is also a need for a dramatic revamping of methods. The traditional economic geography empiricist mindset of collecting numbers and the *ad verbatim* recording of 'expert interviews' no longer cut it. They should be replaced by methodological self consciousness and promiscuity that dramatically widens the definition of appropriate research data and strategies for collection.

The 'cultural turn' was not only conceptual, but a reworking of disciplinary content. In particular, the mid 1990s saw publication of a set of economic geographical monographs demonstrating the importance of taking culture seriously for particular substantive topics. Susan Hansen and Geraldine Pratt in their study of female labor markets in Worcester, MA, *Space, Place and Gender*, showed how gender mattered fundamentally in the supply and demand for jobs. Linda McDowell in *Capital Culture* focused on gender and sexuality in her case study of the practices of merchant bankers in the City of

London. Erica Schoenberger's concern in *The Cultural Crisis of the Firm* was the culture of senior managers in US multinational corporations who sometimes produced stunningly bad decisions, taking companies to the precipice of bankruptcy, and sometimes over it. And Andrew Leyshon and Nigel Thrift in *Space/Money* examined the culture of money within international financial centers such as London, including not only its making but its spending. The larger point was that economic geography required broadening: that economic geography was always more than economic geography. Economic geography should recognize the multiplicities of connection, its joints, hinges, and folds with the noneconomic, its "trailing ands," as William James once put it. This is what the 'cultural turn' was a turn toward.

Summary

Perhaps the most striking feature of the history of economic geography is the lack of overall disciplinary progress. Later approaches rarely take up and refine ideas contained in earlier ones. Rather, much more likely is caustic criticism and rejection. Furthermore, the history is not the classic Kuhnian one of paradigm shift either, with new paradigms developed to explain past anomalies. This is because there is no agreement on what constitutes anomalies, or even whether they exist. Instead of new paradigms emerging and eradicating old ones as envisaged by Kuhn, economic geography is more like a palimpsest, with past versions of the discipline still partially visible, not completely erased, and continuing to contribute to the discipline's present form. This makes for a messy discipline, but clear from its recent history, a lively one, as different approaches continue to jostle and rub against one another. Since the 1980s the discipline has been marked by vibrancy, experimentation, fecundity, and rude health. In particular, its vibrancy is especially evident in eight areas of conceptual discussion and substantive work: theory and methods; globalization and neoliberalism; firms, industry, agglomerations, and networks; innovation and high tech; labor, bodies, and work; retailing and consumption; producer services and finance; and nature and resources. These eight areas are not a definitive list, but exemplars showing contemporary economic geography's range, vitality, creativity, and relevance.

Contemporary Economic Geography

Method and Theory

Perhaps no discussions within economic geography have been more animated than those over its theories and methods. The theoretical controversies have been especially potent, but even issues of methodology, which in the past were subject to a "don't ask, don't tell policy," triggered sometimes heated dispute.

Economic geography continues to be an empiricist discipline in the sense that it is concerned with representing an empirical object. But what counts as empirical information, and the methods used to collect, assemble, and interpret data, have significantly shifted. In the past the central method turned on the assiduous collection of numbers and statistics, and taken to its extreme under spatial science and the quantitative revolution. But since the 1980s a 'qualitative revolution' has occurred, that is, a move to methods concerned with gathering and analyzing non numerical data. It began with 'intensive case study research' pioneered by proponents of critical realism in the early 1980s. But since then qualitative methods have multiplied and include now in depth interviews, focus groups, oral histories, ethnographies, participant observation, discourse and textual analysis, actor network theory, action research, and more beside. In this new methodological environment, nothing is proscribed, everything is permitted. The upside is diversity and rapid change: most of the methodological techniques listed would have been viewed as beyond the pale, or at best, suspiciously *avant garde* when first introduced. The downside is the derogation of quantitative analysis and loss of associated skills. Because of lack of training, economic geographers are now increasingly unable to undertake formal statistical and numerical analysis. As a result, critics complain that the turn to qualitative methods in economic geography has induced slapdash and superficial research. The response is that one needs methods appropriate to the necessary data, and if the necessary data is now qualitative, and reflecting what is important for understanding the object of inquiry, these methods must be employed.

Along with permissive methodology has gone permissive theory. As in the case of methods, now almost anything goes. Economic geography is polycentric, consisting of a set of dispersed theoretical communities. At least five are recognizable: an older mathematical modeling tradition drawn from spatial science, which is sometimes refurbished for the new economic times and spaces, but often is not; a variety of political economic theorizing, loosely grouped around writings of Marx, and frequently focused on theorizing the state or its absence; various stripes of feminist theory that sometimes intersect with political economy, but more likely linked with post structuralism, and concerned with bodies and subjectivities of men and women at work; a range of institutional approaches that draw upon theories developed especially in economic anthropology and sociology (Karl Polanyi's idea of 'embeddedness' and Mark Granovetter's notion of 'social network' have been especially influential); and selective theories drawn from science studies (particularly writings of Bruno Latour and Donna Haraway) and used to understand the brute materialities of economic geographical activities from CAD–CAM technology to the lurching movement of primary re sources along the commodity chain.

As with the new methodological practices, loose uses of theory ('reading around') has been liberating. There is no single authority to kowtow. One simply draws upon the theoretical tradition most appropriate to the task at hand. Critics complain that the result is anarchy, eclec ticism in which nothing fits, producing flighty, sloppy, and sometimes incomprehensible works. The overall result is Balkanization, a discipline of solipsisms and solitudes. The counter response is that fragmented theorizing is necessary to understand the increasingly fragmented geographical economy in which we live and study.

Globalization and Neoliberalism

Both permissive theory and methods are used to under stand an issue increasingly preoccupying economic geography, globalization, and forming a body of research as energetic as its focus of study. Increasingly linked with the study of globalization is work on neoliberalism, the larger political ideology and project averring the im portance of free markets. Globalization and neoliberalism are not necessarily connected, but over the last decade the two have often been twinned.

In many ways, globalization was made for economic geographical examination. And as already discussed, in the beginning it was economic geography. He didn't use the term, but George Chisholm was a student of glob alization. That focus was lost, however, as economic geography later drew inward and investigated only Western industrial economies. The rest of the world was parceled up and given over either to study by regional specialties, for example, Asian or Latin American geog raphy, or to the new subdiscipline of development geography. Processes beginning from the 1970s, however, such as the emergence of a new international division of labor, the increasing number, size, and dominance of multinational corporations, the growth of international financial capital, and new forms of communication and long distance transportation, showed economic geog raphers that they needed to deal with the world as a whole. Further, it became clear that separating economic geographical study of the West from the rest of the world was not just unnecessary but obstructive. Under rapid globalization, the linkages, the 'networks', that connected places across the world were critical. Dividing the globe into separate spheres of study was precisely the wrong strategy, obscuring what was most important. Peter Dicken in his book *Global Shift* was the first contemporary economic geographer to make this point through his study of international investment, and particularly car ried out by large multinational corporations. But Dicken made another point just as important. That globalization was no seamless process, eradicating geographical difference. Globalization represented not the "the end of geography," as economists sometimes said, but the latest form of its continuing importance. Geographical differ entiation was crucial, the very precondition for globali zation's possibility and achievement.

Subsequent work by economic geographers over the last decade demonstrates unassailably that geography matters fundamentally in the processes of globalization. Globalization is not a thin film spreading equally over the globe, but is uneven; it clumps and is spiky, occurring in very particular places. Geographical differentiation makes a difference. It determines where in the globe multinational corporations invest, where different links of the global commodity chains come down to Earth, where international financial transactions occur, and how firms are managed and the spatial interactions they forge. Globalization is not smooth, homogenous, flattening geographical difference as it goes, but utterly dependent upon spatial variegation, disjuncture, and distinction.

Hand in hand with this discussion of globalization has been a complementary critical examination of neoliber alism. It is argued that globalization requires an appro priate regulatory framework. The economy, even the global economy, cannot exist shorn of an apposite gov ernance structure, and which from the late twentieth century has taken the form of neoliberalism, and facili tated by global institutions such as the International Monetary Fund (IMF) and World Bank. By neoliberalism is meant the political ideological project of promoting free markets as the optimal means for undertaking eco nomic activity. While originating as an idea in the 1920s, it came to the fore during the 1980s in the politics and domestic policies of the Margaret Thatcher government in the UK, and the Ronald Reagan administration in the US. Later it was 'exported' to developing countries through the Washington Consensus. Its injunction is that places in the globe where free markets either do not exist, or are insufficiently free, must be transformed to meet neoliberal strictures, subject to policies such as trade and financial market liberalization, deregulation, privatiza tion, and the securing of private property rights.

Whether emerging internally or externally foisted, neoliberalism has increasingly taken over the world, and found not only in high income countries like the UK and the US, but also in the former Soviet Union, China, India, much of SE Asia, and in large parts of Central and Latin America as well as sub Saharan Africa (as a result of the World Bank's neoliberal policies of structural adjust ment). Through neoliberalism's aggressive stance in favor of open trade, investment, and financial capital move ment, the wheels of globalization are greased, making it happen more smoothly, quickly, and easily. This is clear from a growing and diverse set of economic geographical studies examining such topics as international finance, the internationalization of education, the international migration of skilled labor, and international trade in primary resources such as for diamonds, lumber, and water.

Firms, Industry, Agglomerations, and Networks

Studying the economic geography of firms has always been a bulwark of the discipline. In 1910, even Chisholm turned away (briefly) from his global maps of primary commodities to use the work of the German location theorist Alfred Weber to examine the locational factors bearing on 'seats of industry'. Subsequently, each new generation of economic geographers has provided its own interpretation of the firm, and the factors bearing on its geographical behavior. In the 1930s, Michael Wise focused on small firms, and their propensity to cluster geographically in what the economist Alfred Marshall earlier called "industrial districts." The 1960s saw the emergence of 'the geography of enterprise', later morphing into 'corporate geography', and concerned with multisite locational decisions made by very large firms. The 1980s and 1990s saw a return in part to the concerns of Wise, an interest in the tendency of firms to cluster or embed themselves geographically in agglomerations, albeit agglomerations that were themselves interlinked and sometimes half a world apart. And most recently, there is an interest in how firms of different sizes and types cohere together in the form of a network, held in place as much by social as economic relations, and sometimes extending over the entire face of the globe.

The focus on firms in economic geography reflects a deep seated belief that they are principal agents shaping the economic geographical landscape. Understand them, and you understand economic geography. For this reason, industrial geography, the subdiscipline of economic geography principally concerned with firms, and manufacturing firms in particular, has always been important. Given mounting globalization, and industrial restructuring turning on the shift from Fordism to post Fordism, the subdiscipline became central during the 1980s, and defined by several elements.

At first, especially British researchers concentrated on understanding the process of industrial decline particularly in old industrial regions as firms closed down, or moved offshore, or restructured their production operations, with associated large losses of manufacturing employment (deindustrialization). But even in the UK there was recognition that it wasn't only a story of job loss. A new industrial landscape was emerging, triggered by innovative firms. Massey, for example, documented a new spatial division of labor in South Wales created by American and Japanese multinational electronic firms exploiting a spatially trapped, greenfield female labor force.

In the US, Allen Scott focused less on decline than the potential for growth, constructing a new theory of the dynamic firm based on transaction costs, external economies, and spatial linkages, and leading him to rediscover Michael Wise's earlier work, and before him, Alfred Marshall's idea of an "industrial district." Defined as a tight knit agglomeration of small firms with a high degree of specialization and production interconnection, the industrial district became a condensation point for an enormous amount of economic geographical research during the 1980s and 1990s. Interest was given further momentum by Michael Piore and Charles Sabel's influential book, *The Second Industrial Divide*. Piore and Sabel were political scientists but their book was important for economic geography because they made industrial districts central to their thesis of an abrupt shift from mass production (Fordism) to flexible production (post Fordism). Later work within the industrial districts genre emphasized the variety of forms taken by industrial districts (Anne Markusen rechristened them, "sticky places" recognizing four types); global interlinkage (Ash Amin and Nigel Thrift's "Neo Marshallian nodes in global networks"); and the importance of an appropriate institutional framework, in particular, the presence of thick social relations among firms and the larger communities in which they were set (Karl Polanyi's idea of "embeddedness").

Paralleling the work on industrial districts, and drawing partly on Polanyi but also on the economic sociologist Mark Granovetter, was allied research on firms and social networks. Here the relation among firms is determined by social relationships, that is, by the social network among key decision makers who head different firms. Social networks are so important because they are the limits of the decision maker's social world, and thus the limits of who can be relied upon and trusted. Moreover, they are not necessarily confined locally, as in industrial districts, but can stretch across the globe. Their form is topological, independent of proximity, given only by the location of other members of the social network. One particular social network investigated by economic geographers was the Overseas Chinese (particularly by Henry Yeung).

Innovation and High Tech

Ideas of both social embeddedness and networks have been at the forefront of an outpouring of work, again much of it set within industrial geography, on innovation and high tech development. Some of that work was completed by the California school – Anna Lee Saxenian wrote the classic account of high tech development in Silicon Valley and Boston's Route 128 – but a lot is undertaken also by Northern European economic geographers concerned with state led regional development.

The interest in high tech initially focused on the computer (both hardware and software development), but it expanded to include biotech, and more generally all forms of digitally based design industries. High tech emerged as a research topic partly because of its in creasing importance to Western economies (and the seeming antidote to de industrialization); partly because its locational logic was quite different from traditional manufacturing industry – it was frequently suburban, and drawn by entrepreneurs and skilled labor found around university hubs; and partly because it offered the po tential to try out new methodological and theoretical frameworks.

Saxenian's study, for example, rested on the novel (novel, that is, for economic geography) method of long term ethnographic research. Given her method, it was perhaps not surprising that she also stressed the im portance of culture. High tech, Saxenian argued, is a cultural practice as well as an economic one, with culture seeping into the minutiae of everyday business, regu latory, and work routines. Michael Storper subsequently coined the term "untraded interdependencies" in his work on high tech to understand better the role that culture played. He used "interdependence" to stress the collective, communal nature of work in the industry, and "untraded" to signal that high tech was different from other economic sectors because it involved many non market based transactions.

Given the success stories of high tech in places like Silicon Valley and on Route 128, an obvious question was whether they could be replicated elsewhere. This lay behind much economic geographical research especially during the 1990s on 'creative' or 'innovation milieus', or 'learning regions' and 'networks'. The purpose was to find the elements that made successful high tech complexes – for example, particular governance structures, or re lations among firms, or involvement by workers, or the presence of educational institutes – and then reproduce them in new regional contexts. It is a project that con tinues in different forms, most recently in Richard Florida's research on creative cities. Whether one be lieves success is repeatable depends upon the extent to which one thinks that the economic potential of a given place rests upon either elements that are unique and inseparable from the place itself, or are general and de tachable. Certainly many communities that literally bought advice about how to become the next Silicon Valley believed the latter, but the fact that there still remains only one Silicon Valley may suggest the truth of the former.

Labor, Bodies, and Work

Given the emphasis on production by economic geog raphers it is odd that labor and work until recently re ceived so little attention. Perhaps it was because labor as a factor of production was conceived as sedentary, stuck in place, while capital was represented (more interestingly) as mobile and active. As David Harvey once put it, "unlike capital, labor has to go home every night." But the attitude to labor and work has changed and it has become a fertile topic of research.

In part, the new interest stems from globalization which has spurred research at both ends of the labor market. On the one hand, there are studies of the inter national movement of various professionals such as ac countants, bankers, and high end managers, as well as entrepreneurs such as among the Overseas Chinese and Indian high tech engineers. On the other hand, there is also work examining patterns of immigrant workforce segmentation in low end jobs by gender and ethnic ori gin, as well as case studies of particular occupations, such as janitors and back of the house hotel workers.

Additionally, there has been a changed conception of labor, one concerned to recognize its greater agency, its ability to make a difference to economic geographical outcomes. In part this revamped view is associated with research on the geography of labor unions. Andrew Herod, whose work in this area has been important, makes a useful distinction between the geography of labor, the older view of labor as passive, and labor geography, the notion that labor possesses some auton omy. In this newer view, while workers still go home at night, at least during the day, as an organized collective, they possess sufficient force to shape the economic geo graphical conditions in which they find themselves.

An important strand of writing, deriving especially from post structuralism, examines the performance of work, and the role various bodily social markings play. Doreen Massey's writing on *High Tech Fantasies* was the first to recognize the gendered nature of high tech work. While high tech work was often couched as an ethereal, abstract form of rationality, Massey contended the corporeal reality told another story. Based on her study of Cambridge (UK) high tech firm, Massey argued such rationality was achieved by workaholic young, white male bodies that were sustained at the worksite by fe male secretarial and administrative labor and at home by female partners undertaking domestic labor. McDowell's study of *Capital Culture* involved merchant bankers in the City of London. In her case, she was concerned with how the bodies of both men and women, including their markers of sexuality, were contorted in the performances of making large sums of money for their employers (and sometimes for themselves). And at the other end of the employment spectrum, Geraldine Pratt examined *Working Feminism* through the bodies of Filipina domestic workers in Vancouver. The larger point is that labor and work are not neutral, just another technical input into production, but are inscribed by the social, cultural, and geographical contexts in which work is practiced.

Retailing and Consumption

Retail geography first emerged in the 1960s with Brian Berry and the work of his students at the University of Chicago. Following spatial science, the presentation was formal: tangential budget lines and bid rent curves, dis tance minimizing constrained maximization equations, and quasi hexagonal grids of inter and intra urban retail location. As spatial science languished, however, so did retail geography. But it has been recently recouped, and like the contemporary retailing experience itself, re packaged into more enticing forms.

While during the first part of the twentieth century there were retail giants like Sears Roebuck or Woolworth, some of which even internationalized, they did not come close to matching late twentieth century trends of ex ponential growth in corporate retailing and retail global ization. Wal Mart is the classic example. Founded in 1962, within 40 years it was the largest corporation by revenue on the planet, with a million employees, and located in 14 countries. Other retailers, for example, France's Carrefour and the Netherlands's Ahold, while not as large became even more internationalized. Wal Mart also helped pion eer a different form of retailing, the big box store, trans forming the economic geographical landscape of towns and cities first in the West but then everywhere. Big box stores frequently located in the suburbs rather than the city center (the presumed destination for shops in the old retail geography), and took the form of warehouses rather than traditional shops. The locational advantages were cheap land rent, space for on site parking, and highway accessibility. This was not how old retailing was done, but it sparked interest by economic geographers because it fitted with other emerging disciplinary motifs such as corporatization and internationalization.

Another motif with which the new retailing fitted was the cultural turn's emphasis on meaning and identity. We shop till we drop not because it rationally yields the greatest amount of economic pleasure and satisfaction (the presumed motivation in the old retail geography), but because it enables us to express our identity: we are what we consume. We buy things in order to signify who we are. Studies by economic geographers of the new geography of consumption, from drinking Starbucks coffee to trawling the shelves of the local Safeway, show that going shopping is always more than buying goods. It is to participate in the larger cultural performance of affirming social identity.

Finally, linked to the recognition that goods are more than just economic commodities are economic geo graphical studies illustrating the moral geography of consumption, and represented by sites of sweated labor, or clear cut old growth forests. Once we defetishize consumer goods by knowing the details of their social production, including their geography, such studies suggest we no longer see goods as simply good, as only 'things' for consuming pleasure. One technique used by economic geographers to de fetishize is commodity chain analysis. A commodity chain traces the entire trajectory of a good from its conception and design, through pro duction, retailing, and final consumption. Such chains by definition are geographical, and they make it immedi ately apparent, as David Harvey once put it, "where our breakfast comes from." That is, they direct us instantly to other parts of the world where our goods are made, and germane here, they direct us to the moral conditions of production found at such sites whether that be in papaya plantations in Jamaica, cut flower nurseries in Kenya, or virgin cedar and spruce stands in British Columbia.

Producer Services and Finance

Although services as opposed to manufacturing have been the dominant form of employment in high income Western countries, since around 1950 it has taken a long time for economic geography to catch up to this fact. Even during the 1980s, manufacturing was given priority as a study topic. Finally, this is beginning to change.

Initially much of the discussion focused on what are called producer services, that is, high order service firms such as advertising, management consultancy, account ancy, and legal services, that serve the business com munity. Such firms tend to cluster geographically, often around client head offices, are strongly interconnected, rely on face to face contact, and depend upon the edu cational credentials, knowledge, and expertise of the employees hired. At the core of the producer service complex is the generation of specialized, high value knowledge and information, and creating large financial rewards for those engaged in its production.

Subsequently, much of the attention has focused on professional workers undertaking producer services par ticularly in global cities. Through the information they provide, they are critical to facilitating the larger pro cesses of internationalization gripping the economy for the last quarter century. Further, as a workforce they are globalized themselves with a high degree of international mobility. Most recently, it is their culture and the culture of the work they perform that has been scrutinized. For some economic geographers, professional workers are emblematic of wider changes occurring in the economy as the central product becomes knowledge. Producing and selling that product, however, requires very different skills than under the older conditions of work, skills that now emphasize cultural acumen and intelligence. Con sequently how people work, the sites in which they work, the tools they need to work, and even the larger built environment in which they work have changed radically, and producer services are at the center of that maelstrom of change. Those who study them are at an ideal vantage point to understand the emergence of the new economy and its economic geography.

The producer service receiving the greatest con sideration in economic geography since the 1990s is the financial sector, along with its principal focus, money. Money is a commodity like no other; the commodity of all commodities. Given advances in telecommunications, as well as deregulation, money is the commodity with the least friction of distance. Calling it 'liquid' does not come close to representing its ability to move effortlessly around the world, and literally at the speed of light. It was the breakneck speed of money that led the economist Richard O'Brien to announce "the end of geography." Since then economic geographers have been keen to show how geography continues to matter. The existence of international financial centers like London, New York, and Tokyo are one form in which geography still matters. Money does not move on its own, but needs to be made to do so. And where this happens are in the international financial centers: they possess smart machines, smart buildings, and above all the smart people to propel money around the world. Geography matters also in that money does not flow everywhere but usually only along well worn circuits. Accordingly some places are left out, subject to financial exclusion as, for example, in large tracts of sub Saharan Africa, or in inner city US neigh borhoods. And geography matters for the opposite reason that some places possess advantages for holding money such as the Bahamas, or the Canary Islands, or Zurich, with money sticking to such sites. The point is that even money, that most fleeting and slippery commodities, cannot escape geography.

Nature and Resources

Finally, nature and resources have become the focus of an animated disciplinary discussion since the late 1990s. For a long period the larger topic was an intellectual backwater, and almost lost altogether under spatial sci ence. The simplifying assumptions required for spatial modeling meant that the uneven and lumpy nature of resources were treated as complicating factors, the an alysis of which was promised but indefinitely postponed. In the mid 1970s, David Harvey drawing upon Marx made a powerful case for considering nature along with the economy. This was not nature 'red in claw' but social nature, 'the production of nature' as it was called. Pro duction of nature did not mean creating something where nothing existed before. Rather, original nature, 'first na ture' as it was referred, was transformed by industrial capitalism becoming 'second nature': farmed fields, urban climatological hot spots, polluted air, and a scarred manufacturing landscape.

The original Marxist approach emphasized the cen trality of social relations in constituting the economic resource, and in doing so joined discussions with political ecology. But alongside these radical studies of nature and resources also went more traditional accounts, for ex ample, of specific sectors such as the wood products in dustry, or oil and gas, or agriculture. But even here there were attempts to offer relatively sophisticated theoretical renderings drawing upon, say, regulation theory or in stitutional economic theories such as found in the work of the Canadian economist Harold Innis. More recently, two themes have characterized the literature on nature and resources.

First, there is a discussion of the effects of neoliberal regimes of governance on resource ownership, pro duction, and distribution. The argument is that resources are treated differently under the new regime such as, for example, with the privatization of water in many juris dictions, or the deregulation of mining, or the trade liberalization around the export of primary commodities such as coffee or tropical fruits. Second, drawing upon especially work from science studies there is an attempt to rethink the nature of nature. The difficulty with the Marxist idea of the 'production of nature' is that nature appears to drop out, the consequence of only the social. The issue, then, is how to keep nature in the discussion, but also the social too. Useful here is the notion of a hybrid, that is, joining together two different entities to create a third that is related to but nonetheless different from its original constituents. A number of economic geographers have tried to develop hybrid conceptions of nature as both social and natural in studies of water, coffee, lumber, and fresh fruit.

Conclusion and Future Direction

Economic geography even until the late 1970s was rela tively staid and conservative. It was masculinist in in clination as well as membership; it had a clear sense of its object of inquiry, the economy, and which primarily meant manufacturing industry; it favored the tried and true methods of either statistical analysis or the expert interview; and it tended toward straight and narrow theory drawn from economics. As a result, by 1980 ac cording to Nigel Thrift, the discipline was "pretty moribund … at risk of boring its audience to death."

Contemporary economic geography is almost un recognizable by comparison. It is intellectually lively, open, eclectic, pluralist, possibly chaotic and anarchic. Inconstancy is the only constant, inconsistency the only consistency. The last adjective one would now use is boring. Boring would be a welcome respite. Accordingly there is no agreement about economic geography's def inition, or even whether a definition is important. Boundaries between economic geography and other fields are muddy and indistinct. Likewise, the idea of a separate discipline, and a separate empirical object of study, is also contested. Anything can be fodder for its inquiry, and any method can be fashioned appropriately to yield results.

Along with these internal changes the discipline is becoming less masculinist, partly because of the influence of feminist theorizing and methods, and partly because of recent formative contributions made by women such as Kathy Gibson, Julie Graham, Susan Hanson, Linda McDowell, Geraldine Pratt, and Erica Schoenberger along with the continuing influence of Doreen Massey. Economic geography is also becoming even slightly less Anglo American. The Second Global Conference on Economic Geography was held in Beijing in 2007 (the first was in Singapore in 2000) and drew participants from thirty six countries. The biggest contingent continued to come from the US, and the lingua franca was English, but it was a start. And this sense of expanded inclusion applies also to younger economic geographers. Since 2003 there have been bi annual Summer Institutes in Economic Geography held both in the US and the UK designed specifically to socialize graduate students and junior faculty into the discipline.

Not everything is rosy. There are signs of increasing fragmentation and discord, and critics from outside have complained of a hopeless "fuzziness" and a haphazzard approach, while even inside there are grumblings of a lack of rigor, focus, and policy relevance. In short, economic geography's future is not assured.

This is odd. For, the subject matter of the discipline is what so many people outside of the discipline want to talk about. There can be few disciplines whose object of investigation is more germane to the present historical moment. This has even led to other disciplines muscling into economic geography's traditional intellectual turf: such as economics, economic sociology, and even economic anthropology. While there is a temptation to turn one's back on such intrusions, to give the cold shoulder (and which certainly has been given to economics), an alternative strategy, and which may ensure that there is a disciplinary future, is to embrace them, to try to learn from them. The best future for economic geography paradoxically may be the continual dissolution of economic geography as we knew it.

See also: Commodity Chains; Cultural Turn; De-Industrialization; Globalization, Economic; Industrial Districts; Location Theory; Spatial Division of Labor; Spatial Science.

Further Reading

Barnes, T. J. (1996). *Logics of Dislocation: Models, Metaphors and Meanings of Economic Space*. New York: Guilford.

Chisholm, G. G. (1889). *Handbook of Commercial Geography*. London: Longman, Green.

Cook, P. (2004). Follow the thing: Papaya. *Antipode* 36, 642 664.

Dicken, P. (2007). *Global Shift: Mapping the Changing Contours of the World Economy*. New York: Guilford.

Gibson Graham, J. K. (1996). *The End of Capitalism (As We Knew It): A Feminist Critique of Political Economy*. Oxford: Blackwell.

Hansen, S. and Pratt, G. (1995). *Space, Place and Gender*. New York: Routledge.

Hartshorne, R. (1939). *The Nature of Geography*. Lancaster, PA: Association of American Geographers.

Harvey, D. (1969). *Explanation in Geography*. London: Edward Arnold.

Harvey, D. (1982). *The Limits to Capital*. Chicago, IL: University of Chicago Press.

Isard, W. (1955). *Location and Space Economy*. New York: John Wiley.

Jones, C. (1935). *Economic Geography*. New York: Henry Holt.

Leyshon, A. and Thrift, N. (1997). *Money/Space: Geographies of Monetary Transformation*. London: Routledge.

Massey, D. (1984). *Spatial Divisions of Labour: Social Structures and the Geography of Production*. London: MacMillan.

Massey, D. (1992). *High Tech Fantasies*. London: Routledge.

Massey, D. and Meegan, R. (1982). *Anatomy of Job Loss*. London: Methuen.

McDowell, L. (1997). *Capital Culture: Gender at Work in the City*. Oxford: Blackwell.

Peck, J. (1996). *Work Place: Social Regulation of Labor Markets*. New York: Guilford.

Piore, M. and Sabel, C. (1984). *The Second Industrial Divide*. New York: Basic Books.

Pratt, G. (2004). *Working Feminism*. Philadelphia, PA: Temple University Press.

Saxenian, A. (1994). *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.

Schoenberger, E. (1997). *The Cultural Crisis of the Firm*. Oxford: Blackwell.

Scott, A. J. (1988). Flexible production systems and regional development: the rise of new industrial spaces in North America and Western Europe. *International Journal of Urban and Regional Research* 12, 171 185.

Scott, A. J. (2000). Economic geography: The great half century. In Clark, G. L., Feldman, M. P. & Gertler, M. S. (eds.) *The Oxford Handbook of Economic Geography*, pp 18 44. Oxford: Oxford University Press.

Smith, N. (1984). *Uneven Development: Nature, Capital and the Production of Space*. Oxford: Blackwell.

Storper, M. (1997). *The Regional World: Territorial Development in a Global Economy*. New York: Guilford.

Thrift, N. J. and Olds, K. (1996). Refiguring the economic in economic geography. *Progress in Human Geography* 20, 31 37.

Tickell, A., Sheppard, E., Peck, J. and Barnes, T. J. (eds.) (2007). *Politics and Practices in Economic Geography*. London: Sage.

Watts, M. (1994). Oil as money: The devil's excrement and the spectacle of black gold. In Corbridge, S., Martin, R. & Thrift, N. J. (eds.) *Money, Power, and Space*, pp 406 444. Oxford: Blackwell.

Whitbeck, R. H. and Finch, V. C. (1924). *Economic Geography*. New York: McGraw Hill.

Relevant Websites

http://www.clarku.edu
Clark College: the website for the journal Economic Geography, the oldest journal in the field.

http://www.egrg.org.uk
The website of the Economic Geography Research Group of the Royal Geographical Society, Institute of British Geographers and which includes working papers and book reviews.

http://geog.uconn.edu
University of Connecticut, Department of Geography: the website of

the Economic Geography Specialty Group of the Association of American Geographers.

http://faculty.washington.edu/Krumme/gloss
University of Washington faculty web server: a website with an A Z listing of relevant terms for economic geography designed and maintained by Gunter Krumme at the University of Washington, Seattle.

Economic Geography, Quantitative

E. Sheppard, University of Minnesota, Minneapolis, MN, USA
P. Plummer, University of Calgary, Calgary, AB, Canada

Glossary

Complex Spatial Systems These are characterized by a large number of spatially interdependent and heterogeneous elements, where the interdependence between these elements is characterized by nonlinear dependence.

Entropy Maximization A method used to derive the most probable state of a spatial interaction system, given information on total inflows and total outflows and capacity constraints.

Gravity Model An analog with Newtonian mechanics in which the interaction between two places varies directly with the size of those places and inversely with the distance between them.

Methodological Individualism The view that society is best explained by reducing it to the (usually rational) beliefs and actions of the individuals who compose it.

Monte Carlo Methods A simulation exercise designed to explore how a system behaves by running repeated experiments in which the initial conditions of the system and rule-based outcomes are drawn at random from a known probability distribution.

Rational Choice The view that humans make choices that maximize their satisfaction (utility) given the limited set of resources at their disposal.

Stochastic Utility Functions Utility functions expressing individuals' preferences, in which parameter values vary stochastically across an otherwise uniform population according to a known probability distribution.

Quantitative economic geography has, at various times, taken one or more of three forms. First, dating back to the commercial geography of the nineteenth century, is the collection and cataloging of numerical statistics about the geography of economic activities and communications costs. Second, dating back to the 'quantitative revolution' in Anglo American human geography, is the use of statistical methods to explore, and test hypotheses advanced, concerning, spatial relations between economic and other related variables (utilizing quantitative as well as qualitative data). Third, also dating back to the 'quantitative revolution', is the use of mathematics as a language to advance theoretical propositions about the spatial organization of economic activity, and the economic evolution of landscapes. Much of the attention that quantitative economic geography receives in the scholarly literature is devoted to how and whether the economy can adequately be captured through the collection and statistical analysis of economic and other related data, and about the relative merits of quantitative and qualitative empirical research methodologies. Neither of the first two forms raises issues specific to economic geography, however, notwithstanding that quantitative data and methods are of particular relevance in economic geography, due to the nature of economic concepts and information. We thus focus here on the use of mathematics to conceptualize economic geography and advance propositions about the nature of the observable spatial economy. With very few exceptions, mathematical economic geography has concerned itself with analyzing the spatial dynamics of capitalism, notwithstanding the coevolution of a variety of other economic systems (e.g., socialist, subsistence, slavery, feudal, and community economies) at various times, and in all kinds of places.

The use of mathematics to theorize economic geography can be traced back to continental European economists, particularly Alfred Weber (1868–1958), Tord Palander (1902–72), Johann von Thünen (1783–1850), and August Lösch (1906–45), as well as the German geographer Walter Christaller (1893–1969) and the American political scientist Harold Hotelling (1895–1973): 'location and spatial interaction theory'. Within geography, an initial flourishing of quantitative economic geography occurred, starting in the late 1950s, with the initiative of US based geographers Edward Ullman (1912–76), William Garrison (1924–), and Brian Berry (1934–); British geographer Peter Haggett (1933–); and Swedish geographer Torsten Hägerstrand (1916–2004), the quantitative revolution began with economic geography. In the early 1990s, this approach was reinvigorated within economics as a result of new mathematical models pioneered by the American economist Paul Krugman (1953–): 'the new geographical economics'. Within geography, this approach came under significant criticism in the 1970s and 1980s, for both its epistemological agenda and its ontological commitment to mainstream neoclassical economics. Thus, some geographers and regional scientists, notably Trevor Barnes (1956–), Panos Liossatos (1941–), Eric Sheppard (1950–), and Michael Webber (1941–) began experimenting with the use of mathematics for Marxian theorizations of the capitalist space economy: 'regional political economy'.

Under the aegis of the quantitative revolution, and drawing on the preexisting scholarship noted above, economic geographers sought to develop mathematical theorems to describe the spatial organization of a market economy. The bulk of this research, under the label of 'location theory', sought to determine the optimal (i.e., profit maximizing or cost minimizing) point location for individual capitalist enterprises, within the continuous space of a market economy. Industrial location theory sought to identify the point where aggregate transpor tation costs, for both assembling inputs and marketing the commodity produced, are minimized. Central place and retail location theory sought to identify the best locations for individual retailers, taking into account competition from similar firms and the attractions of locating within retail clusters (central places). In both cases, the analysis concentrated on transportation costs, for producers and consumers, as the prime determinant of spatial variation in overall production costs, and therefore total profits. Spatial variation in other costs, such as land, labor, cap ital, and taxes, and the effects of agglomeration econ omies (savings accruing to co location with other economic activities) were typically added as secondary considerations. Capitalists and consumers were con ceptualized as engaged in self interested, fully informed, rational choice. Space was conceived of as a uniform plane, and location theorists sought to understand why economic activities would emerge at some places rather than others on this plane. This research was also ex tended to consider public goods offered by the state, such as parks, education, and emergency and medical facilities. Complex location allocation and public facility location models were developed to determine the point locations that either maximized coverage of a spatially distributed population, or maximized accessibility, or minimized inequities in access.

A second body of scholarship, pioneered by William Alonso (1933–99), examined land use patterns: agri cultural or urban activities that occupy extensive spatial territories or land use zones, rather than point locations. Rational capitalists were assumed to choose, for the land they own or rent, that activity that maximizes profits. By the same token, residents of a certain kind would choose the location that is so desirable that they are willing to pay more for that location than anyone else. These choices were driven by the land market: competition for land means that each location is sold or rented to the highest bidder, who outbids others by using that land in a profit maximizing (or utility maximizing) way. Thus, wheat is grown in regions where it is the most profitable crop, and low income households occupy those parts of the city where they can outbid all other users of land. (Little attention was paid to other space intensive ac tivities, such as coalfields, fishing, or minerals extraction.)

A third body of research, pioneered by John Q. Stewart (1894–1972), William Warntz (1922–88), and Alan Wilson (1939–), examined spatial interaction theory and the economics of transportation. One subset of this research sought to determine cost minimizing transpor tation flows: which modes of transportation would be cheapest over which distances, and which routes through a complex transportation network will minimize time or cost. A second subset, more descriptive than normative in its approach, sought to predict how much spatial inter action (migration, commodity flows, and capital flows) would occur between places given prevailing transpor tation and communications technologies. Gravity models and entropy maximizing models were commonly used as tools for this purpose: seeking to predict the most likely patterns of flows, taking into account the characteristics of origins and destinations (and nearby destinations), the costs of interaction, and any known constraints such as information on total inflows and outflows, and capacity constraints on links. Subsequently, many attempts have been made to demonstrate that the empirically robust empirical predictions of such descriptive models can be grounded in theoretical models of rational choice in the presence of stochastic utility functions. These attempts to provide a rational economic foundation to gravity models are, however, controversial.

A fourth body of research, pioneered by Yorgos Papageorgiou (1943–) and Emilio Casetti (1928–), sought to bring all of these considerations together to simul taneously determine the spatial organization of an entire local economy (often conceptualized as a city): com modity production, consumption, agricultural, residential and commercial land use patterns, and transportation. The approach taken was drawn from neoclassical eco nomics, to which space was added as an extra consider ation. Assuming perfect competition and full information, every economic actor seeks to maximize his economic welfare, with the market determining production levels, consumption, and location. Like neoclassical economics, this scholarship made very strong assumptions about behavior (rationality), methodological individualism (microfoundations), and market structures (perfect competition), adding assumptions about preexisting spa tial structures (an isotropic plane). Importantly, this re search showed that the minor change of including even isotropic space into neoclassical economic theory calls into question the foundational principle of that theory: that a competitive market economy minimizes capitalists' profits and maximizes consumers' welfare.

Alternative comprehensive approaches were de veloped by Leslie Curry (1915–2009) and Michael Webber, who sought to determine the most likely spatial organization of activities, land uses, migration, trade, and labor markets without presuming rational, fully informed choice. Like models of spatial interaction, this approach replaced very strong assumptions about rational choice making with minimalist (even random) assumptions about behavior, showing that predictable spatial patterns nevertheless emerge. Similarly, Alan Wilson and his colleagues sought to extend their systems based spatial interaction methodology to comprehensive models of the local economy. These models explicitly incorporate the types of nonlinear interdependencies and structure changing behavior that are characteristic of chaotic and complex dynamics: 'complex spatial systems'. Typically, scholars following these approaches were more skeptical of the idea that the world must be understood through a set of mathematical theorems, and were willing to experiment with alternative approaches to analyzing the properties of their models through computational experimentation, most notably examining the empirical implications of their theorizations using Monte Carlo simulation.

In the early 1990s, Paul Krugman catalyzed a renaissance of location theory in the spirit of neoclassical economics, the 'new geographical economics'. The impetus for this was a new set of mathematical techniques introduced into mainstream economics by the US based economists Avinash Dixit and Joseph Stiglitz to model monopolistic competition using standard microeconomic tools (rational choice). Krugman's innovation was to apply these, successively, to international trade and the emergence of industrial clusters. Grounded in this highly simplified model of monopolistic competition, the individual rational choice foundations of neoclassical economics, and the assumptions of friction of distance and undifferentiated space inherited from earlier location theories, the new geographical economists sought to extend Krugman's initial insights to account for a set of stylized facts about the spatial configuration of specialization and trade, agglomeration, and regional economic growth. In order to gain some purchase on the behavior of these types of spatial economic systems, conventional proof theorem mathematics was replaced by computational experiments, but always within the methodological bounds set by the methodological individualism of rational choice theory.

Alongside these developments of mathematical spatial economic theory, economic geographers such as Luc Anselin (1953–), Daniel Griffith (1951), Robert Haining (1949–), and Bernard Fingleton (1949–) were active in developing methodologies that would enable such theories to be tested, in good positivist fashion, against empirical observation. Beginning in the late 1960s, enormous progress was made in measuring the confounding effects of spatial pattern on inferential statistics (measuring spatial autocorrelation and determining its effects on statistical inference), in developing methods of inferential spatial statistics that address these problems, and in incorporating these into a methodological program of spatial econometrics. More recently, both spatial interdependencies and spatial heterogeneity were conceptualized as resulting from the nature of the economic process being modeled, in particular, those hypothesized by the new geographical economics, rather than as some 'nuisance' parameter to be controlled to ensure unbiased, efficient, and/or consistent statistical inference. Increasingly, these approaches sought to match the desire of the mathematical theorists to derive foundational theorems about the economy, with definitive hypothesis tests.

The shape taken by this first school of quantitative human geography, as for all intellectual endeavors, was strongly influenced by the time and place within which it emerged. Post 1945 North America and Europe was a period when social science sought to gain prestige by emulating the natural sciences; when the natural sciences saw themselves as providing a powerful value free account of nature by adopting logical empiricism (positivism) as its philosophy; when neoclassical economics was the discipline closest in approach to the natural sciences; and when there was strong belief in the power and ability of quantitative social science to improve society through social engineering. Under these circumstances, it seemed perfectly natural for quantitative economic geography to adopt the twin discursive frames of logical empiricism and neoclassical economics. Both positivism and the capitalist market economy came under increasing criticism, however, in the politically turbulent years of 1968–73. Even enthusiasts of quantitative economic geography, such as David Harvey (1935–) and Allen Scott (1940–), became disenfranchised from the twin ideas that positivism would enable the engineering of a brave new society, and that markets were the key to this. As a consequence, a new school of radical economic geography emerged that was severely critical of neoclassical economics (a critique that was already immanent, given the shortcomings of this theory identified once space was included into neoclassical theory). Radical geographers were equally critical of what was termed the spatial fetishism of this first school of quantitative geography (the assumptions of an isotropic plane, and of space as a variable exogenous to the economy), and its use of mathematical reasoning and quantitative methods. As a result, quantitative economic geography was taken to be equivalent to neoclassical and logical positivist economic geography, and was rejected by most radical geographers as empiricist and bourgeois.

Nevertheless, in the 1980s, Allen Scott, Gordon Clark (1950–), Eric Sheppard, Panos Liossatos, Trevor Barnes, and Michael Webber catalyzed a mathematical economic geography in the spirit of Marxian political economy and post Keynesian economics, influenced by the work of the Italian economist Piero Sraffa (1898–1983), the Japanese economist Michio Morishima (1923–2004), and the American economist John Roemer (1948–). This body of research, which developed into what has become known as 'regional political economy', has extended parallel modeling in economics to account for the flow of commodities, capital, and technical knowledge between regions. It has examined the relevance of a labor theory of value in a capitalist space economy, and has sought to move away from the equilibrium models popular in economics. A persistent conclusion has been that a cap italist space economy is unlikely ever to be in equi librium, and that its spatiality is a destabilizing force. Research has thus been undertaken into the dis equilibrium dynamics of a capitalist space economy.

In the process, it has been shown how the introduction of geographical space – particularly the transportation sector and transportation infrastructure – calls into question some of the theoretical principles of both mainstream economics and mainstream Marxist theory. While rejecting the desire of economists to ground ag gregate models of the economy within microfoundations, research has sought to examine the spatial behavior of capitalists: the adoption of new technologies, and the pricing of commodities in spatially dispersed markets. Empirical work has examined growth dynamics at the national and international scale, the adoption and imi tation of new technologies within and across regions (establishing, for example, that persistent differences in technological mix and the trajectory of technical change exist between regions), the evolution of post Fordist in dustrial districts, and spatial pricing patterns within urban areas.

Methods have included equilibrium and dis equilibrium modeling, evolutionary economic models, and spatial statistics. They also increasingly include the use of computational and simulation modeling, instead of seeking to simplify the economy to the point where it becomes tractable for the generation of general math ematical results. This has demonstrated that the com monly received view that quantitative methods are necessarily positivist or neoclassical is, in fact, erroneous. The study of connections between this research and the interdisciplinary field of complexity theory has un covered parallels between this kind of mathematical economic geography, dialectical analysis, and certain kinds of Deleuzian theory (Sheppard, 2008).

These very different conclusions about how a capit alist space economy works, notwithstanding the use of the same kind of mathematical theoretical reasoning, reflect a very different set of assumptions. Location theory sought to explain the space economy through the rational choices of self interested individuals, took geo graphic space to be exogenous to the economy and de terminant of economic outcomes, and presumed that the economy is tendentially at or near equilibrium. While space undermined perfect competition, the operating principle was that these imperfections and externalities could be identified and corrected through state inter vention and social engineering. By contrast, within re gional political economy different kinds of economic actors are assumed to occupy different class positions within the economic system (defined by the factors of production they own, their position in the hierarchy of economic decision making, and their relative wealth), differentially conditioning the possibilities available to them as individuals. Thus social structures exist which condition action, making methodological individualism an inadequate conceptualization of human behavior. In addition, it is assumed that differences exist between individual actors occupying similar class positions in terms of relative location, resources, preferences, tech nologies, and so forth, and that actors possess only limited information about the current and future economic landscape. Space is not exogenous, because accessibility, transportation technologies, and infrastructures co evolve with the space economy. It can then be deduced that equilibrium is rarely to be expected, and that a capitalist space economy is conflict ridden and characterized by sociospatial inequality and uneven development.

See also: Competitiveness; Complexity Theory, Nonlinear Dynamic Spatial Systems; Dialectical Reasoning and Dialectical Materialism; Economic Geography; Economics and Human Geography; Location Theory; Political Economy, Geographical; Quantitative Methodologies; Spatial Autocorrelation; Spatial Interaction Models.

Further Reading

Alonso, W. (1964). *Location and Land Use*. Cambridge, MA: Harvard University Press.
Anselin, L. (1988). *Spatial Econometrics: Models and Methods*. Berlin: Springer.
Brakman, S., Garretsen, H. and van Marrewijk, C. (2001). *An Introduction to Geographical Economics: Trade, Location and Growth*. Cambridge: Cambridge University Press.
Clark, G. L., Gertler, M. and Whiteman, J. (1986). *Regional Dynamics: Studies in Adjustment Theory*. London: Allen and Unwin.
Curry, L. (1998). *The Random Spatial Economy and Its Evolution*. Aldershot: Ashgate.
Fingleton, B. (ed.) (2007). *New Directions in Economic Geography*. Basingstoke: Edward Elgar.
Krugman, P. (1991). *Geography and Trade*. Cambridge, MA: MIT Press.
Lloyd, P. E. and Dicken, P. D. (1977). *Location in Space: A Theoretical Approach to Economic Geography*. London: Harper and Row.
Losch, A. (1954). *The Economics of Location* (first publ. 1940). New Haven, CT: Yale University Press.
Scott, A. J. (1980). *The Urban Land Nexus and the State*. London: Pion.
Sheppard, E. S. (2000). Geography or economics? Contrasting theories of location, spatial pricing, trade and growth. In Clark, G., Gertler, M. & Feldman, M. (eds.) *Handbook of Economic Geography*, pp 199 219. Oxford: Oxford University Press.
Sheppard, E. (2008). Geographic dialectics? *Environment & Planning A*, 40, 2603 2612.
Sheppard, E. S. and Barnes, T. J. (1990). *The Capitalist Space Economy: Geographical Analysis after Ricardo, Marx and Sraffa*. London: Unwin Hyman.
Webber, M. J. and Rigby, D. L. (1996). *The Golden Age Illusion: Rethinking Postwar Capitalism*. London: Guilford Press.
Wilson, A. G. (2000). *Complex Spatial Systems*. London: Prentice Hall.

Economics and Human Geography

J. R. Faulconbridge, Lancaster University, Lancaster, UK
S. Hall, University of Nottingham, Nottingham, UK

Glossary

Economics The study of rational human behavior (microeconomics) and its implications for economic decision making and the production and behavior of national economies (macroeconomics).

Policy Relevance Work that has a methodology and a message that influences policymakers rather than purely being of interest to an academic audience.

Introduction

Since the work of Adam Smith in *The Wealth of Nations*, work in the discipline of economics has focused upon the study of human decision making in relation to the use of scarce resources. Lionel Robbins' 1932 book entitled *An Essay on the Nature and Significance of Economic Science* provides a more recent discussion of these principles. Here, he reinforces the message of Smith and notes that the core of mainstream economics is (1) the study of rational human behavior (microeconomics) and its implications for economic decision making; and (2) the production and behavior of national economies (macroeconomics). Explorations of the effects of value judgments or aberrant behaviors on decision making are said to fall outside of the realm of such study.

Human geographers meanwhile, and especially economic and urban geographers, have at first glance preoccupied themselves with not dissimilar questions in terms of studying the economy, economies, and their urban manifestations. This is perhaps most obvious in relation to theories of agglomeration and localization that have been used by both geographers and economists to analyze the tendency of people and economic activities to cluster together in cities. However, while on the surface this seems to suggest shared and complementary agendas that might allow interdisciplinary exchanges, the relationship between human geography and economics has at best been complex and contested, and at worst been adversarial, argumentative, and even nonexistent.

Changing Times, Changing Relationships

The often fraught relationship between human geography and economics is by no means a recent development. Dating back to the late 1800s and as indicated above, the relationship has been characterized by the different disciplinary approaches taken to often similar research problematics. For example, Barnes writing in the book *The New Industrial Geography* outlines how the first English language textbook in economic geography, George Chisholm's *A Handbook of Commercial Geography*, sought to position commodity production and trade as fundamentally 'geographical' research concerns. While economists such as David Ricardo and J S Mill had studied trade, particularly through the development of concepts such as comparative advantage, Chisholm sought to demonstrate that it was precisely through geographical differences that such comparative advantages emerged. Moreover, in addition to signaling the importance of geography to economic research agendas, he also forwarded a distinctive methodological approach that privileged detailed empirical research and the presentation of arguments through a distinctly geographical imagination, most notably through maps.

Consequently, while the early relationship between human geography and economics in many ways foreshadows contemporary differences between the two disciplines, at a certain point in the history of human geography, the potential for a close relationship with economics seemed to exist. The 'quantitative revolution' of the 1950s and 1960s ushered in a period of more positivist study in which the modeling of spatial structures dominated the work of most (but not all) human geographers. This was both methodologically and philosophically closer to economics than present day human geography. The chance of a long standing relationship developing diminished, however, with changes in human geography that took place from the 1970s onwards. Two such developments stand out as being particularly significant – developments that echo the earlier relationship between the two disciplines.

First, led by the work of Manuel Castells, David Harvey, and Doreen Massey, human geography underwent a critical and radical turn in the 1970s and 1980s in response to economic changes that were leading to deindustrialization, mass unemployment, and social inequality in 'developed' economies. This gave energy to the work of geographers that emphasized critical, normative politics. Scholarship sought to do more than explain sociospatial phenomena, instead being designed to develop social science theory while actively influencing policy in a way that might reduce inequalities and highlight the exploitative nature of, and inherent contradictions of, capitalism. Research in this period was heavily influenced by Marxist political economy that emphasized the intricate relationships between production and dis tribution and the need for revolutionary change in cap italist systems so as to break down class relations and associated inequalities. This was a very different political economy to that espoused by Adam Smith, mainly be cause of the desire to overhaul rather than explain sys tems of accumulation and the expropriation of surplus.

This opened up a disjuncture between the projects of human geographers and economists. Not only were geographers increasingly radical in their work but a methodological revolution was also further distinguishing their research from economists. New interest in the im pacts of economic structures and systems, the winners and losers and the lived experiences of deindustrializa tion and unemployment that were resulting from capit alism's spatial adjustments of the 1970s and 1980s, led to a burgeoning amount of qualitative research. This meth odological shift was far from total and many human geographers continued and continue today to use quan titative methods to explore a range of geographical questions. For example, Sheppard and Barnes in their book *The Capitalis Space Economy* provide perhaps one of the most thorough analyses of how a quantitative ap proach can reveal the intricacies of location as the interaction of places and labor as a heterogeneous social entity, something that allows a direct critique of neo classical economic work to be made. Nevertheless, a significant switch away from quantitative methods swept the field of human geography in the 1980s and set the foundations for the predominance of qualitative research that now pervades human geography.

Second, this methodological shift was given further momentum by the growing influence of the 'cultural turn' in human geography from the 1990s. Inspired by the growth of cultural studies as an academic discipline from the late 1980s onwards, as well as the rise of diverse and flexible post Fordist regimes of production and con sumption in capitalist space economies, the 'cultural turn' had a significant impact on the ways in which human geographers study the economic. Most broadly, research into 'the economic' began to place emphasis on more 'qualitative' dimensions including: meanings, identities, trust, and knowledges. This was accompanied by research into neglected and relatively new areas of the 'economy', including most notably e commerce, consumption prac tices, and commodity chains and networks. Moreover, there was increasing interest in drawing on post struc turalist thought to critically rethink the economy as an object of analysis. In particular, geographers such as Andrew Leyshon, Roger Lee and Colin Williams in their book *Alternative Economic Spaces* sought to demonstrate that 'the economy' is not necessarily contiguous with 'capitalism' and pointed to the emergence of a range of proliferative and diverse economies.

As a result of these shifts, corporate interviews have been joined by participant observation, participatory methods, and discourse analysis as commonly used re search tools with proponents arguing that such methods allow researchers to access the meanings attached to economic actions in ways that quantitative methods do not facilitate. Therefore, while it is important to note that the growing appreciation of the role of cultural factors in socioeconomic practices is by no means limited to human geography, it is widely acknowledged that for good or for ill, the effect on the discipline of both critically and culturally inflected research into the economic and economies has been substantial, distinguishing it from economics.

In many ways, these developments have been inter preted as a furthering of the 'divide' between human geography and mainstream economics. However, there is nothing inevitable about this. First, as noted above, quantitative geographical studies continue to offer, meth odologically at least, points of convergence that could inform discussions. Second, interest in processes of glob alization and the geographical stickiness of economic ac tivities also creates shared research questions, addressed both by geographers and economists using concepts such as agglomeration and localization through the study of social processes and place specific institutional endow ments. Barnes highlights how, as a result, work in insti tutional economics and 'new industrial geography' effectively converged in the 1990s around five shared questions about: local and reflexive knowledge; the role of institutional analysis; the affects of embeddedness and networks; the power of cultural norms and conventions; and path dependency and 'lock ins' in economic devel opment. Third, this reveals then how economics, in a similar vein to human geography, is now a diverse dis cipline with multiple subdisciplines. Increasingly many of the subgroupings within economics are sympathetic to geographer's beliefs that humans are irrational, socially and spatially variegated heterogeneous actors and ortho dox and neoclassical Economics now sits alongside het erodox economics and emphasizes the historically, socially, and spatially constructed rationalities of agents. Similarly, specialty groupings (e.g., financial economists; inter national economists) approach questions of financial geography and internationalization theory in ways that geographers can relate to.

However, while this emerging plurality within both economics and human geography points to the possi bility for more engaged interdisciplinary discussion in the future, such opportunities have to be put in the context of the politics of both disciplines and the op portunities for and barriers to conversations that emerge when competitive rivalries and ideological positions exist, something that has been particularly significant in the 2000s.

Locating the 'Economy' in Contemporary Human Geography

As will be clear from discussions so far, debates about the relationship between human geography and economics have been principally driven by economic and urban geographers because of their explicit engagement with the 'economy' and 'economies'. The 2000s have seen, though, a number of ongoing debates emerge, principally within economic geography, about whether relationships with economists are appropriate or useful in the context of geographical studies of economies. Two strands cut through these debates that focus on academic approaches to the economy: definitions of the economy/economies themselves; and methodologies for studying the microeconomy.

First, then, 'definitions of the economic'. Amin and Thrift, in a provocative intervention, entitled *What Kind of Economic Theory for What Kind of Economic Geography?* made a call for geographers to abandon all hopes of talking to and working with economists. This argument was made on the basis that economists remain wedded to neo classical, resource based analyses of the 'economic' while human geography, in common with other social scientists such as anthropologists, sociologists, and psychologists, acknowledges the existence of multiple, socioculturally constructed economies. For Amin and Thrift, this fun damental ontological difference means that, following the cultural turn, any engagement with economists will ul timately not prove fruitful. Consequently, human geog raphers should 'play out' with other more like minded social scientists as part of a project to reveal the flaws of even the most heterodox economists.

Others are, however, less sure about this 'prescription' for the future trajectory of human geography's engage ment with the economic and see the focus upon the cultural and failure to fully theorize the processes leading to the construction of economic value as one of economic geography's weaknesses. There is also concern that many geographers are blind to the insights of heterodox eco nomics and the growing body of work on gender and culture that has many parallels with the type of 'econ omy' Amin and Thrift describe. Consequently, there is little agreement about whether human geographers should boycott or engage with (but not 'lie down' and be eaten by) the lions that are economists as part of a project to place economic geography at the center of social sci ence debates about the economy.

A second major debate provoked by Amin and Thrift's intervention surrounds 'the methodology used to study the economic'. On the back of the 'cultural turn', Amin and Thrift argue that micro social studies facilitated by ethnography and long term immersion in the field are vital for understanding the sociospatial constitution of the economy. According to Amin and Thrift, even microeconomists fail to achieve the type of textured analyses being produced by geographers, sociologists, and others. Instead, the scholarship produced by many economists is dominated by desk based theorizing and abstract claims that ignore the messiness of the 'real' world. This, they claim, is another reason to cast aside hopes of relationships with economists.

Reactions to such claims have again been mixed. One response has been to show that many geographers are effectively using quantitative methods and modeling to explore the complexities of sociospatial economies and that this provides the opportunity for useful dialogues to be established. Indeed, Plummer and Sheppard in their paper *Must Emancipatory Economic Geography be Qualitative?* argue that the use of quantitative models by economic geographers is vastly different to that of economists and that the most effective way to highlight the frailties of economist's analyses is to challenge them using the quantitative language they understand rather than through qualitative insights that geographers' rivals quickly rubbish and dismiss. Slightly less provocatively, Sheppard writing about *The Economic Geography Project* calls for the establishment of 'trading zones' through which more effective dialogue can be established be tween researchers with different methodological stand points. Another response has been to question the value of completely jettisoning quantitative approaches from the work of geographers. The dangers of such a move are said to be evidenced by long held concerns about the 'fuzziness' of geographer's concepts and the use of 'thin' empirics to back up theoretical claims. Consequently, an argument has developed in support of the use of exten sive quantitative surveys that will be taken seriously not only by economists but also by policymakers. Amin and Thrift would probably counter that geographers shouldn't want to be taken serious by economists. But others clearly disagree.

Disciplinary Identities, Relevance, and Reproduction

The debates surrounding the intersection between human geography and economics around 'the economy' also go beyond academic questions of defining 'econ omies' and appropriate methodologies for studying them. They have also been couched within the wider context of the contemporary identity of geography (and economic geography especially) in both academic and policy cir cles. Here, two issues stand out as being particularly significant.

First, as hinted above, the need to sustain geography as a distinctive discipline with a unique toolkit able to analyze the effect of space and time on human socio economic practice has raised questions about whether engaging with economics might lead to the loss, rather than the consolidation, of a unique and potentially valuable geographical academic identity. For example, some geographers have argued that there is a need to return to the 'core competencies' of geography – particularly the development of conceptual toolkits to refine theorizations of space and place. These critics claim that the neglect of such endeavors has allowed other social scientists, notably sociologists, to encroach on the heartland of human geography.

Indeed, one of the most notable examples of the rise of geographical imaginations outside of human geography is within the discipline of economics. For example, the economist and special advisor to the United Nations, Jeffrey Sachs, adopts a crude approach to geography using both physical and human geographical differences (such as climate, proximity to coastal zones and population density) as explanatory variables for poverty rates globally. Importantly for our argument here, he frequently explicitly uses the term geography but does not reference the work of contemporary academic geographers. Meanwhile, the 'new economic geography' has been developed within economics itself. This intellectual tradition is most closely associated with Paul Krugman who argues that economists should take the inherently geographical nature of the economy more seriously. This work is particularly concerned with explaining the nature of agglomeration economies by drawing on ideas of imperfect competition and increasing returns to scale. Quite rightly, this growing interest in space and place within economics has been reflected by developments within human geography itself that aim to foster closer links with economists and develop more subtle geographical imaginations beyond human geography. In this respect, the founding of the *Journal of Economic Geography* in 2001 is significant. The journal has editors in both economics and human geography and explicitly states on its website that its aims "are to redefine and reinvigorate the intersection between economics and geography".

However, some geographers have argued that initiating a dialogue with economists, in common with debates with social scientists more generally, represents more of an exercise in importing ideas into human geography from outside, in this case from economics, than genuinely entering into an interdisciplinary discussion where theories and concepts are developed through engaged and open dialogue. This debate has been particularly well developed in terms of the relationship between human geography and one brand of heterodox economics, new economic sociology. Perhaps most closely associated with the work of Mark Granovetter, new economic sociology has been influential across a range of topics in human geography, particularly through the concept of embeddedness. For many human geographers, the attraction of this branch of heterodox economics is its move away from the efficiency and equilibrium principles associated with orthodox economists and its commitment to research a more diffuse set of political and cultural concerns – particularly the multiple contexts of economic action. However, as Jamie Peck in his article *Economic Sociologies in Space* argued that geographers have imported a rather simplistic reading of the sophisticated embeddedness literature within new economic sociology. Moreover, Peck suggests that new economic sociologists have been rather less interested in drawing on ideas from human geography and as such, the flow of ideas within this interdisciplinary dialogue is rather one way.

The second key set of debates surrounding the identity of human geography that impacts upon the discipline's interactions with economics are associated with the relationship between human geography and policy making. Commentators including Ron Martin, Jamie Peck and Michael Samers in a series of articles in *Antipode* and *Transactions of the Institute of British Geographers* have argued that the apogee of human geography's engagement with policymaking circles was in the 1970s and 1980s when the more critical flavor of the discipline's research resonated with the 'big' political questions of the time. However, more recently, it is claimed that work of a more cultural nature has gained far less traction in policymaking communities with economists thus dominating policy relevant research in the social sciences. For example, critics such as Andrés Rodríguez Pose and Henry Yeung argue that the impacts of the cultural turn have left human geography following 'sexy' subjects that are of little interest to broader policy communities and argue that what is needed is a 'policy turn' in the discipline to rectify this position. Moreover, the lack of political engagement by economic geographers is frequently contrasted with economics that is perceived to have had a more influential say in policy than human geography. Consequently, critics argue that while many of the 'big' policy questions are inherently geographical, such as the inequalities associated with globalization, human geographies have been unable to engage policy makers, thus raising major questions about the discipline's identity in the eyes of those making policy decisions.

Of course, it is important to recognize that there are multiple ways of engaging in policy debates. Indeed, on closer inspection, there is evidence of human geographers working across a range of topics on the economy that have been influential in policy circles, particularly in terms of regional governance and financial services. Moreover, it has been noted that engagement with policymaking can also mean being critical of policy communities themselves. Perhaps, the more significant issue in terms of reigniting human geography's interest in policy relevant work is, then, the question of the need to compete with economists in ways in which work is valued both by the academy in economics and human geography but also by the policymaker. Indeed, while economists strive for this form of recognition, human geographers typically devalue it vis à vis other academic pursuits, notably dissemination to their fellow geographical re search stakeholders.

Perhaps the most notable way in which human geographers have begun to think about possible avenues for policy intervention is through an ongoing debate surrounding the nature of both Left politics and what it means to be a left or critical human geographer. This debate has been led by a series of essays published since 2005 in the critical human geography journal *Antipode*. Some commentators have been optimistic about the fu ture trajectories of leftist geographies and its opportun ities for influencing policy in socially progressive ways, something that it is often argued, economists are not so successful at. For example, Amin and Thrift in their provocative piece entitled What's Left argue that rather than a single left, as perhaps characterized the earlier forms of critical human geography discussed above, there are in fact multiple lefts. Moreover, they claim that this is by no means necessarily problematic and could be seen as a significant resource. Others are, however, far more circumspect about the possibilities for rekindling the kind of critical human geography that characterized the 1970s and 1980s. The debate about how to compete or collaborate with economists is, then, likely to rumble on for some time.

Future Directions

Our review here has, in many ways, been rather Anglo American in its focus, primarily because of the domin ance of the work of scholars from the two countries in debates in leading geographical journals. There is, then, probably another story to be told from other perspectives. Indeed, our review has been necessarily selective and can only ever provide a snapshot of the diverse debates taking place about the relationship between human geography and economics. Nevertheless, the discussion in this art icle has revealed the importance of understanding how the relationship between human geography, and eco nomics, at least in the Anglo American world, has been dynamic and hotly contested in recent years. While not necessarily needing to be separate with many shared empirical and theoretical concerns, following the rise of a more critical human geography and the effects of the 'cultural turn', the two disciplines of human geography and economics have tended to develop in comparative isolation to one another, although there are still some important examples of where collaboration has taken place. Without doubt, the 2000s have been characterized by a series of debates surrounding the possible future trajectories that human geography, and particularly eco nomic and urban geography, could and should take and the way a relationship with economics might influence this. These debates have in many ways focused on how economic geographers might best respond to economics perceived dominance in policymaking circles and the potential problems surrounding the more decentered and multiple economic geographies that have emerged fol lowing the cultural turn.

However, while debates such as those surrounding the 'Left' in human geography offer a potential way for human geographers to reclaim the policy ground from economists, and the increasing recognition that more fruitful academic endeavors can stem from a genuine trade in ideas rather than human geographers being 'net importers' from disciplines such as economics are timely and well taken, there are other ways in which new intersections between economics and human geography could fruitfully emerge. First, there are considerable debates across the social sciences concerning the in creasing moves toward more post disciplinary futures. As initiatives such as the launch of the *Journal of Economic Geography* reveal, dialogues between economists and human geographers could certainly be part of this, al though there are important issues for discussion such as the epistemological and methodological convergences or divergences between the two disciplines that would need to be addressed.

Second, and following on, there is an increasing rec ognition that the 'big' research questions currently being addressed by social scientists are inherently post disci plinary and as such demand a range of disciplinary perspectives. This does not necessarily mean that issues previously at the forefront of both disciplines were ne cessarily purely focused on one discipline or the other. Rather, it points to an increasing recognition in both academic and policy circles that many of the issues currently being studied are best approached using a post disciplinary framework from the start. Indeed, echoing earlier developments in human geography in which understandings of the economy were broadened to con sider a range of diverse economies, economists them selves are broadening their own conceptions of the economic, often in ways that have an implicit geo graphical take and are being taken seriously in policy circles. For example, the Stern report on *The Economics of Climate Change* was undertaken by the economist Sir Nicholas Stern and was tasked with explicitly linking economic growth through gross domestic product (GDP) measures to potential future climate change scenarios. Its recommendations are also framed in terms of GDP spend needed to offset potential declines in GDP associated with climate change. Geographers have much to say about sustainable adaptations to climate change and might seek to position themselves in these debates. Meanwhile, the economist Lord Layard's hard hitting *The Depression Report* attempts to measure the cost to the UK economy of depression and chronic anxiety and uses this to urge reform of NHS care for these conditions, again based around the future GDP growth that could be enjoyed if his recommendations were implemented. In many ways, this resonates with the growing literature within human geography that is concerned with ex ploring geographical inequalities in health outcomes.

So in terms of topics such as climate change and health, economists are making significant inroads into political communities. However, their research could be developed in closer dialogue with human geographers as the latter become increasingly concerned to (re)discover their critical and political edge. That said, it is important to remember that future dialogues between human geographers and economists need to be framed around the fundamental differences both methodologically and in terms of disciplinary cultures. If genuine debate could be had around these, then the shared research topics between the two disciplines could provide the welcome opportunity for genuinely engaged post disciplinary work between the two.

See also: Cultural Turn; Economic Geography; Human Geography.

Further Reading

Amin, A. and Thrift, N. (1997). Globalization, socio economics, territoriality. In Lee, R. & Wills, J. (eds.) *Geographies of Economies*, pp 147 157. London: Arnold.
Amin, A. and Thrift, N. (2000). What kind of economic theory for what kind of economic geography? *Antipode* 32, 4 9.
Amin, A. and Thrift, N. (2005). What's left? Just the future. *Antipode* 37, 220 238.
Barnes, T. J. (1999). Industrial geography, institutional economics and Innis. In Barnes, T. J. & Gertler, M. (eds.) *The New Industrial Geography. Regions, Regulations and Institutions*, pp 1 22. London: Routledge.
Barnes, T. J. (2001). 'In the beginning was economic geography' a science studies approach to disciplinary history. *Progress in Human Geography* 25, 521 544.
Barnett, C. (1998). The cultural turn: Fashion or progress in human geography. *Antipode* 30, 379 394.
Castells, M. (1977). *The Urban Question*. London: Arnold.
Centre for Economic Performance (2006). *The Depression Report*. London: London School of Economics.
Chisholm, G. G. (1889). *Handbook of Commercial Geography*. London: Longman, Green.
Harvey, D. (1973). *Social Justice and the City*. London: Arnold.
Harvey, D. (1982). *The Limits to Capital*. Oxford: Basil Blackwell.
Johnston, R. J., Hepple, L., Hoare, T., Jones, K. and Plummer, P. (2003). Contemporary fiddling in human geography while Rome burns: Has quantitative analysis been largely abandoned and should it be? *Geoforum* 34(2), 163 164.
Krugman, P. (1991). *Geography and Trade*. Cambridge, MA: MIT Press.
Leyshon, A., Lee, R. and Williams, C. (2003). *Alternative Economic Spaces*. London: Sage.
Markusen, A. (1999). Fuzzy concepts, scanty evidence, policy distance: the case for rigour and policy relevance in critical regional studies. *Regional Studies* 33(9), 869 884.
Martin, R. (1999). The new 'geographical turn' in economics: Some critical reflections. *Cambridge Journal of Economics* 23, 65 91.
Martin, R. (2001). Geography and public policy: The case of the missing agenda. *Progress in Human Geography* 25(2), 189 210.
Martin, R. and Sunley, P. (2001). Rethinking the 'economic' in economic geography: Broadening our vision or losing our focus? *Antipode* 33, 148 161.
Massey, D. and Megan, R. (1985). *Politics and Method*. London: Methuen.
Peck, J. (1999). Grey geography? *Transactions of the Institute of British Geographers NS* 24(2), 131 135.
Peck, J. (2005). Economic sociologies in space. *Economic Geography* 81(2), 129 174.
Plummer, P. and Sheppard, E. (2001). Must emancipatory economic geography be qualitative? *Antipode* 33(2), 194 199.
Robbins, L. (1932). *An Essay on the Nature and Significance of Economic Science*. London: Macmillan.
Rodríguez Pose, A. (2001). Killing economic geography with a 'cultural turn' overdose. *Antipode* 33(2), 176 182.
Samers, M. (2001). What is the point of economic geography. *Antipode* 33(2), 183 193.
Sheppard, E. (2006). The economic geography project. In Bagchi Sen, S. & Lawton Smith, H. (eds.) *Economic Geography: Past, Present and Future*, pp 11 23. London: Routledge.
Sheppard, E. and Barnes, T. J. (1990). *The Capitalist Space Economy: Geographical Analysis after Ricardo, Marx and Sraffa*. London: Unwin Hyman.
Smith, A. (1776). *The Wealth of Nations*. London: Strahan and Cadell.
Stern, N. (2007). *The Economics of Climate Change*. Cambridge: Cambridge University Press.
Thrift, N. and Olds, K. (1996). Refiguring the economic in economic geography. *Progress in Human Geography* 20(3), 311 337.
Yeung, H. W. C. (2001). Does economics matter for/in economic geography? *Antipode* 33(2), 168 175.

Relevant Websites

http://joeg.oxfordjournals.org/
The Journal of Economic Geography.

Economies, Alternative

S. Healy, Worcester State College, Worcester, MA, USA

Glossary

Alternative Economies (1) Processes of production, exchange, labor/compensation, finance, and consumption that are intentionally different from mainstream (capitalist) economic activity; and (2) an alternative representation of economy as a heterogeneous and proliferative social space.

Class Process The form in which surplus (in labor, value, or product form) is produced, appropriated, and distributed. Class processes theorized by Marx and Resnick and Wolff include slave, feudal, capitalist, independent (ancient), and communal.

Community Enterprise A market- or nonmarket-oriented economic organization concerned to improve community well-being.

Diverse Economy Theoretical proposition that economies are intrinsically heterogeneous spaces composed of multiple class processes, mechanisms of exchange, forms of labor and remuneration, finance, and ownership.

Economic Difference A representation of economy as radically heterogeneous.

Performativity Theory that discourse participates in constituting the reality it purports to represent.

Introduction

Contemporary economic geography has seen a theoretical centering of the relationship among neoliberal governance, economic globalization, and capitalist development. The question within this research program is whether to see neoliberal capitalism as a single force defined by its regularities, or as a referent that describes a variety of relations between the state and the economy (see Harvey's *A Brief History of Neoliberalism*). Against the backdrop of this larger conversation about neoliberal capitalism, a small but growing number of geographers have turned their attention to a variety of economic activities and sites that can be described as alternative in one of two ways. In the first approach, the alternative economy is simply imagined to be something other than a presumptively dominant capitalism and its associated systems of governance. The term 'alternative economies' describes an array of processes of production, ownership, labor, exchange, and consumption that differ from those of the mainstream economy. A second understanding of the alternative economy coalesces around an epistemic break which posits the economy as an always already and intrinsically heterogeneous space. This second approach not only represents the economy as a field of difference but possesses the potential to reconfigure scholarly research within economic geography.

While there has been considerable excitement around the idea of alternative economies, and especially noncapitalist economic activity, the alternative economy remains a marginal concern because of the underlying and frequently unspoken spatial ontology that governs its academic (and popular) representation. In many representations, alternative is synonymous with self-consciously intentional efforts undertaken on a local scale. This association renders what it describes as peripheral and relatively powerless, vulnerable to cooptation or even state repression. From the outset the alternative economy is seen as socially insignificant and thus unworthy of scholarly attention. What is required to combat this double marginality is a theoretical framework that diverges from the conception of the alternative economy as local and intentional. The first step in developing such a framework is to recognize that the marginality of the alternative economy comes from defining something as alternative in the first place. The second step is to produce an ontology of economic difference that highlights the ubiquity (in place) of nonmarket and noncapitalist practices alongside the variety of capitalist forms.

The Trouble with Alternatives

The term 'alternative', by its very nature, underscores a foundational insight from modern linguistic theory – that no term derives its meaning self-referentially. The existence of an 'alternative' economy implies that there is a dominant or mainstream economy against which the alternative is defined. Moreover, the perceived spatial extent and viability of alternatives are shaped by how one understands the dominance of the mainstream economy. Not surprisingly, this issue is a principal point of contention among scholars interested in alternative economies.

Those who understand dominance from a realist epistemological perspective believe that it is possible to gauge relative degrees of power and the extent of vulnerability or powerlessness. They therefore almost always see alternative economies and organizations as weak and likely to be short-lived. Those who understand dominance as performative, on the other hand, see it as predicated upon and produced by the dissemination and repetition of knowledges. Neoliberalism, for example, has become hegemonic in large part because academic knowledges, policy discourses, protest movements, and bureaucratic technologies of enactment have made it the focus of belief and action, bringing it into being in both authoritarian and democratic settings. From this perspective, the con tinual interruptions, large omissions, and widening gaps in the performance of neoliberalism are openings for the other discourses and technologies that are always already participating in bringing different economies into being.

The resolution of this central debate as to whether the alternative economy really is a marginal set of activities or if it is performatively marginalized will determine the future direction of research into alternative economies. One interesting feature of this debate is that adherents of the performative perspective are frequently regarded by those adopting the realist perspective as idealists who imagine that thinking differently – an intellectual com mitment to celebrating noncapitalist spaces and practices – is all that is required to change the world. Celebration here is effectively a code word for an idealist naiveté that ignores the reality of capitalism and a devolved neo liberal state dedicated to its expanded reproduction. But what if the way in which this reality is invoked to reign in the potential study of something other than capitalism is itself part of what keeps capitalism dominant? Perhaps the persuasiveness of the argument that capitalism is the real economy, while the alternative economy is fanciful celebration, is the ultimate confirmation of performativity.

Research based on the realist vision of the alternative economy tends to focus on self consciously alternative economic institutions and practices such as the self employed retro retailers studied by Crewe, Gregson, and Brooks. These enterprises are threatened simultaneously by self employment burnout and by conventional stores targeting the same demographic. The researchers' analysis reflects a frequently expressed concern that the alternative is contingent and vulnerable in comparison with the systemic and encompassing dominant economy. The premise here is clearly a binary opposition in which superior qualities are ascribed to the mainstream.

If the alternative economy is understood from the outset as intrinsically deficient, it is not surprising that alternativeness is always vulnerable to erasure. Market competition, cultural cooptation, and state repression are seen as forces arrayed against the alternative economy that will, almost inevitably, reduce alternativeness to sameness. The continued existence of alternative eco nomic institutions – community based credit unions, for example – is seen as threatened by the dominant ideol ogy that governs finance, the laws of local, state, and national government, and the market forces that favor large capitalist firms. At the same time, it is imagined that alternative economic practices and institutions, such as local economic trading systems (LETSs), can exist only on the margins either in spaces that have been abandoned by traditional financial institutions (poor blighted com munities) or in locations with progressive and affluent populations that can elect to sustain them. While alter native economic institutions are conceived as existing only in contingent circumstances, the presumed domin ance of mainstream financial institutions creates the ap pearance that they exist autonomously, independent of conditions.

Gibson Graham's work attempts to replace the binary opposition of mainstream and alternative with a con ception of the economy as a space of difference. In this vision, self consciously alternative economic activities constitute a fraction of the noncapitalist and alternative capitalist activity within a differentiated economic land scape. The economy is understood as being composed of many different processes of production, exchange, own ership, work, remuneration, and consumption without the presumption of necessary relations of dominance and subordination. The alternative economy as economic difference constitutes a new economic ontology whose performative efficacy depends upon and produces an approach to scholarship that is centered on community engagement and action research.

Theorizing Economic Difference

The commonplace view is that economic alternatives are tiny islands awash in a sea of capitalism. To the extent that difference is recognized – alternatives to capitalism or alternative forms of capitalism – this difference is seen as contained within or conceptually subsumed to capit alism as such. The first problem facing theorists of eco nomic difference is how to define capitalism so that it is no longer seen as that which contains and subsumes difference.

Following Resnick and Wolff, Gibson Graham turns to a close reading of *Capital* in order to more precisely define capitalism. From her perspective, what dis tinguishes capitalism is the specific way in which surplus labor is produced and appropriated, or what Resnick and Wolff term the 'class process'. Capitalism involves the use of free wage labor in the production of goods and ser vices, usually for a market. Wage laborers produce a surplus that is appropriated by nonproducers – a sole proprietor, or the board of directors of a capitalist firm – who distribute this surplus in ways that they wish, though they are usually constrained to direct much of it to re producing the firm and its exploitative class process.

This minimalist definition of capitalism carries with it the implication that particular capitalist firms might operate under a variety of conditions. A capitalist enterprise may or may not accumulate capital (in other words, invest appropriated surplus in expanding plant, equipment, and workforce), may or may not own the means of production (plant and equipment may be leased or borrowed), may or may not be in a dominant position with respect to the wage laborers employed, and may or may not strive to maximize profits or market share. It is the minimal definition of capitalism that allows us to see capitalisms in all of their specificity and particularity at the level of the enterprise.

With capitalism reduced from a systemic entity coextensive with the social space to a type of enterprise scattered (or perhaps concentrated) in a larger economic landscape, it is easier to see the other ways in which goods and services are produced and exchanged. In the United States, for example, it is possible to find instances of goods and services publicly provisioned, or produced for exchange by people who labor without freedom of contract in what might be seen as a slave class process (prisoners, members of the armed forces, indentured migrants, and children). Likewise, as feminists inside and outside of geography have argued and demonstrated for quite some time, household based, noncommodity production of goods and services accounts for 30–50% of total economic activity in both rich and poor countries.

While it is not difficult to find examples of exploitative class relations – capitalist, feudal, and slave – in every country of the world, it is also possible via class analysis to find instances of nonexploitative class relations in which the people involved in producing surplus labor are also its appropriators. Self employment (which Marx referred to as the ancient class process) is ubiquitous in both rich and poor countries, and worker cooperatives (where surplus is communally appropriated) operate in market and nonmarket contexts throughout the world.

Just as productive activities can be read for difference in class terms, it is possible to theorize exchange and labor/compensation occurring in a variety of ways. While many goods and services are produced for exchange on markets, they are also allocated in other ways – via gifts, state transfers, barter, theft, and so on. Similarly, many people in a society may be paid a wage for some of their labors, but likely do other work that is unpaid or alternatively compensated – for example, by payment in kind, or in a reciprocal labor exchange. This heterogeneity of exchange and compensation is a fundamental condition in economies around the world and through time.

The understanding of economy as a space of difference is partially captured in the open ended conceptualization of a diverse economy in **Figure 1**.

This diagram could be expanded to differentiate other aspects of the economy such as property (private, alternative private, and common property), forms of finance, and other dimensions of difference. It can be used for inventorying economic activity in a particular locality or thinking about the ways different elements of the diverse economy are in fact connected to one another in any given practice or social location. For instance, the Community Economies Collective documented the story of a worker cooperative, Collective Copies, which was founded after the original owner decided to close shop

Transactions	Labor	Enterprise
Market	Wage	Capitalist
Alternative market Sale of public goods Ethical ‘fair-trade’ markets Local trading systems Alternative currencies Underground market Coop exchange Barter Informal market	Alternative paid Self-employed Cooperative Indentured Reciprocal labor In kind Work for welfare	Alternative capitalist State enterprise Green capitalist Socially responsible firm Nonprofit
Nonmarket Household flows Gift giving Indigenous exchange State allocations State appropriations Gleaning Hunting, fishing, gathering Theft and poaching	Unpaid Housework Family care Neighborhood work Volunteer Self-provisioning labor Slave labor	Noncapitalist Communal Independent Feudal Slave

Figure 1 A diverse economy. Note: The figure should be read down the columns rather than across the rows. Source: Gibson-Graham, J. K. (2006). *A Postcapitalist Politics*, p 71. Minneapolis, MN: University of Minnesota Press.

rather than allow unionization. The capital used to ac quire the business was donated in the form of outright gifts and a no interest loan from customers and com munity members. Just as primitive accumulation in the form of dispossession has historically given rise to cap italist enterprises, gifting could be seen as providing a source of primitive accumulation for this cooperative enterprise (and perhaps others).

The story of Collective Copies links the success of this enterprise to other elements of the diverse economy, tracing the formation of a nascent community economy in which social interdependence is acknowledged and constructed. To expand this conception of the economy as a negotiated space of relation among diverse elements, it is necessary to conceptualize a flat spatial ontology where differences are provisionally equivalent rather than hierarchically arrayed.

Spatialities of Alterity and Difference

For those who focus on neoliberalism as well as those who are interested in the self consciously alternative economy, the economy is generally structured by a hierarchical spatial ontology in which the local is nested within the regional, national, and global scales. Viewed as containers, activities on a global scale – international fi nancial markets, for example – are assumed to have more determining power than local projects that appear to be contained – a LETS or a worker cooperative. In this way, a hierarchically ordered space effectively affirms the dominance of (global) capitalism while consigning eco nomic experiments to relative powerlessness.

The metonymical pairing of global capitalism and local alternatives structures our understanding of eco nomic space even when an alternative economic practice becomes global in scope. Alter trade (or fair trade), in spite of its expanding presence in the global marketplace, is dismissed by its critics as small in relation to global trade as a whole (as, of course, are many sectors of international trade) and vulnerable to competition and cooptation. What is remarkable about this depiction is that alter trade could just as easily be represented as a powerful innovation, one that has injected an ethical sensibility into trade that did not exist 20 years ago. Alter trade is energized by an ethical dynamic of growth (rather than by a structural dynamic of competition and increasing market share) that works against cooptation, draws increasing numbers of products into fair trade marketing, and links together more and more people. Perhaps in order to see the potential of alter trade, what is required is a different spatial ontology that does not presume in advance the connection between scale, size, and power.

Recently Marston, Jones, and Woodward have pro posed a flat spatial ontology based on the site as an un folding materiality that constructs and reconstructs space through its often uneven and temporary connection to other (distant) sites. In this conception there are no spatial categories or containers that prearrange the world into ordered spaces. Another flat spatial ontology is offered by Gibson Graham, whose feminist political imaginary is premised on a geography of ubiquity. Be cause women are everywhere, local and household based feminist projects can be globally transformative; because diverse economies are everywhere, the projects of local economic activists can have world changing effects. Linked semiotically rather than organizationally, these projects have the potential to configure a place based global movement for economic transformation.

Flat spatial ontologies have several implications for theorizing and enacting economic difference. First, it becomes easier to see economies as self organizing spaces of contingent relationality where there is no presumption of dominance and subordination (though these will, of course, be found to exist in particular settings). When capitalist class relations are no longer to be regarded as necessarily dominant, it becomes more difficult to im agine that other social sites and processes (e.g., house holds) are bound to the task of capitalist reproduction or that economic alternatives are awash in a capitalist sea. Second, a flat spatial ontology allows us to see economic diversity as globally dispersed, while at the same time creating potential connections among disparate locations and processes. In the flat space of economic difference, economic geographers interested in contributing to al ternative economies might play a constructive role in translating experiments from one location to another, formalizing lessons learned from experiments in one place for other places, and working imaginatively with individuals, communities, and regions to produce and disseminate economic innovation. Given the ubiquity of potential sites for these sorts of academic interventions – households, enterprises, communities, and regions – they could conceivably be conducted on a global scale. But in conceptualizing this politics of research and the future direction of scholarly interest in alternative economies, it becomes crucial to return to the concept of performa tivity and the way that it differs from just electing to think differently.

Performativity, Economic Difference, and the Politics of Research

Colin William's recent work on differentiating the pro cess of economic exchange poses a significant challenge to the received wisdom that market exchange/com modification has displaced, or soon will displace, all other forms of exchange. As William notes, the commodification thesis circulates as a widely accepted idea, and yet it remains, essentially, an assertion that has yet to be theoretically or empirically substantiated. His own empirical work in the United Kingdom demonstrates that even in advanced industrial countries (and postindustrial ones), barter, other forms of nonmarket exchange, self provisioning, and mutual aid are widely practiced and may indeed be more integral to the lives of the middle and upper classes than they are for people on the social and economic margins.

The crucial insight from Williams' work is that the persuasive power of the commodification thesis is itself a testament to operations of performativity. Something that is as ubiquitous as nonmarket exchange remains hidden only because it is configured discursively as subordinate. Interestingly, when it is revealed to be ever present it is dismissed as unimportant, dependent upon capitalism, or (from a progressive standpoint) not subversive enough. The perceived expansiveness and dominance of market exchange is performatively constituted even (perhaps especially) by those who are critical of it.

Performativity theory is not simply an invitation to think differently; it also reveals the extent to which the discursive limits the conceivable and, by extension, the doable. In the open and indeterminate space of a diverse economy, the space of academic engagement is correspondingly enlarged; the question becomes how do we understand and engage with economic difference when it is no longer positioned as the subordinate term within a binary opposition? One answer might be to see the space of economic difference as a space of self conscious and unconscious experimentation in becoming, where marginality or dominance, success or failure, cannot be known in advance. As Lincoln points out in his case study of worker owned (cooperative) enterprises, they succeed or fail as ventures for definite reasons just as capitalist enterprises succeed or fail. The vulnerability of cooperative firms is not unique to their form, but reflects a general element of economic risk that is heightened or attenuated in particular situations.

Viewing the economy as a space of experimentation/becoming has another important implication in that the recognition calls forth and depends upon a new form of activist scholarship. The goal of this scholarship is to examine economic practices that are potentially valuable but discursively subordinated, bringing them to light and engaging with the actors to build or strengthen alternative economies. These scholarly interventions highlight the ethics and politics of language and representation, recognizing that acting differently requires thinking differently, and that conscious change begins with the recognition of possibility.

Kevin St. Martin's work, for example, focuses on the discursive politics of resource management in the fisheries industry. Fisheries management is currently guided by the assumption that each individual fisher is a self interested utility maximizer. As St. Martin points out, this leads to a regulatory approach – limits on days at sea – that generates more risk taking on the part of vessels and crews and produces the very behavior it presumes to control. Working closely with fishing communities themselves, St. Martin has put forward an alternative representation of fishers as a community engaged with and invested in a common resource. It is common practice, for example, in the fisheries of the northeastern United States for fishers to share information (map plots) about bottom type and fish habitat; the controlled dissemination of this information constitutes a form of community based resource management that offers, St. Martin argues, an alternative entry point to conservation. What is powerful about St. Martin's work is that it is attempting to initiate a new approach to resource management which includes the community of fishers as potential allies rather than adversaries. His approach to an alternative policy regime depends upon seeing the possibilities in what is already a commonplace, everyday activity.

In another example of the experimental/performative approach to research, the Community Economies Collective is collaboratively drafting a protocol for self study with community based enterprises to document their histories and develop new, more appropriate metrics of success. One observation that has emerged from this work is that community enterprises often change their mandates over time. The Alliance to Develop Power (formerly the Anti Displacement Project (ADP)) started as an organization to preserve affordable housing in the Pioneer Valley of western Massachusetts. At present, the organization owns housing worth $45 million dollars, making it one of the largest property owners in the area. Several years ago the ADP drew on their energetic membership and financial power to start a cooperative enterprise – United Landscaping – that provides worker control and good jobs for ADP members as well as maintenance services for ADP properties. More recently, the ADP has worked with a local union to open a worker education and advocacy center that offers ESOL, adult basic education, campaigns for fair wages and wage restitution, financial services for the unbanked, and many other services to marginalized workers.

In their ongoing efforts to document the successes and setbacks of the ADP and organizations like it, the Community Economies Collective sees its work as a vehicle for validating and disseminating effective strategies of economic activism and community engagement. Via the existing infrastructure of academic and activist channels, the ADP's experiences are readily communicated to scholars and activists worldwide, contributing to the performance of alternative economies in far flung places.

Conclusion

St. Martin and other members of the Community Economies Collective explore the alternative economy as a space of experimentation and difference through a re search process that involves them collaboratively with people in place. Far from being idealist, the performative dimension of these action research projects makes use of what already exists, hidden in plain sight, in order to develop political, ethical, and organizational potential. The performative theory of economic difference is not blind to the realities of power, but emphasizes the power of representation and research.

The examples of Collective Copies, the ADP, and community based fisheries management suggest that the future of alternative economies research within economic geography might revolve around reconceptualizing re search as a process of performatively enacting com munity economies. Activist researchers could engage in disseminating and replicating such experiments in forming community economies, recognizing the powerful role these efforts might play in demonstrating how an other economy might be possible.

Ironically, we might turn to the success of the neoliberal project itself for inspiration. As David Harvey points out, the basic tenets of neoliberalism articulated by Friedrich Hayek operated on the margins of economic orthodoxy for decades until the economic crisis of the 1970s gave pro ponents an opportunity to widely disseminate them. In the same way, perhaps identifying alternative economic prac tices, conceived as part of a larger field of economic dif ference, might become a central part of an activist research agenda for economic geographers. Nearly 40 years after the economic crisis that propelled neoliberalism into a position of discursive dominance, it is neoliberalism that is now widely perceived to be in crisis. The flat rejection of export oriented development and state devolution in Latin America must be seen in the larger context of the failure of the capitalist development project as a whole. When this failure is combined with a deepening awareness of global environmental contradictions, we may find ourselves in an ideal context in which to engage in a performative schol arship and politics of alternative economic development. The expanding presence of the World Social Forum and national forums like the USSF cannot only be read as re flecting a tangible hunger for alternative economies but as a practical context for activists and activist scholars to engage in experimentation and dissemination.

Given the widespread recognition of the failure of development as usual and the ubiquity of alternative economic practices foregrounded in the theory of economic difference, the formation of community econ omies becomes a process that is potentially global in scale while being particular and local in practice. This ap proach to constructing economic alternatives offers no guarantees; it simply foregrounds possibilities, with the understanding that a wider sense of possibility is a first step toward enacting alternative economies. It is the connection between a performative concept of know ledge, an ontology of economic difference, and a trans formative politics of research that will define this area of study and its promise in the future.

See also: Economic Geography; Economy, Informal; Enterprise Discourse; Participatory Action Research; Performance, Research as; Performativity; Poststructuralism/Poststructuralist Geographies.

Further Reading

Community Economies Collective (2001). Imagining and enacting noncapitalist futures. *Socialist Review* 28(3/4), 93 135.

Crewe, L., Gregson, N. and Brooks, K. (2003). Alternative retail spaces. In Leyshon, A., Lee, R. & Williams, C. (eds.) *Alternative Economic Spaces*, pp 55 73. London: Sage.

Gibson Graham, J. K. (1996/2006). *The End of Capitalism (as we knew it): A Feminist Critique of Political Economy*. Minneapolis, MN: University of Minnesota Press.

Gibson Graham, J. K. (2006). *A Postcapitalist Politics*. Minneapolis, MN: University of Minnesota Press.

Gibson Graham, J. K. (2008). Diverse economies: Performative practices for 'Other Worlds'. *Progress in Human Geography* 32(5), forthcoming.

Gibson Graham, J. K., Resnick, S. and Wolff, R. (2001). *Representing Class: Essays in Postmodern Marxism*. Durham, NC: Duke University Press.

Harvey, D. (2005). *A Brief History of Neoliberalism*. Oxford: Oxford University Press.

Lee, R. and Leyshon, A. (2003). Conclusions: Re making geographies and the construction of 'spaces of hope'. In Leyshon, A., Lee, R. & Williams, C. (eds.) *Alternative Economic Spaces*, pp 93 198. London: Sage.

Leyshon, A., Lee, R. and Williams, C. (2003). *Alternative Economic Spaces*. London: Sage.

Lincoln, A. (2003). Alternative work spaces. In Leyshon, A., Lee, R. & Williams, C. (eds.) *Alternative Economic Spaces*, pp 107 128. London: Sage.

Marston, S., Jones, J. P. III. and Woodward, K. (2005). Human geography without scale. *Transactions of the Institute of British Geographers* 30(4), 416 432.

Marx, K. (1977). *Capital*, vol. 1. New York: Random House. Trans. B. Fowkes.

North, P. (2007). *Money and Liberation: The Micropolitics of Alternative Currency Movements*. Minneapolis, MN: University of Minnesota Press.

Pavlovskaya, M. (2004). Other transitions: Multiple economies of households in the 1990s. *Annals of the Association of American Geographers* 94(2), 329 351.

Resnick, S. and Wolff, R. (1987). *Knowledge and Class*. Chicago, IL: University of Chicago Press.

St. Martin, K. (2001). Making space for community resource management in fisheries. *Annals of the Association of American Geographers* 91(1), 122 142.

Relevant Websites

http://autonomista.org
Argentina Autonomista Project.
http://www.ethicalmarkets.com
Ethical Markets a gateway to cleaner, greener twenty first century economies.
http://www.commoner.org.uk
The Commoner a web journal.
http://www.communityeconomies.org
The Community Economies project a place where new visions of community and economy can be theorized, discussed, represented, and enacted.
http://www.schumachersociety.org
The E. F. Schumacher Society linking people, land, and community by building local economies.

Economies, Borderland

M. Tykkyläinen, University of Joensuu, Joensuu, Finland

Glossary

Antiterrorism Measures Comprise the US government's security-related initiatives and regulations affecting the control of transport and foreign trade after the 9/11 attack.

Devolutionary Process The process whereby regions in a state gain political power and growing autonomy at the expense of the central government.

Ethnonationalism A movement among an ethnic group who consider themselves a distinct nation with a right to autonomy or independence in their own territory.

Gateway A borderland or set of borderlands which occupy an important position in trade and transport and the smoothing of institutional differences between larger economic areas.

Institutions Formal institutions are constitutions, rules, laws, property rights, and organizations, and informal or tacit institutions are individual habits, customs, codes of conduct, group routines, and social norms, values, sanctions, taboos, and traditions.

Maquiladora Manufacturing industries located on the Mexican side of the US–Mexican border which utilize the low-cost labor and premises available there and export their production to US markets.

R&D Inputs Research and development inputs constitute labor and funding spent by companies and the public sector on developing new products and production processes to improve the competitiveness of industry.

Shatterbelt A strategically located region consisting of borderlands and states that is under the jurisdiction of a number of conflicting states and groups and is caught between the interests of powers and power coalitions.

Concepts

Although borders are actually social constructs among people, social processes are reflected in the geographical space through the formation of regions. Thus, the world is divided into the layers of regions representing different geographical levels. The geographical system of regions consists of various sociospatial constructs, such as the geographical mosaic that has resulted from administrative and economic decisions. In spite of the talk of a 'borderless' world, various types of regional divisions exist based on administration, ethnicity, economic performance, and welfare, and borderlands are mani festations of these divisions. Borderlands based on nation states are central and generally known in our perception of what the regions of the world are (**Figure 1**). This is the geographical division which is represented in everything from school atlases to government policy documents.

The division is far from firm and stable, however. Continuous border forming processes produced by social processes are reshaping the regional systems, and the media suggest that areas of conflict and turmoil are liable to emerge rapidly, although many borderlands are the results of longer, historical processes based on economic performance, ethnic composition, and the evolution and decline of political, economic, and social organs. Bor derlands are areas on both sides of the edges of territories created by powers, social forces, and economic activities.

Traditionally, borders show the territorial dimension of state power. The Peace of Westphalia in 1648 is re garded as a landmark of the beginning of the existence of the modern state. Since then the principal actors in international relations and in setting up legislation and the terms of trade have been sovereign states. The world was divided into states, and power has been exercised within their borders. Only a few people would call state formation into question, because it is related to our se curity and patriotism, but there are some who speak of a global or cosmopolitan world, and some stress the im portance of regions as the basic units of spatial organ ization. But, in principle, borderlands are commonly considered in a national context.

When emphasizing economic and social interaction, the thesis of the end of the nation state has found widespread acceptance in the minds of citizens in this era of the information society and globalization. Borders have become more related to access to communication

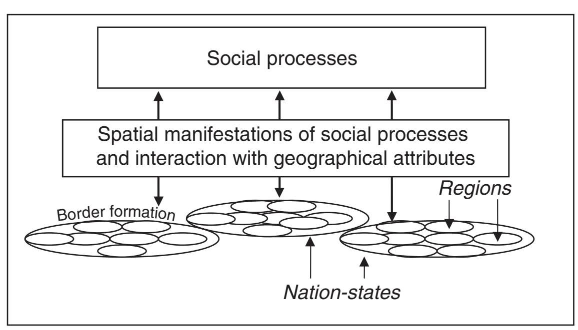

Figure 1 Formation of regions and borders.

systems and economic performance in the advanced parts of the world. As researchers have shown, there are inclusions and exclusions in virtual worlds which are coming more and more to determine individuals' social status. People have uneven opportunities and resources for entering this cyberworld which is not measured in geographical coordinates. Hence, theoretically and increasingly in practice, borderlands can manifest themselves on various spatial scales and in various spaces. In our vocabulary, however, the term borderland still usually refers to areas on the fringe of a national territory.

States and federations have interests in developing and controlling borderland economies. In ancient times colonization was a common development policy in borderlands, and later many infrastructural and regional policy measures were carried out according to the interests of a state in border issues. The European Union (EU) has large programs for abolishing tension in its borderlands and creating safer borders. On the one hand, it aims to increase cooperation between the borderlands of the EU and the adjacent countries, while on the other hand, it is increasing its control over the borders to prevent illegal interaction. A great difference in wealth and opportunities between the two sides of a border can easily lead to increasing control measures, as has been happening in many places in Africa, on the US–Mexican border, and on the outer border of the EU.

Frontex, the EU agency, was created to coordinate operational cooperation in matters of border security, and another task is to follow developments in research relevant to the control and surveillance of external borders. Frontex strengthens the border control capabilities of the border guard institutions of the member states by informing them of modern technologies and products that are available. The agency also provides information on the tools needed for controlling the outer borders of the EU in an attempt to prevent illegal migration and trade. The impacts of such developments in terms of labor are negative rather than positive in vast, remote borderlands, because the implementation of technical surveillance reduces the need for infrastructure and work.

Similar measures have been set up in the US. During the early 1990s the US had instituted programs to protect its borders and stem the flow of illegal workers over the country's southern border, and operations were set up to control the western end of the US–Mexican border more tightly. In addition to increasing the size of the border patrols, the programs included high tech detection equipment and higher and stronger walls along the border. At the same time, the North American Free Trade Agreement took effect there on 1 January 1994. Hence, even in attempting to create higher economic integration, the strategy can lead to strengthening of border controls related to certain segments of the population and trade.

Theoretical Arguments: A Border as a Hindrance or an Opportunity

A geographical space that is divided into states and regions is based very much on the institutional properties of each region. Borders are concrete manifestations of the division of the geographical space into territorial entities with separate institutional settings. From the viewpoint of economic geography, there are lots of arguments which see the constitution of national borders as a hindrance to development. Economies of scale cannot materialize to the extent that they could in a borderless situation, the transactions themselves are more costly across the border, and production factors are less mobile than in a borderless situation. Growing global tension, such as antiterrorism measures, is also increasing the costs of economic cross border interaction. Hence a border that is sensitive to political tensions is an economic hindrance.

A border constitutes a cultural transition zone or barrier. The separating factor can be language, religion, culture, or ethnicity. Ethnonationalism combined with economic interests can be a fundamental force promoting state formation, and hence the emergence of borderlands. Discussion has evolved around the formation of the world's regions based on cultural and economic differences and the potential clash of civilizations in the long term. The tension between the West and the Islamic world, for instance, manifests itself geographically in a division of the world into macrostructures based on religion and ethnicity. This has been criticized as being too rough a division, however, which does not take into account diversity of opinion and the geography of the diversity of ethnicity and religion.

Many borderlands are arenas for conflicts and continuous tension. Strategic positions and natural resources in particular can make these areas unstable, as is the case in Kurd areas, Iraq, Congo, the Cabinda exclave of Angola and Sierra Leone (**Figure 2**). Such high tension borderlands are problematic from the viewpoint of economic stability and development, and their economies are encumbered with high security costs and the threat of violence. Resources are also often used for devolutionary processes, and proximity to major conflicts outside can also lead to instability and chaos. Lebanon is an example of a country which has suffered greatly due to its borderland status between Israel and the powerful Arab countries and groups. It is obvious that there are strong national interests involved in mastering borderland economies in parallel with political interests.

The border formation process reflects in a concrete way the reorganization of society, and the economy of the borderland changes along with it. The setting up of a border usually creates gradual economic decline in a region. An example of this would be Eastern Finland, which has had a stagnant economy for a century since the

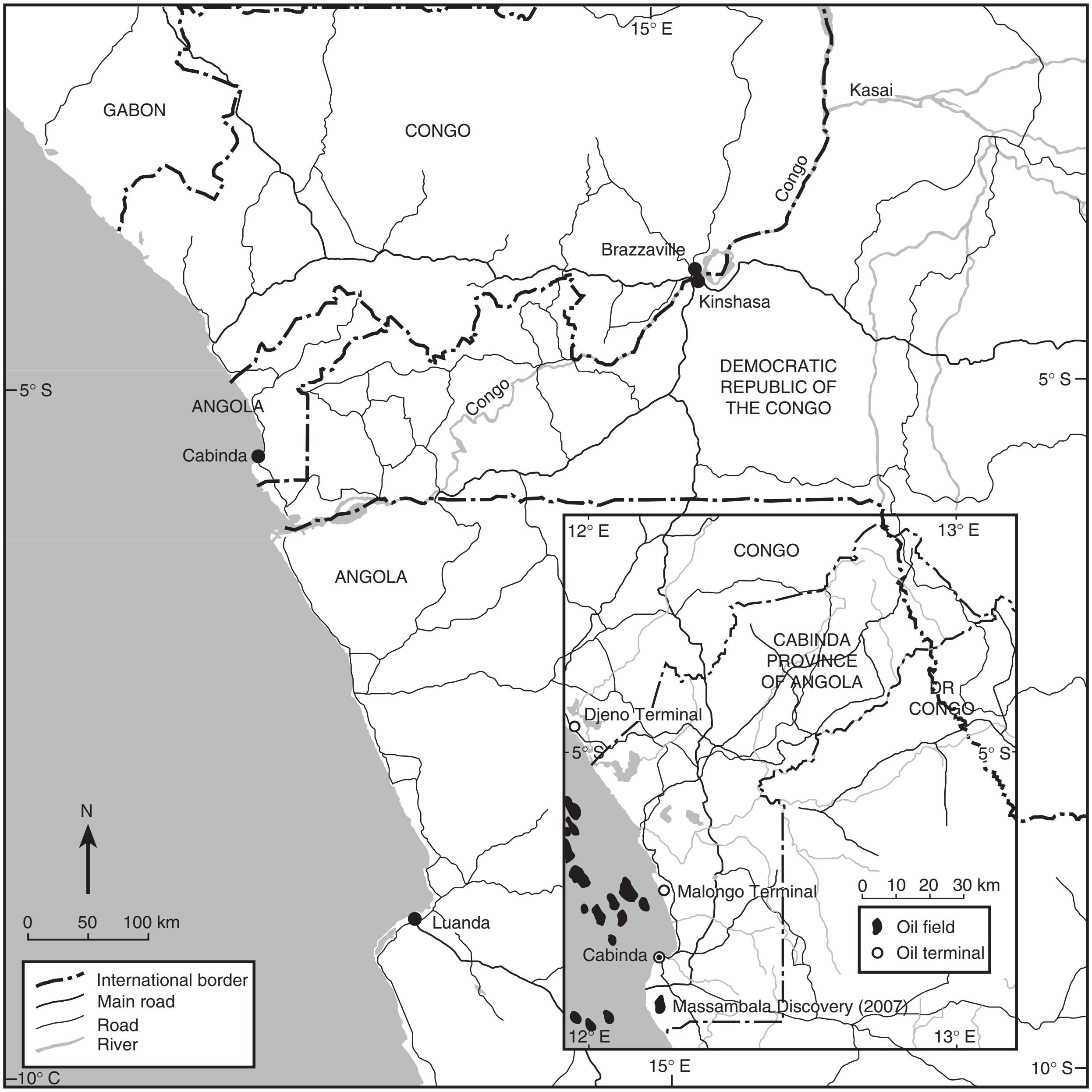

Figure 2 Cabinda, exclave of Angola.

establishment of the border which separated the region from the full impact of the St. Petersburg area in Russia (**Figure 3**). The reconstitution of the border was one of the causes of stagnation. Corresponding stagnant areas can be found in many parts of the world. Post World War II Europe was full of new borderlands, which were de veloping relatively slowly. One of the achievements of European integration is that it has succeeded in abol ishing borders and has hence been able to create larger and more efficient markets in Europe. Nowadays border establishment and state formation processes are taking place in the areas adjacent to Russia, leading to frag mented markets. Many countries would like to be less dependent on Russia, but this can easily lead to the cutting off of economic ties, which creates economic problems. One example of this is the series of energy crises that have emerged in the new independent states adjacent to Russia.

Although the border formation process brings about many regressive results when measured in economic terms, the outcome is not so straightforward. It is also dependent on the scale and viewpoint adopted. The separation of Estonia from the Soviet Union, for instance, is an example of the successful re formation of a bor derland economy. Estonia has transformed itself from a former Soviet agrarian borderland into a modern

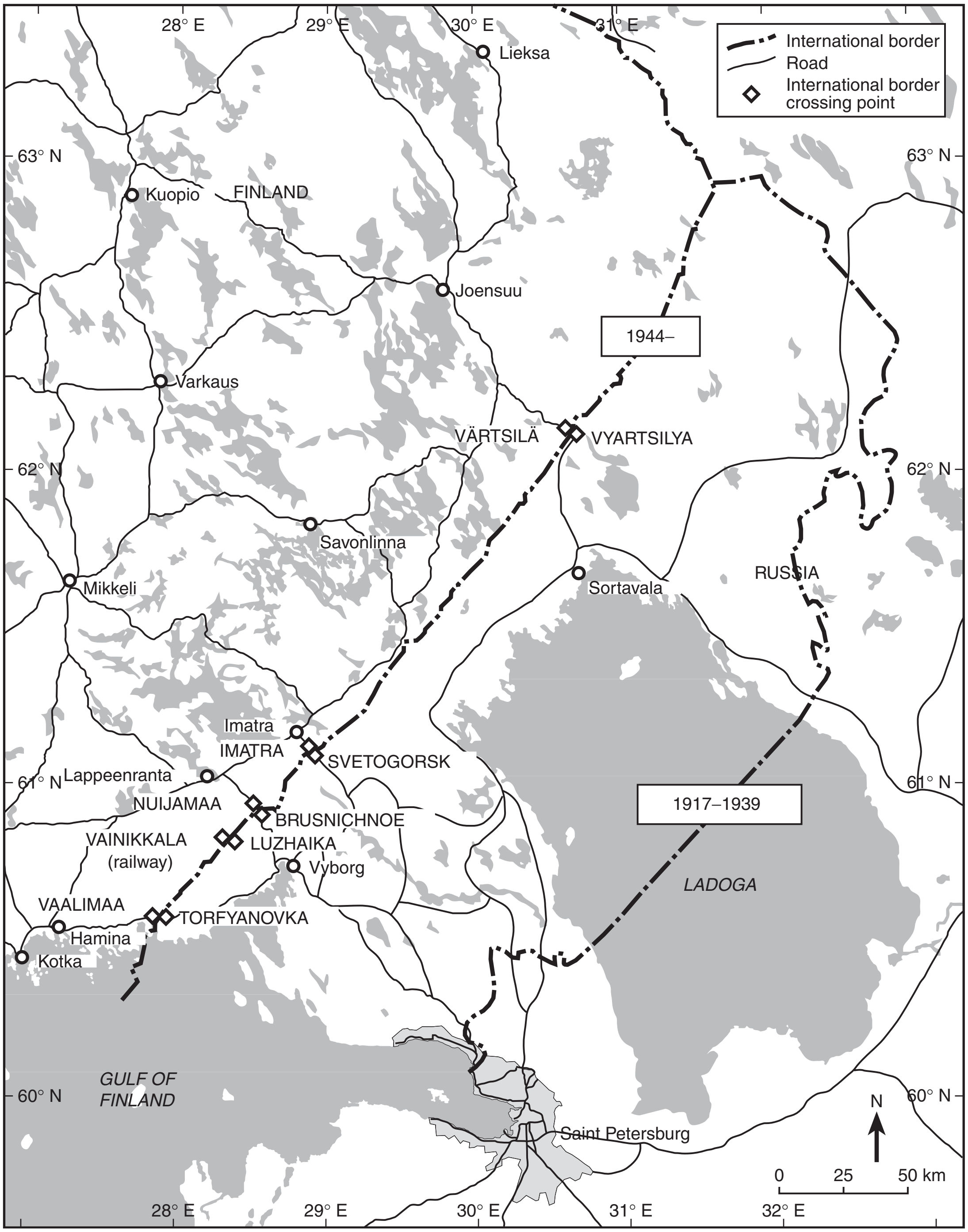

Figure 3 Eastern Finland and Russia.

European industrial state within less than 20 years. Key factors in the transformation were the opportunities created by the division of labor and integration into more advanced economies, cultural ties with market econ omies, and the willingness to undertake radical trans formation of the institutional settings of the economy. This reform also led to a new border formation process, however. The Estonian borderland adjacent to Russia is having to cope with restructuring problems in industry and has ethnic problems related to its original Russian population. The status of a borderland is very much dependent on political decisions and the development of a borderland is steered by changing institutional settings and developing economic conditions. Borderland econ omies are in the last resort the results of the work of social actors: people and companies.

Borderlands and Welfare

The situation of a borderland is usually peripheral and marginal from the economic viewpoint, which easily brings about welfare problems. A borderland is located far away from the core of the country and its economic base may be underdeveloped and one sided. Many borderland economies can be classified into the category of peripheral, stagnant economies. Peripherality may lead to the economic marginalization of an area, or if it has a strategic meaning, the welfare of the region will be strongly supported by the presence of state institutions. The Kola Peninsula is a good example of an area which is too important to remain an underdeveloped borderland due to its strategic location from transport and military viewpoints. One may call such areas sheltered borderland economies, as they are subsidized by the state, which indicates their importance to the state. Fewer areas can utilize their border status to increase welfare to any substantial extent. A more positive welfare effect takes place when the borderland is adjacent to a much wealthier country and economic interaction is possible. Economic impacts are often the results of price differentials in goods and services, which attract customers, and these impacts are substantial when cross border investments are possible. The growth of manufacturing brought about by the lower cost level of the less developed country increases the output of the border area. The maquiladora industry close to the US–Mexican border, for instance, has led to higher and more even incomes in the border states than in the rest of Mexico. Industry of this kind has also been one of starting points for the growth of many borderlands in Asia, such as Guangdong and Johor. These upward transitional borderland economies are merely exceptions, however, and are closely dependent on the more developed region on the other side of the border and on the level of trade regulation.

A borderland can be the area occupied by a national minority, or else there may be various ethnic and linguistic groups along the border. These ethnic issues are often highly visible in former colonies, where borders may have been drawn for strategic reasons, dividing ethnic groups almost randomly from country to country. One example of this is Namibia. This sparsely populated country comprises 28 living languages and several ethnic groups, and includes the 450 km long Caprivi Strip that penetrates from the northeastern inland corner of the country into the middle of Southern Africa in order to guarantee the Namibians a connection with the Zambezi River (**Figure 4**). Caprivi itself is a borderland, and it creates borderlands on both sides of it. The population is a mix of tribes having relatives over the borders. There are also features such as trafficking, smuggling, and health problems related to this border location.

Resource based production is usually over represented in peripheral settings, due to a legacy of concentration on the utilization of natural resources and promotion of the diffusion of settlement. In many cases the borderlands have been target areas for active colonization. Siberia was colonized in order to mobilize all possible resources and to represent Russian, and later Soviet, power in the area. In the Nordic countries, Sweden colonized its vast northern and eastern borderlands efficiently from the seventeenth century onward to make all its corners inhabited and defendable. Hence the political and economic legacy of the past is still visible in the structure of society in many borderlands. The same marking also takes place today. The Caprivi Strip in Namibia was linked efficiently to the main part of the country by the construction of a highway suitable also for an airstrip in the 1990s which is in fact oversized relative to the traffic on it but is an excellent example of an infrastructure intended to show outsiders the readiness of the state to keep the borderland as part of its territory.

In some cases a borderland location has brought some advantages due to economic spillovers. Such borderlands as Johor in Malaysia, the Riau Islands in Indonesia, and Guangdong in China are prosperous although far away from the national centers, and they possess advanced industrial structures and higher incomes. These areas are adjacent to even more advanced areas, however, such as Singapore and the former British colony of Hong Kong (**Figure 5**). In such cases the economy of a borderland develops so that the less advanced border area acquires more mature economic activities such as manufacturing and assembly work from the more advanced area. Industrialization takes place when companies relocate activities to the other side of the border in order to take advantage of lower land rents and labor costs. These developments have proven to be highly changeable, however, as new situations develop and the advantages do not remain the same.

Moreover, in more industrialized areas there is usually a certain daily crossing of labor over the border. This has been the case on the Polish–German border and the Singapore–Malaysia border, for instance. Crossing the border for trading and maintaining social relations is common under various conditions all over the world, but great adaptability is needed on the part of the local population in borderlands affected by tensions and potential conflicts, due to control and regulations.

Impacts of the Closing and Opening Up of Borders

As many examples show, the impacts of closing a border lead in most circumstances to a shrinking of the

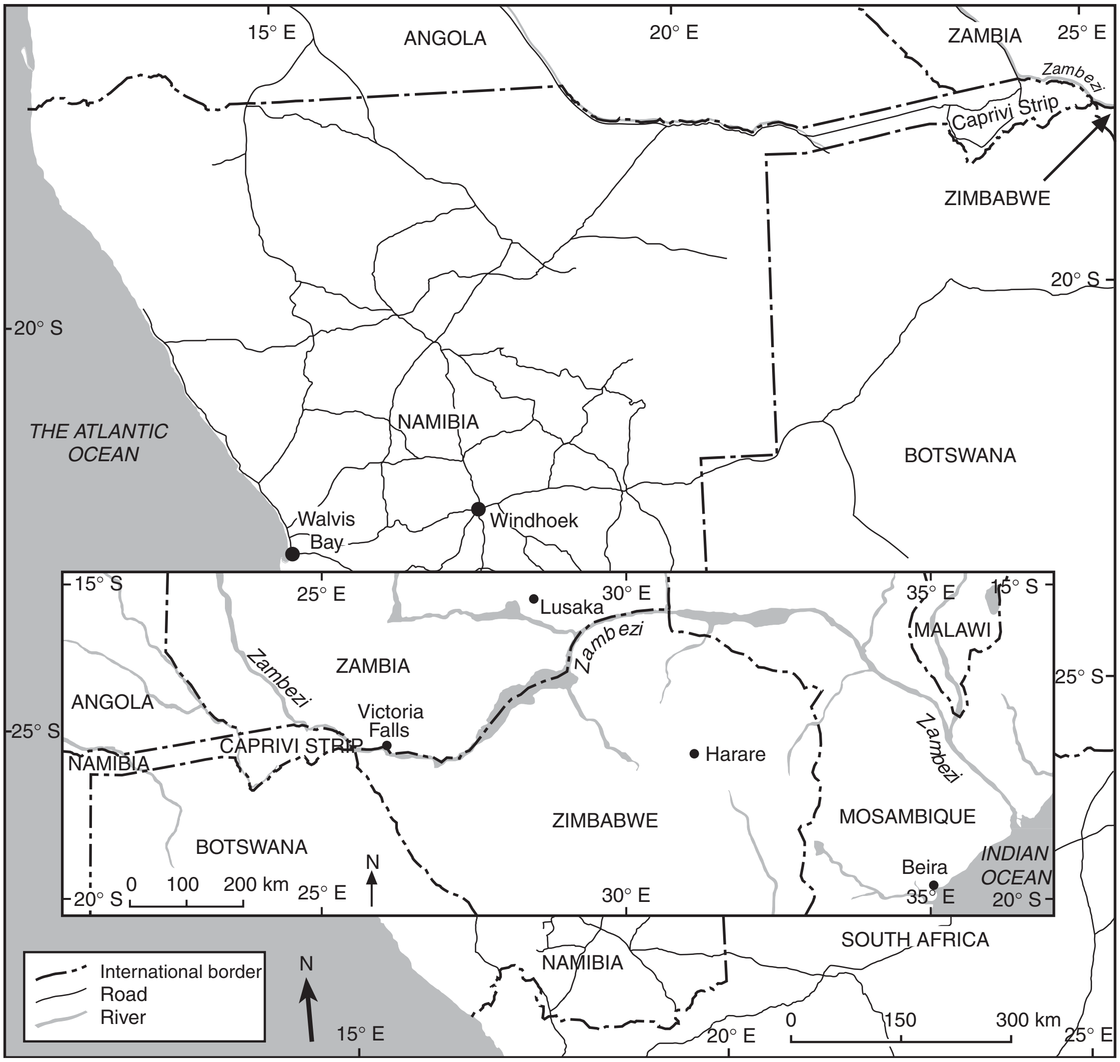

Figure 4 Caprivi strip in Namibia.

borderland economies. Various institutional settings, tariffs, and regulations lead to a situation in which there are less opportunities to use economies of scale and the division of labor. Thus economies become more closed, companies are smaller, and there is less competition. The border becomes a line splitting the market area. Thus integration has been seen in many parts of the world as a panacea for improving the competitiveness of inter regional systems. The EU is a good example of attempts to improve competitiveness by the formation of larger market areas.

The aim of many trade organizations is to abolish trade barriers and protectionist economic regulation. The liberalization of trade and the movement of production factors lead to larger market areas with increasing use of economies of scale, and thus create opportunities for industries to specialize and expand. However, fierce competition leads to closures, mergers, and increasing efficiency, and this may lead to a shift of industries to areas over the border, where it is possible to make use of low cost labor and infrastructure. This means, however, that former borderlands become parts of larger economic markets, and the net effect of the opening up of a border is thus positive.

The enlargement of the EU brought about consider able restructuring in a large number of former East Central European national borderlands. The border areas of Lithuania, Latvia, the eastern and southern parts of Poland, and the northern parts of the Czech Republic and Slovakia are considered to be borderlands that enjoy agglomeration economies but need significant structural adjustment in order to deal with increased competition.

The borderlands in northwestern Poland and most of the borderlands of the Czech Republic, Slovakia, Hun gary, Romania, and Bulgaria are considered to have high development potential due to their favorable

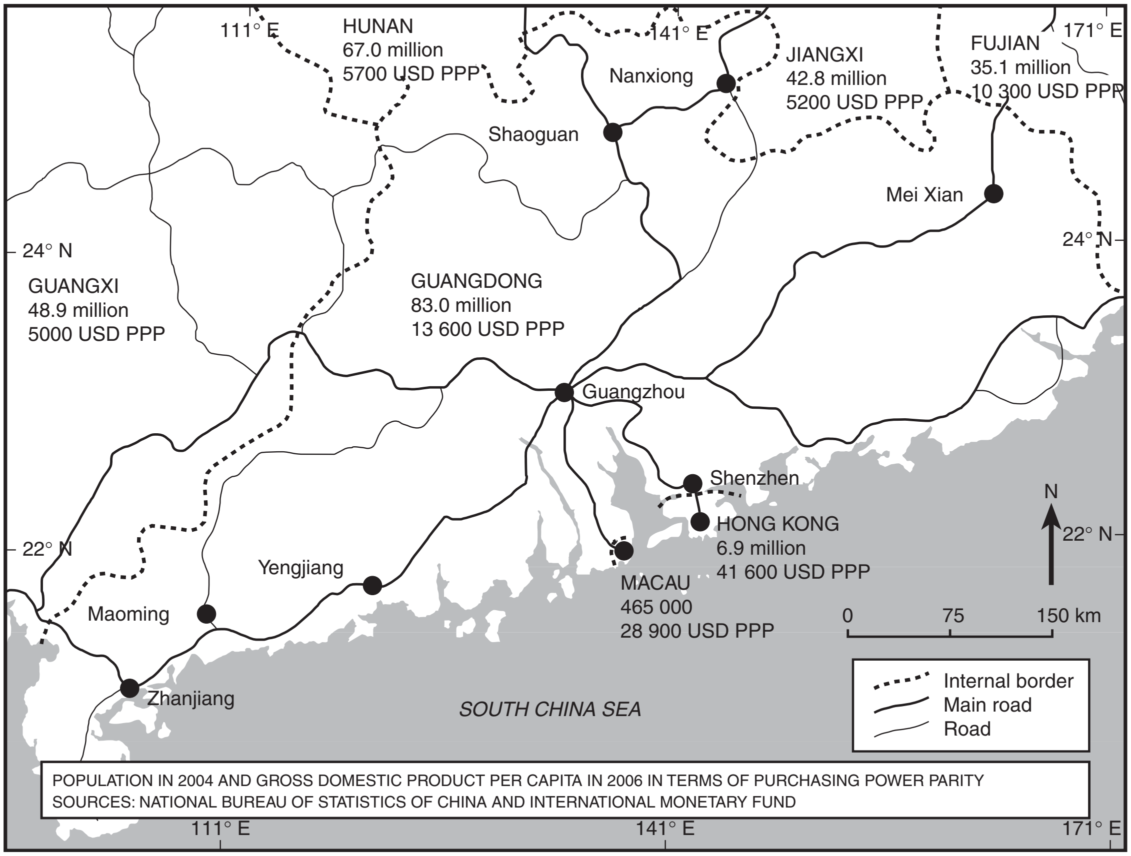

Figure 5 Hong Kong, Macau and Guangdong.

geographical position combined with low economic performance. The findings indicate that integration in Europe is still associated with significant differences between borderlands, and that intensified integration may lead to both booms and busts with respect to border economies. The serious heterogeneity in development processes is a reminder of the importance of initial conditions and geography in shaping economic outcomes, and also an indication of the limited ability of market forces to promote spatial cohesion. The impact of integration on borderland economies is dependent on the results of policy interventions and market dynamics. The enlargement of the EU will for several decades constitute an interesting laboratory for examining the impacts of the opening up of borderland economies.

Economic Projects for Boosting Borderland Economies

Prosperous economic development in a borderland is very much dependent on the stability of the institutional settings and the trustworthiness of the environment. The division of the world into various types of geopolitical regions, such as shatterbelts and gateways, shows the strong dependency of borderlands on international politics and military strategies. Shatterbelts are not areas where economic growth takes place easily, in spite of international agreements, peacekeeping, and aid and support programs. In the worst cases many supranational organizations have had to help borderlands cope with violent crises. This stabilizing task is linked to the policies and duties of the United Nations and other supranational organizations. The aid packages intended to reconstruct borderlands that have suffered due to wars are examples of the programs needed. There is often an intense military presence in borderlands, and power struggles can provoke instability in institutional development and hence lead to poorer economic development.

Due to the devastating world wars, Europe has a long post war legacy of cross border cooperation. In the European context, a Euroregion is a form of transnational cooperation between two or more territories located in different countries. The great majority of these lie within the current borders of the EU. Euroregions are limited to the competencies of the local and regional authorities which constitute them, and they are usually arranged in order to promote common interests across the border and cooperate for the common good of the border populations. There are more than 60 Euroregions in Europe, and the history of this cooperation goes back to 1958, when the first of these was set up, constituting a border land between Germany and the Netherlands. Euroregions are independent organs related to the EU, but many projects in these regions are supported by the Union.

To achieve more balanced spatial development in the EU, cohesion and convergence are required, which puts emphasis on the development of borderlands to create greater stability and integration. Various programs and funding instruments have been in effect within its inner and outer borderlands. Economic revitalization has played a central role in the measures in order to enhance economic competitiveness and sustainable development in borderlands.

The European Commission began in 2003 to set up new programs to boost cooperation across the EU's border with its neighbors. The neighborhood and partnership instruments aim at contributing to economic and social development on both sides of the border, promoting economic and social development in the borderlands in fields such as environment, public health, and the fight against organized crime, ensuring efficient and secure borders, and promoting local, grassroots type actions. The specific priorities and the actions to be supported within this schema are decided upon by the participating countries themselves. This instrument offers solutions to socioeconomic problems in the external borderlands of the EU intended to stimulate outer cross border co operation, good neighborliness, and security.

Conditions of Networking in Borderlands

Interaction across borders has created high expectations for the utilization of border locations, in the sense that factors such as cultural connections, labor, and social relations could be used for promoting economic activ ities. Economic actors become embedded in borderlands in the sense that they absorb and are constrained by the economic activities, cultures, and social dynamics that are embedded in those places. But what are the advantages of borderlands in this setting? One concrete example is the way in which global and national firms may take advan tage of benefits provided by a state or programs for en terprises and local actors intended to establish subcontracting, subsidiary, or R&D operations in a bor derland. Moreover, the location or anchoring down of new firms in border locations may generate a new local or regional network of economic and social relations across the border, involving existing firms as well as attracting new ones. This is what happens in cases when there are a bundle of other factors which make economic activities

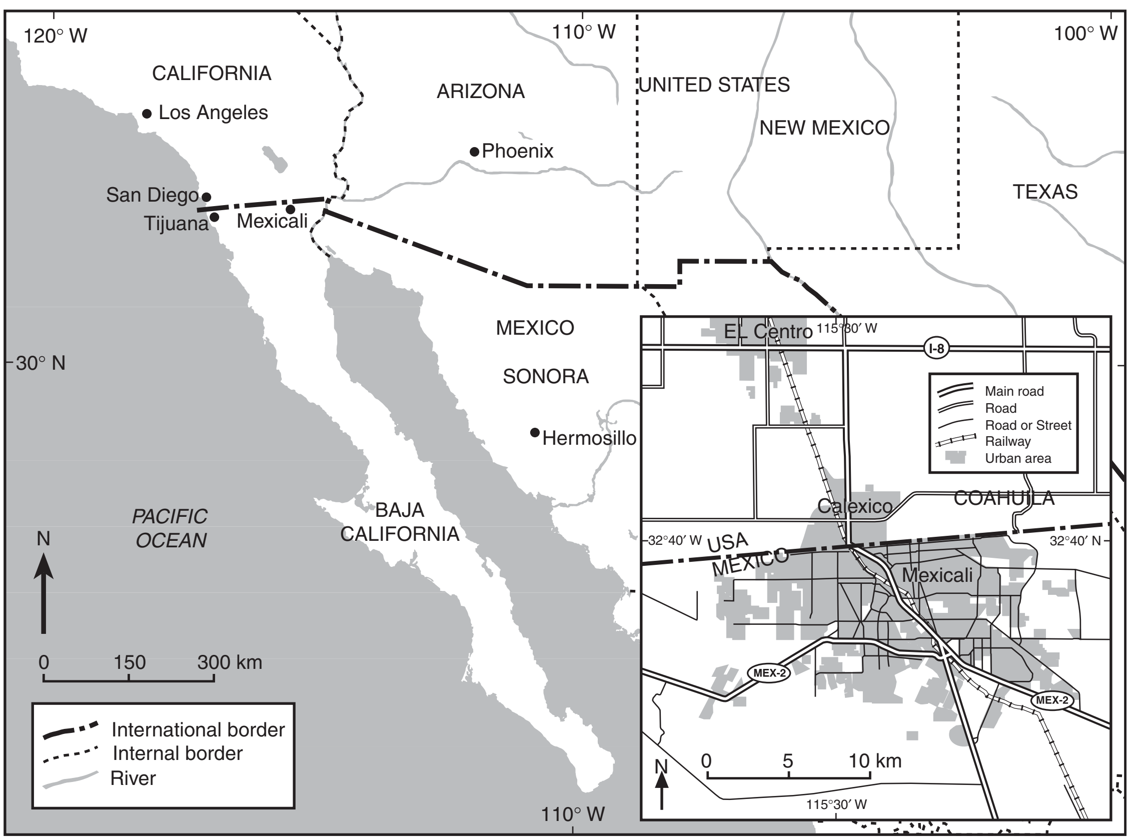

Figure 6 Mexico United States border.

profitable. A strong combination of competitive factors helps to consolidate the transborder embeddedness of industry, which in turn is closely linked to particular regional cultures. One factor is that the hybrid culture of a borderland can prove to be a competitive advantage for a company. Mexicali is an example of a city which has benefited from its hybrid culture. It is the capital of the state of Baja California, Mexico, and situated on the border of the US state of California (**Figure 6**). The residents of Mexicali are from diverse European, Arab, Asian and African backgrounds. It attracts multinational companies to utilize qualified labor, infrastructure, its favourable business environment, and its location on the Californian border. There is social interaction, and assets are often related to culture and languages, which pro mote business in borderlands. In this age of the global izing economy, it seems that many people would wish to do away with a less competitive provincial cultural fac tor, in that companies would prefer to combine yielding cultural assets and space out production within their global networks.

If institutional settings diverge greatly and cultural ties are more or less imagined, the result of this con solidation of joint ventures remains poor. Hence there are insurmountable hindrances which make the utilization of the endowment of difference in the composition and price of production factors less attractive. Adaptation to a new situation and the setting up of economic cooperation takes time and is restricted by supraregional institutional constraints such as differences in laws, rules, and eco nomic regulation.

Borderland economies are heterogeneous, which means that the spatial scales of production must be taken into account in interpretations. Local and transborder markets are sufficient niches for production for a shop keeper or small entrepreneur, but for companies which produce high volumes of consumer products global networking with suppliers and customers is important. Global connections are important for companies which make abundant use of R&D inputs and are innovative, but such an advanced high tech environment is missing from most borderlands. Thus both local and translocal relations are crucial for development and for the globally acknowledged performance of borderlands and the actors involved. Hence, links to research and development and innovation systems created by the state and private actors are just as important for borderland economies as for any other area. Economic development in borderlands is hence a process of glocal network building, creating and maintaining relationships of trust on various interrelated geographical scales.

See also: Borderlands; Competitiveness; Economies, Branch Plant; Embeddedness; Export Processing Zones; Industrial Location; Industrialization; Institutionalism/Institutional Geographies; Local Economic Development; Local Economic Development, Politics of; Networks; Region; Regional Planning and Development Theories; Regional Production Networks; Resource Industries; Territory and Territoriality.

Further Reading

Amin, A. (1999). An institutionalist perspective on regional economic development. *International Journal of Urban and Regional Research* 23(2), 365 378.

Anderson, J. and Wever, E. (2003). Borders, border regions and economic integration: One World, ready or not. *Journal of Borderland Studies* 18(1), 27 38.

Bathelt, H., Malmberg, A. and Maskell, P. (2004). Clusters and knowledge: Local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography* 28(1), 31 56.

Cohen, S. B. (2005). The Eurasian convergence zone: Gateway or shatterbelt? *Eurasian Geography and Economics* 46(1), 1 22.

Hill, F. and Gaddy, C. G. (2003). *The Siberian curse: How communist planners left Russia out in the cold*. Washington, D.C.: Brookings Institution Press.

Ho, K. C. and So, A. (1997). Semi periphery and borderland integration: Singapore and Hong Kong experiences. *Political Geography* 16(3), 241 259.

Holmes, L. (1998). Europe's changing boundaries and the "clash of civilisations" thesis. In Murray, P. & Holmes, L. (eds.) *Europe: Rethinking the boundaries*, pp 51 80. Aldershot: Ashgate.

Huntington, S. P. (1993). The clash of civilizations? *Foreign Affair* 72(3), 22 49.

Ladyz, J. (2006). Chosen aspects of sustainable development on the Polish, Czech and German borderland. *Geo Journal* 67, 1 7.

MacPherson, A. D., McConnell, J. E., Vance, A. and Vancham, V. (2006). The impact of US government antiterrorism policies on Canada US cross border commerce: An exploratory study from Western New York and Southern Ontario. *The Professional Geographer* 58(3), 266 277.

Mollick, A. V. (2003). Employment determination at Mexican maquiladoras: Does location matter. *Journal of Borderland Studies* 18(2), 45 57.

Newman, D. (2003). On borders and power: A theoretical framework. *Journal of Borderland Studies* 18(1), 13 25.

Peach, J. and Molina, D. J. (2002). Income distribution in Mexico's northern border states. *Journal of Borderland Studies* 17(2), 1 19.

Rautio, V. and Tykkylainen, M. (2001). Uneven development of economic interaction across the Finnish Russian border. *Post Soviet Geography and Economics* 42(1), 34 64.

Topaloglou, L., Kallioras, D., Manetos, P. and Petrakos, G. (2005). A border region typology in the enlarged European Union. *Journal of Borderland Studies* 20(2), 67 89.

Relevant Websites

http://www.absborderlands.org
Association for Borderlands Studies (ABS).

http://www.aebr.net
Association of European Border Regions (AEBR).

http://europa.eu
European Agency for the Management of Operational Cooperation at the External Borders (FRONTEX).

http://ec.europa.eu
European Commission's Regional Policy.

http://www.tnris.state.tx.us
Texas Natural Resources Information System, Texas/Mexico Borderlands Information Center (BIC).

Economies, Branch Plant

H. D. Watts, University of Sheffield, Sheffield, UK

Glossary

Filter Down Concept The view that as products age their manufacture filters down the urban system from larger to smaller places and from central regions to peripheral regions.

Manufacturing Belt (USA) The area stretching broadly from the mid-west to New England which in the 1950s and 1960s was the key geographical area in the US manufacturing system.

Product Life Cycle The theory that each product or service has a life cycle running from initial development through standardization/maturity to obsolescence.

Transnational Corporation (TNC) An organization controlling economic activities in more than one country.

Truncated Economy An economy which lacks higher level functions such as corporate head offices and research and development activities.

Introduction

A branch plant is a manufacturing site that is owned by a parent firm with a head office at another location. Conventionally, a parent firm controls a branch plant when it owns more than 50% of the share capital of the plant but share holdings as low as 10% provide opportunities to control an economic activity. In some cases, branch plants are referred to as subsidiaries of their parent companies. The term branch plant includes not only new factories but also formerly locally owned firms which have been acquired by firms based outside the local area. A geographical area is described as having a branch plant economy when branch plants account for a significant proportion of employment and output in the manufacturing sector and when such branch plants are owned by parent firms based elsewhere.

Branch plants can be owned by three different categories of parent firms.

- A transnational company (TNC) with its headquarters (HQ) in another country (a foreign TNC) – Japanese branch plants in the USA and EU countries.
- A TNC with its HQ outside the branch plant economy but in the same country (a domestic TNC) – a Ford plant in the south of the USA.
- A domestic multilocational company with no foreign operations but with its HQ outside the branch plant economy.

Although the role of foreign TNCs are emphasized in some analyses of branch plant economies, a number of studies have shown that their patterns of behavior do not necessarily differ from those of domestic TNCs.

Research has tended to focus on branch plants in the manufacturing sector, but there is no reason why the concept of a branch plant economy should not be useful in the analysis of the primary sector (especially mining and forestry). It might also be applied to areas characterized by a high proportion of service sector activities with parent firms based elsewhere.

Branch Plants and Branch Plant Economies

Although branch plants are often treated in an undifferentiated manner, there are a number of varieties. The local implementer branch plant produces a standard product for a specified market. The specialized contributor branch plant has a specific role within the parent firm and fits into a complex interdependent corporate system. The world mandate plant has worldwide (or at least major regional) responsibility for one (or perhaps more) of a firm's products. The potential impacts of these different branch plants can be very marked. In the discussion that follows, most of the more positive features of branch plants are associated with the latter two types of branch plant and the more negative features with the local implementer category.

An economy without branch plants would almost inevitably be weaker than one with branch plants. Branch plants (especially those on new sites) can bring many benefits to an economy. Incoming plants to new sites not only expand employment opportunities in manufacturing but they can also introduce new industries into an economy, helping to diversify the economic structure or replace older industries with new industrial activities. In some cases, they can provide a platform for the emergence of clusters of new economic activities. The establishment of new industries through the introduction of branch plants may provide a faster route to diversification of an economy than a reliance on the development of locally owned firms.

Branch plants can also provide higher wages and improved work conditions. Even where a branch plant employs only semiskilled or unskilled labor, it may offer higher wages than similar jobs in locally owned firms. This can raise wage levels and enhance work conditions generally in the area surrounding the plant. Another indirect feature of higher wages in incoming branch plants is that they may attract employees from over manned, locally owned firms and these locally owned firms may not replace the employees they have lost. The locally owned firms losing labor to the branch plants then achieve higher output per person as they utilize their smaller labor force more efficiently.

Branch plants have access to the financial resources of the parent firm, resources not available to a locally owned firm which may have to rely primarily on internally generated resources. Ownership of a branch plant by a TNC also offers access to global markets and global supply chains. Wider markets introduce stability by re ducing dependence on a limited geographical range of markets and access to global supply chains might offer a more cost effective purchasing strategy.

Perhaps, most critically, branch plants can also ac celerate knowledge transfer. They provide access to the knowledge of the parent firm and to the knowledge in herent in the wider networks with which it is connected. These links to the parent firm and the parent firm's suppliers and markets provide a conduit by which new knowledge is fed into an economy. In practical terms, this can assist technology transfer and speed the diffusion into the local economy of new processes, new working prac tices, and new products. The diffusion is enhanced where the branch plants enter partnerships with suppliers in the areas surrounding the branch plant. Partnerships may affect not only the suppliers of goods and services but also providers of training and skills for the local work force. Further diffusion can take place by new firms being spun off from the branch plants.

These points make clear that a branch plant in itself need not be a negative feature of a particular economy. A concern arises where an area is seen to have a high de pendence on branch plants, and that is where it is a branch plant economy.

Identifying Branch Plant Economies

It is important to emphasize there is no agreement as to the level at which the proportion of employment in branch plants becomes a cause for concern and at what point an economy qualifies to be termed a branch plant economy. A useful rule of thumb is that where over one third of the employment of the manufacturing sector of an economy is in branch plants, the description branch plant economy might well apply.

Branch plant economies tend to be characteristic of countries, regions, cities, and towns located in areas distant from, or on the periphery of, more prosperous economic areas. For example, the term branch plant economy has been applied at different times to all or part of Canada (peripheral to the USA), Ireland, Northern Ireland, Wales, Scotland, the north east of England (peripheral within the EU), the southern areas of the US (peripheral to the 'manufacturing' belt), and to specific cities such as Aberdeen (UK) and Lexington (USA). Although research has focused on areas in North America and Western Europe, it is probably worth recognizing that some of the issues that have arisen in these economies may appear in the transitional econ omies of Central and Eastern Europe where TNCs play an increasingly important role. The term might also be applied to the export processing zones/special eco nomic zones which are particularly characteristics of industrializing countries in South East Asia and China, although the maquiladoras of Mexico have also at tracted attention. In these zones, special incentives are available to encourage inward investment by foreign TNCs which import materials for processing before being exported.

In the 1970s, 60% of Scotland's manufacturing em ployment in larger plants was in branch plants. In Northern Ireland too, in the early 1970s, 53% of manu facturing employment was in branch plants. Canada is described frequently as a branch plant economy where 41% of employment was in branch plants in the early 1970s. One of the highest figures was recorded in the northern region of the UK where, in the late 1970s, over 80% of manufacturing employment was in branch plants compared with only 16% of manufacturing em ployment in the branch plants of foreign TNCs in the UK economy as a whole at that time. At sectoral level, the proportion of employment in branch plants can be over 90% especially where sectors are dominated by very large branch plants.

There has been a tendency for reliance on branch plants to decrease in many of these areas. In Canada, the proportion has fallen to just less than a third, while in Northern Ireland, in the 1990s, the proportion had fallen to slightly less than 40%. However, in Ireland, the pro portion is still around 50%, but, as shown below, the characteristics of these Irish branch plants have changed.

National governments rarely publish data on the ownership of economic activities, except for identifying activities owned by foreign firms. The latest figures noted above on branch plants in Canada and Ireland are derived from these sources. The most recent available data of this kind are shown in **Table 1**. It should be noted these data relate only to a selection of countries (e.g., data for Germany are not available) and the years for which the latest data are available range from 1999 to 2004. Al though these national data are not comprehensive in their coverage, at regional and smaller scales such data do not exist. The extent to which an economy is dominated by branch plants at regional and local scales can be es tablished only by surveys of individual plants within particular regions. Such surveys are rarely undertaken. In

Table 1 Employment in foreign-controlled manufacturing plants in selected EU countries, various dates

	Manufacturing employment	*Employment in foreign-controlled plants*		*Percent*
	2000	*Employees*	*Year*	
Sweden	764 463	234 785	2000	31.7
France	3 934 309	1 063 752	2001	27.0
UK	3 974 780	914 186	1999	23.0
Austria	608 956	135 115	2003	22.2
Netherlands	857 336	172 147	2000	20.0
Spain	2 429 195	382 043	2004	15.7
Italy	4 016 942	530 399	2001	13.2
Portugal	889 100	105 168	2004	11.8

Source: http://epp.eurostat.ec.europa.eu (Accessed 25 July 2007).

the absence of survey data, many research studies have identified branch plants from commercial directories.

There is an assumption that branch plants are con trolled by their parent firms and that a branch plant lacks control over its economic destiny. This is a simplistic view. Ownership is a useful surrogate measure of external control, but it has two major limitations. First, those few empirical studies that have measured the extent to which external ownership is equated with external control have shown that the control exerted by a parent firm over its branch plant can vary greatly from almost total control to very little control indeed. Much depends upon the in ternal structure of the parent firm and the extent to which it delegates control to its subsidiary organizations. External control will tend to be high in local imple menter branches, while levels of autonomy are likely to be high in world mandate branches. Overall, ownership measures tend to exaggerate the extent of external con trol. Second, many small firms, although nominally lo cally owned and controlled, are in fact dependent on major contracts with externally owned organizations and, *de facto* if not *de jure*, are controlled by them. Here, ownership data underplay the extent of external control.

These data difficulties mean that it is problematic to measure the number and extent of branch plant econ omies. An implication of this is that most analyses of branch plant economies focus on one (or occasionally two) area and there does not appear to have been any attempt to analyze the economic performance of a large set of areas with varying proportions of branch plants. An exception to this assertion is those studies which explore the role of foreign TNCs within national economies.

Emergence of Branch Plant Economies

A consideration of the implications of branch plant economies emerged in the last half of the twentieth century in a debate begun in Canada and developed further in Scotland. Four features of those years pushed a concern with branch plant economies up the agenda of economic geography. First, many areas experienced (and often welcomed) inward investment in new branch plants by foreign TNCs. Second, in 1960s and 1970s, domestic TNCs and domestic multilocational firms located a large number of branch plants in less prosperous regions of the UK (in the north and west) and the USA (the south) as they sought out cheaper labor. Third, and a major force in the emergence of branch plant economies, was the acquisition of locally owned firms by foreign TNCs, by domestic TNCs, and by large domestic multilocational firms. These three trends led to the fourth feature, the emergence of the view that the branch plant economy was a challenge to national sovereignty. Nationalist par ties in Wales and Scotland drew attention to the acqui sition of Scottish and Welsh companies by English (and other foreign firms). Similar concerns were expressed in Canada over the increasing dominance of US firms in the 1960s and 1970s.

It is not surprising there was considerable unease in local, regional, and national communities where major changes in economic activities were decided in distant places.

Measurement of Impacts

As noted earlier, data constraints make the analysis of the impact of branch plants across a wide set of areas very difficult (except where there is a concern simply with branch plants owned by foreign TNCs). In studies based on a single area, the distinctive features of branch plants can be identified by a comparison of branches and locally owned firms controlling for differences in, for example, sector and age. Although a number of such studies were undertaken in the 1970s, more recent evidence of this type is difficult to find. Contemporary studies can iden tify, for example, a low proportion of local suppliers among branch plants but there is usually no comparator to suggest whether or not the branch plants have higher or lower links than a locally owned firm.

Despite the apparent absence of rigorous inter area or intra area analysis, there is sufficient case evidence from a wide range of studies to piece together the impacts of a high proportion of branch plants in an economy. Some of this evidence is drawn from studies of the impacts of foreign direct investment (FDI) which, although sensible in the context of national economies such as Canada and Ireland, presents only a partial picture in the case of areas like the south of the USA or the north of England where domestic TNCs can be as important as, or even more important than, foreign firms. A distinction is also drawn in some analyses between the impacts of branch plants established through acquisition and those set up as new operations. Although the negative impacts of the branch plants have been widely researched, the positive impacts of branch plants, noted earlier, have perhaps been underemphasized. Indeed, it might be argued that while the short run impacts of a branch plants might be negative, the longer run impacts, resulting in the emer gence of a post branch plant economy, may well be more positive.

Employment

Branch plants are seen as introducing an element of in stability into an area's employment profile. As firms can readily switch production to other sites and the distant decision makers are thought to have little attachment to place, branch plants are seen as having high rates of job loss and, in the extreme cases, a high rate of plant closure. Although there are high profile cases of jobs being moved either across regions or, more recently, globally by do mestic and foreign TNCs, the aggregate evidence sug gesting a greater instability in branch plants as distinct from locally owned firms is mixed. The outcomes tend to be contingent upon the sectoral mix of an area and the plant size. In some areas, branch plants show a greater rate of job loss and closure than locally owned activities but in others they do not. However, in other aspects of employment, there is strong empirical evidence that a high proportion of branch plants can weaken an area's employment structure, especially as a result of functional truncation.

Such functional truncation is seen in the fact that branch plants (whether new or acquired) tend to exclude many higher level managerial/professional jobs. Indeed, acquisition of a locally owned firm is frequently followed by the removal of many of the higher level functions to another corporate site. Activities such as sales, marketing, legal services, and higher level management functions tend to take place outside the branch plant economy often at corporate headquarters. Research and develop ment activities are also frequently absent. Further this 'truncation' of an economy reduces the range of occu pations within an area and the lack of higher level pro fessional jobs can result in the out migration of the better qualified from the local economy.

Branch plants (particularly new branch plants) are also associated, in many cases, with a high proportion of unskilled/semiskilled employees, especially in the case of local implementer branch plants. These often undertake primarily, the assembly of standardized products and the branch plants in which they work are sometimes called screwdriver plants. The product cycle concept argues that, in the later stages of the product life cycle, the key input is low skill labor making low wage areas an at tractive location for products in the later stages of the cycle. Tied to this concept is the notion that as a product ages, it filters down the urban and regional hierarchy from larger to smaller places or from prosperous to less prosperous regions. The filter down process creates a spatial division of labor where the less skilled jobs are found in the less prosperous regions. Certainly, many branch plants in peripheral areas do tend to have an over representation of unskilled employees. Accompanying this spatial division of labor and reflecting wider social patterns is the link between less skilled work and the employment of female labor. This is most evident in screwdriver plants employing mainly female assembly workers.

There is also a concern that, from an employee's perspective, the work environment may be unduly con strained. Managers of branch plants may have limited autonomy and the distant decision makers may have limited awareness of local labor market conditions. This might reveal itself in higher frequency of labor disputes, higher labor turnover, and higher rates of sickness and absence among branch plants. This would tend to weaken a branch plant economy's labor productivity and its competitive advantage in the wider economy. However, the empirical evidence to support these arguments is very limited.

Multipliers

Each economic activity has a multiplier effect on its local area. The greatest effect arises from the spending of the employees. *Cetirus paribus*, a branch plant with few skilled employees and dominated by lower wage unskilled em ployees will limit the multiplier effects of employee spending. However, much interest has arisen in the extent to which purchases and sales of branch plants impact upon other firms in the area. Branch plants have been seen as 'cathedrals in the desert', that is, stand alone activities with few or no links to other businesses in the area. In aggregate, branch plants are thought of as being only weakly embedded in the areas in which they oper ate. There appears to be some truth in this view, espe cially with regard to the demand for business services.

It is unusual for branch plants to make use of local business services and this is particularly true where higher level business functions (e.g., marketing activities and legal functions) are absent from the branch plant. Even outsourced routine tasks, such as cleaning and catering, may not be provided by local people but through national agencies with which the parent firm has a national contract. The limited demand for business services restricts the development of such activities in the area and thus reduces the business support available to locally owned firms.

Although it is possible for a branch plant serving intermediate markets to stimulate upstream activities within the area in which it is located (e.g., through a petroleum refinery stimulating the growth of a petro chemical complex) much interest arises in exploring the extent to which branch plants stimulate downstream material supply activities in the area surrounding the plant. Although a few classic (and much cited) examples show such a process at work in motor vehicle assembly plants attracting suppliers of motor vehicle parts, the more general picture is that the stimulation of local supply networks though branch plant supply chains is often limited.

Inputs from the area within which a branch plant is located are often small. This is thought to reflect, in part, the lack of a local supply base or, where a local supply base exists, the limited awareness of purchasing man agers, based outside the area of the potential, of the area as a supply base. In some cases, it has been suggested that external purchasing managers may be aware of the po tential suppliers but, because of their unfamiliarity with those suppliers, they are concerned with the quality of local supply. Local links are further reduced by purchase of inputs from the wider areas within which the parent firm operates and, indeed, through a dominance of intra firm transactions. Even where there are important local purchases, these may be from other branch plants rather than from locally owned small and medium sized firms.

In addition to multiplier effects arising from links to a local economy through material and service inputs, plants can also act as sources of innovation and seedbeds for new firms. However, innovation in branch plants may well be below that of locally owned firms. As noted above, branch plants may lack research and development (R&D) function, thus limiting their innovative capacity. Since they may also lack marketing/sales functions, they may be disconnected from customers and markets and, therefore, less likely to identify market trends and new opportunities. Further, a tendency for branch plants to produce standardized products at the late stage of the product cycle means that the scope for innovation is low. Even where a branch plant successfully develops a new product, there is no guarantee it will be produced on site. It might well be transferred to another site of the firm in another area.

Since entrepreneurial start ups tend to be associated with managerial employees, the limited number and range of managerial jobs in branch plants reduces the entrepreneurial potential of an area. Spin off rates from branch plants tend to be low. Early evidence from Scotland showed few links between branch plants and new firms in the electronics sector and, in Ireland too, in the early stages of development, the evidence of former branch plant managers setting up their own businesses, as a result of plant closures and buyouts, was not very encouraging.

These factors taken together emphasize the limited multiplier effects within a branch plant economy which reduces an area's ability to develop self sustaining growth.

The Post-Branch Plant Economy

Some commentators have noted the emergence of the post branch plant economy. Whereas the latter half of the last century was frequently characterized by an increasing dependence on branch plants, this dependence does not necessarily show a continuous upward trajectory. Further, trends within TNCs have seen a tendency to move toward 'glocalization' where local needs are set more clearly within global operations and the nature of the branch plants can change. It is now recognized that branch plants develop over time through the accumu lation of resources and the development of specialized capabilities. They also change their roles within the parent firms to which they belong. It might be suggested that such developments over time might be more char acteristic of the branch plants of foreign TNCs than those of domestic TNCs.

Branch plants may originate as locally implementer plants but gradually take on more functions and, in the end, become major centers of R&D. The strongest of branch plants may eventually obtain world mandate status. Such a trend is evident in Scotland where there is evidence of plants moving from screwdriver operations assembling parts of a US product range for the European market, through to development of products for the European market and, finally, to engineering develop ment work for a world market. This pattern emphasizes that as a branch plant matures, additional manufacturing or indeed nonmanufacturing functions such as design, engineering, and research can be added. Similarly, within Ireland, branch plants have taken on important new roles within reorganized corporate structures. Branch plants, especially in the information and communications tech nology sectors, now have more significant positions within their wider corporate networks and an emerging cadre of Irish managers is helping to sustain economic development. Although an area may still have a high proportion of its manufacturing employment in branch plants, it is very different in its characteristics to the branch plant economy described earlier.

Although a branch plant economy can evolve into a post branch plant economy, the process of change can be slow. In the future, this evolution may become faster as corporations are hollowed out and more functions are devolved to branch plants.

Policy

The proportions of branch plants in many areas are likely to remain high, despite the increasing significance of local entrepreneurship in some areas. Less prosperous areas still seek inward investment and local firms continue to be acquired by firms based outside the area. Indeed, the more successful entrepreneurial firms may also exit by sale to a larger external competitor.

On the positive side, new branch plants are still seen as providers of new job opportunities and a chance to diversify and strengthen particular economies. More recently, there has been a focus on attracting branch plants in related activities in order to develop a cluster of like activities. In practice, most areas accept most jobs that are offered regardless of overall strategic development aims, although for bigger projects, local administrations will attempt to negotiate benefits additional to job creation. However, there are still concerns over the negative impacts of a high proportion of branch plants.

The inherent instability of some branch plants and their poor employment mix have encouraged a search for ways of changing the role of a plant within its wider organizational structure. It is now clear that managers of individual plants act as plant champions within intracorporate competitions. The role of plant champions can be strengthened by supportive local administrations.

Efforts are also made by local administrations to encourage local sourcing but such a policy is working against, rather than with, current trends to toward global sourcing initiatives. Where there are local suppliers to branch plants, supplier development programs enhance the performance of locally owned firms as the branch plants act as conduits to bring new ideas into the region.

Research Agenda

Since the consequences of a branch plant economy vary in time and space, a key research task is to review the current extent of branch plant economies and the impact of external ownership on employment patterns and local linkages. Much of our knowledge still relies on research studies undertaken one or more decades ago. Replication of earlier studies would prove useful, but, as a result of changes in national, regional, and local economies and a shift to a service economy, more needs to be known about the effects of external ownership on service activities (e.g., call centers, major shopping areas, tourist resorts, and business services). Research on branch plant economies must move away from a focus on the manufacturing sector.

While much can be learned from studies of the impact of inward investment by foreign TNCs, in particular national economies, the role of domestic TNCs should not be ignored in analyses of regional performance within those national economies. Further, any analysis of the impacts of external ownership must be set in the context of other critical elements of an economy. The key research question (certainly in any statistical study) must be whether ownership makes a difference when controlling for other potential variables influencing an area's economic performance.

Summary

The external ownership/control of a branch plant economy is only one factor in a complex system of factors influencing an area's economic performance. It should not be overemphasized. The negative aspects of branch plants are usually highlighted and the positive effects are played down. The evidence that branch plants contribute less to local economies than locally owned firms is contingent on time and place. However, there does seem consistent evidence to suggest that in most branch plant economies, the proportion of professional jobs is small, reliance on local business services is limited, and innovation rates are low. Even if branch plants are problematic, the level at which external ownership of an area's plants becomes a cause for concern is unclear. What is more, the characteristics of branch plants can change and a post branch plant economy can emerge.

See also: Embeddedness; Foreign Direct Investment; Global Production Networks; Spatial Division of Labor.

Further Reading

Birkinshaw, J. and Hood, N. (1998). Multinational subsidiary evolution: Capability and charter change in foreign owned subsidiary companies. *The Academy of Management Review* 32, 773 795.

Birkinshaw, J. and Morrison, A. J. (1995). Configurations of strategy and structure in subsidiaries of multinational corporations. *Journal of International Business Studies* 26, 729 755.

Breathnach, P. (2003). Women's employment and peripheralisation: The case of Ireland's branch plant economy. *Geoforum* 24, 19 29.

Brown, R. (2002). The future of ICT industries in Scotland: Towards a post branch plant economy. In Hood, N., Peat, J., Peters, E. & Young, S. (eds.) *Scotland in a global economy: The 20:20 vision*. Basingstoke: Palgrave Macmillan.

Crone, M. and Watts, H. D. (2003). The determinants of regional sourcing by multinational manufacturing firms: Evidence from Yorkshire and Humberside, UK. *European Planning Studies* 11, 717 737.

Dicken, P. (2007). The multiplant business enterprise and geographical space: Some issues in the study of external control and regional development. *Regional Studies* 41, S37 S48.

Firn, J. (1975). External control and regional development. *Environment and Planning A* 7, 393 414.

Hayter, R. (1997). Canada as a truncated production system. In Hayter, R. (ed.) *The dynamics of industrial location*. Chichester: Wiley.

International Labor Office (2003). *Employment and social policy in respect of export processing zones (EPZ)*. Geneva: International Labor Office.

Grimes, S. and White, M. (2005). The transition to internationally traded services and Irelands emergence as a 'successful' European region. *Environment and Planning A* 37, 2169 2188.

Phelps, N. (1993). Branch plants and the evolving spatial division of labour: A study of material linkage change in the northern region of England. *Regional Studies* 27, 87 101.

Phelps, N. A. and Fuller, C. (2000). Multinationals, intracorporate competititon and regional development. *Economic Geography* 76, 224 243.

Pike, A. (1998). Making performance plants from branch plants? *In situ* restructuring in the automobile industry in the UK. *Environment and Planning A* 30, 881 900.

Watts, H. D. (1981). *The branch plant economy: A study of external control*. London: Longman.

Relevant Websites

Websites tend to focus on the activities of foreign TNCs in national economies. The roles of domestic TNCs in branch plant economies within national boundaries are not identified in public datasets. The two sources below contain useful summary data relating to foreign TNCs only.

http://www.statcan.ca
Global links: multinationals in Canada, an overview of research at Statistics Canada.

http://epp.eurostat.ec.europq.eu
The impact of foreign controlled enterprises in the EU, Eurostat.

Economies, Imagined

A. Cameron, University of Leicester, Leicester, UK

Introduction

The idea that economies are 'imagined' is not a new one. Despite the prevalent assumption in much mainstream social science (particularly economics) that economies – and integral to them their geographies – are both self evidently 'real' (i.e., empirically observable) and the natural outcome of pure economic activity, in practice space economies historically have always been to some degree planned and actively created. Moreover, space economies in practice, however they come into being, are always and necessarily supported by forms and structures that are firmly rooted in the 'imaginary'. This does not mean that the idea of 'the economy' is simply made up and therefore not real, but rather, that we experience its reality only through the medium of the imaginary. If this sounds radically new or like the product of some extreme postmodern philosophy, it is worth remembering that the theory and practice of 'imaginary money' was an entirely normal aspect of the world economy for at least 300 years from the sixteenth to the nineteenth centuries. Indeed the term only went out of use because 'all' money effectively became 'imaginary' as its value became progressively disconnected from precious metals and based instead on circulating credit and legal fiat.

This article will look less at the particular imaginary aspects of economic practices and institutions (money, credit, value, etc.), however, than at the historically imagined relationships between economies and spaces. In particular, it will examine aspects of the historical construction, and alleged demise, of the idea of the geographically distinct and legally sovereign 'national economy'. This is not because the national economy is the only space economy and nor is it the only one that is produced by, and productive of, imaginary spatial concepts. The same is true of any spatially identified economic unit whether it is urban, regional, local, international, etc. The reason for focusing on the national economy is simply that for a little over two centuries, accounts of the existence, nature, and survival of the national economy have been the predominant ways in which economic space has been articulated and debated in and between capitalist societies. It is also because, with the advent of debates about economic 'globalization', the role of the national space economy as a valid unit of analysis and practice is again in question. In many ways, it is suggested here that the debate about globalization is precisely a debate about different and possibly competing imagined economies.

Social and Economic Imaginaries

By 'imagined economies' two distinct, but overlapping things, are being referred to here. First, it is a fact that economies are not naturally evolving or self evident but are actively planned, created, and/or identified by people – individually and collectively. Second, the capacity of economic institutions and practices, once established, to shape the social imaginaries of their subjects is referred to. As such, economic imaginaries are an important subset of, and contributor to, the wider set of social imaginaries that help us make practical sense of the complex world in which we live. The philosopher Charles Taylor defines social imaginaries in the following terms:

> By social imaginary, I mean something much broader and deeper than the intellectual schemes people may entertain when they think about social reality in a disengaged mode. I am thinking, rather, of the ways people imagine their social existence, how they fit together with others, how things go on between them and their fellows, the expectations that are normally met, and the deeper normative notions and images that underlie these expectations (Charles Taylor, 2004: 23).

As Taylor suggests, social imaginaries are not passive reflections on the world but are active processes through which we come to understand our position in the world and shape our normative responses to it. This is of importance to geographers because to conceive of 'an economy' as a spatial entity, and to place ourselves within it, involves some significant leaps of the imagination. Invoking a 'national economy', for example, involves subconsciously linking a disparate series of social, cultural, and political spaces to an economic space, thus according all of these complex systems a common internal geography and a common external boundary. Although we may know intuitively or through experience that these various systems are in practice never wholly contained by these boundaries, these spatioeconomic constructs remain firmly ingrained in our daily lives. This can be seen very clearly in recent British debate over whether to adopt the Euro in place of the Pound Sterling. Those arguing against the Euro appeal to the economic, strategic, political, and cultural significance of a 'national' currency for a 'national' economic space, despite the fact that money and finance have been thoroughly internationalized, even globalized, for many years. Even beyond such extreme debates, most people in their daily lives happily accept the notion of 'the national economy' – for example, as reported in the news media – without questioning the nature of its existence, let alone its geographies.

Although we can find examples of politicians and theorists actively imagining what economic forms and spaces might look like and then enacting them, economic imagination is not confined to these small groups and nor is it a one way process. Economic imaginaries are, rather, integral to everyday life and are enacted as much by the economic activities of ordinary people (consumers, bank customers, credit card holders, taxpayers, etc.) as they are by powerful groups. We could not even begin to function as economically active individuals at any level unless we were already equipped with a set of basic, functional imaginaries and/or assumptions (for example, drawing the very important distinction between a banknote and a photocopy of a banknote). Further, just as 'the economy' can in part be seen to be a product of imagination, once established it too carries within it the potential to imagine 'us'. In other words, when an economy – or a set of economic principles – becomes a reality, that reality to a certain degree also then shapes its subjects – the imaginers are themselves in turn imagined.

Imagining the 'National' Economy

Writing about the relative weakness of Italy as a political and economic force in Europe in the early nineteenth century, the Austrian political economist Friedrich List was very clear about its cause:

> She [Italy] supplied all nations with manufactures, with articles of luxury, and with tropical products, and was supplied by them with raw materials. One thing alone was wanting to Italy to enable her to become what England has become in our days, and because that one thing was wanting to her, every other element of prosperity passed away from her; she lacked national union and the power which springs from it (Friedrich List, 1844: 5).

List's point here is that for all Italy's economic strengths, the lack of a coherent, 'national' economic identity and structure rendered the country weak compared to England. In our terms Italy was not, unlike England, imagined as a single economic space (by its government, industrialists, or population) and thus was unable to capitalize on the economic potential of such unity. England – more accurately Great Britain as a whole – is important here because by the early to mid nineteenth century Britain had achieved an unprecedented degree of national economic and political unity. Much of what we now regard as the basic institutional features of the nation state and national economy were first forged in a series of policies enacted in Britain in the early to mid 1800s which were themselves the products of particular ways of imagining the consolidation and regulation of economic space.

Writing in the late seventeenth century, for example, the English political theorist Sir William Petty proposed the combined use of cartography and 'political arithmetic' (what we would now call statistics) to consolidate, tame, and render more economically efficient Britain's Irish colonies. As he put it in 1671:

> Ireland is such a place which must have so great an army kept up in it, as may make the Irish desist from doing themselves or the English harm by their future Rebellions. And this great Army must occasion great and heavy Leavies upon a poor people and wasted Countrey; it is therefore not amiss that Ireland should understand the nature and measures of Taxes and Contributions.
>
> The parishes of Ireland do much want Regulation, by uniting and dividing them; so as to make them fit Enclosures wherein to plant the Gospel: wherefore what I have said as to the danger of supernumerary Ministers, may also be seasonable there, when the new Geograp[h]y we expect of that Island shall have afforded means for the Regulation abovementioned (Sir William Petty, 1671).

In proposing the introduction of taxation to Ireland through the creation of a 'new geography' (by which he meant a map), Petty was laying the conceptual and practical foundations for what we would now recognize as a modern national economy. Although no such thing as a 'national economy' existed when Petty was writing, he shrewdly anticipated that a militarized and bureaucratic state form, exercising tight control over both land and people, could generate a more efficient and disciplined form of economic practice. By deliberately linking economic and fiscal efficiency with issues of security, morality (his system of taxation was based on a reformed church parish system), and efficient spatial consolidation, Petty is imagining the form of what would later become the ideal of the nation state linked to a spatially consolidated 'national' economy.

By the late eighteenth century, a number of 'classical' political economists (Adam Smith, David Ricardo, David Hume, and others) had built upon the early theorizations of Petty and others, to imagine the forms and structures of a 'national political economy'. These early theorizations of what a national economy 'ought' to look like, combined with a series of responses to particular historical events and in the context of growing industrial and imperial power, made it both possible and necessary for Britain to become the first 'national' economy in the modern sense of the term.

There were many historical dynamics driving the creation of the British national economy, but one of the most powerful and immediate was the threat of war with Napoleonic France. Following the Revolution of 1789, France declared war on Britain in 1793 and threatened to invade. The effects of this on the economic systems of Europe in general and Britain in particular were pronounced, causing, among other things, both the flight of precious metals from British currency markets and placing a considerable defense burden on the British government. The measures taken to resolve these problems entailed the imagination of the British economy in ways never before attempted and further developed the normative possibilities of what a national economy might be like.

First, in response to the run on gold and silver, in 1797 the Bank of England (at this time still an essentially private institution) in consultation with the government suspended payments in gold and silver to people seeking to redeem banknotes. Banknotes were not themselves legal tender at this time, but were promissory notes issued by banks that could be circulated in markets and ultimately cashed in for the precious metals they directly represented. The 'suspension period' was of fundamental importance because for the first time an economic system was seen to function purely on the basis of circulating paper. Earlier experiments to create such a system had failed – particularly the infamous 'John Law experiment' in France in the eighteenth century – leading to considerable distrust of paper money. While the suspension period had many problems – not least counterfeiting and fraud – and while payment in metal was eventually resumed in 1821, it nevertheless demonstrated that a 'national' currency could operate successfully on the basis of nothing more than circulating paper (backed by the national debt) and tight monetary control. This was the model of a national currency ultimately adopted by all modern nation states in the later nineteenth and twentieth centuries.

Second, the costs of fortifying the south coast of England and preparing the fleet were so great that the British Government was forced, reluctantly, to introduce the first ever 'national' income tax in 1799. Prior to this, wars had been funded through a series of *ad hoc* indirect taxes on commodities, periodic levies, and/or borrowing from merchant banks. None of these sources was either regular or productive enough to meet the immediate defense needs of the state in 1799, however, and so it resorted to direct taxation of the whole population in the name of common, national defense. Despite the controversies surrounding the tax, it was highly successful raising significant amounts and was known at the time as 'the tax that beat Napoleon'. Although abolished at the first opportunity, the income tax was soon reinstated more or less permanently and has since been adopted as a normal feature of a national economy. The British income tax set a very important precedent both in terms of the fiscal powers of the state and because it formally 'imagined' the population of the state as a single economic entity with a common responsibility and thus led them to, as Bourdieu put it, "discovering themselves as taxable, as tax payers."

These measures were accompanied by a series of others throughout the nineteenth century which further consolidated the emerging imaginary of Britain as a unified socioeconomic space. The first national census of population was conducted in 1801. The Ordnance Survey (founded in 1795) produced the first highly detailed comprehensive national maps of Britain in the 1830s. Legislation was passed that removed the special status of British cities which, until 1835, were able to run themselves as corporate entities largely independent of the national government. Other reforms introduced between the 1820s and 1870s also helped consolidate a national currency, gave special powers to the Bank of England in the service of national economic stability, consolidated private banking into a few large institutions, laid the foundations of what was later to become the welfare state and, in general, created the spatially unified 'social body' of the nation. This entailed the creation not just of a population that regarded itself as socially and spatially homogenous, but one that understood itself to belong simultaneously to the 'imagined community' of the nation and the 'imagined economy' of the state.

Imagining Economic Citizens

Imagining an economy in terms of institutional arrangements – boundaries, legal provisions, political structures, etc. – is not in itself enough. Economic practices as instituted and spatialized processes only become economies once people actively participate in them. The history of Western capitalist economies is, therefore, one of the continual shaping both of structures and of the people expected to operate within them. As part of the latter process the citizens of the emerging social, political, and economic orders not only had to physically occupy new institutional and regulatory spaces, they had to be educated and/or disciplined to behave in particular ways conducive to the emergent demands of the 'national economy'.

As Michael Perelman has argued, integral to the 'invention of capitalism' in the nascent national economies of Europe was the invention of the capitalist citizen – an individual that would give up traditional agrarian lifestyles to work in the emergent industries of the Industrial Revolution. Perelman argues that the same classical political economists, who imagined the institutional structure of the national economy, also laid down prescriptions for the often brutal disciplining of the people who would work within it. These included the forced relocation and urbanization of agrarian populations, use of the threat of starvation for those unwilling to work even for low wages, debt bondage of industrial workers, the widespread use of slavery, the use of the workhouse to punish the 'un deserving poor' for their economic deviance, and, more generally, the active promotion of a moralizing 'work ethic'.

Such measures also entailed the deliberate shaping of new geographies of economies. This can be seen in particular with the 'Enclosure Acts' passed in Britain between roughly 1760 and 1850, which consolidated British land ownership into large estates and emptied large areas of the traditional rural population. This contributed to the active forging of the new industrial economy in two ways. First, by disrupting traditional systems of small scale, subsistence farming, the En closures created bigger agricultural units which were both highly profitable and operated by a new class of farmer/landowner operating on an industrial scale. Sec ond, by driving people off the land, the Enclosures also provided rapidly growing industrial cities with the workforces they needed.

These processes were also accompanied by the emer gence of new forms of mapping, and a new 'cartographic imagination' which did not neutrally depict the landscape, but presented a powerful image of the emergent national economic space. This was particularly true of the 'cadas tral' maps – maps of legal property owership – that were produced in increasing numbers and with increased levels of detail and precision from the sixteenth century onward. Cadastral maps were both functional and active aspects of the imagination of the new national economic space. They were functional because they permitted levels of sur veillance and regulation over the land that had never before been possible – essential both for establishing and policing property relations and for setting and collecting taxes on property and people. Mapping of this kind helped to create what John Pickles describes as the 'geo coded world' – an ordered image of the landscape in which people and places were immediately and easily legible both to the state and to themselves.

Such maps were also 'imaginary' in that they pre sented an image of the national economic landscape that deployed visual techniques to provide 'simplified' images of people and places. As Pickles argues,

> Mapping, social statistics, public health, moral education and the institution of urban planning and administration were from the beginning thoroughly spatialized practices of identity construction and social engineering. Mapping and statistics made citizens visible in particular ways, rendering them subject to public administration, and it provided useful accounting tools for new practices, such as insurance (Pickles, 2004: 131).

Mapping and planning, as Pickles suggests, were es sential to making the imagined economy of the nation state a reality, because they presented an apparently abstract and neutral image of the land in which emerging economic relations were already embedded. The most apparently neutral and abstract map, therefore, presents an imagined economy (and society) as though it were a 'natural' feature of the landscape.

This naturalization of the imagined economy of the state was further bolstered by the emergence of a plethora of literary and media descriptions of proper (and, therefore, improper) economic practices and be haviors. Such 'imaginary' depictions of economic realities pervaded the popular media and became an important element in public debate over the nature of economic life. Literary historian Patrick Brantlinger has argued that there is a direct connection between the emergence of 'realism' in literary fiction in the nineteenth century and public debate over the 'reality' of certain aspects of the national economy (particularly paper money and credit, both of which were highly controversial because they did not seem to be 'real'). The constant reiteration of the idea of the economy through the popular and literary media, combined with the increasing experience of wage labor and the money economy all helped the process by which ordinary people began to understand themselves as economic citizens.

As a result of all these many, overlapping processes of economic imagination linked, of course, to real economic practices, ordinary people began automatically to asso ciate themselves with particular 'national' economic spaces long before the concept of 'the economy' entered common usage in the 1930s. It is during the inter war period that the idea of 'autarky' – complete national economic self sufficiency – reaches its most extreme and damaging expression in the totalitarian states of Ger many, Italy, Japan, and the Soviet Union. Even beyond these extreme examples, most states during this period imposed very tight controls over economic activity. This very strong expression of economic nationalism persisted to the post war period, albeit in the more benign form of various versions of the Keynesian National Welfare State, when a close association was drawn between individual and collective economic well being and national identity. The welfare state imagines 'the economy' to be a co herent and corporate entity analogous to the nation itself. Although this remains a powerful conception of state/ economy relations, the internationalization and global ization of economic activities starting in the 1960s have produced new economic imaginaries that are challenging the strong idea of the national economy.

Imagining Economic Globalization

Although, as noted above, few people would accept un critically the notion that all national economic activity can ever be contained within a given state space, the idea of economic sovereignty continues to be prevalent both in political and media accounts of economic practice and policy. The normative expectations of an economy and its geographies carried by our 'social imaginaries', this is to say, may not necessarily have too much to do with the 'actual' geographies of economic practices and flows we experience. This is of particular relevance in the globalization debate because much of what is claimed to be distinctive about the emergent global economic order rests on assumptions about economic space.

Most accounts of globalization start from the premise (usually simply assumed) that economic activity used to be national, then became increasingly international and now is rapidly becoming global. This certainly seems to be common sense if one accepts the idea that the natural trajectory of economic growth is to occupy ever larger spatial domains until it becomes, literally, global. However, as has been pointed out by some of the more critical literature on globalization, not only was economic activity never contained within the economies of nation states, in many ways it can be seen to have been more thoroughly 'global' in the past than it is today.

With respect to the idea of imagined economies this has two important implications. First, despite its prevalence in the language of politics and economics, complete economic sovereignty within a specified geographical space has never been completely accomplished and has, as suggested above, a very specific practical and conceptual history. Second, it implies that as the strong conceptualization of the 'national economy' loses its grip on the popular, and political imagination, other economic imaginaries, new and old, become available.

The controversies surrounding the idea of economic globalization, therefore, may stem less from the substantive challenges that emergent global economic structures and institutions may pose than the competing social imaginaries contained within the discourses of national and global economies. This is not to suggest that the global economy is simply a figment of individual or collective imaginations. Nor is it to suggest that the evidence marshaled in defense of the idea of globalization (the Internet, 24 hour banking, e commerce, etc.) is not real or does not represent significant institutional and practical change. On the contrary, economic practices, including the geographies in which they are set and which they help to produce, are continually changing and some of those changes (now and in the past) may be global in extent and/or consequence. How we interpret and represent these changes, however, depends in large part on the sorts of economic imaginaries to which we subscribe.

The ongoing debate over the meaning and existence of globalization is at one level, therefore, evidence of two competing and (possibly) mutually exclusive sets of economic imaginaries: one for which the sovereign nation state established in the nineteenth century remains the most significant geographic element; and another for which the open, global, market is the only rational spatial scale at which economic activity can be properly articulated. Between these two extremes we find politicians (and academics) subscribing both to the continuity of geographies of the states through which they derive their authority and to the new 'realities' of the global marketplace. From this have been derived a range of new descriptors for the state which attempt to capture 'both' its continued salience as an economic space 'and' its inclusion in a globally competitive marketplace. These descriptors include: the 'enabling state', the 'trading state', 'Postnational Schumpeterian Workfare State', the 'market state', and the 'competition state'. In all cases, these new 'descriptive' concepts are also deliberate attempts to 'reimagine' the nature of the relationship between economic practice and state space.

Conclusion

The idea that spatioeconomic categories are, in part at least, imagined has important methodological and conceptual implications for geographers. Conventionally, for example, economic geographers have simply adopted prevailing spatial imaginaries – particularly those of the nation state – and operated entirely within them. Whatever the substantive outcomes of such research, therefore, it has also had the effect of reproducing and reinforcing the imagined economy of the state. More recently, because the conceptual grip of such established imaginaries is declining, geographers and others have started to examine economies in different ways, reflecting different sociospatial imaginaries. Examinations of 'alternative' economies, various types of 'social' and/or gift economies, 'cultural' economies, 'urban' economies, state space, and so on, are all deliberately altering the ways in which the geographies of economies are imagined. This is not because 'national' economies no longer exist – the same plethora of media sources that first helped shape the nation state in the nineteenth century continues to promote the idea, after all – but because there is now a more general recognition that a properly reflective and critical geographical practice can no longer simply take the imagined nature of such economies for granted.

See also: Cultural Economy; Globalization, Cultural; Globalization, Economic; State Theory.

Further Reading

Amin, A., Cameron, A. and Hudson, R. (2002). *Placing the social economy*. London: Routledge.

Amin, A. and Thrift, N. (eds.) (2003). *The cultural economy reader*. Oxford: Blackwell.
Anderson, B. (1991). *Imagined communities: Reflections on the origin and spread of nationalism*. London: Verso.
Bobbitt, P. (2002). *The shield of Achilles: War, peace and the course of history*. London: Allen Lane/Penguin.
Bonney, R. (ed.) (1998). *The rise of the fiscal state in Europe c1200 1815*. Oxford: Clarendon Press.
Bonney, R. (2001). France and the first European paper money experiment. *French History* 15(3), 254 272.
Bourdieu, P. (1998). Rethinking the state: Genesis and structure of the bureaucratic field. In Bourdieu, P. (ed.) *Practical reason*, pp 35 63. Cambridge: Polity.
Brantlinger, P. (1996). *Fictions of state: Culture and credit in Britain 1694 1994*. Ithaca: Cornell University Press.
Brenner, N., Jessop, B., Jones, M. and Macleod, G. (eds.) (2002). *State/space: A reader*. Oxford: Blackwell.
Brenner, N. and Keil, R. (eds.) (2005). *The global cities reader*. London: Routledge.
Cameron, A. and Palan, R. (2004). *The imagined economies of globalization*. London: Sage.
Castoriadis, C. (1987). *The imaginary institution of society*. Cambridge, MA: MIT Press.
Cerny, P. (1990). *The changing architecture of politics: Structure, agency and the future of the state*. London: Sage.
Daunton, M. J. (1995). *Progress and poverty: An economic and social history of Britain 1700 1850*. Oxford: Oxford University Press.
Davies, G. (2002). *A history of money from ancient times to the present day*. Cardiff: University of Wales Press.
Einaudi, L. (1953). The theory of imaginary money from Charlemagne to the French Revolution. In Lane, F. C. & Riemersma, J. C. (eds.) *Enterprise and secular change: Readings in economic history*, pp 229 261. Homewood, IL: Richard D Irwin Inc.
Feavearyear, A. (1963). *The pound sterling: A history of English money*. Oxford: Clarendon Press.
Gilbert, N. and Gilbert, B. (1989). *The enabling state: Modern welfare capitalism in America.* Oxford: Oxford University Press.
Habermas, J. (1991). *The structural transformation of the public sphere*. Cambridge: Polity.
Helleiner, E. (2003). *The making of national money: Territorial currencies in historical perspective*. Ithaca: Cornell University Press.
Hirst, P. and Thompson, G. (1996). *Globalization in question*. Cambridge: Polity.
Ingham, G. (2004). *The nature of money*. Cambridge: Polity.
Jessop, B. (2002). *The future of the capitalist state*. Cambridge: Polity.
Joyce, P. (2003). *The rule of freedom: Liberalism and the modern city*. London: Verso.
Leyshon, A., Lee, R. and Williams, C. C. (eds.) (2003). *Alternative economic spaces*. London: Sage.
List, F. (1844). *The national system of political economy*. London: Longmans, Green & Co.
Mitchell, T. (1998). Fixing the economy. *Cultural Studies* 12(1), 82 101.
O'Brien, D. P. (ed.) (1999). *The history of taxation* vol. 1. London: Pickering and Chatto.
Perelman, M. (1996). *The invention of capitalism: Classical political economy and the secret history of primitive accumulation*. London: Duke University Press.
Pickles, J. (2004). *A history of spaces: Cartographic reason, mapping and the geo coded world*. London: Routledge.
Poovey, M. (1995). *Making a social body: British cultural formation 1830 1864*. Chicago: Chicago University Press.
Poovey, M. (1998). *A history of the modern fact: Problems of knowledge in the sciences of wealth and society*. Chicago: University of Chicago Press.
Rosencrance, R. (1986). *The rise of the trading state: Commerce and conquest in the modern world*. New York: Basic Books.
Scott, J. C. (1998). *Seeing like a state: How certain schemes to improve the human condition have failed*. New Haven: Yale University Press.
Seymour, W. A. (ed.) (1980). *A history of the ordnance survey*. Folkestone: Dawson.
Taylor, C. (2004). *Modern social imaginaries*. London: Duke University Press/Public Planet Books.
Williams, C. C. (2005). *A commodified world?: Mapping the limits of capitalism*. London: Zed Books.
Williams, R. (1973). *The country and the city*. London: The Hogarth Press.

Relevant Websites

http://www.adbusters.org
Adbusters Journal of the Mental Environment.
http://www.neweconomics.org/gen
New Economics Foundation.
http://www.quatloos.com
Quatloos Cyber Museum of Scams and Frauds.
http://www.taxhistory.org
Taxanalysts A public service of Tax Analysts.
http://terranova.blogs.com
Terra Nova A Weblog about Virtual Worlds.

Economy, Informal

S. Turner, McGill University, Montreal, QC, Canada

Glossary

LETS (Local Exchange and Trading Systems) Operate a currency system as an alternative to the formal economy.

Introduction

The informal economy is open to multiple interpretations leading to many, often contradictory, definitions. Very generally, the informal economy can be considered to be that part of the economy where activities, often occurring on a small scale, take place beyond official recognition and record. Such activities may include small scale enterprises and trade, street vending (**Figure 1**), garbage recycling ventures, subcontracting, and homework. Within countries, the nature of the informal economy is highly fragmented by location of work, be it in the home, in a small enterprise or on the street; the specific nature of the activity; and the status of employment. Yet in general, it can be said that all people involved in the informal economy share a lack of formal labor rights and social protection. The informal economy is most commonly found in developing countries where at times up to 80% of the economy might be based upon informal work, much of it self employed; however, such informality certainly also occurs in developed countries.

Two other sectors of the economy are sometimes incorporated within the term informal economy, a practice the International Labour Organization (ILO) argues against. First, it is argued that the criminal economy is not part of the informal economy. While there may certainly be illegal or semi illegal production or labor arrangements within the informal economy, it is suggested that this sector does not cover criminal activities in illegal goods or services such as drug smuggling. Second, the reproductive or care economy, including unpaid domestic work is considered by some not to be part of the informal economy, due to the fact that these activities are not part of the economy for which goods and services are created for sale or remuneration. However, such limits on the sector are debatable as they point to a sizeable grey zone, and others argue just as forcibly that these activities should be included. Likewise, debates continue as to whether subsistence agriculture forms part of this definition, with suggestions that because such activities are so widespread, they need to be analyzed in their own right.

There are many perspectives on the informal economy, and while some have tended to follow one another over time, they are not necessarily in conflict. Yet, while the term 'informal sector' was used in most initial works to describe these activities, it tends to have been

Figure 1 Informal economy street vending, Hanoi, Vietnam.

overtaken in more recent studies by the term 'informal economy'. In the 1990s and current decade, much re search tends to build from a range of previous ideas, pulling elements from different approaches when they are relevant to a particular setting. Indeed, it is important to note that informal activities are highly context con tingent, a fact stressed persuasively in recent work, with numerous different forms of informality occurring in different places and at different times. In addition, this term pertains to different economic systems, from peas ant to wage labor systems, from feudalism to capitalism to communism.

Approaches to the Informal Economy

Dualistic Models

Since their initial appearance in the 1940s and 1950s, dualistic models have tended to occupy center stage in research on the informal economy especially that oc curring in developing countries. Often considered a remnant of precapitalist economies, the informal econ omy was considered by this approach to refer to the activities of 'local' people in colonized countries, not yet integrated into the colonial economic and social struc ture. As a result, the sector was seen as backward, ir rational, and inefficient.

Writing in 1953, Boeke introduced an early concept of dualistic production systems based on an examination of the Dutch East Indies. He defined two social systems: the capitalist, imported from abroad; and the domestic social system. This model was to provide a starting point for much subsequent work including that of Clifford Geertz, who in 1963, identified the 'firm centered' and 'bazaar' economies in his work focusing on two towns in Java and Bali, Indonesia. The 'firm centered' economic sector in volved trade occurring via impersonal social institutions, while the 'bazaar economy' related to local competitive traders often relying upon numerous spontaneous interactions.

In 1971, at the 'Conference on Urban Unemployment in Africa' held at the University of Sussex, UK, Keith Hart coined the specific term 'informal sector' drawing from his work in Ghana. Hart understood the informal sector as being distinct from wage earning employment (the formal sector), the distinction between the two being the degree of rationalization of work. Labor within the formal sector was recruited on a permanent basis for fixed wages, whereas the informal sector consisted of ir regular forms of employment and self employment. Despite its relative complexity, Hart developed a classi fication of the two sectors based on the income oppor tunities available to people in a city. It is interesting however, that his division did make it possible for members of the same household to gain employment in different sectors of the economy. Shortly after, in 1972, the ILO, in a study based upon work in Kenya, built upon Hart's work while scaling it up to examine the structure of the economy as a whole. While maintaining a dualistic approach, one outcome of their research was to suggest that governments should support informal economy ac tivities because of the important role they play in ab sorbing unemployment. This marked a notable shift in attitude, from the informal economy being seen as a backward sector in earlier work, to instead being sug gested by both Hart and the ILO as a site for efficiency, creativity, and resiliency; indeed, perhaps a solution to economic development and poverty alleviation.

Almost immediately after the informal economy/ sector concept first appeared, criticism of the approach also emerged. One of the strongest concerns was that by placing the informal economy in opposition to the for mal, an artificial dichotomy was created. In addition, these approaches were accused of social evolutionism, or having a 'modernization take' on the economy, with as sumptions that informal ways of making a living would eventually be subsumed by the modern capitalist econ omy. Subsequently, a number of modifications to this initial work was made, including the work of Milton Santos who, in 1975, spoke of the urban economy as consisting of two circuits or subsystems, an 'upper' and 'lower' one, structured in a closely linked relationship. He described the upper circuit as using imported capital intensive technology and consisting of banking, export trade and industry, modern urban industry, trade and services, wholesaling, and so on. Most of the relations of this circuit were seen to operate at a national or inter national level. On the other hand, the lower circuit consisted of small scale, non capital intensive activities using locally adapted technologies. This circuit was dominated by the poor, and was well entrenched in the urban realm. Although appearing independent, it was argued that these circuits were closely interlinked with the lower circuit maintaining a dependent relationship with the upper circuit.

Petty Commodity Production Critiques

While the approach based on the informal/formal economy dichotomy was being developed by Hart, the ILO, and others from the 1970s, the increasing wealth of information linking those operating in the informal economy to those involved in other sectors of the economy forced many researchers to seek alternative frameworks. The most prominent of these emerged from a neo Marxist critique of utilizing informal economy terminology in developing countries, and subsequently the 'modes of production' approach became popular.

Neo Marxists saw informality as being a result of the unequal development of capitalism, with the informal economy playing an essential role in providing a low wage reserve of labor, a strategy theorized by Karl Marx as the maintenance of a disposable 'industrial reserve army'. While essentially still dividing the urban economy into two parts – the capitalist and precapitalist modes of production – the neo Marxist 'modes of production' approach paid closer attention to interfirm relationships, while being more pluralistic in nature. Several forms of precapitalist mode of production were recognized, one being labeled 'petty commodity production'. This aspect of the model became increasingly popular as a tool with which to examine the development of, and relationships surrounding, what had previously been called the 'informal economy'. Instead of having a benign, autonomous, or complementary relationship with the formal economy, this approach identified a relationship in which the petty commodity production sector was exploited and subordinated by the formal economy. Two main exploitative relationships were identified: first, the informal economy was considered an extension of the productive network of large firms, providing a pool of cheap and flexible wage labor; and second, it subsidized the formal economy by providing cheap goods and services to the labor force therefore enabling large firms to pay low wages.

However, this neo Marxist approach was not immune from criticism either, critiques including that it viewed socioeconomic transformations as a product of external determination, and as such ignored local histories, structural differences, and diverse possible responses. As such, there was little room in the approach for local agency. Yet at the same time, it did not take long for some authors to feel that the traditional dualistic approach and the neo Marxist approaches were intersecting, given that the features of those undertaking work in the informal economy were basically the same as those undertaking petty commodity production.

Trying to synthesize these approaches, authors such as Mazumadar (1976), and Bromley and Gerry (1979) created a framework based upon the vulnerability of workers. The informal economy was perceived as a labor market phenomenon identifying a specific dichotomy between the unprotected informal economy and the protected formal economy. Formal economy protection arose from the actions of unions, governments, or both acting together. Formal workers tended to have stable wages, protection via labor laws, and perhaps health and pension benefits. In comparison, the informal economy was characterized by industries that were 'unprotected' by company policy, government regulations or trade union action, and to which entry was relatively easy. Such workers, many of whom Mazumadar noted were rural to urban migrants, were often employed on a daily basis, and could be undertaking piecework for larger firms. Some authors placed the self employed in a separate category again, noting that such work created different forms of vulnerability.

Neoliberal/Legalist Approach

In 1989 the seminal work *The Other Path* written by Peruvian Hernando De Soto was translated into English. De Soto argued that informal activities were not due to the dynamics of the labor market but were a consequence of excessive state regulation. To support his arguments, De Soto and his research team undertook a widely cited experiment to try to register a small garment enterprise in Lima, a process that involved over 100 administrative procedures and took almost a year of continuous effort. Neoliberal/legalists see members of the informal economy as the hope for the future for capitalist economic development. They argue that what is needed is a reduction in state interference in the market and the elimination of the maze of regulations that are commonly associated with establishing an enterprise. Alternatively, critics of this approach state that an unregulated economy would not be the best solution, as the approach ignores the structural subordination of small enterprises and traders.

Structuralist Approaches

Given the climate of global economic restructuring during the 1980s and 1990s, it is not surprising that a structuralist approach, rooted in neo Marxist traditions, emerged partly in response to recessions in a number of developed countries, structural adjustment reforms, and increased competition from newly industrializing countries. Indeed, this was a time when post Fordist modes of production were becoming increasingly apparent in developed and advanced developing countries, with many workers pushed from formal employment into the informal economy. This approach, generally speaking – as there are many variants – rejected all notions of economic dualism and focused instead on the linkages amongst formal and informal production. Proponents of this approach argued that the informal economy was very much a result of capitalist firms attempting to evade state regulations regarding taxes and social legislation such as maintaining minimum wages, safety standards, and providing social benefits. Firms attempt this by implementing flexible labor arrangements such as networks of subcontracting, outworking, and piecework. The use of such labor practices, involving growing part time or casual work, tend to impact the most on the labor sector made up of vulnerable workers such as women, youth, and immigrants. In many cases the state has been supportive of such moves, especially in developing countries where informal practices are accepted so that a country can obtain comparative advantages with more regulated nations.

While the structuralists would argue that there are significant differences in the social and economic composition of the informal and formal economies, the neoliberal/legalists argue that these differences are legal, rather than structural, and are of the state's making. Structuralists wish to push for improved worker conditions, while neoliberal/legalists argue that worker protection will lead to a loss of employment opportunities. Nevertheless, proponents of both approaches tend to agree that the linkages between the sectors work to disadvantage those in the informal economy.

Small-Scale Enterprises

Since the 1980s another body of work to have emerged is that focusing upon the economic activities of small scale enterprises. This is a diverse body of literature that tends to focus on concepts such as 'informalization' rather than the 'informal economy'. This literature is highly relevant to discussions on the informal economy because of the fact that small enterprises are so often the setting for informal labor practices, which may often cut across family/enterprise boundaries, as well as often being unregistered and outside formal recognition by the local state. Within this realm of studies, there has been a concentrated body of literature focusing upon clusters of small enterprises, in areas such as the 'Third Italy', as well as developing countries such as Indonesia. Some have drawn from these studies to apply arguments emerging from flexible specialization literature to small enterprises operating in developing countries – with mixed results as to the applicability of such a concept. Work on small enterprises in general, however, unlike many informal economy studies in the past, tends to take a closer look at individual actors, highlighting the importance of the social embeddedness of economic activities. Such an approach allows for the inclusion of important variables such as gender, ethnicity, class, religion, and local politics in studies of how people mobilize resources and use social networks to not only get by, but often to get ahead as well. The important role of such small scale enterprises as a social safety net was especially notable after the beginning of the economic crisis in Southeast Asia in 1997 when thousands of people previously employed in formal work entered the realm of small enterprise and trade in Thailand and Indonesia. Indeed, here, as well as in Latin America in the 1980s, during structural adjustment in Africa, and economic transitions in the former Soviet Union and Eastern Europe, it has become plainly clear that employment in small scale enterprises and trade, and the informal economy as a whole tends to grow during times of economic crisis.

Social Networks and Social Embeddedness

Closely associated with the work on the informal economy and small enterprises is a body of literature that is even more heterogeneous, drawing upon interdisciplinary research from geography, economic anthropology, sociology, and economics. In essence, this research demonstrates that a common feature of informal activities is that they are facilitated by a wide variety of noneconomic actions, including reciprocal social relations involving kinship, friendship and neighborhood networks, and household, community, and other informal ties. Relations of production, consumption, and exchange in informal economies are therefore often deeply entangled in noneconomic relationships of trust and solidarity. The social networks view thus focuses on the organizational role of social ties, and rather than regarding informality as an absence of regulation *per se*, sees it as an alternative form of regulation functioning external to the framework of formal state bureaucracy. In more ways than one, this approach draws upon what social scientists with a substantivist leaning have argued since the 1950s, that is, there is no economy without cultural context, and local practices such as kinship ties and ethnic networks of trade partners are important elements in which the economy is unavoidably embedded.

Social networks play a role in social cohesion, the building of trust, and in defining community identity. They frequently operate within a framework of moral ethics that ensures mutual obligation and the right to subsistence for all members, acting as a 'safety net'. Such work is often closely associated with that on social capital, the networks and ties that can help people 'get ahead'. This approach is also being increasingly drawn upon for recent studies of informal small scale enterprises and trade.

Social network and social embeddedness contributions offer perspectives that contest the traditional economic dualisms of work/nonwork, economic/noneconomic, public/private, and monetary/nonmonetary. Instead, these approaches recognize that informal activities frequently take place in domains that straddle and combine these dimensions. Following such an approach it could be argued that the informal economy includes a number of activities that involve no transfer of money at all, such as barter, the exchange of domestic labor, mutual aid networks, and Local Exchange and Trading Systems (LETS), the latter operating a currency outside of the formal sector, to date most commonly found under this name in the UK.

The Causes of the Informal Economy

Naturally, as there are numerous definitions, approaches, and groups of participants, in turn there are a multitude of reasons put forward as to why the informal economy exists. Some of the more common explanations include government failures, for example, to provide adequate social support and labor regulations, or more broadly to satisfy societal needs; or alternatively, excess regulation of the economy by the state. Over urbanization and the inability of urban areas to absorb rural–urban migrants into 'regular employment' is often argued to play a role, especially in developing countries, China and India being key current examples. Economic restructuring and globalization are also often put forward as key causes of informal economy growth. Finally, another cause is people operating within the informal economy simply as an act of resistance to formal systems and the sur veillance that so commonly comes with them.

Geographical Perspectives

The work of geographers has spanned all of these dif ferent approaches to the informal economy. For example, Terry McGee, one of the most foremost human geog raphers writing on Southeast Asia, has been a key sup porter of the petty commodity production approach arguing that the dualistic approach was too simplistic, masking the realities of economic relationships in cities in the region and beyond. Indeed, geographers are valuable contributors to this field of study, bringing with them a concern for how local, on the ground realities are interwoven with economic and political changes at broader spatial scales, including national, regional, and global transformations; an understanding of the import ance of place based specifics to informal economy attri butes; and an awareness and understanding of the heterogeneity of the livelihoods and survival mechanisms of those working in this sphere.

See also: Economic Geography; Globalization, Economic; Informal Sector; Informalization; Livelihoods; Networks; Networks, Urban; Political Economy, Geographical; Slums; Third World.

Further Reading

Boeke, J. H. (1953). *Economics and economic policy of dual societies, as exemplified by Indonesia*. New York: Institute of Pacific Relations.

Bromley, R. and Gerry, C. (eds.) (1979). *Casual work and poverty in third world cities*. Chichester: Wiley.

Cheng, L. and Gereffi, G. (1994). The Informal economy in East Asian development. *International Journal of Urban and Regional Research* 18(2), 194 219.

De Soto, H. (1989). *The other path: The invisible revolution in the third world*. New York: Harpercollins.

Geertz, C. (1963). *Peddlers and princes: Social change and economic modernization in two Indonesian towns*. Chicago: University of Chicago Press.

Hart, K. (1973). Informal income opportunities and urban employment in Ghana. *The Journal of African Studies* 11(1), 61 89.

International Labor Organization (1972). *Employment, incomes and equality: A strategy for increasing productive employment in Kenya*. Geneva: International Labor Organization.

International Labor Organization (2002). *Women and men in the informal economy: A statistical picture*. Geneva: Employment Sector, International Labor Organization.

Mazumdar, D. (1976). The urban informal sector. *World Development* 4(8), 655 679.

McGee, T. G. (1979). The poverty syndrome: Making out in the Southeast Asian city. In Bromley, R. & Gerry, C. (eds.) *Casual work and poverty in third world cities*, pp 45 65. Chichester: Wiley.

Portes, A. and Sassen Koob, S. (1987). Making it underground: Comparative material on the informal sector in western market economies. *The American Journal of Sociology* 93(1), 30 61.

Rakowski, C. A. (1994). Convergence and divergence in the informal sector debate: A focus on Latin America, 1984 92. *World Development* 22(4), 501 516.

Santos, M. (1979). *The shared space: The two circuits of urban economy in underdeveloped countries*. London: Methuen.

Turner, S. (2003). *Indonesia's small entrepreneurs: Trading on the margins*. London: RoutledgeCurzon.

Relevant Websites

http://www.ilo.org
Informal Economy, International Labour Organization.

http://www.wiego.org
Women in Informal Employment: Globalizing and Organizing (WIEGD).

Ecotourism

D. A. Fennell, Brock University, St. Catharines, ON, Canada

Glossary

Accreditation Programs that provide a means of establishing the extent to which a business offering tourism experiences meets industry-nominated standards.

Alternative Tourism The antithesis of mass tourism defined by: local control, small-scale, benefits distributed more evenly in community, support for other local sectors, preservation of authenticity, and more holistic approach to planning.

Community Ecotourism should be designed according to a local orientation, as regards to benefits, control, and scale, as one of the five defining characteristics of ecotourism.

Consumptive Tourism Activities that have a direct negative impact on individuals, groups, or other aspects of the natural world, for example, hunting and fishing.

Conservation The prevention of resource depletion, species extirpation, or habitat degradation, as one of the five defining characteristics of ecotourism.

Education One of five defining characteristics of ecotourism, such that ecotourism is defined by a degree of education or learning as part of the experience.

Mass Tourism The predominant form of tourism defined by: foreign ownership, marketing based on high volume, seasonal employment; very little of the tourists' money stays in the destination, and with little effort to support local producers.

Nature-Based Tourism (NBT) Any form of tourism that uses natural resources in a wild or undeveloped form. Ecotourism is one form of NBT which is one of the five defining characteristics of ecotourism.

Nonconsumptive Tourism Activities that typically do not have a direct negative impact on individuals, groups, or other aspects of the natural world, for example, wildlife viewing.

Policy Whatever governments choose to do or not to do with respect to the planning, development, and management of the ecotourism industry.

Sustainability Development that meets the goals of the present without compromising the ability of future generations to meet their own needs. One of five defining characteristics of ecotourism.

Introduction

A sector of tourism that has generated a significant amount of interest and controversy over the past two decades is ecotourism. Interest, because it represents an alluring alternative to the mass tourism paradigm that has been so transformative to the social and ecological conditions of destinations throughout the world; and controversy because many feel that it has yet to live up to its position as a more responsible or ethical form of travel. Ecotourism ('eco' as in ecological) is a type of tourism that is founded primarily on a specific interest in the natural history of a region. Aspects such as culture and adventure, however, have also been associated with the term, making assessments of its magnitude very difficult. In 2003, ecotourism is said to have earned US$25 billion globally, representing approximately 2–4% of international tourism. Theorists have tended to focus on the origins and definition of ecotourism, as well as issues related to economic realities and incentives, consumptiveness, ecotourism as mass tourism, control and rights, and accreditation, all of which will be briefly discussed below.

Origins and Definition

Although there does not appear to be any definitive study outlining the precise origins of the concept of ecotourism, there is tacit support for the idea that it emerged from the eco development literature of the 1970s, regarding the harmonization of social and ecological objectives, wise management of environment, thresholds, and the enhancement of people, communities, and justice. Several theorists have adopted this position in illustrating that the ideas behind ecotourism resulted from concern over inappropriate use of natural resources. During the latter part of the 1980s, these ideas were more formally crystallized through the sustainable development movement as a model for structural change within society – one that ventures away from a strictly socioeconomic focus to one where development meets the goals of the present without compromising the ability of future generations to meet their own needs.

The application of sustainable development ideas and principles very quickly entered into the tourism vernacular, with an expanding market of travelers clamoring to take advantage of new eco based opportunities. The face of sustainable development in tourism emerged in the form of alternative tourism (AT) as a natural outgrowth of the dissatisfaction with the mass tourism industry which by the 1980s had adversely transformed many of the world's most beautiful and sensitive regions at the expense of communities and the natural world. The philosophy behind AT – forms of tourism, like ecotourism, that advocate an approach opposite to mass conventional tourism – was to ensure that tourism policies no longer concentrate on economic and technical necessities alone, but rather emphasize the demand for an unspoiled environment, the needs of local people, and the elimination of outside influences (e.g., external stakeholders) in winning back decision making powers.

Although the onset of ecotourism is very strongly tied to eco development and AT, it is important to realize that the concept has been with us for much longer. For example, the travel department of the American Museum of Natural History has conducted natural history tours since 1953. Probably the finest examples of the evolution of ecotourism can be found in the African wildlife based examples of tourism developed in the early twentieth century and, to some, the nature tourism enterprises of the mid nineteenth century. In the present day, some of the icons of ecotourism include many protected areas in Costa Rica like the Monteverde Cloud Forest Reserve, the Galapagos Islands of Ecuador, Iguaçu Falls, the Amazon basin, the Patagonian region of Argentina and Chile, Antarctica, hill tribe trekking in Southeast Asia, Ayers Rock, the Great Barrier Reef in Australia, the Milford Track in New Zealand, the Serengeti Plain, Kruger National Park in South Africa, and polar bear watching in Churchill, Canada. Indeed, ecotourism thrives in all environments including rainforests, mountain regions, polar environments, islands and coasts, deserts and grasslands, and marine regions. Practices in these and other ecotourism destinations include the watching of specific animal groups like cetaceans, birds, and bears. But ecotourism has also branched out into the realm of culture (indigenous practices) and adventure (kayaking and canoeing). While this branching out has increased the market, there are serious issues surrounding the legitimacy of forms of tourism that are more cultural and adventure based than ecological in their orientation.

Much of the literature supports the fact that Ceballos Lascuráin was the first to coin the phrase, defining it as traveling to relatively undisturbed or uncontaminated natural areas with the specific objective of studying, admiring, and enjoying the scenery and its wild plants and animals, as well as any existing cultural manifestations (both past and present) found in these areas. However, the term has been traced further back to the work of Hetzer, who used it to explain the intricate relationship between tourists and the environments and cultures in which they interact. Hetzer identified four fundamental pillars that need to be followed for a more responsible form of tourism: (1) minimum environmental impact, (2) minimum impact on – and maximum respect for – host cultures, (3) maximum economic benefits to the host country's grassroots, and (4) maximum 'recreational' satisfaction to participating tourists.

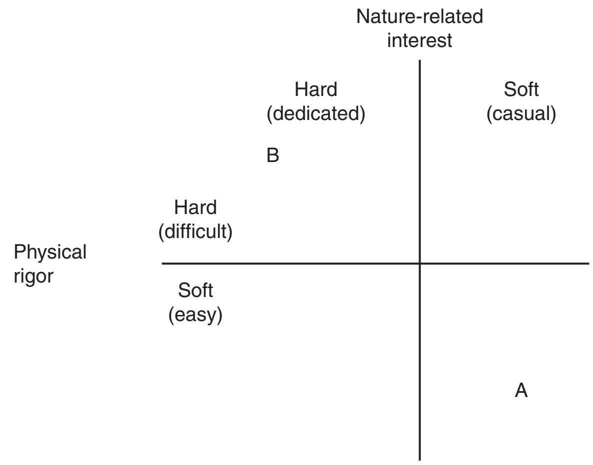

Figure 1 Hard and soft dimensions of ecotourism. From Laarman and Durst (1987).

Other theorists have made reference to nature tourism's (nature tourism being synonymous with ecotourism at the time) hard and soft dimensions, based on the physical rigor of the experience and also the level of interest in natural history (**Figure 1**). The hard and soft path characteristics have also been discussed as deep and shallow ecotourism. Deep ecotourism is characterized according to intrinsic value, small scale development, community identity, community participation, and the notion that materialism for its own sake is wrong. Conversely, shallow ecotourism is characterized as a business as usual attitude to the natural world; nature is seen as a resource to be exploited in maximizing human benefits, management decisions are based on utilitarian reasoning, and sustainability is viewed from a weak or very weak perspective. During the early 1990s ecotourism and nature based tourism (NBT) were differentiated on the basis of the fact that the former was recognized as one small part of a broader NBT defined as any form of tourism that uses natural resources in a wild or undeveloped form, and includes activities such as adventure tourism, hunting, fishing, as well as numerous other consumptive and nonconsumptive activities.

In efforts to provide a clear distinction between ecotourism, nature tourism, and other related forms like wildlife tourism (wildlife tourism is tourism based on encounters with nondomesticated (nonhuman) animals, which can be consumptive, for example, hunting, or nonconsumptive, e.g., wildlife viewing), researchers have attempted to isolate core variables that make ecotourism distinctive. This core includes sustainability, education, a nature base, a conservation mandate, and contribution to community, all of which have been incorporated in the following definition by David Fennell:

> Ecotourism is a sustainable, non invasive form of nature based tourism that focuses primarily on learning about nature first hand, and which is ethically managed to be

> low impact, non consumptive, and locally oriented (control, benefits, and scale). It typically occurs in natural areas, and should contribute to the conservation of such areas. (Fennell, 2008).

As regards to the ecotourist, studies point to the fact that this type of traveler maintains the following characteristics: predominantly male, affiliated with conservation organizations, long staying, well educated, and financially well off. Of interest is the recent trend toward more female representation in ecotourism, described as the 'feminization' of the ecotourism sector. Four studies examined between 2001 and 2005 found 57%, 62%, 70%, and 59.2% of their samples to be female.

Issues

Economic Realities and Incentives

Economists argue that ecotourism cannot hope to accomplish its lofty goals (e.g., biodiversity conservation), because of the inability to control the costs and benefits of market exchanges. Ecotourism is constrained by the fact that it produces a number of negative costs (termed externalities, like damage to the ecology from ecotourism) imposed on others which are often ignored in the operation of the market. It is difficult to internalize these external costs because economic theory tells us that incentives to ignore externalities are quite strong: it is rational, in an economic sense, to impose costs on third parties if this contributes to our own self interest. This means that there is scepticism regarding how successful ecotourism operators will be in balancing profit with the goals of ecotourism – with the former usually taking precedence over the latter. In the face of declining participation, firms will typically increase profits or increase demand, both of which may detract from the firm's faithfulness to the tenets of ecotourism. By reducing costs, efforts to minimize external costs are diminished by deliberately not taking steps to, for example, reduce pollution or remove garbage. While the service provider benefits by avoiding these costs, such are typically borne by a third party, like a community. In the case of increasing demand, the service provider appeals to a wider market by offering additional services or increasing carrying capacities in existing programs (e.g., the Galapagos Islands National Park in Ecuador). What eventuates is an industry that has parallel characteristics of mass tourism with larger numbers who may be less sensitive to the environment.

This seems to be the case with places like SeaWorld which are labeled and marketed as ecotourism, but without ecotourism's philosophical foundation. Researchers argue that the capture and display of animals for touristic consumption should not be considered as bona fide ecotourism because it places the well being of the animal at risk. As such, any form of ecotourism that places stress on the ecology of an area, including individuals and populations, should not be considered as ecotourism. This coincides with the extent to which other more consumptive forms of NBT have been considered as ecotourism. For example, moose hunting in Sweden has been referred to as ecotourism if cultural and social aspects are taken into consideration. In other cases, bill fishing (sport fishing for marlin and sailfish) has been referred to as ecotourism because it directs economic assistance to the local community and has economic advantages over other uses. Theorists, however, argue that such an activity cannot be ecotourism because of (1) the intention to catch the animal (ecotourism should be about minimum disturbance in all cases); (2) the pain and stress that results from catching the animal; (3) consumptiveness (catch and release practices still may be viewed as consumptive); and (4) values, such that ecotourists have a different set of values related to sport and the intrinsic/extrinsic motivations surrounding participation in these activities. The argument follows that the treatment of animals cannot be based on healthy populations (i.e., it is fine to catch animals because of the healthy state of the population), but rather that respect must be shown to individuals comprising these populations. The development and use of weak ecotourism definitions has opened the door to a great deal of misrepresentation (and unethical marketing) and the prospect of more anthropocentric values superseding ecocentric ones. The rationale for including activities like hunting and fishing under the ecotourism umbrella is not clear. One hypothesis is that the ecotourism label offers these activities social acceptability in the face of declining memberships. Accordingly, the best way to increase participation is to make the activity more socially acceptable through normative measures hinged on sustainability and ethics (i.e., catch and release or contributions of meat, skins, bones, and ivory to local people for commercial reasons).

The economic realities of ecotourism have also been used to predict the willingness of local people to support ecotourism and conservation through incentives. Conservation and environmental awareness are often raised because of involvement in tourism, which act as an incentive for a new rationality in traditional resource use. Ecotourism sites, however, that are not charismatic rarely generate revenue at a scale that will induce incentives for conservation. This is the case in Mongolia, where although local people have a strong conservation ethic, they were unwilling to discontinue grazing in certain protected area regions without compensation. Parallel to this is the finding that a chief motivating factor for involvement in the ecotourism industry is a depleting natural resource base based on deforestation or overgrazing activities. In Baja, Mexico, for example, poor fisheries conservation techniques lead to a depletion of the inshore fishery. In an effort to bolster the local economy, whale watching ecotourism emerged as an option on extrinsic (profit) grounds only. And because the whale watching industry evolved with the profit motive only in mind (in the absence of an ethical and well managed approach to ecotourism), it has not been so lucrative as to alleviate the pressures on the inshore fishery. As such, conflict over the right to use marine resources has only intensified as a result of ecotourism.

Community Control and Rights

The ecotourism literature is replete with examples of how the introduction of tourism has lead to the marginalization of populations through the removal of control and rights. Pressure from outside investment often translates into inevitability about the pace and scale of development and the perception of insignificance regarding the role of local people in ecotourism. Taken out of the hands of the people who live in the destination is the ability to control the effects that these large developments have on local resources, resulting in higher tourist impacts, the failure to put money into the hands of local people, and marginal efforts to conserve ecological resources. A logical reaction to exogenous forces is to simply reject outside funded tourism and ecotourism in order to fully control the tourism landscape. However, these efforts are constrained by (1) the inability to organize to the point of political mobilization; and (2) whether the mobilized unit can ever be strong enough to influence government policy, in the face intense development interests from outside. The path of least resistance for cash strapped governments, unfortunately, is foreign investment.

Unhappily, the rush to develop tourism in marginal areas has led to a series of problems that undermine efforts to balance traditional human uses and conservation. For example, the importance of considering local property rights has been demonstrated in Bohol, Philippines, where protected areas were established based on the effects of intensive deforestation, agricultural exploitation, and quarrying. With the establishment of limits to traditional harvesting activities, local people accuse policy makers of ignoring their rights because of the top down approach employed in the establishment of protected areas. This has led theorists to suggest that coexistence between local people, tourism, and protected areas managers should be developed through adaptive governance models that are based on shared involvement and more inclusive objectives.

Two issues emerge from the foregoing discussion. First, there has been a critical lack of meaningful policy in ecotourism to guide industry practices, which should be informed by broader national tourism policy, which itself is a manifestation of an overall government policy. Policy should involve all stakeholders who stand to be impacted by ecotourism development, including protected area personnel, local communities, the tourism industry, NGOs, financial institutions, consumers, and national ecotourism councils. Second, policy might also emerge in response to the inculcation of newer forms of governance which are based on the interactions of actors, structures, and processes through which societal decisions are made and power is exercised, and through institutions and norms. Consistent with this approach is the precondition that a given society be open to influence and change through development vehicles like ecotourism. This aspect of openness should not be treated lightly. If society is to somehow become better, it must be open to different approaches in a climate of cooperation between those groups who stand to be affected.

Ecotourism as Mass Tourism

Another issue that has come to the fore in ecotourism is the likelihood of infusing ecotourism's more responsible ethic into mass forms of tourism in making these other types somehow 'better.' Some ecotourism theorists argue that because scale doesn't matter in our efforts to be sustainable (i.e., small and large scale can either be good or bad) there is no reason to believe that ecotourism cannot occur at a grander scale. The argument follows that because softer path ecotourists restrict their activities to front country regions of protected areas which are hardened sites that can absorb the impacts of numbers (with hard path ecotourists more likely to penetrate deeper into the sensitive back regions of these areas), this group has less of an impact. Furthermore, the softer path ecotourist is more likely to contribute to conservation financially through the revenue generated by larger numbers of tourists. An example of this 'ecotourism as mass tourism' model can be found in Thailand, where two of the region's oldest eco tour operators, Sea Canoe and Siam Safari, are said to uphold the principles of ecotourism even though they have structural connections to the package tourism industry that provides their main market. As such, ecotourism in Phuket has emerged not in opposition, but as a function of the mass tourism industry, where tourists typically stay in four to five star hotels, visit Phuket en route to other destinations or for short holidays, and book through travel agents and tour operators.

Other theorists, however, are not so keen to accept this mass scale linkage because of the belief that ecotourism loses its integrity along the way. The fact that some softer path ecotourists stay in more luxurious accommodations, as with Phuket, above, means that their higher resource demands will contribute to a larger ecological footprint. The nature of this larger footprint often comes at the hand of the economic elite who gain considerable social capital by visiting exclusive softer path ecotourism resorts, even though they may not be very ecologically benevolent. The issue of impact is compounded by the fact that ecotourism involves long haul travel and associated high usage of fossil fuels to satisfy hedonistic ends. The more that participate, the higher the fuel demand.

Accreditation (Certification)

Some pundits argue that if ecotourism as an industry is to become legitimate, it might only do so through the implementation of accreditation (also referred to as certification) schemes which seek to instill confidence that programs in ecotourism are high quality, based on predetermined levels of competency. Accreditation is defined as programs that provide a means of establishing the extent to which a business offering tourism experiences meets industry nominated standards. Adherence to these standards, and the ability to advertise as such through eco logos, provide the organization with the opportunity to be competitive. A good example of accreditation in action is the Nature and Ecotourism Accreditation Program (NEAP) of Australia, where operators are given a core level of accreditation if they satisfy the program's basic criteria. The system also encourages operators to implement measures beyond the standards of the core in earning advanced standing. An important aspect of NEAP is that programs are accredited rather than operators. This means that although operators may run a number of programs, only those that are worthy get accredited.

However, despite the advantages of these schemes, others argue that accreditation has become politicized because of the sheer number of these schemes in operation and the control over those which are most prominent (e.g., Green Globe 21). Just as important in the debate over the significance of such schemes is the basic question of whether accreditation is worth all of the time and energy. They may ultimately be less effective than simple government regulations and codes of ethics, but more research is required to answer this question. Other theorists argue that such schemes cannot do justice to the experiential aspects of ecotourism because there is simply too much of an emphasis on the institutionalization of ecotourism through objective forms of evaluation which serve to strengthen power structures rather than topple them. Consequently, accreditation schemes have turned more toward commodification of ecotourism, rather than well being and participatory democracy – which is really what ecotourism ought to defend.

Conclusion

Issues like economic realities and consumptiveness, mass tourism as ecotourism, control and rights, and certification are only a few of the themes that continue to constrain ecotourism from reaching its potential as a positive agent of chance. These are all ethical issues that need proactive rather than reactive treatment if we are to move forward collectively. Perhaps most damning to ecotourism is the criticism that it is often planned and managed no differently than other more invasive forms of tourism through linear, instrumental modes of thinking that have made the marginalized morally unimportant in the face of industry efficiency. Theorists are only now realizing that the pervasive problems in ecotourism are a function of 'who we (humans) are at the deepest level'; and until we start to better understand the dynamics of this, no name, label, or concept can have any hope of making things better. Ecotourism can still be that form of tourism that we want it to be, but we do need to temper this hope in full view of some fundamental truisms about human nature that need to be explored from an interdisciplinary perspective. What also continues to limit ecotourism, apart from this interdisciplinary void is a disconnection between industry, local people, academics, and government with respect to a collective vision for the future. It is indeed unfortunate that so many competing demands have grossly disfigured ecotourism's most basic philosophical premise – an ecocentric approach to tourism – because of more instrumental values.

See also: Conservation and Ecology; Consumption; Ecology; Nature; Sustainability; Sustainable Development; Tourism; Tourism, Rural.

Further Reading

Fennell, D. A. (2008). *Ecotourism: An introduction* (3rd edn.). London: Routledge.
Fennell, D. A. and Dowling, R. K. (eds.) (2003). *Ecotourism Policy and Planning*. Wallingford: CABI.
Higham, J. (ed.) (2007). *Critical Issues in Ecotourism: Understanding a Complex Tourism Phenomenon*. New York: Elsevier.
Honey, M. (1999). *Ecotourism and Sustainable Development: Who Owns Paradise?* Washington D.C.: Island Press.
Page, S. and Dowling, R. (2002). *Ecotourism*. Essex, UK: Person Education.
Weaver, D. B. (1998). *Ecotourism in the Less Developed World*. Wallingford: CABI.

Relevant Websites

http://www.bigvolcano.com.au
Big Volcano Ecotourism Resource Centre.
http://195.212.4.4
ECONETT.
http://www.ecotourism.org
Ecotourism Explorer.
http://ecotour.csu.edu.au
Ecotourism Information Centre.
http://www.txinfinet.com
Planeta.
http://www.yorku.ca
STRING.
http://www.green travel.com
The Green Travel World Ecotourism Directory.
http://www.tourismconcern.org.uk
Tourism Concern.

Edge Cities

N. A. Phelps, University College London, London, UK

Glossary

Edge City A settlement that has grown rapidly in under half a century, which has more jobs than bedrooms and at least five million square feet of office space and 600,000 square feet of retail space.

Introduction

Washington Post journalist Joel Garreau in 1991 coined the term 'edge city' to speak to the apparently spontaneous growth of new urban nodes at the edge of but not contiguous with major established cities in the United States. More so than a profusion of other terms used by social scientists over several decades prior to the 1990s, Garreau's term edge city managed to impose itself upon the popular consciousness and academic debates in the United States. The term, for instance, found a prominent place in Tom Wolfe's novel *A Man in Full*. It also inevitably was brought into academic discussion and analysis in the United States. In particular, it has come to be associated with the increasingly polycentric, postmodern nature of city regions in the United States where "what was once central is becoming peripheral and what was the periphery is becoming increasingly central … with the intensive urbanisation of the suburbs into outer cities or edge cities …" (Soja, 2000: 152).

In fact, edge city like developments were observed to be a feature of expanding Chicago – the archetypal 'modern' industrial city – during the 1920s and 1930s and, if we see their lineage in processes of suburbanization, we might trace the ancestors of edge cities in the earliest suburbs of modern cities. This article situates a discussion of edge cities with respect to long standing processes of suburbanization. Edge city appears as one among many terms used to try to capture the ever increasing size and complexity of city regions over time. The article also discusses the extent to which the term edge city signifies something distinct about the form, function, processes producing, and ideological content of, contemporary urbanization when compared to earlier processes of suburbanization. The article then considers the mildly paradoxical planned nature of urban sprawl and the rise of edge cities. Finally, the article considers whether edge city is a term that can be deployed outside of the United States.

Edge City

An edge city was formally defined by Garreau in 1991 as: having at least 6 million square feet of office space; having at least 600 000 square feet of retail floor space; having more jobs than bedrooms; being perceived as a separate place; and having been nothing like a city as recently as 30 years ago. On the basis of this definition, Garreau identified most major United States cities to have one or more such edge cities – there being 123 existing and 78 emerging edge cities in the United States at the outset of the 1990s. Beyond these main criteria, Garreau's descriptions of edge city development also drew attention to the locational preferences of owners and managers of companies for an 'edge' location. There is also what he describes as the 'law' that 85% of employment growth in edge cities derives from existing businesses.

Despite the apparent formality of this definition, it conveys little of what most people understand by the term edge city and which is now so firmly invested in the popular and academic consciousness. Perhaps the key defining feature of a United States edge city is the low density, car dependent commercial and residential development apparent at major intersections on the freeway system. Edge cities present themselves at the most accessible points on the interstate road system and/or with close connections to other major transportation infrastructure such as airports.

The process of suburbanization has always been more economically, socially, and racially heterogeneous than often entertained in popular discourse on the subject. Studies have therefore, to an extent, filled in some of the details left blank by Garreau in his descriptions of edge cities. Bingham and Kimble, for example, suggest that the economy of many edge cities is often more diverse than their central city neighbors such that they have replaced a segment of downtown. Edge cities also share some of the mixed social complexion of cities and inner suburbs such that many of the problems of cities stand to be peripheralized with growth of edge cities.

However, in other respects

> The "edge cities" label is misleading. First of all, they are not cities. They are business centers with some but not all, of the functions of the old down towns. Nor is the "edge" part really accurate. Many were nowhere near the edge when they were built. … Few edge cites are anywhere near the edge today. In fact, after the 1980s few new ones have appeared. It seems that the sudden growth

> of these centres might have been only a brief transitional phase in a much longer process of decentralization. (Bruegmann, 2005: 71)

Indeed, two of the three types of edge city identified by Garreau tend to imply that such developments are not entirely novel. Popular and academic interest in Garreau's work has tended to focus on what he terms 'greenfield' edge cities that are the newest and physically most distinct of edge cities. However, there are also what he describes as 'uptowns', which have been built on preexisting settlements of some note, and 'boomers', which have evolved from chaotic suburban development.

Garreau's term is just one of 44 pieces of terminology used more or less successfully to try to capture the ever expanding urban fabric of the United States. In its definitional criteria it echoes earlier terminological innovations which have also drawn attention to the basis of such new urban nodes in employment rather than residential functions. Terms such as edge city have tended to imply a break with past processes of suburbanization – and this remains something of an open question as we consider further below.

Lineages of Edge City

Although Garreau's concern was to stress the novelty of edge cities as the new frontier in American life, it clearly is appropriate to locate edge cities as a recent incarnation of long standing processes of urbanization outside the city proper as understood at any particular historical moment. If we do so, we are presented with at least three interrelated issues around which any historical–geographical analysis should focus, namely, the form, appearance, or morphology of urban development; the processes and actors involved in urban development; and the function of urban development.

The term 'suburb' dates back to the expulsion of unwanted residents and land uses from the walls of the preindustrial city. While a small number of the wealthy in society also resided outside the city proper, there seems little point in tracing the lineage of edge cities to such ancient times. For a start, the meaning and ideological content of suburban development at this time was in almost complete contrast to the meaning and ideological content that has come to dominate in modern times.

Hayden includes edge cities as one manifestation of the 200 year long history of suburbanization apparent among what arguably were from the start modern cities in the United States. Important continuities in the ideological content of various incarnations of suburbanization are apparent in this more limited time frame. Even so, Teaford has suggested that new postsuburban ideological content has been added with the rise of the likes of edge cities. Central "to the emerging post suburban polity is the tension between suburban ideals and post suburban realities. … Economically these areas may have become post suburban but intellectually and emotionally they were solidly suburban" (Teaford, 1997: 5).

Walker and Lewis have recast the growth and outward expansion of North American cities dating back to the 1800s as typically occurring in a series of nodal developments that were then subsequently absorbed into one continuous fabric. That same process was apparent to Louis Wirth commenting on 1920s Chicago – the modern city made famous by the Chicago School of urban sociologists.

> The city and its hinterland represent two phases of the same mechanism. One of the latest phases of city growth is the development of satellite cities. These are generally industrial units growing up outside of the boundaries of the administrative city, which, however, are dependent upon the city proper after the city has inundated them, and thus lose their identity. (Wirth, 1925, cited in Dear, 2003: 502)

In the 1960s, Jean Gottman was drawing attention to the decentralization of manufacturing and service industries and specifically to the interurban location of associated developments within Megalopolis – the heavily urbanized swathe of eastern seaboard of the United States connecting Boston and Washington, DC. Finally, by the time Garreau was writing, others such as Robert Fishman and Jon Teaford were also drawing attention to what were perceived to be novel aspects of urban development at the edge of established cities.

Seen in the context of the historical antecedents described above, the edge cities enumerated and described by Garreau lose some of their novelty. Moreover, the term edge city has been overtaken with recent reference to what Lang calls 'edgeless cities' in the United States and to 'Zwischenstadt' – in between cities – in Europe, as commentators try to come to terms with the economic, social, and cultural basis of the evermore diffuse urban landscapes and systems in which we live and work.

Paradoxes of Edge City

It is tempting to view the growth of edge cities as something of a spontaneous unplanned phenomenon. Indeed, there is a sense in the United States that the generalized urban sprawl of which edge cities are a part has been produced from a planning system which has helped to produce an outcome of nonplanning . Garreau, in fact, appears to have borrowed the term edge city from a Phoenix city region planning strategy which sought, in a context of diffuse urban sprawl, actively to channel development into identifiable urban centers. Thus, planning conceived in several different guises has been instrumental in the rise of edge cities.

Federal state policies have clearly had a major role in the production of urban sprawl and edge cities. The characteristic habitat of edge cities at the intersections of major interstate highways has much to do with the major Federal government expenditures made in the 1950s and 1960s which threw up fresh development opportunities at a host of new highly accessible but undeveloped locations. Federal expenditures – in the shape of housing programs – also incentivized mass tract house building. The articulation of these Federal programs with other state and municipal policies and with developments in the private sector have seen urban sprawl identified as 'the' physical manifestation of the Fordist industrial complex. In a similar analytical vein, albeit that the systemic logic and the precise articulation between sectors and interests within national economies are argued to have metamorphosed, the likes of edge cities and their counterparts in Europe have come to be regarded as the projection spaces of an emergent post Fordist economy.

Within this context of a broad articulation between sectors of the economy, we can also see quite specific instances of 'planning' that have contributed to the production of sprawl and edge cities in particular. In Los Angeles, and pretty much everywhere else in the United States, the limited planning capacities of local governments coupled with their own desire to compete with one another for jobs and wealthy residents has seen land use zoning powers promote urban sprawl. In the absence of any counterbalancing cooperative planning and joint financing arrangements among municipalities, urban sprawl has tended to produce severe income, employment, and collective consumption expenditure inequalities among municipalities in major metropolitan regions.

However, in a technical sense much of the planning expertise exists outside of municipal planning departments in the private sector. Hise, for example, has stressed the role of major private sector companies – either manufacturing companies decanting from inner city locations or housing construction companies – to the master planning of new communities at the edge of major established cities such as Los Angeles. He presents what is, at first glance, a curious juxtaposition when arguing that "the historical record suggests … Edge cities are not a new phenomenon. We can trace conceptual roots back to Ebenezer Howard's garden city and the planned dispersion of the industrial city" (Hise, 1997: 215).

Hogan too highlights how the relative success of what he terms 'big picture planning' has generated the urban sprawl in San Diego that it was supposed to prevent. Here, big businesses, big environmental groups, and municipal and state planners actually thought rather alike. There is a sense then in which planning ideals and plans themselves have had unintended consequences in their part in the production of sprawl and edge cities.

Finally of course, continuities in the ideological content of suburbia and the postsuburbia of edge cities and the like should alert us to the fact that we as individuals – with our residential, work, and commuting preferences – are implicated in the production of sprawl and of edge cities or their functional equivalents elsewhere in the world. While urban development including discrete elements such as edge cities take place within a systemic context, this context has evolved over time with a degree of population decentralization having taken place under each economic system. It is under capitalism – the system most associated with mass affluence and democratic politics – that Bruegmann sees the greatest production of sprawl as a product of the expression of myriad residential and business location preferences.

"For much of the world, the Edge City maxim, that every American city is growing in the fashion of Los Angeles has become much more of a foreboding than a hopeful promise" (Soja, 2000: 401). The foreboding relates to the perceived inequities of the sort of urban sprawl presumed to be represented by edge cities. Here again interesting paradoxes abound. While urban sprawl does present us with some genuine questions, concerns over sprawl have often been aimed – due to the very term sprawl – at a moving target, and one that frequently has been defined on the basis of class and aesthetics. Edge cities – as the term implies – actually display a degree of nodality and density of development which we presently would not associate with the term sprawl. In fact, decreases in the density of United States cities reached their maximum by the 1960s and since then have stabilized or reversed. This has happened because of increases in the density of development in the suburbs. There is no little irony in the fact that Los Angeles is now one of the most densely populated cities in the United States.

California Über Alles?

Despite the generality of the definition proposed by Garreau, the term edge city is one that is now firmly invested with the specifics of the appearance of suburbanization in the United States setting. Also, if anything, the term has been further melded in popular and academic discourse with the archetypal United States sprawl – Los Angeles in the state of California. There seems little sense therefore in looking for edge cities *per se* in Europe and East Asia for example. Yet, literature on Europe and East Asia has begun to focus on urbanization at the edge of established cities. So the question remains, does the apparent uniqueness of the United States edge city obscure potential points of comparison with possibly similar developments elsewhere in the world?

Certainly some of the differences between patterns and processes of suburbanization in the United States and those in Europe, for example, appear to revolve around differences in their timing in these different contexts. Commentators point to a degree of convergence among United States and European and between United States and East Asian patterns of urbanization, including the sorts of decentralization of population and jobs apparent in edge cities. As such sprawl, and presumably the sorts of systemic tendencies and individual preferences crystallized in edge cities, are not peculiarly American.

Thus, "recent development tendencies in European metropolitan regions bear resemblance to Edge City development in several respects" (Bontje and Burdack, 2005: 317) but differences also exist in terms of the involvement of the public sector, for example, that render these developments, specifically European variations, on the edge city theme. Here, while new urban nodes at the edge of major cities may differ in terms of their physical appearance from edge cities in the United States, they may have similarities in terms of their functioning as part of a wider urban system and even in terms of the processes and the politics implicated in their growth.

The partial incorporation of market principles and elements of de and re regulation of economies in East Asia has set in train very rapid urbanization, including at the edge of major cities. This has led some to argue that since the 1980s "a pattern of new town developments integrated with industrial estates, toll roads, ports and airports. … Notwithstanding the very different settlement pattern on which it is being imposed, the new arrangement matches very closely with what Garreau described in Edge city" (Dick and Rimmer, 1998: 2304). However, the significance and place of such developments within the porous and mixed land use extended metropolises typically regarded as rather unique to Southeast Asia remains an open question.

Conclusion

Like it or loathe it, the term edge city is one that continues to excite opinions both popular and academic. The phenomenon of edge cities can usefully be situated in a broader historical and geographical context in order that its value and limitations can be appreciated. As a term it may already be too firmly invested with the specific appearance or morphology of urban development in the United States to be of wider analytical use. Yet edge cities, as incarnations of longer term processes of suburbanization which have been as much nodal as seamless additions to contiguous urban areas, focus our attention on some of the most active and advanced patterns of urban development.

Edge city may already have experienced its high point as the term of choice to refer to developments that many have observed and named. Judging by past experience, the essence of what it crystallizes about the socio economic and cultural complexity of modern urban systems is bound to be the subject of continued popular and academic interest and terminological invention.

See also: Concentrated Deconcentration; Corridor and Axis Development; Los Angeles School of Post-Modern Urbanism; Suburbanization; Urban Planning and Human Geography.

Further Reading

Bingham, R. and Kimble, D. (1995). The industrial composition of edge cities and downtowns: The new urban reality. *Economic Development Quarterly* 9, 259 272.
Bontje, M. and Burdack, J. (2005). Edge cities, European style: Examples from Paris and the Randstad. *Cities* 22, 317 330.
Bruegmann, R. (2005). *Sprawl: Compact History*. Chicago, IL: University of Chicago Press.
Dear, M. (2004). The Los Angeles school of urbanism: An intellectual history. *Urban Geography* 24, 493 509.
Dick, H. and Rimmer, P. (1998). Beyond the third world city: The new urban geography of Southeast Asia. *Urban Studies* 35, 2303 2322.
Garreau, J. (1991). *Edge City: Life on the New Frontier*. New York: Doubleday.
Gottman, J. (1961). *Megalopolis*. Cambridge, MA: MIT Press.
Hayden, D. (2003). *Building Suburbia: Greenfields and Urban Growth, 1820 2000*. New York: Pantheon.
Hise, G. (1997). *Magnetic Los Angeles: Planning the Twentieth Century Metropolis*. Baltimore, MD: Johns Hopkins University Press.
Hogan, R. (2003). *The Failure of Planning: Permitting Sprawl in San Diego Suburbs, 1970 1999*. Columbus, OH: Ohio State University Press.
Keil, R. (1994). Global sprawl: Urban form after Fordism. *Environment and Planning D, Society and Space* 12, 31 36.
Lang, E. (2003). *Edgeless Cities: Exploring the Elusive Metropolis*. Washington, DC: Brookings Institution Press.
Muller, P. (1981). *Contemporary Suburban America*. Englewood Cliffs, NJ: Prentice Hall.
Phelps, N. A., Parsons, N., Ballas, D. and Dowling, A. (2006). Business at the margins? Business interests in urban politics. *International Journal of Urban and Regional Research* 32, 362 383.
Phelps, N. A., Parsons, N., Ballas, D. and Dowling, A. (2006). *Post Suburban Europe: Planning and Politics at the Margins of Europe's Capital Cities*. Basingstoke: Palgrave MacMillan.
Sieverts, T. (2003). *Cities without Cities: An Interpretation of the Zwischenstadt*. London: Routledge.
Soja, E. (2000). *Postmetropolis*. Oxford: Blackwell.
Teaford, J. (1997). *Post Suburbia: Government and Politics in the Edge Cities*. Baltimore, MD: Johns Hopkins University Press.
Walker, R. (1981). A theory of suburbanization: Capitalism and the construction of urban space in the United States. In Dear, M. & Scott, A. (eds.) *Urbanization and Urban Planning in Capitalist Societies*, pp 383 429. London: Methuen.
Walker, R. and Lewis, R. D. (2001). Beyond the crabgrass frontier: Industry and the spread of North American cities, 1850 1950. *Journal of Historical Geography* 27, 3 19.

Relevant Websites

http://www.garreau.com
The Garreau Group: Edge City.

Edge Effects

I. Yamada, University of Utah, Salt Lake City, UT, USA

Glossary

Complete Spatial Randomness (CSR) A point pattern is considered to be completely spatial random when each point has the equal probability of occurring at any location in a study region and there is no interaction between point locations. It is often used as a baseline point distribution with which an observed pattern is compared.

Nearest-Neighbor Distance Analysis A method for point pattern analysis that utilizes inter-point distances from points to their nearest neighboring points. A short nearest-neighbor distance implies potential clustering of those points, while a long one implies a regular (or uniform) point pattern.

Power of a Statistical Test A measure of how likely the statistical test can detect a deviation from the null hypothesis (e.g., CSR in point pattern analysis) when it actually exists.

Ripley's *K*-Function Analysis One of the most common methods for point pattern analysis that is based on the distribution of inter-point distances for all pairs of points observed in a study region. Also known as the reduced second moment measure.

Spatial Filtering A technique that applies a filter function to an imagery so as to obtain an enhanced imagery or to reduce noises in it.

Edge Effects: Definition

Spatial analysis inevitably concerns itself with a finite region, a small bounded segment of an infinite space. Because of this finiteness of a study region, a boundary always exists, while any spatial phenomenon such as spatial distribution, association, interaction, and diffusion observed within the study region is most likely to extend beyond its boundary. In addition, the majority of spatial statistical theories has been developed on the basis of the infinite space assumption. Therefore, analysis confined within a bounded study region may well be biased because of the ignorance of the outside of the study region as well as the inappropriateness of the theories. This problem of potential bias in spatial analysis is referred to as edge effects (or boundary effects). Edge effects are important for any type of spatial analysis, including analysis of point and areal data, because methods for spatial analysis always require that spatial relationships between observations be defined based on their proximity, adjacency, or other criteria, which may be biased due to unrecorded observations located outside the study region.

For point pattern analysis, there are a variety of analytical methods that are based on inter point distances. When points distributed outside the study region are ignored, the nearest neighbor distance for a particular point observed within the study region may be overestimated, which will in turn distort test statistics. For instance, by applying such methods to a CSR pattern without realizing edge effects, one might falsely conclude that it was a regular pattern because of the longer inter point distances than expected for CSR.

Methods for areal data analysis often take into account neighbors of individual areal units because, for example, a crime rate observed in a particular area in a city tends to be influenced not only by characteristics of the area itself but also by those of its neighborhood. Because areal units lying along the boundary of a study region generally have their neighbors outside the study region too, such areal data analysis methods will also be affected by edge effects. If those external influences are simply ignored, results of spatial analysis will be less reliable for areal units close to the study region boundary than for those well inside. The problem of edge effects is often called the boundary value problem in areal data analysis.

Types of Edge Effects

Two types of edge effects can be distinguished in spatial statistical analysis of a finite space. The first type of edge effects occurs when one ignores observations of a spatial phenomenon that are located outside the study region. The second type of edge effects comes into play when a statistical method designed for an infinite space is applied to a finite space. A point based analysis applied to a tree distribution within a small subset of a forest will be affected by both types of edge effects (**Figure 1a**), while that applied to a tree distribution in a whole island will have only the second type of edge effects (**Figure 1b**). On the other hand, in **Figure 1c**, an area based statistic calculated for region A will not have the second type of edge effects, but will have the first type because values observed for border subzones in region A are likely to be related to values of neighboring subzones in regions B, C, D, and E.

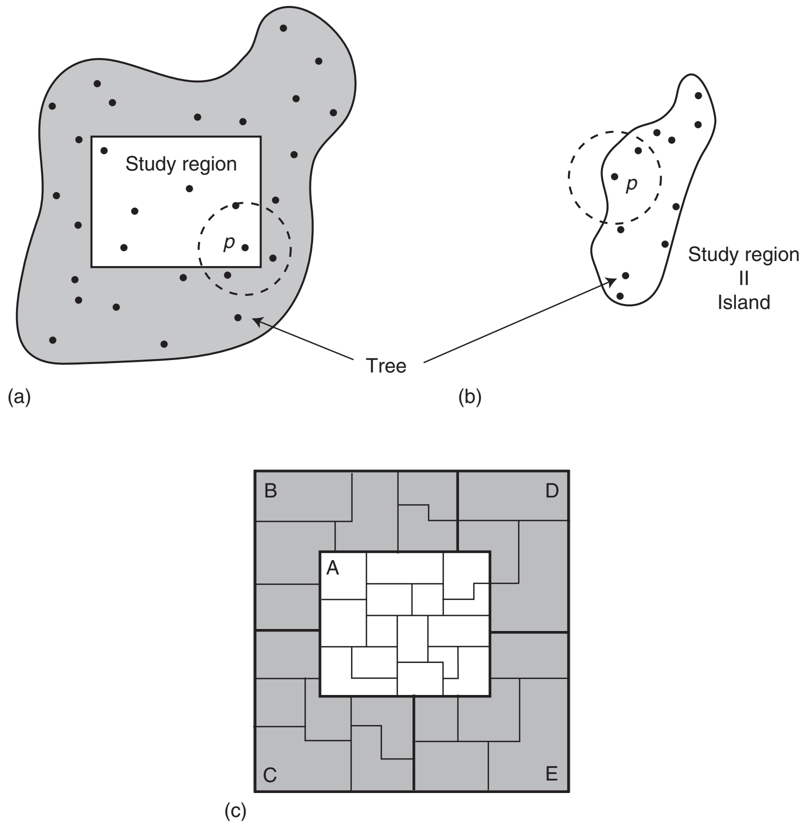

Figure 1 Types of edge effects. The first type of edge effects is due to the ignorance of observations outside the study region and the second type is due to the application of a statistical method designed for an infinite space to a finite space. (a) the first and second types combined; (b) the second type only; and (c) the first type only.

The second type of edge effects may or may not be corrected by modification of the statistical method in use. Although it is not possible to perfectly eliminate the first type of edge effects unless additional data are available for the outside of the study region, edge correction methods to be discussed in the later sections attempt to minimize the effects by modifying structures of the study region and imposing certain assumptions on the distribution of the spatial phenomenon in question.

Magnitude of Edge Effects

The magnitude of edge effects depends on both size and compactness of the study region. Let us assume that one is interested in the number of points for which nearest neighbor distances are less than a particular value, *h*, in a given point distribution. This number may be underestimated because some points may have their true nearest neighbors outside the study region. Points possibly affected by this bias are the ones that are located within distance *h* of the study region boundary, that is, within the shaded areas in **Figure 2**. Comparison of regions (a) and (b) clearly suggests that a larger proportion of points are affected in the smaller study region (b) than in the larger one (a) for the same *h*, given that regions (a) and (b) have the same intensity (or density) of points. On the other hand, the study regions (a) and (c) are equal sized while (a) is more compact than (c). Again the proportion of points affected by edge effects is much larger in the noncompact study region (c) than in the compact one (a). In general, the magnitude of edge effects decreases as the size and compactness of the study region increase. When a spatial pattern under investigation exhibits clustering tendency, locations of clusters relative to the study region boundary will also have an impact on the magnitude of edge effects.

Edge-Correction Techniques

This section summarizes edge correction techniques that are popularly used and applicable to a variety of spatial analysis and modeling methods. Also discussed here are correction techniques that are specifically designed for particular methods.

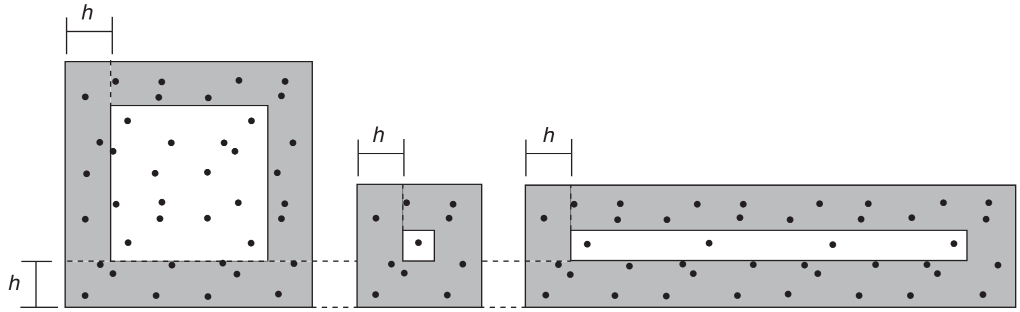

(a) (b) (c)

Figure 2 Magnitude of edge effects in relation to the size and compactness of a study region. Shaded portions represent areas that are possibly affected by edge effects at the scale of *h*. (a) compact and large study region; (b) compact but small study region; (c) noncompact study region with the same size as (a).

Toroidal Edge Correction

Toroidal edge correction, also known as translation, can be applied only when a study region has a rectangular shape. In order to circumvent the constraint of finiteness, this method maps a finite rectangular study region onto a torus, a doughnut like three dimensional shape shown in **Figure 3c**, so that the study region can be considered to be infinite. First, the northern border of the study region is connected to the southern border, which transforms the study region into a cylinder (**Figure 3b**). Then, the eastern end of the cylinder is connected to the western end, which transforms the cylinder into a torus (**Figure 3c**). Another interpretation of this correction technique is that the study region is exactly replicated eight times so as to form a 3 × 3 grid consisting of nine identical rectangles at the center of which the original study region is located (**Figure 3d**). The toroidal edge correction is permissible under the assumption that the underlying process ruling the phenomenon of interest does not differ between inside and outside of the study region.

In addition to an obvious limitation that this method is not applicable to a study region of irregular shape, some researchers have pointed out its possible weaknesses as well. First, errors in a dataset would be propagated in a test statistic because the error associated with the same observation may be used repeatedly in the calculation of the statistic. Second, the torus shape corresponds to no physical geographic situation and thus one may find it unnatural. Third, the toroidal correction as well as reflection correction to be explained below generates artificial periodicity in a spatial pattern, which is often considered to be unrealistic; thus the resulting neighborhood situations are unlikely to mirror realities.

Reflection edge correction can be seen as a variant of the toroidal correction although it is less frequently used. This method also replicates the study region eight times and creates a 3 × 3 grid, but the replicated rectangles are reflected along the four edges of the original study region (**Figure 3e**). The reflection edge correction can be applied only to rectangular study regions.

Guard-Area Edge Correction

Guard area edge correction, also known as border area edge correction, constructs a buffer zone along the study region boundary and uses the remainder in the middle as a new study region (**Figure 4a**). Although in most cases the buffer is created inside the study region (called inner guard area), it is also possible to create the buffer outside the study region (called outer guard area) if data availability allows. In the latter case the study region remains as it is (**Figure 4b**).

In point pattern analysis, points in the guard area are used only to take measurements (e.g., nearest neighbor distances) for points that are located within the study region and the guard area points themselves are discarded from subsequent analysis. The concept is illustrated in **Figure 4c**. Similarly in areal data analysis, areal units in the guard area are used as neighbors of units in the study region, but they do not appear in subsequent analysis.

A fundamental shortcoming of this approach is the information loss; that is, it needs to discard part of the sample and decrease the effective sample size, assuming that the guard area is created inside the study region. When a study region contains a small number of observations, either points or areal units, this problem becomes more severe because a higher proportion of observations will be discarded and the remaining sample size may no longer be sufficient for analysis. Imagine that the shaded part of **Figure 2b** was a guard area, which left only a single point for analysis. Another difficulty exists in determination of the optimal width of the guard area. Too narrow a guard area will fail to sufficiently remove edge effects, while too wide a guard area will lead to an unnecessary amount of data to be discarded. The

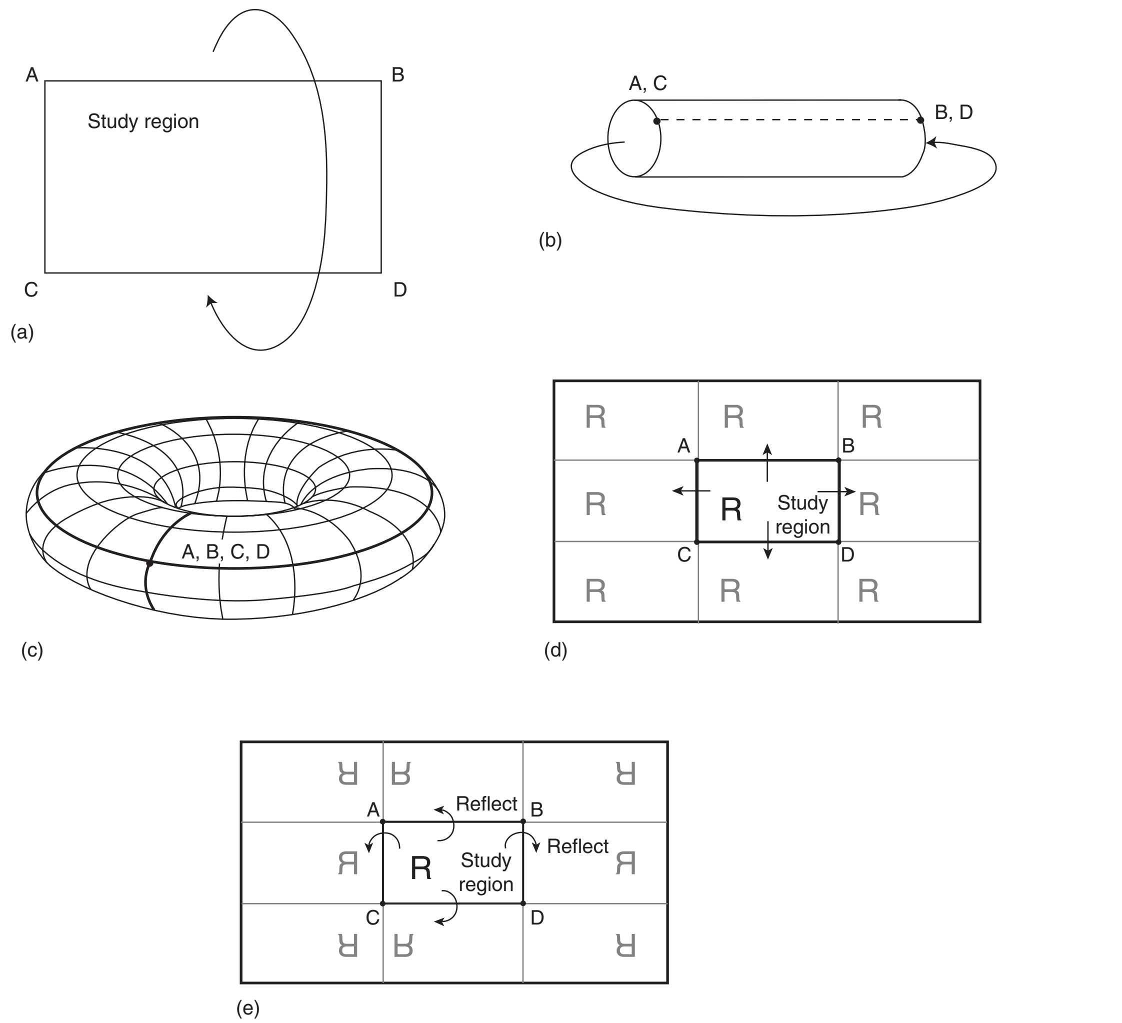

Figure 3 Toroidal and reflection edge corrections. (a) original study region; (b) study region (a) made into a cylinder; (c) cylinder (b) made into a torus; (d) toroidal edge correction; (e) reflection edge correction.

guard area correction, however, is considered to be the most realistic technique because only actually observed data are used while other techniques such as the toroidal correction artificially generate additional data around the study region.

As a special case of the guard area correction, let us consider the following analysis that is based on nearest neighbor distances between n points observed in a study region. A simple method for summarizing the point distribution is to look at the proportion of points whose nearest neighbor distance is less than or equal to a given distance, h, for a range of h. A high proportion at a small value of h, for example, indicates clustering of the observed points. Letting h_i and b_i represent distances from point i to its nearest neighboring point and to the nearest boundary of the study region, respectively, and #( ) mean 'the number of', the proportion denoted $\hat{G}(h)$ is given by $\hat{G}(h) = \#(h_i \leq h)/n$. However, this equation underestimates the actual proportion for small values of h because h_i may be longer than its true value when point i is close to the study region boundary. To eliminate this edge effect, one can calculate the proportion $\hat{G}(h)$ using only points whose distance to the nearest boundary b_i is greater than h to assure that all nearest neighbor distances included in the calculation are valid even when points outside the study region are taken into account. That is, $\hat{G}(h)$ is calculated as $\hat{G}(h) = \#(h_i \leq h,\ b_i > h)/\#(b_i > h)$. This modification can be seen as a variant of the guard area edge correction that changes the width of the guard area in accordance with distance scales under investigation.

Edge Correction Based on an Artificial External Buffer Zone

When analyzing areal data, one may construct a buffer zone outside the study region and assign it a single value such as the mean of the observed attribute values to avoid the information loss resulting from the construction of

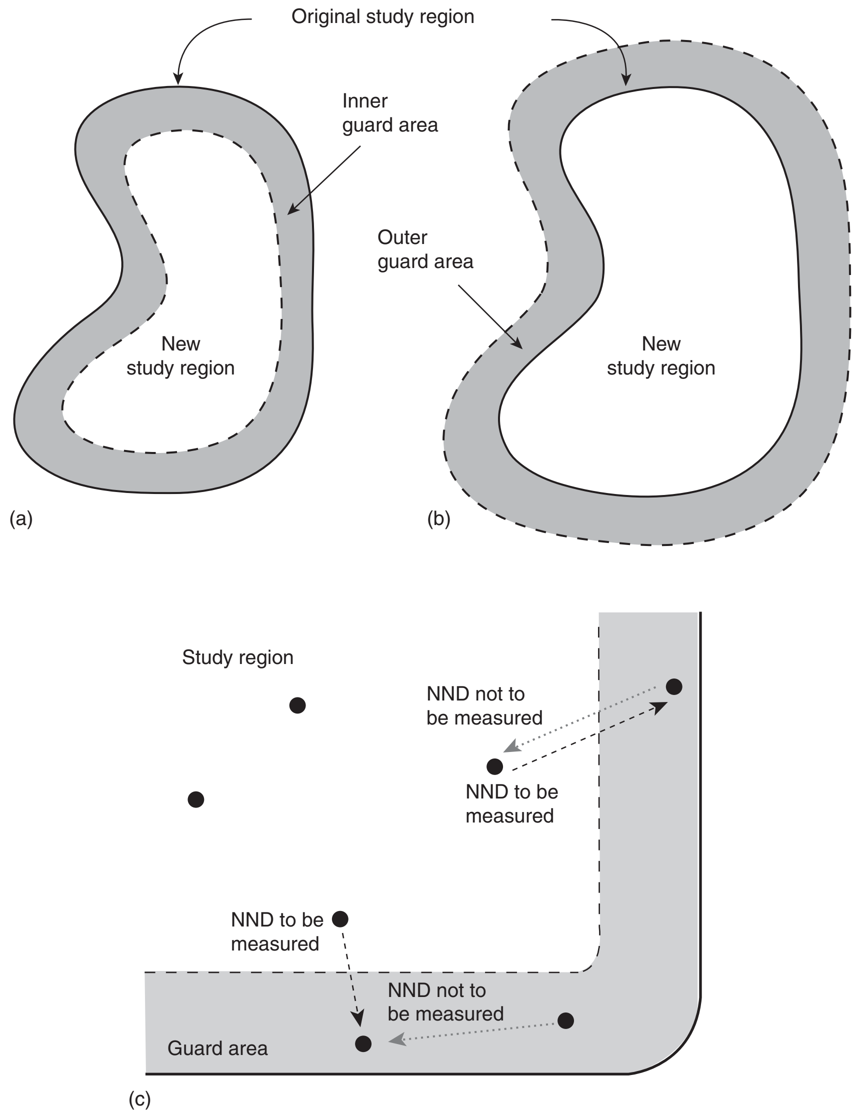

Figure 4 Guard-area edge correction. (a) inner guard-area edge correction; (b) outer guard-area edge correction; (c) application of the inner guard-area correction to nearest-neighbor distance (NND) analysis.

the inner guard area. The external buffer zone can be treated as a single areal unit or partitioned into several units so that interactions between units in the buffer zone can be incorporated. **Figure 5** illustrates the method when applied to a regular lattice. Areal units in the ex ternal buffer zone are handled in the same way as units in the guard area discussed in the previous section. The width of the external buffer zone needs to be determined in relation to the definition of neighbors.

A prominent drawback of this technique is that results are highly dependent upon the value assigned to the external buffer zone. Moreover, it barely makes sense to assume that all areal units in the buffer zone have the same value when spatial variation in the study region is of interest. This drawback may be avoided by extrapolating values to be assigned to units in the buffer zone on the basis of observations in the study region. However, it should also be kept in mind that such extrapolation itself is subject to edge effects.

Edge Correction Based on a Weighting Factor

Generally speaking, an observation that is close to the study region boundary is more strongly affected by edge effects than an observation located in the middle of the study region. It is therefore possible to weight each ob servation in relation to its proximity to the boundary in a way that the weight for an observation serves as a proxy for the degree of missing information and/or the mag nitude of edge effects at its location. It should be noted that specification of the weights may differ reflecting the

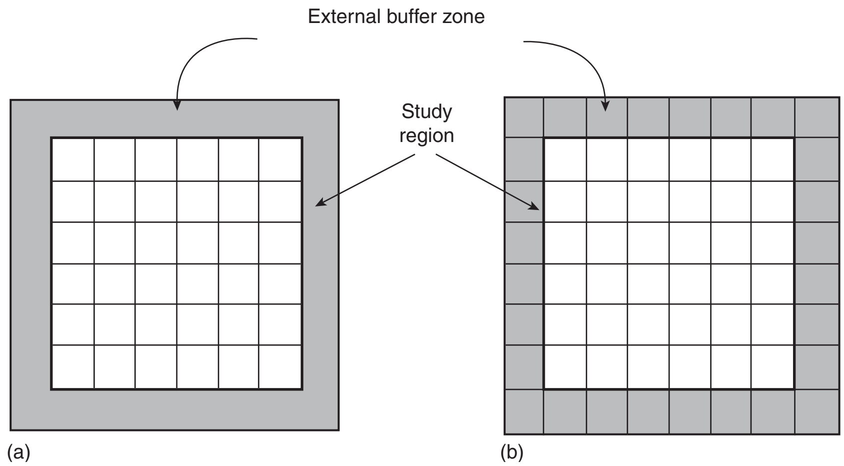

Figure 5 Edge corrections based on an artificial external buffer zone. (a) buffer as a single unit (without inter-unit interaction); (b) buffer as a set of units (with inter-unit interaction).

nature and purpose of the analysis and that it has an inevitable impact on subsequent analysis and results.

A simplest approach to reflecting the proximity to the boundary in the weight structure would be to determine the weight for observation i as a function of the distance between observation i and the study region boundary, denoted d_i. For areal data analysis, d_i may represent the distance between the centroid of areal unit i to the study region boundary. The function should be chosen in a way that an observation with a sufficiently large d_i is assigned a weight of 1 and thus fully incorporated in the analysis, while an observation with a small d_i has a smaller weight value. Another basic weighting scheme that is often ap plied to areal data analysis utilizes the ratio of the pe rimeter of unit i that is in common with the study region boundary, l_{bi}, to the entire perimeter of unit i, l_i, that is, l_{bi}/l_i. An appropriate weight function in this case is a function that takes on 1 for an observation with $l_{bi}=0$ and decreases as l_{bi} increases.

Ripley's circumference method, which is also known as Ripley's local correcting factor, is an example of weight based edge corrections. Although it is originally designed for the K function analysis, this correction method is applicable to other point pattern analysis techniques that are based on inter point distances. The K function at the distance scale of h is defined as the expected number of points within distance h of an arbi trary point in the study region divided by the intensity of points. It takes a higher value when the point distribution is more clustered than CSR, and a lower value when it is more regularly distributed. In essence, for an observed point pattern consisting of n points, one can estimate the value of the K function by first counting, for each point i, the number of other points within distance h of it and then taking the average of those n counts. When deter mining the count associated with point i, Ripley's circumference edge correction inflates the count of point j if a circle centered at i and passing through j crosses the study region boundary as in **Figure 6**, because it implies a possibility that other points at the same distance from i as j exist outside the study region. More specifically, a weight for a point pair i and j, w_{ij}, is defined as the proportion of the circumference of the circle within the study region, l_{ij}; that is, $w_{ij}=l_{ij}/2\pi d_{ij}$, where d_{ij} is the distance between points i and j. Point j is then counted as $1/w_{ij}$ points.

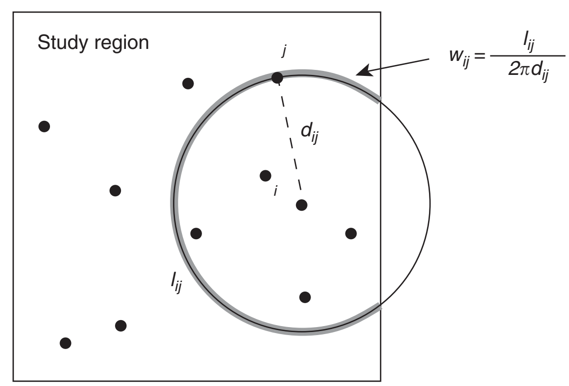

Figure 6 Ripley's circumference edge correction. Based on a circle centered at point i, passing through point j, and therefore with radius d_{ij}, the weight for j around i, w_{ij}, is defined as the ratio of the circumference of the circle within the study region, l_{ij}, to the whole circle.

Diggle, Getis and Franklin, Hasse, and Goreaud and Pélissier provide explicit formulae to implement this correction method for rectangular and/or circular study regions at various levels of details. Goreaud and Pélissier also propose a technique to compute edge corrected K function for study regions of more irregular complex shapes. Some researchers have reported that, although this edge correction provides an approximately unbiased estimator for the K function, it tends to result in an increased variation in the estimated K function compared with its unadjusted counterpart.

Monte Carlo Simulation Approach

Depending on data and/or analytical techniques to be used, edge correction methods described so far may not always be applicable. For instance, the toroidal correction cannot be used for a study region of irregular shape and the guard area correction is not appropriate for a small study region or one with a small number of observations. Under such circumstances, one might wish to simply ignore edge effects while assuming they do not exist. Although this ignorance is not uncommon, it is generally problematic unless empirical evidence suggests the absence of edge effects or the outside of the study region can be assumed to be empty. When testing of a particular hypothesis is of greater interest than obtaining an actual value of a test statistic or model parameters, a practical solution is to use the Monte Carlo simulation that imposes no or few restrictions on the analysis settings, such as the shape of the study region.

The Monte Carlo simulation approach does not attempt to eliminate edge effects. Instead, it repeatedly simulates the phenomenon of interest under the null hypothesis for a given study region, and then estimates the distribution of a test statistic in the presence of edge effects resulting from the given study region.

If an observed point pattern is to be tested against the null hypothesis of CSR, the Monte Carlo simulation approach first simulates a CSR pattern within the given study region and computes a test statistic (e.g., the average nearest neighbor distance) for the simulated pattern without any edge corrections. This process is repeated for a large number of times to generate a set of simulated values of the test statistic, which in turn is considered to be an approximated distribution of the test statistic when the null hypothesis of CSR is true and edge effects are not corrected for. Then the test statistic is computed for the given data, again with no edge corrections. This test statistic is compared with the simulated distribution; the proportion of simulated test statistic values that are greater than or equal to the observed value represents the probability of observing a test statistic that is more extreme than the observed when the null hypothesis is true. If this probability is sufficiently small, the null hypothesis of CSR can be rejected because the observed test statistic value is very unlikely if it is true.

It should be emphasized that no edge corrections are necessary in this procedure because test statistics calculated for the observed pattern and simulated realizations are both biased by the same edge effects associated with the given study region.

Comparison of Edge-Correction Techniques

There are several factors that affect the effectiveness and suitability of edge corrections. Such factors include, but are not limited to, a test statistic and/or model to be used, size and compactness of a study region, and spatial patterns embedded in data. Moreover, application of edge corrections generally imposes an underlying, sometimes implicit assumption that the phenomenon under study has the same (or at least similar) spatial patterns over the study region and its surroundings. When this assumption cannot be justified, any edge correction may result in erroneous conclusions and a possible alternative is to take the Monte Carlo simulation approach that empirically estimates the distribution of a test statistic in the presence of edge effects specific for the given study region.

Several researchers have conducted comparative studies of edge correction techniques for various settings. Some found that an appropriate edge correction could result in improvement of the power of a statistical test because it could substantially reduce instability of the test statistic, while others found opposite cases where a test without edge corrections performed at least as powerfully as ones with edge corrections. This section provides a brief summary of findings from those proceeding studies.

Point Pattern Analysis

As for point pattern analysis, method specific corrections (e.g., Ripley's circumference correction) and the toroidal edge correction were generally found to be more powerful than other corrections discussed above. However, the toroidal correction may not work well when a point distribution under investigation exhibits a regular pattern, or may lead to results that were different from what other types of corrections would lead to, especially for clustering patterns. Multiple studies have also reported that the Monte Carlo simulation approach has at least comparable and sometimes better power than other edge corrections.

When one is interested in comparing two point patterns observed in the same study region, analysis with and without edge corrections tend to yield similar results. For instance, if one examines the difference or ratio of test statistics computed for two point distributions (e.g., trees of two different species), edge effects in the test statistics would mostly be canceled out because they are both attributable to the same study region.

Areal Data Analysis

When applied to areal data analysis, traditional edge corrections do not appear to be effective enough for correcting edge effect biases, and more statistically sophisticated methods have been investigated. Such newly suggested methods include:

1. the dummy variable method that statistically discriminates between border and nonborder areal units while considering edge effects to be a spatial process that possibly works differently in the border and nonborder areas;
2. the generalized least squares method that mathematically determines a transformation matrix to adjust for edge effects in a similar way that a boundary value is treated in time series analysis; and
3. the missing data method that recasts the edge effect problem as a problem of missing values along the study region boundary and statistically estimates values for the external buffer zone; this method can be seen as an extension of 'edge correction methods based on an artificial external buffer zone' discussed above.

The generalized least squares method was found to be a promising approach to removing edge effect biases on the basis of simulation experiments. The missing value method was also reported to effectively reduce edge effects encountered in spatial filtering process.

Discussion

As the previous section suggests, no panacea for the problem of edge effects has been found yet although a wide variety of correction techniques have been extensively studied since 1970s. It appears unlikely that such an optimal solution would be developed because of the complex structure of edge effects. How strong edge effects would be and whether or not they could be removed from analysis depend not only upon the configuration of a specific study region and the structure of an underlying spatial process, but also upon statistical and/or modeling techniques used. Consequently, the effectiveness of any edge correction is also strongly associated with those factors. Nonetheless, it is important to establish some guidelines for selecting an appropriate edge correction in a given situation characterized by the combination of those factors. It would be possible by accumulating empirical findings as discussed in the previous section, while a more systematic, standardized approach is needed to achieve a more comprehensive understanding of the edge effect problem.

In the mean time, the Monte Carlo simulation approach seems to be a promising alternative when the objective of an analysis is to test a hypothesis about a spatial process rather than to obtain an actual value of a test statistic or model parameters. It essentially eliminates the need for edge corrections and also offers an opportunity to test null hypotheses more complex than CSR or other basic ones, for which the distribution of a test statistic may not be theoretically obtained.

See also: Monte Carlo Simulation; Point Pattern Analysis; Spatial Autocorrelation; Spatial Clustering, Detection and Analysis of; Spatial Filtering/Kernel Density Estimation; Spatially Autoregressive Models.

Further Reading

Baddeley, A. J., Kerscher, M., Schladits, K. and Scott, B. T. (2000). Estimating the *J* function without edge correction. *Statistica Neerlandica* 54(3), 315 328.
Cressie, N. (1991). *Statistics for Spatial Data*. New York: Wiley.
Diggle, P. J. (1983). *Statistical Analysis of Spatial Point Patterns*. London: Academic Press.
Ge, Y., Cheng, Q. and Zhang, S. (2005). Reduction of edge effects in spatial information extraction from regional geochemical data: A case study based on multifractal filtering technique. *Computers and Geosciences* 31, 545 554.
Gignoux, J., Duby, C. and Barot, S. (1999). Comparing the performances of Diggle's tests of spatial randomness for small samples with and without edge effect correction: Application to ecological data. *Biometrics* 55, 156 164.
Griffith, D. (1983). The boundary value problem in spatial statistical analysis. *Journal of Regional Science* 23, 3277 3387.
Griffith, D. (1985). An evaluation of correction techniques for boundary effects in spatial statistical analysis: Contemporary methods. *Geographical Analysis* 17(1), 81 88.
Griffith, D. and Amrhein, C. (1983). An evaluation of correction techniques for boundary effects in spatial statistical analysis: Traditional methods. *Geographical Analysis* 15(4), 352 360.
Goreaud, F. and Pélissier, R. (1999). On explicit formulas of edge effect correction for Ripley's K function. *Journal of Vegetation Science* 10, 433 438.
Haggett, P. (1981). The edges of space. In Bennett, R. (ed.) *European Progress in Spatial Analysis*, pp 51 70. London: Pion.
Hasse, P. (1995). Spatial pattern analysis in ecology based on Ripley's K function: Introduction and methods for edge correction. *Journal of Vegetation Science* 6, 575 582.
Lancaster, J. and Downes, B. (2004). Spatial point pattern analysis of available and exploited resources. *Ecography* 27, 94 102.
Lawson, A. B., Biggeri, A. and Dreassi, E. (1999). Edge effects in disease mapping. In Lawson, A. B., Biggeri, A., Bohning, D. *et al*. (eds.) *Disease Mapping and Risk Assessment for Public Health*, pp 85 98. New York: Wiley.
Ripley, B. (1981). *Spatial Statistics*. New York: Wiley.
Ripley, B. D. (1979). Tests of randomness for spatial point patterns. *Journal of the Royal Statistical Society, Series B* 41(3), 368 374.
Yamada, I. and Rogerson, P. A. (2003). An empirical comparison of edge effect correction methods applied to *K* function analysis. *Geographical Analysis* 35(2), 97 109.

Relevant Websites

http://www.spatialanalysisonline.com
Geospatial Analysis (online book).

Education

N. Lewis, The University of Auckland, Auckland, New Zealand

Glossary

Positional Good A good that gives its acquirer advantages relative to others (shuffles its holder up a social hierarchy), as in the value of education at a particular school and hence its quality is dependent upon how that school compares to others rather than on discrete measures of quality (the benefits from consuming that education are entirely at the expense of others).

Production of Space A conceptual framework for understanding geographical configurations as outcomes of the interplay of the lived practices of people in their places, the material relations around which these are arranged, and representations about how relations between people and their places are and should be made (including, importantly, policy settings).

Introduction

In teaching geographies of education as a block in a third year social geography course, the author finds himself compelled to begin the module by posing the question, "why might geographers be interested in education?" While an excellent, discipline affirming way to begin any such module, it shocks the author to think that he needs to convince himself and potentially skeptical students that education is something that should be taught in geography. Such skepticism is clearly unwarranted and generally short lived in the classroom, but is borne of a dearth of geographical scholarship on education. The work of Ron Johnston and William Clark on the role of schools in urban segregation notwithstanding, there has been little sustained geographical interest in education as a site or subject of study. While prominent geographers have made contributions over many years to the rich geographical education literature, education geography has remained marginal.

The marginality of geographies of education is surprising given the centrality of education (formal and otherwise) in social and spatial processes from subject formation; to community and neighborhood change; the production of city, regional, and national identities; the generation of geographical imaginaries; economic growth and advantage; and questions of social order. If cultural geography is about identity and subjectivity and political geography is about influencing people in places, and economic growth is about knowledge production and dissemination, then the absence of thriving geographies of education is a hole in the disciplinary attention. This article draws on writing within the sociology of education as well as geography to consider what we know about the role of education in the production of space and the implication of spatial difference in the generation of education spaces. It argues that geographers might extend this knowledge in productive directions. The article begins by outlining the centrality of education as a social institution, and then highlights its political significance. The subsequent section puts geography into this politics, and identifies key lines of enquiry in what might be termed the political geography of education. The final section confronts the ways in which geographers are tackling these geographies.

Politics of Education

Education is bound up tightly with the production and dissemination of cultural, social, and economic knowledge. Formal education is a formative set of institutions for the production of identities (local, national, class, global, values based, and so on) and formation of subjectivities as well as economic and social resources. Schools, in particular, are perhaps the only social institution through which we all pass, and the core problems of the state are writ large across them. They have important economic functions, and play enduring and central roles in patterns of socialization, social control, and cultural reproduction. If the object of politics is to change attitudes, beliefs, conduct, and knowledge, and the method is to produce a curriculum of political texts and discourses that establishes values, promotes desired conduct, and justifies action then education institutions are crucial political arenas.

Education is implicated in struggles from those over social organization and belief systems right down to micro struggles over neighborhood transportation routes. Education institutions are expected to deliver the subjectivities upon which liberal democracies or religious or revolutionary socialist states are founded. Schools are expected to ensure the safety of children in and around their grounds. There is always an extensive set of social, cultural, and economic purposes claimed for education. Commonly defined poorly or rhetorically and overdetermined by the dominant governmental rationalities of time and place, they are transformed through the political process into particular sets of expectations and policies to achieve them in any setting. They generate multiple contests over models of curriculum, pedagogy, and organization and associated micro practices.

Struggles over who gets taught, what, where, how, and by whom have both a surface politics (education politics) and a deeper structural and ideological politics involving the meaning, purpose, and potential of education (the politics of education). Micro struggles over curriculum, pedagogy, or organization are commonly embedded in, expressions of, framed as, or roughly aligned with, meta ideological positions or meso expressions of them in political projects such as literacy, universal education, the three R's, life long learning, community education, or the knowledge society. Although inevitably influenced by situated practices, surface politics, and conflicting the ories, debates over the content and delivery of education and its spatial organization are always deeply political in inspiration, conduct, and effect. The ideological and political terms of debate tend to be close to the surface.

One key axis of struggle in the politics of education is its contradictory potential to produce knowledges, sub jects, and measures of worth that may be either eman cipatory (however differently conceived) or conservative (even repressive). In the Global South, the challenge has been to extend the privilege of education and realize its potential for positive change. In the discourses of Paolo Freire, the Zapatistas, Islamic states, and more recently the Bolivarian Hugo Cesar, education is clearly under stood as subjectifying, and as knowledge production in very broad terms. Its potential and purpose are seen as revolutionary. In these discourses, education should ideally be imagined, delivered, and governed by com munities. At the other extreme, neoliberal discourses see education as the production of individual opportunity – a pivotal moment in the production and valorization of human capital. They insist that its provision in formal institutions should be driven by consumer preference guided by markets (education and labor markets). To the extent that it generates individual advantage, it should be paid for by individuals who are in effect investing in themselves. For welfare states, education is pivotal but means something different again, the geographies of which are explored below.

States are, as a result, heavily invested in education, whether the political projects in play are imperialist or revolutionary, rooted in the development state or glob alization, promoting the production of citizens and workers for Fordist labor processes or global citizens and creative classes for knowledge societies. States may own, deliver, and/or regulate education, be it schooling, ter tiary education, or (increasingly) preschooling. At the International Geographical Congress in Seoul in 2000, David Harvey made one of the few statements by geographers on the politics of education. He observed that through their education agendas and programs, states orchestrate their efforts to secure their power and impose normative programs on the production of space. In these terms, Harvey argued that state education is a major influence on territoriality, identity, and geo graphical distributions of population, economic activity, social services, wealth, and well being.

Justified by the special case of the public good elements of 'human capital', neoliberal states have re mained heavily invested in education despite bringing to bear various market disciplines. Education has re mained a space where neoliberal subjectivities can be fostered and potentially problematic persons identified and worked upon. In projects of the knowledge economy and the creative city, formal education in schools as well as universities and technical institutes is being integrated into and reconfigured in the terms of national systems of innovation. Students are taught the technologies and instilled with the dispositions that will facilitate the in dividual, city, and national success in the global know ledge economy. Education for citizenship programs seek to graft the environmental ethics of global warming and an ethics of diversity onto the innovative and entre preneurial dispositions commanded by economic na tionalism, at the same time as 'life long learning' is deployed to legitimize the unstable work regimes of postindustrialism.

The Global South is being subjected to the content and rationalities of these policy discourses in terms of good governance conditionalities and expert inter ventions by World Bank consultants, United Nations officials, returning students, and aid donors. At the same time, formal education remains the panacea for enduring discourses of the development project and is cast as community education for an alternative development in more revolutionary settings. In all settings and places, education is a positional good that presents individuals and families the opportunities for social and economic participation in labor markets and paths for social and spatial mobility. It has underpinned patterns of social and spatial mobility over the second half of the twentieth century, presenting pathways of hope and aspiration for poorer families, migrants (inter and intranational), eth nic minorities, and women generally.

However, radical alternatives notwithstanding, edu cation is not a universal, stable, or simply defined good. It is more than certified knowledge acquisition, a bundle of workplace skills, or even a mechanism for sorting and attributing values to human capital. It subjectifies, but does so unevenly and in articulation with other social institutions, individual dispositions, and social distri butions. It delivers highly uneven bundles of skills and identities. There can be no guarantee that achievement is either a measure of capability, or that it will be used as such at any site or moment of ascription. Formal education is a highly problematic basis for social sorting – one that is heavily conditioned by existing distributions of social privilege and that in turn inscribes new inequalities. Spatial unevenness is a key factor in its workings.

Political Geographies of Education: Welfare State Systems as Example

This narrative of the politics of education reveals multiple geographies of education that are both the product of the politics of education and constitutive of the possibilities and projects that define it. Each of the key questions underlying the politics of education (who gets taught what, where, how, and by whom) has pivotal geographical dimensions, while the co resolution of these questions in any education system has a significant bearing on the production of space. Some of this political significance can be teased out by focusing on the geographies of welfare state programs in education.

Geography is central to making decisions about which students attend what institutions and where. These decisions have a certain underlying central place logic, particularly where the state is the supplier of schooling to children, who must, by law, attend schools. Demand for schooling is universal and tied directly to population structure. Informed by welfare state governmentalities of spatial equity, postwar planning in much of the west attempted to lay down such a pattern of schooling. Aided by the flexible thresholds of schooling, schools were located as if the social plain was flat, subject to the irruption of unavoidable topographic features (physical and political). The aim was to ensure equalities of access, with school zones (formal or socially accepted) matching students to schools, and universal curricula, national examinations, standardized teacher training, and other institutions regulating a national equity in access. Universities and technical institutes were also organized to serve, and seen as serving, catchments, but at regional or city scales. They required greater mobility from students, but institutions such as accommodation allowances, hostels, and scholarships evened access, particularly where movement across catchments was necessary for specialist courses.

Educational space was about catchments (formal and perceived), which tied communities to their educational institutes and made them community institutions. With education seen as standard, spatial equity was a question of access. Free and universal education in many ways defined welfare states and the nations built around them. Education was a primary mechanism for delivering spatial equity, leveling opportunity space, and redistributing opportunity across space and social strata. Schools, especially, became sites of a great social engineering project as states intervened in the intergenerational transfer of opportunity, sought to redistribute opportunities, and built national identities. Social advancement was to be achieved by learning, and talent balanced by jobs for all. Welfare state education systems released the great energies and aspirations of the new middle classes and in turn the great high modern advances in knowledge production. At the same time, they provided the social democratic glue of apparent equity and secured national citizenship in different ways (from debating democratic concepts to teaching national constitutions, raising flags, and reciting oaths of allegiance).

However, as sociologists of education recognized and social geographers began to demonstrate in their work in other areas, space is not flat, all places are not the same, schools are power driven institutions, identity formation is complex, and there is no uniform student. Equity in education is not simply a question of central place planning. Neither the social nor the educational experiences of schooling are uniform, and social mobility is not simply ascribed by educational qualifications. Geography and the relationships among places, their social make up, and their schools are important. Even if curriculum and pedagogy could be delivered uniformly across space, education is a social experience generated by the working out of social relations in place (schools). Formal education is a positional good. Who is in the classroom is crucial to the education experience, as are the legacies of earlier groups lodged in the physical infrastructure of the school, the practices of teachers, and the reputation of the school.

'Where', then, is a crucial question in the politics of education. Schools are in large part the sum of their children. Their composition is crucial. If places are different (socially, environmentally, ethnically, and culturally), then so too will education experiences (the mix of pedagogy, curriculum, physical space, and 'hidden curriculum' of place). Universal curricula and messages about identity will be delivered and received differently. Schools become expressions of their communities, and education a product of place. The positional value of education will also have crucial geographies. Like urban neighborhoods, schools are spaces produced by the interplay of their students and teachers, their cultural practices and dispositions, the material resources of schools and parents, particular processes of community formation, and the discourses of education, pedagogy, administrative practices, and curricula imposed on them. Geographers are well placed to understand schools in these terms.

Despite the widespread gains in opportunity under welfare states, educational outcomes were far from equal. Underachievement along class, ethnic, and spatial lines reproduced deeper social inequalities. Sociologists of education contested the sources of educational advantage and disadvantage, reminding us that what counts as education is itself politically constructed. Marxian sociologists drew on Bourdieu's notion of cultural capital to explain the persistent underachievement of certain groups by reference to the class and racial biases embedded in education programs and school practices. Although arguing over the difference that schools can make or how exactly cultural capital relates to learning, they drew attention to the relationship between education and the production, concentration, and reproduction of spatial inequalities.

From a geographical perspective, these debates are over the production of space. Bourdieu's work is underlain by complex geographies of schooling that are pressing and involve the co constitution of education institutions and space. Neoliberal reforms to education systems have in many places canceled commitments to spatial equity and fractured relationships between community and school. With competition and individual responsibility the new dominant governmental rationalities and regimes of choice have replaced zones and the presumptions of uniformity. Now made more fully responsible for their children's success, actively choosing parents clearly recognize education as both social relation and positional good. In selecting for social attributes and educational reputations (often collapsed, where social composition is proxy for educational quality) they are making choices that have resulted in new patterns of segregation and concentrations of educational advantage and disadvantage in schools, and new social and spatial inequalities in place.

Universities and technical institutes now compete openly for students in quasi markets within and across national systems. Schools are forced to compete from positions largely determined for them by their place. State schools now compete with quasi state and increasing numbers of private alternatives, while catchments are confused further by the uneven access to schools of different qualities produced by different endowments of financial and cultural capital. The evidence in many settings suggests that while many have benefited from the opportunity to escape the tyrannies of place (particularly in poor urban neighborhoods), concentrations of underprivileged are becoming more pronounced, while the educational prospects and social futures of many children removed from their communities are uncertain.

Efforts by states to guide these 'markets' and alter the patterns of advantage and spatial difference emerging by voucher systems, charter schools, or reintroducing partial zoning are adding new layers of complexity. In constituencies where education has deeper traditions of private provision, these complexities are already inscribed into institutional and material geographies and educational discourses. In all places, the relationships between education, spatial difference, and its reproduction have produced new patterns of ethnic segregation and governance structures. Events such as the closure of rural schools and amalgamation of urban schools demand that new geographies of education be written and geographers are beginning to respond.

Written Geographies

Education institutions are thus both geographically produced spaces and space forming. Schools are creatures of their communities, neighborhoods, and cities, yet also pivotal in their production, while schooling is a crucial cultural and socioeconomic activity. In this light, surprisingly few geographies of education have been written or research agendas built around education. We might expect geographers to have interests in the way in which the questions of 'who' and 'where' are co constitutive with what gets taught, how, and by whom – and in the consequences. The baseline geographies of delivery are a place to start – who goes where, and to what effect in terms of the production of space. However, we might also expect geographers to have asked how and why these patterns emerged, and how they are related to other processes of social formation, such as class, ethnic segregation, cultural, and economic change and urban form. They might also be interested in what these geographies tell us about broader issues of societal composition and social transformation, or concepts such as equity, justice, or democracy. More generally, education offers a productive entry point into reflections on different dimensions of the production of space, particularly those to do with identity and subjectivity. While geographers have been slow to explore this potential, several emergent research agendas promise to give education greater prominence in the geographical literature.

Geographies of Geographical Education

Geographers have long been concerned with the teaching of geography, both in schools and universities. Both have generated rich literatures exploring the geographical knowledges produced by education institutions and related practices. The *Journal of Higher Education in Geography* (*JHEG*), in particular, has provided the basis not just for reflection on the teaching of geography and the nature of geographical knowledges, but also routinely locates these in the changing geographies of its provision – policy changes, globalization, the relationship between the emergence of ideas and material change, issues of social justice (particularly their spatialities), and projects to decenter the imperialism of Anglo American geography. Commentators have been interested in the geographical imaginaries generated in the production of both knowledge and subjects in schools and universities. Much of what we know as geography is practiced and produced in educational institutions, which are also where the geographical imaginaries that shape the world are co constituted with other knowledges. The geographies of the provision of geographical education are very important.

Geographies of School Composition

Geographers have also studied spatial patterns of school attendance, and in doing so, confronted to some degree the political geographies of school composition. Working with quantitative methodologies, geographers in the 1970s and 1980s examined the role of schooling in processes of urban segregation. Engaging with policies such as 'school bussing', which moved children between neighborhoods to counteract racial segregation in schools within administrative boundaries, the literature provided evidence for policy debates and theorizations of urban processes. It engaged with a politics of social justice around tyrannies of place in poor urban neighborhoods and the virtues of integration (perhaps now more likely framed as diversity). Studies have highlighted the role of schooling in increasing segregation within cities, its interplay with income differentials, cultural attitudes, and real estate markets and the difficulties of intervening through policies such as bussing.

This literature, its protagonists, politics, and methodologies, has resurfaced in recent years in research collaborations between educationists and geographers in Bristol, Cardiff, and Los Angeles. It is now concerned increasingly with a critique of post zoning, competitive, and choice based strategies that seek to match students with schools in state systems. That is, it examines the effects of systems of increased parental choice on school composition, success in schooling, and residential location. The literature is again concerned with measuring who goes where, but must now confront active choices and the strategies of parents as well as the changing ethnic and class composition of different schools. Studies confirm just how important education is in family strategies, and how significant schools are in stitching together the social fabric of cities. They direct attention to the effects of schooling policies on urban forms and life worlds. The British studies reveal, for example, that the preferences of different minority groups for ethnically dominated schools have resulted in greater ethnic segregation in schools. The use of geographic information system (GIS) techniques has enhanced this work, while also offering a valuable field of study for the application of these techniques by critical geographers. Ideally, this literature will now engage in a more sustained dialog with sociologists of education examining what actually happens in schools, the social and educational outcomes generated, and the implications for the production of space.

Schools and Their Communities

In rural communities and tightly bound urban neighborhoods, schools are central social institutions. Where catchment boundaries are more open, relationships between schools and communities will take different forms, but schools are still places where individuals cross paths and share experiences across social groups and generations. They are often the places where the community meets for other social purposes and become repositories of historical practices and community memory and store historical resources in their physical and social fabric. These memories and social exchanges condition the behavior of teachers, students, and parents, such that communities come, in turn, to define the nature of the education experience in their schools. Studies by geographers of programs of school closures in New Zealand and Canada confirm these relationships. Although yet to be studied in the same way by geographers, it is clear that tertiary education institutes can also come to define their communities, especially where they are dominant employers in regional centers, university towns, or even major centers. The geographies of hi tech clusters and creative cities in many parts of the world are in a sense geographies of education.

Children's Geographies of Schooling

The school is, alongside the family, an important institution in the geographical worlds of children and young people. Influenced by feminist perspectives, cultural studies, and a moral stance on the rights of children, these geographers of children concentrate on worlds that take place in and around schools and where schools are often a significant presence. Although often shadowed in the written geographies of children, school experiences are pivotal to the lived geographies of children. They generally have a background presence in this work, although geographers such as Gill Valentine have studied schooling directly. The spatial imaginaries of young people, their classroom experiences, the permeable boundaries between school and after school for younger children, the problems of getting to and from school, children's resistance to various forms of subjectification, the significance of school spaces such as school grounds, and the implication of schools in child health, all appear in this literature but demand more sustained research.

Schools and Identity and the Production of Citizenship

Work emanating from the Canadian scene and from interests in both geographies of school composition and those of mobility and citizenship has in recent years directed attention to the production of citizenship in schools. Katharyne Mitchell and Ranu Basu have both grappled with the implications for identity formation and citizenship of what is taught to whom and how. Mitchell concentrates attention in this regard on the production of social space related both to the experiences of diaspora and political projects of multiculturalism. Echoing work on the politics of education, she shows how education is a sphere of social life in which projects of multiculturalism can be connected to projects of economic development by mobilizing diversity for competitive advantage. In her later work she reconnects these interests to a critique of neoliberalism in the formation of European education policy. This connection with neoliberal critique, echoed in Basu's work, is one of several connections to other emergent geographies of education – children's geog raphies, school composition and closures, international izing education, and neoliberal governmentalities in education.

Internationalizing Education

Geographers have also become interested in the social and knowledge spaces generated by both neoliberal re form and the internationalization of education – com monly in concert. Publishing alongside educationists in the journal *Globalisation, Societies and Education*, they are beginning to ask questions about the emerging networks of flows and their political economy, the involvement of private enterprises in the unbundling of commodity chains, the use of benchmarking, branding, and audit technologies to neoliberalize universities and constitute a global industry, the strategies of education institutions as economic agents, the geographical imaginaries at work and the remaking of geographical knowledges, and the realignments of geographical centers of enquiry and power within the discipline. In their own performance of globalizing education, political and social geographers are now working alongside educationists under the um brella of the World Wide University Network on a re search project labeled *Constructing Knowledge Spaces*.

Geographers interested in changing urban forms, built and social, have begun to explore the implications of large scale student movements in the production of these forms in destination cities – in the investment decisions of property developers, the retail signage that adds color to cities, the peopling of the city, the languages spoken, spectacles performed, and the social networks estab lished. Some have begun to explore education in terms of global city formation. Questions are being to be asked about a global battle for brains, the neocolonialism of the knowledges produced, and the formation of a global elite class. Some of this work lies embedded in the discipline's reflection on its own practices in the *JHEG*, but work in this area would benefit from insights drawn from devel opment geography. In an important cross over with work on the internationalization of education, a network of scholars centerd on Singapore has begun to ask questions about altered family lives, the stretching of families across borders, the writing of international education into migration strategies, and the formation of new global social and economic networks. The aim is to bring re search on the mobility effects of international education into dialog with concerns over the battle for brains in the knowledge economy and creative city strategies of globalizing cities.

Conclusion

The shift by geographers to confront the globalization and neoliberalization of education in both schooling and university systems is an exciting field of geographical enquiry. This interest is often framed by an interest in their own practices as teachers and researchers, and ex tends to the interests of political geographers in edu cation as a space in which to examine the ways in which government agencies influence decision making in mul tiple and dispersed settings. However, this recognition of the treatment of schools and universities as economic agents has yet to tempt economic geographers to con sider many of the themes in which they show great interest – from clusters to knowledge transfer, the boundaries of the firm, governance, and so on. Again surprisingly, geographers have been slow to investigate the micro spaces and the micro practices of the pro duction of space in schools and universities. They are sites where policy representations and agendas come into direct confrontation and articulation with the ingrained practices of education, the social practices of young people, the material social relations of place, and estab lished physical architectures.

A thriving education geography might like that the field of health geography be based on a sustained dialog between different geographical subdisciplines and edu cationists. Alternatively, it might be a more disparate set of explorations of the relationship between education and the production of space in different fields of geographical enquiry (political, economic, cultural, or social), where education places, institutions, and processes are objects for analysis in their own terms or for other enquiries. If this were to be the case, then a meaningful and partially cohesive education geography might be fashioned from periodic dialogs across these enquiries. In either of these scenarios, the value to the discipline of focusing more attention on so central and rapidly changing a social institution cannot be overstated. Education is central to questions of social justice, and its emancipatory potential has not been fully squashed by neoliberalism. The lessons of children's geographies and mobility studies, and work on parental choice in schooling suggest that geographical enquiry from participatory action research to policy critique can make a significant difference. The opportunities to provoke and promote more just out comes from applying the ethics, perspectives, and prac tices of geography to education should not be lost.

See also: Children/Childhood; Citizenship; Community; Cultural Capital; Segregation.

Further Reading

Andersson, A. and Subramanian, S. V. (2006). Explorations of neighborhood and educational outcomes for young Swedes. *Urban Studies* 43, 2013 2025.

Apple, M. (1982). *Education and Power*. Boston, MA: Routledge and Kegan Paul.

Basu, R. (2007). Negotiating acts of citizenship in an era of neoliberal reform: The game of school closures. *International Journal of Urban and Regional Research* 31, 109 127.

Bourdieu, P. and Passeron, J. C. (1977). *Reproduction in Education, Society and Culture* (transl. Nice, R.). London: Sage.

Burgess, S., Wilson, D. and Lupton, R. (2005). Parallel lives? Ethnic segregation in schools and neighbourhoods'. *Urban Studies* 42, 1027 1056.

Clark, W. A. V. (1987). Urban restructuring from a demographic perspective. *Economic Geography* 63, 103 125.

Croft, J. (2004). Positive choice, no choice or total rejection: The perennial problem of school catchments, housing and neighbourhoods. *Housing Studies* 19, 927 945.

Dale, R. (1989). *The State and Education Policy*. Milton Keynes: Open University Press.

Hankins, K. and Martin, L. (2006). Charter schools and urban regimes in neoliberal context: Making workers and new spaces in metropolitan Atlanta. *International Journal of Urban and Regional Research* 30, 528 547.

Holloway, S. L. and Valentine, G. (eds.) (2000). *Children's Geographies: Playing, Living, Learning*. London: Routledge.

Johnston, R. J. (1984). *Residential Segregation, the State and Constitutional Conflict in American Urban Areas*. London: Academic Press.

Johnston, R., Wilson, D. and Burgess, S. (2005). England's multiethnic educational system? A classification of secondary schools. *Environment and Planning A* 37, 45 62.

Lewis, N. (2004). Embedding the reforms in New Zealand schooling: After neo liberalism? *GeoJournal* 59, 149 160.

Mitchell, K. (2003). Educating the national citizen in neoliberal times: From the multicultural self to the strategic cosmopolitan. *Transactions of the Institute of British Geographers* 28, 387 403.

Mitchell, K. (2006). Neoliberal governmentality in the European Union: Education, training, and technologies of citizenship. *Environment and Planning D: Society and Space* 24(3), 389 407.

Novelli, M. (2006). Imagining research as solidarity and grassroots globalisation: A response to Appadurai (2001). *Globalisation Societies and Education* 4, 275 286.

Olds, K. and Thrift, N. (2005). Assembling the global schoolhouse in Pacific Asia. In Daniels, P., Ho, K. C. & Hutton, T. (eds.) *Service Industries, Cities and Development Trajectories in the Asia Pacific*. London: Routledge.

Waters, J. (2006). Emergent geographies of international education and social exclusion. *Antipode* 38, 1046 1068.

Witten, K., Kearns, R., Lewis, N., Coster, H. and McCreanor, T. (2003). Educational restructuring from a community view point: A case study of school closure from Invercargill New Zealand. *Environment and Planning C: Government and Policy* 21, 203 223.

Electoral Cartography

A. Kavanagh, National University of Ireland, Maynooth, Republic of Ireland

Glossary

Cartogram Map which represents different regions not according to their land areas, but according to some other thematic mapping variable, such as population, thus distorting the shape of these regions to take account of other variables.

Electoral Geography A branch of geography, within the subdiscipline of political geography, which is especially concerned with studying the geographical dimension of electoral results, as well as of how elections are organized and conducted.

Gerrymandering The practice by which the shape and size of electoral boundaries is deliberately changed to benefit one political grouping (usually a governing party) at the expense of others.

Geographically Weighted Regression Technique for exploratory spatial data analysis, which allows for the measurement of spatially varying relationships (thus showing how relationships between the dependent and predictor variables might vary over space).

Géographie Électorale Early work in the field of electoral cartography and geography carried out in France in the early 1900s and particularly associated with French geographer, André Siegfried.

Marked Register Turnout Figures Turnout figures calculated, usually for very small areas, based on the study of the marked electoral registers used in specific electoral contests, thus addressing the dearth of small-scale electoral data that emerges in politicians where such data is only available for (relatively large) constituency units. Such figures may be calculated to the level of the smallest areal units used for the purposes of census returns (such as electoral divisions, in the case of the Republic of Ireland).

Redistricting Term particularly associated with American politics, mainly involving the making of changes to the boundaries of election constituencies, usually in order to take account of population change within a state as identified in the publication of census reports. In states such as the Republic of Ireland and the United Kingdom, independent commissions take reponsibility for the redrawing of election boundaries, whereas in other states this responsibility may lie in the hands of government politicians, a factor that may give rise to the practice of gerrymandering (see above).

Electoral cartography, concerning the mapping of the various factors associated with elections, including election results, voter turnout levels, and party mobilization efforts, is a key aspect of electoral geography. Studies within the areas of electoral cartography and electoral geography highlight the key role of geography in elections; stressing that voters do not express their preferences within a vacuum, and that political decision making processes are influenced, to varying degrees by the contexts – national, regional, and local – that such processes are inserted within. Electoral cartography ensures that a spatial, or contextual, viewpoint remains central to studies of electoral participation, voter choice, and the operation of electoral systems, and stresses that place matters in those processes, an argument put forward by leading geographers in this field, such as John Agnew, Ron Johnston, and Chris Pattie.

Much research is concerned with the mapping of electoral patterns – such as area based electoral participation or party support figures – and it attempts to explain these trends with reference to various causal factors. The mapping of electoral patterns particularly formed the key focus of the earlier work within the discipline, as most notably associated with the French regional school and the 'géographie électorale' of André Siegfried in the early 1900s. Siegfried's research, as with many other subsequent studies in electoral geography, involved a technique of map comparison, wherein the degree to which the spatial patterning of election results for different French political parties could be associated with the spatial trends associated for other, causal, factors was studied. Siegfried's (1913) classic study of voting trends in the Ardèche region of France was accused of being environmentally deterministic in nature, given his identification of particular voting patterns for different physical landscapes, but this ignores how linking the physical and political environments allowed Siegfried to highlight how different landscapes helped produce different socioeconomic conditions, which in turn acted to shape the political orientation of voters within different French regions.

Little work was carried out within the field of electoral geography (and electoral cartography) outside of France until the 1960s and the emergence of geography's quantitative revolution. With this, Siegfried's cartographic comparisons gave way to more statistically based techniques at this time, and rigorous statistical analyses of how space impacted on electoral behavior replaced the earlier map comparisons approach and these studies were particularly concerned with voting cartographies, electoral boundaries, and the analysis of geographical influences on voting, including work on the neighborhood and friends and neighbors effects. While other areas of political geography, such as geopolitics, were in decline during this period, electoral geography, with its large data bank of electoral statistics, proved to be particularly suited to the quantitative approach. However, cartographical techniques still remained as a particularly powerful tool; for instance, maps of different outputs produced by the ecological models were often produced to complement the statistical analyses and to illustrate the key findings. Since then new and highly sophisticated computer based applications have added further dimensions to electoral cartography and to its applicability within practical politics. The use of more sophisticated mapping programs has, however, further increased the effectiveness of partisan gerrymanders, as evidenced in the results of congressional elections in the USA in the 2000s, with such practices, in turn, contributing to more general trends of congressional incumbency protection and declining electoral competitiveness. The development of geographically weighted regression techniques, in turn, has added new dimensions to work within the area, as the mapping of localized statistical outputs from such research allows the reader to easily appreciate how relationships between various causal factors and electoral behavior may vary across space, thus offering a statistical means to illustrate how place or local context matters, in terms of influencing how different factors relate to electoral behavior.

The use of cartograms to display geographically based electoral figures has also added new dimensions to electoral cartography, with the use of such cartograms for the purposes of displaying US election results (including the work of Gastner and Newman) giving a more faithful presentation of the relative strengths of the Republican and Democrat parties in recent elections. Conventional maps suggest that US political space is being largely dominated by Republican support, mainly due to this party's strength in the large, but low density, Mid Western states. In using cartograms to display electoral support levels however, the sizes of US states (or counties) are rescaled according to population levels, thus placing greater emphasis on highly populated states where Democrat support is often higher.

New and interesting research continues to emerge, facilitated by the aforementioned developments in statistical techniques and geographical information systems, as well as the availability of better quality electoral and socioeconomic data. Such research has tended to be quite varied in its scope, with different papers focusing on specific elements of the main strands within electoral geography. Recent research has focused on the redrawing of constituency boundaries and the electoral abuses of gerrymandering and malapportionment, while other research in this field has highlighted the impact of first past the post electoral systems in producing disproportional and biased electoral results. The availability of more detailed electoral data at the aggregate level has also acted as the basis for interesting new research, as evidenced in recent marked register analyses based studies of voter turnout levels (as in the Republic of Ireland), focusing either on the national picture or on more localized contexts. The volume of research within this field has remained strong, although various commentators have viewed the bulk of this output as amounting to little more than piecemeal offerings to the general study area, involving isolated studies of specific elections within specific states and regions that fail to help create a coherent body of knowledge that would add to the overall study of political geography. The general consensus among such commentators is that electoral geography is overtly empirical in its approach and fails to underpin these researches with well developed theoretical frameworks, an atheoretical stance that has left it at odds with general trends within human geography and its cultural turn. Commentators, such as Colin Flint, have viewed electoral geography's conception of time and space as being particularly constrained, given that much research tends to focus on a single election or series of elections within a particular region or state. Another critique leveled at electoral geography relates to the manner in which geography is used in such research, with some commentators arguing that it is usually treated as an epiphenomenal or residual factor, a point which has been contested by electoral geographers, such as Johnston and Pattie.

Efforts to promote a stronger theoretical dimension to electoral geography have found strong resonance with realist philosophy and structuration theory, and argue for the need to better situate electoral research within frameworks that locate it within the larger perspective of global economic transition while also placing greater stress on the impact of the local context. This places greater emphasis on the role of geographical scale in electoral research, while highlighting the importance of place, or the local context, in terms of electoral decision making processes. Researchers, such as Taylor and Flint, have advocated a world systems approach to the study of electoral geography that would also address the discipline's failure to study elections in the developing world, while also arguing for the need to understand how electoral behavior is shaped by structural constraints and how voters in specific places are responding to global processes. Creating electoral cartographies within such frameworks allows for place specific electoral behavior to be studied with reference to the political and economic context provided by global economic trends and also the localized institutions and cultural practices that mediate electoral responses to these global processes. In light of this, electoral cartography needs to be based on an approach to electoral geography that marries together a series of different but interrelated geographical scales, as argued by John Agnew, thus refocusing the discipline on a study of how historical–geographical context impacts on political behavior. In stressing the importance of geographic scale and also the need to base researches on a place centered approach, work within the areas of electoral geography and electoral cartography, married to new and ongoing developments in geographical statistics and computer mapping, promises to offer fresh insight not just to the subdiscipline of political geography, but to the larger study of human geography and indeed also to political science and the other social sciences.

See also: Electoral Districts; Electoral Geography; Geographically Weighted Regression; Gerrymandering; Maps; Place; Political Geography; Political Representation.

Further Reading

Agnew, J. A. (1990). From political methodology to geographical social theory? A critical review of electoral geography, 1960 1987. In Johnston, R. J., Shelley, F. M. & Taylor, P. J. (eds.) *Developments in Electoral Geography*, pp 15 21. London: Croom Helm.

Agnew, J. A. (1996). Mapping politics: How context counts in electoral geography. *Political Geography* 15(2), 129 146.

Belanger, P., Carty, R. K. and Eagles, M. (2003). The geography of Canadian parties' electoral campaigns: Leaders' tours and constituency election results. *Political Geography* 22, 439 455.

Flint, C. (2001). A timespace for electoral geography: Economic restructuring, political agency and the rise of the Nazi party. *Political Geography* 20, 301 329.

Forgette, R. and Platt, G. (2005). Redistricting principles and incumbency protection in the US Congress. *Political Geography* 24, 934 951.

Johnston, R. J. (2000). Electoral geography. In Johnston, R. J., Gregory, D., Pratt, G. & Watts, M. (eds.) *The Dictionary of Human Geography* (4th edn.), pp 204 205. Oxford: Blackwell Publishing.

Johnston, R. J. (2002). Mapping manipulating maps and winning elections: Measuring the impact of malapportionment and gerrymandering. *Political Geography* 21, 1 31.

Johnston, R. J. and Pattie, C. (2003). Representative democracy and electoral geography. In Agnew, J., Mitchell, K. & Toal, G. (eds.) *A Companion to Political Geography*, pp 337 355. Oxford: Blackwell.

Johnston, R. J. and Pattie, C. (2004). Electoral geography in electoral studies: Putting voters in their place. In Barnett, C. & Low, M. (eds.) *Spaces of Democracy: Geographical Perspectives on Citizenship, Participation and Representation*, pp 45 66. London: Sage.

Johnston, R. J. and Pattie, C. (2006). *Putting Voters in their Place: Geography and Elections in Great Britain*. Oxford: Oxford University Press.

Johnston, R. J., Rossiter, D. and Pattie, C. (2005). Disproportionality and bias in US presidential elections: How geography helped Bush defeat Gore but couldn't help Kerry beat Bush. *Political Geography* 24(8), 952 968.

Kavanagh, A. (2004). Turnout or turned off? Electoral participation in Dublin in the early 21st century. *Journal of Irish Urban Studies* 3(2), 1 24.

Kavanagh, A., Mills, G. and Sinnott, R. (2004). The geography of Irish voter turnout: A case study of the 2002 General Election. *Irish Geography* 37(2), 177 186.

Siegfried, A. (1913). *Tableau Politique de la France de l'Ouest sous la Troisième République*. Paris: Armand Colin.

Taylor, P. J. and Flint, C. (1990). *Political Geography: World Economy, Nation State and Locality*. London: Longman.

Taylor, P. J. and Johnston, R. J. (1979). *Geography of Elections*. New York: Holmes & Meier Publishers.

Relevant Websites

http://www.uselectionatlas.org/
Dave Leip's Atlas of U.S. Presidential Elections.

http://www.electionsireland.org/
Elections Ireland.org.

Electoral Districts

J. I. Leib, Old Dominion University, Norfolk, VA, USA
G. R. Webster, University of Wyoming, Laramie, WY, USA

Glossary

Electoral District Discrete, geographically bounded territories drawn for the purpose of electing members to a legislative body.
Gerrymandering Manipulation of electoral district boundaries to disproportionately inflate the representation of one group and the under-representation of others.
Ideal District Population The population of a jurisdiction divided by the number of districts that jurisdiction is apportioned.
Majority–Minority District An electoral district in which 50% or more of eligible voters are members of minority groups.
Multimember District A geographic district drawn to delineate a constituency that elects two or more members to a legislative body.
(Re)apportionment The process by which seats in a legislative body are (re)allocated to political jurisdictions.
(Re)districting The process of (re)drawing electoral district boundaries.
Single-Member District A geographic district drawn to delineate a constituency that elects a single member to a legislative body.

Definition and Purpose of Electoral Districts

Electoral districts are discrete, geographically bounded territories drawn for the purpose of electing members to a legislative body. While interest within political geography concerning how districts are drawn can be traced back nearly a century, the process of district drawing drew much study starting in the 1960s and 1970s, coinciding with the rise of research in electoral geography at a time when spatial–analytical perspectives were dominant within political geography. Redistricting is still an area of study within the discipline today, but has branched out to include post structuralist concerns with issues of boundaries, place, power relations, difference, and the social construction of identity.

Different systems of electoral districts are used to elect representatives to national legislatures around the world. In some, each member is elected from a geographically discrete single member district. This districting structure attempts to create a direct linkage between the legislator and the geographic area they are elected to represent. In such a system of 'particular territorial representation', legislators are expected to represent the interests of the constituency that elected them. Among the countries that use single member district systems for elections and representation are Bangladesh, France, Ghana, Trinidad and Tobago, and the United Kingdom.

At the other end of the spectrum, some countries, including Israel and the Netherlands, use 'pure' proportional representation systems in which all members are elected from a single 'electoral district' conterminous with a country's borders. Such systems provide for 'virtual representation' in which members are expected to represent the interests of the country as a whole (though recent evidence suggests that in Israel and the Netherlands geographic representation occurs because the political parties competing select representatives to stand for election from those areas of the country where the parties expect to be most competitive).

A variety of electoral district systems are used with characteristics falling between the two extremes of single member districts and national constituencies. For example, some countries use multimember electoral districts from which two or more members are elected to the legislature. Included in this category are Egypt and Kuwait which elect two representatives from each of their electoral districts. Alternatively, Brazil and Austria are subdivided into a small number of districts with each electing a variable number of representatives (from 8 to 70 per multimember district in Brazil, and 6 to 102 per multimember district in Austria). Still other countries use mixed systems in which some members are elected from single member districts and others from multimember districts (e.g., Japan), or in systems which elect some members from single member districts and others through proportional representation from a national constituency (e.g., Italy, South Korea).

One of the most complicated electoral structures is that of the Palestinian Authority, which uses a mixed system to elect 132 members of its legislature. In the 2006 elections, half of all members were elected from electoral districts, while the other half were elected by proportional representation from its territory as a whole. For the 66 members elected from districts, the Palestinian Authority's territory was divided into 16 electoral districts including 11 in the West Bank and five in Gaza. Among these 16 were 13 multimember districts (with the number of seats per multimember district ranging from two to nine), while the other three were single member districts. Thus, all voters cast two sets of votes, one for representatives from their single member or multi member electoral district, and one for legislators to represent the territory as a whole. Further complicating matters, in four of the 13 multimember districts, one or two seats were reserved for the Christian candidate receiving the highest number of votes in those districts.

As well, legislative bodies are found at multiple scales in many countries. Not all use the same types of electoral system and electoral districts at each scale within the same country. In the United States, where elections to legislative bodies are held at the local, state, and national levels, it is common to find voters residing in different types of electoral districts depending on the scale of election being held. For example, at the local level, Woodbridge Township, New Jersey elects members to its township council through both at large and single member district elections, meaning that each resident resides in both a single member and a multimember district. In state elections, New Jersey elects members to its state legislature's upper chamber from single member districts, while representatives in its lower chamber are elected from two legislator multimember districts. Members of the US House of Representatives are elected from single member districts. Thus, a resident of Woodbridge, considering the elections at these three scales, resides in three single member districts, one two member district, and a township wide constituency electing candidates at large.

Variation in Population Size of Electoral Districts

The strength of the linkage between the residents of electoral districts and their representatives is based in part on the number of persons per district. As is evident in **Tables 1a** and **1b**, the number of persons per electoral district varies greatly around the world. In terms of single member districts, the number of persons per district ranges from under 50 000 in Botswana, Ja maica, and Swaziland, to approximately 100 000 in France, Ghana, the United Kingdom, and Uzbekistan, to much larger average ratios of people per single member district in Bangladesh, Pakistan, the United States, and India. In terms of countries using multimember districts, the size per constituency is a function of both the number of total seats and the number of seats per dis trict. While Kuwait is divided into 25 constituencies for its 2.6 million persons (an average of approximately 100 000 persons per each multimember district), Brazil has only 27 constituencies for its population of 183 million (an average of nearly 7 million people per multimember constituency).

The differences in average constituency populations are in part a result of the number of seats in each national legislature. Thus, the larger the number of seats a legis lature has, the smaller the average constituency size will be. As suggested in **Table 1** national legislatures can range in size from several dozen seats (such as 36 in Trinidad and Tobago) to many hundreds (659 in the United Kingdom's House of Commons).

Table 1a Selected countries using single-member districts

Country	*2006 Pop. (in millions)*	*No. of seats/districts (as of 2000)*	*Population per seat/district*
Bangladesh	147.3	300	491 000
Botswana	1.7	40	42 500
Canada	32.2	301	106 645
France	60.7	555[a]	109 369
Ghana	22.4	200	112 000
Haiti	8.4	83	101 205
India	1 123.9	543	2 069 797
Jamaica	2.7	60	45 000
Kyrgyzstan	5.2	70	74 286
Malawi	12.6	177	71 186
Morocco	31.5	325	96 923
Pakistan	166.9	207	806 280
Swaziland	1.2	55	21 818
Trinidad and Tobago	1.3	36	36 111
United Kingdom	59.9	659	90 895
United States	297.1	435	682 988
Uzbekistan	27.3	250	109 200

[a] Does not include constituencies outside of France proper.

Source: Compiled and calculated from data available from Rose, R. (ed.) (2000). *International encyclopedia of elections.* Washington, DC: Con gressional Quarterly Press, and Population Reference Bureau (2006). Population Reference Bureau DataFinder. http://www.prb.org/DataFinder.aspx (accessed in Aug. 2006).

Table 1b Selected countries using multi member districts/constituencies

Country	*2006 Pop. (in millions)*	*No. of seats (2000)*	*Pop. per seat*	*No. of dis./con.*	*Pop. per dis./con.*	*Mean no. seats per dis./con. (range)*
Argentina	38.7	257	150 584	24	1 612 500	10.7 (2 25)
Austria	8.1	183	44 262	9	900 000	20.3 (6 102)
Belgium	10.4	150	69 333	20	520 000	7.5 (2 33)
Brazil	183.8	513	358 285	27	6 807 401	19.0 (8 70)
Costa Rica	4.3	57	75 439	7	614 286	8.1 (4 21)
Dominican Republic	9.1	120	75 833	30	303 333	4.0 (2 31)
Egypt	76.4	444	172 072	222	344 144	2 in each
Ireland	4.2	166	9 638	41	102 439	4.0 (3 5)
Jordan	5.9	80	73 750	20	295 000	4.0 (2 9)
Kuwait	2.6	50	58 000	25	29 000	2 in each
Latvia	2.3	100	23 000	5	460 000	20 (14 27)
Turkey	73.3	550	133 272	79	927 848	7.0 (2 21)
Vietnam	83.5	450	185 555	158	528 481	2.8 (2 3)

Adjustments to Electoral Districts

The population sizes of districts are frequently highly dynamic, and many countries periodically adjust the numbers, population sizes, and the geographic configurations of their electoral districts to reflect growth or decline in population. As population increases, some countries increase the number of seats in their legislatures to maintain somewhat equal populations in each district over time. In other cases, the number of seats remains constant but districts are redrawn to reflect population shifts within a country, and to keep districts roughly equal in population size with each other.

These processes can be demonstrated by comparing the districting process for legislatures in three Western Hemisphere countries: the House of Commons in Canada, and the House of Representatives in both Jamaica and the United States. All three countries currently elect members to these national legislative bodies from single member districts. To reflect population growth, the number of seats, and therefore the number of districts, has increased over time in each country's legislatures.

Canada

The Canadian Parliament is composed of the House of Commons and the Senate. Of the two, the House of Commons is by far the most powerful, and its 308 members are elected by voters from single member districts. As the population of Canada has grown since its founding in 1867, the number of seats in the House of Commons has also increased from 181 in 1867 to 308 today. However, the increase in the number of seats has not kept pace with the increase in population. As a result, the average constituency size has increased fivefold from an average of approximately 20 000 persons per district in 1871 to nearly 100 000 today. Based on formulas crafted in the Representation Act of 1985, the decennial reapportionment following the 2001 Canadian Census saw nearly 90% of the seats in the House of Commons apportioned to the ten provinces based on their populations. The remainder were allocated (1) to the provinces to insure that no province had fewer seats in the House of Commons today than it did in 1976, or in the Canadian Senate (these provisions added seats to all provinces except for the fastest growing (Ontario, Alberta, and British Columbia) above what they would have received based on their 2001 population alone), (2) and to the three Canadian territories (Nunavut, Northwest Territories, and Yukon Territory), which are apportioned one seat each. As a result, average district sizes among the provinces range from approximately 34 000 for each of the four seats in Prince Edward Island to nearly 109 000 for each of the 36 seats in British Columbia. The number of seats per jurisdiction ranged from one for each of Canada's three territories to 106 for Canada's largest province in population, Ontario.

Once seats are reapportioned among the provinces, electoral district boundaries are readjusted through the process of redistricting (the process of drawing electoral district boundaries), as detailed in the Electoral Boundaries Readjustment Acts of 1964 and 1985. The districts are redrawn in each province by a three member non partisan provincial electoral boundary commission. Commissions are to create districts within a province that are 'reasonably' equal to each other in population, with population deviations generally acceptable to $\pm 25\%$ of the ideal district population (the total population of a jurisdiction divided by the number of districts that jurisdiction is apportioned). Additionally, criteria guiding redistricting include taking into account such factors as communities of interest and identity, and the locations of previous district boundaries. The commissions are also charged to maintain districts of a 'manageable geographic size', especially in the thinly populated northern regions of many of the provinces (in such cases, commissioners may create districts that deviate in population by more than 25% from the ideal district population).

Jamaica

The Jamaican Parliament is composed of the House of Representatives and the Senate. Members of the Senate are appointed, while members of the House of Representatives are elected from geographically defined constituencies. Article 66 of the 1962 Independence Constitution provided the island be divided into no less than 45 and no more than 60 constituencies, with each electing one member of the House. Prior to the parliamentary elections in 1967, the first nationwide vote after independence, the number of constituencies was increased to 53. In 1976 the number of constituencies was further increased to a total of 60.

The Electoral Advisory Committee (EAC) is charged with making recommendations to the House of Representatives on appropriate changes to the boundaries and numbers of constituencies. Changes may be recommended to Parliament for their approval every 4–6 years. Subsequent to approval, changes are effected by an order of the governor general. Jamaica is subdivided into 14 first order civil divisions called 'parishes'. The Jamaican Constitution mandates that all parishes have a minimum of two constituencies, and that the size of the electorate in all constituencies be as "nearly equal to the electorate quota as is reasonably practicable." The 'electorate quota' is calculated by dividing the total electorate by the total number of districts or constituencies. Significant differences from the electorate quota may be justified on the basis of contrasting physical features, existing transportation facilities, and contrasts between urban and rural areas as reflected in different population densities. However, the Constitution limits size differences to no smaller than 66.66% and no greater than 150% of the electorate quota. The EAC has also adopted guidelines for its recommendations such as following natural boundary lines and avoiding the division of communities.

There are two primary parties in Jamaica: the Jamaican Labour Party (JLP) and the People's National Party (PNP). Though the PNP has been the dominant party for nearly two decades, the 2002 elections indicated much greater parity in support for the two parties. Also in 2002, the parliamentary elections in Trinidad and Tobago resulted in a virtual tie causing substantial controversy. These two events led the Jamaican EAC to begin developing recommendations to increase the number of constituencies from the even number of 60 to 65 to prevent the potential for a tie in the number of seats won by the two main parties (no third party in Jamaica has ever won a parliamentary seat). In 2005 the EAC hosted an international boundary conference to seek input from other countries on the redrawing of Jamaica's constituency boundaries. The EAC subsequently recommended that changes in the number of constituencies be effected prior to the parliamentary elections of 2007. A bill constitutionally allowing for an immediate increase to 63 seats and a subsequent rise to 65 seats was debated in the Jamaican House of Representatives in 2006, but tabled due to concerns over gerrymandering by some of the chamber's members. The EAC is also having increasing difficulty making boundary changes to districts in some parishes to remain in compliance with the constitution's prohibition of constituencies being no larger than 150% of the electorate quota. Jamaican election officials reported in 2006 that three constituencies exceeded the constitutional limits for the total number of electors, with two more soon expected to be in noncompliance.

United States

The United States Congress is composed of two chambers: the Senate and the House of Representatives. The House of Representatives apportions seats based upon population by state, with all members elected from single member districts. Membership in the House of Representatives was initially set in the 1780s at 65 seats, subsequently rising to its present level of 435 members. While initially the size of House constituencies was set at one representative per 30 000 persons, in the reapportionment subsequent to the 2000 Census of Population each constituency averaged approximately 647 000 people, an over 21 fold increase in constituency size. Between 1790 and 1910 the number of seats in the House was commonly increased after each census to lessen the effect upon constituency size due to the country's rapid population growth. The size of the US House was last increased following the 1910 Census to a total of 435 seats. In that census the US population was enumerated at 92 million, leading to an average district constituency size of 211 000 people. While the Reapportionment Act of 1929 froze the size of the House at 435, the country's population has tripled since 1920, leading to the comparatively large sizes of House constituencies today.

Each of the 50 states is allocated a minimum of one seat in the House of Representatives, with the remaining 385 seats apportioned to the states based upon their populations. In the apportionment following the 2000 Census, the number of House seats per state ranged from one seat in seven states to California's 53 seats. The range in the constituency sizes was greatest between those states having the constitutionally defined minimum of one seat. Thus, Wyoming's one member represented

494 000 people while Montana's single member repre sented 902 000 people.

The districts electing members of the House of Representatives are generally redrawn after each census, as are most single member districts at all lower geo graphic scales (e.g., county commissions, city councils, state legislatures). In nearly all cases the process of redistricting must be completed prior to the first general election after the release of the census data. Legislative bodies are most often responsible for completing the process. Thus, in most states the state legislature redraws their own districts, as well as those for the US House of Representatives. Due to the almost certain partisanship in the process, some states have vested responsibility in nonpartisan agencies (e.g., Iowa) or bipartisan commis sions (e.g., New Jersey). The purpose of such alternatives is to reduce the likelihood of partisan gerrymandering (the manipulation of election district boundaries to dis proportionately inflate the representation for one group and the resulting under representation of others).

Since the early 1960s the redistricting process in the United States has increasingly come under the scrutiny of the federal judiciary. For example, the federal courts have mandated evermore perfectly equal populations between US House districts in the same state. In some states the number of people per congressional constitu ency may vary by less than a handful of individuals. Although state legislative and local government districts generally have greater flexibility, it is exceedingly rare for districts at these scales to vary by more than 5% from the ideal district population within those jurisdictions.

In the early 1990s those having responsibility for the redistricting process substantially increased the number of districts having racial and ethnic majorities. This trend emanated from the US Department of Justice's legal in terpretation that amendments to the 1965 Voting Rights Act required that majority–minority districts be de lineated whenever possible. Such affirmative redistricting resulted in several majority–minority districts that were geographically extensive and of comparatively lesser geographic compactness. Several court suits were filed against these districts during the 1990s that called into question the constitutionality of irregularly shaped dis tricts drawn to promote minority representation. Gen erally race and ethnicity can be considered in the redistricting process at present, but such demographic considerations must be used in conjunction with other districting criteria such as contiguity, compactness, and maintaining the integrity of local government boundaries.

In the past decade partisan gerrymandering has be come an increasingly important legal issue at the national level due to the similar numbers of Democrats and Re publicans in the US House of Representatives. Since state legislatures generally have the responsibility to delineate congressional districts, the majority party in each state legislature can have an enormous effect upon the party makeup of the US House. Due to redistricting litigation in Texas after the release of the 2000 Census, the state's districts were redrawn by a panel of federal judges in advance of the 2002 congressional elections. In the 2002 election the Texas State Legislature shifted to the control of the Republican Party, and the body again undertook an effort to redraw the state's congressional districts prior to the 2004 congressional elections. The resulting plan sig nificantly increased the number of Republicans elected from Texas, aiding that party's ability to continue as the majority party in Congress. While it is customary that, barring a court order, redistricting is carried out only once a decade in the United States, the US Supreme Court did not find a constitutional violation in the Texas legislature's mid decade redistricting. This leads to the possibility that states might revisit the redistricting process for partisan purposes multiple times during a single decade.

Issues in Electoral Districting

There are a number of issues surrounding the delineation of electoral districts including those pertaining to the data, technology, and criteria used to draw districts, and the fairness of the resulting districts to political and demographic groups. For example, the data employed in the redistricting process is generated through enumer ations of the population in each country. As a result, errors in the population enumeration of a country can result in the under or over representation of geographic areas and/or demographic groups. Generally ethnic mi norities, the poor, and those residing in rural settings tend to be undercounted. For the purposes of districting, population data are broken down into small geographic units that can be aggregated to form districts of nearly equal population. In the United States, the Census Bur eau provided demographic data at the 'block' level, the lowest level in the census' geographic unit hierarchy, for all parts of the country for the first time after the 1990 Census. Such detailed data allows those delineating boundaries to create districts with near perfectly equal populations to insure equitable representation. But they also provide an increased ability for partisan manipula tion of district populations.

The advent and rapid development of sophisticated Geographic Information Systems (GISs) has greatly re duced the historically cumbersome character of the redistricting process. A GIS can allow the consideration of massive quantities of demographic, economic, social, and political data in the districting process. Such systems have arguably democratized the system, as any citizen with access to GIS can create districting plans to submit to districting authorities. At the same time, however, GIS has also created a greater potential for gerrymandering.

The criteria used in the districting process differ depending upon country. Among the most common criterion is equal population. Two additional criteria that are commonplace are territorial contiguity and geographic compactness. Territorial contiguity requires that any portion of a district must be accessible from all other parts without entering a second district. Geographic compactness has long been used as a criterion for multiple purposes. Historically it is has been assumed that citizens residing near one another will have more in common than those living at greater distances. Geographic compactness has also been viewed as a means of limiting the manipulation of districts for political or demographic advantage. However, such an assumption is also increasingly suspect, as more detailed data and GIS allow comparatively compact districts to be formed that facilitate the over representation of some groups by dividing the geographic concentrations of other groups.

It is also common in many countries to consider the representation of religious, ethnic, or racial minorities as a criterion in the redistricting process. Thus, in Lebanon's Parliament there is an established number of seats for each confessional group including among others, Maronites, Sunnis, Shiites, Druze, Protestants, and Armenian Catholics. The equitable representation of ethnic and racial minorities in the United States has been actively litigated for over a half century. Due to the Voting Rights Act, extended in 2006 by the US Congress for another 25 years, racial, ethnic, and linguistic minorities must be considered in the districting process.

Most observers would agree that the electoral districting system should be as fair as possible in providing an equal chance of representation to all segments of the population. However, achieving fair representation can be difficult because of differing definitions of what is 'fair'. In spite of this difficulty, the best electoral districting plan is one that provides the greatest access to representation to the greatest number of demographic or political groups. The districting process is about allocating power to population groups, and a group without access clearly has a limited political voice in the legislative body.

See also: Electoral Cartography; Electoral Geography; Gerrymandering; Political Geography; Political Representation.

Further Reading

Arrington, T. and Ingalls, G. (1998). The limited vote alternative to affirmative districting. *Political Geography* 17, 701 728.

Elections Canada (2002). *Representation in the House of Commons of Canada*. Ottawa: Elections Canada.

Forest, B. (2001). Mapping democracy: Racial identity and the quandary of political representation. *Annals of the Association of American Geographers* 91, 143 166.

Grofman, B. (ed.) (1998). *Race and Redistricting in the 1990s*. New York: Agathon.

Johnston, R. J., Pattie, C. J., Dorling, D. F. L. and Rossiter, D. J. (2001). *From Votes to Seats: The Operation of the UK Electoral System since 1945*. Manchester: Manchester University Press.

Johnston, R. J., Rossiter, D. J. and Pattie, C. J. (1999). *The Boundary Commissions: Redrawing the UK's Map of Parliamentary Constituencies*. Manchester: Manchester University Press.

Latner, M. and McGann, A. (2005). Geographic representation under proportional representation: The cases of Israel and the Netherlands. *Electoral Studies* 24, 709 734.

Leib, J. (1998). Communities of interest and minority districting after *Miller v. Johnson. Political Geography* 17, 683 699.

Leib, J. and Webster, G. (1998). On enlarging the U.S. house of representatives. *Political Geography* 17, 319 329.

Morrill, R. L. (1981). *Political Redistricting and Geographic Theory*. Washington, DC: Association of American Geographers.

Morrill, R. L. (1999). Electoral geography and gerrymandering: Space and politics In Demko, G. J. & Wood, W. B. (eds.) *Reordering the World: Geopolitical Perspectives on the 21st Century* (2nd edn.). pp 117 138. Boulder, CO: Westview.

Population Reference Bureau (2006). Population Reference Bureau DataFinder. http://www.prb.org/DataFinder.aspx (accessed in Aug. 2006).

Rose, R. (ed.) (2000). *International Encyclopedia of Elections*. Washington, DC: Congressional Quarterly Press.

Taylor, P. J. and Johnston, R. J. (1979). *Geography of Elections*. London: Croom Helm.

Webster, G. R. (2000). Playing a game with changing rules: Geography, politics and redistricting in the 1990s. *Political Geography* 19, 141 161.

Webster, G. R. (2004). Evaluating the geographic compactness of representational districts. In Janelle, D. G., Warf, B. & Hansen, S. (eds.) *WorldMinds: Geographical Perspectives on 100 Problems*, pp 43 47. Dordrecht: Kluwer.

Relevant Websites

http://www.ace.at.org
ACE Project, The Electoral Knowledge Network.

http://www.commoncause.org
COMMON CAUSE, Holding Power Accountable.

http://www.fairvote.org
Fair Vote, The Center for Voting and Democracy.

http://www.lib.purdue.edu
Purdue University. Libraries, Access Knowledge Success.

http://www.parliament.uk
United Kingdom Parliament, Elections, Parliamentary Constituencies.

http://www.census.gov
US Census Bureau, Census Redistricting Data Office.

Electoral Geography

C. Pattie, University of Sheffield, Sheffield, UK
R. Johnston, University of Bristol, Bristol, UK

Glossary

Egocentric voting Voting based on evaluations of what is best for oneself, irrespective of the impact on others (often discussed in contrast with sociotropic voting).

Electoral de-alignment A process in which long-standing divisions within the electorate, influencing which parties members of different groups are likely to support, weaken. De-alignment hs typified politics in most western nations since the 1970s.

Logistic regression An extension of the regression technique to take into account situations where the dependent variable is measured using categorical data. The technique is particularly useful for the analysis of survey data.

Modifiable Areal Unit Problem (MAUP) A problem which arises when small areal units can be combined into several different sets of larger units. Combing the smaller units in different ways can substantially alter the apparent geography of an underlying attribute of interest at the level of the larger units.

Multi-level modelling A family of statistical methods designed to take into account situations where data exhibit a multi-level structure (for instance, school pupils' performance may be influenced by factors operating at a variety of nested scales: individual ability; family background; school effects; and so on).

Plurality electoral system In plurality elections, the candidate with the most votes is deemed to win, whether or not s/he has achieved an outright majority (50%+) of all votes cast. Most plurality election systems employ single-member constituencies. Plurality elections are most common in the UK and in former British colonies.

Psephology The academic study of elections and voting.

Regression analysis A statistical technique involving fitting a linear trend to data to uncover how one or more explanatory (independent or *x*-) variables affect the variable the analyst wishes to explain (often termed the *y*- or dependent variable).

Sociotropic voting Voting based on evaluations of what is best for others, irrespective of the impact on oneself (often discussed in contrast with egocentric voting).

Spatial regression An extension of regression techniques to take into account spatial autocorrelation in geographical data.

Valence politics Valence issues are those on which all individuals commonly agree (as, for instance, in the general desire for prosperity or for good health). In valence politics, therefore, public support for political parties is based not on the ideological claims of the parties, but on evaluations of which party is most likely to deliver generally-desired outcomes successfully.

Introduction

Electoral geography, the analysis of the interaction of space, place, and electoral processes, shares a considerable hinterland with the related political science subfield of psephology. Pronounced geographies of party support are to be found in most countries, even in those (such as the Netherlands and Israel) which employ highly proportional electoral rules in a single national constituency. The field's primary research question is therefore 'why do particular parties draw more electoral support from some places than from others?'

Explaining how geographies of party support are created involves the analysis of a variety of factors, which interact in potentially complex ways: the underlying social geography of party support; interactions between citizens living in the same areas; the impact of economic geography on government popularity; geographically targeted party campaigning; and the geographical outcomes of the operation of the electoral system.

Underpinning all of these is a debate, common throughout human geography, about the relative importance of compositional and contextual effects. Do people behave and think as they do because of who they are (compositional effects, based on the social groups they belong to, the upbringings they have had, and so on), or because of where they are (contextual effects, drawing on the wider environments within which they are embedded, the things they see and experience around them, and the – spatially structured – encounters they have throughout their lives)? The geography of any actual election is a combination of both contextual and compositional effects: unraveling how these operate is the central task of electoral geography.

Origins

Electoral geography can be traced back to the early twentieth century. Many of the early contributions were by political scientists. The approaches they adopted, and the insights they developed, have had a lasting influence on the field.

The French political scientist André Siegfried is often seen as the founder of academic electoral geography. His path breaking analyses of voting in France linked social processes to geographies of party support. Famously, he compared party support with underlying geology: in northern France, he noted, support for the right was highest in areas where the local rock was granite, while parties of the left did better in chalk areas. However, this was not environmental determinism. Geology influenced vote via its impact on factors such as local land holding patterns (e.g., whether farms were large or small, which affected whether agricultural workers were employees or self employed peasants, and hence local patterns of social stratification) and industrial location (which influenced local class structures). While the maps included local geologies, the electorally important processes reflected social relations in particular places. This remains a focus of attention for modern electoral geography.

Another pioneer, the Swedish political scientist, Herbert Tingsten, employed statistical data to analyze electoral geographies. By comparing electoral and social data for 55 Stockholm electoral districts at the 1932 second chamber election, he demonstrated a strong positive ecological correlation between the proportion of working class voters in an area and the level of support for parties of the left there: the more working class the neighborhood, the more votes the Socialists won. Following Tingsten's lead, the statistical analysis of voting and social data aggregated into electoral districts has become part of the electoral geography mainstream.

The American political scientist V. O. Key provided another influence. His 1949 study, *Southern Politics*, explained the electoral distinctiveness of the southern states of the USA in terms of their particular political histories. Support for the Republicans, the party of the Union, was weak in the south for almost a century after the Civil War, while southern Democrats were more conservative and appealed more to a white supremacist vote than did their more liberal northern counterparts. In addition, racism was also linked to white voters' participation in elections: the larger the local Afro American community in a district, the greater the turnout among white voters. However, the geography of party support in the south also resulted from what Key termed 'friends and neighbors' voting: voters were more likely to vote for local candidates than for candidates from other parts of the state.

Key's discussion of normal and critical elections was also influential. Critical elections break the mould. Existing patterns of support for parties break down, and new patterns are established. After a critical election, these new patterns of support become entrenched, producing a stable period of normal elections, which persists until the next critical election. Periods typified by normal elections tend to have relatively stable electoral geographies, in which some regions consistently give greater support to particular parties than do others. In a critical election, however, these stable geographies can break down suddenly and, on occasion, irrevocably. One such realignment occurred in the US during the 1970s, when the formerly Democratic south began to swing behind the Republicans.

Key's work influenced geographers at the University of Iowa in the late 1960s, exemplified by J. Clark Archer and Peter Taylor's analysis of the regional geography of US Presidential elections from 1828 to 1980. They argued that socioeconomic interests, spatially expressed in different regions of the US, produce sectionalism, which forms the basis of patterns of party support. They identified different voting 'eras', each covering several presidential elections, each with its own distinctive geography, and each breaking down in time. Others have continued this line of work, identifying areas with consistent voting patterns over time at a variety of scales.

A rather different take, both methodologically and substantively, was offered by sociologists at Columbia University, led by Paul Lazarsfeld and Bernard Berelson. Whereas most previous studies – including Siegfried, Tingsten, and Key – had employed aggregate data to study electoral behavior, the Columbia group utilized detailed social surveys in particular communities, beginning with their study of the 1948 presidential election in Elmira, New York. For them, individual voters had to be seen as embedded within social groups and local communities. Their evidence suggested that political influence spread through contact with friends and acquaintances: voters were influenced by (and also influenced) the views and opinions of those they talked to and interacted with.

The longer term effects of the Columbia studies on electoral studies generally can be seen in the widespread adoption of the sample survey as the dominant source of data for the analysis of electoral behavior. Most subsequent studies have followed the example of the slightly later project launched by political scientists at the University of Michigan in the 1950s (and continue to the present as the American National Election Studies), and have employed national random samples of the electorate. In the process, they have gained the ability to generalize across the entire electorate, but have lost the ability to focus in depth on politics within particular communities. Only a handful of subsequent studies, most notably Huckfeldt and Sprague's analysis of the 1984 presidential campaign in South Bend, Indiana, have repeated the Columbia school's focus on a particular community. For electoral geographers, meanwhile, the Columbia school's enduring legacy has been of interest in the neighborhood effect – the ten dency for individuals living in the same place to vote in the same way.

Despite these precursors, however, initially few geographers took up the study of elections. It was not till the late 1950s that Prescott provided one of the first formal discussions of electoral geography as a distinctive subdiscipline. Prescott's electoral geography was firmly within the regional geography paradigm. It could be used to illustrate Hartshorne's concept of the *raison d'etre* of the state – as, for instance, in Northern Ireland, where the geography of the Unionist vote picked out the zones of greatest support for the Province's *raison d'etre*, continued links with the UK. This vision of electoral geography was strikingly apolitical. Tellingly, Prescott noted that "(p)robably the greatest objection which geographers have to using election returns is the need to acquire background political knowledge." Elections where the primary issues at stake are those of public policy, gov ernment performance, and so on, Prescott deemed of little interest for electoral geography, since they reveal little about the state's *raison d'etre*. Later electoral geog raphers, in contrast, have embraced such 'background political knowledge' as an essential part of adequate explanation.

A more seminal spatial science perspective was sup plied in the 1960s by Kevin Cox, who saw voters as de cision makers influenced by their local contexts. Cox conceptualized voters as nodes in a network, receiving, processing, and transmitting political information. The extent to which individuals might be influenced by the views of those they lived among, he argued, would be influenced by a variety of biases:

- Geographical distance bias: people would be more influenced by the views of their neighbors than by individuals who lived further away.
- Acquaintance circle bias: individuals are more likely to be influenced by the views of others in their ac quaintance circle than by the opinions of relative strangers (though later work by Granovetter showed the importance of casual acquaintances as a means of encountering heterodox views).
- Force field bias: especially before the introduction of the secret ballot, voters might come under pressure from politically more powerful groups and individuals.
- Reciprocity bias: individuals are more likely to be influenced by those they are structurally related to (spouses, employees and employers, children and parents) than by people with whom they have no structural relations.
- Ideological bias: voters with strong political convic tions would be less open to influence than those with weaker opinions.

Albeit shorn of some of its spatial science underpinnings, Cox's work continues to inform electoral geography to the present.

Methodological Issues

From these origins, electoral geography has become in creasingly sophisticated, both theoretically and meth odologically. In this section, we consider the latter.

Elections are mass events, depending on the simul taneous participation of several political parties and many voters. Furthermore, election results are routinely published for subnational areas, enabling geographical analysis of the outcome. At the 2005 UK general election, for instance, almost 44.3 million adults were eligible to vote, and 27.1 million did so: results were announced for 646 constituencies. Understanding the entirety of an election requires methods which can generalize over large populations, therefore. As a consequence, quanti tative methods dominate analysis in electoral geography.

The methods employed range from simple carto graphic representations of election results, to more ad vanced methodologies. Regression models are widely used: a party's constituency vote share in each seat can be regressed against a range of social, political, and economic indicators. Even so, communities of electoral geographers differ in the precise methodologies they apply. Spatial regression techniques, which allow the identification of 'hot spots' of support for particular parties, are most widely applied by North American analysts. American electoral geography is dominated by the ecological an alysis of election results in small districts. Despite the heavy use of sample surveys by American political sci entists, few US electoral geographers made use of them. In the UK, by contrast, spatial regression has rarely been used to study constituency voting patterns: conventional ordinary least squares (OLS) models have been more commonly applied. However, British electoral geog raphers have made heavy use of survey data in recent years, applying not only logistic regression models but also multilevel modeling techniques.

There are pros and cons to the analysis of both eco logical data from electoral districts and survey data. Ecological data allow the analysis of election results in all parts of a country simultaneously, since there is complete coverage for all areas. However, analysts are limited by data availability and the ecological fallacy. Survey data allow analysis of individual voters' decisions. However, since most national surveys interview in only a few constituencies, and contact only a handful of voters in each seat, they are not well suited to uncovering the detailed geography of constituency voting.

Analysts have therefore combined ecological and survey data. Most obviously, data on individual voters can be combined directly with data for their constituencies. Alternatively, constituency level estimates can be made of data which are not available directly at that scale. For instance, there are grounds for suspecting that the strength of class voting (typically, working class voters are more likely to vote for the left, and middle class voters for the right) varies from place to place. However, national surveys do not allow a detailed examination of the geography. And while census data on class and election results are commonly available at the constitu ency level, they are almost never cross classified. Some estimation method, such as entropy maximising, is therefore required to get at the constituency geography.

Electoral geographers are not naïve empiricists, however: they draw on a range of theoretical perspectives. Some theories come from political science. Partisan identification (the Michigan model) describes voters' long term loyalties to particular parties. Sociological accounts emphasized the electoral influence of factors such as class, religion, and core–periphery relations. Rational choice theories, meanwhile, emphasize the cost benefit calculations taken by voters and parties, with implications for the analysis of electoral participation, strategic (or tactical) voting, and electoral campaigning. They have also had an influence on the emergence of the valence politics perspective, which argues that voters decide by evaluating the relative performance of parties when in office.

The traffic of ideas has not been one way, however. Electoral geographers have emphasized the contextual nature of voting decisions. Place based explanations drawing on structuration theory have been applied to the development and long term persistence of distinctive local political cultures, in which individuals learn about politics in distinct environments. Two individuals in otherwise identical social positions might have very dif ferent political beliefs if one grew up in an affluent commuter suburb while the other grew up in a declining industrial city, for instance. Much empirical work in electoral geography is at root about tracing the geo graphical ramifications of these theories.

Social Bases of the Vote

Many political parties develop to represent the interests of particular sections of society, most commonly along class lines: Social Democrat or Labour parties try to mobilize working class votes, while Conservative or Liberal parties appeal for middle class support. The class cleavage, while never absolute, dominated West Euro pean electoral politics for much of the twentieth century. British politics provides a good example: at the 1964 general election, roughly two thirds of middle class voters supported the Conservatives, and around two thirds of working class voters supported Labour (**Figure 1**).

A class cleavage among individual voters has impli cations for the geography of the vote, as housing and labor markets produce social differentiation between areas. Some communities are predominantly working class, others mainly middle class. If the individual class cleavage holds everywhere, therefore, we would expect support for the main party of the right to be highest in middle class areas, and for the main left of center party to be highest in working class areas. Imagine two hypo thetical constituencies, Richtown and Steeltown. Both have electorates of 60 000, and both are contested by the Conservatives on the right, by Labour on the left, and by some minor party candidates. Richtown is middle class: 80% of its voters are in nonmanual occupations. Steel town, meanwhile is working class: 70% of its electorate is in manual jobs. If there is a strong national class cleavage in voting, invariant throughout the country, the Con servatives will win a clear overall majority of the vote in Richtown, as will Labour in Steeltown (**Table 1**). Note that the underlying class cleavage is identical in both constituencies: the different election outcomes are purely a result of the differences in class composition between the two seats.

Real world evidence reveals similar patterns. For in stance, the German Social Democrats are traditionally strongest in industrial areas such as the Ruhr, while the Christian Democrats and their partners have drawn more support from affluent and agricultural areas such as Bavaria. The link between geographies of class and party support is also well illustrated by constituency results at the 1970 British general election (**Figure 2**).

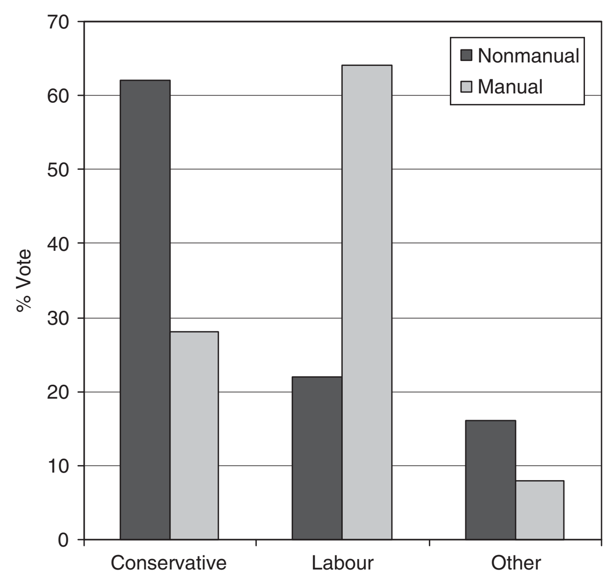

Figure 1 The class cleavage at the 1964 General Election in Britain.

Table 1 The effect of a national class cleavage on voting in two hypothetical constituencies

National class cleavage	*Nonmanual (%)*	*Manual (%)*				
Conservative	62	28				
Labour	22	64				
Other	16	8				
Total	100	100				
Richtown	*Nonmanual*		*Manual*		*Total*	
	(N)	*(%)*	*(N)*	*(%)*	*(N)*	*(%)*
Conservative	29 760	62	3 360	28	33 120	55.2
Labour	10 560	22	7680	64	18 240	30.4
Other	7680	16	960	8	8640	14.4
Total	48 000	100	12 000	100	60 000	100
Steeltown	*Nonmanual*		*Manual*		*Total*	
	(N)	*(%)*	*(N)*	*(%)*	*(N)*	*(%)*
Conservative	11 160	62	11 760	28	22 920	38.2
Labour	3960	22	26 880	64	30 840	51.4
Other	2880	16	3360	8	6240	10.4
Total	18 000	100	42 000	100	60 000	100

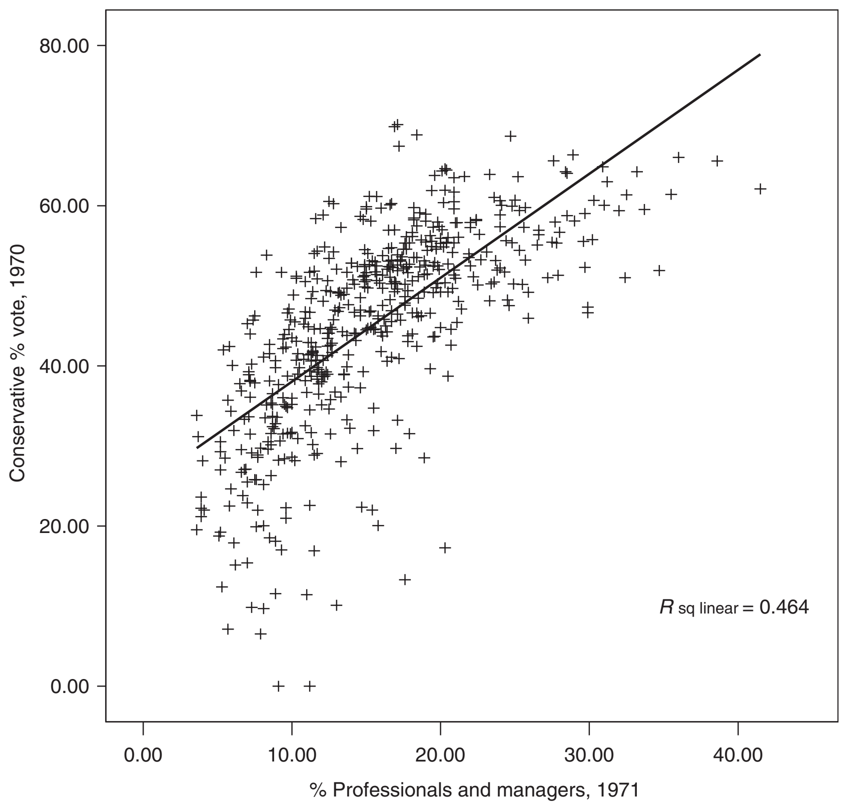

Figure 2 Constituency class composition and the Conservative vote in Britain, 1970.

The geography of the vote can result simply from the interaction of a national electoral cleavage and the geography of different social groups, therefore. Just because party support varies from place to place does not mean that geography has an independent effect on electoral behavior. Indeed, critics have argued that, once the appropriate attributes of individual voters are taken into account, apparent geographic effects will disappear: they argue places vote differently because of the sort of people who live there (compositional effects), not because of the sorts of places they are (contextual effects).

A closer examination suggests contextual effects do operate, however. For instance, Butler and Stokes showed the class cleavage was not uniform throughout

Table 2 Geographical variation in the class cleavage at the 1970 British general election: class self-image and partisan self-image

	Most manual 10% of seats		*Most nonmanual 10% of seats*	
	Middle class (%)	*Working class (%)*	*Middle class (%)*	*Working class (%)*
Conservative	46	21	78	50
Labour	54	79	22	50

Britain but varied systematically from place to place. They contrasted voters living in the 10% of seats with the highest proportions of manual workers in 1970 with those living in the 10% of seats with most nonmanual workers (**Table 2**). In both areas, a higher proportion of middle than working class people voted Conservative than voted Labour, and a higher proportion of working than middle class people voted Labour, as implied by the class cleavage. However, working class voters in predominantly middle class communities were more likely to support the Conservatives than were working class voters in mainly working class areas. And similarly, middle class voters in working class areas were more pro Labour in outlook than other middle class voters in middle class areas. Miller (1977) argued this was the result of 'conversion by conversation': individuals in areas dominated by a particular class would be more likely than average to encounter the views and opinions of that class, and hence would be more open to its in fluence than voters elsewhere. Who people are is not the only influence on how they vote: where they live seems to have an effect too.

Recent cross national trends in electoral cleavages provide even more striking evidence for contextual ef fects. Since the early 1970s, most Western democracies have experienced de alignment. Voters are no longer so strongly influenced by social class, or by long term loy alties to particular parties. We can illustrate this by comparing class voting at the 1964 and 2005 British general elections: far fewer voters voted for their 'natural' class party in 2005 than in 1964, and many more voted either for a third or for the 'wrong' class party (**Figure 3**). Voters are more volatile, and more likely to judge parties by results than in the past.

However, although individual voters are less class aligned than in the past, the same is not true of the geography of party support. Even among de aligned electorates, right of center parties gain most support in middle class neighborhoods and left of center parties in working class areas. For instance, at the 1992 British General Election class had relatively little bearing on the decisions of individual voters. However, at the constitu ency level, little had changed since 1970 (cf. **Figures 4** and **2**). Context matters.

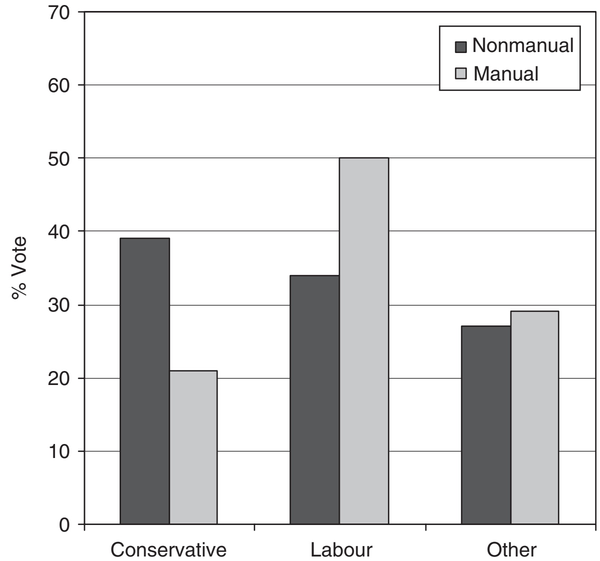

Figure 3 The class cleavage at the 2005 General Election in Britain.

How does context influence the vote? In the following three sections, we look at: the influence of other voters; the impact of local economic conditions; and the effects of party campaigns.

The Neighborhood Effect

The neighborhood effect literature draws on a simple observation: people live, work, and play with other peo ple. Through our daily interactions (some deep, some fleeting) with others we find out what those around us think. Some are likely to be influenced by the views they encounter in these interactions, either by sheer weight of numbers ('if that's what most people think, there must be something in it') or because the opinion is promulgated by individuals who are seen as particularly authoritative ('she obviously knows what she's talking about, so I had better take her view seriously').

This process is contextual: individuals can be influ enced by, and in their turn can influence, the views of those they encounter. The neighborhood effect traces its origins back to (among others) the Columbia school and Kevin Cox. Its core claim is interactions between people living in the same communities are consequential (e.g., as a result of Cox's geographical distance bias), such that a dominant community view will come to be shared by most people living there.

A substantial cross national body of research – much of it inspired by the innovative analyses of Robert Huckfeldt and John Sprague – demonstrates that voters' decisions are indeed influenced by the views of those they talk to. For instance, respondents to the 1992 British Election Study

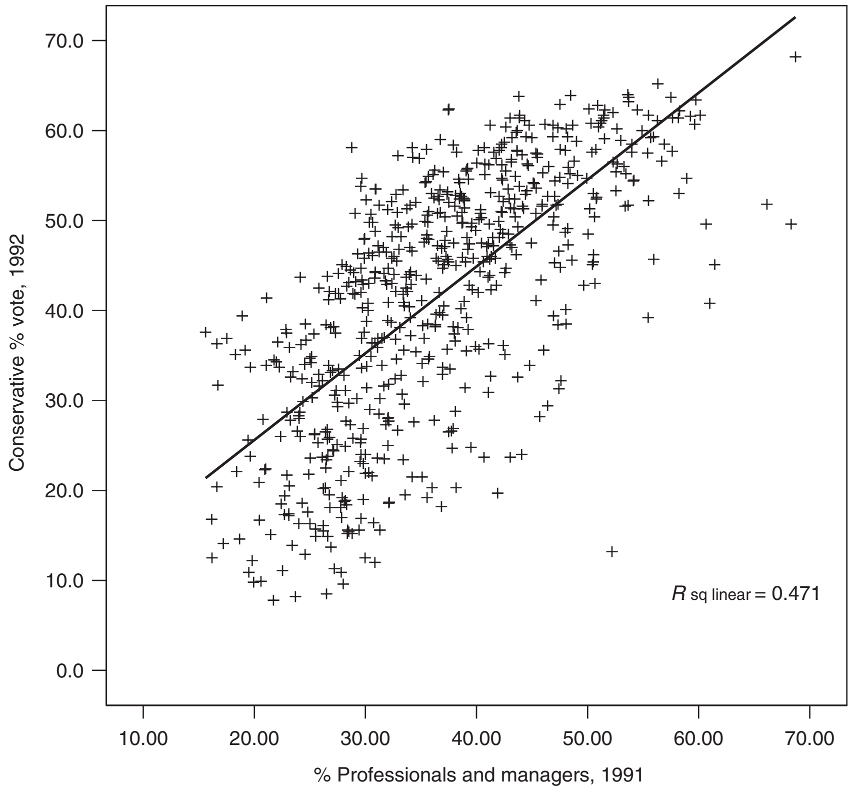

Figure 4 Constituency class composition and the Conservative vote in Britain, 1992.

Table 3 Voter defection and intensity (in terms of % voting in 1992) of partisan reinforcement in conversation networks

Party supported by discussants	*Number of discussants supporting party*			
	0	*1*	*2*	*3*
a) *Respondent did not vote Conservative in 1987*				
Conservative	8.5	26.0	45.5	51.7
Labour	19.4	12.0	5.8	2.1
Liberal Democrat	15.1	15.7	8.9	22.2
b) *Respondent did not vote Labour in 1987*				
Conservative	16.8	8.1	2.7	2.1
Labour	6.9	15.3	37.3	66.7
Liberal Democrat	12.0	6.5	6.2	0.0
c) *Respondent did not vote Liberal Democrat in 1987*				
Conservative	10.5	9.9	8.6	2.1
Labour	9.0	13.6	8.3	6.4
Liberal Democrat	6.5	28.8	40.0	57.1

From: 1992 BES.

were asked to name the three individuals they were most likely to talk to regarding important matters, and were also asked which parties (if any) they thought these individuals supported. Among respondents who had not voted for a party at the preceding (1987) election, the more individuals they reported having talked to who did support that party, the more likely they were to switch their vote to it in 1992 (**Table 3**). Take, for example, those individuals who did not vote Conservative in 1987: while just under 9% of those who did not name any Conservative supporting conver sation partners went on to vote for the party in 1992, 26% of those who had talked to one Conservative did so, as did 46% of those who had talked to two, and 52% of those who talked to three. And the corollary holds too: the more these individuals talked to supporters of other parties, the less likely they were to switch: looking again at those who had not voted Conservative in 1987, while 19% of those who did not name a Labour supporting conversation partner switched to the Conservatives in 1992, only 2% of those who talked to three Labour supporters did so. Conver sations can persuade individuals to change their minds. However, where one's conversation partners agree with one, discussion can reinforce one's views: among those who voted for a party in 1987, the more they talked to other supporters of the party, the more likely they were to stay loyal to it 5 years later, while the more they talked to supporters of other parties, the less likely they were to stay loyal.

The impact of conversations on individuals' views varies from person to person, however. Those who have strong political preferences are less likely to be influ enced than are those who do not (**Table 4**). The chances that those who had not voted for a party in 1987, and who did not generally identify themselves with it, would switch to it in 1992 were almost nonexistent for those

Table 4 Strength of partisanship and the influence of conversation, 1992

(a) Respondent did not identify with Conservatives and did not vote Conservative in 1987:

	% voting Conservative in 1992 Respondent had:	
	No Conservative discussants	*1+ Conservative discussants*
Very strong party identification	0.5	0.0
Fairly strong party identification	2.1	2.7
Not very strong/no party identification	5.1	23.6

(b) Respondent did not identify with Labour and did not vote Labour in 1987:

	% voting Labour in 1992 Respondent had:	
	No Labour discussants	*1+ Labour discussants*
Very strong party identification	0.8	0.0
Fairly strong party identification	1.8	2.7
Not very strong/no party identification	4.7	13.9

(c) Respondent did not identify with Liberal Democrats and did not vote Liberal/SDP Alliance in 1987:

	% voting Lib Dem in 1992 Respondent had:	
	No Lib Dem discussants	*1+ Lib Dem discussants*
Very strong party identification	1.0	5.0
Fairly strong party identification	3.1	11.3
Not very strong/no party identification	6.1	22.1

SDP, Social Democratic Party.
From: 1992 BES.

who identified very strongly with a second party, ir respective of whether they talked to supporters of the first party. However, among those who had not voted for the first party in 1987, did not identify with it, and also did not identify strongly with any other party, the chances of switching to the first party in 1992 were be tween three and five times higher for those who spoke to at least one supporter of the first party than among those who spoke to no supporters of the party. Undecided voters are more easily swayed than are partisans.

However, cross national research suggests the con versation effect is, to some extent, culturally specific. For instance, Japanese voters are less willing than Americans to give their opinions to others, and hence are less likely to encounter the heterodox views of other citizens.

Key's 'friends and neighbors' voting, in which voters in a particular area give greater support to candidates with local connections than to candidates from outside the area, is a distinctive variant of the neighborhood effect. It has been reported at a variety of scales, from local gov ernment contests to US presidential elections. It is rare for a presidential candidate to lose his home state, for instance, even when his campaign stalls throughout the rest of the country. At a more local scale, Parker's (1982) analysis of election returns in the 63 electoral districts of the multimember Galway West constituency at the 1977 Irish general election reveals that the further from the district a candidate lived, the fewer votes the candidate gained there (**Table 5**: the main exception, Molloy, turned out to have family connections to the west of the constituency, where his vote was strong).

Table 5 Friends and neighbors voting in Galway West, 1977: correlation between vote share and distance from candidate's home

Political party/candidate	*Correlation coefficient*
Fianna Fail	
Geoghegan-Quinn	0.20
Loughnane	0.73
Molloy	+0.27
O'Morain	0.66
Fine Gael	
Byrne	0.65
Coogan	0.50
McCormack	0.38
Mannion	0.85
Labour	
Higgins	0.48
Sinn Fein	
Coffey	0.15

From: Parker, A. J. (1982). The friends and neighbours voting effects in the Galway West constituency. *Political Geography Quarterly* 1, 243 263, table 2.

However, the neighborhood effect is neither sufficient nor necessary for the creation of a distinctive electoral geography. As discussed below, voters can also be influ enced by what they see around them, even if they do not talk to others in their local communities.

Valence Politics: Economic Geography and Electoral Geography

Ideally, voters carefully compare political parties' pol icies and ideologies against their own preferences. Sur veys show voters are guided by their ideological predispositions and attitudes. Such 'issue voting' has be come more important over time as de alignment has reduced the influence of class and partisan loyalties. But at the same time, two findings recur in the empirical literature: most voters, most of the time, have little if any detailed knowledge of particular policy areas, let alone where the parties stand on them; and voters' attitudes are often volatile, implying that they are not deeply held. The electorate is not, on the whole, well informed and does not have deep seated, clear preferences.

However, voters' lack of knowledge may be rational. Modern governments are complex, legislating over very many fields, and much of their activity is both highly technical and controversial. Few voters have the re sources (of time, as much as anything) really to get to grips with even a few policy areas. At the same time, the chance that their personal vote will be decisive in an election is small, and (since they vote for a party and its entire program, rather than for individual policy items), the chances of supporting a party with which they agree on every issue are remote. Given these facts, rational ignorance is a sensible strategy: while a broad idea of what one stands for is useful, holding detailed opinions on every policy area is not.

So how are voters, especially in a de aligning elector ate, to choose sensibly between parties? One answer is valence politics: voters decide on the basis of broad judgments of 'what works'. I may not understand economic theory, may be confused by international politics, may be ignorant of social policy, but I do have a useful yardstick against which to judge governments especially: have they improved my standard of living or not? If the government can make me better off, it must be doing something right and hence probably deserves my support. If I become worse off, I might conclude the government is failing, and hence that I should support the opposition. This is the essence of valence politics: results matter.

Much research on valence politics has concentrated on economic performance. Government popularity varies over time in step with the state of the economy. Some studies examine the impact of actual national economic indicators, most commonly the unemployment and in flation rates. Others look at public perceptions of eco nomic performance – do people feel things are getting better or worse over time? Most studies confirm that the better the economy performs, and the more affluent voters feel, the more support the government of the day enjoys. In part, this is due to the effect of general eco nomic movements on voters' personal circumstances (egocentric voting – do I feel better or worse off?). But in part, too, it reflects a more contextual, or sociotropic, effect – irrespective of how I personally am faring, how are other people doing?

Two factors limit the general applicability of these findings, however. First, what 'counts' as a good or bad economic performance is in part contextual. Over time voters can become accustomed to economic conditions which, a few years before, might have seemed highly problematic. Unemployment levels which would have been deemed unacceptably high in the 1960s were often seen as a marker of economic success after the high unemployment of the 1980s. This does not, of itself, invalidate the economic vote model, but it does suggest that models need to be periodically recalibrated.

Second, the extent to which governments are held to account for economic performance depends on how easy it is for voters to hold particular parties responsible for success or failure. Economic performance has a stronger impact on government support in countries with a tra dition of single party majority governments than in countries where multiparty coalition governments are the norm, and where accountability for policy can be less clear cut.

Economic voting has geographical implications. Eco nomic performance varies from place to place. While some areas enjoy prosperity, others suffer austerity – and this geography can change over relatively short time periods. This means that voters' egocentric and socio tropic evaluations may vary, depending on where they live. Individuals are more likely to be out of work themselves in areas where unemployment is high than in areas where it is low. But even where two individuals have identical levels of personal prosperity, they may have different views of the government's handling of the economy if one lives in a depressed region and the other in a thriving one. As a consequence, the changing re gional and local geographies of support for governments are strongly influenced by the waxing and waning of regional and local economies: *ceteris paribus*, voters in prosperous areas support the government, while those in declining areas oppose it.

We can illustrate this by examining recent British elections. A first step is to demonstrate that voters are aware of their local economies. The 1997 British Election Survey (BES) asked respondents how they felt their personal financial situation had changed over the previ ous year, how they thought the national economy had changed over the same period, and how they felt their local area had fared relative to the rest of the country since the 1992 election. Overall, the period between 1992 and 1997 was a good one for the UK economy: un employment fell almost everywhere and the economy grew quite strongly. However, the strength of the re covery varied from place to place. Voters noticed. Their opinions of economic change at all three scales (personal, regional, and national) were sensibly related to actual changes. Those who felt things had improved tended to live in areas with the largest average falls in unemploy ment, while those who felt things had worsened lived in areas with the lowest average falls (**Table 6**). And while local economic conditions were only relatively weakly related to evaluations of personal income, they were re lated more strongly to evaluations of the national, and most strongly of all to evaluations of the regional economy.

Furthermore, voters' party choices are influenced by their evaluations of their local area's economic performance. At the 1997 UK election, a Conservative government was defending its record against the Labour opposition. Among those who felt their local economy was performing much better than the national average, a near majority (49%) voted Conservative, while only a third voted Labour (**Table 7**). However, as perceptions of regional performance worsen, government support falls and opposition support grows. Further analysis demon strates that the effect is strongly influenced by voters' attributions of responsibility for the state of the local economy (**Table 8**). Among those who held the gov ernment responsible for the state of their local economy in 1997, support for the Conservatives was 67% among those who felt their region had become much more prosperous, but only 7% among those who felt it had become much less well off: the equivalent figures for Labour support were 18% and 73%, respectively. But among voters who did not hold the government re sponsible, there was no real change in support for the government and opposition between those who felt the local economy had performed well and those who felt it had fared badly.

Gubernatorial elections in the US reveal another di mension to the economic geography of elections. Vote shares for incumbent governors standing for reelection in states where the local economy was dominated by con ventional industries or by services were higher where the state economy was doing well than where it was per forming badly. However, the reelection chances of gov ernors in agricultural states were unaffected by local economic performance. This reflects the likely attribution of responsibility. Many things which influence profits in agriculture are beyond a governor's control – the weather, commodity prices, etc. – and it would not be reasonable to hold the incumbent responsible for them. But in states dominated by manufacturing and service industries, the incumbent is likely to have a greater direct influence on state economic performance. Attribution of responsibility matters.

Table 6 Retrospective economic evaluations and local economic conditions, 1997

Situation got	*Mean % change in constituency Number unemployed, 1995 1997*		
	Better	*Same*	*Worse*
Area prosperity since last election	30.32	28.99	25.39
UK economic situation last year	30.08	27.30	26.66
Household financial situation last year	29.36	27.96	27.28

From: 1997 BES and constituency data.

Table 7 Perceptions of regional economic performance and vote, 1997 UK general election

Since the last election, local area has become	*% Voting for*		
	Conservative	*Labour*	*Other*
A lot more prosperous	49.0	32.0	19.0
A little more prosperous	36.2	40.0	23.8
Stayed about average	32.2	45.7	22.1
A little less prosperous	18.3	58.7	23.0
A lot less prosperous	10.4	68.3	21.3

The Geography of Electoral Campaigning

The geography of an election can also be influenced by the direct campaigning activities of the parties them selves, which aim to achieve a variety of goals. Campaigns mobilize existing supporters, ensuring they vote for the party. They seek to win over uncommitted voters, whose support may change the election outcome. They try to discourage supporters of a party's opponents from voting, by suggesting that the rival party has no chance of win ning. And they inform voters, of both the main campaign issues and the candidates' identities (which can be useful if, for instance, friends and neighbors voting prevails). In part, this requires campaigns aimed at voters throughout the country. But in part, too, it demands targeting re sources on particular, pivotal contests and communities, where an extra effort might yield electoral rewards. The analysis of electoral campaigning has therefore become an integral part of electoral geography, overturning in the process the accepted wisdom – especially among UK

Table 8 Perceptions of regional economic performance, vote, and attribution of responsibility, 1997 UK general election

Responsibility for local economy	*Government % voting for*		*Other % voting for*	
Since the last election, local area has become	*Conservative*	*Labour*	*Conservative*	*Labour*
A lot more prosperous	67.3	18.4	31.4	43.1
A little more prosperous	47.2	31.8	27.3	47.3
Stayed about average	30.2	49.4	33.8	42.6
A little less prosperous	11.5	66.6	37.8	37.8
A lot less prosperous	7.0	73.1	26.2	47.6

political scientists – that the advent of television cam paigning in the 1950s had rendered local campaigning during a general election a superfluous anachronism.

Effective campaigning requires the efficient allocation of scarce resources across competing demands. In an election campaign, not all areas of the country are of equal importance to a party (especially under a plurality electoral system). In some seats, the party's support is too low to give any realistic chance of winning there. In other seats, its popularity is so high that it can virtually guar antee election, even in a bad year nationally. A rational party will expend few resources fighting in either kind of seat (where the party is unpopular, it may have too few resources to mount an effective campaign, even if it wanted to). Rather, it should concentrate its efforts in marginal areas, where a relatively small number of votes either way could mean the difference between winning and losing. There is substantial evidence that parties do indeed behave in this manner. In Britain, for instance, the main parties identify strategic target seats for extra campaign effort, and largely ignore other seats.

This can be illustrated by the relationship between how much a party spent on its constituency campaign at the 2005 British general election, and how marginal the seat was for the party (we concentrate only on English constituencies in this example). Marginality is defined here as the difference between a party's constituency vote share at the previous (2001) election and (if the party won the seat then) the vote share of the party in second place or (if the party lost the seat) the share of the party which did win. Each party's marginality measure is negative where it lost a seat, and positive where it won. The closer the measure is to zero, the more marginal – and hence competitive – the seat is for the party. Graphs for the Conservatives, Labour, and the Liberal Democrats (the three main parties in the UK) all show similar pat terns (**Figures 5–7**). Campaign spending (as a percentage of the legally permitted maximum in each seat) in 2005 was lowest in seats where the party had lost by a wide margin, reached a peak in the party's most marginal seats, and (for Labour and to a lesser extent the Conservatives) declined somewhat in the party's safer seats.

What is more, constituency campaigning does make a difference to each party's performance (**Table 9**). The more effort parties put into their campaigns locally, the more votes they gather themselves, and the fewer votes go to their rivals. Controlling for how well each party did in each constituency at the preceding election, the more the Conservatives and the Liberal Democrats spent on their local campaigns in 2005, the better they did themselves, and the worse their rivals did. Labour's constituency campaign seems to have misfired in 2005, however. While the party's campaign efforts had the desired negative effect on the Conservative vote, it did nothing to boost either its own vote or its ability to see off challenges from the Liberal Democrats.

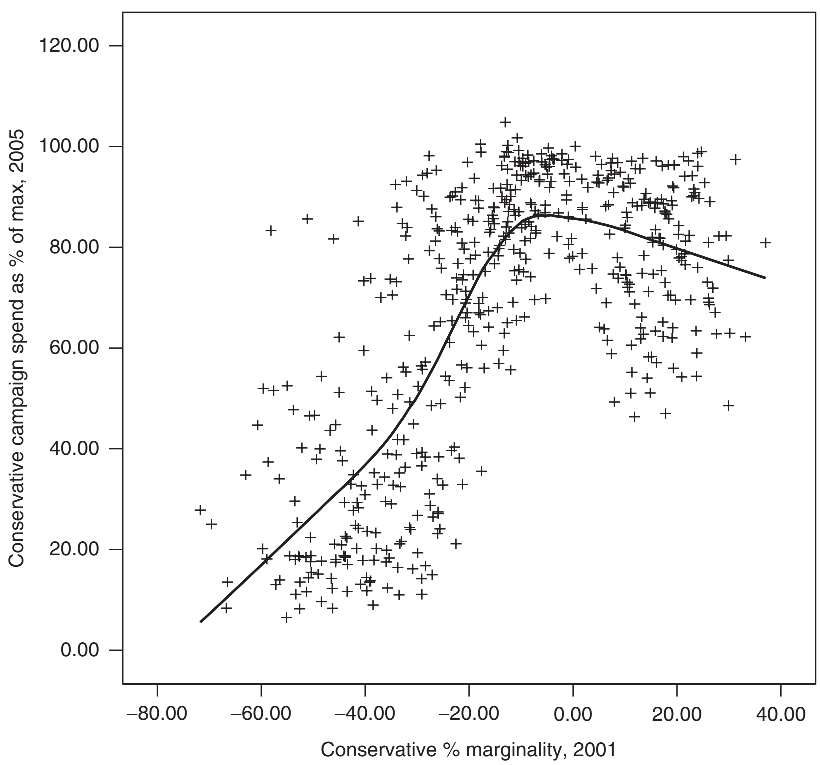

Figure 5 Conservative constituency marginality and campaign expenditure in England at the 2005 general election.

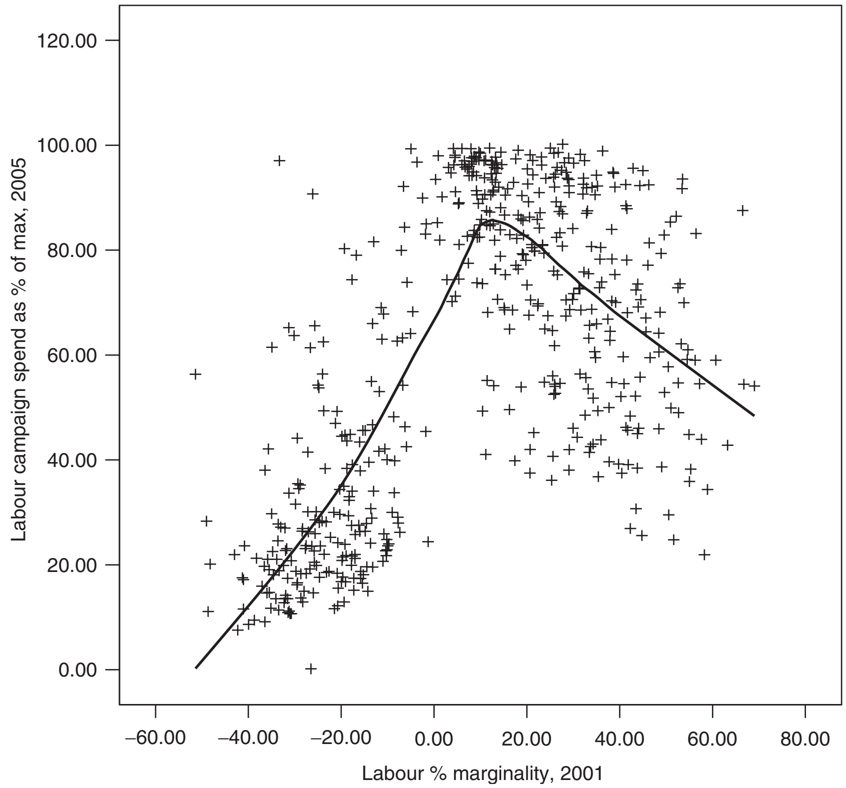

Figure 6 Labour constituency marginality and campaign expenditure in England at the 2005 general election.

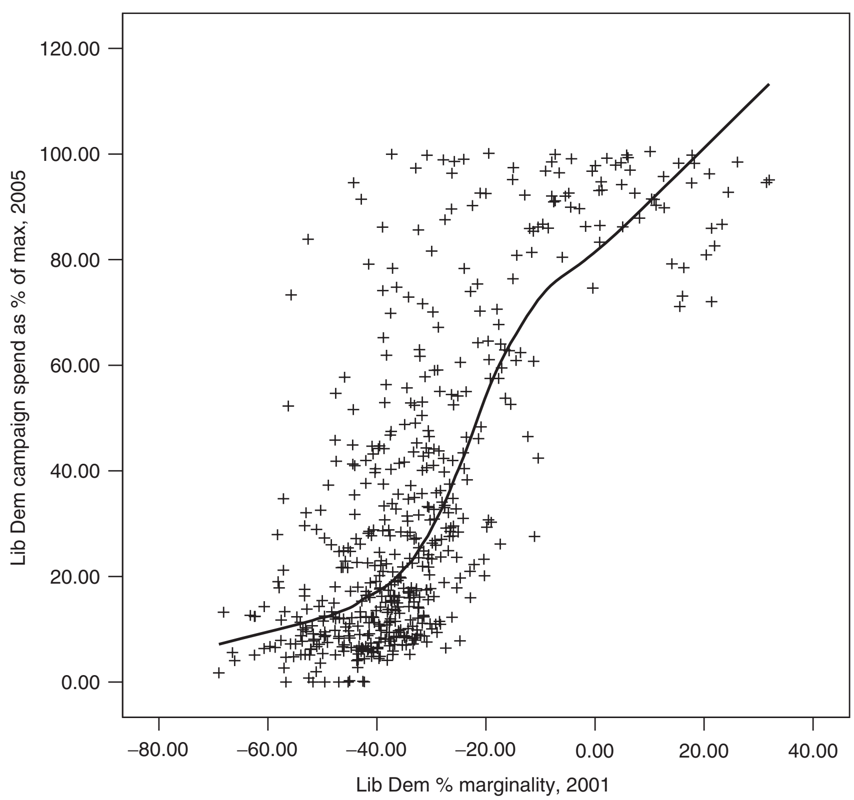

Figure 7 Liberal Democrat constituency marginality and campaign expenditure in England at the 2005 general election.

Table 9 The impact of constituency campaigning on party vote share at the 2005 UK general election in England: regression models

	% vote 2005		
	Conservative	*Labour*	*Lib Dem*
Conservative % vote 2001	0.99[a]		
Labour % vote 2001		0.82[a]	
Lib Dem % vote 2001			0.60[b]
Conservative % campaign spend 2005	0.05[a]	0.04[a]	0.04[a]
Labour % campaign spend 2005	0.01[a]	0.00	0.00
Lib Dem % campaign spend 2005	0.02[a]	0.02[a]	0.12[b]
Constant	1.01	4.87	9.90
R^2	0.95	0.96	0.89

[a] significant at $p = 0.05$.
[b] significant at $p = 0.01$.

Similar results are obtained in US Congressional elections, though here the focus is more on spending by individual candidates than by parties. Incumbent candidates and challengers also differ in the effectiveness of their campaign spending. The more a challenger spends on the campaign, the greater the share of the vote he or she gathers. However, incumbents receive few extra benefits from campaign spending – and may even do worse the more they spend. The reason for this apparent paradox is that, thanks to the publicity of office, incumbents are better known than challengers. Increased campaign spending by challengers raises their profile with the electorate. But increased spending by incumbents is an indication of weakness, since it is likely to be triggered by the threat of a strong challenger. However, this result remains controversial, and its analysis raises difficult methodological questions. Work by Alan Gerber (1998) suggests not only that incumbent spending in US Senate elections has a significant positive impact on incumbent vote share, but that when two stage least squares (TSLS) regression is used in preference to OLS, the impact of incumbent spending on incumbent performance is comparable in size (though opposite in effect) to that of challenger spending (**Table 10**).

This controversy notwithstanding, there is general agreement that campaign spending, especially by challengers, is electorally consequential. There is, however, little evidence that constituency campaigning imparts substantial new information to voters (beyond the identities of the main candidates – though this can of itself win votes). Rather, the main impacts seem to come through a combination of information about the constituency political context and mobilization of one's

Table 10 Campaign spending and incumbent performance in the 1974 92 US Senate elections

	Incumbent % vote share	
	OLS model	*TSLS model*
Challenger spending	5.32[b]	6.50[b]
Incumbent spending	2.58[b]	7.44[b]
R^2	0.57	0.46

[a] Significant at $p = 0.05$.
[b] Significant at $p = 0.01$.
From: Gerber, A. (1998). Estimating the effects of comparing spending on senate election outcomes using instrumental variables. *American Political Science Review* 92, 401 411.

supporters. This can be illustrated by looking at strategic voting and at the geography of electoral turnout.

Strategic voting occurs when a voter's preferred candidate stands no chance of winning a constituency and she votes for her second preference where that candidate has a better chance against a third candidate she strongly opposes. The decision whether or not to vote strategically is contextual, only making sense in the light of one's favored party's prospects in the seat where one votes. Parties can exploit this in their constituency campaigns by drawing attention to the structure of local party competition where this is likely to attract strategic voters to them. At the 2001 UK general election, Labour and Liberal Democrats were ideologically close, and distinct from the Conservatives. In seats where they were in second place behind a Conservative incumbent, or which they held against a Conservative challenge, Liberal Democrat campaign literature pointed out that while a Labour vote there would be wasted (as the Labour candidate had no chance of victory) voting strategically for the Liberal Democrat would provide the best chance of preventing a Conservative victory. Such appeals work. Around 10% of respondents to the 2001 BES said they voted strategically: about 70% of these strategic voters supported either Labour or the Liberal Democrats (**Table 11**). More voters supported the Liberal Democrats for strategic reasons where it was in first or second place in a seat at the previous election than where it was in third place or worse: 30% of Liberal Democrat voters in seats where the party was second to the Conservatives voted strategically, as did 45% of the party's supporters in seats where it was in first place in 1997 and the Conservatives were in second, whereas in seats where the main battle was between Conservatives and Labour, only 20% of Liberal Democrat voters said they voted strategically. In general those who voted Liberal Democrat strategically really preferred Labour – especially in seats where Labour was in third place or worse: almost two thirds of Liberal Democrat strategic voters in seats where a Conservative incumbent faced a Liberal Democrat in second place would have voted for Labour had they not voted strategically.

Table 11 Strategic voting for the Liberal Democrats at the 2001 UK general election

Constituency result		*% voting LD strategically*	*1st preference for LD strategic voters*		
1st	*2nd*		*Con*	*Lab*	*LD*
Con	Lab	20.0	14.3	42.9	42.9
Con	LD	30.0	15.4	65.4	11.5
Lab	Con	20.3	30.8	34.6	23.1
LD	Con	44.9	22.0	54.9	15.9

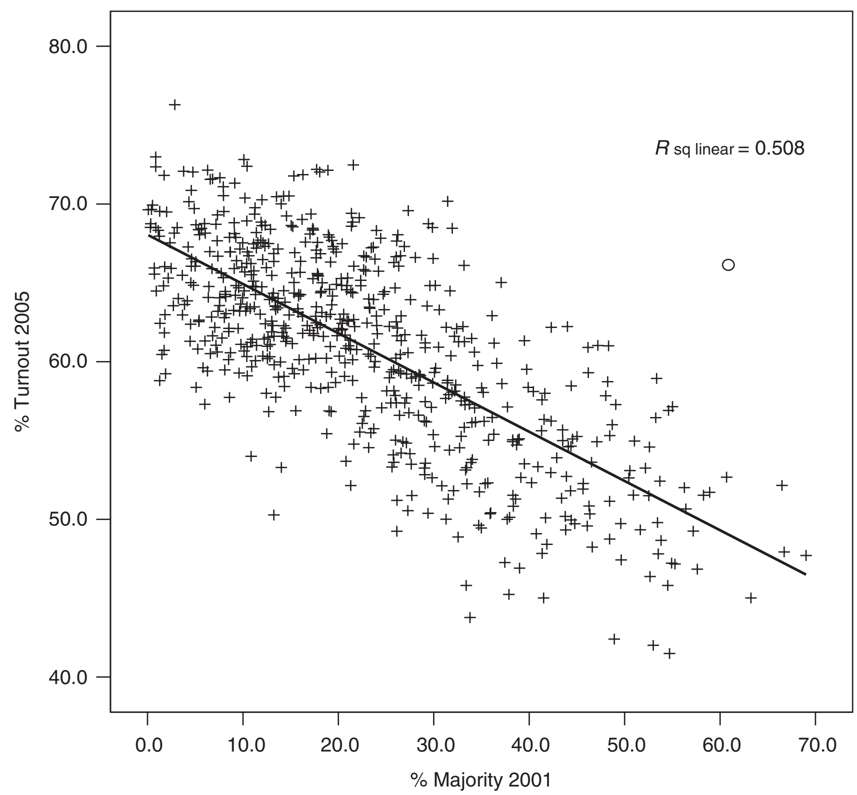

Figure 8 Constituency turnout and marginality at the 2005 British general election.

Constituency campaigns also seek to maximize sup port where it is most needed – in the most marginal constituencies. As the 2005 UK general election showed, turnout is higher in the most marginal seats (where the campaign is fiercest and most effort goes into mobiliza tion) than in safer seats, where there is less to gain (**Figure 8**). The geography of electoral campaigning plays an important role in creating the wider patterns of electoral geography, therefore.

From Votes to Seats: Geography and Electoral Systems

Geography can affect election results in another, more mechanical, manner, through its impact on the con version of votes into seats. This occurs because most electoral systems divide a state up into electoral districts, each of which elects members to the national parliament. However, because party supporters tend to be clustered geographically, the imposition of a grid of electoral dis tricts over a country can affect election outcomes, a consequence of the modifiable areal unit problem. Im agine a country, divided into 9 small areas, each with 1000 voters (**Figure 9a**). Only two parties (L & R) con test elections there, and party R has more supporters than party L. The country employs a plurality electoral system and is divided into three constituencies, each with 3000 voters. There are several ways in which this could be achieved, using the nine small districts as building blocks. Two are shown here. Without any voters chan ging their preferences, we can change the result of the election by using different constituency boundaries. If we define three seats which all run from east to west (panel b:

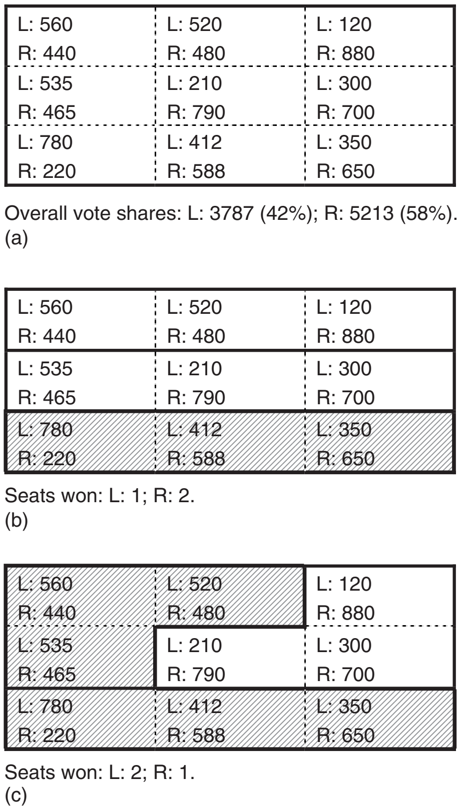

Figure 9 The effect of varying the geography of electoral districts: a hypothetical example: (a) the underlying geography of party support; (b) constituency geography I; and (c) constituency geography II.

the constituency boundaries are the solid lines), party R will win two seats (the two more northerly ones) while party L wins one seat, the shaded constituency in the south of the country. But another possible constituency geography (shown in panel c) would produce only 1 seat for party R and 2 (shaded) for party L: the least popular party would win the election.

Designing real electoral districts is considerably more complex. Taylor (1973) identified a sample of 700 different ways of redrawing Iowa's Congressional districts. In early 1990s Britain, meanwhile, Sheffield's 29 wards could be combined into six constituencies in 1515 different ways, such that no constituency's electorate deviated from the average by more than 8%. But the decision can have partisan consequences. In Sheffield, 1054 of the possible plans would have likely given the Conservatives one seat locally, Labour four, and the Liberal Democrats one. The remaining solutions would have yielded different mixes of partisan outcomes: while some would have left the city with no Conservative MPs, for instance, others would have given the party two seats (**Table 12**).

Furthermore, population change creates a need periodically to revise the boundaries of electoral districts, as some areas grow and others decline. Most modern

Table 12 Simulating the partisan consequences of redistricting in Sheffield, 1992

Distribution of seats			
Conservative	*Labour*	*Liberal Democrat*	*Number of solutions*
1	3	2	64
2	3	1	376
0	4	2	14
1	4	1	1054
0	5	1	7

electoral systems assume that each vote is of equal influence. Failure to revise electoral boundaries in line with population change undermines this aim, however: voters in districts with falling populations end up with more influence, other things being equal, than those in growing districts. Where constituency size varies systematically with party support, this can result in malapportionment. A party which tends to win in the smaller seats is at an advantage over a rival which wins in larger seats as it takes fewer votes on average to elect each of the former's MPs than to elect the latter's. Such malapportionment can occur deliberately or by accident. In the UK, Labour has tended to be the beneficiary of an accidental malapportionment, as it is popular in areas which have lost population steadily since the mid twentieth century, such as the inner cities and Scotland.

Excessive malapportionment is commonly avoided by periodic revisions of constituency boundaries, normally with the aim of making electorates as near to equal in all districts as possible. In the US , for instance, reviews take place after each decennial census. In the UK, Boundary Commissions have reviewed parliamentary constituencies roughly every 11 years since the late 1940s. The rules governing this process vary from country to country, altering the frequency with which reviews take place, the nature of the personnel involved (nonpartisan civil servants or government appointees, for instance), the amount of input from political parties, the extent of public involvement, and so on.

The alteration of constituency boundaries also gives rise to another potential electoral abuse, gerrymandering. In a gerrymander, constituencies are drawn deliberately to maximize one party's representation at the expense of another, as illustrated in **Figure 9**. Gerrymanders can work by concentration – allowing a party to win a few seats by very large margins, hence wasting many of its votes in seats that it wins – or by dispersion – spreading a party's supporters thinly across several seats in order to deny it a winning majority in any one. The abuse is most common in systems which allow parties a strong input – either *de jure* or *de facto* – into the redistricting process. Redistrictings in the US have produced a variety of bizarrely shaped districts which owe their existence to partisan calculations. Richard Morrill's (1973) analysis of

Table 13 The redistricting of Washington State, 1972: expected political impact of alternate plans

	1970 result	*Expected number of seats*			*1972 results*
	1965 districts	*Court*	*Democrat*	*Republican*	*1972 districts*
Senate					
Safe Democrat	8	8	5	9	8
Lean Democrat	13	12	20	10	13
Toss-up Democrat	5	10	6	6	7
Toss-up Republican	5				3
Lean Republican	11	12	10	18	11
Safe Republican	7	7	8	6	7
Total	49	49	49	49	49
House					
Safe Democrat	15	16	10	17	16
Lean Democrat	28	25	40	21	27
Toss-up Democrat	8	19	12	12	14
Toss-up Republican	12				3
Lean Republican	25	24	20	36	24
Safe Republican	14	14	16	12	14
Total	99	98	98	98	98

From: Morrill, R. L. (1973). Ideal and reality in reapportionment. *Annals of the Association of American Geographers* 63, 463 477.

the 1972 redrawing of Senate and House electoral districts in Washington State provides an example. Different proposals for new boundaries were made by him for the courts and by local Democrats and Republicans: those proposed by the parties were designed to maximize the number of seats in which they would be likely to win by relatively narrow rather than large margins (**Table 13**). To minimize the risk of gerrymandering, some countries either exclude parties and the public entirely from redistricting or grant them a relatively minor and restricted role, leaving the final decisions in the hands of independent civil servants (as in the UK).

Gerrymandering and malapportionment are not restricted simply to plurality electoral systems. Boundaries can also be manipulated for partisan advantage in PR systems, for instance by controlling how many MPs are returned from each multimember district. This affects the proportionality of the result. A party might seek smaller constituencies where it is strong (in which case proportionality will decline, and it stands to win more MPs than its vote share strictly entitles it to) and larger constituencies where it is weak (to increase proportionality and hence guarantee it at least some MPs).

The translation of votes into seats can result in electoral bias, whereby one party systematically gains more seats than its vote share strictly entitles it to on a proportional representation basis. A winner's bias is a familiar feature of most plurality systems: the party with the largest vote share normally wins an even larger share of seats in Parliament. The 2005 UK general election provides a striking example. Labour gained the largest vote share of any party, though at just 35%, this was the smallest winning party vote share in the UK in the modern era. Thanks in part to the winner's bias, the party won 55% of seats in Parliament, giving it a clear working majority. But in very close contests, the winner's bias can disappear, and the electoral outcome can be perverse. In the 2000 US Presidential election, Al Gore was marginally the more popular candidate, gaining 48.3% of the national vote compared to George Bush's 48.0%. But the geography of Bush's support was more efficient from the point of view of obtaining support in the Electoral College than was Gore's. While Bush won 271 Electoral College votes, Gore won 266. The slightly less popular Bush became President.

Even where a party loses an election, a biased electoral system can protect its representation. One way of demonstrating bias at a particular election is to apply a uniform swing to all constituencies in order to simulate a result in which both parties tie nationally in terms of vote share. If there is no systematic bias in favor of one party rather than the other, then the two should win identical numbers of seats at equal vote shares nationally. But if, when their votes tie, one party still wins more seats than the other, then the electoral system is biased in favor of the former party. Furthermore, it is possible to break down the sources of bias, in terms of malapportionment, differential turnout, and variations in vote efficiency. If one party is more successful than the other in seats with smaller electorates and lower turnouts, or if it enjoys smaller majorities where it wins and gains smaller vote shares in seats that it loses than does its rival (and hence wastes fewer votes), then a plurality system will tend to be biased in its favor. Electoral bias has now been investigated in many electoral systems, including the US, the UK, and New Zealand.

Scale and Electoral Geography

One of the most striking, though little commented upon, features of the electoral geography literature is that most studies are contained within national or subnational spatial frames. We have studies of American elections, British elections, French elections, national elections, local elections, and so on. But very few have looked at electoral geography on a wider canvas. This is doubly surprising given the rise of comparative politics in the related field of political science. Two notable exceptions stand out.

The international distribution of democracies as op posed to authoritarian governments shows a striking core–periphery pattern. Most core states are functioning democracies while many (though not all – India is a clear exception, for instance) peripheral states are not. Taylor accounts for this in terms of the surplus available for redistribution nationally. In core states, there is sufficient surplus to allow a relatively open political debate over who benefits. But in the periphery limited surpluses raise the stakes: local elites are more reluctant to share their gains with the poor and hence are less likely to allow the poor access to politics through elections. This can lead to a 'politics of failure' in which experiments with electoral democracy in many peripheral states are either short lived, ending in coups, or fail to establish long lasting parties with reliable geographies of support.

Another notable exception can be found in O'Loughlin *et al.*'s work on the spread of democracy, which can be set against the backdrop of initial reactions to the collapse of Communism in the early 1990s. A triumphalist rhetoric emerged among some Western commentators (most notably from Francis Fukuyama) to the effect that democracy had 'won', and would now spread inexorably to all countries. O'Loughlin *et al.* provide an important corrective, showing that while democracy did spread from country to country in a dif fusion like manner, this spread was not inexorable, but was subject to reversals and setbacks. Previous waves of democratization had stalled and been to some extent reversed, and there were no grounds to expect the post Cold War wave to be any different. In the light of international politics post 9/11, his results seem prescient.

These exceptions notwithstanding, however, electoral geography remains primarily concerned with intra national, not international variations, and with elections in established rather than emerging democracies. The development of a global, comparativist electoral geog raphy remains a challenge for the future.

See also: Citizenship; Democracy; Ecological Fallacy; Economic Geography; Electoral Cartography; Electoral Districts; Entropy-Maximising Models; Gerrymandering; Modifiable Areal Unit Problem; Neighborhood Effects; Political Geography; Political Representation; Place, Politics of; Regression, Linear and Nonlinear; Social Class; Social Geography; Social Movements; Spatially Autoregressive Models; Spatial Science; Structuration Theory.

Further Reading

Agnew, J. (1987). *Place and Politics: The Geographical Mediation of State and Society*. London: Allen and Unwin.
Archer, J. C. and Taylor, P. J. (1981). *Section and Party: A Political Geography of American Presidential Elections from Andrew Jackson to Ronald Reagan*. Chichester: John Wiley.
Butler, D. and Stokes, D. (1974). *Political Change in Britain: The Evolution of Electoral Choice* (2nd edn.). London: Macmillan.
Cox, K. R. (1969). The voting decision in a spatial context. *Progress in Geography* 1, 81 117.
Dalton, R. J. (2006). *Citizen Politics: Public Opinion and Political Parties in Advanced Industrial Democracies* (4th edn.). Washington, DC: CQ Press.
Ebeid, M. and Rodden, J. (2006). Economic geography and economic voting: Evidence from the US states. *British Journal of Political Science* 36, 527 547.
Gerber, A. (1998). Estimating the effect of campaign spending on Senate election outcomes using instrumental variables. *American Political Science Review* 92, 401 411.
Gimpel, J. E. and Schuknecht, J. E. (2003). *Patchwork Nation: Sectionalism and Political Change in American Politics*. Ann Arbor: University of Michigan Press.
Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology* 78, 1360 1380.
Gudgin, G. and Taylor, P. J. (1979). *Seats, Votes and the Spatial Organization of Elections*. London: Pion.
Huckfeldt, R. and Sprague, J. (1995). *Citizens, Politics and Social Communications: Information and Influence in an Election Campaign*. Cambridge: Cambridge University Press.
Ikeda, K. and Huckfeldt, R. (2001). Political communication and disagreement among citizens in Japan and the United States. *Political Behavior* 23, 23 51.
Jacobson, G. C. (1978). The effects of campaign spending in Congressional elections. *American Political Science Review* 72, 469 491.
Johnston, R. J. (1983). The neighbourhood effect won't go away: Observations on the electoral geography of England in the light of Dunleavy's critique. *Geoforum* 14, 161 168.
Johnston, R. J. (1986). The neighbourhood effect revisited: Spatial science or political regionalism. *Environment and Planning D: Society and Space* 4, 41 55.
Johnston, R. J. and Hay, A. M. (1983). Voter transition probability estimates: An entropy maximizing approach. *European Journal of Political Research* 11, 93 98.
Johnston, R. J. and Pattie, C. J. (2006). *Putting Voters in Their Place: Geography and Elections in Great Britain*. Oxford: Oxford University Press.
Johnston, R. J., Pattie, C. J., Dorling, D. F. L. and Rossiter, D. J. (2001). *From Votes to Seats: The Operation of the UK Electoral System since 1945*. Manchester: Manchester University Press.
Johnston, R. J., Shelley, F. M. and Taylor, P. J. (eds.) (1990). *Developments in Electoral Geography*. Routledge: London.
Jones, K., Johnston, R. J. and Pattie, C. J. (1992). People, places and regions: Exploring the use of multi level modelling in the analysis of electoral data. *British Journal of Political Science* 22, 343 380.
Key, V. O. (1949). *Southern Politics in State and Nation*. New York: Alfred Knopf.
Key, V. O. (1955). A theory of critical elections. *Journal of Politics* 17, 3 18.
King, G. (1997). *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton: Princeton University Press.

Lazarsfeld, P., Berelson, B. and Gaudel, H. (1948). *The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign*. New York: Columbia University Press.

Lipset, S. M. and Rokkan, S. E. (1967). Cleavage structures, party systems and voter alignments. In Lipset, S. M. & Rokkan, S. E. (eds.) *Party Systems and Voter Alignments*, pp 3 64. New York: The Free Press.

Miller, W. L. (1977). *Electoral Dynamics in Britain since 1918*. London: Macmillan.

Monmonier, M. (2001). *Bushmanders and Bullwinkles: How Politicians Manipulate Electronic Maps and Census Data to Win Elections*. Chicago: University of Chicago Press.

Morrill, R. L. (1973). Ideal and reality in reapportionment. *Annals of the Association of American Geographers* 63, 463 477.

O'Loughlin, J., Flint, C. and Anselin, J. (1994). The geography of the Nazi vote: Context, confession and class in the Reichstag election of 1930. *Annals of the Association of American Geographers* 84, 351 380.

O'Loughlin, J., Ward, M. D. and Lofdahl, C. *et al*. (1998). The diffusion of democracy, 1946 1994. *Annals of the Association of American Geographers* 88, 545 574.

Osei Kwame, P. and Taylor, P. J. (1984). A politics of failure: The political geography of Ghanaian elections, 1954 1979. *Annals of the Association of American Geographers* 74, 574 589.

Paddison, R. (1976). Spatial bias and electoral redistricting in proportional representation systems: A case study of the Republic of Ireland. *Tijdschrift voor Economische en Sociale Geografie* 67, 230 240.

Parker, A. J. (1982). The friends and neighbours voting effect in the Galway West constituency. *Political Geography Quarterly* 1, 243 262.

Prescott, J. R. V. (1959). The functions and methods of electoral geography. *Annals of the Association of American Geographers* 49, 296 304.

Shelley, F. M., Archer, J. C., Davidson, F. M. and Brunn, S. D. (1996). *Political Geography of the United States*. New York: Guilford.

Siegfried, A. (1913). *Tableau politique de la France de l'ouest*. Paris: A. Colin.

Taylor, P. J. (1973). Some implications of the spatial organization of elections. *Transactions, Institute of British Geographers* 60, 121 136.

Taylor, P. J. and Flint, C. (2000). *Political Geography: World Economy, Nation State and Locality* (4th edn.). Harlow: Prentice Hall.

Taylor, P. and Johnston, R. J. (1979). *Geography of Elections*. London: Croom Helm.

Tingsten, H. (1937). *Political Behavior: Studies in Election Statistics*. London: PS King and Son.

Whitten, G. D. and Palmer, H. D. (1999). Cross national analyses of economic voting. *Electoral Studies* 18, 49 67.

Embeddedness

M. Hess, University of Manchester, Manchester, UK

Introduction

Human geography has witnessed a number of epistemological changes over the last half century. These include the quantitative revolution, behavioral approaches, radical political economy, and, more recently, the cultural and relational 'turns', which also resonated in economic geography. In the context of the latter, embeddedness has become a widely used concept. It refers to the social, cultural, and institutional basis of economic activity and has been applied in numerous studies since the 1990s. For many authors, embeddedness has been a key element in explaining the successful economic development of regions, based on specific social arrangements and relationships that help to establish mutual trust and shape collective strategies and goals. By emphasizing the social and institutional foundations of economic behavior, the embeddedness concept – defined as a set of ongoing social relations – moves beyond an atomistic reading of human action and rejects the orthodox economic notion of a 'homo oeconomicus'. Thus, it represents an analytical framework better situated to deal with the complexities of real world socioeconomic development. Like most theories used in human geography, the origins of the embeddedness concept are to be found outside the discipline of geography and can be traced back to the 1940s.

Intellectual Origins of the Concept

The first time the term embeddedness was used in a conceptual way was in the seminal work by the Hungarian intellectual Karl Polanyi, *The great transformation*. Published in 1944, this book was a major stepping stone in rethinking the relationship between economy and society. Polanyi's central argument is that nonmarket systems of economic exchange are embedded in society and enmeshed in institutions. He identifies three different institutional patterns of organizing economic activities: householding, reciprocity, and redistribution. These systems of exchange are based on shared values and norms, and constituted not by monetary goals but by common social and cultural bonds. Hence, economy is seen as a part of the broader society, an instituted process or what Polanyi calls a substantive definition of economics. This he puts in contrast to the formal economics applied by market ideologists and economists at his time and based on ideas of profit maximizing behavior, rational calculation, and unlimited information in a self regulating marketplace. Analyzing the economic history of nineteenth century Europe to illustrate his argument, Polanyi demonstrated how modern market economies became increasingly disembedded from society. Under the free market doctrine, he concludes, instead of economy being embedded in social relations, social relations are embedded in the economic system by monetarizing and economizing the cultural and social elements of society.

Critics have pointed to the fact that Polanyi describes modern market economies as disembedded, although in reality even the most liberal economies arguably are enmeshed in institutions, that is, embedded. This critique, however, is based on a misreading of his embeddedness concept, as Polanyi, whose argument is indeed a complex one and hence easy to misinterpret, sought to expose the utopian nature of pure market societies as advocated by economists and political ideologists at his time. For Polanyi, every exchange system including the market system is socially constructed and embedded, a central insight about the economy–society relationship that subsequently would become highly influential in economic sociology and cognate disciplines. One school of thought that echoes the arguments put forward by Karl Polanyi is the National Business Systems literature as developed by Richard Whitley and his colleagues. Albeit not explicitly using the term embeddedness, they share the view that society and its institutions are crucial in shaping the economy, providing different systems of coordination and control that ultimately embed economies in societies in variegated ways. The most influential writings, however, to firmly establish the notion of embeddedness within the social sciences after Polanyi, stem from the work of economic sociologist Mark Granovetter. He can be credited with coining the term 'new economic sociology' and has since become the most widely cited author of the embeddedness literature.

In his seminal 1985 article, 'The problem of embeddedness', Granovetter picks up Polanyi's main theme, the relationship between the economic and the social. His approach departs from Polanyi – the father of the embeddedness concept – in a number of significant ways. While both authors show their dissatisfaction with the absolutization of the market in orthodox economics, Granovetter arguably attempts to reconcile economics and sociology rather than dismiss the discipline and its intellectual project as ultimately utopian, as Polanyi does. While the latter emphasizes the historically contingent nature of society–economy relationships, Granovetter aims for universal explanations of the social embeddedness of economic action. In his view, the concept of embedd edness serves to avoid both the undersocialized views of economic action such as homo oeconomicus behavior and rational choice assumptions in neoclassical economics and the oversocialized views of human action as based solely on internalized values and norms, a concept drawn mainly from Parsonsian sociology.

At the center of Granovetter's embeddedness concept are not the abstract, structural relations between economy and society, but the concrete personal relations and struc tures of these relations (i.e., networks) that affect economic action, outcomes, and institutions. In other words, it is the relational (dyadic) and structural (network) embeddedness that explains economic action and outcomes. Thus Gran ovetter 'scales down' Polanyi's original concept of embeddedness by switching the focus toward the analytical scale of individual and collective actors in social systems of exchange, rather than the systems themselves.

In the wake of Granovetter's 1985 article, a vast body of literature has developed in the social sciences that used embeddedness as a central analytical category. Many attempts have been made to refine the concept and produce more differentiated understandings of embedd edness by further deconstructing the notion of social embeddedness. One typology, for instance, proposed by Sharon Zukin and Paul DiMaggio, distinguishes four forms of embeddedness: cognitive, cultural, structural, and political. Other classifications followed suit and have been applied in organization and business studies, among other disciplines (see **Table 1** for a summary). At the beginning of the 1990s, the first researchers in economic geography also started to adopt the concept.

Economic Geography and Embeddedness

In hindsight, we are able to identify two major factors that contributed to the growing interest in noneconomic factors by economic geographers over the last 20 years. At an epistemological level, there was an increasing awareness among many geographers of the limitations of mainstream approaches to economic geography. More and more human geographers became concerned about the 'spatial fetishism' of the discipline. This is not to say that social interaction has never been a subject of geo graphical enquiry. Indeed, there have been numerous critical voices in the radical literature of the 1970s and 1980s. However, they often have been neglected or ig nored in a thitherto prevalent tendency of treating social relationships as purely spatial relationships. Over the last two decades, this was increasingly criticized, calling for a different ontology of space as being socially constructed and shaped by human practices (relational space).

At a more empirical level, the emergence of the embeddedness concept in economic geography was as sociated with a growing interest in the analysis of tech nologically advanced and economically competitive regions since the late 1980s. It was this shift of attention away from declining and old industrialized areas and toward these new types of regions – variously called creative milieus, learning regions, or networked regions – that coincided with a conceptual shift away from more traditional explanatory tools such as transaction cost theory or agglomeration economics toward the inclusion of noneconomic factors like embeddedness, social capital, and trust as determinants of regional socioeconomic development. Thus a large body of work evolved that has since become known as 'new regionalism' and has had an enormous impact not only in the academic discipline of human geography but also in policy making circles from the local to the supranational scale.

Two of the most influential articles that helped to firmly establish the embeddedness concept in geography and regional studies have been published in 1992. The first, written by Peter Dicken and Nigel Thrift for the Transactions of the Institute of British Geographers, set out to reconsider the spatial and geographical conditions

Table 1 Different views on embeddedness

	Who?	*In what?*	*Geographical scale*
Polanyi's 'Great transformation'	'The economy', systems of exchange	'Society', social and cultural structures	No particular scale, but emphasis on the nation-state
Business systems approach	Firms	Institutional and regulatory frameworks	Nation-state, 'home territory'
New economic sociology	Economic behavior, individuals, and firms	Networks of ongoing social (interpersonal) relations	No particular scale
Organization and business studies	Firms, networks	Time, space, social structures, markets, technological systems, political systems…	No particular scale
Economic geography	Firms	Networks and institutional settings	Local/regional

of economic organizations and enterprises. The authors argued that business organizations are produced through a historical process of embedding whereby they interact with the specific characteristics of their 'home territories' and the characteristics of other regions they operate in. This echoed Polanyi's analysis of the economy–society relationship and provided a powerful critique of increasingly fashionable 'flat Earth' theories where geography doesn't matter anymore in a globalized world. According to Dicken and Thrift, rather than moving toward global homogenization the territorial embeddedness of business enterprises remains crucial in their production and reproduction. In other words, economic actors like firms do things differently depending on the nature of their cultural, political, social, and territorial embeddedness.

The second paper to have a considerable impact early on, authored by the late Bennett Harrison and published in Regional Studies, is a sympathetic critique of the industrial districts literature, initially based on neoclassical agglomeration and growth pole theories from Alfred Marshall to Francois Perroux and Raymond Vernon. While Harrison acknowledges some explanatory power of these approaches, like Granovetter he criticizes the undersocialized (or disembedded) conception of human action in these theories, ignoring or neglecting the 'historical and cultural vestiges' that shape and structure interpersonal and interfirm relationships and therefore should be regarded as an integral part of explaining the formation of industrial districts and similar territorial economic configurations. For Harrison, the industrial district model represents a particularly strong form of embedding economic or business relations into a broader social and societal fabric, ultimately leading to regional economic success through the seemingly contradictory process of cooperation and competition (or co opetition).

The introduction of the embeddedness concept into economic geography has clearly helped to refocus on the social and cultural foundations of economic activity in order to explain regional socioeconomic development. Whereas traditional economic theories of regional development tended to focus on one set of actors only, namely the firms, the embeddedness view also takes into account other groups of actors and nonfirm institutions that are important to create an 'institutional thickness' that is crucial in generating positive economic outcomes for any region through networked forms of local and regional governance. The four factors that contribute to institutional thickness are: (1) a strong institutional presence (e.g., of education and research institutions, unions, government and development agencies, etc.); (2) a high level of interaction between the different types of actors involved; (3) defined structures of domination and/or coalition; and (4) a mutual awareness of being involved in a common enterprise (or what can be called a common economic agenda based on a specific regional identity or culture). A strong and positive combination of these factors serves to consolidate the local and regional embeddedness of the economic system. Such a reasoning resonated well with policy makers at different levels, as it provided concrete starting points for them to try and increase regional competitiveness in an increasingly neoliberal, competitive global environment.

Trying to replicate the success of highly embedded, successful regions – high tech based and otherwise – has since become a major political endeavor, often under the label of 'cluster policies' to stimulate regional economic growth. Arguably, this has helped to reestablish the discipline of economic geography as policy relevant and being recognized by other disciplines, as well as the wider public. However, to some critics the new regionalism of the 1990s with its emphasis on territorial embeddedness as an important factor in achieving innovation and competitive success has led to a spatial fetishization of the local/regional, while the concept of embeddedness has been inextricably linked with the local and regional scale, at the expense of other possible translocal links. Recent work in economic geography has started to question the idea that the local or regional level is the only spatial logic of embeddedness, and has set out to conceptualize embeddedness in a way that helps to take space and scale seriously in analyzing embedded relationships, while at the same time avoiding the hitherto overterritorialized views of the new regionalism literature.

Spaces and Scales of Embeddedness

It certainly is to the credit of economic geography that space has become a significant element in conceptualizing embeddedness. However, as discussed above, most of the literature has shown a tendency to overemphasize the local and regional scale of embeddedness as the most important. One of the main reasons for this view – apart from the role of institutional thickness and the continued importance of Marshallian agglomeration economies – is the assumption that spatial proximity and hence local embeddedness is essential to generate trust and avoid malfeasance in regional networks of economic activity. The underlying argument is that trust is difficult to achieve over long distances, as face to face interaction is needed to create trust in the first place. Following this very same logic, it is argued that globalization and space–time distance are strongly linked to a state of disembeddedness, whereby social relations are detached from their local and regional context of interaction. That point has been made in Anthony Giddens' analysis of modernity, where localized, personal relationships based on trust have been replaced by trust in reliance upon abstract expert systems like supranational organizations, scientific bodies, or political and financial institutions. Going back to original conceptualizations, this bears a strong resemblance to Polanyi's analysis of the dis embedded market society.

So far, however, no theoretical foundation of the scalar argument has been provided to substantiate the claim that trust – and therefore embeddedness – can only occur in spatial proximity. On the contrary, Giddens himself argues that trust is related to absence in time and space, as trust is not necessary if other peoples' activities are constantly visible. By the same token, he states that the disembedding power of globalization has not eradicated the importance of personal relations, but these, too, now stretch over time and over space, so that personal trust has become de localized as well. In other words, the multiscalar nature of embeddedness has become some thing that needs to be taken into consideration when applying the concept in economic geography and other social sciences. Integrating the insights of original the orists like Polanyi and Granovetter and scrutinizing the spatial, scalar, and not least temporal nature embedding (and disembedding) processes, the very concept of embeddedness has recently been reworked to provide a comprehensive yet consistent typology of embeddedness as a heuristic framework to analyze society and economy in space. Three basic dimensions of embeddedness – analytically distinct but strongly interwoven – have been put forward: societal, network, and territorial embedd edness (see **Figure 1**).

Based on Polanyi's initial ideas, societal embeddedness considers the extent to which economic action is influ enced by the social and cultural background of the actors. As many studies have demonstrated, societal embedd edness or where an actor comes from is important in analyzing economic behavior and economic outcomes. As many studies have shown, companies operating in a global economy still tend to keep many of their 'home territory' characteristics and how they act abroad is – despite globalization – still very much grounded in their societal 'heritage'. Likewise, work on transnational entrepreneurs has demonstrated how migrant com munities – while becoming enmeshed in their host re gions' economy – still keep strong personal and business relationships with their countries of origin, again dis playing a high degree of societal embeddedness.

Following Granovetter's work, network embeddedness refers to the fact that individuals and organizations do not act in isolation, but always within networks of rela tionships, regardless of their countries of origin or the places these actors operate in. In such a relational view, the degree of embeddedness depends on the stability and durability of relationships (relational embeddedness) and networks (structural embeddedness). Network embedd edness therefore emphasizes the fact that economic ac tion – despite considerable autonomy of actors – is neither purely economic and rational, nor is it solely based on an actor's inscribed attributes and internalized social norms and values, but always in relation to some one else. The process of embedding and disembedding thus can be seen as a result of building, retaining, and losing trust between the different network agents.

The extent to which an actor or set of actors is 'an chored' in particular places and regional networks is captured by the notion of territorial embeddedness. While it is not the only spatial logic of socioeconomic relations, as discussed above, it clearly is an important element of the space economy. Actors become territo rially embedded by absorbing and – in some cases – being

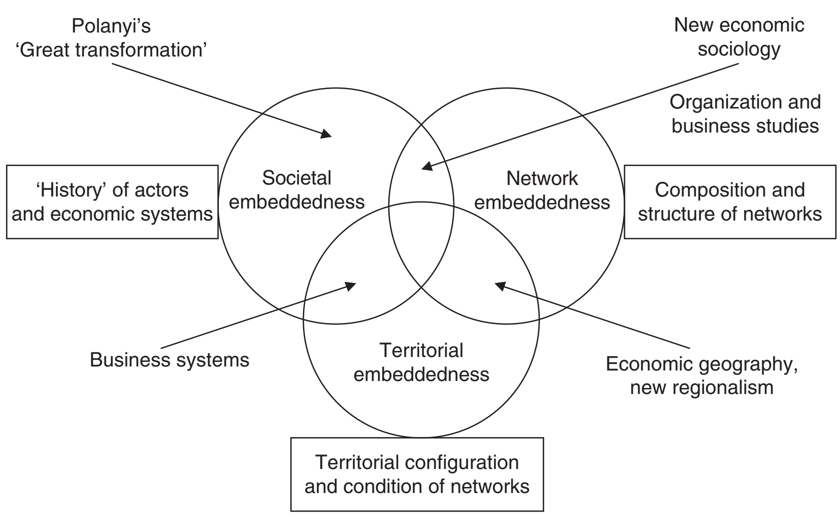

Figure 1 Fundamental categories of embeddedness. Reproduced with permission from Hess, M. (2004). 'Spatial' relationships? Towards a reconceptualization of embeddedness. *Progress in Human Geography* 28(2) p. 178, © Sage Publications, 2004, by permission of sage Publications Ltd.

constrained by the socioeconomic dynamics existing in those places. This latter point indicates that embeddedness is not *per se* and automatically beneficial to a region's economic development, as suggested in much of the literature. Rather, territorial embeddedness can stem from an over reliance on local network resources that ultimately leads to lock in effects, undermining the region's potential for development. This has been the case in many old industrialized regions where such a lock in has resulted in limited adaptability to changing global economic dynamics. Whatever the developmental outcome of territorially embedded networks may be, place obviously still matters in a global space of flows, with regions forming nodes in global networks and thus creating islands in an archipelago economy.

From a human geography perspective, the threefold typology of societal, network, and territorial embeddedness as a spatial–temporal concept can be associated with a typology of space as developed in actor network theory. Thus, societal embeddedness corresponds with what Annemarie Mol and John Law call fluid spaces, formed by actors moving between places like migrant labor or transnational entrepreneurs and enterprises. Network embeddedness reflects the topological space of networks, in which distance and proximity are not measured in Euclidian terms, but are a function of the relations between actors. Territorial embeddedness, finally, links to a notion of space as region, where objects are clustered together and it is possible to draw boundaries, in other words the hitherto most common understanding of embeddedness in economic geography. What is important to note, however, is the fact that it is the simultaneity of all three forms of embeddedness that ultimately shapes networks and economic action in space and time.

Whither Embeddedness?

Without doubt, the concept of embeddedness has gained prominence in the social sciences. Firmly grounded in network ontology, it represents a relational view of (economic) action and interaction. Equally important, embeddedness emphasizes the social and cultural nature of economic action while attempting to avoid the under and oversocialized views of orthodox economics on the one hand and some strands of sociology on the other, with their respective assumptions of atomistic action. Some scholars have noted that the embeddedness concept has become the main paradigm in economic sociology, or what since has become known as 'new economic sociology'. However, its growing impact does not mean the concept remains uncontested. In recent years, it has been criticized from different perspectives for its perceived limitations and contradictions. The main issues questioned by critics of the concept are: (1) the theoretical foundations of embeddedness and its interpretation of 'the market'; (2) the importance of structural factors in economy and society; (3) the nature of the economic and the social, and their relation; and (4) the role of power asymmetries in interpersonal and interfirm relations and the role of politics. Some of these critical comments are addressing not only embeddedness, but economic sociology and its relation to the discipline of economic geography more broadly.

In a recent publication, US sociologist Greta Krippner engages in a critical analysis of Granovetter's embeddedness concept. While recognizing his attempts to develop a 'corrective' to neoclassical economics and 'old' sociology, Krippner argues that Granovetter does not succeed as through his concept he reifies the market as an economic fundamental rather than really overcoming economistic definitions of the market. This stands in contrast to Polanyi's attempts of demonstrating how the market economy and the market society are ideological and ultimately utopian constructs developed in formal economics. Therefore, it is argued, new economic sociology and embeddedness as a concept exist in a symbiotic relationship with mainstream economics and reacts to theoretical developments in orthodox economics rather than transcending it. As Krippner and other commentators like Jamie Peck state, the core argument of markets being socially constructed and embedded has led the proponents of embeddedness to take the market for granted.

That view directly links to a second strand of critique, namely of the epistemological assumptions about the nature of the economic and the social. The embeddedness literature is said not only to reify the market, but to purify the economic from the social thus sustaining an artificial dualism of the social and the economic. This critique has been brought forward from two sides: authors in the political economy tradition have raised it, as well as proponents of the cultural turn that are in favor of a more practice centered approach in social sciences and human geography. While the academic debate on embeddedness currently seems centered around this issue, it is interesting to note that all the above approaches arguably have difficulties in providing an alternative that fully transcends the dualism of the social and the economic, as they all rely on some distinction between the two, at least for analytical purposes.

Another criticism of the embeddedness concept that has been raised by various commentators is its usefulness for critical and radical social science and the associated policy relevance of the concept. As discussed earlier, the concept is appealing for policy makers at different scales, but according to its critics suffers from a depoliticized view of economic development by largely ignoring power asymmetries between actors and institutions. By emphasizing relationships of trust and reciprocity, among other things, embeddedness has come to be signifying a positive, benign form of development based on col laboration and equality. Indeed, as much of the embeddedness literature has empirically focused on successful regions and positive outcomes, there seems little room to consider the dark side of embeddedness. This, however, is not something that is immanent to the concept. The fact that interpersonal or interfirm ties are embedded does not exclude the possibility of power struggles and negative economic outcomes. Examples of corruption, old boys' networks, or nepotism as particular forms of embedded ties may suffice to illustrate this point.

Embeddedness as a concept has come a long way and to date had a considerable career in the social sciences. Economic geography has been among the disciplines taking up the concept to explain developments in the space economy. To this end, embeddedness as a heuristic tool had and still has a lot to offer for analysing socio economic development in space. Recent conceptual developments have made some progress in addressing the criticisms articulated regarding its theorization, and, as with all frameworks, there certainly is room for improvement to strengthen its clarity and intellectual potential. The concept is still widely used and is in forming many studies in economic geography and cog nate disciplines. To what extent it will continue to do so and what its place among a range of different approaches will be in the future remains, of course, to be seen.

See also: Actor-Network Theory/Network Geographies; Local-Global; Locality Debates; Networks; Relational Economic Geography; Structurationist Geography; Transnational Ethnic Networks.

Further Reading

Barber, B. (1995). All economies are 'embedded': The career of a concept, and beyond. *Social Research* 62, 387 413.

Dicken, P. and Thrift, N. (1992). The organization of production and the production of organization: Why business enterprises matter in the study of geographical industrialization. *Transactions of the Institute of British Geographers NS* 17, 279 291.

Granovetter, M. (1985). Economic action and economic structure: The problem of embeddedness. *American Journal of Sociology* 91, 481 510.

Hadjimichalis, C. (2006). Non economic factors in economic geography and in 'new regionalism': A sympathetic critique. *International Journal of Urban and Regional Research* 30, 690 704.

Harrison, B. (1992). Industrial districts: Old wine in new bottles? *Regional Studies* 26, 469 483.

Hess, M. (2004). 'Spatial' relationships? Towards a reconceptualization of embeddedness. *Progress in Human Geography* 28, 165 186.

Hess, M. (2006). Spatial relationships. In Ritzer, G. (ed.) *Blackwell Encyclopedia of Sociology*, pp 4633 4635. Malden and Oxford: Blackwell.

Jessop, B. (2001). Regulationist and autopoieticist reflections on Polanyi's account of market economies and the market society. *New Political Economy* 6, 213 232.

Krippner, G. (2001). The elusive market: Embeddedness and the paradigm of economic sociology. *Theory and Society* 30, 775 810.

Peck, J. (2005). Economic sociologies in space. *Economic Geography* 81, 129 175.

Polanyi, K. (1944). *The great transformation. The political and economic origins of our time*. Boston: Beacon Press.

Embodied Knowing

R. Longhurst, University of Waikato, Hamilton, New Zealand

Glossary

Bodies Bodies are sites of identity, morality, estheticism, practical action, work, play, pleasure, and pain. They have a materiality – skeleton, flesh, organs, nerves, muscles, and blood – but this materiality is always constituted within discourse. Feminists have made the point that there is, of course, no one body – 'the' body is an illusion. Bodies are usually sexed bodies and therefore much discussion focuses on the complex processes through which female and male bodies are differentiated. Geographers have added to debates about bodies that bodies are always 'located', that is, they are inextricably entwined with place.

Corporeal, Corporeality Refers to the physicality, substance, materiality, or matter of the body. As some social scientists have pointed out, however, this matter of the body never sits outside of discourse or systems of social and cultural signification.

Discourse The notion of discourse is complex, mainly because there exist many conflicting and overlapping interpretations from a range of disciplinary and theoretical standpoints. Basically, however, two general understandings of discourse can be identified. First, in linguistics, discourse tends to refer to examples of either written or spoken language. Second, in social theory, discourse tends to be understood in broader terms. Although social theorists recognize the centrality of language to understanding discourse, they use the term to refer to ways of structuring knowledge and social practice.

Dualism When a continuous spectrum is divided into discrete self-contained elements which exist in opposition to each other this is known as a dualism. Dualism is the principle that there are two mutually exclusive types of 'things', for example, white/black, good/bad, rich/poor, public/private, man/woman, and mind/body.

Feminist Geography A body of literature and series of questions prompted by feminist politics and theory about the ways in which gender relations and spaces are mutually constituted. Feminist geography began to emerge in the 1970s. Initially research aimed to document women's absence from the discipline and then to make women visible by examining how women's experiences often differ from men's. More recently feminist geographers have drawn on a broader range of social and cultural theory in order to develop a more complex understanding of the diverse range of women's experiences.

Masculinist Some feminist geographers have argued that the discipline of geography is 'masculinist'. This term describes work that claims to be exhaustive but in actuality forgets about or ignores the worlds of women. It concerns itself solely with the existence of men and masculinity.

Other, Othering A process by which some individuals, groups, and categories become constructed (psychically, discursively, and materially) as the Other, that is, as 'different' from the Self. The Other is understood to be lesser than and/or to lack in some way. The process of Othering is associated with unequal power relations that exclude and marginalize some individuals, groups, and categories.

Post-Structuralist Feminism A version of feminism that emerged mainly from women in ethnic minorities who felt that second-wave liberal feminism placed too much emphasis on sameness and equality rather than on difference. Post-structuralist feminists argue that it is vital to have a theory that accounts for the relationship between the individual and the social, in other words a theory that looks not only at forms of social organization and the values associated with them, but also the subjectivities and actions of individuals.

Reflexivity Critical and conscious introspection and systematic questioning of oneself as a researcher. Researcher reflexivity involves examining one's own embodied subjectivity in order to gain new insights into research.

Situated Knowledge A notion that challenges the idea that knowledge is simply 'out there' waiting to be discovered. Instead knowledge is both 'embodied' and 'situated', that is, it is made by individuals who are situated within particular contexts.

Introduction

Over the past two decades many scholars have critiqued the idea that knowledge is universal, objective, rational, and of the mind, and have instead argued that knowledge arises out of an individual's experience. Knowledge is not simply 'out there' waiting to be discovered. It is partial, situated, and embodied, that is, it is made by individuals who are located within particular contexts. This move toward critically examining the construction of geographical knowledge as partial, situated, and embodied has been particularly strong by those in the discipline who adopt critical perspectives, for example, socialist, antiracist, postcolonial, post structuralist, queer, and feminist geographers. These geographers have argued that focusing on the body is one possible route to changing social, cultural, and economic relations for the better. It is important to recognize that bodies (bodies that have a particular skin type and color, shape, genitalia, impairments, are a specific age, ethnicity, and so on) are always placed in particular temporal and spatial contexts. This has prompted critical geographers, especially feminist geographers, to question the production of knowledge in geography.

Feminist Critiques

Feminist geographers have mounted an argument that a dualism between mind and body underlies, and is integral to, the production of human geography. The mind/body dualism is a presumption, or a foundation, upon which geographical knowledge is constructed. However, not only are the mind and body conceptualized as a dualism but the divisions between mind and body are gendered and sexed. The mind (reason and rationality) has long been associated with masculinity while the body (emotion and irrationality) has long been associated with femininity. What has constituted knowledge in geography is that which appears to be rational, reasonable, public, cultural, productive, masculine, and of the mind. Geography has tended to exclude from its scientific and philosophical discourse that which is thought to be unknowable, irrational, unreasonable, private, natural, reproductive, feminine, and of the body.

Feminist geographers have argued that there is something in the very claim to 'knowing' in geography – a specific notion of knowing, and of knowledge, as objective and disembodied – which tends to have marginalized or excluded women as producers of knowledge. In various and complex ways knowledges have become imbued with a division between rational Man who is thought to be able to pursue universal knowledge unencumbered by the limitations of his body and irrational Woman who is thought to be closely bound to the particular instincts, emotions, and desires of her body. Treating the body as Other in geography serves to marginalize certain individuals and groups, that is, those groups who are thought to be 'tied to their bodies' and, therefore, incapable of reason. Included here, for example, are people who are constructed as physically weak, disabled, ugly, frail, diseased, queer, criminal, degenerate, elderly, black, poor, working class, unemployed, and so on. In other words, almost anyone except young, heterosexual, white, bourgeois, able bodied men are objectified and constructed as belonging on the side of the body as opposed to the mind.

The Body as Other

The privileging of modern reason, rationality, transcendent visions, neutrality, and disembodiment within geography has led to the exclusion and devaluation of women. While it is perhaps questionable to single out a category such as women (since categories such as women, homosexual, elderly, and so on are not mutually exclusive and gender is not necessarily the most salient base of oppression), it has nevertheless been a useful category for prompting a critique of geographical knowledge as disembodied. A geographer who has been at the forefront of mounting these arguments is Gillian Rose. Rose in her book *Feminism and Geography* (1993) explains that masculinist rationality is a form of knowledge which assumes a knower who believes he can separate himself from his body, feelings, values, etc. Rose explains that geography has a history of men attempting to be 'detached explorers' who produce 'transcendent visions'. There has been a privileging of some knowledges over others – the rational, reasonable, public, cultural, productive, and masculine (all that is associated with the mind) over the irrational, unreasonable, private, natural, reproductive, and feminine (all that is associated with the body). The subject who produces knowledge has been bracketed off from the knowledge he or she has produced.

All this is not to say that the body has been entirely absent in the discipline of geography – it has been bracketed off and treated as Other but not entirely absent. Some have described the body in the social sciences (and I would include here geography) as an 'absent presence'. When examining examples of geographies typical of the 1960s, 1970s, and 1980s, it is possible to detect traces of the body. Geographers in the 1990s did not suddenly recover bodies that were previously missing, but rather they shifted bodies from being in the wings (from being Other) to being center stage.

The reason why the body has functioned as Other in geography is because in order to establish rationality, there must be a contrast with the irrational. The exact ways in which bodies have functioned as Other in geographical discourse depends on the particular school of thought under consideration. Therefore, in considering the historical privileging of the conceptual over the corporeal (or mind over body) in the production of geographical knowledge it is necessary to examine more critically exactly which geographical knowledges are to be the subject of critique. To group together, for example, quantitative geography of the 1960s, humanistic geography of the 1970s, Marxist/structuralist approaches to geography in the 1980s, and post structuralist feminist and/or 'new' cultural geographies of the 1990s and 2000s, is not particularly useful for understanding some of the ways in which the body has been Othered in the discipline. Even within these broad subject areas there are likely to be large differences (including national, regional, and individual differences) in the research carried out. It is necessary then to disaggregate the category 'geography' into some of its constituent parts. In an attempt to tease out the idea of the body as Other in geographical discourse it is possible to examine any of the subdisciplinary areas within geography but the three examples discussed below are time geography, humanistic geography, and medical geography.

First, time geography is a perspective on the temporospatial structuring of social life developed by Törsten Hägerstrand. Time geography allowed geographers to focus on the everyday and the ordinary and was adopted by some feminist geographers to examine some of the spaces women tend to occupy. Although bodies 'seem' central to time geography it has been argued by authors such as Gillian Rose (1993) that they are not. The routine actions of individual human agents in time and space are represented by paths – paths that bodies follow. The body itself virtually disappears and in fact becomes its path. It is the 'paths' that define Hägerstrand's subjects, not their bodies. Hägerstrand does not talk about bodies in relation to things such as sickness, violence, and passion. Time geography can be deconstructed in a way that reveals that although the body seems present, in fact, it functions in texts as Other to the mind, the rational, the orderly, and the dispassionate. It is worth noting though that there have been some interesting recent attempts by geographers such as Mei Po Kwan to recorporealize time geographic constructs (such as time–space paths) through post structuralist feminist perspectives.

Second, humanistic geographers, especially in the late 1970s, attempted to understand the ways in which places were perceived by people, arguing that it was impossible to make sense of the social world without paying attention to the interpretations of those who lived in it. Humanistic geographers refuted ideas of objectivity and scientific rationality which characterized positivism. In this way, they placed people center stage. Drawing on phenomenology, they focused not on reason but on human creativity, not on the universal but on the specifics of people's lives. Drawing on phenomenology, humanistic geographers retrieved the body into their studies eliciting dimensions of people's lifeworlds.

Nevertheless, while corporeality played an important role in humanistic geography the body still tended to function as an Other in a number of ways. For example, humanistic geographers failed to acknowledge the 'specificity' of bodies such as observable differences in the rather ordinary ways in which women and men typically perceive, access, move in, use, and experience environments. Also, in the absence of any particular body being specified, a white, masculine, self contained body tended to be presumed. Bodies tended to be treated as though they were complete, autonomous, transcendent, and almost incorporeal. They did not want a body that is messy, incomplete, out of place, and not possessing clear boundaries, for example, the menstruating, pregnant, or birthing body – that which is feminine. In this way, it could be argued that although humanistic geography includes the body in its discourse its masculinism remains intact. The use of Man in humanistic geography makes men the baseline against which Woman's embodied difference is spoken.

Third, one might expect that some consideration has been paid to the body in medical geography; however, this is not the case. Many medical geographers have in the past, and still today, rely heavily on discourses of science in order to map the medical distribution of medical phenomena by counting the number of bodies with a particular disease (note: the more recent critical and post structuralist inspired work on health and/or disability is 'not' being referred to here). These scientific representations serve to textually and socially distance the work from flesh and blood bodies, reducing bodies to mere vectors or points on maps. In this way, the body, while initially appearing present in medical geographies, is actually Othered. Some have argued that medical geographers need to extend their current research agendas by drawing on the possibilities proposed by recent advances in feminist, social, and cultural theory on embodiment.

Embodying Knowledge as a Methodology

Privileging the mind over the body is one of the reasons why, despite critical interventions into the discipline over the past two decades, much contemporary geography continues to be a largely disembodied discourse. The discourses of geography in various and complex ways assert a division between the abstract Self and the embodied Other. In 'reality', we all 'have bodies' (are bodies) but the difference lies in that some are thought to be able to pursue and speak universal knowledge, unencumbered by the limitations of a body placed in a particular time and space/place, whereas Others are thought to be closely bound to their fleshy, located bodies. It has now become apparent to many critical geographers that knowledge is not omnipresent but embedded in specific social and spatial contexts. Knowledge is embodied.

In relation to methodology, geographers have attempted to 'position' themselves in their research usually by acknowledging their gender, ethnicity, or age. But is this enough? Reflecting on our own embodied positions as researchers in order to gain new insights into research practice and knowledge production may be a useful goal but one that in practice is difficult to achieve. How possible is it to fully understand ourselves as researchers, the researched, and the research context? It is important that researchers consider exactly what kind of reflexivity is being recommended and what other kinds of reflex ivity, if any, might be 'better'.

Although it is now permissible, even desirable, within feminist and other critical discourses to acknowledge something of ourselves and our political locatedness in the production of our texts it is still only permissible to acknowledge and reflect on some things about our cor poreality – things we are supposed to unproblematically know and understand such as our gender, ethnicity, and/or age. Many things remain off limits – too private, too emotional, too 'inappropriate', to include. Other (more messy) traces of our embodiment such as our size, food preference, and bodily habits tend to remain absent. These things are significant to the way in which we understand spaces but are not commonly voiced in geographical texts or forums. Some might think that these things belong in the category of the everyday, prosaic, and banal. Perhaps they do. But banal does not necessarily imply that which is intellectually un interesting and unimportant. The banal ought not to escape attention or be sidelined as domestic, feminine, and Other. Sometimes engaging in different kinds of positioning can convey an embodied subjectivity that is shifting.

To date then, it seems there has been a sustained engagement with the bodies of research subjects but less engagement with the bodies of researchers. Although it needs to be noted, some researchers have put their bodies firmly in the research frame by using autobiography. Pamela Moss suggests there are many benefits to 'writing one's life' geographically. There is, however, still a way to go toward embodying knowledge in ways that acknow ledge how researchers learn through their bodies. Un doubtedly, over the past decade geographers have begun to pay more attention to bodies – the bodies of those being researched and to a lesser degree researchers' – and the way in which knowledge is embodied but this line of enquiry is still to have a major impact on the way geographers tend to think about qualitative research practice.

Conclusion

All geographical knowledge is embodied, however, by the time research makes its way into scholarly books and journals all or at least most traces of research participants' and researchers' bodies tend to have dis appeared. Embodied subjectivities (the emotions and the sweat) become transformed into dispassionate, ob jective, rational prose. Knowledges are always situated. They are produced. They come from somewhere. It makes a difference who is doing the research, who participates in the research, where the research is car ried out, and how and why it is being carried out. It also makes a difference how it is turned into a finished product and for whom. The dualism between mind and body still underpins the production of geographical knowledge and as a result the body, to some degree, still remains Othered in the discipline. Geographical knowledge, however, cannot be separated from the minds and bodies that create it. All knowledge, even that which feigns to be objective, is embodied. This is a notion that geographers are likely to continue working with for some years to come.

See also: Emotional Knowing; Feminist Methodologies; Performance, Research as; Performativity; Situated Knowledge, Reflexivity; Subjectivity.

Further Reading

Cook, I. *et al*. (2005). Positionality/situated knowledge. In Atkinson, D., Jackson, P., Sibley, D. & Washbourne, N. (eds.) *Cultural Geography. A Critical Dictionary of Key Concepts*, pp 16 26. London: I.B. Tauris.
Crang, M. (2003). Qualitative methods: Touchy, feely, look see? *Progress in Human Geography* 27(4), 494 504.
Dyck, I. (1993). Ethnography: A feminist research method? *Canadian Geographer* 37, 52 57.
England, K. V. L. (1994). Getting personal: Reflexivity, positionality, and feminist research. *Professional Geographer* 46, 80 89.
Grosz, E. (1994). *Volatile Bodies: Toward a Corporeal Feminism*. St Leonards, NSW: Allen and Unwin.
Jaggar, A. and Bordo, S. R. (eds.) (1989). *Gender/Body/Knowledge: Feminist Reconstructions of Being and Knowing*. New Brunswick, NJ: Rutgers University Press.
Kwan, M. P. (2002). Feminist visualization: Re envisioning GIS as a method in feminist geographic research. *Annals of the Association of American Geographers* 94(2), 645 661.
Longhurst, R. (1997). '(Dis)embodied geographies'. *Progress in Human Geography* 21(4), 486 501.
Moss, P. (2001). 'Writing one's life'. In Moss, P. (ed.) *Placing Autobiography in Geography*, pp 1 21. Syracuse, NY: Syracuse University Press.
Moss, P. (ed.) (2002). *Feminist Geography in Practice: Research and Methods*. Oxford: Blackwell.
Moss, P. (2005). A bodily notion of research: Power, difference, and specificity in feminist methodology. In Nelson, L. & Seager, J. (eds.) *A Companion to Feminist Geography*, pp 41 59. Malden, MA: Blackwell.
Nash, C. (2000). 'Performativity in practice: Some recent work in cultural geography'. *Progress in Human Geography* 24(4), 653 664.
Nast, H. and Pile, S. (eds.) (1998). *Places Through the Body*. London: Routledge.
Rose, G. (1993). *Feminism and Geography: The Limits of Geographical Knowledge*. Minneapolis, MN: University of Minnesota Press.
Rose, G. (1997). Situating knowledges: Positionality, reflexivities and other tactics. *Progress in Human Geography* 21(3), 305 320.
Shilling, C. (1993). *The Body and Social Theory*. London: Sage.

Relevant Websites

http://www.iep.utm.edu
Feminist Epistemology (The Internet Encyclopaedia of Philosophy).
http://www.plato.stanford.edu
Feminist Perspectives on the Self (Stanford Encyclopaedia of Philosophy).
http://www.iep.utm.edu
Foucault and Feminism (The Internet Encyclopaedia of Philosophy).
http://www.thegreenfuse.org
The Green Fuse: Notions of Embodied Knowing.
http://www.teak.fi
The Theatre Academy: Challenging the Notion of Knowledge.

Emigration

B. S. A. Yeoh, National University of Singapore, Singapore

Glossary

Yimin Migrant (Chinese).

A Brief History of Emigration

Usually defined as the departure of persons from their native land to settle permanently in another country, emigration has been part of human experience since ancient times, although it is only with improvements in transportation systems in the sixteenth and seventeenth centuries that the volume of emigration flows became significant. Historically, emigration took different forms and had different demographic effects: for example, the movement of intact families to less populous regions transplanted whole communities and cultures (e.g., the English migration of members of the Puritan religious community to New England in the seventeenth century, or Jewish emigration to escape pogroms in the late nineteenth and early twentieth centuries), while that of young, predominantly male, adults resulted in cultural mixing and the formation of creolized families (e.g., the migration of European men to Spanish and Portuguese America; Japanese men leaving to work in the sugar plantations of Hawaii in the nineteenth and early twentieth centuries; Chinese men as coolie labor to Southeast Asia; and Indian men as indentured workers to Mauritius, South Africa, Malaya, Fiji, and the Caribbean).

The more spectacular emigrations in recent history in terms of geographical and cultural scope have involved the transcontinental movements of people from the Old World to the New. One of the best studied emigration streams in this respect was the flight of a million Irish on so called 'coffin ships' mainly across the Atlantic to the United States during the 1845–52 'potato blight' which also killed another million or more. Not all European emigration stemmed from cataclysmic events of course. Between 1850 and 1914, more than 40 million persons of different occupations and social class left their European birthplaces for a variety of reasons – one tenth of the mean total population of 400 million and perhaps one fourth of the labor force. Large numbers, for example, moved from Britain across the Atlantic to the United States, as well as to settler colonies of the British Empire, including Canada, Australia, and New Zealand. The vast majority of the emigrants paid their own fares, although a small proportion took advantage of government sponsored 'assisted' emigration schemes, including assisted passages under the Empire Settlement Act put in place to divert British and Irish emigrants from the United States to Australia and New Zealand with the aim of populating the white dominions with British stock, as well as the Sao Paulo scheme aimed at recruiting workers – mainly Italians – for the coffee plantations in Brazil. It should be remembered, however, that until the 1830s, the enforced emigration of blacks to the Americas (8 million before 1820) was numerically more important than white emigration (2.5 million before the same date).

Empire building in the nineteenth and early twentieth centuries in the hands of European powers was characterized by the expansion of control over trade and territory in non European parts, which in turn generated a large demand for labor met mainly through the migration of non Europeans. During this period, more than 50 million Chinese migrants and 30 million Indian migrants left their homelands, many to seek work opportunities in European controlled territories. It should be noted that among the Chinese in particular, leaving China was considered a form of 'sojourning', a concept that signified a temporary absence from home which would come to an end with eventual return. Despite the fact that some never returned, in the official Chinese view, there was no concept of permanent emigration; instead for all intents and purposes, those who leave China were, as historian Wang Gungwu noted, "either sojourners who had gone out to trade and would eventually return or defiant lawbreakers who dared not return for fear of severe punishment." In fact, the word *yimin*, the Chinese term for 'migrant', implies 'movement enforced by officials away from one's own land or village'; traditionally, migration to the Chinese is never a voluntary act but rather 'a great evil or calamity to be avoided'.

Mass European emigration reached record levels just before World War I and declined thereafter as a result of the introduction of quotas and restrictions on both emigration and immigration prompted by increasing economic and political nationalism. The rise of totalitarianism in parts of Europe also led to further falls in emigration rates. The intense period of worldwide migration during the age of empire further attenuated with the establishment of nation states in the post–World War II era as political control of the nation state's borders and institutional barriers regulating entry were heightened. The task of national self definition that preoccupied new nation states in the postwar decades depended not only on strategies of inclusion to strengthen group cohesion and belonging to the nation, but also entailed exclusionary strategies which dis tinguished 'self' from 'Other'. Not only were immigrants often thought of as 'out of place' in transgressing the national geobody, emigrants who unmoored themselves from their 'birth' nation states were also considered 'out of place' in betraying their homelands.

The closing years of the twentieth century and the dawn of the twenty first ushered in an increasingly globalized age of quickened mobilities of people, ideas, and commodities as well as a heightened sense of inter connectivity among places. The number of international migrants reached an estimated 175 million in 2000 (i.e., one out of every 35 people in the world is an inter national migrant living in a country other than that in which they were born), having risen by 18 million in 1970–80, 27 million in 1980–90, and 21 million in 1990–2000. Migratory moves are increasingly taking on a transnational character where movement is no longer based on permanent uprooting and settlement, but often multidirectional, complex, provisional, and based on a multiplicity of interconnections and networks sustained between 'home' and 'host'. In this context, the original notion of emigration as a form of permanent departure from a home nation (usually paired with the notion of immigration as a form of full incorporation and per manent settlement) needs to be rethought using frame works of transnationalism that accord more provisionality and flexibility to people movements. As Toro Morn and Alicea noted, "the demarcation line between 'sending' and 'receiving' countries is no longer a clear cut line between core and periphery."

Emigration Dynamics

Emigration is a multifaceted social, economic, and pol itical phenomenon. The current 'emigration climate' is partly shaped by the migration policies of sending and receiving countries, while postcolonial migration path ways often reflect historical links based on ties with former colonial metropolises, military occupation, as well as trade and investment flows. Among African countries, for example, Zairians emigrate to Belgium, Senegalese to France, and Nigerians to the United Kingdom. However, historical and geographical relationships are no longer the primary drivers of migration; instead, there are changing poles of attraction which add considerable complexity to the current patterns of movement.

Globalization dynamics have widened the gap be tween the developed and developing world through the increasing concentration of capital encouraged by foreign direct investment, and this has in turn created new conditions for mobilization that has spurred un precedented population movements. For example, Mexican emigration, one of the largest in volume in the contemporary world, cannot be decoupled from Mexico's insertion into the world economy: on the one hand, high levels of unemployment, low wages, deterioration of agricultural resources, and overall economic instability in Mexico impel movement out of the country, whereas, on the otherhand, greater income inequality, demand for cheap labor, and the territorial and economic reorgan ization of the US economy combined with long standing social networks between communities on either side of the border, result in 'pull forces'. Among the under developed island states in the Caribbean, emigration pressures are clearly linked to structural economic fac tors such as debilitated institutions, strong macro economic imbalances, dilapidated infrastructure, depleted productive assets, and pervasive poverty. How ever, emigration decisions (as in the case of the Anglo phone Caribbean) are not just tied to negative economic pressures but must also be understood in the context of historical linkages, social networks, and the degree of success in establishing transnational communities in host countries. Emigration hence does not signal permanent departure, but forms a part of circular migration based on the formation of highly interactive transnational house holds linked by to and fro movements of people and circuits of remittances, commodities, care, and emotions.

Within the New World Order, emigration in the form of labor migration is undoubtedly the dominant trend in current international migration (illustrated for Asia in **Table 1**). In South Asia, for example, international population movements are numerically dominated by flows of documented and undocumented workers, mainly unskilled young male workers (except in the case of

Table 1 Average annual number of migrants leaving for employment abroad in Asia: 1990 2003

Country	*Number of migrant workers departing (in thousands)*		
	1990 94	*1995 99*	*2000 03*
Bangladesh	174	263	-
China	75[a]	375[b]	-
India	377	360	297
Indonesia	118	328	387
Pakistan	145	118	130
Philippines	489	746	867
Sri Lanka	52	165	195
Thailand	87	193	165
Vietnam	-	13	47
Total	1517	2561	2088

[a] Average for 1990 91 includes workers abroad both under international labour contracts and under project contracts.
[b] Average for 1998 99 includes workers abroad both under international labour contracts and under project contracts.
Source: Adapted from UN Habitat (2006). *Additional Migration Statistics.* http://www.unhabitat.org/downloads/docs/Media/WHD%20 Additional%20Statistics.pdf, 3 (accessed 2 May 2008).

Sri Lanka where over half of emigrant workers are fe male) seeking work in the Middle East, Europe, as well as more developed Asian countries such as Japan and Ma laysia. The importance of the Middle East to Indian workers led the Indian government in 1983 to pass a new Emigration Act which sought to protect the welfare and interest of Indians abroad.

New 'feminized' streams of labor migration have also evolved, in response to changing production and repro duction processes worldwide predicated on a gendered division of labor. Production activities relocated from core economies to the periphery to take advantage of cheaper input costs draw on preexisting gender relations and target cheap and flexible female workers – many from countries with declining employment opportunities – to work in the 'global assembly line' in export pro cessing zones and industrial parks in rapidly industrial izing countries. Women from underdeveloped countries in Asia, Africa, Latin America, and the Caribbean have also left 'home' for overseas jobs as paid domestic workers as a result of the intensification of the 'care deficit' in advanced industrialized countries where reproductive labor is being shifted from the household to the market. The feminization of labor migration has also taken the form of more women migrating to work as 'entertainers' or sex workers in response to the expansion of hospitality and sexual services as male executives and entrepreneurs become more mobile.

While the widening economic and demographic div ide between underdeveloped and developed countries is an important factor in driving emigration dynamics, "there is no cut and dry relationship between poverty, demography and emigration," as noted by Trebilcock and Sudak. Indeed, the impact of war, political oppression, and environmental degradation in many countries often in association with worsening economic conditions are also major contributors. In the underdeveloped econ omies of Africa, for example, a potent cocktail of negative economic growth rates, military or authoritarian rule with poor human rights records, interethnic strife and conflict, the effects of Structural Adjustment Plans on wage freeze and inflation, and the deterioration of the living environment has spurred emigration. Emigration streams are highly complex: they range from cross border or intraregional movements to departures for inter national destinations, and may involve labor seeking migrants using migration as a household strategy to combat poverty and landlessness, whole families escaping political or religious strife and deteriorating living con ditions, or skilled workers seeking professional advance ment. Similarly, among several countries of Central America, emigration is linked not only to economic crises and structural adjustments but also to the dwindling capacity of the state, social, and armed conflict, and human rights violations.

Apart from macro structural factors, emigration de cisions are often dependent on negotiations within the family and community, as well as social networks the aspiring emigrant has with those living in potential host countries. In historical and contemporary times, 'chain migration' where emigrants tend to follow similar path ways and head for similar destinations as family and community members who have emigrated earlier pro vides a mechanism for minimizing risk and uncertainty, and maximizing information and assistance. Once a sys tem of social networks is put in place, migration takes on a self sustaining and self perpetuating character. With the passage of time, emigration generates its own *raison d'etre* and also habituates people into leaving. Sustained emigration flows breed a culture of migration charac terized by a general acceptance of migration as part of the life course and an increasing dependency of the com munity on outside resources and employment. In the Philippines, for example, after 30 years of labor migra tion, emigration as a means of seeking work has become, as Maruja Asis put it, "knitted into the survival or mo bility strategies of many individuals and families, [as well as] a common pathway to achieve the good life."

Changing Discourses on Emigration

The *Oxford English Dictionary* notes that early usage of the term 'emigration' may also apply to the departure of the soul from the body, either as a result of death or ecstatic rapture. Echoing this duality of meanings, official dis courses about emigration today span those carrying connotations of flight, disloyalty, and exile, on the one hand, to those which revalorizes emigration as a modern and even patriotic act on the other. Within academic and popular discourses, a similar variety of representations persists.

In their broad survey of emigration dynamics through recent history, Lowenthal and Comitas identified differ ent stances that home societies may take toward emi gration: those which do not countenance emigration at all (people who leave do so against the will of the com munity, maintain no contacts, and return only at their peril); those which encourage emigration but do not expect aid from emigrants (e.g., prosperous economies such as England in the context of seventeenth and eighteenth century mercantilism); those where emigrants are expected to contribute support but are discouraged from coming home (as in the case of the Irish who viewed emigration as a 'safety valve'); and those where indi viduals are encouraged to leave, to send remittances while away, and eventually return home (see later). These 'social imperatives' of home society affect "the emigrant's attitude toward his [*sic*] homeland, the duration or permanence of his stay abroad, and the speed and ease with which he assimilates to his adopted land."

At the turn of the new millennium, discourses on emigration among home countries in the developing world have generally taken on more positive connotations. While individuals who leave the People's Republic of China were once regarded as 'traitors' within state narratives, such mobility is now encouraged for certain groups as a strategy to tap business potential inherent in the development of diasporic networks. In the case of overseas students, the key policy is encapsulated in the 'Twelve words Approach', '*zhichi liuxue, guli huiguo, laiqu ziyou*' (meaning 'support study overseas, encourage returns, guarantee freedom of (international) movement'). In the case of the Philippine state, turning citizens into labor commodities in the global marketplace by institutionalizing employment abroad through the Philippine Overseas Employment Program has become a major strategy to address poverty and unemployment, and to siphon in remittances vital to the Philippine economy. With over 800 000 Filipinos annually deployed to take up work in foreign countries and roughly 10% of its population abroad, the Philippines has positioned itself within the new international division of labor as a major source country of workers, supplying workers of varying skill levels to over 100 countries. This is accompanied by a consistent state discourse where all the presidents from Marcos to Macapagal Arroyo have lauded overseas contract workers as the nation's 'new heroes'. At the same time, imaginaries of the 'nation' – i.e., the 'transterritorial nation state' in Luis Guarnizo and Michael Smith's words – based on a shared ethnicity continue to be drawn upon to protect the integrity of the nation and sustain its longevity. There is also the assumption here, as Filomeno Aguilar put it, that "at their destinations, emigrants will be transnationalists" who would engage in transnational circuits – both emotional and material – to contribute to the development of their 'homelands'.

Among host countries, however, emigration is viewed in far more ambivalent terms. In the early 1990s, a project on Emigration Dynamics in Developing Countries funded by the International Organization for Migration and the United Nations Population Fund was initiated to address 'widespread concern' about "the increasing numbers of refugees, irregular migrants and asylum seekers" from the South "grasping the emigration option as a last chance for survival and moving, uninvited, to … developed countries."

Consequences of Emigration

While the demographic profile and socioeconomic sustainability of many countries in Europe were vitally affected by transatlantic emigration during the past four centuries, emigration at a level high and sustained enough to lead to depopulation of an entire country is not common. The latter is best exemplified by the nineteenth century exodus of the Irish to North America, which not only seriously depleted the population of the sending country (where the demographic effects lasted for more than a century), but also left behind a country made up of 'residual people' whose young and more productive members had left permanently. Without its most creative and energetic members, Irish society stagnated throughout much of the twentieth century, and only experienced economic, social, and cultural revival in the last two decades of the twentieth century. Irish emigration, however, represents one extreme in terms of its demographic consequences on home society. Under different circumstances, emigration has been used by governments as a deliberate means to relieve population pressure on national resources. In the case of Japan in the 1920s, for example, emigration to South America, and especially Brazil, was encouraged as part of national policy in order to address the twin problems of population growth and economic instability. Today, while many developing countries have been consistently losing population as a result of emigration, the impact on population growth is usually small given the high natural increase rates in these countries.

In weighing the economic costs and benefits of emigration to sending countries in the contemporary world, Trebilcock and Sudak concluded in tentative terms:

> Although emigration poses risks of fiscal loss and human capital depletion, the empirical basis for these fears is unclear given possible offsetting factors such as remittances, return migration, and skills transfer, initial incentives for education, and dynamic investment effects of citizens living abroad…. Emigrants are properly viewed not as a stink cost, but as a resource to be cultivated (Trebilcock and Sudak, 2006).

Among the 'emigration threats' feared, 'brain drain' or 'human capital flight' and its impact on developing countries has been an issue of widespread concern given its scale. As a rough measure, it has been noted, for example, that between a third and half of the developing world's science and technology personnel now live in the developed world. On the one hand, the depletion of skilled, educated, and talented individuals may adversely affect the development of local knowledge and science, retard economic growth, erode tax revenues, and cripple public institutions and service delivery. In small developing countries with high rates of emigration such as the poorer Caribbean island states, the sheer volume of emigration often means that any possible positive effects may be outweighed by negative impacts on economic dynamism, the delivery of key public services, and the depletion of the political classes. It has also been argued, for example, that the hemorrhagic loss of health professionals from African countries has incapacitated health sector reforms and perpetuated a vicious cycle of ill health and poverty.

On the other hand, these negative views have been challenged by those who argue that the emigration of talent is not so much a 'drain' but part of 'brain circulation', and that positive effects for the source countries can include remittance flows, mobilization of foreign capital, exposure to new ideas and technologies, and the creation of diaspora networking that facilitates the exploitation of unique trade and business opportunities. Such an approach to emigration effects is predicated on the presence of transnational linkages among emigrants and their homelands, including the possibility of return. Strong transnational connectivities appear to hold in a range of migration scenarios. For example, the massive out migration from Latin America and the Caribbean to the United States, Canada, and Europe has been characterized by the transnational ties that immigrants maintain across borders between their home and host communities, and that incorporate return and counter flows. More specific examples include the formation of the Hsinchu Science based Industrial Park in Taiwan in the 1980s and 1990s benefited from the return of Taiwanese entrepreneurs and engineers from Silicon Valley while the growth of India's Bangalore and Hyderabad software development and outsourcing centers is tied to business networks facilitated by Indian engineers working in Silicon Valley.

An issue central to understanding the developmental consequences of emigration concerns the complex question as to whether remittance gains might sufficiently spur development to offset human capital loss. Notwithstanding the huge volume of remittances flowing back to source countries of the developing world (US$51 billion out of a total remittance amount of US$63 billion was estimated to have been sent back to developing countries in 2000), the verdict with regard to the development impact of remittances is still an open one. While remittances invested in land purchase, business and technology, or in improving health, education, and infrastructure in source countries may have significant multiplier effects, those used to finance consumption purchases may not generate longer term benefits and instead increase consumerism (e.g., of imported goods) and engender a culture of dependency.

Turning finally to the social consequences of contemporary emigration on source areas in the developing world, scholars have stressed the need to understand the processes operating at the nexus between emigration and the 'left behind' population, and to avoid treating the latter as passive recipients of migration effects. Out migration from rural Southeast Asia, for example, has become so widespread that it has a 'demonstration and emulation' effect on aspirations and achievement throughout communities, making it misleading to assume that the left behind would always hold on to 'traditional values' or remain static in their worldviews. In other instances, far from being deserted by migrating children, left behind parents actively participate in providing financial and emotional support for their children's mobility, leading to the emergence of a translocal extended family linked by multistranded, fluid relations between members across geographical space. Similarly, studies on migrating mothers and their left behind children highlight the (re)enactment of episodes of family interaction through distanced communication as the family takes on transnational dimensions. New work has also called attention to the gender politics implicit in a range of mutually constitutive interactions between migrants and the left behind: in areas such as rural Indonesia and Vietnam where feminized out migration streams predominate, established gender ideologies may either be challenged by changing social practices where women assume breadwinner roles, or continue to regulate traditionally scripted roles for men and women but in new ways.

Conclusion

In defining emigration as 'don't look back exit for a new existence elsewhere' and the 'converse and historical complement of diaspora', Roger Sanjek is able to point to several 'unambiguous instances of historical emigrations', but only with the benefit of hindsight as human intentions at the point of exiting may be unclear or may change. With the proliferation of transnational connectivities, multidirectional mobilities and hybridized identifications in a globalized new order, emigration as a unidirectional people movement of nonreturn is likely to become even more provisional and ambivalent. Instead, emigration as a form of 'home leaving' and 'border crossing' may become a recurring moment, held in tension with 'transiting', 'sojourning', 'settling', 'rooting', 'returning', 'homecoming', 'inclusions/exclusions', and other multiple experiences of migrations.

See also: Immigration I; Immigration II; Migrant Workers; Migration; Remittances; Transnationalism.

Further Reading

Aguilar, F. V. (2004). Is there a transnation? Migrancy and the national homeland among overseas Filipinos. In Yeoh, B. S. A. & Willis, K. (eds.) *State/Nation/Transnation*, pp 93 119. London: Routledge.

Appleyard, R. (1998). *Emigration Dynamics in Developing Countries*, (Vols 1 4). Aldershot: Ashgate.

Asis, M. B. (2005). Caring for the world: Filipino domestic workers gone global. In Huang, S., Yeoh, B. S. A. & Abdul Rahman, N. (eds.) *Asian Women as Transnational Domestic Workers*, pp 21 53. Singapore: Marshall Cavendish.
Baines, D. (1995). *Emigration from Europe 1815 1930*. Cambridge: Cambridge University Press.
Guarnizo, L. and Smith, M. (1998). The locations of transnationalism. In Smith, M. & Guarnizo, L. (eds.) *Transnationalism from Below*, pp 3 34. New Brunswick: Transaction Publishers.
International Organization for Migration (IOM) (2005). *World Migration 2005: Costs and Benefits of International Migration*. Geneva: International Organization for Migration.
Lowenthal, D. and Comitas, L. (1962). Emigration and depopulation: Some neglected aspects of population geography. *Geographical Review* 52(2), 195 210.
Manning, P. (2005). *Migration in World History*. New York: Routledge.
Nguyen, L., Yeoh, B. S. A. and Toyota, M. (eds.) (2006). Special section on migration and the well being of the 'Left behind' in Asia. *Asian Population Studies* 2(1).
Nyiri, P. (2004). Expatriating is patriotic? The discourse on 'new migrants' in the People's Republic of China and the identity construction among recent migrants from the PRC. In Yeoh, B. S. A. & Willis, K. (eds.) *State/Nation/Transnation*, pp 120 143. London: Routledge.
Sanjek, R. (2003). Rethinking migration, ancient to future. *Global Networks* 3(3), 315 336.
Toro Morn, M. I. and Alicea, M. (eds.) (2004). *Migration and Immigration: A Global View*. Westport, CT: Greenwood Press.
Toyota, M., Yeoh, B. S. A. and Nguyen, L. (2007). Special issue on migration and the "Left Behind" in Asia. *Population, Space and Place* 13(3).
Treilcock, M. J. and Sudak, M. (2006). The political economy of emigration and immigration. *New York University Law Review* 81(1), 234 293.
UN Habitat (2006). *Additional Migration Statistics.* http://www.unhabitat.org/downloads/docs/Media/WHD%20Additional%20Statistics.pdf, 3 (accessed 2 May 2008).
Wang, G. (2003). *Don't Leave Home: Migration and the Chinese*. Singapore: Eastern Universities Press.
Xiang, B. (2003). Emigration from China: A sending country perspective. *International Migration* 41(3), 21 48.

Emotional Geographies

J. Davidson and M. Smith, Queen's University, Kingston, ON, Canada

Glossary

Cartesian Dualism A theory that rigidly distinguishes between mind and matter as two distinct substances. This understanding has been extended and associated with other binary distinctions between concepts such as reason/emotion and male/female, where the former is culturally privileged over the latter.

Existentialism A philosophy of existence or being concerned with choice, freedom, and authenticity that holds individuals to be without predetermined essence. Rather, persons are responsible for determining their own lives through acts of individual will.

Phenomenology A philosophy that investigates the experience and significance of phenomena, fundamental elements of perception as disclosed to the human subject. To facilitate purely intuitive analysis of phenomena, phenomenology holds that all naturalistic presuppositions and prior knowledge should be set aside or 'bracketed off'; this suspension of judgement is termed the *epochē*.

Positivism An influential philosophy of science that holds empirical and replicable approaches to the observable natural world to be separate from social context and so capable of producing knowledge that is neutral and value-free. (The approach has been subject to extensive critique since the mid-twentieth century, and its appeal to universal foundations for knowledge has been undermined by critical philosophies that privilege the production of situated knowledges, influenced by context.)

Subjectivity Often used in preference to identity for its less fixed and individualistic, but more fluid and inclusive connotations, subjectivity is concerned with the qualities of being a thinking-feeling self in context. The term is associated with approaches such as feminism that attend to individuals' complexly constituted subjective experiences of the world.

Introduction: Why on Earth Study Emotions?

Emotions are woven into the fabric of our everyday lives – as boredom at work, worry about finances, and amusement at a TV show – and yet they also serve to recompose and transcend these mundane aspects of our existences in intense experiences that mark unique moments, such as the uncontainable exuberance of falling head over heels in love. They are, paradoxically, both inordinately diffuse and all pervasive and yet also heart and gut wrenchingly present and personal; sometimes moods seem to envelop us from without, sometimes emotions seem to surge up from within and comprise the very core of our being here in the world, our human existence.

Emotional responses define places as much as people and yet they seem strangely absent from most geography texts, or at least their import is assumed in such a manner that they become part of the unnoticed background scenery rather than an intimate, indeed indispensable, component of every worldly performance. Even a theoretical term like, for instance, alienation, though definable in terms of modes of economic production, class, the division of labor, and human nature relies for both its descriptive potential and its political bite on the emotional resonances associated with the misery of feeling exploited, the loneliness of detachment from proper human companionship, pride in one's work, and the desire and hope to feel at home and live a fully human life in the face of a discouraging world.

The apparent absence of geographical attention to emotion therefore needs explanation, unless, of course, this lack is more apparent than real. Perhaps, if we reappraise past works in the light of our newly discovered interests we might find geography's hidden emotional treasures lying waiting beneath its seemingly indifferent surface. To some extent this may be so. After all, if emotions are all pervasive then traces of them should be revealed almost anywhere we care to look. But there again, like gemstones, their distribution will be uneven and require more (intellectual) effort to identify, extract, and refine in some instances than others. In some texts emotions are very deeply hidden indeed under tables of data and figures, maps, detached observations, third person or impersonal narratives, and so on. Hence, a geography that takes emotions as its subject matter would still need to reflect upon and interpret this unevenness, suggesting reasons why certain authors, geographic fields, historic periods, methods, and so on, express, or alternatively suppress, emotions in the many various ways they write (about) the world.

This suggests then that emotional geographies might, as it becomes an increasingly familiar term, serve to denote an emergent field of critical geography where a new focus of study also, quite rightly, throws the history and current state of the discipline into relief, illuminating what it excludes and overlooks. Just as feminist geographies continue to critique and recompose geographical accounts of a world where women have been noticeable largely by their absence, so emotional geographies too might be thought of as a recognition and response to a lack, that is, geography's failure to represent our emotional lives. Perhaps we could even understand this term – lack – in ways reminiscent of psychoanalyst Jacques Lacan's (1901–81) usage. If geography has defined its self identity as a discipline partly by excluding the emotional, as something it regards as 'other' than geography (e.g., as belonging to psychology), then the recent emotional turn could be understood as the exercising of a geographical imaginary, the acting out of a desire to make geography more complete by giving graphic (written and symbolic) form to what now, upon reflection, seems lacking in its relations to the world.

This, no doubt, would be a partial and controversial analysis but one that might at least spark further thought about what emotional geographies can and cannot do. For example, in Lacanian terms the human subject's experience of lack and subsequent desire for completion is ultimately unachievable precisely because this subject is originally defined over and against what it is not, what is 'other'. To change the relation to the 'other', to try to own it in some way, inevitably changes and redefines, rather than completes the self. It thereby also reveals to the human subject the dizzying absence of any essential fixed self identity, any final answer to the question of who they really are. In terms of our analogy, recognizing geography's emotional lack makes apparent the fluidity of what might previously have been thought to be hard and fast disciplinary boundaries and the absence of an essential center or core that identifies what geography, as an academic subject, is. To appropriate the words of geographer David N. Livingstone the "idea that there is some eternal metaphysical core to geography independent of historical circumstances will simply have to go."

There is an irony here then. Even if we accept that the motivation for the relatively recent emergence of emotional geographies arises from the recognition that there is something crucial missing in geography as it is currently instantiated, their development will not make geography whole, restore its integrity, or fill a gap. Rather they make this lack available for reflection, make it seem problematic, and on this basis facilitate a form of self critique – what is sometimes referred to in critical theory as a negative dialectics. This threatens to decompose as well as recompose its object, breaking down current disciplinary boundaries and challenging fixed ideas of geography. Some geographers, already feeling their discipline threatened, might not be happy with such a situation, they might try to deny that emotional geographies are real (proper) geography but this, after all, is partly the point. Emotional geographies do not just extend the limit of current geographical research, they recompose what geography might be in terms of emotionally, as well as sociohistorically, situated theories.

On this same analysis, we also need to bear in mind what emotional geographies cannot do. They cannot provide a complete representation of emotions. Why? Because our Lacanian analogy would also suggest another, parallel, kind of lack, a failure of representation that, Lacan argues, operates at the level of the symbolic order itself. Words (indeed all symbols) lack a fixed signifier – something solid and permanent that we might point to as anchoring their meaning or, reciprocally, that might be fully represented by them. Words' meanings, like subjects' identities, are only constituted over and against 'other' words, discourses, and practices. Words too are in flux; they allow us to reflect on the world, to situate our knowledge within various contexts but, they cannot capture the world's essence – if such a thing exists. Something of the world we try to describe always eludes those descriptions. Indeed, this seems paradigmatically the case with (words for) emotions which overlap and melt one into another, that allude to each other and their contexts in complex but often vague and mysterious ways, that operate in close association with other emotional states or exist through attempts to dissociate oneself from them – think of shouting loudly at someone – "I'm not angry!"

Emotions are then conceptually linked together only by our perception of (what the philosopher Ludwig Wittgenstein (1889–1951) originally referred to as) family resemblances, overlapping patterns of similarities and differences between them, and the contexts in which they are used, the situations that give them meaning. Emotions have no common defining feature, a point emphasized by Thomas Dixon's recent claim that "emotions did not even exist until two hundred years ago". That is to say that emotion only comes into use as an over arching category after about 1820, earlier texts tending to refer to the affections, passions, interests, appetites, and so on, each with their own associations. This certainly doesn't mean that previously people didn't feel desolate or happy but that the way particular emotions were understood (and perhaps even something of the ways they were experienced) varied accordingly. For example, the Scottish philosopher Adam Smith (1723–90), like many writing in the eighteenth century, focused attention on what were termed moral sentiments, feelings, like remorse, regarded as intimately associated with ethical responsiveness to 'others' and with judgments about right and wrong. How remorse is defined and how it is occasioned and experienced, perhaps even whether it is experienced at all, are dependent upon historically, culturally, and geographically, varying norms. So, although geographies can, quite literally, add new dimensions to the understanding of these patterns of emotional differences – the spaces between boredom and irritation, anger, rage, and jealousy – they reconstrue rather than complete the definition of them.

Emotional geographies are not then, on this understanding, indicative of a new science that will focus on how to map emotional landscapes, to provide a demographic of despair or cartography of embarrassment. Indeed it works against such attempts to pin emotions down, to define them in such ways that they might become mere objects of quantification, comparison, and manipulation. After all, if geography were to accept such a view it would merely have provided it with a new academic resource and have accepted the very understanding of geography as an entirely rational (in the sense of being emotion free) enterprise that it seeks to disrupt. As Wittgenstein himself pointed out, emotions cannot properly be understood by such mechanistic reductions. "'Grief' describes a pattern which occurs, with different variations, in the weave of our life. If a man's bodily expression of sorrow and joy alternated, say with the ticking of a clock, here we should not have the characteristic formation of the pattern of sorrow or of the pattern of joy".

This is not then to say that emotional geographies are not concerned with empirical matters, far from it. The phenomenal experiences of emotions are the starting points for most researchers, as are the ways emotions are embodied and serve to color our experiential world such that we interpret and value aspects of it in particular ways – we might see red watching Canada's mass slaughter of seal pups; feel dark despair at the emotional disconnection that those involved in such killing exhibit, their absence of compassion for 'other's' lives. But attending to the phenomenology of emotion also leads us away from attempts to reduce the importance of such feelings to simplistic lines of causation or mere numbers, such as the thousands killed, or the paltry few thousands of dollars they represent to the Canadian economy. This experiential aspect is precisely why an emotional geography of seal slaughter would be so very different from an economic geography of the same issue, or a geography framing the question in terms of resource distribution and allocation.

Perhaps we can now return to questions of how the suppression and repression of emotions in contemporary geography originated and how emotional geographies respond to this lack. It should at least be somewhat clearer why certain forms of critical geography, namely, feminist, nonrepresentational, psychoanalytic, and phenomenological, have been most closely associated with the emergence of emotional geographies. This is precisely because these approaches have been concerned with the critique of a world view that accepts the centrality of an essentially rational, unchanging, autonomous, and emotion free or emotionally controlled human subject, (a typically masculine ideal) who has the ability to fully represent the external world within a universally applicable, objective, and rationally determined symbolic order. This view has dominated geography as much as any discipline with scientific pretensions and is often thought of as being exemplified by, and as part of the legacy of, the first modern philosopher, Rene Descartes (1596–1650). (Although certain aspects of this view could, in fact, be traced back much further, e.g., to the Stoics who regarded emotions as clouding our rational abilities.)

Cartesian Coordinates and Emotionally Empty Space

Descartes' philosophy has a number of important geographical and emotional implications. First, in terms of coming to see the world materialistically, as composed of mere objects at the disposal of human subjects. Second, in terms of how subjects come to understand this world through abstract, rational representations of its mechanical properties. Objects come to be defined in terms of their physical extension in a supposedly empty space, a space that can be mapped and measured using mathematical coordinates thereby providing objective mental representations of how material things stand in relation to each other. Third, Cartesian dualism thus gives rise to a representational epistemology, an understanding of knowledge as a mapping or reflection of the physical world within an entirely separate world of mental constructs, concepts, symbols, and so on. As Jonathan Crary has remarked, Vermeer's painting, *The Geographer*, perfectly illustrates this (**Figure 1**). Here a paradigmatically masculine subject mentally surveys the map which represents the outside world from which he averts his eyes. These intellectually coordinated representations are all he needs to know to master that world. Note then, that for Descartes, unlike Lacan, there is no underlying reason why all that the world contains cannot be fully (completely) represented in every material aspect in the transparency of thought. Indeed with the benefit of hindsight Descartes might be regarded as a key source of the rather different view, often associated with scientific positivism, that all that really matters, all that there actually is, can, in theory, be so represented. From this perspective what cannot be measured, manipulated, predicted, or accurately defined by a concept does not exist in any meaningful sense.

So where does this leave the emotions? Interestingly, they seem to be suspended in a kind of no man's land; they inhabit an overlooked territory in between mind and matter, subject and object, resisting attempts to entirely appropriate them by one side or the other. They are not alone here since this interstitial space is also where the senses reside, both having about them a necessary

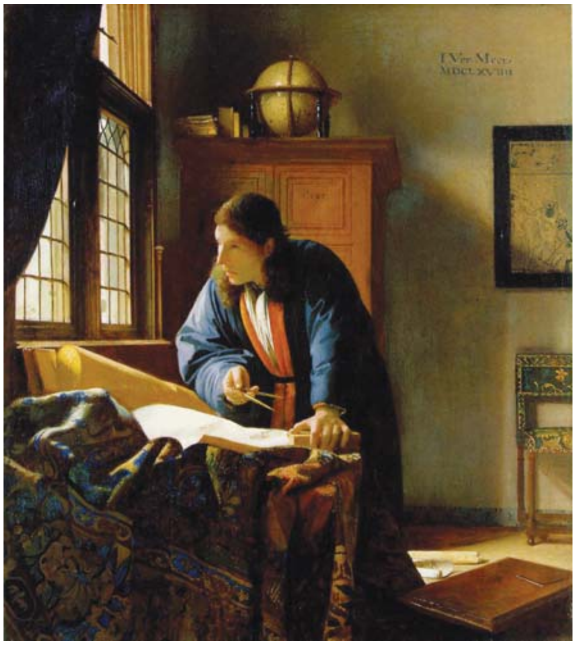

Figure 1 Vemeer's painting. *The Geographer*.

ambiguous status because, given the absolutism of Descartes' dualism, there should be nothing that is not either entirely mind or entirely matter, no in between. And yet without such a ground all connection between these two distinct realms is simply impossible – there can be no representation, no worldly knowledge. If this was so then what would it mean for one to touch something or be (emotionally) touched by what someone does; where exactly do these effects and affects take place? Such questions expose what is lacking in Descartes' conceptions of both his human subject and his theoretical edifice – a space for senses and emotions to operate. The rather bizarre location he does propose, the brain's pineal gland, is hardly convincing!

Although Descartes' substance dualism is largely jettisoned in later philosophical and scientific developments, the basic structure of his thought remains. Today's geographer may sit before a computer screen rather than Vemeer's nautical map but s/he is, if anything, even more dependent on the visualization and manipulation of mathematical representations and coordinates. The emotions find themselves further sidelined, as even the senses, required to play the role of a trustworthy intermediary running back and forth between thought (concept) and object, become cleansed of any remaining emotional encumbrances.

Sight, for example, is reduced to scientific observation, a mode of visual communication with the sole function of information retrieval. But to this extent geography becomes both detached from the real world, and dehumanized. There is no space here for, say, the emotional and ethical insight a loving gaze might offer into the lives of 'others'. Nor, for that matter, can an emotionally cleansed notion of the senses explain the feeling of anxiety or the uncanny (*unheimlich*) that an 'other's gaze can provoke in us, a look that according to existentialist philosopher Jean Paul Sartre (1905–80) makes us aware both of our own individual existence and the fragile, groundless, nature of that existence, the nothingness that "lies coiled in the heart of being – like a worm." Even sight then, the most paradigmatically detached, distanced, and objective of the senses, affects us intimately, magically expressing something of our mode of being in the world and providing insight into 'others'; how else could a stare be aggressive, a look malicious, a gaze lustful or leery, or a glance wistful, capricious, or yearning?

It is just these emotionally mediated senses and relations that compose the fabric of our human existence, that make our lives meaningful, or, in their absence, hopeless. Emotions are vital (living) aspects of who we are and of our situational engagement within the world; they compose, decompose, and recompose the geographies of our lives. And even Descartes recognized this to some extent. After all, in his last book, fittingly entitled *The Passions of the Soul*, he reflected upon what was lacking at the centre of his divisive philosophy, the emotional no man's land in and between the subjective and objective worlds he had broken apart. Here he recognizes that even his rationalistic philosophy is, after all, motivated by desire, a desire for knowledge and that this passion is, in its turn, sparked into being by another, by wonder. "Wonder", says Descartes, "is a sudden surprise of the soul causing it to consider with attention those objects which seem to it novel and unexpected." And while Descartes argues an excess of wonder is harmful and must be suppressed having no remedy "other than acquiring the requisite amount of knowledge," this is, nonetheless, an explicit admission that the "sweetest pleasures of this life" we feel only because we are not pure mind but hybrid beings, profoundly passionate creatures also composed of flesh and blood. Contra to Cartesian dualism, we are always already emotionally engaged beings in the world.

Current Trends in Emotional Geographies

Emotions then are not just a new topic for geographical study. They trouble and yet inspire its very existence just as they trouble and inspire individuals. Without passions, sentiments, and their affects there would be no human geography worth its name. Phenomenologist, feminist, psychoanalytic and nonrepresentational geographers have all recognized this and have consciously tried to situate themselves in this interstitial no man's land, dissolving the interrelated dualisms between thought and world, male and female, reason and emotion, and culture and nature.

Phenomenological approaches achieve this by focusing, as the term suggests, on what is experienced rather than on what those experiences might be supposed to represent (e.g., the feeling of 'therapeutic touch' recently discussed by Mark Paterson in the edited collection, *Emotional Geographies*). The phenomenological world is not the abstract empty space of Cartesian coordinates but the lived world perceived and produced through our emotionally laden activities. Distances, for example, would not be thought of as fixed measurements between objects but as differently experienced movements affected by myriad influences – the frightful depths of the night through which we walk, the wearisome weather we struggle against, and the lightening of our mood and steps in good company. This is not a form of subjectivism, but a recognition of the self's entanglement with the world since all such emotions are intentional, that is, they are inherently about something – the dark, the weather, and so on – even when their root causes are far from clear to us. Being alive we will feel, but never fully own, these emotions which come and go without our having called them into being. It is not surprising then that contemporary geographic phenomenologists, like Yi Fu Tuan and Edward Casey, have often focused on emotional responses and attachments to particular places. Places like people can thus be understood as being constituted within an emotionally charged middle ground, one neither entirely subjective nor objective – this has become increasingly obvious in recent phenomenological studies of our spatially mediated sense of self. As phenomenologist Maurice Merleau Ponty (1907–61) wrote: "Truth does not only inhabit 'the inner man', or more accurately there is no inner man, man is in the world, and only in the world does he know himself."

This quotation leads directly to the potential contribution of feminist geographies, since it is not only man who is in the world! Feminist geographers' interests often overlap with phenomenology but bring a critical awareness of the gendering of emotions (indeed of emotionality in general, so often dismissed as a feminine trait), such as the particular fears felt by women in nighttime urban environments, discussed in a recent work by Hille Koskela, or the gendered incidence of phobias, as revealed in studies by Ruth Bankey, Felicity Callard, and Joyce Davidson.

Feminist theory has been particularly critical of absolute binary distinctions and, though rarely, until recently, explicitly focused on emotion; feminist explorations of embodiment have proved fruitful for emotional geographies. They uncover feelings about experiences of, for example, sexuality, pregnancy, disability, chronic illness, and the consumption of foods and fashions. (Many of these topics are discussed in recent 'emotional geographies' themed issues of *Gender, Place, and Culture*, and *Social and Cultural Geography*.) In investigating these taken for granted emotional aspects of embodied experiences feminists illustrate the intimate connections between physical (material) and mental health. Rather than treating individuals in isolation, feminists also make explicit theoretical connections with broader social systems of politics and power as evidenced by Victoria Lawson's recent work on geographies of fear and hope. Feminist methodologies reflect these theoretical concerns emphasizing relational, reflexive, and intersubjective approaches that are typically qualitative, attending closely to the emotional relations that permeate geographical practices.

Psychoanalytic and psychotherapeutic perspectives offer geographers another field of emotionally relevant theories and practices that are capable of more than simply providing Lacanian analogies (though this hopefully illustrates something of their wider theoretical potential). Geographers like Liz Bondi and David Sibley have paid particular attention to the factors mitigating the creation, maintenance, or dissolution of the more or less stable and permeable psychosocial boundaries that are constitutive of both individual subjectivity and of social identities (the work of Hester Parr and Chris Philo is also instructive here). Emotions clearly play a pivotal role in these formative processes which impact both individual mental health and sociopolitical cohesion, shaping social relations through the mediation of the self's relations to those deemed 'other', drawing and redrawing boundaries between you and me, them and us via feelings of attraction, repulsion, and so on. Psychotherapeutic approaches too, although apparently focused on reaching and treating the contents of individual minds through discursive processes, concern themselves intimately with subjects' relations to other people and places, not least the therapist and therapeutic setting. A key aspect of these, as with psychoanalytic, approaches is the need to recognize the difficulties inherent in composing a necessarily incomplete emotional self understanding and of expressing this (representing it) in language. Composing and expressing oneself emotionally are both necessary and yet, ultimately endless (lifelong and incomplete) tasks.

If psychoanalytic and therapeutic approaches focus on the emotional lack at the heart of subjectivity then, crudely speaking, nonrepresentational theory (NRT) focuses on the necessary incompleteness (the lack) within the act of representation itself, that is, on that which eludes expression in language. NRT's preference for the term affect over emotion is, as Nigel Thrift suggests, linked to this concern for the ineffable; affect is supposedly indicative of those more immediate embodied engagements with the world beyond language. Hence NRT tends to focus on performance, on what people do rather than what they say they do, challenging the Cartesian priority accorded to thinking in and through words as represented in recent writings by Ben Anderson, JD Dewsbury, and Paul Harrison. This is clearly important and NRT offers some radically different insights into emotional flows and conjunctions although there may, ironically, be something of a performative contradiction in NRT's approach, insofar as its practitioners employ some of the densest and most abstract philosophical language (such as that associated with Gilles Deleuze) to theorize and interpret such performances. There is also an important question about how far even nonverbalized bodily movements, in say a ballet performance, are both implicitly and explicitly representational. The dance of the dying swan is actually overburdened with layers of cultural (e.g., gendered) meanings as well as being indicative of presentationally constrained and rehearsed emotional content – something true even of less overtly stylized manifestations. NRT is surely right in that not everything about an emotional performance can be captured in language; actions, like bursting into tears, often speak louder than words. However, the fact that words cannot completely represent emotions is not necessarily a problem for representation, it is, as Lacan might argue, of the nature of representation. And representation in some form or other seems to be an inescapable aspect both of a symbolic order that encompasses far more than words, and of our self understanding. It is not immediately clear that the anger expressed in giving another driver the finger is any less representational (or indebted to a masculine symbolic order) than giving vent to irritability through contextually apt sarcasm. Words too are always performances and all performances, however nonverbally immediate, are meaningless unless they comply with certain (constraining) prefigurations.

There are then important debates still emerging within and between these different approaches to emotional geographies despite their overlapping interests. These debates are an encouraging sign of the creative fluidity of an emergent field that will hopefully continue to resist the desire to present a single unifying theoretical totality, one that would belie the intricate complexities of its subject matter. (Such issues will be explored in interdisciplinary depth by the new journal, *Emotion, Space and Society*.) What makes emotional geographies so fascinating is their potential to critique and reconstrue almost everything that geography has so far taken for granted.

See also: Affect; Agoraphobia; Emotional Knowing; Existentialism/Existential Geography; Feminism/Feminist Geography; Other/Otherness; Positivism/Positivist Geography.

Further Reading

Anderson, B. (2006). Becoming and being hopeful: Towards a theory of affect. *Environment and Planning D: Society and Space* 24, 733 752.

Anderson, K. and Smith, S. (2001). Emotional geographies. *Transactions of the Institute of British Geographers* 26, 7 10.

Bendelow, G. and Williams, S. J. (eds.) (1998). *Emotions in Social Life: Critical Themes and Contemporary Issues*. London: Routledge.

Bondi, L. (2005). Making connections and thinking through emotions: Between geography and psychotherapy. *Transactions of the Institute of British Geographers* 30, 433 448.

Casey, E. (2001). Between geography and philosophy: What does it mean to be in the place world? *Annals of the Association of American Geographers* 91, 683 693.

Davidson, J. and Bondi, L. (2004). Spatialising affect; affecting space: An introduction. *Gender, Place and Culture* 11, 373 374.

Davidson, J., Bondi, L. and Smith, M. (eds.) (2005). *Emotional Geographies*. Aldershot: Ashgate.

Davidson, J. and Milligan, C. (2004). Embodying emotion, sensing space: Introducing emotional geographies. *Social and Cultural Geography* 5, 523 532.

Dewsbury, J. D. (2003). Witnessing space: Knowledge without contemplation. *Environment and Planning A* 35, 1907 1932.

Dixon, T. (2003). *From Passions to Emotions: The Creation of a Secular Psychological Category*. Cambridge: Cambridge University Press.

Harré, R. and Parrott, W. G. (1996). *The emotions: Social, Cultural and Biological Dimensions*. London: Sage.

Harrison, P. (2007). 'How shall i say it'? Relating the nonrelational. *Environment and Planning A* 39, 590 608.

Koskela, H. (2005). Urban space in plural: Elastic, tamed, suppressed. In Nelson, L. & Seager, J. (eds.) *A Companion to Feminist Geography*, pp 257 270. Oxford: Blackwell.

Lacan, J. (1992). *Écrits: A Selection*. London: Routledge.

Lawson, V. (2007). Introduction: Geographies of fear and hope. *Annals of the Association of American Geographers* 97, 335 337.

Lupton, D. (1998). *The Emotional Self*. London: Sage.

Parr, H. (2005). Emotional geographies. In Cloke, P., Crang, P. & Goodwin, M. (eds.) *Introducing Human Geography*, pp 472 485. London: Arnold.

Philo, C. and Parr, H. (2003). Introducing psychoanalytic geographies. *Social and Cultural Geography* 4, 283 293.

Thrift, N. (2004). Intensities of feeling: Towards a spatial politics of affect. *Geografiska Annaler* 86, 57 78.

Relevant Websites

http://www.csun.edu
American Sociological Association, Section on Emotions (Sociology of Emotions), California State University, Northridge.

http://www.elsevier.com
The journal of 'Emotion, Space and Society', Elsevier.

http://www.sseas.org
Society for the Study of Emotion, Affect and Space (SSEAS).

Emotional Knowing

L. Bondi, The University of Edinburgh, Edinburgh, UK

Glossary

Affect Feeling, disposition, or mood that exists prior to cognition or rational thought. Often referred to as a mental state, affect is also bodily or sensory and unconscious or nonconscious in character. To be affected is to feel or to be touched or moved.

Binary Distinctions Concepts that are defined by an either/or structure: one term is defined by the exclusion of the other in a context whether there are only two alternatives. Binary gender is an example: people are categorized as either female or male, with no overlap between the categories and no space outside of them.

Cognition Conscious thought, thought processes, knowledge, and apprehension. Cognition is generally linked to the exercise of mental faculties, such as intelligence reasoning, logic, memory, and perception.

Emotion Feeling or disturbance of feeling. Emotion is often understood as outward and expressive as well as embodied, and therefore as psychosocial (rather than biological or physiological) and as available to thought.

Feeling Bodily sensation or perception, especially through sense of touch. Feeling has also acquired the more specific meaning of bodily sensation associated with affect or emotion.

Post-Structuralist Theory A range of theoretical resources that developed out of critiques of structuralism. Common features include the idea that human subjects are not unified, but are always other to/than themselves; the idea that meanings are not fixed but are produced and negotiated in particular contexts; and an insistence on nondeterministic approaches to knowledge production, which shift the focus of attention away from questions about causes toward questions about how things happen.

Reflexivity Turning back on itself of the awareness of the subject. In relation to research methods, reflexivity refers to the analysis of how the researcher impacts on the research process, and how the subjective realities of researchers and research participants interact.

Subjectivity Qualities of human being that characterize human subjects' situated self experience. Subjectivity is often associated with feelings, perceptions, and beliefs of human individuals. However, a sense of individuality is itself an aspect of subjectivity and is not necessarily inevitable or universal.

Transference A concept used in psychoanalysis to describe how a person's patterns or habits of relating are transferred from one context (typically that of early life with parental figures) to others. Psychoanalysts use the idea of transference to describe key aspects of their experience of how their patients relate to them, which they theorize in terms of early relationships with caregivers.

Introduction

Human geographers have long been interested in peo ple's experiences of the places, spaces, landscapes, and environments in which they live, work, and move. These experiences include people's emotions, that is, their subjective feelings about, and in relation to, these places, spaces, landscapes, and environments. Moreover, emo tions do not necessarily belong uniquely to people: par ticular places, spaces, landscapes, and environments may themselves be redolent with emotions, and may act as powerful sources of affect, in the sense of affecting what is felt and what takes place.

Despite the self evident importance of emotions to the phenomena that human geographers study, until re cently, emotional dimensions of existence have rarely been mentioned or engaged with directly. Silence about emotions has generally been accompanied by apparently dispassionate ways of presenting knowledge about the world. This exclusion of emotions and emotional ways of knowing from human geography is not surprising given that knowledge is widely understood to be grounded in reason as opposed to emotion, whereas emotion is often viewed as intrinsically unreasonable. Consequently, human geographers have been cautious and circumspect about engaging with issues of emotion. Nevertheless, emotion and emotional ways of knowing have found their way into several strands of human geography. Further more, human geographers have also challenged the supposed opposition between reason and emotion, and the notion that knowledge is necessarily emotion free. This has prompted an upsurge of interest in questions of emotion, including the methods appropriate for gener ating knowledge about emotions, and the status of emotion as a source of geographical knowledge. These developments have been inspired especially by feminist, nonrepresentational, and psychoanalytic ideas.

Passionate Traditions

Apparent Dispassion Underlain by Personal Values and Emotions

Traditionally and popularly, scientific knowledge is grounded in value free, objective reasoning and in facts about the world that can be verified by others. Empirical evidence, laws, and theories are linked together by explicit logical rules. From this perspective, emotions are seen as a potential source of bias, liable to undermine logic and objectivity, and to cloud dispassionate judgment, from which scientific knowledge therefore needs to be protected. Emotion, it appears, has no place in the generation of sound, reliable, scientific knowledge about the world.

Accounts of scientific knowledge focus on the methodological frameworks through which the scientific status of knowledge is warranted. These accounts do not address the choices researchers make about the topics they select for investigation, or about forms of scholarship that do not aspire to scientific status. In these domains, less dispassionate and more emotionally infused ways of knowing the world and generating knowledge are evident. Thus, many human geographers (like other researchers) are motivated or inspired to explore particular topics for essentially emotional reasons, including personal values and desires. Examples include anger at injustice inspiring research on geographies of inequality, and passionate attachment to particular places inspiring research about conservation, development, and planning. In these examples, emotion plays a part in initiating the processes through which human geography develops. While dispassionate reason may guide the subsequent production of knowledge, motivating emotions are often essential to inspire the continuing development of the discipline.

Even after research has been initiated, emotions may have a decisive and constructive part to play in the development of knowledge. While reason and logic may frame the formal processes of knowledge production, intuition and inspiration also play a part, perhaps especially in finding new and apparently ingenious solutions to problems. An archetypal example is provided by the story of Archimedes. According to this story, Archimedes had been trying to work out how to calculate the density of a crown in order to determine whether it really was, as claimed, made of gold. The solution to this problem came to him suddenly as he stepped into the bath and observed the displacement of water. In this moment he was not applying logical reasoning to the problem but was relaxing. The moment of realization triggered such excitement that Archimedes is reputed to have leapt out of the bath and run into the streets of Athens naked, shouting “eureka” (I have found it). While the whole problem was solved only when he had worked with his insight to generate a formal, logical explanation of the relationship between volume and density, this formal solution was discovered only through the moment of highly emotional knowing in which Archimedes achieved his crucial insight.

Challenging Exploitation, Inequity, and Discrimination

The idea that knowledge is ever value free has been contested by many human geographers, especially since the 1970s, which witnessed the rise of politically committed approaches to research and scholarship. Scholars informed by Marxist, feminist, antiracist, and antidisablist perspectives criticized all claims to value freedom. Their critiques identified the values implicit in the conduct of research that claimed to be objective, value free, and scientific. For example, models designed to determine the optimal locations of specific facilities might incorporate assumptions about those using the facilities in question that implicitly exclude or disadvantage particular groups, such as those without access to a private car or those with mobility impairments.

Research informed by explicit political commitments eschews any claim to a neutral vantage point and instead aims to challenge exploitation, inequity, and/or discrimination. Contributions to human geography ensuing from these starting points are often palpably infused with powerful emotions, such as anger at injustice, compassion toward the oppressed, and desire for a fairer world. Such research exposes the mechanisms and impacts of discrimination and oppression, laying bare the extent and the consequences of the hurt and injury inflicted, in ways designed to elicit emotional as well as rational responses.

As with more scientific approaches, these traditions have not, until recently, self consciously and explicitly espoused research about emotions, or emotional ways of knowing about the world. However, by arguing for the explicit acknowledgment of values, and through the self evidently passionate commitments they express, antioppressive geographies have further validated the importance of emotional inspirations for knowledge production.

Humanizing the Human in Human Geography

The 1970s also witnessed the rise of humanistic geography, which sought to rehumanize human geography by focusing explicitly on subjective aspects of human experience. This tradition challenged the idea, typical of spatial science, that human action can be modeled and predicted according to the principles of economic rationality. Humanistic geographers argued that people do not have the information or the inclination to behave in such ways, and that human geographers might more usefully explore, rather than making assumptions about, the subjective realities underlying human action. Humanistic geographers also criticized politically committed geographers insofar as they represented people as passive victims of the circumstances in which they found themselves. Against such approaches, they sought to understand the nature of human agency and human meanings. In this context, humanistic geography drew on a range of philosophical perspectives to develop new ways of thinking about the relationship between people and their contexts or environments. For example, the concept of the life world was taken up with a view to transcending oppositions between external, objective realities and internal, subjective perceptions.

Like the other traditions described above, explicit engagement with emotion was quite limited in much of this work. However, the call to understand subjective experience generated new explorations of people's feelings about and for a variety of places and spaces. Studies ensued about such topics as what it was like to live in an old people's home, or about the feelings evoked by places under threat from the building of new roads. These studies drew on methods designed to generate insight about people's subjective worldviews, including in depth interviews and ethnography, which were subsequently taken up and further developed by feminist geographers interested in generating knowledge of, and explicitly informed by, emotion.

Integrating, Embodying, and Narrating Emotion: Feminist Perspectives on Emotional Knowing

Gender, Emotion, and the Critique of Binaries

The approaches discussed in the preceding section have engaged with emotional ways of knowing more by accident than design. One of the first explicitly emotional topics for research was about the geography of women's fear. As well as validating the emotion of fear as an important topic for geographical research, this work was notable for its consideration of fear as something stimulated, exacerbated, countered, or alleviated by environments. This perspective suggested that fear and other emotions, such as joy and anger, might permeate social and physical environments as well as people's interiorized subjective experience. Put another way, emotions could usefully be understood as arising at the interfaces between people and their contexts rather than being thought of as essentially individualized and psychological phenomena.

Some early studies of the geography of women's fear challenged the notion that women's fear is irrational. In so doing, this body of work contributed to the process of rethinking the relationship between emotion and reason, in which feminist geographers sought to undo the idea that rationality and emotion are mutually exclusive oppositions. The binary distinction between reason and emotion is closely linked to a gender binary in which emotion is associated with femininity and reason with masculinity. Since the late 1970s, feminist geographers have sought to challenge such associations, and, drawing on post structuralist theories, they have sought to destabilize the binary structure characteristic of much geographical thought, and they have sought to undo the mapping of emotion onto women, and the treatment of emotion as unreasonable.

According to post structuralist theory, mutually exclusive oppositions in which one term is more highly valued than the other – such as reason and unreason or male and female or mind and body – conceal the dependence of the superior term on the supposedly inferior one. For example, reason is defined by, and therefore dependent on, what it is not, namely, unreason. Reason is therefore never truly free from unreason. Given that emotion is often viewed as intrinsically unreasonable and opposed to reason, this argument suggests that reason is in fact never emotion free. Conversely, emotion is never wholly without reason. Rather, emotion and reason are mutually interwoven with one another. In this context, we might think of dispassionate modes of knowledge production as expressions of emotional attachment to the idea of binary thinking. This approach also suggests that, even if unacknowledged or denied, emotional ways of knowing are always present within the production of knowledge in human geography and other disciplines.

Embodying and Situating Knowledge

The dismantling of binary thinking in feminist geography has developed further as feminist scholars have engaged with questions of embodiment as well as emotion. Emotions are intrinsically sensory. We feel and we are touched by emotions bodily. We smell, taste, see, and hear emotion, whether understood as arising within us or encountered in our interactions with people and places. Through the senses, emotions are inextricably bound up with the body and embodiment, and the idea of emotional knowing is closely linked to sensory, embodied ways of knowing. In this context, feminist geographers have explored a wide variety of intertwined emotional and embodied knowledges. Examples include research focusing on troubled and troubling experiences, such as women's experiences of living with chronic and contested illnesses or conditions such as agoraphobia and fibromyalgia. These studies generate rich accounts of feelings in which the embodiment of emotion and the emotionality of embodiment are powerfully articulated.

Consideration of emotional and embodied aspects of knowledge production has contributed to critiques of universal knowledge claims. If our feelings and bodies are integral to our knowledge of the world, it is not possible to generate knowledge that transcends the perspective of the knowledge producer. All that we claim to know is marked by our specificity, including our location within systems of difference that locate us in relation to places, times, and peoples. Emotional knowing, as it has been taken up and explored by feminist geographers, offers avowedly situated ways of understanding our worlds, in which the subjective positions and contexts of all those involved in knowledge production (researchers and re search participants) are acknowledged and viewed as integral to processes of knowing.

Narrating and Co-Constructing Emotion

At the core of feminist geography is a commitment to giving voice to women and men in relation to their lived experience, especially those whose experiences tend to be marginalized in existing bodies of knowledge, policy, and practice. Consequently, a great deal of attention has been devoted to methodological questions about the conduct and representation of fieldwork.

Feminist researchers reject the idea of the detached, objective, neutral fieldworker and attend instead to par ticular positions they occupy in relation to research participants. They generally seek to establish egalitarian relationships with research subjects, with a view to cre ating trustworthy and safe contexts in which research participants can talk about their lives. Underlying this is the idea that self narration by research participants has a central part to play in the production of knowledge. These narrations communicate emotional dimensions of lived experience. Participants talk about how they feel and thereby make their emotional knowledge available to researchers.

Narration does not take place in isolation, and even when researchers invite participants to produce nar rations in their absence, for example, by keeping written or photo diaries, the data generated are produced in the context of relationships and imagined audiences. Field work data are always, therefore, co constructed, in the sense that their production involves researchers as well as research subjects. Feminist researchers have placed par ticular emphasis on the importance of reflexive self monitoring in order to help understand the character of these co constructions. Researcher reflexivity involves considering the emotional impact of the relative positions of research subjects and researchers, including, for ex ample, how it might feel to talk to someone of similar or different age, gender, and class position, within a par ticular social and environmental context.

The narrations co constructed by research subjects and researchers are analyzed and represented in the process of producing academic knowledge. Feminist re searchers have sought to involve research participants in these phases of research with a view to ensuring that they convey the voices of participants effectively and honor the trust placed in them. Thus, the emotional knowing constructed in the course of feminist fieldwork is understood as delicate, fragile, and vulnerable to ex ploitation and mistreatment. Indeed, feminist geography is deeply marked by awareness of the emotional injuries inflicted through misrepresentation and exploitative re search practices. It is the appreciation of these risks and injuries that has informed much feminist methodological debate about emotional knowing.

Enlivening, Affecting, and Rethinking Emotion: Emotional Knowing Beyond Representation

Escaping Representation and Honoring the Ineffable

While feminist geographers have sought to give voice to the experiences – including the emotional knowledge – of marginalized others, the idea that emotions can be represented effectively through narrative accounts has also been subject to challenge. Thus, some geographers have argued that discursive representations of lived ex perience are profoundly flawed and actively obscure the very vitality researchers seek to convey. According to this critique, geographers need to engage with what people do and feel rather than with what people say they do and feel, and therefore with emotions rather than with nar rations of emotion. Furthermore, this argument suggests that the co constructing emotional narrative replaces emotions with discourses about emotion, and, in so doing, it displaces and even destroys emotional knowing as well as deadening rather than enlivening geographical knowledge.

In order to get beyond representations and discourses, researchers inspired by this argument have often focused their attention on practices that are intrinsically non verbal, such as dance, yoga, meditation, massage, and reverie. It is the embodied feel of such practices that studies explore. However, researchers have also ap proached talk and conversation as performative rather than representational practices. Whatever the practice on which research focuses, attention is given to the sensory, embodied, spiritual, and affective doing, rather than the substantive content or overt meaning of what is said or communicated. Emotional knowing is understood to reside in and to be accessible only via the practices themselves.

As well as seeking to convey emotional dimensions of practices without recourse to narrative accounts of emotion, research from this nonrepresentational per spective has also explored emotion as intrinsically unarticulable and ineffable. Sometimes focusing on specific emotions such as hope, and sometimes on unnamable emotional qualities associated with loss or with love, this work emphasizes what cannot be rendered into words. In so doing, it draws attention to limits of representational modes of knowing and valorizes nonrepresentational emotional knowing.

Beyond Cognitive Knowledge and Beyond Personal Feelings

Research within this nonrepresentational tradition often uses the term affect rather than emotion. The rationale for this preference is an association made in psychology between emotion and cognitive representations of feelings, in contrast to an association between affect and noncognitive experiences. While the distinction between emotion and affect is contested (e.g., by feminist geographers in the context of the critique of binary thinking discussed above), the key point to be made is that nonrepresentational approaches to emotional knowing are concerned with, and focus on, the noncognitive. From this perspective, emotion/affect lies beyond cognition and constitutes a distinctive mode of knowing about the world that is radically different from cognitive knowing. One feature of this otherness is that emotional/affective knowing always exceeds representability.

As well as emphasizing what lies beyond cognition, nonrepresentational approaches challenge the idea that emotions are essentially individualized phenomena, felt by and belonging to human individuals. This underlines the presumption that emotional knowing is not available through self reflection. Moreover, because the idea that emotions belong to individuals is so strong, at least within Western cultures, scholarship within this perspective is often practiced in a way designed to disrupt such assumption. For example, to convey a nonindividualized understanding, emotion/affect may be described as inhuman. Certainly, from a nonrepresentational perspective, the task of research is not to provide insight into how individuals feel about their lives. Rather, alongside, conveying ineffability and nonrepresentability, this work seeks to retheorize emotion/affect in new ways that actively decenter conscious human agency.

Practices, Networks, and Posthuman Ways of Knowing

Nonrepresentational geographers draw on a variety of resources in order to undertake this retheorization. The focus on noncognitive, nondiscursive practices points to ideas about repetitive, unconscious, habitual everyday practices elaborated by writers, including Bourdieu and de Certeau. However, these theorists have not proven particularly popular among geographers concerned with the development of nonrepresentational work on emotional knowing. Two other sets of ideas have been more influential. First, Latour's ideas about the agential properties and circulating effectivities of linkages and networks have proven attractive because of their capacity to radically reframe and reorient debates about agency away from the traditional focus on human beings. In other words, Latour's work helps geographers to rethink the subjects and subjectivity of emotional knowing as nonhuman, inhuman, or transhuman. Second, the writings of Gilles Deleuze have been especially influential within this body of work to such a degree that nonrepresentational approaches are sometimes thought of as intrinsically Deleuzian. Deleuze provides resources for working against the grain of individual subjectivity, describing affects and desires as independent of human subjects.

Although nonrepresentational approaches to emotional knowing contest some features of feminist approaches, including especially the status of narrations of emotion, there are also important points of connection and convergence. For example, some feminist theories of embodiment also seek to write against the grain of body as biological organism and to disrupt assumptions that equate subjectivity with individuals. Indeed, where feminist and nonrepresentational approaches tend to converge, for example, in their insistence that emotion is a form of knowing, they also share a good deal with psychoanalytic perspectives.

Relating, Flowing, and Connecting Emotion: Psychoanalytic Approaches to Emotional Knowing

Unconscious Relational Registers of Emotion

Like nonrepresentational approaches, psychoanalysis is not a singular body of theory, but, among the ideas, its different versions share is that of the unconscious. In psychoanalytic theory, the unconscious is not merely that which is not conscious, but it is related to conscious life in a necessarily conflictual and antagonistic way. It is radically other to conscious life and always present within it, adding layers of meaning to all that is said and done in ways that necessarily exceed conscious meanings, and that are always liable to disrupt conscious intent. The Freudian slip (in which what is said is not what the speaker had consciously intended but which can be understood as expressing an unconscious thought) is a classic example. Expressions of unconscious life are emotionally rich and often highly charged. Indeed, emotional experience often provides clues about the content of unconscious life, as, for example, in moments when we are surprised or disturbed by our feelings. Unconscious life is like a rich reservoir of noncognitive knowledge, often deeply felt, but not directly available to conscious thought.

Different versions of psychoanalysis also share the idea that unconscious life is configured relationally, both intrapsychically and interpersonally. Unconscious life can be thought of in terms of objects located in a variety of places, including inside and outside our own representations of ourselves. These objects are images of (or transfers from) anything that is salient within our unconscious lives. Psychoanalytic theory postulates that images of significant others (classically our parents), or aspects of significant others are configured in this way, and also acknowledge that significant others may be signified in a multitude of different and often indirect ways. Emotionally redolent relationships operate between, and are constitutive of, the multitude of signifiers that forge unconscious life. This is integral to the psychoanalytic concept of transference, according to which the impress of our unconscious patterns of relating is necessarily transferred across different interpersonal relationships, classically but not solely from patterns generated by our earliest relationships (normatively with parental figures) to subsequent ones.

As this account indicates, psychoanalysis and non representational approaches start from very different points, the former arguing that meaning and knowledge of all kinds are inextricably bound up with processes of representation, and the latter seeking to access meaning and knowledge prior to and beyond representation. However, the two approaches unite in their emphasis on noncognitive and non (or un) conscious modes of knowing (including emotional knowing), and in their challenges to assumptions about human subjectivity as agential, intentional, conscious, individualized, and bounded.

Emotional Flows and Fluid Boundaries

According to psychoanalysis, a sense of subjective interiority is produced at least in part through the process of locating salient signifiers inside self representations. But signifiers may be located anywhere, for example, when emotion states such as anger are projected onto others, so that subjective boundaries never correspond directly to self representations. Instead, self making requires endless psychic work, which is inevitably rich in emotion, and emotional knowing is part of what it means to be human. For geographers interested in emotional knowing, psychoanalytic ideas are useful because they provide a way of thinking about the spatiality of subjectivity as well as a spatialized approach to emotional knowing.

Geographers working with psychoanalytic ideas (in common with those drawing on nonrepresentational approaches) have emphasized the mobility and fluidity of emotion. From a psychoanalytic perspective, this fluidity is integral to the idea that the unconscious is constituted by signifiers, the meanings of which are necessarily relational rather than fixed. Emotion is understood as movement between, through which connectivities are expressed, created, and sustained. Emotional knowing enables and requires boundaries to be crossed and reworked. Thus, feelings about selves and environments actively connect people to specific landscapes, blending qualities of places into people's sense of who they are.

Connecting (to) Fields of Emotion

Psychoanalytic ideas present challenges to, and resources for, researchers seeking to generate emotional knowledges. If the omnipresence and radical alterity of unconscious life are taken seriously, then the traditional caution around emotion as source and topic of knowledge may seem to be entirely appropriate. However, as a few geographers have argued, psychoanalysis, and some of the practices it has spawned, also offers useful pointers for exploring emotional life in ways alert to, but not overwhelmed by, the challenges presented by the unconscious. For psychoanalysts and other practitioners who draw on psychoanalytic ideas, one's own emotional experience (or countertransference) provides a point of access to the field of emotional dynamics in which one is located. Psychoanalytic practices create very precisely specified fields in which practitioners use highly refined skills to sense what is happening. If emotions are understood as always relational and as at least potentially symptomatic of unconscious processes, then something broadly analogous might be possible in other contexts. Researchers' uses of self experience may be much less refined than those of psychoanalysts, and their purposes may be quite different, but if emotion always entails movement between, then our own feelings are likely to provide clues about the emotional contexts in which we are embedded. On this account, our own emotional knowing is integral to any attempt to generate knowledge about emotions.

There are connections between this perspective and both the reflexive self monitoring advocated by feminist geographers, and the immersion in practices advocated by those working with nonrepresentational perspectives. With respect to the former, psychoanalytic ideas support and provide conceptual and practical resources for exploring the interpersonal dynamics of fieldwork, including the ways in which researchers and research participants take up, are drawn into and move between different positions in relation to one another. Psychoanalysis shares with the latter the idea that emotional knowing requires embodied, noncognitive immersion within, rather than the production of representations of, practices. In such ways, different approaches to emotional knowing are not necessarily mutually exclusive.

See also: Affect; Agoraphobia; Embodied Knowing; Emotional Geographies; Feminism/Feminist Geography; Feminist Methodologies; Non-Representational Theory/ Non-Representational Geographies; Performativity; Psychoanalysis; Psychotherapy/Psychotherapeutic Geographies; Situated Knowledge, Reflexivity; Subjectivity.

Further Reading

Ahmed, S. (2004). *The Cultural Politics of Emotion*. Edinburgh: Edinburgh University Press.

Anderson, B. (2004). Time stilled space slowed: How boredom matters. *Geoforum* 35, 739 754.

Bennett, K. (2004). Emotionally intelligent research. *Area* 36, 414 422.

Bingley, A. (2003). In here and out there: Sensations between self and landscape. *Social and Cultural Geography* 4, 329 345.

Bondi, L. (2003). Empathy and identification: Conceptual resources for feminist fieldwork. *ACME: International Journal of Critical Geography* 2, 64 76.

Bondi, L. (2005). Making connections and thinking through emotions: Between geography and psychotherapy. *Transactions of the Institute of British Geographers* 30, 433 448.

Bondi, L., Avis, H., Bingley, A. *et al*. (2002). *Subjectivities, Knowledges and Feminist Geographies*. Lanham, MD: Rowman and Littlefield.

Chodorow, N. (1999). *The Power of Feeling*. New Haven, CT: Yale University Press.

Davidson, J. (2003). *Phobic Geographies*. Aldershot: Ashgate.

Davidson, J., Bondi, L. and Mick, S. (eds.) (2005). *Emotional Geographies*. Aldershot: Ashgate.

Harrison, P. (2007). How shall I say it…? Relating the non relational. *Environment and Planning A* 39(3), 590 608.

Hochschild, A. (1979). Emotion work, feeling rules and social structure. *American Journal of Sociology* 85, 551 575.

Laurier, E. and Parr, H. (2000). Emotions and interviewing in health and disability research. *Ethics, Place and Environment* 3, 98 102.

Nigel, T. (2004). Intensities of feeling: Towards a spatial politics of affect. *Geografiska Annaler* 86B, 57 78.

Parr, H. (2005). Emotional geographies. In Cloke, P., Crang, P. & Goodwin, M. (eds.) *Introducing Human Geography*, pp 472 485. London: Arnold.

Empire

M. Power, University of Durham, Durham, UK

Glossary

Decolonization The process by which colonies achieve their political independence from the colonial metropolitan power. This also relates to an intellectual movement known as postcolonialism which seeks to examine the legacies of colonialism and imperialism.

Hierarchization The process by which the expansion of political power establishes dynamics of inequality, geographies of uneven development, or unequal relations of exchange and accumulation.

Imperialism The creation and maintenance of an unequal economic, cultural, and territorial relationship usually between states and often in the form of an empire, based on domination and subordination.

Neocolonialism A means of economic and political control articulated through the powerful states and capitals of 'developed' economies over the economies and societies of the 'underdeveloped' world.

Neoliberalism A political project that entails a particular set of economic and political ideologies based, in part, on eighteenth-century liberal economic theory. Concerned primarily with theoretical economic models, neoliberalism is a theory of how to structure economies through free markets.

Orientalism The scholarly study of and general interest in 'the Orient', which was a predominantly European and American construction, produced within a grid of power, knowledge, and geography. Orientalism is also a discursive formation and corporate institution for the production and domination of the 'Orient' by the 'Occident'.

Postcolonialism A critical political and intellectual formation concerned with the impacts of colonialism and its contestation on the cultures of both colonizing and colonized peoples in the past and the reproduction and transformation of colonial relations, representations, and practices in the present.

Introduction: Geography and Empire

The word 'empire' evokes visions of grandeur and glory, conquest and slavery but, as we shall see, the word has had a complicated history and many different, fiercely contested meanings. It has also often been combined with several other contentious words like imperialism, colonialism, globalization, neocolonialism, and others. To complicate things further, a great deal of compound terms have been added to the mix such as informal empire, cultural imperialism, internal colonialism, postcolonialism, and many more. Projects of empire have also been very varied through time and not always mutually supportive and compatible including, for example, projects of conquest, commerce, settlement, and rule. The difficulties are not just conceptual but also political and emotional since to define something as imperial or colonial today almost always implies hostility to it. Across several imperial formations through history, to be an empire builder was to be an adventurer, a hero, and a selfless laborer for others' well being. Only over the last century has this changed (where today's empire builders are met with near universal distaste). Empires both past and present have profoundly shaped the contemporary world and empires live in memory long after they have faded from power.

In many ways geography should be treated as both a discipline and a discourse of empire: as a set of geographical ideas, institutions, and practices that induced and legitimized territorial expansion; and as a dynamic medium through which European attitudes of dominance and core periphery or metropolitan–colonial relationships were imagined, negotiated, and represented. Empire has thus been central to the making of the discipline of 'geography' which produced and disseminated a wide variety of material and imaginative resources that aided the exploration and colonization of many parts of the non Western world, especially at the end of the nineteenth and start of the twentieth centuries. In the production of imperial imaginaries and in the formulation of visions of empire (including the production of ideologies of superiority and conceptual devices for ordering the world), geography both as a discipline and a discourse contributed to the rising capacity of Europeans to bring the world within their imperial visions. Geography then, at least to begin with, was a kind of 'handmaiden of empire' along with associated disciplines like anthropology. In this sense, geography served the causes of European imperialism, helping to create and maintain an unequal economic, cultural, and territorial relationship between states in the form of empires based on domination and subordination.

Over the last 500 years, imperialism has been a predominantly Western project and form of dominance shaped by expansionist (both capitalist and communist) systems. Europe's rise to global dominance in the nineteenth century stemmed from a much longer history of trade, settlement, exploration, and warfare however. From the fifteenth century explorers of Spain and Portugal through to the 'scramble for Africa' at the end of the nineteenth century and the climax of imperialism in the early twentieth century, the construction of empires has followed intrinsically geographical dynamics that involve extensions of sovereignty in one nation state over the land and lives of distant others. This process has invariably involved military conquest, the imposition of direct rule, colonial settlement, and informal empires of trade and political supervision. As European colonial empires began to collapse and dissolve in the latter half of the twentieth century the term has acquired other connotations, being used to describe the global economic influence of Japan or the USA, the power of transnational corporations, the geopolitical spheres of influence maintained by China or the USSR during the Cold War, and the management of the so called Third World by Western countries and institutions.

Today the predominant motif in investigations into questions concerning the nature and possibilities of the contemporary world order has been that of 'empire'. A significant number of speculations concerning the imperial nature of our times have appeared since late 2001, many influenced by Marxist–Leninist thought where imperialism has been used as a synonym for capitalism. The emergent consensus on the imperial nature of the contemporary world order is not intended to indicate a break between the post Cold War era and a brand new twenty first century imperial age. Speculations regarding the imperial form of the world order after 9/11 arose first of all in commentaries that argued that the contemporary conjuncture had more in common with the political form of the Roman Empire. Such arguments indicate that the concepts of 'empire' and 'new imperialism' are deployed not because of their uniqueness or singularity to the contemporary era, but because of their historical resonances that are seen to afford purchase upon, and thus deeper understanding of, contemporary world politics. Indeed, this resurgence of interest in imperial political forms suggests that a consensus is emerging that empire is a neglected category of political analysis. Many critics thus suggest that we need to conceptualize contemporary empires and imperialisms in more general terms – as subject to logics of power, as systems of violence, or as sets of historical tendencies and cultural perspectives since its effects are pervasive and wide ranging.

We are probably most likely to think of empire in connection with Rome; the term, in fact, comes from the Latin *imperium*, with its root meaning of order and command. Tacitus, the Roman historian, also used it to refer to the immense size and diversity of the Roman world. Empires predate the Romans by several hundred years, however: Egypt, for example, set up an empire in the sixteenth century BC by invading and then incorporating Nubia and the ancient city states of the Levant. The Akkadian Empire of Sargon of Akkad exists as one of the earliest models of a far flung, land based empire. Many different empires beyond that of Rome – for example, Mesopotamian, Ottoman, Persian, Chinese, Aztec, and Mughal – have appeared throughout history and around the world. By exploring the diverse historical, cultural, and geographical settings of empires, we can understand how and why empires emerged, when and where they did, how they were maintained, and how and why they came to an end. In most cases, military power has been critical to empire building. While empires generally introduced new political and administrative institutions, they also frequently adapted to existing institutions and local elites. Further, most empires spent considerable energy and resources in attempting to control production and trade within their realms. Ultimately, however, what is perhaps most interesting about the comparative study of empires is that peoples widely separated by time and place independently created common forms of political and social organization.

Defining Empire and Imperialism

There has been considerable debate about what constitutes an empire. There are thus many definitions of empire and its conceptualization has been the focus of some considerable debate. Some define an empire as a state that extends dominion over areas and populations distinct culturally and ethnically from the culture/ethnicity at the center of power. Other definitions may emphasize economic or political factors. The term generally implies military hegemonic power. Like other states, an empire maintains its political structure at least partly by coercion. Empires can be land based (such as the Mongol Empire) or can be seaborne (such as the Portuguese and the British empires) or they may feature looser structures and more scattered territories. Empire contrasts with the example of a federation, where a large, multiethnic state – or even an ethnically homogeneous one like Japan or a small area like Switzerland – relies on mutual agreement among its component political units which retain a high degree of autonomy.

Historical comparison, however, benefits from conceptual clarity and the application of a 'core' definition. Empire has often been defined as a state that is endowed with particular properties. This state (by contrast with bands, local groups, or chiefdoms) has been defined by a range of observers as a differentiated set of institutions and personnel, embodying centrality. In this sense, political relations radiate outward from a center to cover a territorially demarcated area, over which the empire claims a monopoly of binding and permanent rule making, backed up by physical violence. Empires differ from generic states not so much in terms of size as in terms of diversity and internal political differentiation. Clearly 'empire' intuitively connotes large scale: however, given that scale is necessarily relative in the sense that it is contingent on historical context, it is impossible to set any absolute quantitative thresholds for imperial statehood. Moreover, the meaning of empire has varied over time, which requires us to understand the different historical contexts in which the term has been used. The emergent literature on the 'new imperialism' is often at pains to delineate a distinction between 'imperialism' on the one hand and 'empire' on the other. This distinction is intended to capture a break from twentieth century Marxist understandings of European colonial imperialism. Marxist accounts have been criticized for providing overly reductionist representations of the relationship between the political and the economic and for being bound to the explanation of the historically particular form of nineteenth and twentieth century European colonial empires. Indeed, as we shall see below, this emergent literature implicitly follows the argument that the present world order is something altogether different from 'imperialism'.

Some of the first global empires were a product of the European age of exploration that began with a race of exploration between the then most advanced maritime powers, Portugal and Spain, in the late 1400 s. Over the next four centuries, Europeans began to establish empires across the world that, according to Eric Hobsbawm, ultimately resulted in the 'age of empire' (defined as 1875–1914), where colonial empires were formed and where the economic character of industrial capitalism proved to be critical. Imperialism in this sense is the salient 'inner contradiction' in the phase of capitalist development. Capitalism, according to Marx and Engels, creates a world after its own image and thus compels all nations, on pain of extinction, to adopt the 'bourgeois mode of production'. In this 'age of empire' between 1875 and 1914, a half dozen European powers parceled out colonies among themselves amounting to one quarter of the planet's land surface. For some observers, such as Hobsbawm, imperial expansion was but one response to the crisis of profitability that beset capitalism in the great depression of 1873–96 (the others included protective tariffs, cartelization, and pressures on labor). Though other explanations for the remarkable burst of colonization at the turn of the century have emerged, attempts to divorce the explanation of imperialism from the specific developments of capitalism in the late nineteenth century have often been regarded as ideological exercises.

In theory, capitalism was profoundly internationalist, without loyalty to community or nation, dedicated only to profit wherever or however it might be made, impatient of political frontiers or any other so called artificial constraints on the efficient flow of the factors of production. In practice, however, this model was inadequate given that the evolving world economy of capitalism was a collection of solid national state blocs, as well as a fluid entity. In this sense, the age of empire has been considered a period of intense rivalry between national state blocs. The clash between the universalizing and the nationally competitive tendencies of capitalism formed the central thesis of Lenin's *Imperialism: The Highest Stage of Capitalism* (1917) in which he argued that the resultant imperial rivalries had engendered the fratricidal capitalist war then raging. Lenin quotes Cecil Rhodes as saying that colonialism and imperialism abroad were the only possible ways to avoid civil war at home. Although it was not direct economic confrontation *per se* that led to armed conflict, global capitalist expansion did engender an atmosphere of limitlessness, in which the linkage of national interest to concretely defined objectives was severed. States might no longer be able to specify their own national necessities, but they did believe that any concession, large or small, to any adversary, real or potential, could eventually prove fatal.

More recent definitions of empire touch on the elements of hierarchy and heterogeneity, for example, and have focused much more directly on indirect forms of power and control than has previously been the case. Here empire is a relationship, formal or informal, in which one state controls the effective political sovereignty of another political society and it becomes a system of interaction between two political entities, one of which, the dominant metropole, exerts political control over the internal and external policy – the effective sovereignty – of the other, the subordinate periphery. This form of domination accommodates various forms of 'indirect' rule, (i.e., the exertion of political control through and in conjunction with the elites of the subjugated peripheries) and does by no means presuppose direct control from the center over the actual management of these peripheries. Related definitions of empire tend to stress diversity and alienation. For instance, empire might be seen as a state in which one or more regions have a living memory of being part of a different autonomous polity, which was at some point brought under the rule of the current imperial power. As this memory fades, a unitary state evolves.

Diversity *per se*, however, is of limited value in defining empire simply because all prenation states in history that were bigger than city states tended to be heterogeneous in terms of language, religion, and other defining cultural characteristics: if diversity were the critical variable, most premodern states would have to be subsumed within the category of empire. Many observers have noted that empires almost constantly fight and seek to separate the zone of conflict, the frontier, from the zone of peace, the core – at least in theory. In this view, most nation states begin as empires in that they finish this homogenizing job, imposing linguistic uniformity and a sense of encompassing identity. Empires and nation states may thus be conceived as two ideal types that bound a range of combinations in between. Common to all empires and an institution in its own right, the frontier marks the boundaries between insiders and outsiders and always becomes a contested fault line along which acts of violence accumulate. For some, empire may be regarded as a centrifugal violence transfer, from the interior to the frontier.

These defining qualities of political hierarchy and cultural diversity have also been combined in some definitions where an empire is understood as a large, composite, multiethnic, or multinational political unit, usually created by conquest and divided between a dominant center and subordinate (sometimes very distant) peripheries. The culturally diverse character of the imperial whole, on the other hand, while often a logical and indeed inescapable consequence of conquest, cannot properly be regarded as an essential action, condition, or ingredient of empire. Overall, core periphery centered concepts work well for any empires that are republics, because the citizenry of the dominant core will strive to preserve any privileges associated with its formal separation from the population of the dominated periphery, as well as for any modern colonial empires with their European or North American cores and geographically and culturally distant possessions. Several definitions thus see empire as a form of political organization in which the dominant state – the 'mother country' or the 'metropole' – attempts to create a network of allied elites in regions abroad who accept subordination in international affairs in return for the security of their position in their own administrative unit (the 'colony' or, in spatial terms, the 'periphery'). In this sense, there is an explicit differentiation of empires from nation states, where states can be socially and ethnically homogenous but the empire is differentiated: territorially, religiously, nationally, and occupationally. In this formulation, empires stabilize inequality at home and export social hierarchies.

While modern definitions of empire that emphasize the domination and exploitation of a periphery by a dominant core owe much to the experience of modern colonial empires, this does not limit the utility of such definitions to that particular context. It is hard to envision any kind of imperial expansion that is not sustained by a core or metropole (however diffuse, as in the case of the Mongol or Ottoman conquest empires), which claim certain entitlements. Conquest, by contrast, is not an essential ingredient. This distinction begins to break down in the case of very long lived empires, such as the Roman Empire under monocracy – a period in which the privileges of the core gradually declined to zero – and China after the Qin unification. In other cases, most notably Old Kingdom Egypt, core periphery relationships disappeared much more rapidly and territorial states emerged. This problem underlines the need for a dynamic definition of empire that gives due considerations to two issues: that states we may not ordinarily think of as 'empires' often underwent an 'imperial' phase of core periphery dominance and attendant diversity – say, France or Britain; and that even very large and initially highly heterogeneous empires may, over time, turn into states that retain some of their diversity but lose a distinctive core (such as the mature Roman Empire or China). 'Empire', therefore, is best understood both as a particular type of political entity 'and' as a stage or phase of state formation.

Empire State Building: British Imperialism

Modern European empires were modeled on those of Greece and Rome but states such as Britain built the first truly global empires and grappled with issues that their predecessors did not. The British Empire produced Europe as a sovereign subject seen to be at the pinnacle of human civilization and at the center of an imaginative grid and global hierarchy of civilization that positioned all other continents as subordinate spaces. Imperial expansion was conceived in triumphalist terms (although this was often ridden with anxiety and fear), as a universally beneficent agent of progress and as an inevitable consequence of European history. Similarly, non European lands and peoples were depicted as 'uncivilized' or 'backward' and hence in need of domestication and rule. Not only in Britain, but also in many other European colonial empires, colonialism contributed to the making of European modernity and elaborated 'imaginative geographies' of 'us' and 'them', fictitiously creating binary oppositions between a supposedly dynamic, rational, masculine, and democratic 'Occident' and an eternal, irrational, feminine, and despotic 'Orient'. In this sense, particularly in Britain, travel writers, cartographers, surveyors, professional geographers, geographical societies, and landscape artists contributed to empire and shaped imperial categories of thought and colonial practices in numerous ways.

It is difficult and unwise however to generalize from the experience of the British Empire given that European states developed distinct imperial ideologies and styles of power and dealt with diverse subject peoples in often very different ways. Different agents of empire (such as explorers and merchants, governors and settlers, and engineers and armies) made the world over in the image of particular interests (such as those of gender, race, or class) and drew upon culturally specific ideas and practices. Further, European colonists often forged identities that diverged from metropolitan visions of empire and in some cases played a role in leading colonial societies out of empire. British and other European identities were often renegotiated in the course of imperial passage through other non European landscapes and cultures. European colonial empires were thus geographically variegated systems of power and knowledge. Empires were also heavily shaped by covetous national agendas and were implicated in the production and construction of national identities. British national identity, for example, was (among other things) shaped by a range of imperial ideas that revolved around the utilization of space. Contemporaries argued that Britain's strength and distinctiveness in Europe and the imperial world rested on principles of property and liberty and practices of cultivation and international commerce.

A product of the European age of discovery, the British Empire was the most extensive empire in world history and for a long period was the foremost global power. By 1921, the British Empire held sway over a population of 458 million people (or approximately one quarter of the world's population) and covered 14.2 million square miles (one quarter of the earth's total land area). As early as 1550, a committed policy of colonizing Ireland had been put in place centered upon a number of plantations that would later form the templates of empire, overseen by key figures like Walter Raleigh and Frances Drake who would later also play an important role in the early colonization of North America. The British Empire took shape during the early seventeenth century with the settlement of the Eastern colonies of North America (which would later become the United States), Canada's Atlantic provinces, and the colonization of Caribbean islands like St. Kitts, Barbados, and Jamaica. Slavery was central to the sugar plantation economies the British established in the Caribbean while the American colonies provided rice, cotton, and tobacco among other things. The settlement of Australia (through the formation of penal colonies from 1788) and New Zealand created a major zone of British migration. In the seventeenth and eighteenth centuries, the focus of British expansion shifted from the Americas to Asia and later Africa following the American war of independence (1775–83). In the 1840s and 1850s, self governing colony status was granted to Australia and Canada given that their inhabitants were seen to be closely aligned with the interests of the 'mother country' but this status was not granted to Ireland which had been incorporated into the United Kingdom of Great Britain and Ireland in 1801 following the Irish rebellion against British rule in 1798. During this time and following the abolition of slavery in British colonies in 1834, Britain expanded its political and economic influence through 'informal empire' which was underpinned by notions of free trade and a strategic preeminence based on naval dominance.

One of the first multinational corporations, the British East India company, led the annexation of most of the Indian subcontinent and the conquest of Hong Kong, Singapore, and Malaya and other surrounding Asian countries. Established in 1600, with the intent to favor trade privileges in India, the company originally functioned as a trading concern establishing an informal empire in the subcontinent, but by the 1850s the company functioned more like a state, engaging in conflicts with neighboring Asian countries. By this time India had become an extremely important part of the empire's economy. As the company struggled to maintain control over its conquered Indian territories, the Crown and British government began to assume direct rule over the Indian subcontinent lasting for 90 years following the dissolution of the company. This period of direct rule, known as *The Raj*, saw the nations now known as India, Pakistan, Bangladesh, and Myanmar collectively referred to as British India. Britain's overseas commercial dominance and free trade ideology (which had allowed it to master control of the world's raw materials and markets) was beginning to be challenged in the latter half of the nineteenth century, however, by countries like France, Germany, and the US and by the rising tide of protectionism. While early European colonial empires were seen as mercantile in nature and promoted by European states through merchant companies, this new period of 'classical imperialism' (Hobsbawn's 'age of empire') was rooted in a new phase of capitalist development and interimperial rivalry. There were growing concerns in Britain that existing trade links and overseas markets would be undermined by the increasingly closed imperial trading blocs being established by rival powers.

Forced to join the renewed 'scramble' for formal empire at the end of the nineteenth century, Britain became part of the 'new imperialism' in a period during which there was an unprecedented pursuit of 'empire for empire's sake', involving aggressive competition for overseas territorial acquisitions. During this time Europe's powers added nearly 8.8 million square miles to their overseas colonial possessions. A 'scramble for Africa' resulted given that the continent had been mostly unoccupied by the Western powers as late as the 1880s and the continent's territory was divided up formally between European colonial empires at the Berlin conference of 1884–85. This conference sought to regulate the competition between the powers by defining 'effective occupation' as the criterion for international recognition of territorial claims. Egypt was occupied from 1882 followed by Sudan (1896–98) and South Africa (1899). Between 1885 and 1915, Britain took nearly 30% of Africa's population under its control (compared to 15% for France and 9% for Germany). Nigeria with 15 million people had more subjects than the entire German colonial empire in Africa. Cecil Rhodes, a key figure in British imperial expansion in Africa, dreamed of a British Empire stretching across the continent from Cape Town to Cairo, connected by British technologies like the railway and the telegraph.

The British Empire began its transformation into the modern Commonwealth with the extension of dominion status to the already self governing colonies of Canada (1867), Australia (1901), New Zealand (1907), New foundland (1907), and the Union of South Africa (1910). Full dominion independence was formalized in the 1926 Balfour Declaration and the 1931 Statute of Westminster where each dominion was henceforth to be considered equal in status to Britain and autonomous in inter national relations. The aftermath of World War I had led to an expansion of British imperial possessions involving League of Nations mandates for Palestine and Iraq (fol lowing the collapse of the Ottoman empire) and Britain taking control of territories like Tanganyika and South West Africa (Tanzania and Namibia, respectively) fol lowing the collapse of the German colonial empire in Africa. By the end of World War II, however, the collapse of Britain's own empire was gathering momentum as anticolonial nationalist movements in British colonies began to challenge an imperial power more preoccupied with issues closer to home. India became independent in 1947 and over the following two decades most of Britain's colonies followed suit. The end of the British Empire in Africa took place with remarkable rapidity although withdrawal from the southern and eastern parts of the continent was complicated by the regions' white settler populations – white settler rule continued in South Af rica until the end of apartheid in the 1990s and the white minority in Southern Rhodesia (a self governing colony since 1923) made a unilateral declaration of independ ence rather than submit to the immediate majority rule of black Africans.

Although in 1997 Britain's last major overseas territory (Hong Kong) became a Special Administrative Region of China, Britain still retains sovereignty over 14 territories outside the British Isles which remain under British rule because of a lack of support for independence among the local population, because of a small population size dependent on British economic subsidies or because the territory is largely uninhabited except for transient mili tary or scientific personnel. British sovereignty of two of these territories, Gibraltar and the Falklands Islands, is disputed by their nearest geographical neighbors: Spain and Argentina, respectively. Many former British colonies (and even one former Portuguese colony) are today part of the Commonwealth of Nations.

Toward the Colonial Present and the 'New' Imperialism

The predominant motif in investigations into questions concerning the nature and possibilities of the contemporary world order has been that of 'empire'. For some contemporary observers, this 'new imperialism' appears as nothing more than the revisiting of the old, though in a different place and time. Since the start of the twenty first century, a significant number of speculations on the imperial nature of our times have appeared. These include a range of seminal interventions from authors including David Harvey, Michael Mann, Alain Joxe, Emmanuel Todd, Derek Gregory, Michael Ignatieff, and Niall Ferguson, as well as a range of articles in scholarly journals and news media. This writing has received added impetus from the global popularity of some of these works as they inquire into the contemporary global capitalist world order (particularly that of Michael Hardt and Antonio Negri) and the multiple scholarly critiques of this work that have emerged since 2000. Indeed, this resurgence of interest in imperial political forms suggests that a consensus seems to be emerging that empire is a neglected category of political analysis. Some of these interventions suggest that the world market is globalizing beyond the capacity of nation states to affect and shape it. In the 1990s, after the end of the Cold War, many scholars of international politics wrote about the emer gence of the postinternational order. This is in reference to the international system of nation states that began in Westphalia in 1648. This postinternational order is seen to be one in which the political sovereignty and terri toriality established in the 1648 treaty of Westphalia are surpassed and replaced with a deterritorialized or even supraterritorial order in which states would dissolve in global architectures of governance. The supposedly dis crete human communities hitherto represented by na tion states would thus emerge into a condition of global supraterritorial interconnection.

In this emerging literature on the new imperialism the sovereignty of nation states is thus said to be vanishing, replaced by a newly emerging global sovereignty or 'empire' arising from the coalescence of a series of na tional and supranational entities united under a single logic of rule, with no clear international hierarchy. Here, in some cases, 'empire' does not simply refer to imperi alist domination of the periphery by the center, but to an all encompassing entity that recognizes no limiting ter ritories or boundaries outside of itself. Some of these authors concerned with contemporary forms of empire thus argue that, insofar as they represent restraints on the homogenizing force of the world market, all forms of imperialism are doomed to fail. Empire is thus, in a sense, seen to be both 'postcolonial' and 'postimperialist'. Whereas classical imperialism is understood by some to have been a global machine that channels, codes, and territorializes the flows of capital (blocking certain flows and facilitating others); the world market, in contrast, is seen to require a smooth space of uncoded and deterri torialized flows.

Some observers see the Vietnam war as the final moment of the imperialist tendency and argue that concepts such as center and periphery are now all but useless given that through the decentralization of production and the consolidation of the world market, the international divisions and flows of labor and capital have fractured and multiplied so that it is no longer possible to demarcate large geographical zones as center and periphery, North and South. Empire, the name given to this supposedly New World Order, is argued to be a product of the struggle over sovereignty and constitutionalism at the global level. Many of these authors are skeptical about the prospects for localized acts of resistance against 'empire', believing that the struggle now is simply over the form globalization will take. Even David Harvey, whose work delineates the historical continuities (as well as transformations) in relations between state sovereignty and global capitalism and thus draws extensively upon Marxist understandings of the imperialism of nineteenth and twentieth century European colonialism, distinguishes in the very title of his book between such 'old' imperialism and a 'new' imperialism specific to the contemporary era. This distinction is intended to highlight the specificity of contemporary world politics and imply that while resonances of imperial forms might be discernable, today's empire is different from yesterday's. 'Empire' (or the 'new imperialism') is thus intended to indicate that the emergent contemporary world order is a progression to a new form of order different from that of the classical transcontinental or nineteenth /twentieth century European colonial empires, from the Westphalian states system, and from the post Cold War interregnum of globalization (of capitalism and governance).

David Harvey's examination of the place of American hegemony in a historical continuum of imperial forms draws upon the historical resonances of 'empire' in order to ask about the nature of contemporary world politics. Harvey starts from the assumption that American hegemony is a variety of 'capitalist imperialism', a contradictory fusion of 'the politics of state and empire'. Imperialism is here understood as a distinctively political project on the part of actors whose power is based in command of a territory and capacity to mobilize its human and natural resources toward political, economic, and military ends. For Harvey, imperialism is also a diffuse political–economic process in space and time in which command over and use of capital takes primacy. American power, Harvey argues, needs to be understood in this nexus of territorial command and capital accumulation enabling an exploration of the manner in which American foreign policy seeks to maintain hegemony over the wider global dynamics of capitalist accumulation. In this sense, Harvey's problematic has some similarities with the enquiries into the colonial policies of European powers in the nineteenth century that were similarly rooted in the relationship between state power and capital accumulation. Harvey argues that hegemony is predicated upon the three pillars of money, productive capacity, and military might. Imperialism is thus the deployment of these resources in order to achieve a hegemonic stabilization and conservation of the advantages that accrue from a given set of unequal relationships of exchange. Harvey charts the development of the new imperialism, or American hegemony, through identifiable historical transformations of the deployment of these resources since 1870.

While Harvey's position is firmly located in the Marxian tradition of political economy (indeed, Harvey notes that the so called 'new imperialism' is similar to the old imperialism described by Lenin and others) other observers have relied much less directly on Marxian thought to make related points about the contemporary significance of empire. Michael Hardt and Anthonio Negri, for example, argue that although nation state based systems of power are rapidly unraveling in the face of global capitalism, globalization cannot be understood as a simple process of de regulating markets. Far from withering away, regulations today proliferate and interlock to form a multicentered supranational order which they choose to call 'empire'. The term, as they use it, refers not to a system in which tribute flows from peripheries to great capital cities, but to a more Foucauldian figure – a diffuse, anonymous network of all englobing power. Hardt and Negri claim that the sinews of this diffuse polity – its flows of people, information, and wealth – are simply too unruly to be monitored from metropolitan control centers. Their account suggests that an older, statist world (of ruling class and proletariat, and of dominant core and subject periphery) is breaking down and being replaced by a less dichotomous and more intricate pattern of inequality. The logic of this volatile and diffuse entity evades, blurs, and transgresses all the inherited divisions of political thought: state and society, war and peace, control and freedom, and core and periphery. More than just an important historical episode, the advent of this 'empire' has important consequences for our conceptions of reality and is an event of significant ontological importance.

Other authors, such as Michael Mann, have also focused on the diffusion and incoherence of contemporary forms of empire not by examining politics or economics but by examining the recent history of American military power. While America cannot be said to have universal dominance, it can be said to have global reach and is a military force equipped only for battlefield superiority and not for occupation or colonial garrison. While America is thus able to cajole and put pressure on other states, the implication is that it will struggle to remake the world order in its preferred image. In turning to the economic power of the US, Mann notes that America is a reluctant participant in the multilateralism necessary to keep a globalized economy running. This is similar in some ways to Harvey's argument that American foreign policy seeks to maintain hegemony over the wider global dynamics of capitalist accumulation. Without engaging in multilateralism the US is unable to dominate the global economy and can only steer it with limited success. Moreover, while America brings an ideology of free trade to the global economy, it practices the protection of American interests. Mann thus argues that America's political power is 'schizophrenic' in that American foreign policy oscillates between multilateralism and unilateralism. The US perceives itself to be facing possible threats from either a coalition of states or a single state and is uncertain whether the best approach is multilateral engagement or unilateral invasion. Given the incoherence of American power and its reliance on military dominance, contrary to Harvey, Mann argues that the US does not have imperial political powers and that the 'age of empire' has gone while America's imperial efforts are not hegemonic. It is important, however, to avoid narrow conceptions of the international which focus solely on political–military and economic interactions and overlook the thicker sets of social relations and sociocultural flows.

Conclusions: Imperialisms 'Old' and 'New'

The literature on the new imperialism/empire is an important attempt to come to terms with the manner in which political forms are changing in the early twenty first century. Whether understood as the latest stage in capitalist spatiotemporal expansion or the global extension of military dominance, many authors take 'empire' to be 'the' dominant, determining motif of our times: a central political form of the contemporary world order. Derek Gregory's identification of the 'colonial present', for example, examines the 'war on terror' in Afghanistan and its extensions into Palestine and Iraq and argues that this must be understood in the historical context of British and American involvement in the Middle East, where colonial power continues to cast long shadows over our own present. For Gregory, 9/11 activated a series of political and cultural responses that were profoundly colonial in nature. This kind of work leads us to ask which historical echoes and resonances of past imperialisms shape the contemporary use of the term 'empire'. More specifically, what core attributes are seen to occur in any instance of contemporary world politics to which the term 'imperial' can be applied? Beyond this, what is the novelty, uniqueness, and specificity of these contemporary uses of the term 'empire'? Since empire is deployed to speak about a transformation in world order, what attributes are said to inhere in the world order of the contemporary era that legitimizes references to a 'new' form of imperial power?

The literature on the new imperialism often legitimizes its use of the term 'empire' by noting a range of similarities between the expansive characteristics and extensive reach of classical transcontinental or European colonial empires and the global aspirations of American power in the early twenty first century. Since empire or imperialism indicates an expansion beyond the borders of the state of political influence or control some authors see continuities and similarities between European imperial attempts to open markets around the world and attempts by the American state to maintain a global hegemony of neoliberal capitalism. Others have noted the manner in which the achievement of global military dominance encourages the US to deploy military force well beyond its borders. The dynamic of subordination and exploitation that characterizes the expansion of political power in the contemporary era also has its own historical echoes of past imperialism in that there is arguably a similar dynamic of 'hierarchization'. Some authors have thus argued that 'hierarchization' is a key logic of empire in the colonial present where the expansion of political power establishes dynamics of inequality. Thus authors such as Harvey have argued that the geopolitics of empire comprises a geography of uneven development and unequal relations of exchange and thus establishes a hierarchical geography of accumulation.

It is less clear to see, however, where some authors draw a distinction between old and new imperialisms while the specificity of contemporary empire is often even harder to detect. It is thus not always clear how the form of expansion, hierarchization, or capital accumulation in the contemporary world order is radically different to those found in analyses of classical Roman or modern European colonial empires. Since analyses of the Roman Empire, understandings of imperial power have taken crisis and decline to be central. Empires are said to rise and fall, overstretch themselves, and collapse. The concept of crisis has also been important in the literature on new imperialism/empire. For Harvey, America is a hegemon struggling to maintain its control over global capitalism. More specifically, the invasion of Iraq must be read as an attempt to maintain control over 'the global oil spigot' of the Middle East in order to preserve US hegemony. America, it is argued, thus faces a crisis in its imperial leadership, one that it seeks to postpone by increasingly desperate foreign interventions. Moreover, Harvey argues that this crisis is not simply one of hegemony over the global conditions of capitalism, but also an internal crisis within American civil society. American invasions of Afghanistan and Iraq are thus read as responses to 'internal' crises of civil society. Others such as Michael Mann have argued that the contemporary world order is in a period of crisis in which American militarism threatens peace and order. Moreover, this crisis results from the weakened imperial state of the USA that, in its global militaristic dominance and unilateralism, threatens to overstretch its hegemonic position and plunge the contemporary imperial form into decline. In arguing that the 'age of empire' has gone, Mann suggests that an anarchic order of nation states is reasserting its primacy.

While the Westphalian states system has always had pretensions to timelessness, empires have always been time bound, rising, and falling. In this sense, when looking for ways in which to demonstrate that the contemporary world order could be said to represent a transformation of the mythical Westphalian principles of interstate order, the figure or idiom of empire is a natural choice. It is very difficult to conceive of change in the Westphalian order and so those scholars who continue to be preoccupied with an older, statist world (of ruling class and proletariat, and of dominant core and subject periphery) talk of a resumption of business as usual with unilateralist statism sweeping away the temporary hiatus of the post Cold War enthusiasm for globalization. Contrary to this timeless order of statism, the idiom of empire is one that is rooted in time and always already asks how an imperial power arose and whether it is in danger of decline. The introduction of the concept of 'empire' is therefore intended to suggest that, at least in part, world politics cannot be read through the constraints of the Westphalian interstate system. For some, retrieving the imperial offers a way out of the 'territorial trap' set by Westphalia and alerts us to a range of phenomena occluded by the use of our central categories in geography, international relations, and so on.

Thus, some authors have begun to read the concern with the expansion of political control as an assertion of the primacy of transnational interconnections over international relations, foregrounding the primacy of connections (which transgress established borders) over those relations that are predicated upon preserving such borders. Using the concept of empire to chart the expansion of political power suggests that borders are naturally permeable and permeated and that sovereignty is neither contained in its territorial bounds, nor inviolable. This expansion of power in the 'new imperialism' is said to involve the emergence of deterritorialized networks of global governance that relay norms and regulative ideals throughout global space via institutions lacking territorial coordinates (e.g., the international financial institutions). While all forms of empire have had regimes of government, today's regime of governance is not a territorial incorporation of the periphery into the circuits of capital emanating from and returning to the center, but rather the extension of diffuse, nonterritorial circuits of regulation embodied in intergovernmental organizations, civil society, non governmental organizations and transnational corporations. The circulation of notions of human rights, good governance, competition, and free trade, for example, establishes a global empire of individualist (neo)liberalism. This is a diffuse global hegemony in an empire that aims to appear universal and natural while disguising its imperial nature. Thus the 'new imperialism' or 'empire' of the contemporary world order is seen as wholly other than that of the period of classical imperialism because of the dissimilarity between the territorial basis of previous historical empires and the global ambitions of the deterritorialized empire of the contemporary world. Evidence for such antiterritorial imperial dynamics is said to be found in the evolution of a hegemonic global liberal governance regime.

Whereas nineteenth and twentieth century colonialism brought the values of the colonizer from the metropolitan heartlands of the empire to the colonized at its periphery, there is, however, a distinct difficulty in establishing the peripheries of these 'new' contemporary forms of empire. The global hegemony of (neo)liberalism is achieved not through imposition, but through the manner in which these values come to be considered 'norms' and therefore become the global horizon of political discourse. Thus, supposedly peripheral governments openly embrace ideas of good governance or human rights not because they are being incorporated as 'colonies' but because they are becoming yet another relay in the global circuit of liberal governance. Thus, the use of the term is justified by noting that expansion was a primary feature of previous empires, but the specificity of contemporary empire is noted by arguing that the new imperialism is a nonterritorial, globalizing regime of governance, in which the traditional imperial separation between core and periphery, colonizer and colonized, is rendered indistinct.

Whereas classical understandings of the hierarchization entailed by imperialism viewed the unequal relationships as ones in which the center extracted values from the periphery, the literature of the new imperialism/empire takes empire to be the naturalization of the hegemony of a universal (neo)liberal regime. In this sense, while accumulation often flows from the Global South to the Global North, this is not clearly the same as the flow from center to periphery. Following Hardt and Negri, we could argue that while the 'old' imperialism rested on the blood and tears of conquest and extraction, the 'new' imperial power is characterized by the hegemony of a naturalized set of putatively universal values. The rise of financial capital and the emergence of flexible production, for example, are said to be transforming the logics of state away from preservation of territorial integrity (as per the logic of Westphalia) toward engagement in multilateral institutions of governance and transnational networks of production and accumulation. Such a change leads to the progressive deterritorialization of production and accumulation and hence the dissolution of the classic colonial practices that characterized European imperialism. The state is no longer the tool of acquisition that it was assumed to be by theories of 'old' imperialism. Rather the state, through its multilateral engagement with organs of global governance, establishes the conditions that permit transnational corporations to roam without restriction across global space. For Hardt and Negri, the national colors of the imperialist map of the world have merged and blended into an 'imperial global rainbow'.

Despite this supposed merging and blending of national colors the global role of America is a central thread running through the tapestry of this literature on the new imperialism/empire. This preoccupation with America is also one of the ways in which this literature claims to have identified the specificity of the present world order. Some have argued that since 1945 the US global empire has assumed two forms: the empire of production, from World War II to the 1970s, and the empire of consumption, from the 1970s to the present. Others have pointed to the twentieth century 'empire of the dollar' or point to the US oil interests in Iraq and the emergence of 'petroimperialism'. Other contemporary authors however, like Hardt and Negri, argue that US imperialism is not a central force in the world today in that the US cannot form the center of an imperialist project since imperialism is over and no nation will be world leader in the way modern European nations were. This passage to a new global constitutional regime, it is argued, is shown by the Gulf War, during which the United States emerged as the only power able to manage international justice, not as a function of its own national motives but in the name of global right. In setting itself up as some kind of 'world police' the United States is thus argued to act in the interests of a deterritorialized 'empire'. In this sense the Gulf War did indeed, as George Bush claimed, announce the birth of a New World Order.

Thus while the 'new' empire might resemble the old in many ways, it differs insofar as it is a specifically American empire. America has thus taken its turn as the imperial power. The new empire is an American empire in peril however: either of overstretch and collapse, of decadent decline, or in danger of plunging the rest of the world into a 'global civil war'. The crisis at the heart of the 'new' empire represents the return of history and an urgent call to consider the emergent features of the contemporary world order. The literature on the new imperialism/empire thus offers a significant opportunity to rethink the nature of this world order and the categories and concepts we draw upon to make sense of it. We must take seriously, however, the 'postimperial' character of contemporary American and Western power, and recognize a much wider range of contemporary 'quasi imperial' forms. Imperial power relations are even a common feature of many non Western states, considered 'Westphalian' nation states or 'postcolonial' states in some previous classifications of contemporary statehood.

It seems incumbent upon scholars of world politics then to address the specificity of the imperial form in the early twenty first century and grasp the possibilities for a nonstate centric analysis that it offers. This poses important challenges for geographers in particular who are well placed to contribute to these debates and to the formulation of concepts and theoretical perspectives. We need to keep in mind, however, the versatile power(s) of geography in the construction of empires and imperial hegemony, past and present. It is also important for 'geography' as a discipline to come to terms with its own history of engagement with European colonial empires. The challenge of decolonizing geography is thus important and ongoing. Much recent work in geography, focused on the tensions of empire today, is imbued with the perspectives of postmodernism and postcolonialism and these approaches offer important ways forward. This work pays close attention to the Eurocentric dimensions of empire and imperialism and the ambivalent nature of colonial discourses. Many critics have argued that while colonial empires have been dissolved, global relations are still structured by imperial attitudes and the affairs of postcolonial societies are still shaped by 'Western' frameworks of knowledge. This work has focused debates around the nature of 'otherness', the legacies of colonialism, and how the West continues to engage the world in imperial terms. It also focuses attention on the ways in which imperial and colonial projects were subverted. In order to address the specificity of the imperial form in the early twenty first century and to grasp the possibilities for new forms of analysis that this offers, geographers must come to terms with how the discipline is still shaped by imperial attitudes, by 'Western' frameworks of knowledge and by the postcolonial legacies of imperial conceptions of difference and otherness.

See also: Capitalism; Cold War; Colonialism I; Colonialism II; Imperial Cities; Imperialistic Geographies; Neocolonialism; Postcolonialism/Postcolonial Geographies.

Further Reading

Barkawi, T. and Laffey, M. (2002). Retrieving the imperial: Empire and international relations. *Millennium: Journal of International Studies* 31(1), 109 127.

Brewer, A. (1990). *Marxist Theories of Imperialism: A Critical Survey* (2nd edn). London: Routledge.

Callinicos, A. (2002). The actuality of imperialism. *Millennium: Journal of International Studies* 31, 319 326.

Corbridge, S. (2003). Countering empire. *Antipode* 35, 184 190.

Cox, M. (2004). Empire, imperialism and the Bush doctrine. *Review of International Studies* 30, 585 608.

Driver, D. (2001). *Geography Militant: Cultures of Exploration and Empire*. Oxford: Oxford University Press.
Ferguson, N. (2004). *Colossus: The Rise and Fall of the American Empire*. London: Allen Lane.
Gregory, D. (2004). *The Colonial Present: Afghanistan, Palestine, Iraq*. Oxford: Blackwell.
Hardt, M. and Negri, A. (2000). *Empire*. London: Harvard University Press.
Hardt, M. and Negri, A. (2004). *Multitude: War and Democracy in the Age of Empire*. New York: Penguin Putman.
Harvey, D. (2003). *The New Imperialism*. Oxford: Oxford University Press.
Howe, S. (2002). *Empire: A Very Short Introduction*. Oxford: Oxford University Press.
Ignatieff, M. (2003). *Empire Lite: Nation Building in Bosnia, Kosovo and Afghanistan*. London: Vintage.
Joxe, A. (2002). *The Empire of Disorder*. New York: Semiotex(e).
Kirsch, S. (2003). Guest editorial: Empire and the Bush doctrine. *Environment and Planning D: Society and Space* 21, 1 6.
Mann, M. (2003). *Incoherent Empire*. London: Verso.
Raman, B. (2003). De centering Empire. *Antipode* 35, 194 196.
Robinson, J. (2003). Postcolonising geography: Tactics and pitfalls. *Singapore Journal of Tropical Geography* 23, 273 289.
Shaw, M. (2002). Post imperial and quasi imperial: State and empire in the global era. *Millennium: Journal of International Studies* 31, 327 336.
Smith, N. and Godlewska, A. (eds.) (1994). *Geography and Empire*. London: Blackwell.
Walker, R. (2002). On the immanence/imminence of empire. *Millennium: Journal of International Studies* 31, 337 349.

Relevant Websites

http://www.infoshop.org
An online copy of Hardt and Negri's *Empire* (2000).
http://www.bbc.co.uk
BBC resources on the British Empire.
http://www.scholars.nus.edu.sg
Resources concerning post colonialism.
http://www.empiremuseum.org
The British Commonwealth and Empire museum.
http://www.globalissues.org
Useful resources on geopolitics and US imperialism.

Empowerment

R. Scheyvens, Massey University, Palmerston North, New Zealand

Introduction

"We not only want a piece of the pie, we also want to choose the flavour, and to know how to make it our selves," says Ela Bhatt, founder of the Self Employed Women's Association of Ahmedabad, 1992.

Discourse on empowerment has had an important influence on debates about development especially since the early 1990s. Along with similar catchy terms including 'sustainable development' and 'participation', 'empowerment' has been captured by, and incorporated into, the rhetoric of a wide range of development agencies with diverse social and political aims.

While, at first, empowerment was associated with alternative development approaches endorsing grassroots participation and local level development, it is now applied more broadly to attempts to overcome global inequalities at a wide range of scales. For the poor, the disadvantaged, and the disenfranchised, empowerment implies a significant challenge to current structures of power. Empowerment can thus be seen as essential for any group which wishes to challenge systems of oppression which have curtailed their life opportunities. From a geographical perspective, it is interesting to observe that claims for space – whether demands for recognition of land rights by indigenous groups or the importance for women in having a separate space in which to meet – are often closely associated with empowerment processes.

This article discusses how empowerment can be defined, different approaches to empowerment, and highlights a range of ways in which the concept has been employed.

As will be shown, the concept of empowerment has been utilized particularly by development geographers (especially those interested in gender or indigenous development), feminist geographers, geographers studying other less powerful groups (including the disabled and children), and those advocating participatory or inclusive research methodologies.

Definitions

Empowerment means activation of the confidence and capabilities of previously disadvantaged or disenfranchised individuals or groups so that they can exert greater control over their lives, mobilize resources to meet their needs, and work to achieve social justice.

Before we can fully understand empowerment, however, we need to be clear on the meaning of its source term, 'power'. Power is often seen as a binary, hierarchical concept whereby it is synonymous with domination. This is power 'over' and it is instrumental in that power is understood as a means of influencing others to one's own advantage. However, development geographers generally prefer to see power as an enabling, productive force. This can include power 'to' – generative forms of power which can include resistance; power 'with' – the power of collective action; and power 'from within' – a sense of self belief which leads to confidence to seek changes which could improve one's life. With these latter understandings of power, it becomes clear that a development agency cannot actually 'empower' someone who is oppressed, marginalized, or otherwise disadvantaged; however, it can 'facilitate' empowerment.

There are concerns that, like power, 'empowerment' is often used in a rather instrumental sense as a means by which disadvantaged peoples can be more effectively harnessed to achieve certain organizational or institutional goals set from above. Thus, the notion that empowerment aims to enhance the well being of individuals or communities, as is well understood within development discourse, is at odds with the use of the term in the business world where empowerment is associated with handing over greater responsibility to workers who will then be more efficient, deliver enhanced customer service, or otherwise benefit corporate interests.

Consciousness Raising and Empowerment

Consciousness raising is regarded as an important first step in facilitating empowerment of disadvantaged peoples. If power is to come from within and if people are to realize what they are capable of – something referred to as psychological power – there needs to be a strong awareness of the political and social context of one's oppression, so that people feel that they deserve to have access to resources and to be respected. They also need to know how they can use existing economic, political, and bureaucratic systems to improve their lives. Consciousness of their situation and a sense of self belief can determine, for example, whether people from minority ethnic groups will stand up to outside imposed development programs such as mines and dams, or whether adults who have had no formal schooling will consider enrolling in literacy classes, and whether any marginal ized persons will put themselves forward for election to decision making bodies to ensure that their voices are heard at community, regional, or national level.

Schuler, reflecting on possibilities for women's de velopment, devised a model of levels of empowerment which suggests that consciousness raising is an important first step in moving people toward collective action and the achievement of political power (**Figure 1**).

As an example, the gender and development program of a Maasai run NGO named CORDS (Community Research and Development Services) carries out aware ness raising and training among 14 Maasai women's groups in Northern Tanzania. In 2006, in Kiteto District, this training inspired women to come together in an at tempt to evict illegal settlers who were planting crops on what was previously cattle grazing land. There was a suspicion that male leaders in the area, who made no attempt to challenge the settlers, had received bribes from them. The collusion of their own leaders made it very difficult for other Maasai men to act to have the settlers removed. The women were not willing to be complacent, especially, when they saw that Maasai chil dren were regularly being beaten by the settlers when their livestock roamed onto newly cultivated land. Armed with new knowledge about land rights and women's rights from training provided by CORDS, these women decided to take action, first knocking on the doors of district officials to make complaints. When this had no effect, they decided to send a delegation of one man and one woman to Dodoma to talk to the Prime Minister. The Prime Minister saw them briefly, and later ordered a commission of enquiry into the land grabbing that was occurring. As a result: (1) the District Commissioner of Kiteto was sacked, and (2) the illegal settlers were issued with eviction letters. While they are still trying to resist eviction, the women of Kiteto District are adamant that they will continue to fight to ensure that what they have left will pass on to their children. These women are very keen for the training from CORDS to continue:

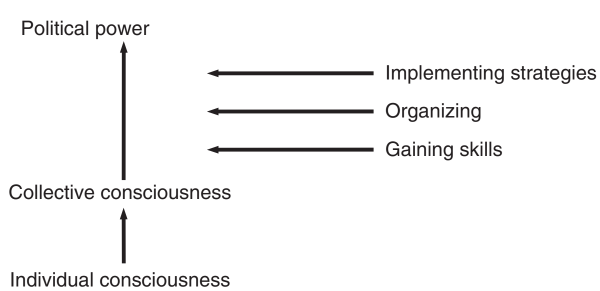

Figure 1 Building empowerment. From Schuler, M. (1986). *Empowerment and the law: Strategies of Third World women*. Washington, DC: OEF International.

"We want more training because that's how we get strength and knowledge. If we weren't trained, we would have stayed in our houses until the intruders came and burned them down," says a member of Irpopongi women's group.

Changing Approaches to Empowerment within Development Geography

Alternative Development

The notion of empowerment can be seen as implicit in social movements, struggles for democracy and in dependence that date back hundreds of years; however, its use in development quarters is more recent. While advocates of women's development and others embraced the concept during the 1980s, empowerment did not enter the mainstream of development discourse until after the publication of John Friedmann's book, in 1992, entitled *Empowerment: The politics of alternative development*. Friedmann understood empowerment to be a complex phenomenon involving psychological, social, and polit ical dimensions. He focused on the level of the household as the locus for an alternative approach to development, suggesting that development could best be achieved by a self empowered group of citizens mobilizing resources.

Friedmann's work inspired supporters of alternative development who espoused bottom up, participatory forms of development. They embraced empowerment as it was seen as central to possibilities for the development of oppressed, disenfranchised peoples and, through its recognition of politics, for self determination. It is thus not surprising to find much talk of empowerment among those interested in the rights and knowledge of in digenous peoples, such as those who were concerned that Eurocentric processes of development were stifling in digenous knowledge and initiative. Empowerment was also taken up in the gender and development field, es pecially after an article by Caroline Moser appeared in *World Development* in 1989 noting that empowerment could be seen as the most advanced of five approaches to women and development. As opposed to other ap proaches which tended to focus on 'practical gender needs' associated with immediate concerns such as the need for food and shelter, an empowerment approach recognized that it was also vital to address 'strategic gender needs' which were more long term in nature, such as women's rights to land, control over their bodies, and encouraging men to share domestic tasks.

In addition, geographers inspired by alternative de velopment approaches have also utilized the concept of empowerment as an analytical tool, as shown in **Table 1** which specifies four dimensions of empowerment (eco nomic, social, psychological, and political). This concept attempts to show what empowerment can mean for host

Table 1 Framework for assessing extent of empowerment of communities involved in tourism

	Signs of empowerment	*Signs of disempowerment*
Economic empowerment	Tourism brings lasting economic gains to a local community. Cash earned is shared between many households in the community. There are visible signs of improvements from the cash that is earned (e.g., houses are made of more permanent materials; more children are able to attend school).	Tourism merely results in small, spasmodic cash gains for a local community. Most profits go to local elites, outside operators, government agencies, etc. Only a few individuals or families gain direct financial benefits from tourism, while others cannot find a way to share in these economic benefits because they lack capital, experience, and/or appropriate skills.
Psychological empowerment	Self-esteem of many community members is enhanced because of outside recognition of the uniqueness and value of their culture, their natural resources, and their traditional knowledge. Access to employment and cash leads to an increase in status for traditionally low-status sectors of society, for example, youth, the poor.	Those who interact with tourists are left feeling as if their culture and way of life are inferior. Many people do not share in the benefits of tourism and are thus confused, frustrated, disinterested, or disillusioned, with the initiative.
Social empowerment	Tourism maintains or enhances the local community's equilibrium. Community cohesion is improved as individuals and families work together to build a successful tourism venture. Some funds raised are used for community development purposes, for example, to build schools or improve water supplies.	Disharmony and social decay. Many in the community take on outside values and lose respect for traditional culture and for their elders. Disadvantaged groups (e.g., women) bear the brunt of problems associated with the tourism initiative and fail to share equitably in its benefits. Rather than cooperating, families/ethnic or socio-economic groups compete with each other for the perceived benefits of tourism. Resentment and jealousy are commonplace.
Political empowerment	The community's political structure fairly represents the needs and interests of all community groups. Agencies initiating or implementing the tourism venture seek out the opinions of a variety of community groups (including special interest groups of women, youth, and other socially disadvantaged groups) and provide opportunities for them to be represented on decision-making bodies, for example, the wildlife park board or the regional tourism association.	The community has an autocratic and/or self-interested leadership. Agencies initiating or implementing the tourism venture fail to involve the local community in decision making so the majority of community members feel they have little or no say over 'whether' the tourism initiative operates or 'the way' in which it operates.

Source: Scheyvens, R. (2002). *Tourism for development: Empowering communities*. Harlow: Prentice Hall.

communities involved in tourism as well as how dis empowerment of host communities may manifest itself.

Neoliberalism

After its uptake by devotees of alternative development, such as members of nongovernmental and community based organizations, rhetoric on empowerment started appearing in documents produced by governments and by bilateral and multilateral agencies which were more likely to approach development from a neoliberal perspective.

A neoliberal view of empowerment is instrumentalist, focusing on enhancing the ability of the poor to meet their own needs and thus reducing their dependence on welfare or state services. This is how the term has often been used to justify reductions in state spending on social services or to utilize the poor as free or cheap labor in terms of service delivery. For example, the post apartheid state in South Africa in some cases exploited the labor of poor township residents in its waste collection schemes, using the rhetoric of black empowerment and gender empowerment.

A neoliberal approach to empowerment has also been the basis of programs to develop small scale entre preneurialism among the poor and thus overcome eco nomic and social inequalities between regions. An example of this was the revitalization program for distressed urban and rural areas in the United States devised in 1993 under the Clinton administration. The Empowerment Zone/Enterprise Community program stood out from other government attempts to uplift social and economic well being in impoverished areas because communities had to compete to join the program, and their acceptance was based on the quality of a strategic plan drawn up by community groups. In the strategic plan, communities were to assess the assets available to them, note constraints to development in the area, give their vision of a better future, and show how this vision could be achieved. The program was thus supposed to be driven by community members at the grassroots and to build upon existing resources. In 1994, 69 urban com munities joined the program and a further 20 were added in 1999. All were provided with federal grants, and tax benefits were also provided to local businesses which engaged in initiatives to meet objectives of the strategic plan. This attempt to engage private sector investors in initiatives to improve the well being of the poor also made the program unique, as this was a clear departure from approaches which suggested that poverty alleviation was a role only for the public sector.

While politicians claimed that these zones empowered anyone who was willing to work hard to improve their lives, there were also critics of the scheme. Such critics argued that the rhetoric of bottom up empowerment was undermined by both involvement of already powerful individuals in the planning phase, and administration of the funding from above, with strings attached. In some cases, therefore, it has been suggested that the programs have had only a marginal impact upon uplifting poorer people.

There has also been broader criticism of the phil osophy behind a neoliberal approach to empowerment because while it may devolve greater resources and rights to individual citizens or community groups, it also charges them with great responsibility for the planning and delivery of services, thus releasing the state of some of its obligations toward the welfare of citizens.

Post-Development

The influence of the cultural turn in the social sciences led to the development of interesting 'new' geographies during the 1990s, and use of new tools of analysis. Al ternative development was not seen as offering a sub stantial challenge to mainstream development; thus, concepts like empowerment lost favor. Instead, many geographers chose to focus on different types of in vestigations utilizing the likes of discourse analysis and deconstruction of texts. While such tools were valuable in adding to critiques of development concepts, categories, and institutions, thus providing fuel for an antidevelop ment perspective, they often presented a rather narrow view of development issues and they could not hope to reveal multiple sites of resistance to development from people around the world. In addition, deconstruction can be seen as offering little hope to those seeking guidance on positive ways forward. A focus on critique can be dismissive of the concerns of many Third World peoples who, while not wishing for externally imposed and controlled development, do want to see infrastructural improvements, and have access to more economic de velopment opportunities and social programs.

While empowerment was out of favor with some academics through the 1990s, it has returned to favor under the guise of a post development approach. Stimulated by the cultural turn, post development wri ters have criticized development binaries (such as tradi tional–modern), challenged the universalisms of grand theories of development which fail to account for social, economic, and geographical diversity, and argued for greater recognition of social diversity. However, they have also taken up the challenge to the 'new' geographies to move beyond critiques of hegemonic forces and in equitable structures and to find ways of transforming these structures so that the poor/oppressed or otherwise disenfranchised may benefit. Not all of the new geog raphies have been as receptive, however, to the notion of a moral imperative for their work to make a contribution to society.

Empowerment sits comfortably with many of the new, more enabling ideas of development endorsed under a post development approach, including post development interpretations of social capital, rights based approaches, and capabilities.

Contrasting These Approaches to Empowerment

While some authors argue that there is now considerable shared ground between different approaches to devel opment and empowerment, there are also some funda mental differences that remain. A neoliberal perspective on empowerment does not seek to change the existing order, and it is likely to see disadvantaged peoples as a resource to meet development agendas determined from above. However, contrary to a strong neoliberal position which advocated the pulling back of the state, there is growing emphasis on the roles states should play in empowering the poor. Post development theorists, meanwhile, share with alternative development advocates a more explicit socio political view of empowerment, rather than an economic view. Under a socio political view, empowerment entails a challenge to existing structures of society in order to challenge inequitable power relations between governments and citizens, men and women, dominant and minority groups.

Thus, those supporting empowerment from a post development perspective generally are somewhat aligned with proponents of alternative development, while critiquing neoliberal perspectives on empowerment. However, post development advocates have also raised concerns about the obsession with the local within alternative development discourse, whereby the local level is highlighted as the site of empowerment. In a related point, proponents of alternative development have been criticized for making assumptions that local communities are homogenous entities whose members have shared values, interests, and face similar disadvantages. Post development advocates stress that communities are diverse; thus, there are often significant power relations and inequalities evident at the local scale which may provide an impediment to participation and empowerment in practice. In addition, glorification of the local scale in alternative development forums has meant that the significance of oppressive national and international political and economic structures has often been overlooked. This focus on the local is illogical to geographers who recognize that social, economic, and political processes at work at a wide variety of scales set the context within which local development does, or does not, occur.

While most of the discussion above has shown the influence of these different approaches on development geography, it is important to note that empowerment has also been utilized by geographers outside of the development field, as the following section explains.

Empowerment Applied in Other Areas of Human Geography

Geographies of Less-Powerful Groups

Empowerment has influenced the thinking of a range of human geographers, as is particularly evident in research on geographies of childhood, ethnic minorities, the poor, geographies of (dis)ability, and in feminist geographies. There has been increasing advancement of the idea that members of less powerful groups are social agents, who should play a central role in planning their own lives, rather than beneficiaries requiring handouts from the state. Furthermore, efforts have been made to ensure that the voices of less powerful groups are heard regarding planning and resource distribution issues.

Technologies for Empowerment

Geography as a discipline has expanded with the development of technology, as is particularly obvious with applications such as remote sensing and geographic information systems. While some human geographers have regarded such change with suspicion, due perhaps to its association with the belief that quantitative methods are more rigorous than qualitative methods of enquiry, many have seized the potential of these new technologies (see the section titled 'Community mapping').

There are also bodies of research within geography which explore ways in which commonplace technologies can have liberating, or empowering, potential for certain groups of people. Researchers have studied topics such as the relationship between cell phones and freedom of movement of teenagers, and the spread and use of the Internet, especially in terms of the access of less privileged groups of society to this supposedly global information technology. Debates still rage concerning whether such technologies are (1) accessible, (2) relevant, and (3) cost effective, as means of improving the lives of disenfranchised groups.

Community Mapping

In times past, map production was the domain of specialist cartographers and to some extent geographical information systems (GIS) technologies have continued to keep map making as a specialist activity. However, in recent years, there has been growing recognition that maps do not always have to represent interests of an elite, rather, what Peluso refers to as 'counter mapping' can be conducted by communities searching for social justice and seeking to claim resources. The empowering potential of community mapping can be realized either by using technology like GIS or by using purposefully low technology techniques associated with participatory rural appraisal (PRA) or participatory learning and action (PLA). Here people may gather together to draw up a social or resource map using minimal inputs – maps have sometimes been drawn in the dirt using sticks and stones as markers, and more commonly on a large sheet of paper.

A community map typically represents a community's efforts to document places and resources which are important to them. They can then use this information to determine ways of responding to threats to their environment. Thus, maps can be seen as rhetorical devices which enable communities to articulate their interests or plans, rather than to present objective, neutral representations of facts. For this reason, community maps are often used by indigenous groups as the basis of land rights claims in order to protect land that their people have resided in for generations, but to which they have no legal title. Similarly, an urban community may identify that there has been significant erosion in public park space as their city has grown and this may provoke initiatives to lobby their local authorities to try to limit future incursions on this space.

Feminist and Participatory Methodologies

In addition, empowerment has become central to certain research philosophies and methodologies which regard research participants as far more than the subjects of the research. Researchers employing feminist methodologies, for example, have been keen to ensure that geographical knowledge is produced in ethical ways by working with, not on, their research participants. In this way homeless women, or women who have experienced violence, can come to be seen as partners in the research exercise. They may be actively involved in research design, data collection, and analysis, and there may be attempts to ensure they share ownership of the research results. Similarly, a number of geographers have studied topics like health, housing, or impairment in ways which intentionally develop the skills of research participants and increase their knowledge. Such research is referred to by Gill Valentine as 'enabling geography'.

However, geographers with a commitment to engaging in an empowering research process should not underestimate ways in which institutional pressures can erode their good intentions. First, access to research participants often requires the permission of an intermediary agency, and if this agency is hierarchical in nature, they may oppose participatory research because they feel threatened by the idea that the research process will raise awareness or empower participants. Second, university funding agencies often display little appreciation of the time and resources required to enact participatory, empowering methodologies. Third, it can be difficult to truly share ownership of the research results in practice. Despite plans by researchers to share their findings with participants and allow them to scrutinize transcripts before their voices are turned into quotations in an article or report, the expectations of the academy may mean that such meritorious aspects of the research as set out in the design phase are pushed aside once that geographer is sitting back in their office amidst the usual pressures to 'publish or perish'.

Empowerment through Claiming Space

Geographers have long been interested in the relationship between power and space. Space has liberating potential, as is evident in some of the examples given from both development geography and human geography more generally. Thus, it is through drawing maps that members of illegal settlements have claimed their space in the city – they have become visible. Similarly, indigenous groups often have a very real appreciation of the political nature of space; for them, displacement from traditional lands is equivalent to disempowerment. In order to resist domination, they may choose to occupy a space – be it a building or a section of contested land – as a central strategy in making their claims.

A number of gender and development writers have also shown how the space provided by women's groups can provide an important forum for the collective empowerment of women. These groups provide a safe space in which women may meet and discuss their burdens and their ideas, share skills and knowledge, learn more about their situation and strategies for change, and become involved in collective action. Thus even though women's groups may appear to be conservative, they have often been very effective in terms of providing a legitimate, socially acceptable reason for women to gather together regularly within their communities, and this has sometimes led to social actions which challenge their subordinate position in society.

Empowerment as a Process and a Goal

This article has examined a number of very interesting applications of the empowerment concept. While some writers draw a clear distinction between empowerment as a 'process' and empowerment as a 'goal' of development, in the examples discussed, the two are often closely related. To elaborate, advocates of community mapping suggest that this can be an empowering experience for participants in two ways. First, the process of coming together and engaging in collective action can enable people to discover that they have shared interests and aspirations with those living around them and that together, they can work to enact positive changes in their communities. Second, community mapping can lead to people gaining greater control over resources. In a related point, urban communities have often drawn the maps in response to attempts by government officials to evict them from illegal settlements, but the process of developing the maps has resulted in better communication between officials and occupants of illegal settlements, and sometimes this has led to programs to assist with upgrading of informal settlements.

However, the process leading to empowerment can be corrupted. There is always the risk that dominant individuals or groups will take control of supposedly empowering processes, or that more powerful individuals will fill decision making roles.

Conclusion

Empowerment is a crucial aspect of development for social groups which face disadvantages related to phenomena such as gender, ethnicity, class, age, and religion. For them, empowerment is best understood as part of a process to enable disadvantaged or marginalized peoples to play a stronger role in society, influence decision making, and access resources according to their self determined needs. When employed by development agencies, however, empowerment can be seen to represent a range of meanings. As with any popular, yet contested, development term, it is thus important that we are critical in our use of the term empowerment, and when analyzing the potential of empowerment to be a positive force for change, we need to consider carefully the approach being adopted, rather than just being seduced by the versatility of the concept.

While a diverse range of agencies and institutions have embraced empowerment rhetoric, the outcomes of empowerment programs have sometimes been disappointing, such as with the Clinton regime's Empowerment Zone/Enterprise Community program. For the disadvantaged to gain greater control over their lives, the agents of development, including multilateral institutions, governments, donors, nongovernmental organizations, and community based organizations, may need to change their *modus operandi* to see the disadvantaged more as partners who wish to direct their own development, rather than drawing on neoliberal logic which regards them as resources to be harnessed to achieve organizational goals set by outside agencies. Empowerment should not be reduced to a process for encouraging poorer groups to take on greater responsibility for meeting their own needs. Studies of the culture of powerful development institutions could help to illuminate whether or not they are structured in such a way as to support such a shift in thinking.

While the early advocates of empowerment spoke from an alternative development perspective and worked mainly with communities for bottom up empowerment, if empowerment of the poor and disadvantaged is to be central to development agendas, then it is vital that we look beyond the local scale. It is important not to overlook the significance of oppressive national and international political and economic structures which help to set the context within which local development occurs. More recent research on empowerment has shown the value of studies which connect local, regional, national, and international levels.

In conceptualizing empowerment, geographers can offer unique insights into the close relationship between space and power. Space is not apolitical; it represents power. This is why processes of empowerment are often strongly associated with claims to space, such as the demand for secure land tenure, or the right to speak in public forums. However, development geographers are also faced with important challenges. One challenge that stems directly from recognition of the value of empowerment is that academics need to move beyond critiques of the language, documents, institutions, and practice of development. Rather, we need to continue to rise to the challenge of ensuring that our research on development issues has legitimacy not just in terms of the academy, but in development policy making and practice as well.

See also: Eurocentrism; Feminism/Feminist Geography; Feminist Methodologies; GIS, Public Participation; Neoliberalism and Development; Participation; Postdevelopment; Resistance; Social Capital.

Further Reading

Allen, J. (2003). *Lost geographies of power*. Malden, MA: Blackwell Publishing.

Bebbington, A., Woolcock, M., Guggenheim, S. and Loson, E. (eds.) (2006). *The search for empowerment: Social capital as idea and practice at the World Bank*. West Hartford: Kumarian Press.

Brohman, J. (1995). Universalism, Eurocentrism, and ideological bias in development studies: From modernisation to neoliberalism. *Third World Quarterly* 16(1), 121 140.

Friedmann, J. (1992). *Empowerment: The politics of alternative development*. Cambridge, MA: Blackwell.

Hickey, S. and Mohan, G. (2005). Relocating participation within a radical politics of development. *Development and Change* 36(2), 237 262.

Kabeer, N. (1999). Resources, agency, achievements: Reflections on the measurement of women's empowerment. *Development and Change* 30, 435 464.

Miraftab, F. (2004). Making neo liberal governance: The disempowering work of empowerment. *International Planning Studies* 9(4), 239 259.

Mohan, G. and Stokke, K. (2000). Participatory development and empowerment: The dangers of localism. *Third World Quarterly* 21(2), 247 268.

Moser, C. O. (1989). Gender planning in the Third World: Meeting practical and strategic gender needs. *World Development* 17(1), 1799 1825.

Peluso, N. (1995). Whose woods are these? Counter mapping forest territories in Kalimantan, Indonesia. *Antipode* 27(4), 383 406.

Pricechalita, P. (1994). Spatial metaphor and the politics of empowerment mapping a place for feminism and postmodernism in geography. *Antipode* 26(3), 236 254.

Rowlands, J. (1997). *Questioning empowerment: Working with women in Honduras*. Oxford: Oxfam.

Scheyvens, R. (2002). *Tourism for development: Empowering communities*. Harlow: Prentice Hall.

Schuler, M. (ed.) (1986). *Empowerment and the law: Strategies of Third World women*. Washington, DC: OEF International.

Valentine, G. (2003). In pursuit of social justice: Ethics and emotions in geographies of health and disability. *Progress in Human Geography* 27, 375 380.

Relevant Websites

www.wateraid.org.uk
Manual by WaterAid, London on 'Community mapping: A tool for community organizing' (2005).

http://www.nativeweb.org
Native Web.

http://cerd.ph
The Center for Empowerment and Resource Development.

http://www.undp.org
Women's empowerment page of the United Nations Development Program.

Enlightenment Geography

C. W. J. Withers, University of Edinburgh, Edinburgh, UK

Glossary

Cosmopolitanism The view that intellectual enquiry and exchange was a shared enterprise and should take place unhindered by religion, rank, or geographical boundary.

Metrology The science and practice of measurement.

Republic of Letters A term used to describe the significance of correspondence and of literary production and exchange for the public good.

Science of Man A term used by contemporaries to signify the systematic observation and empirical examination of human moral and cultural traits.

Utilitarianism In its strict sense, the belief that knowledge should have an economic value; less rigidly, the view that knowledge itself is valuable.

Introduction

The Enlightenment is that period – by convention, the 'long' eighteenth century in Europe *c.* 1680 to *c.* 1815 – when the world was made 'modern'. For contemporaries then and for modern scholars now, the Enlightenment was both a historical moment and an intellectual movement distinguished by challenges to established authority, belief in the power of reason, and by new ways of critical thinking about the world. The Enlightenment is commonly considered in terms of its defining features and principal characters – its 'what' and its 'who'. Great attention has been paid to the origins and motivations of Enlightenment (its 'how' and its 'why') because they fascinate still, and to the Enlightenment's legacy (its 'so what') because of the enduring implications of claims to freedom and the ideals of progress. In these terms, the Enlightenment is most often treated as a historical phenomenon. Yet the Enlightenment and its ideas, personnel, and artifacts were locatable as well as datable: in several important ways, the Enlightenment had a 'where'.

This article takes Enlightenment geography to mean the geographical dimensions of the Enlightenment. Drawing upon a range of evidence and against a summary background of how the Enlightenment has been approached, this topic is explored in three related sections. In the first, attention is paid to the geographies of the Enlightenment, to the Enlightenment's geographical expression at different scales, nationally, transnationally, and locally. The second outlines the ways in which the world was geographically revealed and understood as an object of geography in the Enlightenment – through exploring, for example, in mapping and in travel narratives. In the third, the focus is on the subject of geography in the Enlightenment, with how it was understood and employed in a variety of ways and social spaces. Laid out thus – by considering the geography of, geography and, and geography in the Enlightenment – this essay moves from an understanding of the Enlightenment as a geographical question evident at different scales and places, to the impact of geographical ideas upon Enlightenment thought, and to geography as an Enlightenment subject.

The Enlightenment: Questions of Definition, Matters of Geography

Eighteenth century commentators concerned with what has become known as the Enlightenment focused on the extent to which public affairs should be governed by reason, with the freedom of press, with science and philosophical enquiry as useful, and with the idea of human progress. Key to their interests and works was the shared belief that the application of reason based on careful observation of the world, not faith in ancient authority, would bring practical social benefits. For contemporaries, the Enlightenment was about improvement in the human condition.

Most modern interpretations of the Enlightenment agree that it was an eighteenth century phenomenon concerned with new ways of critical thinking as means to intellectual and social reform. Yet important variations exist in modern Enlightenment scholarship. For some people, the Enlightenment was a largely philosophical phenomenon, predominantly urban and male and by origin French, a movement defined by the writings and lives of thinkers such as Voltaire and Montesquieu. Other commentators discern a broader Enlightenment, with origins in Holland and in England or, more widely, have moved beyond philosophical enquiry to examine the social history of the Enlightenment and the ways in which new ideas and attitudes were evident in topics such as race, science and medicine, and women to name only a few. Further strands of Enlightenment studies are apparent in the work of those who have considered the Enlightenment's contradictions, consequences, and continuing implications. For as we look back on eighteenth century claims concerning the power of reason, it is clear that Enlightenment ideals did not always result in political advance, intellectual liberation, and individual freedom. The power of reason was at times and in the wrong hands a corrupt force: with reason came restraint, with tolerance, terror. Here, for some, the focus is the Enlightenment's uneven legacy, what some have considered 'the dark side of the Enlightenment'.

For all these reasons, the Enlightenment continues to fascinate, albeit differently understood, then and now. One further and vital reason is its geography – for what was a period of philosophical and historical enquiry was also a period of geographical investigation.

Although interpretation of the Enlightenment's 'where' has often been overlooked, the Enlightenment was a geographical matter in several senses. In the Enlightenment, what was known about the size and the content of the world changed dramatically as a result of voyages of geographical exploration. Previous views of the world were in time overturned (**Figure 1**). The Pacific Ocean in particular became Europe's ethnological, botanical, and astronomical laboratory as new peoples and natures were revealed to the gaze of Europeans and island sites became local fixing points for global measurement. As **Figure 1** shows, there was much to know at the beginning of the Enlightenment about the shape of continents, and the dimensions of the 'fourth part of the world', Australasia. Geographical data provided the raw material from which philosophers and historians fashioned their theories about human society. In books, maps, and illustrated travel accounts, audiences across Europe, on the eastern coast of North America and in Spanish Latin America encountered these new geographies with curiosity and wonder. In universities and public academies as well as in schools, geography was widely taught in the Enlightenment, often in association with geometry and with history: the first valuable in learning mathematics, the second in understanding the development of nations, both as useful knowledge. Books of geography incorporated Enlightenment theories about human society, about the shape of the Earth, and about the natural resources of countries. Furthermore, the Enlightenment had geographical expression in the location and reception of its ideas among different publics, and in the movement of those ideas – from the Pacific or Lapland back to metropolitan Europe, from town to country, between different social groups and within different institutions. In short, the Enlightenment was about geography and spatial difference – about the Earth and its geographical variety – and it was something that had a geography, which took place in and over space.

Geographies of the Enlightenment

The Enlightenment in National Context

Where the Enlightenment has been studied geographically, the scale of examination has commonly been the nation. Two principal variants of the Enlightenment exist as a national phenomenon. The first is signaled to by 'the

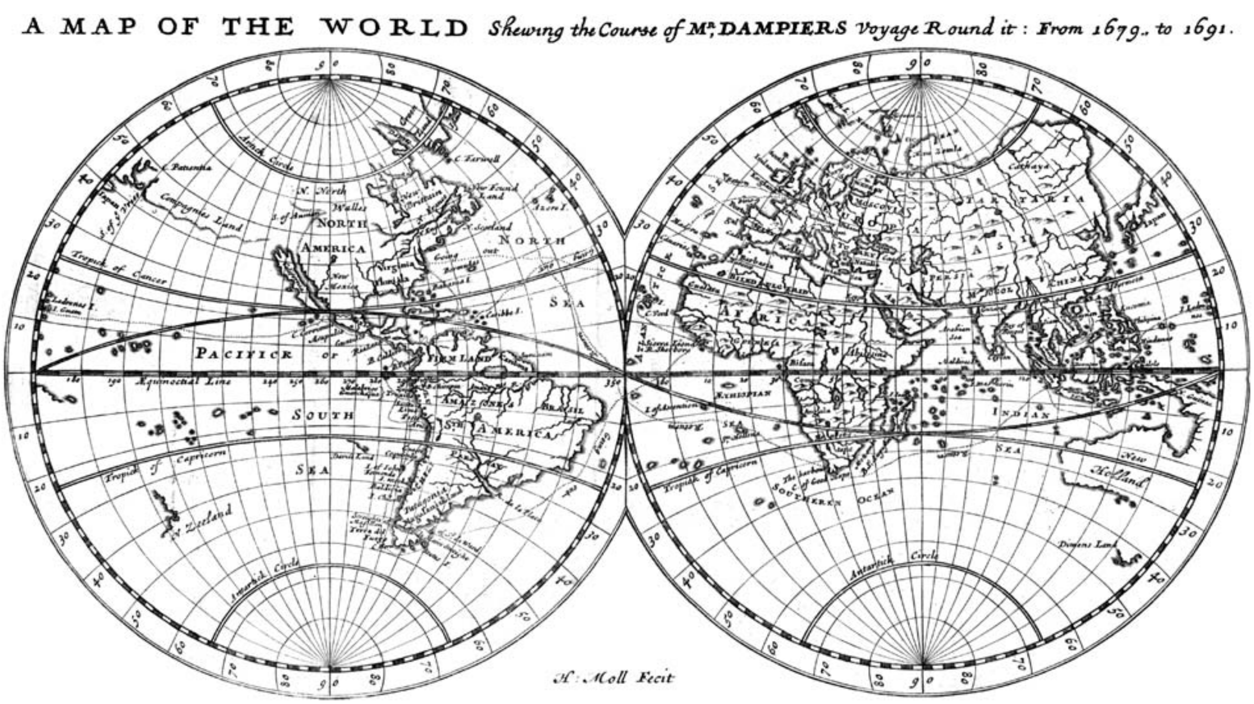

Figure 1 The pre-Enlightenment world: note the uncertainty over the northwest American coastline and the errors over New Guinea and Australia. Reproduced from Dampier, W. (1698), *A New Voyage Round the World*. London: James Knapton, map opp. title page, drawn by Herman Moll, by permission of the Trustees of the National Library of Scotland.

Enlightenment in France', 'the Enlightenment in America', 'the Enlightenment in the Netherlands', and so on. The second is evident in 'the Italian Enlightenment', 'the Scottish Enlightenment', and the like. In the first of these variants, the Enlightenment is made to have both essential defining characteristics and variant national expression by virtue of an unproblematic movement – in the form, for example, of philosophical or scientific ideas, the travels of influential figures, or in the dissemination of their books – from one nation to another. Over time and over space, the Enlightenment travels as an essential entity and, often, is held to do so from an originating Enlightenment 'core' or 'hearth' – very often France or the Netherlands – to a 'periphery' or to 'margins', such as Portugal, Spain, the Greek speaking countries of southeast Europe or to the Spanish speaking colonies in Latin America. The second variant hints at more distinctive experiences resulting from given national circumstances but, in so doing, differences within the nation are obscured even as common experiences above the nation may be neglected. As common as they are, these variants have been subject to criticism and revision in recent years in line with wider changes in Enlightenment studies.

One significant feature has been suspicion of the definite article – a dangerous linguistic tool for geographers and historians. This is important. What matters is not that there may not have been a shared and single Enlightenment apparent at the national level: it is just that scholars should not take it for granted that there was. The use of the definite article encourages the presumption that there was such a thing as the Enlightenment and that how we treat it geographically is simply to trace its origins and its dissemination in different national contexts. Omitting the definite article – even, thinking of diverse Enlightenments rather than any one essential Enlightenment – reminds us of the importance of differences within and above the nation. Two examples will help illustrate this point.

The term 'the Scottish Enlightenment' was first coined in about 1900. Yet what this label speaks to was delimited a century or so before by writers commemorating men whose work gave Scotland a prominent place on the map of eighteenth century philosophical and scientific life. Modern scholarship on the Scottish Enlightenment at one time saw the movement as essentially philosophical and certainly urban, even Edinburgh based, the preserve of a few great thinkers such as David Hume, Adam Smith, and Thomas Reid. More recently, such interpretations have been replaced by studies, which reveal differences between the cities of Edinburgh, Glasgow, and Aberdeen, which point to the importance of chemistry and natural history as well as to moral philosophy and which show that enlightening Scots looked to counterparts in Naples, Boston, and Berlin as well as to each other.

So, too, in Italy. Where the Italian Enlightenment was once reckoned the preserve of a few liberal priests and cultured bibliophiles, more recent study has highlighted important geographical and social distinctions. Enlightenment in the south centered upon Naples, was rooted in political economy, and apparent in topics as diverse as Newtonianism and volcanic activity. In the Papal States by contrast, significant features of Enlightenment included scientific reform in agriculture and the engagement of women in a scientific and courtly milieu akin more to Paris than to Naples.

Thus, while the nation is a convenient unit by which to study the Enlightenment, to do so uncritically is problematic. Pictures of the Enlightenment as a single intellectual movement fracture in the face of differences between nations and the geographical and social diversity within nations. In the eighteenth century, the idea of the nation was anyway relatively new. In France, for example, Enlightenment map makers and military surveyors were much involved in defining the extent of their nation, beginning in the 1680s, and, after the Revolution, in restructuring old regional boundaries in the New Republic. In Greek speaking eastern Europe, Enlightenment notions were imported from western Europe and, used in debates about national identity often by Greek expatriates in Italy and Germany, and were put to a different purpose than had been intended by their progenitors: in Greece, the idea and ideal of the nation and of Enlightenment helped establish one another. In Portugal, Enlightenment was characterized less by its nationals working elsewhere and more by foreign intellectuals working from within. In Spain, the Enlightenment was much restricted by the Church and despotic government. Yet, in Spain's colonies in Latin America, a distinctive Enlightenment emerged that looked to local traditions of historical scholarship and moral authority, not to the uncritical acceptance of European intellectual concerns. Then, rather than think of the Enlightenment as having only national expression, we should also approach its geography in terms of transnational connections.

Cosmopolitan Geographies

Thinking about the Enlightenment above national context allows us to focus on the nature, strength, and directions of the intellectual traffic in ideas that sustained critical commentators in different parts of the world. What connected political economists in Naples and in Edinburgh, for example, was the fact and the process of correspondence. Throughout Europe, epistolary networks and literary exchange bound people together in botanical study, chemical experimentation, and medical thinking, for example, as journals were exchanged, reviews penned, news disseminated, and societies' volumes distributed. Beyond Europe, Enlightenment ideas about natural diversity and human difference were shaped and reshaped by the receipt of new data – plant specimens, cultural artifacts, and on occasion, living humans.

The Enlightenment was thus a Republic of Letters (and of artifacts), defined by correspondence, practices of exchange, and networks of communication. Such transnational or cosmopolitan geographies are apparent, for instance, in the internationalism of the expeditions which undertook to measure the transits of Venus in the 1760s – an important moment in astronomical study and in the development of longitude – and in the coordination of mapping projects between Britain and France even as the two countries were at war. Yet the Republic of Letters was not, as it were, a 'free country', something simply 'above' geography, or national interest. Correspondence and exchange was principally between people deemed credible witnesses and reliable authorities by virtue of their social status and intellectual capacity. Global voyages of exploration such as those of the Briton, James Cook and the Frenchman, Jean François de Galaup, the *Comte de* Lapérouse, pursued commercial objectives for national advantage. Most significantly, those networks of correspondence and exchange that sustained the Republic of Letters centered upon the intellectual interests of urban metropolitan Western Europe more than they did upon the peoples of the Pacific, for example, or of America – or even of Northern Europe.

These facts of social and geographical variation do not weaken the idea of the Enlightenment as a dynamic and cosmopolitan phenomenon so much as they remind us of the need to pay attention to the nature, direction, and strength of its connections and to the fact that the Enlightenment moved above and beyond the nation. This in turn demands that we consider the local institutions and social spaces in which the Enlightenment's cosmopolitan geographies were grounded.

Local Sites and Social Spaces

Consideration of the Enlightenment's geographies at a local scale is consistent with, not separate from, its study at national and transnational levels as well as being part of the more general diversification in Enlightenment studies. For what national and cosmopolitan approaches reveal is less a single national Enlightenment or a transnational Enlightenment in only, say, political economy, than myriad micro geographies of the Enlightenment at work in different social spaces – in salons and societies, in learned academies and on board ship, in café talk and public house lectures, and in bookshops and from pulpits.

Between 1660, and the foundation of the Royal Society of London, and the 1790s, about 70 official academies and societies were in existence across Europe with a few in the Americas and in Asia. These institutions were the Enlightenment's 'hot houses', sites for the debate and discussion of dissenting ideas, for the analysis of scientific data, and sociable conversation on the latest news. Many such bodies, often short lived, also appeared in smaller towns and cities and had a concern with agrarian reform, mineralogy, or with other expressions of economic patriotism. This was especially so in the predominantly Protestant regions of Germany. In France, the Enlightenment gravitated around the Academy of Sciences in Paris and other comparable bodies, but throughout that country, learned societies undertook an enlightened agenda often with local emphases – Montpellier for medicine and mathematics, for example, more mercantilist interests in the port towns.

Anatomy theatres, coffee houses, concert halls, and other such spaces provided further sites for the engagement with and performance of the Enlightenment. In Enlightenment London in particular, coffee houses were commonly sites for turning scientific debates into popular lectures, and Newtonian cosmography into neighborly conversation. In Paris by contrast, café talk was often written down and reported on and was thus more guarded. In salons, talk was not modulated gossip but mannered and informed debate. Salons were important sites for Enlightenment sociability, polite spaces for the domestication of abstract ideas and, in France particularly, afforded a space for the engagement by women with Enlightenment ideas given their exclusion from academies and societies.

In addressing the geographical scale of the Enlightenment, several points may be made. The importance of different geographical scales of analysis should be noted. Recognizing these different scales is not to privilege one over another. It is, rather, to consider the connections between them, to see how the Enlightenment was made and received in different ways in different places and to understand the Enlightenment to have been about the dynamic interchange of ideas over space not just a period or a moment in time. An important consequence of such geographical thinking is to displace the idea of the Enlightenment as an essential thing by a view of it as multiple and variegated, with numerous geographical modifiers. Whatever people held Enlightenment to be, its networks of correspondence, movement, conversation, and experimentation came to Earth in certain places: its cosmopolitan concerns always had local expression.

Geographical Knowledge and the Enlightenment World

Exploring, Travelling, Mapping

Enlightenment exploration and the individuals and institutions that undertook and promoted it sought to bring the globe under the sovereignty of science through the practices of primary survey – voyaging, observing, mapping, collecting, reporting, writing, and disseminating. Charting the globe did not necessarily mean leaving the laboratory and the library behind, for many explorers' ships were, effectively, mobile universities and collecting sites, vessels of modernity for a new understanding of the globe.

Exploration in the Pacific in particular revealed what contemporaries called 'a new, or fifth division' to the world, Europe, Africa, and Asia being the first three, Columbus's discovery of America having provided the fourth. Taken together, the voyages of men like John Byron, James Cook, John Phipps, and George Vancouver for the British, of Louis Antoine de Bougainville, Nicolas Baudin, and Lapérouse for the French, of Alessandro Malaspina (a Genoan working for the Spanish) and Vitus Bering, a Dane employed by the Russians, gave shape to the world in terms of its continental margins and their overall extent. The myth of *Terra Australis Incognita*, an unknown southern continent, was exposed as Australia and New Zealand and strings of Pacific islands began to take recognizably modern cartographic dimensions; the northwest coast of North America likewise; earlier depictions of Baja California as an island were replaced by its status as a peninsula (cf. **Figure 1**).

Even so, Enlightenment voyages did not reveal a geographically diverse world equally or easily or at once. Discovery was often dangerous, new facts often hard won (**Figure 2**). Cook and Lapérouse died in the Pacific, Malaspina was imprisoned by his patrons upon his return, his work never published. The polar regions remained resistant to geographical scrutiny until well into the nineteenth century. Explorers' ships were akin to floating laboratories, yet, as **Figure 2** suggests, the pursuit of scientific knowledge often had to give way in the face of the more immediately practical problems posed by uncertain geographies. What was revealed – natives, fellow brethren or 'savages' depending upon different commentators' views, a bountiful nature, a scientific wealth undreamt of – provided matters for disquiet as well as reflection. Map makers had to rethink their works. At least initially, published reports of Enlightenment navigator explorers have a tone of astonishment and hesitancy: they are not simply scientific accounts with all the language of precision and regularity that this term connotes. Although continental margins were plotted, interiors were uncharted: Africa was the least well known continent in the 'Old World', Australia the same in the 'New World', China figured in travel accounts but its description was often based only on hearsay.

That is why the geographical opening up of the world in the Enlightenment is distinguished by the development of published guides on how to travel and what to collect. From the late seventeenth century onwards, the genre of what is, effectively, the geographical field guide, starts to appear in an attempt to regulate the languages of description, collection, and explanation. Such regulatory texts were not always successful: in the face of different scales for measurement, of cultural diversity and different languages, and given the lack of consistency in longitudinal measurements until the 1770s, it was hard to reveal the world and to represent it according to agreed metrological conventions and standards. That is, again, why we must be cautious about seeing the Enlightenment as an unproblematic move to modernity in which rational practices and languages were everywhere and at once adopted, and why Enlightenment geography matters, for the European encounter with the exotic 'other' varied from place to place and did not always involve the far away.

The Enlightenment brought new mapping projects and new types of map makers into being. Where most European countries earlier had a mix of mariner map makers, geographers, commercial publishers, and land surveyors, the Enlightenment was distinguished by the rise to prominence of the professional military or state surveyor. Many had an institutional base in an academy or, in the case of France's military geographers, the *Dépôt de la Guerre*, and for its hydrographers, the *Dépôt de la Marine*. France took the lead in European nations in being systematically surveyed under the direction of the Cassini map making family in the Enlightenment, but in estate mapping, in topographical survey, in the Military Survey of Scotland and elsewhere in Britain, in the triangulation projects of William Roy that underlay the establishment of the Ordnance Survey, Europe's nations mirrored the French. In Bengal, men such as James Rennell produced atlases and maps that promoted and reflected the commercial expansion of Britain's empire.

Enlightenment map making thus emphasized empirical observation and quantifiable measurement in the service of the state and of science. Yet for many of the peoples who were encountered through geography and mapping, discovery was not a sure route to their proper cartographic delineation. Enlightenment maps are symbols of geographical encounter and of cultural engagement but they are in part also documents of cultural effacement in that they often omit or misrepresent evidence of the peoples discovered. The painterly depictions of foreign geographies in the work of artist naturalists such as William Hodges and Sydney Parkinson likewise strove for depictive accuracy but did so according to European conventions. The move toward realism in representation was associated with the development of standards in measurement, observational training in collection, and systematic explanation in science. Yet encounter and empirical enquiry could also lead to effacement, the writing 'out' of people rather than their always being sketched 'in' (**Figure 3**). George

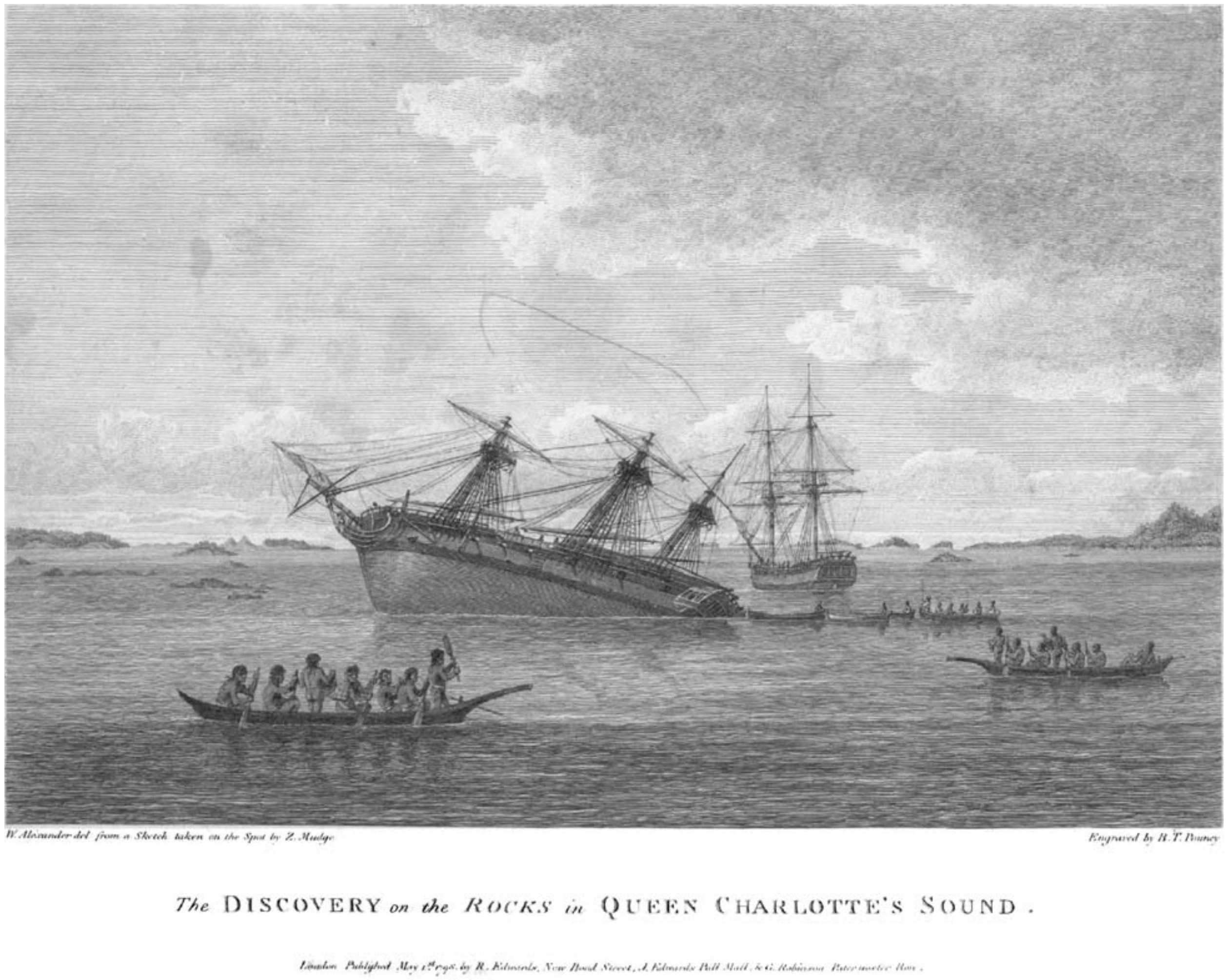

Figure 2 The fragility of Enlightenment geographical exploration is evident in this engraving entitled 'Discovery on the rocks' [in August 1792]. Reproduced from Vancouver, G. (1798). *A Voyage of Discovery to the North Pacific Ocean, and Round the World*. Vol. I, facing page 364. London: G. G. and J. Robinson, by permission of the Trustees of the National Library of Scotland.

Vancouver's work – from which **Figure 3** is taken – is one of many which helped realize and disseminate 'foreign' geographies in word and image to public audiences. Indeed, the facts of exploration, travel, mapping, and picturing only added to the sum of Enlightenment knowledge if they made a final move – the 'voyage into print'. When they did, Enlightenment audiences avidly consumed tales of the far away and strange in different ways and places: by borrowing from libraries and reading, by meeting the Tahitian natives who returned or by witnessing in a London theatre the staged mimicry of Cook's distant geographical 'performances'.

Encountering the Terraqueous Globe

"On the surface of this immense globe we are presented with heights and depths, prairies, seas, marshlands, rivers, caves, chasms, and volcanoes and, on first inspection, we find no regularity, no order in all of that." Writing thus in his *Histoire Naturelle*, the leading Enlightenment natural historian Georges Louis LeClerc, *comte de* Buffon, encapsulated the sheer scale of what there was to know about the world and the role of Enlightenment natural philosophers in systematizing what was to be known. As they set about the investigation of the terraqueous globe – "to discover the true geography of the Earth" as Vancouver put it – Enlightenment commentators challenged established views of the age and origins of the Earth, the authority of religious interpretations and, at the same time, provided the foundation for modern scientific disciplines. The themes of geodetic measurement, botanical classification, earthquake theorization, and meteorology may be used to illustrate these claims.

The 'shape of the Earth' debate was a question central to Enlightenment interests in what we may think of now as terrestrial physics or geodesy. The problem, in essence,

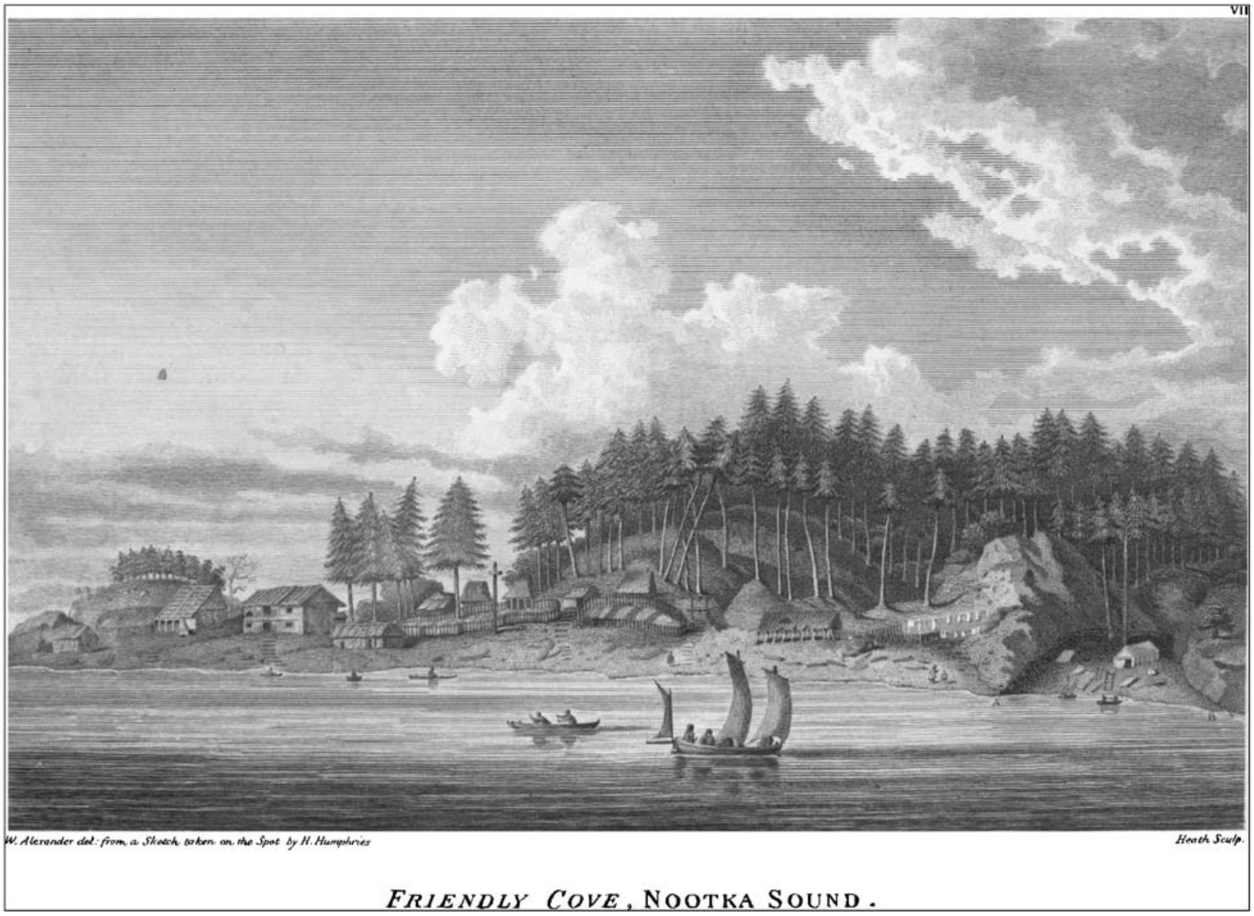

Figure 3 Here native names have been partly displaced by explorers' views of the nature of the encountered human 'other'. Reproduced from Vancouver, G. (1798). *A Voyage of Discovery to the North Pacific Ocean, and Round the World*. Vol. I, facing page 388. London: G. and G. and J. Robinson and J. Edwards, by permission of the Trustees of the National Library of Scotland.

was straightforward: if one followed the precepts of Newtonian gravitational theory, the Earth was elliptical in shape, flattened at the poles. If one subscribed to Cartesian ideas of vortices, the Earth would be extended at the poles, and flattened across the equator. To an extent, these competing mathematical conceptions found national even local expression: the British favored Newton, the French (with one or two significant exceptions) favored Descartes. The solution to the problem demanded geographical fieldwork in those parts of the world key to the different theoretical views, on the equator and as near as possible to the poles, where experiments with pendulums and in magnetic declination would determine the Earth's shape. And so in the early 1730s, two expeditions were undertaken. The first in 1733 headed to Lapland under the guidance of Pierre Louis Moreau de Maupertuis (an advocate of Newtonian reasoning), the other to Peru in 1737 under the direction of Charles Marie de la Condamine. The results of the expeditions proved Newton and Maupertuis right. Geographical fieldwork, from often small sites but with global consequences, both corroborated and refuted book based reasoning (**Figure 4**). Large scale and land based Enlightenment science had its many smaller counterparts – as is evident in the tents at Observatory Inlet in **Figure 4** – and often depended upon maritime corroboration of its instrumental readings: new continents were 'put to shape' by this interplay of instrumentation and navigation at the margins of land and sea.

The mid 1730s were crucial to Enlightenment physical science for another reason, for as geodesists were testing the Earth's shape, botanists were putting the world to order. This is particularly evident in the schema of the Swedish naturalist Carl Linnaeus. His works established, on the basis of plants' sexual characteristics, a standard botanical nomenclature of generic and species names. Linnaeus also proposed five climate zones across the world by which to explain the distribution of plant types. For Linnaeus, botanical work with a utilitarian purpose was central to the Enlightenment: "the sciences are the light that will lead the people who wander in darkness." Yet his ideas were not the only ones nor were

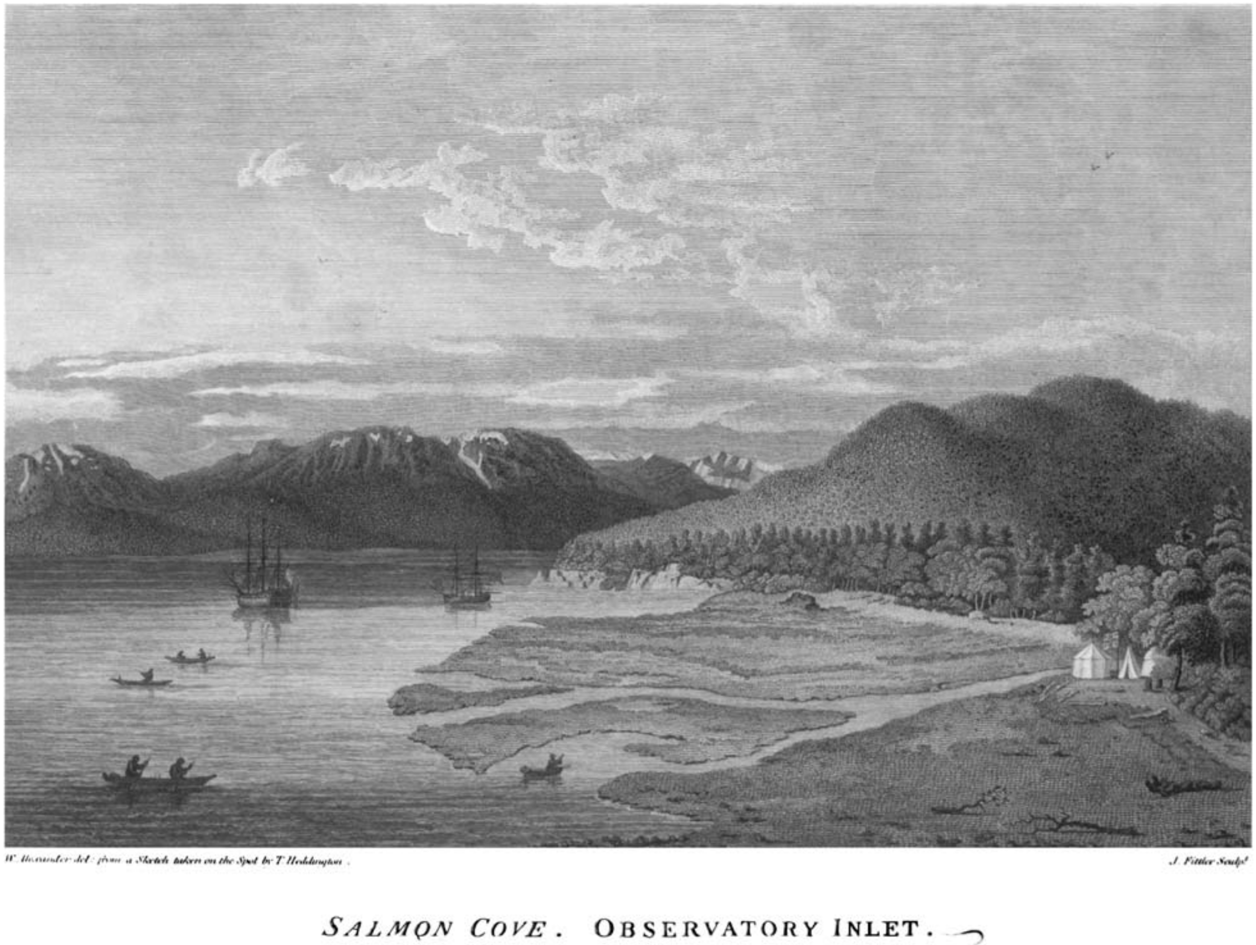

Figure 4 James Cook observed the Transit of Venus from a tent incongruously named 'Fort Venus' on Tahiti. Here the small-scale nature of geography's field sites is evident in one of Vancouver's observatories on the North American coast. Reproduced from Vancouver, G. (1798). *A Voyage of Discovery to the North Pacific Ocean, and Round the World*. Vol. I, facing page 332. London: G. G. and J. Robinson and J. Edwards, by permission of the Trustees of the National Library of Scotland.

they everywhere accepted. Parisian naturalists opposed him. In London, botanists did not think his taxonomy accurate. In Berlin, Edinburgh, and in St Petersburg, alternative readings of the *Book of nature* were proposed. Even as the empirical sciences became established, the map of Enlightenment scientific activity looked different in different places.

One site which profoundly altered conceptions of the Earth's history as divinely disposed and prompted renewed speculation upon the Earth as a dynamic object was Lisbon. On 1 November 1755, one of the largest earthquakes in human history hit the city, followed by a tsunami and fires. The physical shock was felt across Europe: in Scotland, the waters of Loch Lomond were 'much agitated'; in the West Indies, a smaller version of the tidal wave arrived 9 hours later; aftershocks continued for weeks. The mental shock was as enormous and longer lasting. In a world where natural circumstances had theological significance, what could the destruction of a European capital mean? What clerics explained as divine retribution, Voltaire interpreted as the end of enlightened secular optimism and others saw as either a chemical phenomenon or the result of electrical discharge. However understood, the effect of the 1755 Lisbon earthquake was to accelerate existing trends in the Enlightenment toward materialistic and historical philosophies of nature and away from theological ones.

Much the same is true of meteorology in the Enlightenment. In Classical thought, meteorology meant 'reading the skies' for natural portents that could then be used to explain or foretell social circumstances. With the development of mathematical reasoning, physical experimentation, and instrumentation, commentators across the world began to systematize weather records and to correspond with one another over the results. It is not appropriate to see shifts from weather watching to weather recording as the basis to an explanation of climate, far less to see in such work any sense of climate change in modern terms (although people were aware of extreme weather events). Yet late Enlightenment meteorology in particular was a portent of developments in the sciences as a whole. Further, questions of climate mattered crucially in the Enlightenment in relation to questions of human difference.

Charting Human Difference

What emerged from the voyages and books of geographical discovery in the Enlightenment was the 'Shock of the Human New': reports of pygmies, 'cannibals', Pacific natives, American indigenous peoples flooded in, and challenged conventional wisdom even as they were, in many instances, refuted. The impact of this 'shock' was apparent in several ways. New evidence about human variety helped lay the foundations for the human sciences, and, more generally, helped establish what Enlightenment commentators termed the 'Science of Man', namely, bringing questions of science as method and of utility as outcome to bear upon the moral and the human world, not just the physical and natural world. In this respect, the Enlightenment laid the foundations for the emergence of 'modern' human disciplines – anthropology, sociology, even moral philosophy, and linguistic studies to name just a few. And as ideas about human difference and critical methods about evaluating human difference appeared, so theoretical models of human development began to be articulated, theories which commonly posited 'European Man' at the head of stages of social development and positioned other peoples and cultures at different stages of development beneath them (**Figure 5**). Often, as in this depiction of canoes, illustrations served to provide a visual taxonomy of social difference and geographical difference, taxonomies that could themselves be transported home for later further analysis. The Enlightenment was thus characterized by geographies of human difference in time in which the social systems and cultures of different peoples were used to advance theories about relative social development, and, at the same moment, by geographies of human difference over space which underlay what Edmund Burke called the "Great Map of Mankind" – namely, the evidence of different social systems and thus, different stages of human development, being present at once in different parts of the world.

One powerful explanation for these different human geographies was climate. Unlike earlier writers who considered climatic questions either divinely disposed or fixed by planetary influences in arguments about bodily 'humors', for example, Enlightenment theorists turned to climate and climate's influence in order to explain the relationships between custom, law, social systems, and human difference. A key figure in this respect was Charles Louis de Secondat, *Baron de* Montesquieu, whose 1748 *L'Esprit des Lois* (*The Spirit of Laws*) is a

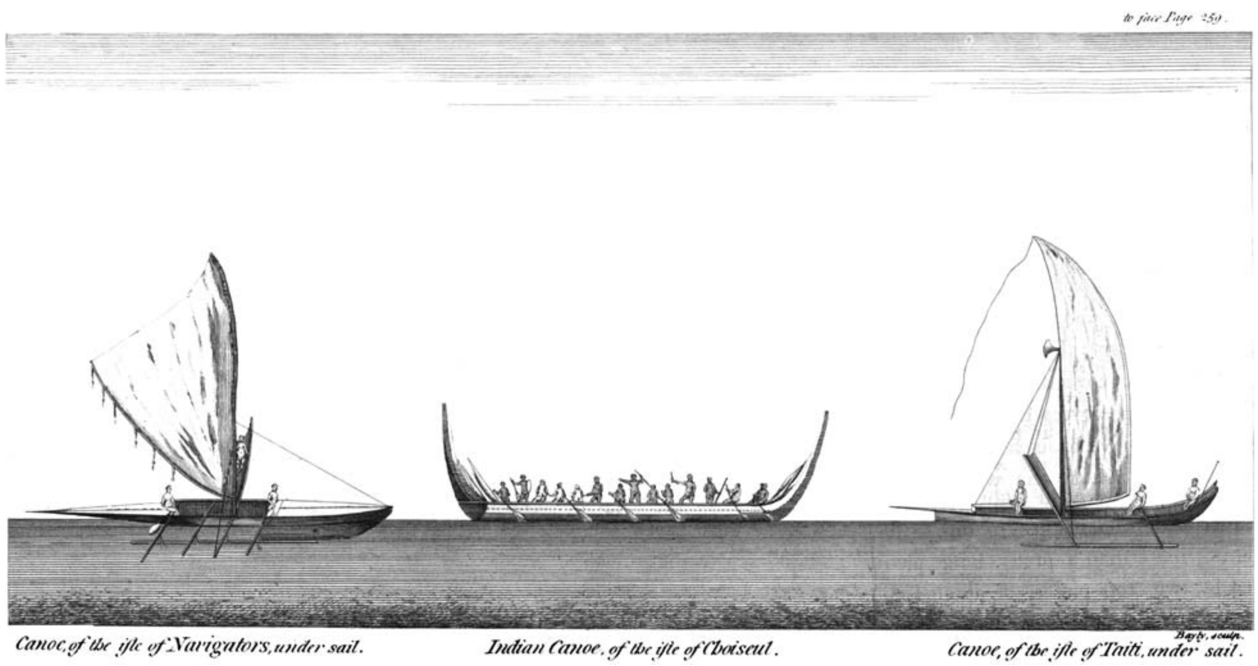

Figure 5 Understanding human social differences involves classifying material artifacts as well as cultural traits and languages here the focus is Pacific islanders' canoes. Reproduced from de Bougainville, L.-A. (1772). *A Voyage Round the World* (*Trans. from the French by J. R. Forster*), facing page 258. London: J. Nourse and T. Davies, by permission of the Trustees of the National Library of Scotland.

comprehensive attempt to understand the moral codes governing human conduct – and a key work on the effects of climate as a physical determinant of social difference. Montesquieu subscribed to the view that climate – by which, essentially, he and others meant temperature variations with latitude – influenced the body's physical state which, in turn, helped constitute different social customs. What geographers and others (such as Linnaeus) took to be the 'temperate zone', Montesquieu saw as a space of moral and cultural equability: 'In temperate climates we find the inhabitants constant in their manners, as well as in their vices and virtues: the climate has not a quality determinate enough to fix them'. But to the North and to the warmer South especially, climate determined different degrees of cultural inconstancy against this supposed European norm. Hence South Sea islanders, South African 'Hottentots' and others could be – and were – regarded as 'savages' as a result of the climate driving their culture.

Montesquieu's work (like others') needs to be understood in context. He was in part advancing an environmental explanation from earlier theories and he was reliant more upon earlier geographical accounts than upon the results of later Enlightenment geographical discovery. But he was doing so as part of novel attempts to explain the nature of the human condition and its geographical variability. Others such as Voltaire and David Hume thought his emphasis on physical causes, misplaced: for them, social agency not temperature or latitude was the basis to human difference. It is for these reasons important that we should not judge Enlightenment commentators by later scholarly or moral standards since what they were searching for was, effectively, a new set of rhetorical instruments, a new analytic vocabulary, by which to explain the Science of Man. And the terms they employed in respect of human differences – 'race', 'nation', and 'species', for example – changed over time and were differently used in different Enlightenment sites and social spaces.

Linnaeus's classifications of the genus *Homo* and its single species, *sapiens*, incorporated four *varietate* (varieties) in his initial formulation in 1735: *Europaeus albus* (white European man), *Americanus rubens* (American red man), *Asiasticus fuscus* (Asian yellow man), and *Africanus niger* (African black man). But his classifications changed as new geographical information brought new data to his attention. What he took to be a world of different fixed varieties, others such as Buffon interpreted as more dynamic, the result of natural relationships with *la variété humaine* changing in its traits over the generations. Where Linnaeus saw a world of human fixity, Buffon saw variation over time. And while variation might mean 'improvement' in the nature of human society, so it could also allow degeneration, the deterioration of the stock. That is why Buffon, looking at the generally smaller, less robust and, to judge from numbers and distribution, less fertile human species in the Americas, interpreted the peoples there as inferior to Europeans.

At the same time, ideas of 'progress' and of the stage by stage development of human society from hunting, pastoralism, agrarianism, to commercialism, helped position the West' and the 'Rest'. As James Dunbar put it in his *Essays on the History of Mankind in Rude and Cultivated Ages* (1780), "The history . . . of some of the South Sea isles, which the late voyages of discovery have tended to disclose, enables us to glance at society in some of its earlier forms." Such ideas about what contemporaries termed 'conjectural history' were particularly prevalent in Enlightenment Scotland, but they were a feature of much Enlightenment writing about human society.

In effect, what people then understood as the 'Science of Man' was in several senses a 'Geography of Man'. For those encountering human difference across the globe, human geography was a matter of location and distribution: who was where? Explaining human distribution as a matter of geography – why were peoples where they where? – elicited geographical answers: climate and latitude. For yet others, geography was not a fixed natural 'base' against which human cultures were determined, but, rather, something that could be moulded by human agency. That is why so many works of economic and agrarian commentary and of historical and philosophical speculation in the Enlightenment are concerned with improvement: they were, in one way or another, outlining future geographies – not what was but what could be.

Taken together, the processes of geographical knowledge at work in the Enlightenment made the world seem older, bigger, more complex and with a more modern 'shape' than had earlier been thought. As geographical examination provided the raw data for emergent disciplines and the physical world was subject to detailed scrutiny (**Figure 6**), so humans individually and socially became the object of theoretical models. As Grüner shows in this engaging and early picture of glaciers, Enlightenment observers placed themselves in nature: direct encounter was an important concern in depicting and in understanding the world's geography. And as they looked for explanation of the issues of climate and social development, Enlightenment thinkers turned to the facts of geography to explain human society.

Geography in the Enlightenment

Books and the Making of Geographical Identity

As the Enlightenment world became bigger and more diverse, so books of geography incorporated and disseminated the new facts and theories. In doing so, they exemplified for those particular concerns the rapid expansion of periodical and printed works that

Figure 6 Glaciers were of interest for their aesthetic merits as well as being understood and labeled as agents of landscape formation. Reproduced from Gruner, G. S. (1770). *Histoire naturelle des glaciers Suisse.* Paris: L'Imprimerie, Plate XII, by permission of the Trustees of the National Library of Scotland.

distinguished the expanding literate markets of the Enlightenment's public sphere in general. In understanding what books of geography were in the Enlightenment, what they did, who were their authors and who their readers, it is helpful to make a distinction between books of geography in a more strict sense – even although there was variation within this category – and books in which geography figured prominently as part of the promotion and diffusion of Enlightenment ideas.

A key work in the latter sense is the *Encyclopédie, ou: Dictionnaire Raisonée* edited by Denis Diderot and Jean d'Alembert, and published in Paris in 17 volumes between 1751 and 1765 (with 11 volumes of plates appearing between 1762 and 1772). For its critical tone, engagement with the unwarranted authority of the Church, and advocacy of powerful reason, the *Encyclopédie* was much prized by contemporaries. It has been seen by later commentators as the embodiment of the Enlightenment, even its 'manifesto'. Its editors saw it as providing a new map of the intellectual world in which the essence of, and the connections between, different categories of knowledge were laid out. Geography figured within this schema, had close connections with mathematics and history and was divided into several parts – physical geography, historical geography, sacred geography, ecclesiastical geography, natural geography, and political geography being the main divisions. As a subject, geography thus embraced the study of the globe and the measurement of its physical characteristics, matters of practical utility resulting from such work – such as the mathematical bases to navigation – and the moral bases to human occupation of the Earth.

In looking at the *Encyclopédie* in another way – at its movement and reception rather more than at its content – we can see how the Enlightenment had a geography through its books. In addition to the volumes of the Paris edition, the *Encyclopédie* was published in Geneva, Lucca, Leghorn, Neuchâtel, Berne, and Lausanne. Many of the authors of the Paris edition were members of local academic institutions, notably so for the authors from Languedoc who were associated, for example, with the *Société Royale des Sciences de* Montpellier. From the figures of the subscription sales we can chart, at least on a city by city basis, the purchase of the *Encyclopédie* and, thus, plot something of the social contours of the Enlightenment's reception. Lyons dominated the geography of the *Encyclopédie*'s sales with 1079 subscriptions, nearly twice that of Paris. Sales were generally higher in France's provincial capitals – Toulouse, Bordeaux, and so on – and in towns with an administrative and cultural focus. The work sold poorly in port towns and in the textile towns of northern France. Beyond France, Geneva dominated with 284 subscribers with towns such as Genoa, Warsaw, and Liège having several dozen subscribers. By and large, the main purchasers were a fairly heterogeneous group of noblemen, clerics, civic officials, and professional persons.

If the *Encyclopédie* thus illustrates in outline the Enlightenment's purchase, mobility, and reception among different social groups in different towns, a glimpse at books of geography reveals a rather different picture. Most books of geography in the Enlightenment emphasize geography as a coherent body of knowledge about a clearly defined object, namely the situation of places on the Earth and the content of those places in human and natural terms. Many geography books were not, however, the work of geographers, at least not in our modern and professional understanding of that term. Most were by compilers of geographical facts, jobbing journalists in effect. This is not to say that the books were not regarded as useful or that they did not incorporate new geographical information. Many did and as Enlightenment discoveries were added to the stock of knowledge, numbers of such books and their print runs increased.

English language books of geography produced in Britain tended to place Britain at the forefront of utilitarian knowledge and as the leading nation in advancing new geographical understanding, but occasionally, they did not privilege Britain or even England. In his *A New Geographical, Historical, and Commercial Grammar* (1770), for example, the Scots born William Guthrie explicitly connected his geography with those contemporary theories, many by his fellow Scots, on the stage by stage development of human societies. In France, Edme Mentelle wrote different sorts of geography books to use in different social settings – first as a Royalist tutor, latterly as an author for the New Republic in works which were intended to mould a new citizenry and a new modernity through the study of geography. In Greece, geographical authors used their books to promote the idea of contemporary Greece not as an Enlightenment 'margin' but as a modern setting for the virtues – in reason, philosophy, and science – of ancient Greece.

The clearest picture of geography books making the Enlightenment work and of geography making national identity work in the Enlightenment comes from America. As was the case in many towns and cities on both sides of the Atlantic, the Enlightenment public had a wide variety of public teaching and instruction in geography to choose from. In America, however, interest in geography depended on books from anywhere but America. Worse still, what was written about America was very often not founded on truth. The response to these circumstances came from Jedidiah Morse, formerly a Congregationalist minister in Charlestown, Connecticut, whose *The American Geography, or A Present Situation of the United States of America* (1789) is the first major book in Enlightenment America to shape America's national consciousness through geography. To understand why this was so, it is helpful to look at Morse's earlier *Geography Made Easy* (1784), at other comparable works such as Thomas Jefferson's *Notes on the State of Virginia* (1787) and to appreciate the intellectual and political context in which books of American geography helped make America.

The new geography of the North American continent represented challenges to established knowledge and to political administrators alike. The gradual East–West movement of map makers slowly rendered America visible as a geopolitical space (**Figure 7**). Powell's map here shows, on its western margins, the then Spanish dominion of Louisiana. After 1804, and the Louisiana Purchase, America significantly enlarged itself to the West. For authors of geography books, new on the ground information was necessary if the emergent nation was to be fully understood and properly governed. The result was the appearance in America of a new 'print discourse of geography', what has even been termed 'a geographic revolution in early America', in which books become the means to know one's self and one's nation. Where Noah Webster promoted the Americanization of place names and sought to provide a national frame of reference for different spoken accents in his *The Grammatical Institute of the English Language* (1783–93), Jedidiah Morse turned to books of geography. And in order to promote America's geography, Morse inverted the established order of European texts by placing America first. *Geography Made Easy* is the first book in Western print history to do so.

Morse's *American Geography* built upon these ideas and upon Jefferson's *Notes on the State of Virginia* which had

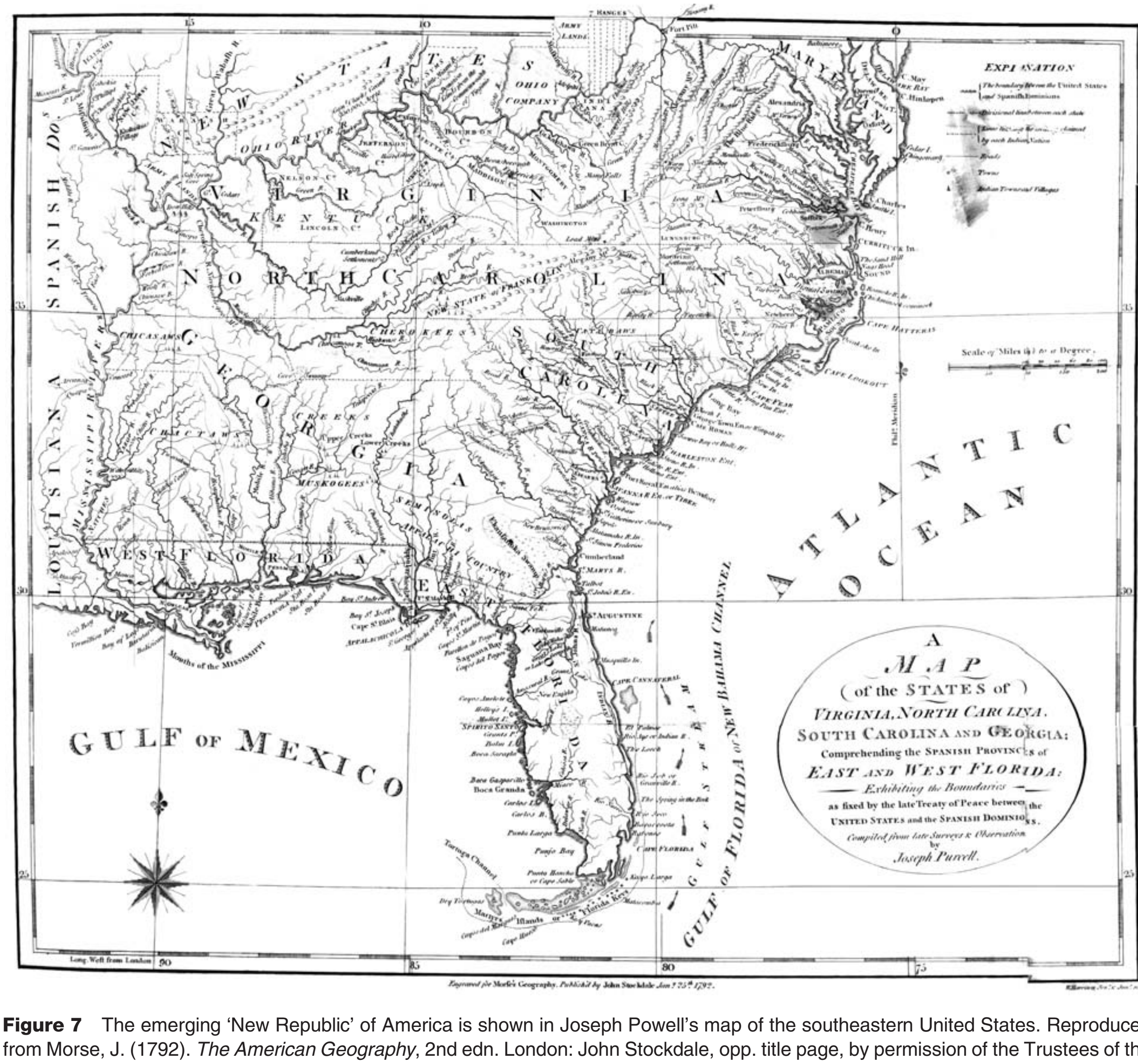

Figure 7 The emerging 'New Republic' of America is shown in Joseph Powell's map of the southeastern United States. Reproduced from Morse, J. (1792). *The American Geography*, 2nd edn. London: John Stockdale, opp. title page, by permission of the Trustees of the National Library of Scotland.

offered a critique of Buffonian views about the limiting effects of America's climate. Again, Morse inverted textual tradition and put America first. He went on to offer a moral view of geographical difference within America, drawing upon notions of climatic influence and of the present and future social development of different states – Connecticut being the most virtuous, New England generally being the standard by which the new republic should advance, the southern states being too prone to indolence through drink, innate sloth, and the effects of climate. To modern readers, such views would be anathema: for Morse, as for Guthrie, books of geography could and did provide the basis for national identity and standards of moral conduct.

Maps, Medicine, and Military Space

Geography worked in other ways, not just through its books. In the Nuremberg Cosmographical Society (1746–54), for example, classes were held on mathematical geography and historical geography and schemes laid for a state geographical bureau to collect geographical statistics and for correspondence networks to collect data from individuals farther afield. In Paris, we can identify the local sites and social networks through which geography and map making was carried on. Before 1790, most geographers – and their associated publisher booksellers – concentrated on the *Quai d'Horloge*, the *Quai des Augustins* and the *Ile de la Cité* in the heart of the city, some like the *géographes du roi*, the geographers to the King, in court appointed residences. After the Revolution, these local geographies of Parisian geography altered as courtly privileges ceased and trade restrictions were lifted.

In France, significant geographical work in the Enlightenment was undertaken by the corps of military engineers, the *ingenieurs géographes*, men trained in foreign languages and survey techniques and, to use modern parlance, in the design and interpretation of questionnaires and other qualitative techniques by which the contents of national space, and not just its measurement and mapped delineation, could be known. Military mapping in America before the American War of Independence was used by both sides in the campaigns.

There was by the late Enlightenment a definable 'medical cartography', even a 'medical geography'. Disease mapping has an established history in the Enlightenment. In 1700, the Bolognese geographer and naturalist Luigi Marsigli had prepared a plague map of the eastern Adriatic coast in which, among other features, he had outlined the time geography of different treatment times from the center of the outbreak. Dot maps of the yellow fever outbreaks in New York in the 1790s helped physicians in their treatment of the disease. In the work of a German physician, Leonhard Ludwig Finke, we can discern an attempt to think geographically about the geography of disease and its treatment and, thus, to go beyond mapping toward explanation. Finke outlined his concerns in his 1772 *Versuch Einer Allgemeinen Medinisch–Praktischen Geographie* (Attempt at a General Medical–Practical Geography) by drawing upon his own experiences as a regional medical officer, a civil servant responsible for preparing what were known as 'medical topographies' – syntheses of meteorological, climatic, and demographic data geared toward more effective political administration. He drew upon the work of others, including the statistician and early zoo geographer Eberhardt Zimmermann and of some French map makers. Finke's reasoning was geographical, his end in view political: here was geography being used as a tool for the practical improvement of peoples' lives – the very ideal of Enlightenment.

Spaces of Geographical Sociability

Finke's book may not have been treasured and read in Enlightenment homes as was the case for Morse's and Guthrie's texts, but books, maps, globes, and even geographical games were a commonplace of Enlightenment sociability in the home. And not just there, for in schools, learned academies, universities' curriculum, and in public classes, the Enlightenment was characterized by a widespread civic engagement with geography.

Institutional membership lists can be used to assess geography in the public sphere but must be examined with care. In terms of listed occupation, for instance, there were only seven *géographes du Roi* enumerated as members of the Parisian *Académie des Sciences* between 1699 and 1793, and a further 17 persons who recorded 'geography' as their principal disciplinary interest. But among the many botanists, natural historians, and mathematicians were men like Maupertuis and La Condamine who we have seen to have had geographical interests. The same is true for the Royal Society in London where geography was with astronomy and mixed mathematics a major component of the papers heard by that body in the years between 1720 and 1779, at least. In such academic societies, there was more geography being done than there were geographers doing it.

The same is true of those Enlightenment universities whose curriculum in this period shows geography being commonly taught as part of civil history, mathematics, classics, and humanities and in association with moral and natural philosophy. Although geography's connections with history and with mathematics was a commonplace of the period, the fact that different professors in different institutions placed different emphases upon different parts of geography means that we can site local geographies of geographical teaching. For that very reason, however, we must be wary of uncritical generalization. In Aberdeen, for example, the moral philosopher

Thomas Reid gave a course, which emphasized geography's utility 'to the important Offices of our Society'. In Glasgow, geography was taught alongside mathematics. In Königsberg, Immanuel Kant concentrated upon physical geography as, too, did Saussure in Geneva. In Göttingen, by contrast, mathematical geography was to the fore given the close connections between that university's teaching staff and the Nuremberg Cosmographical Society.

These remarks about the local sites and spaces of geographical activity apply also to geography's place in schools in the Enlightenment. Different institutional settings mattered: in his dissenting academies, the chemist and radical philosopher Joseph Priestley insisted on geography being taught as an adjunct to history, for instance, for history and geography helped to "free the mind from many foolish prejudices." In most grammar schools in England, by contrast, geography was associated with Latin and Greek as part of a much more conservative humanist tradition.

Most people in the Enlightenment did not receive a formal education. Yet there were many public classes in geography (and in many other subjects) available. Pubs and coffee houses provided a venue for Newtonian philosophy and sociability as well as geography teaching. There were over 20 different geography classes in Edinburgh between 1708 and 1800, a story echoed in Glasgow, London, and overseas in Europe and in North America. In Philadelphia, geography classes for young women were a particular feature of Mr Poor's Young Ladies Academy: there, in December 1789, one Eliza Dolby was given a prize for 'excelling in geography' – in fact, for reading aloud to her classmates from Morse's newly published *American Geography* and for discoursing upon its significance. In school, in public classes, and of course in the home, geography was thus spoken about as well as read and drawn or committed to memory and was learnt through play as well as via formal instruction. Narratives of exploration held the public's interest, perhaps none more so than Mungo Park's 1799 *Travels in the Interior Districts of Africa*, in which the author portrayed himself as dashing, even visionary, in contrast to his textual account of the rigors and dangers of African exploration (**Figure 8**). For women and

Figure 8 The Scottish explorer Mungo Park enjoyed considerable fame for solving, in 1796, one part of the then 2000-year-old 'River Niger problem' which way did the river run? He died in 1806 unsuccessfully trying to answer the other part where did the river end? Reproduced from Park, M. (1799). *Travels in the Interior Districts of Africa*. London: W. Bulmer, frontispiece plate, by permission of the Trustees of the National Library of Scotland.

children, the home was a site of conversation, instrumental instruction, and politeness: geography and sociability were commonly learnt together in the Enlightenment's many social spaces.

Enlightenment's Geography: Matters of Continuing Significance

Thinking about the Enlightenment's 'where' is as important as thinking about its 'when', 'what', 'who' and 'why'. Here, evidence about scale and the importance of geographical difference ought now to allow the dismissal of the idea of The Enlightenment as an essential feature, philosophical in nature, French in origin, and reducible only to the level of the nation. Further, it is clear that the Earth was revealed as an object of geography through practices of geographical enquiry such as exploring, mapping, and reporting. Geographical knowledge was a key component of Enlightenment theories about human culture, in the present, in the past, and in informed speculations about the future. In many of their speculative writings, commentators were, effectively if not implicitly, sketching out what the geographies of the future world might be. Geography was, with many other intellectual concerns, a form of Enlightenment knowledge, a form evident in texts, instruments, polite conversations, and formal instruction in different institutions.

If, then, consideration of the Enlightenment's geography allows us to know the period and the movement better and to be more faithful to its varying social and intellectual expression, so uncritical notions of 'The Enlightenment Project' weaken in the face of geographical difference. The often assumed privileged geographies of metropolitan Europe can be shown to be connected in complex ways with other parts of the world. Particular places – the Americas, Pacific islands – were key sites for the making of world knowledge and were not, as is too commonly depicted, 'margins' to an Enlightenment European 'hearth' or 'core'. These are important matters. From them, two further points follow. The first is that the Enlightenment – and its geographical differences – is with us still. I mean by this that the Enlightenment continues as an object of modern scholarly study because of its political implications, its ideals and languages and for the fact that it is unfinished. The Enlightenment continues also as an object of scrutiny for the legacy of its 'dark side'. Not all persons and nations benefited equally from attention to questions of reason. Doctrines of progress for some were for others route maps to the death camps. This is why some scholars ask of the Enlightenment 'what's left of it?' and, why some postmodern commentators, keen to dismiss the Enlightenment but wrong to do so, ask of it 'so what?'. The second is that the ideals of the Enlightenment, its emphases upon reason, the commitment to tolerance and social well being – what contemporaries often considered the 'quantum of happiness' of a society or a people – are for many persons still values worth committing to. Of course the Enlightenment looks different from the viewpoint of different human cultures and different human geographies and of course it may be interpreted differently from the vantage point of the historic present than from within the period itself. These too are reasons to keep thinking about Enlightenment and its geographies, then and now.

See also: Exploration; Geography, History of; Modernity.

Further Reading

Baker, K. M. and Reill, P. H. (eds.) (2000). *What's left of Enlightenment?: A postmodern question*. Stanford: Stanford University Press.

Bruckner, M. (2006). *The geographic revolution in early America: Maps, literacy, and national identity*. Chapel Hall: University of North Carolina Press.

Canizares Esguerra, J. (2003). *How to write the history of the new world: Histories, epistemologies, and identities in the eighteenth century Atlantic world*. Stanford: Stanford University Press.

Clark, W., Golinski, J. and Schaffer, S. (eds.) (1999). *The sciences in enlightened Europe*. Chicago: University of Chicago Press.

Clayton, D. W. (2000). *Islands of truth: the imperial fashioning of Vancouver Island*. Vancouver: University of British Columbia Press.

Fox, C., Porter, R. and Wokler, R. (eds.) (1996). *Inventing human science: Eighteenth century domains*. Berkeley: University of California Press.

Gavroglou, K. (ed.) (1999). *The sciences in the European periphery during the enlightenment*. Dordrecht: Kluwer Academic Publishers.

Kors, A. (ed.) (2003). *Encyclopedia of the enlightenment*, 4 vols. Oxford: Oxford University Press.

Mayhew, R. (2000). *Enlightenment geography: The political languages of British geography, 1650 1850*. Basingstoke: Macmillan.

Porter, R. (ed.) (2003). *The Cambridge history of science. Vol.4: Eighteenth century science*. Cambridge: Cambridge University Press.

Porter, R. and Teich, M. (eds.) (1981). *The enlightenment in national context*. Cambridge: Cambridge University Press.

Schmidt, J. A. (ed.) (1996). *What is Enlightenment? Eighteenth century answers and twentieth century questions*. Berkeley: University of California Press.

Withers, C. W. J. (2006). Eighteenth century geography: Texts, practices, sites. *Progress in Human Geography* 30, 1 19.

Withers, C. W. J. (2007). *Placing the Enlightenment: Thinking geographically about the age of reason*. Chicago: University of Chicago Press.

Withers, C. W. J. and Livingstone, D. N. (eds.) (1999). *Geography and Enlightenment*. Chicago: University of Chicago Press.

Enterprise Discourse

P. M. O'Neill, University of Western Sydney, Parramatta, NSW, Australia

Introduction

Enterprises are the core organizational entities for capit alist economies. An enterprise is a commercial business and might take the form of a corporation, a partnership, a cooperative, a conglomerate, a franchise or simply a corner store run by a sole trader. The most powerful form of enterprise, of course, is the corporation, and it is the main type of enterprise referred to in this article. Enterprise discourse is the language process that creates and propels the enterprise, forming up its organizational structures, devising its logics and strategies, and enacting its per formances. In this sense, enterprise discourse underpins the creation and operation of the modern corporation.

Constructing the corporation out of the idea of enterprise discourse draws on the epistemology of post structuralism. In this approach, the existence and oper ation of the corporation are co determined by texts and discursive practices, and the social practices and devices that bring them into existence. An enterprise discourse, then, has specific sets of vocabularies, language assem blages, practices and intentions which bring the enter prise into a certain type of existence at any point in time. Importantly, though, while a particular type of enterprise discourse can dominate all or part of a corporation for a period of time, enterprise discourses are never fully stable or enduring since, by their very nature, they generate alternative and oppositional discourses. The corporation, then, is not reducible to a single enterprise discourse. Rather, a corporation is a field of competing discourses each seeking different, often contradictory, operational pathways and distributional outcomes.

Enterprise Discourse and the Nature of the Firm

In one sense an enterprise discourse view of the firm sits uncomfortably with traditional views of the firm, espe cially those from the economic literature. In neoclassical economics the firm is portrayed as an ordered, autono mous and ultimately rational economic subject that op erates according to centrally imposed logic while manifesting predictable (or general) dynamics. A belief that a unifying, ordered logic underpins the capitalist firm similarly dominates Marxist economics, notwith standing the profound differences between neoclassicism and Marxism about the actual substance and consistency of this logic.

In another sense, though, one can find resonance with the idea of enterprise discourse in historical accounts of the firm in economic theory. Beginning with Adam Smith, the firm has consistently been assigned the sub jectivity of an individual with pure economic knowledge and detached agency. According to this prime organiza tional position, neoclassical economics naturalized the firm as the key agent of profit creation and capital ac cumulation within an economic system.

At the same time, though, there is a vast economic literature that explores the way the firm departs from such an idealized position. John Stuart Mill, for instance, wrote of the growing complexity of the firm seeing it as an entity engaged in the resolution of social claims over distributional flows. Mill thus expanded interest in the role of the firm beyond profit making and accumulation. His insights drew attention to the consequences for firm behavior of the separation of business management from business ownership and to the impacts on firms and their managers of the market conditions in which they operate. Through this type of thinking, Mill opened up the firm as an entity actively engaged in both its internal com position and, as much as its market power allowed, its external operating environment.

The works of Michael Jensen and William Meckling are a more recent manifestation of interest in internal firm composition and behavior and the rejection that these can be read from a general theory or exposition of capitalism. So too the works of Alfred Marshall, Edward Chamberlin, Joan Robinson, and William Baumol have demonstrated how a firm's behavioral options and con straints are shaped not by any general laws deducible from the foundational logics of an idealized capitalist firm but from the outcomes of interplay with other firms each with their own specific structures, powers, and ac tions. The theory of the firm developed in this lineage thus involves an exploration of the ways a firm functions as a vehicle for a variety of stakeholders to advance their interests and aspirations.

This theoretical broadening has opened up the firm as a space of contestation about economic procedures and directions and about social desires and obligations. It has also seen a broader interest in the firm by social scientists other than economists, accountants, and economic man agement theorists. There is now a considerable literature on the firm within many discipline domains, including sociology, psychology, labor studies, human geography, and cultural studies. Not accidentally, post structuralist approaches from within these disciplines have been brought to bear on our understanding of the firm especially in exploring the role played by language and its deployment.

Once we step outside the specific analytical language of economics, it is relatively easy to see that there are many different types of language sets, structures, and discourses that bear on a firm's structure and behavior. The following section concentrates on four of these: production and investment, financial, organizational and legal narratives. Like all discourses, each of these carries its own logics, unique vocabularies, dedicated calculus and so on. These four discourses are probably the key four that have impact in today's corporate world, though they compete for prominence with other significant discourse types such as the discourses of technological change, business ethics, management practice, and human resources, such is the complex field of the modern corporation.

Importantly, with an ever expanding range of production possibilities, markets, financing options and organizational structures, and with corporations becoming immensely more complex, traditional top down and hierarchical management practices that seek to rely on a unifying enterprise discourse have become redundant. Management decisions are now fed by highly skilled teams devoted to the creation of legal, financial, and engineering complexities using specialized verbal, symbolic, and mathematical languages, sometimes to expose the performance of the corporation or its parts, for example, to drive improvement; and sometimes to conceal the precise nature of a corporation's activities or structures, for example, from the eyes of competitors or regulators. Moreover, and somewhat paradoxically, rather than stabilizing the corporation, this overlay of monitoring, reviewing, and strategizing narratives builds oppositional enterprise discourses which offer alternative accumulation and distributional pathways for the firm. Management of the enterprise becomes the task or art, then, of assembling and implanting discourses which are capable of resisting opposing discourses because, perhaps, they take a contrary view about the way a firm should perform or because, perhaps, they seek to install different people to the positions of power from which new discourses can be deployed. In the modern corporation, discourse management is no easy task. As we see below, a modern corporation must have stable discourses that drive production and investment. These must intersect with a financial system, itself produced and governed by one of many types of financial discourses. So too the corporation must adopt a legal architecture and produce an organizational structure, each interrelated with the other and spliced inseparably onto the firm's production, investment and financial operations. That said, we now turn to each of these discourse types one by one.

Types of Enterprise Discourses

Production and Investment

Ideas of how to best organize the production of goods and services in large corporations have waxed and waned since the industrial revolution. In 1911, for instance, the construction of a text and an argument for scientific based organization of factory labor by Frederick W. Taylor in his book *The Principles of Scientific Management* led quickly to the overturn of artisan or craft based systems of industrial organization. Taylorism, as it became known, was an industrial discourse that provided the rationale for wholesale changes to the conduct of work. Tasks became subdivided into sets of basic operations allowing workers to be stationed in fixed positions along an assembly line to perform low skilled repetitive tasks. Factories became transformed from workshops into large scale sites designed for the mass production of standardized goods through the installation of specialized capital equipment whose design and function intersected explicitly with divisions of labor designed to maximize worker productivity and factory throughput, however demanding the work process had become.

A similar transformation in industrial production methods was driven by a variant to the Taylorist production discourse, namely, the one advanced by Henry Ford in the middle years of the second decade of the twentieth century, the precursor years to the economic boom years of the 1920s. Fordism, as this version became known, added the idea of mass consumption to the discourse of mass production, Ford arguing that living wages were in the interests of industrial corporations since they provided the pool of domestic demand necessary for the sales of growing industrial output.

One can chart these types of discursive constructions and displacements throughout the twentieth century: Fordism replaced by state driven demand management practices known as Keynesianism; Taylorism undermined by automation and the substitution of labor by capital; developed world industrialization eroded by productive investments in newly industrializing nations like South Korea and Singapore, describable by many as the creation of a new international division of labor; massive restructuring of developed world industrial capital and the depowerment of organized labor under neoliberalist economic management regimes pioneered by UK prime minister Margaret Thatcher and US president Ronald Reagan; high technology, flexible industrial systems, especially in Scandinavia, in search of a high value added, high wage industrial response to deindustrialization of developed world economies; and so on.

In their time, all of these discourses of production and investment convinced industrial audiences that there was a better way to enhance the productivity of capital and labor. Once adopted, a production/investment discourse produced not just the promised changes to output but also corresponding shifts in distributional outcomes – according to how powerful or clever certain fractions of capital or organized labor have been in the negotiations and struggles relating to the industrial changes in question.

Financial

Financial discourses also underpin the operation of the corporation and play an increasingly powerful role in decisions about how production is organized, what investments are desirable and what distributional outcomes should be encouraged or engineered. The package of logics, vocabularies and calculus that constitute financial discourses – like other forms of enterprise discourse, though probably more powerfully – produces three outcomes: first, it creates organizational structures; second, it creates management devices; and, third, it creates regulatory and performance standards, some of which are internal to the financial community and others which are governance measures imposed on it.

Ever since the headline capturing takeovers, mergers and asset stripping ventures of the 1980s, there has been considerable public awareness of the powerful role that financial management plays in building the structures and strategies of firms. Yet, accounting and finance departments have always been integral to firms' structures. They rival production, engineering, and sales departments for their influence on firms' behaviors. The introduction of Taylorism to workplaces in the early twentieth century, though, led to the dramatic insertion of accounting practitioners and their instruments into the management of production processes in a way that saw the nature of production redesigned to take account of what financial, especially accounting, narratives dictated. What the firm looked like changed radically. So too, the discovery that idle cash flows internal to a firm gave the firm much potential as a debt raising and payment device saw profound changes to the structure of firms in the 1980s. These changes which not so much shifted individual modes of production – though these modes were certainly changing for technological reasons and in response to rising international market competition – but brought a consciousness to the production manager of the need to capture and steer value creation as quickly and as efficiently as possible. Once again, new ways of doing financial business changed firms' organizational structures, the internal structures of firms being reorganized to mimic financial, not just production, representations of the firm. Aided by the IT revolution, new systems of financial management were devised which focused managers' attention on specific performance measures (asset values, gearing levels, cash flows, labor costs and so on) which thereby demanded particular behaviors. They also created particular social and political ways of governing individuals, activities and organizations. For example, they might have encouraged more aggressive controls over labor practices in a firm while according zero economic value to philanthropic behaviors. A new role of financial management has, therefore, become the creation of a logic, a direction and a process of change. Each financial management model constructs a new language, replete with its own formation, representations, vocabulary and calculus. Thus in a discursive sense, financial management can be seen as a process of scripting the operations of the corporate organization and the behaviors of its members.

A financial narrative, then, operates as an enterprise discourse which constitutes and is constituted by a field of power relations thereby producing, as we have seen, the preconditions for its contestation. A financial narrative is constructed for the purpose of producing a set of power relations which will deliver and stabilize a particular set of accumulation and distributional outcomes. The enduring success of a hegemonic management narrative lies, therefore, in its capacity to recreate its logics, languages, and calculus, making other narratives and their agents less inviting through time. Of course, any process of suppression will involve appropriating strategic features from alternative narratives through the same processes of networking and talk which maintain the hegemonic status of the dominant narrative. On the other hand, suppression creates the conditions for resistance and the rise of competitive narratives with different claims and aspirations. Financial discourses are thus, like all enterprises discourses, comprised of a multitude of competing stories continuously in the process of formation and, potentially, in competition for dominance.

Notwithstanding their role as a vehicle for the successful deployment of power and control within an enterprise, financial discourses within an enterprise are also part of an external field of financial discourses and of wider societal discourses of social and economic transformation. The legitimacy and survival prospects of a firm are enhanced by adopting key financial management practices and procedures consistent with wider political directions and institutional processes. Successful financial management depends on the societal network of practices and discourses in which it is set. Yet, simultaneously, financial management constructs and transforms social institutions, and the (largely economic) truths that are associated with them. Of course, the intersection of financial narratives and societal talk is enhanced by the status of finance professionals in the wider community, in the political process and in the control of large corporations.

Organizational

The organizational structure of a corporation is the direct outcome of a plethora of talk, especially management talk. This talk has a particular motivation and agenda depending on its authors and disciples. There is a complex debate in economics and organizational theory about which types of management talk have achieved primacy in modern capitalism. The traditional view is that managers talk up maximization goals especially the maximization of profits, sales and asset values. Another view is that a firm's managers will seek to maximize the opportunities and rewards they can accrue personally from running the firm so long as some reasonable level of return to the firm's owners are maintained. Another is that managers are motivated by securing the firm's long term interests and so will seek control over the field of transactions that intersects with the firm's interests. Managers might also be seen as maximizing their power within the organization resulting, perhaps, in actions to accumulate personnel and resources under their direct control. And, it is possible that managers might be motivated by wider social justice and sustainability concerns and so steer the firm toward socially and environmentally responsible activities.

Of course, a large corporation will have long experience of all of these motivations, and more, from within its management ranks. As it embraces a variety of management actions – each and every one of which is devised and propelled within an enterprise discourse – it creates the organizational wherewithal to process actions and responses, most commonly through the engagement of specialists and the formulation of strategies and policies in the key areas from which talk is generated: production, sales, finance, human resources, ethics, IT, and so on. Over time, these 'areas' are institutionalized into specialist departments with defined responsibilities, and so corporate structure is built.

On the one hand, then, the firm is a dynamic field of competing logics and aspirations often situated within wildly chaotic, disconnected fields of economic and social transactions. On the other hand, because firms have long histories of reconciling distributive claims on productive capacities and earnings, their operations are stabilized by their organizational form and by sets of much prescribed and therefore repeated behaviors and transactions. Importantly, these prescriptions draw heavily from observations of other firms and from the vast array of management discourses available almost ubiquitously including at airport bookshops and in the management texts of the world's most elite universities. Successful managers are the ones who devise and institute new operating languages to control and capture the value that firms create without the need for operatives to make constant referral to management for direction and decision. Enterprise discourse instills instinctive responses into a firm's practices and thereafter consolidates them into organizational form and structure.

Legal

The enterprise can also be portrayed as the product of legal discourse, a cluster of legal obligations and relationships, or a nexus of contracts – some formal, some not – which have been spawned to identify and protect the private property and wealth creating rights of the corporation as a social entity. Legal rights and obligations that are seen to be located within the boundaries of the corporation include basic property rights; rules for the use of productive assets; frameworks to maximize and legitimize economic cooperation in internal enterprise relations, such as employer/employee and shareholder/manager relations; rules governing market behaviors; duties and obligations in raising debt and equity finance; and rights and restraints in international transactions.

Just as with production/investment, financial, and organizational discourses, the legal narrative of the corporation is a historical tale of tension, contest, resolution, and crisis, bounded by wider societal knowledges, practices, and expectations. And just as with other types of enterprise discourse, the legal version plays a major role in stabilizing the corporation, legitimizing private production and distribution activity, and directing external relationships now and into the future. Law builds the corporation as the prime economic entity through three major installations: autonomy, limited liability, and fiduciary duty.

Autonomy is the transfer of the rights of an individual directly onto the corporation. The idea of the corporation as an autonomous entity draws directly from the Lockean notion of the autonomous individual subject, and is fundamental to corporate legal discourse.

Limited liability is the condition by which shareholders of a company are legally responsible for the debts of a company only to the extent of the nominal value of their shares. Limited liability was the legal cornerstone to nineteenth century American industrialization enabling risky ventures like railways and water schemes to attract private finance without further call on financiers' other assets. Simultaneously, limited liability reinforces the benefits that flow from the idea of autonomy by effectively evacuating citizenship responsibilities from the list of obligations of the shareholder.

Fiduciary duty is the basic duty of care in supervision and positional conduct that all citizens are required to exhibit. For the corporation, fiduciary duty refers to the relationship of trust that shareholders have with the corporation's directors and managers; chiefly that they will advance the objectives of the company, which invariably involves the maximization of financial returns to shareholders via dividends and capital gains. While open to interpretations that press wider societal and environmental obligations on a firm, court interpretations of fiduciary duty have invariably narrowed its field of meaning to oblige corporate managers and directors to elevate the interests of shareholders above all others. Such a narrowing also reinforces the suite of legal pro tections and powers that emanate from the discursive presence of the ideas of autonomy and limited liability.

Just as the law is a discursive act, we can see that the corporation is in turn constructed by legal discourse; and that as this discourse shifts and bends according to new interpretations, statutory adjustments, and professional practices, so too the corporation adjusts and changes. Often these changes will be required to accommodate changes made within the corporation arising from shifts in production/investment, financial, and organizational discourses. Frequently, though, corporate law – its texts, practices, and professionals – will override, prohibit or adjust processes of change emanating elsewhere, signi fying that while legal discourse may lack the physical force of a production/investment narrative, the cut through power of a financial narrative or the historical inertia of organizational talk, legal discourse sources its unique form of power from the autonomy, distinction, and status that it accords itself and has ceded to it.

Alternate Discourses of Responsibility

Taking an enterprise discourse view of the corporation opens up the corporation to the possibility of it being something else: more environmentally or socially re sponsible, for instance; or more generous to outsiders; or, even, engaged politically in pursuit of something better. By identifying the coexistence of the production/in vestment, financial, organizational and legal discourses of the firm, we can see the enterprise as organized by a number of different logics, some of which are mutually supporting while others are deliberately in opposition; which means that the determining power or logic or normative prescriptions within any one discourse is never guaranteed.

From a management point of view, as we have seen, the latent instability of a corporation arising from the coexistence of both supportive and competing enterprise discourses means that much of the art and task of man agement is the construction of hegemonic narratives of performance within a firm in order to produce a stable strategy capable of driving a firm's accumulation and, probably more importantly, its distributional outcomes. Yet, because there will always be an alternative man agement narrative that offers a different accumulation and a distributional outcome, there will always be a struggle over a narrative's formulation and enactment. Other managers will seek to displace the dominant nar rative with another. Employees will contest it in order to claw back a greater proportion of surplus value in all sorts of monetary and nonmonetary forms. On the edges of the firm, other stakeholders (shareholders, com munities, host environments, and state agencies) will want a say in the narratives that drive the firm's oper ations, and they will focus also on the form and balance of accumulation and distribution options. At any point in time, then, the firm can be shown to be composed of a very deeply institutionalized set of structures, though at every other point in time these will change, driven by a contest of talk over accumulation and distributional outcomes.

Such a discursive and reflexive view of the enterprise raises the broader question about realizing the corpo ration's potential as a public sphere – where it is normal to witness open debates about corporate governance, rights of access to information, mutual and social obli gations, philanthropy, environmental sustainability and so on. Mobilizing these sorts of discourses about the en terprise and attaching them to a field of possibilities that we might know as the corporation has the potential to create a more humane and responsible, perhaps even a more productive, modern corporation.

See also: Corporate Responsibilities; Discourse; Economic Geography; Firms; Poststructuralism/ Poststructuralist Geographies.

Further Reading

Boden, D. (1994). *The Business of Talk*. London: Polity Press.
Brown, J. S. and Duguid, P. (1995). Organizational learning and 'communities of practice'. In Cohen, M. D. & Sproull, L. S. (eds.) *Organizational Learning*, pp 76 92. London: Sage.
Fairclough, N. (1992). *Discourse and Social Change*. Cambridge: Polity Press.
Gibson Graham, J. K. and O'Neill, P. M. (2001). Exploring a new class politics of the enterprise. In Gibson Graham, J. K., Resnick, S. & Wolff, R. (eds.) *Re/Presenting Class: Essays in Postmodern Marxism*, pp 56 80. London: Duke University Press.
Hopwood, A. G. and Miller, P. (1994). *Accounting as Social and Institutional Practice*. Cambridge: Cambridge University Press.
Maskell, P. (2001). The firm in economic geography. *Economic Geography* 77, 329 344.
Miller, P. (1994). Accounting as social and institutional practice: An introduction. In Hopwood, A. G. & Miller, P. (eds.) *Accounting as Social and Institutional Practice*, pp 1 39. Cambridge: Cambridge University Press.
Norton, B. (1995). The theory of monopoly capitalism and classical economics. *History of Political Economy* 27, 737 753.
O'Neill, P. M. (2001). Financial narratives of the modern corporation. *Journal of Economic Geography* 1, 111 129.
O'Neill, P. M. and Gibson Graham, J. K. (1999). Enterprise discourse and executive talk: Stories that destabilize the company. *Transactions of the Institute of British Geographers* 24, 11 22.
Schoenberger, E. (1997). *The Cultural Crisis of the Firm*. Cambridge, MA: Blackwell.
Taylor, M. and Asheim, B. (2001). The concept of the firm in economic geography. *Economic Geography* 77, 315 328.
Yeung, H. W. (1998). The social spatial constitution of business organizations: A geographical perspective. *Organization* 5, 101 128.

Entrepreneurship

E. Stam, Utrecht University, Utrecht, The Netherlands

Glossary

Cluster Geographic concentration of interconnected companies and institutions in a particular field.
Entrepreneurial Opportunities A set of ideas, beliefs, and actions that enable the creation of future goods and services in the absence of current markets for them.
Nascent Entrepreneur Individual who is taking steps to found a new business but who has not yet succeeded in making the transition to new business ownership.
New Firm Formation Rate The number of new firms started, as a percentage of the total number of firms.
Self-Employment Rate The number of business owners, as a percentage of total employment.
Spin-Off A firm that is started by someone who has left his employer.

Introduction

Entrepreneurship has been recognized as an important element in explaining (regional) economic development. This means that the explanation of regional variations in entrepreneurship has also become an important issue. Even more so because there are pronounced differences within and between nations in rates of entrepreneurship.

Historically, the literature has often explained entrepreneurship as the product either of environments (like provision of venture capital (VC), growing demand) or of personal attributes (like risk taking propensity, need for achievement). Individuals are heterogeneously endowed with knowledge (bounded rationality, prior knowledge), attitudes, and preferences and environments are heterogeneously endowed with institutions, inputs, and demand for products. The entrepreneurial process depends on entrepreneurial opportunities in the environment and enterprising individuals that identify and exploit these opportunities. Entrepreneurship is the result of the interaction between individual attributes and the surrounding environment.

When individuals identify an opportunity, they do not react automatically with establishing a new firm (assuming that they have the intention to start one): new firms are created with a sequence of processes like creating a legal entity, product development, and financing. Given prospects of employment, education, and other circumstances differ across individuals, the population is heterogeneous with different individuals facing different opportunity costs when acting to exploit an opportunity they recognized. There are many obstacles at the road from entrepreneurial intentions to a successful business.

Although this article aims to give an overview of the regional determinants of entrepreneurship, it should be kept in mind that results from studies of specific regions are hard to transfer to another region due to the specific, often idiosyncratic institutional context of these regions. This means that it is unlikely that there will be institution free descriptions of best practices in entrepreneurship policy.

Definitions and Measurement of Entrepreneurship

The literature on entrepreneurship recognizes a variety of entrepreneurial roles. Entrepreneurs have been recognized to fulfill different roles in economic change, as:

1. the person who assumes the risk associated with uncertainty (Knight);
2. an innovator (Schumpeter);
3. a decision maker (Casson);
4. an industrial leader (Schumpeter);
5. an organizer and coordinator of economic resources (Marshall);
6. an arbitrageur (Kirzner); and
7. an allocator of resources among alternative uses (Schultz).

These roles all implicitly carry an economically positive connotation with them. However, if entrepreneurs are defined to be persons who are ingenious and creative in finding ways that add to their own wealth, power, and prestige (Baumol), then it is to be expected that not all of their activities will deliver a productive contribution to society.

Three indicators of entrepreneurship are often used: the percentage self employed in the labor population, the number of new firms, and the number of new growing firms. Some studies also take into account people with a preference for entrepreneurship ('latent entrepreneurship'), or people that take steps to start a new business ('nascent entrepreneurship'). The latter two indicators can be seen as potential entrepreneurship. The dominant empirical definition of entrepreneurship in economic geography is the formation of a new firm by (an) individual(s). These individuals are called entrepreneurs, or an entrepreneurial team when a collective starts a firm. There are two general methods for empirically analyzing entrepreneurship as new firm formation. The first method standardizes the (annual) number of new firms relative to the number of establishments already in existence, which is termed the ecological approach, because it considers the amount of start up activity relative to the size of the existing population of businesses. The second method, the labor market approach, standardizes the number of self employed with respect to the size of the labor force. The latter approach is based on the theory of entrepreneurial choice, which explicitly focuses on the occupational choice between wage labor and self employment.

Entrepreneurship as a Regional Event

Entrepreneurs are hardly lone individuals who rely primarily on their extraordinary efforts and talents to overcome the difficulties inherent in the formation of a new firm. The process of starting a new firm is eminently social, as information and resources are to a large extent acquired via the personal networks of the (nascent) entrepreneur. The social ties of the potential entrepreneurs are likely to be localized, and induce entrepreneurs to start their firm in close proximity to their homes and to their current employers. For example, a study of Portuguese manufacturing firms found that Portuguese entrepreneurs were willing to accept labor costs three times higher than in alternative locations to locate the new business in their current region. It is almost a stylized fact that entrepreneurs start their firm in the region where they live and/or work. For them the choice is what kind of firm to start given their location, not so much choosing a location for a given firm.

There are several reasons for this locational inertia of entrepreneurs. First, they can utilize their existing network to seek partners, employees, suppliers, customers, advisors, and investors. This decreases search costs, but it also permits them to build upon credibility and trust developed in past relationships. Second, they can start on a part time basis (often being home based) and delay full time commitment until the venture seems sufficiently promising. Third, a spouse can keep a job so that income continues to flow to the family; other aspects of a founder's life can remain the same. The full energies of the entrepreneur can then be devoted to start up.

Demand and Supply of Entrepreneurship

For explaining the spatial distributions of entrepreneurship, one can look at the demand side and at the supply side of entrepreneurship. The supply side refers to the composition and characteristics of the population of individuals and organizations as producers of entrepreneurship. It starts with the availability of (potential) entrepreneurs. The supply side can be studied from several perspectives: the nature and number of organizations in a region, the regional culture, and the labor market structure in a region. Key elements are the resources, abilities, and preferences of individuals. It explains why in given environments some individuals are more likely to start a firm than others: for example, due to their willingness to incur risk, preference for autonomy and self direction, specific human capital, and experience. The demand side of entrepreneurship represents the opportunities for entrepreneurship. In contrast to the supply side perspectives, demand side perspectives argue that individuals in particular environments are more likely to be entrepreneurs because the availability of opportunities encourages their exploitation by starting a firm. The demand side can be studied by looking at the sources of opportunities: for example, a growing purchasing power, technological change, and regulatory change (see **Figure 1** for a graphical representation of the analytical framework).

Empirically the supply and demand for entrepreneurship can be measured with different indicators. **Figure 2** shows the levels of perceived skills and knowledge to start a business among adults in European regions (supply of entrepreneurship), while **Figure 3** shows the levels of perceived opportunities to start a business in the local area (demand for entrepreneurship). These figures show that there are not only large international variations in the factors related to the demand and supply of entrepreneurship, but also significant regional differences within countries (e.g., within the UK, Germany, and Italy).

Industry Structure

New firm formation across regions can be explained by differences in the regional composition of industries and by differences in one particular industry in specific regions. The latter would indicate that there are context specific differences affecting entrepreneurship rates, while the former would indicate that the explanation should largely be sought in the specific industry structure of the

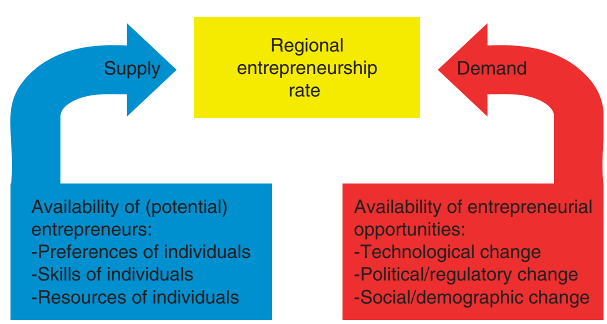

Figure 1 Supply and demand of entrepreneurship.

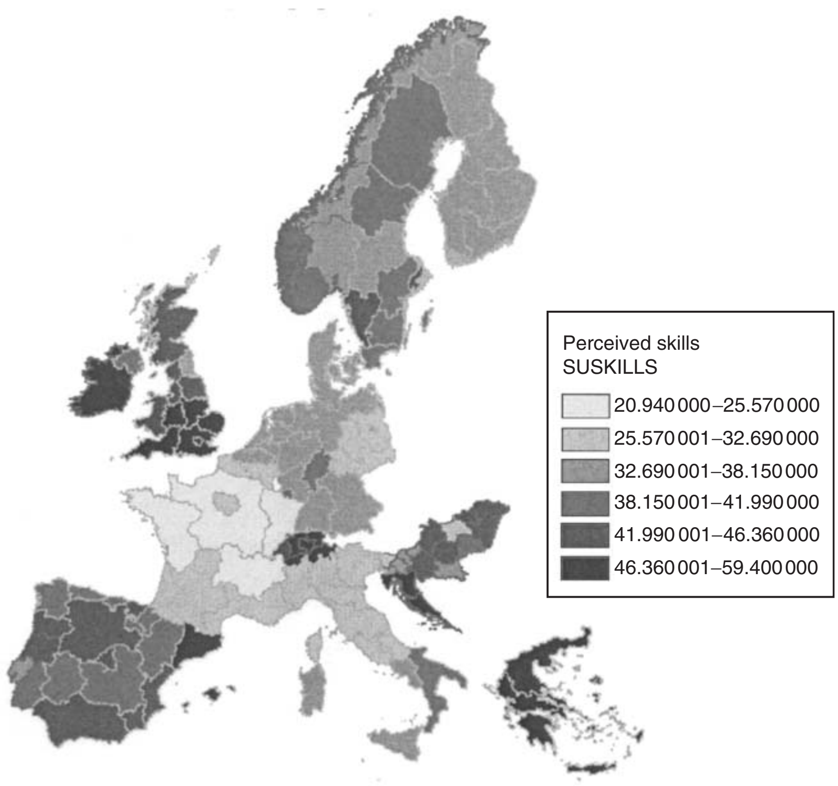

Figure 2 Perceived skills and knowledge to start business, percentage of population between 18 64, 2001 05 (Source: Global Entrepreneurship Monitor.)

region. The industry structure of a region affects the overall new firm formation rates in a region, as industries differ in their degree of contestability (entry barriers) and the extent to which entrepreneurial opportunities emerge (e.g., many in business services and few in mining). In dustry structure may affect the supply of, as well as de mand for, entrepreneurship.

Sometimes both the industry structure and the re gional context are favorable for the new firm formation rate in a region; this can, for example, be found in the South East of the UK, which has both a favorable in dustry mix (especially construction, service, finance, and related sectors) and favorable local conditions. In con trast, regions like Northern Ireland, Scotland, and Wales suffer from a combination of both an unfavorable in dustry mix and unfavorable local conditions for new firm formation. Often the industry mix component dominates the local conditions component in statistical analyses of determinants of regional firm formation rates.

Entrepreneurship as an Organizational Product

Although some individuals become successful entre preneurs without related prior experience, they are the exception rather than the rule. Entrepreneurs are often organizational products, that is, they spin off a firm from their previous employer. Many entrepreneurs are char acterized by ‘sectoral inertia’, that is, they start their firm in an industry with which they already were familiar. Far from the universal choice, entrepreneurial action is relatively constrained: instead of looking around to seek the most profitable opportunity, the potential entre preneur concentrates his attention on a familiar sector. A person working in an industry is more likely to identify a market gap, irrespective of the degree of industry com petition and growth prospects.

This prior experience and personal networks are likely to be acquired during the entrepreneur’s career in existing organizations. This explains why the nature and number of organizations in a region are important de terminants of entrepreneurship in that region.

Empirical research has shown that a population of business organizations dominated by small firms is a fertile breeding ground for new firms. The greater the proportion of an industry’s labor force with direct ex perience of working in smaller firms, the more wide spread the propensity for self employment and hence the greater the propensity to start a new firm. Industries dominated by large plants would be expected to perform poorly as incubators of new business founders.

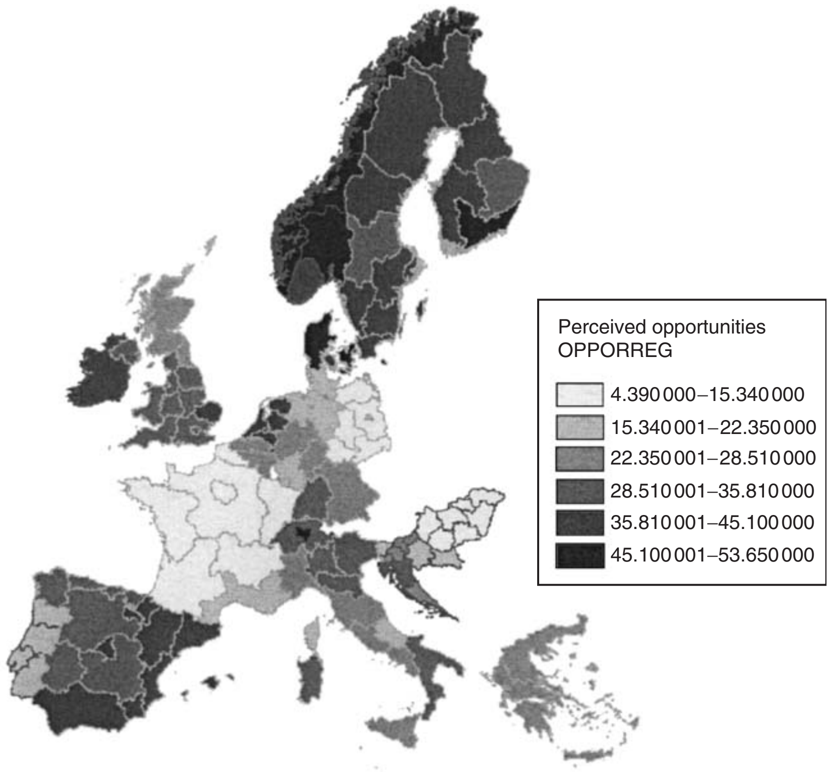

Figure 3 Perceived opportunities to start a business in local area, percentage of population between 18 64, 2001 05. (Source: Global Entrepreneurship Monitor.)

Clusters

Regions that have a concentration of organizations of a certain kind (in a specific industry) tend to generate a relatively large number of new organizations of that same kind. This pattern has been shown for industries in general, and for specific industries like footwear, ac counting, biotechnology, computer workstations, and motorcycles. There are several reasons for why indu strial clusters foster entrepreneurship. Clusters provide established relationships and better information about opportunities. They open up niches of specialization due to the low degree of vertical integration. Clusters foster a competitive climate and strong rivalry among firms that impose pressure to innovate due to the presence of close competitors. They provide role models with the presence of other firms that have 'made it', and a cultural en vironment where establishing one's own business is normal and failure is not a social stigma. Clusters provide access to physical, financial, and commercial infra structure; easing the spin offs of new companies from existing ones. Especially because potential entrants will know how the local industry functions and have the technical skills to operate in it.

Existing organizations expand the pool of potential entrepreneurs available in a region by employing individuals in roles conducive to acquiring information about entrepreneurial opportunities and to developing the social ties necessary for resource mobilization. This all lowers entry and exit barriers and reduces the risk and uncertainty for aspiring entrepreneurs.

Urbanization

Urban areas have important advantages for entrepre neurship. These include the relative ease of access to customers, as well as the inputs required (capital, labor, and suppliers) to produce the goods or services. The classical 'incubation hypothesis' in urban economics states that persons aspiring to go into production on a small scale have found themselves less obviously barred by a high cost structure at the center of the urban area than at the periphery. Urbanized areas are often con centrations of educated individuals with business ex perience in their early and middle adult years, and in that way they are a source of entrepreneurs. Urban areas have important advantages for the demand for entrepreneur ship. The risk of starting a business in urban areas is also relatively low due to the abundant employment oppor tunities, which function as an occupational buffer for the entrepreneur when the firm fails.

There is systematic evidence that growth in demand, measured with either population growth or income growth, positively affects the rate of new firm formation. As demand increases, more new firms will be founded to satisfy this greater demand. Also increased diversification of consumer demand positively affects entrepreneurship. This latter phenomenon is central in flexible specialization theory, which explains such trends in terms of the breakup of the mass market for standardized goods and services and the consequent emergence of a variety of smaller niche markets capable of exploitation by new or small businesses. This growth of demand is likely to be associated with urbanization. Metropolitan areas are often concentrations of (advanced) demand, providing relatively many niches to be served by entrepreneurs (better matching of demand and supply).

Recent research has also acknowledged the importance of urban areas for knowledge spillovers, that is, the relative abundance of new knowledge that might be an important input for entrepreneurial opportunities.

Culture

The regional culture is also important in the supply of entrepreneurship via its effect on the attitude and values that people acquire. Social psychologists have claimed that an individual's attitudes and traits are not inherited but are developed in interaction with the social environment. Perceptions about the desirability of becoming an entrepreneur are formed and revised given the set of information available to each person. The existence of a number of entrepreneurs in a region also legitimizes the activities of nascent entrepreneurs. Regional differences in culture in that way affect the level to which people aspire and think about becoming an entrepreneur, which is an important phase in the process of starting a firm. One example of this is the fear of failure that might deter people from starting a new firm. Cognitive theories have proposed that by observation of entrepreneurial activities in its immediate environment, an individual acquires information and skills that might trigger and enable his/her choice for an entrepreneurial career.

Distinct local cultures can have effects on different types of entrepreneurship. For example, a 'self employment' and a 'career' life mode have different positive effects on entrepreneurship, while a 'wage earner' life mode has a clear negative effect on entrepreneurship. In the self employment life mode the dominant job related motivation is to own the means of production and control the production process. This culture is carried over from generation to generation, and is most frequently found in rural areas characterized by independent and self reliant small scale farmers, artisans, and small business owners. It is rare in areas dominated by large scale operations. The dominant value in the wage earner life mode is the sale of one's labor at the highest possible price in order to maximize the utility of one's leisure time. Such individuals are unlikely to set up new businesses, except when they are 'forced' by unemployment. This life mode is likely to be most common in regions characterized by a narrow industrial base and dominated by large externally owned firms. Finally, the dominant value of individuals with a career life mode is the advancement of their career. They are likely to be well educated and working in large private or public sector organizations. They will start their own businesses if this becomes the best way in which to benefit from their skills, knowledge, and expertise. These businesses are often technologically advanced, innovative, and with good marketing capabilities. Career mode entrepreneurs are often concentrated in large metropolitan areas and smaller attractive cities. This career life mode resembles the life mode of the so called creative class to a large extent. The creative class consists of individuals with relatively high levels of creativity in their work. These individuals are relatively more likely to recognize and realize entrepreneurial opportunities. A spatial concentration of creative class has been shown to positively affect new firm formation rates. Indirectly, amenities in regions affect new firm formation rates, by attracting the creative class.

Unemployment

During the economic downturn in the 1980s there was much attention for the relation between unemployment and new firm formation. It was suggested that when workers are unemployed they might be more likely to start their own business out of necessity. As the unemployment rate rises there is likely to be some increase in the number of potential entrants and, other things equal, a greater propensity to enter.

However, the relationship between unemployment and entrepreneurship is more complicated, as high levels of unemployment are likely to indicate a reduction in aggregate demand (and thus demand for entrepreneurship), thereby putting downward pressure on the rate of new firm formation. Empirical research shows both effects: time series analyses point to unemployment being positively associated with new firm formation, whereas cross sectional studies appear to indicate the reverse. The latter outcome depends on the initial unemployment levels in a region. In regions with low unemployment, relative self employment rises along with relative unemployment, indicating that push effects are dominant. In regions with already high unemployment, relative self employment falls if relative unemployment rises, indicating the dominance of prosperity pull effects.

Growth in (Scientific) Knowledge

New knowledge created at universities and research centers generates opportunities for entrepreneurship, especially in high tech industries. Often these organizations are not able to fully recognize and/or appropriate the ensuing opportunities to commercialize that knowledge. Knowledge workers in these organizations respond to opportunities generated by new knowledge with starting a new firm, and in this way appropriate the expected value of their endowment of knowledge. Geographic proximity to these sources of new knowledge is an asset, if not a prerequisite, to entrepreneurial firms in accessing and absorbing spillovers from universities and research centers. However, geographical proximity should not be seen in a too local sense, as these knowledge spillovers are said to extend up to more than 100 miles.

The degree to which technological change promotes new firm formation (in high tech industries) depends on the institutional environment. The institutional setting affects the nature of technical labor markets, VC markets, and the structure of buyer–supplier ties that are highly relevant for the incentive constraints and appropriability constraints acting on incumbent and start up firms, respectively. For example, institutions enabling a fluid labor market, a well developed VC market, and loose buyer–supplier ties allow new firms to rapidly assemble and deploy experienced engineering talent, and move quickly to commercialize advanced technology. Cross country research has shown that this latter situation can be found in the US, in contrast to Japan, and cross regional research found, for example, that Silicon Valley is much more conducive to new technology based firms than Route 128 (Massachusetts) for similar reasons.

The institutional environment also affects the opportunity costs involved in leaving a (relatively secure) job at a university or research center for self employment.

Financial Capital

Liquidity constraints are an important factor disabling entrepreneurs to realize their business opportunities. This is especially relevant for large new firms that require relatively large scale investments for their initial activities. Small scale firms can often be financed with bank loans or the support of the entrepreneur's family and friends. The availability of collateral, most often in the form of the entrepreneur's own housing, is shown to be the single most important source of such collateral. Indirectly, fluctuations in the local housing market could thus affect the availability of financial capital for new firms. New firms that require large scale investments are more likely to enter the VC market. Providers of VC provide not only financial capital, but also knowledge of developing a business. The provision of financial capital in general is more likely to be bounded to national scales, while the provision of VC is most likely constrained to relatively large regions. The supply of VC is distributed evenly across regions. For example, the VC market in the US is highly concentrated (both in supply and investments) in the east and west coasts of the country, and in the UK it is highly concentrated in the Southeast, in and around London. This uneven regional spread has also been found in other countries. This means that in regions far away from these centers entrepreneurs might be discouraged to start capital intensive start ups. The assumption is that spatial proximity may be necessary for the formation of a VC relationship and that it makes monitoring of investments easier. Face to face contacts between the entrepreneur and the VC provider are necessary to identify the value of the new business and the involvement in business affairs by the VC provider. These contacts are hard to initiate and sustain over a large distance. Recent evidence shows that most of these contacts cannot easily be maintained over a longer distance via telecommunication, as this is only used as a complement to face to face contacts, not as a substitute.

Conclusion

The regional heterogeneity in the determinants of entrepreneurship leads to an even higher heterogeneity in the spatial distribution of entrepreneurship, as can be seen in **Figure 4**. This figure shows the variation in nascent entrepreneurship and young businesses over European regions. In order to explain the regional variations in entrepreneurship rates, both the factors related to the supply of and the factors related to the demand for entrepreneurship should be taken into account.

This article has discussed several important factors in the supply of entrepreneurship, like the industry structure, the population of organizations, and clusters in a region, providing a breeding ground for potential entrepreneurs. Another important factor in the supply of entrepreneurship is the regional culture, which explains the preference for entrepreneurship. A more ambiguous factor is unemployment: unemployment could make self employment a relatively attractive option (necessity entrepreneurship), but unemployment could also decrease the demand for products of new firms and thus also lower the demand for entrepreneurship. Other important factors in the demand for entrepreneurship are the expanding knowledge base of a region, which increases the number of technological inventions to be commercialized by new firms. Urbanization is also likely to increase the demand for entrepreneurship, as it is generally correlated with an increasing diversification

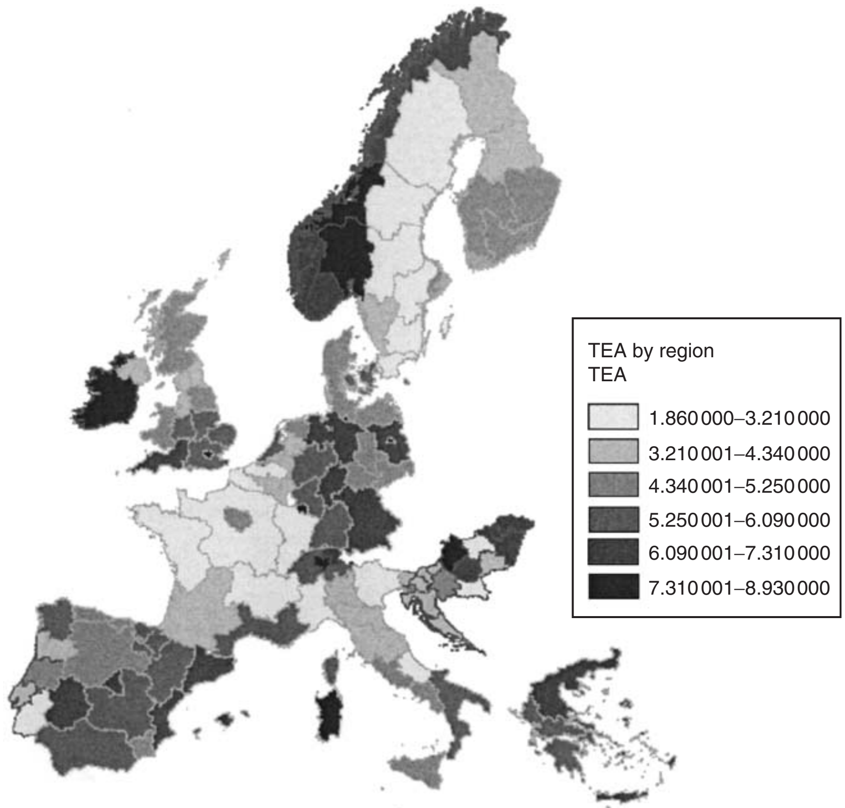

Figure 4 Early-stage entrepreneurial activity (TEA), percentage of population between 18 64, 2001 05. (Source: Global Entrepreneurship Monitor.)

of demand. Finally, the relatively abundant supply of (venture) capital is likely to lower the barriers to develop a new business.

See also: Agglomeration; Capitalism; Competitiveness; Creativity; Firms; High-Tech Industry; Industrial Location; Industrial Organization; Industrial Restructuring; Innovation; Labor Market; Local Economic Development; Location Theory; Networks; Public Policy; Technological Change; Technology and Regional Development; Technology Industries; Venture Capital.

Further Reading

Audretsch, D. B., Keilbach, M. C. and Lehmann, E. E. (2006). *Entrepreneurship and economic growth*. New York: Oxford University Press.

Bosma, N. and Schutjens, V. (2008). Mapping entrepreneurial activity and entrepreneurial attitudes in European regions. *International Journal of Entrepreneurship and Small Business* 5, Forthcoming.

Feldman, M. P. (2001). The entrepreneurial event revisited: Firm formation in a regional context. *Industrial and Corporate Change* 10, 861 891.

Hanson, S. (2003). Geographical and feminist perspectives on entrepreneurship. *Geographische Zeitschrift* 91, 1 23.

Hébert, R. F. and Link, A. N. (1989). In search of the meaning of entrepreneurship. *Small Business Economics* 1, 39 49.

Malecki, E. J. (1997). Entrepreneurs, networks, and economic development: A review of recent research. *Advances in Entrepreneurship, Firm Emergence and Growth* 3, 57 118.

Reynolds, P., Storey, D. and Westhead, P. (1994). Cross national comparisons of the variation in new firm formation rates. *Regional Studies* 28, 443 456.

Saxenian, A. (1994). *Regional advantage: Culture and competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.

Wagner, J. and Sternberg, R. (2004). Start up activities, individual characteristics, and the regional milieu: Lessons for entrepreneurship support policies from German micro data. *Annals of Regional Science* 38, 219 240.

Zucker, L. G., Darby, M. R. and Brewer, M. B. (1998). Intellectual human capital and the birth of U.S. biotechnology enterprises. *American Economic Review* 88, 290 306.

Relevant Website

http://www.gemconsortium.org
Global Entrepreneurship Monitor (GEM).

Entropy-Maximising Models

R. Johnston, University of Bristol, Bristol, UK
C. Pattie, University of Sheffield, Sheffield, UK

Data Paucity and Ecological Inference

In many areas of social science the data needed for testing hypotheses are unavailable and bespoke data collection (by conducting surveys, for example) is not feasible – very often because of finance. Researchers are thus forced to adopt 'second best' strategies, using whatever data are available to generate approximations of the information that they need. Such approaches necessarily carry caveats – these are the best answers that we can provide to our research questions given the available data and our ability to manipulate them.

One area of work in which such methodologies are frequently deployed goes under the collective title of 'ecological inference'. It involves deriving estimates of 'unknown quantities' – such as the number of households of different sizes with two cars in each of a series of areas – from what is known about those areas (the number of households of different sizes in each and also the number of households with two cars – but not the cross classification of the two).

Entropy

A form of ecological inference developed and applied by human geographers is entropy maximizing. 'Entropy' – a concept developed in thermodynamics but since widely applied in information theory – is a measure of the amount of uncertainty in a probability distribution, subject to constraints. The greater the uncertainty, the greater the entropy. Thus, for example, if you have a population of 100 persons, 98 of whom are male, the probability of randomly selecting a male is very high – or the uncertainty associated with the outcome of such sampling is very low. If 48 are male, however, the uncertainty is much greater – and hence the entropy is larger.

Entropy is linked to another important mathematical–statistical concept – 'maximum likelihood'. In the study of probabilities, the maximum likelihood estimate is that most likely to occur. (In the example in the preceding paragraph, with 98 males out of 100 the maximum likelihood expectation is that a randomly drawn individual will be male.) Mathematically, the two procedures – entropy and maximum likelihood – are equivalent.

Entropy maximizing methods have been used in two major areas of research in spatial science. These are discussed here, following an introductory section setting out the mathematics behind the procedures.

The Entropy-Maximizing Procedure

Based on the thermodynamic concept of entropy as maximum disorder, the entropy maximizing procedure (EMax) seeks the most likely configuration of elements within a constrained situation. It is widely used in a range of disciplines, but its potential for the social sciences has been only slightly recognized. It can be illustrated by a simple example.

Take a 2×2 ($N \times M$) matrix, in which the row and column totals (shown by the $\sum$) are known but the internal cell values are unknown (the ?):

	M_1	M_2	$\sum$
N_1	?	?	4
N_2	?	?	6
$\sum$	3	7	

This is termed the system's 'macro state'.

That macro state can be generated by a number of 'meso states', distributions of individuals across the four internal matrix cells which are consistent with the system constraints – the macro state's row and column totals. There are four feasible meso states:

	M_1	M_2	M_1	M_2	M_1	M_2	M_1	M_2
N_1	3	1	2	2	1	3	0	4
N_2	0	6	1	5	2	4	3	3

EMax identifies which of those meso states is most likely to occur – given no other constraints or influences on which individual is allocated to which cell. To do that involves identifying the 'micro state' most likely to occur because it has the largest number of different allocations of the individuals in any one meso state. For example, the first meso state above has four individuals in row N_1 – *a*, *b*, *c*, and *d*, say. There are four different ways in which they can be allocated to the two cells in that row. (1) *a* appears in column M_2 and the other three, *b*, *c*, and *d*, in column M_1; (2) *b* in M_2, *a*, *c*, and *d* in M_1; (3) *c* in M_2, *a*, *b*, and *d* in M_1; and (4) *d* in M_2, *a*, *b*, and *c* in M_1. Whichever allocation occurs, all six of the individuals in row N_2 (*u* *z*, say) must be allocated to column M_2. Hence there is a total of four micro states associated with the first meso state.

The second meso state has six different ways of allocating the four individuals in row N_1 to the two columns (the first pair refers to M_1 and the second to M_2): (1) *ab*, *cd*; (2) *ac*, *bd*; (3) *ad*, *bc*; (4) *cd*, *ab*; (5) *bd*, *ac*; (6) *bc*, *ad*. In addition, there are six different ways in which the six individuals in row N_2 can be divided between the two columns in the ratio of 1:5 ([u; $v-z$], [v; u, $w-z$], etc.). This gives a total of 36 different micro states associated with that meso state (i.e., the product of the two separate allocations). There are 60 different micro states associated with the third meso state (four different ways of splitting the four individuals in N_1 in the ratio 1:3 and fifteen of splitting the six in N_2 in the ratio 2:4) and 18 associated with the fourth (there are 18 ways of splitting the six individuals in N_2 in the ratio 3:3 and just one way of allocating the four in N_1 in the ratio 0:4).

The maximum likelihood solution is the third of the four meso states: this has the most micro states associated with it and is thus the most likely to occur. If all we know is that ten individuals are distributed across the four cells of the macro state, then the most likely meso state associated with that configuration is the third of the four identified, because it has the largest number of separate ways (micro states) in which it can be created, given the constraints of the row and column totals.

The Mathematical Basis

Application of the method here involves a three dimensional data cube for which the three sets of sums are known. Assume a matrix V with m rows and n columns which applies to the entire country (this is matrix ${}_mV_n$). That country is divided into k districts, each with its own matrix ${}_mv_n$. Thus v_{ijk} is the value for the ith row, in the jth column, and the kth district and $\sum v_{ijk} = V_{ij}$. If the total number of individuals in the country is T, then the number of different combinations of individuals in the first cell – v_{111} – of the data cube is

$$T!/[v_{111}!(T - v_{111})!]$$

For the next cell – v_{112}, say – the number of combinations is

$$\{T - v_{111}\}!/[v_{111}!(T - v_{111} - v_{112})!]$$

The procedure continues to identify all the possible ways of allocating T individuals across the ijk cells, until all of the cells are filled. The total number of possible combinations is W, defined as

$$W = T!/\prod v_{ijk}!$$

which is the factorial of the total number of individuals divided by the product of the factorials of all of the individual v_{ijk} values. Stirling's approximation can be used to derive this as

$$\log W = (T \log T - T) - \sum (v_{ijk} \log v_{ijk} - v_{ijk})$$

This is an unconstrained solution, however. We need the cell values to sum to the relevant row and column totals – $v_{i.k}$, and $v_{.jk}$. Thus $\sum v_{ijk} = v_{i.k}$, over all m rows in each district and $\sum v_{ijk} = v_{.jk}$, over all n columns in each district.

In addition, the sum for any one cell across all districts must equal the sum in the national matrix (V: the contents of this national matrix may be derived from a variety of sources, for instance from a special census tabulation or – commonly – from the results of a national random survey):

$$\sum v_{ijk} = V_{ij}, \quad \text{over all } k \text{ districts}$$

These three constraints are then used to limit the distribution of all T individuals across the v_{ijk} cells of the data cube, giving a statistic L, defined as

$$\begin{aligned} L = (T \log T - T) &- \sum (v_{ijk} \log v_{ijk} - v_{ijk}) \\ &+ \sum \lambda_{ij}(V_{.j} - v_{ij.}) + \sum \gamma_{ik}(V_{i.} - v_{i.k}) \\ &+ \sum \Phi_{jk}(V_{ij} - v_{ijk}) \end{aligned}$$

in which the three parameters λ, γ, and Φ constrain the distribution of T to the known totals

$$\exp(-\lambda_{ij}) = A_{ij}V_{ij}$$

$$\exp(-\gamma_{ik}) = B_{ik}V_{i.}$$

$$\exp(-\Phi_{jk}) = C_{jk}V_{.j}.$$

These are scalars for each row in each state (B_{ik}), each column in each state (C_{jk}), and for each cell in the national matrix (A_{ij}) which ensure that the constraints are met. The maximum likelihood value for each allocation of T across the v_{ijk} cells is then calculated as

$$v_{ijk} = A_{ij}V_{ij}B_{ik}V_{i.}C_{jk}V_{.j}$$

Fitting the Model

Calculation of these v_{ijk} values is achieved using the iterative biproportionate matrix scaling procedure (or 'Mostellerisation'). The three dimensional data cube is assembled, and initial estimates of the internal cell values – that is of each v_{ijk} value – are added: the easiest way of starting this to designate all cell values as 1.0. These are then summed across one of the dimensions in the cube, and those sums are compared with the relevant constraint values (the known values). If the rows of the matrix for each district were chosen, this would give a matrix of row estimates (ER)

$$\mathrm{ER}_{i.k} = \sum v_{ijk} \quad \text{for all } j \text{ columns in the } k\text{th district}$$

These are then compared with the relevant constraint for each row, and a row scalar (SR) produced as a ratio be tween the sum of the estimates and the actual value:

$$\mathrm{SR}_{i.k} = V_{i.k}/\mathrm{ER}_{i.k}$$

The cell values are then multiplied by the relevant scalar to give a new estimate of their values

$$e_{ijk} = v_{ijk} * \mathrm{SR}_{i.k}$$

These are then summed across the other two dimensions of the cube, to produce estimates of the column sums (matrix EC) and the slice sums (matrix ES). These are compared with the relevant constraints, to produce a new set of scalars (SC and SS):

$$\mathrm{SC}_{.jk} = V_{.jk}/\mathrm{EC}_{.jk}$$

$$\mathrm{SS}_{ij.} = V_{ij.}/\mathrm{ES}_{ij.}$$

The process is then repeated by producing a new set of cell value estimates – e_{ijk} – using one of these two scalars – either

$$e_{ijk} = e_{ijk} * \mathrm{SC}_{.jk}$$

or

$$e_{ijk} = e_{ijk} * \mathrm{SS}_{.jk}$$

The decision on which to use is determined by a goodness of fit statistic, for instance the average difference between the estimated and constraint sum, expressed as a percent age of the constraint sum.

The procedure continues until a stopping rule is ap plied – either after a set number of iterations or when each of the two goodness of fit statistics is below a pre defined maximum – an average of 1%, perhaps. (The third will always be zero, since the cell values have been scaled to fit that set of constraints.)

Gravity Models and Transportation Flows

The first use of entropy maximizing models in geog raphy was by Alan Wilson, in research on transportation planning. Before his work, most researchers had applied the ‘gravity model’ – based on an analogy with Newton’s theory of interaction between two bodies separated by distance. This had the form

$$T_{ij} = f(X_i,\ Y_j,\ d_{ij}) \tag{1}$$

where X_i is the number of travelers originating in area i; Y_j is the number of travelers whose destination was in area j; d_{ij} is the distance between areas i and j; and T_{ij} is the total volume of movement originating in i and moving to j.

As X and Y are assumed to have a positive impact on the volume of movement and d a negative effect (i.e., distance is a deterrent to moving), this model was usually tested (and applied in planning studies) as a regression equation with the following form:

$$T_{ij} = a[\{(b_1 X_i) * (b_2 Y_j)\}/\{d_{ij}{}^{e}\}] \tag{2}$$

However, Wilson pointed out that this model does not conform to the method of bounds, and can generate ‘nonsense’ estimates/predictions – for example, more people traveling to j than there were destinations there. Thus, gravity models fitted using eqn [2] could be of little use in estimating travel flows in models developed for planning purposes.

In their place Wilson proposed a constrained gravity model in which the estimated/predicted total outflow from each X equals the total number of travelers known to originate there, and the estimated/predicted total in flow to each Y equals the known destinations there. (In a traffic model, for example, X may be the number of workers living in each zone and Y the number of jobs there.) The macro state was thus a matrix with N rows (the origin areas) and M columns (the destinations), and the constraints were the row and column sums. A large number of meso states would be associated with this, each of which had an associated set of micro states. The EMax procedure identified the meso state with the largest number of micro states as the most likely pattern of flows, given the two sets of constraints.

A third constraint was then added, the distance de terrence factor – the total number of miles traveled, for example – which, in predictive models especially, would result in the least overall effort solution (which would be the same as the maximum likelihood estimate). The model thus allocated movements between home and workplace areas so that no area generated more or less journeys than there were workers living there, no area attracted more or less workers than there were jobs there, and the total number of miles traveled did not exceed the third constraint. It was the ‘least effort solution’ given the constraints.

The constrained gravity model was thus an extension of the ‘normal’ specification (model [1] above) with the form

$$I_{ij} = f(X_i,\ A_i,\ Y_j,\ B_j,\ e^{-\beta dij}) \tag{3}$$

In which the A_i and B_j are balancing terms to ensure that the twin constraints – the number of workers and jobs – are met, and $e^{\beta d_{ij}}$ is a multiplier which ensures that the total amount of interaction does not exceed the sum of miles traveled in the entire system. This model cannot be rewritten as a regression equation (as in eqn [2] above) but instead has to be fitted by the iterative biproportional method described above (or some equivalent procedure).

Wilson's goal – and that of the large research teams who worked with him at Leeds in the 1970s–1980s – was not to test hypotheses regarding spatial structure and spatial behavior (as was the case in much of the 'quantitative and theoretical geography' being undertaken at the time), but rather to build models from available data to indicate not only the most likely patterns given current constraints but also those which would emerge if the constraints were changed, by forecasting future patterns of residential and workplace location, and from these – using EMax methods – to deduce the most likely flow patterns on which planning (of transport routes and service provision, for example) could be based.

A detailed example of this approach was provided in a book length study of the West Yorkshire region in the UK. Separate strategies were used to model the region's future population geography, plus the location of work places, shopping and other services, and of demand for those services. This geography of supply and demand for transportation was then fed into EMax models to derive predicted trip patterns, by type and by mode of transport ('modal split'). Further work extended these exercises in a variety of ways (by, for example, subdividing households according to their income and other characteristics in order to predict their trip behavior).

Geographies of Voting

The only other research question for which EMax has been widely deployed by geographers concerns spatial patterns of voting at the individual scale. Whereas election returns provide aggregate data, answering the question 'how many people voted for what, where?', they almost invariably fail to answer questions such as 'what type of people voted for what, where?'. In such cases, researchers have had to rely on methods of ecological inference, estimating answers to the second questions with data that only explicitly answer the first. There is a long history of exploratory analyses addressing this issue: King claimed that he had solved the ecological inference problem, but later debate has shown that although his EI procedure has made major advances possible in certain situations (especially for the simplest problems, involving sets of 2 × 2 matrices), many technical problems remain and alternative procedures continue to be sought.

Spatial Variations in Voting Behavior

The use of EMax to address this issue was first promoted some 15 years before King's work. It has since been used to tackle a substantial number of issues regarding voting patterns in the UK, Australia, New Zealand, the USA, and Germany, but has not been widely adopted by researchers other than the original authors.

The original problem which led to the development of this EMax application was an observation that the concept of uniform swing then widely employed in the analysis of changing voting patterns in the UK was both theoretically and empirically untenable. If party *A* and *B* received 54% and 46% of the votes cast nationally at the first election in a pair and 52% and 48% at the second, this involved a net swing of 2 percentage points away from *A* and to *B*. Empirical analyses showed that a similar swing occurred in most of the country's Parliamentary constituencies, which led to the inferred conclusion of a uniform political culture throughout the country: people were moving away from *A* towards *B* in similar proportions everywhere.

This could not be the case, however. If parties *A* and *B* get 72% and 28% of the votes respectively in constituency *X* at the first election, then with a uniform swing the allocation at the second will be 70 and 30 – with *A* losing the support of 2.78% of those who voted for it at the first contest. In constituency *Y*, where *A* and *B* got 30% and 70% respectively of the first election votes, on the other hand, the result of the swing would be 28 and 72 at the second, with *A* losing 6.67% of its supporters to *B*. There was thus a paradox: a uniform absolute change was produced by a non uniform relative change. But what was the pattern of shifts between the two elections across the country's constituencies? The EMax matrix for each constituency can be written as

	Election 2		
Election 1	*A*	*B*	$\sum$
A	*u*	*u*	*k*
B	*u*	*u*	*k*
$\sum$	*k*	*k*	*k*

The cells represented by *k* indicate known quantities (the number of votes for each party at each of the two elections), whereas those represented by *u* indicate unknowns. EMax can be deployed to estimate the *u* values for each constituency, and thus portray variations across constituencies in what is known as the flow of the vote. If each *u* value is expressed as a percentage of the row $\sum$, then the value of the top left cell shows – for example – the percentage of those who voted for party *A* at the first election and who voted for it again at the second, whereas that in the top right cell shows the percentage who switched from *A* to *B*. The national constraint (represented by V_{ij} in the above algebra) is normally provided by evidence from survey data, which allows the estimation of the total number of voters in each cell of the national transition matrix.

Further observations identified another issue. Analysts assumed that the pattern of voting by class – the dom inant cleavage in UK politics at the time, with middle class voters tending to vote Conservative and working class voters preferring Labour – was the same every where. If the national matrix of voting by occupational class was applied to each of the country's constituencies (using census data), however, this produced very poor predictions of the election outcome in most. The con clusion was that similar people must vote differently in different places – but how differently?

Both issues are variations on the same ecological problem. We have a three dimensional data matrix, in which each slice is an area – in the UK, a Parliamentary constituency. If each (k) of those two dimensional matrices comprises $N=5$ rows (e.g., a division of the electorate into five social classes) and $M=4$ columns (i.e., the three parties contesting each election, plus ab stentions), then we know the row sums (the number of people in each social class) and the column sums (the number of votes for each party), but not the matrix's internal cell values. For the country as a whole we also know, from voter surveys, the estimated contents of that 5×4 matrix summed over all k constituencies. Using the method of bounds – the foundation of King's EI method also – we know that:

1. In any one constituency, k, the number of people in each social class voting for the four parties must equal the total number of people in that class (the known row total) (throughout, we treat abstention as a party for ease of discussion).
2. In any one constituency, k, the number of votes for each party across the five social classes must equal the recorded total number of votes cast for it (the known column total),
3. Across all constituencies, $\sum k$, the number of people in each social class voting for each party must equal the national number of people in that matrix cell (i.e., the known number of votes cast for each party by each class as divulged by the voter survey).

The technical problem was how to use the three sets of known values to estimate the unknown values – the in ternal cells of the matrix for each constituency – so that they were consistent with those known values.

Wilson's EMax approach was adapted and presented in a series of papers. A number of applications to UK electoral analysis followed, looking at spatial variations across constituencies in both the pattern of voting by class and the flow of the vote between elections. The conclusion was clear: there are very substantial spatial variations across the UK in patterns of voting by class and in inter election voter transition matrices.

One specific application of EMax in recent years has been in the study of split ticket voting, as in the United States where at any one election voters may have to choose both a President and a Representative and at MMP elections in Germany, New Zealand, Scotland and Wales in which voters cast ballots both for a constituency MP and a party list. Some may vote a straight ticket – that is, for the same party's candidates for both President and the House, or for both a constituency member and her/his party in the list contest – whereas others may vote a split ticket. EMax has been used to estimate spatial variations in the volume of split ticket voting (as in Germany), and the accuracy of those estimates has been evaluated against New Zealand data, thereby enhancing the method's credibility.

Conclusions

Entropy maximizing has been deployed in the study of two major areas of research in human geography, but although its potential usage across the spatial sciences, where the issue of ecological inference arises, is wide, it has received little attention outside those two appli cations. Apart from an early paper on its value in mi gration studies, for example, it was not deployed for studies using census data until Simpson and Tranmer's paper on using the iterative proportional fitting pro cedure (the analytic algorithm – Mostellerisation – de ployed to fit EMax, as indicated above) to estimate car ownership levels by housing tenure in small areas using census data.

The EMax approach is just one of several promoted to resolve the missing data problem which besets many researchers – others include micro simulation, which is based on the same principles, and linear programming (which produces the same outcomes as EMax in certain circumstances): Wilson has also argued that EMax out comes can be interpreted using the economic concept of consumer's surplus: maximum likelihood estimates also maximise that surplus – the economic gain accruing to consumers when they engage in trade (or, in this case, interaction). It allows researchers to estimate, with some certainty, aspects of human spatial behavior which would otherwise remain unknown (and recent work has ex tended the method by calculating error terms for the estimates), thereby not only uncovering patterns that would otherwise remain hidden but also to suggest so lutions to problems of spatial allocation and planning.

See also: Electoral Geography; Quantitative Methodologies; Transport Geography.

Further Reading

Bacharach, M. (1970). *Biproportional Matrices and Input Output Change*. Cambridge: Cambridge University Press.

Berrington, H. (1965). The general election of 1964. *Journal of the Royal Statistical Society, Series A* 128, 17 66.

Chilton, R. J. and Poet, R. W. W. (1973). An entropy maximising approach to the recovery of detailed migration patterns from aggregate census data. *Environment and Planning A* 5, 135 146.

Cleave, N., Brown, P. and Payne, C. (1995). Evaluation of methods for statistical inference. *Journal of the Royal Statistical Society, Series A* 158, 55 72.

Dempster, A. P., Laird, N. M. and Robinson, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B* 39, 1 38.

Elff, M., Gschwend, T. and Johnston, R. J. (2008). Ignoramus, ignoramibus: The principle of maximum entropy and uncertainty in ecological inference for RxC tables. *Political Analysis* 16.

Evans, S. P. (1973). A relationship between the gravity model for trip distribution and the transportation problem in linear programming. *Transportation Research* 7, 39 61.

Gould, P. R. (1972). Pedagogic review. *Annals of the Association of American Geographers* 62, 689 700.

Gschwend, T., Johnston, R. J. and Pattie, C. J. (2003). Split ticket patterns in multi member proportional election systems: Estimates and analyses of their spatial variations at the German federal election, 1998. *British Journal of Political Science* 33, 109 128.

Johnston, R. J. (1983). Spatial continuity and individual variability. *Electoral Studies* 2, 53 68.

Johnston, R. J. (1985). *The Geography of English Politics: The 1983 General Election*. London: Croom Helm.

Johnston, R. J. and Hay, A. M. (1982). On the parameters of uniform swing in single member constituency electoral systems. *Environment and Planning A* 14, 61 74.

Johnston, R. J. and Hay, A. M. (1983). Voter transition probability estimates: an entropy maximizing approach. *European Journal of Political Research* 11, 93 98.

Johnston, R. J. and Hay, A. M. (1984). The geography of ticket splitting: A preliminary study of the 1976 elections using entropy maximizing methods. *The Professional Geographer* 36, 291 296.

Johnston, R. J. and Pattie, C. J. (1993). Entropy maximizing and the iterative proportional fitting procedure. *The Professional Geographer* 45, 317 322.

Johnston, R. J. and Pattie, C. J. (1999). Constituency campaign intensity and split ticket voting: New Zealand's first election under MMP, 1996. *Political Science* 51, 164 181.

Johnston, R. J. and Pattie, C. J. (2000). Ecological inference and entropy maximizing: An alternative estimation procedure for split ticket voting. *Political Analysis* 8, 333 345.

Johnston, R. J. and Pattie, C. J. (2002). Campaigning and split ticket voting in new electoral systems: The first MMP elections in New Zealand, Scotland and Wales. *Electoral Studies* 21, 583 600.

Johnston, R. J. and Pattie, C. J. (2003). Evaluating an entropy maximizing solution to the ecological inference problem: Split ticket voting in New Zealand 1999. *Geographical Analysis* 35, 1 23.

Johnston, R. J., Pattie, C. J. and Allsopp, G. J. (1988). *A Nation Dividing?* London: Longman.

King, G. (1997). *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton, NJ: Princeton University Press.

King, G., Rosen, O. and Tanner, M. A. (2004). *Ecological Inference: New Methodological Strategies*. Cambridge: Cambridge University Press.

Simpson, L. and Tranmer, M. (2005). Combining sample and census data in small area estimates: Iterative proportional fitting with standard software. *The Professional Geographer* 57, 222 234.

Wilson, A. G. (1970). *Entropy in Urban and Regional Modelling*. London: Pion.

Wilson, A. G. (2000). *Complex Spatial Systems*. Chichester: Wiley.

Wilson, A. G. and Bennett, R. J. (1985). *Mathematical Methods in Human Geography and Planning*. Chichester: Wiley.

Wilson, A. G., Rees, P. H. and Leigh, C. M. (1977). *Models of Cities and Regions: Theoretical and Empirical Developments*. Chichester: Wiley.

Environment

S. Eden, University of Hull, Hull, UK

Glossary

Commons Originally a piece of land used by local people for access, food, and pasture, especially in the Middle Ages in Europe; more recently, an environmental resource open to use (and abuse) by many different groups, such as the ocean.

Cultural Turn A move in human geography toward cultural issues and approaches in the early 1990s.

Darwinism Charles Darwin's theories of evolution, especially that only the fittest survive the struggle for existence and that species (and human races) evolve in response to their environmental conditions.

Dualism A problematic way of thinking that divides everything in a set into two opposite and mutually exclusive categories, not allowing for any middle ground, exclusions, or overlaps. Examples are nature–culture, mind–matter, human–nonhuman, and female–male.

Ethical Turn A move in human geography toward ethical issues and approaches in the late 1990s.

Humanistic An approach in human geography that values the empirical study of human subjective experience, feeling, and thought, usually at the level of the individual.

Instrumental Value/Thinking Where the environment is valued or thought about only in terms of its practical use or as a means to an end, rather than the end itself.

Intrinsic Value/Thinking Where the environment is valued or thought about in terms of its existence, as an end in itself, rather than a means to an end.

Environment

The environment is the context in which things happen: it makes us what we are but we also change it through everything we do. This relationship between us and our environment is essential to life, but it is also a challenging and complex thing to manage. Moreover, how we think about the environment is often dependent on the times we live in.

We can understand this better if we consider how people in different places and times made sense of their environments and of environmental change. Assumptions and preconceptions about environmental influence and change are central to human geography, but are often buried deep within our collective psyche. Instead, we need to bring them into the light to recognize how they shape our thinking about environmental issues. Human geography research suggests that we can group these diverse ways of thinking about the environment into three: the environment as a powerful resource outside society; the environment as fragile and damaged by society; and the environment as part of society. This article examines each of these in turn to show how they arose historically, and then considers some themes that cut across these and are particularly important in recent times.

The Environment as a Powerful Force and Resource Outside Society

The first way of thinking about the environment is as an external force or object. We might call this a 'modern' way of thinking about the environment, because, unlike previous 'premodern' thinking associated with undeveloped cultures, it sees the environment as outside society, as the external, inanimate context within which society must work, rather than as intimately intertwined with human life. This arose during what is variously referred to as the Scientific Revolution, the Age of Enlightenment, and the Age of Reason in medieval Europe. This started well away from the environment with astronomy, when Copernicus, Galileo, and others argued that the Sun, not the Earth, was at the center of the universe, challenging the long held dogma that the Earth was the center, and thus main purpose, of creation. However, such early science was also a search for laws to explain how the environment worked, and gradually led to the environmental sciences with which we are familiar today.

The scientific method is also associated with objectivity and neutrality in understanding the environment. In the seventeenth century, René Descartes theorized that 'mind' and 'matter' were different, because 'matter' was subject to primary observation, being measurable, empirical, objective, or otherwise corporeal, and 'mind' was only subject to secondary observation through logic or theory, not being measurable but theoretical or subjective (hence his famous aphorism: *I think, therefore I am*). This division of mind from matter – now referred to as 'Cartesian dualism', after Descartes – still underlies the division often assumed today between the natural or physical sciences, which are seen as providing 'hard' data and objectively true results, and the social sciences, which are seen as wooly, uncertain, and prone to subjective bias or distortion. Both of these assumptions are problematic, but they also underlie the division between physical and human geography, in that the former is dominated by natural science and the latter by social science.

More generally, accepting this division allowed people to classify the environment as merely 'matter', thus stripping animals, ecosystems, and indigenous people (who were often regarded as being nearer to nature than civilization, with typical European arrogance) of both feelings and rights. As merely 'matter', the environment became a lifeless thing, a machine, or, more in line with the thinking of the day, like clockwork, all cogs and wheels but no feelings or soul, a mechanical system that we can tinker with, model, adjust, and improve. This is sometimes referred to as a mechanistic view of the en vironment and is important because it supported later arguments about environmental exploitation, as ex plained below.

Environmental determinism

The role of the environment as an external force was also important in human geography. In the late nineteenth century, geographers were still trying to establish their subject in universities and borrowed from the more established sciences to give it more credibility and in fluence. Ideas from Darwinism were used to explain the puzzle of why human cultures differ so much, by arguing that they were produced by environmental conditions. This had long been a theme in natural history: for ex ample, the ancient Greeks thought that people who lived in cold environments, especially in the near Arctic, were fierce and cruel because they had to struggle to survive there, whereas people who lived in very hot environ ments were lazy because food was easy to come by in the lush tropics. But, like Goldilocks, they thought that people who lived in temperate environments, like themselves, were just right in terms of their morals, politics, and cultures.

Such arguments are now termed 'environmental de terminism' and underpinned much later European atti tudes toward how 'warmer climes' made people lazy and willful, such as eighteenth century views of Africans which were used to justify the slave trade and nine teenth century views of India held by British occupiers. In the nineteenth century, such notions were seen as scientific explanations for human differentiation and were widely studied, through analysis of facial charac teristics of different races, for example. Today, we can see the racism and ethnocentrism in such notions, but they also suggest how the environment was seen as a crucial force in either supporting or constraining human devel opment took a long time to ebb away. Even in the early twentieth century, geographers were debating how far environmental conditions explained social differen tiation, although they had moved away from the crudest forms of environmental determinism in favor of more complex models of how the environment and human society influenced each – an idea that.

Environmental resources for exploration and imperialism

So, science underpinned the first main way of thinking about the environment that was introduced in the be ginning of this article, particularly how the environment influenced human development. This way of thinking also saw the environment as a set of material resources for exploitation for the benefit (and profit) of growing societies, from timber, furs, meat, and land for growing good crops, to metal ores, coal, oil, and gas for fuelling industry and development. Human societies have fought wars to control environmental resources and built em pires upon them, again illustrating how the environment can be seen as a source of power. Because of this im portance, human geography has frequently studied the natural resources provided by the environment and also how best to manage and conserve them.

This way of thinking about the environment was supported by two other factors in the early modern period. First, human populations exploded in Europe, multiplying three or four times over between 1500 and 1850. Europeans began to emigrate across the world, especially to colonies in the New World, trading and empire building on the way. Often, this was prompted by the search for one's fortune or even 'paradise on earth' (in a religious evocation of the return to the biblical Garden of Eden for Christians). Such ideas were used not only deliberately in advertisements for sailors and colonizers, but also more implicitly in how people imagined other places to be – their imagined geographies often saw the environment as an inexhaustible resource which people had a (literally) God given right to exploit. This can be seen at its most extreme, perhaps, in the European col onization of North America, where European immigrants thought that the environmental resources they saw all around them were too plentiful ever to be exhausted. Consequently, the environment was widely and rapidly plundered, forests were cut down for agriculture, for settlement, and sometimes just for the fear that they would harbor wild animals or bandits or Indians bent on attack. Sometimes it was just for the fun of sport, such as massacring a flock of passenger pigeons, so that the bodies piled up in great mounds. The result of such thinking, in the case of the passenger pigeon, was that there were maybe eight billion passenger pigeons in the seventeenth century, flying 'so thick that I could see no sun', according to one person at the time, but none by 1914. This idea of having no limits to environmental exploitation proved long lived and it was not until the late twentieth century that it was seriously challenged.

Second, capitalism and industrialization expanded as part of empire building, and tapped into increasingly powerful techniques to exploit environmental resources, from smelting technologies to calculating longitude accurately for navigation, to plantation agriculture, railways, and surveying. Together with the movement of people, these knowledges and technologies fundamentally changed the environment in many countries across the world. This colonial period also saw the entrenchment of private property rights over what had often been, before then, collective environmental resources or 'commons' that could be used by anyone for collecting firewood, hunting game, or just passing through.

Such privatization of the environment supported capitalist development and also reinforced the separation of nature from growing (and urbanizing) human societies. Instead of living on the land and close to the animals and plants (e.g., often sharing buildings with livestock), people were increasingly urbanized and separated from the environment through new technologies and land use practices. By making the environment separate from and subordinate to society, and regarding it as a machine or inorganic system subject to human exploitation, some have argued that the environment ceased to be a living, symbolic entity in people's minds: a forest stopped being full of meaning and life, animals and plants, food and game, and contemplation and fear, as in premodern cultures, and instead became (merely) a quantified amount of timber for logging, selling, shipping, and processing.

This represents 'instrumental' thinking about the environment, where an environmental resource is valued because humans can make use of it directly, especially for profit, such as land, whale oil, buffalo skins, or timber. Environmental economists call this 'user' or instrumental value and it supports a way of thinking about the environment as an object, a resource for exploitation. This is not necessarily only exploitation for individual gain – it can also be 'utilitarian' in arguing that the environment should be used for the greater public good, often in supporting a growing national economy, for instance. Overall, therefore, this first way of thinking about the environment is often practical and functional. Moreover, we can still hear it used today, to justify environmental exploitation.

The Environment as Fragile and in Need of Protection

The second way of thinking about the environment comes much later and was a reaction against the first. With the rising power of human technology and the dominant Enlightenment ideal of progress, more and more commentators questioned the environment's power to shape human destiny. They also questioned whether the resources provided by the environment were truly inexhaustible and began to wonder if such resources, especially fossil fuels but even basic resources like food and water, might eventually run out. Such questions fostered ideas of environmental degradation and damage and also the need to manage the environment better, leading to the development of the modern environmental movement.

The Romantic view of nature and the antiurban backlash

But the origins of this way of thinking lie a bit further back, in a backlash against urbanization, industrialization, imperialist science, and environmental exploitation from the late eighteenth century. The 'Romantic' movement rejected urban life because of its immorality, industry, pollution, and loss of civilized values and instead sought solace in the environment, valuing an esthetic response to nature. In England, the Romantics were epitomized by William Wordsworth, and in the USA, by Henry David Thoreau and John Muir. They revered mountains, rivers, waterfalls, and moorland; praised natural forms through the arts of poetry and painting, representing the environment as powerful, forceful, and everlasting; and encouraged human humility toward nature, rather than the control and exploitation of the resource perspective.

This way of thinking can be seen as a fanciful idealization of the environment, imagining the countryside to be a pastoral 'idyll' of wholesomeness and beauty, sanitized and purified of all the city's social problems and activities. But it was influential in challenging the instrumental view of the environment because it represents a way of thinking about the environment based on 'intrinsic' value, rather than instrumental value. Also called 'existence' value by environmental economists, this is where an environmental resource is valued merely because it exists, even if it is not used as a resource nor is ever likely to be. Whales, buffalo herds, Californian redwoods, and a clear view of the stars above our heads can therefore be valued not because they can be turned to a profit, but because they have a right to exist independent of human needs and because the world would be a less meaningful place without them.

Environmental damage and the beginnings of environmental concern

As well as valuing the environment differently, this second way of thinking also takes a different perspective on human power, seeing it as capable of inflicting environmental damage. This was a challenge to the dominant ideas of environmental abundance and inexhaustibility in the nineteenth century (as outlined earlier), and turned the tables of the first way of thinking by seeing the environment as fragile and humans as powerful, capable of damaging the environment and thereby also threatening their own destruction.

Although we tend to think of concern about environmental degradation as modern, ancient Greek writers considered problems of soil erosion and laws against air pollution, especially smoke, were enacted in London, England, in the thirteenth century. Again, it was in the nineteenth century that concerns about the impact of increasing urbanization and industrialization on the en vironment and people's health solidified into campaigns for public sanitation, personal hygiene, nutrition, exer cise, and access to open space. As outlined above, the Romantic antiurban backlash was part of this process of revising society's views about the value of the environ ment, leading to early environmental groups like the 1865 Commons, Open Spaces and Footpaths Preser vation Society being developed to preserve open land for workers in cities and to correct the lack of 'wholesome' countryside. In 1895, the National Trust was set up in England to buy land and thus protect it from industrial development and, in early twentieth century in Africa, the protection of big game species for hunting prompted colonial administrations to designate land for protection. These movements, therefore, identified human activity as a powerful and frequently damaging influence on the environment, rather than solely the reverse.

A touchstone for geographers is *Man and Nature*, by George Perkins Marsh, published in 1864 in the USA, sometimes cited as the first 'modern' expression of such environmental concern. Marsh's book was before its time in challenging the common assumptions that environ mental resources were inexhaustible and that the en vironment was resilient to human impact. But although Marsh's book helped to inspire US forest protection in the late nineteenth century, it fell into neglect until re discovered in the 1930s, when another US environmental problem – soil erosion from agriculture – gained public and political attention.

Environmental strategies of preservation, conservation, and amenity use

The early environmental movement was therefore prompted by fears of losing resources, species and sport, and consequently sought to hold back the tide of human development. Attempts to do this fell into three main categories: preservation, conservation and amenity use. Preservation sought to protect land and wildlife from human exploitation by designating an area with en vironmental value (often because it was seen as 'natural') as a national park or nature reserve, thus emphasizing intrinsic value over instrumental value. The first National Park was officially designated in 1872 in Yellowstone, USA, to protect (1) geological interest, (2) esthetic interest (scenery that people liked), and (3) a preserve for wild animals, and in 1890 Yosemite National Park was the first park explicitly designated to preserve its wilderness.

A second way to protect the environment was con servation. This grew out of a move toward sustainable yields and the professionalization of forestry (under the influence of Gifford Pinchot in the USA in particular) in the late nineteenth and early twentieth century. Con servation was not merely about preserving environments, but about using them productively and efficiently, while also having an eye on their future sustainability and re plenishment. Hence, conservationists not only saw the environment as producing resources (instrumental value), but also acknowledged that these resources were not inexhaustible but had to be carefully managed to ensure they were replenished. Thus, conservation is a forerunner of many forms of state led environmental management today and also a forerunner (at least par tially) of concepts such as sustainable development.

A third way to protect the environment was to use it for public amenity or leisure purposes. The late nine teenth century saw a craze for mountain climbing in the Alps and the Rockies, and Victorian philanthropists in vested in urban parks to provide environments as a public good for their workers, but it was the inter war period that saw a big increase in interest in outdoor exercise, such as walking and cycling, across all income groups, especially in Europe. The Ramblers Association was set up in the UK in 1926 and mass trespasses by walkers in the Peak District in 1932 protested against the lack of access to private land. Here, we see two environmental ideas come into conflict: the environment as a private property resource for the owner to manage and exploit for profit and the environment as a public property re source for the nation to enjoy for the good of their mental and physical health. Both are forms of instrumental value, but they benefit quite different social groups. This interest led to a drive for public access to land and for national parks in the UK, linked after World War II to the nationalization agenda to make the UK 'a land fit for heroes' who had fought in the war. This resulted in the 1949 National Parks and Access to the Countryside Act that set up national parks in the UK for both protecting the environment as a national asset and supporting public access. But this tension between protection and use has remained problematic ever since, because it fails to re solve the inherent conflicts between environmental pre servation, conservation, and use for amenity.

The rise of modern environmentalism in science and culture

In the 1960s, we finally see the origins of what we probably think of today as 'modern' thinking about the environment – as fragile and in need of protection from (and by) humans. This was partly nurtured by science, especially ecology. Haeckel used the term 'oecologie' in 1866 to represent the Earth as a household ('oeconomie') of interrelated organisms that interact with their en vironment, but he did little more to establish the content of the science that he named. It was not until later that the term 'ecology' came to mean the interrelationship between plants, animals, and the inorganic environment Early on, it was associated with the notion of the 'climax community' as the stable state into which systems of plants and animals mature over time, unless arrested on the way by natural disaster or by human intervention, and later with the notion of the 'ecosystem' that we are familiar with today. By the 1950s, this 'new ecology' had become the orthodoxy in scientific thought and uni versities ran courses in environmental management that tended to take a mechanistic view of the environment as an ecosystemic machine that needed to be properly managed.

Ecology also lent scientific authority to notions of environmental damage and threat, by measuring the ef fects of human activity on ecological quality and species loss. Rachel Carson's 1962 book *Silent Spring*, about the effects of pesticide residues on bird populations, was a bestseller and is often credited with spurring the debate about the environmental effects of pollution, buttressed by subsequent and well publicized accidents with oil tankers, nuclear power plants, and industrial waste dumps. In contrast to earlier periods, when the en vironment was seen as powerful and its resources in exhaustible, the twentieth century increasingly delivered evidence of the environment as fragile, as the victim of (human) violence, and threats to its functioning, which is a very important change. Nearly a century after Marsh's *Man and Nature*, a leading cultural geographer, Carl Sauer, led a symposium that was published in 1956 as *Man's Role in Changing the Face of the Earth*, which was perhaps the first step in modern geography in this way of thinking.

By the 1970s, two key concepts rose to importance: in environmental terms population and environmental limits. In 1798 in England, Thomas Malthus had pub lished his *Essay on Population*, arguing that there was a 'natural law' wherein the human population tended to reproduce faster than its food supply, so that population growth would only be slowed or reversed by outside forces, such as bad weather causing crop failure or out breaks of disease or war. Although Malthus was much disputed, his ideas reemerged in the 1960s on a global stage in relation to population growth in developing countries. Such 'neo Malthusian' ideas are shown in Paul Ehrlich's book *The Population Bomb*, published in 1968, which sought to raise the alarm about rising population levels, and Garrett Hardin's essays, *The Tragedy of the Commons* and *Living in a Lifeboat*, published in 1968 and 1974 respectively, which criticized the overexploitation of environmental resources because they were limited by nature. The most famous example of this way of thinking was *The Limits to Growth* model and book, commissioned by a group of business leaders, The Club of Rome, in 1972. The 'limits to growth' was a computer model of all the major global trends – human population, food supply, and industrial pollution – projected to 2100. Although it was crude and we could spend plenty of time critiquing its assumptions and data, the point is that it generated debate instead of complacency about the future. The time was also appropriate: the global oil price crisis highlighted the finite nature and precarious supplies of essential environmental resources.

But there were challenges to such thinking, which was derided as 'doom mongering'. Marxist critics in human geography and other disciplines argued that there were no 'natural' limits to human growth, because growth depends upon human ingenuity (to create technology) and human justice (to distribute resources fairly). If anything, limits would be relative to human development. These two ways of thinking directly clashed in 1980, when Paul Ehrlich made a bet with the economist Julian Simon (who disputed the idea of limits) about what would happen to the prices of five key metals over the next 10 years. Ehrlich lost and had to pay Simon US$576 in 1990. Simon's ideas fit what geographers have called 'technocentrism' (or 'cornucopianism'). This arises from classical science and modernism and argues that eco nomic growth (usually measured in gross domestic product/gross national product (GDP/GNP)) is the in dicator of a healthy society and that science and tech nology 'create' resources by finding ways to turn a shiny bit of rock into something that can be used and sold. Such thinking rejects the idea of fixed (natural) limits to en vironmental resources and argues that technology and science will figure out answers to our environmental and social problems. Hence, this is sometimes criticised as a superficial or 'technological fix' to our problems, where we develop a new gadget or upgrade our old one in order to redress environmental damage, rather than change more fundamental aspects of how we eat, live, work, and travel.

Modern environmentalism as politics and culture

But not only science nurtured modern environmental thinking. The environmental movement also arose within a wider movement in the 1960s in science and the arts to express feelings that were antiestablishment, anti indus trial, and antimaterialist – often called, collectively, the 'counterculture'. Grassroots political movements sprang up on many fronts: peace, antinuclear, civil rights and Black power, students, women, and the environment. Like the Romantics in the previous century, this was another backlash against social trends, in this case against tech nology and sociopolitical oppression, rather than against urbanization and its ills.

Moreover, modern thinking had not wholly sup planted premodern thinking about the environment. In deed, researchers have been increasingly interested in how cultures other than the industrialized West see the environment, especially where they see the environment not as a separate realm from society, but as part of their everyday life through both practice and meaning. These ways of thinking about the environment are often associated with religion and spirituality. In Native American culture, symbolic rituals are part of practical actions, such as hunting animals by showing them respect, 'courting' them with dances and songs before the hunt commences, and in some African cultures, there is a strong individual identification with animals. Geographers find it particularly interesting when such cultures emphasize the importance of place, as a very particular and well known environment. Aboriginal Australians think that where people are born (as well as to whom) is very significant to their status and future prospects – 'the country' therefore gives both life and meaning and also is a physical and spiritual product of the lives of their ancestors. In many premodern cultures, spirits or gods lived throughout the environment, in animals, waterfalls, caves, and so on, and deities also took animal form. Such notions are also important in non Western traditions today, such as in Egyptian and Indian religions, causing scholars to study different religious traditions to understand and compare their environmental ethics. Indeed, some have argued that the Judeao Christian theology, so dominant in the industrialized West, has had damaging consequences for the way that Western cultures (and those in areas colonized by the west) think about and treat the environment and that other religions may offer better approaches.

The environmental movement therefore has a diverse philosophical base drawing from ecology and other sciences, but also from ethics, religion and culture. One key environmental philosophy is 'deep ecology', sometimes also referred to as 'ecocentrism' in contrast to technocentrism (see above). Deep ecology emerged from the countercultural reaction against increasingly polluting industries and from a turn in the West in the 1970s toward eastern religions and philosophies such as Taoism, Buddhism, yoga, and meditation, as well as new scientific ideas like chaos, complexity, and relativity. Deep ecology's two key principles are self realization and biocentric equality, based on the intrinsic value of all life. So, it is not just about managing the environment better, but about managing oneself better and self improvement through reexamining one's relationship with the environment. Other important examples of environmental philosophy are ecofeminism, bioregionalism, and social ecology.

But again, environmental philosophies are not just about analysis, they also inspire practical action. Indigenous ways of thinking about the environment in non Western cultures are often held up as practical examples of better environmental management, because they draw on both indigenous knowledge and a closer connection between people and their environment, whether this is spiritual or philosophical in character. Sometimes this risks idolizing premodern cultures as environmentally perfect, when they may simply have been less damaging because their technology was less extensive or their numbers were far fewer than today. However, analyzing alternative ways of thinking about the environment can also be very useful in challenging the dominant way of thinking, empowering indigenous people and arguing for change. Good examples from India are the Chipko women's movement and the arguments of Vandana Shiva for more culturally informed environmental management and, from North America, Native American tribal groups that have campaigned for better management of sacred sites. In the 1970s, Schumacher's notions of Buddhist Economics also inspired the 'appropriate technology' movement, a movement that had a lasting impact on the politics of international development.

Overall, this second way of thinking about the environment was generally and self consciously more political than the first, which focused more on practical benefits. However, both influence practice, because how we think about environmental degradation and its causes can strongly affect how we create and implement possible solutions.

The Environment as Part of Society

So far, we have looked at the two main ways of thinking about the environment today and they often clash over the best environmental policies and forms of management. For example, the dispute over logging in the Pacific Northwest of the USA in the 1980s and 1990s was often polarized as jobs versus owls – in other words, the instrumental use of the forest for economic development versus the preservation of the forest for the intrinsic value of the species within it, especially spotted owls.

But there is another way of thinking about the environment that has developed recently as a reaction against the other two, mainly within academia. This third way rejects the separation and (often negative) opposition between the environment and society implied by the first two ways of thinking. It both refuses to ignore the environment's influence on human development and also rejects the human arrogance implied by assuming that humans are all powerful. Instead, it emphasizes how the environment and society are intertwined, rather than separated or in opposition, using both policy concepts like sustainability and urban wildlife and theoretical concepts like hybrid natures. It is also more philosophical in character, compared to what we might think of as the more practical and political characters of the first two, respectively.

There is much more written on this topic in other articles, but it is worth noting here that I have arranged these three ways of thinking chronologically, to emphasize that the first underlies the development of modern society, the second was a consequence of that development, and the third responds to the gaps in both. But this oversimplifies, because one did not easily replace another as time passed; rather, they come frequently into conflict with each other, even today, over how best to understand and manage the environment in different times and places. This is a very important point to re member whenever we engage in environmental debates.

Other Themes in Environmental Thinking

Having outlined the history and development of three ways of thinking about the environment, the rest of this article will consider some crosscutting themes in human geography more recently. The last 20 years saw the most activity in environmental research in the social sciences generally, as a consequence of environmental thinking rising up the political agenda and becoming an important social phenomenon in itself.

The Contemporary Environmental Movement

However, the product of this rising attention was not a single, clear block of environmental thinking, but rather a wide and loosely defined set of groups and ideas about environmental policy and practice. A lot of work on environmentalism has been done outside geography, es pecially in political science (studying the groups) and in philosophy (studying their ideas). In geography, en vironmental groups are sometimes studied to see how much influence they have had on a particular environ mental policy. Such research usually studies what we can loosely call 'reformist' or 'light green' groups, which tend to be centrally organized and controlled and good at lobbying politicians and developing policy proposals for environmental reform (hence the name) – WWF is a good example. By contrast, more 'radical' or 'dark green' groups tend to be less centrally organized, less hier archical, and more interested in grassroots protesting, adopting artistic, confrontational, or even illegal forms of protest as well as nonviolent direct action through oc cupying threatened sites – Earth First! is a good example. These groups have been studied for what they tell us about environmental lifestyles or subcultures, rather than what they tell us about policy and politics.

Of course, this dichotomy of 'reformist' and 'radical' oversimplifies a complex political arena. Many environ mental groups do not fit solely into one side or the other, because they have multiple strategies of lobbying and influence. For example, Friends of the Earth in some re spects is reformist, through lobbying parliament and working with business, whereas in other ways it is more radical through its structure of locally autonomous groups. But conceptually, this dichotomy helps human geographers and others to distinguish between environ mental thinking that seeks mainly to change politics and policies and environmental thinking that seeks to change cultural norms, such as about recycling and car use, and promote alternative lifestyles. We could see the latter as 'cultural politics' or a 'politics of resistance' to dominant systems, because it seeks political change through cultural activities, rather than through lobbying politicians. For example, the antiroad protests in the 1990s in England established camps on sites threatened by road building and these drew in a diverse range of protestors and de veloped alternative, non consumerist lifestyles (living in homemade shelters) and expressive, political art.

This diversity of ideas and personalities within the environmental movement often results in conflicts over the best environmental pathway. For example, those more inclined toward reformism might favor improving the energy efficiency of car engines to mitigate against cli mate change, but more radical campaigners might want to ban certain forms of cars or their use in certain places and times. A reformist may see a radical as too extreme, revolutionary, and unrealistic, trying to change the world from a tipi, whereas a radical might see a reformist as only interested in small, incremental changes, in wishy washy, sticking plaster solutions that do not address the root causes of environmental problems, in 'end of pipe' cleanup of a problem, rather than its prevention. A good example recently is how the environmental movement has split in the UK over wind farms, some environmental groups seeing these as necessary forms of renewable energy and others seeing them as damaging wildlife (especially migratory birds) and scenic value.

Environmental Policy and Decision Making

Another major area for analysis by human geographers has been environmental policy on everything from re cycling to managing wild species in the environment. Although diverse, there are two foci worth highlighting briefly.

First is the dominance, over the last 15 years or so, of the concept of sustainable development (or sustain ability), especially in Europe. This was originally pro moted by the UN sponsored Brundtland Commission, especially through their report, *Our Common Future*, published in 1987. Sustainable development brings together economic development and environmental protection and thus attempts to resolve some of the tensions between the first two ways of thinking at the beginning of this article especially through ideas of ecological modernization and environmental regulation. This is clearly a great challenge and has generated a lot of research into what sustainable development means, how it might be operationalized and how it can best be communicated. Research at different levels, from the global to the local and from the highly abstract to the eminently practical, suggests that the tensions between the economy and the environment cannot easily be resolved within existing systems of decision making and economics.

Second, human geographers have been increasingly interested in how the public can be involved in environmental policy. This reflects the 'participation' agenda that developed in the 1990s, as part of *Agenda 21*, a blueprint for worldwide sustainable development agreed at the Rio Earth Summit in 1992, and in the UK in the Local Agenda 21 strategies developed by local authorities. Emphasizing participation is part of what people have called a move from environmental government, where managing the environment is largely done by the state in a top down approach, to 'environmental governance', where managing the environment involves a wide range of nongovernmental organizations (NGOs), uses nontraditional channels and policy instruments, such as public–private partnerships and emphasizes the need to involve 'ordinary' people in formulating and implementing environmental policy. The argument is that, for environmental policy to be truly democratic, it should not be dictated top down by government to the public but should reflect the concerns of everybody in a bottom up or grassroots process. Moreover, if environmental policy is to work in practice, everybody must be able to adopt it or fit into their daily lives.

But involving the public is easier said than done. Environmental policy has a tradition of being quite technocratic and exclusionary, with decisions being made by closed circles of scientists and advisers. Some policies were traditionally made on a corporatist or bipartite basis, decided between government and business, so that environmental groups and the public were usually excluded. Changing this has been difficult and requires us to improve processes of consultation over environmental decision making, particularly to encourage more interactive dialog between decision makers (including environmental scientists and others seen as 'experts') and the public. Attempts to achieve this have included 'planning for real' exercises, deliberative mapping, citizen juries, and the countrywide GM Nation? process in the UK in 2002–03.

Environmental Awareness and Concern among the Public

A central theme for many geographers researching environmental issues is how the public see the environment and how environmental concern is produced and patterned. This is because the public are not merely passive recipients of environmental facts, but actively respond to and create their own environmental meanings from a diversity of sources, including not only the mass media but also their own firsthand experiences of their environments and the ideas of their friends and families.

There are also tensions between environmental awareness (factual knowledge about environmental change) and environmental meanings (how people feel about the environment) and between both these things and environmental behavior (what people actually do to put their knowledge and feelings into action – or not). Public opinion polls have consistently reported that concern about environmental change has been (relatively) high since the 1990s. Yet, people's behavior rarely matches their concerns, in terms of recycling, reducing emissions, voting for 'green' parties, or other pro environmental behaviors. This is referred to as attitude–behavior dissonance in psychology – we say or believe one thing but do another – and it remains a trying problem for researchers to explain.

Politicians and campaigning groups often try to provide information about environmental change in the hope that this will cause people to change their behavior. For example, adverts encourage us to switch lights off, to car pool, and to turn the temperature down on our washing machines, to help reduce climate change. In sociology, this is referred to as the 'deficit model' in the public understanding of science, because it assumes that people behave badly because they lack information and if that lack (or deficit) is remedied through new and better information, people will change their behavior. But sociologists and human geographers have shown that things are never so simple and that information does not necessarily change behavior, because many other factors – perceived moral norms, convenience, and personal agency – are involved. Cigarette smoking is one simple example.

An expanding field of research is therefore into green and sustainable consumption. Unlike research into environmental policy and regulation, this focuses not on national or local government activity, nor on groups within the environmental movement, but on individual choices and lifestyles and their aggregated effects on natural systems, agricultural production, and environmental damage. Key issues here are how aware people, particularly householders, are about environmental change and their role in exacerbating or curbing it. Our purchases of everything from energy efficient boilers to fast food can be indicative of how we think about environmental issues (or not) in our daily lives. Studies suggest that people have complex and multilayered notions of environmental change and value. Simple correlations between higher levels of environmental concern and age, gender, educational level, location, lifecycle, income, and occupation have been suggested, especially that women, younger people, those with more education, and those in more socially supportive occupations (teaching, healthcare, and local government) tend to be more sympathetic to environmental messages. However, the results of empirical studies remain inconclusive or contradictory.

Corporate Environmentalism and Environmental Commodification

Business is another key actor in environmental thinking and management, traditionally associated with the first way of thinking about the environment in seeing the environment as raw material for exploitation and profit. But in the last couple of decades, many companies have expressed concern for environmental protection, in line with the second way of thinking, under pressure from the public, government, and the environmental movement. Some would argue that this move has been superficial rather than substantial, a case of 'greenwashing' for public relations, rather than important operational change.

We can trace this move back to the 1970s and the 'limits to growth' argument (above) about rising pollution because of economic growth, which cast business as the villain of environmental damage. This caused a backlash from business groups and other commentators against such 'doom mongering', but increasing environmental regulation in the 1970s and 1980s and environmental campaigning that targeted big business, such as early campaigns against nonreturnable Schweppes bottles by Friends of the Earth, added pressure to respond. But it was only in the 1980s that business really grasped the green nettle, because of more global attention and science about environmental problems like ozone depletion, climate change, and deforestation. Moreover, this period saw business increasingly repositioned as part of the solution, rather than solely part of the problem. For example, the UN sponsored Brundtland Commission (see above) on sustainable development said that industrial growth was needed, but that it should change its form to be less resource intensive and more environmentally and socially beneficial.

There was then a general change, with big corporations in particular seeking green credentials and a role as the environmental solution or savior in the manner advocated by the Brundtland Commission, rather than the villain. In particular, companies set out to implement 'good housekeeping', which meant improving their own processes to reduce pollution (and save money), and also corporate social responsibility (CSR), which meant paying attention to their environmental and social impacts as well as developing ethical policies and philanthropic activities. For some industries this was seen as a way to reestablish their social legitimacy and 'license to operate' by overturning a history of bad press about incidents like benzene in Perrier water, Nike using sweatshops, Shell's disposal of the *Brent Spar* platform, and the Exxon Valdez oil spill in Alaska.

These attempts have been widely criticized by human geographers and others, as 'greenwash', public relations exercises and simplistic technological fixes rather than true changes in the relationship between business and the environment. Marxist geographers have used political economy approaches to analyze corporate environmentalism and its commodification of nature, seeing the 'greening of business' as a response to a crisis within capitalism, rather than a response to environmental damage. This argues that business has *not* moved away from the first way of thinking introduced at the beginning of this article: that environmental resources exist to be exploited, particularly in order to accumulate wealth through notions of neoliberalism, free trade, and market led development. Further, recent interest as part of the 'ethical turn' in cultural and economic geography has considered fair trade, ethical standards, alternative food networks, and environmental certification processes as examples of changing business practices. Although still often critical of market solutions and of big corporations, such analyses are attempting to see how far environmental governance can involve business people, as well as householders and NGOs, in pursuit of more sustainable technologies and practices.

Analyzing Environmental Thinking through Social Construction

Underlying a lot of what has been written in this article is an argument about how we frame environmental issues. Before any policy or action can be taken about an environmental issue, before anyone can get concerned or campaign about it, we must first identify it and give it a name. Climate change is a good example: if the US Government under George W. Bush argues that there is no conclusive evidence that climate change is a problem, then there is no pressure on them to take action to solve it. Analyzing such processes of identification can use the theory of the social construction of reality, which posits that how we think about the environment is not a simple reflection of that environment but a social construct. We may know about the environment through firsthand information (observing the weather, walking through the local park, or visiting the seaside) and secondhand information (watching television or chatting to other people) and we often change our minds from time to time. But we change our minds about the environment not primarily because the environment changes, but because we do. Hence, there is no single environmental 'truth' to discover, because our environmental issues depend on how we investigate and consider the environment.

The idea of social construction has been around for 30 years now and seeks to understand how society makes sense of the world, particularly through endowing it with meaning and changing that meaning as society changes. It is sometimes misinterpreted to mean that there is no reality, no external environment, except in our collective imagination, but the social construction of environmental issues does not mean that such issues are imaginary. It does mean that we can only think about them in certain ways, often ways that we learn through our interactions with particular social groups in specific times and places. Hence, two people may think of the same hot day differently – one as normal summer weather and one as evidence for global warming – and this matters because of what it tells us about social understandings and also because it may affect what different people do in response, whether they put on more sunscreen or think about how they can reduce their carbon dioxide output. So, the social construction of the environment is about social meaning but it is also about the consequences for human action, and in that sense social construction can have a very real impact.

Social construction is a theory applicable to a range of social issues, not just the environment, but the environment is a particularly interesting construct because it is often framed at a distance or a scale beyond people's everyday lives, especially for global environmental issues like climate change, ozone depletion, and deforestation. In consequence, it is frequently framed through science, which may make it more difficult for people to connect these global ideas with their own, often more local, experiences. Human geographers have particularly analyzed such links across scales, not least to problematize the rather easy slogan of the environmental movement in the 1990s: think globally, act locally.

Applying social construction to environmental issues also makes us think about how these issues are framed for us by various 'claims makers', such as environmental groups, scientists, big corporations, and journalists, who identify and define an environmental issue and bring it to our attention through the media. All these actors may 'encode' the environment in particular ways to suit their medium and their own agenda about what they want their audience to think. Hence the coverage of climate change will vary greatly between the BBC *News at Ten O'Clock*, Al Gore's film *An Inconvenient Truth*, and the pages of *New Scientist*. Often, a distinctive name or metaphor is coined, like 'the greenhouse effect' or 'Frankenstein foods', to convey very complex environmental ideas through a simple and often highly emotional shorthand.

But claim makers cannot control their audience's reaction. Some people will 'decode' the environmental meanings presented to them quite differently from how the producers intended. Geography matters: familiarity with and proximity to the environments being described may influence how people respond. People living around big factories and nuclear power plants understand the environmental impacts and risks of those sites in different ways, influenced negatively by 'folk memories' of smells, accidents, and company publicity and positively by the employment provided, so their views may differ substantially from those of someone living far away. And research shows that, whatever the message about environmental change, we tend to distrust the messenger more if they come from industry or government than if they are seen as 'independent' scientists or pressure groups. Moreover, how we construct the environment can change rapidly, with some environmental issues going through a 'life cycle': they rise from relative obscurity to reach a high public profile and enthusiasm for action, but then decline from prominence to a backseat in people's minds. So, how environmental meanings are both produced and consumed is critical, because many layers of encoding and decoding accumulate over time.

More recently, some researchers have criticized social construction for failing to appreciate environmental agency and seeing society as exclusively framed by human actors. In trying to move away from the perils of environmental determinism, some have argued that geographers and other social scientists have gone too far the other way, into social construction and the perils of social determinism by denying any power to environmental processes. So, there has recently been a strong push to put 'nature' back in and not consider merely 'social' construction. In human geography, such thinking has been driven by ideas of cyborgs, actor networks, and hybrid natures from analyses in the sociology of science. Such ideas share a desire to collapse or blur dualisms like nature–culture and human–nonhuman in order to posit a world of networks and hybridity whereby entities of heterogeneous types come together to form and reform environmental systems. Rather than dividing up the world between nature and culture, between the rural and the urban, and between the wild and the tame, these ideas consider such divisions as delusions of modernity and seek instead a more inclusive ecological politics, whereby all actors – human, animal, plant, machine, and environment – are intimately interrelated and through those interrelationships continually reconstitute the world. This reflects the third way of thinking that was outlined above: one which does not seek to choose between environmental power and human power, but sees power as a product of that interrelationship.

Scale: Global and Local Environments

A key theme in all human geography is how things change across space and how places are differentiated. Environmental thinking reflects this well because many environmental issues are constructed as global issues, requiring global solutions and global collaboration. In the Space Age of the 1960s, photographs of the Earth taken from space became hugely popular, selling as posters for student rooms and being used as symbols by groups like Friends of the Earth, projecting the idea of a fragile planet floating in space, a Spaceship Earth operating within strict environmental limits (like Hardin's lifeboat mentioned above). Ecology also scaled up to consider ecosystem ideas on a planetary scale and used the metaphor of Spaceship Earth to consider how best to manage and recycle the planet's finite resources.

There is a very specific example of such planetary thinking. The Gaia hypothesis was developed by James Lovelock in the 1970s, out of his contributions to the US Jet Propulsion Laboratory's studies of the atmospheric composition of other planets in our solar system. His work made him think about why the Earth's composition was so different, with far less carbon dioxide and so much more oxygen and nitrogen that are beneficial to life. He came to argue that the Earth self regulates through its ecosystems, whereby all organisms contribute to the at mospheric balance and maintain things at a level that supports life. Although this started out as a scientific hypothesis, it has become much more than that – it has become a metaphor for planetary symbiosis and inter dependence, and Lovelock has also gone on to become an important commentator on environmental change more generally.

Moreover, this global scale of environmental thinking is reflected in policy. Even early in the conservation movement, environmental groups sought international collaboration. The 1900 Convention for the Preservation of Wild Animals, Birds and Fish in Africa (to prevent export to Europe) and the 1902 convention to protect birds useful to forestry and agriculture were signed by 12 nations and, although not strong on implementation, were the first on paper to adopt an approach now very familiar to us. Today, many international environmental policies are led by the United Nations, such as the famous Earth Summits at Rio de Janeiro in 1992 and at Johannesburg in 2002. Most high profile environmental issues are global, as are their policies, like the 1997 Kyoto Protocol that was set up to reduce global greenhouse gas emissions.

At the other end of the scale, human geographers have been busy trying to understand how people think about local environments. Driven by the humanistic approach and 'cultural turn', human geographers have explored the local and familiar, especially within environmental con flicts, to illustrate how the environment is differently imagined and valued and to relate this to gender, edu cational background, and personality. We can see that in this sense there is not one 'environment' but multiple environments.

At the same time, the recent 'ethical turn' has re asserted the importance of linking ideas and action and in pursuing environmental justice. Morally, the connection that each individual has with the (especially local) environment is argued to be useful for imbuing en vironmental citizenship and encouraging sustainable lifestyles. Environmental awareness and education are seen as valuable not only in democratically involving publics in environmental decision making, but also in implementing environmental policies through citizen action. In this way, the environment is internalized in the individual, even as global constructions of environmental change dominate the media headlines.

Related to the arguments above about public partici pation, researchers have also increasingly sought to value lay knowledge about the environment, knowledge that is held by people not commonly regarded as 'experts' but who may be intimately involved in environmental change and at risk from its impacts. It is therefore important to analyze how the knowledge of people like farmers, users of a local park, and amateur naturalists can be legitimated and brought into environmental policymaking more ef fectively. This has been particularly important in devel opment studies, where lay knowledges about the environment in the Global South are seen as devalued by colonialist perspectives, but having great potential for scientific learning and even for future exploitation of species for agriculture and medicine.

Creating Environments, Restoration, and Terraforming

Most of this aricle so far is about environmental damage and the loss of environmental resources. So let us end with the opposite: the creation of environmental re sources. In order to redress environmental degradation, some are seeking to create lost environments and restore damaged ones. This is an exciting but controversial practice, for many reasons.

Often referred to as 'ecological restoration', the pro cess of creating particular habitats and ecosystems, in whole or in part, has been increasingly researched by natural scientists, including physical geographers, since the 1940s, beginning with restoring prairies in North America. Such activities are sometimes seen as having great potential for ecological learning, through using the environment as a natural laboratory for testing theories about species recolonization and prey–predator rela tionships, reflecting the first way of environmental thinking introduced in this article. But they are also often used to reflect the second way of thinking in seeking to protect valued habitats by recreating them in one place after they have been lost in another, because of human impacts like road construction or sea level rise. They can also reflect the third way of thinking by claiming en vironmental restoration as a holistic activity, involving science, morals, culture, and lay people.

Examples of environmental restoration are recreating wetlands, remeandering rivers, and replanting forests.

Reintroducing species is also a highly controversial example. In the early twentieth century, the USA sought to eradicate the wolf in order to protect livestock for agriculture and deer populations for sport. The authorities classified wolves as 'vermin' and put a bounty on each wolf's head delivered to them, with the result that the wolf was soon largely eradicated as a species within the coterminous USA. After the policy ceased in 1936, support grew for the intrinsic rights of such animals to exist and fulfill their natural role as predators in controlling herbivore numbers. This prompted a movement to restore wolves to Yellowstone National Park in the 1990s, where they are now an attraction for wolf watchers and scientists. They are also a target for critics: several were shot after being reintroduced, probably by local people who saw them as a threat to livestock and game species.

Part of the controversy is what such restorations mean. Can they truly correct past environmental damage or are we simply deluding ourselves about the possibility of erasing our mistakes? Environmental philosophers have spent decades analyzing the (im)possibility of restoration and the morality of artificial environments, comparing restored environments to paintings faked in the style of a famous artist and passed off as the genuine article. More prosaically, geographers have considered the impact on policy of restoration, especially where it has been used to legitimate environmental damage caused by a new development such as a motorway or quarry. Proponents have argued that the habitats and species damaged would be moved or recreated elsewhere as compensation for the damage, and that thus there would be no net loss and possibly even some planning gain through the new development. Using ideas of restoration in this way is not merely about the philosophical possibility of faking the environment, but about how we think about the environment and make policy to manage it. In the most extreme form, restoration has been referred to as 'terraforming', echoing *Star Trek II*: *The Wrath of Khan* in which a new planetary environment was recreated through a nuclear style bomb called 'genesis'. Such ideas are thus an extreme expression of human power over the environment and its resources.

Conclusion

A lot of material and a lot of time have been covered in this article, from early ways of thinking about the environment as a powerful force and a resource for human development, through fears of environmental damage to human powers of creating new environments. Human geographers have therefore found many aspects of the environment to analyze over the decades and this will only continue, as environmental science and policy change and as we as a society continually reshape and reflect upon our ways of thinking about the environment.

See also: Climate Change; Conservation and Ecology; Culture/Natures; Ecology; Environment, Historical Geography of; Environmental Justice; Environmental Policy; Environmentalism; National Parks; Nature, Social; Nature-Culture; Radical Environmentalism; Rio Summit; Social Studies of Scientific Knowledge; Sustainability; Sustainable Development; Wetlands and Reclamation; Wilderness.

Further Reading

Adams, W. M. (2003). *Future Nature*. London: Earthscan.
Benton, L. M. and Short, J. R. (1999). *Environmental Discourse and Practice*. Oxford: Blackwell.
Burgess, J. (1990). The production and consumption of environmental meanings in the mass media: A research agenda for the 1990s. *Transactions of the Institute of British Geographers* 15, 139 161.
Callicott, J. B. and Ames, R. T. (eds.) (1989). *Nature in Asian Traditions of Thought*. Albany: State University of New York Press.
Cosgrove, D. (1994). Contested global visions: One world, whole earth and the Apollo space photographs. *Annals of the Association of American Geographers* 84(2), 270 294.
Dobson, A. (1990). *Green Political Thought*. London: Unwin Hyman.
Evans, D. (1997). *A History of Nature Conservation in Britain* (2nd edn.). London: Routledge.
Macnaghten, P. and Urry, J. (1998). *Contested Natures*. London: Sage.
McCormick, J. (1989). *The Global Environmental Movement*. Chichester: Wiley.
O'Riordan, T. (1981). *Environmentalism* (2nd edn.). London: Pion.
Palmer, J. (ed.) (2001). *Fifty Key Thinkers on the Environment*. London: Routledge.
Pepper, D. (1996). *Modern Environmentalism*. London: Routledge.
Shiva, V. (1989). *Staying Alive: Women, Ecology and Development*. London: Zed Books.
Shiva, V. (1997). *Biopiracy: The Plunder of Nature and Knowledge*. Boston: South End Press.
Worster, D. (1989). *Nature's Economy: A History of Ecological Ideas*. Cambridge: Cambridge University Press.

Relevant Websites

http://www.sd commission.org.uk/
UK's Sustainable Development Commission.
http://www.un.org/esa/sustdev/
UN Division for Sustainable Development.
http://unfccc.int/kyoto protocol/items/2830.php
UN's Kyoto Protocol website.
http://www.nps.gov/
US National Parks Service.
http://www.wwf.org.uk/core/index.asp
WWF global network.

Environment, Historical Geography of

S. W. Trimble, University of California, Los Angeles, CA, USA
M. P. Conzen, University of Chicago, Chicago, IL, USA
E. K. Z. Zygmunt, University of Silesia, Sosnowiec, Poland

Glossary

Environmental History It is the study of environment over time, usually with more emphasis on political and economic aspects rather than landscape.
Historical Geography of the Environment It is the study of past landscapes with the purpose of identifying and understanding environmental changes and problems, especially those caused by human agency.
Human Impacts They are the changes in the environment caused by humans.
Intended Impacts They are those so designed, such as building a water reservoir by damming a stream.
Unintended Impacts They are those not expected along with human changes to landscapes. An example might be water loss from evaporation in the case of the reservoir.

Historical geography of the environment is the study of past landscapes with the purpose of identifying and understanding environmental changes and problems, especially those induced by human actions. In short, this approach shows how and why humans have transformed the natural landscape and what the consequences have been (see, for example, **Figure 1**). It is a tradition going back to ancient times and one that was warmly embraced and developed by nineteenth century geographers, such as the German natural scientist Alexander von Humboldt, the English geologist Charles Lyell, the French geographer Élisée Réclus, and the American natural scientist George Perkins Marsh. Such work has clearly demonstrated that severe environmental disruption was created at times by some of the earliest societies and has continued through the ages. Present environmental problems may be different but, relatively speaking, not necessarily worse than those created by humans in the past. The destruction of cropland or forest, for example, may have had similarly severe impacts on populations centuries ago as global warming is said to have on populations today.

During the early modern period of geography (*c.* 1900–70), historical geography of the environment was a mainstay. Perhaps the most powerful voice was that of Carl Sauer who wrote widely and deeply on the subject. Sauer had a profound understanding of the physical as well as the cultural landscape and was especially concerned about the long term effects of soil erosion among other environmental problems. Another geographer prominent in the field was H. C. Darby who, with his students, explained many present day physical features of the British landscape as having been created during historical time by human modification. Yet another prominent geographer was Clarence Glacken whose philosophical *Traces on the Rhodian Shore* (1967) remains a classic statement of changing cultural perceptions of the Earth environment and people's resulting choices about land use.

A benchmark of sorts was achieved in 1956 with the publication of *Man's Role in Changing the Face of the Earth*. Edited by W. L. Thomas, it featured often inspired essays by prominent geographers like Carl Sauer and Arthur Strahler as well as many other natural scientists and has served as an inspiration for work since then. Among the human geographers who have carried on this tradition are Michael Williams, Hugh Prince, Robin Butlin, David Lowenthal, William Denevan, Leslie Hewes, and John Hudson.

But starting in the 1960s, human geographers started showing less interest in this line of inquiry and more physically oriented geographers picked up the theme. Especially important were works of Karl Butzer (*Early Hydraulic Civilization in Egypt*), Claudio Vita Finzi (*The Mediterranean Valleys*), and I. G. Simmons (*Changing the Face of the Earth: Culture, Environment, and History*).

The neglect of environment by social science trained historical geographers was influenced by pressures for methodological specialization within the discipline. These pressures relaxed the expectation that geographers would continue to be similarly well grounded in both physical and human geography. This estrangement paralleled a growing interest among human geographers in social theory, and the narrowed belief that, in terms of human activity, the social sciences had more to teach them than the physical sciences, and that the physical side of geography could be regarded as some kind of black box that could be left unexamined by them. This neglect is reflected in the magisterial four volume series by D. W. Meinig, *The Shaping of America*, which, while being the most ambitious and accomplished single synthetic work ever written by a historical geographer, nevertheless pays little direct attention to the specifically physical aspects of landscape. This same trend can be seen in the content of geography journals during the last decades of the twentieth century. Further evidence is

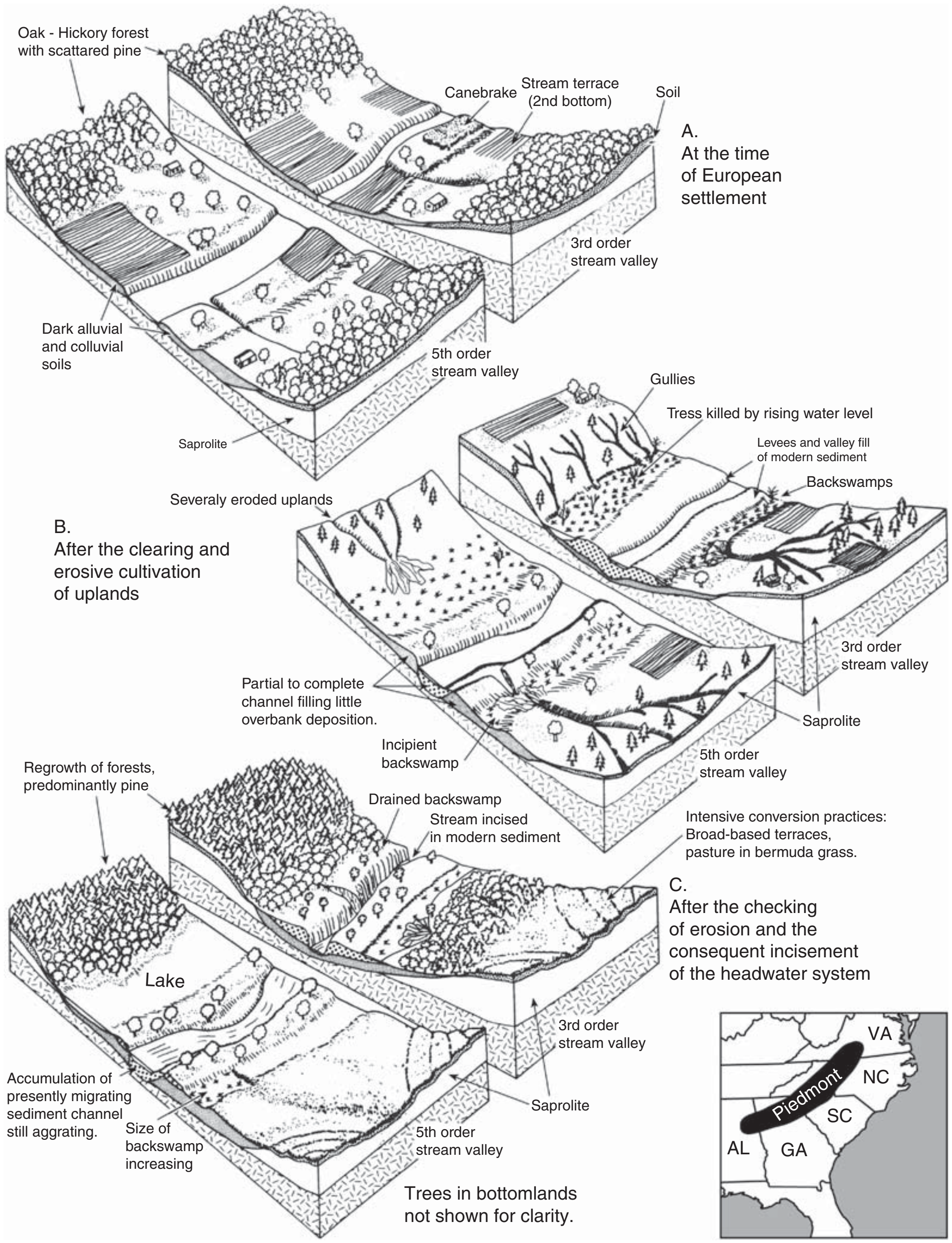

Figure 1 Generalized evolution of the Southern Piedmont landscape 1700 1970 showing the changes of land use, landscape, and stream morphology. From Trimble, S. W. (1974, 2008). *Man-Induced Soil Erosion on the Southern Piedmont, 1700 1970.* Ankeny, IA: Soil Conservation Society of America.

supplied by the massive German compendium, *Geography, Physical Geography, and Human Geography* just published. From hundreds of pages of text, historical geography merits only 13 pages and environment receives a mere passing notice.

Nevertheless, there are human geographers still pur suing the genre, and indeed there has been something of a renaissance in recent times. In addition to those already mentioned, these include, in the United States, for ex ample, Lary Dilsaver's writing on national parks, Craig Colten's work on toxic wastes and New Orleans, Brad Baltensperger's studies of the US Great Plains, and John Hudson's work on settlement and environmental trans formation in the American Middle West. More generally indicative of this renewed interest, Alan Baker gives the genre a solid chapter on environment in his recent and definitive international survey, *Geography and History: Bridging the Divide*.

The vacuum created by the divergence of and non communication between human and physical geog raphers did not escape notice in other quarters, and historians have rushed into the void. Sometimes rela tively untutored in the physical aspects of landscape, they have nonetheless plunged into environmental work. Historians, of course, are primarily interested in the way people have perceived and made decisions about the physical environment around them and how this has af fected social relations, and not surprisingly the emphasis is more on the changing state of people's awareness of environment than on the changing state of the environ ment itself, particularly in its spatial interconnectedness. Some of this historical work has an overtly political tone and embraces advocacy, and sometimes falls short of the expected environmental understanding. Nevertheless, this potent movement contains many gifted practitioners, among whom the most notable are William Cronon, Carolyn Merchant, and John Stilgoe. Accompanying and facilitating the growth of environmental history has been the rapidly growing professional group, the American Society for Environmental History. Its journal, *Environ mental History*, now in volume 13, succeeds *Environmental History Review* which began in 1976.

Recent decades have seen the genre of historical en vironmental geography become to a large extent the domain of physical geographers and other natural sci entists, primarily in journal articles. Here, more emphasis is placed on the physical aspects, but the historical landscape dimension remains strong. Important con tributors have been Andrew Goudie, Heather Viles, H. Lamb, Andrew Warren, J. Grove, Angela Gurnell, Tim Burt, Ron Cooke, Janet Hooke, James Knox, and W. L. Graf. Graf's work has been especially influential because he has written several books and served on committees of the National Research Council of the US National Academy of Sciences.

The genre, therefore, has continued to evolve. Andrews's Goudie's historically sensitive *The Human Impact on the Natural Environment* was first published in 1981. Now in its sixth edition, it has been read by thousands of college students, faculty, and interested lay people. In 1990, a new compendium, *The Earth as Transformed by Human Action: Global and Regional Changes in the Biosphere over the Past 300 Years*, was proffered with much fanfare as a successor to the 1956 classic, *Man's Role in Changing the Face of the Earth*, but has not achieved the same visibility. The last decade saw the emergence of the impressive work, *From Coastal Wilderness to Fruited Plain: North America from 1500 to the Present,* by Gordon Whitney. While the title and subtitle are overblown, as it does not cover all of North America nor even the United States, it does set a very high standard in its coverage of the northeastern region. It serves as an outstanding model for future work on the remainder of the continent, and, in deed, for other parts of the world.

The increasing volume and methodological diversity of recent writing in the historical geography of the en vironment reveals a broad grouping along three lines, distinguishable more by precept than topical focus. In the first instance, there are physical studies of environmental change in which humans have played a significant role, including, especially, studies of changing ecosystems. For the most part, these contributions adhere to the direct, positivist assumptions of the exact sciences and have come to a large degree from geomorphologists and bio geographers. A striking exception, however, is Michael Williams' ambitious synthesis, *Deforesting the Earth: from Prehistory to Global Crisis*. This group has enlarged dra matically in the decade or so surrounding the turn of the millennium and now represents perhaps the strongest sector of the subfield.

A second grouping concerns studies that examine the role of values, attitudes, and other perceptual influences on human behavior toward the physical environment. Some of the work explores unfiltered human engagement with environmental circumstances, while other studies focus more on public discourses in relation to selected perceptions of the environment, the full range of which is well illustrated in Murphy and Johnson's collection of critical essays, *Cultural Encounters with the Environment*. Overwhelmingly, both types deal with rural and so called wild contexts, but there is a growing interest in the limits of human technological subjugation of nature in urban settings, for which Blake Gumprecht's fine study of the Los Angeles River is representative. An insightful ex ample of the assertion of aggressive values more esthetic than utilitarian is Stephen Daniel's engaging study of Humphry Repton and the landscape architecure of Georgian England.

Lastly, a third grouping concerns the periodization and shifting directions of environmental politics, both as a reflection of social conflict over resource use and its environmental costs and as an arena for policy develop ment. Beloved of historians, this dimension has received growing attention from historical geographers, often with a more explicit grasp of the science behind environ mental understanding. Physical geographers contributing to this thread have often used their models of ecosys temic change to prognosticate future problems and needs, as has, for example, Andrew Marcus in the case of Chesapeake Bay on the American Atlantic seaboard. Social science historical geographers have examined political events and movements of the past to highlight critical mistakes and missed opportunities in managing the environment. Representative of this type of study are Colten's recent examination of New Orleans, Wescoat's study of water management in the Indus Valley, and, on a grander scale, Marcus Hall's intercontinental treatment of environmental restoration.

The importance of studying the earth environment from the perspective of historical geography has been validated by the quality and interdisciplinary acceptance reflected in the modern literature discussed. Most significant has been the trend to reunite the aims and methods of physical and human geography toward better explanation of the complex interactions of environmental and social systems, in which scholars working in the historical geography of the environment have played a beneficial part.

See also: Environment; Historical Geography; Nature; Nature, Historical Geographies of.

Further Reading

Baker, A. R. H. (2003). *Geography and History: Bridging the Divide*. Cambridge: Cambridge University Press.

Butlin, R. and Roberts, N. (1995). *Ecological Relations in Historical Times: Human Impact and Adaptation*. Oxford: Blackwell.

Butzer, K. W. (1972). *Early Hydraulic Civilization in Egypt*. Chicago, IL: University of Chicago Press.

Colten, C. E. (2005). *An Unnatural Metropolis: Wrestling New Orleans from Nature*. Baton Rouge, LA: Louisiana State University Press.

Colten, C. E. and Skinner, P. N. (1996). *The Road to Love Canal: Managing Industrial Waste before EPA*. Austin, TX: University of Texas Press.

Cooke, R. U. and Reeves, R. W. (1976). *Arroyos and Environmental Change in the American Southwest*. Oxford: Clarendon.

Cronon, W. (1983). *Changes in the Land: Indians, Colonists, and the Ecology of New England*. New York: Hill and Wang.

Daniels, S. (1999). *Humphry Repton: Landscape Gardening and the Geography of Georgian England*. New Haven, CT: Yale University Press.

Dilsaver, L. M. and Colten, C. E. (1992). *The American Environment: Interpretations of Past Geographies*. Lanham, MD: Rowman & Littlefield Publishers.

Dix, A. and Schenk, W. (2006). Historische Geographie. In Gebhardt, H., Glaser, R., Radthke, U. & Reuber, P. (eds.) *Geographie, physische Geographie und Humangeographie*, pp 817 829. Munich: Elsevier.

Glacken, C. (1967). *Traces on the Rhodian Shore: Nature and Culture in Western Thought from Ancient Times to the End of the Eighteenth Century*. Berkeley, CA: University of California Press.

Goudie, A. (2005). *The Human Impact on the Natural Environment: Past, Present and Future,* (6th edn.). Oxford: Blackwell Publishing.

Graf, W. L. (1994). *Plutonium and the Rio Grande*. New York: Oxford University Press.

Gumprecht, B. (1999). *The Los Angeles River: Its Life, Death, and Possible Rebirth*. Baltimore, MD: Johns Hopkins University Press.

Hall, M. (2005). *Earth Repair: A Transatlantic History of Environmental Restoration*. Charlottesville, VA: University of Virginia Press.

Lowenthal, D. (2000). *George Perkins Marsh: Prophet of Conservation*. Seattle, WA: University of Washington Press.

Marcus, W. A. (1994). The Lower Potomac Estuary from tidewater to the Chesapeake Bay: Past trends and future needs. *Potomac Valley Chronicle* 1, 36 47.

Meinig, D. W. (1980 2004). *The Shaping of America*, vols. 1 4. New Haven, CT: Yale University Press.

Murphy, A. B. and Johnson, D. L. (eds.) (2000). *Cultural Encounters with the Environment*. Lanham, MD: Rowman and Littlefield.

Prince, H. (1997). *Wetlands of the American Midwest: A Historical Geography of Changing Attitudes*. Chicago, IL: University of Chicago Press.

Sauer, C. O. (1941). Foreword to historical geography. *Annals of the Association of American Geographers* 31, 1 24.

Simmons, I. G. (1996). *Changing the Face of The Earth: Culture, Environment and History* (2nd edn.). Oxford: Blackwell.

Stilgoe, J. R. (1982). *Common Landscapes of America 1580 1845*. New Haven, CT: Yale University Press.

Thomas, W. L. (ed.) (1956). *Man's Role in Changing the Face of the Earth*. Chicago, IL: University of Chicago Press.

Trimble, S. W. (1974, 2008). *Man Induced Soil Erosion on the Southern Piedmont, 1700 1970*. Ankeny, IA: Soil and Water Conservation Society of America.

Turner, B. L. (1990). *The Earth as Transformed by Human Action: Global and Regional Changes in the Biosphere over the Past 300 Years*. Cambridge: Cambridge University Press.

Vita Finzi, C. (1969). *The Mediterranean Valleys*. Cambridge: Cambridge University Press.

Wescoat, J. Jr., Halvorson, S. and Mustafa, D. (2000). Water management in the Indus Basin of Pakistan: A half century perspective. *International Journal of Water Resources Development* 16(3), 391 406.

Whitney, G. C. (1996). *From Coastal Wilderness to Fruited Plain: North America from 1500 to the Present*. Cambridge: Cambridge University Press.

Williams, M. (1989). *Americans and Their Forests: A Historical Geography*. New York: Cambridge University Press.

Williams, M. (2003). *Deforesting the Earth: From Prehistory to Global Crisis*. Chicago, IL: University of Chicago Press.

Environmental Hazards

G. A. Tobin, University of South Florida, Tampa, FL, USA
B. E. Montz, Binghamton University, Binghamton, NY, USA

Glossary

Disasters The result of (usually) extreme geophysical events interacting with the human-use system such that there is a negative impact on society – the actual event itself.

Environmental Hazards The potential interaction between forces of the physical environment and the human-use system such that there is a negative impact on society – the potential for disaster.

Vulnerability A human-induced condition governed by situational factors that limit social, economic, and political power such that individuals or communities are less able to cope with disaster.

The global toll from natural hazards in 2005 amounted to approximately 250 000 deaths, affected at least 150 mil lion people, and caused an estimated US$150 billion in damages. This was not an unusual year. Annual records show an increase in the incidence of disastrous events and significant increases in economic losses to these events. These data reflect an underlying global hazardousness that presents many ongoing challenges. Addressing these challenges requires a multidisciplinary approach, one that encompasses all components of environmental haz ards, from an understanding of geophysical processes, such as the flood, earthquake, volcanic eruption, or storm, to an examination of those human forces that make up society, the complex web of culture, economics, and politics, and technological resources and development. The ultimate goal of these studies has been to understand fully those processes that create, foster, and exacerbate environmental hazards with a view to informing effective mitigation planning. Geography, then, is well situated to take a leading role in such investigations.

Hazard Definition

The literature on environmental hazards is confusing in part because of differential use of the terminology re ferring to hazards and disasters. In this context, en vironmental hazard refers to the interaction between the human use system and extreme geophysical events; in other words, hazard represents the likelihood of a dis astrous event occurring. By this definition, a hazard constitutes the threat to society; it is the potential for disaster to occur. Once the extreme geo physical event transpires, then this may (or may not) constitute a di saster (**Figure 1**). The disaster designation only applies when there is a 'significant' negative impact to society. However, even this terminology is vague since the scale of any impact is indeterminate, there being no generally accepted threshold in terms of deaths, damages, or peo ple affected, when this level has been reached. Does a disaster constitute 25, 50, or 100 dead, does it mean over $1 million in losses, or is it actually something more amorphous?

From an environmental hazard perspective, humans must be involved in some way; a volcanic eruption in the Aleutian Islands or a major earthquake in Siberia where few people live would not normally be classified as a disaster under such a definition. Furthermore, the en vironmental and human use spheres do not merely overlap, but are integrated phenomena that together create the hazardous conditions and precipitate or ameliorate disaster impacts. An illustration of such haz ardousness is found in communities situated along the coast of the Gulf of Mexico and in the Caribbean Islands. These communities are prone to hurricanes. When a hurricane makes landfall in this area, a disaster is almost certain to occur given the number of people exposed. For example, New Orleans, Louisiana is perhaps the epitome of such a setting, with much of the city located below sea level, thus requiring protection in the form of levees and flood walls; Hurricane Katrina in 2005 demonstrated clearly how such hazardousness of place can evolve into disaster. More than 1000 people died and many thou sands were made homeless because of failure to

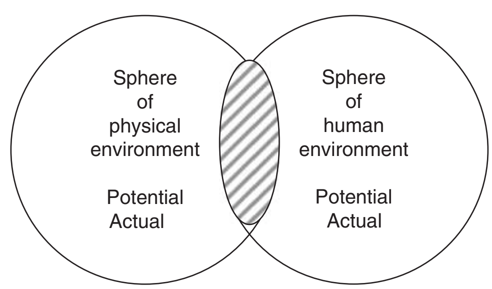

Figure 1 The environmental hazard framework. The shaded area represents the hazard where forces of the physical environment conflict with the interests of the human-use system. When an extreme geophysical event actually occurs causing negative impacts to the human-use system then this may constitute a disaster. In reality, of course, both spheres are fully interlocked, interacting in all human endeavors.

comprehend fully the interaction of physical processes and human forces.

Temporal and Spatial Trends

Over the last 100 years there has been an increase in the prevalence of disasters and a corresponding increase in economic losses. These temporal patterns can be explained in part through better global monitoring and reporting of the occurrence of extreme events. However, the trend is also a product of the definition of disaster. Since disaster is often defined by the number of people killed or affected, or by a simple measure of economic loss, global processes have led to increases in disaster incidence. The growth in global population alone means that more people are exposed and hence live in hazardous locations. Furthermore, with more people at risk, so is more property at risk, thus increasing economic loss potential.

Deaths from environmental disasters are difficult to determine accurately although there is evidence that deaths have declined over time, at least in wealthier nations. Global spatial patterns of disasters show that most disaster related deaths occur in poorer nations, whereas 75% of economic loss accrues in the wealthy nations. This is hardly surprising. Wealthy nations have more resources than poorer nations with which to mitigate disaster impacts and protect their populations, but at the same time more property and goods are exposed to potential loss. However, simple statistics do not tell the whole story. Losses must be examined proportional to overall resources. For instance, to a Bangladeshi farmer, the loss of an animal in a typhoon, while probably not worth much in the global context, could be devastating if it means that farmer can no longer provide for his or her family. In contrast, the destruction of a car by a tornado to a wealthy European or American may constitute little more than a minor irritant. Such arguments can be extended to a global scale looking at the relative impacts among nations; it is important to stress that a million dollar loss is not the same throughout the world.

Modeling Environmental Hazards

Much hazards research has sought to provide a model or framework through which disaster processes might be examined. Indeed, understanding the web of geophysical and human factors that together exacerbate or attenuate disaster has been an ongoing challenge requiring an interdisciplinary approach. The framework, shown in **Figure 2**, outlines those themes that are of concern. It is the combination of physical characteristics and political factors that serves to define risk. It is from the physical characteristics that we can determine probability of occurrence while the political and economic factors determine severity of harm or outcome. Social characteristics also come into play, but only as they are related to the other factors. By contrast, vulnerability is determined by all the elements in various combinations. However, it is through this lens of vulnerability, as shown in **Figure 2**, that environmental hazards can be viewed most productively. Although the figure may suggest that all elements are of equal concern in all situations, this is not the case; indeed, not all variables may be relevant in any given situation, and some are certainly more significant than others. The conditions are dynamic. For instance, political and economic factors, over which an individual has little control, dominate in some hazardous situations or in post event activities and priorities; in other instances, it is the relationship between the physical environment and social characteristics that defines what happens, either pre or post event. In addition, the context in which an event occurs makes a difference for all dimensions, as illustrated by the examples in the previous section.

Vulnerability

As suggested above, a key concept in understanding environmental hazards is vulnerability and its role in exacerbating or ameliorating the effects of extreme events. It is the interaction of physical and human environments that determines exposure and risk, but it is the combination of risk and vulnerability that reflects the degree to which societies and individuals are threatened by, or alternatively protected from, the potentially disastrous impacts of these events. Vulnerability has been described as a human induced situation that results from the availability and distribution of resources throughout a society. This vulnerability may come from simple exposure to the geophysical event or, more likely, is linked to prevailing social, economic, and political forces. It is abundantly clear that hazards do not affect everyone equally; some people and places are considerably more vulnerable than others. For instance, research has repeatedly demonstrated that some groups invariably suffer most in disasters, with higher levels of vulnerability correlated with higher levels of poverty, with the politically disenfranchised, and with those excluded from the mainstream of society for whatever reasons. The outcomes of Hurricane Katrina in 2005, the 2004 Asian tsunami, and the 1995 Kobe earthquake would have been very different if significant mitigation measures had been implemented or those in place had been effective, if social structures of communities had been different, or if individuals had access to additional resources. In this context, therefore, vulnerability is seen as a human condition, not simply as exposure to extreme geophysical processes. Yet, the event is a precipitating factor.

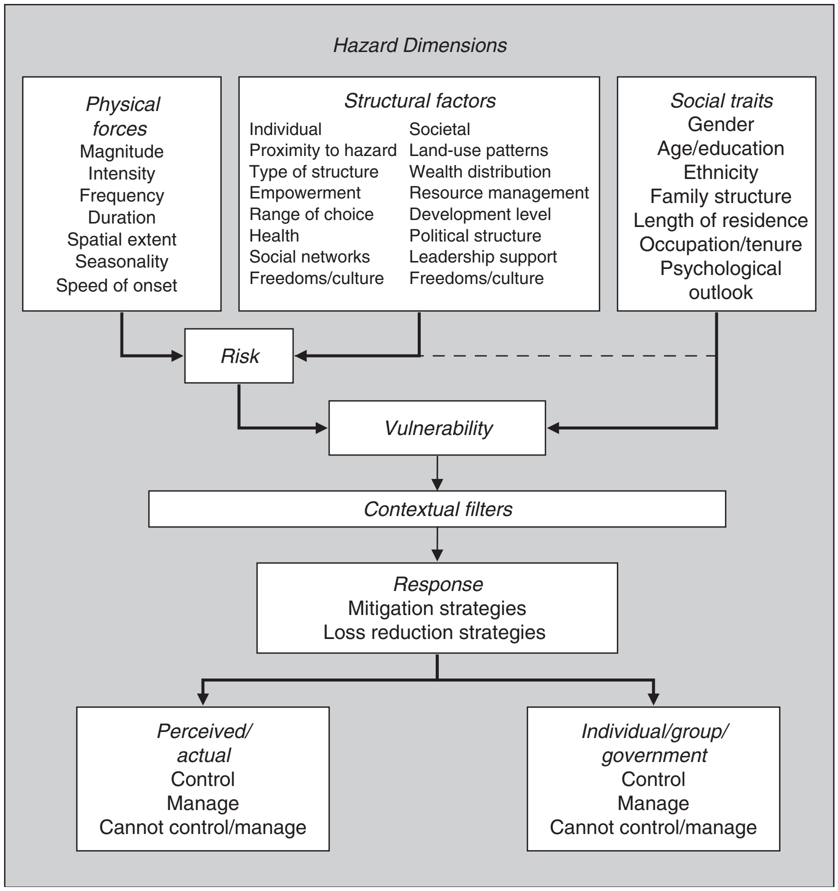

Figure 2 Natural hazards and human response in context. Reprinted with permission from Tobin, G. A. and Montz, B. E. (1997). *Natural Hazards: Explanation and Integration*. New York: Guilford Press.

Physical Dimensions

Hazards research has gleaned much from examining the physical forces that trigger disaster impacts and, in so doing, has leaned heavily on the natural sciences (**Table 1**). There are many common features shared by natural hazards, even those as apparently divergent as floods and droughts or tornadoes and earthquakes. For example, it is possible to classify natural hazards according to different geophysical measures, each of which can have a direct impact on human behavior. These include estimates of probabilities of occurrence for hurricanes and floods, calculations of differential magnitudes of geophysical forces for such events as earthquakes and volcanoes, assessments of intensity of damage from tornadoes, and estimates of potential durations of droughts and flash floods.

The greater the magnitude and/or duration of an event, the greater the impacts are likely to be. In some cases, such as drought, action can be taken even while the event is ongoing, while the same cannot be said for quick onset events like earthquakes. For some events, speed of onset makes it possible for warnings with sufficient lead times to allow for evacuation. In other events, such as earthquakes, warning is impossible, and in others, such as tornadoes and flash floods, it may not be possible to provide location specific warnings or sufficient time for evacuation. On the other hand, intensity is related to local context. The same magnitude earthquake will cause different levels of damage depending upon geological characteristics, building construction type, and organizational response capabilities.

Table 1 Disasters classified by geophysical events

Meteorological	Hurricanes	*Geological*	Liquefaction
Tropical storms/ cyclones	Tornadoes	Earthquakes	Lava flows
	Hailstorms	Volcanoes	Ash falls
Thunderstorms	Frost	Lahars	Landslides
Lightning	Blizzards	Tsunami	Mudflows
Windstorms	Heat waves	Subsidence	
Ice storms		Sinkholes	
Cold waves	Fog	*Extraterrestrial*	
Avalanches	Drought	Meteorites	
Hydrological			
Flooding			
Wildfire			

Although it is not possible to develop a scientifically predictive model of human behavior and response based solely on physical or environmental criteria because of the many other variables that influence decision making, the significance of physical criteria should not be overlooked. Generally, appropriate responses are contingent upon a full appreciation of physical processes. For example, hazard warning systems may be ineffective if the speed of onset is more rapid than anticipated, though in other circumstances they have saved many lives and much property. For example, hurricane warnings throughout the Caribbean have substantially reduced deaths in recent years. Further, the greater the frequency and the more intense the hazard experience, the louder the calls for action; it is these calls that often capture the attention of the political machinery and ultimately stimulate remedial action. Having said this, many mitigation measures for hazards have been implemented based solely on physical criteria. Building codes are often associated with projected earthquake or tornado magnitudes, and sea walls are designed to protect against given heights of tidal surges. Certainly, design standards must take into account the physical forces associated with natural hazards at a place, but determining the appropriate design level for safety is part of the human realm, and managers ultimately must make decisions about the value of life and property.

Human Dimensions: From Society to the Individual

Whether during pre or post disaster periods or as long term responses, hazard zone activities are constrained by the social, economic, and political realities of a location. Perhaps the most basic example is the money needed to provide relief and undertake mitigation, which may not be available or forthcoming in poorer communities. Yet, it is not just the absolute availability of money for development, planning, or hazard mitigation that is a problem; more often, it is the unequal distribution of resources that affects vulnerability.

Research indicates that inadequate attention is often given to environmental hazards where the expense to protect against disaster cannot be justified, or where the risk of an area is not recognized or acknowledged. Indeed, the hazard represents only one of a series of interlocking issues facing society; all demands must be assessed according to need. Overpopulation, land use patterns, availability of materials and technology, and the need for industrial development are all issues that are important and that require planning and the commitment of financial, human, and natural resources. Embedded at a deeper level, these social, economic, and political factors interact to determine how society functions and eventually define levels of vulnerability and marginalization for various groups. These factors create constraints that affect every level of society and they become clearer after an event has occurred and the distribution of impacts can be analyzed.

Sometimes termed socioeconomic traps, these constraints perpetuate risk and foster vulnerability. It has been shown that not all individuals are equally vulnerable; the risk to all inhabitants of hazard prone areas is high (as indicated by the frequency of disasters), but some groups and individuals are more vulnerable than others. At the global scale, floods, droughts, earthquakes, and other hazards recur with great frequency, bringing repeated incidents of death and destruction. Take, for instance, health impacts of environmental hazards where the same geophysical event can generate multiple outcomes. Volcanic ash can be particularly debilitating for the very young whose lungs may not be fully developed and for the elderly whose lungs are under stress, while other adults may suffer fewer problems. Hence, different characteristics, including demographic and socioeconomic, lead to differential vulnerabilities. The same is true of other situations related to the human condition and physical well being.

The logical solution would seem to be to ‘get out of harm’s way,’ but unfortunately, this option is not open to all individuals and, indeed, often is inappropriate from a societal perspective. For instance, rural land tenure issues and the decreasing ability of the rural poor to provide for themselves and their families, combined with increased investment in urban areas as well as increased opportunities for employment, can lead to the development of urban squatter settlements which are frequently located in the most hazardous sites. These changes, often generated through globalization forces external to the community concerned, exacerbate vulnerability. Therefore, the locus of social, economic, and political power is such that many individuals and communities cannot effect change to better their own situations or to promote structural change. This has important ramifications for environmental hazards where the traditional response to disasters has been to return communities, along with prevailing socioeconomic structures, to those conditions existing prior to the event, thus continuing a cycle of disaster–damage–repair–disaster and doing little or nothing to reduce vulnerability. While exposure to some events may be reduced by mitigation measures, differences in vulnerability remain and the same individuals suffer the consequences. Blaming such individuals for their predicament is not constructive. Instead, we should examine the underlying reasons, and the structural constraints of people inhabiting such hazard prone areas to understand fully why they continue to inhabit high risk areas.

Arguments that people should not continue to live in hazardous areas are moot when they have little or no choice. Instead, planning for hazards and even disaster relief programs must include sufficient resources for all victims so that adequate new locations can be found, or so that rebuilding and new development are undertaken with the risk in mind. Too often, planning for hazards is nonexistent and disaster relief is concentrated on rebuilding quickly and in place. Consequently, many individuals are usually unable to initiate changes in their living conditions to truly mitigate the hazard. Control rests with those having political and economic power who too frequently fix the symptoms of the problem (such as a flood or an earthquake) by repairing structures and returning the community to the pre event *status quo*, but who do nothing to remedy the root causes. Thus, the hazard is perpetuated.

There may also be perceived economic advantages and compelling incentives to remaining in the hazard prone area. On one hand, property values may be lower than in 'safe' areas because the risk has been capitalized into the value of the land. The consequences of this may mean that hazard prone areas develop different socioeconomic traits compared to other areas; for example, many floodplains are characterized by low income households, the elderly, young families, people on assistance, and mobile homes. On the other hand, many hazard prone areas, such as coasts, offer amenities that attract those able to afford such locations. Communities are loath to discourage such development because of perceived financial benefits.

Decision Making

When analyzing hazardousness as it varies from place to place or decisions that are made or not made to mitigate hazardousness, the political, economic, and social factors that put boundaries around individual and societal choices cannot be ignored. For example, the next earthquake in Afghanistan will almost inevitably lead to more female deaths than males purely because women have little access to hospital care. Hence, the key elements relate to individuals' situational contexts, frequently associated with societal norms, which relate specifically to those variables that influence vulnerability.

Many of these social factors affect perceptions of a hazard, which in turn determine human behavior. Indeed, it is through the perceived world that individual decision making occurs. This perceptual context is dynamic and variable over time and space, and is perhaps most difficult to model from a scientific perspective. Questionnaire surveys of disaster victims have attempted to elicit perceptual worldviews, but these provide merely snapshots of a changing and evolving scenario. What people say they will do under certain circumstances does not always coincide with what actually happens later. For instance, people behave according to perceived rather than objective reality; if a tornado hazard is considered a low risk or a hurricane is perceived as survivable, the individual concerned may take little or no action to mitigate impending events. Consequently, further work is needed to put perception and situation in context.

Disaster experience affects individuals' perceptions and in some circumstances generates what is called gambler's fallacy. After a disaster has occurred, a common consensus is that another one is not expected for some time, that is, bad luck will be followed by a run of good luck. This is common in the United States with respect to floods. It is not unusual, for instance, to hear flood victims' state that they have experienced the 100 year flood, and consequently will not experience a similar flood in their lifetime. As many victims of the 1993 floods of the upper Mississippi River found out, this is a fallacious argument; severe flooding occurred several times that summer in some communities and again just 2 years later in others.

The implementation of mitigation measures can also lead to a false sense of security. The construction of a dam or levee to control flood waters, the building of a sea wall to repel tidal surges, or elevating homes above expected storm surge levels generate feelings of safety such that all risk has been eliminated. However, such projects are designed to specific standards; so the risk remains that one day a larger event will occur. This false sense of security generates behavior that is not always conducive to minimizing losses. For instance, construction of structural measures or even the implementation of nonstructural adjustments, such as zoning ordinances, can initiate increased development on the 'safe' side of the project, a process termed the levee effect. Of course, all projects have limitations as defined by the design standards, which will fail when compromised. Consequently, losses can be disastrous and, over the long term, possibly greater than if the project had never been implemented because of the increased development; there are many examples of increased losses, such as when levee systems fail, and New Orleans in Hurricane Katrina is a prime example. Therefore, attempts to control or even live with events influence individuals' and communities' perceptions of risk and hazardousness.

From Mitigation to Resilience

Individuals may or may not be willing or able to protect themselves from extreme events. As a result, various mitigation strategies are used in an attempt to protect communities and to increase resilience. This has traditionally been provided by structural measures, using engineering works to protect an area or buildings, but increasingly, nonstructural strategies, usually involving land management and policy oriented approaches, have been implemented to reduce risk by avoiding development in hazard prone areas.

In either case, those evaluating hazard mitigation strategies must consider not only the broader impacts of mitigation alternatives, but also the context in which strategies will be implemented. Structural mitigation makes places safe up to a design standard, as noted above. However, such mitigation does not eliminate the risk completely; so people remain vul nerable to disaster. In the United States, it is recom mended that those located in areas protected by levees still develop evacuation plans and purchase flood insurance in case the levee fails. In all cases, though, the political will to implement such measures is not the only requirement; human, financial, and technological resources are needed to ensure that appropriate measures are taken, that they are imple mented properly, that they do not burden different sectors of society, and that they improve community resilience, in addition to reducing risk. Because this involves the sharing and probable redistribution of these resources away from other priority areas, pol itical motivation is sometimes lacking.

Mitigation policies that take land out of development whether to build a mitigation structure or to restrict the use of land and property at risk raise other issues. Land use regulations (such as zoning) limit what can and cannot be done with land. In some nations, land can be taken with little or no legal repercussions, but in others, the ramifications can be quite significant depending on political situations. In the United States, land can be assumed under public ownership through eminent do main if it benefits the larger population and just com pensation is paid. However, such a denial of land use rights to property owners has been challenged as un constitutional; some legal cases have demonstrated that the state has authority to do this under hazard miti gation strategies, while other court decisions have as serted the rights of the landowner. The rights of individuals in this instance must be evaluated against the common good. By comparison, centralized political systems may have fewer problems removing people from hazardous environments or keeping them from moving there in the first place. For example, the Chi nese government has relocated thousands of people to construct dams and other water resource projects. In other instances, strong disincentives exist to keep people from moving to urban areas. While not all of these policies were enacted with natural hazards in mind, the success of hazard mitigation policies will differ de pending on political context.

Integration

Environmental hazards incorporate physical and human elements of geographical research. Both are important to understanding the nature of the events, their impacts, and any means for mitigation. The characteristics and impacts of Hurricane Katrina along the Gulf Coast of the United States, and in New Orleans in particular, are illustrative of this intersection of geographical themes. In this regard, natural hazards research has made tremendous progress, moving from almost exclusive concern with natural phenomena and technological fixes to one that recognizes processes of the human use system, and specifically the complex web of social, political, and economic forces that exacerbate or attenuate the impacts of events. This has led us to the following conclusions:

1. Disasters are ubiquitous, and spatial patterns of particular events are generally well addressed and understood.
2. Risk or exposure to hazard is well defined though not always acted upon since preparation and mitigation strategies are usually limited.
3. The temporal and spatial scales of events must both be evaluated in order to develop a comprehensive understanding of hazards and disaster.
4. Disasters involve a complex web of factors, including physical processes and human elements; social, eco nomic, and political forces in various forms and at local, national, and international levels interact with the geophysical world to exacerbate or attenuate hazardous conditions.
5. A clear understanding of vulnerability in all its con texts and an appreciation of the role of marginal ization are the keys to effective mitigation strategies and reduction of losses.
6. Individuals take actions to protect themselves from events based on situational and cognitive factors and are often constrained by circumstances often beyond their control.
7. Mitigation measures, whether structural, nonstructural, or both, can lead to a misplaced sense of safety. At the same time, they can be very effective at protecting communities and improving their resilience.

See also: Environment; Environmental Justice; Globalization, Cultural; Globalization, Economic; Nature.

Further Reading

Alexander, D. (1993). *Natural Disasters*. New York: Chapman Hall.
Bankoff, G., Frerks, G. and Hilhorst, D. (eds.) (2004). *Mapping Vulnerability: Disasters, Development and People*. London: Earthscan.
Cutter, S. L. (2006). *Hazards, Vulnerability and Environmental Justice*. London: Earthscan.
Hewitt, K. (1997). *Regions of Risk: A Geographical Introduction to Disasters*. Singapore: Longman.
Montz, B. E., Cross, J. and Cutter, S. L. (2004). Hazards. In Gaile, G. L. & Wilmott, C. J. (eds.) *Geography in America*, ch. 29, pp 481 493. New York: Oxford University Press.
Oliver Smith, A. and Hoffman, S. M. (1999). *The Angry Earth: Disaster in Anthropological Perspective*. New York: Routledge.

Slovic, P. (ed.) (2000). *The Perception of Risk*. London: Earthscan.
Smith, K. (2001). *Environmental Hazards: Assessing Risk and Reducing Disaster* (3rd edn.). London: Routledge.
Tobin, G. A. and Montz, B. E. (1997). *Natural Hazards: Explanation and Integration*. New York: Guilford Press.
Wisner, B., Blaikie, P., Canon, T. and Davies, I. (2004). *At Risk: Natural Hazards, People's Vulnerability and Disasters* (2nd edn.). London: Routledge.

Relevant Websites

http://www.cred.be
CRED: Center for Research on the Epidemiology of Disasters.
http://www.crid.or.cr
CRID: Regional Disaster Information Center, Latin America and the Caribbean.
http://www.em dat.net
Emergency Disasters Data Base.
http://www.colorado.edu
Natural Hazards Center: University of Colorado at Boulder, Colorado, USA.
http://www.reliefweb.int
OpenForm Relief Web.
http://ochaonline.un.org
United Nations: Office for the Coordination of Humanitarian Affairs.

Environmental Health

S. J. Elliott, McMaster University, Hamilton, ON, Canada

Glossary

Confounding A variable is said to confound a relationship between an exposure and an outcome when it is a plausible, third variable that could potentially explain the relationship. For example, there is a suspected relationship between exposure to polychlorinated aromatic hydrocarbons and low birth weight babies. However, if the population of interest being studied also had high rates of smoking cigarettes, that third variable – smoking – may actually be the strongest explanatory variable. In short, the confounding variable(s) confound(s) the suspected relationship between exposure and outcome.

Epidemiology The study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to the control of health problems.

Hazard A source of potential harm, damage, or adverse effect.

Odds Ratio The likelihood of an outcome occurring, given the exposure of interest. For example, an odds ratio of 1.5 in a population exposed to high levels of smog means that an individual in that population has a risk of one and a half times that of an unexposed individual of dying prematurely.

Risk The chance (or probability) that a person will be harmed or experience an adverse health effect if exposed to a hazard.

Introduction

The relationship between human health and the environment is of great interest to researchers, policy makers, and the general public. Despite the realities of the situation, environment and health represent a rather nebulous construct: hard to see, hard to define, hard to measure. Einstein defined the environment as everything that is not me; hence, impossible to measure. This article describes the evolution of the geographer's interest in environment and health research; articulates the difficulties inherent in measuring the impacts of the environment on human health; and explores salient issues related to psychosocial health, environmental justice, and developing policy under scientific uncertainty.

From Medical to Health

The history of medical geography is well documented and need not be repeated here. It is important to set the stage, however, for the role of geographers in health research in the present day. Briefly, our current place within health research is the result of a tripartite evolution of: models and definitions of health; medical to health geography; and, the rise of the population health perspective. The evolution of thinking about health as more than simply the absence of disease to complete physical, social, and emotional well being (World Health Organization (WHO)) set the stage for a revisiting of the biomedical paradigm and for the development of models of health and well being that emphasized not only multiple determinants of health but the role of the social, cultural, political, physical, and economic environments within which health and well being are shaped. Alternative health frameworks began to emphasize the role of socio environmental factors, in addition to lifestyle and biology, in the production of health and illness, while at the same time diminishing the role of the formal healthcare system. Health is now a resource for everyday living that allows individuals to manage, cope with, and even change their environments.

During the late 1980s geographers began to explore their own role in research, relative to those evolving models and definitions of health. In the Canadian context, this took the form of lengthy but friendly debates and a change in the name of our professional association from medical geography to the geography of health and healthcare. Our UK colleagues soon followed suit. Overall, this was not seen as a move to exclude our colleagues working in traditional, disease ecology areas of research. Rather, the intent was to broaden our focus by formally embracing the evolving models and definitions of health. A substantial component of this exploration of our role within medical/health research was contained within a parallel exploration of the role of medical/health geography within the larger subdiscipline of social/human geography.

Concomitant with this shift was a shift from quantitative/positivist approaches to more theoretically informed, qualitative approaches. In this context, the questions have determined the method. That is, explorations of the meaning of health and illness in place require an alternative epistemology, one that recognizes individual situated experience as a valid way of knowing.

Within the broader health research community, at around the same time, we saw increasing empirical support for a population based framework for under standing the broader determinants of health. That is, there was increasing recognition of the decreasing roles of biomedicine, as well as the formal healthcare system in the production of health and illness, while the import ance of socioenvironmental factors and their interaction with individual level biological and behavioral variables took on increasing importance. These notions are codi fied in the population health perspective. Although not without its critics, it has become the guiding framework for much Canadian, US, and European Union (EU) health and health promotion policy. This health frame work moves beyond traditional definitions of health as simply the absence of disease; embraces the evolving definitions that expanded to include social and emo tional, as well as physical properties of health; and opens the door to consideration of broader determinants of health that include the social, cultural, and physical en vironments, as well as individual lifestyle behaviors.

The population health perspective was embraced by health geographers; questions of the intersection of in dividual level biological and behavioral variables with social and environmental factors are central to the health geography research agenda.

From Health to Environment

Clear articulation of what we mean by health is a useful point of departure for our understanding of the relation ships between environment and health. Defining the en vironment is complex. In the context of environment and human health, we often think of the physical environment (e.g., air quality) and its impacts on human health (e.g., asthma). But what of the social environment (e.g., social capital), the economic environment (e.g., unemployment rates), the cultural environment (e.g., role of women), or the political environment (e.g., access to opportunities and institutions)? These environments, too, have been shown to influence human health. This is why the WHO, in a meeting in Sophia, Bulgaria in 1993, crafted a very broad definition of environmental health that includes quality of life in the concept of health and refers to the influence of the physical, biological, chemical, social, and psychosocial environments on human health. This definition has a further, prescriptive component addressing the correction, control, and prevention of those factors in the environ ment that can (potentially) adversely affect the health of present and future generations.

The classic example of an environment and health relationship found in virtually every health geography text book, and expounded upon in every introductory health geography lecture, is the ingenuity of Dr. John Snow of London, England, who solved the mystery of the cholera outbreak of 1858 by mapping the location of deaths due to this infectious, waterborne disease. At the time, of course, most households did not have piped running water. Rather, there were – scattered throughout the urban neighborhoods of inner city London – stand ing water pumps from which local residents drew their daily requirements. When cholera started claiming so many lives in the mid 1850s, there was no apparent ex planation. At this time, there was no knowledge of the links between cholera and contaminated water. There were, rather, ill defined notions that cholera – and other infectious diseases – were due to miasma or bad airs. Snow was obviously suspicious of these notions and de cided to map the deaths from cholera occurring across the city. What he found was that there was a cluster of deaths in the vicinity of the Broad Street water pump. The local water supply was drawn from the Thames River, commonly known to be essentially an open sewer for both industrial and human wastes. Several water companies drew their raw materials from the Thames, and 'treated' the water before distribution to local households. Clearly, some water companies did a better job than others. Snow ceremoniously cut the handle off the Broad Street pump in order to seal his discovery.

Natural Hazards

The work of early geographers (Burton, Kates, White) on the human impacts of natural hazards (e.g., floods, earthquakes, volcanoes) forms the foundation for the modern study of geographies of environment and health. Human impacts of these hazards focused on the loss of life and property, loss of community and social func tioning, loss of home, as well as perceptions of risk, the social amplification of risk, and the role of individual choice with respect to risk and coping with/adapting to risk. This large body of literature can be summarized in a very few points: natural hazards and natural events are two very different concepts; natural hazards and their impacts are unequally distributed across space; natural hazards and their impacts are increasing with time. With respect to the first point, there are catastrophic natural events that occur in our world on a regular basis: fires, floods, droughts, earthquakes, and tornados. Indeed, some catastrophic natural events result in very positive benefits (e.g., a lightening strike that causes a forest fire can result in badly needed ecosystem regeneration). Natural events become natural hazards when they impact upon human populations. Further, the extent of the impact on human populations is affected by actions taken by humans in the social, political, economic, and cultural realms. For ex ample, the extent of the flood damage wrought by Hur ricane Katrina in the US south in 2006 was exacerbated by the poor state of the levees originally constructed to protect residential areas from rising flood waters. Another example is the Bangladesh Cyclone of November 1970, wherein one third of the population of low lying coastal areas/reclaimed islands died. There were a number of factors that contributed to this death toll, some very simple (i.e., the cyclone occurred at 11:00 p.m.; the only radio and telegraph station in the area closed regularly at 11:00 p.m.) and some more complicated (i.e., the gov ernment reclaimed low lying coastal areas and islands through sophisticated engineering technology and en couraged agricultural settlement in these areas).

It is clear from the two examples cited above that natural hazards and their impacts are unequally experi enced across the globe. In the aftermath of Hurricane Katrina in the US south, for example, there have been heated debates about the response to the people of the area, claiming that the response was slow in coming and inadequate when it did arrive, primarily because those affected were primarily poor people of African American descent. The poor farmers of the low lying coastal zones of Bangladesh may have felt the same way. Indeed, be tween 1947 and 1981, as a result of floods, earthquakes, and cyclones, approximately 85% of all lives lost were in Asia; 4% in the Caribbean and Central America; and 4% in South America. Only 2% of lives were lost in Europe and only 1% in North America.

Unfortunately, the toll (of human lives, social func tioning, property damage) of natural hazards will increase over time, not decrease. There are essentially three rea sons for this. The first is the growth and changes in distribution of human populations. That is, not only is the human population growing exponentially over the sur face of the globe, but its distribution is rapidly changing from one of a rural–urban mix to one that is primarily urban. These large concentrations of human populations create the opportunity for hefty tolls from natural dis asters. The second reason is a simple one: we are getting better at recording the existence and parameters (i.e., magnitude, frequency, duration, spatial dispersion) of natural hazards, as well as their impacts (i.e., numbers of casualties, property damage, etc.). Finally, modernization of some developing countries has actually made them more vulnerable to acute natural events. For example, much commercial development in Kingston and other parts of Jamaica is taking place on flood plains, as well as alluvial and unconsolidated soils, making them even more susceptible to earthquakes, as well as floods.

Technological Hazards

Much of what we study in the present day from an en vironment and health perspective deals primarily with technological, human made hazards, as opposed to the natural hazards discussed above. These hazards may be acute, such as an accident (e.g., the chemical plant ex plosion in Seveso, Italy; the nuclear accident at Cher nobyl; the train derailment at Bhopal, India causing a major chemical exposure) or they may be chronic (e.g., the virtual disappearance of the Aral Sea in Uzbekistan; the contamination of the Great Lakes Basin in North America). Hazards may be point source (e.g., the Sydney Tar Ponds in Sydney, Nova Scotia, Canada – the most contaminated industrial site in Canada) or ambient (e.g., air pollution across the EU and much of the Western industrialized world). They may be known, well under stood, and well documented, scientifically (e.g., the effects of lead exposure on the intellectual and physical devel opment of children; the relationship between asbestos exposure and asbestosis of the lung) but many, if not most, are suspected, not well understood, the subject of equivocal science, and hotly debated in the literature, the policy realm, and the media. These include, for example, the relationship between low level exposure to electro magnetic fields (EMFs) and brain cancer. EMFs come from a range of small appliances found in most homes in the world (i.e., refrigerators, hair dryers, clock radios, microwave ovens) but the major source of public concern, and focus of most research, is overhead power lines that carry electricity. While some studies have shown a rela tionship between exposure and outcome, others have not. These are studies done by credible scientists and pub lished in respected, peer reviewed journals. However, the science remains equivocal. Other suspected hazards come into and out of fashion. For example, the potential for links between radon and lung cancer was of tremendous interest in North America and Europe in the 1970s and 1980s. Radon is a by product of the natural breakdown of uranium. As a result, it is naturally emitted from rocks, as well as in the uranium refining process. Given that it is naturally occurring, its distribution is unequal across the globe. There are some areas of the world where it is naturally occurring in high quantities (e.g., Pennsylvania in the United States). However, it went out of fashion for a while; people forgot all about it as a potential risk for lung cancer. It has recently been making media headlines again, however, and home buyers are being encouraged to purchase radon detectors (at a cost of about US$20) to take into any new home they may be considering pur chasing. Further, some exclusive tourist resorts in Greece now produce warning pamphlets for their visitors on the potential relationship between radon exposure and lung cancer. This is due to the relatively high levels of radon being detected in Turkish baths and hot springs often frequented by European and North American tourists. Other environment and health relationships are seem ingly common knowledge. Included in this category would be the relationship between adverse air quality or air pollution and asthma. Again, however, a thorough search of the research literature finds only equivocal evidence for this relationship. Indeed, a study conducted in the late 1970s, early 1980s in Hamilton, Ontario, Canada (a city located on the western end of Lake On tario that built its economy on steel production and other heavy industry), often held up as the gold standard for air quality and health research, firmly concluded that air pollution did not cause asthma; rather, air pollution could be responsible for exacerbating an already existing re spiratory health problem in young children. Indeed, re cent research about asthma from leading world investigators tells us two things: first, that we know very little about what causes asthma; second, if we had to choose a culprit in terms of causation, we should focus as much on the indoor as opposed to the outdoor en vironment. This is where difficult questions of policy arise: we still have to make decisions to set policy and guidelines, for example, around air quality in order to protect the health of the public. How do we do that based on such equivocal science?

Design, Measurement, and Evaluation

At this point, the key question arises: given the level of interest and activity around environment and health re lationships, why is the science so equivocal? Essentially, this has a lot to do with the complexity of designing environment and health research, measuring exposures and outcomes, and evaluating the evidence.

Design

There are several different study designs to choose from in any health research endeavor. The strongest of these is the randomized controlled trial (RCT). Herein, the re searcher has control over both the exposure of interest, as well as the sample exposed. Typically, RCTs are under taken to test a new treatment for a known illness, often a newly developed pharmaceutical. As such, the researcher would select a control group (who did not get the newly developed pharmaceutical or would receive a placebo) and an intervention group (who would receive the newly developed pharmaceutical) and assess the impacts, (ideally in a double blind manner; i.e., neither the re searcher nor the subject would know if they were re ceiving the newly developed pharmaceutical or the placebo). Widely heralded as the strongest study design, yielding the most credible results, it is impossible for use in an environment and health study. For many reasons, including ethical ones, it would be impossible to expose, say, a group of children to lead for a period of time and then assess the impacts compared to a control group not exposed. As an aside, we often have access to 'natural laboratories' in environment and health research. For example, the 1976 explosion at a chemical factory in Seveso, Italy allowed researchers to follow the workers and residents of the surrounding community for the next 30 years to assess the long term health impacts. However, because this is not an RCT, the researchers cannot control for all of the potential confounding factors that may also be influencing the health outcomes of interest.

Since we cannot do RCTs, then, in environment and health research, we are limited to observational studies – essentially, observing what has already happened. The first form of observational study is the case control de sign. Herein, a number of cases are identified – for ex ample, a cluster of leukemia cases – and these individuals are studied for their exposure histories. This was the study design used in the famous case of Woburn, Mas sachusets. In this case, a seemingly inexplicable cluster of childhood leukemia cases occurred in a residential neighborhood in this small town. Area residents sus pected the local water supply as the culprit. The water had not been right for some time – it smelled of chem icals and caused rashes and burning on human skin. These area residents worked long and hard to convince local authorities and physicians that there was indeed a link between the leukemias and the water supply. A local industry was subsequently convicted of contaminating a city well, and the entire subject was made into a Holly wood movie, *A Civil Action*, staring John Travolta.

A second type of observational study is the cohort study. Herein, we sample on the basis of exposure, as opposed to outcome (as in the case control study used as an example, above). Cohort studies can be one of two types: retrospective or prospective. In the context of the former, a cohort of individuals are identified who are thought to have experienced an exposure. The cohort is then studied to determine what, if any, adverse health effects they might have experienced in the past. This is the typical type of study done in environment and health research. Often, this is because a community realizes it may have been exposed to something (an air pollutant, a leaking landfill, a contaminated water supply) as a result of some ill defined symptomatology or an inexplicably large number of cancer cases in a neighborhood. This self identified cohort may then be studied to assess po tential health effects experienced in the past during the exposure. An example here is the Sydney Tar Ponds in Sydney, Nova Scotia, Canada. This site, known as the most heavily contaminated industrial site in Canada, saw approximately 100 years of coal smelting and steel pro duction, characterized by inappropriate waste disposal practices. The site is now contaminated with 700 000 tonnes of known carcinogens. Area residents living within 5 km from the site are very concerned about the impacts of the exposures on reproductive health outcomes. An ecdotal reports in the area reported higher than normal levels of miscarriages, still births, infertility, and low birth weight babies. A retrospective cohort study of the local residents documented the reproductive histories of 500 women living within 5 km of the site in order to assess whether the adverse reproductive outcomes they experienced happened at levels higher than would have been expected in an unexposed population.

A prospective cohort study would also be characterized by sampling on the basis of exposure, but would follow the cohort forward in time to document any adverse health outcomes. An example of a prospective cohort study in environment and health would be the explosion at the chemical factory in Seveso, Italy in 1976. Both workers and members of the local community were exposed to high levels of dioxin. A team of occupational epidemiologists followed this cohort forward for 30 years to assess whether or not adverse health outcomes occurred in the cohort at rates higher than would have been expected in an unexposed population.

Finally, from a humanistic perspective, we can conduct qualitative studies in environment and health. For example, the study referred to above in the Sydney Tar Ponds not only involved a large scale epidemiologic survey of women's reproductive histories, it also involved a smaller set of qualitative interviews with families living in the area in order to try to understand the psychosocial impacts of the exposure on their daily lives, as well as anxieties around reproductive health and decision making.

Measurement

Complicating things further are issues of measurement; that is, it is no easy task to measure exposures, outcomes, and ascertain the pathways through which individuals are exposed.

Exposure

When individuals report to their friends, family, physician, local public health official, and so on that they feel they have been exposed to some environmental hazard, it is very difficult to measure that exposure. First, the exposure has happened in the past and recall bias (do you remember how often you spent time in your basement in the last 5 years? How much time you spent in the shower or bath? How much tap water you consumed? How much time you spent in the garden or yard of your house?) will impact on how accurately we are able to measure the exposure. Second, often the exposure comes from an illicit, undocumented source; for example, chemical contamination by an industry to the public water supply in Woburn, Massachusetts. Third, we rarely know what the exposure contains. For example, when Lois Gibbs first discovered yellow ooze seeping into her basement in her Love Canal neighborhood home, all she knew was it was yellow ooze. Toxicologists then had to deconstruct the ooze; this is very difficult to do. When a toxicologist goes fishing for a toxin, s/he needs to know what s/he is fishing for so that it can be identified. Are we looking for polyaromatic hydrocarbons? Dioxin? Dieldren? Trichloroethylene? Each of these has a unique identifying character. We can't just run yellow ooze through a chemical scanner and get a read out of what is in it. Fourth, we know very little about the synergies between chemical contaminants. That is, we may have some confidence of the relationship between trichloroethylene (a chemical primarily used in the process of dry cleaning) and bladder cancer, but what if the trichloroethylene is mixed with polycyclic aromatic hydrocarbons (PAHs)? Is there a synergistic effect that can actually cause more harm to human health than if one were exposed to these chemicals in isolation? This is why we often see the phrase chemical soup used when referring to the yellow ooze seeping out of Love Canal, or the waters of the Great Lakes Basin. Finally, all of this science is so very new to us. It is not that long ago (1950s, 1960s) that Dow Chemical spent much money and energy developing television advertisements about the wonders of chemicals and how they were going to change our lives. We still have much to learn. Sometimes our technology outstrips our knowledge – that is, we can now measure contaminants in such small amounts (from parts per million to parts per trillion) yet have very little if any knowledge of the impact of exposures at that level on human health.

Outcome

If we can ascertain what an individual or a community has been exposed to, we must then determine what health outcomes we should be documenting. This is a very difficult task, for a number of reasons. First, we simply do not know how community based (as opposed to occupational) exposures to these chemicals will impact on human health. Therefore, we don't know what to look for. In the highly regarded Upper Ottawa Street Landfill study undertaken in Hamilton, Ontario, Canada, researchers documented entire health histories of area residents who unknowingly had been exposed to (illegally dumped) liquid industrial waste for over 5 years. Every organ system was examined. This was extremely time consuming, and resource intensive. But not knowing what to look for, these researchers looked for everything. What did they discover? Not much – some red, itchy eyes, some psychosocial stress. Why? It could be that the dose that area residents were exposed to was not sufficient to have a health impact (it is the dose that makes the poison). Second, it could also be that, as in many cases, the health impacts experienced are ill defined symptoms that are difficult to trace to any specific exposure. For example, we often hear reports of headaches, rashes, stomach problems, sleeplessness, fatigue, psychosocial distress. These symptoms could be due to any number of factors; again, how do we control for all the confounding factors that exist in a community laboratory? The third reason why these researchers, despite their very thorough examination of the population, may have found very little in the way of health effects, relates to the issue of latency. That is, if we are concerned about cancers or some other type of chronic illness, it takes time (15 years? 20 years?) for such illnesses to develop. Following an exposed cohort forward for 20 years is very expensive research and fraught with difficulties (e.g., control for confounders, sample attrition).

In short, the measurement of health outcomes is no easy task. This is further complicated by the fact that we are often very interested in measuring not only the physiological health impacts of an exposure, but also the psychosocial health impacts. Psychosocial impacts can be defined as the combination of distress, dysfunction, and disability manifested in a wide range of social, behavioral, and psychological impacts in individuals, groups, and communities as a consequence of actual or perceived environmental contamination. Psychosocial impacts are important to study for at least two reasons. First, they are a health impact in and of themselves (recall the definition of health cited above); one cannot have full health when one is anxious, worried, distressed. Indeed, despite whether the exposure is actual or perceived, these are real effects that have real consequences. As the sociologist, W. I. Thomas expounded in the 1960s, if a situation is defined as real, it will be real in its consequences. For example, the study in the Sydney Tar Ponds in Sydney, Nova Scotia, Canada found no excess levels of adverse reproductive outcomes (i.e., still births, miscarriages, low birth weight). However, the perception of area residents of the impacts of the Tar Ponds on reproductive health was so strong that they were making decisions about fertility based on this perception: some families were putting off having children until the site was cleaned up, others were deciding not to have children at all, still others were choosing to adopt from outside the community. The second reason why it is important to study psychosocial impacts of exposure is the relationship they have with physiological health. For example, when the chemical factory in Seveso, Italy blew up in 1976, workers and local residents received high doses of dioxin exposure; researchers who followed this cohort for over 30 years now expected to find elevated rates of cancers, particularly liver cancer. They did not. They found, however, a fivefold increase in cardiovascular disease. Why? These researchers did not really know. However, they speculate that the exposed cohort worried so much about whether or not they would develop cancer, that they worried themselves into cardiovascular disease.

Pathways

There have been many situations where an actual documented environmental hazard exists, and individuals and communities are worried about their health. But before we can claim a relationship, there has to be a clear pathway for the exposure. For example, the contamination in Sydney, Nova Scotia from the Tar Ponds site was to a tidal estuary. A moratorium on fishing was established; area residents do not obtain their water from this source. So, if there is an exposure, what is the pathway? Such pathways are clearer in the case of Bhopal, where a train derailment led to the release of 40 tonnes of methyl isocyante into the air. Individuals could then breathe this in; this is a pathway for exposure. Another clear example is the radiation exposure at Chernobyl. It was relatively straightforward to document immediate exposure through documentation of radiation sickness and later through the consumption of contaminated crops and livestock. In short, a contaminant somehow has to have a route into the body – through ingestion, inhalation, or through the dermis.

Evaluation

Once we have measured exposure and outcome, it behooves us to evaluate the existence and strength of the relationship between them. We do this by calculating odds ratios; that is, the likelihood of an individual experiencing the health outcome of interest given the exposure of interest. For example, at Seveso, the researchers documented an odds ratio of 5 for cardiovascular disease, indicating that those exposed were 5 times more likely to develop cardiovascular disease than those not exposed. An odds ratio of 5 in environmental health research is very high; typically, we see odds ratios between 1.5 and 2. While these odds ratios may be statistically significant, they may or may not be clinically significant. For example, studies between air pollution and premature mortality in Montreal, Canada indicate an odds ratio of about 1.03 of dying prematurely on a bad air day. While this is borderline statistically significant, we might consider it to be clinically significant because we are applying that odds ratio to a large urban population (i.e., approximately 2 million people). Thus, this environmental exposure may not be a high risk but affects a lot of people. Perhaps then there should be a public health response.

In addition, even if we have a relatively robust odds ratio, we still need to ascertain whether or not there is a causative relationship between x and y. In the 1970s, Sir Austin Bradford Hill gave an after dinner speech to the Royal Association of Physicians wherein he outlined what are now euphemistically known as the Bradford Hill criteria or the nine diagnostic tests of causation. Briefly, Bradford Hill asserted that one could not claim causation unless most if not all of these criteria were met. They include:

1. strength of association,
2. consistency of association,

3. specificity of association,
4. temporality (i.e., that the cause comes before the effect in time),
5. existence of a biological gradient or dose–response curve,
6. biological plausibility of the association,
7. coherence of the relationship with what we already know,
8. experimental data, and
9. analogous evidence in the literature.

Conclusion

Society finds itself, therefore, in a seemingly untenable position. In the face of existing (air pollution, waste, contaminated water, nuclear waste); (re) emerging (West Nile virus, influenza); and new (global climate change, avian flu) environmental risks, the public is demanding zero tolerance for risk. Indeed, in this postmodern society in which we live, the public feels an entitlement to an environment free of risk. There is a substantial literature on risk and risk perception. Suffice it to say the public is guided in its perception of risk by two factors: what is unknown and what is dreaded. Further, there is sub stantial difference between the factors that affect the risk perceptions of experts (probabilities of mortalities) and the lay public (the dreaded possibilities; the potential impacts for self and family). Regardless, the public is just acquiescing anymore, particularly given the unequal na ture of the distribution of risks and benefits in the context of environmental risk. The growing literature around environmental justice has documented the unequal treatment of some members of society with respect to those who are exposed and those who reap the benefits. And of course, as in all other inequities it is the mar ginalized (the poor, those of color, Aboriginal persons, women, the disabled) who bear the brunt of the risk. While some have suggested that the risks we experience in our postmodern world have indeed become dem ocratized, there are others who strongly disagree. While residents of the entire globe may now be at risk for the public health impacts of global climate change, it is those in the marginalized groups who are least able to cope with the impacts.

So, how do we move forward? Some would suggest that policymakers take the best of what they can from the research and science that is available and follow the precautionary principle. That is, Article 15 of the 1992 Rio Summit on Building a Sustainable Earth essentially declares that we don't have to have all the science in place before we are compelled – morally and ethically – to make a reasonable decision in order to protect the health of the ecosystem, including human health. In 1964 the US Surgeon General declared that smoking cigarettes caused lung cancer. Over the next 40 + years, we have expended much public health effort to get that message out. However, it was not until 1994 that we finally understood the biological mechanisms involved. Should we have waited?

See also: Development I; Development II; Health and Development; Health Geography; Health Inequalities.

Further Reading

Beck, U. (1992). *Risk Society: Towards a New Modernity*. Newbury Park, CA: Sage Publications.
Bullard, R. D. (1994). *Dumping in Dixie*. Boulder: Westview Press.
Burton, I., Kates, R. W. and White, G. F. (1993). *The Environment as Hazard*. New York: The Guilford Press.
Greenberg, M. (ed.) (1987). *Public Health and the Environment*. New York: Guilford Press.
Hill, A. B. (1965). The environment and disease: Association and causation. *Proceedings of the Royal Association of Medicine* 58, 295 300.
Klinenberg, E. (2002). *Heat Wave: A Social Autopsy of Disaster in Chicago*. Chicago, IL: University of Chicago Press.
National Institutes of Health (2004). *Progress and Promise in Research on Social and Cultural Dimensions of Health: A Research Agenda*. Washington, WA: National Institutes of Health.
Slovic, P. (2000). *Perception of Risk*. London: Earthscan Books.

Relevant Websites

http://www.cdc.gov
Centers for Disease Control and Prevention, Atlanta, GA, USA.
http://www.europa.eu
European Union Environmental Policy.
http://www.hc sc.gc.ca
Health Canada.
http://www.un.org
United Nations Millennium Development Goals.
http://www.epa.gov
United States Environmental Protection Agency.
http://www.who.int
World Health Organization.

Environmental Justice

A. A. Lehtinen, University of Joensuu, Joensuu, Finland

Glossary

Communities of Justice Alliances or co-associations, both local and translocal, that (re)produce the criteria, practices, and relations central to the shared formation of a sense of justice.

Corrective Justice An ethical or juridical approach that aims at compensating the past injustices.

Discursive Regulation Both informal and formal negotiations developed through market competition and socio-environmental pressure. Practices of local–global co-governing where all the concerned stakeholders, with their diverse expectations, condition specific operational spaces.

Distributive Justice An ethical or juridical approach that focuses on the distribution of benefits and burdens, including the outcomes of decisions or events.

Modernization Development phase and paradigm that emphasizes human reason and rationality, accompanied by the ideas of linear progress and absolute truth as well as a confidence on human ability to control nature. The modern world view also contains strong ideals of equality and liberty.

Neoliberal Environmentalism The commodification and privatization of natural resources by favoring the market mechanisms.

Power Regimes The arrangements and conventions of governing the complex interrelations between public and private interests, also characterized by a vested presence of principles and norms that support the existing forms of domination.

Procedural Justice An ethical or juridical approach that is concerned about the degree of inclusion and exclusion in decision making.

Retributive Justice An ethical or juridical approach dealing with penalty or punishment for unjust conditions or actions.

Risk Society A society where the production and distribution of social and environmental risks, including their latent and unpredictable transformations, have become a central characteristic. In risk progression the occurrence of a hazard meets the vulnerability of the community.

Social Vulnerability The culturally inherited and socially produced skills, capacities, and conditions of an individual or community to cope with the changes in the environing world. The ability to resist and recover from harmful transformations.

The Environmental Justice Movement in the USA

The Environmental Justice Movement emerged in the early 1980s in North Carolina, USA, due to a local dispute over toxic waste dumping near a neighborhood of African American people. The movement emphasized from the beginning that environmental problems cannot be solved without unveiling the practices maintaining social injustices. Many of the traditional environmental organizations were targeted by this critique.

The North American debate on environmentalism and justice has developed via deep contradictions, which reflects the delicate historical nature of the issue. Racial and other social questions are often intentionally avoided by dedicated nature conservationists and this frames the whole tradition, initially established through the struggles for nature parks and wilderness areas. The history of the Western idea of nature is part of the history of the white middle class that had learned to appreciate the esthetic value of wilderness. Purified and thus white areas of nature therefore symbolize the areas of white power. This was made clear during the formative years of nature conservation in the USA when the indigenous First Nations were forced to leave their homelands which were overlapping with the proposed national parks of Yellowstone and Yosemite.

Environmental history in North America is rooted to the expansive colonial control over the resources of newly settled areas. The European colonization of the continent turned into a cruel genocide of the First Nations. The frontier of the settlers progressed through the wilderness and this was considered synonymous with the dawn of civilization. The colonial success was completed by the transatlantic slave trade and later immigration to urban ghettos colored by unfair divisions of welfare. Environmental justice issues turn therefore repeatedly into questions of environmental racism. The evidence telling that environmental risks tend to accumulate on ethnic minorities starkly reminds the North Americans of their violent ethnic past.

The historically specific sense of justice, built on the awareness of the interethnic violence behind the founding of modern North America, is present in the continuous re articulation of social and environmental inequalities, both local and global. The environmental justice issues are accordingly dealt with in two main forums. They are routinely taken care of by law experts in courtrooms while the critical alternatives are presented by the activist networks worrying about local–global injustices. The pragmatic lawyers and the forward looking activists share with many Europeans the ideal of just decision making and the belief in change for the better. However, despite the common background, the practical conditions and formulations of environmental justice are different in Europe.

The Sense of Environmental Justice in Europe

The specific way of reacting in Europe to environment and justice questions is connected to the recent en vironmental history of the continent. The hardships in human–nature relations met during the last decades have profoundly modified the senses of justice. Concerns about highly modified food products are, for example, widespread in the European Union, and are bound to the continental food crises of the 1990s. The risks of high tech food industry have become evident both in the kitchen routines of families and in industrial food production, and this has fueled the debate on the unjust results of tightening technological dependencies. The handling of animals in the food industry has also brought up the question of animal justice.

A new type of ethical environmental debate took shape in Europe in the aftermath of the explosion of a nuclear reactor in Chernobyl, Ukraine, in 1986. The dramatic event reminded Europeans about the vulner ability of the technological age and it also revealed the alarming risks embedded in industrial culture. During the late spring weeks of 1986, radioactive rain polluted large areas of Europe. The clouds and rain showers fol lowed the main wind directions and thus did not respect any jurisdictional borders or varying state commitments to nuclear energy. The challenge to shelter oneself from the radioactive pollution largely unified people, but the degree of individual protection varied according to the ability or willingness to change one's daily routines bound, for example, to employment histories, variations in wealth and health settings, risk awareness, home place attachment, and age identification. The catastrophe also sharpened the claims for critical and ethical reflection concerning our duties toward those innocents suffering from our errors, including the coming generations.

The drama of Chernobyl was a shock to Europe, and it led to critical rethinking about matters of industrial pro gress, environmental change, and social justice. Nuclear energy was regarded as a political, technical–economic, and moral question, and it was also seen as a condition without escape. It became an example of intergenerational (in)justices: whether positive or negative, we have to learn to live with the path dependencies developed by those before us.

Chernobyl left a mark on the European environmental consciousness, and the concern about the direction and justification of modernization grew in importance. The critique of modernization challenged those popular postmodern promises that aimed at freeing us from the historical worries of the modern. The twentieth century ended with insecurity in Europe. The Risikogesellschaft, or the risk society, had emerged as a continuation to, and a reminder of, the earlier traumas of the continent: the birth of the risk society further confused Europeans still suffering from the deliriums of the colonial–imperial past, and it brought back the nightmares of the world wars and the Holocaust.

The risk society debate re examined the ambivalent lessons of Enlightenment and progress. Finally, nature – the material sphere of our earthly being – was thoroughly contaminated, both the one surrounding us and the one in the form of intimate embodied experience. The nat ural realm, which had been carefully othered in the margins of Europe, returned in the form of radioactive rains, accumulations in various ecosystems, and new health risks. The boomerang of nature followed the logic that the othered Third World had earlier landed on the hearth of the continent. The idea of pure Europe was again challenged.

The environmental history of late modern Europe helps us to understand the way that Europeans tend to deal with environmental justice issues, such as problems with human and animal rights, quality of food, rural and fringe areas, ethnic minorities, and ecological hazards and strains. The past is present in Europe either in the form of ethical re thinking, or alternatively, clothed in pro tective security measures at the gates of Fort Europe. This tension forms the core of the debate on the con stitutional renewal of the European Union. The vision of a strong federal bastion strikingly contrasts with the idea of democratic co jurisdiction. The dispute deals with the sociospatial dynamics and control of the communities of justice.

Global Environmentalism and Justice

In the late twentieth century, global environmental concern became a key framework wherein a broad variation of changes in human–nature relations were interlinked. The global approach resulted in expansive programs in several fields of environmental governance, stretching from biodiversity conservation and atmos pheric emission regulations to certificates of green trade. Despite its widely welcomed launch, the global en vironmental turn has met serious opposition. Accusations of expert domination, ideological bias, technocratic re ductionism, ecological imperialism, and green devel opmentalism have surfaced, both from outside and from within environmental groups and organizations. Internal criticism has become the most influential as it has forced environmentalists to rethink their own role at the other end of the general antagonism between conservationists and developers.

The practical methods and tools of global environmentalism were questioned in particular. Its ideological framework also became a target of criticism: global management mentality, when tracing specific ecological causalities and necessities, was seen as resulting in social and cultural ignorance.

The development of environmentalism has evidently strengthened the role of Western ecological expertise in identifying and solving local and global environmental problems. Ecological expertise is increasingly accepted as the guiding knowledge technology in environmental management and it takes place at the cost of traditional and lay understandings of nature. It has, among other things, resulted in communicative problems. The vocabulary of experts often crucially differs from the lay modes of communication. The programs of nature conservation and environmental regulation repeatedly violate lay senses of justice. Due to the globalization of environmental concern, the question of justice is thus frequently faced. Whose environments are conserved? Whose expertise is ranked high? Who can afford, and who has the skills, to participate in the formulation of goals? Those asking these types of questions often underline the shortages of equity in decision making processes. These concerns are more or less explicitly inspired by the underlying philosophical debates on procedural justice or process equity.

Due to globalization, and the following intensification of contacts across major cultural borders, variations in the historical development of the sense of justice have emerged as matters of conflict. Tensions in environmental appreciation are central to understanding the current forms of cultural confrontations, ranging from contested ethnic spaces in urban milieus to the defense of homelands by indigenous folks, and extending across all the spatial scales, from local disputes to global contests of the leading geopolitical allies.

Human Geography and Environmental Justice

Human geographers have conducted both ethical and politico ecological studies on environmental discrimination and justice. The ethical orientation has, for example, aimed at formulating general principles to the fair division of good and bad environments. This has developed toward a meta ethical debate about the moral lessons that humankind needs to adopt to change the current alarming social and ecological trends on Earth.

A healthy environment is seen in this type of ethical thinking as a foundational right for humanity. In addition, fair treatment of humans and nonhumans is often included and then it is seen as important to recognize the value that environment has for all the inhabitants of our planet. The idea of a shared environment is central to the ethical geographical tradition.

Ethically oriented geographers have in general followed the philosophical ideas and traditions of justice and applied these to issues of environment and development. These connections explain the popularity of the models of procedural, distributive, corrective, and retributive justice in geographical ethical thinking.

On the other hand, the politico ecological tradition has emphasized the uneven characteristics of concrete social and environmental conditions. Environmental changes have been examined by focusing on the processes that cause inequality and injustice both in the politico economical practices and in the routines of everyday life. The scholars linked to this tradition have examined the transformations of natural–cultural ecosystems and communities under exploitative practices, the changing socio environmental roles of individual consumers and consumer groups, non and (inter)governmental forms of environmental regulation, as well as the features of neoliberal reorganization of local–global natures.

The geographers working with political–ecological justice have a common interest related to the question of how the use of nature is intertwined with power. Their approaches include a critical engagement founded on historical analyses of uneven socio environmental relations, often followed by suggestions of policy changes and compensatory measures. A connection to the philosophical discussion on corrective and retributive justice can sometimes be notified. Environmental racism has then been studied from indigenous and minority perspectives and they are often accompanied by proposals of compensating the past humiliation. Critical regional studies, on the other hand, have revealed the weakening of local living conditions caused by translocal development projects overlooking the local preferences. Unequal progression of social and ecological vulnerability has been analyzed in a similar manner, identifying and criticizing certain patterns of risk production.

Some political geographers have in parallel called attention to the multiscale nature of environmental justice problems. They emphasize the role of social and spatial processes in connecting individual articulations of injustice to broader changes in society. Local pollution, for example, can become an international question when linked to more general issues of industrial restructuring. Similarly, global biodiversity programs, as with any questions of inter species justice, rarely work without adjustments to meet basic human rights questions in regional implementation. In general, the scalar ambiguity of justice topics has been taken by political geographers as a challenge to develop analyses of political struggles within and across different sociospatial scales. The ways in which matters of justice are inserted into various geographical scales have in particular appealed to polit ical geographers.

New animal geography has brought up the specific question of inter species justice: socially accepted and legitimized indifference in matters of nonhuman suf fering has gained critical attention and publicity. More over, intergenerational aspects of environmental justice have been introduced both by historical geographers concerned about the rights of the past generations in the current heritage restructuring and by those participating in sustainable development discourse. The worries about future generations are then seen as signs of sensitivity to the trajectories of power across the present and coming generations.

Many critical geographers and political ecologists share the motive to reveal the practices of discrimination linked to the formation of sociospatial and ecological peripheries. The critical view concentrates on the shortages of participation and representation caused by the silencing and ignoring of certain members of the society by more or less explicit measures of ethnic, age, gender, regional, or class based discrimination, or by overlooking the nonhuman neighborhood. The practices of participation are then tied to the critical moments of inclusion and exclusion in decision making and daily routines.

When analyzing the hindrances of participation, some human geographers have underlined the critical role of communities in (re)producing the notions of nature and the sense of justice. Community is accordingly seen as the idea, place, and the relations and practices that generate what is considered socially just and ecologically sound. The community, in the form of a locality, alliance, co association, or network, thus lays both the value basis and practical frames that condition the formation of environmental preferences and sense of justice. The community is the forum of co identification by which the ideas and practices of culture–nature are formulated. This forum becomes comprehensible only while seen as developed from within the ideas, values, and experiences shared by the members of the community.

The community based environmental justice approach requires sensitivity to the uneven distribution of risks, including fears of them, faced in the daily lives of communities. The questions are then: How do major environmental changes, and the culturally varying responses to these changes, take shape in different for ums? How and by whom are these changes articulated and justified? What are the outcomes of these changes? How are they distributed? These questions bring environ mental research close to issues of distributive justice.

Late Modern Communities of Justice

Local power relationships have gone through profound transformations during the centuries of modernization. Translocal contacts have increased and this has brought along specific changes in local regime formations, eco nomic clustering, gender divisions, and intergenerational relations. This translocalization has in general differen tiated the local social groupings and increased local vulnerability to occasional interventions.

Locality is in this respect perhaps a weaker actor today than ever before. This is much due to the non predictable clustering of interventions constantly dis placing the individuals and their communities. These processes are part of our everyday life world while making our daily purchases, for example. The certificates of sustainability, demanded by critical consumers and focused both on the products and the processes behind them, are presently under contested definition. The reading of these certificates has consequently become a necessity for the careful consumer.

This development has not, however, brought along the disappearance of local identities and neighborhood ethics. Local communities are still important forums today for the development of the sense of justice. This has become evident in the criticism of the lengthening of product chains that has resulted in rising social and environmental costs of transporting. Regional preferences in production and consumption patterns, such as local food campaigns, have therefore been introduced as part of the sustainability repertoire.

Human geographers have participated in the debate on community displacement by rethinking their approaches on places, identities, and situated ethics, and this work has contributed to the understanding of the dynamics of the ethical communities. These studies have informed us about the background of the local criticism against the pressures of straightforward trans localization. Criticism often rises from the experience of humiliation based on sociospatial marginalization, and it is often strengthened by deep, heritage like experiences of injustice. Social and environmental policy measures are confronted in the peripheries in the context of the local sense of justice which is more or less internally shared, and this often tends to differ from the external translocal principles of justice derived from more uni versal standards.

On the other hand, the culture–nature relationship is also produced within the postlocal network com munities that we participate in through our daily prac tices. Today the impacts of these more or less fluid co associations can even exceed the local community effects as conditioning factors of our life worlds. Within this postlocal framework, or sphere, the basic formulation of environmental values takes place without the personal experience of the immediate strength of the environing social nature. Nature increasingly turns into a product or a symbol that is meaningful to us in the form of questions about pollution control and green trade. In this view, the postlocal communities are stretched across the local–global chains of production and consumption, both material and immaterial, dominated by the networks of globally oriented organizations and corporations.

Careful monitoring of these postlocal changes is important for the suppliers and clients of globalization leaders, but it has also become a necessity for the principal local–global actors themselves. They cannot, being thoroughly embedded in their operative networks, avoid the risks caused by continuous assessment of their competitive potential. Today, companies with global strategies need to be sensitive to the voices of their suppliers, consumers, shareholders, and other partners. A more respectful attitude toward social and environ mental pressure groups has also become a necessity. A bad image easily turns into credibility problems and worsening economic options. The postlocal networks of environmentalists need, in a similar manner, to be aware of the risks of becoming labeled as green imperialists or busy professionals for whom the planet Earth is primarily a management object.

The postlocal actors participating in the global contest have faced new risks based on the fluidity of their sociospatial environments. The credibility problems caused by wrong moves or bad images are easily reflected in the shrinking of their operative space. This type of new vulnerability has become an integral part of the local–global forums of discursive regulation, and many of the most successful global actors have become increasingly sensitive to questions of social and environmental responsibility.

Just Geography

The community based environmental justice approach can also be extended to cover the questions of academic engagement. Research often finds legitimacy by aiming at bringing up facts and viewpoints that are only partially visible in the everyday life world. Research on human communities, for example, is in general supposed to add something analytical, interpretative, or summary like to the articulations and actual views produced by the subject actors of the case studies. Research might even become emancipating while opening new views on the questions under investigation. The political relevance of research depends on its capability to cope with the concrete social and environmental factors conditioning the action spaces of the subject groups. Here the question is, how much are the research motives shaped by the academic premises of the scientific communities and how well does the researcher identify her or himself as part of the communities in the study focus? Different identifi cations and priorities result in different rankings of the most important research issues. The preliminary research design has thus a decisive impact on the political rele vance of the questions studied. Geographers sensitive to justice concerns cannot afford considering the subject groups merely as interesting cases for theory testing or instrumental means to succeed in the local–global merit markets.

See also: Environmental Security; Ethnic Conflict; Human Rights; Place, Politics of; Social Movements.

Further Reading

Abramson, A. and Theodossopoulos, D. (eds.) (2000). *Land, Law and Environment*. London: Pluto Press.
Berglund, E. (1998). *Knowing Nature, Knowing Science*. Cambridge, MA: White Horse Press.
Blaikie, P., Cannon, T., Davis, I. and Wisner, B. (1994). *At Risk*. London: Routledge.
Cronon, W. (1996). *Uncommon Ground*. New York: Norton.
Harvey, D. (1996). *Justice, Nature and the Geography of Difference*. Oxford: Blackwell.
Howitt, R. (2001). *Rethinking Resource Management*. London: Routledge.
Jones, M. (ed.) (2006). Special Issue: Essays on landscape, law and justice. *Norwegian Journal of Geography* 60, 1 127.
Lehtinen, A. (2006). *Postcolonialism, Multitude, and the Politics of Nature*. Lanham, MD: University Press of America.
Low, N. P. and Gleeson, B. J. (1997). *Justice, Society, and Nature*. London: Routledge.
Martinez Alier, J. (2002). *The Environmentalism of the Poor*. Cheltenham: Edward Elgar.
McCarthy, J. and Prudham, S. (eds.) (2004). Neoliberal nature and the nature of neoliberalism. *Geoforum* 35, 275 393.
Ossenbrugge, J. (1993). *Umweltrisiko und Raumentwicklung*. Berlin: Springer Verlag.
Sachs, W., Loske, R. and Linz, M. (1998). *Greening the North*. London: Zed Books.
Walker, G. and Bulkeley, H. (eds.) (2006). Geographies of environmental justice. *Geoforum* 37, 655 738.
Wolch, J. and Emel, J. (eds.) (1998). *Animal Geographies*. London: Verso.

Relevant Websites

http://www.ejrc.cau.edu/
Environmental Justice Resource Center at Clark Atlanta University.
http://www.justiceandenvironment.org/
Justice and Environment Network, Central and Eastern Europe.
http://www.cas.uio.no/Groups/lsclawjus/Landscap.htm
Landscape, Law and Justice Group at the Center for Advanced Studies, Oslo, Norway.
http://www.snowchange.org/
Snowchange Network/Climatic Change and Northern Indigenous Nations.
http://www.unl.edu/geography/IPKRC/
The Indigenous People's Knowledge's and Rights Commission of the International Geographical Union.

Environmental Policy

C. Cocklin, James Cook University, Townsville, QLD, Australia

Introduction

It is widely acknowledged that 'policy' is a term that evades precise and agreed definition. Indeed, there is the often cited metaphor that defining policy is like trying to describe an elephant – easy enough to recognize if you have seen one before, but difficult to describe to those who have not. Most definitions refer to policy, in one way or the other, as agenda setting combined with action.

One useful way of representing policy is by estab lishing its relation to other aspects of decision making and action, as in the following hierarchical arrangement:

- Policy – the setting of direction and broad intent, objectives (e.g., an objective to reduce greenhouse gas emissions);
- Program – a strategy and management framework for achieving policy (e.g., to achieve the policy by reducing carbon dioxide (CO_2) emissions from in dustry); and
- Plan – a specific element or aspect of a program (e.g., encourage emissions reductions by providing subsidies to implement energy efficiency in industry).

It is common also to define policy with reference to a decision making process in which it is imbedded. **Fig ure 1**, for example, illustrates a 'policy cycle' that shows a sequence originating with the identification of an issue or problem, agenda setting, and the enlistment of support, implementation, and monitoring and review. The clear implication is that policy is a systematic, ordered process by which issues are defined and attended to in a logical sequence.

This is what is sometimes referred to as the 'rational' or 'prescriptive' planning model. The risk in such rep resentations, though, is that they suggest that policy statements are systematically implemented in such a way as to give direct force to the stated policy imperatives. Alternative representations of policy point to the fact that there are uncertainties and ambiguities that typify planning and preclude the implementation of the so called rational model. For example, Lindblom's well known representation of policy as 'disjointed incre mentalism' highlights the fact that policy problems are not always clearly defined, that there may be conflict among goals and objectives, that alternatives are not fully

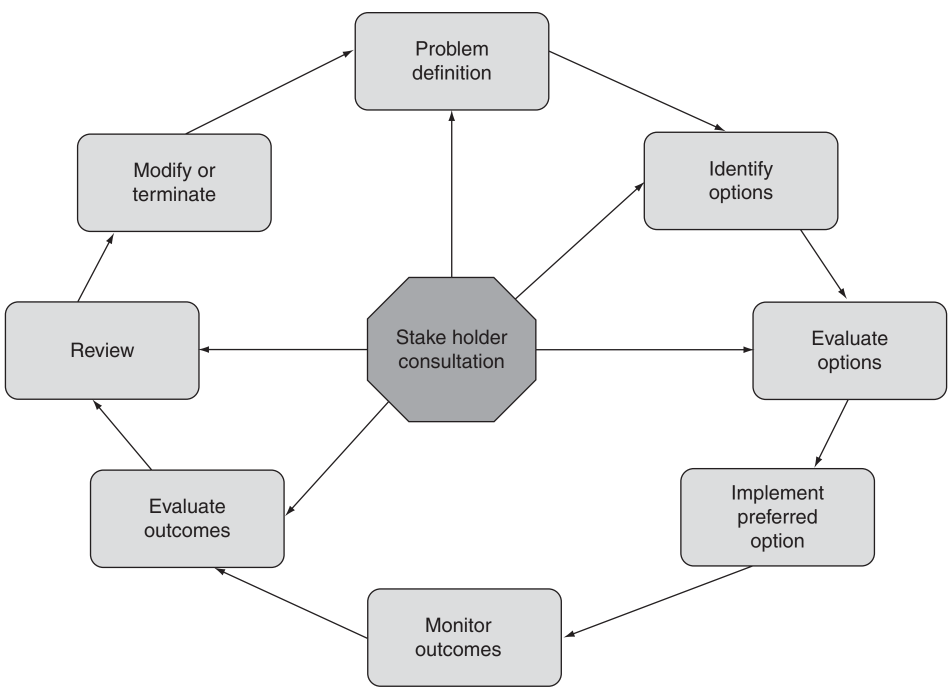

Figure 1 The policy cycle. Modified from Thomas, I. (2005). *Environmental Management: Processes and Practices for Australia*. Sydney: The Federation Press.

considered, and that rarely is there any one clear cut solution. It is also the fact that policy statements are in many instances instruments of 'politics' (in both public and private realms) as opposed to being genuine statements of intent.

Environmental Policy and Sustainability

In the time since the term 'sustainable development' achieved widespread exposure through the publication of the Brundtland Report – 'Our Common Future' – environmental policy almost everywhere has been aligned with this term. The Brundtland definition – sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs – is undoubtedly the most widely cited. However, the voluminous literature around sustainability and sustain able development has spawned numerous others. For example, the 'natural step', a framework for organizations to orient their activities in support of sustainable devel opment, establishes four 'system conditions' for sustain ability: do not mine materials faster than they replenish, do not produce substances faster than they can be broken down by natural processes, do not harvest renewables faster than they are replenished, and human needs must be met fairly and efficiently.

The interlocking circles representing the three arenas of environment, economy, and society – the 'fellowship of the rings' (**Figure 2**) – has also become a mainstay in defining sustainability/sustainable development. The notion of the 'triple bottom line' is an immediate de rivative of this early conceptualization of the terms.

These and the numerous other definitions that have been advocated have given rise to the comment that the terms sustainability and sustainable development have become effectively meaningless. However, while some people have lamented the lack of consensus on definition, to do so overlooks the fact that communities and societies will always contest the relative emphasis that is given to economic growth, social justice, and environmental protection. It should therefore be taken as given that sus tainability will be subject to mediation through social contest and debate. The objective then shifts from a fruitless search for agreed meaning to one of understanding the many and often competing perspectives on sustain ability and to understanding how these contests play out in social space – the role of actors, the multiple constructions of meaning, power relationships, the authority of com peting knowledges and discourses, and the implications of the many and varied pathways toward sustainability. Thus, we can acknowledge that:

- sustainability involves balance and compromise among social, economic, and environmental priorities, or what is now popularly referred to as the 'triple bottom line'.
- sustainability refers to 'progress toward' (or 'retro gression from') preferred futures, rather than to a defined endpoint. This means that sustainability is not an absolute concept, but a relative one.
- definitions and interpretations of sustainability are socially contested, because people assign different priorities to social, economic, and environmental assets and outcomes.
- some indicators may point to communities that are progressing toward or retrogressing from sustainability (e.g., rising/falling incomes, increasing/decreasing population, and increasing/decreasing employment), but these indicators are not consistently reliable markers of direction.
- sustainability implies equity, both within con temporary society (intragenerational) and in terms of the legacy for future generations (intergenerational).

While it is important to acknowledge the vast literature that has emerged around the definitional aspects of sustainability, the important point here is that 'sustain able development', as imprecise as it may be, has become 'the' point of convergence for environmental policy internationally. It would be difficult today to find state ments of environmental policy that do not enlist the term in one way or another.

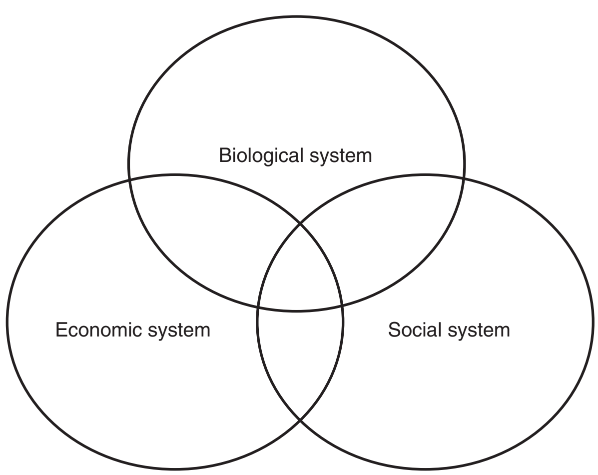

Figure 2 The three dimensions of sustainability.

The Internationalization of Environmental Policy

Environmental policy entered the main stage in the late 1960s and early 1970s, following the publication of a number of influential works that documented the extent and severity of environmental degradation around the world. Rachel Carson's 1962 book, *Silent Spring*, which documented the use and effects of agricultural chemicals is often held to have been the first of these significant contributions. Others of note include *The Population Bomb* by Paul Ehrlich, Barry Commoner's *The Closing Circle*, and Garret Hardin's famous essay 'The tragedy of the commons'. Amidst a precipitous rise in social concern about the environment, governments were pressured to respond through the development of environmental policy and its enactment through legislation and environmental management strategies. In this regard, the United States' Environmental Policy Act was a landmark piece of legislation, followed rapidly by essentially similar regulations throughout much of the Western world.

From these early days and through until about the late 1980s, the impetus came from concerns for environ mental degradation at national and regional scales and almost all of the environmental policy and legislation that came into force was, accordingly, national in its origins and scope. However, from about the 1980s, interest began to shift from environmental change that was nationally scaled to observed changes in environ mental quality that are transnational or have global reach. Acid deposition, depletion of the ozone layer, and human induced climate change are the archetypes in this regard.

The history of international agreements on the environment can be traced back to the 1940s, but it was post 1980 that the number of agreements, treaties, and conventions proliferated, leading to the international ization of environmental policy. International policy in any domain can be difficult since, as Timothy O'Riordan (1995) pointed out, it requires nations to surrender, in some measure, their national sovereignty – that is, "the freedom to act as a sovereign state must be circumscribed by an obligation to respect the legitimate interests of other sovereign states." In order for this to occur, he suggests that three conditions must be met:

Mutual advantage, which requires recognition by all states that cooperation and compliance is better than if one or more states does not observe the collective agreements;

Credible threat, suggesting that there is accepted science in regard to the environmental threat and that noncompliance is not in the interest of nations individually nor collectively; and

Credible enforcement, being recognition among nations that compliance will be enforced and non compliance will be penalized.

This concise structure helps us to comprehend why, at a simple level, it has been easier to get international agreement on some issues than others. In the case of ozone depletion, for example, the science was relatively unequivocal (credible threat), and it is reasonably clear that most nations were at risk to some degree from the further enlargement of the holes in the ozone layer (mutual advantage). In contrast, there was equivocation for a long time about the science of human induced climate change, and even now there is an acknowledg ment that some nations will benefit in some ways from global warming (e.g., some Northern Hemisphere nations, such as Canada and Russia). These two 'ambi guities' go a long way in explaining why it has been so difficult to achieve international agreement on limiting greenhouse gas emissions.

The issue of enforcement is a persistent concern in terms of international environmental policy and regu lation, and there is widespread acknowledgment that compliance is essentially voluntary – there is little in a material way that the international community can do to penalize noncompliance. Even so, international influ ences on policy relating to the environment have increased markedly over the last three decades; with so many of the major threats to environmental quality being sourced internationally and having global consequences, the environmental policy agenda will inevitably remain an important expression of globalization.

From Government to Governance

Environmental policy, conservation, and the management of natural resources were once the function of appointed officials in the public sector. More recently, environ mental policy and management has been constructed as a broader project, requiring the direct involvement of communities, individuals, nongovernment organizations, and the corporate sector. Thus, a tendency has emerged for responsibility for environmental protection and management in support of the public interest to be as signed more broadly.

Changes in the role of government as regards environmental policy have occurred in response to social, economic, and technological changes that societies have undergone in the past several decades. This includes a shift from 'government' defined strictly by the nation state to a more inclusive concept of 'governance' which recognizes the contributions of various levels of govern ment (global and transnational/regional/local) as well as the roles of the private sector, nongovernmental actors, and civil society.

The emergence of these new forms of governance has been attributed to the need for new institutions to ad dress the more complex problems of present day society. Ideology and economic globalization have also played a role in the shift from government to governance. Free market competition is now perceived by many actors as superior to government direction of the economy and commercial activities. This is partly the consequence of the fact that trade liberalization and privatization have diminished the ability of governments to directly control their economies. While some believe that globalization has made national governments less powerful, others argue that rather than simply eroding government power, globalization has changed the ways in which governments operate and influence situations. For example, the three key institutional sectors of society – government, market, and civil society – have begun to work in closer collaboration, partnering with each other in multiple and diverse ways when their goals are common. Of course, this is not to imply that they always, or even mostly work in partnership or have priorities that are aligned, but it does mean that they have so more often than they did, including in terms of environmental policy objectives. The nature of global governance on a range of issues, including the environment, is today best understood not only as what states do but also as a combination of that which the state, civil society, and markets do, or do not do.

The concept of sustainability has raised many questions about the traditional development model and the relationship between society, economy, and the environment. It has also raised questions about the suitability of the traditional government model to meet the demands of managing the environment and economy sustainably. Thus, responsibility for environmental management has been shifted upward to international bodies and to transnational companies, and downward to local governments, businesses, and resource users. In addition, individuals play a greater role in the environment as consumers, private owners of land, and participants in the policy discourse, as does civil society.

The more prominent role that businesses and civil society groups have played in governance has not been without controversy. Some believe that only the state can act in the public interest, while industry and citizens are motivated by self interest. Others see all actors as motivated by self interest and, in this context, believe that competition and the market ensure the best outcomes – public and private. According to this view, civil society, consumers, and industry bear greater responsibility and share the risks, while the state maintains a role in setting standards and auditing performance.

While the roles, responsibilities, and powers assigned to the respective actors remain a contested subject, it is widely acknowledged that the responsibility for the environment and sustainability has become a much broader project, no longer primarily the preserve of governments, but one involving civil society, the private sector, and the state. Thus, success in achieving environmental policy objectives and sustainable development depends on the institutional capability of people in all arenas – the state, the market, and the community.

Policy Cultures and Styles

The acknowledgment that the governance of the environment and natural resources is evolving has been accompanied by an increasing recognition of the differences among policy cultures and styles. One of the main lines of distinction that has been drawn is between the command and control approach to environmental policy and regulation and the 'cooperative' approach to policy design, which relies more on voluntary behaviors than mandatory legal instruments.

Further, it has been demonstrated that there are distinctive national 'policy styles' which, among other things, reveal different propensities to adopt coercive or cooperative policy designs. In seeking to explain the incidence of policy styles, Richardson *et al.* (1982) suggested that it arises from the convergence of: "(a) the government's approach to problem solving and (b) the relationship between government and other actors in the policy process." From these constructs, they derived a typology of styles that distinguish national approaches according to whether they are anticipatory or reactive and whether the approach is consensual or coercive.

Policy styles will manifest themselves in specific aspects of policy design and implementation. Rayner (1993), for example, suggested that variations in political culture exert an influence on the selection of policy instruments. Thus, countries will differ in their reliance on regulatory, voluntary, and market based instruments according to the prevailing policy style, which in turn may be influenced by the power exercised by different stakeholders (such as civil society, industry, and government) and political culture. Differences in policy cultures and styles may also manifest themselves in the nature of 'policy learning'. Dovers (2005) distinguishes four different forms of policy learning:

Instrumental learning, which refers to reflection on the performance of policy tools;
Government learning, being an evaluation of the performance of administrative arrangements and processes;
Social learning, which evaluates constructions of policies and goals; and
Political learning, an evaluation of how to influence political and policy processes.

These forms of learning are essential to the effective monitoring and evaluation of policy. Clearly, the policies to which these processes of evaluation are applied vary with policy styles and cultures, and they will variably reinforce or reorient policy design and implementation in accordance with those same styles and cultures.

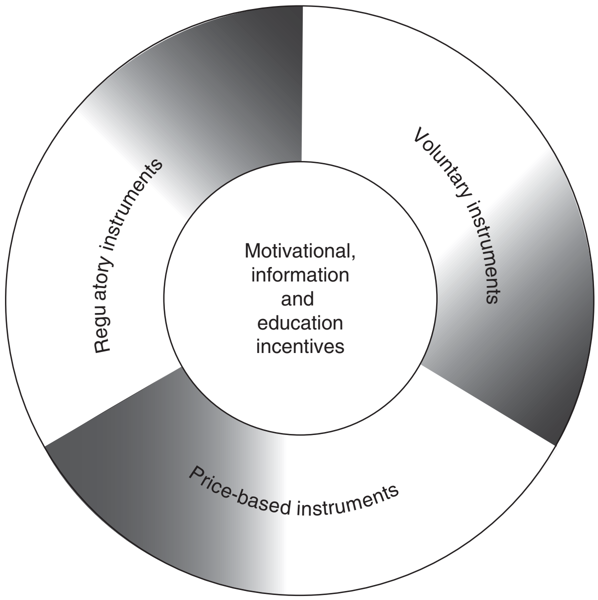

Figure 3 Policy instruments. Adapted from Young, M., Gunningham, N., Elix, J. *et al.* (1996). *Reimbursing the Future: An Evaluation of Motivational, Voluntary, Price-Based, Property-Right and Regulatory Incentives for the Conservation of Biodiversity.* Canberra: Department of Environment and Heritage.

Policy Instruments

Policymakers can call upon on an array of mechanisms in support of environmental policy (**Figure 3**). The main categories of policy instruments are regulation (com mand and control), market based (economic) instru ments (MBIs), voluntary approaches, and education and information. The approaches are not mutually exclusive: all rely to some extent on education and information provision, which is signified in **Figure 3** by the central placement of this category. For example, voluntary approaches include self regulation, while economic instruments might be underpinned by regulation.

In concert with the changing nature of environmental governance generally, the balance in the use of these instruments is changing from what was once a pre dominant reliance on regulatory methods to a much greater resort to market based and voluntary approaches. Accordingly, greater responsibility is being transferred to the individual, either directly through the market as an instrument of policy or through policy prescrip tions that individualize the responsibility for the environment.

In most nations, environmental policy has historically relied on the use of public regulation, which is often associated with command and control instruments, such as zoning, discharge standards, licensing, bans or limits on inputs and outputs, and requirements in terms of tech nology and design. Despite their widespread use, regu latory measures are considered by some people to be inflexible, intrusive, and inefficient; in many cases, they have failed to change behavior with respect to the environment. Accordingly, in recent years, there has been a trend in many Western nations toward self regulation, in which private interests (such as farmers, companies, and individuals), rather than governments, regulate to protect the public good. Examples of private interest regulatory mechanisms include industry codes of prac tice, environmental certification, eco labeling programs and agreements between interest groups and industry. There is a growing interest in quality control codes and certification systems, particularly in manufacturing, agriculture, and food processing industries. Although there is no legal enforcement, governments can influence the design, implementation, and impact of private regu lation in a variety of ways. Thus, governments are showing interest in a greater reliance on self regulation and voluntary standards, together with a statutory duty of care for the environment, on the basis that everyone who could influence the risk of environmental harm should be required to take all reasonable and practical steps to prevent any foreseeable harm from their actions. This is an important shift in the approach to policy implementation.

The progressive shift of responsibility to individuals, firms, and communities has been accompanied by an emphasis on the development of voluntary or partnership programs involving individuals and community groups undertaking local projects for environmental remedia tion. The strategies of partnerships, self help, and com munity empowerment have been used to encourage participation and to promote the idea that environmental problems are best addressed through communities working together and with government and industry. Advantages might include increased public participation in decision making, which is now sometimes associated with what is called 'deliberative democracy'. The positive rhetoric surrounding partnerships and civic environ mentalism is counterbalanced by a wide range of concerns, however. These refer to the implications of imbalances of power in partnerships, volunteer burnout in community based programs, an absence of strategic direction in terms of environmental outcomes, and a lack of adequate resources.

To overcome the limitations of both voluntary and regulatory approaches, there has been an increasing interest in, and acceptance of, the use of economic in struments in support of environmental policy objectives. These market like mechanisms aim to internalize nega tive environmental externalities, for example, by creating markets in carbon, biodiversity, and salinity. Economic instruments provide financial incentives/disincentives that, in many cases, are likely to engender better responses to environmental issues, while avoiding the complexities and prescriptive nature of legislation.

Although the enthusiasm for economic instruments is widespread, they have also been subject to critical assessment. There are equity issues and prospective operational problems arising from the definition of property rights and organizational capacity. A review of the use of economic instruments in the Organisation for Economic Co operation and Development (OECD) nations revealed 'no black and white picture of their general success'.

The success of policy instruments relies on matching them to the particular ecological, political, and economic situation and to the capabilities of institutions and stake holders. It is not necessarily a matter of developing new tools and instruments, but designing a 'mix' of policy instruments that is best suited to the circumstance. It has also been established that the attitudes and outlook of stakeholders (e.g., industry, individuals, and land man agers) are fundamental considerations in bringing about changes in environmental practices. To succeed, new policy approaches require the cooperation of these stake holders. Thus, the effective and efficient implementation of policy and management strategies by government relies, in part, on an understanding of what stakeholders regard as the key issues that influence those decisions that in turn impact on the environment, how they are responding to these issues, their disposition toward various policy and management tools, and their past experience in working with government agencies.

See also: Environment; Environmental Justice; Environmental Regulation; Environmental Security; Governance; Public Policy; Sustainability; Sustainable Development.

Further Reading

Barde, J. P. and Smith, S. (1997). Do economic instruments help the environment? *OECD Observer* 204, 22 26.
Dovers, S. (2005). *Environment and Sustainability Policy: Creation, Implementation, Evaluation*. Sydney: The Federation Press.
Lindblom, C. (1968). *The Policy Making Process*. Englewood Cliffs, NJ: Prentice Hall.
O'Riordan, T. (ed.) (1995). *Environmental Science for Environmental Management*. Harlow: Longman Group.
Rayner, S. (1993). Introduction to the special issue: National case studies of institutional capabilities to implement greenhouse gas reductions. *Global Environmental Change* 3(1), 7 11.
Richardson, J., Gustafsson, G. and Jordan, J. (1982). The concept of policy style. In Richardson, J. (ed.) *Policy Style in Western Europe*, pp 11 16. London: George Allen and Unwin.
Thomas, I. (2005). *Environmental Management: Processes and Practices for Australia*. Sydney: The Federation Press.
Young, M., Gunningham, N., Elix, J. *et al*. (1996). *Reimbursing the Future: An Evaluation of Motivational, Voluntary, Price Based, Property Right and Regulatory Incentives for the Conservation of Biodiversity*. Canberra: Department of Environment and Heritage.

Environmental Regulation

P. McManus, The University of Sydney, Sydney, NSW, Australia

Glossary

Beyond Compliance A term used to describe a situation where a corporation exceeds the regulatory standards, possibly for ethical reasons, to gain competitive advantage or to improve public relations.

California Effect A situation where a dominant trading partner can introduce more stringent regulations that other countries eventually follow so as to maintain trade links.

Command and Control Regulation The direct regulation of an industry or activity by legislation that states what is permitted and what is illegal.

Dual Standards Where a corporation applies higher standards in the country of its head office, or in countries with similar environmental regulations and enforcement mechanisms, and applies lesser standards to their operations beyond their home base.

Firm–Based Global Environmental Standards A uniform set of product and process requirements that must be adhered to by all of a firm's activities around the world.

Public-Sector Regulation This is the regulation of government departments and their activities.

Race to the Bottom A scenario where environmental regulation is reduced in order to attract or retain businesses in a particular location.

Regulatory Capture A situation where a government regulatory agency serves the interests of the industry it is supposed to regulate.

Regulatory Convergence A situation where the regulations of a dominant trading partner are adopted in other jurisdictions in order to maintain export markets.

Self-Regulation (Voluntary Regulation) Regulation of environmental, safety, or other issues by corporations, or an industry representative association, instead of government regulation.

Introduction

Environmental regulation refers to the imposition of limitations or responsibilities on individuals, corporations, and other entities for the purpose of preventing environmental damage or improving degraded environments. In this article, the term refers to both the use of legislation through parliaments and the regulations established by organizations mandated to protect the environment. Environmental regulation is one possible approach to environmental management and is often used in conjunction with other compatible approaches.

Regulation occurs in many spheres of activity. Environmental regulation has a long history, but it became increasingly important in the latter half of the twentieth century. The advantages and disadvantages of environmental regulation are common to many sectors where regulation has been adopted. These advantages and disadvantages have meant that while many people support environmental regulation, there are also groups lobbying to reduce the amount of environmental regulation and to change the character of environmental regulation.

Environmental regulation limits the unfettered growth of an individual business, but it provides the conditions for overall growth over a long period of time. This may be achieved by protecting the quantity of environmental resources (e.g., fish and trees) or the quality of the resources (e.g., by preventing pollution flowing into water bodies). More recently, it has been recognized that higher environmental standards may enable a country to develop beneficial domestic industries and achieve economic growth through export production, or provide the clean up technology that can be exported to other countries that have degraded their environment.

History

Early History of Environmental Regulation

Many of the early examples of environmental regulation are either undocumented or subject to interpretation about the exact nature of the regulation and its effectiveness. One of the earlier examples of environmental regulation that is often cited, usually inaccurately, is the banning of coal burning in the reign of Edward I of England.

Air pollution caused by the burning of cheap surface coal that had been imported to London by sea (hence the term sea coal) was considered a serious problem in London by the late thirteenth century. It appears that there were restrictions on burning coal from 1273. It is difficult to say whether King Edward I or his successors favored regulating the burning of coal, or whether it was pressure from citizens that led to a proclamation in 1306 that prohibited the use of sea coal. The extent of the ban – whether this was limited to when parliament was in session, whether it included selling and burning, and whether it applied to both tradespeople and householders trying to keep warm – appears to vary in the literature. The enforcement mechanism, that is, the burning of sea coal was made a capital offence, is also disputed, as is the record of one offender being executed. Supposedly, citizens continued to burn sea coal in violation of the law, because few could afford the cost of wood. Whether accurate in all details or not, this story highlights many of the issues that need to be considered when imposing environmental regulations.

Air pollution was a target for early environmental regulation because the perpetrators were relatively few, and there were many affected people. It is possible to discriminate against certain people through environmental regulation. For example, in the regulation of London coal burning in the early fourteenth century, the lime burners who were blamed for early air pollution were also often accused of unfairly raising prices. Many colliers and some lime burners were Northerners, and this may have made them a definable target.

The environmental regulation of air quality in the US arose because of health issues and concerns about the reputation of cities such as Pittsburgh, Cleveland, and Milwaukee. In 1869, an ordinance was passed in Pittsburgh that made it illegal to burn soft coal in locomotives within the city boundaries. This was ineffective. In Milwaukee, a technical fix was attempted, again without success. It was not until the 1890s that legal action started to become effective, and polluting industries often relocated beyond the city boundaries. By 1897, many large US cities had passed ordinances where the responsibility to control smoke emissions rested with officials from the municipal health department. Regulatory responsibility resided with people, who had limited understanding of the processes that generated the smoke, but it avoided early regulatory capture as the interests of the industrialists and health officials were not easily aligned. Apart from the removal of certain industries beyond the city borders, another important solution to the smoke problem was not regulation but substitution, whereby steam locomotives were replaced with electric rail systems in most of the larger US cities in the early twentieth century. Problems of smoke in cities persisted well into the twentieth century and have partly been addressed by the decline of some industries and their replacement by imports from countries such as China.

Other early forms of environmental regulation were concerned with the separation of incompatible uses to avoid odors, noise, and loss of amenity to people. This often had a negative effect, however, as undesirable facilities were isolated and could cause environmental damage without being observed. An example of moving facilities away from people was the regulation of slaughterhouses and abattoirs. Formerly, the slaughterhouses operated in residential areas, and blood was washed away down the streets. Beginning in Paris under the reign of Napoleon Bonaparte in 1806, the slaughterhouses were reorganized as abattoirs outside of the city. The expansion of Paris necessitated further reorganization when new abattoirs were built at La Villette in the late 1860s and Vaugirard in the late 1890s. Other cities around the world adopted this example. In the case of Sydney (Australia), in 1849 the Slaughter Houses Act forbade new slaughtering operations within, and within 2 miles (3 km) of, the City of Sydney's boundaries.

In essence, early environmental regulation addressed visible environmental problems, particularly where there were few sources of the problem, many people impacted, health was an issue, and there were opportunities to reduce the problem by spatially separating the source from the receiving population. It is often referred to as command and control regulation.

More Recent Environmental Regulation

Environmental regulation occurred in the early part of the twentieth century, but it was often enforced at the city or state/provincial level. This led to discrepancies between jurisdictions, which could be exploited. The overall quality of the environment was not being addressed because of a lack of consistency in regulations. It also meant that companies operating in more than one jurisdiction had to follow a myriad of regulations. One example of different environmental regulations, and the use of environmental regulations in conjunction with other approaches, was how the City of St. Louis, in 1939, regulated against the use of high volatile coal. In the winter of 1940–41, the city secured low volatile coal from Arkansas and made arrangements with a railroad for its transport, so it could be available for domestic use at low cost. In contrast, the city of East St. Louis across the Missouri River continued to burn high volatile coal and suffered serious smoke problems.

There was a significant growth in environmental regulation in the 1970s, following the increase in environmental awareness in the 1960s. In the USA, the rise of the Environmental Protection Agency (EPA) saw a proliferation of new environmental regulations. The implementation of various measures (e.g., pollution control, waste management, and water issues) meant that the states were responsible for planning and implementing these measures, albeit with EPA funding (which was a means of controlling the operations of potentially recalcitrant states). This approach has been criticized by many commentators, ranging from authors writing from a libertarian perspective who critiqued the rise of big government through to sympathetic critiques from authors who participated in this legislative expansion, but believed that it added unnecessary costs to corporations, was unpredictable in its impacts, lacked effective coordination and sometimes simply transferred the environmental issues to other categories (a shell game) rather than effectively addressing environmental concerns.

The 1980s were a time of reducing the role of the national government in the USA. President Ronald Reagan's administrations (1981–89) attempted to reduce the role of the federal government in environmental matters through deregulation, defunding federal agencies and introducing tax cuts so that there would not be money available for future expansion and devolution of many functions to state and local levels. Hence, de regulation was achieved not just by changing or deleting the regulations but also by reducing the budget and staff available to implement and enforce the regulations. This example was followed in a number of other countries.

Despite the ideologies and rhetoric of various political leaders, almost without exception, there has been an increase in regulatory activity, regulatory personnel, and regulatory budgets. The rare exception (see **Figure 1**) was the decline in real administrative costs of federal regulation in the USA during the early to mid 1980s. In the figure, the social category includes consumer safety and health, homeland security, transportation, workplace, environment, and energy. The formation of the EPA in the USA in 1970 partly contributed to the rise in social expenditures during the 1970s. The importance of en vironmental protection through regulation increased during the presidency of Bill Clinton (1989–97), but the biggest increase in social expenditure in recent years has been the regulation related to homeland security, par ticularly following the airplane attacks on prominent sites in the US on 11 September 2001.

The level at which environmental regulation occurs has also changed, with greater emphasis on international, supraregional (as in the European Union), and national regulation, in contrast to city or state based regulation. In some cases, the expansion in environmental regulation has also resulted from changes in economic structure, where previously state owned enterprises have been privatized and sometimes split into competing corpor ations. The privatization process has highlighted the need for regulation given that the government can no longer control the operations of the entity through ownership mechanisms.

Challenges for Environmental Regulation

There are numerous challenges to ensuring effective environmental regulation. These challenges range from ideological opposition to regulators interfering in market economies through to challenges derived from the characteristics of the environment and activities to be regulated. The ideological position is particularly chal lenging, because it influences the level of support for regulation among the public, and it may inspire certain governments (generally Republican in the US and con servative parties in other countries) to not regulate activities or to deregulate existing regulated activities. This challenge is refuted by environmental protection organizations who argue that effective environmental regulation is important in supporting a clean, competitive economy that creates jobs, reduces business risk, and generates markets for environmental goods and services.

There are also challenges within the field of en vironmental regulation. A number of these challenges are

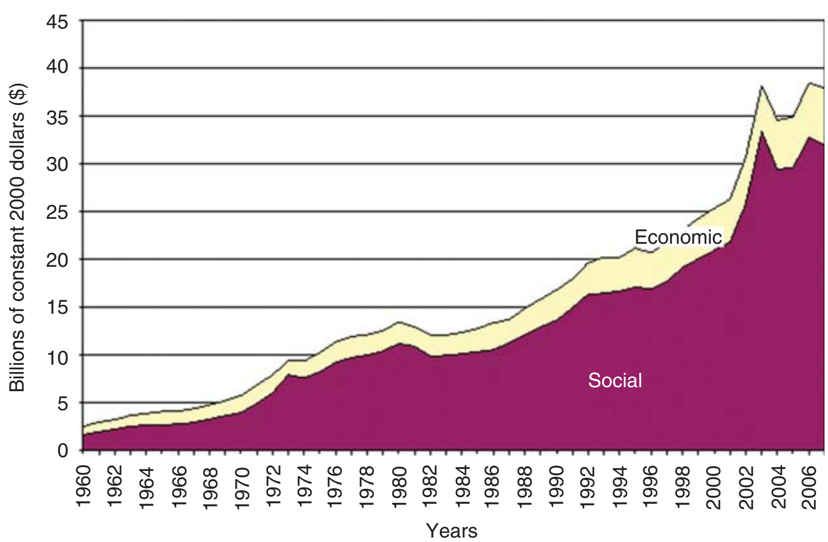

Figure 1 Administrative costs of federal regulation, inflation adjusted using $2000. Reprinted from Dudley, S. E. and Warren, M. (2005). The regulators' growing budget. *Regulation* 28, 6 7.

discussed in the subsections below. These challenges include the global consequences of particular issues (e.g., activities contributing to climate change), activities that may severely impact the health of people (especially if there is a lag time between the cause and effect), activities that involve widely dispersed but very small sources of pollution, the issue of regulatory capture, the challenge of transnational corporations spanning regulatory jurisdictions, the increasing need to regulate public sector activities as the nature of government departments changes, the challenge to the legitimacy of environmental regulation resulting from clauses in international trade agreements, the rise of indigenous issues, and the legitimacy and jurisdiction of environmental regulation by governments (particularly in countries such as Australia and Canada), and the characteristics of particular environments and activities to be regulated.

The issue of regulating particular environments is complex. The effective regulation of fishing requires an understanding of fish breeding, habitats, numbers, and so on, in addition to an awareness of fishing activity and the capacity and likelihood to overfish. Given that the fish are underwater and do not stay still, there is a potential to overestimate fish numbers. Once regulations have been set, there is also the need to enforce these regulations because the nature of the industry (including unpatrolled oceans) means that there is greater potential for illegal fishing, which has cumulative impacts on fish numbers.

The environmental regulation of products, such as steel production, may be challenging for other reasons. If two pieces of steel appear similar, how does anybody know if one piece has been produced with significantly more environmental degradation than the other? The production process needs to be regulated to ensure that environmental issues, as well as working conditions and payments, are addressed satisfactorily.

The range of environmental issues to be regulated (including water, biodiversity, climate, pollution, noise, etc.), the forms and levels of regulation, and the divergent interests of various corporations, countries, and international organizations mean that environmental regulation is a contested and controversial activity.

Compliance and Beyond Compliance

Compliance is a challenge, but there are many reasons why firms comply with environmental regulation, or in some cases go beyond what are the minimum environmental regulations. Corporate executives are likely to weigh up the costs and benefits of compliance and noncompliance in relation to profits, corporate image, and future regulation activity. There is also the risk of detection, the severity of corporate and personal penalties (including fines and imprisonment), and the benefits to the corporate image of being perceived as a good corporate citizen. There have been many debates about the impacts of environmental regulation on corporate behavior, including whether firms seek out locations of minimal environmental regulation. Recent evidence from the US suggests that rather than moving to pollution havens, many pollution intensive industries also have high demands for human and physical capital, and needs such as a highly skilled workforce are best met in developed countries which often have greater political stability and more stringent environmental regulations.

The literature on corporate adoption of beyond compliance approaches highlights the advantage of being a first mover and possibly setting the parameters for the regulation of other players in the industry. One outcome of the move toward incentives as an approach to environmental protection is that corporations have greater choice than simply whether they should follow regulations or risk detection and penalties. This greater choice has encouraged research into the internal dynamics of corporations, with one finding being that noneconomic factors are important in influencing whether a firm adopts a particular environmental course of action. The corporation, in this analysis, is not the decision making entity but the internal dynamics between executives who hold different values is important in the decision to move beyond compliance. This approach is important because if it is effective, it can be used in conjunction with environmental regulation to encourage better environmental outcomes in a way that is more efficient in terms of the economic costs of implementing regulation.

Of course, it is easier to comply with environmental regulations if those regulations are not very stringent. This may be the case where regulatory capture has occurred. Corporations could be complying with environmental regulations, but the environment is still being destroyed.

Regulatory Capture

One of the potential problems of all regulation, including environmental regulation, is the phenomenon of regulatory capture. This occurs when a government regulatory agency serves the interests of the industry it is supposed to regulate. It is believed that many organizations will try to reduce the effectiveness of their regulators by gaining political control of the regulatory body, or by obtaining more information and employing more effective staff to counter the regulatory body. Over time, there is often a two way transfer of staff between the regulator and regulated industries, thereby bringing them closer to each other. The historian, Gabriel Kolko, claimed that regulatory capture was a normal part of the relationship between business and regulatory bodies. It happens because the businesses that are regulated have a high stake in the outcome of the regulatory practice and, therefore, devote attention and resources to ensuring the outcomes are good for their business. This often means developing relationships between the regulators and those who are being regulated.

Eventually, it may happen that the businesses being regulated actually desire this regulation, because it can reduce the likelihood of new entrants entering the industry. This can be seen in some Australian states with regulation of the sex industry, where police may enforce a policy of containment, which allows existing premises to continue but restricts the establishment of new premises. Similar examples occur in environmental regulation, where there is no intention of destroying an industry, even an unpopular one, because it still contributes to economic growth, employment, and makes products that people use. The regulation of the industry is, however, problematic because regulation is needed, but if regulatory capture has occurred then it is very difficult for the regulation to be effective.

Transnational Corporations (TNCs)

One of the most challenging issues for governance is the operations of TNCs whose operations span many national regulatory jurisdictions. It is often argued that TNCs have grown beyond the control of national governments and that national governments should deregulate to remain competitive with other countries. This approach is also seen as a race to the bottom in terms of environmental, worker safety, and other standards. The growing importance of the environmental movement in the 1980s and 1990s, at a time of the new right political ascendancy in the United Kingdom, the US, and other countries, meant that environmental issues were often regulated despite the rhetoric of deregulation espoused by some governments. This regulation extends to the operations of TNCs, particularly in relation to poorer countries. For example, in 1994, the dumping, in poor countries, of hazardous waste produced in wealthier countries was banned by the Basel Convention, largely out of concern that toxic traders were acting as brokers for the hazardous waste. It was also easy for TNCs with operations in many countries to ship waste to the site of poorest environmental regulation and enforcement, especially if accepting the waste was perceived as a way of generating income for the poorer countries.

TNCs have particular characteristics that limit their ability to be regulated. They possess greater flexibility, mobility, and leverage that local companies do not enjoy. By way of contrast, their activities also have a higher profile, and they are often scrutinized more than local companies. TNCs also have the financial, technological, and managerial resources to devote to environmental issues and the ability to absorb the costs of regulation within their operations. They are also more likely to be aware of environmental management developments abroad.

The concern of many environmentalists and corporate watchdogs is that, despite these advantages, some TNCs operate dual or varying standards, generally applying higher standards in the country of their head office or in countries with similar environmental regulations and enforcement mechanisms, and applying lesser standards to their operations beyond their home base. The extent to which this occurs is debated, especially given the importance of other factors such as political stability in some countries with lesser environmental standards. It is, however, a particular concern for environmental regulation when the resource in question is located in poorer countries with limited environmental regulations and little means of enforcing them. This situation occurs in African countries in relation to oil exploitation, with the example of the operations of the Royal Dutch/Shell Group and the impacts on the Ogoni people of Southeast Nigeria being well documented. It also occurs in forestry, where tropical rain forests, which are mostly located in developing countries, have been cleared by TNCs. Forestry is an example of the spread of TNC activity, where TNCs from countries such as Malaysia have been instrumental in the destruction of tropical rainforests in countries such as Papua New Guinea. The resultant impacts include loss of biodiversity, erosion on steep slopes when heavy rainfall occurs, and negative social impacts on indigenous people.

The tragic case in Bhopal, India, in 1984, is an example of the consequences of poor management of dangerous operations in a developing country. The US headquartered company Union Carbide operated a pesticide plant that leaked deadly gas which killed thousands of people and caused illness for many thousands more local residents. The public relations disaster that may affect a corporation involved in such damaging environmental activity, which would therefore affect its share price, is partly offset by the ability of TNCs to engage issue management corporations to reduce the impacts of the problem on the TNC.

The debates about varying standards of operation by TNCs in respect of environmental regulation involve two main arguments. On the one hand, commentators argue that this is an approach by TNCs to increase their profits. On the other, commentators argue that TNCs are increasingly applying best practice in developing countries, because it is to their advantage to standardize their operations across the corporation. This means that they are operating beyond compliance in developing countries. The issue of systematic attempts to evade higher levels of environmental regulation is an issue that is still debated.

Recent studies argue that increasingly stringent end market environmental regulation, plus concern to protect a firm's reputation and operating legitimacy, is leading to firm based global environmental standards that exceed the environmental regulations of many poorer countries. This means that TNCs are applying firm based standards uniformly to all plants worldwide, rather than attempting to meet the local regulatory requirements where a plant is located. This is seen as a way of promoting improved performance within a corporation at a time when environmental regulations are proliferating at local, national, and international levels. Despite the existence of firm based global standards, environmental improve ment is still primarily driven by national and inter national environmental regulations.

International Trade Agreements

As discussed in relation to TNCs, trade can potentially lead to an improvement in environmental regulations and, if they are appropriate and administered effectively, an improvement in the environment. Trade is also seen as promoting regulatory improvement if an exporter wants to market products where there are many potential customers and relatively strict environmental regulations. The exporter must meet the requirements of this market, and, in doing so, is likely to have their home country adopt similar standards so the exporter can have uniform production requirements (thus achieving economies of scale) and not be undercut by domestic competitors complying with lower environmental standards. This has been termed the California Effect, where a dominant trading partner can introduce regulations that other countries eventually follow so as to maintain trade links. It is a case of regulatory convergence, rather than a divergent regulatory scenario which would characterize 'a race to the bottom' as countries compete with each other to lower environmental regulations in order to attract and retain businesses.

The California Effect is only possible if trade agree ments do not restrict the right of states to introduce environmental regulations. The environment may be protected through international trade agreements, or these agreements may be used to prevent environmental regulations being enacted in particular countries, where it is deemed that an environmental regulation is being used to restrict trade.

These trade agreements include global agreements, such as those made through the World Trade Organ ization, and regional agreements between a limited number of countries, such as the North American Free Trade Agreement (NAFTA) between the US, Canada, and Mexico. While these agreements may contain environmental provisions, and in the case of NAFTA an environmental side accord and a Commission for Environmental Cooperation based in Montreal, there is concern about how to apply environmental standards within the confines of these agreements. One such dis pute between Mexico and the US was about protecting dolphins when fishing for tuna.

Voluntary (Self-) Regulation by Corporations and Industries

Considerable attention has been given to the notion of self regulation by corporations of environmental, health, and safety matters. Advocating the use of self regulation is often a move by corporations that anticipate the introduction of government regulation, or the strength ening of existing regulation. The desire to self regulate is a tactic to either avoid this form of regulation, or possibly to be prepared for it by phasing it in through a process of self regulation. The advocacy of self regulation may also be based on a belief that TNCs are beyond regulatory control and, therefore, this would be a waste of resources, but it may be possible for the corporations themselves to develop and enforce their own rules to meet public ex pectations. There is a danger in environmental depart ments relying on data provided by corporations, as was demonstrated in Britain by the US owned chemical company Sevalco discharging cyanide into the Severn Estuary between 1998 and 2002. The company was later fined for the discharges and for falsifying records to conceal the offences.

While some commentators favor self regulation and incentive based approaches to environmental manage ment, other commentators argue that there is a need for an effective regulator, backed by sufficient legislative power and financial resources, to ensure environmental protection.

Conclusion

Environmental regulation is necessary. It can prevent the worst excesses, and the existence of regulation and enforcement is itself an important deterrent. Environ mental regulation has resulted in environmental im provements, but it sometimes does impose what could be perceived as excessive costs on businesses. The costs and benefits of regulation relative to other approaches should always be considered when contemplating how best to manage a particular environmental issue. Regulation is one of a number of tools to improve environmental quality, and it should be used effectively with other approaches such as incentives. While the history of environmental regulation shows that it has contributed to improved environmental outcomes, scientific evidence indicates that numerous environmental problems are worse now, and new issues emerge through technological developments in areas such as biotechnology, nuclear energy, and wind power that require some form of regulation. To leave each of these sectors unfettered may encourage competition and profits within the sector, but at great expense to the rest of society and the environ ment. The economic costs of effective environmental regulation are significantly lower than the environmental costs of irreversible environmental damage to the life systems of the planet.

See also: Environment, Historical Geography of; Globalization and Transnational Corporations; Governance, Corporate; Industrial Location; Industry, Historical Geographies of; Location Theory; Resource Industries; Transnational Corporations in Developing Countries; Waste Management.

Further Reading

Anderson, T. L. (ed.) (1997). *Breaking the Environmental Policy Gridlock*. Stanford, CA: Hoover Institution Press, Stanford University.

Andrews, R. (1999). *Managing the Environment, Managing Ourselves: A History of American Environmental Policy*. New Haven, CT: Yale.

Angel, D. P. and Rock, M. T. (2005). Global standards and the environmental performance of industry. *Environment and Planning A* 37, 1903 1918.

Brimblecombe, P. (1987). *The Big Smoke: A History of Air Pollution in London since Medieval Times*. London: Methuen.

Cochrane, K. and Doulman, D. (2005). The rising tide of fisheries instruments and the struggle to keep afloat. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360, 77 94.

Cole, M., Elliott, R. J. R. and Shimamoto, K. (2005). Why the grass is not always greener: The competing effects of environmental regulations and factor intensities on US specialization. *Ecological Economics* 54, 95 109.

Dudley, S. E. and Warren, M. (2005). The regulators' growing budget. *Regulation* 28, 6 7.

ENDS (2004). Sevalco: Driving a coach and horses through self regulation. *ENDS Report* 359, 23 26.

Fowler, R. J. (1995). International environmental standards for transnational corporations. *Environmental Law* 25, 1 30.

Hood, C., James, O. and Scott, C. (2000). Regulation of government: Has it increased, is it increasing, should it be diminished? *Public Administration* 78, 283 304.

MacGill, I., Outhred, H. and Nolles, K. (2006). Some design lessons from market based greenhouse gas regulation in the restructured Australian electricity industry. *Energy Policy* 34, 11 25.

Network of Heads of European Environment Protection Agencies (2005). *The Contribution of Good Environmental Regulation to Competitiveness*. Copenhagen: European Environment Agency. http://org.eea.eu.int/documents/prague statement/prague statement en.pdf (accessed Aug. 2008).

Prakash, A. (2001). Why do firms adopt 'beyond compliance' environmental policies. *Business Strategy and the Environment* 10, 286 299.

Vogel, D. and Kagan, R. (eds.) (2004). *Dynamics of Regulatory Change: How Globalization Affects National Regulatory Policies*. Berkeley, CA: University of California Press.

Relevant Websites

http://www.citra.duq.edu
Center for International Regulatory Assistance: European Environmental Regulations.

http://www.epa.gov
United States Environmental Protection Agency.

Environmental Security

J. Barnett, University of Melbourne, Melbourne, VIC, Australia

Glossary

Environment The biological, physical, and chemical components and systems necessary for the process of life.

Environmental Change The short- and long-term changes in biological, physical, and chemical components and systems resulting from both human activities and natural processes.

Environmental Security The assurance that individuals and groups have that they can avoid or adapt to environmental change without critical adverse effects.

Insecurity The risk of loss or damage to a thing that is important to survival and well-being.

Security Freedom from the risk of loss or damage to a thing that is important to survival and well-being.

The Environment and Security

The concept of environmental security has different meanings and is contested. This arises from the merging of two powerful yet equally ambiguous concepts – environment and security – and the diverse array of disciplines and schools of thought that contribute to the study of these two concepts. The way a person or a group understands each of these concepts informs their understanding of the combined concept of environmental security.

In general, the environment refers to the biological, physical, and chemical components and systems necessary to sustain life. Yet, this is a broad agenda within which there are multiple issues, such as resource scarcity (diminishing supplies of inputs into human systems) and pollution (the contamination of inputs into human systems), occurring at multiple scales (from global to local) and in different ways and to different degrees in places around the world.

Environmental problems are now prominent political issues as a consequence of the increased scale of consumption and pollution in modern high energy societies, which has caused large decreases in primary forest cover; biodiversity losses; depletion of fish stocks; land degradation; water pollution and scarcity; coastal and marine degradation; the contamination of people, plants, and animals by chemicals and radioactive substances; and climate change and sea level rise.

Security, too, is a nebulous concept. It can apply to many different things that are valued (referent objects such as jobs, health, the state, and territory) and refer to many different kinds of risks (such as unemployment, hunger, change of government, and invasion). Yet, the most powerful discourse about security concerns national security and the risks to it that originate from 'Other' living both beyond and within the territorially bounded nation state. As geographers and critical international relations scholars have shown, national security discourse is a discourse that manufactures and sustains particular kinds of identity through constructions of difference. Its primary function is to secure the nation state by justifying disciplining institutions.

However, the end of the Cold War coupled with increased interdependence between societies and countries created space for new ways of thinking about security as they rendered problematic the simple imaginaries of world space that national security institutions perpetuated. These changes have opened space to consider myriad local, national, and global interactions that create security and insecurity and the way some people's security occurs at the expense of others. The concept of environmental security has been a central issue in this process of deepening the scale and broadening the content of security.

The end of the Cold War triggered a proliferation of articles and new research projects on environmental security. Yet, there was another broad change in world affairs that was also important in the development of environmental security as a research and policy issue. The growth of environmental consciousness in developed countries was necessary for environmental problems to be seen as serious enough to warrant calling them a security issue. This began in the 1960s with some high profile studies of environmental problems and growth in the number of environmental nongovernmental organizations, including the creation of large, international, environmental nongovernmental organizations, such as the World Wildlife Fund, Friends of the Earth, and Greenpeace. The 1970s saw the beginning of international summits on environmental issues and an associated proliferation of international agreements on the environment. The first major global environmental summit was the 1972 United Nations Conference on the Human Environment. Prefiguring the contemporary idea of environmental security, in the 1970s and 1980s, a number of peace and environmental scholars began to highlight the inability of national security institutions, and in particular the military, to manage common environmental problems that pose threats to international stability and national well being. Then, in 1987, the World Commission on Environment and Development released its landmark report titled *Our Common Future*, which popularized the concept of sustainable development, and introduced the term 'environmental security'. This report in turn lead to the 1992 United Nations Conference on Environment and Development, which has had follow up conferences in 1997 and 2002 and which gave rise to major multilateral environmental treaties on climate change and biodiversity.

During the Cold War, the discipline of international relations dominated the study of security. This influence grew in part because of the retreat of geographers from the study of world politics due to the perceived role of geopolitical theories in justifying the world wars. However, as security has become more pluralized – away from states and war and toward people and the multitudinous risks they must manage – it increasingly becomes a general concept of social science, and geographers and in particular critical geopolitics have increasingly been engaged with this rethinking of security. The ways in which old and many new interpretations of security imagine the world as a series of bounded spaces that contain 'Others', and the problematic practices that flow from this simple geopolitical vision have been powerfully critiqued by geographers such as John Agnew, Simon Dalby, Klaus Dodds, and Gearóid Ó. Tuathail. These insights carry over into the study of environmental security, particularly through the influential work of Simon Dalby. Geographers have also long been engaged with the study of environmental problems, a tradition that can be dated to early founders of the discipline such as Alexander von Humboldt and Peter Kropotkin. Research from human–environment geographers, such as Steve Lonergan, Michael Redclift, and Michael Watts, has also informed the study of environmental security. In these and other ways, geographers have been and continue to engage with research on environmental security.

The plurality of meanings of environmental security can be categorized into six principal interpretations. First, environmental security can be seen as being about the impacts of human activities on the environment. Second, environmental security can be seen to be about the impacts of the military–industrial complex, including war, on the environment. Third, environmental change can be seen as a security problem common to all states, therefore requiring collective action. Fourth, environmental change can be seen as a threat to national security. Fifth, environmental change has been identified as a possible cause of violent conflict. Sixth, environmental change can be seen as a risk to human security. The following discussion focuses only on the last three of these interpretations, as they have been the most prominent in terms of research and their influence on policy.

Environmental Change and National Security

The majority of interpretations of environmental security work from existing theories of international relations – which by and large take the preeminence of national security for granted – and then factor in environmental issues. This was particularly the case in the early writings on environmental security, and it is one of the main reasons why there has been a lot of attention given to links between environmental change and violent conflict (since violent conflict is the preeminent concern of international relations). However, irrespective of whether environmental change is a factor in violent conflicts, it can in many less subtle ways undermine national security.

In the first instance, environmental change can weaken the economic base that determines military capacity. In many countries, natural resources and environmental services are important to economic growth, and income from primary sectors like agriculture, forestry, fishing, and mining, and from environmentally dependent services like tourism, can be lessened by environmental change. For example, China's rapid economic growth – which has funded its military modernization program – is increasingly being seen to be ecologically unsustainable as water shortages, water pollution, air pollution, and land degradation all appear to be increasing.

Most environmental problems that countries face are caused by endogenous activities. However, some problems are largely beyond their control. Examples of these include global emissions of ozone depleting substances increasing incidences of skin cancer in countries in the Southern latitudes, the long lived environmental (and social) impacts of nuclear weapons tests conducted during colonial times in French Polynesia and the Marshall Islands, the impacts of the Chernobyl nuclear reactor accident on countries in Eastern Europe, the impacts of forest fires in Indonesia on air pollution and health in Malaysia and Singapore, and the impacts of global emissions of greenhouse gases on low lying countries and countries with high climatic variability. It seems appropriate to call such transboundary flows national security problems in as much as their impact can be equal to, or greater than the magnitude of impacts associated with war (measured in terms of, say, the effective loss of territory, the loss of life, increasing morbidity, or economic contraction).

Nevertheless, there are problems associated with idea that some environmental problems are national security issues. The military threats that traditionally concern national security theory and policy are different from environmental threats in that they are deliberately imposed, and the cause of the threat is easily identifiable. In contrast, environmental risks are typically externalities of economic activity, and their causes are often uncertain. Linking environmental issues to national security also may not galvanize the kind of attention and action on environmental problems that traditional military threats do. Indeed, it may serve to strengthen existing security logic and institutions through, for example, strategies to engage the military in environmental activities, and preparations to fight the so called environmentally induced conflicts. As Simon Dalby has shown, environmental issues can be appropriated by dominant security discourse in this way, because its uncritical categorization of space as national territory is deeply rooted, robust, and adaptable, and because this discourse serves state power. Nevertheless, in as much as some environmental problems cross boundaries and scales of analysis, and stimulate political action above and below the state, such as occurs on the issue of climate change, then environmental change also challenges the power of national security discourse and its institutions.

Environmental Change and Violent Conflict

One of the major ways that environmental change may undermine national security is if it increases the risk of violent conflicts that can have dramatic local impacts, repercussions on trade and may demand international intervention on humanitarian grounds. While such conflicts may not be the most serious outcome of environmental change on most countries, because of the influence of national security institutions on the security research agenda, there has been a lot of research on the cause–effect relationships between environmental change and violent conflict. Yet, for as long as this has been a focus of research, critics have argued that the connections between environmental degradation and violence are ambiguous because of the methodological and theoretical difficulties in distinguishing environmental factors from the many other causes of violent conflicts. These difficulties are somewhat irreducible given the nature of violent conflict. Nevertheless, emerging from this research are some compelling arguments that environmental change is one of the many factors – albeit rarely the most important – in the generation of violent conflicts within, if rarely between, countries.

Perhaps the most influential work on the connections between environmental change and violence is by political scientist Thomas Homer Dixon and his colleagues at the University of Toronto. Homer Dixon examined cases where there was violent conflict and then sought to determine the influence of environmental factors in their generation. This method has its limitations, in particular, it ignores the many instances where environmental problems do not result in violent conflict, and its search for environmental factors leads to a downplaying of other many, varied, and arguably more important causal factors in the conflicts in question. There is also a tendency in Homer Dixon's work to ignore previous studies of the particular conflict and the very large body of work on environment, resources, and social responses conducted by geographers and sociologists. The findings of this research include that: scarcity of renewable resources can at times lead to violent conflict and social instability; unequal consumption of scarce resources is a critical factor in violent conflicts; violent conflicts where environmental scarcity is a factor are more likely in low income, resource dependent societies; population pressure is a contributing factor to violent conflicts; and when mechanisms that enable adaptation to environmental scarcity fail violent conflict is a more likely outcome.

An alternative approach to the connections between environment and violence entails quantitative analyses of aggregated data to test the relationships between various environmental and social variables, including resource scarcity, population growth, resource abundance, democracy, and poverty. A common finding of many of these studies is that 'strong states' with a high trade/GDP ratio tend to be less prone to internal conflicts and are more capable of managing environmental degradation and change. In contrast, states undergoing significant economic and political transitions are relatively more prone to internal violent conflict. Most of this quantitative research finds that relative and absolute poverty are important variables. Many studies find that it is the abundance of natural resources as much their scarcity that drives conflict. However, none of the findings of these quantitative studies are particularly conclusive, as they are significantly constrained by the quality of data they use. More importantly, while they show weak correlations among variables, they are unable to describe how the associations that they find occur.

In contrast to the effort to identify environmental factors in violent conflicts, there has been some research that seeks to learn from and explore peaceful responses to environmental change. For the most part, this research concerns cooperation between states over shared resources, in particular transboundary rivers and seas. Geographers such as Tony Allan, Steve Lonergan, and Aaron Wolf have been very influential in demonstrating that the connections between water and conflict are not as straightforward as commonly assumed, and that transboundary waters may be more likely to cement peace between countries than stimulate conflict. This explicit focus on peace has yet to focus on smaller scales of cooperation, and the endeavor as a whole has yet to reach a stage of critical mass whereby clear findings can emerge.

Geographers have played an important role in qualifying and contextualizing the connections between environmental change and violent conflict. Research firmly grounded in long standing environment–society studies conducted by geographers, as well as anthropologists, and sociologists, has offered more detailed insights into environmental problems and violence. These studies have highlighted the importance of unequal outcomes of social and environmental changes, including the inadequate distribution of the returns from resource extraction activities, declining agricultural production, and unequal access to economic and political resources. Many of these studies show that communities affected by externally imposed resource extraction activities rarely comply with these developments without some exercise of resistance. They suggest that the success or failure of nonviolent resistance strategies is critical to understanding the generation of violence (or peace). In these more qualitative studies, a range of economic, political, and cultural processes that structure both material and institutional forms of power are seen to be more important that scarcity *per se.* That these studies have more explanatory power suggests that ways forward in this field lie in research that grounds the effects of environmental changes in the local contexts in which agents sustain their lives and livelihoods, and that much can be learned from researching processes as they happen, and regardless of whether violent conflict is an outcome.

Environmental Change and Human Security

Global environmental change impacts most often and most severely on people that have less material wealth, for example, on people who are undernourished and/or without a reliable supply of safe water. Environmental insecurity – the double vulnerability that arises from both poverty and environmental change – is at its most obvious in low income societies. This is not to say that people in high income societies are immune from the risks of global environmental change, as was clearly revealed in the damage caused to some middle income households in New Orleans by Hurricane Katrina in 2005. Nevertheless, as is the case with all environmental hazards, Katrina showed that it is the poor who tend to suffer the most, and that vulnerability to environmental change is socially differentiated across gender, class, race, and age.

The poor and the marginalized tend to be more vulnerable to environmental change for numerous interlocking reasons. They are typically heavily dependent on natural resources and ecosystem services whose abundance and quality is at risk from environmental changes. They are more likely to live in places that are affected by environmental degradation and hazards because land is cheaper or free in these places. Their homes frequently do not have adequate waste disposal services and water supply. They cannot afford insurance and rarely receive compensation for losses they incur. The poor and marginalized are often more food insecure due to smaller landholdings and poorer quality land, and/or because they have little purchasing power in food markets. Their property and labor rights are nonexistent or minimal and variable. They have less capacity to exercise their skills and act on their values and aspirations because of barriers to accessing information, health care, and education. They have little voice in the decisions that affect them, be they the allocation of food at mealtimes, the allocation of land among villagers, the formulation of social welfare and land use policies, the development of large scale infrastructure projects, and the formulation of disaster and environmental management policies. They therefore tend to be more exposed, more sensitive, and less able to adapt to environmental change. Geographers such as Hans Bohle, Tom Downing, and Michael Watts have made critical contributions to understanding the ways in which these social determinants structure vulnerability to environmental change.

It is therefore the larger context of modernity and its contrapuntal forces – the limited alleviation of environmental constraints for some and the exacerbation of environmental problems for others, the attainment of a certain kind of material well being and power for some and the exacerbation of poverty and powerlessness for others – that gives rise to the problem of global environmental change and as a risk to human security. Human security is a variable condition where people and communities have the capacity to manage stresses to their needs, rights, and values. Environmental change is a human security problem when it undermines the needs, rights, and values that are important to individuals and communities.

The growing attention paid to environmental change as a human security issue does not fit so well with the national security policy community. It does, however, have some appeal to the development policy community and to environmental groups and organizations. In the future, it is likely that the concept of environmental security as human security will become more central in the fields of environmental studies and development studies, and figure more prominently in their respective policy domains. A key institution in this respect is the Global Environmental Change and Human Security project, in which geographers have been influential.

Conclusions

When a problem is identified as a security issue, it can lead to dramatic state action to address that problem – at least as evidenced by the large proportion of national budgets spent on defence. This is one of the intentions of those environmentalists who promote the idea of environmental security; they seek to make it a matter of 'high' politics that warrants extraordinary responses from governments equal in magnitude and urgency to their response to more orthodox security threats. As a consequence, environmental security has to some degree raised the profile of environmental issues among foreign and security institutions, as well as development and aid institutions, so that there is now a general recognition that environmental changes can in some sense be considered as security issues. However, calling something a security problem can also lead to state monopolization of solutions.

The most influential interpretations of environmental security are those that fit well the dominant approach to national security. The security policy community and the armed forces, especially in the United States, have largely accepted the arguments that environmental change may be a cause of violent conflict, and that environmental problems in other countries are threats to national security. So environmental security is still largely understood to be about threats to the nation state rather than to the environment *per se*, to other states, or to individuals. This suggests that while environmental security may have broadened the idea of security, it has been less successful in changing its scale of reference.

There are nevertheless some good reasons for continuing to use the concept of environmental security. Understood in terms of human security, the concept has gained some purchase with development agencies as it helps to capture the environmental risks that create vulnerability. It better communicates the critical nature of environmental problems such as climate change on countries most at risk, or the health impacts of water pollution, than standard concepts like sustainability or vulnerability. It can also help integrate local (human security), national (national security), and global (international security) levels of environmental change and response. It also serves as a boundary object that helps link international relations with development studies and environmental studies, and people from diverse arms of government, civil society, and academia across both the developed and developing worlds to produce new fusions of knowledge and awareness. Finally, there is value in continuing to contest the legitimacy of the dominant security paradigm by pointing to the contradiction between simple state based and military approaches to national security, and the complex, multiscale, and transboundary nature of environmental flows. Indeed, the relevance of environmental security will most probably increase as the impacts of certain environmental problems seem set to increase: concentrations of greenhouse gases in the atmosphere are ever increasing, and so will sea levels and the intensity of climatic hazards, such as cyclones, floods, and droughts; problems of nuclear waste have not been dealt with and the stockpile is still growing; and water demand is increasing but supply is relatively fixed.

See also: Critical Geopolitics; Environment; Livelihoods; Rio Summit; Sustainable Development.

Further Reading

Baechler, G. (1999). *Violence Through Environmental Discrimination: Causes, Rwanda Arena, and Conflict Model*. Dordrecht: Kluwer.
Barnett, J. (2001). *The Meaning of Environmental Security: Ecological Politics and Policy in the New Security Era*. London: Zed Books.
Conka, K. and Dabelko, G. (eds.) (2002). *Environmental Peacemaking*. Baltimore, MD: John Hopkins University Press.
Dalby, S. (2002). *Environmental Security*. Minneapolis, MN: University of Minnesota Press.
Deudney, D. and Matthew, R. (eds.) (1999). *Contested Grounds: Security and Conflict in the New Environmental Politics*. Albany, NY: State University of New York Press.
Diehl, P. and Gleditsch, N. (eds.) (2001). *Environmental Conflict*. Boulder, CO: Westview Press.
Homer Dixon, T. (1999). *Environment, Scarcity, and Violence*. Princeton, NJ: Princeton University Press.
Myers, N. (1996). *Ultimate Security: The Environmental Basis of Political Stability*. Washington, DC: Island Press.
Peluso, N. and Watts, M. (eds.) (2001). *Violent Environments*. Ithaca, NY: Cornell University Press.
WCED (World Commission on Environment and Development) (1987). *Our Common Future*. Oxford: Oxford University Press.

Relevant Websites

http://www.cusa.uci.edu
Center for Unconventional Security Affairs (CUSA) at the University of California, Irvine.
http://www.envirosecurity.org
Institute for Environmental Security.
http://www.wilsoncenter.org
The Environmental Change and Security Program at the Woodrow Wilson International Center for Scholars, including the influential Environmental Change and Security Project Report.
http://www.gechs.org
The Global Environmental Change and Human Security (GECHS) Project.
http://www.unep.org
The United Nations Environment Programme.

Environmental Studies and Human Geography

C. Sneddon, Dartmouth College, Hanover NH, USA

Glossary

Environmental Policy The realm of public policy that deals explicitly with environmental concerns through creation, negotiation, and implementation of environmental legislation and regulations at local, national, and global levels.

Environmental Politics The realm of politics that deals explicitly with social conflicts over the access and use of resources, conflicting discourses over nature, and debates over solutions to environmental deterioration.

Human–Environment Geography One of the core areas of the field of geography, focusing on the dynamic interactions, at multiple scales, between the social sphere and a range of biophysical systems.

Interdisciplinarity An approach to education and research that emphasizes the need to employ and integrate multiple approaches to knowledge production in order to confront the full complexity of societal problems.

Sustainability The assortment of theories and practices that define and put into practice a normative vision of how to reconcile human activities with the long-term integrity of the planet's biophysical systems.

Introduction

As academic disciplines, the fields of geography and environmental studies share common themes of study, theoretical perspectives, and methodological approaches. They also differ according to how their relative coherence as a distinct academic discipline, how they define and employ interdisciplinarity, and how they relate to public opinion and public policy. Human geography is the part of geography that explores the patterns and processes that shape social relations and delineates human interactions with the environment. Human–environment geography deals directly with the latter, and typically requires some understanding of the dynamics of climatology, geology, ecology, hydrology, biogeography, and geomorphology, in addition to the multiple historical and social processes that produce the diversity of cultural, economic, and political institutions across the planet. Human–environment geography, as with human geography, employs a diverse array of qualitative and quantitative methods and methodologies, including case studies, historical research, textual analysis, survey research, and statistical analysis among others. Environmental studies, in a fashion similar to human–environment geography, examines how human interactions produce environmental changes and how those changes in turn influence social dynamics. It is a broad field of inquiry that includes within its purview both the natural environment and human produced environments such as cities and agricultural systems. Environmental studies is strongly oriented toward the study of current environmental problems and the complex set of biophysical, economic, political, and cultural dynamics that have produced them. The problem-oriented nature of environmental studies has produced a strong set of linkages between its practitioners and a variety of policymakers and advocates in government agencies, the popular media, independent research institutes, nongovernmental organizations, and community groups.

Roughly 400 colleges and universities in North America offer an undergraduate program in environmental studies, and there are over 220 graduate programs in environmental studies. According to Educational Directories Unlimited, an online guide to graduate programs, there are an additional 150 graduate programs in environmental studies and hundreds of undergraduate programs with an environmental studies orientation beyond North America. This represents a remarkable rate of growth for a discipline that originated in the late 1960s. While environmental studies is by definition an interdisciplinary field of study that embraces education and training in the natural sciences, social sciences, and humanities, a majority of programs at both the undergraduate and graduate level are firmly oriented toward environmental science. The variety of disciplinary expertise and course offerings in environmental studies is quite broad. Topics of study in the biophysical sciences include environmental chemistry, environmental toxicology, geophysical sciences, conservation biology, wildlife ecology, and host of applied fields oriented toward natural resources such as fisheries, forestry, water resources, and the like. The breadth of social science approaches within environmental studies is similarly wide, spanning anthropology, environmental sociology, environmental politics, environmental and resource economics, and geography. While fields from the humanities are not as well represented in environmental programs, many programs do include a focus on environmental literature, ecocriticism, or similar areas of study. Finally, a significant number of programs are oriented toward or include aspects of environmental engineering and public health, while many also have linkages with environmental law and public policy programs.

The Historical Roots of Disciplinary Identities

The historical trajectories of human–environment geog raphy and environmental studies as academic disciplines and as approaches to the study of human–environment relationships provide a foundation for interpreting their current place within academia and within the public eye. Human–environment geography, as a core concern of the broader discipline of geography, has waxed and waned following the fortunes of its parent discipline from the time of geography's formalization within European academic structures in the late nineteenth century and throughout its development, particularly in the United States, during the twentieth century. By contrast, environmental studies emerged as a response to mounting societal concerns over environmental pol lution, resource depletion, and nature conservation in the latter half of the twentieth century. Many of the earliest environmental studies programs in Canada and the United States came about in conjunction with the enactment of a significant amount of national environ mental legislation in the late 1960s and early 1970s. The key differences that drive how scholars in human geography (and subsequently human–environment geography) versus those in environmental studies inter pret environmental change and human–environment relations derive from the quite different historical development of the two fields. A closer examination of these histories reveals, on the one hand, human–environment geography's maturation almost entirely within the strictures of academic institutions and, on the other, environmental studies' emergence in response to an array of external actors and events located within broader society.

Geography, and as a result human geography, traces its roots to the nineteenth and early twentieth centuries, when it became more or less established as a scholarly endeavor in a variety of European and North American academic settings. Geography as the study of human–environment relations owes its early academic standing to the efforts of European practitioners – active in the late nineteenth and early twentieth centuries – such as Alexander von Humboldt and Frederich Ratzel in Germany, Elisée Reclus in France, Joachim Schouw in Denmark, Vidal de la Blache in France, and Halford Mackinder in Great Britain. Oftentimes, these efforts to inculcate geography within academic insti tutions were directly linked to the nationalist sentiments and geopolitical aspirations of these scholars' nation states, where the descriptions of the physical and human characteristics of new places were useful to ex panding colonial empires. These geographers, through their research, writings, and participation in public forums, sought to define geography's unique place in academia as both a spatial–chorological discipline and a field concerned with explaining the dialectical rela tionship between human activities and environmental processes.

In the United States, the human–environment tradition in geography was carried forward in the early twentieth century by scholars such as William Morris Davis, Ellsworth Huntington, and Ellen Churchill Semple, all of who strongly advocated the idea that geographical circumstances determined human behavior. This perspective, now recognized as environmental determinism, was challenged later in the century through the work of Harlan Barrows, who called for stronger links between geography and the budding field of ecol ogy, and Carl Sauer, whose unique vision of landscapes as mutually constituted by cultural and biophysical processes dominated discussions of human–environment geography certainly up to and including the 1960s. From the late 1960s to the present, however, geography's concern with the human–environment condition has witnessed an efflorescence of theoretical and methodo logical approaches. These approaches – ranging from older traditions such as human ecology, cultural ecology, and hazards research through to more recent iterations such as political ecology and the social construction of nature school – have been influenced by a variety of social theories that became incorporated into the dis cipline during this period. The most prominent of these, from the standpoint of human–environment geography, include Marxist political economy, ecological Marxism, feminist theory, post structuralism, social constructivism, and more recently postcolonialism, science studies, and cultural studies. These perspectives have at times rested uncomfortably against human–environment geography's continuing engagement with ecological theory and a broad array of biophysical sciences. During the past two decades, scholars working within human–environment geography have also widened their geographical purview. The array of approaches collectively identified as polit ical ecology have been particularly ardent about in vestigating human–environment relations in the Global South, where a majority of research has focused on the intersections among resource dependent communities, developmental states, and an evolving global economy. Tellingly, efforts to define the essential character of human–environment geography and its place within the social sciences by its practitioners throughout the twentieth century have taken place almost entirely within academic settings.

Environmental studies, although its roots can in part be traced back to the nineteenth century and the seminal work of George Perkins Marsh, is a relatively young discipline in terms of academic recognition, becoming more or less firmly established in a variety of North American and European institutions in the 1970s and 1980s. In contrast to human–environment geography, the origins of environmental studies and the continuing activities of its practitioners are the result of a series of dialogs between academic institutions and individuals and events located within broader societal contexts. Authors whose scientifically grounded arguments were explicitly targeted at a broader audience – including Rachel Carson's *Silent Spring*, Barry Commoner's *The Closing Circle*, Paul Ehrlich's *The Population Bomb*, and a host of others – exerted significant influence over the early development of environmental studies. Environ mental studies owes its existence, in large part, to the efforts of students and faculty who became intellectually and politically engaged with the debates over environ mental quality and resource exhaustion ignited by these widely read works and the ensuing public debates over their ramifications.

In the United States, as support both within and outside of academia grew for programs of higher edu cation focused on environmental problems, a number of undergraduate programs emerged in the 1970s. These early programs, such as those at Dartmouth College and Middlebury College in the northeastern United States during the late 1960s and early 1970s, linked together a loose confederation of faculty whose primary research and teaching responsibilities tended to be identified with traditional disciplines. The primary goal at these and other liberal arts institutions was to raise awareness of the rapidity and intensity of environmental changes con fronting the planet and to educate undergraduate stu dents in both the fundamentals of environmental science and how environmental knowledge is applied in society at large through policy and advocacy mechanisms. At the graduate level, environmental studies became interlinked with applied sciences, focusing on the training of resource professionals who upon graduation typically found employment with state and federal resource management agencies or specific resource based indus tries. For example, a number of natural resource and forestry schools and colleges embedded within broader university settings, including those at Yale University, the University of Minnesota, the University of Michigan, and many others, broadened their graduate educational pro grams to embrace growing societal concerns over the long term effects of pollutants, the global degradation of ecological systems, biodiversity loss, and, significantly, the need to address these problems through policy measures. In addition, research and writing in environ mental studies have both critically analyzed and been influenced by a range of philosophical perspectives, in cluding the ecocentrism of the 'deep ecology' movement, James Lovelock's Gaia hypothesis, and ecofeminism.

Programs in environmental studies, in North America, Europe, and elsewhere, have increasingly taken a global perspective. This has manifested in two distinct ways. The first concerns recognition on the part of scholars in environmental studies that the scope of planetary prob lems such as climate change, transboundary pollution, degradation of marine ecosystems, changes in biogeo chemical cycles, and biodiversity loss are genuinely global. Again, foundational publications identified with environmental studies – for example, *The Limits to Growth* by Donella Meadows and her co authors, the *World Conservation Strategy* by IUCN (World Conservation Union), and the assortment of publications put out by the Intergovernmental Panel on Climate Change (IPCC) – bolster the claim of environmental studies scholars that biophysical processes transcend political boundaries and the resolution of global environmental problems' demands an international perspective. The second global perspective focuses on questions com monly grouped under the rubric of environment and development studies. As questions over North–South relations, intergenerational equity, and sustainable development were catapulted into the public sphere by publication of the Brundtland Commission's *Our Common Future* in the late 1980s and global fora such as the United Nations Conference on Environment and Development (UNCED) and the follow up World Summit on Sustainable Development (WSSD) in 1992 and 2002, respectively, environmental studies began to embrace difficult questions regarding rapid ecological changes in the developing regions of Latin America, Africa, and Asia in the context of industrialization, agricultural intensifi cation, and diverse political settings.

Converging Identities and Themes

Human geography and environmental studies converge in their embrace of a plurality of approaches to their core subject matter, and in their mutual concern with material transformations of the environment, human drivers of such transformations, and human conceptions of nature. Perhaps the most striking similarity between human geography and environmental studies is that both are fundamentally interdisciplinary in their outlook on knowledge production, although this becomes evident within the two fields in different ways. Both fields draw on multiple epistemological and methodological tradi tions – ranging from positivist to constructivist and from quantitative to qualitative – to examine the dialectical relations between social and biophysical changes. There are several hybrid approaches, inclusive of scholarly traditions beyond geography and environmental studies, which defy easy categorization, but include fields such as cultural ecology, human ecology, political ecology, environmental geography, and numerous others. The acceptance of interdisciplinarity as a fundamental condition of human–environment geography and environmental studies has provided a unique opportunity for researchers in both fields to employ multiple methods and transcend narrow discipline based topics and approaches. In environmental studies, in particular, it also signifies recognition that the resolution of problems brought about by anthropocentric forms of environmental degradation demands an approach able to understand and interpret the economic, cultural, and political aspects of biophysical change.

Despite the similar claims of interdisciplinarity by advocates of both human–environment geography and environmental studies, scholars typically come to the fields via quite different educational pathways. Most faculty members in geography enter into academic positions thoroughly socialized into the traditions and conventions of the discipline, broad as these might be in terms of the physical and human branches. As a result, human geographers, and in particular human–environment geographers, tend to have at least tangential experience in reconciling their methodological predilections with the substantially different and largely quantitative orientations of their colleagues in physical geography and the mapping sciences. Scholars within environmental studies programs, by contrast, tend to settle within their academic bases after attaining far more specialized degrees from a wide variety of natural and physical sciences (e.g., ecology, conservation biology, and biogeochemistry), social sciences (e.g., political science, economics, sociology, anthropology, history, and geography), and humanities (e.g., philosophy and literature programs). Moreover, this fragmentation of disciplinary identities in environmental studies is in many cases reinforced by the institutional milieu in which programs reside. Many, if not a large majority of environmental studies programs, consist of loose confederations of individual faculty members whose primary intellectual attachment remains with the academic field within which they received their graduate degrees. Institutionally, many faculties in environmental studies programs are more strongly linked to their primary departmental home – whether it be in chemistry, the biological sciences, sociology, political science, or within a professional school – than to their colleagues and students in environmental studies. While the reverse may also be true, the implications of this are profound. At one level, there are fewer institutional incentives within environmental studies to undertake the necessary theoretical and philosophical work of carefully defining, and on occasion defending, a shared intellectual identity. Such identity crises have also plagued human geography, and both fields at various times and places in their histories have had to defend their continuing relevance when threatened by dwindling resources and academic hierarchies antagonistic toward their fundamentally interdisciplinary worldview.

Thematically, environmental studies and human–environment geography encompass a similar collection of research topics. A first revolves around the actual material transformations of landscapes and waterscapes precipitated by various human activities. A second area of shared inquiry concerns the creation and maintenance of institutions (e.g., social conventions, regulations, and policies) for governing, exploiting, and managing the environment. A third focuses on the roles played by different actors, collectively and individually, in facilitating, compounding, or mitigating biophysical changes that are ecologically and socially disruptive. Finally, a fourth region of shared concerns relates to human perceptions of nature and how conceptions of nature have shifted over time. Within these broad areas, researchers in human–environment geography and environmental studies have carried out a host of interrelated research projects characterized by theoretical sophistication and empirically detailed accounts of human–environment relations in diverse locations and at household, community, national, regional, and transnational scales. More notable examples include: the multiple political and economic drivers of deforestation in regions as diverse as the northwest coast of North America, the Amazon River basin, and the island of Borneo; the multiple historical determinants of environmental change in the Sahel region of western and northern Africa; the transformation of various river systems across the planet via the construction of large dams and irrigation infrastructure to promote both industrialization and agricultural intensification; the role of different state agents in facilitating, inhibiting, and reversing harmful ecological changes brought about by human production and consumption; and the shifting understandings of what constitutes nature exhibited by different cultural traditions at different times and places. While the relative bias toward the environmental sphere versus the social sphere may be pronounced in many of these studies, all endeavor to balance knowledge of biophysical changes with knowledge of the human societies whose internal and external interactions with a host of cultural, economic, and political forces produce and shape these changes.

Diverging Identities and Approaches

In many ways, the divisions within environmental studies mirror the broad division within geography between the realms of the human and the physical. If anything, environmental studies is pulled in even more directions, attempting to straddle multiple fault lines within and between the natural sciences, social sciences, and humanities. As a result, much of the work that comes to be associated with environmental studies might more accurately be described as the work of a chemist, Earth scientist, anthropologist, ecologist, or sociologist working within the broad rubric of environmental studies. Furthermore, most research produced by practitioners of environmental studies tends to be published in journals that are more specialized in orientation. There are relatively few outlets for publication that are unequivocally geared toward a field broadly defined as environmental studies. Moreover, there are comparatively few, if any, professional associations or regularly scheduled conferences targeted specifically to scholars and educators within (in the broadest sense) environmental studies. Thus, it is impossible to identify with any precision what a typical researcher in environmental studies might produce. Much of the interdisciplinary work of environmental studies is carried out primarily through curriculum development and teaching, not necessarily through collaborative research, although there are major exceptions. Human–environment geography certainly suffers from a similarly fragmented identity, but the discipline of geography has a wide array of institutional norms and mechanisms (e.g., journals, professional societies, and regular meetings) that generate and maintain research affinities for scholars of human–environment relations.

Despite the observation that both environmental studies and human–environment geography champion a range of epistemological and methodological positions, the fields manifest this intrinsic interdisciplinarity in distinct ways. At the broadest levels, a majority of environmental studies programs in North America, Europe, Australia, and elsewhere emphasize environmental science over approaches grounded in the social sciences or humanities. Within those programs that seek a greater divisional balance, faculty members in the natural sciences and social sciences are more equally represented. With relatively few exceptions, programs with a significant number of scholars approaching environmental studies through a humanities oriented lens are rare. Inside those programs that have achieved a relative balance between environmental scientists and social scientists, the social scientists in such settings tend to be grounded more in quantitative social science traditions such as environmental and resource economics, and the policy sciences. Indeed, while a plurality of disciplinary training (e.g., the ecological sciences, environmental health, environmental economics, ecological anthropology, and environmental history) is well represented in environmental studies programs, methodological pluralism and a willingness to accept qualitative approaches as viable and crucial are more contentious issues.

In human–environment geography, one of the most obvious manifestations of interdisciplinarity is through an engagement with a diverse array of social theories and the methodological options such theories imply. The recent history of political ecology is instructive in this regard. At the time of its materialization as an important research area within geography and cognate disciplines in the 1980s, political ecology undertook a melding of political–economic approaches grounded primarily in Marxist theories and a refined understanding of ecological processes. The result was an approach that sought explanations of environmental transformations within a host of causal mechanisms, ranging from the land management decisions of individual farmers, to state initiated development programs and the policies of international financial institutions. In subsequent years, human–environment geographers working within a political ecology framework expanded their theoretical repertoire to include, for example, feminist theories to interpret the gendered character of resource use and management, and post structuralist theories to demonstrate the historically and socially contingent character of how different actors perceive and act upon environmental changes. Concerns over power relations, culturally mediated understandings of nature, and the highly complex character of explanations of environmental transformations led political ecologists to develop and employ an array of qualitative research designs and methods, including case study approaches, ethnographies, interviews, oral histories, participant observation, and textual analyses. Thus, while interdisciplinarity in environmental studies is best represented by specialists from the natural and social sciences engaging in collaborations with colleagues with different methodological tool kits, interdisciplinarity in human–environment geography is best represented by the diverse approaches, epitomized in the variety of social theories incorporated into geographical research, employed by single researchers and their collaborators.

In part, due to specific institutional histories, there are important differences in how human geographers and practitioners of environmental studies, in general, approach the production of knowledge and, once produced, communicate geographical or environmental knowledge to different audiences. While these differences should not be overstated – there are individuals and specific programs that transcend this dichotomy – they are important to an understanding of each field's institutional cultures and broad outlook on research and education. Due to its historical development in response to public calls for research leading to resolution of environmental problems, environmental studies has a long tradition of direct engagement with representatives of government agencies, nongovernmental organizations, and to some degree the private sector. For example, a good deal of the early research conducted by scholars in environmental studies during the 1970s was directly aimed at effecting changes in environmental policy. This also explains the strong presence of resource economics, environmental eco nomics, and, more recently, ecological economics within environmental studies programs, since many of the battles over environmental policy both past and present revolve around perceived trade offs between economic growth and ecological sustainability. The aims of policy oriented work in environmental studies revolve around the desire to insert the results of environmental research into policy discussions at multiple levels. By contrast, human–environment geography (and human geography more broadly) has largely abdicated formal engagement with public policy, oftentimes assuming a more critical stance toward the policy realm, particularly where gov ernment agents are concerned. While this can in part be explained by human geography's intellectual history within academia, it is also a reflection of many scholars' insistence that the state, whether through the actions of specific agencies or broad policy initiatives, is a central actor in contributing to environmental degradation. Under this view, a formal engagement with policymakers is perceived as unnecessarily restrictive in terms of the types of questions under investigation.

Lessons and Opportunities

Human geography and environmental studies, as both disciplinary identities and broad approaches to the study of human–environment relations, share a great deal. Both are concerned with identifying, describing, and analyzing a broad range of biophysical and social processes, frequently with the goal of demonstrating more ecologically sustainable and socially just modes of human interactions with the natural world. Both are inherently interdisciplinary, attempting to carve out some semblance of internal coherence while embracing a plurality of epistemological and methodological approaches. This has led both fields to engage in a certain level of intellectual hand wringing – undoubtedly more prominent within geography – regarding the rela tive status of their professional identity within the broader contours of the academy. Although these identity crises play out differently depending on institutional context, all are reflective of both fields' continuing struggles to come to grips with the tension between intellectual clarity and unity on the one hand and an embrace of pluralism on the other.

The interrelated notions of sustainable development and sustainability constitute a potentially fruitful intel lectual region for collaboration between the two fields. Discourses and policies concerning sustainable develop ment, following formalization of the concept within a wide range of governmental and nongovernmental forums during the 1990s, have blossomed in recent years. The broad goal of sustainable development is to recon cile human social and economic practices with the long term integrity and resilience of the Earth's biophysical systems. The struggle to define and institutionalize sustainability takes place at many institutional levels and geographical scales, and this effort holds significant consequences for the way people are able to exert control over resources that influence the sustainability of their livelihoods. The explicit linking of human activities and biophysical dynamics within the broad rubric of sus tainability studies would seem to offer productive avenues of research, advocacy, and teaching for both human geographers and practitioners of environmental studies.

Taking account of the quite specific intellectual his tories, different approaches to interdisciplinarity, and contrasting relationships with the nonacademic world ex hibited by environmental studies and human–environment geography, there are several ways in which the fields might benefit from the experiences of the other. Environmental studies would benefit from a deeper engagement with the wide variety of social theories and methodological ap proaches currently employed within human–environment geography. Human geography would profit from a more direct effort to address its research toward a diverse audience of policymakers and citizens. Practitioners of environmental studies have been more adept in this regard. As a result, scholars in environmental studies can more readily make the claim that the knowledge they produce has a more direct impact on how environmental concerns are understood, and how biophysical entities are governed and managed. Yet, because of their general disinterest in theorizing the multiple social roots of environmental change, environmental studies scholars risk presenting an oversimplified and partial explanation of environmental degradation to policymakers and citizens.

As public concern over environmental issues con tinues to increase into the twenty first century and the complexity of human–environment relationships con tinues to increase, interdisciplinary modes of knowing will be crucial in increasing public knowledge of a host of biophysical processes and helping to apply this know ledge to the array of local, national, and global en vironmental problems that confront human–environment relations in the twenty first century. The promise of the architects of research, education, and outreach initiatives in both human–environment geography and environ mental studies is that these highly interdisciplinary fields of inquiry are immanently qualified to address these problems.

See also: Ecology; Environment; Environmental Policy; Environmental Regulation; Geography, History of;

Interdisciplinarity; Land Change Science; Political Ecology; Space I; Space II; Sustainability; Uneven Development.

Further Reading

Adams, W. M. (2001). *Green Development: Environment and Sustainability in the Third World* (2nd edn.). London: Routledge.

Dickens, P. (1992). *Society and Nature: Towards a Green Social Theory*. Philadelphia, PA: Temple University Press.

Maniates, M. F. and Whissel, J. C. (2000). Environmental studies: The sky is not falling. *BioScience* 50, 509 517.

Nelissen, N., van der Straaten, J. and Klinkers, L. (eds.) (1997). *Classics in Environmental Studies: An Overview of Classic Texts in Environmental Studies*. The Hague: International Books.

Neumann, R. (2005). *Making Political Ecology*. London: Hodder Arnold.

Robbins, P. (2004). *Political Ecology: A Critical Introduction*. Malden, MA: Blackwell.

Turner, B. L. II. (2002). Contested identities: Human environment geography and disciplinary implications in a restructuring academy. *Annals of the Association of American Geographers* 92, 52 74.

Environmentalism

A. Davies, Trinity College, Dublin, Republic of Ireland

Glossary

Environment Circumstances or conditions that surround an organism or ecosystem; the social and cultural conditions that affect individuals and societies.
Environmental Of or pertaining to the environment such as environmental factors; environmental pollution; environmental disaster; environmental deterioration.
Environmentalist Someone who is concerned with environmental issues. It is usually applied to those who are actively working to protect the environment from destruction or pollution.
Environmental Ethics Reflection on the role and responsibility of humans to the environment and the question of whether the environment can be treated rightly or wrongly.

Environmentalism

Definitions

Despite general agreement that environmentalism refers to attention to interactions between natures and cultures there remains considerable contestation about the possibility of establishing a universally acceptable and unambiguous description of the term and its associates (such as environment, environmental, and environmentalists). Given that the words nature and culture have been identified as being as complex as life itself, it is unsurprising to find a similar level of complexity in definitional debates surrounding environmentalism. While most general definitions propose that environmentalism refers to concerns for environmental protection, particularly from the harmful effects of human activity, beneath the veneer of simplicity lies a host of ethical, political, and scientific questions such as what is the 'environment' and are humans part of, or apart from, it? What constitutes 'protection' and for what purpose? And finally, how is 'harm' established and by whom? Some analysts have replaced 'concern' with 'advocacy' in order to emphasize the politicized nature of environmentalism, while others have introduced organizational or social movement dimensions to their definitions. However rather than resolving ambiguities these modifications to general definitions serve to highlight the spectrum of positions that have come to be collected under the expanding banner of environmentalism. This spectrum of environmentalism is considered below following attention to its origins and evolution. These temporal and ideological reflections on environmentalism set the scene for a discussion of current debates that reveal the significance of geographical inquiry in environmental matters.

Origins and Evolutions

A lack of definitional clarity has implications for the identification of precise roots and patterns of development within the field of environmentalism. It has been suggested that environmentalism emerged only relatively recently in response to tensions between economic development and its use of natural resources (or environmental goods), particularly through the resulting production of 'pollution' (or environmental bads) and resource depletion that became the focus for numerous analysts during the 1960s and 1970s. However, while it is widely acknowledged that popular attention to the potentially negative ramifications of human activities did increase during the latter part of the twentieth century, identifiable elements of concern for human–environment relations were evident in earlier periods. They can be seen in debates about preservation and conservation of nature through the formation of national parks – as illustrated in the works of John Muir and Gifford Pinchot – and in the romanticism of poets and artists – such as William Blake or William Wordsworth – who called for the protection of the inherent beauty of nature throughout Europe and North America during the nineteenth century. Excavating the foundations of modern environmentalism even further has led authors to identify diverse roots in scientific enquiry of the Enlightenment, political ideologies of ancient Greece, and interpretations of religious doctrines.

Eras of Environmentalism

With such a long and diverse history, attempts have been made to delineate specific phases or eras of environmentalism. There is common ground in many of these analyses in terms of an identification of concern with industrialism and urbanism (and allied discussions of the need for conservation and preservation of rural and agrarian landscapes) during the late nineteenth century as the precursor of the modern environmentalism of the 1960s and 1970s. Key features of the modern era of environmentalism were the incorporation of societal considerations in debates, a widening of the social base of environmentalists and the formation of pressure groups during a period of relative calm in western countries following the end of World War II. At the same time the effects of environmental catastrophes were being disseminated to a wider public through the media and popular texts. Rachel Carson's *Silent Spring*, published in 1962, documented the impacts of the pesticide DDT on ecosystems and a decade later the *Limits to Growth* manifesto, developed by a team of systems scientists and computer modelers, challenged conventional wisdom of unfettered expansion and warned of impending limits – especially environmental limits – to human development and economic growth. Beyond this there is less coherence across analyses about distinct phases within environmentalism. Some authors have emphasized its emergence as a mainstream feature in society and politics following the recognition of global environmental problems. Others focus on the growing diversity of organizational forms and concerns of environmental groups contrasting large transnational environmental groups with grassroots organizations or categorizing the different forms of action utilized from formal lobbying to civil disobedience. Contemporary environmentalism is then typified by diversity in ideas, practices, and organizational cultures to the extent that a number of subforms of environmentalism have been identified. A selection of these forms is detailed in **Table 1**.

Environmentalism or Environmentalisms?

Considerable attention has been paid to deconstructing the various strands of environmentalism in an attempt to provide analytical frameworks for theoretical and empirical inquiry. Arne Naess produced a *Deep to Shallow Ecology* typology, while Robyn Eckersley constructed a classification demarcating 'ecologism' from 'conservatism' and Andrew Dobson established a 'green–green' dichotomy. However the most cited framework is that of

Table 1 Illustrative diversity of contemporary environmental movements

Forms of environmentalism
Deep ecology views nature as having intrinsic value and adopts a strong ecocentric system of environmental ethics. It is effectively a philosophy that calls for a profound shift in our attitudes and behavior based on voluntary simplicity; rejection of anthropocentrism; reconnecting with nature; decentralizing power; celebrating cultural and biological diversity; a belief in the sacredness of nature. There is often a strong commitment to direct personal action to protect nature, improve the environment, and bring about fundamental societal change.
Social ecology takes the position that most ecological problems arise from deep-seated social problems and that therefore ecological problems cannot be clearly understood, much less resolved, without resolutely dealing with problems within society. The current ecological situation is seen to be the product of capitalism. As such social ecology presents a radical critique of current social, political, and anti-ecological trends. It claims to be a reconstructive, ecological, communitarian, and ethical approach to society.
Ecoterrorism is typified by acts of violence, vandalism, property damage, or sabotage against individuals and companies in the name of environmentalism. The term has also been applied to crimes committed against companies or government agencies in order to prevent or interfere with activities allegedly harmful to the environment. Another example of ecoterrorism, sometimes referred to as environmental warfare, consists of the deliberate and illegal destruction, exploitation, or modification of the environment as a strategy during periods of armed conflict.
Ecofeminism is commonly characterized as an intellectual and practical movement based around the interactions of women and nature, and specifically concerns relating to the oppression of both women and nature through history. It is generally portrayed as a biocentric environmental movement although there are internal differences between ecofeminists about the relative importance of cultural and social concerns. Despite these internal politics the view that the oppression of women and nature are linked historically, materially, culturally, and ideologically remains the central tenet of ecofeminism.
Bioregionalism is focused on a concern to reconnect humans with bioregions that are characterized in terms of the unique overall pattern of ecological, geological, and historical characteristics that are found in a specific place. People are also considered an integral aspect of a bioregion if they attempt to live and work in harmony with the bioregion. In terms of the practices of bioregionalists there is an emphasis on practical actions rather than protest and on lifestyle habits rather than calling for government legislation.
Prometheanism (or cornucopianism/contrarianism/environmental scepticism) runs counter to the claims of other environmentalists that people are irreversibly damaging the environment. Environmentalists with this perspective believe that the benefits of economic development, even if it has negative environmental implications, are more important to society than the damage caused. Within this movement concern has been expressed that the environmental crisis identified by other environmental groups is exaggerated and not based on sound science. Others believe that technological developments will resolve any damage caused.
Free market environmentalism is an ideology that argues the free market is the best tool to preserve the health and sustainability of the environment. This position differs from the most common contemporary approach of looking to government intervention, through mechanisms such as taxation or legislation, to help prevent environmental degradation and is in stark contrast to social ecology positions.
Political ecology prioritizes the interconnections between politics and the environment. Includes study of political struggles over natural resources or where outcomes of political struggles are defined by access to natural resources.

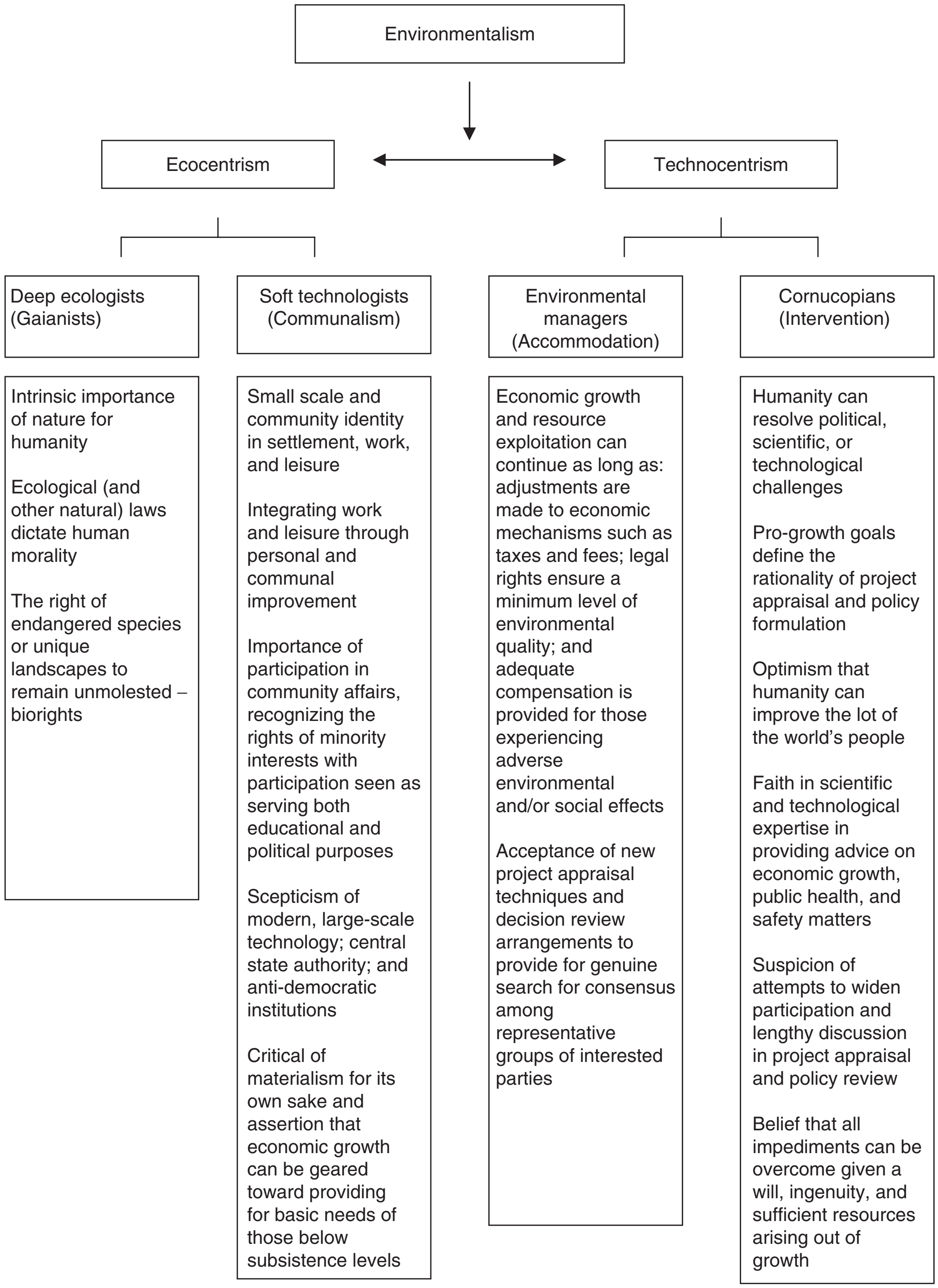

Figure 1 Ecocentrism and technocentrism. From O'Riordan, T. (1976). *Environmentalism*. London: Pion.

ecocentrism and technocentrism produced by Timothy O'Riordan (summarized in **Figure 1**).

O'Riordan made several amendments to his original dualism of ecocentrism and technocentrism, but the fundamental distinction between the desire to put the environment first, and modify societal practices to achieve this (ecocentrism), and the commitment to human control and management of the environment for societal benefit (technocentrism) remains intact. It is the ecocentric position, most frequently espoused by some nongovernmental organizations and civil society actors, that presents the most significant challenge to prevailing socioeconomic arrangements of contemporary society. Technocentric arguments, frequently articulated by big business and governments, tend to support the *status quo*. This distinction is also highlighted by Robert Garner in his analysis of approaches to environmentalism in which he seeks to simplify the plethora of analytical categories into two main approaches – radical and reformist – that each contains particular economic, political, and philosophical perspectives. In essence he suggests that reformist positions are fundamentally human centered, seeking protection of the environment for the benefit of humanity, and believing that this can be done without dismantling the structures of modern industrial society (similar to technocentrism). It is essentially an optimistic position where economic growth and environmental protection are seen to be compatible objectives and faith is placed in scientific and technological innovations to provide satisfactory pathways out of any tensions encountered when trying to achieve both goals; an idea that has become synonymous with the term 'ecological modernization'. In contrast, the radical perspective proposes much more dramatic change to political, economic, and social structures in order to create a harmonious and sustainable relationship between society and its wider environment.

It is however rare to find individuals or organizations conforming completely to any one of the positions identified by analysts of environmentalism and there is a danger of overemphasizing a simple dichotomy between radical and reformist environmentalism. Indeed as O'Riordan noted, while reflecting on his own typology, the world does not divide neatly in ecocentric environmentalists on the one hand and manipulative professionals and administrators on the other. Rather a blurring of boundaries occurs depending on the particular issue under consideration, the socioeconomic context, and the institutional arrangements that prevail. Inevitably much complexity is lost when diverse thoughts and practices are polarized. Nonetheless, the two positions (eco and technocentrism) are still recognizable in many public debates about environmental concerns in the twenty first century and nowhere more so than in the debates about 'sustainable development'.

As with environmentalism there has been a proliferation of attempts to specify exactly what sustainable development means since the first official declaration of sustainable development by the World Commission on Environment and Development that society should meet the needs of the present without compromising the ability of future generations to meet their own needs. Yet there are parallels in the debates about sustainable development and environmentalism that result from them both having common features of political concepts that call for some degree of social and economic change. There are those who call for a 'strong' interpretation of sustainable development and see environmental protection as a precondition for economic development (which would be allied toward the radical or ecocentric positions of environmentalism) and those who turn this statement on its head and suggest that economic development is a precondition for environmental protection as detailed in the technocentrism or the reformist positions.

As with sustainable development, environmentalism seems to have a generally understood meaning in social and political circles (a first level meaning), but beneath this general acceptance lies a deep core of contestation. Equally this deeper contestation is not restricted to critical engagements across the broad ecocentric–technocentric divide, but can also be found within apparently quite narrowly defined environmental subgroups such as ecofeminism, deep ecology, and bioregionalism. For example, while there is general agreement that ecofeminism refers to an intellectual and practical movement based around the interactions of women and nature, and specifically concerns relating to the oppression of both women and nature through history, there are significant internal politics among those who would identify themselves with the movement. Just as there is no one agreed form of environmentalism or feminism, ecofeminism remains a contested concept. The main division within ecofeminism is encapsulated by a distinction between cultural and social ecofeminism. Cultural (or radical) ecofeminism argues that women have a special relationship with nature because of their reproductive role such that their connections with nature emerge through their biological characteristics. Additionally cultural ecofeminists suggest that there are strong spiritual bonds between women and nature which are suppressed by the patriarchal conditions that dominate environmental protection programs and policies. These positions have stimulated a raft of criticisms, including some from other ecofeminists, that the core cultural ecofeminist views are guilty of essentializing women, effectively reducing them to their biology without engaging with wider social and political issues. In contrast social ecofeminist positions emphasize the social and political aspects of women–nature interactions seeing nature as a political rather than personal category and rejecting biological determinism. Parallels are drawn between exploitation of nature and exploitation in society according to class, race, or gender. Therefore for social ecofeminists, connections between women and nature lie in the socially constructed practices, characteristics, and roles relating to gender. Social ecofeminists are not without their detractors either though for a while avoiding essentialist criticisms other concerns have been voiced regarding the validity of making universal statements about women–nature interactions across cultures. This is particularly the case when many of the ecofeminists writing in western contexts have tended to be well educated, white, and affluent. Despite these criticisms the central tenet of ecofeminism – that the oppression of women and nature are linked historically, materially, culturally, and ideologically – remains solid.

It is not the case, however, that there is even common agreement that environmental problems require human attention at all. Recognized theoretically as an extreme position of the technocentric camp of environmentalism, there have been cases made, sometimes called Pro methean or cornucopian, which deny the existence of acute environmental problems. The argument of this position is that not only do humans have the ability to overcome any hurdles that face them; but that these challenges actually stimulate innovation and that many of the environmental concerns expressed by more radical environmentalists lack a rational, scientific basis. These positions emerged initially in response to the limits to growth debates of the 1970s where the suggestions of an imminent future of crisis and catastrophe were attacked. More recently the voices of a few high profile contra rians, such as Wilfred Beckerman and Richard North, alongside the work of a number of think tanks including the Institute of Economic Affairs and the Cato Institute, have maintained a Promethean presence in environ mental discussions. Such reactions to environmental movements and ideas could be seen as a symbol of the successes made by environmental groups in gaining popular and political attention. Certainly claims of 'greenwashing' (see **Table 2**) within large corporations and the mimicking of environmental group tactics by pro development lobbies, such as the wise use move ment, suggest that environmentalism in its many guises is perceived as influential. As a result some researchers, such as Andrew Rowell, have urged environmentalists to be wary of a backlash against environmentalism.

Geographies of Environmentalism

Concern with nature and society relations has been a central component in geography since its formation as a discipline in late nineteenth century Europe and North America. It was a part of what David Livingstone de scribed as the 'geographical experiment' to bring together nature and society under a single conceptual umbrella. In this way geographers have contributed significantly to analyses of environmentalism whether through the definitional debates described above or through detailed examination of human interactions with the environment. Other key geographical concepts such as scale and place are also relevant to understanding environmental ism. For, in addition to embracing a wider variety of environmentalisms (both ideologies and practices), there has also been a conceptual shift in attention to different scales of environmentalism. Recognition of global en vironmental problems was matched by the emergence of global environmental organizations such as Greenpeace, World Wide Fund for Nature, and Friends of the Earth, but transnational environmental advocacy net works linking grassroots, citizen based organizations (e.g., GAIA, the Global Anti Incinerator Alliance) across the globe have also emerged.

The scalar nature of environmentalism is perhaps most evocative in the context of climate change. At a global scale there are organizations such as ICLEI (International Council for Local Environmental Initia tives) who work with local authorities through their 'Cities for Climate Protection' campaign. Equally there are coalitions of groups campaigning for climate change action at national levels such as Australia's 'Climate Change Coalition', the 'Campaign against Climate Change' in the UK, and the Bangladesh 'Cam paign against Climate Change'. Many of these national campaigns seek to pressurize simultaneously national governments and international governance structures while also providing supportive structures for local cli mate initiatives. Irrespective of international or national support, innovative action is emerging at the local level with initiatives such as the 'Big Clean Up' in Auckland, New Zealand, and 'Transition Towns' in the UK. The transition movement currently involves around 20 groups of people living in towns, villages, districts, cities, and even a forest that have committed themselves to addressing issues related to climate change and peak oil through practical activities. Individuals are also being encouraged to consider their lifestyle choices through programs like 'Global Action Plan's Ecoteams' and the Sydney based sustainable living program 'Green Home'. Such actions, both across scales from the global to the local and from the grassroots to the transnational, can be seen as indicative of environmentalism's incursion into mainstream politics and lived environments.

In addition to recognizing scalar geographies within environmentalism there is a growing acknowledgement that both culture and identity play a significant role in forming relationships between societies and environ ments. Up until the 1980s it was common for North American and European geographers to travel to other locations and report on their experiences. Most geo graphical studies drew on western philosophies of environmentalism to structure the interpretations of

Table 2 Defining greenwash[a]

Greenwash
Green*wash: (n) Disinformation disseminated by an organization so as to present an environmentally responsible public image. Derivatives greenwashing (n). Origin from green on the pattern of whitewash.
Green*wash: (gr∼en-w^osh) -washers, -washing, -washed (1) The phenomenon of socially and environmentally destructive corporations attempting to preserve and expand their markets by posing as friends of the environment and leaders in the struggle to eradicate poverty. (2) Environmental whitewash. (3) Hogwash.

[a] www.corpwatch.org.

what they uncovered. Assumptions were often made that nonwestern cultures had not developed sophisticated environmental ideologies because they lacked awareness of the value of environments or because they did not have the resources to invest in protecting the environment (or indeed both). More recently there has been greater at tention to the nuances of indigenous environmentalisms in different locations and cultures. In particular the strengthening of political ecology and environmental justice movements has helped bring issues of environ ment and development in the global south to wider attention. The work of Arundhati Roy in relation to the Narmada Dam project in India and the activism of Ken Saro Wiwa, whose nonviolent campaign to protect the rights, livelihood, and environment of the Ogoni people ended in his execution by the military government of Nigeria are particularly visible examples, but there are many more organizations, issues, and perspectives across the global south. Together, philosophical theorizing and practical activism of many individuals and organizations, form a growing movement to transcend westernization in environmentalism and challenge assumptions that con cerns about environmental quality are the preserve of affluent, industrialized societies.

The Future of Environmentalism

It is still a moot point as to whether the future of environmentalism will be illustrated by a convergence of perspectives from the current spectrum of positions or a further proliferation of activities, philosophies, and organizations. Equally there is no agreement about whether either of these two paths will lead to the death of environmentalism, as proposed by Michael Shellenberger and Ted Nordhaus, or its elevation to more influential levels. According to Bill McKibben, the future of en vironmentalism will likely depend on what happens to the environment in the future and whether that will lead to the development of some new metaphor, leader, or vision for a better world. Whatever happens it is likely that more unity in diversity within current environ mentalism and among environmentalists will be needed to tackle the big issues, such as climate change, that require changes to entrenched patterns of trade and governance.

See also: Conservation and Ecology; Culture; Ecology; Environment; Nature; Political Ecology; Radical Environmentalism; Sustainable Development.

Further Reading

Baker, S. (2006). *Sustainable Development*. London: Routledge.
Bauer, J. (2006). *Forging Environmentalism: Justice, Livelihood, and Contested Environments*. New York: ME Sharpe.
Dobson, A. (1995). *Green Political Thought* (2nd edn.). London: Unwin Hyman.
Dryzek, J., Downs, D., Hernes, H. and Schlosberg, D. (2003). *Green States and Social Movements: Environmentalism in the United States, United Kingdom, Germany, and Norway*. Oxford: Oxford University Press.
Eckersley, R. (1992). *Environmentalism and Political Theory*. London: UCL Press.
Guha, R. (2000). *Environmentalism: A Global History*. New York, Harlow: Longman.
Milton, K. (1996). *Environmentalism and Cultural Theory*. London: Routledge.
Naess, A. (1973). The shallow and the deep, long range ecology movement. A summary. *Inquiry* 16, 95 100.
O'Riordan, T. (1976). *Environmentalism*. London: Pion.
Pepper, D. (1996). *Modern Environmentalism*. London: Routledge.
Pepper, D., Webster, F. and Revill, G. (2003). *Environmentalism: Critical Concepts*. London: Routledge.
Plumwood, V. (1993). *Feminism and the Mastery of Nature*. London: Routledge.
Robbins, P. (2004). *Political Ecology: A Critical Introduction*. Oxford: Blackwell.
Rowell, A. (1996). *Green Backlash: Global Subversion of the Environmental Movement*. London: Routledge.
Simmons, I. (1993). *Environmental History: A Concise Introduction*. Oxford: Blackwell.
Sutton, P. (2000). *Explaining Environmentalism: In Search of a New Social Movement*. Aldershot: Ashgate.

Relevant Websites

www.climatechangecoalition.com.au
Climate Change Coalition.
www.globalactionplan.org.uk
Global Action Plan.
www.iclei.org
ICLEI.
www.transitiontowns.org
Transition towns.

Epidemiological Transition

K. McCracken Macquarie University, Sydney, Australia
D. R. Phillips, Lingnan University, Tuen Mun, Hong Kong

Glossary

Acute Disease A disease of rapid onset and usually of brief duration. The term carries no connotations about the severity of a disease (cf. chronic disease).

Chronic Disease A disease that persists for a long time. The term carries no connotations about the severity of a disease (cf. acute disease).

Compression of Morbidity The proposition that, as mortality declines, ill health and disability will be compressed into the later years of life (due to improving living standards and medical and public health advances).

Degenerative Disease A disease that involves the biological deterioration of the body (heart disease, some cancers, rheumatoid arthritis, etc.), usually, but not always, associated with aging – sometimes referred to as noncommunicable disease (cf. infectious disease).

Epidemic A mass outbreak of a disease in a particular geographic area which spreads and then disappears relatively quickly.

Epidemiological Polarization The coexistence of 'old' (pre-transition) and 'new' (post-transition) diseases in a country.

Epidemiological Transition The shift in the major causes of death from infectious diseases (e.g., tuberculosis, diarrhea and enteritis, and pneumonias) and also famine to chronic and degenerative ailments (e.g., heart disease, stroke, and cancers) associated with the long-term decline in mortality from high to low levels.

Infectious (Communicable) Diseases Diseases caused by pathogens (e.g., bacteria, viruses, protozoa, and endoparasites). Infectious diseases which are spread by direct contact with infected persons are often termed contagious (e.g., sexually transmitted diseases and ebola), although some writers use contagious more loosely as a synonym for all infectious diseases.

Pandemic An epidemic that spreads over a very wide area, which can be continent-wide or worldwide (e.g., the 1918–19 Spanish Flu; the mid-fourteenth-century bubonic plague outbreak across Europe).

Introduction

The term 'epidemiologic(al) transition' refers to the shift during the decline in mortality from high to low levels, from an infectious disease dominated cause of death profile (e.g., tuberculosis, smallpox, typhus, diarrhea, and dysentery) to one in which chronic degenerative ailments (e.g., heart disease, stroke, and cancer) predominate (**Tables 1** and **2**). The term was coined by Abdel Omran in a 1971 paper that by dint of the research it has subsequently generated has come to be regarded as a public health classic. While Omran is deservedly credited with the foundation statement on the transition, the notion of such a transition though can be found in earlier work. For example, a 1962 United Nations study demonstrated how cause of death patterns change with changing life expectancy, while a 1969 paper by Frederiksen developed a typology outlining 'predominant patterns of disease, mortality, and fertility' in association with 'state of society'. In turn, the initial concept and framework enunciated by Omran have been both substantially extended and widened. Today, the epidemiological transition is recognized as part of a wider 'health transition' and one strand of a suite of interconnected other transitions (**Figure 1**) – in fertility, age structure, nutrition, urbanization, technology, and social and economic development.

Omran's Original Formulation of Epidemiological Transition Theory

The 'theory' of epidemiological transition was described by Omran in 1971 as conceptually focusing on "the complex change in patterns of health and disease *and* on the interactions between these patterns and their demographic, economic and sociologic determinants and consequences." Five basic propositions were advanced:

1. that mortality is a fundamental factor in population dynamics;
2. that during the transition, pandemics of infection are gradually displaced by degenerative and man made diseases as the chief form of morbidity and primary cause of death;
3. that the most profound changes in health and disease patterns during the transition occur among children and young women;
4. that the shifts in health and disease patterns that characterize the transition are closely associated with the demographic and socioeconomic transitions that are part and parcel of modernization; and

Table 1 Epidemiological transition in (a) Sydney, Australia and (b) Mexico

Cause of death	*% of death*	*Cause of death*	*% of death*	*Cause of death*	*% of death*
(a) Sydney, Australia	*1901*		*1947*		*2000 02*
Respiratory diseases	14.4	Diseases of the heart	33.6	Malignant neoplasms (cancer)	27.2
Diarrhea and enteritis	11.0	Cancer	13.6	Heart disease	26.0
Tuberculosis	10.4	Cerebrovascular disease (stroke)	11.2	Cerebrovascular disease (stroke)	10.7
Organic heart disease	6.6	Diseases of respiratory system	5.8	Diseases of respiratory system	8.8
Cancer	5.4	Accidents	4.4	Accidents and adverse effects	3.4
(b) Mexico	*1955 57*		*1980*		*2002*
Gastroenteritis	17.5	Heart diseases	11.6	Heart diseases	13.9
Influenza and pneumonia	15.5	Accidents	11.0	Malignant neoplasms (cancer)	12.7
Childhood diseases	10.4	Influenza and pneumonia	8.8	Diabetes Mellitus	11.9
Heart diseases	7.1	Enteritis and other diarrheal diseases	8.6	Unintentional injuries	8.1
Malaria	5.1	Malignant tumours (cancer)	6.1	Cerebrovascular disease (stroke)	5.7

Sources: Sydney McCracken, K. and Curson, P. (2000). In sickness and in health: Sydney past and present. In Connell, J. (ed.) *Sydney: The Emergence of a World City*, pp 96 118; 352 356. Melbourne: Oxford University Press; Australian Bureau of Statistics, Unpublished data. Mexico Frenk, J., Bobadilla, J. L., Sepulveda, J. and Cervantes, M. L. (1989). Health transition in middle income countries: New challenges for health care. *Health Policy and Planning* 4(1), 29 39; World Health Organization, Burden of Disease Project (Web access).

Table 2 Leading cause of death profiles, selected WHO member states at different stages of epidemiological transition, 2002

Cause of deaths	*% of deaths*	*Cause of deaths*	*% of deaths*	*Cause of deaths*	*% of deaths*
Pre to early transition					
Ethiopia		*Niger*		*Zambia*	
Lower respiratory infections	14.4	Lower respiratory infections	14.0	HIV/AIDS	41.6
HIV/AIDS	11.5	Diarrheal diseases	10.3	Lower respiratory infections	11.6
Diarrheal diseases	6.0	Malaria	9.6	Malaria	8.3
Tuberculosis	3.9	Measles	5.5	Diarrheal diseases	6.7
Measles	3.7	Low birth weight	4.8	Tuberculosis	3.1
Mid transition					
Bangladesh		*East Timor*		*Yemen*	
Ischemic heart disease	11.7	Tuberculosis	9.8	Lower respiratory infections	14.2
Lower respiratory infections	11.2	Ischemic heart disease	9.3	Diarrheal diseases	11.1
Tuberculosis	6.7	Lower respiratory infections	9.0	Ischemic heart disease	9.5
Diarrheal diseases	6.2	Cerebrovascular disease (stroke)	4.6	Low birth weight	5.5
Cerebrovascular disease (stroke)	5.8	Meningitis	2.9	Road traffic accidents	4.4
Late transition					
Japan		*Russian Federation*		*United States of America*	
Cerebrovascular disease (stroke)	13.8	Ischemic heart disease	29.6	Ischemic heart disease	21.3
Ischemic heart disease	9.6	Cerebrovascular disease (stroke)	22.2	Cerebrovascular disease (stroke)	6.8
Lower respiratory infections	9.4	Poisonings	2.8	Trachea, bronchus, lung cancers	6.5
Trachea, bronchus, lung cancers	5.8	Self-inflicted injuries	2.5	Chronic obstructive pulmonary disease	5.3
Stomach cancer	5.2	Trachea, bronchus, lung cancers	2.4	Alzheimer and other dementias	3.8

Source: World Health Organization, Burden of Disease Project (Web access).

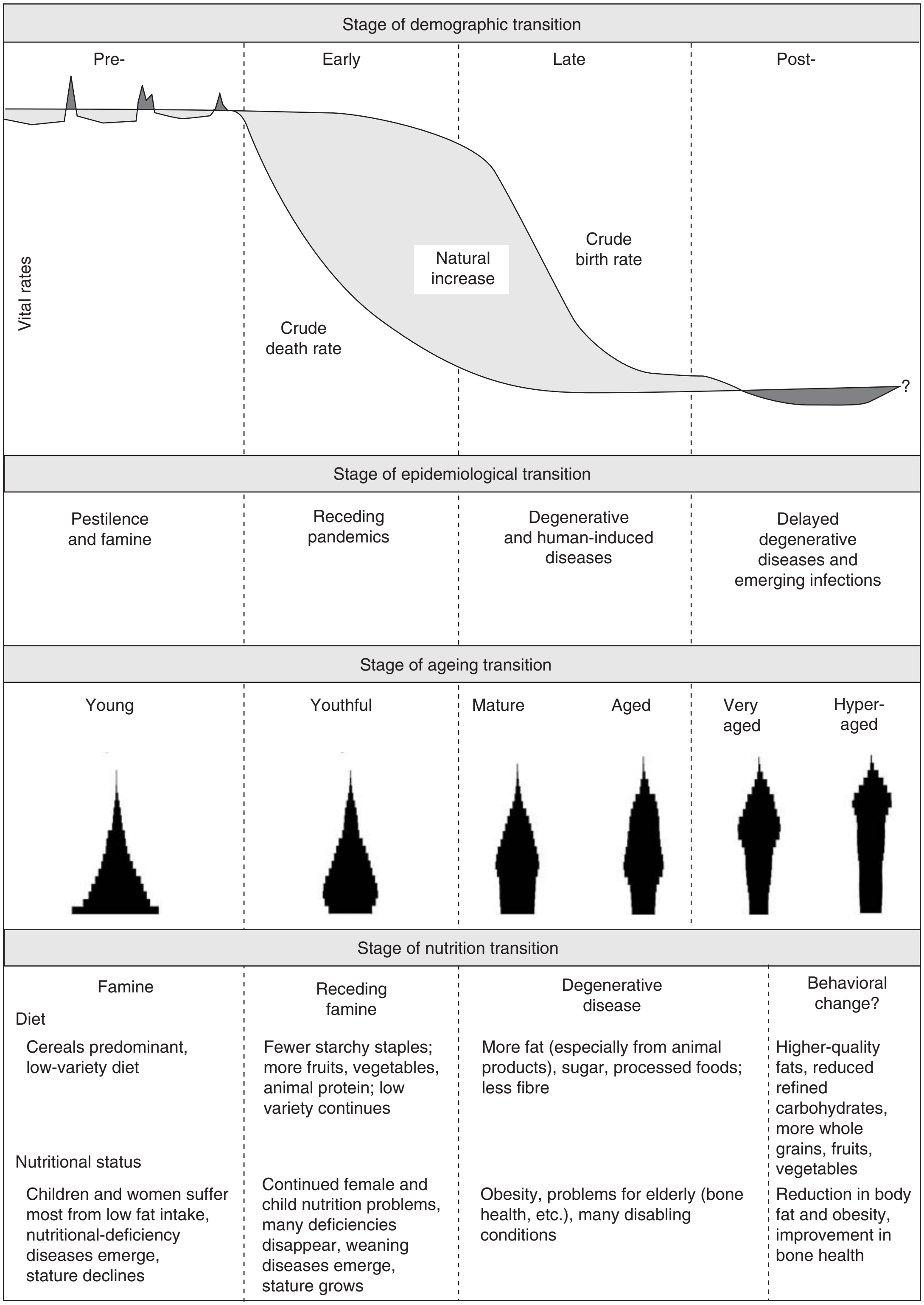

Figure 1 The demographic, epidemiological, aging and nutrition transitions. Sources: McCracken, K. and Phillips, D. R. (2005). International demographic transitions. In Andrews, G. J. & Phillips, D. R. (eds.) *Ageing and Place*, pp 36 60. London: Routledge.; Popkin, B. M. (2006). Global nutrition dynamics: The world is shifting rapidly toward a diet linked with noncommunicable diseases. *American Journal of Clinical Nutrition* 84, 289 298.

5. that variations in the pattern, pace, determinants, and consequences of population change differentiate three basic models of epidemiological transition.

Stages of Epidemiological Transition

The shift in diseases referred to in proposition 2 was argued to typically involve three major successive stages:

- Stage 1. 'The age of pestilence and famine', the era before the transition, when mortality was high and fluctuating and dominated, as its name was intended to evoke, by the Malthusian 'positive checks' of infectious disease, famine, and wars.
- Stage 2. 'The age of receding epidemics', "when mortality declines progressively and the rate of decline accelerates as epidemic peaks become less frequent or disappear". With the decline of infectious diseases the average age at death rises and degenerative diseases become more important.
- Stage 3. 'The age of degenerative and man made diseases', when mortality continues to decline and the move from infectious to degenerative disease predominance is completed.

Models of Transition

Three basic models of epidemiological transition were identified, although other variants have also been proposed as noted below:

1. The 'classical (Western) model' in which the shift in total and cause specific mortality patterns occurred gradually and progressively over a century or more. Most Western European societies were seen as typifying this pattern.
2. The 'accelerated model' in which the fall in the crude death rate and shift from an infectious to degenerative disease dominated mortality profile occurred over a far shorter period of time. Japan was cited as especially exemplifying this model.
3. The 'contemporary (or delayed) model', characterizing the "relatively recent and yet to be completed transition of most developing countries", Chile and Sri Lanka (Ceylon) being chosen as representative examples.

Disease Determinants

In turn, three major categories of disease determinants were outlined – (1) 'ecobiologic'; (2) 'socioeconomic, political, and cultural'; and (3) 'medical and public health' – and ascribed varying importance in the respective models. The transition from infectious to degenerative disease predominance in the classical (Western) model was attributed primarily to ecobiologic and socioeconomic factors with medical factors "largely inadvertent until the twentieth century, by which time pandemics of infection had already receded significantly." In the case of Japan and other countries fitting the accelerated model, the mortality revolution was seen as stemming from 'sanitary and medical advances' along with 'general social improvements'. Imported medical technology and public health measures were meanwhile deemed paramount in the delayed model.

To summarize, the picture painted was a linear one of more or less uninterrupted progression from stage 1 through to stage 3 in which degenerative and man made diseases displaced infections as the leading causes of mortality and life expectancy at birth rose from low vacillating levels in the 20–40 years range to over 50 years. Whether fully intended or not, the image conveyed was of a finite process, with stage 3 seeming to be the endpoint of human epidemiological evolution. Similarly, implicit was the idea of the transition as a universal process, an epidemiological destiny that all countries would ultimately experience.

Omran subsequently produced further papers on epidemiological transition, some on the transition in general, others applying the theory to the United States. The only real elaborations upon the original formulation though were, one, to define the vaguely labeled 'man made diseases' (better termed 'human induced') as things such as "radiation injury, accidents, occupational hazards, carcinogens in the environment and in industry, and food additives" along with mental illness and drug dependency and, two, to add a 'transitional variant of the delayed model'. This additional model was to differentiate those developing countries that had experienced rapid mortality decline in the 1940s, as per the delayed model, and which over ensuing decades had maintained the momentum of decline (in contrast to the slackening off found in many parts of the developing world) and also registered substantial fertility reduction.

Critiques and Extensions of the Theory

The validity and universality of epidemiological transition as a concept in general and Omran's theory in particular have been widely debated across a range of disciplines including geography, demography, epidemiology, public health, medical and economic history, and health planning. Aspects that have attracted particular attention have been the notion of unidirectional sequential change, the seeming end state accorded the third stage and its implied vanquishing of infectious disease, the relative importance of different disease determinants, the limited attention given to morbidity, and the broad social and spatial brush of the model.

Linear Progression?

The impression conveyed by Omran's model is of smooth, uninterrupted advancement through the transition, the three eras occurring in clear, by implication, one way sequence. Subsequent empirical research, however, has largely invalidated this conceptualization, with demonstration that epidemiological reversal ('counter transition') can equally occur. The life expectancy declines associated with the African HIV/AIDS epidemic (HIV, human immunodeficiency virus; AIDS, acquired immune deficiency syndrome) are perhaps the most powerful illustration of this and the falsity of any notion of seemingly inevitable epidemiological improvement and progression through Omran's stages. Zimbabwean life expectancy at birth for instance is estimated by the UN to have fallen from around 60 years in the mid 1980s to 37 at the end of the century; Botswana's from 63 to 48; and Zambia's from 51 to 34. A corollary of the AIDS epidemic has been a resurgence of tuberculosis morbidity and mortality following from the increased prevalence of immune system impairment. From having advanced to varying points within stage 2, these countries have been catapulted back to the pretransitional 'pestilential' era.

Another well documented example of counter transition life expectancy decline is the experience of several of the former Soviet Union republics where the economic collapse, social disruption, and political upheaval following the dissolution of the USSR caused death rates, particularly among males, to soar. Similar reversal has also occurred among the United States inner city young black male population where a deadly epidemiological cocktail of violence, accidents, and HIV has sometimes emerged.

Further Stages of Epidemiological Transition?

The suggested finality of stage 3 in Omran's model has subsequently also been proven premature by continuing epidemiological evolution in developed countries. This evolution has seen a very large decline in cardiovascular disease mortality, a pronounced delaying of deaths from the leading degenerative diseases (heart disease, cancer, and stroke) to later ages (the 'rectangularization of the survival curve'), the rise of 'social pathology' deaths, the dual emergence of new infectious diseases (e.g., HIV/AIDS and hepatitis C), and the resurgence of old ones (such as tuberculosis, in drug and multidrug resistant forms). Also, as the bottom panel in **Table 2** indicates, while 'late stage' epidemiological transition countries all share degenerative disease dominated mortality profiles, important variations have evolved beneath that broad commonality. A critical component of modern day transition is the morbidity, as well as mortality, from psychosocial conditions, such as depression and stress, with associated physical ill health, sometimes suicides or deliberate self harm, and loss of quality of life for individuals and families. Moreover, the often noted modern 'epidemic' of dementias associated with aging populations is crucial as these dementias, which have various underlying causes, result in huge morbidity and the need for care, place burdens on families and communities, and often lead to the death of the individuals.

The increasing incidence and prevalence of many of these conditions have seen the transition model expanded to incorporate a fourth stage. This new era has been termed the 'age of delayed degenerative diseases', and some have suggested the label 'hybristic stage', to reflect the increasing influence of social pathological and life style related mortality (such as substance abuse, cirrhosis of the liver, suicide, homicide, and accidents).

The development of this modern and emerging phase has varied both socially between groups in the population and geographically across regions and within urban areas. Lower socioeconomic groups, for instance, have often not enjoyed such large declines in heart disease and stroke as those of higher status. Likewise, hybristic social pathology mortality displays marked racial group differentiation in the United States, Australia, and many other developed and middle income countries. For example, in the United States, accidents, assault (homicide), and HIV accounted for 14.1% of black male deaths in 2000, more than twice the equivalent proportion in the white male population.

Yet, further stages of possible epidemiological evolution have been touted. For example, a distinct new stage centered around the continuing re emergence of infectious and parasitic diseases has been suggested. Meanwhile, more recently speculation has been voiced about rising childhood obesity rates leading to an era of rising nutritional disorders and declining life expectancy in developed nations. Obesity is also becoming a major problem in developing nations in earlier stages of the transition and may slow down life expectancy progress there below what might have previously been expected. Research on the nutrition transition by Popkin and others offers many valuable insights in this area (**Figure 1**). In the light of developments over the past three to four decades, it would thus seem unwise to think of epidemiological transition as ever being complete.

The Conquest of Infectious Disease?

The move through the transition to an era in which infectious and parasitic disease has been essentially vanquished as a force in mortality has not transpired to the extent heralded by Omran. Indeed, the WHO has pointed out that numerous 'new' infectious diseases, or variants on known infections, are regularly discovered. Moreover, a million people a year, mainly children, die from malaria and almost 2 million from tuberculosis; the majority in poor countries, and these two conditions alone cause many times more episodes of acute and chronic ill health. Certainly, degenerative conditions have replaced infections as the predominant killers in the more developed nations, but the optimistic 1960s pro nouncements of the era of epidemic disease coming to an end even in developed nations have proved seriously premature. Readily recognized and dramatic reminders are the HIV/AIDS epidemic and the emergence of drug resistant tuberculosis and pneumonia. Similarly, today, many parts of the world are braced for what the WHO and some others expect to be a pandemic of avian influenza if the H5N1 virus mutates into a form easily transmissible between humans. But also important is the continuing, though often overlooked, contribution of less spectacular infections to present day developed and especially developing world mortality.

The International Classification of Diseases (ICD) in advertently potentially encourages this underestimation. Over its various revisions, the first chapter of the classi fication has been 'infectious and parasitic diseases', and commentators have frequently taken deaths in that chapter (now down to close to 1% of total deaths in most developed nations) as representing virtual elimination of infectious diseases as killers. However, buried away in the classification are other deaths with infectious etiologies; those directly identified, such as influenza and pneu monia, but also numerous others – for example, liver cancer deaths resulting from hepatitis B and C, cervical cancer deaths due to human papillomavirus (HPV), some genito urinary conditions, rheumatic heart disease, and the like. Apart from this hidden side of infectious disease, there is the 'associated' role infectious conditions (such as influenza, pneumonia, and septicemia) play in many deaths and many drug resistant forms of infections can be transmitted within hospitals. Also overhanging all nations, more developed and less developed alike, are the threats of future pandemics of influenza (avian or other types) and other infectious outbreaks, along with the possibility of bioterrorism events.

Explaining the Classical (Western) Transition

The classical (Western) model transition from infectious to degenerative disease predominance has been ascribed as primarily due to ecobiologic and socioeconomic factors, rating medical factors as of scant importance (i.e., 'largely inadvertent until the twentieth century'). This view was also very powerfully argued, both before and after Omran's initial paper, in the writings of Thomas McKeown, with improvements in nutrition specifically nominated as the key factor. These claims have sub sequently become probably the most hotly contested aspect of the transition theory and the orthodoxy now is that both Omran and McKeown failed to give medical factors due recognition. Both authors were probably correct in that curative medical advances played rela tively little part in reducing mortality until well into the twentieth century, but in viewing 'medical factors' in that narrow curative sense they are now generally seen as having undervalued the role of public health inter ventions – sewerage and drainage reforms, improvements in water and milk supplies, the removal of waste, laws against overcrowding, etc., although some argue that there is little evidence that nutrition improved.

The Place of Morbidity?

Although proposition 2 of Omran's thesis speaks of both morbidity and death and elsewhere reference is made to 'health and disease' patterns, in fact very little attention was given to the morbidity side of the transition or to health in a wider sense than as indicated by changing cause of death patterns. This has left his theory open to the charge of being 'insufficiently epidemiological' and the broadening into a more genuine 'health transition' model. Riley and Alter were among the first to extend Omran's formulation to incorporate morbidity behavior during the transition, importantly showing that 'mortality cannot be used as a proxy for morbidity'. While the mortality dimension of the transition involves gains in life expectancy, and in that sense health improvement and certainly quality of life, they made the point that the associated morbidity transition "is characterised by a shift from acute to chronic conditions, which can bring about an overall increase in the prevalence of poor health." Riley and Alter saw their paper as contributing to a more complete theory of epidemiological transition and retained that descriptive title. Working around the same time however, Caldwell and colleagues introduced the term 'health transition' to widen the focus from just death and illness to health in its full positive as well as negative sense and to the cultural, social, and behavioral de terminants of changing health. The attention drawn to cultural, social, and behavioral determinants was a de liberate attempt to highlight factors other than medical interventions and income.

One aspect of this widening of focus has been the development of the concept of 'healthy life expectancy' (or 'health adjusted life expectancy' – HALE), that is, the average number of healthy years of life remaining to a person of a given age in a particular population. This concept broadens the conventional mortality based measure of life expectancy to take into account the years of life free of disability. HALE is a conceptually more attractive and informative measure of health status through incorporating both morbidity and mortality, but caution needs to be exercised in interpreting published HALE figures for many less developed nations due to data availability and quality problems. For example, the

WHO's estimate of HALE at birth (years) for Afghani stan in 2002 was 35.3, with an uncertainty interval of 26.7–40.4; for North Korea, 58.0 and 50.9–64.2; and for East Timor, 47.9 and 40.8–53.6.

A Compression of Morbidity?

With the widening of transition theory to incorporate morbidity has also come considerable debate about the quality of health over the years of additional life ex pectancy. Are the extended years of life principally years of good health, or simply extra years of poor health as postulated by Verbrugge in 1984 in a classic paper? The two scenarios have respectively become known as the 'compression' and 'expansion of morbidity' theses. The compression thesis, the optimistic perspective, argues that as mortality declines ill health and disability will be increasingly compressed into the later years of life due to improving living standards and medical and public health advances. The alternative view, the expansion scenario, argues instead that, while death rates from many chronic diseases have fallen, their incidence and prevalence remain largely unchanged, leaving increasing numbers of unhealthy survivors. Unfortunately, detailed data internationally are often too scanty to give definitive support to either scenario although HALE Project data from Europe suggest that, while European populations are aging, the proportions of older persons with disability are decreasing. These findings differ across various countries suggesting that dynamics of functioning may differ across cultures and indicating the need for research to clarify which potentially modifiable and culturally determined factors protect against functional decline.

In specific cases, especially in the developed world, evidence is rapidly mounting for the compression thesis and changes (mainly decreases) in disability have been noted in countries such as the USA in specific groups. There is nevertheless debate about methodological issues and data interpretation, suggesting that declines in morbidity may be more gradual than some analyses suggest. Fries, in 2003, revisits his 1980 compression of morbidity paradigm, and notes USA data, which show morbidity decline has been about 2% per year, con trasted with mortality rate decline of about 1% per year. This supports a compression of morbidity argument, at any rate in the United States, at the population level.

Perhaps the most important feature is that longi tudinal research now strongly links good health habits with long term reductions in cumulative lifetime dis ability. Fries notes that people with few behavioral health risks have only one fourth the disability of those who have more risk factors and it appears that the onset of disability is postponed from 7 to 12 years, far more than any increases in longevity in the respective groups. A crucial thing for policy, which links compression of morbidity to aspects of epidemiological transition, is that rigorous research on health enhancement programs in older populations shows a combination of reduction in health risks, improved health status, and (key for policy) decreased medical care utilization. This means that current health policy initiatives might help increase and consolidate health gains, especially for older people and probably for all age groups. The exact role of lifestyle/ behavioral factors in compression of morbidity is nevertheless blurred; while behavior such as smoking has decreased and there has been improved screening for hypertension and some types of cancers, at the popu lation level their health gains might be offset by, for example, increases in obesity and, in later life, dementias, which are more or less worldwide phenomena. Another important practical and policy issue is that some argue the current improved trends in mortality are unlikely to continue without major breakthroughs in the basic sci ence of aging. Others have a more positive view and argue that mortality experts who have asserted that life expectancy is close to an ultimate ceiling have repeatedly been proven wrong. They suggest that, for 160 years, best performance life expectancy has steadily increased by a quarter of a year per year, an extraordinarily con sistent human achievement. In any case, compression of morbidity does seem to be happening (in the USA and presumably in many other developed economies), and this has important policy and financial implications, especially in aging populations, as noted below.

Intra-national Epidemiological Variety

The picture drawn in the original statement of the epidemiological transition was essentially one of nation states at distinct epidemiological stages. Some recog nition that underneath the national aggregate pattern different groups in a population might experience epi demiological change differently was shown in proposition 3's claim that children and young women experience the most profound changes in health and disease patterns during the transition. Later, in work on the transition in the United States, Omran extended the discussion of differentials to include race, outlining how the transition "among the non white population of the U.S. was slower than among the whites." Beyond that, however, sub national epidemiological heterogeneity was not considered.

Such heterogeneity, however, is intrinsic to the notion of transition, the passage from one epidemiological stage to another necessarily involving a period of time when different types of diseases coexist strongly in a popu lation. Cumulative cause of death graphs in some of Omran's papers in fact clearly show that overlap stage in England and Wales and the United States, but the emphasis in those papers was on succeeding stages, not overlap.

However, this overlap has become of major signifi cance to many present day less developed countries with the coexistence becoming entrenched. As described by Frenk *et al.*, a new model of epidemiological transition is occurring where "the changes do not fully take place and where different types of disease coexist in the same population." They labeled this as 'protracted epidemi ological transition'. Countries in such overlap thus bear a double burden of disease – old pre transition style dis eases and new 'modern' health problems. In many cases this has led to 'epidemiological polarization', where dif ferent social classes and geographic locations have dis tinct epidemiological profiles. Richer groups and locations tend to be characterized by 'modern' disease profiles, while poorer groups and places often suffer more from old style diseases. This shows in many countries in a rural (old style dominance)–urban (new style domin ance) cleavage. Significant polarization though is also frequently found within cities. **Table 3** presents the ex ample of Cape Town, South Africa, highlighting the very different cause of death profiles of the poor Cape Flats district of Khayelitsha and the affluent South Pen insula area. Caldwell notes that this is also happening in many societies of the Eastern Mediterranean Region where the less affluent sectors have a high incidence of infectious diseases, especially among children, while the wealthier sectors show completely different patterns of illness. A focus on national averages thus masks the different 'epidemiological worlds' citizens of the same country can inhabit, doing a disservice to virtually all people. Indeed, the WHO's Commission on Social Determinants of Health is paying special attention to inequalities in health between and within countries.

Epidemiological Transition and Health Planning

Can a broad and long term model or concept such as epidemiological transition have any use for health and social care planning? Certainly in the 'short term' this utility is limited, but there are certainly some important implications for longer term policy that can be drawn. In some ways, various policy implications of epidemi ological transition can be clearer in the developing world, where rapid transition is often occurring and where the overlapping stages of transition prove great challenges to healthcare and other human services planning. Take for example some sub Saharan countries which are facing the epidemiological challenges of the earlier types, in volving malnutrition or food shortage, infectious diseases (such as HIV/AIDS), multiple drug resistant tubercu losis and malaria, yet at the same time are facing epi demics of cancers and circulatory diseases, as well as growing populations of older people with new long term care requirements. These countries have to plan for a protracted period in which they will have to face this mixed transition, with huge implications for personnel and facility provision, as well as for their health tech nology and essential drugs needs.

Today, in many countries, much of the future need for public taxation and retirement savings is predicated on the 'moral panic' scenario, which predicts an avalanche of older populations, by implication disabled, dependent, and costly to care for. However, if we look forward with knowledge of epidemiological transition, forewarned may be forearmed. Current trends in morbidity and mortality data suggest that we are doing something right. People are often living longer and, it seems, sometimes in healthier states. This will potentially have huge policy impacts. This is even if the most important policy has not yet systematically been implemented in many health systems – the postponement of disability through primary prevention. Future policy must therefore focus on ways to reduce morbidity from mainly nonfatal (but potentially fatal) diseases such as osteoporosis, rheuma toid arthritis, cognitive disorders (such as depression), and then dementias, especially Alzheimer's disease. It is clear that long term spending on public (and private) healthcare and associated long term care may very well be affected by trends in epidemiological transition.

Late stage 'emerging' epidemiological transition in particular therefore highlights how we should be looking at the conditions that will be the major challenges of tomorrow. Implicitly, of course, where there is profit to be made, this is already being done by businesses.

Table 3 Leading causes of death (%) in Cape Town and selected subdistricts of Cape Town, 2001

Cape Town		*Cape Town subdistricts*			
		Khayelitsha (poor)		*South Peninsula (affluent)*	
Cause of death (%)					
Homicide	(10.5)	Homicide	(18.7)	Ischemic heart disease	(12.8)
Ischemic heart disease	(8.1)	HIV/AIDS	(15.0)	Hypertensive heart disease	(7.8)
HIV/AIDS	(7.4)	Tuberculosis	(14.9)	Diabetes Mellitus	(6.9)
Hypertensive heart disease	(6.4)	Lower respiratory infections	(5.2)	Homicide	(5.7)
Tuberculosis	(5.9)	Road traffic accidents	(4.9)	Stroke	(5.2)

Source: Groenewald, P., Bradshaw, D., Nojilana, B. *et al*. (2003). *Cape Town Mortality, 2001.* (*Part I, Cause of Death and Premature Mortality; Part III, Cause of Death Profiles for Each Sub District*.) Cape Town: City of Cape Town, South African Medical Research Council, University of Cape Town.

We see many examples in which pharmaceutical com panies are looking for the 'holy grails' of dementia treatments and those for cancers, and health technology researchers and corporations are stepping up the devel opment of early detection equipment, especially of the safe and noninvasive kind. By contrast, nonprofitable treatments and prophylaxis for poorer country conditions such as malaria are often overlooked by big business.

In particular, the area of future health and welfare personnel needs and training is important. Staffing re quirements in various medical and welfare specialisms will change over time. More important, perhaps, is the fact that today's doctors and nurses often receive rather limited initial training in the needs of aging populations and the detection and management of many common chronic aliments, especially psychosocial and cognitive impairments which cause huge medical and social care workloads. Their needs are often as much social and supportive as purely clinical. Medical, nursing, and social work syllabi therefore can sensibly be reviewed in the light of the conditions that today's students are likely to meet when they are working 10, 20, or 30 years later.

In summary, the concept of epidemiological transition does seem to have utility and importance if used as a long term planning advisory framework. It is by no means an infallible framework, but used judiciously it is almost an essential component of long term health and welfare planning.

See also: Demography; Diffusion; Disease Diffusion; Disease Mapping; Health Geography; Medical Geography.

Further Reading

Caldwell, J. C. (2001). Population health in transition. *Bulletin of the World Health Organization* 79(2), 159 160.

Caldwell, J. C., Findley, S., Caldwell, P. *et al*. (eds.) (1990). *What We Know About Health Transition: The Cultural, Social and Behavioural Determinants of Health*, 2 vols. Canberra: Australian National University Press.

Frederiksen, H. (1969). Feedbacks in economic and demographic transition. *Science* 166, 837 847.

Frenk, J., Bobadilla, J. L., Sepulveda, J. and Cervantes, M. L. (1989). Health transition in middle income countries: New challenges for health care. *Health Policy and Planning* 4(1), 29 39.

Fries, J. F. (1980). Aging, natural death, and the compression of morbidity. *New England Journal of Medicine* 303, 130 136.

Fries, J. F. (2003). Measuring and monitoring success in compressing morbidity. *Annals of Internal Medicine* 139, 455 459.

Gribble, J. N. and Preston, S. H. (1993). *The Epidemiological Transition: Policy and Planning Implications for Developing Countries*. Washington, DC: National Academy Press.

Groenewald, P., Bradshaw, D., Nojilana, B. *et al.* (2003). *Cape Town Mortality, 2001*. (*Part I, Cause of Death and Premature Mortality; Part III, Cause of Death Profiles for Each Sub District*.) Cape Town: City of Cape Town, South African Medical Research Council, University of Cape Town.

Jamison, D. T., Breman, J. G., Measham, A. R. *et al*. (2006). *Disease Control Priorities in Developing Countries* (2nd edn.). New York: Oxford University Press and World Bank.

Manton, K. G., Gu, X. and Lamb, V. L. (2006). Long term trends in life expectancy and active life expectancy in the United States. *Population and Development Review* 32(1), 81 106.

Marmot, M. (2005). Social determinants of health inequalities. *The Lancet* 365, 1099 1104.

McCracken, K. and Curson, P. (2000). In sickness and in health: Sydney past and present. In Connell, J. (ed.) *Sydney: The Emergence of a World City*, pp 96 118; 352 356. Melbourne: Oxford University Press.

McCracken, K. and Phillips, D. R. (2005). International demographic transitions. In Andrews, G. J. & Phillips, D. R. (eds.) *Ageing and Place*, pp 36 60. London: Routledge.

McKeown, T. (1976). *The Modern Rise of Population*. London: Edward Arnold.

Olshansky, S. J. and Ault, A. B. (1986). The fourth stage of the epidemiologic transition: The age of delayed degenerative diseases. *The Milbank Quarterly* 64(3), 355 391.

Olshansky, J., Carnes, B., Rogers, R. G. and Smith, L. (1997). Infectious diseases new and ancient threats to world health. *Population Bulletin* 52(2), 52.

Olshansky, S. J., Passaro, D. J., Hershow, R. C. *et al*. (2005). A potential decline in life expectancy in the United States in the 21st century. *New England Journal of Medicine* 352(11), 1138 1145.

Omran, A. R. (1971). The epidemiologic transition: A theory of the epidemiology of population change. *Milbank Memorial Fund Quarterly* 49(4), 509 538.

Omran, A. R. (1977). Epidemiologic transition in the US. *Population Bulletin* 32(2), 42.

Omran, A. R. (1983). The epidemiologic transition theory: A preliminary update. *Journal of Tropical Pediatrics* 29, 305 316.

Phillips, D. R. (1990). *Health and Health Care in the Third World*. London: Longman.

Phillips, D. R. (1994). Does epidemiological transition have utility for health planners? *Social Science and Medicine* 38(10), vii x.

Phillips, D. R. and Verhasselt, Y. (eds.) (1994). *Health and Development*. London: Routledge.

Popkin, B. M. (2002). An overview on the nutrition transition and its health implications: The Bellagio meeting. *Public Health Nutrition* 5(1A), 93 103.

Popkin, B. M. (2006). Global nutrition dynamics: The world is shifting rapidly toward a diet linked with noncommunicable diseases. *American Journal of Clinical Nutrition* 84, 289 298.

Riley, J. C. and Alter, G. (1989). The epidemiologic transition and morbidity. *Annales de Demographie Historique* 199 213.

Rogers, R. G. and Hackenberg, R. (1987). Extending epidemiologic transition theory: A new stage. *Social Biology* 34(3 4), 234 243.

Smallman Raynor, M. and Phillips, D. (1999). Late stages of epidemiological transition: Health status in the developed world. *Health and Place* 5, 209 222.

United Nations (1963). *Population Bulletin of the United Nations, No. 6: The Situation and Recent Trends of Mortality in the World*. New York: United Nations.

Verbrugge, L. (1984). Longer life but worsening health? Trends in health and mortality of middle aged and older persons. *Milbank Memorial Fund Quarterly* 62(3), 475 519.

Wolf, D. A., Hunt, K. and Knickman, J. (2005). Perspectives on the recent decline in disability at older ages. *Milbank Quarterly* 83(3), 365 395.

Relevant Websites

http://www.cdc.gov/
Centers for Disease Control and Prevention.

http://reves.site.ined.fr/en/
REVES Network on Health Expectancy.

http://www.who.int/healthinfo/bodproject/en/index.html
The World Health Report; Burden of Disease Project.

http://www.who.int/social determinants/en/
WHO Commission on Social Determinants of Health.

http://www.who.int/classifications/icd/en/
WHO International Classifications of Diseases (ICD).

Equity

J. Wiles, University of Auckland, Auckland, New Zealand
A. Kobayashi, Queen's University, Kingston, ON, Canada

What Is Equity?

Equity is about social justice and fairness, and reflects a concern to reduce systematic discrimination and mar ginalization. Some have described equity as the absence of systemic disparities (e.g., in health outcomes, in major social determinants of health, access to health, access to employment, employment outcomes) between social groups who have different levels of underlying social advantage or disadvantage. For example, groups may be socially disadvantaged by virtue of being poor, female, disabled, ill, or a member of a racialized or minoritized ethnic, cultural, religious, or age group. The ethical principle of equity is that selective priority should be given to improving the situation of the most socially disadvantaged groups in a society, which may mean the readjustment of strict rules or structures, or the redistribution of the resources and other processes that drive a system of inequality. Others have described equity as inclusiveness and the affirmation of opportun ities for all on the grounds that there is no such thing as a 'level playing field'. Still others associate equity with principles of human rights. The difficulty with defining equity is that specific definitions have implications for the conceptually rigorous measurement and monitoring of, and accountability for, equity.

Equity is not synonymous with equality. Whereas equality is generally about fairness defined as sameness, equity is about addressing unfair difference. In a human rights context, 'formal equality' refers to the principle of treating everyone equally, regardless of the outcomes. 'Substantive equity', in contrast, refers to the principle of achieving a fair outcome. In order to achieve substantive equity, it is necessary to take positive measures to affect outcomes.

Even so equality is to some extent value based, because people are not the same, and all treatment is based on some socially defined and normative choice concerning the standard for sameness, especially when that standard is defined according to the able bodied, white, middle class male. In a typical city, for example, everyone may have equal right of access to a service or employment, but they are not the same because a host of barriers (poverty, lack of transportation, racial discrimination) may prevent people from actually gaining access.

In employing these different terms with respect to health for example, it is a value based decision to decide when the more frequently used term 'health inequality' becomes a health inequity. Margaret Whitehead dis tinguished between health inequalities as those differ ences relating to physical variation and freely chosen health damaging behavior, and health inequities as dif ferentials due to health damaging behaviors not based on informed choices (such as exposure to unhealthy living or working conditions, and inadequate access to health and social services). Health inequalities are either un avoidable or 'fair', whereas inequities are judged to be avoidable and unfair. Of course the difficulty in this is establishing agreement on precisely which differences in health outcomes and service uses are avoidable, to what degree they are so, and whether this is the same for everyone. As Whitehead concludes, equity in health "is concerned with creating equal opportunities for health, and with bringing health differentials down to the lowest levels possible."

As **Table 1** suggests, while equity and equality are not synonymous, they are complexly related. Equity might be considered a specific or even ethical form of equality (for example relating to the third, fourth, and fifth of Torstein Eckhoff's principles of equality). Equity implies something more than the fairly straightforward sameness implied in Eckhoff's first and second principles of equality. Eckhoff's third and fourth principles refer to equality of outcomes, which are more easily measured

Table 1 Eckhoff's principles of equality

1	Objective equality	Equal amounts to each recipient (traditional equality)
2	Relative equality	The 'equity rule', equality of outcome/input ratios, or equality relative to individual contributions
3	Subjective equality	Needs rule equality of outcomes taking into account need and/or desert
4	Rank order equality	If costs or investments of one group are higher than those of another, distributive justice requires their rewards should be higher too. Condition of equilibrium 'felt justice' normative expectations
5	Equality of opportunity	Complex because of complexity of term opportunity, historical antecedents of inequality

After Eckhoff, T. (1974). *Justice: Its Determinants in Social Interaction*. Rotterdam: Rotterdam Press; Cook, K. S. and Hegtvedt, K. A. (1983). Distributive justice, equity, and equality. *Annual Review of Sociology* 9, 217 241.

and more popular amongst policy makers than the fifth principle of equality of opportunity, which is difficult to measure and thus to monitor. Equal opportunity (also known as affirmative action) describes an approach in tended to ensure people are not excluded from activities of society (education, employment, healthcare) on the basis of 'immutable traits' or social characteristics. For example, this might be practiced as equal oppor tunity employment, which includes practices that are ethnic identity or gender blind, or that involve affirmative action or positive discrimination. In practice, methods for fulfilling equal opportunity can be difficult, especially since it is difficult to measure opportunity. Therefore equal opportunity is often said to exist when equality of outcomes exists, since outcomes can be measured more easily than opportunities. Equality of outcomes might be defined as being when people of similar abilities reach similar results, or the attempt to reduce differences in material condition between in dividuals, households, or groups in a society. This is still problematic: how are similar abilities measured? And how are we to account for inequalities that are passed from one generation to another? Even so, equity of outcomes might take into account the varying need of different groups, or the distributive justice requirement so that where costs or investments are higher for some groups, their outcomes or rewards should be greater too. For example, a social group who have been systematically and historically racialized or minoritized could be said to have borne a greater social cost than that of privileged groups, and thus be entitled to measures that will remove barriers to achieving outcomes that are equivalent to other groups.

It is possible to distinguish between two broad cat egories of justice principles: (1) those that rely on char acteristics of recipients (micro justice principles), and (2) those that do not (macro justice principles). In this entry, the focus is on macro justice principles related to power, social justice, and systemic fairness, as the term 'equity' has come to be used in geography. In evolutionary science, psychology, and economics, how ever, equity and equality are often defined in micro justice principles. For example, equity is defined as 'you get out what you put in', that is, more if you put in more, whereas equality is defined as 'everyone gets the same regardless of what they put in'. A large body of primate and human research deals with this exchange approach to equity, particularly with reactions to injustice. For example, individuals treated inequitably experience distress, especially those who are unrewarded. They respond by restoring equity, for example, they may alter their own inputs, alter their own outcomes, cognitively distort their own or their partner's inputs or outcomes, or leave the situation (recent work by Sarah Brosnan on fairness and co operation amongst Capuchin monkeys, and the Swiss economist Ernst Fehr on economics of human equity, co operation, and altru ism reflects this). There is also a very specific financial sense in which equity is used to mean the exchange value of a good or a resource controlled or held by the 'owner' above the value of any indebtedness on that good. The concept of equity as fairness stems from the historical legal sense of 'maxims of equity' from the old English form of law and the Court of Equity. In the latter sense, equity is associated with 'natural justice' or a specific set of legal principles which supplement strict rules of common law where their operation would impact harshly. This is the most closely related to equity as we are using it here.

The implementation of equity programs in public policy is extremely uneven geographically. In addition to major differences internationally in the degree to which the concept of equity is recognized and policies imple mented, scale differences occur within national juris dictions from the local to the national level, between the public and private sectors, and depending upon form of equity. In Canada, for example, there have been national policies in place for more than two decades to require equity in hiring (called 'employment equity') in those areas under federal jurisdiction and for those groups (women, racialized minorities, persons with disabilities, and aboriginal persons) designated as under represented. Equivalent programs for employers under provincial jurisdiction or in the independent private sector, however, are extremely uneven. And, while antidiscrimination clauses exist in almost every aspect of legislation, con stituting formal equality, there are few other areas where positive equity policies are in place, including education and, in all but one province, accessibility for persons with disabilities.

In the United States, in contrast, there is national legislation, the Americans with Disabilities Act, that covers accessibility for persons with disabilities, but in tense public debate over questions of affirmative action in other areas. For example, the principle of affirmative action in education became widespread during the 1970s in response to the civil rights movement, as universities, recognizing the systemic discrimination faced by mem bers of minoritized groups, established quotas for entry of those groups. The practice was challenged in a series of court cases, most notably that of Allan Bakke who, having been denied entry to the University of California medical school, took his case to the Supreme Court, which ruled in 1978 that he should be admitted to medical school, but also confirmed that affirmative action policies undertaken by universities, and by inference employers, may consider race and ethnicity as one of a number of factors when offering admission or hiring. Students or employees could not be denied entrance or employment if race was the only factor considered. Both sides of the debate claimed victory. In recent years, however, affirmative action policies have virtually disappeared in American universities, as state legislation has dismantled the prac tice based on the neoliberal principle of individual merit. Opponents of affirmative action presume that the *status quo* is fair and equal; discrimination can and does occur, but must be proven on a case by case basis. Claims for 'group rights' or systemic programs of positive redress, rather than leveling an uneven playing field, are seen from such a perspective to grant 'special interest groups' particular privilege, regardless of skill or merit.

Proponents of equity often refer to John Rawls' liberal concept of justice as fairness, or distributive justice. There are two principles to Rawls' concept:

1. each person has the same indefeasible claim to a fully adequate scheme of equal basic liberties, which scheme is compatible with the same scheme of lib erties for all (the liberty principle), and
2. social and economic inequalities are to satisfy two conditions: first, they are to be attached to positions open to all under conditions of fair equality of op portunity; and second, they are to be to the greatest benefit of the least advantaged members of society. (Rawls, 1971, 2001: 42)

The first principle, or the liberty principle, emphasizes sameness and collective equality. The second part of this is often called the difference principle, reflecting the idea that inequality is only justified if it is to the advantage of those who are least well off. This places priority on reducing systemic disadvantage. As Rawls described this in the 1990s

> the first [principle] requires equality in the assignment of basic rights and duties, while the second holds that social and economic inequalities, for example inequalities of wealth and authority, are just only if they result in compensating benefits for everyone, and in particular for the least advantaged members of society. These prin ciples rule out justifying institutions on the grounds that the hardships of some are offset by a greater good in the aggregate. It may be expedient but it is not just that some should have less in order that others may prosper. But there is no injustice in the greater benefits earned by a few provided that the situation of persons not so fortu nate is thereby improved. (Rawls 1995: 13)

For Rawls, then, we cannot think about sameness or equality as a collective goal or ideal without a broader judgement of how that goal affects those who are already least well off or most disadvantaged. Inequalities are only just if they result in some benefit for the least advantaged members of a group. The inequality between a doctor's salary and a cleaner is only acceptable, for example, if this is the only way to encourage the training of sufficient numbers of doctors, preventing an unacceptable decline in the availability of healthcare services which would disadvantage everyone.

In contrast to the distributive justice approach to equity which tends to emphasize the absence of socially unjust or unfair disparities, a rights based approach to equity emphasizes a more positive sense of equity beyond the absence of socially unjust disparities. The latter im plies the recognition of the need to consider the rights and needs of those who are vulnerable or disadvantaged, and for privileged groups to recognize these needs and their own advantages. The philosopher Amartya Sen has also been influential in inflecting nuance into discussions about equity and equality. For example, he argues that we cannot begin to discuss 'why equality' unless we also consider 'equality of what?' (e.g., incomes, wealths, opportunities, achievements, freedoms, or rights). This is also important in the equity discussion, since we might be discussing equity in terms of access to opportunities, experiences of systems and practices in institutions, material equity, equity in access to resources, equity in distribution of public goods, or equity of freedoms. This also raises the difficult question of 'equity for whom', since if equity is about improving the conditions of the most marginalized groups, many groups can potentially lay claim to the position of being 'most' marginalized. In practice, strategies to improve the condition of one disenfranchized group (such as affirmative action for employment matters) do not always improve the con ditions of other groups, and in the worst cases the success of one group in gaining ground in equity may actually be at the detriment of other systematically marginalized groups. This difficulty needs to be balanced against the strategic issues of getting equity on the institutional and policy agenda in the first place.

Sen's ideas therefore promote policies that emphasize equity in a positive sense (rather than a focus on reducing inequity) in order to increase human capacity to flourish as human beings. For example, the right to health as set forth in the World Health Organization's (WHO) Constitution and international human rights treaties is the right to 'the highest attainable standard of health.' Paula Braveman and Sofia Gruskin suggest that the highest attainable standard of health could be understood as the standard of health enjoyed by the most socially advantaged group in a society. These at least reflect levels that are physically attainable, and minimum standards for what should be possible for everyone in that society in the foreseeable future. There are, however, problems with this principle: first, what is the highest attainable standard – is it to be measured by life ex pectancy, longevity, quality of life, or access to prevent ive/primary or secondary care? Second, can we justifiably expect everyone to have those levels of consumption that the highest level of society has? How does this principle relate to equity for nonhuman agents? Also, how do we define the most socially advantaged group? Is it to be a decision measured by economic wealth, longevity, quality of life, social status, security, freedom from conflict, opportunity, power to, power over, etc.?

Equity thus also has the potential to be co opted, since it is likely that those in power are also those who determine at a societal level what is equitable and what is not. For example, in some situations where women are disenfranchized, those in power have argued that conditions are not unfair but are appropriate given the different capacities and roles of men and women. They argue that it is a principle of equity, rather than equality, to treat women and men differently according to their different needs. Similar essentialist arguments have been used in the past and present to justify discrimination, unfair treatment, and even oppression on the basis of ethnicity, age, and disability. In these contexts equality can become a crucial reference point.

Equitable procedures do not guarantee equitable outcomes. We need to think about equity in terms of procedures for access to opportunities and resources (e.g., to employment, health resources, environmental processes, and resources), and outcomes (e.g., experiences within workplace environments, policies and barriers that sustain unfair practices etc.). Equity could thus be thought of as procedural and consequential. Procedural (or instrumental) equity is about access to opportunities (e.g., equitable participation in decision making, equitable access to resources such as healthcare services, employment opportunities, or education), whilst constitutional equity is about equitable ability to enjoy the benefits and disadvantages of outcomes of decisions, and use of resources. Instrumental equity relates to the means to the end, a strategy or procedures rooted in social justice, for example, equity structuring how decisions are made, such as who should be interviewed and selected for an employment position, or how a resource should be used. Consequential equity is more related to an end, a goal, or something to be achieved. This might mean the experience of individuals from marginalized groups once they are in employment, so that they work in an environment that is socially just, or the socially just distribution of outcomes of decision making. Environmental equity is a good example. Procedurally, all groups, but particularly those most systematically marginalized and/or vulnerable to the impacts of climate change (and/or to strategies to address climate change), should have equitable involvement and participation in processes and decision making. Consequentially, all groups, including those most privileged by the decision, should share the distribution of costs, burdens, and responsibilities of climate change and strategies, as well as the distribution of resources, benefits, and opportunities.

Measuring Equity

There are a variety of approaches to assessing equitability. These might include parity (or equality) of procedure or outcome, proportionality (e.g., according to input), priority (e.g., according to need), classical utilitarianism (e.g., greatest good to greatest number), and Rawlsian distributive justice (e.g., inequalities are only just if they result in some benefit for the least advantaged members of a group). Further, with respect to consequential equity there are different perceptions of what a 'fair' or 'just' opportunity or outcome is and who is to define this. Often self interest influences the selection of criteria and determination of equitable decisions, and there is potential for great conflict over the assessment, definition, comparison, and aggregation of the welfare of different groups, parties, and nations.

Conflicts between equity and economic efficiency also arise. Efficiency often implies maximization of output subject to resource constraints, but this can result in inequitable income distribution, and may lead to inequities across generations. The debate on sustainability, or the need to balance short term interests with the protection of the interests of future generations (such as the responsibility to provide them with opportunities), is a classic example of both the conflict around how to define efficiency and the short term versus long term interests.

Further, there is tension between whether equity is to be defined (and measured and monitored) as existing between individuals or groups within a population. In health and public policy, for example, researchers increasingly define equity as between social groups, especially between worse off and better off groups. Researchers such as Barbara Starfield have been critical of definitions that measure differences between individuals because these disregard the ethical concern for distributive justice between groups, or that priority should be given to improving the situation of the most socially disadvantaged in a society. In 2000, for example, the WHO used the term 'inequities' to describe differences between individuals in a population. This was perceived as unethical because it masked the systemic injustices that lead to inequitable health outcomes. Improving overall average health status may have no impact on the difference between the worst and least advantaged groups in a society, and may even exacerbate this difference. In contrast the focus in a group related definition of inequity is on social disadvantage, implying a political commitment to social justice rather than a more individualistic focus on survival of the fittest. This is also an issue of scale and measurement of equity. Whereas most research on determinants of health uses data at the level of individuals, equity research requires modelling the influences on the health of populations at various levels (community, local, state, or national) to encourage consideration of the impact of factors in the policy and political realms on factors at the community level, which in turn influence the factors at the individual level.

Achieving Equity?

Equity is more effectively achieved by improving distribution between groups (of resources, of access, of outcomes) than by improving overall averages. Examples of national level strategies for achieving equity through redistribution include: progressive taxation systems; social safety nets to protect the very poor; providing social services for the population at large (particularly primary, generalized services that are accessible, e.g., affordable and high quality housing, public access to high quality education, preventive and primary health care rather than specialist tertiary care, good nutrition, and food security); responding to identified hardship; targeting specific groups for support; emphasizing political and civil rights; promoting genuine democracy; and widespread participation in decision making processes.

Institutional strategies for achieving equity include affirmative action and positive discrimination at a procedural level, to remove barriers that have limited people who have been systematically disadvantaged. Institutional change also requires more structural changes such as pay parity within workforces and between occupations, changes to policy, and particularly changes in behavior and practice. There is a tension between having equity policies in place in an institutionalized environment and practical compliance with these, or enforcing compliance.

Starfield usefully synthesizes evidence from a range of case studies presented at a workshop on Using International Experiences in Translating Health Equity Research into Policy, all of which suggest that addressing (health) inequities requires:

- recognition and documentation of the problem;
- existing principles of justice and explicit goals for attaining it (a clear and coherent ethical framework);
- specification of alternative policy strategies (e.g., addressing a particular disease may be more effectively approached through providing widespread basic primary services that are more acceptable to the affected population and its representatives);
- widespread participation (including those who are especially affected, e.g., engaging diverse communities);
- support at high levels of government and/or strong and dynamic leadership; and
- focus on distribution (between social groups) rather than improving overall averages. (Starfield, 2006: 25)

While this workshop focused on equity and health specifically, these principles could feasibly be applied to the issue of equity in a number of situations, including equity in labor markets, education, health, environmental issues, justice issues, and access to resources, and at a variety of spatial and temporal scales from localized to global, and immediate to long term.

Equity and Geography

There are several ways in which geographers can contribute to discussions about equity. For example, spatial and temporal scale are critical equity issues. There is a tension between inter and intragenerational equity, just as there is a tension between inter and intraregional equity. The aggregate geographical level at which equity is measured will have an impact on the degree of inequity (hence the political process of gerrymandering, or dividing an area unfairly to one's own political advantage by concentrating the opposition's strength into as few districts as possible – or distributing it over many districts to dilute the strength of a social group). Equity needs to be understood as a systemic process operating at interconnected scales, so that equity at a global scale has implications for equity at national, regional, community, family, and individual levels. It also needs to be understood as a form of strong discourse that produces normative notions of the human body and the places that human bodies occupy.

In an affirmative and very practical sense, work by geographers on the inter relationship of physical and social environments has shown how adjustments to physical environments can promote social equity. For example, functional limitations associated with physical impairments can be changed with basic measures and environmental design (e.g., installing railings, ramps at appropriate angles, appropriate flooring or street materials, GPS navigation systems, etc., and providing physical training to increase strength). Socially, the degree of impairment associated with many psychological and physical conditions is highly related to the degree of social stigmatization or acceptance of people with those conditions, a concept fundamental to the social model of disability. Other forms of physical access, such as access to transportation, also emphasize that equity can be a spatialized condition.

Finally, the issue of equity appears increasingly in discussing access to the discipline of geography. The discipline has been dominated by males of European background since the Enlightenment and recent studies show that although the academy is becoming more diverse, change happens slowly. Analysts who have questioned why change occurs slowly find that it is partly a question of scale: decisions are made at the level of individual departments, often on the basis that all candidates are treated equally, but the cumulative result when scaled up is a discipline that shows the systemic results of years of differential access to education, of the location (in the world) of those who have enjoyed access to education, and the sociospatial reproduction of priv ilege. As geography struggles to achieve equity, so equity struggles over geography.

See also: Environmental Justice; Gerrymandering; Human Rights; Inequality; Scale; Social Justice, Urban.

Further Reading

Braveman, P. and Gruskin, S. (2003). Defining equity in health. *Journal of Epidemiology & Community Health* 57, 254 258.
Burman, E. and Chantler, K. (2004). There's no place like home: Emotional geographies of researching 'race' and refuge provision in Britain. *Gender, Place and Culture* 11(3), 375 397.
Cook, K. S. and Hegtvedt, K. A. (1983). Distributive justice, equity, and equality. *Annual Review of Sociology* 9, 217 241.
Eckhoff, T. (1974). *Justice: Its Determinants in Social Interaction*. Rotterdam: Rotterdam Press.
Gender, Place and Culture (2006). *Special Issue on Equity. Gender, Place and Culture* 13(1).
Hall, J., Murphy, B. L. and Moss, P. (2002). Focus: Equity for women in geography. *Canadian Geographer Geographe Canadien* 46(3), 235 240.
Munasinghe, M. (2000). Development, Equity and Sustainability (DES) in the context of climate change. *IPCC Guidance Paper*, Oregon State University.
Rawls, J. (1995). *A Theory of Justice* (rev. edn.). Cambridge, MA: Harvard University Press.
Rawls, J. (2001). *Justice as Fairness: A Restatement* (first publ. 1971). Cambridge, MA: Belknap Press.
Sen, A. (1995). *Inequality Re Examined*. Oxford: Oxford University Press.
Starfield, B. (2006). State of the art in research on equity in health. *Journal of Health Politics, Policy and Law* 31(1), 11 32.
Whitehead, M. (1992). The concepts and principles of equity and health. *International Journal of Health Services* 22(3), 429 445.

Error (Propagation and Modeling)

S. Gopal, Boston University, Boston, MA, USA

Glossary

Accuracy It is the degree to which information in a geographic information system (GIS) matches true or accepted values. Accuracy pertains to data quality and the number of errors contained in a map or GIS database.

Ambiguity It arises when objects are assigned different labels by different national or cultural groups.

Error or Confusion Matrix This is a statistical method of error analysis used in remote sensing in which a reference set of classification data is compared to another dataset or ground truth.

Error Propagation When maps contained in a GIS database are used as input to a GIS operation, such as in map overlay, then the errors in the input data layer will propagate or spread to the output of the operation. Errors continue to propagate when the output from one operation is used as input to an ensuing operation.

Fuzzy Sets It is used as a representation scheme and calculus for representing the concept of a class with unsharp boundaries. It is basically a multivalued logic that allows more precise representation of transition zones and imprecise categories in GIS.

Precision It refers to the number of digits used to report a measurement that is often related to the instrument's measurement accuracy. It could also refer to the performance of an instrument in making repeated measurements of the same phenomenon.

RMSE (Root Mean Square Error) It is the square root of the average squared difference between recorded measurement and the truth; it is roughly interpreted as the average difference or as the standard deviation of the error distribution.

Uncertainty It is used to describe situations in which the digital representation in a GIS is incomplete or inaccurate in representation of reality.

Vagueness It denotes the uncertainty that characterizes the positions of the boundaries of a zone and in its attributes.

Introduction

In the last decade, there has been an explosion in the use of geographic information system (GIS). There is a widespread tendency among users to believe GIS outputs (e.g., maps) as being inherently accurate. Many researchers have raised the question about the quality of GIS end products since GIS data and analysis are subject to error. No map output and model produced from a GIS is truly error free since errors and uncertainty creep into a GIS in a variety of ways. Ultimately, errors undermine the purpose of GIS analysis and lead to increased uncertainty in the validity of the conclusions that are drawn. Thus, there is a need to understand how errors arise and propagate through the GIS and impact the conclusions.

In GIS context, the term error is associated with uncertainty, ambiguity, accuracy, and precision. Error refers to the difference between actual and represented value of a variable in GIS while uncertainty refers to our imperfect and inexact knowledge of the world. The two are often used synonymously. Uncertainty in GIS is present in the stages of conception, measurement and representation, and analysis. Conception errors arise since it is impossible to capture or create a perfect representation of the world in a GIS database due to uncertainty in conception. Many geographic data and definitions are characterized by ambiguity, vagueness, and spatial uncertainty. For example, mapping wetlands has proved to be a legal challenge since wetlands can be classified differently according to their respective functions, values, and benefits. In human geography, it is not easy to represent the true spatial coordinates of residents' perceived neighborhoods. Nor is it easy to represent the true meaning of everyday spatial semantics such as 'near', 'far', 'on', 'at', and 'in'. These terms are culture specific and are fuzzy concepts presented in the seminal work of Zadeh.

Measurement uncertainty is inherent in both raster and vector data models of GIS. The Earth is not a perfect sphere, and hence, a range of mathematical approximations are made of it leading to many types of measurement errors. Positional error in GIS refers to the measurement of position, where every coordinate (x, y, and z) may be subject to errors. Positional accuracy refers to the degree to which the measured positional information matches true or accepted values. The mapping standards employed by the United States Geological Survey specify that:

> requirements for meeting horizontal accuracy as 90 per cent of all measurable points must be within 1/30th of an inch for maps at a scale of 1:20,000 or larger, and 1/50th of an inch for maps at scales smaller than 1:20,000.

The referenced Java applet shows a simulation of uncertainty in the survey of a square parcel of land.

Assuming that each corner point is subject to an in dependent error in both coordinates, a root mean square error (RMSE) of 2 m, the Java applet simulation tracks the changes in the average area, standard deviation, and other statistics as well as a display of a histogram of areas. The Java applet is platform dependent (and hence can be run on personal computers (PCs) and Macs) and enables a simple visualization of simulating uncertainty.

A distinction is often made between accuracy and precision in measurement since they both refer to dif ferent aspects of the measuring instrument's ability to capture reality. Accuracy is the degree to which infor mation in a GIS matches true or accepted values while precision refers to the performance of an instrument (such as a Global Positioning System (GPS)) in making repeated measurements of the same phenomenon. Pre cision in measurement does not guarantee accuracy. The measurement could be highly precise but inaccurate. Another definition of precision is more commonly used and refers to the number of digits used to report a measurement. In many GIS calculations, answers are overly precise (i.e., contain many decimal points) that are often spurious. While accuracy relates to validity errors, reliability refers to the consistency or repeatability in measurement, for example, a GPS instrument at a given location should consistently produce same latitude and longitude readings to be considered reliable. Note that the GPS could consistently produce same results and hence be reliable but not accurate.

In human geography, almost always, we sample only a few individuals of the total population often resulting in sampling error. For example, we might take a sample of shoppers to measure shopping patterns in a local super market. The sample may not be a true representative of the total population. Sampling errors are more difficult to characterize and correct than measurement errors. Most frequently, we assume a normal population. The type and magnitude of these errors can be determined by repeated subsampling or by more exhaustive sampling.

There is extensive statistical theory for dealing with both measurement and sampling errors. Most often, an 'expected value' (usually given by the mean) and a 'variance' of a normal or Student's distribution can be used to express the uncertainty of the expected value. If the distribution of errors is known to have a non normal form, the parameters can be estimated using a different distribution.

Uncertainty may be introduced in representing the data in a GIS. In a vector database, error may be intro duced in map generalization process since detail is lost due to generalization at smaller spatial scales. For ex ample, a coastline is conceptualized as a discrete line object. The position of the coastline is uncertain at coarser scales. At a scale of 1:1 000 000, there is gross generalization as many irregularities and indentations may be generalized. **Figure 1** shows the details of the Florida coastline and **Figure 2** shows the generalization of the same coastline map. It can be seen that the second map does not contain details in the coastline, including inlets, islands, and other irregularities in the shape of the coastline as in **Figure 1**.

In a raster database, much research has been devoted to errors in thematic or categorical maps. Thematic maps are made for a specific purpose and portray information using some system of classification for the landscape such as soil taxonomy, vegetation classes, and land use categories. In a thematic map, each location on the ground has to be assigned to a category (or class). Typically, the categories are easily differentiable in their pure states and become more ambiguous at the bound aries between the categories. For example, it is difficult to distinguish cultural classes such as golf courses and parklands from natural vegetation using an automated classification procedure. Similarly, it is difficult to assign a single label to meadows and grassland since the dis tinction may be primarily a function of annual water balance and may vary year to year. This results in mis classification errors due to uncertainty.

The last type of uncertainty is in the analysis of geographic phenomena, particularly relevant in human geography, where the object of study is the individual. Most GISs use aggregate data about the individual due to privacy concerns or cost associated with data collection at the individual or household level. As a consequence, there is uncertainty in ascribing even the most dominant characteristics of areas to an individual. For example, if an individual is located in a cancer cluster, does it mean he/she has cancer? This is often referred to as the ecological fallacy where an incorrect inference is drawn about the characteristics of the individual using aggregate data.

Related problem in GIS is called the modifiable areal unit problem (MAUP) described first by Openshaw who did seminal work related to MAUP, and its two mani festations, zoning and scaling effects. Zoning effect refers to the variability in statistical results obtained within a set of areal units as a function of the various ways these units can be grouped at a given scale. Scaling effect refers to the variability in statistical results obtained from the same set of data grouped at different levels of spatial resolution (e.g., census blocks, tracts, counties, states, and regions). Regional geography, a branch of human geog raphy, is concerned with the delineation of regions, zones, or states, such that internal homogeneity within zones is minimized while between zone heterogeneity is maxi mized. MAUP introduces uncertainty in the character ization of the positions of the boundaries of a zone and its attributes. Statistics, including regression, correlation, and spatial autocorrelation, may vary markedly due to scale and zoning effects.

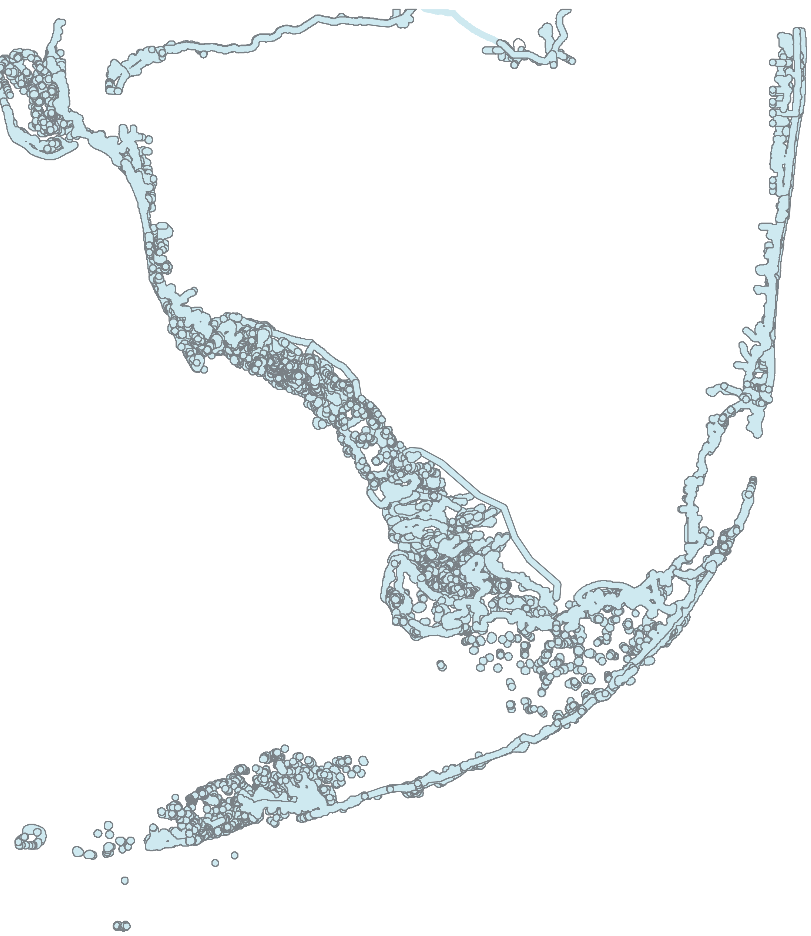

Figure 1 Detailed coastline of Florida.

Scale effects can be visualized by mapping categories at each scale. For example, in a recent study on addiction and serious mental illness (SMI) in the US, local spatial autocorrelation was identified using local Moran's *I*. Since the local Moran statistic decomposes Moran's *I* into contributions for each location, I_i, the sum of I_i for all observations is proportional to Moran's *I*, an indicator of global pattern. Thus, local Moran statistics can have two interpretations, as indicators of local spatial clusters and as a diagnostic for outliers in global spatial patterns. In the study on addiction and SMI, the following questions were addressed:

1. Do local clusters (regions where adjacent areas have similar values) or local spatial outliers (areas distinct from their neighbors) exist in the pattern of addiction and SMI in the US?
2. Are these local clusters and spatial outliers statistically significant?

Figure 3 shows the state level and substate level analysis for addiction to marijuana. Large areas of high use in California, Oregon, and Colorado are present in the substate (substate geography in the US census contains aggregations of counties) local indicators of spatial association (LISA) map but not at the state level LISA map. If only data at the state level were considered, one would mistakenly assume that marijuana is only a problem in the Northeast and not the West as well. In addition, the state distributions did not hold up to the rigorous significance tests of the Moran's scatterplot. Such findings have implications for health policy and funding.

Error Modeling

Error modeling deals with statistical tools for describing and modeling error and uncertainty. In this section, we describe the error modeling for categorical (nominal) and continuous (interval and ratio) data scales. Classes shown in a thematic map are categorical data and traditionally, the accuracy of a thematic map is determined empirically by comparing the map with corresponding reference or ground data. Often in remote sensing, an error or confusion matrix is built, in which a reference set of classification data are compared to another dataset or ground truth. The product is a square 'error matrix' where the

Figure 2 Generalized coastline of Florida.

columns and rows are the categories in the classification. The reference classification is represented in the columns and the ground truth in the rows. The diagonal elements of the matrix represent frequencies of agreement be tween the two classifications and measure accuracy in classification, while off diagonal elements give the fre quencies of misclassified items by each category. **Table 1** shows the confusion matrix for five land cover classes in a dataset.

Columns in **Table 1** represent ground truth and rows represent the land cover classes in a map. The accuracy of the classification is the sum of the elements of the di agonal divided by the total number of sites or pixels ($10+6+3+10+5=34/50$ or 68%). For example, there were only three instances of agreement between the GIS map and ground truth (of field data) in the case of brush. Columns represent the entries that truly belong to a class that are actually recorded as such in the GIS. For example, of the seven instances of brush found in the field, three were misclassified as conifer and one as hardwood. The columns of the error matrix represent the producer's ac curacy because the producer is concerned with mini mization of error while the rows have been called the users accuracy because the user is interested in the contents of the GIS database. In the case of brush, producer's accuracy is 3/7 or 43% while user's accuracy is 3/10 or 30%.

Omission error refers to the exclusion of an instance or a pixel that should have been included in the class (i.e., omission error $= 1 -$ producers accuracy $= 1 - 0.428 = 0.57$). Commission error refers to including a pixel in a class when it should have been excluded (i.e., commission error $= 1 -$ consumer's accuracy $= 1 - 0.30 = 0.70$). A more useful measure of accuracy is kappa statistic, which estimates the number expected by chance in each diagonal cell by multiplying the appro priate row and column totals and dividing by the total number of cases. Kappa $K=(d-q)/(N-q)$, where d is the number of cases in the diagonal cells, q is the number of expected cases in diagonal by chance, and N is the total number of cases or pixels. In this context, $q=23.3$. The kappa statistic for **Table 1** is $K=(34-23.3)/(50-23.3)=0.40$. Kappa varies between 0 and 1 and values closer to 1 suggest greater accuracy. There are other ways of quantifying error in a thematic map including one based on fuzzy sets.

Error Propagation Modeling

Error propagation modeling allows the GIS users for assessing the effects of known degrees of error in a model's inputs and producing measures of confidence in model outputs. Error propagation is an important issue in GIS map overlay or other operations, since input data from disparate sources are overlaid and each input layer may have a wide range of errors associated with it. For example, suppose two maps with accuracies of 0.75 and 0.80 are overlaid, the accuracy of the resultant map is the product of the errors associated with each input map layer, $0.75*0.80=0.60$.

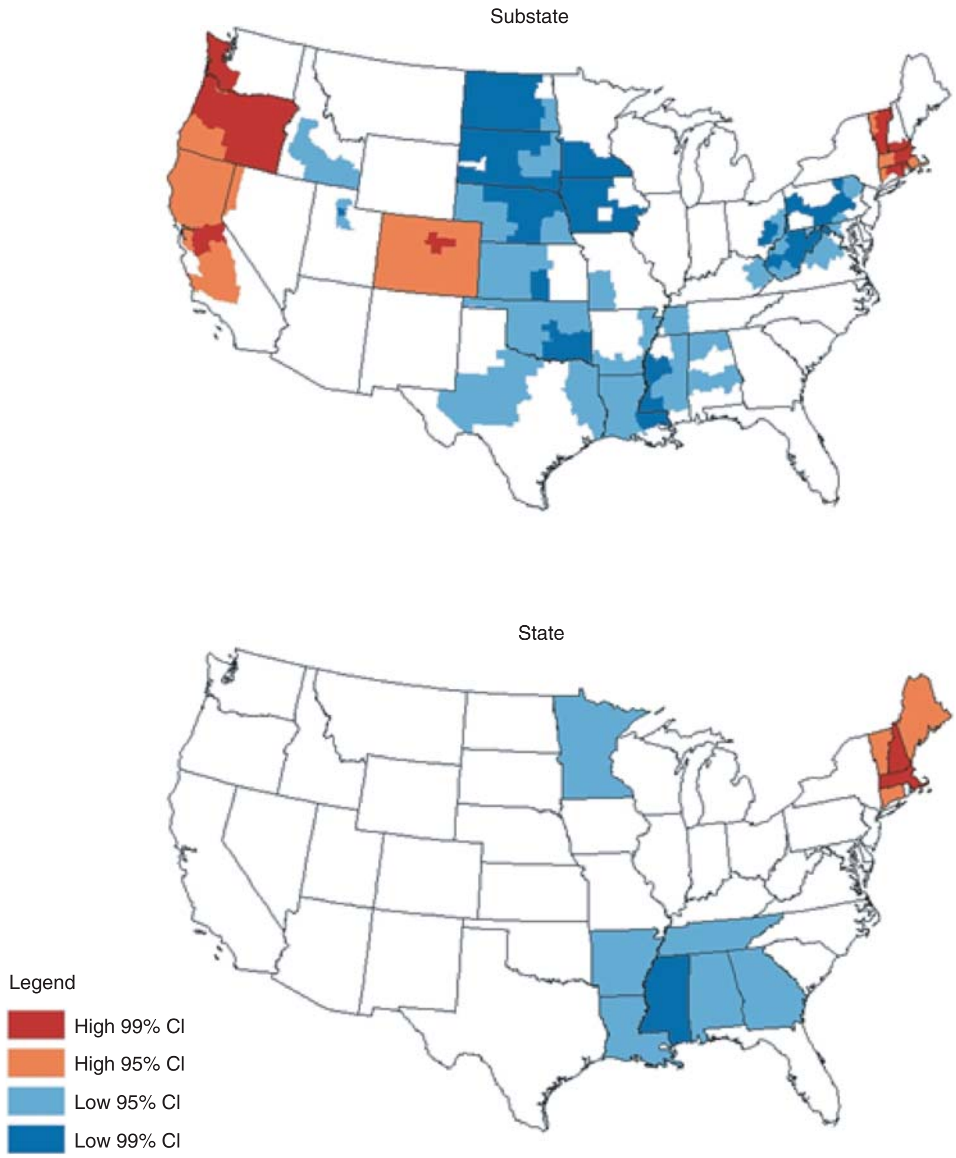

Figure 3 LISA analysis of marijuana addiction at the state level and substate levels of US (based on National Survey on Drug Use and Health Survey Data 1999 2001).

Table 1 Error or confusion matrix estimating categorical error

Class	*Number of pixels or instances of a class*					
	Conifer	*Hardwood*	*Brush*	*Water*	*Meadows*	*Total*
Conifer	10	2	3	0	0	15
Hardwood	3	6	1	0	0	10
Brush	4	2	3	0	1	10
Water	0	0	0	10	0	10
Meadows	0	0	0	0	5	5
Total	17	10	7	10	6	50

Error in a quantitative attribute (continuous data) can be conveniently defined as the difference between 'reality' and our representation of 'reality' shown on the map. For instance, if the rainfall at some location equals 20 mm, while according to the map it is 15 mm, the error is $20 - 15 = 5$ mm. Generalizing this example, let the true value of a spatial attribute (rainfall) at some location x be $a(x)$, and let the representation of it be $b(x)$. Then, according to the definition, the error $v(x)$ at x is simply the arithmetical difference $v(x) = a(x) - b(x)$. Note that the error $v(x)$ is never exactly known, since if it was, then it could simply be removed. Rather, what we do know about $v(x)$ is limited to specifying a distribution of possible values and can best be conveyed to the GIS user by representing the error as a random variable $V(x)$. Statistical error modeling allows us to characterize uncertainty in distribution of the variable.

The error propagation problem can be formulated mathematically as follows. Let U be the output of a GIS operation g on the n input attributes A_i: $U = g(A_i, A2, An)$.

Figure 4 Land cover map of North America using fuzzy ARTMAP neural network classification. Data from the moderate resolution imaging spectroradiometer sensor.

The objective is to determine the error in the output U, given the GIS operation g (such as map overlay) and the errors in the input attributes A_i. The uncertainty of U is contained in its variance t^2.

The error propagation problem is relatively simple when g is a linear function since both mean and variance of U can be analytically derived. However, the propa gation of errors through complicated dynamic spatial

Figure 5 Confidence in land cover classification using ten fuzzy ARTMAP neural networks.

models can only be assessed with powerful computational techniques. Most of these techniques model errors as stochastic variables. Analytical solutions for error propagation exist only for a limited number of relatively simple spatial functions. As a consequence, error propa gation in dynamic spatial models is mostly calculated by Monte Carlo simulation modeling.

Error propagation has profound effects on GIS prod ucts and operations. For example, if there is an under representation of suburban regions (due to difficulty in assigning class labels to suburban and woodlands or grasslands) in a GIS map, the error may propagate through many GIS operations such that final results in the GIS map may not truly represent suburban sprawl.

Visualization of Error and Uncertainty in GIS

Error visualization is an important topic by itself. Improvements in multimedia technology, virtual reality (VR) environments (e.g., Maya), and programming lan guages (e.g., Java, Python), and increasing computation power available in personal computers and desktops have all contributed to an interest in visualization. Uncertainty data layer can be visualized as animated movies of simulations that give a graphic rendering of error land scape. Animation capabilities are present in ArcGIS, and Geoda. While animation provides an easy way to see all renditions of the uncertainty model, it is not impossible to infer quantitatively where the uncertainties are large or small. For example, Hengl and his colleagues used saturation intensity (HSI) color model to visualize multiple memberships and confusion between thematic categories. In the case of continuous variables, both predictions and the prediction error can be visualized simultaneously. These visualization techniques use principles of cartographic representation to produce the uncertainty landscape.

Figure 4 shows the land cover map of North America using Fuzzy ARTMAP, neural network classifier (Liu et al., 2004). There are 14 land cover classes. Due to various uncertainties, the automated neural network was not completely confident in all its predictions. **Figure 5** shows a final confidence level map of ten independent neural networks obtained by pooling their classification results together. One can think of each neural network classifier to be an expert in a panel of experts whose consensus establishes the winning class for each pixel. There are levels of agreement (shades of green) and disagreement (shades of orange) among the experts shown in the confidence map. There is some agreement over 60% of the map while there is disagreement over mixed forest, croplands, and woody savannas. Such a confidence map gives additional information to the user of the map. The user may place less reliance on classes that exhibit greater degree of uncertainty.

Managing Error in GIS Practice

Currently, there are three methods to cope with errors in GIS – ignore the issue, characterize uncertainty with statistical measures, for example, RMSE, probability confidence, and variance, and finally, simulate equally probable realizations of the data (using Monte Carlo (MC) technique). Because of improved access to com putational power, many researchers are using the stat istical and computer simulation method to improve our understanding of errors in GIS. Software packages such as Geoda, IDRISI, and ArcGIS are capable of performing statistical error analysis as well as help in visualization.

Managing error in GIS datasets is now widely recognized since failure to control and manage error can invalidate the results of a GIS analysis. The following procedures are prescribed for practical management of error – setting standards for procedures and products, making periodic data quality reports, measuring and testing products (such as kappa statistic), and calibrating a dataset (via sensitivity analysis) to ascertain how error influences solutions, and report results in terms of the uncertainties of the data.

Conclusion and Future Directions of Error Research

Uncertainty and error analysis in GIS is important since it provides validation. Every GIS map should have an uncertainty layer associated with it, which will provide the user with additional information. The development of formal theoretical frameworks, programming languages like Python and Java, and the creation of new visual representations of error and uncertainty can help validate new theoretical models, enable better understanding of data, and facilitate better decision making using GIS. GIS is not about making pretty maps but rather about making valid decisions and robust spatial models. Thus, statistical modeling of errors, MC techniques, and visual repre sentation of errors and uncertainties should become the norm rather than the exception.

See also: GIS and Cartography; GIScience and Systems; Neural Networks; Uncertainty.

Further Reading

Burrough, P. A. (1986). *Principles of Geographical Information Systems for Land Resources Assessment*. New York: Oxford University Press.

Goodchild, M. F. and Janelle, D. G. (eds.) (2004). *Spatially Integrated Social Science*. New York: Oxford University Press.

Goodchild, M. F. and Janelle, D. G. (2004). Thinking spatially in the social sciences. In Goodchild, M. F. & Janelle, D. G. (eds) *Spatially Integrated Social Science*, pp 3 22. New York: Oxford University Press.
Goodchild, M. F., Yuan, M. and Cova, T. J. (2007). Towards a general theory of geographic representation in GIS. *International Journal of Geographical Information Science* 21(3), 239 260.
Gopal, S., Adams, M. and Vanelli, M. (2007). Modeling the spatial patterns of addiction in the US. In Thomas, Y. F., Richardson, D. & Cheung, I. (eds.) *Geography and Drug Addiction*, Chap 25, pp 415 435. New York: Springer.
Gopal, S. and Woodcock, C. (1994). Theory and methods for accuracy assessment of thematic maps using fuzzy sets. *Photogrammetric Engineering and Remote Sensing* 60(2), 181 188.
Herskovits, A. (1987). *Language and Spatial Cognition*. New York: Cambridge University Press.
Heuvelink, G. B. M. (1998). *Error Propagation in Environmental Modelling with GIS*. London: Taylor and Francis.
Liu, W., Gopal, S. and Woodcock, C. E. (2004). Uncertainty and confidence in land cover classification using a hybrid classifier approach. *Photogrammetric Engineering and Remote Sensing* 70(8), 963 972.
Longley, P. A., Goodchild, M. F., Maguire, D. J. and Rhind, D. W. (2005). *Geographic Information Systems and Science* (2nd edn.). New York: Wiley.
Openshaw, S. (1984). *The Modifiable Areal Unit Problem (CATMOG 38)*. Norwich: GeoBooks.
Story, M. and Congalton, R. (1986). Accuracy assessment: A user's perspective. *Photogrammetric Engineering and Remote Sensing* 52(3), 397 399.
Veregin, H. (1994). Error modeling for the map overlay operation. In Goodchild, M. F. & Gopal, S. (eds.) *Accuracy of Spatial Database*. London: Taylor and Francis.
Zadeh, L. (1965). Fuzzy sets. *Information Control* 8, 338 353.

Relevant Websites

http://www.geog.ubc.ca
Department of Geography at The University of British Columbia.
http://www.itc.nl
International Institute for Geo Information Science and Earth Observation (ITC).
http://www.ncgia.ucsb.edu
National Center for Geographic Information & Analysis (NCGIA).
http://www.colorado.edu
University of Colorado at Boulder.

Ethical Issues in Research

R. Dowling, Macquarie University, Sydney, NSW, Australia

Glossary

Ethics It is concerned with perspectives of right and proper conduct. Discussions of ethics in relation to research methods are a form of applied ethics.
Institutional Ethical Review An institutional review is a review process held by an independent review board that provides protection for human research participants through the initial and ongoing review of research studies. Boards usually comprise highly qualified individuals representing a variety of scientific and nonscientific disciplines.
Research Ethics Perspectives on the proper conduct of researchers during their interactions with research participants and the social, emotional, and political consequences of research methods.
Situational Ethics Contextual ethical discussions, taking account of specific circumstances.
Universalist Ethics Ethics or standards which can be applied and upheld universally.

Introduction

Ethics are not abstract ideas that float above people's activities but permeate human behavior, relations, and decision making. As such, research methods in human geography necessarily involve ethical considerations. Decisions about which research topics to pursue, appropriate and worthwhile methods of investigation, 'right' ways to relate to sponsors and participants of research, and appropriate modes of writing and com munication of results involve ethical questions. These questions include: how researchers 'ought' to behave, the role of research in the pursuit of social change, and whether and how research methods are 'just'. A number of different ethical perspectives on methods are adopted in human geography. Universalist ethics are used in processes of ethical review and in research adopting a scientific–positivist mode of enquiry. Situational ethics are more common in qualitative research, and more recently, relational ethics have been advocated, especially in critical human geography. This article provides an overview of how human geographers approach and ad dress the ethical aspects of methods. It begins with a brief history of methodological ethics in human geography and their variation across the discipline. It then outlines the key ethical issues and the diverse means through which research ethics are governed. A consideration of ethical issues that pertain to the research methods of human geography raises more questions than answers. This article does not provide definitive solutions, but provides an overview of the ethical routes human geographers have taken.

Background History and Context

Geographical research methods have always involved ethical considerations, but it is only in the last 40 years that published and formal reflections on ethics have become widespread. The publication of Mitchell and Draper's *Relevance and Ethics in Geography* was a defining moment in human geography's engagement with ethics. Not only was it the first book length treatment of research ethics in the discipline, but it also identified the two principal methodological aspects of ethics that have been of enduring concern to human geographers. The first issue identified by Mitchell and Draper was 'relevance', which can be more broadly encapsulated as the social purpose of research. In whose interest is research in human geography? What is the societal role of human geographical research? Are human geographers disinterested observers who chronicle life's happenings? What is the 'right' purpose of human geography research? What is the role of justice and advocacy for a more just society in the practices of research? The second key methodological aspect of ethics identified by Mitchell and Draper was ethical behavior in the actual conduct of research – researchers' interactions with those they research, and the consequent issues like privacy and harm that arise.

The issues raised and solutions proposed by Mitchell and Draper remained little examined outside a narrow field for much of the ensuing decades. During the 1980s and 1990s, for example, ethical issues remained in the background of methodological discussions in human geography. While there was continued discussion about the social role of research, the ethical aspects of many research methods did not garner much interest. The mid to late 1990s saw a deeper and more widespread engagement with research ethics, linked to renewed interest in ethical philosophy, the widespread use of feminist and other critical theories to re evaluate geographical methods, and stronger concern with geo graphical methods in general. Journals such as *Ethics, Place and Environment* emerged, with a broad remit that includes research ethics, and ethical debates occurred within a number of subdisciplines, including, but not confined to, geographic information system (GIS), cultural geography, and medical geography. Currently, the profile of research ethics within the discipline is strong, with geographers contributing to debate beyond the discipline.

Within the context of this stronger profile of methodological ethics, human geographers' engagement with research ethics is noteworthy in the diversity of approaches adopted and solutions advocated. The spectrum of philosophical and political stances embraced within the discipline engenders key differences in the social purpose of research and moral codes regulating the interactions of researchers with the people and social worlds they research. A simple yet powerful encapsulation of this diversity is presented in **Table 1**. For heuristic purposes, it divides human geographers' ethical stances into three main groups, and illustrates the research ethics that follow.

In positivist human geography (as practiced, for instance, in most of contemporary GIS), researchers document the social world rather than being immersed in it, or play a role in changing it. Research stands outside politics and society, and is neutral, value free, and devoid of power relations. This stance is summed up in the use of the term 'subject' to describe people involved in research projects: they are the subjects of investigation, and the research process and researcher are objective. Quantitative methods generally endorse universalist codes of ethical review that focus on issues of privacy, anonymity, and informed consent. Qualitative human geography, in particular, as practiced in contemporary cultural geography, envisages research to be a part of, and constructive of, societal processes. Those researched are hence participants or respondents, and the research relationship requires recognition and negotiation of intersubjectivity rather than objectivity. Power relations, cultural and other stereotypes, and issues associated with personal interactions such as emotions are defined as key ethical issues for these methods. Finally, in critical human geography, research, by its very nature, does and should effect social change. Action research methodologies emanate from this framework, with subjects/participants identified as co researchers and co creators of knowledge. The key ethical issue in this framework is the purpose and use of research, and its potential to challenge and transform dominant power structures and relations.

These three broad frameworks are neither mutually exclusive nor internally homogeneous. There exists some agreement on ethical issues across the spectrum. For example, elements of positivist human geography, and critical GIS, share critical geography's position on what counts as 'good' and 'just' research. Similarly, both quantitative and qualitative researchers recognize the potential of these research methods to be disempowering. Further, qualitative methods may be used in an interpretive, critical frame, or from a scientific basis. With the increasing adoption of mixed methods in human geography, such intersections will also invariably increase.

Relevance or the Social Purpose of Research

For many human geographers, ethical issues arise in the definition of research questions and specification of the methodological framework, before any specific methods are implemented. The issue of relevance became hotly contested in the 1970s, where criticisms were made of the interests served by research using a spatial science framework. In response, calls were made for geographers to do more than study society, and that they should be advocates who contribute to, and provide solutions for, socially significant problems. The Detroit Geographical Expeditions, where university geographers sought to use university teaching and research to forge better social and economic outcomes for disadvantaged Blacks in Detroit, was an important early example of geographers as advocates.

Table 1 Perspectives on research ethics in human geography

	Positivist human geography	*Qualitative human geography*	*Critical human geography*
Role of research	Observation documentation	Socially, politically, and culturally embedded	Action research
Power	Absent or unimportant	Recognized; critical reflexivity	
Representative term for those researched	Subject	Participant	Co-researcher
Researcher researched relationship	Objectivity	Subjectivity Intersubjectivity	Co-creation of knowledge
Key ethical issues	Anonymity Privacy Consent	Power relations of research	Transformation of power relations
Principles of ethical review	Universalist	Situational	Situational/Interactional

Debates about relevance have continued since then, across divergent subfields. In general, scholarship is criticized for its lack of normative ethics. In studies of disability, for example, calls have been made for geog raphers to explicitly adopt a normative ethics stance and use the aim of a more just society to frame their choice of methods. In contrast, the ethical basis and outcomes of GISs and geodemographics have been widely questioned. Is it possible for GIS to be used for socially beneficial ends and avoid undesirable outcomes, especially given the methods' original and ongoing links to military ap plications? Whilst some like Lafferty argue that picturing people as numbers and points in space is ethically problematic, others like Kwan and Pavlovskaya have developed a case for the role of GIS in critical and feminist geographies.

Key Ethical Issues in the Methods of Human Geography

Mitchell and Draper's second sense of ethics as research ethics – or researchers' relations with research subjects/ participants – has been the focus of considerable geographical engagement. The conduct of researchers – how they relate to, respect, negotiate with research par ticipants – has been termed research ethics, with two broad topics attracting most attention.

Consent, harm, and power relations

In most traditions of geography, research involves inter actions with people: in interviews, focus groups, obser vations, or filling in questionnaires. Many of the ethical discussions relate to understandings of how participants should be treated, and of the consequences of the research relationship. There is widespread agreement that ethical research practice in human geography in volves informed consent: that research subjects consent to being involved in the specific research project. Exceptions to informed consent, and in particular de ception, are rare in human geography. Unlike other fields such as psychology where quantitative methods are just as likely to involve deception, it is within qualitative frameworks that informed consent has been waived in human geography. Research on those in positions of relative power has occurred without fully informing participants of the purpose of the research. The justifi cation given is that the research could not have proceeded with informed consent, that there was no potential harm to respondents, and that the social good that would result from the research outweighed the de ception. Ethnographies of sites like workplaces or tourist resorts have also involved deception, with the researchers not disclosing their research purposes but instead pre senting themselves as employees, tourists, or the like. In these cases, the necessity of having an 'authentic' and 'true' experience is seen to outweigh the deception.

Most traditions of human geography acknowledge that research interactions are not divorced from relations of power. Researchers may be more privileged than their respondents, or perhaps in a position of subservience if research involves elites. More generally, research in human geography is potentially exploitative – it uses other people's time, opinions, etc., for personal or institutional gain. Moreover, the potential for exploitation is greater in some research methods and projects than in others. For example, researchers may bring and perpetuate ethno centrism and/or racism to their cross cultural research encounters. Research involving vulnerable populations such as children, people with mental illness, people who are homeless, communities in the developing world, and indigenous people has greater potential for exploitation. These potentials exist across the spectrum of quantitative, qualitative, and action research methods. Exploitation is an ethical issue, because it raises questions about what is 'good' or 'right' research practice and the potential for research methods to be socially or psychologically harmful to participants. The general consensus is that exploitation should be avoided, though very different mechanisms of avoidance or management of exploitation are advocated. One option is to make the research more of a two way relation through giving something back to participants. This could include reports of the research findings, transcripts of their interviews to review, and perhaps training as researchers. Financial payments are sometimes made, though these are controversial. While they are an acknowledgement of the time people give to the research, they do not obviate unequal power relations. In some cross cultural research settings, or action research frame works, exploitation is directly confronted through the in volvement of participants in the formulation of research questions, choice of method, and interpretation of the findings.

Certain qualitative research methods such as in depth interviews raise additional questions about consent, harm, and power because of the more intense social and personal interactions involved. For feminist researchers, in depth interviews have been advocated as appropriate feminist research methods because they allow women to tell their stories in their own words. But they have also been criticized as unethical in exploiting gender stereotypes and women's vulnerabilities. Similarly, it is acknowledged that interviewing couples (either separ ately or together) may create tension for the couple. Qualitative researchers also acknowledge that partici pation in research may involve emotional responses, and that ethical conduct involves attempts to "manage the research such that the emotions that are provoked do not harm the research participants or the researchers or the research itself" (Ansell and Blerk, 2005: 72). Across these examples, a balancing of harm and benefits is advocated as an appropriate ethical response.

Privacy, confidentiality, and anonymity

There is wide agreement in human geography and in legislative and professional codes of ethics that people participating in research projects have the right for their personal details to remain private. Research participants are guaranteed anonymity, that researchers will not release their details into the public domain, and that the publications resulting from the research will not allow them to be directly or indirectly identified. Pseudonyms for informants are almost universally used, and increasingly common is the masking or omission of other identifying characteristics such as specific occupation, age, and family circumstance. Human geographers' ethical stance in relation to the anonymity of well known people, and/or those who represent public organizations or corporations tends to depart from the principle of anonymity. Research using interviews with CEOs of large corporations, for example, has identified informants on the grounds that permission has been granted and also that it is in the public interest for such individuals/corporations to be identified. This is more common in political and economic geography.

A peculiarly geographical take on anonymity is the debate on whether to identify the specific places that are being studied. The uniqueness of place – such as neighborhoods, cities, and villages – is often a key component of geographical research, and for some researchers places need to be identified to assure the relevance and validity of the research findings. For others, identifying places is ethically problematic for a number of reasons. Knowing the location is one way of identifying people, especially if it is a small place or a commercial establishment such as a restaurant. Thus, in Crang's study of work cultures in an English restaurant, he gave a pseudonym to the restaurant and did not identify its exact location. Another argument is that the identification of places characterized by high levels of poverty, crime, or other elements of social marginalization is not ethical because of the possible consequences of such naming. Naming these sites may work to further disadvantage already marginalized groups and places, and thus it is not uncommon for studies in human geography to mask the identity of such research locations.

Privacy and confidentially take somewhat different forms in quantitative research that uses large scale secondary data at a fine geographical scale. Publicly collected data such as the census, labor force surveys, and travel surveys are typically coded for geographical location. The agencies responsible for the collection and release of these data are concerned that individuals may be identified when details at small scales are made available. The most common response to maintain privacy is that cells in data matrices are sometimes randomized to avoid such identification.

Recent technological advances have made privacy concerns more acute. In GIS, the processing and linking of a number of different datasets may enable individuals to be identified, thus losing their privacy. Remote sensing capabilities, and in particular, the increasing prevalence of unobserved satellites that repeatedly take images of the same location similarly infringe privacy. Rather than ban the use of such data, various masking techniques are advocated as a solution to the privacy concerns raised.

Governance of the Ethical Basis of Research in Human Geography

Who ensures that geographical research meets agreed ethical standards? Four major answers to this question have been given within human geography, with the answer chosen dependent on the historical and geographical circumstances as well as the researcher's ethical stance. The first two means of making certain ethical research are informal, while the last two are legislative and institutional. One answer is that oversight of the ethical conduct of human geography research rests with individual researchers and a community of peers; that researchers' capacity to act in ethically defensible ways is produced and guaranteed by their position as professionals and by peer review processes. Research processes undergo various processes of peer review – by funding bodies (e.g., Economic and Social Research Council and National Science Foundation), external collaborators, doctoral committees, and research supervisors, and these processes include review of ethical conduct. The argument is that these procedures ensure that unethical research will not receive funding or endorsement. In qualitative research, a particular peer review guarantee of ethical conduct is advocated: 'critical reflexivity'. Critical reflexivity, as summarized by England, requires researchers to be alert to the social and power relations that are enacted during qualitative research, and to make attempts to change the nature of these interactions during the research process. There have been criticisms of critical reflexivity, including the impossibility of total self reflection, its inability to ameliorate power relations, and its limitations in ensuring ethical conduct. Nonetheless, it has attracted support as a key ethical stance in an increasing cross section of the discipline, including social and cultural geography.

A second form of oversight of ethical research practice, and one that became more pronounced in the late twentieth century, is the guidance and guarantee of ethical conduct by codes formalized by professional associations. For example, the British Sociological Association and the American Sociological Association have developed 'statements of ethical practice' to guide the ethical conduct of research. In 1998 (and with minor revisions made in 2005), the Association of American Geographers (AAG) published its own 'statement on professional ethics', which included guidelines for the conduct of not only research, but also for relations with colleagues, students, and the broader community. The AAG statement is explicitly nonjudgmental and nonprescriptive. For the association, the statement:

> should be seen as the beginning of an ongoing, collective project aimed at promoting reflection on ethical matters and ethical accountability in geography. It is designed … to encourage consideration of the relationship between professional practice and the well being of the people, places and environments that make up our world. (American Association of Geographers, 1998: 8)

Common to such statements is their lack of binding force over the conduct of research; their efficacy is instead achieved through moral suasion and voluntary compliance.

Legislative changes in the late twentieth century, coupled with more legalistic research environments across nations in which human geography flourished, saw the formalization of ethical review. The third and fourth guarantees of ethical conduct in human geography are of this type. One is statements on ethical research by government agencies charged with the oversight of research. In Australia, for example, the guidelines of the National Health and Medical Research Council are binding over all university and other research and similar guidelines exist in other countries as well. These statements provide both general guidance and prescriptive rules about research conduct, consent, harm, deception, etc. They are universalist in nature, prescribing ethical practices based on a set of moral standards that are generally not context specific. They do, however, recognize the specificity of cross cultural research and research with other marginalized and vulnerable groups, and include specific rules for any type of research with these groups.

A fourth type of ethical review, and one that has attracted most recent attention from human geography is review by human ethics committees. Most human geography research falls under the auspices of university ethics committees, though many human geography research projects will also need to engage with the ethics committees of hospitals, schools, charitable associations, or indigenous organizations. These committees differ from peer review, in that they are typically comprised not only of researchers but also of lay people and sometimes representatives of religious organizations. University and other institutional ethics committees have two main roles. The first is an educative role, with most having a charter to provide guidance and information on research ethics, which often include concrete examples of ethical research practices. Second, these committees have an evaluative role, charged with assessing the ethical issues raised by research projects and specific methods, and judging the adequacy of proposed measures for addressing these issues.

Review of geographical research by such committees can be wide ranging, encompassing many of the issues outlined in the opening section of this article, though typically the social purpose of research is not evaluated. They explicitly consider relations with research participants, and in particular, informed consent, harm, privacy, anonymity, and reciprocity (see previous section). These committees also consider additional issues. Potential conflicts of interest, especially in relation to the funding or sponsorship of research are of particular concern. These committees may also consider the personal safety of the researcher.

Just as definitions of ethical methods and ethically sound geographical research are far from settled, the place of institutional ethical review remains the subject of some debate and sometimes intense criticism. The universalist ethical stance underpinning these codes has been the most frequently identified problem, because it is at odds with the situational ethics more commonly used in human geography. In applying and enforcing universal rules, it is at best difficult, and at worst, impossible, for institutional ethical review to be situation specific. The validity of medical–scientific models, as the basis for evaluating the ethical aspects of research social sciences such as human geography, is also questioned. This criticism more commonly emerges from human geographers using qualitative methods, who are critical of the medical scientific assumptions of objectivity and disembodied researchers. Nonetheless, institutional ethical review is likely to assume greater rather than lesser importance in the future of human geography. Attention has, hence, most recently turned toward ways of critically engaging with and transforming institutional review.

Summary

What is ethical research in human geography is far from a settled question, and is unlikely to be so in the near future. What is settled is that all methods involve ethical questions, and that answers to these questions are dependent on the broad philosophical framework of the research. Across the frameworks, however, there is agreement on three important issues. The first is that ethical issues can and do arise at any stage of the research process – in formulation of research topics, choice of methodology, interactions 'in the field', analyzing results, and in writing. The second is that ethical issues rarely have simple solutions and require ongoing critical reflection. Finally, responses to ethical dilemmas in human geography are just as likely to involve a balancing of potential harm and benefit as they are the application of universal ethical principles.

See also: Critical Geography; Critical GIS; Cross-Cultural Research; Ethnography; Interviews: In-Depth, Semi-Structured; Quantitative Methodologies.

Further Reading

American Association of Geographers (1998). *Statement on Professional Ethics*. http://www.aag.org/Publications/Other%20Pubs/Statement%20on%20Professional%20Ethics.pdf Accessed on Jul. 2008.

Ansell, N. (2001). Producing knowledge about Third World women: The politics of fieldwork in a Zimbabwean secondary school. *Ethics, Place and Environment* 4, 101 116.

Ansell, N. and Blerk, L. V. (2005). Joining the conspiracy? Negotiating ethics and emotions in researching (around) AIDS in Southern Africa. Ethics, *Place and Environment* 8, 61 82.

Crang, P. (1994). 1994 It's showtime!' On the workplace geographies of display in a restaurant in South East England. *Environment and Planning D: Society and Space* 12, 675 704.

Dowling, R. (2005). Power relations, subjectivity and ethics in qualitative research. In Hay, I. (ed.) *Qualitative Research Methods in Human Geography* (2nd edn.). Melbourne, VIC: Oxford University Press.

England, K. (1994). Getting personal: Reflexivity, positionality, and feminist research. *Professional Geographer* 46, 80 89.

Kitchin, R. and Wilton, R. (2000). Disability, geography and ethics. *Ethics, Place and Environment* 3(1): 61 102.

Kwan, M.P. (2002). Feminist visualization: re envisioning GIS as a method in feminist geographic research. *Annals of the Associations of American Geographers* 92(4): 645 61.

Howitt, R. (2005). Human ethics, supervision and equity: Ethical oversight of student research. *Journal of Geography in Higher Education* 29, 317 320.

Israeil, M. and Hay, I. (2006). *Research Ethics for Social Scientists*. London: Sage.

McLafferty, S. (2004). Women and GIS: Geospatial technologies and feminist geographies. *Cartographica* 40, 37 46.

Meth, P. and Malaza, K. (2003). Violent research: The ethics and emotions of doing research with women in South Africa. *Ethics, Place and Environment* 6, 143 159.

Mitchell, B. and Draper, D. (1982). *Relevance and Ethics in Geography*. London: Longman.

Pavlovskaya, M.E. (2006). Theorizing with GIS: A tool for critical geographies?*Environment and Planning A* 38(11): 2003 20.

Proctor, J. D. and Smith, D. M. (1999). *Geography and Ethics: Journeys in a Moral Terrain*. London: Routledge.

Wilton, R. (2000). Sometimes it's OK to be a spy: Ethics and politics in geographies of disability. *Ethics, Place and Environment* 3, 91 97.

Relevant Websites

http://www.aag.org
American Association of Geographers.

http://www.nhmrc.gov.au
NHMRC news and events.

http://pre.ethics.gc.ca
Selected human research ethics norms, organizations and resources.

Ethnic Conflict

O. Yiftachel, Ben Gurion University, Beer-sheva, Israel

Glossary

Diaspora An organized ethnic community of long-term residence away from the historical homeland.

Ethnicity Group identity based on belief in a common past and future, a shared culture, and association to a specific place.

Ethno-Class A social position based on a combination of ethnic affiliation and economic position.

Ethnocracy A regime under which a dominant ethnic group appropriates the state apparatus and shapes most public policies.

Ethnonationalism Political mobilization aimed at achieving, consolidating, or protecting ethnic territorial sovereignty.

National Self-Determination The ability of a nation to shape its collective identity and determine its preferred form of government.

Spatiality The articulation of a human phenomenon in spatial terms, both discursive and material.

Introduction

The world's political map has been profoundly shaped by ethnic identities and conflicts over space. This article reviews and critically analyzes the political geography of ethnicity, focusing on the modern epoch. The main perspective advanced here is that the construction and evolution of ethnic identities are closely related to spatial conflicts. Hence, the very nature of ethnicity cannot be fathomed without 'opening up' its underlying geographical and political foundations. The article also reviews the historical construction of ethnicity; its relations with systems of spatial power, especially the colonial, national, and capitalist state; and recent transformations in the wake of accelerating urbanization and globalization.

Defining Ethnicity

Ethnicity gained prominence during the 1970s, with the growing interest in social diversity and justice, and has since become one of the central questions of contemporary social science. Ethnicity is based on the Greek root 'ethnos', meaning 'blood connection', as distinct from the 'demos', which was a territorial–civil association. Over the centuries, ethnicity changed its meaning, alternatively being associated with tribal, regional, religious, class, and national affiliations.

In this article, as in the majority of contemporary writings on the subject, ethnicity will be defined as group identity, based on common cultural affiliation and a belief in a shared ancestry and a common future. Ethnicity is predominantly linked to specific places, either existing or mythical, as a central pillar of identity construction. Ethnic identities – like all social constructs – are fluid and malleable, tending to rise in social and political importance during periods of conflict, and wane during times of peaceful ethnic coexistence.

It is useful to sharpen the above definition of ethnicity by highlighting two ideal types: 'immigrant' and 'homeland'. The purpose is not to create a (false) dichotomy of ethnic experiences, but to highlight dimensions of the phenomenon, which help explain the wide variety of ethnic groups, their boundaries, and goals.

Immigrant ethnicity is based on a long standing distance from the homeland, and is typically characterized by a gradual process of integration and assimilation into the host society, although the extent of this process varies greatly between groups. Immigrant ethnic groups commonly struggle for civil rights and equality in the allocation of public resources. The gradual assimilation and mobilization of most immigrant groups into host societies has often transformed their identity into what sociologist Herbert Gans termed 'symbolic ethnicity'. Such identity maintains several cultural markers, such as holidays, religion, and food, which do not threaten the main values and practices of the host society, especially the state's cultural orientation, its legal, economic, and political systems. Symbolic ethnicity imposes negligible constraints on major life decisions of minority members, being expressed in their relative freedom on matters such as place of residence, marriage, work, politics, and leisure. Immigrant ethnicity thus generally marks a gradual erosion of the group's structural distinction from the host society and from other immigrant minorities.

On the other hand, homeland ethnicity is held by groups residing on the territory believed to be the 'cradle' of their identity and history. Drawing on the group's geographical continuity and the corresponding preservation of cultural and social boundaries, the distinction of the group is maintained, often quite intensely. Residence on homeland territory is often accompanied by the establishment of ethnic institutions (in areas such as education, culture, and religion) which reproduce cultural and social boundaries between the group in question and neighboring groups. Needless to say, homeland ethnicity is also subject to change, but its relative stability has been a source of cultural and political power. This power has been used extensively to transform homeland ethnic groups into national movements, by making inseparable the links between collective identity, power, and land, as powerfully illustrated by the (now replaced) national anthem in pre 1994 South Africa.

The Call of South Africa

Out of the blue heaven, out of the depth of our sea
Over our eternal mountains, where the cliff tops echo again…
We will live, we will die   we are yours, South Africa…
In the marrow of our bones, in our heart and soul and spirit
Fatherland we will carry the renown of your name with honour
True and faithful as Afrikaners   children of South Africa

Yet, as the South African case demonstrates, what appears to be an 'eternal' identity is actually reshaped ceaselessly by political events, public discourses and practices, material processes, and cultural trends. It cannot be analyzed in isolation from other social relations, because it is constantly mediated by gender relations, class affiliation, politics, and religious practices. Ethnic identities are often perceived as 'threatened' by the dynamics of assimilation, globalization, and immigration. Hence, the construction of ethnicity requires a constant practice of boundary maintenance, often achieved through social conflict, territorial struggles, and racism. The tension between the fluidity of identity and the attempts by social powers (particularly national states and cultural elites) to maintain and deepen ethnic boundaries is a notable hallmark of ethnic identity and societal structure in the modern era.

Colonizing Ethnicity

Modernity has brought about fundamental changes in the nature of ethnicity. As philosopher Charles Taylor, sociologist Anthony Smith, and anthropologist Ernest Gellner explain, the political and spatial reorganization of human society (marked by urbanization, industrialization, nationalism, and colonialism) have profoundly broadened and politicized cultural identities.

European colonialism, which saw the conquest and control of most of the globe by between the sixteenth and twentieth centuries, had a central role in this transformation. In parallel, the modern nation state was created in Europe, elevating ethnicity to the level of state ideology, and fueling the stratification of colonized identities, as discussed in the next section. The colonial period is critical because it set the scene for ethnic relations and conflicts to the present day. European colonizers did not simply attempt to control and co opt local populations, as commonly practiced by earlier empires, but to comprehensively manage their identity, economy, and geography. European colonial rulers, aided by modern technologies and new forms of deadly violence, systematically registered and classified conquered populations as both ethnically (and often biologically) different and inferior. This created the foundation for centuries of systematic forms of global racism, based on the putative superiority of the white 'race'.

New geographies lay at the heart of the colonial period. As shown by geographer Derek Gregory, this was both physical and mental. Colonialism rearranged colonized societies by facilitating European settlement and exploitation of local labor and resources, destroying and displacing local societies and their living environments, augmenting local urbanization, and introducing urban and regional planning. In addition, colonized regions were exploited for geopolitical military reasons, and 'modern nation states' were carved out of a pre colonial diverse and complex political space, particularly in Africa, Asia, and the Middle East. These measures created a geography of long term ethnic conflict, whose ruptures and dislocations still fuel tensions across the globe, in locations such Peru, Rwanda, South Africa, and Sudan, through Lebanon, Iraq, and Israel/Palestine, to Pakistan, India, and Malaysia, to name just a few examples. Moreover, colonial geographical practices were accompanied by a discursive mental framing, which 'scientifically' rationalized European conquest as modernizing a wide range of 'primitive' tribes and nations through the imposition of European specific and predominantly Christian values.

Colonial rule, framed as it was within the rising of global capitalist economy, also typically created patterns of uneven development and long term structures of economic dependency between the metropolitan (European) economies, and the colonial peripheries, as well as class stratification within the colonies. It was common for colonial rulers to sharpen, and often invent, ethnic differences within colonized territories, with the upwardly mobile local groups typically undergoing a process of Europeanization, later conceived as Westernization. The formalization of these identities as 'ethnic' (or 'racial', or 'religious', or 'sectarian'), institutionalized social difference around the notion of ethnicity. To be sure, the colonial influence also brought with it ideas of democracy, equality, and citizenship, but these egalitarian concepts seem to have often remained secondary to the manipulation of primordial sentiments and mobilizations.

The remaking of ethnicities during the colonial era was thus couched in patterns of cultural (European) domination. It established the foundation for Western control over what became the 'Third World', generating pervasive anticolonial insurgencies, resulting in the liberation of many colonized societies. Widely documented and analyzed in postcolonial studies, European colonialism has left deep footprints in postcolonial societies, often generating new rounds of ethnic conflict due to the prevailing logic of partition and state structures putatively 'belonging' to a dominant ethnicity. As noted, this idealized political geography imported and imposed European concepts of bounded ethno national identities, and forcefully superimposed them over a complex, dynamic, and fluid space of intertwined identities in the colonized world.

Another conspicuous legacy of colonialism has been a mass immigration from ex colonies to metropolitan centers, such as London, Paris, and Amsterdam. This postcolonial 'ripple' has recently destabilized the political geography of 'metropolitan' European and North American societies, and unsettled the image of their own nation state. Yet, the evolution of ethnic identities and conflicts cannot be solely attributed to European colonialism, as they have re emerged in different guises during the postcolonial period. New forces have propelled the reformulation of ethnicity, including globalization, increased internal and international migration, economic restructuring, the rise of indigenous peoples, and internal political struggles for control of the nation state, to which we now turn.

Nationalizing Ethnicity

The modern nation state has emerged as one of the most important loci of power during the modern epoch, if not the most important. As analyzed by political geographer Peter Taylor and legal geographer David Delaney, the state developed into a powerful 'container' of compartmentalized human societies, by combining the increasingly important political principle of self determination with highly effective forms of human territoriality.

The demarcated political territory was constructed as a 'sacred homeland', belonging to a people defined generally in ethnic terms; that is, a cultural group holding a myth of common origin and kin relations, and a long historical affiliation with 'its' territory. This sentiment was echoed in the famous call by Chancellor Otto von Bismarck during the nineteenth century attempt to unite Germany: "Germans – think with your blood".

In immigrant and settler societies the emphasis was naturally different, but also centered on the framing of cultural cum ethnic identity, through the promotion of notions such as 'creed', 'founding fathers', and 'state ideology'. These were translated into the application of assimilation ('melting pot') policies, designed to create a homogenous nation.

On the state level then, ethnic identities were widely transformed and institutionalized into state nationalism. Groups thus politicized and mobilized their members by strengthening, recreating, or inventing the historical connection of culture, people, and territory. The rise of nationalism began mainly in Europe and, as noted above, was associated with European colonialism, during which cultural and 'racial' identities became widely studied, formalized, and stratified. There were cases in which genuinely equal citizenship developed through the promotion of civic nationalism, but until the middle of the twentieth century, these were relatively scarce.

The popularity of the nation state order grew steadily, and reached a peak during the Interwar period (1918–39), when it was introduced (often imposed) to eastern parts of Europe and on large parts of the colonized continents. Proponents of this order argued that the division of the world into a mosaic of discrete nation states ('a state to each nation, and a nation to each state') would enhance stability, democracy, and prosperity. However, as detailed below, this system has often exacerbated, rather than abated, ethnic conflict.

The association between ethnic nationalism and conflict is often profoundly geographical, being linked to claims for exclusive control over mixed territories. For this kind of politics, then, national territory is a zero sum resource, which unlike other resources (e.g., economic growth, culture, or environmental assets) cannot be shared. This has provided the locus of numerous ethno national conflicts, with rivaling claims for sovereignty over the same territory. To complicate matters, the meaning of territory is often related to fractured and contested histories, drawing legitimacy from periods of varying ethnic control over the territory in question. These multilayered histories cannot be easily accommodated in the current uni dimensional political geography of state power, generating conflicting claims, aggressive wars, and ethno national tensions. The recent examples of Kosovo, Bosnia, Northern Ireland, Israel/Palestine, Tamil Elam, the Baltic Soviet States, and Kashmir, all illustrate the enduring potency of ethno national territorial disputes based on fractured histories and contested visions over the same territory.

The geography of ethnic conflict is closely linked to the political concept of self determination, which has been widely accepted as a legitimate political expression of peoplehood. It has thus provided an important organizing principle for world politics over the last century, by constructing a desired state of affairs whereby each nation can exercise its right for self determination on 'its own' homeland territory.

Self determination created a direct link between identity, territory, and power, and has politicized the notion of the 'homeland'. It had the effect of fueling, and even creating, political aspirations among many ethnic groups. The spatiality of the self determination principle has been critical to this development, as the attachment of the group to 'its' land has developed into a highly valuable asset in the pursuit of state power. The proliferation of ethnonationalism is forcefully illustrated by the constantly rising number of UN member states – from 50 in 1945 to 192 in 2007.

Marginalizing (Minority) Ethnicities

The dominance of the nation state order encountered an enduring complication – the persistence of minorities. Nationalist ideologies were commonly premised on the disappearance of substate ethnic differences, or alternatively on their permanent marginalization. This was augmented by the rising influence of capitalism and liberalism, both focusing on the individual as a producer, consumer, and political agent, thereby drawing attention away from the identity and plight of minorities.

The seemingly 'natural' connection between ethnicity, homeland, and nationalism often involved the appropriation of the state by a dominant ethnicity which usually crowned itself as representing the 'true' national spirit. This gave rise to a range of homogenizing and domineering policies. However, this approach also created 'trapped minorities' – those facing discrimination due to their identities, but at the same time prevented from exercising their own self determination in the state's homogenizing public sphere. The 'entrapment' of minorities, and the ensuing conflict is one of the most significant legacies of state nationalism to this very day.

This constellation led to continuing attempts to control and marginalize minorities. Geographers such as Alexander Murphy, David Sibley, and Oren Yiftachel, and sociologist Michael Mann have documented the widespread implementation of discriminatory policies. These may take many forms ranging from a denial of recognition of distinct minority cultures, such as practiced in France, Turkey, or Germany, to more overt discrimination (also termed 'derogatory recognition') evident in 'ethnocratic states' such as Israel, Malaysia, and Estonia.

In extreme cases, oppression and conflict may lead to ethnic cleansing, as occurred in Cyprus, Palestine, Bosnia, or Sudan; and even to genocide, as horrifically witnessed during the Nazi extermination of Jews, and in more recent times in Cambodia and Rwanda. The common denominator has been the constructed 'eternal ownership' of the homeland by the dominant ethnic nation and the questionable status of minorities. The concept of an exclusive homeland can also cause a 'chain reaction' and mobilize minorities against one another. The logic of ethnic spatial purity is exemplified by the following quotation from an ancient Sinhalese text, the Pujavaliya, often used in the current ethnic struggle in Sri Lanka to enshrine the dominance of the Buddhist Sinhalese over other communities residing in the island state.

> This island belongs to the Buddha himself… Therefore, the residence of wrong believers in this Island will never be permanent… Even if non Buddhist ruled Ceylon by force for a while… Lanka is destined to be ruled by Buddhist kingdoms.

But the oppression of minorities is not universal. Some states have opted for maintaining a fragile status quo of majority domination without overt oppression. Other more democratic states have introduced policies of accommodation, as documented by the thorough works of political scientists McGarry and O'Leary. These policies have ranged from accepting minority driven partition, as in the case of Malaysia–Singapore, Czechia–Slovakia, and some of the post Soviet states, or more commonly introducing power sharing and forms of spatial coexistence embedded in the concept of 'consociationalism' or 'new federalism'. In recent times a more open ended approach of multiculturalism has also gained prominence, attempting to base cultural and regional policies on recognition of the identities and aspirations of minorities. These politics of recognition have been most evident in immigrant societies such as Canada, New Zealand, and Australia, and also increasingly in policies toward minorities in Europe, Asia, and Africa.

Table 1 summarizes the spatialities of three main approaches to policies toward ethnic minorities – oppression, status quo, and accommodation. These are not mutually exclusive, and the table treats them as 'ideal types' for analytical purposes. It is acknowledged that in reality, policies and spatialities are more complex and multilayered.

Policies toward minorities, however, have never remained static. In recent years, following evidence that ongoing oppression of minorities generally spawns resistance and possible insurgence, many multi ethnic states have increasingly introduced accommodation policies. Despite this move, other states continue to oppress and marginalize minorities, thus maintaining a high level of ethnic conflict.

Urbanizing and Globalizing Ethnicity

Urbanization has been one of the major geographical transformations of the modern era. In 2005, for the first time in history, more people resided in cities than in rural areas, while in the West the figure is around 90%. Needless to say, urbanization means more than a mere spatial transformation, harboring profound economic, cultural, and political implications. These have intensified in recent years with the acceleration of globalization, which elevates the scale of human activities beyond the state. Cities are vital connecting nodes in this process, as they facilitate the actualization of globalization processes (known as 'regional integration' or Europeanization)

Table 1 Policy approaches to minority ethnic spaces a schematic presentation

Ethnic space	*Policy options*		
	Oppressive	*SQ*	*Accommodation*
Body	Forbid/marginalize	Apathetic	Facilitate
Home	Demolish/confine	Confine	Facilitate
Property	Reduce value	'Market forces'	Develop
Neighborhood	Undermine/confine	Restrict control	Indifference
Ghetto/enclave	Create/preserve	Preserve	Protect/open boundary
Turf	Undermine	Restrict	Facilitate
Sacred Sites	Minimize	Condone	Expand
Shanty town	Criminalize	Ignore	Selectively legalize
City	Majority hegemony	Maj. control	Power sharing
Land	Expropriate/restrict use	Little change	Proport. allocation/restitut'n/compent'n
Region	Deny, undermine	Restrict	Facilitate
Territory/homeland	Deny	Deny	Autonomy
Diaspora	Sever links	Restrict links	Indifference

through their economic and cultural roles. The impact of urbanization on ethnic identities has been profound. It has changed the manner in which ethnic boundaries are constructed, the nature of interaction with other groups, and the characteristics of political mobilization.

Cities constitute the hubs of economic, political, and cultural power, and the sites in which development and globalization actually take place. As such, they are vitally important for understanding the dynamics of ethnic identities and conflicts. Notably, urban ethnic dynamics are shaped by opposing and simultaneous forces, en hancing both assimilation and fragmentation. On the one hand, as shown by political geographer Peter Taylor, state regimes promoting national unity and economic growth saw the cities as critical for their project of creating a unified nation and economy. Urban sociologist Manuel Castells opened a new perspective by noting that cities became centers of collective consumption (for housing, education, health, and the like), thereby enabling the regime to integrate urbanizing populations into a rela tively homogenous collectivity.

At the same time, however, as highlighted by geog raphers and urbanists Jane Jacobs, Doug Massey, and Fredrick Boal, countervailing forces of fragmentation have been at work, making the city a springboard for a new political geography of ethnic mobilization and conflict. While in the rural areas ethnic identities have been relatively (though of course not entirely) stable, in the city, the identity sphere became more dynamic. In the rural areas the construction of identities was closely associated with national claims for territorial control, while in the cities issues relating to civil status, economic position, and service provision became more prominent.

This rural–urban difference should not be treated as dichotomous, but as a matter of degree and contingency. The relative openness, fluidity, and instability of cities – in terms of housing, employment, or lifestyles – is associated with a corresponding (relative) flexibility of politics and identities. In addition, due to their global izing economic and cultural roles – and unlike rural areas – cities are centers of intra state and international im migration, which constantly reshapes the nature of ethnic spaces and relations. Therefore, as noted by urban scholars such as Scott Bollen, Ira Katznelson, Don Mitchell, and Kian Tajbakhsh, the city possesses 'a promise' to transform identities and society, and creates new forms of spatial coexistence, even in cases of pro tracted conflicts. Katznelson thus notes:

> Cities are porous… Liberalism is inconceivable… without the urbanization of early modern Europe, which became the location for political emancipation, from racial, ethnic and religious oppressions…

Yet, structural forces such as expansionist or racist state nationalism, and uneven globalization and economic growth have often worked to thwart this promise. As shown by Oren Yiftachel these forces often produce 'ethnocratic cities', whereby dominant ethnic groups use the mechanisms of urban governance to further its power and material interests. Such cities are marked by high levels of coerced segregation, racial ghettoes, and per sistent conflicts, as shown by the examples of Paris, Jerusalem, Colombo, or Berlin. Significantly, powerful ethnic groups use the development and accumulation processes governed by urban policy to buttress their dominant class position. Minorities are typically left on the margins of urban based accumulation, leading to the demarcation of segregated urban 'ethno classes' as elab orated below.

In some cases, segregation may be largely voluntary, and acts as a form of conflict management. As shown by geographer Yizhak Schnell, this occurs mainly in cities accommodating groups separated by religious or racial affiliation, without explicit claims to a national control over space, such as Muslim groups in most European cities, Blacks and Hispanics in American cities, and ultra orthodox groups in Israeli cities. Conflicts abound in such cities, but their potential harm is reduced by patterns of voluntary segregation.

Ethnic relations are particularly volatile in cities divided on ethno national lines, most notably in urban areas occupying sites of collective symbolic value. Planning scholar Scott Bollen has shown how protracted urban conflicts in cities such as Belfast, Jerusalem, and Sarajevo, are associated with broader national claims for territory. These tensions are commonly reproduced by urban policymakers, working under the assumptions of their own national historical narratives. Yet, the influence is never uni directional, and urban policy can also aggravate, or appease, the larger national conflict.

Other types of ethnic and racial conflicts are documented by geographer Jenny Robinson, who claims that 'ordinary urbanism', not associated with sacred or national locations, also typically produces local conflicts, thereby sharpening ethnic, class, and local identities. Hence, in a wide variety of settings, the city gives rise to the formation of 'ethno classes', being partially porous social groupings based on a combined cultural and economic affiliation. Quite often, the more economically mobile sections of minority groups integrate and even assimilate into the mainstream and voluntarily lose their ethnic affiliation, or alternatively form a wide range of 'hybrid' identities.

The economically deprived parts of the community typically remain quite segregated from the mainstream and subsequently maintain higher degrees of ethnic difference. These segments are likely to face barriers to assimilation, stemming from their marginalized position in the labor and housing markets. This is often enhanced by popular and institutional prejudice and by a politics of identity – all working to maintain the ethno class boundary. As shown by geographers Ash Amin, in immigrant cities, ethnic identities tend to strengthen with economic marginality, causing alienation and protest based on the coexistence of difference and deprivation. This alienation erupts in times of mass protest and violence, as occurred in Paris's immigrant neighborhoods in late 2005, in northern cities of Britain during 2003, and Los Angeles in the early 1990s.

Urbanizing globalization has also mobilized diasporas as a significant political geographical factor. Improved communication technology and mobility have facilitated the maintenance of close contact with the homeland, and have coincided with the politicization of difference to create active diaspora communities. These communities tend to have a dual orientation: they struggle for their equality in the new location, and also maintain material, political, and emotional contact with their countries of origins. This setting creates new global, transnational networks, mediated through urban ethnic communities.

Another critical feature of today's geography of ethnic conflict is the emergence of mass urban informalities. As illustrated by urbanists Ananya Roy and Mike Davis, globalizing urbanization governed by neo liberal states often results in the coerced and rapid rural to urban migration, and in the subsequent creation of unplanned and un serviced shanty towns, usually at the edge of existing urban centers. In parallel, large disenfranchised and often rightless populations 'disappear' into the urban mass, typically as domestic and illegal workers. To illustrate the process, it may be enough to quote a 2006 report by the Chinese government, which notes that Beijing and Gunagzhou have each absorbed 3–4 millions people classified as 'floating' (unregistered and unauthorized) during the last two decades.

This process creates in effect a new situation which can be conceptualized as 'urban apartheid', where citizenship is constructed on the basis of people's origins, ethnicity, and class. Migrants fuel the urban economy with their labor and spending, but are prevented from attaining full urban citizenship, that is, full access to urban resources and representation. This emerging landscape of profound inequality forms the basis for enduring urban conflicts, and appears to sketch the political geography of ethnic relations for years to come.

Overview

As shown above, the making of ethnic identities has been inseparable from the geography of political and social conflict. Ethnicity has been politicized during the era of European colonialism, and institutionalized by the parallel emergence of the nation state as the backbone of the world's political order. Within this political geography, ethnic and cultural affiliations have become directly linked to the hierarchies of power and wealth. The nation state has both assimilated and marginalized minorities, generating conflicts and the sharpening of identities. In recent decades, accelerating urbanization and globalization have spawned new geographies of urban 'ethno classes'. While ethno national and regional struggles over territorial control continue to rage, a new terrain of urban ethnic struggles has emerged over issues of segregation, economic mobility, and urban informalities. The dynamic interaction of ethnic identities and human geography is therefore likely to shape world politics well into the future.

See also: Diaspora; Ethnicity; Immigration I; Immigration II; Nation; Nationalism; Nationalism, Historical Geography of; Political Geography; Segregation; Space I; Space II; Territory and Territoriality; Urbanization; War.

Further Reading

Amin, A. (2002). *Ethnicity and the Multicultural City: Living with Diversity*. London: Department of Transport, Local Government and the Regions; Government Printers.

Bollen, S. (2007). *Cities, Nationalism and Democratization*. London: Routledge.

Castells, M. (1999). *End of Millennium*. Oxford: Blackwell.

Delaney, D. (2005). *Territory: A Short Introduction*. London: Blackwell.

Jacobs, J. (1996). *Edge of Empire*. London: Routledge.

Mann, M. (2004). *The Dark Side of Democracy: Explaining Ethnic Cleansing*. Cambridge: Cambridge University Press.

McGarry, J. and O'Leary, B. (2004). *The Northern Ireland Conflict: Consociational Engagements*. Oxford: Oxford University Press.

Mitchell, D. (2003). *The Right to the City: Social Justice and the Fight for Public Space*. New York: Guilford.

Murphy, A. (2002). The territorial underpinnings of national identity. *Geopolitics* 7(2), 193 214.

Robinson, J. (2006). *Ordinary Cities: Between Globalization and Modernity*. London: Routledge.

Sibley, D. (1996). *Geographies of Exclusion*. London: Routledge.

Tajbakhsh, K. (2002). *The Promise of the City: Space, Identity and Politics in Contemporary Social Thought*. Berkeley: University of California Press.

Taylor, C. (1992). The politics of recognition. In Gutman, A. (ed.) *Multiculturalism: Examining the Politics of Recognition*, pp 25 73. Princeton: Princeton University Press.

Taylor, P. (2003). Radical political geography. In Agnew, J., Mitchell, K. & Toal, G. (eds.) *A Companion to Political Geography*, pp 47 58. London: Blackwell.

Yiftachel, O. (2006). *Ethnocracy: Land and Identity Politics in Israel/Palestine*. Philadelphia: University of Pennsylvania Press.

Ethnic Economies

L. Lo, York University, Toronto, ON Canada

Glossary

Affirmative Action Policies Affirmative action means positive steps taken to increase the representation of certain demographic groups such as women, ethnic minorities, persons with disabilities, and the queer community in areas of employment, education, and business from which they have been historically excluded. The goal is to counteract past and present discrimination sufficiently that the power elite will reflect the demographics of society at large.

Ethnoburb Ethnoburbs are suburban ethnic clusters of residential areas and business districts in large metropolitan areas. Created through deliberate efforts of an ethnic group within changing global/national/local contexts, ethnoburbs have their own internal socioeconomic structures, and at the same time are connected to the mainstream society through information exchange, business connections, and social activities. They contain some features of an enclave, and some features of a suburb lacking a specific ethnic identity. They coexist along with traditional ethnic enclaves in inner cities, but are rapidly replacing them as more important new 'ports of entry' for immigrants.

Institutional Completeness It is measured by the number and diversity of an immigrant group's social organizations. It is achieved whenever the ethnic community could perform all the services required by its members, and members would never have to make use of native institutions for the satisfaction of any of their needs, such as education, work, food and clothing, medical care, or assistance.

Middleman Minorities They are those ethnic entrepreneurs who trade in between a society's elite and the masses. In the past, they most commonly established businesses in poor minority neighborhoods deserted by mainstream retailers and service industries and business owners belonging to the dominant group. In recent years, they have been found to open up businesses in affluent and middle-class neighborhoods and have shown up in the primary sector of the mainstream economy.

Rotating Credit Association It is an association formed upon a core of participants who make regular contributions to a fund that is given in whole or in part to each contributor in turn.

Sojourning Orientation It refers to the tendency of those immigrants who have no wish to be permanent settlers but to eventually return to their place of origin.

Introduction

In an age of migration, ethnicity appears to have a role in transforming the social and economic landscapes of major cities around the world. At the turn of the twenty first century, immigrants or the foreign born account for about 12% of the total population in the US. The pro portion is even higher in Canada (18%) and Australia (22%). These immigrant populations are diverse. As an example, immigrants from some 169 countries lived in the Toronto Census Metropolitan Area, making Toronto the most multicultural region of Canada, if not the world, and one of the highest immigrant populated cities in the world. Often, different immigrant groups settle in dif ferent parts of the city. Their rates of self employment and/or participating as employers in the economy also vary. They may or may not contribute to visible ethno commercial landscapes known as Chinatown, Little Havana, Little Portugal, or Curry Mile. Employed or self employed, they may or may not dominate certain niches of the economy as Italian men in construction and Filipino women as domestic caregivers in Toronto, and Chinese and Indians in the computer industry in Silicon Valley. Indeed, recent decades have observed an ethnic division of labor in the paid labor market and a pro liferation of small and medium size businesses owned and operated by recent immigrants. Ethnic economies emerge and/or revitalize. As immigrant self employment rate rises above that of the native born, several questions ensue. Does ethnicity make a difference in the geog raphies of economic life? Why do some immigrant and ethnic groups develop thriving business activity and notable economic landscapes? Are the factors internal to the group or the result of context? This says that to understand the emergence of ethnic economies, we need to know the causes of ethnic entrepreneurship.

Origin of Interest in Social Sciences and Human Geography

The concept of ethnic economy is rooted in historical sociology and the thinking of some black Americans over a century ago. Classical sociologists such as Max Weber and Karl Marx believed that businesses in precapita list societies practiced favoritism, nepotism, communalism, and exceptionalism in their operations, and that their decision making was shaped by ethnoreligious relation ships rather than profit maximizing motives. This thinking called attention to the notion of ethnic capitalism, referring to marginalized ethnic minorities such as Jews and overseas Chinese who specialized in market trading in traditional societies. The notion of ethnic capitalism was specifically advocated by some black American thinkers such as Booker Washington and Malcolm X; they em phasized economic power over political power, hence business ownership, as strategies for black advancement.

The interest of social sciences in ethnic economies began in the early 1970s with the publication of *Ethnic Enterprise* in America by Ivan Light and the theory of middlemen minority by Edna Bonacich, both sociologists. Light's empirical work, comparing Chinese, Japanese, and black self employment activities between 1880 and 1940, emphasized the importance of social trust and rotating credit associations in explaining differential self em ployment rates, thereby introducing the notion of ethnic entrepreneurship and ethnic capital. The theory of middlemen minorities that discusses how internal co hesion or solidarity among members of an ethnic group can foster their own trading society provides the back bone for developing the concept of ethnic economy. The concept was formally introduced a decade later by Edna Bonacich and John Modell. In '*The Economic Basis of Ethnic Solidarity*', they refer to an ethnic group's self employed, employers, their co ethnic employees, and unpaid family workers as comprising an ethnic economy.

Geography was notably absent until the notion of enclave economy emerged. The idea of an enclave economy transpired in the early 1990s out of an ensuing debate caused by the work of Kenneth Wilson and Ale jandro Portes on Cuban immigrant workers in Miami. Whereas ethnic economy is derived from the theory of middlemen minorities, ethnic enclave economy origin ates from dual labor market theory used to explain per sistent inequality in employment of marginalized groups. An ethnic enclave economy stands for the economic advantage due to the locational clustering of businesses established by members of an ethnic group. The belief is that clustering allows higher monetary returns for par ticipants in an enclave economy than they would have been able to obtain without that enclave structure. The implication is that an ethic enclave economy manifests itself as a monopoly sector which recaptures co ethnic spending as a result of both vertical and horizontal in tegration along ethnic lines.

Ethnic enclave economy is a special case of ethnic economy. Spatial clustering is a major part of the attributes that make up an ethnic economy. This makes ethnic economy geographic in nature despite the recent begin ning of geographic research in this area. Geography has a long tradition in examining residential segregation and immigrant settlement/resettlement patterns. In a similar vein, just where immigrants work, where ethnic businesses locate, and how these relate to international migration, transnational capital, urban morphology, economic structure, and the location of other groups are of particular relevance to geography. This is especially so when we see that most taxi drivers in Ottawa appear to be Middle Easterners, whereas those in Toronto are South Asians; that housekeeping staff in North American hotels are mostly women from the Philippines, the Caribbean, or Latin America; that most street vendors in Lisbon are of African origin; that most corner convenience stores and dry cleaners in Toronto are run by Koreans; that signage in suburban shopping malls in Vancouver and Los Angeles is predominantly Chinese. This implies that the distri bution of people among different occupations, industries, sectors, and neighborhoods is not the outcome of a ran dom process of allocation. The underlying processes, complex as they may be, are inherently geographical, re flecting the ways immigrant groups live or experience in the places they now call home, where they come from, how they are represented, and the practices they use to overcome barriers to social and economic integration in their adopted home.

A Typology

Often the term ethnic economy is used rather loosely. It denotes the many different ways in which ethnic groups intersect with any sort of economic activity. The fol lowing summarizes what recent researchers come to terms with what has been a confounding literature.

Ethnic-Controlled Economy

An ethnic controlled economy encompasses a situation when co ethnic employees in a workplace, an occupation, or an industry in the general economy exert significant and persistent economic power such that they influence wages, hiring, and working conditions. It denotes an ethnic economy where workers have *de facto* control due to their number, their clustering, their organization, or government mandates. Examples include how Irish, Italian, and Jewish immigrants, respectively dominated construction, sanitation, and school teaching jobs in the municipal offices of New York City in the twentieth century, and how African Americans now dominate public sector jobs in the US as a result of government mandate to improve their economic chances.

Ethnic Ownership Economy

An ethnic ownership economy, distinctive from an ethnic controlled economy, consists of co ethnic self employed and employers, their co ethnic employees, and unpaid family workers. It designates a business and employment sector that coexists with the general metropolitan econ omy. This is the most generally known concept of ethnic economy and the focus of this article. To complicate things, there are, however, different types of ethnic ownership economies (or ethnic economy hereafter). The mix of ethnicity of their workers and clients, the kind of goods and services they offer, the market space they command, the locale they are situated in, and the degree of formality in such organizational behavior as capital fi nancing and hiring practices are criteria for distinguishing an enclave economy, an ethnic (non enclave) economy, and a mixed economy.

Ethnic enclave economy

An ethnic enclave economy is one in which all business owners, employees, and clients are of the same ethnicity. It largely serves the ethnic market with low order ethnic goods and/or mid order services. Firms are generally small and cluster around a territorial core. Informal economic practices are commonly alleged to be an in tegral part of an enclave economy. The most interesting aspect of an ethnic enclave economy is its spatial clus tering. The territorial core is usually where the ethnic group resides. Examples abound in Chinatowns and Greektowns. However, there are the rare cases where co ethnic residential concentration is absent. The Little India in Toronto, a commercial strip lined with South Asian restaurants, sari stores, and jewelers, is an example. While touted a tourist site, local South Asians living elsewhere in the metropolitan area still visit in droves on weekends.

Ethnic (non-enclave) economy

An ethnic economy differs from an enclave economy in two aspects. Businesses neither target co ethnic clients nor cluster spatially. They locate within and outside ethnic areas and serve as intermediaries selling cultural products to tourists or the general population. They may also sell non cultural goods and services to other underserved minority populations. The Chinese green grocers of an earlier era, Korean convenience stores, and Thai restaurants of contemporary times are examples of ethnic economy.

Mixed economy

A mixed economy consists of hybrid firms that utilize both ethnic and non ethnic resources and embody both formal and informal organizational behavior. Their markets are spatially and ethnically unbounded. Their businesses, not confined to general retailing, may include manufacturing, specialist retailing, and service provision for consumers, as well as producers. Mixed economies are outcomes of maturing ethnic economies responding to a host of globalizing/transnational, and political/ institutional forces. They can simultaneously host an enclave component and an integrated component.

The above describes ideal types. In reality, an enclave economy is a special case of an ethnic economy, and a mixed economy cannot exist without some traces of an ethnic economy. While no standard measures exist, knowledge of the size, composition, and market orien tation of ethnic firms are often used to infer the type of economy to which an ethnic economy belongs. In that sense, an ethnic ownership economy can be broadly re defined as an ethnic based economic structure consisting of a set of enterprises owned and managed by members of the same ethnic group, and running parallel to the larger metropolitan economy.

The Geography of Ethnic Economy

The role geography plays is twofold. One concerns the notion of spatial capital recently emerged as a factor for the creation and sustenance of an ethnic economy. The other involves the impact that ethnic economic activity has on the metropolitan landscape.

A geographical perspective looks at the relationship between ethnic economies and other geographically identifiable phenomena, such as residential segregation, underserved markets, or institutional distributions. It examines how location helps, hinders, or shapes ethnic economies. As discussed earlier, different types of ethnic economy fulfill different functions and have different spatial orientations. Firms that cater to co ethnics tend to congregate at its ethnic neighborhood. Otherwise, they can be more footloose. The spatial context shapes the practices and experiences of ethnic entrepreneurs, and the manifestation of ethnic economies.

Ethnic communities trigger our imagination partially because of the businesses they contain. We all know of the ethnic places that figure prominently in our cultural imagination. They often form a major component of a city's marketing emphasis. However, ethnic economic imprints go deeper than a few celebrated 'towns' and 'villages'. With ethnic spaces continuing to flourish under the guise of diverse groups and in unexpected locales, the implications of ethnic economies are more than visual. As immigrants become wealthier and more educated, ethnic economies expand to the suburbs and occur in multiple locales within the city. Suburban ethnic shopping malls, as in the case of Toronto and Vancouver, have been built to showcase Chinese businesses. Coexisting as they may be, these ethnic suburban malls have taken over their traditional downtown cousins as ethnic central places. These emerging landscapes speak to changes also in the way that neighborhoods in the city are structured and organized.

Ethnic groups have the ability to imprint landscapes with meanings in ways both evident and subtle. With architecturally distinct buildings, particular institutions, statuary or other markers, signage, language, and spe cialty goods which are ordinary to the insiders yet awed to the outsiders, ethnic businesses make and mark ethnic places. The place making quality of ethnic economies stamps an ethnic identity onto a district and creates a special landscape. To members of the group, an ethnic place is a place of residence or place of work where socialization is conducted in their own language. It reinforces their ethnic identity. To nonmembers, it is a tourist spectacle, a mystery of a distant land.

Ethnic economies operate at different scales, from local dependence to transnational linkages. Spatially, some are tightly bound and clearly defined. Others may concentrate over several square miles. Ethnic economies have also proven to be a dynamic and mobile phenomenon. Long marked as an ethnic enclave, the character and landscape geography of Little Tokyo in Los Angeles has altered over the years by spatial processes operating at a variety of scales. In recent years, with a new generation of wealthier immigrants, the role of Chinatowns is diminished by the emergence of ethnoburbs. In Los Angeles, these new concentrations of ethnic business depend to a large extent on the ability of ethnic banks to finance their development through both mortgage lending and commercial loans. In Vancouver, the transformation was sparked by transnational capital. In Toronto, both small local Chinese developers and big mainstream developers took the lead to change the city's northeastern suburbs into a landscape of Chinese malls.

Ethnic Economies and Ethnic Entrepreneurship

Ethnic economies are varied. They can range from a few restaurants and grocery stores mainly serving a specific ethnic group to one that has a full range of economic activities. Ethnic economies are sometimes visible in a single concentration. Others extend across several separate locations. Alternatively, ethnic businesses may be dispersed throughout the metropolitan area. For example, in Toronto, Korean owned convenience stores are all over the city despite the presence of a Koreatown in the city core. Similarly, Chinese owned restaurants serving Chinese, Japanese, Korean, Thai, Vietnamese, Malaysian, and European continental and fusion food are everywhere, although there are as many as five Chinatowns in the Toronto metropolis.

Hidden behind the spatial manifestation of ethnic businesses are the varying rates of self employment and entrepreneurship across ethnic groups. Both historical and contemporary records show considerable disparities in business formation among various ethnic groups. Today, high entrepreneurial groups in the US are Arabs, Armenians, Chinese, Greeks, Italians, Japanese, Jews, Indians, Pakistanis, Lebanese, Koreans, and Persians in that order; Blacks, Mexicans, Vietnamese, and Puerto Ricans are the least entrepreneurial. In Canada, the most entrepreneurial groups originate from Europe, including Jews and Italians, followed by those from the Middle East and East Asia. The least entrepreneurial are mostly visible minority immigrants from the Caribbean, Southeast Asia, and Africa. Ethnic entrepreneurs also appear sectorized, with Poles and Italians having a high stake in the manufacturing and construction industries; Korean, Chinese, Greek, Israeli, and Middle Eastern immigrants concentrating in the distributive services; and those from the US, Hungary, and South Africa focusing on business, personal, and public services. Immigrant self employment/entrepreneurship not only varies among ethnic groups, it also differs between men and women, with men more likely to be self employed. In Toronto, the proportions of self employment for immigrant men range from 37.1% to 9.7%, those for immigrant women range between 20.5% and 2.7%.

To understand these variations in form and function of ethnic businesses, we need to look at: Why do immigrants/ethnic minorities become entrepreneurs? Why is entrepreneurship more common among some groups than others? Why are some groups doing better than others who have entered the same country at the same time? How do different ethnic groups deal with the challenges of entrepreneurship? Answers to these questions are not independent of each other. The following explanations address the issues in varying degrees.

Cultural Characteristics

The 'cultural thesis' suggests that some ethnic groups are conditioned with certain characteristics that make their people more inclined toward business. These cultural characteristics can be the habit of savings, an appreciation of deferred gratification, the propensity to take risk, values often acquired through the process of socialization. They may also include norms that facilitate obtaining capital, sharing resources, or working together. With traditional middlemen minorities such as Chinese and Jews, the belief is their ancestral history provides them with a 'tool kit' that facilitates business development. In recent decades, the business immigration programs in Australia, Canada, and the US have brought in many migrants with business acumen, and these business immigrants tend to come from certain regions of the world like Hong Kong, Korea, and Iran. The general argument is that there are cultural differences across space. The dynamics of social relations and cultural processes channel certain ethnic groups more readily into business development.

Structural Push

Disadvantage or block mobility theorists argue that discrimination in the regular labor market imposes limitations on an ethnic group's economic mobility, and the disadvantage they face forces their group members into self employment or petite businesses. In the past, Chinese arriving in San Francisco and Vancouver and their des cendents were openly barred from professional practices and relegated to working in the gold mines or railway yards. Now immigrants arriving in Los Angeles, Toronto, London, and Sydney find themselves disproportionately concentrated in low status, low paid, and insecure jobs even though they are more educated than the native born population. This is because their foreign credentials are not recognized and employers often demand local ex perience. Contemporary devaluation of human capital across international space has the same effect as the place based exclusionary practices of the past. When certain groups are excluded from the upper echelon of the labor market, they seek self employment or entrepreneurship as a survival strategy or a defense mechanism.

Resource Mobilization

The ability to open and operate large and successful businesses depends on a group's entrepreneurial capacity. Entrepreneurial capacity is measured by the availability of ethnic and class resources. Whereas ethnic resources, generated as a part of normal group life, are features (e.g., solidarity, sojourning orientation, values, and traditions) of a group that co ethnics utilize in economic life or from which they derive economic benefit, class resources, in the form of cultural, human, and financial capital, provide the necessary cultural and material endowments to ad vance entrepreneurialism. Class and ethnic resources are ideal types, easy to separate on paper, but often em pirically intertwined. Development of ethnic resources is often linked to one's class position which affects one's ability to mobilize financial and/or human resources.

Ethnic resources

Ethnic resources, or the more trendy terminology of social capital, refer to ethnic ties and ethnic social networks. Close ethnic ties are important to those entrepreneurs lacking access to mainstream source of credit or assistance. A larger and more established ethnic/immigrant group provides the critical mass necessary to breed churches, mutual aid societies, rotating credit associations, business associations, and other types of ethnic institutions that provide formal as well as informal organizational resources for the group to get ahead than a small newcomer group. To the entrepreneur, ethnic social networks are sources of information, capital, and labor. They have been shown to be significant in the development of businesses among the Cuban immigrants in Miami and Korean immigrants in Los Angeles. Social networks are critical to starting and maintaining a business, especially small and low barrier to enter types. They are community and often neigh borhood based. A consequence of such a geographical process is that it perpetuates an inward outlook of ethnic businesses and strong competition among themselves.

Communities with a longer settlement history and with enclave economic activities promote community building and social capital formation. They are more likely to have a stronger ethnic network complete with their own set of institutions – cultural, religious, political, educational, and financial. These institutions, in return, support a more matured business structure moving be yond 'mom and pop' operations of earlier immigrant communities. These enclaves have historically tended to locate in traditional inner city immigrant reception areas in compact sociospatial units in which a large number of immigrants from the same ethnic background have clustered. In these areas may be found most of the im migrant community's cultural and religious institutions, as well as its services and businesses. Echoing Raymond Breton, we argue that given the number and diversity of these institutions and businesses, these groups have cre ated institutionally complete communities. This being said, however, institutional completeness as manifested by these groups was neither uniform nor static over time. In fact, the visibility and longevity of these enclaves is dependent upon not only the development of social networks, but also on the continuing construction of an institutional infrastructure.

Thus, varying levels of institutional completeness create differences among immigrant groups in the avail ability and access to neighborhood based and group based resources. Concentrated ethnic economic activities broaden the base for social interaction and provide a critical economic base for the immigrant community to function effectively. In contrast, institutionally incomplete communities make it more difficult for their members to gain formal credit and financing.

Class resources

There are two sides to class resources – material and cultural. On the material side, private property and wealth form the basis of financial capital; wealthy groups respond more successfully to opportunities for formal sector entrepreneurship than poor groups. Wealthy immigrants need not rely on informal co ethnic loans to start a business, and can build larger firms. Similarly, possession of human capital, in the form of education, destination country official language proficiency, and work/business experience, makes more effective entrepreneurs.

Cultural capital refers to the knowledge, skills, atti tudes, and values necessary to start and run a business in the market economy. It is an occupational culture trans mitted in the course of socialization both at home and in schools. Having an entrepreneur parent, working in a relative's business, or studying business administration at school contributes to the culture of entrepreneurship. Evidence shows that among all ethnoracial groups, human capital increases the rate of entrepreneurship. Class re sources have become increasingly important with the rising international migration of skilled workers, pro fessionals, and business people.

Opportunity Structure

Opportunity structure refers to market conditions and access to ownership. Business opportunity is affected by the local social–economic environment, as well as the larger political–institutional environment.

On the socioeconomic side, market conditions may be 'open' or 'closed'. An open market, which allows immi grant businesses to serve outside its own ethnic com munity, offers more entrepreneurial opportunities than a closed market serving cultural products to the ethnic community. Access to business ownership is largely de termined by the number of business vacancies, the degree of competition for those vacancies especially from the native born population, and government policies regu lating immigrant entrepreneurs. It is also impacted by the lack of interest from the native born population in cer tain businesses. Government policies setting the terms on which immigrants enter, deciding where in the country they settle, and affecting the ease with which businesses can be started affect the rate of immigrant entre preneurship. In this regard the treatment of Turkish workers in Germany as temporary residents, and the Dutch requirement that prospective entrepreneurs must show a need for their business, has a very different effect on entrepreneurship from the business immigration programs in Canada and Australia.

Immigrants' abilities to establish firms are determined not only by the extent to which they are legally permitted to set up a business, but also by factors such as taxation system, labor relations, residency status, job quality, health and safety laws, as well as their enforcement. In this re gard, the larger politico institutional context includes an array of regulatory structures and institutional contexts – not just those targeting immigrants – that promote certain economic activities and inhibit others. Immigrant entre preneurship may be higher in France because non citizens are excluded from most public sector jobs, yet lower in Canada because Canada has adopted affirmative action policies that do not differentiate between citizens and landed immigrants in all public sector hiring. Also, while virtually everyone can establish a private business in the US, Germany and Austria require the application of spe cial licenses even to sell flowers in restaurants and bars. Generally speaking, with the economic dynamics of a market in North America being different from those in Europe, these markets offer different opportunities and obstacles, demand different skills, and lead to different outcomes in terms of immigrant self employment rate and business success.

The Two Sides of Ethnic Economy

The self employed immigrants and their businesses con tribute to our metropolitan and national economies in several ways. First and foremost, ethnic businesses gen erate employment opportunities. Conservative estimates done in Toronto at the turn of this century found jobs created by the Somali ethnic economy, one of the smallest and most recent, accounted for about 7% of its labor force, whereas the Chinese with a longer history created employment opportunities equivalent to 34% of its labor force or 2% of the total labor force in the city. The job impact of ethnic economies cannot be ignored. Besides, another study found self employed immigrants who ar rived between 1980 and 1995 accounted for 1.7% of all self employed individuals in the country and generated 1.4% of the nation's self employment income or 0.07% of the national income. Given the recency of this group, this is by no means a small economic contribution especially as some of them also generate employment income for others and add to the public coffer through other forms of tax contribution. More importantly, as an ethnic economy grows, it provides jobs not only for some of its co ethnic workers, but also for the general labor force.

Although ethnic economies create jobs, there are often concerns about the quality of these jobs, in particular if they offer appropriate returns to human capital, and hence social and economic mobility. Many suggest that ethnic employers exploit especially their co ethnic em ployees, and that participation in the enclave ethnic market may prevent immigrant workers from ever reaching the general labor market. But evidence, while inclusive, shows that immigrant workers can move across ethnic boundaries and market sectors and away from the informal ethnic domain to the formal market that offers better working conditions and higher income.

On a more positive tone, we can view enclave em ployment as a way earlier immigrants help later arrivals to establish themselves. The advantage of such employment is that barriers such as non recognition of credentials, lin guistic limitations, and discrimination are much less likely to present problems than in the mainstream labor market. Even though the enclave labor market pays lower wages, workers earn more than they would if they were un employed. Labor market segmentation theory also says that immigrants, as one marginalized group, are often employed in dead end, low skill jobs in the secondary market. Even when employed, immigrant workers may end up with little career mobility. An enclave economy at least offers im migrant workers a protected niche with some opportun ities for upward mobility, including self employment. Where spatial clustering occurs, agglomeration economies are in effect. Enclave businesses like grocery store and movie rental clearly benefit from the concentration of potential customers. Co ethnic consumers, attracted to the neighborhood as a whole, bring in business. Distinct en clave economies such as Chinatown in New York or Curry Mile in Manchester are often tourist attractions in their own right and provide access to a wealthy tourist clientele.

As an ethnic economy expands, ethnic business organ izations emerge. In addition to promoting entrepreneur ship among their members and forging internal cohesion, the organizations often open trade links and promote business interaction between the communities where im migrants settle and those in their countries of origin.

Ethnic economies can be seen as contained and lead to social isolation and ethnic segregation. However, ethnic businesses can act as important agents for urban renewal, and help the neighborhoods to develop a place identity and revive their local economy. An enclave neighborhood collectively forms a tangible basis for the nurturing and maintenance of an ethnic identity. To be Chinese in Chinatown is to be reminded of one's 'Chineseness'. This in turn makes it more likely that one will continue to demand Chinese goods and services. In this way, ethnic enterprises collectively benefit from their visible presence on the urban landscape in ethnic places.

See also: Entrepreneurship; Ethnicity; Immigration I; Segregation; Segregation, Urban.

Further Reading

Barrett, G. A., Jones, T. P. and McEvoy, D. (1996). Ethnic minority business: Theoretical discourse in Britain and North America. *Urban Studies* 33, 783 809.

Bonacich, E. and Modell, J. (1980). *The economic basis of ethnic solidarity: Small business in the Japanese American community*. Berkeley, CA: University of California Press.

Collins, J., Gibson, K., Alcorso, C., Tait, D. and Castle, S. (1995). *A shop full of dreams: Ethnic small businesses in Australia*. Sydney: Pluto Press.

Hiebert, D. (2001). The spatial limits to entrepreneurship: Immigrant entrepreneurs in Canada. *Tijdschrif voor Economische en Sociale Geografie* 93, 173 190.

Kaplan, D. (1998). The spatial structure of urban ethnic economies. *Urban Geography* 19, 489 501.

Kaplan, D. and Li, W. (eds.) (2006). *The landscapes of ethnic economy*. Lanham, MD: Rowman and Littlefield.

Kloorsterman, R. and Rath, J. (2001). Immigrant entrepreneurs in advanced economies: Mixed embeddedness further explored. *Journal of Ethnic and Migration Studies* 27, 189 202.

Kloorsterman, R., van der Leun, J. and Rath, J. (1999). Mixed embeddedness: Informal economic activities and immigrant businesses in the Netherlands. *International Journal of Urban and Regional Research* 23, 253 267.

Light, I. (1972). *Ethnic enterprise in America: Business and welfare among Chinese, Japanese and Blacks*. Berkeley, CA: University of California Press.

Light, I. (2004). Immigration and ethnic economies in giant cities. *International Social Science Journal* 56, 385 398.

Light, I. and Gold, S. J. (2000). *Ethnic economies*. New York, NY: Academic Press.

Preston, V., Lo, L. and Wang, S. (2003). Immigrants' economic status in Toronto: Stories of triumph and disappointment. In Anisef, P. & Lanphier, M. (eds.) *The world in a city*, pp 211 225. Toronto: University of Toronto Press.

Rath, J. (ed.) (2000). *Immigrant businesses: The economic, political and social environment*. Basingstoke: Macmillan Press.

Rath, J. and Kloosterman, R. (2000). Outsiders' business: A critical review of research on immigrant entrepreneurship. *International Migration Review* 34, 657 681.

Waldinger, R., Aldrich, H., Ward, R. and Associates (eds.) (1990). *Ethnic entrepreneurs*. Newbury Park: SAGE Publications.

Zhou, M. (2004). Revisiting ethnic entrepreneurship: Convergencies, cocntroversies and conceptual advancement. *International Migration Review* 38, 1040 1075.

Relevant Websites

http://www.equalityhumanrights.com
Employment data on Britain are available from the Publications and Resources page of the Equality and Human Rights Commission website.

http://cmd.princeton.edu
Its Data Archive page contains data collected by The Comparative Immigrant Entrepreneurs Project, Centre for Migration and Development, Princeton University.

http://www.cis.org
Reconsidering Immigrant Entrepreneurship: An Examination of Self Employment among Natives and the Foreign Born by Steven A. Camarota, available from the Publications page of the Centre for Immigration Studies website.

http://www.statcan.ca
Statistics Canada's E stat System provides extensive data on migration, ethnicity and labor markets.

http://international.metropolis.net
The Immigrant and Ethnic Entrepreneurship Bibliography Database compiled by Jan Rath, University of Amsterdam is available from the Research and Policy page.

http://www.eeoc.gov
The US Equal Employment Opportunities Commission has Statistics for all US cities and states along the lines discussed for Los Angeles.

Ethnicity

W. Li and E. Skop, Arizona State University, Tempe, AZ, USA, and The University of Colorado at Colorado Springs, Colorado Springs, CO, USA

Glossary

Ethnic Groups Are defined as a unique set of individuals who share common interests based on mutual understanding, beliefs, and values, including a common homeland, a mutual language, a shared religion, and/or other contemporary and historical experiences.

Ethnic Identity Is seen as a fluid process by which individuals identify and are identified with a particular ethnic group based on key commonalities. Both primordial group ties and contemporary experiences affect how ethnic identities will be created, maintained, and manipulated in varying contexts.

Ethnoscapes Are those material manifestations of ethnicity that are reflected in the built environment, such as commercial enterprises, social and religious institutions, building styles, and streetscapes that signify a particular ethnic group. Oftentimes, particular places acquire emotional significance and symbolic importance for an ethnic group, who then consider the space an important component of their ethnic identities.

Pan-Ethnicity A term that connects the role of government policy and the general public in the rise and reconfiguration of ethnic identities. Local and parochial identities (Gujarati, Oaxacan, and Sicilian) give way to broader nationality groups (like Asian Indian, Mexican, and Italian), and even more general labels (like Asian, Hispanic, and non-Hispanic white). These umbrella classifications serve to ‘fit’ seemingly divergent groups into broader categorizations created for administrative and resource-allocation purposes. The process also facilitates the process of racial formation, since ethnicity and national origin become conflated with racial classification schemes. Minority or immigrant groups also maximize pan-ethnicity for economic pursuit and political mobilization.

Reactive Ethnicity A perspective that argues that ethnic identities are created dynamically in relation to the host society. Structural influences (including residential segregation, economic competition, and levels of discrimination) take precedence over primordial attachments as the source of ethnicity and ethnic identity.

Situational Ethnicity Results from multiple memberships in different social groups when an individual can identify himself/herself differently depending on the particular situation in question.

Symbolic Ethnicity The idea that individuals (especially third- and fourth-generation descendents of European immigrants in the US) voluntarily (re)construct the sense and feeling of belonging that ethnicity provides. Ethnic identity emerges as a nostalgic allegiance to the culture of the old country whereby individuals carefully and continually (re)choose traditions and ideals that serve them in varying contexts.

Introduction

Though there is no authoritative definition that is universally accepted, most scholars agree that ethnicity is a social classification that categorizes different groups of people by particular cultural characteristics. Ancestry is the most common identifier of ethnicity, though many other cultural features have been used to differentiate ethnic groups, including a shared homeland, language, or dialect; religious faith or faiths, traditions, values and symbols, literature, folklore and music, food preferences, social and political ties that transcend kinship, neighborhood, and community boundaries, and/or migratory status. While some researchers continue to assert the primordial origins of ethnicity, increasingly, most scholars agree that ethnicity is also the product of structural forces, social organization, and cultural representation, rather than solely the result of innate characteristics and hereditary factors. In other words, ethnicity derives from both an internal sense of distinctiveness and an external perception of difference. Because there is a complex relationship between ethnicity as an analytical concept and ethnicity as it relates to various forms of social relations, the term is subject to ongoing debate. Three main approaches have emerged in the enormous literature on ethnicity to understand how groups conceive, maintain, and manifest their ethnic identities: the culturalists see ethnicity as primordial and unproblematic, the structuralists regard ethnicity as strategic and intentional, and the constructionists view ethnicity as situational and dynamic.

Culturalists tend to view ethnicity as a primordial attachment to collective memories and a common cultural heritage. Particular cultural features (like religion, language, food, dress, and daily custom) are attributed to all the members of an ethnic group. As a result, ethnic groups can be easily identified and cataloged using key cultural features. Unfortunately, culturalists tend to treat ethnicity as unproblematic, stable, and inherent, whereby ethnicity acts as a unifying, powerful, and influential force in which all group members emotionally invest. Culturalists argue that ethnicity is persistent and pro vides structure, continuity, and meaning for members. Familial, economic, social, organizational, religious, and political traditions are maintained and recreated by members of the ethnic group through time and space.

The structuralist paradigm that gained popularity during the social upheavals of the 1960s challenged the idea that ethnicity is somehow fixed, natural, and in herited. Unlike culturalists, structuralists treat ethnicity as a reactive device utilized by ethnic groups in response to structural inequalities present in mainstream society. Structuralists argue that it is not enough to examine the internal 'cultural content' of ethnicity; more important are the external structural constraints in which ethnicity develops. The most critical structuralists argue that the staging of particular ethnic identities is more dependent on the benefits collectivization may bring. Ethnic group actions are based, not on some cultural affinity, but on the rational calculation of interest. Indeed, parochial loyalties (to linguistic–regional–religious subgroups) are dis regarded in the face of the competing claims and pres sures of other similarly constructed ethnic and racial groups. Ethnic solidarity is seen as being driven by eco nomic and political necessity, rather than as an inherent feature or natural bond of the ethnic group. The role of the individual is downplayed in the face of the more powerful claims of the group and group rights.

The constructionists scrutinize the uneven and dif ferential ways in which individuals sustain and assert their ethnic identities in their everyday encounters. Constructionists argue that culturalists and structuralists oversimplify the actions of individuals and society; they suggest that ethnicity should be understood as a complex dialectical relationship between ethnic group members and the social structures that surround them. Con structionists recognize that while individuals are active agents in defining their ethnicity, the process must be negotiated within a reactive, shifting environment. Con structionists propose that ethnic identity is situational and dynamic – in other words, ethnicity may be con ceived of differently by various members of the group where individuals may use the same ethnic label, but they construct their ethnicity based on shifting notions and interpretations of their cultural heritage. At the same time, ethnicity is not necessarily stable over time, and can be manipulated, altered, and transformed since ethnic categories are dependent upon broader social, economic, and political dynamics. Ethnicity, then, is a creative and complex response to both individual and societal forces. The formation and expression of ethnic identity is not a static process, but the result of both historical circumstances and individual negotiations to endow ethnicity and ethnic symbols with meaning.

The Role of Migration in Ethnic Identity Formation

Ethnic identity formation is oftentimes intertwined with the process of migration, since immigrants frequently arrive in a new place where the dominant society exer cises much power in defining difference. As they struggle not only to survive, but to prosper and build their lives with meaning and dignity, immigrants oftentimes utilize various ethnic attachments to create a sense of belonging amongst group members. Yet, because the host society utilizes its own ethnic categories to differentiate and label recent arrivals, with the country of origin typically serving as the most common delineator, the diversity that characterizes individuals when they emigrate (in terms of region, dialect, class, and religion, among other things) is disregarded for the categories that the host society uses to classify different groups. As a result of interactions with members of the dominant group and other minority groups in the host society, pan ethnic labels generally emerge and immigrants of disparate origin begin to unite around these simulated categories, primarily to protect their practical interests.

Once members of disadvantaged immigrant groups utilize ethnic collective action to rally against economic and political exploitation by the dominant group, the process becomes known as 'reactive ethnicity'. That is, the development of ethnicity may have its origins in a common heritage, but the persistence of ethnicity is the result of a group's attempts to withstand structural disparities. Immigrants (and their descendents) fashion pan ethnic identities to enable greater influence in the economic sphere and facilitate political organizing. The Irish provide a powerful historical example of the ways in which immigrants take on the ethnic/racial structure in the US and use it to their own strategic advantage. As a minority population victim to racial and economic op pression present in the US during the late nineteenth and early twentieth century, immigrants from Ireland created their own 'Irish' political identity to break through dis crimination barriers. With the help of the Democratic Party, and the incorporation of that group's pro slavery and racist ideology, the 'Irish' differentiated themselves from 'blacks', the other predominant minority population in the US. In claiming 'Irish ness', immigrants from Ireland facilitated their economic and political in corporation and also engaged in the process of reactive ethnicity.

According to the assimilation–acculturation model, eventually, distinct ethnic identities and ethnic attach ments should disappear as the descendents of immigrants (especially the second and third generations) take on the core culture and begin interacting with members of other ethnic groups, particularly on the intimate levels of friendship and family formation, without regard to ethnic differences. Ethnicity should remain important in only the most superficial ways after assimilation occurs, as individuals choose symbols (like ethnic foods and festi vals) that require little effort to recreate, and that do not interfere with other aspects of their lives. This symbolic ethnicity (as the process is known) emerges as a nostalgic allegiance to the culture of the 'old' country whereby individuals carefully and continually (re)choose those ethnic traditions and ideals that serve them in varying contexts. This idea appears to be valid with regards to ethnic groups that 'blend in' with the host society (such as white European groups that migrate to immigrant receiving countries with a white majority). But the reality is that groups that appear to be different (particularly in terms of physical attributes, religion, and class) must continue to rely on ethnic collective action because so ciety is organized and ruled along racial and ethnic lines.

Ethnicity and Race

It is impossible to talk about ethnicity without also talking about the issue of race. In the US, for instance, recent research encompasses the decisive role of official racial classification schemes in reinforcing race and ethnicity as meaningful social categories. Identified specifically is the key role of the state's recognition of race and ethnicity as the basis for incentive programs. The classification scheme created by the US government for administrative and allocation purposes in 1977, for instance, resulted in the consolidation of the 'ethno racial pentagon'. This 'ethno pentagon' not only institutionalizes particular ra cial/ethnic categories (including non Hispanic white, black, Hispanic, Asian, Pacific Islander, and American Indian), but also reinforces the racialization process, whereby racial meaning and racial categories are ex tended to previously unclassified groups. All non white groups are known as racial minorities. Take, for instance, the term Hispanic, which was formerly thought of as an ethnic designation, but which is now a separate cat egory alongside other 'racial' groupings. Or consider that extremely broad ranges of national origin, ancestry, cul ture, and physical types are now lumped together into seemingly arbitrary categories that nonetheless end up having very real consequences in people's everyday lives. And what happens to Middle Eastern Americans, or multiracial and multiethnic Americans, who reflect a wide variety of cultural and physical differences, but for whom a place in this official classification scheme is not entirely resolved? Similarly, in Canada many non white ethnic groups are classified as 'visible minorities' in the census; whereas in the United Kingdom until not too long ago, all non white groups were simply referred to as black.

Concomitantly, scholars argue that the emergence of equal rights legislation and affirmative action programs induces minorities to rally as groups to make them eligible for certain structural benefits. Thinking as a group be comes a reactive strategy to advance the interests and privileges of individuals. Caught in this system of ascrip tive racial and ethnic categorization, individuals come together to forge emotional links and alliances. Primordial ties do play a role in ethnic identity formation but within the context of the racialization process, which involves the sociohistorical dynamics by which racial categories are created, inhabited, transformed, and destroyed.

When individuals are defined as racialized minorities, the dominant group can exercise much power and priv ilege over individuals cast as the racialized 'other'. Some scholars even argue that given historical colonial legacies and contemporary Western hegemony, white European groups, regardless of where they reside, possess the power to either assert or ignore their racial identities (i.e., their whiteness) whereas other groups often do not have that privilege. Indeed, researchers note the considerable range of alternative identities that whites can enjoy, suggesting that race allows them flexibility in choosing their ethnic identities. Yet 'visible' minorities, even as they gain socioeconomic mobility, are subject to stigmatization and marginalization.

Ethnicity, Space, and Place

Within human geography, a significant literature has de veloped with regards to ethnicity, space, and place. Much of this research focuses on mapping the spatial distributions of racial and ethnic groups; whether by studying ethnic homelands or by analyzing the social and spatial assimi lation of ethnic minorities in a new destination – with special attention paid to the development of ethnic ghettos and ethnic enclaves. But recent critiques and theorizations have exhibited a greater sensitivity to the social con struction of ethnicity and the dynamics of racialization, with a particular focus on the sociospatial structuration of ethnic and racial communities and the spatialization of ethnicity and race in the contemporary world. Scholars articulate how various groups experience places as inclusive or ex clusive of them and others, and how the regulation of space reflects and reinforces the identities, privileges, and interests of some groups over others.

Many geographers argue that, by mapping the spatial distribution of racial and ethnic groups, broader patterns and processes of settlement and adaptation can be understood. Also important, scholars suggest, is the de gree to which racial and ethnic distribution patterns change over time. This geographic research illustrates that racial and co ethnic distributions are not necessarily stable, and that spatial patterns and levels of segregation amongst and between racial and ethnic minority popu lations somehow measure the degree to which groups have become spatially assimilated. At the heart of this research is the idea that residential change and social mobility are inextricably interwoven. The assumption is that the movement of ethnic and racial minorities away from co ethnic concentrations indicates that social and structural barriers have weakened, thus suggesting eventual assimilation into the dominant society. There are many classic collections that study ethnic groups from this particular geographic perspective. These volumes generally rely on culturalist understandings of ethnicity, since ethnic attachments are taken as inherently given and primordial in nature, at the same time that ethnic communities persist because of tight solidarity and the unifying force of ethnicity.

In these discussions, ethnic ghettoes and ethnic enclaves are distinguished by their homogeneity, density, and old country 'flavor'. The links of shared history, family, emotions, interests, language, and organizational life bind residents together and reinforce the existence of ethnic communities. Each distinctive ethnic group creates and occupies a geographic area that becomes intimately associated with that group. These ethnoscapes develop into centers of ethnic life concentrating residents, businesses, and places of worship within a delineated ethnic space. Neighborhoods are identified with particular groups, where members create their own social worlds, with cultural and social institutions that satisfy residents' desires to re create 'home'. Ethnic ghettos and ethnic enclaves sometimes acquire place names symbolic of a state of origin, such as Little Italy, Chinatown, and Irish Town.

Recently, geographical studies of ethnicity have experienced important changes in their use of social theory, the foci of research projects, and the breadth of methodologies utilized. New questions have emerged around the issues of ethnicity and race, particularly in regard to (1) the forms of racism and racial inequality that operate through geographical patterns, processes, and ideas; and (2) the ways in which ethnic and racial boundaries (both material and ideological) become struggles of both territory and positionality within society. In other words, recent works argue that the study of ethnicity cannot be understood apart from the spaces within which it takes place. Thus, scholars illustrate how ethnoscapes are more than a passive backdrop for the negotiation of ethnic and racial identities. Rather, they suggest that ethnic ghettos and ethnic enclaves (along with ethnic citadels, which are the ethnoscapes of the dominant culture) must be considered the sites and mediums of privilege, power, and prestige. Ethnic groups positioned in the poorest and most transitional areas, as well as those within the wealthiest neighborhoods, are spatial reminders of broader ethnic and racial divides (and socioeconomic hierarchies as well). Ethnoscapes are concrete spaces around which ethnic and racial categories are inhabited, naturalized, reinforced, and destroyed.

These new theorizations articulate how ethnic and racial identities are not only socially constructed but also spatially constituted. Within this sociospatial dialectic, there are several key contentions. First, space and place are fundamental to any understanding of ethnic and racial identity formation. Second, there are important territorial dimensions to the study of ethnicity and race; in that particular spaces both produce and reflect specific social ideologies. Finally, it is the mutual constitution of these spaces with social, cultural, economic, and political elements that consequentially interweave ethnic identity formation and the racialization process. As some scholars argue, the process of racialization is an essential aspect of how everyday geographies are made, understood, and challenged. Thus, geographers have much to contribute to the literature on racialization. Likewise, locating space in contemporary ethnic and racial formation processes can enrich our understanding of ethnic groups and their experiences in the global social order.

Within this new scholarly arena, the decisive role of dominant groups in the social construction and classification of ethnic groups and racialized minorities is emphatically maintained. How the dominant group 'sees' 'others' is as important, if not more important, than how these groups identify themselves. This is primarily because the dominant group reserves for itself the luxury of dismissing racial identity while more oppressed groups are made daily aware of their ethnic and racial identities. Thus, the ways in which ethnospaces in turn both produce and reflect specific racial ideologies becomes critical. Instead of relying on the 'mapping' of ethnoscapes as a way to identify the spatial distributions and characteristics of 'others', some researchers outline how dominant groups formalize all ethnic and racial categories through institutional practice and material form. They reveal how certain ethnic groups become racially categorized during the time when the neighborhoods they live in become spatially segregated communities. But they also point out how the ethnospaces of the dominant group, which are usually taken for granted as invisible, ordinary, and mundane, systematically, though sometimes unconsciously, create particular spatial boundaries and spatial codes of exclusion. This process of differentiation occurs because the dominant group cannot, or refuses to, recognize how they themselves embody particular ethnic and racial categories. By maintaining invisibility, the dominant group remains racially undefined, and social hierarchies can be rigidly reinforced.

Another recent avenue of geographical inquiry articulates how ethnicity and race (and the ethnoscapes that physically manifest these categories) become commodified. That ethnicity and race are potential products that possess a quality consumers might desire is the result of a variety of complex and highly disputed factors. But many scholars agree that in a world of globalization and diminishing spatial barriers to communication, movement, and exchange, originality, exoticism, and authenticity become prized possessions. Some would even go so far as to say that the intensification of contemporary neoliberalism and technology induced capitalist globalization means that ethnic identities will become even more important and salient in the coming years. Thus, ethnoscapes (which draw upon a variety of ethnic and racial identities, including whiteness) offer a kind of 'ethnic tourism' for individuals looking for a diversion from the homogeneity of everyday life. This process encourages the ethnic business community to reproduce ethnic and racial identities in profitable ways to attract clientele. This process of commodification has become an ever more recognized trend in which people's actions and experiences as consumers have an increasingly formative role to play in maintaining ethnic and racial difference.

Conclusion

It is important to point out that despite its wide usage, ethnicity as an analytical concept in academia still lacks a universal definition. While most scholars agree that ethnicity is the product of structural forces, social organization, and cultural representation, rather than solely the result of innate characteristics and hereditary factors, the term is used with varying meanings in different contexts, and those understandings change over time and across space. Moreover, the complexity of ethnicity is continually reasserted as transnational connections become more prevalent and the cross national movement and linkages of people, goods, information, and financial capital increase. Thus, it is imperative that we study and analyze ethnicity under such changing dynamics, and take into consideration the differences between Western and non Western countries, between immigrant home and host countries, and between domestic indigenous groups/nations and migrant populations and their descendants. Certainly, ethnicity affects societies deeply and in multiple ways that are not always easily identified, separated, and cataloged. Part of the task of understanding the influence of ethnicity is to try to identify and tease apart the many influences of this categorization while remaining conscious of the ways in which this classification scheme is constantly transforming. In human geography in particular, the spatiality of ethnicity should be given high priority in our understanding and analyzing of ethnicity in historical and contemporary contexts. In this way, we can begin to understand how ethnicity becomes a central determinant in perpetuating inequity in society.

See also: Identity Politics; Immigration I; Other/Otherness; Race.

Further Reading

Allen, J. A. and Turner, E. (1988). *We the people: An atlas of America's ethnic diversity*. New York: Macmillan.

Anderson, K. (1991). *Vancouver's Chinatown racial discourse in Canada, 1875 1980*. Montreal, ON: McGill Queen's University Press.

Appadurai, A. (2003). Disjuncture and difference in the global cultural economy. In Braziel, J. E. & Mannur, A. (eds.) *Theorizing diaspora: A reader*, pp 25 48. London: Blackwell Publishing.

Bonnett, A. and Nayak, A. (2003). Cultural geographies of racialization: The territory of race. In Anderson, K., Domosh, M., Pile, S. & Thrift, N. (eds.) *Handbook of cultural geography*, pp 300 312. London: Sage Publications.

Hiebert, D. and Ley, D. (2001). *Working paper series no. 01 08: Assimilation, cultural pluralism and social exclusion among ethno cultural groups in Vancouver*. Vancouver: Research on Immigration and Integration in the Metropolis.

Ignatiev, N. (1995). *How the Irish became white*. London: Routledge.

Jackson, P. (ed.) (1987). *Race and racism: Essays in social geography*. Sydney: Allen & Unwin.

Kobayashi, A. and Peake, L. (2000). Racism out of place: Thoughts on whiteness and an antiracist geography in the new millennium. *Annals of the Association of American Geographers* 90(3), 392 403.

Li, W. (ed.) (2006). *From urban enclave to ethnic suburb: New Asian communities in Pacific Rim countries*. Honolulu: The University of Hawaii Press.

Nash, C. (2003). Cultural geography: Anti racist geographies. *Progress in Human Geography* 27(4), 637 648.

Omi, M. and Winant, H. (1994). *Racial formation in the United States from the 1960s to the 1990s* (2nd edn). London: Routledge.

Peach, C. (2000). Discovering white ethnicity and parachuted plurality. *Progress in Human Geography* 24(4), 620 626.

Sibley, D. (1995). *Geographies of exclusion: Society and difference in the west*. London: Routledge.

Skop, E. (2008). *Saffron suburbs: Lessons learned from an Indian American Community*. Chicago: The Center for American Places.

White, P. and Jackson, P. (1995). (Re)theorising population geography. *International Journal of Population Geography* 1(2), 111 123.

Wright, R. and Ellis, M. (2006). Mapping others. *Progress in Human Geography* 30(3), 285 288.

Ethnicity and Resistance, Historical Geographies of

D. P. Nally, University of Cambridge, Cambridge, UK

Glossary

Accumulation by Dispossession A neologism coined by David Harvey to refer to the historical geography of capital accumulation through imperialism.

Contact Zone The hybrid, improvisational space where different cultures meet and contend with one another.

Ethnocentrism A deeply entrenched belief in the superiority of one ethnic group over another.

Manichean A binary or dualistic mode of thinking and categorization.

Primitive Accumulation The original process by which markets are established, capital is acquired, and capitalist relations of production are mobilized.

Reification The ideological process whereby human relations are abstracted from their social contexts and re-presented as 'things'.

Introduction

Although the word 'ethnic' dates back to Greek antiquity, deriving from *ethnos* meaning people, the term 'ethnicity' really only entered popular usage in the wake of World War II. Ironically the term was first deployed as an acceptable alternative to the concept of 'race', which was mostly abandoned in the social sciences because of its association with Nazi policies. However, the word ethnic had an exclusionary meaning from the very beginning, serving to distinguish Christian and Jew from their pagan and Gentile neighbors in ancient times. Therefore, historically speaking, the growth of a collective sense of solidarity and group cohesion depended upon the elevation of anthropological differences and the segregation of human populations. In its more dramatic and virulent form these beliefs and 'imagined geographies' could spiral into fear, interethnic conflict, and collective violence. Today the term ethnicity is primarily used to refer to common religious, linguistic, or cultural characteristics that help differentiate one group of people from another. An ethnic group thus differs from a 'community' in so far as the signifiers of difference are usually perceived to be in some sense hereditary or pertaining to historical kinship.

Thus, as in the distant past, an epistemology of 'difference' tends to define the calibration and promotion of 'sameness'. It is hardly surprising therefore, that ethnicity is often used synonymously with the more hotly contested term 'race' and why race wars are frequently indistinguishable from ethnic conflicts. Where the terms are qualified, however, and nuance and distinctions recognized, most academics note the attributed 'natural' or biological aspects of race in contradistinction to the imputed 'cultural' moorings of ethnicity. However, in a more critical vein post structuralist scholars observe the social construction of both nature and culture and hence the ultimate artificiality of racial and ethnic inscriptions. In popular parlance at any rate, both race and ethnicity tend to bleed into one another.

It is also clear that the notion of territory proves central time and again to ethnic claims, controversies, and conflicts. This holds true on a number of fronts. For example, religious beliefs are very often associated with holy sites, chosen lands, and sacred spaces. Languages and dialects also tend to be geographically circumscribed. Similarly the ideology of the nation state (frequently, though not necessarily, promoting ethnic homogeneity) requires a sufficient territorial basis in order to advance sovereign independence and guarantee future stability. In each case, boundary promulgation is an important aspect of ethnic selection and affinity.

Of course when considering the territorial dimensions of ethnicity there are other, more material concerns. Resources are obviously abundant in some places and not in others and ethnic conflicts all too often map the contours of resource allocation. In this sense ethnic antagonisms are frequently the result of larger geopolitical strategies and power plays that generate and sustain human cleavages for material personal benefit and gain. Referring to the 'Age of Exploration' in the fifteenth century Karl Marx described the mass expropriation of foreign resources and the extirpation and enslavement of indigenous peoples as 'primitive accumulation' and more recent literature examines the ongoing specter of 'accumulation by dispossession' in neoliberal economics and the US led 'war on terror'.

Marx also famously claimed that people make their own history, but not necessarily under the conditions of their own choosing. We might take this and indeed the foregoing discussion as implying that all identities – including ethnicity – are both *in*scribed and *a*scribed. In other words, on the one hand, we can find historical cases that demonstrate the ways in which ethnicities (or ethnic categories) are attributed to certain peoples in order to morally justify coercive actions, while, on the other hand, we can locate historical moments where these selfsame categories and identities are resisted, and other identities inscribed in their place, sometimes in order to advance alternative political goals. Hence ethnicity can be understood dialectically as a relentless juxtaposition and contestation over group identities forged in the mael strom of human interactions.

In the subsequent sections of this article, three im portant aspects of ethnicity will be sketched out. First, the relationship between colonialism, ethnicity, and re sistance will be historically examined. Next, the contours of economic globalization will be explored in order to illustrate how uneven geographical developments can produce resistance movements based on ethnic dis tinctions. Finally, the case of ethnic differentiation and resistance will be considered in the context of regional and transnational migrations, diasporic cultures, and polyethnic immigrant societies.

Colonialism and Ethnic Identities

The exercise of colonial and imperial power often rests on the social construction of ethnicity. We may define colonialism as the exercise of one country's sovereignty – however provisional or incomplete – over another peo ple, their resources, and territory. Imperialism, on the other hand, commonly refers to more informal mech anisms for maintaining dominance over indigenous peoples. These definitions are fraught of course, but for our purposes they help distinguish control wrested by military force and economic leverage from the more subtle and dynamic ways in which cultural influence and dominance are asserted. Throughout history empires have sought to marshal superior strength to plunder other nations, polities, and regions. However, in the process of absorbing foreign territories empires fre quently try to impose sociocultural, religious, and lin guistic structures on the conquered population in the hope of securing submission.

Colonialism and imperialism aid ethnic differentiation in two significant ways. In the first place, imperial and colonial powers often create or exacerbate differences between 'colonial' and 'colonized' subjects in order to justify *la mission civilisatrice*. In other words, the extension of sovereignty over other populations, their resources, and territories relies extensively on the 'ethnocentric' belief that the morals and values of the colonizer are in every way superior to those of the colonized. Indeed some observers have characterized the colonial world as essentially Manichean – or 'divided into compartments' – that serve to perennially partition 'us' over 'here' from 'them' over 'there'.

Frantz Fanon and other early anticolonial writers (such as Aimé Césaire, Ngugi Wa Thiong'o, and Jean Paul Satre) pointed to the ways in which Europe was quite literally the creation of the Third World: its enormous wealth stacked on colonial plunder and booty, so much so that many of the great capitals of Europe continue to reflect this stolen opulence. It is also argued that this level of expropriation extends to the psyche of colonized peo ples, who develop a sense of inadequacy that becomes the correlative to the colonizer's feeling of superiority. In ad dressing the radical asymmetries that divorce the 'native' from the 'settler colonizer', postcolonial scholars have at tempted to theorize the ways in which ethnic identities are mutually constructed – and contested – through colonial encounters. So, for example, if French colonizers depicted Algerians as lazy, dependent, fickle, licentious, and sub human this gesture allowed the French to conceive of themselves as precisely the opposite: industrious, autono mous, dependable, restrained, and civilized. Many scholars link such beliefs to the racist, pseudoscientific theories developing in Europe in the eighteenth and nineteenth centuries, but whatever the origins, these representational strategies were an obvious fulcrum around which ethnic identities were forged.

It is possible, therefore, to examine the ways in which ethnicity is produced or imposed at the 'contact zone' of two cultures. Indeed the charge of radical 'otherness' very often encourages the development of ethnic conscious ness and solidarity among colonized peoples. This growth of ethnic solidarity is frequently referred to as 'ethnic nationalism'. While it ought to be noted that this form of nationalism is not confined to imperial and colonial processes, it is nonetheless remarkably common to such historical encounters.

In the case of English and later British colonization of Ireland, the Irish could choose to reject outright the im perial projections of ethnic inferiority and posit an alter native identity in its place. Therefore Irishness was often narrowly defined as everything un English or British. Al ternatively the colonized Irish could turn those selfsame charges on their head, making a pejorative attribution a source of ethnonational pride. Thus, for example, the Irish re engineered the accusation of being a rebellious and warring population into proof of their martial prowess and guile. Moreover, if rural impoverishment and primitive living in Ireland enabled the projection of an urban, so phisticated colonial identity, Irish nationalists refracted those attributed faults so that rural Ireland came to sym bolize rustic and romantic living. Indeed during the late nineteenth century Irish 'folk life' was consistently rep resented as frugal, light hearted, and carefree, and always offset against a projection of the phlegmatic, doleful Anglo Saxon. The mutual constitution of 'Paddy' and 'John Bull' was always a dialectical shuffle between mi mesis – the desire for sameness – and alterity – the as piration to be quintessentially different.

This colonial 'mirror dance', as the process has been called, is a common outcome of agonistic encounters and regularly creates the basis for anticolonial and post colonial ethnic affinities. The assessment of these histories and processes continue to inspire controversy and critical debate. For instance many scholars lament the essentialist understandings of identity that undergird colonial, anticolonial, and postcolonial movements. These so called ‘antifoundationalist’ accounts critique the presumption that any authentic identity stands outside social analysis. Ethnicity, in this sense, is a historical formation, forged at the juncture of colonial imposition and indigenous re sistance. These critiques draw attention to the ex clusionary nature of ethnonationalism, which frequently reproduces the Manichean chauvinism it so stringently rejects. As an alternative political message, cultural the orists such as Homi Bhabha highlight the contingent, ‘hybrid’ nature of colonized subjectivities and argue this as an appropriate platform to envisage a postcolonial identity of resistance that nonetheless goes beyond the strictures of an exclusionary ethnonationalism. ‘Hybridity’ is simultaneously symbol and intervention. The message is that we don’t have to ‘polarize to polemicize’; that Western modernity was never a unilateral project; and that from the beginning minority groups and cultures (read ‘ethnicities’) run to the core of colonial and imperial identities. In contrast, other writers have argued that at certain moments ethnic or cultural differentiation can be politically important, and even necessary. The use of ethnic categories in order to further certain perceived emancipatory goals (to advance anticolonial struggles, for example) is sometimes referred to as ‘strategic essential ism’. Hannah Arendt captured the spirit of this concept when she reasoned that a Jew should always respond as a Jew when attacked as a Jew. Nevertheless, adopting es sentialist identities for ‘strategic’ purposes remains a fraught and complex position. In practice where ethnic identities are cultivated, and human differences under scored, there is often a vast gap between emancipatory and egalitarian politics.

Globalization and Ethnicity

A number of scholars have argued that ‘economic glob alization’ is responsible for a surge in ethnic identities around the world. Economic globalization is sometimes used to suggest an increasingly internationalized econ omy, coupled with evermore interconnected modes of capital production. Thus promoters of new era economic globalization tend to celebrate the ‘flattening’ of the world – through ‘time–space compression’ – as an in trinsic force for good, whereas critics of economic glob alization argue the unevenness of this restructuring and how it continues to reproduce social and economic inequalities.

Although economic globalization might be acceler ating spatiotemporal convergence, as some commentators insist, the significance of ethnic cleavages seems far from waning. Indeed, ethnicity is frequently mobilized as the basis for resisting the homogenizing forces or particular local effects of globalization. Indeed, as ‘borders’ are rendered more porous (by capital flows, institutional structures, multilateral policies, international forums, global travel, etc.) it is arguable that ‘boundaries’ become evermore important and social stratification more en demic. The continued salience of ethnic identities is sometimes considered to be a knee jerk reaction to global communications technology, the preponderance of ‘Hollywood’, and the so called ‘Americanization’ of the world. For instance in 2010 it is predicted that India will become the country with the most English speakers in the world. To some commentators, the emergence of the Hindustani language film industry in India, so called ‘Bollywood’ cinema, is an exemplary countercultural phenomenon that raises a significant challenge to the dominance of Western media. Similarly the legalization of French as the sole official language in the Canadian province of Quebec can be viewed as a reaction to the increased globalization of the English language (some times referred to as ‘linguistic imperialism’). Thus an ideology of ethnic distinction regularly informs the practices that resist dominant or hegemonic cultures.

One particular manifestation of ethnicity as resistance is the promotion and defense of indigenous rights in the face of the local effects of global economic activities. Indigenous peoples around the world have organized into local, regional, and global associations, councils, and other alliances to collectively challenge development activities that they believe contravene their interests. Where collective action has not worked, or has not ap pealed, some groups have resorted to systematic violence or terrorism to assert their rights or elevate their political status.

The exploitation of oil resources by transnational corporations has inspired these different styles of resis tance. Two well publicized examples include the Huaorani people from the Ecuadorian Amazon and the Ogoni people of the Niger Delta. Since the 1970s, Christian missionaries, the petroleum industry, and un fettered logging have encroached on Huaorani land and only in recent years, as a result of indigenous political organization, have the Huaorani begun to claw back rights to their ancestral lands. This history is not unique to the Huaorani, but mimics the experience of many indigenous peoples in Latin America and around the world. Like the Huaorani, the people of Ogoniland fought a long battle against the interests of global petrocapitalism both within and beyond the Nigerian state. Most famously, the Movement for the Survival of the Ogoni People (MOSOP) was established with the aim of resisting multinational oil (notably the Shell Petroleum Develop ment Company), promoting indigenous control over indigenous resources, redressing the effects of political marginalization, and securing recompense for environ mental contamination. Significantly, the best known statement of this indigenous political program – known as the Ogoni Bill of Rights and submitted to the government and people of Nigeria in November 1990 – established the above claims on behalf of the Ogoni people, described as a 'separate and distinct ethnic nationality'. The docu ment also reaffirmed the responsibilities of the Federal Republic of Nigeria to respect the rights of all ethnic groups, no matter how small, and irrespective of global crude prices. Nonetheless, since the execution of Ken Saro Wiwa and eight Ogoni leaders by the Nigerian state in 1994, this conflict has descended into violence. Today, the people of the Niger Delta continue to stake their claim to oil profits by mobilizing discourse around ethnic entitlements and using militant tactics of resistance, such as kidnapping foreign oil workers, thieving oil, and at tacking production facilities.

The foregoing examples illustrate very well the practical difficulties faced by activists, governments, and international organizations seeking to protect indigenous peoples and minority ethnic groups faced with systematic discrimination and violence. So common are assertions of abuse by indigenous groups that corporations and de velopment agencies have created guidance frameworks, policies, and principles to ensure good relations between projects and indigenous peoples. For example, the International Finance Corporation's Equator Principles and the World Bank's Policy on Indigenous Peoples aim to ensure 'free, prior, and informed consent' of affected indigenous peoples. These approaches are premised on a 'human rights' framework and the duty of governments and corporations to respect these.

However, the interpretation of these conflicts within the framework of human rights is not entirely un problematic. Governments frequently disregard inter national law to suit their own national agendas. Indeed the vast schism between rhetoric and practice has led many scholars to question the presumed distinction be tween civil rights enjoyed by most Western populations and the human rights usually ascribed to the world's poor and oppressed. On the back of this, other critics believe the critical and political emphasis should lie elsewhere. Researchers working on India, for example, tend to prefer a class based analysis that attends to ways in which peasantries and small ethnic groups are forced to undergo proletarianization. Large scale development projects, like hydroelectric dams (sometimes organized in partnerships between the national government and de velopment institutions such as the World Bank) often sacrifice local land based livelihoods in the name of na tional progress. In addition to highlighting the material basis of exploitation this work suggests that in some cases ethnic identities might serve as an instrument of class oppression by preventing the refinement of class consciousness among subjugated peoples. Clearly eth nicity means different things to people at different his torical junctures and in different geographical locations. Moreover, competing theoretical and political con structions of ethnicity can have a profound impact on indigenous cultures, including their efforts to instigate counter hegemonic strategies.

Multiculturalism, Diasporic Cultures, and Ethnicity

The tenacity of ethnic identities is in no small part due to historical migrations of human populations, the develop ment of multiethnic immigrant societies, and the persist ence of 'diasporic cultures' retaining a shared sense of distinctiveness associated with a homeland, however mythical that designation/destination might be. In the so called multiethnic immigrant societies a key issue tends to be the intrinsic tension between ethnic identification and citizenship. The former is almost by definition exclusion ary whereas the latter frequently appeals to the inclu sionary values of 'multiculturalism'. To some degree this replays the familiar distinction between 'ethnic' and 'civic' nationalism. While ethnic nationalism is judged to em brace a hierarchical model of identification, its civic vari ant is thought to assume a more horizontal social status order. Although this characterization has been criticized as too simplistic, its neat compartmentalization between 'good' and 'bad' nationalisms seems to parallel more recent debates on multicultural values, usually defined against ethnic identifications. Thus multiculturalism is usually represented as progressive, whereas ethnic identification is considered to be 'atavistic' and discriminatory. However, as often is the case, these designations obscure a more complex reality.

Two obvious examples of multiethnic immigrant soci eties are the United States of America and Australia. In both cases difficulties and tensions surface during com memorative events that recall, or more appropriately 'perform', political or civic foundation myths. In the case of the US, the Columbus Quincentennial commemor ations in 1992 were originally conceived as a celebration of the 'Age of Discovery' in which the 'New World' pioneer myth was repeatedly promoted. However, the com memoration proved immensely controversial as political activists used the occasion to highlight grievances of the indigenous peoples of the Americas, including their his tory of exploitation and their own sense of themselves as founding peoples. Similarly the slow recognition of Abo riginal land rights in Australia jars with the official mantra of multiculturalism. As in the US the existence of eth nically and racially based hierarchies stands at odds with the self projected image of multiethnic diversity and tol erance. Furthermore some critics have claimed that multiculturalism, and to certain degree cosmopolitanism, constitute a drive toward national identity defined ex clusively by a dominant elite class. In this sense, we can read the above resistance to 'corporate multiculturalism' and official 'national' history as a determined reassertion of ethnicity from groups and peoples that continue to feel excluded and marginalized despite popular platitudes of cultural pluralism.

In addition to domestic cultures within polyethnic immigrant societies, the historical migrations of popu lations have also created unique diasporic cultures that transcend state boundaries and often span the globe. The meaning of the term 'diaspora' has dramatically expanded in recent years. Traditionally, the word was associated with the expulsions of the Jews from Palestine, but more recently the concept has been generalized to refer to any ethnic population that has migrated to another country and yet continues to retain its own distinct identity. Moreover, because diasporic cultures are primarily anti assimilationist, they present a serious challenge to the ideology of the nation state.

The migration that produces diasporic communities can be voluntary or compelled and this often relates to the degree of ethnic cohesion among the migrant com munity. Nevertheless there are important political and historical factors that need to be registered. The creation of the Jewish diaspora in the US, for instance, owes a large part to the genocidal policies of the Nazis. Over time American Jews consolidated their position and were able to advance their own domestic 'and' international interests. The forced transportation of Africans into slavery is another conspicuous example of compelled migration. However, in contrast to the Jewish experience, the imposition of slavery and racial segregation meant that American blacks remained excluded and largely disenfranchised. This exclusion in turn nurtured a long, largely nonviolent struggle for civil rights, which in a sort of Rawlsian way, was an entitlement claim by one group of people over the oppression of another.

These distinct migrant histories and experiences have led one scholar to distinguish between 'proletarian' and 'mobilized' diasporas. Mobilized diasporas refer to mi grant ethnic populations with access to social and eco nomic capital. Their skills or knowledge are seen as valuable to native elites and they quickly achieve pos itions of significant social status. Proletarian diasporas, in contrast, often migrate from developing countries be cause of poverty or political turmoil. They often lack basic education and skills and are therefore excluded from elite positions and consigned to low end jobs. The distinction between 'proletarian' and 'mobilized' dias poras helpfully historicizes some of the factors that nurture ethnic identities and conflicts. For example, the Irish in America began as a proletarian diaspora, but over time organized and extended their influence primarily at the expense of other proletarian ethnic groups. The historic mobilization of Irish Republicanism depends to a large extent on the successes of the Irish diaspora in the same way that Zionism continues to rely heavily on the political activities of American Jews. Militant Islam also relies on a transnational network of training camps, international financial banking, and a galvanizing sense of jihad that might be loosely described as diasporic. In this sense, the old association between transnational com munities, ethnic identities, and political activity con tinues today.

Recent developments in transportation and infra structure, as well as cross border cooperation, mean that sizeable diasporas can develop in a relatively short period of time, as in the case of the Polish community in Ireland and Turkish community in Germany, for instance. In such cases there are many factors shaping ethnic for mation and potential intercommunal conflict including occupational segmentation, civic inclusion, residential segregation, language and religious differences, as well as broader geopolitical conflicts. It is arguable, for instance, that the 'war on terror' has mobilized ethnic violence against British Muslims and in some cases British Sikhs have been violently attacked because of their apparently 'similar' apparel. Such vicious behavior illustrates all too well the continued salience of ethnicity, not to mention the optical constructions that mark one human being as radically different from the next.

Issues

It would be too simplistic to say that historically, con quered populations fought for independence on ethnic grounds, whereas in the present ethnic groups fight for inclusion in their host societies. Certainly the politi cization of ethnicity can be a powerful means of resisting inequalities, but it can also socialize individuals in ways that produce, maintain, and exacerbate violent inequal ities. Ethnicity continues to form the structural basis for genocidal policies throughout the world, and other vio lent and discriminatory practices including prohibitions on intermarriage, inequalities in schooling, unequal ac cess to the means of production, the global 'war on terror' and the waging of Islamic jihad, show signs of an en trenched ethnic status order. The fact that the vast ma jority of the world's nation states are polyethnic seems to have done little to lessen the resilience of ethnic con sciousness. As the above examples illustrate, ethnic identities are frequently bound up with other forms of subjectivity and recent scholarship has begun to examine how ethnicity is interpellated with gender, sexuality, and class across different spatial and temporal contexts. Other scholars have pointed out that majorities are no less ethnic than minorities and that white identities, for ex ample, are socially constructed through a range of dominant discourses and practices that are equally important to interrogate.

The political goal of much of this work is to de naturalize and de ontologize ethnicity by illustrating its inherent constructedness. However more recent research, building on 'postrace' theorists' attempts to undermine the representational economy of race, has urged for a more radical approach for working through and against 'identity politics'. As well as critiquing the way contemporary theory is complicit in the reification of ethnicity this emerging body of work proposes an alternative political platform based on resistance to 'ethnicity' itself. In other words, this work is committed to practicing a new political imaginary that goes beyond ontological determinism and opens the way toward what one scholar has termed 'planetary humanism'. To some critics, this notion of existence without essence sounds airily removed from the fault lines of political struggle, whereas for others a new and radically different epistemological structure is absolutely central to contesting the oppositional identities and affiliations of political ethnicity.

See also: Cultural Economy; Diaspora; Identity Politics; Migration; Multiculturalism; Nationalism; Race.

Further Reading

Abizadeh, A. (2001). Ethnicity, race, and a possible humanity. *World Order* 33, 23 34.

Anderson, K. J. (1991). *Vancouver's Chinatown: Racial Discourse in Canada, 1875 1980*. Montreal: McGill Queen's University Press.

Anderson, J. and Shuttleworth, I. (2004). A new spatial fix for capitalist crisis? Immigrant labour, state borders and the new ostracising imperialism. In van der Pijl, K., Assassi, L. & Wigan, D. (eds.) *Global Regulation: Managing Crisis After the Imperial Turn*, pp 145 161. Basingstoke: Palgrave Macmillan.

Armstrong, J. (1976). Mobilized and proletarian diasporas. *American Political Science Review* 70, 393 408.

Bhabha, H. (1994). *The Location of Culture*. London: Routledge.

Eriksen, T. H. (1993). *Ethnicity and Nationalism*. London: Pluto Press.

Esman, M. J. (1986). Diasporas and international relations. In Sheffer, G. (ed.) *Modern Diasporas in International Politics*, pp 333 349. London: Croom Helm.

Fanon, F. (1963). *The Wretched of the Earth*. New York: Grove Press.

Gilroy, P. (1998). Race ends here. *Ethnic and Racial Studies* 21, 838 847.

Hall, S. (1992). The new ethnicities. In Donald, J. & Rattansi, A. (eds.) *Race, Culture, Difference*, pp 252 259. London: Sage.

Harvey, D. (2003). *The New Imperialism*. Oxford: Oxford University Press.

Hobsbawm, E. (1990). *Nations and Nationalism Since 1780*. Cambridge: Cambridge University Press.

Hutchinson, J. (1994). *Modern Nationalism*. London: Fontana Press.

Lustick, I. (1993). *Unsettled States, Disputed Lands: Britain and Ireland, France and Algeria, Israel and the West Bank Gaza*. Ithaca: Cornell University Press.

Mallon, F. E. (1992). Indian communities, political cultures and the state in Latin America, 1780 1990. *Journal of Latin American Studies* 24, 35 53.

Nayak, A. (2006). After race: Ethnography, race and post race theory. *Ethnic and Racial Studies* 29, 411 430.

Olund, E. (2002). Public domesticity during the Indian reform era; or Mrs Jackson is induced to go to Washington. *Gender, Place and Culture* 9, 153 166.

Taussig, M. (1993). *Mimesis and Alterity: A Particular History of the Senses*. London: Routledge.

Wilford, R. (ed.) (1998). *Women, Ethnicity and Nationalism: The Politics of Transition*. London: Routledge.

Ethnography

K. E. Till, Virginia Tech University, Alexandria, VA, USA

Glossary

Coding An iterative and analytical practice used to identify patterns, elaborate upon insights, and refine ideas generated in the research process. Originating from the sociological tradition of grounded theory, there are many types of coding approaches used by ethnographers to analyze notes, documents, and material culture.

Creative Analytical Ethnographies Such projects perform, rather than simply document, theoretical claims and include a range of representational strategies, including autoethnography, fiction stories, poetry, performance texts, polyvocal texts, visual presentations, allegory, conversation, and layered accounts.

Fieldnotes Ethnographers write descriptive, reflective, and interpretive notes; create sketches and maps; and formulate questions about the everyday geographies, emotions, social spaces, and material encounters in the settings being researched. Fieldnotes may later be treated as textual objects to be coded, interpreted, and/or analyzed. The ethics of recording (through notes, video, audio tapes, and maps) remain heavily debated.

Memos Ethnographers write analytic essays to reflect, interpret, and record their responses to situations and the research process. This research and writing practice encourages ethnographers to question underlying assumptions that emerge in specific contexts and to scrutinize their varied relationships with research fields and partners.

(Participant) Observation Ethnographers pay attention to, and may partake in, everyday geographies to become familiar with how social spaces are constituted in various settings. Observation contexts and knowledge production include social interaction, emotions, and embodiment.

Transnational Ethnographies Projects that analyze the forms and effects of globalization through countertopographies, networks, and multiple sites of: consumption and production; tourism and whiteness; citizenship; critical, black, and/or queer political and philosophical thought; and movements of peoples, including refugees and expatriates across political territories.

Ethnography

Ethnography is a methodological and practice based approach to understanding and representing how people – together with other people, nonhuman entities, objects, institutions, and environments – create, experience, and understand their worlds. Denzin and Lincoln define ethnography as a research strategy rather than methodology *per se*, because it entails a variety of research designs, methods, analytical tools, theoretical perspectives, and representational forms. Ethnography is time intensive, iterative, and open ended, and includes observing, listening, reflecting, experiencing, writing, and learning from peoples, places, situations, institutions, landscapes, and things. The questions ethnographers ask always change during the research process, because a significant part of any study entails being sensitive to the political relationships and ethical responsibilities associated with framing, generating, co creating, and representing knowledges.

In geography, ethnographic research has long been a significant methodology in cultural ecology, development studies, and feminist geography; more recently, it has become established in social, political, cultural, and nature society geography, as evidenced by the 'Doing ethnography' sessions sponsored by the Qualitative Research Specialty Group at the annual Association of American Geographers meetings (ongoing since 2003) and similar sessions at Royal Geographic Society/Institute of British Geographers meetings (in 2007). Geographer ethnographers examine everyday processes of meaning making and materiality in a range of settings by focusing on spatial practices, including place making, inhabiting social spaces, consumerist spatialities, becomings, memory, transnational citizenship, and grassroots activism. While geographers still draw upon anthropological and sociological traditions, they now are making significant contributions to transnational, collaborative, and creative analytical ethnographic research. The article overviews research techniques and approaches in critical ethnography and highlights recent work by geographers.

From Traditional to Critical Ethnography

Historically, the ethnographer was a heroic figure, a detached observer of 'foreign' cultures who collected material culture, stories, even 'primitive' peoples, to advance scientific research on behalf of the colonial project. The ethnographer observed culture groups in their 'natural' settings, wrote scientific reports, took photographs, and gathered artifacts. As scientists, ethnographers were to maintain some level of distance, and hence objectivity, so as not to be disruptive of the activities and people under observation. By not interfering with 'normal' practices of 'a' culture, there was also a reduced chance of the eth nographer 'going native'.

New ethnographic forms have emerged over the past 20–30 years that are considered 'critical' of underlying assumptions of traditional ethnographic research prac tices, documentary styles of representation. Since at least the 1970s, for example, ethnographers have questioned claims of scientific objectivity and of 'fields' existing 'out there' to be studied (colonized). The ethics of collecting data in a foreign place and returning to the (Western) university to 'write up' a study remains heavily debated. While multiple methods remain central to ethnography, the notion that a knowable object of study can be trian gulated has been challenged. Gillian Rose also argues that ethnographic practices such as reflexivity and posi tionality problematically assume that the researcher can render oneself fully transparent or place oneself on an even terrain of power with those researched. Mary Thomas has recently questioned the assumption that a researcher can 'read' personal narratives and interview transcripts as transparent representations of unconscious processes. The researcher's body is also understood as a contested site of knowledge production. Critical eth nography, in other words, is an intersubjective form of qualitative research, whereby the roles of researcher/ researched, insider/outsider, self/other, body/environ ment, and field/home are understood as shifting and multiple, and always negotiated through complex power relations. The sections that follow outline recent dis cussions that have emerged within critical ethnography and geography about research methods, approaches, and forms of representation.

Ethnographic Fields as Contexts for Social Interaction

Ethnographers are often described as bricoleurs because they use multiple methods, depending on the project, including: interacting with others in social settings; writing field notes and memos; sketching maps; working for an organization or community group; reflecting upon and engaging in conversations about being in and ex periencing social spaces and particular places; gathering visual, material, and documentary materials related to the research setting; asking others to document and describe their everyday worlds (such as through sketches, photo graphs, videos, soundtracks, and diaries); and conducting in depth to semi structured interviews, focus group interviews, and/or surveys.

Justin Spinney, for example, has experimented with video ethnography, including collaborative video and music co editing, to research geographies of mobility. Spinney, a cyclist and commuter, filmed his and other cyclists' journeys, as well as asked other cyclists to film their own journeys, using helmet cameras and portable microphones. In addition, he randomly interviewed cyclists by distributed flyers about his projects on people's bikes throughout London. Spinney's multiple methods not only attempts to gain more information about both the range of people who cycle, from commuters and bike messengers to leisure cyclists and bicycle motocross (BMX) and trials riders, and the skills these riders use to create spaces of freedom and flexibility in the city. Through collaborative video editing with some of the people he has filmed and talked to, he wishes to under stand how the bike is both a social place and object/ assemblage that allows cyclists to enjoy embodied feel ings of riding.

Observation, writing fieldnotes, collecting material culture, and in depth interviews are the methods most commonly identified with ethnography and are also the most controversial, given ethnography's colonial legacy. The status of the ethnographer as the expert arbiter of knowledge is not only questionable – analyses are widely read and challenged by the peoples studied – but so too is the assumption that a local 'scene' observed is typical for any social group. For this reason, Michael Angrosino and Kimberly Mays de Pérez suggest that observation should be thought of as a context for social interaction rather than as a social science method *per se*. As Dydia DeLyser explains, by partaking in everyday life as it is lived over many months and oftentimes years, even at tourist places such as ghost towns, the ethnographer is more likely to understand the ways that people experience and create their worlds. Through this close attention to lived and unspoken geographies about particular places, eth nographers may better understand the overlapping and often abstract processes that connect and separate, to borrow from Lila Abu Lughod, domestic, national, bio graphical, daily, spontaneous, and sociogeological spaces and times.

The researcher's own body, for example, is a signifi cant place through which ethnographic research is con ducted. Geographers, Alison Bain and Catherine Nash, Heidi Nast, Hester Parr, Arun Saldanha, and Andy Tucker, in quite distinct ways, discuss how social inter actions, and viscous assemblages of human and nonhu man bodies, create public and private spaces in cities, neighborhoods, homes, and sites of leisure. A significant part of their work stems from their experiences and in tuited understandings as embodied, sexed, gendered, and racialized researchers. Their observations of, and inter actions with, others helped them write about the ways bodies interact, meld, and constitute social spaces. Their work is different from traditional forms of participant observation, because they consider emotions and phys icality as important forms of knowing. By actively en gaging in social settings rather than observing events from a distanced perspective, these scholars also discuss the limits and possibilities of observation as an embodied research practice in particular situations.

Parr's observations of what constitutes normal bodily movements in public urban spaces, for example, allowed her to change her bodily presence to interact with and learn from marginalized peoples in streets, parks, and squares. Her work raised significant insights about embodied social relationships, including gender, sexuality, and health, as well as the problems with assuming that a researcher's self/body is contained in the research process. It also resulted in formulating more participatory research strategies with individuals that had mental health problems. Yet, Parr also acknowledged the limits and ethical concerns of using the body as a covert research tool to observe and communicate with vulnerable peoples in public space. She argues that some social processes and interactions cannot always be researched with informed consent or through collaborative processes. Some individuals actively seek to avoid dominant institutional spaces; most of us also communicate in nonlinguistic ways. In such situations, working intersubjectively on an agreed upon project may not be possible, appropriate, or even ethical. Stuart Aitken and Doreen Mattingly also discuss the moral difficulties of speaking for vulnerable groups, such as children and teenagers, respectively, who cannot or may not be able to speak for themselves. Children and teenagers see things in environments that researchers do not see and experience worlds that adults may not understand. As Mattingly describes, while empathy and being open minded helps, well intended efforts, including community based projects, may not be enough.

In addition to treating observation as a form of social interaction and embodiment, ethnographers learn about everyday spatial tactics, practices, and strategies from listening to people telling stories about the places and things that are important in their everyday lives. Richa Nagar's 'ethno geographies' combine oral histories within geographically specific contexts and relations, making unexpected connections across stories and spaces. Anna Secor used focus group discussions and observations to understand how individuals experienced being classified as either 'citizen' or 'stranger' in Istanbul. Some described everyday urban tactics, such as anonymity, as a form of resistance, whereas others co opted dominant strategies by identifying themselves as political subjects in schools, neighborhoods, and workplaces. Patricia Ehrkamp similarly learned about the everyday spatial strategies of Turkish women living in Marxloh, Germany, by asking participants to draw mental maps of their neighborhoods. When asked why certain parts of their neighborhood were highlighted, crossed out, or not sketched in, these women responded that they avoided prominent male only Turkish teahouses and frequented female only and mixed social spaces that challenged masculinist forms of power relations in their neighborhoods.

Ethnographers also pay attention to how material forms of situated knowledges embody the emotional geographies and worlds of others. Luke Dickens, for example, followed graffiti crew members, individuals, artistic institutions, found objects, and events as they traveled through and create urban spaces, including materially, through film, on the Internet, and by telling stories. Dickens tracked the journeys of graffiti writers, stencillers, artists, and others involved in London's graffiti scene to understand how, through their various practices, they remade the street, art gallery, and city. Katie Walsh explored notions of belonging by British expatriates in Dubai through their attachments to things. She noticed that some objects, from an ugly, orange plastic bowl to a landscape painting, to pictures and postcards taped to walls, were given a special status in domestic spaces: many stories of (dis)location and attachment were embodied in them. In a different setting, Mary Thomas used autophotography as a technique to learn more about teenage women in Charleston, South Carolina. Considered neither child nor adult, teens do not have many options to exist in cities without being considered disruptive, or even dangerous. By inviting teens to take pictures of the places they hang out in and then to tell stories about these pictures, Thomas was able to interact with young women through a familiar and playful medium – disposable cameras and photo albums. They pointed to these pictures in their homes, often laughing at themselves and their friends in the images, as they retold stories about taking the photos, the places, or memories triggered by the images. Through these varied forms of communication, these teens expressed (in contradictory ways) their fears and desires about difficult topics such as teen sexuality, subjectivity, and belonging.

As these examples suggest, when observation is understood as an interactive context, then writing field notes about everyday geographies, emotions, fluid social spaces, and material encounters should be considered a practice of discovery and analysis rather than a form of objective reporting that assumes that language is a transparent signifier of witnessed events. Consequently, the fieldnotes that ethnographers write will contain a range of descriptive, reflective, and interpretive notes, sketches, found objects, lists of people to contact, and questions that emerge as they work. To provide a space for reflection, interpretation, and their fears, ethnographers also regularly write memos that question their experiences and assumptions, pay attention to processes, consider the material and visual cultures in their study, and scrutinize their varied relationships with research fields and partners.

Ethnographers also treat their notes as textual objects that can be interpreted and analyzed. Approaches used to analyze ethnographic fieldnotes, documents, and other forms of visual and material culture may include:

open ended coding, qualitative or quantitative content analyses, narrative analysis, conversation analysis, forms of discourse analysis, semiotic analysis, hermeneutics, psychoanalytical analysis, and phenomenological approaches. Coding in particular is an analytical practice used to identify patterns, elaborate upon insights, and refine ideas. While its origins stem from the sociological tradition of grounded theory, coding is no longer understood as an objective scientific practice. Instead, open coding is a form of brainstorming, whereby the researcher revisits materials in order to think about possible ideas, themes, and issues. Different types of focused forms of coding are used to identify variations and patterns in relationships. Overall, coding is an iterative analytic practice that directs the ethnographer to examine closely the materials generated in the research process, breaking down a larger whole into parts, that will later again be scrutinized in different parts.

What is not so apparent in the published texts of ethnographers is the almost obsessive manner in which ethnographers revisit their writings, sketches, codings, and other materials during the research process. They understand their writing as creative and imaginative spaces to explore, think through, and represent knowledges, desires, and fears. For Arun Saldanha, writing gives 'existential embeddedness'. For Miles Richardson, 'poetics in the field and on the page' are inseparable. As described in the next section, ethnographers often experiment with forms of writing, such as trying out metaphors, first and third person points of view, different narrative structures, or fonts and typeface possibilities within the space of the page, as well as use images and other media to communicate complex temporalities and spatialities of the worlds they wish to represent.

New Directions: Transnational, Collaborative, and Creative Analytical Ethnographies

Geographers are particularly sensitive to the different forms of power/knowledge that enable access for some individuals and prevent movement for others. Because of the kinds of questions geographers ask, they are beginning to contribute in significant ways to new forms of ethnography, in particular about transnational and multitemporal spaces. Transnational ethnographic projects in geography analyze forms and effects of globalization through networks and multiple sites of: consumption and production; tourism and whiteness; transnational citizenship; black philosophical thought; and movements of peoples, including refugees and expatriates, across political territories.

Cindy Katz's concept of 'countertopographies', for example, provides a political and methodological approach to understanding and countering the uneven effects of global capitalism. Based upon many years of research, Katz juxtaposed the circumstances of children in two places – Howa (Sudan) and New York City – to make unexpected connections across space and time between the simultaneous processes of displacement for youth experiencing their childhood during periods of economic transitions. Such an approach challenges the abstract mappings of global capitalism methodologically by producing 'thick descriptions' of global economic restructuring processes such as deskilling; disinvestment in social reproduction; or the exoticization of 'remote' places (such as Howa) as 'objects' of capitalism. For Katz, the concept of countertopography is more than comparing one locale to another or even building coalitions between diverse places: it offers a powerful alternative spatial abstraction to globalization. By tracing connections among particular historical geographies and abstract social processes, countertopographies may enable new political possibilities and spatial imaginaries.

Katherine McKrittick's concept of 'cartographies of struggle', while not explicitly ethnographic, also offers a powerful approach to transnational ethnographic research. She looks to the literatures, philosophies, and material geographies of black women to connect their diasporic locations in Canada, the Caribbean, and the United States, both during and after the transatlantic slave trade. McKrittick analyzes locations such as slave auction blocks, black Canada, and new France, as well as conceptual spaces of feminism to explore the ways that particular women created powerful senses of place and sites of struggle to inhabit, negotiate, and challenge dominant mappings of their bodies and worlds. While marked by the legacy of violence and slavery, McKrittick argues that these black geographies are alterable and can provide the opportunity for change.

If ethnographic research is a practice of responsibility, to borrow from Marjorie Wolf's simple but prolific insight, than our work is a form of knowledge production that is never done in isolation. Indeed, ethnographers do not simply 'collect' forms of visual and material culture, or conduct interviews; they participate with others in its creation. Some geographers, drawing upon the work of feminist and local scholars, advocate participatory and community based research designs to develop cogenerative inquiry processes that decolonize knowledge productions and locate the research process outside of academic settings. Richa Nagar's collaborative work with the Sangtin Writers is an example of a many year process that included sharing personal diaries, hosting public and collective conversations (within and across countries), conducting interviews with numerous grassroots activists, and collaboratively writing about their experiences with nongovernmental organizations (NGOs) and development ideology in the Uttar Pradesh region of India. The

Sangtin Writers, Anupamlata, Ramsheela, Reshma Ansari, Richa Singh, Shashibala, Shashi Vaish, Surbala, and Vibha Bajpayee, are village level mobilizers who worked with Nari Samata Yojana (NSY) in the 1990s, a well known women's organization working with rural communities that later became the Sangtin group. As a scholar based in the US, Richa Nagar had come into contact with many of these individuals through her re search. After identifying their mutual frustrations about the problematic ways that outside agencies and re searchers take the ideas and voices of grassroots activists for their own agendas, they agreed upon trying out a longer term relationship to explore their mutual inter ests. Their story details this collaborative journey as well as their personal struggles and group successes. The Hindi edition of the book led to controversial public discussions about the issues confronted in the book, in cluding the sexist and patronizing attitudes rural women face when interacting with officials of large NGOs.

Ethnographers also try to establish responsible and respectful relationships with the peoples, communities, groups, nonhuman natures, and places they study through their writing as a practice of decolonizing methodologies. Jackie Huggins, Rita Huggins, and Jane Jacobs created a polyvocal text to describe their experiences returning to Carnarvon Gorge National Park in Australia. Jane Jacobs first experienced this place as part of a national heritage and interpretive rock art trail that guides tourists to an aboriginal sacred space. For Jackie Higgins, a historian, and her mother Rita Higgins, this is as a place of homecoming, known to them and other Bidjara people as Koor amindanjee. It is a desert oasis in the sandstone tablelands of the Great Dividing Range, where their families and the ancestors who went before them would rest, get water, eat, and perform ceremonies. Their text describes the colonial power relations that continue to shape the meanings and possibilities of dwelling by problematizing their roles as historian, mother, daughter, geographer, or friend.

In quite a different form, Annette Watson and Orville Huntington wrote a collaborative account about a Koyukon Athabascan moose hunting excursion that also departs from the ethnographic tradition of a Western scholar reporting upon local forms of knowing. But rather than intersperse personal accounts, Watson and Hun tington move back and forth between descriptive narra tives and theoretical insights (from both Western academic institutions and indigenous spiritual discourse), unresolved conversations between and moments of revelation. Their enfolding and unfolding narrative ab ruptly and often switches from Watson's to Huntington's first person points of view to communicate how their ideas evolved. They experiment with co authorship not so much to 'co produce' knowledge from Western and indigenous perspectives, but to perform the ways that the hunt reproduces the spiritual practices that connect hunter and prey. Significantly, their writing practice il luminates the more abstract concept of Koyukon Atha bascan epistemic space, a space wherein the worlds of humans, nonhuman lives, objects, and environments must be understood as more than any individual self: together they become an event.

Watson's and Huntington's work demonstrate yet an other way geographers are contributing to ethnography: through experimental forms of representation, or what Laurel Richardson describes as creative analytical practice. Such ethnographies perform, rather than simply docu ment, larger theoretical and conceptual claims, and may include: autoethnography, fiction, stories, poetry, drama, performance texts, polyvocal texts, readers' theater, re sponsive readings, comedy, satire, visual presentations, allegory, conversation, and layered accounts. The work of Miles Richardson, a particularly eloquent and creative writer, poet, and ethnographer (and adopted geographer), demonstrates multiple genre writing practices. His most recent work draws from 40 years of research about Catholics in Spanish America and Baptists in the Ameri can South, and is tied to his personal story of becoming an ethnographer and anthropologist. Part autoethnography, part poetry, part performance text, Richardson's vivid descriptions and intimate passages about his personal and ethnographic journeys with worshipers to holy places, pilgrimages, and ceremonies invites readers to consider, with the author, how death might be understood within the discourse of human evolution and in the context of post structuralist thinking.

Karen Till's 'geo ethnography' is an approach to re searching the complex spatial stories and social relations involved in place making, spatial hauntings, national identity formation, and memory work in Berlin. Using book sections, pages, narratives, and images in un conventional ways, she attempts to avoid a false sense of narrative coherence and thereby challenge traditional claims to ethnographic authority. Interlude writings, in the form of fieldnotes and from a first person point of view, provide an autocritque while connecting ideas across chapters; readers are introduced to places and mundane sites (fences, newspaper articles, citizen actions, and conceptual projects) that offer a different urban topography than the memory district of public places described in the formal chapters. These performative ethnographic fragments highlight how mundane experi ences in the city are also a way to understand spatial hauntings. Till understands the creative potential of ethnography as a place based form of memory work.

In addition to written texts with maps, sketches, and photographs, geographers are now experimenting with public artistic practices as part of their ethnographic re search. Toby Butler created an audio walking tour called *Memoryscape* about the local histories of two parts of the river Thames in London that is available free of charge online and through city agencies. In one of the tours, *Driftings*, he followed a piece of wood along the river and created 'stops' where it hit the bank. After conducting extensive research at each stop, which included talking to people, interpreting the landscape, getting archival ma terials, and other methods, Butler created a guided tour with music, interview passages, and commentary, to invite local residents and tourists to experience their environ ments in new ways. Holly McLaren's collaborative project, *Bordering*, was a scattered site exhibition and series of events in and around Oswestry, a small market town near the Welsh–English border. McLaren invited artists Ruth Jones, Stefhan Caddick, and Simon Whitehead and Tea to respond to this cultural and historic borderlands through new work that explores the ways that national borders shape knowledges and experiences of place. In addition to the curatorial role McLaren assumed in this project, she also set up an Internet blog where she posted her field notes and memos; invited artists to respond to her and other artists' ideas as they evolved; asked artists to post their concepts and ideas before and after the installation and exhibition; and invited locals and actual and virtual visitors to post their impressions during the process. The webpage, performance installations, exhibitions, walks, and the catalog explored the creative possibilities of transdis ciplinary collaboration about understanding the sensuous and fleeting lived experiences of place through the cre ation and languages of site specific and performance art.

As these examples demonstrate, ethnography as a methodology raises significant questions about the status and situations of knowledges used and created in geog raphy. Ethnographers do more than consider what methods generate the best data in their research design, which is a common way that social scientists use multiple methods, such as structured interviews, survey results, or census data. As a research approach, ethnography makes explicit processes and power relations that are involved in knowledge creation and selection, and explore inter subjective and responsible approaches to conducting research with partners.

See also: Auto-Photography; Embodied Knowing; Fieldwork; Interviews: In-Depth, Semi-Structured; Participatory Action Research; Representation and Re-presentation; Situated Knowledge, Reflexivity.

Further Reading

Clifford, (1988). *The Predicament of Culture. Twentieth Century Ethnography, Literature, and Art. Cambridge*, MA: Harvard University Press.

Clifford, J. and Marcus, G. (eds.) (1986). *Writing Culture: The Poetics and Politics of Ethnography*. Berkeley, CA: University of California Press.

Emerson, R., Fretz, R. and Shaw, L. (1995). *Writing Ethnographic Fieldnotes*. Chicago, IL: University of Chicago Press.

Katz, C. (2004). *Growing up Global: Economic Restructuring and Children's Everyday Lives*. Minneapolis, MN: University of Minnesota Press.

Parr, H. (1998). Mental health, the body and ethnography. *Area* 30(1), 28 37.

Richardson, L. (2003). Writing: A method of inquiry. In Denzin, N. & Lincoln, Y. (eds.) *Collecting and Interpreting Qualitative Materials* (2nd edn.) pp 499 541. Thousand Oaks, CA: Sage.

Richardson, M. (2003). *Being in Christ and Putting Death in Its Place: An Anthropologist's Account of Christian Performance in Spanish America and the American South*. Baton Rouge, LA: Louisiana State University Press.

Rose, G. (1997). Situating knowledges: Positionality, reflexivity and other tactics. *Progress in Human Geography* 21(3), 305 320.

Smith, L. T. (1999). *Decolonizing Methodologies: Research and indigenous Peoples*. London: Zed Books Ltd.

Spinney, J. (2006). A place of sense: A kinaestheitc ethnography of cyclists on Mont Ventoux. *Environment and Planning D: Society and Space* 24(5), 709 732.

The Sangtin Writers and Nagar, R. (2006). *Playing with Fire*: *Feminist Thought and Activism Through Seven Lives in India.* Minneapolis, MN: University of Minnesota Press.

Thomas, M. (2007). The implications of psychoanalysis for qualitative methodology: The case of interviews and narrative data analysis. *Professional Geographer* 59(4), 587 605.

Till, K. E. (2005). *The New Berlin: Memory, Politics, Place*. Minneapolis, MN: University of Minnesota.

Watson, A. and Huntington, O. H. (2008). They're *here*, I can *feel* them: The epistemic spaces of indigenous and Western knowledges. *Social and Cultural Geography* 9(3), 257 281.

Relevant Websites

http://www.borderingart.org.uk/about.html
Holly McLaren, *Bordering* webpage.

http://stefanobloch.wordpress.com/
Stefano Bloch, *Urban Aesthetics Research* blog webpage.

http://www.memoryscape.org.uk
Toby Butler, *Memoryscape* webpage.

Ethnomethodology/Ethnomethodological Geography

E. Laurier, University of Edinburgh, Edinburgh, UK

Glossary

Accountability All human actions are essentially observable and reportable. What is unusual in ethnomethodology is that, not only can people tell us about their activities, but also that those activities are produced so as to be describable in certain ways; in other words, they are 'reflexively' accountable.

Epistopics A series of epistemic themes and classical sources of theoretical problems in natural and social sciences such as observation, replication, measurement, imagination, and explanation. Ethnomethodology seeks a return to them as ordinary practices (e.g., birdwatchers observing eagles, neurologists redoing an experiment, economists measuring hospital efficiency, landscape gardeners imagining a new shrubbery, and explaining why one was so upset at work today) and in doing so remove the metaphysical aura of these grand topics.

Haecceity What makes an object uniquely what it is, or, the 'just thisness' of a thing. In ethnomethodology, there is a desire to get away from assuming we know what even the most ordinary things are and follow what all the 'this' and 'that' are of any action, event, conversation, and so on.

Indexicality A long-standing and underpinning term in ethnomethodology. Simply put, the relevance, meaning, appropriateness, and correctness of any expression varies as it is put to use in different settings, by different agents, and at different times. Indexicality causes endless problems for those who would try and replace subjective expressions with objective expressions. Nevertheless, and confusingly to many who would assume otherwise, indexical expressions can, and do, nevertheless have rational properties.

Ironic Attitude A common stance of the social sciences toward members of societies' understanding of events, that what is happening is 'not' what they think is happening; it only seems to be like that. From the point of view of the skeptical social scientist, something else is going on. Such an attitude usually substitutes the intelligibility of action at source with explanations from theory that thereby ironize the words and deeds of members.

Reflexivity Related to accountability and indexicality, the concreteness, sense, and intelligibility of human actions are tied to the settings in which they occur, yet also, whose sense they produce. Ethnomethodology diverges from many of the social sciences who have treated a concern with reflexivity as allowing them to have a superior knowledge over the common sense of societies' members.

Respecification Rather than building new theories or models of society, ethnomethodology respecifies existing epistopics in the light of ordinary practice. There is a desire to investigate foundational matters in human geography and the social sciences more widely through carrying out empirical studies. Bewilderingly, for social scientists that hope for laws or unchanging rules that govern society, ethnomethodological studies do not lead to generalization, instead they surprise from actuality and, sometimes, a form of therapy for the troubles of generalization in the social sciences.

Skepticism Related to the ironic attitude, ethnomethodology is a response to the skeptical attitude in the social sciences which constantly questions whether members of society see things as they really are or are being duped by other (usually larger) actors or foundational ideologies. Ethnomethodology does not seek to defeat the powerful skepticism of the social sciences, rather to show how it reduces, or just misses, all manner of wonderful sense and sense making in human practice.

Unique Adequacy Worried about the social sciences' tendency to miss the core of what constitutes numerous practical fields of activity, ethnomethodology suggests that its students should achieve at least a vulgar competence in the skills of the groups they are studying. If students remain incompetent in the practices they wish to depict, or to put it more politely, retain a disengaged perspective, they will miss what the activity consists of in terms of its skills, rules, morals, reason, order, and more.

Definitions

Ethnomethodology is the study of folk or members' methods for producing recognizable and reasonable so cial orders, its title deriving from related terms such as ethnobotany which is the study of folk knowledges of plants. It is as happy studying how street sellers pitch their wares as it is examining how social scientists code interview transcripts. Or, as Garfinkel put it during the Purdue symposium,

> [T]here are now quite a number of persons who, on a day to day basis, are doing studies of practical activities, of commonsense knowledge, of this and that, and of

> practical organizational reasoning. That is what ethno methodology is concerned with. It is an organizational study of a member's knowledge of his ordinary affairs, of his own organized enterprises, where that knowledge is treated by us as part of the same setting that it also makes orderable. (Hill and Stones Crittenden, 1968: 10)

This simple definition belies a long standing conten tious relationship with the social sciences and an en gagement with ordinary practice in the face of theorization and modeling that is as powerful as it is puzzling.

Origins

Ethnomethodology as a distinctive approach is commonly accepted to have been initiated by Harold Garfinkel in the 1950s in response to a series of problems he had come upon in the 1930s and then in pursuing his PhD in the 1940s under the supervision of Talcott Parsons. His initial work was contemporary with that of Ludwig Wittgenstein and C. Wright Mills, sharing their concerns with the limitations of rules, references, and individualism that had been handed on as solutions to the problem of meaning and intelligibility. It was the publication of 'Studies in ethnomethodology' in the late 1960s that really brought what had by then already become a congregation of practitioners (including Egon Bittner, Ed Rose, and Aaron Cicourel) to the attention of the social sciences more widely. At this stage it was centered around the University of California's campuses and attracted a reputation for being both a committed community of scholars, and something of a cult, that it has never quite lost. It was also at this stage that Harvey Sacks was collaborating with Harold Garfinkel, a fruitful partnership that led to the initiation of conversation analysis as a fraternal twin of ethnomethodology. Ethnomethodology caught the atten tion of a number of young scholars who went on to be come significant figures in sociology, such as David Sudnow, Lawrence Wieder, Don Zimmerman, Wes Shar rock, Peter McHugh, Jeff Coulter, and Rod Watson. During this period it also spread beyond the Anglophone academy, becoming established in France and Japan.

While ethnomethodology does not appear to have been adopted as an approach by any particular human geographers during its inception, it was, and remains, congruent with the contemporary concerns of humanistic and phenomenological practitioners in the discipline. Indeed, with the return to prominence of cultural geography within human geography, there has come a renewed commitment to, and concern with, place over more abstract conceptions of space. Human geography today has a still deeper affinity with the early and later concerns of ethnomethodology in its attendance to the conjointness of human practices and particular places. Moreover, they are both concerned with how it is that similar forms of agency or subjectivity or identity emerge differently in each and every place.

Key Concepts

One should be slightly, no 'very', cautious, about pro ducing headings that would count as a list of 'key concepts' in ethnomethodology. A central pursuit in ethno methodological studies is respecification of concepts in the light of detailed investigations of particular cases and settings. Notionally, any concept is therefore liable to respecification in the same way that concepts from social or cultural theory are open to their deconstruction by Derrideans. In that light, we should note the further har monies with a human geography influenced by post structuralism, in its many shapes and forms, that has sought to replace singular and overarching grand theory with more modest and plural theories. Ethnomethodology diverges from many of the more philosophically grounded post structural geographies by prioritizing detailed in quiries into practice. This empirical commitment and fo rensic investment in the details is not to everyone's tastes.

What is 'methodological' about ethnomethodology is the focus on the logic of methods, or in more familiar terms, the logic of practice. This focus is applied across the board from ordinary social situations such as queuing for a bus to social scientists doing standardized surveys. The lack of acceptance of an in principle distinction between social science methodologies and other more mundane practices had led to some hostility toward ethnomethodology. It has been taken to be bringing the social sciences down from their privileged epistemo logical position into the muddle of ordinary affairs, whereas what it was doing was raising mundane methods up to the level of more formal methodologies.

In providing a brief list here, the aim is not to limit the concepts; it is instead to offer a flavor of the sorts of concepts that have already been the subject of rigorous study by ethno inquirers.

Accountability

Where we might be used to thinking about accountability as definitively characteristic of institutions such as gov ernments and professional bodies to have to provide justifications of their collective actions, ethnomethod ology explores accountability as an accomplishment, background expectation, and ongoing concern of human action. It is used somewhat interchangeably with the compound word observable–reportable. What we do be it as, for instance, a local councilor or Simon's girlfriend, has the characteristic of being always observable and reportable. All human practices, be they of the investigator or the member of cultures, are taken to be more, or less, accountable to those involved in them. How we 'see' and speak of the reason, morality, and motives of others (and of ourselves) is thoroughly bound up with the observability and reportability of those practices and equally our seeing and speaking has its observable and reportable characteristics.

Reflexivity

Where we left off with accountability is where we take up reflexivity. In the social sciences the dominant view of re flexivity is that it is a privileged maneuver of social scientists that allows them one step up above naïve uses and understandings of language. The social sciences are aware of how certain words or images are ideologically loaded, incorporate hidden meanings, and investigate their speech and writing for similar investments. Ethnomethodology puts reflexivity at the heart of its study of human affairs by treating it as incarnate in those human affairs. In line with raising members' practices onto the level with the social sciences, members are equally capable of taking one step up and commenting on their and others modes of expression. Ethnomethodology cuts itself adrift by taking social order as locally produced in whatever settings, descriptions (or representations) of particular courses of action which are reflexive to those self same settings. Reflexivity thus over laps with accountability in the ongoing production of order in each and every place in the world.

Mind

Hopefully there will now be some familiarity with the ethnomethodological drift when it is added that a series of topics related to the study of mind such as motives, rea sons, intentions, perceptions, imagination, memory, and cognition are taken by ethnomethodology to be locally observable–reportable features of human action. Setting itself against 'mental' explanations of how people re member, reason, imagine, see, and so on, ethnomethod ology is concerned with how such apparently 'private' processes are publicly available. Not only are they publicly available, it is our ongoing concern to make them so.

Programs

A recent collection of Garfinkel's papers was entitled 'The ethnomethodological program'; it might have been better expressed in the plural. Since its inception, ethno methodology has sired a number of more or less legitimate offsprings. To begin, once again, where we left off, with the concept of mind, this run of studies is concerned centrally with practical action and practical reasoning. In some ways it has had the longest run given that it begins with Gar finkel's famous 'breaching experiments' which were designed to cause the collapse of social order. Therein Garfinkel had his students carry out experiments which disrupted the intelligibility of various places, such as act ing as lodgers within their own home or refusing to pay the display price of items in shops. Rather than social order collapsing, as certain social theories had posited, it was kept in place, though not without all manner of trouble, complaints, and some pleasing discounts for Garfinkel's students in department stores.

As noted earlier, the most famous offshoot of ethno methodology is conversation analysis. Its concern is with the ways in which those who converse are also analysts of the conversation they are ongoingly having with one another. The name of this approach is in a classic ethnomethodological manner the study of the analysis done in, by, and with conversation rather than a pro fessional and/or scientific abstraction of conversation. However, and somewhat confusingly, conversation an alysis has also ended up becoming a technical study of language in interaction that in various ways supplants the methods of members with its own more technical lexicon.

Conversation analysis began with Harvey Sacks looking at the detailed methods used by speakers during phone calls. His studies were unusual at the time for making extensive use of audio recordings of actual phone calls to a suicide helpline and between friends, along with other sources of conversation on the phone. Drawing on these materials he respecified a number of major philosophical topics such as measurement, observation, and, once again, mind. Of great importance were how speakers relied on and put to use membership categorization devices and the sequential properties of conversation. This work led ultimately to one of the most cited papers of all time in the social sciences on the turn taking mechanism. For analyzing transcripts con versation analysis has been taken up in human geography as an alternative to software assisted coding. Where the latter is driven by content alone, the latter is a meeting point be tween content and process. Conversation analysis helps us grasp not only what people have said but how they said it and what they were trying to achieve at the time.

A third branch of ethnomethodology, and one that is perhaps the most obviously geographical, is workplace studies which takes as its concern the production of order in each particular setting of work. While these range across scientific laboratories, truck garages, martial arts studios, court rooms, classrooms, banks, and air traffic control centers, they have a common concern with the routine features of each workplace as they are produced ongoingly, day by day. These features are not merely routine; they are also taken to be the objective, if disputable, phenomena of these workplaces. The constitutive gap between their ob jective qualities and their indexicality remains the sig nificant focus for workplace studies.

For a while it seemed like ethnomethodology, with its ethos of having its practitioners become their phenomena (e.g., to study law become a lawyer, to study jazz learn to play jazz piano), might disappear into other professional fields entirely. However, in two interdisciplinary realms it has reappeared as a significant school: computer supported cooperative work (CSCW) and science and technology studies. Whether ethnomethodology will become an es tablished program in geography is not yet clear.

The grounds for the establishment of an ethno methodological program in geography would at first glance be its harmonies with the archeological approach of Michel Foucault, the underpinnings it has provided in social studies of science and technology, and its ways of accessing and describing everyday life and the ordinary. There are at least two further passages from ethno methodology to human geography. The first is in its treatment of speech as part and parcel of human action which allies it with the nonrepresentational work of Thrift and others. The second is in its revivification of empirical work and field studies which are there to respecify and put to rest a number of theoretical prob lems that beset human geography.

Central Misunderstandings

Ethnomethodology seems to have been beset by more misunderstandings than most approaches in the social sciences. Had ethnomethodology come into being now rather than in the midst of the post war positivist dom inance in the social sciences, there might have been far fewer. To mention three here:

1. Ethnomethodologists collect and document various sorts of 'folk wisdom' that either underlie or distort more scientific knowledge of the world. From the Purdue Symposium:

   > ANDERSON: … There is still the question of whether your concern is that of an outsider studying folk wisdom.
   >
   > GARFINKEL: NO! Once and for all, no! We are not studying folk wisdom in an ironic way. I am not saying that I know better. I am not armed with resources that would permit me to say, no matter how discreetly, "Look, does the botanist believe there are salt water fish in a fresh water lake? Get that!" Nor am I saying things like, "The Catholics believe that whatever it is; the Jews have the inside track on that one." There is no irony. (Hill, 1968: 28)

   An ethnomethodological geography would not supply 'folk geographies'; to do so would be to ironize how knowing certain things constitutes certain com munities and equally how certain communities make sense of their actions.
2. Ethnomethodology is a method. Unsurprisingly, given the positioning of methodology in its title many as sumed that ethnomethodology is a methodology of the social sciences or is a variant on ethnography. It is neither. As noted at the beginning of this article, it is the study of methodologies, be they those of members of a queue at a bus stop or members of a neuroscience lab examining images from electron microscopes.
3. Ethnomethodological studies are microstudies. Be cause ethnomethodology remains stubbornly attached to studying the details of diverse local sites of action, such as looking for a book in a library, a conversation among teenagers, or learning to playing piano, it is often taken to task for missing structures that operate at a larger scale. However, the idea of how a larger context or scale exists in ethnomethodological studies is the same as the treatment of scale or context in actor network theory. Large scale structures are, in their haecceities, inevitably and reflexively en countered locally. The questions raised in ethno methodology and actor network theory are: how do they go about 'localizing' themselves and how do they go about 'extending' themselves into other places. More specifically the concern is whether, and how, larger contexts are made locally available and locally recognizable in any particular course of action.

Future Directions

To maintain its vigor ethnomethodology has always hy bridized with other disciplines and crafts. As noted earlier the meeting of computing science with ethnomethod ology in the interdisciplinary zone of CSCW gave fresh impetus to its workplace investigations. The craft skills of writing software and building prototypes required the forms of detailed description of work practice that ethnomethodologists were exceptionally good at offering. Equally, the marrying of science studies and ethno methodology in the work of Lynch, Livingston, Bjelic, and others has brought a number of new concerns such as, to name but three, epistemology, evidence, and equipment.

Where human geography might entwine with ethno methodology is, firstly, in its epistopics, some shared with the other social sciences and humanities, and others less so; context, observation, representation, and imagination to name but a few. Second, there are a number of 'wild geographies' which are of interest to human geographers and ethnomethodologists. The vernacular geographical knowledges emergent in endless varieties, such as what neighbors know about their neighborhood, how tourists find their way around cities, how browsers search for books in libraries, and how people formulate their location during mobile phone calls. Ethnomethodology could offer a way for human geography to draw succor from these wild variants rooted in other life worlds than the academy.

Critiques

Over the years Bruno Latour has grown increasingly sympathetic to ethnomethodology. His initial responses were somewhat more critical than we now find him. In particular he was uneasy about the lack of a 'big picture' or, rather, finding small things rendered large. Here, commenting on an ethnomethodological study of a laboratory.

> In reading the book, one has the same feeling as reading a newspaper through a microscope. Somehow the focus does not seem right, and one is tempted to ask, 'Please, Mike, couldn't you zoom a bit the other way, I can't see a damn thing here'. (Latour, 1986: 545)

Latour is not suggesting that the details are not important. What he wants is the ethnomethodologist to travel away from the local worksite to see how it is connected to other worksites. There is no strong reason why ethnomethodologists do not follow chains of connection between worksites, though they would be looking for a group of some sort whose daily business is to follow those sorts of connections rather than trying to invent a new methodology for the social sciences that provides a privileged perspective on the action.

Perhaps the most substantial critique comes from Alan Blum and Peter McHugh whose school of analysis grew out of ethnomethodology. Theirs remain, like Latour's, an understanding of ethnomethodology that is sympathetic, informed, and respectful. What formed their point of departure from ethnomethodology was a disinterest in doing field studies and a return to theorizing which, at first brush, appears antithetical to the spirit of ethnomethodology. Yet, the sense of being a respecificatory response to theorists in the social sciences is retained. Their critique of ethnomethodology resides in an argument more complex than be rehearsed here on whether irony is always invidious, what the nature of theorist and member relationship is, and the limits of convention. Ethnomethodology, argue Blum and McHugh, reaffirms convention when it could question it and refuses to exercise its authority to formulate excellence for members.

See also: Actor-Network Theory/Network Geographies; Ethnography; Non-Representational Theory/Non-Representational Geographies; Regional Geography I.

Further Reading

Bjelic, D. (2004). *Galileo's Pendulum: Science, Sexuality and the Body Instrument Link*. New York: State University of New York Press.
Blum, A. and McHugh, P. (1984). *Self Reflection in the Arts and Sciences*. Atlantic Highlands, NJ: Humanities Press.
Brown, B. A. T. and Laurier, E. (2005). Maps and journeys: An ethnomethodological investigation. *Cartographica* 4(3), 17 33.
Button, G. (ed.) (1991). *Ethnomethodology and the Human Sciences*. Cambridge: Cambridge University Press.
Carlin, A. P. (2003). Pro forma arrangements: The visual availability of textual artefacts. *Visual Studies* 18(1), 6 20.
Crabtree, A. (2003). *Designing Collaborative Systems: A Practical Guide to Ethnography*. Heidelberg: Springer.
Crabtree, A., Nichols, D. M., O'Brien, J., Rouncefield, M. and Twidale, M. B. (2000). Ethnomethodologically informed ethnography and information system design. *Journal of the American Society for Information Science* 51(7), 666 682.
Garfinkel, H. (1963). A conception of and experiments with trust as a condition of stable, concerted actions. In Harvey, O. J. (ed.) *Motivation and Social Interaction*, pp 187 238. New York: Ronald Press.
Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice Hall.
Garfinkel, H. (ed.) (1986). *Ethnomethodological Studies of Work*. London: Routledge and Kegan Paul.
Garfinkel, H. (2002). *Ethnomethodology's Program, Working Out Durkheim's Aphorism*. New York: Rowman and Littlefield Publishers.
Harrison, P. (2008). Corporeal remains: Vulnerability, proximity and living on after the end of the world. *Environment and Planning A* 40(2), 423 445.
Hill, R. J. and Crittenden, S. K. (eds.) (1968). *Proceedings of the Purdue Symposium on Ethnomethdology*. Institute for the Study of Social Change, Purdue.
Latour, B. (1986). Will the last person to leave the social studies of science turn on the tape recorder. *Social Studies of Science* 16, 541 548.
Latour, B. (2003). *Paris*: *Invisible City*, Paris, http://www.bruno latour.fr/virtual/EN/index.html, Photos by Emilie Hermant, Web Design by Patricia Reed.
Laurier, E. (2001). Why people say where they are during mobile phone calls. *Environment and Planning D: Society and Space* 19(4), 485 504.
Laurier, E. and Philo, C. (2004). Ethno archaeology and undefined investigations. *Environment and Planning A* 36, 421 436.
Laurier, E., Whyte, A. and Buckner, K. (2002). Neighbouring as an occasioned activity: Finding a lost cat. *Space and Culture* 5(4), 346 367.
Livingston, E. (1986). *The Ethnomethodological Foundations of Mathematics*. London: Routledge and Kegan Paul.
Livingston, E. (1987). *Making Sense of Ethnomethodology*. London: Routledge.
Luff, P., Hindmarsh, J. and Heath, C. (eds.) (2000). *Workplace Studies*. Cambridge: Cambridge University Press.
Lynch, M. (1993). *Scientific Practice and Ordinary Action: Ethnomethodology and Social Studies of Science*. Cambridge: Cambridge University Press.
McHugh, P., Raffel, S., Foss, D. C. and Blum, A. F. (1974). *On the Beginning of Social Inquiry*. London: Routledge and Kegan Paul.
Mondada, L. (2007). Deixis spatiale, gestes de pointage et formes de coordination de l'action. In Barberis, J. M. & Manes Gallo, M. C. (eds.) *Parcours dans la ville*. *Descriptions d'itinéraires piétons,* pp 211 230. Paris: L'Harmattan.
Sacks, H., Schegloff, E. A. and Jefferson, G. (1974). A simplest systematics for the organization of turn taking for conversation. *Language* 50, 696 735.
Thrift, N. (1996). *Spatial Formations*. London: Sage.
Weilenmann, A. (2003). "I can't talk now, I'm in a fitting room": Formulating availability and location in mobile phone conversations. *Environment and Planning A* 35(9), 1589 1605.
Wylie, J. (2005). A single day's walking: Narrating self and landscape in the South West Coast Path. *Transactions of the Institute of British Geographers* 30(2), 234 247.

Relevant Websites

http://www.mc.murdoch.edu.au
Australian Institute for Ethnomethodology and Conversation Analysis, School of Media Communication & Culture.

http://www2.fmg.uva.nl
Information on Ethnomethodology and Conversation Analysis, University of Amsterdam.

http://www.iiemca.org
The International Institute for Ethnomethodology and Conversation Analysis.

Eurocentrism

J. Sundberg, University of British Columbia, Vancouver, BC, Canada

Glossary

Epistemology A theory about what constitutes valid or legitimate knowledge.

Ontology The study of being or the nature of knowable things.

Introduction

Eurocentrism has been variously defined as an attitude, conceptual apparatus, or set of empirical beliefs that frame Europe as the primary engine and architect of world history, the bearer of universal values and reason, and the pinnacle and therefore model of progress and development. In Eurocentric narratives, the superiority of Europe is evident in its achievements in economic and political systems, technologies, and the high quality of life enjoyed by its societies. Eurocentrism is more than banal ethnocentric prejudice, however, as it is intimately tied to and indeed constituted in the violence and asymmetry of colonial and imperial encounters. Euro centrism is what makes this violence not only possible, but also acceptable or justifiable. As such, Eurocentrism is the condition of possibility for Orientalism, the discursive and institutional grid of power/knowledge integral to the production and domination of the Orient as Other. Sig nificant critiques of Eurocentrism emerged in the context of post World War II shifts in geopolitical power, in cluding anticolonial and anti imperial revolutionary movements. Even so, Eurocentric epistemologies con tinue to haunt the production of knowledge in geography in significant and disturbing ways.

In conventional Eurocentric tellings, Europe is the engineer and architect of modern agricultural, cultural, economic, political, and scientific innovations, including capitalism, democracy, and industrial, medical, and green revolutions. Concepts like 'the rise of Europe' and 'the European miracle' exemplify Eurocentric models of history and development. Europe's so called rise is ex plained in terms of superior social and environmental qualities deemed internal to it: inventiveness, rationality, capacity for abstract thought, outward looking, freedom loving, along with advantageous climate and geographies. Many of these cultural traits are said to be inherited from the Bible lands and ancient Greece and Rome – framed as Europe's ancestral hearths – though their highest de velopment is said to have been achieved first in imperial England and then the United States of America – hence the term 'Euro Americanism'. In these narratives, pro gress and development ride what James Blaut calls 'the westbound Orient Express'.

As a consequence of the perceived historical move ment of the westbound express, 'Europe' has morphed into the 'West' and now the 'Global North'. These fluid geographic imaginaries may refer to not only Europe and white settler societies like the United States, Canada, and Australia, but also Japan and any other region or group that envisions itself as the possessor or inheritor of European culture, values, and academic, political, and economic systems. At the same time, however, particular places within the West such as the United States are privileged as the source of universal theory, while others like New Zealand are framed as limited by their par ticularities. Latin America and the Caribbean were col onized by Europeans, but are rarely included in the West. In short, it may not always be clear to what exactly these geographical imaginaries refer, but they are used as though they correspond to a commonsensical external reality. Through their repetition in everyday speech and academic and institutional narratives, that reality is continuously brought into being.

Eurocentrism in Action

People the world over encounter Eurocentric images, attitudes, categories, and stories everyday. Whether glaringly obvious or mundane and unremarkable, Euro centric expressions constitute the texture of daily life for many people. Indeed, it is difficult to encounter an arena of socioeconomic and political life that has escaped its hold. Although Eurocentrism is not a stable and trans historical mindset, an archive of codes, categories, and conceptual structures has emerged over time from which historically and geographically situated social actors may draw. As individuals recite and reconfigure these elem ents in new yet similarly structured ways, they put Eurocentrism into action. In other words, Eurocentric narratives are also doings: they produce the cosmologies and conceptual orderings they purport to describe.

Stories about history and human development are prime sites of Eurocentrism in action. Murals gracing the halls of many a government building in white settler so cieties are especially obvious examples. The small town of Magdalena in Sonora, Mexico hosts a vivid illustration. Magdalena is an important destination for religious pil grims from northern Mexico and the southwestern United States, for it is where a beloved Italian Jesuit missionary, Father Eusebio Kino is buried. The mausoleum built around Father Kino's grave in 1966 is adorned with a mural depicting his work in the region. A small plaque explains Father Kino's life and describes his missionary efforts from 1687 to his death in 1711. Father Kino, the plaque indicates, founded 30 pueblos (settled communities of native peoples) and was a "defender of the indigenous peoples." In addition: "He taught them to build homes and to carry out farming and herding."

The mural painted by Nereo de la Peña tells a classic Eurocentric story about history, development, and colonialism. In the segment pictured here, Father Kino is on the left, a Native American group on the right (see **Figure 1**). Each is portrayed prior to their encounter, which is depicted in the space between them. Father Kino is shown sitting on his horse, gazing into the distance, surveying the lands claimed by the Spanish Crown. The maps looming above Father Kino are indicative of his assignment as Royal Cartographer; as explained in the plaque, Father Kino was a "respected scientist who discovered that Baja California was a peninsula and not an island as the cartographers of the time erroneously believed." Despite the heat of the Sonora Desert, Father Kino is fully clothed in heavy black robes. As evidenced by his solitary stance in the desert, Father Kino is a fearless explorer, missionary, and colonizer.

To the far right of Father Kino is pictured a group of Native Americans – portrayed as a nuclear family – from Pimeria Alta, the Spanish colonists' name for northern Sonora and southern Arizona. In contrast to Father Kino, the native people are representatives of their kind rather than individuals with names. They are scantily clad: the children are completely naked while the adults cover their genitals with rough skins; the woman's prominent breasts are bare. Lying at the feet of this native group are the fruits of their technological achievements: baskets and pottery. Their agriculture is portrayed as a chaotic patch of maize stalks; they appear to have no other food. The man and boy hold small, rustic looking weapons. Directly above the adult male's head is a native symbol (although it looks more like a brand once used to mark slaves and cattle as property).

The encounter between them is pictured between these two sets of images. Here, Father Kino is portrayed as a teacher: he instructs a native man, now wearing pants, how to use a horse drawn plow, a technology used in the Mediterranean. The result is a tidy, neatly arranged agricultural field. Through religious instruction and technological diffusion, the image suggests, Native Americans are civilized and transformed into orderly and productive individuals.

In what ways does this mural represent Eurocentrism in action? In other words, how does this mural reflect and reproduce a Eurocentric story? First, the mural suggests that history and development are driven by Europe. America is nothing – naked and unorganized – until awoken by a European man. Like many of its kind, this Eurocentric story of development reflects gendered hierarchies: America is feminized, a Sleeping Beauty awaiting the touch and tutelage of her Prince. Second, Father Kino appears to explore and colonize alone, without the assistance or participation of native peoples as guides or allies. Hence, the cartographic knowledge he produces is portrayed as the solitary outcome of his explorations.

Figure 1 Father Kino's mausoleum, Magdalena, Sonora, Mexico.

Europeans appear to be innately capable of conquest, colonization, and accurate cartography. Third, development is unidirectional: innovation flows from Europe to others. It is Father Kino who brings technology to the Americans, who only have baskets and pottery, which are associated with women's work. In this way, native technologies are further feminized, while European technology is pictured as a horse drawn plow driven by a man. Fourth, Europe gives, but is not transformed by the process. As the embodiment of European civilization, scientific knowledge, colonial power, and religious authority, Father Kino is already a fully formed individual. It is the naked, unnamed, and undeveloped native peoples who need formation. Through such framings, European and Indian become fixed identities, hierarchically organized racial categories that purport to tell the truth about their inner qualities. Fifth, the relationship between adult and child is used as a metaphor for development and mapped onto the racial categories of European and Indian. In the mural, native peoples are infantilized and pictured as Father Kino's pupils; in religious terms, he is literally their father. Through his instruction, they are guided through the stages of growth to become mature adults. Finally, colonialism is pictured as a benevolent process and is about spreading the fruits of European knowledge and development. The violence of colonialism is concealed and rendered unimportant to the overall story of progress.

The mural points to key elements of Eurocentrism. First, Eurocentric modes of representation produce polarized and hierarchical stories featuring the West and the Rest as the primary characters. Regions outside of Europe only come on the stage of world history when they are colonized by Europe; their cultural and technological achievements may be appropriated, but their contributions to modern formations are occluded. As a consequence of such representational practices, world regions and cultures appear as autonomous, bounded units, each with their own characteristics. The interconnections between them are rendered invisible or unimportant. Second, Eurocentric representations make the West the beacon of enlightenment and progress by concealing and denying the violence, genocide, and dispossession of colonialism. Western myths of modernity, Enrique Dussel argues, justify colonialism in terms of the gifts of enlightenment and civilization given to its victims.

These elements of Eurocentrism are alive and well in schools and universities today. They are put into action through the continuing use of dualistic models that make Europe or the West the referents of analysis, the yardsticks by which to analyze and represent other peoples and places. In the nineteenth century, Western scholars deployed dualisms like civilized/barbaric or advanced/backward to organize the world's people through reference to the racial superiority of Europeans. Since the mid twentieth century, knowledge is organized through dualisms like modern/traditional, developed/undeveloped, core/periphery, or developed/developing. These dualisms replace notions of racial difference with cultural conceptions of development and progress. However, the dualisms of the past are not so different from those commonly used today: both sets of configurations make the West the referent of analysis; both treat each side of the dualism as a bounded and discreet entity unto itself; both posit evolutionary schemas through which societies inevitably progress; and both have their genesis in colonial or imperial relations of power. The discursive formations may have shifted but the underlying presumption of a superior white Western self as referent of analysis remains the same. The continuing authority of these dualisms as valid tools to organize and understand the world pivots on the erasure of their provincial origins.

In addition, the intersection between academia and geopolitical power means that Western scholarship defines the frames of reference with which others must engage to be intelligible/legible. European or Western theories, concepts, and models are held up as universally applicable and treated as valid tools with which to analyze the world in all times and places. Social scientists commonly deploy concepts like capital, class, and patriarchy as though they can be applied universally without a prior analysis of if and how they are constituted and given meaning in particular places. As a consequence, argues Dipesh Chakrabarty, European or Western thinkers are the only ones alive in the social sciences, no matter how long they have been dead. For Chakrabarty, this form of Eurocentrism is evident when scholars the world over must refer to European history and concepts to obtain international recognition, but Westerners are not obliged to reciprocate. In short, ignorance of the non Western world and its scholarship does not affect the quality, objectivity, or reception of Eurocentric work.

Critiques of Eurocentrism and Their Limitations in Geography

There have been critiques of colonialism and imperialism ever since there were colonial and imperial subjects. However, critical analyses of Eurocentrism as a totalizing meta narrative about world history and development emerged in the era of anticolonial and anti imperial revolutionary movements after World War II. Initially, these critiques were elaborated in the form of empirical debates about the privileged position of Europe in world history. The rise of Europe thesis postulates that Europe's dominance on the world stage is the result of special qualities internal to it, which led it to catapult ahead of everyone else in the late 1400s by reaching the previously unknown continent now called America. Inspired by the revolutionary and anti imperial sentiment of the day, non European and European scholars alike drew on emergent empirical evidence to argue that Europe was not more progressive or technologically advanced than other societies in the 1400s and in fact, was on the margins of the world system at that time. Researchers also contested the argument that the Eurocentric capit alist world system is the result of European superiority and demonstrated how this system came into being through European colonial expansion and the systematic dispossession and underdevelopment of its colonies. Such studies treat the developed and underdeveloped regions, not as mutually exclusive and static categories, each with their own cultural traits, but as intimately linked in his torical, politico economical, and epistemological terms.

Although these influential critiques have sparked new lines of inquiry in geography, Eurocentric versions of his tory and development continue to overshadow the discip line, structuring silences and limiting the kinds of questions that will be asked. In James Blaut's view, the continued acceptance of Eurocentric interpretations must be under stood in the context of the West's powerful position in global geopolitics. Eurocentrism, Blaut argues, is simply "the colonizers model of the world." As such, criticizing Eurocentrism on the grounds that it is built upon factual errors, untruthful beliefs, or ahistorical visions of develop ment will not lead to its undoing. Contesting Eurocentrism and challenging its on going hegemony in geography ne cessitates analyzing the discipline itself at institutional and epistemological levels.

Eurocentrism and Geography's Hallowed Halls

Anthropology, geography, linguistics, and other social sciences were constituted in and through a colonial social order. Western academia speaks from this privileged position, a place of enunciation rooted in imperial power. Geography was instrumental in colonial and imperial exploration and mapping, as well as strategies of and justifications for dispossession. For instance, the Spanish Crown made the production of geographical knowledge integral to the consolidation of colonial power when it commissioned the *Relaciones Geográficas* in 1577. Spanish authorities obtained important knowledge about popu lations, languages, and physical terrain in New Spain, but they considered the maps they obtained to be useless failures because many reflected a native cosmology and therefore were deemed unintelligible. Geographical modes of looking at and representing the world, then, have been historically centered in and central to Euro pean colonialism; they also are Eurocentric.

This history is key to the discipline's formal in corporation into Western academic structures and the ethnic and gender composition of its practitioners. The continuing dominance of Western white men of elite or middle class standing in geography departments is evidence of this history. Indeed, geography as a discipline presumes and universalizes a white, Western masculine subject position as the norm, the frame of reference, and referent of analysis. Feminist geographers challenge geographical claims to universality on the grounds that they mask and protect very particular masculinist inter ests and perspectives; however, feminism has not been immune to Eurocentrism. Indeed, feminist geography has been accused of projecting a white, Western middle class feminine subject as the frame of reference and referent of analysis.

Moreover, institutional practices in graduate pro grams, tenure and promotion processes, publication re view, as well as funding in Anglo American geography tend to privilege Anglo American scholarship and Euro centric ways of knowing and writing. European or Western scholars are the mainstay of scholarship. Latin Americanist geographers, for instance, are not obliged to refer to Latin American scholarship to produce work that will be given the stamp of approval in the field. Geog raphers situated in other parts of the world, however, find that they must engage Western scholarship to be recog nized as international. In Anglo American institutions, international is a circumscribed imaginary. Publishing in Latin American journals, for example, will not be given the same intellectual weight by tenure and promotion committees or funding agencies as Anglo American or European journals.

In addition, the discipline of geography has difficulty attracting and retaining scholars of non European des cent. Indeed, individuals with established careers along with promising graduate students have left the discipline. As the dominant sociospatial experience in geography, whiteness distorts our intellectual production and en sures that Eurocentrism lives on in the discipline. Al though geography is not alone in its disciplinary history, fields such as anthropology have made more systematic and systemic efforts to transform their institutional structures, the gender and ethnic composition of their departments, and therefore the kinds of knowledge produced.

Eurocentric Epistemologies

To comprehend the continuing power of Eurocentric ways of knowing and conceptual strategies in the dis cipline, geographers are increasingly pointing to the hegemony of specific epistemologies and the ontologies they enable. As a discipline built on Western phil osophies, geography is structured by a set of ontological ruptures; specifically, the Cartesian split between mind and body, reason and material world, wherein rationality is framed as external to the body and embodied experience. In addition, this framework empties the body and material world of meaning and significance while objectifying them as available for interrogation, measurement, and analysis. Given that Descartes positioned the mind in an instrumental relationship to the body and material world, this dualism also inscribes hierarchy and domination into the relationship between things.

These ontological ruptures underwrite conventional notions of objective knowledge production as a de corporealized and de contextualized process. Claims to objectivity and therefore universality are founded in the assumption – as in the act of taking on and taking for granted – of an unmarked rational gaze that is stripped of and separated from the body, material social relations, and geographical location. The 'geographer as objective observer', then, is fashioned as surveying the world from an elevated vantage point that is out of reach for those seen as closer to the ground, caught up in the corporealities and banalities of daily life. This practice of self fashioning is built upon acts of exclusion: historically, only white Western men were deemed capable of bodily transcendence. Women and non Western peoples have been constituted as incapable of transcending their bodies and therefore seen as incapable of rationality and knowledge production. Knowledge is framed as the exclusive property or expression of the geographer as objective observer rather than something that emerges from embodied interactions between geographers and other subjects or entities. These practices of self fashioning in geography ensure the perpetuation of Eurocentrism.

Claims about 'what' the world is like – and therefore who 'we' are – are increasingly understood as grounded in specific epistemologies. In geography, certain epistemologies are validated and given preference over others. D. P. Dixon and J. P. Jones III identify the epistemology of the grid as particularly important, for it is the procedure for segmenting a dynamic, complex, and interconnected social reality to capture, measure, and analyze it. Imposed upon the world as a way of knowing, this grid has become inseparable from the social order it is meant to describe and analyze. Reality has come to seem segmented; the ontological boundaries between, for example, mind and body, culture and nature, structure and agency, public and private appear natural to the social order, rather than the outcome of specific representational practices. Categories are treated as fixed and associated with specific characteristics. When privileged in geography, these epistemologies enable Eurocentric ontologies. When imposed upon other ways of knowing, they become tools of Eurocentric violence.

Undoing Eurocentrism

Undoing Eurocentrism in geography is no small task. The colonial order of things constitutes the present and therefore will surely inform the future. A wealth of geographical scholarship in the last decade has gone a long way in challenging Eurocentrism on empirical grounds. Rather than sustaining polarizing accounts that center on the West and the Rest, new research focuses on unearthing connections and co productions between people and places. Geographers increasingly shed light on the dark side of modernity and illustrate how the everyday discourses and practices of colonialism are integral to the emergence of modern sociospatial formations. In short, efforts to de colonize knowledge are now widespread and may soon be mainstream.

However, empirical efforts alone will not undo the prevalence of Eurocentrism in geography, for it is entrenched in our institutional structures and methodological frameworks. Although racial identity is not coterminous with a particular perspective, concerted efforts to transform the gender and ethnic composition of geographical institutions will create space for other experiences and interests. Following from this are strategies to de colonize pedagogical conventions and content. Do curricula in geography at undergraduate and graduate levels perpetuate Eurocentric histories, geographies, ontologies, and epistemologies? Do we encourage learning about other ways of knowing and being? Do we teach students to see how the privileges of Eurocentrism and whiteness work to sanction ignorance of the ontological and epistemological pluralism that make up our world? To what extent do we insist that researchers engage with a diverse range of scholarship? Do our pedagogical practices police students' writing styles in order to produce de corporealized and de subjectivized narratives wherein the geographer is everywhere and nowhere? Are diverse methodological approaches encouraged, so that instead of studies on, geographers might collaborate with others to co produce knowledge?

The subdiscipline wherein these questions are most actively engaged is feminist geography; women of color in particular push the discipline in new directions. Antiracist feminist research is built upon relational ontologies grounded in embodied epistemologies that move beyond the segmentation and hierarchy of the grid. In Katherine McKittrick's work, for instance, black women's geographies are treated as integral rather than marginal to regulatory systems of domination and classification. In addition, antiracist feminist geographers have spearheaded and institutionalized the notion of situated knowledge, which implies recognizing that knowledge comes from somewhere: geography, history, social position intersect to constitute knowledge that is necessarily partial, selective, and incomplete. As a methodological practice, situating knowledge enables the production of nongeneralizing knowledges that engage with and learn from other knowledges. To this end, antiracist feminist geographers also challenge the discipline to pursue alternative research methods that enable engagements with other ways of knowing. One such strategy is collaboration and a number of antiracist feminist geographers have produced innovative works that topple the geographers' position of authority and create spaces for exchange and surprise. Collaboration across social, political, and economic boundaries breaks up alliances between whites (whether conscious or not) that historically have worked to sustain white hegemony in academia.

Situating knowledges and collaborating are strategies for undoing the Eurocentrism symbolized by Father Kino, sitting on a horse, elevated and separated from his subjects, whose authority as a producer of knowledge is founded in his position as a colonizer. The notion of walking together, feet on the ground, offers a more humble, corporeal metaphor for geographical research and promises to enable rich and exciting engagements with the ontological and epistemological pluralism that constitutes our world.

See also: Feminism/Feminist Geography; Green Revolution; Orientalism; Postcolonialism/Postcolonial Geographies; Whiteness.

Further Reading

Abu Lughod, J. L. (1989). *Before European Hegemony: The World System A.D. 1250 1350*. New York: Oxford University Press.

Amin, S. (1989). *Eurocentrism (transl. Moore, R.)*. New York: Monthly Review Press.

Blaut, J. M. (1993). *The Colonizers Model of the World: Geographical Diffusionism and Eurocentric History*. New York: The Guilford Press.

Chakrabarty, D. (2000). *Provincializing Europe: Postcolonial Thought and Historical Difference*. Princeton, NJ: Princeton University Press.

Coronil, F. (1996). Beyond occidentalism: Toward nonimperial geohistorical categories. *Cultural Anthropology* 11, 51 87.

Dixon, D. P. and Jones, J. P. (1998). My dinner with Derrida, or spatial analysis and poststructuralism do lunch. *Environment and Planning A* 30, 247 260.

Dussel, E. (1995). Eurocentrism and modernity. In Beverley, J., Aronna, M. & Oviedo, J. (eds.) *The Postmodernism Debate in Latin America*, pp 65 76. Durham: Duke University Press.

Landers, E. (2000). *La Colonialidad del Saber: Eurocentrismo y Ciencias Sociales. Perspectivas Latinoamericanas*. Buenos Aires: CLACSO.

Mahtani, M. (2004). Mapping race and gender in the academy: The experiences of women of colour faculty and graduate students in geography departments in Britain, the US and Canada. *Journal of Geography in Higher Education* 28, 91 99.

McKittrick, K. (2006). *Demonic Grounds: Black Women and the Cartographies of Struggle*. Minneapolis: University of Minnesota Press.

Mignolo, W. (1995). *The Darker Side of the Renaissance: Literacy, Territoriality, and Colonization*. Ann Arbor, MI: University of Michigan Press.

Nagar, R. (2006). *Playing with Fire: Feminist Thought and Activism through Seven Lives in India*. Minneapolis, MN: University of Minnesota Press.

Pulido, L. (2002). Reflections on a white discipline. *The Professional Geographer* 54, 42 49.

Slater, D. (1992). On the borders of social theory learning from other regions. *Environment and Planning D: Society & Space* 10, 307 327.

Wolf, E. (1982). *Europe and the People without History*. Berkeley: University of California Press.

Relevant Website

http://www.lib.utexas.edu
Relaciones Geográficas Collection, Benson Latin American Collection, University of Texas, Austin.

Europe

M. Kuus, The University of British Columbia, Vancouver, BC, Canada

Glossary

The European Union The European Union (EU) is a political project and a form of legal organization. Growing out of four decades of political integration in Europe, the EU was established by the Treaty on European Union (The Maastricht Treaty) in 1992. It exists "to organize, in a manner demonstrating consistency and solidarity, relations between the Member States and between their peoples" (Article 1 of the Treaty on European Union). The EU currently has 27 member states. Although the various treaties that form the legal basis of EU's activities do not officially confer legal personality on the EU, the Union does have a single institutional structure and a *corpus* of law (*acquis communautaire*). Adopted by the author from the glossary section of the EU homepage.

Introduction

Europe is not simply a (sub)continent located between the Atlantic and the Urals. It is also a complex and contested idea and political project. The daily uses of the term 'Europe' are inextricably tied to a geographical imaginary in which Europe occupies a privileged position as a space of progress and enlightenment. Parallel with this valorization of Europe, as the term is increasingly used as shorthand for the European Union (EU), it has also become a codeword for faceless bureaucracy that undermines the continent's deeply rooted national identities. This view also involves a geographic imaginary, one in which the continent of Europe is, or should be, a patchwork of territorially and culturally distinct nations. Close analyses of the (re)demarcation of Europe's external borders and the (re)definition of political identities inside these borders are key themes in geographic scholarship on Europe.

The Borders of Europe

The concept of Europe as a territorial unit with a distinct political identity is relatively recent. Although the term 'Europe' entered the vocabulary of the continent's pol itical elite in the late Middle Ages, it denoted the realm of Christendom rather than a geopolitical or cultural entity. The modern conception of Europe as a territorial unit governed by the balance of power among key states and clearly set apart from non Europe emerged after the religious wars of the seventeenth century. It was, indeed, only in the eighteenth century, in the context of Euro pean exploration and colonialism, that secular theories about Europe's borders, identity, and historic mission became mainstream. The burgeoning inflows of infor mation about distant places, the increasing trade and travel within Europe, and the intellectual currents of the Enlightenment created a sense of a common identity among Europe's political elite.

From the beginning, the modern idea of Europe emerged as a territorial concept. In particular, Europe was defined in contrast to what was conceived as non Europe, especially the East. The political border of the emerging entity of Europe had no clear physical markers – although the cartographers of the Russian Czar Peter the Great placed it in the Ural Mountains in the 1700s. Defining Europe, therefore, involved making place out of space and imbuing that place with an idealized notion of European ness. In particular, it involved defining the East as fundamentally different from, and an external threat to, (Western) Europe. Even today, evoking Europe involves drawing explicit or implicit boundaries between the putative enlightened, peaceful, and progressive West – and Europe as its birthplace and a key center of gravity – and various relatively backward Easts.

The concept of Eastern Europe is a key example of that process. It was invented in the eighteenth century, when the Ottoman Empire declined in power and Western Europe needed new markers against which to demarcate itself. Eastern Europe was defined as being of Europe by geography, but as still being in the process of becoming European in a political and cultural sense. It emerged not as irredeemably alien, but as a halfway house between Europe and Asia: not simply backward, but a learner, an experiment, and a testing ground. A similar process of creating internal repositories of un European ness occurred with respect to the Balkans; that region, too, was invented, especially in the nineteenth century, as a part of European identity construction. To highlight these processes of boundary making is not to deny that the eastern and southern parts of Europe differ from the western part of the continent, and they have been markedly poorer than (the core of) Western Europe for centuries. It is rather to highlight a framework of meaning in which complex economic and political dif ferences are conceptualized in terms of proximity to, or distance from, the core or essence of European ness.

Europe thus emerges not as a place but as a political idea. There has never been one clearly defined unit of

Europe. We can rather think of its meaning in terms of a process of inscription in which places are included in an idealized European sphere by inscribing them with shades and degrees of European ness. Even during the Cold War, when the division of Europe into East and West was perhaps the most clear cut, Europe and European ness remained ambiguous and contested ideas. On the one hand, the Cold War geopolitical discourse tended to lump east central Europe together with the Soviet Union as the constituting 'other' of the West. Analyses of Central Europe frequently exoticized the region quite similarly to other 'non Western', and hence, 'backward' regions. At the same time, the Soviet Union and its satellite states were also viewed as capable of industrial, scientific, and technological development. They did not fully fit the image of the backward and fundamentally un European East. They were never fully outsiders; they were rather conceived as having not yet begun to fully participate in European politics.

After the end of the Cold War, questions about the borders of Europe intensified. The Cold War had imposed a political order that had muted questions about the essence and borders of European identity. These questions were re opened in the 1990s. The borders between East and West were not erased, however, but reconfigured. On the one hand, as the Soviet Bloc was no longer extant and east central Europe became the object of the Western led transition effort, the region was 'downgraded' on the scale of development and European ness. It became a space of Western tutelage, conceptualized as a variant of the Third World. On the other hand, the erstwhile Soviet satellite states came to be viewed as, at least potentially, a part of Europe. 'Return to Europe' became the central theme in the foreign and domestic policies of the postsocialist states, and these states' integration with the EU was widely viewed as a unification of Europe (**Table 1**). Europeanization as a move toward Europe became the key lens through which complex societal problems came to be understood.

The concept of Central Europe is illustrative of the post Cold War re drawing of Europe's political and cultural borders. That concept was invented in the 1980s by Polish, Hungarian, and Czechoslovakian dissident intellectuals. They defined Central Europe in terms of Europe, as 'the kidnapped West' – a region that is culturally in the heart of Europe, but politically incorporated into an alien and ultimately Asiatic Soviet sphere. They differentiated Central from Eastern Europe in terms of the presumed moral superiority of the fully European Central Europe over the less European Russia. Central Europe's chief reason for being was – and is – to be distinct from Eastern Europe. The concept undermines the Cold War era demarcation of Europe, but not the overarching framework of Europe and the East. It is a

Table 1 Brief chronology of the development of the European Union, 1952 2007

1952	Creation of the European Coal and Steel Community (ECSC) by Belgium, France, Federal Republic of Germany, Italy, Luxembourg, and the Netherlands (Treaty of Paris)
1957	Two treaties of Rome, creating the European Economic Community (EEC, later EC for European Communities) and the European Atomic Energy Community (EURATOM)
1973	Accession of Denmark, Ireland, and the UK to the EC
1981	Accession of Greece to the EC
1986	Accession of Portugal and Spain to the EC
1993	Foundation of the European Union (EU) as successor to the EC (Treaty of Maastricht, signed in 1992)
1995	Accession of Austria, Finland, and Sweden to the EU
1999	The Euro is officially launched (comes into common circulation in 2002)
2004	Accession of Estonia, Latvia, Lithuania, Poland, The Czech republic, the Slovak republic, Hungary, Slovenia, Malta, and Cyprus the EU
2005	European Union opens accession negotiations with Turkey and Croatia. The Former Yugoslav Republic of Macedonia is granted candidate status
2007	Accession of Bulgaria and Romania the EU

Source: Author compilation from the European Commission website.

mechanism of shifting Eastness further east. Places like Croatia, Turkey, and Macedonia, to name the current applicants, can increase their European ness and shift themselves closer to Europe if they follow EU recommendations. Thus, even though the EU and North American Treaty Organization (NATO) enlargement processes have undermined, perhaps even dissolved, the Cold War era division of Europe into two, they have simultaneously fueled a more complex division of the continent into the European core, and various eastern and southern peripheries that are inscribed with lesser degrees of European ness.

In the West, Europe's political and cultural border used to be relatively unimportant for decades following World War II, as Atlanticism provided the framework in which European identity was conventionally understood. However, the traditional Atlanticist conceptions of close political and cultural proximity between Western Europe and North America have come under pressures in recent years, most notably since the US led war in Iraq in 2003. Tensions between US foreign policy and European public opinion have fueled questions about whether Europe has developed, is developing, or should develop its own identity as a global political actor outside the framework of Atlanticism. Whatever the substantive differences between the policies of the United States and its European allies, these debates have shown that the US and Europe still tend to define themselves in relation to each other. In the West, too, the borders of European identity are central to political debate about Europe.

In the south, the Mediterranean still provides a seemingly unambiguous geopolitical and geo economic border. Yet, even there, questions about Europe's political and cultural borders are gaining prominence. Increased immigration, itself a part of Europe's colonialist and imperialist legacies, problematizes the traditional defi nition of political and cultural Europe. The EU's own external policies, most notably the European Neigh bourhood Policy, are altering the EU's interactions with its southern interior. It is certainly premature to speak of open and porous borders, as the EU is keeping its insti tutions tightly closed. However, EU's external policies in the Mediterranean increasingly deploy the concept of European norms and European neighborhood to effect democratization and influence political affairs in the re gion. This could propel a certain opening of the concepts of European identity, European values, and ultimately, Europe in a political and cultural sense.

The key point of these examples is that the political project of European integration is closely tied to the idea of Europe as a geographical, geopolitical, and cultural entity. The bounding of Europe in a political and cultural sense is not an abstract matter; it is a part of the very fabric of Europe's relations with the rest of the world. EU practices toward its eastern and southern exterior in creasingly frame differences not in terms of fixed pos itions of inside or outside the Union, but in terms of a continuum of European ness. They deploy the argument of 'all but membership', advocating the operation of EU norms without formal membership. Until the 2007 en largement, the EU's influence on the applicant states involved the carrot of eventual membership. It appears that no such direct promise is offered to today's outsiders, as the enlargement processes have slowed down since the 'big bang' expansion in 2004. As a consequence, the EU's external relations in its immediate neighborhood will be effected less through administrative mechanisms of ac cession negotiations and more through narratives of European identity and values. When political actors like Russia or Ukraine conform to the so called European norms, they are framed as moving to Europe; when they resist, they are framed as still distant from Europe and still in the process of learning European ness. The EU's neighbors are not framed as 'outside' the European sphere; they are rather cast as further away from the core of the European cultural area. Yet these neighbors are not promised an 'inside' position either. Rather, they are in corporated into a spatial logic that assigns them onto locations on a scale of European ness. The metaphor of learning European norms denotes not simply a har monization of technical standards; it also inscribes a political and cultural hierarchy of European ness. To expose and unravel this hierarchy, current scholarship on the concept of Europe increasingly foregrounds the presence of the East in Europe and vice versa. The scholarship is firmly situated in broader critiques of, and developments in, area studies and postcolonial theory. As a result, as John Pickles points out, the cartographies and geographies of 'Europe' are being destabilized and de colonized. Europe is becoming a more ambiguous and potentially a more open category.

The Nation-State in European Political Identity

Parallel with the re demarcation of Europe's external borders, there has also been a re scaling and re territor ialization of political identities inside these borders. This raises difficult questions about the role of Europe's tradi tionally strong nation states in European identity con struction. The growing influence of EU institutions is a catalyst here. True, the EU is a community of nation states, and intergovernmentalism is a key principle in EU insti tutions, especially in the spheres of foreign and security policy as well as justice and home affairs. The EU budget makes up less than 2% of the combined gross domestic product (GDP) of the member states. However, although the nation states maintain decisive role in a number of matters, EU regulations play an important role in 'how' the nation state regulates these issues. This role cannot easily be measured, but it is estimated that over 50% of national legislation of the member states is adopted in reference to EU laws and programs. Furthermore, Europeanization involves not only the technical transplantation of EU standards and procedures; it also transforms economic, political, and social processes. It not only regulates be havioral and policy outcomes, but also reshapes the pro cesses by which political subjects are produced. It is not imposed on political subjects, like nation states, from above; it is an integral part of decision making and identity formation at the state and local level.

Europeanization destabilizes the territorial imagina tion in which political space consists of a patchwork of culturally distinct communities – nations. It challenges the traditional framework of discussing European identities in terms of categories like 'the Finns', 'the French', or 'the Dutch'. They re territorialize identity upward toward the EU, downward toward cities and regions, and horizontally through cross border regions. There is a vibrant geo graphic scholarship on such initiatives. It shows a fascin ation with, and insurgence of, premodern and nonstate regionalisms, ranging from the large transnational net works around the Baltic Sea region, the North Sea region, and the Danube region, to the much smaller cross border regions like Galicja or Catalonia. This scholarship also foregrounds the deep entrenchment of the nation state in European political identities. The role of central state in public expenditures has not declined substantially, and the nation state remains the key mechanism of political legitimation. Nationalism is reproduced not only official state policies, but also, and more importantly, through everyday life. Cross border cooperation does not neces sarily undermine nationalism; it can also entrench national boundaries. Although the state centric view has been de bunked, geographic research cautions against uncritical celebrations of new 'neo medieval' spatialities with over lapping spheres of authority. Regional and local identities are not necessarily more 'natural' or more progressive than national ones.

Europe's deeply territorialized national and regional identities have an uneasy relationship with European integration. The claims that EU rules contradict and undermine local traditions have been used for decades by the left and the right alike. Right wing populist move ments, in particular, view the EU as a force of top down globalization and rootless cosmopolitanism, one that erodes the essential link between land and people in Europe. The EU emerges in such arguments as the ul timate 'other' of such Europe of nation states. Across Europe, the deepening of integration and the expansion of the Union have been accompanied by greater emphasis on national and state sovereignty. For example, UK, Denmark, and Sweden have stayed out of the Eurozone in marked part due to popular concerns over sovereignty. The relative weakening of state authority has therefore not necessarily increased identification with Europe; it has also fueled voter apathy and Euroskepticism. In the new member states, sovereignty has emerged as the rhetorical touchstone by which to express the dis illusionment with national governments as well as EU institutions. Populist politicians have used social issues like same sex marriage, abortion, and the role of religion in public life to mobilize sentiments of national excep tionalism and Euroskepticism. Ironically, the social issues listed above are in the jurisdiction of the nation state anyway; the rhetoric of sovereignty shows not just an attachment to the *status quo* but also a widespread ig norance of how the EU operates. EU institutions have taken notice of the absence of a European demos; at tempts to create a common European identity are among their policy priorities.

Immigration is one of the most prominent and sen sitive political issues in today's Europe. Throughout most of the post World War II era, most of Europe's nation states have been highly homogenous. Europe is certainly diverse in ethnic and linguistic terms, but the individual nation states are often highly homogenous (**Table 2**). In the last few decades, there has been a relatively rapid increase of immigrant population, and even greater public perception of such increase. Although the majority of immigrants in Europe are from other EU member states (and are, hence, nominally 'European'), migration has fueled tensions over the role of non nationals in the political and cultural life of the nation states. All EU members imposed restrictions on the free movement of labor from the new member states in 2004 (and again in 2006, in anticipation of the 2007 enlargement). The marked presence of Muslims in Europe has highlighted unresolved questions about the meanings and limits of multiculturalism and republicanism. States are increas ingly demanding affirmation of loyalty and belonging, and placing greater emphasis on the obligations of citi zenship. Immigration has also fueled the securitization of national identities and state institutions, that is, the framing of these identities as under threat, and therefore, in need of emergency assistance from the state. Immi gration and security concerns have thereby become powerful mechanisms of legitimizing state power. The underlying point here is that the re territorialization of political identity in Europe is not simply a trend in which the national scale loses and the local scale gains. Rather, multiple scales of identity construction, including the national, are deployed and reconfigured simultaneously.

Table 2 Foreign-born population in selected European countries

Country	*Foreign-born population as share of total population, 2005*
Austria	15.1
Belarus	12.2
Belgium	6.9
Bulgaria	1.3
Cyprus (Republic of)	13.9
Czech Republic	4.4
Denmark	7.2
Estonia	15.2
Finland	3.0
France	10.7
Germany	12.3
Greece	8.8
Hungary	3.1
Iceland	7.3
Ireland	14.1
Italy	4.3
Latvia	19.5
Lithuania	4.8
Luxembourg	37.4
Malta	2.7
Netherlands	10.1
Norway	7.4
Poland	1.8
Portugal	7.3
Romania	0.6
Russia	8.4
Slovakia	2.3
Slovenia	8.5
Spain	8.5
Sweden	12.4
Switzerland	22.9
Turkey	1.8
Ukraine	14.7
United Kingdom	9.1

Source: United Nations 'World Migrant Stock' Database, 2005.

Conclusion

Geographic research shows that Europe is not simply a descriptive term but a powerful normative concept – closely bound up with geographical and territorial definition – that should be interrogated in its own right. Since the end of the Cold War, the concept of Europe has moved center stage in political rhetoric in Europe, and yet, evoking that concept activates several, sometimes contradictory, geographical imaginations. Europeanist geography picks apart the sim plistic narrative of a linear movement toward a unified Europe, in harmony with national identities and with the various outsiders, which represents Europe's 'true' core identity. It shows that the bounding of Europe is best viewed in terms of a multi layered and contested process in which economic and cultural integration are sometimes on different pathways. The re inscription and re demarcation of Europe is not a practice that emanates from the center; it is also practiced on the margins, like the new member states. The normative space of EU influence, that is, the space in which the Union has a constitutive influence on social relations, is expanding, even if the Union's membership remains unchanged for years to come. This process is open ended; it could both open up the concept of Europe and entrench current spatial hierarchies. As a range of complex social and political processes are explicitly or implicitly conceptualized in terms of European ness, the borders of Europe will not dissolve but multiply.

See also: Communist and Post-Communist Geographies; Nationalism; Sovereignty; State.

Further Reading

Agnew, J. (2005). Bounding the European project. *Geopolitics* 10, 575 580.

Agnew, J. (2005). Territoriality and political identity in Europe. In Berezin, M. & Schain, M. (eds.) *Europe Without Borders: Remapping Territory, Citizenship, and Identity in a Transnational Age*, pp 219 242. New York: Columbia University Press.

Bialasiewicz, L. and Minca, C. (2005). Old Europe, New Europe: For a geopolitics of translation. *Area* 37, 365 372.

Dunn, E. (2005). Standards and person making in east central Europe. In Ong, A. & Collier, S. (eds.) *Global Assemblages: Technology, Politics, and Ethics as Anthropological Problems*, pp 173 193. Oxford, UK: Blackwell.

Heffernan, M. (1998). *The Meaning of Europe: Geography and Geopolitics*. London: Arnold.

Heffernan, M. (2007). *Geography and the Idea of Europe* (Hettner Lecture). Heidelberg, Germany: Springer.

Jones, A. (2006). Narrative based production of state spaces for international region building: Europeanization and the Mediterranean. *Annals of the Association of American Geographers* 96, 415 431.

Kaplan, D. and Hakli, J. (eds.) (2002). *Boundaries and Place: European Borderlands in Geographical Context*. Lanham, MD: Rowman and Littlefield.

Kofman, E. (2005). Citizenship, migration and the reassertion of national identity. *Citizenship Studies* 9, 453 467.

Kuus, M. (2004). Europe's eastern enlargement and the re inscription of otherness in east central Europe. *Progress in Human Geography* 28, 472 489.

Mamadouh, V. (2003). 11 September and popular geopolitics: A study of websites run for and by Dutch Moroccans. *Geopolitics* 8(3), 191 216.

McNeill, D. (2004). *Urban Change and the European Left: Tales from the New Barcelona*. London: Routledge.

Moisio, S. (2006). Competing geographies of sovereignty, regionality and globalisation: The politics of EU resistance in Finland 1991 1994. *Geopolitics* 11(3), 439 464.

Murphy, A. B. (1999). Rethinking the concept of European identity. In Herb, G. & Kaplan, D. (eds.) *Nested Identities: Nationalism, Territory, and Scale*, pp 53 73. Boulder, CO: Rowman & Littlefield.

Neumann, I. B. (1999). *The Uses of the Other: 'The East' in European Identity Formation*. Minneapolis, MN: University of Minnesota Press.

Paasi, A. (2001). Europe as a social process and discourse: Considerations of place, boundaries and identity. *European Urban and Regional Studies* 8, 7 28.

Painter, J. (2002). Multilevel citizenship, identity and regions in contemporary Europe. In Anderson, J. (ed.) *Transnational Democracy: Political Spaces and Border Crossings*, pp 93 110. London: Routledge.

Pickles, J. (2005). 'New cartographies' and the decolonization of European geographies. *Area* 37, 355 364.

Scott, J. W. (2002). A networked space of meaning? Spatial politics as geostrategies of European integration. *Space and Polity* 6, 147 167.

Shore, C. (1999). *Building Europe: The Cultural Politics of European Integration*. London: Routledge.

Sidaway, J. (2001). Rebuilding bridges: A critical geopolitics of Iberian transfrontier cooperation in a European context. *Environment and Planning D: Society and Space* 19, 743 778.

Stenning, A. (2005). Out there and in here: Studying Eastern Europe in the West. *Area* 37, 378 383.

Stolcke, V. (1995). Talking culture: New boundaries, new rhetorics of exclusion in Europe. *Current Anthropology* 63, 1 20.

Relevant Websites

http://www.coe.int
Council of Europe.

http://www.ccre.org
Council of European Municipalities and Regions.

http://europa.eu
European Union: In particular, see the Public Opinion Analysis section of the European Commission; Glossary section of the European Commission site.

http://www.osce.org
Organization for Security and Co operation in Europe.

Europe of Regions

J. W. Scott, University of Joensuu, Finland

Glossary

Cohesion Principle of promoting economic prosperity and social justice within the European Union (EU) in order to avoid socioeconomic fragmentation and greater regional disparities.

Federalism Form of political organization defined and guaranteed by national constitutions, in which the activities of government are divided between central and subnational governments in such a way that each level of government is responsible for specifically defined tasks on which it makes final decisions.

Multilevel Governance A political concept that interprets contemporary political relationships within the EU, Europe as consisting of overlapping authorities and competing competencies at the supranational, national, and subnational levels.

Regionalism The advocacy of a political system that divides national territories into partially autonomous regions and that involves a sense of loyalty to and an active representation of the interests of a particular region or of groups of regions.

Regionalization A process of dividing national territories into subunits for purposes of political decision making and/or public policy.

Regional Policy A policy that aims at decreasing socioeconomic differences within national territories (and the European Union) by providing incentives for investment and growth in specifically designated regions.

Subsidiarity A principle according to which each level of political decision-making authority is responsible for those public tasks that it is best suited to fulfill.

Structural Policy A policy that addresses the long-term improvement of growth and economic performance of states and regions through a variety of incentives, investments, and sectoral measures.

Introduction

The development of the European Union (EU) has been closely associated with and influenced by regional scale and the evolution of regional perspectives on various political, economic, and social issues. Indeed, the modernization of regional development policies has been one of the most visible successes of European construction. During the late 1980s and early 1990s, the idea of a Europe of Regions was a defining element of the European Community's grand political project of interstate integration. In a strict sense, the Europe of Regions suggested a redoubled focus on socioeconomic equity (cohesion) within an enlarging political community as well as a shift in decision making powers away from the nation state. Conflating notions of economic competitiveness, democratic governance, subsidiarity, and cultural identity, official policy discourse fed expectations that regions would be given a greater voice and more active role in formulating community policy.

The Maastricht Treaty of 1992 did indeed establish a basis for regional representation within the EU. However, many of the regionalist scenarios envisaged for post Maastricht Europe have not materialized; ultimately, nation states remain principal actors within the EU and, generally speaking, there is little evidence of a serious challenge to a nationally defined sense of identity and citizenship. Furthermore, the ability of regions to actively participate within systems of multilevel governance is as varied as the political functions, roles, and territorial dimensions of the regions themselves. While several regions within the EU (e.g., in Spain and Germany) have become important political actors and have been able to enhance their influence within Europe, many others are little more than consultative bodies or quasi nongovernmental development agencies. Measured against initial expectations, the creation of new institutions representing regional interests within the EU has been rather modest as well. Reflecting a loss of momentum in European regionalization processes, contemporary debate on the future of the EU rarely mentions a 'Europe of Regions' in explicit terms. At present in fact, questions of national sovereignty and identity and the definition of specific roles of individual member states within the EU appear to have pushed regional issues somewhat to the sidelines. Despite this, however, regional scale remains a powerful leitmotiv of governance and development within the EU. This is due to the sheer scope of governance challenges facing an enlarged and extremely heterogeneous EU. Both political demands from the subnational level as well as a perceived need for more effective regional development policies and coordination mechanisms necessitate some form of regional participation.

Europe of Regions as a Political and Sociospatial Project

Regions in Europe can be understood as subnational territories where ethnic, linguistic, cultural, and/or economic structural characteristics help create (actively or passively) a sense of common identity. Regionalism in Europe is thus in principle a political ideal that seeks to maximize the participation, influence, and development potential of regions within the wider context of interstate integration. This ideal, furthermore, is premised on the assumption that regions – because of the dense social, economic, and cultural relationships that provide them a sense of identity – are an essential locus of growth and democratic governance. Admittedly, this notion of 'the region' largely reflects a regionalist understanding that is historical in nature. Indeed, many regions within Europe have a powerful sense of local identity and distinctiveness that has emerged as a result of both experiences with and struggles over political autonomy. Ethnic, linguistic, and cultural traditions have also fostered a sense of regional uniqueness in many parts of Europe. Examples of such 'historically distinct' regions include Scotland and Wales in the United Kingdom. Scotland, for example, is in many respects a nonsovereign nation within the United Kingdom with its own parliament and social and religious institutions. Europe abounds with other examples of historically and culturally defined regions that claim (and to an extent have received) a greater voice in political matters such as Wallonia and Flanders in Belgium; Catalonia, Galicia, and Valencia in Spain; Alsace, Bretagne, and Occitania in France; and Southern Tyrol in Italy. Even the trinational Swiss–French–German Dreyeckland centered on the city of Basel has lobbied for a certain degree of recognition as a culturally cohesive and functionally interdependent region. Furthermore, in some areas regional separatist movements present a serious challenge to national authority. The Basque country which straddles parts of Northeastern Spain and Southwestern France (the País Basco and Pays Basque, respectively) is a case in point. Corsica's demands for autonomy from France are another. Another type of politically visible region in Europe is characterized by subnational levels of government in federal states. The German states (Länder), which share power with the national government, also possess a certain degree of regional identity although many of them were created as amalgamations of historical regions.

Originally, the notion of a Europe of Regions was informed by a set of demands formulated by advocates of subnational representation within European institutions. This set of demands included (1) recognition of the cultural diversity of regions, (2) creation of a regional parliamentary body with formally defined decision making powers and guarantees of participation in decision making processes, (3) the formal recognition of subsidiarity as a political principle (this demand was in fact enshrined in Article 3b of the Maastricht Treaty and later in both the Amsterdam and Nice Treaties), and, finally, (4) guarantees of the right of subnational governments (regions, provinces, etc.) to address grievances directly to the European Court of Justice. Europe of Regions thus describes a political project that aims both to broadly promote regional development and to strengthen the governance role of subnational units (regions) within the EU. Indeed, the EU Commission suggests that a more direct experience with European integration and greater understanding for the importance of promoting integration can only develop out of a sense of regional–local empowerment. It must be emphasized that Europe of Regions, both as a concept and a political project, has evolved in response to long standing tensions within the EU. Debates over the remits of EU institutions and the degree of federalization of decision making processes continue to mark divisions between member states. At the same time, however, regions with increasing or well established political bases and/or economic clout have clamored for greater self determination and have criticized what they see as centralizing tendencies within the EU.

Governance and Cohesion: Policy Roles of Regions within the EU

From an official EU standpoint, the achievement of cohesion and good governance are central goals of interstate integration. With the objective of economic and social cohesion, the EU aims at "reducing disparities between the levels of development of the various regions and the backwardness of the least favoured regions or islands, including rural areas." Cohesion policies, such as those defined in the EU's 2000 Lisbon Agenda, seek to promote growth, competitiveness, and employment while emphasizing environmental sustainability. With good governance, on the other hand, a responsive and democratic institutional architecture are understood to be prerequisites for integration. The 2001 white paper on European governance has defined five principles that underpin good governance: openness, participation, accountability, effectiveness, and coherence. With governance the EU aims at a process of community building based on common rules as well as on adherence to a comprehensive set of political and ethical values. Within this metadiscursive context regional scale is judged to be of central importance. It is here – at least theoretically – where the EU's multiple objectives of economic dynamism, efficiency, democracy, and empowerment translate into concrete development projects.

Indeed, for over 30 years regions have represented a central focus and addressee of EU policies. The European Regional Development Fund was established in 1975 in order to address industrial crises and underdevelopment within member states. During the first years of the Fund's existence, the European Commission acted as an administrator on behalf of national development interests, disbursing regional assistance according to fixed quotas.

However, progressive reforms of the ERDF in the 1980s, accompanied by the accession of Greece, Ireland, Portugal, and Spain to the Community, established a new regional development agenda with a much more differentiated ap proach. With the establishment in 1988 of the Structural Funds (in effect consolidating all regional and sectoral aid programs within one policy instrument) regional policy was essentially denationalized; the EU assumed a major role in defining specific target areas and regional problems to be addressed. Furthermore, many regions have been able to develop policy partnerships directly with Brussels, drawing up strategic plans and projects that subsequently have re ceived support from Brussels. Since 1988, numerous pro grams and initiatives have been launched with the express goal of creating new development perspectives for per ipheral areas, agricultural regions, regions in industrial de cline, areas undergoing rapid structural transformation, etc. In addition, European policy established a framework for cooperation and knowledge exchange between different regions in various areas of economic, social, and spatial development. Some of the – evocatively named – structural initiatives have included RESIDER (restructuring of steel making regions), LEADER (development strategies for rural regions), KONVER (conversion of military instal lations to civilian use), RENAVAL (assistance to regions characterized by large shipbuilding industries in crisis), and RECHAR (assistance to mining regions). With the current programming phase (2007–13) of the European Structural Funds, emphasis has been put on the notion of territorial and socioeconomic cohesion and a streamlining of com munity initiatives. Cohesion focuses on economic devel opment in poorer regions ('convergence'), the enhanced economic performance of more prosperous regions ('com petitiveness and employment'), and cross border net working between different types of regions within the EU ('territorial cooperation'). **Figure 1** provides an overview of the eligibility of European regions for convergence and competitiveness programs.

In addition to regional policies, spatial planning at the EU level also promotes a decidedly regional perspec tive within the larger post 1990 geopolitical context of European development. This has culminated in the elaboration of a European Spatial Development Per spective (ESDP). Although not a community level policy in the sense of agriculture or regional development, ESDP is a policy framework of an advisory nature agreed by the European Ministers of Spatial Planning in 1999 that enshrines sustainable economic development and socioeconomic cohesion. Central to ESDP is a focus on regional urban systems, urban–rural relationships, access to development opportunity structures, and a concern for a diverse natural and cultural heritage.

In terms of the policy objectives mentioned above, Europe of Regions suggests a long term strategy of cre ating flexible territorial contexts for governance and socioeconomic cohesion within the supranational frame work of the EU. By the same token, it is important to mention that since the mid 1990s there has been a dis cursive shift increasingly away from empowerment and cultural variety cum identity toward a more neoliberal interpretation of cohesion, namely, one that favors entrepreneurial spirit, the role of small and medium size enterprises (SMEs), regional innovation, and culture as an economic asset.

Regional Institutions and Organizations in Europe

One of the most important demands of European re gionalists is that of a greater regional voice with regard to various societal issues – both within member states and the EU. As a result, the political project of a 'regional Europe' has taken concrete institutional form. Presently, a variety of organizations dealing with different policy agendas operate at the European level, several of them including regions outside the EU 27. Among these bodies are the Association of European Border Regions (AEBR, established in 1971), the Conference of Peripheral Maritime Regions of Europe (CPMR, created in 1973), the Assembly of European Regions (AER, founded in 1985), and the Congress of Local and Regional Author ities of the Council of Europe (founded in 1994). The most authoritative regional institution to date is the Committee of the Regions (CoR), a political assembly made up of 317 representatives of local and regional au thorities. Established in 1994 as a result of the Maastricht Treaty (Article 198a), the CoR is a consultative body that facilitates subnational representation in the development of EU legislation. In addition, CoR attempts to provide a direct institutional link between local governments, citi zens, and the EU. As stipulated in the Treaty, the Euro pean Commission and Council are required to consult CoR on legislation with clear regional and local ramifi cations. Within the Maastricht framework five policy areas for local/regional consultation were identified (economic and social cohesion, trans European infrastructure net works, health, education, and culture). With the Amster dam Treaty of 1999 these policy areas were expanded to encompass employment policy, social policy, the en vironment, vocational training, and transport. Outside these areas, the Commission, Council, and European Parliament have the option to consult the CoR on issues judged to have important regional or local implications. The CoR can also draw up an opinion on its own ini tiative, which enables it to put issues on the EU agenda.

Another important level of institution building in the sense of a Europe of Regions are semiformalized networks that operate as lobbies, discussion forums, and platforms for joint projects. There are numerous examples of these networks including the EUROCITIES network of major

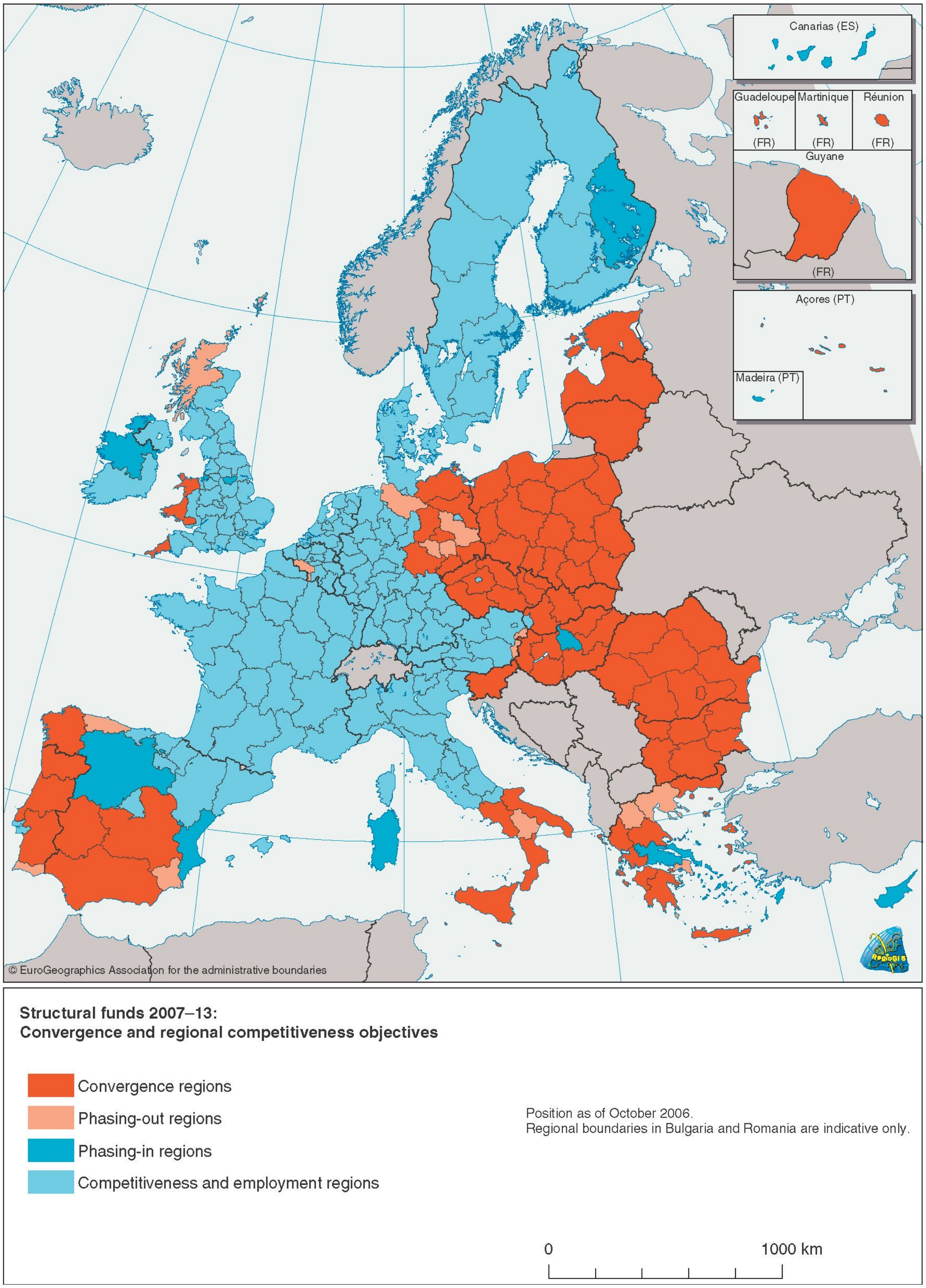

Figure 1 European regions and their eligibility for EU Structural Funds. Source: EU Commission, INFOREGIO.

European cities and METREX, The Network of European Metropolitan Regions and Areas. In addition, many subregional networks flourish. Prominent among these are those that focus on cooperation between Baltic Sea regions and regions (e.g., the Baltic Sea portal, the Baltic Islands network, and the Baltic Euroregions network).

Regionalization Processes and Patterns

In addition to regional policies and institution building, the Europe of Regions has also been characterized by a process of regionalization: that is, of a gradual creation of regional administrative units entrusted with various public policy remits. European regional policy began in the late 1980s to direct development aid to specific subnational and cross border contexts. This reform was reinforced by parallel processes of regionalization in which subunits of member states, such as France, Spain, Italy, and the UK, were ac corded a larger role in defining their futures and influ encing European policy. Regionalization has proceeded apace with the most recent EU enlargements in 2004 and 2007 as new member states with generally centralist tra ditions have experimented with the reorganization of ad ministrative space. In this latest regionalization phase, EU requirements have played an essential role; administrative decentralization has been made a prerequisite for the disbursement of structural development funds.

The process of region building within the EU has, of course, been uneven. This is partly due to the lack of a unitary framework that defines regions as such, for ex ample, in political and functional terms. While often based upon historical and cultural traditions, other regionalizing policies have been adopted by states as a means of ra tionalizing administration, managing internal change, and satisfying EU demands. As a result, the participation of regions within national and European policymaking pro cesses is dependent on their status within national con texts; Spanish regions, the *Comunidades Autónomas*, might have considerably more powers than *Régions* in France but lack the basic and constitutionally defined co decision making powers of German Länder. At the same time, there are few regions below the level of the German Länder that enjoy political legitimation. Devolution in the UK has been an issue of according Wales and Scotland greater autonomy but rather less one of empowering the English regions in administrative political terms. The situation is similarly heterogeneous in the new member states of the EU. In 1999, Poland opted for a dual system similar to that of France in which the regions, *województwo*, are represented by regionally elected officials and ap pointed agents of the central government. Hungary, on the other hand, has created seven new regions that are ba sically vehicles for regional development policies and managed largely by Quasi Nongovernmental Organisa tions (QUANGOs) financed by the central government.

Cross-Border Regions as a Case in Point

Regions on the internal and external borders of the EU represent a specific case of regionalization. Here, the EU has played a crucial role in supporting local and regional cross border cooperation as these are seen to be important aspects of interstate integration and a mechanism for deepening relations with non EU neighbors. It has also been argued that transnational regionalism contributes to community building through promoting processes of so cial learning and that political support of cooperative projects addressing specific regional problems will help create a citizen's Europe. INTERREG, now in its fourth phase (2007–13), has supported numerous transboundary and transnational cooperation projects between regions (**Figure 2** indicates the regions eligible for INTERREG IV funds). Financed out of the EU's structural funds, INTERREG has disbursed well over €10 billion making it the community's largest structural initiative. In addition, programs targeted for Central and Eastern Europe and the former Soviet Union, most prominently PHARE and TACIS, have provided supplemental funds for cross bor der projects in regions on the EU's external boundaries.

So called Euroregions were pioneered and developed as locally based cooperation initiatives in the Dutch–German border regions as early as the 1960s. Indeed, the Dutch–German EUREGIO, a Euroregion with its own local council and close ties to German and Dutch state agencies, has served as a model of sorts for the devel opment of border region associations both within and outside the EU. **Figure 3** depicts one of the German–Polish border regions where Euroregions have been established since 1993 in anticapation of Poland's EU membership. The Euroregion 'Pro Europa Viadrina' has its administrative center in the twin cities of Frankfurt (Oder) and Slubice. The main goal of Euroregions and similarly named organizations is to promote binational initiatives that address specific economic, environmental, social, and institutional problems affecting their re spective regions. These associations, many with cross border local government associations, represent an additional tier of regional administration and play a vital role in channeling European regional development aid into the border regions. According to the organizational principles of the original EUREGIO, community based associations designed along these lines have established secretariats to deal with day to day management. In order to structure their long term operations and, at the same time, satisfy new European requirements for re gional development assistance, Euroregions periodically define Transboundary Development Concepts (TDCs) that identify principal objectives of transboundary

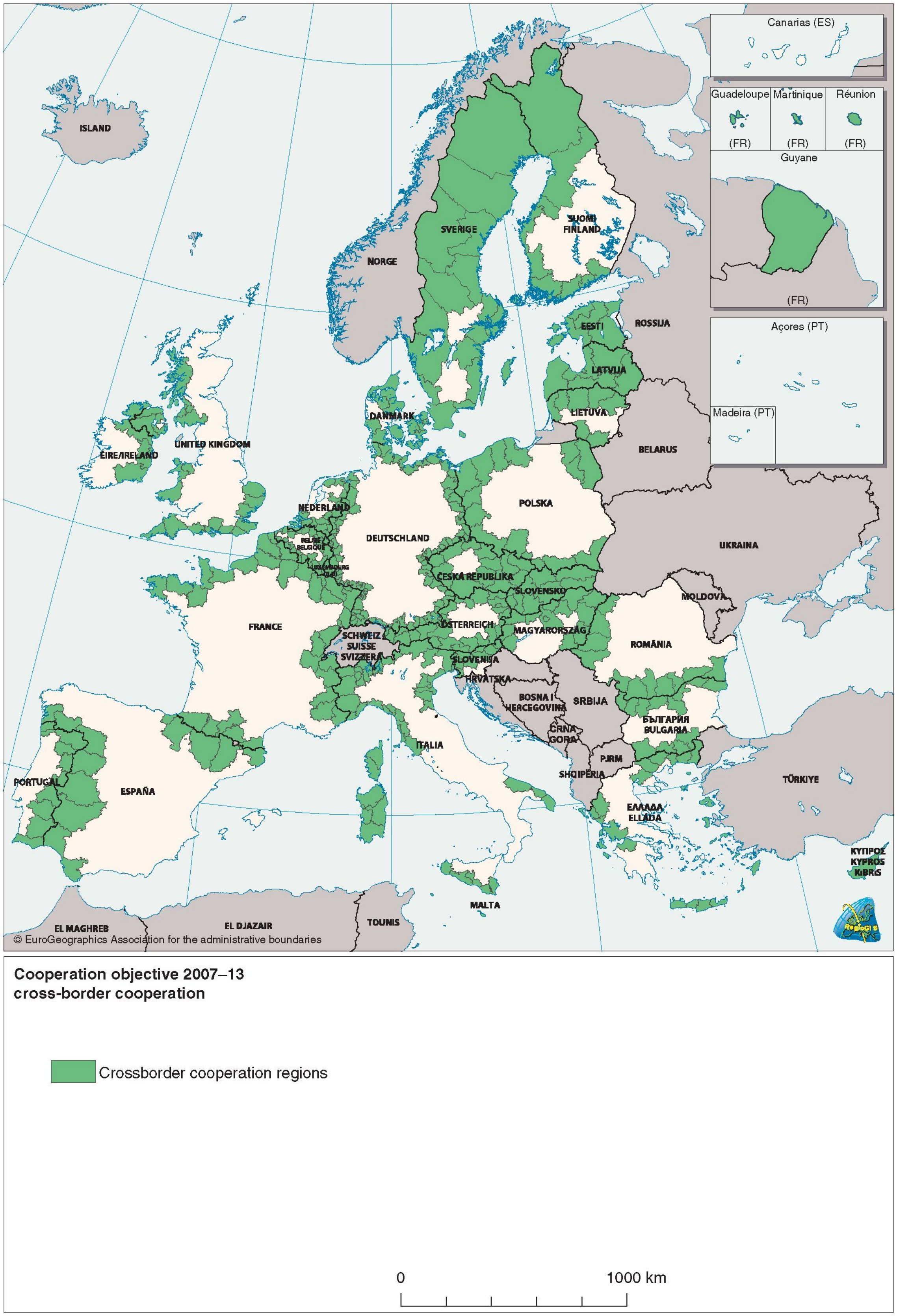

Figure 2 European regions eligible for INTERREG funds (cross-border cooperation). Source: EU Commission, INFOREGIO.

Figure 3 German Polish Border Region (Euroregion Pro Europa-Viadrina with “twin cities” on the border. Source: Brandenburg Ministry for Infrastructure and Spatial Planning).

cooperation and define possible courses of action. TDCs build the basis for concrete projects, proposals for which can then be submitted to the EU, national governments, or other funding sources for support.

Criticism of the ‘Europe of Regions’ Concept

European regionalism is not without its contradictions. Despite decades of regional support there remain great regional differences in the levels of development and socioeconomic well being. As the EU Commission’s Third and Fourth Reports on Social and Economic Cohesion document, despite increases in general welfare, the imbalances between Europe’s core areas and its vast peripheries remain and that depopulation of many rural zones continues unabated. Furthermore, regional disparities as well as cultural and political heterogeneity are certain to increase as a long term result of enlargement. Discourses of interregional competition and economic efficiency, often informed by neoliberal ideology, clash with principles of solidarity and spatially balanced economic development.

The normative assumptions that inform the EU’s regionalist perspectives have thus been subject to ample critical scrutiny. One issue in this context is the one sided influence of powerful and prosperous regions in Western Europe and their attempts to define the EU’s regionalist agenda. For example, the regions of Catalonia, Lombardy, Baden Württemberg, and Rhone Alpes which form the self styled ‘Four Motors’ of economic development in Western Europe, maintain a club like character, and have attempted to influence European discourses on governance and regional policy. As a result, such regional strengths have contributed to exacerbating core–periphery differences not only in economic but also in political terms through the emergence of powerful lobbies centered upon prosperous urban regions. For the weaker or rather artificially created regions, many of them that are little more than arbitrarily defined administrative units, it seems doubtful that they can develop

Figure 4 Campaign poster of the Coalició Valenciana denouncing illegal immigration, insecurity, Catalan influences, and a lack of attention to 'Valencian' interests.

an autonomous basis for economic growth and political power. The experiences of new member states, such as Hungary, Romania, and the Czech Republic, are indicative of this situation; regionalism in post socialist Hungary has not been the outcome of internal political, economic, or cultural developments. Instead, it is a political process initiated in response to integration into the EU and that has resulted in *ad hoc* regionalist arrangements. Under such circumstances, regional economic dynamism and competitiveness could, ironically, largely depend on levels of state support and state involvement in promoting specific initiatives. In addition, the cases of Greece, Portugal, and other countries indicate that regional boundaries can be shifted for politically expedient reasons, such as responding to changes in EU eligibility rules for development aid.

Finally, regionalism is not by definition a necessarily progressive concept. On the contrary, it can give sustenance to reactionary, particularistic, and exclusionary political forces that share little of the political idealism of Brussels. While perhaps not part of the mainstream, these regionalist groupings influence debates regarding identity and citizenship and the general openness of European society. The Coalició Valenciana (see **Figure 4**) is a good example of this tendency; in their 2007 local election campaign in the Autonomous Community of Valencia they have defined corruption, illegal immigration, and insecurity, and, curiously, the advance of 'Catalanism' as major threats to Valencian interests.

Conclusions: Where does the Europe of Regions Stand Today?

Measured against initial expectations, the concept of a Europe of Regions has provided rather mixed results. On the one hand, regions have become a major focus of EU structural policies and billions of Euros have been allocated for development projects in a wide variety of regional contexts. On the other hand, power sharing and endogenous economic dynamism appear reserved for the most prosperous and politically well organized regions. In addition to this, progressive enlargements of the EU, and the historic 2004 accession of 15 new member states in particular, have added new layers of complexity to the regional map of Europe. Continued fiscal retrenchment will increase divergence between prosperous and poor regions. Furthermore, the extent to which decentralization might provide greater powers to the new regions of Central and Eastern Europe is as yet unclear. What economic, social, cultural, and political roles might then regions realistically fulfill in terms of the construction of a European political community? There is no clear answer to this. A Europe of Regions can be interpreted as an attempt to reconcile seemingly antagonistic tendencies of centralization and local empowerment and, at the same time, promote the idea of an integrated European social, economic, political, and cultural space. However, it should be emphasized that there is no single regional idea or uniform concept with which to frame regions as politically meaningful actors within the EU. Europe of Regions can therefore be characterized as an uneven political project, as a long term debate over the role of local societies and changing perceptions of the governance role of the state; it is indicative of a negotiated and by no means always harmonious rescaling and reconfiguring of European governance upward toward international organizations, downward to cities and regions, and horizontally across nations, regions, and communities.

See also: Europe; New Regionalism; Region; Regional Development, Endogenous; Regional Integration; Regionalism.

Further Reading

Allen, J., Massey, D. and Cochrane, A. (1998). *Rethinking the Region*. London: Routledge.

Anderson, J. (2001). The rise of regions and regionalism in Western Europe. In Guibernau, M. (ed.) *Governing European Diversity*, pp 35 64. London: Sage.

Ash, A. (1999). An institutionalist perspective on regional economic development. *International Journal of Urban and Regional Research* 23, 365 378.
Ash, A. and Tomaney, J. (1995). The regional dilemma in a neo liberal Europe. *European Urban and Regional Studies* 2, 171 188.
Brenner, N. (2000). Building 'Euro Regions'. Locational politics and the political geography of neoliberalism in post unification Germany. *European Urban and Regional Studies* 7, 319 345.
European Communities (2004). *A New Partnership for Cohesion, Convergence Competitiveness, Cooperation. Third Report on Economic and Social Cohesion*. Luxembourg: Office for Official Publications of the European Communities.
European Communities (2007). *Growing Regions, Growing Europe. Fourth Report on Economic and Social Cohesion*. Luxembourg: Office for Official Publications of the European Communities.
Hadjimichalis, C. and Hudson, R. (2007). Rethinking local and regional development: Implications for radical political practice in Europe. *European Urban and Regional Studies* 14, 99 113.
Hueglin, T. (1986). Regionalism in Western Europe: Conceptual problems of a new political perspective. *Comparative Politics* 18, 439 458.
Jones, B. and Keating, M. (eds.) (1995). *The European Union and the Regions*. Oxford: Clarendon Press.
Keating, M. (2001). Rethinking the region. Culture, institutions and economic development in Catalonia and Galicia. *European Urban and Regional Studies* 8(3), 217 234.
Keating, M. and Hooghe, L. (2005). By passing the nation state? Regions and the EU policy process. In Richardson, J. J. (ed.) *European Union. Power and Policy Making* (3rd edn.), pp 239 255. London: Routledge.
Loughlin, J. (ed.) (2001). *Subnational Democracy in the European Union: Challenges and Opportunities*. Oxford: Oxford University Press.
Painter, J. (2002). Multi level citizenship, identity and regions in contemporary Europe. In Anderson, J. (ed.) *Transnational Democracy: Political Spaces and Border Crossings*, pp 93 110. London: Routledge.
Pálné Kovács, I. (2008). Regionalisation in Hungary: Options and scenarios on the 'road to Europe'. In Scott, J. (ed.) *New Regionalism in Socio Political Context*: *Space Making and Systemic Transformation in Central Europe and Latin America.* Aldershot: Ashgate (in press).
Sagan, I. and Lee, R. (2006). Spatialities of regional transformation in Central Europe and the administrative spaces of the EU. In Sagan, I. & Halkier, H. (eds.) *Regionalism Contested. Institution, Society and Governance*, pp 163 176. Aldershot: Ashgate.
Scott, J. (ed.) (2006). *EU Enlargement, Region Building and Shifting Borders of Inclusion and Exclusion*. Aldershot: Ashgate.

Relevant Websites

www.a e r.org
Assembly of European regions.
www.aebr.net
Association of European border regions.
www.ben.org
Baltic Euroregion network.
www.balticsea.net
Baltic Sea Region Portal.
www.cpmr.org
Conference of Peripheral Maritime Regions.
http://www.coe.int
Congress of Local and Regional Authorities.
http://www.cor.europa.eu
Committee of the Regions.
http://ec.europa.eu
EU Cohesion Reports and spatial planning documents website.
http://www.4motors.org
Four Motors.
http://europa.eu
Interactive maps of EU and regions in member states.
http://forum.europa.eu.int
Multimedia link of Eurostat Portrait of the Regions.
http://ec.europa.eu
Multimedia link of the European Union DR REGIO.

Evans, E. E.

M. C. Keane, St Mary's University College, Belfast, UK

Evans, Estyn E (1905–1989)

Emyr Estyn Evans was the most influential geographer working in Ireland in the twentieth century and, although the core of his work was the interpretation of the people and landscapes of Ireland, he was in no sense parochial. A pioneering researcher and charismatic teacher, his ideas continue to have an international reach.

He was born on 29 May 1905 to the Reverend George and Elizabeth Evans in Shropshire, England, where his father ministered to a Welsh speaking Presbyterian congregation. The family soon returned to rural Wales and many of the ideas which later informed his schol arship may be traced to these borderland beginnings and to his experiences of the cross currents between non conformist and Anglican beliefs and between Welsh and English cultures. His familiarity with the countryside and with country people convinced him that they are the principal bearers of heritage and they, thus, became the focus of his life's work.

The years 1922–25 at the University College of Wales, Aberystwyth, where Evans studied geography and an thropology under Professor H. J. Fleure, were formative. Although Fleure's background was in natural history, he promoted a humanistic trilogy of geography, anthropology, and history which reflect an integrated worldview and excluded contemporary ideas of racial purity. Evans later observed that Fleure had taught him to study regional social evolution as a continuous interaction between people and their environment and, through field work in Wales and Wiltshire, had initiated his lifelong commitment to prehistoric archeology. For his MA (1931) thesis he ex plored the late Bronze Age in Europe, and in 1939 he was awarded a DSc for his accumulated scholarship.

Earlier, in his outstanding BA dissertation on the Pyrenees, which prefigured his belief in the processes of cultural continuity and synthesis, Evans showed his fa miliarity with the ideas of the French geographers, es pecially Paul Vidal de la Blache. The thinking of the geohistorians of the *Annales* school also resonated with him, particularly their analyses of *mentalités*, Marc Bloch's insistence that land, life, and history were in separable, and Fernand Braudel's genius for synthesis. An early reference to mental outlook in his portrait of *France* later resurfaces as regional personality in his ac claimed *Mourne Country*. For Evans, the interconnection between habitat, heritage, and history, what he called human ecology, was the dynamic which gave each place its unique personality.

Evans' core work concentrated on elucidating the ecology of the cultures of Ireland, and from this concern grew a mutual and lasting bond between himself and Carl Sauer. His preferred method was field study, weighing up what he could discover on and in the landscape with what he could glean from talking to the people who inhabited it. *Irish Heritage* and *Irish Folk Ways* established his dis tinctive voice and gained him a wide and, ultimately, international readership. His archeological excavations up until 1939 demonstrated the reality of the Irish Neolithic. He was convinced that Ireland's island status and geo graphical isolation on the Atlantic fringe had ensured the continuity of artifacts and institutions from prehistory. His insistence that the intermingling of newer arrivals with settlers from the prehistoric past had been continuous rather than episodic was at variance with a nationalist orthodoxy, particularly that which argued that a Celtic invasion had created a discontinuity with the prehistoric. More recent excavations and improved scientific dating techniques have shifted the balance of evidence towards Evans' continuity thesis, and a more generous view of the soundness of his scholarship has emerged.

Evans had a pluralistic vision of culture in which na tional identities exhibit regional variations. These are cre ated by the adaptation and absorption of new arrivals and modulated by habitat, heritage, and history. He envisaged the Irish province of Ulster (with its individual personality) as one strong regional variant: Ireland in microcosm with its fragmented, regionalized subcultures. His explanation, corroborated by other studies, lay in the isolating effect of the border landscape, meaning that it remained the most Irish of Irish regions until the seventeenth century plan tation. However, in a society of contested identities and spaces, such an explanation has been read as scaffolding Ulster nationalism, in spite of Evans' self imposed avoid ance of the controversies of politics and religion.

In 1928, Evans founded the first geography department in Ireland at the Queen's University of Belfast. He struggled against resistance to the emerging discipline and the lack of resources afforded to it until an honors school was approved in 1947, two years after he had been appointed professor. Apart from sabbaticals in the US, Queen's was to be his base for 40 years, after which he spent 2 years as the first director of the Queen's Institute of Irish Studies, finally retiring in 1970. Evans was a member of the Royal Irish Academy, president of the geography (1957) and anthropology (1959) sections of the British Association for the Advancement of Science and president of the Institute of British Geographers (1970).

He was awarded six honorary degrees, the Victoria medal of the Royal Geographical Society (1973), a merit award from the Association of American Geographers (1979), and a CBE (1970). He produced seven major books, an array of academic papers, archeological surveys, excavation reports, and more ephemeral public talks, field excursions, and media contributions. Some of these were published posthumously in *Ireland and the Atlantic Heritage*. His material legacy, the Ulster Folk and Transport Museum (1958), records for posterity shared rural roots of a divided society in diverse parts of a complex region.

Estyn Evans was an influential figure in the development of cultural geography as a coherent field of enquiry in the twentieth century. Although his concerns were local, they were part of a universal process and the human ecological approach he evolved to illuminate regional personality was more than just the synthesis of habitat, heritage, and history; it extended to the psychological and spiritual realms. He was no stranger in his latter years, then, to geography's cultural 'turn'.

See also: Anthropology and Human Geography; Cultural Geography; Geohistory; Heritage and Culture; Heritage and Identity; Historical Geography; Land Change Science; Landscape; National Schools of Geography; Nationalism, Historical Geography of; Place; Possibilism; Regional Geography I; Sauer, C.; Sense of Place; Vidal de la Blache, P.

Further Reading

Campbell, J. A. (1996). The writings of E. Estyn Evans 1926 88. In Evans, E. E. (ed.) *Ireland and the Atlantic Heritage: Selected Writings*, pp 261 268. Dublin: Lilliput Press.

Evans, E. E. (1937). *France: A Geographical Introduction*. London: Christophers.

Evans, E. E. (1942). *Irish Heritage: The Landscape, the People and their Work*. Dundalk: Dundalgan Press.

Evans, E. E. (1951). *Mourne Country: Landscape and Life in South Down*. Dundalk: Dundalgan Press.

Evans, E. E. (1957). *Irish Folk Ways, London*. London: Routledge and Kegan Paul.

Evans, E. E. (1966). *Prehistoric and Early Christian Ireland: A Guide*. London: Batsford.

Evans, E. E. (1968). *The Irishness of the Irish*. Belfast: The Irish Association for Cultural, Economic and Social Relations.

Evans, E. E. (1981). *The Personality of Ireland: Habitat, Heritage and History* (2nd edn.). Belfast: Blackstaff Press.

Evans, E. E. (1984). *Ulster: The Common Ground*. Mullingar: Lilliput Press.

Evans, G. (1996). Estyn: A biogeographical memoir. In Evans, E. E. (ed.) *Ireland and the Atlantic Heritage: Selected Writings*, pp 261 268. Dublin: Lilliput Press.

Evans, E. E. (1996). *Ireland and the Atlantic Heritage: Selected Writings*. Dublin: Lilliput Press.

Fleure, H. J. (1971). Emyr Estyn Evans: A personal note. In Buchanan, R. H., Jones, E. & McCourt, D. (eds.) *Man and His Habitat: Essays Presented to Emyr Estyn Evans*, pp 1 7. London: Routledge and Kegan Paul.

Graham, B. J. (1994). The search for the common ground: Estyn Evan's Ireland. *Transactions, Institute of British Geographers, New Series* 19, 183 201.

Relevant Websites

http://www.uftm.org.uk
This is the website of the Ulster Folk and Transport Museum which highlights a memorial to the work of E Estyn Evans.

http://www.ulsterhistory.co.uk
Ulster History Circle: this website provides a picture of Estyn Evans and another of the blue plaque which identifies his home in Belfast as that of Estyn Evans, the geographer.

Evolutionary Algorithms

N. Xiao, The Ohio State University, Columbus, OH, USA

Introduction

Efficiently solving optimization problems has been a great challenge to researchers from a wide range of disciplines, including computation science, mathematics, operations research, and geography. A typical example of such optimization problems is the traveling salesman problem that requires searching for a path such that a salesman can visit each of a set of cities exactly once and return to the starting city in the least cost (measured by the overall time or distance). The difficulty of solving this type of problem is that the time required to find an optimal solution may become exceedingly prohibitive for problems with large sizes (measured by the number of input). It has been proved, for example, that the computation time used to solve the traveling salesman problem grows exponentially with the number of cities, which makes it one of the most difficult optimization problems to solve.

For many optimization applications, though a global optimal solution is useful, finding a high quality solution (i.e., optimal or near optimal) in a reasonable time frame may become more important in a practical sense. Based on this belief, researchers have developed a variety of heuristic approaches, which can be used to find good solutions efficiently though the optimality of solutions found cannot be guaranteed. A particular type of heuristic approach is called metaheuristics, which provide general frameworks to address a variety of optimization problems. Typically named after their metaphors or origin, metaheuristic approaches include artificial neural networks, evolutionary algorithms (EAs), Lagrangian relaxation, simulated annealing, and tabu search.

Different from other metaheuristic approaches, EAs are based on the maintenance of a population of solutions to the problem being solved. EAs represent a problem solving paradigm in artificial intelligence with an aim to develop computer programs that can be used to search for solutions with limited human input. Since the 1960s, researchers in different countries started developing a family of solution approaches that share a set of characteristics. The work of Rechenberg and Schwefel has led to the emergence of evolution strategies (ESs) that can be used to solve real valued parameter optimization problems. Fogel and his collaborators developed evolutionary programming (EP) that is suitable to search for finite state machines in artificial intelligence domain. Holland developed genetic algorithms (GAs) using binary strings as a generic representation of solutions. Koza's genetic programming (GP) evolves computer programs as parsing trees. These procedures have been found robust in solving optimization problems in various fields such as engineering design, operations research, and spatial optimization. Three major academic journals have been dedicated to this area, including *Evolutionary Computation*, *IEEE Transactions on Evolutionary Computation*, and *Genetic Programming and Evolvable Machines*.

Design of Evolutionary Algorithms

To develop an EA, regardless of its flavor, one needs to include the following basic components:

1. a population of individuals, each of which represents a solution to the problem being solved;
2. a fitness function that can be used to evaluate each individual; and
3. a reproduction process that selects (parent) solutions from a current population and creates (offspring) individuals that form the next generation of the population.

With the above components, an EA can be generally outlined as follows:

1. randomly generate a population of individuals
2. repeat
   (a) compute the fitness of each individual in the current population
   (b) create a new population of individuals by selecting parent individuals from the current population and recombining them to form new individuals
   (c) modify individuals in the new population using a mutation operation
   (d) replace the current population by the new population
3. until termination criteria are met.

The general EA procedure can be illustrated using an example that maximizes function $f(x_1, x_2, x_3) = x_1^2 - x_2 - x_3$, where x_1, x_2, and x_3 are integers in the range from 0 to 7. To design an EA for this problem, a binary representation can be used so that each of the three variables is represented using three bits and the length of an individual solution is therefore 9. Individual 011001010, for example, represents a solution of (3, 1, 2) that yields an objective function value of 6. In this example, the objective function is directly used as the fitness function. Suppose a population contains four individuals. A hypothetical initial population of four individuals, along with their corresponding decoded variables and fitness function values, are shown in **Figure 1a**. **Figure 1b** shows a scenario where parent solutions are selected from the initial population and are used to generate four new individuals. Note that the initial individual with the highest fitness value is selected twice, reflecting the concept of 'survival of the fittest' – better individuals have a higher chance to be selected. **Figure 1c** shows the impact of a mutation operation. It can be noted that the best individual in the resulting offspring has a higher fitness value.

A commonly used termination criterion is the total number of generations that can be determined by prior trials of the algorithm. Another useful termination condition is based on the improvement: an EA stops when it cannot achieve significant improvement. The mutation process is often conducted with a relatively low probability, meaning only a small portion of the population is altered.

It should be noted that most of the terms (e.g., population, mutation, and selection) used in an EA have their biological roots. An individual is often called a chromosome; for each element of an individual, a possible value is called an allele; the recombination operation is called crossover in many contexts, especially in GAs.

Representation and Evolutionary Operations

The binary representation (**Figure 1**) has its limitations when real numbers or other types of data must be used. Though this can be addressed by specifying the precision of each variable and then transforming them into integers, a more straightforward (and perhaps more natural) way is to directly use real numbers. For the example problem in **Figure 1**, if the integer requirement is relaxed, a string of three real numbers (x_1, x_2, x_3) can be used to represent a solution. In this case, the one point recombination method (see **Figure 1b**) will become naïve because merely switching the elements between two parent solutions will not provide enough search power. New recombination methods can be derived using a general form: $x_i' = f(x_{i1}, x_{i2})$, where x_i', x_{i1}, and x_{i2} are the i th element in the new and two parent individuals, respectively. A commonly used form is $x_i' = (1 - a_i)\, x_{i1} + a_i\, x_{i2}$, where a_i is a random number between 0 and 1. A number of mutation operations can be designed for a real valued representation. A uniform random form can

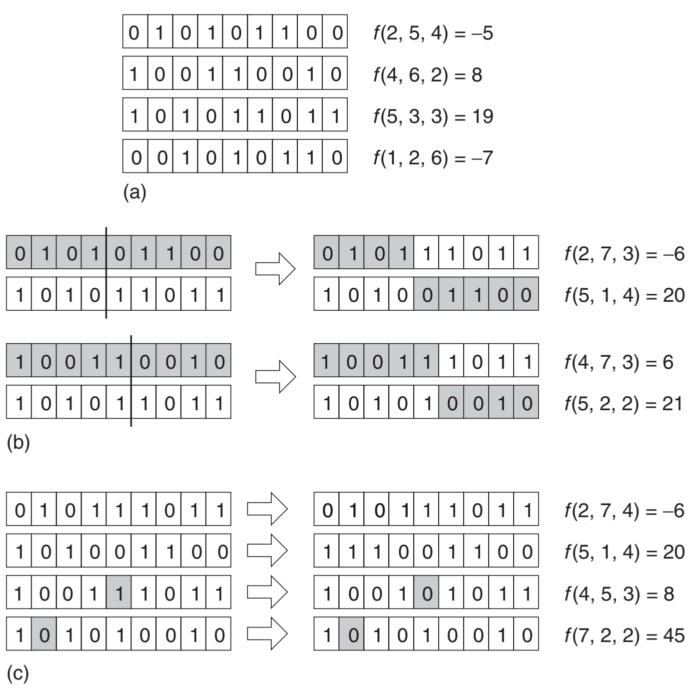

Figure 1 An example of EA. (a) An initial population of four individual solutions using a binary representation. For each binary string, every three bits starting from the left represent an integer ranging from 0 to 7. The fitness value of each individual is also shown. (b) A recombination operation called single point crossover for two pair of solutions selected from (a). The recombination process works by switching the bits after a randomly determined position. The shaded cells are used to illustrate the effect of the operation. (c) A mutation operation applied to the four new individuals in (b); note that only two bits (shaded) are undergone the mutation process.

be used to yield a random number uniformly distributed between the lower and upper bounds for an element of the solution that undergoes a mutation process. A normally distributed mutation can be designed as $x_i' = x_i + N(0, \sigma_i)$, where σ_i is a user specified standard deviation for the i th element.

Fitness Evaluation and Selection

Given the representation of a solution, an evaluation function must be defined to compute the fitness of each individual. Though this function is not necessarily the objective function of the original problem, it is normally derived from the objective function so that a solution that is closer to the optimum is assigned a higher fitness value and a solution further away from the optimum has a lower fitness value. In some circumstances, one can use the rank of each individual as a fitness value.

The selection operation is inspired by the notion of 'the survival of the fittest' of the Darwinian evolutionary theory. Two types of selection procedure are commonly used in EAs. The first type is the proportional selection in which the probability that the i th individual is selected can be calculated as $p_i = f_i / \sum_{j=1}^{N} f_j$, where f_i is the fitness function value of the i th individual and N is the number of individuals in a population. This procedure is also called the roulette selection because it is similar to the mechanism of a roulette wheel. The second type is called tournament selection in which a fixed number (defined *a priori*) of individuals are randomly picked from the population and the one with the highest fitness value is selected.

In general, a selection operation is directly related to the reproduction process by determining which individuals in the current generation will survive. In GAs, the population size is normally maintained constant and the new generation of individuals is used to replace the current generation. For example, the four individuals in **Figure 1c** will become the individuals in the next generation. In a more general context (especially in evolutionary strategies and evolution programming), however, the following two types of selection and reproduction methods can be defined.

1. $(\mu + \lambda)$ selection. Here, the population size is μ, and λ new individuals are created using recombination and mutation operations. Then, the best μ individuals are selected from $\mu + \lambda$ (parents and offspring) individuals to form the next generation of the population.
2. (μ, λ) selection. With μ individuals (old generation), a total of λ ($\lambda > \mu$) individuals will be created. Then the best μ individuals are selected from the λ individuals to form the new generation.

Constraint Handling

For many real world optimization problems, the feasibility of their solutions is often defined by a set of constraints. The traveling salesman problem, for example, requires that a tour should not repeat any city that has already been visited and that the tour should include all cities.

In EAs, constraints can be handled in three different ways. The first method is based on a penalty function that decreases the fitness function value of individuals that violate constraints and therefore decrease their chances of survival. The second method is to use algorithms to force feasibility of every individual in the population; these algorithms normally repair infeasible solutions during the recombination and mutation processes to make them feasible.

The third type of method is based on constraint preserving evolutionary operations that will not create infeasible solutions; these approaches are often based on a particular encoding/decoding strategy such that infeasible will not occur during evolutionary processes. A typical example is a particular EA designed for the traveling salesman problem. Let n be the length of a sorted list of cities. When $n = 5$, the cities can be listed as $C = (1\ 2\ 3\ 4\ 5)$. A chromosome is a string of n integers, where the value of the i th element ranges from 1 to $n - i + 1$. For a chromosome of (2 3 2 1 1), the first element is 2, meaning the first city to visit is the number 2 element in C. Now, remove the number 2 element from list C and look at the next number in the chromosome. This number is 3, which refers to the third number in the remaining list of C, and therefore the second city to visit is 4. This process continues until the tour is constructed as 2 4 3 1 5. This encoding/decoding strategy will guarantee the feasibility of solutions created by a traditional cross point recombination and random mutation operations.

Exploitation, Exploration, and Diversification

The impact of selection techniques on the overall convergence of an EA is measured by selection pressure, the preference given to the best solutions in the current population. A low selection pressure may make an EA similar to a random search where good solutions found are treated equally as low quality solutions. A high selection pressure, on the other hand, often causes the best solutions found to be trapped into some local optima because a few good solutions at the early stage of the evolutionary process may quickly take over the entire population.

To address this dilemma, one must consider the balance between the exploitation of information obtained so far and the exploration of new information that is yet to discover. A selection operation is exploitative because it tends to choose good solutions and eliminate undesirable ones. Recombination and, especially, mutation operations are explorative because they are designed to introduce new individual solutions to the population.

A key concept for the design of EAs is diversity, measured by the overall difference among all individuals in a population. A diverse population will reduce the chance of an EA to be trapped in local optima and will increase the chance of exploring new information. Though diversity can be achieved by simply increasing the use of mutation, doing so may not be able to preserve a number of good solutions in the population. A common way of maintaining diversity is through a sharing function:

$$\mathrm{sh}(d) = \begin{cases} 1 \quad (d/\sigma_{\mathrm{share}})^{\alpha} & \text{if } d < \sigma_{\mathrm{share}} \\ 0 & \text{otherwise} \end{cases}$$

where d is a distance measure between any two solutions, σ_{share} is a user specified distance threshold, and α is a user specified coefficient (often set to 0.5, 1, or 2). Using this function, a niche count can be computed for each individual solution: $nc_i = \sum_j \mathrm{sh}(d_{ij})$, where d_{ij} is the distance between individuals i and j, and nc_i is the niche count of the i th individual in a population. For an individual, a high niche count means there are many similar solutions (defined by d) and the solution is overrepresented. Therefore, by calculating a shared fitness as $f_i' = f_i/nc_i$, where f_i is the original fitness of the i th individual, it is possible to give underrepresented solutions (with low niche count) a high chance to survive.

A drawback of using the sharing function is the possibility of losing the best solution found. To address this issue, a mechanism called elitism is often employed so that the best solution found so far is always maintained in the population throughout all generations.

Schema Theorem

To explain why GAs (and EAs in general) work, Holland defined a schema as a similar template that describes a subset of strings with similarity at certain string positions. He introduced a 'don't care' symbol (*) to represent a string element that can be either 0 or 1. For example, the schema 11**1 is a template of four possible strings: 11111, 11011, 11101, and 11001. The order of a schema is defined by the number of its fixed positions (0's and 1's), and the defining length of a schema is the distance between the first and last fixed positions. According to these definitions, 11**1 has an order of 3 and a defining length of 4. In a GA, if a schema can yield a high fitness during an iteration, it will have a greater probability of being selected into the next generation. That is, the instances of this schema will increase in the next generation. This observation leads to the schema theorem: short, low order, above average schemata receive exponentially increasing trials in subsequent generations. Based on this theorem, the building block hypothesis can be derived: a GA seeks near optimal performance through the juxtaposition of short, low order, high performance schemata, called building blocks.

Multiobjective Evolutionary Algorithms

Many real world optimization problems involve multiple, often conflicting objectives. Political redistricting, for example, requires subdivision plans satisfy population equality and compactness, and possibly other objectives. Different from a single objective optimization problem, in which an optimal solution can be considered to be better than all other feasible solutions, a multiobjective optimization often has multiple optimal, or nondominated, solutions. **Figure 2** illustrates an example with two objectives where the thick curve represents the locations of the nondominated solutions in which preferring a solution over the other means a gain in one objective but a simultaneous loss with respect to the other objective. To successfully solve such problems, it is important to find many (if not all) optimal solutions and examine the trade offs among them.

Two major types of EAs have been developed to solve multiobjective optimization problems. Aggregating approaches are the first type that often uses a weighting scheme to scalarize the vector of multiple objective values into a single value (fitness) and then a regular EA is applied. To implement this approach, the scalarization functions must be based on the relations among objectives, which may not be known in many real world situations.

The second type of approach to multiobjective optimization has attracted more attention in the literature, especially when the goal is to search for a variety of trade off solutions. The majority of these methods takes the notion of nondomination and then maps it to fitness evaluation. In general, an individual will receive a high fitness value if it is not dominated by any other individuals and it is located in a neighborhood (in a space formed by the objectives) with a relatively small number of other individuals. Suppose the black dots in **Figure 2** represent the individual solutions in an EA population for one generation, solution B will be more favorable than A, though both are nondominated. In the context of multiobjective optimization, the diversity of a population has two perspectives. First, the solutions need to be different in the solution space formed by the parameters to be optimized. Second, solutions are also preferred to be different in the objective space formed by the objectives.

Figure 2 The trade-off between two objectives. Both objectives are to be minimized. The shaded area represents the objective space formed by all possible solutions, where each location represents a possible solution to the problem.

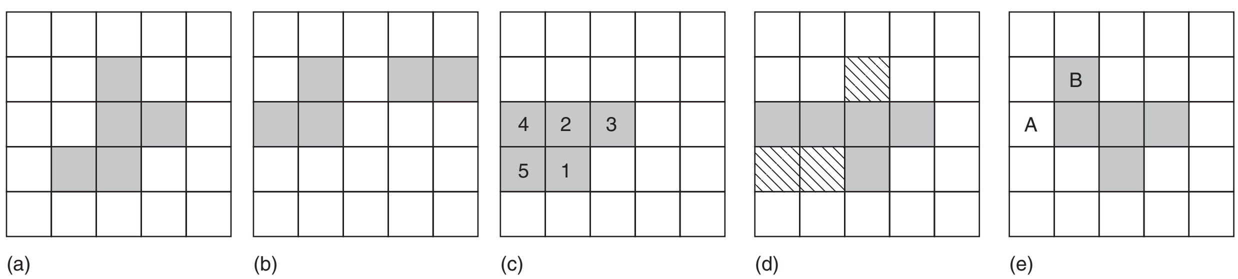

Figure 3 An EA for solving site search problems. Each cell represents a land parcel that is often associated with an acquisition cost.

For multiobjective EAs, the latter version of diversity seems to be more important because one may hope a diverse population in the objective space will promote the chance to reach the optimal front (e.g., the thick line in **Figure 2**). More importantly, a diverse population will lead to a relatively even distribution of EA solutions on the front.

Spatial Optimization Using Evolutionary Algorithms

A significant number of optimization problems in geography involve searching for an optimal configuration of discrete spatial entities or activities with respect to a set of objectives. Examples include location–allocation problems, regionalization problems (school or electoral redistricting problems), and land acquisition problems. Many of these problems also have multiple objectives, especially for decision problems in the public sector that involve a variety of social, economic, and environmental factors.

For some spatial problems, it is straightforward to formulate them in an EA framework. To solve the *p* median problem, for example, a solution can be simply represented using a string of unique integers, each referring to a facility location. For other spatial optimization problems, however, their spatial constraints will require a more careful EA design. Site search problems, for example, are often solved when a contiguous set of land parcels must be acquired so that the acquisition cost is minimized. To solve this kind of optimization problems with spatial constraints, it is important to incorporate spatial topological relationships into the evolutionary operations.

Figure 3 illustrates the operations of an EA that can be used to solve a site search problem. Here, a contiguous set of five land parcels are to be acquired. **Figure 3a** shows a feasible solution where acquired parcels (shaded) are contiguous. **Figure 3b** shows an infeasible solution where acquired parcels are not contiguous. An initialization process is demonstrated in **Figure 3c** where a seed parcel (marked as 1) is used to start a growth process (in a sequence indicated by the order of the numbers) that continues until a full solution is created. **Figure 3d** shows the creation of a new solution by recombining solutions in **Figures 3a** and **3c**. Note that the recombined solution often contains more than the required number of parcels and a repair process is needed to ensure a feasible solution. The repair process continuously removes par cels that can be excluded from the solution without violating the continuity requirement. Parcels in the parent solutions but not in the new solution (after repair) are marked in a fill pattern. **Figure 3e** shows a mutation operation that creates a new feasible solution based on the solution in **Figure 3d**. Here, parcel *A* can be removed from the current solution without creating an infeasible partial solution. Parcel *B* can be added into the current partial solution (after *A* is removed) without violating the contiguity requirement.

Critics and Future Research

Schema theorem has been regarded as a foundation of EAs. A major problem with this theorem is the implicit assumption that the building blocks of a problem are independent, which may not hold for many GAs. Though some efforts have been made to understand EA dynamics using alternative theoretical analyses (e.g., Markov models and statistical mechanics models), a conclusive prediction of EA convergence is still yet to emerge.

As a heuristic approach, an EA cannot guarantee that the optimal solution be found for every problem inci dence. For some engineering problems where the opti mum is required in a controlled system, this limitation may be significant. It should be noted, however, that for many real world problems, the problems are often ill structured, meaning that many important social, eco nomic, political, and environmental factors are not in cluded in the optimization model. The optimal solution to such an ill structured problem may not be optimal when those factors are considered. Therefore, finding suboptimal solutions in an efficient manner may be de sired in many circumstances. As a benefit of using EAs to solve these problems, it is possible to yield a diverse set of high quality solutions that can be subsequently examined by analysts and decision makers. In this way, EAs are search procedures that can be used to help humans learn about the problems being solved before a final decision can be reached.

Finally, it is worth noting the no free lunch theorems, which suggest that 'for any algorithm, any elevated per formance over one class of problems is offset by per formance over another class'. In the context of EAs, it is implied that EAs are not superior to other search algo rithms, when all optimization problems are considered. Though these theorems are not necessarily suitable to a subset of optimization problems, EAs do not run without limitations. An obvious drawback of EAs is their com putation complexity. Though EAs are more efficient than many exact approaches, they may not be as efficient as other metaheuristic methods (e.g., tabu search and simulated annealing) that do not perform on a population of solutions. For a particular set of problems, it will be a useful practice if the analyst is aware of the limitations and select the 'fittest' approach.

See also: Location Analysis; Location Theory.

Further Reading

Bennett, D. A., Xiao, N. and Armstrong, M. P. (2004). Exploring the geographic consequences of public policies using evolutionary algorithms. *Annals of the Association of American Geographers* 94(4), 827 847.

De Jong, K. A. (2006). *Evolutionary Computation: A Unified Approach*. Cambridge, MA: The MIT Press.

Deb, K. (2001). *Multi Objective Optimization Using Evolutionary Algorithms*. Chichester: John Wiley.

Fogel, L. J., Owens, A. J. and Walsh, M. J. (1966). *Artificial Intelligence through Simulated Evolution*. New York: Wiley.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA: Addison Wesley.

Holland, J. H. (1975). *Adaptations in Natural and Artificial Systems*. Ann Arbor, MI: University of Michigan Press.

Koza, J. R. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA: The MIT Press.

Michalewicz, Z. (1996). *Genetic Algorithms + Data Structures = Evolution Programs*. Berlin: Springer Verlag.

Rechenberg, I. (1965). *Cybernetic Solution Path of an Experimental Problem*. Library Translation No. 1122. Farnborough: Royal Aircraft Establishment.

Schwefel, H. P. (1981). *Numerical Optimization for Computer Models*. Chichester: John Wiley.

Simon, H. A. (1960). *The New Science of Management Decision*. New York: Happer and Row.

Wolpert, D. H. and Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation* 1(1), 67 82.

Xiao, N. (2008). A unified conceptual framework for geographical optimization using evolutionary algorithms. *Annals of the Association of American Geographers* 98(4), in press.

Existentialism/Existential Geography

D. Seamon, Kansas State University, Manhattan, KS, USA
J. Sowers, Missouri State University, Springfield, MO, USA

Glossary

Body-Subject The preconscious intelligence of the lived body manifested through action; plays a major role in the habitual, routine aspects of the lifeworld.
Dasein Human being-in-world and referring to the lived fact that human beings are always and already inescapably immersed in the world in which they find themselves.
Existential Geography A range of geographic approaches that insist human experience, awareness, and meaning must be incorporated into any study of peoples' relationships with space, place, and environment; closely related to humanistic geography and phenomenological geography.
Existentialism A way of philosophy that works to understand the basic structures of human existence so that individuals can create useful meaning in their lives through free and informed understanding and action.
Lifeworld The tacit, taken-for-granted context, tenor, and pace of daily life to which normally people give no reflective attention; a major focus of phenomenological investigation.
Phenomenology A philosophical approach that examines and describes phenomena – that is, things and experiences as human beings experience those things or experiences.
Place A fusion of human and natural order and any significant spatial center of a person or group's lived experience.

Introduction

Existential geography refers to a range of geographic approaches that insist human experience, awareness, and meaning must be incorporated into any study of peoples' relationships with space, place, and environment. Existential geography is most closely related to the larger movement of humanistic geography, which arose in the 1970s as a critical response to positivist research that, at the time, dominated the discipline. In emphasizing the conceptual and applied need for understanding the lived relationship between people and their geographical worlds, humanistic geographers argued that positivist geography reduced environmental and place experience and meaning to tangible, measurable, quantitative units, and relationships expressed spatially and materially – for example, symbols on maps, points on graphs, physical linkages in flow diagrams, or figures in equations.

Today the term 'existential geography' is rarely used because the point of view and subject matter are subsumed in different ways by more specific topical and conceptual labels that include phenomenological geography, phenomenological ecology, environmental hermeneutics, ecopsychology, environmental philosophy, and place studies. Geographers of postmodernist, poststructuralist, and critical theory persuasions claim that their approaches reinterpret, improve upon, and supplant earlier existential–geographical work, though geographers and other environmental researchers who continue to work in the existential tradition strongly disagree.

Existentialism and Phenomenology

Philosophically, existential geography is most closely related to the continental traditions of existentialism and phenomenology. Associated especially with the twentieth century philosophers Jean Paul Sartre, Simone de Beauvoir, Albert Camus, Karl Jaspers, Martin Buber, and Gabriel Marcel, existentialism is a way of philosophy that works to understand the basic structures of human existence so that individuals can create useful meaning in their lives through free and informed action. Existentialism insists that philosophy must arise from a person's own life and from his or her own particular individual, historical, and societal situation. "Existence precedes essence," proclaimed Sartre, meaning that human beings experience their world first and only then endow it with self conscious meaning, whether personal, cultural, ideological, or scholarly.

Existentialists claim that earlier Western philosophy was too much concerned with reason, ideas, and abstract theory. Instead, philosophy must start with one's own experience and personal understanding, which can be studied, clarified, and thus enriched. Reason can contribute to this effort, but all other aspects of human being – for example, bodily, sensuous, sexual, emotional, interpersonal, cultural, societal, and transpersonal dimensions – must also be considered and given a place.

The philosophical tradition of phenomenology became important to some existentialists because it offered a practical method for probing and understanding the nature and quality of human existence, though this was not the original aim of phenomenology's founder, philosopher Edmund Husserl, who instead had developed the phenomenological method as a means to identify invariant structures of cognitive consciousness.

Because Husserl viewed consciousness and its essential structures as a pure 'region' separate from the flux of specific experiences and thoughts, his style of phenomenology came to be known as 'constitutive' or 'transcendental'. Eventually, however, other phenomenological philosophers such as Martin Heidegger and Maurice Merleau Ponty reacted against Husserl's transcendental structures of consciousness. These "existential phenomenologists," as they came to be called, argued that such transcendental structures are questionable because Husserl based their reality on speculative, cerebral reflection rather than on actual human experience taking place within the world of everyday life.

In 1927 in his *Being and Time*, Heidegger argued that consciousness was not separate from the world and human existence. He called for an existential correction to Husserl that would interpret essential structures as basic categories of human experience rather than as pure, cerebral consciousness. In 1945 in his *Phenomenology of Perception*, Maurice Merleau Ponty broadened Heidegger's correction to include the active role of the body in human experience. Merleau Ponty sought to reinterpret the division between body and mind common in traditional Western philosophy and psychology. This 'existential turn' of Heidegger and Merleau Ponty shifted away from Husserl's focus on pure intellectual consciousness and toward a reflexive understanding of everyday human life and its lived meanings.

Central Existential Claims

As the preceding account suggests, there are great range and variation in the approaches and conclusions of existential and phenomenological philosophers. Both Camus and Heidegger, for example, repudiated the 'existentialist' label, while Sartre would eventually attempt to reconcile his existentialism with Marxism. These contrasting understandings of existentialism have led to different uses and points of view on the part of geographers claiming some degree of existential perspective in their work, thus, for example, Tim Cresswell, in his 2004 *Place: A Short Introduction*, finds a continuation of earlier existential perspectives in current post structural and postmodernist geographies, while Stuart Elden, in his 2001 *Mapping the Present*, argues that Heidegger's philosophy played a major role in the historical theories of critical theorist Michel Foucault, particularly his understanding of what Elden calls 'spatial history'.

In spite of the wide range of ways in which existentialism has been interpreted – both by its original founders and by geographers later drawing on the tradition – one can identify several conceptual claims that mark out a certain amount of common ground. These claims have been central for many (but not all) geographers seeking to make use of existential principles.

1. To study human beings existentially is to study human experiences, meanings, actions, situations, and events as they happen spontaneously in the course of daily life. On the one hand, the world in which we find ourselves is inexhaustible and much more than each of us as unique individuals can ever experience or know. On the other hand, existentially, the world is only what each of us uniquely experiences and understands it to be.
2. There is no world 'beneath' or 'beyond' the world of our primordial lived experience, which is always and already before us as the world for which we have no choice but to have. Through knowledge, effort, and will, a person or group may over time have some possibility to change the constitution and tenor of their world, but in any particular moment of experience, our world is as it is; it cannot be otherwise.
3. Human awareness and consciousness are always intentional – that is, necessarily oriented toward and finding their significance in a world of emergent meaning. We are never just aware but always aware of something, whether an object, living thing, idea, or the like. This always present quality of intentionality is an invariant feature of human existence and demonstrates that human experience, action, and meaning always unfold in relation to the world in which we find ourselves.
4. This intentional relationship with the world means that person and world are not separate and two but indivisible and one. Human beings are inescapably immersed and enmeshed in their world – what Merleau Ponty called 'body subject' and Heidegger called '*Dasein*', or 'being in the world'. Because of this lived reciprocity between self and world, one cannot assign specific phenomena to either self or world alone. Everything experienced is not only 'given' but also 'interpreted', is 'of the world' but also 'of the person'.
5. This lived reciprocity means that each person's existence gives meaning to his or her world, which in turn gives meaning to that person's existence as it emerges for both himself or herself and for others. This lived reciprocity also means that traditional philosophical dualities – for example, self–world, subject–object, body–mind, perception–cognition, thinking–feeling, personal–transpersonal, people–environment, individual–society, nature–culture, and so forth – must be called into question, revisioned, and rephrased through existential–phenomenological explication. One of the most difficult challenges is to describe the lived reciprocity of person and world in

language and conception that do not fall prey to traditional philosophical dichotomies.

6. The everyday structure through which this lived reciprocity unfolds is the lifeworld – a person or group's day to day world of taken for grantedness that is normally unnoticed and therefore concealed as a phenomenon. One aim of existential study is to disclose and describe the various lived structures and dynamics of the lifeworld, a focus that has become central to existential–geographical research.
7. There are three dimensions of lifeworld that all must be considered in a thorough existential–phenomenological understanding of human existence, experience, and meaning: (1) a person or group's unique personal situation – for example, one's gender, sexuality, physical and intellectual endowments, degree of ableness, and personal likes and dislikes; (2) a person or group's unique social, cultural, and historical situation – for example, the time and place in which one lives, economic and political circumstances, religious and societal background, and technological infrastructure; and (3), a person or group's situation as it involves their being typical human beings sustaining and sustained by a typical human world – for example, the lived fact that we are bodily beings who are always and already 'emplaced' physically in our world.

Existential Perspectives in Geography

Foreshadowed by the humanist work of earlier geographers such as Johannes Granö, J. K. Wright, Clarence Glacken, and David Lowenthal, the existential tradition was initiated in the early 1970s largely by three geographers: Edward Relph, Yi Fu Tuan, and Anne Buttimer. The first explicit discussion of the value that an existential approach might have for geography was in a 1970 *Canadian Geographer* article by Relph, who drew on the insights of existential phenomenology and emphasized that all knowledge arises from the world of experience, including peoples' everyday understanding of their geographical world. In his popular 1974 *Topophilia*, Tuan examined the positive, affective bond between people and place, while Buttimer, in a 1976 article in the *Annals of the Association of American Geographers*, considered how the phenomenological notion of lifeworld might offer insights into sense of place, social space, and time–space rhythms.

The most influential study in this early work was Relph's 1976 *Place and Placelessness*, an existential phenomenology of place, which he defined as a fusion of human and natural order and any significant spatial center of a person or group's lived experience. The existential crux of place experience, Relph claimed, was insideness – the degree to which a person or group belong to and identify themselves with a place. Relph argued that the existential relationship between insideness and its experiential opposite outsideness is a fundamental dialectic in human experience. Through different degrees of insideness and outsideness, different places take on different meanings and identities for different individuals and groups (see **Table 1**). For Relph, 'existential insideness' is the foundation of the place concept because, in this mode of experience, place is experienced without any directed or self conscious attention, yet is laden with significances that are tacit and unnoticed unless the place is changed in some way – for example, one's home and community are destroyed by natural disaster.

The first comprehensive work to demonstrate the considerable topical range that existential approaches could offer geography was David Ley and Marwyn Samuels' 1978 *Humanistic Geography: Problems and Prospects*, an edited collection that included Samuels' chapter, 'Existentialism and human geography'. In this and later work, Samuels drew on Martin Buber's ideas regarding the significance of spatiality in human life to call for geographical research on what he called a "biography of landscape" – the lived foundations of geographical and spatial patterns, interconnections, and attachments. In practice, conceptual and applied efforts to identify and explore such 'existential origins' continued to be conducted by geographers and other environmental researchers drawing on the perspective of existential phenomenology.

Major Themes in Existential Geography

Since the 1970s, partly through the conceptual foundations established by Buttimer, Ley, Relph, Samuels, and Tuan, existential work examining human beings' lived relationship with the geographical world has steadily continued, though overshadowed in the last two decades by the more vocal efforts of postmodernist, post structural, and critical geographies. Most broadly, the aim of current existential–phenomenological work is to understand how qualities of the geographical world, such as place, home, journey, mobility, habitual embodiment, natural landscape, and so forth, found and contribute to human existence, both broadly, in relation to human nature generally, and, specifically, for persons and groups living in particular places, cultural contexts, and historical moments. Three themes are highlighted here to exemplify this work: (1) lived body, (2) place, and (3) a lived environmental ethic.

1. Research on the lived body examines ways in which our existence as bodily beings contributes to our lived geographies. The fact that we are solid, upright bodies gives us a sense of immediate centeredness to space

Table 1 Relph's modes of insideness and outsideness[a]

1.	*Existential insideness*
	A situation involving a feeling of attachment and at-homeness. Place is "experienced without deliberate and self-conscious reflection yet is full with significances" (Relph, 1976: p 55). One feels this is the place where he or she belongs. The deepest kind of place experience and the one toward which people probably all yearn.
2.	*Existential outsideness*
	A situation where the person feels separate from or out of place. Place may feel alienating, unreal, unpleasant, or oppressive. Homelessness or homesickness would be examples. Often, today, the physical and designed environments contribute to this kind of experience 'unintentionally', e.g., the sprawl of suburban environments, the dissolution of urban downtowns, or the decline of rural communities.
3.	*Objective outsideness*
	A situation involving a deliberate dispassionate attitude of separation from place. Place is a thing to be studied and manipulated as an object apart from the experiencer. A scientific approach to place and, typically, the approach most often taken by planners, designers, and policymakers.
4.	*Incidental outsideness*
	A situation in which place is the background or mere setting for activities, e.g., the landscapes and places one drives through as he or she is on the way to somewhere else.
5.	*Behavioral insideness*
	A situation involving the deliberate attending to the appearance of place. Place is seen as a set of objects, views, or activities, e.g., the experience one passes through when becoming familiar with a new place figuring out what is where and how the various landmarks, paths, and so forth all fit together to make one complete place.
6.	*Empathetic insideness*
	A situation in which the person, as outsider, tries to be open to place and understand it more deeply. This kind of experience requires interest, empathy, and heartfelt concern. Empathetic insideness is an important aspect of approaching a place phenomenologically.
7.	*Vicarious insideness*
	A situation of deeply felt secondhand involvement with place. One is transported to place through imagination through paintings, novels, music, films, or other creative media. One thinks, e.g., of Claude Monet's paintings of his beloved garden Giverny, Thomas Hardy's novels describing nineteenth-century rural England, or John Sayles' films portraying conflicts within and between particular American communities and regions.

[a] Source: Relph, E. (1976). *Place and Placelessness*, pp 51 52. London: Pion.

and – in conjunction with upright posture and binocular vision – orients us within a sixfold directional axis of up–down, front–back, and left–right. We experience the world largely through the five senses, and existential–phenomenological research has studied how each sense offers a different but complementary way of knowing the world. Existential–phenomenological studies of the lived body have also examined specific environmental experiences of particular individuals or groups, thus Jonathan Cole and Kay Toombs explored how less abled persons' loss of mobility leads to a changed interaction with the surrounding physical and human world, while sociologist Chris Allen examined the bodily encounters and resulting social geography of visually impaired children.

Beyond bodily structure and process, there is research on the prereflective but learned intentionality of the body, which Merleau Ponty termed body subject—the preconscious intelligence of body manifested through action. For geography, one of the most significant aspects of body subject is its automatic, taken for granted ability to work in extended ways over time and space. One can ask how routine behaviors of individuals coming together regularly in space can transform that space into a place with a regular dynamic and character. David Seamon has explored how such 'place ballets' can be established and sustained through environmental design and policy.

2. The theme of place has been a central focus in existential–phenomenological work since the seminal studies of Buttimer, Tuan, and Relph highlighted above. In philosophy, Edward Casey, Jeff Malpas, and Robert Mugerauer have written book length accounts arguing for place as a central ontological structure founding human experience. As Casey wrote in his 1993 *Getting Back into Place*, "place, by virtue of its unencompassability by anything other than itself, is at once the limit and the condition of all that exists… To be is to be in place." (Casey, 1993: 15–16).

Much of this place research has focused on particular groups' lived relationships with particular places; one early example was David Ley's study of African American neighborhoods in inner city Philadelphia presented in his 1974 *The Black Inner City as Frontier Outpost*. More recent efforts include Ray Oldenberg's 1989 *The Great Good Place*, which highlights 'third places' – public or semipublic establishments where people informally gather and socialize; Mindy Fullilove's 2004 *Root Shock*, a study of the traumatic impact that urban renewal had on African American neighborhoods in several American cities; and Douglas Rae's 2003 *City*, an examination of traditional American urbanism grounded in lively streets and place based community, using the city of New Haven, Connecticut, as a study focus.

3. Research involving a lived environmental ethic considers ways whereby people might become more sensitive to and involved with the natural world so that their wish and will to care for the environment becomes stronger and more durable. Mugerauer, Relph, and philosopher Ingrid Stefanovic have all drawn on Heideggerian thinking to suggest possible ways whereby people might better care for the natural world. In his 1981 *Rational Landscapes and Humanistic Geography*, for example, Relph developed the concept of environmental humility – a way of seeing and understanding that is responsive to the best qualities of the Other and that might foster a compassion and gentle caretaking for places, people, and the things of nature. In a related way, Stefanovic drew on Heidegger's understanding of place and emplacement as means to reconsider the nature of environmental and economic sustainability.

Criticisms of Existential Geography

With the development in the 1980s and 1990s of post modernist, post structural, and critical geographies, existential and humanistic geography has faced numerous conceptual and ideological criticisms that can most broadly be described by charges of: (1) essentialism, (2) authoritativeness, (3) voluntarism, and (4) an ideological bias toward bounded, static, exclusionary places.

As indicated by the writings of Buttimer, Relph, Samuels, and Tuan, early existential work in geography focused largely on human typicality with the result that critical and social constructionist critics label existential studies as essentialist – that is, presupposing and claiming an invariant and universal human condition only revealed when all 'nonessentials' such as culture, economics, or history are stripped away, leaving some inescapable core of human experience and existence. A second, related criticism, leveled especially by feminist geographers, is that the existential approach is authoritative, in that it privileges the interpretive sensibilities of academic experts who arbitrarily claimed to provide an accurate depiction of the geographical situations and experiences of more ordinary human beings. This depiction, the feminist critique argues, is in fact grounded in an implicit masculinist point of view assuming that the experience of (mostly) academically trained men could represent all human situations – for example, the experiences of women, gays and lesbians, the less advantaged, particular racial and ethnic groups, and so forth.

As already pointed out, however, any thorough existential presentation of the geographical lifeworld requires that, in addition to human typicality, researchers must also recognize and examine the personal and cultural dimensions of particular environmental and place experiences; the already cited work of Allen, Cole, Fullilove, Oldenburg, Rae, and Toombs are convincing examples of the conceptual and lived effectiveness of existential work dealing with specific individuals, groups, and places. Existential researchers claim that their people and place portrayals, grounded in the openness of phenomenological method, are more accurate and fair than social constructivist or critical accounts that too often force fit their subjects through some predefined set of cultural, political, gender, or socioeconomic filters, and thus end in lifeworld misrepresentation and distortion.

A third criticism of the existential approach is that it is voluntarist – that is, that it tacitly views society and the world as a product of intentional, willed actions of individuals. This criticism arises from critical geographers, especially neo Marxists and feminists, who give attention instead to the broader societal structures and power relations that underlie particular lifeworlds and places. As a way of philosophy that highlights self awareness as a vehicle of personal and societal change, the existential perspective places greatest emphasis on individual freedom, understanding, and agency. On the one hand, there is an underlying ideological assumption that individual will and personal growth provide the major means for making a difference in the larger world. On the other hand, an existential approach can examine power, exclusion, resistance, justice, political process, and other central emphases of critical geographers, though little work has been done in this direction, perhaps because most existential thinkers instinctively favor an interest in experience, selfhood, and personal and group autonomy; one exception is political theorist Daniel Kemmis's 1990 *Community and the Politics of Place* and his 1995 *The Good City, the Good Life*, two books that explore how citizens' sense of responsibility for the place in which they live might facilitate a civilized politics.

A fourth criticism of the existential approach is that it portrays lifeworlds and places as bounded, static, exclusionary, and reactionary; that place, insideness, dwelling, and rootedness are ethically favored over nonplace, outsideness, journey, and mobile, shifting modes of life and styles of identity. This criticism has been brought forward by post structuralists, who emphasize relativist, shifting meanings, and distrust any kind of permanent conceptual or lived structures. Existential thinkers respond that these post structuralists misread existential conclusions and that, throughout existential–geographical work, there is a flexibility of conceptual expression – a recognition that an excess of place can lead to a provincialism and callousness to difference just as an excess of journey can lead to a loss of identity or an impartial relativity that allows for commitment to nothing. One central question that thinkers like

Kemmis, Oldenburg, and Seamon address is whether, through design and policy, progressive, connected places can be made that draw together human diversity through everyday environmental co presence and co awareness – for example, the potential that a small grained, permeable pathway system has for bringing people of difference together in city districts through informal, serendipitous sidewalk and street encounters.

Post structural critics are also concerned that existential work too often seems a nostalgic paean to pre modern times and places. How can authentic places, at homeness, and dwelling exist in a postmodern era of human diversity, rapid technological change, globalization, and geographical and social mobility? This criticism, however, ignores a central conclusion of existential thinkers such as Relph, Casey, Malpas, Mugerauer, and Stefanovic: that regardless of the particular cultural, technological, or historical situation, people will always need place because having and identifying with place are integral to what and who we are as human beings.

From this perspective, the postmodernist devaluation of place as it supports individual and group identity and solidarity is questionable existentially and potentially devastating practically. Instead, the crucial theoretical and real world question is how a conceptual and applied understanding of place might lead to the making of actual places that draw together diverse individuals and groups through a shared sense of environmental involvement and attachment. In our mobile, ever changing, globally dominated world, one of the few real world ways to accommodate a pluralistic society may be the power of place to gather human differences together spatially, and thereby sustain the autonomy of individuals and groups as, at the same time, they feel a sense of belonging and commitment to the larger place of which they are a part.

See also: Body, The; Critical Geography; Cultural Geography; Feminism/Feminist Geography; Humanism/Humanistic Geography; Phenomenology/Phenomenological Geography; Place; Positivism/Positivist Geography; Postmodernism/Postmodern Geography.

Further Reading

Allen, C. (2004). Merleau Ponty's phenomenology and the body in space encounters of visually impaired children. *Environment and Planning D: Society and Space* 22, 719 735.

Casey, E. (1993). *Getting Back into Place*. Bloomington, IN: Indiana University Press.

Cloke, P., Philo, C. and Sadler, D. (1991). 'Peopling' human geography and the development of humanistic approaches. In *Approaching Human Geography: An Introduction to Contemporary Theoretical Debates*, ch. 3. New York: Guilford Press.

Cole, J. (2004). *Still Lives: Narratives of Spinal Cord Injury*. Cambridge, MA: MIT Press.

Hay, P. (2002). Reclaiming place: Seeking an authentic ground for being. In *Main Currents in Western Environmental Thought*, ch. 6. Bloomington, IN: Indiana University Press.

Hubbard, P., Kitchin, R., Bartley, B. and Fuller, D. (2002). *Thinking Geographically: Space, Theory and Contemporary Geography*. London: Continuum.

Johnston, R. J. (1997). Humanistic geography. In *Geography and Geographers*, ch. 6. London: Arnold.

Malpas, J. E. (1999). *Place and Experience*. Cambridge: Cambridge University Press.

Mels, T. (ed.) (2004). *Reanimating Places: A Geography of Rhythms*. Burlington, VT: Ashgate.

Mugerauer, R. (1994). *Interpretations on Behalf of Place*. Albany, NY: State University of New York Press.

Relph, E. (1976). *Place and Placelessness*. London: Pion.

Seamon, D. (ed.) (1993). *Dwelling, Seeing, and Designing: Toward a Phenomenological Ecology*. Albany, NY: State University of New York Press.

Seamon, D. (2000). A way of seeing people and place: Phenomenology in environment behavior research. In Wapner, S., Demick, J., Yamamoto, T. & Minami, H. (eds.) *Theoretical Perspectives in Environment Behavior Research*, pp 157 178. New York: Plenum.

Stefanovic, I. L. (2000). *Safeguarding Our Common Future*. Albany, NY: State University Press of New York.

Toombs, S. A. (1995). The lived experience of disability. *Human Studies* 18, 9 23.

Relevant Websites

http://www.phenomenologycenter.org
Center for Advanced Study in Phenomenology.

http://www.arch.ksu.edu
Environmental and Architectural Phenomenology Network, College of Architecture, Planning and Design at Kansas State University.

http://www.environmentalphilosophy.org
International Association for Environmental Philosophy.

http://www.uri.edu
Merleau Ponty Circle, University of Rhode Island.

http://www.phenomenologyonline.com
Phenomenology Online.

http://pegasus.cc.ucf.edu
Research on Space and Place Website, University of Central Florida.

http://www.spep.org
Society for Phenomenology and Existential Philosophy.

http://pages.slu.edu
Society for Phenomenology and the Human Sciences, Saint Louis University.

http://plato.stanford.edu
Stanford Encyclopedia of Philosophy: Existentialism, Stanford University.

Experimental Design

S. Bell, University of Saskatchewan, Saskatoon, SK, Canada

Glossary

Between-Subjects Experimental Design An experiment in which each case is exposed to only one level of the independent variable.

Dependent Variable Those variables in the experiment that are measured and/or observed. Dependent variables will respond to predetermined variation in the independent variable in ways will either support or reject the hypothesis.

Hypothesis A statement about the relationship between experimental variables that can be tested by conducting a controlled scientific experiment. Hypotheses are often associated with a researcher's best guess as to the outcome of an experiment.

Independent Variable Independent variables are set or determined by the researcher prior to the experiment's outset.

Population The universe from which samples are taken and to which results are generalized.

Sample The subset of the population that is involved in the experiment. The existence in the population of a pattern observed or identified in the sample can be determined using the theory and practice of statistics.

Within-Subjects (Repeated Measures) Design An experiment in which each case is exposed to all levels of the independent variable.

Introduction

Experimental design is the process of carrying out research in an objective and controlled fashion so that precision is maximized and specific conclusions can be drawn regarding a hypothesis statement. Generally, the purpose is to establish the effect that a factor or independent variable has on a dependent variable. The principles of experimental design play an important role in research that does not follow the strict tenets of hypothesis testing and this article holds relevant information for this type of research. However, this article is primarily focused on planning research that will utilize experimental design in order to test and validate the relationship between and among experimental variables. Important topics germane to experimental design include hypothesis statements, experimental control, specifying independent and dependent variables, selection and assignment of samples or participants to conditions, collecting data, and selecting valid statistical tests. Through accurate and precise empirical measurement and control an experimental design increases a researcher's ability to determine causal relationships and state causal conclusions.

Experimental design as a subset of scientific investigation is a popular and widely used research approach. The essence of experimental design and perhaps the most important reason researchers choose to design and conduct experiments is the precision with which one can analyze the relationship between and among variables and to make that analysis as objective as possible. To look at it from another perspective, experimental design minimizes ambiguity and attempts to eliminate confusion. A true experimental design relies on testing the relationships between and among variables; generally speaking, one variable, the independent variable, is controlled in order to measure its effect on other, dependent, variables. A central concern of any researcher using experimental design must be control; in experiments, the researcher chooses an intervention, associated with the independent variable, and controls how that intervention is applied, or introduced, into the research setting. If the experimental design is applied correctly, then a causal relationship can be established between the independent variable and dependent variable(s).

There are several steps in the development of an experiment. The researcher must generate a research question, state a testable hypothesis, determine how to control variability during the experimental process, select or develop intervention conditions, sample from a population in order to assign them to experimental conditions, and determine what empirical measures will be made (and how data will be recorded). There is often an important and intimate relationship between the experimental design and the type of data that is collected and statistical approach that will be used to analyze the data. In geography, and cognate disciplines that are interested in spatial variation, there is a subset of research that deals with data that does not conform to all of the assumptions central to traditional statistical tests. As a result spatial statistical tests have been developed to account for the unique character of spatial data; these differences will be covered briefly in a later section detailing spatial statistical approaches in geography.

Developing and Stating a Testable Hypothesis

Initially, developing and stating a testable hypothesis is the principal requirement of using an experimental design to conduct research. A hypothesis is based on a careful reading and consideration of past research and represents a researcher's 'best guess' as to how something happens or how two or more variables will interact with one another. An experiment is an attempt to create an environment that tests the hypothesis such that the outcome is as close to reality as possible. There are several components of experimental design that play a role in maintaining experimental control and allowing researchers to make statements regarding the generaliz ability of their results beyond the confines of the con trolled experimental setting. Immediately preceding the development of an experiment, one must decide upon a testable hypothesis; this, in turn, is based on a the se lection of a general topic of interest and perhaps the development of a 'science question' that is the product of reasoned exploration of theory and past research, previ ous experimental outcomes, and considerable thought by the researcher. The choice of a research topic and even the statement of a research question (or a science ques tion) is a relatively exciting and loosely governed process; it is here that a researcher can let their mind wander and consider all the possibilities that scientific inquiry offers. As these exciting times coalesce into more formal and clearly stated questions, the researcher must then do the 'heavy lifting' associated with turning their wild ideas into testable hypotheses.

Jeffrey Lee suggests that hypotheses can be developed in one of four ways: analogy, induction, deduction, and intuition; each offers its own advantages and presents specific challenges. A prerequisite for any scientific en deavor that relies on the development of a testable hy pothesis is the expertise of the researcher. Developing a hypothesis through 'analogy' relies on being able to see the similarity between different but related phenomena or processes. Knowing how water moves through one type of soil might help develop a hypothesis regarding its flow through other materials, soil types, or under dif ferent conditions. Through 'induction' a researcher can exploit their scientific and theoretical knowledge to de velop a hypothesis to test the generalizability of a dis crete observation. 'Deductive' hypotheses are based on a sound understanding of the processes and properties that underlie observable and testable phenomena. When 'intuition' is the basis for hypothesis generation, the researcher looks for what 'seems right' with little back ground to guide the reasoning behind the stated hypothesis.

Once a testable hypothesis is developed and stated, the work on experimental design can begin. Independent of how much time and effort went into the generation of the hypothesis, the researcher must accept that the quality of the hypothesis remains unknown until it is tested. While there are several different approaches that can be used to develop a testable hypothesis, performing the test requires close adherence to the principles of science. Furthermore, careful and controlled thought and the deliberate application of reason will be central to the success of one's experimental design and their ability to test the stated hypothesis (in experimental jargon the researcher wants to accept the alternative hypothesis by refuting the null hypothesis).

Experimental Variables

One of the first steps in the experimental design process is the selection of variables, both independent and dependent, associated with phenomena that will be measured. Dependent variables are those that are meas ured and/or observed during the experiment while independent variables are set or determined by the re searcher prior to the experiment's outset. When designed appropriately the independent variables are intended to have some bearing on the dependent variable during the experiment (the hypothesis states there will be an interaction or effect). In most scientific experiments, the independent variable, or factor, in an experiment has predetermined levels, usually nominal, which are also called conditions. This fact generally leads to the use of analysis of variance (ANOVA) statistical tests. With the introduction of more than one factor, each with two or more conditions, the research can establish not only the effect of a single factor (usually called a main effect) but also interaction among factors (called interaction effects). For instance, let us consider a study interested in how age and spatial ability are related to one's ability to point to locations in a recently experienced environment. An equal number of female and male participants of three distinct ages (9 years, 13 years, and adults) are asked to follow a multi turn path through an urban environment. At the end of the path, participants are asked to point to several locations along the path. This experimental design has the potential to test main effects of age (is pointing accuracy greater for older participants?) and sex (is point accuracy different for males and females?) as well as more complicated interaction effects (do sex differences in pointing accuracy vary with age?). Dependent variables can often be measured at a variety of levels, nominal, ordinal, interval, or ratio; in some situations the researcher can choose the level of precision they wish to have, in others the variable can only be measured in one way. In the example above, the dependent variable is pointing accuracy, which could be measured by the amount of difference in degrees between a target's actual direction and the participant's indication of its direction. Constants are another important factor in experimental design; these are elements in the experiment that do not change over its course.

When putting all of these elements together, it is important to keep in mind that the best way to establish relationships between and among variables, particularly causal relationships, is to establish an experiment that allows for empirical observation resulting in measures that can be compared among or between the variables.

On a related and complicating topic, geographers are often faced with a lab versus field dilemma, most of what we are interested in studying or explaining happens in a naturalistic setting which has much more in common with a field based experiment. However, establishing an experiment in a naturalistic setting is complex and the researcher runs the risk of introducing confounding variables that are not controlled and might not even be known by the experimenter. The experiment might be affected by variables that cannot be controlled yet are related to and change the value of the dependent variable(s). There is no correct answer to the question, "should my experiment be based in a lab (more control) setting or a naturalistic (field) setting that might offer greater ecological/external validity?" A lab setting offers more control, but the extent to which one can generalize lab based findings will always be debated; on the other hand, establishing causality between two variables is easier in a lab setting. Some researchers will call experiments that lack complete, or lab based, control quasi experiments, even though it is possible to create field based experiments that include manipulations of independent variables in order to observe measurable changes in a dependent variable.

Direct and Indirect Measurement

In geography, as is the case in many disciplines, the measurements we make in an experiment may be direct or indirect. Indirect measures are those that rely on an inference between what is observed and the phenomena of interest. For instance, the decisions associated with how to navigate from one location to another are generally made in one's head as a result of a variety of cognitive structures and processes loosely associated with our cognitive map. In order to understand the nature of this decision making, it is often only possible to observe (and measure) the spatial behavior that results from the decision making process. If we are interested in examining whether humans are spatially rational decision makers when navigating we might state a hypothesis like "humans will consistently choose the shortest path between two points when navigating through the environment," in turn we can have participants in our study walk from an origin to a destination and observe what paths are selected. The final step would be examining how many participants chose each path (shortest and others) or to measure the amount of difference (in length) there was between the various paths chosen and the shortest path. The former is an example of choosing a nominal measurement approach (different paths), whereas the latter is a ratio measure (metric distance of each path traveled).

Within and between Case Experimental Design

The nuts and bolts of experimental design deals with how cases are assigned to levels of one or more independent variables. When an independent variable is manipulated so each case is tested at all levels of the variable, then the experiment is a repeated measures design. If each case is only exposed to one level of the independent variable, and other cases are exposed other levels, then the experiment is a between subjects or between case design. Repeated measures designs require slightly different statistical tests, require fewer cases/participants for the same statistical power, and provide the research with more control over potentially confounding variables. In spite of these important benefits of a repeated measures design, between subjects designs are required in many situations where a repeated measures design is impossible or inadvisable. Inadvisable situations include settings in which exposure to one condition may affect the state of the case for later conditions or in an experiment using human participants where exposure to a test condition might cause learning, or will benefit them in later conditions. Impossible repeated measures experiments might include situations in which the cases have unique characteristics that cannot be held constant. Such is the case in many comparative studies that are interested in comparing two distinct places or populations.

Spatial Statistics and Experimental Design

While traditional experimental design involving the use of hypothesis statements, controlled interventions, objectivity, dependent and independent variables, sampling techniques, correlation and causality, measurement, and bias, there are aspects of what we study that have required geographers and cognate researchers to develop tools that do not strictly conform to the tenets of experiment based scientific exploration. The use of census and other population surveys, remotely sensed imagery and other 'samples' that represent the known population or reflect a single point in time are not necessarily intended to be a representative sample of other times or places, and the presence of spatial dependence between observations results in an inability to meet some of the assumptions associated with traditional stochastic statistical methods that are normally applied to experiment based data collection. The systems that geographers study are often complex and hard to replicate; furthermore, when our sample is the population (census and other population level datasets), it is virtually impossible to obtain (determine) the data distribution under the null hypothesis, each of which is an assumption of traditional statistical methods and an assumed component of experimental design. Geographers have developed two responses to these problems that generally apply one of two strategies: new spatial statistical methods (spatial autocorrelation and spatial regression) that account for spatial dependence, and the application of traditional experimental design and stochastic statistics in ways that are congruent with the traditional assumptions of both to as great an extent as is possible. In the latter case, geographers might relax, or ignore, some of the assumptions of traditional experimental design or stochastic statistics in order to benefit from the control offered by these techniques.

Spatial analytic and spatial statistical methods are arguably the most important approach to overcoming statistical assumptions that are to sampled nonspatial data. For geography they are becoming an increasingly important contribution to researchers both within and beyond geography. Geographic theories are gaining exposure in disciplines not traditionally associated with a spatial or geographic perspective. Furthermore, geographers are offered an opportunity to challenge and develop new theories that are spatial and geographic in nature. Researchers in natural, social, and medical sciences are relying on exploratory spatial data analysis, spatial analysis, spatial statistics, and cartography to explore important research topics far beyond the capacity of researchers in geography. Some of these developments can be traced to the emergence of GIS, integrated computing systems for handling, analyzing, and representing spatial data. However, for the most part spatial statistics have been the product of geographers demanding and developing new techniques for problems that have an important spatial component. Without an appreciation for the role that space plays, it becomes more difficult to verify the role played by other nonspatial variables. Geographers have been essential for the development of techniques (statistical and otherwise) that allow for the measurement and control of spatial effects in univariate and multivariate analysis.

See also: Hypothesis Testing; Map Perception and Cognition; Mental Maps; Mixed and Multiple Methods; Naturalistic Testing; Quantitative Methodologies; Sampling; Scale Analytical; Scientific Method; Statistics, Inferential; Statistics, Overview.

Further Reading

Lee, J. A. (1999). *The Scientific Endeavor: A Primer on Scientific Principles and Practice*. San Francisco, CA: Addison Wesley Longman.

Montello, D. R. and Sutton, P. C. (2006). *Introduction to Scientific Research Methods in Geography*. Thousand Oaks, CA: Sage Publications.

Exploration

S. A. Royle, Queen's University Belfast, Belfast, UK

Glossary

Age of Discovery A period from the fifteenth to the seventeenth centuries when European maritime nations explored the world in search of trade goods such as gold and spices, supplementing earlier penetrations overland. The Age was associated with technical developments in ship design, which permitted safer open ocean travel; improvements in cartography and navigation; also the development of capitalism and other advances in this early modern era in Europe.

Chomolungma The Tibetan name for Mount Everest, meaning 'goddess mother of the world'.

Circumnavigation The act of traveling around the globe and returning to one's starting point. The route should include crossing two regions at opposite ends of the globe (antipodes). Traditionally circumnavigations referred to voyages by sea; in modern times air travel has also been included. The first undisputed circumnavigation was the expedition led initially by the Portuguese sailor Fernando Magellan in 1519–21.

Merchant Companies Chartered European trading companies active in the commercial penetration into Asia and elsewhere during the Age of Discovery and later. Examples include The Company of Merchants of London Trading into the East Indies (the East India Company, chartered 1600) and the Dutch Vereenigde Oostindische Compagnie (chartered 1602).

Northwest Passage A sea route connecting the Pacific and Atlantic Oceans via the north of Alaska and the Canadian mainland. European explorers sought this route to the riches of Asia from as early as the fifteenth century but never successfully and only in 1906 was the first passage made, and that in a small ship by Roald Amundsen. Ships capable of dealing with ice can navigate the route, but it has never proved to be commercially viable.

Spice Islands A name used for the Molucca Islands (Malaku), although sometimes others in the region are included such as Java, Sumatra, Borneo, and Celebes. Their products, particularly spices like cinnamon, cloves, mace, and nutmeg, were prized during the early European penetration into Asia.

Third Pole The highest point on the Earth's surface, thus Mount Everest or Chomolungma.

This article considers the early commercial and then colonial imperative for exploration, moving from the exploration of encounter, the 'discovery' of place to exploration in detail, and the knowledge of place. This latter was often associated with the administration and/or exploitation of territory through colonialism. Later, a search for scientific knowledge became more prominent, joined but not entirely replaced by an attitude which regarded exploration as something of a sport – the Mallory tradition – although a quest for glory of some sort had accompanied much exploration throughout its history.

Travelers' Tales

From well before the age of print, adventurers would return and scribe accounts of strange lands for curious people. The publication by Marco Polo (1254–1324), his *Description of the World*, written after his experiences in Asia in the thirteenth century was only the most exotic of the genre. There were more prosaic works. For example, Giraldus Cambrensis (*c.* 1146–*c.* 1223), as well as writing on his own country of Wales, composed *Topographica Hibernica* in 1188 after a visit to nearby Ireland during the era of Norman expansionism from Britain into that island. Sometimes people followed in the explorers' footsteps (or wakes), moving to the new (to them) lands to trade or settle, stimulated by the travelers' tales. It is even thought that Greenland may have been given its inappropriate name as an inducement to migrants after the Vikings had found and settled in parts of that icy land mass in AD 982, living there for about 500 years. They continued to explore westwards and there was a Norse settlement on Newfoundland, at what is now called l'Anse aux Meadows for at least a few years after the year 1000. Christopher Columbus' party were certainly not the first Europeans to see what became North America.

'Because It Is There': The Mallory Tradition of Exploration

One traveler who wrote an account of his journey was Robert Falcon Scott (1868–1912) who memorably penned "Great God, this is an awful place" in his tent in Antarctica shortly before his death in 1912 on the return of the British expedition from the South Pole, where they had been only the second party there, beaten by the Norwegians led by Roald Amundsen (1872–1928). Scott epitomized a certain strand of exploration: relatively wealthy people taking themselves to the furthest or highest places for the sake of it. Echoes of Scott are still heard: the British adventurer Sir Ranulph Twisleton Wykeham Fiennes, third baronet (usually known as Ran Fiennes (born 1944)), was the first man to reach both poles by land as part of the Transglobe Expedition of 1979–82 and, with Mike Stroud (born 1955), the first to walk across Antarctica in 1993. A decade after Scott's death, Englishman George Mallory (1886–1924) was to try to climb Mount Everest 'because it is there', although, it seems, he did not actually recall making his famous remark. He and Andrew Irvine (1902–24) died on the mountain in 1924, having perhaps been the first people to reach the 'third pole', the world's highest point. The earliest confirmed ascent was not until 1953, an expedition sponsored by the Alpine Club of Great Britain and the Royal Geographical Society (RGS).

The announcement of the achievement of this British expedition was made the day before the coronation of Elizabeth II and the news was swept into the celebrations. Elizabeth II is also queen of New Zealand and could thus knight one of the successful climbers, New Zealander Edmund Hillary (1919–2008). Hillary's companion on the final ascent was a Nepalese man who climbed Everest not just because it was there but because he had been working as a high altitude porter to expeditions since the 1930s. The Nepali atop, to him, not Everest but Chomolungma was a Sherpa. Sherpas are actually an ethnic group found in Tibet and northern Nepal but the name is usually associated with the occupation of portering, and for many years Hillary's companion was popularly known as Sherpa Tenzing (1914–86); the second of the paired names: Hillary and Tenzing. More recently, in the era of postcolonial awareness, references to Tenzing tend to drop 'Sherpa' in favor of Tenzing Norgay and he takes first place sometimes, as with the annual Tenzing Hillary Everest Marathon. Hillary himself spent much of his later life involved with development and environmental work in Nepal, but one implication from the way in which the story of 1953 was originally portrayed is that exploration is an activity undertaken by people from rich countries and that poor countries, with less noteworthy citizens, are places to be explored.

Exploration and Economic Gain

That exploration takes place from relatively rich countries is a commonality throughout its history, but the Mallory tradition is relatively recent, a product of an age and society able to free up the resources necessary to fund such nonproductive activity. Exploration's story is more than filling in blanks on the map or seeking entries in record books. Rather, it was for long associated with economic gain. One of exploration's great sagas is the search for the Northwest Passage, a route sought to facilitate trade between Europe and Asia via the north of what is now mainland Canada and Alaska. This beguiled Arctic explorers from the time of John Cabot (Giovanni Caboto, *c.* 1450–*c.* 1499) in the late fifteenth century, through a series of men who have left place names to mark their failed attempts: Martin Frobisher (*c.* 1535–94), John Davis (*c.* 1550–1605), Henry Hudson (*c.* 1570–*c.* 1611), Robert Bylot (fl. 1610–16), and, later, John Franklin (1786–1847). Ironically, after all that British effort it was none other than Roald Amundsen in *Gjøa*, who first completed the journey along the length of the passage in 1903–06.

Louis Antoine de Bougainville (1729–1811), the notable French maritime explorer of the eighteenth century, was a mathematician who made important astronomical observations on his circumnavigation of the globe but also claimed territory for France, including Tahiti in 1768 (fallaciously, as an English navigator, Samuel Wallis (1728–95) was there a few months earlier). de Bougainville's contemporary, British captain James Cook (1728–79), surely the man who 'discovered' more of the world than anybody else (a ship on his third expedition was even called *Discovery*), had a scientific mission for his three great voyages (1768–71; 1772–75; 1776–79) but he also claimed much territory for Britain, while his notable achievements in surveying and improvements in navigational techniques led to commercial benefit. 'Discovered' was placed in inverted commas above to denote that most of Cook's and other explorers' discoveries are merely the identification for a domestic audience of places hitherto unknown to them. Local people knew they were there. Only a few places such as some uninhabited Atlantic islands were truly discovered by Europeans, examples being Bermuda (*c.* 1505, Juan de Bermudez, Spanish, died 1570), St. Helena (1502, João da Nova Castella (Juan de Nova), a Spaniard working for Portugal, died 1509), and the Falkland Islands (1592, John Davis, English), although there may have been earlier sightings back to Amerigo Vespucci (an Italian working for Portugal, 1454–1512) in 1501 or 1502.

Zheng He

It was not just Europeans who were exploring in early periods; the best known of the Asian travelers is the Chinese admiral Zheng He (Cheng Ho, 1371–1433) who commanded or organized seven voyages from China to Arabia, India, Southeast Asia, the Persian Gulf, and the east coast of Africa between 1405 and 1433. A recently discovered map purports to be a copy of a 1418 original and may indicate that Zheng He circumnavigated the world a century prior to Fernando Magellan's accepted claim of having commanded the first expedition to do so from 1519 to 1522. Zheng's voyages were exploration of a type that was to become common elsewhere. There was government involvement; the state on whose behalf the voyages were made was relatively more advanced than the places to which sailed her admiral (or his commanders, Zheng did not always go himself). Further, the purpose of the voyages was to enrich the sponsoring state. Among the largest ships in the huge Chinese fleets were those designed to carry back the treasure. Another element was the beginnings of colonialism, including the establishment of a Chinese outpost at Malacca on the Malay Peninsula and Zheng described the people of the countries beyond the horizon as ‘subjects’ of China. However, shortly after the last of these voyages, the Chinese rulers became focused on their homeland and both the voyages and the nascent colonialism they incorporated ceased.

The Age of Discovery

Instead of China, it was Europe, which from the fifteenth century until into the seventeenth century developed the classic exploration model in what has become known as the Age of Discovery.

Henry the Navigator

One important actor was Henry the Navigator (1394–1460), son of the King of Portugal. A devout Catholic, Henry as a young man was involved in crusades against Muslim holdings in North Africa. He went on to organize voyages further down the coast of Africa and into the Atlantic, many of whose islands were truly discovered by Portuguese seafarers and some of which, such as the Azores, were colonized and exploited for timber and agricultural products. Henry sponsored many voyages to Africa without significant return, until in the 1440s the use of an advanced craft, the caravel, enabled his captains to reach further along the west coast of Africa where gold and slaves were obtained. This stimulated the interest of the private sector in exploration and trade and commercial expeditions began to be operated with some success under license from the Portuguese authorities. By 1454, other trading peoples such as the Venetians and Genoese were taking part in Portuguese voyages, while the pope had recognized the claims of Portugal to the islands and coastal lands her captains had found. In the decades after Henry’s death his country’s explorers pushed Portugal’s influence ever further east. Bartolomeu Dias (*c.* 1450–1500) sailed round the Cape of Good Hope in 1488; Vasco da Gama (*c.* 1469–1524) led a fleet to India in search of spices in 1497, and Goa was established as a permanent Portuguese settlement on the west coast of India in 1510, in the manner of African settlements such as Portuguese Guinea (now Guinea Bissau) taken in 1446. Later, in 1557, the Portuguese were to establish a trading enclave on the Chinese coast at Macau, which was governed by Portugal until 1999.

Gold, God, and Glory

Meanwhile, Portugal’s neighbor, Spain, was exploring islands and coasts on the other side of the Atlantic, employing Spaniards and other Europeans such as Amerigo Vespucci of Florence (after whose first name it is thought America is named) and Christopher Columbus (Cristóbal Colón, 1451–1506) of Genoa. The Spanish were seeking a western route to Asia and Columbus thought that in 1492 he had reached that continent, not that he had (re)‘discovered’ the new world. As with Portugal, the Spanish established colonial settlements, such as at Hispaniola in 1493. The Treaty of Tordesillas in 1494 formally divided the two Iberian nations’ spheres of influence on either side of a north–south meridian drawn to the west of Portugal’s Cape Verde Islands. Land to the west, namely America except for Brazil, was for Spain; that to the east was for Portugal.

The motivation for these early voyages of exploration has long been summed up in the alliterative phrase, ‘gold, god, and glory’. From the early activities under Henry the Navigator, European Christians, particularly Catholics, were convinced that by spreading the word of their God they were doing his work. Later, souls ‘saved’ were perhaps some compensation for those lost to the spread of Protestantism in Europe. The pursuit of glory was significant in the Age of Discovery and remained a motivation throughout the history of exploration: who now knows the name of the third man atop Everest, despite his effort and personal achievement being identical to Tenzing and Hillary? (It was Jürg Marmet (born 1927) of the 1956 Swiss Mount Everest and Lhotse Expedition.) Gold itself was a key driver in the early European penetrations of South America as in the Spanish Conquest of Mexico and the Aztec Empire under Hernán Cortés (1485–1547) from 1519 to 1521. Perhaps the Aztec leader, Moctezuma (Montezuma II, Motecuhzoma Xocoyotzin, *c.* 1466–1520), had been foolish to have given the Spaniards extravagant gifts in gold when he first encountered them. In a similar vein, the second expedition to Peru in 1526 of Francisco Pizarro (1475–1541) came across gold and during the actual Conquest of Peru in 1532; famously the ransom for the Inca, Atahualpa (*c.* 1502–33) involved the filling of a large room with gold and silver artifacts.

Early Circumnavigations

Often, usually, trade goods other than gold were the reasons why voyages of exploration/exploitation were sponsored. The famous circumnavigation of Portuguese explorer Ferdinand Magellan (Fernão de Magalhães, 1480–1521) of 1519–21 at the service of Spain was actually another attempt to find a westward route to the

Spice Islands (Malaku, formerly called the Moluccas in what is now Indonesia). Magellan himself did not make it all the way round – he was killed in the Philippines in 1521 – and only one of the five ships that had left Seville in 1519 returned there, under Juan Sebastián Elcano (1476–1526). Magellan discovered the route from the Atlantic to the Pacific between the South American mainland and Tierra de Fuego, now called the Strait of Magellan rather than the name Magellan gave it, Estreito de Todos los Santos. Further, while Magellan cannot be said to have discovered the Pacific Ocean in any sense, his party were the first explorers to have entered it from the east and the name of the ocean is that which Magellan gave it. However, despite its considerable achievements regarding exploration, this had been a trading voyage. Magellan had not set out to circumnavigate the globe and it was probably Elcano who actually took the decision to push on from the Spice Islands ever westwards via the Indian Ocean in *Victoria*, rather than returning the way they had come. The other surviving ship, *Trinidad*, attempted to return via the Pacific but was captured by the Portuguese and was wrecked in a storm. *Victoria* reached Seville with just 18 survivors, but with 26 tons of spices, mainly cloves (actually worth more than their weight in gold) and cinnamon. The expedition turned in a small profit.

The English sailor, Francis Drake (1540–96), made the second circumnavigation from 1577 to 1580, a venture described as a voyage of discovery, but there was also a profit motive, this time from plunder of Spanish treasure ships in the New World. Drake also benefited from the 'glory' aspect of exploration, being knighted and becoming a popular hero. His ventures against the Spanish were part of a confrontation that brought England and Spain to war. After the defeat of the Spanish Armada in 1588 in which Drake played a key part, Spanish maritime power waned and nations from Northern Europe began to play a greater role in exploration and the trading to which it was linked.

Merchant Companies

Some of the activity was not by states themselves, but by chartered merchant trading companies such as those established in the Netherlands, France, Brandenburg, and Denmark. One of the most significant was The Company of Merchants of London Trading into the East Indies, known just as the East India Company (EIC), founded in 1600. The EIC became a massive trading organization, which also administered the subcontinent of India, before being liquidated in 1858. EIC activities by the end of the seventeenth century had become set as the routine (if never banal, given the dangers) transportation from the Orient to Europe of spices and other comestibles such as tea (the famous tea clipper, *Cutty Sark*, was an EIC ship). However, in its early period, EIC voyages were scoping ventures, searching out possibilities for trade, and might be seen as exploration. The English had made a reconnaissance expedition to the East from 1591 led by James Lancaster (*c.* 1554–1618). Lancaster went on to command the first voyage of the new EIC from 1601 to 1603 and was knighted after he brought it to a successful conclusion. Like Sir Henry Middleton (fl. 1580–1606) who commanded the second EIC voyage in 1604–06, Lancaster left a record of his exploits and, like much literature on early exploration, these accounts have been republished by the Hakluyt Society.

The trading companies had, on paper, mighty powers. Thus, the Dutch Vereenigde Oostindische Compagnie (VOC, founded 1602) was entitled to annex land in foreign countries, build forts and settlements, maintain an army and navy, declare war, and make treaties with foreign rulers. The companies usually co operated with local rulers about their land holdings given the difficulty of protecting these small enclaves; the main Dutch base in the east at Bantam in Java was held under treaty with the sultan from 1596. Such arrangements were unsatisfactory given the constant need for collaboration with local authorities; better that land be directly controlled. This gave an impetus to further colonization, a process already established by Spain and Portugal. Some of the seventeenth century colonies actually belonged to commercial companies operating under charter, including in what later became the United States. Company colonies were not a success. For example, while Virginia was at first a colony under the control of the Virginia Company (chartered in 1606), ownership was transferred to the crown in 1625 because of problems with its sharply differentiated society, high mortality rates, and a massacre by indigenous people in 1622. Colonialism later, with just a few exceptions, became again a state venture.

Colonialism and Exploration

Colonialism was and had always been associated with exploration given that colonialists need to investigate their actual or potential territories. There is a change in emphasis here from what might be termed exploration as encounter to exploration in detail, from claiming territory now known to exist to getting to know the territory, the better to administer and exploit it.

One of the first examples is that of Hernando de Soto (*c.* 1496–1542), a Spaniard who had been with Pizarro in the Conquest of Peru and had become wealthy on the plunder therefrom. In 1539, he led an expedition to what is now the south of the United States seeking, as always, a passage to the Orient and gold but also with the aim of establishing colonial settlements. His party traveled through areas, including present day Florida, Georgia, the Carolinas, Tennessee, Mississippi, Alabama, and Texas, and they may well have been the first Europeans to see the Mississippi. de Soto, who died on the exped ition, failed in all his objectives. His surviving men limped back into Spanish Mexico in 1642 in penury, without gold, in poor health, and wearing only animal skins. However, the records of the exploration are in valuable scientifically as they detail aboriginal culture prior to contact with Europeans other than their own party, ironically societies which they helped in large part to devastate, not through conquest but rather the trans mission of diseases to the aboriginal peoples against which they had no immunity. Finally, de Soto's explor ations were the basis for the Spanish claim to large parts of what is now the southern USA.

In similar vein was the expedition led by Meriwether Lewis (1774–1809) and William Clark (1770–1838) of 1804–06. In 1803, the recently established USA had ac quired from France through the Louisiana Purchase over 2 million square kilometers of territory west of the Mississippi. President Thomas Jefferson set up this ex pedition party, the Corps of Discovery, with the principle aim of establishing practicable water communication across the continent for the purposes of commerce, with subsidiary aims of studying local people, botany, geology, etc. Starting in Illinois in May 1804, the party reached the Pacific Ocean and, after overwintering at Fort Clatsop in Oregon, began the return journey in March 1806 reaching St Louis in September carrying significant in formation, including scientific samples and maps re garding the geography of the new territory of the United States. The Lewis and Clark expedition helped open up the new land for the fur trade initially and later more extensive settlement and also strengthened the US claims to further areas beyond the Louisiana Purchase, namely Oregon Territory.

A less well known but equally valid example of ex ploration associated with the acquisition of territory re lates to the British occupation of the Falkland Islands from 1833. In that year, after a complex history of settlement and ownership, the British navy established authority over the islands. There was a need to create a more substantial presence than the handful of people supported by naval rations who lived in the one settle ment there in the 1830s. The Colonial Land and Emi gration Commissioners were instructed to report on economic prospects for the Falklands and suggestions included use as a penal settlement, a refreshment station for ships rounding South America, a military base, and an agricultural colony, suitable, it seemed, for a small body of people of frugal and industrious habits – Scottish is landers were suggested. A new lieutenant governor, Richard Moody (1813–87), was sent down replacing the last of the naval officers in charge and tasked to explore the area under his command. In 1842, within 3 months of his arrival, Moody presented an extensive document of 15 000 words of detailed topographical and geological description including a land classification. This was ex ploration with a focus on economic opportunities: pas toralism, hunting and fishing, and growing vegetables. He identified plenty of building stone, peat for fuel, and kelp for manure near the coasts.

A generation later when the British, at first through the Hudson's Bay Company, were settling Vancouver Island off the west coast of Canada, people there went to explore, to see what there was. It is perhaps appropriate that this ex ample of exploration in detail was carried out in a place named for George Vancouver (1757–98), whose voyages in this area were exploration as encounter. The first private settler, W. Colquhoun Grant (fl. 1820–57), published in the *Journal of the Royal Geographical Society* in 1857 a lengthy account of the island and its potential, including timber and coal supplies, harbors and fisheries, and prospective agricultural land, also sections on the native peoples; all 'the results of personal observation', one of the sections being a 'trip round the island'. Grant looked to Vancouver Island 'being settled by bodies of independent freeholders accustomed to rely on themselves for support'. Within 3 years *Facts and Figures Relating to Vancouver Island and British Columbia Showing What to Expect and How to Get There* by the island's surveyor general, J. Despard Pemberton (1821–93) had appeared to further encourage such migrants.

Exploration and the Search for Knowledge

By the late nineteenth century one can begin to recog nize an alternative trend to the type of economic ex ploration epitomized by The Corps of Discovery, as exploration entered what might be termed its search for knowledge phase, which itself slipped into the Mallory tradition of the twentieth century.

What might seem to be a parallel to Lewis and Clark is the 1860–61 Australian expedition of Burke and Wills. However, there are significant distinctions. One was that this expedition was organized not by the government but privately by the Royal Society of Victoria. Also, in con trast to Lewis and Clark, the aim here was somewhat nebulous, being principally to cross the country from Melbourne to the north coast. Robert Burke (1821–61) and John Wills (1834–61) were not particularly experi enced or equipped with the skills they would need for such a venture and on the return leg from the Gulf of Carpenteria both leaders and all but one other member of the main expedition died a slow but according to Wills' last journal entry a not particularly unpleasant death, the ngardu they found to eat providing substance but insufficient nourishment. The sole survivor, John King (1838–72), was rescued by a relief party. The expedition certainly saw more white Australian penetration into the interior of the continent than hitherto, but its utility to science and geography was largely limited to a negative, showing that there was no inland sea. The distinctions between Lewis and Clark and Burke and Wills are important. The first was a state sponsored expedition tasked with scientific but also utilitarian purposes; it was exploration to practical ends in a tradition stretching back centuries. The latter seemed to have been more a journey for its own sake and it was a society venture.

At this period geographical societies became more involved with exploration. Particularly notable in this regard is the British RGS. One of the recipients of a gold medal for exploration from the RGS was the polymath Francis Galton (1822–1911) who wrote *Narrative of an Explorer in Tropical South Africa* in 1853. The motivation for his expedition he had made clear in an article in the *Journal of the Royal Geographical Society* in the first paragraph of which he wrote "urged by an excessive fondness for a wild life, I determined to travel for a second time in Africa": exploration as personal fulfillment. The RGS is most known at this period for the sponsorship of the search for the source of the Nile by John Hanning Speke (1827–64) in 1860–63, which stimulated a bitter dispute between Speke and his former traveling companion, Richard Burton (1821–90). The day Speke was due to have a public debate with Burton, he was killed in a hunting accident, with some suggestion that it was suicide. Speke's travels around Lake Victoria did have some impact on the later scramble for Africa. However, searching for where the Nile starts was not of itself of direct economic utility, it was not the equivalent of looking for potential trade routes to the Spice Islands or for the Northwest Passage. Not all late Victorian African exploration was for intellectual curiosity or to satisfy personal vanity. Indeed, David Livingstone (1813–73) went to Africa for missionary work and to campaign for an end to illegal slavery, but was also involved in seeking for routes across the continent to help open it up for trade and other outside influences. This particularly relates to his Zambezi Expedition of 1858–64, which was sponsored by the British government and took place after he had left the London Missionary Society. However, it is hard not to see source of the Nile type exploration as other than the pastime of a then wealthy and arrogantly confident nation, which had profited from less altruistic and more profitable foreign journeys in earlier eras, including those associated with the slave trade. This imperial grandeur perhaps died with Scott (1913) who wrote "for my own sake I do not regret this journey, which has shown that Englishmen can endure hardships, help one another, and meet death with as great a fortitude as ever in the past;" an heroic end which meant he did not come home beaten by a Norwegian, a man often scorned by Britons because he ate his sled dogs. However, 'Englishmen' or at least those descended from Britons like Edmund Hillary were to have some success in exploration's Mallory tradition, though of course Mallory himself (and Irvine) also had heroic deaths engaged in the pursuit of glory.

Postcolonialism

Early literature on exploration was dominated by the reports of the explorers themselves, including the logs of navigators (many now available in modern editions through the offices of the Hakluyt Society, named for the English travel writer Richard Hakluyt (*c.* 1552–1613)). Scholars would compile collections of such accounts, perhaps adding commentaries. *Purchas, His Pilgrimes*, published from 1613 to 1625 by the clergyman Samuel Purchas (*c.* 1575–1626), who did not travel himself, is an example. Such works celebrated exploration, as did the later imperial narratives of adventurers such as Henry Morton Stanley (1841–1904). Though assuring the reader of his *How I Found Livingstone* that he 'had met in the United States black men, whom I was proud to call friends', he was nonetheless prepared to have recalcitrant African members of his party to "flogged with 'the Great Master's' donkey whip." It is easy to render oneself appalled by the arrogance and cruelty of a man like Stanley looking back from another era, but certainly literature has since set such attitudes firmly into a postcolonial critique. Modern writing such as that by Miles Ogborn considers rather the operation of the exploration and subsequent colonization process, in his case with a focus on the EIC.

The change to the world brought about by exploration has been immense and can never undone, despite modern recognition of the negative impacts it had upon people whose lands were explored and often subsequently colonized. However, there is at last a recognition that the take on the world of "the Great Master" need not be the last word through the reversion of many places to their pre exploration, pre colonial names, including Everest as Chomolungma; New Zealand as Aotearoa, Easter Island as Rapa Nui, and the Queen Charlotte Islands as Haida Gwaii. George Dixon (*c.* 1755–1800) named these islands for his ship, itself named for the wife of the British king, George III. Such tribute presumably meant little to the local people, the Haida. The original and now restored name Haida Gwaii means 'our land' or 'islands of the people', a people who were reduced in number by introduced disease from around 12 000 upon European 'discovery' to about 700 by 1900.

See also: Imperialistic Geographies; Trade, Transport and Communications, Historical Geographies of.

Further Reading

Ambrose, S. E. (1996). *Undaunted Courage: Meriwether Lewis, Thomas Jefferson, and the Opening of the American West*. New York: Simon and Schuster.
Amundsen, R. (1927). *My Life as an Explorer*. London: Heinemann.
Bonyhady, T. (1991). *Burke and Wills: From Melbourne to Myth*. Balmain, NSW: David Ell Press.
Cameron, I. (1974). *Magellan and the First Circumnavigation of the World*. London: Wiedenfeld and Nicholson.
Dreyer, E. L. (2007). *Zheng He: China and the Oceans in the Early Ming Dynasty, 1405 1433*. New York: Pearson Longman.
Fernández Armesto, F. (2006). *Pathfinders: A Global History of Exploration*. Oxford: Oxford University Press.
Galloway, P. (1997). *The Hernando de Soto Expedition: History, Historiography, and "Discovery" in the Southeast*. London: University of Nebraska Press.
Hordern, N. (1971). *God, Gold and Glory*. London: Aldus Books.
Markham, C. R. (ed.) (1877). *The Voyages of Sir James Lancaster, Kt., to the East Indies, with Abstracts of Journals of Voyages to the East Indies, during the Seventeenth Century, Preserved in the India Office. And the Voyage of Captain John Knight (1606), to Seek the North West Passage*. London: Hakluyt Society.
Ogborn, M. (2007). *Indian Ink: Script and Print in the Making of the English East India Company*. Chicago, IL: University of Chicago Press.
Riffenburgh, B. (1993). *The Myth of the Explorer: The Press, Sensationalism, and Geographical Discovery*. London: Belhaven Press and Scott Polar Research Institute.
Scott, R. F. (1913). *Scott's Last Expedition. Vol.1. Being the Journals of Captain R.F. Scott, R.N., C.V.O. Vol. 2. Being the Reports of the Journey and Scientific Work Undertaken by Dr E.A. Wilson and the Surviving Members of the Expedition*. London: Smith, Elder.
Speke, J. H. (1863). *Journal of the Discovery of the Source of the Nile*. Edinburgh: W. Blackwood.
Struzik, E. (1991). *Northwest Passage: The Quest for an Arctic Route to the East*. Toronto: Key Porter Books.
Wallis, J. P. R. (ed.) (1956). *The Zambezi Expedition of David Livingstone, 1858 1863*. London: Chatto and Windus.
Williams, G. (ed.) (2004). *Captain Cook: Explorations and Reassessments*. Woodbridge: Boydell Press.

Relevant Websites

http://www.win.tue.nl/~engels/discovery/
Discoverer's Web: this is a website managed by Andre Engels of the Netherlands which has itself extensive collections on exploration and many links to material on discovery and exploration in all areas of the world.
http://www.hakluyt.com
Hakluyt Society: as its website states, "the Hakluyt Society seeks to advance knowledge and education by the publication of scholarly editions of primary records of voyages, travels and other geographical material." It has been publishing books on exploration since 1847. The Society is named after Richard Hakluyt (1552 1616), who published collections of material on English exploration.
http://www.rgs.org
Royal Geographical Society: housed in London, it holds an extensive archive of exploration material especially of the nineteenth and twentieth centuries. The holdings are open for public consultation.
http://www.spri.cam.ac.uk
Scott Polar Research Institute: named for Captain R. F. Scott, it is part of the University of Cambridge. It is a center for research into polar regions and has extensive library and on line collections of material.

Exploratory Spatial Data Analysis

S. Dall'erba, University of Arizona, Tucson, AZ, USA

Glossary

LISA Statistics Indication of the extent to which there is a significant spatial clustering of similar values around a specific observation.

Moran's *I* Statistics that measure global spatial autocorrelation. It corresponds to the degree of linear association between the value of a variable at one location and the spatially weighted average of neighboring values.

Moran Scatterplot Graphical display of the value of a variable at one observation (on the horizontal axis) against the standardized spatial weighted average (average of the neighbors' value) on the vertical axis.

Spatial Autocorrelation Correlation between the value of an observation and the value of neighboring observations.

Spatial Heterogeneity Variability or difference between the values of observations over space. This variation may follow specific geographical patterns such as east and west, or north and south.

Spatial Weight Matrix Matrix whose cells display the degree of connection between all the observations of our sample.

Introduction

Exploratory spatial data analysis (ESDA) has greatly benefited from the integration of spatial statistical methods to the geographic information science (GIS) environment. A substantial collection of spatial data analysis software is currently available to the practitioner, which should sustain the level of success this field has experienced over the last few years. ESDA allows considering the spatial context in which most geographically referenced events take place. These events cannot be viewed as disjoint and spatially independent, otherwise conclusions about the data and, more importantly, theoretical considerations regarding the social processes that occur across space would be overlooked.

Traditionally, there are two spatial effects that ESDA is able to highlight. They are called spatial autocorrelation (or spatial dependence) and spatial heterogeneity. Spatial autocorrelation can be positive or negative. Positive spatial autocorrelation refers to the idea that nearby observations tend to display similar attributes. It refers to a geographic concentration, similarity, or clustering of a particular phenomenon. Spatial autocorrelation arises from different factors. As stated in Tobler's 'first law of geography', nearby locations are more likely to have more in common than locations that are more distant. Spatial autocorrelation may also come from the fact that the data are affected by processes spread over different locations and/or from a variety of measurement problems, such as a mismatching between the administrative boundaries used to organize the data and the actual boundaries of the processes under study. Spatial heterogeneity means that the value of observations is not homogeneous over space. This variation may follow specific geographical patterns, such as east and west, or north and south, also called spatial regimes or spatial clubs. Both spatial effects are related since the same influences captured by spatial autocorrelation can account for spatial heterogeneity.

ESDA is a step by step process which often starts with aspatial tools, such as choropleth maps and box plots that display the distribution of a variable and identify outliers. However, both tools are considered aspatial because they do not pay attention to the spatial linkages that influence the distribution of a variable. In order to fill this gap, we need to turn to the definition of spatial interactions via a spatial weight matrix and then to associated spatial ESDA tools. Practitioners often choose first a Moran's *I* statistics for global spatial autocorrelation. It reveals the degree of linear association between the value of a variable at one location and the spatially weighted average of neighboring values. A more detailed picture of the spatial distribution is then offered by Moran's scatterplot, local indicators of spatial association (LISA), and Getis–Ord statistics which measure local spatial autocorrelation and, eventually, reveal spatial heterogeneity. These last three tools compute, for 'each' observation, the degree of spatial clustering of similar values around that observation. While all these steps are performed and developed below on area data, it is important to note that ESDA can be applied to other types of observations such as point data, flow data, etc.

Nonspatial Tools

Choropleth Map

The first step of an ESDA is to perform a so called choropleth map, which simply displays the distribution of the variables under study. For the sake of clarity, we use the example of the distribution of per capita gross domestic product (GDP) across the 145 regions of 12 countries of the European Union in 1989. Their distribution, relative to the European average, is shown in **Figure 1**.

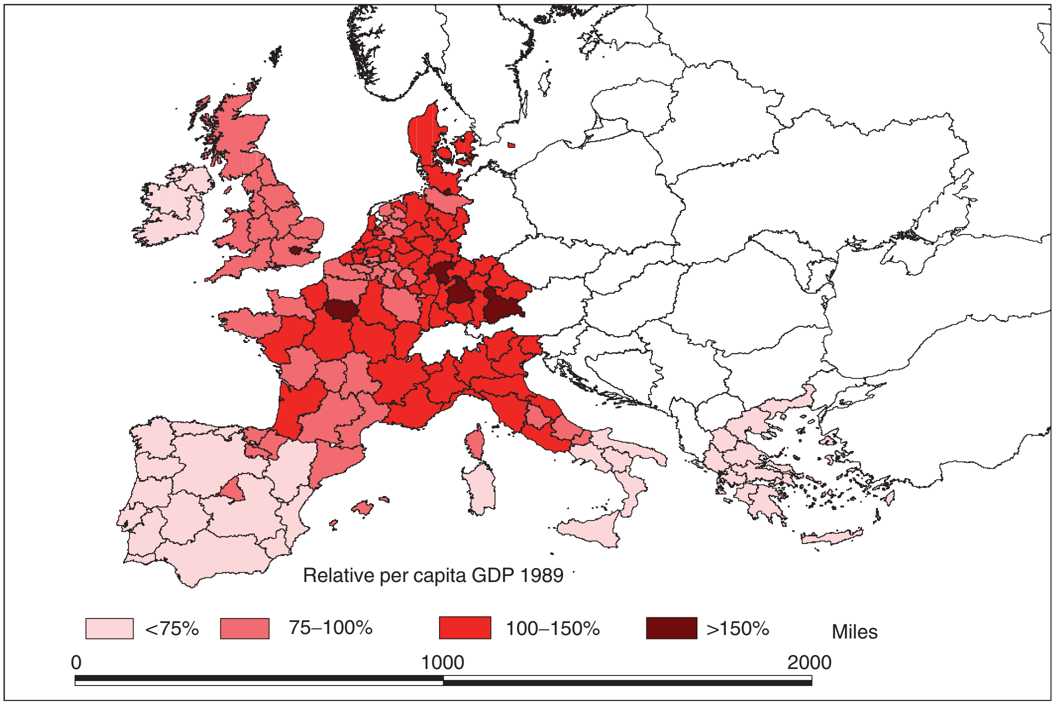

Figure 1 Choropleth map: relative distribution of per capita GDP across the regions of Europe in 1989 (in 1990 Euro prices). This figure has been realized using ArcView GIS 3.2 (Esri). Reprinted from Dall'erba, S. (2005). Distribution of regional income and regional funds in Europe 1989 1999: An exploratory spatial data analysis. *Annals of Regional Science* 39, 121 148.

Europe is marked by a clear core–periphery pattern, with the core (the darker color) composed of the richest regions, whereas the peripheral regions (Spain, Portugal, Greece, and the southern part of Italy) are the poorest ones (per capita GDP is below 75% of the average). Note that the choice of four categories is simply due to the political concerns typical of the case under study. Intervals of the categories could also have been chosen to be equal or according to the quantiles of the distribution of our variable. While a choropleth map allows visual izing the distribution of a variable, it does not tell us much about the persistence over time of this distribution nor the significant presence of the spatial effects we introduced earlier. Before we go any further, we need to say a few words on the spatial weight matrices that will be used in the spatial analysis.

Visualizing Extreme Values

Like the choropleth map above, the box plot is another basic tool that depicts the aspatial distribution of a variable. A box plot displays the median, first, and third quartile of the distribution (respectively the 50%, 25%, and 75% points in the cumulative distribution) as well as outliers. An observation is called outlier when its value is above or below a given multiple of the difference be tween the first and third quartile. The standard multiples are 1.5 or 3. Here is the corresponding formula with the case of 1.5 as a multiple: outlier $< Q_1 - 1.5*(Q_3 - Q_1)$ $> Q_3 + 1.5*(Q_3 - Q_1)$, where Q_1 and Q_3 are respect ively first and third quartile of the distribution. **Figure 2** displays an example with the same variable as above and a multiple of 1.5.

The bar in the middle of the dark area corresponds to the median while the upper (resp. lower) part of the dark area is the third (resp. first) quartile of the distribution. The thin line on the upper part of the figure is called the hinge, here corresponding to the default criteria of 1.5 times the difference between the first and third quartile. The figure indicates that five regions display some ex treme values of per capita GDP. They are the outliers: Hamburg and Damrstadt (Germany), Ile de France (France), Brussels (Belgium), and Luxembourg. They all belong to the core regions of Europe.

Visualizing outliers can be done by linking the box plot to a box map or by representing the distribution according to the standard deviation of the variable. In the latter case, outliers are observations outside of the standard $[m - 2\sigma, m + 2\sigma]$ range, where m is the mean and σ the standard deviation.

Spatial Weight Matrix

A spatial weight matrix is a matrix whose cells display the degree of connection between all the observations of our sample. Traditionally, connections are based on pure

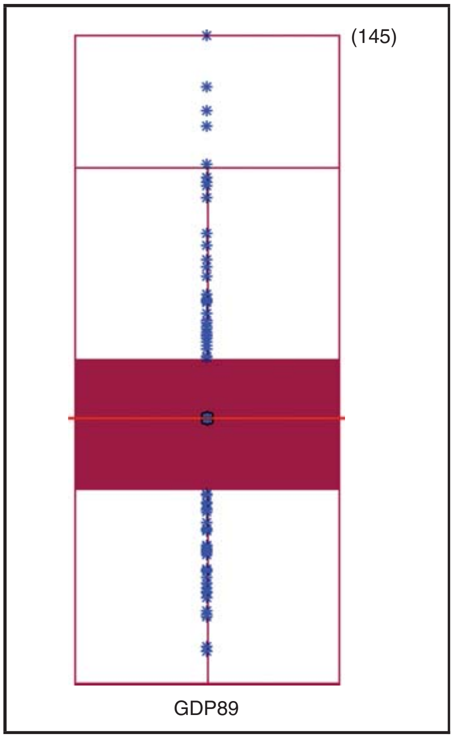

Figure 2 Box plot of the distribution of European regional per capita GDP in 1989. This figure has been created in Geoda.

geographical distance as the exogenous character of distance is unambiguous. This aspect becomes necessary when, at a later stage, spatial econometric regressions are run. (Note that in the case of per capita GDP, the true connections between regions are not based on dis tance but on factors such as interregional migration, trade flows, knowledge spillovers, etc. However, such data do not exist for the European regions yet.) We base our spatial weight matrices on the great circle distance for dependence to be considered in any direction. More precisely, we base our weight matrices on the $k = 10$, 15 nearest neighbors. The form of the spatial weight matrix is the following:

$$\begin{cases} w_{ij}(k) = 0 \text{ if } \ i = j \\ w_{ij}(k) = 1 \text{ if } \ d_{ij} \leq D_i(k) \quad \text{for } k = 10, 15 \\ w_{ij}(k) = 0 \text{ if } \ d_{ij} > D_i(k) \end{cases} \qquad [1]$$

where d_{ij} is the great circle distance between centroids of regions i and j. $D_i(k)$ is the critical cutoff distance defined for each regions i, above which interactions are assumed to be negligible. In other words, $D_i(k)$ is the k^{th} order smallest distance between regions i and j such that each region i has exactly k neighbors. In the spatial weight matrix above, the weights (1 or 0) capture the spatial relationships between geographic observations. It is worth mentioning that in the European context, the minimum number of nearest neighbors that guarantees international connections between all regions is $k = 7$. Different kinds of weight matrices can be built (great circle distance with a cutoff, contiguity matrices, share of common borders, etc.), and their choice is always some what arbitrary. In the case of a k nearest neighbors weight matrix, we recommend limiting the number of neighbors to 10% of the total number of observations. It is common practice to use the inverse of the distance $(1/d)$ or of the squared distance $(1/d^2)$ in order to reflect the greater degree of connection between closer observations. An other common practice consists in normalizing the out side influence upon each region by row standardizing each matrix (i.e., the row elements of the matrix are divided by the sum of the elements by row) so that each row sum is one. Row standardization means that it is relative and not absolute distance that matters.

Global Spatial Autocorrelation

Join Count Statistics

The simplest measures of spatial autocorrelation are for qualitative (binary) variables that take the value 0 or 1. These measures are called join count statistics because they measure the number of times a join corresponds to a similar or dissimilar value in the neighboring location. Because areas with observations 1 (resp. 0) are refereed as black (resp. white), there are three types of join counts:

$$\begin{aligned} \mathrm{BB} &= \tfrac{1}{2}\sum_i \sum_j w_{ij} x_i x_j \\ \mathrm{BW} &= \tfrac{1}{2}\sum_i \sum_j w_{ij} \left(x_i x_j\right)^2 \\ \mathrm{WW} &= \tfrac{1}{2}\sum_i \sum_j w_{ij} (1 \quad x_i)(1 \quad x_j) \end{aligned} \qquad [2]$$

There is positive spatial autocorrelation if the BB or WW joins dominate; there is negative spatial auto correlation if the BW joins are more numerous. Inference is based on the approaches which are listed below for Moran's I and Geary's c.

Moran's *I* and Geary's *c*

Moran's I and Geary's c are the best known statistics to measure global spatial autocorrelation of a quantitative variable. They correspond to the degree of linear asso ciation between the value of a variable at one location and the spatially weighted average of neighboring values. Formally, the Moran's I is given by

$$I = \frac{N}{S_0} \frac{\sum_i \sum_j w_{ij}(x_i \quad x)(x_j \quad x)}{\sum_i (x_i \quad x)^2} \qquad [3]$$

where N is the number of observations, w_{ij} is the degree of connection between the spatial units i and j, and x_i is the variable of interest in region i and

$$S_0 = \sum_i \sum_j w_{ij}$$

that is, the sum of all the weights. If the spatial weight matrix is standardized, $N = S_0$, the Geary's c is given by

$$c = \frac{N - 1}{2S_0} \frac{\sum_i \sum_j w_{ij}(x_i - x_j)^2}{\sum_i (x_i - \bar{x})^2} \qquad [4]$$

with the same notation as above.

Moran's I values larger (smaller) than the expected value $E(I) = -1/(n-1)$ indicate positive (negative) spatial autocorrelation. In our case, with $k = 10$ neighbors, $I = 0.481$ (calculations have been performed in SpaceStat) while $E(I) = -1/144 = -0.00694$. This indicates the presence of positive spatial autocorrelation, that is, in our case study, wealthy regions tend to be geographically clustered and poor regions tend to be close to each other as well. The distribution of the regional per capita GDP in Europe is therefore certainly not random. This result is confirmed by the Geary's c coefficient (0.511) which is smaller than its expected value, 1. Had the c coefficient been larger than 1, we would have concluded that a negative spatial autocorrelation exists. This would have suggested a chess board pattern of spatial dissimilarity that is rarely found in spatially referenced data. It is good practice to perform both statistics with various weight matrices and over several years (if available) to confirm your results.

Significance tests for the I and c statistics are often based on a standardized z value whose mean and standard deviation can be calculated according to different approaches: normal distribution, the randomization assumption, or the permutation approach. For instance, the results above were calculated using a permutation approach with 10 000 permutations. It means that 10 000 resampled datasets were automatically created for which the I (or c) statistics are computed. The value obtained for the actual dataset has then been compared to the empirical distribution obtained from these resampled datasets.

While the previous spatial statistics can detect global spatial autocorrelation, they are not able to identify local patterns of spatial association, such as local spatial clusters or local spatial outliers. In addition, positive global spatial autocorrelation does not indicate whether one is dealing with a cluster of high or low values. The identification of observations that belong to a clustering of high (low) values is based on the results of a Moran scatterplot.

Local Spatial Autocorrelation

Moran Scatterplot

The idea of the Moran scatterplot is to display the value of an observation (on the horizontal axis) against the standardized spatial weighted average (average of the neighbors' value, also called spatial lag) on the vertical axis. Expressing the variables in standardized form (i.e., with mean 0 and standard deviation equal to 1) allows the assessment of both the global spatial association, since the slope of the line is the Moran's I coefficient, and local spatial association (the quadrant in the scatterplot). As shown in **Figure 3** the Moran scatterplot is therefore divided into four different quadrants corresponding to the four types of local spatial association between an observation and its neighbors:

1. Quadrant I (on the top right corner) displays the observations with a high value (above average) surrounded by observations with a high value (above average). This quadrant is usually noted as HH. In our case study, this would correspond to rich regions surrounded by rich regions.
2. Quadrant II (on the top left corner) shows observations with a low value surrounded by observations with high values. This quadrant is usually noted as LH. This would correspond to poor regions surrounded by rich regions.
3. Quadrant III (on the bottom left) displays observations with low values surrounded by observations with low values, and is noted as LL. This corresponds to poor regions surrounded by poor regions.
4. Quadrant IV (on the bottom right) shows observations with high value surrounded by observations with low values. It is noted as LL. This corresponds to rich regions surrounded by poor regions.

Figure 4 displays the Moran scatterplots of regional per capita GDP for 1989, with $k = 10$ nearest neighbors. Regions located in quadrants I and III refer to positive spatial autocorrelation, the spatial clustering of similar values, whereas quadrants II and IV represent negative spatial autocorrelation, the spatial clustering of dissimilar values. Positive global spatial autocorrelation, which was previously detected by the value of Moran's I, is reflected by (1) the positive slope coefficient of the linear regression of the spatial lag value on z, (a slope of one, passing through the origin, would indicate perfect spatial association) and (2) the fact that most of the regions are located in quadrants I and III. Those are regions respectively located in the core and the periphery of regions. Conversely, HL and LH type regions are often regions located at the border between peripheral and core regions. Overall, the results of the Moran scatterplot indicate that the spatial distribution of regional income is more complicated than the simple core–periphery framework previously detected in the choropleth map.

Moran scatterplot is also a useful tool for the detection of spatial heterogeneity. Indeed, from **Figure 4** it is clear that most European regions are either in the HH or in the LL quadrants which means that the distribution

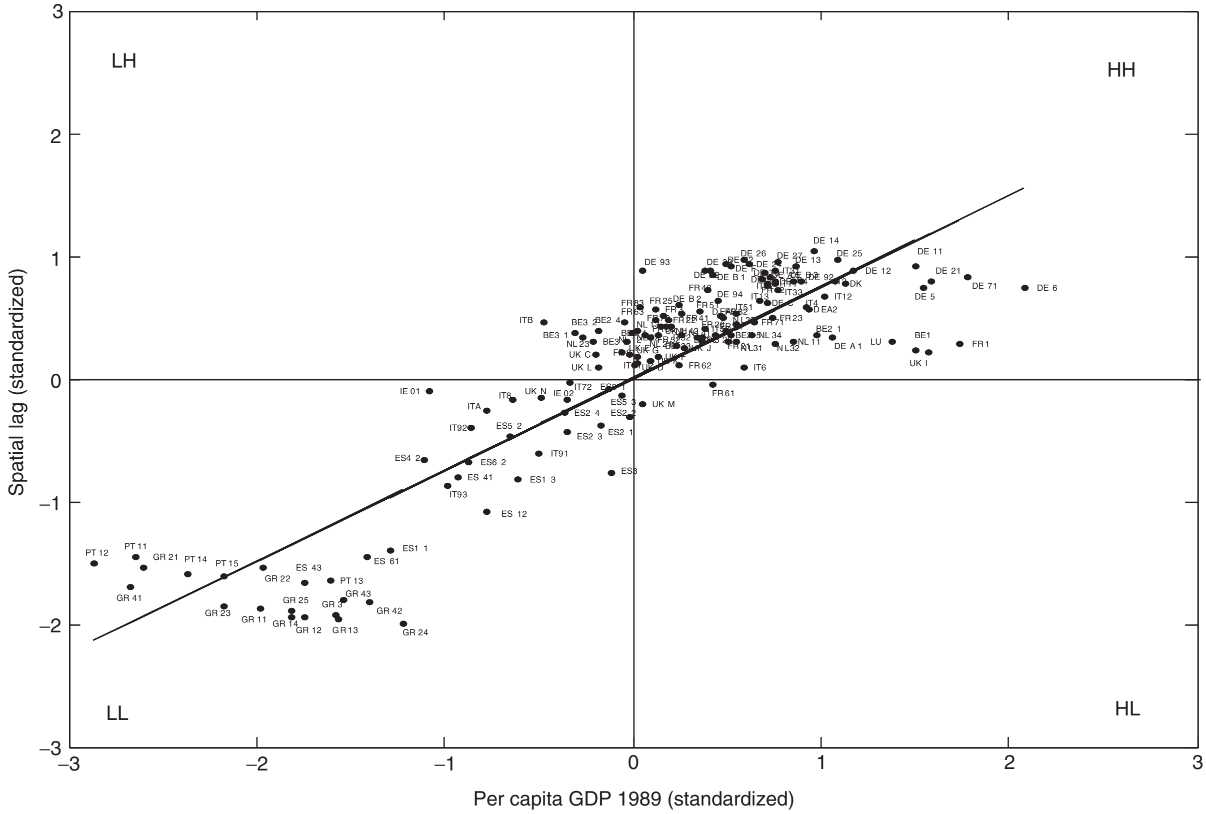

Figure 3 Moran scatterplot of regional per capita GDP in 1989. This figure has been performed in Matlab. Reprinted from Dall'erba, S. (2005). Distribution of regional income and regional funds in Europe 1989 1999: An exploratory spatial data analysis. *Annals of Regional Science* 39, 121 148.

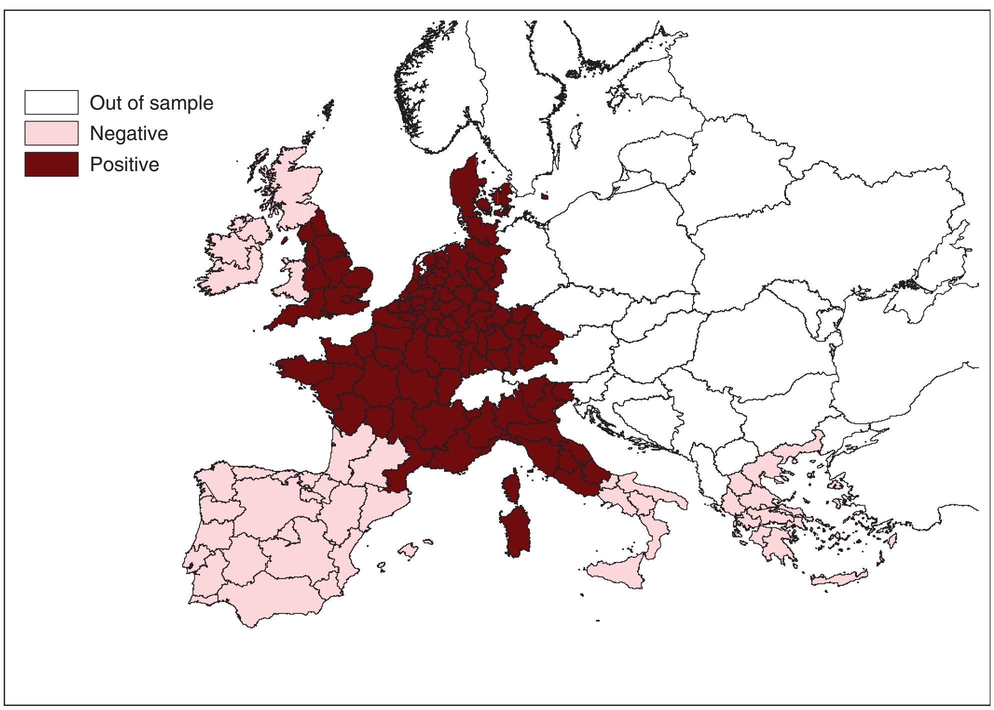

Figure 4 Getis Ord statistics on the per capita GDP in 1989. This figure has been realized using ArcView GIS 3.2 (Esri) and the Getis Ord statistics have been computed in SpaceStat.

of the variable studied (per capita GDP in 1989) is not homogeneous over space. Further tools are used below to detect whether these two spatial clubs are consistent.

Finally, note that **Figure 4** corresponds to a so called univariate Moran scatterplot because we use the same variable on the x and the y axis. A multivariate Moran scatterplot can also be done. It corresponds to a case where the y axis is calculated on another variable (a different year, a different variable or both) than the variable used in the x axis. The case where the two axes are based on different years is referred to as spatial–temporal dependence.

Getis–Ord statistics

The G statistics developed by Getis and Ord provide an alternate measure of global spatial association (G) and observation specific (G_i and G^*) measures of local spatial association. These last statistics detect the presence of local spatial autocorrelation: a positive value of this statistic for observation i indicates a spatial cluster of observations with high value of the variable of interest, whereas a negative value of the G_i or G_i^* statistics indicates a spatial cluster of observations with low value around observation i.

Formally, the G_i statistics can be written as follows:

$$G_i = \frac{\sum_{j \neq i} w_{ij}(d) x_j}{\sum_{j \neq i} x_j} \qquad [5]$$

where $w_{ij}(d)$ are the elements of the matrix whose distance cutoff is d. The matrix used for the calculation of this statistic does not necessarily have to be symmetric and binary. The difference between G_i and G_i^* is that the latter statistics provide a measure of spatial clustering that includes the observation under consideration, while G_i does not. In other words, $j = i$ is included in the sum (in the denominator) in G_i^* statistics.

These statistics imply a two way split of the sample which can be useful for the determination of spatial regimes or spatial clubs. Applied to our case study, the two clubs (rich surrounded by rich and poor surrounded by poor) are represented in **Figure 4**. Spatial heterogeneity, previously detected in the Moran scatterplot, seems to be consistent with the Getis–Ord statistics as well.

Note, however, that we did not pay attention to the significance level of the statistics for **Figure 4**. If we had included the significance level, 68 of the 145 statistics would have been nonsignificant.

Local Indicator of Spatial Association

The Moran scatterplot displayed above does not give any indication on the 'significance' of spatial clustering. We therefore need to calculate LISA statistics for each observation to obtain an indication of the extent to which there is a significant spatial clustering of similar values around a specific observation. LISA statistics, developed by Anselin, are used for the detection of significant local spatial clusters (also called 'hot spots') as well as for the diagnostics of local instability, significant outliers, and spatial regimes. The local Moran's statistics for each region i can be formalized in the following way:

$$I_i = \left(\frac{x_i}{m_0}\right) \sum_j w_{ij} x_j \quad \text{with} \quad m_0 = \sum_i \frac{x_i^{\,2}}{n} \qquad [6]$$

with x_i (x_j) is the observation in region i (j) (each measured as a deviation from the mean value) and the matrix w is row standardized.

The sum of LISAs for all observations is proportional to a global indicator of spatial association. Since we use a row standardized matrix, the average of local Moran statistics is also equal to the global Moran's I statistics.

The results from the application of LISA with $k = 10$ nearest neighbors are displayed in **Figure 5**. The significance level is based on a conditional permutation approach with 10 000 random permutations of the neighboring regions for each observation. The pseudo significance level is 5%.

Regions revealing a significant spatial association of per capita GDP (HH or LL) are mostly those previously detected as core (HH cluster) and peripheral (LL cluster) regions. Indeed, in the figure, 83% of the local statistics that are significant are either HH or LL type, reflecting the global trend of positive spatial association and spatial heterogeneity since the clusters represent different values of per capita GDP. The figure also indicates that some regions display signs of significant negative spatial autocorrelation (LH or HL type). As explained above, those are regions with a level of per capita GDP significantly different than the one of their neighbors, such as Madrid in Spain for instance (HL type).

Finally, it is important to pay attention to the problem of multiple statistical comparisons that the LISA statistics above may present. This problem comes from the fact that the neighborhood sets of two regions may contain common regions. In order to correct for the error on the appropriate significance level due to multiple comparisons, we need to compute a Bonferroni pseudo significance level (= a 5% pseudo significance level in our case). The Bonferroni procedure consists in adjusting downward the alpha level of 'overall' significance (5%) by the number of comparisons ($k = 10$) in order to find the 'individual' significance level (= 0.5% or 5% Bonferroni pseudo significance level).

Software

We seek to provide the reader with a brief review of the software currently available to perform ESDA. We realize this section is doomed to be quickly obsolete given the

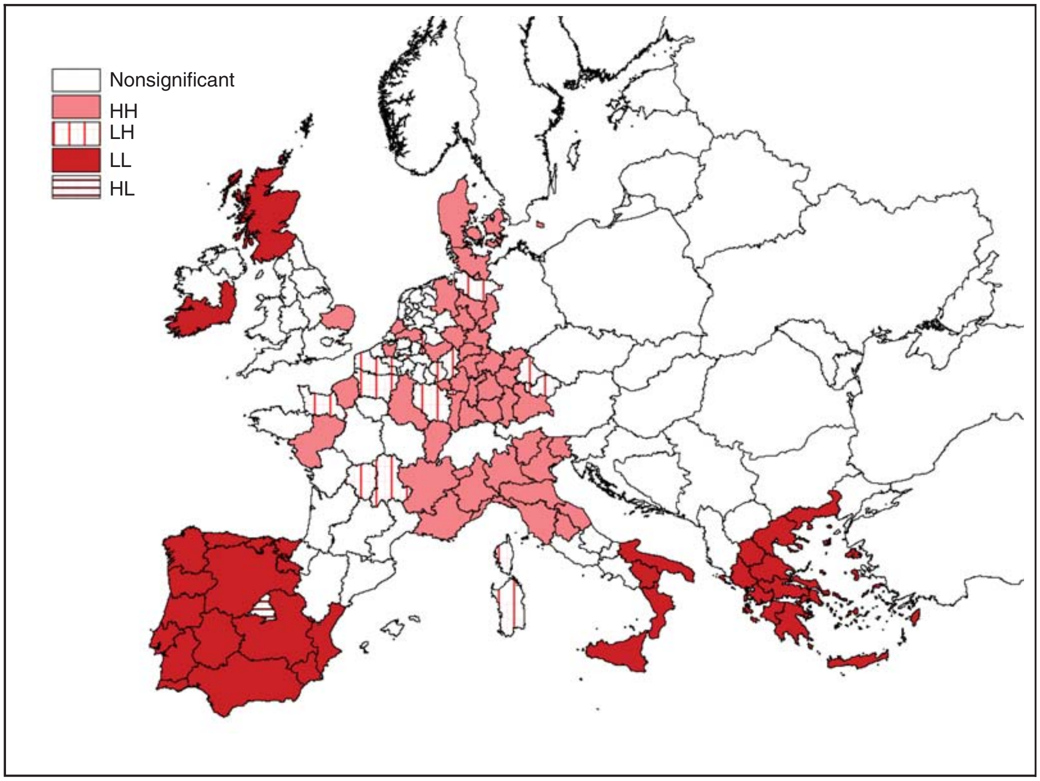

Figure 5 LISA statistics on the per capita GDP in 1989. This figure has been realized using ArcView GIS 3.2 (Esri) and the LISA statistics have been computed in SpaceStat.

speed at which this field evolves. Some of the oldest software includes the 'Spider/Regard' toolboxes which allow 'geographic brushing'; SAGE which adds ESDA capability to the standard ArcInfo GIS environment; and the DynESDA extension for ArcView. ArcGIS Geosta tistical Analyst allows exploring the spatial distribution of data, but not necessarily with all the tools described above. Currently, the most complete software is Geoda and SpaceStat. Despite its old fashioned interface and performance constraints for large datasets, the latter provides very useful capabilities. It requires an ArcView extension for visualization. Geoda is a more recent and user friendly software that allows both ESDA and some (limited) spatial modeling capabilities. CrimeStat is a Windows based software that interfaces with most desk top GIS programs. Primarily devoted to fighting crime, it has been very popular with police departments around the country. ArcGIS Desktop 9.0 Spatial Statistics Geo processing Toolbox Patch contains performance im provements for cluster and outlier analysis, hot spot analysis, and spatial autocorrelation. Other toolboxes, but which require some programming skills, include the 'tmap' function developed in Stata; the 'spdep' and 'DCluster' functions in R; PySpace and STARS (space– time analysis of regional systems) which are written in Python and, finally, the Matlab Arc Mat and econo metric routines.

Looking Ahead

We live in an exciting period when spatial statistics have tremendously benefited from technical and theoretical advances in GI science. It is very likely that these prow esses will continue in the decades to come. However, with the use of spatial statistics becoming more popular across different fields, there is a clear need to tighten the interaction between the methodologies, the computa tional capabilities, and the specificities inherent to each discipline. For instance, many disciplines deal with linkages that are not based on pure geographical distance but, instead, rely on more sophisticated concepts of space (such as economic, social, or political distance). It is not an easy task though since it would require a rethinking of the current statistical methods. Another aspect to be developed focuses on the exploration of patterns of flow data, since reducing transportation costs and more gen erally technological improvements have tremendously increased mobility. In addition, previous ESDA tools do not necessarily allow the user to capture the full extent of a variable distribution as the value in one observation is often compared to the average of the sample or neigh boring observations. With the size of the samples becoming bigger, the chances for a distribution to be multimodal (i.e., displaying signs of multiple clusters) increase. In this case, we recommend complementing an

ESDA with the increasingly popular tools such as density functions, Markov chains, stochastic kernels, and their spatial version. Finally, the outcome of an ESDA will never be more than a detailed description of the distri bution of a variable over space. Its capacity to measure the relationships between several variables is very lim ited. The natural subsequent step is therefore a thorough estimation of the workings of the system under study. This could be done, for example, using spatial econo metric techniques where the interaction between variables and spatial effects are formally tested and measured.

See also: Spatial Autocorrelation; Spatial Clustering, Detection and Analysis of; Statistics, Descriptive; Statistics, Spatial.

Further Reading

Anselin, L. (1995). Local indicator of spatial association LISA. *Geographical Analysis* 27, 93 115.
Anselin, L. (1999). Interactive techniques and exploratory spatial data analysis. In Longley, P. A., Goodchild, M. F., Maguire, D. J. & Rhind, D. W. (eds.) *Geographic Information System: Principles, Techniques, Management and Applications*, pp 251 264. New York: John Wiley.
Anselin, L. (2005). *Exploring Spatial Data with Geoda: A Workbook*. Urbana, IL: University of Illinois at Urbana Champaign.
Anselin, L., Florax, R. G. J. M. and Rey, S. (2004). *Advances in Spatial Econometrics: Methodology, Tools and Applications*. Heidelberg: Springer Verlag.
Cliff, A. and Ord, J. K. (1981). *Spatial Processes, Models and Applications*. London: Pion.
Dall'erba, S. (2005). Distribution of regional income and regional funds in Europe 1989 1999: An exploratory spatial data analysis. *Annals of Regional Science* 39, 121 148.
Fotheringham, A. S., Brunsdon, C. and Charlton, M. (2000). *Quantitative Geography, Perspectives on Spatial Data Analysis*. Thousand Oaks, CA: Sage Publications.
Getis, A. and Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24, 189 206.
Haining, R. (1990). *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge: Cambridge University Press.
Le Gallo, J. (2004). Space time analysis of GDP disparities among European regions: A Markov chains approach. *International Regional Science Review* 27, 138 163.
Le Gallo, J. and Dall'erba, S. (2006). Evaluating the temporal and spatial heterogeneity of the European convergence process, 1980 1999. *Journal of Regional Science* 46, 269 288.
Ord, J. K. and Getis, A. (1995). Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis* 27, 286 305.
Rey, S. J. and Montouri, B. D. (1999). US regional income convergence: A spatial econometric perspective. *Regional Studies* 33, 143 156.
Upton, G. and Fingleton, B. (1985). *Spatial Data Analysis by Example*. New York: Wiley.

Relevant Websites

http://www.csiss.org
Center for Spatially Integrated Social Science.
http://www.esri.com
ESRI: GIS and mapping software, ArcGIS Geostatistical Analyst.
http://www.geoda.uiuc.edu
Geoda.
http://www.icpsr.umich.edu
Inter University Consortium for Political and Social Research: CrimeStat.
http://www.spatial econometrics.com
Matlab spatial toolboxes.
http://www.spatstat.org
Open source software for spatial statistics.
http://regionalanalysislab.org
Regional Analysis Laboratory: STARS.
http://www.terraseer.com
Terraseer, Inc.: SpaceStat.
http://www.shef.ac.uk
The University of Sheffield: SAGE.

Export Processing Zones

F. Wu, Cardiff University, Cardiff, UK

Glossary

Backward Linkage The linkage of the enterprises (in the EPZs) through using local materials and services.

Free Trade Zones It is equivalent to export processing zones. In some cases, in more developed economies, a free trade zone may be an exclusive duty-free trade area without manufacturing activities.

Multi-Fibre Agreement (MFA) The rules of the MFA set textile and clothing quotas through bilateral negotiation, which allowed selective quantitative restrictions when an increase in imports threatened or damaged the industry of the importing country. On 1 January 1995, it was replaced by the WTO Agreement on Textiles and Clothing.

Introduction

According to the World Bank, export processing zones (EPZs) are industrial estates that are fenced in for producing manufactured goods for export. In short, they are trade enclaves that import raw materials, process them, and then export to the world market. The intermediate goods and materials are treated as duty free. EPZs usually enjoy more flexible regulation control and tax concessions. Publicly funded EPZs are developed through government funded infrastructure, although infrastructure may also be developed through development corporations. The goods produced are not allowed to be sold in the domestic market, although some (such as the Dominican Republic's EPZ and Mexico's EPZ) allow a limited amount to be sold in the domestic market, or even free sale to the domestic market (Manaus EPZ of Brazil).

EPZs are a prelude to the wide adoption of liberal trade policies. In the early stage of development, EPZs combine industrial parks and free trade zones. The terms are interchangeable. Similar terms include free zone, special economic zone (SEZ), and maquiladoras. Because they originated from the use of labor by multinationals in a confined zone, the diffusion of technology may be limited. Many EPZs perform well in terms of employment generation and foreign exchange earnings but not in terms of backward linkage. The operation of manufacturing activities may be isolated from domestic producers.

According to the International Labour Organization (ILO), EPZs in total employed 41.9 million workers worldwide, among which Asian EPZs employ 36.8 million, and Central America and Mexico 2.2 million (**Table 1**). China alone employed 30 million workers in its EPZs. However, because Chinese EPZs include SEZs, including in fact whole cities opened up for the world economy, the actual number of workers in strictly defined EPZs may be smaller than the figure cited. Because of rising labor costs, Central America and Mexico have lost employment in recent years to lower cost economies such as China. Many EPZs in Central and Eastern Europe will disappear once their economies join the EU. The geographical distribution of EPZ employment is shown in **Table 2**.

EPZs have become an important export base for developing countries. Major exports from EPZs include electronics, garments, telecommunications equipment, and auto parts. These goods are particularly labor intensive, and their assembly requires repetitive but nonstandard labor input that has not been totally automated. According to the World Trade Organization (WTO), China, Malaysia, and Mexico export telecommunications equipment from their EPZs, and Bangladesh has experienced a significant increase in garments export, partially due to the Multi Fibre Agreement (MFA).

The emergence of EPZs reflects the shift in development strategy from an import substituted one to an export oriented one. EPZs are used to increase export. Often the development of EPZs is associated with trade liberalization. China designated four cities as SEZs and many more specific development zones on a smaller scale, known as economic and technological zones (ETZs). Other zones include high tech development zones and free trade zones.

Table 1 Employment in EPZs in the world estimated by the ILO (2003)

	1975	*1986*	*1995*	*1997*	*2002*
No. of countries with EPZs	25	47	73	93	116
No. of EPZs	79	176	500	845	3000
Employment (millions)				23	43
of which China				18	30
of which other	0.8	1.9		5	13

Note: The table is compiled by the ILO from its various publications. The information on EPZs is summarized from a total of 108 countries. Data from International Labour Office (ILO) (2003). *Employment and Social Policy in Respect of Export Processing Zones (EPZs)*. Geneva: ILO.

Table 2 Geographical distribution of EPZ employment in 2002 03

Region	*No. of EPZs*	*Total employment*
North Africa	16	440 515
Sub-Saharan Africa	50	431 348
Indian Ocean	4[a]	127 509
Middle East	38	691 297
South Asia	48	36 824 231
Southeast Asian	100	
Mexico and Central America	82	2 241 821
South America	134	226 130
Transition economies of Central and Eastern Europe	93	245 619
Pacific	14	13 590
Europe	65	50 830
North America	2 (+366 parks)	330 000
Total	502	41 934 133[b]

[a] Including the whole island of Mauritius.
[b] Including the employment of 30 000 000 in China.
Data from International Labour Office (ILO) (2003). *Employment and Social Policy in Respect of Export Processing Zones (EPZs).* Geneva: ILO.

In recent years, EPZs have developed more in the form of high tech industrial parks, which emphasize technological development rather than low skilled material processing. For example, in Shanghai, Jingqiao EPZ and Waigaoqiao Free Trade Zone are only two of the zones within the vast Pudong new district, which now spearheads high tech industrial development and financial and trade functions. The national level zones in China increased to 54 in 2003. When China joined the WTO in 2001, these zones began to lose their competitive advantages. The government tried to transform the manufacturing EPZs into multifunction zones.

Rationale

The development of EPZs is justified by a number of motivations: to generate employment, to produce foreign exchange earnings, to promote exports, to provide a catalyst effect on local firms about how to export to the world market, to use trade enclaves to diffuse knowledge, know how and management skills to local firms, and to stimulate industrial development in the host country. All these objectives are not without their caveats. For example, the objective of diffusing technology cannot be easily achieved, because EPZs are fenced off from the domestic sector. The gap between local suppliers and the enterprises operating in EPZs may be too large to bridge, and thus it is difficult to diffuse technological development. Job generation is also not necessarily true. Most jobs are limited to low skilled and low income sectors.

EPZs may not be the only way to generate foreign exchange earnings. This effort is not always successful. Such a goal may be better achieved through opening up the whole economy rather than isolated trade enclaves such as EPZs. Firms in EPZs, although they may increase exports, may import more than domestic firms, because they are more familiar with international suppliers. Thus, the increase in exports may not produce net foreign exchange earnings. EPZs promote gross exports, but, measured in terms of net exports, the promotion of exports may not live up to expectations. Because of the poor quality of local suppliers and weak backward linkage, the effect of diffusion of export knowledge may be limited.

Technology transfer is much hoped for in EPZ development. Through the spillover effect, export oriented firms in EPZs may develop backward linkage with local firms, for example, using raw and intermediate materials from the host country. Thus, local firms may initially act as suppliers for foreign firms and then become capable of exporting the processed product to the world market. But foreign managers are more familiar with international suppliers and tend to have a low propensity for using local suppliers. Special regulations may prevent them from accessing local suppliers. For lower rank processing, such as the garments industry, production is labor intensive and low tech. The technology is mature and has been stable for many years. There is very little effect of technology advance and diffusion. EPZs are often used as a way of circumventing trade quotas, besides utilizing cheap labor. For higher rank goods, it is also questionable whether local firms are able to develop the technology to produce these products, because most foreign firms tend to locate their production sections in EPZs while keeping their research and development (R&D) in developed countries. But more recently, along with the transition of EPZs to high tech zones, multinational companies (MNCs) such as Microsoft have begun to develop their R&D facilities in developing countries such as China and India.

Besides the transfer of ‘hard’ technology, it is also hoped that EPZs can help to introduce ‘soft know how’ related to knowledge about selling export goods in the world market. Although EPZs have clear physical boundaries, foreign firms inside EPZs may produce a catalyst effect by demonstrating entrepreneurial skills. This may be very helpful for those countries that are not familiar with international trading. EPZs thus may pro mote trade, not only by foreign firms but also by local firms. Although EPZs are initially developed as enclaves, the host country often hopes to use EPZs to bring know how into the local economy. EPZs are often justified by their catalyst effect through demonstrating the way to export manufactured goods to the world market. But in order to achieve this goal, EPZs should be connected through backward linkage with the domestic sector. This often proves to be difficult, because the local sector may not have the emulation capacity. Thus, the spread of export knowledge across the boundary is not effective. Because of poor infrastructure outside EPZs and some times unstable political situations, such capacity may not exist in developing countries. As a result, EPZs continue to exist as trade enclaves and fail to be integrated with the local economy as a whole.

The diffusion of EPZ know how depends upon overall progress in economic development in the host economy. If EPZs are used as an initial step to open up the whole economy, the establishment of EPZs then may help to diffuse knowledge of the world market. If EPZs are used only as a temporary measure to fence off competition from the world market while producing foreign exchange, the diffusion of EPZs may be limited. The overall eco nomic performance of the host economy also affects the absorption of know how. The most successful diffusion of EPZ know how is in the newly industrialized economies, where the local economy is export oriented. The devel opment of EPZs thus facilitates such a direction.

Labor Issues

Because of the concern over labor issues, the ILO spe cially recommends that the development of EPZs should support both economic and social advancement. The issue of labor is critical for EPZs because export pro cessing is based on the intensive use of labor. EPZs are especially useful when there is a high level of surplus labor, regardless of the number of jobs generated. Because of the existence of unemployment and surplus labor in many developing countries, there is a need to generate jobs. The development of EPZs may be seen as a useful policy. For example, EPZs in the Dominican Republic and Mauritius are helpful, because high unemployment coexists with a significant number of surplus workers. The local economy simply cannot cope with such a scale of job generation.

But the quantity of employment generated by EPZs is limited, compared with the unlimited supply of labor. While one major objective of EPZs is job creation, the actual jobs created by EPZs are usually low quality, using low cost and cheap labor. Excessive working hours are another issue. EPZ employment has seen a highly un balanced gender structure. A significant proportion of these jobs use female workers. In garment industries, over 80% of the jobs can be female, because the enterprises seek a lower wage. Because of labor intensive operations, firms located in EPZs use female workers. The real effect of job creation is not as strong as expected. Tax con cessions are often justified by employment creation. But it is still debatable whether EPZs are effective in employ ment generation.

Some evidence suggests that, despite all the govern ment investment in infrastructure, the number of jobs generated is limited. If the same amount of resources were invested in alternative domestic small and medium enterprises, the effect of employment generation could be more significant. On this front, Asian EPZs seem to perform better than African EPZs in that they do absorb surplus migrant labor from rural areas. In sub Saharan African EPZs the goal of generating employment is not exactly achieved, due to political instability; most African EPZs perform poorly, except for Madagascar and Mauritius. In Kenya, for example, a survey shows that payment to zone workers is even lower than outside EPZs. This phenomenon is not seen in East Asian countries such as China. In theory, the enterprises in the zone contain modern premises and should provide a better working environment.

The jobs created in EPZs are of low security. This is associated with the footloose nature of firms in EPZs. They choose to locate in EPZs because of the low labor cost or beneficial trade situation (e.g., to use a trade quota). These factors may disappear, and the enterprises may then relocate to competing EPZs. Foreseeing the increase in labor costs and unionization, some enterprises have begun to reduce their workforces through voluntary resignation.

The wage for female workers is lower because they are temporary workers. Many may return to rural areas when they marry. The employer tends to discriminate against female workers by suppressing their wages. Women's jobs are more routine and labor intensive, which does not help to develop their skills. To maintain flexibility, some EPZ firms also try to subcontract their work to local firms that use lower paid workers. This further aggravates the problem of employment insecurity. Some jobs are sea sonal and are only created when there is an order.

To attract investors, the host country may also relax labor laws in EPZs, prohibit labor unions, or restrict labor union activities. In developing countries, even when such a labor law does exist, its implementation is often ineffective. There is a lack of supervision. The govern ment may have a misperception that more stringent labor regulation would reduce their competitiveness. It must be recognized that for EPZs productivity, rather than the cost of labor, is what matters. Government mediation and support services should not be seen adding to labor costs. Good labor relations enhance productivity and help to develop a stable labor force. The major constraint in the initial stage may be because the employer tends to stress flexibility instead of long term development. In general, because of restriction of union activities, unions are weaker in EPZs. There is a lack of long term social security in EPZs.

In the long run, the cost saving strategy, that is, simply constraining the cost of labor, may not enhance the competitiveness of EPZs. Competitiveness is built on many factors. Some EPZs have a poor physical and institutional environment. Thus, policies to provide training and labor development so as to enhance the quality of the labor force are important. Employment security helps to maintain a skilled and stable labor force, which is critical for the development of EPZs into potential drivers for the economy.

Establishing local labor unions for zone workers may cause capital flight. As they expect a trade enclave not to have unions, enterprises may relocate to other places. To solve this problem, regional labor unions may be set up to strengthen regional labor solidarity. This may be effective in retaining firms that aim at the regional market, as shown in South Korea's experience. The firms are concerned about their corporate image and the marketability of their products in the region, and decide to remain in the EPZ despite the rising cost.

Environmental Protection

The lack of environmental protection in EPZs is another concern. This is because control over industry is more lax and also the concentration of industrial activities in the zones causes the concentration of pollution. EPZs are set up for job creation and generating foreign exchange earnings rather than as places for high quality living. Environmental quality is thus not emphasized in EPZ management. On the other hand, the problem of lax environmental protection is not specific to EPZs but exists in many developing countries. There is a lack of any system of comprehensive environmental protection for domestic enterprises. The problem of pollution is serious in some cases. Given modern infrastructure, EPZs may in fact bring in an opportunity to enhance environmental protection. Building the capacity for treatment into EPZ infrastructure is important, although this may be a cost to the EPZ. But along with the transformation of EPZs into economic development zones, environmental quality is becoming essential for the competitiveness of zones.

Setting up too many EPZs can also cause a problem over the inefficient use of land in developing countries. In China, development zones are developed not only for promoting exports but also for land acquisition. Through setting up zones, the local government can get more land leasing premium through real estate development. This causes the problem of encroachment on rural land. As many zones are developed in more developed and densely populated costal areas, the fast expansion of development zones jeopardizes fragile ecosystems.

The Life Cycle of EPZs

Like all products, EPZs have a life cycle. The timing of the setting up of EPZs is important. Many successful EPZs were set up in the 1970s when there were fewer competing zones. Some EPZs, such as the Mauritian EPZ, benefit from preferential trade agreements to enter the European and US markets under MFA. When an EPZ reaches a mature stage, the cost of labor becomes higher and is no longer competitive. After operating for some years, tax incentives may no longer be applicable to some firms. This raises the question of whether these firms will remain in the EPZ. Gradually, the comparative advantage of cheaper labor is lost. Firms in EPZs respond to rising labor costs by reducing their employment and using subcontracts and flexible labor contracts to reduce their commitment to the place. They also update their production by using automation and product differentiation. The attractiveness of EPZs in terms of labor cost is then diminished. The relocation of standard labor intensive processing work seems inevitable. However, not all firms do relocate to other places. Some enterprises may shift to more technologically intensive products. For the production of complex products, other advantages may become more important, such as geographical proximity to the major market. For example, the EPZs in Mexico have seen rising costs recently but still have the advantage of being close to the US market.

Without new advantages, EPZs may be forced to compete with each other by reducing labor costs, a phenomenon known as the 'race to the bottom'. Supporting the development of high tech industries and/or nonstandard and just in time products may be the way forward. Through economic restructuring, EPZs may evolve into high tech industrial parks or production zones producing higher valued added goods. Tax concessions may not be the only way forward, because this may undermine the capacity of the local government to upgrade their science and technological infrastructure, in turn constraining the potential for upgrading the economic structure of EPZs.

Evidence from South Korea's EPZ (Masan Free Export Zone), for example, shows that some firms may remain in the zone when faced with rising costs, because of other advantages. When an EPZ is fully developed, the infrastructure improves, and administration services be come more experienced. Investment in EPZs is a sunk cost for investors, preventing them from relocating. Cultural and language affinities are also important factors to draw investors into a place. For example, the Masan FEZ is targeted at Japanese investors and the Japanese market. Through long term Japanese investment, many managers and workers in the zone can understand Jap anese and the Japanese management style, and thus can communicate with investors and customers in Japan more smoothly. Similarly, Taiwanese firms benefit from their language affinity with the mainland of China. The EPZ in Madagascar is known as the *Zone Franche* because there is a large French speaking population.

Along with overall trade liberalization, some of the specific advantages of free trade policies possessed by EPZs in developing countries may no longer be an at tractive point. The development of the WTO gives wider access to the US and European market. For example, ETZs in Madagascar are used to circumvent certain trade quotas. Asian countries, having used up their quotas for the US market, began to use the Madagascan ETZs as a production base. The phasing out of MFA may ser iously undermine their competitiveness. For many late comers in African EPZs, it becomes difficult to emulate Asian export oriented growth strategies to develop manufacturing industries for the global economy. The WTO rules out export subsidies, but some lower income economies (per capita national income below US$1000) enjoy exemption until 2007. The rule prohibits direct and main tax exemption in more industrialized developing economies. For some African countries, this may provide a chance to develop EPZs into high value added devel opment zones.

It is not the right time to set up EPZs that mainly rely on advantages in labor costs. EPZs should no longer be confined to the processing of imported raw and inter mediate materials. Newly established EPZs are facing tougher competition, and are more likely to have a comparative advantage in more complex products such as electronics. However, moving to complex products re quires a higher quality of labor force in the host country. Education will be crucial for this objective. Also, main taining a stable and skilled labor force is important. The government should emphasize better labor training, ra ther than simply constraining the labor cost by, for example, reducing labor union activities. Development zones should be used on a manageable scale to experi ment with new economic regulations that suit com petition in the world market. The host country should explore other advantages of EPZs besides tax and pref erential trade treatment, which could include geo graphical location within regional integration. The availability of regional production networks may also be a reason for some firms to remain in mature EPZs, as the production of complex products requires just in time production and geographical proximity.

Nearly all mature EPZs face the problem of rising costs. It is therefore critical for them to upgrade their structure. But investment in infrastructure requires some initial input. Foreign transnational corporations are un likely to initiate this upgrading process because they are the most likely to reduce employment and use pro duction automation. The contraction of employment can be disruptive, not only reducing productivity but also causing social instability. Labor disputes can undermine the strength of EPZs and is detrimental for upgrading the economies of EPZs. The capacity of the government is important to sustain the competitiveness of EPZs and help them transform into high tech parks.

The integration of EPZs with the overall economy is the key to success. The government should review policy periodically to suit new conditions and foster backward linkage. EPZs should be used as a window to further expand trade and production capacity rather than re maining as isolated trade enclaves. Generating employ ment and foreign exchange earnings is now becoming secondary to technology transfer and the introduction of management and knowledge through EPZs. To achieve this objective, it is important to develop the domestic sector to become partners with firms in EPZs. The overall trend is that EPZs will become comprehensive development zones.

The Future of EPZs

EPZs are created under specific circumstances – de veloping countries have abundant labor resources, while capital becomes mobile in the global economy. The combination of labor and capital in EPZs provides a chance for developing countries to absorb foreign direct investment (FDI) and be linked to the global economy with the minimum impact on the domestic economy, as the goods produced in EPZs are exported. The con sequence is that this policy indirectly protects domestic firms that do not have international competitive capacity.

For developing countries that are in the earlier stages of opening up their economies, the EPZ policy is a useful experiment. The demonstration effect can help the re form of the whole economy toward more export oriented growth. China seems to suggest this route of growth, with EPZs being part of the overall reform package, which will eventually spread over the coastal region and to inner and Western regions. But such a spreading out effect may not easily be achieved. According to United Nations Conference on Trade and Development (UNCTAD), the production of EPZs focuses on low skills and low technology. The sectors are concentrated in garments, toys, footwear, electronic assembly, and light machinery goods. The material processing industries are low value added sectors, and the FDI absorbed in these sectors is 'low quality' in terms of the potential for technological advancement. Even when productive capacities are de veloped, with separation from the local economy, low technology transfer, and weak backward linkage, EPZs may not be able to overcome this growth trap and fully develop into engines for future growth.

Because WTO rules are against subsidies for export, direct exemption from taxes in EPZs in countries that exceed US$1000 annual per capita income for 3 years is seen as a subsidy for exports. This requires the phasing out of EPZs' tax incentives. Some EPZs are taking this chance to evolve into industrial and science parks and beginning to integrate with their local economies.

See also: Foreign Direct Investment; Globalization, Economic.

Further Reading

Boyenge, J. P. S. (2003). *ILO Database on Export Processing Zones*. Geneva: International Labour Organization (ILO).

Cling, J. P., Razafindrakoto, M. and Roubaud, F. (2005). Exporting processing zones in Madagascar: A success story under threat? *World Development* 33, 785 803.

Dowla, A. (1997). Export processing zones in Bangladesh: The economic impact. *Asian Survey* 37(6), 561 574.

International Labour Office (ILO) (2003). *Employment and Social Policy in Respect of Export Processing Zones (EPZs)*. Geneva: ILO.

Jauch, H. (2002). Export processing zones and the quest for sustainable development: A Southern African perspective. *Environment & Urbanization* 14, 101 113.

Jayanthakumaran, K. (2003). Benefit cost appraisals of export processing zones: A survey of the literature. *Development Policy Review* 21(1), 51 65.

Johansson, H. and Nilsson, L. (1997). Export processing zones as catalysts. *World Development* 25, 2115 2128.

Landingin, N. and Wadley, D. (2005). Export processing zones and growth triangle development: The case of the BIMP EAGA, Southeast Asia. *Journal of International Development* 17, 67 96.

Lee, Y. S. (1999). Labor shock and the diversity of transnational corporate strategy in export processing zones. *Growth and Change* 30, 337 365.

Madani, D. (1999). *A Review of the Role and Impact of Export Processing Zones*. Washington, DC: The World Bank.

Mireri, C. (2000). The impact of export processing zone development on employment creation in Kenya. *Singapore Journal of Tropical Geography* 21, 149 165.

Sargent, J. and Matthews, L. (2001). Combining export processing zones and regional free trade agreements: Lessons from the Mexican experience. *World Development* 29, 1739 1752.

Sargent, J. and Matthews, L. (2004). What happens when relative costs increase in export processing zones? Technology, regional production networks, and Mexico's maquiladoras. *World Development* 32, 2015 2030.

Schrank, A. (2001). Export processing zones: Free market islands or bridges to structural transformation? *Development Policy Review* 19, 223 242.

Zee, H. H., Stotsky, J. G. and Ley, E. (2002). Tax incentives for business investment: A primer for policy makers in developing countries. *World Development* 30, 1497 1516.

Relevant Websites

http://www.ilo.org
International Labour Organization.

http://www.wto.org
World Trade Organization.

Extended Metropolitan Region

V. F. S. Sit, University of Hong Kong, Hong Kong, People's Republic of China

Glossary

Global City Region A term invented by Scott to reflect the post-1980 'large scale metropolitan urbanization' gravitating around key metropolitan sites in the emerging global economic order. Among them many have been named megalopolises and had been significant in national and global influence before the recent round of economic globalization. They also include key metropolitan regions and their extending periphery in Pacific Asia, and Central and South America.

Megalopolis A term coined by Gottmann to mean a very large and spreading urban complex, with some open land and farm land, formed when built-up areas, widespread over an extensive area due to the popular use of the automobile, enlarge to such an extent that they become linked together. Megalopolises are formed in the northeastern seaboard of the US, or along the northern shore of the Inland Sea of Japan. The component centers are not distinct from each other in income level or level of economic development. The megalopolis is thus typical of mega-urban region development in advanced economies.

Metropolitan-Based Region (MBR) In parts of the industrialized world, like Israel, a new metropolitan region, has emerged. It is composed of the metropolis and a surrounding territory. While the metropolis seems to be stagnant or declines in growth, particularly in population, the surrounding territory has been growing rapidly, both economically and in population due to intensified linkages with the metropolis. It is this wider spatial process of consolidation within the MBR that has strengthened the economic role and influence of the metropolis. Such growth in the surrounding territory is not suburbanization but a result of the unique combination of agglomeration diseconomies of the metropolis and the agglomeration economies that attracted activities to its hinterland.

Metropolitan Interlocking Region (MIR) It is a term coined by Zhou to refer to the six large urban clusters in China. The three largest clusters have been the basis for the three extended metropolitan regions (EMRs) in this article. Conceptually, the MIR is regarded as similar to the megalopolis, though it exhibits 'Chinese characteristics' such as: each contains two or more cities with a population of over 1 million; a large international seaport with over 100 million tonnes of annual throughput; an international airport; and a total population of 25 million. The component parts of the MIR are linked together by efficient one-day transport.

Definition

Pacific Asia had been the world's fastest growing region in the decades of 1980–2000. Admist their hectic economic growth, countries in the region such as China, South Korea, Thailand, and Indonesia had experienced rapid urbanization around their major cities, often the national capitals and port cities, that took the form of 'regional urbanization', that is, the urbanization of former rural areas around these urban cores. Such a pattern of concentrated growth has led to the formation of a large mega urban region around the former metropolis, that is, the formation of an extended urban region from the core metropolis that is dubbed as "extended metropolitan region" (EMR).

The EMR is a term coined by McGee and is believed to be a local response in Pacific Asia to increasing economic globalization, for the EMR doubles as a workbench for the global production system and a local logistics and marketing hub for multinational corporations (MNCs). The EMR shows a unique spatial structure with different growth patterns in its component spatial parts. Within it, the old urban center, or a few neighboring and functionally closely linked large cities, forms the core (**Figure 1** and **Table 1**). While the core is growing slowly with lower fertility and lower net migration, the peripheral and/or interstitial areas around the core, originally rural areas and small towns, have experienced rapid economic development leading to hectic structural transformation and population growth. The latter have been represented by urban expansion, rapid population growth, and large net migration increase. In short, greater involvement with global economic processes in the core has led to an overspill of lower skill and land intensive activities into peripheral areas which has caused the transformation of such areas.

In the regional scale in social science research, the EMR represents a new regionalism which has importance in economic planning and political governance, largely due to the territoriality of globalization that has led to the ever more intensive concentration of capital, people, institutions, and technology in localized urban clusters. These 'new' urban regions are assuming important roles as: (1) a fundamental basis for economic and social life, (2) a vital relational asset for distilling

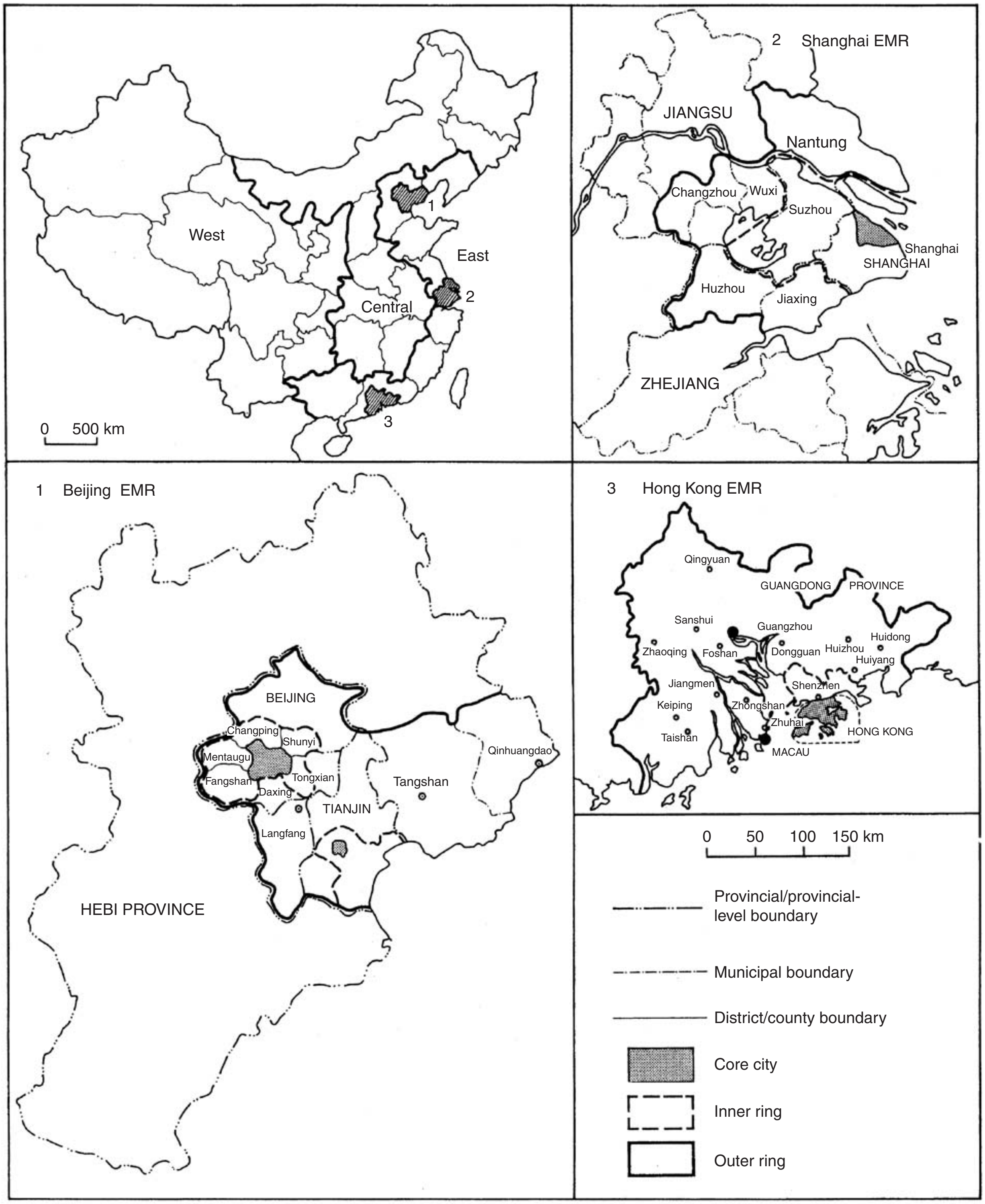

Figure 1 The three coastal EMRs of China. Source: Sit, V. F. S. (2005). China's extended metropolitan regions: Formation and delimitation. *International Development Planning Review* 27(3), 297 331; Jones, G. W. (2002). Southeast Asian urbanization and the growth of mega-urban regions. *Journal of Population Research* 19(2), 119.

learning based competitive advantage, and (3) a meso level of analysis within which to examine the new era of flexible capitalism. They are thus different from other mega urban regions such as the megalopolis, metro politan based region, and global city region. As the focal point of economic growth in many such nations that

Table 1 Selected extended metropolitan regions in Asia

EMR	*Area (km²)*	*Population (in 1000)*			*Pop. and growth rate (%)*		*Density (Pop./km²)*	
		1980	*1990*	*2000*	*1980 90*	*1990 2000*	*1980*	*2000*
Bangkok								
Core	1565	4691	5876	6362	2.28	0.80	3000	4038
Ring	6200	1947	2706	3760	3.35	3.34	314	606
Total	7765	6644	8582	10 080	2.59	1.62	855	1298
Jakarta								
Core	664	6481	8223	8385	2.41	0.20	9760	12 689
Ring	5312	5413	7676	12 749	3.55	5.20	1019	2388
Total	5916	11 894	15 899	21 134	2.94	2.89	1990	3536
Manila								
Core	636	5926	7948	9933	2.98	2.25	12 497	15 617
Ring(1)	2927	2820	4107	5049	3.83	2.09	1403	1724
Ring(2)	9394	2932	3908	4806	2.92	2.09	312	511
Total	12 967	11 678	15 963	19 788	3.18	2.17	901	1526
Seoul								
Core	615	8364	10 613	9853	2.41	0.74	13 600	16 021
Ring	3800	4934	7974	9463	4.92	1.73	1298	2490
Total	4415	13 298	18 587	19 316	3.41	0.39	3012	4375

Source: Sit, V. F. S. (2005). China's extended metropolitan regions: Formation and delimitation. *International Development Planning Review* 27(3), 297 331; Jones, G. W. (2002). Southeast Asian urbanization and the growth of mega urban regions. *Journal of Population Research* 19(2), 119.

possess comparative advantages for harnessing the processes of globalization, the EMR will continue to attract attention in discussions on economic development and urban growth in the less developed countries (LDCs).

Formation and Spatial Structure

In the late 1980s, some researchers had noted the development of large areas of 'desakota' (semi rural, semi urban, or rural–urban transitional zones) around major urban centers in Southeast and East Asia, leading to the formation of a new type of mega urban region – the EMR. The latter is regarded as a distinct component of recent urbanization in Pacific Asia, a dynamic transactive region that defies all levels of conventional rural–urban thinking. The key cities in Southeast and East Asia on which the EMRs were often based were already economic foci of their respective countries before the 1980s. They possess obvious advantages for foreign direct investment (FDI) and export oriented economic growth after these nations de regulated and embraced economic globalization since the 1980s. In particular, the 'global shift' associated with the New International Division of Labor (NIDL) has given momentum to their growth. Agglomeration economies underpine the comparative advantages of these major metropolises such that they continue to exert a powerful magnetic attraction for investment and services toward themselves, while at the same time stifling the potential for urban and industrial dispersal. The resultant increase in export oriented manufacturing and logistics (which include the warehousing, distribution, and transportation for both exported and imported goods for the national market) have fuelled economic and urban growth around them, leading to the formation of the desakota zone whose name derives from the Indonesian terms for town and village.

The desakota contains not just the conventional range of rural and agricultural operations but a mixture of activities of a nonagricultural nature and some of urban characteristics, with comprehensive sets of economic, social, and cultural links with the nearby metropolis. Thus 'regional urbanization' takes the spatial form of extension of the core metropolis in desakota growth, often joining together into one urban region of functionally linked towns and cities in the vicinity. 'Glocalization', that is, the functional and spatial processes that enable the metropolis to continue to grow in increasing economic globalization, is therefore the major process behind EMR formation. Through 'glocalization' or EMR formation, these countries are able to engage and, more importantly, to retain their stake in the global and regional economies.

The Hong Kong EMR (Hong Kong as its core and the Pearl River Delta its periphery) and the EMRs in Indonesia, Thailand, and the Philippines demonstrate that economic globalization in the form of FDI in labor intensive, export oriented industries has been the main force behind EMR formation. The spatial characteristics of these EMRs are also distinct from those of a megalopolis. Within each EMR there is a socioeconomic and population gradation that descends from the core to the periphery to reflect the functional/economic division of labor between these two component parts, and the spatial difference between them, to underline their different economic and infrastructural milieux. For example, FDI in the tertiary sector of the EMR is mostly attracted to the core, while FDI in labor intensive industries, ware housing, transport, and recreation is largely located in the periphery. These two spatially distinct roles within the EMR are integrated in the 'front shop–back factory' model as revealed in the Hong Kong EMR case. More precisely, service activities, including producer services and international trade, are located in the core city to support the overall role of the EMR as a production and marketing platform in NIDL and international sourcing, while production processes depending on cheap and unskilled labor are largely found in the periphery.

As illustrated by **Table 1**, in Thailand, the Philip pines, Indonesia, and South Korea (countries which vary in size from about 90 000 to over 1 million sq km) the respective EMRs cover a territory of 4415–12 967 sq km, with a population of 11–21 million. In terms of area, each occupies less than 5% of the country concerned. How ever, they are the national capitals, the most important international ports, and the most highly urbanized and developed regions of these countries. They are also the main recipients of FDI (e.g., 46.8% in Bangkok; 35% in Jakarta), and national nodes for integrating with the global economy in the form of international trade and passenger flows.

The spatial expansion of the core city into its im mediate former countryside to form the EMR has been a means for keeping down the critical costs of land and labor, and for avoiding the growth restraining concerns of congestion, pollution, and shortage and high cost of space in the core metropolis. As referred to before, it is a form of glocalization, by which these metropolises have re mained competitive and capable of attracting further inflows of FDI in increasing economic globalization to function as low cost production centers and local marketing and logistics hubs for MNCs. The Bangkok EMR, which comprises the Bangkok Metropolitan Area (1565 sq km) as the core city, and the surrounding five provinces (6193 sq km) as its periphery, may illustrate such a glocalization process. Between 1980 and 1995, the core city was rapidly 'Manhattanizing', that is, becoming more specialized in high level producer services such as banking and financing, and taking on an urban landscape characterized by a skyscraper style of modern intelligent office blocks. Measured in investment terms, it had registered a 150% growth in its banking, finance, and real estate sectors. It also had absorbed 83% of Thailand's total investment in services. The periphery, comprising the surrounding cities and the desakota area, contained 64% of the EMR's investment in manufacturing and accounted for 23% of Thailand's total investment in manufacturing for the same period.

Indeed, glocalization reflects the functional and spatial processes of EMR formation as: (1) the core city becomes increasingly specialized in producer services; and it also includes a concentration of hotels and tourist facilities, modern international transportation and communication infrastructures such as airports, seaports, and 'info ports', which enable it to serve as the national/regional hub for international and domestic cargo, passenger, and infor mation flow and (2) labor and land intensive activities, including manufacturing, warehousing, and recreational facilities have been decentralized and expanded into the periphery. In summary, EMR formation represents a more thorough going integration of Pacific Asia and its cities into the global system while the traditional core–periphery relation was an inadequate way of encapsu lating the present role of these city regions in the post 1980 world economy – a trend that was said to have been accelerated in the 1990s.

The Chinese EMRs are typical in terms of the dy namics of EMR formation and its spatial processes and characteristic features, as among the Asian Pacific countries, China has been vigorously de regulating, pursuing reforms to attract FDI, and actively partici pating in the global economy. These efforts have been generalized by the rubric of the 'open and reform policy' that started in 1978. At present, the country is the world's second largest host to FDI, with the US being the world leader. Out processing, which is a form of LDC partici pation in NIDL based on cheap labor and efficient access to global markets, accounted for over 50% of China's total international trade and has been a key dynamic in the nation's rapid economic growth in the past two decades. The last two processes have been concentrated in the country's coastal region where three EMRs have been taking shape (**Figure 1**). They reflect the locations of China's major port cities and the earlier and privileged policies in the open and reform policy that these locales have enjoyed. However, like some Western scholars, many Chinese researchers have confused the Chinese EMRs with megalopolises and created a new term – the metropolitan interlocking regions (MIRs) – for these urban clusters of China. They nevertheless admit the connection between the global economy (particularly trade), dense and interactive urban growth within the region, and the coordinating role of the central city in MIR formation.

EMR Delimitation

Existing EMR studies usually intuitively take the largest city of the country as the core city, and its adjacent ad ministrative units, very often provinces, municipalities, or counties that are believed to have received the dispersal of economic activities from the core city, as the periphery or ring. In short, they use the central metropolis as the core, while contiguous administrative units that satisfy selected demographic, density, and employment criteria are in cluded as the ring. The latter comprises sometimes the inner ring (30–50 km from the core) and the outer ring (50–100 km from the core). Such definition is not a problem in itself, though it is not scientific. In the four cases in **Table 1**, the core city is small (615–1565 sq km in size) while the periphery may be four to six times the core area. Population growth and the 'regional urbanization' in these EMRs, take place mainly within the periphery, whose population growth rate varies from twice to 20 times that of the core city. With the exception of the Seoul EMR, which lies in a country with a very high level of urbanization of over 77%, the EMR periphery in Pacific Asia in general comprises a substantial rural population.

Like other Asia Pacific studies, Chinese scholars have also paid little attention to the scientific delimitation of the EMR. The first serious attempt to delimit a Chinese EMR quantitatively was made by Ning. For him the EMR is composed of a number of metropolitan regions contiguous to each other. Each metropolitan region should have a central city with a population of over 200 000. The ring of the region comprises surrounding rural counties that: (1) are intensively linked to the central city, and (2) have achieved a high degree of ur banization and industrialization, measured by the con tribution of nonagricultural activities to GDP (>75%) and employment (>60%), and transportation connect ivity with the central city. Sit's 2005 attempt is so far the most scientific. Based on submunicipal, that is, urban district and county data, he used variables on economic growth, economic globalization, (measured by FDI inflow and international trade), producer services, and rural–urban use mix to delimit the core, inner, and outer rings of the EMRs in China (**Table 2**).

As shown in **Table 3**, the Chinese EMRs are of roughly the same size, about 30 000–40 000 sq km in area, with a radial extent of about 150 km, and a population of 20–40 million. They are comparable with each other, though a few times larger than other EMRs in Asian Pacific. All three have very high overall population densities. In terms of size of their economy measured by GDP, both the Shanghai and Hong Kong EMRs exceed US$100 billion. This is comparable to a medium sized and medium income level country. The gradation from the core to the outer ring in population density, per capita GDP, and intensity of export has been obvious (**Table 3** and **Figure 1**).

In the past 20 years, the three EMRs have dominated China's export oriented economy. In 1999, they ac counted for 73% of China's FDI inflow, 73.1% of its total exports, and contributed 30.7% of the country's GDP, although they occupied only 1.24% of the nation's ter ritory and accommodated 7.53% of its total population (**Table 3**). Of the three EMRs, the Hong Kong EMR has an economic size about 70% larger than the Shanghai EMR as measured by GDP. It is also a much more open economy in terms of utilized FDI and exports. This is contrary to the findings of many proto EMR studies by mainland Chinese researchers, who regard the urban cluster in the Yangtze Delta as the most important eco nomic region in the Chinese economy. This message is easily understood as the Hong Kong EMR possesses Asia's largest international port cluster and airports cluster. As **Table 3** illustrates, in 1999, the EMR handled 56.6% of China's total exports (US$67.4 billion in value) and absorbed 57% (US$37.3 billion) of its total FDI inflow.

Favorable geographical locations, appropriate open and liberal policies, and modern infrastructure have en abled the three EMRs, particularly the Hong Kong EMR, to be hubs for domestic and international flows of people, materials, and capital, and production centers of export oriented industries. They are at present a global factory for the production of light manufactures of low skill inputs, for example, textiles, garments, toys, footwear, consumer electrical and electronics, and computer parts and peripherals. They are equally significant bases for China's import substitute consumer industries, which are

Table 2 Criteria for delimitating the three EMRs

	National	*Three coastal clusters*[a]	*Delimitation indice*		
			Core	*Inner ring*	*Outer ring*
Per capita GDP (¥)	6387	11 728	> 23 456	> 17 592	> 11 728
Actualized FDI per 10 000 ¥ GDP (¥)	535	655	> 1310	> 983	> 655
Export/GDP (%)	19.7	35.5	> 53.3	> 35.5	> 19.7
Tertiary sector/GDP (%)	33.0	39.1	> 60.0		
Primary sector/GDP (%)	15.7	11.0	< 5	< 7.5	< 15.7

[a] The criteria have been applied to official statistics of various provinces and subprovincial regions for 1999 based on the administrative divisions of that year. Data computed by author from various official economic yearbooks of 2000.

The three coastal clusters refer to the provincial units included in step 2. Using these more globalized provinces' averages as baselines generates areas of higher level of development and of more intensively globalized economy.

Source: Sit, V. F. S. (2005). China's extended metropolitan regions: Formation and delimitation. *International Development Planning Review* 27(3), 297 331; Jones, G. W. (2002). Southeast Asian urbanization and the growth of mega urban regions. *Journal of Population Research* 19(2), 119.

Table 3 Key economic and social data for the three EMRs, 1999

EMR	*Area (sq. km)*	*Population (10 000)*	*GDP (¥100 mil.)*	*Industrial output (¥100 mil.)*	*Utilized FDI (US$100 mil.)*	*Export (US$100 mil.)*
Beijing: Core	1112	947	540	1394	7.2	3.6
Inner ring	11 066	684	930	1736	29.8	113.4
Outer ring	19 038	860	2632	2227	4.0	40.0
Subtotal	31 216	2512	4102 (31.8%)	5357	41.0	157.0
Shanghai: Core	812	738	1153	1755	15.0	100.0
Inner ring	11 811	914	4006	4786	36.9	144.5
Outer ring	28 983	2280	3395	8844	12.1	81.5
Subtotal	41 586	4032	8554 (31.6%)	15 385	64.0	326.0
Hong Kong: Core	1100	684	8091	2510	252.0	1000.0[a]
Inner ring	9478	903	4038	3018	60.0	360.0
Outer ring	32 244	1402	2400	8436	61.0	314.0
Subtotal	42 822	2989	14 529 (95.6%)	13 964	373.0	1674.0
Three EMRs as % of national total	1.24[b]	7.53[b]	30.7[b]	26.7[b]	73.0[b]	73.1[b]

[a] Include half of Hong Kong's re exports, that is, US$75.8 billion.
[b] Include Hong Kong and Macau.
Bracketed figures are ratios of export value to GDP.
Source: Sit, V. F. S. (2005). China's extended metropolitan regions: Formation and delimitation. *International Development Planning Review* 27(3), 297 331; Jones, G. W. (2002). Southeast Asian urbanization and the growth of mega urban regions. *Journal of Population Research* 19(2), 119.

equally highly globalized in the form of joint ventures and technology transfer, for example, the auto and household appliances industries in Shanghai and the Pearl River Delta (PRD). The Hong Kong EMR has exemplified the path of EMR formation centered around a major port city of long experience in producer services, supported by a legal system with international credibility, being an international finance and trading center, an entrepot with excellent port facilities, besides having post war industrialization experience that provides it with accumulated expertise in marketing, design and production management, while the PRD's large labor and land resources complement the manufacturing process that is decentralized from yet dependent on, as well as supportive of the core city.'

EMR: Management

EMRs are expanding rapidly in Pacific Asia due mainly to glocalization there in the past two decades. They are at present in the midst of a structural change to meet in tensified inter urban competition across nations, par ticularly nations within Pacific Asia which has been the world's most dynamic growth engine since the 1980s. With worldwide economic integration, the accelerated 'regional urbanization' in the form of the EMR has made traditional concepts of the city and the related urban and regional planning and development strategies increas ingly problematic and present many challenges to re searchers and policy makers in the twenty first century.

Their growth and development have also enhanced regional competence by bringing together nearby administratively independent territorial units of different comparative advantages into a single urban economic region. So far, the growth has mostly been the outcome of economic forces rather than of direct government promotion. Keen competition among them, however, has pushed some of them toward further improving their competitiveness through a new level of intra EMR cooperation in which governments have played a leading role, such as coordination in pan EMR projects and policies in infrastructure, trade and tax, and in place marketing in the global market through trade fairs, exhibitions, and trade promotion. For example, Hong Kong's intra EMR initiative has resulted in closer eco nomic partnership agreement (CEPA), with the mainland of China, an important step forward (Hong Kong SAR Government). In Indonesia, the Special Region of the National Capital of Jakarta has been created to coordinate development and planning of the Jakarta EMR or Jabotabec. Intra EMR cooperation with a co ordinated and unified development vision for the com ponent administrative areas and ensuring intraregion economic policies has now been generally accepted. The setting up of pan region institutions for such purposes, for example, under CEPA, for implementing the agreed policies, has become an urgent planning agenda in the central and relevant local governments of respective nations.

Although the term EMR was raised as early as 1989, our understanding of the EMR is still inadequate, both conceptually and quantitatively, for addressing pressing issues of planning and policy formulation raised by it for achieving development through harnessing the forces of economic globalization. Their obvious competitive ad vantages in the era of economic globalization means that EMRs will continue to be the major growth centers for many LDCs. How to better understand them and to harness their special features and advantages for further growth should therefore be an important item on the nation's drawing board. For example, with the deepening of economic globalization and competition among Asia Pacific's mega urban regions, China's EMRs have to stay competitive internationally, in order to be the country's primary future growth platforms and key to further economic development. The Chinese central and local governments should not only provide a stable and fair legal and social environment for enterprises and individuals within the EMRs. In addition, they should take action to induce more polarization, that is, concentrated development for scale economies and specialization economies, and economic globalization by: (1) abridging administrative hurdles for the flow of people, materials, and capital between different com ponent parts of the EMR and to achieve integrative planning within the EMR based on a common develop ment vision, and (2) adopting the right policies for attracting FDI and for achieving further international competitiveness.

Besides adding impetus to national economic growth, as fast growing large urban units, EMRs have also created quality of life and environmental problems, as well as problems of regional disparity. Firman has listed a number of environmental issues caused by the rapid growth of Jabotabec in Indonesia: (1) loss of farmland, (2) degradation of water resources, (3) commercial agri culture in upland areas, (4) replacement of mangroves by housing, (5) rapid increase in urban solid wastes, and (6) air pollution. Accompanying and making these issues increasingly problematic are also faults in local govern ance, for example: (1) violation of land use plans by local governments for encouraging business activities, (2) ab sence of metropolitan authorities above local govern ments to coordinate sectoral planning in the EMR and lastly, (3) in plan implementation, local governments compete, instead of cooperate with one another over metropolitan development planning. Thus, it is generally agreed that concerns over future growth of an EMR are not simply ones of economic development and ability in harnessing globalization forces. Environmental and social sustainability are equally important for its continued growth and development. The authorities, too, have to pay attention to redressing regional disparity caused by polarized development in the form of EMR growth. The large scale influx of population into the EMR's ring has caused strains on the urban regions' resources, environ ment, and social stability. The pending rapid structural change of EMRs economically into economies of higher order and higher value added activities, and further spatial dispersal of the low skill and labor intensive ac tivities beyond the EMR into remoter provinces, such as the case of China, have also cautioned the need for new and coordinated policies both at the regional and na tional level. Quite obviously, new institutional changes above what has been mentioned are required for coping with both these economic and noneconomic concerns within and outside the EMRs.

Mega urban region growth in developing countries represents the best and worst of the development process. EMRs are not only places where highly productive and innovative economies are present, but where multi faceted market failures will likely happen in future, leading to issues of environmental and social stability and overall sustainability of development. EMR authorities may submit to these pressures passively, or engage actively in institution building and policy making in an effort to avoid the negative side effects of development and of rigid management, and thereby continue to turn globalization as far as possible to their maximum benefit.

See also: Foreign Direct Investment; Globalization and Transnational Corporations; New Regionalism; Region; Transnational Corporations in Developing Countries; Transnationalism and Labor Geography; Urbanization.

Further Reading

Bar El, R. and Parr, J. B. (2003). From metropolis to metropolis based region: The case of Tel Aviv. *Urban Studies* 40(1), 113 125.

Firman, T. (1995). The emergence of extended metropolitan regions in Indonesia: Jabotabec and Bandung metropolitan area. *Review of Urban and Regional Development Studies* 7, 166 188.

Gottmann, J. (1961). *Megalopolis: The urbanized northeastern seaboard of the United States*. New York: K.I.P.

Jones, G. W. (2002). Southeast Asian urbanization and the growth of mega urban regions. *Journal of Population Research* 19(2), 119.

Lo, Fu chen and Marcottullio, P. J. (eds.) (2000). Globalization and urban transformations in the Asian Pacific region: A review. *Urban Studies* 37(1), 77 111.

Macheod, G. (2001). New regionalism reconsidered: Globalization and the remaking of political economic space. *International Journal of Urban and Regional Research* 25(4), 804 829.

McGee, T. G. (1989). Urbanisasi or kotadesasi? Evolving patterns of urbanization in Asia. In Costa, F. J. (ed.) *Urbanization in Asia*, pp 93 108. Honolulu: UHP.

McGee, T. G. (1995). Metrofitting in emerging mega urban region in ASEAN: An overview. In McGee, T. G. & Robinson, I. M. (eds.) *The Mega urban regions of Southeast Asia*. Vancouver: UBC Press.

McGee, T. G. (1998). Globalization and rural urban relations in the developing world. In Lo, F. & Yeung, Y. M. (eds.) *Globalization and the world of large cities*, pp 471 496. Tokyo: UNU.

Ning, Y. (1998). New urbanization process: On China's urbanization dynamics and characteristics. In Xu, X., Xue, and Yen, X. (eds.) *China's Rural urban Transition and Coordinated Development*, pp 201 207. Beijing: Science Press.

Scott, A. J., Agnew., Soja, E. W., and Storper, M. (2001). Global city regions. In Scott, A. J. (ed.) *Global City Regions: Trends, Theory, Policy,* pp 11 32. Oxford: Oxford University Press.

Sit, V. F. S. (2001). Increasing globalization and the growth of the Hong Kong extended metropolitan region. In Lo, F. & Marcotullio, P. (eds.) *Globalization and the sustainability of cities in the Asia Pacific region*, pp 199 238. Tokyo: UNU Press.

Sit, V. F. S. (2005). China's extended metropolitan regions: Formation and delimitation. *International Development Planning Review* 27(3), 297 331.

Zhou, Y. X. (1991). The metropolitan interlocking regions in China: A preliminary hypothesis. In Ginsburg, N., Koppel, B., and McGee, T. G. (eds.) *The Extended Metropolis: Settlement Transition in Asia*, pp 89 111. Honolulu: University of Havwaii Press.

F

Factor Analysis and Principal-Components Analysis

F. Wang, Louisiana State University, Baton Rouge, LA, USA

Glossary

Communality Proportion of a variable's variance shared by the retaining components in principal-components analysis or by the common factors in principal-factor analysis.

Data Standardization The process of converting a series of data to a new series with mean equal to 0 and standard deviation equal to 1.

Latent Variables Unobserved variables that combine many original observed variables based on principal components analysis or factor analysis.

Multicollinearity Linear relationships between two or more explanatory (X) variables that may cause difficulties in interpreting regression results.

Oblique Rotation In contrast to orthogonal rotation, oblique rotation allows correlations between factors in order to achieve greater simple factor structure in factor analysis.

Orthogonal Rotation This rotation uses linear functions of independent (uncorrelated) factors to reconstruct the scores on the original variables in factor analysis.

Varimax Rotation An orthogonal rotation that maximizes the variance of the squared loadings for each factor, and thus polarizes loadings (either high or low) on factors for easy interpretation.

In many applications, variables are often correlated with each other, and thus contain duplicated information to some extent. When the number of variables is high, it is also desirable to find a few latent (unobserved) variables that describe those original variables. Principal components analysis (PCA) and factor analysis (FA) use fewer components or factors to represent the original variables, and thus consolidate the data for easy interpretation. For example, the term 'socioeconomic status' is composed of many highly correlated variables such as income, education, occupation, and others. PCA and FA are used to come up with a smaller number of components or factors that represent the shared correlation among these variables. They are sometimes used together for the common objective of data reduction. Despite the commonalities, there are important differences between the two. PCA attempts to model all the variance in the original variables, uses the same number of components to simply transform the original data, and thus is strictly a mathematical transformation. FA attempts to model only 'common' or shared variance and assumes that some variance is 'unique' to each variable. FA uses fewer factors to capture interrelations of the original variables with some residuals, and thus is strictly a statistical method.

Principal-Components Analysis

For convenience, the original data of observed variables Z_k are first standardized so that variables do not depend on measurement scale and thus are comparable to each other. PCA transforms original data on K observed variables Z_k (likely correlated with each other to some extent) to data on K principal components F_k that are independent from each other:

$$Z_k = l_{k1}F_1 + l_{k2}F_2 + \cdots + l_{kj}F_j + \cdots + l_{kK}F_K \qquad [1]$$

Conversely, the components F_j can also be expressed as a linear combination of the original variables Z_k:

$$F_j = a_{1j}Z_1 + a_{2j}Z_2 + \cdots + a_{Kj}Z_K \qquad [2]$$

The components F_j are constructed to be uncorrelated with each other, and are ordered such that the first component F_1 has the largest sample variance (λ_1), F_2 the second largest, and so on. The variances λ_j corresponding to various components are termed 'eigenvalues', and $\lambda_1 > \lambda_2 > \cdots$. A larger eigenvalue represents a larger share of the total variance in all of the original variables captured by a component and thus indicates a component of more importance. In other words, the first component is more important than the second, and the second is more important than the third, and so on.

For readers familiar with the terminology of linear algebra, the eigenvalues λ_j $(j=1,2,\ldots,K)$ are based on the correlation matrix **R** (or, less commonly, the covariance matrix **C**) of the original variables **Z**. The correlation matrix **R** is written as

$$\mathbf{R}=\begin{pmatrix} 1 & r_{12} & \ldots & r_{1K} \\ r_{21} & 1 & \ldots & r_{2K} \\ \cdot & \cdot & \ldots & \cdot \\ r_{K1} & r_{K2} & \ldots & 1 \end{pmatrix}$$

where r_{12} is the correlation between Z_1 and Z_2, and so on. The coefficients $(a_{1j}, a_{2j}, \ldots, a_{Kj})$ for component F_j in eqn [2] comprise the eigenvector associated with the jth largest eigenvalue λ_j.

Since standardized variables have variances of 1, the total variance of all variables also equals the number of variables such as

$$\lambda_1 + \lambda_2 + \cdots + \lambda_K = K$$

Therefore, the proportion of total variance explained by the jth component is λ_j/K. Since the variances of principal components equal the different eigenvalues, the principal component values F_j represented in eqn [2] are not standardized. For various reasons, it is often desirable and more convenient to work with standardized components. For instance, factor loadings and factor scores, discussed in the next section on FA, are based on standardized components.

In PCA, the number of components equals the number of original variables, so no information is lost in the process. The newly constructed components are a linear function of the original variables and have two important and useful properties: (1) they are independent of each other and (2) their corresponding eigenvalues reflect their relative importance. The first property ensures that multicollinearity is avoided if resulting components are used in regression, and the second property enables us to use fewer components to capture most of the information in a multivariate data set. In general, the first few components account for a large fraction of the total variance.

By transforming the original variables to the same number of components, the PCA does not serve the purpose of data reduction. One may simply discard the less important components and keep only the first few components to achieve data reduction. If we retain only the J largest components ($J<K$), eqn [1] becomes

$$Z_k = l_{k1}F_1 + l_{k2}F_2 + \cdots + l_{kJ}F_J + v_k \qquad [3]$$

where the discarded components are represented by the residual term v_k such as

$$v_k = l_{k,J+1}F_{J+1} + l_{k,J+2}F_{J+2} + \cdots + l_{kK}F_K \qquad [4]$$

Equations [3] and [4] are sometimes termed a 'principal components factor' model. However, it is not a true FA model.

Factor Analysis

FA is based on a different assumption than PCA. FA assumes that each variable has common and unique variance, and the method seeks to create factors that account for the common, not the unique, variance. FA uses a linear function of J unobserved factors F_j, plus a residual term u_k, to model each of K observed variables Z_k:

$$Z_k = l_{k1}F_1 + l_{k2}F_2 + \cdots + l_{kJ}F_J + u_k \qquad [5]$$

The same J factors, F_j, appear in the function for each Z_k, and thus are termed 'common factors'. Equation [5] looks like the principal components factor model in eqn [3], but is very different. The residuals u_k in eqn [5] are unique to variables Z_k (termed 'unique factors') and uncorrelated. In other words, after all the common factors are controlled for, correlations between the unique factors become zero. On the other hand, residuals v_k, defined as in eqn [5], are linear functions of the same discarded components, and thus correlated.

In order to understand this important property of FA, it is perhaps helpful to introduce the classical example used by psychologist Charles Spearman, who invented the method about 100 years ago. Spearman hypothesized that a variety of tests of mental ability such as scores in mathematics, vocabulary and other verbal skills, artistic skills, and logical reasoning ability could all be explained by one underlying factor of general intelligence (g). If g could be measured and controlled for, there would be no correlations among any tests of mental ability. In other words, g was the only factor common to all the scores, and the remainder was the unique factor indicating the special ability in that area.

When both Z_k and F_j are standardized, the l_{kj} in eqn [5] are standardized coefficients in the regression of variables Z_k on common factors F_j, also termed 'factor loadings'. In other words, the loadings are correlation coefficients between the standardized Z_k and F_j, and thus range from -1 to $+1$. For example, l_{k1} is the loading of variables Z_k on standardized component F_1. Factor loading reflects the strength of relations between variables and factors. Estimates of the factors are termed 'factor scores'. For instance, the estimate for F_j $(j = 1, 2, \ldots, J,$ and $J<K)$ is

$$\hat{F}_j = c_{1j}Z_1 + c_{2j}Z_2 + \cdots + c_{Kj}Z_K \qquad [6]$$

where c_{kj} is the 'factor score coefficient' for the kth variable on the jth factor.

The computing process of FA is more complex than that of PCA. A commonly used procedure for FA is 'principal factor analysis', which utilizes the results from PCA in iterations to obtain the final factor structure and factor scores. The PCA extracts the principal components based on the eigenvalues of correlation matrix **R**, and its computation is straightforward. The principal factor analysis extracts principal factors of a modified correlation matrix $\mathbf{R}^*$. Whereas elements on the major diagonal of **R** are 1, these elements in $\mathbf{R}^*$ are replaced by estimates of 'communality' (a variable's variance shared by the common factors). For instance, the communality of variable Z_k in eqn [5] equals the proportion of Z_k's variance explained by the common factors, that is, $\sum_{j=1}^{J} l_{kj}^2$. After an initial number of factors is decided, the communality and the modified correlation matrix $\mathbf{R}^*$ are estimated through an iterative process until there is little change in the estimates. Based on the final modified correlation matrix $\mathbf{R}^*$ and extracted principal components, the first principal component defines the first factor, the second component defined the second factor, and so forth. In addition to the principal factor analysis method, there are also several alternative methods for FA based on 'maximum likelihood estimation', which is beyond the scope of this article.

How do we decide the number of factors to use? Eigenvalues provide a basis for judging which factors are important and which are not. In deciding the number of factors to include, one has to make a tradeoff between the total variance explained (higher by including more factors) and interpretability of factors (better with fewer factors). One may follow a rule of thumb that only factors with eigenvalues greater than 1 are important. Since the variance of each standardized variable is 1, a factor with $\lambda < 1$ accounts for less variance than exists in a single variable. Such a factor thus does not serve the purpose of data reduction.

Using one as a cutoff value to identify 'important' factors is arbitrary. A 'scree graph' plots eigenvalues against factor number, and provides a more useful guidance. For example, **Figure 1** shows the scree graph of eigenvalues from the case study presented later in this article. The graph levels off after component 3, indicating that factors 4–11 account for relatively little additional variance. Therefore, three components (factors) are retained. Other techniques such as statistical significance testing based on the maximum likelihood FA may also be used to choose the number of factors. However, most importantly, selecting the number of factors must be guided by understanding the true forces underlying the research issue and supported by theoretical justifications.

Initial results from FA are often hard to interpret as variables load unevenly across factors. While fitting the data equally well, 'rotation' generates simpler structure by attempting to find factors that have loadings close to ± 1 or zero. This is done by maximizing the loading (positive or negative) of each variable on one factor and minimizing the loadings on the others. As a result, we can detect more clearly which variable(s) make up each factor(s) latent variable and therefore label the factors adequately. These latent variables represent underlying dimensions that are combinations of many observed variables.

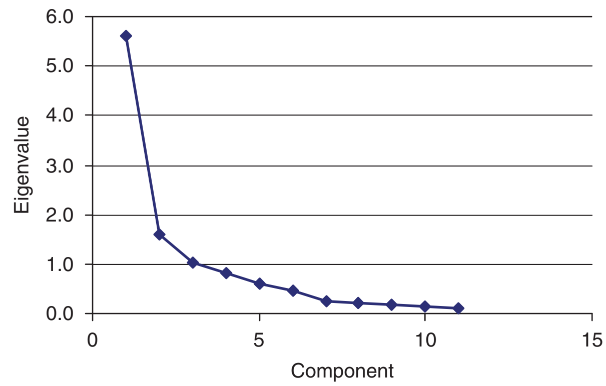

Figure 1 Scree graph for PCA.

'Orthogonal rotation' generates independent (uncorrelated) factors, an important property for many applications (e.g., avoidance of multicollinearity in regression). A widely used orthogonal rotation method is 'varimax rotation', which maximizes the variance of the squared loadings for each factor, and thus polarizes loadings (either high or low on factors). 'Oblique rotation' (e.g., promax rotation) generates even greater polarization but allows correlation between factors.

As a summary, **Figure 2** illustrates the major steps in PCA and FA:

1. the original data set of K observed variables with n records is first standardized;
2. PCA then generates K uncorrelated components to account for all the variances of the K variables;
3. FA generates only J (J< K) principal factors to model interrelations of the original variables; and
4. a rotation method can be used to load each variable strongly on one factor and near zero on the others for easier interpretation.

In essence, PCA attempts to explain all of the variances in the original variables, whereas FA tries to explain their intercorrelations (covariances). Therefore, PCA is used to construct composite variables that reproduce the maximum variance of the original variables, and FA is used to reveal the relationships between the original variables and unobserved underlying dimensions. Despite the differences, PCA and FA share much common ground. Both are primarily used for data reduction, uncovering latent variables and removing multicollinearity in regression analysis. Many people consider PCA a special case of FA.

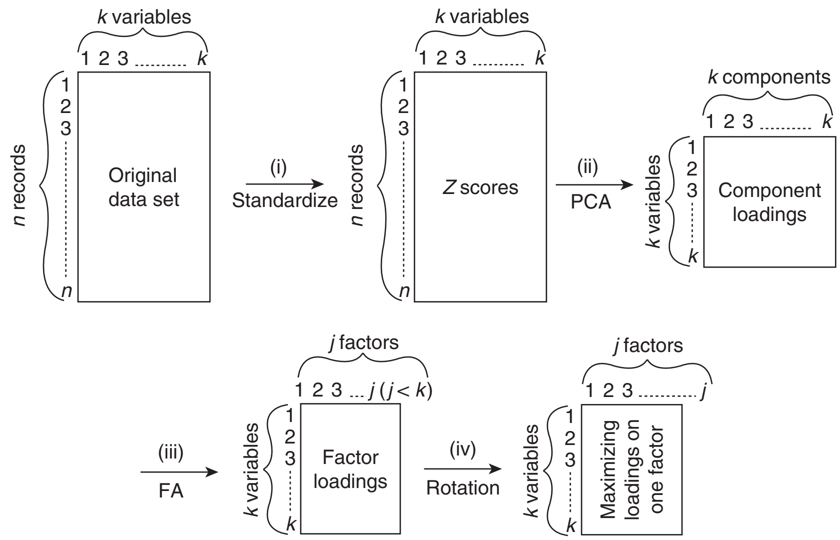

Figure 2 Major steps in PCA and FA.

An Illustrative Example

This section uses an example to illustrate the process of implementing PCA and FA. This example is extracted from a case study of measuring healthcare access. People's access to healthcare is not only affected by their proximity to healthcare providers ('spatial factors') but also by their socioeconomic and demographic characteristics ('aspatial factors'). PCA and FA methods can be used to summarize these socio demographic characteristics of populations that influence access to healthcare. Numerous population characteristics such as age, income, and ethnicity have been suggested to affect healthcare access. These variables are often highly correlated with each other. Using PCA and FA, one can consolidate these variables into a few factors, and then map the resulting factor scores to examine spatial patterns.

The following 11 variables are considered:

1. population with high health needs, including seniors of ages above 65, children of ages 0–4 and women of child bearing ages 15–44;
2. population in poverty;
3. female headed households;
4. home ownership;
5. median income;
6. households with an average of more than one person per room;
7. housing units lack of basic amenities;
8. non white minorities;
9. population without a high school diploma;
10. households linguistically isolated; and
11. households without vehicles.

All the above variables are measured in percentages ex cept for median income which is in dollars. As shown in **Figure 2**, the first step of the analysis is to standardize each variable to a series of z values with a zero mean and a standard deviation of one.

The strong associations among these variables are shown in the correlation coefficient matrix (**Table 1**). Correlation coefficient values higher than 0.50 are high lighted. The table shows a clear cluster of five variables on the top that are highly correlated to each other (all cor relation coefficients higher than 0.5): female headed households, population in poverty, non white population, households without vehicles, and homeownership. Note that a higher homeownership ratio indicates a socio economic advantage and all other four variables represent disadvantages, and thus the variable 'homeownership' correlates with the other four negatively. In addition, we can also identify a secondary cluster of three variables (households with linguistic isolation, households with >1 person per room, and population without high school diploma), with all three correlation coefficients >0.5. This important structure of the data will be further clarified using PCA and FA.

The second step is to use the PCA to help decide the number of component or factors to retain. **Table 2** presents the eigenvalue and corresponding portion of variance associated with each component. Based on the PCA result, three components with eigenvalues >1 are retained, and capture 75% of total variance of the ori ginal 11 variables. The scree graph in **Figure 1** also supports the selection of three components.

After deciding on three factors, FA is conducted to represent each of the three factors by a linear function of the 11 variables (i.e., step (iii) in **Figure 2**). **Table 3** shows the coefficients in the regression of factors on variables, or factor loadings, which indicate the strength of relations between variables and factors. Most variables

Table 1 Correlation coefficients of original variables

Variables	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
1. Female headed households (%)	1.00										
2. Population in poverty (%)	**0.78**	1.00									
3. Non white population (%)	**0.77**	**0.68**	1.00								
4. Households w/o vehicles (%)	**0.67**	**0.78**	**0.72**	1.00							
5. Homeownership (%)	**0.54**	**0.72**	**0.61**	**0.80**	1.00						
6. Housing units lack of basic amenities (%)	0.29	0.36	0.31	0.39	0.35	1.00					
7. Households with linguistic isolation (%)	0.02	0.13	0.20	0.18	0.32	0.19	1.00				
8. Households with >1 person per room (%)	0.45	**0.50**	**0.57**	0.46	**0.51**	0.32	**0.69**	1.00			
9. Population w/o high school diploma (%)	**0.56**	**0.63**	**0.59**	**0.58**	**0.52**	0.35	**0.53**	**0.72**	1.00		
10. Population with high needs (%)	0.13	0.20	0.07	0.24	0.28	0.02	0.08	0.07	0.04	1.00	
11. Median income ($)	**0.52**	**0.65**	0.43	**0.53**	**0.57**	0.25	0.13	0.34	**0.61**	0.36	1.00

Note: Coefficients higher than 0.50 are highlighted in bold.

Table 2 Eigenvalues from principal-components analysis

Component	*Eigenvalue*	*Proportion*	*Cumulative*
1	**5.5944**	**0.5086**	**0.5086**
2	**1.6088**	**0.1463**	**0.6548**
3	**1.0138**	**0.0922**	**0.7470**
4	0.8078	0.0734	0.8204
5	0.6013	0.0547	0.8751
6	0.4594	0.0418	0.9169
7	0.2544	0.0231	0.9400
8	0.2285	0.0208	0.9608
9	0.1622	0.0147	0.9755
10	0.1466	0.0133	0.9889
11	0.1226	0.0111	1.0000

Note: Components with an Eigenvalue >1 are highlighted in bold.

Table 3 Factor patterns without rotation

Variables	*Factor 1*	*Factor 2*	*Factor 3*
Female-headed households (%)	0.7918	0.2767	0.3538
Population in poverty (%)	0.8789	0.2215	0.1167
Non-white population (%)	0.8204	0.0458	0.3013
Households w/o vehicles (%)	0.8609	0.2030	0.0642
Homeownership (%)	0.8238	0.1033	0.1566
Housing units lack of basic amenities (%)	0.4789	0.1219	0.1134
Households with linguistic isolation (%)	0.3854	0.7888	0.3820
Households with >1 person per room (%)	0.7196	0.5640	0.0446
Population w/o high-school diploma (%)	0.8129	0.3066	0.0826
Population with high needs (%)	0.2161	0.5576	0.7059
Median income ($)	0.7064	0.2630	0.2994
Variance explained by each factor	5.5944	1.6088	1.0138

have the highest loading on factor 1 except for two variables: 'households with linguistic isolation' has the highest loading on factor 2, and 'population with high needs' has the highest loading on factor 3.

The final step in PCA (i.e., step (iv) in **Figure 2**) is rotation. After rotation, the loadings are polarized (i.e., the loading of each variable is maximized on one factor but minimized on other factors) for easier interpretation. **Table 4** shows the rotated factor structure by the var imax rotation. The three factors are labeled to reflect major variables captured by each factor. Note that the variance explained by each factor in the rotated factor pattern is different from that in the unrotated factor pattern.

1. Factor 1 is by far the most important factor explaining 40.40% of total variance (i.e., accounting for 54.08% of the variance explained by the three factors). This factor captures six variables: the first five have already been identified as the primary cluster of correlated variables in **Table 1**, and the sixth variable 'housing units lack of basic amenities' has the smallest loading. This factor is a comprehensive indicator of 'socio economic disadvantages'.
2. Factor 2 explains 20.98% of total variance (i.e., ac counting for 28.09% of the variance explained by the three factors). Three variables, including linguistic isolation, load highly on this factor, forming a com prehensive indicator of 'socio cultural barriers'.
3. Factor 3 explains 13.32% of total variance (i.e., ac counting for 17.83% of the variance explained by the three factors). It mainly captures the variable 'popu lation with high needs', as well as 'median income', and therefore, it may be labeled 'high healthcare needs'.

Table 4 Rotated variable loadings on factors

	Factor 1. Socioeconomic disadvantages	*Factor 2. Socio cultural barriers*	*Factor 3. High healthcare needs*
Female headed households (%)	**0.9089**	0.0058	0.0504
Population in poverty (%)	**0.8662**	0.1642	0.2405
Non white population (%)	**0.8481**	0.2153	0.0153
Households w/o vehicles (%)	**0.8231**	0.1905	0.2699
Homeownership (%)	**0.6686**	0.3362	0.3922
Housing units lack of basic amenities (%)	**0.4278**	0.2703	0.0323
Households with linguistic isolation (%)	0.0479	**0.9561**	0.0164
Households with >1 person per room (%)	0.4464	**0.7966**	0.0631
Population w/o high school diploma (%)	0.5800	**0.6406**	0.1219
Population with high needs (%)	0.0316	0.1050	**0.9186**
Median income ($)	0.5491	0.2053	**0.5605**
Variance explained by each factor	4.4436	2.3080	1.4654

Note: The highest loading of each variable on a factor is highlighted in bold.

Application in Social Area Analysis

PCA and FA have been widely used in quantitative human geography for measuring complex concepts like socioeconomic disadvantage. One of the important early applications was for social area analysis, the study of social and spatial differentiation in cities. Social area analysis uses statistical methods to investigate these patterns of residential differentiation within cities. Based on a variety of demographic and socioeconomic variables measured for small areas in a city, social area analysis uses the PCA and FA to consolidate these observed variables into underlying basic constructs (factors). For instance, an idealized factor loading matrix in social area analysis may look like **Table 5**. Based on the table, oc cupation and education form the first factor character izing 'economic status'; fertility rate, women's labor force participation, and single family houses form the second factor labeled '*family status*'; and percentage of minorities makes the third factor 'ethnic status'.

Geographers have made important contributions to the social area analysis by analyzing the spatial patterns associated with these dimensions. Factor scores are mapped to show spatial patterns. Geographers in the 1960s and 1970s demonstrated that in many North American cities socioeconomic status factor tends to exhibit a sector pattern. Family status factor tends to follow a concentric zone pattern in which inner zones are dominated by single young adults and elderly resi dents, and outer zones are most occupied by families with young children. The ethnic status factor tends to form clusters, each of which is dominated by a parti cular ethnic group. PCA and FA provide important tools for summarizing social and spatial variation in complex social, economic, and demographic variables, and the methods have also been used to study cities outside the North America that display very different social patterns.

Table 5 Idealized factor loadings in social area analysis

	Economic status	*Family status*	*Ethnic status*
Occupation	I	O	O
Education	I	O	O
Fertility	O	I	O
Female labor participation	O	I	O
Single-family house	O	I	O
Minorities	O	O	I

Note: 'I' denotes a number close to 1 or 1; 'O' denotes a number close to 0.

Conclusion

PCA and FA methods are primarily used for data re duction. Both seek a few underlying dimensions that account for patterns of variation among many observed variables. Revealed latent variables indicate how the original variables are combined, and thus simplify an alysis and subsequent interpretation. Major advantages of PCA and FA include more parsimonious models, sim plification of data structure, uncovering hidden patterns, and convenience for graphic displays. In most cases (e.g., based on orthogonal rotation), factor (component) scores are independent from each other, and thus avoid multi collinearity in regression analysis.

However, one major challenge for PCA and FA is whether the factors or components generated by the methods are truly underlying dimensions. It is not un common to have variables that spread the loadings on two or more factors (components), leading to difficulties in interpreting the results. Like any statistical analysis methods, the quality of PCA and FA depends on data collection. Often the major factors (components) are dominated by variables that tend to measure the same element of the study object and account for majority of variances, and are susceptible of criticism of 'garbage in and garbage out'.

Further Reading

Bailey, T. C. and Gatrell, A. C. (1995). *Interactive Spatial Data Analysis*. Harlow: Longman Scientific and Technical.

Cadwallader, M. T. (1981). A unified model of urban housing patterns, social patterns, and residential mobility. *Urban Geography* 2, 115 130.

Griffith, D. A. and Amrhein, C. G. (1997). *Multivariate Statistical Analysis for Geographers*. Upper Saddle River, NJ: Prentice Hall.

Hamilton, L. C. (1992). *Regression with Graphics*. Belmont, CA: Duxbury.

Rencher, A. C. (1998). *Multivariate Statistical Inference and Applications*. New York: Wiley.

Wang, F. and Luo, W. (2005). Assessing spatial and nonspatial factors in healthcare access in Illinois: Towards an integrated approach to defining health professional shortage areas. *Health and Place* 11, 131 146.

Relevant Websites

http://www.statsoft.com/textbook/stathome.html
Electronic Statistics Toolbox.

http://cfs.nrcan.gc.ca/subsite/wulder/components
Principal Components Analysis and Factor Analysis.

Fair Trade

L. T. Raynolds, Colorado State University, Fort Collins, CO, USA

Glossary

Agri-Food System The overarching arena encompassing agricultural production and subsequent activities all the way to food consumption; sometimes called the agri-food system.

Alternative Trade Organization A group that purchases items from disadvantaged producers at favorable prices, transports them, and sells them directly to consumers.

Certification Process by which items are assessed for compliance with previously established standards and where compliance is formally acknowledged often through a label or seal.

Global North/South A set of terms distinguishing between a highly industrialized wealthy region and a less industrialized poorer region of the world, which roughly aligns along a North–South axis.

Global Commodity Chain The interlocking production, processing, and distribution activities inherent in moving an internationally traded product from the point of conception to that of consumption; sometimes called a global value chain.

Introduction

Fair trade as a concept refers to a critique of the his torical inequalities inherent in international trade and to a belief that trade can be made more socially just. In the current era, fair trade refers to a set of initiatives that challenge global inequalities and create more egalitarian commodity networks linking marginalized producers in the global South with progressive consumers in the global North. Fair trade seeks to alleviate poverty and empower producers in the global South through the provision of better prices, stable market links, and other material and informational resources. Fair trade also seeks to bolster responsible consumption practices among Northern consumers by encouraging the purchase of items that are produced and traded under more equitable conditions. At the broadest level, fair trade seeks to eliminate North–South trade inequalities.

In creating alternative commodity networks fair trade joins a growing array of market based initiatives that promote social and environmental concerns through the sale of alternative often certified, commodities. Trade in certified fair trade products is relatively modest in relation to the global market, with sales worth somewhat over $1 billion. But sales are expanding rapidly as fair trade extends into new commodities, new production regions, and new markets. In its advocacy to make global trade more equitable, fair trade forms part of the growing alternative globalization/trade justice move ment. In fact fair trade has emerged as one of the key rallying points and constructive engagements within this broad movement.

Since fair trade addresses issues of inequality in ex changes between distant peoples and places, it is of central interest within geography as well as related social science disciplines. Fair trade lies at the intersection of economic, political, and social geographies and has been fruitfully explored from all these angles. Because fair trade markets and movements are centered in the agro food sector, scholarship within this domain has proved particularly insightful. In analyzing the importance of fair trade, this article focuses on: (1) fair trade initiatives, ideas, and institutions; (2) fair trade distribution and consumption; and (3) fair trade production and exports.

Fair Trade Initiatives, Ideas, and Institutions

The modern fair trade movement grows out of variety of post World War II (WWII) initiatives, which have sought to challenge the inequalities of international trade and assist disadvantaged populations through a strategy of 'trade not aid'. Initial efforts by church and development groups focused on assisting impoverished peoples by providing preferential markets for their handicrafts. Al ternative trade organizations (ATOs) were established, which purchased items at favorable prices, shipped them, and sold them to dedicated consumers. These efforts blossomed with growing Third World solidarity move ments in the 1960s and 1970s, evidenced particularly in Europe with the proliferation of 'World Shops'. ATO initiatives were informed by popular theoretical views of the time regarding the role of unequal trade in main taining Third World dependency. Within this context, ATOs created direct trade channels eliminating exploit ative international merchants and increasing the returns to producers.

A distinct new strand of fair trade was established in the late 1980s with the introduction of product certifi cation and labeling. Fair trade groups initiated this strat egy to move beyond the craft sector and solidarity shops to sell major food exports like coffee in conventional supermarkets. The move into the certification and labeling of food items was fueled by the recognition of the con tinued centrality of agro export production for poor people and regions of the global South and the rise of specialty markets in the global North for products dif ferentiated based on their social or ecological standards. Fair trade certification has been highly successful in ex panding market and production arenas.

By 2004, fair trade labeling initiatives operated in 20 countries across Europe, North America, and the Pacific coordinated via the umbrella organization, the Fair Trade Labeling Organizations International (FLO). Eighteen commodities were sold with FLO labels, produced by 433 groups across Latin America, Africa, and Asia. FLO cer tification guidelines are based on a set of core require ments for producers and buyers. Fair trade producers must be organized into democratic groups of small growers or estate workers who uphold high social standards and more modest environmental standards. Fair trade buyers must guarantee payment of FLO minimum prices, pay a social premium, offer producer credit and long term contracts, and trade as directly as possible with producers. While fair trade certification resembles other labeling initiatives regulating ecological conditions in food or timber or labor conditions in garments or textiles, fair trade is the only such initiative that goes beyond establishing production standards to address trade conditions. Though the ATO and labeling strands of fair trade are distinct, an informal alliance (FINE) promotes their common interests and has forged a joint definition of fair trade.

"Fair trade is a trading partnership, based on dialog, transparency, and respect that seeks greater equity in international trade. It contributes to sustainable devel opment by offering better trading conditions to, and se curing the rights of, marginalized producers and workers – especially in the South. Fair trade organizations (backed by consumers) are engaged actively in sup porting producers, awareness raising and in campaigning for changes in the rules, and practice of conventional international trade."

As this statement suggests fair trade operates both 'in and against the market', working through market channels to create new commodity networks for items produced under more favorable social and ecological conditions and simultaneously working against the conventional market forces that create and uphold global inequalities.

Research on fair trade's operation within the market focuses largely on the alternative international networks created and maintained by these initiatives. Fair trade establishes a new spatial ordering of international eco nomic relations. From a political economy perspective, global commodity chain analysis provides a very in sightful avenue for analyzing the institutional and spatial configurations, power relations, and patterns of resource distribution within fair trade networks. This approach reveals the internal architecture of the relations linking fair trade producers and consumers. From a sociocultural perspective, fair trade networks are understood in more relational and fluid terms, emphasizing the ways in which diverse actors negotiate the movement of commodities and associated material and nonmaterial resources across space. Fair trade can thus be seen as a dense network of connectivity co created by locally situated, but globally connected, people, places, ideas, and things. Together political economy and sociocultural analyses of fair trade illuminate an alternative global geography, which is strikingly different from the conventional corporate driven global topography.

Network analyses highlight the degree to which fair trade resembles other alternative, often local, agro food networks, which promote alternative values, practices, and institutions to 'shorten the distance' between pro ducers and consumers. These alternative food networks hinge on personal relations of trust and dense infor mation flows which humanize economic transactions. Rather than being abstracted from their roots, commod ities are embedded with information regarding producers' ideas and practices and the places/spaces of production. While ATOs are able to bolster fair trade relations dir ectly via face to face interactions with producers and consumers, within certified fair trade networks transac tions are mediated by increasingly formalized rules, standards, and product labeling procedures which may threaten these alternative relations.

Research on fair trade's operation against the market focuses on the location of these initiatives within the broader alternative globalization/trade justice movement. In recent years, a number of collective efforts have emerged that coalesce around a critique of corporate globalization and a set of engagements to forge an alter native globalization based on embedded, territorially specific, social and ecological rights and values. From a social movement perspective, fair trade should be analyzed within this broader terrain. Fair trade has clear insti tutional and ideological ties with related efforts like antisweatshop and social movement union initiatives in manufacturing, and agro ecology, food sovereignty, and land reform initiatives in agriculture. What is striking about the new movement cartography is the importance of nongovernmental organizations in local, national, and transnational arenas. Nongovernmental organizations as sociated with fair trade and related initiatives have become a powerful force pressuring corporations to improve their practices and creating noncorporate alternatives.

Fair Trade Distribution and Consumption

The total world market for fair trade products, including ATO and labeled sales, was valued at roughly US$1.2

billion in 2004. Though this represents a minor fraction of global trade, this market is growing at roughly 30% per year due to mounting consumer interest and the expansion of fair trade circuits. The fair trade boom is part of a worldwide increase in conscientious and green consumption. There has been a dramatic rise in pur chases of ethical and eco friendly products across major markets in the global North. A growing body of research demonstrates the importance of consumption as a space for individual and collective identity creation and of political action. A political economy of consumption is particularly important in researching national and international food supply chains. This approach chal lenges the historically productionist focus in geography as well as other related social science fields.

The political nature of consumption ideas and prac tices is particularly evident in agro food arenas where ecological and food safety threats have undermined trust in the quality and safety of conventional agro industrial production and products. Consumption of organic, local, fair trade, and other specialized food items is growing rapidly. As in fair trade, many of these new food products are certified to guarantee their special quality attributes and labeled to bring these attributes to consumers' at tention since typically they are not visible in the product itself. While some certification and labeling efforts rep resent nothing more than a corporate face lift promoted by branders and retailers, others are overseen by com mitted nongovernmental organizations and involve a fundamental refashioning of the conventions upon which commodity networks are based. Rising consumer unease about the social and environmental conditions embedded in their purchases has fueled the market for third party certified products, particularly for those sourced inter nationally where distance heightens consumer concerns over production impacts and over potentially unfounded corporate ethical and ecological claims.

Food commodities certified as fair trade under the FLO system account for the vast majority of the recent rapid growth in fair trade sales. As noted in **Table 1**, the fair trade certified market was worth over US$1 billion in 2004. Though ATO sales continue to be important largely in nonfood items and are even growing in some regions, analysis focuses on certified fair trade distri bution and consumption patterns given this sector's much larger size and greater data availability. Europe has the most well developed and largest fair trade market, led by the United Kingdom and Switzerland. A range of fair trade products are readily available in government offices, universities, supermarkets, and spe cialty outlets. While fair trade markets are growing only modestly in some European countries, in countries like the United Kingdom sales continue to rise rapidly. North American fair trade markets are much less well developed than those in Europe. Certified products were introduced more recently, in the late 1990s, and there are still fewer fair trade products available and fewer venues which carry these items. Yet, the United States has emerged as the largest national fair trade market. The United States fair trade sales were valued at close to US$300 million in 2004, with earnings doubling each year. Fair trade labeling organizations are just beginning to get established in Pacific markets, in Japan, Australia, and New Zealand.

Coffee, the first labeled commodity, forms the core of the fair trade system. Fair trade certified coffee accounts

Table 1 Fair trade certified sales by region and commodity (2003/04)

	Volume (metric tons)					
	Coffee	*Tea*	*Cocoa*	*Bananas*	*Total*[a]	*Total*[a] *Value (US$ 1000)*
Europe						
UK	2889	1090	903	18 177	24 212	256 000
Switzerland	1550	36	275	19 002	23 336	169 000
Netherlands	3096	60	147	2610	5998	43 000
Germany	2865	156	343	116	4217	72 000
France	2368	52	227	829	4059	87 000
Italy	230	10	346	2038	3330	31 000
North America						
US	3574	52	92		3718	267 000
Canada	648	6	54		738	22 000
Pacific						
Japan	22	8			30	3000
Total[b]	19 895	1990	3473	51 336	83 297	1 033 000

[a] Includes sales of all fair trade labeled commodities, not just the listed core commodities.
[b] Includes countries not listed on the chart.
Source: Compiled by the author from FLO data.

for roughly a quarter of all earnings and is by far the most widely available product. Rising consumption of fair trade coffee is linked to the growing differentiation of products, including the growth of gourmet and organic certified varieties, and the differentiation of consumption experiences, including the ubiquitous café chains that define the upscale urban marketplace. The United States has by far the largest and most dynamic fair trade coffee market. The Netherlands, United Kingdom, Germany, and France also have substantial markets for fair trade certified coffee.

Cocoa and tea are also well established core fair trade certified commodities. Fair trade cocoa has the larger market with sales concentrated in Europe, particularly the United Kingdom. Fair trade cocoa is not yet widely available in North America, though the market is grow ing rapidly in the United States. Fair trade tea sales are more modest and are growing relatively slowly except in the United Kingdom, which dominates the market. Though markets are quite small, FLO certification has also long been available for sugar.

Bananas are now the second most valuable fair trade product, despite the fact that certification was introduced much more recently than in other major commodities and was not initiated until 2004 in the United States. Switzerland and the United Kingdom have the largest fair trade banana markets, and this product is widely available in major supermarket chains in these countries. The growth potential for fair trade in this popular fruit is enormous, though the distribution of highly perishable bananas is complex. Bananas represent the entry of fair trade certification and labeling into the fresh produce sector, and success in this area has spawned the launch of a number of additional FLO labeled fresh fruits like citrus, grapes, and mangoes.

Fair Trade Production and Exports

Fair trade production and exports have grown dramat ically across the global South. Bourgeoning demand in major markets has spurred the growth and diversification of fair trade exports. This growth has been fueled also by neo liberal policies around the world, which have in creased the competition in conventional markets and necessitated a search for nontraditional niche markets. Within this context, producers across the global South have sought out avenues for differentiating their products according to social or ecological attributes and where possible gaining access to circuits that can provide some semblance of security within volatile global markets. Certified fair trade agro food production and exports have grown within this context, alongside organic food, certified timber, and related sectors. While the vast majority of fair trade certified items are shipped from the global South to major markets in the global North, there are nascent national markets developing in Mexico, Brazil, South Africa, and other countries, which can provide market openings for domestic and regional producers.

There are roughly 544 certified fair trade producer organizations, representing hundreds of thousands of small scale growers and farm workers, in 53 countries across Latin America, Africa, and Asia. **Table 2** outlines fair trade's production geography. Latin America repre sents the clear hub of certified fair trade production with by far the largest number of certified groups and export volumes. This is perhaps not surprising since fair trade certification was developed in this region, with the initial certification of fair trade coffee from Mexico. Latin America has 187 FLO certified coffee farmer groups, which together supply 83% of all global fair trade coffee

Table 2 Fair trade certified production by region and commodity (2004/05)

	Coffee	*Banana*	*Cocoa*	*Tea*	*Other*[a]	*Total*
Latin America						
Certified producer groups	187	19	15	1	82	314
Production volume (metric tons)[b]	24 900	64 700	1900	0	11 200	102 700
Africa						
Certified producer groups	33	1	4	17	97	152
Production volume (metric tons)[b]	4400	1300	2100	1600	21 200	30 600
Asia						
Certified producer groups	8	0	0	48	22	78
Production volume (metric tons)[b]	800	0	0	800	700	2300
Total						
Certified producer groups	228	20	19	66	211	544
Production volume (metric tons)[b]	30 100	66 000	4000	2400	33 100	135 300

[a] Includes sugar, honey, other fruits and juices, rice, wine, nuts, flowers, spices, and sport balls.
[b] Since some items are not measured in tons, the other and total columns are suggestive only.
Source: Compiled by the author from FLO data.

supplies. Mexico remains the region's and the world's largest fair trade coffee producer. Latin America is also the dominant producer of fair trade certified bananas, accounting for 98% of world supplies. There are 19 FLO certified banana groups representing both small scale farmers and workers on plantation enterprises. Latin America is also a major supplier of fair trade certified cocoa. A growing number of groups in the region are becoming involved in the production of new fair trade commodities like mangoes, orange juice, and quinoa.

Africa is the second most important supplier of fair trade products with 152 FLO affiliated groups. This re gion is the largest source of fair trade certified cocoa and tea and an important producer of coffee, with Ghana and Tanzania leading the production. Yet the importance of Africa in fair trade circuits in the future looks not to be in these core commodities but in the production of newly certified product areas. South Africa has recently emerged as the region's fair trade leader with 52 FLO registered groups and is now the world's largest exporter of certified fresh and processed fruits, juices, and wine. Though fair trade producers in Africa are engaged in small and large enterprises, almost all of the recent growth is within the plantation sector.

For small scale producers, the most direct benefits from fair trade come from the higher guaranteed prices. The importance of these price guarantees is clearest in the case of coffee where the FLO minimum price has far exceeded the world market price for most of the past 15 years. This price floor has meant the difference between survival and bankruptcy for many small scale coffee growers. In addition to protecting producers from world price slumps, fair trade provides a social premium to be invested in community projects. The fair trade premium supports critical education, health, food self sufficiency, transportation, and farm improvement projects. For lar ger enterprises, the FLO price floor provides economic security. But the social premium is much more important for workers, since this is what provides the funding for purchasing ownership shares and supporting educational, health, and housing projects. Research suggests that for both producers and workers, the most important benefits of fair trade engagement appear in the long run to come from the multifaceted informal and formal support pro vided for organizational capacity building.

Though the potential benefits of fair trade for pro ducers and workers in the global South are substantial, the increased geographic spread, product diversification, and enterprise variation within certified networks makes realizing those benefits increasingly difficult. The spread of certified fair trade production across dozens of coun tries requires adjusting this model to varied local political economic circumstances. Local farmer and worker or ganizations vary tremendously as does fair trade's fit within local social movements. There are important synergies between fair trade and various social justice and ecological initiatives around the world. But there are also critical tensions across the global South over fair trade's continued emphasis on export production for Northern markets, instead of local food self sufficiency goals.

Conclusions

At heart, the fair trade concept refers to a critique of the historical inequalities inherent in international trade and to a belief that trade can be made more socially just. This article has focused in large part on outlining the par ameters of the recent challenges raised against neo lib eral globalization and the types of alternative material and ideological networks created by fair trade. The true significance of fair trade lies not in its albeit impressive markets, but in the extent to which this movement de stabilizes conventional market principles. Conventional capitalist markets are guided by prices that understate the full ecological and social costs of production, and thus encourage the degradation of environmental and human resources, particularly in the global South. The fair trade movement makes visible the ecological and social relations embedded within a commodity and asks that consumers shoulder the true costs of production. Fair trade shortens the distance between Southern pro ducers and Northern consumers through dense infor mation as well as resource flows. By demystifying global trade and creating more equitable relations of exchange, the fair trade movement goes further than other social or ecological certification initiatives in challenging con ventional market practices. Fair trade is without a doubt a critical component of the current and future cartography of alternative globalization.

See also: Dependency; Development I; Development II; Food Networks, Alternative; Gender and Rurality; Neoliberalism and Development; Nongovernmental Organizations; Regional Production Networks; Social Justice, Urban; Social Movements.

Further Reading

Barrientos, S. and Dolan, C. (eds.) (2006). *Ethical sourcing in the global food chain*. London: Earthscan.
Brown, M. (1993). *Fair trade*. London: Zed Press.
Gereffi, G., Garcia Johnson, R., and Sasser, E. (2001). The NGO industrial complex. *Foreign Policy* July/August: 56 65.
Murdoch, J., Marsden, T. and Banks, J. (2000). Quality, nature, and embeddedness: Some theoretical considerations in the context of the food sector. *Economic Geography* 76, 107 125.
Mutersbaugh, T. (2002). The number is the beast: A political economy of organic coffee certification and producer unionism. *Environment and Planning A* 34, 1165 1184.
Nicholls, A. and Opal, C. (2005). *Fair trade: Market driven ethical consumption*. New York: Sage.

Ransom, D. (2001). *The no nonsense guide to fair trade*. London: Verso.
Raynolds, L. (2000). Re embedding global agriculture: The international organic and fair trade movements. *Agriculture and Human Values* 17, 297 309.
Raynolds, L. (2002). Consumer/producer links in fair trade coffee networks. *Sociologia Ruralis* 42, 404 424.
Raynolds, L. (2004). The globalization of organic agro food networks. *World Development* 32, 725 743.
Raynolds, L., Murray, D. and Taylor, P. (2004). Fair trade coffee: Building producer capacity via global networks. *Journal of International Development* 16, 1109 1121.
Raynolds, L., Murray, D. and Wilkinson, J. (2007). *Fair trade: The challenges of transforming globalization*. London: Routledge.
Renard, M. C. (1999). The interstices of globalization: The example of fair coffee. *Sociologia Ruralis* 39, 484 500.
Renard, M. C. (2003). Fair trade: Quality, market and conventions. *Journal of Rural Studies* 19, 87 96.
Whatmore, S. and Thorne, L. (1997). Nourishing networks: Alternative geographies of food. In Watts, M. & Goodman, D. (eds.) *Globalising food: Agrarian and global restructuring*, pp 287 304. New York: Routledge.

Relevant Websites

http://www.colostate.edu/depts/sociology/cfats
Center for Fair & Alternative Trade Studies.
http://www.fairtrade.net
Fair Trade Labelling Organizations International.
http://www.ifat.org
International Federation for Alternative Trade.

Famine

P. J. Atkins, Durham University, Durham, UK

Glossary

Coping Strategies The strategies and tactics that people employ to minimize the stress associated with livelihood crises and other events.
Famine A food crisis resulting in major excess mortality and widespread severe, acute malnutrition.
Food Insecurity Unavailability of food due to insufficient purchasing power, inappropriate distribution, or inadequate utilization at the household level.
Livelihoods The ways in which people make a living, including various means of earning income, goods in kind, or access to resources and services.
Resilience The ability of people to recover from stress and to protect themselves against future stresses.
Vulnerability The degree to which a group (family or community) is likely to experience harm due to exposure to stress.

Introduction

As a word, 'famine' is so emotive that it affirms the affectual power of language and, along with iconic images of starving children, it can change the world. An example is Michael Buerk's BBC/NBC television report from Korem on the Ethiopian famine in 1984. An estimated 470 million people around the world watched horrified as the voice over described the scene: Dawn, and as the sun breaks through the piercing chill of night on the plain outside Korem it lights up a biblical famine, now, in the 20th century. This place, say workers here, is the closest thing to hell on earth. Thousands of wasted people are coming here for help. Many find only death (Michael Buerk, 1984).

And yet famine's sensational headline potential can get in the way of thinking about the more general issues of malnutrition and food insecurity. Recent research on famine sees it as the extension of structural, long term vulnerabilities in society rather than as the result of unpredictable 'acts of God'.

Definitions: The F-Word

Popular definitions of famine such as 'an increase in death rates due to regional mass starvation' have proved inadequate for a number of reasons. First, increased mortality is not a universal feature of famines, and deaths anyway are more usually the result of disease than of starvation. Second, famines are certainly limited in space but at what scale? Are food shortages in a few villages enough to declare a famine or should we reserve the term for a region, province, or a whole country? Third, famines vary in their causality and structure, so should an element of explanation be added to that of description, for instance 'drought famine', or should the terminology encompass 'complex emergencies', which may have sev eral elements, including war? Fourth, famines are time limited. Most commentators seem to agree that they last only for a matter of months, or perhaps a year or two, whereas the chronic global problem of undernourishment is a long term one affecting 800 million people, many of whom have accompanying deficiency diseases and im paired immune systems. So, is famine qualitatively dif ferent from hunger and not just the opposite of food security? Fifth, famines sometimes happen when there is food available but it is expensive or inaccessible to those in need. Finally, famine is often a symptom of broader issues of poverty and is therefore embedded in livelihood systems generally. A better definition might be 'a wide spread, acute crisis of livelihoods'.

However, such academic cavils and caveats need to be presented carefully if they are not to seem churlish. Those of us debating definitions need to remember that the politicians who control the purse strings of food aid are most clearly moved by the label famine, coupled with high media visibility. This leads to the argument some times articulated that famines should only be declared for large scale, acute events because to do so more fre quently would 'debase the coinage'.

Definitions remain difficult and controversial. Recent food emergencies, concentrated mainly in sub Saharan Africa (SSA), have shown that consensus is difficult to achieve even among non governmental organizations (NGOs), let alone between governments and inter national aid agencies such as the World Food Program (WFP). This is partly because these organizations have different, and sometimes contradictory, interests, but there are also technical issues of measurement that need to be addressed.

Measurement: Intensity, Magnitude, and Early Warning

The social dislocation accompanying famines, especially when they are associated with war or disasters such as floods, often means that the measurement of consequences is difficult. The spatial characteristics of famine, such as extent and localized concentration, are complex and usually dynamic, for instance when migration follows livelihood failure. A crisis may move from hunger to famine through a series of stages, and if the early signs can be recognized, then there is a chance that excess mortality can be averted later on. In the early stages of a famine, people's primary goal is to avoid the loss of land, livestock, and equipment. The commonest coping strategy at this point is self imposed food rationing. Second, there is dietary change, such as the consumption of wild foods, immature crops, and even seeds. Third, families may send away some of their number, for instance children to be fed by neighbors or relatives who still have food. Fourth, taking loans or selling assets may become necessary, which may not be reversible at a later stage, followed possibly by begging or migration to locations where external support is available (food aid). The unfolding of these coping strategies is context specific, so their use as general famine descriptors and predictors is problematic. A recent development is CARE/WFP's Coping Strategies Index, which is being used for both monitoring and intervention purposes.

It has been suggested that measurement should focus on intensity (**Table 1**) and magnitude (**Table 2**). The criteria (mortality, wasting, and edema (swelling due to fluid retention)) are well known indicators of food problems, although the thresholds between the various categories are arbitrary to some extent. Such an approach is subject to all of the problems that afflict the measurement of socioeconomic phenomena in space, the most obvious being scale of analysis.

Famine early warning systems (FEWSs) are an aspect of measurement. The Global Information and Early Warning System of the Food and Agriculture Organization (FAO) was established in 1975, and the interagency Food Insecurity and Vulnerability Information and Mapping Systems in 1996. The United States Agency for International Development (USAID) sponsored FEWS NET started in 1985. In addition to global FEWS, many countries have their own systems, as do the large NGOs. These provide regularly updated lists of regions said to require assistance, based upon various methods of declaring alerts. Geographical Information Systems (GIS) cartographies of famine and food insecurity have become central to these methodologies, along with satellite remote sensing which is used to show rainfall and also to delineate areas of drought anomaly on 'greenness maps' of vegetation.

Most FEWS products have hitherto been based on a belief that food availability decline (FAD) causes famines and the application of the methodology has been top down. However recently, livelihood mapping and food economy reports have been added, enabling grassroots needs assessments of vulnerable groups within a regional geography of zones that are homogeneous in terms of

Table 1 Famine intensity

Level	*Crude mortality rate (per 10 000 per day)*	*Wasting (%)*	*Edema*	*Livelihoods: food security descriptors*
0. Food security	<0.2	<2.3		Social system cohesive; prices stable; negligible adoption of coping strategies
1. Food insecurity	0.2 0.5	2.3 10		Social system cohesive; price instability, and seasonal shortage of key items; reversible 'adaptive strategies' employed
2. Food crisis	0.5 1.0	10 20	✓	Social system significantly stressed but remains largely cohesive; dramatic rise in price of food and other basic items; adaptive mechanisms start to fail; increase in irreversible coping strategies
3. Famine	1 5	20 40	✓	Clear signs of social breakdown; markets begin to close or collapse; coping strategies exhausted and survival strategies adopted; affected population identify food as the dominant problem in the onset of the crisis
4. Severe famine	5 15	>40	✓	Widespread social breakdown; markets closed or inaccessible to affected population; survival strategies widespread; affected population identify food as dominant problem in onset of crisis
5. Extreme famine	>15	>40	✓	Complete social breakdown; widespread mortality; affected population identify food as dominant problem in onset of crisis

From Howe, P. and Devereux, S. (2004). Famine intensity and magnitude scales: A proposal for an instrumental definition of famine. *Disasters* 28, 353 372.
Note: wasting defined as the proportion of the children 6 months to 5 years who are below 80% of the median weight for height.

Table 2 The magnitude scale

Category	*Phrase designation*	*Mortality range*
A	Minor famine	0 999
B	Moderate famine	1000 9999
C	Major famine	10 000 99 999
D	Great famine	100 000 999 999
E	Catastrophic famine	≥1 000 000

From Howe, P. and Devereux, S. (2004). Famine intensity and magni tude scales: A proposal for an instrumental definition of famine. *Dis asters* 28, 353 372.

natural resources and livelihood profiles. This approach is based upon the insight that famine does not begin to kill people until it has destroyed their livelihoods. Spatial analysis, then, has been enthusiastically embraced by a range of institutions, and maps are regularly used in food emergencies. Since 1994, the WFP has refined its Vul nerability Analysis and Mapping to the point where it is used regularly for food aid targeting and famine relief. Baseline vulnerability and food security assessments feed data into WFP disaster mitigation programs.

The scientific prediction and identification of famine has, of course, to be backed up with the political will to act upon the warning, but this has not always happened, bringing the technocratic, cartographic paradigm of fam ine relief into disrepute. The WFP rarely receives all of the funding it requires for its flash appeals (80% is the average) because politicians need incentives to act.

A final aspect of measurement, retrospective dem ography, comes if mortality cannot be estimated in real time. Apart from the obvious increase in deaths, there is usually also an immediate reduction in fertility due to starvation induced amenorrhea (menstrual failure). Both are clearly visible in the population pyramid of China, which suffered a horrific famine between 1958 and 1962. Postfamine populations also tend to have a higher pro portion of girls and women than normal, for complex reasons that are probably to do with bodily resistance to stress. There are also more in the age groups 10–45 be cause of the adverse impact upon vulnerable younger and older people.

Famine: A Conceptual Genealogy

The late eighteenth century argument of Thomas Mal thus was that populations (he meant working people) were at risk of famine because they expand faster than their means of sustenance. Although disproved in the nineteenth century by increased crop yields and the es tablishment of a worldwide trade in foodstuffs, this idea was recycled in the late twentieth century, when it be came a negative commentary on the ability of post colonial states in the Third World to meet the basic needs of their rapidly growing populations. There are subsidiary, ongoing debates that global resources are somehow insufficient to accommodate further economic and demographic expansion, and that the ever growing human imprint will irrevocably damage the environment.

Unfortunately for the many scholars with genuine concerns about sustainability, or who wish to empower women with regard to choices about family size, the ex cesses of the neo Malthusians have undermined the 'too many people' strand of thinking. Their predictions of dire consequences have often been exaggerated and their ethics questionable. Examples include the frankly reprehensible 'Lifeboat ethics: the case against helping the poor' (1974) of Hardin, and the triage advocated in *Famine 1975! America's decision: who will survive?* (1967) by the Paddock brothers. Even the science assumed by Malthus and his followers, that famines represent a 'check' on population growth, is disproved by demographers. The evidence overwhelmingly demonstrates that fertility rebounds after famines and that numbers eventually recover.

In the nineteenth century the discussion of famine was influenced by catastrophes in Ireland and India. In the former, the fungal disease *Phytophthera infestans* (potato blight) ruined the crops of 1845, 1846, and 1848, causing an excess mortality estimated at 1 million. The FAD for individual households (at the same time that food was being exported from Ireland to Britain), coupled with widespread evictions, caused destitution that the state only half heartedly dealt with through public works. The consequences, such as a diaspora of 2 million Irish emi grants, are well known but the recent cluster of publi cations inspired by the 150th anniversary of the 'potato famine' has added value through a number of powerful 'famine imaginaries'.

The Indian famines were in the Delhi region, 1860–61; Cuttack, 1865–66; West Bengal, 1873–74; and south ern India, 1877; and various experiments with relief eventually culminated in the systematic response of the Famine Codes. These were first formulated in 1883 and went through various guises in the different provinces. In essence, they were a precursor of the modern FEWS, using a variety of indicators of food insecurity (crop failure, market prices) as triggers for preplanned re sponses, such as food for work, food distribution through handouts, soup kitchens, or relief from taxes. Under standable questioning of the motivations of such colonial surveillant biopolitics should not blind us to the con ceptual significance of this development.

Ironically, although in the modern age the capacity to deal with famine has been maximized through enhanced production, better transport, and the initiation of moni toring, nevertheless the twentieth century saw the greatest number of famine deaths in history. Famine re lated mortality (not all from starvation) has been vari ously estimated, but a total of 70 million seems to be very likely (**Table 3**).

Table 3 Twentieth-century famines: excess mortality in thousands

Years	*Location*	*Deaths*	*Causes*
1899 1901	NW India	1000	Drought
1903 06	Hausaland, Nigeria	5	Drought
1906 07	S Tanzania	37.5	Conflict
1913 14	Sahel, W Africa	125	Drought
1917 19	C Tanzania	30	Conflict and drought
1920 21	Gansu and Shaanxi, China	500	Drought
1921 22	USSR	9000	Drought and conflict
1927	NW China	3000 6000	Natural disasters
1929	Hunan, China	2000	Drought and conflict
1932 34	Ukraine, USSR	7000 8000	Government policy
1943	Henan, China	5000	Conflict
1943	Bengal, India	2100 3000	Conflict
1943 44	Rwanda	300	Conflict and drought
1944	Netherlands	10	Conflict
1946 47	USSR	2000	Drought and government policy
1957 58	Tigray, Ethiopia	100 397	Drought and locusts
1958 62	China	30 000 33 000	Government policy
1966	Wollo, Ethiopia	45 60	Drought
1968 70	Biafra, Nigeria	1000	Conflict
1969 74	Sahel, W Africa	101	Drought
1972 73	Maharashtra, India	130	Drought
1972 75	Wollo and Tigray, Ethiopia	200 500	Drought
1974 75	Somalia	20	Drought and government policy
1974	Bangladesh	1500	Flood and market failure
1979	Cambodia	1500 2000	Conflict
1980 81	Uganda	30	Conflict and drought
1982 85	Mozambique	100	Conflict and drought
1983 85	Ethiopia	1000	Conflict and drought
1984 85	Darfur and Kardofan, Sudan	250	Drought
1988	S Sudan	250	Conflict
1990 2003	Iraq	200 1000	Sanctions
1991 93	Somalia	300 500	Conflict and drought
1995 99	North Korea	2800 3500	Flood and government policy
1998	Bar el Ghazal, Sudan	70	Conflict and drought
1999 2000	Ethiopia	20 98	Drought and conflict

Devereux, S. (2000). *Famine in the twentieth century, IDS working paper 105*. Brighton: University of Sussex.

Recent work on famines in geography and cognate disciplines can be divided into four categories. First, a dominant theme has been that of 'natural' hazards: famines associated with the destruction of crops or disruption of infrastructure that comes with floods, droughts, or plagues of locusts. The word natural was eventually questioned as we learned more about human impact, for instance of overgrazing upon desertification or deforestation upon floods. In the new millennium, hazards analysis has renewed in popularity due to the need to understand and mitigate the extreme events predicted to come with 'global warming'.

Second, we can identify a humanitarian turn in attitudes to famine. The formation of Oxfam in 1942 and other international NGOs after the war was paralleled by an increasing commitment by Western governments to food aid, and the formation of the WFP by the United Nations (UN) in 1961. It is possible to be cynical about the strategic aims of food aid donors and to argue that they should pay more attention to protecting livelihoods, but the need for short term emergency interventions remains. An interesting development of antifamine action in recent decades has been the involvement of pop musicians in high profile events, such as the Concert for Bangladesh (1971), BandAid (1984), and LiveAid (1985).

The humanitarian theme has recently achieved greater intellectual rigor as the result of a debate on the 'right to food'. This is already an inalienable human right under international law, recognized in the Universal Declaration on Human Rights and the International Covenant on Economic, Social, and Cultural Rights, but its practical application is still at an early stage of development. In 2004, the FAO Council adopted the Right to Food Guidelines but these are voluntary. Some activists have suggested that modern famines are always political to some extent and that it is therefore possible to argue legally that there are accountable parties who are liable, either for negligence or in the worst cases for genocide, requiring prosecution in the International Criminal Court. This has dramatically shifted the center of gravity of famine studies from 'what is the cause of famines?' to 'who benefits from famines?'. One possible crime might be famine denial. There was no word, for instance, from China during the Great Leap Forward (1958–61) to indicate that Mao's radical restructuring of the economy was so disruptive that it led to upward of 30 million deaths, one of the greatest famines in human history. In Sudan, on a much smaller scale, the govern ment similarly failed to acknowledge a drought related famine in 1984–85, a tragic mistake that cost a quarter of million people their lives.

Third, political economy interpretations of famine have tended to look at the dysfunctional role of market mechanisms (taxes, hoarding speculation, excess profits, market failure) or at the role of class, privilege, and re pression in the global South. Increasingly, these factors have been tied into the processes of globalization, such as the terms of international trade or the geopolitical am bitions of the great powers. One of the best writers in this tradition, Michael Watts, has made a series of telling interventions. His work is especially important because of its contextual exploration of the nature of food insecurity. He has conceptualized this as a function of 'spaces of vulnerability' (**Figure 1**), based on: (1) entitlement and capability – market perturbations (economic exchange), coping thresholds (socioeconomics of resilience), and social security limitations (informal moral economies or formal welfare institutions); (2) empowerment and en franchisement; and (3) political economy – class and crisis. Watts's conclusion is that we need a realist theory of famine as sketched in **Figure 2**. Political ecology is another version of this approach, representing in effect the environmentalism that was absent from Marxist geography until the 1990s.

Fourth, the FAD argument, that hitherto had seemed to be the commonsense explanation of famine, was chal lenged in 1981 by Amartya Sen. He observed that famines are often documented in places where food is available. His examples were the Bengal famine of 1943; Ethiopia, 1972–74; the Sahel, 1973; and Bangladesh, 1974. Sen's insight was that people's (individuals, households) access to food, their 'exchange entitlement', was what mattered and that this could be undermined by a failure of subsistence, a lack of bargaining power in a market, or the erosion of traditional community support mechanisms. Sen's work was influen tial and there followed a surge in micro level empirical work that put flesh on the bones of his framework of an alysis. This has shown that contingent reality is even more complex than he had imagined. Plausible though his case study analyses are, the evidence suggests that they work because they fit a particular type of famine in a particular socioeconomic context. In SSA, where migration to large feeding centers is a common feature of food emergencies, mortality tends to be less from starvation than hygiene related infections in the camps such as diarrhea or

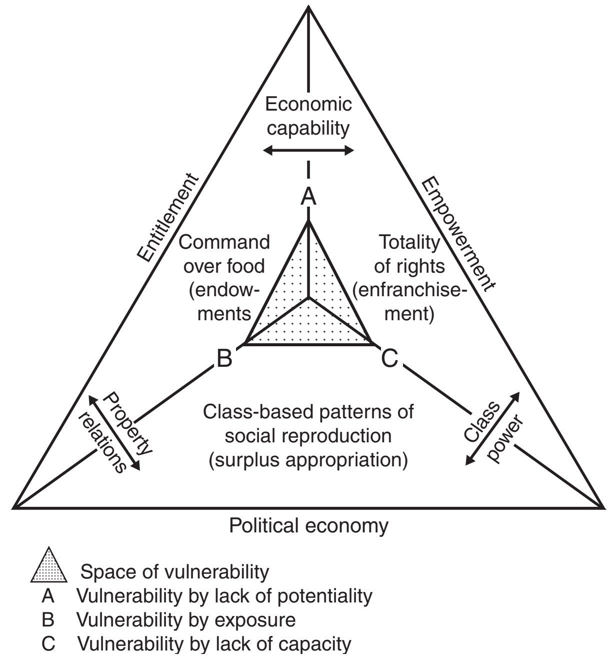

Figure 1 Reproduced from 'The space of vulnerability: the causal structure of hunger and famine' by Watts, M. J. and Bohle, H. G. in *Progress in Human Geography* 17, 43 67, Sage Publications 1993.

contagious diseases like cholera. Here the poor and the better off suffer together. Sen is most relevant to peasant societies with private property and some engagement with market based exchange, but less predictive for precapitalist societies dominated by common property resources or for famine under socialism. He is at his weakest when the bedrock of his theory, legal based entitlements, is flouted, as in times of civil conflict and war.

Sen's later work has argued that the presence of liberal democracy in a country is a powerful antifamine tool. The argument runs that effective political opposition and a critical media will hold governments and their agencies to account and that indifference to starvation, inefficiency in relief delivery, and the deliberate creation of famine conditions are, as a result, less likely. But does the evi dence fit? The great famine in Ireland happened under the control of UK parliamentary democracy. India is lauded as the world's largest democracy and a miracle of multi culturalism, yet her accountable politicians have not sys tematically addressed their problem of malnutrition, with the result that 27% of the world's hungry live in that one country. Also, 70% of the nonpregnant women in India suffer with iron deficiency anemia and 60% of preschool children have vitamin A deficiency. This is a wretched failure for the liberal developmental state. However, it is less the presence of a constitutional democracy that matters than the effective operation of institutions, cou pled with the politicization of famine in a manner that commits politicians to an 'antifamine political contract'.

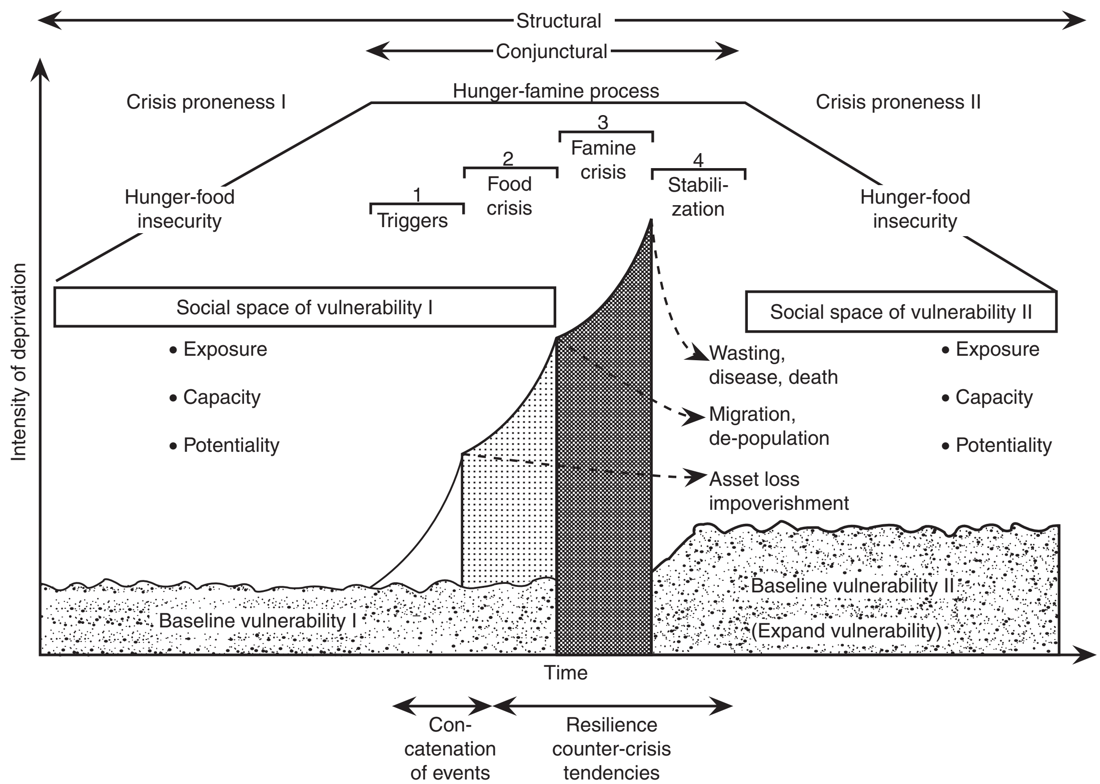

Figure 2 Reproduced from 'The space of vulnerability: the causal structure of hunger and famine' by Watts, M. J. and Bohle, H. G. in *Progress in Human Geography* 17, 43 67, Sage Publications 1993.

New Famines

Jenny Edkins has suggested that famine is 'produced' by modernity. It is not surprising that in the latest phase, globalization, we can see some new variants. The first is due to the erosion of capabilities and emergence of new vulnerabilities that has accompanied the AIDS epidemic. One estimate is that the economy of South Africa, for instance, is 22% smaller than it would have been without AIDS, and at least one recent famine was exacerbated by AIDS morbidity through a reduction in productive cap acity, a rise in dependency ratios, and an increase in the burden of care for the sick. Often family decision makers are affected by the disease but it is their dependants, including a generation of orphans, who are most sus ceptible to malnutrition.

Second, there are hungry cities, where poor slum dwellers may not have the vegetable growing and small livestock keeping possibilities of their rural counterparts, and they are therefore vulnerable to price spikes and labor market imperfections in times of economic stress. An example that tipped over into urban famine was Antananarivo, the capital of Madagascar, in 1985–86. This was a famine hidden from the outside world by the government then in power, but its impact can be re constructed from demographic data. Mortality increased especially amongst children and this coincided with a period of price instability in the staple commodity, rice.

Third, although there have been benefits of global ization, ranging from an expanded trade in foodstuffs and the creation of buffer stocks against times of famine, they have often been nullified by the presence of insurgencies and civil wars in many African countries. The typical modern African famine involves a military/political di mension that may have links to international geopolitics. An example is the 1998 famine in Bar el Ghazal, Sudan. Here the production and export of oil from the south of the country was the trigger for rebel groups to object to the exploitation of 'their' resources and for the govern ment to reinvest in military conquest.

Finally, in Iraq there have been an estimated 200 000 to 1 million deaths attributable to what Haris Gazdar has called 'postmodern famine'. This is defined as "excess mortality in spite of protected food entitlements but due to non food crises: notably the stresses of macro shocks on relatively sophisticated health and social welfare systems." Sanctions were imposed on Iraq by the UN in 1990, causing immediate problems in a country that had previously been dependent on imports for two thirds or more of its food. This was a famine, in the broad sense, in a market economy that continued to function. The key to understanding this type of famine is that a country that was closely integrated into the global economy suffered a severe shock when that access (export of oil in return for food and manufactures) was cut off. Although the prox imate cause was political, arguably the structure is similar to that of other economic crises, such as Indonesia (1997), Russia (1998), and Argentina (2001–02), where hunger related morbidity and mortality increased far beyond the normal. Maybe this is a glimpse of one type of future famine.

Famine Futures

The first of the UN Millennium Goals is to halve hunger from 800 to 400 million by 2015. Will this step to utopia be achieved through developmentalism? The evidence since World War II has not always been encouraging but famine in its acute sense does now seem to be confined to Africa. The only recent aberration (1995–99) was in North Korea where an authoritarian and secretive regime presided over perhaps 3 million starvation deaths at a time of flood induced chaos in their fragile rural econ omy. Lest complacency about world prospects sets in, we should note three points.

First, there is an industry of dire predictions about the outturn of global warming. Some of these involve a commentary about perturbations in the Asian monsoon or the spread of aridity in Africa and the Mediterranean, and the implication is of increased vulnerability in countries (such as India) that are densely populated. As the climate modeling becomes more sophisticated, a consensus is growing that the risk of famine will increase over the coming decades.

Second, who will feed China? This is a question posed by Lester Brown, a well known pessimist about world population growth. He has moved on from the cruder realms of neo Malthusianism to a new position that recognizes China's stunning economic success since the death of Mao and argues that it will need to import more and more food in order to meet rising living stan dards and to plug shortfalls in its own production. With India and other parts of the global South likely to share in this new prosperity, we are invited to wonder whether ultimately there will be enough food in international trade to prevent famine elsewhere.

Both of these first two points may be contested but, third, there is no doubt that at the beginning of the twenty first century there are more people dying of malnutrition than of famine. Many of the arguments raised in this article apply equally as well to the chronic food problems as they do the acute. Given the world community's very welcome new commitment (in prin ciple) to food as a human right, what is needed is an ethical and legal geography of the extent to which this right is being flouted, directly and indirectly, by various stakeholders. Let us hope that the present century will be one, not just of liberal hand wringing about hunger, but of accountability for the famine crimes that are being committed everyday against the rights of the vulnerable and the powerless. In short, we need a repoliticization of famine.

See also: Africa; Aid; Developmentalism; Human Rights; Plant Geographies; Vulnerability.

Further Reading

Braun, J., von Teklu, T. and Webb, P. (1998). *Famine in Africa: Causes, responses, and prevention*. Baltimore, MD: Johns Hopkins University Press.
Campbell, D. (2003). Salgado and the Sahel: Documentary photography and the imaging of famine. In Debrix, F. & Weber, C. (eds.) *Rituals of mediation: International politics and social meaning*, pp 69 96. Minneapolis: University of Minnesota Press.
Davis, M. (2001). *Late Victorian Holocausts: El nino famines and the making of the third world*. London: Verso.
Devereux, S. (1993). *Theories of famine*. Hemel Hempstead: Harvester Wheatsheaf.
Devereux, S. (2000). *Famine in the twentieth century, IDS Working Paper 105*. Brighton: Institute of Development Studies, University of Sussex.
Devereux, S. (ed.) (2007). *The new famines: Why famines exist in an era of globalization*. London: Routledge.
de Waal, A. (1997). *Famine crimes: Politics and the disaster relief industry in Africa*. Oxford: James Currey.
de Waal, A. (2005). *Famine that kills: Darfur, Sudan* (2nd ed.). Oxford: Oxford University Press.
de Waal, A. and Whiteside, A. (2003). New variant famine: AIDS and food crisis in southern Africa. *Lancet* 362, 1234 1237.
Drèze, J. and Sen, A. (eds.) (1990). *The political economy of hunger, Vol. II: Famine prevention*. Oxford: Clarendon Press.
Dyson, T. and Ó Gráda, C. (eds.) (2002). *Famine demography*. Oxford: Oxford University Press.
Edkins, J. (2000). *Whose hunger? Concepts of famine, practices of aid*. Minneapolis: University of Minnesota Press.
Howe, P. and Devereux, S. (2004). Famine intensity and magnitude scales: A proposal for an instrumental definition of famine. *Disasters* 28, 353 372.
Keen, D. (1994). *The benefits of famine: A political economy of famine and relief in Southwestern Sudan, 1983 1989*. Princeton: Princeton University Press.
Sen, A. (1981). *Poverty and famines: An essay on entitlement and deprivation*. Oxford: Oxford University Press.
Swift, J. (2006). Why are rural people vulnerable to famine? *IDS Bulletin* 37(4), 33 40.
Watts, M. J. (1983). *Silent violence: Food, famine and peasantry in Northern Nigeria*. Berkeley: University of California Press.
Watts, M. J. and Bohle, H. G. (1993). The space of vulnerability: The causal structure of hunger and famine. *Progress in Human Geography* 17, 43 67.

Relevant Websites

http://www.fivims.net
An Interagency Initiative to Promote Information and Mapping Systems on Food Insecurity and Vulnerability.
http://www.fews.net
USAID Famine Early Warning System.
http://www.wfp.org
World Food Program.

Fatherland/Homeland

R. Kaiser, University of Wisconsin–Madison, Madison, WI, USA

Glossary

Constitutive Outside The Others against which identities are constructed, and which is therefore incorporated within as the constitutive core all identification discourses and practices.

Instrumental Constructivism The approach to collective identity formation which views it as a top-down, center-out project through which elites reshape the public into new sociospatial communities, and that tends to treat nationalization as a once-and-for-all process that results in the construction of mass-based nations.

Memoryscape The complex or network of sites of memory that symbolically constitutes the homeland, and helps to ground or sediment the fiction of the nation as a natural and timeless entity rooted in the soil of the homeland.

National Symbolic Landscape A landscape image that stands as the visual representation of the entire national homeland.

National Territorialization The approach that views identity as a category of discourse and practice, rather than as an essentialized thing in the world; identity is understood as constituted in and through a set of discursive practices that people narrate and enact in their everyday lives.

Nonessential or Post-Structural Constructivism The approach that views identity as a category of discourse and practice, rather than as an essentialized thing in the world; identity is understood as constituted in and through a set of discursive practices that people narrate and enact in their everyday lives.

Primordialism The approach that views identity categories as natural, timeless essences; the dominant discourse adopted by nationalists that works to naturalize and reify nations as organic entities or beings in the world.

Territorial Nationalization The nationalization of space, or the process of re-imagining a place as a national homeland, which may include mapping, constructing sites of memory, re-narrating the past of a place, and the myriad banal practices through which places are nationalized as homelands in everyday discourse.

One recent demonstration of the power contained in the word 'homeland' is its invocation in the United States following 11 September 2001. In a country whose founding myths and legends emphasize immigration to a new land in search of freedom, homeland as such has not played an overly or at least overtly significant role in the construction of American national identity. Yet when an attack penetrated the borders of the US, almost imme diately, homeland talk and homeland politics rose to a dominant position in the security discourses that fol lowed. Not only through the creation of the Department of Homeland Security, but also in and through daily use in speeches and reports on national security, in academic journals and books, in school textbooks, and on the daily news and weekly television programming, representing the US as the homeland of American ness became second nature for politicians, journalists, academics, and the general public.

In turn, the use of 'homeland' in US security discourse following 9/11 has come to dominate the English lan guage literature devoted to homeland studies. In a recent survey of the 234 volumes held at the University of Wisconsin–Madison library with the word 'homeland' in the title, 186 were directly related to post 9/11 US homeland security topics, and only 48 treated homeland in its other uses. A search of journal article titles showed a similar semantic capture of homeland by US security discourse. This appropriation of the word 'homeland' has altered the meaning of the term itself, even while its use has transformed the US into a homeland for millions of Americans. Since homeland has frequently been invoked by those seeking to secure the boundaries of national identity against outside threats posed by 'others', we may say that the term 'homeland' was symbolically deployed in the United States to restore national security after 9/11 – by securing state territory as homeland, the home land image could provide a security blanket for Ameri cans and American ness.

Introduction

'Homeland', 'motherland', and 'fatherland' are terms used to symbolize the deep emotional connectedness that people are said to feel toward their places of origin, as well as toward more geographically expansive and so cially constructed birthspaces such as national home lands. The terms themselves are saturated with power, because they contain within them cultural claims to be longing in, and to, particular places and, implicitly at least, political claims to preeminent status for in group members in such spaces. Beyond this, narrating the homeland as motherland and/or fatherland also helps to establish and maintain gendered ideals and social norms of behavior for members of the in group, and contrasting sets of gendered images for out group 'others', and so works as an integral part of governmentality.

This article focuses on four aspects of the changing understanding and usage of homeland, motherland, and fatherland. First, is an examination of the reconceptuali zation of these terms in the contemporary social sciences, including their use in human and political geography. Next we consider homeland making discourses and practices in nationalism – the hegemonic source of their regulation and normalization for at least the past century. Third, the article explores the interrelationship between the three terms, and the conditions under which en gendering the homeland as motherland and/or fatherland occurs. The article concludes with an assessment of the impact that globalization and discourses of homelessness are having on representations of places as homelands.

Reconceptualizing Homeland

The first thing that needs to be said is that homeland, fatherland, and motherland have not been terms that human or political geographers have paid a great deal of attention to. They do not show up, for example, in any edition of the *Dictionary of Human Geography*, have not been the focus of chapters or articles in political or human geography textbooks, and rarely make an ap pearance in broader geographic studies of power and place. Where we do tend to find greater attention to homeland, fatherland, and motherland is in political geographic studies of nationalism and territoriality. This concentrated treatment within geographies of national ism is reflective of the critical role played by nationalists in the discourses and practices that constitute homeland making, as well as the significance of homeland talk and homeland politics to nationalism.

Academic research into the homeland group identity bond has shifted decisively during the past half century. Earlier primordialist treatments of ancestral homelands as a fundamental part of basic group identity, with their idealized sociospatial imaginary of a place for every people and a people for every place, have given way to more critical analyses of homeland claims. During the 1970s and 1980s, structural approaches rejected such primordialist claims, and sought to provide evidence for the recency of homeland sentiments felt by ethno national groups toward places more geographically ex tensive than localities. This more constructivist research perspective emphasized nation and homeland making as instrumental, political processes beginning in the late eighteenth and nineteenth centuries with the national ization and territorialization of elites, and extending to the masses by the late nineteenth and early twentieth centuries. The development of popular homeland senti ments toward the territorial state was especially central to this instrumentalist agenda, and political and cultural elites seeking to represent the territory of the state as the ancestral homeland, motherland, and/or fatherland of the nation used textbooks and schooling, hiking and touring clubs, map making, and other popular media such as poster art and music to socialize the public toward spatial identification with the territorial state as home. At the same time, these instrumentalities were also deployed to nationalize the population. In this way, nation and homeland making were mutually constituted through the processes of territorial nationalization and national territorialization.

Though closer to a critical engagement with nations and homelands not as primordial things but as historically contingent processes, this first wave of constructivist re search did suffer from two serious limitations. First, an instrumentalist framing tended to imagine a top down, center out process of nationalization and territorialization that treated the population being territorialized and na tionalized as an inert mass waiting to be reshaped, and so ignored the interactive and iterative nature of these sociospatial processes. Second, even though this approach views both nations and homelands as modern social constructs, much of the research conducted in this vein does tend to depict social construction as a once and for all process producing nations and homelands as real things that were made in the world at some point in the not too distant past. The problem here is that analysts continued to treat nations and homelands – once made – in relatively static, essentialized ways, as reified actors whose nature could be studied, understood, and predicted as such. Popular and even many academic understandings of na tions and homelands (e.g., the German, Irish, Japanese nation thinks this and feels that; Germany, Ireland, Japan believes this and acts like that) continue to suffer from this reification of nations and homelands as conjoined, unitary actors or beings in the world that can be categorized and studied in this way.

More recently, the social construction of places and identities as mutually constitutive processes has focused on the discursive practices of everyday life through which nations and homelands become. This more nonessential or post structural treatment of homelands as processes of becoming and as narratives and enactments through which people emplot themselves in the places they call 'home' resolves the more glaring problems associated with earlier critical analytical approaches, and shifts the focus of re search attention in a number of ways. First, post structural or nonessential homeland studies investigate the dis courses through which places are represented as home lands, as well as the daily citational practices that produce homelands as material effects. This shift also focuses attention on the actors and the sociospatial stances that they take as they narrate and enact or 'flag' the homeland in their daily discursive practices. These first two redir ections have helped to de naturalize and de essentialize homelands, and to emphasize homelands as ongoing say ings and doings. Beyond this, post structural studies move the analytical gaze from the center (i.e., from homeland as a unifying, even unitary analytical category) and toward the boundaries that are constructed through the discursive practices of homeland making to differentiate and ex clude. This shift highlights both the centrality of the constitutive outside in the social construction of home lands, and also the fact that homelands do not grow or ganically from within or from the center out, but always materialize as a set of interiorizing and exteriorizing boundary effects.

Overall, in the process of this transformation in homeland studies, the principal research question has shifted from 'what is the homeland/fatherland/mother land?', to 'when is the homeland?', and finally to 'how are homeland images naturalized and sedimented through daily discursive practices?'. Under what circumstances do homeland discourses and practices work? That is, when do homeland talk and homeland politics resonate among their intended audiences and socially mobilize these publics? Interestingly, this new set of post structural re search questions has brought primordialism back in – al beit in a new way – since it is these more primordialist narratives and enactments of homeland that are frequently deployed in the naturalization and essentialization of homeland representations. However, rather than treating homelands as ancestral places of primordial groups, non essential approaches explore how the contestable fiction of blood and soil connectivity used to establish both ethno national identity and ethno territorial rootedness in the soil of the homeland is turned into the unassailable, un questioned truth by and for those becoming nationalized and territorialized. In the section below, we provide a more detailed discussion of the discourses and practices used to naturalize and sediment the homeland as the ground on which the nation is constituted.

The Discourses and Practices of Homeland-Making

Places successfully represented as ancestral homelands are critical to the naturalization of nation ness, since homelands are socially constructed to provide visible, tangible evidence of nations as ancient, even timeless, fixtures on the landscape. The fiction of the nation's primordial and organic essence is given concrete ex pression and materializes as fact in and through the homeland. Mapping and other cartographic presentations of national space are – along with censuses and museums – the chief technologies through which nations are so cially constructed and naturalized in place.

Physical representations of the territorial state or other politically bounded spaces as homelands, nationalized toponyms, and bordering processes – both delimiting on

Figure 1 Cover of *Eesti Atlas*. The cover of this Estonian atlas uses two of the homeland representations discussed: an outlined or logoized map of Estonia, and the symbolic national landscape of Estonian-ness embedded within it. This combination promotes the nationalization of this sociospatial imaginary which helps both to localize the national and nationalize the local in the nation and homeland (Avita, Tallinn, 1996). Reproduced with permission of Avita Publishers.

maps and demarcating on the ground – all help to ground truth the national sociospatial imaginary. Logoized representations of the homeland's boundary outline and other uses of the territorial state/homeland shape on coins and currency, stamps, flags, posters, souvenirs, and so on reinforce the geovisualization of the nation in and through homeland imagery (**Figure 1**). Symbolic national landscapes, whether of manicured agrarian scenes or stark and untamed nature, conjure up singular visual representations of the national homeland, ones that frequently also make their nightly appearances in folk tales and other bedtime stories. However, there is nothing natural about the choice of symbolic national landscapes, which are frequently selected as a counternarrative to symbolic landscape images of the national 'others'. For example:

> The West of Ireland came to symbolize the whole of Ireland to some Irish nationalists in the early twentieth century. Nationalist Irish writers initially developed a great interest in the region towards the end of the nineteenth century, partly because it was seen as the area least affected by the processes of Anglicization, but also because its bare and rugged landscape seemed to contrast so breathtakingly with the more pastoral landscapes through which contemporary England was imagined (Rose 1995: 91). (**Figures 2** and **3**).

In other words, not only are homelands always already constructed in relation to the 'other', even something as natural in appearance as a symbolic national landscape that represents the homeland as a rugged, unpopulated wilderness is imagined through interaction with the constitutive outside.

Specific commemorative sites, monuments, museums, and other *lieux de memoire* (sites of memory) all work to materialize nationalized narratives of the past in a network of memoryscapes that symbolically constitute the homeland. Publics experience the naturalization and essentialization of the nation and homeland through these sites, and these visitations take place in a wide variety of contexts. Schoolchildren regularly take fieldtrips to these places of collective memory, and also visit them virtually through textbook images and narratives, educational and popular films, and increasingly through Internet websites devoted to the production and dissemination of knowledge about these important nodes in the timespace fabric of the nation. This memory work is a crucial part of the socialization and spatialization of each generation into the nation. Special commemorative events are also staged at many of these sites of memory – typically on national holidays – during which nation and homeland are frequently invoked. Beyond these official ceremonies, informal cultural events such as folk festivals, concerts, plays, and more carnivalesque activities often also use sites of memory as their backdrop, and so help to naturalize and sediment nation and homeland into the public's sociospatial imaginary through these more banal discursive practices.

However, rather than seeing these sites as stable national anchors through which generation after generation is socialized and spatialized into nation and homeland, it is more accurate to see them as dynamically changing public narratives and enactments whose storyline shifts over time to accommodate new understandings and interpretations of the nation homeland imaginary. Additionally, of course, there are sites of memory that contest

Figure 2 *Mr. and Mrs. Andrews*, painting by Thomas Gainsborough *c.* 1750. This picture of a domesticated rural scene with estate owners posing before worked fields and gently rolling hills came to be identified as the symbolic national landscape of English-ness. The selection by Irish and Scottish nationalists of stark, undomesticated wilderness for their symbolic national landscapes was, at least in part, done to differentiate Irish-ness and Scottishness from what it meant to be English (see Figure 3). Reproduced with permission of the National Gallery, London.

Figure 3 Photograph of West Ireland. Irish nationalists chose West Ireland for the symbolic national landscape of Irish-ness both because it was the region that had experienced the least Anglicization, and also because it provided such a strong visual counterpoint to the symbolic national landscape of English-ness (see Figure 2). Reproduced with permission of the photographer, Eric Zuelow.

as well as support dominant national discourses, and memoryscapes are perhaps more appropriately seen as a multilayered disjointed circuitry with many breaks in the line, patches, and switches, rather than as a cohesive network of smooth memory flows. Monuments con structed during one discursive era are at times so vio lently at odds with superceding dominant identity discourses and practices that they must be removed, as in much of the former Eastern Europe following the breakup of the socialist bloc and the emergence of ter ritorial nationalism as the dominant ideology during the 1990s. More frequently, these earlier monuments are left standing; though their positions within the memoryscape are undoubtedly transformed. This process of disjointed accretion also includes alternative sites of memory cre ated to represent subaltern understandings of the past in place. Beyond this, each visitor brings her/his own set of discursive practices through which s/he employs herself/ himself into the public narrative of nation and homeland constructed in and through these nodes. The flexibility of these shifting storylines and variable individual readings and emplotment typically occurs within the loosely bounded parameters of established nationalized norms, and is clearly a strength in that it allows the discourse and practice of nation and homeland to change with the times – becoming more or less inclusive of 'other ness' in the process – while seeming to remain permanent and unchanging touchstones of the nation's timelessness.

Maps, symbolic national landscapes, and sites of memory are three ways that places become homelands that in turn help to convert the fiction of the nation's solidity and primordial essence into fact – at least for the nationalizing public. Allegorical figures are another set of powerful images that not only represent the homeland as the nation (e.g., Mother Russia), but in doing so, also gender the homeland as fatherland or motherland. Male statues most often depict national heroes – military men, political leaders, and cultural figures – whose biographies serve as allegories of the national past. Less frequently, male images are used as allegorical figures of the nation itself. Allegorical representations of the homeland most often take a feminine form, and may be presented as map images of the territory itself as a woman (**Figure 4**), or alternatively, as a female statue as the motherland (**Fig ure 5**). Allegorical figures are also reproduced in paintings, on posters and postcards, in books, and on the Internet.

In an elision of gendered identity and the gendering of place, the nationalist vision of women primarily as mothers reproducing the nation extends to motherland images as the birthplace of the nation. The motherland, therefore, is more often associated both with more rural blood and soil imagery, and also with more localized regions or places of origin, the more homely side of homeland. Those who invoke the motherland are fre quently calling on the nationalized sons – and less fre quently daughters – of the soil to love, honor, and defend the land as they would their own mothers. Not sur prisingly, then, motherland images are often used to mobilize the public during wartime, as in this World War II poster from the USSR (**Figure 6**).

Gendering space as motherland and/or fatherland is central to any nationalization process. While nationalists

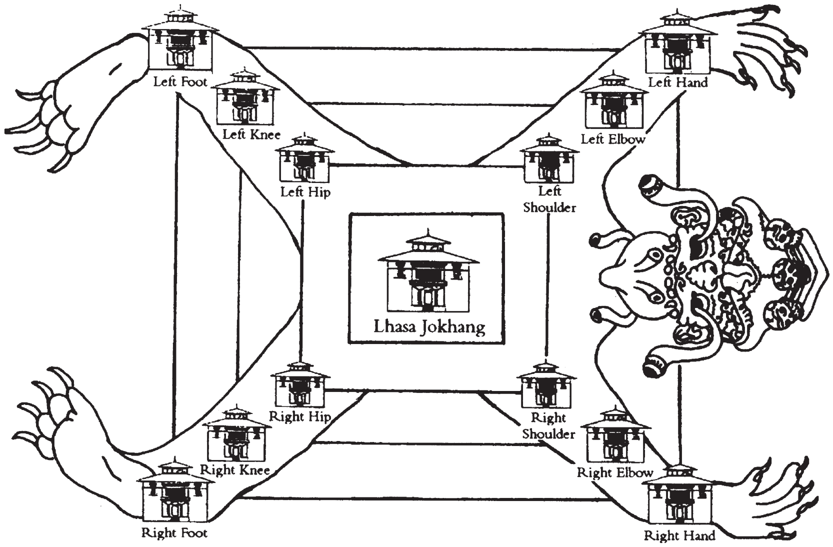

Figure 4 Tibet as the embodiment of the demon goddess Srin mo. This image represents one form of the gendered representation of homeland in this case, the homeland as a female goddess. As legend has it, Buddhist temples anchored by Lhasa Jokhang were located in geographically strategic positions to pin her down and subdue her. From French, R. (1995). *The GoldenYoke: The legal cosmology of Buddhist history*. Ithaca: Cornell University Press. Reproduced with permission of Cornell University Press.

Figure 5 Motherland Statue, Volgograd, Russia. The motherland statue was built to commemorate the battle of Stalingrad in World War II, and represents a grand allegorical image of the homeland as the mother of the nation to be defended. This allegory to the motherland speaks of the power of such sociospatial images to mobilize national publics, especially given the context in which it was produced a Soviet Union that, in principle, had rejected such place-identity politics in favor of socialist internationalism. Reproduced with permission of the photographer, Martha Waites.

tend to use motherland images to promote a more in ward looking, security minded public, representations of the fatherland promote a re territorialization of identity toward the paternalistic state as a national homeland that must be obeyed. Also, because motherland narratives and enactments tend to draw on a more localized sense of homeland, they help to ground nationalized discourses of place and identity in localized legends and landscapes, and in the process work to localize understandings of the nation. Fatherland narratives evoke more masculine, martial, militarized understandings of the nation–home land relationship, and tend to connect the homeland image to politicized national space – as the territory of a state already created or a state to be – and so tend to nationalize the local. In these ways, motherland and fatherland imagery work in tandem to stitch together the sociohistorical fabric of nation and homeland by drawing in and nationalizing localized discursive practices that

Figure 6 'Motherland Calls!' by Irakliy Toidze, 1941. This image was part of the exhibit: "Soviet Propaganda Posters: The Power of Art, Art of Power", sponsored by the Center for Russia, East Europe and Central Asia (CREECA), University of Wisconsin Madison. Reproduced with permission of CREECA. This Soviet poster from the outset of World War II speaks even more poignantly of the power of the allegorical motherland image to mobilize the public in times of war. In this poster, the motherland holds a "war oath", which seeks to employ this symbolic power in the name of the USSR as motherland. In one part of the oath, the soldier swears "to protect her (the motherland) courageously (*muzhestvenno* in a manly way), skillfully, with dignity and honor, not sparing my own blood and life for the achievement of complete victory over the enemy."

constitute the collective experience of becoming. Taken together, the gendering of homeland both as mother and father has helped to naturalize and sediment the linkages between birthplace and territorial state in the popular perceptions of those being socialized and territorialized into the nation and state, and so has served as a powerful instrumentality in the social construction and stabilization of states as the natural containers of national identity.

If motherland representations of home emphasize the sense of belonging aspect of national territoriality and focus more attention on an inward looking we belong here and nowhere else sentiment, fatherland images tend to stress a sense of exclusiveness toward the state as the national homeland, a more outward looking no one else belongs here sentiment, and so place the 'other' to be exteriorized at the center of attention. While both gendered aspects are always present within homeland narratives and enactments, the balance between motherland and fatherland discourses and practices continually shifts with changing circumstances. In the United States, for example, the more prominent we here motherland discourse that followed the end of the Cold War was rapidly transformed into a much more 'other' dominated and exclusionary fatherland discourse following 9/11.

Homeland's End?

Today, in a certain sense, we live in an increasing condition of homelessness, and many analysts of cosmopolitanism and transnationalism have written obituaries not only for the nation state, but also for home and homeland. According to this literature, a sense of homelessness is growing, not only among people living away from their places of origin, but also as a result of alienation from those places imagined as homelands among people who have never left, as these places undergo processes of transnationalization. Identity itself is frequently depicted as deterritorialized, as something detached not only from homeland and state, but from place more generally. Do the discourses and practices that have come to be labeled globalization signal the end of homelands?

In thinking through the implications of globalization for homelands, it is important to keep in mind that homelands are socially constructed through daily discursive practices rather than natural pregiven things in the world, and that homeland narratives and enactments are constantly changing and becoming. Similarly, it is important to see globalization as a set of discursive practices that are affecting people's identification in a multiplicity of ways that are highly variable spatially. When seen in this light, a number of problems with the globalization as deterritorialized homelessness argument may be identified. First, this assertion seems to rest on a fairly simplistic equating of leaving home and homeland with the deterritorialization of identity. In many ways, this literature reinforces the naturalized, essentialist view of home and homeland – as backward looking, traditional, static, and reified places that are being left behind by the hypermobility of late modernity. Yet, as we have seen above, dynamically unfolding representations of homeland symbolically and discursively connect people not only to places, but also to one another, and serve as the memoryscapes through which knowledge of collective identities is produced. Once away, collective memories and nostalgic longing continually produce homeland thoughts, images and talk, as well as homeland politics. Homelands are not only symbols and memories, they materialize through the everyday practices of those who are away just as they do for those who remain behind – both in the social net works that stretch homelands globally, as well as in the social construction of homelands in new places. Home lands – albeit in decidedly reterritorialized forms – are popping up all over the place and seem quite resilient in a globalizing world.

More broadly, all identification is also a process of differentiation that is centrally about constructing, con testing, and renegotiating borders with the constitutive outside. Home and homeland – as two of the most powerful interiorizing discourses and practices through which identity comes into being – are unlikely to dis appear in a more mobile world where the processes of boundary construction and contestation between identi fication and differentiation are growing in intensity.

Finally, there is a value judgment apparent in much of the globalization as homelessness literature that needs to be questioned. The term 'homeland' – treated in much of this literature as a holdover from the pre modern past and as the place in whose name exclusionary, even genocidal, politics are practiced – carries a strong nega tive valence. Home has similarly been treated by at least some feminist and postcolonial geographers in negative terms, as a masculine defined space of containment and control. On the other hand, diasporic identity, hybridity, and homelessness carry an equally strong positive valence – as deterritorialized life on the margins – in a kind of libratory thirdspace. While home and homeland can certainly be places of confinement, and even of ex termination, they need not be, and for many they have been places of refuge, security, and freedom, places of liberation. By the same token, living a homeless life in diaspora is, for many, not a choice freely made but is rather an imposed condition. The value of home and homeland as well as hybridity and homelessness are not good or bad in their own right, but rather are dynamic, situated, and context specific, and need to be studied as such.

Understanding homelands as socially constructed through daily discourses and practices opens new possi bilities to explore both the power relations that materi alize and stabilize the homeland and its gendered forms, and also the ways that leaving home both de and reterritorializes identity as it reconfigures home and homeland as imagined places in the world. In not ac cepting homelands as things in the world but as ongoing discursive practices, we are freed to reenvision what kinds of homes and homelands we desire, and then to set about narrating and enacting them.

See also: Cultural Geography; Cultural Politics; Diaspora; Ethnic Conflict; Ethnicity; Ethnicity and Resistance, Historical Geographies of; Habitus; Heritage; Heritage and Culture; Heritage and Identity; Home; Homelessness; Human Geography; Hybridity; Identity Politics; Indigeneity; Indigenous Mapping; Mapping, Race and Ethnicity; Maps and the State; Memorials and Monuments; Memory; Nation; National Spatialities; Nationalism; Nationalism, Historical Geography of; Other/ Otherness; Place; Place, Politics of; Political Boundaries; Popular Culture; Representation, Politics of; Sense of Place; Sovereignty; State; Territory and Territoriality.

Further Reading

Billig, M. (1995). *Banal nationalism*. London: Sage.
Brah, A. (1996). *Cartographies of diaspora: Contesting identities*. London: Routledge.
Brubaker, R. (1996). *Nationalism reframed*. Cambridge: Cambridge University Press.
Connor, W. (1986). The impact of homelands upon diasporas. In Sheffer, G. (ed.) *Modern diasporas in international politics*, pp 16 45. London: Croom Helm.
Diener, A. (2004). *Homeland conceptions and ethnic integration among Kazakhstan's Germans and Koreans*. Lewiston, NY: Edwin Mellen Press.
Edensor, T. (2002). *National identity, popular culture and everyday life*. Oxford: Berg.
French, R. (1995). *The GoldenYoke: The legal cosmology of Buddhist history*. Ithaca: Cornell University Press.
Gupta, A. and Ferguson, J. (eds.) (1997). *Culture, power, place*. London: Duke University Press.
Hage, G. (1996). The spatial imaginary of national practices: Dwelling domesticating/being exterminating. *Environment and Planning D: Society and Space* 14, 463 485.
Hooson, D. (ed.) (1994). *Geography and national identity*. Oxford: Blackwell.
Kaiser, R. (2001). Geography. In *The encyclopedia of nationalism*, vol. 1, pp 315 333. San Diego: Academic Press.
Kaiser, R. J. (2002). Homeland making and the territorialization of national identity. In Conversi, D. (ed.) *Ethnonationalism in the contemporary world*, pp 230 247. London: Routledge.
Morley, D. and Robins, K. (1996). No place like heimat: Images of homeland in European culture. In Eley, G. & Suny, R. (eds.) *Becoming national*, pp 456 480. Oxford: Oxford University Press.
Paasi, A. (1996). *Territories, boundaries, and consciousness*. New York: Wiley.
Paasi, A. (1999). Boundaries as social practice: Territoriality in the world of flows. In Newman, D. (ed.) *Boundaries, territories and postmodernity*, pp 69 88. London: Frank Cass.
Pratt, G. (1999). Geographies of identity and difference: Marking boundaries. In Massey, D., Allen, J. & Sarre, P. (eds.) *Human geography today*, pp 151 167. Oxford: Polity Press.
Rose, G. (1995). Place and identity: A sense of place. In Massey, D. & Jess, P. (eds.) *A place in the world? Places, cultures and globalization*, pp 87 132. Oxford: The Open University.
Smith, G., Law, V., Wilson, A., Bohr, A. and Allworth, E. (1998). *Nation building in the post soviet borderlands*. Cambridge: Cambridge University Press.
Yiftachel, O. (2001). The homeland and nationalism. In *The geography of nationalism*, vol. 1, pp 359 383. San Diego: Academic Press.

Feminism and Work

J. Pollard, Newcastle University, Newcastle, UK

Glossary

Discourse A set of practices through which the world is rendered meaningful and intelligible, including narratives, concepts, ideologies, and signifying practices.

Essentialism The belief in the existence of fixed, invariable properties ('essences') that define a person or entity.

Globalization The increasingly international integration, connectivity, and hybridity of a range of economic, cultural, political, and environmental activities.

Human Capital Theory Argues that individuals invest in their own human capital by improving their skills and qualifications for which they are rewarded accordingly.

Liberal Feminism Views women and men as equal, as deserving of equal rights in economic, social, and political life, and the use of the political and legal system to secure such rights.

Queer Theory Challenges essentialist notions of homosexuality and heterosexuality and understands sexuality as constructed and shifting in different historical and cultural contexts.

Radical Feminism Argues that inequalities between men and women are rooted in patriarchy, a system that constructs rights, structures, and practices in ways that privilege men.

Socialist Feminism Argues that there is a link between the class structure in capitalism and the oppression of women.

Definitions

Feminism is a broad, complex, contested term that comprises both an intellectual and a political project that seeks to identify, understand, and dismantle inequalities between men and women. As an intellectual movement, different strands of liberal, Marxist/socialist, radical, and post structuralist feminist thought have demonstrated how sexual differentiation and the social construction of gender differences are intricately linked with questions of power. As a political movement, feminism has waxed and waned and evolved through different eras of activity and achievement. Early, so called 'first wave' feminists in the US and UK, inspired in part by Mary Wollstonecraft's Vindication of the Rights of Woman (1792) and its appeal to the power of reason and equality, fought for women's political rights, notably the vote. A second wave of feminist activism in the 1960s fought a broader array of economic and social inequalities, including discrimin ation in the workplace and earnings inequality. This era of activism also asserted that 'the personal is political', that women's rights extended to control of their bodies, their sexuality, and reproductive rights. Feminist theor ists, writers, and activists have long struggled with ten sions between competing claims around universalism (that women have common experiences of inequality and discrimination) and particularism (that women are not homogeneous, that they occupy different social, eco nomic, cultural, political, and geographical locations). Arising from black feminists' and others' critiques of the class and racial bases of second wave feminism, 'third wave' feminists have explored the diversity of women's lives, experiences, and politics. Key to this has been an interest in the politics of intersectionality, the relations between gender and its co constitution with other axes of difference, like race, ethnicity, and sexuality. Drawing on queer theory, postcolonial theories, and post structuralist scholarship, feminist theorists have explored the signifi cance of representation, subjectivity, language, and iden tity. Many feminists now critique discourses concerning the singularity of knowledge and truth and note the partiality of projects that treat 'women' as an already constituted, homogeneous group or that treat 'gender' as a stable, social category.

The philosophical, political, and theoretical diversity that now characterizes feminist thought makes it difficult to generalize its key aspirations and achievements and it is now more accurate to talk about 'feminisms' in the plural. Nevertheless, through their strong ethical, polit ical, and utopian impulses, feminists have sought to highlight and resist the systematic production of social difference around gender and challenge the persistence of inequalities between men and women. Work – what it is, who does it, and where – has been a key arena of struggle. There is now a vast literature on feminism and work and some excellent reviews are available. In what follows, some key feminist and geographical contri butions are highlighted.

Origins of Interest in Human Geography and the Social Sciences

Just as there is no single definition of feminism, there is no single origin of social scientific interest in women's de mands for equality that fueled feminist scholarship. Part of the impetus came from feminist writers such as Charlotte Perkins Gilman, Emma Goldman, and Simone de Beau voir who identified work and class as important sites in the struggle for women's liberation. Yet, this emphasis on work also reflected the reality that women all over the world spend the major portion of their lives working, be it paid or unpaid. The twentieth century involved a transforma tion in the economic landscapes of many economies and the gradual feminization of waged employment, especially in service industries and occupations. For many women in North America and Europe, World War II marked a turning point as they were drawn into paid work, often for the first time, on a massive scale. The demand for labor was such that traditional gender divisions of labor were swept aside, at least temporarily, as women filled jobs traditionally occupied by men in engineering, steel, autos, and other manufacturing industries.

Since the 1950s in particular, a raft of social and cultural shifts and, in many places, economic necessity have driven rising female participation in labor markets in the global south and north. While it is difficult to find consistent data, most especially for many of the world's poorest countries, Joni Seager's Atlas of Women in the World presents a wealth of data on women's work. UN figures reveal that in large parts of Africa and Central Europe and some parts of northern Europe, over 60% of all women were working for pay in 1999. While the overall trend has been toward rising female participation in waged labor, there is a good deal of geographic vari ation, with lower rates of female participation in Catholic and Islamic countries. In many parts of the world, women are concentrated in agriculture or informal, largely un regulated sectors like domestic service or market trading (see **Table 1**).

Key Contributions

Contesting the Definition of Work

One of feminist theorists' major contributions to the study of work has been to critique how economics, sociology, geography, and other social science disciplines have defined 'work'. Many social scientific analyses have equated 'work' with 'paid employment', an assumption visible in long histories of different governments' stat istics keeping practices that define and categorize work in particular occupational and industrial classifications. Feminist scholars have argued that defining 'work' to include only formal waged labor represents a masculine norm and ignores a raft of unpaid, paid in kind, and informal work often performed by women, notably domestic labor and childcare.

Early research on work largely ignored domestic labor because homes were not viewed as sites of work or production but rather as places where women engaged in what were viewed as their 'natural' and biologically given roles of caring and nurturing. Yet this analytical and spatial separation of home and work is geographically and historically specific; it is a description that applies largely to a suite of industrialized capitalist economies after the industrial revolution. An important feminist contribution was thus to challenge the analytical separ ation of work (the public sphere) and home (the private sphere) in industrialized capitalist economies that sug gested that women's reproductive roles either limited their work to the private sphere or else meant a 'second shift' in the public sphere of paid work. Feminists have sought to understand housework as a form of work, to analyze it like other forms of work, and to have it taken seriously in terms of its role in producing essential goods and services.

Anne Oakley's *The sociology of housework* was a major contribution that argued that unpaid work in the home could be understood in the same terms as paid work. For liberal feminists like Betty Friedan, the unpaid status of domestic labor meant that is was undervalued. Other Marxian and socialist feminists joined in what became known as the 'domestic labor debates' of the 1970s and 1980s and argued that homemakers were fundamental for the reproduction of capitalist social relations and, like their partners in paid work, were being exploited. The oretically, this emphasis on the household was important because it did more than simply add the experiences of women into traditional analyses; instead, it challenged prevailing approaches that explained sexual divisions of labor in terms of individual choices and argued that issues of dependence, tradition, and most especially power over resources were crucial. Key schisms in these literatures concerned the analytical status of class and gender and whether capitalism and patriarchy were conceived as a single or 'dual system'. Politically, this scholarship highlighted the precarious situation of many married women who were full time domestic laborers and thus economically dependent on waged partners. Moreover, women's domestic labor, unpaid and under valued, was not subject to various forms of statutory protections, like social security, redundancy pay, or un employment benefit that accompanied waged work. At various times then, this scholarship translated into de mands for wages for housework and attempts to measure the economic value of domestic labor, although this in turn created many debates about how to draw the line between work, leisure, and play in, for example, activities such as child rearing.

Explaining Occupational Segregation

A second contribution of feminist scholarship has been to confront and interrogate one of the most pervasive and enduring features of women's participation in the

Table 1 Women's employment by economic activity[a] 1995 2003, selected countries

Country	Agriculture[b] 1995 2002 (%)		Industry[c] 1995 2002 (%)		Services[d] 1995 2002 (%)		Contributing family members[e] 1995 2003 (%)	
	Women	*Men*	*Women*	*Men*	*Women*	*Men*	*Women*	*Men*
Argentina		1	12	30	87	69	59	41
Australia	3	6	10	30	87	64	58	42
Bangladesh	77	53	9	11	12	30	81	19
Bolivia	3	6	14	39	82	55	63	37
Botswana	17	22	14	26	67	51	45	55
Canada	2	4	11	33	87	64	65	35
France	1	2	13	34	86	64		
Japan	5	5	21	37	73	57	81	19
Kenya	16	20	10	23	75	57		
Mexico	6	24	22	28	72	48	49	51
Norway	2	6	9	33	88	58	63	38
Pakistan	73	44	9	20	18	36	33	67
Russian Federation	8	15	22	36	69	49	42	58
Singapore			18	31	81	69	76	24
Slovenia	10	10	29	46	61	43	62	38
South Africa	9	12	14	33	75	50		
Sri Lanka	49	38	22	23	27	37	56	44
Thailand	48	50	17	20	35	30	66	34
UK	1	2	11	36	88	62	68	32
Yemen	88	43	3	14	9	43	26	74

[a] Defined as all employed and unemployed persons (including those seeking work for the first time), employers operating unincorporated enterprises, people working on their own account, employees, unpaid contributing family workers, members of producers cooperatives, members of the armed forces, those producing primary products (e.g., foodstuffs for their own consumption), and certain other nonmonetary activities. In principle, any such work for as little as 1 h a week is taken to define a person as economically active. Note that elements of the standard concept of economically active population may differ substantially from country to country, such as the choice of time reference period and the determination of minimum hours of work and unpaid family work, including production for own consumption. These differences may result in the undercounting of women who are economically active. Also, in most countries, the statistics of the economically active population relate only to employed and unemployed persons above a specified age, while in some there is no such age provision in the definition of economic activity. Source: United Nations Women's Watch 2000, Table 5a Economic Activity, http://unstats.un.org/unsd/demographic/products/indwm/tab5c.htm

[b] Agriculture refers to activities in agriculture, hunting, forestry, and fishing.

[c] Industry refers to mining and quarrying, manufacturing, construction, and public utilities (gas, water, and electricity).

[d] Services refer to wholesale and retail trade; restaurants and hotels; transport, storage, and communications; finance, insurance, real estate, and business services; and community, social, and personal services.

[e] Contributing family member is defined as a 'person who works without pay in an economic enterprise operated by a related person living in the same household' p. 354 UNDP 2005, Technical notes.

Source: Compiled from UN Human Development Report (2005) International cooperation at a crossroads: Aid, trade and security in an unequal world, Table 28, United Nations, New York (http://hdr.undp.org/reports/global/2005/pdf/HDR05 HDI.pdf).

waged labor force, namely their segregation both vertically (in the lowest levels of seniority) and horizontally (in particular, industries, typically clerical and other personal service work and in some forms of manufacturing like textiles, clothing, and footwear). Men and women typically perform different jobs, in different industries, with women dominated occupations often classified as 'unskilled' or 'semiskilled', while men predominate in occupations labelled 'skilled'. Women are also disproportionately represented in home work, part time work and other forms of temporary, flexible, and casualized work. For human capital theorists like Gary Becker, this state of affairs is predictable given that men are, in general, better educated and command higher rates of return in the labor force than women. Taking a household as the unit of analysis, human capital theorists would argue that it is rational for men to undertake waged work and for women to perform domestic labor and if they do undertake waged work, to work part time and relatively close to home. While accounts such as these describe some common patterns, they take as given, rather than explain, prevailing gender divisions of labor.

Many feminists have rejected explanations based on individual choices made in the context of family responsibilities. Many early analyses turned to Marxian theory for inspiration and sought to understand how women's labor helped to reproduce capitalist social relations. Socialist feminists sought to reconstruct Marxist theory in order to introduce analyses of male domination as a key axis of power relations, alongside class. Early Marxist and socialist feminist analyses of women's occupational segregation noted the similarities in the kinds of work being performed by women in the home and workplace and argued that occupational segregation benefited employers in that, women's unpaid domestic labor minimized employers' costs of reproducing labor power and, in turn, gave employers a strong incentive to restrict women to low paid jobs (to maintain their dependence on other wage earners for whom they perform domestic labor). As such, women were conceived as a 'reserve army' of labor, a disposable, underemployed group whose numbers wax and wane with the requirements of employers.

Later work of labor market segmentation theorists examined the correspondence between the sexual division of labor and stable, well paid ('primary' sector) jobs and those more precarious, low paid jobs with high turnover ('secondary' sector jobs). Marxian inspired variants of segmentation theory argued that employers had a vested class interest in dividing and ruling workers, just as the working class had an interest in unity and, by implication, resisting segmentation. Yet, the difficulty faced by Marxian and other accounts of segregation is its sheer persistence and its resilience to changing labor market demographics, competition, new forms of work, and constantly shifting occupational and industrial structures. Although employers have an interest in 'divide and rule', they also have strong material incentives to substitute cheap(er) female labor for its more expensive (male) equivalent. Heidi Hartmann made an important contribution to these debates by arguing that men's patriarchal interests in maintaining the subordination of women gave them a material incentive for supporting segmentation in the workplace. In essence, Hartmann argued that capitalism and patriarchy operated as 'dual systems' which articulate in different ways at different times.

In addition to explaining a rationale for persistent occupational segregation, feminist scholars have sought to identify the mechanisms that perpetuate occupational segregation. This body of work has tended to look at different occupations, workplaces, and organizations to understand how work is labeled 'male' or 'female' in different contexts. Feminists' insistence on establishing a distinction between sex as a biological category and gender as a social category has been crucial in dismantling biological determinism as an explanation for occupational segregation. Generalizations about women's strength, height, or dexterity and their suitability for some, but not other occupations, have foundered on their conspicuous lack of historical, geographical, and occupational consistency. Instead, feminists have argued that jobs are created as 'appropriate' for men or women, that jobs are socially constructed and maintained through practices that embody the socially accepted (and very variable) attributes of femininity and masculinity.

There have been many excellent studies in this area. One particularly important contribution has been the argument that 'skill', a key component of job classification, is not an objective characteristic of work, but one that is socially constructed to value highly certain kinds of work and to undervalue other kinds of work. Thus, the capacity of employers and institutions like labor unions to define job and skill categories and to shape recruitment and retention policies plays a crucial role in reproducing the segregation of men and women in the workplace. Cynthia Cockburn's *Brothers: Male dominance and technological change* (1983), presents a fascinating insight into how skilled male compositors in the British printing industry fought to maintain their skilled status in the face of technological changes that replaced their hot metal compositing with computerized typesetting. The men were retrained with new keyboard skills (often by young women) and represented themselves as skilled, experienced printers in contrast with their new female colleagues who were merely 'typists'.

Another significant suite of studies looking at the maintenance of segregation focused not on particular occupations and the social and discursive construction of particular skills or workers, but rather the manner in which firms and corporations organize and reorganize work. A classic of this genre was Rosabeth Moss Kanter's *Men and women of the corporation* (1977), which argued that organizational structures and the management of bureaucracies, and not the characteristics of workers *per se*, perpetuates gender differences in the workplace. Kanter argued that in large bureaucracies, the need for clear communications encourages managers to opt for 'homosociability', to stick with familiar and common experiences. Thus, male managers in male dominated parts of organizations are likely to appoint other men to their teams at the expense of those in the numerical minority (the so called 'tokens'). Kanter suggested that increasing the proportion of 'tokens' in any group was the key to changing the social dynamics of workplace groups, whatever their age, gender, ethnic, or racial composition. Kanter's arguments generated significant debate and critique because of her focus on organizational hierarchy rather than gender, yet her work has been important in illustrating how organization and hierarchy can produce different workplace behaviors and attitudes.

Some of the critiques of Kanter's work pointed to the fact that women entering traditionally male occupations faced greater hostility than men working in female dominated sectors and thus that gender does matter, beyond the organization of work, and requires systematic analysis. Moreover, feminists like Joan Acker have argued that organizations are saturated with gendered meanings, practices, and ideologies that shape the definition of jobs, job evaluation, promotion criteria, and other managerial practices. Labor market and other institutions, in turn, reflect and reproduce masculinist values which are institutionalized, normalized, and taken for granted. As Robin Leidner has argued in her examination of interactive service work in her book *Fast food, fast talk* (1993), one of the striking accomplishments of the social construction of gender is the illusion that gender variations in behavior, attitudes, appearance, and competence are 'natural' and not created. Leidner's work was also part of a shift to taking seriously the embodied nature of work, particularly in the context of the growth of service work in which the characteristics and performance of workers – their appearance, demeanor, and gender identity – is fundamental to the product being sold, be it a hair cut, some fast food, or a piece of legal advice. Gender and sexuality are thus 'performed' in the workplace as culturally embedded notions of masculinity and femininity shape both workers and their occupations.

In this vein, Rosemary Pringle's book *Secretaries talk: sexuality and power at work* (1988) made two important contributions to the literature. First, while many feminist analyses of women's work have examined the gendering of occupations by studying male dominated manufacturing jobs, Pringle examined a stereotypically feminized occupation. Second, she was innovative in using post structural theories of subject and discourse to weave together discursive and structural understandings of women's work. Pringle started out by noting the difficulty in defining what a secretary is, given the vast and often ambiguous range of tasks she undertakes in different contexts. She argued that attempts to define 'secretary' are repeatedly driven back to stereotypical notions of 'who she is'. She explored how three discourses – 'the office wife', 'the sexy secretary,' and 'the career woman' – played out in various media and everyday office life, shape and reproduce gender inequalities in the workplace, leaving male embodied sexuality in the workplace largely invisible. One of the key analytical shifts and contributions here was the recognition that institutional, cultural, and structural factors, and not simply the attributes and qualifications of individual workers, shape the gendering of employment. Feminist researchers have demonstrated that men and women do not enter the labor force with fixed and stable gender characteristics, rather gender identities and traits are assigned to workers on the basis of the jobs they do.

Challenging Earnings Inequalities

A third key contribution of feminist scholarship has been to translate academic scholarship into political action designed to improve women's position in the labor force. To take the above example, occupational segregation has attracted attention from feminist scholars not only because of its pervasiveness and its durability but, crucially, because of its relationship with another truism of women's work, namely their lower earnings compared to men. UN data from 2002 showed that in 27 of 39 countries surveyed – in OECD and industrializing countries – women's wages were 20–50% less than men's for work in manufacturing (see **Table 2**). Scholarship has thus been used to fuel campaigns to increase women's representation in high wage occupations, to challenge the so called 'glass ceilings' that constrain women's advancement in organizations, and to increase wages in women dominated occupations and sectors. Feminists have argued that earnings inequalities between men and women cannot be explained solely by differences in skills, experience, and educational attainment, and scholars in economics, sociology, and economic history have examined the gender wage gap, how it has varied historically and spatially and its sensitivity to educational attainment, hours, ethnicity, and occupational segregation. This scholarship has in turn fueled campaigns for 'Equal Pay' legislation and, in the USA for example, 'Affirmative Action' legislation (which seeks to improve women and minorities' access to higher paying jobs).

Further links between feminist academic and political work can be seen in analyses of gendered organizations and institutions discussed above. While early attempts at equal pay relied on job evaluations to assess the 'worth' of different jobs, scholarship on gendered organizations argues that evaluation methods contain gendered biases that downgrade jobs with any 'nurturing' content. Thus, labor unions and other institutions have turned to look at how pay is assigned to particular jobs and, in so doing, to go some way to addressing gender bias in how different jobs are valued.

Table 2 Women's wages in manufacturing as a percentage of men's wages, selected countries, various years

Country	*Year*	*%*
Australia	2002	89
Bahrain	2002	44
Botswana	2003	52
Brazil	2002	61
Colombia	2003	65
Egypt	2002	68
France	2002	78
Hungary	2002	74
Iran	2001	80
Ireland	2003	69
Japan	2003	60
Kenya	1997	123
Lithuania	2003	77
Mexico	2001	70
Norway	2003	88
Singapore	2003	61
Sri Lanka	2003	81
Thailand	2001	72
UK	2003	79
Ukraine	2003	69

Source: Compiled from United Nations Statistics Division http://unstats.un.org/unsd/demographic/products/indwm/ww2005/tab5g.htm.

Feminist Geographies of Work

This section focuses on feminist readings of work that embody a geographical imagination, spanning from the body to the globe. Geographers have actively engaged with feminist scholarship and sought to understand how the binary, and hierarchical, male/female has been written through the social production of space. The section includes reference to theoretically informed work which explores local and international labor markets and the gendered construction of globalization.

Gendered Local Labor Markets

Many economic and sociological analyses of women's occupational segregation paid little attention to the spatialities of gendered behaviors and divisions of labor. Some of the early geographic research on gender divisions thus focused on how individuals' daily time–space paths are shaped by domestic responsibilities and access to physical infrastructure, like transport, schools, and workplaces that shape and constrain mobility. Studies thus focused on, for example, commuting patterns and whether different groups of women undertake shorter or longer commutes than their partners. Literatures in geography, urban studies, and planning also analyzed the rise of suburbanization and the changing spatial separation of 'home' and (paid) 'work'.

In addition to exploring the spatialities of daily life that shape labor markets, geographers have also drawn attention to the uneven spatial distribution of different sectors and occupations. Given the pronounced occupational segregation discussed above, the geographies of men and women's participation are shaped, differentially, by the growth and decline of particular types of work. Thus, Susan Hanson, Geraldine Pratt, Doreen Massey, and others have examined the gendered implications of economic change, especially manufacturing decline and the growth of service sectors and occupations, in industrial areas characterized by unionized, male dominated jobs. Returning to the discussion about the social construction of skill, labor process theorists like Harry Braverman have argued that as work is deskilled by employers, women would be employed as the new 'unskilled' laborers, replacing men. Extending some of the Marxian accounts of women's work discussed earlier, feminist geographers argued that employers seeking to make use of divisions between workers to reduce their costs may thus be attracted to regions where there is a plentiful supply of cheap 'green', female (and indeed child) labor and that these workers constitute a reserve army of labor. Similarly on an international scale, literatures on the new international division of labor in the 1980s drew attention to the feminization of poorly paid manufacturing work as industrialized economies in the north started to relocate labor intensive manufacturing work, like electronics assembly and garment production, to lower wage economies in Southeast Asia and other parts of the globe.

Gendered Globalization

Another important product of a geographical feminist imagination has been engagement with critical development studies to explore the experiences of women in diverse social and geographical contexts and, specifically, the experiences of women beyond the (for the most part) privileged confines of North America, Australia, and Western Europe. This engagement entails an empirical shift in that it brings into focus the often neglected places and subjects of economic geography. More than this, however, it entails an analytical shift away from western centered knowledges that privilege capitalism, class, and globalization and toward grounded analyses of the experiences of marginalized groups and spaces that opens up possibilities for different ways of theorizing globalization. The research on changing international divisions of labor described thus, for example, generated vigorous debate around two areas. First, there were concerns that stereotypes surrounding women's employment – that women represented a nimble fingered, nonunionized and, relatively docile, labor force to be exploited by capitalists – again reared their head, with little recognition of women's agency and their capacities and strategies for resistance. Lesley Salzinger, for example, has undertaken ethnographies of work that reveal the embodied, performative context dependent production of gender identities in four macquiladoras in Chihuahua, Mexico. Salzinger demonstrates the variability and flexibility of gender identities and argues that women's productivity and docility did not predate the arrival of transnational firms in Mexico, but was instead produced by firms that moved to the region. Second, however, women's growing visibility in factory production in industrializing economies also fueled debate about the qualities of jobs being created and whether women's labor force participation improved or worsened the quality of life for these women. Many analyses of women's growing engagement with transnational nature of capitalist production – which is often in a geopolitical context of high levels of indebtedness and neoliberal structural adjustment policies imposed by western lending institutions argue that women's participation in low paid manufacturing jobs strains household dynamics and further increases women's workloads. These analyses drew attention to the necessity of recognizing and exploring the relationship between household and workplace, between production and reproduction, the formal and the informal and, in essence, questioning the supposed dominance of capitalist forms of production. An important accomplishment of feminists' research is a broadening of analysis to acknowledge that 'the eco nomic' takes place not only in the factory, but also in the home, the street, and elsewhere.

More recent feminist literatures on globalization ex plore issues of marginalization, displacement, and struggle through analyses of how globalization is pro ducing new transnational movements of women and children, particularly from industrializing countries in the south. Saskia Sassen has explored the links between the deteriorating economic position of different indus trializing economies and the growth of what she terms 'counter geographies', women's growing involvement in alternative economies of survival based on prostitution, illegal trafficking, migration, and remittances. Inter national networks now organize the transnational move ments of women who are pushed, lured, or coerced into formal and informal work as maids, nurses, mail order brides, sex workers, and so forth. Joni Seager estimates that about 500 000 women and children are trafficked into Western Europe each year, many from China, other parts of Southeast Asia, and parts of East and Central Europe where unemployment and debt have pushed households into poverty and fueled the growth of trafficking as a profitable industry. While poorer regions typically be come source regions for trafficking, rising disposable in come in other regions fuels demand for these women and children. Research on trafficking, prostitution, and the global sex trade and on women's mobilities as migrants and refugees thus reveals how these complex counter geographies are intrinsic to the economic, cultural, and political connections and flows which link work, eco nomic restructuring, and livelihoods in the north and south.

Debates, Intersections, and Concluding Comments

Different waves of feminist theory and practice have made incisive and lasting contributions to understanding the gendered construction of work. As such, discussion of feminist approaches to work is becoming a regular fea ture in textbooks and readers in economic geography, sociology, and development studies. Analyses have moved well beyond simply emphasizing women's (hitherto neglected) contributions to different forms of work and have expanded the definition of 'work' and sought to explain the social, cultural, and economic trends that have fueled the feminization of labor forces in many parts of the globe. Feminists have moved away from viewing gender as a fixed attribute of individuals and have instead identified the mechanisms, practices, and performances that construct occupations as more or less appropriate for (typically very variable) characteristics of masculinities and femininities. Feminist geographers have made im portant contributions to these debates by demonstrating the geographically constituted gendering of work in dif ferent sites (household, factory, office, and market) and at scales ranging from the body to international divisions of labor. More recent work on globlization demonstrates some of the complex links between women in different, yet connected, circuits of labor in the global north and south.

In terms of its future contributions, feminist scholars continue to explore the tensions between women's commonality of experiences of inequality and dis crimination and the importance of recognizing their different locations – in terms of age, ethnicity, class, sexuality, and so forth – in divisions of labor. Different feminisms will continue to broaden analyses of work, exploring the heterogeneity of women and their prac tices, paid and unpaid, which constitute their economies in diverse geographical, political, and cultural contexts. Yet, this sensitivity to heterogeneity will proceed alongside the political commitment to translate know ledge into political action. As feminists have long argued, any analysis of gender is also an analysis of power. In rapidly shifting, increasingly international divisions of labor in which women still perform the bulk of childcare and domestic labor, in addition to their other waged and unwaged work, feminist scholars and activists have im portant roles to play in the struggle for social justice and greater equality.

See also: Capitalism; Development I; Labor Market; Labor Unionism; Marxism/Marxist Geography I; Political Economy, Geographical; Postcolonialism/Postcolonial Geographies; Poststructuralism/Poststructuralist Geographies; Social Class; Spatial Division of Labor; Structural Adjustment.

Further Reading

Beechey, V. (1987). *Unequal work*. London: Verso.
Cockburn, C. (1988). *Machinery of dominance: Women, men and technical know how*. Boston, MA: Northeastern University Press.
Fuentes, A. and Ehrenreich, B. (1983). *Women in the global factory*. Boston, MA: South End Press.
Game, A. and Pringle, R. (1983). *Gender at work*. London: George Allen and Unwin.
Goldin, C. (1990). *Understanding the gender gap: An economic history of American women*. Oxford: Oxford University Press.
Hanson, S. and Pratt, G. (1995). *Gender, work and space*. London: Routledge.
McDowell, L. (1991). Life without Father and Ford: The new gender order of post Fordism. *Transactions of the Institute of British Geographers* 16, 400 419.
McDowell, L. (1997). *Capital culture: Women at work in the city*. Oxford: Blackwell.

Pringle, R. (1988). *Secretaries talk: Sexuality, power and work*. London: Verso.
Sassen, S. (2000). Women's burden: Counter geographies of globalization and the feminization of survival. *Journal of International Affairs* 53, 503 524.
Walby, S. (1986). *Patriarchy at work*. Minneapolis, MN: University of Minnesota Press.
Women and Geography Study Group (1997). *Feminist geographies: Explorations in diversity and difference*. Harlow: Longman.

Relevant Websites

http://www.un.org
Directory of UN Resources on Gender and Women's Issues.
http://www.unifem.org
United Nations Development Fund for Women. http://unstats.un.org
United Nations Statistics Division.
http://w3.unece.org
United Nations Economic Commission for Europe.

Feminism, Maps and GIS

M. Pavlovskaya, Hunter College – CUNY, New York, USA

Glossary

Cartographic Visualization or Geovisualization A set of techniques for representing spatial data for visual examination in a computerized environment.

Epistemology Part of social theory that establishes the relationship between the knower and the known or the subject and the object.

Feminist Geography Part of human geography inspired by feminist scholarship and focused on representing adequately women's worlds and transforming research practice in accordance with principles of feminist scholarship.

GIS (Geographic Information System) Computer-based technology for storage, management, display, and analysis of spatial information.

Visuality Use of visual information and images in production of knowledge and other social practices.

Introduction

Until recently, there was not much connection between feminism and cartography or GIS (geographic information science and/or system) but today they are increasingly intertwined. The oldest body of knowledge among the three, cartography, has a long and respected history as both the 'art and science' of mapmaking. It represents a unique way to communicate and understand spatial phenomena, relationships, and patterns. The meaning and purpose of mapping, however, have significantly changed in recent decades due to developments both within and outside the discipline of cartography. For example, the mapping process has become computerized, virtually all spatial information is now digital, and GIS has emerged as essential when working with spatial data. At the same time, women have increased their visibility and presence in science and technology and geospatial technologies play an increasingly important role in a variety of activist movements, including that of women. Finally, feminist scholarship has led to a profound rethinking of the practices of knowledge production in social science and geography, and it has come to inform and influence the fields of cartography and GIS.

The impact of feminist scholarship on mapping parallels its impact on knowledge production more generally. For centuries, women were missing from the practices of science as both its subjects and objects but in recent decades gendered economic, social, and cultural analyses have become more common. Furthermore, new approaches to research have developed rooted in feminist scholarship. In geography, too, women became more visible as objects of mapping and now as mapping subjects. Simultaneously, the rapidly growing field of GIS has had (as in other high tech fields) a contradictory impact on women. While presenting barriers to the entry of women into this field, GIS also provides feminist and other critical scholars with new opportunities for non positivist mappings.

Feminists analysis of and impact on science occurred just as GIS (and computer cartography) displaced traditional cartography. Simultaneously, the humanities and social sciences experienced a 'spatial turn' expressed in the increased use of spatial and cartographic metaphors. These movements have allowed feminist scholars to address the intersections of GIS technology, mapping, and cartography in a number of fruitful ways. They have examined cartography's legacy in the production of authority through mapping people and territories from a position of power (e.g., colonial, state, surveillance, or military). They have addressed the role that GIS plays in the continuation of these practices. They have reflected upon the professional role of women in cartography and GIS. And they have begun creating feminist cartographies – both representational of the landscapes and metaphorical.

Below, the undervalued contribution of women to both cartography and GIS will be examined and major feminist critiques of these fields shall be presented. New developments that demonstrate an ongoing engagement of feminism with GIS and practices of mapping will then be highlighted. Feminist scholars working to reclaim the power of vision and mapping by creating a range of feminist cartographies that utilize mapping in non hierarchical and non objectifying ways will be shown. In the last section, the broader impacts of geospatial technologies on women's daily lives, an issue of concern for feminist scholars shall be discussed.

Women in Cartography and GIS

Cartography and, more recently, GIS remain male dominated fields. Women's entry into these professions has been limited and, often, their contribution ignored. Yet, there is a history of women in cartography that, although unacknowledged, is significant and important to recount. According to scholars Alison Hudson (head of the map division of the New York Public Library) and

Mary Ritzlin, women have long worked in the map making trades in North America. Initially they were in volved mainly with printing (e.g., map engraving) or the map publishing businesses of their husbands, which they often ran after their husband's death. Because most women lacked training in the mathematics necessary to construct maps and could travel only in their husband's company, becoming explorers or surveyors themselves was difficult. However, women, as educators, are often the prominent disseminators of geographic knowledge. For example, in the nineteenth century, they authored many geography books, atlases, and histories for students and general readership. An ongoing history project that aims at recovering women's contribution to cartography in North America has identified as many as 300 such women.

Women's overall contribution is also undervalued because so few women have been in highly skilled or leadership positions in either the cartographic pro fessions or GIS. There are complex reasons for this situation that range from having to face many challenges upon entering the field to the observed gender differ ences in spatial cognition and mapreading skills (now attributed to socialization rather than inherent ability). In most countries, women remain a minority while being trained in cartography at universities and other schools. They commonly face masculinist model of teaching that further intimidates female students. Later they see themselves as 'intruders' into the men's world of geo spatial technologies. The conflict between personal as pirations (e.g., family) and the masculinist ethics of career oriented professional jobs also plays a role.

Despite recent changes, GIS remains male dominated in the US and Canada; it would seem that women face the same barriers to entry that exist in the information technology (IT) fields more generally. The impact of feminism on GIS, too, was limited in the past as feminist scholars initially felt the technology was not suited for feminist epistemologies. Today, the situation is changing. Both women in the workforce and feminist ideas more directly interact with geospatial technologies. Most ob viously, women are entering the field as professionals in greater (although still small) numbers. Their very pres ence slowly transforms the profession because women's more complicated career trajectories do not follow men's linear career paths. Beyond this, however, is the profound change in the very practices of GIS, mapping, and related knowledge production. First, there is a growing group of academic geographers who use GIS and are inspired by feminism (both women and men). Second, feminist scholars outside geography increasingly use spatial and cartographic metaphors in their work; they seek to create new nonhierarchical and nonexploitive spaces, spaces transformed by feminist idea(l)s. Together with the growth of GIS based feminist research, they constitute 'feminist cartographies'. Third, through access to infor mation and communication technologies and the Inter net, women who are not GIS specialists also gain access to geospatial technologies that change their daily lives. The sections below provide examples of these developments.

Feminist Critiques of Cartography and GIS

Inspired by feminist critiques of visuality and power to look in Western science by Donna Haraway and other theorists, feminist geographers began by examining the role of the masculine gaze in shaping modern geography and cartography. They also drew upon post structuralist critiques of cartography as a medium of power/know ledge pioneered by Brian Harley. Geographer Gillian Rose argued that the privileging of sight and looking as a primary mode of observation – whether assisted by modern technology such as satellites or GIS or not – elevated visual examination of the landscape such that it dominated geographic fieldwork. Such visual observation is premised upon maintaining a distance between the observer and the observed (the subject and the object), the practice that produced the supposedly detached, disembodied, and objective scientist whose only passion is pursuit of knowledge. At the same time, nature and the unknown in the Western tradition have become associ ated with the feminine – both threatening and beautiful, to be tamed and adored, to be looked at and understood. Thus, a male geographer simultaneously experienced the exploration of new lands or cities as a rigorous scientist, a conquering hero, and a pleasure seeking admirer. Therefore, Rose argues, modern geographic readings of the landscape not only express class or colonial power (as demonstrated by Marxist and postcolonial critiques) but are profoundly masculinist.

Furthermore, post structuralist, postcolonial, and feminist scholars revealed a particular connection be tween cartography and European imperialism. Cartog raphy sustained the imperial gaze through mapping colonized landscapes that, in the imperial imagination, were simultaneously equated with the feminine. In the words of Alison Blunt and Gillian Rose, "Maps are central to colonial and post colonial projects." They are "graphic tools of colonization" serving to make sense of the colonized spaces that are perceived empty and without meaning. The power to inscribe colonized landscapes and people with meaning is the power to establish control. "Imperialist maps not only describe colonies: they also discipline them through the discursive grids of Western power/knowledge." This power was amplified by the feminization of colonized landscapes. In imperial discourses, indigenous women symbolized the land and both women and the land became legitimate objects of colonization; both were seen as equally dan gerous and seductive and in need to be controlled and taken care of.

The advent of GIS and automated cartography led to the rapid integration of mapping with other geospatial technologies and digital information (e.g., Global Pos itioning System (GPS), remote sensing, tracking devices, cellphones, digital databases, and the Internet). It soon became clear that, together, such contemporary tech nologies could also support structures of power, imperial ambition, and unrestricted surveillance practices. In addition, feminist scholars exposed the gendered nature of GIS and examined the visual power of this mapping technology.

Donna Haraway has argued that satellites and other technological devices enhance the power of modern science by extending the limits of visual observation thereby making it seem more objective and disembodied. A similar observation applies to GIS. The computer enables the user to quickly access, display, and manipu late images of places and the visual excitement of GIS is magnified by the illusion of possession. As the fulfillment of the quantitative and spatial science revolution and its dreams of rational control, some have suggested that GIS may be incompatible with nonpositivist and feminist scholarship. Critics thought that 'as is' GIS was only good for supporting corporate and state interests as well as global imperial ambitions.

Reclaiming GIS as a Research Tool for Feminist Geographies

Despite the initial powerful critiques, feminist geog raphers are now rethinking GIS instead of rejecting it entirely. A feminist geographer Susan Hanson argues that, as with quantitative methods, GIS should be reconceptualized as a technology that, despite its limi tations, can be used by women and in feminist research. Indeed, Nadine Schuurman and Gerry Pratt suggest that rather than use social theory (e.g., feminism) to distance GIS, construct it as "other," we should develop mutual knowledge and "care for the subject." Further, Nadine Schuurman and Mei Po Kwan argued that although the social theorists have productively applied feminist cri tiques of visual technologies to GIS, they overlooked Donna Haraway's point that technologies are not in herently masculinist but constructed as such, and femi nists should learn to use them in an emancipatory manner. Warning that technological determinism may prevent critical scholars from using GIS, Kwan put forth arguments that GIS based visualization may be re constructed as a feminist method. In her 'A cyborg manifesto for GIS', Nadine Schuurman called for greater involvement of women in GIS. Sara McLafferty looked for the ways to link the dynamic but separate worlds of GIS and feminist geography by combining reflexive feminist methodologies with layered GIS representation of the world. Marianna Pavlovskaya demonstrated that contrary to common perceptions, GIS functionality is no more quantitative than it is qualitative, and critical geographers, including feminists, should seek new strat egies to integrate nonquantitative information and em ploy the visual power of GIS to advance socially responsible research agendas.

Although relatively new, theses developments have made a significant contribution to both the fields of GIS and feminist geography. Sara McLafferty even talks about the ongoing 'feminization of GIS'. In her view, its gen dered construction is changing toward a greater affinity with feminist practices of knowledge production. This involves a shift from feminist critiques of GIS to feminist analyses of its impacts on gendered social relations that involve gendered identities, spaces of everyday life, and women's activism.

Feminist Cartographies

Feminist engagement with cartography and GIS can be traced to several locations. First, cartography and map ping have become important metaphors in the human ities. They allow feminist artists and scholars to examine and rethink social practices and experiences in important ways. The South Asian Muslim artist Zarina works with maps to create prints of borders she crossed and that divide people (e.g., between India and Pakistan). She also creates prints of places destroyed by war. In another in stance, a feminist artist Ursula Biemann overlays satellite images with personal stories of women recruited into world sex trade to visualize 'female geobodies' – hidden global flows of female bodies from Southeast Asian or Eastern European countries to the countries of con sumption in Europe and North America. The Irish artist Kathy Prendergast creates maps of the female body in ways that are similar to how people map the land. Her work invokes representations of Ireland as femi nine in popular, nationalistic, and scientific discourses. While reminiscent of control and exploitation, these mappings also – as the artist attempts to do – suggest nonpatriarchal representations of both women and the land. Feminist artists also use nonpossessive mappings of the male body which subvert the primacy of the mas culine gaze and re inscribe mapping practices as non hierarchical. In so doing, they reclaim vision and mapping and reconstructs them as nonmasculinist practices.

Second, feminist geographers have turned to mapping and cartography as a way to reveal gendered experiences at various scales. Traditionally, maps hide women's worlds behind general population indicators but new research (and new maps) reflects their experience as well as the differences between experiences of women and men. To do this, feminist scholars use spatial data, statistics, and spatial analysis to construct social, economic, cultural, and political landscapes that make these gendered experiences visible. The most comprehensive example of such feminist cartography is *The Penguin Atlas of Women in the World* by Joni Seager, now in its revised third edition. Similar to other world atlases, it provides a global overview of various socioeconomic indicators by country but in this case international statistics are compiled to show phenomena that specifically affect women in different parts of the world. It presents maps which reflect the state of women's equality, health, political involvement, domestic violence, lesbian rights, employment, body politics, and the other dimensions of their lives. **Figure 1** is an example of a map that shows the unequal opportunities of women across the globe.

In addition, mapping and GIS based analysis focusing on women and gender have recently emerged in many disciplines including health related research. Examples include using distance as a factor in calculating the effects of prenatal exposure to pollutants from the World Trade Center disaster; modeling the distribution of rates of HIV infection among rural pregnant women in a region of South Africa in relation to the geography of the local road network that supports male migration; and examining a relationship between access to childcare and women's labor force participation in the Netherlands. In this latter case, mothers living in areas with a higher number of daycare slots per 100 children were more likely to be employed. **Figure 2** illustrates the very uneven geography of childcare availability mapped by the researchers. Visualizing such social landscapes that shape women's labor market experiences would not be possible without the computing power of GIS or the theoretical insights of feminism and its commitment to gender equality.

Third, a group of scholars has emerged that engages both GIS and feminist social theory. To the analysis of gendered experiences, they added a feminist rethinking of the role of GIS in knowledge production. In their research, social and feminist theory directly intersects with the analytical and visual power of geospatial technologies opening new possibilities for both feminist scholarship and GIS. For example, Mei Po Kwan has pioneered the modeling of gendered time–space geographies in a GIS environment. Drawing on Hagerstrand's ideas, she combined datasets that included street networks and the distribution of urban opportunities with data from travel diaries completed by individual women. Kwan analyzed these data using a dedicated algorithm that enabled her to

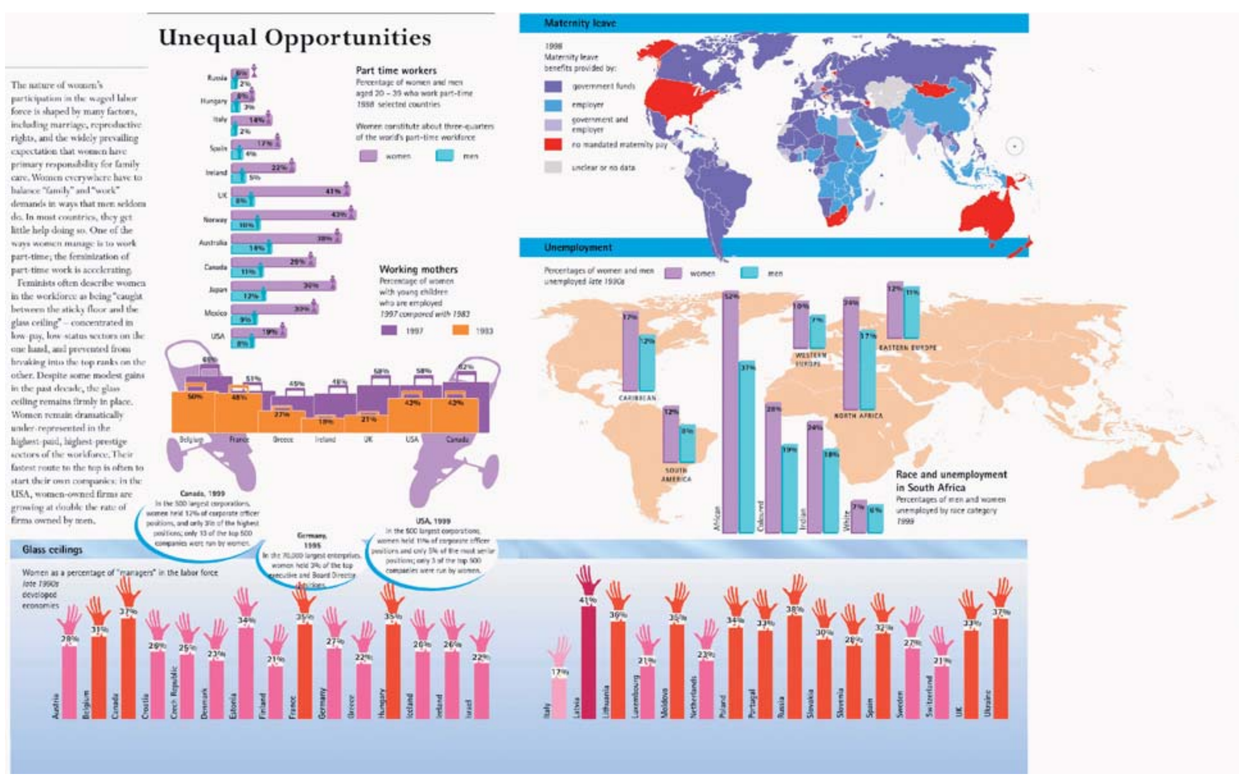

Figure 1 Unequal opportunities. From Myriad Editions and Seager, J. (2003). *The Penguin Atlas of Women in the World* (revised 3rd edn.). New York: Penguin Books. Reprinted with permission.

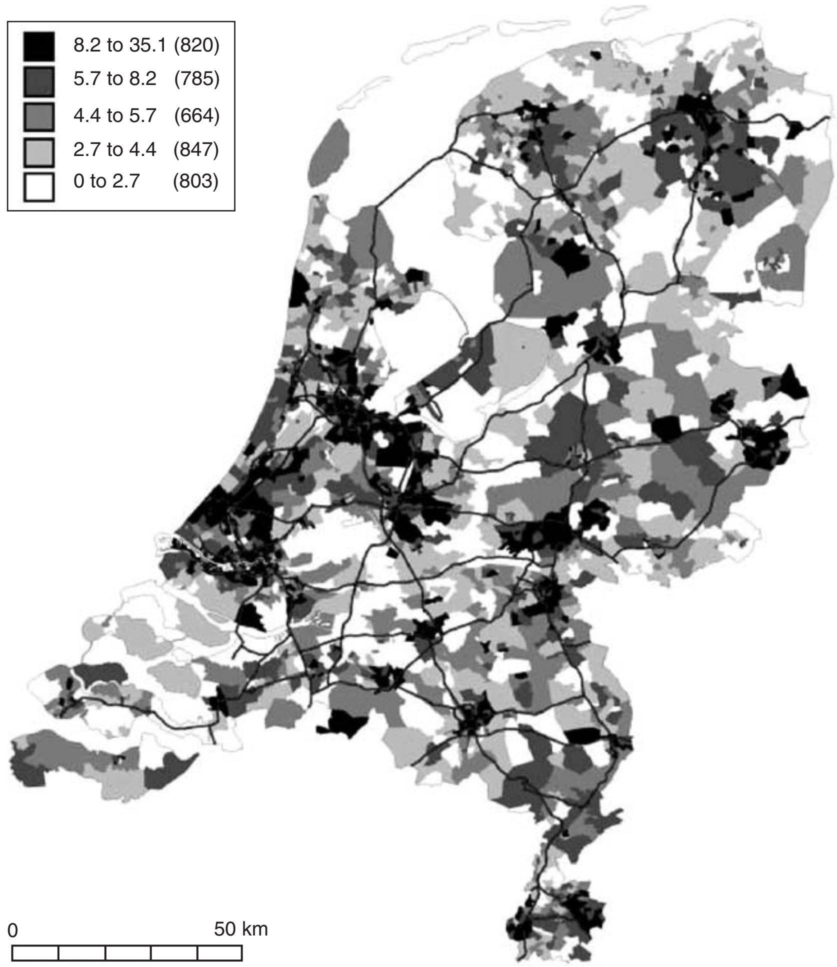

Figure 2 Geographical access to childcare within 10 minutes (slots per 100 children). Calculations using data from Monitoring Agency Childcare Provision (NUK 2001). From van Ham, M. and Mulder, C. H. (2005). Geographical access to childcare and mothers' labour-force participation. *Tijdschrift Voor Economische En Sociale Geografie* 96(1), 63 74. Reprinted with permission.

model the life paths of women as embodied experience. In particular, her three dimensional (3 D) visualization of women's daily movement through urban space takes into account the multiple stops they make (as opposed to the home–work commute of most men) and makes clear the gendering of access to urban opportunities. Furthermore, Kwan was able to account not only for the differences in uses of urban space between men and women but for differences between groups of women as well. In par ticular, her visualization has eloquently demonstrated that the life paths of African American women are more constrained compared to other women due to their rela tive economic disempowerment manifested in lower ac cess to automobiles.

Marianna Pavlovskaya has used GIS to visualize the multiple economies of the daily lives of Moscow house holds based on qualitative interview data. The economic landscape revealed by the interviews included not only commonly analyzed formal and monetized economic activities such as paid employment but also informal and unpaid economic practices such as work for cash, do mestic work, childcare, and exchanges of goods and services via social networks. W Pavlovskaya's analysis showed that contrary to mainstream economic theories, such informal economic practices persist and, moreover, underpin the everyday lives of households in modern socialist and post socialist 'capitalist' economies. She has argued for the greater use of GIS in critical geographic research to advance social theorizing based upon in novative visualization techniques, the incorporation of primary data that captures human worlds not reflected in standard datasets, and the merging of qualitative data with GIS.

Fourth, feminist GIS scholars have conducted re search on women's activism that involves geospatial technologies and they also have provided GIS based scientific expertise for women's community groups. For example, Sara McLafferty worked with a group of ac tivists from Long Island to map in a GIS the distribution of breast cancer occurrences in their communities. The data was collected by women themselves to make a case for the need to study the potential environmental causes of this disease as opposed to the dominant focus on in dividual risk factors. As a result, substantial government funding was secured for this research.

Geospatial Technologies and Women's Lives

Finally, there is the issue of the impact of geospatial technologies on women's everyday lives. Broadly defined, they include not only GIS but also the Internet, location based services such as cellphones, navigation tools, and other types of communication technologies. In particular, patterns of Internet (which is an immense source of geographic information) and technology use are gen dered and the evidence suggests that the impact of geographic information accessed through these means is also gendered; although at this point little is known about how this unrestricted information is used. Furthermore, these technologies have radically transformed the everyday lives of people and the effect on women, in particular, needs further investigation. For example, these technologies change the way we work, as in the case of telecommuting and the way we look for jobs, which lead to new types of gendered employment. Also, shopping and other household chores – typically the work of women – are increasingly done using the Internet. In addition to these individual uses, the Internet has become an important means for establishing and maintaining social networks beyond local communities and extending across national boundaries. These networks continue to serve as important sources of information for women in many domains of life, including prenatal care. The communities also extend to the virtual 'bridgespace' that supports flows of people, capital, information, and goods between, for example, North America and South Asia by facilitating the identification of marriage partners as shown by Paul Adams and Rhina Ghose.

The Internet, geospatial technologies, and information technologies in general, however, have uneven and con troversial impacts on women. The combined effects of class, gender, age, and race perpetuate the digital divide and large groups of low income, elderly, and women of color do not have access to such technologies and, therefore, are increasingly marginalized in terms of em ployment, consumption options, health outcomes, and political empowerment. The Internet also plays a crucial role in the sexual exploitation and trafficking of women and children and enables global flows of women's bodies from Third World countries and Eastern Europe to Western Europe and North America.

Another issue generated by geospatial technologies is that of privacy and surveillance in many domains including the relationship between the state and its citizens, corporations and their employees, producers and consumers, and individuals in their own private lives. The research suggests that these impacts are also gen dered as they follow already established patriarchal pat terns. For example, women are more likely to be targeted by advertisers owing to their greater role in consumption for their households or have their movements monitored by tracking devices installed by their partners.

Conclusion

In conclusion, the use of geospatial technologies and geographic information has rapidly grown in the last two decades. It is used by academics and scholars including those with a feminist sensibility, by professionals in many fields, and in the everyday lives of men and women. Digital maps and other types of geographic information are increasingly accessible to those outside the fields of cartography and GIS, including women in different professions and of different economic status, ethnic and racial identity, as well as geographic location. Feminism continues to make important impacts on mapping, GIS, and geospatial technologies more generally, but more work awaits ahead.

See also: Critical Cartography; Critical GIS; Geovisualization; GIScience and Systems.

Further Reading

Biemann, U. (2002). Remotely sensed: A topography of the global sex trade. *Feminist Review* 70, 75 88.

Gilbert, M. R. and Masucci, M. (2004). Moving beyond "gender and GIS" to feminist perspectives on information technologies: The impact of welfare reform on women's IT needs. In Seager, J. & Nelson, L. (eds.) *A Companion to Feminist Geography*, ch. 21, pp 305 321. Oxford: Blackwell.

Harley, J. B. (1988). Maps, knowledge, and power. In Cosgrove, D. E. & Daniels, S. (eds.) *The Iconography of Landscape: Essays on the Symbolic Representation, Design, and Use of Past Environments*, pp 277 312. Cambridge: Cambridge University Press.

Harris, L. and Harrower, M. (eds.) (2005). *Special Issue: Critical Cartographies. ACME: An International E journal for Critical Geographies* 4(1).

Hudson, A. and Ritzlin, M. M. (2000). A preliminary checklist of pre twentieth century women in cartography. *Cartographica* 37(3), 3 8.

Kwan, M. P. (2002). Is GIS for women? Reflections on the critical discourse in the 1990s. *Gender, Place and Culture: A Journal of Feminist Geography* 9(3), 271 279.

Kwan, M. P. (2002). *Special Issue: GIS and Feminism*. *Gender, Place and Culture: A Journal of Feminist Geography* 9(3).

McLafferty, S. L. (2005). Women and GIS: Geospatial technologies and feminist geographies. *Cartographica* 40(4), 37 45.

Pavlovskaya, M. E. (2006). Theorizing with GIS: A tool for critical geographies? *Environment and Planning A* 38(11), 2003 2020.

Pavlovskaya, M. E. and St. Martin, K. (2007). Feminism and GIS: From a missing object to a mapping subject. *Geography Compass* 1(3), 583 606.

Samantrai, R. (2004). Cosmopolitan cartographies: Art in a divided world. *Meridians: Feminism, Race, Transnationalism* 4(2), 168 194.

Schuurman, N. and Pratt, G. (2002). Care of the subject: Feminism and critiques of GIS. *Gender, Place and Culture: A Journal of Feminist Geography* 9(3), 291 299.

Seager, J. (2003). *The Penguin Atlas of Women in the World* (revised 3rd edn.). New York: Penguin Books.

Sheppard, E. S. (2005). Knowledge production through critical GIS: Genealogy and prospects. In Harvey, F., Kwan, M. P. & Pavlovskaya, M. (eds.) *Special Issue: Critical GIS. Cartographica* 40(4), 5 21.

van Ham, M. and Mulder, C. H. (2005). Geographical access to childcare and mothers' labour force participation. *Tijdschrift Voor Economische En Sociale Geografie* 96(1), 63 74.

Relevant Websites

http://lcweb2.loc.gov/ammem/awhhtml/awgmd7/
Library of Congress, American Women, Geography and Maps Division.

Feminism/Feminist Geography

L. C. Johnson, Deakin University, Geelong, VIC, Australia

Glossary

Patriarchy A theoretical conceptualisation of male power and control over women.

Women's Liberation Movement Emerged in the 1960s in Western Europe and North America as a political movement which defined women's oppression as the problem, identified men as the enemy and the personal as political in a quest to achieve women's autonomy and equality.

Introduction

Feminist geography is a relatively recent discourse in human geography. Emerging from the Women's Liberation Movement and feminist academic interventions in the 1960s and 1970s, it has involved a critique of geography – as a discipline with particular bodies of knowledge and set of research, field, institutional, and teaching practices – as well as its reconceptualization. Not only have key geographical concepts been rethought – such as class, work, mobility, migration, development, and even space itself – but feminist geographers have developed whole new subject areas – such as the home, the body, and sexuality in space – and modes of working within the discipline that are reflexive, collaborative, cross disciplinary, and situated. While varying across time and space, feminist geographers are united in their concern for the importance of gender in the constitution of space and social relations, in their indebtedness to feminist theorizing, and their commitment to understanding and ameliorating the inequitable place of women. Current debates cohere around transnationalism, economic inequality, and multiple axes of difference and identity. Feminist geographers approach these and other issues in uniquely critical, synthetic, and progressive ways that see gender as a fundamental analytical and organizing sociospatial category.

The discussion which follows charts the emergence of feminist geography from particular contexts in a number of countries. It delimits three trajectories in the discourse from the 1970s – historical, spatial, and theoretical – before considering the significance of feminist interventions into the discipline, and some current debates and sites of innovative practice. For this latter task, the article relies on a reading of the one journal devoted to feminist geography – *Gender, Place and Culture*. However, before proceeding, this overview needs some qualification.

For the very idea of writing an encyclopedic overview of a knowledge field, accepting by implication that all can be known, understood, and presented in one neat 10 000 word essay, undercuts one of the fundamental insights of feminism. In the 1980s, postmodern feminists convincingly argued that quests for the all encompassing theory or bodies of knowledge were in themselves patriarchal and suffused with modernist delusions of impossible grandeur. Such a context produces both a reluctance and ambivalence toward the task of producing this encyclopedia article. Having taken it on, the implications of the critique have been integrated into the account, such that the following is necessarily selective and can never in any sense be seen as complete or final. Consistent with a further feminist insight, the interpretations presented here also need to be personally and politically positioned. So this account is by an able bodied, Anglo Celtic academic who experienced the passion and personal resonance of discovering feminism, fought the academy for the recognition of women, and moved through a range of theoretical positions and empirical studies that were all immensely important and convincing at the time. Her personal trajectory is not unlike many other feminist geographers: coming to socialist feminism from 1970s Marxism, over the 1980s she worked hard to make women fit into its schema in a study of the Australian textile industry while also fighting for and then teaching women's studies in Australia. Moving to New Zealand in the late 1980s, she then embraced radical sexual politics both personally and professionally; accepted not only the insights but also the conservative politics of the postmodern turn in the 1990s to study Australian suburbia, and then engaged with the fraught politics of difference in studies of cultural capital as the new millennium dawned. Such positions do not of themselves excuse the absences in this account in predictable ways, for there remains agency, empathy, other identities, and an ability to traverse foreign intellectual terrains. But such a positioning does explain the highlighting of work from the edge of the Anglo American center – in Australia and New Zealand – and the emphasis on literature and debates that have appeared in English. The further focus on economic, social, urban, and cultural geographies and with Marxist, postmodern, and postcolonial theorizations, arise from these academic choices and contexts, though this does not preclude an awareness of vital work in, for example, development, rural, and historical geography, or in languages other than English. Rather, it is to suggest some of the limits of scale and scope in the process of creating encyclopedias of (all!) feminist geography – a will to power as breathtaking in its ambition as it is impossible in practice.

Emergence of Feminist Geography

Feminist geography as a critical discourse has an ongoing relationship with feminism and a range of other pro gressive political movements that emerged in the 1960s. In the context of a long post war economic boom and mass higher education, there were mobilizations of young people demanding black and gay civil rights in the US, left wing reform in Europe, and an end to imperialism in the Third World. Enmeshed in movements demanding social justice, newly freed from the necessary connection of sexual activity to procreation and in the paid workforce in ever larger numbers, 'women' came to see themselves as a group that also experienced disadvantage. The Women's Liberation Movement defined women's oppression as the problem, identified men as the enemy, and the personal as political. Across Europe, England, North America, and Australasia, inspired by Marxism and radical feminism, middle and working class women mobilized to demand equal pay, an end to sexual violence, and dedicated pol icies, programs, and services for women. Middle class, primarily white women, urged and achieved a greater presence in the institutions of power. Within the academy, a few women and some men began to critically examine the demographics, contents, methods, and theories of the humanities and social sciences, including geography, to document and demand changes.

As centers of political radicalism and a hegemonic Western culture, it is not surprising that it was in the US and England that the first systematic feminist cri tiques of human geography occurred. The earliest work drew not only on socialist feminist and radical feminist theorizing, but also on liberal feminist traditions to reflect differences in national cultures and an early diversity. Thus, in 1973, through the pages of *The Professional Geographer*, Wilbur Zelinsky documented 'The strange case of the missing female geographer' in the US. Acknowledging the vital trigger of the Women's Liber ation Movement and the presence of women's caucuses in other disciplines, Zelinsky saw geography as lagging behind, a consequence he suggested of wider patterns of sexism as well as particular institutional rules, tradi tions, and biases. His survey was followed by others in Britain, Canada, the US, and more recently in Japan, South Korea, Norway, Israel, Australia, and Germany, though such counts in what could be labeled as 'core' countries of Anglo American discourse are much less frequent now. In addition to the surveys of the gender imbalance within the academy, there have been overviews of publications, major conferences, and professional organizations which document and question the absence of women. Despite the range of sites, the timing, and the national contexts, virtually all of these surveys charted a similar pattern: of women's presence as geography undergraduates, but then a growing male dominance in the numbers of graduate students, and finally in university faculties, on editorial boards, in mainstream journals, as keynote speakers, and as 'eminent' in the discipline. So in 1990, between 0% and 10% of geography professors were women across the globe, with the situation improving somewhat in some countries – such as Britain, Australia, Canada, the US – up to 20% or 30%, into the new millennium, but still not to the level of demographic parity. The political task of securing women's equitable place within the discipline is therefore far from complete, though its urgency and prominence as an issue has lessened.

In addition to the demographic absences of women in geography, there were also early critiques launched into the systematic biases in the methods, empirical work, and even the theorizations used by human geographers. Thus, Alison Hayford in the newly established 'radical' journal of geography – *Antipode* – wrote in 1974 of how women were invisible within the discipline, assumed to either have no distinct role, be subsumed by a dominant masculine order, or to be continually adjusting to that order. Building on this in the first mainstream piece in England (in the journal *Area* in 1978), Jacqueline Tivers argued for women's spe cific role within urban society – as paid workers, as shop pers, and as carers – needed to be taken seriously. Four years later, Mary Ellen Mazey and David Lee, under the auspices of the Association of American Geographers, published the first monograph on feminist issues in geog raphy with their descriptive cartography of women across the US. Beginning a strong empirical tradition, *Her Space, Her Place* involves a series of maps and discussions which located women, documented their migration patterns along with their attitudes to and impact on the environment. This mapping of women's place began a tradition that has been taken up in Joni Seager's regular maps of women. Along with the use of geographical information systems (GIS) technologies, these later works engage increasingly with contentious political issues – such as Seager's recent atlas which charts women's status, civil rights, domestic violence, lesbian rights, literacy, and beauty cultures – across the world. Mapping the place of women, then, suggests that through such a technique, powerful political points can be made, depending on the questions asked and theoretical perspective adopted.

At the time that Mazey and Lee began the process of documenting women's presence in the US, Janice Monk and Susan Hanson wrote of the sexist bias in the content, methods, and purpose of Anglo American geography. Thus, in a 1982 article also in *The Professional Geographer*, they discussed studies of migration, development, suburbanization, and urban travel which did not consider female as well as male experience. Echoing Tivers, they suggested that subjects of direct importance or relevance to women – such as child care or unpaid work – were not deemed worthy of study, while any direct discussion of women assumed traditional gender roles. In their overview of research methods, they noted how the experiences of male respondents – as 'heads of households' – were often generalized to whole populations, while interview techniques further entrenched the sexist assumptions which, they concluded, pervaded the production of geographical knowledge.

If the work by Zelinsky, Hayford, Tivers, Monk, and Hanson critiqued the absence of women in geography departments, research, and methods, an exchange be tween Irene Bruegel and Pat Burnett in the increasingly Marxist journal of radical geography – *Antipode* – was the first to systematically connect feminist politics to left wing theorizing in the discipline. Thus, in 1973, Burnett argued that existing models of urban form were static, conservative, and assumed the sexual division of labor. She saw a Marxism enlivened by radical feminism as the alternative to an approach that ignored the structural relations between class, sex, and race. In her frame work, capitalism is built upon the family, an unequal gender division of labor, and a 'male psychology of power'. In her response, Breugel argued for a need to recenter the analysis on class and the Marxist framework, beginning a discussion which lasted for over a decade on just how socialist feminists could remain true to Marxist theoretical precepts and a woman centered, feminist politics.

The earliest statements of what came to be called feminist geography therefore emerged from the Women's Liberation Movement and from wider feminist critiques of academic knowledge and practices. Informed by the full range of feminist theorizations – including liberal, radical, and socialist feminism – a few key human geographers, women and men, critiqued the absence of women in the discipline, the ways geographical research was conducted, subsequent bodies of sexist knowledge, and left wing theorizations. Levels of activity varied, with more in the US than elsewhere, and national political traditions informing the approaches taken. Thus, a liberal feminist mapping of women predominated in the US, while Tivers brought a more left leaning approach to bear on the place of women in England, though such a position was also present in the work of Burnett and Breugel in the US. The critique of geographical know ledge, research practices, and theorizations as well as a focus on women's distinct worlds became the mainstays of much subsequent feminist geographical work, which also maintained an ongoing debt to feminist theories and politics.

Development of Feminist Geography

Charting the development of feminist geography can occur within at least three different organizing schema: an examination of the major shifts over time, theoretical emphases, and national differences. There are, of course, many other ways of mapping out the discourse – the matically, by subject matter, via approaches to space, within the subdisciplines of rural, economic, social, his torical, or cultural geography, through the work of key individuals, and so on. The approach used here is therefore an idiosyncratic one, though it does mirror at least some other summaries, personal experience, and serves to highlight important dimensions – of temporal and theoretical shifts across space. How these various developments have impacted on the discipline will be considered in a subsequent section.

Time Shifts

The nature of feminist geography has shifted dramatic ally over time, a reflection of changes in feminist theo rizing and politics, as well as of changes in the discipline of geography. In many surveys of the field, feminist geographers have noted a general temporal move: from a 1970s' and early 1980s' critique of absence and bias, to a 1980s' mapping of women which documented their in equitable experiences, to a greater concern with differ ence, representation, and identities from the 1990s.

Such shifts are well represented by work published by the Women and Geography Study Group of the Institute of British Geographers. Created in the 1970s, the group was a response to the growing level of activity on women in other disciplines and dissatisfaction with the sexist nature of geographical knowledge and the patriarchal structuring of their work environments. Agreeing by the early 1980s on the need to become more visible, in 1984 the group published the first major text book in the field – *Geography and Gender: An Introduction to Feminist Geog raphy*. It was a book where individual authorship was subsumed by the group on the cover but attached to individual essays. Within, feminist geography was de fined, given a history, and then used to inform work on gender in the urban environment, especially women in bingo halls and the problems women with young children faced in relation to mobility. In these studies, women were constituted as a group, differentiated by class, but basically united by an oppressive gender order which could be successfully apprehended, critiqued, and challenged by large and small scale political actions.

Comparable work then occurred within the develop ment field as Janet Momsen and Janet Townsend built on the path breaking work by Esther Boserup to demand that women be considered in any study of, and also projects for, economic development. So too in urban geography which had been interrogated through a special issue of *Antipode* in 1984 on 'Women in cities'. Much work had already been done by feminists in design, planning, and architecture – such as Dolores Hayden and Gerda Wekerle – who had only documented the absence of women in their various professions but then went on to study the ways in which cities and neighborhoods in the US and England variously built an inequitable gender order into them, primarily by the separation of homes from paid workplaces, the creation of a gendered distinction between public and private spaces, and the operation of transport and other service systems. This work then informed feminist geographers such as Jo Little, Linda Peake, and Pat Richardson in their 1988 study of *Women in Cities: Gender and the Urban Environment*.

This focus on women as a relatively homogeneous group united by their gender oppression, which was the mainstay of the early Women's Liberation Movement and work in the social sciences, was to undergo sustained critique from the late 1980s by minority women of color within Britain and the US and from those writing from the 'Third World'. In addition, theoretical developments from within feminist and humanities scholarship raised serious questions about any notion of unified subjects and placed fractured notions of power, social difference, and issues of representation squarely onto the academic agenda. The cultural turn within geography was also registered in feminist geography as problems were increasingly defined in cultural as well as material terms and textual analyses joined interviews and observation as vital sources of research data. Such shifts were registered in later editions of key feminist geography books; so that in their 1997 book, the Women and geography study group of the IBG RGS now focused on *Feminist Geographies: Explorations in Diversity and Difference* while the updated version of Momsen's development tome also reflected the change in its title: *Different Places, Different Voices: Gender and Development in Africa, Asia and Latin America*. Not only were the subjects different – with more emphasis on diversity and differences between women – but the approaches have also fundamentally altered. In the later Women and Geography Study Group book, there is an interactive approach as readers are invited to draw their own conclusions about key issues, consistent with the authors' approach of challenging the taken for granted assumptions of the discipline. The aim is to give multiple versions of feminist geography.

So from the 1990s, studies are infused with theorizations that direct attention away from economic and gendered structures and centers of power toward more dispersed, representational, and identity oriented studies of difference between women. A new generation of feminist geographers emerges, such that personal histories embedded in the 1970s Women's Liberation Movement give way to those who are reaping the gains of women's inclusion in the discipline, as well as engaging with very different theoretical debates. Contemporary studies thus become more focused on gender relations than the position of women *per se* with as much emphasis given to the place of men as women, on the experiences of racism and sexuality as much as sexism and the complex mutual constitution of these relationships within and across space. In such moves, dominant theoretical paradigms in the discipline and across the humanities and social sciences have been vital in shaping the timing and direction of change, as well as informing their politics, which has tended to become less about public actions than discursive interventions.

This trajectory has taken only slightly different forms across the subdisciplines. So, for example, Jo Little and Ruth Panelli in 2003 delimited three phases in feminist rural geography:

1. 1970s–1980s: politically motivated work that was informed by women's studies and focused on gender relations in rural areas, especially women's subordination compared to men, and a politics that aimed for greater equality;
2. Late 1980s to 1990s: concepts focused on feminist theories and gender relations in rural areas which valorized rural women, such that equality was replaced by a focus on and celebration of gender difference; and
3. 1990s onward: work focused on theories and concepts within rural studies and gender studies which looked at masculinities, femininities, and ruralities, that is, multiple identities with an emphasis on difference, diversity, and change along with studies of the social construction and representation of the categories studied. Such latter work is well represented by Jo Little's 2002 *Gender and Rural Geography: Identity, Sexuality and Power in the Countryside*.

So too with development geography, an initial focus on *Women's Role in Economic Development* informed the UN Decade for Women from 1975–85. Such research provided massive empirical detail and had the effect of making women in the South visible and showed how development programs impacted negatively upon them. Subsequent work focused on 'gender relations' rather than women and, while acknowledging differences, saw patriarchy as dominant and universal. In contrast, postmodern development work argues that both approaches disempower and universalize women and thereby focuses on difference, diversity, 'multiple identities', and articulations of power leading to contingent and contextual representations of gender.

Feminist interventions into rural, urban, and development geography are therefore comparable, up to at least 2000. After that time, from a reading of *Gender, Place* *and Culture*, multiple trajectories emerge: of consolidation and stocktaking, a mainstreaming of feminist theory, and concerns into geography's subdisciplines, and an array of cross disciplinary projects which transcend preexisting categorizations. Such work draws on psychoanalytic and post structural theory and feminist theorists such as Judith Butler, especially her notion of performativity; Iris Marion Young and her reformulations of social justice; Doreen Massey's view of fluid, open, and multifaceted space; and Elspeth Probyn's ideas of sexualized and spatialized identities; as well as newer nonrepresentational and emotional geographies, to produce very different feminist geographies.

In the new century, feminist geographers – or perhaps publishers – begin a process of consolidation and autocritique, producing a new generation of textbooks – such as Mona Domosh and Joni Seager's *Putting Women in Place* and Louise Johnson's *Placebound* – as well as drawing on their own history, to produce a consolidated collection of readings – such as Linda McDowell and Joanne Sharp's *Space, Gender, Knowledge: Feminist Readings*. The many interventions into geographical methods have also been collected into books, such as Pam Moss's *Feminist Geography in Practice: Research and Methods*. In addition, within the pages of *Gender, Place and Culture* there have occurred regular autocritiques, as the impact of feminist ideas into, for example, rural and historical geography, as well as on the doing of geography, is assessed along with the discourse of feminist geography itself. In the process, before postcolonial questioning, feminist geography is found to be wanting in its inclusion of non English speaking authors and voices from outside of the Anglo American center. Thus, in a fine evaluation piece published to mark the tenth year of *Gender, Place and Culture*, the Spaniard Maria Dolores Garcia Ramon joined the Dane Kirsten Simonsen and Greek Dina Vaiou to note how from 1994 to 2005, of the 242 authors of articles and viewpoints, only 19% were not based in Anglo American universities or research centers – which included Australia and New Zealand. Authors based in the US and the UK constituted 64% of writers, while of the 320 books reviewed, only seven were not written in English (2%) and only five reviewers were non Anglo American. Subsequently publishing all abstracts in Spanish as well as in English, with a real effort to include non Anglo Americans on its editorial board, accept papers in languages other than English, and as subjects of articles and book reviews, Audrey Kobayashi could still write that feminist geography faces the ironic situation that as more studies of racism and other forms of discrimination appear, the number of women of color doing those studies remains static and abysmally low at 2%. The shift then is from a consideration of women's absence from the discipline and its subjects, to the absence of those from ethnic minorities as writers, teachers, and subjects of geographical knowledge. A central contemporary issue for feminist geographers then, is 'difference', what it means, how minorities are constructed and confined to margins, and how this marginality can be ameliorated.

As well as general books which bring the insights of earlier decades of feminist geography scholarship together, there are those which venture into the newer territory opened by these postmodern and postcolonial perspectives. Perhaps best captured in 2004 by Walid Falah and Nagel's *Geographies of Muslim Women: Gender, Religion and Space*, the focus is on how difference and the complex intersections of space, identities, and power mutually constitute social groups in places. It is these subjects – of transnationalism, femininity, masculinity, cross gender bodies, the constitution of marginality, and social exclusion – newly interrogated by theoretical frames incorporating performativity, hybridity, sexualities, emotion, and difference which are informing current feminist geographies. In such work, women as dedicated subjects who are automatically subject to oppressive social relations are no longer central. Rather the place of women is now a question to be explored across time and space with a progressive, liberatory politics. Such studies explore the relationship of ethnicity, race, sexuality, religion, and ability to the definition and negotiation of gendered identities across space, while also documenting and challenging the complex ways in which oppressive social relations are constituted.

Theoretical Orientations

The preceding account has already suggested that theoretical frames have been critical to feminist geographers, shaping approaches to the discourse, to research and informing political actions. The following theoretical transect is somewhat traditional but these frameworks have structured many of the debates as well as much of the politics and practice that make up feminist geography. While not always deliberately invoked, these frames help to make sense of the diversity and abstract underpinnings of feminist geography. Having been considered earlier, the liberal perspective is only briefly considered. While radical feminism has a somewhat marginal role to play in feminist geography, it has nonetheless been vital in heightening the issue of sexuality in the discipline. But here, most attention will be devoted to those frameworks that have generated most activity and debate – socialist feminist, postmodern, and postcolonial feminism – before the section ends with speculations on where the current work is drawing its theoretical sustenance. The following theoretical orientations do not map neatly onto time and, while there are periods when one dominates over another, and times of course when, for example, the postcolonial critique did not exist while the frameworks do not neatly supplant each other but coexist in the present and at different locations.

The geography of women/liberal feminist geography

As noted earlier in the discussion of origins the earliest manifestations of feminist geography involved documenting the absence of women within the discipline and in positions of power as well as where and how women inhabit the world. Inspired by the Women's Liberation Movement, such approaches were indebted to nineteenth century liberal feminism in that they constructed women as a unified group, strove for women's visibility within existing structures and discourses, and pursued a reformist politics of inclusion. The framework therefore impelled the counting and mapping of women's locational disadvantage and spurred a range of measures to ameliorate these, including equal employment opportunity and affirmative action policies within geography departments, separate women's organizations in national and international associations, and planning interventions to ensure the better distribution of services.

Despite the somewhat dated nature of this agenda, it made a vital contribution to feminist geography; for the women's caucuses and action it inspired, it meant that women have been appointed and promoted, while women's issues have been legitimated in research. It meant that research could and did detail the effects of gender inequality while the constraints on women – as caregivers, housewives, and workers – via existing spatial structures, house designs, patterns of accessibility, service provision, and transport having been identified, could be reformed. Such research and action is ongoing, though it is intriguing that there are very few assessments of women's presence within the discipline and the Gender and Geography Study Group within the Institute of Australian Geographers was disestablished in 2004. While such caucuses continue in a host of other countries and within the International Geographical Union, it is worth speculating on the future of a 'geography of women' in the current political and intellectual climate, one which is not only informed and for many corrupted by neoliberalism, but also inspired by a focus on culture, ethnic, racial, and sexual minorities rather than on women *per se*.

The work conducted within the liberal feminist tradition has been criticized for being too descriptive, with women constructed as victims of their sex roles rather than as active agents. For many, inspired by different paradigms, women are positioned within larger structures of patriarchal, economic, and political power. These Marxist, radical, and psychoanalytic frames offer some at least greater explanatory power as well as a more radical politics.

Radical feminist geographies

Radical feminism underpinned much of the early political actions of the Women's Liberation Movement. With its focus on women as a united and oppressed group, male violence, and patriarchal culture, the impacts on feminist geography have been limited, primarily confined to work since the 1980s which considers fear and sexual violence in the city – such as by Susan Smith, Gill Valentine, Jalna Hanmer, and Mary Maynard in England, and Gerda Wekerle and Carolyn Whitzman in Canada and Australia, respectively – though the emphasis on sexuality and embodiment has inspired, if not directly informed, later work on sexuality and the body.

The work by Gill Valentine in Reading detailed not only the ways in which women's fear was constructed but also how it could inform the better design of public spaces. She documented how the elevation of public space as a site of inherent danger for women excluded them from a legitimate place and gave a false sense of safety in a far more dangerous place – the home – where most rape and violence against women occurred. Valentine not only argued for domestic violence to receive more media attention but also offered an array of practical planning actions that could make public space safer for women. This approach toward the redesign of public space – echoed by many feminist planners – has had major impacts across urban planning throughout the Western world, informing the rethinking and redesign of many public spaces to ensure a better sense of safety for women.

While radical feminism focused primarily on patriarchal or male power and sexual violence against women, its concern with personal–political issues, such as fear, oppression, and identity, also embraced sexuality and the body. These subjects have been pursued by an increasing number of geographers who have also adopted postmodern theorizations and post structuralist methods. Following the temporal trajectory already noted, there began in the 1970s and 1980s a mapping of gay spaces in some cities. In 1983, Manuel Castells identified an inner city, gay urban territory in San Francisco through the presence of gay bars, businesses, stores, and professional offices and votes for gay political candidates. He concluded that their presence reflected a profound gender difference: where men sought to dominate and delimit space as their own, while lesbian women were more scattered and had less localized relationships and networks. Comparable studies done of other cities in Australia, Canada, and England not only confirmed these patterns but also, as the postmodern critique developed, the idea of fixed sexual identities along with distinct gay geographies was challenged through the reworking of 'queer' and more fluid notions of sexualized identities and bodies in space.

Socialist feminist geography

It is not without significance that one of the earliest feminist geography interventions into theory focused on

Marxist geography. For it was the early 1970s and the radical critique of the discipline's conservatism and preoccupation with quantitative methods was just beginning through the pages of new journals such as *Antipode*. But it was to be another 10 years, however, before feminist geographers were seriously engaging with socialist politics and concepts, drawing on socialist feminist work on paid and unpaid labor to create totally new geographies of British, Australian, and Canadian cities.

The Canadian Suzanne Mackenzie set the scene in 1980 by critiquing the *status quo*, descriptive emphasis of the geography of women and calling for a new geography which examined how women shaped the environment via the social relations of productive and reproductive work. In the context of a vital socialist feminist movement, Linda McDowell in England pursued this approach in her historical analysis of British cities and their suburbs. Thus, she argued – following one line of socialist feminist theorizing – that production and reproduction were part of a single process which varied across space and time. She charted the development of British suburbs as the outcome of a set of historically specific, contested, and actively created relations between male and female paid work, housework, and domestic ideologies. The result of these changing social relations, she argued, was that suburbs were quite different in the 1950s compared to the 1960s and 1980s, as the nature of work and relations between paid and unpaid work, as well as ideological constructions of gender roles, altered.

In a series of locality studies generously funded across the UK in the 1980s, the relationships between productive and reproductive work were systematically studied by geographers, sociologists, and economists. Along with a critique of Marxist inspired studies of the capitalist labor process went a new approach to paid work, one which owed much to feminist theoretical and political interventions. Across the UK, studies of manufacturing and mining decline detailed not only the ways in which this process impacted disproportionately on men but also how it was accompanied by a growth of new industries in green field sites where women were mobilized into paid labor. Explaining such patterns focused on how the sexual division of labor within the home, in paid work, within trade unions, and in localities was altering. Doreen Massey in her 1984 *Spatial Divisions of Labour* was critical to this new approach to space, labor, and gender, as she documented the fate of different industries, their international as well as their national and internal organization, in these terms.

Working with Linda McDowell, Massey also examined how capitalism and patriarchy articulated to create a set of distinctive conditions for the maintenance of male dominance at various localities, be it in the mining villages of Northern England, the Lancaster cotton towns, in the London rag trade, or rural life in the Fens. In this 1984 essay, Massey and McDowell argued that capitalism presented patriarchy with different challenges across space which in turn shaped gender, class, and labor relations, thereby prioritizing a key feminist term over those derived from Marxist theory.

The invocation of patriarchy and its complex relationship with capitalism and Marxist categories became the subject of ongoing discussion within feminist geography in the late 1980s. Such discussion culminated in 'the patriarchy debate' which raged over 1986 and 1987 in the pages of *Antipode*. Here then was a core feminist concept, one derived primarily from a radical feminist agenda, that numerous feminist geographers were trying to connect in some meaningful way to their Marxism and to studies of space. In their opening essay, Jo Foord and Nicky Gregson moved to reconceptualize patriarchy within a realist framework, isolating its necessary (men and women, biological reproduction, and heterosexuality) and contingent elements (patriarchy experienced at particular places and asserted within marriage and the nuclear family in 1980s Britain). It was primarily the consignment of heterosexuality and gender to the status of ahistorical universals that generated the most vehement responses.

The intensity of 'the patriarchy debate' was real and typified many of the dilemmas faced by socialist feminists as they tried to reconcile Marxist and feminist concepts and political priorities. However, the urgency was to quickly dissipate as the iron curtain came down and the end of ideology was trumpeted. Before a rising cacophony of voices from minorities demanding to be heard, neoliberals triumphed across the Western world, overseeing an economy, society, and set of technologies now defined in terms of the modern and the postmodern rather than the capitalist and socialist. Feminist geographers too were increasingly engaged with these critiques and the very different approaches coming from postmodernity and postcolonialism.

Feminist geographies of difference – postmodern and postcolonial

If the late 1980s was distinguished by the demise of ideology, it was also associated with global economic and spatial restructuring, as developed economies experienced a crisis of profitability and some 'Third World' countries attracted huge capital investments and became massive exporters of consumer durables. The introduction and widespread use of computer technologies in communication, production process, and for creating new consumption sites revolutionized industry and popular culture. Subsequently labeled late capitalism, New Times, postindustrial and disorganized capitalism, this shift was also typified as a move from a modern to a postmodern era. For geographers such as David Harvey and Edward Soja, the 1980s were distinguished by a more pervasive globalization of Western economic, social, and political relations, a shift from a Fordist form of mass production toward a disorganized, flexible, or post Fordist way of creating goods and a change from manufacturing to service economies dominated by relations of consumption rather than production. Accompanying these structural economic shifts went changes in a esthetics and philosophy, as accepted truths, certainties, and frameworks derived from the Enlightenment around rationality, the unified individual, and science were thrown into question.

As part of this critique, in historical studies of medicine, the prison, and sexuality, Michel Foucault argued for the importance of discourses – ways of writing and speaking by those in particular positions of power which constructed subjects and discourses – to revolutionize the ways in which not only the past but also the present was studied. The widespread application of this approach in literary, cultural, and media studies and latterly human geography demonstrated the importance of textuality, sign systems, and deconstructive techniques to the constitution and apprehension of the postmodern world and to those who lived within it, now conceived as multifaceted and discursively constituted subjects. Such work, along with the psychoanalytic theories of Freud and Lacan and the philosophy of Derrida, Deleuze, and Guattari entered feminist thinking – via such critical writers as Judith Butler, Elizabeth Grosz, and Elspeth Probyn – with the result that desire, the sexed body, language, and culture became new priorities for feminists and feminist geographers.

In addition to the textual and cultural turn in social theory, went another which was more grounded in politics than the academy. For the crisis of postmodern representation was also a crisis of European cultural authority in the face of critics from colonized countries and marginalized minorities. In this critical milieu, feminism was far from exempt. Thus, in the late 1980s, feminism was subjected to a postcolonial critique as its claims to universality were exposed as illegitimate. Many feminist claims and theories were exposed as focusing on and representing the voices of white, Western, heterosexual, and middle class women. Launched by minority and colored women within the West as well as by women from Third World countries, this critique forced a rethinking and reorientation of feminist geography as well as feminism as a whole. Thus in her engagement with *Feminism and Geography*, Gillian Rose argued in 1993 that despite over 10 years of feminist work, it remained profoundly masculinist, as feminist work is brushed off and the place of gender in space and geographical knowledge is rendered marginal by those who refuse to acknowledge their privileged positions.

The critique offered by feminists of established knowledge, research methods, and disciplines, as well as alternatives offered by minorities and those from (de) colonized countries, can readily be seen as part of the broader crisis of representation. However, in geography, in the analysis and theorizing of postmodernity, globalization, and late capitalism, the most 'important' works did not acknowledge any debt to feminist or postcolonial thinking or recognize the particular place of women in the new economic and social orders being created, be they in Third World export processing zones or in the newly restructured industries and cultures of the First World. In this postmodern era, it was to be the new tool of textual deconstruction that was to be turned upon such key works, while the complex ways in which cities, workplaces, and workforces were discursively created were to be documented by postmodern feminist geographers.

Thus, in a spirited engagement with some key postmodern theorists Liz Bondi, Rosalind Deutsche, Doreen Massey, and Linda McDowell amongst others noted how David Harvey's *Condition of Postmodernity* and Edward Soja's *Postmodern Geographies* were founded on unexamined sexism and profound phallocentrism which made these books partial and deeply limited. They variously argued that Harvey and Soja constructed a view of the postmodern condition that ignored women and feminism. It was argued that both books failed to acknowledge that patriarchy and racism were fundamental aspects of oppression within both the modern and postmodern eras and that feminist, antiracist, and postcolonial struggles were significant political movements in shaping these periods. Further, neither author recognized their positions of discursive power as white, heterosexual, male geographers, or acknowledged that they themselves were offering a totalizing framework rather than accepting that such a claim was no longer possible.

Such deconstructive techniques were also central to the first solo authored feminist geography book by Gillian Rose. Thus, *Feminism and Geography* goes beyond simply noting the absence of women in mainstream geography texts and traditions to implicate time geography and socialist feminist geography in a set of binaries which privilege culture over nature, the mind over the body, reason over emotion, work over home, and production over reproduction, and therefore men over women in a deep and formative way. Rose thereby argues for an alternative feminist geography, one which does not just reverse such binaries but goes beyond them to create new spaces for women, seen not as a unified but as a highly differentiated, discursively and politically constituted grouping.

Other feminist geographers have joined the critical interrogation of the post Fordist and globalization frameworks. Thus, the American Julie Graham noted in 1992 how these narratives were totalizing and exclusionary and, in line with a material conceptualization of discourse, further argued that such ways of thinking had real political effects. For in presenting an analysis where capitalism is the only pathway and globalization the only script, any sense of agency is removed and any resistance doomed from the start. Graham joined with the Australian Katherine Gibson to present not only a fundamental critique of this conceptualization of capitalism but also to offer an alternative economic and social vision. Thus, in 1994, Gibson Graham wrote of *The End of Capitalism as We Knew it* by deconstructing Marxist and postmodern theorizations. In this book and in their later *Postcapitalist Politics*, they use feminist and other liberation theorizing to construct an alternative economics. Beginning with a general concept of labor application and the appropriation of its products to highlight sites of economic activity – such as the home, the informal sector, crime, and exchange systems – in First and Third World countries – they argue that these various places are every bit as vital theoretically and practically as capitalist forms of production and exchange. Just as importantly for their political project, such noncapitalist forms of production and exchange are often less exploitative, violent, and unstable and allow very different identities, alliances, and economies to flourish. Here then is an alternative feminist economic geography, grounded in a progressive politics as well as in postmodern and postcolonial theorizing, one which is discursively powerful and politically mobilizing.

If geographical knowledge was thereby subjected to a new round of politicized deconstruction from the early 1990s, feminist geographers have also applied these frameworks to the analysis of cities and workplaces. Thus, in 1991, Linda McDowell engaged directly with the post-Fordist regime of production, but to the analysis offered by Harvey and others, highlighted the role of women as casualized and sexualized members of the rapidly expanding service industries. She takes this analysis further when in later research, she documents the importance of gender relations and, in particular, the construction of various forms of masculinity to the nature of London's banking and finance sector for her 1997 book *Capital Cultures*. In this work, the nature of the gendered body in space becomes critical, as dress codes, demeanor, and the very nature of the work task becomes integrated into particular bodily types, activity patterns, and looks.

While radical feminism had highlighted the importance of sex and sexuality to the oppressive gender order of the 1970s, it was not until the cultural turn in geography and the widespread use of deconstruction, queer theory, and the approaches offered by Judith Butler and Elizabeth Grosz that feminist geographers focused on the sexed body in space. Thus, in 1996, Nancy Duncan edited a book on *Body Space* in which a number of feminist geographers engaged with the new questions associated with postmodern research: how the cultural construction of gendered identities impacted on material spaces how sexualities intersected with space to produce 'the heterosexual street', gay communities, and masculinities in various work spaces. Accepting Grosz's arguments – drawn from Foucault – that the body carries traces and inscriptions from a range of medical, legal, and advertising discourses, feminist geographers subsequently studied how women's bodies move in space and how they assume particular forms through sport, weight training, or expectations associated with reproduction. Thus, in New Zealand, Robyn Longhurst's *Bodies: Exploring Fluid Boundaries* detailed how, for example, the pregnant body is constructed as frail, disorderly, and transgressive to challenge the mind–body distinction not only within philosophy but also in everyday urban design.

In a detailed study of Worcester, Massachusetts, over the 1990s, Susan Hanson and Geraldine Pratt apply a comparable framework to a place to document the ways in which class, gender, race, and ethnicity variously come together to create communities, particular home–work relationships and workforces, as well as complex individual identities. Thus, in *Women, Work and Space* as well as in other studies of workplaces, cities, and workforces, notions of multiple identities, decentered power structures, and the mutual as well as discursive constitution of places and people are utilized by feminist geographers to create not only incisive and insightful studies but also alternative spaces for analysis and progressive action. Such a focus on the vital role of ethnic difference in the creation and changes in places has been echoed in a host of more consciously postcolonial geographies. Emanating unsurprisingly from those countries which were either at the center or on the edges of empires, work by Jane Jacobs and Kay Anderson in Australia, Canada, and England, Alison Blunt and Gillian Rose in England among many others, have transformed historical geography into a systematic interrogation of the ways in which the complex vectors of colonial power constituted not only gendered but also racialized peoples and places, be they an abandoned brewery in Perth, Western Australia, Chinatowns in Sydney and Vancouver, or Victorian women and men traveling across colonial Africa and India.

If deconstruction offered a new tool for analysis, postmodernity a new vision of identity, and textuality and postcolonialism a very different interrogation and politics of difference, feminist geographies in the 1990s expanded well beyond the dualism of women and men to admit race, ethnicity, age, disability, nationality, religion, and sexuality into their works. Feminist geography therefore now documents how multiple axes of difference are realized in and through space. Feminist geography also draws on a broader range of social theory to better understand the research process and to inform research methods, which have shifted away from objectivist epistemologies toward situated knowledge claims and alliances.

National Differences

Documenting national differences in feminist geog raphies on the one hand appears a straightforward quest, as place clearly matters to geographers! However, any expectation of major cleavages between nations flies in the face of academic traditions which ensure that only those engaged with international scholarship are pub lished and promoted, while the international job market as well as conference circuit ensures that at least those geographers with access to resources are usually highly mobile between countries. As well as a tradition of Third World countries going to the First for higher degrees, there is also a growing propensity for international col laborations, primarily between English speaking aca demics, between the First and Third World scholars and across linguistic divides. So in this increasingly globalized world, national differences in academic cultures are not fixed or pronounced. But they do exist and relate to discursive power centers which form national and local traditions, which in turn are shaped by key individuals.

Alongside tendencies toward convergence, there is a differentiation to be made between the First and Third World feminist geographers as well as between those within the developed world who have English as their first language and those that do not. Thus, as noted earlier, there is a preponderance of Anglo Americans within the discourse as well as those who are part of the English speaking diaspora in Canada, Australia, and New Zealand. Those countries which have English as part of a multilingual tradition – usually the result of colonization – such as India, Singapore, and South Africa – attain some visibility within the mainstream literature, but for those geographers based in Europe, Asia, South and Central America, and Africa, they are truly on the edge of the feminist as well as geographical discursive spaces – a power geometry of knowledge circulation that neces sarily renders invisible much work which may well be relevant to this English language review.

Within those English speaking countries that have seen published feminist geography activity, some national tendencies can be discerned. There are local intellectual traditions, centers of academic work, key individuals, and groupings as well as national traditions of enquiry and politics. Within the UK, there is a strong tradition of socialist feminist activity as well as a more recent ten dency to engage with the implication of the postcolonial critique in this once mighty imperial power. In contrast, the US has a less pervasive left wing tradition and con trasts with Canada in this respect. In Australia and New Zealand, consciously drawing on these centers, there has been significant work done not only within socialist feminism but also more recently within postcolonial frameworks from the very different position of a colon ized country.

But there are also particular feminist geographies which emerge from the geography, feminist traditions, and particular individuals in countries. The role of key people cannot be underestimated, for they have provided not only the intellectual leadership, mentoring, higher degree supervision, networking skills, institutional, and personal support for the creation but also the shaping and sustaining of specific feminist geographies. Thus in Britain, the Women and Geography Study Group of the IBG (latterly combined with the Royal Geographical Society as the RGS IBG) as well as the presence, pub lishing prowess, and networking skills of Liz Bondi, So phie Bowlby, Alison Blunt, Jo Little, Linda McDowell, Doreen Massey, Linda Peake, Gillian Rose, Joanne Sharp, Jacqui Tivers, and Gill Valentine have been vital. These individuals have also shaped the theoretical trajectory of British feminist geography, which owes a particular debt to socialist feminist and postcolonial thinking. The par ticular bifurcated professional role of Liz Bondi as both psychoanalyst and feminist geographer has also impacted on the way in which British geography has moved from post structuralism toward nonrepresentational and then emotional geographies.

So too in Canada, the role of Suzanne Mackenzie in shifting the emerging radical feminist discourse de cisively toward a socialist feminist perspective, effect ively changed the course of the discourse in North America as did the presence of Geraldine Pratt at the University of British Columbia in Vancouver and Linda Peake and Gerda Wekerle at York in Toronto make these two cities real centers of feminist geography re search and practice. In the US, Mona Domosh, Jan Monk, Karin Morin, and Joni Seager have all been es sential to the ongoing presence of feminist geography, utilizing a liberal feminist framework to good effect and also by drawing on and developing other approaches to engage with fundamental issues of transnationalism for workers and migrants as well as neoimperialism by the US in Latin America. In Australia, Ruth Fincher and Katherine Gibson have been critical leaders and mentors for younger women while offering decisive intellectual interventions into economic and urban geography, while the internationally mobile Jane Jacobs and Kay Ander son have been vital contributors to postcolonial dis course around the world. In New Zealand, an ongoing commitment not only to look outward but also to forge autonomous intellectual spaces have been sustained by key individuals within very few institutions – especially Anne Magee, Robyn Longhurst, and Linda Johnston at the University of Waikato. Their work on sexuality and space and on the body in space has impacted feminist geography around the world, and they have been in strumental in networking younger women across the country to continue pushing the boundaries of feminist geography theory and practice.

Within Europe, the picture is less clear, partly because of my own linguistic limitations, and also because the feminist geography project within particular European countries has been focused on their own economies and societies and, for some, such as Germany, has not impacted significantly at the institutional level. So, for example, there has been an emphasis in Sweden on paid work, while other Nordic countries have a particular interest in work and family relations as well as the role of the welfare state. In the Netherlands, feminist geography has connected strongly to planning, while in Spain it has engaged strongly with rural geography. In German speaking countries, a number of working groups exist which have explored a range of theoretical issues – added to by the queer emphasis of the Doreen Massey Weekend Reading Group – but feminist issues are not on the official agenda of German geography and remains marginal, with few faculty and an 'add women/gender and stir' approach. Even those who see themselves as feminists, pursue feminist studies as a background rather than a core activity. There is also a lack of institutionalized networks equivalent to the RGS IBG and IGU groups, though this facilitates informal networking. Within Asia, again, there are key individuals within particular countries and institutions – such as Hae Un Ri in South Korea and Saraswati Raju in India. In Singapore, Brenda Yeoh and Lily Kong have focused on transnational migration and investment as well as on the gendering of state policies, work that has clearly emerged from the particular history and geography of their island nation.

National differences in feminist geography as a whole are echoed within subdisciplines. Thus, in their overview of feminist historical geography, Karin Morin and Lawrence Berg discern particular national traditions – in the US, there is a focus on the frontier and work emanating from the landscape school of Carl Sauer, which provides the vital context for Vera Norwood and Jan Monk's collection: *The Desert Is No Lady*. In contrast, work in Britain focuses on imperialism, leading to studies of Africa, India, and women travelers as well as the nature of geographical knowledge as integral to and complicit in the colonial project.

In this academic work feminist geographers have effectively entered the academy, academic journals, conferences, and the public arena transforming the discipline of geography, including its demographics, its content, methods, and theoretical frameworks. But in highlighting the fundamental role of key individuals in particular countries, this account also notes a limitation of feminist geography – its ongoing indebtedness to relatively few scholars, who came to the cause via the Women's Liberation Movement in the 1970s and who are now relatively senior and secure within the academy. Mentoring, supervising, publishing, teaching, and networking, these scholars are also aging, increasingly drawn into management positions and will soon also leave their institutions and abdicate leadership roles within the organizational structures supporting feminist geography. What remains will lie in the hands of their successors, who are coming of academic age in a very different political and theoretical era, one where gender is no longer fundamental and where issues around race, ethnicity, religion, sexuality, economic inequality, identity, and emotion are now central. What is now emerging is therefore a very different feminist geography, and in the case of Australia at least, such changes have seen the demise of the Gender and Geography Study Group within the Institute of Australian Geographers and the evacuation of the category by many. This may or may not be the future for other places, especially Britain, the US, Canada, Spain, and New Zealand, where there is a critical mass and ongoing senior leadership, but it does highlight the vulnerability of small and ultimately marginal group of geographers, if not discourse.

Significance of Feminist Geography

The previous transects have highlighted aspects of the long and complex history of feminist geography. It now comes to discuss more systematically its practical, discursive, methodological, and political significance.

Practices

Fundamental to the emergence of feminist geography and its earliest concerns in the 1970s and 1980s was the low representation of women – in geography departments, publications, research, and teaching – and women's exclusion from geographical knowledge, along with the assumption that women were subsumed by studies of men/heads of households/humanity. Alongside such observations went a critique of fieldwork, as a highly masculinized activity.

What has occurred subsequently has been the organization of women's caucuses within the discipline, a growing presence of women in academic geography, and a transformation of the ways in which geographical research was conducted and geographical knowledge produced and circulated, such that a more gender aware curriculum, staffing profile, and set of teaching and research practices have emerged. It can be safely assumed that in those places where these activities have cohered – for example, at the University of Waikato in New Zealand, at Melbourne University in Australia, at the Autonomous University of Barcelona, at the University of British Columbia and York University in Canada and at the NUS in Singapore, Cambridge and Open University in England – that the changes will be permanent and gender balances will be maintained at the graduate and faculty levels and perhaps within their national–geographical canons. In other institutions and countries, change in practices may or may be maintained, through initiatives to work cross culturally and transnationally may facilitate the ongoing dissemination of feminist geography practices as well as discourses. In addition to these structural positions, there has been a related curriculum presence in tertiary teaching across the world, including Australia, Britain, Spain, South Korea, India, Japan, US, Canada, New Zealand, Singapore, and Norway. Indeed, some feminist geography has become part of the normal curriculum in most departments of geography.

In addition to changes in institutional teaching and hiring practices, feminist geographers have challenged and transformed the ways in which knowledge has been generated and presented. In line with feminist practices in other disciplines, first person narrations/authorial identification have been central to the doing and pre senting of feminist geographies. But there have also been new ways of writing/personal positioning. Thus with the visibility of J. K. Gibson Graham, the Women and Geography Study Group of the RGS IBG and the Feminist Pedagogy Working Group, collaborative and collective ways of researching and publishing have been successfully developed. The place of the feminist aca demic as someone who works within and from the academy has been broadened to embrace the collabora tive agent, who actively creates academic knowledge in the process of engendering positive social change. This can be seen in the work of Gibson Graham in the Phil ippines and of Gerry Pratt's with Philippina women in Vancouver.

Discourses

In addition to altering the demographics and practices of geography, feminist geographers have been critical in rethinking key terms, frameworks, and methods in the discipline. Across the subfields, feminist geographers have offered their unique perspective to ensure that historical, economic, political, urban, development, and rural geography are all not only aware of gender differ ences but are engaged with feminist theorizations. But feminist geographers have also interrogated their own work as much as that of others, such that in the 1980s' feminist studies of fear and the journey to work came under fire from minority positions, leading to its reworking.

Feminist geographers have also intervened to rethink key concepts and approaches in the discipline, including 'development' – to include women's unpaid and caring labor – 'gentrification' to ensure that its gendered nature was acknowledged; journey to work studies which now not only highlight the gender but also the racialized and ethnic differences, and migration studies which now ac knowledge and focus on the multifaceted identities of those involved. These studies join other analyses of public space, domestic life, and occupational segregation where there has been a reworking from within by femi nist geographers, with a focus on patriarchal environ ments in which male and female subjectivities are formed and maintained, in order to present these spaces anew from a nonsexist perspective.

Feminists have extended the theorization of 'class' to include not only the mode of reproduction as well as production but also other forms of labor appropriation – including paid and unpaid, 'productive' and invisible, apparently unproductive, such as caring labor, clerical work, secretarial work, and home based work – while feminist geographers have integrated the vital role of such social relations in place and communities to rethink the very ways in which 'space' is thought about and studied.

In addition to the critique and reworking of existing categories and subdisciplines, feminist geographers have brought entirely new subjects and approaches to the discipline. Thus, for example, a highlighting of sexuality and space led not only to a consideration of gays and lesbians across cities, but also more recently to studies of the active negotiation of heterosexuality as well as di vergent and hybrid sexualities. Feminist geographers have also systematically examined domestic spaces – as sites of work, desire, and design – and also sexualized bodies in space – be they pregnant, engaged in body building, or negotiating taboo practices and spaces.

In addition to new spaces, feminist geographers have constructed entirely new approaches, meshing feminist theorizing with other developments in geography and beyond. Thus, Doreen Massey and Linda McDowell have not only completely rethought the process of gendering work places but Hanson and Pratt have also documented how gender interconnects with and mutually constitutes home, community, and work places, while Gibson Gra ham, in their studies of alternative economic spaces, and Gregson and Crewe, on car boot sales, offer new per spectives that can and should revolutionize economics and economic geography. So, too, studies of colonial and postcolonial encounters are now highlighting the inter connection of gender, race, ethnicity, and space to illus trate how material realities, corporeal differences, and subjective understandings of place, race, and racism are mutually constructed in colonial and postcolonial settings.

Methods

In addition, feminist geography has been part of a sus tained reworking of research methods. Thus, from the first critiques launched by Monk and Hanson to more recent evaluations of positivist and empiricist methods in terms of their faith in objectivity, value neutrality, and universal applications, there has been a concerted feminist critique and reformulation of research methods. Drawing on feminist assumptions as to the inherent value of women's perspectives and on the intersection of the personal and the political, all methods used in the 1970s and 1980s were critically evaluated. In their place went more emphasis on qualitative approaches, which valued women's position, the role of the researcher, and the political nature of the research transaction.

In this exercise, positionality becomes a crucial dimension of any research process, while the knowledge produced was increasingly seen as partial, relative, provisional, and grounded rather than absolute, definitive, and generalizable. Thus, in place of empirical research not only by gendered but also by relatively objective scholars, went knowledge quests by subjects, occupying particular gendered, classed, racialized, sexualized, and also dynamic positions, rendering what was known, limited, and bounded by these dimensions. Doing feminist geography research – as Pam Moss's 2002 book detailed – was therefore not about adopting qualitative over quantitative methods, but about positioning the research question, the researcher, the techniques, and the researched within complex frameworks of power and knowledge.

Qualitative and other feminist methodologies have now been incorporated into the discipline along with collaborative and nonexploitative approaches to research to challenge the unequal power relations which underpin many research practices. But quantitative methods have now also been admitted to the canon if they are deemed useful as choice of method is related to purpose. Thus, research is now conducted 'with' rather than 'on' subjects, with politically positioned academics engaged in conversations and collaborations rather than observations using an array of techniques in a highly reflexive manner.

Politics

As a consequence of highlighting women's absences from the discipline and conducting gender sensitive research, feminist geographers have been engaged in an ongoing political project which extended well beyond the discursive to involve the hiring practices of universities, the professional organization of the discipline, the operation of editorial boards, and ultimately the challenging of inequitable gender relations across a vast number of sites and countries. Feminist geography has always been far more than a discursive intervention, but integral to it has been a progressive and active politics of institutional change.

From the earliest head counts of women in the discipline, there were organized as well as informal efforts to alter universities, geographical curriculum, and the operation of the profession. In addition, feminist geographers have been active in designing better and safer public spaces, lobbying for gender sensitive and better services for women – especially transport and also child care and health services. As a discourse, then, feminist geography has always been highly political and engaged, though this has lessened with the cultural and textual turn. This potential disengagement of feminist geography constitutes one of a number of current issues.

Possible Directions in Feminist Geography

Attempting to distil current priorities in feminist geography (**Table 1**) encounters the problem not only of understanding the present while within it, of gaining a perspective and transnational experience where time and space remain real barriers, but also of surveying a field which might well have died within Australia. But the task is made possible through the existence of *Gender, Place and Culture* (*GPC*), and it is from a reading of is output over the last 14 years that the following speculations derive.

Discerning the future from the present suggests a thickening of the diversity and complexity that has always been present in feminist geography and also a new dimension has entered: namely, a willingness, even a predilection to transgress and easily move across established typological boundaries. Much current work published in *GPC* moves across genres, subjects, and frameworks. Thus, there are studies of the discursive construction of the migratory Filippina entertainers which sit readily within migration, political, feminist, and cultural geography; studies of women's household strategies in rural Appalachia which draws heavily on economic, cultural, and rural geography; or of the paintings of men as landscapes by women.

In addition to these transgressive works, there are those which continue to explore the new terrain opened by feminist geographers, with studies of embodiment, material cultures, and First–Third World intersections of domestic labor, sex tourism, and representations. Such work focuses on the ways gender differences work within and through economic, political, cultural, and sexual differences across international space. Within this now familiar quest goes the need to better interconnect the material, corporeal, and the discursive and develop emancipatory spaces. There is a need and emergent actuality in the pages of *GPC* to revisit the notion of 'oppression' and to reexamine the processes by which black, Islamic, minority, lesbian, working class, etc., women still hold less power, are still significantly disadvantaged, and subjected to gross levels of physical and sexual violence around the world. Such (re ?)examination is emerging from various margins, especially

Table 1 Feminist geographies traditions and trajectories

Feminist geographies	*Mapping women's place*	*Gender relations*	*Multiple/intersecting identities*
Feminist theoretical tradition (beliefs and strategies)	Liberal feminisms Women are potentially equal to men. Equity is possible in existing institutions Remove barriers Lobby and organize	Radical and socialist feminisms Women's oppression by men (patriarchy). Women the focus/source of knowledge. Sisterhood is powerful/ separatism. Built on Marxist theory Prioritizes production, social reproduction, and class. Structural analysis demands structural vs. reformist solutions	Postmodern and postcolonial feminisms Individuals are now discursively constituted and embodied subjects Psychoanalytic and post-structuralist methods Imperialism is a primary constitutive process of countries, people, and discourses Need to deconstruct colonial texts and practices, document colonialism while giving voice to the colonized
Precursors/key theorists	Mary Wollstoncraft J.S. Mill, H. Taylor. Betty Friedan	Kate Millett, Shulamith Firestone, Adrienne Rich, Susan Brownmiller, Susan Griffin, Mary Daly Dolores Hayden Marx, Engels Juliet Mitchell Christine Delphy, Lise Vogel, Heidi Hartmann, Zillah Eisenstein	Michel Foucault Sigmund Freud Jacques Lacan Jacques Derrida Deleuze and Guattari Liz Grosz Judith Butler Elspeth Probyn Iris Marion Young Eve Sarkorvsky Sedgewick Edward Said Bell hooks HomiBhabha Stuart Hall Gayatri Spivak
Key feminist geography texts	Mazey and Lee *Her Space. Her Place.* Seager and Olsen *Women in the World Atlas.*	Hanmer and Maynard *Well Founded Fear* Valentine 'Geography of women's fear', (with Bell and Binnie) 'All hyped up and no place to go' Women and Geography Study Group *Geography and Gender* Massey *Spatial Divisions of Labour*	G Rose *Feminism and Geography* Gibson-Graham *The End of Capitalism as We Knew it* Davidson, Bondi, Smith *Emotional Geographies* McDowell *Capital Culture Gender at Work in the City* Blunt and Rose *Writing Women and Space: Colonial and Postcolonial Geographies* Falah and Nagel *Geographies of Muslim Women*
Key feminist geographers	Zelinsky Mazey Lee Monk Hanson Seager Olson	Gill Valentine Mary Maynard Jalna Hanmer Doreen Massey Suzanne Mackenzie Linda McDowell	Liz Bondi Gillian Rose Katherine Gibson Julie Graham Alison Blunt Jane M. Jacobs Heidi Nast Kay Anderson Brenda Yeoh

India, Indonesia, and Africa and increasingly conducted by feminist geographers working across First and Third World countries.

So while feminist geography may well be in abeyance in Australia – a result of the aging, mellowing, and pro motion of key individuals as well as of very different theoretical and practical priorities by a newer generation of cultural/postcolonial/urban rather than feminist geog raphers – in most other countries, one cannot yet see an end of a gender agenda. Rather as Geraldine Pratt notes in a 2004 essay on dealing with difference in feminist geography, various dimensions of difference are contin gent, while gender remains a fundamental organizing category with a radical feminist politics still connected to ending oppressions wherever they occur.

See also: Critical Geography; Cultural Geography; Economic Geography; Feminist Geography, Prehistory of; Historical Geography; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Quantitative Methodologies; Queer Theory/Queer Geographies; Rural Geography; Social Geography.

Further Reading

Anderson, K. (1990). Chinatown re oriented: A critical analysis of recent redevelopment schemes in a Melbourne and Sydney enclave. *Australian Geographical Studies* 28(2), 137 213.

Blunt, A. and Rose, G. (eds.) (1994). *Writing Women and Space: Colonial and Postcolonial Geographies*. New York: The Guildford Press.

Bondi, L. (1990). Feminism, postmodernism, and geography: Space for women? *Antipode* 22(2), 156 167.

Boserup, E. (1970). *Women's Role in Economic Development*. New York: St Martin's Press.

Breugel, I. (1973). Cities, women and social class: A comment. *Antipode* 5, 62 63.

Burnett, P. (1973). Social change, the status of women and models of city form and development. *Antipode* 5, 57 61.

Castells, M. (1983). *The City and the Grassroots*. London: Arnold.

Deutsche, R. (1991). Boy's town. *Environment and Planning D: Society and Space* 9, 9 30.

Domosh, M. and Seager, J. (2001). *Putting Women in Place: Feminist Geographers Make Sense of the World*. New York: Guildford Press.

Duncan, N. (ed.) (1996). *Bodyspace: Destabilizing Geographies of Gender and Sexuality*. London: Routledge.

Falah, G. W. and Nagel, C. (eds.) (2005). *Geographies of Muslim Women: Gender, Religion and Space*. New York: Guildford.

Foord, J. and Gregson, N. (1986). Patriarchy: Towards a reconceptualisation. *Antipode* 18(2), 186 211.

Garcia Ramon, M. D., Simonsen, K. and Vaiou, D. (2006). Guest editorial: Does Anglophone hegemony permeate. *Gender, Place and Culture* 13(1), 1 5.

Gibson Graham, J. K. (1994). *The End of Capitalism as We Knew It*. Oxford: Blackwell.

Gibson Graham, J. K. (2006). *A Postcapitalist Politics*. Minnesota, MN: University of Minnesota Press.

Graham, J. (1992). Post Fordism as politics: The political consequences of narratives on the left. *Environment and Planning D: Society and Space* 10, 393 410.

Hanmer, J. and Maynard, M. (1984). *Well Founded Fear: A Community Study of Violence to Women*. London: Hutchinson.

Hanson, S. and Pratt, G. (1995). *Women, Work and Space*. London: Routledge.

Harvey, D. (1989). *The Condition of Postmodernity*. Oxford: Blackwell.

Hayden, D. (1981). *The Grand Domestic Revolution. A History of Feminist Designs for American Housing*. Cambridge, MA: MIT Press.

Hayford, A. (1974). The geography of women: An historical introduction. *Antipode* 6(2), 302 305.

Jacobs, J. (1996). *Edge of Empire*. London: Routledge.

Johnson, L., Huggins, J. and Jacobs, J. (2000). *Placebound: Australian Feminist Geographies*. Melbourne, VIC: Oxford University Press.

Kobayashi, A. (2006). Why women of colour in geography? *Gender, Place and Culture* 13(1), 33 38.

Little, J. (2002). *Gender and Rural Geography: Identity, Sexuality and Power in the Countryside*. Harlow: Prentice Hall.

Little, J. and Panelli, R. (2003). Gender research in rural geography. *Gender, Place and Culture* 10(3), 281 290.

Little, J., Peake, L. and Richardson, P. (eds.) (1988). *Women in Cities: Gender and the Urban Environment*. Basingstoke: Macmillan Education.

Longhurst, R. (2001). *Bodies: Exploring Fluid Boundaries*. London: Routledge.

Mackenzie, S. (1980). Women's place women's space. *Area* 12(1), 47 49.

Massey, D. (1984). *Spatial Divisions of Labour: Social Structures and the Geography of Production*. London: Macmillan.

Massey, D. (1991). Flexible sexism. *Environment and Planning D: Society and Space* 9, 31 57.

Mazey, M. and Lee, D. (1983). *Her Space, Her Place*. Washington, DC: Association of American Geographers.

McDowell, L. (1984). Towards an understanding of the gender division of urban space. *Environment and Planning D: Society and Space* 1, 59 72.

McDowell, L. (1991). Lie without father and Ford: The new gender order of post fordism. *Transactions Institute of British Geographers, N.S.* 16, 400 419.

McDowell, L. (1997). *Capital Culture Gender at Work in the City*. Oxford: Basil Blackwell.

McDowell, L. and Massey, D. (1984). A woman's place? In Massey, D. & Allen, J. (eds.) *Geography Matters! A Reader*, pp 128 147. Cambridge: Cambridge University Press.

McDowell, L. and Sharpe, L. (1997). *Space, Gender, Knowledge: Feminist Readings*. London: Arnold.

Momsen, J. and Kincaid, V. (1993). *Different Places, Different Voices: Gender and Development in Africa, Asia and Latin America*. London: Routledge.

Momson, J. and Townsend, J. (1987). *Geography of Gender in the Third World*. Albany, NY: State University of New York.

Monk, J. and Hanson, S. (1982). On not excluding half of the human in human geography. *Professional Geographer* 34, 11 23.

Morin, K. and Berg, L. (1999). Emplacing current trends in feminist historical geography. *Gender, Place and Culture* 6(4), 311 330.

Moss, P. (ed.) (2002). *Feminist Geography in Practice: Research and Methods*. Oxford: Blackwell.

Norwood, V. and Monk, J. (eds.) (1987). *The Desert Is No Lady: Southwestern Landscapes in Women's Writing and Art*. New Haven, CT: Yale University Press.

Pratt, G. (2004). Feminist geographies: Spatialising feminist politics. In Cloke, P., Crang, P. & Goodwin, M. (eds.) *Envisioning Human Geographies*, pp 128 145. London: Arnold.

Rose, G. (1993). *Feminism and Geography*. Minneapolis, MN: University of Minnesota Press.

Seager, J. (2003). *The Penguin Atlas of Women in the World*. New York: Penguin Books.

Seager, J. and Olsen, A. (1982). *Women in the World: An International Atlas*. London: Pan Books.

Soja, E. (1989). *Postmodern Geographies: The Reassertion of Space in Critical Social Theory*. Cambridge, MA: Blackwell.

Tivers, J. (1978). How the other half lives: An historical study of women. *Area* 10(4), 302 306.

Valentine, G. (1989). Geography of women's fear. *Area* 21(4), 385 390.

Wekerle, G., Peterson, R. and Morley, D. (eds.) (1980). *New Space for Women*. Boulder, CO: Westview Press.

Wekerle, G. and Whitzman, C. (1995). *Safe Cities: Guidelines for Planning, Design and Management*. New York: Van Nostrand Reinhold.

Women and Geography Study Group of the IBG (1984). *Geography and Gender*. London: Explorations in Feminism Collective in association with Hutchinson.

Women and Geography Study Group (1997). *Feminist Geographies: Explorations in Diversity and Difference*. Harlow: Addison Wesley Longman.

Zelinsky, W. (1973). The strange case of the missing female geographer. *Professional Geographer* 25(2), 101 105.

Feminist Geography, Prehistory of

S. Bowlby and J. Tivers, University of Reading, Reading, UK, and Nottingham Trent University, Nottingham, UK

Introduction

In Britain, the history of feminist geography began in the late 1970s. This article is about the largely unwritten 'prehistory' of feminist geography in Britain. The first stirrings of concern for the geography of women seem to have taken place in North America in the early 1970s, but did not impact the United Kingdom until the mid 1970s. What did established women geographers feel about the male dominance of the discipline before the mid 1970s? Many of them, like Monica Cole and Alice Coleman, found important roles within the geographical community, and did not question the male bias of its subject content. In the early 1970s, however, a new generation of geographers began their study. They had been brought up in a post war world of growing women's employment, when women's education and academic achievement were becoming accepted among the middle classes. Feminist geography began to develop among this new generation.

We start our story by outlining the state of feminism in Britain in the late 1960s and early 1970s and then discuss the intellectual and social context within which geographers in Britain were working.

Second-Wave Feminism

The so called 'second wave feminism' is generally held to have developed out of the Civil Rights and leftist movements that were active in the 1960s in North America, Europe, and Australasia. This was a time of increased prosperity for adolescents and young adults that encouraged a 'youth culture', evidenced in consumption of new styles of music and dress, which approved opposition to the ideas of the older generation and sought sexual freedom, facilitated by the development of 'the pill'. The Civil Rights Movement in North America and the opposition to the war in Vietnam encouraged ideas about liberation – for oppressed colonial peoples, from racial prejudice and from outmoded social conventions. Many women who engaged in these movements became disillusioned with the treatment they received from male activists – relegated to routine work and organization, while men played the more glamorous activist and leadership parts – summed up in the image of women 'making the tea'. Furthermore, sexual liberation often meant that women activists were expected to provide sexual services for male leaders as part of their delivery of 'sexual liberation'.

It was out of such tensions and concerns that the women's movement began. In North America and Britain, groups of women began to meet to share ideas and experiences and to develop political demands that were expressed through a variety of demonstrations and public statements. At this period, the majority of women activists were white, educated, and middle class: as many feminists have pointed out, this resulted in later claims that the experience and oppressions of non white and working class women had been excluded from the framing of feminist politics and theory. These claims were later to result in intense debate within feminism and major changes in feminist theory and practice. However, from its inception the women's movement was politically disparate – it encompassed a wide variety of views, from liberal rights based demands for equality to radical questioning of women's sexuality and separatist movements. Between 1970 and 1978, the women's movement, nevertheless, articulated seven demands: (1) equal pay; (2) equal education and equal opportunities in work; (3) financial and legal independence; (4) free 24 hour nurseries; (5) free contraception and abortion on demand; (6) a woman's right to define her own sexuality; and (7) freedom from violence and sexual coercion.

Anglo-American Geography in the Late 1960s and Early 1970s

The so called 'quantitative revolution' in geography occurred in the 1960s in Britain and North America. A new concern to develop and use rigorous 'scientific' theory and methodology, adopt positivist methodology, and statistical analysis of quantitative data became widely accepted. Much of the methodological and theoretical inspiration for this approach came from neoclassical economics. As behavioral geography developed, ideas and methods from psychology were also drawn in. This was a discipline which talked about genderless 'individuals' and 'households' whose decision making resulted in the creation of settlement, travel, and economic activity patterns that might be predicted and modeled. Other traditions still continued and alternative ideas and approaches were also in evidence. As Hanson asks us to imagine, however, this was a world in which:

> women's subordinate status is vividly evident through gender differences in educational attainment (years of formal education), employment (job type, hours of

> employment, wages), and political power (presence/ab sence in governmental decision making positions) (…) in which geography journals and university geography curricula are practically devoid of the words "gender" and "women" and in which all instructional materials implicitly, if not explicitly, assume that all salient geo graphical actors are male. Imagine university geography teaching staffs, graduate student cohorts, and national geography conferences composed almost entirely of white men. (Hanson, 2008: 1)

Hanson describes the type of geographical education that both authors of this article received as under graduates and, in the next two sections, we tell our stories of 'discovering feminism' as we now recollect them. In doing so, we have struggled to re create for ourselves the intellectual and social climate of the period as well as details of times, places, and people. As many commen tators on 'narrative approaches' have remarked, such re call in not always accurate and we apologize to others who were there at the time if we have misrepresented or omitted events or people. Drawing on the material of our stories, we will finish with a discussion of how feminist ideas entered geography in the UK.

Sophie's Story

In 1966, I (Bowlby) left Cambridge, eager to train as a postgraduate student at Northwestern University in Chicago. At the time, the Geography Department at Northwestern was in the vanguard of the 'new' quanti tative geography. In my second year as an undergraduate, I had been inspired with enthusiasm for this approach by reading William Bunge's *Theoretical geography* and by the ideas of Peter Haggett and Dick Chorley, lecturers in the Geography Department at Cambridge. And I was not disappointed – courses on math, economics, and stat istics, and discussions of research on Monte Carlo simulations of the spread of innovation and modeling of travel behavior seemed exciting and novel (if difficult!). It never occurred to me to question the validity of this approach or to think that geographers should pay at tention to the differences in the lives of men and women; however, I was becoming aware of some of those differ ences in my own life. Specifically, some members of staff in the Department thought that women 'were not worth taking on as postgrads' or that we were there to 'catch a husband' and some fellow male students resented being outshone by their female peers. I was angry and surprised that my, and other women's, desire to learn and excel academically was not taken seriously (and thought their attitudes old fashioned and ridiculous). Although in my personal life I was adopting feminist principles, I knew nothing of feminist politics or theory and, certainly, it did not occur to me to question the gender blind nature of geographical study – indeed, I had no idea of such a concept.

In retrospect, I see that during my stay at North western I also encountered ideas, politics, and approaches that moved me in the direction of feminist theory. I be came more committed to left wing politics through my reactions to the Chicago race riots, the political up heavals of the Chicago Democratic convention of 1968, and the Kent State shootings in 1970. I was introduced to sociological thinking through a wonderful course run by Janet Abu Lughod. In 1971, I went back to Northwestern, from the UK, to spend a summer writing up my thesis and sharing a house with Susan and Perry Hanson, who were analyzing data for their theses from a diary study of household travel behavior in Uppsala, Sweden. This allowed them to analyze the differences in the travel behavior of employed men and women – one of the first geographical analyses that explored the implications of the gender division of labor.

Back in Britain as a lecturer at Reading University in the early 1970s, I was asked to lecture on Social Geog raphy and began to read literature in urban sociology, as well as the new ideas about Marxist geography. A reading group of social scientists at Reading was exploring Marxist literature and I quickly became a convert to this (to me) new type of approach. At the same time, my experience in job hunting and at conferences had made me more aware than before of the prejudices and pitfalls facing woman academics: at one interview I was asked if I could guar antee not to have a baby before I was 35. I was fortunate in ending up in one of the rare Departments that welcomed women, and where other women lecturers were soon ap pointed. I was also introduced further to feminist political ideas by John Silk who started lecturing in the Department the year before me, and by his wife Cathy Silk. I got a grant, in 1975, from the Social Science Research Council (SSRC, a government funded public body), to extend the work I had done for my thesis on knowledge of, and access to, shopping opportunities. In this research, I surveyed each sample household's 'principal grocery shopper' most of whom were, of course, women. My thesis had stemmed from an interest in mental maps and the ways in which everyday travel in the city might lead to the development of knowledge of shops and leisure opportunities. By the mid 1970s, I was beginning to re frame the problem in terms of inequality and class differences but it was only through doing this study that I became increasingly con cerned with the fact that it was 'women' who were doing the shopping, and began to think about the implications of inequalities in domestic labor. These ideas were given further impetus when, in 1975, Laurie Pickup came to the Department to do a thesis on women's travel behavior and employment, supervised by Peter Hall and myself. His exploration of the then scanty literature on women's travel to work broadened my appreciation of feminist issues in geography.

It was not until I met other people working with feminist ideas at successive Institute of British Geographers (IBG) conferences, from 1977 to 1979, that I became properly aware of the wealth of feminist work that had already been done in other disciplines and the geographical work that was underway in the United States. Some of the women I met had been actively involved in feminist politics and were better read than I in feminist analysis – notably Jo Foord, Jane Lewis, Suzanne Mackenzie, Linda McDowell, and Jacky Tivers. Many of them, like Jacky, were already working on aspects of women's lives. It was wonderful to meet people of like mind – interested in similar ideas and political concerns, who understood the problems of being a feminist woman in a male dominated profession. We were strongly influenced by socialist feminist ideas and excited by the prospect of making a difference to geography and to the place of women academics in the discipline. These encounters led to my engagement with the Woman and Geography Study Group. This early period laid the foundations for developing my research beyond the initial focus on gender alone, to engage with issues of age, disability, and race. The development of this later research was also aided by involvement with ReadiBus (a bus service for those who cannot use ordinary buses) and Berkshire Women's Aid; by interaction with other feminist postgraduates at Reading (especially Jo Little, Linda Peake, and Pat Richardson); and by supervision of postgraduates who have studied a range of issues, from child care to binge drinking by young women, and extended my knowledge of feminist approaches and issues (including Gill Valentine, Ruth Butler, Fiona Smith, Lisa Doyle and, at present, Melissa Stepney and Cheryl Nosworthy).

Jacky's Story

In early January 1977, I (Tivers) attended my first Annual Conference of the IBG, held that year in Newcastle. At this stage, I was one term into my PhD studies at Kings College, London (funded by SSRC). Only a couple of months earlier, my supervisor, Barrie Morgan, had urged me to focus on a specific topic for research from among the broad field of leisure studies, which was my main interest within geography, and had suggested that I look at the particular interests and experiences of a population subgroup. Since, at the time, I was in my mid thirties with an 18 month old son, he had proposed that I should study the leisure experiences of women with young children. It is to Barrie that I owe the original idea for the research project, which grew into the first completed thesis in feminist geography in the UK. By the time I got to the IBG conference at Newcastle, the topic had broadened into a study of the weekday activity patterns of women with young children and I was reading whatever I could find which was remotely relevant. I 'discovered' the work of Ann Oakley, the British sociologist, and this opened my eyes to feminist issues for the first time, but I began to realize, with a sinking heart, that British geography had so far failed to grapple with even the simple differences between men's and women's lives, let alone the underlying reasons for these differences.

So, it was with some despondency about my chosen topic that I found myself among only about 40 women delegates (out of about 600 delegates in total) at Newcastle IBG. In the queue to register I met a young woman lecturer, Bronwen Walter (now a Professor and well known expert on Irish immigration into the UK), who fell on me with undisguised enthusiasm as another woman in the male dominated place. Of greater immediate significance to my research project was my meeting with three geographers from Reading University – Sophie Bowlby, John Silk, and Laurie Pickup. Each, in their own way, encouraged and challenged me. With Sophie and John, I talked about the idea of studying women, specifically, within the discipline of human geography. Laurie was a postgraduate student, like myself, and was researching mobility and transport issues for women. I was delighted to find that someone else (albeit a man!) was interested in women's lives, from a geographical point of view. At the end of the conference I was warmly invited to go to Reading for further discussions at any time I would like to do so.

The next few months brought further advances. I uncovered some American literature, relating to the 'geography of women'. This mainly took the form of seminar notes and unpublished departmental papers, not easy to track down in the pre Internet era. I met Doreen Massey for the first time; she was lecturing in London, shortly before her move to the Open University. I contacted her and requested a discussion with her. With characteristic generosity, she took me to a café and bought me lunch, and was incredibly supportive and encouraging, urging me to pursue my line of thinking and stressing its importance to geography. My field work (a large scale questionnaire survey of 400 women with young children) was conducted in September 1977 by 14 interviewers, including myself. The geography 'fraternity' heard about my research, probably through Barrie Morgan, my supervisor, who was a committee member of the Urban Geography Study Group (UGSG) at the time, and I was approached to present a paper at the following year's Annual Conference, in a session organized by the UGSG. In preparation for this, I was asked to write a departmental paper, and present a seminar, at Kings College. The atmosphere within the department was supportive and professional, and did not at all prepare me for the onslaught of negative opinion which I was to face at Hull in January 1978.

I still have the conference handbook, and it records that on the morning of Thursday, 5 January 1978, I gave a paper entitled 'Constraints on spatial activity patterns: the case of women' within the UGSG session 'Socio spatial aspects of urban life'. This, I believe, was the first paper ever presented from a feminist perspective at a British Geography conference, although I was not really aware of making history at the time. I did call for the explicit consideration of women in geographical studies, but did not feel that the argument was anything other than 'perfectly straightforward'. Despite the presence at the conference of people like Mayer Hillman and Anne Whalley, Peter Jones, John Raine, and Malcolm Moseley – all thoughtful academics who were researching household activity patterns and the transport needs of individual household members – with whom I had fruitful dis cussions, continuing long after the conference ended, my overwhelming experience of the conference was of being patronized, marginalized, and made to feel that my re search lacked validity and importance. The term 'gender' was not even understood by many people, as became apparent later.

I returned from Hull in a bad mood, stormed into the postgraduate room at Kings College on the following Monday morning, and sat all day writing a polemical article for the journal, *Area*, before my annoyance and frustration had time to dissipate. Without discussing or revising it, I sent it off straight away to Jeremy White hand, the editor. To my shame, I assumed that he would reject it, and that would be another black mark against male dominated geography. But, within a very short time, he wrote back, accepting the paper for publication. It was in print within 6 months of submission – the first paper in a British geography journal to talk specifically about women. I am not particularly proud of the content – it is all talk and few facts, the editor had changed all the references to 'gender' into 'sex', without consulting me, and I felt too insecure, as a humble postgraduate, to alter the page proofs when they were sent to me. Neverthe less, it was a start. I had laid down a challenge:

> By accepting the household as the predominant unit of study in human geography and by neglecting the explicit study of the geography of women, geographers have implicitly accepted the existing structure of society and prevented the examination of one of the most persistent and pervasive forms of inequality in our society (Tivers, 1978a, 305).

At the 1978 Hull IBG conference, the British collective of the Union of Socialist Geographers (USG) was formed, the inaugural meeting taking place (for some reason, which I have now forgotten) in my bedroom. In the spring of 1978, I received my first copy of the USG newsletter, containing an interesting review article by Bonnie Loyd, entitled 'The quiet revolution in geography'. This opened my eyes to research which had already taken place in the United States on the 'geography of women' and to the number of dedicated courses which were already being taught there. Loyd talked about the 'quantity of studies' and a 'steady stream of research', but I felt that the focus of these studies was not on the real constraint which underlay women's lives; and so I wrote 'Yet another view on the geography of women', which was published in the USG Newsletter at the end of the year. In this paper, I argued that geography had been too concerned with production, at the expense of reproduction, and that it was necessary to focus on the 'gender role constraint', which was both supportive of capitalism and the frame for women's lives.

By the end of 1978, I was a committee member of the IBG Social Geography Study Group, and had met both Linda McDowell and Suzanne Mackenzie. Suzanne ar rived from Canada to study at Sussex University, bringing with her a wealth of experience in feminist geography from North America. Together with Sophie Bowlby, we organized the first dedicated session about women and geography at an IBG conference, which took place in January 1979 at Manchester University. It was not an unqualified success; heavy snowfall caused a lot of problems and our principal speaker, Pat Burnett from the US, was unable to attend. Suzanne and I gave papers, hers being a review of North American work in the field, and Sophie read Pat's paper. As I have recorded elsewhere, 'the debate tended to descend into acrimony, with tight corners being fought by the Marxists who saw feminism as a distraction from the 'true cause' of socialism'. This was quite ironic, considering that feminist geography in the UK had been born, primarily, out of Marxist geography.

Nevertheless, we were undeterred, and two further articles appeared in *Area* in 1979 and 1980, the first being Linda McDowell's analysis of women in British geog raphy and the second, a joint article by Suzanne, Linda, and Jo Foord, which sought to extend and perpetuate the discussion on women and geography. At the 1980 Annual Conference of IBG, the Women and Geography Working Party was set up, with Linda McDowell acting as the guiding force. After some delay, during which the group was required to demonstrate the level of support for its activities, the Women and Geography Study Group (WGSG) was inaugurated in 1982. One of its first actions was to enter publishing, since there was a complete lack of material in the UK suitable for teaching feminist geography. It was decided that a textbook would be written collectively; Linda McDowell, Sophie Bowlby, and I were appointed editors. The joint editing was something of a nightmare, in the era before e mail, but the book (*Geography and gender: An introduction to feminist geography*) was very well received and has become a 'classic' in human geography. With this book, the 'pre history' of feminist geography could be said to have ended, as feminism then began to be part of the 'history' of the discipline; increasing numbers of papers and books were published, and increasing numbers of conference sessions organized, from a feminist perspective, mainly through the auspices of the WGSG, which has gone from strength to strength in the last 25 years. As I have re corded elsewhere, 'I am proud to have been a part of the beginnings of feminist geography, and am proud too of the movement which has grown and expanded beyond the wildest dreams of the early days'.

Concluding Thoughts

Clearly, our stories provide only two glimpses of the processes that led to the early development of feminist geography. The other people we have mentioned might tell it in very different ways. But our stories demonstrate that the development of early feminist geography was not the heroic creation of a few pioneers. Rather, in many places during the 1970s, women and some men in geography began to explore the geographies of women, and in time, this led to the examination of a range of explanations for the subordination of women and to the development of more sophisticated feminist geographies. Those of us working in or entering the geographical discipline, in the 1970s, were in a social and economic environment in which ideas of equal opportunities for women, of sexual equality, autonomy, and choice were widely debated and relevant to young educated women entering the academic profession. We addressed issues drawn from our own experience and raised new questions linked to the new ways of life that women were adopting. Although some came to these questions after involve ment in feminist politics, many did not and only began to explore the wealth of feminist literature outside geog raphy after initiating their research.

Despite the contribution of some men, this history also suggests how important it was for the creation of a feminist geography that women were entering the dis cipline. It was women's interest in and experience of aspects of everyday life affecting women that led to the rapid expansion of the early geography of women and the subsequent development of feminist geographies. This reminds us of the importance of striving to ensure that the discipline is open to disadvantaged and minority groups. For example, the developments of geographies of sexuality and of disability have been led largely by people who are bisexual, gay, lesbian, or disabled. It is vital that we work to ensure openness at a time when the pressures of the research assessment exercise (RAE) and work intensification in universities may limit the access of such groups and the development of challenges to contemporary ideas.

See also: Feminism and Work; Feminism, Maps and GIS; Feminist Groups within Geography; Feminist Methodologies; Feminist Political Economy; Possibilism.

Further Reading

Bowlby, S. R. (1979). Accessibility, mobility and grocery shopping. In Goodall, B. & Kirby, A. (eds.) *Resources and Planning*, pp 293 323. Oxford: Pergamon Press.
Bowlby, S. R. (1984). Planning for women to shop in postwar Britain. *Environment and Planning D: Society and Space* 2, 179 199.
Bunge, W. (1966). *Theoretical geography*. Lund: C W K Gleerup.
Foord, J., Mackenzie, S. and McDowell, L. (1980). Women's place, women's space. *Area* 12(1), 47 49.
Haggett, P. (1965). *Locational Analysis in Human Geography*. London: Edward Arnold.
Harvey, D. (1973). *Social Justice and the City*. London: Edward Arnold.
Hanson, S. (2008). Geography and gender: 1984 women and geography study group. In Hubbard, P., Kitchin, R. & Valentine, G. (eds.) *Key Texts in Human Geography*. pp 91 980. London: Sage.
Hanson, S. and Hanson, P. (1974). *Daily Activity Patterns of Working Women and Working Men: Are They Different*? Clarke University: Mimeo.
Lewis, J. (1984). *Postwar Regional Development in Britain: The Role of Women in the Labour Market*. PhD Thesis, University of London.
Lewis, J. and Foord, J. (1984). New towns and new gender relations in old industrial regions: Women's employment in Peterlee and east Kilbride. *Built Environment* 10, 42 51.
Loyd, B. (1978). The quiet revolution in geography. *Union of Socialist Geographers Newsletter* 3(3), 30 32.
McDowell, L. (1979). Women in British geography. *Area* 11(2), 151 154.
McDowell, L. (1983). Towards an understanding of the gender division of urban space. *Environment and Planning D Society and Space* 1, 59 72.
Pickup, L. (1983). *Travel Issues in Women's Job Choice: An Activity Based Approach*. PhD Thesis, University of Reading.
Relph, E. (1970). An enquiry into the relations between phenomenology and geography. *Canadian Geographer* 14, 193 201.
Tivers, J. (1977). *Occasional Paper No. 6: Constraints on Spatial Activity Patterns: Women with Young Children*. London: Kings College, Geography Department.
Tivers, J. (1978). How the other half lives: The geographical study of women. *Area* 10(4), 302 306.
Tivers, J. (1978). Yet another view on the geography of women. *Union of Socialist Geographers Newsletter* 4(2), 25 26.
Tivers, J. (1982). *Weekday Spatial Activity Patterns of Women with Young Children*. PhD Thesis, University of London.
Tivers, J. (1997). In Women and Geography Study Group of the Royal Geographical Society (ed.) *Feminist Geographies*: *Explorations in Diversity and Difference*, pp 36 38. Harlow, UK: Longman.
Tuan, Y. F. (1971). Geography, Phenomenology and the study of human nature. *Canadian Geographer* 15, 181 192.
Wilson, E. (1986). *Hidden Agendas*. London: Tavistock.
Women and Geography Study Group of the IBG (1984). *Geography and Gender: An Introduction to Feminist Geography*. London: Hutchinson.
Women and Geography Study Group of the IBG (1997). *Feminist Geographies: Explorations in Diversity and Difference*. Harlow: Longman.

Feminist Groups within Geography

K. M. Morin, Bucknell University, Lewisburg, PA, USA

Glossary

Feminism It has many different meanings across space and through time; therefore, we can speak of many feminisms, each with its own history, geography, and central problematic. 'Western' feminism typically includes: liberal feminism, which focuses on equal rights for women, especially in the paid labor force; standpoint or radical feminism, which argues for a challenge to patriarchy by valorizing feminine characteristics that have been relegated to inferior status within masculinist social norms; and feminisms that focus on difference, which take an explicit interest in the processes by which our identities as gendered individuals are produced.

Feminist Group For purposes of this article, refers to an organization, society, or other group of academic geographers whose research and work advances feminist and/or gender geography, or women's status in the discipline of geography.

Gender Generally refers to the socially constructed characteristics of femininity and masculinity, or what might be called the social organization of sexual difference. In some societies, no differentiation is made between the terms 'sex' and 'gender'. In other contexts, gender is understood to be in contrast with sex, with gender referring to socially defined gender behaviors, characteristics, and identities, and sex referring to the natural biological differences between women and men. Judith Butler has argued, however, that the notion of two sexes is itself a product of the gendering process – in her words, the 'regulatory fiction' of two genders – and therefore, that the gendering of bodies themselves as masculine or feminine must be open to analysis. Moreover, because gender is understood to intersect with other axes of social difference (race/ethnicity, class, age, sexuality, geography, and so on), we can say that numerous gender(s) (masculinities and femininities) exist, not simply two, and that there are dominant, hegemonic ones, and subordinate ones.

Introduction

In taking inventory of the organizations and societies through which information about feminist and gender geography, or women's status in the discipline of geography, is discussed, supported, and disseminated, one immediately confronts the many layers of social organization that limit or make possible a feminist 'group' in geography. The larger political context of feminism and the question of what constitutes an important or worthwhile women's or gender issue varies considerably from place to place, as does the type of geography practiced in universities and research centers around the world. Shifts in gender politics are fundamentally relevant to the kind of geography practiced, and this, in turn, to the types of networks and communities of geographers that are created, and for what purposes.

The 'waves' or stages of women's liberation movements often align with the purposes and goals of feminist groups in geography. Today, we see simultaneous efforts worldwide to both improve women's visibility and status as geographers, informed by the tenets of the mid twentieth century 'second wave' women's liberation movement in the West (the struggle for equal rights for women), as well as promotion of research on women and gender that is informed by what some call third and fourth wave feminisms. The latter are underlain with new understandings about how gender and other social identities and differences are produced in the first instance, and with social justice concerns that reach beyond 'women's lib'.

Moreover, the capacity for like minded feminist geographers to group themselves together and share information via technology, conferences, publications, and so on is deeply embedded within uneven economic development around the world. In addition, there are networks of significance around the world whose work contributes to feminist geography though they are not necessarily comprised of geographers, for example, in sub Saharan Africa. Even within North America and Europe where formal organizational structures for feminist geography are well developed, feminist geographers group themselves and network informally in countless ways.

Within this context it should be noted at the outset that for practical reasons, this article does not include much discussion of the many feminist NGOs, informal networks, and other groups of feminists whose work contributes to feminist geography but who are not affiliated with a geographical organization or society. Let it also be acknowledged that Anglo American hegemony in the feminist geography academy, English as the *lingua franca* in publication outlets, the question of what constitutes useful feminist theory for challenging material inequalities, and so on are some of the many issues that confront feminist geographers as they attempt to build a transnational feminist praxis through feminist geography groups. Some feminist groups, such as the IGU

Commission on Gender and Geography, frequently bring such transnational issues to the forefront of their work, whereas these issues receive little play in others.

The following outlines feminist geography groups active on the contemporary landscape. These often mirror places where gender and geography courses are taught at universities and/or where feminist geography has become a relatively mainstream branch of geo graphical research and scholarship. Active feminist geography groups can thus be noted in the US, UK, and Canada; in Spain and many Western European countries such as the Netherlands, Germany, Austria, and Switz erland; and in Japan. Feminist geographers in Southeast Asia, Eastern Europe, parts of Africa, Israel, and Brazil, among others, tend to work with like minded individuals outside of geography's more mainstream organizations and structures.

International Groups

IGU Commission on Gender and Geography

The Commission on Gender and Geography of the International Geographical Union (IGU or Union Géo graphique Internationale), based in Brussels, Belgium, represents the main, if not the only, feminist geography group whose mission and purpose is international in scope. The IGU itself, designed to initiate and foster international cooperation in geographical study and re search was formally established in 1922, though Con gresses had been meeting since 1871. Of its numerous commissions and study groups, the IGU's Commission on Gender and Geography is a group of 350–400 geog raphers from approximately 60 countries in all parts of the world. The Commission organizes several meetings each year and maintains its intellectual network through academic publications, newsletters, and its website.

One of the main ways the Commission sustains international networks and collaboration among feminist geographers is through the hosting of workshops and paper sessions at IGU Congresses, as well as symposia outside the Congresses. Gender Commission partici pation in IGU Congresses began in Sydney, Australia in 1988, and since then have had a presence in Beijing (1990), Washington, DC (1992), Prague (1994), Havana (1995), Den Haag, the Netherlands (1996), Lisbon (1998), Seoul (2000), Durban, South Africa (2002), Glasgow (2004), and Tunis (2008). The 2004 Congress in Glasgow featured daily sessions organized on gender themes.

Since 1989, the Commission has also co sponsored symposia outside of the IGU Congress sites in combi nation with other commissions and study groups. These include a symposium on 'Women and work' in Sao Paolo, Brazil, in 1991; one on 'Gender and environment' in Legon, Ghana, in 1995; 'Women in the Asia Pacific region' in Singapore, in 1997; 'Beyond tolerance: social and spatial coexistence' in 2002, in Tel Aviv; and in Rome, in 2003, a symposium on the theme of 'Gendered cities'. Scores of other programs, workshops, and sym posia have been held in other places in other years. For the 2004–08 period, the central theme of the Commission was 'Crossing borders', with events held in Zürich, Switzerland; Taipei, Taiwan; and Tunis, Tunisia. The 'crossing borders' theme focused on international mi gration and transnationalism on a global scale, among other issues.

The Commission also supports international feminist geography study and research through a number of publications, including its 'Working papers series' (35 have been published to date), and newsletter. The Commission's newsletter is an important venue for international communication – it reports individual and group 'news from around the world', lists important publications on topics in feminist and gender geography, announces programs of study and conferences and events throughout the world. Janice Monk compiles this news letter, and has done so twice a year since 1988. Janice Monk is one geographer who has made invaluable con tributions connecting feminist groups worldwide, par ticularly through her work in the IGU and as editor of the Gender Commission's newsletter.

The Commission's steering committee is meant to include representatives from all major regions of the world, in order to foster worldwide communication and participation in geographical study and research. Monk (2007) outlines the history and current networks that have developed within the IGU Gender Commission, noting its success in making connections internationally, though questioning how and to what extent the Com mission actually fosters 'worldwide' communication and participation. Heaviest participation is from Europe (es pecially Spain and Italy), Anglophone North America, and the Asia Pacific, with major absences of participants from Eastern Europe, East Asia (except Japan), Africa (except South Africa), and Latin America (except Argentina). Monk emphasizes the significance of tech nology and geography's infrastructure in any location in shaping international communication. That noted, the Commission's website highlights the value of bringing together scholars from different parts of the world. The Canadian scholar Linda Peake notes that:

> [m]eeting new colleagues from all parts of the world made me too acutely aware of the stereotypical ways in which I was still thinking. It allowed me the opportunity to listen, first hand, to what non Western geographers thought of North American and Anglo European geog raphy: that it was preoccupied with theory, but more importantly, theory built upon concepts that often had little applicability to their everyday lives.

Judit Timár, from Hungary, likewise noted that:

> [d]iscussions that showed us that some ideas worked out in advanced capitalist countries cannot be applied in other regions spurred us to rethink some terms and concepts [such as production reproduction and masculine city feminine suburb].

Society of Women Geographers

The US based Society of Women Geographers (SWG) is a professional and social organization established in 1925 by women who were accomplished explorers (Marguerite Harrison, Blair Niles, Gertrude Shelby, and Gertrude Emerson Sen). At that time, no organization existed to encourage women in geographical exploration and research and to share women's achievements and knowledge. The Society interprets 'geographer' broadly, and has grown to include women from allied disciplines such as anthropology, geology, oceanography, and ecology: its website declares that these are women who 'love adventure … and who meet to share discoveries and adventures'. About 20% of the SWG memberships are academics. Janice Monk has written on a number of women who worked at the New York based American Geographical Society in the mid twentieth century and who held SWG offices, delivered lectures, encouraged and supported women's movement into geography as a career, and advocated for women's research to be included as geography.

The Society's organized groups are located in New York, Washington, DC, Chicago, Los Angeles, San Francisco, and Florida. In addition, members at large are distributed throughout the US and corresponding members live overseas. Meetings, held regularly during the academic year in the larger groups, provide members a forum for presenting the results of their travels or work, and a social atmosphere in which to exchange information informally. The entire Society meets triennially. The Society's records, archives, and library are maintained at its international headquarters in Washington, DC. The headquarters also houses a museum displaying the work of Society members, and bestows a number of awards for women who have made distinguished contributions to knowledge about the world.

National Groups

Subsequent discussion organizes feminist geography groups at the national scale, and focuses primarily on groups that are affiliated with larger geographical organizations.

United Kingdom

One of the most active feminist geography groups is the Women and Geography Study Group (WGSG), organized in the 1970s as an informal group and then as a 'working party' within the Institute of British Geographers (IBG), the then professional organization of academic geographers in the UK. The 'working party' became a formally recognized 'study group' within the IBG in 1980s, and eventually a 'study group' of the Royal Geographical Society (RGS)/IBG when the RGS and IBG merged. The group's aims are to encourage the study of the geographical implications of gender differentiation in society and geographical research from a feminist perspective; and to encourage and facilitate the exchange of information and ideas with reference to research and teaching in these areas. (An earlier mission of the group was to work to improve the position of women within the discipline of geography, but that was omitted from the group's stated mission in the late 1990s.)

The WGSG published their first book in 1984, *Geography and Gender: An Introduction to Feminist Geography*, and both this one and the later one, *Feminist Geographies: Explorations in Diversity and Difference* (WGSG 1997) were published under the name of the group rather than individual authors, with royalties paid to the group. This was an explicit attempt to foreground the feminist value of cooperation. In 2004, the WGSG also published a CD ROM, *Geography and Gender Reconsidered*. The group publishes a newsletter, maintains a website, and sponsors sessions at the annual RGS/IBG conference. It also organizes occasional study weekends. Among other such groups in the UK is a feminist geography reading group at Edinburgh University, similar to the informal reading groups or research clusters that have evolved at many institutions worldwide.

United States

Two feminist geography groups are affiliated with the Association of American Geographers (AAG). The Committee on the Status of Women in Geography monitors and attempts to improve and enhance the status of women in the profession, through programs such as career counseling and mentoring. The Geographical Perspectives on Women (GPOW) specialty group of the AAG is one of the largest such groups of the AAG, with over 500 members in 2007. The purpose of the group is to promote geographic research and education on topics relating to women and gender, investigate ways in which gender is geographically constituted, and understand differences between lives of men and women in different places. The group has an active listserv for communication among members, Geogfem, although as with most such lists, one need not be a member to join. GPOW publishes a newsletter, maintains a website (which links to a bibliography of feminist geography scholarship), sponsors gender and feminist geography related sessions at the annual AAG meetings, bestows a number of awards, and hosts receptions for feminist geographers, mostly honoring feminist geography books published during the year. GPOW has recently added an 'international liaison' pos ition to its board of directors in an attempt to make better connections between AAG members and feminist geog raphers internationally.

A number of more loosely affiliated groups support feminist geography in the US. The National Council for Geographic Education (NCGE) has a women's special interest group that provides awards for research in geography education by women graduate students. Among other initiatives, this group helped publish the Historical Calendars of Women Geographers in 2005 and 2006. The regional women's group of the Association of Pacific Coast geographers also provides travel grants and prizes to women students, though not solely for feminist work. Also, a number of 'supporting women in geog raphy' (SWIG) chapters exist across the USA, typically associated with individual universities (such as the ori ginal group at University of North Carolina at Chapel Hill, the University of California Los Angeles, and Pennsylvania State University). A Canadian chapter is also active at the University of British Columbia in Vancouver. SWIGs are loose associations of students, faculty, alumni, and community members who seek to promote the participation of women in academic or professional geography, via training, career development, resources networks, and through personal support. These goals are pursued through a diverse range of strategies, and operating norms vary significantly from group to group. Attempts are underway at the University of North Carolina Chapel Hill to create a Global SWIG network, via a listserv called 'swig global', designed to keep people in contact with one another and facilitate the planning of joint events.

Canada

Like GPOW, the Canadian Women and Geography Study Group (CWAG), of the Canadian Association of Geographers (CAG), sponsors conference sessions, social events, awards, a newsletter, a listserv (cwag@geog.ubc.ca), and a website. As a group, the CWAG also ad ministers the Suzanne Mackenzie lecture, whereby a feminist scholar is invited every 2 years to present the plenary lecture at the annual CAG conference.

The stated aims of CWAG are to provide a forum for people interested in issues of access, equity, and the well being of women in geography; and for those whose re search is linked to questions of gender as they relate to class, race/ethnicity, sexuality, nation, and physical abil ity. As such, the purpose of this group has vacillated between a study group for feminist research, and a sup port and advocacy group for women geographers. In 2004, the group surveyed its members on 'who we are and where we are going'. Respondents noted that members are united by shared research interests, common strug gles (such as achieving tenure), and connections to aca demia. Among other things, students noted a greater desire to participate face to face at CWAG functions, but are limited in financial resources to do so. Travel grants, low cost opportunities to connect with other members, and informational resources were sought most. Faculty responded that they maintain standing in the group 'to keep abreast of new research, keep in touch with others in the profession, and to know who is working on what'. Connections sought by both faculty and students were noted to be personal as well as professional. Through the survey, the group also discovered that 'female geog raphers' feel less at home in the organization than do feminist geographers, which the former hoped could be resolved through the hosting of additional informal social events.

Antipodes

New Zealand/Aoteorea feminist geographers have been at the forefront of postcolonial feminist work in Anglo phone geography, reflecting the presence of issues about race, colonialism, identity, and power in New Zealand. Feminist geographers working especially in the depart ment of geography at the University of Waikato in Hamilton, Zealand, have been key to this development. In addition to their extensive contributions in feminist research, the New Zealand feminist geographers have published a number of articles about feminist geography in their country. Without the large membership base of an organization such as the AAG or CAG, feminist geography groups in New Zealand and Australia today function on a more informal rather than formal basis. There are many active feminist geographers in both places, and they informally 'group' themselves pro fessionally and socially.

Western Europe

Spain has a long standing feminist geography group lo cated at the Autonomous University of Barcelona, under the leadership of Maria Dolors Garcia Ramon. The group has an official designation as a formal research group within the university, and as such has a different organizational structure than those in the Anglophone world. The group has taken an active leadership role in feminist geography worldwide, organized an array of conferences, and published works of feminist geographers from Spain and elsewhere. The Barcelona group's first conference was held in 1986, a regional meeting of the IGU, which was followed by many subsequent events including an international conference on feminist geog raphy in February 2006. Publication of the conference papers appears with Belgeo (2007).

Information on the active, German language feminist geography group linking Germany, Switzerland, and Austria can be found in the digital newsletter Feministiches Geo Rundmail, started in Bern in 1988 and published four times per year. The newsletter offers calls for papers, conference and event announcements, information on degree programs, lists of publications, relevant links, and discussions of topics of interest (such as the status of gender studies at various universities). Gender and geography study groups are active in all three countries, and several articles have appeared (in English) on German speaking feminist geography and feminist geography groups.

Bäschlin (2002) has written on the history of German feminist geography groups. There were no central organizing groups until 1987 when students from Frankfurt organized a feminist geography workshop, the result of which was an official 'Working Circle on Feminist Geography'. The Circle then held its first conference in 1989 in Saarbrucken, articulating its desire for a 'geographical women's movement'. It has met at every German Geography Day since 1989. Women students formed their own informal group to discuss their situation, holding meetings every 6 months, and without any formal institutional support. According to Bäschlin, research projects of German feminist geographers include topics on spatial structures from which results dominance or exclusion of social groups; questions about the definition of 'labor' and the gendered division of labor; and social constructions of 'nature' and 'culture'.

Sybille Bauriedl outlines the debate about feminist geography in Germany, Switzerland, and Austria. She notes that many of the younger German students studied in British or Scottish universities, attended the IBG conferences, and participated in its WGSG. As a result, she observes, "there was a direct transfer of the British debates into German speaking feminist geography." Eventually, several subnetworks emerged among students, for example, a post structural feminist geography group, and at their initiative, reading weekends with Doreen Massey emerged between 1999 and 2004.

Finally, a network of feminist geographers funded by the European Union for graduate teaching – known as Erasmus – brought an array of universities together for 9 years, 1990–98. Coordinated by Joos Droogleever Fortuijn, the network comprised of individuals from six universities in five countries (Britain, Spain, Denmark, the Netherlands, and Greece). They organized intensive courses on gender and geography each year, in Durham (1991), Barcelona (1993), Athens (1994), Roskilde (1995), Durham (1996), Amsterdam (1997), and Barcelona (1998). Fortuijn (2002) notes the problems the group encountered, however, when the majority of the materials available for the courses were from an Anglo American perspective.

In the Netherlands, other kinds of organizations sponsor feminist geography work, such as the Emancipation Section of the national professional organization of planners – NIROV. This was an active group of female academics and professionals in geography, planning, and architecture who organized seminars and workshops, and published books and brochures, on gender issues in planning and housing.

Eastern Europe

Timár and Jelenszkyné (2004) document the small number of feminist geographers working in Hungary today (perhaps only two or three), and the absence of any formalized feminist geography groups there. One journal issue, approximately 10 articles, and one textbook chapter have been produced by feminist geographers in Hungary to date. They argue that this situation is similar to that of other Eastern European countries.

One shared feature of postsocialist countries and the Commonwealth of Independent States is the strength of an applied, empirical type of economic geography, with little value attached to gender and feminist geography. One of the explanations for this is the pattern of feminism itself in Eastern Europe. Within the socialist countries, the gender equality 'declared' by the state made it difficult for feminism to flourish. Moreover, when women's and gender studies did appear in Eastern and Central Europe in the 1980s, the state sponsored version of Marxism was of little help as a critical tool. Today, feminist geography in Hungary has less interest in the body and gendered identities more typical of the West, than in the serious gender inequalities produced by the social relations of class politics produced especially during the transition from socialism to capitalism.

Voiculescu and Lelea (2003) discuss the situation of feminist geography in Romania. Here again, the authors argue that it was not until after 1989 that scholars began to acknowledge that women's and gender problems had not been solved by communism. Today, there are a number of researches published, and at least one course on gender and geography (at West University, Timisoara), but little in the way of formalized feminist geography groups.

Africa

Awumbila (2007) outlines feminist geography in Anglophone African countries, and argues that gender equity has been acknowledged, since the early 1990s, as key to sustainable development. Higher education courses in women, gender, and place have increased since that time, as has networking among feminist scholars and activists. In sub Saharan Africa, feminist geography and gender and geography courses primarily appear in the Anglophone countries of Nigeria, Ghana, South Africa, Kenya, and Uganda. In Kenya, there are no organized feminist groups in geography; most feminist geographers are members of other feminist groups such as the Association of African Women Researchers.

Monk (2007) overviews the importance of the IGU Commission on Gender and Geography, for developing feminist geography in Africa. The Commission held a training workshop in Ghana in 1995, for example, and two IGU Gender and Geography meetings have been held in South Africa – in Pretoria in 1995 and in Durban in 2002. There appear to be no explicitly feminist geography groups in South Africa, although many aca demics focus their research on feminist and gender geography. Other affiliated groups work on similar problems and issues, for example, the International La bour Research Group is engaged in and with feminist groups locally and regionally.

Israel

Blumen and Bar Gal (2006) discuss the situation of feminist geography and feminist geography groups at meetings of the Israeli Geographical Society (IGS). The first gender and geography session at an IGS meeting was in 1992; yet, today only about three individuals identify themselves with gender and feminist research, much of this is related to planning. The authors assert that the conference itself is an important node in the potential for feminist geography groups to take shape – 'the confer ence serves as an important informal meeting place of real people' – though the gender composition of the speakers is uneven, women speakers have significantly lower prestige than men, and gender segregation of knowledge in the various subfields of the discipline is apparent. Due to the common perception that women are more closely aligned with their social group than with each other, their marginalization as a group remains concealed.

Asia

The Southeast Asian Geography Association (SEAGA) does not support a distinct feminist geography group. Yeoh and Huang (2007) overview gender and geography in Southeast Asia, in the countries of Singapore, Malaysia, Indonesia, Philippines, Thailand, and Cam bodia, arguing that the type of geography practiced in the region does not lend itself well to development of feminist geography groups. While women's and gender studies programs and centers are vibrant in the region, geography is 'a poorly resourced, mainly descriptive, conservative, task oriented subject valorized primarily for training teachers for school geography and producing policymakers who can contribute to national socio economic development'. Strategies to legitimize and strengthen women's and gender studies have included the establishment of networks, such as the Feminist Research Network (FEMNET) in Thailand.

The Association of Japanese Geographers also does not have a formal feminist geography group, but a group of feminist geographers have nonetheless been working and publishing together since the late 1990s. Partici pation in the IGU Gender and Geography Commission is also high, mainly as an outgrowth of the conference held in Seoul, Korea in 2000. According to Hiroo Kamiya: "We have tackled housing problems of Japanese single working women, socioeconomic conditions of one parent families, and women's role in the Japanese management system." The group has published *Urban Space of Single Women in Japan* (2004), and a recent issue of the Geo graphical Review of Japan (2006) was devoted to 'Resi dential choices and housing problems of working women in Tokyo'. The group has also published an article in Geographische Rundschau International, and the Gender Atlas of Japan. Their common concern lies in con temporary Japanese women within the context of glob alizing world economy.

Latin America

Among the Latin American countries, feminist geog raphy is most alive in Brazil and Argentina, owing to the strong ties to the Barcelona group. Cross Atlantic lec tureships between geographers in Barcelona and Buenos Aires, among others, have especially nurtured the de velopment of gender and geography in Argentina since the 1990s.

Feminist geography in Brazil and Argentina was shaped primarily by working class activism and local and global political economic currents, and remains situated within a type of geography that is positivist and empir ical. Special journal issues have been published and meeting sessions organized in feminist geography, but the field remains to become institutionalized in curricula and professional organizations. In Brazil, there are no feminist geography groups within the Association of Brazilian Geographers; however, feminist geographers work within a Research Group on Urban Analysis linked to the Ministry of Education's National Council for Techno logical and Scientific Development.

Conclusion

Uneven development of feminist geography and feminist geography groups worldwide is apparent. Feminist geography groups are most active where the production of research on women, gender, and feminism has become mainstream within larger institutional contexts, in uni versities, and within other geographical circles. Modest advances have been made recently by feminist geography groups in places where they heretofore were absent, most notably expressed through journal issues devoted to feminist geography, through courses taught in gender and geography; and through links with women's and feminist NGOs. The diversity of feminist geography and thus the extent of formal or informal feminist geography groups reflects the tremendous diversity of ideals, practices, obstacles, and goals of geographers in various places; and these in turn must be understood with respect to the type of knowledge production valued at the local and national levels.

See also: Feminism and Work; Feminism, Maps and GIS; Feminism/Feminist Geography; Feminist Geography, Prehistory of; Feminist Methodologies; Feminist Political Economy; Possibilism.

Further Reading

Awumbila, M. (2007). Challenging contexts: Gender studies and geography in anglophone African countries. *Belgeo: Feminist Geographies around the World* 3, 261 274.

Baschlin, E. (2002). Feminist geography in the German speaking academy: History of a movement. In Moss, P. (ed.) *Feminist Geography in Practice: Research and Methods*, pp 25 42. Oxford: Blackwell.

Blumen, O. and Bar Gal, Y. (2006). The academic conference and the status of women: The annual meetings of the Israeli Geographical Society. *Professional Geographer* 58, 341 355.

Da Silva, S. V. (2007). Geography and gender studies: The situation in Brazil and Argentina. *Belgeo: Feminist Geographies around the World* 3, 371 382.

Garcia Ramon, M. D. and Monk, J. (eds.) (2007). *Belgeo: Feminist Geographies around the World*. Heverlee, Leuven: Belgeo Geo Institut.

Longhurst, R. and Johnston, L. (2005). Changing bodies, spaces, places, and politics: Feminist geography at the University of Waikato. *New Zealand Geographer* 61, 94 101.

Monk, J. (2007). Generizando la geografía: Personas, lugares e ideas. (Engendering geography: Connecting people, places, and ideas). *Documents d'Anàlisi Geogràfica* 49.

Moss, P. and Al Hindi, K. F. (eds.) (2008). *Feminisms in Geography: Rethinking Space, Place and Knowledges*. Lanham, MD: Rowman and Littlefield.

Timár, J. and Jelenszkyné, I. F. (2004). Female representation in the higher education of geography in Hungary. *Journal of Geography in Higher Education* 28, 101 110.

Voiculescu, S. and Lelea, M. A. (2003). A review of gender and geography and its development in Romania. *Analele Universităţii de Vest din TimiŞoara, Geografie* 13, 123 144.

Women and Geography Study Group. (1997). *Feminist Geographies: Explorations in Diversity and Difference*. London: Longman.

Yeoh, B. S. A. and Huang, S. (2007). Placing 'gender' in 'geography' across Southeast Asia. *Belgeo: Feminist Geographies Around the World* 3, 383 397.

Feminist Methodologies

D. Thien, California State University, Long Beach, CA, USA

Glossary

Feminism A dynamic set of theories, practices, and social movements broadly committed to social change in the areas of enhancing women's social and political status, and gender equity.

Method A technique or tool used in the research process, for example, conducting interviews, administering questionnaires, or taking photographs.

Methodology The overarching rationale or theory for choosing and using methods in a research process, that is, a methodological framework that describes the ways researchers do research and why.

Positionality The particular context of any individual generated in diverse and uneven ways by the intersections of one's multiple locations (e.g., social, cultural, sexual, geographical, economic, and political).

Post-Structuralism A set of ideas comprising a theoretical critique of structural epistemological frameworks; a stance which encourages deconstructing unified narratives and acknowledging the diversity of multiple knowledges which may shift within and across space and time.

Power Within a post-structural framework, power is understood as a dynamic process, which circulates unevenly between people and places, creating material and symbolic effects.

Reflexivity Within research methodology, reflexivity is the practice of explicitly attending to one's positioning or location, and acknowledging how this affects the research process.

Relational The notion that identities (of people, places, and entities) are created always 'in relation to' something or someone else.

Situated Knowledges The production of knowledge is understood to be interdependent with one's particular situation or positioning, a stance which admits the existence of multiple knowledges.

Feminist Methodologies in Geography: A Brief Historical Overview

Feminist methodologies encompass a range of research techniques, practices, and theoretical frameworks. Collectively, feminist methodologies have made a major contribution to the field of geography by critiquing ontological and epistemological claims, which privilege masculinist worldviews, knowledge, and knowers; reshaping dominant research theories and practices, bringing attention to the presence and absence of particular bodies, subjects, and knowledges in geographical knowledge; and highlighting critical reflexivity as an important research process. The use of feminist methodologies within geography is usually, but not exclusively, associated with the subdiscipline of 'feminist geography'.

Counting Women

Early feminist research within geography was mainly quantitative. Influenced both by the 'quantitative revolution' within geography, and second wave feminist activism and scholarship, feminist researchers sought women's equal representation both as scholars and subjects of geographical research. Feminist geographers gathered empirical data to document women's material experiences, map their spatial representation, and empirically demonstrate their existence and importance as subjects of geographical research. Such feminist empiricism is often reductively referred to as the 'add women and stir' strategy of early feminist revisions to scientific scholarship. Yet, in addition to bringing women into geographical focus, this feminist empirical work contributed to an important epistemological shift: establishing women as knowledgeable researchers. Positioning women as producers of geographical knowledge created a profound challenge to the long standing separation of rationality (ascribed to masculinity and masculine bodies) from irrationality (ascribed to femininity and feminine bodies) within geography's 'positivist' scientific tradition, and so set in motion a significant critique of those epistemological foundations.

Qualifying Gender and Space

In the 1980s and into the 1990s, feminist scholars in geography shifted from the empirical mapping of women's spatial experiences to examinations of the significance of gendered spatial divisions which work to keep women and men in particular places. For example, attention turned toward landscapes and the built environment as tangible reflections of gender on urban and work spaces. The previously predominantly quantitative body of feminist research became more qualitative, as quantitative methods which counted and categorized women were viewed as unsatisfactory for detailing the complexity of women's everyday lives. The universal category of 'woman' as the foundation for feminist politics was rethought by feminist scholars, in response to critiques of feminism as a white, Western and middle class phenomenon. 'Woman' was exposed as an essentialist category, and feminist theorizing began to shift toward feminist politics underpinned by radical differences. Methodologically and substantively, with the category of 'woman' deconstructed, feminist scholars shifted from a strongly material set of concerns with women and their lives to a focus on more symbolic issues of gender inequalities.

As geography's 'cultural turn' took hold in the 1990s, a post structurally influenced agenda within geography, together with ongoing interactions with critical feminist scholarship, engaged feminist geographers in more critical theoretical positions and a more challenging, and increasingly qualitative methodological enterprise. Gender as a unit of analysis was reconsidered; for example, as constructions of particular masculinities and femininities; as performative; and as discursive or textual moments. The spaces of the geographic discipline were also critically reexamined. Gillian Rose's *Feminism and Geography* offered an important critique of geography's masculinist claims to knowledge. This influential text develops an analysis of scientific objectivity, with its unmarked and unremarked male subjects, and exposes the resulting disguised and partial nature of what is presented as scientific, geographic truth. This text marks a pivotal moment in feminist geography, because it articulates the gendered nature of geography's disciplinary practice – a validation of earlier feminist geographers' challenges and an encouragement for those to follow.

Bodies of Influence

Many contemporary feminist geographers give central weight to theorizations which acknowledge power relations as a key aspect of gendered processes. Influenced by feminist theorists such as Judith Butler and Donna Haraway, many feminist geographers no longer consider gender as an adjective, but rather as a lively, spatial process; that is, we 'do gender', and we do it differently through the enactment of multiple possible masculinities and femininities in our everyday practices, in our relations with others, and with the spaces of our actions, in our use of language, and use of space. These ontological and epistemological shifts have encouraged and necessitated multiple methodological possibilities for investigating gender identities for their different meanings, spatialities, and investments of power, including intersections of class, ethnicity, sexuality, and/or ability, as well as encouraging attention to bodies, emotions, and peformativity.

Feminist Theories and Geographic Practices

Feminist thinkers have had a major impact on feminist methodological developments in geography. Feminist philosophical, political, and epistemological critiques have assessed the ways of knowing intrinsic to positivist scientific research and identified a masculinist framework infused with the Enlightenment ideals, questioned constructions of authoritative representations of knowledge, and reconsidered who counts as a legitimate knower. Importantly, feminist scholars also have formulated other, valid ways of knowing. Such epistemological repositionings include feminist empiricism, standpoint theory, feminist *postmodernism, situated knowledges*, and relationality.

Feminist empiricism, feminist standpoint theory, and feminist postmodernism are epistemological reorientations of science famously elaborated by Sandra Harding in *The Science Question in Feminism*. Empiricism values experience as the only basis for knowledge. However, feminist empiricism illuminates the existing sexist biases in scientific inquiry, questioning the neutrality of what is deemed as scientifically valid experience. Methodologically, feminist empiricism seeks to reform science by proposing the inclusion of feminist and other value systems as a corrective for biased scientific methods – effectively, to make 'bad' science, 'good'.

Standpoint theory advances the idea that an individual has a particular 'standpoint' from which s/he is able to know the world, and which gives that person epistemological privilege. Influential proponents of standpoint theory include Patricia Hill Collins, noted for her analysis of Black women's standpoint, and Nancy Hartsock and Dorothy E. Smith, both of whom build from Marxist theory to explore the consequences of evaluating knowledge from the perspective of women's lived realities. For example, women's labor may be invisible or incomprehensible when considered within a male biased labor analysis; however, research grounded in the perspective of women ensures their labor is counted.

Standpoint theory, with its claim for knowledge constructed from a woman's way of knowing, has proven controversial because of its implication that women form a coherent, homogeneous group. In addition, standpoint theory's legitimacy has been tested when employed to justify privileged knowledge of a research population, for example, when women interviewing women are understood to exploit their shared feminine experiences to manipulate female participants. As a methodological strategy, standpoint has also been questioned, for example, when female researchers are presented as better qualified to interview women because all women share the experience of sexism.

Feminist postmodernist or post structural theories reject the universalism presumed by standpoint theories. Donna Haraway, for example, has influentially extended the notion of standpoint in her vision of feminist objectivity: 'situated knowledges'. She argues that to be situated is more complex than having a standpoint;

instead, it is key to acknowledge that any location is limited and also that knowledge is embodied, that is, grounded in real bodies. Such embodied knowledge comes from being in place, and from acknowledging the partiality of that location, rather than from claiming a special privilege or knowledge of oppression.

In contrast to a structuralist emphasis on social and cultural systems of organization presided over by a disembodied master subject, this post structural ren dering of scientific knowledge questions the unity and stability of any subject. Importantly, this questions the processes of becoming a subject. Without ignoring the potentially powerful effects of deeply entrenched boundaries, such as the long standing separation of 'mind' from 'body', or 'man' from 'woman', post struc tural feminists have sought a more relational, or 'inter subjective' understanding of subject formation. Methodologically, this post Cartesian rethinking has made explicit the significance of spatial and embodied aspects of identity. Feminist methodologies have privil eged relationships, and as such, intersubjective or relational dynamics of knowledge production and re production are highlighted.

Efforts to theorize 'subjectivity' have also enlivened understandings of gender and gender identity while complicating methodological endeavors. In *Gender Trou ble*, Judith Butler influentially proposes that the substance of gender is constituted in its very performance: gender is performative. That is, gender is not about who we are, it is not *a priori*, nor is gender an aftereffect of pre established identity; rather, identity is a function of so cial, cultural intelligibility, and as such, who we are always in the process of becoming in relation to (gen dered, masculinist, and heterosexist) regulatory dis courses of personhood.

Considering subjectivity as a process has allowed for a thoughtful examination of how this process is both in extricable from power relations and also, as geographers such as Doreen Massey and Isabel Dyck have noted, intertwined in places; that is, subjectivity is also a spatial process. Feminist geographers Pamela Moss and Liz Bondi have each emphasized not only the specificities of particular placings, but also the oscillation of our mul tiple positionings, a series of ever shifting relational contexts within which we are variously embedded and which must be considered as part of our methodological framework. Liz Bondi emphasizes relationality in her assessment of processes of empathy and identification in interview encounters. Bondi offers the interview relation as an oscillation, a movement of meaning making, within which the interviewer moves between positions of ab sorption to ones of observation. Bondi argues that this oscillation is necessary for the creation of a 'psychic' space which allows the similarities and differences of the interviewer and interviewee to be realized, if mostly unconsciously. Rather than position emotional involve ment and analysis as mutually exclusive, the relational encounter requires acknowledgment and management of emotional content as an explicit part of the research process.

The emphasis on identity as a process and as know ledge produced 'in relation to' an other has methodo logical implications. What are gathered as data, what 'counts' as knowledge, the necessarily limited possibilities for self awareness in any research encounter, and the partial quality of all information must be accounted for in methodological decisions. Reflexivity refers to the pro cess of naming or 'locating' a researcher's various subject positions, as a means of acknowledging the partiality of ones' positioning and so the limits of one's knowledge and contribution to knowledge, and of knowledge itself. This process stands in intentionally marked contrast to historical scientific claims to objective truth, presented as a seamless, disembodied, neutral process of collecting facts. Within feminist geography research, self reflexive research has become a methodological standard. Reflex ivity is characterized by the researcher's close self examination of how s/he shapes and is shaped by re search encounters. In particular, feminist geographers consider the effects of gender and power upon/in re search encounters, for example, on what kinds of research questions are asked, what resulting research data are gathered, and where, and how those data are interpreted. Debates about reflexivity within feminist qualitative re search range from the criticism that this practice is overly self centered to the insistence that it is a necessity of ethical research.

Feminist Methodologies in Geography: Practices and Politics

Feminists in geography made early methodological in roads by highlighting what was lacking in existing geo graphical research and redefining what was suitable for study. Where data about women's experiences were ei ther nonexistent, or buried in assumptions and mis perceptions about women's lives, feminist geographers sought to remedy the situation. Existing data were dis aggregated to discover data on women, and data ex plicitly about women's experiences were collected. But, encouraging appropriate and relevant data collection and management is only a partial story of feminist meth odological approaches within geography. These ap proaches reflect their feminist roots, in that feminist research has been characterized not just by 'what' it studies, but also 'how', 'why', and 'where'.

Feminist methodologies in geography emphasize the importance of how research is done (methods), why it is done (methodologies and epistemologies) – underscoring the politics of a feminist research process, and where (e.g., the 'field' and the academy). Shulamit Reinharz's question, "is there a feminist method?" can be extended to "what makes a methodology feminist within geography?" The emphasis is not on particular techniques, as much as it is on how such methods are used; how and why particular research questions are chosen; and how, why, and where feminist researchers are engaging with the multiple and dynamic subjects of geographic research, including themselves. Although by no means limited to the following, feminist methodologies often incorporate some common themes: (1) a critical attention to issues of social injustice, (2) an agenda for social change, (3) an emphasis on ethical research relations, and (4) an explicitly subjective research process. Within geography, these are still present, but an additional preoccupation with the gendered spaces and places of research also affect feminist methodological choices in geography.

Methods

Feminist methodologies have incorporated both quantitative and qualitative methods, with preferences which have shifted over time. In the 1990s, the bulk of feminist research within geography employed qualitative research methods. Indeed, the entry for 'Qualitative methods' in *A Feminist Glossary of Human Geography* simply says "See Feminist Methodology." This qualitative slant reflects feminist geography's sustained critique of positivist trends in geographic research which emphasize quantitative methods designed to garner factual, testable data. Qualitative methods, in contrast, are deemed especially useful for gathering subjective data about perceptions, experiences, and representations. Furthermore, qualitative methods have been seen to offer greater potential for fulfilling feminist political aims; for example, by making space for women and their experiences, by inviting a collaborative research process, and in so doing, making a deliberate stand against the required objective distance of scientific research. This epistemological stance has been regarded as a strong 'fit' with feminist research frameworks which sought first to highlight women's particular experiences in the world, then to document inequalities and their effects, and later to investigate the ways in which power relations, gender, and spaces intersect. Finally, qualitative methods also invite a specific, embodied awareness of the spaces and places of geographic research.

The particular qualitative methods employed by feminist researchers in geography include interviews, autobiographies, oral and life histories, ethnography, participant observation, focus groups, participatory action research, role play, photography, video, and critical, textual, and discourse analyses. This diversity of techniques reflects both the range of feminist geographies and the interdisciplinary roots of feminist research within geography.

Qualitative Methods

The interview, a one to one structured, semi structured, or unstructured research dialog between researcher and participant is favored as the classic feminist research technique. This qualitative method is valued for putting the production of knowledge into the hands of the participant. Interviews privilege respondents' voices, according authority to their experience, rather than subjecting them to scientific scrutiny. However, as more nuanced theorizations of power have circulated, claims that the interview method straightforwardly validates women and women's experiences have met with resistance. Feminist researchers have probed the way in which both interviewer and interviewee do and do not have power, the multitude of ways in which power is exercised, for example, by refusing to answer questions or withholding key information, and the ways in which relations of power within any research encounter may be concealed or revealed, accessible or inaccessible to either party.

Similarly to interviews, autobiographies, life histories, or oral histories have been employed as feminist methods which gather women's untold, unheard, or undervalued stories. Within geography, such narratives have been used as a method for reading the relationships between gender and place, for example, in feminist reorientations of specific landscapes such as the home, and as a means of placing the researcher. Narratives can demonstrate the specificity of individual lives in particular places, while also revealing the powerful effects of wider sociocultural, political, or economic forces. However, the notion of a seamless, coherent, 'true' account of any life has been deconstructed, raising questions of authority and agency with regard to the collection and representation of life narratives. These feminist and post structural challenges to the universality or linearity of narrative claims to knowledge have been met in part by reflexive research practices which call for an awareness of one's own and others' always situated positions.

Ethnography encompasses a range of classic anthropological techniques, including participant observation, interviews, and associated textual and/or archival analysis. An ethnographic study is an in depth study of a particular group of people, usually carried out in a fieldwork context. While ethnography is not inherently feminist, the ethnographic suite of methods has been seen as a good fit with feminist aims, because these methods privilege subjective accounts, and encourage engaged dialog between researchers and communities. As such, an ethnographic approach offers an interpretative alternative to the focus on objective fact collection prioritized in positivist research. A methodological concern is the ethnographic focus on small groups of people in par ticular places, precluding the application of research conclusions to other populations or contexts. However, many qualitative researchers willingly accept the limi tations of a small sample size in favor of the depth and richness of an ethnographic study. Additional concerns associated with ethnographic methods include issues of ethics and emotion which may arise as a consequence of the potentially intimate nature of fieldwork relations. However, feminist debates suggest that such issues arise in any research context and an explicit attention to these concerns enhances research quality.

Focus groups have been utilized as a feminist method because they offer a potentially more flexible and less hierarchical arrangement than a one to one interview. Feminist arguments in favor of the focus group suggest that the larger number of participants in a focus group allow participants to engage more safely in discussions about matters of concerns to them, as opposed to in one to one conversation directed more explicitly by the goals of an interviewer. In addition, the focus group environ ment is argued to demonstrate more closely the pro duction of knowledge as a process of collective, relational meaning making, thus offering the potential for obser vation of this process in action.

Participatory action research (PAR) explicitly seeks active social change through collaborative process. As a methodology, PAR is organized to respect and prioritize the knowledges and contributions of diverse research participants in all steps of the research process. Col laborative forms of research can be challenging, par ticularly in cross cultural research settings, due to differences in conceptual starting places or because dif ferent research outcomes are expected, needed, or wan ted. Seeking appropriate means for cross cultural work has required 'decolonizing methodologies' (a phrase coined by Linda Tuhiwai Smith in her eponymous book) in order to determine appropriate postcolonial, and antiracist methodological frameworks. However, the challenges of difficult engagements necessitate what can also be construed as the rewards of participatory re search: careful and respectful attention to the lives of others, to relationships, and to ethically sound research practice. These become ongoing reiterations of how to do collaborative and cross cultural research well.

Quantitative Methods

Quantitative methods are neither absent nor rejected in feminist methodologies. Quantitative methods are used in a variety of ways by feminist researchers in geography, despite the assumption that feminist research is purely or most appropriately qualitative. As the special issue of the *Professional Geographer*, *Should Women Count?*, exemplifies, quantitative methods are not, in and of themselves, antifeminist; rather it has been the underlying politics and the epistemological positionings of such methods within masculinist claims to objective truth which have resulted in feminist critiques. Quantitative methods em ployed by feminist researchers in geography include statistical measures, such as frequency counts and log linear models, as well as spatial analytic measures and geovisualization techniques represented by geographic information systems (GISs).

Mixing Methods

The use of multiple methods, known as triangulation, offers the opportunity for a thorough understanding of a research topic from multiple data sources, and also in vites validation of one source of data by another (e.g., comparing interview transcripts to survey results). As such, triangulation may be employed as a strategy to increase the validity and rigor of research. However, within feminist methodological discussions, the use of mixed methods as a means of garnering a more complete representation of reality is debated. Instead, mixed methods research offers an excellent demonstration of the partiality of any one method, and exemplifies the dynamic and multiple possibilities of geographical knowledge.

A rise in mixed qualitative and quantitative methods research has followed the qualitative trend in the 1990s in human geography. The move toward mixed methods research emphasizes the respective strengths of both qualitative and quantitative research. While qualitative research can illuminate socially constructed and indi vidually lived meanings, quantitative research can high light key points of difference on a broadscale.

Troubling Feminist Methodologies in Geography

Feminist geographers have strenuously and productively debated the politics and ethics of who does research, and how research is done and thought about, for whose benefit, and where. Such discussions consider the changes brought as analyses of power dynamics, and gender re lations have become more nuanced; as the possibilities and complexities of feminist research within a mascu linist disciplinary tradition have been articulated; and as the political claims of those in research processes have grown more divergent. In the process, feminist method ologies in geography have undergone rigorous (self) examination.

Researchers, by making a number of choices of in terpretation, and decisions about representation, dem onstrate the power invested in the construction and production of knowledge. The methodological dilemmas for feminist researchers include: how to subvert the masculinist authority invested in scholarly knowledge production, how to achieve feminist political aims within the context of scholarly expectations, and how to make space for women's identities while challenging existing subject positions available to women. Feminist methodological debates illustrate the complexity of negotiating feminist scholarship, geographical disciplinary frameworks, political engagements, and personal commitments.

Feminists have been acutely conscious of the issues of power that necessarily arise between differently situated researchers and those with whom research is done, as well as from the ways in which knowledge is differently validated depending on who is producing the knowledge and about whom. Feminist geographers have been encouraged to practically address and debate these ethically complex issues of power and equality, including the adequate and appropriate representation of research 'subjects', the recognition that researchers are not self evidently in more powerful positions than the research participants with whom we work, what counts as whose knowledge, and the importance of a reflexive research process.

Representing Research 'Subjects': Whose Voices, Whose Power?

Respect for research participants remains a core concern in feminist methodological work in geography. In research practices, feminist researchers continue to seek ways to respectfully acknowledge and represent embodied identities, while remaining sensitive to issues of ethics and power, and remaining consistent with an underlying feminist commitment to social change research. The complexities of such representations have laid bare conflicting agendas. The politics around 'giving voice' to those outside research roles, while also claiming an institutional, professional, or personal research authority, epitomize the competing agendas of feminist research.

Feminist researchers struggle between the desire to validate the knowledges of the 'researched' and to politicize that knowledge in political and/or academic domains. The desire to validate women's experiential knowledge is a historically integral component of feminist methodologies which have a long standing social change agenda. However, this notion of affirming women's realities as political act, evocative of second wave feminist consciousness raising activities, has been extensively critiqued. In particular, Black American feminists have expressed concern that validation of women's experiences by academic researchers promotes White feminist agendas and de emphasizes women's diverse representations. Despite such perceptive efforts to question the fixity of power, and to trouble the capacity of a researcher to 'empower' research participants, researchers often subscribe to an 'empowerment' agenda. Postcolonial, antiracist, and indigenous perspectives continue to challenge Western feminist representations and to highlight the limits of feminist epistemological challenges derived from a dominantly White Western academy. The increasing mobility of Third World women scholars and the resulting ongoing challenge to any uniform notions of researcher privilege is just one example which underscores the importance of this discussion.

Whose Knowledge?

The demand for reflexivity as a feminist methodological strategy has raised ethical and moral questions. To what extent should a researcher disclose personal, political, or academic alliances to research participants? Strategic concealment of personal or political details while in the 'field' may be practical or comfortable; however, intentional misrepresentation raises ethical considerations, such as the possible imbalance between what is asked for and what is shared with a participant. As well, any attempt to deliberately shape a research relationship inevitably alters the research outcomes. Sharing research outcomes with participants has become a common practice for feminist researchers; however, decisions about how to present data to a participant audience (e.g., by providing reports which are different from produced scholarly texts) raise complex questions about the production of and power to control and translate knowledge. These issues of representation and translation are ongoing methodological considerations for the feminist researcher.

Feminists in the Field(s)

Geography has long privileged fieldwork as an essential and distinctive component of geographical research. The 'field' is the place where geographical knowledge has been discovered, often in the form of imperial conquest of colonial spaces. Feminist critiques have assessed this place as one of pernicious repetition of masculinist normative research practices which leave feminine bodies out of place in a masculine field. Fieldwork reconsidered as a feminist endeavor has been the subject of much feminist methodological discussion. In particular, feminist researchers have challenged the gendered nature of fieldwork, the notion of the field as a clearly defined, bounded area, and the conventional distinctions such a place sets up between bodies who conduct research and those who participate or are discovered. In addition to the 'field', the academy, and to a much lesser extent, the classroom, have been considered by feminist researchers for their effects on where and how research takes place.

Feminist Methodologies in Geography: New Directions

Feminist methodologies in geography are established enough to have achieved a certain orthodoxy; however, this remains a dynamic area of activity within feminist and human geographies. Theorizations of gender, power, difference, and their attendant and shifting spatialities continue to matter to feminist politics and feminist research practices. Current trends in feminist methodological practice in geography demonstrate an explicit focus on extending theorizations of identities and embodiment, including 'raced' and 'sexed' bodies; expanding critical geographical methodologies to consider (e.g., antiracist geographies, decolonizing methodologies, and emergent transnational identities); and incorporating both qualitative and quantitative methodologies into feminist research (e.g., into geographical explorations of affect and emotion).

Identities and Embodiment

Pamela Moss, an important voice in feminist methodological discussion in geography, puts forward the notion of 'specificity' as an intentionally troublesome reworking of partiality. Specificity details the ways in which bodies become specific in active relationships to other bodies and within multiple possible spaces. As a methodological intervention, specificity unsettles claims to clarity which may accompany a declared partial perspective; instead, it asserts the dynamic and multiple nature of participants' situated locations in research settings. Methodologically, specificity challenges a single (partial) view of research, of method, of methodology itself, and this has consequences for how research is presented, how the value of research is assessed, and what and how we constitute as feminist methodologies in geography.

Gill Valentine also attends to bodies in considering the methodological enactment of intersectionality, a theorization of the multiple relations between social categories. She proposes intersectionality must be flexed in more than theoretical ways; instead, it must be more explicitly empirical and embodied as a means of attending to power and inequalities. The deliberate 'practice' of intersectionality invites an attention both to bodies and the mutually constituted spatialities of lived experiences, as well as to the workings of power which lead to exclusionary or inclusionary spacings.

Critical Methodologies

Linda Tuhiwai Smith's *Decolonizing Methodologies* is an important text for feminist methodologists. While crediting feminist scholarship for its profound impact on scholarship, especially regarding the consistent epistemological critiques of positivist research traditions, Smith nonetheless 're marks' the limits of the critique – namely, as women of color have argued since the 1970s, feminist revisioning is still bound by Western or European worldviews. Such persistent cultural orientations continue to lie beneath, and so limit feminist methodological endeavors. Smith urges researchers to go beyond a reflexive accounting of individually held beliefs, and instead to take deliberate care to examine the effects of knowledge as a legitimating process, and to thoughtfully and explicitly consider whose agendas and beliefs are legitimized in any research project. To these ends, Audrey Kobayashi outlines an ongoing antiracist feminist and geographical project with methodological implications. She argues for activist engagement, the use of positionality as the beginnings of social transformation (not reflexivity as a merely descriptive exercise), and the acknowledgment of processes of racialization in daily practice and scholarship. Clearly, it remains critical for researchers to acknowledge and continually problematize power relations in research. Feminist methodological debates about power and gender have continued relevance in the context of increasingly transnational iterations of feminist research.

Mixing Methodologies

Mei Po Kwan and other feminist geographers have extended feminist methodologies in geography by collectively offering a set of post structural, post positivist critical perspectives on GIS. Inspired by feminist theorists, and also by qualitative analyses of emotion in geography, key elements of this feminist geographic work are the emphases on the significance of the spatial as embodied, as emotional geographies, and on the political. Simultaneously, this work reclaims the power of maps and reconfigures the epistemological underpinnings of GIS, reorienting their potential as critical, political tools for intervention and feminist critique.

Enacting, Embodying, and Embracing Feminist Methodologies

These diverse feminist methodological directions continue to raise questions about what makes methodological endeavors feminist, and to explore what constitutes the spaces of feminist research methodology with geography. This very ability to call methodological actions, enactions, and spaces into question is a strength of feminist geographical scholarship. Feminist methodological approaches in geography are inspired by a collectivity of feminist scholars who demonstrate a vast capacity to diversify, an imaginative creativity for rethinking methodological dilemmas, and a desire for the long standing social change agenda of feminist scholars. Feminist methodologists within geography continue to extend the repertoire for effective, ethical, and productive practices of feminist methodologies.

See also: Embodied Knowing; Emotional Geographies; Emotional Knowing; Feminism/Feminist Geography; Situated Knowledge, Reflexivity; Subjectivity.

Further Reading

Bondi, L. (2003). Empathy and identification: Conceptual resources for feminist fieldwork. *ACME* 2(1), 64 76.

Bondi, L., Avis, H., Bankey, R. *et al*. (2002). *Subjectivities, Knowledges, and Feminist Geographies: The Subjects and Ethics of Social Research*. Oxford: Rowan & Littlefield.

Butler, J. (1990). *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge.

Davidson, J. and Bondi, L. (2004). Spatialising affect; affecting space: An introduction. *Gender, Place & Culture: A Journal of Feminist Geography* 11, 373 374.

Haraway, D. J. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies* 3, 575 599.

Harding, S. G. (1986). *The Science Question in Feminism*. Ithaca, NY: Cornell University Press.

Katz, C. (1996). The expeditions of conjurers: Ethnography, power, and pretense. In Wolf, D. L. (ed.) *Feminist Dilemmas in Fieldwork*. Boulder, CO: Westview.

Kobayashi, A. and Peake, L. (2000). Racism out of place: Thoughts on whiteness and an antiracist geography in the new millennium. *Annals of the Association of American Geographers* 90(2), 392.

Kwan, M. P. (2007). Affecting geospatial technologies: Toward a feminist politics of emotion. *Professional Geographer* 59, 22 34.

McDowell, L. (1997). Women/gender/feminisms: Doing feminist geography. *Journal of Geography in Higher Education* 21, 381 400.

Moss, P. J. (2002). *Feminist Geography in Practice: Research and Methods*. Oxford: Blackwell.

Moss, P. J. (2005). A bodily notion of research: Power, difference, and specificity in feminist methodology. In Nelson, L. & Seager, J. (eds.) *A Companion to Feminist Geography*, pp 41 59. Oxford: Blackwell.

Parr, H. (2001). Negotiating different ethnographic contexts and building geographical knowledges: Empirical examples from mental health research. In Dwyer, C. & Limb, M. (eds.) *Qualitative Methodologies for Geographers*, pp 181 195. London: Arnold.

Reinharz, S. (1992). *Feminist Methods in Social Research*. Oxford: Oxford University Press.

Rose, G. (1997). Situating knowledges: Positionality, reflexivities and other tactics. *Progress in Human Geography* 21, 305 320.

Sharp, J., Browne, K. and Thien, D. (2004). *Women and Geography Study Group of the RGS IBG, Geography and Gender Reconsidered*. London: WGSG.

Sharp, J. P. (2005). Gender and geography: Feminist methodologies in collaboration and in the field. *Progress in Human Geography* 29, 304 309.

Smith, L. T. (1999). *Decolonizing Methodologies: Research and Indigenous Peoples*. London; Dunedin: Zed Books; University of Otago Press.

Valentine, G. (2007). Theorizing and researching intersectionality: A challenge for feminist geography. *Professional Geographer* 59, 10 21.

Feminist Political Economy

R. B. Sarma, Cotton College, Guwahati, India

Glossary

Economic Restructuring The process of structural transformation, from manufacturing to services.
Feminization of the Labor Force It is an ongoing process through which women in increased numbers enter the labor market.
Feminine A set of social attributes associated with women or female as opposed or inferior to masculinity or masculine.
Human Capital Human capital can be defined as the ability, skill, or knowledge acquired by an individual.
Labor An activity that can be either waged (usually performed for an employer) to gain an income or unwaged (performed usually) within the household.
Nimble Fingers Thesis that argues to demonstrate that wage differentials in the market factories all over the world tend to be male biased. It also illustrates that nimble fingers are socialized skills and not naturalized.
Occupational Segregation The extent to which particular groups of workers are concentrated within different occupations.
Political Economy An approach that treats economics as an extension of power and politics required to study any activity that connects to the production and distribution of wealth.
Power The capacity of controlling influence to do something in a specified way.
Work An activity that may involve either mental or physical force or both to carry out a task or manufacture a good.

Introduction

The study of the geography of feminist economics has become a central concern to virtually all branches in human geography, sociology, and beyond. In the sub discipline's recent history, over the last two and a half decades, this study has made contentious developments. Both economic and feminist geography have had am bivalent relationships to Marxist theory and scholarship. With the publication of the stirring books by David Harvey, Doreen Massey and Neil Smith, Marxist schol arship became widely used to redefine the research frontiers of economic and feminist geography. While economic geography's core aim is to explore and re explore the economic dimensions of the people–place relationship embedded in the dynamics of capitalism, urbanization, imperialism and development, the pro duction of nature, restructuring, globalization and neo liberalism, the geography of feminist economics aims to study the gendered nature of economic processes, con fined largely to the issues of work and labor alongside the aforementioned processes of economic dimensions. Many feminist economic geographers have used a political economic framework to discuss the shift in the gendered nature of economic development. This article addresses the notion of work and labor that has been central to the feminist political economy.

Feminist Political Economy: Notion of Work and Labor

In the nineteenth century, women's rights advocates and writers such as Harriet Taylor Mill (1807–58) and Charlotte Perkins Gilman (1860–1935) paid attention to the exclusion of women from economic activity and their continuous dependence on men. It was actually since the 1970s that the notion of work and labor has been a highly contested and is a wholly pressing debate within feminist geography with many different claims to what, where, and how these notions should actually be studied to signify the politics of gender. It is, therefore, essential to unpack these two notions as a concept to study the geography of feminist political economy. Marxist schol arship that has revolutionized the working class politics, while analyzing the socioeconomic system of capitalist society, has included women's struggle for equality within this class politics. At the same time, this scholar ship has also tended to negate the significance of gender politics as an independent cleavage of this class politics. Nevertheless, in the early twentieth century, many Marxist writers such as Clara Zetkin (1857–1933), Rosa Luxemburg (1871–1919) and Alexandra Kollantai (1872–1952) averred that women's struggle for liberation cannot be flagged as an independent cleavage; instead they fo cused on the inclusion of this struggle with that of so cialism while remaining within the Marxist scholarship. Second wave feminism that emerged in the 1960s, as a universalized and radical expression of women's rights in capitalist societies as a result of the anti Vietnam war movement and civil rights campaign, was also inclined to Marxist philosophy as a radicalized platform to refocus on the notions of work and labor.

Traditional economic geography defines an economy as consisting of only firms and markets, thereby failing to acknowledge the home economics, which is often based on structures of power as well as social and political relationship, embracing domestic division of gender segregated work, including the care of children, other household members, and shopping. Arguably, the activities of unpaid labor of the household provide hugely in the maintenance of household standards of living and to the sustainability of the overall economy. Feminists, therefore, have attacked the notion of work and divided the same into 'remunerated work' (employment or paid labor) and 'household chores' (constituted of domestic labor and child care or unpaid labor). Feminists thereby argue that the spatial structure of remunerated work is based only on the production and reproduction (or manufacture of goods and services for the market), which render invisible the tensions and contradictions making up the notion of 'home'. Therefore, feminist economic geographers have urged an 'add on' and 'count in' strategy to address the gendered nature of development. For example, in 1988, Waring urged adding a monetary value to the economics of unpaid domestic labor and then counting its value as a part of gross domestic product (GDP). Again, Ironmonger in 1996 has argued that the value of unpaid labor at home is approximately equal to the valuation of the goods and services manufactured for the market by paid workers. Ironmonger, therefore, urges an inclusion of this valuation of unpaid labor as gross household product (GHP). In an ideal world, however, the reality of feminizing the economy is far more complex. In 2003, Cameron and Gibson Graham urged the possibility of an alternative economy. Drawing alongside the works of many other contemporary feminists, Cameron and Gibson Graham supported the arguments of Waring and Ironmonger and examined the various ways, approaches, and hidden predicaments in which women face exclusion from the economy and called for the possibility of an alternative economy. Nevertheless, the key insights to the inaccurate representation of the gendered nature of development lie within the spatial polity of public/private divide.

Contemporary debate about the public sphere is drawn mainly from the seminal works of Jürgen Habermas. In 1989, Habermas, for example, argued that the public domain represented the citizen's power to actively participate and debate in the political life, while the private domain represented the lack of such power. Public/private dichotomy is, of course, far more complex than this mere distinction would suggest. But much of this complexity would serve to elaborate this simple division. Its literature has expanded rapidly in recent years, diffusing to diverse regions and periods over space and time that reflects mainly on current issues, social processes, and events, such as class, democracy, anticorporate protests, and the emergence of the Internet. Feminists too have joined this project to challenge the notion of public/private dichotomy. Habermas is criticized by the feminist school of thought. In 1989, Fraser, for instance, was criticized for overlooking the significance of the gendered and the unequal social relationship insinuated in the actual or conceptual institutional segregation of the family from the market relationship or from the political. It is well documented that such a conceptual framework divulging the conflicts between women's use of public and private space has reinforced women's oppression. Earlier, in 1983, Garimarnikov *et al.* argued that the amount of domestic labor performed by women at home, although crucial, is at the same time implicit in enabling men to devote all their energies to a competitive labor market. Arguably, the gender relations in the labor market impinge strongly on the constitution of gender relations in the family and the household. Importantly, this dichotomy has played a powerful role in the construction of gender relations. The notion of work came to be enforced spatially under the metaphorical division of public (paid work synonymous with employment) and private spheres (unpaid labor) since the Industrial Revolution that began in Great Britain in the second half of the eighteenth century. The emergence of this division brought new divisions of labor between working class men and women and idealized separate boundaries for middle class men and women. The materialization of such spatial binary dualism came to be associated with that of the 'home' (feminine) for women and 'work' (masculine) for men. Arguably, the hidden politics of public/private dichotomy is rooted within the notion of power which itself is a difficult notion to convey. Power is experienced in everyday life. In 2003, Allen tried to situate the notion of power from its spatial vocabularies to its spatial constitution. In doing so, he argues that space determines the role of power. Such tangled notion of power is generally expressed under four divisions: power over (power by men over men; men over women, that is, ability to influence); power with (power from collective action); power to (arrange to change the existing hierarchy); and power within (individual consciousness, that is, sense of self worth and self knowledge of a person). Feminists have continued to argue that woman's subordinated status is a universal fact. Arguably, power over features 'home' as a private and quiet space for women to perform unpaid labor, while 'paid work' is featured as a public space for men to perform paid labor. Many feminist scholars argue that increasing participation of women in the paid employment accentuate the public/private distinction as fluid and blurred. Arguably, many women empower themselves, moving beyond the power plays ('power over' and 'power with') to participate in paid employment with their 'power to' and 'power within' assets. Nonetheless, the thesis of public/private dichotomy is discredited for being hegemonic, Euramerican, and enthnocentric. For example, many social scientists, including Momsen in 1991, argue that this thesis fails to acknowledge the fact that in developing economies, the majority of women's labor is absorbed into the informal sector, although they are remunerated, but such work is unrecognized in the official statistics on employment.

It is well known that, since the end of World War II, there has been a steady growth in the number of women in paid employment, although spatially uneven across the globe (**Figure 1**). Due to the absence of statistics re garding global trend of women's employment since World War II, only the years 1993, 1996, 2003, and 2006 are mentioned (as produced by the International Labour Office), and similarly for **Tables 1** and **2**, while **Table 3** mentions the data only for the years 1996 and 2006. According to statistics produced by the International Labour Office (ILO), women constituted approximately 49% of the employment to population ratio in 2006, compared to 50% in 1993 – ILO's definition of em ployment to population ratio refers to a particular economy's capacity to absorb the working age population (generally 15 years or above). During the same period, male employment to population ratio too declined from 76% in 1993 to 74% in 2006 (**Table 1**). It has been observed that labor force participation rate (either in work or actively looking for a job) of younger women has decreased gradually because of their increased partici pation in education – ILO's definition of work includes self employed, employed, employers, and unpaid family members (including unemployed women). However, adult women's participation rate in the labor market has increased slightly. This finding reinforces the previous findings of Anglo American scholars where they argued that, in many countries, there has been a gradual increase in the numbers of married women (or women with children) participating in the labor market than there were in the past.

Nonetheless, the picture presented in **Table 1** and **Figure 1** seems fudged. It does not explain the reasons for women being employed, nor does it explain the quality of jobs. Further, it fails to indicate the percentage of women actually pursuing education. Moreover, it is documented that women remain unemployed at the rate of 6.6%, which is higher than men's unemployment rate of 6.1% (**Table 2**). Besides, employment to population ratios are lower for women than for men (**Tables 1** and **3**).

However, it is well documented that the increasing incorporation of women into paid employment is re inforced by the ongoing and uneven process of economic restructuring: a shift from manufacturing to service economy. This growth is referred to by academics, in cluding economic geographers and development prac titioners, as feminization of the labor force. The feminization of paid labor, therefore, is a process by which changes in global economies and market led growth alters the demand for (and supply of) women's paid labor in situations where new job opportunities in the growing service sector emphasize skills (such as communication and nurturing). These jobs remain typi cally 'sex typed', those that demand the so called 'nimble fingers' and female type professions with low wage, low status, insecurity, and classified as semi skilled, flexible, and mostly part time. Arguably, feminization of the labor force is accompanied by the fact that women's share in industrial employment is much lower than that of men and has decreased over time.

In the contemporary society, especially in advanced industrial economies, women are slowly gaining access to those areas traditionally dominated by men, such as the legal profession, science and technology, business, bur eaucracy and government, and more senior appointments in healthcare. Importantly, this increasing presence of

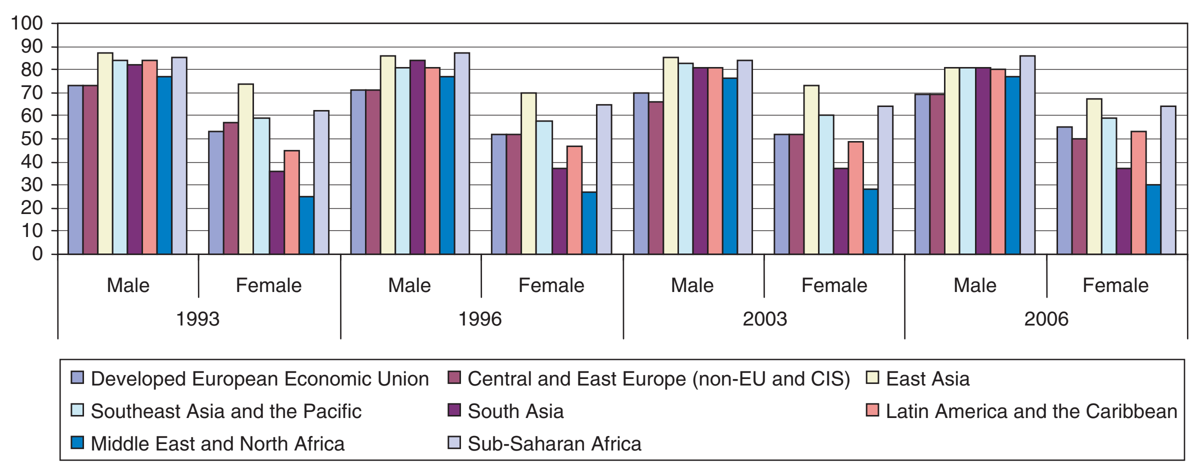

Figure 1 Labor force participation rates, by sex and region (1993, 1996, 2003, and 2006). Source: Global Employment Trends for Women (2004). International Labour Office and Global Employment Trends for Women, Brief, March, 2007, International Labour Office.

Table 1 G oba abor market nd cators[a] (1993, 1996, 2003, and 2006)

	Female				*Male*				*Total*			
	1993	*1996*	*2003*	*2006*	*1993*	*1996*	*2003*	*2006*	*1993*	*1996*	*2003*	*2006*
Labor force (m ons)	1006	1052	1208	1238	1507	1592	1769	1852	2513	2644	2978	3090
Emp oyment (m ons)	948	985	1130	1157	1425	1497	1661	1738	2373	2483	2792	2896
Labor force part c pat on rate (%)	53.5	53.0	53.9	52.4	80.5	80.5	79.4	78.8	67.0	66.7	66.6	65.5
Emp oyment-to-popu at on rat o (%)	50.4	49.6	50.5	48.9	76.1	75.7	74.5	74.0	63.3	62.6	62.5	61.4

[a] According to LO labor market indicators include labor force participation unemployment sector and status of employment wages/earnings and education and skills

Source Global Employment Trends for Women (2004) nternational Labour Office and Global Employment Trends for Women Brief March 2007 nternational Labour Office

Table 2 G oba unemp oyment rates (1993, 1996, 2003, and 2006)

	Female				*Male*				*Total*			
	1993	*1996*	*2003*	*2006*	*1993*	*1996*	*2003*	*2006*	*1993*	*1996*	*2003*	*2006*
Unemp oyment (m ons)	58.2	66.7	77.8	81.8	82.3	94.7	108.1	113.4	140.5	161.4	185.9	195.2
Unemp oyment rates (%)	5.8	6.3	6.4	6.6	5.5	5.9	6.1	6.1	5.6	6.1	6.2	6.3

Source Global Employment Trends for Women (2004) nternational Labour Office and Global Employment Trends for Women Brief March 2007 nternational Labour Office

Table 3 Global male and female employment-to-population ratios (1996 and 2006)

Regions	*Female employment-to-population ratio (%)*		*Male employment-to-population ratio (%)*	
	1996	*2006*	*1996*	*2006*
Developed economies and European Union	46.2	49.2	65.8	64.7
Central and Eastern Europe (non-EU) and CIS	46.4	44.9	64.2	62.2
East Asia	68.8	64.2	81.1	78.1
Southeast Asia and the Pacific	55.3	55.0	80.0	77.4
South Asia	34.9	33.8	80.6	78.3
Latin America and the Caribbean	41.5	47.0	76.2	74.3
Middle East and North Africa	20.4	24.5	68.3	69.3
Sub-Saharan Africa	58.7	56.8	79.2	77.4

Source: Global Employment Trends for Women (2007). Global Employment trends for women, brief, March, 2007, International Labour Office. http://www.ilo.org/public/english/employment/stroat/download/getw07.pdf (accessed Jul. 2008).

women at senior levels, particularly in bureaucratic roles in companies/service sectors, could be perceived as a direct threat to patriarchal management structures as well as the production and (re)production of gender inequalities in the workplace. Therefore, such trends in women's employment patterns epitomize a significant transformation in their labor market behavior and positions.

The feminization of the labor force in developing economies differs from that observed in Western countries. For instance, in developing economies, feminization is associated with a process of proletarianization, whereby the majority of women's labor, as discussed above, is absorbed into the informal sector. This is evident in the deregulation of labor markets, a fragmentation of production processes, and the creation of new areas of export orientation that demand casual (less secure), low skill, informal contracts, and ultimately low pay for most female labor. Further, it has been observed that although women's share in wage and salaried work grew from 42.9% in 1996 to 47.9% in 2006, this share still remains smaller for women than for men, and nowhere near to being closed, especially in the world's poorest regions.

Over the last decade, scholars in the West have observed gender segregation in employment and the influence of post structural and postmodern feminist thinking on female participation in wage labor. Contemporary feminist scholarship has seen growing emphasis of the heterogeneous nature of gendered social relations within work organizations, workplaces, and within the institution of the family. These spheres are contemplated as sites of negotiation, interaction, and conflict between social actors. In 1997, McDowell argued that in contemporary society in the West it is highly recognized that in order to conform to a particular set of gender attributes, jobs themselves are constructed socially. However, she also explains that these occupations should not be considered as 'empty slots' that are replenished by workers or as 'gender neutral'. Instead, jobs are formulated and developed to suit either men or women.

Catherine Hakim was one of the first authors to identify and define occupational segregation by sex in the labor market. The way in which both men and women tend to be clustered in different occupations and industry sectors is known as horizontal segregation, while vertical segregation refers to the way in which men dominate the 'high profile' and 'well paid' jobs. For this, feminist economic geographers have developed the framework of 'supply and demand' and 'individual preference' for explaining and analyzing the patterns of gender segregation in the labor market, mainly through human capital theory. Education and training being the basic economic assets of human capital, it was seen as one of the key determinants of economic performance in 1975 by Gary Becker. Human capital implies the collection of skills, qualifications, and experiences possessed by individual workers. This approach relies on the theory that women have less human capital than men; they are paid less because of their low skills, their low investment in education and training, less labor market experience, and motherhood. Importantly, the human capital theorists stressed the role of the individual agency and economic rationality in interpreting gender segregation in employment.

Fundamental criticism stems from the treatment of women as a homogeneous group. From a human capital propensity, Catherine Hakim stresses instead the different investments that women make in their education and career with respect to individual 'preferences', either as 'home centered' or 'work centered'. Hakim explains that women as a group are heterogeneous, diverse, and divided. This way, human capital theory accounts of gender segregation emphasize voluntarism by the significance of individual tastes and choices in explaining gender divisions of labor. However, Hakim urges that a more distinct version of human capital theory is required to explain the patterns of gender segregation. Human capital theory has in turn been criticized for ignoring the structural constraints and systematic gender biases in the valuation of skills, the role of patriarchal gender relations at home and the labor market, and other institutional factors such as employment hierarchies and recruitment practices.

Nevertheless, there is a general consensus that women's employment is segregated both vertically and horizontally. Many feminist geographers have also sought to focus on occupational segregation on the basis of locality studies. In India, for example, the supply of women's wage labor is more complex as a consequence of the way marriage is used as the principal determinant of women's social position. In the West, by contrast, the greatest single determinant of variation in female labor force participation by hours worked and occupation type is the presence of young children. Importantly, the proximity of paid work to the home and the hard constraints of everyday coordination (e.g., chaperoning children to school) play a vital role in circumscribing what opportunities are actually available to women in the way they reconcile paid employment and family reproduction work. Arguably, the causes for women's continued occupational segregation emerge from either within the family or in the home, or through the practices of the workplaces.

Conclusion

The geography of feminist economics has been central to initiate the notion of work and labor. Although since the nineteenth century, women's rights activists have addressed women's exclusion from the economy, it is ironic that even in the twenty first century, women's incorporation into paid employment is spatially uneven, with diverse explanations of women's segregated position in the labor market. Such explanation has served to legitimate and reinforce the existence of a blurred image carrying unacceptable inequality across the globe.

See also: Feminism and Work; Gender in the City.

Further Reading

Allen, J. (2003). *Lost Geographies of Power*. Oxford: Blackwell.
Allen, S. and Walkowitz, C. (1987). *Homeworking Myths and Realities*. London: Macmillan.
Austin, A. E. and Pilat, M. (1990). Tension, stress, and the tapestry of faculty lives. *Academe* 76, 38 42.
Becker, G. S. (1975). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. New York: Columbia University Press, National Bureau of Economic Research.
Bradley, H., Erickson, M., Stephenson, C. and Williams, S. (2000). *Myths at Work*. Malden, MA: Polity Press.
Crompton, R. (1997). *Women and Work in Modern Britain*. Oxford: Oxford University Press.
Dunlop, J. E. and Velkoff, V. A. (1999). Women of the world women and the economy in India. http://www.census.gov/ipc/prod/wid 9802.pdf (accessed Jul. 2008).
Elson, D. and Pearson, R. (1981). Nimble fingers make cheap workers: An analysis of women's employment in Third World export manufacturing. *Feminist Review* 7, 87 107.
Fraser, N. (1989). *Unruly Practices: Power, Discourse and Gender in Contemporary Social Theory*. Minneapolis, MN: University of Minnesota Press.
Garmarnikov, E., Morgan, D., Purvis, J. and Taylorson, D. (eds.) (1983). *Gender, Class and Work*. London: Heinemann.
Gibson Graham, J. K. and Cameron, J. (2003). Feminising the economy: Metaphors, strategies, politics. *Gender, Place and Culture* 10(2), 145 157.
Global Employment Trends for Women (2004). International Labour Office. http://www.ilo.org/public/english/employment/strat/download/trendsw.pdf (accessed Sept. 2008).
Global Employment Trends for Women (2007). Global employment trends for women, brief, March, 2007, International Labour office. http://www.ilo.org/public/english/employment/strat/download/getw07.pdf (accessed Jul. 2008).
Habermas, J. (1989). *The Structural Transformation of the Public Sphere*. Burger, T. (transl.). Cambridge, MA: MIT Press.
Hakim, C. (1979). *Occupational Segregation*. London: Department of Employment.
Hakim, C. (2000). *Work Lifestyle Choices in the 21st Century Preference Theory*. Oxford: Oxford University Press.
Hanson, S. (2007). People and place in economic geography: Learning from women's entrepreneurship, plenary lecture. *Second Global Conference on Economic Geography.* Beijing, China.
Hanson, S. and Pratt, G. (1995). *Gender, Work and Space*. London: Routledge.
Harvey, D. (1969). *Explanation in Geography*. London: Arnold.
International Labour Organization (2002). Women and men in the informal economy: A statistical picture, International Labour Organization. http://www.ilo.org/public/english/employment/gems/download/women.pdf (accessed Jul. 2008).
Ironmonger, D. (1996). Counting ouputs, capital inputs and caring labor: Estimating gross household product. *Feminist Economics* 2(3), 37 64.
Jarvis, H. (1999). The tangled webs we weave: Household strategies to co ordinate home and work. *Work, Employment and Society* 13(2), 225 247.
Latour, B. and Weibel, P. (eds.) (2005). *Making Things Public: Atmospheres of Democracy.* Cambridge, MA: MIT Press.
Loutfi, M. F. (eds.) (2001). *What Is Equality and How Do We Get There? Women, Gender and Work*. Washnigton, DC: International Labour Office.
Massey, D. (1984). *Spatial Divisions of Labour: Social Structures and the Geography of Production*. London: Methuen.
McDowell, L. (1997). *Capital Culture: Gender at Work in the City.* Oxford: Blackwell Publishers.
Menon Sen, K. and Shiva Kumar, A. K. Women in India: How free? how equal? United Nations India. http://www.un.org.in/wii.htm (accessed Sept. 2008).
Momsen, J. (1991). *Women and Development in the Third World*. London: Routledge.
Rendall, J. (1990). *Women in an Industrializing Society: England 1750 1880*. Oxford: Blackwell.
Reskin, B. and Padavic, I. (1994). *Women and Men at Work*. London: Pine Forge Press.
Rowlands, J. (1995). Empowerment examined. *Development in Practice* 5(2), 101 107.
Rubery, J., Smith, M. and Fagan, C. (1999). *Women's Employment in Europe: Trends and Prospects*. London: Routledge.
Smith, N. (1984). *Uneven Development: Nature, Capital and Production of Space*. Oxford: Blackwell.
United Nations (2003). Putting gender mainstreaming into practice, Economic and Social Commission for Asia and the Pacific, United Nations. http://www.unescap.org/esid/GAD/Publication/Gender Mainstreaming.PDF (accessed Jul. 2008).
Walby, S. (1990). *Theorizing Patriarchy*. Oxford: Blackwell Publishers.
Walby, S. (1997). *Gender Transformations*. New York: Routledge.
Waring, M. (1988). *Counting for Nothing: What Men Value and What Women Are Worth*. Auckland: Allen and Unwin.

Relevant Websites

http://www wds.worldbank.org
Equity and Development, World Development Report 2006.

http://hdr.undp.org
Gender and Human Development, Human Development Report 1995.

http://siteresources.worldbank.org
Gender working towards greater, IDA (International Development Association at Work).

http://siteresources.worldbank.org
Women's work and movement into the public sphere.

Fertility

Y. Underhill-Sem, University of Auckland, Auckland, New Zealand

Glossary

Below-Replacement Fertility Low levels of fertility where the people are not producing sufficient numbers of offspring to replace the numbers who are dying.

Fertility The ability to conceive and give birth.

Malthusian-Inspired Theories Arguments that begin from the basis that populations tend to increase faster than the means of livelihoods to sustain them and therefore 'preventative' and 'positive' checks are needed. Examples of preventative checks are moral restraint, postponed marriage, and birth control. Examples of positive checks are war, disease, and poverty.

Reproduction The various ways in which people produce offspring.

Introduction

Fertility is understood most widely in geography as one of the three determinants of population change alongside migration and mortality. Everyone is born, everyone moves geographically in some manner, and everyone eventually dies. These self evident understandings however belie the importance of the conceptual framing of these processes, and none more so than the notion of fertility. A fertile concept, idea, or process is one that promotes further enquiry, knowledge, and understanding. Yet, within geography, connecting these understandings of fertility through research and analysis is neglected. Instead, fertility is incorporated in a routine uncritical way with the result that geographic understandings of fertility have been a taken for granted and under researched area of study in geography. A partial reason for this is that within geography there has been a pre occupation with migration as the main demographic process that distributes people spatially. This, however, overlooks the vital fact that one needs to be born to be able to first move and this fact is something that occurs in diverse ways and with far more problematic con sequences for all involved. A more complete under standing of fertility involves inquiry into a range of other areas such as reproductive practices, sexualities and embodiment, and notions of family.

As with much of population geography, the analysis of fertility needs to move beyond method and concern over technique and toward being able to articulate new visions of alternatives and just futures. Advancing analysis from the classic fertility narratives and statistical analysis of survey and census data will allow for new ways of seeing fertility that will contribute to debates in both within the discipline and also between other disciplines. For in stance, when the global mobility of domestic workers into Singapore is also understood not just as groups of women moving to care and nurture the aged and young in dis tance places, but as sexualized subjects who carry with them particular fertility potentials which change their childbearing possibilities, then it is also possible to look for new fertility patterns and their implications. For in stance, in what ways is their fertility controlled in new work places? In what ways might it appeal to have chil dren born in places where they might have new possi bilities for citizenship? And in what ways might new informal sexual relations in different places affect formal sexual relations in home countries? Similarly, with new reproductive technologies available in low fertility countries, such as Europe, the social relations between the different biological contributors to a particular pregnancy are increasingly complex and variable. In what ways are new configurations of parents affecting under standings of households and families? In what ways does the particular scale of new reproductive technologies affect analysis of small scale population dynamics? And in what ways does this produce new options for citizen ship and inheritance?

Mainstream Understandings of Fertility and Reproduction

Fertility routinely refers to the way in which groups of people produce offspring. Some demographers argue that the term refers to a more refined analysis of 'natality' but because it is so common they prefer the less committed term 'natality'. This framing has not, however, flowed into other disciplines like geography or anthropology that also appropriate demographic analysis. Universal assumptions about biology mean that similar measures of fertility are used in nonhuman populations such as plants and animals. In human populations there are many measures of fertility and each one attempts, at increasing levels of 'accuracy', to represent the rate at which a population is growing by relating the number of live births to a specific population and time period. Unfortunately, even identifying a live birth is difficult, let alone how many there are.

The simplest is the 'crude birthrate' which represents the number of live births in a given year divided by the number of people in that population in the middle of the year. Slightly more refined and commonly used measures include 'general fertility rates' (the number of births divided by the number of women of childbearing age), 'age specific birthrates' (the number of births divided by the number of women in a given age group), and 'total fertility rates' (the sum of age specific birthrates over all ages of childbearing). More sophisticated measures include 'completed fertility rates' which show the fertility progress of a cohort through the childbearing years and 'age specific birth probabilities'. The former provides some insights into historical shifts in the timing of childbearing and the latter provides for the calculation of population projections.

The variability of fertility among subgroups of a population means the more sophisticated or refined measures of fertility specific to age, marriage, and other social and temporal measures are increasingly important. Further, because of the total absence or unreliability of records about birthdates or total populations in many parts of the world – especially in the demographically largest and fastest growing parts of the economic South such as sub Saharan Africa and China – continual efforts are made to develop the best techniques to estimate fertility rates. Despite continued and often justified skepticism as to the accuracy of fertility rates, they are still ubiquitously used and are relevant for comparative accounts of progress toward social equity. All fertility rates have time and spatial dimensions and should be used carefully, with due regard to the assumptions and methods of data manipulation, the comparative analysis of fertility rates provides a useful representation of change. For this reason it is useful to calculate fertility rates, refining them as best one can, but keeping in mind that they are often a crude tool for analyzing more complex processes.

Closely related to the notion of fertility is the term 'reproduction' or the extent to which a group of people are replacing themselves by natural processes. The term 'reproduction' has a wider scope than the term 'fertility' and is defined demographically as the act or process of producing offspring, giving birth, bringing forth offspring, producing, yielding, making, manufacturing, working, and/or creating unity. Used in demography and population studies, including population geography, this term is often limited to the biological event of producing babies. It clearly has a wider meaning linked to the production of human beings and is the focus of a range of social sciences, especially those interested in class differentiation. The main problem with this concept is that like 'fertility', biology tends to be treated as an uncontested universal that takes agency away from people, especially women.

Central to the shortcomings of these mainstream conceptualizations of fertility and reproduction is the unproblematic treatment of the body as exclusively a biological given. Bodies are now accepted as being both biologically configured and also socially and culturally constituted by places and people. This provides for nuanced analysis of the complex layering of meanings how groups and individuals produce offspring.

Reconceptualizing Fertility: Of Epistemologies, Bodies, and Desires

Interest in fertility within demography is matched by a similar long standing interest within anthropology in documenting ethnographic accounts of childbirth in different places. However, in the process of documenting the many different cultural interpretations of the 'universal' process of birth, it becomes clear that childbirth is as different as the overlapping and frequently contradictory discourses that are used to describe it. For example, in the English speaking world, it is not unusual to hear fathers to be saying 'we are pregnant'. Clearly, there are solid grounds for reconceptualizing another epistemological position from which to interrogate notions of fertility. Population geographers have only recently begun to consider the implications of an epistemology in which demographic events and processes are socially constructed and therefore partial. Following the work of feminist anthropologists and demographers, population geographers need to consider how 'places' construct bodies so that they are not simply understood as a biological given. To do this requires a research process that combines critical ethnographies and geographies with statistical analysis.

There is a potent sharpness of an epistemology based on feminist theorizing of the body because it requires the precise analysis of the 'marking' of unmarked' categories. Thus, fertility analysis would require consideration of what precise body is being represented in mainstream fertility analysis. An embodied analysis would give corporeal form to the otherwise abstract definitions and by doing so would subvert taken for granted understandings whose privileging has suppressed other ways of knowing. This is likely to include other discursively constructed realities such as the existence of 'specialists', 'sorcerers', 'deceased ancestors', and even God. Placing the analysis of fertility within specific historical places gives explanatory preference to the complex interplay of structural features like the dynamics of social, cultural, and political issues. Thus, a study of fertility would not simply focus on the event of the birth of a child but would also examine how the mother and the others are involved in the talk about the child, the pregnancy, and the birth. In this way, fertility can be understood as relating to not only bodies as biological constants but also bodies situated in historical processes in particular places. This requires looking beyond the 'total fertility rates' by also examining the discursive processes that constitute the body. This approach opens up the possibilities of looking at the discursive practices that constitute other fertile bodies such as, for instance, the bodies of trafficked women, of sexually active people who live with physical impairments, disabilities or chronic illness, or women living in isolated villages in the economic South.

The intersections and interrelationships of material and discursive fields in relation to bodies and places are critical places from where the analysis of fertility can expand. Even when one begins to capture the workings of communities and family groups, enmeshed in the overall agreements as to the aim of particular fertility outcomes, there will be different gendered versions. For instance, the reasons for women wanting children and especially sons in some places is not so much to maintain lines of male descent but also to provide comfort and loyalty in one's daily life. These fertility desires and sentiments last as long as the mothers and caregivers do, but they play an important part in understanding the otherwise contradictory fertility practices

Repoliticizing Fertility: Abortion, International Conference on Population and Development (ICPD), and Human Immunodeficiency Virus (HIV)/Acquired Immune Deficiency Syndrome (AIDS)

Fertility studies in geography are an important route into working with contemporary social justice issues around women's sexual and reproductive health and rights. Abortion remains a controversial and complex issue that incorporates moral positioning and political irresponsibility in ways that threaten overall development in the economic South. In some places, such as the United States and the Caribbean, this is combined with increasing use of the morning after pill, 'virgin clubs' (i.e., young people committed to saving sex for marriage), and the return of the big white wedding to recast societal norms around families and households. In other places, such as the Sudan, and Central African Republic, death and chronic illness as a result of continually unwanted pregnancies – often in places of civil war and disasters with the subsequent displacement of domestic life – also have the effect on shifting social norms around families and households. So too has the effect of HIV/AIDS with a growing number of deaths to young adults and especially women threatening to radically alter notions of family and household because different members of the household become the ones whose fertility is most sought after – especially young uninfected girls. Combined with culturally abhorrent practices such as female genital mutilation, sex work, and the cross country adoption of the girl child, the examination of fertility from a geographic perspective is sorely wanting.

On the global policy arena, the ICPD in Cairo in 1994 was a huge shift away from the demographic targets of population processes toward upholding sexual and reproductive rights. But the decennial international population conferences reflected more than this. These conferences (Bucharest in 1974, Mexico in 1984, and Cairo in 1994) had all concerned themselves with the fertility of women – especially from the economic South. Over time the debates moved from Malthusian inspired theories that targeted family planning in 1974 to neo Marxist inspired theories of economic development in 1984. However, in 1994, the relationship between demographic theory and population policy, especially as it concerned women's fertility, witnessed another ideological and political shift. The more explicit gender focus of the Cairo Program of Action argues that the empowerment of women is the way to reduce gender and other social inequalities. Few fertility studies within geography pick up this framework with an apparent comfort in sticking to conservative gender agendas and continuing to work on the science rather than the politics of fertility research. This reluctance will stifle the recognition by policymakers of geographic thinking as providing useful research based policy.

While few geographers have actively engaged with the ICPD process, many more have taken on issues concerning HIV/AIDS and new reproductive technologies. Closely linked to patterns of mortality, the analysis of HIV/AIDS provides fresh insights into the ways in which fertility intersects with the mortality of sexually active populations. Reduced fertility due to shifts in traditional partnering patterns, biochemical reactions from medication, and changing sociocultural values over sexual reproduction (e.g., polygamy, slavery, bride stealing, and sexual trafficking) are issues that are experienced and embodied in very diverse places and hence provide huge possibilities for new approaches to fertility analysis.

Conclusion

The traditional assumption that groups of people are best understood when they are studied as collective outcomes of individual population actions is no longer the way to analyze fertility in geography. New approaches to analyzing fertility might begin by mapping or recording patterns of fertility using standard demographic techniques. However, other texts need to be incorporated into the analysis. This is not just because it is highly likely that standard techniques cannot capture complex human processes, but also because there is greater access to new texts. If an equivalent investment was directed into supporting research using these new texts as has been in supporting the analysis of census and vital statistics, the future looks bright for an invigorated geography of fertility. Drawing on interdisciplinary insights in, for in stance, anthropology, history, and literature, will ensure researchers are fully cognizant of the limitations of par ticular texts or measurements. Patterns of fertility are not just the result of biological processes, but they also cap ture more complex and interesting social, cultural, and political processes that create subjects such as adopted children, orphans of war and HIV/AIDS, and children conceived by *in vitro* fertilization. These new subjects do not just appear but are created through differences in time and place; so we can expect many more new subjects to emerge, and much more, therefore, for geographers of fertility to think about.

See also: Census Geography; Migration; Pregnancy and Childbirth.

Further Reading

Bailey, A. (2005). *Making Population Geography*. London: Hodder Educational.

Graham, E. (2000). What kind of theory for what kind of population geography? *International Journal of Population Geography* 6, 257 272.

Greenhalgh, S. (1996). *Situating Fertility*. Cambridge: Cambridge University Press.

Jordan, B. (1993). *Birth in Four Cultures*. Prospect Heights: Waveland Press.

Neumayer, E. (2006). An empirical test of a neo Malthusian theory of fertility change. *Population and Environment* 27(4), 327 336.

Palmore, J. A. and Gardener, R. W. (1994). *Measuring Mortality, Fertility and Natural Increase: A Self Teaching Guide to Elementary Measures*. Honolulu: East West Population Institute.

Ram, K. and Jolly, M. (1998). *Maternities and Modernities*. Cambridge: Cambridge University Press.

Riley, N. and McCarthy, J. (2003). *Demography in the Age of the Postmodern*. Cambridge: Cambridge University Press.

Shyrock, H. S. and Siegel, J. S. (1976). *The Methods and Materials of Demography* (condensed edition by Stockwell, E. G.). New York: Academic Press.

Sporton, D. (1999). Mixing methods in fertility research. *Professional Geographer* 51, 68 76.

Underhill Sem, Y. (2001). Maternities in out of the way places: Epistemological possibilities for retheorising population geography. *International Journal of Population Geography* 7, 447 460.

Festival and Spectacle

M. Duffy, University of Melbourne, Carlton, VIC, Australia

Glossary

Community A social network of interacting individuals, usually concentrated in a defined location, although this definition is challenged by the rise in use of telecommunications and cyberspace.

Event Tourism The systematic planning, development, and marketing of festivals and special events as tourist attractions, catalysts, and images builders.

Post-Structuralism Movement of thought in various fields (e.g., literary criticism, cultural studies, sociology, psychoanalysis) which developed out of, and was a response to, French structuralist thought of the 1950s and 1960s. Post-structuralism envisages a 'subject-in-process', arising within a space that is plural, multiple, and constantly destabilized, and where identity can no longer be defined in 'essentialist' terms such as class or gender.

Social Capital Those characteristics of social structure or social relations that facilitate collaborative action and, as a result, enhance economic performance.

Overview of the History of Festivals and Spectacles

The structure and role of festivals and spectacles in the contemporary world have their origins in religious celebrations and public cultural practices. Many such religious events continue today, attracting both devotees maintaining centuries old traditions, as well as tourists seeking to experience some notion of an authentic 'other'. These historical forms are important in creating and maintaining communal bonds to particular places, re inforcing social hegemony through symbolic practices, and controlling potential aggression or challenges to authority by providing an emotional outlet. Even so, any such public displays of conformity may hide transgressive and subversive messages, as the activities of carnival demonstrate.

Origins in Pagan and Religious Events

Festival and spectacle have been conceptualized in social anthropology as events apart from that of the secular and everyday. Such events can be traced as a derivation or continuation of pagan and religious celebrations and rituals, with an emphasis on the struggles between the sacred and profane realms of society.

Historical antecedents of contemporary events are found in religious pilgrimage, the feast days of saints, and carnival. These events are explicitly geographic activities; connecting places through the devotees' or pilgrims' pathways across regions, countries, and continents, and also in the ways in which specific cultural landscapes are constructed around certain places and activities.

Origins in Traditional Cultural Practices (Folk, Rural, Urban)

Historically, traditional cultural practices were connected to religious or pagan celebrations. Concepts of time and place as part of a cycle of destruction and regeneration ensured that particular activities relating to community survival, such as planting and harvesting, were under taken correctly. The importance of a fertile earth in pagan communities also led to elaborate practices asso ciated with certain places, dates, and even hours of the day, for example, celebrations around the summer and winter equinoxes. Remnants of such fertility rites can be discerned in some contemporary carnival activities.

Public Space and Public Culture

Public space is a highly contested space. The social and spatial relations that occur within this sphere serve to constitute and regulate identity through categories such as race, class, gender, and sexuality. Although public space may be an arena of open debate and social inclu sion, it is also a site of social exclusion. Yet festivals may also provide opportunities for addressing problems em bedded within a community's social relationships, either through official broadcasting of issues, or through un official challenges and subversive activities.

Liminal or everyday practices?

Festivals and spectacles formalize space, time, and be havior in ways that distinguish these events from every day events. Although strongly connected to the rhythms of daily life, they bring groups and communities together in order to mark out particular religious, sociopolitical, and historical affiliations. In this way the festival supports notions of boundedness to particular place(s) and people, internal coherence, and the relative unity of the group. These notions of coherence are created through a heightened sense of belonging brought about by rituals and associated symbolic practices. Thus, these events are an important means of framing and representing the relationship between people and place, communicating relatively coherent ideas about social order, albeit one of acceptance or challenge.

Ritual and Symbolism

Ritual may be understood very broadly as a set of formal behaviors within a specific context of time and place, outside that of the everyday. Understanding the purpose of ritual ranges from Durkheimian notions of binding people together; the Freudian psychoanalytic position that such practices are symptomatic of unresolved neuroses; or that ritual practices express the symbolic meanings embedded within social and cultural structures. Festival conceived as a ritual practice thus simultaneously maintains traditions while allowing the possibility of change and difference.

Liminality and communitas

Social anthropology has provided two concepts – those of 'communitas' and 'liminality' – significant to understanding the role of the festival and spectacle in historical and contemporary life.

'Communitas' refers to a sense of social interrelatedness – a spontaneous feeling of connectedness and belonging that arises out of involvement in a communal event. These strong, yet often transient, feelings of attachment to others can operate across different social vectors such as class, ethnicity, or gender, important in constituting strong social bonds.

Important to the processes of sociality are an event's associated rituals, which can lead participants to a state outside that of the everyday, of in betweenness or ambiguity, to a state of liminality. The liminal period temporarily suspends conventional social rules, so opening up a space of potential change, exemplified in anthropological literature by ceremonies of initiation. Liminal events are understood to offer some contact with the sacred or divine, and hence facilitate the individual's transformation into a full participant of the community or group. Some argue that truly liminal events can only occur within small scale, integrated, or indigenous societies. However, contemporary performative events such as the modern Olympics are understood by some researchers as deliberate responses to contemporary feelings of alienation and instability. The ritual practices and pageantry of these large scale events offer the possibility of a reconnection to the profound, and so a re emergence of what has been called 'neoliminalty'.

Spectacle and parade

The high degree of visual display and theatricality of spectacle draws on pageantry and fanfare as a means to create a sense of belonging. As Debord argues, spectacle is an illusion, a collection of images that mediates social relationships between people, so bringing about a sense of unification. Feelings of belonging to place may be constructed through positive feelings of celebration or the negative emotions of fear. In either case, spectacle is an effective means of social control, and the social elite who control such events exert a strong influence in the social realm and within popular consciousness.

Parades are a popular and effective means of constructing spectacle and its ideological intent (**Figure 1**). Geographic analysis has demonstrated how the use of particular motifs, such as iconic or symbolic images of national belonging, help constitute a ritual space and 'imagined community' of belonging. Moreover, studies of the movement of bodies and parade motifs show how public space is territorialized, thus reiterating or realigning public space within particular notions of state and community. However, spectacles and parades are also a means to challenge or subvert such hegemonic practices, their symbolic and ritual aspects intimately linked in the political, social, and economic processes of community and place.

Carnival

Strongly associated with Lent in the Catholic Church's calendar, carnival (or more commonly Mardi Gras) is a secular festivity, with some anthropologists arguing that carnival is perhaps the oldest of all Western pagan festivals still observed. Carnival permits the inversion of normal societal order, allowing the expression of powerful emotions and impulses normally repressed in everyday life. Bahktin argues that the ambivalent nature of carnival – its simultaneous use of abuse and praise, its images of death and birth – are fundamental to continuity and renewal. Carnival is, therefore, understood as a paradoxical social process, both reinforcing participants' sense of their place within society (in classical Marxist class struggle terms), but also a means of reaffirming fragile social bonds, dramatizing and reinforcing collective ideas about society, as well as being a release for repressed feelings of frustration. What is significant is that carnival and the carnivalesque activities that take place in the public arena are the community's own representations and performances of common goals and shared values, rather than a framework imposed on the community by its civil and religious leaders.

Festivals and Spectacles as Community Building Processes

Concepts of 'place', 'community', and 'belonging' have been addressed extensively in geographical and sociological literature on festivals. One of the focuses in geographical studies has been on the underlying assumption that festivals function as community building activities, be that of a community with a long established connection to a particular place, one that has had a more

Figure 1 Parade, Swiss Italian Festa, Daylesford (Australia) 2006. During the period of the festival, the space of the event is marked out through visual and aural signs, for example, in the use of flags, banners, masks, and certain genres and styles of music. (Photograph by Michelle Duffy.)

recent migration and settlement, or one that differs in some way from the dominant group. Such events are a means to express and celebrate core values of a particular community, with representations of community identity mediated through distinctive cultural artifacts and activities, such as music, food, and folk dancing.

Creating notions of belonging

Underlying the focus on a celebration of community is the view that social cohesion is the event's necessary goal. The official discourses of those controlling the festival are used to produce an official 'imagined' community, which the festival is then planned to address. Celebrating days of national importance, parades for national or sporting heroes, festivals that celebrate the culture and history of particular ethnic groups – all such activities respond to the politics and social dynamics of belonging. As recent cultural geographic studies demonstrate, the framework of the festival transforms disparate individuals into collective abstractions – community, nation, even race – through engagement of the senses, by inducing action, and giving access to knowledge and experience of these public events.

Drawing on Bourdieu's notion of taste as a marker of social position, feelings of belonging or alienation are articulated through esthetic discourses which differentiate individuals and social groups through practices of consumption. In turn, cultural capital confers status on an individual who recognizes and consumes particular esthetic products, such as art, music, literature, theater, or even sport.

Yet the transitory nature of the event – the brief encounters and exchanges occurring within the festival space – produces different and often conflicting configurations of identity, place, and belonging. Festivals are, therefore, complex sites for localness and belonging, celebrating diversity and connections beyond that of the locally defined community. One example of this paradoxical creation of community and belonging is that of the multicultural festival. The goal of such a festival is to create a site in which potentially divisive differences are reduced and feelings of belonging are heightened. Not surprisingly, tensions arise where certain ideals of community diverge from those of the everyday, heterogenous lived experiences of people. Moreover, once entered into the public realm, events such as spectacle, carnival, and festival become part of the public culture, and those who have organized the event may lose control over what then happens in this domain. This leads to a paradox; these events function as a form of social integration and cohesion while simultaneously they are sites of subversion and protest. This paradox is the source of such an event's sociospatial and political significance.

The roles of emotion, embodiment, and cognition in community and place

With the advent of post structuralism in human geography, much attention has been given to the project of mapping the subject, and, more recently, the affective and embodied processes that constitute subjectivity in ways that lie outside of language. These affective and cognitive processes are significant to feelings of being 'in' or 'out' of place. Festivals and spectacles are structured so as to produce certain responses which then give rise to feelings of belonging. Our emotional responses to such events therefore help foster social capital in that we are encouraged to develop strong emotional ties to other participants, to an imagined community and its connection to ideas of place. In this way, festivals are important in social strategies seeking to address the growing sense of alienation and insecurity felt in public space, or in places with diverse resident groups.

Recent studies in cultural geography on the role of music have been an important means of accessing and examining our emotional and affective responses and the part they play in the constitution of identity and community. Music is an important representational practice, with lyrics, instrumentation, and style evoking images of particular places and identities (**Figure 2**). Music is also an experiential process, and this opens up the possibilities of understanding the role of immediacy and presence in creating social and spatial relationships. Place, identity, and belonging are constituted in the very acts of people engaging with the musical performance. However, music has the potential to reconfigure the festival as a site of subversion, playing against the meanings ascribed by the festival's official framework.

Figure 2 Street scene, Brunswick Music Festival (Australia). While entertaining participants of this street party, the resident performers of the local Spanish restaurant create a 'Spanish' soundscape through song, instrumentation, dance, and costume. (Photograph by Michelle Duffy.)

Festivals as Subversive Practices

The community building processes of festivals, and the very powerful symbolic forms used to mark out relationships between identity and place, also make festivals contested sites of identity, community, and belonging. Different sections of society utilize the festival structure so as to present, constitute, and celebrate particular values, lifestyle choices, and identities. Participating in festival activities enables the participant to reaffirm membership to a particular group. The festival therefore provides a means through which people can advocate or contest certain notions of identity and ideology.

Particular festivals (pop festivals such as Glastonbury, or 1990s technodance parties) may raise issues around notions of community control, perceptions of public nuisance, and environmental damage because of assumptions about the genre of music performed and its assumed audience. On the other hand, strong opposition and a sense of moral panic with regards to such festivals reflect a wider privileging or exclusion of certain social groups, such as youth. The use of cultural and performative strategies within festivals therefore challenges notions of identity, community, and belonging, continuing the traditions of inversion exhibited in carnival.

Geographic research has explored the ways in which the festival subverts the everyday functions of public space. The use of symbols such as banners, posters, and clothing marks out the site of the event, permitting certain forms of behavior that would otherwise be unacceptable, such as demonstration of homosexual affection, overtly political shows, or simply talking to strangers (see **Figure 3**). Yet this also raises contentious issues over definitions of the public and private realm, and how particular identities may come under closer surveillance.

Figure 3 ChillOut, Daylesford (Australia). The main street of a rural country town is remade and reclaimed through particular images associated with queer culture. (Photograph by Michelle Duffy.)

Yet, activities once understood as subversive can also become part of the mainstream. One such example is the Sydney Gay and Lesbian Mardi Gras, originally a march in support of gay rights and a challenge to the violent homophobia of 1970s inner urban Sydney. Continuing to support and promote the gay and lesbian community, this event is now a very successful tourism event, not simply in economic terms – for example, in 1998 the event was estimated to have generated AU$99 million (although it has recently encountered financial difficulties) – but, as some argue, in a growing acceptance of homosexuality. Fringe festivals are another form of response to mainstream cultural practices. Deliberately popular events, in contrast to the arts festivals they are often paired with in the festival calendar, the fringe festival, such as Canada's Edmonton and Winnipeg Fringe Festivals, reclaims public space for forms of public culture. These subversive practices reorganize the rationalized spaces and activities of the everyday, providing new opportunities for increased interaction between individuals, and so a renegotiation of social identities and everyday practices.

Economic Underpinnings

There has been a dramatic increase in the number of festival events in the last decade. One key reason for this is that festivals are recognized as a means to make a unique and valuable contribution to a community's culture, attracting local, national, and even international audiences. Festivals can be important sources of revenue for communities, and have therefore become key components of economic development strategies for cities, rural towns and regions, and even countries.

Festivals and Tourism

Recent geographical studies have explored how the festival has become an increasingly important strategy for both rural and urban regions as a means to initiate economic renewal. Festivals help create and promote particular cultural landscapes that are then readily incorporated into the tourism industry. Festivals help create and consolidate certain place images and their associations with leisure and educational activities (as places of adventure, or relaxation, or heritage, and so on) or with industries closely associated or even developed for that specific place (gourmet trails and vineyards have become popular contexts in which an annual festival event may be located). Effective tourism promotion through these events can increase the number of visitors to a particular locality, and the revenue generated is very important to the local economy. An increase in numbers attending a festival can provide business opportunities for the locale, showcasing the expertise of the host community, and thus generate interest in potential investment.

Yet the popularity of the festival as an economic savior, particularly for regional areas, has in some instances meant a lessening of connection between event and place as entrepreneurs, local businesses, and communities seek to tap into this bourgeoning market. While festivals do construct and reinforce the reputation of the locale, as well as anchoring tourism economies and associated industries in place, the resources and any profit are not always effectively redistributed towards the smaller centers. Moreover, such events can exacerbate the ways in which people and resources are concentrated within urban centers at the expense of more peripheral locations. Although hallmark and mega events hosted in urban locations associated with sporting activities, such as the Olympics or the FIFA World Cup, are understood as important revenue sources, most are not profitable. Their presence can also increase pressure on urban infrastructure and resources as the population increases for the period of the event.

Place Marketing

Festival and special events can be important motivators of travel, not only for the event itself but also in that it contributes to the place's range of tourist attractions, facilitates media coverage for the destination, and promotes awareness of the destination for future visits. Government and nongovernment institutions and organizations draw on a range of factors – cultural planning, leisure, lifestyle sampling, the so called creative industries – to influence the branding of a particular place as a desirable destination for investment and expenditure. Event tourism, and increasingly the importance of the festival within this, has become a significant means of assisting such strategies of economic development and place regeneration. Mega events, such as the Venice Biennale or the arts festivals associated with the summer Olympics or Commonwealth Games, serve to construct and reinforce the reputation of certain places as significant sites within global economic and cultural networks.

The role of place branding is very important. Per ceptions play a significant role in the tourist/consumer's attitude and use of, and hence the value attributed to, place. Festivals commodify traditional narratives, tradi tions, and the presentation of cultural products such as music, dance, pageantry, visual art, and theater, in order to construct and promote the uniqueness of a place or community. In some instances, such as the Edinburgh Festival, the festival itself becomes a tourist destination in its own right.

Impacts on the Sociocultural and Natural Environments

The rise of new tourism – a particularly Western phe nomenon and including a diverse range of forms such as ecotourism, green tourism, responsible tourism, heritage tourism, and cultural tourism – corresponds to transfor mation within a Western society characterized by heightened levels of anxiety. In this secularised society, Nature has become the repository of wholesomeness, valued for its implied critique of contemporary consumer society. Festivals are often promoted as a more benign component of this new tourism, with limited impact on the sociocultural or natural environment. For example, the Woodford Folk Festival, held outside of Brisbane (Queensland, Australia), includes within its program scheduled sessions and workshops about the environment and sustainability, as well as activities such as tree planting, as a means to counter its negative environ mental impact. However, festivals do have considerable impact on the environment. They often involve a con siderable number of people located in a small geo graphical location for a relatively short period of time, and success is equated with maximizing the number of visitors with the assumption that this will generate higher revenue. This in turn raises other related issues such as traffic control and congestion, noise, crowding, and dis ruption to everyday life.

Recent studies in tourism have sought to examine and determine the impact of festivals on the environment, measuring such factors as the stress on endangered spe cies, key resource consumption, environmental planning, and reviewing the effectiveness of environment pro tection measures. In response to these issues and con cerns, calls for more sustainable forms of tourism practice have also been directed at festival organization and management.

Impact on Culture

Besides being important in terms of economic benefit for communities, the festival is also significant in ways that are seen to add to the value of everyday life, such as in the promotion of community participation, the en hancement of community creativity, and the fostering of community well being. These factors are important in helping maintain cultural networks and traditions given the concerns about the globalization of culture, while opening up links with the broader community.

The search for authenticity: Memory, history, and nostalgia

In terms of heritage tourism, cultural or cultural heritage festivals have become a major focus, and notions of au thenticity are intimately intertwined with such a festival's success. Tourists seek an authentic experience, and in the framework of authenticity, the festival and its associated cultural products are defined as authentic or inauthentic depending upon whether these are made or performed by local people according to traditional practices.

Yet festivals framed around notions of ethnicity or culture, as well as the so called multicultural festivals, are often criticized for their superficial dealings with the concepts of identity, culture, and diversity. It has been argued that these festivals often present rather limited views of community and identity, and the emphasis on costume, food, and music is only shallow representations of the complexities of cultures. Such festivals are also criticized because they seem to ignore practical appli cations to issues of social justice, such as economic se curity, access, or equity.

Health and well-being of communities

Festivals are important in that they can demonstrate how symbolic notions of social justice are translated into the material and social spaces of the residents of local council. They have come to be a popular form of showcasing local government and other nongovernmental organization responses to the need to instil a sense of social cohesion, given the culturally diverse communities that characterize much of contemporary life. On the other hand, such festivals also provide opportunities for various minority groups to express and even challenge their difference to the mainstream.

Festivals are also important sites of cultural recovery and education, not only within the specific ethnic or cultural community, but also to noncommunity members. The importance of this more specifically intangible heritage is recognized in UNESCO's commitment to assisting communities with the maintenance of their cultural practices. Other foundations and organizations also recognize the need to help maintain cultural prac tices, and the benefits of a festival framework with which to do this. For example, the Yothu Yindi Foundation supports the annual Garma Festival of Traditional Culture, a festival celebrating the culture of the Yolngu (the Aboriginal people of northeast Arnhem Land, Australia) that is also a forum in which Australian Indigenous and non Indigenous people may come together for cross cultural learning.

Future Directions in Research

Current human geographical research into festivals has drawn on a range of theoretical and methodological approaches, including psychoanalysis, nonrepresentational theory, and geographic information system (GIS). Productive areas of future study draw on these and other quantitative and qualitative methods so as to examine such diverse issues as belonging, exclusion, migration, economic regeneration, as well as a more nuanced analysis of community creation.

See also: Affect; Anthropology and Human Geography; Belonging; Community; Cultural Capital; Difference/ Politics of Difference; Economic Development, Rural; Emotional Geographies; Emotional Knowing; Environmental Policy; Identity Politics; Performative and Embodied Mapping; Place, Politics of; Popular Culture; Public Space; Public Spaces, Urban; Queer Theory/ Queer Geographies; Race; Religion/Spirituality/Faith; Representation, Politics of; Resource and Environmental Economics; Sexuality; Sound and Music; Soundscapes; Tourism; Tourism, Rural; Tourism, Urban; Youth/Youth Cultures.

Further Reading

Bahktin, M. (1984). *Rabelais and his world,* Isowlsky, H. (trans.). Bloomington: Indiana University Press.
Debord, G. (1995). *The society of the spectacle*. New York: Zone Books.
Getz, D. (1991). *Festivals, special events, and tourism*. New York: van Nostrand Reinhold.
Gibson, C. and Connell, J. (2005). *Music and tourism*. Clevedon: Channel View Publications.
Jackson, P. (1988). Street life: The politics of carnival. *Environment and Planning D: Society and Space* 6, 213 227.
Jago, L. and Shaw, R. (1998). Special events: A conceptual and differential framework. *Festival management and Event Tourism* 5(1 2), 21 32.
Kong, L. and Yeoh, B. (1997). The construction of national identity through the production of ritual and spectacle. *Political Geography* 16(3), 213 239.
Ley, D. and Olds, K. (1988). Landscape as spectacle: World's fairs and the culture of heroic consumption. *Environment and Planning D: Society and Space* 6, 191 212.
MacAloon, J. (1984). Olympic Games and the theory of spectacle in modern societies. In MacAloon, J. (ed.) *Rite, drama, festival, spectacle: Rehearsals toward a theory of cultural performance*. Philadelphia: Institute for the Study of Human Issues, 248 280.
MacCannell, D. (1979). Staged authenticity: arrangements of social space in visitor settings. *American Journal of Sociology* 79(3), 589 603.
Marston, S. (1989). Public rituals and community power: St Patrick's parades in Lowell, Massachusetts 1841 1874. *Political Geography Quarterly* 8(3), 255 269.
Stallybrass, P. and White, A. (1993). Bourgeois hysteria and the carnivalesque. In During, S. (ed.) *The cultural studies reader*, pp 382 390. London: Routledge.
Thrift, N. (2003). *Performance and … Environment and Planning A* 35, 2019 2024.
Turner, V. (1984). Liminality and the performative genres. In MacAloon, J. (ed.) *Rite, drama, festival, spectacle: rehearsals toward a theory of cultural performance,* pp 19 41. Philadelphia: Institute for the Study of Human Issues.
Zukin, S. (1995). *The cultures of cities*. Cambridge: Blackwell.
Willems Braun, B. (1994). Situating cultural politics: fringe festivals and the production of spaces of intersubjectivity. *Environment and Planning D: Society and Space* 12, 75 104.

Relevant Websites

http://www.garma.telstra.com
Garma Festival.
http://www.ifeaeurope.com
International Festivals and Events Association Europe.
http://www.ifea.com
International Festivals and Events Association (USA).
http://www.crctourism.com.au
Sustainable Tourism, Cooperative Research Centre (Australia).
http://www.mardigras.org.au
Sydney Gay & Lesbian Mardi Gras.
http://www.unesco.org
UNESCO, United Nations Educational, Scientific and Cultural Organization.
http://www.world tourism.org
WTO, World Tourism Organisation.

Feudalism and Feudal Society

H. Bradley and M. Taylor, Mary Immaculate College, University of Limerick, Limerick, Republic of Ireland

Glossary

Feudalism A postmedieval construct developed to describe a system of devolved political, military, social, and economic control from a king or lord to a vassal in the Middle Ages.

Feudal Society A military-based society, dominated by a system of reciprocated relationships, between a lord and his vassal. This relationship centered on personal fealty, service, and protection. It can also describe a hierarchical social order which emerged in Europe in the 'Middle Ages' based on the legal subjection of the peasantry by a landholding elite. These landholding elite were protected, sustained, and reinforced by their domination of the political and judicial systems together with the economic mode of production.

Fief A gift, usually land but on occasion an annuity or public office, granted by a lord to his vassal in return for military, judicial, and administrative services and homage. A fief was originally granted for the lifetime of the vassal only. By the tenth century it had become hereditary.

Fealty The loyalty owed by a vassal to his lord on becoming his 'man'. An oath taken by a vassal in which he promises to protect his lord's military, economic, and political interests.

Homage A feudal ceremony in which a vassal submits to his lord and promises to show respect and be his 'man'.

Lordship Position or authority enjoyed by a feudal lord or the geographical area within which a feudal lord controlled the political, military, social, and economic landscapes.

Manor A subdivision of a lordship with social, economic, administrative, defensive, settlement, and judicial functions.

Primogeniture A system of inheritance found in feudal Europe, whereby land was deemed legally indivisible and could, in the first instance, only be passed from the father to the eldest son.

Vassal A person under the protection of a feudal lord to whom he has promised homage and fealty. It is also the name given to the holder of a fief. In return for the granting of a fief, the vassal is obliged to attend the lord's court, administer the lord's justice, serve in the lord's army, and promise homage and fealty to his lord.

Vassalage A feudal contractual relationship found among the upper classes in feudal Europe. It was a system whereby a free person (vassal) voluntarily accepted the overlordship of another (lord) in return for the granting of a fief. Both parties entered into a feudal contract. The lord promised to protect the vassal and his family and settle disputes in return for homage, fealty, and military and personal services. A vassal could enter into a contract with more than one lord; this practice was known as multiple vassalage.

Introduction

Feudalism is an elusive concept and considerable ambiguity exists in how it is defined and subsequently employed by academics. This article discusses the meaning of the term feudalism and evaluates its impact on medieval society. The key characteristics of feudalism are identified and discussed. These include: vassalage; homage and loyalty to lord; the granting of fiefs; hierarchical society, land tenure, and the devolution of public power into private hands; and military tenure and service in arms in the form of the knight's fee. Definitions and theories associated with feudal society, such as 'feudal revolution' and 'feudal mutation' are presented. The temporal and spatial impact of feudalism on the landscape is also discussed. Particular reference is made to the impact of feudalism on the social, military, economic, and settlement landscape in the county of Limerick, Ireland. Finally, the end of feudalism is outlined.

As a term, feudalism is derived from the French word *féodalisme*, which originated from the Latin word *feodum*. Over time it has been coupled with terms such as feudal system, feudal regime, feudal law, feudal mode of production, feudal society, feudal monarch, feudal lordship, feudal revolution, and feudal mutation, to name but a few. The term is, therefore, difficult to define precisely. The French historian, Bloch, maintained that nearly every medieval historian understands and defines feudalism as he/she pleases. Economic historians, for example, place great emphasis on the exploitation of land and tenants by landholding classes. Social historians examine the dependency on vassalage and the fief. Constitutional historians study the exaction of taxes and the enforcement of the law. For others, the term has become synonymous with tyranny and oppression.

In its simplest format, feudalism describes a system of devolved political and economic control from a king or lord to a vassal. It describes the political and social structures, whereby landowning elites governed ordinary people. It is said to have originated in the area between the Loire and Rhine rivers, in the ninth and tenth centuries, in response to barbarian invasions and the resulting inability of the Carolingians to assert their authority and protect their vassals. From there it is said to have gradually diffused into Western, Central, and Eastern Europe.

The feudal system centered on vassalage, homage, and the fief. Vassalage and homage were not however new institutions. The term *vassus* makes its first appearance in the sources in the Merovingian period in reference to a minor court official. It is however used with greater currency from the end of the eighth century to the eleventh century. During this period, it came to describe as a freeman in the service of his lord. In return for the granting of a fief, the freeman or vassal entered into a binding contract with the king or lord by agreeing to be his 'man' and perform certain military, administrative, and personal services. The vassal, for example, was obliged to render 'aid and counsel' to his lord. This involved supporting the lord in all internal and external political and legal disputes and rendering him aid if necessary. These personal obligations or services could, over time, be modified but could not be set aside. In return, the lord offered his vassal protection. Homage also has its antecedents in the Merovingian period and refers to a public ceremonial act where the vassal knelt unarmed before his lord, laid his hands in his lord's, and promised to be his man.

A fief could take many forms. It could, for example, be given as an annuity or allowance, granted by the lord to his vassal following a successful military campaign. It could also have involved a vassal taking up public office in his lord's administration. More often than not, however, it was centered on a piece of land, which was granted to a vassal. Initially, the fief was granted only for the lifetime of the vassal, but by the tenth century it had become hereditary, on the condition that the new vassal swore homage and fealty to the lord. Only in extreme cases, where the vassal was disloyal, died without an heir (escheat), or his heir was a minor (wardship) did the land revert back to his lord, and in some cases only on a temporary basis. The permanency of tenure, based around the fief, ensured the continual expansion of the feudal system and the confidence of its lords. Primogeniture, whereby the fief became indivisible and passed to the oldest son, on the death of his father, became the preferred method of inheritance. As a result, family lineage attained greater prominence than the extended kinship unit.

The Study of Feudalism and Feudal Society

The Belgian historian, Ganshof, centers his definition of feudalism on a legalistic and military explanation. In his definition, he places great emphasis on the possession of landed property (fief) and the role it played in defining social relationships between lord and vassal and lord and peasantry. He describes feudalism both as a form of society and as a body of institutions. When dealing with feudal society or his own preferred term the 'feudal regime', Ganshof identifies a number of innate characteristics. These included, first, an emergence of personal dependency in society, which was hierarchical in nature, with knights occupying the upper stratum. Second, it included a subdivision of the rights to property. Third, a graded system of rights to land created by the subdivision of property, and finally, devolution of powers normally associated with a state, to a 'hierarchy of persons' at the local level. Ganshof's study clearly advocates the union of fief and vassal. The lord–vassal relationship quickly became a land based one – a vassal was granted a tract of land in return for service. The holder of the fief became a vassal of his lord, rendering him homage and fealty. The relationship between a lord and his vassal was a personal one and entered into freely, by both parties, while the relationship between the lord and his peasantry was viewed as impersonal and coercive in nature. This created a hierarchical society, based on a paradox of equity and inequality and differing rates of dependence to the feudal lord. Central to this hierarchical society were the key institutions of vassalage, homage, and the fief. These institutions regulated the lord–vassal relationship, while also preventing social mobility among those denied vassal privileges. Ganshof's definition is in line with the British historian, Maitland's, earlier definition. Like Ganshof, Maitland emphasized the legalistic aspects of feudalism, emphasizing the role, which the fief, vassalage, military service, and the judicial system played in maintaining the feudal system.

Others, such as the American scholar, Stephenson, view feudalism primarily as a political phenomenon centered on a military tenurial system. While the motives for the creation of feudal holdings varied both spatially and temporally, the resulting feudal institutions provided a very practical, simple, and stable means of localized government in an age of political instability. Feudalism could also be seen as a system of localized government where the primary relationship was between lord and vassal as opposed to ruler and subject.

Bloch expands on the legalistic and military description by incorporating sociological dimensions. For the most part, his definition included most of the fundamental elements of feudal society as opposed to simply the lord–vassal relationship and its associated features. These included (1) a subject peasantry, (2) use of a service tenement (fief) instead of a salary, (3) the emergence of a class of specialized mounted knights who became more powerful than centralized institutions, (4) emergence of ties of obedience and protection which bound the elite in society – vassalage, (5) fragmentation of authority leading to disorder, and (6) survival of other forms of association, for example, the family and the state. Of primary concern was the resultant society which feudalism fashioned. According to Bloch, feudal society was not simply an upper class phenomenon but rather one that was centered on lords and vassals and lords and peasants. The ensuing society was stratified both politically and socially and the resulting system was one of appropriation and exploitation. This is disputed by others who limit the definition of feudal society to those who controlled society. Central to Bloch's definition was the issue of societal continuity.

Marx, in his stage or epochal model of societal development viewed feudalism as simply a stage in the economic development of a given society. For him, the feudal mode of production was an essential component in the progression toward a socialist system. The feudal mode of production succeeded the classical and preceded the capitalist mode of production. At its very heart, feudalism was a system of exploitation, whereby the peasantry were exploited economically, through the extraction of the surplus, by the landowning classes. For many medievalists, this definition is too broad as it could include any society with a privileged landholding class and a dependent peasantry. Others prefer to view these economic practices of exploitation as part of the manorial system, and hence the term manorialization is used to describe the economic arrangements which existed between the lord and the peasantry based around the spatial unit of the manor.

While these classical definitions of feudalism are still utilized, many modifications have emerged. Criticism has centered on their spatial, temporal, and structural specificity. There can be no doubt that the impact of the feudal system varied both spatially and temporally and appeared in various different guises, at different times, across the European landscape. It appears to have been most advanced in France and Germany in the tenth and eleventh centuries, where public authority was ineffective and where members of the aristocratic class maintained a personal grip on power and authority. While all vassals ultimately held land from their king, the king was often unable to enforce his authority. Bloch defines this initial period as the 'first feudal age'. This period was characterized by the individualism of lords and the spontaneous growth of basic feudal institutions. During this period, kings effectively lost control over their realms with the formation of independent local seigneuries. The situation, however, changed during the late eleventh century, the period termed by Bloch as the 'second feudal age'. This period was characterized by a reaffirmation of sovereign authority, through the reclaiming, by the king, of core public functions, for example, the administration of justice and the issuing of legislation. This was evident in England in the mid twelfth century. This period also witnessed a controlled systematic spatial expansion of feudalism into Central, Eastern, and Western Europe. This was necessitated by, among other things, population growth and the lack of land based opportunities for younger sons of aristocratic families due to the laws of primogeniture in the feudal core. In these newly acquired territories, kings ensured that grants of land were only given to their loyal followers. In all these landscapes, regional variations existed and the system was adapted to suit unique situations. The very flexibility and pliability of the system was perhaps one of its greatest strengths, in particular in frontier locations in Wales and Ireland in the twelfth and thirteenth centuries, respectively.

Some scholars, for example, the British historian Maitland, have suggested that feudalism had the potential to manifest itself differently across the European landscape. In his work on English feudalism, he highlighted the potential for spatial and temporal variations in the effectiveness of the feudal system. He states that feudalism in France was radically different to the feudalism encountered in England and that feudalism of the eleventh century was very different to that practiced in the thirteenth century. He claims that the English landscape did not experience the development of a classic feudal system. Ganshof also recognized the spatial limitations in his study, which concentrated primarily on the Carolingian Empire to the detriment of England and Italy. Feudalism could, consequently, be seen as more a method of organization and as such no static definition should be applied; rather, recognition should be given to the fact that like other political systems, it was constantly developing and had the ability to adapt to changing situations.

Feudalism should, therefore, be viewed as a response to a particular stimulus, that is, the decay or weakening of a highly organized and centralized political system in the Carolingian Empire in the tenth and eleventh centuries. According to some historians, this weakening in centralized authority resulted in a feudal revolution. The presence or absence of a feudal revolution is a debatable issue in medieval studies. This revolution is said to have come about as a result of a geopolitical shift in power in medieval Europe. It originated in the collapse of the Carolingian Empire in the tenth century and the resulting weakening of public centralized authorities. This initiated structural changes in the geography of power in medieval Europe. With the weakening of centralized authority, private power bases, focused on the exploitation of the local landscape, emerged. The establishment of these new territorial or seigniorial lordships, armed with the apparent authority to rule localities, together with the multiplication of knights and their castles have been described as the feudal revolution. These lordships and their lords became the new leviathans or powerbrokers in feudal Europe. Some scholars, in pre ferring the use of the term mutation as opposed to revolution, proposed that the process involved changes at a political and cultural level and the emergence of dominant small scale forms of organization. The stability of centralized power and the judicial system which en forced it in the pre eleventh century was viewed as the direct antithesis of the instability witnessed in the same institutions in the eleventh century.

The French historian Duby was the first to suggest this revolution in his classical study of the Mâconnais in France. In this study, Duby describes the breakdown in public order in Mâconnais, which occurred during the years 980–1080. His study inferred that discontinuity with the past was the key, and the end result was an abrupt change in the character of society. Other scholars have since advanced on the work of Duby by expanding on the significance of the feudal revolution to include not just the judicial institutions of change, but also the mechanism or vehicle of that change – the territorial seigniorial lordship and its violent characteristics. Vio lence in tenth century society was not new. What had changed, however, was the way in which violence affected lordship and social relationships, in particular between the lord and his peasantry. Violence was exploitative in nature, enforcing its will on the powerless in society. At its core was the quest for status and power. Violence and its changing nature had the ability to mold a new order of power. Lordship was predatory in nature and was seen as a means to an end. The key to public authority lay not in the holding of public office but in the possession of land as represented by the fief or lordship. Land was, however, a finite resource in the Middle Ages, hence, the expansion of lordship, in particular the exploitative or tyrannical variety, into unchartered territories, on the fringe of feudal Europe, in the eleventh and twelfth centuries. Violence, as a means to an end, to christianize or civilize other cultures and religions was deemed justifiable and acceptable in the circumstances. The disruption, how ever, occurred when the violence of lordship could no longer be contained.

The potential spatial influence of this feudal revo lution or mutation has also been disputed. Barendse ar gues for the existence of a feudal mutation in world history. His feudal mutation model depicts feudalism both as a useful term in world historic analysis and as a process which left an imprint on vast areas of Eurasia from the tenth to the thirteenth century. He identifies three essential criteria in his feudal mutation model. First, an increase in peasant productivity as a result of an increase in agricultural productivity, colonization, and trade. Second, the emergence of a mounted aristocratic class intent on expanding their power and authority to the detriment of the peasant classes. Third, changes in sociopolitical structures as a result of changes in both the nature of power and how it was expressed. The feudal process, therefore, centered on the subjugation of the peasantry by a military aristocratic class. This military aristocracy were more powerful than any existing central institution and for that reason they had complete control over all aspects of peasant life. The resulting system was, therefore, not unique to feudal Europe. Accordingly, he maintained that European feudalism was just a de centralized variant of a world historic system. For medievalists, this definition is too broad and if used, the term loses any real analytical usefulness for the study of medieval Europe.

The Characteristics of Feudal Landscapes

The American geographer Sauer suggests that each culture has the ability to create its own cultural land scape. This cultural landscape can be influenced by, among other things, the dominant ideology of the time. In the case of medieval Europe, feudalism represented the dominant ideology. The British geographer Dodg shon was one of the first to question whether feudalism created its own unique landscape and whether the feudal institutions of fief and vassalage created unique spatial patterns, which differentiated landscapes in the Middle Ages from earlier kin based landscapes.

Feudalism, through the spatial unit of lordship and the social unit of vassalage, was used as a device of govern ment, or as the mechanism by which the king or lord controlled diverse territories within his jurisdiction. Vassalage, while it originated as a purely social rela tionship was transformed into a social–spatial relation ship, with vassals of the king granted both land and lordship in return for fealty and military service. This, according to Dodgshon, was the paradox that was feu dalism. It required power to be concentrated in the hands of a king, but in order for the king's control to filter down to the lords and in turn to the peasants, the king had to devolve power to his vassals through grants of land. Lords, in turn, subdivided or subinfeudated their lands among their tenants, thereby duplicating the feudal sys tem on individual manors. An example of this can be seen in the context of the medieval lordship of Ireland where the king subinfeudated lands to his loyal followers in return for their loyalty and military service. These lords in turn further subinfeudated lands to their loyal fol lowers, an example of which can be seen in the twelfth century Limerick landscape (see **Figure 1**). As a result, power became socially and spatially regulated at various different scales.

Feudalism, therefore, anchored the relationship be tween the lord and vassal to specific blocks of land. The land held was spatially regulated within the territorial

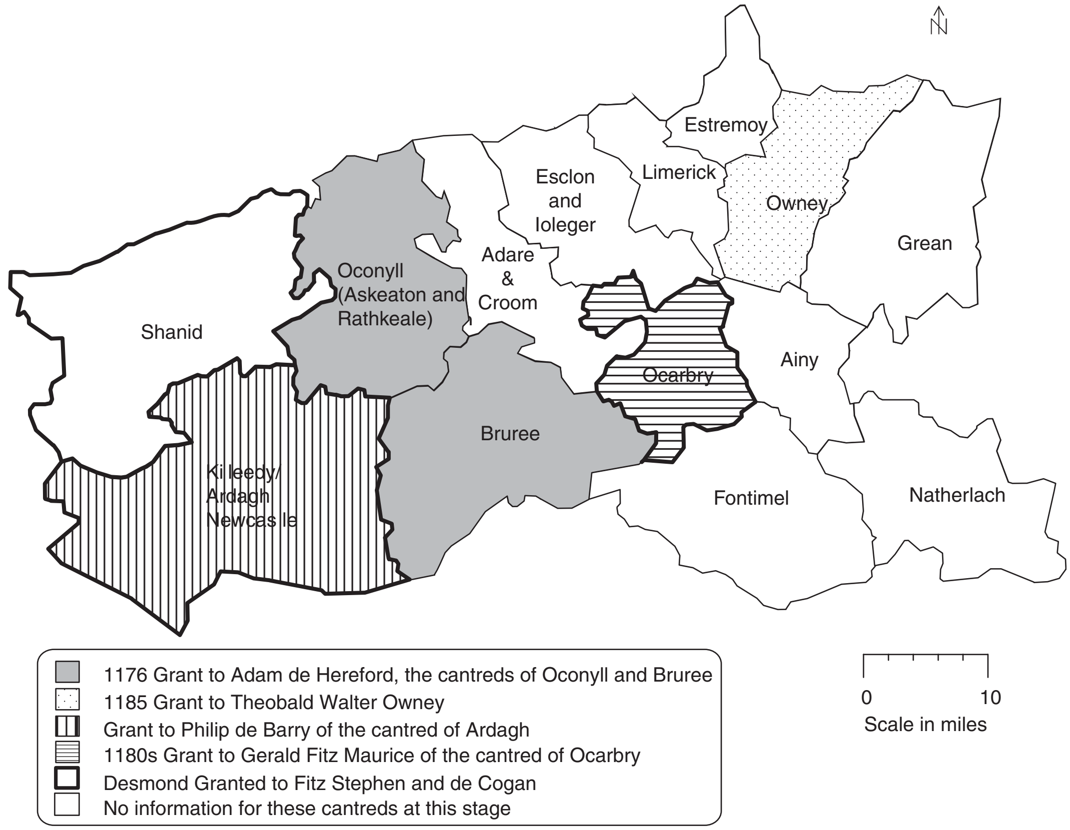

Figure 1 Twelfth-century land grants by cantred in the county of Limerick, Ireland. From Taylor, M. (2007). *Anglo-Norman Settlement in Rural County Limerick 1175 1321. A Geographical Analysis.* Unpublished MA Thesis, Department of Geography, Mary Immaculate College.

framework of the lordship and manor. The lord's place in the social hierarchy was based on the size of his land grant and the nature of his tenurial relationship with his overlord. Lordship represented the spatial basis of feudal power in the Middle Ages. It defined a spatial area within which the feudal lord had feudal rights. However, not all lords or lordships were equal in status or power, and therefore a hierarchical system of lordships emerged. This hierarchical system was subsequently mirrored in geography of the manor. (See **Figure 2.**)

The social landscape: The feudal landscape was socially regulated. People were locked into a particular place in society with little opportunity for social or geographical mobility, in particular from the bottom up. The rela tionship between the lord and his free tenants differed. It was hierarchical and was interlinked with the tenant's tenurial relationship and his social standing in manorial society. Generally, the greater the amount of land a tenant held, the higher his social status. Social status was also based on the type of service owed to the lord. Those free tenants who attended the manorial court or 'did suit' were of a higher social status than those free tenants who were rent paying. The land divided among the tenants was measured in a variety of ways and the payments due to the manorial lord differed according to the status of the tenant.

The military landscape: The feudal system of land holding was centered on the need for localized military defense. As a result, a lord was granted land in return for military service or knight's service. A tenant who held land by knight's service was obliged to either serve in person in his lord's army or provide the services of a knight to serve in his lord's army. When the military needs of the lord became more specialized, for example, in Ireland in the twelfth century, this was commuted to a money payment, usually £20 a year, called scutage. The knight's fee as a unit of measurement was intended to express the potential value of the unit of land in terms of the military service it might be expected to yield. As the purpose of military tenure was the creation of a garrison, the size of a fee depended on the security of its location.

The core of the fief was centered on the location of the castle. Castles symbolized feudal lordship and fur nished upon the lord the means to control people through the garrison of mounted men that were housed within. The castle was a military innovation that was a real, but importantly not an absolute, part of the feudal landscape. Castles were seen as instruments of defense

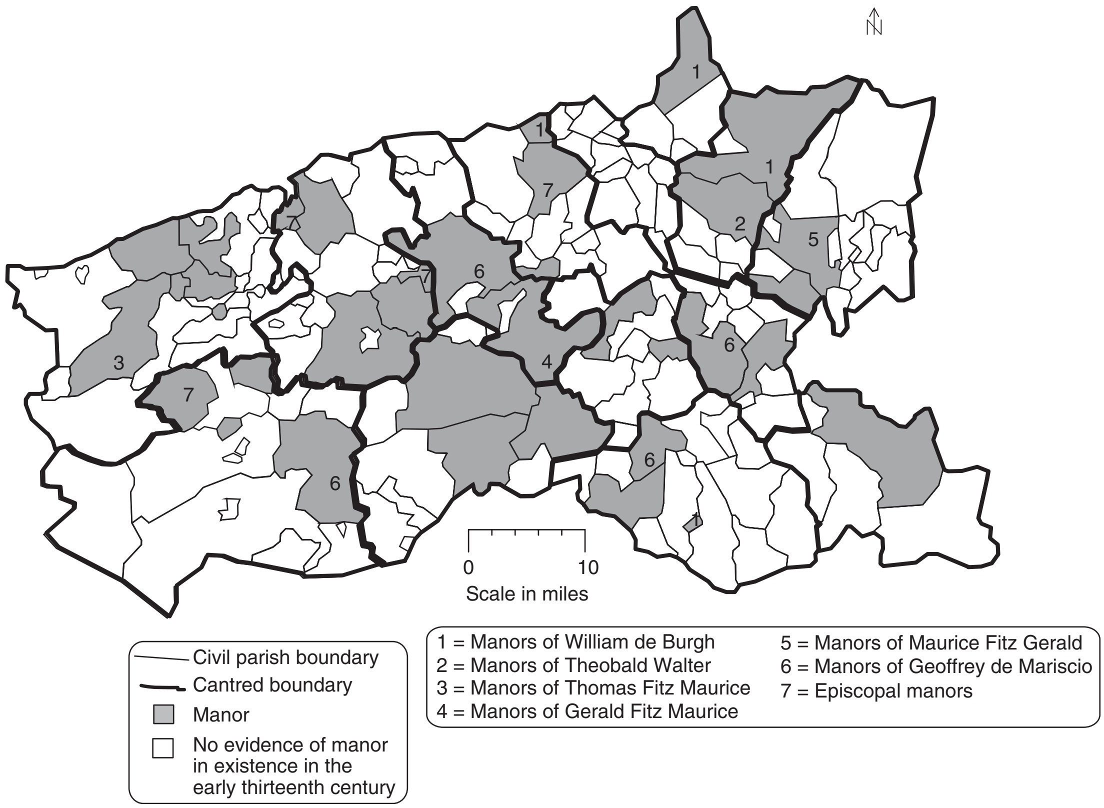

Figure 2 Thirteenth-century manorial landscape in the county of Limerick, Ireland. From Taylor, M. (2007). *Anglo-Norman Settlement in Rural County Limerick 1175 1321. A Geographical Analysis*. Unpublished MA Thesis, Department of Geography, Mary Immaculate College.

and of territorial control. They gave families of the noble class a specifically located center of power, a family seat which they would defend at all costs. In this way, the localized feudalized landscape became personalized.

A castle could take on many forms. In medieval Ireland, the earliest feudal fortifications were earthen motte and bailey castles. These provided for the localized defensive needs in the early stages of the colonization of Ireland. Many of these were replaced in the late twelfth and thirteenth centuries by masonry castles. The position of an individual manor within the overall social hierarchy in Ireland was often reflected in the location and type of castle constructed on these manorial lands. For example, the major lords built large and impressive baronial castles in the center of their manors which catered to the administrative and defensive needs of their lordships. The lesser manorial lords constructed smaller castles which catered to the needs of a more localized landscape. The resulting hierarchy of castle structures mirrored the social hierarchy of the lordship (see **Figure 3**).

In the past, castles were often studied in isolation from the settlement pattern around them. They are now more correctly viewed as just one component of that settlement pattern. This can be seen in the many functions of the castle. Apart from being a defensive stronghold and residence of the lord, a castle was also the administrative and judicial center of the lord's manor and held the manorial court. Ironically, a stone castle could communicate a number of messages which included power and authority on the behalf of the lord and fear and alienation on behalf of his dependent tenants. A castle was an indicator of wealth, of power, and of endurance. The construction of a stone castle was symbolically important as it indicated that the lord's intentions were long term; they were there to stay.

The economic landscape: The theory that feudalism created its own landscape can be seen in the spatial organization of the feudal manor. The manorial structure provided the institutional framework for the lords and the free tenants to organize and administer their territories. It was the manorial structure which allowed feudalism to reach down to the ordinary person in society. The importance of the manor to the success of feudalism should, therefore, not be underestimated. Dodgshon outlines how through the manorial structure the relations between the king or lord and his vassals were set down firmly in space, and in turn the relations between the vassals and their free tenants were anchored to specific blocks of manorialized land. There was a high degree of uniformity in the spatial organization of a manor. The

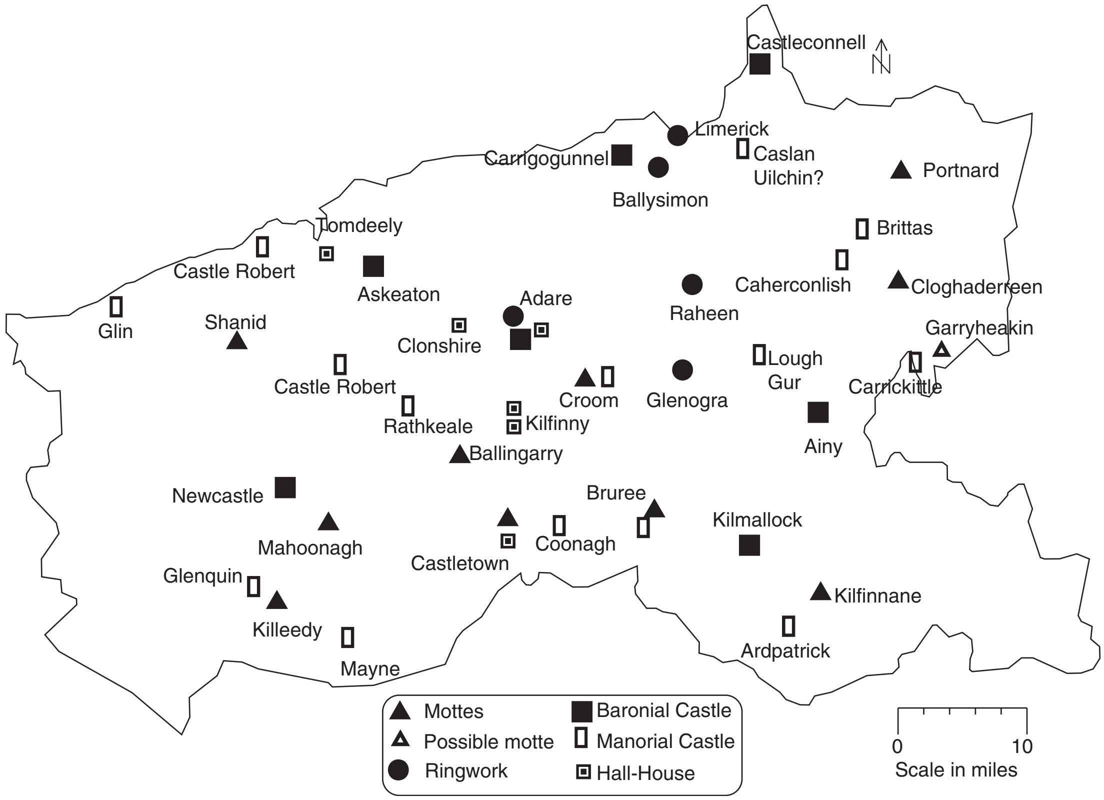

Figure 3 Twelfth thirteenth century defensive structures in the county of Limerick, Ireland. From Taylor, M. (2007). *Anglo-Norman Settlement in Rural County Limerick 1175 1321. A Geographical Analysis*. Unpublished MA Thesis, Department of Geography, Mary Immaculate College.

manor was divided into two sections: the demesne land for the lord's own use and the land rented out to the lord's tenants (see **Figure 4**). In this way, the land became increasingly economically viable.

The settlement landscape: Feudalism created a new framework for both rural and urban settlement. According to the Irish geographer, Graham, the settlement landscape evolved from the requirements of feudalism, in particular its military, political, and economic needs. In Ireland, the emergence of manors facilitated the creation of a new layer of nucleated settlement on the landscape, which differed considerably from what had preceded it. This can be seen principally in the spatial organization of the feudal manor, where relations between the crown and his lord and the lord and his tenants were set down in land. Settlement of the manor varied temporally and spatially. Some manors developed greatly and became a focus for settlement. Defensive structures such as motte and baileys and stone castles were built at manorial centers. Many of these manorial centers with their defensive, religious, economic, administrative, and residential functions attracted more settlement and grew into manorial villages. Their central focus was on the defensive structure and the church (See **Figure 5**). Many of these settlements were also granted borough status. This involved the granting of certain privileges on a group known as burgesses. In Ireland, these privileges were generally modeled on the Breteuil charter of the Normandy town of Breteuil sur Ilon. It granted the settlement a certain status and would have been an inducement for further settlement. Some of the boroughs would not have developed further. A number, however, would have developed into towns. As a result, nucleated settlements became the norm, for example, the planned villages of Eastern Europe or the towns of Anglo Norman Ireland.

The End of Feudalism

The point at which the feudal system ceases to exist is shrouded in controversy. By and large, it is dependent on how feudalism is defined. If a political definition is applied and feudalism is seen to be dependent on the decentralization of authority, then the reemergence of centralized power, through the instrument of the territorial state in the fifteenth century, signals the end of the feudal system. On the other hand, if feudalism is viewed as simply a stage in the economic development of a given society, the introduction of the capitalist mode of production in the sixteenth century signals the end of the

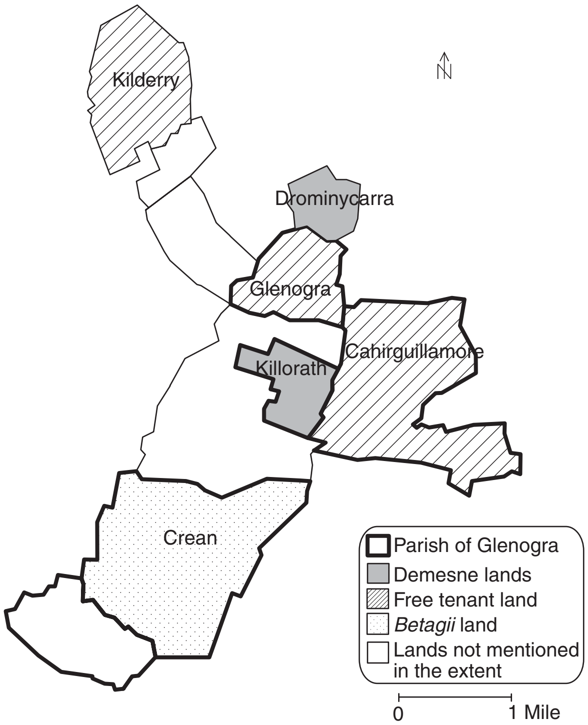

Figure 4 The internal organization of the manor of Glenorga in the county of Limerick, Ireland. From Taylor, M. (2007). *Anglo-Norman Settlement in Rural County Limerick 1175 1321. A Geographical Analysis*. Unpublished MA Thesis, Department of Geography, Mary Immaculate College.

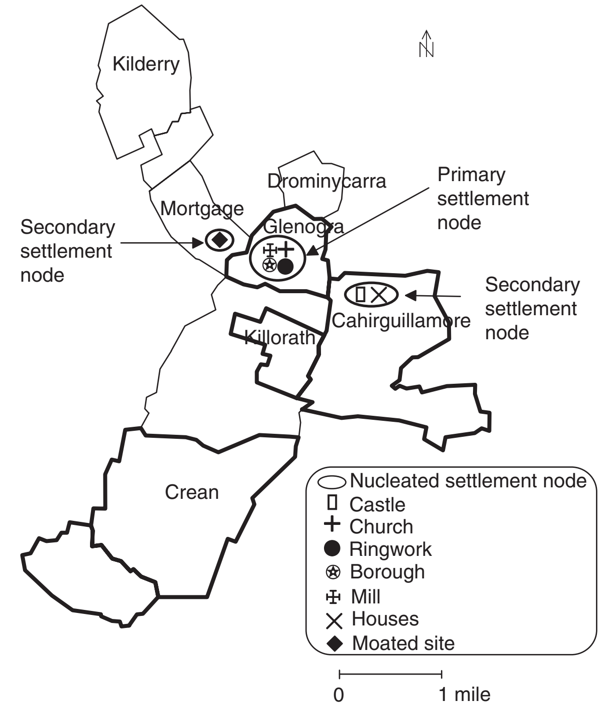

Figure 5 The settlement geography of the manor of Glenogra in the county of Limerick, in the thirteenth and fourteenth centuries. From Taylor, M. (2007). *Anglo-Norman Settlement in Rural County Limerick 1175 1321. A Geographical Analysis*. Unpublished MA Thesis, Department of Geography, Mary Immaculate College.

feudal system. Conversely, if a definition of feudalism centers on the suppression of the peasant classes by the aristocracy in France, then the French Revolution of 1798 defines the exact moment when feudalism was condemned to the history books.

While a definitive definition of the term feudalism may remain elusive, the key institutions of the fief, vassalage, and homage are central to all definitions. However one defines the term, there can be no doubt that feudalism remains a very useful analytical tool which allows geographers to compare and contrast varying landscapes and cultures in Europe in the Middle Ages.

See also: Field Systems and Enclosure; Geohistory; Historical Geographies, Rural.

Further Reading

Barendse, R. J. (2003). The feudal mutation: Military and economic transformations of the ethnosphere in the tenth to the thirteenth centuries. *Journal of World History* 14(4), 503 529.

Bartlett, R. (1993). *The Making of Europe. Conquest, Colonisation and Cultural Change 950 1300*. Princeton, NJ: Princeton University Press.

Bisson, T. N. (1994). The feudal revolution. *Past and Present* 142, 6 42.

Bloch, M. (1961). *Feudal Society*, 2 vols. New York: Routledge & Kegan Paul.

Brown, E. A. R. (1974). The tyranny of a construct: Feudalism and historians of medieval Europe. *American Historical Review* 79, 1063 1088.

Butlin, R. (1996). *Historical Geography. Through the Gates of Space*. Oxford: Oxford University Press.

Coulborn, R. (ed.) (1963). *Feudalism in History*. Connecticut, CT: Archon Books.

Dodgshon, R. A. (1987). *The European Past: Social Evolution and Spatial Order*. London: Macmillan Education.

Dodgshon, R. A. (1998). *Society in Time and Space: A Geographical Perspective on Change*. Cambridge: Cambridge University Press.

Duby, G. (1980). *The Three Orders: Feudal Society Imagined*. London: University of Chicago Press.

Ganshof, F. L. (1952). *Feudalism*. London: Harper & Row Publishers.

Herlihy, D. (1970). *The History of Feudalism*. New York: Walker.

Hilton, R. (1990). *Class Conflict and the Crises of Feudalism*. New York: Verso.

Hollister, C. W. and Bennett, J. M. (2001). *Medieval Europe: A Short History*. London: McGraw Hill.

Le Goff, J. (1988). *Medieval Civilization*. Oxford: Basil Blackwell.

Mac Lean, S. (2007). Review article: Apocalypse and revolution: Europe around the year 1000. *Early Medieval Europe* 15(1), 86 106.

Maitland, F. W. (1968). *The History of English Law*. Cambridge: Cambridge University Press.

Reuter, T. (1997). The 'feudal revolution'. *Past and Present* 155, 177 195.

Reynolds, S. (1994). *Fiefs and Vassals: The Medieval Evidence Reinterpreted*. Oxford: Oxford University Press.

Sauer, C. (1925). The morphology of landscape. *Geography* 2, 19 54.

Stephenson, C. (1941). *Medieval Feudalism*. Cambridge: Cambridge University Press.

Field Geographies

D. Abbott, University of Derby, Derby, UK

Glossary

The British Survey Movement This movement emerged toward the end of the nineteenth century both in the UK and the USA. In a 'movement' which included several thousand amateurs and professionals, many 'regional surveys' based on 'scientific' methods were undertaken mainly by sociologists (but also by those now known as human geographers) on 'fact finding' missions. These were based on specifications of survey designs published in several pamphlets and articles by people such as Geddes, Victor Branford and Le Play House. Challenges to the positivist nature of the methodology which emerged in the late 1950s and then 1960s eventually saw the disappearance of the 'movement'.

Le Play Society It was formed in the UK in the1920s, originally through study groups that developed around sociological issues such as social finance and credit for social needs. Eventually, geographers too began to participate in organised tours and fieldwork (Le Play tours) that were undertaken to carry out regional surveys around these issues. Soon, there was a need to separate out the study and travel focus, and a breakaway society, Le Play Society, was formed focusing solely on student tours or fieldwork.

Standpoint Theory With its roots in feminist discourse, began as a critique of 'rational', 'objective' methodologies and explanations. Standpoint theorists consider knowledge derived from this as biased to Western cultures and male perspectives which altogether present a limited picture of the world. They thus argue for knowledge that derives from multiple perspectives, a knowledge that takes into account communications which take place in a wider context and allow real voices to come through, for example, those of women in poverty.

Situatedness/Posionality Arguments are about the production of wider knowledge through the development of a methodology that is capable of relating to the context and position of actors within diverse settings. Feminist geographers, for example, have studied 'work' and 'home' as locations of situated knowledge for women. Others have considered race and patriarchal contexts in discussing post-colonial positionalities of historically colonized peoples.

Epistemological discourses such as those that have emerged from women's studies (particularly feminism) and development studies have, however, shown geographers that there is a need to challenge the power assumptions embedded in the whole process of research, including methodological choices that can include or exclude. By tracing these discourses and using examples from these two disciplines, this article demonstrates how contemporary geography has taken on board some of the new methodological approaches that have thus transpired. In turn, this has enriched geographical enquiry, which is now, much as the subject itself, seen as a social construct requiring critical reflection and challenge.

Introduction

Often students are attracted to geography because it promises them exciting opportunities of 'fieldwork' both at home and abroad. Teachers too see fieldwork as essential in teaching geography and this is reflected in the definitive geography curricula of many higher education institutions all over the world. However, what constitutes 'the field' and defines 'fieldwork', fieldwork practices, and field methodologies has been fiercely contested, particularly in recent times. An old fashioned, 'traditional' view is summed up, for example, in the following quote:

> Field work is essentially personal observation and recording; it brings reality to geographical study; it helps the geographer to acquire his all important understanding 'eye for the country'; and thus it enriches his descriptive and explanatory powers. I would say that an essential part of the training of a young geographer is for him to choose some small accessible unit area that attracts him; acquire a pair of stout boots, perhaps the geographers first item of equipment, study in the area itself the association of physical and human conditions which there prevail, and in fact give the area its individuality; and record the information he collects in a series of original maps. (Monkhouse, 1955 in Desylers and Starr, 2001, p. iv)

There are many suppositions underlying both the meaning of geography and geographical fieldwork in this quote. These include that fieldwork is about 'going out there' in all types of weather, that it involves collection and recording of information from a small and accessible geographical area, and that the geographer is a man, entering a tough terrain, being prepared for the masculine worlds of geographers.

As hinted in the first paragraph, in recent years however, particularly since the 1960s, all these suppositions have been challenged through a long and ever growing debate starting with the very premises of geography as a discipline, and what constitutes its 'fields', leading onto the nature and meaning of 'fieldwork' and how best this can be done. Geographical 'fields' are no longer viewed as accessible small area studies, as sug gested in the Monkhouse quote above, or 'fieldwork' as based on 'objective' quantifying and mapping of 'facts'. Rather, contemporary geographers may argue that 'the field' occupies political and social spaces around us in our everyday locations, whether on our doorsteps or at a global distance. These 'expanded' fields can be accessible or not, in distant places or not, and visible or not. If anything, in doing 'fieldwork', we need to look beyond the visible in order to discover geographical meanings and human use of physical and social spaces.

This requires a rethink of methodological and 'field work' practices which go beyond the 'objective' and 'distant' methods of mapping boundaries of the physical 'geographical terrain' to the recognition of the 'social terrain' where power relationships embedded in the daily politic define how maps are shaped . For example, the might of the British Empire was often portrayed to school children through (dark pink or red) colonial maps that juxtaposed Britain as the center of the world – a notion that has of course since been turned upside down by projections that have centered on others, for example, Peters projection centering on Africa. This article dis cusses these points further by giving a brief insight into some of the arguments that have shaped debates on contemporary field geographies. It particularly draws on two perspectives which take the notions of power and powerlessness as their central tenet, that is, a feminist one and a postcolonial one. It is suggested that these per spectives have reframed the meanings of geographical investigation, calling for a 'subjective' return to 'the field' which allows geographers to 'explore' the real stories of power relationships that define the use of space and place, in contrast to supposed neutral political spaces of 'distant' fieldwork as illustrated by the quote at the start of this section.

Exploring 'the Field'

It could be said that the 'traditional' model of the 'field' as something out there waiting to be discovered and re corded scientifically is associated with Alexander von Humboldt (1769–1859), the founding father of field studies. Humboldt was a Prussian naturalist who ex plored and wrote about much of Central and South America. He is famous for his diverse activities which include an extensive collection of plant, animal, and mineral specimens which he acquired during his travels; the extensive mapping of northern South America; and the recordings he made of scientific phenomena such as altitudinal and observed astronomical events. Within the context of, and opposition from, divinity and religious based explanations dominant at the time, Humboldt championed 'scientific reasoning', an explanation that was also to influence Darwin's *Origin of the Species* published soon after his death.

However, as Doreen Massey has pointed out, Hum boldt's methods of 'scientific' investigation immediately drew criticism, for instance from Georges Cuvier (1769–1832) who carried out his naturist anatomical studies in the dissection rooms of a museum in contrast to Hum boldt's exploration led field research. In comparing the desk based work of a 'sedentary naturalist' who is able to minutely examine single objects with that of an explorer who takes a 'broad brush' look at the natural environ ment, Cuvier questioned the true location of science. Is this indoor or outdoors? Where can superior knowledge be gained, in natural habitats or through desk based anatomical studies? Cuvier's questioning led to what Massey identifies as a definitive moment, shaping ques tions of epistemology, spatiality, and the general nature of science for geographers as it essentially asks where 'the field' is located, and how 'fieldwork' should be carried out.

Echoes of similar questions are evident in con temporary debates on 'the field', continuing since the Victorian times when the sons of the rich (Humboldt was a baron and women 'explorers' all but invisible) could discover exotic locations and attempt to provide a sci entific rationale for geographical phenomena. Questions include key issues of the distance of locations as well as the 'objective' distance between the researcher and those researched; knowledge gathering; and appropriate 'field' methodologies including how best to represent the 'en counters' in the field in order to seek the 'truth' and represent accurately.

In fact, over the years, there has been much batting to and fro over these questions with increasingly sophisti cated philosophical argument regarding various facets of 'the field' and 'field' practice. For instance, there has been much criticism of earlier influential geographical prac tices of 'viewing' the other from a distance as exemplified by Patrick Geddes who stood on the 'Outlook Tower' he established in 1892 in Edinburgh (Scotland), talking and pointing at a height. This teaching technique has, of course, been replaced with other interactive methods today, yet it remains a familiar method for many geog raphy teachers in British schools. Another example is the fondness for 'objective' survey methodologies practiced in British schools and universities, the influence for which can be traced to at least a hundred regional fieldwork 'expeditions' that were carried out by the Le Play Society in Britain between 1930–60 in order to 'scientifically' collect data, identify causation, later utilized to feed into prescriptive policies.

The need for distance and 'objective' disengagement from 'the field' that is exemplified above has come under much criticism since the 1960s. This is partly due to a broadening of the discipline itself. Contemporary geography, particularly human geography, has seen a major shift in recent years from an awkward attempt to fit into a 'scientific' world of physical geography shaped by scientific tools, techniques, and justification. It is now a much wider discipline, crossing boundaries with several other disciplines such as ethnography, anthropology, sociology, and media. In fact, in the mid 1980s, as methodological boundaries between other disciplines and contemporary human geographies became blurred, geographers such as Eliot Hurst have been provoked in questioning the very existence and future of geography as a subject *per se*!

Apart from this crossing of interdisciplinary boundaries, the 1960s political/societal movements, such as the global women's and black movements, challenged geographical assumptions of 'fieldwork', assumptions built around power relationships that either misrepresented or left out the real voices of people who were marginalized or on the edges of society. This challenge was fierce and vocal. It began to question the historical and material positioning of geographical explanations. For instance, for the millions who were on the receiving end of geographical explorations, the story was different. The Royal Geographical Society, which was founded in 1830 within the context of British colonialism, relied on geographical 'explorations', mapping of boundaries, and hierarchical classifications of 'darker continents' and 'darker races' helped to legitimize forced rule over millions. As Kobayashi and Peake in citing Livingstone suggest, "the strongest of imperial geography's metaphor was that of the moral climatic idiom," which by naturalization of racial differences according to climatic classifications placed those of the 'dark races' at the bottom of geography's terrain.

'The field' is then a complex political arena which spreads beyond a distant, accessible location 'out there'. It is in fact where power relations between the researcher and the researched are played out, whether these are defined by class, race, and/or gender. To get into what lies beyond the surface, and to understand how power shapes space and place, requires less distancing and more engagement with 'the field', and a constant search for more appropriate, inclusive methods of enquiring into the relationship between physical environments and human interaction. In this it is important to understand the historical and social contexts within which we negotiate 'the field'. For instance, as the anthropologist Johannes Fabian suggests, time differentiates and transforms how we view 'the other'. African people, for example, within the context of historical missionary practices during the 'discovery' of Africa were often viewed as 'savages' ready for religious conversion. Global political awareness over time has forced a transformation of such derogative stereotyping of many who were caught up in the story of colonialism.

This article discusses further the point that 'the field' is a political arena through a review of a feminist critique that has defied the androcentricism (consciously or otherwise fixating on the male as the center of the world) and distancing of the whole research process in seeking a return to 'the field' where 'subjective' engagement can be celebrated. This is followed by a review of development methodologies that have equally strongly attempted to include the invisible 'voices of the poor' that have previously been lost or dismissed in historically racialized spaces of a postcolonial world.

A Feminist Return to 'the Field'

As suggested earlier, one of the reasons for focusing on a feminist perspective on 'the field' is that, particularly since the 1960s, feminists have made a huge impact in redefining the disciplinary boundaries and understanding of geography. Feminist intervention in geography has to be understood within the context of an exciting, optimist women's movement of the 1970s when feminist literature (such as Kate Millet's 1970 *Sexual Politics*, Simon de Beauvoir's 1978 *The Second Sex*) brought politics into the personal, questioned public spaces, patriarchal structures and the world of work (as with Anne Oakley's 1972 *The Sociology of Housework*, Maria Mies' 1986 *Patriarchy and Accumulation at a World Scale*), paving paths toward direct collective actions (e.g., with the UK Greenham Common Women's Peace camp 1981–2000).

An important strand of feminist intervention at the time has been the questioning of how we 'find out'. Feminists argued that the processes of enquiry are essentially embedded in gendered power relationships, well summarized in a book entitled *Doing Feminist Research*. This book showed how all processes of research, starting with what is included on research agendas, right up to the end process of publication reflected a male bias of power in universities and other teaching institutions. In turn, this meant that only the research that met the criteria of a dominant male rationality received status and visibility, whereas other important topics were simply left off the agenda or dismissed/sidelined through a questioning of methodological significance.

By the mid 1980s, such challenge to the androcentrism of research processes led to a severe and intense attack particularly on the rational, 'objective' scientific methodology of 'the field'. An example is Lloyd's argument that the notion of 'ideal rationality' is embodied in seventeen century ideals of manhood rather than those of humankind, as evident in the philosophical assumptions of Descartes and Spinoza. Thus, feminists argued that knowledge in modern Western culture is verified pri marily through men's experiences. Any 'truths' generated are therefore based on a limited knowledge of reality in which they circumvent or bypass the real conditions, and the ontological foundation of gender, race, class, and culture. For instance, in apartheid South Africa, was the 'truth' the story that of the segregated black housemaid living in poverty in Johannesburg or that of a well to do white 'madam' for whom she worked?

Writers such as Nancy Hartsock and Susan Harding in 1986 and 1987 therefore argued for a feminist stand point, the framework of which is a nonandocentric, nonidealist methodology that is contextually grounded, and enables a differing understanding of social relation ships of power and powerlessness. For example, in the author's own work with poor women who make meals for migrant workers in the horrendous and vast slums of Mumbai, her 'field' was located within the context of the extreme poverty and vulnerable lives and livelihoods of the slum dwellers. Within this unpredictable and harsh world, there was no neat methodological fit. The author was also so disturbed by what she saw, it was impossible to stay at an 'objective' distant, as a woman researching circumstances of other women. A feminist standpoint gave her the confidence to adapt methodology to the particular context and to use subjectivity and feelings as evidence.

Standpoint theory, as with other politico ontological questions, drew much criticism, particularly from women who saw their boundaries of oppression as defined dif ferently. This included black and Asian women, lesbians, and women from the 'Third World' whose daily reality was shaped through race, color, homophobia, and poverty. This, they argued, was very different from the often privileged and ethnocentric backgrounds of Western academic women. This argument came home to the author sharply during the research she mentions above. While, for instance, she shared her role as a wife or a mother with the women in the slums, there was no comparison between the advantageous background in which she carried out these roles, with those of the wives and mothers from lower caste, poverty ridden back grounds who struggled daily just to feed their children.

Over the years, debates on a feminist standpoint and difference have become very complex and refined, giving rise to further subdebates such as on positionality and situatedness which have led to alternative understandings on space, place, and scale. The range is wide and stret ches from microlevel local studies, including body mapping and gendering of everyday spaces, to macro level studies on interconnectedness between regions, nations, and global linkages, supported by feminist geographers such as Massey, McDowell and Sharp, and Bell and Valentine. As we go into the twenty first century, dominant questions include those of identity and place, especially with those who have 'no place'. This includes the millions of displaced around the world through conflict, through global capitalist restructuring, through environmental destruction, and through re ligious fundamentalism that undermines women's pos ition in society.

Feminist epistemological discourses that began as a challenge to male power, scientific rationality, and 'ob jectivity' have therefore turned a full circle. Instead of a 'distancing', feminist standpoint celebrates subjectivity, thus demanding a return to 'the field' where the 'truth' tells the real story behind the lives of the 'researched' rather than a partial one. For instance, which is the partial and fuller truth behind infant mortality rates in the slums of Mumbai? Is it through the 'distant' census records of infant mortality, or is it through the grief displayed in a mother's story, which will inevitably draw a 'subjective' researcher into her life, particularly if this researcher is a woman and a parent herself?

A Postcolonial Return to 'the Field'

Like the feminist paradigm, those whose 'fields' are the worlds of post 1950s developing and transitional econ omies, question the value of 'Western' field practices and methodologies within very different and highly complex social situations of poverty. While a major issue for 'field' studies in development (which is a branch of geography specializing in mostly ex colonial economies as above) has also centered on a methodology that engages rather than distances the participants, a primary starting point of the discourse is whether 'fieldwork' should be done in the first place, particularly by researchers from the 'West'.

Underlying this dilemma is the acute awareness of how a colonial hegemonic ideology has represented the 'other' (evident in many, often erroneous, anthropological studies of tribes and 'darker' races as discussed earlier). Development practitioners and academics share 'the field' of time and space, not even a full century apart, between 'field truths' that fed the colonial imagination, and research that counts toward poverty alleviation based on meaningful cultural interactions. Fear of the intrusive nature of fieldwork, fear of exploitation, and fear of re inforcing patterns of power and privilege underpin the conflict that is so created. Therefore, since the 1980s, a very rich discourse has emerged on the ethical dilemmas of development especially from 'postdevelopment' critics of Latin America, Asia, and Africa, who question devel opment intervention of any type led from the 'West'. One such is Escobar who argues that the debates around 'development' that have emerged from the West are re inforcing the legitimacy of Western 'experts' (such as

World Bank consultants), and undermining local people and indigenous knowledge.

Related to this is the realization that 'fieldwork' (on which 'experts' feed) continues to either misrepresent or underrepresent the needs of people in developing countries, particularly marginalized people. Therefore policy and practice, which are integral to the development agenda, have at best left out those who are most burdened by poverty and inequality, and at worst subject them to bad development projects that have further impeded progress. Two influential, groundbreaking interventions came from Boserup who pointed to the omission of gender and women's significant contribution from a development analysis, and Chambers who argued for an inclusion of local voices in research through a participatory approach.

Participatory approaches are about development from 'below' rather than 'above', informed by knowledge of those who are at the receiving end of it, rather than 'experts' whose knowledge is remote. Theoretically, this enhances a fuller knowledge base and empowers grassroots participants through the research process as they gain confidence in identifying their problems and designing solutions. Participatory methodology as both theory and practice has thus allowed development studies to break away from the traditional mold of the 'objective'/expert researcher and the subject/respondent. Embedded in this is a recognition that quantitative information (e.g., surveys, census, baselines) is useful as background or complimentary data, but the real information about people's lives and livelihood comes from qualitative responses, the oral histories, the untold stories, case studies, and an understanding of the subjective where research responses are of prime value.

Since participatory approaches have generated much interest from many disciplinary strands, there are several textbooks on how to actually conduct participatory research. There is also much critique on its idealism and limitation, for example, that these approaches are still a little more than another information technique and that participation is not always equal. Nevertheless, participatory approaches over the years have increasingly demonstrated that research roles are actually integrated with goals of development. This is because the 'field' is where researchers intervene in social processes that shape other people's lives and culture. How that intervention is carried out requires introspection and critical reflection, particularly as the history of development is littered with examples where well meaning, but uncritical initiatives have left local people worse off. Because of this, critical reflection aims to include as many stakeholders involved in the process of development as possible so that the knowledge base of progress can be extended. Extended knowledge thus informs wider debates and policy decisions with lesser margins of error. In the process, participants can expect to be further empowered in shaping their own futures rather than be dictated 'from above'.

In recent years, the idea of participatory 'fieldwork' has taken on new meanings in what has become known as 'action research'. Action research relies on 'field' activity that brings together a number of stakeholders who share interests in generating an outcome. These stakeholders can include a range of local people, voluntary or funding agents, academics, and development practitioners, all of whom share a wealth of acquired lived knowledge and are best placed to generate lived theories of how to negotiate action in 'the field'. Action research is, therefore, that which is generated through both lived and critical knowledge amassed by stakeholders who act together to challenge and change their own circumstances in a manner which will generate sustainable, grassroots development.

An example of this is a women's farming project (The Banjundling Women's Vegetable Garden) in the Gambia where women farmers, local activists, interested volunteers, and the author herself as an academic have come together to develop initiatives on how to generate income by diversifying crops. The women farmers have lived knowledge of the soil, farming techniques, and cropping, while local activists, that of internal marketing and sales, and the author, as an academic, has skills which help with funding applications or writing up of reports. In this way, a differentiated group of people shared action over a farming project that over the years has generated income for some 150 previously displaced women farmers.

It is important to note that action research is often a lengthy process which is not always smooth. For instance, in the above example, the project took some 10 years to establish itself and there were many trials and errors during this time. It is also important that the project remains sustainable and can build capacity to teach others at a grassroots level. Action research therefore has a social aim in that ideally this should empower the participants. In the example given above, the women farmers involved in the original project have now helped to establish another nine such projects. They have also learned to negotiate for more land, enter markets, and diversify into animal husbandry. Action research also has a political aim in that it initiates change in power relations. For example, women in the Gambia work the land but mostly do not own any. They live in a fairly patriarchal society, often in polygamous marriages. Participation and working collectively has given them strength to address some of these structural biases that affect their daily struggle.

Discourse around developmental 'fieldwork' has therefore attempted to decolonize understandings based on historical colonial power relations which have presented European knowledge as central. It takes lived knowledge of people at grassroots level as its mainstay in 'the field'. Development geography remains concerned with questions of ethics, representation, historically tangled relationships, and methodological significance. However, at the end of the day as development is about action, full engagement with 'the field' where lived experience and knowledge create action that betters lives is of central importance.

Conclusion

Contemporary geography no longer exists as knowledge of lists, classifications, and worldly facts – acquired by traipsing around in muddy boots, mapping, and measuring of a vast expanse of homogenized space of the 'conquered' world. The geographical field in the twenty first century is much stretched since the 1960s Pattison definitions of the four traditions of geography, that is, geometrical boundary mapping spatial tradition, the area studies tradition, the man–land tradition, and the Earth science tradition.

The geographical 'field' is now understood as vast, being further extended through globalization and interconnectedness through real and virtual worlds. This has meant that difference within spaces and cultures has become increasingly apparent while at the same time globalization has brought new questions in relation to power.

Within this, 'fieldwork' remains an important part of geography, and lends it some distinct identity. However, as the feminist and developmental geographers have argued, there is a need to engage actively in 'the field' in order to open up knowledge and spaces that have often remained invisible through a search for distancing and 'objectivity'.

See also: Developmentalism; Feminism/Feminist Geography; Fieldwork; Participatory Action Research; Postcolonialism/Postcolonial Geographies; Quantitative Methodologies; Regional Geography I; Spatial Science.

Further Reading

Abbott, D. (2007). Doing 'incorrect' research: The importance of the subjective and the personal in researching poverty'. In Thomas, A. & Mohan, G. (eds.) *Research Skills for Policy and Development: How to Find Out Fast*, pp 208 228. London: Sage.
Bonnett, A. (1997). Geography, 'race' and whiteness: Invisible traditions and current challenges. *Area* 29(3), 193 199.
Boserup, E. (1970). *Women's Role in Economic Development*. London: Allen and Unwin.
Chambers, R. (1983). *Rural Development: Putting the Last First*. Harlow: Longman.
Cloke, P., Cook, I., Crang, P. *et al*., (eds.) (2004). The changing practices of human geography: An introduction. In *Practising Human Geography*, pp 1 32. London: Sage.
DeLyser, D. and Starrs, P. (2001). Doing fieldwork: Editors' introduction. *The Geographical Review* 91(1 2), January April.
Escobar, A. (1995). *Encountering Development: The Making and Unmaking of the Third World*. Princeton, NJ: Princeton University Press.
Harding, S. (ed.) (1986). *The Science Question in Feminism*. New York: Cornell University Press.
Harstock, N. (1987). The feminist standpoint: Developing the ground for a specifically feminist historical materialism. In Harding, S. (ed.) *Feminism and Methodology, Social Science Issue*, pp 157 180. Bloomington, IN: Indiana University Press.
Katz, C. (1994). Playing the field: Questions of fieldwork in geography. *Professional Geographer* 46, 67 72.
Kobayashi, A. and Peake, L. (2000). Racism out of place: Thoughts on an anarchist agenda for geography in the millennium. *Annals of American Geographers* 46(1), 73 80.
Livingstone, D. (2003). A brief history of geography. In Rogers, A. & Viles, H. (eds.) *The Student's Companion* (2nd edn.), pp 257 286. Oxford: Blackwell.
Massey, D. (1995). *Spatial Divisions of Labor: Social Structures and the Geography of Production* (2nd edn.). New York: Routledge.
McDowell, L. and Sharp, G. (eds.) (1997). *Space, Gender and Knowledge: Feminist Readings*. London: Arnold.
McNiff, J. and Whitehead, J. (2006). *All You Need to Know About Action Research*. London: Sage.
Nast, H. (1994). Women in the field: Critical feminist methodologies and theoretical perspectives. *Professional Geographer* 46, 54 66.
Roberts, H. (ed.) (1981). *Doing Feminist Research*. London: Routledge and Kegan Paul.
Valentine, G. (2001). *Social Geographies: Space and Society*. Harlow: Pearson Prentice Hall.

Field Systems and Enclosure

J. Chapman, University of Portsmouth, Portsmouth, UK

Glossary

Enclosure (Sometimes Inclosure) Legally, the division of communal land, or any land subject to common rights, between the various partners. The land did not necessarily have a physical boundary, though this was normally the result.
Field System The manner in which the land of a community is organized and run.
Furlong A bundle of strips with a constant orientation.
Hitch Crop A crop grown on part of a field which should normally have been left fallow in that year.
Levant and Couchant An arrangement whereby only stock which could be supported on the right-owner's other land during the close season were entitled to be kept on the common.
Partible Inheritance The division of property equally between the (usually male) co-heirs.
Stint The number of animals an individual could keep on the pasture or waste.
Strip (also Selion) The smallest unit of occupation in a field. In Northern Europe, typically long and narrow.

Introduction

Though field systems and enclosure have been a specific research interest of historical geographers, their legacy has profound significance for modern landscapes and the structure of modern agriculture. Whether holdings are consolidated or fragmented, the efficiency of field size and shape for modern agriculture, and even rural social structures, owe much to the details of past field systems and the way in which enclosure was carried out.

Research during the second half of the twentieth century established a broad, if not universal, consensus on a theoretical explanation of field systems, though the details of local variations and the timing of their emergence remain to be determined. On the other hand, the methods, dating, and social impact of enclosure remain controversial, notably in England where the central government left the process largely in the hands of local communities.

Field Systems

The term 'field systems' could theoretically be extended to include the latifundia and hacienda systems which historically dominated much of Mediterranean Europe and Latin America, and in which estate owners worked large areas of land with a paid labor force but let out tiny plots to tenant smallholders. Similarly, it might be extended to cover the situation common in communist Europe, whereby the land was split between state farms (run by the state with paid labor), collective farms (leased by the state to 'farmer shareholders' who worked the land cooperatively), and private plots (farmed by smallholders for their individual benefit). More normally it is restricted to the community controlled systems which developed throughout much of Europe during the Middle Ages. These show an almost infinite range of local variations, and many lack one or more of the classic features, which may be summarized as follows:

Common Fields

In the classic open field system the arable land of the community was divided into three large fields. Individual holdings consisted of many strips (a hundred or more was not unusual) theoretically scattered evenly through the fields, though in practice often irregularly distributed (see **Figure 1**). The strips were normally organized into furlongs, and bundles of furlongs made up the fields. Cropping patterns were controlled by the manorial court, and in Northern Europe typically involved a three course rotation of wheat and barley, with a fallow year to allow the fertility of the land to recover.

A two field system occurred on many poorer soils, and multifield systems existed in many communities (see **Figure 1**). Rotations, too, were subject to variations, the furlong sometimes acting as the rotation unit. Some courts allowed different crops within a single field, or permitted a hitch crop on the fallow. The fallow field was used by the landholders for pasturing their stock, as were the other fields after the harvest.

Open fields and common fields were theoretically different, the latter being subject to the communal controls and rights mentioned above, whereas the former were not. Legally the landholders in the former were merely involved in a mutual agreement not to exclude each other from their specific strips, and this carried an implied right to withdraw from the system unilaterally. In practice the distinction between the two was often unclear, and disputes arose when one party attempted to withdraw.

Common Pastures

Common pastures played a vital role in ensuring adequate fodder for the stock. They were often restricted to

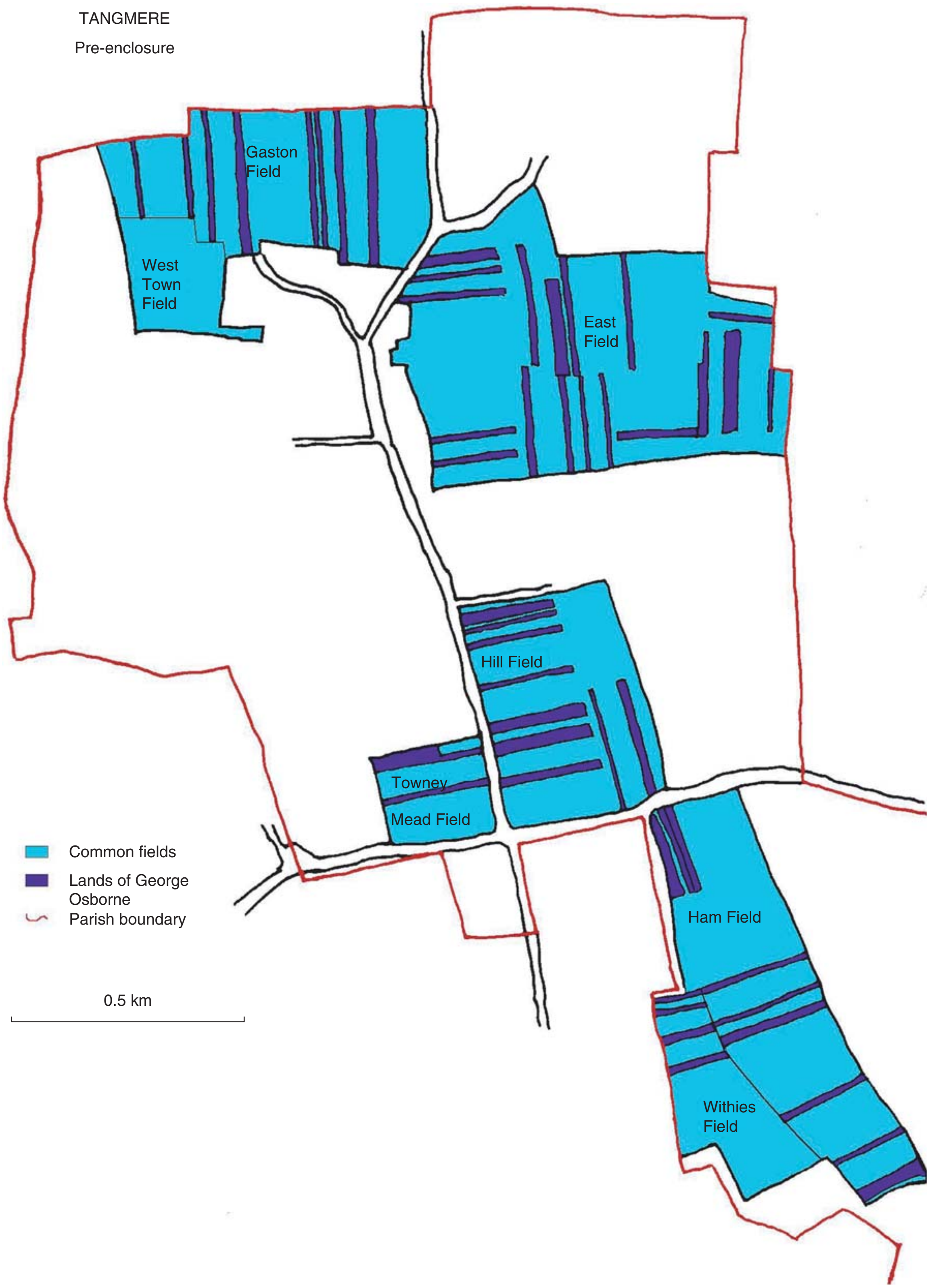

Figure 1 A typical scattered holding before enclosure: Tangmere, Sussex.

plow animals, usually oxen and horses. These pastures were normally strictly controlled, their use being confined to those who held land in the fields, with a specific stint assigned to each individual.

Common Meadows

Common meadows provided hay for supplementary feed, and were normally held in strips like the open fields. Local custom determined how many hay crops should be taken before the meadow was thrown open as a common pasture. In many areas the holdings in the meadows were redistributed on a regular basis, the partners drawing lots to determine which strips they would hold in a given year. Such lot meadows survived locally well into the nineteenth century.

Common Wastes

Common wastes (commons) were significantly different. While individuals held specific pieces of land in the fields and meadows, and known shares in the pastures, common wastes were frequently subject to a whole range of overlapping rights of indeterminate extent. The lord of the manor owned the land, but was not legally entitled to interfere with the rights of others to use it according to custom. These included the right to graze animals, and the right to take fuel in the form of wood, turf, or peat. The right to take stone for repairs to buildings, walls, and roads was normal, and other rights existed in particular communities.

Use of the common wastes caused much conflict. The detail of the rules varied, but only the owners or tenants of land in the community, together with the occupants of certain ancient tenements and anyone with a specific grant from the lord, had a legal right to use the common. In practice, stewards often turned a blind eye to unqualified locals collecting fuel or keeping poultry there.

The principal problem arose from the keeping of stock. Many larger commons were unstinted; others relied on the 'levant and couchant' rule. Commons could thus become overstocked, not least because of the possibilities of fraud. The normal seasonal buying and selling of stock made the levant and couchant rule difficult to enforce, and complaints about owners passing off someone else's stock as their own were frequent.

The Distribution of Field Systems

This has been studied in far greater detail in some areas than in others, but the broad pattern in Europe is clear. The classic form of the open field system was found in a wide band across the North European Plain from northeast France to parts of the former Soviet Union, stretching as far south as the Massif Central and the Alps, and north into southern Sweden. In Britain it occupied a distorted triangle with an apex in County Durham and a base from Dorset to West Sussex. This system was closely associated with large nucleated settlements.

The situation in Southern Europe has been subject to differing interpretations; Dion emphasized its enclosed and individualistic nature, Bloch its communal aspects. Modern thinking has tended to support the latter, while recognizing the widespread significance of old enclosed land. In this interpretation, irregular open fields, subject to a degree of communal control but lacking the rigidity of the northern system, were a frequent feature. The fields were subdivided, though typically into roughly rectangular plots rather than long strips, a difference attributed to the type of plow. The heavy plow, or caruca, plowed deeply, but was awkward to turn, so long strips were advantageous: the light weight aratrum of southern Europe stirred only the surface layers, thus preserving moisture, but cross plowing was needed, so a roughly square plot was preferred. The use of the fallow field for pasture varied from area to area, but communal pasture rights over the waste were widespread. In contrast to Northern Europe, rights were often open to those without a holding in the communal arable.

In the northwestern fringes of Europe, from northwestern Spain to northwest Scotland, a further system existed. Isolated compact farms were frequent, but many communities had a single common field. This was held in strips and received all available manure, enabling it to be cropped every year. This arable was sometimes supplemented by taking in a different portion of the waste each year, the crop relying on the accumulated fertility built up by the pasturing stock and by the ash produced from burning the natural vegetation. In this form the system was known as the infield–outfield system. Traditionally a 7 year cycle was used for the outfield, though much depended on the quality of the land.

Theories of Origins

The origins of these systems have been much debated. Gray in *English field systems* regarded the systems which he identified in Tudor England as being completely separate, seeing irregularities as evidence of a decline from an earlier perfect pattern. He offered an essentially ethnic/tribal explanation, distinguishing the infield–outfield and related systems as Celtic, the Midland open field system as Germanic, and the irregular East Anglian system as a consequence of the Danish invasions. This explanation appeared to fit the Continental evidence available at that time. In France, for example, the open field area of the northeast coincided with the area of greatest Germanic influence, Brittany showed the expected Celtic influence, and the more southerly areas could be attributed to the strong Roman influence.

Subsequent research undermined this explanation, as investigations produced increasing numbers of systems in the 'wrong' ethnic areas, making it impossible to explain these away as chance survivals of ethnic enclaves. Fur thermore, research on the Medieval and earlier periods failed to find the 'perfect' systems which Gray had predicted.

Thirsk offered an alternative explanation, whereby systems show an evolution from a more individualistic toward a more communally controlled arrangement, propelled by population pressure but constrained by environmental factors and contemporary technology. If partible inheritance was operating, a rising population would lead to diminishing farm size and hence the need either to reclaim more land or to increase productivity from the existing area. The former could only be achieved at the expense of the waste, which played a vital part in supporting the stock. Any consequent reduction in stock numbers might result in smaller quantities of ma nure, which, spread over a larger area, could result in total arable production actually falling, thus upsetting the balance of the system.

Fallowing offered an answer, since stock could graze on the fallow, maintaining numbers while simultaneously resting and manuring it. However, long narrow strips made this impossible without communal agreement on cropping arrangements, since otherwise stock could damage crops growing on adjoining strips. Where only two major field crops were grown, the optimum solution was to create three large fields, one of which could be left fallow each year in rotation and thrown open to the stock of all the partners. However, this would involve a wholesale reorganization, since otherwise anyone with land largely within the fallow field would be unable to sustain themselves during that year.

How and when such a reorganization took place has itself been the subject of academic debate. Authorities such as Campbell have produced evidence of imposed reorganization from above, for example, by the lord, and highly regular systems might imply this. However, there is also evidence of periodic redistributions, with the partners drawing lots to determine their holdings, much as in the lot meadows. This system continued well after the Medieval period in parts of Scandinavia, where it was known as solskifte, and there is similar evidence from elsewhere, for example, Northumberland.

In this explanation, the observed distributions repre sent stages in a process of evolution, with the system failing to advance to the next stage where the environ mental limits were reached or where, as in northern Scandinavia, population pressures were never high enough to require further adjustment. Thus individual block farms, single small communal fields, and the in field–outfield system would be associated with areas of limited agricultural potential – uplands, and areas of impoverished soils such as the north German heaths or the Brecklands of East Anglia. A fully developed three field system would occur only where local soils could maintain productivity with limited fertilizer input and a recovery period of 1 year in 3. The thin chalk soils of east Yorkshire and Lincolnshire would sustain yields only if fallowed every other year, hence the prevalence of two field systems, the cropped field being subdivided between the two crops.

Though this explanation is now the most widely ac cepted, ethnic factors seem to have played some part. The peculiarities of the systems in Normandy have been attributed to the influence of the Scandinavian invaders, and it has been argued that the oddities of the East Anglian systems owe their origins to Danish lordship over a subjugated Anglo Saxon peasantry.

Enclosure

Reasons for Enclosure

The open field system and its variants were essentially geared to subsistence, aiming to ensure the necessary minimum for survival each year. Pressure for its abolition occurred when maximum productivity was the aim, and thus when farming became strongly market orientated. The timing varied greatly from place to place, and often became a source of tension within a community as some saw opportunities for market exploitation where others did not. Beresford, for example, argued that enclosure in England in the later Middle Ages was driven by the major landowners' desire to remove the open fields and their small peasant proprietors to exploit the market for wool. Some central governments took a leading role, for eco nomic, political, or strategic reasons. The royal govern ment in Prussia encouraged holding consolidation as part of the modernization of the state from the 1770s on. Others were concerned at a reliance on food imports, and threats to public order from escalating food prices. The 1801 General Inclosure Act in Britain arose from such fears during the Napoleonic Wars.

Kjaergaard has argued that many open fields in Denmark became unsustainable due to declining fertility, and that there was no option but to enclose, an argument which might apply to many other areas.

Methods of Enclosure

The methods of enclosure varied greatly according to national and local laws and customs, and were affected by the type of land involved. Though the literature often concentrates on the abolition of open and common fields, common wastes frequently contributed a larger acreage. The distinction is of considerable importance, since the former was concerned with the consolidation of holdings in the existing farmed area (see **Figures 1** and **2**), while the latter was a reclamation movement, intended to in crease the area of intensively farmed land, and not infrequently increased fragmentation, since the new land might not adjoin the owner's existing holdings. Many countries of Europe treated the two processes separately, concentrating first upon the fields, which were seen as a serious impediment to productivity. The urge for waste enclosure was often driven by improvements in agri cultural techniques and the introduction of new fertil izers, which in many countries had little impact before the mid nineteenth century. Some central government intervention was aimed at one aspect, only, of the

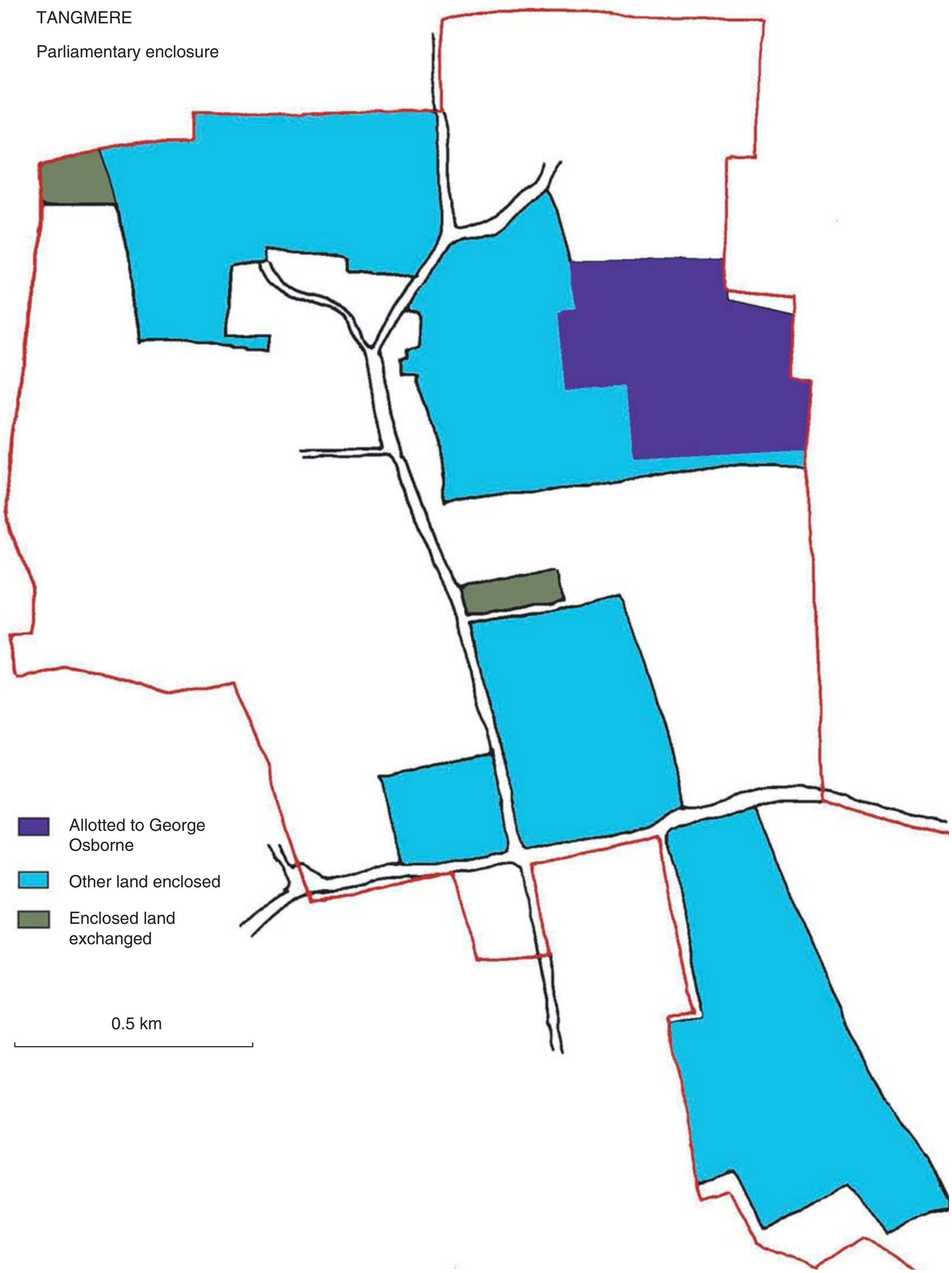

Figure 2 The same holding after re-distribution at a Parliamentary enclosure.

problem, leaving a legacy of a half completed process. In France the right to withdraw unilaterally was enshrined in law in 1793, but effective action to consolidate holdings did not begin until after World War II. In the Netherlands state sponsored consolidation was not completed until well into the second half of the twentieth century. Sweden went through a series of stages, passing a law permitting enclosure in 1749, and following this in 1757 by one reducing the number of plots in the solskifte. Finally, in 1783, an act enshrined the idea of consolidation of holdings into a single plot.

Enclosure took place over a long period, even overlapping with the creation of new open fields in the same system. Much followed the Black Death of 1347–49. Though few villages seem to have been lost as a direct result of the plague, the death rate allowed a redistribution of population, either voluntary as tenants sought better holdings left vacant elsewhere, or compulsory, as lords evicted the survivors in order to create large commercial farms. In either case the vacant land was reorganized into compact units and common rights vanished.

Enclosure was extensive during the seventeenth and early eighteenth centuries. This might be instigated by peasant communities themselves, as in several German states where they could apply to the local prince for permission to reorganize, or enforced from above, as in parts of southern Sweden, where lords carried out the reorganization, dispersing the tenant farms from the village to new enclosed compact farms. In parts of England, Lincolnshire, for example, there was widespread enclosure by agreement, often using external referees to carry out the consolidation of the land and division of the commons.

Enclosure in England and Wales

The most comprehensive attempts at enclosure in the eighteenth and nineteenth centuries occurred in England. These have been intensively studied, and have provoked great controversy over motivation and results, though much of this has focused solely on Parliamentary enclosure, and on the classic Midland open field parishes in the eighteenth century. There is an assumption in some of the literature, sometimes explicitly stated, that other means of enclosure ceased to be of significance by the later eighteenth century when enclosure by private act had become firmly established. This is a gross oversimplification. While little of the Midland core was enclosed without an act after about 1730, elsewhere non Parliamentary means might form a high proportion of the total, exceeding the Parliamentary total in Hampshire and Sussex, for example.

The means of enclosure may be classified into three: Parliamentary, formal agreement, and informal means. The first required an individual Act of Parliament or a Parliamentary Order under the General Act of 1845 and its subsequent amendments. Conventionally, agreements quoting the General Acts of 1836 or 1845 as their authority are also regarded as Parliamentary. Formal agreements required the unanimous consent of all parties, and were a written record either of the division of the lands as agreed, or of an agreement to abide by the decisions of the appointed referees. 'Informal methods' are discussed below.

Parliamentary enclosure affected about 25% of the land area of England, and 11% of Wales. Of this, 35% consisted of open field, 58% of common and waste, and the remainder of common meadow, redistributed old enclosures, and pieces of intermixed land. The proportions varied enormously from county to county, Wiltshire being closest to the national average. The first enclosure act is usually given as Radipole, Dorset, in 1604, but Arthur Young, the agricultural writer, regarded Ropley, Hampshire, in 1709 as the first genuine one. The movement picked up pace in the 1720s and 1730s, but the mean date tended to become later the further from the Midland core, ranging from the 1760s in Leicestershire through the 1810s in the West Riding and Wiltshire to the 1850s in East Sussex and Cardiganshire.

Before passing an act, Parliament required a level of support of, variously, two thirds or three quarters, but calculated by acreage owned rather than number of individuals, allowing one major landowner to override the opposition of many smaller ones. However, tenants or smaller landowners were not necessarily hostile, and many acts faced little if any formal opposition. They were needed because one or more parties could not legally sign a binding agreement, for example, because they were minors or tenants for life. Criticism has been leveled at the fairness of the commissioners who allocated the land, and there are some well documented cases of dishonesty, but these appear to be the exception; most worked fairly and honestly within the law as it stood. Small land or right owners generally got their proportionate share. The main losers were those who may have made use of the commons, but who had no right to them under the law.

The amount of non Parliamentary enclosure during the eighteenth and nineteenth centuries is not accurately known, though there is ample evidence of its significance in areas as widely scattered as Hampshire, Lincolnshire, and Cumberland. A recent estimate suggests that it contributed at least a third of all enclosure in England during this time and that this "may ultimately prove to be an underestimate" (Kain, Chapman and Oliver, 22). Even the amount enclosed by formal agreement is difficult to ascertain, since many have been lost, and others give no acreage. The proportions of field and common are equally obscure. Field enclosure should have been easier, since everyone owned known pieces of land, but

Dorset, for example, used agreements almost exclusively for commons.

An agreement, rather than an act, normally greatly reduced the costs, and smaller landowners had a stronger bargaining position, since unanimity was required. Obviously pressure could be brought to bear: the Crown agent in Norfolk threatened leaseholders with non renewal, and all opponents with the extra costs of an act, if they continued to resist. Nevertheless, many agreements show favorable treatment for less wealthy individuals, for example, by exempting them from all costs.

Informal enclosures involved various processes of varying legality, and left little direct documentary evidence; often the only evidence is that a field or common existed at one date and had gone by another, sometimes with a considerable chronological gap. One process was unity of possession, landowners gradually acquiring all land and rights by allowing tenancies to lapse and buying out freeholders. Lords could also legally enclose parts of the common, providing sufficient was left for the right owners, though wrangling over what was sufficient was frequent. Unilateral withdrawal of one's land from a common field, a legal grey area, was another. Finally, much enclosure was undoubtedly illegal. Individuals simply encroached commons, fencing in parts of them and denying access to others. This was a source of much hostility in some Midland areas, where commoners saw an essential part of their economic welfare being diminished, though in upland areas where commons were very large it was often accepted without protest.

Recent evidence suggests that consolidation of fields often took place at tenant level. Tenants renting whole furlongs or fields from the various owners could ignore communal restraints, though technically these still existed. The intermixed ownership vanished later as owners sold up.

In summary, the process of enclosure was far more complex than is often suggested and general theories about its origins, motivations, and results need to be approached with great caution. What applied to a Parliamentary enclosure of fields in a Midland parish in the 1740s might have little relevance to a moorland enclosure in 1800s Yorkshire, or to a piecemeal enclosure in 1860s Sussex.

See also: Agrarian Transformations; Feudalism and Feudal Society; Historical Geographies, Rural; Peasant Agriculture.

Further Reading

Baker, A. H. R. and Butlin, R. A. (eds.) (1973). *Studies of field systems in the British Isles*. Cambridge: University Press.

Chapman, J. (1987). The extent and nature of Parliamentary enclosure. *Agricultural History Review* 35, 25 35.

Chapman, J. and Seeliger, S. (2002). *Enclosure, environment and landscape in southern England*. Stroud: Tempus.

Dahlman, C. J. (1980). *The open field system and beyond: A property rights analysis of an economic institution*. Cambridge: University Press.

De Moor, M., Shaw Taylor, L. and Warde, P. (eds.) (2002). *The management of common land in north west Europe, c.1500 1850*. Turnhout: Brepols.

Gray, H. L. (1915). *English field systems*. Cambridge, MA: Harvard University Press.

Kain, R. J. P., Chapman, J. and Oliver, R. R. (2004). *The enclosure maps of England and Wales, 1595 1918: A cartographic analysis and electronic catalogue*. Cambridge: University Press.

Kerridge, E. (1992). *The common fields of England*. Manchester: University Press.

Kjaergaard, T. (1994). *The Danish Revolution, 1500 1800: An ecohistorical interpretation (trans. D Hohnen)*. Cambridge: University Press.

Neeson, J. M. (1993). *Commoners: Common right, enclosure and social change in England, 1700 1820*. Cambridge: University Press.

Rowley, T. (ed.) (1981). *The origins of open field agriculture*. Beckenham: Croom Helm.

Thirsk, J. (1966). The origin of the common fields. *Past and Present* 33, 142 147.

Turner, M. (1980). *English Parliamentary enclosure*. Folkestone: Dawson.

Relevant Websites

http://hds.essex.ac.uk
The Enclosure Maps of England and Wales, 1595 1918, AHDS History.

Fieldwork

F. J. Bosco and C. M. Moreno, San Diego State University, San Diego, CA, USA

Glossary

Field/Field Site A singular location traditionally marked off in time and space where a researcher operates. From a critical perspective, the demarcation of the field is fuzzy and unstable.

Politics of Fieldwork The questions and issues arising out of the uneven relations of power existing between researchers and research subjects in the field.

Positionality The recognition that a researcher always operates and produces knowledge in relation to multiple relations of difference, such as gender, ethnicity, class, age, and sexuality.

Reflexivity The process of critically reflecting about oneself as a researcher.

Situated Knowledge The knowledge marked by the context in which it is produced and the partiality of the knowledge producer.

Introduction: Fieldwork as a Method of Geographic Inquiry

Fieldwork typically involves the collection of primary data through a variety of means. The practice of geographic fieldwork can range from the collection of data through systematic observations of places and landscapes and inquiry based on predesigned research instruments, such as surveys, questionnaires, and structured interviews to more open ended, interpretive, and intimate forms of inquiry, such as participant observation, ethnography, focus groups, the collection of life histories and participant diaries, other types of listening strategies, and even some forms of archival research. Surveys and structured questionnaires given to particular populations are typically associated with the fieldwork practices of quantitative and positivistic human geography, while systematic observations of landscapes can be positioned in the fieldwork domain of traditional, regional, and cultural geography. In both cases, such ways of doing fieldwork are considered to be more detached approaches to data collection in which the researchers remain in control of the questions asked and the features recorded and mapped. On the other hand, more participative approaches to fieldwork such as ethnography place the researchers in more intense contact relative to the research subjects. Ethnography and other participatory methods of fieldwork are typically associated with more contemporary and critical approaches in human geography, including Marxism and feminism. However, no philosophical tradition is necessarily linked to a particular way of doing fieldwork. For example, feminist geographers conduct surveys and collect both quantitative and qualitative data through fieldwork, yet the final research product can still be feminist and critical; the same (or the reverse) can also be the case regarding the association of particular philosophies of geography with particular fieldwork practices.

Regardless of the way in which fieldwork is conducted, the majority of human geographers today acknowledge the subjectivity of fieldwork inquiry. No single method of conducting fieldwork provides unmediated, unbiased, and privileged access to the topic, place, and people under investigation. Presently, human geographers recognize that researchers play a defining role in the constitution of the field as a site of inquiry, in the act of recording and interpreting geographic observations and information, in the transformation of such observations and information into geographic data, and in the translation of fieldwork data into a final written product that follows certain accepted (though also dynamic) academic conventions. Overall, the current literature in human geography suggests that fieldwork is as much an analytical, technical, reflexive, and interpretive endeavor as it is a social, relational, political, and emotional one.

The practice of fieldwork in human geography has evolved in relation to broader philosophical changes in the discipline and together with a widening of the scope of inquiry of human geography. As human geographers embrace ever more diverse subjects of study from a number of philosophical and theoretical perspectives, the nature of geographical fieldwork continues to change, reflecting current scientific, sociopolitical, technical, and moral issues. To begin with, there are more geographic fields in which to conduct fieldwork than ever before. Presently, fieldwork in human geography ranges in settings and scales of analysis from the urban to the rural, from the city to the neighborhood, from the village to the agricultural field, from the mundane to the exotic, from the home to the workplace, and from the corporeal, emotional, and affective to the virtual spaces of the Internet and the media landscapes of television and cinema, to name but a few. Moreover, contemporary human geographers embrace the notion that geographical fields are no longer somewhere 'out there' to be explored by a detached researcher/observer who is required to leave a place of work to begin fieldwork – and then return home or the office to write about findings in the field. It is now common to hear geographers assert that one is always already in the field. This assertion has done away with the notion that it is possible to separate and bound the field ‘out there’ in space and time, and furthermore, hints at the notion that the boundaries that separate scholars and their subjects of study are also difficult to delineate.

The last three decades of fieldwork in human geography have seen a dramatic change in the ways human geographers approach the ‘doing’ of fieldwork itself. Particularly relevant has been the recognition of the role of the geographer’s positionality (i.e., the geographer’s own and personal location in the field, and in a place, in relation both to self and to others) in the generation of geographic data and knowledge. This change also has been described as a transition from geographers being in a position of detached, disembodied, and neutral observers and recorders of facts and data about people and places to one in which geographers recognize their own positions of power in the context of social relations and the partiality and subjective nature of the data collected in the field and of the knowledge generated through the practice of fieldwork. In the field, the geographer is never fully inside or outside; as many have argued, in the field, the geographer occupies the spaces of betweenness. In sum, geographers confronted with the prospect of doing fieldwork today are presented with the challenging tasks of thinking about what precisely fieldwork is (and how is it defined), how to conduct fieldwork, and how to conduct oneself in the field (given the recognition of our own positionality), and how to deal with the relations between the constitution of the field and the constitution of the self.

The Evolving Contexts of Fieldwork in Human Geography

From Explorations to Simulations

The concept and practice of fieldwork have had a long, somewhat contested, history in the development and evolution of human geography. Vestiges of fieldwork as we know them today can be traced all the way back to the European ages of navigation and exploration during the sixteenth, seventeenth, and eighteenth centuries. These periods were marked by the development of new travel and shipbuilding technologies as well as advances in navigation and cartography. These developments assisted European nations in their search for new trading routes, trading partners, natural resources, and trade goods. The importance of these early explorations in the development of geography as a discipline should not be underestimated. European exploration not only led the way to encounters of far off people and places, but also gave rise to the need to map previously unexplored seafaring paths and lands. Throughout this time, the thirst for exploration and the power of mapping proved to be the mainstays of Western society and part of the defining characteristics of early geography.

The emergence of geographical societies in Europe, North America, and even South America continued promoting geography through exploration and fieldwork. By the nineteenth and early twentieth centuries, the discipline of geography solidified its presence across European and American universities. The activities of professional organizations altered the development and importance of field research and field data in geographic research. The challenge put to many regional geographers was to develop more refined mapping, surveying, and descriptive techniques. In the United States, a notable geographer in the early twentieth century was Carl Sauer, who claimed geography a “science of observation.” Sauer set out surveying and performing numerous regional descriptions of the Earth’s natural and cultural landscapes. Methodologically, he argued that the best way for geographers to know the Earth’s surface was to go out, perform on site investigations, and meticulously describe what they saw. These sentiments were most evident in his seminal 1923 paper ‘Morphology of landscape’ and his 1925 article ‘The survey method in geography and its objective’, where he outlines that regional studies in geography should be part scientific and part holistic, but always field based.

Many regional geographers regarded the field as their primary source of data, but others felt that the observation and description of landscapes and regions produced discrete and particularistic geographies. Over time, the focus shifted and geographers moved away from the regional approach, finding more applied ways to use their field data, such as informing governmental policies on land use issues. With the rise of positivistic geography and spatial science in the 1950s and 1960s, regional geography was criticized for its lack of scientific rigor. There were also rumblings from quantitative geographers that Sauer’s survey approach was merely an adventure in sightseeing. In response, human geographers began to pursue quantitative research platforms consisting of nonfield and secondary data sources analyzed through statistical methods and modeling. This movement away from field research did not last for long, however. During the late 1960s, behavioral geographers reincorporated field methods into their scientific and mostly positivistic research toolkit. Behavioral geographers pursued behavioral experiments outside the laboratory; the idea was to control conditions in the world as if it were the laboratory. Thus, behavioral geographers used field simulations to collect real world data (even if it was controlled) and to measure and generate scientific laws about human behavior, spatial learning, and decision making, spatial knowledge, and spatial cognition.

Finding and Questioning Meanings in the Field

During the 1970s, with the advent of humanism in geography, a new type of geographic fieldwork, more participatory in nature, took on a much more prominent role. Humanism drew much of its theoretical toolkit from phenomenology. As such, fieldwork at this time moved away from being an experimental spatial scientific venture in search of laws to a qualitative and interpretive one in search of cultural and spatial meaning. Humanist geographers practicing fieldwork turned Sauer's cultural landscapes into layered social texts describing people's lives, and pursued hermeneutic interpretations. Humanists emphasized the subjectivity of both the geographer and the observed phenomena. Research questions turned from the land to the people. Field research became a more intimate way in which to be with people and learn about how people assign meaning to the social relations and places that make up their lives. Humanist geographers believed that in order to understand the world, research needed to be more geared toward trying to understand how people perceive the world. Humanists were also interested in the social and spatial construction of fieldwork, and as a result brought forward important discussions that questioned the experience of space, place, and landscape that were shared by geographers and others they sought to study.

To get at this type of experiential information, human geographers began employing more refined uses of qualitative research methods in their fieldwork. Human geographers also borrowed from other disciplines, such as anthropology and sociology. For example, in anthropology, Branislaw Malinowksi's notion of 'being there' (in his case, among the Trobriander Islanders) brought about a fundamental shift in the way in which social scientists working in far off places, including human geographers, practiced fieldwork. Malinowski's notion of 'participatory fieldwork' challenged the Sauerian School that believed surveys of cultural landscapes should be done through detaching oneself from the landscape they studied. For Malinowski, the only way to come to a more holistic understanding of a culture or place was to live among the 'natives' and 'places' for a duration of time. Malinowski's 'going native' not only speaks to the importance of long term field research projects and first hand participatory observations, but also opens an important dialog about researcher positionality as well as discussions about the importance of objectivity and what that means to the practice of field based research. Also in anthropology, Clifford Geertz took Malinowski's notion of eye/I witness accounts in field research toward a 'thick description' approach where the goal for the ethnographer is to engage with and describe the minutiae of daily, mundane, cultural, and place based activities.

The eruption of postcolonial theory in the late 1970s (from the work of Edward Said and others) also forced some human geographers to move beyond the relationship between methodology and representation in their fieldwork. Initially, postcolonial thinking showed fieldwork as an active extension of the discipline's colonial thinking and put forth a challenge for human geographers to be more introspective about the power relationships embedded in their research. This included critically examining the role of power and ethics in the production of information and knowledge emanating from their research. In doing so, human geographers took on serious discussions of power, ethics, and representation in connection to fieldwork. These included examining the language and methods human geographers use, the motivations and sources of funding for their work, as well as the nature of involvement that research subjects have in fieldwork design and practice. The contention was that fieldwork, or the embodied act of documentation, interpretation, and articulation of the world, has the potential to take away people's rights to their own history, land, communities, as well as their ability to problematize and come up with solutions to their own problems. In many ways, postcolonial thinking dismantled the long held Western belief that in order to save the world, human geographers needed to speak for the world and control it. Postcolonial thinking still plays an important role in the evolving context of fieldwork in human geography today. It raises serious questions about the ownership of field research and the need for field researchers to examine the social, economic, emotional, environmental, and political impact that fieldwork has on the people and places they study. While not always met with open arms, it forces many field researchers to be more caring and responsible in their work and seek out alternative and more collaborative approaches.

Fieldwork as Politics and the Politics of Fieldwork

Much like the humanist and postcolonial inspired geographers, Marxist geographers have radically altered the way in which geographers have come to understand the importance and practice of fieldwork. Marxist geographers are keenly aware of the injustices generated by a world that privileges capitalist relations and profit margins, and thus study and often interpret field sites as hegemonic landscapes of domination and oppression. For Marxist geographers, fieldwork can be a means to uncover injustices and to bring about social change; in this case field sites can be landscapes of resistance. Because Marxism is inherently a political project that is aimed at imagining and creating a more just social world that goes beyond exploitative capitalist relations, Marxist geographers see fieldwork as a method of engagement for social activism. For instance, Marxist geographers have recently shown an interest in engaging with social movements, particularly anticapitalist and antiglobaliza tion movements, throughout the world. This interest is a prime example of the meshing of their political concerns with fieldwork practices. At the same time, other Marxist geographers advocate more involvement with issues of economic injustices that are more local and embedded in the material conditions of daily life of disempowered people in the Global North, who may just happen to be co workers, friends, or neighbors. For example, Marxist geographers have argued that even the universities where geographers work can be the sites for fieldwork research that works for social change if geographers engage with the struggle of lower paid workers (such as maintenance crews, janitorial personnel, and other workers who per form essential tasks that allow for universities to func tion) for better wages and job protection.

Together with Marxism, feminism has proven to be one of the most important philosophies from which human geographers have rethought the practice of fieldwork, though the relations between Marxists and feminist geographers have been many times strained. A main and enduring concern of feminism regarding fieldwork is an examination of how gender relations and other differences, such as race, ethnicity, sexuality, and (dis)abilities, shape field practices and the politics of field research. Feminists also are concerned with critically analyzing the conditions that enable the production of knowledge in the field. One of the initial tasks of feminist geographers regarding fieldwork was to challenge the masculinist notion of objectivity that surrounded most research in human geography. Epistemologically, mas culinism is related to ways of knowing that uncritically define objectivity through a hierarchical separation be tween researcher and research subjects, and an associ ation with positivistic thinking and practice that relies on detached observation and empiricism. Feminists call these ways of thinking and practicing science masculinist because of their association with the dominant subjects of modern social science, which for most of the twentieth century were embodied by white, heterosexual, and bourgeois males. Regarding fieldwork in particular, feminists argue that traditional fieldwork practices in human geography, such as those written and practiced by Carl Sauer and his followers in the earlier part of the twentieth century, were masculinist because they placed the observer at a hierarchical distance from the subject of study – a higher vantage point. From that elevated vantage point, the geographer (unmarked, but assumed to be male) could observe the world as it is and through his 'objective' and 'detached' observations reveal the land scape's hidden order. Feminists have been critical of this type of traditional way of surveying landscapes as a masculinist way of conducting fieldwork and seeing the world. In addition, they have leveled criticism on other more critical human geography approaches, such as Marxism, because of their neglect of issues of gender, race, and sexuality in geographic research, and their privileging of places of production as sites for research (e.g., the factory floor) over other sites, such as the home, where issues such as family work and family care can be studied in the context of, for example, female, unpaid work.

Feminist geographers also see fieldwork as a political project that is is much about interpreting social change as it as about effecting social change – here, feminist geographers come closer to the aims of Marxist geog raphers. Because both feminism and Marxism are con cerned with larger structures of social oppression, such as class and patriarchy, Marxist and feminist geographers have also clarified that conducting fieldwork often in volves broader motivations than the study of a single site. Along these lines, both Marxist and feminist geographers have worried about the limitations of purely interpretive field research, because they understand that the processes that structure class, gender, racial, and even sexual re lations operate at multiple scales. Both Marxists and feminist geographers have pointed out that there are important methodological implications resulting from this move: geographers who focus on the intricacies of the local (i.e., the site of fieldwork) should not lose track of the impact of broader, nonlocal social relations and processes in their research projects.

To the political aims of fieldwork that they share with Marxist geographers, feminists also add a concern with the politics of fieldwork, here coming closer to the re search concerns and intentions of postcolonial thinkers. This concern includes not only attention to the power relations between the researcher and the researched and the conditions that enable the practice of research and the production of knowledge in the first place, but also to the politics of representation surrounding fieldwork (i.e., the challenges posed by being in the field and speaking with and for others who many times are in unprivileged positions of power relative both to dominant structures in society and to the researchers themselves). The critical attention that feminism grants to issues of difference, positionality, and reflexivity in fieldwork have led feminist geographers to cast doubt on the insider status of anyone in the field (including women) and even to question the researchers' own right to speak for others. Along these lines, feminist geographers directly linked the practice of fieldwork in geography to broader debates surrounding the politics of representation in con temporary postmodern and critical social and cultural theory. Engagement with this debate made some geog raphers question their ability to represent others through fieldwork, regardless of how reflexive and critical it is, and some suggested even retreating and withdrawing from social research altogether. This challenge, however, has also encouraged geographers to establish greater connections with the communities, people, and places that they study, enriching fieldwork based social research. Through fieldwork, feminist geographers continue at tempting to break down the boundaries between aca demics and the people that geographers want to represent. Along these lines, feminist geographers have been actively pursuing more democratic forms of field work practice. Such democratization of fieldwork and field knowledge involves, for example, presenting the research subjects with the texts produced by geographers' experiences in the field to ascertain participants own opinions about the geographers' interpretations. Other geographers have experimented with different and more accessible forms of writing that are attuned to different audiences beyond the academy, such as community groups and activists. Finally, and in tandem with the political orientation of feminism, some geographers have connected their academic research to the actual practices of political activists who work for social change.

Conclusion: Some Present Challenges for Fieldwork in Human Geography

In recent years, not only has the practice of fieldwork changed in relation to a wider theoretical engagement in human geography, but precisely because of that engage ment the study of fieldwork itself has enjoyed increased attention as well. Critical discussions have surfaced in relation to the constitution of geographical fields, how they are produced, how human geographers come to know/understand, affect, and are affected by them. Moreover, the thinking and practice of fieldwork in human geography have not remained tightly bound to any one philosophical tradition. Rather, geographers' engagement with different philosophies over time has helped produce new hybrid forms of field practices and new geographies to investigate – witness, for example, the recent emergence of research on emotional geog raphies and the attention paid to the emotional and af fective fieldwork, a move that cannot be understood without reference to the interactions between humanist, feminist, and post structural thinking in geography.

The critical discussions of recent years present geographers with invaluable knowledge regarding field work that would not have been available just a few decades ago. At the same time, human geographers are confronted with new challenges regarding fieldwork in the not so distant future. There are still some uncharted fields that pose new questions and require more reflection and more action. For example, there have been some good discussions lately about the privileging of the tex tual over the visual in fieldwork. Most of the techniques involved in field based research involve verbal exchanges, and result in textual information, but more and more geographers are engaging with visual methods in field work, from the production of video diaries with partici pants to the production of films as the final output of field based research. While visual methods are enjoying increasing attention in human geography, there is still the question of whether the final work of field based visual research and interventions, such as a movie or a docu mentary, is considered a valid academic (and scientific?) endeavor, an issue which goes right back to questions regarding the production of knowledge and the politics of fieldwork.

Finally, some geographers have recently argued – rightly so – that fieldwork in human geography has been defined, at least implicitly, in terms of the co presence of bodies, that is, we have privileged bodily presence and bodily experiences in the practice of fieldwork. But new questions are arising about the possibility of fieldwork through electronically mediated communications, in particular through the Internet and the World Wide Web. Not much has been written in human geography re garding human geographic research of online and cyber communities, but there are ample opportunities to move research forward along these lines. For example, recent developments in the area of emotional geographies pose new questions about the role that emotions play in re search conducted in virtual fields/spaces. At the same time, researching online groups and cyber communities present new practical, methodological, and ethical chal lenges to human geographers wishing to engage with participants who seem so close and, yet, could be phys ically very distant – but again, not necessarily so. For instance, there is the issue of conducting fieldwork from one's own office, from home, or from a café with Internet access, and the possibility of creating a new generation of 'armchair geographers', albeit in a completely different way than those who were called armchair geographers in the 1960s. Does this matter or does it make a difference, given that geographers already agree that we are always in the field and that the boundaries of the field are always fuzzy and contested? Furthermore, there is the issue of overt or covert research when dealing with online com munities. For example, does a geographer join an online community, perhaps through a chat room, and identify himself/herself as a researcher and engage with partici pants in a kind of cyber participant observation – which in turn can be either active or passive? Or does the re searcher assume a new cyber identity (a practice very common in the online world of chat rooms and cyber communities) and actively participate in the online life of a group in a kind of cyberethnography? Just thinking about these two issues (and there are many more) poses incredibly powerful questions regarding the ethics of research in human geography, in particular regarding the overt/covert and passive/active distinctions in field based research. Moreover, besides the important personal and professional ethical considerations, there are institutional issues arising as well, specifically regarding requirements about informed consent and others that have been put in place by institutional/ethical review boards as a way to protect research participants in situations of unequal power relations. In sum, these are just some of the new questions and challenges that present geographers with an opportunity and requirement to keep redefining fieldwork as one of the central methods of geographic inquiry.

See also: Content Analysis; Cross-Cultural Research; Discourse Analysis; Embodied Knowing; Emotional Knowing; Ethical Issues in Research; Ethnography; Feminist Methodologies; Gender and Rurality; Heritage; Participatory Action Research; Situated Knowledge, Reflexivity.

Further Reading

Cultural Geographies (2003). *Special Issue on Field Cultures. Cultural Geographies* 10, 251.

Clifford, J. and Marcus, G. (1986). *Writing Culture: The Poetics and Politics of Ethnography*. Berkeley, CA: University of California Press.

Crang, M. (2003). Qualitative methods: Touchy, feely, look see? *Progress in Human Geography* 27(4), 494 504.

Cupples, J. (2002). The field as a landscape of desire: Sex and sexuality in geographical fieldwork. *Area* 34(4), 382 390.

Driver, F. (2000). Editorial: Fieldwork in geography. *Transactions, Institute of British Geographers* 25(3), 267 268.

England, K. (2006). Producing feminist geographies: Theory, methodologies and research strategies. In Aitken, S. & Valentine, G. (eds.) *Approaches to human geography*, pp 286 297. London: Sage.

GR (Geographical Review) (2001). *Special issue on doing fieldwork. The Geographical Review* 91(1 2), iv XX.

Marcus, G. E. (2002). Beyond Malinowski and after writing culture: On the future of cultural anthropology and the predicament of ethnography. *Australian Journal of Anthropology* 13(2), 191 199.

Nast, H. J. (1998). The body as place: Reflexivity and fieldwork in Kano, Nigeria. In Nast, H. J. & Pile, S. (eds.) *Places through the body*, pp 93 116. London: Routledge.

PG (Professional Geographer) (1994). *Special issue on women in the field: Critical feminist methodologies and theoretical perspectives. The Professional Geographer* 46(1), 54 102.

Powell, R. (2002). The sirens' voices? Field practices and dialogue in geography. *Area* 34(3), 261 272.

Pratt, G. (2000). Research performances. *Environment and Planning D: Society and Space* 18(5), 639 651.

Rose, G. (1997). Situating knowledge, positionality, and other tactics. *Progress in Human Geography* 21, 305 320.

Rundstrom, R. A. and Kenzer, M. S. (1989). The decline of fieldwork in human geography. *The Professional Geographer* 41(3), 294 303.

Sauer, C. (1924). The survey method in geography and its objectives. *Annals of the Association of American Geographers* 14(1), 17 33.

Sauer, C. (1925). The morphology of landscape. *University of California Publications in Geography* 2, 19 54.

Sundberg, J. (2003). Masculinist epistemologies and the politics of fieldwork in Latin Americanist geography. *The Professional Geographer* 55(2), 180 190.

Relevant Websites

http://www.emporia.edu/socsci/fembib/
Gender in Geography Bibliography: K, Department of Social Sciences at Emporia University.

http://www2.glos.ac.uk/gdn/disabil/fieldwk.htm
Projects: Fieldwork in Geography, Earth, and Environmental Sciences Higher Education Curriculum, The Geography Discipline Network (GDN).

http://www.uvm.edu/~mcope/QRSG/index.html
Qualitative Geography Specialty Group's Qualitative Research Bibliography, Meghan Cope, Qualitative Research Bibliography, Department Of Geography, University at Buffalo, The State University of New York.

Film

C. Lukinbeal, Arizona State University, Tempe, AZ, USA

Glossary

Crisis of Representation Refers to the debates in the late 1980s and early 1990s which challenged the centrality of mimesis in geography. Mimesis is the belief that absolute realism through representation is possible.
Cultural Economy Refers to sectors of the economy which produce culturalized goods laden with symbolic meaning. With regard to film, it relates to an approach where the focus is on the geographic processes and practices of production, distribution, and consumption.
Cultural Studies Broadly relate to the study of cultural texts, whether literary or nonliterary, and how these texts constitute or challenge meanings of culture, society, identity, and social–spatial practices. With regard to film geography it can relate to performative logics, scripting, 'ways of seeing', and hermeneutics.
Hermeneutics It is the study of theories and methods for interpreting any cultural texts, whether in the form of a book, film, or landscape.
Relativism As it relates to film geography, it refers to the idea that any individual 'reading' of a film text is only one interpretation among an infinite number of interpretations.

Introduction

During the last quarter century there has been a proliferation of articles and books on film geography. Although approaches vary considerably, recognition of the need to increase our understanding of the power of film in the production of meaning, identity, culture, and economy are generally shared. Geographers began examining broader ramification of the media is the book *Geography the Media and Popular Culture* and *Place Images in Media*. According to Burgess and Gold, popular media is important to geography because it mediates social knowledge, reinforces ideological constructions of the status quo, and is an active agent of hegemony. In *Place Images in Media*, the essays focused on addressing the production of meaning and the role of filmmakers in the depiction of place. *Place*, *Power*, *Situation and Spectacle* was the first book by geographers to focus strictly on film. This collection of essays offers a diverse array of approaches to film, ranging from semiotics, literary theory, film criticism to transactionalism. More recently, geographers have engaged the role of film as it relates to the city, mobility and identity, pedagogy, the audience, landscape, political geography, economic geography, and the frontiers of cinematic geographies. This article provides an overview of this burgeoning subfield, and to do so, it explores the roots of film geography as it relates to humanistic and Marxist criticism toward the dominance of quantitative and regional geography in the 1960s to 1980s. Then, it also examines how the 'cultural turn' and 'crisis of representation' affected geographic engagement with film in the early 1990s. Finally, it frames a discussion of current film geography research through the use of the categories 'cultural studies' and 'cultural economy.'

Roots of Film Geography – Humanism and Political Economy

In J. K. Wright's 1947 presidential address to the Association of American Geographers, he developed the idea of 'geosophy', or the study of everyday texts and images found in "books of travel, in magazines and newspapers, in many a page of fiction and poetry, and on many a canvas" (Wright, 1947: 10). While the term never caught on, the idea that geography should expand its purview and explore the *terra incognitae* of popular texts and images was well before its time. It was not until the late 1960s and early 1970s that humanistic and Marxist geography began to challenge the dominance of regional and quantitative geography. While these approaches differ greatly, both saw quantitative and regional geography as insufficient when it came to explaining issues of ideology, agency, imagination, perception, and power relations.

According to Buttimer, humanism in geography strives to restore a human element to a field where scientific objectivism has dominated. In his often cited lines, Wreford Watson argued quite succinctly for a humanistic approach.

> Not all geography derives from the earth itself; some of it springs from our idea of the earth. This geography within the mind can at times be the effective geography to which men adjust and thus be more important than the supposedly real geography of the earth. Man has the particular aptitude of being able to live by the notion of reality which may be more real than reality itself. (Watson, 1969: 10)

Early forays into media, text, and representation by humanistic geographers focused on literature and painting rather than popular culture and film. By ignoring popular culture, Burgess and Gold argued that it gave geography the appearance of being an elitist pursuit wherein popular culture was not worthy of inquiry. One notably exception to this is the work of J. B. Jackson whose influence on humanistic geography via landscape studies encouraged geographers to examine everyday items like film and popular culture and its impact on vernacular landscapes. In the now classic book, *Interpretations of Ordinary Landscapes*, Donald Meinig examined the role of cinema in forming three American symbolic landscapes. He suggested that the cinema was an "unprecedently powerful propaganda medium" that aided the spreading of the idealized image of California's suburbs to the nation and the world (Meinig, 1979a: 171). From this perspective, cinematic representations create and perpetuate symbolic landscapes. Meinig suggested a starting point to understanding how landscapes become symbolic may lie in inventorying their depiction in Hollywood's Golden Age (1920s to 1950s). He believed that such an inventory of sets and backlots would allow for ready inferences, and act as a foundation for understanding how cultural meanings are appropriated onto specific landscapes. The newer manifestations of humanism are discussed later in this article, but first attention turns to the roots of the political–economy approach to film geography.

A central component to Burgess and Gold's book was Marxist and other political economy approaches to film texts. Burgess and Gold (1985: 1) argue that "the media have been on the periphery of geographical inquiry for too long." They argued that media research in geography was polarized with Americans using "…cognitive and social psychology for both theory and methodology" (Burgess and Gold, 1985: 7) and Europeans employing critical theories that engage ideological, social, and economic praxis. While this is a generalization of academic trends of the time, Kennedy and Lukinbeal argued that this simplification missed an important distinction between behavioral and environmental perception research. Where the former focuses more on quantitative and reductionist theories and methods, the latter is rooted more firmly in qualitative humanistic analysis. In the 1980s and 1990s, transactionalism, a branch of cognitive psychology and environmental perception, provided a useful framework to engage transactions between individuals and media landscapes, film structure, and rhythm, and person–environment relations within a film's narrative.

Early political economy approaches drew from Marxism and the Frankfurt School to examine the role of ideology within representations. Revealing the spatial dimensions of hidden ideologies and power relations, as well as contesting realism in documentary films, are key aspects of this early work. Natter and Jones essay on *Roger and Me* attacks all three by first arguing that Michael Moore's is more objective than most filmmakers, in that he clearly states his ideology and lets the viewer decide how to deal with it. They contrast how public space versus private space is reified by class, race, power, and economics.

Cultural Turn and the Crisis of Representation

In the late 1980s, as cultural studies experienced a 'spatial' turn, geographers began experiencing a concomitant 'cultural' turn. New theories around the concept of postmodernity placed an importance on space and spatial activities, and in geography, made the discipline reappraise how we conceptualize the concept of culture. According to Aitken and Zonn (1994: 2), postmodernists see little difference "between 'real life' and 'reel life'," and because of this "cinematic representation needs to be a key part of geographic investigation (Aitken and Zonn, 1994: 2)." Films are not merely images, but rather actively reconstruct our daily lives. According to Denzin, films are social texts that connect an individual's life world with the larger "textual cultural systems which create and impose their meaning on everyday life" (Denzin, 1991: 150).

'New' cultural geographers emphasized nonmaterial cultural formations, like film, in their analysis. In contrast, Sauer School and Marxist geographers have traditionally emphasized material cultural formations. This distinction led to the 'crisis of representation' debate which centered around mimesis or the belief that an accurate representation of the world is possible. Harvey argued that "in the final analysis" film is "a spectacle projected within an enclosed space on a depthless screen" (Harvey, 1989: 308). For Harvey, material conditions should always be privileged over representations, a normative belief in much of academic geography. The belief is that film is a re presentation of an absolute material reality that pre exists, and as such it is a secondary rather than primary source of spatial data, information, knowledge, and wisdom. To contest this normative belief with its privileged position of material space, new cultural geographers and post structuralists stressed the importance of representations and how they play an active role in producing and reproducing reality.

Lukinbeal argued that the backlash of these debates was to link film to the negative end of each binary, consigning it to the fringes of cultural geography research. Dixon and Grimes suggest that the distinction between 'reel' and 'real' is an arbitrary binary lacking analytical power, and these binaries constrain our potential knowledge on film. Furthermore, a focus on representation occludes the other possible lines of film geography inquiry.

Cultural Studies and Cultural-Economy Approaches

Current research in film geography can be loosely grouped into two categories: cultural studies and cultural economy. These categorizations are used as a pedagogic devices rather than as immutable structural classes. In reality, research in film geography elides these categorical differences. A cultural studies perspective explores how film, space, and meaning are mutually constituted when it comes to phenomena, such as cultural identity, race, gender, and sexuality, and even concepts, such as the city and geopolitics. Some cultural studies approaches have sought to move beyond textual readings of film. Along these lines, Carter and McCormick (2006) argue that while research on film focuses on representational analytics that "rehearse or disrupt particular discursive codes and scripts" it also "involves a consideration of how they amplify the passage and transition between affect and emotion (Carter and McCormick, 2006: 236)." In short, the affects of film extend beyond interpretation and subjectivity because it is a "matter of the production and circulation of affects prior to yet also providing the grounds from which distinctive intensities of feeling emerge (Carter and McCormick, 2006: 240)." Rather than querying the representational logic of a film's internal properties, a focus on affect seeks to show how film amplifies and modulates affects across broader cultural spaces and practices.

Carter and McCormack postulate four different ways in which to engage geopolitics through film. These methods are emblematic of boarder engagements by geographers on film and are thus worthy of our attention. The four methods include performative logics (discussed above), scripting, 'ways of seeing', and hermeneutics. Scripting relates to how codes and signs are articulated in a film's narrative. Frequently, stereotypes are used in film to encapsulate larger perceptual generalizations about people, social groups, social practices, and places. Stereotypes are not neutral, but rather, frame a dominant group's perspective as 'natural.' As Dixon and Zonn (2005: 92) point out, when "the meaning of an idea or object becomes commonly accepted, we have what Barthes calls a 'myth,' that is a construction that is taken for granted, or necessary." Research from a humanistic perspective has frequently sought to examine how scripting affects our perception of place. For instance, Hausladen and Starrs essay shows how the film noir genre scripts ambiguity and impermanence onto a labyrinthian geography of Los Angeles.

'Ways of seeing' is a frequently invoked phrase when it comes to the concept of landscape. According to Carter and McCormack, (2006: 232) when applied to film, the phrase is a "matter of the production and reproduction of particular logics of perception." In seeking to expose how film develops and reflects particular ways of seeing, Godfrey examines the role of the filmmaker in shaping the representational character of nature–culture relations in the Amazon. He found that films made by First World directors framed nature as either pristine wilderness or as chaotic *terrae incognitae* for colonization. In contrast, a Brazilian filmmaker presents a variegated representation of a country mediating between the needs of development, modernization, and traditional Brazilian values.

Hermeneutics focuses on the interpretation of cultural texts and their context with regard to other texts, practices, and conditions. This has been the dominant mode of inquiry in film geography, because of its flexibility with regard to how films can be interpreted. However, it is susceptible to criticism through the issue of relativism. Relativism relates to the fact that any textual interpretation is relative to the interpreter's particular viewpoint. Thus, any film could have countless 'readings' performed upon it. The use of hermeneutics in geography was an outgrowth to the crisis of representation, because it aided in questioning the mimetic belief of film as a representations of reality, and also, in that it granted equality to all forms of cultural texts, whether cityscapes, landscapes, books, music, or film. It is through this approach that film is conceived as a discursive formation, in that its meaning and existence is interwoven through representation, social practices, social formations, and economic activities.

A cultural economy approach centers on film as part of an industrial complex responsible for creating culturalized goods laden with symbolic meaning. According to Scott, (2005: 3) "the cultural economy can be broadly defined as a group of sectors…that produce goods and services whose subjective meaning, or, more precisely sign value to the consumer is high in comparison with their utilitarian purpose." Like the other approaches, this one is not cohesive, but rather includes traditional structuralist economic geography and political economy perspectives. An excellent example of a political economy approach to film is found in *Global Hollywood 2*. Miller *et al.* (2005: 5) posit that "unlike textual reductionists, we do not assume that it is adequate to interpret a film's internal qualities or the supposed 'positioning' of mythic spectators." They suggest that structures of meaning (via cultural studies) and economic structures do not have to be treated as mutually exclusive. Similarly, Macdonald argues that:

> …all cinema is produced and consumed in a highly interconnected global industry of capital, personnel, and ideas, and that film discourse, everything written or said about film, is an integral part of the production consumption matrix. As such, geographic analysis must derive from the underlying economic, cultural, and aesthetic process of this communication and entertainment medium. (Macdonald, 1994: 27)

The cultural economy approach brings to light film's triple ontology: film as an affective agent in the circulation and amplification of ideas, behaviors, emotions, and logic; film as a representational and discursive formation involved with meaning production and re production; and film as an economic commodity involved with the global production, distribution, and con sumption of an industrial complex.

Conclusion

Geographic research on film is a new and growing en terprise. Research can be broadly characterized by cul tural studies and cultural economy approaches. Work from humanistic geography has focused on how our perception of place, our 'way of seeing' and the symbolic, scripted meanings appropriated onto place are influenced by film. Work from the political economy perspective has sought to uncover the naturalized ideologies and spa tialized myths presented in film. The crisis of represen tation attacked the centrality of mimesis within geography. According to Duncan and Ley (1993: 2) mi mesis is "the belief that we should strive to produce as accurate a reflection of the world as possible." Following this debate, hermeneutical engagements reconstituted the textual metaphor allowing film to become an essential component of geographic research. According to Aitken and Dixon, anti essentialist epistemologies, or those that do not believe in mimesis, have taken this one step further by "…emphasizes the fact that objects do not exist in a vacuum; instead, they are given meaning through the actions and thoughts of people" (Aitken and Dixon, 2006).

Research on the cultural economy of film con textualizes it within an economic, historical, and indus trial framework. Current research seeks to address the industrial complex of production, distribution, and con sumption, and its role in globalization. According to Miller *et al*., (2000: 15) "Los Angeles culture and New York commerce dominate screen entertainment around the globe." People around the world are flooded with Hollywood's cultural products and their web of signs, images, and commodities that reinforce the capitalist–patriarchal–Western hegemonic dominance. Recent en gagements by geographers have sought to move away from Hollywood by examining Third Cinema, French cinema, Vancouver, Hong Kong, and other national and regional cinema complexes. Clearly, all approaches are valid lines of inquiry and cross pollination of ideas from different approaches will surely prove to advance future research in film geography.

See also: Popular Culture; Visuality.

Further Reading

Aitken, S. (1991). A transactional geography of the image event: The films of Scottish director, Bill Forsyth. *Transactions: Institute of British Geographers* 16, 105 118.

Aitken, S. and Dixon, D. (2006). Imagining geographies of film. *Erdkunde* 60(4), 326 336.

Aitken, S. and Zonn, L. (eds.) (1994). *Place, Power, Situation and Spectacle: A Geography of Film*. Totowa, NJ: Rowan and Littlefield Publishers.

Burgess, J. and Gold, J. (1985). Place, the media and popular culture. In Burgess, J. & Gold, J. (eds.) *Geography the Media and Popular Culture*, pp 1 32. New York: St. Martin's Press, Inc.

Buttimer, A. (1990). Geography, humanism, and global concern. *Annals of the Association of American Geographers* 80(1), 1 33.

Carter, S. and McCormack, D. (2006). Film, geopolitics and the affective logics of intervention. *Political Geography* 25, 228 245.

Clarke, D. (ed.) (1997). *The Cinematic City*. New York: Routledge.

Cresswell, T. and Dixon, D. (eds.) (2002). *Engaging Film: Geographies of Mobility and Identity*. Lanham, MD: Rowman & Littlefield Publishers.

Denzin, N. (1991). *Images of Postmodern Society*. London: Sage.

Dixon, D. and Grimes, J. (2004). Capitalism, masculinity and whiteness in the dialectical landscape: The case of Tarzan and the Tycoon. *GeoJournal* 59, 265 275.

Dixon, D. and Zonn, L. (2005). Confronting the geopolitical aesthetic: Fredric Jameson, the perfumed nightmare and the perilous place of Third Cinema. *Geopolitics* 10, 290 315.

Duncan, J. and Ley, D. (1993). Introduction: Representing the place of culture. In Duncan, J. & Ley, D. (eds.) *Place/Culture/Representation*, pp 1 21. New York: Routledge.

Escher, A. (2006). The geography of cinema A cinematic world. *Erdkunde* 60(4), 307 314.

Godfrey, B. (1993). Regional depiction in contemporary film. *Geographical Review* 83(4), 428 440.

Harvey, D. (1989). *The Condition of Postmodernity: An Inquiry into the Origins of Cultural Change*. Oxford: Basil Blackwell.

Jancovich, M., Faire, L. and Stubbings, S. (eds.) (2003). *The Place of the Audience: Cultural Geographies of Film Consumption*. London: British Film Institute.

Kennedy, C. (1994). The myth of heroism: Man and desert in Lawrence of Arabia. In Aitken, S. & Zonn, L. (eds.) *Place, Power, Situation and Spectacle: A Geography of Film*, pp 137 160. Totowa, NJ: Rowan & Littlefield Publishers.

Kennedy, C. and Lukinbeal, C. (1997). Towards a holistic approach to geographic research on film. *Progress in Human Geography* 21, 33 50.

Lukinbeal, C. (2004). The map that precedes the territory: An introduction to essays in cinematic geography. *GeoJournal* 59(4), 247 251.

Lukinbeal, C. and Arreola, D. (2005). Exploring cinematic landscape. *Journal of Cultural Geography* 23(1), 1 2.

Lukinbeal, C. and Zimmermann, S. (2006). Film geography: A new subfield. *Erdkunde* 60(4), 315 326.

Lukinbeal, C. and Zimmermann, S. (eds.) (2008). *The Geography of Cinema A Cinematic World*. Stuttgart: Franz Steiner Verlag.

Lukinbeal, C. and Zonn, L. (eds.) (2004). Cinematic geographies. *GeoJournal* 59, 247 333.

Macdonald, G. (1994). Third cinema and the third world. In Aitken, S. & Zonn, L. (eds.) *Place, Power, Situation and Spectacle: A Geography of Film*, pp 27 46. Totowa, NJ: Rowan & Littlefield Publishers.

Meinig, D. (ed.) (1979). Symbolic landscapes. In *The Interpretation of Ordinary Landscapes*, pp 164 192. Oxford: Oxford University Press.

Meinig, D. (ed.) (1979). *The Interpretation of Ordinary Landscapes*. Oxford: Oxford University Press.

Miller, T. (2000). Hollywood and the world. In Hill, J. & Gibson, C. (eds.) *American Cinema and Hollywood: Critical Approaches*, pp 145 155. Oxford: Oxford University Press.

Miller, T., Covil, N., McMurria, J., Maxwell, R. and Wang, T. (2005). *Global Hollywood 2*. London: BFI Publishing.

Natter, W. and Jones, J. (1993). Pets or meat: Class, ideology, and space in *Roger and Me*. *Antipode* 25, 140 158.

Shiel, M. and Fitzmaurice, T. (eds.) (2001). *Cinema and the City: Film and Urban Societies in a Global Context*. Oxford: Blackwell Publishers.

Storper, M. and Christopherson, S. (1987). Flexible specialization and regional industrial agglomeration: The case of the US motion picture

industry. *Annals of the Association of American Geographers* 77(1), 104 117.
Scott, A. (2005). *On Hollywood: The Place, The Industry*. Princeton, NJ: Princeton University Press.
Watson, W. (1969). The role of illusion in North American geography: A note on the geography of North American settlement. *The Canadian Geographer* 14(1), 10 27.
Wright, J. (1947). Terrae incognitae: The place of imagination in geography. *Annals of the Association of American Geographers* 37, 1 15.
Zonn, L. (1984). Landscape depiction and perception: A transactional approach. *Landscape Journal* 3(2), 144 150.
Zonn, L. (ed.) (1990). *Place Images in Media*. Savage, MD: Rowan & Littlefield Publishers.
Zonn, L. (ed.) (2002 2003). Teaching cinematic geographies. *Journal of Geography* 101, 228 270.

Relevant Websites

http://www.aetherjournal.org
Aether: The Journal of Media Geography. Started in 2007, *Aether* is a new online journal solely dedicated to media geography.
http://www.kcl.ac.uk
BA in Geography with Film Studies, Undergraduate Programmes, Department of Geography, School of Social Science and Public Policyat King's College London.
http://www.public.asu.edu/~clukinbe/cinema.html
Chris Lukinbeal's Cinematic and Mediated Geographies webpage. A resource for those pursuing research in the subfield.
http://www.eidc.com
Entertainment Industry Development Corporation. This is the non profit corporation that does film permitting in Los Angeles. Their website provides production data for the greater Los Angeles area.
http://www.pbs.org
Frontline's "The Monster that Ate Hollywood": Excellent overview of the economic logic of the "New" Hollywood; Frontline's "The Merchants of Cool" Great charts which break down the media conglomerates and their holdings; Frontline's "The Persuaders" While focused on advertisement, it highlights key concepts of cultural economy and discusses integrated production advertisement in films and TV.
http://memory.loc.gov
The American Library of Congress' Online Motion Picture Collection Excellent resource of early actuality films from around the U.S. and world; The Life of a City: Early Films of New York, 1898 1906 (First collection: Online Motion Picture Collection); Inventing Entertainment: The Motion Picture and Sound Recordings of the Edison Companies (Second Collection: Online Motion Picture Collection).
http://www.afci.org
The Association of Film Commissions International. If you are wondering what filming occurs in your area, the best place to start is with your local film commission.
http://www.imdb.com
The Documentary: Backstory: Cleopatra: The Film That Changed Hollywood. The film charts Hollywood's transition from vertically integrated to a disintegrated complex through the epic film Cleopatra.
http://geography.sdsu.edu
The Geography of Film Theater, Department of Geography at San Diego State University. Has not been updated since 1998, but still provides a resource for those interested in research in film geography.
http://www.IMDb.com
The Internet Movie Database. This is an excellent resource that references most film and television productions. Under their advanced search options, one can use "location" as a search criterion. Locational attribution is very well done and full of cross references.
http://www.library.yale.edu
Research Guide in Film Studies, The Film Studies Site at Yale: Yale University Library. Excellent place for those starting research on film.

Finance, Historical Geographies of

M. Purvis, University of Leeds, Leeds, UK

Introduction

In common with the rest of the discipline, historical geography has until recently paid little overt attention to the development of financial systems and institutions. Yet finance has been vital in both facilitating and encouraging the societal and environmental transformations that are historical geography's established concerns. Attention to financial history, for its part, reveals the value of a geographical perspective. The evolution of monetary systems cannot be divorced from the challenge of time–space coordination and the aspiration to extend economic and social connections over ever greater distances. This apparently universal challenge has, however, inspired complex geographical variations in monetary forms, commercial institutions, and the conventions of financial dealing.

Using Money

Money's Core Functions

As a means of displaying wealth early money often served ends that were as much cultural and religious as they were economic. Most accounts, however, focus on its utility as a means of payment and a unit of account. Money both constitutes a common scale against which all things can be valued, and embodies that value. It thus removes the constraints of direct barter, facilitating more complicated exchanges over greater distances. Money also allows value to be stored, so that payment received for goods sold, including an individual's own labor, need not be spent immediately. In facilitating trade money has, in turn, encouraged a division of labor and geographical specialization in production that have been key drivers of economic growth.

Exploring the Consequences of Monetization

Profound social and cultural changes have also been linked to increasing money use. Theorists including Marx and Simmel stress money's role in the rationalization of modern social life. For Marx, monetary transactions render exchange impersonal, eroding social bonds and deepening alienation. Moreover, monetary expressions of value help to conceal the origins of goods, including the exploitation of nature and labor involved in their production. In turn the pursuit of monetary wealth and the access to goods which it promises become an incentive to the continued sale of labor, perpetuating its exploitation. Money thus determines a society's moral attitudes, social relationships, and the status of individuals. By contrast, Simmel sees the loosening of social bonds as potentially liberating. However, he too warns that the apparent freedom of choice conferred on the consumer by money can be disorientating and that the pursuit of money may threaten moral order.

These ideas imply that money's impact will be identical wherever it penetrates. Its increased use is thus seen as advancing spatial integration, reflecting both increased trade and common socioeconomic values. But in its lack of historical and geographical specificity such theory is at odds with monetary development in practice. This has been based not on the establishment of a single, universal currency, but on a multiplicity of money specific to particular times and places. What constitutes money, the use made of it and by whom, the meanings it conveys, and hence its societal impacts may all vary between specific contexts.

Money, Territory, and Trust

This understanding of money as place bound, as well as space transcending also accords with recent reassessments of its social foundations. Money may change the character of social relations, but to discharge its economic functions it must circulate within a population. This requires that individuals feel a shared confidence in the authenticity of a specific currency, even when its value is ostensibly confirmed by the worth of the commodity from which it is constituted. Originally such confidence rested upon personal knowledge and trust within a social network. Most such networks were spatially constrained, setting limits to the extent of exchange. Over time successful dealing itself extended networks of trust, allowing currencies to be used more extensively (**Figure 1**). Wider currency use came ultimately to rest on trust in the authority of state and banking institutions to maintain its value as a means of payment. In time institutional authority itself could become a sufficient guarantee of a currency's worth, facilitating the issue of fiduciary money, that is, forms, including paper money, lacking inherent commodity value. Yet the remit of even the most powerful institutional guarantors remained spatially circumscribed, often by state boundaries.

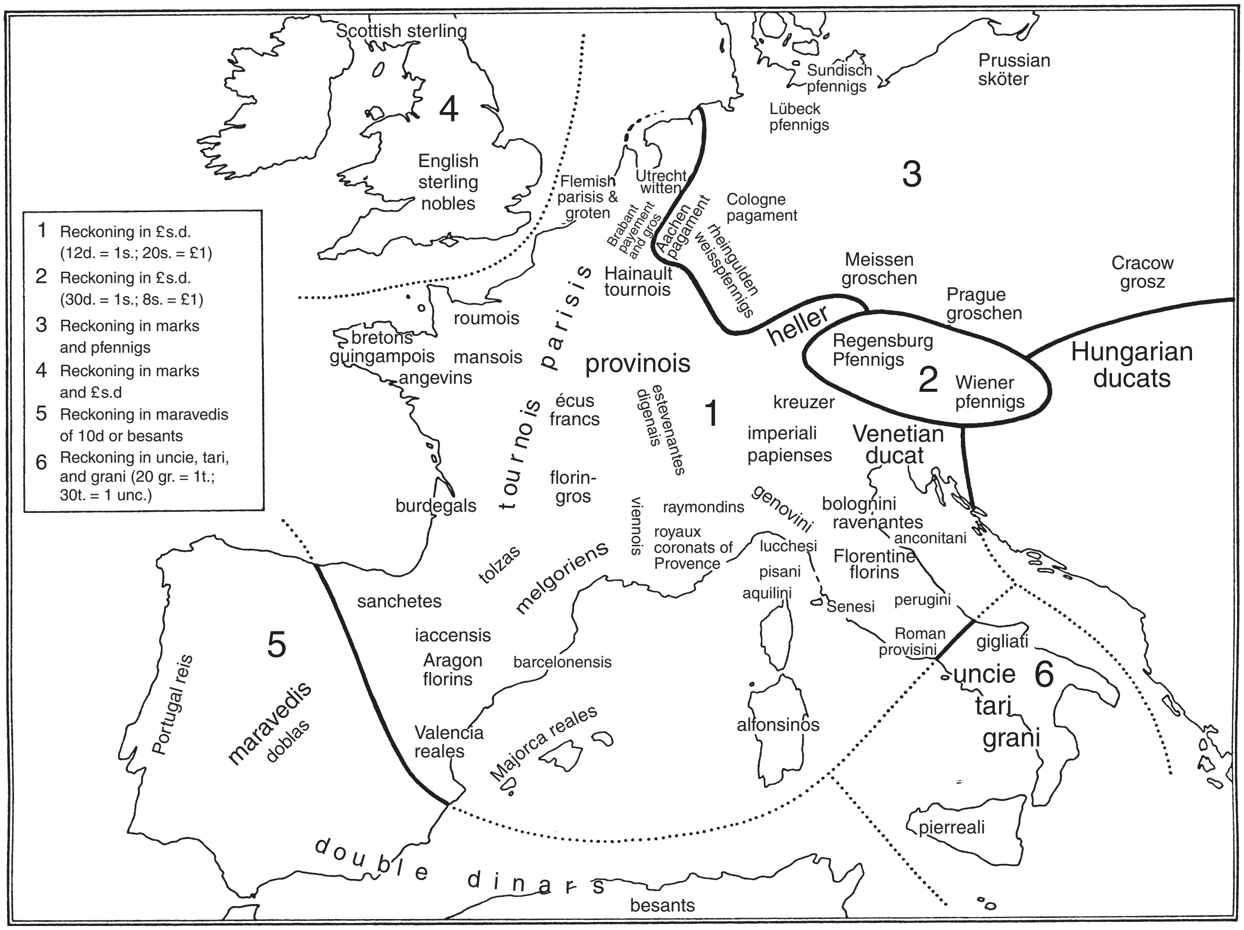

Figure 1 Principal currencies of late medieval Europe. Source: Spufford, P. (1988). *Money and / its Use in Medieval Europe*. Cambridge: Cambridge University Press, Map 32; p 294.

The Evolution of Financial Systems

Currencies and Commodities

Many objects and commodities – including shells, cattle, and salt – have historically served as locally specific primitive moneys. Metallic money often prevailed, however, as trading systems became more complex. Use of irregular lumps of precious metals gave way to standardized rods and ingots, and ultimately to coins. Most early mints were small and served only a localized area. However, consolidation of territorial authority by individual rulers both enabled and required establish ment of a standard currency valid throughout the realm. Currencies' fortunes were thus linked to those of their political sponsors, while currency design became a ve hicle for expressing a regime's power. Expansive regimes enlarged the area across which their currency circulated, often displacing weaker alternatives. State activity itself encouraged the growing use of money, through the dis tribution of pay to armies and officials and by demanding tax payments in coin.

Yet economic growth exposed coin's limitations. Its production was vulnerable to fluctuations in bullion supply. Moreover, state financial management and long distance trade were difficult and dangerous while pay ment required large scale transportation of coins or ingots. The intrinsic worth of the gold or silver contained in high value coins allowed some to be used in inter national payments, but dealers and financiers faced a confusing mix of coins of uncertain authenticity. New financial instruments and institutions were therefore necessary to increase money supply, and to enable funds to be transferred efficiently across distance and between currency areas.

Banking

Increased coinage use highlighted the role of money changers, who served traders by exchanging foreign coins for local currency, and the state by distributing new issues to replace worn and mutilated coins. Gradually, some changers began to act as bankers, accepting cash deposits and granting loans. These new deposit banks increased the availability of commercial credit. They also extended the money supply, making loans in excess of their coin reserves, in the expectation that they would not face demands for the simultaneous withdrawal of all cash deposits. Such business required increasingly sophisticated financial management, including written records of clients' accounts. From this developed facilities for customers to transfer debits and credits directly between accounts. Banks also opened accounts with each other, enabling transfers between institutions and across distance. The new services offered convenience and enabled increased business to be conducted without requiring comparable growth in the coin supply. Transferable deposits – bank money – thus further extended money supplies.

Activities, particularly trade, requiring long distance money transfers and currency exchange prompted parallel developments in merchant banking. This allowed exchange contracts, particularly bills of exchange, as written records of entitlement to payment to be substituted for transportation of coin or bullion. A merchant incurring a trading debt could purchase a bill of exchange from a banker in his local vicinity, to be dispatched to a distant creditor in lieu of coin. The recipient collected his money by selling the bill to an equivalent local banker. Exchange rates were set to yield a profit for the bankers who offset flows of debt in opposite directions against each other and settled any remaining imbalances. Such bills both eased fund transfer, and increased money supply because of the delay between their sale and redemption. Initially this reflected the time taken in carrying bills between urban centers. Payment times became standardized, however, and bills were used to extend credit for purposes other than long distance trade.

Paper Money, National Currencies, and the State

Bills and other promissory notes were one basis for the development of paper money. Rather than being redeemed by the recipient, right to payment could be reassigned, so that bills became a transferable means of settling debts. Deposit banks were a further source of paper money. Notes were initially deemed to have value because they were exchangeable on demand for a stated amount in coin. The promised payment, however, came to be indefinitely postponed, further distancing money supply from bullion reserves. This was not a wholly positive development. Problems of exchange resurfaced where many different local banks issued notes. Without central control excessive note issue could ruin individual banks, undermine wider financial confidence and fuel inflation. State regulation frequently followed, with the right to issue notes being reserved for the state central bank. The design of such notes often also assumed an increasing political importance; their portrayal of iconic landscapes and national heroes being intended to foster popular identification with the state through use of an explicitly national currency.

The remit of the modern state has extended to control over interest and exchange rates, and the transfers of funds between territories. Today, however, state authority is subject to growing challenge. Major financial institutions constitute an international network not easily regulated by individual states, while information and communications technologies allow funds to be moved with increasing speed and ease.

Geographical Contrasts: Europe and China

No simple outline can capture the complexity of monetary evolution in practice. The sequence above broadly reflects European experience, beginning with the emergence of coins during Classical Antiquity. However, subsequent developments were neither consistently progressive nor geographically uniform. In Western Europe coinage use temporarily diminished from the fifth century AD as trade and urban activity declined following the retreat of Roman imperial authority. After fitful beginnings from the seventh century, monetary systems achieved renewed stability by the tenth century. Only at this stage did coin production become firmly established in Northern Europe. The foundations of modern financial services were laid from the twelfth century, as sustained commercial growth prompted innovations in banking. Paper money entered increasing use from the sixteenth century, with many state banks asserting a monopoly over its issue during the nineteenth century. The era of state financial control which this ushered in reached its height in the decades after World War II.

Although contemporary international financial systems owe much to Europe's experience, other global regions displayed their own distinctive path of monetary evolution. In China, for example, lacking ready sources of gold or silver, the imperial state issued paper money in substantial quantities as early as the twelfth century. Use of coin in trade and tax payments consequently diminished through the thirteenth and fourteenth centuries. Confidence in paper money declined, however, as its supply grew excessively during the sixteenth and seventeenth centuries. Uncoined silver and foreign silver coins thus came to be used in substantial transactions. Petty commerce continued to employ low value bronze coins, the demand for which grew during the eighteenth century as the market economy penetrated rural China.

Hybridity in Monetary Systems: Rural–Urban Contrasts

Most monetary systems also exhibited some degree of hybridity, as different forms of money were used for particular purposes, or in specific contexts. As the Chi nese example indicates, different types of money were frequently used in wholesale and retail transactions. Evidence from early modern Europe also shows that the extension of a monetary economy did not readily elim inate barter. Rural money use, in particular, was often seasonal. Coins entered the countryside in payment for harvest, providing the means to pay taxes and rents, and to purchase goods from urban dealers. This spending meant, however, that coins left rural communities rapidly and at other times exchange frequently reverted to barter. In some remote areas this seasonality survived into the twentieth century. Barter was not unknown in urban contexts, but here coinage use was year round. Low value coins and petty credit serviced the bulk of trade, which was small scale and localized. The distinctive needs of major wholesale traders, however, prompted innovation in banking, the uneven diffusion of which closely re flected geographies of long distance trade.

European Banking and Trade: Italian Innovations and Influence

As the initial focus of the enormous trading expansion experienced in thirteenth century Europe, Northern Italy was also at the forefront of developments in deposit and merchant banking. From origins in cities including Genoa, deposit banking had spread to trading centers across Italy, Southern France, and Spain by the late thirteenth century. Diffusion – often through Italian initiative – to commercially less sophisticated regions north of the Alps chiefly followed during the fourteenth and fifteenth centuries. But here deposit banking pros pered only in the largest commercial centers, including Bruges, Antwerp, and Cologne. In Scandinavia and Central Europe the limited concentration of commercial activity in specific urban centers restricted demand for deposit banking.

The foundations of merchant banking were laid by partnerships, originally family based, established in North Italian cities, including Siena and Florence, from the mid thirteenth century. The largest of these de veloped international networks of branches and cor respondents in the principal trading centers of the Mediterranean and along the main north–south arteries of trade linking Italy with the Low Countries (**Figure 2**). Most partnerships had interests in trade and manu facture, but the growing scale of commerce encouraged some to specialize in financial services, including the insurance of vessels and goods in transit, as well as banking. Initially, the main northern outposts of this merchant banking network were the great trade fairs of the Champagne region in France. Subsequent develop ments were linked to other fairs, including Geneva and Lyon, and to growing commercial communities in Bruges and Antwerp. Even in the fifteenth century, however, northern cities lacked merchant banking facilities to match those of Italy or other Mediterranean centers, including Montpellier and Barcelona. East of the Rhine banking development lagged still further. Yet by the sixteenth century German cities, including Augsburg and Nuremberg, which lay astride the principal eastern route for north–south trade in Europe, were challenging Italian control of international finance.

Funding the State

European rulers were also amongst the clients of medi eval Italian bankers. A role in papal finances – including the transfer of income from across Europe to the papal treasury – provided leading banking houses with lucra tive business and substantial funds for investment. Loans from Italian bankers also helped to fund the political ambitions of English, French, and Spanish monarchs. As paymasters for military and diplomatic campaigning, bankers thus exercised an influence upon the political map. Yet their own fortunes were vulnerable to the dis ruptions of war and the nonrepayment of political loans, both of which contributed to Italian banking's decline.

Capital Markets and Flows

Meeting Business Needs: Small-Scale Capital Mobilization

Premodern commerce and manufacturing was pre dominantly small scale, its modest demand for capital being met from savings and the reinvestment of profit, sometimes augmented by family money or the creation of business partnerships. Bankers and moneylenders pro vided only short term credit to fund particular transac tions. Long after the onset of Industrial Revolution in eighteenth century Britain small family operations and private partnerships still constituted the vast majority of businesses. Any additional capital required was chiefly secured through personal contacts or the local client networks of intermediaries such as bankers and attorneys. Capital flows from countryside to town, funding trans port infrastructure and urban construction as well as industry, were thus predominantly intraregional. Even in the early twentieth century there were active stock ex changes in some 20 British provincial centers.

Creating Capital Markets: Funding Enterprise and the State

Only a handful of major businesses required more ex tensive investment. Amongst the earliest were companies trading with the Americas and East Indies founded in Europe during the sixteenth and seventeenth centuries.

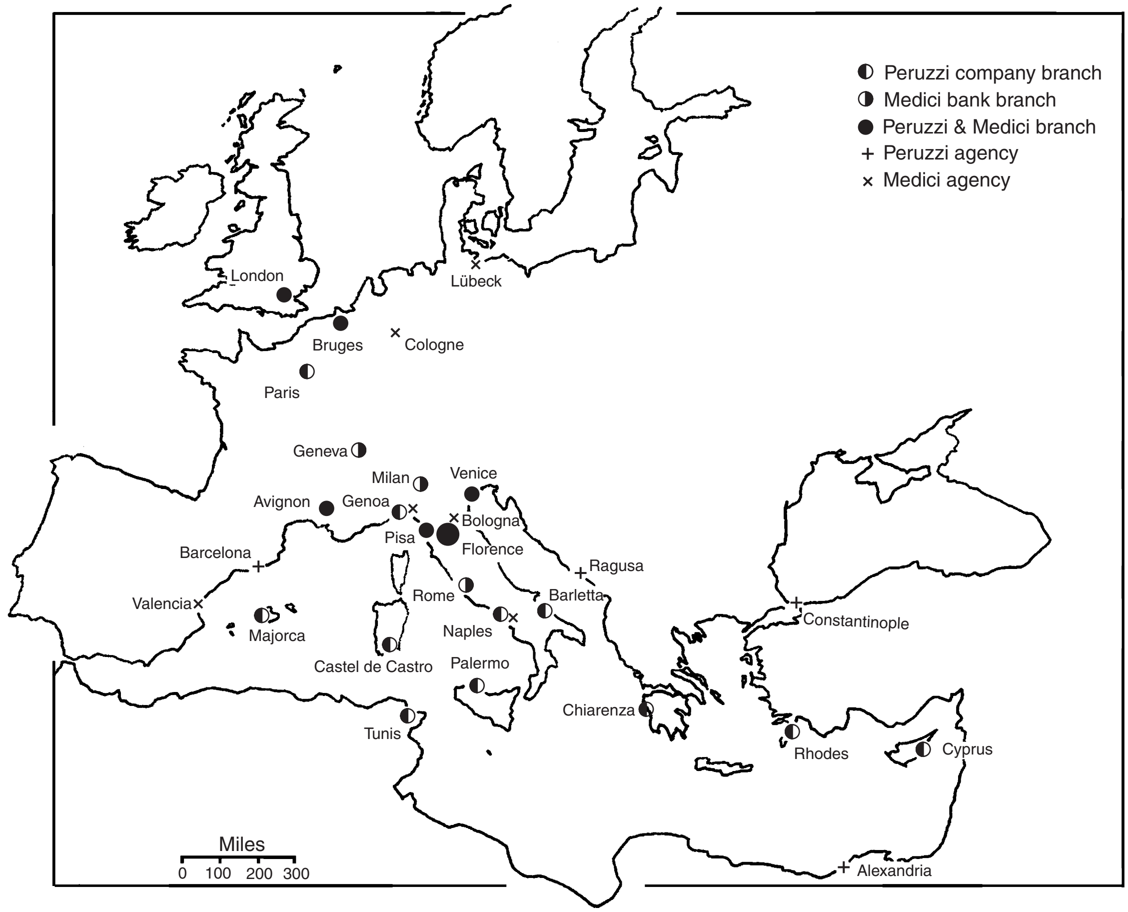

Figure 2 Italian commercial finance in medieval Europe: the operational networks of two Florentine companies. The pattern of branches and agencies reflected the interlinked functions of the Peruzzi and the Medici companies as merchants and bankers. Sources of original data: de Roover, R. (1948). *The Medici Bank: Its Organization, Management, Operations and Decline*. New York: New York University Press. Hunt, E. S. (1994). *The Medieval Super-Companies: A Study of the Peruzzi Company of Florence*. Cambridge: Cambridge University Press.

The efforts of the promoters of the Dutch East India Company of 1609 to sell shares in their business – yielding purchasers a corresponding proportion of its profits – were credited with establishing a pioneering capital market in Amsterdam. Most early capital markets, however, dealt chiefly in government securities. This reflected the translation of war loans and other advances to European states into permanent national debt from the seventeenth century. Rather than loans being repaid, creditors were rewarded with interest as a long term source of income. Shares in national debt thus became tradable commodities and markets grew as escalating military and administrative expenditure prompted rising state demand for funds.

Trade in government securities was instrumental in the rise of eighteenth century London as Europe's principal capital market and in the growth of smaller markets elsewhere in Europe and the US. During the nineteenth century, such trade became international, with London, Paris and, latterly, Berlin emerging as dealing centers for foreign government bonds. From the 1840s railway shares became a second focus of trade, boosting growth in London, Amsterdam, and Paris, and in emerging markets including Brussels, Frankfurt, Milan, New York, and a host of smaller centers. The popularity of railway shares – born of expectations of a safe return – and the proliferation of limited liability joint stock companies gradually encouraged greater dealing in other business shares during the later nineteenth century. In Britain and the US most were owned by wealthy individuals. But across continental Europe, major banks – some expressly established as vehicles to promote economic growth – became substantial industrial stakeholders. Bank investment in production also had an international dimension. For example, the Parisian arm of the Rothschild banking empire – itself a transnational entity that embodied Jewish financial prowess – acquired interests including metal mining in Spain and New Caledonia, and Russian petroleum production.

International Investment: Impacts at Home and Overseas

No other European state could, however, rival the scale of British overseas investment. By 1914 Britons held foreign assets worth £4 billion, confirming London as the primary focus of international capital flows. Overseas investments were particularly directed toward Europe and European settled territories, including substantial interests in the US and Latin America (**Figure 3**). As at home, most money was invested in government bonds and railway shares, funding the development of transport infrastructure, urban utilities, and housing. Much of the remainder flowed into agriculture, mining, financial services, and commodity trading.

The causes and consequences of the export of European capital have been widely debated. Its impact was arguably greatest in European settled territories where the extension of capitalism engendered unprecedented societal and environmental transformation. The continuing underdevelopment of the global South can also be linked to its partial and inequitable integration into world trade as a result of the penetration of European capital. Marxist scholars, in particular, have argued that European economies postponed domestic crises by exporting surplus capital, without regard for its destructive impacts on the non European world. Yet capital exports were not necessarily advantageous if they led to a shortfall in investment at home, as may have begun to happen in late Victorian Britain.

Debate about the logic of Britain's international investments has highlighted some of its likely domestic effects. Overseas agricultural interests and investment income allowed nineteenth century Britain to import increasing quantities of cheap food. This eroded domestic agricultural prosperity, changing rural landscapes, and accelerating rural–urban migration. The relative decline of British industry after mid century may also have been hastened by a lack of investment in innovation and new technologies. In contrast, overseas investment reinforced the wealth of a tiny minority, concentrated in London and Southeast England. Their dominant share of investment earnings, totaling £200 million annually by 1913, was a major stimulus to the expansion of a service based economy centered on the 'gentlemanly capitalism' of London's financial sector. This regional prosperity contrasted with the stagnation of much provincial industry, reinforcing perceptions of a north–south divide in wealth, social status, and political influence.

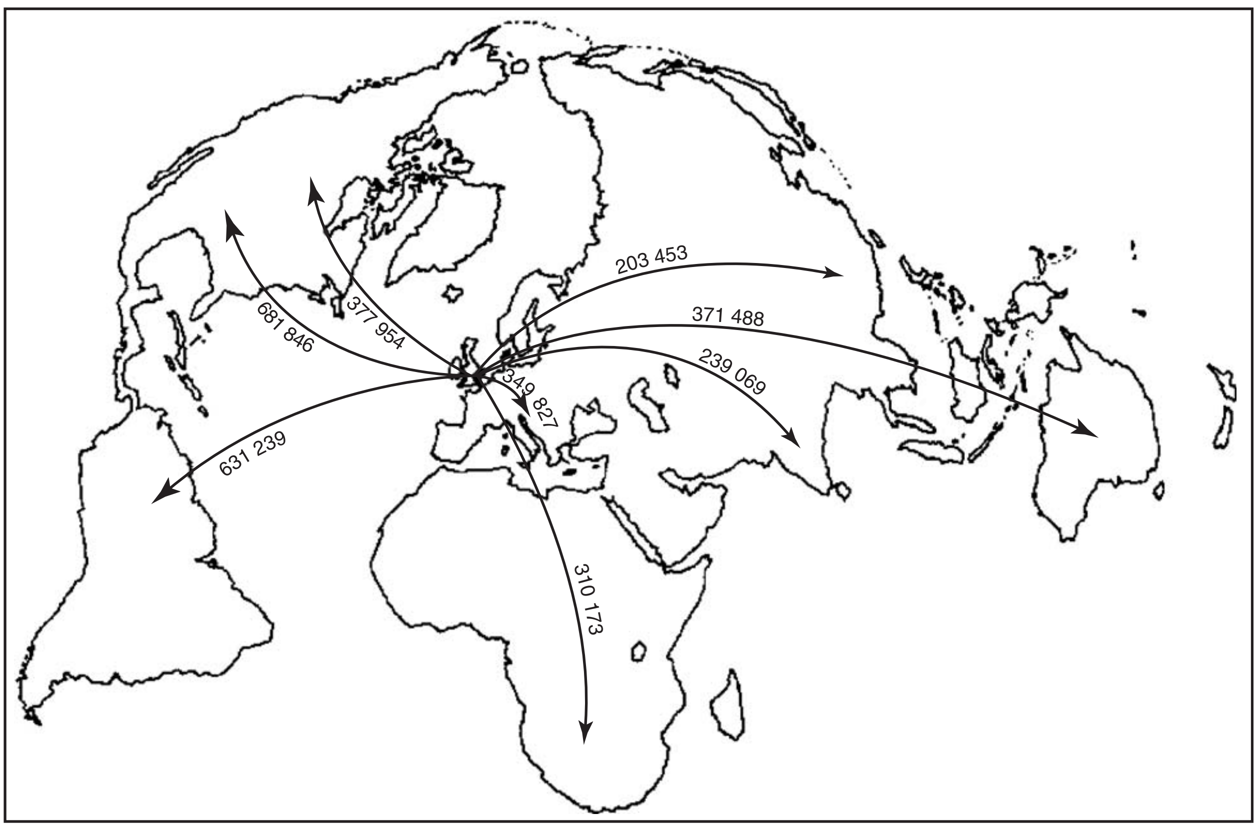

Figure 3 British overseas investments 1865 1914 total capital called up (in £000s). Source of original data: Davis, L. E. and Huttenback, R. A. (1986). *Mammon and the Pursuit of Empire: The Political Economy of British Imperialism, 1860 1912*. Cambridge: Cambridge University Press.

Financial Centers

An Elite of Cities

Primacy within international financial networks has historically been claimed by remarkably few cities. Florence and Bruges achieved initial prominence in Europe. Antwerp, however, emerged as the financial center for an early modern European economy increasingly linked to Atlantic and Indian Ocean trade. During the seventeenth century primacy passed to Amsterdam. Nineteenth century London, in turn, became the focus of a worldwide financial system. All these cities maintained their status as financial centers for some time after the commercial and industrial eclipse of their national economies. Contemporary London thus retains much of its international importance, despite the twentieth century rise of New York and Tokyo as global financial centers.

Networking within the City and Beyond

The functioning of major financial centers necessarily rests on the construction of extensive external links. Information and communications technologies – including the telegraph and telephone in the nineteenth century and computer based systems in the twentieth century – have proved vital in facilitating flows of funds and commercial intelligence. Equally important, however, are the localized links forged through social and professional interaction – often almost exclusively masculine – within a specific urban milieu. Geographical proximity has helped to cement powerful alliances, for example, between the financial and political elites frequently found together in capital cities, and to sustain the regular round of trading in financial markets.

An expression of the enduring importance of internal networking for the financial community is the clustering of activity in particular urban districts. Originating in an era when deals were struck face to face, this spatial concentration remains evident in the contemporary City of London, New York's Wall Street, and many smaller financial centers (**Figure 4**). Stress continues to be placed on access to information and trust deriving from personal interaction as a basis for doing deals. Combined with the pull of key institutions this has limited relocation, even in the face of soaring property prices and instantaneous communications.

Urban Financial Districts: The City of London

The location of London's financial district changed little during the city's international rise. Anchor points were provided by the Royal Exchange on Cornhill and the adjacent Bank of England. But these sixteenth and seventeenth century foundations were themselves sited in an area that had housed the London offices of Italian merchant bankers in the fourteenth century. The modern financial district that emerged in the seventeenth century contained a diverse population, including suppliers of legal services, news, and entertainment. These various activities were frequently intertwined, as in the taverns and coffeehouses of the city, which were places for deal making and information exchange, as well as recreation. Some such premises became associated with particular trading activities, giving birth during the eighteenth century to Lloyd's insurance business and the London Stock Exchange. This locational segregation of specific activities became more marked during the eighteenth and nineteenth centuries. Most stockbrokers' offices came to cluster around the Stock Exchange, while major bank headquarters were concentrated in Lombard Street close to the Clearing House through which their mutual debts were discharged daily.

During the nineteenth century the City of London, in common with financial districts across Europe and North America, lost most of its residential population. Premises that had performed both domestic and commercial functions were entirely converted to business use, or replaced by purpose built offices. These were not simply functional working spaces. Often designs echoed the architecture of classical temples or the Renaissance palaces of earlier Italian financial centers. Such buildings were intended to foster public confidence in the wealth and probity of the companies they housed, distinguishing them from the speculators and fraudsters who also populated the city. Collectively, they also proclaimed the importance of the financial sector to London and its role in promoting the international standing of Britain's capital city. Such thinking was particularly explicit in early twentieth century designs for the reconstruction of the Bank of England and other bank headquarters which formed part of wider efforts to give architectural expression to London's imperial status. Imperial links were equally evident in the export of the palatial architecture of major European banks to cities across the world as a tangible symbol of a powerful alien presence.

Low Finance

Payment and Dealing

The finances of ordinary people, while linked to commercial and governmental decisions about investment or taxation, spawned a further diversity of instruments and institutions. Barter, payment in kind, and localized currencies persisted in everyday dealing. Sometimes this reflected continuing shortages of coin, but payment in kind or by tokens exchangeable only at employer run truck shops was also a deliberate means of intensifying labor exploitation by lowering the real value of wages.

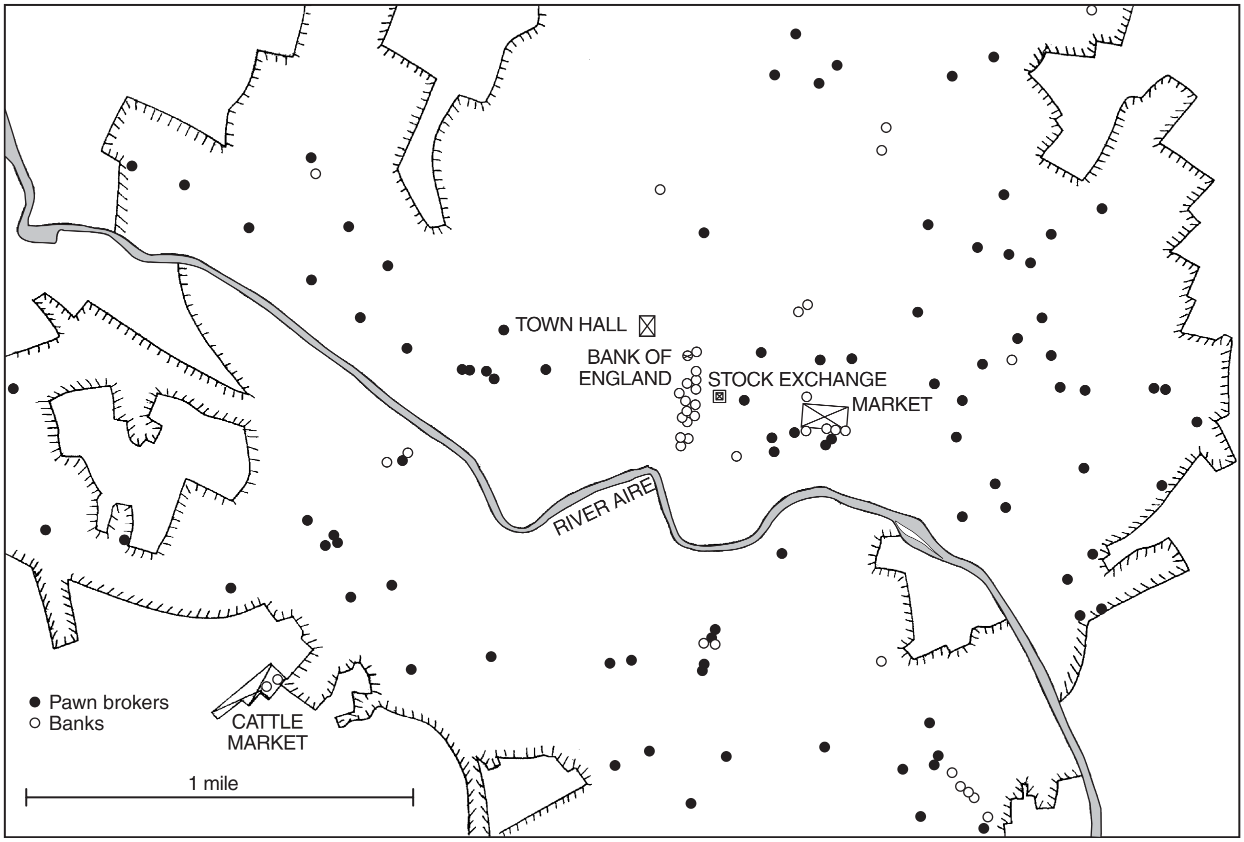

Figure 4 Two financial worlds in a British industrial city bank branches and pawnbrokers in Leeds, 1905. The principal bank branches clustered in central Leeds close to other key financial institutions; smaller clusters reflected specific foci of client demand, including the city's markets. Pawnbrokers, often also trading as jewelers or clothes dealers, were particularly common in the poorer residential districts of east and south Leeds. Source of original data: Kelly's Directories (1905). *Kelly's Directory of Leeds.* London: Kelly's Directories.

Petty Credit

Lending to the poor, often as pawnbroking where money is advanced on the security of deposited items, has a near universal history. In Europe it emerged as an organized trade alongside other banking during the thirteenth century. Recorded first in Southwest France, licensed pawnshops – granted exemption from religious prohibitions against usury – spread across France, Italy, and to the Low Countries during the thirteenth and fourteenth centuries. Italian involvement – evident in the term 'Lombard banking' for the sector – was widespread. But the moral taint associated with circumventing usury laws meant that marginalized minorities, particularly Jews, were also prominent. In the industrial era pawnbroking became ubiquitous in poorer urban communities across Europe and North America (**Figure 4**).

Help for the Poor and Self-Help by the Poor

The recognition that pawnbroking, while exploitative, provided a useful service in extending credit to the poor, prompted alternative initiatives. Charitable efforts to provide loans at reasonable interest rates included the *monte di pietà* first established by Franciscan monks in fifteenth century Italy. Other organizations such as the rural credit cooperatives promoted in nineteenth century Germany by F. W. Raiffeisen were more the product of individual initiative. Equivalent schemes in British India and elsewhere were state backed. Industrial and political revolution in Europe and North America, however, fostered popular institutions – mutual banks, building societies, and insurance providers – that substituted self help for patronage.

Financial service provision by and for the poor often reflected contrasting urban and rural needs. In the countryside credit was chiefly provided to allow peasant farmers to obtain seeds and other inputs, with repayment being expected after harvest. Consumption, however, was the main stimulus of urban demand for petty credit, driving the weekly cycle of pawning and redemption which characterized the budgetary management of many poorer households during the industrial era. Consumers' cooperation, as an alternative means of stretching meager incomes, was also strongest in industrial towns and villages. Its roots were often deepest in smaller centers, for – as with high finance – initial success often rested upon a foundation of trust between individuals who knew each other personally.

See also: Capital and Space; Capitalism; Capitalism and Division of Labor; Central Business District; Financial Centers, International; Financial Knowledge; Historical Geographies, Urban; Imperial Cities; Industry, Historical Geographies of; Networks; Networks, Urban; Street Names and Iconography; Trade, Transport and Communications, Historical Geographies of; Uneven Development; Urban Architecture.

Further Reading

Black, I. (1999). Imperial visions: Rebuilding the Bank of England, 1919 39. In Driver, F. & Gilbert, D. (eds.) *Imperial Cities: Landscape, Display and Identity*, pp 96 113. Manchester: Manchester University Press.

Davis, L. E. and Huttenback, R. A. (1986). *Mammon and the Pursuit of Empire: The Political Economy of British Imperialism, 1860 1912*. Cambridge: Cambridge University Press.

de Roover, R. (1948). *The Medici Bank: Its Organization, Management, Operations and Decline*. New York: New York University Press.

Diederiks, H. and Reeder, D. (eds.) (1996). *Cities of Finance*. Amsterdam: Royal Netherlands Academy of Arts and Sciences.

Gad, G. and Holdsworth, D. (1984). Building for city, region and nation. Office development in Toronto 1834 1984. In Russell, V. L. (ed.) *Forging a Consensus. Historical Essays on Toronto*, pp 272 319. Toronto: University of Toronto Press.

Gilbert, E. (2005). Common cents: Situating money in time and place. *Economy and Society* 34, 357 388.

Gilbert, E. and Helleiner, E. (eds.) (1999). *Nation States and Money: The Past, Present and Future of National Currencies*. London: Routledge.

Hunt, E. S. (1994). *The Medieval Super Companies: A Study of the Peruzzi Company of Florence*. Cambridge: Cambridge University Press.

Kindleberger, C. P. (2005). *A Financial History of Western Europe*. London: Routledge.

Leyshon, A. and Thrift, N. J. (1997). *Money/Space: Geographies of Monetary Transformation* (esp. 'Introduction', pp 1 38). London: Routledge.

Michie, R. C. (1997). Friend or foe? Information technology and the London Stock Exchange since 1700. *Journal of Historical Geography* 23, 304 326.

Spufford, P. (1988). *Money and Its Use in Medieval Europe*. Cambridge: Cambridge University Press.

Spufford, P. (2005). *From Antwerp to London. The Decline of Financial Centres in Europe*. Wassenaar: Netherlands Institute for Advanced Study in the Humanities and Social Sciences.

Von Glahn, R. (1996). *Fountain of Fortune: Money and Monetary Policy in China, 1000 1700*. Berkeley: University of California Press.

Zelizer, V. A. (1994). *The Social Meaning of Money: Pin Money, Paychecks, Poor Relief and Other Currencies*. New York: Basic Books.

Relevant Websites

http://www.bankinghistory.de
Relevant Academic and Professional Bodies Include the European Association for Banking and Financial History.

Finance, Offshore

S. M. Roberts, University of Kentucky, Lexington, KY, USA

Glossary

Offshore Finance Financial activity that takes place in less-regulated jurisdictions, often with low or no applicable taxes.

Offshore Financial Center (OFC) A jurisdiction that has a deliberately less-regulated and less- (or un-) taxed financial sector and offers a range of financial services (corporate and personal banking, insurance, securities, financial management, trusts, and so on). Bigger OFCs host offices of major international banks and the big four global accounting firms.

Tax Haven A jurisdiction that either exempts foreigners from paying tax on savings or income from abroad, or has minimal rates of taxation so as to attract foreign companies and individuals seeking to avoid taxes.

Introduction

Offshore finance has become a major component of the international financial system. Although there are no precise figures available, it is estimated that at any time, over half of the world's money will be traveling in offshore circuits or located in offshore centers. Indeed, depending on how they are defined, there are now between 14 and 70 offshore financial centers (OFCs) scattered around the globe (see **Figure 1**). The Bank for International Settlements (BIS) estimates that for 2006 the largest five OFCs (Bahamas, Cayman Islands, Netherlands Antilles, Hong Kong SAR, and Singapore) together accounted for 12% of all international bank liabilities – a greater share than the US. In addition to offshore banking activities of a traditional sort, the offshore financial markets include a large and diverse financial services business offering innovative financial products or instruments in the bond markets, in the mutual funds sector (including hedge funds), in the trust business, and there is also a significant and very active offshore insurance industry. Every major bank, each of the big four global accounting firms and most, if not all, transnational corporations (TNCs) have extensive offshore networks.

By hosting offshore finance – turning themselves into OFCs – a number of small states have managed to establish successful economies. Some OFCs are in places (often small islands) that have few other options for economic development and hosting offshore finance has proven an attractive option. For example, the Caribbean islands with the highest gross national products (GNPs) per capita include the OFCs of the Bahamas, the Caymans, and the Netherlands Antilles. In Europe, a set of states have historically positioned themselves to host flight capital and now offshore finance, including Switzerland, Luxembourg, and Liechtenstein. Each has profited from its place in the offshore financial world. However, as is discussed below, offshore finance does not only take place in these havens, it also takes place in the world's supposedly 'onshore' financial centers – notably London.

Defining Offshore

As is already clear, offshore finance is not a sphere of activity that is completely separate from so called onshore activities. Most analysis of offshore finance points out that offshore is deeply connected to onshore through flows of money and through institutions and other actors that work in both realms. The interesting attempts by BIS researchers to map international banking flows show that many international banking transactions between 'onshore' centers are routed through OFCs. More profoundly, though, offshore finance only exists as a relational category. Offshore is defined by its difference from onshore. Specifically, offshore finance takes advantage of less (or differently) taxed and regulated spaces in the world economy. OFCs are, as Ronen Palan describes, deliberately structured juridical spaces. Offshore finance is, in this way, distinctly geographical. Through innovative uses of offshore markets and centers, capital exploits the uneven global topographies of taxation and regulation to minimize its liabilities to the state (taxes) and accumulate profits. In general, offshore finance may be thought of as the 'constitutive outside' of finance. That is, while it may be regarded as very different from or even as working in opposition to onshore finance, even the briefest of analysis reveals that offshore is very much a part of onshore.

Offshore finance has recently come under a great deal of pressure and the pressure has largely been aimed at the key juridical spaces through which it works: the OFCs. Long suspected of providing shelter to money launderers and super rich tax dodgers, the OFCs are now suspected of providing terrorists with a financial network, and with assisting corporate fraudsters whose companies pay tiny amounts of taxes in their major countries of operation while hoarding corporate and personal fortunes offshore.

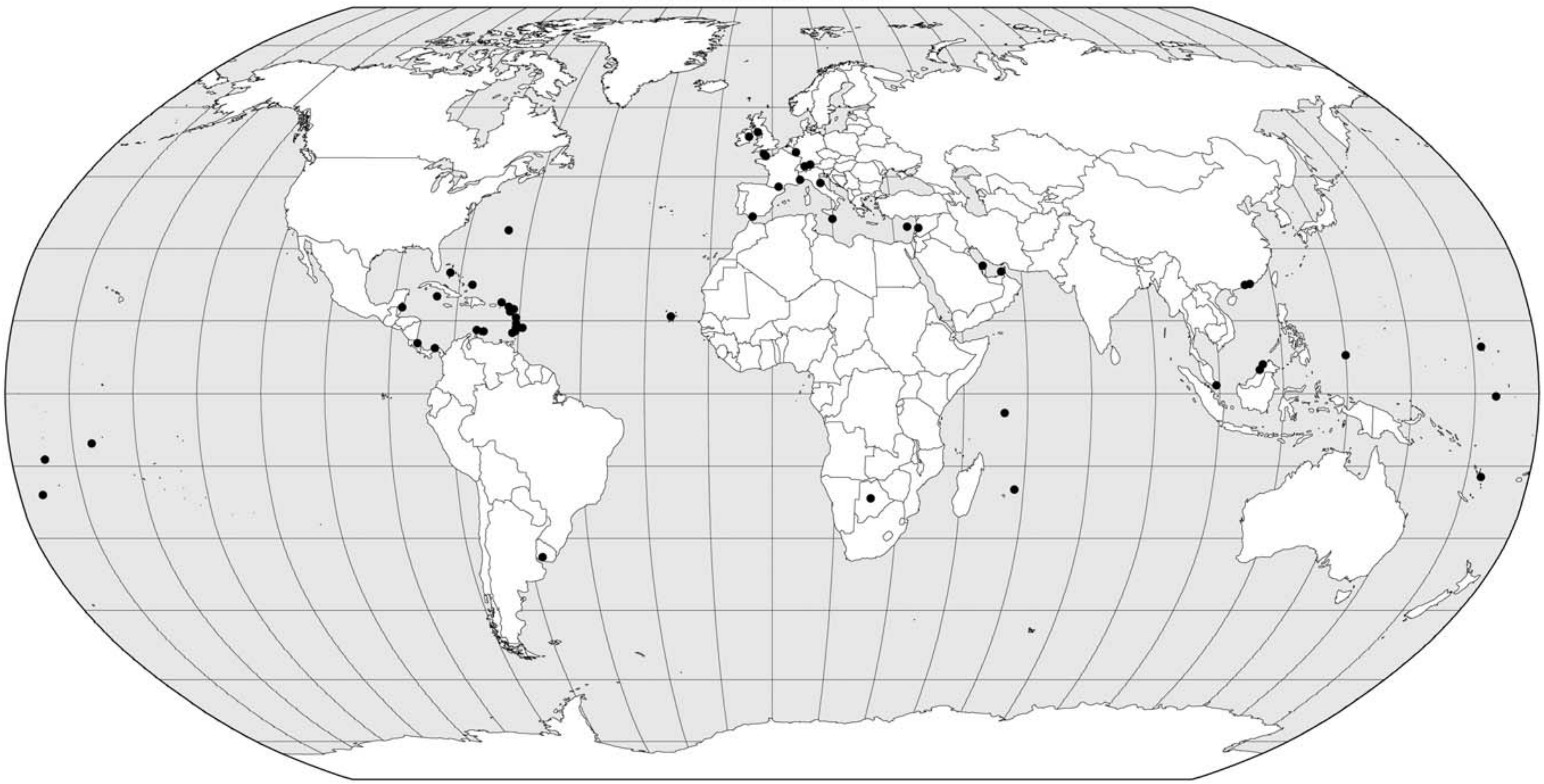

Figure 1 Offshore F nanc a Centers. Source data: MF 2006.

The Rise of Offshore Finance

Offshore financial centers developed, in many cases, out of tax havens. Tax havens themselves arose as places in which wealthy families could shelter their money in trusts or investment companies and avoid paying different kinds of taxes in their home countries. The revenue authorities in Great Britain and in the United States investigated such practices in the 1920s and 1930s. In subsequent decades, as the number of TNCs rose, corporations rather than individuals or families became the main users of tax havens. Indeed, the rise of TNCs during the 1950s and 1960s and more particularly the growth in the levels of foreign direct investment (FDI) activity by TNCs during these decades may be partly explained by the incentives provided for FDI under prevailing tax regimes. Essentially, TNCs' profits made abroad were not taxed in the companies' home jurisdictions until they were repatriated. Thus TNCs had every reason to reinvest profits made overseas in further rounds of FDI. To defer or avoid taxes on profits made overseas, TNCs established various subsidiary corporate entities in less taxed (or differently taxed) jurisdictions. By using a suite of conduit and holding companies, along with old fashioned transfer pricing techniques, a TNC could effectively minimize its tax liabilities worldwide. If that was the 'demand side', the 'supply side' was found in the growing ranks of tax havens where enterprising (often expatriate) lawyers, bankers, and accountants worked with legislators to structure the legal and regulatory systems of their countries to facilitate and attract TNCs wishing to set up various types of corporate entities. Many aspiring OFCs (or 'international financial centers' as these new and former tax havens often prefer to be called) found they could reap certain rewards, in the shape of fees and revenue, from the business of being an OFC. For otherwise quite desperately poor places, often without much scope for enacting economic development strategies, this potential revenue stream was very appealing.

A further and decisive boost to the fledgling offshore finance sector came during the late 1960s and 1970s. This was the period during which the international financial system underwent a profound transition. From the end of World War II until that time, exchange rates had been more or less fixed under the so called Bretton Woods system. The US dollar acted as the *de facto* world currency and it was tied to gold (at $35 an ounce). The tremendous pressures on this relatively inflexible (but stable) system built up in the post war decades as economic growth occurred unevenly. The US found itself in a deficit position (in part due to the costs of the Vietnam War) and in August 1971 President Richard Nixon suspended the convertibility of the dollar to gold, put a surcharge on imports, and demanded that the dollar be devalued. Eventually, the system of fixed or pegged exchange rates broke down and the world's currencies began to more or less float against one another. The rigidity of the post war Bretton Woods currency regime was broken. Floating exchange rates were a source of uncertainty and risk for TNCs to some degree, but the movement of currencies against one another also provided scope for further profit making through speculation. Both aspects further fuelled the rise of offshore finance. To deal with currency risk, TNCs could park sums denominated in different currencies in OFCs. To take advantage of the potential for arbitraging in the currency or foreign exchange markets, TNCs could establish financial subsidiaries as profit centers in their own right and base them in OFCs.

As the world economy globalized further through the last three decades of the twentieth century with international finance taking an ever more prominent role, OFCs became homes to financial institutions and other TNCs seeking to take advantage of uneven topographies of regulation, as well as taxation. Offshore markets became sites of innovation as new types of entities and new types of financial products or instruments were developed to allow corporations, including banks, to take advantage of the spatial differences in regulations on financial activities such as interest rate restrictions, reserve requirements, and reporting requirements. As the leading sectors in the US and EU became more identified with services and with intellectual property (rather than with production or trade in more tangible goods), the OFCs became central to the global strategies of technology and pharmaceutical corporations, for example. A corporation holding intellectual property (patents, copyrights, etc.) can structure itself so that the patents are held by a subsidiary based in an OFC and then the onshore corporation will pay its own subsidiary licensing fees or royalties and very effectively minimize taxes due onshore. For instance, by setting up an Irish subsidiary, Google arranged its finances so that a high proportion of its earnings were through the Irish entity, meaning they were taxed at 12.5% rather than at 35% as they would have been if they had been earned in the US.

OFCs are still used by rich individuals (known as 'high net worth individuals' in the trade) and wealthy families. Private wealth stashed in OFCs by rich individuals and families is reckoned by some estimates to be worth over US$11.5 trillion. This would generate an untaxed income of about US$850 billion a year. With the rise of a super rich class most recently typified by, but not restricted to, the new cadre of cosmopolitan Russian billionaires, OFCs have become routine elements in transnational wealth management strategies. With political instability threatening economic insecurity, middle and upper classes have long sought 'bolt holes' should they need them. For example, during the build up to both World Wars, Switzerland and Liechtenstein were perceived as safe places to which those in other parts of tumultuous Europe could send their assets. Most OFCs still host flight capital of all sorts and many offer offshore services such as ship or aircraft registries that, while they are not financial, appeal to super rich individual users of offshore finance.

The Geography of Offshore Finance

Identifying places that may be categorized as OFCs is quite difficult. In part, this is because the very term 'offshore' is somewhat ambiguous and in part because, as is explained below, offshore, while it operates in a distinctly geographical way, creates a convoluted geography. This is seen in the development of the onshore offshores in the early 1980s. In 1981, US authorities allowed banks to set up 'international banking facilities' (IBFs) in New York (and eventually other states) which operated free of the key restrictions on interest rates and reserve requirements – requirements that previously could only be escaped by setting up in an OFC. These efforts and related ones in Japan, designed to lure offshore business back onshore, plus the long standing dominance of London in the Eurocurrency markets resulted in a major portion of offshore finance actually being conducted onshore. Indeed, there is much truth to the observation that London is the world's biggest offshore center. With the recent politicization of offshore finance (see below), the identification of places as OFCs has become a loaded issue. However, even acknowledging the definitional difficulties (political and otherwise) it is possible to identify places that host offshore finance.

Researchers at the Tax Justice Network (TJN) have compiled an expansive list of OFCs which contains over 70 jurisdictions they identify as tax havens (see **Table 1**). They usefully categorize the various centers into groups: major financial centers; premier havens; midrange havens; minor havens; and notional havens. Major financial centers are those that are typically regarded as 'onshore' and have a diversified financial services industry, but where certain tax and/or regulatory structures exist to permit their use as havens or OFCs as well. The premier group of centers are those which are the biggest in terms of the volume of offshore business they handle and, while they may be specialized (for example, the British Virgin Islands (BVI) in offshore companies formation) they offer a suite of offshore services and products. Every premier center also hosts offices for each of the big four global accountancy firms (KPMG; Deloitte; Ernst & Young; and PricewaterhouseCoopers). The midrange centers, in many cases, aspire to be premier centers but may lack the volume of business or the presence of major international players in the accounting or banking fields. The 'notional' centers are those whose legal and regulatory frameworks make them potential OFCs, but for various reasons they have attracted only a limited amount of offshore business. There is some concern that this class of centers is the most likely to attract illicit finance or those seeking to evade

Table 1 Typology of offshore financial centers and tax havens

Major financial centers	*Premier havens*	*Midrange havens*	*Minor havens*	*Notional havens*
Belgium	British Virgin Islands	Aruba	Andorra	Alderney
City of London	Cayman Islands	Bahrain	Anguilla	Campione d'Italia
Frankfurt	Cyprus	Barbados	Antigua and Barbuda	Ingushetia
Hong Kong	Dubai	Bermuda	Belize	Liberia
Netherlands	Guernsey	Costa Rica	Grenada	Marshall Islands
New York	Ireland	Dominica	Madeira	Melilla
South Africa	Isle of Man	Gibraltar	Monaco	Montserrat
Switzerland	Jersey	Hungary	St. Lucia	Nauru
Tel Aviv	Liechtenstein	Iceland	St. Vincent and the Grenadines	Niue
	Luxembourg	Labuan	St. Kitts and Nevis	Samoa
	Singapore	Lebanon	The Cook Islands	Sao Tome e Principe
	The Bahamas	Macau	The Maldives	Sark
		Malta	Trieste	Somalia
		Mauritius		The Marianas
		Netherlands Antilles		The Seychelles
		Panama		Tonga
		Taipei		Turkish Rep. of N. Cyprus
		Turks and Caicos Islands		Vanuatu
		Uruguay		
		US Virgin Islands		

Source: Tax Justice Network (2007). Closing the floodgates: Collecting tax to pay for development. Report commissioned by the Norwegian Ministry of Foreign Affairs. Edited by R. Murphy. TJN London (http://www.taxjustice.nl/dialogs/download.aspx?oid=8ad013e9 1df9 4f2a a400 4fbd6700f0fd) (accessed 3 August 2007).

taxation. However, in terms of their overall contribution to offshore finance, these centers are negligible.

There are identifiable clusters of OFCs. Using the shorter list of OFCs used by the International Monetary Fund (IMF), the map in **Figure 1** shows the global distribution of OFCs (see also **Table 2** above). The major clusters are found in the Caribbean, Europe, and the Pacific. In terms of volume of business, the Caribbean and European centers outrank the Pacific centers, although if Asia and the Pacific are taken together it is reasonable to expect that the centers in this region will continue to grow as Asian economies experience high annual economic growth rates. This is because the clusters of OFCs each tend to operate in connection to their nearby onshore economies. Thus the Caribbean centers are most closely linked into North American circuits of capital and in much the same way the European centers tend to operate as adjuncts to the EU even as they attract business from all over the globe. In part this clustering may be explained by the time zone factor since OFCs tend to be closely associated with onshore financial centers and economies with which they share some portion of the business day. However, it is also a result of relic geographies of empire – many of the Caribbean OFCs are Dependent Territories of Britain, for example – in which small islands found themselves with minimal legal structures ripe for adaptation to the needs of offshore finance and with little democratic or political pressure to do otherwise. Within each cluster there has tended to emerge a hierarchy of centers. For example, in the Caribbean, the Cayman Islands and the Bahamas are much more significant OFCs than St. Lucia or Dominica, say. Each center has developed specializations within the broad category of offshore finance. Bermuda, for example, is well known as a global center for the 'captive' insurance business and as such competes with regional rival The Caymans but also with Guernsey, the Isle of Man, Luxembourg, and Vermont. The BVI, on the other hand, have specialized in offshore company formation and are home to over 300 000 so called international business companies. They compete not only with other Caribbean centers but also with Ireland, Gibraltar, Luxembourg, and Delaware in this niche. Despite the fact that individuals and corporations establishing offshore entities typically establish several in different locations so as not to 'put all their eggs in one basket', centers within a cluster also tend to compete for business. They keep a careful watch on the other centers within the region in order to counteract (or preempt) any innovations or adjustments others might be making with regard to the products or services they offer and the fees they charge. This intense competition has led some commentators to regard this as another example of a 'race to the bottom' since it is most commonly expressed as competitive deregulation.

Table 2 shows all the OFCs under IMF scrutiny. The IMF is one of a number of institutions seeking to exert some control over offshore finance. Since the 1990s, offshore finance has come to be seen as a problem by many, and there was an intense effort to clamp down on OFCs in the 1990s. It is to this issue that we now turn.

Table 2 Offshore financial centers under IMF scrutiny

Europe	*Middle East, Africa, and Indian Ocean*	*Caribbean, Atlantic, Central, and South America*	*Asia and Pacific*
Andorra	Bahrain	Anguilla	Brunei
Cyprus	Botswana	Antigua and Barbuda	Cook Islands
Gibraltar	Dubai	Aruba	Hong Kong SAR
Guernsey	Lebanon	Belize	Macao SAR
Ireland	Mauritius	Bermuda	Marshall Islands
Isle of Man	Seychelles	Barbados	Malaysia (Labuan)
Jersey		British Virgin Islands	Nauru
Liechtenstein		Cape Verde	Niue
Luxembourg		Cayman Islands	Palau
Malta		Costa Rica	Samoa
Monaco		Dominica	Singapore
San Marino		Grenada	Vanuatu
Switzerland		Montserrat	
		Netherlands Antilles	
		Panama	
		St. Kitts and Nevis	
		St. Vincent and The Grenadines	
		The Bahamas	
		Turks and Caicos Islands	
		Uruguay	

Source: International Monetary Fund (2006). Offshore financial centers: The assessment program a progress report (http://www.imf.org/external/np/pp/eng/2006/020806.pdf) (accessed 3 August 2007).

The Crackdown of the 1990s and the Reconfiguration of Offshore Finance

Since the mid to late 1990s there have been a plethora of efforts to 'crack down' on OFCs. The origins of the crackdown lie in the huge expansion of the offshore markets through the 1980s and 1990s. As the flows of offshore capital grew, various onshores started to become concerned about offshore markets and specifically about the role of OFCs. The onshores had two major classes of concern. First, they were upset that they were losing business to OFCs, despite considerable onshore offshores in centers such as London in the UK and Delaware in the US. This competitive concern evolved into the discourse of 'unfair tax competition'. Second, onshore elements were worried about the uses of offshore markets and centers for hiding and washing dirty or illicit money. This concern was not new and continued, as it had been, to be typically referenced by the label 'money laundering'. Since 9/11 this issue has been allied with a generalized concern about the funding of terrorism with which it is now always dis cursively twinned. Both discourses – unfair tax com petition and money laundering – cast OFCs as particular 'problems' or 'suspects' to be disciplined.

In an attempt to reduce the opportunities for criminals to hide their money behind the confidentiality guarantees of the OFCs, the US authorities exerted considerable pressure on Caribbean OFCs through the 1970s and 1980s. In 1986 they succeeded in badgering the Cayman Islands, for example, into signing a Mutual Legal Assistance Treaty. However, in the 1990s the pressure was stepped up and became more multilateral in character. In 1996 the Offshore Group of Banking Supervisors (a group formed in 1980 and connected to the BIS) issued a report critical of OFCs. Then in 1989 the Financial Action Task Force (FATF) was established by the G7 with the main focus on money laundering. The Organization for Economic Co operation and Development (OECD) began their in vestigations into 'unfair tax competition' in 1996, producing the report '*Harmful tax competition: An emerging global issue*' in 1998. In 1999 the BIS established the Fi nancial Stability Forum (FSF), apparently following the wishes of the G7 finance ministers and central bank governors, and this group (inevitably) issued a report of its own. Since 2003 the IMF has had its own Offshore Fi nancial Center Program which, in coordination with the FATF, the FSF, and the International Organization of Securities Commissions (IOSCO), monitors OFCs com pliance with 'supervisory standards'.

Along with the rise in the number of multinational organizations scrutinizing the OFCs, onshore states con tinued their own efforts to discipline them. The UK Government produced the Edwards Report in 1998, fo cused on the OFCs that were British dependencies. This was followed by an officially commissioned report in 2000 carried out by KPMG that focused again on Bermuda and British overseas territories in the Caribbean. The EU Savings Tax Directive, which came into effect 1 January 2005, is a system for exchanging information between member states about savings held by EU nationals in ac counts based outside their jurisdiction of residency in order to collect taxes on interest earned. This directive affected not just EU member states but also nonmember European OFCs such as Andorra and Monaco, as well as the Caribbean OFCs that are dependencies of Britain (Anguilla, BVI, Turks and Caicos, Caymans) and the Netherlands (Aruba, Netherlands Antilles). Despite pro tests, particularly from the Caymans and from Switzerland, the directive's terms have been adopted by most OFCs in Europe and the Caribbean dependencies. The EU Savings Tax Directive is just one of a set of recent efforts by the authorities in rich countries to counter cross border tax evasion and crackdown on OFCs.

These initiatives, despite pitting the richest countries on Earth against a motley collection of small states and quasi states, did not succeed in shutting down the OFCs. In fact the results, so far at least, of all these efforts have been rather minimal. There was, for a start, considerable backlash against the 'name and shame' strategies adopted by the authorities who, in their various reports, generated lists of uncooperative jurisdictions. Many in the OFCs pointed out that the lists smacked of older geographies of empire with their racist undertones as to who was, and who was not, on the so called 'black lists'. Caribbean leaders pointed out that it seemed unfair at best that Switzerland and Luxembourg were often left off the lists while The Bahamas and The Caymans were nearly al ways on them. With support from the Commonwealth, Caribbean OFC leaders established the International Tax and Investment Organisation which became the Inter national Trade and Investment Organisation (ITIO). Its members are 17 'small and developing states', 11 of which are Caribbean OFCs. The ITIO has issued its own re ports to counter the OECD led crackdown, pointing out that all the efforts are led by essentially the same group of powerful countries, and that they target offshore business only as it occurs in non OECD member states. As Jason Sharman explores, the OFCs' objections to the crack down were bolstered by the intervention of several prominent libertarian and neoliberal think tanks, espe cially in the US, whose arguments about the benefits of and moral righteousness of tax competition and privacy resonated with many in the US administration. The views of groups such as the Cato Institute, the Center for Freedom and Prosperity, the American Enterprise Insti tute, and Canada's Fraser Institute also appealed to many in the corporate world for whom OFCs provided lucra tive opportunities. For all these reasons at least, the rich countries pulled back from any direct action and appear, for now, to have settled for a regulatory regime that permits offshore finance to continue playing a major role in international capitalism and may have actually strengthened the position of the OFCs.

Recent Developments

Meanwhile, the OFCs have become the focus of attention for campaigners of a different variety. Citizens' groups in France and Britain, among others, have increasingly mobilized the discourse of corporate social responsibility to object to the use of OFCs by corporations that do business in their countries but pay little or no tax in them. Tax dodging by the newly numerous cadre of cosmopolitan super rich also fuels populist complaints that hardworking middle and working classes are shoul dering an increasing share of the tax burden; dutifully paying their taxes and effectively subsidizing the super rich and the corporations who can avail themselves of OFCs. Organizations such as the Association for the Taxation of Financial Transactions to Aid Citizens (ATTAC), which was established in France in 1998, and the TJN, set up in Britain a few years later, have publicly denounced OFCs and the multinational accounting firms that promote them. Such groups found common cause with antipoverty and development organizations. For example, Oxfam has worked to raise awareness of the links between offshore finance and poverty in the de veloping world. In their 2000 report titled '*Releasing the hidden billions for poverty eradication*', they estimated that tax havens were responsible for at least US$50 billion in revenue losses from developing countries. This amount, the report points out, is roughly equivalent to annual development aid flows to the developing countries. In 2005 the UK charity Christian Aid, working with the TJN, issued a briefing paper making very similar points. In the US, populist politics concerning the uses of OFCs by individuals and corporations waxes and wanes. Every corporate scandal seemingly contains an offshore ele ment. Enron, for example, is reported to have had 881 offshore subsidiaries, 692 in the Cayman Islands alone. Such corporate behavior coupled with tax dodging by individuals results in an estimated loss of over $100 bil lion a year in tax revenues due to offshore. Citing such figures, US Senators Carl Levin, Norm Coleman, and Barack Obama have introduced a Stop Tax Havens Abuse bill to counter the industry that they say permits tax cheats to offload their responsibilities onto ordinary working people. With such efforts raising awareness of the offshore finance, it may be that OFCs will come under renewed attack. On the other hand, they are places that the world's biggest corporations and wealthy people are very literally invested in and whether they will allow their greed to be compromised is another issue.

See also: Capitalism; Financial Centers, International; International Organizations; Transnational Elites.

Further Reading

Baker, R. (2005). *Capitalism's achilles heel*. Hoboken, NJ: Wiley.
Brittain Catlin, W. (2005). *Offshore: The dark side of the global economy*. New York: Farrar, Straus and Giroux.
Christensen, J. P. and Hampton, M. P. (eds.) (1999). *Offshore finance centers and tax havens: The rise of global capital*. Basingstoke: Macmillan.
Komisar, L. (2007). *Corporate profits take an offshore vacation*. Global Policy Forum. http://www.globalpolicy.org/nations/launder/haven/2005/0404ashore.htm (accessed 3 August 2007).
International Monetary Fund (2006). *Offshore financial centers: The assessment program a progress report*. http://www.imf.org/external/np/pp/eng/2006/020806.pdf (accessed 3 August 2007).
McGuire, P. and Tarashev, N. (2006). Tracking international bank flows. *BIS Quarterly Review* December 27 40.
Oxfam (2000). *Tax havens: Releasing the hidden billions for poverty eradication*. Oxfam GB Policy Paper. http://publications.oxfam.org.uk/oxfam/display.asp?searchterm=Tax+havens%3A+Releasing+the+hidden+billions+for+poverty+eradication&K=20040623 2316 000 034 (accessed 3 August 2007).
Palan, R. (2003). *The offshore world: Sovereign markets, virtual places, and nomad millionaires*. Ithaca, NY: Cornell University Press.
Piccioto, S. (1999). Offshore: The state as legal fiction. In Christensen, J. P. & Hampton, M. P. (eds.) *Offshore finance centers and tax havens: The rise of global capital*, pp 18 42. Basingstoke: Macmillan.
Rawlings, G. (2007). Taxes and transnational treaties: Responsive regulation and the reassertion of offshore sovereignty. *Law and Policy* 29(1), 51 66.
Roberts, S. M. (1994). Fictitious capital, fictitious spaces: The geography of offshore financial flows. In Corbridge, S., Martin, R. & Thrift, N. (eds.) *Money, power and space*, pp 91 115. Oxford: Blackwell.
Sharman, J. C. (2006). Norms, coercion and contracting in the struggle against 'harmful' tax competition. *Australian Journal of International Affairs* 60(1), 143 169.
Tax Justice Network (2005). *The price of offshore*. Briefing paper prepared by Tax Research Limited for the Tax Justice Network UK. http://www.taxjustice.net/cms/upload/pdf/Price of Offshore.pdf. (accessed 3 August 2007).
Tax Justice Network (2007). *Closing the floodgates: Collecting tax to pay for development*. Report commissioned by the Norwegian Ministry of Foreign Affairs. Edited by R. Murphy. TJN London. http://www.taxjustice.nl/dialogs/download.aspx?oid=8ad013e9 1df9 4f2a a400 4fbd6700f0fd (accessed 3 August 2007).
The Economist (2007). Survey: Offshore Finance. http://www.economist.com (accessed on 22 February 2007).

Relevant Websites

http://www.visar.csustan.edu
Association for Accountancy and Business Affairs.
http://www.caymanfinance.gov.ky
Cayman Islands Financial Services.
http://www.imf.org
International Monetary Fund.
http://www.itio.org
International Trade and Investment Organisation.
http://www.oecd.org
Organisation for Economic Cooperation and Development.
www.taxjustice.net
Tax Justice Network.

Financial Centers, International

D. R. Meyer, Brown University, Providence, RI, USA

Glossary

Brokerage Firm Buys and sells stocks, bonds, currencies, commodities, and other financial instruments for customers.

Commercial Bank Diversified financial firm that takes deposits from individuals, firms, governments, nonprofit institutions, and so on and loans to those groups, as well as provides them financial services such as offering advice and exchanging currencies.

Derivative Financial instrument traded privately between parties or on exchanges in which the price of the instrument depends on the value of some underlying security such as stocks, bonds, currencies, or commodities. The rights or obligations based on the underlying product are traded, not the underlying product. Derivatives include futures, options, and swaps (exchanges) of payments on currencies or interest rates.

Gross Domestic Product The total market value of goods and services produced in a country within one year. It equals consumption, investment, government spending, plus the value of exports, minus the value of imports.

Hedge Fund Private investment fund that invests and trades in various assets such as stocks, bonds, and currencies on behalf of clients, including pension funds, universities, and wealthy individuals.

International Financial Center Cities that house major financial firms such as commercial and investment banks, brokerage firms, money management firms, private equity firms, and hedge funds which operate internationally.

Investment Bank Sells new stocks and bonds for other firms, provides financial services such as trading currencies and commodities and advising on mergers and acquisitions, and invests its own capital in assets.

Private Equity Investment in an asset such as a manufacturing firm, retailer, office building, and airline, and this equity does not trade on a public stock exchange.

Social Networks of Capital The relationships or linkages among firms and individuals which are paths for exchanging capital. The 'social' emphasizes that these relations occur in business and nonbusiness venues.

In international financial centers, pivotal actors (individuals and firms) make decisions about the exchange of global capital. These places house commercial and investment banks, brokerage firms, money management firms, private equity firms, hedge funds, and so on. The actors in a small number of these centers account for a large share of the exchange of capital around the world, but they do not function alone. Actors outside the leading financial centers are integrated with the pivotal actors through exchanges of capital. Some of these linkages are internal to the organization of a firm, such as a large multinational bank that has regional offices in various financial centers. The majority of the linkages, however, are between actors in different organizations.

Broad world regional collections of financial centers exist, and these places have a high degree of linkage among themselves, reflecting the economic exchanges within that world region. At the same time, many linkages among financial centers cross world regions. Some of these have hierarchical characteristics, such as the linkages from a financial center to one of the leading global financial centers. Other connections tie financial centers in different world regions which reflect economic exchanges such as flows of manufactures from a producing country to a distant country which consumes these goods. Although politicians often portray their city as competing with other financial centers, the major financial actors do not interpret intercity relations as competitive. Instead, they engage in cooperative relations through intrafirm linkages, and their interfirm ties consist of both cooperation and competition.

Past and Present

The legacy of the past remains a striking feature of international financial centers. London is unequalled as the greatest financial center because firms from all over the world base major offices there to share knowledge about global finance, and their operations extend to every part of the world economy. London solidified this status in the eighteenth century when British merchants usurped Amsterdam's dominance over world trade, and the strength of the British economy and imperial reach over the subsequent centuries enhanced that position.

The United States took over from Britain as the most dynamic economy by the late nineteenth and early twentieth centuries, and New York City became a great financial center. British financial firms remain important yet many firms from other countries are larger. The United Kingdom's economy as measured by gross domestic product ranked fifth in the world in 2004 at $1.6 trillion (US$) just below China at $1.7 trillion. The

United States' $10.8 trillion economy, the first ranked, was almost seven times larger. Nevertheless, London remains the unquestioned leading global financial center, demonstrating that it is the concentration of financial intermediaries which determines importance in global finance and not simply the size of the national economy.

In a similar vein, Amsterdam, the Netherlands and Zurich, Switzerland continue to rank among the top five to ten financial centers based on their premier businesses. Yet the world rankings of their home countries according to gross domestic product are much lower; the Netherlands ranks 15th and Switzerland ranks 19th. Amsterdam was the leading international trading and financial center in the sixteenth century before losing that position to London two centuries later. Today its pivotal financial intermediaries include leading global banks headquartered there (e.g., ABN Amro and ING) and more than 70 foreign financial firms with offices in the city. Zurich became a merchant trading center by the thirteenth century, and many of its modern financial institutions emerged in the nineteenth century. It now houses the global headquarters of such top flight firms as Zurich Financial, Swiss Re, UBS, and Credit Suisse, as well as many other firms, and it has over 100 offices of global financial firms headquartered elsewhere.

Financial centers can break into the top ranks. Tokyo emerged as a major international financial center during the first half of the twentieth century, displacing Yokohama which had been the key center for Japan in the late nineteenth century. Tokyo ranks among the top five or so centers based primarily on its large financial institutions that serve the Japanese economy and function as bridges for Japan to the global economy. Hong Kong did not exist until the 1840s when it became the headquarters for British political, economic, and military activities in Asia. Britain had taken Hong Kong from China after defeating it in the Opium War. Under the protection of the British military, leading British firms established their Asian headquarters in Hong Kong, especially for trade with China. Foreign trading firms quickly joined these premier firms, and top Chinese firms from Asia moved key operations to Hong Kong. By the early 1860s when the Hongkong and Shanghai Bank was established, Hong Kong was on its path to become the trade and financial center of Asia based on the large agglomeration of firms from all over the world which linked Hong Kong to the global economy.

The Social Networks of Financial Capital

The existence of international financial centers cannot be questioned; the huge agglomerations of financial firms in London, New York, Hong Kong, Tokyo, Zurich, and so on testify to that. Nonetheless, their continued resilience seems puzzling, given the innovations in modern telecommunications and information technology such as the trunk lines of fiber optic cables and the satellite systems which transmit vast volumes of voice, text, and data globally at blinding speeds. Why do firms refuse to disperse their global or regional headquarters to widely separated locations where they can enjoy pleasant pastoral settings or small towns?

The mechanism of the social networks of financial capital helps explain the paradox that continued centralization coexists with extensive decentralization of the distribution of information. Financial intermediaries make more sophisticated use of the swelling amounts of complex financial information which is transmitted. They employ this information when they participate in social networks with other financiers, and this enhances their effectiveness in controlling the exchange of capital. These networks are built and reinforced by intense personal exchanges of deep knowledge, and the networks operate at local levels. Daily work life and many other settings, including local governmental and nonprofit boards, social groups and clubs, and religious organizations, as well as social gatherings such as dinners, parties, and other entertainment settings, are venues for these exchanges.

As financial intermediaries participate in these social networks they manipulate their network connections to access complex knowledge about financial opportunities and risks. These networks also reach among financial centers, and long distance travel is used to maintain the network linkages. The financiers move among agglomerations, and each cluster is an intense setting of social networks of financial capital. Failure to participate in these social networks becomes a competitive disadvantage because the individuals or firms who are isolated from the networks miss out on opportunities for profitable financial exchanges.

The Rising Scale of Financial Intermediation

The scale of international financial intermediation has grown dramatically. Since the mid 1990s the external assets of international banks more than doubled, reaching almost $20 trillion (in United States dollars) by the early twenty first century (**Figure 1**). A substantial amount of the flows of global capital consist of savers (e.g., individuals, firms, and governments) providing capital to borrowers. This is indicated by the fivefold rise in international debt securities over the 1995–2005 period to about $15 trillion (**Figure 1**). These capital flows are equal to one half or more of the world's domestic product, which equaled about $35 trillion in 2004.

The sophistication of international financial intermediation is increasing as complex derivatives become the key mechanisms to manage risk, such as countries following different economic trajectories, currencies fluctuating in value, and interest rates varying across countries and over time. International banks create de rivative contracts and exchange them, as well as providing them to their customers (e.g., individuals, firms, and governments). Over the counter derivative contracts are those traded directly between two parties without going through an exchange or intermediary, and their nominal or notional value (gross value of the deal) far exceeds the amounts traded on exchanges. The total notional value of these derivative contracts has soared; between 1998 and 2004 it more than tripled to over $250 trillion, an amount that is over eight times the value of the world's gross domestic product (**Figure 2**).

There are many types of derivative contracts, in cluding foreign exchange, interest rate, equity linked, and commodity, and each of these has variations. Large amounts of foreign currency contracts are exchanged, but the most important derivative contracts are interest rate contracts which are revalued continuously and the timing of gains and losses cannot be readily manipulated. The value of these contracts almost quadrupled during the 1998–2004 period to a total of $252 trillion (**Figure 2**).

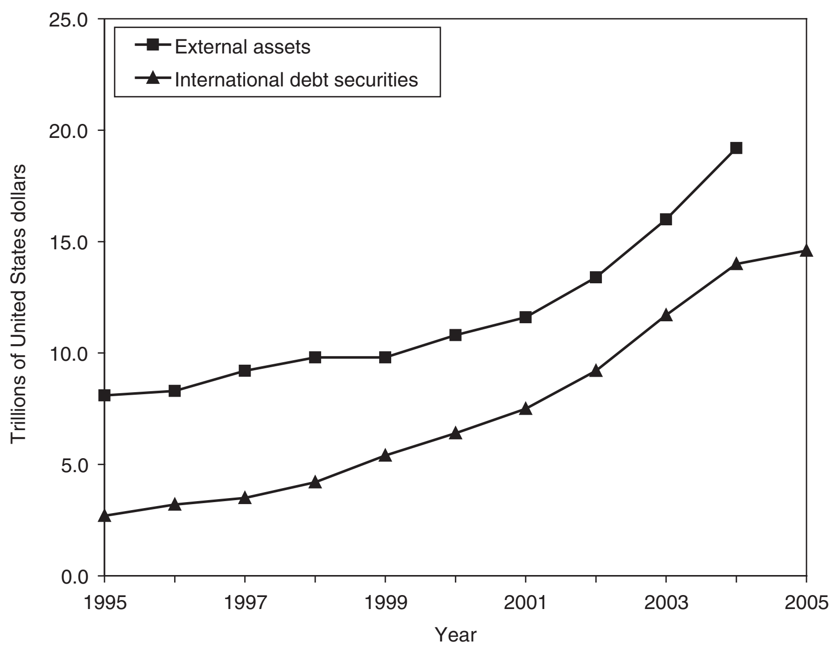

Figure 1 Financial positions of international banks, 1995 2005.

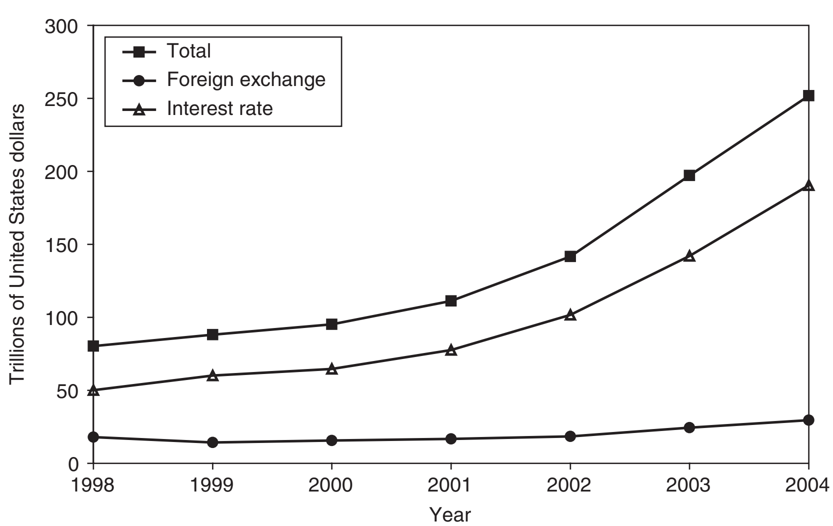

Figure 2 Outstanding over-the-counter derivative contracts of international banks, 1998 2004.

The international financial centers are the headquarters of these managers of risk; and London, New York, and Hong Kong dominate as the decision making centers of global capital.

The Pivots of Financial Capital

London

Since the eighteenth century, the migration of international financiers to London has always been a key feature of its strength. Émigrés from Europe and elsewhere in the world fleeing turmoil in their home countries, as well as financiers seeking their fortune in the great financial center, contributed capital, talent, and their networks of capital to the London agglomeration. These connections built London's hub position in the global economy and buttressed the imperial linkages which were encapsulated in the old saying that the "sun never set on the British Empire." Throughout the twentieth century and continuing into this century the London hub attracted top financial firms requiring a presence in the city's social networks of capital. Because it also has deep connections to the huge economy of Western Europe, any global firm with extensive business in that world region needs a London address.

The scale of London's financial sector is vast. Large United Kingdom banks have their global headquarters in London, including HSBC (US$1277 billion assets in 2004), Barclays ($992 billion), Lloyds ($540 billion), and Standard Chartered ($142 billion), and the other big domestic banks headquartered elsewhere have major international operations in London. The foreign banks in London, however, are the main reason that the city is the pivot of global capital. At least 260 foreign banks have significant offices in London, and they account for over half of all financial assets of banks in the United Kingdom. The foreign banks employ almost 70 000 in the city, and United States banks, especially those headquartered in New York, account for one third of them, indicating the tight bond between these great financial centers.

London specializes in financial services for global clients. It is the world center of over the counter derivatives trading, and one of the top places for exchange traded derivative contracts. It is the second largest center of the world's hedge fund managers (after New York), and the leading European hedge fund center. A wide array of specialized business services such as legal services, accounting and consultancy services, administration, and prime brokerage services support the hedge fund industry. The large concentration of global investment banks in London provides the prime brokerage services. Most of European wide private equity investments are managed from London, and it is second only to the leading places in the United States, especially the New York City area. London is the global leader in over the counter trading of foreign exchange, and this is an outgrowth of its financial networks which reach wider than any other financial center. This gives London exchange traders the best access to knowledge about the direction and causes of changes in currencies around the world. Furthermore, exchange trading brings together a wide range of specialized actors – such as foreign exchange brokers, banks, clients, and securities brokers – whose decisions need to be coordinated.

British financial firms headquartered in London have their regional headquarters for North America in New York, just as United States firms headquartered in New York have their regional headquarters for Europe in London. The agglomeration of the same sets of firms in both cities, firms which both collaborate and compete with each other, enhances the tight financial bonds between the two top cities of the global financial network.

New York

By the early nineteenth century, New York had become the premier financial center of the United States, and its firms maintained close ties to London, a legacy of the long standing merchant trading bonds between the two cities dating from the seventeenth century. The commercial and, later, investment banks, as well as many other specialized financial firms in New York City which emerged over the course of the nineteenth century and throughout the next century, grew to world class status because the United States became the largest national economy. As of the early twenty first century, it accounted for 30% of the world's gross domestic product, twice as much as second ranked Japan. United States financial firms headquartered in New York built their scale and sophistication on the basis of their dominance of the vast domestic market, and they have leveraged that strength in their global expansion as they initially followed the large domestic corporations that expanded internationally. This began in the late nineteenth century and accelerated in the twentieth century, especially after World War II. Now these financial firms operate as exchangers of capital for clients from everywhere in the world.

The world's two largest banks as measured by capital, Citigroup and JP Morgan Chase & Company, are headquartered in New York. The world's fourth ranked by capital, Bank of America, headquartered in Charlotte, North Carolina, houses its global corporate and investment banking, wealth and investment management, and consumer and commercial banking businesses in New York, making it the *de facto* global headquarters of the bank. The city is the global headquarters of six of the ten largest investment banks in the world – JP Morgan, Citigroup, Goldman Sachs, Morgan Stanley, Merrill

Lynch, and Lehman Brothers. Almost 220 foreign banks operate a major office in New York, thus providing a powerful local social network of capital to all of the financial institutions and giving them network ties to every international financial center. Because most of these foreign banks also have a presence in London, the local social networks of capital in each of the world's top two financial centers have tight intra and interfirm ties. Similar to London, New York houses a wide array of specialized financial businesses and of the specialized service firms who assist them. The New York Stock Exchange is the largest in the world, and the city is the leading location for hedge fund managers. It is the most important headquarters for international law firms, and Europe, in particular London, is the principal center for the branch offices of these firms.

The intraorganizational global linkages of Citigroup and Morgan Stanley reveal the sweep of New York's ties (**Table 1**). Citigroup operates in over 100 countries, and it divides the world into six regions, though Mexico and Japan are countries. The list of selected international financial centers in three of the regions reveals that Citigroup embeds itself into the financial networks of the leading metropolis of the respective countries. The firm's offices in the top international financial centers are the headquarters of the corporate and investment banking units of Citigroup which house those financiers who offer the most specialized and sophisticated banking services. The bank operates world regional headquarters. The European headquarters is based in London, and this gives the corporate and investment bankers, private equity investors, and wealth management groups access to their peers in other firms who also supervise European operations. At the same time, Citigroup's corporate and investment banking offices such as in Frankfurt, Germany, while much smaller, still have about 500 employees who work with the 100 global corporations headquartered in Germany. Citigroup reaches into risky yet potentially lucrative financial markets by having a large office in Moscow.

Consistent with the way many global financial firms approach Asia, Citigroup treats Japan separately; its Tokyo office reports directly to the global headquarters in New York City. They take this approach because the Japanese economy is the second largest in the world; thus the sophistication of the bank's financiers must be quite high to deal with financial exchanges within Japan. Equally as important, however, Tokyo is not a pivotal meeting place of the social networks of capital in Asia because the Chinese business networks – the dominant group of Asia – do not intersect in Tokyo with the networks from outside Asia. Hong Kong is that pivotal meeting place of the Chinese and foreign networks of capital. Citigroup operates its Asian headquarters (ex Japan) from Hong Kong, as do most foreign firms with Asia wide operations. The other selected Asian financial centers for Citigroup's operations are premier business centers, including Singapore, Shanghai, and Mumbai. In Latin America, Citigroup operates from the top business centers of each country, most of them the political capitals.

Table 1 The global ties of Citigroup and Morgan Stanley by world region and selected countries and cities, *c.* 2006

Citigroup	*Morgan Stanley*
North America	North America
Mexico	Mexico
Mexico City	Mexico City
Europe, Middle East, and Africa	Europe
Amsterdam, Netherlands	Amsterdam, Netherlands
Budapest, Hungary	Dublin, Ireland
Copenhagen, Denmark	Frankfurt, Germany
Frankfurt, Germany	Geneva, Switzerland
London, United Kingdom	Lisbon, Portugal
Madrid, Spain	London, United Kingdom
Moscow, Russia	Luxembourg
Paris, France	Madrid, Spain
Vienna, Austria	Milan, Italy
Warsaw, Poland	Moscow, Russia
Zurich, Switzerland	Stockholm, Sweden
Japan	Zurich, Switzerland
Tokyo	Latin America
Asia Pacific	Buenos Aires, Argentina
Bangkok, Thailand	Sao Paulo, Brazil
Hong Kong, China	Africa
Kuala Lumpur, Malaysia	Johannesburg, South Africa
Mumbai, India	Middle East
Seoul, Korea	Tel Aviv, Israel
Shanghai, China	Asia Pacific
Singapore	Bangkok, Thailand
Sydney, Australia	Beijing, China
Latin America	Hong Kong, China
Bogota, Columbia	Melbourne, Australia
Buenos Aires, Argentina	Mumbai, India
Caracas, Venezuela	Seoul, Korea
Lima, Peru	Shanghai, China
Quito, Ecuador	Singapore
Santiago, Chile	Sydney, Australia
Sao Paulo, Brazil	Taipei, Taiwan
	Tokyo, Japan

As an investment bank Morgan Stanley is much more specialized than Citigroup which has a large consumer component to its international business. With few exceptions, Morgan Stanley concentrates its activities in financial centers in North America (mainly the United States) and Europe, and its Asian presence is expanding (**Table 1**). While Citigroup has a presence in over 100 countries, Morgan Stanley operates in 30 countries. As with Citigroup, Morgan Stanley has its European headquarters in London, and it has smaller offices in the major national financial centers. Again, like Citigroup, Morgan Stanley's Asian business is supervised from a regional headquarters in Tokyo whose primary responsibility is Japan, and the Hong Kong headquarters oversees each of the offices in the international financial centers in the rest of Asia.

New York thus has tight integration with the world's international financial centers, and its leading firms provide intraorganizational linkages globally. Because the major offices interact with their peers in other firms in each of the centers, this provides another cross cutting level of integration of financial centers. Hong Kong is the third great center of global financial management and exchange.

Hong Kong

Based on measures such as the size of financial institutions with their global headquarters in Hong Kong, the city never has counted high on the world ranking of financial centers. The one bank of renown which had its global headquarters in Hong Kong, the Hongkong and Shanghai Banking Corporation (HSBC), shifted its world headquarters to London prior to the return of Hong Kong to China in 1997. Nonetheless, the bank still derives the majority of its profit from Asia, and its regional headquarters in Hong Kong supervises all of Asia and particularly the rapidly expanding China operations which now use Shanghai as the country headquarters.

Hong Kong is the pivotal financial center of Asia because it is the greatest meeting place of the Chinese and foreign social networks of capital. The nonfinancial firms from around the world house their Asia Pacific regional headquarters in Hong Kong. It has almost 1200 regional headquarters and over 2600 regional offices of international firms. The countries and territories whose domestic corporations choose Hong Kong as their regional headquarters for Asia read like a list of the richest, most economically important places in the world (**Table 2**). The United States leads the list, and the countries of Western Europe are heavily represented. Within Asia, Japan is the leader and second rank overall. Many Japanese corporations choose to run their Asia Pacific business from Hong Kong rather than from Tokyo or other major Japanese cities because it possesses the best social networks of capital for business in Asia. Mainland China firms are shifting their Asia Pacific decision making offices to Hong Kong rather than operating them from their corporate headquarters city, making China the fourth ranked home for firms placing regional headquarters in Hong Kong. It also is the *de facto* global headquarters for a growing share of China's largest firms. This vast agglomeration of firms needing financial services for their operations in Asia constitutes a strong incentive for financial firms from Europe and North America to house their Asia wide regional headquarters in Hong Kong. The leading Overseas Chinese banks in Asia from outside Hong Kong, such as those from Singapore, Malaysia, and Thailand, also place a major office in Hong Kong.

Table 2 Regional headquarters of global firms in Hong Kong by country/territory of incorporation, 2005

Country/territory of incorporation	*Number of regional headquarters*
United States	262
Japan	204
United Kingdom	115
Mainland China	107
Germany	75
Netherlands	54
France	49
Singapore	45
Switzerland	38
Taiwan	33
Italy	20
Sweden	18
Australia	18
South Korea	17
Canada	16
Other	96
Total	1167

The agglomeration of 263 foreign banks in Hong Kong, including 71 of the top 100 banks in the world, is comparable to the clusters in London and New York. Over 70 financial institutions use Hong Kong as their regional headquarters which means they supervise offices in other financial centers in Asia. Another 225 financial institutions have regional offices in Hong Kong which are responsible for all or parts of Asia, although they do not directly supervise offices in other financial centers of Asia. The Hong Kong Stock Exchange is quickly growing into one of the world's largest stock exchanges as Mainland Chinese companies expand their listing on the Hong Kong exchange. As of 2005, their numbers had reached over 300. Foreign exchange and derivative trading likewise are expanding rapidly.

Hong Kong's position as the third global financial center – dominant in Asia as London is dominant in Europe and New York in North America – is being fueled by the growth of China. The Beijing government promotes Hong Kong as China's window to global capital, and its officials steadfastly support that stance with explicit policies. All of the leading Mainland Chinese banks have their key global banking financiers operating in Hong Kong. As India's economic growth continues the Hong Kong financiers are leveraging their Asia wide ties, which have roots reaching back over 150 years, to participate in the economic integration of India and China.

The Future

The leading financial center of every country in the world operates as an international financial center, and some large countries have multiple centers. The domestic financial firms of these countries have the best access to social networks of capital within their respective countries, and these firms have linkages to the leading world financial centers as capital is exported and im ported. Nonetheless, a relatively small number of inter national financial centers house the headquarters of the major global firms that dominate the exchange of capital. Some of these firms have multinational office structures through which they participate in the social networks of capital in countries and world regions. These intraorga nizational networks interface across firms, thus leading to extensive collaboration, as well as competition.

London, New York, and Hong Kong are the key global decision making centers for the exchange of capital, with linkages to most of the world's important international financial centers. The firms in each city also dominate exchanges of capital within their world region, but this does not mean that the exchanges are one sided. Instead, extensive collaboration occurs partly because the leading firms in these cities operate their own intrafirm networks. Innovations in telecommunications, computers, and software increase the capacity to exchange information, allowing the leading financial firms to centralize their most sophisticated decision making in the top inter national cities. These innovations also permit firms in lesser cities to become more sophisticated in their de cision making. These seemingly countervailing trends exist because the flows and complexity of financial ex changes continue to increase at high rates as part of the growth of the global economy (**Figures 1** and **2**). In large countries with rapid economic growth such as China and India we can expect to see some of their international financial centers become more important in the global economy. Cities such as Shanghai in China, Mumbai in India, and Sao Paulo in Brazil will become major head quarters of their domestic international firms, as well as sites for many foreign firms to set up their financial op erations. Differential economic growth and development, therefore, contribute to alteration in the importance of international financial centers, a process which has been going on for centuries.

See also: Finance, Historical Geographies of; Financial Knowledge; Financial Risks and Management.

Further Reading

Bryson, J. R., Daniels, P. W. and Warf, B. (2004). *Service worlds: People, organizations and technologies*. London: Routledge.
Clark, D. (2003). *Urban world/global city*. London: Routledge.
Corbridge, S., Thrift, N. and Martin, R. (eds.) (1994). *Money, power, and space*. Cambridge, MA: Blackwell.
Knox, P. L. and Taylor, P. J. (eds.) (1995). *World cities in a world system*. Cambridge, England: Cambridge University Press.
Lo, F. C. and Yeung, Y. M. (eds.) (1998). *Globalization and the world of large cities*. New York: United Nations University Press.
Meyer, D. R. (2000). *Hong Kong as a global metropolis*. Cambridge, England: Cambridge University Press.
Sassen, S. (2001). *The global city: New York, London, Tokyo* (2nd ed). Princeton, NJ: Princeton University Press.
Sassen, S. (2002). *Global networks, linked cities*. New York: Routledge.
Scott, A. J. (ed.) (2001). *Global city regions: Trends, theory, policy*. New York: Oxford University Press.
Taylor, P. J. (2003). *World city network*. London: Routledge.
Yeung, H. W. C. (2004). *Chinese capitalism in a global era: Towards hybrid capitalism*. London: Routledge.

Relevant Websites

http://www.bis.org
Bank for International Settlements.
http://www.citi.com
Citigroup.
http://www.lboro.ac.uk
GaWC, Globalization and World Cities.
http://www.ifsl.org.uk
International Financial Services London.
http://www.imf.org
International Monetary Fund.
http://www.investhk.gov.hk
InvestHK.
http://www.morganstanley.com
Morgan Stanley.
http://www.worldbank.org
World Bank.

Financial Exclusion

A. Leyshon, University of Nottingham, Nottingham, UK

Glossary

Financial Citizenship A concept that recognizes the significance of the financial system to everyday life and confers a right and ability on individuals and households to participate fully in the economy and to accumulate wealth.

Financial Exclusion The process by which people of low and moderate incomes are directly and indirectly excluded from the formal financial system and denied access to mainstream retail financial services.

Financial Infrastructure Withdrawal The closure of financial services infrastructure, typically branches.

Financial Literacy and Capability The ability of consumers to manage money, keep track of their finances, plan ahead, choose financial products, and stay informed about financial matters.

Prime and Subprime Markets The standard industrial classification of retail financial markets based on the perceived risk of customers. Prime markets are made up in the main of low-risk individuals and households with good credit records. Subprime markets are made up of either low-income and/or credit-impaired individuals and households who are deemed higher risk. As such they pay higher prices for the products they are able to access, which are inflated by a *de facto* insurance premium.

Introduction

Research on the geography of financial exclusion emerged out of concern with the equity effects of the financial services sectors upon communities and localities. The origins of work of this kind can be traced back to the 1970s, and the initial forays of David Harvey into the political economy of urbanization, and the role played by finance in reshaping the built environment through the actions of the institutions that approved or denied mortgage funding, as well as the role of realtors and/or estate agents in acting as gatekeepers and managers of urban and suburban residential property markets. This work suggested that some areas of cities were effectively embargoed, or redlined, a phrase that reflected the practice of drawing a redline around areas on a map within which borrowers and properties were to be denied funding, because they were deemed either to have too high a risk of default or that the quality of the area was such that the properties that were borrowed against might decrease in value. Research of this kind flourished briefly but petered out in the 1980s only to reemerge in the 1990s with a more explicit focus of the issue of financial exclusion.

The catalyst for a renewed interest in such issues was research undertaken on the financial dynamics of the Los Angeles economy. Against a background of the neoliberal reform of the US banking system and the potent racial politics of the Los Angeles metropolitan area, economists Gary Dymski and John Veitch sought to illustrate the role of retail banking in the everyday construction of uneven development in the urban environment. The concept of financial dynamics sought to illustrate how retail banks and other credit granting institutions were responsible for the circulation and redirection of capital within urban economies. In short, Dymksi and Veitch argued that financial institutions used their branch networks to gather in savings and funds from a wide population but would lend it out more selectively according to strategies of risk aversion and profit seeking. Thus, although individuals and households in poorer inner city areas may be able to deposit money through branches located in their neighborhoods – although significant proportions of the population in such areas would have little if no contact with mainstream financial services – there was much less chance of them being granted mortgages and loans for houses purchase and maintenance. Conversely, those individuals and households living in more affluent suburban areas were more likely to be granted access to credit, which means that theoretically they were able to access the intermediated savings and deposits of those less affluent financial customers who were permitted to run current/checking accounts with institutions but were, for reasons of risk aversion, unable to draw down credit. In this way, the inhabitants of more affluent and suburban areas were able, via the intermediation of the financial system, to access funds that not only maintained the market value of properties in such areas – by making purchases possible – but also helped to increase such values by providing funds that paid for the upkeep and improvement of the built environment through maintenance, upgrades, extensions, etc. Similar processes of uneven access to credit and debt operated for local businesses too, generating significant spillover effects within the communities effected.

In this way, then, the financial system was seen to play an active role in the deepening of uneven development within urban economies. It did so by making evaluations of the prospects of individuals, households, and neighborhoods which, through either denying or granting access to credit, become constitutive of such outcomes as suburban areas flourished through an ongoing process of development while inner city areas become neglected and impoverished through a downward spiral of urban decline as residents of such areas find it hard to both sell, buy, or upgrade property while local businesses may be unable to obtain working credit.

This initial foray into the socioeconomic consequences of financial dynamics and financial infrastructure was highly influential in the emergence of a geographical literature on 'financial exclusion' in the UK in the mid 1990s. Financial exclusion was understood as a process by which certain individuals, households, and communities were denied access to the financial system. This work was immediately concerned with closure of bank and building society branches – financial infrastructure withdrawal – influenced as it was by the material evidence of large numbers of branch closures taking place against the background of regulatory reform and a more competitive retail financial service market. Empirical research on the reorganization of bank and building society branch networks in Great Britain revealed that the overall number of branches declined by almost one third between 1989 and 2003 (**Table 1**). During the period 1989–2003 banks closed 36% of their branches. In comparison, the closure rate for building societies – organizations mutually owned by and on behalf of their members – was less than half that of banks, with 17% of their branches closing over the period, while the top ten converted building societies – that is, those that bought out their mutuality during the 1980s and 1990s to convert to public limited company status – closed 22% of their branches over the same period. This suggested that shareholder pressure – the requirement of public limited companies to make sufficient returns to maintain share prices and pay dividends to the legal owners of the business – was a major factor in driving branch closures which were closed more vigorously to cut costs and increase profits.

Closures were also found to vary geographically, as firms adjusted their branch networks to spatial variations in markets (**Table 2**). The average branch closure rate for bank and building society branches for all areas in Britain between 1995 and 2003 was 20%. However, the highest rate of closure – almost 24% – was experienced in 'multicultural metropolitan' areas, which included poor inner city areas. Higher than average rates of branch closure were also experienced in areas defined as 'prospering metropolitan, traditional manufacturing, built up areas', and 'student communities'. Meanwhile, areas that experienced lower than average branch closures tended to be more affluent, which could safely be described as typically Middle England: these were 'suburbs and small towns, coastal and countryside' areas, and 'industrial hinterlands'.

By the late 1990s, work which focused on the reorganization of retail financial services branches was also being undertaken in other economies such as Canada and Australia where similar processes of financial infrastructure withdrawal were observed. For example, in Canada, a series of mergers between large banks had a serious impact on rural branch networks: 75% of rural communities in a study of British Columbia were classified as either being 'moderately' or 'highly' vulnerable to closures according to a bank closure vulnerability index. Meanwhile, the number of bank branches in Australia declined by over 20% between 1981 and 1998. Such observations seemed to confirm that the wave of bank branch closures sweeping industrialized economies was the product of an international process of learning and adaptation to a set of industrial and organizational conventions of how to manage change within the retail financial services sector. Moreover, the outcomes of such conventions were seen to vary over space and, in particular, according to the nature of the institutional context in which they unfold. Thus, in both Canada and Australia, the existence of a richer and more varied financial ecology – such as a vibrant credit union sector, for example – lessened the impact of bank branch closure compared to economies such as the UK which has a thinner and more centralized mix of financial institutions.

Table 1 Branch networks of top six bank groups, top ten converted building societies and top ten building societies, and, Great Britain, 1989 2003

	Branches			*Change (%)*		
	1989	*1995*	*2003*	*1989 95*	*1995 2003*	*1989 2003*
Top 6 bank groups	12 659	10 406	8077	17.8	22.4	36.2
Top 10 converted building societies	3473	3348	2702	3.6	19.3	22.2
Top 10 building societies	1699	1478	1403	13.0	5.1	17.42
Total	17 831	15 232	12 185	14.5	20.0	31.7

From French, S., Leyshon, A. and Signoretta, P. (2008). All gone now: The material, discursive and political erasure of bank and building society branches in the UK. *Antipode* 40(1), 79 101.

Table 2 Branch closures and openings by banks, converted building societies and building societies, by supergroup area, Great Britain, 1995 2003

'Supergroup'	*Total branches 1995*	*Total branches 2003*	*Branch closures, 1995 2003*	*Branch openings, 1995 2003*	*Net change*	*Net change (%)*
Industrial hinterlands	1873	1524	479	130	349	18.6
Traditional manufacturing	1677	1303	499	125	374	22.3
Built-up areas	1832	1424	508	100	408	22.3
Prospering metropolitan	1431	1111	480	159	321	22.4
Student communities	1829	1442	579	192	387	21.4
Multicultural metropolitan	1040	795	329	84	245	23.6
Suburbs and small towns	2651	2209	628	180	448	16.9
Coastal and countryside	2341	1942	497	98	399	17.0
Accessible countryside	164	128	42	6	33	22.0
Total	14 838	11 871	4041	1074	2967	20.0

From French, S., Leyshon, A. and Signoretta, P. (2008). All gone now: The material, discursive and political erasure of bank and building society branches in the UK. *Antipode* 40(1), 79 101.

As work on financial exclusion has progressed, the domination of the field by geographers has declined as researchers from other fields have become interested in the problem. As a result, the understanding of financial exclusion has widened, and it is now recognized that there are at least five kinds of financial exclusion. While the earliest geographical studies were concerned in the main which what might be described as 'access exclusion' – such as branch closure – it is now recognized that there are at least four additional forms of financial exclusion: condition exclusion (where potential customers are excluded for failing to conform to defined product preconditions, such as failing a credit scoring test); price exclusion (where products are priced out of the reach of some consumers); marketing exclusion (where undesirable customers are avoided by aiming marketing campaigns at potentially profitable sociodemographic groups); and self exclusion (where customers do not even apply for certain financial products, either because they have been refused in the past, they assume that they will be denied, or they are unaware of products that they might apply for).

Having outlined the dimensions of financial exclusion, and some of its consequences and effects, the next section considers the wider context within which the social process of financial exclusion has arisen.

Financial Exclusion in Context

Financial exclusion is a feature of a new financial land scape that has emerged in many core capitalist countries over the past 30 years or so which, in turn, has been forged by at least two interrelated processes. First, the re regulation of financial services has been in the vanguard of processes of neoliberalism which have swept market based economies over the past 30 years or so. Initially, processes of neoliberalism were associated with the rollback of the state, as markets were subject to processes of re regulation to make them more open and competitive. Financial sys tems in core capitalist economies such as the US, the UK, and Japan were remade in successive waves of regulatory reform from the late 1970s onwards. For example, in the UK during the 1980s the Conservative government's neoliberal program of regulatory reform dismantled long standing structural forms of industry regulation, which encouraged risk aversion and ensured that different retail financial markets remained institutionally distinctive, in favor of a more prudential regulatory framework designed to encourage competition. As firms began to enter each other's markets in search of market share and profits, the widespread cross subsidization of consumers across dif ferent financial products was swept away and each con sumer increasingly became evaluated in terms of their potential value to the business. Subsequently, neoliber alization has proceeded through new policies by which states have sought to gradually abrogate responsibility for many areas of social life which are now left to the market and to individual and household responsibility. This in cludes greater financial responsibility, and expectations of higher levels of financial capability. This has had the im pact of generating significant new markets for the financial services industry.

Second, the process of neoliberalization has been accompanied by an empowering of financial markets, which has led some commentators to argue that contemporary life is increasingly subject to a process of financialization. Financialization is seen to have agency at a range of scales, including higher levels of instability within the economic system as a whole, and pressure exerted on corporations by capital markets. However, significantly, the process of financialization also increasingly has equity effects as it works its way through the financial system to connect to individuals and households. In particular, retail consumers are being connected to global financial markets as debt repayments are pooled and securitized to create investment products that are sold to international investment institutions. In turn, securitization, and financialization more generally, has been enabled by rapid innovations in information communication technologies, developments which have also driven the emergence of telephony and Internet based retail financial services delivery channels, the centralization of processing tasks, as well as the replacement of traditional face to face systems of risk assessment and management with sophisticated, at a distance credit scoring systems.

The outcome of these processes of neoliberalization and financialization has been the creation of an uneven landscape of financial citizenship. Financial citizenship confers on individuals and households a right and ability to participate fully in the economy and to accumulate wealth. Levels of financial citizenship are uneven, both between and within national economies, and have significant impacts on life chances and upon overall levels of economic development. For example, there are marked divides in the possession of these rights and in levels of financial participation in the UK. At one extreme there is the 40% of the population that might be considered 'super included'; they have savings and investments and are becoming increasingly wealthy because of it. These are the idealized financial citizens of the neoliberal state. Second, there are those that may be considered to be less privileged financial citizens, in that while they are fully inside and included within the financial system, they do not have significant savings or wealth, and while they may have full access to financial products they may have limited assets and be heavily indebted. Finally, there are those that are financially excluded or only partially connected to the financial system; they live in a cash economy, use doorstep financial services, or mainstream but limited products like basic bank accounts.

Thus, the process of financial exclusion is a product of a broader bifurcation of the market for retail financial services. Sociotechnologies such as credit scoring systems enable financial institutions to sort prime from subprime customers. Prime financial markets are made up of individuals and households that possess socioeconomic and geodemographic profiles that make them targets of the marketing and financial strategies of retail financial services firms. These middle and high income customers are actively pursued by retail financial services firms, and may be described as the financially super included, benefiting from intense competition between institutions for their business. One of the drivers of this strategy is the tendency toward the securitization of retail financial products, whereby lenders aggregate the loans made to low risk customers and sell them to investors in international securities markets. Subprime customers, meanwhile, have low to moderate incomes and/or financial assets and are either excluded from mainstream financial marketing campaigns for new products or are denied access to services if they apply.

The geography of prime and subprime financial markets follows established geographies of income and wealth. Thus, for the most part prime retail financial customers may be found in affluent urban and suburban areas, whereas subprime markets are concentrated in areas of low and moderate income, typically in inner city areas (and in Britain, at least, on public sector housing estates). In the absence of mainstream financial services, which continue to close branches in such areas, a host of specialist subprime or 'fringe' retail financial institutions ply their trade. They provide similar services to the mainstream but at a much higher cost. It is now possible to identify pronounced financial ecologies, made up of distinctive combinations of markets, customers, and institutions which are distributed over space.

Financial Exclusion and Public Policy

Financial exclusion has been recognized by policymakers because of the ways in which it disproportionately affects those of low incomes and can exacerbate problems of social exclusion. Early examples of legislation passed to tackle financial exclusion – although it was not identified as such at the time – include the Home Mortgage Disclosure Act (HMDA) and the Community Reinvestment Act (CRA) in the US. The HMDA was passed in 1975 and requires mortgage lenders to make their loan data publicly available for scrutiny. It makes it possible for communities and activists to determine whether financial institutions are serving local housing needs, help local government and other officials attempt to attract private investment to areas where it is needed and, significantly, make it possible to identify possible discriminatory lending patterns. The CRA, meanwhile, was passed in 1977 to ensure that banks and other deposit taking institutions served the credit needs of the communities in which they operate, including low and moderate income neighborhoods, albeit consistent with safe and sound banking operations. As a result, financial institutions are periodically evaluated according to CRA criteria to en sure that credit needs are being met. In many ways, the HMDA and the CRA remain high water marks of anti financial exclusion legislation, although even in the US they have not served to prevent the advance of processes of financial exclusion, and have waxed and waned ac cording to the political predilections of presidential ad ministrations. Both measures were responses to the problems of civil unrest in the US cities in the 1960s that revolved around problems of discrimination and uneven development. A similar policy response took place after the 1992 Los Angeles riots, which saw the growth of a movement which encouraged the growth of community development financial institutions.

Since the late 1990s, the UK government has pion eered legislation that has explicitly sought to combat financial exclusion. The central tenets of the UK finan cial exclusion policy have been fourfold: (1) pressuring banks and leading building societies into offering basic bank accounts, which are stripped down versions of normal current/checking accounts, but without overdraft facilities, specially designed for low income consumers; (2) lining up the post office as a third party distribution network in order to take advantage of the fact that it has more branches than the leading banks and building so cieties put together; (3) seeking to tackle low levels of financial literacy and capability through the promotion of financial education; and (4) to promote third party lenders, such as credit unions, which in some cases are able to provide community based alternatives by lending in areas that are otherwise subject to financial exclusion.

However, what is significant about all these measures is that they were market orientated and sought to avoid any strong forms of government interventions within retail financial services markets. Moreover, in some in stances the measures contradicted other arms of gov ernment policy. For example, the plan to use the thousands of post office branches as *de facto* banks and building society branches to compensate for the closure of actual branches may have been a good idea in prin ciple but failed to take into account that elsewhere the government was seeking to turn the quasi state post of fice into a self sustaining economic organization capable of reproducing itself without government subsidy. Thus, in order to cut operating costs, the post office embarked upon its own round of branch closures replicating the actions of the banks and building societies for which it was supposed to substitute. Moreover, in an effort to open up new income streams the post office signed a co operative agreement with the Bank of Ireland to develop its own range of branded financial products, thereby putting itself in direct competition with the institutions for which it was supposed to be acting as a third party distributor.

Future Research

The growing amount of public policy attention to fi nancial exclusion will have an important bearing on the kind of work being done in this area. There are initiatives on tackling financial exclusion underway in large num bers of Western economies and increasingly at the level of the European Union. As a result, the volume of policy related material has increased markedly in recent years, much of it produced by or for government departments, regulatory bodies, commissions, and policy review bod ies. However, the focus of public policy is toward market based solutions, based on constructions of individuals as consumers, which tends to exclude more radical or interventionist strategies.

An area of work that does need urgent attention is a critical reevaluation of the concept of financial exclusion itself. While initially highly valuable as an idea that drew attention to the ways in which financial marginalization acts as an agent that helps to produce social deprivation, the exclusion–inclusion binary has steered policy toward producing outcomes that can deliver higher levels of ‘inclusion’. While a worthwhile outcome, it is increas ingly recognized that financial inclusion on its own is not necessarily a panacea, and that finer grained and more nuanced understandings of the relationship between in dividuals and households and financial services markets are required. Ways forward here include the development of concepts such as financial citizenship and financial ecologies, which have the advantage of examining fi nancial markets as a whole, not just the boundaries. These concepts may have a better purchase on the relationship between social position and geography in the production of the position of individuals and households within fi nancial markets.

See also: Public Policy; Uneven Development.

Further Reading

Dymski, G. A. (2005). Financial globalization, social exclusion and financial crisis. *International Review of Applied Economics* 19(4), 439 457.

Dymski, G. A. and Veitch, J. M. (1996). Financial transformation and the metropolis: Booms, busts and banking in Los Angeles. *Environment and Planning A* 28, 1233 1260.

French, S., Leyshon, A. and Signoretta, P. (2008). All gone now: The material, discursive and political erasure of bank and building society branches in the UK. *Antipode* 40(1), 79 101.

Leyshon, A. and Thrift, N. (1997). *Money/Space: Geographies of Monetary Transformation*. London: Routledge.

Leyshon, A., Burton, D., Knights, D., Alferoff, C. and Signoretta, P. (2004). Towards an ecology of retail financial services: Understanding the persistence of door to door credit and insurance providers. *Environment and Planning A* 36, 625 645.

Leyshon, A., Thrift, N. and Pratt, J. (1998). Reading financial services: Texts, consumers, and financial literacy. *Environment and Planning D: Society and Space* 16, 29 55.

Marshall, J. N. (2004). 'Financial institutions in disadvantaged areas: A comparative analysis of policies encouraging financial inclusion in Britain and the United States. *Environment and Planning A* 36, 241 261.

Pollard, J. S. (1996). Banking at the margins: A geography of financial exclusion in Los Angeles. *Environment and Planning A* 28, 1209 1232.

Relevant Websites

http://www.citizensadvice.org.uk
Citizens Advice Bureau.

http://www.ffiec.gov
Federal Financial Institutions Examination Council.

http://www.fsa.gov.uk
Financial Services Authority.

http://hm treasury.gov.uk
HM Treasury.

http://www.which.co.uk
'Which?' Magazine.

Financial Knowledge

G. L. Clark, Oxford University Centre for the Environment, Oxford, UK
J. C. Marshall[†]**,** University of Oxford, Oxford, UK
K. Strauss, Oxford University Centre for the Environment, Oxford, UK

Glossary

Cognitive Psychology Branch of psychology that stresses individuals' information-processing capacities and the constraints thereon.

Context Used here to indicate the significance of the environment in which people make decisions including the social resources at hand.

Financial Information Data, indicators, and related materials representing the performance of financial markets.

Financial Knowledge Substantive interpretations of the underlying causes and consequences of financial market performance.

Intuition A complex automatic response involving cognition, experience, and emotion rather than conscious calculated response to an issue or situation.

Liberalism Branch of political philosophy that stresses the autonomy and significance of the individual, most often attributed to British writers of the eighteenth and nineteenth centuries.

Neoliberalism A term used by contemporary writers to signify the retreat of the state and the re-emergence of individualism driven by market imperatives.

Rationality An idealized conception of the process whereby individuals make decisions consistent with their best interests.

Risk Estimated probability of a loss-making event or class of events occurring over a well-defined period of time (linked with the rate of return in financial markets).

Uncertainty Unknown events and circumstances that may affect individual welfare (especially relevant to nonstationary financial markets).

Introduction

In the space of about 25 years, the welfare state and its attendant principles and practices has been heavily eroded in virtually all Western advanced economies. Left behind is a commitment to maintaining adequate standards of living for the population as a whole (especially in the Anglo American economies) and a dawning realization that the state cannot guarantee welfare in the face of globalization and demographic ageing (especially in continental European countries). Liberalism has re emerged as a core political ethos prescribing the respective responsibilities of the individual, the corporation, the community, and the state in relation to social welfare. Notice, however, that the liberalism associated with John Stuart Mill combined a commitment to individual autonomy and responsibility with a recognition that not all people were equipped by reason of intelligence, education, and moral character to be fully responsible. By contrast, neoliberalism has little patience with state paternalism; individuals (and their families) are deemed responsible for their well being in the market economy.

With political neoliberalism has come the hegemony of a certain type of economic theory predicated upon individual rationality and utility maximization. Being both a recipe for theory and a normative vision for society, this form of economics has been associated with the de regulation of financial markets, the fostering of shareholder (as opposed to stakeholder) value, and the global integration of trade and currency transactions. Compared to the broad sweep of the twentieth century post World War I, individuals are increasingly responsible for their own long term welfare, itself a function of their performance in financial markets. And yet, at its moment of victory, the rational actor model has seen its theoretical hegemony slip away in the face of critiques from behavioral finance and cognitive psychology. Empirical in focus and subject to the exacting protocols of shared testing regimes, cognitive science has raised significant doubts about the capacity of people to play the role assigned to them by neoliberalism. We may be rational, but not rational enough.

If people are to be the rational actors of theory and policy, they need to process information in a manner consistent with their own interests and in relation to the interests of others in a well functioning market economy. Given their dependence upon the burgeoning financial markets of the late twentieth and early twenty first centuries, individuals need a certain quality and quantity of financial knowledge such that their decision making will promote their long term (if not short term) well being. By this logic, financial knowledge is the bedrock of neoliberalism; it is a necessary ingredient in any government policy that attempts to shift responsibility for social welfare to the individual. As such, financial knowledge is a matter of substance (knowing how

[†] Deceased, 22 August 2007.

markets function, for example), and a matter of social identity and even morality (being responsible, informed citizens).

Scope and Significance

Financial knowledge clearly matters in many aspects of everyday life, some of a short term nature and some with profound long term consequences for welfare. Financial knowledge is used for rudimentary decisions such as discriminating between consumption items on the basis of cost and quality. So, for example, when we assess washing machines and dryers the 'rational' model suggests that we should do so by comparing the attributes of various options assuming a budget constraint and some base line preferences of expected use and location in the home. As summarized, this assessment process could be quite formal and explicit. In fact, the evidence suggests that people use shortcuts or heuristics to sort the available options into manageable packets where, in the last instance, issues of cost and value may be used as tie breakers between equally desirable products. In these situations, substantive financial knowledge may not be necessary to make a decision. There are, nevertheless, a variety of public and private agencies whose mandate is to encourage 'informed' financial decision making.

There are, however, other more important decisions that cannot be easily reversed once taken and the costs of which cannot be recouped in any simple sense. For example, in many countries people purchase healthcare insurance plans where the long term costs of under insurance may be so significant that short term cost effectiveness should be ignored. Likewise, in many countries, the middle class are increasingly significant consumers of education for their children; cost effectiveness must be balanced against the expected long term rate of return of human capital. In these situations, where past decisions affect long term welfare, financial knowledge is closely related to individual's discount functions – the value attributed to short term consumption against expected long term well being. People might reasonably apply some rudimentary financial knowledge and decision techniques in all the above cases. However, buying a washer–dryer is different in 'type' from buying healthcare, education, and pensions.

The distinction drawn here is between financial knowledge that is relatively shallow, being contingent on current information about short term benefits, and financial knowledge that is relatively deep, being contingent upon long term expected prospects and their consequences. The former refers to knowledge and decisions that are typically reversible (at modest cost), whereas the latter refers to knowledge and decisions that are in large part irreversible (albeit with exceptions in extreme cases). It can be observed that the former is almost always about events given individuals' background circumstances – the only knowledge needed to make a decision is that inherited from past decisions (e.g., other products of manufacturer X have been reliable and good value for money). The latter may require knowledge only partially available from past decisions and is, more often than not, dependent upon financial expertise not shared widely through society. For example, calibrating the risk adjusted rate of return on one kind of healthcare policy over other kinds of healthcare policies requires a wide range of detailed knowledge (including knowledge about the relationship between healthcare and lifetime earnings).

Whether about rudimentary or complex time dependent issues, financial knowledge must be assembled and then applied through decisions. Therefore, financial knowledge has more value for its instrumental use than for its intrinsic quality. This suggests two implications. First, individuals must decide on how much and of what quality information to collect, recognizing that the costs of assembling data must be balanced against its value in decision making. Second, given the costs of decision making (time, effort, and other opportunities foregone) people have incentives to apply decision templates rather than approach each decision afresh on its own merits. Inevitably, much financial knowledge is generic but, in some instances, it may also be case specific. There are significant incentives in the financial markets for commercial vendors of information to blur the two together, treating the former as the solution to the latter.

Rationality and Knowledge

Rationality is a universal characteristic of people whatever their context or culture: it could hardly be otherwise given biological imperatives. However, individuals who share a basic cognitive capacity as human beings will also vary considerably in terms of their cognitive performance, including their ability to logically assess and evaluate alternatives. More significantly what counts as sensible (or acceptable) in some societies will not be necessarily the same in other societies; context and culture (being here synonyms for environmental factors) are crucial when people evaluate or judge behavior against social standards. Rationality represented by the subjective expected utility (SEU) maximization model makes two specific assumptions about the utilization of financial knowledge. First, given the price of knowledge it is rational to economize on its collection and use and, second, outcomes (positive and negative) are symmetrical in that they are equally valued for their consequences.

Unfortunately for the SEU model, neither of these assumptions hold true in real life. It is self evident that agents vary in terms of the financial knowledge they can afford. The knowledge they can afford thereby affects the options they consider, and the options considered may be suboptimal in terms of the maximization of individual welfare. With experience calibrated using past decision metrics, individuals may become isolated from the best options and even from the better options they can afford. This type of behavior may be legitimated by cultural preferences which serve to justify the favoring of some options over other options. It is also apparent that many people whatever their sociodemographic status are risk averse, preferring the certainty of a known but small ‘win’ over a much larger but risky potential ‘win’. Being risk averse may be more or less valued in different settings, attracting the admiration of some and the approbation of others. These cultural cues are likely to constrain, reinforce or, in some instances, determine behavior.

Most importantly, SEU models of rationality have been criticized for their shallowness regarding social identity and the significance of the emotions. It is frequently assumed that social identity and the emotions adversely affect reason because they filter what is observed and the implications drawn thereof. Therefore, to the extent that reason is informed by information and financial knowledge, social commitments and emotions are thought to prompt wishful thinking and biased or capricious decision making. However, cognitive psychologists have sought to counter these theoretical assumptions with empirical evidence suggesting that the emotions may be a valuable intuitive device for first order response to new situations. If financial knowledge has a formal quality (perhaps learnt from or explained by others) such that it is a means of assessing the virtues or otherwise of competing options, the emotions may provide a simple mechanism for presorting options by relevance. By this account, explicit financial knowledge could be added to the mix of decision techniques at the end of a sequence of more intuitive judgments rather than at the start of the process.

In practice, people are more or less rational; they approach problems from two intersecting vantage points – their intrinsic cognitive capacity *and* their experience in certain settings (Herbert Simon’s scissors). Simon believed in bounded rationality and his famous metaphor created a picture of individual decision making in which the twin blades of cognition and environment come together to produce a decision outcome. People are also innately emotional in the sense that they bring to situations intuitive judgments based upon fear, anger, happiness, and love (see below). For some theorists, however, financial knowledge is an antidote to the confounding affects of emotion and commitment. However, this seems naïve: the results of decision making – behaviour – have cultural and social significance including, crucially, emotional significance. Being purely rational is more a normative statement of supposedly ideal behavior rather than a positive statement of fact. It is also often an ideologically driven assessment of socially acceptable (even laudable) behavior in the context of market based late capitalist societies.

Risk and Uncertainty

Financial markets are awash with information. There are information markets for vendors and for buyers – in fact, the problem is not so much the lack of information so much as the lack of means to discern relevant information and our inability to discriminate between information and information sources in terms of their quality. Directly or indirectly, financial products inevitably involve risk and uncertainty. So, for example, the purchase of a home mortgage may involve the vendor and buyer in complex time dependent calculations of expected inflation and interest rates. Vendors who underestimate interest rates over a specified period of time effectively subsidize mortgagees (and vice versa). Moreover, many people do not appropriately weigh low incidence but costly risks. Continuing the mortgage example, during the 1980s and 1990s many UK home buyers took out endowment mortgages, betting that a low risk event (collapse of the stock market) would not occur or would not be compounded by collateral threats to household income and wealth.

Many people are unaware of the risks they face in everyday life. They certainly do not appreciate the fact that risks are related such that one event may cascade to affect a person’s entire well being. To deal with these issues in regard to finance, governments have vigorously encouraged disclosure policies, transparency in product design and management, and the adoption of risk assessment by formal decision trees that deliberately expose contingent risks (see below). Most importantly, governments have encouraged ‘plain English’ disclosure policies using terms and concepts anchored by average rather than expert competence. However as suggested above, these policies are compromised on three counts. First, rationality is idealized, eschewing the complications of context and culture. Second, rationality is portrayed as an issue of substance rather than of process. Third, rationality is conceived as unbounded notwithstanding the fact that people seek to limit the scope of a problem rather than see it as a whole.

Naïve investors assume that financial information is representative of stable economic processes. Financial information is taken on face value with little appreciation of the motives of those who produce and market such information. In part, this reflects the lack of alternatives – independent cost effective expert information is hard to obtain. By necessity, risk is calibrated using past information and extrapolated into the future. Therefore, naïve investors tend to use financial information in ways that reinforce past commitments; information at odds with individual predispositions is often discarded or ig nored, suggesting that the costs of assimilating such in formation make it easier to wait for exogenous events to force through reconciliation or a change in tactics. In practice, näive investors tend to associate cost with reputation; rightly or wrongly, the presumption is that the higher the cost of financial information the more reliable the information.

Sophisticated investors closely scrutinize the integrity of financial information, seeking evidence of contamina tion by competing interests. Rather than extrapolating surface trends, they look for changes in the variance of underlying time series, believing that this is one indi cation of market instability. Sophisticated investors be lieve that uncertainty is endemic; finely calibrated risk profiles of financial products, the core of advertising programs by the finance industry, suggest a level of cer tainty that is not justified by the performance of financial markets. For sophisticated investors, financial knowledge is more important than financial information. However, there is a paradox: financial knowledge must be con tinuously assessed in the light of changing circumstances. It must be constructed and deconstructed at every turn.

Stylized Facts

If it has become apparent that people are not the efficient information processing machines of economic theory or, for that matter, government policy (although we should recognize that government policy is hardly ever coherent on this issue given the vestiges of paternalism that infuse the modern nation state). On average, people are risk averse, eschewing financial opportunities that a so called fully rational person would or should assume. On aver age, people are inefficient users of information, often backing ill informed opinion when they should collect more information or, perversely, collecting more infor mation that is warranted by the scope of the problem. On average, people overvalue the near future and undervalue the long term future. On average, people do not carry through on past plans, 'jumping at shadows' when they ought to back informed commitments. On average, peo ple are poor at data analysis, befuddled by even the most elementary notions of probability and contingent risk.

However, the results of cognitive science do not tell the whole story. Research on the financial behavior and planning of different sorts of people suggests a more nuanced picture. While the precise details vary by country, especially if we were to include developed economies outside of the Organisation for Economic Co operation and Development (OECD), it appears that financial literacy is distributed on a 20/60/20 basis in most advanced capitalist countries. That is, 20% of the population appear to have such poor financial com petence, knowledge, and access to advice that their im mediate and long term welfare is imperiled. At the other end of the distribution, 20% of the population are best characterized as sophisticated investors combining access to expertise with knowledge and understanding of fi nancial markets. In between, however, 60% of the population are, at best, naïve investors subject to many of the cognitive shortcomings noted above but with the resources to be active consumers of financial products.

It is tempting to correlate cognitive ability with age, gender, and socioeconomic status. There is, for example, a clear relationship between older age and some types of cognitive decline. On the other hand, folk law in the finance industry would have it that men and women are different in that the former are analytical and the latter are emotional (left brain/right brain dominant, etc.). This claim is empirically ill founded and makes gross gener alizations about people from a specific environment with a distinctive and hierarchical segmentation of tasks and functions.

Social scientists believe that sociodemographic status counts because status is a good proxy for income and educational resources – the assets needed to acquire fi nancial knowledge and expertise so as to compensate for acknowledged cognitive shortcomings. Moreover, cogni tive psychologists tend to at best ignore, and at worst discount, the significance of sociodemographic charac teristics and their testing regimes often involve groups of undergraduates and MBA students from elite uni versities. Studies have shown that risk aversion, for ex ample, is frequently correlated with income and household wealth, which chimes with the concentration of equity ownership along class lines (notwithstanding the 'shareholder society' supposedly heralded by the dot com boom). The 'interaction' of sociodemographic characteristics in the context of financial knowledge and decision making is understudied.

Most importantly, it appears that formal education, professional qualifications, and task specific training can make a difference to people's financial expertise. University level education in subjects demanding quan titative skills and the attainment of professional qualifi cations postuniversity in areas of related knowledge do make a difference, and long term task specific training seems to reinforce the advantages of education and professional qualifications. Inevitably, education, pro fessional qualifications, and training are correlated, albeit imperfectly, with household income. It is also apparent that, not withstanding the massive post war improvement in educational attainment, in Western societies high quality education, quantitative skills, and training are socially stratified not widely distributed.

Context of Knowing

By this account, financial knowledge is intimately related to individual decision making. Just as plainly, the quality and quantity of financial information available to individuals is related to their socioeconomic status. There also appears to be a relationship between expertise, education, qualifications, and training. Most importantly, many people have neither the cognitive ability nor the acquired skills to be expert decision makers. They must therefore rely upon others through social networks and organizations to improve the acquisition and use of financial knowledge. The evidence suggests, in fact, that most people depend upon family and friends when seeking to extend their financial knowledge and better calibrate their decision making. By this logic, the household is a very important resource for information, advice, and decision clues. It need not contain any more expertise than the sum of its parts; nonetheless, it may be able to rule out extremes and fanciful assumptions.

However, small groups may be also subject to the influence of dominant individuals (whatever their expertise or lack thereof) and may converge on conventional or even catastrophic solutions rather than the best or even second best solutions. This is an argument for broadening the decision environment to include the workplace and wider sections of society at large. Some 30 or 40 years ago, the workplace combined paternalism with class specific entitlements such that the available knowledge was allocated by virtue of job classification. However, evidence on the value of workplace information exchange is mixed; surveys suggest that employees rarely share financial information with one another. More often than not, employers provide the relevant information and brief programs on the issues and options. While employees often indicate that this information is useful, sign up rates for information briefings are reported to be low with default options dominating deliberate employee choice. Research on pension plan participants has shown that the behavioral effects of such training programs are limited despite positive attitudinal responses.

Another source of financial knowledge is peer imitation and advertising. If conventional class related social aspirations are less meaningful in (post)modern societies, it is apparent that many people make financial decisions according to their role models (public figures) and their peers (represented in the media by television programs and the like). In part, consumption by imitation may be driven by workplace relationships and professional ties. In part, consumption by imitation may signal aspirations of social identity and recognition. Consequently, financial knowledge is nothing special; consumption (the accumulation and assessment) of financial knowledge is likely governed by the same behavioral customs and conventions that govern the purchase of fashion items such as clothes and automobiles, where you live, and the organizations to which you belong. In fact, it's hard to believe otherwise.

In this world of image, media, and consumption, the internet and web based resources are sometimes identified as important alternative (good and bad) sources of financial information. It appears that access to this medium is heavily dependent upon an individual's work environment and job tasks. It also appears that younger people are more willing to use the web to search for information and to make purchasing decisions based on web information than older people (a finding that will not surprise those knowledgeable of the adoption of innovation). However, most websites, especially those sponsored by government and not for profit consumer advocates, utilize an abstract display format rather than a media driven format, emphasizing a claimed difference between (important) financial knowledge and other kinds of (less important) consumer information. Whether this is effective remains to be seen.

Emotional Knowledge

The 'contextual' model of financial knowledge suggests that it is made at the intersection between individual cognition and the environment (the family, place of work, society at large, etc.). By this logic, the exercise of cognition provides incomplete solutions if not supplemented by specific learning (formal and informal). Clearly, then, social and cultural context matters both in terms of what counts as financial knowledge and in terms of how individuals reach economic and financial decisions. The worlds of image and media consumption, of work, and of peer and professional association discussed in the previous section are not geographically and culturally uniform.

Financial knowledge is socially constructed and often locally embedded, despite the much heralded effects of globalization on information, as well as capital flows. Much of the research discussed in this contribution originates from, and relates to, the global North, but financial knowledge is not hegemonically constructed by the firms and institutions of the 'developed' world. The emerging economic strength of China and its embrace of certain forms of capitalism and entrepreneurialism under the rubric of communism is but one example of a substantially different financial knowledge context. The different responses of Asian governments to the financial crises of the late 1990s suggest others, as do the socialist and populist Latin American regimes of Hugo Chavez and Evo Morales. Moreover these countries are characterized by huge social and economic inequalities: sophisticated, globally integrated, financially wealthy investors exist alongside small (though in many cases growing) middle classes and substantial populations of very poor individuals, many of whom depend on subsistence farming.

In this sense geography also matters, particularly in relation to the mutable scales at which individuals experience financial decision making and the construction and dissemination of financial knowledge. Geography is not only key to an understanding of the context of decision making because it describes the role of place and the importance of spatial embeddedness: geographers' insistence on the entwined nature of culture and economy (without one being reducible to the other) is also valuable. The research programs of economists and cognitive scientists are strong on behavioral modeling, the identification of nonrational behaviors in experimental settings, and the extrapolation of behavioral effects from empirical data. Geographers have the capacity to contribute to a nuanced understanding of context that fleshes out the environment of decision making.

However, there is another sense in which cognition is incomplete – the emotions also play a significant role in determining 'knowledge' and behavior. To agree with this statement is to move a significant step away from the rational actor paradigm that dominated the social sciences over the second half of the twentieth century and embrace a twenty first century theory of the mind–brain that stresses cognitive integration and functional interdependence. It is also to accept, as suggested above, that the rational individual is not an island: cultural and societal interactions, personal relationships, and environmental and geographical factors matter. Thus, it provides a clue as to the nature of intuition – the moment when people reach a solution by some 'unconscious' mix of cognitive skill, social resources, and emotion.

Basic emotions have a built in physiological origin; emotions such as fear, disgust, happiness, and anger are mediated by innate brain circuits in higher primates, we included, although their expression can be modulated by social learning. Other emotions like love are even more modulated by culture and learning and more ascriptive in origin; ascribed behavior draws social approbation or admiration. Following Aristotle, the emotions can also be thought of as a form of visceral judgment made about a situation and the behavior that follows. In this sense, the emotions 'value' information by visceral judgment, providing a response that may or may not be further conditioned by cognition and social resources. This way of integrating the emotions with decision making and behavior recognizes that most behavior is responsive rather than reflective.

So, for example, fear may be the response of many people to a financial market crisis. As a basic emotion, but with a large overlay of acquired triggers and meanings, it represents foreboding about a sudden reversal of fortune bringing to the fore the likely negative consequences of such an event. It may follow a stock market bubble characterized by euphoria and the suspension of disbelief ('irrational exuberance'), equally deep seated emotions from our biological past but applied to very different material circumstances. It is little wonder, therefore, that risk aversion is so systematic amongst humans across the whole range of experience, whether they are educated or not, skilled or not, or reflective or not. Taking this one step further, fear and happiness may (respectively) amplify or discount adverse financial information, just as these emotions may select amongst the available financial information to fashion an interpretation of current circumstances. Fear and happiness are shared human experiences that can lead to 'herd behavior'.

For those less convinced about the role and significance of the emotions in financial knowledge, consider a more social emotion: love. Clearly, love is associated with biological drives such as sexual desire and reproduction. On the other hand, love is also associated with socially desired behavior such as care for the young, respect for the old, and a commitment to the welfare of family and friends. Not all people are bound by love; some are entirely selfish. However, the comparison is revealing. Those committed to others may have very shallow discount functions, whereas those who are wholly selfish may have very steep discount functions. Love may give rise to hyperbolic discount functions, whereas selfishness may manifest as weakness of will. As feminists have long pointed out, caring behavior is essential for both social reproduction and social cohesion. Either way, people's need or otherwise for approval may be a powerful determinant of how they respond to financial information and knowledge.

Future Research

Financial knowledge is a field of research and an area of public policy that is bound to grow in significance over the coming years. Its place at the intersection between a retreating state and the growing responsibilities of individuals for their own welfare suggest that it is one of the most important determinants of long term social justice. At the same time, so many of us appear to be naïve decision makers and investors; we lack the skills of financial professionals and will never have enough information and knowledge to makeup the difference. That relatively few people have the capacity to structure decision making in ways that manage cognitive shortfalls, the costs and benefits of social resources, and the play of emotions suggests that some market agents will benefit from exploiting those differences. Here lies an issue of political economy in that the welfare consequences of such inadequacies are bound to flow back into the political system.

Cognitive science has made heavy inroads into the coherence and plausibility of the rational actor model. Research findings have emphasized the inherent nature of many of the anomalies and biases taken by some to be aberrations easily managed by self conscious planning. At the limit, the revolution in cognitive science threatens to overturn the notion of agency and responsibility. How ever, this is an extreme interpretation: rationality remains a human trait. Most often, human cognition fails or proves inadequate in circumstances that require a degree of precision that few people in the course of everyday life would recognize as realistic. One of the most important research frontiers is at the interface between cognition and the environment where Herbert Simon's scissors are in play – where decisions have to be taken that require not just cognitive capacity but also an appreciation of the circumstances of specific situations.

This also requires an appreciation of the breadth and diversity of those situations, culturally, socially, and geographically, beyond the advanced capitalist econ omies. The contribution of a group of anthropologists, psychologists, and economists studying game theory across a large number of small scale 'traditional' societies (and in comparison with the results of similar experi ments carried out in university undergraduates in Europe and North America) has been to show that the norm of utility maximization is not universal. On the other hand the results of their analysis demonstrated that social learning was a statistically significant factor in observed behavioral differences. Cross cultural accounts of the construction and dissemination of financial knowledge are important to a nuanced understanding of the context of individual decision making.

As described, financial knowledge is vital to the well being of adult men and women. Less understood are the possible payoffs of financial education and literacy pro grams aimed at young children and teenagers. In play, in this regard, are a series of complex issues having to do with the cognitive development of children as they ma ture to adulthood and the degree to which they are able to assimilate information from the environment in their putative decision frameworks. At the other end of the age distribution, there are issues of similar significance, not ably the consequences of ageing for cognitive functioning and the capacity of the elderly to respond to changing circumstances. At its core, liberalism values human au tonomy and responsibility whatever people's ages. However, there are counterarguments for the desirability of some level of paternalism, especially on the side of those who are simply not able to play the roles assigned to them by neoliberalism.

In any event, there remains a most troubling issue: how individual choice based on financial knowledge is exercised in the context of culture and the emotions. It was suggested above that intuition is one expression of their interrelationships. It was also suggested that a single focus on cognition at the expense of cultural expectations and emotional commitments would be incomplete as a matter of empirical reality and incomplete as a matter of policy relevance. While easily expressed as such, there are few frameworks available that allow for a means of integration. Furthermore, given the significance of fi nancial information and knowledge for well being, the academic literature has preferred silence on these issues to a full fledged research program that would set the shape of things to come.

See also: Emotional Geographies; Finance, Historical Geographies of; Financial Risks and Management; Neoliberalism; Psychoanalysis; Psychoanalytic Theory/ Psychoanalytic Geographies; Uncertainty.

Further Reading

Ainslie, G. (2001). *Breakdown of Will*. Cambridge: Cambridge University Press.

Baron, J. (1994). *Thinking and Deciding* (2nd edn.). Cambridge: Cambridge University Press.

Bertrand, M., Karlan, D., Mullainathan, S., Shafir, E. and Zinman, J. (2005). *What's psychology worth? A Field Experiment in the Consumer Credit Market. Working Paper 11892*. Cambridge, MA: National Bureau of Economic Research.

Brocas, I. and Carrillo, J. (eds.) (2003). *The Psychology of Economic Decisions, vol. 1: Rationality and Well being*. Cambridge: Cambridge University Press.

Bruner, J. (1990). *Acts of Meaning*. Cambridge, MA: Harvard University Press.

Clark, G. L., Caerlewy Smith, E. and Marshall, J. (2007). The consistency of UK pension fund trustee decision making. *Journal of Pension Economics and Finance* 6(1), 67 86.

Clark, G. L., Munnell, A. and Orszag, M. (eds.) (2006). *The Oxford Handbook of Pensions and Retirement Income*. Oxford: Oxford University Press.

Damagio, A. (ed.) (2003). *Looking for Spinoza*. London: Vintage.

Gigerenzer, G., Todd, P. M. and The ABC Research Group (1999). *Simple Heuristics that Make us Smart*. New York: Oxford University Press.

Gross, D. M. (2006). *The secret history of emotion: From Aristotle's Rhetoric to Modern Brain Science*. Chicago: University of Chicago Press.

Henrich, J., Boyd, R. and Bowles, S. (2006). 'Economic Man' in cross cultural perspective: Behavioural experiments in 15 small scale societies. *Behavioural and Brain Sciences* 28(06), 795 815.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica* 47, 263 291.

Leyshon, A. and Thrift, N. (1997). *Money/space: Geographies of Monetary Transformation*. London: Routledge.

Nussbaum, M. (1994). *The Therapy of Desire*. Princeton: Princeton University Press.

Ortony, A., Clore, G. L. and Collins, A. (1988). *The Cognitive Structure of Emotions*. Cambridge: Cambridge University Press.

Shiller, R. (2001). *Irrational Exuberance*. Princeton: Princeton University Press.

Simon, H. A. (1982). *Models of Bounded Rationality*. Cambridge, MA: MIT Press.

Thrift, N. (2005). *Knowing Capitalism*. London: Sage.

Wagner, R. K. (2002). Smart people doing dumb things: The case of managerial incompetence. In Sternberg, R. J. (ed.) *Why Smart People can be so Stupid*, pp 42 63. New Haven: Yale University Press.

Relevant Websites

www.behavior.org/econ
behavior.org.
www.fsa.gov.uk
Financial Services Authority.
www.nber.org
National Bureau of Economic Research.
www.oecd.org
Organisation for Economic Co operation and Development.

Financial Risks and Management

P. Bennett, University of Edinburgh, Edinburgh, UK

Glossary

Derivative A collective term for a range of financial products, including options and futures, that are derived from something else, such as real commodities and indexes of virtually anything.

Future A financial product which takes the form of a contract to buy something (which does not have to be a real commodity) in the future at a price agreed in the present.

Hedging An act which protects against financial loss and future fluctuations in prices, such as by purchasing a derivative.

Moral Hazard The controversial idea that people, institutions, or even states will take less care over their financial well-being if they know their risk of falling into financial difficulty is insured.

Reinsurance Insurance for insurers. Reinsurers normally take on the more extreme portions of risks covered by ordinary insurance companies.

Retroactivity The application of current laws to events that took place before the law was passed.

Strict Liability Legal responsibility for an event which applies even where no fault or negligence is proven.

Vertical Disintegration The outsourcing of parts of the supply chain to other firms, rather than managing the entire production chain in-house.

Financial Risks and the Risk Society

Risks are not events in themselves, be they accidents to individuals, natural hazards, or global stock market crashes, but rather the ways in which the potential negative effects arising from such events are interpreted. Risk, then, is the interpretation of the potential for suffering a harm of some kind – be that physical, emotional, financial, or any other kind of harm. Inevitably, in a capitalist society, risks are commonly interpreted in financial terms, and their multifaceted dimensions are often reduced to the potential for suffering financial losses.

This was reinforced in 1921 by Frank Knight who influentially defined risk in relation to its management through the insurance industry. He used the term risk for situations in which, while the outcome of an individual event was unknown, the probability of the various potential outcomes could be calculated based on the history of previous occurrences and/or mathematical calculation. This was in contrast to uncertainties, which were situations in which even the odds were unknown. Of course, one individual's uncertainty is an insurance company's risk, because while an individual may lack historical data (not to mention actuarial training), the insurer can aggregate together the outcomes of many different events over both time and space. This allows the probability of individual events occurring to be established, and premiums to be set for all that cover the expected losses of the few. For Knight, risks were something for which insurers were prepared to offer insurance while uncertainties were uninsurable.

More recently, insurability (or lack of it) is a key factor in sociologist Ulrich Beck's depiction of the Risk Society. For Beck, the Risk Society is characterized by uncertainty (in Knight's terminology) with the calculus of risk breaking down and insurers find it increasingly difficult to calculate probabilities and offer insurance against risks that are perceived as new, large scale, and uncertain. Environmental and technological hazards that have arisen (or that we have become aware of) since the middle of the twentieth century often have no reliable risk history. Well established weather related risks may have their risk histories overturned by climatic change. Nor do they seem to possess other characteristics that render them insurable, such as a clearly defined beginning and end, or a clearly limited set of victims. How, then, is the insurance industry supposed to offer cover to people against the hazards arising from climate change or cover for companies' environmental liabilities?

It's not just insurers, but all the traditional sources of risk expertise and management, most notably the state, that face difficulties in the Risk Society. Increased economic uncertainties since the 1970s combined with the apparent lack of capacity and/or willingness of the state to fully fund national insurance and welfare systems have led to a greater emphasis on the responsibility of the individuals to manage their lifetime of risks. The prudent, responsible citizen is one who takes out private insurance against unemployment, ill health, old age, and death. However, with the insurance industry also becoming more circumspect about the risks it takes on, this raises concerns about those deemed too risky or too poor (or both) to access these market based systems.

And of course, economic uncertainty affects not only individuals, but also corporate actors who have established various strategies to help manage their financial risks in a volatile global marketplace.

The rest of this article will further explore these three areas of research into financial risks and management, focusing on environmental, welfare, and corporate risk management.

Environmental Risks and Insurance

Insurance companies, although willing to take on the risks of others, are just as risk averse as any other company. However, by pooling many risks, predictability is increased and the insurers remove the risk to themselves of unexpected liabilities. There are, however, various conditions that risks have to be able to fulfill in order to make their insurance commercially viable. Any financial losses covered must be clearly delimited over time, in space, and in value. They should be caused accidentally. The financial losses covered should accrue to the insured party. There should be a large number of similar, but crucially independent, insured units. And they should not have catastrophic potential for the insurer. At the same time the price of insurance must be low enough to attract customers.

The changing physical – and economic and legal – nature of environmental risks has implications for their management through both first and third party insurance sectors. In the first party sector, property insurers will be affected if new, large scale risks result in unpredictable changes to the frequency and magnitude of damage caused. Climate change could render well established historical records of floods and hurricanes useless for calculating future insurance premiums. Conversely, if an increasing frequency of extreme weather events is experienced, some areas could become so prone to particular events that insurance cover becomes economically unfeasible.

There are some clear research opportunities for physical geographers here, particularly those carrying out research that purports to predict future climate and weather patterns. Reinsurance companies, who by insuring the insurers bear the brunt of extreme events, have invested substantial sums in climate related research. Such research also raises ethical issues where the results are used to enhance risk assessment and, potentially, exclusion from insurance cover; here there is a role for human geographers in examining the social consequences of such research by their physical geography colleagues. Nevertheless, increasing climate related claims are hardly just (or even mainly) due to physical environmental change. That virtually all of the 50 costliest insurance events over the last 30 years occurred in more developed countries is not a reflection of the relative severity of physical processes. Rather, processes that encourage people and firms to locate in risk prone areas, that cause insured values to rise, and, crucially, that further the interconnectedness of economic activity (the 'loss' of a factory being destroyed is hardly limited to the cost of the factory itself, nor even to any one company's business), have arguably had the greater impact on insurers' fortunes in recent times.

Even with regard to physical processes, economic geographers also have a role to play in examining more innovative ways in which companies are dealing with both extreme and mundane weather events. The futures market has been long used to hedge risks of commodity price fluctuations, but can also be enrolled to hedge the risks of fluctuations in events. By purchasing catastrophe derivatives, linked to a quarterly index of catastrophe in a particular region, insurers can smooth out their exposure. If catastrophes are high, the price of the future increases, allowing insurers to make money with which to offset their higher payouts. Conversely, if catastrophes are low, the insurers lose money on the future, but their payouts are also much lower. Overall, the risks are smoothed out over time. Human geographers have also investigated the use of weather derivatives, which are used to displace the risks of the more everyday weather that insurance companies will not usually cover but which might cause significant losses for a company running outdoor attractions, for instance. While the buyers may be hedging their risk, the sellers may be engaged in speculation, and the use of weather derivatives can be seen as an example of the commodification of nature. This raises important questions about how and why this commodification occurs, and what the consequences of it are, such as the privatization of knowledge about the weather.

Third party insurers are affected by the changing sociolegal construction of environmental liability risks that has increased the uncertainty and potential magnitude of claims. Two key research questions arise here. The first question concerns the continued viability of commercial insurance in compensating for environmental financial risks, given their changing scientific, social, and legal production. Environmental issues pose challenges for the most fundamentally held legal principles, that enable individuals to be held responsible for clearly defined events. Most famously, the US site clean up program known as Superfund, originating in 1980, established controversial principles of strictness, retroactivity, and joint and several liability. Strictness and retroactivity meant that insurers were liable for risks caused during the period of insurance, even if the insured company obeyed all the laws of the time and used the best available technology. In many cases risks built up unseen over time, only becoming apparent as sites were cleaned up many years after the policy year had ended (and often after the policyholder had gone out of business). In common with other environmental and health risks, insurers found themselves exposed to long tail financial liabilities. Joint and several liability meant that the authorities could sue any company that had worked on the site for the costs of cleanup, irrespective of their responsibility in causing the pollution in the first place. Insurers argued this led to a 'deep pocket syndrome' in which the companies with the richest insurers were sued. These principles, which risked undermining the very conditions of insurability, led to a contraction of the environmental liability insurance market in the late 1980s, and raised doubts over whether the insurance industry was up to coping with a new era of environmental concern and associated legislation (much as Beck had suggested).

Of course, environmentally risky production didn't cease in 1989. Many large companies found it viable (and cheaper) to self insure. Insurers themselves began to offer claims made policies, in which payouts were made if an environmental risk was discovered, rather than caused, during the policy year, so limiting their responsibility and redefining the temporality of environmental risks so that insurers' concern over their production was deflected into the future. More exclusions and lower limits made it harder to claim. A veneer of business as usual masked the limits faced by insurers in coping with new environmental risks.

More positively, insurers' concerns over their liabilities led to them becoming more circumspect about who they insured. The second key research question around liability insurance concerns the extent to which insurers can be enrolled in programs of environmental governance. Insofar as insurance can facilitate the externalization of environmental risks from the production process and so even undermine the polluter pays principle increasingly embodied in environmental liability legislation, this may seem unlikely. Once a polluting firm has purchased insurance, it can 'remove' the incentive to take care because it knows the insurer will pay up, creating moral hazard. On the other hand, precisely by translating an intangible environmental risk into a financial value, the insurer potentially has some sway over the environmental behavior of the firms it insures. Corporations with good pollution control programs and good risk records get lower premiums. Where possession of environmental liability insurance is a ticket to trade, insurers can act as gatekeepers to entire industrial sectors. However unfair insurers might have found legislation like Superfund, it means they now think twice about which companies they are prepared to cover. Managing financial risks becomes the quasi policing of firms with which you have a relationship. Understanding the advantages and disadvantages of what might be considered to be either the governmentalizing of private relationships or, conversely, the privatizing of government regulation, is an important research agenda.

The experience of the environmental insurance sectors suggests that the Risk Society poses considerable challenges for traditional forms of risk compensation and management, although physical environmental change is perhaps the least of these sectors' concerns. However, it also suggests that Beck has somewhat overstated his case, and that insurers can find new, profitable opportunities in environmental risks that help to sustain business as usual.

Personal Financial Risks and Exclusion

A refusal to offer flood insurance to households in flood prone areas is one way property insurers have reduced their exposure to environmental risk. Human geographers have long studied the impact of insurers' redlining in urban areas perceived to be at high risk from crime. More recently, the rolling back of the welfare state, while taking many forms in different countries, has often produced a greater reliance on the private insurance sector for the management of personal financial risks arising from major life events such as unemployment, ill health, old age, and death. Insurances that have historically mainly been provided privately, such as life insurance, have taken on a greater significance in those countries where private housing has expanded at the expense of social housing, and more people require it to protect their dependents from mortgage obligations. New insurances, such as critical illness cover, mortgage payment protection insurance, and forms of private unemployment insurance, have also been developed.

Human geographers have examined the contrasts between state and private insurance provision in attempting to understand the implications for people who are excluded from these private financial products. Insurance provision tends to work on the basis of one of two principles: solidarity or mutuality. Private insurance systems are almost always organized on the principle of mutuality. State insurance systems can be organized on either principle, but tend to maintain at least some element of solidarity. In a solidarity insurance system, contributions are mandatory, cover is paid for collectively out of taxation, and is provided through the monopoly of the state. Those who are better off pay most, thus covering claims from those who cannot afford to pay. Risks are calculated for the national population as a whole and premiums reflect individuals' share of this national risk, weighted according to individuals' ability to pay (their taxable income).

Mutual systems differ in two key respects. First, individuals choose whether to purchase the insurance or not. Second, the market is competitive, which affects the price of premiums. Fail to offer a good price to low risk individuals and they'll find a competitor who does, leaving you with the high risks and the big claims. For the insurance underwriter, then, competition doesn't just mean making premiums as 'low' as possible; it means differentiating them so that they are as 'accurate' as possible with respect to the risk posed by each applicant. Participants are grouped into pools of similar or mutual risks and are charged a commensurate premium. In this system, those deemed to be of high risk (and, of course, most in need of the cover) pay the most or may even find themselves excluded altogether.

Of course, the financial/social exclusion agenda can be argued to deflect attention away from more funda mental issues of poverty and structural inequality. Refusal of access to private insurance markets might be con sidered to be a middle class concern, but of little rele vance to those who can't contemplate the financial outlay required to purchase even standard rates, or, for that matter, get a home or mortgage or job to insure in the first place. However, lack of access to private insurance can be a *de facto* barrier to accessing housing markets (life insurance may be considered a prerequisite for taking out a mortgage by those with dependents), can exacerbate health concerns and, if the risk does become reality, can potentially plunge households into serious financial dif ficulty and even poverty.

In life and illness insurance, an applicant's riskiness relates to their health prospects and so the actual underwriting of risks, in which a premium is calculated, is preceded by the collection and assessment of health risk information. While health risk assessment by in surers is well established, this does not mean it is un controversial. The practice of health risk assessment raises questions about the ethical and appropriate use of medical information and the consequences for those people who cannot access insurance. Personal medical details, not only of the applicant but also of their families, are accessed as a condition of even applying for the product. Research by human geographers has found that currently 8% of life insurance applicants in the UK receive inflated premiums and 1.4% are excluded altogether, with considerable variations between insurers. The recent development of predictive genetic tests has also raised questions about a new class of asymptomatic 'healthy ill' and their possible exclusion from insurance markets.

The way in which risk knowledge is constructed by financial institutions has also undergone some profound changes with both positive and negative implications for financial inclusion. Banks (as well as insurers) place in creasing reliance on computerized systems of risk rating, and less on traditional systems of vetting applicants which depended more upon personal, tacit knowledge. Positively, such systems of producing and assessing knowledge at a distance reduce the risk that the personal prejudices of the assessor will influence the decision over whether or not to offer a loan or insurance. Negatively, credit scoring systems have set new definitions of riski ness based on statistically measurable and describable characteristics that can be assessed at a distance, which may further exclude those already disadvantaged.

Human geographers, concerned with the geography of knowledge, as well as social exclusion, have investigated the ethical and policy implications of these develop ments, armed with participatory research techniques that open up such debates to those without detailed scientific or commercial knowledge.

Corporate Risks and Management

Many financial risks faced by corporations are not in surable, but the ways in which these risks are managed often have profoundly geographical consequences. Fi nancial turbulence, particularly in the post Bretton Woods era, has long been of interest to economic geog raphers. Rapid fluctuations in the prices of currencies (and the relative prices of commodities and financial products) present clear financial risks to any firm that operates or trades internationally (and, in the longer term, also affects those that don't). Other fluctuations – cultural, psychological, and political – have implications for market conditions and patterns of demand.

Firms can attempt to manage the financial risk of these fluctuations in demand and supply costs through spatial, temporal, and structural changes (the latter two just as geographical in their implications as the former). The irony of all these strategies is that, while they ameliorate some risks they almost always create new ones.

Altering the spatial structure of the firm is one way to manage those risks (both before and after they have arisen). A decentralized spatial structure, particularly where different production facilities cater for different markets, can reduce the risks of fluctuations in supply and demand through diversification, and help stabilize rev enue across the firm as a whole. On the other hand, it can also lead to increased transaction costs and difficulties in monitoring the behavior of spatially dispersed managers. Relocation may also be an option for dealing with dif ficult market conditions that have already arisen, where the option exists to relocate to areas of potentially lower labor and regulatory costs and/or greater productivity.

The existence of sunk costs, however, means that spatial strategies are both costly and risky in themselves. Closing down an operation may mean writing off much investment that has been made in plant facilities, in the training and accumulated experience of labor, and in the wider local business relationships that have been estab lished. Beyond writing off existing sunk costs, actual expenditure may be required at the point of exit and even for a considerable time afterward (most significantly in terms of pension obligations), while relocation means beginning a new round of sunk costs elsewhere. And therein lies the inherent risk to any spatial fix to the risks of capital accumulation: new investment often requires significant long term commitment to make sunk costs worthwhile, yet such commitment is difficult in an un certain market environment. Relocation can seriously dent a company's profit line to the extent that exit via bankruptcy may be the only viable option. Sunk costs, then, are both inherent to, and potentially signifi cant constraints on, corporate spatial restructuring, and form the basis of significant counterarguments to the more extreme proponents of globalization. Further re search is required into their characteristics (which are by no means solely negative for firms) and their implications for both creating and managing financial risks faced by companies.

Structural changes have been associated with risk minimizing, so called flexible firms, not least because they reduce exposure to sunk costs. The vertical dis integration of the production chain refers to the out sourcing of parts of the production process. This may, of course, be done for cost minimization purposes and, in such cases, vertical disintegration is bound up with spa tial relocation as those parts of the production process deemed less skilled are sourced from lower cost lo cations. However, risk management is another key mo tivation behind vertical disintegration and here it need not be bound up with spatial decentralization (indeed, it can have quite the opposite effect). In a nutshell, out sourcing externalizes the bulk of the sunk costs involved in the activity from the buying firm. Notwithstanding the significant penalties that may lie in breaking contracts, the sunk costs invested in capital, labor, and pensions become the responsibility of the outsourced firm. Once again, however, new financial risks arise for firms on both sides of the relationship as they become ever more dependent on linkages and transactions with other firms. For instance, the subcontracted operator might not share the same commitment to (an image of) social and en vironmental responsibility as those firms with famous brand names to protect. In many vertically disintegrated sectors, not least the financial industry, trust and the acquisition of tacit knowledge become ever more im portant in making risky transactions, and the associated need for face to face meetings and networking is one (of many) reason why vertically disintegrated sectors like financial services still tend to be agglomerated in par ticular locations.

Firms and investors can also manage risk temporally through the purchasing of financial derivatives (including the weather and catastrophe derivatives mentioned above), such as futures, options, and swaps, allowing them to reduce their exposure to volatile financial values. As with other strategies of financial risk management, the use of derivatives brings with it its own risks. While de veloped as a means of hedging risks, derivatives are also used speculatively, as a means of taking risks in order to (hopefully) make profits, and it is this that has been re sponsible for the huge growth in trading of derivatives, and the range of financial products on offer, over the last three decades. Yet this speculation, and the use of in creasingly risky products, means that the volatility in the system which necessitates the use of derivatives in the first place is increased, not lessened. As well as the sys temic risk, any firm that trades heavily in derivatives, can face major risks itself, as exemplified by the collapse of Barings Bank in 1995. It's not just firms that are at risk. Other institutions, including governments, can trade de rivatives, most famously Orange County in California, which lost $1.6 billion in the derivatives market in 1994, with major implications for the funding of public facil ities. The real world implications of these virtual prod ucts, in terms of their impact on trade, investment, financial integration, and people's welfare, is an import ant question for economic geographers, as are the chal lenges of regulating such risky financial products by countries which are locked in a competitive struggle for the business of the financial sector.

Conclusions

The issue of how to manage financial risks, and the im plications of this management for personal, environ mental, and corporate welfare, is relevant to all areas of economic geography, and many areas of human geography beyond. In this article three areas have been outlined in which concern with financial risks and management are a central part of research agendas, but it is important never to assume that profit maximization is the only motivation behind firms' behavior, a crucial insight with implications for all research into individual and corporate economic decision making.

See also: Agglomeration; Climate Change; Corporate Responsibilities; Environmental Hazards; Financial Centers, International; Financial Exclusion; Fordism, Post-Fordism and Flexible Specialization; Redlining.

Further Reading

Beck, U. (1992). *Risk Society: Towards a New Modernity.* London: Sage.

Bennett, P. G. (2000). Environmental governance and private actors: Enrolling insurers in international maritime regulation. *Political Geography* 19, 875 899.

Bennett, P. G. and Smith, S. J. (2007). Genetics, insurance and participation: How a Citizens' Jury reached its verdict. *Social Science and Medicine* 64, 2487 2498.

Brown, R. L. (1991). Environmental liability insurance: An economic incentive for responsible corporate action. *Alternatives* 18, 18 25.

Clark, G. L. and Wrigley, N. (1995). Sunk costs: A framework for economic geography. *Transactions of the Institute of British Geographers NS* 20, 204 223.

Ericson, R., Barry, D. and Doyle, A. (2000). The moral hazards of neo liberalism: Lessons from the private insurance industry. *Economy and Society* 29, 532 558.

Knight, F. (1921). *Risk, Uncertainty and Profit*. Boston, MA: Hart, Schaffner & Marx.

Leyshon, A. and Thrift, N. (1997). *Money/Space: Geographies of Monetary Transformation*. London: Routledge.

Leyshon, A. and Thrift, N. (1999). Lists come alive: Electronic systems of knowledge and the rise of credit scoring in retail banking. *Economy and Society* 28, 434 466.

Pollard, J. S., Oldfield, J., Randalls, S. and Thornes, J. E. (2007). Firm finances, weather derivatives and geography. *Geoforum* 38 (in press).

Swyngedouw, E. (1997). Producing futures: Global finance as a geographical project. In Daniels, P. W. & Lever, W. F. (eds.) *The Global Economy in Transition*, pp 135 163. Harlow: Longman.

Tickell, A. (2000). Dangerous derivatives: Controlling and creating risks in international money. *Geoforum* 31, 87 99.

Relevant Websites

http://www.swissre.com
Sigma Research Database, Swiss Re.

Firms

M. Taylor, University of Birmingham, Birmingham, UK

Introduction

The firm in economic geography is fundamental to understanding how local economies and communities function and evolve in a globalizing economic environment. In their infinite variety, they are at the very heart of modern day life: as transnational corporations (TNCs), corporations, branch plants, and subsidiaries; as joint ventures, subcontractors, and franchisees; as small firms and family firms; as partnerships and sole proprietorships; as manufacturers, service providers, and retailers. They are pivotal actors and sites of collective agency at the core of the economic geography project that seeks to understand how they shape economic and social growth, development, decline, and change in local and regional economies. They are the sites of enterprise and innovation, of decision and indecision, and of growth, obsolescence, and decay. They are the creators and destroyers of jobs, products, and services; the creators and manipulators of markets; and, simultaneously, the allies and the enemies of governments. At the same time, they are interlocked one with another in a myriad of ways as they compete, control, and collaborate to create wealth for their owners and managers. Simultaneously, there are limits to managerial discretion and degrees of autonomy within firms that reflect inter and intra organizational power relations, asset bases, and competencies. In addition, it is these layers of complexity that make firms' boundaries increasingly fluid and hard to define.

From the perspective of economic geography, firms can be regarded both as creators of and created by the societies within which they operate. Within those societies, they are the loci where social and economic processes meet, interact, and are played out against the backdrop of the drive for personal wealth creation that underpins capitalist societies. Against this background of the dynamic complexity and plurality of firms, there is a constant academic search for the processes that create and maintain both enterprise and enterprises in capitalist societies: a search for a theory of the firm.

Theorizing the Firm

The search for that body of theory is now more important than ever in the wake of rapid internationalization and economic globalization, the constantly widening and deepening reach of TNCs, and the rapid industrialization and growth of the Indian and Chinese economies. There is a new urgency to understanding what firms are and how they interact with places and people to shape, mold, transform, and change the way economies work. In this changing environment, it is equally important to better understand local processes of enterprise and entrepreneurship in both corporations and small firms as these are seen as central to building and maintaining the competitive advantage of places. These processes are now regarded as being increasingly dependent upon the local embedding of firms in networks of trust within places: communities, localities, and regions. This has thrown into prominence the need to understand more fully the interconnection and idiosyncrasies of firms, along with their competencies, capacities, and capabilities, and their weaknesses, limitations, and vulnerabilities.

Attempts to understand and theorise the firm have developed and evolved across a range of disciplines, including economics, economic sociology, and management science, to understand the processes of enterprise and collective agency at the core of capitalist systems of production. These separate but complementary perspectives on the nature of the firm have added layers of new thinking by exploring 'rationalist' perspectives, focused on efficiency, and 'socioeconomic' perspectives, concerned with the social construction of economies. They explore processes of maximizing, transaction costs, uncertainty and risk taking, social relations, networking, embeddedness, institutionalization, governance, and learning. Each contributes a distinctive layer of meaning to a progressively deeper stylizing of the firm.

Economic geography has borrowed and built on these theories to develop a fuller understanding of the role of place, space, and locality in the development and functioning of firms – to develop an appreciation of the symbiotic relationship between firms and places, with each being, at least in part, constitutive of the other. This focus of place, space, and locality is a key difference between geography and the other social sciences in the way the firm is theorized. Economic geography has emphasized location decision making, the spatiality of transaction patterns, networking, and the local impacts of culture and society as they shape commercial organizations. Within these parameters, firms in economic geography are interpreted in very different ways. At one extreme, they are seen as no more than 'capital', a homogeneous group in constant conflict with 'labor'. Most commonly, they are treated as a phenotype, the smallest unit of analysis that does not need to be unpacked. Increasingly, they are seen as network members defined primarily by their relational ties, with porous and open boundaries that make them hard to define and delimit as they merge and meld one with another. Equally, they are seen as some form of collective organism engaged in a form of biological evolutionary struggle into which they are born, and within which they age, mature, and die. Alternatively, at the other extreme, they are interpreted as being engaged in a performative discourse or dance, a point and counterpoint of interpersonal relations and relational ties as key individuals and decision makers jostle for power, control, and ascendancy.

From Contracts to Behavioralism

The foundations on which conceptualizations of the firm have developed in economic geography can be identified in the neoclassical, rationalist ideas of the complete and incomplete contracts theories of the firm from economics, and their extension into the less economistic, behavioral approaches in management science. The 'complete contracts' approach assumes that, within a capitalist economy, all economically significant contingencies can be foreseen and incorporated into enforceable contracts. The 'incomplete contracts' approach, most frequently referred to as the transaction costs model, assumes, more realistically, that there are costs associated with using the market mechanism and that it is either impossible or too costly to specify all the unforeseen contingencies in a transaction. As a result, contracts between firms are necessarily open ended or incomplete. Both approaches also assume that firms strive for efficiency and competitiveness by maximizing profits and optimizing productivity. The transaction cost approach attempts to leaven the rationalist, maximizing assumptions of the complete contracts model by introducing the concept of bounded rationality, that firms do not have all the information they need on which to make decisions. However, issues of efficiency and enforceable contracts are always important and central in shaping firm behavior, though in more recent conceptualizations of the firm in economic geography that focus on the social construction of economies, they have too readily been relegated and neglected to the point of being rejected.

The first attempt formally to conceptualize the firm in economic geography was built on Herbert Simon's ideas on 'satisficing' by Alan Pred in *Behavior and Location* published in 1967. The aim, as in a work 30 years later, was to understand how knowledge and behavior affected location decision making. In this groundbreaking work, the firm was seen as an information processor conditioned by the information available to the entrepreneur or the enterprise to inform decisions on the commitment of scarce resources to generate profits, and their ability to use it. Benchmarked against neoclassical maximizing principles and, coupled with Smith's elaboration of Rawston's concept of spatial margin's to profitability, an interpretation of spatial patterns of economic activity was offered that went far beyond the untheorized empirical description that characterized the economic geography of the 1960s.

This was, in effect, a behavioral, knowledge based theory of the firm that was well ahead of its time in recognizing the spatial consequences of significant variations in the knowledge and ability between decision makers. It was, however, only a beginning. It was extended through the 1970s and into the 1980s by a range of complementary empirical strands of research on three main elements of knowledge acquisition and decision making. These elements were: (1) the linkages and transaction structures of firms; (2) the information flows (including face to face meetings) of individuals and management teams; and (3) the changing corporate forms as TNCs emerged and evolved, bringing the need to better understand the spatiality of intra organizational as well as inter organizational decision making.

This stream of behavioral research in economic geography was most fully summarized by McDermott and Taylor in 1982 in *Industrial Organisation and Location.* Some of the major elements of this stream of thinking were that:

- firms learn through their transactions (linkages) and information ties (flows) with other organizations;
- the spatiality of these ties varies between types of firm, with the smallest and technically least sophisticated having the most localized transaction structures and information flows – that is, they depend on the proximity and local linkage that is now labeled as 'clustering';
- there are size thresholds in the development processes of firms that also match with the funding thresholds of the 'venture capital gap', the 'equity gap', and the 'international finance gap' as they appeared 25 years ago;
- the external commercial environment impacts differently on different types of enterprise, be they corporate headquarters, branch plants, innovative small firms, subcontractors, family firms or commercially unsophisticated 'refuge' firms;
- different types of firms, especially branch plants and innovative small firms, bring different structures and dynamics of employment to the places where they locate or where they are created; and
- firms that survive and grow disengage from local agglomerations of production – that is, 'market makers' rather than the locally linked achieve commercial success.

This is, of course, a brief caricature of a now largely ig nored genre of behavioral economic geography research which finds perhaps its fullest expression in the work of Taylor and Thrift in 1983 and the enterprise segmentation model of the firm that they began to elaborate. The model drew on structural contingency theory and recognized the different operating environments faced by the different segments of corporate sector enterprises and smaller firms, which, in turn, impacted on their use of funds, skills, technology, inputs, production, outputs, and their con tractual relations with counterpart organizations. In the corporate sector, 'leaders' developed new products, assets, and services to secure future revenue streams, 'inter mediates' were the main generators of current incomes, while 'laggards' extracted residual income from assets approaching obsolescence. Corporate sector 'supports', meanwhile, provided management and technical supports, controlled finance, information, and the appointment of key personnel: all the transfer pricing mechanisms to maximize revenues, minimize the global tax take, and maximize executive salaries. In the smaller firm segments, 'leaders' were again the innovators, though frequently being no more than externalized research and develop ment (R&D) for the corporate sector. The 'loyal oppos ition' occupied segments of the economy too small to excite corporate sector interest, while the subcontractors and franchisees were the 'captured capitalists' comprising the 'satellite' segments subordinated to corporate sector interests. The 'satisfied', in contrast, were smaller, often family owned firms, recognized in the management lit erature as a significant proportion of the small and me dium enterprise (SME) populations whose principal goal was to remain independent and to achieve satisfactory profits. Finally, those small firms with technical skills but limited commercial acumen were labeled 'craftsmen'.

Unfortunately, this essentially agency based behavioral approach to the firm was, in the words of Nigel Thrift, 'mugged' by Marxism. That mugging saw the firm reduced to 'capital' (in opposition to 'labor') and declared an ir relevant analytical category by Walker . This was an ex treme position that was softened in work on the spatial divisions of labor and, more generally, in thinking within geography on localities.

Developing Socioeconomic Perspectives on the Firm

An alternative, more strongly socialized approach to the theory of the firm developed in parallel with what some would say were the overly economistic ideas of the complete and incomplete contracts theories in eco nomics, and it is these ideas from the business, manage ment science, and strategic management literatures that have informed more recent, innovative thinking on the firm in economic geography.

The inspiration for this socioeconomic thinking is clearly evident in Polanyi's ideas on reciprocal exchange and the interpretation and extension of these ideas by Granovetter. It is found in Simon's ideas on 'satisficing' and Veblen's ideas on the entrepreneur as being first and foremost a 'moneymaker'. It draws on behavioral ideas; evolutionary ideas, and, in particular, resource based ideas initiated by Penrose in 1959 but reinterpreted and extended by later researchers.

The behavioral strand of thinking, popularized by Cyert and March, focuses on decision making in firms. From this perspective, the firm is a site of decision making in the face of conflict, uncertainty, problem stimulated search, learning, and adaptation. It is tied into a framework of relationships, involving, among others, managers, workers, stockholders, suppliers, customers, lawyers, tax collectors, and regulatory agencies. Most have little interest in the specific objectives of the firm as long as they receive satisfactory side payments (wages, dividends, satisfactory products, and good treatment of creditors). However, conflict arises because, depending on their areas of responsibility, different managers pri oritize different goals and judge the firm's success by those goals. When goals are not met, search behavior is instigated, initiating learning and adaptation. What this approach provides is a set of linked mechanisms to ex plain how decisions are made within firms when futures are vague and opaque, how resources are deployed, and how knowledge is created and used.

The evolutionary strand of thinking is drawn from evolutionary economics as outlined in Nelson and Winter's seminal work on the firm and economic growth. These authors started from the position that economic change involves firm specific, endogenously generated competencies and resources combined with exogenous pressures for change generated in the environments within which firms operate. At the core of the firm are its 'routines' – the way it does things that can be as in appropriate as they can be appropriate. External pressure necessitates a constant 'search' for new routines and technologies so that the firm can cope. The 'selection environment' comprised everything that affects a firm's well being: conditions outside the firm in a particular industry or sector and the behavior of other firms. This was the environment that was not only the arena within which the firm sought opportunities, but it was also the arena within which the firm itself was selected.

Resource based views (RBV) of the firm began with Penrose in 1959, who interpreted the firm as a set of resources mobilized to cope with the pressures of the external commercial environment. The firm is there to create and mobilize tangible and intangible assets to create value and fuel growth. Learning in management teams is seen as central to this process as it combines with the organizational slack of excess resources and human capital, constrained by managerial capacities. It is, in essence, a supply driven approach to the firm centered on the strengths and capacities of individual enterprises.

The RBV is an interpretation of the firm that has been extended in later elaborations to incorporate issues of inimitability and the isolation of resources in order that firms can maximize returns to their owners. In these later interpretations, resources are viewed as isolating mech anisms used to generate rents and prevent imitation; with ambiguity, property rights, and the high costs of learning being used to prevent imitation by competitors. In part of this literature, managerial competence is regarded as a resource – the resource of imagination built on an initial stock of knowledge. Importantly, this line of thinking leads directly into the knowledge based view of the firm which interprets firms as social communities designed to create and transform knowledge into moneymaking products and services. In this capacity, firms are thought to work better than markets, especially when internal structures are flat rather than hierarchical, because this is thought to promote learning. Weeks and Galunik have further extended the RBV to develop a cultural view of the firm. By this interpretation, firms are cultures created by people working within the limits of bounded ration ality to estimate rent streams from products and services. As such, the culture the firm wants to create guides the optimal design of the firm.

These streams of thinking have shifted the ground rules for thinking on the theory of the firm in economic geography in three ways: (1) they have relaxed the as sumption of rationality, (2) they have seen decision makers as satisficers working in the face of uncertainty with imperfect knowledge, and (3) they have viewed learning and knowledge as the keys to commercial suc cess or failure. These shifts in thinking have been drawn on extensively in economic geography's treatment of the firm.

Socioeconomic Perspectives on the Firm in Economic Geography

Socioeconomic perspectives on the dynamics of local economies have created new conceptualizations of the firm in economic geography that draw on and extend those perspectives emphasizing a geographical cum spa tial dimension. Here it is possible to identify six different but complementary approaches: (1) an institutionalist perspective; (2) an embedded networks perspective; (3) a resource based perspective (including the concept of the 'learning firm'); (4) a relational perspective; (5) a dis cursive perspective; and (6) a temporary coalitions perspective.

The 'institutionalist' perspective interprets the firm as a refuge from market pressures and an institutional framework that supersedes issues of cost. It is an organ ization within which rules and routines can be used to fashion strategies to counter competitive uncertainty. Skills and information underpin traditions and habits within firms which gives those enterprises permanence beyond market speculation. Those habits and skills act like genes to reproduce an organization that will evolve and change, much as is envisaged in the evolutionary economics that is currently gaining popularity in eco nomic geography.

The 'embedded enterprise' perspective envisages the firm as enmeshed in socially constructed networks of reciprocity and interdependence involving untraded inter dependencies and unequal power relations. Embeddedness itself comprises distinct elements: cognitive embeddedness as shared ways of thinking; cultural embeddedness as common cultural values; social embeddedness created by immersion in shared social relationships; political embeddedness as a common framework of governance; and the structural embeddedness of buyer–supplier relation ships that build to form networks that channel market exchange. This is a perspective on the firm that has influ enced views on, among other things, 'industrial districts', 'innovative milieu', 'power networks', and 'learning regions', and resonates strongly with Porter's views on clusters.

The RBV of the firm emphasizes competencies and 'learning'. Firms are seen as bundles of activity specific resources combined into firm specific competencies. Firms learn through the application of their resources and become processors and generators of knowledge. Processes of 'ubiquification' in a globalizing world have been argued to have made knowledge, especially tacit knowledge, the last remaining localized resource, trans ferred through face to face contact. Hence, the RBV of the firm in economic geography has emphasized the importance of the 'learning firm' in the 'learning region'. However, this interpretation has not gone unchallenged, with Amin and Cohendet maintaining that the place premium attached to externalities, trust, and social cap ital, for example, have been overplayed. Instead, it is argued, it is relational not geographical proximity that is important for the functioning of the firm.

The RBVs and knowledge based views of the firm that find their origins in the strategic management literature underpin a set of interrelated and interwoven ideas on the firm in economic geography that have been labeled variously as 'discursive' and 'relational'. Here the firm is seen as bringing together diverse social relations, in cluding social, family, and interpersonal ties.

The relational perspective regards the firm as a con stellation of network relationships controlled by social actors with activities spread across space but embedded in particular places. In line with the literature on global networks by Dicken *et al.* (2001) these networks are the "foundational unit of analysis for … understanding … the global economy, rather than individuals, firms or nation states (p. 89)." The relational power and the power geometries of those networks together with their terri torial organization mold the economic geographic out comes of these network relationships. The actors themselves draw on and bring into the firm the diverse social relations within which they are themselves em bedded. Those relationships, in turn, define the bound aries of the firm which are recognized as being extensive, variable, and fuzzy. According to Yeung, adopting this perspective involves shifting attention away from the capitalist logic of the firm to the relations among firm specific actors. As such, the conceptualization of the firm needs to be more open and fluid, and the economic conceptualization of the firm needs to be extended to embrace place based politics, cultural practices, and so cial interactions. As it currently stands, this relational perspective from economic geography is no more than a heuristic, a description of collections of relationships that by Yeung's admission is fragmented and fuzzy. Critics of this relational perspective argue that it offers no pro cesses through which combinations of people, assets, competencies, and knowledge arise, prosper, decline, fail, or evolve and change, just that power and relationships exist. With power untheorized, it does not engage with competitive or collusive strategies, and it has been said that it is no more than process less description that is neither in time nor of time.

The 'discursive' view of the firm in economic geog raphy draws on actor network theory and emphasizes the importance of representational analysis in building the knowledge of the firm. In this project, it also mirrors aspects of the cultural evolutionary perspective on the firm that has been developed in strategic management. The firm in this view is more than a limitless domain of networks and flows, and the meaning of the firm is de rived from the common understandings among owners and managers that are needed to secure legal rights, together with property and territorial jurisdiction, to balance self interest and cooperation, to appropriate surpluses generated by the firm, and also to legitimize the right to exploit labor. This is the firm as discourse – a discourse of managerialism centered on information, knowledge, and talk. What the firm is appears and fades according to the narrative that holds sway at any time; a narrative of production and investment, or narratives of organization, finance, or legalism each with their own logics, vocabularies, and calculus. In essence, the firm by this interpretation is the medium that becomes the message that becomes the reality.

Finally, the temporary coalitions perspective on the firm goes beyond the legally defined 'shell' of the firm to disentangle processes of people being enterprising from the narrower operations of enterprises as legally defined firms. By this view, firms are temporary coalitions of people who identify commercial opportunities to create personal wealth, shoulder risk, apply assets to tasks, and trust in theirs and their partners' competencies to achieve their goals. Firms, irrespective of whether they are SMEs or corporations, are, therefore, time limited and vulner able, temporary crystallizations of commercial endeavor. Coalition members may be coopted, imposed, recom mended, tolerated, or be there by virtue of birth. They are there to achieve personal wealth, and trust is not a prerequisite for membership. They are held together by currently poorly understood processes of collective agency through which they combine their skills to achieve a common purpose. They wield their combined power internally to achieve a cost effective product or service that has market value, and they wield it externally to secure resources and access markets through whatever mechanism is expedient, be it competition or collusion.

Conclusion

The conceptualization of the firm outlined here dem onstrates a progressive layering and deepening of think ing as efforts have been made to engage with issues of spatiality, change, and globalization. However, a signifi cant agenda for future research remains comprising, in the first instance, seven central elements. First, theory needs to deal better with time, going beyond the bulking up of micro processes and placing the firm centrally in time. Second, economic geography needs to understand better the proprietorial control of knowledge by firms as part of their inimitability strategies to retain and main tain competitive advantage and to secure future streams of profits. Third, theorizing in economic geography needs to more fully consider the nature of the boundaries of the firm. Current thinking sees those boundaries as fuzzy, permeable, and multiple, but, in so doing, makes the firm an awkward if not irrelevant analytical category, not withstanding the legal/contractual structures that em brace and surround it. Fourth, economic geography currently underplays the significance of entrepreneur ship processes; not only when firms are first created, thus missing the first crucial point in the development of a firm, but also at later stages in firm development as coalitions of owners and managers strive to cope with change. Fifth, a related issue is the need to more fully incorporate into economic geography's treatment of the firm the impact of demand and pressures of the market. How do firms cope with external market pressures and the power relationships they involve? Indeed, as a sixth issue, theory needs to consider more fully power rela tionships both between and within firms, especially in multi sector, multi site, and multinational enterprises.

Finally, economic geography's treatment of the firm needs to explore more fully the supply of funds to firms. As with economics and management science, economic geography is preoccupied with the use of funds by firms – their investment strategies – but not the way they acquire and garner the financial resources they employ. The commodification of money since the 1980s has relegated the supply of funds into firms to just one among a number of necessary inputs. The fiduciary responsibility of fi nancial institutions, once central to the finance literature, now goes largely unmentioned, as the current credit crisis shows only too clearly. This issue of the supply of funds impacts strongly on the first six items of the re search agenda on the firm outlined here. Finance for enterprise, it can be argued, is at the very heart of any understanding of the firm in an environment of economic globalization.

See also: Agglomeration; Competitiveness; Creativity; Embeddedness; Enterprise Discourse; Entrepreneurship; Globalization and Transnational Corporations; Industrial Location; Industrial Organization; Innovation; Knowledge Economy; Location Theory; Networks; Relational Economic Geography; Technological Change; Technology and Regional Development.

Further Reading

Amin, A. and Cohendet, P. (2000). Organisational learning and governance through embedded practices. *Journal of Management and Governance* 4, 93 116.

Cyert, R. M. and March, J. G. (1963). *A Behavioral Theory of the Firm*. Englewood Cliffs, NJ: Prentice Hall.

Maskell, P., Eskilinen, H., Hannibalsson, I., Malmberg, A. and Vatne, E. (1998). *Competitveness, Localized Learning and Regional Development Specialization and Prosperity in Small Open Economies*. London: Routledge.

McDermott, P. and Taylor, M. (1982). *Industrial Organisation and Location*. Cambridge: Cambridge University Press.

Nelson, R. and Winter, S. (1982). *An Evolutionary Theory of Economic Change*. Cambridge, MA: Harvard University Press.

Penrose, E. (1959). *The Theory of the Growth of the Firm*. New York: Wiley.

Pred, A. (1967). *Behaviour and Location. Part 1. Lund Studies in Geography, Series B*, No. 27.

Schoenberger, E. (1997). *The Cultural Crisis of the Firm*. Cambridge, MA: Blackwell.

Storper, M. (1997). *The Regional World. Territorial Development in a Global Economy*. New York: Guilford Press.

Taylor, M. and Asheim, B. (2001). The concept of the firm in economic geography. *Economic Geography* 77, 315 328.

Taylor, M. and Oinas, P. (eds.) (2006). *Understanding the Firm: Spatial and Organizational Dimensions*. Oxford: Oxford University Press.

Taylor, M. and Thrift, N. (1983). Business organisation, segmentation and location. *Regional Studies* 17, 445 465.

Walker, R. (1989). A requiem for corporate geography. *Geografiska Annaler B* 71, 43 68.

Yeung, H. W. C. (2005). The firm as social networks: An organisational perspective. *Growth and Change* 36, 307 328.

First Law of Geography

M. F. Goodchild, University of California, Santa Barbara, CA, USA

Glossary

Geostatistics The theory of regionalized variables, or spatial variables subject to spatial dependence.
Idiographic Science The scientific description of the unique properties of places and times.
Nomothetic Science Science that stresses the value of general principles that apply everywhere in space and time.
Positivism A philosophy of science that stresses the confirmation and refutation of hypotheses through experiment.
Semantic Interoperability Successful sharing of data as a result of terms having shared meaning.
Spatial Autocorrelation A measure of spatial dependence.
Spatial Dependence The tendency for nearby things to be more similar than distant things.
Spatial Heterogeneity The tendency for conditions to vary over a space.

Introduction

In science generally, it is widely accepted that the most valuable truths about the natural world are those that apply everywhere in space and time. There would be little value, for example, in a periodic table of the elements that was valid only in North America, or only on Tuesdays. Such knowledge is described as nomothetic, and philosophies of science such as positivism see it as the end product of a lengthy but well recognized process of inference and experimentation. By contrast, knowledge about the specific properties of places and times is de scribed as idiographic, and many of the terms used to characterize such knowledge are pejorative – consider, for example, descriptive, anecdotal, and journalistic.

The notion that there might be nomothetic truths within the domain of the discipline of geography has risen and fallen in popularity over the years. In the 1950s Hartshorne and Schaeffer championed the idiographic and nomothetic positions respectively, and Schaeffer in spired a generation of graduate students, many of them enrolled at the University of Washington, to direct their research toward nomothetic knowledge, in areas such as central place theory and theoretical geomorphology. Bunge's 1962 book *Theoretical Geography* (and its second edition of 1966) captured this movement most forcefully, arguing that geographic knowledge "must meet certain standards including clarity, simplicity, generality, and accuracy" (p. 2) and that the purpose of geographic re search was to discover "these patterns, these morpho logical laws…so that (our) planet, Earth, fills (our) consciousness with its symmetry and ordered beauty" (p. xvi).

While the nomothetic impulse flourished for more than a decade, it fell into decline in the 1970s with the waning of the quantitative revolution. But in the 1990s a degree of revival took place, and much of the earlier work became newly fashionable, motivated at least in part by the rise of geographic information systems (GIS). One of the early observers of this shift was Peter Taylor, who argued in a 1990 critique that GIS represented 'the positivists' revenge'.

GIS is a generic tool designed to represent a vast array of distinct types of geographic information. As such it must recognize the nature of geographic information, and exploit its significant features in order to achieve both generality and acceptable performance. Thus, the nature of geographic information is an important topic in GIS texts, and the growth of interest in GIS has prompted new research into principles, regularities, and laws that can be found through experimental investigation. Of these, by far the most important has turned out to be a principle suggested almost tangentially in a paper written in 1970 by Waldo Tobler, one of the leading members of the University of Washington school, on the topic of modeling the urban growth of Detroit. Tobler suggested somewhat tongue in cheek that the principle might be termed the first law of geography, but in recent years as its generality and utility has become recognized a con sensus has emerged that it does indeed justify that title.

The next section describes the law in detail, and discusses its generality and utility. The following section reviews recent commentaries and examines the case for additional laws.

Tobler's First Law of Geography

The Earth's surface is almost infinitely complex, and it would be impossible to characterize even a small part of it with perfect fidelity. Instead, geographers and others often indulge in spatial sampling, selecting a comparatively small number of locations at which to collect data, and assuming that the gaps between such observations can somehow be filled. If every location had characteristics that were completely independent of those of its nearby locations, this would clearly be impossible. But in reality characteristics tend to vary fairly slowly over the Earth's surface, such that the characteristics at one location tend to be similar to those at nearby locations. Of course, what exactly is meant by 'nearby' and 'similar' remains to be seen, and depends on the particular characteristics of interest. Weather, for example, tends to vary very little over distances less than 1 km, but tends to vary greatly over distances of 1000 km; soils, on the other hand, can vary substantially over distances as short as 10 m.

Tobler (1970) originally stated the first law (TFL) in the form "All things are related, but nearby things are more related than distant things". The important message is in the second part, and because 'related' carries un necessary connotations of causality TFL is better stated as "nearby things are more similar than distant things". TFL thus forms the basis of the process known as spatial interpolation, or the formal process by which the gaps produced by spatial sampling can be filled.

The word law is very powerful, suggesting the kind of universality that only a second law of thermodynamics can possess. Nevertheless, there are clearly exceptions even to the second law of thermodynamics, as Maxwell's Demon demonstrates. Newton's laws of motion were found to have exceptions at both ultrasmall and ultra large scales, but despite quantum mechanics and the special law of relativity it is still common to refer to them as laws. In short, as commonly used in science the term law does not necessarily imply perfect predictability.

In the social sciences the concept of law is much more controversial. Many would argue that law like behavior will never be found among human populations; and even that search for it is unethical. The experimental physicist Ernest Rutherford is once said to have remarked, perhaps after having been annoyed by some social scientist col leagues, that the only truth that could possibly be dis covered in the social sciences is "some do, and some don't".

Against this background it seems fair to label Tobler's statement as a law, even though it is far from determin istic, and even though its parameters vary from one phenomenon to another. There are clearly exceptions to the law, when phenomena reveal less similarity locally than at a distance. The checkerboard is often cited as an example, since every white square is surrounded by the opposite color, but note that in order for this to occur it is necessary for every square to have a uniform color; in other words, TFL applies at the within square scale, but not at the between square scale. On the economic landscape, the territories established by certain organisms similarly defy TFL, since the presence of one organism makes others less likely in the immediate vicinity. The cliffs of the Grand Canyon represent another exception to TFL, since the property elevation varies dramatically over short distances.

One way to gain insight into the significance of TFL is through a simple thought experiment. In a world in which TFL was absent nearby places would be as dif ferent as distant places. It would be necessary to step only a vanishingly small distance away from one's current location to encounter the full range of conditions on the Earth's surface, from the height of Mount Everest to the depths of the Marianas Trench, and from the summer temperatures of Death Valley to the winter temperatures of the South Pole. A world without TFL would be an impossible world from the perspective of human existence.

Several methods have been devised for characterizing TFL and the property it addresses, which is known generally as spatial dependence. The field of geostatistics, also known as the theory of regionalized variables, was devised by statisticians as a way of formalizing the principle known informally as TFL, and of optimizing the process of spatial interpolation that relies on it. The variogram (or more correctly semivariogram) is defined as a function describing the increase of variation in a phenomenon with increasing distance. In the absence of TFL, variation over short distances is as large as variation over long distances, and the variogram is flat. With phenomena that obey TFL, however, variation is gener ally observed to rise monotonically until it reaches a maximum value known as the sill, at a distance known as the phenomenon's range (**Figure 1**).

TFL can also be measured at fixed scales for patterns such as the checkerboard, or for data distributed over irregularly shaped regions such as states or counties. Two indices of spatial dependence are in common use: Moran's I and Geary's c. Both are described more specifically as metrics of spatial autocorrelation. Moran's I scales from positive values when TFL applies, to zero when neighboring values are as different as distant values, to negative when neighboring values are more different than distant values. The respective ranges for Geary's c are <1, 1, and >1. Despite the existence of two competing statistics, however, the terms used to describe spatial de pendence typically follow the Moran ranges, and patterns for which TFL is valid are commonly described as having positive spatial dependence. **Figure 2** shows such a data set and its associated Moran statistic.

Spatial dependence can be described, largely cor rectly, as the most important property of any spatial pattern. It can be useful in distinguishing the impacts of hypothesized processes, since the presence of strong, positive spatial dependence at a particular scale implies that the processes causing the phenomenon are similarly persistent at that scale. Smoothing processes (such as glaciation) and diffusion processes (such as migration) both result in patterns with strongly positive spatial de pendence, while sharpening processes such as economic competition result in the opposite. But this kind of

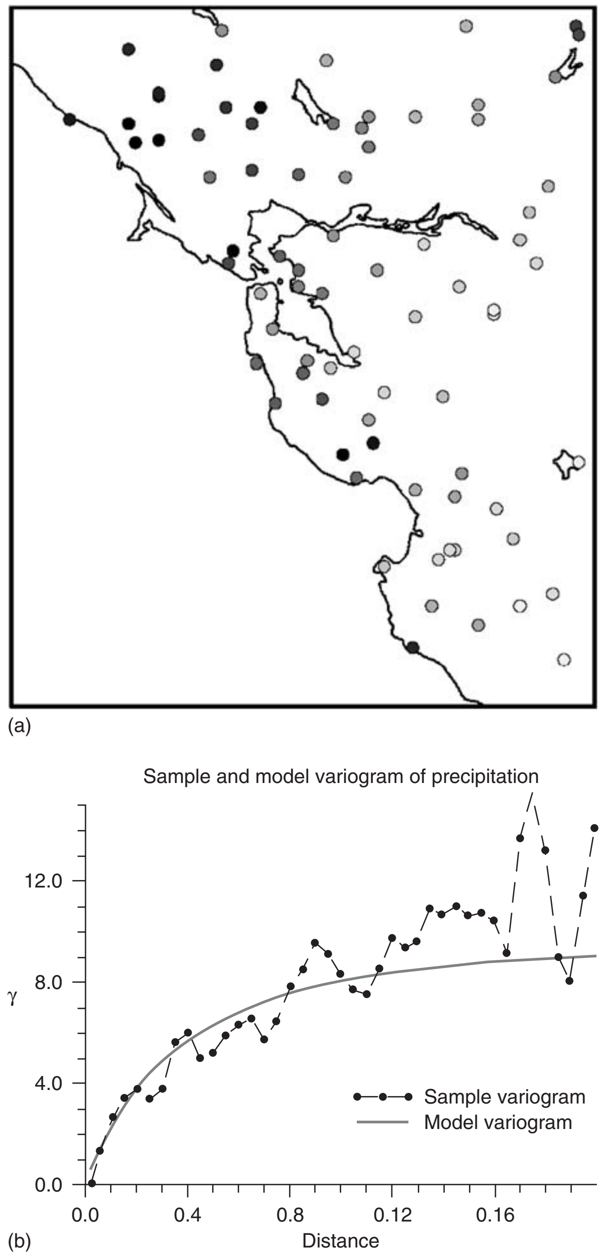

Figure 1 A typical variogram analysis of Tobler's first law. (a) The map shows precipitation amounts at points for California's Bay Area. (b) The graph plots variance against distance, showing the familiar monotonic rise.

inference is scarcely sufficient to explain the growth of interest in TFL in recent years, which relies instead on much more utilitarian arguments concerned with GIS design.

In a world without TFL every point's characteristics would bear no relationship to those of its neighbors, and in order to represent any geographic data set in a GIS it would be necessary to characterize every point. The re sult would be an impossibly large data set. TFL allows spatial patterns to be captured by sampling, since spatial

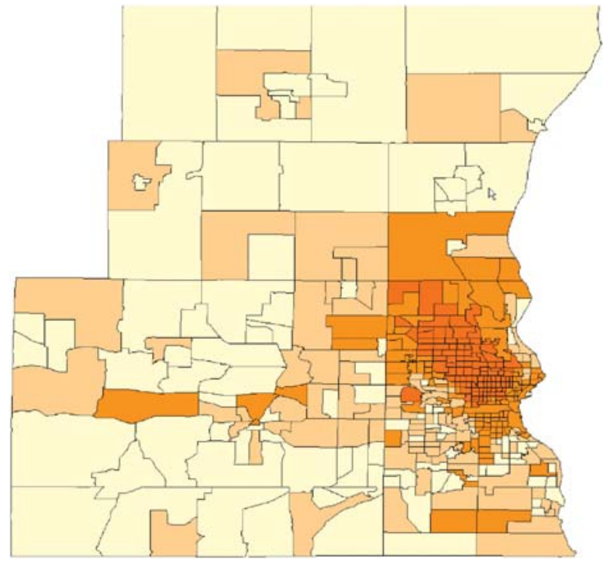

Figure 2 Percent black mapped for Milwaukee census tracts from the 1990 census. Moran's *I* statistic is 0.8971, indicating very strong, positive spatial dependence (and adherence to TFL).

interpolation can always be used to fill in the gaps. TFL is the basis on which all contour maps are made, and the basis on which each day's weather maps are compiled from point data. It allows large areas to be characterized as homogeneous, and represented in a GIS as polygons rather than as multitudes of points, achieving massive degrees of data compression. In other words, TFL is the basis of the armory of tricks with which GIS databases represent what is in principle an infinitely complex world. The specific details of any phenomenon's spatial behavior, as represented for example in its variogram, are the basis on which the density of sampling is determined, since points spaced closer together than the range of the phenomenon will yield observations that are to some degree redundant. It is also the basis for parameters of GIS representation such as pixel size, minimum mapping unit, and spatial resolution. No wonder, then, that the growth of GIS stimulated a renewed interest in TFL and its implications.

Commentary

In 2004 the *Annals of the Association of American Geographers* published a Forum on TFL, with five distinct per spectives, an introduction by the Forum's editor Daniel Sui, and a response by Waldo Tobler. Jonathan Phillips reinterpreted TFL in the context of physical geography, while Harvey Miller discussed the potential for greater formality in TFL as scientific understanding of social behavior advances. Among the skeptics, Trevor Barnes argued that TFL, as all science, must be understood as a social construction that is intimately related to the places of its origin and the life of its author, while Jonathan Smith examined TFL against three require ments – universality, necessity, and synthesis – and found it lacking in all respects. It is clear that TFL is a contentious idea within the discipline, particularly among cultural geographers, and that debates over its meaning will continue.

The present author's contribution to the Forum ad dressed two questions: does TFL apply to spaces other than the geographic, and are there other laws of geog raphy waiting to be discovered? On examination, it is clear that TFL applies to the spaces of other planets, but interesting insights can be gained by asking whether it applies to the space of the human brain, the human genome, or the digits of pi. Similar benefits to those obtained by exploiting TFL in the representation of geographic information, including compression and in terpolation, might apply to branches of informatics that address these other spaces.

Several candidates for additional laws have been suggested. Perhaps the strongest candidate is spatial heterogeneity, the observation that conditions on the Earth's surface vary, in what a statistician would describe as nonstationarity. If it is impossible to conceive of perfect explanation in the social sciences, and indeed in many of the environmental sciences as well, then it fol lows that the unexplained variation will display geo graphic patterns, and that models estimated for different parts of the geographic world would have different par ameters. This concept has been addressed recently through the development of a range of place based statistics, or statistics that measure the properties of places, with no expectation that the results will gener alize to other places. They include geographically weighted regression, a technique that allows the par ameters of a linear model to vary spatially, leading to interesting insights in many applications.

Spatial heterogeneity has an interesting implication for another aspect of GIS, the problem of interoperability. Two systems are said to be interoperable if data from one can be read and processed by the other, and clearly issues of data format are at the forefront of the problem. More difficult to address, however, are issues of semantics, or the variation in the meaning of terms from one part of the world to another. A number of projects, such as the European Union's INSPIRE, have attempted to overcome such variations through the establishment of common standards, and techniques for translation and cross walk ing between classifications. But the problem remains a serious impediment to the integration of geographic data, and to studies that span several jurisdictions.

Lack of semantic interoperability, through the adop tion of different classification systems and different meanings of terms in separate jurisdictions, can be readily understood as a necessary consequence of spatial heterogeneity. Left to their own devices, researchers in any limited area of the Earth's surface will adopt systems that work well for that area, and because of spatial het erogeneity are necessarily different from the systems that work well in other areas. Global standards will never work as well, and there will always be resistance locally to their adoption.

Conclusion

Seen from half a decade later, the discipline of geography of the 1950s appears to have been remarkably homo geneous. The debates of that period led to the divisive quantitative revolution, which was followed by an equally divisive retreat. Today, however, the discipline is thriving, at least in the US, as a polyglot of vastly different per spectives and methodologies. The reaction to TFL, as reflected in the *Annals* Forum of 2004, is only one indi cator of the degree to which the discipline has agreed to tolerate, or at least agreed to differ. Within the GIS community TFL is regarded as a cornerstone of repre sentation and a key to analytic insights, while large sec tors elsewhere in the discipline may raise fundamental objections to the very idea of laws in geography.

Any effective use of GIS requires an understanding of the nature of geographic information, and TFL stands as the most compelling summary of that nature. It demon strates once again that the discipline of geography is the logical home of GIS, because only with a deep under standing of the nature of the real geographic world can one hope to make sense of the bits and bytes of its digital representation.

See also: Geographically Weighted Regression; Kriging and Variogram Models; Spatial Autocorrelation; Spatial Interpolation.

Further Reading

Bunge, W. (1966). *Theoretical Geography* (2nd edn.). Lund, Sweden: Gleerup.

Cliff, A. D. and Ord, J. K. (1981). *Spatial Processes: Models and Applications*. London: Pion.

Fotheringham, A. S., Brunsdon, C. and Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Hoboken, NJ: Wiley.

Isaaks, E. H. and Srivastava, R. M. (1989). *Applied Geostatistics*. New York: Oxford University Press.

Longley, P. A., Goodchild, M. F., Maguire, D. J. and Rhind, D. W. (2005). *Geographic Information Systems and Science* (2nd edn.). New York: Wiley.

Sui, D. Z. (2004). Tobler's first law of geography: A big idea for a small world? *Annals of the Association of American Geographers* 94(2), 269 277.

Taylor, P. J. (1990). Editorial comment: GKS. *Political Geography Quarterly* 9(3), 211 212.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46, 234 240.

First World

K. D. Willis, Royal Holloway, University of London, Egham, UK

Glossary

Global Governance Forms of rule making and policymaking at a global level which involve a range of institutions, including national governments, nongovernmental organizations, and multilateral organizations, among others.

Gross Domestic Product (GDP) The value of all goods and services produced within a country regardless of who produced them.

Gross National Income (GNI) The value of all goods and services claimed by the residents of a particular country regardless of where those goods and services were produced. It includes income from abroad, such as remittances and profits repatriated by companies from overseas subsidiaries.

Introduction

'First World' is widely used as an umbrella term for the countries of Western Europe, as well as the United States of America, Canada, Japan, Australia, and New Zealand. The term's origin was political and dates from the Cold War period. Despite the collapse of the Communist bloc, the rise of the so called 'newly industrializing countries' (NICs) and the growing interconnectedness of many parts of the world due to processes of globalization, the idea of the First World continues to be used as a short hand for economically rich capitalist countries that have liberal democratic political systems.

This article discusses the changing meaning of the term First World and considers the usefulness of cat egorizing many countries and millions of people in this way. In particular, issues of economic and human de velopment diversity will be considered, both between the countries of the First World and within these countries. Changing global political power relations and new forms of global governance at the start of the twenty first century will also be considered, including the rise of China and India as potential global superpowers in both economic and political terms.

Origins of the First World

As with the terms 'Second World' and 'Third World', the concept of the First World only makes sense when related to other categories. Rather than being a self ascribed identity, the concept of the First World came to the fore at the meeting of the nonaligned countries at the Bandung Conference in Indonesia in 1955. This conference of largely newly independent African and Asian countries sought to lay out a development path which would follow neither the capitalist route of the First World, nor the communist strategies of the Second World. The threefold categorization was therefore not meant as a hierarchy, but rather a division between different forms of political system. Based on this definition, the First World of the 1950s consisted of Canada, the United States of America, the countries of Western Europe, Australia, and New Zealand (see **Figure 1**).

Despite its original use as a political term, First World soon became synonymous with countries with high levels of economic development, high levels of urbanization, low mortality rates, widespread access to basic services, and high levels of consumerism. In other words, based on Eurocentric ideas of 'development' the First World was 'developed' and First World status became a goal for many countries in the Global South. Even though Japan had been one of the countries at the Bandung Confer ence, its rapid post World War II economic development along noncommunist lines led to its inclusion in many classifications of the First World from the 1960s onward.

The End of History

The collapse of the Soviet bloc in the late 1980s/early 1990s marked the end of the Second World and what Francis Fukuyama termed the 'end of history', as the Cold War framing of world events had come to an end with the perceived victory of capitalism and liberal democracy. The collapse of the Second World should have, in theory, meant the end of the ideas of the First World and the Third World as they only made sense as part of a tripartite division of the world. However, the terms had become so embedded in common usage that they continue to be used, although other divisions, such as Global North/Global South and Developed/Developing countries, have become popular.

Most research which considers the regional groupings, focuses on defining or problematizing the Third World or Global South, rather than considering the First World. This reflects a common process within categorization whereby what is viewed as 'normal' or 'desirable' is not interrogated, despite the fact that it is as socially con structed as any other category. Eurocentric ideas of 'de velopment' and 'progress' have often meant that the First

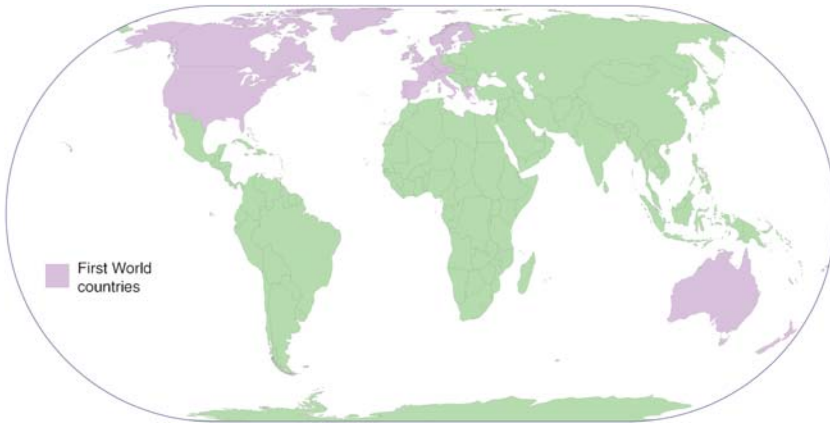

Figure 1 First World *c.* 1950. From map data © Maps in minutes™ (1996).

World is taken as the norm and other parts of the world are contrasted with it.

The countries which emerged from the Soviet bloc, both through the disintegration of the Soviet Union and the move toward capitalism of the Eastern European countries, are sometimes referred to as 'transitional' economies. However, within this category, there are countries such as the Central Asian republics of Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan which have levels of economic and human development similar to many countries in the Third World, while others such as Latvia, Lithuania, Estonia, and Russia are much more similar to the First World nations. Politically, ten of the ex communist states or republics are now members of the European Union (EU).

Economic Diversity

As with all attempts to classify countries, there are vast differences between countries within one category, not to mention major differences within countries. In 1987, for example, Gross Domestic Product per capita (GDP p.c.) in the First World ranged from US$5500 in Greece and US$5597 in Portugal to US$17 615 in the USA and US$16 375 in Canada.

High levels of national income per capita are still used as a form of classification, but this does not necessarily just encompass the countries which are usually referred to as First World. The World Bank classifies countries into four categories based on Gross National Income per capita (GNI p.c.) (see **Table 1**). In 2007, the high income country classification included the long standing First World countries, but also included other countries, most notably South Korea, Singapore, Israel, and a number of the oil rich states in the Middle East, such as Saudi Arabia and the United Arab Emirates (see **Figure 2**). Thus, the countries with First World levels of national economic wealth vary greatly in terms of the structure of their economies, and also their political systems. They are also involved in different political groupings and alliances. Coherence within this group of countries is therefore unlikely and demonstrates the problems with classification systems.

Table 1 World Bank country classification criteria 2007

Classification	*Gross National Income per capita (GNI p.c.)*
Low income	US$935 or less
Low middle income	US$936 US$3705
Upper middle income	US$3706 US$11 455
High income	US$11 456 or more

From World Bank (2007). World development report 2007, *Report of the World Bank*. Washington, DC: World Bank.

Within countries there can also be significant differences. The First World classification was, and continues to be, based on national level indicators, but in many cases wealth does not 'trickle down' equally to all members of the population. This may be related to spatial differences, for example, between rural and urban areas, but there may also be social differences, between men and women, between different classes, and also income distinctions based on ethnicity. The term 'Fourth World' has sometimes been used to refer to economically and socially marginalized indigenous groups in the First World. These included First Nations peoples in Canada and aboriginal populations in Australia. Income inequalities are often measured using the Gini Index which runs from 0 to 100 with 0 being absolute equality in expenditure or income and 100 being absolute inequality. Thus, the higher the index, the more unequal the wealth distribution within that country. Among the World Bank classified 'high income countries' in 2007, Gini Index figures ranged from 24.7 for Denmark and 24.9 for Japan to 42.5 for Singapore and 40.8 for the

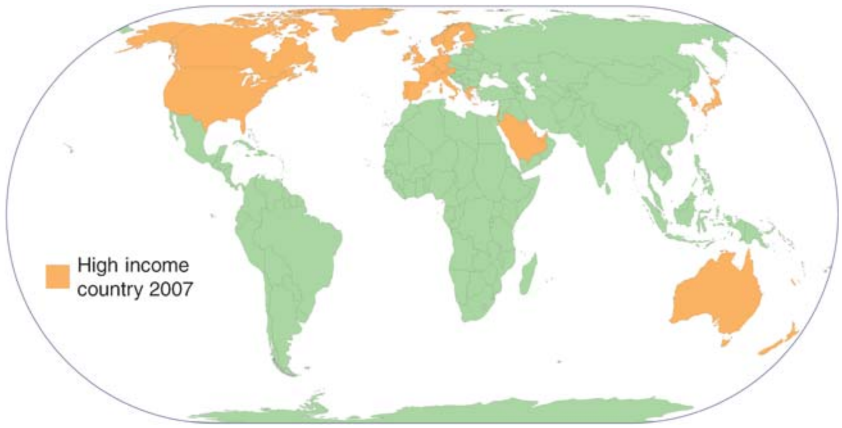

Figure 2 High income countries, July 2007. Based on data from World Bank (2007) and map data © Maps in minutes™ (1996).

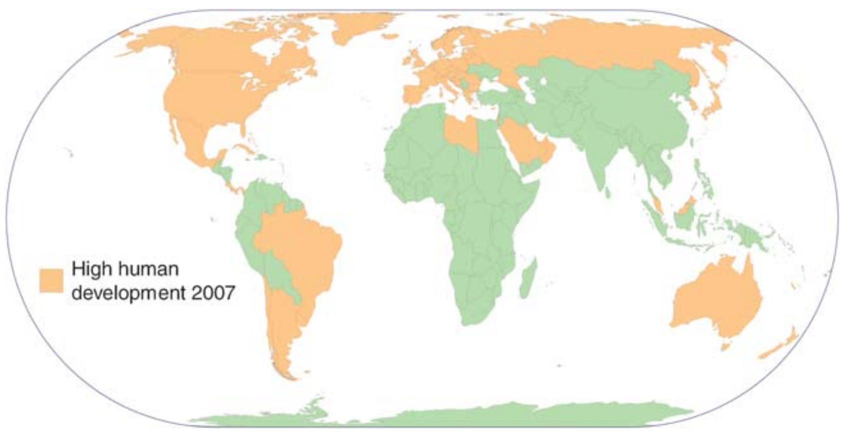

Figure 3 Countries with high HDI, 2007. Based on data from UNDP (2007) and map data © Maps in minutes™ (1996).

USA. Figures were not available for the oil rich states of the United Arab Emirates, Bahrain, Kuwait, and Saudi Arabia, among others.

Levels of Human Development

The use of First World as synonymous with 'developed' needs to recognize the diversity of development meas ures. In the late 1980s, the United Nation Development Program (UNDP) devised the 'Human Development Index' (HDI) as a measure of development which goes beyond simple economic measures such as GNI p.c. The concept of 'human development' encompasses standard of living (measured using GDP p.c.), education (meas ured using adult literacy rates and gross enrolment rates), and health (using life expectancy at birth). Using these measures, the UNDP calculates a composite index – the HDI – which runs from 0 to 1; the higher the figure, the higher the level of human development.

According to the UNDP, countries with a high HDI in 2007 were much more widespread than those with high incomes according to World Bank criteria (see **Figure 3**). This reflects an understanding of development which does not see education and health levels as highly cor related with income levels. Thus, while all countries with a high income had high HDI in 2007, there are many others which had achieved high levels of human devel opment at a national scale, without a GNI p.c. figure of over US$11 456. Many Eastern European 'transition' countries fall into this category, as do many Latin American countries. Malaysia was classified as a high human development country in 2007, as was Libya – the only African country to achieve this status.

Achieving high levels of human development can come through economic wealth, but the patterns of high HDI also reflect the way in which state policies, par ticularly in health and education, can create conditions for improved human development despite lower levels of national wealth. For example, Cuba achieved an HDI

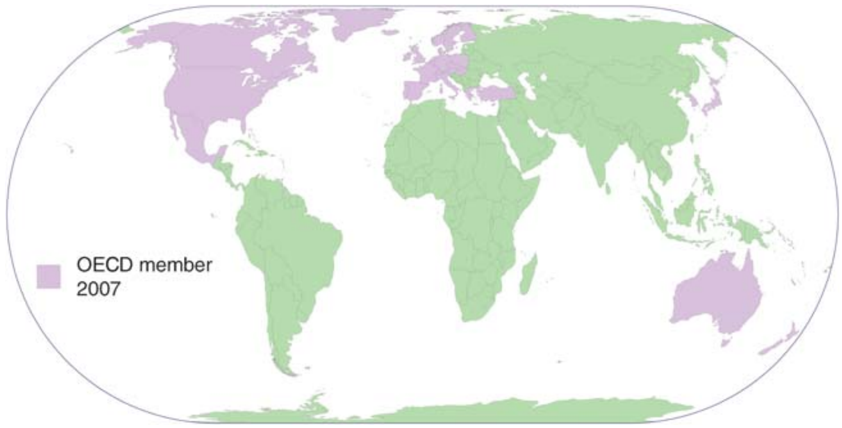

Figure 4 OECD members, 2007. Based on data from www.oecd.org and map data © Maps in minutes™ (1996).

figure of 0.838 with an estimated GDP p.c. figure of US$6000 and Libya's HDI of 0.818 was achieved with a GDP p.c. figure of US$10 335. These figures take into account purchasing power parity (PPP) to recognize different costs of living within different economies. These mismatches between levels of human development and economic wealth, demonstrate how problematic the concept of First World is, once the original political categorization has broken down.

Global Political and Economic Power

In the post Cold War period, new groupings have emerged on the global stage representing new forms of alliances. In some cases, these groupings have strong representation from earlier First World countries, but increasingly countries which were previously in the Second or Third Worlds are making their mark and challenging the idea of First World meaning economic ally and politically dominant.

The G8 is an organization with its roots in the 1970s Cold War world. Government representatives of the original G6 of advanced industrial democratic countries (France, the USA, the UK, West Germany, Japan, and Italy) met to discuss key economic and political chal lenges both domestically and internationally. By the late 1970s this group had become the G7 with the addition of Canada, and representatives of the European Com munity also attended meetings. Thus, most of the countries of the First World were part of this organiza tion. Notably, exceptions included Australia and New Zealand. Following the collapse of the USSR, Russia joined in 1998, so the group became the G8. In some sense, this group of countries represents a new First World in terms of economic and political power. But such a position should not be taken for granted.

The Organization for Economic Co Operation and Development (OECD) has a similar history. Its members have high levels of economic development and are broadly classified as capitalist democratic states. This again has overlaps with the original definition of the First World, but as of 2007, membership was more diverse, including Mexico, Turkey, South Korea, Hungary, Poland, the Czech Republic, and the Slovak Republic (see **Figure 4**). Countries apply to become a member and the OECD Council decides whether to extend mem bership and on what conditions. The organization aims to promote economic development in its member countries by sharing expertise.

Since World War II, global governance institutions such as the United Nations, World Bank, and World Trade Organization (WTO) have been set up as a way of promoting peace and economic stability. Of course, the agendas set by these organizations are not neutral and have often reflected the priorities and ideologies of the First World, particularly the United States and countries of Western Europe. However, the economic rise of some countries in the Global South, combined with the in creasing mobilization of alliance within the Global South, has meant that First World dominance in certain fora has been challenged.

This has been particularly noticeable in the WTO trade talks. Since it was established in 1995, the WTO has promoted reductions in trade barriers between its members and sought to uphold a rules based trading system. Countries which become members of the WTO have to abide by the rules and members can appeal to the WTO in situations where they believe other mem bers have broken the rules. Despite such a system, criticisms have been leveled at the WTO regarding the possible uneven application of the rules in favor of First World or Global North countries. In 2003, the WTO members met at Cancún, Mexico for trade talks. Prior to the meeting a group of Third World or Global South countries created an alliance called 'G21' as a negoti ating bloc against the US and EU agricultural subsidies

Table 2 Economic and population indicators

	GDP (US$ bn) 2005	*GDP p.c. (PPP) (US$) 2005*	*GDP p.c. annual growth rate (1990 2005) %*	*Population 2005 (millions)*
Brazil	796.1	8402	1.1	186.8
China	2234.3	6757	8.8	1313.0[a]
Germany	2794.9	29 461	1.4	82.7
India	805.7	3452	4.2	1134.4
Japan	4534.0	31 267	0.8	127.9
Russia	763.7	10 845	0.1[b]	144.0
USA	12 416.5	41 890	2.1	299.8

[a] Figure includes Taiwan.
[b] Figure refers to shorter period due to changing national boundaries following break up of the USSR.
From UNDP (United Nations Development Programme) (2007). Human development report 2007/8, *Report of the United Nations Development Programme.* Oxford: Oxford University Press. Tables 1, 5 and 14.

policies. G21 members included the increasingly economically important China and India as well as many Latin American countries, Pakistan Egypt and the Philippines, among others. Rather than agreeing to terms proposed by the Global North which had often happened in the past, the G21 countries continued to lobby for fairer policies on agricultural subsidies. No agreements could be made and the talks broke down. These new forms of alliances have continued to make their presence felt within the WTO and other institutions, meaning that assumed power of the First World countries is not always exercised in reality.

The economic rise of the so called BRIC countries (Brazil, Russia, India, and China) in the early twenty first century also demonstrates the problems with a bounded idea of an economically dominant First World based on categories from the 1950s (see **Table 2**). These countries contain a significant percentage of the world's population and are growing in importance in relation to both manufacturing production and natural resources. If economic development patterns continue to follow the same trends, then some, or all, of these countries will dominate the world economy in the middle of the twenty first century in a way that the previously designated 'First World, countries did in the middle of the twentieth century.

Summary

The First World was originally identified in relation to the Second and Third Worlds in the 1950s and had a political basis. Over time, the concept has taken on other meanings, most notably to do with levels of economic development and progress. This has meant that the concept continues to be used despite the disappearance of the Second World with the collapse of the communist bloc. It is now used to refer to countries with high levels of national income per capita, high levels of urbanization, and high standards of living. However, the number of countries which fall into this category is increasing depending on the criteria used and the countries have diverse forms of political system, as well as different economic structures. There are also disparities within the countries; the population is not all at First World status. Finally, assumptions that First World countries have the power to shape global economic and political processes are being challenged in some settings.

Classifying countries is always open to debate and there are concerns about diversity within national boundaries. The use of the term First World is indicative of this. The debates outlined in this article also stress the ways in which categories are based on assumptions which are often left unstated, such as 'development' being defined in relation to levels of national income or other economic measures. Since the end of the Second World, other terms have become increasingly common to refer to First World regions, such as Global North or Developed World. While they may be based on different criteria, they too suffer from the problems of diversity and the impossibility of classifying the world's people and places into a small number of categories.

See also: Cold War; Development I; East/West; Eurocentrism; North-South; Second World; Third World.

Further Reading

Antipode (1981). Special issue on the 'Fourth World', *Antipode* 13: Issue 1.
Auty, R. (1979). World within worlds. *Area* 11, 232 235.
Blaut, J. M. (1993). *The Colonizer's Model of the World*. New York: The Guilford Press.
Bonnett, A. (2004). *The Idea of the West: Culture, Politics and History.* Basingstoke: Palgrave Macmillan.
Eckl, J. and Weber, R. (2007). North South? Pitfalls of dividing the world by words. *Third World Quarterly* 28, 3 23.
Friedmann, J. (1992). The end of the Third World. *Third World Planning Review* 14, v vii.
Fukuyama, F. (2006). *The End of History and the Last Man*. New York: Free Press (originally published 1992).
Langhammer, R. J. (2005). China and the G 21: A new north south divide in the WTO after Cancún. *Journal of the Asia Pacific Economy* 10, 339 358.
UNDP (United Nations Development Programme) (1990). Human development report 1990, *Report of the United Nations Development Programme*. Oxford: Oxford University Press.
UNDP (United Nations Development Programme) (2007). Human development report 2007/8, *Report of the United Nations Development Programme.* Oxford: Oxford University Press.
Williams, G., Meth, P. and Willis, K. (2009). *Geographies of Developing Areas: The Global South in a Changing World*. Abingdon: Routledge.
World Bank (2007). World development report 2007, *Report of the World Bank.* Washington, DC: World Bank.
Worsley, P. (1979). How many worlds? *Third World Quarterly* 1, 100 108.
Worsley, P. (1984). *The Three Worlds: Culture and World Development*. London: Weidenfeld and Nicolson.

Relevant Websites

http://www.g8.utoronto.ca
G8 Information Center.
http://www.oecd.org
Organization for Economic Co operation and Development.
http://www.undp.org
United Nations Development Program homepage. Follow links to the Human Development Report for national statistics.
http://www.worldbank.org
World Bank homepage. Follow links for statistics on national economies.
http://www.wto.org
World Trade Organization homepage.

Flâneur, The

A. Latham, UCL, London, UK

Glossary

Arcades An architectural form that emerged in Paris at the end of the eighteenth century and was popular until the second half of the nineteenth century. Arcades consisted of a pedestrian passage, flanked by a row of facing shops, open at both ends and roofed in glass and iron roof. They usually connected two parallel streets. They were a historical precursor to both the department store and the enclosed shopping mall.

Baudelaire, Charles (1821–67) A Parisian poet, journalist, and publisher. Seen by many as the first poet of the modern city. Closely associated with the figure of the *flâneur*.

Benjamin, Walter (1892–1940) A Jewish German cultural historian, social critic, and philosopher. Key member of the Frankfurt School of critical social theory. Most well known for his writings on the emergence of a commodified, capitalist, metropolitan culture in nineteenth-century Paris.

Commodification The process though which something of no previous market value becomes part of the market economy, that is to say, becomes a commodity. The term originates from the writing of Karl Marx.

Flâneurie The activity of being a *flâneur*. In contemporary Western cities individuals are constantly switching between a kind of *flâneurie* and other more instrumental ways of navigating through urban space.

Modernity The sense that the world is defined by constant social, economic, and technological change.

Situationist International/Situationism An international left-wing avant-garde artistic and philosophical movement active from late 1950s to the early 1970s. The work of the situationists centered on critiquing the commodification of everyday life. Central to this critique was the artful generation ‘situations’ that ruptured the otherwise smooth surface of everyday commodity culture.

Surrealism An avant-garde artistic movement that originated in Paris in the early 1920s. Surrealism sought to tap directly into the creative powers of the human unconscious.

Introduction

The *flâneur* was a nineteenth century literary figure. Originating from the French verb *flâneur*, which literally translates to ‘stroll’, the *flâneur* has come to personify a key moment in the emergence of a consumption driven, modern, capitalist urban economy. This importance is largely a product of the emphasis that two key writers placed on the *flâneur*: Charles Baudelaire, a mid nineteenth century Parisian poet and critic, and Walter Benjamin, a mid twentieth century interpreter of Baudelaire, and the rise of consumer capitalism. Both Baudelaire and Benjamin saw the *flâneur* as being not simply defined through their style of movement, but by the fact that they are walkers or pedestrians. The flâneurs are much more defined by their stance toward – and their relationship to – the urban crowd within which they find themselves. The *flâneur* represents the urban individual who has come to terms with the anonymity and confusion of the urban crowd to such an extent that he is genuinely at home within it. As such, the *flâneur* is significant in three ways. First, he (as will be seen, the *flâneur* is almost always a man) embodies an exemplary way of experiencing and making sense of the energy and confusion of the commodified spaces of the modern metropolis. Second, and simultaneously, the *flâneur* also represents a particular artistic and critical sensibility that manages to reach beyond the ephemeral nature of urban life and somehow captures its essence. The *flâneur* in the figure of the artist, the writer, or the journalist represents the possibilities of creative becoming within the capitalist city. Third, if the *flâneur* represents the possibilities for self becoming within the capitalist city, he also maps out the limits of this self fashioning, and this self becoming.

The Emergence of the Flâneur

The *flâneur* first emerged as a recognizable social form in 1830s’ and 1840s’ Paris. Closely associated with the dandy – an English term from the Regency period that referred to aristocratic men who cultivated an exaggerated interest in fashion and manners – the *flâneur* was a person who delighted in the experience of being among urban crowds. What was distinctive about the *flâneur*, and which marks him off from the regency dandy is his absorption in the everyday minutiae of the city. Whereas the dandy was principally concerned with his own self stylization (again like the *flâneur*, the dandy was always a man), and as part of that self stylization always finding himself in the most fashionable places, the flâneur’s focus is more democratic: he is interested in everything and anything that is going on in the city. As Charles

Baudelaire, the most eloquent and committed champion of the *flâneur* and *flâneurie*, wrote of the *flâneur*:

> The crowd is his domain, just as the air is the bird's, and water that of the fish. His passion and profession is to merge with the crowd. For the perfect idler, for the passionate observer it becomes an immense source of enjoyment to establish his dwelling in the throng, in the ebb and flow, the bustle, the fleeting and the infinite. (Baudelaire, 1972: 399 400)

This is an interpretation of the *flâneur* that is in some senses idiosyncratic and distinctive to the writing of Baudelaire. If, along with Baudelaire, novelists such as Honoré de Balzac and Victor Hugo saw dignity in the *flâneur*, for many other writers he was, in fact, little more than an idler. Indeed, a popular French dictionary published at the start of the nineteenth century de scribed the *flâneur* as "a lazybones, a loafer, a man of insufferable idleness, who doesn't know where to carry his trouble and his boredom" (Ferguson, 1994: 24). But for Baudelaire, in finding a home in the urban crowd the *flâneur* makes himself a paradigmatic figure of modernity: he is a heroically modern individual.

This notion of the *flâneur* as a heroic figure is crucial to the *flâneur's* enduring intellectual significance. In elevating the *flâneur* from a simple 'idler' or blasé dandy, Baudelaire transforms the *flâneur* from an incidental figure to someone who stands at the intersection of three forces that defined nineteenth century modernity: (1) the rise of the large industrial city, (2) the dissolution of traditional social bonds that anchored social identity, and (3) the ever in creasing commodification of everyday life. The *flâneur* as an ideal type manages both to give form and meaning to the maelstrom of change that defined the nineteenth century metropolis, and he defines his own identity through his interaction with the social forces pressing that change. Equally, however, the experience of the *flâneur* also embodies the precariousness of the social relationships that define modernity: the *flâneur* as someone who enjoys the urban crowd and who is also dependent upon it to make a living as a writer or an artist. While Baudelaire himself drew inspiration for his own descriptions of the *flâneur* from a range of contemporary Parisian literary and artistic figures, most notably the illustrator Constantine Guys, it is his own life and his own writing that most tellingly express the experience of the *flâneur*.

Charles Baudelaire, the Urban Crowd, and Nineteenth-Century Paris

Baudelaire was born in Paris in 1821 and, apart from spending part of this childhood in Lyons and 2 years near the end of his life in Belgium, he lived his entire life there. Paris was, so Baudelaire claimed, the only place where he managed to obtain either happiness or inspiration. It was only in Paris, its streets, cafes, parks, and apartments, that Baudelaire found what he needed to live anything like a worthwhile life, and the form of life that he pursued owed almost everything to that city. Although Baudelaire had a small inheritance, he earned his living as a journalist, publisher, and poet. Yet, what is striking about his work is how he reinvented the role of the poet, as that of the poet as *flâneur*. Baudelaire sought to be a poet of the now, the everyday, a poet of the modern city. He wrote how "Parisian life [was] rich in poetic and wonderful subjects" and how in the city "[t]he marvellous envelops and sat urates us like atmosphere." Indeed, the very form and rhythm of his poems are fundamentally embedded within the labyrinth of the city. For example, in describing the stirrings of prostitutes and criminals in the early evening in the poem 'Dusk', Baudelaire wrote:

> One hears the hissing kitchens close at hand,
> The playhouse screech, the blaring of the band.
> The tables at the inns where gamesmen sport
> Are full of swindlers, sluts, and all their sort.
> Robbers who show no pity on their prey
> Get ready for their nightly work a day
> Of cracking safes and deftly forcing doors,
> To live a few days more and dress their whores.
> (Baudelaire, 1993: 193)

This scene could only be of a city. But, while dependent upon the urban, Baudelaire's poetry does much more than simply reflect what is already there. In the poem above Baudelaire throws image upon image, the plain (the kitchens) upon the sinister (swindlers, sluts, and all there sorts), all the while twisting every image toward a demonic pitch (the kitchens hiss, the playhouse screeches, and the perhaps homely inn is full of swind lers), and then carelessly lets these demonic forces out into the city night (where, in the next verse, they set to slaughtering), as after all – and to return to the ordinary – the criminal has to work to live like everyone else. The movements of Baudelaire's poems through the urban landscape transform that landscape: his verse throws it out of balance, and reconfigures it. They also do some thing else that is quite remarkable. They take ordinary urban scenes, ordinary urban inhabitants, and make them the legitimate subject of art.

What is striking about Baudelaire's writing is how it involves a kind of mimesis: his poetry literally takes on the qualities of the city it describes. This empathy for the world around oneself is one of the defining features of the *flâneur*, and it is one of the key abilities that distinguishes him from the simple '*badaud*' or 'gaper'. In his prose poem 'The Crowds', Baudelaire (1991: 44) describes how "the Poet benefits from an incomparable privilege which allows him to be, at will, himself and others." [...] "He makes his own all the professions, all the joys, and all the sufferings that chance presents to him." Following the argument of the German cultural historian and social critic Walter Benjamin, one might say that in his poetry Baudelaire attempted to get hold of his world mimetic ally. To do this required a remarkable, and carefully cultivated, modality of interacting with the urban crowd: a perceptual modality that involved a subtle and complex play between distance and closeness, empathy and boredom. The *flâneur* had to open himself up to the crowd, while simultaneously protecting himself from the danger of sensory overload that the urban crowd pre sented. Baudelaire described how entering a crowd was like moving "into an enormous reservoir of electricity." It is the *flâneur's* ability to channel this energy through his own empathy that gave him his artistic power, and it was also his control of this empathy that protected him from being consumed by the crowd. The *flâneur* and Baude laire's empathy was founded on an ability to turn his empathy back in on the self so that it did not escape back on to the original empathic object. It was this distance that allowed Baudelaire to remain, despite the poetically empathic communions occurring within him, apart from the world he observed. In addition, the *flâneur's* mobility was one more defence: at any time he could detach himself from one scene and move on.

Much of Baudelaire's poetry speaks of the distinct pleasures, and possibilities for momentary transcendence, that come through the play of surface within the urban crowd, most famously in the sonnet 'To A Woman Passing By':

> Around me roared the nearly deafening street.
> Tall, slim, in morning, in majestic grief,
> A woman passed me, with a splendid hand
> Lifting and swinging her festoon and hem;
>
> Nimble and stately, statuesque of leg.
> I, shaking like an addict, from her eye,
> Black sky, spawner of hurricanes, drank in
> Sweetness that fascinates, pleasure that kills.
>
> One lightning flash … then night! Sweet fugitive
> Whose glance has made me suddenly reborn,
> Will we not meet again this side of death?
>
> Far from this place! too late! never perhaps!
> Neither one knowing where the other goes,
> O you I might have loved, as well you know.
> (Baudelaire, 1993: 189)

In these lines, the urban crowd exists as a source of excitement and mysterious surprise. Indeed, the crowd's movement defines the very way in which this figure is encountered. Both the poet and the woman are caught within the crowd's momentum, their meeting can only be a fleeting one, could only be 'one lightning flash'. This ephemerality, with all the unspoken mysteries that reside within it, is the very pleasure of the crowd; in it, no tedium, no mundanity has any chance to gain hold. It was this kind of ephemeral, yet moving moment that Bau delaire sought to give form to in his writing. As Benjamin (1973: 125) wrote commentating on 'To a woman passing by': "This is the look […] of the object of a love which only a city dweller experiences, which Baudelaire cap tured for poetry, and of which one might not infrequently say that it was spared, rather than denied, fulfilment."

If Baudelaire welcomed and indeed championed the powers of this momentarily shared gaze, it is also clear that something was absent within it. Baudelaire describes the women in 'To A Woman…' as aloof and distant, and unlike the poet she does not appear to be touched by the encounter as he is. Indeed, there is a vein of coldness that runs through 'To A Women Passing By'. To understand this coldness one must return to the dynamics of the *flâneur's* empathetic gaze. Baudelaire talked of the artist among the crowd as "like a mirror as vast as the crowd." Indeed, a process of narcissistic mirroring lies at the very heart of the *flâneur* poet's gaze. It is clear that the women in 'To a woman passing by' makes eye contact with the poet. But the shock felt by the poet, and its resonance as expressed within the lines of the poem, are profoundly monological. The closing "O you I might have loved, as well you know" has nothing to do with the sighted women's thoughts, and everything to do with the poet's.

How are we to understand this apparent failure of the *flâneur* to really connect? The most common reading of this failure is that it represents less the failure of the poet *flâneur* himself, and to read his lack of connection as an allegory of the profound social and economic transfor mations within which Baudelaire the *flâneur* was bound. Following an argument first put forward by Walter Benjamin in 1973, in his book *Charles Baudelaire: A Lyric Poet in the Age of High Capitalism*, the gaze of the *flâneur* can be seen as bound to the process of commodification that was coming to pervade all relationships in the nine teenth century metropolis. Within this world, the joys of looking enjoyed by the *flâneur* were the joys of the market, of the value free fall of the commodity:

> The intoxication to which the *flâneur* surrenders is the intoxication of the commodity around which surges the stream of customers. If the soul of the commodity which Marx occasionally mentions in jest existed, it would be the most empathetic ever encountered in the realm of souls for it would have to see in everyone the buyer into whose hand and house it wants to nestle. (Benjamin, 1973: 55)

But for the *flâneur*, his free roaming gaze is only one half of the commodification process: the other was the individual's entrapment within the very forces on which that gaze played. As Benjamin wrote: "Baudelaire knew how things really stood for the literary man: As *flâneur*, he goes to the market place, supposedly to take a look at it, but already in reality to find a buyer" (quoted in Buck Morss 1989: 185).

The Historical Decline of the *Flâneur* and His Intellectual after Life

It is important to stress the achievements of Buadelaire's writing. As Benjamin (1973: 154) wrote, the singular achievement of Baudelaire was that he took "something lived through" and gave it "the weight of experience." But in any case it was not the failure of the *flâneur* as writer or artist that led to the demise of the *flâneur* in the pure form expressed in Baudelaire works. It was the changing structure of Paris itself that led to the *flâneur's* decline: the construction of the city's grand boulevards and the rise of the department store in the later half of the nineteenth century robbed the *flâneur* of his natural habitat. The construction of the grand boulevards gen erated large increases both in the volume and speed of nonpedestrian traffic throughout Paris. These increases made the act of strolling the streets evermore precarious. At the same time, the development of the department store and the parallel decline of Paris' many arcades robbed the *flâneur* of another of his key haunts. Depart ment stores were places to loiter in, but only for a certain purpose: that of shopping and consumption. This was definitively not what the *flâneur* and his *flâneurie* was about – even if the department stores had been prepared to be accommodating of his presence.

So what has happened to the *flâneur*? If he has dis appeared from the streets of Paris, the *flâneur* continues to have a significant place in much contemporary writing on the contemporary city. There are three main ways that the *flâneur's* enduring significance can be traced. First, the kind of emphatic gaze described by Baudelaire and carefully explicated by Benjamin can be seen as the original or 'Ur' form of the consumerist gaze that defines a great deal of everyday life. What is significant about the gaze of the *flâneur* is that it has become a taken for granted, natural, way of looking and relating to much of the contemporary world. All sorts of ways of looking and interacting with people and things are defined through the kind of emphatically distanced, mobile, gaze like that of the *flâneur*. Television, cinema, shopping, driving, the Internet, all involve a modality of relating to the immediate environment that is very close to that of the *flâneur* – even if it has been stripped of the heroic element that so concerned Baudelaire. What is inter esting about this interpretation of the afterlife of the *flâneur* is the way in which a certain kind of visual *flâneurie* has become universal. It is no longer simply associated with the male gaze – although that gaze does in all sorts of ways remain privileged – and also with the visual practices of women. Indeed, some commentators such as the sociologist Zygmunt Bauman see the decline of the nineteenth century *flâneur* as intimately tied up with the commercially driven feminization of urban space: urban public space comes to be thoroughly defined by the commercial calculus of salesmanship and profit, and the shopper – who is ordinarily a woman – becomes a pas sively empathetic vessel responding only to the carefully crafted phantasmagorias constructed by marketers. Thus, as the cultural historian Susan Buck Morss writes:

> The *flâneur* becomes extinct only by exploding into a myriad of forms, the phenomenological characteristics of which, no matter how new they appear, continue to bear his traces as ur form. This is the 'truth' of the *flâneur*, more visible in his afterlife than in his flourishing. (Buck Morss, 1989: 346)

Second, the *flâneur* remains significant because the kind of heroic stance toward the seductions of the modern city fashioned by Baudelaire continues to be seen as full of artistic – and somewhat ironically – critical possibilities. The list of artists and artistic movements who have been influenced by Baudelaire (or indeed Benjamin's inter pretation of Baudelaire) is too long to list in detail. Nonetheless, within human geography, three schools of artists are of particular importance: surrealism, the Situ ationist International (SI), and contemporary performance art. Each of these schools has taken Baudelaire's refusal to be fully seduced by the pleasures of consumer capitalism, along with his intense conviction that the everyday has a power to reveal the beautiful, as the starting point for a wide ranging series of explorations of urban space. Taking the techniques and insights developed by the surrealists, SI, and working with artists and political activists, geog raphers have explored the possibilities of engendering new ways of looking at and engaging with the ordinary spaces of cities. In a more traditionally academic vein, other writers have seen *flâneurie* less as an actual everyday practice, and more as "an analytic form, a narrative de vice" (Jenks, 1995: 148). In these accounts, the researcher becomes, like Benjamin, a kind of intellectual *flâneur*, who attempts through observation, historical detective work, and carefully constructed texts, to guide the reader through the spaces of the contemporary city. Within geography, perhaps the best examples of this type of work are Jon Goss' essays on the Mall of America – America's biggest shopping mall at the time Goss published his essays – and Allan Pred's writing on modernity in Sweden.

Finally, the *flâneur* has remained a central starting point for discussions about the emergence of a con sumption dominated, capitalist, urban society in the nineteenth and early twentieth century. In this context, the focus is less on the figure of the *flâneur* and more on the visual practices that he embodied, that is to say, his *flâneurie*. For social historians and historical geographers working in this area, *flâneurie* has become "shorthand for the mode of modern urban spectatorship that emphasizes mobility and fluid subjectivity" (Schwartz, 1999: 9).

Two strands of this kind of work are worth high lighting. The first concentrates on exploring the role of women in the nineteenth and earlytwentieth century metropolis. Early feminist commentators on Walter Benjamin's writing on the *flâneur* stressed the degree to which the gaze that Benjamin was writing about was profoundly masculine. However, more recent work has emphasized the degree to which women were in fact actively involved in the making of the urban crowd. In particular, a great deal of writing has demonstrated the very real freedoms and possibilities that the new con sumptions spaces of the nineteenth and early twentieth century presented to women. Women might have been precluded from being *flâneurs* in the idealized style of a Baudelaire, but they found ways of engaging in all sorts of types of *flâneurie* that were liberating in quite profound ways. The second strand of work attempts to move be yond the *flâneur* and *flâneurie* and examines how the urban crowds that emerged in the nineteenth century metrop olis were more than just an incoherent mass. In fact, these crowds came to constitute themselves as a kind of public: a kind of imagined community. Through the crowd's interaction with a range of novel forms of mass media that were emerging at the same time – from illustrated newspapers, popular theaters, to wax museums and cinema – the second half of the nineteenth century saw the emergence of new forms of mass spectatorship. In these accounts, the '*badaud*' or 'gaper' that Baudelaire and Benjamin were so dismissive of regains his or her his torical dignity. They become players in the construction of a new kind of urban community. In this work the *flâneur* does not disappear, but he does rather lose his central place in the story of nineteenth century urban modernity.

See also: Consumption; Cultural Geography; Marxism/Marxist Geography I; Public Space; Retail Geographies.

Further Reading

Amin, A. and Thrift, N. (2002). *Cities:* Reimagining the *Urban*. Cambridge: Polity.

Baudelaire, C. (1972). *Selected Writings on Art and Literature*, Charvet, P. (trans.). London: Penguin.

Baudelaire, C. (1993). *The Flowers of Evil,* McGowan, J. (trans.). Oxford: Oxford University Press.

Bauman, Z. (1995). *Life in Fragments. Essays in Postmodern Morality*. Oxford: Blackwell.

Benjamin, W. (1973). *Charles Baudelaire: A Lyric Poet in the Era of High Capitalism*. London: Verso.

Buck Morss, S. (1986). The flaneur, the sandwichman and the whore: The politics of loitering. *New German Critique* 39, 99 140.

Buck Morss, S. (1989). *The Dialectics of Seeing: Walter Benjamin and the Arcades Project*. Cambridge, MA: MIT Press.

Featherstone, M. (1998). The flaneur, the city and virtual public life. *Urban Studies* 35(5 6), 909 925.

Goss, J. (1993). The 'magic of the mall': An analysis of form, function, and meaning in the contemporary retail built environment. *Annals of the Association of American Geographers* 83(1), 18 47.

Goss, J. (1999). Once upon a time in the commodity world: An unofficial guide to the mall of America. *Annals of the Association of American Geographers* 89(1), 45 75.

Jenks, C. (1995). Watching your step: The history and practice of the flaneur. In Jenks, C. (ed.) *Visual Culture*, pp 142 160. New York: Routledge.

Knead, L. (2001). *Victorian Babylon: People, Streets and Images in Nineteenth Century London*. New Haven, CT: Yale University.

Latham, A. (1999). The power of distraction: Distraction, tactility, and habit in the work of Walter Benjamin. *Environment and Planning D: Society and Space* 17(4), 451 473.

Lees, L. (ed.) (2004). *The Emancipatory City: Paradoxes and Possibilities?* London: Sage.

Pinder, D. (2005). Arts of urban exploration. *Cultural Geographies* 12(4), 383 411.

Pinder, D. (2005). Urban encounters: Dérives from surrealism. In Adamowicz, E. (ed.) *Surrealism: Borders/Crossings*, pp 39 64. Oxford: Peter Lang.

Pred, A. (1995). *Recognizing European Modernities: A Montage of the Present*. London: Routledge.

Prendergast, C. (1992). *Paris and the Nineteenth Century*. Oxford: Blackwell.

Rappaport, E. (2000). *Shopping for Pleasure: Women in the Making of London's West End*. Princeton, NJ: Princeton University Press.

Savage, M. (1995). Walter Benjamin's urban thought: A critical analysis. *Environment and Planning D: Society and Space* 13(2), 201 216.

Schwartz, V. (1999). *Spectacular Realities: Early Mass Culture in Fin de Siecle Paris*. Berkeley, CA: University of California Press.

Shaya, G. (2004). The flaneur, the badaud, and the making of a mass public in France, circa 1860 1910. *The American History Review* 109, 1.

Tester, K. (ed.) (1994). *The Flaneur*. London: Routledge.

Wilson, E. (1991). The invisible *flâneur*. *New Left Review* 191, 90 110.

Woolf, J. (1985). The invisible *Flâneuse*: Women and the literature of modernity. *Theory Culture and Society* 2(5), 37 46.

Fluidity–Fixity

P. Adey, Keele University, Keele, UK

Glossary

Mobility–Moorings Dialectic A phrase coined by John Urry and Henri Lefebvre that describes the mutually reinforcing relationship between fixity and fluidity. Fluidity and fixity are enabled and sustained by each other.

Obduracy A term which is suggestive of permanence, rigidity, or fixity. It is less about immobility but an inability to change – the capacity for something to endure and remain the same.

Pointillism Developed by Marcus Doel, the term denotes a style of thinking which presupposes that social action occurs 'in' places of a container-like character.

Relational Materialism An approach developed from science and technology studies and actor-network theory that views objects and things as both social and material entities. Matter, from this perspective, is the achievement of a matrix of relationships.

Sedentary and Nomadic Metaphysics Two ideas or imaginations of mobility in social and cultural thought. These ways of seeing the world cast fixity or fluidity as their starting points, or their norms.

> [T]here's no holding the world still and looking at it. (A. N. Whitehead, 1979)

> [T]he human intellect feels at home among inanimate objects, more especially among solids, where our action finds its fulcrum and our industry its tools. (H. Bergson, 1911)

Introduction

Fluidity and fixity seem fairly banal and material notions. Indeed, for human geographers, they may appear more relevant to the study of rivers, coasts, and lava flows. However, fluids and fixities are also powerful metaphorical devices that we use on a daily basis. Consider this: when describing someone who we might think of being very able to physically protect themselves, "he's (or she's) hard" might be a popular phrase used to describe them. Likewise, we might be involved in a conversation where we get carried away with what we are saying (something that occurs rather commonly for an academic!). If someone stops us in mid sentence, we might protest, "I was just getting into the flow of things." The first example suggests a certain kind of immobility through the toughness, or persistence of the person we might consider 'hard'. Conversely, the second example draws upon the metaphor of a fluid. Being 'carried away' implies that we are being pulled or drawn along down a stream – the momentum of the academic in full flow. For most people these metaphors might be used in everyday life as a means to describe or comprehend our world. This article shows how geographers deploy them in exactly the same way – as a means to make sense of processes, relations, and geographies.

This article sketches out how conceptions and metaphors of fluidity and fixity are embedded and routinely enacted in geographical theory and research. Given the interdisciplinary nature of the discipline, the article does not limit itself to the words of just geographers, but also pertinent works which have influenced them. The article covers how conceptions and metaphors of both fluidity and fixity are embodied within geographical epistemologies, ontologies, and research methodologies.

This article starts by tracing out the contradictions between approaches which have privileged fixity within their analysis, and conversely those which have given primacy to fluidity. The prioritization of either of these two perspectives, it is suggested, imbues such approaches in moralistic meanings and significance. The rest of the article complicates these points of view. Perspectives which presuppose a dialectic of fluidity and fixity are examined, as is the understanding of these terms within research methodology. The article finishes with a consideration of the relationality of fluidity and fixity, and further research directions being developed now or imminently.

Sedentarism and Nomadism

Lisa Malkki once argued that the inclination to fixity surmounted to a 'sedentarist metaphysics': the propensity to see the world in fixed and bounded ways. It assumes a sedentary way of life, one based on routines and habits in particular places. From the humanistic geographical tradition, a number of authors saw that one's fixity within a certain place could furnish senses of social attachment. These sedentary lifestyles could enable the creation of what is known as a 'sense of place'. Through fixity the socially barren landscape of space would be turned into a place of meaning. The home was characterized most commonly in this way. However, casting place and home in fixed and bounded ways serves to construct a binary opposition between fixity and mobility. Posed as a polar opposite to fixity, fluidity tends to be given the reverse meanings; it opposes the fixity of a meaningful place by its inherent supposition of change. Furthermore, by moving out of or beyond places meant the loosening of the social bonds one might form with a specific place. In a sense, fluidity and mobility were everything fixity and immobility were not.

These sorts of perspectives have been questioned by approaches coalescing around more fluidic principles, underpinned by what we consider to be a 'nomadic metaphysics'. More recent post structural thought, drawing upon the processual thinking of philosophers, such as Whitehead and Bergson, has driven efforts to rethink the fixity of the world anew. A more mobile world is understood alongside more mobile concepts. Cultural theorists discuss how mobility offers a way out of the fixed power structures of authority. For many authors, the airport, the rhizome, or the nomad provide useful metaphors which critique essentialist and foundational ways of thinking. Here fluidity and mobility are romanticized as acts of freedom and democratic choice, whereas fixity and immobility are seen as the product of domination and subjugation. For De Certeau, the idea of resistance was described by the process of a fluid – a wave washing over the solidity of the fixed and enduring power of the rock.

The work of many geographers has resembled these styles of thought. The powerful frictionless fluidity of the nomad even featured in Halford Mackinder's 1904 essay on the *Geographical Pivot*, which discussed the 'mobilities of power' generated by transportation technologies in relation to the fixed land they moved over. In essentializing fluidity with power, and fixity with geopolitical vulnerability, for Mackinder the topographical characteristics of the Asian steppe offered the free flowing mobilities that would permit an army to move swiftly across the continent, and seize global power.

Geography's quantitative revolution saw its own romantification of fluidity. Flows and fluids took on economically productive associations with the movement of people, goods, and capital in circuit models of transportation networks and location analysis. Indeed, motorways were compared to the hierarchical structure of river systems. The object of the fluid was even given ontological status as the basis for the human individual, understood as nothing but a flow for behavioralists such as George Zipf. Although as Emily Martin describes in more recent scholarship, the body as flow is not always interpreted in such a positive way, as leaky flows and fluids threaten to overcome the integrity of the skin.

Given the more recent preeminence given to nomadism and fluidity within post structural geography, several scholars have warned of the problems given to prioritizing fluidity over fixity, or nomadism over sedentarism. Two of the main critiques in this area have revolved around freedom and transformation. We can deduce that the act of equating nomadism with escape works to essentialize fluidity to a universal. Of course, this is incredibly problematic given that we do not all have the same access to mobility. Furthermore, by allocating these kinds of meaning to fluidity one forgets how some mobilities are entirely forced and restricted. Indeed, the intersecting flows of the airport terminal have been repeatedly used to characterize new kinds of nomadic existence, and even thought, regardless of the inequity of the airport experience.

However, Tim Cresswell has gone far to show us that while these ideas and concepts might have found the focus of so many academic discussions, they have not remained in this domain. Rather, it has been suggested that both nomadic and sedentary metaphysics can inform opinions surrounding immigration or strategies of protest and state techniques of control. For instance, David Atkinson shows how both sedentarist and nomadic ways of thinking and living were enacted in the Bedouin Sannusi resistance to the Italian fascist regime in Libya. The conflict saw sedentarism and nomadism pitted against each other through the different styles of warfare. For the Bedouin, guerrilla style tactics were employed which relied upon freedom of mobility, independence, and little concern with a territory to defend. Conversely, the Italians enacted a technique of fixity and containment. By constructing concentration camps they confined the nomads into an enclosed and disciplined space characterized by immobility. In the contemporary world, incoming illegal migrants and asylum seekers are interpreted, given meaning, and potentially treated along similar lines. Immigrants are described as a flood, a rising tide that threatens the stability and homeplace of particular localities. Comparable terms have been used to describe the in movement of tramps, homeless people, and travelers. Rendered a leaky container, place is transgressed.

Dialectics

Other approaches have worked less to separate fluidity and fixity but understand how they are intimately related. For David Harvey, the bounded and fixed senses of place favored by some humanistic geographers may, in fact, be seen as reactions to the threatening and unsettling world of globalization and the fluidity of capital. Harvey's approach to place runs in parallel to the kind of dialectical relationship between fixity and fluidity that, he argues, is so fundamental to the perpetuation of capitalism.

For Harvey, and several of the authors who have followed his work, the hallmark of the capitalist system is the dialectical relationship between fluidity and fixity. Drawing on Karl Marx, Marshall Berman, and later Alfred North Whitehead, Harvey argues that the circulation of global capital is dependent upon an apparent tension between fixity and fluidity. This tension is evident in urban spaces which function as 'spatial fixes'. The city is a site in which production and consumption are located in just such a way that capital accumulation can continue. For Harvey, these fixities are only ever temporary states of equilibrium. Crisis points meet the city with creative destruction, restructuring it for the next surge of capital accumulation. In other words, the perpetual running of capital cannot occur without the fixities around which it will accumulate and later depart.

In his reading of Whitehead's notions of enduring objects in *Justice Nature and the Geography of Difference*, Harvey is at pains to show that neither can there be any fixity without the generation and movement of capital to construct and maintain it. Such processes always meet crises. At such a point the fixities created by the accumulation of capital melt into thin air torn down repeatedly like the casinos in Las Vegas. New locations and spaces upon which accumulation occurs will be developed once more.

Neil Brenner has probably done most to extend Harvey's and Henri Lefebvre's thought on this topic through the notion of spatial scales. According to Brenner, scales as territorial entities offer another dimension to that of flat spatial fixities; and capital's need to territorialize and accumulate within fixed and discrete geographical spaces is repeated in a scalar dimension as various geographical scales offer hierarchical and bounded territories for capital accumulation. Furthermore, Brenner shows how throughout the last three centuries, such scales have been reworked and reorganized as the economic *status quo* reaches crisis point. Consequently, the reconfiguration of territorial spatial scales allows the processes of territorialization, de territorialization, and re territorialization to successfully occur on new relatively immobile geographical infrastructures.

Running alongside this work, scholars writing from the perspective of travel and mobility have noticed similar relationships occurring in the organization of societies and lives that occur on the move. It is considered that mobilities and immobilities are dialectically linked, and actually need each other in order to sustain various systems and processes. Borrowing ideas from complexity theory, John Urry defines this relationship as the 'mobility–moorings dialectic' in which complicated and large scale mobility systems rely upon rigidly fixed infrastructures to support them. For Urry, and later explored by Adey, the infrastructural mooring of the airport offers an obvious example of this correlation. The airport is a system of both relatively immobile materials such as bricks, concrete, and mortar; infrastructural pipelines of lead, copper, as well as spatially sunk and located institutional arrangements and contractual obligations. Such a fixed entity provides mooring to the vast matrix of the global airline system. Without this immobility, the airline system would simply not work. One might make similar comparisons to the immobility of the telecommunication and satellite communication infrastructures that permit the technologies of mobile and satellite phones and television, or even wireless broadband.

No Holding Still

Another strand of work which orbits the fluidity–fixity debate concerns discussion about the world which we study. How geographers simply comprehend and make sense of the world and the very research processes they engage in, can be seen to operate according to pretty well fixed understandings. We might think of this as a particular kind of ontology – a theory of a world that resembles a snapshot image – a world captured, fixed – a world that 'is'. Such perspectives have undergone several critiques in recent years.

In terms of research methodology, various efforts have been critical. Sarah Whatmore borrows from science and technology theorist Isabel Stengers to rethink the 'field' as a passive world out there that lies down and plays dead for researchers to capture and probe. Such a prospect is reminiscent of Whitehead's famous maxim that, "there's no holding the world still and looking at it," echoed in Dicken and Lloyd's rather overlooked complaint, "we can't just go out and 'stop the world' (Whitehead, 1979)." As an alternative approach, drawing on the writings of Simonden, Deleuze, Guattari, and others, geographers have begun considering not the world as it is, out there for us to study, but rather a world that is in becoming. This is a world that cannot just be, but that is brought forever into being. Conceiving the 'field' as a fluidic, active, complicated, and messy terrain to encounter problematizes the issue, as does an understanding of the mobilizing process of research itself. Drawing on Bruno Latour, Whatmore outlines the 'energetic exchanges' set in motion by research practice. Practices which literally put the world into play, as researchers take evidence back to the lab, compress earth and soil underfoot, or cause pedestrians to bypass their questionnaire.

While this relationship between fluidity and fixity may provoke geographers and academics to rethink their research methods, it has also disturbed conceptions of the spatial relationships geographers are concerned with. In particular, it de solidifies the more or less container like ideas of space and place that provide stages to social action. Every space is in motion says Nigel Thrift. Taking an 'ontogenetic' approach is not to suggest that things and events are understood to happen in place, a conception

Marcus Doel describes as 'pointillism' which we can re late back to sedentary thought. Rather this kind of ap proach is to understand how social action does, in fact, 'take place'. For Doel it is to understand the 'myriad currents' that traverse and indeed make space. It is not a question of ontology, but an ontogenetic animation of space, action, or subjectivity. It is an engagement with a world or a subject always turning into something else.

Body centered, phenomenological ontologies that orbit around body practice, are encompassed in some of the ideas deployed by geographers who have rethought the relationship between subjectivity and landscape. J. B. Jackson had started this some time ago as he pondered the flexible life of the mobile home dweller in the American landscape. But more recently John Wylie's rendering of his ascendance of Glastonbury Tor postures a less than static individual, but rather an 'interinvolve ment' between landscape and the body. One's subjectivity is not pre given *a priori*, just as space, is never simply 'there'. It is not out there, fixed, for us to find. Rather through the act of walking up the Tor one is assembled and performed in the very practice of strolling along or hiking upwards. For Mike Crang, this kind of space–time acts less like a container of action, but is deformed and made by it; as one's 'self' is assembled, so to is the world. Geographers influenced by Ingold's phenomenological attitude to landscape suggest that practices of walking and dwelling do not occur on top of the landscape, acting 'upon' it, but they occur with it. Fixity in this sense be comes an illusion taken from the academic attempting to separate themselves from it.

However, it is important to consider that this kind of genesis may be helped and nudged along. Dodge and Kitchin, borrowing from Adrian Mckenzie, argue that the in becoming status of the world is brought about by its augmentation – its manipulation by various software, electronic, and information communications systems. This 'technicity' serves to marry up space with software code. In this way, almost every action and movement we make becomes recorded, tracked, traced, and modified by software in one form of another. It also means that space is increasingly more reflexive. Spaces react to presence and shape and change accordingly.

These sorts of outlooks appear grounded in a world always in motion. The illusionary nature of fixity is ex plained and surpassed by notions of connection, rela tionality, and transformation. Even notions such as identity are seen as fluid and ephemeral, coalescing into semipermanent forms before moving on again. Having said that, some authors have become increasingly dis satisfied with such an emphasis. In comparing the thinking of Judith Butler and Luce Irigary, Jenny Robinson questions Butler's favoritism of instability, of identities constructed anew through performativity. While suggestive of potential social transformations, Robinson asserts that Butler elides the sticky and hard to change fixities of identity which do not alter and transform, or are simply replayed rather than replaced. Similarly, Hester Parr makes the point that many people with mental health problems seek not fluidity but sta bility – safe psychological and social boundaries. More over, Rose and Wylie complain that efforts to embrace the 'relational', 'mobile', and 'fluidic' in fields like land scape, can go too far, flattening out arrested topographies of disconnection, isolation, and immobility. Fixity, it seems, is making a comeback.

Special Effects

The final work we can dwell on concerns relationships and effects. In fact, for a group of authors the charac teristics of fixity and fluidity emerge as the effect or an accomplishment of a series of relationships. Perhaps the most exemplary instances of this work can be found in the recent focus upon architecture within human geog raphy. Drawing on theorizations of 'relational material ism' from science and technology studies and theorists, such as John Law, work in this area suggests that we think of buildings as 'building events'. In this view, the apparent stasis or movement of a building is brought about by an effect, by the sum of a particular relationship. Geog rapher Lloyd Jenkins follows Law and Anne Marie Mol's example of a World War I bunker through which the materiality of the bunker is not only given by the com position of the concrete it is made of, but rather by the relations it is involved in, such as the weather, reinforcing rods, and more. The interactions of a host of people, things, and climate act in such a way that the very old building is maintained and made durable.

Consider a wall. Clay, sand, cement, water, carbon, the energy from the fire that dries the bricks, rain, the various acids in the rain, particulates of dust and more intersect with a shoe, a hand, a football, the seat of one's pants, skin, blood, light, to create what we know as a wall. In deed, the wall may be protected through legislation, ownership rights, and a host of social forces that stop the wall being vandalized or knocked down. While these relationships hold the wall together and create its sense of fixity, other kinds of relationship can make it much more fluid. Jane Jacobs expands on the example of the Ronan Point high rise disaster in the 1970s to consider the complex webs of associations and materialities that added up to the big 'event' of the high rise apartment, and the event of its collapse. Gas, pipes, matches, load bearing flanks, Mrs Ivy Hodge, the designers, building regu lations and codes of practice, all interacted together to produce the building's fluidity.

Failure or breakdown is not the only time when flu idity is apparent within the networked sociality of buildings or individual objects. In Law's example of a Portuguese merchant ship, he describes how the various people, objects, and things that make up the ship hold together in relatively solid relationships. The network immutability realizes the task of holding together a working sea going vessel, thereby, giving the ship its mobility. However, these kinds of immutable mobiles have come under question because of the absolute fixity given to the internal space of the network, or its topology. Mol and Law discuss how some objects work precisely because of the internal fluidity of their network. Thus, something like a Zimbabwean bush pump continuously changes. Its internal structure can be modified and altered. Parts break off it and can be replaced; its shape and form alters slightly as it is used differently in different places. Its network is fluid, a mutable mobile.

Fixed and Fluid Futures

As Lackoff and Johnson put it, fixity and fluidity seem to be metaphors we live by. Their characteristics have been applied to subjects as diverse as the workings of capitalism to the meanings given to asylum seekers; from understandings of landscape to the form of buildings. They seem to be terms geographers have put to work in order to better comprehend systems of knowledge, to get at how the world functions, or to theorize the research methodologies we might employ. They are not, furthermore, just academic for they are metaphors rooted in thought, daily language, and practice. Let us now conclude with some of the directions research in this area looks to be going.

Degree, Phase, and State

Most simply, there appears to be an increasing effort, and need, to begin to differentiate these terms. Like concepts such as mobility, fluidity and fixity can become rather opaque and all encompassing, hiding the innumerable differences between one fixity and another, one fluid and another. Work in this area might begin to see fixity and fluidity as two ends on a continuum, with varying degrees in between. Alternatively, notions, such as 'phase' and 'state', as recently delineated by Martin Jones, Law, and Mol, and explored in fields, such as pedestrian dynamics, can take into account the ability for fluids or fixities to shift and change state from one moment to the next, phasing in and out. Moving beyond the binaries of fluids and fixities, Mol and Law also suggest that the properties of fire may helpfully describe the flickering presences of certain relationships.

Obduracy, Endurance, and Persistence

We also need to ask, why is it that some things do not move? Why do some things endure or attain a sense of persistence? Some of the most innovative work at the interface of fluidity and fixity debates concerns this very question. Anique Hommels' study on obduracy and the unbuilding of cities may be one such area geographers and social scientists will follow. Building upon ideas surrounding relational materialism from the social construction of technology (SCOT) and actor network theory (ANT), Hommels provides a comprehensive analysis of what holds the form of cities together. Focusing on several urban areas in the Netherlands, Hommels argues that a lack of participation and the momentum of tradition can make a city hard to budge. Citizens, closed out of the planning process may find their urban spaces particularly rigid and unmalleable. Conventions and traditions in city design and use can work to solidify structures further.

Affect, Materiality, and Sensation

Another indication of the future of research in fluidity and fixity may be found in the writings of geographers such as Ben Anderson who discuss and theorize the notion of affect. Arguing that we take account of the phenomenological and affective dimensions of subjectivity, David Bissell suggests that we consider the corporeal experiences of fixity and immobility. Such an approach elides dualisms of fluidity and fixity, by exploring how the most immobile acts of waiting for a train, listening to music, or being bored, can evoke innumerable affective resonances and transitions in the body's activity or passivity; anger, agitation, excitement, tiredness, fatigue, and hunger are some among many examples. In this sense, the corporeal affective modes of activity go beyond fluid displacement or a fixed position, and overcome romanticized and productivist associations.

See also: Actor-Network Theory/Network Geographies; Humanism/Humanistic Geography; Poststructuralism/Poststructuralist Geographies; Social Studies of Scientific Knowledge.

Further Reading

Adey, P. (2006). If mobility is everything then it is nothing: Towards a relational politics of (im)mobilities. *Mobilities* 1, 75 94.

Anderson, B. (2004). Time stilled space slowed: How boredom matters. *Geoforum* 35, 739 754.

Atkinson, D. (2000). Nomadic strategies and colonial governance: Domination and resistance in Cyrenaica, 1923 1932. In Sharp, J., Routledge, P., Philo, C. & Paddison, R. (eds.) *Entanglements of Power: Geographies of Domination and Resistance*, pp 93 121. London: Routledge.

Bergson, H. (1911). *Creative Evolution*. New York: H. Holt.

Bissell, D. (2007). Animating suspension: Waiting for mobilities. *Mobilities* 2, 277 298.

Brenner, N. (1998). Between fixity and motion: Accumulation, territorial organization and the historical geography of spatial scales. *Environment and Planning D: Society and Space* 16, 459 481.

Certeau, M. D. (1984). *The Practice of Everyday Life*. Berkeley, CA: University of California Press.
Crang, M. (2001). Rhythms of the city: Temporalised space and motion. In May, J. & Thrift, N. J. (eds.) *Timespace: Geographies of Temporality*, pp 187 207. London: Routledge.
Cresswell, T. (2006). *On the Move: The Politics of Mobility in the Modern West*. London: Routledge.
Dicken, P. and Lloyd, P. E. (1990). *Location in Space: Theoretical Perspectives in Economic Geography* (3rd edn.). London: HarperCollins.
Doel, M. A. (1999). *Poststructuralist Geographies: The Diabolical Art of Spatial Science*. Edinburgh: Edinburgh University Press.
Harvey, D. (1996). *Justice, Nature and the Geography of Difference*. Cambridge, MA: Blackwell Publishers.
Hommels, A. (2005). *Unbuilding Cities*. Cambridge, MA: MIT Press.
Ingold, T. (2000). *The Perception of the Environment: Essays on Livelihood, Dwelling and Skill*. London: Routledge, London.
Jacobs, J. M. (2006). A geography of big things. *Cultural Geographies* 13, 1 27.
Jenkins, L. (2002). Geography and architecture: 11, *Rue du Conservatoire* and the permeability of buildings. *Space and Culture* 5, 222 236.
Jones, M. (2005). Towards phase spatiality: Regions, regional studies and the limits of thinking space relationally. Paper presented at the Regional Studies Association Conference, Aalborg, Denmark.
Law, J. and Mol, A. (2001). Situating technoscience: An inquiry into spatialities. *Environment and Planning D: Society and Space* 19, 609 621.
Mackinder, H. J. (1904). The geographical pivot of history. *Geographical Journal* 170, 298 321.
Malkki, L. (1992). National Geographic the rooting of peoples and the territorialization of national identity among scholars and refugees. *Cultural Anthropology* 7, 24 44.
Martin, E. (1998). Fluid bodies, managed nature. In Braun, B. & Castree, N. (eds.) *Remaking Reality: Nature at the Millennium*, pp 64 83. London: Routledge.
Massey, D. B. (2005). *For Space*. London: Sage.
Parr, H. (2006). Mental health, the arts and belongings. *Transactions of the Institute of British Geographers* 31, 150 166.
Thrift, N. (2006). Space. *Theory, Culture and Society* 23, 139 146.
Urry, J. (2000). *Sociology beyond Societies: Mobilities for the Twenty First Century*. London: Routledge.
Urry, J. (2003). *Global Complexity*. Cambridge: Polity.
Whatmore, S. (2003). Generating materials. In Pryke, M., Rose, G. & Whatmore, S. (eds.) *Using Social Theory*, pp 89 104. London: Sage.
Whitehead, A. N. (1979). *Process and Reality. An Essay in Cosmology*. London: Free Press.
Wylie, J. (2002). An essay on ascending Glastonbury Tor. *Geoforum* 33, 441 454.
Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge, MA: Addison Wesley Press.

Focus Groups

A. J. Secor, University of Kentucky, Lexington, KY, USA

Glossary

Dialogic Data Data that consist of the interchanges between research participants.

Focus Groups Group discussions for the purposes of research that are comprised of a moderator, a set of questions, a number of recruited participants, and a determined time and place within which the discussion takes place.

Focus Group Guide An outline of questions that is used to structure group discussion around particular issues.

Moderator The person who leads focus group discussion.

Snowball Technique A sampling method in which research subjects are selected by a chain of referral, with one set of contacts providing further contacts.

Focus groups are group discussions constructed for the purposes of research. Their key elements are a moderator, a set of questions, a number of recruited participants, and a determined time and place within which the discussion takes place. Focus groups have also been called group interviews or in depth groups. Researchers using these terms have emphasized that, in adapting focus group methods from marketing research, social scientists can modify group techniques to facilitate more in depth and informal discussion among participants. Focus groups have become a notable method in geography since the mid 1980s. They can be used as a primary research method or in conjunction with other methods, such as interviews, surveys, or ethnographies.

The strengths of focus group research arise from the interaction of participants. Focus groups provide a forum in which research participants can discuss and debate issues of common concern. Researchers have the opportunity to observe the range of statements and positions taken up and the ways in which they are deployed in relation to one another. Focus groups allow researchers to observe social interaction, albeit in a setting of their own construction. Group dynamics are not a flaw of the method, but rather the point of it. Depending on the goals and techniques of the moderator, focus group conversations can work to decenter the researcher and thus destabilize some of the power inequalities that inhere in the research relationship. In a focus group discussion, participants typically interact more with each other than they do with the researcher. Participants are able to shift the terms of the discussion and question each other, allowing new framings of the central research questions to emerge. In addition, researchers have found that, by encouraging participants to share experiences and form solidarities, focus groups can have positive effects. For these reasons focus groups have also been adopted as a feminist research method.

Focus groups also have weaknesses. Focus group research is not the best way to collect individual level data. A focus group with eight participants should not be thought of as eight interviews, but rather as a singular event with the unit of analysis being the dialog between participants. Focus groups are not the right method for learning people's life stories in depth. In addition, group discussions are not appropriate for investigating sensitive topics, especially in closed communities where participants may fear that their comments will travel beyond the group. While group dynamics are an important element of focus group inquiry, care should be taken to recognize and interpret the silences of certain participants and the dominance of others. Finally, focus groups should not be thought of as a panacea for problems of inequality within research practices.

Designing focus group research involves decisions about group composition, recruitment, and procedures. The number of focus groups needed for a study is variable. Usually, four to five groups on a single topic with a single population are enough to reach saturation, that is, the point where no new themes or modalities of discourse emerge. The relatively low number of groups needed to reach this point is a function of the research method itself, which provides access to discursive productions circulating within particular social contexts. Focus groups usually consist of 4–12 participants, with the optimal number being 8–10. Groups are usually composed to maximize the commonalities among participants and the differences between groups, though the dimensions of difference (age, gender, race, class, employment, etc.) vary depending on the study. While it is often advocated that focus groups be comprised of participants who were not previously acquainted, many researchers have conducted focus group research with preexisting groups. Recruitment of participants will vary depending on the kind of focus groups being conducted. While snowball techniques (in which a set of contacts is asked to provide additional contacts) will work to produce groups of people who know one another to varying degrees, the construction of anonymous groups requires wider sampling techniques.

Focus groups are led by a moderator who may be the researcher or an assistant employed for that purpose. The moderator leads the group to a greater or lesser degree, depending on the goals of the research. A guide, or outline of questions, is used to structure discussion around particular issues. Group discussions usually begin with a statement about the research project and intro ductions from all participants, after which the moderator poses general questions conducive to the free flow of discussion. Sometimes, activities such as writing or commenting upon images or stories are included in the design. Focus groups usually last from 90 to 150 min and are tape recorded with the permission of the participants.

In geography, focus groups have been used to in vestigate a wide range of topics, from the experiences of immigrants to the ways in which retailers talk about environmental responsibility. Geographers have em phasized the importance of where group interviews are conducted, since different places may affect both the comfort level of certain participants and the positioning of the researcher in relation to the participants. In her research project with domestic workers in Canada, Geraldine Pratt used focus groups to investigate the ways in which immigrant women and service providers talk about their experiences of work, race, and urban space. Some of these groups were conducted with immigrant women who already knew one another from the Philip pine Women Center. This context was an important part of the focus group research. Because these groups took place within an activist organization, the focus groups easily became spring boards for collective action around issues of shared concern.

Analysis of focus group discussions begins during the group meeting and continues through the stages of transcription, coding, and writing. Focus group tran scription is challenging because speakers often speak over one another, interrupt, and use nonverbal cues to complete sentences. Notes taken during the group meeting can be very helpful in deciphering transcripts. Focus group transcripts can be coded in the same way as interview transcripts, with themes and interpretations being flagged and cross referenced. Given that focus group research produces dialogic data (i.e., the inter changes between participants), it lends itself to discourse analysis. Such an analysis is attentive to the positions taken up, the framing of issues and debates, and the si lences, omissions, and negated statements that structure the conversation. In short, focus group analysis takes into consideration the ways in which group discussions function as social contexts within which positions and meanings are produced and contested.

See also: Discourse Analysis; Interviews: In-Depth, Semi-Structured; Mixed and Multiple Methods.

Further Reading

Barbour, R. and Kitzinger, J. (eds.) (1999). *Developing Focus Group Research: Politics, Theory and Practice*. London: Sage.

Burgess, J. (1996). Focusing on fear: The use of focus groups in a project for the community forest unit, countryside commission. *Area* 28, 130 135.

Goss, J. and Leinbach, T. (1996). Focus groups as alternative research practice: Experience with transmigrants in Indonesia. *Area* 28, 115 123.

Krueger, R. A. (1994). *Focus Groups: A Practical Guide for Applied Research*. Thousand Oaks, CA: Sage.

Limb, M. and Dwyer, C. (eds.) (2001). *Qualitative Methodologies for Geographers*. London: Arnold.

Montell, F. (1999). Focus group interviews: A new feminist method. *National Women's Studies Association* 11, 44 71.

Morgan, D. (1996). *Focus Groups as Qualitative Research*. Thousand Oaks, CA: Sage.

Pratt, G. (2002). Studying immigrants in focus groups. In Moss, P. (ed.) *Feminist Geography in Practice*, pp 214 229. Oxford: Blackwell.

Wilkinson, S. (1998). Focus groups in feminist research: Power, interaction, and the co construction of meaning. *Women Studies International Forum* 24, 111 125.

Food Networks

P. Richardson and S. J. Whatmore, University of Oxford, Oxford, UK

Glossary

Actant/Actantiality An entity that is potentially active; the potentiality of an entity to become active, to emerge as an actor through interaction within a network; the recognition of agency as a relational effect constituted by interactions within a network.

Agribusiness The agglomeration of all operations involved in the production of a final food product for consumption, such as manufacture and distribution of farm supplies, on-farm production, storage, processing, distribution, marketing, and retailing. Agribusiness is increasingly characterized by vertical integration, whereby individual transnational corporations coordinate (and profit from) activities in all spheres of exchange.

Connectivity A term borrowed from mathematics and electrical engineering, now used in post-structuralist social theory, to refer to the existence of pathways between different entities/spaces. Things can be considered as more or less connected, and so the emphasis here is on the intensity and density of connections, rather than on the metric distance between objects. Conceiving of the relations between entities in terms of connectivity results in a nonlinear and multidirectional conception of space–time.

Emergence/Emergent Referring to a behavior, property, or pattern that appears through the interaction of a number of entities or actants, and in doing so forms more complex behaviors, properties, or environments as a collective. Emergence is typically associated with an ontology of 'becoming' rather than 'being'.

Relational Materialism A philosophical approach that understands relations to be a condition central to the experience of embodiment. Material forms or assemblages are considered to be co-dependent, co-produced, and embedded within ecological networks. Within this framework, all bodies are interdependent and are always constituted through and meditated by interaction with other bodies, be they human or nonhuman.

Topology The visualization of space as an active, folded, and relational component of socioecological relations rather than as a flat, passive surface on which they take place. Topological spaces are formed interactively, through convergences, connections, and continuities, and by the tensions that exist between fixity and movement.

The notion of food networks emerged in popular, policy, and academic discourses in the 1990s and is closely allied with the terminologies of 'alternative food networks' and 'commodity networks'. Less a development of agricultural geography or food geography *per se*, the notion of a food network has become a key heuristic device within the broader interdisciplinary realm of agro food studies. Conceptualizing the connections between the diverse activities and arenas of food production, distribution, retailing, and consumption in terms of a 'network' signals a shift in thinking about the space–time of agro food (away from the systems based approaches associated with political economy, which had been dominant). In an influential article, 'The social construction of international food: A new research agenda', Arce and Marsden argued that the "political economy perspective on the food system has reached its empirical and conceptual limits." This statement marks a definitive moment in the turn toward 'network' approaches in agro food studies.

In advocating for network based approaches, Arce and Marsden pointed to the deterministic character of approaching agro food through the lens of political economy. Pre 1990s debates, they argued, had been preeminently concerned with the application of economistic, industrial restructuring theories to agricultural geography. By the end of the 1990s, however, political economy had developed a wide repertoire of concepts for approaching agro foodstuffs; including agro food systems, food regimes, agro food complexes, agri business, and commodity chains. The 'commodity chain' approach (largely inspired by Freidland *et al.*'s 1984 seminal work on lettuce) was perhaps the most popular of these and had a tendency to focus analysis on relations between farm businesses and agri chemical suppliers and/or food processing industries. By contrast, the terminology of 'food regimes' signaled an analytical focus on the regulatory apparatus sustaining global agricultural markets and prices. What these approaches share in common is a central concern with industrial food production and an engagement with the corporation and/or the nation state as the principal units of analysis. Within agro food studies, and geography in particular, these approaches came to be heavily critiqued on a number of fronts. The limitations that political economy faced in thinking about (1) the materiality of foodstuffs and the question of 'nature', (2) social agency and the (un)making of so called 'macro scale' processes, (3) the spatiotemporal contingency of food production and consumption, and (4) the complex role of food consumption practices, are most relevant to the develop ment of network based approaches.

'*Globalising food*: *Agrarian questions and global restructuring*', an influential collection of essays published in 1997, is a key reference that captures the early tensions between political economy and network approaches. The tensions between these approaches have often been distracting, but have sometimes been creative, and efforts to recognize and work through them mark one of the major contributions of geographers to the development of agro food studies. As the '*Globalising food*' collection foreshadowed, the analytical disagreements and silences between political economy and network based approaches map onto a growing divergence between North American and European agro food research (and politics). These divergences are linked to distinctive theoretical influences, analytical foci, and policy engagement. As such, and also because of the locational bias of the case studies presented in most of the food networks literature, the following discussion is perhaps most salient in the European context.

The Specificities of a Network Approach to Food

This section outlines the distinctive features of a network approach to agro food. But first, it must be pointed out that the vocabulary of food networks is not confined to academic domains; food networks are conceived and discussed in activist and policy circles with equal vigor. Nor is this vocabulary deployed from a single philo sophical position. The word 'network' is often used as spatial metaphor or visualization device without articu lating the clear ontological or empirical stance that would follow from an engagement with actor network theory (ANT), (which is generally considered to be the most radical network approach). However, despite the variety of ways in which networks are conceptualized, there are four analytical commitments that are distinctive of a 'food network' approach: (1) relationality and co production, (2) embeddedness, (3) materiality and embodiment, and (4) topological spatialities.

Relationality and Co-Production

The network approach is a relational approach. In con trast to approaches that presume modernist divisions (between nature and society, for example), relational approaches understand these divisions as co produced and emergent. That is, instead of assuming autonomous and stable identities, relational approaches assume a radical indeterminacy and aim to avoid imposing *a priori* definitions and distinctions upon actors/entities. Actors or entities as such are mutually constituted through the relationships that (in)form networks. Importantly, net works, entities, and relations are understood not as simply being there, but as in a continual process of interaction and 'becoming'.

Embeddedness

An important aspect of the food network approach is an attentiveness to embedded practices, experiences, and situations. Food networks destabilize notions of global and local, or macro and micro, by focusing on embedded interactions. The notion of embeddedness specifically lo cates the interactions that constitute a network. The con cept of embeddedness captures the particularity and context of interconnections, giving consideration to the places and socionatures that inform (and are formed through) a food network. In one of the earliest studies articulating a network approach, Thorne and Whatmore argued that the relations between actors in a Fairtrade coffee network (including farmers, certificatory institutions, consumers, retailers, etc.) should be understood not as in evitable consequences of the systemic globalization of an agro food commodity, as is often portrayed, but as em bedded performances that are highly contingent. By at tending to the embeddedness of food practices and interactions, the network perspective "problematises global reach, conceiving of it as a laboured, uncertain, and above all, contested process of 'acting at a distance' (Whatmore and Thorne, 1997: 290)." In other words, looking at em bedded practices enables a critical stance on totalizing discourses and metanarratives, such as globalization.

Materiality and Embodiment

Networks always have physical dimensions; they are made of bodies that are always interactive and multiple. This is not incidental but integral. Playing with the analytic concerns of ANT, feminism, phenomenology, and environmentalism in various ways, food network approaches engage with the material embodiment of food relations, that is to say, with the metabolic relations between the eater and the eaten. Analytic concern is focused on how the process of eating connects different bodies – the stomach, the laborer, the foodstuff – and on how these connections are integral to the unfolding of the food network itself. For example, when thinking about 'quality' food networks, it is clear that the discursive 'qualities' of food are always bound up with material relations. This materialist aspect of the network approach contributes to posthumanist debates by considering food as a 'fleshy vector' through which heterogeneous entities interact and are incorporated into sociomaterial assem blages. This perspective argues that materiality 'matters' in the sense that it makes a difference or affects others, in other words that it has agential capacity. A valuable contribution to this debate is Emma Roe's concept of 'things becoming food', in which she adopts a relational materialist approach to interrogate the process of how things become (conceived as) edible. In thinking about the specific material transformations undergone as 'things become food', she emphasizes the importance of taking materiality seriously and asks us to address the meth odological challenges that are entailed.

Topological Spatialities

A key distinction of the network approach is the way in which space–times are mapped and imagined. In contrast to conventional understandings of space as a flat surface over which things move in a measurable way, networks articulate an alternative spatiality that is folded and nonlinear. The location of food production and con sumption are not mapped in terms of the metric distance between them, but in terms of the density and diversity of connections that constitute them. Moreover, whereas a non network analysis of agro food relations might understand the space–time of these relations to be the consequence of a particular economic logic, a network approach interrogates this logic and disrupts the deter minism of such analyses. Following the analytical com mitments to relationality, embeddedness, and materiality, the topological spatialities, (or folded spaces), of a food network are conceptualized as emerging through the situated relation of bodies to other (multiple) bodies. In this way, space–time is considered as relational and dynamic rather than given and static.

Food Networks in Practice

What emerges from the food network approach is an overarching interest in connectivity, and from the more radical perspectives, an interest in new (anarchic) forms of connectivity. As discussed, the vocabulary of food networks is not confined to agro food studies or aca demia. In fact, food networks are an excellent example of the ways in which concepts and discourses are mediated through academics, policy makers, social activists, and retailers. This section will point to a few practical mobilizations of the food network approach and flesh out the analytical tropes discussed above.

Co-Producing Organic Foods

It can be argued that the principles of relationality and co production are well articulated by the discourses and practices of organic farming. Through examining the activities of organic box schemes or organic community supported agriculture, a practical articulation of rela tionality can be discerned. As opposed to traditional farmers' unions and organizations, these schemes are performative modes of ordering that explicitly enroll multiple actors (including producers, processors, con sumers, and distributors) into the food network, and in doing so form producer–consumer alliances that are necessarily interconnected and interdependent. In such schemes, for example, the Riverford Organic Vegbox Home Delivery operation in Devon, UK, the commit ment of consumers to producers, and vice versa, anchors a strategic alliance that enables the network to be sus tained. Through consumer commitment, flexibility, and support in the form of regular payment to Riverford for a seasonal and variable vegbox, producers have a depend able and direct market for their produce. Producers and consumers experience a material and discursive con nectivity through the organic vegbox that is central to the constitution of the food network. Riverford's website tells us of the "desire to grow something that was not dependent on subsidies but was actually wanted by cus tomers (http://www.riverford.co.uk; accessed 15 April 2009)." Similarly, the popular philosophy articulated by the UK Soil Association regarding the production and consumption of organic foods entails a diffuse relational understanding of health and environment. That is, the Soil Association argues that organic food connects eco logical health and human health: "Eating food grown without unnecessary chemicals is both good for you and the countryside (http://www.whyorganic.org; accessed 15 April 2009)." It is purported that "Organic farmers work with nature. Trees, hedges, and wide un farmed field edges are important on organic farms. These provide habitats for natural predators such as beetles, spiders and birds which control pests (http://www.whyorganic.org; accessed 15 April 2009)." These arguments demonstrate how the notion of relationality is expressed by organiza tions that suggest and perform a co production of natures, hedges, habitats, spiders, foods, and human well being in the making of their food networks.

Embedding Food

The concept of embeddedness, which is central to the network approach, is clearly articulated through agro food movements that seek to localize production–consumption relations. The notion of local food is not the same everywhere – the idea of the 'local' is usually full of contradictions and tensions spurred by ethical debates and the politics of scale. Here, a brief discussion of an alternative food movement – FoodShare in Tor onto, Canada – will serve to convey what embeddedness can come to look like in practice. FoodShare was founded in 1985 around a concern for the growth of hunger in the city of Toronto following economic re cession. FoodShare was originally mandated to co ordinate emergency food services, but soon began to explore more restorative methods for addressing issues of food poverty, health, and community well being. As FoodShare's website states, "FoodShare tries to take a multifaceted, innovative and long term approach to hunger and food issues. This means that we're involved in diverse actions: grassroots program delivery, advocacy for social assistance reform, job creation and training, nutrition education, farmland preservation and campaigns for comprehensive food labelling are just a few examples of the areas we work in (http://www.foodshare.net; accessed 15 April 2009)." By promoting policies around social and environmental sustainability through community based outreach projects and education, FoodShare explicitly articulate a concern with 'embeddedness'. That is, they seek to 'make food a priority at all levels of society' by addressing the particular issues faced in their area, embedding more just and health giving food practices in the everyday lives of situated individuals and communities in Toronto.

Materializing Animal Welfare

As discussed above, some network approaches – particularly those influenced by feminist scholarship – take a special interest in embodiment and the shifting materialities of agro food networks. An engagement with the bodily currency of food networks explicitly connects the health of people, animals, plants, and ecologies at large. In 'quality food' campaigns, for instance, the materiality of food is explicitly acknowledged in arguments around food quality and safety. Furthermore, the bodily relation of consumer health, producer health, and environmental health is pivotal to their thought and practice. The UK Soil Association implies these connections when making the argument that "Organic food is good for your health. No food scares have ever included organic food (http://www.whyorganic.org; accessed 15 April 2009)." Another acknowledgment of the importance of materiality can be seen in the marketing campaign of 'Rachel's Organic' in Wales, who imply that it is the food itself that physically conveys 'quality' through how it tastes. Rachel's Organic explicitly connect tastiness to the conditions of production: "We believe that caring for our land and animals in a sensitive way results in far better tasting products, which are good for our customers and their families (http://www.rachelsorganic.co.uk; accessed 15 April 2009)." Similarly, the Soil Association claims that "The welfare of animals is not only important for their happiness but also to produce quality food (http://www.whyorganic.org; accessed 15 April 2009)."

Through these practical expressions of materialist modes of thought, health is understood as a material condition that is co produced through bodily relations of the soil–cow–milk–human. As 'Rachel' states: "I strongly believe that the cultivation of healthy soil leads to production of wholesome food which in turn promotes the health and well being of those who eat it (http://www.rachelsorganic.co.uk; accessed 15 April 2009)." That ecological degradation and animal cruelty lead to the transmission of ill health through a network of relations is a well versed notion in the discourses of alternative food networks. For further examples of the ways in which embodied connectivity is evoked by alternative food movements, see the activist film 'The Meatrix' and the Rachel's Organics promotional film 'Dancing Cows'.

The Topological Spatiality of Fairtrade

Fairtrade is a coded practice and a brand that focuses specifically on the relationship between producers and consumers. There are over 2000 Fairtrade certified products in the UK, which generally come from developing countries in the global South, such as the coffee farmer cooperative, Coopeldos, in Costa Rica. Producer organizations that supply Fairtrade products are inspected and certified by the Fairtrade Labelling Organisations International. The label should ensure that producers receive a minimum price covering the cost of 'sustainable production' and premium for support of social or economic development projects. In discussing Fairtrade coffee, Whatmore and Thorne argued that strengthening and embedding certain points of connection, such as farmers' cooperative institutions, agricultural practices, and consumer knowledge, is important to maintaining flows and connectivities throughout the network. These points of connection are encoded in technologies and legislation – in particular, through labeling and certification practices – which are imperative to the 'success' or durability of the Fairtrade coffee network. Traceability and certification practices work by territorializing, or stabilizing, particular social and technical arrangements in order to facilitate the movement of products and strengthen the reliability of the ethical claims that they embody. The Fairtrade label thus makes certain forms of relation durable. In this process, actors are brought into specific material and ethical relations that remap the coffee network. What is important is that the spatiality of Fairtrade coffee is not understood in terms of the metric distance traversed by the coffee as it travels from the locality of the (globally sourced) producer to the locality of the (globalized) consumer. Rather, different actors are connected through specific material and ethical relations, and it is the form that this connectivity takes that is important to us, and to the producers/consumers. The binary concepts of 'local' and 'global' then seem senseless to the spatiality of this coffee network: through the ethical and material connectivity articulated by the Fairtrade label, the spaces of the global and the local are enfolded.

Conclusion

At this point it is important to remember the diversity of network based approaches that are employed by food geographers. However, it will by now be clear that the network approach to studying and engaging with the geographies of agro food has lent itself to, and been pioneered by, movements and critiques that consider themselves to be alternative or radical. What is it about the network approach that makes it so amenable to a politics of 'alternativeness'? Although the network approach is entirely applicable to studying any set of food relations, and does indeed de stabilize the alternative versus conventional distinction that we might crudely imply, there is a particular kind of ethics that is articulated through the idea of food networks. It is these ethical implications that make it useful to alternative food movements and radical scholarship. In particular, thinking about relations brings the co production and complexity of materiality to the fore and in doing so, calls for a more than human relational ethics. Not only does this approach challenge exploitative relations, it also gives credence to the active participation of 'nonhumans' in food networks This relational approach to ethics swells the company of others to whom our responsibilities extend. As an analytical approach to agro food relations, the network approach facilitates an empirical understanding of the material ways in which human and nonhuman health become intimately connected through the interbodily relations that bind us all together.

Although the more radical approaches to agro food have been seen as interrogative of the geometry of power in networks (by showing how struggles build and maintain particular asymmetrical relationships), it is important to note the critique to which it has been subject. Some are troubled by ANT's abandonment of the social categories associated with inequality – such as gender, race, and class – without which, they believe, no analytically or politically adequate means of addressing questions of power is possible. Others are more troubled by the opening up of politics to more than human others which is seen as consequence of a loss of privilege attributed to human intentions through ANT's 'symmetrical' approach to agency. These critiques reveal the deep seated and often unexamined humanist commitments of orthodox social science approaches. Readers should consider the different ways in which power and agency are understood and situated from the perspective of ANT and recognize the challenging and, perhaps unfamiliar, philosophical resources on which it draws to adequately assess these critiques.

See also: Actor-Network Theory/Network Geographies; Agriculture, Sustainable; Animal Welfare, Agricultural; Commodity Chains; Embodied Knowing; Environment; Food Networks, Alternative; Food Regimes; Globalization and Transnational Corporations; Human-Nonhuman; Local-Global; Moral Economies; Nature; Nature-Culture; Space-Time.

Further Reading

Arce, A. and Marsden, T. K. (1993). The social construction of international food A new research agenda. *Economic Geography* 69(3), 293 311.

Bryant, R. L. and Goodman, M. K. (2004). Consuming narratives: The political ecology of 'alternative' consumption. *Transactions of the Institute of British Geographers* 29(3), 344 366.

Fitzsimmons, M. and Goodman, D. (1998). Incorporating nature: Environmental narratives and the reproduction of food. In Braun, B. & Castree, N. (eds.) *Remaking reality: Nature at the millennium*, pp 194 220. London: Routledge.

Friedland, W. H. (1984). Commodity systems analysis: An approach to the sociology of agriculture: Focus on agriculture. In Schwarzweller, H. K. (ed.) *Research in Rural Sociology and Development*, pp 221 235. Greenwich, CT, USA: JAI Press.

Goodman, D. (1999). Agro food studies in the 'age of ecology': Nature, corporeality, bio politics. *Sociologia Ruralis* 39(1), 17.

Goodman, D. (2001). Ontology matters: The relational materiality of nature and agro food studies. *Sociologia Ruralis* 41(2), 182.

Goodman, D. and DuPuis, E. M. (2002). Knowing food and growing food: Beyond the production consumption debate in the sociology of agriculture. *Sociologia Ruralis* 42(1), 5.

Goodman, D. and Watts, M. (1997). *Globalising food: Agrarian questions and global restructuring*. London: Routledge.

Ilbery, B. and Maye, D. (2005). Alternative (shorter) food supply chains and specialist livestock products in the Scottish English borders. *Environment and Planning A* 37(5), 823 844.

Lockie, S. (2002). The invisible mouth: Mobilizing 'the consumer' in food production consumption networks. *Sociologia Ruralis* 42(4), 278.

Lockie, S. and Kitto, S. (2000). Beyond the farm gate: Production consumption networks and agri food research. *Sociologia Ruralis* 40(1), 3.

Morgan, K., Marsden, T. and Murdoch, J. (2006). *Oxford geographical and environmental studies series: Worlds of food: Place, power and provenance in the food chain*. Oxford, New York: Oxford University Press.

Murdoch, J. (2000). Networks A new paradigm of rural development? *Journal of Rural Studies* 16(4), 407 419.

Murdoch, J. (2006). *Post structuralist geography: A guide to relational space*. London: SAGE.

Raynolds, L. T. (2002). Consumer/producer links in Fair Trade coffee networks. *Sociologia Ruralis* 42(4), 404.

Roe, E. (2006). Things becoming food and the embodied, material practices of an organic food consumer. *Sociologia Ruralis* 46(2), 104 121.

Sage, C. (2003). Social embeddedness and relations of regard: Alternative 'good food' networks in south west Ireland. *Journal of Rural Studies* 19(1), 47 60.

Whatmore, S. (1995). From farming to agribusiness: Global agrifood networks. In Watts, M. & Johnston, R. J. (eds.) *Geographies of global change*. Berlin: Blackwell.

Whatmore, S. (2002). Transgressing objectivity: The monstrous topicality of GM food. *Geographies: natures, cultures, spaces*, pp 120 145. London: SAGE.

Whatmore, S., Stassart, P. and Renting, H. (2003). What's alternative about alternative food networks? *Environment and Planning A* 35(3), 389 391.

Whatmore, S. and Thorne, L. (1997). Nourishing networks: Alternative geographies of food. In Goodman, D. & Watts, M. (eds.) *Globalising food: Agrarian questions and global restructuring*, pp 287 304. London: Routledge.

Relevant Websites

http://www.fairtrade.org.uk
Fairtrade Standards, The Fairtrade Foundation, London, UK.

http://www.foodshare.net
FoodShare.

http://www.rachelsorganic.co.uk
Rachel's Organic.
http://www.riverford.co.uk
Riverford Growers, Riverford Organic Vegetables.
http://www.soilassociation.org
Soil Association.
http://www.themeatrix.com
The Meatrix Films.
http://www.whyorganic.org
Providing information on Organic Issues.

Food Networks, Alternative

D. Goodman, University of California – Santa Cruz, Santa Cruz, CA, USA
M. K. Goodman, King's College London, London, UK

Glossary

Alternative Food Networks New and rapidly mainstreaming spaces in the food economy defined by – among other things – the explosion of organic, Fair Trade, and local, quality, and premium specialty foods. In these networks, it is claimed that the production and consumption of food are more closely tied together spatially, economically, and socially; however, the politics and practices of alternative food networks have more recently come under critical scrutiny from geographers and others as a narrow and weakly politicized expression of middle- and upper-class angst.

Convention Theory An influential theoretical approach to studying Alternative Food Networks. Identifying the norms, qualifications, and organizational forms involved in network coordination which uphold different conventions of 'quality', convention theory offers a general typology to distinguish product quality in terms of 'orders of worth' that specify the different logics orchestrating their production and governance.

Conventionalization Thesis With particular reference to California, the conventionalization thesis suggests that large-scale organic producers are integrated into national and global commodity networks, while holistic, 'movement farmers' or 'artisanal' growers serve low-volume, localized markets. This is also known as the 'bifurcation' thesis of structural change in US organic agriculture.

Embeddedness Drawn from a dialog between Granovetterian economic sociology and heterodox economics the so-called embeddedness in alternative food networks refers to the 're-placement' of food within its social, cultural, economic, geographical, and environmental contexts in response to the 'dis-embedding' forces of conventional food networks.

Ethical Trading Initiative Formed in the 1990s in the UK through a coalition of supermarkets, transnational corporations, activists, and development NGOs, the Ethical Trading Initiative focuses on the enforcement of international labor standards at the points of commodity production in the global South across food and a number of other commodities, such as flowers and clothing.

Fair Trade Developing from political 'solidarity' movements in the 1960s and networks of 'alternative trading organizations' created by development NGOs, church groups, and charitable institutions, fair trade was conceived originally as an alternative political project intended to replace the conventional international trading system; it now extends to all manner of foods, handicrafts, and manufactured goods. Constructed as a 'politicized' movement and market, fair trade is distinguished by its commitment to more-equitable development for poor Southern farmers and producers through direct trading links with Northern consumers who pay a premium price and 'social dividend' for goods such as coffee and bananas.

Quality Foods These are foods – such as organic commodities, those with definable (and often certifiable) places of origin, and 'distinguished' by taste – that, through both materially and socially constructed means, affect a host of assignable characteristics that attempt to set them apart from 'other' foods.

Reflexive Localism This refers to a theoretical project, *vis à vis* local food movements and their conceptualizations, that works to get beyond the typical normative and potentially conservative/reactionary localisms that have become *de rigeur* in local food activism and scholarly work.

Spatial Dynamics of Care This is where, in alternative food networks, an ethics of care is constructed across the places and spaces of food production and consumption to connect, for example, fair trade producers and consumers, and organic consumers to a pesticide-free countryside. These connections are fostered through both material and discursive means, the latter often through highly 'meaningful' labels of origin and/or images of producers designed to invoke an affective response.

Sustainable Agriculture Movements These are heterogeneous and increasingly fragmented congeries of NGOs, food and agricultural activists, government institutions, and academics working to make conventional agriculture more environmentally and socially sustainable. Some argue that the more 'radical' early incarnations of sustainable agriculture have been 'captured' by a techno-centered and flat political agenda at the expense of a more robust and holistic food democracy.

Introduction: Alternative Food-Scapes

Several interrelated processes, not entirely synchronous, but with cumulative effect, are transforming and di versifying modern food provisioning in Western Europe, North America, and many other parts of the world. These processes have created economic and cultural spaces, often designated as niche markets, for alternative food networks (AFNs), whose products – organic, Fair Trade, local and quality, premium specialty foods – are differentiated from those typically furnished by mainstream food manufacturers and retailers (see **Figures 1–5**). In their infancy, AFN products were distributed through new institutional forms parallel to mainstream channels, such as charity shops, food cooperatives, farmers' markets, community supported agriculture (CSA), or box schemes.

However, the rapid growth of AFN sales, particularly to more affluent income groups, attracted the attention of the large scale retail supermarkets, which now provide shelf space for AFN products, often under their own labels, and increasingly subject AFN producers to the rigorous cost price disciplines of supply chain management. The differentiated products supplied by AFNs and their spectacular growth has reinforced perceptions that 'quality', in its various socially constructed and material dimensions, rather than price, is becoming the new basis of competition in food provisioning. The 'turn' to quality, and the scope it provides for further product differentiation, neatly complements the competitive strategies of firms in the highly oligopolistic sector of food retailing. The neatness of fit between nonprice quality attributes and oligopolistic product differentiation strategies is illustrated perfectly by organic products in the UK, where supermarkets currently account for 70% or more of sales.

The interface between 'alternative' and 'conventional' food provisioning thus is an increasingly permeable and highly contested terrain. While the distinction between these two systems retains some heuristic value and provides insight into the eco social imaginaries and materialities of these provisioning schemes, this article highlights the competitive and cultural political dynamics unfolding along this interface as 'alternative' food products enter mainstream channels of provision. Additionally, we emphasize the contested nature of the spatial, political economic, and cultural conceptualizations of what 'AFNs' constitute in popular and academic discourse.

Quality, its material and discursive construction in farm fields, the market place, and on the table, is the fulcrum of this struggle to control modes of integration between food production and consumption. The turn to quality associated with the rise of alternative food practices and the proliferation of AFNs thus potentially represents a significant reconfiguration of producer–consumer relations and the spaces where these are performed. Although the production 'moment' in these alternative agro food circuits dominates academic research, the quality 'turn' and the growth of AFNs typically are attributed to processes whose effects work largely through the demand side as consumer knowledges, perceptions, and economies change.

These processes include heightened consumer anxieties about the safety and nutritional value of standardized industrial foods as the result of episodic food scares (e.g., BSE; *Escherichia coli*; salmonella; foot and mouth disease; and avian influenza). This crisis of confidence in mass produced 'placeless and faceless' foods is articulated particularly by higher income consumers – the 'worried well' – with the means to 'opt out' from mainstream provisioning. In this climate, rising

Figure 1 Scenes from Inside London's Borough Market: An Organic Statement. Here, in the iconic Borough Market in London, the 'practice' of AFNs is readily apparent. On one of the banners hanging over the 'Total Organics' food stall inside the Market is a declarative statement that "quality" foods "won't bite back"; this goes to the heart of AFNs' charm offensive through their often combined sense of 'safer' and ethical food.

Figure 2 The 'Vision' of AFNs. One of the important storylines of AFNs is that they provide the knowledge to consumers of where their food is coming from and how it has been produced. For the Ginger Pig, this vision means consumers knowing they are getting free-range and traditional pig breeds but also, in a slightly humorous twist, free-range 'farmers and butchers'. Also note the other trappings announcing the quality of the goods given that they supply a well-known 'foodie' café and have garnered attention from the BBC and *The Observer Food Monthly*.

Figure 3 'Knowing' Northfield Farm and its Rare Breeds. Northfield Farm's bid for quality comes through this proud statement: "Our aim at Northfield Farm is to produce and source premium quality Beef, Lamb & Pork; naturally reared & humanely killed. We specialise in rearing and sourcing Rare and Traditional breeds of cattle, sheep & pigs. We believe the producer & consumer must strive together to ensure traceability and excellence © 2004." Thus, rare and traditional-bred livestock is one of the 'alternative' selling points of Northfield Farm and one of the ways they work to differentiate themselves in the kinds of quality meats they provide.

affluence and growing income inequalities have translated these omnivore food anxieties into effective monetary demand for quality products of known provenance, transforming cultural norms of 'good food', taste, and social distinction.

The prominence of provenance and place – 'terroir' – in new constructions of quality has received further impetus from the parallel, closely related rise of ethical consumption, and a 'moral imaginary' of food which encompasses ecological sustainability, social justice,

Figure 4 "How Did it Get Here?": Tracing the Cheese. This is an hour-by-hour photo-display of where the Comté cheese wheels in the right of the picture come from; the traceability of the cheese is foremost in the presentation of this quality food.

cultural integrity, and animal welfare, for example. This has certainly been the case with transnationally sourced fair trade foods, such as coffee, and is encapsulated in the so described novel 'moral economy' created by the discursive and material producer–consumer connections that characterize fair trade networks. Here, moral geographies work to stitch together the various 'poor' places of production in the underdeveloped South with those of the 'spectacular' places of consumption in the North in the name of transnational development for small farmer communities. Price premiums and guaranteed price floors mix with the emplaced solidarity images of farmers to facilitate a 'developmental consumption' that allows Northern consumers the opportunity to convert their 'labor' of choosing fair trade goods into on the ground Third World development. This morally charged 'charm offensive' of developmental consumption is echoed in Northern AFN discourses in the precepts of consuming to support the livelihoods of small farmers, rural communities, and ecological farming through the purchase of organic and local foods.

Finally, in the European Union (EU), meso and macro level developments are creating new opportunities for AFNs and quality food products as the gradual reform of the Common Agricultural Policy (CAP) is slowly attenuating the pronounced sectoral emphasis on agricultural production and commodity support mechanisms. Impelled by the trade liberalization mandate of the World Trade Organization (WTO), this narrow and sectoral productivist focus is giving way

Figure 5 Re-producing AFNs: Organic and Artisinal Goods in Erlangen, Germany. A modern German incarnation of an AFN in the well-known form of the farmers' market. Organic and artisanal goods, like this bread maker, are at the core of this AFN in Erlangen's city center.

to a pluralistic, regionally based and more endogenous concept of integrated rural development. Thus, the in stitutional space and budgetary resources created by CAP reforms have encouraged some scholars to see AFNs as the vanguard of the 'new' model of rural de velopment, as advocated in the 1996 Cork Declaration, the Agenda 2000 reforms, and more recent EU policy documents. That is, innovative expressions of a new paradigm of European rural development are marking the transition away from the productivist, crisis ridden model of conventional industrial agriculture into AFN lands 'beyond'.

These disparate processes, operating at different scales and conjunctures, have combined to give AFNs 'room for manoeuver' in food provisioning systems, despite the continuing hegemony of mainstream industrial foods and the retail multiples. The quality 'turn' manifest in AFNs has stimulated a significant body of academic research and a wide repertoire of theoretical approaches.

Approaching AFNs: Understanding the Spaces of Quality, Embeddedness, and Ethics

The liminal, yet integral, situatedness of food in agri cultural nature, social labor, and culture gives rise to multiple material and discursive constructions of quality. This polysemic character of food, in turn, explains the complex analytics of the quality turn to alternative food networks. The main conceptual categories used to expli cate this turn in food provisioning and consumption are the interrelated characteristics of quality (e.g., taste, 'dis tinction', etc.), embeddedness, trust, place, and the local.

The contested and contingent nature of these con structs, and the associated difficulties of deciphering their complex expression in material environments, social be havior, organizational forms, discourse, and power re lations, defines the theoretical and empirical challenge posed by AFNs.

The literature on AFNs has drawn but sparsely on meso level theories and concepts. These include the network paradigm, actant network theory, conventions theory, and culture economy approaches. More avowedly sectoral or industry perspectives have extended com modity chain analysis in elaborating concepts such as 'systems of provision' (SOPs), short food supply chains (SFSCs), and global value chains (GVCs), to give greater analytical purchase on issues of governance, power re lations, and the competitive dynamics of economic rent. Other allied but more culturally inflected approaches draw loosely on the 'commodity cultures' approach in order to 'follow the thing' as these alternative foods move from and among sites of material and symbolic pro duction and consumption.

Of the meso level approaches, convention theory has been by far the most influential in AFN research since it speaks directly to the 'economy of quality'. It offers a general typology to distinguish product quality in terms of 'orders of worth' that specify the different logics or chestrating their production and governance. It thus identifies the norms, qualifications, and organizational forms involved in network coordination and which up hold different conventions of quality. The 'worlds of production' variant of this perspective has been used in the AFN literature to conceptualize the quality 'turn' as a contested process of transition. That is, from the 'com mercial' and 'industrial' worlds, with conventions based respectively on price and standardized quality norms, to the 'domestic' world, with quality conventions founded on trust, tradition, and place, and the 'civic' world, where conventions incorporate social justice and environmental sustainability.

The concept of social embeddedness has been drawn, somewhat uncritically, from the dialog between Grano vetterian economic sociology and heterodox economics, united by their critique of the asocial abstractions of neoclassical economic theory. This productive dialog influenced the 'new' industrial geography and research on Marshallian industrial districts in the late 1980s, and also informs more recent debates on the economy–culture divide in human geography. However, this genealogy and recent reworkings of the concept of embeddedness have received scant attention in the AFN literature. Similarly, major theoretical advances in human geography, notably relational conceptualizations of place, space, economy, and the politics of scale, find little re flection in AFN research, despite the critical importance attributed to the local and provenance.

The meso level perspectives deployed in AFN re search are production centered and focus primarily on the characteristic features of different modalities of economic coordination. Consumers and consumption are seen as mere appendages to the production process or as abstract inhabitants of 'spaces of quality' in typologized 'worlds of production'. Reconstructive efforts have sought to overcome this neglect by theorizing consumers as social, relational, and reflexive constituents in AFNs and deploying the notion of consumer knowledges. These efforts attempt to transcend the false dichotomy between production centered perspectives of society and cultural approaches, which respectively theorize consumption practices in terms of Marxian commodity fetishism or Durkheimian cultural totemism. This 'economic/cultural faultline' is also reflected in the unhelpful academic division of labor between agricultural 'production' soci ology and the 'cultural' sociology of food. Other authors have addressed this issue by developing notions of con sumer aesthetics to analyze cultural formations of pro duction and consumption in comparative perspective, as in the polar cases of US fast food and Italian slow food or standardized and artisanal culinary traditions.

In further efforts to get beyond these false dichotomies, recent work on AFNs has sought to explore and charac terize the expansive 'moral geographies' that inhabit and often underpin these networks. Extending the so called normative turn in human geography, research has focused on the 'economy of regard' fostered in farmers' markets and the 'spatial dynamics of concern' in constructing the spaces and places of fair trade. Here it is argued that an 'ethics of care' is woven in the tissue of the imagined and material connections between knowing consumers and distant 'others' in the diverse forms of free range chickens and cattle, pesticide free streams, artisanal bread makers, and Ghanaian cacao growers. There remains great scope for deeper theorization along these lines, in particular through a careful and critical exegesis of the limits and possibilities of these nascent conceptualizations of ethic ally drenched 'alternative economic spaces' for concrete political action and social justice.

In the main, the AFN literature has neglected theo retical development in favor of empirically grounded, case study analyses of alternative food production and provisioning networks, new economic forms, and insti tutional mechanisms of governance and policy. This emphasis has yielded 'thick' ethnographic descriptions of individual local networks and detailed analyses of the burgeoning number of food accreditation, labeling, and promotional schemes devised to provide new bases for rural development.

These generalizations extend broadly across AFN scholarship. Nevertheless, despite a shared microanalytic toolkit, the US and Western European literatures diverge significantly on several points. These two branches of scholarship project different sociopolitical imaginaries, which can be attributed to their respective research constituencies. A parallel literature has arisen on glob alized AFNs, but one heavily framed by the precepts of Third World development and the quality turn's pen chant to 'certify everything'. Thus, an implied project of this article is to suggest that further conceptual and empirical cross fertilization needs to open up between Northern focused AFN work and that on Southern 'de velopment focused' AFNs.

The American Experience: Conventionalization, Entrepreneurialism, and Localism

In the US, researchers are in conversation with activist communities and typically have strong normative com mitments to the social movements contesting the hegem ony of the mainstream, corporate industrial food system and the allied agricultural technoscientific complex. Correspondingly, AFNs and, more generally, sustainable agriculture movements (SAMs) figure as material and symbolic expressions of eco social imaginaries of alter native, local and regional 'foodsheds', whose scaling up, it is argued, would create an ecologically sustainable, egali tarian, and socially just food system. In the US literature, AFNs and SAMs thus are assessed in terms of their op positional status and 'transformative potential' to deliver progressive systemic change in food provisioning.

Measured against this standard, SAMs have been castigated persistently by academic scholars for their neglect of labor relations, social justice, and the needs of low income consumers. Despite SAMs' origins in the radical civil rights and antiwar politics of the 1960s and 1970s, contributors have emphasized the primacy of an alternative technological managerialism and related ef forts to achieve scientific legitimacy over social justice and a progressive transformative politics.

In this managerialist representation, it is argued, the organic is framed as a competing system of efficient resource management, buttressed by modernist epis temology and scientific knowledge claims. This repre sentation privileges ecological sustainability, food safety, and health concerns, placing SAMs on the neoliberal, green consumerist terrain of individual market choice and consumer sovereignty. Such technocentrism, con tributors have emphasized, has preempted progressive socioecological politics that give prominence to questions of social justice, food security, class, gender, and race. This social agenda, by default, has fallen to an uneasy coalition of Community Food Security movements and antihunger activists, who advocate local entrepreneurial initiatives and public food entitlement approaches, respectively.

These shifts in the political ambition of organic AFNs charted by US scholars have also been theorized in terms of commodity fetishism. Thus, by contrast with con ventional industrialized foods, the organic intentionally unveils its ecological production relations, 'farming in nature's image', and philosophical values of stewardship and a 'land ethic' of care are ascribed to this 'moral ecology' of transparency. However, as several authors have demonstrated, this transparency does not extend to labor, working conditions, citizenship, or equality of ac cess to nutritious food. In short, the commodity fetish has been reworked to support food safety and health claims, while continuing to obscure unjust social relations of production and consumption. The ethical standing of organic AFNs and SAMs arises from their embeddedness in sustainable ecologies and not, as in the case of Fair Trade, in moral economies of social justice.

This 'technology led vision' was endorsed by the leading SAMs in negotiations over the USDA's Proposed Organic Rule in the late 1990s and is codified in the technocentric regulatory criteria used to define acceptable production practices for the USDA's 'Organically Grown' label. This narrow conception, which effectively translates 'organic' into a market brand, a mode of technological competition, and one among many differentiated food products, has fragmented SAMs and dissipated hopes invested in organic agriculture as a progressive force.

Most recently, activist ambition and academic research has focused on the development of localized food systems, local farmer knowledge, and territorial labels of origin. This research tracks smaller organic growers, excluded by scale requirements from interregional and export markets, who are seeking new sources of economic rent and livelihood by going 'beyond organic' and finding niches in local direct marketing and local food networks. Some US contributors have denounced the economic privilege and WASP ish ethnocentrism that currently characterizes these networks. However, efforts to develop a food justice movement to mitigate the class and racial complexion of AFNs are isolated and incipient at this point.

In following this trajectory of SAMs, US scholars have analyzed the decline of progressive politics and structuralist critique and the emergence of a more circumspect, incrementalist analytics of change, more accommodated to the hegemonic neoliberal discourses reifying the market and consumer choice. One recent empirical study of 37 alternative food initiatives in California reveals that they are characterized by an entrepreneurial culture and a politics of engagement explicable only by neopopulism. These Californian initiatives are no longer 'oppositional' but rather 'alternative' organizations articulating narrowly proscribed claims of localism, community food security, and sustainable local agriculture.

The trajectory of fragmentation is formulated in political economic terms in debates evoked by the so called 'conventionalization' and 'bifurcation' theses of structural change in US organic agriculture, and with particular reference to California.

These formulations contend that the rapid growth of markets for organic produce, reinforced by the minimalist 'allowable inputs' approach to federal codification, has engendered a dualistic process of change. Large scale, often monocultural producers are integrated into national and global commodity networks, leaving holistic, 'movement farmers' or 'artisanal' growers to serve low volume, localized markets. Empirically, the large scale specialized segment of organic production shares many attributes with its conventional industrial counterpart, and these similarities are even more pronounced in marketing and distribution as Earthbound Farms meets Whole Foods, Inc., accentuating the industry's dual structure.

Other contributors have argued that the division between 'industrial' and 'artisan' growers 'within' the organic sector is problematic and may conceal unwarranted normative and ideological assumptions. For example, that 'artisanal' maps directly onto more progressive, movement oriented ideologies and holistic agroecologies, and that 'industrial' equates with both economic behavioral and conservative ideological elements of the 'conventionalization' thesis. This debate clearly calls for more empirical research, not least because agro industrial dynamics and capitalist imperatives of social reproduction are pervasive.

This discussion of binary structures and unexamined normative assumptions leads neatly to agrarian localist discourses and portrayals of AFNs as sites of resistance to time space distantiation and corporate power articulated through the oligopolistic, globalizing industrial food system. In contributing to those discourses, US scholar activists have formulated normative concepts such as 'food in social context', 'commensal community', and 'foodshed' as counter narratives to corporate hegemony. In this binary opposition, the local is represented as a potentially 'insulated space', which affords opportunities to re embed food provisioning in local institutions and the ethical norms of a 'moral economy' characterized by mutuality, reciprocity, and equity. These norm based, neopopulist narratives propagate ideas of 'secession', disengagement, and what some authors term the 'principle of succession', which involves a gradual transition from the industrial food system to the foodshed.

Several recent contributors have sought to problematize the conflation of social relations and spatial relations in those idealized formulations of local food networks. Such spaces, it is argued, may accommodate a reactionary 'defensive localism' marked by conservative politics and nativist sentiments, rather than the progressive, ethical, and associative economy so often imputed uncritically to localized AFNs. In short, eco social embeddedness can obscure difference, inequality, and social injustice.

Other authors have challenged such 'normative localism', where the local is cast as the realm of resistance to anomic, globalizing capitalist forces, for its erasure of the politics of the local. That is, it fails to recognize that contested issues of distributive justice, human rights, and identity can arise in these idealized, insulated spaces. The notion of 'reflexive localism' accordingly has been proposed in order to focus attention on the social relations and power asymmetries subsumed in the rhetorics of 'sustainable local development' and 'local control'. A reflexive conceptualization of localism recognizes the 'differential (or adverse) incorporation' of social classes into the market economy, civil society, and state. Social classes are differentially positioned to benefit from 'sustainable local development' but these distributive consequences are lost in the discursive trappings of normative localism.

Activist narratives of a normative localism embedded in a secessionist moral economy have also failed to problematize scale construction at the local level. This omission ignores potentially negative outcomes of the dynamic, contested interactions between local forms of social spatial organization and translocal actors and in stitutions. Rather than situating the local within a re lational politics of scale, it is framed as a social space where new economic forms incorporating ethical norms are allowed to grow and flourish. In short, the local is reified, obscuring the contested sociospatial processes involved in its construction and the practicalities of se cession and local control. Although these analytical limitations and erasures arguably reflect the pre occupation of US scholarship with the 'movement' as pects and 'transformative' potential of AFNs, similar shortcomings are also found in the European literature.

Niche Markets, Relocalization, and 'Alternative' Policy: AFNs and the Western European Experience

Although reified local–global dichotomies, norm based localism, and neopopulist rural imaginaries inhabit the European literature, it differs from its US counterpart in the prominence given to AFNs as territorialized value chains, a source of diversified farm livelihoods and, more broadly, as catalysts of revitalized rural economies. The focus is more economistic: 'just value' rather than 'just values' and communitarian discourses of social justice. The instrumental market centered tenor of AFN re search in Europe intersects with wider political economic processes, ranging from CAP reform, the transition to more pluralistic visions of rural development, and de volved rural governance to public debates on food safety and the conservation of rural socioecological environ ments. In this context, AFNs are envisioned as integral elements of a European eco social imaginary, that is articulated politically by claims of rural cultural excep tionalism and finds institutional expression as 'multi functionality' in EU rural development policy and international agricultural trade negotiations.

In buttressing claims of European exceptionalism, a complementary discourse of economic performance and competitiveness has emerged, which frames AFNs/SFSCs as new sources of value added that can be captured locally and so stimulate rural economic regeneration. These value streams arise from the capacity of AFNs to re embed and re socialize food provisioning in local ecol ogies and communities. In this way, as urged by the 2002 Curry Commission report, 'The future of food and farming', published in response to the UK outbreak of Foot and Mouth Disease, AFNs are 're connecting' with consumers through new markets of quality local produce.

In this market led, 'economic' localism, re connection is made by valorizing signifiers of the local that can be commodified and translated into premium prices; that is, into streams of economic rent. Several contributors have drawn attention to the dynamism of commodification processes as new attributes of territoriality are 'internal ized' by the discourse of food quality and subsumed by the value form, including the conservation of rural landscapes, wildlife habitat, and sustainable rural en vironments. Other authors suggest that the spatial pro liferation of territorial valorization processes will lead to 'competitive localism', threatening economic rents and accentuating the uneven development patterns impli cated in these 'new rural geographies' of food.

European AFN/SFSC research has grown rapidly in recent years and a wide selection of local and regional case studies is now available. Some contributors have made preliminary schematic attempts to assess the geo graphical distribution and economic significance of quality food networks. However, the dearth of com prehensive data is such that only 'educated guesses' are possible about the economic impacts of AFNs, their temporal, spatial, and evolutionary dynamics, and their stability in the longer term. Accordingly, the following discussion concentrates on several lacunae and erasures in European AFN research.

This literature reaffirms the role of farmers as key actors in the relocalization of food provisioning and privileged beneficiaries of the economies of scope asso ciated with the diversification of on farm production activities. Yet farm level innovation and AFN centered rural development strategies have largely escaped critical interrogation. Thus AFN case studies typically fail to analyze relations of power within the farm enterprise, whether in the labor process, the household, or gendered property structures. Similarly, although farm based multifunctional activities are seen as the engine of the 'new' rural development model, there has been little systematic evaluation of how this model will address entrenched rural problems of income equality, low paid employment, and rural poverty or reconfigure spatial processes of uneven development. By neglecting to examine relations of domination at these different scales, this literature leaves unanswered the question of 'who gains and who loses' from this market led, farm centric response to the quality 'turn'.

The erasure of relations of domination and, by ex tension, politics in analyses of economic localism is complemented by the limited attention given to the rent seeking behavior of downstream actors, notably retail multiples. In addressing this neglect of food supply chain relationships, a recent study analyzes the production of new quality food 'spaces' in the UK and the concomitant struggle to dominate the material and discursive con struction of quality. Competitive control of quality confers power to delineate markets and draw boundaries between retailer led commodity chains and AFNs. In the UK case, the entrenched regulatory hegemony of re tailer led food governance, sanctioned by the state, places AFNs at a tremendous disadvantage in this struggle. This analysis of power relations across the spaces of food provisioning will hopefully stimulate further research on this theme.

Other contributors have approached this question more generally by emphasizing the dynamic nature of economic rents and their competitive erosion by rent seekers. These pressures are exemplified by EU Regu lation 2081/92 regarding PDI/PDO designations, which has led to a profusion of quality assurance schemes, logos, and labels to valorize territorial identity and local prov enance. In one study, these differentiation strategies are described in terms of the 'qualification' of place, process, and product, or the three Ps. These new territorial constructions of quality are open to imitation and stra tegic convergence, and retail multiples have responded by developing own label, locally sourced product lines and quality food brands.

Such moves by leading retailers to dominate the emerging 'quality food spaces' suggest that 'main streaming' and brand proliferation may so dilute the meaning and imagery of 'local', 'organic', and 'specialty' foods that these terms will become meaningless as in dicators of 'alternative' food provisioning networks. These dynamic changes at the interface between 'con ventional' and 'alternative' provisioning clearly threaten the foundations of the territorial value added model of rural development insofar as these competitive pres sures erode price margins, and rents are captured by extralocal, downstream, supply chain actors. Several authors have identified the need for a better under standing of the dynamics of economic rents and their distribution as a major research gap in the AFN/SFSC literature.

This gap and the neglect of social relations, power dynamics, and the politics of the local arguably are symptomatic of a deeper theoretical problem. That is, the ontology of the local is taken as a given, and not as a category that is socially constructed, contingent and therefore in need of explication. AFN/SFSC scholarship thus does not problematize the local as 'place', as a scale of accumulation nor its genesis in dynamic and contin gent interactions with translocal actors and institutions. The static, bounded, and reductionist conceptualization of the local in this literature entirely ignores theoretical developments in human geography, notably relational perspectives of place, as in Doreen Massey's work, and analyses of the politics of scale. Close engagement with such work would expose the reification of the local and its naturalized social relations to much needed critical analysis.

The failure to acknowledge consumers as active partners in the emergence of AFNs is paralleled by the limited analysis of the social relations of consumption in both the US and Western Europe. With its markedly higher prices, quality food provisioning has been char acterized as a 'class diet' of privileged income groups, who can afford this form of personal 'quarantine' from conventional industrial foods. Thus the US natural foods chain, Whole Foods, is satirized as 'Whole Paycheck'. Several recent studies of public sector procurement policies, notably school meals, community food co operatives, and sustainable urban food systems, speak to this issue of democratic access.

Such initiatives, however, almost surely will be over taken by the accelerating incorporation of quality foods into product lines of mainstream retailers. The 'Wal Mart effect' on the conventional alternative provisioning interface represents the new frontier for the analysis of 'spaces of quality' and their alterity. This raises the interesting question of how mainstreaming and more democratic access will unsettle the eco social imaginaries and moral geographies of quality foods, perhaps by di luting the moral charge of connection as these com modities are popularized and no longer garner social distinction.

Globalized AFNs: Qualities, Ethics, and Mainstreaming

So called quality foods and foods invested with particular 'qualities' are driving the growth of globalized AFNs in forms of transnational organic food chains, the Ethical Trading Initiative (ETI), and fair trade networks. All three have – in one way or another – captured the geo graphical imaginations of activists and scholars, polit icians and businesses, and consumers and producers as pathways to ameliorate damage to the livelihoods of workers and peasant communities in these neoliberal 'times'.

Certified organic AFNs comprise the growing on and off season movement of tropical fresh and processed foods to the North from the South; one of the largest and fastest growing exporters is China. While research to date has been highly descriptive, excellent accounts of one such AFN exist in the particularly fine grained ethno graphic analyses of a set of Mexican organic coffee co operatives. Attention here is to the practice of the governance 'work' of organic certification and the po tential problems this writes into the eco social relations fostered amongst coffee, community members, certifiers, and international regulatory bodies. This research points to the paradox of the 'roll out' of certification schemes throughout poor farming communities in the South: greater development possibilities open up but at the expense of often unforeseen economic and social costs in internecine competition and conflicts among communities and their members. In short, certification for producer communities – in organic and fair trade networks – is not a neutral practice. A related lesson is that these political economies of certification can often only be gleaned through extensive ethnographic research.

The ETI, developed in the late 1990s and located in the UK, was formed from a coalition of supermarkets, transnational corporations, activists, and development nongovernmental organizations (NGOs). With its focus on the enforcement of international labor standards at the points of commodity production in the global South, research on the ETI has focused on food, flower, and clothing production. Some of the best work on ETI AFNs – again done at the ethnographic scale – has focused on the international trade of 'veg' from Africa to the UK. Contextualized against the canvas of wider food 'scares' from imported goods, one scholar conceptualizes the ETI as a form of 'soft' colonialism working to produce 'pure' African spaces through the regulation of nonexploitative labor and chemical free foods. Here, the unintended and non neutral consequences of the ETI certification have worked to marginalize some of the poorest farmers, alter community social relations through the creation of 'benign dictators' invested with certificatory power, and further entrench the economic power of supermarkets more than half a social, economic, and cultural world away.

Fair trade AFNs have quickly become an expanding 'sacred' space for activists, scholars (much beyond geography), and many multinational corporate entities (see **Figures 6–9**). Some of the earliest geographical work utilized fair trade coffee as a backdrop around which to wrap a series of conceptual and theoretical threads; here it was argued the 'nourishing networks' of fair trade coffee were better conceptualized through the Latourian metaphors of 'longer' versus 'shorter' networks of 'immutable mobiles'. This was put forward in an attempt to shed the essentialized scalar categorizations of the local, regional, and global in describing the 'mode of ordering of connectivity' in the producer–consumer connections of fair trade coffee AFNs. Unlike AFN scholarship, work (geographical and otherwise) following this initial beginning has been more conceptually and theoretically driven, and ethnographic research on fair trade networks has been relatively neglected.

Several lenses have been used to work through fair trade and its bid to 'guarantee a better deal for Third World producers'. An early developmental lens for fair trade has given way to wider discussions of how fair trade must constantly negotiate the contentions of the conventions it articulates: those of domesticity (e.g., trust) and civic norms (e.g., global citizenship) *vis à vis* those of commercial, industrial and market ness (e.g., price, quality, labeling, etc). This articulation is encapsulated in the rather empty twin platitudes – both of which have been uttered with little geographically minded acumen

Figure 6 The Beginnings of the Fair Trade Cacao Commodity Network: A Farmer and his 'Farm' in Costa Rica. Fair trade works through organized cooperatives of small farmers; this fair trade cacao farmer in the figure is a part of the cooperative known as the *Asociación de Pequeños Productores de Talamanca* (APPTA) on the southeastern coast of Costa Rica (http://www.appta.org/). Fair trade works to produce goods based as much as possible on sustainable agroecological principles. Thus, the farmer is standing in the midst of his inter-cropped pineapples, cacao, and other crops. At the same time, though, even with this image of a fair trade cacao pod the yellow fruit hanging on tree branches in the picture it 'hides' the different relationships of power, economy, and politics 'present' in this cacao that will eventually end up in a fairly traded chocolate bar.

Figure 7 Fair Trade Cacao in the Dominican Republic. These are processing facilities for CONACADO, the largest exporter of fair trade cacao in the world located in the Dominican Republic. One of the things to note here, besides the massive size of the operation, is that the processing facilities are more cooperatively run and owned than in the Costa Rican market where most of the early processing is done by the farmers. The image here is of one of six *bloques* of Dominican farmers and processing facilities scattered around the country.

Figure 8 Drying Cacao in the Dominican Republic. Note the denuded hills in the background; often fair trade farms and cooperatives become islands of green in the midst of environmentally-damaging 'business-as-usual' farming.

or verve – that fair trade operates 'in the market but against it' and/or that it is 'in the market but not of it'. Newer and more grounded work on the so called de velopment 'case' for fair trade has sought to explore the ability of these AFNs to build small producer capacities (at a number of levels) and bolster livelihood strategies within and beyond fair trade commodity chains. On the other hand, dissection of the governance structures of fair trade certification and regulation is beginning to reveal some problematic aspects of this certificatory regime in the growing inequalities of power between producer communities and regulators.

There is also a 'business' case being made for fair trade, particularly through the publication of the book *Fair Trade: Market led Ethical Consumption.* In simple terms, the arguments being made are that fair trade is simply another expression of consumer choice and is in fact a neoliberal solution to the problems of international trade and development. While the validity of these claims is somewhat questionable from a number of different perspectives, it is a rather politically savvy, 'realpolitik' line of argument that attempts to preemptively defend the fair trade model from the critiques and gerry mandering of more powerful so called 'free trade' advo cates. In many ways, it is these conceptualizations that have paved the road for fair trade to move smoothly into the mainstream of supermarkets and larger retail outfits, and become part of the established repertoire of some of the transnational giants of the conventional food system, such as Nestlé, McDonald's, and Starbucks.

The final lens discussed here is that of 'consumption': work has concentrated on conceptual development that entrains production and consumption and space and place in the exploration of fair trade AFNs. This has come at the expense of greater engagement with 'live' consumers. Rather, one strand of the discussion has

Figure 9 A Cacao Nursery Funded through the Fair Trade Premium. This cacao-plant nursery is owned and operated by CONACADO. In addition to more and better infrastructure for example, roads, home gardens, housing, etc. the nursery was financed through fair trade's 'social dividend' and is used to promote the economic and social development of its members.

focused on the use of the images and discourses of producers' 'em placed' livelihoods designed to link the spaces and places of fair trade production and consumption. It is argued that these 'political ecological imaginaries' have given poor farmers, fair trade companies, and NGOs the opportunities to 'seize the fetish' and rework it toward more advantageous Third World development. More recently, it has been suggested that the reworked fetish has damaging effects through an essentialization of the 'other' that maintains the deeper inequalities of representation and political economy that quite rightly still haunt the global economy; soft colonialism might just be rearing its ugly head in fair trade AFNs.

Another strand of engagement with fair trade's consumption scape explores the sort of ethical key to which fair trade is tuned. Conceptualized as a 'solidarity seeking' moral economy, it has been suggested that fair trade moves beyond the requirement that one must be socially and geographically 'near' to care about others. Instead, these AFNs shed the confines of 'place' to open up an 'ethics of care' that is similarly fostered across distant social, economic, and geographical 'spaces'. And, yet, counterarguments are emerging to suggest that, precisely because of the far flung distances between producers and consumers in the global economy, it is only through the images, realities, and connections of place that fair trade can and does work.

A number of further avenues are opening up for greater empirical and theoretical work on globalized AFNs. For instance, research is beginning to focus on the sort of Faustian bargain fair trade has made to enter the mainstream: on the one hand, markets have expanded rapidly and thus more money is being transferred from the well off North to poor farmers in the service of livelihood empowerment and economic development. On the other hand, this growing presence in conventional retail spaces has come at the price of a lower political register that is starting to see fair trade networks first and foremost as 'quality' products rather than as advancing a potentially transformative political project. More research needs to focus on the wide ranging effects of the mainstreaming of fair trade and other ethical goods from a number of geographically informed angles. Not least of all, researchers should engage with the growing bifurcation of the fair trade movement – similar to that in organic AFNs – along the faultlines of the so called 'movement' or '100%' fair traders versus those that advocate a more mainstreaming perspective.

See also: Agriculture, Sustainable; Agri-Environmentalism and Rural Change; Food Networks; Food Regimes.

Further Reading

Cook, I. *et al*. (2006). Geographies of food 1: Following. *Progress in Human Geography* 30(5), 655 666.

Goodman, D. (Guest ed.) (2002). *Special Issue: Rethinking Food Production Consumption: Integrative Perspectives. Sociologia Ruralis* 42(4), 271 277.

Goodman, D. (Guest ed.) (2003). *Special Issue: International Perspectives on Alternative Agro Food Networks: Quality, Embeddedness, Bio Politics. Journal of Rural Studies* 19(1).

Goodman, D. (2004). Rural Europe redux? Reflections on alternative agro food networks and paradigm change. *Sociologia Ruralis* 44(1), 3 16.

Goodman, D. and Watts, M. (eds.) (1997). *Globalising Food: Agrarian Questions and Global Restructuring*. London: Routledge.
Goodman, M. (2004). Reading fair trade: Political ecological imaginary and the moral economy of fair trade foods. *Political Geography* 23(7), 891 915.
Guthman, J. (2004). *Agrarian Dreams? The Paradox of Organic Farming in California*. Berkeley, CA: University of California Press.
Harvey, M., McMeekin, A. and Warde, A. (eds.) (2004). *Qualities of Food*. Manchester: Manchester University Press.
Marsden, T. and Murdoch, J. (eds.) (2006). *Between the Local and the Global: Confronting Complexity in the Contemporary Agri Food Sector*. Oxford: Elsevier.
Maye, D., Holloway, L. and Kneafsey, M. (eds.) (2007). *Constructing Alternative Food Geographies: Representation and Practice*. Oxford: Elsevier.
Morgan, K., Marsden, T. and Murdoch, J. (2006). *Worlds of Food: Place, Power, and Provenance in the Food Chain*. Oxford: Oxford University Press.
Mutersbaugh, T., Klooster, D., Renard, M. and Taylor, D. (Guest eds.) (2005). *Special Issue: Certifying Rural Spaces: Quality Certified Products and Rural Governance. Journal of Rural Studies* 21.
Nicholls, A. and Opal, C. (2005). *Fair Trade: Market Driven Ethical Consumption*. London: Sage.
Raynolds, L., Murray, D. and Wilkinson, J. (eds.) (2007). *Fair Trade: The Challenges of Transforming Globalization*. London: Routledge.
Van der Ploeg, J. (Guest ed.) (2000). *Special Issue: Rural Development: From Practices and Policies towards Theory. Sociologia Ruralis* 40(4).
Zaccai, E. (ed.) (2007). *Sustainable Consumption, Ecology and Fair Trade*. London: Routledge.

Relevant Websites

http://www.appta.org
Asociación de Pequeños Productores de Talamanca (APPTA).
http://www.sustainweb.org
Sustain: The Alliance for Better Food and Farming.
http://www.colostate.edu
The Center for Fair and Alternative Trade.
http://www.foodsecurity.org
The Community Food Security Coalition.
http://www.fairtrade institute.org
The Fair Trade Institute.
http://www.isealalliance.org
The International Social and Environmental Accreditation and Labelling (ISEAL) Alliance.
http://ofrf.org
The Organic Farming Research Foundation.
http://www.fairtrade.org.uk
The UK Fairtrade.
http://www.foodethicscouncil.org
The UK Food Ethics Council.
http://www.soilassociation.org
The UK Soil Association.
http://www.farma.org.uk
The UK National Farmers' Retail and Markets Association (FARMA).
http://www.ams.usda.gov
The USDA National Organic Program.

Food Regimes

B. Pritchard, University of Sydney, Sydney, NSW, Australia

Glossary

Doha Round A set of World Trade Organization (WTO) negotiations which commenced in 2001 with the purpose of developing a new body of international trade agreements.

Regulation Theory A theory to explain the emergence of extensive periods of economic and political stability. In regulation theory, capitalism was stabilized when a mode of regulation locked into a regime of accumulation.

Uruguay Round A series of negotiations which took place from 1986 to 2003 that established a new set of international trade agreements, including, in particular, the formation of the WTO and the Agreement on Agriculture.

World Trade Organization An international organization with responsibilities to administer international trade agreements.

World-System Theory An approach which seeks to understand and explain economic and political processes in their global and historical terms.

Introduction

The food regime concept was developed in the 1980s as a way of understanding the relationships between the es tablishment and evolution of nation states on the one hand, and the international political economy of food, on the other. Upon its seminal presentation in a 1989 article by Harriet Friedmann and Philip McMichael, it struck an immediate chord with a wide array of researchers seeking to identify the historicized logic to the complexities of agri food economics and politics. Moreover, coming at a time when international rules on agricultural production and trade were being incorporated into the multilateral trading system, it provided a mechanism by which re searchers might interpret these developments. It indi cated that contemporary negotiations over agriculture could be seen as components of a restructuring of the world food order in line with an epoch of crisis and (attempted) resolution.

In the years since its initial presentation, it needs to be said, the food regimes concept has come under con siderable challenge. In the mid 1990s, researchers influ enced by 'post stucturalism' suggested that the food regimes approach offered an overly mechanical inter pretation of the history of food. Allegedly, the macro scale orientation of the approach gave insufficient acknow ledgment to local and regional diversity. Alternative perspectives oriented with an eye to emphasizing diversity were presented most comprehensively by Michael Watts and David Goodman in their 1997 edited collection *Globalizing Food*. Notwithstanding the debates over these issues, the food regimes concept remained an attractive and widely used method for investigating the global pol itics of food. In the first decade of the twenty first century, moreover, it has garnered new found credentials by way of providing a framework to explain global scale recon figurations of the world food order centered on private regulation of food trade (notably by transnational super markets) and the appropriation of green credentialing by these same actors. A prescient article by Harriet Friedmann in 2005 poses the suggestion that the instabilities of the World Trade Organization (WTO) regime during its first decade of existence (notably, fo cused around intractable disputation on agricultural trade) reflect the collapsing fragments of the second food regime, with emergent 'corporate environmental' trade systems as the formative basis of a new regime in processes of construction.

Origins and Significance

The importance of the food regimes concept lies in its capacity to provide a holistic representation of the international agri food system without resorting to market determinism. It argues that the struggles faced by actors within agriculture and food systems can be understood through recourse to a world historical per spective that positions such processes within global po litical epochs. This approach stands in stark contrast to the orthodox approach, dominant within most branches of economics, which treats politics as an exogenous variable within market relations.

This formulation was built from the merger of two broad approaches in international political economy. The first of these was the world system theory approach pioneered by Immanuel Wallerstein in the 1970s. The world system approach seeks to understand and explain economic and political processes in their global and historical terms. This gives emphasis to, and makes ex plicit, the world historical conditions that underpin the individual circumstances of economic and political actors. In other words, the multifaceted struggles con fronting actors in the world economy are seen as de riving, ultimately, from a system wide logic of global capitalist processes. The second broad approach influencing food regimes concept is regulation theory. Developed initially by French researchers interested in charting the institutional, political, economic, and social factors that underwrite capitalist development, regulation theory posed a means to explain the emergence of extensive periods of economic and political stability. In regulation theory, capitalism is seen to be stabilized when a mode of regulation (regulatory practices and settings) coincides with a regime of accumulation (the conditions for profit making).

Friedmann and McMichael welded together these two approaches in their seminal articulation of the concept. Their objective was to develop a historicized appreciation of how and why global systems of food provisioning connected to world political systems. Arguing that for considerable periods of history, there is relative stability in the characteristics of large scale food production and consumption and their relation to the state system, the approach connected a series of disparate arguments about national and international transformations in food and agriculture, to a historically anchored framework based around the metatrends of international political economy. Hence, as suggested by Farshad Araghi in his 2003 article published in the *Journal of Peasant Studies*, the food regimes approach is ultimately concerned with the economics of imperialism and power in world politics.

Friedmann and McMichael identified two food regimes over the preceding century. The 'first food regime', which Friedmann and McMichael date from 1870 to 1914, originated in the food politics of Europe's urban working classes and coincided with British Imperial hegemony. It was characterized by enhanced levels of commodity production as the colonial 'frontier' was opened up and the demand for food and fiber increased in the burgeoning economies of a newly industrializing Europe and North America. In the process, an international division of labor was established, based on the production of tropical (cocoa, palm oil, sugar, coffee, and tea) and temperate (beef, grains, and dairy) commodities in the colonial territories of the European powers, and manufactured goods within the industrialized centers. In the United States, the same process was underway as the 'moving frontier' expanded and facilitated extensive ranching and grain farming, and led to the increased production of agricultural products to feed the swelling populations of the industrializing Eastern seaboard, continually being replenished by mass migration from Europe. Effectively, therefore, this process amounted to relocation of the British (and to a lesser extent, European) food supply to colonial territories, altogether underpinned by a 'free trade' rhetoric anchored by the gold standard. On this basis, the first food regime has also been dubbed the 'colonial diasporic' food regime. The class relations which underpinned this regime were the urban working populations furnished by cheap food and the settler farmers who supplied this. It depended also on the environmental bounty coming from the farming of lands previously nonsubjected to Western farming practices.

After World War I, the indebtedness of the colonial powers and the economic instability associated with the Great Depression resulted in these colonial relations losing momentum as a driving force of global capital accumulation in agriculture. The food system, correspondingly, was characterized by an unfolding of tensions in the inter war period, of which the 'dust bowl' crisis of the US Midwest represents the most widely represented illustration. Coming out of this crisis, governments framed a set of policies to rein in agricultural market volatility. This included the Roosevelt New Deal, which established the basis of US family farming anchored within state supports.

Connecting these developments to the rise of a new world political order, Friedmann and McMichael date the emergence of a stabilized 'second food regime' with the ascendance of US political, economic, and military might in 1945. Two conjoined political–economic processes underpinned the establishment of a stabilized food order in this period. First, production supports and price stability systems furnished originally as part of President Roosevelt's New Deal underwrote an expansion of the soybean hogs and grains livestock complexes of Midwestern and Prairie agriculture. Second, this regime of domestic interventions sat hand in glove with an assertion of US political hegemony, buttressed by the production of large agricultural surpluses that were incorporated into the Marshall Plan and disbursed globally through food aid programs. The decision by the US Congress in 1947 not to ratify the proposed International Trade Organisation kept agriculture out of the multilateral trading system, and hence made domestic US farm policies largely immune from the pressures of international trade liberalization which, by the 1970s and 1980s, impacted heavily on the manufacturing sector. Consequently, over the 1950s and 1960s, American agriculture (protected at home and supported abroad) emerged as the most vital force in the global food economy. Unlike the first food regime, which was framed around the growth of trade, in the period of the second food regime, trade was politically constructed around the provisioning of food aid, and the subsidization of commodity exports to serve domestic political agendas.

These arrangements were supported by the insertion of Fordist modes of production and consumption within food and agriculture. The second food regime was an 'intensive' food regime characterized by the mass production of significant transformed (durable) agro food products for mass consumption. Increases in production were generated from the development of new hybrid seeds, agricultural technologies, and mass food production systems. (This is in opposition to the first food regime which can be characterized as 'extensive' because growth in output was predicated on bringing new areas of land into the global production system.) Rapidly rising real incomes were matched by a rapid increase in the production and consumption of 'Fordist' food, typified in the popular imagination by Andy Warhol's Campbell Soup cans. The emphasis was on the mass production of standardized products, increasingly organized by large scale agri food companies and marketed through channels of mass distribution, such as supermarkets and fast food chains. The Fordist logic of the second food regime, however, came together most spectacularly with the feed livestock complex, which girded a global set of interconnections based around the production and export of (US subsidized) feed; its supply to cattle producing regions (notably Latin America), and the production of low cost meat for mass markets. An important article by Steven Sanderson in 1986 coined the phrase 'the world steer' and brought together these ideas for the first time. It had an influential role within Friedmann and McMichael's crystallization of the food regime idea a few years later.

The tensions and contradictions of the second food regime came into the open in the 1980s in the form of trade wars between the United States and Europe, a prolonged crisis in US agriculture, and initiatives to regulate these systems internationally via the General Agreement on Tariffs and Trade (GATT). The institutionalization of the European Common Agricultural Policy (CAP) generated significant agricultural surpluses in Europe, challenging US hegemony as the global provider of food. At the same time, the costs of agricultural subsidization were made increasingly visible to national constituencies thus providing the basis for attacks on the status quo from neoliberal economic groups and parties with vested interests in reducing global agricultural subsidies, such as Australia. The decision in 1986 to commence a new round of trade negotiations (the Uruguay Round), with the explicit objective of incorporating agriculture into the multilateral trade system, represented an agenda to resolve these tensions through international political means. Ultimately, this gave rise to an emergent 'global' regulatory architecture for governing the politics of food. The conclusion of the Uruguay Round in 1994 implemented the establishment of the WTO as a permanent institution to promote international accord on the rules of trade, to encourage trade liberalization, and to arbitrate trade disputes. Through this, key elements in the national regulation of food and agriculture were subsumed to global institutional parameters. For food regime researchers since that time, the vital issue on which much analysis has turned has related to the question of whether these institutional transformations can be categorized in terms of a supposedly emergent 'third food regime', or whether they reflect inchoate elements of transition that are yet to take full form.

A Third Food Regime?

In various contributions during the period of GATT/WTO negotiations in the first half of the 1990s, Friedmann and McMichael were careful to interpret these developments in terms of 'the contours' of a potential new food regime. While asserting how these processes were reflective of the unravelling of the second food regime, their analysis suggested that the jury was out, so to speak, on that regime's successor.

Other analysts, however, were quicker to interpret agriculture's inclusion within multilateral trade rules as evidence of an emergent third food regime. One set of ideas, explored in Richard Le Heron's 1993 book *Globalized Agriculture, Political Choice* foresaw the incipient third food regime as an 'integrated regime' of global agri food production, characterized by flexibility, specialization, and globally footloose production trade complexes. As outlined by Le Heron, this regime would recompose national centered food complexes within the logic of international trade regulation. Consistent with these trajectories, moreover, would be enhanced freedoms for transnational agri food companies undertaking production and intrafirm trade within a global 'level playing field'. Although expressed as an ideal type, inspiration for this model seems to have come from the development of the intercontinental, counterseasonal, fresh fruit and vegetables trade, which grew dramatically in the early 1990s. In an influential 1995 article, Richard Le Heron and Michael Roche applied these ideas to the 'clean and green' agri food exports of New Zealand.

Clearly, the fresh fruit and vegetables sector is reflective of powerful new impulses concerning the organization and regulation of the global food system. Nevertheless, it remains a commodity specific model at variance to the way that much of the rest of the global food system is organized. The neoliberal 'New Zealand experiment' remains an outlier in the regulation of agri food systems. Self evidently, the concept of an emergent third food regime implies there is a hegemonic political order which is underwriting a particular system of agri food production and trade. Equally self evidently, the pioneering New Zealand neoliberal model of the mid 1990s does not fulfil this requirement. Bill Pritchard's research on the international restructuring of the dairy sector highlights the continuing role of state centered intervention which complicates attempts to generalize upon the contemporary character of agri food globalization. His later work with Niels Fold, brought together in the 2005 edited collection *Cross continental Food Chains*, asserts the sector specific and territorially embedded conditions which need to be acknowledged in discussion of these issues.

In seeking to distill the key elements of the global politics of food, and by extension positing the vital signs of any incipient 'third food regime', from the late 1990s McMichael turned his attention to how subsidized US agricultural exports were undermining the domestic food economies of developing countries. This perspective drew into focus the basic contradiction of the WTO in its treatment of agriculture: firstly, that it enshrined and accepted the rights of developed Northern economies to continue to protect their domestic agricultures, and secondly, it imposed on developing countries mandatory economy opening regulations that subjected them to subsidized agri exports from the North. Focusing on the first of these contradictions, the WTO has belied implied suggestions in some parts of the food regimes literature of the early 1990s that the WTO would instil a free market in world agri food trade. While some researchers had looked toward the conclusion of the Uruguay Round of GATT and the formation of the WTO as a defining moment of transformation in world agriculture, the 'on the ground' impacts of these developments were far from momentous. Indeed, the first decade of the WTO seemed to perpetuate, in different institutional guise, the con tradictions of the global food order of the 1980s. How ever, in the case of agriculture and the WTO, levels of protective support within Northern countries remain little changed from those at the commencement of the Uruguay Round in 1986. What has happened is that agricultural policy makers in Europe, North America, and Japan have been adept in finding 'wriggle room' that enables continued support for farmers yet remains legally consistent with the complexities of international law as prescribed in WTO agreements. As such, the WTO Agreement on Agriculture (AoA) broadly sanctioned the continuation of agricultural support policies, albeit in restructured formats to accommodate WTO provisions.

For developing countries, in contrast, the WTO system has codified and delimited their scope for man euver. The WTO framework has encouraged developing countries to extensively restructure their agricultural sectors in line with agendas to open domestic economies, liberalize land laws (thus facilitating large scale agri culture), and to ease restrictions on foreign investment. McMichael phrased this state of affairs as producing a situation where 'comparative advantage becomes com parative access to subsidies'. The forced agricultural liberalization for developing countries combined with sanctions to the North has imposed a progressively heavy burden on global development over the past decade. One key focus of McMichael's work on these issues has been the 'east Asian import complex', with particular reference to the US–Japan agricultural agreements opening Japan's beef and citrus sectors to US exports, and the role of US maize exports in undercutting and destabilizing the do mestic Filipino food system.

Certainly the food and agricultural situation facing developing countries in the WTO era has assumed pressing global significance. By the mid 1990s half of the foreign exchange of the FAO's 88 low income food deficit countries went to food imports (FAO – Food and Agri culture Organization). Subsidized Northern agricultural exports are having considerable displacement effects on farmers, from Mexico to Malawi. Faced within an in creasing inability to secure domestic food security through local cultivation of food staples, (instead these are imported), developing countries have turned to high value agri exports. The sale of these products to wealthy consumers in first world nations underpins export earn ings that are used, to some extent at least, in purchasing the imports required to ensure local food security. Interestingly, McMichael's attention to these issues was not explicitly couched in the framework of a third food regime. In a chapter published in 1999, McMichael commented that the food regimes concept was a 'child of its time', meaning that it spoke to the emergent tensions in world food politics of the late 1980s, as the post 1945 order fell apart. Implicitly, what seemed to be apparent in McMichael's analysis of the injustices of the WTO food system was that it remained part and parcel of a de stabilized interregnum. His recent research on new social movements, in particular *La Via Campesina*, engages with the question of the 'politics of resolution' coming out of this period.

Friedmann has taken the furthest steps toward pro posing the elements of a prospective third food regime with her theorization of the lineaments of a corporate environmental food regime. Building on broadly con temporaneous work by David Burch and Geoffrey Lawrence, which sought to connect the restructuring of supermarket supply systems to the food regime concept, the central feature of Friedmann's recent work is what she sees as a convergence of environmental politics and retail led reorganization of food chains. In the shadow of intractability of WTO negotiations on agriculture, key elements of the regulatory scene have shifted to private companies. In their quest to control key aspects of food safety risk and to present a green face to their consumers, transnational supermarket chains are now establishing grades and standards which often exceed internationally sanctioned requirements. As Friedmann suggests, this is cultivating a new set of class relations within the global food system; between affluent consumers (shopping in supermarkets) who have access to clean and green food; and those materially excluded from such possibilities. In a word, state anchored international and national institutions have been outflanked by private capital, creating the basis for an emergent new food regime.

Considering these issues in global context, Friedmann asserts that the fundamental cleavage in world trade talks has shifted from the US–EU trade impasse to a wider set of differences between Northern and Southern countries. This is occurring in the backdrop to declining US global influence, particularly with the ascent of China within the WTO and the economic rise of India. Together with Brazil and South Africa, these countries displayed their power for the first time comprising a formidable negotiating bloc that derailed the WTO Ministerial Conference in Cancun, Mexico, in 2003.

Conclusion

Friedmann's 2005 research provides the most com prehensive and compelling application of the food re gimes concept to the conditions of the present time. Nevertheless, she is careful to acknowledge that the emergent corporate environmental food regime will have its countercurrents. The essential feature of the food regimes approach is that it is best used as a tool of hindsight. It can help order and organize the messy reality of contemporary global food politics, but its ap plications are necessarily contingent upon an unfolding and unknowable future.

However the current age evolves, the food regimes concept will provide a conceptual architecture to shape interpretations of change. Since the late 1980s, it has opened up new ways of thinking about the relationships between food, politics, and society. Used by geographers and other social scientists alike, it is one of the most influential conceptual innovations in the field of agri food studies over the past three decades.

See also: Dependency; Food Networks; Green Revolution; Neoliberal Economic Strategies.

Further Reading

Araghi, F. (2003). Food regimes and the production of value: Some methodological issues. *Journal of Peasant Studies* 30(2), 41 70.

Burch, D. and Lawrence, G. (2005). Supermarket own brands, supply chains and the transformation of the agri food system. *International Journal of Sociology of Agriculture and Food* 13(1), 1 18.

Fold, N. and Pritchard, B. (eds.) (2005). *Cross Continental Food Chains*. London: Routledge.

Friedmann, H. (1993). The political economy of food: A global crisis. *New Left Review* 197, 28 59.

Friedmann, H. (1994). Distance and durability: Shaky foundations of the world food economy. In McMichael, P. (ed.) *The Global Restructuring of Agro Food Systems*, pp 258 276. Ithaca: Cornell University Press.

Friedmann, H. (2005). From colonialism to green capitalism: social movements and emergence of food regimes. In Buttel, F. H. & McMichel, P. (eds.) *New Directions in the Sociology of Global Development: Research in Rural Sociology and Development* Vol. 11, pp 227 264. Oxford: Elsevier.

Friedmann, H. and McMichael, P. (1989). Agriculture and the state system: The rise and decline of national agricultures, 1870 to the present. *Sociologia Ruralis* 29, 93 117.

Le Heron, R. (1993). *Globalized Agriculture: Political Choice*. Oxford: Pergamon.

Le Heron, R. and Roche, M. (1995). A "fresh" place in food's space. *Area* 27(1), 23 33.

McMichael, P. (1994). Introduction: agro food system restructuring unity in diversity. In McMichael, P. (ed.) *The Global Restructuring of Agro Food Systems*, pp 1 18. Ithaca: Cornell University Press.

McMichael, P. (1996). Globalization: Myths and realities. *Rural Sociology* 61(1), 25 55.

McMichael, P. (1998). Global food politics. *Monthly Review* 50(3), 97 122.

McMichael, P. (1999). Virtual capitalism and agri food restructuring. In Burch, D., Goss, J. & Lawrence, G. (eds.) *Restructuring Global and Regional Agricultures: Transformations in Australasian Agri Food Economies and Spaces*, pp 3 22. Aldershot: Ashgate.

McMichael, P. (2000). A global interpretation of the rise of the east asian food import complex. *World Development* 28(3), 409 424.

Pritchard, B. (1996). Shifts in food regimes and the place of producer co operatives: Insights from the Australian and United States dairy industries. *Environment and Planning A* 28, 857 875.

Pritchard, B. (1998). The emergent contours of the third food regime: Evidence from Australian dairy and wheat sectors. *Economic Geography* 74(1), 64 74.

Sanderson, S. (1986). The emergence of the world steer: internationalisation and foreign domination in Latin American cattle production. In Tullis, F. L. & Hollist, W. L. (eds.) *Food, the State and International Political Economy*, pp 123 148. Lincoln, NB: University of Nebraska Press.

Fordism

R. Hudson, Durham University, Durham, UK

Glossary

Division of Labor This term describes the ways in which work is divided between people. It can refer to the organization of work in individual workplaces (technical division of labor), between companies (social division of labor), or between places (spatial division of labor). Fordism as a particular way of organizing the economy has its own distinctive technical, social, and spatial divisions of labor.

Limits to Growth This concept reflects the fears that some key natural resources (notably hydrocarbons and metallic minerals) would become exhausted and pose 'limits to growth'. Such resources undoubtedly are absolutely limited in supply. However, a variety of technical and social innovations have, for the moment, kept this issue relatively low on the agenda.

Mass Production and Consumption Mass production and consumption are central to Fordism. Mass production leads to the lowering of average production costs via achieving economies of scale – the reduction in the average cost of production as a result of progressive reductions in marginal production costs as production volumes increase. Mass consumption is the necessary corollary of mass production. The sale of commodities such as automobiles, washing machines, and shirts enables companies to exchange the things that they have produced for money, which in turn enables them to buy raw materials, components, and labor and produce further commodities.

NIMBY ('Not in My Back Yard') A term that denotes opposition to locating unwanted activities such as polluting industries or waste disposal facilities in particular locations. However, not all people have equal capacity to mount such opposition and in some parts of the world people compete to attract such activities as a source of jobs and incomes. As a result, the geography of such activities is uneven at various spatial scales.

Regulationist Approaches Regulationist approaches, first developed by French political economists such as Michel Aglietta, Robert Boyer, and Alain Lipietz, strive to understand how the inherently crisis-prone capitalist economy manages to avoid slipping into serious crises. A regime of accumulation refers to a relatively stable aggregate relationship between production and consumption. Mode of regulation denotes the regulatory mechanisms within the state and in civil society, the body of beliefs, habits, laws, and norms consistent with and supportive of a regime of accumulation. Regulationist approaches emphasize the correspondence within a given mode of development between regimes of accumulation and modes of regulation necessary for social systemic stability. Fordism represents one such coupling, typically seen as dominant for much of the twentieth century.

Scientific Management Closely associated with the work of F. W. Taylor (hence 'Taylorism'), this refers to the application of time-and-motion studies to the organization of work in the workplace so as to maximize labor productivity and minimize the amount of time that workers are not actually working while at work. Henry Ford built upon this in producing automobiles by introducing the moving production line and controlling the speed of work via the speed of the line.

Suburbanization The process through which the consumption and residential landscapes of Fordism were altered via the spatial extension of cities and towns, initially in the United States in the 1920s and 1930s. This process was crucial in creating markets for the new mass-produced consumer goods central to Fordist development.

Third World A term coined in the 1960s to denote former colonies that achieved political independence in the 1960s, *in contra* to the more advanced capitalist development of the First World and the state socialist Second World. Subsequently some Third World states attracted industrial investment from the Fordist core of the First World and became seen as 'newly industrializing countries' and as the sites of peripheral Fordism, though many were excluded from these processes as uneven development became more sharply etched at the global scale.

Introduction

The term 'Fordism' has come to have many meanings. It originated as a description of the radical new way of organizing the mass production of automobiles that was developed by Henry Ford in the early years of the twentieth century. From there it diffused to a great variety of other places and industries. Later it became used to describe a particular macro scale model of the organization and linking of mass production and consumption, often seen as the canonical model of capitalist development in the second half of the twentieth century. At both micro and macro scales, Fordism was seen to have distinctive geographies, of production, exchange, and consumption. If the twentieth century can be characterized as that of Fordism, it is also the case that by the end of that century there was a lively debate as to whether Fordism had run its course and was being replaced by some sort of 'post Fordist' successor. As of now that debate remains to be settled one way or the other but the evidence of continuing mass production and consumption, and indeed its expansion into the new spaces of the emerging economies such as China and India is difficult to discount.

Fordism and the Organization of Work in the Workplace

Fordism as a specific form of micro scale organization of mass production first emerged in the US in the early years of the nineteenth century at Henry Ford's automobile plant at Highland Park. While undoubtedly a revolutionary shift in methods of manufacturing, Ford's mass production model built upon previous advances in methods of manufacturing, in particular Taylor's work on scientific management, a detailed technical division of labor in the production process, and the precise measurement of the time needed for workers to undertake a particular task on a production line. Taylorism in turn built upon the emergence of the American System of Manufacture specifically as a way of wresting control over the production process from craft workers, based on the use of specialist machines to produce interchangeable parts. This was a key development in manufacturing, with profound implications for workers and the organization of work but the full implications of this development for deepening the technical division of labor and organization of the labor process were not developed until Taylor's work on scientific management and Ford's revolutionary mass production automobile plant with its mechanised assembly lines.

Mass production is based around selling low cost standardized products in large homogeneous markets. When initially introduced into automobile production by Ford, it led to dramatic increases in labor productivity. The labor time needed to produce a Model T Ford fell from 12 h and 8 min in October 1913 to 1 h and 30 min 6 months later. Mass production thus conferred great competitive advantage but required workers to be organized and to work in specific ways, based around a separation of mental from manual work, extreme specialization of tasks, and a deep technical division of labor within the workplace. Workers typically perform simple, repetitive de skilled tasks with very short job task cycles (often defined in seconds) on the production line. The moving line delivers materials to them at speeds determined by management. Increases in labor productivity are achieved via increasing the line speed because of managerial decisions. As such, alienated workers perform repetitive routine tasks on an uninterrupted basis at a pace dictated by the speed of the line. Shift systems allow maximum utilization of machines. There are well known problems associated with Fordist mass production as a result of both labor and product market changes. The initial challenges arose due to increasing resistance by workers in the 'full employment' conditions of the core urban and regional labor markets that emerged in the advanced capitalist countries in the 1960s as tensions emerged in the macro scale Fordist mode of regulation (see below). There was increasing resistance to speedup and intensification of work, leading to industrial disputes and disruptive official and unofficial 'wild cat' strikes. One neo Fordist managerial response to this was to increase automation within mass production factories (e.g., by using robots in production) to alleviate problems of labor control.

Some commentators have proclaimed the demise of Fordist mass production in many branches of manufacturing and the emergence of a post Fordist production system but others point to the extension of the principles of Fordist production into activities such as agriculture and mining and to an expansion of both 'downgraded manufacturing' and 'downgraded services' as the principles of Fordist mass production have been extended to encompass more occupations and locations (and the latter point is discussed more fully below). Principles derived from Fordist approaches to production have been translated into, and to a degree, reworked and extended within a diverse range of routine service activities, including the fast food sector and call centers and other activities that involve processing large amounts of data and/or paper or dealing with customer enquiries.

Fordism as a Regime of Accumulation and Mode of Regulation: Balancing Mass Production with Mass Consumption and Mass Destruction

As well as being seen as a micro scale approach to organizing work and mass production in the workplace, Fordism has also come to be seen in macro scale terms, as a particular way of regulating economy and society and normalizing particular conceptions of social and economic relationships. Indeed, it would be fair to say that Fordism is generally acknowledged as the defining mode of capitalist organization in the second half of the twentieth century, despite claims as to its demise and its replacement by 'post Fordism'. This macro scale conception of Fordism is particularly linked to the rise of the French regulation school, while recognizing variations in emphasis within this approach. Such variations notwithstanding, the shared central concern of the regulationist approach is to understand how the inherently unstable and crisis prone capitalist economy manages, improbable though it may seem, for most of the time to avoid serious crises. From a regulationist perspective, a regime of accumulation refers to a relatively stable aggregate relationship between production and consumption – Fordism represents one such coupling. In relation to understanding Fordism in particular, the significance of the national state in balancing increases in productivity and consumption within the national territory is perhaps the primary insight of the regulationists. They place particular emphasis upon the role of the state in equilibrating productivity increases with increases in consumption levels and patterns. Achieving and maintaining this temporal stability implies a definite degree of correspondence between changes in consumption and the lifestyles and living conditions of waged workers and their families and changes in production. More specifically, the rising labor productivity and the expansion of output that is central to Fordist mass production were related to the emergence of new forms of housing and the growing ownership of various 'white goods', consumer electronics and, above all, of automobiles, the symbolic commodity of Fordist mass consumption (see below). In addition, however, the emergence of the advertising industry – "the magic system," as Williams (1980) famously described it – also played a crucial role in changing patterns and norms of consumption and persuading people to purchase the new consumer goods. Mode of regulation refers to the regulatory mechanisms within the state and in civil society, the body of beliefs, habits, laws, and norms consistent with and supportive of a regime of accumulation and which, in fact, make it possible. Regulationists advocate an approach which privileges neither production nor reproduction but which emphasizes the correspondence within a given mode of development between regimes of accumulation and modes of regulation necessary for social systemic stability.

In regulationist terms, then, Fordism is a regime of accumulation in which the productivity gains resulting from the widespread diffusion of mass production methods are balanced by the growth of mass consumption and rising consumption norms as the national state sought agreements over the rate and composition of economic growth and the distribution of the benefits of such growth with the major collective interests of capital (such as national employers' associations) and organized labor (notably trades unions). By seeking to ensure that national economic growth rates are both sufficient and sustainable (economically, politically, and socially – though significantly not ecologically: see below), with the national territory appropriately located within the overall accumulation process, the national state can fund its own activities and avoid the dangers of fiscal crisis. By seeking to guarantee that changes in conditions of production and consumption are mutually consistent and compatible (via taxation and income redistribution policies, for example) and that the mode of regulation is itself successfully reproduced (e.g., through educational and health policies), the national state seeks to ensure smooth and sustained economic growth. While increasing mass consumption was a central element in sustaining Fordism as a regime of accumulation, it is also important to emphasize the role of the production of armaments, weaponry, and the means of destruction as central to the Fordist regime, exemplified most sharply in the military industrial complex of the US, with an estimated 30% of the US economy accounted for in this way.

While at one level of abstraction it is possible to speak of a canonical national Fordist regime of accumulation and mode of regulation, in practice 'actually existing Fordisms' exhibit variations around this canonical model. This is partly a reflection of the fact that regimes of accumulation and modes of regulation do not simply exist, preformed and awaiting discovery by national states. Moreover, Fordism became established in different national political and cultural contexts, which shaped the particular ways in which distinctive national Fordisms emerged and developed. There are, therefore, national variations around the canonical generic Fordist regime of accumulation. The contrast between the UK and US, West Germany and Sweden, and Japan in the 1980s provides a striking example of the different ways in which national state strategies can develop in the context of the same global economic and political environment. The rise of Fordism to dominance in the US was central to both the transformation of the sociospatial structure of the US and the rise of the US as the globally dominant capitalist power in the twentieth century. The rise of China in the first decade of the twenty first century provides evidence of the continuation of Fordist mass production in new sociopolitical contexts and the continuing influence of the Chinese Communist Party in an emerging capitalist economy. This particular variant of mass production and consumption is having dramatic effects, at a variety of spatial scales and registers. It is, *inter alia*, leading to China becoming the major center of production in the global economy for a wide range of goods, especially consumer goods. This in turn is linked to major changes in the international division of labor and patterns of international trade. It is also having fundamentally transformative effects within China: for example, regional inequalities have sharpened enormously as the new export oriented industries are concentrated on the eastern coast, leading to internal migration on an unprecedented scale and massive urban expansion. Industrial expansion coupled with the rise of a mass consuming middle class is in turn leading to massive environmental transformation, devastation of the environment, and serious problems of environmental pollution. The coexistence of these national variants was and is permitted by the international monetary system, which left scope for national autonomy in selecting the precise form of Fordist mode of regulation. It is, however, also important to note that some of the most successful – in terms of conventional indicators of growth such as GDP per caput – economies of the latter part of the twentieth century certainly involved mass production and consumption but cannot plausibly be conceptualized as Fordist, such as that of South Korea.

Fordist Spatial Divisions of Labor

Initially, companies using Fordist mass production methods were located in the cities and major urban areas within the US and later in Western Europe that provided both masses of cheap labor and markets for the new mass produced consumer goods. However, by the latter half of the 1960s such locations were becoming increasingly problematic as sites of production as 'full employment' conditions led to growing labor resistance to the discip lines of Fordist production and the physical and mental stress resulting from demands for constant increases in lines speeds and in labor productivity, which are crucial to corporate competitiveness.

In response to this, and enabled by changes in both production technologies and transport and communi cation technologies, big companies using Fordist mass production strategies increasingly began to restructure the social and spatial organization of production. The growing size of corporations enabled them to split up production processes, functionally and spatially, while technological changes allowed de skilling of jobs. Con sequently, companies could locate specific parts of the overall production process in spaces with 'appropriate' labor market conditions, and in particular shift de skilled work to areas characterized by large masses of people prepared to undertake such work, usually for lower wages. Since workers require little training for routine production, companies can pursue aggressive 'hire and fire' tactics, often with rapid labor turnover as retention is not an issue in spaces in which there is an abundant supply of people looking for work. In contrast, high level managerial and control functions, R&D, and skilled manufacturing activities tend to be retained in their original locations, although there is now evidence of changes in the locations of some of these more highly qualified 'white collar' activities to emerging economies with cheap but highly skilled labor such as India. Geographies of Fordist mass production have thus been recast at various scales, stretching the social relations of production over successively greater distances. While initially these new geographies were recast within na tional boundaries, increasingly they are now recast as parts of new international divisions of labor in what is typically represented as a globalizing economy as com panies have increasingly become multinational or trans national in their scope and mode of operations. Indeed, the shift of routine assembly and component production to peripheral locations leads some commentators to speak of 'peripheral Fordism'. Such 'spatial fixes' thus enable Fordist manufacturing to be preserved by con taining its inherent contradictions within new intra national and international divisions of labor – for a time at least.

Fordism and Fordist Geographies of Exchange and Consumption

The emergence of mass consumption and new con sumption norms for the relatively affluent middle classes and fractions of the working class was associated with the construction of new spaces of exchange, sale, and con sumption. Just as Fordist geographies of production re quired new and changing geographies, so too did those of exchange, sale, and consumption of commodities.

The development of Fordist mass production was linked with a series of innovations in the ways in which retailing and spaces of exchange were organised. There is an identifiable historical geography to their evolution – Miller *et al.* (1998: 3) describe this in terms of "an arc from the arcades and department stores of Paris through to the shopping malls of the United States." While de partment stores emerged in the central areas of Paris, London, and New York in the nineteenth century, they both spread spatially and proliferated in number in North America and Western Europe as Fordist mass production took off in the early twentieth century. For Nava (1997: 56), they constitute one of the "key iconic aspects of modern urban society," targeted deliberately at female shoppers, fusing aspects of entertainment and tourism with 'hard sell', as geographies of retailing and sale were recast as an integral part of the expansion of Fordism.

The emergence of the shopping mall represented a further step change in the organization of spaces of sale. While planned shopping centers first emerged in the US in the early twentieth century, shopping malls as now commonly understood first appeared in 1956 with the opening of the Southdale Mall in Minneapolis. This was to be the first of many, all over the world and increasing in size (with the emergence of 'megamalls' in the 1980s). Once again, mall developments were associated with significant changes in geographies of retailing. On the one hand, these were typically located out of town, requiring access by automobile, while, on the other, their internal spatial arrangements were designed to increase sales by manipulating the movements of shoppers within the mall.

While related, retailing geographies of sale and geographies of consumption within Fordism are not synonymous. The house – or home – is a key site of consumption and one of the distinctive aspects of Fordism is the way in which new residential areas – above all the suburbs – developed from the 1920s and 1930s, initially in the US, as locations for the affluent middle class and, later, for fractions of the working class. Suburbanization involved creating new residential living spaces, containing a self service society dependent upon household appliances produced and purchased as commodities. Magazine advertisements, promoting the growing diffusion and adoption of 'white goods' – vacuum cleaners, refrigerators, and other household appliances – were central to the construction of this new consumption space. The sale of these commodities was promoted in a very gender specific way – allied to allusions (or maybe illusions?) of the rationality of science and a modernist belief in scientific progress. This was projected as providing the basis for new forms of household management, replacing domestic labor by new consumer durables, outputs of the burgeoning new Fordist mass production 'growth industries'. Purchasing and using such devices would, allegedly, lead to cleaner homes, reducing the risks of ill health because of improved standards of food hygiene.

This emphasis on the consumption of new mass produced household goods amplified a tendency that had already emerged in the US in the late nineteenth century increasingly to commodify the home, replacing unpaid domestic labor with machines. By the end of the nineteenth century, food, shelter, clothing, and home furnishings had all become commodities. This was a process that was subsequently to expand greatly in terms of both scale and spatial reach. Referring specifically to furnishings and the ways in which these artefacts are deployed in creating particular environments and spaces, such as living rooms and drawing rooms, Sack suggests that, as consumption in the home involves both material and symbolic aspects, then each time a product is brought into the home, that home is transformed physically, as something is there which was not there before, and transformed in terms of values and meaning. The point can be generalized to all the other rooms and spaces within the home. Such rooms are then both material and discursive constructions, embodying and expressing the identities and values of the people who make and perform them, as homemaking has become very closely related to identity formation, and once again the activities of the advertising industry are critical in mediating relations between producers and consumers and creating markets for new commodities.

Fordism and the Destruction of the Natural Environment

A corollary of Fordism as a system of mass production and mass consumption is that it was (and to a considerable degree still is) a system of the mass production of wastes and ecological mass destruction. Since economic activities all involve energy and materials transformations that are subject to the laws of thermodynamics, they inevitably have unintended environmental effects. Fordist production, like all capitalist production, regards the natural environment as a source of natural resources to be exploited efficiently in search of profits. At various times fears have been expressed that some key natural resources (notably hydrocarbons and metallic minerals) would become exhausted and pose 'limits to growth'. Such resources undoubtedly are absolutely limited in supply but a variety of technical and social innovations as Fordist production has evolved have, for the moment, kept this issue relatively low on the agenda. However, while there have been some improvements as a result of government legislation and state regulation, the way in which the natural environment continues to be regarded as a sink of infinite capacity for deposition of unwanted wastes, the unpriced by products of production, poses very serious problems. In addition, once the products of Fordist mass production are sold, consumers of them treat the environment as a waste dump for the unwanted by products of consumption or for commodities at what they deem to be the end of their socially useful lives. There is abundant evidence of such impacts. Limits to this waste dumping can be set by national states placing a total prohibition of particular types of waste discharged, by setting maximum permissible limits, by creating prices for them via taxation systems, or via creating trading systems in pollution permits. A consequence of this is that pollution standards vary between countries, with important implications for the location of pollutant industries and for patterns of trade in wastes.

As environmental standards rose in the core countries of the Fordist economy, so too did pressures to relocate environmentally polluting production processes to peripheral locations. From around the 1960s there was growing concern about the environmental pollutant effects of certain sorts of industrial production (including those involving asbestos, arsenic trioxide, benzidine based dyes, certain pesticides and carcinogenic chemicals, chemicals such as polyvinyl chloride (PVC), and some basic mineral processing activities such as those for copper, lead, and zinc). As a result, enabled by advances in transport and communication technologies, companies began to relocate hazardous and polluting production activities. Initially this involved moving to peripheral regions within their home national territories but increasingly to parts of the Third World, thereby constituting a formative moment in the creation of the (so called) newly industrializing countries. Companies were often encouraged in this by financial inducements and low (or no) levels of environmental regulation as national governments were eager to encourage the per ceived benefits of modernization via industrial growth and employment. This has greatly extended the com modification of nature and greatly exacerbated local causes and levels of environmental degradation. Cru cially, however, this simply displaced the location of pollutant industries as part of the new Fordist spatial divisions of labor rather than effectively dealing with the issues of polluting production.

An alternative to exporting polluting industries from the core territories of the Fordist economy is to export their pollutants. The international trade in pollutants is complex, however, with locations such as La Hague in France and Sellafield in the United Kingdom importing nuclear waste from other parts of the world for re processing there. More often, though, destination lo cations for wastes are in countries on the periphery of the global economy, such as Kassa Island, off the coast of the former French African colony of Guinea, or the Ivory Coast.

Thus, as people in more economically developed countries came to understand the dangers posed by noxious pollutants, and environmental standards were increased, pressures rose to find ways of dealing with such pollutants. Exporting them to other areas desperate for foreign currency was often cheaper than dealing with them at home. It was made easier when the recipient countries were misled about the nature of the wastes and/or had authoritarian nonelected governments who neither knew nor cared about the risks to the environ ment and the health of their populations. Dealing with such wastes within their home territory could, on the other hand, involve considerable financial costs but, perhaps more significantly, political costs, in the face of NIMBYism ('not in my back yard') and opposition by local communities to such wastes being treated in 'their place'. However, not all local communities have equal capacity to resist so that such waste dumps are often located in areas of working class or ethnic minority housing. Indeed, poorer communities within the ad vanced capitalist world and peripheral states within the global economy have engaged in bidding wars, seeking to act as destinations for hazardous wastes in return for monetary payments and incomes.

The Limits to Fordism?

There clearly are limits to Fordism, at both the micro and macro scales, as a particular way of organizing capitalist production and, more generally, economic and social life in capitalist societies. Some of these limits would apply to all forms of political and economic organization, notably those arising as a result of the laws of thermodynamics, others arise from the general contradictions of a capitalist political economy (e.g., those that reflect the contradictory character of capitalist social relations) and some are spe cific to Fordism itself (e.g., the tensions inherent in the state seeking to balance mass production with mass con sumption within the national territory). There has been much written about the demise of Fordism and its re placement with some form of post Fordist successor but while there is considerable evidence of experimentation with micro scale organization of production processes, it is by no means clear that this constitutes a non Fordist al ternative and there is scant evidence of a new and stable post Fordist macro scale model of political and economic organization. One has to only look at China and India to recognize that the age of mass production and mass con sumption is far from dead. On the other hand, the failure to recognize that Fordist production, exchange, and con sumption activities involve material transformations that have unintended consequences and have created a legacy that could well have globally catastrophic consequences is becoming perhaps the key issue of the twenty first century.

See also: Consumption; Environmental Regulation; Fordism, Post-Fordism and Flexible Specialization; Industrialization; Regulation; Spatial Division of Labor; State.

Further Reading

Aglietta, M. (1979). *A theory of capitalist regulation: The US experience*. London: New Left Books.
Aglietta, M. (1999). Capitalism at the turn of the century: Regulation theory and the challenge of social change. *New Left Review* 232, 41 96.
Beynon, H. (1973). *Working for Ford*. Harmondsworth: Penguin.
Dicken, P. (2007). *Global shift* (5th edn.). London: Sage.
Hudson, R. (2001). *Producing places*. New York: Guilford.
Hudson, R. (2005). *Economic geographies*. London: Sage.
Hudson, R. (2007). Region and place: Re thinking regional development in the context of global environmental change. *Progress in Human geography*, 31(5), 827 836.
Jessop, B. and Sum, N L. (2007). *Beyond the regulation approach*. Chelthenham: Elgar.
Lash, D. and Urry, J. (1987). *The end of organised capitalism*. Cambridge: Polity.
Lipietz, A. (1987). *Mirages and miracles*. London: Verso.
Miller, D., Jackson, P., Thrift, N., Holbrook, B. and Rowlands, M. (1998). *Shopping, place and identity*. London: Routledge.
Miller, R. (1991). 'Selling Mrs. Consumer': Advertising and the creation of suburban socio spatial relations. *Antipode* 23, 263 301.
Nava, M. (1997). Modernity's disavowal: Women, the city and the department store. In Falk, P. & Campbell, C. (eds.) *The shopping experience*, pp 56 92. London: Sage.
Sack, R. (1992). *Place, modernity and the consumer's world*. Baltimore: John Hopkins University Press.
Sassen, S. (1991). *The global city: New York, London, Tokyo*. Los Angeles: University of California Press.
Williams, R. (1980). *Problems in materialism and culture*. London: Verso.

Fordism, Post-Fordism and Flexible Specialization

J. Lovering, Cardiff University, Cardiff, UK

Introduction

A well known narrative in human geography enshrined in many textbooks tells the story of the rise and fall of Fordism and its replacement by post Fordism. In the 1990s, this came to occupy center stage in economic geography in particular. It has since lost some of its drawing power, but its influence remains significant, particularly in analyses of the changing role of regions and cities, and in policy debates. It is also implicit if not explicit in many accounts of the geographies of postmodernity.

The Narrative: The Historical Transition from Fordism to Post-Fordism and the Rise of Flexible Specialization

The narrative of Fordism into post Fordism suggests that economic development has proceeded through a distinct sequence of ages or epochs. The last of these was Fordism. We are now living under its successor, post Fordism. The story also supplies an explanation for this transformation. It brings a set of cognitive and normative messages. It tells us how to think about the past and the present, and what to do to prepare for the future. The lessons are political. The advent of post Fordism entailed not only economic but also social, cultural, and political shifts.

The concept of Fordism originated with writings on Henry Ford, whose industrial achievements in mass car manufacture greatly impressed his contemporaries. An tonio Gramsci, reading about them at considerable social and geographical distance (he was in an Italian prison), came to the conclusion that Ford's huge factories were not only the industrial corollary of mass production, but they also represented a new paradigm for how to organize labor in capitalist society in general, using a subtle combination of coercion at work and seduction in the world of consumption. Henry Ford had deployed the techniques of sociological inquiry and scientific management to develop a method for managing a largely immigrant workforce with diverse skills, origins, and languages. His combination of a mechanically regulated workplace, relatively high wages, and ideological in struction promised miracles of productivity in the factory, and unprecedented levels of consumption in the home. Gramsci saw Fordism as offering workers a Faustian deal; ever growing consumption was traded for dehumanizing work and political quietism. It implied the construction of an entirely new kind of working class. But he also believed that in the longer term this arrangement would undermine itself. Fordism would provide workers with more leisure than ever before. The more advanced workers would use this to develop a broader critique of capitalism. They would come to recognize their collective strength, and would challenge once again the domination of the capitalist class system. This time around the confrontation would take place increasingly through 'cultural' struggles in civil society.

Gramsci's writings on Fordism (or the American System) were revisited in the 1960s and 1970s by a new generation of West European intellectuals in or on the fringes of the Communist Party. They were searching for a means of conceptualizing the class dimensions of what seemed at the time to be series of profound upheavals in economic and political life. Gramsci's suggestions con cerning Fordism were expanded to develop the theory that capitalism developed through a series of well defined stages, conjunctures, or 'epochs'. The develop ment of capitalism was shaped by a sequence of strategies seeking to contain the tendency of the working class to form a self aware unified political force. While each such strategy provided a degree of stability for a time, it eventually disintegrated and a crisis prevailed until an alternative was arrived at. Each phase of stability de pended on the one hand a distinct pattern of economic growth (the 'regime of accumulation') and on the other a distinctive set of social institutions and behavioral norms (the 'mode of regulation'). Fordism as sketched by Gramsci had been the latest and greatest of such com binations, but by the 1970s it was in terminal decline. Capital was trying to modify it to create instead neo Fordism. The availability of new computerized infor mation and control technologies would enable firms to reimpose control by redistributing work across different groups of workers, and different geographies. The huge plants employing vast labor forces on routinized tasks that had characterized Fordism would give way to a di versity of workplaces. Some of these would employ labor on more flexible terms, giving workers the oppor tunity to develop high skills and conduct a wider range of tasks. These workers would gain greater autonomy, but their contribution to the collective task of production would be ensured by new technologies and the discip lining effects of competition. Others would labor on much as before.

The concept of Fordism thus entered political dis course just as the object it purported to describe was disappearing. Its function was to lay the ground for an analysis of what followed; neo Fordism. As this narrative became academicized and began to attract a wide range of contributors drawing on different analytical per spectives and political agendas, predictions of the arrival of neo Fordism generally gave way to talk of post Fordism. Sometimes there is a lot in a name. Talk of post Fordism suggested that capital was now seeking to do more than patch up tried and trusted institutions by using new geographies. Instead, it was embarking on a much more comprehensive project, rejecting all the key institutional and cultural elements of Fordism. American writers on post Fordism tended to devote more attention to technology and markets and the creative potential of the liberated individual worker, and less to the re structuring of the working class, and the contradictions in this which had been Gramsci's main concern. With the addition of the theory of 'flexible specialization' in the 1980s the story of life after Fordism took a radical new turn, apparently offering lessons very different from those intended by its original authors.

By the 1990s these insights and suggestions had crystallized into a well defined orthodoxy among neo Marxist social scientists. Here the term 'Fordism' came to be largely a descriptive one, designating the institutions and social patterns of advanced capitalist countries (above all the USA) in the period beginning roughly in the late 1930s and ending in the mid 1970s. The key institutions were oligopolistic corporations, insti tutionalized systems of national collective bargaining, and high levels of national taxation and public spending. These all entailed huge bureaucracies, which in their turn created new fields of service sector employment, also organized on largely Fordist principles. These were the institutional correlates of a new set of macro economic relationships: regularized large scale invest ment in physical capital and plant, steady increases in productivity and output, mass consumerism, monopo listic economic governance, and nation centered circuits of spending and investment (overseas trade forming a relatively small part of national consumption). All this underpinned what came to be called the Golden Age of Capitalism, which saw not only an unprecedented rise in output, productivity, and material living standards but also a general reduction in economic and spatial inequalities.

Writers on management tended to call this managerial capitalism, economists were more likely to call it Keynesianism, while political analysts might refer to an era of welfare nationalism. To describe it as Fordism was to indicate membership of a particular intellectual and political community drawing its cognitive and normative inspiration from a broadly Marxist analysis of the causal centrality of capitalism. From this perspective a central feature was the feedback loop that Fordism provided between growing productivity and rising spending. This banished overproduction and underconsumption, the ghosts that had haunted capitalism in the 1870s and the 1930s. It achieved in reality the principle of recycling incomes so as to expand domestic working class pur chasing power that critics of the economics of imperial ism such as Rosa Luxemburg or J. A. Hobson had argued for theoretically half a century earlier. By expanding domestic markets, the economic motor of the tendency to imperialist expansion and war was switched off. Fordism provided the framework within which warfare capitalism came to be displaced by welfare capitalism.

The Working Class under Fordism

The central organizational feature of Fordism, according to Gramsci and the Regulationists, was the insti tutionalization of relations in the workplace. Different stages of the labor process were rigidly separated: con ception (the organization of methods and engineering based on elite skilled labor), manufacturing (deploying a skilled and disciplined workforce), and unskilled as sembly and execution (drawing on labor with no skill and allowing extensive mechanization). This followed the principles of scientific management set out in by F. W. Taylor, who famously claimed that the most boring tasks could as well be performed by a trained gorilla. For later writers on Fordism, especially in the American tradition influenced by Piore and Sabel in the 1980s, the Taylorist workplace was the cellular unit encoding the essential principles and drivers of the system of mass production. It imposed constraints on individual expression and creativity and these eventually dragged that system into crisis. The European neo Gramscian approach gave ra ther more emphasis to the class relations implicit in the institutions of Fordism, emphasizing national procedures shaping collective agreements in industry, full employ ment policies, the empowerment of trades union as bargaining entities, the cultivation of working class con sumerism, and the effects of the welfare state in sus taining the feedback links of consumer capitalism. Keynesian macroeconomics had provided for the first time an apparatus able to regulate the national economy as a whole. Under Fordism the space of capital came to correspond to national political space to a degree unequalled before or since. High taxes and a large public sector were the price paid by national capital as a trade off for assured if unremarkable profits and minimal disruption.

The Rise of Fordism

This conception of Fordism was based on an image of the US. But different variants of 'Fordism' were to be found in most advanced economies in the 1950s and 1960s. A cloned and exaggerated version in some respects flourished in the Soviet Union, symbolized in the anec dotal literature that was very popular among writers in this field by the famous barber shop with 100 seats. Some of the industrial and labor process ingredients of Fordism proved to be exportable to Third World countries such as Mexico, Brazil, Turkey, and South Korea. The global ization of Fordism was promoted by US multinationals, US political influence through transnational bodies, and by local development elites seeking rapid industrializa tion by copying Western methods. In the latter cases, the working class was not large or affluent enough to create the self sustaining macroeconomic dynamics of the core Fordist countries, so the Taylorist and state protectionist aspects usually dominated.

This story of Fordism concentrated on the functional relationships which it entailed. This was complemented by a historical narrative which sought to explain how this remarkable assemblage came into being. Private firms discovered the production line and Taylorist labor process. Governments discovered the advantages of state deficit financing and macroeconomic management. Eventually, especially in the US during the 1940s, the construction of Fordism become a conscious goal (for Gramscians a 'project') on the part of identifiable factions within capital and within the state. In the US motor industry interests and the European oriented foreign policy establishment proved especially important.

The Demise of Fordism

But by the same token Fordism contained the seeds of its own eventual demise. For some this lay in its destructive impact on individuals and their creativity. For others the problem lay in the unsustainability of its institutions and systemic connections. Over time, firms used to profiting from largely secure domestic markets lost the incentive to maximize efficiency or innovation and began to lose market shares to new competitors. They turned increas ingly to national state protection in the 1960s. The state underpinned empowerment of workers under Fordism also meant that they too experienced declining com pulsion and became increasingly inefficient and expen sive. The late 1960s saw a rapid growth of grass roots level industrial militancy and the 'youth revolt' against capitalist discipline. From all sides the demands upon the state expanded until they reached the limits of the tax and spend made possible by the *de facto* class accord, and the technical capacities of state bureaucracies. A fiscal crisis broke out. In the US and later Britain this was followed by a middle class tax revolt which put an end to Keynesianism and the growth of the welfare state by the late 1970s. The demise of Fordism seemed to explain much of the crisis of the Western capitalist world: a combination of inflation and economic stagnation, or 'stagflation'. New political leaderships (Reagan, Thatcher, Mulroney, Hawke, Ozal, etc.) committed themselves in varying degrees to purging their national Fordist insti tutions. In so doing they deepened and globalized the recession. In this austere context, corporations experi mented with new styles of labor management, technol ogies, marketing, and locations. By the 1990s, these had coalesced to form the key elements of post Fordism.

In the late 1970s, Regulationist writers had anticipated increasingly austere neo Fordist control over the working class and a corresponding new global geography of pro duction. The dominant account of post Fordism 10 years later presented rather a different scenario. It emphasized all round fragmentation: technological, economic, social, cultural, political, and geographical. And it emphasized the corporate and geographical opportunities this would bring. The central feature of the emergent post Fordist regime of accumulation – in these new accounts – would be a reorientation from mass produced standardized commodities to differentiated product markets. Industrial investment would increasingly focus on creating the capacity to introduce new products (and to a lesser extent processes) rather than ever larger economies of scale. Corporate managements would move away from top down authoritarian management and seek instead to mobilize the creative resources of their staff through flatter management hierarchies and new individualized incentives. The market share of firms could only be de fended by permanent revolution in product innovation, processes, marketing, and thence employment. The key to this upbeat new reading of 'life after Fordism' was the arrival of what its enthusiasts called 'flexible specialization'.

Flexible Specialization

By the mid 1990s 'flexible specialization' supplied a hugely influential new final chapter to the story of Fordism. In their landmark New Industrial Divide of 1985, Pirore and Sable claimed that new technologies were making possible a new social and technological model of production ideally suited to the demands of the new age of permanent industrial and market revolution. That revolution was portrayed as a technological and cultural inevitability to be celebrated, rather than a de fensive response to governmental and corporate crises, and as such to be inspected for weaknesses. For the new technological optimists the increased volatility of markets reflected, quite properly, the fickle tastes of affluent consumers and the inevitable, but welcome, growth of competition. Flexible specialization was the formula for success in addressing these challenges. Bespoke batch production, on a new high technology basis, provided the technological corollary to the break up of standardized consumption. Specialized goods could be produced for rapidly changing markets using flexible general purpose machinery and high skill labor.

For its enthusiasts, many of whom found a home in economic geography departments, flexible specialization was not merely a new corporate strategy; it was a desirable model for the construction of an entirely new system of production and consumption. It unleashed the creativity and collaborative capacities of the workforce which the dismal routines of Fordism had long suppressed. Instead of a mass force engaged in boring routine tasks and earning institutionally set incomes from rigidly prescribed Taylorist work, post Fordist competitiveness required a dynamic and flexible labor force able to make constant active contributions to ensure ceaseless innovations to both products and process. Rather than relying for their careers on a stock of established technical skills, workers would increasingly need to create new careers by nurturing the skills of adaptability and the ability and willingness to contribute their creativity and imagination to work teams. Where the icon of Fordism had been Henry's quintessential mass production car plant, the corporate icon of post Fordism was another but very different car manufacturer – Toyota. Toyota (so the story went) had transformed production and employment by deploying new design, production, and communication technologies. It distributed work among closely linked subcontractors and maximized the use of just in time to minimize inventories. Workers themselves took responsibility for quality control, and helped generate new innovations. Toyota embodied the post Fordist future.

Workers would no longer be interchangeable lumps of labor. Particular groups of skilled workers, rather than vast supplies of anonymous labor, would be the key to corporate competitiveness. In Piore and Sabel's vision, flexible specialization would enable to rise of a new class of yeoman crafts workers and entrepreneurs. They would be linked through relations not only of commerce but also of trust. But this would take time, and in the immediate future the spread of flexible specialization would widen the unhappy differentiation between different groups of workers (and consumers). The manual industrial working class would decline to become a rump surviving only in the low wage, low skill margins where flexible competitors had yet to arrive, or anticipated profits to be too low to bother. The economy and labor market would, for the foreseeable future, be characterized by a growing industrial divide.

The Fordist Region

The Fordism into post Fordism narrative was particularly attractive to geographers because it constructed each term in ways that had immediate spatial implications. It promised both a powerful way to explain empirically observable changes in economic geographies, and suggestions as to how future policies might best capitalize upon emergent trends. The idea that the geographical changes of the 1970s–1990s represented a restructuring of historical significance – the spatial dimension of a fundamental reconstruction of capitalism as a system – became an almost unchallenged orthodoxy.

The economic geographies of Fordism were shaped by its characteristic separation between the different functions of the labor process, projected across space. As the sustained growth of national markets led local firms to become national corporations, a new interregional division of labor had formed, guided by the Taylorist separation of functions. Different types of work were allocated to different regions. Some specialized in skilled production tasks and others in de skilled production and assembly. Less advantaged places specialized in routine production and assembly. But the background context of high national growth rates and full employment policies ensured a degree of convergence in terms of overall and per capita gross domestic product (GDP) indicators. The era of Fordism was in general one of declining regional and urban disparities. In aggregate terms, if not qualitative ones, employment opportunities spread out from the overheated center to the cooler periphery. At the local level, this often created a *de facto* territorial compact between places and firms.

But the demise of Fordism implied the reversal of these trends. In the versions of post Fordism that gave pride of place to flexible specialization, a very different conception of the causes and requirements of regional (and later urban) development became influential in the 1990s. The relationships between places would now be shaped by competition between them, rather than by their functions in a corporate and state administered division of labor. Post Fordism means place competition. For some writers, notably the Los Angeles School of Economic Geographers at UCLA and its many followers, this raised the exciting prospect of development of a set of new industrial spaces based on clusters of innovatory firms interacting both through the market and (more importantly) in civil society. Coinciding with – and targeted on – the emergence of a new stratum of policymakers at the regional (later city regional) level, such cheering claims became very prominent in the 1990s, acquiring the label of the New Regionalism. This highlighted the positive potential of post Fordism and the new proliferation of subnational units of economic governance (small countries, regions, and city regions).

The Fordist City

The concept of Fordism also suggested new ways of understanding the urban. Where the Victorian industrial city had been defined by its role as a center of pro duction, the Fordist city was essentially a site for mass consumption. Under Fordism, urban development was assimilated to the process of generating new consumption demands to absorb the increasing output. The most vis ible and revolutionary dimension of this was an un precedented wave of suburbanization, and the radical reconstruction of cities around car use. And as subur banization attracted higher income groups outward the inner cities changed their social composition. At first they became more narrowly working class. Later, as Fordism and skilled manual employment declined, they became the main locus of deindustrialization, un employment, and poverty. In racially divided societies such as the US this took on the charged dimension of 'white flight'.

The demise of Fordism accordingly signified the start of an epochal change in the regional and urban dynamics of capitalism. And this had cultural dimensions. As post Fordist flexible accumulation became the key to renewed growth (accumulation) postmodernism became the dominant mode of representation. In this account, the post Fordist city witnessed the replacement of industrial society by urban society. Many claims originating in the Fordism–post Fordism analysis went on to influence postmodernist ideas of the city more widely. Hardt and Negri, for example, claimed that the Fordist worker was a 'mass worker' while the post Fordist worker is a 'social worker', self valorizing and intellectually equipped, whose appearance on the historical stage augurs the (spontaneous) arrival of a new post socialist revolution ary 'multitude'.

Post-Fordism in Policy

At a less abstract level the thesis of the imminent arrival of post Fordism was thought to have major and clear policy implications. A whole new quasi academic indus try emerged to spell them out. Industrial policy should aspire to generalize flexible specialization, especially by creating smaller firms and more innovative ones. Vertical disintegration, networked production, and the adoption of flexible technologies should be encouraged. Labor market policies should break up collectivist institutions and encourage flexibility and training for adaptability. Learning and knowledge should be key policy goals for firms, workers, and places. Spatial policies should en courage place competition, especially through the de velopment of industrial districts and other forms of cluster wherein the players interact dynamically through both traded and untraded connections. The old spatial policies of the Fordist era were now both unworkable and unnecessary. For many urban and regional political leaders, Silicon Valley and the Third Italy came to be seen as offering transferable lessons.

Strengths and Weaknesses of the Narrative

The narrative of Fordism and its transition into post Fordism has been hugely influential in human geography, especially economic geography. But in the course of its migration from a European Left intellectual perspective with an implicitly radical agenda into a transatlantic academic orthodoxy with policy implications in line with the wider political and corporate policy rhetoric, it underwent a number of subtle but important changes. The initial idea that Fordism would give way to neo Fordism was largely forgotten. And the differences of focus between the neo Marxist Regulation approach and the American technological–institutional approach ten ded to be overlooked as the narrative became a neat pedagogic orthodoxy. As such, it provides students with a sweeping introductory historical outline that has the virtue of seeming to explain much. It certainly inspired a wave of productive new research. But on closer exam ination it was always riddled with empirical, conceptual, and theoretical problems. Talk of Fordism and post Fordism, and especially of flexible specialization, has only been free of controversy or challenge when it has been comfortably vague. For it was, and still is, beset by a lack of clarity and agreement as to exactly what the key concepts describe, how appropriate they are to different national and regional situations, and what it is that they really explain.

Empirical

When and when did Fordism exist?

The first and most basic problems were empirical. The narrative tells a neat story, but history and geography are never so obliging. Few parts of the world ever experi enced full blooded Fordism–Taylorism of the kind which occupies central place in the story. The Ford car plant itself, for example was never very representative of US industry. In many advanced countries many industries, including automobiles, were never thoroughly Taylorist even in the Fordism's supposed heyday. Many companies experimented with Taylorism and abandoned it in favor of more humanistic techniques even in the Fordist era. The tidy world conjured up by Fordist storytellers never really existed, even in the American heartland.

The narrative of the decline of Fordism was no less empirically questionable. The claim that mass production was coming to an end was wildly overstated when it was first made in the 1970s. It is perhaps even more at odds with reality today. In the twenty first century, standard ized goods and services are being produced and con sumed on a more global and massive scale than ever before. That some of these are modified to form local bespoke packages does not speak to the end of mass production, but its continuation as the platform for strategies of market differentiation in an era of incipient overproduction. In the more developed countries recent employment growth has been concentrated in the service industries. And much of this has taken a highly Fordist form in terms of the labor process and the production of standardized commodities (Walmartization and Starbu kization). In the fastest growing newly industrializing countries such as India and China, much of the recent expansion has been based on techniques of mass pro duction and labor processes that echo Fordism and Taylorism rather than post Fordist flexible specialization.

Many of the foundational assumption of the post Fordism thesis turned out to be at best overstated, at worst parochial. Even when these claims were first made, researchers cast doubt on their empirical validity. In the Third Italy, for example, larger firms were in fact growing faster than small ones. The much publicized Italian craft model was limited to a few special cases, and these de pended on the existence elsewhere of much more Fordist kinds of employment. Large US firms were reasserting their dominance over supposedly post Fordist smaller ones, and this was giving rise to increased labor market segmentation and worker inequality. Today, in the world as a whole the number of people producing mass prod ucts and working under more or less Fordist–Taylorist conditions is probably far greater than when the death sentence on Fordist mass production was formally an nounced in seminar rooms two decades ago.

How pervasive is 'flexible specialization'?

The suggestion that flexible specialization provides the descriptive and explanatory key to post Fordism has not been confirmed by empirical or theoretical research. Empirically, there are clearly many different models of 'post Fordism', including not only Toyotaism but the Alpine and Nordic model based on high technology consumer goods. And some of the supposedly arche typical examples of post Fordism, such as Japanese car firms, have turned out to be much less exemplars of flexible specialization than early commentators believed. Surveys of small firm development have not identified a general trend to flexible specialization, even in sup posedly classic case such as subcontracting in Japan and Germany.

The theory of flexible specialization claimed that it was much more than just another form of corporate adaptation; it was destined to become hegemonic. Some seekers after flexible specialization and of the dynamic regional clusters associated with it have managed to find sufficient examples to sustain this vision. But the non converted have not been persuaded. While empirically it is clear that various forms of flexible specialization have been adopted by a number of prominent firms and in fluenced a number of significant locations, it requires a very selective focus to make these appear as prototypical. And it requires a rather tendentious kind of technological determinism to cast flexible specialization as the key factor explaining contemporary growth and geographical tendencies in general.

Labor flexibility – post-Fordist speculation versus neoliberal reality?

An early empirical critique of post Fordism focused on the vagueness in its discussion of labor flexibility. Post Fordist accounts often failed to distinguish carefully enough between increased flexibility in work and flexi bility in employment. Clearly, some fields of work were indeed becoming more pleasant. More liberal conditions of employment were typical for the higher qualified and more 'creative' work force. But these were often made possible only because elsewhere labor was engaged under broadly Fordist conditions (e.g., in subcontracting or supplier production units located in the emerging economies). If the quintessential Fordist worker was tied to a production line in a car plant, the post Fordist equivalent turned out to be tied to a cappuccino machine and the erratic demands of customers, not necessarily a freer or more creative situation. The idea that post Fordism necessarily involved flexible specialization, and that this meant labor markets and employment con ditions would generally become less Taylorist, could be sustained empirically only by looking at selected trees, not the forest. Critics objected that advocates of the post Fordist were not discovering new trends so much as lending a new cheerful gloss to the age old reality of fragmented and inegalitarian labor markets.

Conceptual and Theoretical

If the empirical and historical status of the claims central to the Fordism–post Fordism narrative were contentious, their explanatory value was even more contestable. The narrative was from the start criticized for a lack of clarity in conceptual terms and apparent contradictions. These critiques focused on the tendency to explanatory re ductionism in the form of technological determinism in the versions of post Fordism emanating from one side of the Atlantic, and disguised vulgar Marxist economic determinism in those coming from the other.

In the technology centered and individualistic per spective (signified by Piore and Sabel), the explanatory weight is carried by claims concerning new technological possibilities in the organization of production and labor markets. Some detected in the rosier expressions of this the influence of a lingering romantic Jeffersonian vision of an economy of independent small businesspeople, close to that of classical liberal economics. The flexible specialization thesis could be seen as a Left ish version of the prevailing neoliberal tendency. The legacy of the

European post Gramscian approach on the other hand was criticized for a tension between its official intention to develop a multicausal perspective and a *de facto* ten dency to portray events as superstructural adaptations to the requirements of a new pattern of accumulation. This tradition smuggled in a functionalist, if not simply tele ological, pseudo explanation.

In both streams of post Fordist writing the original Gramscian notion of capitalism as a class system and as a cultural construct had been lost. By focusing on sup posedly enduring factors of production and employment both accounts gave more emphasis to the apparent structural dimensions of capitalism (institutions, both formal and informal) than the perpetual and amorphous class struggle that, for Gramsci at least, had been at least equally important. The original Regulationist question, asking how an enduring way of managing class relations and conflicts comes to be established, and how it gives rise to new sets of contradictions, had slipped off the research agenda. Although much ink would eventually be spilt by Regulationist writers insisting that an analysis of Fordism and post Fordism enables analysts to grasp na tional or subnational specificities the contents of this tin have rarely matched the description on the outside. The best historical accounts of the development of accumu lation and regulation in particular countries have drawn only lightly on the Fordism narrative, treating it as a stylized ideal typical heuristic starting point rather than rigid framework. The Fordism into post Fordism story does not really explain much about recent history, and much of recent history fails to confirm the story.

What is the connection between post-Fordism and flexible specialization?

By the mid 1990s even some of its earlier enthusiasts felt that so many criticisms of the flexible specialization ex planation for industrial districts had accumulated that its use in future should be much more cautious and muted. The claim that post Fordism in general would be char acterized by flexible specialization has become increas ingly implausible. Jessop suggested that post Fordism should be seen only as a very general historical category the journey toward which could take a variety of forms. The normative assertion that policy should aim to maximize flexible specialization enjoyed considerable influence in certain elite policy networks, especially at the regional and urban levels, but has been controversial in more intellectually rigorous circles.

The relationship between flexible specialization and economic restructuring asserted by early writings on post Fordism arguably rested on misreading changes in degree as changes in kind, and on misinterpreting con tingent corporate behavior of the period as signs of ne cessary general tendencies. In the early 1970s, Hymer argued that firms increasingly focused their product innovations and marketing on new expensive products sold to affluent elites (vertical accumulation), rather than continuing to expand the output and reduce the price of products consumed by working class consumers (horizontal accumulation). This tendency was reinforced by the effect of subsequent neoliberalizing policies, including greater openness to imports and inward in vestment, and the accompanying slowdown in growth of per capita output and incomes in former Fordist coun tries. Firms have sought to protect market share by driving up expenditure on marketing and greater product differentiation. This suggests that the corporate emphasis on innovation – so prominent in post Fordist analysis and policy debates – is more a response to changed market conditions than their prime mover. Some degree of flexible specialization has long been an element in the repertoire of corporate strategies, contingently useful or not according to case. The post Fordism claim that it is necessarily much more than this is not anchored in a firm grasp of economic history.

Methodological problems in the use of case studies

The academic literature invoking flexible specialization as the key to post Fordism has drawn heavily on case studies. But these have often attracted criticism on methodological grounds. A typical approach took the form of listing alleged exemplars of flexible specialization (Emilia Romagna, Baden Wurttemberg, the Salzburg region, and Silicon Valley used to be favorites), inferring common characteristics, attributing these with explana tory significance, and then drawing policy conclusions. This method risks treating regions as semiautonomous entities making their own way in the world. And indeed, many criticisms argued that the particularities of each example and their connection to specific national and international contexts were not properly explored. The New Regionalism was characterized by a voluntaristic conception of what regions can do for themselves, carrying the tendentious implication that particular re gions show the rest of the world its destiny. The sug gestion that Silicon Valley, for example, was some kind of post Fordist prototype that could be emulated elsewhere became enormously popular. Millions of air miles were consumed relaying visiting politicians and consultants from other valleys ambitious to learn its secrets. But its real achievements were much more ambivalent, and much more dependent on nontransferable locational and historical factors than the fashionable view pretended.

The most celebratory accounts of the new geographies of post Fordism have been marked by a very casual use of ideal types in the theoretical analyses, and a lack of at tention to epistemological and methodological issues concerning case studies. Many empirical studies also adopted an actor focus, emphasizing the perspective of managements in particular. This may have encouraged the tendency to detach the particular from the general, the microeconomic from the macroeconomic. It is always difficult to know how much weight to attach to actors' claims for their own novelty and significance.

Alternative explanations of phenomena attributed to post-Fordism

As time has passed, the post Fordist narrative has become steadily less persuasive. Many factors other than those stressed in these accounts have clearly had a major im pact on industrial, employment, and geographical change. These include the continued vulnerability of capitalism to booms and slumps (1997, 2002, and 2007–) and the rising significance and globalization of finance capital and consequent property market, and the wider question of the role of modernization strategies in shaping national economic trajectories. Macroeconomic and international trade and investment policies since the 1980s have slo wed rates of growth of aggregate demand while opening up national economic space. This has had a far greater quantitative and qualitative impact on the geographies of employment than flexible specialization or any of the other institutional dimensions of post Fordism.

Leaving aside the massive changes in the economic impact of a few major countries – especially China and India – the era after Fordism has been characterized by a general stagnation in the growth of productivity, output, and output per capita. In this sense it has been less dy namic than the Fordist era, despite the impression given by the flood of glittering new products in the shops. A larger share of output goes to capital, but the growth of consumer demand has not been sufficient to generate enough investment opportunities to absorb it, so a smaller share of profit is invested. An increasing share is diverted to acquiring rent generating assets (land, prop erty, and artifacts) rather than industrial productivity. This underpins the explosion of the property market which has shaped the global proliferation of urban re generation strategies. This is probably both empirically and theoretically a much more ubiquitous and significant influence on development in many cities and regions than the emergence of flexible specialization.

The post Fordism literature paid little attention to neoliberalism and especially to the associated growth of finance capital (the words are conspicuously missing from the index to Amin's reader on post Fordism). But the rise of finance capital would seem to go a long way to ex plaining why many large firms have in effect transformed themselves from organizational units embracing physical assets in the form of production capacities (factories) and large scale in house labor, and have become instead financial units coordinating the distributed division of labor from conception to production to marketing. Neoliberal industrial restructuring has transformed capitals from bounded units of production to units of organizational power. Many older manufacturing cor porations have shed much of their production capacity and have evolved into entities managing the conception, production, and marketing process through a series of subcontractors, filliers, strategic alliances, and so on. Case studies of particular units characterized by flexible specialization, or of regions, throw little light in them selves on the dimensions and dynamics of change in this corporate context.

The post Fordist account also completely overlooks some prominent macroeconomic and labor market in stitutions of the literally post Fordist (or neoliberal) era. One such is the Bretton Woods II system whereby US consumption is subsidized to the extent of $1.7b daily by funds from Asia. This bizarre contingency of inter national economic relations underpins the expansion of Chinese 'neo Fordist' production on the one hand and the 'post Fordist' specialization of some American pro ducers on the other. Another is the fact that the global commodity producing labor supply has grown by of 1 billon – between one third and a half – following the demise of the Soviet Union and the state led opening of China. This has had a profound impact on corporate strategies, for example, making it much more feasible to redistribute routine labor to cheaper workforces for example.

The Fordism–post Fordism story, despite intentions, has taken a very one sided and economistic view of the construction of market forces which overemphasizes technology and underemphasizes political and socio cultural factors. One casualty has been gender, largely invisible in most accounts of Fordism and post Fordism. But changing gender relations, household structures, as pirations, and labor supply have a fundamental effect on economic restructuring, especially in countries experi encing neo Fordist growth such as Turkey, Brazil, or China. This is both an empirical and a theoretical omission.

Policy and Politics

Any doubts about the empirical and theoretically plausibility of the Fordism–post Fordism narrative filter through to its usefulness for policymaking. For critics the policy conclusions drawn from this account were in adequate from the start: first, they would be unlikely to succeed even in their own terms. Second, they would be likely to be partial in the sense that they prioritize par ticular economic interests and social groups over others for spurious reasons.

One line of criticism suggests that flexible special ization was in fact little more than the particular response of a particular set of Northern firms to local contin gencies. It was a local reaction to specific influences, including the general neoliberalization of policy. Post Fordist accounts misrepresented this as signifying a new historical trend in capitalism in general. In so doing they provided an ideological gloss legitimizing the particular response to crisis adopted by those American and European firms and their governments, puffing up a local contingency as a universal prototype. The effect was to suggest that a US centric 1990s model was nothing less than the ideal type for the rest of the world for the foreseeable future.

If post Fordism is a Gramscian project to reestablish capitalist stability, then the flexible specialization story tended to become part of this project rather than a critique of it. It licensed a focus on, and political support for, those who stand to benefit from globalization and neoliberalization. This emphasis on flexibility (and by derivation, on territorial production complexes) diverted attention from many other questions relevant to geographical change. In particular, it overlooked macroeconomics, although the state of aggregate demand remains much more important to the economic welfare of working people in general than the competitiveness of particular firms or sectors.

The post-Fordism narrative legitimizes the emergence of a global capitalist class?

Arguments such as these implied that the politics of post Fordism, despite claims made for it in the 1980s was by no means automatically progressive or Left wing at all. And indeed the post Fordist narrative proved to be very useful to particular fractions of capital and particular governance interests, playing a major role in the construction of a shared discourse within the new global networks of policymakers and corporations that proliferated from the 1990s. It spoke of and to the new networks of spatial policymakers, rather than communities, working people, or small entrepreneurs in general. Stephen Hymer argued in his pioneering analysis of globalization in the early 1970s – largely ignored by Regulationist writers – that the shift from horizontal to vertical accumulation, which post Fordist accounts now take as a given fact of contemporary economic life, in fact reflected the political emergence of a fraction of capital whose interests lay in the global enforcement of private property rights, the free movement of capital, and increased income inequalities, rather than the steady increase of domestic working class incomes. The post Fordist thesis, by taking these as foundational assumptions and proposing policies that adapt to them rather than challenge them, might perhaps usefully be seen as one of the academic reflections of the emergence of a new global capitalist class associated with the rise of neoliberal policy biases at a range of spatial scales.

The problematic legacy of the Regulation School

These suggestions take us back to the original political concerns of the early writers on Fordism. How is it that the concepts devised by Gramsci and the early Regulation School with the aim of exposing the dynamics of capitalism came to be so enthusiastically deployed in their own interest by powerful fractions of capital and the state? Part of the answer seems to lie in the epistemological and ontological problems the Fordist account inherited from that diverse group of pioneers.

Regulationist analysis (of which there have been many variants) was characterized from the start by the attempt to glue together a structuralist interpretation of the Marxist theory of the economic dynamics of capitalism (with in at least some versions an inexorable tendency to crisis), and a more conventionally empirical notion of the significance of politics and political culture and elite control of institutions. Recognizing that each of these domains has its own set of distinctive dynamics was an attempt to avoid crude economic determinism. But this was often little more than a well intentioned gesture. For the tendency to determinism lingered on in the way that the economic domain tended to be treated as if it was the original source of economic problems, while at the same time a Gramscian tendency to celebrate culture as autonomous played itself out in ideas about the mode of regulation. Explanations of the decline of Fordism and the construction of post Fordism which built on these on incompatible foundational principles unsurprisingly tended to oscillate between treating outcomes on the one hand the inevitable results of objective mechanistic determinants and on the other as the effects of contingent struggles and discourses.

Gramsci's original writings on the American system drew attention to the role of culture, and the capitalist penetration of leisure time through commodification and profit seeking. Fordism's corollary, the rise of a consumer society and the development of pleasure as a market, would bring about a long 'war of position' within civil society. Out of this, he hoped, the ideological and institutional conditions would emerge (organic intellectuals and new hegemonic forces) for a renewed working class challenge to capitalist hegemony. Gramsci thought that Fordist capitalism made political concessions to labor, empowering trade unions and social democratic political parties. It replaced military models of control with discipline imposed through the seductions of commodity consumption. But this only meant that it would eventually be confronted in new and more multidimensional ways as the working class developed new ways of life and a political morality fundamentally at odds with the capitalist elite. On the last point at least Gramsci's high minded expectations (along with Hardt and Negri's elitist fantasy of the politicization of the 'multitude') proved to bear little relation to real events.

Whatever the merits of the Gramscian conception of Fordism, it had a very different purpose and resonance than that of the Regulationist and technology based theories which crystallized out half a century later. The thesis that Fordism had to be replaced took on a life of its own as it became academicized, accreted new em pirical and theoretical claims, and transmuted into a rhetoric supplying a plethora of local policy suggestions to institutions of governance. In Britain, in particular, the institutionalization of the story of the arrival of post Fordism in the late 1990s was closely bound up with the development of an intellectual anti Thatcherite position, especially in the pages of *Marxism Today*, the magazine of the disintegrating Communist Party. Central to this narrative was the idea that the transition to post Fordism was the latest in a sequence of chapters in the history of capitalism. Although this story greatly exaggerated the extent of change, while leaving the explanation for it profoundly unclear, it served for some to provide a theoretical justification for a new conception of social democracy and the invention of 'new labor'. Several of its authors later became key players in, or theorists of, the Blair government of 1997–2007.

Conclusion: Post-Fordism Today

Over the same period the Fordism into post Fordism narrative became a pedagogic orthodoxy in geography, especially economic and urban geography. But the stron ger claims it makes beg too many empirical and theore tical questions. The descriptive and explanatory usefulness of the concept of Fordism has been thrown into doubt by disputes over how geographically general it was (e.g., differences between US and other countries); over the extent to which mass production prevailed and was para digmatic; over the link between the labor process and macroeconomic conditions in different cases; over how much lingers on or is even growing. Whatever the chronologically post Fordist era is, it is not simply an inversion of everything Fordist. Indeed, much of what defined that earlier era seems to be still prevalent today. Post Fordism as it was advertised in the 1980s and 1990s never really arrived. And the changes it purported to ex plain are often more usefully understood in other terms.

Perhaps the most striking fact about recent economic geographical history is that the richest countries and regions of the Fordist era are for the most part the richest countries and regions of the post Fordist era too. The exceptions concern some parts of a very small number of rapidly growing former peripheral economies, notably China and India. And here the story seems to be gen erally less about the arrival of post Fordist flexible spe cialization than about the large scale imposition of Fordist–Taylorist techniques. The Chinese case in particular raises the question of whether the Fordist and post Fordist 'models' should be seen not as sequential historical scenarios but rather as competing paradigms of socioeconomic management to be selectively deployed by ruling elites (corporate or national) committed to bringing about what they consider to be 'modernization' by allowing the expansion of capitalist social relations.

The post Fordist narrative, in this perspective, has diverted excessive attention to local geographies of change at the cost of broader geographies of continuity. Big stories have been overlooked in the search for smaller ones which tell the lessons that its exponents have set out to confirm. But the Fordism–post Fordism narrative is still ensconced in geography textbooks as if it describes simple facts and reliable explanations. Part of the ex planation for this academic curiosity must lie in the usual time lag between the emergence of academic ideas and their codification in textbooks. In part it also perhaps reflects the transformation of the university in recent years. Academics are under increasing pressure to es tablish their individual brand identity and competitive ness by associating with theories and claims that are widely debated and appear 'relevant' in the political context of the day. As a result, many of the pedagogic orthodoxies of human geography in the 2000s, reflected in this encyclopedia, are essentially the radical political claims of the 1980s, codified, elaborated, academified, depoliticized, and stamped with professorial authority. To this extent the best tools for an explanation of the strange history of the Fordism–post Fordism narrative may perhaps be found in Bourdieu's analysis of the role of habitus, self interest, group behavior, and fashion among Homo Academicus.

Although research, critique, and policy disillusion have robbed the Fordism–post Fordism story of its initial luster it nevertheless remains a suggestive simplification of current trends and imperatives from which to set out. And in the absence of an equally bullet point friendly alternative it is likely to continue to enjoy popularity on the academic fringe and its interface with the world of policy discourse and journalism for some time yet.

See also: Fordism; Industrialization; New Regionalism; Regulation; State.

Further Reading

Aglietta, M. (1979). *A Theory of Capitalist Regulation The US Experience*. London: New Left Books.

Amin, A. (1994). *PostFordism: A Reader*. Oxford: Blackwell.

Bonefeld, W. (1993). Crisis of theory: Bob Jessop's theory of capitalist reproduction. *Capital and Class* 30, 25 47.

Cooke, P. and Morgan, K. (1998). *The Associational Economy: Firms Regions and Innovation*. Oxford: Blackwell.

Curry, J. (1993). The flexibility fetish. *Capital and Class* 50, 99 126.

Davis, M. (1986). *Prisoners of the American Dream*. London: Verso.

Glyn, A. (2006). *Capitalism Unleashed: Finance, Globalization and Welfare*. London: Oxford University Press.

Harvey, D. (1985). *The Urbanisation of Capital*. Oxford: Blackwell.

Hui, W. (2003). *China's New Order: Society Politics and Economy in Transition*. Cambridge, MA: Harvard University Press.

Hymer, S. H. (1972). The multinational corporation and the law of uneven development. In Bhagwati, J. (ed.) *Economics and World Order*, pp 113 140. London: Macmillan.

Jessop, B. (1993). Towards a Schumpeterian workfare state? Preliminary remarks on post Fordist political economy. *Studies in Political Economy* 40, 7 40.

Kenny, M. (1999). Marxism and regulation theory. In Gamble, A., Marsh, D. & Tant, T. (eds.) *Marxism and Social Science*. London: Macmillan.

Lipietz, A. (1987). *Miracle and Mirages: The Crisis of Global Fordism*. London: Verso.

Lovering, J. (1990). 'Fordism's unknown successor. *International Journal of Urban and Regional Research* 14, 159 174.

Lovering, J. (1999). Theory led by policy: The inadequacies of the New Regionalism (illustrated from the Case of Wales). *International Journal of Urban and Regional Research* 23, 379 395.

Massey, D. (1995). *Spatial Divisions of Labour: Social Structures and the Geography of Production* (2nd edn.). London: Macmillan.

McDowell, L. (1991). Life without father and Ford: The new gender order of post Fordism. *Transactions of the Institute of British Geographers* 16, 400 419.

Murray, R. (1989). Fordism and postFordism. In Hall, S. & Jacques, M. (eds.) *New Times: The Changing Face of Politics in the 1990s*, pp 38 53. London: Lawrence and Wishart.

Pijl, K. V. D. (1984). *The Making of an Atlantic Ruling Class*. London: Verso.

Piore, M. and Sabel, C. (1985). *The Second Industrial Divide: Possibilities for Prosperity*. New York: Basic Books.

Pollert, A. (1988). Dismantling flexibility. *Capital and Class* 34, 42 75.

Sayer, A. (1989). PostFordism in question. *International Journal of Urban and Regional Research* 13, 666 695.

Scott, A. (1988). Flexible production systems and regional development: The rise of new industrial spaces in North America and Western Europe. *International Journal of Urban and Regional Research* 12, 171 185.

Stewart, P. (ed.) (1996). *Beyond Japanese Management: The End of Modern Times?* London: Frank Cass.

Storper, M. (1995). The resurgence of regional economies ten years later. *European Urban and Regional Studies* 2, 191 221.

Tickell, A. and Peck, J. (1992). Accumulation, regulation and the geographies of post Fordism: Missing links in regulationist research. *Progress in Human Geography* 16, 190 218.

Williams, K., Cutler, T., Williams, J. and Haslam, C. (1987). The end of mass production? *Economy and Society* 16(3), 405 439.

Foreign Direct Investment

A. Lagendijk and B. Hendrikx, Radboud University, Nijmegen, The Netherlands

Glossary

Export processing zones An area in which a particular set of institutions, buildings and services is set up in order to attract (foreign) businesses for export production. It is often endowed with more favorable institutions and regulations for production (taxes, labor regulation, tariffs) than within other parts of the country.

Greenfield investment Direct investment in new facilities on a location where no previous facilities existed.

Internalization advantages The advantages for production arising from the internalization of production and other owner specific advantages by firms to avoid market imperfections (i.e., avoiding import tariffs; avoiding property rights enforcement costs; etc.).

Location specific advantages The advantages for production arising from country specific assets and resources, such as the presence of a skilled labor force, an optimized institutional context, and low transportation costs.

Owner specific advantages The advantages for production arising from assets specific to a firm that give it an comparative advantage over other/domestic firms, such as the ownership of specific technologies, trademarks or patents, specific management and organizational systems, access to spare capacity, etc.

Product (life) cycle theory Theory developed by Raymond Vernon that links each of the phases in the development of products to particular spatial strategies. The theory is based on the premise that at each stage of the product-life cycle different types of countries (ranging from the developed to the developing countries) will have comparative advantages in the production of the product. The product life-cycle encompasses four stages: (1) the product's introduction and development (in the home country); (2) the products growth and maturation (high income countries); (3) the standardization of production (low income countries).

TRIMs (Trade Related Investment Measures) National policies aimed to place conditions on foreign investors. TRIMs include policies aimed at bolstering FDI spillovers by creating incentives, regulations, and import barriers to stimulate the local economy. TRIMs may also include measures to ensure a certain degree of sustainability and social responsibility from foreign investors.

A Characterization

Foreign direct investment (FDI) is a cross border corporate governance mechanism through which a company obtains productive assets in another country. Such a link can involve the acquisition of (or merger with) already existing production facilities by getting hold of a majority share, or the establishment of a new facility. The first case is often abbreviated as an 'M&A' investment; the latter case is commonly described as a 'greenfield' investment (even if it takes place on a 'brownfield' location). FDI, as a consequence, differs from portfolio foreign investments and (bank and nonbank) loans, in that it results in a long term relationship between the parent company and the 'host country' facilities. Quintessential to this relationship is not the financial transfer, but the establishment of managerial control through the transfer of a bundle of factors and cross border organization of production. Because establishing control generally takes place in a step by step process, it is denoted as an 'establishment chain' generally unfolding over longer periods of time.

Like other forms of cross border investment, FDI is a major driver and constitutive of globalization. FDI is what makes a multiplant cross border company, better known as the multinational or transnational corporation (MNC or TNC; see item on globalization). It is thus a key ingredient in the shaping of global networks of production, trade, finance, and transfer of technology, knowledge, managerial practices, and labor. In doing so, it also contributes to the shaping of economic interdependencies and resulting in uneven relations, practices of cultural transfer and dominance, and the establishment of political hierarchies and 'power geometries', that is, intricate webs of domination and subordination across space. What characterizes FDI is that it tends to stretch out further than the flows of financial transfer. While financial globalization remains confined to a limited number of well connected urban hot spots across the globe, FDI also reaches the corners of lesser developed regions and countries. As a result, many peripheral regions depend heavily on the local subsidiaries of foreign companies. Like financial capital, however, FDI is hardly targeted at the poorest countries of the world. The latter continue to rely, to a large extent, on state transfers of technology and finance from donor countries.

Not only does FDI reach out further than other forms of capital, it is also less mobile. Since the investments target productive assets, they are 'bolted down', tied not only to fixed capital but also to a local production environment. This results in dependencies between the foreign investor, a particular pool of labor, local state and nonstate agencies, suppliers, distributors, etc. While there is always the threat of disinvestment, the image of being 'global' and 'footloose' is generally used by investors to increase their leverage over local communities and state agencies, more than as a prelude to actual relocation. Only certain types of investment, such as labor intensive production of staple commod ities, tend to respond quickly to shifts in relative cost levels of labor and inputs by relocation. Rather than full fledged disinvestment, companies often engage in rearranging the division of labor between their affili ates, accompanied by partial restructuring of the local labor pool and asset base.

What appears more of a 'threat' is the continuing appeal to companies in more developed countries to shift productive activities out to other parts of the world. Al though such outward moves have always been part of spatial process of industrialization, improving connect ivity and the rapid growth of highly skilled labor forces in countries like India and China have given a new impetus to what is now labeled as 'offshoring' of activities from 'West' to 'East'. In addition to the more traditional industrial activities by large companies, offshoring in creasingly involves the relocation of back office activities by smaller service companies. The background and im pact of these trends will be discussed in more detail below.

FDI Growth Patterns

FDI goes back a long time in history. The practice of founding in company trading representatives in foreign locations is actually as old as the emergence of pecuniary trading networks. The seventeenth and eighteenth century – the heydays of mercantilism and colonialism – provided a major impetus to the development of multi plant companies operating at a global scale through the institutionalization of the 'firm', notably in the form of incorporated companies. This did not only result in the rapid development of cross border banking networks in Europe, but also in the establishments of global networks by the British and Dutch (East) India Companies. In addition, many adventurous, and often short lived, trad ing companies emerged that hoped to benefit from new economic relationships with the East (Indian) and West American world as well as Africa.

The nineteenth century and twentieth century saw a further growth in foreign expansion, targeting both for eign markets and resources, from minerals and cattle to labor and technology, punctuated by periods of stag nation and decline. This started with the rise of modern 'multinational' companies, many of which are still around today (some subsumed in larger companies), such as Lever, Siemens, General Electric, Michelin, Hoechst, Nestlé, and Nobel Dynamite. In the nineteenth century, however, FDI was still dwarfed by portfolio investments that sought to benefit from difference in interest rates and financial investment opportunities in rapidly growing economies. Moreover, investment levels, both direct and portfolio, dropped dramatically during the Depression, starting in 1929, and the period of war that followed. Investment levels started to pick up slowly in the 1960s and 1970s, as a prelude to rapid growth notably in the 1990s up to the boom year 2000. FDI flows stalled in the early 2000s, after which they recovered to the level of 2000 in 2006 notably through the impact of M&A activity.

In the post war period, most investment flows continued to take place between developed countries, although the share of less developed countries as re cipients showed a gradual growth, from 20% to 40%, in the 1990s. In the late 1990s and early 2000s, this growth was dented as a result of the financial crises in Southeast Asia and Latin America (see **Figure 1**). In terms of donor countries, while recent figures show some diversification, what is striking is the persistent dominance of a small number of countries. After UK dominance in the nineteenth century, the USA was the lead spender in the early 1980s (over 40% of the global FDI stock). In the late 1980s and 1990s, the European Union (EU) and Japan have grown in importance, in contrast to the rest of the world (of which the share in world FDI stock dropped from 15% in 1980 to below 10% in the early 1990s). Yet, due also to an increase in flows between developing countries (so called south–south investments), this trend has been reversed in the last decade.

While such aggregate figures say a lot about the broad changes in FDI flows and stocks, putting exact figures to FDI is far from easy. Employment figures are relatively straightforward and easy to access, but financial statistics are troubled by a myriad of definitional and classificatory problems. They are also more difficult to access and publish (due to confidentiality issues). Most information is available at an aggregate level through the statistical offices of international organizations such as the United Nations Conference on Trade and Development (UNCTAD) and the Organization for Economic Co operation and Development (OECD), and national stat istical offices. Yet, even at this level data tend to differ markedly, due for instance to other ways of defining territorial ownership (e.g., based on the company regis tration vs. the owners' nationality), and to varying solu tions to cope with different systems of accountancy and taxing. Such problems need to be taken into account when interpreting FDI data.

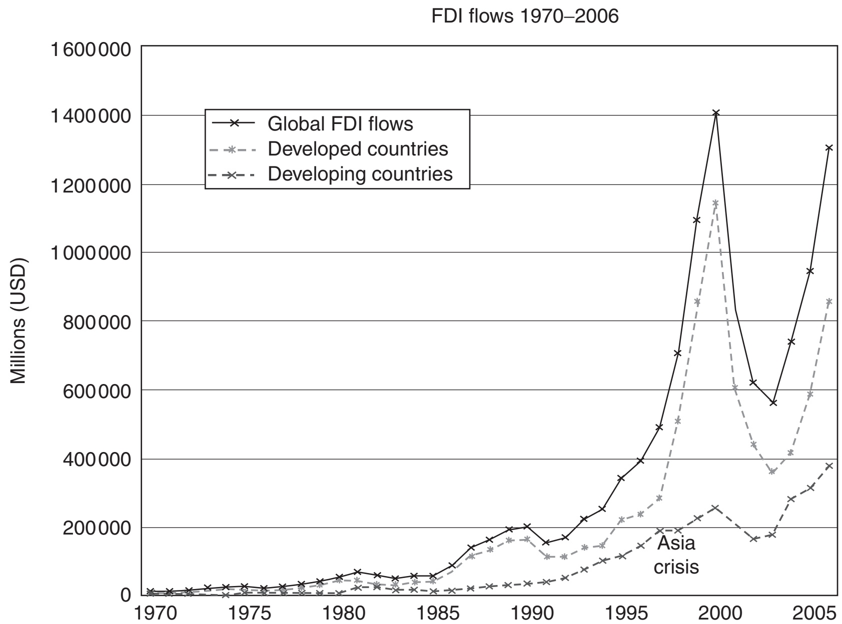

Figure 1 FDI flows, global, developing, and developed countries 1970 2006. Source: UNCTAD FDI database 2007.

Explaining the Flow

FDI flows are, in part, triggered by the same factors as international trade and flows of technology, labor and finance. Given the present distribution of resources, markets, and financial conditions across space, a profit can be made by establishing a transnational transaction. From a locational perspective, therefore, there are three fun damental explanations: (1) market opportunities abroad, resulting in 'horizontal' expansion; (2) the availability of (cheaper and better) resources (material, services, labor, or technology), leading to 'vertical' expansion; and (3) the exploitation of differences in financial conditions and investment opportunities (as manifested by divergent interest rates and exchange rate fluctuations). Initial ex planations of FDI focused, in particular, on market op portunity and resource availability invoking the concept of the 'product cycle' (see **Figure 2**). According to this concept, each phase in the development of a product would correspond to a particular set of spatial strategies. Initial product development will generally take place within the context of a single home country because of the need of proximity to knowledge producers and ad vanced markets. In the next phase, product maturation opens the window for market expansion abroad. Finally, triggered by the rise of competition, further standard ization of production allows for the relocation of pro duction to areas with low cost (and low skill) labor and cheap resources.

While necessary, locational factors and endowments do not suffice to explain FDI. It is not only the trans national transaction that needs to be clarified, but also the specific form of governance adopted. Why is FDI pre ferred to exporting or the licensing of the use of tech nology and brand names to a local firm? Why is, in other words, the transaction 'internalized' within a (multi national) company? The international economist John Dunning addressed these questions in his seminal 'eclectic' approach that combines market based and firm based perspectives. His approach specifies the way the three sets of components work together in the shaping of particular flows of FDI (see **Table 1**).

First, the abovementioned factors were subsumed under the rubric of 'location specific advantages'. In addition to markets and labor, two types of advantages of growing significance are, on the one hand, technology and knowledge, epitomized by 'clusters of innovation' and, on the other, concentrations of advanced producer (notably financial) services (increasingly confined to the world's major urban agglomerations). Much in line with the 'product cycle' approach, the search for labor, tech nology and capital plays a dominant role in specific sectors in certain phases of development. For more 'standardized' products and process, the continuing search for low cost production sites ('offshoring') remains an important part of the battle for (global) market shares. For 'high tech' activities, on the other hand, inter nationalization has become more important than in the past. Since most 'high tech' sectors tend to manifest strong concentrations of knowledge dynamic in different hot spots around the world (like ICT, multimedia, or bio tech clusters), it is important for both upcoming and established firms to have roots in these various locations. Anchoring in various 'places to be', as a strategy to

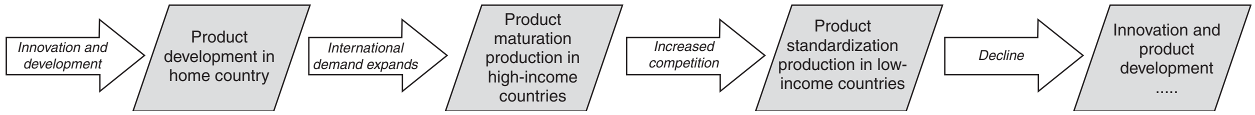

Figure 2 Product cycle and spatial strategies of MNCs. Source: Vernon, R. (1966). International investment and international trade in the product cycle. *Quarterly Journal of Economics* 80, 190 207.

Table 1 Examples of location-specific, ownership-specific, and internalization advantages

Location-specific advantages	*Ownership-specific advantages*	*Internalization advantages*
Accessibility and proximity to clients and domestic markets	Access to markets and transnational clients	Quality control over international standards
Access: need to be close to local media	Access to international capital and financial markets	Cognitive market failure and asymmetry of knowledge/ fear of underperformance by licensees
Availability of skilled/unskilled labor	Supportive role of government	Need to pursue international investment strategy
Government regulations	Goodwill	To reduce transaction costs with foreign agencies/ coordination gains
Adaptation to local standards and procedures	Economies of size and specialization	Managerial capacities
Customization to local tastes and needs	Financial strength	Protection against exchange/ political risks
Suppliers (e.g., advanced producer services)	Control over standards	Importance of international arbitraging
Agglomerative economies (knowledge)	Modern management practices	In case of core assets, need for protection from dissipation
Oligopolistic strategies among dominant domestic companies	Knowledge intensive technology	
	Professional skills	

Source: Adapted from Dunning, J. H. (1993). *The Globalization of Business*: *The Challenge of the 1990s*. London: Routledge.

search for, rather than exploit, knowledge, thus also triggers FDI, often in the form of joint ventures.

Second, companies possess certain advantages, such as technology, brand names, and marketing knowledge, that give them an advantage over 'local' companies in host areas, so called 'ownership specific advantages'. 'Owner ship specific advantages' mark out what a parent com pany intends to exploit by going abroad. If this entails a strong brand name or a technology well protected by a patent, it is easier to use arm's length contracts (licenses and franchises) than when advantage rests with more tacit forms of marketing or production knowledge. When, on the other hand, a company is thriving primarily on managerial skills and financial assets, 'in house' expan sion through the extension of managerial control is an obvious choice. Indeed, the 'excess' of managerial and financial capacities is generally seen as a major driver in the transformation in business governance from unitary (business departments like R&D, marketing, and finance) and functional (task units) to multidivisional and multi national (organizational and spatially decentralized product divisions). In knowledge intensive sectors, moreover, it is not only the exploitation of ownership specific advantages, but also the wish to 'improve' such advantages that may trigger foreign investments and ventures targeting technological hot spots.

Finally, companies want to employ their ownership specific advantages in the most (cost) effective and secure way. A problem is that international markets generally do not provide suitable mechanisms to meet these criteria through transactions at arm's length. Mechanisms such as exporting, licensing, and even joint ventures fail to pro vide a (cost) effective way to employ ownership specific advantages in a targeted area. The investor thus has to resort to an 'in house' solution, reaping the benefits from what are called 'internalization advantages'. What is particularly difficult, when crossing judicial boundaries, is to design a governance structure or contractual ar rangement at arm's length that prevents the undue ap propriation of profits or the unwanted spillover of proprietary knowledge. Firms thus opt for 'full' control by establishing or acquiring a subsidiary abroad, or sometimes by creating a joint venture under close supervision of the parent company. The option of a joint venture, in which the acquisition of foreign assets is only partial, is sometimes preferred when targeted location specific advantages of the targeted area are owned by a local company. An example is when a local company has built a strong local marketing network (say in food re tailing) or complementary knowledge (say in medical technology) that a parent company would like to use. By retaining the current managerial control of the targeted company, the preservation and development of such assets may be secured better (and more cheaply) than by full acquisition.

Of the triple set of ownership specific, location specific and internalization advantages, the third/last set produces the most complex picture. In addition to the reasons already given for the lack of market based solutions – no effective governance structure at hand and lack of judicial backing for cross border transactions – nation states play an important and often ambivalent role. On the one hand, nation states encourage FDI by creating barriers to other forms of transactions, notably trade. When a parent company cannot export its final products to a host market due to trade restrictions, it needs to resort to import substitution by investing in local production. When, furthermore, the local assembly of a parent company's product is subjected to a minimum amount of 'local content', also a certain amount of components needs to be bought or produced in the host country. A more subtle form of limiting the trading of final or intermediate products is through voluntary restraints. This has been adopted, for instance, by Japanese companies selling in the US, which induced substantial flows of direct investments from Japan into the US. Which mode of import substitution is chosen will depend on other conditions, such as the availability of companies to buy or to work with in a joint venture, the kind of knowledge and assets that are to be transferred, the incentives for investing in greenfield locations and the broader macroeconomic, political and legal context. It is through the last two categories that the state plays additional direct and indirect roles. Much FDI has been directly attracted by a range of state subsidies, investment support measures (like the building of infrastructure, business parts, training programs, etc.) and tax benefits, including the establishment of free trade zones (or export processing zones, EPZs). However, what has often received less attention is the significance of indirect factors that bear upon FDI, including monetary and fiscal stability, political stability and an adequate protection of proprietary knowledge. The latter is associated not just with patent and (intellectual) property rights, but also with its broader legal and social–political system and the general attitudes toward private property. For high tech companies, an adequate level of property protection is, in general, a necessary condition for any kind of involvement, whether contractual or in house. While state induced barriers to trade and investment incentives may encourage FDI, the lack of legal protection of assets and knowledge – or the perceived possibility of that in the near future – tends to deter investors.

'Internalization' also refers to the way companies form part of 'webs of enterprises'. Over time, firms have developed ever more sophisticated forms of trading and contracting with other companies, in the search for the optimal balance between control and flexibility. The result is highly complex vertical chains, in which the relationship between primary suppliers, providers of parts and services, main assemblers and brand owners, and distributors is governed by a myriad of (hybrid) market, network and ownership relationships.

This (inter)organizational competitive world compels companies to swiftly respond to and anticipate each others' moves in the search for the best and cheapest resources and suppliers, new market opportunities (and their point of access) and seedbeds of new ideas and practices. The playing field for this competitive game is essentially spatial. At a 'global' level, companies are chasing each other in the race for similar gains and pole positions. The result is that, notably in mature phases of the industry cycle, waves of FDI inflow are concentrated in particular areas, because of perceived market opportunities or cost advantages. Over the last two decades, examples of such FDI 'growth peripheries' include countries integrated within the EU (e.g., Southern and Eastern Europe), 'opening' economies in Latin America (e.g., Chile and Brazil), Asia (Southeast, and recently China and India) and even Africa (South Africa, Egypt, and Nigeria). However, also the more central areas of the EU and North America host hot spots for investment from each others' as well as Japanese investors. At a more 'local' (national or regional) level, however, investors often seek to carve out their 'own' space of production, establishing exclusive access to a local pool of labor and suppliers, and to institutional resources (such as land use and other forms of regulation and support). Waves of investment thus result in the spread out of subsidiaries over a large space, each creating their own domain of control. Such patterns are particularly manifest in sectors with a preference for greenfield locations, such as the inflow of FDI in the automotive sector in countries like Spain (with presence of nearly all major car companies dispersed through the country), USA (European and Asian investors), Eastern Europe and China. FDI, accordingly, may induce dramatic changes in regional economic divisions of labor within and between countries. Only when there is a stronger inclination toward M&A investments, or when the main incentive is the search for technology, FDI tends to concentrate in major industrial agglomerations, reinforcing existing regional economic geographies.

Local Impacts

The impact of FDI on a host region is threefold.

First, there is a 'direct' effect resulting from the inflow of capital, assets and technology, and the recruitment of local staff, boosting employment. The employment impact can be gauged in quantitative and qualitative terms.

The first refers to questions like how many (full time or equivalent) jobs are created, at what pay levels and how many of those are taken up by 'local' job seekers. More qualitative issues are: what kind of skill levels are demanded, what kind of training is provided, and what kind of contracts are offered (e.g., part time/full time, open ended/fixed term, flexible, etc.).

Second, there is an 'indirect' effect caused by the demand for intermediate goods and services, producing an additional stream of finance, technology and demand for labor. This effect is confined to the supply chains in which the foreign investor is locally inserted, and is generally calculated through a kind of input–output analysis on a detailed level of value added flows.

Finally, there is a broader economy wide 'induced' effect caused by the consumptive, and potentially productive, expenditures by the labor force working within the subsidiary and its supply chains, as well as the spillover of technology. Income effects can be measured through input–output analysis that encompasses the backward links of the household sector. Technology and knowledge effects can either be assessed through case studies and surveys, or through advanced forms of modeling. So, taken together, such linkage dynamics may underpin virtuous cycles of investment, knowledge development, and strong export growth, resulting in local cluster formation.

What compounds these various effects is that they partly depend on the state, notably through the raising of direct, indirect and company taxes. Considerable research has been undertaken to measure and explain these various impacts, although most has been confined to either macroeconomic econometric models or to single country case studies. Apart from the more trivial observations, such as that local content requirements are generally accompanied by a higher level of in house production or local procurement, or that greenfield investments create more jobs than comparable M&A activities, empirical studies show rather mixed results. Given the fact that FDI generally unfolds as an 'establishment chain' taking place over a long span of time, driven by highly variable business strategies, locational opportunities and state practices and strategies, this should come as no surprise. A final reason why it is so difficult to pin down the impacts of investment is because of the lack of a contrasting case, that is, a comparable setting in which a certain investment has not taken place. Moreover, in exploring the role of impact, the interest has shifted somewhat from impact on employment and capital flows to technology and knowledge flows, items which are more difficult to measure. Such shifts reflect changes in prevailing academic and policy agendas.

Although the focus has tended to shift, employment remains one of the most visible and sensitive issues. Especially in the more traditional sectors of resource extraction and manufacturing, and also in back office activities, foreign investment often takes the form of large establishments, such as mines, automotive or electronics assembly plants, or office towers, generating large numbers of jobs. As a result, in terms of numbers, and primary (wage and contract) as well as secondary (benefits) labor conditions, foreign investors are able to put their mark on a local (or even national) economy. When the main motive for investors is the search for low cost resources, they are often accused of exploiting a dependent labor force or of depleting natural resources. The evidence for such exploitation is not very strong, however, at least not in a direct sense. While obviously investors will benefit from lower wage levels in host territories, from a local perspective they tend to be good employers, offering relatively high salaries and good primary and secondary working conditions. In general, additional investments are made in training and support such as in health, housing and transport. It is not in the 'direct' but in the 'indirect' impact of foreign investment where stronger evidence of exploitation is found. Overall, wage levels and working conditions in the supplier firms tend to be much less advantageous than in the foreign subsidiaries. In some cases, however, notably in the textiles and footwear sector, conditions can be truly appalling. Workers face very long working days, under poor working conditions, lacking any basic support or security. A very serious issue is the use of child labor. The critical question is, to what extent foreign subsidiary deliberately or at least knowingly benefits from, or even creates, such conditions through forcing down the prices of intermediate goods and services. What follows is a deeply ethical question of as to what extent foreign investors, when causing indirect job formation, can be held responsible for the qualitative aspects of jobs that are organizationally and spatially distant.

Impacts on 'capital flows' strongly depend on the various financial decisions and steps made along the establishment chain. While FDI, by definition, comes with an investment from the parent company into productive assets in the host region, some capital can also be raised locally. In the latter case, an increase in the capital stock under foreign control is achieved without an inflow of FDI. Because of the volumes involved, there is a possibility that this reduces the possibility for other investments. Such an allocative effect also applies to state finance, when the state provides subsidies (either direct or in the form of infrastructure and supporting facilities) to the investor. Subsequently, once a facility becomes (again) fully operational under foreign control, the distribution of profits becomes a thorny issue. How much is reinvested locally? How much is locally retained in the form of savings of consumption? And how much is 'repatriated'? Because the latter is generally subjected to tight forms of state regulation, this points not only to formal and direct transactions, but also to more covert ways through which parent companies seek to bank in on their activities abroad. Further down the establishment chain, the continuity of the asset structure of the sub sidiary is at stake. Further investments can be made, based on either FDI, the recycling of profits or the raising of local capital. It is also possible that assets are sold with the gains repatriated. The result is an outflow of capital and a corresponding reduction in the capital stock under foreign control. When they result in the closure of an entire plant or industrial complex, such disinvestments may cause societal and political upheaval such as that occurred around the dramatic closures of automotive plants in Belgium. In many cases, however, disinvest ments occur in a step by step manner, resulting in a gradual downgrading of a facility in the overall corporate structure of the parent company.

The debate on local impact is strongly centered on the role and transfer of 'technology'. For recipient areas, foreign investors are seen as the harbingers of efficiency, new technology and knowledge, and possibly local in vestment in R&D. Much research has gone into the question of to what extent FDI has brought on a 'spill over' of knowledge and technology into the local econ omy. Whether such an effect can happen depends, in the first instance, on the extent to which FDI is accompanied by the inflow of technology and knowledge that could be of use to local companies and labor. This is only the case when there is not too big a gap between the kind of productive activity established by FDI and activities already existing in the host area. There must be, in other words, sufficient 'absorptive capacity' in order to benefit from the inflow of technology and knowledge. When a state of the art car assembly or electronics plant is erected in a largely rural, peripheral area, or in a backward industrial area, the scope for technology 'copying' or the emulation of 'global' business practices by local companies is minimal. In general, absorptive capacities tend to be better in open economies with a higher level of industrialization. Nevertheless, even in more closed and backward areas, certain mechanisms of spillover might still materialize, for instance, through in company training of local employees and management training provided to local suppliers. Spillover may also be constrained by a 'lack' of a knowledge gap. Cross border investments that are primarily geared to the ex ploitation of the ownership of brand names, for instance, often do not carry new technologies with them (save perhaps new marketing practices), comparable to li censing agreements. In some cases, such companies in vest in local R&D, in order to tailor to local market conditions, but this will be of little relevance to local companies. Indeed, what results then is often a reverse flow of knowledge: the subsidiary provides the parent company with strategic knowledge about the host market, including information on competitors. Such a reverse flow also occurs when upcoming firms in high tech sectors invest in 'antenna' affiliations in the world's technological hot spots.

A crucial factor in creating spillovers is the capacity of the foreign investor to control for 'leakage'. As argued by internalization theory, FDI is often preferred because it provides the most secure vehicle to prevent the spillover of knowledge. However, this protection only bears on those ownership specific advantages that a company sees as the core of its competitive edge and that are strategically employed in a cross border transaction. Nevertheless, there may still be spillover from other types of knowledge a company possess, such as best practices on supplier linkages, marketing, human re source management (HRM), financial management, etc. When such knowledge is nonstrategic to the foreign company and complementary to host companies, spill over may be seen as beneficial to all parties, notably in the context of supplier links. Especially large multi national companies often have a keen interest in edu cating local workers as well as local companies through dedicated training programs. Not only does this yield direct economic benefits, it is also a sign of goodwill and hence a source of reputation for the investor. Particularly in the case of large investments with prolonged estab lishment chains, continued engagement of a subsidiary within the local economy may yield social benefits to the community even in excess of economic benefits to the investor. Not surprisingly, there is much interest from researchers as well as policymakers to the way sub sidiaries become 'embedded' in local economies and play an 'educational' role.

The extent to which firms can really control pro prietary knowledge – especially when not protected by regulatory systems such as patenting or intellectual property rights (IPR) – remains a subject of debate. Nevertheless, what is clear is that for all firms in cap italist economies, finding the right balance between deploying and protecting their assets is critical to their survival and growth. How this is done, furthermore, depends on the organizational structures and govern ance practices of the corporate organization. Sub sidiaries under tight central control tend to be more protective and less venturing than subsidiaries enjoying a considerable degree of local autonomy. Some auton omy is essential, indeed, for the build up of local research and development activities. The extent of au tonomy is not a matter of principle, however; it is a strategic function determined by the corporate organ ization. By granting subsidiaries their own space for development, in an organizational as well as spatial sense, they can act as sites of learning and experi menting. Learning, on the one hand, often involves the acquisition of knowledge on local market and pro duction conditions, allowing subsidiaries to tailor their activities to local circumstances and the corporate organization to improve the company's international division of labor. Experimenting, on the other hand, involves developing and trying out new products and production practices (techniques, labor relations, logistics, procurement, marketing, etc.) as part of the wider corporate innovation strategy. In home countries, such experiments are often difficult to instigate because of institutional, cultural and organizational impediments. Seemingly paradoxically, it is in 'far off' host locations that more radical forms of change can be practiced. Examples of such forms of 'distance' learning are the ways in which US and EU automotive and electronics companies have experimented with 'Japanese style' 'just in time' procurement and new employment practices in more peripheral locations. Currently, 'emerging' economies in Asia and Latin America present seedbeds, notably for product development. However, despite strategic intentions, attempts to bring home the message have often been met by the same kinds of resistance that impeded change in the first place. So while major benefits in terms of (direct and indirect) employment growth, production growth and knowledge spillover accrued from such experiments locally, to the advantage of the host region, effects for the home region often proved to be less than expected. What this shows is that impacts in host areas cannot be seen in isolation from the overall corporate strategies and practices.

Besides spillovers, finally, there is also another knock on effect with wider implications on host regions' economies. By settling themselves in a local production structure, and by becoming active in a local market setting, foreign investors impose 'global' standards upon a local economy. Especially when the host economy is relatively open, such standards, and the competitive setting in which they are employed, will provide a major incentive to local companies to improve productivity and raise their standards. The latter may be effectuated not only by spillover of knowledge from the foreign subsidiaries, but also by other means of knowledge acquisition and incentives. Indeed, the mere 'exemplary' function of foreign subsidiaries, and the capacity to break open oligopolistic market structures and induce competition, is sometimes deemed more important than the channeling in of knowledge. Such a 'bootstrapping' effect can only be successful, however, when local firms have the time and space to adjust, for which some degree of temporary protection and support may be warranted. An accompanying approach may be to foster the move of domestic capital and workers from affected, less productive to other more productive sectors through a program of structural readjustment. Overall, the evidence on to what extent and in what ways local companies are able to benefit from FDI remains rather mixed and ambiguous.

National Policies and Regulations Targeting FDI

Given the significance of FDI and the scope for local benefits, FDI constitutes a specific target of national policymaking. In an increasingly integrated world, foreign investments are seen as a core vehicle to create jobs, value added and knowledge. In sectors with a high presence of multinational companies, FDI may effectively represent the primary source of productivity, innovation and market access. National and regional economies specializing in these sectors have thus become more and more dependent on foreign investments. State authorities are making strenuous efforts to lure investors to their territory, to retain them and to maximize the benefits from their activities. This dependency is not one directional, however. Once 'bolted down', a foreign investor also depends on support from the state and other organizations (such as unions). The support an investor receives is always politically contingent, depending, among others, on issues such as the position of other foreign and domestic companies, prevailing views on foreign ownership and the protection of the 'national interest', and the development of the state regulatory system. One could say that, after each investment flow, the recipient state and the investing company are capable of holding each other at ransom. While, at first glance, the investing companies, as vehicles of global capital, may be the dominant partner, once established and embedded, the state can exert considerable influence on local operations. It is in this political–economic context that the role of FDI policies and regulation should be assessed.

Policies and regulatory measures aiming at attracting FDI can be divided into two groups. On the one hand, there are 'specific' policies that are designed to attract and support targeted investments and improve their local impact. This includes direct subsidies and incentives (tax reductions, exemptions from trade regulation, etc.), as well as indirect forms of support and privileges (provision of infrastructure, training of the company's workforce, benefits to suppliers, etc.). Certain countries provide such benefits in the form of a package in spatially delimited EPZs and the so called 'offshore financial centers' representing an exclusive regulatory space separated from the national economy. Specific policies can also direct and restrict local operations through formal or informal requirements to undertake certain productive or support activities, to locally procure certain items, or to transfer certain types of technology.

On the other hand, there are 'general' policies that change the context for productive activities under foreign control. This includes monetary and fiscal policy (setting of exchange rates), trade policy (notably imposition of trade and nontrade barriers to imports), industrial policy (support for sectoral development to businesses and labor and research centers) and dedicated investments in infrastructure, education, etc. It is through these general policies that certain critical factors for the attraction of FDI and the enhancement of local benefits can be targeted. Industrial and education policy can help to improve the local economy's absorptive capacity expanding the scope for spillover and indirect growth effects. A well attuned industrial and trade policy can help to foster openness and innovation while giving local firms the opportunity to adjust. Monetary and fiscal policies can help to create a stable financial environment both within the national economy and vis à vis other major markets, as to provide a secure basis for exports.

In addition to these incentives, governments use a range of measures to force foreign investors to comply with certain performance requirements and trade restrictions, the so called trade related investment measures (TRIMs). TRIMs include policies aimed at bolstering FDI spillovers by creating incentives, regulations and import barriers to stimulate sourcing from domestic producers. The idea behind such policies is that active government involvement and regulation is necessary to achieve spillovers and domestic economic development from FDI.

FDI policies and regulations can also be classified based on the extent to which they are market conformable or not, that is, to the extent they are decreasing or increasing market imperfections. Because the way FDI is affected by location specific and internalization advantages, both directions may effectively encourage as well as discourage FDI. On the one hand, 'increasing' market imperfection, manifested for instance in a lack of competition policy, the privileging of particular national champions, a deficient judicial regime, or an instable currency, will enhance the value of internalization advantages and hence a preference for FDI. Such imperfections will have negative impact on localization advantages, however. So if market imperfections become too stifling – for instance, in the form of strict ownership, price and remittance controls – this may tip the balance and put off investors. On the other hand, various methods of improving the operation of markets, and hence the 'reduction' of market imperfections, may induce more FDI, now pulled by localization advantages. Such methods include the liberalization of trade and financial flows, deregulation of markets, antidumping measures, privatization of utilities and strict enforcement of (intellectual) property rights. Again, there is also a downside to such methods because they tend to lower the significance of internalization advantages so that firms may opt for trade or licensing instead of FDI. While both increasing and decreasing market imperfection may boost FDI, the nature of such investments differs markedly. Increased market imperfections will appeal to low tech production destined for the local market, in which a company can extend the lifeline of obsolete assets. A more open market, in contrast, is likely to attract export oriented producers which will be compelled to import and use state of the art technology. In other words, where the first case will only yield direct, and some indirect employment, the second will provide scope for knowledge spillover. To what extent spillover really occurs depends on the absorptive capacity of local companies and the inclination and competencies of employees working in foreign subsidiaries to spin out and establish new firms. The scope for the latter will, in turn, depend on the extent to which general policies in the field of education, sectoral development, and finance are able to support domestic economic activities.

Global Regulation

This 'carrot and stick' nature of FDI incentives and regulations reflects the schizophrenic nature of the relationship between national governments and FDI. It is also one of the most controversial issues in the debates on global FDI regulation, even for the free market organizations such as the World Bank, International Monetary Fund (IMF), and World Trade Organization (WTO, successor of the General Agreement on Tariffs and Trade, GATT). On the one hand, these organizations hold that only an 'open economy' route focusing on improving the general conditions for both foreign and domestic investor will produce a long and lasting impact. Notably, in its recommendations for developing countries, the IMF has called for an approach that improves the investment climate in a fundamental, nonselective manner, for all kinds of capital. In line with a neoliberal doctrine, structural change opening up domestic economies is preferred to, and considered much more effective, than incentives targeting specific firms. The IMF claims, more specifically, that investing in a sound regulatory environment, including the build up of credible enforcement mechanisms, is much more (cost) effective than policies and incentives designed to lure FDI. It may, in particular, prevent 'races to the bottom' in which foreign investors play out different (national and regional) authorities against each other in seeking a new investment site.

What is not denied, on the other hand, is that state policy and action need to respond to more than calls for economic efficiency. In particular, with respect to FDI, tensions may arise between the aim of economic efficiency and the protection of national sovereignty. Regardless of the openness of their economies, all nation states conceive certain economic activities as 'strategic', to which access from abroad should be restricted. Such restriction can be based on the view that activities are of military significance, or that they draw on national resources of which the benefits should accrue primarily to domestic actors and the state. The latter includes energy resources, minerals, and fishing grounds. Indeed, while there may be a general trend toward liberalization, restrictions on FDI within activities deemed 'strategic' remain rife. In the developed world, the fear for terrorist attack has produced a significant growth in restricted practices. In the developing world, on the other hand, state authorities continue to nationalize the assets of foreign companies in the name of the 'national interest' notably in extraction activities. Such expropriations are, at best, partially compensated for, thus posing major risks to companies with a high exposure in politically less stable countries.

Despite the perceived need, a regulatory framework concerning state policy and FDI at the global level is actually still wanting. What does exist today is an intricate web of bilateral, regional, and multilateral agreements, which operates in a rather uncoordinated and restricted way. Most of the international multilateral agreements on FDI focus on the restraints and requirements imposed upon MNCs by national governments. They particularly address the trade barriers and performance requirements (TRIMs) aimed to capture FDI spillovers and to protect national economies from negative influences of MNCs. Other agreements (bilateral and regional agreements) focus upon the protection of international property and the attraction of FDI. The debate on the international regulation of FDI and TRIMs started in the early 1980s, when protective policies and performance requirements became increasingly criticized by investors and MNCs for significantly distorting global trade and investment flows. Supported by the governments of the industrialized countries (particularly the OECD countries), those critics sought to develop an international system of regulation which would achieve a 'level playing field' in which foreign investors are accorded the same treatment as domestic businesses. At the same time, the growing belief that FDI had a significant positive impact upon national economies increased efforts by national governments to attract FDI. To prevent a 'race to the bottom', in which no country would benefit, it was also argued that an international regulatory framework for FDI was needed. So far, such frameworks have been developed notably at the level of regional common markets such as the North American Free Trade Agreement, APEC, the Mercosur and the EU economic zone. Agreements here generally intend to bolster FDI inflows by establishing a secure multilateral economic policy environment as well as providing access to a larger market. Recent studies on the role of regional cooperation have shown that the level of FDI typically increases when countries join common markets.

Less successful have been the attempts to build a regulatory framework to govern FDI on a global level, as manifested by the problems faced by the WTO in its pursuit for a new agreement on TRIMs. So far, the TRIMs negotiations did not manage to yield more than a weak compromise. The main result of the TRIMs agreement within the WTO was that it affirmed that already existing GATT/WTO regulations now also applied for FDI. The most pressing problem in establishing a global regulatory framework for FDI remained the issue of to what extent FDI regulation should reconcile 'open economy' aspirations aiming at the free flows of goods and capital with the wish to protect the particular interests of states, labor and the environment.

Whereas most developed countries aim to support their 'national' industries and firms by arguing for a free and open global economy, many of the developing countries are hesitant to enforce policies that aim for trade liberalization. Besides the critique that trade liberalization is often rather one sided (aimed at the singular opening up of developing economies), the incompatibility of liberalization policies with the more protective import substitution developing strategies of the developing countries remains an important point of conflict. Other groups, such as labor organizations and nongovernmental organizations (NGOs), are also critical of the 'open economy' aspirations of developed countries. Labor organizations, such as the International Labor Organization (ILO), the World Federation of Trade Unions and a wide variety of national labor unions, refer to the negative effects that free trade and investment have upon labor standards. They fear that an increased race to the bottom over FDI might lead to the 'social dumping' of international workers' rights and labor standards. Environmentalists and environmental NGOs share similar fears for the environment. They argue that protective policies and regulations at both national and international levels are necessary to prevent pollution and environmental deterioration.

As a result of these tensions, the most recent attempt to establish a multilateral regulatory framework for FDI culminated in a complete breakdown of the FDI regulation debate. In the late 1990s, the OECD started negotiations on a multilateral agreement on investment (MAI). The goal was to provide a strong framework for FDI regulation which had market liberalization, the protection of (foreign) investment and international dispute settlement as central issues. The MAI would become a counterpart of the GATT/WTO, where GATT would provide the rules and regulations for international trade and the MAI would provide the rules and regulations for international investment. This free standing nature of the MAI was underlined, in particular by the US, and it was stated that the MAI would be open to all OECD and non OECD members could also enter. The choice for the OECD as the setting for the negotiations was quite controversial since it excluded the participation of the majority of developing countries. However, OECD members argued that since 85% of all FDI flows were constituted by OECD member states, the negotiation of the MAI within the OECD would produce the best results. Besides that, the preference for the OECD as the main venue to discuss the MAI can also be connected to the disappointment of, primarily, the USA over the results of the TRIMs negotiations within the WTO.

The MAI generated a fair amount of criticism, in particular from labor and environmentalist NGOs, leading to many demonstrations and protests across the world. In countries such as Canada and France, MAI was seen as threatening local cultures and national sovereignty. This ultimately led to the withdrawal of France from the MAI negotiations in 1998. After the withdrawal of France many countries followed and in 1999 the draft treaty was rejected. Today, the TRIMs agreement within the WTO remains the primary regulatory framework for FDI on the multilateral front.

Outlook

Given recent setbacks on the road toward 'full' liberalization, regulatory aspects will remain a key issue in the debate and research on FDI. From an economic geography perspective, this issue has received little attention. There is considerable scope for exploring the links with political geographies of regulation and work on power geometries. Research could focus, in particular, on the persistent, yet changing role of legal borders, the way these borders delimit 'spaces of regulation' and how this is changing under the impact of new forms of global and regional (suprastate) agreements. Financial, resource (including labor) and production settings remain strongly spatially differentiated, with nation states continuing in pole positions. As a result, trends toward more integration and global standardization are far from automatic, but present, to a large extent, the hard one results of regulatory struggles and arduous institutional building. This will continue to provide a major research agenda for the future.

So far, economic geography has made more inroads in the debate on location and organization strategies, in the context of globalization and technological change. A fruitful alliance has developed, accordingly, with the field of international business studies. Economic geography has gained a more dominant role because of the sophistication that firms have built up in the slicing of activities over space. Not only location strategies, but also ownership (technology and capital) and internalization (control) strategies have evolved quite dramatically in response to the need to operate flexibly in an increasingly volatile global context. While externally firms have developed refined forms of outsourcing, licensing and joint ventures, internally manufacturing and supply structures have been further opened up. This has produced a shift further away from hierarchical control systems toward network structures, further blurring the boundaries between internal and external organization. The result is a separation between brand owners, the production network consisting of contract assemblers and part suppliers (engaging in what is called 'distributed manufacturing') and local distribution companies. Such network configuration allows companies to tap into multiple sources of innovation, including 'clusters of innovation' dominated by concentrations of (hi tech) small and medium sized enterprises (SMEs). Originally economic geography contributed mainly through insights on the spatial division of labor within large corporations and the mapping of location advantages across space. Geographers are able to bring in vital expertise on the spatial embedding and the development and impact of networks ('webs of enterprises' and Global Production Networks) to study these developments.

The emphasis on network does not make the older forms of location and impact analysis irrelevant, however. In particular, the relationship between FDI and innovation continues to present a major, and highly influential, research agenda. The combination of network perspectives and more classical location and impact analysis, moreover, provides a powerful apparatus to elucidate the role of FDI in unfolding geographies of innovation and production. FDI is global in reach but, as explained above, it (re)produces particular geographies, ranging from the bolstering of existing agglomerations to the establishment of new 'growth peripheries'. FDI contributes, in particular, to the webs of relations 'between' growth centers and peripheries and hence to the archipelago structure of today's global economic geography.

See also: Agglomeration; Global Commodity Chains; Globalization and Transnational Corporations; Globalization, Economic.

Further Reading

Buckley, P. J. and Ghauri, P. N. (2004). Globalization, economic geography and the strategy of multinational enterprises. *Journal of International Business Studies* 35, 81 98.

Dicken, P. (2007). *Global Shift: Mapping the Changing Contours of the World Economy* (5th edn.). New York: Guilford Press.

Dunning, J. H. (1993). *The Globalization of Business: The Challenge of the 1990s*. London: Routledge.

Hoekman, B. M. and Kostecki, M. M. (2001). *The Political Economy of the World Trading System: The WTO and Beyond*. Oxford: Oxford University Press.

Lagendijk, A. (1995). The foreign takeover of the Spanish automobile industry: A growth analysis of internationalisation. *Regional Studies* 29, 381 393.

UNCTAD (2007). *World Investment Report 2007: Transnational Corporations, Extractive Industries and Development*. United Nations Conference on Trade and Development, Geneva.

Vernon, R. (1966). International investment and international trade in the product cycle. *Quarterly Journal of Economics* 80, 190 207.

Relevant Websites

http://www.twnside.org.sg/
Anti/alter globalist thirdworld network.
http://www.citizen.org/trade/issues/mai/
Anti/alterglobalist website on the MAI.
http://www.imf.org
International Monetary Fund.
http://www.oecd.org/daf/mai/
Official OECD account on the MAI.
http://www.oecd.org
Organisation for Economic Co Operation and Development.
http://www.unctad.org
United Nations Conference on Trade and Development.
http://www.worldbank.org
World Bank.

Foucauldianism

M. Huxley, University of Sheffield, Sheffield, UK

Glossary

Archaeology A historical method for examining systems of thought and regimes of truth in a given society at a given time.

Biopower Management of the biological and demographic aspects of a population as a whole – knowledge of, and attempts to regulate, rates of reproduction (births and deaths), health and disease, and risk and accident.

Discipline A form of power producing and regulating normalized bodily comportments and individual behaviors.

Discourse Archivally identifiable collections of statements and descriptions of practices connected in their expression of underlying historically contingent 'truths'.

***Dispositif* (Variously Translated as: Deployment, Apparatus, Arrangement, and Disposition)** Network of spaces, plans, buildings, institutions, practices, regulations, philosophies, laws, discourses, and associated with particular configurations of power, knowledge, and government.

Genealogy A method for bringing to light the exercises of power involved in the installation of particular regimes of truth and the devalorization of subjugated knowledges.

Governmentality A dispersed form of power aiming to act on the actions of others and to incite individuals to govern themselves; techniques of government of individuals and groups.

Heterotopia A form of space different from the everyday routinized spaces surrounding it and embodying potentials for subverting dominant *dispositifs* of power, knowledge, and space.

Panopticism Bentham's Panopticon as the 'diagram' of a form of 'disciplinary' power that incites subjects to monitor their own behaviors: an element of a *dispositif* that exemplifies modern forms of subjectification of the individual.

Power (*pouvoir* – to be able) As a productive network that runs through the whole social body; contrasted with, distinct from, *puissance* ('power over'; sovereign power), and domination and repression. (Productive) power produces discourses, knowledges, and subjectivities.

Power/Knowledge (*Pouvoir/savoir*) The interrelation between power and knowledge, so that knowledge is made possible through the operation of power in the production of dominant discourses, while power is exercised through the appropriation, distribution, or restraint of knowledge. (*Savoir* - 'know-how' embodied not only in practices and institutions, but also linked to theoretical texts.)

Subject, Subjectification The subject is the historically contingent product of technologies of governmental power that operate to install truths about the self. Processes of subjectification work to form particular subjectivities in individual human entities. Subjects also have capacities to carry out work on the self through knowledge and self-awareness.

Introduction

Michel Foucault's (1926–84) works are diverse and wide ranging, appearing in dispersed forms (books, articles, lectures, and interviews), and defying easy categorization. They could be classed as the works of a philosophically inclined, sociospatial historian, but such a clumsy designation barely hints at their richness and complexity. These works engage with histories, philosophies, and philosophical histories of modern knowledge and 'truth'; of science, psychiatry, medicine, and of the prison; of the state and its legal, administrative, and political institutions; of modern European sexuality, and Greek and Roman practices of self formation; and the ways that time, space, and architecture are intertwined in the interrelations of practices and truths.

In addition to this diversity of interests, Foucault continually reworked his ideas and revisited earlier studies from different perspectives, disrupting traditional notions of the progressive development of a body of work attributable to a singular 'author'.

Similar indeterminacy exists when attempting to locate Foucault's writings within a conventional spectrum of philosophical–social–theoretical perspectives: humanism–antihumanism; structuralism–post structuralism–postmodernism; social constructivism–phenomenology; behavioralism–psychoanalysis. This is a problem which derives from seeing his works as attempts to provide accounts of a social whole, and to see Foucault as a 'grand social theorist'. But while elements of all these approaches may be identifiable in Foucault's work, Foucault does not lay claim to the comprehension of a social totality.

> My work takes place between unfinished abutments and anticipatory strings of dots. I like to open up a space of

> research, try it out, and then if it doesn't work, try again somewhere else. … My books aren't treatises in philosophy or studies of history: at most, they are philosophical fragments put to work in a historical field of problems. (Foucault, 1991a: 73 74)

The difficulties in apprehending Foucault's work and its theoretical position are further exacerbated by the order and rate at which it has been published in France and elsewhere; the problems of translation, retranslation, and yet to be translated into various languages; and the ways in which it has been taken up at different times in different places and by different disciplines. 'Foucauldianism' continues to be, very literally, 'works in progress'.

Foucault's influence has been extensive but polymorphous – for instance, in France, his work is much more prominent in philosophy than it is in geography, while in Germany there is a wealth of Foucauldian inspired analyses of the state. In the English speaking humanities and social sciences, there is hardly a discipline that has not been touched, to some extent, by the debates around Foucault's ideas or their stimulus to research: from the history and philosophy of science to accounting; from the 'psy' disciplines to law; from cultural studies to economics; from political theory to architecture; from education to geography – to indicate but a few.

Questions of space are implanted in the breadth and depth of Foucauldian concerns, and the development of Foucauldian perspectives on 'the spatial', again, takes diverse forms. In Anglophone geography, interest in Foucault's work has grown since the early 1980s, when *Discipline and Punish* became the entry point for many geographical engagements with institutional spaces of discipline, control, and reform, especially through Derek Gregory and the department of geography at Cambridge University. Another influence was Ed Soja's book, *Postmodern Geographies*, which "gave prominent place to Foucault as a social and cultural theorist helping to correct the long standing marginalization of space in critical social theory" (Hannah, 2007: 91). These entry points have given a particular caste to the uptake of Foucault's works in geography – a point that is elaborated below.

The intellectual, temporal, spatial disjunctions and dispersions of Foucault's influence resonate with his conscious attempts to radically subvert conventional categories of thought, authorship, and academic disciplines. There are thus many 'Foucaults', and 'Foucauldianism' should not be understood to refer to any kind of endeavor to produce a unified theory of history or society. Nevertheless, it is possible to identify a number of themes that run through these works and their developments, and the next section outlines some of these key issues and their significance for geography.

Key Themes

Archaeology and Truth; Genealogy and Power

Foucault identified one of the continuing strands of his work as a 'critical history of thought' that sought to question the taken for granted status of knowledge at a given time in a given society, and to explore under what conditions something could become an object of knowledge about which true and false statements could be made. There are two aspects of such a critical history: archaeology and genealogy.

Archaeology, as undertaken in *The Order of Things* and reflected upon in *The Archaeology of Knowledge*, is an approach or methodological strategy which seeks to excavate the 'subsoil' of implicit knowledge in a society to expose the commonalities among disparate philosophical ideas, public opinions, commercial interactions, moral discourses, political practices, social theories, the regulation of institutions, and legal decisions. It is a "history of all that 'contains thought' in a culture, of all in which there is thought. For there is thought in philosophy, but also in a novel, in jurisprudence, in law, in an administrative system, in a prison" (Foucault, 1996: 18). Archaeology places theories, institutions, and practices on the same plane of analysis: it is a synchronic disinterment and assembly of layers of underlying knowledges in a society.

In this way, archaeology interrogates 'truth' as a historical category, a culturally specific system of rules through which true and false can be distinguished. Each society has its 'regime of truth' that enables distinctions between the Same and the Other; allows certain statements to be accepted or rejected; determines methods by which such assessments may be made; and valorizes the positions of those who are acknowledged as being able to enunciate the truth.

Genealogy, on the other hand, examines how knowledges and truths are linked to struggles over the exercise of power that sustain and extend, modify and overturn regimes of truth and networks of knowledge. These struggles are simultaneously discursive and material and are present in the mundane programs and practices of prisons and asylums as much as they are in the weighty debates of philosophers and politicians. In Foucault's rereading of Nietzsche, genealogy traces contingent, indeterminate 'descents', and 'emergences' of discourses and practices, truths and knowledges, rather than seeking 'origins' that respond to some predetermined necessity (such as 'the march of progress' or 'the needs of capital').

Foucault's major genealogical works are *Madness and Civilization*, *The Birth of the Clinic*, and *Discipline and Punish*. They are detailed historical studies which trace practices of psychiatry, clinical medicine, and penal reform as they produce and reproduce truths about the mad, the unhealthy, and the criminal, formulate divisions between them as normal/abnormal objects of knowledge, and locate them in relation to normative and spatial boundaries. These genealogies are neither narratives of great events nor histories of ideas: rather, they are illuminations of specific, historically contingent ways of knowing the truth; of commonplace, yet fateful practices, codes, and regulations that are, at one and the same time, the bases of 'truths' and reflections of them; and the local spatial and architectural forms – conceptual and physical – that instantiate and modulate such truths.

The deployment of knowledges and the circulation of truths involve the exercise of power. For Foucault, power is not a monolithic, monocentric imposition of re pression. It "traverses and produces things, it induces pleasure, forms of knowledge, produces discourse. It needs to be considered as a productive network that runs through the social body, much more than a negative in stance whose function is repression" (Foucault, 2001a: 120). Power and knowledge are implicated in incessant maneuvers, strategies, and tactics around truths and discourses, local knowledges and practices, and actions on the actions of others.

If archaeology is the historical examination of underlying 'regimes of truth' that make knowledge pos sible in a given society at a given time, and which brings to light 'subjugated knowledges', genealogy traces the ways in which dominant truths and knowledges are never stable or uncontested, but are constantly threatened by subversion and recalcitrance.

These are conceptions that underpin Foucault's studies of discipline, biopower, and governmentality.

Discipline, Biopower, and Governmentality

Discipline is a form of power focused on the body, which regulates bodily actions, comportments, and aims to in culcate productive, docile behaviors. It is most famously described in *Discipline and Punish*, Foucault's detailed historical account of the emergence of the modern prison in the nineteenth century. The dramatic opening de piction of the dismemberment of the body of Damiens the Regicide in 1757 is contrasted with a description of Faucher's prison timetable from 1838, which, Foucault argues, marks a historical shift in the treatment of criminals (and other categories of 'abnormal' individuals). These vignettes are given as exemplars of the contrast between sovereign power that uses violent punishment as a spectacular deterrent, and disciplinary power that seeks the reformation of individuals through incarceration, spatial segregation, bodily training, temporal routines, and constant surveillance. The divisions of space in buildings and the divisions of time in daily routines reinforce the bodily capacities being sought by the disciplinary regime. These techniques are mirrored in military organizations, schools, factories, hospitals, and similar institutions.

The 'diagram' of Bentham's Panopticon is a central motif in crystallizing the differences between sovereign and disciplinary power: the Panopticon, with its unseen watcher in the central tower, exemplifies the aspirations of disciplinary power to fabricate subjects in whom social duty and individual inclination correspond without fur ther need for training, surveillance, or punishment. In this way, disciplinary regimes and surveillance intersect with, and affect the form of, wider transformations such as the emergence of the social or human sciences, the rise of industrial capitalism, and the making of the modern liberal subject (see below).

Discipline is also connected to the emergence of 'biopower': while discipline (or 'anatomo power') relates to the training of productive individual bodies, biopower focuses of the management of populations – births and deaths, health and reproduction, and environments – and statistical and actuarial ways of comprehending popu lations as objects of knowledge and intervention. At the level of processes that affect populations, biopower is concerned with problems of disease, illness, and death; with old age, infirmity, and incapacity; and with the causal effects of, especially urban, environments or milieux.

Most crucially, techniques of power over life – the politics of the body and the politics of populations – intersect around the control of reproduction and the medical and moral supervision of sexuality. Sexuality is a fundamental problem for the production of appropriate moral, social, economically productive bodily comport ments and behaviors, and it is an equally fundamental problem for the reproduction of some populations and the restriction of others. Like discipline, sexuality is implicated in the historic constitution of the modern subject.

Foucault's notion of 'governmentality' traces another aspect of the intersection of the discipline of bodies and the management of populations to examine how methods of sovereign rule, primarily concerned with external threats to the boundaries of territory and the mainten ance of lineage, shifted into modes of government of populations, concerns with the security of the population, and exercises of power through of law and administration. Initially, governmental 'police power' of the state mir rored disciplinary powers over the body in seeking to know and regulate all aspects of social and economic interactions in order to increase the wealth and strength of the state. Subsequently, these attempts at the exercise of an all encompassing police power were overshadowed by discourses of biological, social, economic, and political 'processes' that signaled a shift in emphasis toward liberal formulations of the relations between strategies of rule, subjects, society, and economy.

Foucault also shifted his emphasis on governmentality as a mode of state power and expanded the notion to include, more generally, the ways in which individuals govern others and themselves. Governmentality is ‘the conduct of conducts’ – a form of power that, in more or less systematic ways, attempts to act on the actions of others and the self, and which is exercised at dispersed sites – in the family, in schools, in workplaces, in the confessional, and on the psychiatrist’s couch. It encompasses multiple and dispersed practices and ‘mentalities’ that aim to produce and incite various bodily comportments and self monitoring behaviors; to regulate populations and their health, reproduction, and exposures to risks; and to manage environments in which bodies and populations flourish or decline. Times, spaces, and environments are thus central to how programs of government seek to bring about desired outcomes.

But the achievement of governmental aims is never guaranteed, since things, subjects, and events are constantly recalcitrant and the aspirations of diverse governmentalities are not necessarily consistent. In particular, liberal forms of governmentality are underpinned by tensions between twin conceptions: on the one hand, of the individual who preexists society and is essentially free willed; and on the other, of the subject as the ‘creation’ of various forms of government seeking to foster appropriate exercises of autonomy. In these ‘technologies of government’ lies a link between Foucault’s conceptions of discipline and governmentality, and his conceptions of the historically constituted subject.

The Subject and the Self

In Foucault’s concerns with the nonessential nature of the subject, subjects are seen as having capacity for action, while nevertheless not only being fabricated externally with various attributes of identity, but also being able to modify that identity (or identities) through working toward the attainment of self knowledge. These ideas are more than usually scattered, but are most clearly present in Foucault’s later works on ancient sexual practices as forms of self fashioning and an ethics of the self.

The constitution of the modern subject is based in humanism, in political understandings of the autonomous individual, and in the privileging of sex as a site of identity. In processes of ‘subjectivation’, subjects come to understand themselves and their social identities through governmental (in the broadest sense) incitements to monitor their own actions in acceptable ways. However, these social identities are not fixed, both because they are formed at situated confluences of disparate processes (the sexual subject and the democratic voter), and because there exists the capacity for the self to work on the self, out of which resistances or counter conducts to attempts to govern through specific subjectivities (woman, homo sexual, black) can be born.

Here again, space and time are elements in the constitution of subjectivities and the performance of identities. In their interconnections in projects of government of others and of the self, local spaces and specific times are actively implicated in evoking or constraining repertoires of conduct.

Space and *Dispositifs*

Foucault’s claim (2001b; 361) that “space is fundamental in any form of communal life: space is fundamental in any exercise of power” most obviously not only informs his studies of institutions of confinement and reform, but is also a continuous thread in his works.

As already noted, the studies of disciplinary, reformative, and healing institutions provide detailed accounts of how space and time operate in these regimes of control to produce normalized bodily comportments and behaviors. Less restrictive arrangements, such as the Mettray colony, give glimpses of how space is implicated in projects of government of others and of the self in more diffuse settings – not just in factories, schools, or offices, but also in bedrooms, houses, streets, towns, or territories.

The term *dispositif* can be used to indicate how truth, knowledge, governmental technologies, individual conducts, bodily attributes, built forms, and spaces are enmeshed elements in governmental identification of problems and formation of solutions, to be examined in genealogies of, for instance, sexuality or madness. The study of how such *dispositifs* are assembled and deployed, therefore, involves the study of ‘thought’ about space as well as the intended and unintended effects of such arrangements and interrelations. Explorations of the importance of space to situated relations of power have been central contributions of geography to the development of Foucauldian perspectives.

Foucauldian Geographies

As discussed above, the uptake of Foucault by Anglophone geographers was influenced by the appearance in English of *Discipline and Punish* and by Soja’s *Postmodern Geographies*. These two strands – broadly, one historical and the other conceptual – shaped debates in geography for some time and produced much groundbreaking research in historical geography into crucial spatial elements of disciplinary and reformative institutions. However, despite Driver’s (1985: 426) early cautions that “any attempt to reduce *Discipline and Punish* to a set of simple ground rules will fail to recognise Foucault’s abhorrence of ‘totalizing’ forms of explanation,” many subsequent studies tended to see ‘discipline’ as fundamental feature of modern society, that had 'escaped' the prison walls. The puzzle of that 'escape' was in part a consequence of a concentrated focus on the physical spatial configurations of the prison and the Panopticon (see below).

In a parallel fashion, despite the originality of Soja's arguments about the radical, theoretical, and critical potentials of 'postmodern geographies', his claim that Foucault was a postmodern geographer and critical theorist '*malgré lui*' (despite himself) gave rise to debates about the adequacy of Foucault's conceptions of space and in relation to social and political theory and action. In particular, Soja's (1996: 159) 'heterotopology', and his claim that "… space was as central to Foucault as it was to Lefebvre, the former inflecting primarily through the nexus of power what the latter persistently parsed through the meanings of social production," aligned Foucault with the search for grand theories of space, power, and society that subsequently affected the way Foucauldian perspectives have been geographically framed.

As Philo (1992: 139) notes, "One of Soja's weaknesses is his failure to look much more closely at [the] historical treatments of substantive matters." Or as Gregory (1994) puts it: Soja "severs Foucauldian analytics of space from the genealogy of the subject (p. 297)" and fails to acknowledge that the relations "between specific rationalities, local knowledges, and forms of subjectivity form a conceptual triangle that has to be central to any serious appropriation of [Foucault's] work (p. 298)."

The concept of 'heterotopia' caught the attention of geographical theorizing, because of its radical subversion of the primacy accorded to time over space in much Western philosophy, and because of its tantalizing hints of the potentials for positive difference secreted in these 'other' spaces. But Foucauldian approaches can be taken further than this, since Foucault "calls forth a world in which one sees only spaces of dispersion: spaces where things proliferate in a jumbled up manner on the same 'level' as one another" (Philo, 1992: 139), and all of which are potentially both repressive and liberating. "… I do not think that it is possible to say that one thing is of the order of 'liberation' and another is of the order of 'oppression'. … Liberty is a *practice*" (Foucault, 2001b: 354).

These conceptual difficulties have continued to attend the reception and use of Foucault's work in some fields of geography. Latterly, however, and with ongoing publication in English, understandings of how the philosophical and historical aspects of Foucault's work can be joined with attention to the spatial have emerged. Elden demonstrates this in his important study of the influence of Heidegger on Foucault's attention to space.

> Given that Foucault uses power relations to investigate society, … and that power relations are in and through space, we can see that *questions of space are inherently political* … [and] … *politics is inherently spatial.* … Understanding how space is fundamental to the use of power and to historical research into the exercise of power allows us to recast Foucault's work not just as a history of the present but as a mapping of the present. (Elden, 2001: 150, 152) (original emphases)

Within and beyond these points of emergence and continuing debates, Foucault's works have stimulated extensive applications and developments in almost all fields of human geography. In terms of substantive research, for instance, the emphasis on the deployment of space and time in categorization, training, and normalization of bodies in disciplinary institutions and through governmental regulations is the source of significant studies in historical geography of asylums, hospitals, workhouses, and prisons; of the planning and administration of the sanitary, health, circulatory, commercial, and moral problems of urban areas; and of territorial demarcation, state formation, and the management of populations, including technologies of statistics, demography, and cartography. The influence of *Discipline and Punish* extends into social geographies of policing and surveillance, particularly in studies of current policies and technologies of social control of public space, of terrorism, and of 'socially excluded' or other marginalized groups.

'Governmentality' inspires much work in political geography, both in studies of the internal organization and political management of the spaces of the nation-state, and in the geopolitics of international relations and globalization. In economic geography, work on neoliberal spaces and subjectivities draws on ideas of post-disciplinary societies and the exercise of 'advanced liberal governmentalities' that "link individuals into a society through acts of socially sanctioned consumption and responsible choice, through the shaping of a lifestyle according to the grammars of living that are widely disseminated …" (Rose, 1999: 166).

Foucauldian approaches are also taken up in postcolonial studies of the distributions of discourses of colonial 'centers', practices of colonizers, and counter-practices of indigenous peoples; in studies of the government of postcolonial states; and how practices of exploration, mapping, and city planning are all implicated in colonial and postcolonial power relations.

Questions of the interrelations of spaces, subjects, and identities are particularly pertinent to queer geographies and geographies of sexuality, gender, and feminism. Whether spaces can be designated either as wholly 'queer' or wholly 'straight', for instance, continues to be debated. For instance, Howell argues that a Foucauldian approach suggests that

> It would be better to say that all space carries traces of heteronormative spatiality, and/or that all space is actually *queer*, in the sense that the norms of heterosexual society are unstable and incoherent. This means, for one thing, that the identification of spaces as either *gay* or *straight* is fundamentally mistaken. (Howell, 2007: 310) (original emphases)

Such studies productively highlight ways in which dominant discourses are implicated in the materialities of spaces, even as such spaces offer possibilities for ‘counter conducts’.

Current developments thus see all of Foucault’s studies as thoroughly spatialized – not just histories of particular kinds of spaces, but spatial histories and reinvigorated historical geographies. Conceptualizations of space as historically contingent pay attention to the significance of diverse, localized spaces and struggles over their configurations in exercises of power, governmentality, and formations of the self; and examine regimes of spatial truths underpinning such configurations. In Foucauldian framings, questions of space contribute to unsettling theorizations of its ontology, but equally, and possibly more importantly, they are analytical and political lenses through which to critically examine histories and geographies of the present.

Critical Debates

This subdisciplinary diversity in Foucauldian approaches is evidence of lively and productive deployments and developments of these ideas – ongoing explorations and extensions, rather than static applications of a set of routinized theories and methods. Foucauldian conceptions taken up in various fields of human geography have been based in quite different readings and different emphases on different aspects of the work, so that it can seem difficult to discern commonalities. So, for instance, queer and sexuality studies focus on subjectivities, identities, and performativity, while surveillance studies have tended to stress the overwhelming presence of the state in disciplinary projects for the control of conducts. While such diversity is to be celebrated, there might also be a case for greater interaction between subdisciplines (or indeed, between human geography and other disciplines), especially in exploring the nature and potentials of their different interpretations.

Interfield dialog might also be a productive source of debate and counterargument in relation to critiques of Foucault’s works. Critical commentaries tend to read them as (1) flirting with the ‘dangers’ of relativism, particularly in attempting to bypass theories of the subject; (2) focusing on the ubiquity of discipline and repressive relations of power; (3) writing histories from the standpoint of the privileged white European male; (4) making only either metaphorical or restrictively descriptive historical references to space; and (5) failing to provide ethical political stances or analytical tools for struggles of resistance or progressive politics of place, class, race, gender, and sexuality. Fostering wider dialog about these issues is particularly important, since developments of Foucauldian geographies are dynamic explorations, especially in the light of the fact that many of Foucault’s works still await publication and/or translation.

Nevertheless, it is the case that many critiques are founded in partial readings of the material that is available. The dispersed and extensive nature of these works makes them particularly prone to selective uptake, such that *Discipline and Punish*, or the essay on ‘heterotopias’, or the lecture on governmentality come to stand for the totality of Foucault’s thinking. Equally, many troublesome aspects of these notions have been ‘tamed’ by their attachment to more generally circulating (dominant) theories, methods, and discourses: generalized ‘concepts’ of ‘governmentality’ or ‘Panopticism’, for instance, are abstracted from space and time, and applied to current events.

For example, processes of subjectification are sometimes seen as forms of social constructivism, but this move underplays the ‘thinness’ of Foucault’s conception of the subject and his focus on regimes of truth, discursive formations, and *dispositifs*, which are radically different from conceptions of a constructing subject. Similarly, the search for theories of resistance and struggle become trapped in notions of resistance to repressive power, and by the romance of struggles of marginalized ‘victims’ against such power. In such analyses, instead of being provocations to rethink power and resistance, Foucauldian conceptions of ‘counter conducts’ and strategies of self formation become diluted and inadequate versions of other theorizations.

Most notably, both Panopticism and governmentality have been adopted in ways that subvert much of what is different and difficult about Foucault’s treatment of these concepts. The Panopticon was ‘invented’ by Jeremy Bentham (1748–1832) as the ideal and economical form of a building that would reform criminals by the simple method of never allowing them to know whether or not they were being watched, so that inmates would come to behave in corrected ways without the need for costly and inefficient gaolers. Foucault’s use of the qualities of the Panopticon is thus not to emphasize the punitive, repressive architecture of a real prison, and to see it as such, as many commentators do, is to miss the purpose of its description in *Discipline and Punish*. It is, rather, a ‘diagram’ of a productive form of power.

> He who is subjected to a field of visibility, and who knows it, assumes responsibility for the constraints of power; he makes them play spontaneously upon himself; he

> inscribes in himself the power relation in which he simultaneously plays both roles; he becomes the principle of his own subjection. (Foucault, 1979: 203)

The Panopticon is the "figure of a political technology that may and must be detached from any specific use" (p. 205). In it can be seen the shift toward a discourse of 'disciplinary' (not 'disciplined') society of subjects in whom, as Bentham hoped, duty and interest would coincide – a modern liberal society in which subjects are ruled and rule themselves as if they are free.

Such a reading of the significance of the Panopticon resonates with notions of governmentality and the ways in which subjects are brought to invigilate their own comportments and behaviors. But governmentality, too, is often attached to other, more readily available, conceptual positions. It is, for instance, conflated with 'governance' in which the state divests itself of many of its welfare and collective consumption functions, and operates through 'partnerships' with 'stakeholders' in the private, voluntary, and community sectors. In these formulations, governmentality is continuous with the state, and becomes another way of looking at state–civil society relations, but with the added advantage of highlighting ways in which the state produces – fairly unproblematically – new subjectivities of a neoliberal kind. However, as outlined above, governmentality is not principally aligned with the state, even if the state is now becoming increasingly 'governmentalized'. It is a form of power seeking 'the conduct of conducts', dispersed across multiple sites and programs, with no necessarily unified set of objectives and no certainty of intended outcomes.

Debated and often contradictory as these developments are, it is now hard to imagine a geography that does not in some way engage with the issues raised by Foucauldian approaches. Discussions over the significance of Foucault's works and their deployments in multiple geographic fields will continue and mutate as further lectures and essays are published and translated into other languages. Happily, however, differences are unlikely to be resolved, since Foucauldian projects do not seek such closure.

Conclusion

In his historical studies of the interrelations and struggles of truths, powers, knowledges, and spaces, of practices, conducts, and counter conducts, and in unsettling and politicizing notions of the subject, Foucault opens up a space that geographers enter with enthusiasm. Geographical 'Foucauldianism' is producing much valuable work, both theoretically and in projects of substantive research, in foregrounding the centrality of spatial concerns that are often neglected in other disciplines. On the other hand, what are missing from many of the docile 'Foucaults' to be found in certain developments of his work, are genealogies of particular sites and localized practices of regulation, surveillance, of subjectification, and of government.

But 'Foucauldianism' is, if nothing else, work in progress. As more lectures are published and translated, further directions for research and debates are opened up. Studies of postdisciplinary biopower and new medical technologies; explorations of geopolitical governmentalities and 'advanced liberal subjectivities'; race and nation; postcolonial powers and spaces, all draw on, and contribute to, the emergence of new Foucauldian geographies.

In all of these, Foucauldian sensibility requires the careful excavation and construction of critical and effective spatial histories that provide tools for thinking and acting differently in the present.

> The things which seem most evident to us are always formed in the confluence of encounters and chances, during the course of a precarious and fragile history. … They reside on a base of human practice and human history; and … since these things have been made, they can be unmade, as long as we know how it was that they were made. (Foucault, 1988: 37)

See also: Biopolitics; Governance; Governmentality; Structuralism/Structuralist Geography; Surveillance.

Further Reading

Burchell, G., Gordon, C. and Miller, P. (eds.) (1991). *The Foucault Effect: Studies in Governmentality*. London: Harvester/Wheatsheaf.
Crampton, J. and Elden, S. (2007). *Space, Knowledge and Power: Foucault and Geography*. Aldershot: Ashgate.
Driver, F. (1985). Power, space, and the body: A critical assessment of Foucault's *Discipline and Punish*. *Environment and Planning D: Society and Space* 3, 425 446.
Driver, F. (1993). *Power and Pauperism: The Workhouse System, 1834 1884*. Cambridge: Cambridge University Press.
Elden, S. (2001). *Mapping the Present: Heidegger, Foucault and the Project of a Spatial History*. London: Continuum.
Foucault, M. (1967). *Madness and Civilization*. London: Routledge.
Foucault, M. (1972). *The Order of Things*. London: Tavistock.
Foucault, M. (1972). *The Archaeology of Knowledge*. London: Tavistock.
Foucault, M. (1976). *The Birth of the Clinic*. London: Tavistock.
Foucault, M. (1979). *Discipline and Punish*. New York: Vintage Books.
Foucault, M. (1979). *The History of Sexuality, Vol. 1: An Introduction*. Harmondsworth: Penguin.
Foucault, M. (1980). *Power/Knowledge: Selected Interviews and Other Writings 1972 1977.* (Rabinow, P., ed.). Brighton: Harvester.
Foucault, (1984). The order of discourse. In Shapiro, M. (ed.) *Language and Politics*, pp 108 138. Oxford: Basil Blackwell.
Foucault, M. (1986). Of other spaces. *Diacritics, Spring* 22 27.
Foucault, M. (1988). Critical theory/intellectual history. In Kritzman, L. (ed.) *Michel Foucault: Politics, Philosophy, Culture*, pp 17 46. New York: Routledge.

Foucault, M. (1991). Questions of method. In Burchell, G., Gordon, C. & Miller, P. (eds.) *The Foucault Effect: Studies in Governmentality*, pp 73 85. London: Harvester/Wheatsheaf.
Foucault, M. (1991). Governmentality. In Burchell, G., Gordon, C. & Miller, P. (eds.) *The Foucault Effect: Studies in Governmentality*, pp 87 104. London: Harveser/Wheatsheaf.
Foucault, M. (1996). The order of things. In Lotringer, S. (ed.) *Foucault Live*, pp 13 18. New York: Semiotext(e).
Foucault, M. (2001). Truth and power. In Faubion, J. (ed.) *The Essential Works of Foucault 1954 1984, Vol. 3: Power*, pp 111 133. London: Penguin.
Foucault, M. (2001). Space, knowledge and power. In Faubion, J. (ed.) *The Essential Works of Foucault 1954 1984, Vol. 3: Power*, pp 349 364. London: Penguin.
Foucault, M. (2003). *Society Must Be Defended: Lectures at the College de France 1975 76*. London: Allen Lane.
Foucault, M. (2007). *Security, Territory, Population: Lectures at the College de France 1977 78*. London: Allen Lane.
Gregory, D. (1994). *Geographical Imaginations*. Oxford: Blackwell.
Hannah, M. (2000). *Governmentality and the Mastery of Territory in Nineteenth Century America*. Cambridge: Cambridge University Press.
Hannah, M. (2007). Formations of 'Foucault' in Anglo American geography. In Crampton, J. & Elden, S. (eds.) *Space, Knowledge and Power: Foucault and Geography*, pp 83 106. Aldershot: Ashgate.
Howell, P. (2007). Foucault, sexuality, geography. In Crampton, J. & Elden, S. (eds.) *Space, Knowledge and Power: Foucault and Geography*, pp 291 316. Aldershot: Ashgate.
Huxley, M. (2006). Spatial rationalities: Order, environment, evolution and government. *Social and Cultural Geography* 7(5), 771 786.
Matless, D. (1998). *Landscape and Englishness*. London: Reaktion Books.
O'Farell, C. (2005). *Michel Foucault*. London: Sage.
Ogborn, M. (1998). *Spaces of Modernity: London's Geographies 1680 1780*. New York: The Guildford Press.
Philo, C. (1992). Foucault's geographies. *Environment and Planning D: Society and Space* 10(2), 137 162.
Philo, C. (2004). *A Geographical History of Institutional Provision for the Insane from Medieval Times to the 1860s in England and Wales: The Spaces Reserved for the Insane*. Lampeter: Edwin Mellen.
Rose, N. (1999). *Powers of Freedom: Reframing Political Thought*. Cambridge: Cambridge University Press.
Soja, E. (1989). *Postmodern Geographies: The Reassertion of Space in Critical Social Theory*. London: Verso.
Soja, E. (1996). *Thirdspace: Journeys to Los Angeles and Other Real and Imagined Places*. Malden, MA: Blackwell.

Relevant Websites

http://www.cartome.org
Bentham texts and images, Cartome.
www.ucl.ac.uk
Bentham Project UCL Bentham Project, University College of London.
http://www.michel foucault.com
Extensive sources and links, michel foucault.com.
http://www.foucault.info
Extracts from texts, extensive links, Michel Foucault, Info.
http://www.foucault studies.com
Online Journal Foucault Studies.
http://www.surveillance and society.org
Online Journal Surveillance and Society.
http://images.google.com
Panopticon, Google Images.
http://www.ucl.ac.uk
University College London's ongoing project for the publication of the complete works of Jeremy Bentham (including little known conceptual image of the Panopticon idea).

Fractal Analysis

N. S. Lam, Louisiana State University, Baton Rouge, LA, USA

Glossary

Fractal Brownian Motion (fBm) A mathematical model governed by a zero mean Gaussian process with different levels of stationary increments that can be used to simulate curves and surfaces of various fractal dimensions.

Fractal Dimension A parameter similar to the Hausdorff–Besicovitch dimension used to describe the complexity of forms and processes; the more complex the form, the higher the fractal dimension.

Hausdorff–Besicovitch Dimension Noninteger dimension values first extended by mathematicians Hausdorff and Besicovitch to describe forms and processes.

Self-Similarity The property where each portion of a form or pattern looks like a reduced-scale of its whole.

Steinhaus Paradox The dilemma of measuring complex forms that occurs when the result of the measurement depends on the measurement scale used; for example, the smaller the step size used to measure the length of a curve, the longer the length of the curve.

Introduction

It has been three decades since Mandelbrot coined the term fractals in 1977. Since then, fractal analysis dom inated the scientific literature for more than two decades. In the 1980s and 1990s, studies on fractals can be found literally in every major discipline, ranging from physics, chemistry, engineering, geosciences, and geography to music, arts, movies, and video game industry. Its impacts have been far and deep, and its significance is comparable with, if not exceeding, the development of the concepts of entropy, chaos, and Gaussian distribution. It was once claimed that no one would be considered scientifically literate if the meaning and scope of fractals was not understood.

Fractal analysis has a very close tie with geography and the broader environmental science. In fact, fractal geom etry was derived by Mandelbrot mainly because of the inefficiency of classical geometry in describing complex shapes and phenomena that we observe everyday, such as clouds, terrains, settlement patterns, coastlines, and river networks. In his 1967 influential article that subsequently led to the derivation of fractals, Mandelbrot posted a question that geographers and cartographers have been asking: how long is the coastline of Britain? The answer to this question depends mainly on the step size of the walking divider used to measure the coastline from the map; the smaller the step size, the longer the coastline. This 'Steinhaus paradox' is related to a number of fun damental issues that have long plagued geographers, ecologists, and geoscientists, such as the issues of scale, scale dependency, map generalization, and landscape measurement, representation, and characterization.

The fractal theory is intimately related to chaos, complex theory, and nonlinear dynamical systems. Lit erature on these is abundant and can be overwhelming, with new research continuing to appear. While there have been many successful applications and valuable contributions, there are also major misunderstanding, problems, and misuses of fractals. The following focuses on fractal analysis as applied in geography; it explains the key concepts, major applications, issues and challenges, and future directions. We hope this brief overview will facilitate a better understanding and appreciation of fractal analysis in geography.

Key Concepts

Fractal Dimension and Self-Similarity

The main parameter in fractals is the fractal dimension *D*, which is used to represent the complexity of objects, forms, and processes. In classical geometry (i.e., Euclid ean geometry), a point has a dimension value of zero, a line is one dimensional, an area has two dimensions, and a volume three dimensions. This is called topological dimension, and is defined by integer values. In fractal geometry, the fractal dimension of a point pattern can be any noninteger value between zero and one, a curve between one and two, and a surface between two and three, depending on the complexity of the pattern. The more geometrically complex the pattern, the higher the fractal dimension. This concept of fractional dimension was first formulated by mathematicians Hausdorff and Besicovitch. Mandelbrot later coined it as fractal di mension and defined fractals as "a set for which the Hausdorff–Besicovitch dimension (*D*) strictly exceeds the topological dimension (D_T)."

In defining the fractal dimension, Mandelbrot em ployed a concept called self similarity. Many spatial forms such as curves and surfaces are self similar, meaning that each portion can be considered as a re duced scale of the whole. The degree of self similarity is used to define the theoretical fractal dimension as follows:

$$D = \frac{\log N}{\log(1/r)} \qquad [1]$$

where r is a similarity ratio (i.e., scale reduction factor) and N is the number of steps needed to traverse a curve. The relationships between fractal dimension (D), number of steps (N), and similarity ratio (r) are illustrated in **Figure 1**. **Figure 2** shows an example of Peano curve that is self similar. This type of curve is also space filling, meaning that by increasing the density of grids, the same curve can be generated that will eventually fill in the space, going from one dimensional to two dimensional.

Since most natural forms, curves, and surfaces are not strictly self similar but rather statistically self similar, the fractal dimension will need to be estimated statistically through the regression technique. In general, the empirical D value of a curve (e.g., coastline) is estimated by regressing between the length of the curve and the step size used to measure the length of the curve, using the following equations:

$$\log L = K + B \log \delta \qquad [2]$$

$$D = 1 \quad B \qquad [3]$$

where L is the length of the curve, δ is the step size, B is the slope of the regression, and K is a constant. D is a function of the regression slope B. The more irregular the curve, the greater the increase in length as step size decreases, hence the steeper the negative slope, leading to higher fractal dimension.

The D value of a surface can be estimated in a similar fashion, but using the following equation instead of eqn [3]:

$$D = 2 \quad B \qquad [4]$$

There are quite a few algorithms in the literature for estimating the fractal dimension of curves and surfaces,

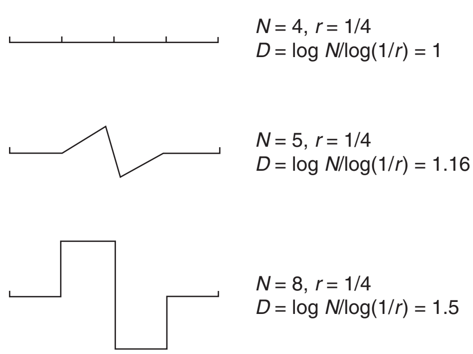

Figure 1 Relationships between similarity ratio (r), number of steps (N), and fractal dimension (D).

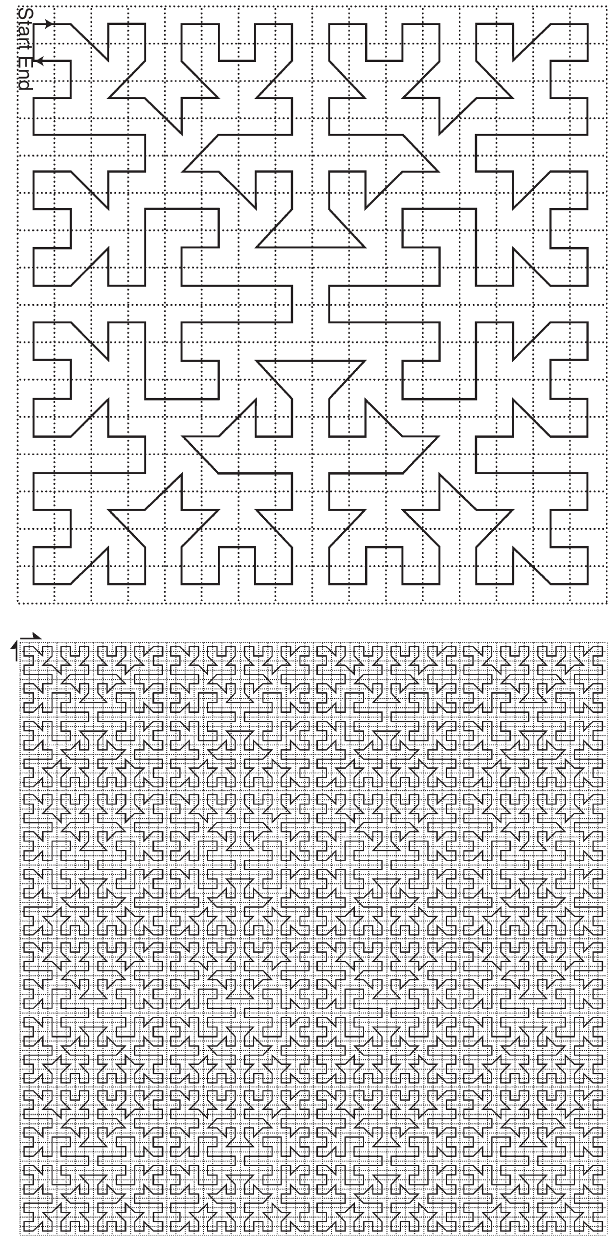

Figure 2 Space-filling, self-similar Peano curves generated using the ArcInfo GIS software on a 16 × 16 and a 256 × 256 grid.

and the results from applying these algorithms can vary substantially. This discrepancy is a result of different realizations of the fractal theory into practice by different researchers. As discussed below, this has been a major problem in utilizing fractals effectively for measurement and characterization of real curves and surfaces.

Fractional Brownian Motion

The other side of the coin in fractal analysis is how to simulate the processes and patterns that the fractal model is meant to describe. Measurement using the fractal dimension is the first step toward scientific understanding of complex forms and processes. Simulation often follows, which can further our understanding if it successfully mimics reality. In searching for a stochastic model that can mimic fractal patterns closely, Mandelbrot employed the Brownian motion model in physics. Brownian motion refers to the irregular movements of small particles, the time series of which often shows a highly autocorrelated sequence, indicating long time dependence of previous movements. Because of the similarity between fractals and the Brownian process, the latter can be used as a model to simulate fractal curves and surfaces. Brownian trails are in fact fractal sets, with a fractal dimension of 1.5 for a Brownian curve and a fractal dimension of 2.5 for a Brownian surface. Mandelbrot later modified the ordinary Brownian functions using the concept of self similarity, and the combination of these two leads to a generalization called fractional Brownian functions. From these functions, curves and surfaces of other dimensions can be generated; hence, they are called fractional Brownian motion (fBm) curves and surfaces. Examples of fBm curves and surfaces are shown in **Figures 5** and **6**.

Algorithms

There are quite a few algorithms for measuring and simulating the fractal dimensions of curves and surfaces for geographical analysis. The following describes an algorithm for each type of operation. The algorithms selected for discussion below are based on their simplicity, applicability, and performance. The computer programs of all the algorithms discussed can be found in Lam and De Cola's edited volume *Fractals in Geography*, first published in 1993.

Fractal Curve Measurement

For curve measurement, the walking divider method is the most straightforward and intuitively simple, as it corresponds directly with the original definition of fractals. The method is sensitive to the irregularity of the curve and hence it is considered the most accurate. The method generally takes more computational time than other 'approximate' methods such as the box counting and equipaced polygon methods, but computation time is no longer considered a major problem now given that fast computers are generally available and quite affordable.

Given a curve (e.g., coastline), the walking divider method will begin walking the dividers along the curve using an initial divider length (step size). The number of steps it takes to traverse the curve is counted. The initial step size is usually one half of the average distance between every pair of adjacent points of the digitized curve. Intermediate points are interpolated using ordinary trigonometric functions and distance calculations. The next

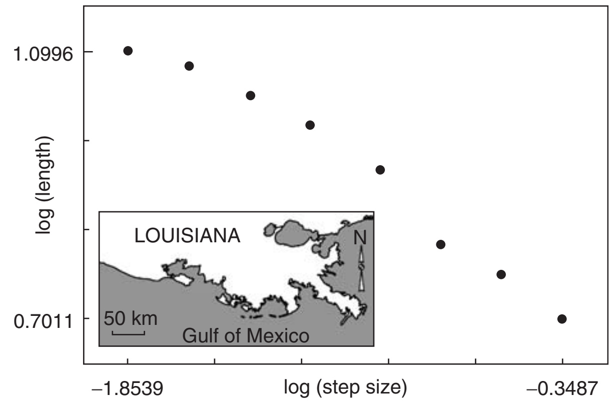

Figure 3 Computation of the fractal dimension of the Louisiana coastline, digitized with 485 points from a map of scale 1:1 350 000. The measured D is 1.2 with R^2 0.98.

step size doubles the preceding one, such as 2, 4, 8, 16, and so on, and the number of steps needed to walk through the curve is recorded for that step size. These two variables are plotted on a double log form, and a linear regression is applied to determine the slope of the regression line and the goodness of fit R^2 value. For a reliable regression estimate, at least five step sizes should be used. **Figure 3** shows the computation of the Louisiana coastline and its corresponding log–log plot. The computed D value for the coastline is 1.2 with R^2 of 0.98, compared with a computed D value of 1.0 for a circle with a low R^2 of 0.49, and 1.2 for Koch snowflake with an R^2 also of 0.98. The theoretical D values for circle is 1.0 and for Koch snowflake is 1.26. This simple example shows that there are some discrepancies between the theoretical and computed D values. Ambiguity in this method arises from decisions on two parameters: how many steps one should use, and how to determine the break point where the log–log plot is no longer linear (i.e., scale range).

Fractal Surface Measurement

Measuring the fractal dimension of a surface presents a greater challenge than measuring curves, because adding a dimension makes the task of measurement much more complex and ambiguous. Similar to measuring the length of a curve, the area of a surface is measured to indicate surface complexity. The isarithm method for measuring the surface dimension is a direct extension of the walking divider method, except that the curves measured are now isarithms (e.g., contours) derived from the surface. In extending the method from curve to surface estimation, two additional considerations are needed: how many isarithms should be used to represent the surface, and how should these isarithms be averaged to derive a composite D value. Although the isarithm method is intuitively simple and quite accurate, because of these ambiguities, another method, the modified triangular prism method, is presented. The modified triangular prism method was found to be the most accurate and reliable method for estimating the fractal dimension of remote sensing image surfaces, and there is no additional parameter needed to be specified other than choosing the number of steps to be used in deriving the fractal dimension.

Given a surface defined by a mesh of grids with x, y, and z values, the modified triangular prism method cal culates the total surface area defined by triangular prisms at each step size. For a given step size (defined by the length of a square grid, step size is 1 for 1×1 grid, 2 for 2×2 grids, and so on), the triangular prisms are con structed by connecting the heights or z values of four adjacent pixels to the center height, which is the average z value of the four adjacent pixels (**Figure 4**). The sur face area of the top of the prism is then calculated, and the procedure is repeated across the image to derive the total surface area. This procedure is repeated for each step size. The logarithm of total surface area is then re gressed against the logarithm of step size, and the fractal dimension calculated as 2 minus the slope of the re gression line as in eqn [4].

Fractal Curve Simulation

Fractal curve (and surface) simulation is often used for both visualization and analytical purposes (which will be elaborated in the next section). Simulation of regular objects such as the Koch curve can be achieved by a set of equations based on the principle of strict self similarity. Simulation of realistic curves such as coastlines will need to incorporate a certain degree of randomness, since strict self similarity is rare in nature. A commonly used algorithm is the midpoint subdivision method introduced by Carpenter in 1981. Given a curve defined by a few points, the method adds details by recursively sub dividing the curve by adding midpoints in the line seg ments. The x, y coordinates of the midpoint are the average x, y values of the endpoints of the segment, with the exception that the y coordinate is displaced by a random component. The random component is con trolled by the following parameters: a roughness factor with values ranging from 0 and 1, a scale factor specific to the scale and resolution of the sample data, and a random value between -1 and $+1$. Statistical self similarity is maintained by ensuring that the deviation between the interpolated and the actual midpoint is proportional to the distance between the end points. **Figure 5** compares the results of the generated curves using two different roughness factor values. Unlike the surface simulation method discussed below, this technique does not provide a mechanism whereby a D value can be specified at the beginning of the simulation process.

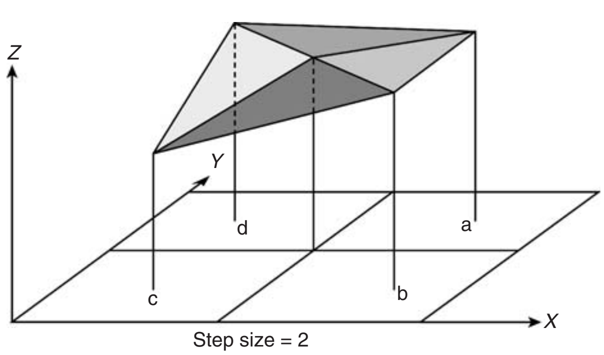

Figure 4 An illustration of the modified triangular prism method for calculating the surface area at each step size.

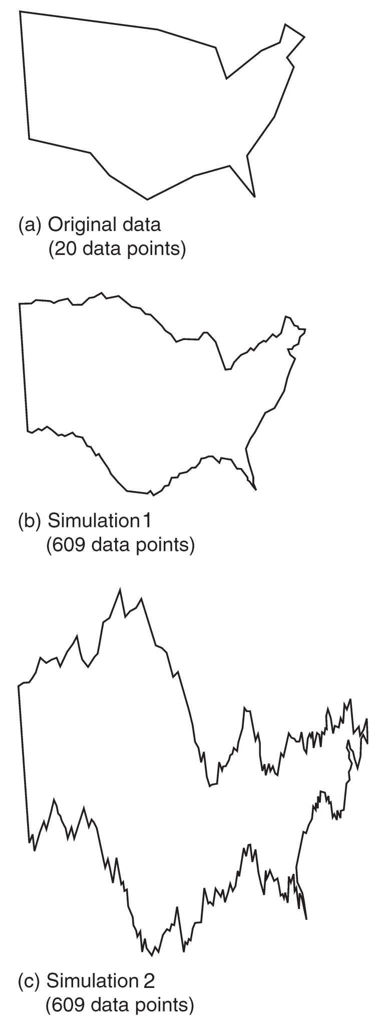

Figure 5 Two fractal curves (609 points) generated from (a) a simplified US boundary defined by 20 points; (b) curve with roughness factor 0.2, estimated *D* 1.07; and (c) curve with roughness factor 0.8, estimated *D* 1.48. Reprinted from Lam, N. S.-N. and De Cola, L. (eds.) (1993). *Fractals in Geography*. Englewood Cliffs, NJ: Prentice-Hall.

Fractal Surface Simulation

A widely used method for surface simulation is the shear displacement method. The method starts with a matrix of square grids with zero z values. A succession of random cuts across the surface is then generated, and the surface is displaced vertically along each random line to form a cliff. This process is repeated until several cliffs are created between adjacent sample points. The set of random lines generated has the property that their points of intersection form a Poisson point process, while the angles of intersection are distributed uniformly between 0 and 2π. The heights of the cliffs are controlled by the parameter H, which has values between 0 and 1. H determines how the variance between two points relates to their separation distance, and is defined as

$$E[z_i - z_{(d+i)}]^2 = |d|^{2H} \qquad [5]$$

where $E[z_i - z_{(d+i)}]^2$ is the expected variance between two points having a distance d. The fractal dimensions of the simulated surfaces are equal to $D = 3 - H$. **Figure 6** illustrates three simulated surfaces ($D = 2.1$, 2.5, and 2.9) generated from this algorithm.

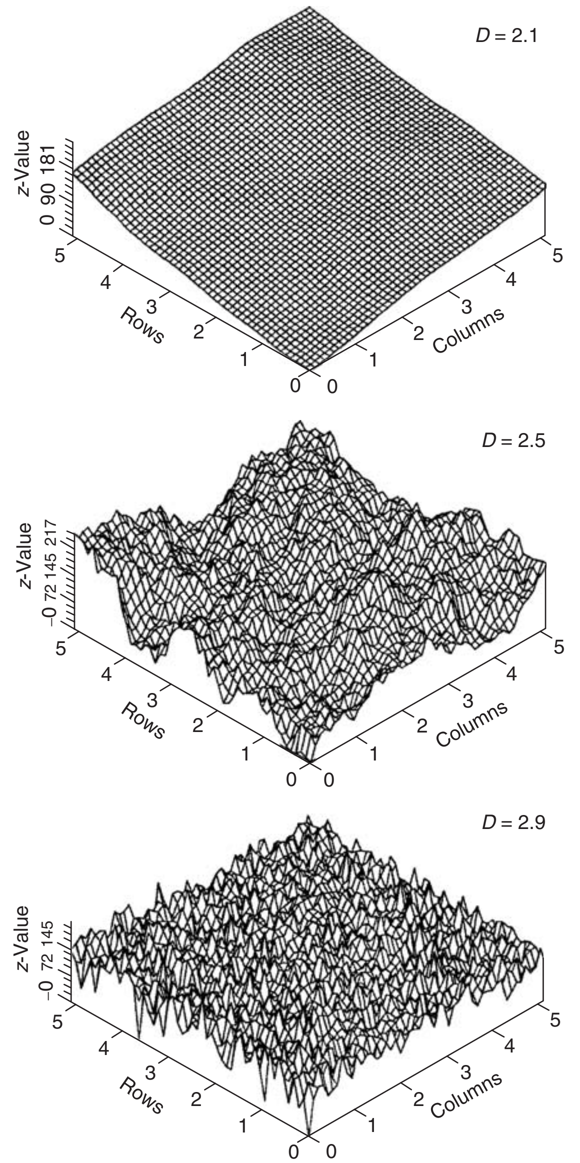

Figure 6 Three simulated surfaces (D 2.1, 2.5, and 2.9) generated from the shear displacement algorithm using a 50×50 grid.

Applications

Applications of fractals that are broadly related to geography generally fall into two groups. The first uses the fractal model to simulate real world objects and patterns for both analytical and display purposes. Typical objects or patterns simulated include coastlines, ecotones, river networks, time series, terrains, clouds, natural landscapes, settlement patterns, urban hierarchy, diffusion patterns, and remote sensing images. Fractal simulation is frequently used in computer graphics and movie industry because they can simulate life like landscapes and objects for video games, movies, and a variety of virtual reality applications. Spectacular images generated by the fractal technique and enhanced by computer graphics rendering have fascinated many people and made the fractal model a very popular tool. Famous motion pictures such as Star Wars and its sequels have used the fractal model to create artificial landscapes and objects. The Mandelbrot set, derived in 1980 and now bears his name, was reportedly the most complex mathematical set discovered in the literature and has since become an icon of fractals. In addition to their sheer beauty and visual impacts, the simulated objects and patterns help unveil some of the properties of complex dynamical systems that have seldom been visualized before.

Analytical uses of fractal simulation have been valuable. At least three types of analytic applications have been made. First, simulated fBm curves and surfaces are regarded as ideal theoretical test data sets for testing a number of geographic models and algorithms. In cartography and geographic information systems (GIS), for example, fractal surfaces have been used as test data sets to examine the performance of various spatial interpolation methods and the efficiency of different data structures. Second, the fractal simulation technique has been employed as an interpolation method itself to generate coastlines and curves based on a few sample points (**Figure 5**). This curve generation process can be seen as an inverse of curve generalization, and could be implemented in a GIS for automated cartographic displays with varying levels of cartographic generalization.

In human geography, fractal simulation technique has been applied to simulate central place hierarchy, urban morphology, and urban growth. In physical and environmental geography, elaborated algorithms for deriving river networks from simulated fractal surfaces were developed to help understand the various properties of geomorphic processes. The third type of analytical uses of fractal simulation involves hypothesis testing. Fractal surfaces could serve as a null hypothesis terrain or norm whereby further simulation of various geomorphic processes can be made to investigate the effects of different landform processes.

The second group of applications is related to fractal measurement. Fractal dimension has been frequently used as an index for describing the complexity of curves and surfaces, which can then be used to interpret and infer the underlying spatial processes. For example, fractal dimensions computed for different parts of a coastline can be utilized to infer the underlying coastal processes that shape the coastline. A smooth coastline may indicate the dominance of large scale coastal process such as ocean currents, whereas a contorted coastline is likely a result of local deltaic deposition and land subsidence. The power functions, as typified in fractal analysis, have been employed to describe many phenomena that exhibit similar power relationships, such as the power function between stream length and basin area, the time series of river bedload transport, and the city rank size rule. In addition, in a pioneer article in 1980 that established the first application of fractals in geography, Goodchild demonstrated that fractal dimension can be used to predict the effects of cartographic generalization and spatial sampling, a result which may assist in determining the optimum resolution of pixels and polygons used in remote sensing and GIS studies. The fractal model has been applied to distinguish different types of land cover from remote sensing images, and the results indicate that urban land cover generally has higher dimension than agricultural or forest land cover. Other studies found that fractal dimension could be used to indicate the degree of human disturbance in natural landscapes; the higher the fractal dimension, the least disturbed the landscape by humans.

Major Issues

The fractal model is no doubt a fascinating tool for simulation. However, its use as a technique for measuring and characterizing spatial phenomena has raised criticisms. First, reducing the description and analysis of complex phenomena into a single fractal dimension value will always run the danger of oversimplification and overgeneralization. On the other hand, one could argue that fractal dimension as a descriptor is no different from any other single statistical descriptors such as mean and standard deviation. Using just the mean or standard deviation value alone is not sufficient for interpreting a data set or a spatial pattern. The same is true for fractals where the fractal dimension reflects the complexity of forms and patterns only. In fact, one could further argue that why not use the fractal dimension as a standard descriptor for any spatial patterns, just like the functions of mean and standard deviation for a statistical distribution.

The second problem, which is more specific to fractals, refers to the fundamental concept of self similarity. Strict self similarity assumes that the form or pattern of a spatial phenomenon remains unchanged throughout all scales, which further implies that one cannot infer the scale of the spatial phenomenon from its form or pattern. This has been considered unacceptable in principle and has been rejected by a number of researchers. Empirical studies have shown that most real world spatial phenomena are not pure fractals with a constant *D*, but instead, *D* varies across a range of scales. Rather than using *D* in the strict sense as originally defined by Mandelbrot, many researchers, including Mandelbrot himself, realize that strict self similarity is rare in natural phenomena, and that self similarity must be estimated statistically and at certain scale ranges. Instead of regarding the scale dependency of fractals as a shortcoming, some researchers suggested that information on the changes of *D* with scale could indeed be used to reveal the scale dependency characteristics of the phenomenon and help interpret the underlying processes at specific ranges.

The third problem refers to the technical aspect of fractals. First, the fact that self similarity exists only within certain ranges of scales makes it difficult to identify where the break points are during the computational process. The location of break points affects the final *D* value, which is then used to characterize the curve or surface. Second, extension from curve measurement to surface measurement is not straightforward. Different researchers employ different interpretations and sampling strategies in their surface measurement algorithms. The end results are not encouraging, as these algorithms frequently yield different fractal dimension values for the same surface. How to determine which computed value is the most accurate one? This leads to the third problem, which is how to determine the statistical significance of a computed fractal dimension value. If two surfaces are computed with *D* values of 2.21 and 2.29, would they be considered to be statistically the same, thus indicating the same degree of complexity? What are the confidence limits and their corresponding statistical significance? Monte Carlo simulation may ultimately be needed to determine statistical significance. Fourth, a major impediment to utilizing fractal measurement is that there are very few algorithms readily available to researchers for experimentation. Although this problem is somewhat alleviated recently because of increased availability of ready to use algorithms, the same frustration remains that the results from applying differing algorithms often contradict with each other. A thorough evaluation of the various fractal measurement algorithms is necessary be fore they can be used to reliably characterize curves and landscapes.

Recent Advances in Geography

Significant progress regarding fractal analysis in geog raphy has been made recently. First, computer software for fractal analysis is increasingly available, which makes testing of existing algorithms and exploration of new applications easier. The development of a software package called Image Characterization and Modeling System (ICAMS) by Lam and others is one example. ICAMS provides algorithms for computing fractal di mensions for surfaces and images, generating fractal surfaces, as well as computing a number of related spa tial/textural statistics such as wavelets, lacunarity, spatial autocorrelation, and variograms. Furthermore, bench mark studies comparing fractal measurement algorithms have also been conducted, which lead to further improvement or correction of existing algorithms. The modified triangular prism method, as described above, is an improved algorithm and is a direct outcome of being able to utilize the software for such studies.

Fractals and scale are two inseparable concepts. In response to the problem of strict self similarity and uniform scale implied by the original fractal concept, recent applications have focused on using fractals to determine the scale ranges of different spatial phenom ena. Another significant advance has been the develop ment of multifractals and local fractals, both are designed to better represent complexities. In particular, local fractal dimensions computed for small moving windows in remote sensing images were used to detect land cover changes. This new local approach is designed to over come the problems of using a single fractal dimension to represent an entire study area. By applying the calcu lation in a small window area, the variations and com plexity can be better captured. These local modules have already been implemented in ICAMS.

Recent studies have also advocated the need to use fractal dimension as part of the metadata for spatial data, especially remote sensing images. The argument is that if mean and standard deviation are basic nonspatial de scriptors for spatial data, why not use the fractal di mension as an index of spatial complexity to add the missing spatial information that conventional nonspatial statistics do not provide. Such information will help speed up the mining of large image databases and the selection of appropriate imagery for many applications.

Increasingly, fractal analysis is used as part of a group of texture indices that can be used for image classification and change detection, which has resulted in dramatic increase in image classification accuracy. Unlike many spatial indices used in landscape ecology, fractal dimen sion can be computed directly from unclassified images. This property makes fractal analysis a useful tool for changing detection without the need for prior image classification, which then paths the way for future auto mation in image classification and change detection.

Future Directions

The beauty of the fractal model lies in its simplicity, which is to use a simple index to represent complex forms and systems. Useful modifications such as multi fractals and local fractals have significantly expanded its applicability. At the same time, it is recognized that using a single index may never be sufficient to represent all aspects of a complex system. Adding more complexity to the original fractal model, however, could defeat the original purpose. This ‘fractal paradox’, as coined here, is the dilemma we are facing that may inhibit the progress. What are the best ways to utilize such a fantastic concept and make progress toward utilizing fractals to advance geographical sciences?

Here we suggest three directions for future research. First, more research should focus on developing stat istical vigor for the computed dimension values by identifying the confidence limits and statistical signifi cance. Second, fractal dimension alone is not sufficient to represent all aspects of spatial complexity, additional related indices such as wavelets, lacunarity, spatial autocorrelation, and others, are needed to help in de scribing complex system. Hence, the complementarities among these various indices should further be explored. Third, with improved and corrected algorithms be coming increasingly available, more studies are needed to explore the relationships between fractal patterns and their underlying processes. Further studies will be nee ded to determine whether fractal dimension alone, or in conjunction with other indices, can be used to suggest various conditions of curves, surfaces, and landscapes, such as the degree of human disturbance in natural landscapes, the degree of biodiversity after major dis turbance, or the social and health conditions of a region. Fractal analysis, deeply grounded in the concept of complexity, should in turn be used to unveil and explain complexities.

See also: Chaos and Complexity; Complexity Theory, Nonlinear Dynamic Spatial Systems; Scale Analytical.

Further Reading

Arlinghaus, S. (1985). Fractals take a central place. *Geografiska Annaler* 67B(3), 83 88.

Batty, M. (1985). Fractals geometry between dimensions. *New Scientist* 105(1450), 31 35.

Batty, M. and Longley, P. (1994). *Fractal Cities: A Geometry of Form and Function*. London: Academic Press.

Clarke, K. C. (1986). Computation of the fractal dimension of topographic surfaces using the triangular prism surface area method. *Computers and Geosciences* 12(5), 713 722.

Feder, J. (1988). *Fractals*. New York: Plenum Press.

Goodchild, M. F. (1980). Fractals and the accuracy of geographical measures. *Mathematical Geology* 12, 85 98.

Goodchild, M. F. and Mark, D. M. (1987). The fractal nature of geographic phenomena. *Annals of the Association of American Geographers* 77(2), 265 278.

Lam, N. S. N. (1990). Description and measurement of landsat TM images using fractals. *Photogrammetric Engineering and Remote Sensing* 56(2), 187 195.

Lam, N. S. N. and De Cola, L. (eds.) (1993). *Fractals in Geography*. Englewood Cliffs, NJ: Prentice Hall.

Lam, N. S. N., Qiu, H. L., Quattrochi, D. A. and Emerson, C. W. (2002). An evaluation of fractal methods for measuring image complexity. *Cartography and Geographic Information Science* 29, 25 35.

Lam, N. S. N. and Quattrochi, D. A. (1992). On the issues of scale, resolution, and fractal analysis in the mapping sciences. *The Professional Geographer* 44(1), 89 99.

Mandelbrot, B. B. (1967). How long is the coast of Britain? Statistical self similarity and fractional dimension. *Science* 155, 636 638.

Mandelbrot, B. B. (1983). *The Fractal Geometry of Nature*. New York: WH Freeman.

Peitgen, H. O. and Richter, P. H. (eds.) (1986). *The Beauty of Fractal Images*. New York: Springer Verlag.

Peitgen, H. O. and Saupe, D. (eds.) (1988). *The Science of Fractal Images*. New York: Springer Verlag.

Francophone Geography

G. Benko, Université Panthéon-Sorbonne, Paris, France
C. Desbiens, Université Laval, Québec City, QC, Canada

Glossary

***Espace Vécu* (‘Lived Space’)** Armand Frémont’s concept offers an alternative to the abstract logic of spatial analysis by proposing that humans do not live in space ‘as it is’ but rather in space ‘as they represent it to themselves’. As a result, a distinction emerges from the set of spaces individuals use on an everyday basis (*espace de vie*) and the social relations they form in these spaces (*espace social*).

GREMI ‘European Research Group on Innovative Environments’ (*Groupe de recherche européen sur les milieux innovateurs*). This European/North American research team was created in 1984 by Philippe Aydalot, professor at Paris 1, to study the relations between technical innovations and territories. According to these researchers, companies do not preexist in the local milieu but are generated by it.

Innovative Environments Innovative environments (*milieux innovateurs*) enable the creation of innovation networks and are the source of their dynamism by generating a context for their formation, development, and diffusion.

Proximity The ‘French School of Economic Proximity’ is based on the hypothesis that geographic proximity is not just a matter of physical distance, that is, it is not solely conditioned by natural constraints but is socially constructed, if only through the presence of infrastructures.

Semiotics Semiotics is the study of signs and symbols, both individually and grouped in systems. It includes the study of how meaning is constructed and understood. Drawing from this approach, cultural geographers have used semiotics to analyze landscapes as sign systems.

The Formation of the French School of Geography: An Overview

Francophone geography traces its origins to the end of the nineteenth century, which also coincides, in Europe, with the emergence of human geography as a separate discipline from physical geography. In the early days of the French school, German thinker Carl Ritter’s (1779–1859) attempt to underscore the unity of diverse phe nomena across the surface of the Earth exerted a strong influence. In addition, Friedrich Ratzel (1844–1904), and his concept of anthropogeography, represented a signpost for this early period. The publication, in 1859, of Charles Darwin’s *The Origin of Species* raised new questions concerning the links between human diversity and evolutionism; these issues called for original ways of re flecting on the relationship that ties humans not only to their environment, but also to each other. Early French geography eventually drew more of its insights from neo Lamarckism, which posits the existence of a feedback effect within evolution, but the political context that prevailed in France at the end of the nineteenth century must also be accounted for as a factor in the creation and expansion of a school of human geography: as Paul Claval (1932–) notes in his *Histoire de la géographie française de 1870 à nos jours* (1998), France’s defeat in the Franco Prussian War (1870–1871) which resulted in the annex ation of Alsace Lorraine by Germany gave rise to a wave of patriotism. During this movement, the production of knowledge about France’s people and territory became synonymous with duty toward the mother country (*patrie*) and the State was particularly active in developing a geography curriculum, as well as in fostering the cre ation of tools for teaching this curriculum.

In this first generation of French geographers, Émile Levasseur (1828–1911) was a notable figure. Due to his background in economics, demography, and statistics, his approach to geography relied on density maps as sources of information and tools of analysis. Studied at different scales and in a comparative perspective, patterns of human clustering in space pointed to key aspects of the nature of the human relationship with the various elements that made up the environment. To this, he added an economic analysis of the networks of pro duction and distribution that also model human density across the globe. Though the cultural aspect is largely absent from his work, he provided a useful system for mapping the placing of human settlements and the re lationship between them.

Élisée Reclus (1830–1905) was another important figure from this period. A student of Carl Ritter, his holistic approach to geographical knowledge is evident in the title of his major works such as *Nouvelle géographie universelle: la terre et les homes* (1875), or *L’Homme et la terre* (1905–08). Despite the extensive breadth of these ana lyses, his approach also encompassed an attention to micro scale phenomena such as was demonstrated in his book *Histoire d’un ruisseau* (1882), which can be regarded as a precursor to the ecological sensitivity that became widespread in the latter part of the twentieth century. In studying the Earth, his chief interest was to understand humans and the various inequalities that exist among them. As a result, power relations and political geography are important dimensions of his contribution to geography, which partly explains the renewed interest for his work at the dawn of the second millennium.

Vidal de la Blache

Despite the importance of these – and several other – contributions, Paul Vidal de la Blache (1845–1918) is regarded as the chief artisan of the consolidation of the French School of Geography at the turn of the twentieth century. Like Reclus and others around him, he found much of his early inspiration in the works of German geographers, most notably Carl Ritter and his concept of geographic unity. If all geographic phenomena are unified at their root, one important element nevertheless comes to the forefront in Vidal de la Blache: it is that of 'positioning'. Indeed, the position of one element of human geography relative to others near and far largely determines its role in a totality. With this perspective, landscapes emerged as layered spatial units where every object acts and is acted upon differently within the whole.

Spanning more than three decades, Vidal de la Blache's work is multifaceted. If fieldwork was an integral part of his approach, his contribution must also be understood in light of his development of a pedagogy specifically adapted to geography. True to Ritter, physical geography and ecology inform his analysis of human geography, but it is the geopolitical dimension that reaches most widely across his major works. At a time when European nations were in flux, the construction of national coherence and unity out of heterogeneous populations and lands fueled much of his research on the national character of France. His work *Tableau de la géographie de la France* (1903) created a portrait of France as a gathering of distinct but complementary regions that strongly anchor the country in its territory and history. By describing the processes of human adaptation specific to each region, the main goal of Vidal de la Blache was to represent the 'geographic personality' of France.

The two concepts that form the overarching structure of his work are those of 'milieu' and '*genre de vie*'. The sum of the physical characteristics that shape human adaptation to, and exploitation of, a given locale, his notion of the milieu which is the sum of the physical characteristics that shape human adaptation to, and exploitation of, a given locale offered a counterpoint to determinism. Indeed, the concept of milieu is influenced by Friedrich Ratzel's idea of possibilism, which posits that "Nature proposes and humans dispose." As a result of his extensive research in the Mediterranean region throughout the 1880s, his concept of *genre de vie* became more precise, referring essentially to how agricultural and pastoral people develop the tools and techniques necessary to draw their sustenance from their environment. These processes, Vidal de la Blache (1911) proposed, accumulate and sediment themselves to give rise to landscapes that are original and specific to the people who, through interaction with their milieu, make them:

> As [the *genres de vie*] have prevailed over broad stretches of land, they are highly evolved formations that, without having taken on the fixity of animal societies, represent a series of accumulated efforts that eventually become cemented. Man is a creature of habit more than a creature of initiative (*Annales de géographie*).

Toward the end of his life, Vidal de la Blache attempted to synthesize his vision of human geography in his *Principes de géographie humaine* (1922) but passed away before its completion. The sum of his *oeuvre* set the agenda for Francophone geography for decades to come; yet, critical perspectives started to emerge with the fastening pace of industrialization in France and around the world as the twentieth century unfolded. Imbued with a certain traditionalism, his concept of *genre de vie* seemed ill adapted to the analysis of modern ways of exploiting the resources of the environment and to the growing role of cities as places of human dwelling and structures of interaction with the milieu.

The Interwar and Post-War Period: Emergent Themes and Subdisciplines

Even with the momentum imparted to the discipline at the turn of the century, World War I significantly slowed down the blossoming of human geography in France. As the State was busy with national reconstruction, few resources were available for the expansion of geographic education or the hiring of geography professors in institutions of higher learning. Still, various people worked to give the discipline a more formal structure, chiefly through the role of associations and scientific journals. The creation of the International Geographic Union (IGU) before the war brought about the formation of a French National Committee; the latter was initially presided by Emmanuel de Martonne (1873–1955) who eventually served several mandates as head of the IGU. The work of the IGU and its national committees helped focus geographic research in France and foster its links with other countries but it was the continued expansion of the *Annales de géographie* (created in 1892) that created a forum and a means of diffusion for professors and researchers in the field. Founded in 1821, the *Société de géographie* de Paris also represented a meeting ground for

French geographers but its focus on exploration was at times at odds with the production and diffusion of scientific knowledge outside the logic of colonial expansion.

Due to the continuing influence of Vidal de la Blache, rural geography remained a dominant theme for French geographers during this period. Although traditional methods and approaches remained important, new issues – among them, the mechanization of techniques, the depopulation of rural areas, or the growing linkages between countryside and urban centers – began to draw attention and called for original modes of inquiry. As the countryside seemed to lose some of its economic vitality and human resources to urban centers, territorial management and the rational development of planning structures became an important area of teaching and research for French geographers. The period following the events of May 68 also saw the beginning of a diversity of foci and methods that led to the gradual firming up of different specializations. These pulled researchers into increasingly well established subdisciplines in the mid and latter decades of the century, including, but not limited to, economic geography, political geography, and cultural geography.

Economic Geography

At the turn of the twentieth century, there were two schools of thought in economic geography. Long before the advent of geographers, some economists had integrated into their reflections, the role of space through different aspects of economic life such as communications, circulation, or the establishment of industry. On the other hand, geographers described, but rarely explained, the distribution of economic activity in the world. Economic geography was often linked to specific demands, such as the establishment of business and distribution frameworks. Many geographers worked in business schools and provided useful information to companies, a practice known as utilitarian geography.

Yet in this period, the works of Vidal de la Blache, especially his *Principes de géographie humaine* (1921), provided a rich departure point for economic geography, particularly his suggestion that every society is confronted with the same dilemma: individuals group together to share the advantages of the division of labor, but at the same time they must face difficulties inherent in this grouping. Vidal de la Blache was also interested in the role of transportation and the concentration (or polarization) of economic activities and cities as structuring elements of regional space. Geographers insisted on the role of competition and agreements altering it, as well as on the forces of conglomeration and dispersion.

During these early years, geographers contributed little in the way of textbooks or research on economic analyses. The work of Pierre George (1909–2006) was the one most frequently referenced: his *Précis de géographie économique* (1956) went through five editions between 1956 and 1970. Its first sentence provided the author's view of the subdiscipline: "Economic geography is the study of different types of production and location of consumption of various products existing in the world." It was essentially a catalog of various economic activities grouped by sectors. George presented an annotated statistical table that became quickly outdated without analyzing the economic logic that transforms the world. There was no mention of any work on 'new geography', nor was regional science mentioned. It was during this period that Paul Claval emerged as a leading figure of human geography, and its various subdisciplines, in France. Through his contact with Anglo Saxon geography on the one hand and modern spatial economics or regional science on the other, he wrote a text in 1963 entitled *Géographie générale des marchés* and, in 1968, he wrote *Région, nations, grands espaces*. With these works, he introduced new ideas in spatial analysis and proposed a renewal of the discipline through spatial economics.

In the 1970s, France, like many Western countries, was hit by an economic crisis and the limits to the Fordist development model. It became increasingly hard for the State to produce policies of regional development. By combining development with the 'environment', geographers working within the *Groupe de recherche européen sur les milieux innovateurs* (GREMI) research group analyzed how innovative environments are embedded in territories and local economies (among others, Gilly and Torre (2000); Pecqueur and Zimmermann (2004)). Gabriel Colletis (1955–) and Bernard Pecqueur proposed an analysis of spatial competition factors that put forward the idea of territorial specificity in supply. Economic geography was undergoing such a transformation that it was being referred to as a 'new social economic geography'. Research was meant to respond to social demands and explain the spatial behavior of capitalism. Interacting with these various influences, research was conducted by a team from Grenoble on 'localized industrial systems' under the supervision of Claude Courlet and Bernard Pecqueur. The sum of these contributions yielded a new school of thought on the notion of proximity emphasizing geographic and organizational proximities through institutions and collective learning.

Political Geography

Within the general framework of Vidalian thought, one finds a geopolitical dimension. By painting a broad picture of France's geography, he answered some essential questions relating to the existence of this country and its frontiers. Also of note during this period was the work of

Camille Vallaux who published a two volume book entitled *Géographie sociale*: the first volume was *La Mer* (1908), and the second *Le sol et l'État* (1911). Much of his inspiration came from Ratzel, but Vallaux took care not to fall into the determinism that French readers criticized in his *Politische Geographie*. Between the two world wars, French geographers offered an original contribution to geopolitical thought by approaching the problem of international relations from a perspective of balance and peace rather than from a standpoint of power.

With *Espace et pouvoir* (1978), Paul Claval helped broaden the framework of political geography by proposing that each set of institutionalized relationships is structured, in part, by power relations. His interpretation was structured around the Weberian distinction between pure power, authority, and different forms of influence and domination. Jean Gottmann's (1915–94) book, *The Signifiance of Territory* (1972), showed that the dominant forms of the political organization of space in the world emerged only in the seventeenth century. This theory was embraced by Bertrand Badie (1950–) in *Fin des territoires* (1995), but he added a new theme: the practice of sovereignty, the foundation of the classic state, was eroded by the multiplication of players on the international scene and by the appearance of supranational sovereignties. New avenues of inquiry in political geography were in the making.

Yet, the spread of a new political geography belongs to Yves Lacoste, in large part through the launch, in 1976, of a new journal known as *Hérodote*. During the same year, he published a book called *La géographie, ça sert, d'abord, àfaire la guerre*. Lacoste focused on the geographical position of a given area and analyzed it from several perspectives. He described its natural and social characteristics, which was the classic part of the approach. Then he presented the players, exploring the way they viewed situations, perceived challenges, and developed their strategies. In the French geographic community, nobody remained indifferent to this controversial study. Claude Raffestin (1936–) also proposed a global perspective on power in his book *Pour une géographie du pouvoir* (1980) where he focused on techniques of space management to analyze the territorial nature of power, which is in large part determined by modalities of access to resources.

Starting in the 1980s, political geography made inroads into the field of the geostrategy of nuclear deterrence. Jacques Lévy (1952–) and his colleagues also emphasized the significance of globalization and expanded their analysis to reflect on global political ethics. In the 1980s and 1990s, electoral studies multiplied and were renewed: with the rise of the urban metropolis and its expanding political role, the perspective on internal geopolitics developed by Béatrice Giblin Delvallet (1947–) and Yves Lacoste led to questions on the strategies of parties and elected members, as well as on the evolution of voting.

Cultural Geography

Notwithstanding the continued influence of the Vidalian approach, cultural geography in France expanded its perspectives in the latter part of the twentieth century to envision the concept of *genre de vie* within a larger analytical framework. In line with their Anglo Saxon colleagues, French geographers attempted to develop a more complex notion of culture in order to grasp intangible realities such as lived and symbolic space, or landscape as both a material and symbolic object. Many cultural geographers drew from semiotics in order to study the importance of language, myths, symbols, and images in the relationship between humans and their environment. Landscape was no longer seen simply as a component of a biological or regional whole but was also approached as a cultural production: as a result, geographers needed to understand the role of:

> perception, imagination, sign systems, power strategies, and the play between the different orders of reality (Claval, 1998: 362).

In doing so, they developed a multifaceted perspective on humanized spaces. In his book, *La région espace vécu* (1976), Armand Frémont (1933–) proposed that humans do not live in space 'as it is' but rather in space 'as they represent it to themselves'. He draws from literature in order to analyze how spaces are narrated and how these narrations (*récits*) in turn influence the experience of everyday environments. Frémont's approach brought social geography to the center of cultural research while also offering an alternative to the abstract logic of spatial analysis. Other cultural geographers studied collective representations of space in order to explore notions of place and identity, as well as culturally specific geographical perceptions and concepts: notable among them were Joël Bonnemaison (1940–97) and his studies of Vanuatu as well as Augustin Berque who has written extensively about Japan. Among many other contributions, their work allowed geographers to bring dominant conceptions of nature, space, landscape, territory, or the environment into a globalized dialog between different cultures. By making the human experience of space a central component of their analyses, French cultural geographers have, in the last 30 years, greatly expanded the purview of their discipline. In this regard, Jean Robert Pitte (1949–) charted a wholly original course through his research on sight, taste, and smell as central components of how humans apprehend, use, and represent space, an approach that is also strongly anchored in historical geography.

Indeed, cultural geography is not isolated from other subdisciplines in France: its impact is also found in tropical geography, led by Pierre Gourou (1900–99) whose extensive body of work contributed to defining the field of tropical studies in France as well as elsewhere. Gourou's initial research centered on the Tonkin region of Vietnam but eventually spanned the rest of Asia as well as Africa in his efforts to draw a global image and synthesize the human and economic dynamics of the 'tropical countries' (*Les pays tropicaux* 1947). The work of Gourou is strongly anchored in the colonial politics of his time: after World War II, he contended that the colonial enterprise had accomplished the triple objective that, in his view, legitimized it: it maintained order and peace; developed a communications network; and kept disease epidemics under control (*Remarques,* 1952: 97). In *Pour une géographie humaine* (1973), he posited that the human interaction with the landscape proceeds through a series of *encadrements* (enframings) much more than through the tools and techniques that are used to transform the environment. In this regard, his vision opposes that of Vidal de la Blache, despite his early association with the Vidaliens. In the postcolonial era, French geographers had been ill at ease with Gourou's legacy as his work was seen to have driven a wedge between the temperate and tropical worlds by presenting the former as rational, organized, and progressive while the latter, following the dynamics of orientalism described by Edward Said, was overly represented as its exact opposite. Thus, French cultural geographers in the late twentieth century challenged the static view of culture as the 'unique' quality of a place and people, to apprehend it instead as an evolving – and often subjective – set of practices, perceptions, and power relations that shape the surface of the Earth.

Francophone Geography Today: The Continued Broadening of an Intellectual and Physical Space

Throughout these subdisciplines, urban geography and historical geography were also well represented. In addition, Francophone geography has been developed over the last century through the contribution of Francophone geographers outside of France, whether in Switzerland, Belgium, Quebec, or other locales of the *francophonie*, notably in Africa, Asia, and the Caribbeans. In Quebec, Louis Edmond Hamelin – who co founded the geography department at Laval University – has been a leading figure since the 1950s chiefly through his work on *nordicité*, a term he coined to describe both the physical, cultural, and political characteristics of Canada. Notwithstanding its European influences, Francophone geography in Quebec is strongly North American due to its cross fertilization with English Canadian and American universities. Also, the presence of aboriginal cultures in Quebec and Canada means that native and postcolonial studies have been better represented than in France. Whether in Quebec or in France, however, one of the subdisciplines that has been left aside by Francophone geographers is that of feminist geography. Although many female geographers – among them Caroline Desbiens, Cynthia Ghorra Gobin, Béatrice Giblin Delvallet, Anne Gilbert, Damaris Rose et Marie Claire Robic – have contributed to the blossoming of French human geography, their work has not yet emerged into a well established feminist core in Francophone geography. Paired with emerging interests in studies of the environmental, global dynamics or even geopoetics, Francophone geography therefore will have no shortage of fruitful research avenues in the years to come.

See also: Anthropogeography (After Ratzel); Cultural Geography; Determinism/Environmental Determinism; Ecology; Economic Geography; Environment; Geopolitics; Historical Geography; Lamarck(ian)ism; Landscape; Political Geography; Possibilism; Reclus, E.; Regional Science; Ritter, C.; Rural Geography; Social Geography; Spatial Analysis, Critical; Tropical Geography; Urban Planning and Human Geography.

Further Reading

Badie, B. (1995). *Fin des territoires: Essai sur le désordre international et sur l'utilité sociale du respect*. Paris: Fayard.

Benko, G. (1998). *La science régionale*. Paris: Presses Universitaires de France.

Benko, G. and Lipietz, A. (eds.) (2000). *La richesse des regions. La nouvelle géographie socio économique*. Paris: Presses Universitaires de France.

Berdoulay, V. (1981). *La formation de l'école française de géographie (1870 1914)*. Paris: Bibliothèque nationale Comité des travaux historiques et scientifiques.

Berque, A. (1982). *Vivre l'espace au Japon*. Paris: Presses Universitaires de France.

Bonnemaison, J. (1996). *Les fondements géographiques d'une identité L'archipel du Vanuatu: Essai de géographie culturelle*. Paris: ORSTOM.

Bowd, G. and Clayton, D. (eds.) (2005). *French tropical geographies*. *Singapore Journal of Tropical Geography* 26(3), 271 374.

Brosseau, M. (1996). *Des Romans géographes: Essai*. Paris: L'Harmattan.

Claval, P. (1968). *Régions, nations, grands espaces. Géographie générale des ensembles territoriaux*. Paris: M. Th. Génin.

Claval, P. (1969). *Géographie générale des marchés*. Paris: Belles Lettres.

Claval, P. (1978). *Espace et pouvoir*. Paris: Presses Universitaires de France.

Claval, P. (1998). *Histoire de la géographie française de 1878 à nos jours*. Paris: Nathan.

Claval, P. (2001). *Histoire de la géographie*. Paris: Presses Universitaires de France.

Collignon, B. (1996). *Les Inuits: ce qu'ils savent du territoire*. Paris: L'Harmattan.

Fall, J. (2007). Lost geographers: Power games and the circulation of ideas within Francophone political geographies. *Progress in Human Geography* 31(2), 195 216.

Frémont, A. (1976). *La région espace vécu*. Paris: Presses Universitaires de France.
George, P. (1956). *Précis de géographie économique*. Paris: Presses Universitaires de France.
Gilbert, A. (1999). *Espaces franco ontariens*. Ottawa: Le Nordir.
Gilly, J. P. and Torre, A. (eds.) (2000). *Dynamiques de proximité*. Paris: L'Harmattan.
Gottman, J. (1973). *The Significance of Territory*. Charlottesville: University Press of Virginia.
Gourou, P. (1947). *Les pays tropicaux*. Paris: Presses Universitaires de France.
Gourou, P. (1952). Remarques sur les plans de mise en valeur des régions économiquement attardées. *Symposium intercolonial*. Bordeaux: Delmas, 97 10.
Gourou, P. (1973). *Pour une géographie humaine*. Paris: Flammarion.
Hamelin, L. E. (1975). *Nordicité Canadienne*. Montreal: Hurtubise HMH.
Lacoste, Y. (1985). *La géographie, ça sert, d'abord, à faire la guerre*. Paris: La Découverte.
Lefebvre, H. (2000). *La production de l'espace* (first publ. 1974). Paris: Anthropos.
Pecqueur, B. and Zimmerman, J. B. (eds.) (2004). *Économies de proximité*. Paris: Lavoisier/Hermes.
Pitte, J. R. (1983). *Histoire du paysage français*. Paris: Tallandier.
Raffestin, C. (1980). *Pour une géographie du pouvoir*. Paris: Litec.
Robic, M. C. (ed.) (2000). *Le Tableau de la géographie de la France de Paul Vidal de la Blache: dans le labyrinthe des formes*. Paris: Comité des travaux historiques et scientifiques.
Reclus, É. (1982). *L'homme et la terre* (first publ. 1905 1908). Paris: Maspero.
Reclus, É. (1995). *Histoire d'un ruisseau* (first publ. 1869). Aix: Babel Actes Sud.
Staszak, J. F. (1995). *La géographie d'avant la géographie: Le climat chez Aristote et Hippocrate*. Paris: L'Harmattan.
Vidal de la Blache, P. (1911). Les genres de vie dans la géographie humaine. *Annales de géographie* 20(111 112), 193 212, 289 304.
Vidal de la Blache, P. (1922). *Principes de géographie humaine*. Paris: Armand Colin.
Vidal de la Blache, P. (1979). *Tableau de la géographie de la France* (first publ. 1903). Paris: Tallandier.

Relevant Websites

http://www.editions sedes.com
Annales de géographie, Editions SEDES.
http://www.cgq.ulaval.ca
Cahiers de géographie du Québec.
http://www.espacesetsocietes.msh paris.fr
Espaces et sociétés.
http://www.enec.paris sorbonne.fr/index.html.
Espaces, nature et culture.
http://www.ville saintdie.fr
Festival international de géographie Saint Dié des Vosges.
http://www.herodote.org
Hérodote.
http://www.socgeo.org
Société de géographie de Paris.

Functionalism (Including Structural Functionalism)

J. E. Castro, Newcastle University, Newcastle upon Tyne, UK

Glossary

Dysfunction The opposite of function: mechanisms, roles, effects, utilities, and conditions that do not contribute to or may even negatively affect structural stability.

Function Mechanisms, roles, effects, utilities, and conditions that contribute to the maintenance of a given structure.

Functional Differentiation Process of specialization of the parts or mechanisms of a given structure to perform different functions.

Functional Equivalence When the same function can be performed by different parts or mechanisms of a given structure.

Functional Requisites The critical conditions needed for the existence of a given structure.

Holistic Functionalism Functionalist approaches that place the emphasis on the explanation of macro-structural social processes.

Individualistic Functionalism Functionalist approaches departing from the assumption that social structures exist for the benefit of the individual.

Social Darwinism The application of Darwinian concepts and principles for the explanation of social processes.

Teleofunction A function that is goal oriented, intentional, and consciously performed.

Introduction

Functionalism is one of the major traditions in scientific analysis and explanation. It has its main origins in biology and has been very influential in the development of the social sciences, including human geography. Functionalist approaches have also been very important in a wide range of other disciplinary fields, including psychology, engineering, architecture and planning, and archeology. This article places emphasis on functionalism in the social sciences, looking at the development of its main strands: functionalism and structural functionalism. It also provides a brief discussion on the concept of function, as well as a reference to the reemergence of functionalism in system theories, and more recently in neofunctionalism.

The main tenet of functionalism is that scientific analysis and explanation must focus on the functional integration of the differentiated component parts of a given structure or whole. It places the emphasis on the role played by functions or consequences of substructures and processes in fulfilling the needs of the system of which they are part. The earliest forms of functionalist explanation were developed in biology and focused on the functions played by the different organs in the maintenance of vital equilibrium in living beings. Biological or organismic functionalism became the model that inspired the development of functionalism in other scientific areas, most particularly in sociology and anthropology. However, although functionalism is often understood to be a unified school or approach, in fact, the functionalist field is home to divergent and even mutually contradictory positions, which has fueled protracted and still unresolved controversies. The divergence concerns in particular the balance in the emphasis that different strands of functionalism place either on functions or structures. Thus, when the focus is mainly on the former, the analysis seeks to explain the role that a certain component plays in the maintenance of the totality of which it forms part; while in the case of the latter, the aim is to explain the persistence of the structure under consideration by focusing on the functions played by one or more component parts. Moreover, the concept of function itself is the object of continued controversy, as it has often been applied with very different and even contradictory meanings. Functionalism, and its reformulation as structural functionalism, provided the prevailing paradigm for the social sciences during the 1950s and 1960s, but became the object of sustained criticism since the 1970s in a context of theoretical and methodological diversification. Although functionalism lost its dominant position, functionalist principles continue to inspire the development of a rich diversity of theoretical frameworks, including systems theory, some strands of Marxist analysis, and neofunctionalist approaches in the social sciences.

The Development of Functionalism in the Social Sciences

Although the genealogy of functionalist theorizing can be traced back to Greek philosophy, the development of systematic functionalism in modern scientific thought is associated with the work of the French paleontologist Georges Cuvier (1769–1832) who eventually became the father of functionalist biology. In his study of animal fossils, Cuvier developed the notion that animal organs can be explained by their functions as parts of the animal's integrated whole. He rejected the evolutionary theories of the time, such as Jean Baptiste Lamarck's, because he believed that any transformation of an animal's component parts would adversely affect the functional integration of its organism and hamper its very survival. Cuvier argued that there exist certain functional conditions ('conditions of existence') that restrict the form that animals can take given that the form and function of each animal part is integral to the organism as a whole. A different conceptual source of functional analysis in social science can be traced back to the Romanticist and historicist traditions, and particularly to Georg Wilhelm Friedrich Hegel (1770–1831). Hegel's concept of the uniqueness of the absolute spirit inspired an idealistic notion of structure, with component parts, interrelations, and functions derived not from organismic analogies but from a philosophical conceptualization of human history. Along these lines, the sociologist Karl Mannheim (1893–1947) developed an influential body of work where the concepts of structure and function played a central role.

Nevertheless, functionalism in the social sciences has been mainly shaped by biological thinking. Thus, it would be another Frenchman, August Comte (1798–1857) who would get inspiration from biological functionalism to develop a functionalist science of society, which he called initially 'social physics' and later 'sociology'. Comte's main preoccupation was to understand the principles underlying social order in the context of rapid social transformation characterizing nineteenth century Europe. He proposed that the study of society should distinguish between social dynamics (social change) and social statics (social structure), and he adopted the language of biological functionalism mainly for the latter. He assumed that there is a correspondence between social and biological matter in analytical terms, and comparing society with a biological organism he argued that social structures can be explained by focusing on how their component parts (i.e., individuals, families, religions, and social classes) contribute to the maintenance of the body social.

The systematic application of biological functionalist principles to explain social order was further advanced by Herbert Spencer (1820–1903), a leading exponent of social Darwinism. Spencer also established an analogy between social and animal organisms as systems that evolve through a process of increasing structural complexity and functional differentiation. According to him, social institutions mirror animal organs and would have the function to meet the survival needs of the social system. However, though Spencer shared Comte's concern with identifying the functional requisites that make social order possible, he departed from different assumptions and produced a very different response. He accepted that in the case of organic matter the parts have the function to preserve the systemic whole, but he counter argued that in the case of society the whole exists for the benefit of the individual members. Spencer's social Darwinism was a combination of evolutionary theory, functional analysis, and individualistic liberalism.

Emile Durkheim (1858–1917) rejected Spencer's social Darwinism but relied both on Spencer and Comte for developing a theory of society modeled on the main principles of organismic functionalism. Durkheim's main concern was centered on identifying the mechanisms of social integration that are needed for the preservation of social systems, which he believed could be classified as the cultural, structural, interpersonal, and cognitive functions of integrated social structures. For him, the explanation of social facts, which are the objects of study in sociology, requires identifying both their causes and their functions in relation to their relevant social structures. Durkheim applied this approach to a wide range of 'social facts', such as the incest taboo or crime punishment, seeking to identify existing correspondences between these facts and the needs of the social organism. Thus, for instance, according to Durkheim the punishment of crimes is caused by the intensity of the collective sentiments offended by such acts, but in turn punishment is a social fact that has the function of maintaining the level of intensity of such sentiments in the body social, which might otherwise be weakened and affect the integrity of the social system. For Durkheim, the main focus of sociological analysis should be on the mechanisms of solidarity and cohesion that help to explain the persistence of harmonious social systems.

A central principle derived from biological functionalism for the study of society is the notion that social systems tend to maintain a state of internal equilibrium (homeostasis), which underlies the works of Comte, Spencer, and Durkheim. One of the key social theorists that developed this principle further was Vilfredo Pareto (1848–1923), who built on Léon Walras' (1834–1910) economic theory of general equilibrium. Pareto argued that equilibrium was the normal state of social systems, and focused on the functions or 'utilities' of particular social phenomena in the achievement of overall systemic stability. He emphasized the importance of (a) social elements, (b) their distribution, (c) their change over time, and (d) their combination and interdependence with other elements in the process leading the social system toward its normal state of equilibrium. The heteronomous interaction between different elements of the system would lead to unintended, incidental effects that may have utility for the equilibrium of the system but are independent from the interests, intentions, sentiments, or desires of social actors. An additional contribution of Pareto to functionalist analysis was the further specification of the notion of functional alternatives or equivalents, that is, that the same systemic function can be fulfilled by different social mechanisms. He applied this concept in his theory of political elitism, where he argued that the maintenance of political equi librium by the ruling elite takes place through a number of mechanisms, including co optation, diplomacy, fraud, corruption, and force (complex forms of coercion). Through these alternative but nonexclusive mechanisms the ruling elite enforces its monopoly of power, which is the natural state of the system, although this equilibrium is not enduring and is subject to periodic revolutionary changes leading to the replacement of the ruling elite (elite circulation).

The development of a systematic functionalist an alysis in anthropology was mainly due to Bronislaw Malinowski (1884–1942), who both drew on and intro duced significant modifications to Durkheim's frame work. In particular, Malinowski went beyond establishing an analogy between organic matter and society to ac tually ground his analysis on the biological needs of human beings. For him, social institutions and cultural traits are functions developed by society to fulfill primary biological needs, such as nutrition or reproduction, thus reversing Durkheim's efforts to base the explanation of social facts exclusively on other social facts. Moreover, Malinowski also rejected Durkheim's derivation of functions from their role in maintaining social structures and adopted a utilitarian and psychologistic approach to the definition of social functions. Thus, Malinowski's functionalism postulated that traditions, social activities, and collective processes were in fact developments ori ented at the satisfaction of individual needs, which to some extent restated the principles of Spencer's indi vidualistic functionalism.

Functionalism and Structural Functionalism

Although the interrelation between the notions of func tion and structure is part and parcel of functionalist thought, the relative weight given to them in the analysis has led to the distinction between functionalism and structural functionalism, where the latter places greater emphasis on structural aspects such as systemic inte gration. The foundations for the emergence of structural functionalism can be traced back to Durkheim, and es pecially to the work that he carried out in collaboration with his disciple and nephew Marcel Mauss (1872–1950). In particular, in their work on human forms of classifi cation and symbolic thought they developed the argu ment that the human mind derives its systems of mental classification from existing social structures. In their perspective, the existing groupings and divisions in the social structures determine the modes in which the human mind develops its own system of classification. For instance, the notion of time in this perspective is shaped by the particular social structures characteristic of a given social group, and therefore it will differ between different societies. Thus, for Durkheim a calendar reflects the actual rhythm of social activities in a given social system, while in turn it has the function of ensuring the regularity of such activities. This work would exercise a powerful influence on the development of the structural functionalist approach, which is primarily associated with English anthropologist Alfred R. Radcliffe Brown (1881–1955). He claimed that there is no functionalism as such but only structural functionalism, and argued that the concept of function necessarily implies that of structure, namely a system of interrelated components that is maintained by the activities of the different parts. Rad cliffe Brown also held that social structures are natural systems of their own, composed of the mutual inter relations between individual members. In addition to Radcliffe Brown, the British school of structural functionalist anthropology was also populated by other key figures such as Meyer Fortes (1906–83), Edward E. Evans Pritchard (1902–73), and Max Gluckman (1911–75).

However, structural functionalism reached its height with the work of American sociologist Talcott Parsons (1902–79), who had become acquainted with the work of Durkheim during the 1920s as a student of Malinowski. Parsons went on to elaborate a complex synthesis of the work of Durkheim, Max Weber, Vilfredo Pareto, and Alfred Marshall, aimed at providing an integrated and holistic theory of society. Parsons' holistic functionalism brings together elements of a voluntaristic theory of human action, derived mainly from Weber, Pareto, Marshall, and a particular reading of Freud, with Dur kheim's concern with the integration of social systems. Thus, Parsons recognized the importance of individual agency and will in the pursuit of goal oriented behavior, with individual human beings having the capacity of making choices to achieve their objectives, but he maintained that the choices available to individual agents are circumscribed by the physical, cultural, and social environment, and shaped by social values and norms. Parsons defined functions as those mechanisms that allow the adjustment of the system to the external environ ment, and postulated that all social systems have four functional requirements, namely (1) adaptation, (2) goal attainment, (3) integration, and (4) latency (systemic equilibrium). According to him, complex socie ties are structured as a hierarchy of systems and sub systems, where each specific system must meet the four functional requirements in order to (1) adapt to the physical environment by securing the resources needed to achieve its goals, (2) appropriately manage such re sources, (3) provide for the effective coordination of the system's components and interrelationships to secure integration, and (4) maintain the system's overall equilibrium. Complex social systems develop specialized, highly differentiated means to respond to these four functional requirements, where the key mechanism is the adaptation of individual members of society to the system's social values and norms through socialization. Parsons normative functionalism and his emphasis on system stability informed much urban planning and design since the 1950s, heavily influencing the physical shaping and social organization of urban landscapes.

Robert K. Merton (1910–2003) reviewed Parsons' theory in a number of areas. First, he provided a counterbalance to what was perceived to be an overemphasis on structure in Parsons' approach, and placed more weight on the concepts of 'manifest' and 'latent' functions that give individual agency a higher profile in the overall configuration. Merton further specified these concepts: manifest functions are those mechanisms that contribute to the adjustment and adaptation of the system and are the result of intentional, conscious human actions, while latent functions are those that are neither intentional nor consciously carried out by the actors. Second, Merton also reviewed a number of implicit assumptions in classical functional analysis, such as that all components in a social structure fulfill certain functions needed for the stability and survival of the system. He argued that there exist certain components that do not fulfill any functions, while other elements that he called 'dysfunctions' may actually have a negative effect on the system (also termed malfunctions). Conversely, he also reasserted the principle of functional equivalence, namely that the same function can be performed by a number of different components of the structure.

Through the work of Parsons and Merton, structural functionalism became the dominant sociological orientation in the US during the 1950s and 1960s. It came to include a significant number of prominent sociologists, such as Kingsley Davis (1908–97), Wilbert Moore (1914–87), David Aberle (1918–2004), Lewis Coser (1913–2003), and Marion Levy (1918–2002) among others. The structural functionalist school also became very influential in the development of the social sciences in developing countries during the post World War II period. For instance, the Italian sociologist Gino Germani (1911–79), who spent a period at Harvard University working with Parsons after escaping from Fascist Italy in the 1940s, significantly contributed to the establishment and development of sociology in Latin America since the late 1950s from his base at the University of Buenos Aires, Argentina. During the same period in Brazil, Florestán Fernández (1920–95) developed a highly respected school at the University of Sao Paulo, blending classical Marxist and Weberian sociology with functionalist analysis.

The influence of the functionalist and structural functionalist traditions in human geography has been less conspicuous in terms of prominent figures or the development of schools of geographical thought. It is perhaps in the field of political geography, and particularly in the study of territorial boundaries and international relations, where an explicitly functionalist geography can be found. Thus, keynote geographers, such as Russian born French geographer Jean Iona Gottman (1915–94), developed functionalist explanations in the fields of urban and political geography, and particularly in his studies on the roles (functions) played by geographical (topography, hydrographic networks, etc.) and social (economic, cultural, and political) forces on international political stability. Gottman argued that the political partitioning of the world can be explained as the balance resulting from the interaction of two main forces, circulation (communication and movement) and iconography (the shared system of symbols of a given population). Along similar lines, the American geographer Richard Hartshorne (1899–1992) put forward his theory of territorial integration where he postulated that the main role of states would be to maintain stability by counteracting destabilizing centrifugal forces, such as disruptive physical features (e.g., mountains, forests) or social characteristics (e.g., low population, language barriers), that provoke the fragmentation of political units and foster intranational conflicts. Building on both Gottman and Hartshorne, American geographer Stephen Barr Jones (1903–84) proposed a unified field theory aimed at bridging the gap between morphology and function in political geography.

Functionalism has also been influential in the fields of urban geography and planning, where an explicitly functionalist conceptualization has been applied to a range of issues, including the study of neighborhoods, communities, and urban areas, housing policies, and territorial governance. Despite the fact that there are few other references in the literature to an explicitly 'functionalist' or 'structural functionalist geography', and that an in depth debate about the impact or implications of functionalism and structural functionalism for human geography is largely missing, many principles and concepts associated with the approach have been integrated in different ways in the work of human geographers. In particular, functionalist concepts have informed much urban geographical work on the spatial dimensions and scales of social order and disorder, organization, and conflict.

Criticism, Demise, and Revival of Functionalism

Since the 1970s functionalism, and in particular structural functionalism, became the subject of strong criticism on a number of fronts. First, because of its organismic analogies in the conceptualization of social systems and its emphasis on structural equilibrium and social order functionalist analysis was charged with been unable to account for social change. Second, for similar reasons functionalism was criticized as being an intrinsically conservative approach because of its emphasis on system integration, stability, and social consensus that would render unobservable the internal contradictions of social systems. Third, according to the critics functionalists would be unable to account for the causes driving the process of structural and functional differentiation that is at the center of functionalist analysis and they would have conflated the concepts of cause and function. Moreover, most functionalist explanations would be either tautological (the existence of a given function is explained as a response to structural needs, while the maintenance of the structure is explained as being the result of the functions) or teleological (the emergence of structures is explained as a response to the ends or needs existing in a given system). These and other criticisms of functionalism became increasingly louder during the 1970s, at the same time that the social sciences were experiencing a rapid diversification and fragmentation of theoretical and methodological approaches. According to some commentators this process would have led to the demise of functionalism to the point that it would have almost disappeared from the social sciences by the 1980s.

However, despite its critics functionalism has survived in several areas of scientific analysis and explanation. For instance, the work of functionalist political geographers like Hartshorne has been influential in the development of research on comparative border studies and regional integration (and geopolitical studies at large), a very dynamic interdisciplinary field, particularly in continental Europe. More generally, the German sociologist Niklas Luhmann (1927–98) developed a radical reformulation of Talcott Parsons' approach aimed at providing a comprehensive theory of society by bringing together elements from cybernetics, neurobiology, and German idealist philosophy that has also influenced developments in human geography. Luhmann proposed a theory of social systems whereby human beings are conceptualized not as part of society but as part of its environment. For this reason Luhmann's theory has been labeled posthumanist, as its main thrust is to de link the social from the human, breaking away from classical anthropocentrism. In turn, society is defined as a closed communication system, self referential and reflexive, and the function of the (economic, legal, political, and other) subsystems is that of reducing the increasing complexity of information flows to ensure system stability and survival. He also stressed the centrality of functional equivalence in the maintenance of systemic equilibrium. Luhmann's theory has informed new work in human geography, particularly in Europe and especially in Germany, aimed at developing new theoretical approaches to sociospatial analysis, although these authors do not draw as much inspiration from Luhmann's functionalism as they do from his sophisticated intertwining of systems theory and neurobiology.

While Luhmann's functionalism has clear conservative overtones, Oxford philosopher Gerry Cohen has argued for the application of functionalist analysis within the framework of Marxist theory. For Cohen, although functional analysis could not help to explain the original cause of a certain social structure (i.e., an institution or cultural trait that contributes to the reproduction of the capitalist social order), it helps to explain the permanence of such structure over time. An ongoing debate prompted by Cohen's work focuses on the compatibility of functionalist principles with the Marxist analytical framework. In this connection, some commentators have pointed out the influence of functionalist concepts in the work of Marxist geographer David Harvey, for instance in his explanation of suburbanization processes and their role in contributing toward the stability of the capitalist system by fostering consumption, commodity fetishism, and preempting revolutionary change. However, perhaps the most explicit and comprehensive attempt to revitalize and reformulate functionalist analysis has been carried out by sociologists Jeffrey Alexander and Paul Colomy, who have actually termed their approach 'neofunctionalism'. These authors aim at reinvigorating Talcott Parsons' ambitions of producing an integrated and multidimensional theory of society by relating it to other theoretical frameworks, preserving and at the same taking critical distance from Parsons' functionalism. These efforts have encountered significant resistance and prompted and ongoing and open debate about the usefulness and soundness of functionalism for contemporary social science.

The Concept of Function

One of the main criticisms to functionalism is the fact that the very concept of function is ambiguous, a carrier of multiple meanings, and is often applied in an unrigorous fashion without properly defining its specific content. Already in the 1940s Robert Merton warned about the terminological confusion affecting functionalist approaches, which often used 'function' as synonymous for a range of terms, including use, utility, purpose, motive, intention, aim, effect, and consequences. However, it is worth noting that this confusing application of 'function' has not been a preserve of the social sciences, and has also been identified in other areas, including biology and techno scientific disciplines. As discussed above, in the particular case of the social sciences one the main precisions introduced to specify the concept of function is that of manifest (conscious or intentional, also called teleofunctions) and latent (unintended and inci dental) functions. While the former places emphasis on agency and goal oriented social action, the latter is concerned with structural, unplanned effects, on social systems. In a recent methodological debate Mahner and Bunge have identified several distinct but logically interconnected meanings of 'function' used in biology, the social sciences, and techno scientific disciplines. An important specification of the concept proposed by these authors is that 'function' can refer either to the mech anisms that are responsible for a given fact or process or to the roles and effects of a given social element in re lation to the social system. The distinction is that while the former would allow to develop proper scientific ex planations, as it identifies a mechanism that explains how and why a given fact or result has occurred, the latter would at most offer descriptive accounts or exemplifi cations of known patterns of phenomena, but not scien tific explanations because they lack a reference to the mechanism responsible for the fact or process under analysis.

See also: Darwinism (and Social Darwinism); Determinism/Environmental Determinism; Gottmann, J.; Hartshorne, R.; Harvey, D.; Lamarck(ian)ism; Posthumanism/Posthumanistic Geographies; Structuration Theory; Structurationist Geography; Systems; Systems Theory.

Further Reading

Alexander, J. C. (ed.) (1985). *Neofunctionalism*. Beverly Hills, CA: Sage.
Cohen, G. A. (1978). *Karl Marx's Theory of History. A Defence*. Oxford: Oxford University Press.
Cummins, R. (1975). Functional analysis. *Journal of Philosophy* 72, 741 764.
Davis, K. (1959). The myth of functional analysis as a special method in sociology and anthropology. *American Sociological Review* 24, 757 772.
Gottman, J. (1951). Geography and international relations. *World Politics* 3(2), 153 173.
Gregory, D. (1978). *Ideology, Science, and Human Geography*. London: Hutchinson.
Hartshorne, R. (1950). The functional approach in political geography. *Annals of the Association of American Geographers* 40(2), 95 130.
Jones, S. B. (1954). A unified field of political geography. *Annals of the Association of American Geographers* 44(2), 111 123.
Knox, P. and Pinch, S. (2000). *Urban Social Geography. An Introduction* (4th edn.). Harlow: Prentice Hall.
Koch, A. (2005). Autopoietic spatial systems: The significance of actor network theory and systems theory for the development of a system theoretical approach of space. *Social Geography* 1, 5 14.
Lippuner, R. (2007). Coupling, steering, differentiation. The geography of social systems. *Erdkunde* 61(2), 174 185.
Luhmann, N. (1982). *The Differentiation of Society*. New York: Columbia University Press.
Mahner, M. and Bunge, M. (2001). Function and functionalism: A synthetic perspective. *Philosophy of Science* 68, 75 94.
Merton, R. K. (1949). *Social Theory and Social Structure*. New York: Free Press.
Pahl, R. E. (1967). Sociological models in geography. In Chorley, R. J. & Haggett, P. (eds.) *Models in Geography*, pp 217 242. London: Methuen.
Parsons, T. (1937). *The Structure of Social Action*. New York: Free Press.
Radcliffe Brown, A. R. (1965). *Structure and Function in Primitive Society*. New York: Free Press.

Relevant Websites

http://www.intute.ac.uk
Intute: Social Sciences.
http://www.sociosite.net
The Socio site.
http://www.socioweb.com
The Socio Web: A guide to sociology and sociological research.
http://www2.lib.udel.edu/subj/soc
University of Delaware Library: Sociology.

Fuzzy Set and Fuzzy Logic

Y. Leung, The Chinese University of Hong Kong, Shatin, Hong Kong

Glossary

Fuzzy Logic Nonstandard logic for approximate reasoning with imprecise premises.
Fuzzy Set Set with imprecise memberships, often characterized by a membership function.
Linguistic Variable Variable which takes on linguistic terms rather than numbers as its value.
Regional Boundary Line or transition zone, often imprecise, that separates regions.
Regionalization Grouping of spatial units into regions.
Rough Set Set with rough boundary, often characterized by a pair of lower and upper approximations.
Spatial Decision Support System Computer system that supports decision making by coupling human intelligence with the capabilities of the computer.

Introduction

Geographical systems are generally complex systems with omnipresence of incompleteness, imprecision, and subjectivity. While fuzziness is common in these systems, it has been avoided for the sake of precision and value free principle cherished in physical science. In the pursuit of precision, geographers conventionally assume that the systems are well defined, and we describe spatial behaviors by rigid mathematical models, which make no provision for the imprecision and subjectivity in the cognition, perception, and decision making processes.

In the description of spatial behaviors, we commonly use terms with artificial precision which fail to account for the intrinsic fuzziness of our natural language. In the treatment of space, precision in the subdivision of, and grouping for, regions is usually enforced, although regional boundaries are generally fuzzy. In spatial economic analysis, it is generally assumed that there is no fuzziness in our rational preference–utility–choice structure so that different goods or places can be precisely discriminated and selected. In achieving optimal choice, information on the availability of resources is usually assumed to be exact and perfect so that objectives and constraints can be precisely specified. In the construction of and operation on spatial databases such as geographic information systems (GISs) and remote sensing systems, two valued logic is commonly employed, although imprecision is rampant in spatial information processing. In spatial reasoning, truth of a statement is absolute, and fuzziness of valuation is not permitted. Binary logic again is employed as the backbone of inference, albeit our reasoning is mostly based on common sense embedded with imprecise terms or expressions.

With such relatively unrealistic assumptions, conventional spatial models often tend to be oversimplified, too mechanical, or too inflexible to explain and plan for the real world. To obtain a better approximation to and control of our spatial systems, it is essential to build theories or models which mimic human values, reasoning, and behaviors. Instead of de emphasizing or avoiding fuzziness, spatial scientists should directly deal with it in a rigorous manner. Such a view leads to a fundamental shift of paradigm in geographical analysis.

In recent decades, researchers have argued for the significance of formulating spatial models through fuzzy sets theory so that human spatial behaviors can be better approximated. It appears that fuzzy sets models are more general, realistic, and flexible in analyzing human spatial reasoning and behaviors. The purpose of this article is to provide a state of the art scrutiny of the fuzzy sets and fuzzy logic approach to spatial analysis over the years. The discussion ranges from the analysis of fundamental spatial concepts – the very basic aspect of spatial analysis – to the construction of intelligent spatial decision support systems – the most recent advance in geoinformatics and spatial decision analysis. A paradigm for spatial uncertainty analysis and directions for further research are given in the conclusion.

Fuzzy Spatial Concepts

Every scientific discipline has a set of fundamental concepts on which analyses are based, and geography is no exception. Concepts such as 'distance', 'direction', and 'connection' have long been identified as basic terminologies for the analysis of spatial phenomena. Traditionally, the formalization of these terms are based on the definitions given in the theories of mathematical analysis, such as those in the Euclidean or Minkowskian space. Under this system, the concepts are exact and their operational definitions have precise connotations. However, human behaviors have been demonstrated to be a function of the cognitive evaluative structure in which imprecision prevails. Consequently, traditional mathematical definitions of such concepts are inadequate in conveying the imprecision of human cognition and valuation. In place of the physical length of a given distance, the meaning of that length to the decision maker becomes more important in determining his behavior in space. A given cognitive term, such as 'long', places a restriction on the range of the possible values a base variable, such as physical distance, can take. Consequently, the study of spatial connotations of our common language is necessary in the understanding and prediction of human behaviors in space. The concept of a linguistic variable based on fuzzy sets theory is thus appropriate for such a task.

Since fuzzy terms such as short or long can be employed to describe cognitive distance, the fuzzy concept 'distance' can then be treated as a linguistic variable with linguistic terms, such as long and short, defined on the base variable (physical distance) as its values. Linguistic hedges such as not, very, somewhat, and extremely are employed as operators in modifying the primary terms to create new linguistic values. Operations such as union and intersection can also be applied to combine terms together to form new terms. All of these constitute the syntactic rule. The meaning of each linguistic value is a fuzzy set defined on the base variable.

In spatial analysis, the study of connectedness of objects often involves terms such as contiguity and adjacency. The concept of contiguity can be exact or fuzzy. In the spatial context, when we say two objects are contiguous we usually mean they are in actual contact along a boundary or at a point. Thus, either two objects touch or they do not. There is no fuzziness involved. However, when we say two points are near, the connotation is then fuzzy. Similarly, in the temporal context, contiguous in time ordinarily means next or near in time. This definition then becomes the definition of adjacency. In describing adjacency, words such as 'close' or 'remote' often have fuzzy spatial connotations. Therefore, regardless of whether the context is spatial or temporal, linguistic characterization of the concepts of 'connection' appears to be a more appropriate method in defining the degree of connectedness between objects. The method of linguistic variable is thus appropriate. Like distance and connection, the cognition of 'direction' also possesses fuzziness. Thus, linguistic characterization of this concept is again appropriate.

In addition to distance, connection, and direction, other spatial concepts such as cognitive points, lines, paths, and polygons of various geometric shapes can also be represented by the similar method. In addition to defining concepts, linguistic variables may also be employed to unravel spatial images on the location of points and lines, as well as shapes of different geometric figures. Leung and Yan, for example, have formulated a framework for representing fuzzy points, lines, and polygons in GISs.

Regional Concepts and Regionalization

Regions are fundamental analytical units on which most spatial analyses are based. Conventionally, a region is treated as a spatial construct which can be precisely identified and delimited. Therefore, boundaries separating regions are assumed to be clear cut. Contrary to this common conceptualization or practice, regions however may not be precisely identified and boundaries generally exist as gradients and should appear as zones rather than lines. The implication then is that a region is generally a fuzzy concept and a boundary is generally a fuzzy line.

The justification for such an argument is given by Leung in 1985. First, it is difficult to delimit precise boundaries separating regions and nonregions over a space on which phenomena distribute or vary in a more or less continuous manner. Regions merge into one another along the boundaries to a certain extent, because regions are part of a continuum in space. It is thus difficult to find a precise line that can separate two neighboring regions in a clear cut manner. Hence, a precise boundary does not appropriately reflect such a phenomenon. Second, terms such as 'hot' and 'dry' are adjectives with imprecise physical denotations. When they are used to classify regions, the corresponding boundaries become fuzzy. Third, regional differences are differences in degree rather than kind. Intermediate areas usually exist between regions. Therefore, regional boundaries are generally fuzzy and they separate regions to a certain extent.

Under such situation, regions can be characterized by linguistic propositions with mathematical translations which are essentially fuzzy sets. Since such characterization has fuzzy spatial connotations, the core and boundary of a region may not be precise. Under this situation, the core of a region is the points or area in space whose characteristics are most compatible to the linguistic proposition characterizing it. The boundary of a region is in fact a zone including all points in space whose characteristics are more or less compatible to the characterization of the corresponding region. That is, we are indeterminate on the precise value at which an individual characteristic can be considered insignificant with respect to a designated region. Therefore, boundaries exist as gradients rather than lines. Nevertheless, the boundary becomes a crisp line if imprecision vanishes.

Since boundaries are fuzzy, regions can only be separated to a certain extent. Applying the separation theorem, Leung gave in 1987 the concept 'degree of separation' which shows the extent to which regions can be separated. If they are completely different, then they can be fully separated. If they are similar to a certain extent, then they can only be separated to a certain degree. Thus, attached to a boundary is a measure which shows the degree of separation between two regions. Such conceptualization is more realistic and informative. It has led to a series of studies on spatial objects with indeterminate boundaries. Under fuzziness, defuzzification is possible if one wants to force a complete separation of regions even though they cannot be completely separated. Nevertheless, information loss will be experienced as a consequence.

Once regions have been characterized, we can assign spatial units to appropriate regions. A regional assignment procedure consists of a rule, or a set of rules, through which a discriminatory assignment of elementary spatial units to a set of predetermined regions can be performed. When both spatial entities are fuzzily characterized, it is natural to expect that a spatial unit can be assigned to several regions with varying degrees of possibility.

In additional to regional characterization and regional assignment discussed above, fuzziness also exists in the grouping for regions. Differing from fuzzy isodata analysis, partitioning of fuzzy relations via the concept of a fuzzy graph was proposed by Leung in 1984 to construct a flexible regionalization framework that groups spatial units into regions which are homogeneous or functional, hierarchical or nonhierarchical, disjoint or nondisjoint with either qualitative or quantitative information. It should be noted that the fuzzy set regionalization procedure becomes the conventional zero–one grouping procedure when there is no fuzziness involved in the similarity or distance measure.

Spatial Economic Behavior and Planning under Fuzziness

Conventionally, spatial economic theories and models are constructed on the assumption that individuals are capable of making clear cut discrimination of different goods or places. Consumers and producers are assumed to possess perfect information and precise cognitive and decision making capabilities so that individual or group preference is exact and the associated utility and choice structures are precise. In general, such preference–utility–choice structures have unrealistic presumption. The information available is usually imperfect and individuals are often vague in making a decision because they are unable to discriminate perfectly between goods or places. It has been argued that preferences and perceived utilities are often vague and the all or nothing preference and utility structures are just a special case of the more overwhelming fuzzy structure found in reality.

Employing fuzzy relations concepts, Ponsard defined in 1979 a consumer's strict preference structure as a binary fuzzy relation which is antisymmetric and max–min transitive, that is, a fuzzy nonstrict order. The indifference structure is expressed as a similitude relation which is reflexive, symmetric, and max–min transitive, and a consumer's preference indifference structure can be defined as a fuzzy preorder. The fuzzy preference structure allows for the gradation of preference intensity, and it is a generalization of its counterpart in classical theory.

In classical spatial economic analysis, a consumer maximizes his utility with respect to a budget constraint, while a producer maximizes his profit or utility of profit subject to a technological constraint. Due to imperfect information, inexact cognitive and decision making processes, and incomplete control of inputs and outputs, the budget and technological constraints often cannot be precisely specified. Thus, a consumer may have to maximize his fuzzy utility function subject to a fuzzy budget constraint, and a producer may have to maximize his fuzzy utility of profit subject to a fuzzy technological constraint. In general, this specification involves the maximization of a fuzzy objective function subject to a fuzzy constraint.

In addition to spatial equilibrium analyses, a variety of fuzzy spatial economic problems has also been analyzed. By modeling fuzziness in a rigorous manner, flexibility and realism of spatial economic models can be substantially enhanced. Analytically, its higher level of generalization makes the fuzzy sets approach both philosophically and mathematically appealing.

Urban and regional planning is similar to the regionalization and spatial equilibrium problems in that it often involves quantitative and/or qualitative information, subjective valuation, and imprecise decision making processes. Conventional methods are not flexible enough to effectively analyze problems with inexact objectives and constraints. For example, instead of imposing a maximum permissible resource usage α, a real number, decision makers may only be able to restrict it to a tolerance region 'approximately' α, a fuzzy subset. Instead of maximizing an objective function, planners may prefer to specify a fuzzy satisfaction region within which the optimal value of the objective function may be achieved. The maximization of a fuzzy objective function subject to fuzzy constraints is naturally a fuzzy optimization problem. Such an approach is especially important when the original nonfuzzy problem fails to yield an optimal solution due to the incompatibility of objectives and constraints, or when the time varying parameters are to be incorporated into the optimization framework. The framework of fuzzy mathematical programming has been employed to solve spatial optimization problems with fuzzy and/or nonfuzzy objectives and constraints.

Spatial Information and Intelligent Spatial Decision Support Systems

With respect to logical foundations, present day GISs are predominantly based on Boolean logic. It is especially so in data retrieval and overlay operations. It is well known in the GIS community that overlay by Boolean logic generally ends up with loss and/or misrepresentation of information. It forces artificial precision on the otherwise imprecise thought processes. It arbitrarily screens out intrinsic imprecision in the database which might be essential in the analysis. It fails to communicate the extent of imprecision or error to users. Therefore, results of overlay unravel more questions about data quality and boundary mismatch than they solve. Similar problems also exist in other GIS operations and cartographic modeling exercises. It is thus imperative to have a more appropriate logical system for the design of GIS. Such a system should be able to handle and communicate precise and imprecise information. It should also be instrumental in the design of intelligent spatial decision support system. Fuzzy sets in general and fuzzy logic in particular can play an important role in this area of research. Fuzzy logic is a nonstandard logic within which an inference is a deduction of imprecise conclusions from a set of imprecise premises. It can in a way be regarded as a 'fuzzification' of standard logic because it introduces fuzzy predicates and fuzzy truth values into the classification system. It reinstates the importance of impression in natural language and human cognition, perception, and thought.

Fuzzy logic, like standard logic, depends on some rules of inference to deduce conclusions from a set of premises. In standard logic, the antecedent in the minor premise (usually a fact) has to confirm or match exactly with that of the major premises (usually a rule) before the consequence of the major premises can be regarded as the conclusion. This, however, is not flexible and realistic enough to model human inference in which the antecedents of the major and minor premises might be approximately equal. In this situation, the best we can say is that the conclusion could only be approximately equal to the consequence of the major premises.

In knowledge based GIS, the following is a typical example:

- Major premise: If temperature is high, then pressure is low.
- Minor premise: The temperature in pixel X is quite high.
- Approximate conclusion: The pressure in pixel X should be quite low.

Here, the major and minor premises are imprecise because they involve fuzzy terms high, low, and quite high which can be represented as fuzzy subsets with nonsharp boundaries. The conclusion would not be possible under stand logic because high $\neq$ quite high and low $\neq$ quite low. However, such a reasoning process agrees with our intuition and can be handled by fuzzy logic through approximate reasoning.

Unlike standard logic, fuzzy logic allows the existence of fuzzy predicates in both the premises and the conclusion. We can deduce through fuzzy logic possibly imprecise conclusions from a set of imprecise premises. Given a major premise, we can theoretically deduce from minor premises with approximately equal antecedents, conclusions which are approximately equal to that of the major premise. This is a natural, flexible, and efficient way to manage inference in knowledge based GIS. Through fuzzy logic, we can implement essential features such as data structure, knowledge representation, a query language, rules of inference, acquisition of knowledge, and systems and user interfaces in intelligent systems.

To make GIS a more effective tool for spatial analysis and decision making, human intelligence has been regarded as a must, and knowledge based GIS has thus emerged in the research agenda. Domain specific expert systems for geoinformation processing and inference have been discussed or constructed in recent years. The possibility of employing fuzzy sets and fuzzy logic to build knowledge based GIS has been extensively investigated.

Conclusion

This article examined the development and applications of fuzzy sets theory and fuzzy logic in spatial analysis and planning. It has been demonstrated that the fuzzy sets approach provides an effective means to model human behaviors in space. It is especially versatile when the systems are embedded with intrinsic fuzziness. We have seen major theoretical and practical advancements in the last few decades with impressive results.

As argued by Leung in 1988, uncertainty in space and time can be due to randomness, fuzziness, or both. While probability theory is the paradigm for the analysis of randomness, fuzzy set is a theory for the formalization of imprecision. It is further realized that uncertainty can be caused by incompleteness and roughness of spatial information that can be formally described by the concept of rough set. Therefore, geographical uncertainty can be a result of randomness, fuzziness, or roughness with different frameworks of analysis. They are philosophically and mathematically different. However, uncertainty can be due to randomness and/or fuzziness and/or roughness. Under this situation, integration of the three paradigms will construct a comprehensive picture of uncertainty in geographical phenomena and processes. Such integration is depicted in **Figure 1**. Therefore, fuzzy set and fuzzy logic play an important role in geographical analysis under uncertainty.

It should also be noted that the successful integration of fuzzy sets and fuzzy logic with other paradigms such as neural networks and evolutionary computation is a

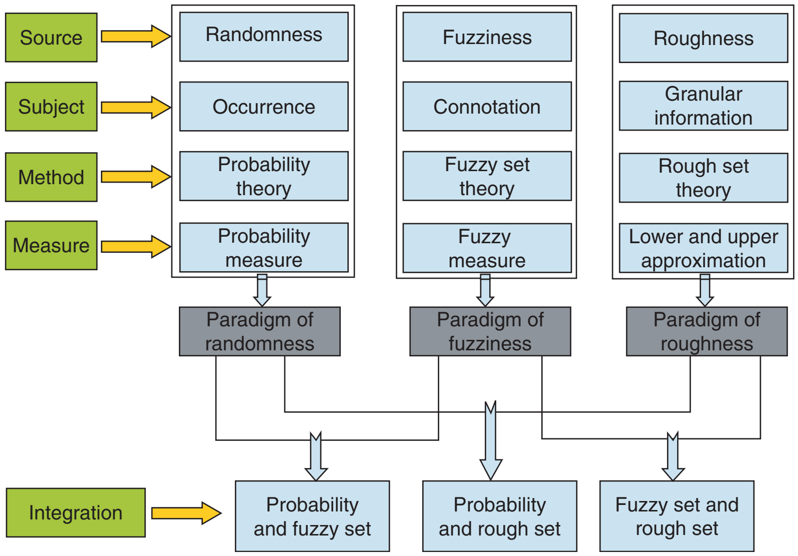

Figure 1 Integrated framework for the analysis of spatial uncertainty.

profitable direction for further development of the field. It leads to the formulation of more effective frameworks for system complexity analysis. For example, the inte gration of fuzzy logic and neural network for spatial in ference, spatial data classification, and rule learning have been demonstrated to be successful in recent years. The development of genetic algorithms for fuzzy rule ex traction has also been investigated. To make spatial de cision making efficient and effective, a general framework can be constructed for the development of hybrid con nectionist expert system which integrates rule based in ference, neural networks, and genetic algorithms into a seamless whole. Based on the trend of development, more significant results in the applications of fuzzy set and fuzzy logic in spatial knowledge acquisition, spatial learning, spatio temporal change detection, data com pression, and spatial reasoning are expected in the years to come.

See also: Artificial Intelligence and Expert Systems.

Further Reading

Alexandra, G., Vassilios, A. T., Evangelos, V., Christos, P. and George, S. (2006). Combining geographic information system, multicriteria evaluation techniques and fuzzy logic in siting MSW landfills. *Environmental Geology* 51, 797 811.

Altman, D. (1994). Fuzzy set theoretic approaches for handling imprecision in spatial analysis. *International Journal of Geographical Information Systems* 8, 271 289.

Bellman, R. E. and Zadeh, L. A. (1970). Decision making in a fuzzy environment. *Management Sciences* 17, B141 B164.

Burrough, P. A. and Frank, A. U. (eds.) (1996). *Geographic Objects with Indeterminate Boundaries*. Taylor and Francis: London.

Fisher, P., Wood, J. and Cheng, T. (2004). Where is helvellyn? Fuzziness of multi scale landscape morphometry. *Transactions of the Institute of British Geographers* 29, 106 128.

Guesgen, H. W. and Albrecht, J. (2000). Imprecise reasoning in geographic information systems. *Fuzzy Sets and Systems* 113, 121 131.

Leung, Y. (1982). Approximate characterization of some fundamental concepts of spatial analysis. *Geographical Analysis* 14, 29 40.

Leung, Y. (1984). Towards a flexible framework for regionalization. *Environment and Planning A* 16, 1613 1632.

Leung, Y. (1985). A linguistically based regional classification system. In Nijkamp, P., Leitner, H. & Wrigley, N. (eds.) *Measuring the Unmeasurable*, pp 451 486. Dordrecht: Martinus Nijhoff.

Leung, Y. (1987). On the imprecision of boundaries. *Geographical Analysis* 19, 125 151.

Leung, Y. (1988). *Spatial Analysis and Planning Under Imprecision*. Amsterdam: North Holland.

Patrick, B., Donald, K. and Fred, P. (2005). Fuzzy sets in database and information systems: Status and opportunities. *Fuzzy Sets and Systems* 156, 418 426.

Ponsard, C. (1979). On the imprecision of consumer's spatial preferences. *Papers in Regional Science Association* 42, 59 71.

Robinson, V. B., Petry, F. E. and Cobb, M. A. (2003). Special issue on incorporating fuzzy sets in geographic information systems. *Transactions in GIS* 7, 1.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control* 8, 338 353.

G

Gardens and Gardening

David Crouch, University of Derby, Derby, UK

Introduction

The garden has been an informing metaphor for geographical thought for sometime and as an affective material object and gardening as a process in the figuring and refiguring of space. It has represented an ideal environment and culture, a rather pre cultural, pre human state in a number of world religions, and continues to reappear in contemporary geographical discussions of the sacred. These leitmotifs of human geography are significantly theorized through ideology, discourse, and power, where 'the garden' becomes iconic. Signifying identity as well as status, cultural capital and social difference, as well as social/cultural relations, the garden and ways of gardening emerge as expression. A more complex conceptualization of the garden and gardening emerge in debates concerning consumption, commodification, and identity. In recent decades, the garden as artifact has been increasingly transformed to gardening as practice and as significant in developing critical conceptual approaches to a range of 'new' cultural geographies. These shifts and developments accompany the increasing geographical interest in process, practice, and performance. The 'nature' dimensions relating to, and perhaps informed by, gardens and gardening emerge in new ways in terms of reconceptualizations of nature where significance and meaning may emerge through practice, and in relation to the nonhuman; and debates concerning the ethical and moral in human geography, including shifting symbolism of the garden and of gardening in relation to war and peace. These developments in human geographies have been enmeshed with wider humanities and social science thinking and beyond these, from art theory and social anthropology to environmental debate.

Garden Iconographies, Materiality, and Sacred Space

The significance of the garden in world religions is pervasive, demonstrating its power to define space. In ancient Mesopotamia and in Muslim traditions, the date palm, material object, was considered the tree of life in the Garden of Eden. Ancient summer palaces of China incorporated in their gardens shapes to reflect and to contain cosmological visions, magic powers, and *feng shui*. The garden, as Eden, symbolized the originary and pre Fall, pre cultural, a particular notion of nature. Paradoxically combining reference to such appeals to purity with excuse, or opportunity, for erotic display, Renaissance painting frequently used religious or classical imagery in rendering the garden as idyllic playground, its own version of the garden.

More secularly, the garden became a signifier of Englishness, generally conservative and redolent of organic rural values. It was used and understood to symbolize English identity and its contribution to global civilization during the English country house period of the last three centuries. Such design power sustained into the twentieth century, during which gradually a more mechanized engagement with modernity was developed, symbolizing a new iconography of progress that became increasingly widespread across Europe. Sometimes echoing this European version, the garden as yard in North America displayed status; idealistically combined individuality with community at the front: family and home values at the rear. Persistent return to the sacred dimensions of what the garden signifies is prevalent in photography, art and literature; gardens as sacred spaces were depicted in cultures diversely across the world and in the geographies of world religions. A particular version of nature as perfected by cultivation's civilizing imprint pervades such models, but there can be also a more relational vision where the garden is used to represent a more relational mix with culture and human action, including a mixture of control and liberation.

The Garden, Ideology, Discourse, and Power

As the Italian Renaissance garden, later copied but refigured in English garden design, was set aside as apart from the world, its iconography was also significantly deployed in the pursuit of power by wealthy landowners. Particular components of layout and design, iconographic in themselves of a particular version of high culture; of orderly adaptation of nature; and their security and private character open only to select guests ensured their role in display of importance to those permitted to view them, as they became associated with the ordered leisure time of the upper class. They were used to confer status. Beyond their boundaries these significant garden spaces emitted their power through their associated delimitation of boundaries, and clearances of ordinary inhabitants in their wake, to provide more space and cleanse [sic] the view. This iconography was affirmed through a very visual manner – in their physical design and visual consumption. Thus the garden played a political discourse of class and its appropriated culture. This was the garden as landscaped for purpose. The focus upon the garden in these ways helped progress critical discourse in cultural, humanistic geographies and drew attention to the compound of vision, space, and power. These features combined with critical reappraisals of art, especially painting, and its deployment in garden design, to which they were often prototypes. Paintings as objectification of discourse played a symbiotic role with the objectification of discourse in the garden. While holding on to notions of the sacred even if formalized culturally, these gardens, like the paintings, also served as opportunity for erotic display, for example, in sculpture, often drawing upon classical motifs.

A wider range of iconography of the garden has more recently emerged, problematizing earlier versions of landscaped garden to reveal new geographies. These formal gardens have been identified as key influences on everyday contemporary gardens. However, these interpretations may be reductive as other influences deserve note. Such gardens include: the garden as necessitous to feed the household and at the same time provide sensory delight; and the iconography of individuals' small, mixed, and cottage gardens, community gardens known variously worldwide as allotments, vacant lot gardens, *kleinegarten*, or *petits jardins*. The latter may be considered 'compensatory' for the absence or lack of either a garden space adjacent to the home, as rented rather than owned, or 'replacing' larger cultivable spaces following its appropriation (enclosure of common lands, urbanization). Moreover, they may be considered gardens of particular social and cultural opportunity.

Together, these more avuncular spaces provide distinctive examples of the vernacular that combine cultural context with a distinctive iconography: the mixed use and unselfconscious layout of small parcels of land to sustain the household; and the plot of land usually separate, or distant, from the house. More broadly the the vernacular character of household garden spaces speaks of distinctive identities, values, meanings, and attitudes. Their design and outward appearance can make a narration of complex social and cultural influence and projection, everyday production and popular imagery.

Each of these 'mundane' gardens and their work produces distinctive ground for interpreting space and culture; and present a complexity of significance, context, and action in their geographies. They contain their own character of nature, sacred identity, and political discourse. In Cuba, the garden has become resonant of political struggle, holding onto Cuban identity and subverting trade embargoes. In Berlin, New York, Russia, and Britain the community garden is significant in the politics of land, development, and identity.

Garden, Gardening, and Cultural Identity

Acknowledging the complexity and diversity of the garden, especially in contemporary modernity, continues to be a focus of geography and across the humanities and social sciences as well as depictive of landscape in artwork, in increasingly diverse ways. For example, the garden has been a focus of debate concerning gender, the domestic and 'the home', and ethnicity.

Class underlined the distinction of the landscaped garden across Europe and its colonies, notably India. In the twenty first century, class remains significant, demonstrated in the use and deployment of distinctive design, plants grown, and evidence of display.

Gardens bear distinctive forms of ethnic identity. In their design and in plants grown, ethnic identity can be conferred. Caribbean and Bangladeshi gardens in England, for example, contain once exotic crops only just becoming popular and commonplace in once more temperate gardens. In Oslo, Norway, allotments can be geographically separated according to Turkish growers with high intensity clear ground food cultivation and a Norwegian middle class white population more intent on decorative planting, with accompanying cultural–political tensions.

Gender, household, and familial relations linked with ideas of the domestic, the home, emerge in geographies of the garden. Gendering in the garden, curiously stereotyped by men growing food and women flowers [or by men for 'the wife'], includes also ways in which the articulation of garden and indoor space are constituted in gendered ideas of the domestic, expressed in the use, design, apportionment, and demarcation of activity and distribution of labor. During the nineteenth century, gardening was familiarly and habitually part of women's work in the domestic division of labor among both working class households and where land work as small scale farming was done by men. More recently, the wider evolution of garden from space essential to the welfare of the family/household to spaces of display has made this domestic, gendered relationship more complex and nuanced, and potentially contested. It is possible too that age and generation can be distinguished in the way in which these elements are distributed.

However, the garden and ethnicity and other social divisions and distinctions are not merely significant in terms of different, distinctive ways of cultivation but in terms of values and gardening as process. Social and cultural identity and distinctiveness remain just as significant in the practice of gardening as in the garden as idea and object.

Consumption, Commodification, and the Identity Process

The study of space in terms of consumption has prompted and been informed by work on commercialization on the character of participation in gardening and what the garden means in contemporary culture. Attention has been drawn to the increasing importance of commodification of the garden. The garden has become objectified as a commodity underlined and contextualized by its commercial, marketing imagery, open to manipulation as material culture. In this way the garden arguably becomes narrowed to its purpose in status making – particular styles of garden and plants, planting, and cultivation an exercise in achievement. Garden objects, as plants, furniture, and design contextualized, are given value and meaning through the construction of style in the commodity – monetary value, fashion, and the act of purchase conferring significance, cultural identity, and status. Such extension of the garden connects notions of domesticity and the increasing significance of the home in 'commodity culture'. Commodification of the garden and gardening is accompanied by more widely mediated cultures in specialist gardening and gardens, and more popular magazines and other literature, through the global spread of television 'garden makeover' programs, as well as both more earnest and more stylized bespoken garden information. Thus the garden and gardening are transformed along commercialized and stylized discourses, transplanting significance and materiality from other more popular and everyday impulses.

However, this emphasis on commodification in consumption may deliver an over reductive discourse. Consumption includes refiguring of the value of commodities, and their reworking in relation to the individual's life, and its intersubjective territory. In this more active process of the consumer doing work, constituting meaning and value with the purchased objects and ideas, individuals may engage these prefigured values associated with purchased products in their own ways, through their own lives, and identities confer their own values and meanings.

Moreover, there is a significant current in gardening and in the use and signification of gardening that derives its significance from intersubjective and cooperative practice. Gardening is familiarly, still in the twenty first century, an act of friendship and communality; informal networks of knowledge and practice; and of sharing ideas and informal, cooperative, soft competition. Competition as evidenced in flower shows can be highly competitive but not always so. Ideas on garden design, what to grow, and how to grow can be shared and convivial, partly informed by the garden media and commercial seduction, but refigured socially and culturally. Communal networks through gardening practice convey further social and collaborative practical ontology.

Practicing gardens includes garden visits possibly to encounter demonstrations of 'high culture' (such as landscaped gardens typical of much of Europe and their transplants in North America and elsewhere) and everyday gardens often included in locally available 'gardens open to the public' schemes. Such may be regarded as exemplars of stylization and accompaniments to commodification. Yet, this may not necessarily be so. Their respective significance may be considered as exercises in sightseeing and in relation to exploring identities, rediscovering individual or shared, local and national histories and heritages, and in complementing leisure interests in gardening.

Gardens emerge from these discourses as both private and public. Familiarly, the front garden, at the front of the house is on view to the public anytime as a site for display. The back garden is a back space concealed from wider observation and surveillance. However, the back garden may be also the site of significant social intercourse, barbecue, and relaxed intersubjectivity where friends are invited. The front garden may play a similar role in sociality as the site where gardening combines meeting people passing by on the street.

The community garden is significantly a public and a private space; a site of public encounter and private reflection. Typically situated among anywhere between 20 and 100 small garden plots, often rigidly gridded in orderly delivery of land ownership, these spaces may profer relative isolation in their opportunity for individuality in cultivation, wider practice, and relaxation. However, they are frequently also open to collective view among other plot holders; and they may offer view from surrounding areas. While they continue a heritage of domestic relations (traditionally a male exercise, less so today), the informality of community gardens and their location outside notions of domestic space and bespoken designer mediation make their opportunity for alternative, even subversive design and cultivation possible. Thus, these small spaces are characterized by an often makeshift opportunity as a distinctive self expression, a vernacular landscape, and as an allotment aesthetic.

Gardening as Embodied Practice and Performative Process

In both a distinctive departure and yet bearing traces of earlier garden discourses in human geography and re lated fields, more recently, the idea of the garden has become reactivated anew in cultural geographies and geographies of embodied practice and performance. The garden has been verbalized into gardening, cognizant with the increasing geographical attention to process and subjectivities. Through writing on practice, performance, dwelling, and lay geographies, these contributions reen gage the active process of human encounters; the re working of cultural context and refiguring of geographies, and conceptualizations of nature as well as the non human.

Moving somewhat beyond the visually geared inter pretations of the garden as iconography, new cultural geographies of gardening articulate the importance of practice, in consideration of so called nonrepresentational geographies, in acknowledgment of the complex and nu anced iteration between embodied practice and contextual representations. Phenomenologically encountering the world and its spaces bodily subjectively and inter subjectively, multisensually, and expressively, the world becomes perceived through both feet as well as eyes. Practice is enlarged through a consideration of perform ance: ritualistic, stylized, encountered in a life journey of exploration as well as self identification in performativ ities, with potential to 'hold on' and to 'go further' in one's life through mentally reflexive embodied encounters and interplay with memory.

Gardening participates in a making of lay geography, temporal, complex, and contingent. It holds the potential of difference, of becoming, incompletely constrained by context and prefigured iconography and significance. Gardening holds potential for creativity, rather than merely mundane continuity and confinement in available garden categorizations. Objects encountered, made or responded to, in gardening may be significant in the ways in which doing gardening may be felt to be important. Objects convey their material culture in ways that may bear prefigured significance. Individuals may negotiate and reconstitute these prefigurations in the process. Gardening is thus a bodily process of active reflexivity.

This process may include a component of well being through physical activity, mental stimulation and re flection, and particular rhythms. Private and community gardens can be used as therapeutic schemes for offenders and the mentally ill. Arguably working with and along side plants may confer an open experience of relational care. Yet, countering idealization, garden space may be used and valued very differently. It may be the location of storage, accumulation of domestic materials, including the discarded, where no interior space is available, and gardening a chore of simply keeping tidy, a site for coping or the expression of not coping with life.

Gardening may offer opportunity for attention to physical matter of material culture – and nature – in a more visceral way. The practiced lay geographies of gardening provide an example of practical ontologies, constituting and negotiating identities relationally with the formally prefigured identities of cultural contexts. Furthermore it may offer opportunities for relational ontologies with nature that circumvent the formal ontologies of nature as idealised, set apart, pre cultural. Thus nature may be constituted through gardening and its embodied, visceral performance with imagination, metaphor, and fleshy materiality that are familiarly en countered in garden practice.

Back to Nature and Beyond

The garden emerges historically in powerful icon ographies of purity, cosmology, religion, and the sacred; a moralistic existence of the pre Fall; and the pre cultural. It was the garden not wilderness that signified perfection. It represented a particular contract of pre cultural, pre human perfection through cultivation that combined a complex discourse concerning human–nature relations. This powerful story persists in world religions, and nonreligious beliefs, even in contemporary artwork, with distinctive moral and ethical import. Over the last cen tury, the reemergence of a more ecologically sensitive and aware discourse repositions this thinking. With ref erence to the garden as more than private space, as generic symbolization of human–nature relations, a re valuation of ethical responsibility emerges. Gardening can offer opportunity for ethically sustainable practices and nature relations, yet not exclusively. The last 100 years of experience suggests its equal opportunity to af fect an ethics of control and exploitation even at a do mestic level. Recent discussion includes the relation between gardening as a human directed act, and the ways in which the nonhuman, plants, and their growth interact and influence human activity and reflection. The non human is conceptualized as deserving ethical respect. Wilderness accompanies this garden vision as a softening of technocratic, modernized, mechanical domination. Gardening and the garden familiarly involve negotiations of threads of control, domination, and power; liberation and enabling, soft management; cooperation; and active relation and respect.

Nature's effect on and through performance and sig nification in gardening can be significant too. Distinct hybridities of practice, feeling, and possibility emerge through the multiple interplay of human and nonhuman actants. Nature can be constituted through doing; through relationships built and imagined with the materiality and felt experience of the nonhuman. Thus gardening is not static, but its action and meaning con stituted in the process of its doing. This process includes a practiced relationship with nature and wilderness – a negotiated control and cooperation.

Moral geographies can pervade thinking on the gar den and gardening in terms of their association of correct ethical use, appropriate display, and iconography: the landscaped garden of the wealthy landowner and the painted landscape; the ordered domestic garden; and the owner authorized tidy rented plot. Because of the semiotic depth of gardening, these moral judgments and conflicts persist. Gardening is familiarly linked with reverie; sacred significance, and poetics. Poetics of gar dening include reference to gardens' iconography and emerge through action. Poetics can be conferred in art, literature, and formal design. It can also emerge through performance, practice, and human–nonhuman relationality.

The garden, and gardening, may be interpreted as a foundational trope of fixity – originary human civil ization in settlement and continuity. Yet, this is only one version. Gardening, in its processual and potentially performative modes, can offer an exploration of life – identity and creativity, memory and possibilities – in perpetual contingency and active refiguring as well as an imagined permanence.

The iconographic power of the garden to signify peace as well as practical and poetic possibility in pro viding nourishment, offering exercises of beauty, and expression of hope has been rendered in gardens made in war time, in war zones of extreme social, political, eco nomic, and cultural conditions made in, for example, Warsaw and other ghettos during World War II; in prisoner of war and civilian internment camps; the recent Gulf Wars; and behind the trenches, by soldiers, during World War I, as both humanized landscapes and ex perience. Again the garden metaphor and materiality of performance is echoed in similar ways in the cemetery, where the significance of the garden and gardening the cemetery and graveyard are given acute significance in the working of memory.

See also: Body, The; Landscape; Nature.

Further Reading

Alfrey, N. and Daniels, S. (2005). Exhibition catalogue *The Garden*. London: Tate Gallery.

Chevalier, S. (1998). From woollen carpet to grass carpet: Bridging home and garden in an English suburb. In Miller, D. (ed.) *Material Culture: Why Some Things Matter*, pp 26 46. London: Routledge.

Cosgrove, D. and Daniels, S. (eds.) (1988). *The Iconography of Landscape*. Cambridge: Cambridge University Press.

Crouch, D. (2003). Performances and constitutions of natures: A consideration of the performance of lay geographies. In Szerszynski, B., Heim, W. & Waterton, C. (eds.) *Nature Performed*, pp 15 30. Oxford: Blackwell.

Crouch, D. (2003). Spacing, performance and becoming: Tangles in the mundane. *Environment and Planning A* 35, 1948 1960.

Crouch, D. (2004). Gardening as a construction of landscape. In Shneider, U. & Wolshke Bulmahn, J. (eds.) *Gegen den Strom*, pp 57 72. Hannover: University of Hannover.

Crouch, D. and Ward, C. (1988). *The Allotment: Its Landscape and Culture*. London: Faber and Faber. [5th edn., revised 2007, Nottingham: Five Leaves Press.]

Francis, D., Kellaher, L. and Neophytou, G. (2005). *The Secret Cemetery*. Oxford: Berg.

Groening, G. and Shneider, U. (1999). Design versus leisure: Social implications of functionalist design in urban private gardens of the twentieth century. In Crouch, D. (ed.) *Leisure/Tourism Geographies*, pp 149 163. London: Routledge.

Hitchings, R. (2005). At home with someone non human. *Home Cultures* 1, 169 186.

Helphand, K. (2006). *Defiant Gardens: Making Gardens in Wartime*. San Antonio, TX: Trinity University Press.

Ingold, T. (2004). Culture on the ground: The world perceived through the feet. *Journal of Material Culture* 9(3), 315 340.

Jackson, J. B. (1994). *Sense of Place/Time*, pp 119 134. New Haven, CT: Yale University Press.

Jones, O. and Cloke, P. (2001). *Tree Cultures: The Place of Trees and Trees in Their Place*. Oxford: Berg.

Lee, R. (2000). Shelter from the storm? Geographies of regard in worlds of horticultural consumption and production. *Geoforum* 31(2), 137 157.

Longhurst, R. (2007). Plots, plants and paradoxes: Contemporary domestic gardens in Aotearoa/New Zealand. *Social and Cultural Geography* 7(4), 581 593.

Westmacott, R. (1993). The gardens of Afro Americans in the rural south. In Hunt, J. D. & Wolshke Bulmahn, J. (eds.) *The Vernacular Garden: Dumbarton Oaks Colloquium on the History of Landscape Architecture 14*, pp 77 105. Washington, DC: Dumbarton Oaks.

Whatmore, S. (2002). *Hybrid Geographies: Natures, Cultures and Spaces*. London: Sage.

Garrison, W.

R. Morrill, University of Washington, Seattle, WA, USA

Garrison, William (1924–)

Bill Garrison is perhaps the single geographer who had the greatest influence on the transformation of the discipline into a modern science in the latter half of the twentieth century (**Figure 1**). Born in 1924 in Tennessee, he started college at Peabody in Nashville in 1940, but service with the US Air Force in the Pacific as a meteorological officer intervened. He returned to Peabody for a BSc and then an MA in geography, then on to Northwestern, receiving his PhD in 1950 in urban geography – a quite traditional geographical training.

Indeed, the then radical notion of developing theories of location and movement, and the use of statistics and mathematical models, did not develop until some time after he came to the University of Washington in 1950.

A visiting research appointment at the University of Pennsylvania (1952–53) was the trigger for a vision of a revolution for economic and urban geography, following Garrison's interaction with economists and statisticians. Back at the University of Washington he offered a geography course in statistics in 1954 (probably the first of its kind in the world), and in location theory in 1956, as well as revising transportation geography along these lines. The underlying goal was the search for explanations of the economic landscape, to understand the patterns of urban and rural land use, of economic activities and of transportation and movement, as well as the spatial processes that created them, and to test these theories/explanations with appropriate methods. We stress this as Garrison was (correctly) uncomfortable with the term 'quantitative geography' that later became attached to the endeavor, as it was really a concern with theories of spatial process.

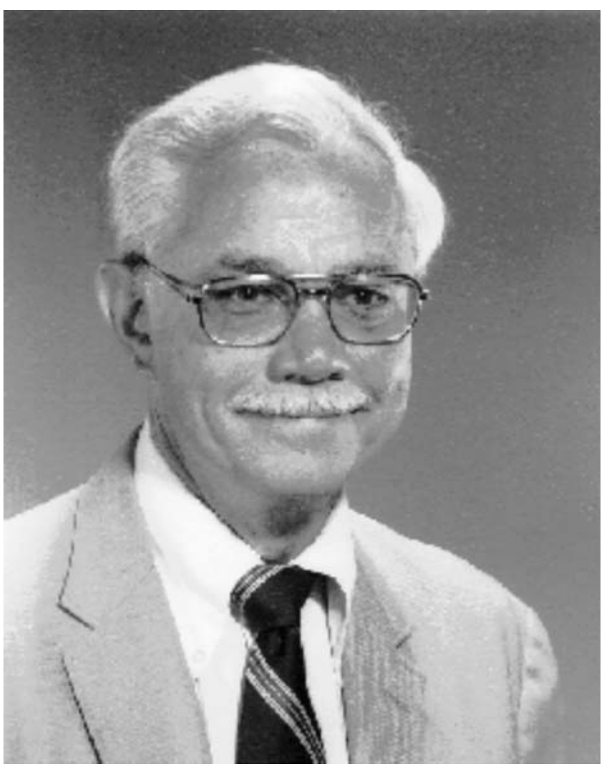

Figure 1 W. Garrison. Courtesy of Dr. Garrison, W.

Garrison's new direction was supported by other key faculty, and by one of those accidents of academic history, the arrival, between 1954 and 1960 of a group of graduate students who became the so called 'space cadets', who fervently accepted the theoretical–statistical challenge, and took the responsibility to proselytize the discipline as they took positions across the country: for example, Brian Berry, Duane Marble, John Nystuen, Richard Morrill, Michael Dacey, William Bunge, Arthur Getis, and Waldo Tobler. Internal interaction within the group was intense, and external communication was furthered by Garrison's critical support of the informal departmental discussion papers that circulated around the world.

Garrison was devoted to interdisciplinary study, and expected his students to take advanced courses in economics, business, sociology, and other fields. This interest took the form locally of interdisciplinary seminars and research groups and externally by joining the new field of regional science. Garrison and his students (e.g., Marble and Tobbler) were pioneers in the early beginning of computer use and programming, including what later became geographic information systems and computer cartography. Most significant in his future career – and to these early students – was active participation in interdisciplinary funded transportation research (geography, business, civil engineering, and planning). From this came Garrison's seminal publications: 'The benefits of rural roads to rural property' in 1956, 'Economic impacts of highway improvements' in 1957, 'Geographic impacts of highway improvements' in 1958, 'Estimates of the spatial parameters of spatial interaction' in 1956, 'Connectivity of the interstate highway system' in 1960, and a book – *Studies of Highway Development and Geographic Change* in 1960 – which included chapters from student dissertations. These pioneering studies solidified Garrison's lifetime concentration in transportation. His concern for a much wider conception of the costs and benefits and spatial consequences of transportation innovation and investments proved perhaps as unorthodox to the tradi tional transportation establishment, as did his dedication to science in geography.

Garrison helped further change geography through joint publications with his students, sometimes providing the first theoretical–statistical study in a given field, as was the case with Brian Berry (four famous studies on central place systems in 1958), with Duane Marble (agriculture in 1957), and Richard Morrill (models of movement in 1960). Particularly important was Garri son's introduction into geography of the revolutionary ideas of spatial theory and statistical method, as in *Ap plications of Statistical Inference to Geographic Research* in 1956, and *Spatial Structure of the Economy* in 1959 and 1960.

In 1960 Garrison moved to Northwestern University, with even stronger programs in transportation, not to mention better access! His 7 years at Northwestern brought more converts and aided in the growth of many Midwestern centers of the new geography, furthered by a series of National Science Foundation sponsored sym posia and workshops in scientific geography. Key contri butions of the work at Northwestern were in the analysis of transportation networks. Influential publications in cluded *Toward a Simulation Model of Urban Growth and Development* in 1963, *Quantitative Geography* (edited with D. Marble in 1967), and *The Structure of Transportation Net works* (with D. Marble in 1968).

A joint appointment in civil engineering, becoming Director of the Transportation Center and involvement with large transportation research studies and national transportation research committees helped bring about his move from geography to transportation by his move to civil engineering, at the University of Illinois, Chicago (1967–69). One important activity from this period in cluded starting a long term involvement and service with the National Research Council and the Transportation Research Board, and with a broad interest in the role of science and technical innovation (including but not just transportation) in regional development and social change.

In 1969 Garrison moved to become Professor of Civil and Environmental Engineering at the University of Pittsburgh, a period of continuing concern with inno vation and of the management of technology (e.g., 'Technology and urbanization' in 1971), and the early implementation of environmental impact analysis. Geography was not forgotten, however, and he contrib uted a significant essay 'Future geographies' in Chorley's 1973 collection *Directions in Geography*.

Garrison's final and longest tenure was in the Trans portation Program at the University of California, Berkeley, from 1973, through and past his retirement in 1991. Broad interests, exemplified by such publications as 'Future evolution of urban transportation systems' in 1976, 'Innovation for transportation' (in *Space and Time in Geography*, 1981), *Tomorrow's Transportation Changing Cities, Economies and Lives* (with Ward in 2000), and *The Trans portation Experience: Policy, Planning and Development* (with Levinson in 2005), balanced heavy involvement in crit ical urban transportation issues and decisions, and more technical publications like *Transportation Systems Man agement* in 1977, *Transportation* in 1984, and *Transportation Planning*, 1985. Although working in centers for trans portation studies in his later career, Garrison remained a loyal and dedicated geographer. An award to this effect was created in his honor by the Association of American Geographers (AAG): the biannual award for best disser tation in computational geography.

Garrison was a quiet revolutionary, who led a funda mental transformation of human geography during the period 1955–70. Although the movement was dubbed the 'quantitative revolution', as it indeed led to widespread adoption of statistical methods and of mathematical modeling in human geography, the greater significance was the emphasis on geography as a normal science, on the search for theory and explanation of human spatial behavior, and the evolution of the landscape. Within the broad field of human geography, Garrison's greatest substantive and policy contributions have been in the field of transportation, especially in metropolitan regions. In the present era, in which science has become devalued in geography and technology attacked, Garrison's career and writing remains inspirational and creative, a contri bution highly valued by his students and friends, and in the long run, the discipline as a whole.

See also: Berry, B.; Central Place Theory; Quantitative Revolution; Race.

Further Reading

Garrison, W. (1959). Spatial structure of the economy. *Annals of the Association of American Geographers* 49, 232 239; 471 482.

Garrison, W. (1960). Connectivity of the interstate highway system. *Proceedings of the Regional Science Association* 6, 122 137.

Garrison, W. (1960). Spatial Structure of the Economy: III, *Annals of the Association of American Geographers* 50, 357 373.

Garrison, W. (1963). Toward a simulation model of urban growth and development. *Lund Studies in Geography, Lund Haaken Ohlssons, Series B* 24, 91 108.

Garrison, W. (1973). Future geographies. In Chorley, R. J. (ed.) *Directions in Geography*, ch. 11, pp 237 249. London: Methuen & Co., Ltd.

Garrison, W. (1977). *Transportation System Management: Promise, Performance and Prognosis*. Berkeley, CA: Institute of Transportation Studies, University of California, Berkeley.

Garrison, W. (1979). Playing with ideas. *Annals of the Association of American Geographers* 69, 118 120.

Garrison, W. (1981). Innovation for transportation. In Woldenberg, M. J. (ed.) *Space and Time in Geography*, pp 215 236. Lund: Gleerup, C.W.K.

Garrison, W. (1984). Transportation technology. *Transportation Research A* 18, 267 276.

Garrison, W. (2000). Innovation and transportation's technologies. *Journal of Advanced Transportation* 34, 31 63.
Garrison, W. (2002). Lessons from the design of a life. In Gould, P. & Pitts, W. (eds.) *Geographical Voices*, pp 99 123. Syracuse, NY: Syracuse University Press.
Garrison, W. and Berry, B. J. L. (1958). A note on central place theory and the range of a good. *Economic Geography* 34, 304 311.
Garrison, W. and Berry, B. J. L. (1958). The functional bases of the central place hierarchy. *Economic Geography* 34, 145 154.
Garrison, W., Berry, B. J. L., Nystuen, J. and Morrill, R. (1959). *Studies of Highway Development and Geographic Change*. Seattle, WA: University of Washington Press.
Garrison, W. and Levinson, D. (2005). *The Transportation Experience, Policy, Planning, and Deployment*. New York: Oxford University Press.
Garrison, W. and Marble, D. (1967 1968). *Quantitative Geography, Studies in Geography, no. 13 and 14*. Evanston, IL: Northwestern University.
Garrison, W. and Marble, D. (1968). *The Structure of Transportation Networks*. Washington, DC: Office of Technical Services, U.S. Department of Commerce.
Garrison, W. and Souleyrette, R. (1994). The relationship between transportation and production. *Transportation Quarterly* 48, 257 265.
Garrison, W. and Ward, J. (2000). *Tomorrow's Transportation: Changing Cities, Economies, and Lives*. Norwood, MA: Artech House.
Spatial Structure of the Economy: II (1959). *Annals of the Association of American Geographers* 49, 471 482.

Relevant Websites

http://cutc.tamu.edu
CUTC; Awards; Award Recipients; Award Recipient Biographies.
www.aag.org
The biannual William Garrison Award for Best Dissertation in Computational Geography.
www.ce.berkeley.edu
UC Berkeley Department of Civil Environmental Engineering.
www.ce.berkeley.edu
William Garrison: research and teaching areas. Policy; New Technology; Planning; Freight.

Gated Communities/Privatopias

R. Atkinson, University of Tasmania, Hobart, TAS, Australia
S. Blandy, University of Leeds, Leeds, UK

Definition

Defining gated communities is difficult and contentious. There are two aspects of such residential neighborhoods which make them distinct from other examples of secured accommodation (such as high rise condominiums with concierge staff or individual houses with gated access). The first aspect is, of course, the physical constitution of such neighborhoods, but even then there may be differing degrees of 'gatedness'. In some gated communities we may find gates or booms across the road and in others it may be that gates block car access (or theft) while adjacent pedestrian thoroughfares are not controlled in the same way. The main differentiating physical feature of gated communities is that where access would ordinarily be expected in an 'open' neighborhood, it is restricted or available for control in a gated community.

The second key hallmark of gated communities is their legal constitution. As developers have built gated communities they have effectively created privately or ganized neighborhoods with their own infrastructure, including internal roads, common spaces, and services such as refuse collection. Unlike open neighborhoods, where roads are managed and repaired by the local state, gated communities are run by residents' management committees, known as homeowners' associations in the US. These governance organizations are often controlled by the developer in the first instance, and are then taken over by residents who pay an annual contribution for the maintenance of infrastructure and services provided to residents. These services may also extend to the em ployment of maintenance staff and security personnel and the use of closed circuit television (CCTV) systems. Many gated communities also have their own amenities such as gyms and swimming pools, or are built around golf courses or sailing lakes, which are also paid for and managed by the residents.

The Development of Gated Communities

It is important to remember that freedom of access has always been a negotiated aspect of urban life. In medi eval towns throughout Europe, as in Italian city states of around the eleventh century onward, the surrounding walls provided a fortified barrier to entry by outsiders. However, unlike today's gated communities, this barrier applied to the city as a whole, and access to streets and neighborhoods within the city was a right enjoyed by all citizens. Just as walled cities no longer operate to ex clude outsiders, true gated communities established in the past may be seen with hindsight to have been a temporary phenomenon. For example, in the eighteenth and nineteenth century, gated squares and private streets were common in London and other English cities. By 1900 most of the barriers had been removed in order to allow freedom of movement for an increasing volume of traffic.

The relatively more recent trend toward purpose built gated developments as the preserve of the affluent began in nineteenth century North America, and has accelerated over the last 50 years to the point where in some US states it is now almost impossible to purchase a new dwelling that is not located in an enclosed private neighborhood. It would be wrong to conclude that the parallel growth of gated communities worldwide has necessarily been influenced by developments in the USA. However, such developments are now common both in countries where this type of built form represents a relative novelty, as well as in regions, notably China and the Middle East, where inward looking, multi occupied residential developments have always been part of the architectural heritage.

Drivers for the Increase in Gated Communities

It is clear that in some instances the state, at both national and local levels, has played an important role in the de velopment of gated communities. For example, McKenzie has shown how the growth of gated communities in North America was orignally driven by developers and builders but was later significantly boosted by the state's en couragement of homeowner associations as the legal framework for collective ownership and management of property. These developments are in the interest of American municipalities, which derive fiscal advantages from affluent private neighborhoods combined with low demand for local state services from the residents. Even where government is concerned to regulate the manage ment of multi occupied developments, as in Taiwan, it also benefits from the devolution of neighborhood man agement to the homeowners associations as a devolved and private form of local governance. In the absence of gov ernmental or financial encouragement to create self managed gated communities, developers are nonetheless keen to build them as gating adds a cachet which translates into higher property values, and hence profitability. This has led to some concern that the growth of gated communities may be driven by supply side marketing, rather than genuine demand from potential purchasers though disentangling these issues remains difficult.

Gated communities are also increasingly common in countries undergoing economic and political transition, such as China and some Eastern European countries, where a desire for security has grown within contexts of economic and social change. However, enclosed and guarded housing developments have also been emerging in stable Western European countries with low crime rates, requiring an explanation other than, or in addition to, pure security. One view is that private neighborhoods with their legal frameworks which give residents control over private resources, reflect a growing disillusionment with the ability of government to provide the desired quality of services. Hence, drawing on club goods theory, gated communities have been characterized as private, localized economies which respond to desires for private aggregations of common interest to achieve the provision of distinctive goods, in this case secured neighborhoods not provided by local or central states. This explanation in turn has been criticized as inadequate to explain the historically and regionally specific development of gated communities. Gated communities appear to vary in their form and overall extent according to the national context they are found in, and there appear to be complex reasons for such variations.

Gated communities can be broadly differentiated between three types of development, each meeting different consumer needs. 'Prestige' gated communities enable their residents to enjoy an urban lifestyle while providing complete security. 'Lifestyle' developments incorporate exclusive leisure and other facilities, while the third type is the 'security zone', an inner city or inner suburban area retro fitted with walls, gates, and other security features, usually at the request of residents. Subsequently, analysis of the 2001 American Housing Survey has undermined the usual assumption that gated communities are the preserve of wealthy homeowners, by showing that low income, non white, renters are now more likely to live in gated developments than owner occupiers.

A national survey in England established that there is considerable diversity in terms of the built form of gated communities, the type of housing provided, and their location. So far as the motivation of purchasers is concerned, research has found that the issues of security, exclusivity, and prestige were often subsumed into a more general desire by purchasers to buy property which will maintain its value. Marketing materials, unlike in America, very rarely refer to the 'community' aspect of living in a gated community. In England, gated communities appear to appeal to predominantly young affluent singles and older couples. However, this varies according to particular locations and the relative availability of certain types of accommodation. For example, in London many new build and converted spaces that have been turned into gated communities consist largely of flats which are unlikely to be suitable for families. In both England, North America, and Australia there are growing markets for gated developments catering exclusively to retired people.

Further research is needed on the profile of gated community residents in other countries. It is apparent that in all global regions, however, gated communities are popular with expatriate communities and transnational elites, providing a secure and insulated residential base for their business dealings, as well as luxury leisure pursuits.

Crime and Other Effects

From residents' points of view, in high crime societies like Latin America, the USA, and South Africa, the recent increase in gated communities can be seen as a relatively rational response to a fear of violent disorder and personal harm. This has meant that stronger forms of fortification have been witnessed. In particular the gated communities of South Africa and cities like Sao Paolo have shown a marked tendency toward fortification in affluent areas where armed guards checking entrants and private policing of such spaces, outside the view or regulation of publicly monitored forces, has become common.

The small amount of evidence on whether gated communities prevent crime or not for such neighborhoods is contradictory. Where some have found that such 'target hardening' has reduced crime, others have found that crime rates are no different to other similarly high status neighborhoods. In South Africa, emerging evidence has suggested that gated communities have displaced crime to less affluent neighboring areas leading to the impression of a geography of risk connected with relative affluence. In short, this solidifying of patterns of segregation may reinforce positive effects for residents of gated communities while further disenfranchising residents outside the gates. Nevertheless, evidence in both the UK and California has suggested that many gated communities exist within surrounding areas which are also highly affluent, if not necessarily gated. Finally, gated communities can be seen to provide spaces of contained wealth which may be attractive to potential house breakers seeking neighborhoods rationally containing potentially greater rewards, though systematic evidence on these patterns has not yet emerged.

Life Behind the Walls

Gated communities appear to be growing in their prevalence by appealing to people with concerns about crime, as well as delivering prestige and privacy for those motivated by such issues in their residential decisions. However, the limited anthropological evidence available so far tends to show that, in fact, the residents of gated communities are highly susceptible to fear of crime directed at those outside the boundaries of gated com munities, as well as at service personnel who continue to connect these spaces of relative privilege with less well off communities outside, and who are therefore also viewed as a possible threat.

Since gated communities have not been a regular feature of urban life until perhaps the last 15 years, and then predominantly in Latin America, South Africa, and North America, the longer run implications of social life in these kinds of neighborhoods have not been fully thought through. However, some writers have begun to ask what will happen to children who grow up in the kind of predictable, racially homogenous, and privileged spaces of gated communities. These fears, of a with drawn, shy, fearful, and affluent class have recently been realized in cities like Moscow where the growth of a super rich social elite has led to the protection of their children in gated communities, as well as being trailed by personal bodyguards. Again, the implications of these lifestyles for the future views and social politics of af fluent classes brought up in protected neighborhood environments is unlikely to be positive, with possible impacts on a lack of empathy with people from different social backgrounds, as well as a fear of such difference.

Although internal relations between residents may be good, the fact that gated communities are self managed does not seem to inevitably lead to strong feelings of community cohesiveness. Everyday life within a gated community is largely regulated by legal documents, which set out the rules with which residents must comply, and the arrangements for self governance by the home owners. These non negotiable legal instruments under mine the concept of gated neighborhoods as voluntary communities able to develop their own informal controls and sanctions. Existing evidence on life within gated communities thus suggests a high degree of regulation that must be accepted in order to find distinctiveness and safety for the purchaser's household. Such developments may offer security or privacy but in a context in which, ironically, the freedoms of residents need to be sacrificed to achieve the apparent benefits of 'gated' living. North American examples of these problems abound, including restrictions on the maximum size of dogs and cases in volving residents prevented from flying flags outside their homes. Research in England found instances of disputes and tension, including families subjected to letters of complaint from the residents' management committee for leaving their childrens' toys on lawns. The self regulating aspect of gated communities may also mean that children are seen as less desirable to the extent that some of the legal documents governing these developments include a prohibition on children playing anywhere other than in the designated areas.

A further problem for the residents of gated com munities has been that many lack a full awareness of the ways in which private government may impinge on their daily lives. The necessary legal documents are signed in order to complete the purchase, but interview data sug gest that few purchasers read these thoroughly or ap preciate that their own conduct and use of their property will be constrained by the contents of these legal docu ments for years to come. Evidence is emerging, for ex ample, in North America, Taiwan, and England, that gated communities are not necessarily havens of like minded residents. Disputes do arise, between owners, between individual owners and homeowners associations or residents management organizations, and between owners and professional managing agents.

Further, as some analysts have begun to point out, the apparent benefits of these secessionary modes of private neighborhood governance have in certain instances been undermined by the increasing costs of private mainten ance. Individual residents can be faced with a large annual charge which if not paid might lead to their eviction. Examples where the residents are legally bound to pay their contribution include expenditure arising from un foreseen occurences such as minor earthquakes destroying internal roads, or more frequent instances of large but foreseeable items of expenditure like re roofing a swim ming pool, but where no sinking fund has been put aside. In some of these cases a return to public governance has been mooted by the dissatisfied homeowners, and more than one North American state has now passed legislation requiring vendors to warn purchasers of the legal impli cations of buying a property in a gated community.

Gated Communities in Literature

Gated communities have figured in significant literary examples. For J. G. Ballard, collectively privatized neighborhoods have formed a significant setting for many of his novels, particularly *Running Wild*, *Super Cannes*, and *Cocaine Nights*. In fact the idea of enclosure and rarified environments in which often affluent residents run amok outside the normative restraints of mass society has be come a recurring theme for Ballard. Much of this is drawn from and inspired by Ballard's own confinement during the Japanese invasion of Shanghai, which formed the basis for his semi autobiographical *Empire of the Sun*, also turned into a film. More recently T. C. Boyle has used the desires of an affluent community to gate their estate as a literary device in *The Tortilla Curtain*, in which an affluent young couple are confronted by their own fear and prejudices as they encounter desperate poor Mexican migrants on the US border. As fears of the Mexicans increase, fueled by the prejudices of a minority of resi dents, a surrounding fence and gates are finally installed against their instinctive rejection by the story's ambiva lent hero. Dystopian fictional treatments like these highlight the way in which gating can be seen as a zeit geist through which we understand a series of wider social and physical transformations affecting an increas ing range and number of cities.

Wider Implications

This privatization of what would otherwise be public spaces has driven wider debates about the relative influ ence of gated communities on social life in urban areas which have, particularly in the European context, been associated with the diversity and democracy of the street. This raises a broader question about the implications of forting up for the character of Western urban life. If we take away freedom of access to the street what does a city become? For example, some commentators have that this form of hyper segregation and fortification represents a new and critical moment which has transformed cities with earlier histories of open and democratic public spaces into a series of enclaves which protect affluent residents, while leaving an envious and poorer class of residents outside these protected bubbles. Poorer groups in these ghettoized spaces remain vulnerable to a variety of forms of street crime and violation in the home in economies which have seen freefall generated by austerity measures imposed by international financial institutions.

Gated communities are also notable for their growth in societies characterized by lower prevailing crime rates and higher levels of social cohesion. In this sense such 'communities' may be seen as social barometers indi cating much deeper undercurrents of social fear and aspiration toward ex territoriality as the signifier of membership to an affluent and secure class. In this con text the significance of gated communities lies less in their number and more in what they say about a wider bundle of social forces that are directing where and how people live. Nevertheless, the continued growth of gated communities suggests that they are an increasingly sig nificant proportion of dwellings, both responding to and perhaps also generating anxiety.

For sociolegal scholars, gated communities provide a clear illustration of both the power of property law and the ease with which it can be pressed into the service of developers and homeowners seeking seclusion, security, and prestige. The legal framework which creates and maintains these private neighborhoods places an invisible perimeter around the self managing residents, echoing the line of the physical boundary, both of which serve to demarcate the residents of the gated community and its exclusive services and amenities from outsiders who re main reliant on publicly provided facilities.

The Future

It seems likely that perceptions of rising crime will lead to more of this kind of secured development, in the absence of public planning interventions and along with rising real incomes for the middle classes of many de veloping countries – seeking prestige and the securing of their new found wealth. As transnational property markets have been opened up this has fuelled interest in an urban form capable of offering security 'in absentia' for owners, as well as insulated tourist bubbles for those visiting cheap but hostile zones. The impression of growing disorder in some contexts in the West and sig nificant and real increases in crime and vulnerability in parts of the majority world have undoubtedly raised the prospect of further gating and securitization in the future.

As gated communities have become common, the growth of gated cities in Southeast Asia has also been observed. In the resurgent tiger economies, gating has been used to create secure compounds for foreign workers. In these contexts it is less clear whether fear or a desire for social distinction is the driving force for such development but the longer run impacts in terms of a division between an emerging middle class and wider society appears to be increasingly running in tandem with processes of security which demonstrate a high degree of emerging social segregation. This appears to be a depressing outlook, given the impact of gated com munities on urban sustainability and the continued sep aration of affluent classes from the social life of cities touched by this kind of development.

See also: Community; Defensible Space; Governance, Urban; Law and Law Enforcement; Neighborhoods and Community; Private/Public Divide; Public Spaces, Urban; Segregation, Urban; Social Justice, Urban; Transnational Elites; Transport and Accessibility; Urban Regimes.

Further Reading

Atkinson, R. and Blandy, S. (eds.) (2006). *Gated Communities*. London: Routledge.
Atkinson, R., Blandy, S., Flint, J. and Lister, D. (2004). *Gated Communities in England*. London: Office of the Deputy Prime Minister.
Blakely, E. J. and Snyder, M. G. (1997). *Fortress America: Gated Communities in the United States*. Cambridge, MA; Washington

DC: Lincoln Institute of Land Policy; Brookings Institution Press.
Bohle, H. G. (2002). *Special Issue: Vulnerability Geographica Helvetica* 57(4).
Caldeira, T. (2000). *City of Walls: Crime, Segregation and Citizenship in Sao Paulo*. Berkeley, CA: University of California Press.
Geographica Helvetica (2003). Special issue on gated communities. *Geographica Helvetica* 58(4).
Gero, J. and Purcell, T. (eds.) (2002). *Special issue: Cognitive Modelling in Design*. *Environment and Planning B* 29(3).
Glasze, G., Webster, C. and Frantz, K. (eds.) (2006). *Private Cities: Global and Local Perspectives*. London: Routledge.
Housing Studies (2005). Special issue on gated communities. *Housing Studies* 20(2).
Low, S. (2003). *Behind the Gates: Life, Security and the Pursuit of Happiness in Fortress America*. London: Routledge.
Marcuse, P. (1997). Walls of fear and walls of support. In Ellin, N. (ed.) *Architecture of Fear*, pp 101 114. New York: Princeton Architectural Press.
McKenzie, E. (1994). *Privatopia: Homeowner Associations and the Rise of Residential Private Government*. New Haven, CT: Yale University Press.
Sanchez, T. W., Lang, R. E. and Dhavale, D. (2005). Security versus status? A first look at the census's gated community data. *Journal of Planning Education and Research* 24, 281 291.

Relevant Websites

http://www.cf.ac.uk
Cardiff School of City and Regional Planning, Cardiff University.
http://www.gatedcomsa.co.za
Gated Communities in South Africa.

Gay Geographies

V. Miller, University of Kent, Canterbury, UK

Glossary

Cottaging A term used in the Gay community to describe the act of picking up partners in a public lavatory for the purposes of sexual activity.

Cruising A term used to describe the act of deliberately seeking a sexual partner in a public place, often for anonymous sex.

Gay Ghetto A term sometimes used to describe an area, usually urban, in which a large number of gay people live and focus their commercial and leisure activities. Gay ghettoes are focused around a significant number of gay-oriented institutions, commercial establishments and leisure spaces, and normally include a surrounding residential population that is substantially gay.

Heteronormative The assumption that heterosexuality is the only 'natural' or 'normal' sexual orientation. Thus, implying that all other forms of sexuality (such as homosexuality) are abnormal or unnatural, and therefore wrong.

Homophobic Discrimination against, fear of, or aversion to, homosexuals or homosexuality in general.

Liminal Space Spaces characterized by openness, indeterminacy, or ambiguity.

Stonewall Riots A series of violent conflicts in New York City that lasted the week of 28 June 1969. The conflict was brought about by a police raid of the popular gay club, the Stonewall Inn. It is seen as a watershed in the worldwide gay rights movement.

In past centuries, extreme prejudice and repression meant that gay life was restricted to only the largest cities, and to clandestine networks of private and public spaces which were largely unknown to 'straight' society. However, the urbanization process of the nineteenth and twentieth centuries, and the social and legal reforms of the late 1960s onward, have allowed spaces which are much more visibly oriented to the gay community to develop. These spaces have now become an increasingly common feature within the inner areas of medium sized and larger cities of North America, Europe and Australasia, as well as the largest cities of South America and Asia.

Introduction: Invisible Cities – Gay Geographies in Historical Context

Historically, the two most defining characteristics of gay geographies are their urban locations and their 'invisible' or 'clandestine' qualities. Due to the subordinate or taboo status of homosexual activity in a largely hetero normative Western world (a status that was more often than not enforced through sometimes very harsh legal repression and punishment), gay geographies have ten ded to consist of a network of both private spaces and commercial establishments. Constellations of private homes, clubs, café's bars, hotels, as well as public spaces such as lavatories, parks, and certain streets that estab lished a local reputation (through personal links, rumor, and social networks) as 'cruising' or 'cottaging' areas, have tended to be the focal points for gay geographies through the centuries.

These areas tended to be more clandestine or semi clandestine in that their use by gay men was usually not widely known by the public, nor were they advertised in a public domain. Chauncy stated this situation eloquently:

> Gay men had to contend with the threat of vigilante anti gay violence as well as with the police. In response to this challenge, gay men devised a variety of tactics that allowed them to move freely about the city, and to construct a gay city in the midst of, yet invisible to, the dominant city. (Chauncey, 1994, p. 180)

The understanding of these places as 'gay' only by 'those in the know' was intentional due to legal re pression, and also because of the desire for anonymity among many gay men. Anonymity was an often necessary condition to escape not only legal punishment but also social ostracism for their sexual activity. Brown refers to these types of spaces as 'closet spaces', an appropriate metaphor to describe the schizophrenic spatial lives of gay men who are not (or not yet) visible or 'out' in public spaces.

As a result, a primary characteristic of gay geographies throughout history (and still largely true today) is that they were a combination of specific private spaces where gay men could feel comfortable expressing their sexu ality, and specific public spaces (such as cottages and cruising sites) where men could meet in anonymity.

Higgs' anthology *Queer Sites* has many examples of how gay geographies were constructed historically through these networks of 'invisible' public and private spaces in several major cities. For example, in that vol ume, Michael Siblais charts the shifting historical geog raphy of gay Paris from 1600 to the present. Illustrating this mix of closed and private spaces of café's, bars, and public spaces centered around the Tulieries Garden, the wooded parklands of the original Champs Elysees, the banks of the Seine, and, from 1830, the famous 'vespa siennes' or public urinals.

The second characteristic of gay geographies within a historical context is that they were (and still are) over whelmingly 'urban' geographies, and occurred primarily in large cities such as Paris, London, Amsterdam, New York, Sydney, or Toronto. There are several reasons for this. First of all is the simple issue of numbers. Relatively larger populations are needed to sustain gay oriented commercial establishments because not all gay men are public or 'out' enough to openly visit an establishment like a gay bar. Thus, proportionally more gay men are needed to support an establishment. For example, it has been suggested by American research that roughly 50 000 people are needed to support one gay bar, and by British research that 50 000 men over the age of 20 are needed. These figures are significantly higher than the population levels needed to support similar 'straight' establishments.

Second, heteronormative discourses in society, along with antigay oppression, meant that it was only in large cities where:

1. There existed enough anonymity among the popu lation, and enough diversity within the population to create an environment in which there is a threshold of cosmopolitan tolerance among the mainstream population.
2. There were enough liminal or marginal areas (such as 'red light' districts) for gay spaces to exist in either 'open' or 'clandestine' forms. In these liminal zones, there tends to be an increased tolerance for all manner of 'unconventional' or 'undesireable' behavior on the part of the general populace and, more importantly, law enforcement.

As a result, the more marginalized areas of larger cities such as docks, working class areas, and, in general, places which were seen as a locus for other marginal activities such as prostitution or vice areas, have tended to tradi tionally be the foci for gay geographies as well.

The period of urbanization associated with the In dustrial Revolution and the early modern period of course saw a significant rise in the numbers and sizes of large cities throughout the Western world. The creation of the great industrial cities in Europe and America, and the rapid expansion of settlement and urban development to the western half of North America all helped to create cities with large, mobile, and diverse populations. This gave many cities the kind of cosmopolitan backdrop which helped to sustain more pronounced gay spaces.

This can be seen in urban places that had acquired a link to alternative lifestyles, such as bohemianism and the cultural avant garde of the early twentieth century. Often these areas provided gay, or at least gay friendly, spaces within the urban nexus. For example, Chauncey charts the development of Greenwich Village, New York in the early 1900s into America's most famous Bohemian community, populated by 'unmarrieds, artists, free lovers, and anti materialists', making the Village a safe and even congenial place for gay men and lesbians to live. This then led to the development of a substantial gay enclave and a somewhat visible gay presence.

Theoretically, the Chicago School of Sociology was one of the first to systematically study these liminal urban zones and their relationship to gay space. For ex ample, Nels Anderson argued that the social and spatial configuration of the transient, male dominated 'hobohe mia' district of Chicago in the 1920s gave rise to a set of same sex sexual practices unique in the urban context. Overall, the Chicago school tended to link mobility with 'social disorganization' and the proliferation of non normative sexual practices and spaces. Thus, in the famous 'co centric zone' model of urban ecology by Er nest Burgess, the co centric zones themselves suggested an outward progression of increasing social and moral organization in which non normative sexualities, in cluding homosexuality, were confined to the inner zones and in particular the 'zone of transition'. These were the areas with the most mobile populations: bohemian en claves, racialized ghettoes, vice districts, and 'the world of furnished rooms'.

Although there were periods (such as the 1920s) and occasional spaces where homosexuality was more or less tolerated, gay geography still tended to adopt the 'in visible city' model given that the repression of gay men was still very much a concern. The marginal legal and social status of gay men meant that the toleration of open homosexuality in an area like Greenwich Village was always provisional and thus subject to moral panics and crackdowns on 'immoral' behavior.

It was not really until the late 1960s that this situation changed. This was largely due to the civil rights move ments of the 1960s, along with the gay liberation move ment, which were largely successful in helping to change the legal status of homosexuality in many countries, le galizing homosexuality in countries like the UK in 1967, and Canada in 1969. Several US states also saw changes in such discriminatory laws. In addition, the 'Stonewall riots' in New York in 1969 were a fundamental inspir ation in the gay liberation movement and signaled a more militant desire to attain equality legally and socially, and perhaps more importantly, it coalesced a desire for visi bility and openness. In this respect, it is really only after 'Stonewall' and the legal reforms of the late 1960s that 'gay ghettoes' as such can properly emerge.

The Idea of the 'Gay Ghetto'

One of the first systematic discussions of the idea of a 'gay ghetto' within an academic context was by Martin

Levine. Levine used the criteria set out by Robert Park and Louis Wirth of the Chicago School of Sociology, who used the term ‘ghetto’ to describe neighborhoods inhabited by people possessing a distinct culture and who are socially segregated from the mainstream community. Park and Wirth were primarily referring to ethnic groups in their discussion; however, Levine applied their criteria to the burgeoning gay areas which had gained some notoriety in the large American cities of Boston, New York, Chicago, San Francisco, and Los Angeles through the late 1960s and 1970s.

Park and Wirth had argued that ghettos had four key features:

1. Institutional concentration: that is, a concentration of social spaces, commercial and religious establishments.
2. Culture area: referred to the dominance of the minority culture within a particular urban area.
3. Social isolation: in which the ghettoized group in question is segregated from meaningful social relations with the wider mainstream community.
4. Residential concentration: which signified a significant spatial concentration of homes or residences of the minority group in question.

Levine applied these criteria to cities and the neighborhoods within them which, at the time, were reputed to have large populations of gay men: Boston, New York, Chicago, San Francisco, and Los Angeles. ‘Institutions’ (such as bars, clubs, bookstores, steam baths, restaurants, churches and movie houses catering to the gay community, as well as cruising areas) were mapped to demonstrate institutional concentration. ‘Culture area’ was determined though a qualitative analysis of the prevalence of ‘gay culture’ manifested through a domination of males in public places, billboards, posters, shop signs, and ambient conversation which utilized gay cultural traits and terminology. To this, Levine also added the ubiquity of ‘gay fashion’ in certain areas. ‘Social isolation’ was also determined qualitatively through interviews which demonstrated that many of the gay men in these areas only had sporadic contact with ‘heterosexual society’ and that their social lives revolved primarily around gay spaces. Lastly, ‘residential concentration’ was inferred through secondary media sources such as gay oriented magazines and anecdotal evidence, both of which implied a large concentration of gay males in certain areas.

Levine argued on this basis that there were three fully fledged ‘gay ghettoes’ in the United States: the West Village in New York City; The Castro Village in San Francisco; and ‘Boys Town’ (aka West Hollywood) in Los Angeles. In addition, there were a further 12 neighborhoods in the cities studied that could be considered on their way to achieving such ‘gay ghetto’ status.

While methodologically problematic, Levine’s work was fundamental in articulating an increased visibility and social prevalence that gay life was having in certain (particularly North American) cities. He argued that such spaces were born of two social processes: the continuing societal antipathy toward homosexuality, and the modicum of tolerance afforded to gay men under the more liberal social and legal environment of the late 1960s and 1970s as compared to the almost total suppression of homosexuality in the decades before.

In terms of methodology, Levine’s work also highlighted a fundamental problem with studies of sexuality and urban space: the measurement of residential concentration. Ultimately, secondary and official data sources such as the census do not provide data on sexual orientation. Thus, perhaps the most fundamental part of the ghettoization process, residential segregation, when applied to gay communities, must be inferred from other indicators such as single or dual male households, or collected though primary (almost always qualitative) data, which is often dubious from a quantitative standpoint. In short, gay ghettos are often hard to ‘prove’ using the traditional yardsticks of urban geography.

A short time after Levine, the spatial expression of the gay community of San Francisco was investigated by Manuel Castells. Similar to Levine, Castells charted and mapped the spatial dimensions of gay institutions and facilities but expanded on the relationships between space and power, arguing that an area like the Castro not only manifested increased visibility, but also served to aid in political struggle. In addition, Castells work also posited a link (elaborated below) between gentrification and the development of gay neighborhoods.

While bringing the discussion of gay space more firmly into mainstream discussions of urban space, Castells work was heavily criticised by subsequent researchers for two reasons. First was Castells rather essentialist conception of the difference between gay male and lesbian spatial structure, which he argues was the result of a ‘male territorial imperative’ versus a female tendency to concentrate on relationships and networks. Second was the criticism that Castells analysis was somewhat simplistic, neglecting larger structural aspects which meant that the spatial expressions of the gay community were not chosen as much as they were a result of the marginality of status gay men within society, and the marginality of the spaces they occupy within the urban fabric.

The idea of a ‘gay ghetto’ was seen as a peculiarly American (and then North American) phenomenon. In European major cities such as Paris and London, gay space still tended toward the ‘invisible city’ model of closed and private spaces, as well as clandestine public spaces, usually concentrated around several nodes in a given city. For example, Siblais argues that gay night life in 1960s and 1970s Paris revolved around the areas of the Rue St. Anne, as well as Monmartre and Saint Germain des Paris, with a number of parks and public conveniences thrown into the mix. Residentially, the gay population of Paris was still scattered throughout the metropolitan region, drifting into the above areas during leisure hours. However, Siblais demonstrated that by the mid 1990s, Le Marais had be come established as Paris' first gay ghetto, and had become the focal point of gay life in Paris both commercially and residentially.

The Functions of a 'Gay Ghetto'

Early studies of gay ghettoes and enclaves took the ap proach of viewing such spaces in a similar manner to spaces dominated by oppressed ethnic minorities. As such, there was a focus more on the development of gay ghettoes as partly (or even primarily) a response to the prejudice of heterosexual society. Typical of this is David Bell's use of Fredrick Boal's argument that ethnic resi dential clusters provided four basic functions:

- 'Defense', in terms of physical and mental security which is produced in areas of ethnic or racial homo geneity. The feeling that there is 'security in numbers', and the ability to organize defenses on such a basis.
- 'Avoidance' for social adaptation, serving as 'ports of entry' into the host society. Such enclaves provide low pressure areas in what can be considered a hostile and alien world.
- 'Preservation' of culture, way of life through insti tutions present (one could include here the notion of institutional completeness).
- 'Attack' in terms of political and social struggle through political lobbying and voting power, or through insurrectionary activity in the case of violent struggle.

Such functions were expressed in a number of studies, where, for example, many argued in terms of 'defense' that gay urban space becomes a locus of safety from verbal, physical, and even legal harassment of which gay men have historically been the victims. Security is pro vided simply through there being 'safety in numbers' as well as through the establishment of a zone of liminality in which more marginalized sexual behaviors and dis courses are tolerated. In this secure environment, gay men and lesbians can test the limits of their sexual rights and experience the comfort of expressing a controversial identity among those they feel are similar to themselves, or are at least open minded and sympathetic to their plight.

Under the rubric of 'avoidance' functions, gay ghettoes also act as 'ports of entry', in the sense that such spaces allow for 'closeted' gay men to 'come out' or introduce themselves overtly to a larger gay community from which that may have been isolated both physically and metaphorically. For example, it is not uncommon for gay media, particularly newspapers and magazines, to have a significant impact in easing isolation for those who are not yet 'out' or who live in areas where no obvious gay presence is felt. Gay oriented print media such as *The Advocate* in the United States or *The Pink Paper* in Great Britain contain descriptions of places and communities as a major component of their content. Often it is through such media where the reputations of 'gay ghettoes' are established and brought to the attention of people out side a local context. The end result is that places which acquire these reputations become akin to pilgrimage sites and migration destinations for closeted gay men who wish to lead a more open life. This has certainly been the case for San Francisco, West Hollywood, and the West Village in the United States, London and Manchester in Great Britain, Sydney, Australia, and Paris, France. Sev eral studies within the geographic literature have dem onstrated how gay enclaves in large urban centers become a focus for migration, while others have exam ined the role of media in the process of 'defining' certain spaces as 'gay'.

Gay enclaves also function to provide specialized commercial and social services. From the work of Levine onward, such institutional provision has been seen as the most fundamental characteristic of the gay ghetto. In particular, social spaces such as bars, clubs and café's, as well as bookstores which stock gay oriented media are seen crucial. In addition, more overtly sexual establish ments such as bathhouses, pornography outlets, and cruising areas are usually referenced. Of course, the larger and more well established gay spaces will often have a commercial core that houses the full range of services that any small community or neighborhood would provide – clothing retail, lawyers, taxi services, hairdressers, charity headquarters, real estate agents and so on – but will usually be gay owned and focus on a gay clientele.

Themes of 'preservation of culture' within studies of gay space are often related to the commercial and social infrastructure they provide. As a result of oppression, lesbians and gay men have traditionally adopted strat egies to pass as straight in the outside world. This can make it extremely difficult for them to meet friends, sexual partners, and lead an open life in general. Gay spaces provide, or consist of, social networks where gay men and lesbians can meet each other as gay, and lessen the isolation people can feel in an overtly heterosexual society. Thus, Gay ghettoes provide their users with the freedom not to act straight. In that way, these places have been said to be inextricably bound with gay identity and its expression.

This idea of expression is important, as a basic and fundamental issue for gay men and lesbians is visibility. As a result, one can speak of the role of such urban spaces in helping to 'create' culture and identity, as opposed to merely preserving it. As many have argued, the devel opment of gay ghettoes plays an integral part in gay culture building, demonstrating the expressive potential which space can hold.

Thus, while it can be argued that these enclaves do provide defense and attack functions in rare cases, gay enclaves certainly do not 'preserve' culture in the same sense as an ethnic enclave might be said to. Their efforts are more likely to 'create' a common gay culture or identity. This desire for territory from which a common identity can emerge can be seen in the more 'proactive' collections such as *Queers in Space: Communities, Spaces, Sites of Resistance* and *Mapping Desire* which more or less invite gay men and lesbians to 'take' space and culture.

Another function of gay enclaves often discussed in the literature is related to political struggle. Residential enclaves often serve as a basis for power for gay men in that by concentrating in certain areas, a 'gay voting block' can be established, thereby giving gay men a collective voice and a greater influence on local politics. The broader literature on 'gay urban studies' evidences many examples of how a marginalized group has been able to 'colonize' space and gain access to economic and political power. Knopp depicted how local gay entrepreneurs, press and political agents in Minnesota mobilized their political–economic power to establish social reforms at the local scale. More generally, Lauria and Knopp depict how gay men mobilized voting in residential concen trations in gentrified areas to obtain sympathetic ears on city councils in eight American cities.

The Political Economy of Gay Spaces in a Postindustrial Age

Studies such as those above, often inspired by Marxist thought and political economy, began to examine gay enclaves in the wider context of urban space in general, noting that urban space was not 'natural' but instead 'constructed' by the people and forces embedded in a capitalist and increasingly global society. In that sense, studies of gay urban spaces became less essentialist, and viewed gay men as actors who are influenced by capital processes, production, consumption, and symbolism.

Gentrification

The late 1970s and 1980s saw the rise of the so called 'urban renaissance' of which gentrification was a major process. Gentrification (briefly) referred to a new type of settlement pattern in which persons of high socio eco nomic status (professionals, management and adminis tration workers) migrated into what would traditionally have been lower income inner city neighborhoods and turn of the century industrial districts. Investments of capital would follow these 'urban pioneers' resulting in an overall upgrading of many inner city neighborhoods into high income enclaves which enjoyed the amenities of city center living.

Early on, gentrification had been linked with gay homeowners who were often among the first wave of these 'urban pioneers'. Even Castells' early work on San Francisco pointed to the upgrading and renovation of inner city housing stock by gay men connected to a relatively prosperous service economy and relatively few financial responsibilities in terms of family. Lauria and Knopp's groundbreaking work demonstrated this link in the greatest detail. Looking at the political economy of the gentrification process, they argued that gay men and lesbians were over represented in the main demographic characteristics of gentrifiers: people (and couples) with out children, and working in the burgeoning new ad ministrative, managerial, and service sector employment growth occurring in large city centers. This, combined with affordable, low cost, renovatable housing in de pressed city center neighborhoods, gave gay men the opportunity to develop a territorial and economic base within the urban structure.

In general, the argument that links gay men with the gentrification process revolves around three main points:

1. Gay men and couples tend to have high disposable incomes due to a tendency to be involved in (com paratively well paid) professional and service sector employment. This combines with a tendency to have fewer family responsibilities in terms of children. Thus, gay men are more likely to have the financial resources to renovate older properties to add value.
2. Gay men often prefer to reside in city center areas as such neighborhoods are often more socially diverse (and therefore more tolerant) than suburban areas. In addition, their proximity to the amenities and night life of downtown makes these areas more vibrant and therefore preferable.
3. The marginalized status of gay men gives them fewer options but marginalized areas in which to build a community.

However, it must be maintained that these points are quite simplistic generalizations. One must consider that, just as any community, the gay community is eco nomically and socially diverse and includes not only people of low income, but couples with children and those who prefer a more suburban as compared to urban lifestyle. In addition, a study of gentrification in Toronto's 'Cabbagetown' area demonstrated an interesting dis tinction between gentrified areas populated by a sub stantial number of gay men, and 'gay ghettoes', both of which, it was argued, were the results of substantially different lifestyle ambitions within their respective populations.

As we saw in the previous section, several studies have depicted how the gentrification process helped local gay entrepreneurs and gay men in general to gain access to local power structures. Perhaps the most impressive example of the relationship between gentrification and the empowerment of gay men on the local scale is of West Los Angeles, where gay entrepreneurial participation in local redevelopment schemes, along with gay and lesbian movement coalitions with other lobby groups of renters and senior citizens, led to the incorporation of West Hollywood as a city in its own right in 1984. This ultimately resulted in the election of the first West Hollywood city council which was made up by a majority of gay members. Following this, there was a rapid subsequent expansion of gay commerce in the new city.

Postindustrial Revitalization, Urban Renewal, and 'Gay Villages'

The link between gentrification and the development of gay ghettoes demonstrated how the establishment of gay spaces could be linked to processes of urban renewal. In recent years, the link between gay men and urban renewal more broadly has become explicit enough that *Governing* (a magazine for state and local government officials in the United States), ran a cover story in August 2003 entitled 'Chasing the rainbow: Is a gay population an engine of urban revival?'. Citing the work of Richard Florida among others, the article is premised on the argument that there is a link between the size and visibility of a local gay population, to the size of a 'creative class' in a given locality, and therefore to the potential economic well being of a locality in a postindustrial, globalized age.

Political–economic work on gay space in the 1990s has gone beyond accounts of gentrification to consider the place of gay men (and in particular, the spending of gays, or the 'pink pound' as it is often referred to), within a wider urban and global context of urban revitalisation, cosmopolitan branding, and tourism in a cultural economy.

In this respect, more recent studies have focused on the development of gay spaces under a less acrimonious atmosphere and within a context where local councils and entrepreneurial powers (both gay and straight) actually cooperate with their local gay communities in creating gay oriented urban spaces. This is done in order to attract inward investment, migration, and tourism from national and even international sources into struggling postindustrial cities.

For example, Jon Binnie examined the development of gay spaces in London's Old Compton Street and in central Amsterdam within the context of the early 1990s economic recession. He depicted both of these revitalization strategies as 'an alliance between queer politics and pink capital' whereby inner city businesses in London were keen to court gay custom in a competitive recession era environment. Similarly, in Amsterdam, local government support of gay rights and businesses was linked to the desire to market Amsterdam internationally as a gay capital to the lucrative gay tourism market.

The development of the Gay Village in Manchester, UK illustrated this process very well. Quilley argues that the creation of 'The Village' as the UK's only gay village by local authorities was heavily linked to the Central Manchester Development Corporation's desire to redevelop the inner city Canal Street district and promote the Village, and Manchester more generally, to international tourists under a banner of 'cosmopolitan diversity'. Plans for the development and marketing of similar districts are now being devised for other UK cities such as Newcastle and Birmingham, and similar efforts have been made in many cities around the globe.

This shift in attitude (embodied in the shift in terminology from the popular usage of the defensive and pejorative term 'gay ghetto' to the somewhat more positive 'gay village' phrase) in which gay visibility has come to be seen in many circles as a positive, revenue enhancing feature of a city is undoubtedly due to the increasing recognition of the economic value of the gay community as a whole. Thus, gay men have managed to use existing political and economic power structures to carve out spaces, rights, and visibility within a postindustrial context. Such an achievement, within an often unsympathetic society, is no doubt a positive accomplishment. However, as many have pointed out, such victories can be seen as tenuous, based on whether 'pink capital' is wanted or needed at a given point in time.

Issues

The previous sections have shown an overall historical shift within gay geographies from clandestine or underground networks of spaces, to the more visible and open 'gay ghettoes', and finally to more contrived and formally recognized 'gay villages'. These shifts have been largely due to the larger economic processes of industrialization, urbanization and postindustrialism, as well as sociopolitical changes, namely an increased societal and legal tolerance, and in some cases open acceptance, of homosexuality.

Clearly, the development of visible gay ghettoes and villages provides many benefits to the gay community in terms of expression, visibility, and power; however, these developments have not been without controversy. Some argue that on the one hand, ghettoization opens up new possibilities in terms of a collective strength to be visible and for gay men to express themselves in public, but on the other, ghettoization as a process continues to set gay men apart from their fellow citizens. Thus, the continued development of gay ghettoes runs the risk of helping to continue the social marginalization that created the con ditions for their development in the first place, making the marginalized status of gay men a permanent one.

Others within the gay community make the opposite point, and are critical of the increasingly visible, socially acceptable, consumption oriented gay culture that is being embodied within these new gentrified 'Village' spaces. Many argue that these consumption oriented spaces glorify a 'mainstream lifestyle' which is pro gressively eroding any sense of a distinct gay identity. Many also point to the exclusionary aspects of these new consumer oriented spaces which are dominated by young, clean cut, wealthy, white men at the expense of the poor, older men, ethnic minorities, and those gay men who are less 'mainstream' in general. In this respect, many feel as though the diversity, identity, and radicalism of previous decades have been lost in the name of con formity and consumerism.

While such discussions will no doubt continue, it is nevertheless clear that spaces oriented around aspects of gay culture and/or lifestyle continue to play an ever more significant part in the fabric of contemporary cities.

See also: Chicago School; Gentrification; Heteronormativity; Lesbian Geographies.

Further Reading

Bell, D. (1991). Insignificant others: Lesbian and gay geographies. *Area* 23, 323 329.

Bell, D. and Valentine, G. (eds.) (1995). *Mapping Desire: Geographies of Sexualities*. London: Routledge.

Binnie, J. (1995). Trading places, consumption, sexuality and the production of queer space. In Bell, D. & Valentine, G. (eds.) *Mapping Desire: Geographies of Sexualities*, pp 182 199. London: Routledge.

Bouthillette, A. (1994). Gentrification by gay male communities: A case study of Toronto's Cabbagetown. In Whittle, S. (ed.) *The Margins of the City: Gay Men's Urban Lives*, pp 65 83. Aldershot: Arena.

Brown, M. (2000). *Closet Space: Geographies of Metaphor from the Body to the Globe*. London: Routledge.

Castells, M. (1983). *The City and the Grassroots*. London: Edward Arnold.

Chauncey, G. (1994). *Gay New York: Gender, Urban Culture, and the Making of the Gay Male World, 1890 1940*. New York: Basic Books.

Heap, C. (2003). The city as a sexual laboratory: The queer heritage of the Chicago School. *Qualitative Sociology* 26, 457 487.

Higgs, D. (ed.) (1999). *Queer Sites: Gay Urban Histories since 1600*. London: Routledge.

Ingram, G., Bouthillette, A. and Retter, Y. (eds.) (1997). *Queers in Space: Communities, Public Places, Sites of Resistance*, p CA. San Francisco: Bay Press.

Kenney, M. (2001). *Mapping Gay LA: The Intersection of Place and Politics*. Philadelphia, PA: Temple University Press.

Knopp, L. (1987). Social theory, social movements and public policy: Resent accomplishments of the gay and lesbian movements in Minneapolis Minnesota. *International Journal of Urban and Regional Research* 11, 243 261.

Knopp, L. (1994). Social justice, sexuality, and the city. *Urban Geography* 15, 644 660.

Knopp, L. (1995). Sexuality and urban space: A framework for analysis. In Bell, D. & Valentine, G. (eds.) *Mapping Desire: Geographies of Sexualities*, pp 149 164. London: Routledge.

Lauria, M. and Knopp, L. (1985). Towards an analysis of the role of gay communities in the urban renaissance. *Urban Geography* 6, 152 169.

Levine, M. (1979). Gay ghetto. *Journal of Homosexuality* 4, 363 377.

Siblais, M. (1999). Paris. In Higgs, D. (ed.) *Queer Sites: Gay Urban Histories since 1600*, pp 10 37. London: Routledge.

Swope, C. (2003). Chasing the rainbow: Is a gay population an engine of urban revival? *Governing*. http://governing.com/textbook/gays.htm (accessed on February 2008).

Quilley, S. (1997). Constructing Manchester's 'new urban village': Gay space in the entrepreneurial city. In Brent Ingram, G., Bouthillette, A. & Retter, Y. (eds.) *Queers in Space: Communities, Spaces, Sites of Resistance*, pp 275 293. San Francisco, CA: Bay Press.

Whittle, S. (ed.) (1994). *The Margins of the City: Gay Men's Urban Lives*. Aldershot: Arena.

Gender and Health

J. Wiles, University of Auckland, Auckland, New Zealand

Introduction – General Trends

Geographers are increasingly interested in how interconnected physical and social dimensions of health play out in and through place. Recent geographical work highlights connections between place, aspects of identity and social status (such as gender), and health. Gender is generally used to refer to socially constructed differences, such as the set of qualities and behaviors expected from a male or female by a society, or one's subjective sense of self as masculine or feminine. Gender roles are thus socially determined; they are learned and vary widely within and between cultures, and over time. In contrast, sex is generally used to refer to biology and anatomy, and the objective categorization of a person's physiological status as male or female. Of course, such a simple distinction (social vs. biological) is arbitrary and problematic, even though it is useful for the purpose of clarification for this article. For example, gender can also be thought of as something that is 'practiced', or performed, within and against the constraints of social norms as well as physiological structures. Moreover, gender is always influenced by both social and physical aspects of identity, such as age, sexuality, region, and ethnicity.

Disease prevalence, and lay and official responses to disease, differ across geographic locations and for different social groups, and gender differences in health are a good example of this. Moreover, meanings attached to places and regions themselves also influence how specific health issues are understood or represented. A good example is Elizabeth Chacko's work on Pellagra, a disease associated with low income and poor nutrition whose victims were mainly black, low income women in the American south. Chacko argues that perceptions of the South as a 'regional' 'Other' were combined with reluctance to acknowledge widespread deprivations in the region by Southern leaders and inadequate scrutiny of the racial and gender dimensions of the disease. This meant that those most adversely affected by the disease were those who experienced the combined effects of a Southern regional identity, low socioeconomic status, an African American heritage, and female gender. This relationship worked at different scales to confuse the way officials and health professionals interpreted and therefore dealt with the disease. This is a good illustration of how place is not simply a container or backdrop to the interplay of gender and health, but actively involved as an important factor in and of itself, and why geographers have important contributions to make to better understanding of gendered health issues.

As Isabel Dyck and her collaborators have argued in a work on *Geographies of Women's Health*, health geographers recognize the interconnections between biological, physiological, social, cultural, economic, and political dimensions of health and illness. Far from being neutral scientific and medical phenomena, health and disease are understood within specific social, cultural, and spatial frameworks. As one among many overlapping frameworks, gender is not only associated with different risks and protections for specific illnesses, but also produces different behavior when ill. As the sociologist Judith Lorber has suggested, gender elicits different responses in health personnel, affects the social 'worth' of patients, and influences priorities of treatment, research, and financing. Gender not only has an impact on health, but gender relations are in turn impacted by health and wellbeing. The provision of both paid and unpaid healthcare is gendered, so that who provides what kind of care when and where, and the meaning and value that is given to that care, contributes to formal and informal economies of healthcare and social status.

Recognition of the socially and spatially contextual nature of health and healthcare moves our consideration as geographers firmly away from a biomedical focus. While a biomedical approach regards health as the opposite of illness and disease, a social approach to health understands health as a complex set of social constructs, recognizing wellness as something more than the simple absence of disease. This redefinition of health has broadened the topics covered by geographers interested in health and gender, necessitated new methods to explain and understand relationships between gender and health, and led to a more explicit engagement with critical theory.

These developments reflect the growing theorization of gender with respect to health, and a general progression from 'add women and stir' approaches of the 1970s and early 1980s to more complex analyses and interpretations of the meaning of gender, sex, and sexuality. Early studies of geographies of gender and health typically focused on women's health. Frequently research developed out of a need to challenge the traditional focus on men or 'gender neutral' studies which ignored either sex or gender differences. While some work focused on understanding trends in women's health in comparison and contrast to men's health (such as female and male patterns of morbidity or mortality), others focused on women's reproductive health (most often maternity and infant issues). Less frequently, geographical researchers have studied male specific health issues, although there is a growing body of work related to masculinities, health, and place.

Early attempts to explain health related differences between women and men, whether in health outcomes or access to and use of services, tended to focus on different gender roles between men and women. For example, certain occupations such as nursing and teaching are more typically done by women than men. This is partly because of social norms and expectations relating to gender roles, such as traditional stereotypes that women are 'naturally' more nurturing and better equipped for providing care than men. Women are also more likely to juggle multiple roles including paid and unpaid work and responsibilities for caring for children and older people. These role differences have implications for women's health, including the types of work that are available to them and the relative income they earn. For example, health outcomes may be related to specific occupational health risks or implications. Alternatively, the greater stress of juggling multiple roles or holding specific roles may be related to increased morbidity. Later explanations focused on gender relations and inequalities, querying gendered wage differences, or how gender relations in households shape distributions of labor, responsibility, and power. Recently, researchers have theorized gender as even more complex, interacting with factors such as ethnicity, age, stage in life course, wealth and class, and sexuality, among others. This has led to a destabilizing and de centering of gender as a category for under standing health differences, so that gender is not neces sarily the most important category of difference but one of many interacting aspects of individual and group identities and social structures. Gender is understood to be embedded in social systems and practices, whether family, political, medical, or legal systems, or cultures. The premise in such work is that the ways we think about and 'do' gender have a significant effect on the physical, social, emotional, and mental health, well being, and illness of individuals and social groups.

These more complex theorizations are reflected in the range of research in the area of gender, health, and geography. While in the past the primary focus of re search on gendered health was issues related to repro ductive health, this has now broadened to include other aspects of health and healthcare. Moreover, while the primary focus was traditionally on women and health, only more rarely was there a focus on men. There is now a growing interest in masculinities and health which is quite different to the research which assumes the 'clinical norm' is male. Geographers have moved beyond docu menting patterns of difference on the basis of sex to explore how social constructions and powerful norms of masculinity, difference, and power have (frequently negative) impacts on men's health. More rarely, re searchers are finally moving beyond the binary of male/female to explore different forms of gendered and sexualized identities as they relate to health and place, such as transgendered spatialities, or more multiple and fluid forms of sexuality and gendered identities.

The remainder of this entry will cover four main arguments:

1. there are gendered differences in health experiences and outcomes;
2. access to and use of services is gendered;
3. roles in provision of healthcare (as formal care pro viders or as informal care providers) are gendered; and
4. gender differences in health occur at a range of interrelated scales.

Health Outcomes

Health could be considered in terms of the spatial and temporal spheres in which individuals and groups interact on the basis of their gender. For example, Nancy Davis Lewis and Edith Kieffer, writing about the relationship between gender and health for women in developing countries, point out these might include biological re production, agricultural production and formal and in formal work sectors, social reproduction and unpaid work, and time allocation and different responsibilities. In less developed countries, women (especially poor women) face health burdens in reproductive health, care for young, literacy, access to clean water, fertility issues, and power issues. In situations of poverty, the work women typically do has negative implications for their health. Activities such as hauling water for long distances, or tending toxic or smoky fires based on coal, dung, or crop residue indoors using stoves that do not combust well or in places with poor ventilation, mean women are exposed to health risks on a prolonged and daily basis. Health risks include respiratory infection and disease such as tuber culosis and asthma, and cancer including lung cancer.

In wealthier countries, male mortality rates tend to be higher than female mortality rates at all ages. Yet female morbidity and use of medical care are usually reported as being higher than those of men. There are important characteristics to this data. Although men have typically tended to have higher rates of conditions such as is chemic heart disease, leading to their higher mortality rates, women have tended to have more acute short term illnesses and more chronic non fatal conditions such as arthritis. The latter are less likely to result in death but are associated with need for medical care. Recent statistics suggest that even some of these trends are showing convergence, for example as women show higher rates of heart disease, perhaps related to trends in smoking and types of stress.

These sex differences in health have been explained in a wide range of ways. They may be linked to biological risk or some intrinsic difference between men and women conferring differential risks of morbidity. They may be related to acquired risk, such as risk of injury associated with work or recreation, lifestyle and health habits such as smoking, exercise and nutrition, stress, and unemploy ment. A third explanation relates to illness behavior, or the psychosocial aspects of how people perceive symp toms, assess what to do about them and are able to do this. A fourth explanation is health reporting behavior, in cluding how people report their health to interviewers, while a fifth explanation focuses on prior healthcare re ceived and how this influences subsequent health and illness. Broader structural explanations focus on the interaction of aspects of difference and context. For ex ample, power relations between men and women are shaped by class, culture, age, and context. These messy relations shape gender roles and relations (e.g., role status, role satisfaction, and stresses), and gendered inequalities, for example in access to resources, or in responsibilities and opportunities. However while we consistently see gradients of health related to relative wealth, these do tend to be shallower for women than for men. This sug gests the greater reported morbidity rates of women relate mostly to acquired risks such as stress and their gender related roles, and much less to their attitudes toward health and illness and greater use of healthcare services (which may in fact protect them from greater morbidity).

One outcome of this is that women tend to live longer than men, and are more prone to disability and illness associated with older age. Age is associated with ill health through accumulated risks up to present (social) and gradual physiological decline (biological). Older women in general are a particularly vulnerable group, with high levels of poverty (much of which is kept 'hidden'). As populations in wealthy industrialized countries are aging, older people in general are healthier and have a higher quality of life. But we are also seeing greater incidences of diseases and illnesses associated with aging, over more sustained periods of time.

Access to and Use of Health Resources

Margaret Stacey and Virginia Olesen, introducing a special issue of *Social Science and Medicine* on women's health, submit that most healthcare takes place in a social domain located between the domestic sphere of family and the public realm of the marketplace and state. It is difficult to separate out these formal and informal, or public and private places, of healthcare. For this reason it is suggested that it is more useful to talk about gendered access to and use of health 'resources' rather than more specific health 'services'. Health resources encompass unpaid care as well as knowledge about health and access to healthy behaviors (such as nutritious food habits or healthy exercise), in addition to the range of formal and informal (e.g., alternative or traditional) services and materials provided by health professionals and paid providers. Health resources extend from state provided services (including housing and housing regulations, a justice system, and education in addition to more direct public health and health service provision) to the specific resources to which an individual has access (whether this is knowledge and information, social support, the time to practice healthy behaviors, or access health services, or material resources such as money, or a healthy body and mind). The possession of, access to, and use of health resources are inherently gendered and have specific geographies. Not only do men and women typically have differential access to these resources, but other social relations interacting with gender further shape access to and use of health resources.

Statistically, women use health services more often than men, and they are more likely to use a wider range of health services and resources. Middle aged, well educated, relatively affluent women in particular are high users of alternative, traditional, or complementary therapies, as are those suffering from chronic illnesses or terminal illnesses. There are a range of proposed reasons for this greater use of services. One is that, given their higher reported rates of morbidity, women are in greater need of health services. This may be because they are more vulnerable or sus ceptible to illness and disease (perhaps for physiological or biological reasons such as physical reproduction, but also perhaps as a result of inequalities, gendered hierarchies, gender roles and responsibilities, and power relations). The latter types of explanation suggest that in 'typical' roles as carers and nurturers women are likely to be supporting and facilitating others' use of services as well as their own, or may be more likely to report illness because they are more attuned to the efficacy of care and prevention. Other ill ness behavior explanations suggest that women's use of services is preventative, or at an earlier stage in the pro gression of illness, whereas men tend to present later but pose an even greater burden on the healthcare system. Others argue that it is more to do with the intersection of gender and socioeconomic status, so that it is actually persistent wage gaps, and higher poverty levels among women and particularly older women, that are responsible for the greater need for and hence use of services.

Provision of Healthcare Resources

The provision of healthcare is gendered. The majority of healthcare providers are women, even more so if we include informal, 'lay', and/or unpaid healthcare. Even within the formal healthcare system, there is a strong hierarchy to healthcare provision. The majority of formal (paid, professional, and trained) healthcare providers is women, yet women are disproportionately represented in the lower status and lower paid healthcare professions, while men are disproportionately represented in higher prestige positions. Even those women who are in professions such as medicine are more likely to be clustered in certain specialties such as obstetrics and gynecology, pediatrics, and general practice. The concept of the glass ceiling, where women can rise only so high in a career relative to their male colleagues, is well known in medicine, where expectations about other roles such as parenting are placed on women, just as supposedly essentially 'female' characteristics such as nurturing and caring are used as reasons for encouraging women into specific specialties and professions such as nursing. An equally gendered phenomenon, the 'glass elevator', has also been described whereby those few men who do enter predominantly 'female' occupations such as nursing move swiftly to positions of seniority and status. The historical evolution of gendered health provision and power is deeply bound up in these social constructions of gender, power, and identity.

Given these gendered differences in the provision and use of health services, the medical encounter itself is also structured by gender relations and power. Typically, the relationship between patients and providers is one where providers are viewed as having the most knowledge and therefore power, and as an asymmetrical relationship. This is challenged by suggestions that while health professionals have expert knowledge about disease and wellness, patients contribute expert knowledge about their own bodies and personal contexts. Some work has shown different communication styles between male and female health professionals, although this research walks a difficult line between 'essentializing' characteristics of men and women. For example, there is often little recognition of the performative nature of gender whereby individuals conform to (or challenge) gendered norms but also have these norms imposed on them. Assumptions about the gendered nature of health and of social relations filter into the arena of health provision. For example heart disease is frequently associated with men even though it is among the leading causes of death for women, and often goes undiagnosed among women. Assumptions about the work roles of men and women mean that men may be encouraged to go back to work whereas women are encouraged to stay at home after an illness.

By far the majority of healthcare takes place through informal care or on a lay basis through 'medical consultations' with family, friends, and acquaintances, through knowledge sources such as the internet, or through unpaid care. Health service use statistics reveal only 'the tip of the iceberg' of health problems, and many problems are informally treated by self, family or friends, or use of alternative medicines whether from trained providers or self administered supplements. With shifts in the geographical location and the qualitative nature of care (e.g., from institutions to 'community based' care, and from in hospital to day surgery and ambulatory care), the importance of 'lay', unpaid, or kin or friend based care is increasingly obvious. Research confirms that the majority of informal care providers are women, and that the experience of providing care on an informal basis is inherently gendered. Even those men who provide informal care report that they frequently experience discrimination on the basis of their gender, such as assumptions that they are somehow less competent carers. The reasons for women providing the majority of informal care are related to ways societies tend to think about gender, such as perceptions and stereotypes of women as inherently more nurturing, or that informal care reprises many of the tasks, skills, and competencies that are typically associated with women, who may therefore simply feel more competent or able to take on the role. At the household level, decisions about who provides care may be based on women typically earning lesser wages in a household or being more likely to have a flexible or part time job and therefore able to incorporate the role. At an individual level, many carers – both male and female – report that they derive considerable personal satisfaction and a source of identity and validation from this role, in addition to struggling with the psychosocial and physical problems associated with 'care giver burden'.

Health and Gender at Different Scales

The material discussed above describes gender and health issues at the scale of households, neighborhoods and communities, cities, and regions and nations. In this section, how gender differences in health occur at a range of interrelated scales from the body to the global scale is discussed.

The body, or the 'geography closest in', has become a key focus for health and gender. Drawing from the strengths of work on disabilities, researchers have explored the relevance of work on bodies for gendered health geographies. For example, while pregnant bodies are often treated as though pregnancy is a medical condition, feminist geographers have argued that there are problematic consequences in thinking of pregnancy in this biomedical frame, and explored the social and spatial aspects of pregnancy. Others, in exploring the social construction of health and illness, have critiqued assumptions that experiences of particular illnesses are universal, or that access to a healthcare service is universally the same for those at a certain distance away. Similarly, they contest the notion of such a thing as a 'neutral' dispassionate observer or healer. Instead they argue for understanding knowledge as embodied and material, through the corporeality of bodies and an increasing emphasis on experience. Recently, researchers have been attentive to the idea of embodied conscious ness, or embodied subjectivity. Rather than focusing on either the purely biomedical approach, or a purely social constructionst way of thinking, they argue for under standing the materiality of bodies as they are 'produced' in specific historical and geographic contexts. They see individuals as 'part of' larger assemblages, whether net works or relationships, and the body itself as a phe nomenon of physical constraints, options, and choices.

Violence must also be understood as an aspect of gendered health. Gendered and sexualized violence oc curs from the micro scale (such as interpersonal violence, domestic violence, and street crime) right through to the macro scale (such as ethnic aggression, seizure of land, and other resources such as water or oil, or war and na tional aggression), and for social, economic, or political reasons. Cynthia Cockburn, discussing gender and vio lence in a paper for the World Bank, argues that the gendered effects of violence are produced in three interrelated ways: through the bodily specificity of males and females, through the characteristic roles in society of males and females, and through gender ideologies. In armed conflict, for example, while the highest rate of direct death is among young men, women and children typically bear high rates of indirect death. Indirect death as a result of conflict is caused by damage to health sys tems, whether because health facilities and workers are targets, or through the effect of sanctions on economically vulnerable groups, and especially women and children. Other impacts are epidemics and the spread of disease from one place to another, which is exacerbated by mal nutrition, poor sanitation, and inadequate water supplies. Environmental harm through exposure to chemical war fare and substances, whether nuclear testing which has an effect on reproductive health, or exposure to landmines, may affect women and children more through their ac tivities and the spaces they occupy. More specific gender issues include sexual violence, with rape, torture, muti lation, and sexual exploitation; prostitution and slavery are common occurrences in conflict areas. In addition to the mental health issues involved in gendered responses to war, trauma and displacement is the inordinate care burden placed on women by the ravages of war.

Finally, it is important to consider gendered health issues at the global scale. The benefits of globalization, for example, have been unequally distributed across the globe and between different groups. The feminization of poverty is one such example of the unequal distribution of wealth, whereby increasing proportions of the world's poor are women and female headed households. This is due, among other things, to persistent wage gaps and dis crimination and stereotyping in the labor force, to en forced gender stereotypes and distributions of power in general, and to conflict and violence at the individual right through to international level. Another example is the global shift in the spaces and nature of production. This has gendered implications such as the migration of job opportunities from the central city to the suburbs, from more industrialized to lesser developed nations, or from more industrialized to more service oriented economies. The resulting change in employment patterns has severe impacts on health – for those men who lose jobs (and their families and communities), and the increasing numbers of women who are employed in 'flexible' positions with de teriorating work conditions and unsafe environments with little job security, whether in 'maquilladora' style factory sweatshops or in various precarious forms of employment. A further example of international trends affecting gender and health is the internationalization of neoliberal market based policies for social services and free trade in general. The privatization of formerly state owned services and decreased public spending on social services dis proportionately affects women, both as employees and as those in need of services.

Summary

The relationship between gender and health is complex, and is mediated by and shapes many social relations (age, sexuality, and socioeconomic status) and various socio spatial contexts (environmental, political, social, and economic). While we can observe a number of gendered patterns of health and health outcomes across space, there are a wide variety of ways to conceptualize and explain these, drawing on different ways of understand ing gender itself, in addition to increasingly complex approaches to health.

See also: Care/Caregiving; Gender, Historical Geographies of; Health Geography; Masculinities; Medical Geography; Nature, Social; Pregnancy and Childbirth; Welfare Reform.

Further Reading

Chacko, E. (2005). Understanding the geography of pellagra in the United States: The role of social and place based identities. *Gender, Place and Culture* 12(2), 197 212.

Cockburn, C. (1999). Gender, armed conflict and political violence. The World Bank. http://www.genderandpeacekeeping.org/resources/3 gender armed conflict and political violence.pdf (accessed in April 2008).

Doyal, L. (1995). *What Makes Women Sick*. London: MacMillan.

Dyck, I. (2003). Feminism and health geography: Twin tracks or divergent agendas? *Gender, Place and Culture* 10(4), 361 368.

Dyck, I., Lewis, N. D. and McLafferty, S. (eds.) (2001). *Routledge International Studies of Women and Place: Geographies of Women's Health*. London and New York: Routledge.

Earth, B. and Sthapit, S. (2002). Uterine prolapse in rural Nepal: Gender and human rights implications: A mandate for development. *Culture, Health and Sexuality* 4(3), 281 296.

Geoforum (1995). Special Issue: Geographies of Women's Health. *Geoforum* 26(3), 239 323.
Laws, G. (1995). Understanding ageism: Lessons from feminism and postmodernism. *The Gerontologist* 35, 112 118.
Lewis, N. D. (1998). Intellectual intersections: Gender and health in the pacific. *Social Science and Medicine* 46(6), 641 659.
Lewis, N. D. and Kieffer, E. (1994). The health of women: Beyond maternal and child health. In Philips, D. & Verhasselt, Y. (eds.) *Health and Development*, pp 122 137. London: Routledge.
Lorber, J. (1997). *Gender and the Social Construction of Illness*. Thousand Oaks, CA: Sage.
Macintyre, S. (1993). Gender differences in longevity and health in Eastern and Western Europe. In Platt, S., Thomas, H., Scott, S. & Williams, G. (eds.) *Locating Health: Sociological and Historical Explorations*, pp 57 73. Aldershot: Avebury.
Pearson, M. (1989). Medical geography: Genderless and colourblind. *Contemporary Issues in Geography and Education* 3, 9 17.
Stacey, M. and Olesen, V. (1993). Introduction: Special Issue on Women's Health. *Social Science and Medicine* 36(1), 1 5.
Verbrugge, L. M. (1989). The twain meet: Empirical explanations of sex differences in health and mortality. *Journal of Health and Social Behavior* 30, 282 304.

Gender and Rurality

J. Little, University of Exeter, Exeter, UK

Glossary

Sub-discipline Geography as a discipline has long been broken down into recognised sub-disciplines reflecting particular areas of research and writing. Such sub-disciplines are fluid and tend to have widely varying rates of instiutional acceptance as well as theoretical and intellectual development.

Introduction

Discussions of the development of rural geography in the 1970s and 1980s have frequently noted the seeming reluctance within the subdiscipline to embrace new theoretical perspectives. While human geography, in general, was discovering a new social relevance and political awareness, exploring inequality and power relations through the application of Marxist and political economic perspectives, rural geography remained largely descriptive, slow to apply and develop the kinds of 'cutting edge' theoretical debates characterizing much of the rest of human geography. The study of gender was no exception and the examination of difference and social inequality based on gender only 'arrived' in rural studies once it was well established in human geography. Similarly, key debates surrounding the relevance of feminist theory to the study of human geography were firmly rooted in urban studies and only rehearsed in a rural context some years after being adopted by urban geographers.

In other disciplines, notably rural sociology, two areas of research had identified the relevance of gender to the study of rural social and economic relations. The 'community studies' common in the early and mid twentieth century had highlighted the different family and community roles of men and women in the ethnographies of daily life in the English village. Similarly, studies of agriculture had recorded the role played by women in petty commodity production and the survival of the family farm. In neither area, however, was gender the central focus of research and any acknowledgement of gender difference was simply in recognizing the particular characteristics of the lives of women and men as they contributed to the operation and organization of the rural society.

Both areas, the organization of domestic and community life and the operation of the family farm, were, however, to prove important foci of gender studies as they emerged in rural geography. As rural geography began to pay more attention to issues of welfare, inequality, and resource allocation in the 1980s and to begin to place such work in theoretical frameworks of power, studies started to build on these early descriptions of men and women's roles in the development of specific gendered and feminist perspectives. The following two subsections will outline the development of work on gender in the two key areas of farming and community. Such a separation of work is, in part, reflective of a division in rural geography from the mid 1980s between agricultural and nonagricultural research. These were not, however, discrete areas of intellectual endeavor, at least not in the area of gender research, and the exchange of ideas between those working in agriculture and those working on the rural community was vital to the development of understanding of gender relations and identities and to the application of feminist perspectives. The material discussed in these sections is drawn mainly from the UK and the US but with a recognition of the more recent and growing contribution of studies from Australia and New Zealand and parts of Europe, notably Scandinavia.

Gender and Agriculture

In common with other areas of geography, the introduction of gender issues in the study of agricultural geography began by documenting the roles of men and women and drawing attention to the neglect of women's lives by academic perspectives that tended to subsume their activities into those of men. This documentation, while essentially descriptive, implicitly sought to challenge the assumption of men's work and men's lives as 'the norm' against which women were measured, judged, or simply ignored. The fundamental inequality of such an approach was further highlighted in the case of agriculture and the family farm, in particular, by an assertion that in neglecting or downplaying gender difference, the essential role of women in the economic profitability and viability of the agricultural enterprise was, in fact, being ignored. A focus on women in agriculture was needed, it was argued, both to understand how women lived their lives and how their activities and responsibilities differed from men's, but also to fully comprehend the economics of the farm business and the ways in which the pressures and demands of agricultural production on the family run farm were managed.

The initial recognition of women's 'invisibility' within the agricultural production process gave rise to a number of studies in the UK, the US, and Europe which sought to identify the nature of women's contribution to agricultural work and to calculate its economic centrality to the farm business. Women, it was argued, occupied widely differing positions regarding the amount and type of labor they performed but, regardless of what they actually did, their labor was universally treated as less important than that of the man. Just as importantly, women were found to undertake a whole range of tasks that, although critical to the running of the farm (tasks such as secretarial work, feeding the farm workers, caring for small animals, and providing emergency labor), were not recognized within the wider definition or evaluation of agricultural 'work' and were certainly not regarded as important to the business itself.

As was the pattern in geography more generally, this initial concern to document the lives of farm women started to give way to the development of theoretical frameworks to explain the gender difference. In particular, studies questioned classic Marxist and political economic interpretations of the labor process and argued that the particular position of women in agriculture needed to be conceptualized more broadly in terms of gender relations as well as (or instead of) labor relations. Researchers argued for the adoption of theoretical frameworks that focused on the operation of patriarchal gender relations in order to explain the lives of farm women and to draw attention to the largely unvalorized role of the reproductive sphere in the organization and operation of the family farm. What was required, it was argued, was some understanding of the gendered power relations that were central to all aspects of the economic and social reproduction of the family farm. Rather than simply examining the work of farm women, studies turned to look at the relational nature of men and women's roles, showing how gendered power relations were woven through not only the practical organization of agricultural tasks but also the underlying recognition of the status and responsibilities of men and women on the farm.

During the 1990s, the emphasis of feminist geography shifted away from 'grand theories' of explanation toward more small scale, localized, and fluid notions of gender difference. A questioning of what were seen as the priorities of white, middle class, Western feminism in the academy encouraged a recognition of the multiple and varying lives of women and men and steered academic work toward the study of gender identities and the different ways in which the roles of men and women were constructed. In rural geography, studies of gender in agriculture in Western economies recognized the construction of a dominant 'farming woman' identity (through, e.g., showing how women are associated with particular parts of the business and women's role in farm diversification). They also explored ways in which some women's identities differed from the dominant form, showing how women carve out alternative identities and challenge expectations about their role within and contribution to the farm business. While new work on farming and gender identities has encouraged important work on difference and identity construction, some researchers have argued for the continued relevance of patriarchy in understanding common experiences of men and women involved in agriculture and articulating broad structures and patterns of power within farming.

Gender and the Rural Community

As noted above, work on gender and rurality also developed in another direction in a separate series of studies on the rural community and nonagricultural household. A similar pattern unfolded in which initial work sought to draw attention to the particular issues faced by rural women – specifically in relation to access to resources and opportunities. It was argued that rural women were doubly disadvantaged in a rural context; as rural residents, they suffered the problems of remoteness and poor services but these problems were exacerbated by gender and the fact that, as women, they were responsible for the domestic reproduction of the household. Studies of employment claimed that rural women's disadvantage was particularly evident in the labor market where a lack of available jobs combined with an absence of childcare and other services further limited opportunities, particularly in full time and professional employment.

Following this articulation of gender difference and of the issues and problems faced by rural women, work moved on to attempt to explain gender roles through the application of feminist theoretical perspectives. Such work focused on gender relations and sought to place rural women's roles within a broader understanding of power and, specifically, the operation of patriarchy. In so doing it drew widely on feminist geography, asserting that it was not the rural location that led to women's inequality (e.g., in relation to job opportunities or access to services) but rather the power relations that existed between men and women and, in particular, the assumption that women were chiefly responsible for the domestic sphere and for associated issues of childcare. This work also argued, however, that while it was gender, and not rurality, that was responsible for women's inequality, rural society imposed a particular set of assumptions on the operation of gender relations that ensured the endurance of conservative attitudes to the roles of men and women. Families, households, and wider rural society were organized along traditional lines that saw women as the 'lynch pin' of the rural community and the center of the domestic realm and, as a result, any involvement in paid labor by women was seen as secondary to their main role.

Through the 1990s, as questions of gender identity became more central to feminist studies (as described above), rural geographers shifted away from the examination of the domestic division of labor and the operation of the community in the development of work on the constructions of rural femininity and masculinity. Taking up ideas on the operation of a particular form of rural patriarchy, studies argued that it was possible to identify specific rural constructions of gender identity. Initially, such work concentrated on the examination of what several authors termed the 'cultures of rural womanhood', showing how constructions of rural femininity articulated certain values and expectations about the rural community and that such constructions created/reinforced a hegemonic form of rural femininity that downplayed difference and resistance. They described how highly conservative assumptions about the domestic work of women and the importance of the family dominated, particularly among older women, and how ideologies of rural motherhood were central to the gender identities of rural women at the individual and household level. Such studies linked the construction of traditional femininity with dominant cultures of rurality, arguing that the idea of a rural idyll served to keep in place and valorize moral assumptions surrounding rural gender identities and the family.

Gender and Rural Sexuality

With the development of work on rural gender identities and 'otherness' came an explicit focus on sexuality. Again, rural geographers tended initially to lag behind urban geographers in recognizing the relevance of space to the construction and performance of sexuality. Such tardiness was only in part due to a lack of engagement with the wider debates around sexuality and space, also reflecting, it may be argued, the construction of the rural as a less consciously sexualized space. While urban geographers explored the increasing visibility of gay spaces and examined the emergence of new kinds of relationships between sexuality and space, rural geographers found little comparable challenge to the traditional sexual identities in rural spaces. Thus it was only when notions of otherness and marginalization began to be taken seriously, that rural geography began to take note of hidden identities, among them gay and lesbian identities.

This absence of visible 'gay spaces' in rural areas meant that studies of lesbians and gays were characterized initially by an absence. While some rare exceptions identified gay rural communities, particularly in the US, the majority of studies examined problems faced by those identifying as gay or lesbian in finding networks and establishing friendship groups. They concluded that the lack of a visible presence reinforced and was reinforced by an absence of services, activities, and support for gays and lesbians, exacerbating a sense of marginalization and a belief that homosexuality had no place in the rural community. Kramer's study of gay men living in the Midwest of the US, for example, revealed the difficulties they faced in small conservative rural towns in finding places to meet potential partners. The research went on to suggest that the absence of an identifiable gay space forced many men away from rural areas to live in more tolerant and diverse societies.

What emerged from these studies of social exclusion of rural gays and lesbians was the importance of placing the examination of homosexuality within a broader understanding of rural sexuality. Research on the experiences of those living (or trying to live) gay and lesbian lives in rural communities needed to be conceptualized in a recognition of the dominance of a particularly conservative and family based heterosexuality in which there existed a lack of tolerance toward other, 'deviant' sexualities. Thus studies moved from a rather narrow focus on the identification and examination of the lives lived by rural homosexual men and women to an exploration of constructions of heterosexuality and its reproduction in all areas of rural society.

Rural society, it was argued, is dominated by a belief that the natural sexuality of rural communities is heterosexuality. Many of the practical and ideological characteristics of the rural community are based on an assumption of the implicit superiority of the nuclear family as the dominant form of social organization. Such arguments showed how central to the rural idyll is the interrelationship of family and community. Just as previous research had highlighted the importance of traditional gender relations to the operation and survival of the rural community (reliant as it was on women's unpaid labor), so more recent work showed how heterosexuality and community are mutually dependent and reinforcing.

Assumptions about the dominance of the family in rural society help to create a moral geography in which any variation from hegemonic forms of heterosexuality is seen as deviant. Thus, it is not only homosexuality that is seen as inappropriate in a rural context but other forms of non family based sexuality. The rural is constructed as a space of morally superior and benign sexual relations, in contrast to the deviance of the urban. Acknowledgement of the rural as a place of active sexual relations tends to stress the cheery 'love among the haystacks' image of jocularity and inexperience. While the countryside may be a place of sexual experimentation and relative freedom among the young, the sexual experiences that take place are seen as largely innocent and unthreatening and far removed from more seedy and potentially dangerous urban encounters.

Research has shown how this very powerful construction of the rural as heterosexual and unthreatening is reflected in dominant versions of rural masculinity and femininity. Much of this work has focused on masculinity and on the links that can be drawn between the rural landscape and environment and a certain form of traditional and stereotypical masculinity. Thus work on agriculture, forestry, and on military training has shown how ‘real men’ are constructed, in a rural context, as men who are fit and tough and can cope with the physical challenges of the outdoors (whether in relation to making a living as a farmer or surviving as a soldier). As Woodward points out, rural nature is seen as a ‘test’ for rural masculinity and control over nature (through survival) evidence of the ability to perform as a man. She describes how, among soldier recruits, to be beaten by nature is to fail as a man and be accused of being soft or a fairy. A real man in this rural context, then, is definitely a heterosexual man.

Fewer studies have looked directly at rural femininities, yet there is frequently an assumption in work on rural sexuality, in general, of an equally strong set of rural ‘qualities’ surrounding the construction of the rural woman. Work by Little, in 2003 and 2007, for example, explores the expectations articulated by single farmers in their attempts to find a partner/wife. Strong assumptions about the ‘right’ sort of woman are clearly articulated by the farmers – thus, true country women possess a knowledge of nature and rurality that make them acceptable partners. But, interestingly, this understanding of the countryside is linked to other qualities in the rural woman – a stability, homeliness, and practicality. Rural women are low maintenance and sensible and often, it seems, have a style and appearance that reflects these qualities.

Other work on rural femininities has explored how the idea of ‘acceptable’ rural femininities becomes established within the identities of young girls growing up in the countryside. Such work includes oral histories of women growing up in the rural UK in the 1960s, and shows how rurality imposed particular constraints on the performance of femininity and how young girls resisted and negotiated those constraints at times, particularly when they moved beyond the spaces of the village. Other research has looked at the relationship between femininity, space, and rurality in the identities of rural teenagers and shown how strong moral assumptions surround the association between place and sexuality with a belief among both young men and women of the moral superiority of the countryside in the context of sexual identities.

The various studies of rural sexuality have demonstrated the relational nature of constructions of masculinity and femininity. The rural is shown not only as both the site for the performance of sexuality but also as a factor in its construction. Certain rural structures and institutions, such as the family, church, and community, are seen to reinforce and anchor heterosexuality, offering little option of deviation from highly conservative norms and expectations. The ideas that support these more concrete structures relate, it is argued, to a wider rural ideology that underpins an idyllic construction of rurality. Thus, heterosexuality is firmly embedded in rurality as an idea and a way of life.

The final short section of this discussion of gender and rurality concerns recent developments in the study of embodiment as they have translated to rural geography. Such developments have been partly inspired by a greater emphasis on performativity within cultural geography and, in particular, within gender studies on the related notion of gender as reiterative and constantly becoming. The body, as part of the performance of gender, has started to occupy a prominent position in the work of geographers whose attention has turned, unsurprisingly, to the role of space in shaping and mediating the embodied performance of gender.

Gender and the Rural Body

As noted, some work has made reference to the rural body in drawing attention to the stereotypical embodiment of masculinity and femininity associated with dominant constructions of rural sexuality. In addition, other authors have begun to look at the body as part of the growing importance attached to the understanding of the role of emotions in constructions of place. For example, work on adventure tourism agues that new forms of tourism (such as bungee jumping and whale watching) provide a different form of experience in which the tourist is allowed/required to be much closer to nature. The emotional engagement with nature is enhanced, it is suggested, when the tourist becomes, bodily, part of the nature they are experiencing. While arguably not directly gendered, the ability to engage in some of the ‘opportunities’ of adventure tourism is strongly inscribed with assumptions about the gendering of bodies and of the competences required for more extreme activities. Work on surfing and snowboarding has also extended these debates about the relationship between gender, the body, and competence within outdoor, rurally based activities.

The body is brought back into the mainstream discussion of rural social relations and the cultural construction of rurality in an examination of the representations of the rural body. It is argued that such representations have served to reinforce the rural community as a place of stable and unthreatening gender relations. The unthreatening portrayal of the rural body in the (now flourishing genre) of naked calendars is further evidence, it is suggested, of the uncontested and benign nature of rural heterosexuality. Representations of the rural body in the calendars tend to be jocular, often older and rounder than the classic images seen in the traditional calendars. Poses are tasteful, demure, or saucy, reinforcing the ‘love among the haystacks’ image discussed earlier and there is no space for more raunchy or seedy depictions of the body.

This article started with an assertion that rural geographers tended to be slow to respond to new developments in the study of gender and place. There was a reluctance, in particular, to strive for new theoretical directions and the emphasis of early rural work on gender seemed always to be on applying ideas initiated in an urban context. While distinctive rural dimensions were noted to, for ex ample, the organization of gender roles and the operation of patriarchy, the rural itself did not seem to offer par ticularly fertile ground for more original discussions of gender difference. Recently, however, there has been more of an interest in developing distinct and innovative dir ections in rural gender studies. Work on nature and gen der, for example, has brought novel perspectives to issues of sexuality and the body, while debates around the pro duction and consumption of local food have also seen the development of new avenues of work on gender in relation to health, nutrition, and eating practices. Yet, there remain many underdeveloped areas of rural gender research and the rural offers considerable potential for new agenda setting work in the field of gender studies.

See also: Agriculture, Sustainable.

Further Reading

Bell, D. and Valentine, G. (1995). Queer country: Rural gay and lesbian lives. *Journal of Rural Studies* 11, 113 122.

Brandth, B. (1995). Rural masculinity in transition: Gender images in tractor advertisements. *Journal of Rural Studies* 11, 123 133.

Campbell, H., Bell, M. M. and Finney, M. (eds.) *Country Boys: Masculinity and Rural Life*. University Park, PA: Penn State Press.

Elswood Hollard, K. (2005). 'We used to scrump the apples, we used to have our knickers full of 'em'. Growing up in the countryside: Forging femininities in rural Somerset *c*: 1950 1970. In Little, J. & Morris, C. (eds.) *Critical Studies in Rural Gender Issues*, pp 123 140. Aldershot: Ashgate.

Fellows, W. (1996). *Farm Boys: Lives of Gay Men from the Rural Mid West*. Madison, WI: University of Wisconsin Press.

Hughes, A. (1997). Rurality and "cultures of womanhood": Domestic identities and the moral order in village life. In Cloke, P. & Little, J. (eds.) *Contested Countryside Cultures: Otherness, Marginalization and Rurality*, pp 123 137. London: Routledge.

Liepins, R. (2000). Making men: The construction and representation of agriculture based masculinities in Australia and New Zealand. *Rural Sociology* 65, 605 620.

Little, J. (2002). *Gender and Rural Geography: Identity Sexuality and Power in the Countryside*. Harlow: Pearson Education.

Little, J. (2003). 'Riding the rural love train': Heterosexuality and the rural community. *Sociologia Ruralis* 43(4), 401 417.

Little, J. and Austin, P. (1996). Women and the rural idyll. *Journal of Rural Studies* 12, 101 111.

Little, J. and Leyshon, M. (2003). Embodied rural geographies: Developing a research agenda. *Progress in Human Geography* 27(3), 257 272.

Sachs, C. (1983). *The Invisible Farmers: Women in Agricultural Production*. Totowa, NJ: Rowman and Allenfield.

Saugeres, L. (2002). The cultural representation of the farming landscape: Masculinity, power and nature. *Journal of Rural Studies* 18, 373 404.

Woodward, R. (2000). Warrior heroes and little green men: Soldiers, military training, and construction of rural masculinities. *Rural Sociology* 65, 640 657.

Gender in the City

Linda Peake, York University, Toronto, ON, Canada

Glossary

Affect A term referring to an emotional state – love, hope, fear, anger – without any specific reference to an object (such as love of one's neighbor or hope for peace), indicating that it is a nonrepresentational mode of thought.

Gender The social construction of masculinities and femininities. It is increasingly recognized as a problematic term that does not translate easily into other languages.

Global South Underdeveloped countries that have been locked out of economic and social advancement by processes of imperialism and the economic policies of wealthier nations.

LGBTTQ This acronym refers to lesbian, gay, bisexual, transgendered, transsexed, and queer folk. It is increasingly seen as a problematic term, referring to both gender and sexuality for its defining features.

Masculinity Studies An interdisciplinary field of study concerning men, masculinity, gender, sexuality, and their associated politics.

NIDL New international division of labor.

Urbanization The process of a country's population becoming urban and of cities becoming the engines of growth of the economy.

Introduction

From the outset, gender in the city has been an interdisciplinary field that has attracted both academics and professional practitioners, with the earliest conference session probably being the 1976 UN Habitat conference in Vancouver that led to the production of the *Women and Environments* magazine, which has recently celebrated its thirtieth anniversary. In the discipline of Geography, gender in the city is a field of study that has chiefly interested feminist geographers. Although research about gender and the city need not necessarily focus on women (it can, for example, examine the structures of power and material discourses that work to produce inequality between women and men), the predominance of feminists in this field goes a long way to explain why the vast bulk of research has centered on women, although this is now starting to change given the inroads that are being made by LGBTTQ and Masculinity Studies. And it has been the predominance of white feminists, and the hegemony of Anglo American geography, that explains the focus on white women in Western cities, although this too is changing. So in the twenty first century, when the interests of feminist geographers and other feminist scholars, including historians, political scientists, sociologists, planners, and architects, can no longer be neatly defined or categorized, the field of gender in the city refuses easy overview.

Something Old, Something New?

Feminist perspectives are relatively new to Urban Studies, a field which started in continental Europe in the 1870s. Anglo American feminist geography emerged a century later in the early 1970s – a result of second wave feminism – and there is widespread agreement that its longest standing field of enquiry has been gender in Western cities, especially its social, economic, and political dimensions. Thus, the first feminist geography publications (in *Area* and *Antipode*) were on urban topics, respectively, a study of the childcare needs of single mothers and a critique of Marxist conceptions of production, social reproduction, and urban form. These articles were also illustrative of the diverse interests that would come to preoccupy urban feminist scholars, ranging from simply enumerating and comparing the everyday lives of women and men, to making distinctions between sex and gender, to developing feminist urban theory. The specific factors behind the development of the topic of gender in the city were, on the one hand, the influence of second wave feminism on (Anglo American based) feminist geographers and, on the other, the overwhelming focus by the critical left in Geography on urban economic based theories of capitalism – itself a result of the fact that in advanced industrialized societies most aspects of everyday life took place in cities and towns. But this is not the whole story. It is often forgotten that at the same time feminist scholars looking at issues of development in the Global South also turned to studies of gender in the city. Moreover, these studies of gender in urban places in the Global South have, increasingly, come to encompass studies that originate in the South, giving rise to a vibrant literature. Literatures on women in Western cities and women in the urban Global South, however, have tended to follow separate trajectories and rarely acknowledge each other. This is by no means unusual; the discipline of Geography is replete with accounts of subdisciplines talking past each other. In this case, it may be explained by Anglo American feminist scholars' preoccupation with the development of critical urban theory while the focus in the Global South has been primarily on poor women's lives in towns and cities.

A surficial review of the research in this field might lead to the assumption that much remains the same. Certainly, some of the topics of enquiry appear reassuringly familiar and date back to the late 1970s, as **Table 1** shows. The first studies of women in Western cities developed from theoretical engagement with processes of urbanization and of the gendered dimensions of urban form, especially issues of city centers being regarded as sites of masculinity/ production/ paid work and suburbs as sites of femininity/ social reproduction/ unpaid work. As so much of the work was initiated by feminist planners and architects aspects of built form, such as urban planning, design, and architecture, including those related to domestic spaces and the home, were also of interest. Studies also focused on labor market and work–home links, as well as access to facilities such as childcare and transportation. Much of this work was already theoretically sophisticated, building as it did on critiques of Marxist, neo Marxist, and Weberian accounts of the development of urban form. Given the focus on critiques of capitalism it was hardly surprising that, at least in Britain, feminist research in Geography came to be dominated by studies of the interlocking axes of power of gender and class, with socialist feminism being the predominant theoretical focus. From this socialist, feminist theoretical base, urban feminist geographers in the mid 1980s expanded their repertoire into urban politics (including women's participation in formal and informal spheres of politics and in urban social movements), housing (especially in relation to homelessness and gentrification), urban based identities (including gender, class, 'race' and ethnicity, and urban geographies of patriarchies), and women's fear in, and of, urban places.

Other research clusters, not having as long a heritage in feminist work, established themselves from the early 1990s onwards, as a result of feminist geographers' increasing engagement with theoretical developments in the social sciences such as postmodernism, poststructuralism, psychoanalytic theory, third wave feminism, and critical race theory, resulting in critiques of earlier work for its lack of engagement with feminisms outside the socialist orbit. Urban geographies of women of color started to appear, especially of immigrant and refugee women, First Nations and Aboriginal women, as well as a focus on specific groups of women workers such as domestic workers and sex workers. Other topics emerged because of the increasing attention being paid to alternative sexualities to heterosexuality such as to queered urban spaces. The work of geographers of sexuality also led to an intense interest in embodied urban geographies. Accounts of urban public space joined those of urban form. And there has, increasingly, been space for studies of women's agency in relation to urban consumption and women's spaces of pleasure/leisure, as well as those of spaces of urban poverty.

Table 1 Gender in the Western city research foci

Gender in the Western city research foci	*1970s*	*1980s*	*1990s*	*2000s*
Gender and urbanization; urban form: city center versus suburbs; processes of production (paid work) versus reproduction (unpaid work); urban restructuring	x	x	x	x
Labor markets and work home links, including childcare; transportation and access to facilities; (as well as from the late 1990s with studies of parenting, childhood and children, care giving and the ethics of care)	x	x	x	x
Urban planning, design, and architecture	x	x	x	x
Women and urban politics; urban social movements; women's political participation (as well as from the late 1990s with issues of citizenship).	x	x	x	x
Domestic spaces; the home		x	x	x
Urban-based identities of gender, race, ethnicity, and sexuality; urban geographies of patriarchies; social constructions of difference, especially in relation to femininities and masculinities (with the introduction in the late 1990s of issues of racism, whiteness, and transnationalism)		x	x	x
Women's fear and urban places (with the introduction in the 2000s of issues of surveillance)		x	x	x
Urban poverty		x	x	x
Housing: homelessness and gentrification		x	x	x
Immigrant women, First Nations, Aboriginal women in cities; a focus on specific groups of workers such as domestic workers and sex workers; (with the introduction in the late 1990s of issues of racism, whiteness, and transnationalism)			x	x
Lesbian and gay urban geographies; LGBTTQ spaces			x	x
Embodied urban geographies			x	x
Urban public space (though earlier studies on, e.g., women's access to parks and women-only buildings date back to the late 1970s)			x	x
Women's spaces of pleasure/leisure, the flaneur			x	x
Urban emotional geographies				x

Two trajectories, however, lead to a critique of the view that the field of gender in the city in the new millennium is merely more of the same. First, since the late 1990s new topics that emphasize the diversity of women's experiences have made their way into the lexicon of feminist geographers including urban citizenship, racism and whiteness, children's urban geographies, urban emotional geographies, and visual methodologies. Second, established topics are being addressed through new tensions. A renewed focus on the work–family nexus through studies of parenting, childhood, caregiving and the ethics of care is replacing earlier studies on home–work linkages and women's dual role. In addition, the consideration of transnationalism has necessitated a blurring of boundaries and the imploding of seemingly fixed hierarchies; a focus on migration, household reproduction, family policy, and marginalized workers is indicative that the Global South is no longer the sole reserve of development geographers as Western cities have become increasingly integrated and inculcated with global flows of people, capital, goods, and ideas.

The field of gender in the city in relation to cities in the Global South commenced, in the 1960s, with studies of rural women's migration to urban centers, as well as on the role of women's work in the informal sector in urban areas (see **Table 2**). In the 1970s, research started to make inroads questioning the representations of women in these earlier studies, especially in relation to their sexuality and engagement in prostitution. The focus also turned to women's roles within households but it was not until the late 1970s that feminist research developed momentum. In the 1980s, studies continued with a focus on women's triple role as mothers, workers, and community organizers. In particular, women's efforts to secure housing and to upgrade their homes, as well as accounts of their daily lives in urban slums and squatter settlements, united with studies of urban politics to investigate how women organized to secure services for their communities. Studies of household structures and accounts of urban poverty formed another set of interests, especially investigations of how resources were distributed within the household. In the 1990s, this theme merged with studies of the impact of structural adjustment policies on low income households. The survival strategies that households devised to cope with the economic violence unleashed on their lives emerged alongside studies of women's experiences of domestic violence. The latter overlapped with studies that looked at the health concerns of women, especially reproductive health, and the lives of urban youth and children. An interest in specific groups of women workers led, in particular, to studies of sex workers and transnational migrant workers. Studies also began to emerge on the gendered nature of urban space and on urban environmental issues, such as women's engagement in urban agriculture, the impact on women's lives of the degradation of urban environments, and urban feminist political ecology. Most recently, in the 2000s, studies of alternative sexualities have started to appear.

Table 2 Gender in the urban Global South research foci

Gender in the urban Global South research foci	*1960s*	*1970s*	*1980s*	*1990s*	*2000s*
Women and urbanization processes, especially rural to urban migration, urban marginalization/culture of poverty (with its focus on the immorality of women and an emphasis in the 1960s on modernization policy)	x	x	x	x	x
Women and work: formal and informal sectors; women as active in production and reproduction (in the 1970s, with a focus on basic needs policy and, in the 1980s, with a focus on the implications of the NIDL); feminist critiques of prostitution		x	x	x	x
Land, housing, and human settlements; urban poverty; women's triple role			x	x	x
Household structures; household survival strategies; distribution of resources within the household (with a focus from the 1990s on transnational households)			x	x	x
Urban politics, especially squatter movements; urban service provision such as water, childcare, transportation			x	x	x
Urban crisis survival strategies; impact of SAPs (structural adjustment programs); urban citizenship				x	x
Urban violence, domestic and economic				x	x
Specific groups of workers, especially transnational migrant workers and sex workers;				x	x
Urban environments: degradation of urban environments; women farmers in urban areas; urban feminist ecology				x	x
Gendered nature of urban space				x	x
Urban health; women's reproductive health; AIDS/HIV				x	x
Urban children and youth				x	x
Sexualities and queered urban spaces					x

NIDL = New International Division of Labor.

Gendered Understandings of Urban Form and of Gendered City Lives

In the second half of this entry I turn to brief accounts of some of the major current themes that have been high lighted by feminist scholars in the field of gender in the city.

The Restructuring of Urban Form and of Women's Lives

Feminist concerns have focused on both the role played by social relations, especially those of gender and class, in the production and reproduction of urban form, and the extent to which the social position of women has im proved with urbanization. Data on sex ratios in urban areas reveal that women outnumber men, but for cen turies there have been various attempts (not least sym bolically) to assign women to the private sphere of the home and even to prevent them from entering public spaces. While the association of women with the private and men with the public is now recognized as more of an ideal, or ideology, as opposed to a reality, the concept of 'separate spheres' has been a powerful one. Work on the effect of segregated land uses – of home and work – in post World War II Western cities, emphasized how women tended to be housewives and men, breadwinners. Many (white) feminists have thus, argued that Western urban forms, in particular the mid twentieth century divide between city center and suburbs, have served to increase women's subordination by facilitating the re moval of job opportunities from women and relegating them to the domestic world of the home. More recently though, postmodern feminists have emphasized the eco nomic, social, and sexual freedoms and opportunities that have arisen from urbanization. The rise of department stores in the nineteenth century, for example, gave women a legitimate right to be in public spaces, as did libraries, museums, and galleries. Critiques of the dichotomy of male/public and female/private spaces also came from African American feminists who argue that women of color have a long engagement with the public sphere through waged work. The destabilization of these het erosexualised male and female spaces is also evident in other various ways; from the gentrification of neighbor hoods and lesbian and gay pride parades, that LBGTTQ folks assert their presence in the urban environment.

Besides the focus on waged work, research on women (historically and in the present) providing housing for themselves and their children, has also contributed to studies of changing urban forms. Perhaps, in recent years, gentrification has been the major topic related to housing that has drawn the attention of feminist geographers. The regeneration of run down, inner city areas has afforded single women and female headed households the op portunities to live downtown in mixed use areas where accessibility to job and services is maximized. The lit erature on gentrification, however, tends to downplay the libratory potential it holds for women; despite the op portunities central downtown locations hold for women to live outside the confines of the nuclear family, the diversification of household forms in the suburbs and the apparent increase in paid domestic help in high income dual earner households, suggests that changes in gender relations may well be taking place faster in places that were, traditionally, considered bastions of the nuclear, heterosexual, middle class family.

Women's increasing entry into the labor force in the latter half of the twentieth century further added to urban restructuring through the emergence of new spatial div isions of labor. Firms have been shown to change their locations to attract female employees, such as relocating to suburbs or to public transportation hubs (given that women use public transport more than men), and also moving to countries where wages are lower. However, research on how the latest round of neoliberal, urban governance is reconstituting urban space and gendered identities in Western cities is only at an early stage. Studies have examined how women who are mothers juggle life–work balance in urban arenas. Men's and women's roles as par ents have come under increasing scrutiny as have the schedules and strategies that households develop to cope with restructuring (and uncertainty). The ways in which households are becoming more implicated in each other's social reproduction, through both paid and unpaid rela tionships is also under investigation. The rise of com modified caring is being examined both in relation to children and elder care. And although the majority of women live lives very much rooted in place, studies of the ways in which women maintain transnational households, both from positions of privilege and disadvantage, have increased. What these studies reveal is that it is clearly at the micro urban level that we begin to understand how race, family structure, class, and stage of life cycle affect women's mobility and opportunities.

What is known about urban restructuring under neoliberalism, characterized by public–private sector partnerships (what some refer to as the 'shadow state'), is that it has resulted in increasing social and economic polarization and has greatly increased the burdens on many women's lives, due, partially, to the rise of pre carious employment. There have also been major impacts on women's service organizations which have been forced to focus more on service delivery and to relinquish advocacy roles to maintain state funding. The divides between women in Western cities with professional, well paid, unionized jobs with pensions, versus those in low paid, part time, and (in some cases) home based jobs with no recourse to benefits have substantially increased, while in cities in the Global South, there has been a rise in work in informal sectors as the formal sectors in many countries have collapsed under the burdens of structural adjustment policies and neoliberal governance, resulting (in both the West and the Global South) in the feminization of poverty. Increasing engagement in the informal sector, though, has also included, for example, more women starting home based businesses based on food services, attempting to sell produce grown in community or home gardens, and providing care for children and the elderly. Thus, the collapse of formal sectors, also presents 'opportunities'. For example, women in the barrios in Lima, Peru, formed community kitchens and the Glass of Milk program for children (that was ultimately taken over by the municipal government).

Urban Planning and Design

The literature on urban form has been very closely aligned with that of 'women and planning' where the focus has been on the practicalities of women's everyday lives as mothers, household members, workers, and activists. Some feminist planners have argued that urban planning is preoccupied with the need to control the city and, *inter alia*, women; it is, they argue, as much an ideology as a technical practice. With its origins in nineteenth century scientific rationalism, planning has searched out logical solutions to the allocation of scarce resources in attempts to create 'order out of chaos'. However, the rational comprehensive planning focus has long been usurped by other paradigms, including communicative planning, radical planning, insurgent planning, and by feminist practitioners whose emphasis is on the emancipatory and just city and on democratic participation in reshaping the city.

Many feminist planners consider the interrelationship between land use and transportation to be a key factor in determining the degree of ease with which women can live their daily lives. The rigid imposition of zoning, the separation of land uses, and the development of urban sprawl in Western cities have created arduous time/space activity patterns for women, given their responsibilities for childcare and the home. Studies have shown that inequities in access to transportation still constrain the life chances of large numbers of women. For example, while the mobility of disabled, working class, and immigrant women is severely restricted and they are 'spatially entrapped', professional women often spend the same amount of time commuting in their journeys to work as their male counterparts. Transportation has also been the focus of urban movements such as the Bus Riders' union in Los Angeles, led by women demanding equity for poor and visible minority women without access to automobiles. Although regulatory mechanisms are less rigid in many cities in the Global South, the problems experienced by women are exacerbated by nonfunctional social and physical infrastructures, as well as an increasing inability of national and local governments to manage and regulate urban life.

Since the 1970s, there have arisen vibrant, loosely knit women and planning movements in both Western and Global South cities in which women have often tried to bridge the informal/formal sectors through activism that occurs at a range of scales from the home/neighborhood to the city, region, and transnational/international. Often combining professional and community based women, these movements have given rise to a sustained feminist critique of urban planning issues such as housing, child care, disability, retailing, safety, sport and leisure, technology, and transport. They have illustrated the positive role feminist planners can play in the reproduction of social relations through alternative visions of urban life and the creation of new urban spaces. They have dealt with issues as mundane (but vital) as the provision of public toilets, as well as creating city wide movements focused on reducing violence against women. They have also emphasized women's historical contribution to urban planning, for example, in North America where women's contributions to planning and the construction of nineteenth century cities (often linked to temperance and abolitionist movements) led to the development of settlement houses, shelters, and women's clubs. But women centered planning policies are few and far between. Currently, in many places, the professional (and invariably masculinist) culture of planning, combined with a scarcity of resources and a feminist backlash, has further served to marginalize feminist perspectives in planning as, ironically, have the fashionable tropes of environmental sustainability and human rights in the urban Global South. Poverty, economic restructuring, armed conflict, and environmental degradation are all resulting in increased mobility and moves to urban areas where millions of people live in squatter settlements and shanty towns in the Global South. Here the focus has been on studies of the regularization of illegal/informal settlements and the everyday difficulties facing women such as securing the provision of services and employment.

Urban Public Space and Urban Citizenship

Some of the most analytically exciting work on urban public space, originating in the 1990s, has drawn attention to the changing contours of the relationship between gendered public and private spaces. Feminist geographers have questioned recent concerns about the decline of urban public space and the consequences of this for democratic practices, and have also questioned the fixed meanings that have been attached to public space, in particular, that of the correlation between the public and masculinity. Focusing on specific groups of women in various historical periods, they have revealed the constant slippages between feminine–masculine and public–private divides. Some have argued that as women traversed the boundaries of the private, the so called masculine spaces into which they were intruding became feminized in order to neutralize such a potentially disruptive act. Hence, what is commonly perceived to be the decline of public space is the feminization of such space, usually as commercial space. However, while women may increasingly be included in such spaces, others are not, such as the poor, the homeless, and people of color. Feminists have thus investigated the implications of urban public space for identities and citizenship, pointing out how urban spaces, throughout time, have been regulated in order to control the social reproduction of groups considered in need of control – women, the working class, the homeless. The attempts by groups to reclaim these urban spaces have also been well documented, from such organized resistance such as Take back the Night Marches by women, and gay and lesbian pride parades, to routine everyday practices, such as breastfeeding in public.

Feminists have also highlighted the contradictory nature of women's experiences of urban space with historical research revealing how middle class women's experience of public urban spaces – through their physical presence in spaces of leisure and shopping – was imperative to capitalist reproduction while working class women were heavily engaged in production. Recent studies on the revitalization of old urban centers through landscapes of consumption and spectacle have explored the impact on gendered identities, emphasizing that, even in landscapes that actively promote diversity and democratization, traditional gender roles, unequal gender relations, and normative sexualities can still hold sway. Even the glamorous, liberated, affluent (and white) single women in the hugely popular US TV show Sex and the City ended up happily married. Feminist research on moral geographies of the street, which has focused on the regulation of sexuality through research with and about sex workers, has also challenged the notion of urban public spaces as ones of democratic encounters. Most recently, there has been increasing interest in how bodies move through, and occupy, changing urban spaces and how the form and ideology of the public are mutually constitutive of shifts in cultural inscriptions of the body (see **Figure 1**).

Closely aligned with work about urban public space is research concerned with urban politics, and most recently with the concept of urban citizenship. Feminist work on urban politics has been longstanding, but spotty. While there has been increasing recognition of women as political actors, given their incursions into the public spheres of local and national politics, so much of what women do – for example, urban community activism – has not been considered worthy of analysis. Hence, feminists have argued that it is necessary to separate the content of activities from the spaces in which they are conducted. Doing so not only reveals the broad range of women's activism but also sharpens understanding of the public–private divide and of what is considered political activity. For example, marginalized women using place based identities to engage in urban environmental justice movements are not to be dismissed as women engaging in domestic activities, but rather as an example of a mobilized civil society engaging in democratic practices. Feminist urban political research has also turned to new possibilities for oppositional politics in urban public space and the notion of citizenship as a sense of belonging. In both Western cities and those of the Global South over the past 30 years, women as urban citizens have formed and led movements to acquire resources to meet the needs of everyday life, have been able to enter local state politics, and have created space for feminist policies and women's rights to representation, as well as forging transnational networks that have expanded concepts of citizenship.

Housing/Homelessness/Homes

Research has emphasized the various ways in which women house themselves, including through women's co operatives, and the impact in Western cities of neoconservative agendas on women and children's access to housing. In the eyes of some though, the field of women and housing in Western cities, has become moribund, preoccupied with women's access to the neoliberal dominant form of owner occupation, and lacking in analytical potential. Hence, some have recently argued for a rethinking of housing as 'living room', and for an exploration of the entanglement between culture and nature – between human and nonhuman actors – in the creation of housing. Indeed, there is a nascent field of urban feminist political ecology that is explicitly exploring the ways in which, for example, water supplies and gardens form affective links between humans and housing stocks and how houses are saturated with links between the organic, material, and emotional realms. In the Global South, however, where rates of urbanization exceed those of the West, and where millions of people are accommodated in informal housing, living in slums and squatter settlements, the focus has been on women's attempts to upgrade their homes and secure the provision of services for their neighborhoods.

Unlike housing, the topic of women and homelessness in larger and smaller cities, in both the West and the Global South, has attracted increasing interest as the numbers of homeless women increase and as their relative invisibility in relation to men becomes better understood. In the Global South, research is increasingly focusing on the situation of homeless (and often stateless) refugees, millions of whom live in squalid and dangerous conditions in camps that often resemble large towns in their size, although often lacking any services and employment.

Figure 1 Some sympathetic and less sympathetic embodied representations of women in the city. (a) Alison Lapper Pregnant, a sculpture by Marc Quinn occupies the ‘fourth plinth’ in the Northwest of Trafalgar Square, London, England. The sculpture was unveiled on 15 September 2005 by the Mayor of London, Ken Livingstone, at a public ceremony with both the artist and Alison Lapper in attendance. (b) ‘The Endless Bench’ by Lea Vivot, Unveiled 7 July 1984, donated by the artist to the Hospital for Sick Children, Toronto, Canada. (c) Neon Yellow Lady Sign for Club Paradise (Gentleman’s adult entertainment club), 1313 Bloor St., Toronto, Canada. Image courtesy: Kathryn Travis.

More recently, the attention of some Anglo American feminist researchers has focused on the complex and contradictory emotional geographies of home and domestic spaces. The home is predominantly still viewed as a site of domesticity and social reproduction despite its increasing incorporation into worlds of employment, whether high paid computer based work or low paid domestic or contract work. But there is also increasing recognition of the role of technological developments, particularly access to the Internet and telecommuting, in reconfiguring home–work spaces. Although this access gives working women more flexibility in the timing of activities over the working day, research shows that rather than challenging normative gender roles it has tended to cement them.

The home has also received attention as both a material and an affective place – a site of belonging and of alienation, of loving relationships and of violence. Recent research has focused on the affective nature of particular places in the home, such as kitchens, and the ways in which particular artifacts, such as photographs, reveal the symbiotic nature of the construction of female subjectivities and urban spaces. There is also an escalating research trajectory relating to transnational geographies of home(s).

Women’s Fears and Pleasures in the City

Within this small but burgeoning field, the dominant theme has been that of how cities contain and oppress women, enhancing their fear of violence. Yet, there has also been substantial and sustained women’s resistance to this framing from the early 1980s to the present; going beyond women’s fear it focuses on women’s interventions in policy and programs and in the creation of alternative service responses. While research on urban women’s fear started with a focus upon improving urban design to make urban places safer for women it has now moved on to the mutual construction of identities and urban space.

Researchers are also seeking to further conceptualize understandings of an apparent ‘spatial mismatch’, namely, while studies have revealed that women have a greater fear of being assaulted in a public place they are at greater risk of being sexually assaulted in the 'privacy' of the home by a male of their acquaintance. Most recently post structuralist understandings of subjectivities, which resist the notion of gendered roles and identities as fixed, and thus, see femininities as being discursively produced through a variety of material practices, have been employed to show how an individual woman may or may not be enticed to challenge dominant discourses such as that of patriarchal social control. Others have illustrated how constructions of white women's fears fail to take into account feelings of fear dealing with encounters with racialized men. Race shapes women's experience of fear and safety, both at the scale of the city and in discrete public spaces, but there is no simple reading of 'race' and fear. Racialized experiences of fear are strongly mediated by class identities and urban form, adding to our understanding of the complex and multiple origins of fear. Most recently, literature on women's fear has turned to two different trajectories. First, studies in the post 9/11 world have turned from the oppressive nature of the male gaze to the technological gaze. In urban space women are more likely to be looked at than men, both in terms of face to face encounters and mediated gazes, such as that of the surveillance camera, revealing the deeply gendered nature of new urban surveillance technologies. Other studies have also turned to the very real consequences of the male gaze, focusing on male violence against women, such as recent studies of the disappearance and murders of *maquiladoras* in Mexican cities and of sex workers in Vancouver.

This analytical view of the city as a space of containment for women has been challenged by those who put forward a view of the city as a contradictory and shifting space that allows women opportunities. The enjoyment that the anonymity of the city affords, for example, is a theme that has been explored through the trope of the flaneur. This paradigmatic figure has been reappropriated by lesbian analysts who have explored how women throughout time have been able to explore the city in ways that are not mediated through men. Research has also shown how women have made numerous attempts to create their own urban spaces, from nunneries and women's housing co operatives to lesbian bars. Probably the most significant in current times are the initiatives that have emerged around women's safety, such as half way houses and refuges for battered women. But it is not only through engagement in material city spaces that women have challenged the patriarchal structuring of urban social space; women have also created imaginary urban places, such as the urban utopia, Herland, of Charlotte Perkins Gilman and Dolores Hayden's Non Sexist City model.

See also: Feminism and Work; Feminist Geography, Prehistory of; Gay Geographies; Gender, Historical Geographies of; Gentrification; Ghettos; Home; Homelessness; Housing; Immigration II; Modern City; Patriarchy; Planning, Urban; Poverty; Public Spaces, Urban; Slums; Squatter Settlements; Urban Design.

Further Reading

Bashevkin, S. (2006). *Tales of Two Cities: Women and Municipal Restructuring in London and Toronto*. Vancouver: UBC Press.

Bell, D., Binnie, J., Longhurst, R. and Peace (eds.) (2001). *Pleasure Zones: Bodies, Cities, Spaces*. Syracuse: Syracuse University Press.

Bondi, L. and Rose, D. (2003). Constructing gender, constructing the urban: A review of Anglo American feminist urban geography. *Gender, Place and Culture* 10, 229 245.

Dandekar, H. C. (1993). *Shelter, Women and Development: First and Third World Perspectives*. Ann Arbor, MI: George Wahr Publishing.

Fincher, R. and Jacobs, J. M. (eds.) (1998). *Cities of Difference*. New York: Guilford.

Garber, J. A. and Turner, R. S. (eds.) (1995). *Gender in Urban Research*. Thousand Oaks, CA: Sage.

Greed, C. (2003). *Inclusive Urban Design: Public Toilets*. Oxford: Architectural Press.

Grosz, E. (1992). Bodies cities. In Colamina, B. (ed.) *Sexuality and Space*, pp 241 254. Princeton: Princeton Architectural Press.

Hayden, D. (1981). *The Grand Domestic Revolution: A History of Feminist Design for American Homes, Neighborhoods and Cities*. Chicago, IL: University of Chicago Press.

Little, J., Peake, L. and Richardson, P. (eds.) (1988). *Women and Cities: Gender and the Urban Environment*. Basingstoke: Macmillan.

MATRIX (1984). *Making Space: Women and the Man Made Environment*. London: Pluto Press.

McDowell, L. (1997). *Capital Culture: Gender at Work in the City*. Oxford: Blackwell Publishers.

Miranne, K. B. and Young, A. H. (eds.) (2000). *Gendering the City: Women, Boundaries and Visions of Urban Life*. Lanham, MD: Rowman and Littlefield.

Moser, C. O. N. and Peake, L. (eds.) (1987). *Women, Human Settlements and Housing*. London: Tavistock.

Moser, C. O. N. and Peake, L. (1995). Seeing the invisible: Women, gender and urban development. In Stren, R. & Kjellberg, B. J. (eds.) *Perspectives on the City, vol. 4: Urban Research in the Developing World* pp 279 348. Toronto: Centre for Urban and Community Studies, University of Toronto.

Pile, S. and Nast, H. (eds.) (1996). *The Body and the City: Psychoanalysis, Space, and Subjectivity*. New York: Routledge.

Rendell, J. (2000). *Gender and Architecture*. Chichester: Wiley.

Sheldon, K. (ed.) (1996). *Courtyards, Markets, City Streets: Urban Women in Africa*. Boulder, CO: Westview Press.

Walkowitz, J. (1992). *City of Dreadful Delight: Narratives of Sexual Danger in Late Victorian London*. Chicago: University of Chicago Press.

Wekerle, G. (1984). A woman's place is in the city. *Antipode* 16, 11 20.

Wekerle, G., Peterson, B. and Morley, D. (eds.) (1980). *New Space for Women*. Boulder, Co: Westviwe Press.

Wilson, E. (1991). *The Sphinx in the City; Urban Life, the Control of Disorder, and Women*. Berkeley: University of California Press.

Wilson, E. (2001). *The Contradictions of Culture: Cities, Culture, Women*. London: Sage.

Relevant Websites

http://www.gendersite.org
This website comprises a cross disciplinary database of bibliographic references, policy documents and web resources relating to issues around gender and the built environment.

http://www.shehr.org
This website Shehr is Urdu for city addresses gendered urban research in the Middle East, Africa and South Asia.

Gender, Historical Geographies of

L. Price, Queen's University Belfast, Belfast, UK

Glossary

Dualistic Gender The belief in a binary or categorical distinction between men and women.

Essentialism A biologically determinist argument determining gender inequalities to be an essential biological difference between men and women.

Feminist Geography Scholarship dating from the late 1960s which, in general, aims to demonstrate the construction and significance of sexual differentiation as a key organizing principle and axis of social power, as well as a crucial part of the constitution of subjectivity.

Gender and Geography The study of the various ways that geographies and genders are mutually constituted.

Gender Order A regime or order consisting of distinguishing relations of power, production, and emotional attachment.

Gender Relations Inequitable relations of gender based on material social practices.

Gender Roles The spatial and temporal socially defined arrangement of gender characteristics.

Patriarchy Within feminist scholarship, the system by which men as a group are assumed to be superior to women as a group and so to have authority over them.

Postmodern Gender Identity A hybrid notion of unstable, mixed, and multiple ideas of individual identity.

Representation The ways in which meanings are conveyed or depicted.

Separate Spheres The separation across time and space of the social worlds of men and women into arenas such as public/private and home/work.

Sexuality A term emerging in the late nineteenth century to refer to patterns of behavior associated with sexual expression.

Introduction

Historical geographies of gender can be said to have four key strands. First, the historiography of the geographical discipline itself from its inception in the nineteenth century onwards has been critiqued from feminist, gender perspectives as encompassing a white, male, Eurocentric gaze which has impacted on ways of knowing and practicing historical geography. Second, gender perspectives have led to a revisiting of themes and connections between space and time with the 'adding in' of women to historical lenses of enquiry from the medieval period onwards. Third, feminist theorization of spatially variable gender relations from the 1970s onwards had led to new and exciting historical geographies illuminated by gender which are rethinking the links between the development of Western industrial capitalism, urban spatial evolution, and relations of gender. Finally, ontological and methodological approaches to historical geography and the ways in which the past impacts upon the present are being influenced by postmodern ideas of gender identity and representation. Thus, historical geographies of gender continue to draw on conceptualizations of gender from across the social sciences providing new ways of linking the past and the present over space and time.

The Geographical Discipline

A key impetus to reexamining the historiography of human geography arose in the 1990s when an article by Mona Domosh criticized the discipline for ignoring feminist theory and the formulation of geographic knowledge from women's ways of knowing. A rebuttal of Domosh's assertions by Stoddart in the same year sparked a debate as to the extent to which geography's historiography required feminist infusion. Importantly in 1995, Gillian Rose highlighted how the geographical tradition had reflected paternal lines of geographers which served to erase women from the structures of the discipline. This, it was suggested, rendered women invisible in historical geographic accounts of places over time. Hence, Livingstone's highly acclaimed 1992 book of the geographical traditional was criticized for only mentioning two women in the 500 years of Western geographies it covered.

Much work in historical geography has been subject to reexamination and shown to be highly influenced by the male perspectives of its members. For example, it was contended that Victorian societal structures precluded women from traveling alone; to publishing under their own names or to having independent incomes and that this would naturally exclude them from geographic time specific practices and organizational structures. Therefore, as they were not 'there', a rewriting of geography's historiography is considered unnecessary from such a perspective. A view now largely consigned to history.

It has been suggested that as the early European nineteenth century geographers attempted to establish the academic boundaries of the discipline around scientific, rational, and exploratory endeavors in the field of colonial imperialism, geographers in the Western World developed complicated hierarchies of empiricist classification and ordering of artifacts and ideas from far and distant places. A key point made by the Women and Geography Study Group in their 1991 book is that the knowledge derived from such exploratory endeavors is male centered as it developed from institutional disciplinary practices which largely excluded women. An initial impetus to highlighting historical erasure of women from the geographical tradition has found researchers revealing a number of women such as Isabella Bird and Mary Kingsley who, when funds were available, were able to engage in nineteenth century travel and ethnographic fieldwork. They made social connections and revealed realities of colonized people in new and uncharted territories across the globe. Such exploits, revealed through personal diaries, continue to be discovered and published by researchers. Such revelations contribute to the ongoing revision of the historiography of the geographical discipline. Thus, criticism remains that it was the ethnographic practices themselves that were seen as nonscientific and nonmale, failing to fit into the scientific, quantitative, and cartographic rigidity of the newly institutionalized Royal Geographical Society. Despite women being eligible for entry into the Scottish Royal Geographical Society (SRGS) in 1884 and initial entry into the London based Royal Geographical Society (RGS) in 1892, opposition by committee members to the 1892 RGS entry meant that women in the United Kingdom remained largely excluded from the society until 1913. An institutional legacy and bias which many of those involved in writing the Women and Geography Study Group book argue continues to this day within the experiences of female, professional geographers.

Adding Women to Historical Geographies of Gender

As American civil and equal rights movements began to influence early feminist geographers from the 1970s onwards, historical geography became infused with new gendered perspectives critical of male centered knowledges and was keen to offer new descriptions of changing places over time. Gender initially equated to 'women' and thus the traditional foci of historical geography on human–environment relationships began to be infused with gender analysis. This challenged previous views of geographies and their relationships to the present by adding women into fields of vision. Everyday spatialities of women's lives in the past began to be revealed as new conceptions of 'work', for example, began to influence political and socioeconomic historical geographies. This is sometimes referred to as the 'geography of women' and has a close relationship with the development of feminist geography more generally. Drawing on ideas of essentialism, where social roles are derived from perceived biological differences between men and women, historical geographies based on how such gendered male/female roles differentiated spheres of operation began to be revealed. Traditional topics of interest to historical geographers such as medieval agrarianism and economic settlement patterns have been shown to be highly reliant on the work, organization, connections, and experiences of women. Jane Humphries, for example, has demonstrated how enclosure and the erosion of common rights in the eighteenth century contributed to the proletarianization of families and the development of men as the industrialized familial breadwinner. Drawing on market records, it was shown how previously women were dealers and manual workers as their menfolk worked in other occupations, often within the home.

Many other traditional areas of historical geographical enquiry which focus on the socioeconomic labyrinth of global connections over space and time have been reexamined with a gendered lens of enquiry. Examination of gender embedded in complicated, global relationships has been shown to be crucial to an understanding of the temporal rise of colonialism and global industrialization. For example, justifications for slavery and imperialism have been exposed as being framed within the banner of the civilizing of savage peoples which was persistently based upon a vision of non Western women as more suited to manual work. The perceived lack of morality of non Western women has also been shown to have led to a justification for sexually deviant activities by male colonial plantation owners. Western imperialism was reexamined, therefore, for the ways in which colonization of so called 'savage' peoples was justified in opposition to 'civilized' people and presented as such in nineteenth century geography school texts. Notions of civility have been illuminated with the aid of gender perspectives. This demonstrates that it was the idealization of women as the 'angel in the house', producing superior offspring which contributed to the idea of colonial improvement. Incorporating gender into historical geographies has also opened up new ways of seeing which challenges notions of migration as predominantly led by men. When women's migration patterns are examined via a gender lens women's colonial practices are often revealed as being pioneering. For example, women played a vital reproductive role in providing and socializing a new breed of colonials loyal to a superior vision of the home country, thus contributing to colonial institutions and practices. Such practices by women associated with Western diplomatic service have a wide audience in a variety of media. Joanna Trollops' 2006 novel, for example, *Britannia's Daughters* draws on historical letters and diaries to provide a new view of geographies of women's journeys in the past as diplomat's wives, nurses, and even prostitutes.

Historical Geographies of Gender Relations and Capitalism

As feminist geography continued to theorize gender into the 1980s, a move away from gender conceptualized as natural biological sex differences between male and female to ideas of socially constructed gender characteristics based on a dichotomous and symbiotic relationship between male/female, masculinity/femininity began to influence historical geography. The concept of hegemony became crucial to an understanding of such theorization as the male side of this dualistic relationship was deemed to be hegemonic over the female. Influenced by critical human geography derived from Marxism, feminist researchers highlighted how a re reading of historical geographies might now be engaged in to challenge geographies of gender which were seen to be culturally, spatially, and temporally specific. Therefore, depending on when and where you lived, it is suggested that the experience of being male or female would have been differentially guarded by physical and social boundaries of operation and behavior. Key thinkers in this area, such as Linda McDowell, add to our knowledge of historical geographies of gender by revealing the gendered organization of space which reinforces unequal gender social relations. A key link between gender relations and the evolution of capitalism was made by groundbreaking theorists, such as Sylvia Walby, as it was now realized that capitalism was dependent upon gendered economic and consumptionist processes which relied upon the temporal relations between Western men and women being relational, differentiated by gender and by being patriarchal.

Hence, rather than merely a description of where women were, what they were doing, and their contribution to topics and incorporation into historical geographies, a gender relational explanation of global economic temporal development could now be engaged in. Such engagement aids historical geographers to understand the changing nature of gender orders over time and space. The sociologist Rob Connell has referred to such time–place geographies of gender relations as 'gender orders' evolving social relations of power, production, and emotional attachment that have influenced thought within feminist, historical geography. Historical geography now came to look at such topics as globalization, industrialization, and urbanization utilizing the idea of patriarchy as based in Marxist historical materialism whereby individual's thoughts and relationships derived from their gender positioning within relations of materialist production. Hence, debates mirrored those in sociology as to the extent to which capitalism and patriarchy remain single, dual, or intertwined systems.

The integration of ideas of patriarchy and power formulated by social theorists such as Doreen Massey has been highly influential in demonstrating how, historically, male power over women is embedded in patriarchy. This power is seen to derive from the place specific variations in the gendered spatial division of labor and relations of production. In an important text, considering the nineteenth century coal, linen, and agricultural industries in the United Kingdom, McDowell and Massey reinforced that place certainly does 'matter' when it comes to men and women accepting their societal position and relationships and that this acceptance is essential to industrial evolution. What such academics have also achieved, however, is to demonstrate how such historic place specific operations of gender relations have left a legacy which is visible to this day. For example, restructured Western industries such as mining and agriculture have left a legacy of gender relations. Working class men are now struggling to cope with the loss of their manual industry breadwinner role with the ensuing status and patriarchal power that accompanied it. The impact of such a historic legacy of geographic gender relations is being viewed today, therefore, with the aid of such gender perspectives which makes an understanding of past and present geographies of gender richer and deeper. Researchers are noting, for example, how men and women are now struggling to redefine their relational masculinities and femininities as capitalism in the twenty first century is taking advantage of the plethora of female, traditionally lower paid workers in the North of England and the valley towns of Wales. Thus, the legacy of Victorian industrialization is a present day ready supply of cheap feminized labor which is increasingly being drawn on in service industries. For example, call centers are purposely locating to capitalize on an abundant female workforces in areas of heavy industrial decline. Capitalism, it is suggested, relies on men and women accepting their gender differentiated work roles and the relations of gender which ensue and is derived from Victorian ideas of separate spheres of operation for men and women. Hence, historical geographers have reexamined the rise of the Western city, industrialization, and globalization.

Historical geographers utilizing epistemologies of gender have suggested, therefore, that the rise of industrialized gendered relations of production and consumption facilitated and promulgated Western industrialization. However, these areas remain open to future research. The 1999 book *Modern Historical Geographies*, while not specifically containing work on gender, provides chapters where present and future integration of gender research would be useful. Researchers acknowledge, for example, how little attention has been paid to historical geographies of the patterns of use of preindustrial cities or the impact of Victorian ideologies of the gendered separation of home and work on urban spatial planning. Therefore, exciting work is awaited by historical geographers now fully embracing gender as a key socioeconomic devisor and organizer. For example, the rise of industrialization has been considered to be dependent upon the shrinking of space by time in a factory system that required huge rural to urban migration and the reorganization of home and work. Victorian beliefs per taining to gender orders, reflected in ideals of the home as a feminized space of consumption, socialization, and respite from the rigors of capitalism have also been acknowledged. The separation of work from the home and the social ization of new capitalist workers and managers, it has been suggested, derived from a middle class gendered ideal largely dependent on women remaining in the home in an increasingly suburbanized city. The integration of ideas of space and place into women's histories is adding to an understanding of Western industrialization as requiring a middle class ideal of morality and home ownership where the production lines increasingly provided technologies and trappings for the home.

However, such accounts informed by theorization of gender relations became heavily criticized in the 1990s for relying on a historical geography where theorization was 'top down' and incorporated into a metanarrative where differentiation within and between women was lost. Hence, working class women could be said to have little in common with the middle class 'angel in the home' ideal and more in common with working class men. But what geographers continue to suggest is that whatever their social positions, as Western indus trialization took hold from the eighteenth century on wards, women experienced patriarchy, a loss of power, and that this was as a result of the social relations of production.

Postmodern Influences

The final set of theoretical influences upon historical geographies of gender occurred from the 1990s onwards and can be situated within the so called postmodern and spatial turn within social sciences generally. These new perspectives have had an influential impact on historical geographer's ways of knowing, asking, and writing. The whole notion of the concept of gender has come to be criticized as a 'performative' fiction where identities are mutually constructed through, for example, race or age. Theorists such as Judith Butler, influenced by psycho analysis, have contributed to new understandings of how geographical identities are formed and experienced. Thus, a fundamental shift occurred in the 1990s, both in the theorization of gender itself and its relevance to the his toricity of geographies of gender. Hence, sexuality as a matter of choice away from essentialist dualisms has de stabilized the concept of male/female genders. This has promoted a move toward situated knowledges away from objectivist epistemologies. Postcolonial geographers, for example, have been critical of the Eurocentric feminist imposition of colonial geographies. Such geographies fail, it is argued, to acknowledge the individual histories, stories, and acts of agency which resisted colonialism. Thus, a range of examples of how gender intertwined with such resistances, as, for example, within Bedouin women has been explored. Thus, the influence of post structuralist thought and the perceived breaking of ties with socioeconomic modernist, industrial structures of gender relations attached to specific places have given way to a flourish of research interested in destabilizing writing on geographies of the past written from the Western center. Such work traces both the impacts of the processes of colonization in the place of the colonizer but also identifies the agency and subversive powers of the colonized in the making and transformation of colonial societies. This research has drawn on new sources chal lenging the perceived masculinity of previous know ledges. Ways of knowing, it is contended, cannot be separated from ways of asking and writing. Hence, re search as process with an emphasis on the nonprioritiza tion of the researcher over researched has been promoted in collaborative projects aimed at revealing new geog raphies of the past in oral testimonies, diaries, and auto biographies thus resisting official and high cultural textual historical sources.

A challenge has ensued, therefore, for researchers in historical geography. This challenge poses itself as to the extent to which revealing patterns of gender inequities of the past situated within wider socioeconomic, spatial patterns, as outlined in the previous section, can be reconciled alongside a new view of past people as pos sessing their own historical situated knowledge's which are contingent upon context. Such a challenge is em bedded in work developed around postmodern gendered, cultural representations which revisit how we view the past according to the sources available. Since the 1990s, feminist geographical journals such as *Gender, Place and Culture* have been at the forefront in leading the post modern challenge of retheorizing capitalism and pro cesses of social class. Spatial metaphors are now the vogue as in between spaces and ideas of hybridity have largely replaced a two dimensional gendered view of past geographies as possessing processes and patterns ex perienced across the Western world.

The integration of the 'cultural turn' has been crucial to this new way of knowing and approaching gendered geographies of the past. Whereas the Berkeley School of cultural geography drew on the collision of social and cultural geography to describe patterns and cultural in fluences on the landscape, other historical geographers have been engaged with the contested nature of culture and its representation in a variety of texts. Thus, theo rization within a newly encultured geography recognizes the politics of representation. Culture is seen as a sig nifying system, whereby power relations are enacted and meanings contested. This has particular resonance for historical geographers working with concepts of gender identities. Hence, the historical geographies of landscape and nature and their representation have been interrogated for the ways in which they reflect ways of seeing. Historical landscapes, for example, from the sixteenth century onwards are deemed to be both physically and symbolically drenched with gendered symbolism which in turn shapes the physical and social world around us. This means that landscape art is not just a product, but that it aids a focus on the visual senses acting as a politics to gendered representations of the past. This, it is considered, impacts on representations and social realities of the present.

This culturally infused 'turn' has led to a reexamination of the Western landscape esthetic. Ideas of gender have been integrated into an arena where historical geographies have been influenced by gender perspectives. If a landscape is to be viewed, this view is deemed to have come from somewhere. Hence, the dominant gaze has been considered to be male. It is at this point that theorization again divides some researchers as they contend that women can, in fact, subvert the male gaze, particularly white women adopting a free and superior position in colonial lands. Other researchers, drawing on spatial metaphors of belonging and place, feel that the Western landscape esthetic reflects historical geographies where landscapes reflect masculine traits of prioritizing the visual, the lineal, and the feminization of landscape and nature. For example, gendered reading of Mary Kingsley's *Travels in West Africa* published in 1897 could be read as a partial view of the world around her from within the confines of ideas of feminine sensibilities where women lacked the freedom to be outdoors, to draw, and to paint openly. Or they could be said to offer a different view of the botanic world which subverts a male gaze.

But what historical geographers have emphasized is that at different times Western societies in different places have shared views about ways of seeing, organizing, and feeling about the environment based on ideas about gender. Thus, the much oft studied Gainsborough painting of Mr. and Mrs. Andrews's from 1789 has been read as a text, associating ideas of nature with women and ideas of culture with men. Thus, Mrs. Andrews is seated in an iconographic, submissive position reflecting her position as wife. Mr. Andrews, gun in hand, is positioned within the agrarian landscape he owns, projecting power and authority over the virgin Earth behind him. Mrs. Andrews has been viewed as being part of nature, a static figure, symbolizing fertility beneath an oak tree. Thus, it is contended, that this vision of the female from the male gaze continues to be reflected in contemporary art where the male gaze is privileged over the feminized 'other' by a male spectator.

The Future

Other exiting areas are beginning to be developed from this perspective where, for example, ideas of the body, imaginations, memory and heritage, and war are impacting on physical memorials to the past which draw on past and present understandings of male and female bodies and their relationships to society and economy. For example, we might like to consider where statues are located, how men and women are remembered, and if such representations really provide a view into the past gendered geographies of men and women. Heritage industries in Ireland have been noted to be particularly apt to situating statues and memorials of migration according to the perceived historical gender contributions and movement of men and women, while work in the rural arena is beginning to consider how gendered ideas of the past impact on present idealizations of rural countryside, economy, and society.

Historical geographies of gender remains an area, therefore, open to new and exciting explorations. Views of the present impacted by the past and connections over time and space are revealed and interrogated via a range of theoretical gender perspectives. While theoretical evolution may appear to be linear, in reality it is not such a neat progression. A variety of gender perspectives continue to debate ideas of gender roles, relations, and identities and the ways in which historical geographies are constituted. New topics aligned to these perspectives continue to be investigated in debates, for example, over the extent to which research into historical geographies of gender is feminist; can include men; or is transgendered. Discussions of an era of post feminism continue to be tempered by radical perspectives highlighting gendered legacies of industrial capitalism on the present day landscape, new patterns of global immigration and labor, along with the gendering of work roles. This remains, therefore, an area in human geography open to new debates and lenses of enquiry where immense scope to investigate the impact of the past on the present and connections across space and time exists.

See also: Critical Geography; Cultural Turn; Empire; Feminism/Feminist Geography; Feminist Groups within Geography; Feminist Methodologies; Finance, Historical Geographies of; Geography, History of; Hegemony; Hybridity; Landscape; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Masculinities; Medieval Historical Geographies; Nature; Patriarchy; Poststructuralism/Poststructuralist Geographies; Representation and Re-presentation; Sexuality.

Further Reading

Blunt, A. (1994). *Travel, Gender and Imperialism: Mary Kingsley and West Africa*. London: Guildford Press.

Butler, J. (1990). *Gender Trouble*. London: Routledge.
Connell, R. (2002). *Gender*. Oxford: Blackwell.
Domosh, M. (1991). Toward a feminist historiography of geography. *Transactions of the Institute of British Geographers* 16, 95 104.
Graham, B. and Nash, C. (eds.) (1999). *Modern Historical Geographies*. Harlow: Pearson Education Limited.
Humphries, J. (1990). Enclosures, common rights, and women: Proletarianization of families in the late eighteenth and early nineteenth centuries. *The Journal of Economic History* L(1), 17 42.
Livingstone, D. (1992). *The Geographical Tradition*. Oxford: Basil Blackwell.
Massey, D. (1984). *Spatial Divisions of Labour*. London: Macmillan.
McDowell, L. (1999). *Gender, Identity and Place: Understanding Feminist Geographies*. Oxford: Blackwell.
McDowell, L. and Massey, D. (1984). A woman's place. In Massey, D. & Allen, J. (eds.) *Geography Matters!*, pp 28 147. Cambridge: Cambridge University Press.
Rose, G. (1995). Tradition and paternity: Same difference? *Transactions of the Institute of British Geographers* 20, 414 416.
Stoddart, D. (1991). Do we need a feminist historiography of geography and if we do, what should it be? *Transactions of the Institute of British Geographers* 16, 484 487.
Trollope, J. (2006). *Britannia's Daughters: Women and the British Empire*. London: Pimlico.
Walby, S. (1990). *Theorising Patriarchy*. Oxford: Basil Blackwell.
Women and Geography Study Group (1997). *Feminist Geographies: Explorations in Diversity and Difference*. Harlow: Longman.

Relevant Websites

http://www.tandf.co.uk/journals/titles/0966369X.asp
Gender, Place and Culture.
http://www.rgs.org/OurWork/Research+and+Higher+Education/ResearchGroups/Research+Groups+N+ +Z/Women+and+Geography.htm
The Women and Geography Study Group.

Genealogy and Family History

S. M. Otterstrom, Brigham Young University, Provo, UT, USA

Glossary

Genealogist One who works to gather, organize, and compile names and records of related persons or families who are deceased.

Microfiche A small sheet of plastic film that has a large number of pages of printed matter that have been shrunk to fit onto the dimensions of the film. Microfiche sheets are read using special machines that magnify their image.

Microfilm The process and product of photographing records and shrinking them to a much-reduced size for more space-efficient storage. Microfilm rolls are viewed using a machine that magnifies their image.

Pedigree A listing of family lineage or ancestry beginning with a single individual.

Proxy Someone who acts on behalf of another.

Vital Records Data giving specifics concerning the births, marriages, and deaths of members of a population.

Introduction

Every person who has lived on this Earth has imprinted a personalized geographic history on the landscape. Each individual is born in a discrete location and dies in an exact place. These spatial markers can begin to tell the story of one inhabitant, but the narrative is much richer. With each additional shred of vital history the portrait of the studied becomes more detailed. Such data can be overtly geographical such as the location of homes and workplaces over the life course, place of marriage, and city of birth for children. But helpful biographical facts can also be informational such as recorded life experiences, occupational history, and personal and family names, including the names of the parents, spouse, and children. These additional clues concerning the character of the individual in question are often couched within a cultural context that encourages an investigation of the historical geographies that the person passed through in life.

This article explores how historical geography intersects with family history and genealogy. It outlines the development of the practice and the growth of seeking out the past geographies of deceased relatives over the past 160 years. It also shows how this cultural phenomenon has spawned its own human and spatial dimensions. Family history and genealogy are related terms. Family history is the documentation of the life events and experiences of individuals and families, while genealogy is often more narrowly focused on determining pedigree relationships using vital record data. In this article the two terms are used interchangeably.

Interest in genealogy is not new. Different cultures around the world have preserved the memory of their ancestors by the written word and orally. The Chinese with their jiapu system of recording family histories began many centuries before Christ. The Old Testament in the Bible contains recorded genealogies of many generations of the Israelites and the early patriarchs. Royal lineages in Japan, Europe, and elsewhere were also faithfully recorded through the middle ages. Additionally, oral traditions in parts of Africa and the islands of the Pacific have preserved the names of some forebears in these areas.

Notwithstanding the historical practices of recording family relationships by various means, the development of genealogical related activities into a growth industry with millions of hobbyists has been of relatively recent date. It was not until 1844 that the world's first genealogical society was established in Boston (the New England Historic and Genealogical Society). It was followed by the New York Genealogical and Biographical Society (1869) and the Genealogical Society of Utah (1894). The following year the Genealogical Society of Pennsylvania was organized. Since that time, a multitude of other societies have been founded around the world to promote the preservation of local genealogical records and historical records, or to strengthen the practice of genealogy among certain ethnic or racial groups. These historical societies are key resources for genealogists.

Popular Family History and Genealogy

Substantial interest in family history and genealogy activities is exhibited by a growing segment of the public that spends vast amounts of time in genealogical pursuits as a hobby or for religious purposes. Three broad types of people exist who use genealogy and family history data. The first is the hobbyist who pursues his or her own family history and pedigrees because of an abiding personal interest. These people range from the weekend enthusiast to the highly knowledgeable family expert. The second type of genealogist is the professional who is paid to research and trace other people's relatives. He or she can be certified by an accrediting organization, and may have a specialty such as Ohio (USA), English, or German records. The third group is the segment of historians, geographers, demographers, and other researchers who use genealogical data to explicitly study the past. This section looks at various facets of genealogy that the amateur and paid genealogist participate in along with the relationships that develop between people and places from different times. The following section focuses on academic research that reveals diverse historical geographies utilizing genealogical data.

The practice of family history and genealogy is intrinsically historical geography, because it necessarily deals with people within the context of their past experiences both socially and locationally. Those who pursue this avocation or career find themselves often referring to historical maps and gazetteers as they attempt to locate ancestors. For example, it matters to the genealogist that Du Page County in Illinois was created from Cook County in 1839, if they are looking for a progenitor who lived in that area before 1839. They also discover that knowledge of historical, economic, and societal conditions for their areas of interest can benefit their search. For instance, if an ancestor who lived in Cook County, Illinois in the 1840s disappears from the local records, it may help to know about the migratory tendencies during the period to nearby and distant western territories.

When a family history hobbyist discovers that an ancestor emigrated from a far off place there is often a desire to learn about the history and geography of the locale. They may ask a variety of questions such as: What did my ancestor do for a living? Did he play any important role in the local society? Why did he leave the community? Of his sisters or brothers, who stayed in the area and who migrated? They may also find out about the religious persuasion of the people there and the town government structure. In essence, while reconstructing the picture of the life of an ancestor who resided in some distant (or nearby) land, the family historian actually produces an individualized mini historical geography for one time period of a country, village, or city. A genealogist who combines many of those family biographies ends up with geographical and genealogical documentation that covers widely separated times and places.

Of course, a full life history on all ancestors is not a realistic goal. Often, little more than names and dates can be recovered from the incomplete records that fill the shelves of government and genealogical society offices. In those cases, the genealogist has to find some measure of contentment in the information that is available even while hoping for more to materialize. Generally, the further back in time one reaches, the greater the likelihood that only partial data on an ancestor will be found, if anything is to be found at all. However, at some fortunate junctures a progenitor is discovered who was either famous or infamous, and who therefore warranted more than a cursory notation in the local church registers. At other times, a family Bible or a pertinent local history with many generations of names turns up to aid the genealogical enthusiast. Thus the process of constructing a personal ancestral historical geography is, by its nature, always a work in progress whose spatial and historical limits are continually being tested and refined.

As genealogists extend their temporal record of family relationships they often develop the desire to travel to their ancestors' countries, towns, and neighborhoods in search of records and/or to enhance or create an emotional attachment to an ancestor through being exposed to the same physical environment where the progenitor lived, although time delayed. This is true even though exponential growth of online historical records databases has reduced the need to be on site to gather family data. The reasons for this phenomenon are twofold. First, not all local records are accessible on the Internet. Second, increased knowledge about an ancestor spawns curiosity about his home and environment. As a result, a subtype of heritage tourism has developed where enthusiastic descendants of often obscure citizens in little known villages flock to these places of ancestral origin to see what remains of the geography that existed in the past life of the relative. Local government courthouses, libraries, churches, and cemeteries are some of the places where these tourists visit, which set of places contrasts markedly from the itinerary of a more traditional visitor.

Another way that people celebrate and strengthen their familial connection with their ancestors is the formation of family societies. These organizations usually garner membership by recruiting the descendants of one prominent progenitor or couple. Activities of these groups often include periodic family reunions, family newsletters, and book writing projects. In the process, a shared identity is developed among cousins, aunts, uncles, grandparents, and siblings. In some cases, families may gather near the birthplace of a particular ancestor or near his or her childhood home. Intense research into the lives of common ancestors can occur so that very detailed descriptions of past life geographies emerge. Because of natural increase, over the years the family organizations have ever increasing numbers of potential members, people that have a direct family tie who can in turn rediscover the historical geography of someone who matters to them. The membership geography of such societies also shows how much the descendants of particular people have dispersed across the countryside.

Data sources for family historians and genealogists are diverse but are almost always place specific. This association with geographic locales is crucial because the place attaches a marker to a person that helps identify them positively. Thus a successful search through disparate sources, such as censuses, cemetery indices, probate and other court records, county history biographies, newspaper indices, immigration lists, newspaper obituary notices, land records, voter lists, church membership rolls, ship passenger lists, county vital records, and other historical records, requires one to proceed with some knowledge or even an educated assumption regarding ancestor location. Consequently, even a beginning genealogist soon learns that geography is inextricably connected to their ancestors, and that knowledge of these past geographies increases their potential for success in finding the ancestor in question.

Besides the millions of genealogical hobbyists, there is one group of people who pursue the accurate collection of their family records and ancestors with a religious objective. Members of the Church of Jesus Christ of Latter day Saints (whose members are commonly called Mormons, or Latter day Saints) have been interested in preserving pedigrees and other genealogical records since soon after the church's founding. Beginning in the 1840s in Nauvoo, Illinois where the church was headquartered, members were taught by their church president, Joseph Smith, that they should be baptized in proxy for relatives who had died without being baptized. This belief was tied to both Jesus' teaching in the Bible (John 3:5) that all should be baptized and to Paul's words (1 Corinthians 15:29) when he discussed the practice of baptizing for the departed in New Testament times.

Feeling a religious responsibility to perform church ordinances for their deceased ancestors, Mormons have developed one of the most extensive genealogical record depository and library systems in the world. Thus the Genealogical Society of Utah was established (1894) to help Latter day Saints in Utah seek out their own genealogies. Not only was it inaugurated early in the diffusion of genealogical societies, but the Utah society, because of its religious orientation, has sought to collect records from all over the United States and world, in contrast to many other genealogical societies that are more localized in their focus.

At first, the Genealogical Society of Utah library was only open to paying members, but after many years the church incorporated it into its administrative structure (1944), and the library's holdings became available to the public. Subsequent growth of this genealogical library in Salt Lake City has been significant. Now called the Family History Library, it has over 742 000 microfiche, 310 000 books, and 4500 periodicals, and 2.4 million rolls of microfilm, with some 4100 rolls of microfilm and 700 books and 16 electronic resources being added to the collection every month (in 2003). Much of the ongoing growth in microfilm collections comes from the efforts of church missionaries and employees who have traveled to over 110 countries, territories, and possessions, to film vital records and other historical documents.

Besides the main Family History Library in Salt Lake, the Church has established some 4500 branch family history libraries around the world. Most of these libraries are located within Mormon church meeting houses, and are staffed by volunteers from the local church membership. They are open to church members and the public, and these patrons are assisted by librarians in both doing basic and more advanced research. One can also order any one of the microfilms from the main library in Salt Lake City, and search them using the microfilm reader in the local Church family history library.

Historical Geography and Genealogy

Many academic avenues tied to historical geography can be explored related to family history and genealogy. The potential of applying geographic approaches to historical problems using genealogical data has been recognized for some time. As technology and record availability have radically improved over the last few decades, even more opportunities for understanding past people and regions are now available. The topics that are studied in this area fall into two basic categories. The first is research in the historical geography of migration, other demographic factors, kinship ties, and the cultural markers of personal names. The second topic is the exploration of the historical geographies related to the process and cultural implications of the practice of genealogy and family history.

Geographers and other academics usually have a broader goal in mind than the genealogical hobbyist. They want to reveal larger social and environmental processes, while a genealogist is interested in only certain limited groups of people, namely sought for ancestors. This difference in approach underscores the fact that genealogical and family history data sets must usually be reconfigured to meet the needs of academics before they can be successfully used. However, once the limitations of incomplete and incorrect genealogical associations are recognized and accounted for and the data reformatted, genealogical records can be very powerful resources for those interested in reconstructing past places.

Historical migration researchers have used a variety of data sources to depict patterns of past population movements, but tracking these migrations has been difficult because few primary resources have sufficient information to show more than general migratory trends. For example, beginning in 1850 the United States Census started recording state or country of birth for each individual in their tabulations. These data were helpful for those interested in knowing the birth origins of populations, but the data do not reveal where the person might have lived in the intervening years between birth and the census. That is where efforts to match people between censuses or between other data from other years have been used. The matching exercise is very tedious and time consuming, even with modern technological ad vances, so the potential appeal of using extant geneal ogies to illustrate and analyze migration is apparent.

Although genealogical data sets tend to be narrower in scope than matched records between census years, they can show multigenerational migration streams. This is not possible with other methods (see **Figure 1**). With pedigrees that are now available digitally, the migration patterns of multiple people who settled in one area and their descendants can potentially be mapped and ana lyzed. As multiple genealogical data sets that contain birth, marriage, and death dates and locations are com bined and reordered to reflect temporal, relational, and spatial order a powerful analytical tool will result. The potential of such an effort to describe historical migration streams across the sea and over land, and to extrapolate larger general patterns is great.

Other historical geographies tied to demographic characteristics can also be constructed. A related factor in migration research is the role that kinship plays in whether one decides to migrate. For example, geneal ogies can contain data that show multiple siblings moving to the same neighboring county or emigrating to the same distant country. Additionally, these same relational ties could help explain why extended family members did not migrate in some communities. For those who study fertility, marriage, mortality, and family structure among other demographic factors published genealogies of certain communities and families can depict a detailed local portrait of life events that are often missed through the traditional reliance on other historical sources such as censuses and local church birth and death records. When using genealogical data with migration research, poten tial statistical errors and biases (e.g., omissions of children who died young, or false family links) must be taken into account when applying population analysis methods to genealogical tabulations.

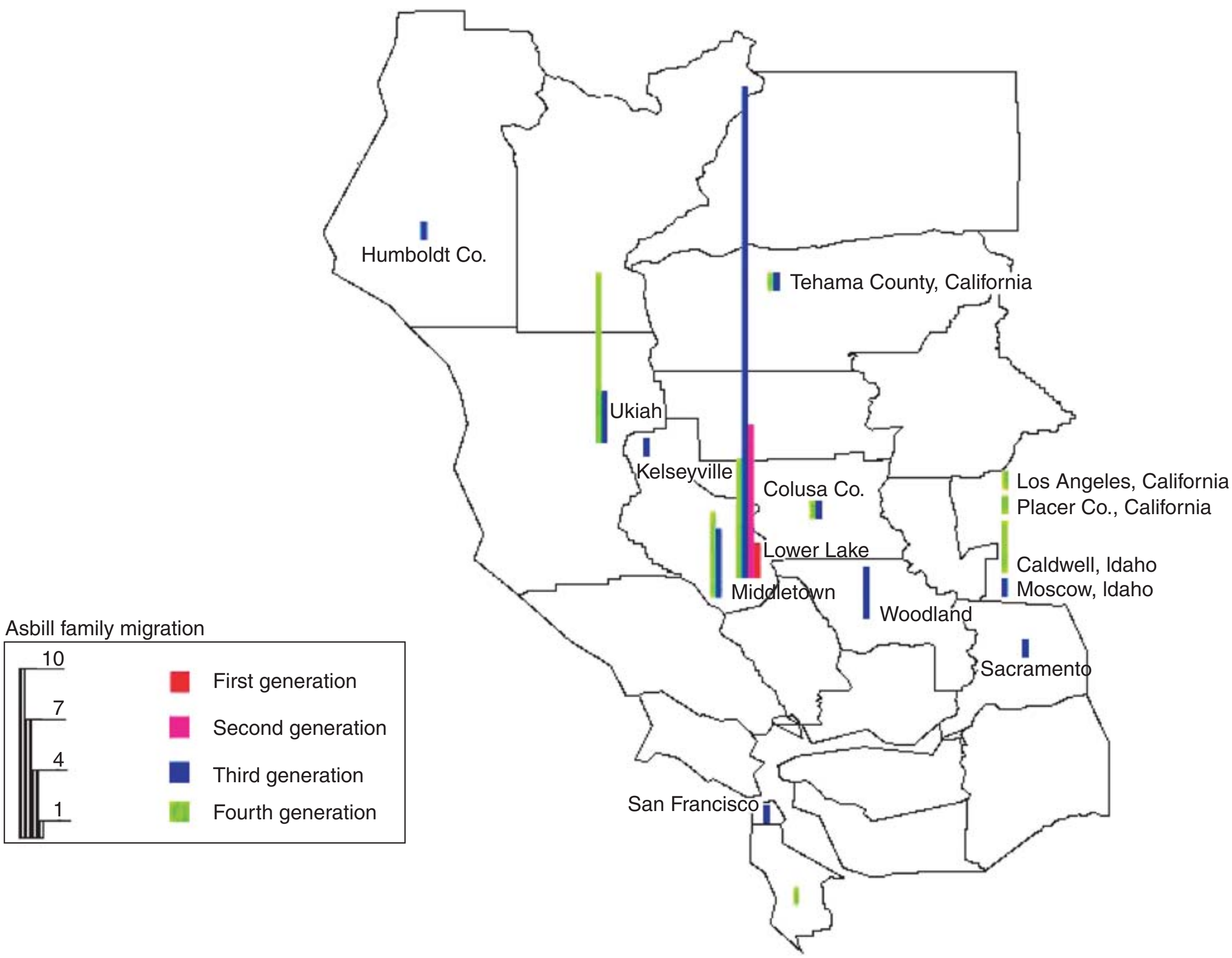

Figure 1 Locations of four generations of Asbill family descendants in northern California using genealogical data.

Closely allied with migration studies is the utility of name data to show the diffusion of certain family or ethnic groups. For example, tracking the spread or diaspora of individual surnames of French origin, such as Thibodaux, across the state of Louisiana in the United States can show the multiplication of family descendants and their spatial spread, especially when the name can be traced to one or two original settlers in a region. Additionally, certain surnames can be studied to find their locational origin. Depending on the type of surname, this research approach can be especially helpful in illustrating the historical geography of family clan areas (e.g., the MacLeans in Scotland), or physical geography features at the point of origin (e.g., Nordstrom (north stream)) in Sweden, or the occupational structure of the earliest people in those source areas who had these names (e.g., Fisher, Shoemaker, Schmidt), although the first bearers of these names may be lost from recorded memory.

Studies of the experiences of people attempting to connect with their past through genealogy is the second major area of historical geography research. Some of the issues that can be explored in this vein may include the experiences themselves of those participating in heritage tourism activities; how collective memories constructed by genealogists shape the accepted historical geographies of past people; the process of people seeking a particular ancestral identity through genealogical research and the effect that has on their personal cultural conception; and the spatial genealogical interconnections that have arisen from the fertile combination of ongoing technological innovations with an active cohort of family historians eager to interact with others holding information on a common family relation. Additionally, with the growth of heritage tourism tied to the locations of ancestors, communities, small and large, with long histories and rich historical records can be identified and even encouraged to take advantage of their potential attraction to traveling genealogists.

Effects and Geographic Ramifications of New Technologies on Family History Research

A dizzying explosion in the digital and other technological areas has expanded the research possibilities and accessible resources for genealogists and historical geographers studying family history. From the advent of the earliest genealogical societies in the mid 1840s to the early spread of the personal computer, tracing ancestral ties was a very slow and laborious process. Letters had to be written or phone calls made to distant relatives, historical archives, family societies, or government offices in search of tidbits of family information. Long waiting was often a requisite to obtaining new and helpful records. Large paper indices of censuses or microfilms had to be scoured in the hopes of finding the correct relative. It often required a large dose of patience on the part of the researcher.

All of these slow methods are still used and have value, but computerized innovations, especially the Internet, have spawned a tremendous surge in digitally available resources and powerful online search mechanisms. Complete censuses of the United States and other countries have been digitized and indexed, thus drastically cutting the need to consult the original microfilmed records. Censuses are just one of a multiplying list of sources that are being made available to the general public, both freely and in some cases only to a paying membership base. Genealogical societies, government agencies, for profit companies, family organizations, and enthusiastic individuals are all involved in expanding this data bank. Practically any type of primary genealogy data that one can name is found on the Internet in one form or another, ranging from scanned newspapers from the 1800s or before to obituaries published only days before a search. Of course, availability of different types of data varies by location, but new sources are being added daily. Additionally, whole websites are dedicated to the relations of certain family groups or individuals where many people with common research interests can interact and cooperate successfully. So too can one search for specific living relatives around the world using a variety of online databases. The result is an international collection of individuals collaborating across space to better reconstruct the historical geographies of shared ancestors.

Digital historical maps and gazetteers have also added a beneficial research resource to genealogists and historical geographers working with family history data. Old scanned maps provide a wealth of knowledge about the past environment that can give family historians clues to the whereabouts of the people they are seeking. Digital mapping projects of county, state, and other historical boundaries have made it much simpler for one to see what governmental jurisdictions used to cover a progenitor's home. So also have benefits come from online gazetteers that have links to previous place names and their period of use formatted in an easily searchable platform.

An example of the great push to increase record availability digitally is that of the Mormons. With the rapid development of the Internet, the Church of Jesus Christ of Latter day Saints has constructed a free website that gives the public access to hundreds of millions of names of people who have died (FamilySearch.org). One of the distinct searchable databases that is available at the site is the International Genealogical Index (IGI). The IGI is essentially the complete database of names (with dates, places, and family relationships) of people who have died and who were Mormons or the dead whose ordinances have been performed for them by proxy in one of the church's temples. The website now has over 1 billion names that can be searched for possible family ties.

Not only have the Mormons made available to the public their digital records through their website, and their genealogical microfilms through their system of family history libraries, they are now working to put their microfilms and other primary records online for more efficient family history searches. Their church owns the largest collection of genealogical data in the world, and their microfilm records contain over a billion names copied from censuses, probate records, church records, and other historical documents. The original films are stored in the Granite Mountain Records Vault, which was constructed by blasting tunnels into the side of a canyon east of Salt Lake City. The church's end goal is to have their 2.4 million plus rolls of microfilm records scanned, indexed, and then placed on their Family Search website.

Efforts to connect people with their ancestors using genetic evidence instead of written records have been growing. For example, adopted children searching for their parents can use DNA tests to determine whether someone who they have made contact with is actually their father or mother. Additionally, these tests can also show if two people with the same last name share a common ancestor. At a more expansive scale, projects tracking the historical migrations of people using genetic markers have revealed important geographic connections among various cultures and time periods around the world. These large scale efforts are less interested in tying people to their direct ancestors by name than the cases where direct blood relatedness is in question. Shared genetic markers are tracked on the straight paternal line using Y chromosome DNA tests, while direct maternal ancestors are followed using mitochondrial DNA tests. Tracing one's paternal or maternal ancestry using key genetic markers ties their unique DNA to certain hearths in the distant past allowing a DNA link into the main human migration stream. Research collecting this type of data includes the National Geographic Society's 'Genographic Project', which initiative is working to illustrate historical migrations over the millennia using a large collection of genetic data.

Genealogy and Historical Geography and Its Future

Rapid developments in computer and Internet technology have combined to transform genealogy and family history research. The ability of genealogists and researchers to use computers to quickly search through a growing collection of names and places has accelerated the pace of successful searching and has more efficiently united far flung individuals interested in common ties. Thus, multiple matching criteria have made searching for specific ancestors a faster process and have greatly opened the avenues for population related research by historical geographers.

There are two main directions that genealogists track relatives of individuals: back through time with ever expanding cohorts of ancestors, and forward in time from a progenitor and his multiplying descendants. Regardless of temporal direction of the search, location and place are vital to the success of this type of detective work. Thus, historical geography both adds local context to the everyday life experienced by deceased relatives and provides clues concerning the possible movements of these people across regional and national borders. As the popularity of genealogy grows, and as digital resources and search technologies multiply, the natural overlap that exists between historical geography and family history will most likely catalyze greater success in both arenas of endeavor.

See also: Demography; Emigration; Heritage and Identity; Immigration II; Migration.

Further Reading

Bieder, R. E. (1973). Kinship as a factor in migration. *Journal of Marriage and the Family* 35, 429 439.

Dennis, W. J. and Lynch, K. A. (1982). Genealogies and population registers. In Dennis, W. J. & Lynch, K. A. (eds.) *Sources and methods of historical demography*, pp 109 131. New York: Academic Press.

Dyke, B. and Morrill, W. T. (eds.) (1980). *Genealogical demography*. New York: Academic Press.

Earle, C. V., Otterstrom, S. M., and Heppen, J. (1999). *HUSCO 1790 1999*. (Historical United States County Boundary Files on CD Rom). Baton Rouge: Geoscience Publications, Louisiana State University.

Haley, A. (1977). *Roots*. New York: Dell.

McWhiney, G. and McDonald, F. (1983). Celtic names in the Antebellum Southern United States. *Names* 31, 89 102.

Macpherson, A. G. (1984). Migration fields in a traditional Highland community, 1350 1850. *Journal of Historical Geography* 10, 1 14.

Morgan, P. (1990). Locative surnames in Wales: A preliminary list. *Nomina* 14, 17 23.

Nash, C. (2002). Genealogical identities. *Environment and Planning D: Society and Space* 20, 27 52.

Smolenyak, M. and Turner, A. (2004). *Trace your roots with DNA: Using genetic tests to explore your family tree*. Emmaus, PA: Rodale.

Timothy, D. J. and Guelke, J. K. (eds.) (2008). *Geography and genealogy*: *Locating personal pasts*. Aldershot, England: Ashgate.

West, R. C. (1986). *An atlas of Louisiana surnames of French and Spanish origin*. Baton Rouge: Geoscience Publications, Louisiana State University.

Widdis, R. W. (1998). *With scarcely a ripple: Anglo Canadian migration into the United States and Western Canada, 1880 1920*. Montreal: McGill Queen's University Press.

Zhao, Z. (2001). Chinese genealogies as a source for demographic research: A further assessment of their reliability and biases. *Population Studies* 55, 181 193.

Relevant Websites

http://www.ancestry.com
Ancestry.com.

http://www.cyndislist.com
Cyndi's List of Genealogy Sites on the Internet.

http://www.familysearch.org
Family Search.
http://www.gendir.com
Genealogy Directory and Message Board.
http://www.onegreatfamily.com
One Great Family.
https://www3.nationalgeographic.com
The Genographic Project, A Landmark Study of the Human Journey, National Geographic.
http://www.usgenweb.org
US GenWeb Project.
http://www.werelate.org
We Relate.

Genealogy Method

Ú. Crowley, National University of Ireland, Maynooth, Co Kildare, Ireland

Glossary

Discourse A social boundary defining what can be said and what cannot be said.

Episteme Systems of thought and knowledge (or conceptual frameworks) which define a system of conceptual possibilities that determine the boundaries in a given period.

Subjugated Knowledge Knowledge that has been discounted, disqualified, or low ranking.

Situating Genealogy

In order to situate and better understand genealogy as method, this article briefly considers some of Michel Foucault's early works. When engaging in such a periodization or categorization, however, it is important to stress that Foucault himself regarded his work as intensely unified, part of a single enterprise, though marked by transitions.

The first period of Foucault's published works deals with a series of historical case studies that are mainly concerned with the emergence of a range of modern human sciences. In *Madness and Civilisation* (1961) (the history of the asylum and the emergence of the modern concept of mental illness), *The Birth of the Clinic* (1963) (a critique of modern clinical medicine), and *The Order of Things* (1966) (an examination of the evolution of thought, reason, and unreason), Foucault introduced a method of analysis he referred to as archaeology. *The Archaeology of Knowledge* (1969) provides a methodological summation of archaeology as method and a challenge to the ways in which knowledges are traditionally analyzed in the human sciences.

Archaeology is a process for working through the archives of a society (parliamentary debates, prison records, chronicles, diaries, journals, logbooks, official records, grand theories, popular knowledge, subjugated knowledge, and so on) and, like genealogy, is concerned with 'the history of systems of thought'; the history of societal structures (or epistemes in Foucault's terminology) that have produced and shaped the boundaries of knowledge, ideas, truths, representations, and discursive formations in different historical periods. Archaeology as method isolates and deconstructs components of accepted knowledge. It exposes the randomness of interpretation, the ordered procedures that made discourses possible and what conditions their unity by providing alternative accounts and uncovering popular knowledges, local beliefs, and understandings that traditional history has disqualified. Foucault was not so much concerned with determining whether the knowledge systems of the human sciences were true but rather with contextualizing and historicizing notions of truth, knowledge, and rationality. He examined their conditions of emergence, how and why a society in a given era considers some things knowledge, how and why some procedures are judged rational and others not. Foucault was interested in tracing out an historical ontology of the individual in relation to concepts of truth, for it is through such concepts that people constitute themselves as subjects of knowledge.

In *Discipline and Punish* (1978) and *The History of Sexuality* (1979) (historical critiques of the discourses of criminology and sexuality) Foucault rethinks and refines his analytical method, shifting from an attempt to develop a theory of rule governed systems of discourse to a more explicit focus on power, knowledge, and the body. Foucault deploys Friedrich Nietzsche's genealogical analysis of the development of morals through power as a starting point for developing a method that includes an examination of complex power relations between institutional practices, bodies, and systems of thought. This more refined analytical method is not a replacement method for archaeology – the analysis of discourse, the demonstration of discontinuity and shifts, and the rejection of a totalizing view of history remains central to the genealogical analytic – rather archaeology and genealogy are two halves of a complementary approach, alternating and supporting each other.

Genealogical Analytics

Discipline and Punish and *The History of Sexuality Volume 1* demonstrate Foucault's most thorough application of his genealogical analytics. In these works Foucault redefines the problematic of power and marks a fundamental break with conventional theorizations. Foucault was particularly interested in the way in which power operates through the construction of particular 'knowledges' and argued that it is through discourse that power/knowledge is realized. Conceptions of truth and knowledge are fundamentally products of power. This combination of power/knowledge and the embedding of reason diffuses through the social body producing what people are and what they can do, structuring the ways things are thought about, how people see themselves and others, and how they relate to the world around them. There is, however, no general theory of the character of power in Foucault's genealogies or indeed any attempt to provide one. He instead provides a tool kit for the analysis of power relations, writing 'for users, not readers'.

In traditional conceptions of power, power is ascribed to and exercised by agents. Power is equated with domination, cause and effect, oppressed and oppressor. For Foucault, however, there is no binary and all encompassing opposition between rulers and ruled at the root of power relations, rather it is the multiple of relations that take shape and come into force throughout society. Power is not an institution, a structure, a strength people are endowed with, something that can be held or transmitted. Power is not simply imposed in a repressive way from the top of a visibly demarcated social hierarchy, through the state or major dominations (e.g., religious authorities, dictators, and so on) in society.

Power works from diverse places and at different levels, working in and through space, among individuals, families, social groups, in institutions, simultaneously constraining and enabling. Power is incessant, constant, and wholly relational. That power is permanent and self producing is simply the overall effect that emerges from the multiplicity of force relations. For Foucault the sovereignty of the state, of legislation, of rule in society is not a 'given', something that has been there from the outset; rather they are simply the 'terminal form power takes'.

Within a genealogical perspective power is viewed as much more diverse, subtle, and complex than the negative and transparent relations of oppression, prohibition, and censorship (it is important to note, however, that Foucault does not reject the occurrence of repression but does reject its theoretical primacy). Power is productive; it produces subjects, it makes things happen, and achieves outcomes. Power operates discretely and subtly, as well as ambiguously and through ostensibly freely adopted practices, determining individuals' behavior not simply by coercion or repression but rather by controlling individuals' decisions to behave. Power does not have a single identifiable point; the power is the network itself.

Discipline and Punish

In *Discipline and Punish* Foucault maps the historical transformation in the exercise of power (and by inference the use of space) through an examination of the underlying principles, aims, and organization of the prison. It charts the changes in theory, practice, and social function of disciplinary institutions and practices that emerged in Western society in the eighteenth and nineteenth centuries. In particular, it provides an insight into how space is used politically in relation to deviant and nonconforming others and focuses on a microphysics of power where bodies are targeted as the site in which the most mundane social practices and local power relations feed into the constitution of large scale social and institutional power relations. Much of this book, however, is not about the punishment of criminals rather it is concerned with new modalities of regulation, with surveillance, control, and policing of modern Western society.

Foucault opens his genealogy of the penal system with a description of the public torturing, dismemberment, and execution of a regicide, Damiens, in Paris in 1757. The detailed, shocking, and gruesome contemporary account of his torture and execution is immediately followed by a description of a meticulous, tedious, and regimented routine, where young offenders are subjected to constant and pervasive supervision and observation in a reformatory in Paris 80 years later. These very different methods of punishment frame the parameters of Foucault's method, namely to explore the implications of the change from the feudal regime of power based on collective fear and built on violence and spectacle to one that is more subtle and gentler, regulating and disciplining individuals through procedures of observation, judgment, standardization, calculated punishment, or correction of perceived behavioral abnormalities. This transformation from feudal forms of power to disciplinary power in the eighteenth century is epitomized in the Panopticon (as designed by Jeremy Bentham), a 12 sided polygon prison building with a central tower through which the supervisor (unobserved) can observe and monitor the behavior of inmates. The Panopticon was designed to effect the moral transformation of its inmates by carefully controlling time, space, and bodies. Although the Panopticon was never built, Foucault argues that its conceptual construction and operation reveal a lot about the use of space and nature of social control in modern Western society.

Foucault was particularly interested in the transformation in the scale and continuity of the exercise of power, the involvement of much greater knowledge of detail, and the uninterrupted constraints imposed in practices and processes of discipline and training. While acknowledging the element of genuine progressive reform during this period of transition, Foucault mocked as superficial supposedly enlightened views that purported to show the emergence in the West of a humane and liberating reason. Instead of uninterrupted progress toward an increasingly caring society Foucault identifies differing styles and practices of social control in response to changing socioeconomic circumstances. He argued that the complicity of power/knowledge created a plethora of experts (doctors, social workers, probation officers, and so on) and institutions of discipline – prisons, schools, factories, hospitals, asylums, and so on, which though often promoted in the name of improvement, in reality consolidated and legitimated administrative authority and bureaucratic regulation. This new model of discipline became a model of control for an entire society and applicable to all forms of social governance.

A crucial element of Foucault's account of disciplinary power is that it functions positively (and also oppressively) to shape individuals as particular subjects. Power through torture, blood sanctions, and executions destroys their target while power through discipline and training targets the soul, and the submission of bodies through the control of ideas reconstructs its subjects to produce new actions, habits, and skills and ultimately new kinds of people, reformed, normalized, rehabilitated, and prepared to return to society. Discipline and surveillance is interiorized to the point that each individual is his or her own overseer. For Foucault it is the formation of subjectivity that constitutes the whole rationale of disciplinary and punishment techniques.

The History of Sexuality

Foucault's other major genealogy, *The History of Sexuality*, commences an inquiry into the conditions of emergence and success of sexuality as an organized discourse or apparatus for constructing and controlling human subjects – a strategy for directing social relations, classifying and examining bodies, authorizing and legitimizing specialized knowledges and experts (e.g., medicine, psychotherapists, psychiatrists, and so on) and multiplying sexual perversions (the masturbating child, homosexuality, nymphomania). *The History of Sexuality* keeps continuity with the microphysics of power, the concerns of governments and societies, and a focus on the relationship between power and knowledge. This work, however, places less emphasis on direct surveillance and the creation of docile bodies through discipline and provides a greater insight into how individuals' internalize norms laid down by the social sciences and truth discourses. The emphasis is on how people participate in their own subjectification and come to recognize themselves as particular types of subjects, judging and monitoring themselves and others in an attempt to conform to the norms of society. It is this focus on the body as a site of self scrutinization, self regulation, and self subjectification; as a site where both power and struggles are enacted that opens up a role for resistance that was not made explicit in his genealogy of the prison.

In *The History of Sexuality* Foucault attacks the 'repressive hypothesis'; the common thesis that sexuality was repressed in Western society (particularly during the nineteenth and early twentieth centuries) and argues that on the contrary there was a fixation with sex. He charts how governmental practices came to explicitly focus on sexual morality as a key target of social intervention, organization, and normalization. During this period there was, in Foucault's words, a 'political, economic, and technical incitement to talk about sex' – an explosion of discussion about sex (including both social and criminal acts). There was a flood of reports, policy documents, medical guidebooks, religious treatises, and legislation concerning sexual and moral matters, the emergence of social purity organisations, a strengthening of the network of institutions dealing with sexual 'deviants' (penitentiaries, reformatory schools, asylums, prisons, and so on). Illegitimate sexualities would only be tolerated in specified locales – the brothel, the mental asylum.

The claims to truth inherent in these discourses (e.g., the Catholic Church claimed an infallibility in regard to its moral teaching that was binding under pain of sin) offered knowledge to people about how to behave, how to be moral, how to care for themselves and for others. Subjects were increasingly educated to make themselves the subject of their own gaze, constantly monitoring theirs and others' bodies, behaviors, actions, and feelings and thus perpetuating and reproducing the authority of the regime's discourses. What materialized, for the first time in modern history, was an intimate symbiosis between government and civil society. It involved a move from government by the state to self government, self regulation, and self discipline.

The basic thesis of Foucault's genealogy on sexuality is that sexuality is not a natural reality but a creation of a system of discourses and practices that form part of the increasing surveillance, regulation, and control of individuals in modern society.

If, however, *Discipline and Punish* and *The History of Sexuality* are interpreted in too simplistic, rigid, or extreme a fashion, one risks connecting and conflating power with domination in the very manner Foucault was questioning. His historical exploration of the prison and sexuality is merely an interpretive perspective, offering insights into our social worlds. The basic disciplinary logic (practices of normalization and internalization of discipline) materializes differently and unevenly in different circumstances and is always only imperfectly realized in practice. Foucault's conception of a disciplinary society should not be mistaken for a disciplined society where discipline is seen as an insidious, monolithic force enveloping all relations, where power has a repressive and relentless hold on its subjects, where resistance seems futile, and citizens conform to the demands placed upon them. Foucault was interested in the techniques of power more than in the nature of power and rather than viewing power as something to be overthrown, he provides the conceptual tools to utilize, exploit, and transform power. Genealogy as method is a mind tool – a tool to stimulate thinking and cognitive curiosity.

Genealogy in Practice

Over the last two decades geographers and spatial theorists have been influenced by Foucault and have used his works as reference points for further theoretical arguments. His writings have been particularly significant for those critical human geographers seeking to define power, mainly the way it is played out in and through space and the importance of disciplinary techniques (both discursive and nondiscursive) of the body in the production and deployment of power has been comprehensively elaborated. Unlike much of Foucault's work, which is almost completely deficient in empirical material to support his theorizations, these writers use detailed case studies, policy documents, censuses, and so on to support their hypotheses.

More recently, human geographers and others have begun to explore ways in which personal details are stored, checked, transmitted, and used as a means of influencing and managing individuals and populations. Many of our everyday activities are, as in the Panopticon, regulated as much by the 'threat' of observation as by 'actual' observation, for example, surveillance cameras, police check points, threat of a tax audit, not carrying a licence (failure to be readily identifiable), being caught without car insurance, and so on.

Geographers and social theorists have also drawn on Foucault's genealogical method of inquiry in order to elucidate ways in which discourse, power, and knowledge have come together in different spaces (including, e.g., schools, industrial settings, offices, neighborhoods, community centers, hospitals) to mould human subjectivity and morality. His work has also appealed to academics interested in the historical geographies of marginalized groups (such as Irish Travellers, Native Americans), how these groups have been historically constructed through authoritative discourses and scientific practices.

The History of Sexuality fundamentally challenged traditional frameworks for understanding sexuality and has opened up the field to critical analysis, assessment, problematization, and development and has been enormously significant in shaping recent thought about the relationship between sexuality and society. Feminists and queer theorists have drawn extensively on Foucault's genealogies to extend notions of sexuality, to uncover silences and to further challenge orthodoxy, to destabilize and denaturalize traditional histories, categories, and theories. In particular, writers have complimented and extended the theoretical organization and methodological perspective of genealogy in order to examine how discursive practices and processes turn notions of sexuality and sexual categories (heterosexual, gay, lesbian, bisexual, and so on) into objects of study and analysis. Over the last two decades geographers have also drawn on and modified Foucault's method to dispute heteronormative assumptions, to examine issues of sexual citizenship, explore how externally enforced regulations come to be internalized, and to expose how meanings and choices surrounding sexuality are fluid, ever changing, unstable, diverse, and complex and structured through a myriad of discourses, spaces, and social relations.

See also: Body, The; Queer Theory/Queer Geographies; Sexuality; Surveillance.

Further Reading

Butler, J. (1999). *Gender trouble*. New York: Routledge Press.

Dean, M. (1994). *Critical and effective histories: Foucault's method and historical sociology*. London: Routledge.

Dodge, M. and Kitchin, R. (2005). Code and the transduction of space. *Annals of the Association of American Geographers* 95(1), 162 180.

Elden, S. (2001). *Mapping the present*. London: Continuum.

Foucault, M. (1978). *The history of sexuality: An introduction*. Hardmonstown: Penguin.

Foucault, M. (1979). *Discipline and punish: The birth of the prison*. London: Allen Lane.

Halperin, D. (1995). *Saint Foucault: Towards a gay iconography*. New York: Oxford University Press.

Hannah, M. (1993). Space and social control in the administration of the Oglala Lakota ("Sioux"), 1871 1879. *Journal of Historical Geography* 19(4), 412 432.

Jackson, N. (1998). Labour as Dressage. In McKinlay, A. & Starkey, K. (eds.) *Foucault, management and organization theory*. London: Sage.

Kitchin, R. and Lysaght, K. (2003). Hetrosexism and the geographies of everyday life in Belfast, Northern Ireland. *Environment and Planning A* 35, 489 510.

McNay, L. (1994). *Foucault: A critical introduction*. Cambridge: Polity Press.

Miller, P. (1987). *Domination and power*. London: Routledge and Kegan Paul.

Prada, C. G. (1995). *Starting with Foucault: An introduction to genealogy*. Oxford: Westview Press.

Rabinow, P. (ed.) (1984). *The Foucault reader*. New York: Pantheon.

Valentine, G. (1993). (Hetero)sexing space: Lesbian perceptions and experiences of everyday spaces. *Environment and Planning D: Society and Space* 11(4), 395 413.

Relevant Websites

http://www.foucault studies.com
Foucault Studies.

http://www.foucaultsociety.org/resources
The Foucault Society Resources.

Generalization

W. A. Mackaness and O. Z. Chaudhry, The University of Edinburgh, Edinburgh, UK

Glossary

Cartographic Generalization A range of techniques applied to ensure optimal portrayal of geographic phenomena according to task (theme and scale).

Cartometric Analysis A set of spatial analysis tools used to make explicit and measure metric, topological, and shape characteristics within and among a set of geographic phenomena.

Map Generalization A process of effectively portraying changing levels of detail among geographic phenomena in order to reveal their various properties.

Model Generalization Selection, simplification, and aggregation of features stored in a database for improved processing, dataset integration, and prior to cartographic generalization.

MRDB A database used to store multiple representations of geographic phenomena (created via the model generalization process).

Scale, Pattern, and Geographic Process

It is not possible to describe a geographic process without reference to ideas of scale. The scale of observation is critical to the discernment of pattern and the identification of various types of relationships implicit among the representation of a set of geographic phenomena. Viewing and analyzing geographic space at various levels of detail is common practice in the geosciences. The activity helps to discern the operational scales of geographic phenomena, the extent and permanence of patterns, which in turn sheds light on the underlying processes and their interactions. The scale of observation scopes the problem – the map acting as a filter and, in a digital context, an interface by which we can further explore attributes of the data. Cross scale analysis enables us to connect together processes operating across a continuum from the fine scale (large scale mapping) to the coarse scale (small scale mapping), and is a start point from which we can theorize about causal processes, and from this, make generalizations about the world, and go on to test and develop predictive models (**Figure 1**)– the evolution of our models and theories often revolving around the visual form.

The creation of paper map series (typically at scales of 1:10 000, 1:50 000, 1:250 000, 1:500 000) is testament to a requirement that we are able to view the world in an abstracted, thematic form, at various levels of detail, in order to discern fundamentally different (yet connected) properties among a specific set of geographic phenomena. For example, general circulation models help us understand climate change at the global scale. They are (almost by definition) inadequate in helping us to explain subtle regional differences that might point to localized anomalies, yet which collectively contribute to a global effect. Being able to represent geographic phenomena at a range of scales lies at the heart of the cartographic discipline (**Figure 2**).

Traditionally, the cartographic task was the preserve of the human cartographer. The human acted to interpret the requirements of the user, and produce a paper based map of high quality – void of ambiguity, often tailored to the expectations of the user. Information technology (IT) has disrupted this relationship, offering, in the first instance, opportunities to support this human process (via computer aided cartography), but now going further – seeking to emulate the process of design, thus fundamentally changing the role of the cartographer. Beyond this, IT has created a paradigm shift in how we explore geographic data, leading to developments in the field of scientific visualization, in which the digital map acts as a window onto the underlying database, by which we can manage, visualize, and analyze the data.

Not surprisingly, early attempts at automation treated map series products as being quite distinct from one another – reflecting the discrete production process for each of the scales within national mapping agencies (NMAs). This resulted in different databases, representing (in many cases) the same phenomena but at different scales. But a line of thinking soon began to emerge which challenged the wisdom of this redundancy. The question became: could we not store the phenomena once (at a very high level of detail), and then apply a

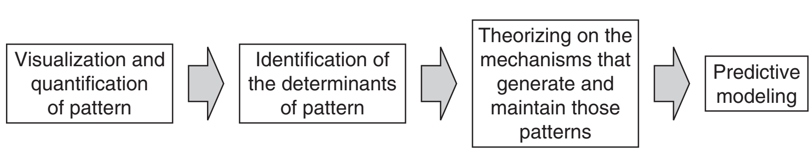

Figure 1 From scrutiny of pattern to predictive statements.

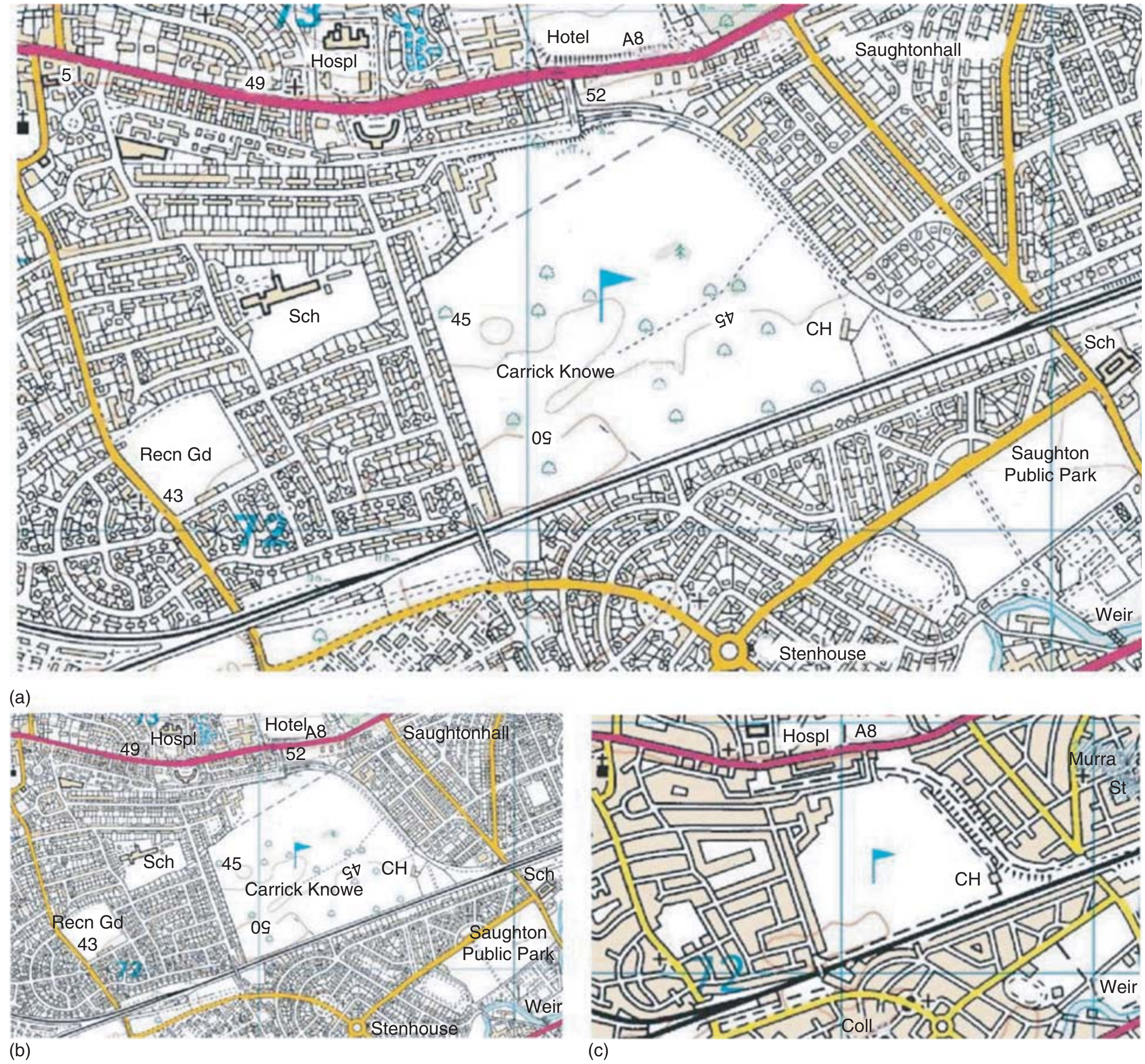

Figure 2 Need of generalization. (a) Source map at 1:25 000, (b) nongeneralized map at 1:50 000, and (c) generalized map at 1:50 000. (Ordnance Survey © Crown Copyright. All rights reserved).

range of algorithms that controlled their selection and representation according to the desired scale and theme?

There were significant benefits to this line of thinking. Maintaining a single database was more cost effective than maintaining multiple databases; a high level of consistency could be maintained between different datasets; duplication of storage could be avoided, thus obviating the need to make multiple updates each time something was changed or added into the geographic landscape; most importantly, it offered the opportunity to integrate data from disparate sources, captured at different scales (or levels of detail).

These benefits were premised on the existence of a set of algorithms that would, with minimum intervention from the user, control selection and representation of geographic phenomena according to a specified scale and theme. The science of 'map generalization' is all about designing such algorithms, algorithms that manipulate and symbolize the geometric primitives stored in the database that represent various real world phenomena. Map generalization is also concerned with (1) modeling the process of design (how generalized does the map need to become), (2) devising evaluation methodologies by which it can assess the quality of the design solution, and (3) development of interfaces that support users who may well be cartographically illiterate, or who struggle to specify their requirements. We begin by describing the techniques used to manipulate objects within the database. We then describe some of the frameworks designed to support their application in the overall design of the map. The discussion that follows this argues that high levels of automation can only be achieved if the automated environment includes methods of evaluation. The article concludes with a brief discussion of the changing context of map generalization within developing applications (such as exploratory data analysis and location based services).

Tools and Techniques for Map Generalization

The goal of map generalization is to give emphasis to salient objects and their properties while omitting less important qualities with respect to the scale and the purpose of a map. We therefore need a system that supports manipulation of map objects and their relationships, and more generally supports the representation of phenomena at different scales. For example, at the finest scale, we might wish to represent each individual building, street light, and pavement that we find within a city. But at a coarse scale, all of this might be subsumed by a single 'dot' (with say, the word 'London' next to it), representing the idea of 'city' in which all those buildings are contained. Therefore, the requirements for a map generalization system are: (1) a database containing some abstraction of the real world, (2) a set of algorithms for aggregating objects in that database (model generalization), (3) a library of symbols with which to render the objects according to various themes, and (4) a set of algorithms focusing on improving the legibility of those symbolized objects (cartographic generalization). The database containing that first abstraction is typically called a digital landscape model (DLM – **Figure 3**). The DLM might be created by digitizing paper maps or from photogrammetric techniques applied to remotely sensed imagery. Typically, a notional scale is associated with the DLM database though it is perhaps more apposite to talk of the level of detail. Data from the database can be symbolized and visualized directly via cartographic techniques. Alternatively, a database of lower semantic and geometric resolution can first be derived (via model generalization) – creating different digital cartographic models (DLMs – **Figure 3**) before cartographic generalization techniques are applied to produce different maps (DCMs).

Altering the theme and level of detail enables different phenomena and different properties to be portrayed. Sometimes the emphasis is on precision of location, or of shape (important in the map interpretation process). In other circumstances, we may wish to emphasize connectivity at the expense of other properties and qualities. Maps of transportation networks (such as the London Underground) are nice examples of the need to emphasize connectivity over geographical location. Irrespective of theme, in all cases a map (digital or paper) reflects a compromise in design – a compromise between

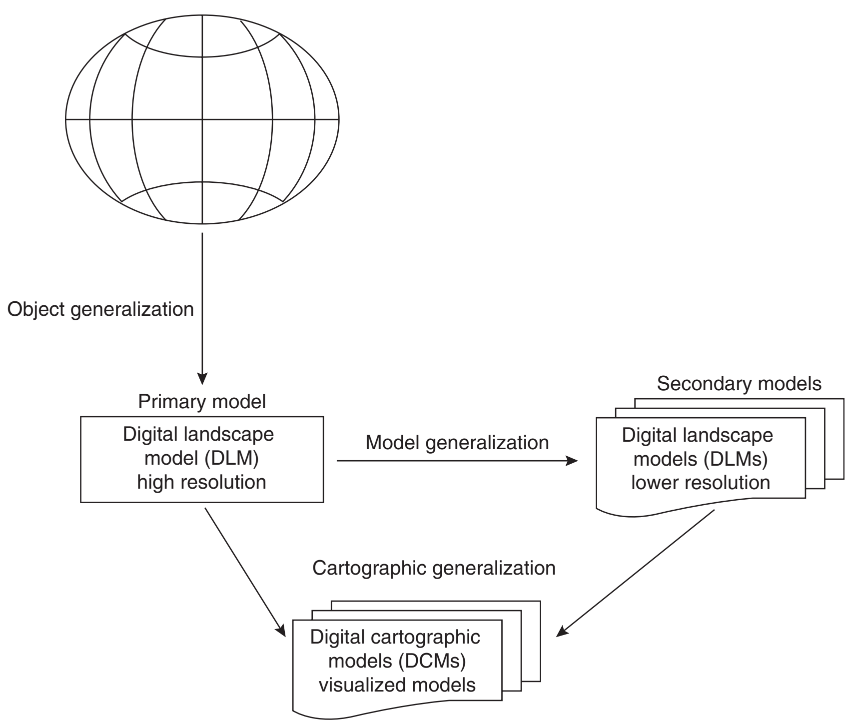

Figure 3 The first abstraction of reality creates the primary digital landscape model (DLM), from which a digital cartographic model (DCM) can be produced either directly from the DLM or via the process of model generalization and the creation of secondary models (DLMs).

wanting to convey information unambiguously, but not having enough room (given the minimum size of symbology) to show all information. In this sense, the process of design is about making sense of things – the cartographer perhaps working from a mental thumbnail sketch by which their solution reflects the requirements of the user in terms of their need, which in turn govern and constrain the representation of each feature in the map.

Various methodologies have been proposed for trying to capture this design process within an automated environment. Via close observation and interrogation of the human cartographer at work, researchers have distilled out a set of techniques used to manipulate and create more generalized yet recognizable forms of geographic phenomena. Considerable research effort has gone into creating algorithms that mimic these human techniques. These techniques are not applied in isolation, but rather in concert, and in varying degree, across the map, depending on the density of information, the type of phenomenon being mapped, and of course, the theme and scale. Therefore, in addition to algorithms that mimic these techniques, we need some framework in which we can orchestrate this whole design process, and we need some evaluation methodologies that enable us to assess the quality of the solution produced within such a framework. We begin with a review of generalization techniques under the headings of model and cartographic generalization.

Model Generalization

Typically we can divide the techniques under two headings – model and cartographic generalization. The objective of model generalization techniques is to reclassify and reduce down the detail, and give emphasis to entities associated with the broader landscape – thus enabling us to convey the extent of the forests rather than see the trees, or to see the island chain along the plate margin, rather than the individual island. The model generalization process is not concerned with issues of legibility and visualization; is more useful to view it as a filtering process; a set of techniques concerned with (1) selection of phenomena according to theme, and (2) the classification and aggregation of phenomena. Typically, model generalization precedes cartographic generalization; alternatively, model generalization may be required in response to a nonvisual query, or as a prerequisite to data analysis. For example, the question 'what modes of travel exist between the cities of Edinburgh and Glasgow?' requires us first to aggregate together phenomena at the fine scale (in this case dense regions of buildings) in order to define the extent and general location of these two entities. Only then can we identify, for example, the major roads that connect these two urban centers.

Composite or higher order objects are formed via the processes of thematic and spatial abstraction. In thematic abstraction, the number of distinct attributes of objects in the database is reduced. In spatial abstraction, the number of objects are reduced by means of aggregation or elimination. Thematic abstraction often triggers spatial abstraction. For instance, objects having similar attribute structure can be categorized into classes under the process of classification. Each object then becomes an instance of a particular class and that class defines an object's properties in terms of its attribute structure. If different classes share some attributes, then a superclass or parent class can be created whose attributes are the common attributes of its child classes. This creates a hierarchy where complex classes are present at the detailed low end of a hierarchy and increasingly abstracted classes are present as we go up the hierarchy. This type of hierarchy is called a taxonomy or classification hierarchy (**Figure 4a**). The creation of these hierarchies is an important way of modeling the changing levels of detail and provides a basis for creating the generalized map (**Figure 4b**).

Another complimentary hierarchy useful in the creation of composite objects is a partonomy. Whereas a taxonomy refers to an 'is–a' relationship, a partonomy refers to 'part–of' relationships between parent and child classes – reflecting more of a functional and conceptual division of geographic space. Over large changes in scale, it is necessary to aggregate objects belonging to different classes in order to create composite objects. A prototypical view of a city might be defined as a dense collection of municipal and industrial buildings, and multimodal transportation infrastructures. Once represented in partonomic form, it can be used as a basis for combining objects together – in this case, moving from the detail of the house, land parcel, and pavement, to a simple city block (**Figure 5**and **Figure 2**).

In addition to these techniques of aggregation, there are two other techniques that fall under 'model generalization' – selection and simplification. As the name suggests, selection is the (straightforward) process of selecting a subset of all classes of objects falling within a specified region. The selection process is governed by task, which in turn tends to define the scale and the theme. The long tradition of topographic and thematic mapping often acts as a basis for specifying content, and thus which classes of objects are selected. The other technique is called 'simplification' – and is defined as the process of reducing the number of geometric points used to store the physical location or extent of a geographic object. One can envisage many points being used to record the detail of the outline of a gothic cathedral, or the sinuous path of a low lying river. The challenge of simplification is to reduce the number of points used to store the representation of such features, but in a way that still conveys their essential shape and location. Successful

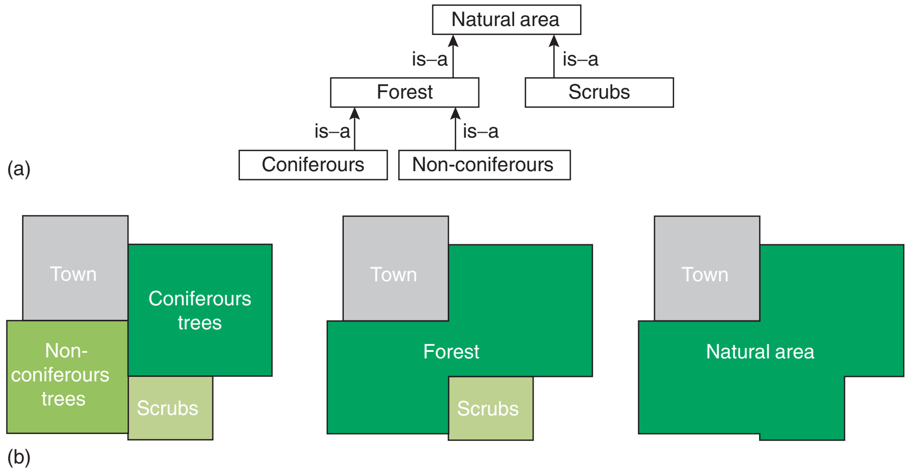

Figure 4 (a) Example of a classification hierarchy, that is, a taxonomy, and (b) class-driven generalization based on that taxonomy.

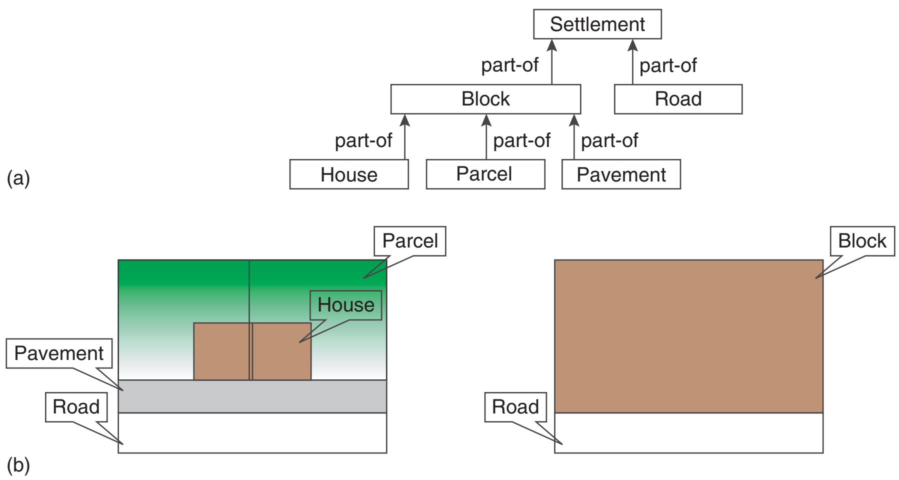

Figure 5 (a) Example of a partonomic structure for settlement, and (b) aggregation of data based on that partonomy.

simplification reduces storage requirements and processing time. Successful algorithms have been those that have identified and retained critical changes in direction (such as the corners of a building, or the main changes in the course of a river) while removing those points that are not critical to conveying the essential qualities of the object (e.g., the extent and angular nature of a building, or the connecting properties of a river to the sea). Once the model generalization process has taken place, the challenge is then to render those objects into some map space (either for paper production, or as part of a digital interactive environment – a desktop or mobile environment).

Cartographic Generalization

Cartographic generalization is a set of techniques concerned with increasing the efficiency with which the map is interpreted – thus the techniques aim to resolve ambiguity, and to retain those qualities of a representation that best fit with the user's expectations. Symbols used to represent spatial objects from the source database must be visible to the naked eye; but as the scale reduces, the amount of space available decreases thus creating competition for space. To retain the clarity and to represent the information effectively, a range of cartographic generalisation techniques are applied such as symbolization, smoothing, simplification, grouping, enhancement, displacement, and text placement (**Figure 6**).

When we come to apply these techniques (often in combination), they must be applied such that, irrespective of the scale of portrayal, the essential qualities of the feature are still conveyed (that rivers retain their sinuous and connected form, and buildings retain their anthropogenic qualities (such as their angular form)). Different combinations, amounts of application, and different orderings of these techniques can produce different, yet aesthetically acceptable, solutions. The focus is not on making changes to information contained in the database, but upon avoiding ambiguity in interpretation of the image. The process is one of compromise reflecting the long held view among cartographers that making maps involves telling small lies in order to convey the truth! **Figure 7a** metaphorically represents the contents of the database. The symbolization and enhancement process has led to overlap (**Figure 7b**) such that displacement is required (**Figure 7c**), which has improved the clarity of the map, but at some cost to the locational accuracy of these features.

Analysis, Synthesis, and Evaluation of Cartographic Solutions

For any given cartographic conflict (such as the one presented in **Figure 7b**), one can envisage a number of viable solutions (of which **Figure 7c** is just one). The choice of solutions will depend, among other things, on: the number of features, their position relative to one another, and their importance relative to the intended theme. Trying to create viable solutions (synthesis – **Figure 8**), and then choosing a solution among that choice requires two things: (1) an initial analysis phase in which the conflicts are identified (analysis – **Figure 8**), and (2) a form of evaluation such that we can assess the quality of the solution (evaluation – **Figure 8**). Failure to find an adequate solution might either result in further analysis of the conflict or flagging unresolved conflicts and drawing these to the attention of the user.

The analysis phase is akin to the eyes of the cartographer and involves making assessment of the degree of severity of the conflict (extent and complexity and composition). A broad and extensive set of cartometric techniques have been developed to measure the various qualities inherent among the map features. This is because we wish to ensure minimum disruption in those qualities during the cartographic generalization process. For example, an unacceptable solution to **Figure 7b** would be one that placed the church to the right of the road, or one that significantly distorted the shape of the road. Maintaining adjacency relationships requires the use of Delaunay and Voronoi diagrams (**Figure 9**) which are ways of exhaustively tessellating the space between geographic features, and making explicit shape and distribution parameters.

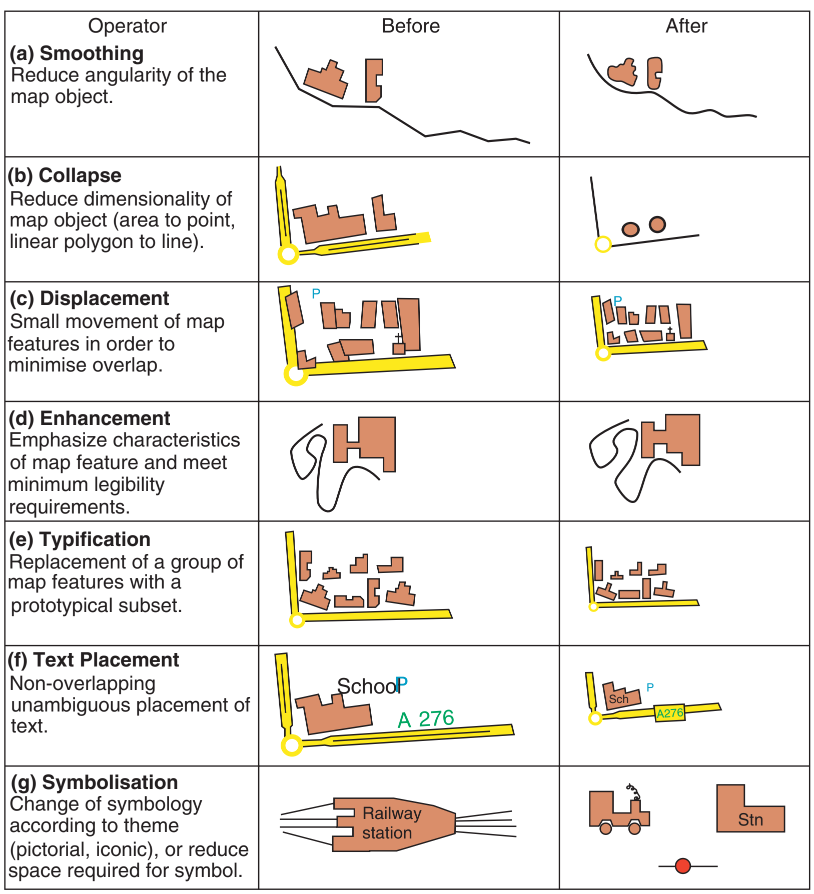

Figure 6 Definition of generalization operator, initial data, and result after application of the technique.

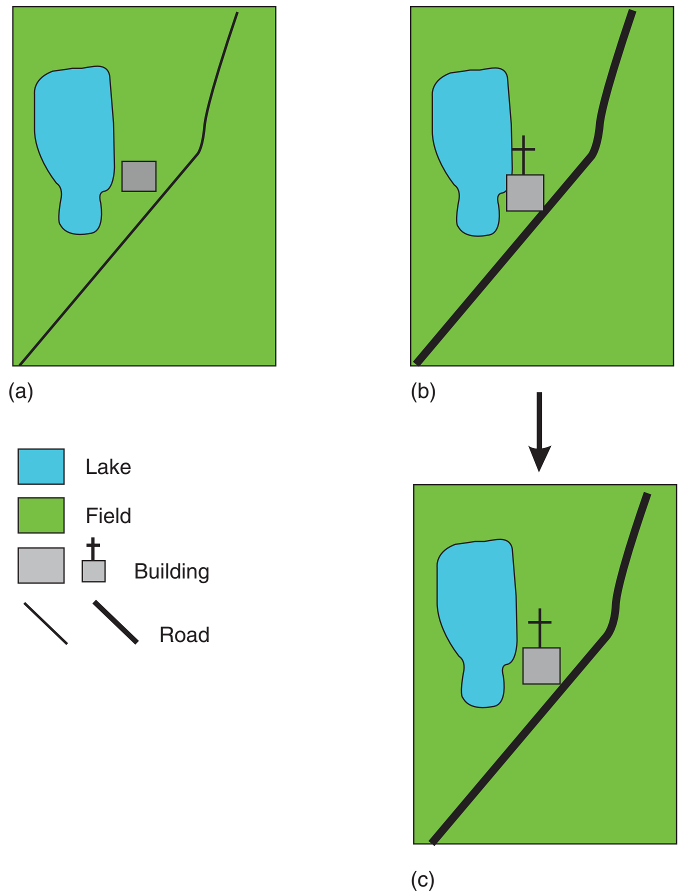

Figure 7 Difference between cartographic and model generalization. (a) Result of model generalization, (b) results of cartographic generalization (symbolization, enhancement, and simplification) has led to overlap, and (c) displacement is needed to remove ambiguity.

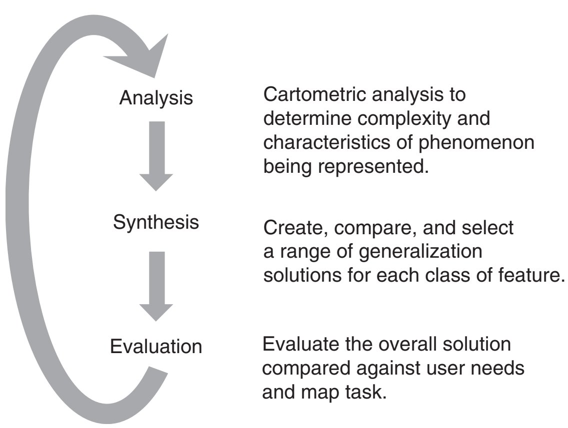

Figure 8 Essential components of a map generalization system: analysis, synthesis, and evaluation.

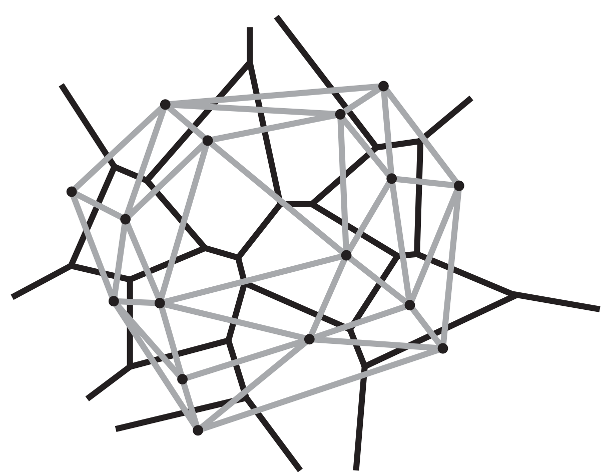

Figure 9 The use of Delaunay (gray) and Voronoi (black) diagrams enables relative proximity and adjacency properties to be measured.

It might also be necessary to ensure connectivity among features (such as road networks) during the process of generalization. Graph theory is often used to model this process (both in model generalization when selecting subsets of roads and during cartographic generalization) (**Figure 10**).

Many other shapes and pattern metrics have been proposed – often applied in the analysis phase, and again in the evaluation phase. If we assume that synthesis involves creating a number of candidate solutions, then the

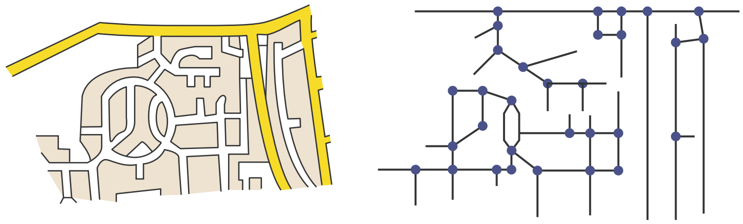

Figure 10 Graph theory can be applied to networks to ensure conservation of connectivity relationships and to control the 'pruning' of networks. (Ordnance Survey © Crown Copyright. All rights reserved).

best solution might be one that has resolved the conflict (improved its legibility), while producing the least amount of change among the various cartometric meas ures (in terms of topology, area, shape, and distance).

Modeling the Generalization Process

Having a set of techniques for manipulating objects in the database and in the portrayal stage is not, by itself, suf ficient when it comes to designing an autonomous map generalization system. The selection and application of techniques, the creation of candidate solutions, and their evaluation requires some framework in which this can all take place. Because of the interdependent nature of geographic phenomena, it is rare that changes can be made without having to consider the broader context. For example, the solution in **Figure 7c** is only appropriate because there is sufficient space for the objects to be displaced into. If buildings have to be aggregated in one part of the map (perhaps because of the density of fea tures), then for reasons of consistency, this needs to be applied to other occurrences. Procedural and heuristic knowledge needs to be incorporated within these frameworks so that the solutions most likely to be suc cessful can be applied first. Among the various 'frame works' explored, two are worthy of mention: rule based approaches, and constraint based approaches.

Rule-Based Approach

Since the cartographic design process appeared to in volve decision making and use of 'rules of thumb', it was assumed that knowledge based approaches (expert sys tems) could be used to model the process – using a rule based approach. These systems used either a predeter mined rule execution sequence or an inference engine to control the execution sequence in applying various techniques. They consisted of three main parts: a know ledge base, an inference engine, and a user interface (**Figure 11**). The knowledge base contained a set of rules, facts, or procedures. The inference engine controlled the generalization process by making use of the rules and procedures in the knowledge base. The user interface contained menus for selecting the datasets and intended

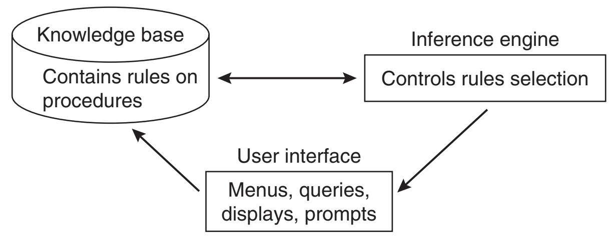

Figure 11 The essential components of a rule-based system.

theme and scale, and a mechanism for adding or updating rules in the knowledge base.

An example rule might be:
If important (building) AND size (building, $< 50\ m^2$) THEN enhancement (building).

This simple example of a rule captures the idea of cartographically enhancing buildings if they are small yet important to the intended theme. There are examples of partial successes with this approach (the OSGEN system from Ordnance Survey and CHANGE from the Uni versity of Hannover), but several weaknesses were iden tified: it was hard to formulate rules for all the exceptions and circumstances in which a particular conflict might occur, they tended to operate over quite small changes in scale, and for limited classes of features. It was hard for the human to know what the consequences might be for any change or addition made to the rules. It had been assumed that rules could readily be formulated from map specification documents, but solutions were often dem onstrated by illustration, and cartography has never de vised the syntax or semiotics (a system of signs) that could be used to facilitate the rule formulation process. When cartographers were asked to explain their actions as they worked, they often found it hard to articulate their reasoning and were found to be inconsistent in their application of a technique.

Constraint-Based Systems

More recently, generalization research has focused on a holistic view of the process acknowledging the knock on effects of generalization and the interdependent nature of

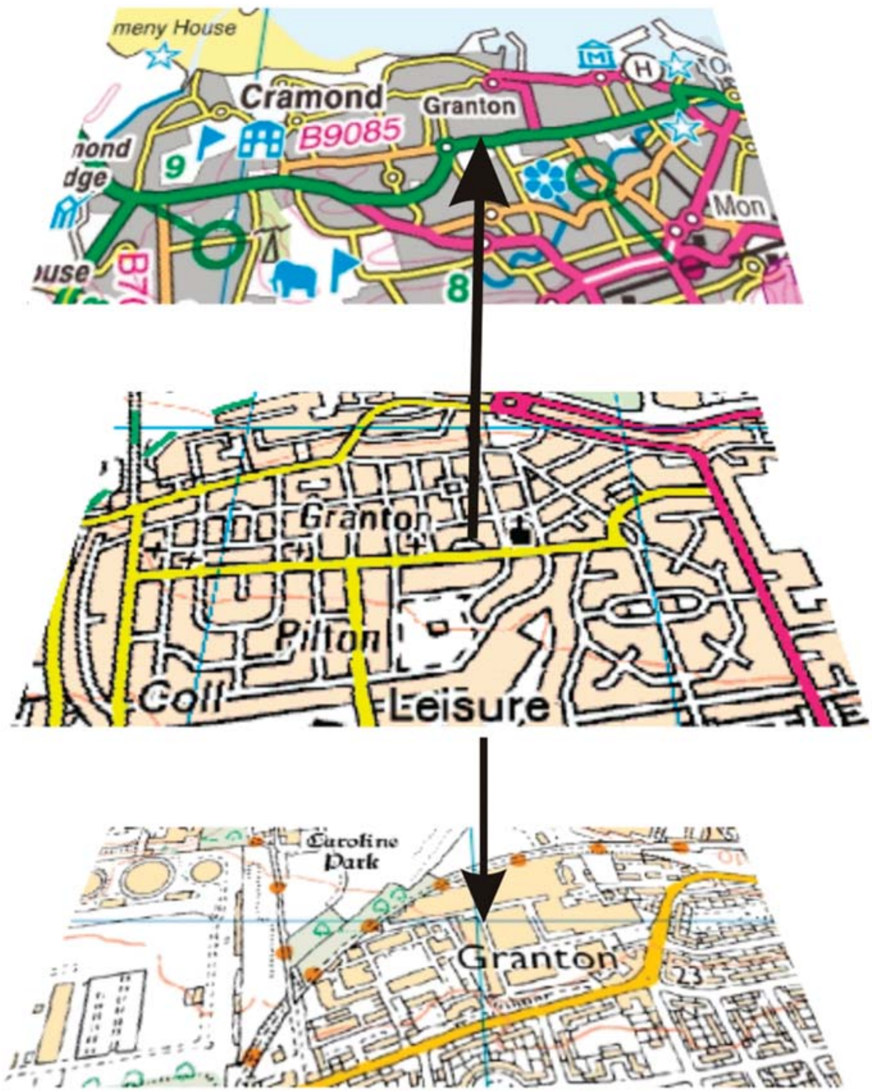

Figure 12 MRDB storing multiple representations of different geographic phenomena. (Ordnance Survey © Crown Copyright. All rights reserved).

the solution. Currently, there is much interest (and promise) in using a constraint based approach – where the aim is to find a state whereby the maximum number of constraints can be satisfied. In this context, there has been much research effort devoted to agent based methodologies – in which each object in the database is modeled as an agent – an object oriented concept in which the object has goals, behaviors, and a capacity to communicate with other agents. These are referred to as 'multi agent systems'. The goals reflect those of the generalization process – namely to efficiently render the object without ambiguity. The agent makes decisions about its representation based on its own goals while considering the goals and constraints of its neighbors. Ideas have included a hierarchy of agents in which higher order agents are concerned with broader contexts and distribution of agent classes, while agents at the individual object level are concerned with the specific representation of individual objects. The AGENT project is one project which has been developed into a commercial system that now supports a number of national mapping agencies, notably the National Mapping Agency of France (IGN).

As a corollary to this discussion, by partially incorporating the decision making process within rule based and agent based systems, the balance of decision making has shifted away from the human to the machine. This has presented some real challenges in the design of interfaces that are intuitive to use, allowing the user to specify their mapping requirements in a simple and efficient manner within very complex systems. Researchers have challenged the idea of totally autonomous solutions, arguing that interaction is critical to ensuring that the user remains part of the design process. The idea of semiautonomous generalization techniques, involving the user in critical evaluation tasks reflects a more col laborative approach to design. Coupled with machine learning techniques, this scenario might enable capture of design heuristics – thus gradually improving the so phistication of proffered solutions.

Multirepresentation Databases

For databases potentially containing millions of objects, the processing overheads for even the simplest of gen eralization tasks is huge. Delivery of digital maps over the Internet, generalized in real time is not currently achievable. The obvious solution is to generalize maps in batch mode in anticipation of demand, or at least store in the database multiple generalized representations of each object, or groups of objects. This has given rise to the idea of multirepresentational databases – databases that ex plicitly store objects in different generalized states, such that generalization is more of a selection and com positional process – bringing together different classes according to scale and theme (**Figure 12**).

The data model must connect together these different representations such that when changes are made in the primary DLM, these changes can be automatically re flected among the more generalized versions of the ob ject. This same hierarchical structure can then be traversed according to the intended level of detail as the user zooms in and out.

Conclusion

When we look at maps at different levels of detail or scale, it is not that we see 'less' information but that we see 'different' information. The aim of generalization is to support this process – to give emphasis to attributes and relationships at the broader scale. Attempts at automation have revealed just how challenging and creative the art and science of cartography is. Developments in visual ization methodologies have not obviated the need for generalization methodologies that abstract geographic space in order to reveal the patterns and inter dependencies inherent among geographic features.

See also: Geovisualization; GIS and Cartography; Maps; Mapping, Distributed; Spatial Databases.

Further Reading

Brassel, K. E. and Weibel, R. (1988). A review and conceptual framework of automated map generalization. *International Journal of Geographical Information Systems* 2, 229 244.

Joao, E. M. (1998). *Causes and Consequences of Map Generalisation*. London: Taylor and Francis.

Lamy, S., Ruas. A., Demazeau, Y., Jackson, M., Mackaness, W.A. and Weibel, R. (1999). The application of agents in automated map generalisation. In Keller, C. P. (eds.) *Proceedings of the Nineteenth International Cartographic Conference*, pp 1225 1234 Canada: Ottawa

Mackaness, W. A., Ruas, A. and Sarjakoski, L. T. (2007). *Generalisation of Geographic Information: Cartographic Modelling and Applications*. Amsterdam: Elsevier.

McMaster, R. B. and Shea, K. S. (1992). *Generalization in Digital Cartography: Resource Publication in Geography*. Washington, DC: Association of American Geographers.

Molenaar, M. (1998). *An Introduction to the Theory of Spatial Object Modelling for GIS*. London: Taylor and Francis.

Muller, J. C., Lagrange, J. P. and Weibel, R. (eds.) (1995). Generalization: State of the art and issues. In Masser, I. & Salgé, F. (eds.) *GIS and Generalization: Methodology and Practice (Gisdata, No. 1)*, pp 3 18. London: Taylor & Francis.

Ruas, A. (1999). *Modèle de Généralisation de Données Géographiques à Base de Constraints et D'autonomie*. Thèse de doctorat. L'universite de Marne La Vallée, Paris. http://recherche.ign.fr/labos/cogit/pdf/THESES/BARD/These Bard 2004.pdf (accessed in April 2008).

Sheppard, E. and McMaster, R. B. (eds.) (2004). *Scale and Geographic Inquiry: Nature Society and Method*. Oxford: Blackwell.

Weibel, R. and Dutton, G. (1999). Generalising spatial data and dealing with multiple representations. In Longley, P. A., Goodchild, M. F., Maguire, D. J. & Rhind, D. W. (eds.) *Geographical Information Systems*, pp 125 156. New York: Wiley.

Relevant Website

http://ica.ign.fr/
ICA Commission on Generalisation and Multiple Representation.

Genetics

E. Hall, University of Dundee, Dundee, UK

Glossary

Bioprospecting The search for and exploitation of valuable compounds from the genetic resources of, for example, the plant life of tropical forests, for use in the development of medical and cosmetic products. The ownership of and rights to prospect and use such material is commonly contested between commercial companies and national governments.

Geneticization The increasing tendency to view most illness conditions, behaviors, and physiological differences as determined (wholly or in part) by genes (adapted from definition by Abby Lippman, 1992).

Genetic Map Genetic maps represent the relative and absolute positions of genetic sequences on the 23 pairs of chromosomes that sit within the nucleus of almost every cell in the body. Detailed genetic maps are critical in the identification of genes that contribute to disease susceptibility.

Hybrid A term used both to critique the long-standing nature–society binary and to describe the new forms of mixing of social and natural elements in, for example, genetically modified plants and animals.

Introduction

Stimulated by a series of biotechnological developments, human geographers have begun a tentative research engagement with genetic science. We are now in the so called 'biotech century', a period of transformation of the understanding of plant, animal, and human bodies, and the possibilities to reconfigure these bodies through biological manipulation. From genetically modified plant crops to the identification of genetic 'causes' for chronic illnesses, and from the global trade in genetic resources to the detailed tracing of early human migrations, the changing nature of seemingly immutable biological entities is now a major focus of public and academic attention.

Although this may seem the stuff of natural and medical science, geography's interest should perhaps come as no surprise. The myriad of developments – encompassing bodies, environments, science, politics, corporations, and ethical concerns – has at its heart the interface of nature and society. All of the developments we are witnessing as the biotech century unfolds involve the transformation, or at least the reframing, of the natural–social relation, either through material change (e.g., the genetic modification of a plant's biological makeup, to produce more flavorsome tomatoes or disease resistant soybeans) or metaphorical change (e.g., the recasting of the causation of disease, such as the identification of the BRCA 1 and 2 'genes for' breast cancer). This cross penetration of the social into the natural and the natural into the social challenges embedded binary understandings. As such, the key debates generated by genetic science are centered on the disciplinary core of geography – understanding the interface of the natural and social worlds. Echoing David Stoddart's words of two decades ago, geography is arguably well placed to 'claim the high ground' of the theoretical framing and analytical investigation necessary to understand and critique the revolution in the relationship between natural entities and processes and social structures and actions, that genetics presents.

Sarah Whatmore and Noel Castree, in separate editorials in 1999, kick started geographical interest in genetics. Whatmore insisted that geography's integrative positioning spanning the natural and the social sciences makes it perfectly suited to understanding the reconfigurations of the natural–social relation produced by genetic science. At the same time, however, the challenge presented by genetics has highlighted the theoretical and methodological divergence of human and physical geographies, where science has been largely rejected within human geography and, more broadly, how the powerful sense of integrated 'lived worlds' that were once studied and celebrated within the tradition of geography, comfortably combining human and physical sciences, has been lost. Castree shares a similar frustration with the inability of a geography separated on natural and social science lines to tackle the challenges posed by genetics. Encouragingly, in response, not only has geography begun to rediscover its 'synthetical' approach to the world, albeit in small measure, but it has also begun to apply both existing and innovative theoretical tools to the new questions raised by genetics.

Castree initiated a neo Marxist framing of genetics as a new manifestation of the long standing production and commodification of nature. Arguably, all that has changed is the scale of the focus and penetration of capital into nature. In short, there is little difference between the exploitation of the resources of forests, rivers, and land, and that of the genetic material of plants, animals, and human bodies. In a globalized economy, capital accumulation through nature draws together resources from a myriad of places, bodily and organic sources, to sustain this accumulation. An example is 'bioprospecting', the search for genetic material in, particularly, the plant life of tropical forest regions of developing countries (e.g., Costa Rica) that can be potentially utilized in the development of medical and commercial products. Economic benefits accrue to the source country for the materials under bioprospecting agreements, but the developed West gains far more in its exploitation of this biological material and potential. What is perhaps different about such exploitations of nature is that in this new genetic regime of accumulation, nature is not damaged or destroyed as the DNA is extracted. Rather, once returned to the laboratories of the West, the natural resource is 'unpacked' to reveal its genetic constituents and then, through combination with other natural (e.g., another plant) and social (e.g., a pharmaceutical or cosmetic compound) entities, it is remade into an object that is a form of natural–social mix or 'hybrid'.

There is a powerful sense that genetic science, in its identification and manipulation of the genetic material of plants, animals, and human bodies, is redefining the relationship between nature and society and, as such, demands a new conceptualization. A binary of fixed, immutable opposites, limited to interaction, rooted in Enlightenment scientific rationality, has defined the way that human geographers from all traditions have regarded nature and society. While realists see nature as material and social constructionists understand nature as a product of society, *both* see nature and society as separate spaces and entities. The cultural turn in geography squeezed any lingering sense of 'life' out of the conceptual frame, though more recently there has been a welcome resurgence of interest in materiality. The realist versus constructionist positionings on nature of course reflect the nature–society binary itself, and the underpinning sense of 'pure' forms of the natural and the social, a sense that persists despite many years of environmental transformation by social action. Society cannot escape nature, and nature has quite clearly not escaped society. Genetics, with its deliberate mixing of the natural and the social – from genetically modified crops to medical gene therapy – demands a different conceptualization, where binaries can be discarded and a sense of hybridity embraced. Social theorists Donna Haraway and Bruno Latour have provided inspiration for this, with their stress on the mixing of the human and the nonhuman and, in particular, a desire to include nonhumans, from plants and animals, to technology and machines, in social worlds on an equal footing. The focus of interest then shifts from interactions between the separate entities of nature and society to the networks and relations between a whole set of people, places, objects, and organisms within which hybrid entities are produced and maintained. Genetically modified plant crops and gene therapies are two such natural–social hybrids, the products of strongly geographical networks of scientists, government, pharmaceutical corporations, medical clinicians, technologies and machines, bodies, and the microbiological chemicals of DNA themselves.

Networks of Genes

A 'relational' sense of nature–society has permitted geographers to engage effectively in the challenging and intense debates on the new genetics. Most importantly, a geography of relations places the hybrid object – the gene, the genetically modified plant, and the gene therapy – in the context of the networks that produced it, forcing an engagement with the full range of natural and social factors. Just as importantly, the materiality of the chemical mix of DNA in the human or nonhuman body that is named the 'gene' is emphasized. Each of the elements in the network is an 'actor' or 'actant' and all are caught in a productive tension, able to act, but also subject to reaction (a mix of social decisions and natural nonintentional actions), and as they relate to one another, so a collective understanding is produced. So, in theory, no one actor – human or non human – can dominate the network, with any outcome being the product of relations between all of the actants in the network.

Two examples of this approach are provided by Whatmore and Edward Hall. Whatmore describes how plant genetic resources are within and are the product of a complex network or 'assemblage' of developed and developing world governments, multinational corporations, peasant farmers and technologies, patent laws, and international agreements on the exploitation and trade in such resources. 'Actor network theory' encourages a reading of this as neither an ordered and governed process of trade nor as a simple power imbalance of domination and exploitation, but as a 'coming together' of factors with unpredictable outcomes. Through adopting such an approach, we can see that genetic technologies raise some important questions. First, the global trade in plant genetic resources has tried to make sense of the genetic material by wishing to delimit its biological purity, that is, state with clarity its discrete genetic 'root', but has found this an impossible task given the long history of human cultivation that plants have undergone, and the complex ecological contexts within which they are located and are to a large extent defined. Second, delimiting the geographical origin of the plant genes has become a source of tension, between the notion of genes as a common entity, to which all humanity has common access, and the notion that developing countries (where most plant genetic resources of interest to corporations are located) have ownership over these indigenous species and their biological material. Third, are the genes the materiality of the chemical mix of DNA or are they the intellectual and technological rigor that extracts and utilizes them? This is important, for it decides where the gene is located (and owned), in the laboratories of the West or in the fields and forests of the South. Through all of this, the impossibility of separating the natural from the social is clear; plant genetic resources, indeed all genetic material, are necessarily at the same time a social artifact and a natural entity.

Innovations in healthcare, and in particular new cures and treatments for chronic and fatal illnesses such as cancers, heart disease, and Parkinson's disease, have been at the forefront of the promises of the new genetics, heralded by the 'completion' of the Human Genome Project (HGP) – the mapping and sequencing of all of the human body's genes – in 2003. Geographers (and others) have seen this as a potentially significant shift in the medical and social discourse of health and disease causation and therapy. Dominant understandings of the causation of health and illness, that is, individual behaviors conditioned by external social and spatial contexts, have been challenged by the claim of internal causation located within an individual's genes. The emergent notion of individual behaviors and environments as either irrelevant or providing simply the trigger for the development of a disease condition, has become known as the 'geneticization' of disease. The importance of geographical contexts to the explanation of health is increasingly marginalized. Through the adoption of Actor Network Theory, Hall has challenged this seemingly smooth transition to geneticization (of heart disease, in this instance). By conceptualizing geneticization as a managed product of a highly spatialized network, including, crucially, not only scientists and the genes on which they conduct research, but also funding, technology, and even experimental rats (and the unpredictability of their genetic material), a fuller picture of the 'messiness' of science emerges, exposing the impossibility of separating social (environment) and natural (genes) factors, and the remaining powerful influence of certain actors (geneticists), contexts (government and charity science policies), and geographies (biotechnological data and medical research) in explaining the causation of and the development of treatments for heart disease.

Genetic Cultures

Other geographers have examined the ways in which genetics highlights how the biological makeup of the body has resonances amplified far beyond its chemical constituency, reaching into debates around what it means to be human, of a particular ethnic or sexual identity, and the broader tension between biology (nature) and human agency (society) in determining futures. The HGP revived and intensified a long standing, if up until then scientifically rudimentary, genetic imagining of the body, in particular in its process of the 'mapping' of the genome. The HGP saw the production of a complete genetic map of the human genome – the identification of the positions of genetic sequences on the 23 chromosomes that sit within the nucleus of almost every cell in the human body. The genetic map is the central analytical tool of the new genetic knowledge, operating both as an information pool in the gathering of data, and as a framework for the subsequent interpretation of genetic information and the linking to physiological and psychological conditions. It is no terminological accident that geneticists used the language of cartography to describe the gathering of knowledge of the body's DNA, nor that the media so readily adopted this metaphor. Maps retain a popular metaphorical power, suggesting the complete and accurate representation of an unknown landscape, in this case the genome. Further, the map presupposes that there is a landscape there to be known (and that it is possible, through technology, to fully know that landscape), and that once mapped, the landscape becomes a defined object, possible to occupy, own, and exploit. The purpose of the genetic map is to imagine the gene, to transform its chemical materiality into a set of individual objects set within an overall spatial structure. The map makes sense of the gene, indeed it 'makes' the gene – breaking the body down into atomized pieces, opening up the body and the genes (like a colonized land) to domination and ownership (such as the commodification of genetic information by pharmaceutical and biotechnology companies). So, while the map now perhaps stands on the fringes of human geography as an analytical tool, genetics reminds us both that it retains metaphorical power and that as geographers we should reawaken our long held knowledge of the map. With the map, as with the natural–social interface, genetics speaks to the disciplinary core of geography.

The complete knowledge of the body promised by the so called 'book of life' of the human genome map, and the subsequent linking between particular genes and bodily states, has penetrated the geographical study of identity. The HGP and other genome projects, most notably in Iceland, have used gene patternings within indigenous, and therefore supposedly pure, populations as benchmarks for the assessment of ethnic difference and migratory histories. This draws on the notion of the gene as somehow our innate humanity, reaching back into our evolutionary past and unaffected by socio environmental change over time. So, discussions of racial and genealogical identities, and the origin and spread of peoples and families, have been heightened by genetic technologies and claims of absolute purity and complete knowledge. Such supposed purity of identity of ethnicity, sexuality, and nationality can be a dangerous set of claims to make, as discrimination (utilizing the same genetic technologies) can flourish. However, there is, more positively, the potential for genetics to reveal the complex histories – in Stuart Hall's memorable phrasing, the 'routes' rather than the 'roots' – of people now defined and confined by a national state border, shifting toward a possible hybrid notion of identity.

In the case of Iceland's genome project, in which large numbers of people have donated genetic material to be triangulated with medical and genealogical records to better understand gene–environment–health relationships, and so further the search for 'genes for' chronic illnesses, there is understood to be a very particular genetic resource containing a bounded, stable, and homogenous population (termed an 'island laboratory'). Importantly, not only was the genetic material of thousands of individuals depersonalized and transformed into the resource of a 'national' genome with generalizable characteristics, but also the genes were very much the target of the research, not the individuals that host them. The human body, and the broader contexts within which bodies exist, is replaced by the gene code (and the associated medical record data). And, further, this genetic information can then be enrolled and commodified into the genetic technologies of the organizations managing the projects (and become bioinformatic information to transact globally). However, it may not be as straightforward as it seems to pull apart the gene from the body and its contexts; through the consent forms, medical records, and public opinion, individuals retain a grip on the ways in which the genetic material is used, and reassert the importance of environments in understanding health and disease, and human development.

The HGP and other national genomic collection programs and the dominant symbol of genetics – the double helix of DNA – emphasize a common humanity: we all, irrespective of ethnicity, gender, sexuality, or location, share a very large percentage of our genetic material with other humans and organisms. At the same time, the spinning helix also invokes the commodification of the body's biology, the distancing of people from their biology, and the ability to alter 'life', which seems to undermine any sense of common humanity. Charting a way through this debate, geographers emphasize the often marginalized social contexts, natural environments, and natural–social networks within which the body and identity is produced, reproduced, and contested.

The Ethics and Regulation of Genetics

The ethical issues raised by genetics in one sense echo those of other areas of contemporary science – such as nuclear energy – in raised risk and uncertainty; in another sense, they herald something quite different, as the natural–social boundary, usually so tightly marshaled, is unsettled and disrupted as genetics continues to push the boundaries of what is 'natural'. As such, much thinking on the ethics of genetics has been provoked within geography. Genetics plays on our greatest hopes – of cures for chronic illness and disease resistant plant crops to feed the poor of the world – and our greatest fears – of producing entities, such as genetically modified crops and transgenic animals (e.g., pigs and sheep developed to produce human proteins in their milk, and 'oncoMouse' altered to develop cancerous tumors), that are not 'natural', potentially dangerous to health, and for many deeply disturbing.

Geographers have tackled these issues in ways that avoid making a judgment between what is 'right' and what is 'wrong' – as such judgments tend to be formulated around maintaining distance between the natural and the social – instead, emphasizing the new complexities of natural–social relations, and tracing the many connections, networks, and interests involved, spread across a range of spaces. So, the ethical challenge becomes something rather different: traditional ethical discourse is framed in terms of the dualism of nature and society, with ethical values and decisions lying solely within society (and nature as untainted by such dilemmas). The capacity for ethical judgment is possessed solely by the rational human individual within passive social and environmental contexts. One ethical response to this narrow position is the extension of rights to nature, to animals, and to plants. But in the world of hybrid entities, where the natural and the social are entangled in new and ever more fantastical formulations, a different approach is perhaps necessary and possible. The notion, then, of 'ethical communities' of people, bodies, plants, institutions, and others may allow us to recognize that ethical discourse is something produced by the actions and reactions of a whole series of actors. So, the ethics of genetics, rather than being a set of predetermined codes of what is 'right' and 'wrong', instead emerge within this network of relations. The difficulty that society is finding at present when faced with technological developments in plant and animal hybrids, and ever more innovative medical and pharmaceutical treatments (such as animals genetically modified to allow animal to human organ transplantation), that seemingly revel in the possibilities of injecting the social into the natural and vice versa, and in deciding whether such developments are ethically acceptable or not, could be bolstered by a reconceptualization of ethics. If we recognize that a transgenic animal is neither (clearly) purely natural nor purely social, but is the product of multiple connected processes, then our geographically informed ethical eyes can be trained on the spaces and agents where these entities are being produced.

While the products of genetic science are novel, the structures, organizations, and interests within which they are produced are arguably not (albeit in new combinations). We do not have to decide whether we are 'for' or 'against' genetics; our energy is better spent unearthing how, where, and why it is happening, who determines the rules, and who (humans and nonhumans) benefits. For example, the bioprospecting for plant or human genetic resources in Costa Rica or Iceland need not be rejected on principle as neocolonial exploitation of natural resources (although it can be), rather it can be located and then questions asked about why such activity is occurring in that particular place, what are local understandings and responses, and who benefits and how. The assessment then becomes less of ethics and more of critical practicality. This is part of a growing spatializing of genetics within geography. Through 'placing the gene' within the 'uneven and unequal' contexts where genetic science is being produced, applied and regulated, the complex, dynamic, and contested landscape of genetics is mapped, and geography's unique contribution identified. In the case of medical genetics, these contexts include laboratories funded by corporate and charitable science, clinical services unequally provided within national healthcare, and the social and spatial inequalities of health, and government policy and legal frameworks. For genetically modified crops and seeds, these networks are transnational and transcultural, with genetic products being mapped onto existing unequal relations of knowledge, technology, trade, and politics.

The study of genetics is dominated by natural and medical science. Geographers, however, are beginning to forge a particular and incisive approach. By mapping out the networks of the production and application of genetic science, revealing the many natural–social actors, interests, and spaces, geography is reasserting the importance of (uneven and unequal) contexts and environments. Further, genetics is an opportunity for geography to reconnect to its disciplinary roots: the concern with the materiality of 'life' at the interface of the natural and social worlds.

See also: Nature; Nature-Culture.

Further Reading

Bridge, B., McManus, P. and Marsden, P. (2003). The next big thing? Biotechnology and its discontents. *Geoforum* 34, 165–174.

Castree, N. (1999). Synthesis and engagement: Critical geography and the biotechnology century. *Environment and Planning A* 31(5), 763–766.

Castree, N. (2003). Bioprospecting: From theory to practice (and back again). *Transactions of the Institute of British Geographers* 28(1), 35–55.

Castree, N. and Braun, B. (eds.) (2001). *Social Nature: Theory, Practice and Politics*. Oxford: Blackwell.

Davies, G. (2006). The sacred and profane: Biotechnology, rationality and public debate. *Environment and Planning A* 38, 423–443.

Greenhough, B. (2006). Towards a geography of bodily biotechnologies. *Environment and Planning A* 38, 416–422.

Greenhough, B. (2006). Decontexualised? Disassociated? Detached? Mapping the networks of bioinformatics exchange. *Environment and Planning A* 38, 445–463.

Hall, E. (2003). Reading maps of the genes: Interpreting the spatiality of genetic knowledge. *Health and Place* 9, 151–161.

Hall, E. (2005). The 'geneticisation' of heart disease: A network analysis of the production of new genetic knowledge. *Social Science and Medicine* 60, 2673–2683.

Lippman, A. (1992). Led (astray) by genetic maps: The cartography of the human genome and health care. *Social Science and Medicine* 35(12), 1469–1476.

Nash, C. (2002). Genealogical identities. *Environment and Planning D: Society and Space* 20, 27–52.

Nash, C. (2005). Geographies of relatedness. *Transactions of the Institute of British Geographers* 30(4), 449–462.

Whatmore, S. (1999). Geography's place in the life science era? *Transactions of the Institute of British Geographers* 24, 259–260.

Whatmore, S. (2002). *Hybrid Geographies*. London: Sage.

Gentrification

D. J. Hammel, University of Toledo, Toledo, OH, USA

Glossary

Devalorization In Marxist urban theory, it is the destruction of a parcel of land's ability to produce increased return on capital invested in it.

Displacement The forced movement of a lower-income individual out of a neighborhood usually due to increasing rents or property taxes brought about by gentrification.

Land Rent It is a form of economic rent accrued through the control (or ownership) of land in productive use.

Postindustrial City As it relates to gentrification, it is the city created by the move away from a strong reliance on manufacturing as an economic base, particularly throughout the 1970s, and an increasing emphasis on the provision of high-level business and financial services as the basis for economic growth.

Revanchism It is a nineteenth-century French political movement that attempted to take revenge on the working classes for a series of perceived national humiliations. As applied to gentrification, it describes the series of neoliberal policies aimed at poor and minority urbanites to ensure that central cities were more palatable for middle-class habitation and consumption.

Definitions

Gentrification refers to a process of neighborhood transformation in which working class and poor residents are displaced by an influx of middle class residents. This change results in improvements in the area's private housing stock and public infrastructure with a con comitant increase in house values and contract rents. Gentrification, the processes that create it, its effects, and its political ramifications have been the focus of a large amount of social science research and substantial political rhetoric. The importance of gentrification rests not only upon the dramatic effects it has upon both the residents of a neighborhood and the neighborhood's landscape, but also upon the linkages that have been created between gentrification and theories of urbanization.

Origins

Sociologist Ruth Glass originally used the term in 1964 to describe a process of class based neighborhood change in a handful of London neighborhoods like Islington. The process clearly predates the name and what might be termed as modern gentrification began in the mid 1950s as bohemian enclaves in sections of several global cities. The earliest identified gentrification in the United States was the discovery of Georgetown in Washington, DC, by bureaucrats brought to the city to assist in the adminis tration of New Deal programs of the 1930s – although the process appeared contemporaneously in some Boston neighborhoods as well. In Europe, the first significant evidence of gentrification was in 1860s Paris. 'Embour geoisment', as it was termed, was one result of the wholesale destruction of neighborhoods inherent in Haussmann's infamous modernization scheme. Despite the precursors, the coining of the word gentrification does represent an important watershed because it marks the beginning of a period of cyclical but sustained gen trification activity in cities around the world.

Shifts in Meaning

The term 'gentrification' has been subject to shifts in meaning over the four decades since it was first created, and debate about the meaning of the term has been a part of the process since the beginning. The gentrification that Glass observed in London involved the basic upgrading of older homes and mews dispersed through several inner city neighborhoods. As the process expanded in London and was identified in other advanced capitalist cities, it began to involve the redevelopment of retail properties as well as residential ones. In addition, some areas experienced demolition and new construction. Little attention has been paid to retail redevelopment, but most scholars seem to acknowledge that it is part of the gentrification process, although clearly secondary to the activity in the housing market. A more robust and longer lasting debate developed around the legitimacy of what is now termed new build gentrification. Some ob servers confine the use of the term gentrification to the process of rehabilitating old housing stock, or what is sometimes termed traditional gentrification. This usage is somewhat more common in work emanating from the design fields, and despite the substantial limitations that it places on both the extent and impact of gentrification, there are still those who adhere to this traditional view. Very few areas, however, have been gentrified purely through the redevelopment of old housing stock, espe cially in the North American context. Even Society Hill in Philadelphia, the locus of some of the earliest and most theoretically significant work on gentrification, involved both the redevelopment of historic structures, and clearance and construction of large condominium towers.

The political context of gentrification has also created substantial debate over the use and meaning of the term. While there was little early reaction to Glass' usage, by the mid 1970s the term had significant competition. Reinvestment, revitalization, renewal, and renaissance were all preferred over gentrification. The authors of one of the early scholarly collections on the topic chose to use the phrase 'neighborhood renovation' in the title, and the special section of the *Journal of the American Planning Association* appearing in 1979 used the title 'Symposium on neighborhood revitalization'. There has been no systematic investigation of the evident discomfort that American scholars had with the term, but the class relationships explicit in gentrification may have played a role. In addition, early North American research often viewed gentrification as a generally positive process, and alternate terms, particularly renaissance, may reflect that bias. Throughout the 1980s, research on gentrification took a much more critical approach and organized resistance to gentrification became much more common. The term was increasingly used by academics in Europe and North America, but had acquired a strong enough negative connotation that policymakers, developers, and gentrifiers themselves actively avoided it. Perhaps the most acute example of this avoidance in the Urban White Paper completed in 2001 as a guide to the future of the cities of the United Kingdom. The term renaissance appears in the subtitle of the implementation document, and regeneration is used commonly to describe the goals of the policy. Some critics have pointed out that the document is one of the most comprehensive policy statements on gentrification ever written, but the term gentrification never appears in the hundreds of pages of text. For many, the term gentrification is one to be avoided for fear of a political backlash. For others, its explicit class connotations make it the preferred label for the process.

Urban scholars have also expressed some difficulty with the term on a conceptual level. Since the early 1980s, there has been concern that gentrification has been a catch all term used to describe an increasingly chaotic concept. In an attempt to introduce more specificity, various modifying phrases have been added. New build and traditional gentrification have been joined by rural gentrification and super gentrification as ways to describe newly theorized aspects of gentrification. Rural gentrification was first identified in the United Kingdom in the early 1980s and most of the research on the topic has been conducted there. The term is used to describe the class transformation of large portions of small rural villages. New telecommunications technologies have made it possible for a small portion of the urban middle class to leave the city and establish rural residences. The transformation in some villages is similar to that in gentrifying urban neighborhoods and the concept of rural gentrification has gained recognition as another form of the gentrification process. At the same time the process in urban areas has continued to develop, and after a history of over four decades in some cities gentrification in some areas has intensified to the extent that a previous generation of gentrifiers has been displaced from their homes by a set of very wealthy gentrifiers who may have multiple residences and reside in the neighborhood on a part time basis. This super gentrification appears to be confined to major global centers, and is yet another facet of the larger process of gentrification.

History of Gentrification

Gentrification Waves

Like most urban processes, gentrification is temporally uneven. In the European and North American context it has been closely linked with economic cycles, first as a counter cyclical process, and later as a cyclical one. The first significant wave of gentrification appeared in the late 1960s and early 1970s. Jason Hackworth and Neil Smith observe that gentrification during this period was led by state redevelopment efforts and was confined to major US and Western European cities. A second wave of gentrification began to gain momentum in the late 1970s. At this point, it became evident that gentrification was no longer confined to global cities, but was present in moderate amounts in cities well down the hierarchy. Although the state still was actively involved in promoting the process, the role of private actors, who had begun to see the market potential of gentrification, was more prominent in creating gentrification. This time period, particularly in the mid 1980s also saw coalescence of organized resistance to gentrification, albeit mostly at the local levels. In the United States, this period also saw one of the largest office building booms in the country's history due in part to tax policies of the Reagan administration. Much of the new construction was focused on the core of major American cities and the combination of public/private partnerships to create large gentrification projects and substantial office tower construction made a highly noticeable impact on the landscape of the American downtown. The third wave began in the mid 1990s as the economies of the industrialized world began a long period of sustained economic growth that lasted throughout the decade and into 2001. Another mild recession slowed gentrification activity, but unusually low mortgage interest rates, especially in the United States, allowed urban housing markets to navigate the recession with little difficulty. The result has been nearly a decade and a half of substantial growth and expansion of gentrification. This wave of gentrification is distinguished from prior bursts of activity not only by its duration, but also by the expansion of the process to cities outside the industrialized world. Some research has also suggested that this new gentrification has been financed more by large, global sources of capital than in previous periods.

Geographic Extent

Over the four decades that gentrification has become a significant feature of the urban fabric, its extent and intensity have changed dramatically. By the late 1970s gentrification seems to have been identifiable in most major cities in Western Europe, North America, and Australia.

Throughout the 1980s published reports of gentrification in Johannesburg, Kyoto, Sao Paulo, Buenos Aries, and Tel Aviv appeared. Shortly after the fall of communist regimes in Eastern Europe, the gentrification process began in earnest in cities such as Prague and Budapest. It was even identifiable in Warsaw whose historic core was almost completely destroyed in World War II. By the late 1990s, gentrification was evident even in Moscow. Given the strong movement toward a market based economy in East and South Asia it is unsurprising that gentrification and displacement became major issues in Chinese and Indian cities, especially, Shanghai and more recently in Mumbai. While the spread of gentrification continued, its intensity in North American and European cities increased. By 2001, large sections of London and New York were heavily gentrified. The processes had proceeded to the point that long standing areas of poverty were gentrifying: Harlem and the Lower East Side in New York, and sections of Brixton and the East End in London. Manhattan had become such a center of activity that gentrification spilled over into substantial sections of Brooklyn and even Hoboken, New Jersey. In Chicago, gentrification reached deep into the south side, and northwest of the Loop has surrounded Cabrini Green, one of the largest public housing projects in the country. In several more spatially limited central cities like Boston and San Francisco, gentrification, combined with an overheated housing market, had almost eliminated any affordable housing in the central city. Toronto, Vancouver, and Sydney also faced similar issues.

At the same time, gentrification was moving down the urban hierarchy both in Europe and North America. By 2000, urban programs explicitly designed to foster gentrification had created small but viable areas of redevelopment in the United Kingdom's old industrial centers. The remote Spanish city of Bilbao was selected as the site for the European branch of the Guggenheim museum, the now famous structure helped start a wave of gentrification that penetrated well away from the city's redeveloped waterfront. In the United States, gentrification has slowly begun to transform small areas of cities near the bottom of the urban hierarchy with row house redevelopments and warehouse conversions in places like Lancaster, Pennsylvania, or the creation of lofts in small downtowns like Bloomington, Illinois. Some critics have cautioned that gentrification has been defined so broadly that if you look for it you can find it almost anywhere, and there are conceptual difficulties with conflating redeveloped houses in a city of 50 000 people with the massive flows of capital that have rebuilt large portions of a number of global cities. Nonetheless, there is an identifiable trend toward both a spatial expansion and intensification of the gentrification over the last 30 years of the 1900s, and the trend appears to be continuing into the new century.

Causes of Gentrification

Early Explanations

The emergence of gentrification throughout the 1960s and 1970s caught the attention of urban scholars for several reasons. It was evident on the urban landscape, it offered some solution to the problems of urban decline, and it seemed to contradict nearly 50 years of urban theory that focused upon the out movement and suburbanization of the middle class. This challenge to the theoretical canon required serious consideration of the causes of gentrification. While the early research on gentrification covered a range of issues from displacement to other social and policy considerations, it was the attempts to theorize gentrification that dominated the scholarly research throughout the 1980s.

The earliest explanations for gentrification developed out of a view that it was a temporary phenomenon rooted in demographic factors present in both North America and Europe. Brian Berry suggested the major recession of the early 1970s followed by large numbers of baby boomers entering the housing market created a significant shortage of housing, rapidly driving up housing prices beyond the means of many members of the large cohort of first time homebuyers. A few of these individuals turned to lower cost housing in declining areas of central cities. These gentrifiers, Berry suggested, would soon leave their central city location and move to more typical suburban locales when their incomes rose, housing became more affordable, and they began to have children. This demographic explanation was echoed by many others and established a general view of gentrification as a highly ephemeral event. As it became clearer that gentrified areas were expanding throughout the 1980s instead of contracting, Berry suggested that gentrification was more an issue of contagious abandonment of inner city housing brought about by overbuilding in suburban housing markets. This substantial shift from a demand side to a supply side explanation has been largely ignored because of Berry's famous contention that gentrified areas were "islands of renewal in seas of decay."

While early explanations of gentrification emphasized demographic factors, many also noticed the correlation between gentrification and emerging high order service centers. Starting in the mid 1970s numerous scholars linked gentrification with an increasing concentration of white collar jobs in the central business district. By the late 1970s, some were trying to empirically evaluate this link. At the same time, a number of other suggestions were made about the root causes of gentrification: high fuel prices, increasing numbers of women in the work force, etc. While all these factors probably played a role in bringing about gentrification, they were frequently offered in an *ad hoc* fashion. Critics of this early phase of explanation accurately pointed to the lack of coherent theories to contextualize the process. However, it is somewhat inaccurate to refer to the explanations as *ad hoc*. Both of the Berry's explanations of gentrification, although mutually exclusive, were consistent with and linked to prevailing neoclassical theories of urbanization (although he claims to have challenged these same theories). It is accurate, however, that most of the early research on the causes of gentrification failed to make explicit links to established urban theory. This situation changed dramatically in 1979 and 1980.

Capital and Culture

With the publication of a short paper dedicated to moving toward a theory of gentrification, Neil Smith introduced what has become known as the rent gap hypothesis. In doing so, he identified gentrification as the leading edge of a process of urban restructuring driven by the demands of advanced capitalism. He suggested links between the rent gap and Marxist theories of urbanization being honed by Henri Lefebvre and David Harvey, and cemented those links in his later work on uneven development. The rent gap hypothesis emphasized the supply side causes of gentrification and focused on movements of capital in the gentrification process. The subtitle of Smith's paper was 'A back to the city movement by capital and not people'. The rent gap theory focused on a cycle of devalorization, or destruction of value, that all urban land is subject to. The five stage cycle suggests that while some land maintains its value, other land proceeds through phases of devalorization, perhaps eventually to abandonment. As devalorization occurs, the actual land rent (termed capitalized land rent) declines as less and less value can be extracted from the land given its current use. At the same time, in a metropolitan area with continued investment in the built environment, or valorization, the potential land rent of the devalorized parcel remains high because of their high levels of accessibility. This difference in capitalized and potential land rent is the rent gap. In land rent terms, gentrification represents the closing of the gap. The rent gap hypothesis stood in stark contrast to the demographic explanations of gentrification because it emphasized the supply of gentrifiable land and housing and not the demand for gentrification growing out of the rise of a particular cohort of urbanites.

The postindustrial city hypothesis formed another new strand in gentrification theory, emerging at the same time as the rent gap. David Ley was the first to fully elaborate the impact of this hypothesis on gentrification in a seminal paper in 1980. Ley emphasized the effect of a new liberal ideology in changing the inner city landscape in Vancouver, British Columbia. As the city's economy was transformed by the increase in high level service functions, a growing group of elite white collar professionals began to assert power in local politics, adopting a liberal ideology that Ley suggests was characterized by a concern with humane governance and emphasized issues involved with the urban esthetic. As this political transition played itself out, it became clear that the new liberal elite were not terribly concerned with issues of social justice, but instead, placed issues of urban livability above all other concerns. The urban orientation of this new class was in part a result of their employment in the urban core and in part a function of the liberal ideology. Ley outlines many of the emphases of Vancouver's liberal movement in a case study of False Creek, an old industrial area that was the subject of city led redevelopment. Ley never used the term gentrification in the 1980 paper, but it was clear that he was outlining a different view of the process of gentrification. He explicitly links gentrification to larger political and economic forces that help create a new class of potential gentrifiers. He also suggest that liberal ideology was so focused upon livability and esthetics that some of its results, including the False Creek, project were uneconomical. In this sense, the liberal ideology explanation of gentrification would seem to be in direct opposition to the rent gap hypothesis, and in some respects, it is.

The debate surrounding these explanations of gentrification continued through the 1980s, and included a lively exchange between Smith and Ley in the *Annals of the Association of American Geographers*. Other gentrification theorists began to portray these explanations of gentrification as being based either in capital or culture. In this sense, the theories surrounding gentrification became a microcosm of larger debates and theorization in geography and other social sciences. Indeed, much of the attention that gentrification received was not due to its empirical extent, or even its effects on working class and poor neighborhoods. Gentrification was understood as a symbolic process representative of sweeping changes in industrialized cities. Theorizing gentrification was a key component in rejecting positivist theories of urbanization and supplanting them with alternative theories. However, the capital and culture debates were not actually about capital and culture. The rent gap hypothesis was often simplified to a deterministic diagram representing the gap between potential and capitalized land rent, but ignoring the politics of neighborhood change and class constitution that drove land rent changes. Conversely, while the liberal ideology explanation certainly contained a cultural component, it also relied heavily upon concepts common in political economy. Both Smith and Ley seemed to move in this direction in later work, and most scholars who wrote about gentrification theory seemed to suggest the need for some combination of both explanations, but often with the clear intent of privileging one of them.

Revanchism and Neoliberalism

By the early 1990s, the so called capital and culture debates had run their course, and gentrification itself was slowed by a recession in Europe and North America. Given the culmination of these two events, it is unsurprising that several scholars began to speculate about the end of gentrification, and processes of degentrification. Once again they raised questions about its usefulness as a concept. The rapid reappearance of gentrification in many global centers, however, silenced many of those questions (even when they raised important and difficult issues). As academic interest renewed in gentrification, the focus changed. Jan Van Weesep suggested that after the experiences in gentrified areas during the 1980s more attention needed to be paid to issues of displacement and the inequalities that gentrification was inscribing in the landscape. The research that followed focused initially upon documenting and eventually upon understanding and theorizing processes of gentrification as they unfolded after the recession of the early 1990s.

Gentrification in the 1990s can be differentiated from early processes in several ways. Oddly, there appeared to be less political resistance to the process in gentrified areas, especially in contrast to the sharp and sometimes violent conflicts of the 1980s. In the United States, part of this difference may be explained by shifts in a range of housing policies. Throughout the 1980s the residential lending industry was deregulated in ways that allowed an increase in both the products that mortgage lenders could offer and the type and characteristics of the lenders themselves. A secondary mortgage market had existed in the United States for decades, and had expanded massively in the 1970s, but in the 1980s private secondary market added to the growth. Historic discrimination in lending had been addressed in a series of legislation throughout the 1970s, which the Clinton administration began to enforce with some vigor. These developments, combined with market pressures, began to open up inner city neighborhoods to mainstream mortgage capital for the first time. The intended result was that low and moderate income individuals would have increased access to mortgages, and this did occur, but these shifts also made gentrifying areas the target of increased mortgage lending. Home buying and renovations in gentrifying areas had traditionally been funded through processes of sweat equity, personal savings, and informal loan networks. While mortgages were made in gentrifying areas, they were difficult to obtain, and were often at unfavorable rates. Access to mainstream mortgage capital, initially in the 1980s and in massively increasing amounts throughout the 1990s helped finance gentrification at a pace that had not been experienced previously. Resistance to gentrification at the local level was handicapped by an increasingly direct link between the process and global capital markets.

The extent of gentrification in many major US cities was significant enough that David Ley's suggestion, made in the early 1980s, that gentrification may one day turn North American cities inside out creating a morphology much more similar to other cities around the globe, seemed plausible. The amount of gentrification in places like Manhattan, Chicago, Boston, Washington, DC, and San Francisco made a significant mark on local politics. Neil Smith borrowed the term revanchism to outline a theory of middle class revenge upon the poor in an attempt to take back the city and make it more palatable for middle class consumption habits. The vigorous policing championed in New York, with its occasional overt violence inflicted upon marginalized groups, effectively pushed the poor and dispossessed out of Manhattan. It also served as a model for other US cities. Similar tactics were adopted in European cities, although initially they were somewhat less aggressive.

Revanchism was quickly linked to a political and economic movement in advanced capitalism, termed neoliberalism. A number of scholars have argued that gentrification is a clear example of neoliberal policies and its expansion during the 1990s was brought about by those policies. Philosophically neoliberalism is a return to a deeply held belief in the efficacy of the free market, albeit with substantial assistance from the public sector that is typically used to guarantee a high return on private investment. Gentrification throughout the 1990s was often aided by local neoliberal regimes that saw it as a free market solution to urban problems. There are numerous examples in the United States, but the United Kingdom may offer the single most sweeping example of neoliberal gentrification policy in the Blair administration's plan for regenerating the country's central cities.

The Process of Gentrification

Stage Theories

There has been significant investigation into the process of gentrification, and a few useful generalizations have been made. Early work suggested a first group of gen trifiers who were risk oblivious and thus were willing to live in neighborhoods that probably lacked many of the typical amenities they would expect, and from a real estate perspective posed significant investment risks. These early gentrifiers often had specific reasons, fre quently economic constraints, that moved them toward gentrification. College students have been identified in this group, but the arts community is often seen as the most prominent participants in the early stages of the gentrification process. Artists' need for studio space, their low incomes, and their counter cultural outlook led them to neighborhoods that were, in real estate terms, off the map. Some research has suggested that gay men and lesbians have also been key actors early in the gentrifi cation process. During the earliest waves of gentrification, the social barriers to living openly as homosexuals in many parts of the city made economically marginal areas attractive as safe havens. In addition, single working mothers, another group with limited housing choices, may have chosen gentrified areas because of their prox imity to the workplace, which helped them maintain the difficult balance between domestic and professional re sponsibilities. Discussion and theorization about the role of these groups introduced an important gender com ponent into gentrification and while it has never been the primary concern of most research on the topic, theories explicating processes of gendering space have been pre sent in work on gentrification for several decades.

These early gentrifiers have often been the subject of debate about their role in the gentrification process. If these groups manage to start the process of gentrification, do they bear some responsibility for the resulting dis placement from later, more intense gentrification? The subject has received much discussion on web forum and electronic mail lists, and the issue has received serious consideration by some members of the subject com munities that are often characterized by progressive pol itics. However, from a policy standpoint, it may be myopic to lay blame on socially marginalized groups for a process that other research has clearly indicated is largely driven by mainstream economic and cultural forces.

Gentrification in the second stage is characterized by risk prone or risk aware in movers. These individuals are attracted to gentrifying areas because of the social diversity and the counter culture trendiness that result from the first stage of activity. The gentrifiers may be professionals with somewhat limited financial resources, but are well aware of the investment risks of purchasing in areas in early stages of the gentrification process. On the other hand, they are just as aware of the potential financial benefits, although most work focusing on stages of gentrification typically emphasizes the cultural aspects of gentrification in its early stages. It is also in this risk aware stage that investors and developers begin to pay attention to the area.

The final stage of gentrification is often referred to as the risk adverse stage. At this point a neighborhood is heavily gentrified and will likely have many of the typical amenities upper middle class urbanites expect in their neighborhoods. In movers, at this point, will pay sub stantial prices for homes and expect that investment to be safe. Gentrified neighborhoods at this point have taken on a particular economic profile and may be solidly middle class, or in some cases an elite urban enclave that is comparable to some of the traditional outposts of urban wealth.

While stage theories of gentrification are useful in understanding the potential forms that gentrification may take, it is difficult to effectively apply them to individual neighborhoods due to the diverse forms that gentrifi cation may take and the incredible range of urban contexts that occurs within. There has been some in vestigation into various types of gentrified areas, but it has typically been constrained to neighborhoods within a single city, or at most selected neighborhoods in cities from one country. The diversity of international contexts poses a significant challenge to anyone attempting to create some form of classification of gentrified neigh borhoods, but the task might provide an *entrée* into more fully understanding the multitude of processes that cre ate and maintain gentrified areas.

Profile of Gentrifiers

Questions about the identity of gentrifiers and their mo tivations have been addressed in a wide array of research projects. The initial view of gentrifiers as a 'back to the city' group was shown to be inaccurate by early empirical studies. By and large, gentrifiers lived in households that were either new, or that had moved from other areas of the city. A more sophisticated picture began to emerge in the 1980s that tied gentrifiers to both liberal ideology and a culture of consumption. More recently, some have argued that gentrifiers should be understood as part of a trans formed middle class who seek to identify themselves through the process of gentrification. Theorists in this vein have invoked Bourdieu's concept of habitus in an attempt to link the structural aspects of gentrification to individual beliefs and dispositions.

Consequences of Gentrification

Displacement

The potential of gentrification to displace lower income residents from a neighborhood seemed obvious very early after the identification of the process. Some of the earliest research focused on displacement, and while the findings were based on relatively small scale studies, the general consensus was that gentrification in a traditional neighborhood setting would result in the displacement of many of the original residents of the neighborhood. Most of the displacement was the direct result of economic issues, but the stress and strain of class differences and neighborhood change were also identified as factors.

More recently, some controversy has arisen over the issue of displacement, as part of a larger movement that seeks to suggest that gentrification is a largely positive phenomenon and the only realistic way of revitalizing many inner city areas. Several studies conducted by Lance Freeman have indicated that gentrification does not directly contribute to displacement of low income individuals. There is a high degree of residential mobility in poorer households and the background rates of mobility are high enough and the typical process of gentrification slow enough that it is impossible to attribute the out movement of low income individuals directly to gentrification. On the other hand, there are a number of case studies that clearly identify incidences of displacement directly caused by gentrification. In addition, there can be little doubt that gentrification can dramatically reduce the housing options for low income people, forcing them to direct their search elsewhere.

Displacement is an on going concern with gentrification and is likely to be the source of a great deal of debate. The debate, however, has tended to focus on the link between gentrification and displacement. In academic research, at least, there is little argument about the severity of the process and the difficulties displacement creates for low income people.

Policy Consequences

The consequences of gentrification reach beyond the neighborhoods in which the process is operating, because it has become the centerpiece of a range of urban policy initiatives. In much of the industrialized world, gentrification has been seen as a possible solution to a range of urban problems. There is little doubt that the massive gentrification of Manhattan and parts of Brooklyn contributed to the strong fiscal position that New York City enjoyed throughout the 1990s and was in strong contrast to the debt ridded metropolis of the 1970s. In the US, the purported fiscal benefits have provided a justification for many cities to subsidize gentrification. While the process has long been subsidized and even initiated by the state in myriad ways, it was not until the 1990s that gentrification became fully linked with large scale redevelopment policies and theories.

In the United States, ideas surrounding the importance of a so called creative class in initiating and expanding economic development enmeshed gentrification and development policies. In the creative class hypothesis, it is the educated, urbane, and highly skilled work force that drives the postindustrial economy. This creative class is often also a gentrifying class and this link has driven a set of neoliberal policies that have attempted to attract the creative class through subsidizing developments and projects that would improve the lifestyle of the urban middle class and attract members of the creative class. It has also driven a series of revanchist policies aimed at removing the poor from the areas that are likely to be most attractive to the creative class. The hypothesis does emphasize the importance of diversity and tolerance in local governance, but this emphasis is rarely translated to policy.

One of the most direct links between gentrification and public policy is the Urban White paper issued by the UK government in 2002. Formally titled, 'Our towns and cities – The future: Delivering an urban renaissance', the paper was based on a government commissioned study of urban problems, 'Towards an urban renaissance'. Both the report and the ensuing policy recommendations place heavy emphasis on the need for the regeneration of the urban core, particularly in the country's regional metropolises. The document places emphasis on public/private partnerships in creating gentrification and makes scant mention of a range of social issues that plague the urban cores of the country's most troubled cities. In short, the document poses gentrification and commercial redevelopment as a solution to a vast range of urban issues. As Loretta Lees has observed, however, the term is not used and renaissance, regeneration, and sustainability are substituted, thus avoiding the deeply imbedded class constitution of the gentrification process. The discourse makes the intent clear. The new urban core will be a place for the British middle classes. Given the influence of the central government in the United Kingdom, the implementation of these policies, as was intended, has triggered gentrification often as a major facet of central city or waterfront redevelopment.

Perhaps, one of the more unusual effects of gentrification has been its influence on public housing policy in the United States. In the early 1990s, after a report on the troubling conditions in many of the country's public housing projects, the HOPE VI program was launched to improve the conditions in the nation's worst public housing. HOPE VI was a multidimensional program and its origins were complex, but there is a distinct link with several major aspects of the program and the intense gentrification of many American inner cities. There had been a growing concern about the effects of concentrated poverty in American cities and public housing was seen as one of the greatest contributors to the issue of concentration. The results of several decades of gentri fication had led to situations in some US cities where public housing projects and gentrified neighborhoods were sitting side by side. The HOPE VI program had adopted a policy of encouraging income mixing when possible, and the presence of gentrification in some lo cales made it possible. Chicago provides one of the most stunning examples, where several of the country's largest projects were slated for demolition and replacement with a mix of as much as 50% market rate housing, and only vague plans for relocating a majority of public housing residents. Gentrification then has created a new and troubling option for policymakers who seek to create a new form of urbanization that excludes all but the privileged.

See also: Gentrification, Rural; Habitus; Housing; Land Rent Theory; Ley, D.; Neoliberalism, Urban; Regeneration to Renaissance; Smith, N.; Uneven Development; Urban Policy; Waterfront Development.

Further Reading

Atkinson, A. and Bridge, G. (2005). *The New Urban Colonialism*. London: Routledge.

Butler, T. and Lees, L. (2006). Super gentrification in Barnsbury, London: Globalisation and gentrifying global elites at the neighbourhood level. *Transactions of the Institute of British Geographers* 31, 467 487.

Duany, A. (2001). Three cheers for gentrification. *The American Enterprise* 12, 36 39.

Freeman, L. (2006). *There Goes the 'Hood': Views of Gentrification from the Ground Up*. Philadelphia, PA: Temple University Press.

Hackworth, J. and Smith, N. (2001). The changing state of gentrification. *Tijdschrift voor Economische en Sociale Geografie* 92, 464 477.

Lees, L., Slater, T. and Wyly, E. (2007). *Gentrification*. New York: Routledge.

Ley, D. (1980). Liberal ideology and the post industrial city. *Annals of the Association of American Geographers* 70, 238 258.

Ley, D. (1996). *The New Middle Class and the Remaking of the Central City*. Oxford: Oxford University Press.

Rose, D. (1989). A feminist perspective on employment restructuring and gentrification: The case of Montreal. In Wolch, J. & Dear, M. (eds.) *The Power of Geography*, pp 18 138. Winchester, MA: Unwin Hyman.

Smith, N. (1979). Toward a theory of gentrification: A back to the city movement by capital not people. *Journal of the American Planning Association* 45, 538 548.

Smith, N. and Williams, P. (eds.) (1986). *Gentrification of the City*. Boston, MA: Allen and Unwin.

Smith, N. (1996). *The New Urban Frontier: Gentrification and the Revanchist City*. London: Routledge.

Wyly, E. and Hammel, D. (1999). Islands of decay in seas of renewal: Urban policy and the resurgence of gentrification. *Housing Policy Debate* 10, 711 771.

Zukin, S. (1989). *Loft Living: Culture and Capital in Urban Change* (2nd edn.). New Brunswick, NJ: Rutgers University Press.

Relevant Website

http://members.lycos.co.uk/gentrifications/
Gentrification Web maintained by Tom Slater, Lycos.

Gentrification, Rural

M. Phillips, University of Leicester, Leicester, UK

Glossary

Exurbia A phrase widely used in North America to refer to settlements which although located in the countryside are inhabited by residents whose lives are principally linked to urban settlements.

Hegemony A term widely associated with the Italian Marxist Antonio Gramsci and used to refer to the exercise of power or influence without the use of force or sanction. It is often used to suggest that certain social groups can exercise power through cultural means, such as getting people to accept the exercise of power as legitimate or unavoidable or getting people to adopt and internalize the views of the dominant group. The concept of counterhegemony is also often used to refer to situations where groups reject the culture of dominant groups and/or express different cultural values and practices.

Petite Bourgeoisie French terms widely used to refer to the people who own a business in which they work and which employs only a small number of employees, many of whom might well be family members. Businesses are classically seen to encompass shopkeepers and craft trades, but the term also encompasses family farms and a range of self-employed workers. It is sometimes also used to refer to a middle class that adopts conservative values, consumption habits, and tastes.

Postproductivist Countryside A term that is used to signal some break or difference from a countryside dominated by expansive capitalist production centered within agriculture. It is seen to emerge from a shift in agriculture away from a focus on expanding output and/or the release of land/properties from agricultural use. There has been considerable debate as to the value of the term in characterizing rural areas in contemporary advanced capitalist societies given the continuities in agricultural production and because of uncertainty as to the precise contours of any emergent postproductivist countryside.

Rent Gap A term used to refer to situations where the actual payment, or rent, for a piece of land or property is lower than could be received under some alternative usage. It is suggested that at the point of capital investment, actual and potential rents are close to each other, because investors will seek to maximize their returns, but that thereafter, there is a tendency for actual rents to fall and potential rents to rise, creating rent gaps, which may be of such extent that redevelopment of the area becomes profitable. It is in these situations that capital investment is seen to flow in, establishing gentrification.

Reproductive Labor A phrase that is used to refer to the everyday work people do that enables them to be able to participate in the paid economy. It is commonly seen to encompass domestic work, through which people reproduce the conditions of their own existence (by, for instance, creating food to sustain themselves) and also provide childcare, which can be seen to be reproducing future generations of workers.

Service Class A term often used to refer to people in managerial and professional occupations, and also sometimes used interchangeably with the term new middle class. A range of debates have surrounded the meaning and value of term, with a series of definitional criteria having been proposed, including receipt of salary rather than wages, performance of intellectual and administrative work, autonomy from supervision, and the presence of trust. It has also been suggested that important divisions may exist within the service class, and indeed that there may be no overall coherence to the class.

Introduction

Gentrification has long been the subject of research and debate within urban geography, with it often being described as an inherently urban process. It has, however, also been identified in rural areas, and has had almost as long a history in rural studies as in urban. The term is often traced back to Judith Glass, who in a study of London suggested that a process of gentrification was occurring, involving both the refurbishment of properties in an area, and also changes in its class composition through both in movement and displacement. Although there has subsequently been considerable debate about the meaning of gentrification, and the relative emphasis that should be given to in migration, refurbishment, and social displacement, Glass' definition is still incorporated within many recent definitions of gentrification, even though many have sought to add additional features, often of the basis that the nature of gentrification has itself changed.

Although Glass used the term gentrification in 1964, it was really only in the 1970s and early 1980s that it became more widely used to refer to change occurring across a range of urban, and indeed rural, spaces. Gentrification in the countryside has not, however, been subject to the same level of attention as that paid to gentrification in urban centers, with relatively few rural researchers using the term and even then often providing little by way of accompanying commentary or justification. This is despite growing popular reference to rural gentrification, particularly in North America, and evidence that the use of the term in rural studies is becoming more akin to that within urban arenas, where a series of differing theorizations of gentrification have emerged. By the early 1990s, it was possible to identify at least five distinct, albeit overlapping, approaches to gentrification studies, each of which has some applicability to rural studies.

Gentrification as Class Colonization and Displacement

Glass' initial commentary clearly highlighted notions of middle class in migration and working class displacement, and these dual changes have also been readily identified in rural as well as urban studies. From the late 1960s, a series of studies identified middle class in migration into rural areas of Britain and documented a series of associated displacements. Often the term gentrification was not used, with studies often phrased using terms such as incomer/local and urbanization/counterurbanization, whereby counter urban migration was seen to produce urban incomers who displace locals and their rural way of life, via mechanisms such as outbidding rural locals in housing markets and dominating local political and social organizations. Similar comments could be made with respect to rural studies in many other countries. In North America, for instance, the foci of many rural studies have been on population and economic growth, encapsulated by notions such as rural development and rural renaissance, and on the impact of urban to rural migration, particularly in so called exurbia settlements, with very few studies making connections to concepts such as gentrification or class colonization/displacement.

Even studies that refrained from use of the term gentrification and focused on the spatialized identities of incomer–local/rural–urban sometimes did contain recognition that these might involve in migration of middle class people, occurring at the expense of working class people. Such connections became much more explicit with the emergence of critical/political–economy perspectives in rural geography and associated rural studies. Work by people such as Paul Cloke and Jo Little began to use to term rural gentrification, viewing it in terms of middle class in migration occurring at the expense of other, lower class groups, as well as considering the degree to which settlement planning practices might foster gentrification in certain rural locations.

The equation of incomer with middle class and local as working class was clearly open to criticism. First, many of those considered local are often seen to be far from working class, as with, for instance, farmers who are in Britain widely perceived as wealthy landowners and/or middle class small scale employers. Likewise, there is clear evidence that not all incomers are members of an affluent middle class. Second, although there is considerable contestation over the meaning of class, with a wide range of seemingly incommensurable concepts being advanced, contemporary class structures are widely seen to be more complex than implied by the dichotomous middle class/working class division. In the late 1980s, studies began to speculate that there might be a series of middle classes, or middle class fractions, moving into rural areas, and hence rather than talk about an incoming middle class, it might be better to speak in terms of incoming middle classes. Related to this, given that middle class in migration may have been occurring for many years, it may be that rather than seeing rural social change in terms of a middle class replacing a working class, as the term gentrification might seem to imply, in many instances it might be more appropriate to speak in terms of one middle class, or middle class fraction, replacing another.

The suggestion that rural areas might be being colonized by a series of middle classes has been subject to some empirical investigation. Keith Hoggart, for example, explored the growth of three middle class groups in England between 1971 and 1991, suggesting that quite distinct patterns of growth were apparent, with a service class of professionals and managers growing greatest in the Southeast, in both rural and non rural areas, while the number of clerical workers in England remained relatively static, and the size of a *petite bourgeoisie* declined overall, although they were a significant presence in some remote rural areas. These findings were broadly confirmed in a later analysis using the 2001 census by Martin Phillips, although here attention was also focused on differences between professionals and managers as well as between these groups and a *petite bourgeoisie*. Taken together, these studies suggest that there is a complex geography to patterns of middle class colonization, an impression confirmed by local scale studies, such as Darren Smith and Deborah Phillips' study of the hill top and valley locations around Hebden Bridge, Yorkshire, England. They argued that the former locations were undergoing gentrification by high income legal, business, and finance professionals, while people moving into town and surrounding valley locations tended to work in teaching and other, what might be deemed lower professional, occupations.

Connections can be drawn between studies of rural gentrification and debates within class analysis, economic geography, and studies of the 'geography of gentrification' within urban studies. Smith and Phillips, on the other hand, focus attention on the attraction that particular locations hold for incoming gentrifiers, proposing that the term rural gentrification be replaced by the term rural greentrification, to reflect the significance that 'green space' plays in stimulating in migration. Sylvain Paquette and Gérald Domon develop quite similar arguments in their studies of population flows into southern Quebec, suggesting that rural areas with open views and remoteness attract wealthy householders seeking an esthetic relationship with rurality, while rural areas lacking scenic attraction become sites for low income in migration and residential property deterioration.

As Eliza Darling has commented, the emphasis given to environmental influences on rural in migration shows clear parallels with so called consumption side theories of urban gentrification in that both see consumer preferences as the key determinant of gentrification processes, and indeed, in some cases posit these as constitutive of differentiation between rural and urban gentrification. It is also a focus that connects rural gentrification to a long established, and much more widely expressed, research interest in migrational motivations. However, while many studies of residential migration to rural areas have highlighted the significance of the perceived naturalness of the countryside, as many, if not more, have also highlighted other constituents in the desire for country living, such as its perceived sociability and distance from social ills such as crime, poverty, homelessness, and conflict. Studies of rural gentrification employing a consumption perspective can clearly draw on studies of migrational motivation and the processes underpinning counterurbanization, even though, as noted earlier, many of these studies have been reticent about adopting concepts of class colonization and displacement often seen as central to notions of gentrification.

Besides linking to studies of migration, rural gentrification also connects quite directly to research on social marginalization and exclusion. Rural researchers and commentators have long highlighted how incursion of affluent residents who commute to work can adversely impact rural retailing, transport, welfare services and housing availability; it is also suggested that these incoming residents can direct the course of development in rural areas by exerting political, intellectual, and cultural hegemony over other classes and social groups in the countryside through a combination of direct and indirect influence on the rural housing market and planning systems. While popularly often described in terms of NIMBYism (NIMBY, not in my backyard), a more nuanced argument is provided in the concept of hegemonic blocs outlined by Paul Cloke and Mark Goodwin in 1992 and subsequently incorporated into a stage model of gentrification (see **Table 1**).

Gentrification as Capital Investment

While gentrification has been readily linked to issues of population movement, this emphasis has been questioned, notably by Neil Smith who argued that urban gentrification should not be viewed as 'a back to the city movement of people'. He suggested that this view was mistaken because not only were many inner city gentrifiers actually moving from other urban areas, but also, and more significantly in his view, because gentrification was a process of capital rather than population movement. Smith developed what has widely been characterized as a production side theory of gentrification, in which gentrification is seen as a manifestation of the uneven circulation of capital, and specifically the product of processes of devalorization and revalorization in which so called rent gaps emerged where current use of an area yields financial returns significantly lower than those available under some alternative land usage. The existence of such gaps means that there are opportunities for higher level profits for those people or institutions that invest capital – in the broad sense of products and labor power as well as money – in this new use. Gentrification, for Neil Smith, is a process of property investment, and gentrifiers are in effect property developers who reap the rewards of converting properties from some lower value activity toward one associated with middle class residences.

Although this concept has been widely explored and debated in the urban context, until recently little use of it was made within rural contexts. This is despite there being discernible rural parallels, such as studies highlighting rapid gains made from buying and selling houses in the countryside, and rural residents undertaking substantial building work on their property to heighten its exchange value. One particularly prominent instance of rural refurbishment in Britain is the barn conversion, and while this is by no means the most empirically extensive instance of rural refurbishment it does clearly illustrate processes of devaluation and revalorization, in that conversion of these properties in many cases occurred after they fell out of agricultural use, although they are now such sought after properties that so called neo barns have been built, which simulate a converted barn but are themselves new builds. The barn conversion, and rural gentrification more generally, connects to the concept of the postproductivist countryside which, in at least some interpretations, highlights de valorization of land and buildings with respect to agricultural production and their revalorization with respect to other, more consumption orientated, capital networks, including those associated with residential development.

Processes of devaluation are not restricted to just agricultural land and properties but can span a range of economic sectors. Likewise, revaluation is not restricted to the flow of capital into residential development but can encompass retail, leisure, and other facilities. This has clear parallels with Neil Smith's concept of 'gentrification generalized' – whereby gentrification is seen to have evolved in scale and diversity such that it often now

Table 1 A stage mode of rura gentr ficat on ncorporat ng flows of cap ta , agenc es of gentr ficat on, and cu tura or entat ons

Types of capital and associated practices of gentrification			*Phase in the circuit of capital*	*Type of agency gentrification*	*Cultural orientation*
Labor/products	*Physical infrastructure*	*Finance*			
Dec ne of agr cu tura workforce	Deva or zed propert es	Centra zat on of cap ta ownersh p, product ve cap ta nvestment n agr cu ture	Deva or zat on of cap ta	Mu t p e	Deva or zat on of ong-stand ng hegemon c b ocs
Rebu ds and restorat ons	Refurb shment of rust c propert es	nvestment n res dent a propert es	D rect cap ta nvestment	nd v dua zed	Gentr fiers w th h gh degree of countercu tura or entat on
Product on of gentr ficat on products – .e., rep ca products and bu d ngs	Creat on of ensemb e of fac t es to create a rura scene'	Creat on of oca rea estate market; nvestment n consumpt on serv ces	ntens ficat on of cap ta	nd v dua zed but more empowered	Low ncome res dents and potent a res dents
Gentr fied festy e magaz nes; v age pub cat ons	Demarcat on of h stor ca bu d ngs, vernacu ar house sty es, conservat on zones	Promot on of area as des rab e res dent a and perhaps tour st and reta dest nat on; fict ous cap ta (cred t) encourages further expans on	Symbo zat on of cap ta	Nat ona and mu t nat ona firms	ncreas ng use of commod fied cu tura textures
C rcu at on of deas and personne	Pressure for gentr fied new bu d	Profess ona zed agenc es of rura gentr ficat on emerge	D ffus on and corporat zat on of cap ta	Large corporate and spec a st cap ta	Emergence of new gentr fied hegemon c b oc

From Phillips M (2004) Other geographies of gentrification *Progress in Human Geography* 28 5–30 involves the formation of cross sectoral developments involving retailing, leisure, and entertainment as well as housing – and with Sharon Zukin's concept of gentrification as a 'sociospatial complex' formed out of a series of flows of labor, commodities, and finance, a concept which has been explicitly linked into rural gentrification (see **Table 1**).

Eliza Darling also adopts a production side focus in her study of wilderness gentrification, suggesting that notions of devaluation and revaluation are relevant to areas such as the Adirondack State Park, USA, which has seen forestry, mining, and agriculture become supplanted by tourism and public service related activities. Darling suggests that the wilderness character of this area itself relates to broad scale devalorization of the region as a consequence of land ownership structures and state conservation regulations acting to discourage large scale capital investment. The consequence of this was an underutilization of potential ground rent and a preservation of landscape and habitats which subsequently become highly valued by tourists, whose desire to visit the area eventually stimulated capital investment in properties, which in turn has fostered the displacement of earlier residents. Work in North America, Australasia, Northern and Southern Europe, and South Africa has indeed suggested that tourist activity may be a key driver of gentrification in rural areas, highlighting a further stream of capital investment to add to those of retailing, recreation and residence which have come to be recognized in both rural and urban gentrification studies.

Capital centered accounts of gentrification have not gone uncontested, with at least two lines of debate of relevance to rural studies emerging. First, attention has been paid to the extent to which it implies that economic motives underpin gentrification activities, an issue which adds a further dimension to studies of migrational motivation that have long been a concern within rural studies. Second, it has fostered consideration of the agents involved in gentrification, as in Martin Phillips' stage model of gentrification (**Table 1**) and his earlier discussion of gentrification in the Gower Peninsula, which highlights both the use of professional agencies, such as architects, builders, building societies and estate agents, in the gentrification process and also the presence of so called marginal gentrifiers who rely largely on their own labor, or sweat equity, to gentrify their properties. This last group has also been invoked by Paul Chaney and Ken Sherwood in a study of changes following the sale of former council houses in Northamptonshire.

Agents of Gentrification and Social Difference

In urban studies, consideration of agents of gentrification emerged not only from production side theorizations but also within so called consumption side perspectives. Examination of locational decision making by incoming householders, for instance, revealed a range of motivational influences, which not only questioned the profit motivation arguably implied in the gentrifier as developer concept, but also suggested that residential decision making might be influenced by issues such as access to retail, leisure, and entertainment facilities. This in turn was suggestive that urban gentrifiers were relatively young and perhaps living alone or in childless households and raised the issue of whether there was a need to disaggregate the notion of gentrifier beyond the simple designation as middle class, or indeed from a particular middle class, and recognize that a variety of forms of social difference might be of significance in the formation of gentrification. It was argued, for instance, that gentrification might relate to demographic as opposed to class change, with changes in marriage and child rearing patterns both leading to the creation of more single person households and childless couples.

There are clear rural correlates to these arguments. Counterurbanization studies, for instance, not only recognized a class dimension to rural in migration, but also, and often more directly, identified other social differences as significant. Age relations, for example, were often seen as important, with counterurbanization being associated not only with gentrification but also with geriatrification, a term used to refer to situations where there is an increasing proportion of elderly in an area's population. Studies in the US also frequently highlighted the racial character of rural migration streams, an issue which is being increasingly recognized in British and European contexts, while notions of national identity have long been prominent in studies of migration to rural Wales, Scotland, and France. The rise of postmodernism within rural studies promoted recognition of neglected difference and otherness, concerns that were explicitly drawn into the study of rural gentrification in Darren Smith and Louise Holt's study of lesbian gentrifiers.

The rise of postmodern perspectives also invited reflection of the general meaning and value of the term gentrification. Recognition of social differences such as age, race, nationality and sexuality raised questions as to the degree to which gentrification should, by definition, be seen to involve processes of class recomposition. Responses have included abandonment of the term or dilution of its class connotations, to suggestions that attention needs to be paid to how other social differences are interwoven with socio economic differences in the constitution of gentrification. Questions have also been asked about the extent to which gentrification is a modernistic grand concept or theory, lumping together a range of events, practices, and relations that should be viewed as quite distinct. While these latter arguments may lead some people to reject the concept of gentrification, there have also been attempts to adopt more postmodern/interpretive approaches to the study of gentrification.

Gentrification, Gender and Reproductive Labor

Recognition of difference is of significance to, and has been very much fostered by, considerations of gender, with gentrification studies being clearly influenced by feminist as well as postmodern perspectives. Within studies of urban gentrification, three distinct lines of ar guments relating to gender can be discerned.

First, gentrification has been seen to be, at least in part, an outcome of rising female participation in paid employment which has led to increasing numbers of both high income, single women households and dual income households who can afford to engage in gentrification activities. Second, the 'back to the city movement' of people commonly associated with urban gentrification has been viewed as a response to increased pressure placed on the performance of reproductive labor by the rise of female paid employment, as a consequence of which householders have sought to live in locations which minimize journey to work costs and facilitate the substitution of market produced commodities for household labor. Third, it has been argued that women, and some other gentrifying groups, may be drawn to areas where there is a perceived lessening or a dilution of patriarchal attitudes and relations.

Each of these arguments has an explicit urban dis position, with research pointing to urban areas as having higher rates of female paid employment; greater acces sibility, service, and retail provision levels; and lessening patriarchy. The three aspects of gender, however, still have rural relevance. Performance of reproductive labor may, for instance, be impacted by rural gentrification given that studies have suggested that rural in migration is often stimulated by the onset of child rearing, with many households moving when they had, or were in tending to have, children. While a move to countryside may be perceived as good for children, it is often ac companied by female disengagement and/or dislocation from the paid labor market, and a spatialized and gen dered separation of productive and reproductive labor. While women may formally stop paid work for only a short time after childbirth, many return on a part time basis and/or take a new job that can be more easily in tegrated into childcare. These jobs can be difficult to find in rural areas, and as a consequence, women who intend to return to paid employment may not do so, or else do it by forsaking their previous careers. The demands of re productive labor may hence create the strong and highly gendered class asymmetry that has been observed within rural households, it being suggested that a feminized service proletariat might be emerging in the countryside alongside the migration of affluent male service and other middle classes.

Transport and service provision in rural areas can be considerably lower than that found in urban areas, which may well act to heighten disengagement and dislocation. On the other hand, rural areas are often perceived to have high levels of personal and community service provision that might help ameliorate these tendencies, and indeed may also form part of the attractiveness of rural localities. However, community and personal sup port services are often both conceived and practiced as feminized activities, features that stem from, and re inforce, patriarchal social relations and identities. Rural gentrification may hence be in part sustained by, and itself help sustain, a patriarchal rural gender regime or order, although the studies of Hebden Bridge by Darren Smith, Deborah Phillips and Louise Holt have high lighted how rural locations can be sites for gentrification because they are perceived as localities where non conventional household and partnering relations are widely practiced. Such findings, along with the identifi cation of the significance of childcare in the locational decision making of some urban gentrifiers, and detailed examination of the class composition of households across inner city, suburban and rural locations, raise questions as to the significance of urban/rural differences as opposed to differences within urban and rural space.

Gentrification and Culture

Preceding discussion has largely been framed in terms of rural studies being able to draw upon ideas circulating within studies of urban gentrification. There is, however, one area of contemporary discussion of gentrification, its relationships to culture, where it can be argued that rural studies have been somewhat more advanced, at least in temporal terms, than their urban counterparts.

From the late 1980s, for instance, considerable atten tion has been paid to eliciting how cultural attachments to the countryside fuel rural in migration. While much of this discussion was not framed with regard to gentrifi cation, Paul Cloke and Nigel Thrift's discussion of middle class fractions did involve consideration as to whether these might connect to different attachments to the countryside, issues that have been followed up in work considering how different rural identities and life styles might both reflect and actively constitute particular class relations. Early emphasis was given to the role that consumption plays in the establishment of social position, with ideas of cultural capital and distinction, advanced by Bourdieu, being employed. Subsequently, and as with some urban gentrification studies, use has been made of

Bourdieu's concept of habitus, as in Smith and Phillips, discussion of hill top and village gentrification. However, critical comments have also been made about this concept in rural studies, it being argued that when linked to concepts such as cultural capital, culture tends to become reduced to the pursuit of personal advantage. Studies by Paul Cloke, Nigel Thrift, Martin Phillips, John Holmes and Kate Hartig have all suggested that less reductive concepts of culture need to be employed. Attention has been paid to the practices through which the countryside comes to be represented as a place of middle class desire and habitation, and how culturally constructed notions of competence and morality are used in everyday life to devalorize the practices of gentrification.

Concluding Comments

While rural gentrification has attracted much less attention than its urban counterpart, the two concepts can be seen to exhibit clear parallels, as well as some points of difference. The extent of the parallels between the terms raises at least three important questions. The first concerns whether differentiation of gentrification into urban and rural is warranted. Eric Clark is one of the few people explicitly arguing that it is not, although there has been growing recognition that gentrification is appearing in places other than the inner city. This extension is seen by some to stretch the concept of gentrification close to, if not beyond, its meaningful limit. On the other hand, conceptions of gentrification within urban studies have been wide ranging and can each be quite readily applied to changes occurring in rural space. The degree of conceptual overlap itself raises a second question, namely, whether studies of rural gentrification can simply draw on concepts circulating in urban studies, or whether there are issues that are specific to, or at least are of greater significance within, rural rather than urban spaces. One issue that has been identified as being of particular, although not unique, rural significance is the role of nature or green space in gentrification. On the other hand, it has also been argued that gentrification has become generalized, although this is often framed as top down, urban out dispersal, a viewpoint which can be questioned for overlooking that studies of rural gentrification have almost as long a history as do those of urban gentrification. This is not to say that there may not be processes of serial reproduction of gentrification across space, but rather that attention also needs to be paid to whether gentrification may emerge from spatially discrete processes of change. Finally, questions have emerged about the degree to which differences in gentrification can be mapped onto broad spatial categories such as urban or rural, or whether more complex geographies of gentrification are required, which recognize differences within and between categorizations of rural and urban space.

See also: Ageing and Mobility; Ageism and Age; Anti-Urbanism; Belonging; Census Geography; Cloke, P.; Counterurbanization; Culture; Demography; Dwelling; Economic Development, Rural; Ethnicity; Feminism/Feminist Geography; Gated Communities/Privatopias; Gender and Rurality; Gender in the City; Gentrification; Greenfield Development; Habitus; Home; Homelessness; Homelessness, Rural; Housing; Housing, Rural; Identity and Otherness, Rural; Land Rent Theory; Leisure; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Migration; Multicultural City; Patriarchy; Population Geography; Post-Productivist and Multifunctional Agriculture; Postmodernism/Postmodern Geography; Poverty, Rural; Race; Regeneration to Renaissance; Rural Communities; Rural Geography; Rural Populations; Second Homes; Services, Rural; Smith, N.; Social Class; Social Geography; Suburbanization; Tourism, Rural; Tourism; Transport and Social Exclusion; Transport, Rural; Urban Habitats/Nature; Whiteness.

Further Reading

Agg, J. and Phillips, M. (1998). Neglected gender dimensions of rural social restructuring. In Boyle, M. & Halfacree, K. (eds.) *Migration into Rural Areas: Theories and Issues*, pp 252 279. London: Wiley.

Chaney, P. and Sherwood, K. (2000). The resale of right to buy dwellings: A case study of migration and social change in rural England. *Journal of Rural Studies* 16, 79 94.

Clark, E. (2005). The order and simplicity of gentrification A political challenge. In Atkinson, R. & Bridge, G. (eds.) *Gentrification in a Global Context: The New Urban Colonialism*, pp 256 264. London: Routledge.

Cloke, P. (1979). *Key Settlements in Rural Areas*. London: Methuen.

Cloke, P. and Goodwin, M. (1992). Conceptualizing countryside change: From post Fordism to rural structured coherence. *Transactions of the Institute of British Geographers* 17, 321 336.

Cloke, P., Phillips, M. and Thrift, N. (1995). The new middle classes and the social constructs of rural living. In Butler, T. & Savage, M. (eds.) *Social Change and the Middle Classes*, pp 220 238. London: UCL Press.

Cloke, P. and Thrift, N. (1987). Intra class conflict in rural areas. *Journal of Rural Studies* 3, 321 333.

Cloke, P. and Thrift, N. (1990). Class change and conflict in rural areas. In Marsden, T., Lowe, P. & Whatmore, S. (eds.) *Rural Restructuring*, pp 165 181. London: David Fulton.

Darling, E. (2005). The city in the country: Wilderness gentrification and the rent gap. *Environment and Planning A* 37, 1015 1032.

Ghose, R. (2004). Big sky or big sprawl? Rural gentrification and the changing cultural landscape of Missoula, Montana. *Urban Geography* 25, 528 549.

Glass, R. (1964). *London: Aspects of Change*. London: MacGibbon and Kee.

Hoggart, K. (1997). The middle classes in rural England, 1971 1991. *Journal of Rural Studies* 13, 253 273.

Holmes, J. and Hartig, K. (2007). Metropolitan colonisation and the reinvention of place: Class polarisation along the Cessnock Pokolbin fault line. *Geographical Research* 45, 54 70.

Little, J. (1987b). Rural gentrification and the influence of local level planning. In Cloke, P. (ed.) *Rural Planning: Policy into Action?* pp 185 199. London: Harper and Row.

Paquette, S. and Domon, G. (2001). Trends in rural landscape development and sociodemographic recomposition in southern Quebec (Canada). *Landscape and Urban Planning* 55, 215 238.

Paquette, S. and Domon, G. (2003). Changing ruralities, changing landscapes: exploring social recomposition using a multi scale approach. *Journal of Rural Studies* 19, 425 444.

Parsons, D. (1980). *Rural Gentrification: The Influence of Rural Settlement Planning Policies. Department of Geography Research Paper No. 3*, Brighton: University of Sussex.

Phillips, M. (1993). Rural gentrification and the processes of class colonisation. *Journal of Rural Studies* 9, 123 140.

Phillips, M. (2002b). The production, symbolisation and socialisation of gentrification: A case study of two Berkshire villages. *Transactions of the Institute of British Geographers* 27, 282 308.

Phillips, M. (2004). Other geographies of gentrification. *Progress in Human Geography* 28, 5 30.

Phillips, M. (2007). Changing class complexions in and on the British countryside. *Journal of Rural Studies* 23, 283 304.

Phillips, M. (ed.) (in press) *The Gentrified Countryside*. Aldershot: Ashgate.

Phillips, M., Page, S., Saratsi, E., Tansey, K. and Moore, K. (2008). Diversity, scale, and green landscapes in the gentrification process: Traversing ecological and social science perspectives. *Applied Geography* 27, 54 76.

Smith, D. (2002a). Extending the temporal and spatial limits of gentrification: A research agenda for population geographers. *International Journal of Population Geography* 8, 385 394.

Smith, D. (2002b). Rural gatekeepers and 'greentrified' Pennine rurality: Opening and closing the access gates? *Journal of Social and Cultural Geography* 3, 447 463.

Smith, D. and Holt, L. (2005). Lesbian migrants in the gentrified valley and other geographies of gentrification. *Journal of Rural Studies* 21, 313 322.

Smith, D. and Phillips, D. (2001). Socio cultural representations of greentrified Pennine rurality. *Journal of Rural Studies* 17, 457 469.

Smith, N. (1979). Toward a theory of gentrification: A back to the city movement by capital not people. *Journal of the American Planners Association* 35, 538 548.

Smith, N. (2002). New globalism, new urbanism: Gentrification as global urban strategy. *Antipode* 34, 428 450.

Visser, G. (2004). Second homes and local development: Issues arising from Cape Town's De Waterkant. *Geojournal* 60, 259 271.

Walker, P. and Fortmann, L. (2003). Whose landscape? A political ecology of the 'exurban' Sierra. *Cultural Geographies* 10, 469 491.

Whitson, D. (2001). Nature as playground: Recreation and gentrification in the Mountain West. In Epp, R. & Whitson, D. (eds.) *Writing off the Rural West: Globalization, Governments, and the Transformation of Rural Economies*, pp 145 164. Edmonton: University of Alberta Press.

Zukin, S. (1990). Socio spatial prototypes of a new organization of consumption: The role of real cultural capital. *Sociology* 24, 37 56.

Geocomputation

M. Birkin, University of Leeds, Leeds, UK

Glossary

Abduction A process of scientific reasoning in which hypotheses are generated to best explain the available data.

Cyberinfrastructure Latest generation computational technologies which support the storage, manipulation, sharing, analysis, and visualization of data.

E-Science The development and testing of scientific theories using computational methods: broadly the European equivalent of cyberinfrastructure.

Grand Challenge A fundamental problem in science or engineering which is amenable to solution by computational methods.

Grid Computing An approach to computation which involves the combination of multiple processors into a grid or network.

Microsimulation A research methodology which aims to model the behavior of individuals over time.

Introduction

The first use of the term 'geocomputation' in the lexicon of human geography can be traced to the first of a series of international conferences on that topic in Leeds in 1996.

Despite its relative novelty, the term geocomputation has already been defined in at least three different ways by its advocates:

1. A crusading definition has been put forward by Openshaw, who seeks to establish nothing short of a new set of approaches – a 'computational paradigm' – for doing human geography. In this definition, geocomputation takes geographic information science (GIS) and moves it beyond the traditional emphasis on gathering and representation of data toward a focus on analysis, modeling, and finding solutions to real world problems. Geocomputation seeks to combine spatial data, computational technologies, and high performance computing (HPC) hardware. In this view, the subclauses of the term are deliberately capitalized in order to emphasize the combination of spatial data and analysis methods with novel computational approaches.
2. An epistemological definition is preferred by Helen Couclelis, focusing in essence on the prior existence of a scientific discipline – computation – and applications within a geographical environment. In this view, geocomputation becomes a universe or 'grab bag' of geographical analysis methods applied to spatial problems. The problem of articulating a research agenda for geocomputation can be pursued in relation to the tools and frameworks of computation as a parent discipline.
3. Paul Longley prefers an ostensive definition, in which the subject matter of geocomputation is defined according to the conduct of its researchers and practitioners. This view places a much stronger emphasis on the continued development of GISs and techniques, in contrast to Openshaw's emphasis on applications and real world problem solving. We return to this discussion later, pending a more detailed outline of the subject matter of geocomputation.

Six important features of the approach will be discussed in further detail below: inductive emphasis, spatial orientation, technical novelty, computational intensity, policy focus, and multidisciplinarity.

Inductive Emphasis

The availability of increasingly rich spatial datasets provides an important part of the rationale for geocomputation. Indeed there is a clear association between approaches such as the cluster hunter (i.e., geographical analysis machine, or GAM) and the geographical explanations machine (GEM) and the computational discipline of data mining, in which a barrage of statistical techniques are applied to large datasets in the search for patterns. According to the philosophy of this approach, only after patterns have been detected does it become helpful to articulate and refine potential hypotheses which might explain these observations. Such depths of inductivism have not met with universal approval within the geocomputation community. Paul Longley quotes with an apparent mixture of admiration and horror the Openshaw dictum that "prior knowledge or ignorance was rendered equally irrelevant" in the search for patterns in spatial data. Mark Gahegan has perhaps done most to try and reconcile such rampant inductivism with established approaches to the scientific method, arguing for a process of map based abduction in which the process of data exploration, pattern extraction, and theory generation are combined together. Penn State's GeoVista software platform therefore provides an environment for the exploration of spatial datasets through large scale visualization together with spatial analysis and statistics.

The importance of inductive methods has led adherents of geocomputation into an interest in metadata, or data about data. The argument here is that as data sources continue to expand and multiply, then it is not simply the analysis of data which becomes the problem, but the identification of which datasets exist in relation to specific phenomena in the first place. It is axiomatic that much of geocomputation is concerned with the reuse and combination of secondary datasets. The essence of the metadata problem is therefore to provide descriptions of the form and content of spatial data which allow potential users to find them (through conventional search engines – the equivalents of Google or AltaVista), and ideally to query, link, or model their constituent fields. This is clearly a nontrivial task in a world in which spatial representations vary, and entity definitions can be fluid even across small geographical distances – for example, as neighboring cities or local authorities may define land use categories in completely different ways.

One sense in which the 'data deluge' is currently more acute than previously is in the capture and interpretation of real time data. For example, environmental sensors can now detect variations in air quality across multiple points within an urban area on a continuous basis and these can provide a basis for spatial analysis applications focused on causal mechanisms, environmental justice, or monitoring and forecasting. A particularly interesting class of problems arises where there is also a capacity for real time feedback. For example, Kai Nagel and colleagues have recently proposed a real time model of the diurnal journey to and from work across the whole of Switzerland. Queues at individual junctions can be simulated and monitored, in relation to road conditions and control mechanisms such as traffic lights. There is then no reason why the timing of traffic lights (for instance) cannot be adjusted to maximize the smooth flow of traffic throughout the network. Similar applications have been envisaged in relation to problems such as crowd control (at festivals or sports events), retail optimization, or even culture and recreation within cities.

Spatial Orientation

Clearly the focus on spatial problems is emphasized in the notion of geocomputation. For example, Helen Couclelis argues that geocomputation needs to find a coherent perspective on geographical space – to justify the term 'geo'. For most practitioners the question goes beyond an application of computational techniques to geographical problems, and toward the establishment of computational methods which recognize and exploit the unique properties of spatial data. While Couclelis has argued that geographers have failed to get beyond a dearth of theory, it is at least possible to point to a number of practical examples of techniques which combine the geographical and the computational in powerful and exciting ways.

Openshaw's original 'GAM' remains the standout application in this regard. In this work, the author and associates set out to find a method to determine clusters in spatial data, without regard to any *a priori* expectations as to the location or cause of such clusters. The technique works by the repeated application of standard statistical tests across space and at a multitude of geographical scales. For any dataset, it is then possible to answer questions including – where is the greatest concentration of occurrences?, how significant is this concentration?, and what is the spatial extent of each cluster? In a famous application to a dataset of childhood clusters, GAM identified not only a well known preexisting disease cluster around the Sellafield nuclear storage facility in Cumbria, North West England, but also a previously unrecognized and more significant cluster in Gateshead, Tyne and Wear. In later work, attempts were made to produce a GEM in which a target data set is modeled statistically against a number of spatially independent datasets in order to find the most appropriate set of linkages. Unfortunately this work remains relatively underdeveloped following Openshaw's enforced early retirement from academic life. Both GAM and GEM are freely available as Internet applications. An example is shown in **Figure 1**. A dataset of retail customers has been input to GAM. The data consists simply of a list of x and y coordinates of customers on the ordnance survey grid. The output of GAM displays a number of clusters and serves as an indication for further exploration or development of theories in relation to the retail network, its competitors, and customers.

Another successful example of an explicitly spatial approach to the problem of geocomputation is Stewart Fotheringham's geographically weighted regression (GWR). GWR recognizes that statistical techniques, specifically regression analysis, change when they are applied to spatial data. The model coefficients are therefore modeled as continuously varying in space. GWR is provided as a desktop application which is distributed by compact disc, and has users in many different countries. There are published applications in relation to house prices, healthcare, educational attainment, and many others.

A third technique which introduces an explicitly geographical emphasis to a classical problem, in this case with its roots in economics, is spatial microsimulation. Microsimulation is widely used in modeling the impact of financial strategies (e.g., tax and benefit regimes) on a population represented as a list of individuals or households with unique characteristics. Spatial microsimulation has recognized that population characteristics are unevenly distributed across cities and regions, and that this

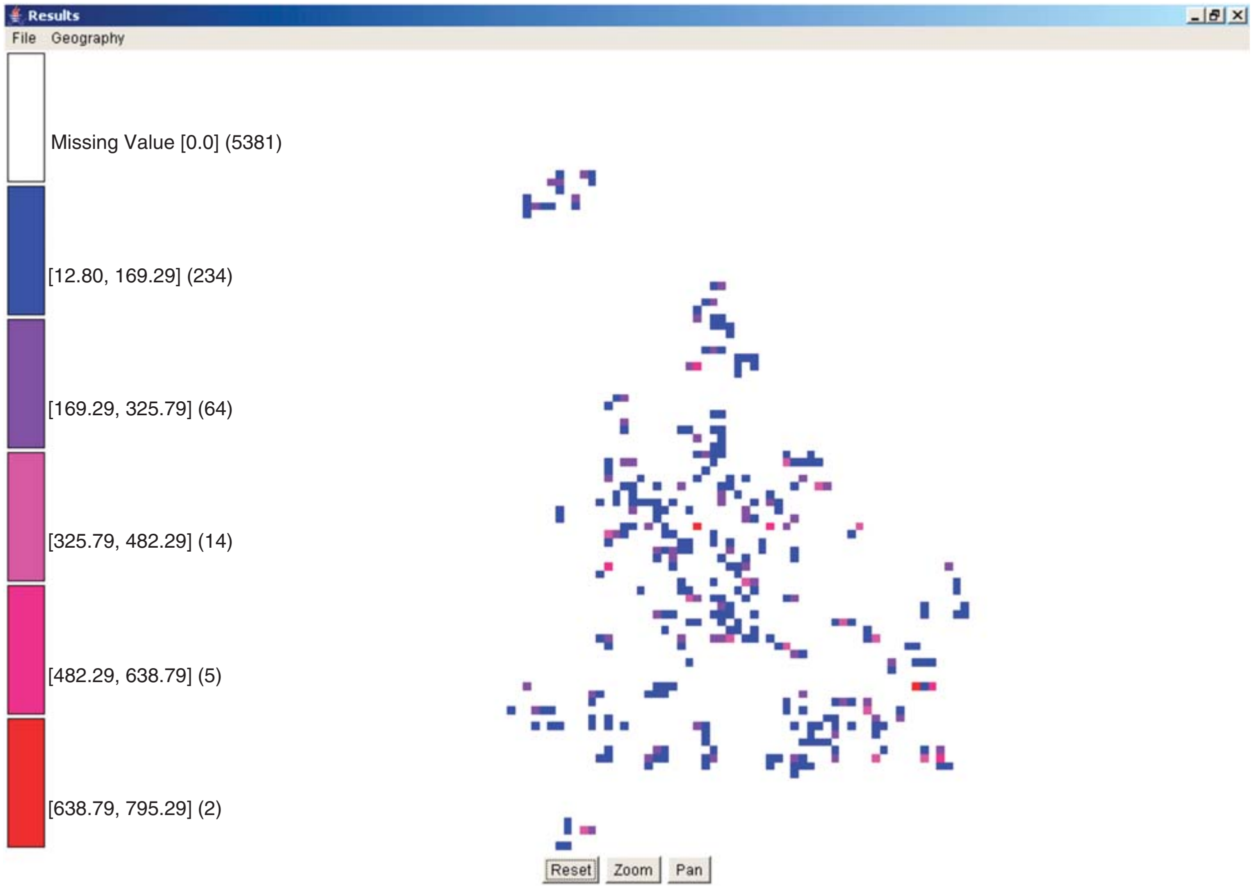

Figure 1 Clustering of spatial data with the geographical analysis machine.

has crucial implications not only for financial policy, but for the delivery of services such as education, retailing, and healthcare.

A further sense in which geocomputation has adopted an explicitly spatial emphasis is in the visualization of the spatial data. Applications range from the extension of classical spatial visualization methods through the transfer of visualization methods from the physical and mathematical sciences (like recurrence plots and relative distribution methods), to the creation of entire software studios for the visual representation and interpretation of spatial and statistical data (e.g., GeoVista at Penn State, and GeoDa at Illinois).

Technical Novelty

We have noted above that there are standout applications in which geocomputation has sought the redevelopment of classical techniques within a geographical environment. However, it is equally true that geocomputation has given an impetus to the introduction of many computational techniques from scientific disciplines into geography. In the late 1990s, cellular automata (CAs) became popular as a means of representing spatial distributions and processes. A CA is an abstract entity which can be represented in a variety of states. The states of neighboring automata are interdependent so that change in one can generate further changes in its neighbors. CAs are particularly suited to land use problems, in which the CA represents a parcel of land and the state is the use to which that land is put (e.g., agricultural use, housing, recreation, and so on). Problems of urban growth are particularly tractable by this method.

If movement is added to CA models, then we begin to enter the even more popular area of agent based simulation. In this context, an agent is an entity whose behavior is governed by certain well defined rules. Advocates of agent based modeling approaches argue that complex geographical phenomena can be replicated using agents with relatively simple behavioral rules. A popular example is Schelling's model of residential segregation, in which the behavior of agents with one characteristic – membership of an ethnic group – are located on a grid, and move around that grid in accordance with a desire to be close to neighbors of a similar ethnic type. This model can be shown to generate distinct ethnic clusters, even when the propensity to locate in proximity to similar ethnic groups is relatively weak. Agent based simulations have also become popular as land use models, but have also been explored in a variety of economic, sociological, and geographical con texts, such as the behavior of agents in a retail market, and the movement of individuals within public spaces.

A wide variety of techniques from artificial intelli gence have also been popularized by members of the geocomputation community. Examples include the use of neural networks to forecast river levels and flood risk, deployment of genetic algorithms for the optimization of retail market areas, derivation of alternative spatial interaction models using genetic programming, network optimization from artificial life, and fuzzy logic classifi cation of residential neighborhoods.

Computational Intensity

The availability of massively powerful computational resources is one of Openshaw's three pillars of geo computation. In early experiments with the GAM, a single national level simulation of disease clusters was estimated to need 9 days of processor time. Bearing in mind that typical applications will typically require multiple runs, of say fifty or a hundred datasets, then some sort of parallel processing capability becomes ne cessary. Parallel processing occurs when computational tasks are shared out across many processing units within a single computer or machine group. Massively parallel computers are simply those with very large number of processors, often several hundred. The good news for geographers is that spatial problems can often be de composed quite easily for parallel execution as a similar task is frequently repeated across multiple locations. A similar application of massive parallelism to a network optimization problem in the 1990s reported increased speeds from 3 months to a few seconds through a com bination of code parallelization and the use of an enor mous 'data vault' for virtual storage. The British Broadcasting Corporation's (BBC) ClimateNet project has recently enlisted 50 000 home personal computers (PCs) on a computational grid for long term environ mental predictions.

Access to massive computational resources has re cently become an intense battleground between discip lines. This is not surprising given the size of the resources now being made available. For example, in the UK alone the e Science program distributed £340 million to re searchers in the years 2001–04, a sum comparable to the total social science research budget. e Science is a gen eric term covering the application of state of the art computational resources to big science problems, some times referred to as 'grand challenges'. A broader case for 'human systems modeling' as a grand challenge for e Science has been sketched – inevitably by Stan Openshaw! Geon and Moses are two significant research projects which explore the deployment of massive computational resources within a geocomputational environment.

Geon is a multidisciplinary collaboration involving geographers, earth scientists, and computer scientists. A network of users contribute data as well as sharing tools for global database search, seismological modeling, and topographical reconstruction from remotely sensed data. Construction of elevation models with billions of data points are facilitated through the National Science Foundation's TeraGrid network.

Moses is a project within the UK National Centre for e Social Science (NCeSS) which develops data rich simulations for the strategic analysis of social policies. Using the technology of the White Rose Grid, a centre of excellence on the National e Science Grid, Moses rep resents a synthetic national population of individuals and households. The changing behavior if this population is modeled under a variety of scenarios to provide decision support capabilities for healthcare, transport, housing, and other planning applications.

While both Moses and Geon share a capability to piggyback on computational grids, the e Science logic for these projects extends beyond a capability for parallel processing. In both cases, access to virtualized data re sources is a second key feature, and a third is the pro vision of a collaborative infrastructure for resource sharing. Data integration and collaborative infra structures are both important features of many e Science projects (in the US, the term 'cyberinfrastructure' tends to be preferred). However the number of applications remains small, because of difficulties in accessing re sources and the skills required to deploy such systems. The more commonplace use of massive geocomputa tional resources depends on the availability of a shared infrastructure of hardware and software. While such shared infrastructures are currently the subject of re search, whether they will ever be made universally available remains an open question.

Policy Focus

Geocomputation seeks to take GIS to a new level, not least in solving real world problems. Policy relevance may be regarded as the ultimate test of real world ap plicability. For example, the Moses project introduced above is explicit in its desire to address policy issues regarding resource allocation, service delivery, efficiency, and equity of access across a variety of domains, in cluding transportation, health, and housing. While Geon is perhaps more research oriented, related GeoScience projects at San Diego Supercomputer Center (SDSC) have a more obvious policy dimension. For example, TeraShake seeks to simulate the impact of the next big earthquake across southern California.

Episims is a particularly good example of policy relevant geocomputational research. Episims is an epi demic simulation model, which is constructed with ref erence to the social network relationships within a population of individual demographic agents. In other words, the model seeks to recognize that transmission of diseases takes place from individual to individual, and that these individual contact probabilities are mediated through social structures such as workplaces, school playgrounds, cinemas, and so forth. From a policy per spective this means, for example, that the spread of a disease might be restricted by school closures. Episims is itself a development of the earlier Transims model, in which agent based simulations of a transport network are constructed. In early research, the developers considered a hypothetical terrorist attack leading to release of the smallpox virus into the city of Portland and were able to demonstrate, for example, that mass vaccination would be one of the most dangerous response strategies (because of the likelihood of increased contact rates among already infective individuals). This research is now being heavily funded by governments in the US and elsewhere in preparing plans to counter potential outbreaks such as H5N1 (Asian bird flu).

Of course the spread of diseases has been a favorite concern of geographers since the quantitative revolution. Nevertheless, geocomputational approaches demonstrate theoretical novelty in their treatment of individual interactions and the mediation of those individuals through social networks, as well as utilizing advanced visualization capabilities for the simulations, and har nessing the power of HPC for realistic models at a re gional, national, or even international scale.

Multidisciplinarity

In early writings, geocomputation is sometimes described as the search for a new paradigm in human geography. However, it is clear that the approach has developed to become inclusive of both human and physical geography. Physical geography applications appear to span most, if not all, of the features of geocomputation which are described above. Algorithms have been deployed in ductively to search for patterns relating to phenomena such as climate change, land cover, and the distribution of ice masses, techniques such as neural networks (to model hydrological discharge), and wavelets (in the simulation of avalanche behavior) have been deployed with en thusiasm, while applications such as Geon show the importance of HPC within the GeoScience domain. Applications such as flood forecasting or recognizing patterns in hydrocarbon formation have clear policy relevance for both commercial and public organizations. Some of the most popular applications of agent based simulation technologies in particular have focused on land use modeling, and can therefore be considered to cross the divide between human and physical divide.

Geocomputation in human geography can also be considered to be multidisicplinary in its substantive emphasis. Many of the examples above, such as Episims, Moses, and disease clustering applications of GAM have a strong health sciences focus. Among agent based simulations, the focus on artificial societies and social networks provides a link to the discipline of sociology, and often a means to explore theories which are more qualitative in nature. There is also much interest in this area from economists seeking to explore concepts like imperfect information and subrational decision making, with clear links to classical behavioral geography.

A third sense in which geocomputation promotes a multidisciplinary approach is in sharing techniques with disciplines such as artificial intelligence, computing, and mathematics. Indeed Helen Couclelis in particular has argued that the essence of geocomputation is to explore the metaphors of computer science within a spatial context. Examples like GWR and GAM serve to em phasize that geographers have not simply sought to re hash new third party techniques close to home, but to add a more distinctive methodological spin. Perhaps a true measure of success in this context would be the extent to which the methods of spatial science can themselves become embedded within other disciplines, particularly those of the hard sciences. A nice example of this would be something like the 'mouse brain atlas', a project in which the various parts of the mouse brain, their functions, and interrelationships were represented using GIS techniques. Nevertheless, it is still striking in the extent to which projects like AstroGrid, which seek to integrate astrophysical data, to map such data, and then automatically detect astronomical phenomena, deploys techniques entirely without reference to GIS, let alone cluster hunting or other methods, for detection of spatial patterns. While geocomputation may have achieved something in lifting the credibility of quantitative geog raphy among the mathematical, computational, physical, and biological sciences, there is clearly some way to go in this regard.

Conclusion

In reflecting on the development of geocomputation within a relatively early volume, Bill Macmillan draws a historical analogy with the development of astronomy in the Renaissance period. He argues that new techniques in the sixteenth and seventeenth centuries gave scientists the ability to make all kinds of new empirical discoveries about the heavens and from this to lift their science to new levels. For Middle Age telescopes read GIS, and for astronomy read geocomputation! The argument is that GISs provide us with the capability to represent and explore spatial environments in hugely powerful ways; what we need is to articulate geocomputational approaches which help us to make more sense of all this.

The metaphor is an interesting one, but perhaps needs to be embellished by the observation that the process of scientific discovery is cyclical rather than linear. The invention of new telescopes is surely stimulated to some extent by a desire to explore the skies, and the need to answer deeper questions arising from initial observations will provide a stimulus to further development of the basic technology. Similarly with geocomputation, it is not simply a case of taking GIS and using these techniques to answer new and hopefully interesting questions. Rather geocomputation should be seen as the evolution of a set of scientific methods (including, but not limited to, GIS) through attempts to apply those methods to hard problems. This view perhaps helps to explain why in practice the subject matter of geocomputation meetings extends right through from relatively theoretical papers about GIS and spatial methods to highly applied studies of a specific problem domain, and why it is right that this should be the case.

The discipline of geocomputation is still young. It might be argued that despite a body of interesting applications, the grand challenges in human geography which geocomputation can bring into view are yet to be fully articulated let alone addressed. Yet the field has made substantial progress in its early years and provided common ground for a new generation of spatial scientists with a passion for a geography which is at once empirical, relevant, and scientifically rigorous. If such enthusiasm can be maintained in the coming decade, then surely its achievements will begin to multiply exponentially.

See also: Artificial Intelligence and Expert Systems; Cellular Automata; Computational Human Geography; Geographically Weighted Regression; Simulation; Spatial Clustering, Detection and Analysis of; Spatial Data Mining, Cluster and Pattern Recognition; Spatial Data Mining, Geovisualization.

Further Reading

Barrett, C. L., Eubank, S. G. and Smith, J. P. (2005). If smallpox strikes Portland … . *Scientific American* 292(3), 54 61.
Couclelis, H. (1998). Geocomputation and space. *Environment and Planning B* 25th Anniversary Issue, 41 47.
Ferguson, N. M., Cummings, D. A. T., Cauchemez, S. *et al*. (2005). Strategies for containing an emerging influenze pandemic in Southeast Asia. *Nature* 437, 209 214.
Foster, I. (2003). The grid: Computing without bounds. *Scientific American* 288(4), 78 85.
Longley, P., Brooks, S., McDonnell, R. and Macmillan, B. (eds.) (1998). *Geocomputation: A Primer*. Chichester: John Wiley.
Martin, D. (2005). Socioeconomic geocomputation and e social science. *Transactions in GIS* 9, 1 3.
Openshaw, S. (1995). Developing automated and smart spatial pattern exploration tools for geographical information systems applications. *The Statistician* 44, 3 16.
Openshaw, S. and Abrahart, B. (eds.) (2000). *Geocomputation*. New York: Taylor & Francis.
Openshaw, S., Charlton, M., Craft, A. W. and Birth, J. M. (1988). Investigation of leukaemia clusters by the use of a geographical analysis machine. *Lancet* I, 272 273.
Openshaw, S. and Openshaw, C. (1997). *Artificial Intelligence in Geography*. Chichester: John Wiley.
Xie, Y. and Brown, D. G. (eds.) (2005). *Proceedings of the 8th International Conference on Geocomputation, Geocomputation CD ROM*. Ann Arbor, MI: University of Michigan.

Relevant Websites

http://www.ccg.leeds.ac.uk
Centre for Computational Geography at the University of Leeds.
http://www.geongrid.org
Geon project website.
http://www.geocomputation.org
Home site for the Geocomputation Conference series, including downloads of papers from recent events.
http://www.ncess.ac.uk
National Centre for e Social Science: Moses project.
http://ncg.nuim.ie
National Centre for Geocomputation, Maynooth.
http://www.mbl.org
Overview of the mouse brain library, including a description of the mouse brain atlas.
http://www.geovista.psu.edu
Pennsylvania State University Geographic Visualization Science, Technology, and Applications Center.
http://www2.astrogrid.org
Portal to the AstroGrid Virtual Observatory.
http://www.nesc.ac.uk
Portal to the UK National e Science Centre.

Geodemographics

M. Birkin and G. P. Clarke, University of Leeds, Leeds, UK

Glossary

Cluster Analysis A statistical approach to grouping data. When applied to small area census data, this becomes a means of identifying areas which are similar to one another.

Ecological Fallacy The idea that applying a descriptive label to an area or group of individuals and households implies homogeneity where there is actually diversity.

Factor Analysis A way of determining the main axes of differentiation within a dataset: for example, in what way can one group of areas be distinguished from another.

Instrumental Rationality The idea that by targeting policies on certain groups of customers or service users, those groups will automatically become the most significant.

Lifestyles Data Information about people including behaviors and interests in addition to classic descriptors such as age and social class.

Profile A structured comparison between a dataset and a parent population, identifying over-representation in certain groups and under-representation elsewhere.

Segment A group of people or a set of geographical areas of a similar type.

Social Area Analysis An approach to the study of urban areas which was popular in the US in the middle part of the twentieth century. It tried to determine the extent to which certain social groups, such as young people, ethnic minorities, or affluent families, would tend to group together within cities.

Introduction

Geodemographics is a classification of the entire population of an area (e.g., the UK) according to the type of area in which they live. Such classifications are usually based on a broad range of indicator variables that cover housing tenure, socioeconomic information, and demographic characteristics (classic census variables).

The geodemographic industry as a whole is a very lucrative one. For example, Claritas Europe – one of at least a dozen major players in the global market – had annual turnover of $100 million in 2003. The vast majority of businesses in retailing, financial services, forecourt retailing and utilities, as well as service providers, local and central governments now make extensive use of the technology, making geodemographics represent the largest end user of census data and applied spatial analysis. Geography has become an extremely marketable subject in recent decades!

This article looks at the history of geodemographic systems, the make up of contemporary systems, and then focuses briefly on the data and methods needed to build and apply such a system. Finally, we consider some drawbacks and criticisms associated with these methods.

Early Developments of Geodemographics

The history of geodemographics is frequently traced back to the Charles Booth poverty maps of Victorian London. However, the intellectual roots of the approach lie most clearly in the work of the Chicago School of urban sociologists, led by Park and Burgess in the 1920s. The human ecology of the urban sociologists developed naturally into the formal classification of areas within cities by means of social area analysis. In statistical terms, this work amounts to a factor analysis of multidimensional data structures built around census variables, and was stimulated by the availability of small area data for census tracts in the order of 4000 people in size from 1910 onward. Commercial applications of the technology date back to at least 1961, and Jonathan Robbin's foundation of General Analytics Corporation (GAC), later renamed as Claritas. The first geodemographic system proper, Potential Rating Index for Zip Markets (PRIZM) did not appear until the mid 1970s, however.

At first the development of similar techniques in the UK was hampered by a lack of the equivalent small area data. From the 1950s onward, however, and perhaps at an accelerated rate in association with the quantitative revolution in geography, a number of classifications of UK cities began to appear. Eventually a full national classification known as A Classification of Residential Neighbourhoods (ACORN) was produced by Richard Webber at Centre for Environmental Studies (CES). Initially this work was targeted at local authorities, with an interest in the distribution of people, housing, and social deprivation within their areas. The application of the ACORN system in the private sector resulted from experimental work conducted in the early 1970s for the British Market Research Bureau (BMRB). Webber's ward based classification was used and cross tabulated with information collected about product purchases and media use, to produce what became known as the Target Group Index (TGI). The study revealed a surprising ability to discriminate more effectively (e.g., between readers of the Guardian and the Telegraph) than 'traditional' indicators, such as social class, age, or education.

In 1979 the consumer analysis company, CACI (originally California Analysis Centers Inc.) recruited Richard Webber, who developed his system for them into a new Enumeration District (ED) level 36 cluster classification of neighborhood types, and it was at this point that it was renamed ACORN and linked into CACI's InSite retail catchment characteristic description system. Webber later left CACI in 1985 to join CCN Marketing, where he developed the MOSAIC geodemographic system.

In recent years, academics have continued to be interested in technical enhancements to the methodology of geodemographics, offering new approaches based on fuzzy clustering, self organization, and genetic algorithms. First Superprofiles, and lately Output Area Classification (OAC), the official Office for National Statistics (ONS) classification of output areas (OAs) in the UK, have been derived as the products of university research. Academics have also had an interest in case study applications of geodemographics, such as crime and policing, education, and healthcare. In the commercial world, however, the major impact has been from the assimilation of new data sources, especially massive lifestyle databases which have allowed systems to be enriched and refreshed on a much more frequent basis. This has made geodemographics progressively more lucrative, with the result that most providers are now embedded within global market research businesses (e.g., Claritas within Nielsen (US) and Acxiom (Europe), CCN as part of Experian, and most recently the acquisition of CACI by ESRI). Geodemographic systems now exist not only across North America and Europe but for many international markets such as Japan and the Philippines (Cameo International), China, New Zealand, and Australia (MOSAIC Global).

Examples of Geodemographics

Major commercial systems in the UK include MOSAIC, ACORN, and Cameo. The major system in the North American market is PRIZM. Typically, these systems adopt a hierarchical classification structure, in which a broad set of area types can be disaggregated into a more refined set of profiles. In many systems, a third level of disaggregation is also possible. For example, ACORN neatly identifies six broad categories from 'thriving' to 'striving' (see **Table 1**), which can then be broken down into a further 17 second level headings ('groups') and 54 third level headings ('types'). Each area type is accompanied by a 'pen profile', which constitutes an intuitive portrait of a geographic neighborhood which summarizes its main characteristics. The pen profiles vary in character from the factually accurate but perhaps bland (e.g., Cameo) to the more evocative, but arguably fanciful (e.g., MOSAIC). The area names themselves are a clue to this dichotomy: Cameo's level 2 areas are labeled simply 1A–10C, whereas MOSAIC lists areas like 'pebbledash subtopia' and 'bohemian melting pot'!

Both MOSAIC and Cameo are also available in international versions. This means that it is not only possible to apply market profiling techniques to global markets, but also to make between country comparisons of cities or regions. Many providers have also experimented with sector specific versions of their classifications, like Grocery MOSAIC and Financial Acorn, in which the basic classifications are fine tuned for a particular purpose, using specialized data about retailing

Table 1 Area classifications within ACORN

ACORN category	*ACORN groups*	*ACORN types*	*% UK households*
B. Expanding	4. Affluent executives, family areas	4.10 Affluent working families with mortgages	1.8
		4.11 Affluent working couples with mortgages, new homes	1.3
		4.12 Transient workforces, living at their place of work	0.3
	5. Well-off workers, family areas	5.13 Home-owning family areas	2.5
		5.14 Home-owning family areas, older children	2.6
		5.15 Families with mortgages, younger children	1.9

ACORN A classification of Residential Neighbourhoods.
- Thriving: 19.8% of UK households (3 groups and 9 types)
- Expanding: 11.6% of UK households (2 groups and 6 types)
- Rising: 7.8% of UK households (3 groups and 10 types)
- Settling: 24% of UK households (2 groups and 7 types)
- Aspiring: 13.7% of UK households (2 groups and 6 types)
- Striving: 22.6% of UK households (5 groups and 16 types)

or financial markets. However, the breadth and general applicability of geodemographics is one of its main ap peals, and the sector versions have tended to be less popular.

The technique of geodemographics has been deployed in many contexts spanning both commercial and public sector domains. A large number of commercial case studies are maintained on corporate web pages. The main application areas include market research, customer tar geting, trade area analysis, and resource allocation. Geo demographics is appealing for market research as a sampling method, in order to determine a set of locations where interviews or other primary research can take place to provide balanced coverage of the population. Customer targeting is a staple commercial application. A typical approach would be for a company to undertake a local trial for an expensive direct marketing campaign, and then to analyze customer responses according to demographic groups. The national campaign would then be targeted at the types of areas in which customer responses were most favorable in the trial. Trade area analysis can be used by retail organizations to profile local markets. For example, this technique was used in appraisals of the Meadowhall Regional Shopping Centre to estimate trading volumes and customer preferences in advance of the development. This technique can also be used by retailers for processes such as sales benchmarking (am I doing well here?) and format optimization (have I got the right products here?). Resource allocation decisions can be supported for com mercial and service organizations so that spending is matched to need. For example, in relation to crime pre vention, neighborhood policing might be a priority in areas of a particular geodemographic type. Funding for healthcare is regularly prioritized to areas of greatest need using geodemographics or alternative indicators of deprivation.

In the rest of this section we offer two examples of geodemographics in action, first, in a commercial en vironment, second, in a public service planning context. Our first example concerns a travel business. One of the beauties of geodemographics is that the only essential requirement is a list of customer addresses. These ad dresses are matched individually to census geographies using a point in polygon operation or lookup table, and this allows geodemographic codes to be assigned to customers based on their place of residence. Now the customer database can be 'profiled', that is, compared to the expected or national average distribution. For ex ample, when the travel agent considers customers who book 'flight only' holidays (no accommodation, car hire, or other products) then 4% of customers are located in Cameo Group 2D (described as urban living professional singles and couples), whereas only 2% of the national population are in this category. These customers are twice as likely to buy 'flight only' than the total market. This relationship would usually be expressed as an index around 100; in this example, the precise value is 213. The top ranking geodemographic groups for flight only, along with cruises and family holidays are shown in **Table 2**.

An observer considering **Table 2** might conclude that all this analysis reveals is what one might know or expect anyway. At one level this is true, but many would argue that a great strength of geodemographics is that it does tend to present an intuitively reasonable representation of the world. For example, individual users (like the market research manager of our travel agent) can carry out tasks such as querying their own postcode to find the kind of neighborhood that they live in, and this profile nearly always makes sense. However, we are also able to quantify quite accurately the levels of variation in cus tomer uptake, and then to take specific action to target new products or promotions. For example, in looking at next year's cruise packages the travel agent might elect to target brochures to all customers of type 5G both existing and potential – sending flyers by direct mail or door drops, newspaper or television advertising in regions with a high concentration of 5G customers, stocking extra brochures, or prominent display of posters at branches with a high 5G customer base. In this way, the travel agent would expect to become more effective in targeting

Table 2 Penetration indices by Cameo segment for a travel business

Type	*Description*	*Index*
2.1	*Flight only*	
2D	Urban living professional singles and couples	213
4D	Mature owner occupiers in rural and coastal neighborhoods	207
1A	Opulent couples and singles in executive city and suburban areas	205
4C	White-collar home owners in outer suburbs and coastal areas	190
2.2	*Cruise*	
5G	Older couples and singles in suburban family semis	243
8C	Older households with school-age children in towns and suburbs	239
2B	Affluent mature families and couples in large exclusive detached homes	231
2A	Opulent older and retired households in spacious rural properties	230
2.3	*Family holidays*	
4E	Couples and families in modern rural and suburban developments	210
8C	Older households with school-age children in towns and suburbs	189
3D	Wealthy mixed households living in rural communities	175
2B	Affluent mature families and couples in large exclusive detached homes	172

and recruiting new customers (as well as retaining the ones they have already) and thus generating real financial benefits from the application of geodemographics.

Our second brief example is taken from work undertaken with the UK Home Office. This arose from the need to benchmark police force performance and the possibility of using such analysis to help better allocate existing resources. This type of work stems from the recent government emphasis on performance indicators in public services and the increasing interest in geodemographics for public sector applications. Often schools, hospitals, police authorities, etc., are listed in a league table based on such key performance indicators. Working with the Home Office, researchers at Leeds built a model of 'crime expectation' based on a geodemographic classification which uses a raft of socioeconomic variables. This creates a number of clusters of police force areas which are similar in socioeconomic terms. Then, police authorities working in similar types ('families') of community environment could be directly compared. **Figure 1** shows variations in burglary rates within family 4, large city areas with many socioeconomic problems. These new rankings and families proved useful policy alternatives to the traditional crime lists which always had leafy England at the top (lowest crime rates) and the forces working in the most deprived estates at the bottom. Using the 'distance from expected value', Knowsley, for example, emerged as an area where, despite some of the highest poverty rates in the UK, crime rates were lower than expected (they were still high of course). Subsequent media attention focused on the good things the police force was doing to reduce drug related crime in particular. The new method of classification proved popular with police forces. For example the commander of Knowsley Police, Supt. Stuart Kernohan, was quoted as saying:

> This is a thought provoking analysis of crime statistics and highlights the importance of the force's move to neighbourhood policing across Merseyside.

Data

The most important source of data for all geodemographic systems is the Census of Population, which is conducted at 10 year intervals in the UK, US, and many other countries. The census provides data which is wide ranging in its social and demographic contents, it covers the whole population, and is available for very small geographical areas.

The major kinds of data which are captured within the census and are useful within a geodemographic classification include demographics (age, sex, marital status, ethnicity), housing (tenure, number of rooms, central heating), socioeconomic status (economic activity, occupation), household composition (number of children, dependents, lifestage), and health (limiting long term illness, overall health and fitness).

For commercial purposes in particular, however, a number of important things are missing from census data. The topic of income has been debated for a long time in the UK, and indeed the trial census from 2001 included such a question, but ultimately personal and household income was not collected as part of the 2001 Census. Furthermore, it has often been argued that spending depends more on behaviors and attitudes rather than demographics and class. Thus hobbies and interests could be just as important as age and job type. In order to capture these dimensions, 'lifestyles' data are particularly important. Lifestyles data are collected by a variety of means, including newspaper questionnaires, product

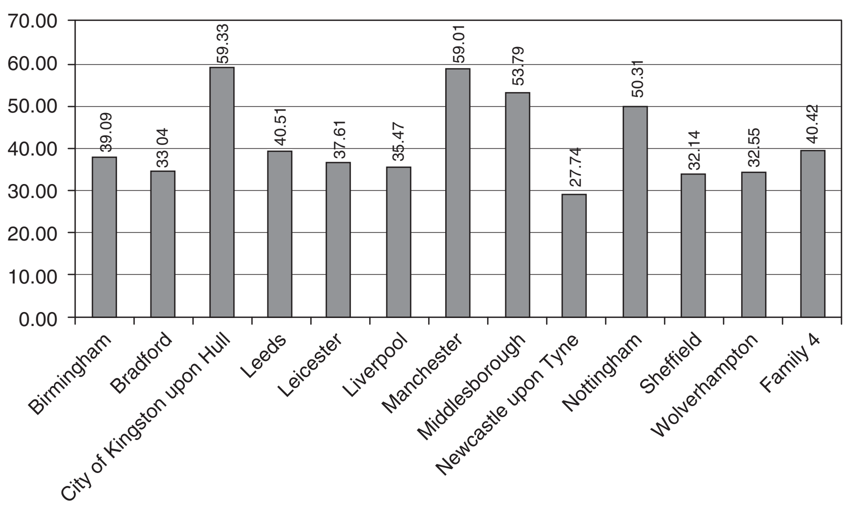

Figure 1 Family 4 from the Home Office classification: a comparison of domestic burglaries rates, 2000 (per 1000 persons).

guarantee forms, and door to door leaflets. The questionnaires typically ask people about their hobbies and interests, household income, and product ownership (e.g., type of vehicle owned or driven) in addition to basic demographic and household data. These data are therefore a good match for the most important missing elements from census data. Although not as extensive as the census, companies like CMT and NDL have collected tens of millions of responses to lifestyles questionnaires. It could also be argued that lifestyles data are superior to census data to the extent that the information tends to be continually refreshed and so is more up to date, and it is usually available for individual respondents, and so classifications may be produced for individual households, or at very fine spatial scales (such as unit postcodes or zip plus four in the US).

In addition to lifestyles data, many geodemographics companies will also incorporate third party data such as property prices, shareholder data, or County Court Judgements into their profiling process. For example, EuroDirect produces seven different versions of Cameo using a variety of data sources of this type, so that while Cameo UK is purely census based, Cameo Property is based on a profile of house prices by area. The idea is that the different classifications can be combined flexibly for different purposes. On the other hand, Experian uses the names and addresses of company directors as a direct input to the classification process for MOSAIC, so here the emphasis is on using a wide variety of data to get the best possible core product.

Finally, pure spatial data are used in some classifications. For example, Experian has developed measures of accessibility to retail and other services, which are used in addition to population density measures in area profiling. There is no reason why measures of green space, or even variables like pollution and air quality, might not become increasingly used as such information becomes universally available.

In summary, then, census outputs are augmented within geodemographic systems by three other types of data – lifestyles, spatial data, and third party registers. In addition to 'thickening' the source data, potential advantages of these sources are that they may be bang up to date and allow geodemographic profiles to be disaggregated to very fine spatial scales. However, none of these secondary data sources can match the census for overall quality and density of coverage, and so classifications which respect census timescales and geographies will always inspire the most confidence.

Methods

In order to build a geodemographic classification, it is clear that some kind of method for data reduction is required. In other words, we need to take a whole raft of data elements (described above) and to collapse them into a single code which represents a type of neighborhood. The most suitable technique for this problem is cluster analysis – in essence, we take a set of areas and divide them into groups of areas which are of a similar type. In a two or three dimensional space, these might be visualized as clusters within a traditional sense – see **Figure 2**, for example. In a higher dimensional space, the concept is similar but the visualization is a bit more difficult!

There are three basic flavors of cluster analysis, known as agglomerative, hierarchical, and relocation. To fix ideas, we will assume that the problem is to divide 200 000 census OAs into 10 clusters, although the specific numbers have no impact on the logic. Using an agglomerative method, we would start with 200 000 OAs and define a measure of similarity between areas based on the data available. For example, if area A has 10% unemployment, area B has 9% unemployment, and area C has 3% unemployment, then area A is more similar to area B than area C on this criterion. With appropriate assumptions, these similarities can be summed or averaged across all the data inputs. Next we join together the two most similar areas. In effect, this gives us 199 999 clusters. The characteristics of the new cluster are compounded from its constituent areas. The process is simply repeated until we arrive at the desired set of 10 clusters.

Using hierarchical methods, we would start with 200 000 OAs allocated to a single cluster. Then we would split the cluster into two pieces, in such a way that the two pieces are as different as possible. Again, in two dimensions we can visualize this process as simply drawing a boundary between two groups of areas on a graph (**Figure 2**). The process is exactly analogous in higher dimensions. Next, we try and cut each of our two superclusters into two pieces, and accept the solution which gives maximum differentiation. Then we keep going until ten clusters are reached.

A relocation method would take ten OAs at random to form cluster centers, and then allocate each OA to its nearest cluster. Once this has been done, the initial area is

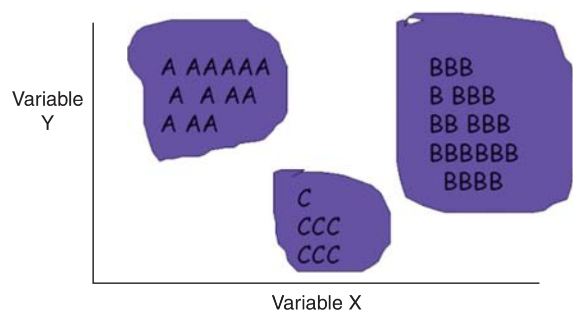

Figure 2 Three clusters in two dimensions (schematic).

no longer at the geometric center of each cluster, so the cluster centroid must be recalculated. Then the OAs are reallocated, and we proceed until the cluster centroids stop moving.

In practice, both agglomerative and hierarchical methods are hugely expensive computationally and cannot be implemented for problems of a reasonable size (although there is a hierarchical process known as Ward's algorithm which is available in Statistical Package for the Social Sciences (SPSS) and can be operationalized for clusters composed from several hundred cases). For example, with 200 000 OAs, generation of the first pairing in the agglomerative algorithm requires evaluation of similarity, maybe across several hundred variables. Then this process needs to be repeated 199 999 times. This is too much work even for a supercomputer! In practice, therefore, relocation based methods are used. The most popular of these is k means clustering within SPSS. Here 'k' is simply a reference to the number of clusters, and 'means' is a reference to the arithmetic mean, because the cluster centroids are calculated using the arithmetic mean of each variable in each cluster.

K means is not a perfect technique. In particular, solutions are not unique and will vary according to the starting conditions (i.e., which random cluster centroids are initially chosen). Furthermore, cluster analysis itself is limited by the need to select a number of clusters at the outset, and it is difficult to propose satisfactory and objective measures to compare a solution with (say) 8 clusters to a solution with 10 clusters. Other practical problems include which data to include, and the possibility of weighting certain variables to increase their importance (e.g., suppose we have ten demographic variables and two health variables: should we increase the weighting of the health variables because there aren't very many?). Nevertheless k means cluster analysis has been shown to provide robust classification systems with distinctive area types at a number of levels of resolution (number of clusters). Refined versions of the k means technique have been proposed, including the use of genetic algorithms, or even self organizing maps. Arguably these methods may lead to slightly 'better' classifications, but logically they are relocation based and very similar to k means clustering.

A slightly more fundamental refinement of k means clustering is the idea of fuzzy clustering. This idea starts from the observation that the distinctions between clusters are not necessarily clear cut. Thus OAs which are near to cluster centroids may be relative distinct, but those which are close to boundaries might easily be reclassified, for example, if slightly different data were chosen, or due to sampling errors in the data itself. Fuzzy mathematics is essentially a discipline which recognizes that boundaries are inexact, and fuzzy clustering therefore proceeds by recognizing explicitly that cluster membership is probabilistic rather than deterministic. However, while it is conceptually interesting and methodologically valid, fuzzy clustering is problematic simply because it complicates a technique, that is, geodemographics, of which the most appealing feature is its simplicity.

Once a set of small areas has been grouped into clusters, the procedures associated with the preparation of a geodemographic system are fairly straightforward. The key step is to inspect and interpret the distribution of data means across the clusters. Usually these means are presented in the form of indices, so that 100 represents the average across the whole population. Suppose, for example, a cluster is indexed at 120 for infant population (aged 0–4). Thus, areas in this cluster have 20% more infants than the population as a whole. Following this logic, this cluster also has 20% less unemployment than the population as a whole, and so on. By repeating this process across the full set of variables in the classification, we then arrive at a view of what typical areas in this cluster look like, and the narrative form of this analysis forms the basis for the pen profiles discussed earlier. All that remains as part of the descriptive process is to produce pithy names for the clusters!

Some form of 'ground truthing' is important in the validation of the clustering process. Ground truth is a concept with its origins in remote sensing which involves comparing the results of a modeling or estimation process with an actual observation on the ground. So what we would do is to select some areas from the classification and actually see what they look like. In particular, do neighborhoods of the same type share similar characteristics in real life, and are they different to neighborhoods of other types? Ultimately, many systems would then try to represent typical neighborhoods using photographs and embed these into product descriptions or pen profiles.

Once all this has been done, then it is to be hoped that the geodemographic system can be used to profile populations in a distinctive and helpful way. A typical process here is simply to compare the distribution of a 'target market' against the national population. Let us suppose that the target in question is students entering higher education. Postcodes for all students are taken from the Higher Education Statistics Agency, and profiled to show that 26% of all students come from affluent professional backgrounds. In the national profile, 13% of school leavers have similar backgrounds. Entry to higher education is represented at 200 within this database. The effectiveness of the profiling can be assessed informally by comparing these penetration indices across groups. An effective application will have a number of values both much greater than and well below 100. This effectiveness can be formally measured using a measure such as the gini coefficient, more commonly known simply as 'lift' within the market analysis industry.

Figure 3 shows an example of the recent UK ONS ‘Output Area Classification’ for the city of Leeds in the UK.

Issues

In a searing critique of geodemographics, Jon Goss has objected to the ‘instrumental rationality’ of this methodology. In our interpretation, this means that geodemographics becomes a self fulfilling prophecy: if products are exclusively targeted at customers (in neighborhoods) type x, then customers of type x will start to demonstrate a strong propensity to buy these products, and so on. Therefore, one objection to geodemographics is that it creates or exacerbates existing divisions within markets, and potentially within society more broadly. According to Goss such marketing sciences exercise ‘surveillance and control’ over the field of consumption and take no account of those who are ‘excluded from the pleasure even of playing with consumer identity’.

Goss has a number of other ethical objections to geodemographics, revolving around notions of ‘dataveillance’ and ‘consumer espionage’ – the lengths to which organizations will go in piecing together customer profiles from disparate sources and the extent to which individual privacy is potentially infringed. On the other hand, one could also argue that there are extensive legal frameworks governing the use of personal data, and furthermore that one of the attractions of geodemographics is that it presents profiles of areas or groups and not of individuals.

Alongside these ethical objections there also sit important methodological reservations about geodemographics. The most significant of these are the ecological fallacy (the implication that all residents of a

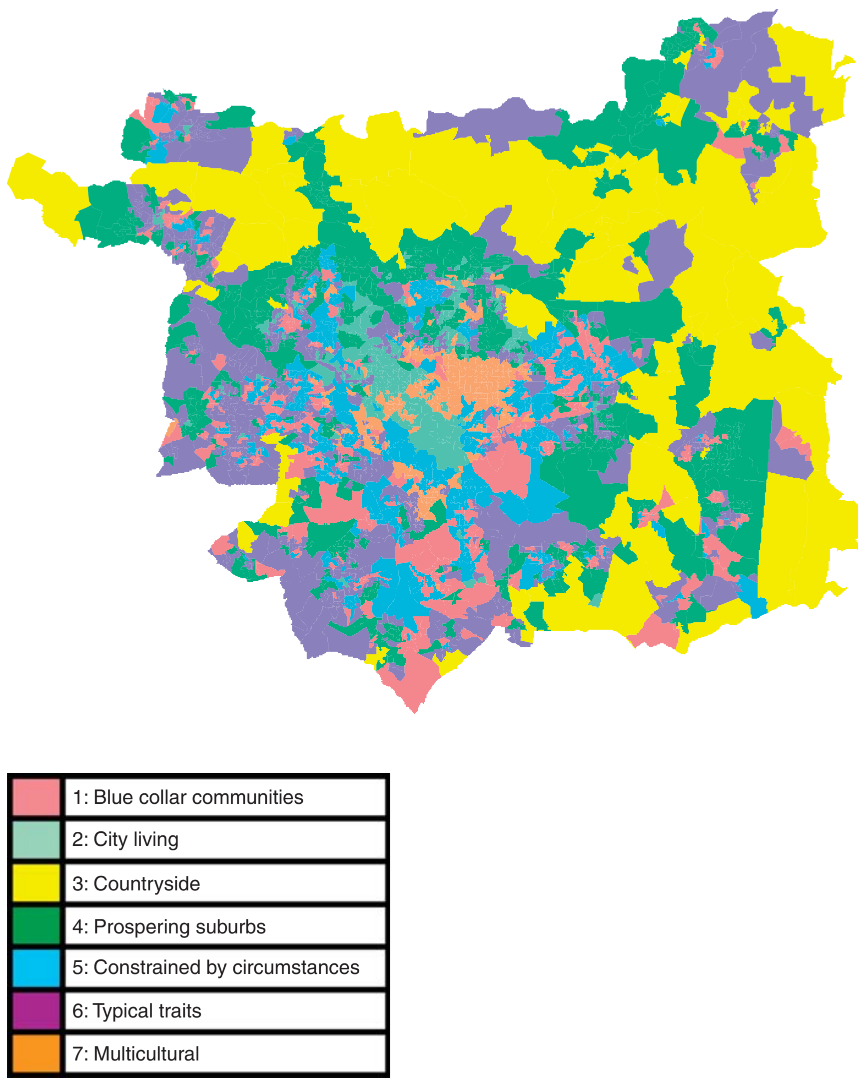

Figure 3 ONS Output Area Classification for Leeds, UK.

neighborhood share the same characteristics), up to dateness, and data accuracy.

At the end of the day, one potential rebuttal to the methodological problems at least is that geodemographics appears to 'work' in a robust and meaningful way. Even this benefit has been turned around as a weakness by some, arguing that however meaningful empirically, geodemographics lacks a robust theoretical foundation. In constructing models of educational attainment, Richard Harris and Ron Johnston have argued that at best area profiles should be used as only one potential explanatory variable, and perhaps better left in the role of exploratory analysis.

However, it is undoubtedly true that geodemographics is one of the most important applied techniques in the world outside academia. For the current authors, engagement with such techniques is therefore important in our teaching, and what is important in our teaching is important in our research. Important research questions about geodemographics should continue to be asked, for example, in relation to scale effects (how do patterns vary at different spatial scales, what does this mean about neighborhood geographies, and what are the practical implications?), in relation to dynamics (how do neighborhoods change over time, are there general models, and how do these differ from place to place?), and in relation to spatial inequalities and environmental justice (do resource allocation mechanisms, e.g., in schooling or healthcare, reflect variations in neighborhood types, and is it right that they should do so?). The study of geodemographics will continue to be of importance to academics, students, and policymakers.

See also: Census Geography; Ecological Fallacy; Education; Factor Analysis and Principal-Components Analysis; Health Inequalities; Urban Morphology.

Further Reading

Ashby, D. and Longley, P. (2005). Geocomputation, geodemographics and resource allocation for local policing. *Transactions in GIS* 9(1), 53 72.

Feng, Z. and Flowerdew, R. (1998). Fuzzy geodemographics: A contribution from fuzzy clustering methods. In Carver, S. (ed.) *Innovations in GIS 5*, pp 119 127. Taylor & Francis: London.

Goss, J. (1995). We know who you are and we know where you live: The instrumental rationality of geodemographics systems. *Economic Geography* 71, 171 198.

Harper, G., Williamson, I., Clarke, G. P. and See, L. (2002). *Family Origins: Developing Groups of Crime and Disorder Reduction Partnerships and Police Basic Command Units for Comparative Purposes*. London: Home Office.

Harris, R., Sleight, P. and Webber, R. (2005). *Geodemographics, GIS and Neighbourhood Targeting*. Chichester: Wiley.

Longley, P. (2005). Geographical information systems: A renaissance of geodemographics for public service delivery. *Progress in Human Geography* 29(1), 57 63.

Openshaw, S., Blake, M. and Wymer, C. (1995). Using neurocomputing methods to classify Britain's residential areas. In Fisher, P. (ed.) *Innovations in GIS 2*, pp 97 111. Taylor & Francis: London.

Sleight, P. (1997). *Targeting Customers. How to Use Geodemographics and Lifestyle Data in Your Business* (2nd edn.). Henley on Thames: NTC Publications Limited.

Tonks, D. and Clarkson, S. (1997). A geodemographic analysis of degree performance. *Higher Education Review* 30(1), 5 19.

Tranmer, M. and Steel, D. G. (1998). Using census data to investigate the causes of the ecological fallacy. *Environment and Planning A* 30, 817 831.

Vickers, D. W. and Rees, P. H. (2006). Introducing the national classification of census output areas. *Population Trends* 125, 15 29.

Voas, D. and Williamson, P. (2001). The diversity of diversity: A critique of geodemographic classification. *Area* 33(1), 63 76.

Relevant Websites

http://www.statistics.gov.uk
Access Point for the Office for National Statistics Output Area Classification, UK Statistics Authority.

http://booth.lse.ac.uk
Charles Booth Online Archive, Charles Booth Poverty Maps.

http://www.eurodirect.co.uk
Description of the Cameo Family of International Classifications, EuroDirect.

http://www.casa.ucl.ac.uk
Description of the 'Geodemographics and Geo Genealogy' Project at University College London.

http://www.claritas.com
Home Page of Claritas Describing Geodemographic Products and Applications.

http://www.casa.ucl.ac.uk/
Home page for David Ashby, Describing Research into Geodemographics and Policing, UCL Centre for Advanced Spatial Analysis.

http://www.experian.co.uk
Home Page of the MOSAIC System and Experian Business Solutions, Experian.

http://www.caci.co.uk
Overview of ACORN.

http://geodemographics.org.uk
The 'Geodemographics Knowledge Base', a Portal to the Geodemographics Industry Maintained by the Market Research Society.

Geodesy

J. Hessler, Library of Congress, Washington, DC, USA

Glossary

ECEF Coordinates One of the forms of geocentric coordinates where the axes of the coordinate system are attached to the rotating solid Earth.

Ellipsoid Purely geometrical idealized reference shape of the Earth that has two or three axes of differing size.

Equipotential Surfaces Surfaces of equal gravitational potential of which the geoid is a special case.

Geodetic Datum Any numerical or geometrical quantity or set of such quantities that serves as a reference or base for other quantities and distances on the Earth's surface.

Geoid Figure of the Earth abstracted from its topographic features and specified by a model of the equipotential gravitational surface. This to a first approximation is defined by the idealized mean sea level.

Global Positioning System (GPS) A constellation of at least 24 medium Earth orbit satellites that transmit precise microwave signals enabling a GPS receiver to determine its location, direction, and time.

Inertial Coordinates One of the forms of geocentric coordinates in which the axes of the coordinate system are associated with the fixed stars.

Point Positioning The accurate determination of the coordinates of a point on land, at sea, or in space with respect to a coordinate system.

Geoid, Ellipsoid, and Gravity Field

The geoid is essentially the figure of the Earth abstracted from its topographic features. It is an idealized closed equilibrium surface that is formed by the surface of sea water. This is the mean sea level surface in the absence of currents, air pressure variations, and other perturbing phenomena. The geoid, unlike the Earth's reference ellipsoid, is irregular and too complicated to serve as the computational surface on which to solve geometrical problems like point positioning. The geometrical separation between the geoid and the reference ellipsoid is called the geoidal undulation and is defined positively above and negatively below the reference ellipsoid. To account for this difference the notion of geodetic height is used and is calculated in terms of the Earth's gravitational potential. These heights are specified according to an equipotential surface whose calculation determines the shape of the geoid. In geodesy, a point on the surface of the Earth having a height that is 'above sea level' forms an irregular, physically defined surface. Height is generally not thought of as a coordinate but rather a physical quantity. There are three types of heights normally associated with geodetic computation: normal heights, geopotential heights, and orthometric heights. Both orthometric and normal heights are defined in meters above sea level, whereas geopotential heights are measures of potential energy (**Figure 1**) and therefore not metric or distance related. Geodetic heights that define the geoid are referenced to the deformation potential of the Earth's gravity and are not constants but vary with time. The variation is caused by forces acting on the earth and things such as the time varying loading of the atmosphere. The global geoid shape and the gravity potential of the Earth are now well defined and accurately expressed using the coefficients of spherical harmonic expansions and other types of approximations that utilize the data from Earth orbiting satellites.

To calculate the shape of the geoid, the gravity field of the Earth must be calculated along with both its internal and external gravity potential. A particle on the surface of the Earth is acted on by two forces, the gravitational force represented by

$$\rho_g \mathrm{grad}\, u(x, y, z)$$

and the inertial force represented by

$$\rho_i \mathrm{grad}\, \omega^2 \left(x^2 + y^2\right)$$

in a geocentric coordinate system which rotates with angular speed ω of the Earth. The actual computation of the geoid is based on the solution of the field equation of gravitation which describes gravity by

$$\mathrm{div}(\mathrm{grad}\, u(x, y, z)) = 4\pi g \rho(x, y, z)$$

The source of gravitation in the above equation is the mass density field $\rho(x, y, z)$ which is multiplied by the coupling constant g of the inertial and gravitational forces. There are many solutions to this field equation that effectively define the gravitational potential.

The reference ellipsoid is usually chosen to be the same volume as the geoid and is described by its semimajor axis defined by the equatorial radius and the flattening that makes it appear oblate. The quantity $f = (a - b)/a$, where b is the semiminor axis (polar

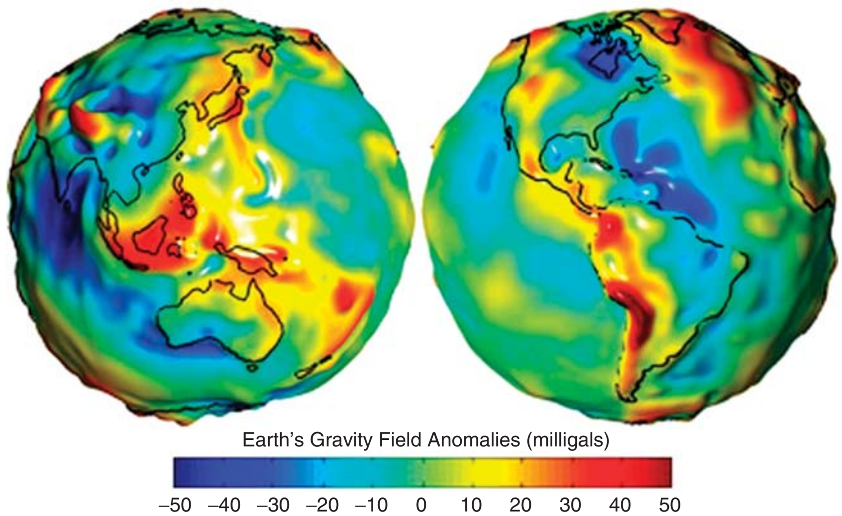

Figure 1 Gravity field anomalies from GRACE. Courtesy NASA.

radius), is a purely geometrical relationship. The mech anical ellipticity of the Earth which is the dynamical flattening caused by its rotation has been determined to a high degree of precision by observation of satellite orbit perturbations. Its relationship with the geometric flat tening is indirect and depends on the internal density distribution of mass in the Earth. The 1980 Geodetic Reference System (GRS80) posited a 6 378 137 m semi major axis and a 1:298.257 flattening. This system was adopted at the XVII General Assembly of the Inter national Union of Geodesy and Geophysics (IUGG) and formed the basis for geodetic positioning by the Global Positioning System.

Coordinate Systems

Finding the accurate locations of points on the surface of the Earth is an important aspect of geodesy and these points are most conveniently described by three co ordinates, X, Y, and Z, in space. Since the advent of satellite positioning, such coordinate systems are typi cally geocentric where the Z axis of the coordinate sys tem is aligned with the Earth's axis of rotation. In geodesy two types of data are necessary, a horizontal and a vertical. The horizontal datum forms the basis for the computations of surveys that consider the curvature of the Earth, while the vertical datum is the baseline that elevation determinations refer to. Before satellite geo desy, the coordinate systems associated with geodetic datum attempted to be geocentric, but their origins differ from the geocenter by hundreds of meters, due to re gional deviations in the direction of the plumb line. These regional geodetic data, such as ED50 (European Datum 1950) or NAD83 (North American Datum 1983), have ellipsoids associated with them that are regional 'best fits' to the geoids within their areas of validity, minimizing the deflections of the vertical over these areas. It is only because Global Positioning System (GPS) satellites orbit about the geocenter that this point be comes naturally the origin of a coordinate system defined by satellite geodetic means, as the satellite positions in space are themselves computed in such a system. Satellite data are defined in a slightly different way, based on the positions of tracking stations, the geopotential model used to define the Earth's gravitational field and physical parameters that do not appear in traditional data. These include things such as the velocity of light, the rotation of the Earth, and clock corrections. Geocentric coordinate systems used in geodesy can be divided naturally into two classes. First, there are inertial reference systems, where the coordinate axes are oriented relative to the fixed stars. Second there are co rotating, also called ECEF ('Earth Centered, Earth Fixed'), coordinates, where the axes are attached to the solid body of the Earth.

Satellite Geodesy and Point Positioning

Point positioning is an important application of geodetic data and is the accurate determination of the coordinates of a point on land, at sea, or in space with respect to a coordinate system. Point positions are found by compu tation from measurements linking known positions with unknown terrestrial positions. This may involve trans formations between or among astronomical and terrestrial coordinate systems. The known points used for point positioning can be triangulation points of a higher order network, GPS satellites, or some other form of satellite geodesy. The use of satellites in geodesy has a long his tory and is now standard for most geodetic measurements.

Historically, there have been a wide variety of techniques employed in satellite geodesy and several of the most important are discussed briefly in the following.

Geodetic Use of Global Positioning Satellite Systems

A GPS receiver calculates its position by measuring the distance between itself and three or more GPS satellites. Measuring the time delay between transmission and re ception of each GPS microwave signal gives the distance to each satellite, since the signal travels at a known speed. These signals also carry information about the satellites' location and general system health (known as almanac and ephemeris). By determining the position of, and distance to, at least three orbiting satellites, a receiver can compute his position using trilateration. Receivers typi cally do not have perfectly accurate clocks and because the positioning in the GPS is based on very precise time measurements between satellites, a receiver normally tracks one or more additional satellites to correct for possible clock error.

Laser Ranging to Satellites

In satellite laser ranging (SLR), a global network of ob servation stations measures the round trip time of flight of ultra short pulses of light to satellites equipped with retro reflectors. This provides instantaneous range measurements of millimeter level precision which can be accumulated to provide accurate measurement of orbits and a host of important scientific data. It is the most accurate technique currently available to determine the geocentric position of an Earth satellite, allowing for the precise calibration of radar altimeters and separation of long term instrumentation drift from secular changes in ocean topography. Its ability to measure the variations over time in the Earth's gravity field and to monitor motion of the station network with respect to the geo center, together with the capability to monitor vertical motion in an absolute system, makes it unique for modeling and evaluating long term climate change by providing a reference system for post glacial rebound, sea level, and ice volume change.

Interferometric Synthetic Aperture Radar

Interferometric synthetic aperture radar, also abbreviated as InSAR or IfSAR, is a radar technique used in geodesy and remote sensing. This geodetic method uses two or more synthetic aperture radar (SAR) images to generate maps of surface deformation or digital elevation, using differences in the phase of the waves returning to the satellite. The technique can potentially measure centi meter scale changes in deformation over time spans of days to years. It has applications for geophysical monitoring of natural hazards, for example, earthquakes, volcanoes, and landslides.

Satellite Laser Altimetry

Ice, Cloud, and Land Elevation Satellite (ICESat), part of NASA's Earth Observing System, is a satellite mission for measuring ice sheet mass balance, cloud and aerosol heights, as well as land topography and vegetation characteristics. The satellite uses a laser system to measure distance, a Global Positioning System (GPS) receiver, and a star tracker attitude determination sys tem. The laser transmits short pulses of infrared light and visible green light photons reflected back to the space craft from the surface of the Earth and from the atmos phere, including the inside of clouds, that are being collected in order to precisely determine distances.

Satellite-to-Satellite Tracking

The goal of the Gravity Recovery and Climate Experi ment (GRACE) space mission is to obtain accurate global and high resolution determination of both the static and the time variable components of the Earth's gravity field (**Figure 1**). GRACE enables precise measurement of the Earth's shifting water masses by detecting their effects on the gravity field.

To determine accurate benchmarks or reference positions, geodetic data from any source is used to build a hierarchy of networks to allow point positioning within a country. Highest in the hierarchy are traditional trian gulation networks used in surveying. These are made into networks of polygons, into which local mapping surveying measurements, usually with measuring tape, corner prism, lasers, and the familiar red and white poles, are tied. Most measurements today are performed with GPS. The higher order networks are measured with static GPS using differential measurements to determine vectors between terrestrial points. These vectors are then adjusted in the traditional network fashion to determine local distance and height variations. A global polyhedron of permanently operating GPS stations under the aus pices of the IERS is used to define a single global, geo centric reference frame which serves as the 'zero order' global reference to which national measurements are now attached. One purpose of point positioning is the pro vision of known points for mapping measurements, also known as horizontal and vertical control. In every country, thousands of such known points exist in the terrain and are documented by the national mapping agencies.

See also: Global Positioning/GPS; Photogrammetry/Aerial Photography.

Further Reading

Grafarend, E. W. (1994). In Varicek, P. (ed.) *What Is a Geoid? Geoid and Its Geophysical Interpretations*, pp 3 32. Boca Raton, FL: CRC Press.
Grafarend, E. W., Krumm, F. W. and Schwartz, V. (eds.) (2003). *Geodesy: The Challenge of the Third Millenium*. Berlin: Springer Verlag.
Hoffmann Wellendorf, B. (2006). *Physical Geodesy*. New York: Springer Verlag.
Rapp, R. H. and Sanso, R. (eds.) (1991). *Determination of the Geoid, Present and Future*. New York: Springer Verlag.
Smith, J. (1996). *Introduction to Geodesy*. NewYork: Wiley.
Torge, W. (2001). *Geodesy*. Berlin: W. de Gruyter.
Vanicek, P. and Christou, N. (eds.) (1994). *Geoid and Its Geophysical Interpretations*. Boca Raton, FL: CRC Press.

Relevant Websites

www.agu.org
Geodesy Section of the American Geophysical Union.
www.iugg.org
International Union of Geodesy and Geophysics.

Geographical Journals

L. D. Berg, University of British Columbia, Kelowna, BC, Canada

Glossary

Academic Capitalism It refers the conjuncture of processes under neoliberalism that is forcing both academic institutions and academic faculty members to behave in more capitalist-like ways. Neoliberalism has seen the reduction in state block-funding of universities, and a concomitant emphasis being placed on both targeted state funding and external nonstate funding. Funding is being streamed to higher-performing institutions. At the same time, such 'performance' is being measured in more finance-like terms, drawing on financial audit systems as the metaphor for measuring educational and research performance. The change in funding regimes, along with the rise of audit cultures in the academy, have both acted to pressure academics (and the institutions they inhabit) to behave in more entrepreneurial ways than in the past. The embedding of such processes into everyday academic life is a key aspect of the rise of academic capitalism.

Citational Consciousness Arising within both neoliberal understandings of the academic Self as a unified individual and academic capitalist valuations of individual labor, citational consciousness involves the overemphasis of the value and importance of academic citations (i.e., when one academic cites the publication(s) of another academic in his/her own published work). Citational consciousness also involves the conflation of citations with quality, such that more highly cited works are seen to be of higher quality than less-cited works. Similarly, more highly cited works are seen to be of greater importance than less-cited works. All of this draws on the science model of knowledge production, which is closely connected to enlightenment reasoning and the notion of single direction for scientific 'progress'. At the same time, it ignores the geographic contexts of knowledge production, the problematic nature of English-language hegemony in the sciences and social sciences, as well as less-obvious issues such as the differing intercitational rates of arts, humanities, and social sciences. Citational consciousness has become more embedded in academic life through processes like the UK Research Assessment Exercises and the New Zealand Performance-Based Research Funding assessments. It is likely to be even more entrenched in the lives of UK academics with moves to the new 'metrics-'based (read citations) Research Excellence Framework. Notwithstanding these official uses of citation counts, they have been an important unofficial art of many academics at North American research universities for sometime now.

Introduction

While geography has a long history as an intellectual pursuit, its formal institutionalization as an academic discipline with its own separate identity in Western universities is a phenomenon of the late nineteenth and early twentieth centuries. Even then, according to Johnston, "full institutionalization [of Geography] did not occur until well into the twentieth century." A wide range of interrelated processes were part of the extended institutionalization of Geography: the establishment of Geography departments in universities; the introduction of Geography degree programs; the allocation of professorial (lecturing) posts in Geography to teach and undertake research in the discipline; the establishment of learned societies to promote Geography; and, the establishment of publication outlets to publish research in Geography (note the use of upper and lower case letters to signify the difference between Geography as a discipline and geography as practice).

While all of these phenomena were important to the development of Geography, of interest here is the establishment of scholarly publication outlets in the discipline, although this facet of what James Sidaway calls the 'production of Geography' cannot be easily separated from others. Over the years, geographers have come to have a wide range of outlets available to them for publication of their scholarly work. These include edited books, monographs, reports, and scholarly journals whose contents have been subject to peer review. Scholarly journals have come to dominate publishing in Anglophone Geography.

Academia has long operated within a central conceit of dispassionate and impartial 'objectivity'; one that is reflected in a key metaphor that is often used by supporters and detractors alike to locate and describe the university – 'the ivory tower'. Peer reviewed geographical journals are one of the primary means by which academics, geographers among them, have attempted to maintain the ideals of impartiality and objectivity. This article seeks to describe geographical journals, their role in the production of geographical knowledges, and their place in the messy social relations of the 'real world' that their contributors seek to describe in their research. In this regard, the article can be seen as part of the movement to develop a political and cultural economy of geographical journals (ACME Editorial Collective, 2007; Barnett and Low, 1996; Moss, Berg and Desbiens, 2002). It is important to note two points: (1) this is an interpretation that draws from critical geography for its inspiration (and thus it is open to contestation from other perspectives); and (2) the focus is on Anglo American geographical journals.

Geographical Journals: A Brief Description

For many people, as Rob Kitchin and Duncan Fuller, (2003: 313) observe, publication in geographical journals is best encapsulated by the metaphor of the 'black box': "it has inputs (e.g., submitting a paper or abstract) and outputs (referees comments, acceptance, rejection), but a lack of clarity exists concerning what happens between one and the other" (Kitchin and Fuller, 2005). Given the general lack of knowledge about geographical journals, it is useful to briefly review the development of the peer reviewed journal in Geography and the processes to which manuscripts are submitted prior to publication.

The Learned Societies

The first scholarly research journals in Anglophone Geography were launched by learned societies such as the Royal Geographical Society or the British Association for the Advancement of Science (Section E, Geography). Over time, some of these learned societies came to be known by other names and other new societies were formed to represent academic Geography. In the Anglo American context, these 'newer' societies include the Royal Geographical Society with the Institute of British Geographers in the UK (publishers of *Transactions, Institute of British Geographers*), the Association of American Geographers in the USA (publishers of *Annals, Association of American Geographers*), the Institute of Australian Geographers in Australia (publishers of *Geographical Research*), the New Zealand Geographical Society in Aotearoa/New Zealand (publishers of *The New Zealand Geographer*), and the Canadian Association of Geographers in Canada (publishers of *The Canadian Geographer*). Their journals had relatively similar characteristics, the most significant feature being that they were 'peer reviewed'. The term peer review refers to a number of related processes: review of proposals for research funding, evaluation of ongoing or completed research projects, evaluation of teaching, and the review of manuscripts for publication in academic journals. In this case, then, peer review refers to scholarly review by one's peers of manuscripts submitted for consideration for publication in a scholarly journal. In most cases, such review involves what is termed a 'blind' review, in that neither the author nor the reviewers know the other's identity. There is, however, considerable disagreement among geographers over just how anonymous blind review is, especially in the light of the relatively small numbers of academic geographers extant in the Geographical subdisciplines today. Some journals in Human Geography, including *Antipode* and *ACME* (two radical geography journals), thus operate open by choice reviews in which authors and reviewers can choose to be identified in the review process.

The Rise of Commercial Journal Publication

As a result of both demographic changes (the post war baby boom) and moves by the state to democratize university systems, university education expanded significantly in places like Australia, Canada, New Zealand, the UK, and the USA during the 1950s and 1960s. With these demographic and educational policy changes came significant growth in both the size and number of academic Geography departments, and subsequently, increased pressure on the limited journal publication venues offered by the learned societies. Commercial publishers moved in to fill this vacuum, and by the 1990s a significant majority of journals were being published by commercial publishers, many of which saw an opportunity to extract significant profits from the academic publishing market (on profits in scholarly publication).

During the 1990s, changes in research policies and, in particular, the rise of new audit systems to manage 'research quality' put even greater pressure on geographical journals. The Australian Department of Education, Science, and Training, for example, implemented policies that more or less forced Australian geographers to publish in refereed 'international' journals. There were (and continue to be) significant career consequences for those who did not do so. The story is the same for New Zealand which has implemented its own performance based research funding exercise, modeled extensively on the UK research assessment exercises (RAEs). Similar stories can be found in many other academic settings, as neoliberalization of the academy continues apace, and academic audit and assessment leads geographers to publish more and more in geographical journals as opposed to other forms of publication available to them. This is all part and parcel of what has been termed 'academic capitalism', in which market like behavior and the principles of performance and control become crucial. Anssi Paasi argues that the rise of academic capitalism also results in the homogenization of publication practices in Geography, and there is strong evidence to support his contentions. Moreover, academic capitalism transforms, repositions, and regulates the activities of researchers. Ultimately, "this process increases the power of management and diminishes the autonomy of professional academics" (Paasi, 2005: 773).

The Dominance of Journals in Scholarly Communication

Within academic capitalism, scholarly journals have become crucial to the institutional survival of academic disciplines. While authors may choose to publish in monographs, edited collections, or reports, journal articles are the current stock in trade of geographical publishing in Anglo American Geography. The so called 'international' journals are also coming to dominate the publishing terrain open to geographers from outside the Anglo American metropole, as state funding agencies in other countries encourage their researchers to become 'international' (read: 'published in English') as well.

These processes are all part of the homogenization of opportunities for Geography publishing that Paasi speaks of. Paradoxically, it is within this larger process of homogenization that increased specialization has also occurred in the publishing activities of geographers. Thus, it is important to note the distinction between general Geography journals (some of which are society journals and which publish both human and physical geography), general human Geography journals, and specialist subdisciplinary journals. With a few exceptions (e.g., *Annals of the AAG* and *Transactions of the IBG*), general Geography journals have become less influential in the discipline. Much the same can be said for general human Geography journals (again with exceptions, such as *Progress in Human Geography*). The last one and a half decades have seen a proliferation of specialist journals in Geography, and for the most part these have tended to become more and more influential in (and outside) the discipline. While many serve specialist communities of geographers, a number of them (e.g., *Society and Space* and *Gender, Place and Culture*) have also created interdisciplinary spaces for nongeographers to contribute to spatial understandings of the world.

While the international terrain of journal publication is uneven at best, it is clear that in some Geography publishing cultures, journals have become almost the sole publishing outlets that 'count' as credible research 'outputs' in the calculative rationality of research assessment (of both the formal and informal varieties). For example, less than 10% of submissions to the UK 2001 RAE were books, monographs, or other forms of research publication; 91.1% of all submissions were scholarly journal articles. While the RAE might present a skewed understanding of research 'outputs', primarily because journal articles are easier to assess and they might be deemed by authors and university departments to be more likely to be ranked higher than other forms of publication, it is still safe to assume that scholarly journals are the most important single form of publication in UK Geography. At the same time, given an understanding of the similarities and differences in geographically differentiated cultures of publication, it is logical to infer that journal articles are also the primary form of publishing for academic geographers working in places that comprise the rest of Anglo American Geography: Australia, Canada, New Zealand, and the USA. Other places not normally thought of as Anglo American – and Singapore comes immediately to mind – also rely on scholarly Geography journals to publish the majority of their work (and primarily 'international', English language journals at that). It is important to keep in mind that many other cultures of geographical publishing differ from that found in Anglo American universities, although, as suggested in the discussion in the previous section, in recent years with the almost worldwide rise of new and more coercive forms of 'accountability' in academia, the 'international' journal is fast becoming the *sine qua non* for publication of work by all academic geographers.

The 'Black Box' of Journal Publication

In order to further develop an understanding of the political and cultural economies of Geographical journals, it is useful to outline some of the more mundane and banal processes of journal publication. 'Publish or perish' may be a hackneyed cliché in academia, but that doesn't make the forces behind such statements any less 'coercive'. Most academic geographers have three key components to their university workload: teaching, research, and service. In the administrative discourse of most universities, all three components of academic workloads are valued equally. In practice, however, most geographers know that things related to research – or more specifically, obtaining external funding for research, carrying out that research, and publishing one's findings – matter most. Accordingly, academic geographers are only too aware that they must prioritize these activities over other aspects of their work if they are to 'get ahead' in a university position. Of course, this is an overly schematic representation of these processes, and the balance between research and teaching varies greatly both within and between universities. Notwithstanding these differences and complexities, it is very clear that publication of research 'in the right venues' is a fundamental part of being an academic geographer, and peer reviewed scholarly journals are invariably the venue of choice for geographers.

The manuscript submission and review process

So how does the 'black box' of journal publishing that Kitchin and Fuller (2003) refer to, function? To better understand this, it is useful to follow a 'typical manuscript' through a standard journal submission process.

When an author submits a manuscript to a journal for consideration for publication, they must usually attest to a number of things. First, it must be original scholarly work completed by the author and her/his co authors. Second, it must not have been published elsewhere previously, nor can it be under consideration for publi cation with another journal (or other medium) at the same time. If these initial conditions are met, then the journal editor will quickly review the manuscript to ensure that it meets minimum standards of publication and that it fits with the editorial policies of the journal. If the manuscript meets these minimum requirements, the editor will send a confirmation of receipt to the author and assign the manuscript to reviewers, usually two or three, but possibly as many as six, for peer review. The editor often draws on the expertise of her editorial board members, both to provide the names of potential referees, and to act as possible referees themselves. Some journals (such as *ACME*) have explicit policies that manuscripts be reviewed by at least one editorial board member in addition to other external referees.

Peer review takes time, especially in light of the ever expanding workloads referees face, and the fact that they do peer review as a service that gets little recognition in the university system of merit and promotion – in spite of its absolutely pivotal role in academic knowledge pro duction. Most journals specify a review period of 6 weeks, although most editors are forced to allow 8–10 weeks before they search for an alternate reviewer. Authors should expect to wait at least 4 months for an editorial decision, although some journals take up to a year to provide such decisions.

While such waits are problematic, it should be understood that managing manuscripts, reviews, and re viewers is one of the most time consuming (and trou blesome) aspects that an editor and her editorial staff face. Just to put things in perspective, an accepted manuscript can generate as much as 30 separate pieces of correspondence from submission to publication. Given that most editors manage at least 100 manuscripts in any given year, they and their staff are managing somewhere between 2500 and 3000 letters and e mails relating to manuscript management. Put another way, if each piece of correspondence takes just 5 min to write or read, an editor can spend between 200 and 250 h yr just dealing with manuscript management correspondence. This workload does not include reading manuscripts, referee reports, and writing decision letters, the latter of which can take between three to four hours to complete.

Once the reviews are in, the editor must then review the referee reports, review the manuscript (again), and then write a decision letter outlining the reasons for their decision and linking it explicitly to the referee reports. Editorial decisions are based on more than a simple tallying of the referee decisions for or against publication, but instead attempt to balance a host of issues. While there are many places that an author may succeed or fail, it is much less common in human geography to have straightforward reasons to reject a manuscript for publi cation. The issues in human geography are rarely about technical correctness (although this still happens in some approaches to the discipline), but instead focus on questions relating to the quality of argument, sophisti cation of theoretical approach, internal consistency, and the nature of the contribution to the literature the manuscript might make. A good editor will be working to protect the reputation of both the journal she edits, as well as that of the authors submitting their work. Journal articles become a very public and permanent record of both the journal and the scholars who publish in it, so it is in everyone's interest to have the editor make careful acceptance decisions.

In the majority of cases, manuscripts are rejected with encouragement to revise and resubmit for further review. In the author's 10 years experience as either an editor or editorial board member on six journals, there has prob ably been only one instance of a manuscript being ac cepted for publication without revision. So, let us follow the more typical manuscript that gets rejected with en couragement to revise and resubmit. This manuscript goes back to the author, who then needs to decide whe ther or not to revise and resubmit to the original journal, or whether to revise and submit the manuscript else where. If the editor's letter is at all encouraging, it is good strategy to do the revisions as requested (within reason) and to resubmit to the original journal. If, however, an author decides to submit his/her manuscript to a second journal, they must inform the editor of the first journal of their decision to do so. Failure to do so, results in a manuscript technically being under review at two journals simultaneously, since the editor at the first journal will keep the manuscript file active for approximately one year after sending out a decision letter to an author. Most editors will ask authors to provide some indication of how they wish to respond to the editorial decision.

Once a manuscript has been revised and resubmitted, it is then up to the editor to decide if the revised manuscript must go out to the original referees for a second review. Often, editors have specified in their original decision letter to the author whether or not the manuscript will go back out to reviewers for a second round of revisions. Usually, this decision is aided by referees who specify whether or not the revisions required are of the magnitude that the manuscript needs to be reviewed again. Editors can choose to heed such advice or not, and sometimes their decisions may be based on their own expertise on an issue, or on their need for manuscripts. Most journals have an abundance of submissions, and often a backlog of copy, but publication schedules, juggling of special issues, and other technical matters around page counts, issue size, etc., can mean that the editor must take issues other than solely manuscript 'quality' into account when making acceptance or rejection decisions. If the manuscript goes back out to referees for review, it often goes to just one or two of the original reviewers. This can add a further 6–8 weeks to the process. At various times, the editor's other duties (since most editors do their editorial tasks in addition to teaching, research, and university service, often with only a minimal teaching reduction) can slow the process further.

Once a manuscript has been accepted for publication (whether after the first or second set of reviews), it must then be prepared for submission for production. It seems rare for authors to submit their manuscripts in the required format, so this adds one further step to the process – revision to fit the journal's style guidelines. After the manuscript has been prepared according to submission guidelines, the editor must then make a de cision about when it will be published. This decision is based on the existing backlog, the juggling of special issues, the yearly page limits specified by the publisher (publishers usually have in place substantial financial penalties for going over yearly page allotments, so this is a significant issue for an editor), and other matters that one might not think are relevant to scholarly editorial practice. For example, given current 'citation conscious ness' among journal editors, one issue that concerns many editors is ensuring their best manuscripts get published in early issues, thus giving more time for them to be cited, counted in citation counts, and thus raising their 'impact factor' in journal ranking systems (for a discussion of these issues).

Once the editor has decided on the manuscripts (and other publication materials such as book reviews) that will make up an issue, it is standard practice to submit the manuscripts and materials that make up an issue to the publisher in digital format. Some publishers have FTP sites to which editors may upload their manuscripts, while others allow submission via CD ROM or DVD. It is now becoming quite commonplace for many publishers to use web based submission (and review) systems, such as 'Manuscript Central'. Most commercial publishers require about 6 months lead time to produce an issue. In addition, many journals have significant backlogs of ac cepted manuscripts, and the resulting lag between ac ceptance and publication can be as long as 18–24 months. The hegemony of the geography journal as a form of publishing geographical knowledge and the resulting inability of journals to publish in a timely manner is now a central contradiction in the processes of academic knowledge production.

Power Geometries of Geographical Journal Publication

Of course, the above discussion about the review and editorial decision making process makes the process seem like a series of primarily 'technical' issues that are dealt with by rational, impartial, and objective editors, assisted by similarly disinterested editorial board mem bers and journal staff. This, of course, is what the metaphor of the 'ivory tower' implies. Certainly, it would be inaccurate to suggest that journal editors (and edi torial board members) have an axe to grind. Most editors are fair and even handed in their support for the work that all authors submit to their journals. At the same time, editors clearly have specific editorial policies to fulfill that favor some kinds of knowledge over others.

It is important to remember that regardless of the good intentions of editors, there exist powerful structural relations that interpellate the production of geographical knowledges in Geography journals. To paraphrase Marx, for academic geographers, publishing in Geography journals is all about making geographies, but not under the historical–geographical conditions of their own making. In this regard, it should be clear by now that Geographical journals are enmeshed in the very power geometries of the world that their contributors seek to describe and understand in their journal articles. Geo graphical journals, then, are best understood as 'situated' within both the politics and practices of academe. In the remainder of this article, the author briefly canvasses some of these power geometries and their impact on Geographical journals.

Academic Capitalism and Citation Consciousness

While not all geographers agree on the specific character of academic capitalism, there is growing consensus that there has been an intensification of capitalist like pro cesses arising as part of the neoliberalization of academia. As a result, 'impact factors' and ranking of journals in 'league tables' are now commonplace processes in the working lives of geographers. This is because such things have become an important component of what Michel Foucault (1988) termed 'governmentality', or the gov ernance of individuals from afar.

> The techniques of governmentality disguise as mere technical processes a set of highly political neoliber alizing practices in this case, journal ranking based on impact factors designed to ensure that the work of academics aligns with the objectives of the institutions in which they work (*ACME* Editorial Collective, 2007: 132).

Geographers are thus being asked to become re sponsible for the academic capitalist accumulation strategies of universities, and one way of attempting to ensure that they do so is through the disciplining prac tices of impact factors and journal ranking systems. Geographers have long had an informal ranking of the status and quality of Geography journals. However, the institutionalization of such ranking processes means that Geographical journals are themselves having to align with the calculative desires of impact factors, rankings, and league tables. This is highly problematic for a number of reasons.

First, there is presently only one source of data for impact factors and rankings:

> [t]he effective monopoly supplier of such statistics is the Philadelphia based ISI, formerly known as the Institute for Scientific Information, and now owned by the Thomson Corporation (*Nature*, 2002: 101).

In a number of cases investigated by editors of the journal *Nature*, the raw citation data collected by ISI Web of Science were incorrect, and further mistakes were found in impact factor calculations. The editors of *Nature* (2002: 101) concluded that such problems "…highlight an even greater need… to check the ISI's data. Researchers, policy makers, and publishers who depend heavily on citation statistics should be urged to treat them with greater caution." The problem is that in the current climate of audit and accountability driven by research assessment of both the formal and informal kind, academic administrators are becoming ever more reliant upon citation statistics, impact factors, and journal rankings, to evaluate the merit of scholarly publication. Thus status, funding, and career advancement for geographers are all becoming more closely linked to publication in only the top ISI ranked journals. Personal communications with many UK based colleagues make it clear that more than a few Geography departments have internalized this 'citational consciousness', and they now require academic staff to publish in journals ranked in the top ten of the ISI ranked Geography journals.

For-Profit Publishing and the Narrowing of Choice

If citation consciousness linked to ISI rankings has a tendency to reduce the number of journals to which geographers submit their work, then the growing trend toward commercial publication has further reduced the availability of journals in academic research libraries. Human Geography journals are less affected by this process, as most titles cost less than USD$1000.00 per year. This problem is most acutely felt in the physical sciences where some journals can cost tens of thousands of dollars per year for subscriptions. Studies commissioned by the Association of Research Libraries (a non profit organization of 123 research libraries at comprehensive, research extensive institutions in the US and Canada) almost a decade ago suggest that there was a growing crisis of affordability in serials publications. Mary Case notes that, in the decade between 1988 and 1998, the average cost increases to research libraries for serials was 9.5% per annum. Inflation during that period was usually less than 5% per annum.

As a result, libraries have had no choice but to cancel significant numbers of journal subscriptions and to reduce monographic purchasing, decimating their collections. Many believe the exploitative practices of a few large commercial companies operating in a near monopolistic market are the fundamental cause of high prices.

An earlier 20 year longitudinal study of scientific journal publishing in the USA found that neither increases in journal pages and number of issues, nor inflation, could adequately account for price increases made by commercial publishers. In other words, it appears that commercial publishers are making significant increased profits on the backs of a 'captured market' – academic research libraries. Commercial publishers are also exploiting state funders of the universities that pay the salaries of academics who contribute their manuscripts and their editorial expertise to journals as absolutely necessary components of the (commercial) journal publication process. Finally, commercial publishers exploit state research councils that fund the research that forms the basis upon which scholarly publications are built.

The issue, of course, is that because of significantly increasing costs, journal collections, including those of Geographical journals, are shrinking in university libraries. Not only are there fewer venues in which to publish – because of the pressures arising from academic capitalism and citational consciousness – but now there are fewer journals being picked up by academic libraries. This is a positive feedback loop in which fewer journals are available to academics, thus reducing the number of journals with higher citation counts, thereby leading to fewer journals for academics to choose from in order to submit their work.

Academic capitalism has another, closely related set of implications for scholarly journal publication, namely the hegemony of Anglo American approaches in the academy, which arises from the domination of Anglo American journals in Geography (and all other disciplines too). The issue of Anglo American hegemony in Geography publication has been well canvassed elsewhere, so my comments here are brief. The key point is that both citation consciousness and the dominance of commercial publishers seeking profit maximization have led to a concentration of English language journals published from either the UK or the USA predominating the scholarly publishing terrain in Geography. One small example should illustrate this point: of 39 Geography journals ranked by ISI, just two – *Mitteilungen der Osterreichischen Geographischen Gesellschaft* and *Geographische Zeitschrift* – can be said to be truly non Anglo journals in that they publish in a language other than English and their editorial boards are dominated by scholars working outside the Anglo American 'core' of Geography. Just as Anssi Paasi suggested, we are indeed seeing the hom ogenization of Geographical journal publishing outlets.

Open Access?

Recognizing these problems, there have been some movements to contest commercial control of journals through the so called 'open access' movement. Open access journals can be defined as peer reviewed scholarly journals that are available free to readers (and insti tutions), usually on the Internet, and at the time of initial publication. The Directory of Open Access Journals is the best available source for locating open access jour nals, and its database currently holds 26 entries for 'geography'. Of these, 18 journals publish at least some editorial materials in English (and these journals are listed in the list of open access journals below). Un fortunately, the potential of an independent, open access system of scholarly publication (via the internet) has not been realized to date. As Mike Sosteric concluded:

> In the context of the crises in the scholarly communi cation system, proselytizing a future for independent publication was a laudable project. However, the vision of independent communication has not been realized pri marily because scholars have been reluctant to take up the call to independent publication. Despite the powerful potential of information technologies to wrest control of the scholarly communication system from the big com mercial publishing houses, relatively few individual scholars have initiated journal projects. It is perhaps this failure on the part of the academic community to move quickly to independent publication that has encouraged commercial houses to grow bolder in their demands. (Mike Sosteric, 1998)

Knowing the power of citation consciousness and the lack of rewards within the academy for scholarly journal editing, it is completely understandable that academics would be reticent to engage the open access movement more fully. However, it is this lack of commitment to developing a system of open access scholarly communi cation, as well as the desire on the part of academics for quick and easy access to journal articles via the internet, that has given commercial publishers even more 'mon opolistic powers in scholarly publication. Commercial publishers have developed pricing structures for their journals that involve requirements to purchase 'both' print and online versions of their journals, and at prices that the International Coalition of Library Consortia "fears will eventually add as much as 40% or more to the cost of scholarly material in journals (Sosteric, 1998)."

Conclusion

This article has outlined, albeit briefly, a political and cultural economy of Geographical journals. Such journals became a prominent feature of the publishing landscape in Geography in the mid twentieth century, and by the end of the century had come to dominate academic publishing in Geography. At the same time that the journal article became the *sine qua non* of publication for geographers, commercial publishing houses came to dominate the journal publishing market. These processes were embedded within wider processes in the academy that saw the rise of academic capitalism and the creation of a form of citation consciousness. In turn, this has handed a great deal of control of the scholarly publishing system in Geography to commercial publishers who now operate both the actual publication of journals, and the measurement of those journals' 'value' through impact factor calculations and journal rankings. Most geog raphers are strongly discouraged from publishing in anything but highly ranked journals, as measured by ISI Web of Science Journal Impact Factors. As a result, and in spite of the proliferation of journals being published in the latter half of the twentieth century, the terrain of Geographical journals is becoming less varied and more strictly policed by structural constraints imposed by the capitalist like practices arising within academic Geography.

See also: Citation Geography; Research Funding Bodies.

Further Reading

ACME Editorial Collective (2007). The politics of indexing and ranking journals. *ACME: An International E Journal for Critical Geographies* 6(2), 131 134.

Barnett, C. and Low, M. (1996). Speculating on theory: Towards a political economy of academic publishing. *Area* 28, 13 24.

Berg, L. D. (2001). Masculinism, emplacement and positionality in peer review. *Professional Geographer* 53(4), 511 521.

Berg, L. D. (2004). Scaling knowledge: Towards a *critical geography* of critical geography. *GeoForum* 35(5), 553 558.

Berg, L. D. (2006). Hierarchical space: Geographers and neoliberalism. *Progress in Human Geography* 30(6), 764 769 and 778 782.

Berg, L. D. and Kearns, R. A. (1998). America unlimited. *Environment and Planning D: Society and Space* 16, 128 132.

Case, M. (1998). Views of the current marketplace for scholarly journals. *ARL: A Bimonthly Report*, no. 200. No pagination. http://www.arl.org (accessed 15 July 2008).

Castree, N., Aspinall, R., Berg, L. D. *et al*. (2006). Research assessment and the production of geographical knowledge. *Progress in Human Geography* 30(6), 747 782.

Castree, N. and Sparke, M. (guest eds.) (2000). *Special Issue: Professional Geography and the Corporatisation of the University. Antipode* 32(3), 322 339.

Ferguson, R. (2003). Publication practices in physical and human geography: A comment on Nigel Thrift's 'The future of geography'. *Geoforum* 34, 9 11.

Foster, J., Muellerleile, C., Olds, K. and Peck, J. (2007). Circulating economic geographies: Citation patterns and citation behaviour in economic geography, 1982 2006. *Transactions of the Institute of British Geographers* NS 32, 295 312.

Foucault, M. (1988). Technologies of the self. In Martin, L. H., Gutman, H. & Hutton, H. P. (eds.) *Technologies of the Self: A Seminar with Michel Foucault*. Amherst, MA: University of Massachusetts Press.
Gutiérrez, J. and López Nieva, P. (2001). Are international journals of human geography really international? *Progress in Human Geography* 25, 53 69.
Johnston, R. J. (2003a). Geography: A different sort of discipline. *Transactions of the Institute of British Geographers* NS 28, 133 141.
Johnston, R. J. (2003b). The institutionalization of geography as an academic discipline. In Johnston, R. J. & Williams, M. (eds.) *A Century of British Geography*. Oxford: Oxford University Press.
Kitchin, R. (2006). Research assessment in Ireland. *Progress in Human Geography* 30(6), 749 751 and 778 782.
Kitchin, R. and Fuller, D. (2003). Making the 'black box' transparent: Publishing and presenting geographic knowledge. *Area* 35(3), 313 315.
Kitchin, R. and Fuller, D. (2005). *The Academic's Guide to Publishing*. London: Sage Publications.
Lee, D. and Evans, A. (1984). American geographers' rankings of American geography journals. *Professional Geographer* 36(3), 292 300.
Lee, D. and Evans, A. (1985). Geographers' rankings of foreign geography and non geography journals. *Professional Geographer* 37(4), 396 402.
Mitchell, K. (1999). Scholarship means dollarship, or, money in the bank is the best tenure. *Environment and Planning A* 31, 381 388.
Moss, P., Berg, L. D. and Desbiens, C. (2002). The political economy of publishing in Geography. *ACME: An International E Journal for Critical Geographies* 1, 1 7.
Nature (2002). Errors in citation statistics. *Nature* 415(6868), 101.
Paasi, A. (2005). Globalisation, academic capitalism, and the uneven geographies of international journal publishing spaces. *Environment and Planning A* 37, 769 789.
Sheppard, E. (2006). Geographies of research assessment: The neoliberalization of geography? *Progress in Human Geography* 30(6), 761 764 and 778 782.
Slaughter, S. and Leslie, L. (1997). *Academic Capitalism: Politics, Policies and the Entrepreneurial University*. Baltimore: Johns Hopkins University Press.
Sosteric, M. (1998). At the speed of thought: Pursuing non commercial alternatives to scholarly communication. *ARL*: *A Bimonthly Report*, no. 200. http://www.arl.org (accessed 15 July 2008).
Stoddart, D. (1983). Progress in geography: The record of the I.B.G. *Transactions, Institute of British Geographers* NS 8, 1 13.
Tenopir, C. and King, D. W. (1997). Trends in scientific scholarly journal publishing in the United States. *Journal of Scholarly Publishing* 28(3), 135 170.
Withers, C. W. J., Finnegan, D. and Higgit, R. (2006). Geography's other histories? Geography and science in the British Association for the Advancement of Science, 1831 *c.*1933. *Transactions, Institute of British Geographers*, NS 31, 433 451.

Relevant Websites

www.acme journal.org
ACME: An International E Journal for Critical Geographies.
http://www.zrc sazu.si
ACTA Geographica Slovenica.
http://www.anuario.igeo.ufrj.br
Anuário do Instituto de Geociências.
http://www.sunysb.edu
Coordinates: Online Journal of the Map and Geography Round Table of the American Library Association. Series A.
http://www.cybergeo.eu
CyberGeo: European Journal of Geography.
www.doaj.org
Directory of Open Access Journals.
http://www.ff.uni lj.si
Dela.
http://www.geo focus.org
GeoFocus: International Review of Geographical Information Science and Technology.
www.geo publishing.org
Geo Publishing: A Publishing Resource for Geographers.
http://www.geotropico.org
GeoTrópico.
http://igeographer.lib.indstate.edu
Industrial Geographer.
http://redalyc.uaemex.mx
Investigaciones Geográficas.
http://www.journalofmaps.com
Journal of Maps.
http://metropoles.revues.org
Métropoles.
http://nome.unak.is
Nordicum Mediterraneum.
http://revistaurbanismo.uchile.cl
Revista de Urbanismo.
http://www.soc geogr.net
Social Geography.
http://www personal.umich.edu
Solstice: An Electronic Journal of Geography and Mathematics.
http://redalyc.uaemex.mx
Tabula Rasa.

Geographical Masking

G. Rushton, The University of Iowa, Iowa City, IA, USA

Glossary

Data Enclaves Supervised work areas where use and access to individual personal records is closely monitored to ensure that the privacy and confidentiality of personal records is ensured.

Geographic Masking Any trasnformation of spatial data disigned to protect the privacy of individuals by concealing their identity from anyone with access to geocoded information about them.

***k*-Anonymity** Identification of a given number of individuals, *k*, among whom a given individual's personal information is merged.

Mask Metadata Specific information about the mask that has been used to guarantee personal privacy in the release of information about an individual.

Reverse Address-Matching The coordinates of a symbol on a map are linked to a street address file to determine the likely street address of the map symbol.

Spatially Adaptive Mask Identification of an area, based on local information about it, that is of sufficient size to merge to given individual's personal data and preserve the confidentiality of their information.

Spruill Coefficient The proportion of records that are located closer to their parent record than to any other record.

Introduction

There is a great deal of interest in human geography in examining human behavior in space and time, and for this purpose many activities of individuals are now routinely geocoded. However, there is a risk that individuals will become identified from location information about them and so methods have been developed to protect their privacy. One of the methods is known as 'geographical masking', and is the subject of this article.

The term 'geographical masking' refers to any transformation of spatial data designed to protect the privacy of individuals by concealing their identity from anyone with access to geocoded information about them. Examples where such masks are used include the health information of individuals for disease maps, the locations of crimes for crime maps, or the store visited in a study of household spatial behavior for making models of spatial retail behavior. Masking methods are thus methods for concealing locations, or other data associated with locations to protect personal privacy and assure confidentiality.

When spatial data have been masked, typical questions asked are how effective is the mask at protecting the true locations from discovery by others and how the purposes for which the data are released may be met as well after being masked as before. Geographic masking is thus one of a class of methods for limiting the disclosure of personal information. Interest in geographical masking has increased recently with the widespread availability of fine grained spatial data that can be used to link geographic coordinates to demographic, social, and economic variables in a geographic information system. Matching the addresses of people in a database to a digital street map to find their location or, knowing their place on a map, recovering their addresses are examples of how spatial information is acquired or used.

Threats to Privacy and Anonymity: Disclosure Risk

The need for geographical masking arises from the observation that in many cases the location of an individual is a unique identifier. The ethical conduct of research in the social sciences demands that individuals are not harmed by either the conduct of the research, its dissemination, or its use. Professional academic societies have codes of conduct that spell out responsibilities to protect the confidentiality of research subjects or any individual explicitly covered by the research. Most countries have laws that specify standards of privacy and confidentiality to which holders of personal data or distributors of such data must conform. In addition, institutions that engage in research using information provided by individuals follow standards and procedures for ensuring that research is done according to ethical principles. The United States, for example, has institutional review boards (IRBs) that review proposed research and ensure that it follows established procedures to protect the privacy and safety of individuals. Confidentiality considerations relate to how information about individuals is handled by others who have access to it.

Does the Location of an Individual Identify the Individual?

The degree to which knowledge of the location of individuals potentially identifies them depends on the degree of specificity of the location provided, other information characteristics of the individuals that could be linked to the respective individuals, and the number of other people with whom an individual could potentially be confused. In the world of digital information, knowing the latitude and longitude of a person in decimal degrees accurate to six decimal places, identifies a person in absolute terms to within approximately 2 m. Whether such absolute location coordinates are known with this accuracy depends on the method by which the coordinates were established. Possible sources are global positioning system (GPS) measurements which vary in accuracy by the technology used and the quality of the instruments used; geocodes that are read off properly registered orthophoto maps; and address matching technologies which associate the postal address of an individual with digital road files containing either individual addresses, given ranges of addresses, or land parcel files. In most cases, any of these sources of location information are sufficiently accurate either to uniquely identify an individual or to associate a small group of individuals with the data source.

Spatial Aggregation as a Geographic Mask

Throughout the world, the disclosure of personal information in census data is most commonly accomplished by aggregating data into groups or classes and releasing the data for regions. In such cases, spatial aggregation is the method that accomplishes the purpose of the geographic mask. In such cases, individual data are geocoded to known geographic entities. With US Census data, this might be a census tract number; in the United Kingdom it might be a Census Ward. Although this approach can be effective for the purpose of protecting individual information, it often reduces the usefulness of the information for many important purposes. Studies of relationships between environmental exposures and health, for example, often rely on estimates of exposure to harmful contaminants made at particular locations and estimates of the locations of people close to these locations. If information on exposures are averages for areas and information on health outcomes are also averages for areas, it may be impossible to determine whether the exposures at particular locations led to the poor health outcomes at those locations because the size of the geographic mask conceals a relationship that really exists. In other cases, a researcher might want to know whether there is a relationship between poverty and health but using spatially aggregated census data, such correlations suffer from the ecological fallacy in which relationships observed between census values for areas are different from relationships measured between individuals in these areas.

Cartographic Display

It is not uncommon for data to be shown as dots on a map where each dot refers to an event. It is possible technically to identify the address associated with the dot by a process known as ‘reverse address matching’ in which the coordinates of the symbol on the map are matched against the coordinates of the street center line data set. With the likely address identified, links can be made to other data sources that identify a person linked to the address. A number of studies have shown, using empirical examples, that a very large percentage of the addresses illustrated in such maps can be exactly recovered by this process. To obviate this disclosure risk, a geographic mask can be used that relocates any information to prespecified locations. Any information from a given street segment may be shown at the midpoint of the street segment, no matter where, in fact, it is located. If insufficient numbers of other people live on the street segment, the rule might be modified and the information may be shown at the closest street intersection.

One negative aspect of applying a geographical mask is that it raises the risk of false identifications. That is, when individual data are mapped, any attempt to protect an individual by attaching his or her information to a location which in fact is not their location runs the risk that some other person who lives at the location identified in the masked data may be identified by others as the person whose data are shown. Especially when the activities portrayed in the data are activities that a person does not wish to be associated with such as crimes, rapes, infectious diseases, and many other activities, a person falsely identified as the individual associated with the event will likely object and have a fair argument that they have been harmed by the false disclosure. To overcome such resistance to portraying the locations of individual events, it is not uncommon to place disclaimers on maps advising users that the locations shown are not the real locations of the events. One publicly available web map showing the locations of registered sex offenders had a note attached “PLEASE NOTE: The symbols identifying the street location DO NOT represent the exact location of where the offender lives. The symbols have been enlarged and offset to keep an exact location from being determined.”

Principles of Masking

Scale, Translation, and Rotation (Affine Transformations) Masks

This commonly used mask moves each location a particular distance and direction from its true location. This mask is easy to implement and can be effective in concealing the true locations of each person shown. Should even one person become correctly located, however, the locations of all other persons immediately become compromised – which is an acknowledged weakness of this mask. When the scale of the map is changed, the information shown can only with difficulty be re connected to other geographic information. When using a scalar transformation, if the map with information on individuals that needs to be protected shows any add itional information beyond the point locations, it can often be rescaled using the ancillary information and then the mask is no longer effective. In short, these transformations can be effective in limited circumstances but they are inherently weak.

The Random Perturbation Mask

The random perturbation mask displaces each location by a randomly determined amount within a specified range and by a direction selected by a uniform random angle between 0° and 360°. An alternative specification of this mask which has the desirable property of placing the displaced location with equal likelihood within the area of the mask is to select the displaced location from a uniform random distribution within a given distance range east–west and then north–south of the real lo cation. The displaced location is any location selected at random from the set of points computed by this process. Sometimes constraints are added to ensure that all dis placed locations are feasible locations for the event in question, or to ensure that the locations are within a given shape chosen for the mask. If, for example, the displaced event is a residential location, the constraint may ensure that the displaced location is located where residences are found. The random perturbation mask is a commonly used mask and, providing parameter values for the mask are selected with care, this mask is robust and risks of disclosure of personal identities are low when it is used.

A Spatially Adaptive Mask

There are many factors that affect disclosure risk and it is often not immediately clear how these factors, working in combination, contribute to disclosure risk. VanWey *et al.* show how a research study that used a sample of US schools, each of which had been guaranteed protection from identification, could determine the size of the geographic mask each should use to perturb the location of the true school so that the approximate pattern of sample schools could be shown on a map. In making such a map available to the public, the study has the additional task of conveying to the public that the schools in the sample were not being identified – or more accurately misidentified – on the map. A possibility in this case was to use the principle of *k* anonymity to protect the iden tity of the schools in the sample. This principle selects the number of other schools that may be selected to represent each school. If $k = 20$, for example, then the number of miles distant from the sample school that is needed to find 20 actual schools becomes the size of the geographic mask for the school in question. Since this size would be different for each school in the nation, the result is the definition of a 'spatially adaptive geo graphic mask'. The ability to implement such a mask does depend on the availability of a data file of all schools eligible to be included in the sample, their geographic coordinates, and an analysis system capable of making the measurements in question. The spatially adaptive mask is becoming more widely used as potential users recognize that it has all the strengths of the random perturbation mask with additional protections in some specific situations.

Attribute Perturbation Mask

For some kinds of research, knowing the attribute of a location is sufficient for it to be identified using external data sources. An example might be knowledge gained from a detailed map that estimates contaminant ex posures by people living at any location. A data set that supplied health data and contaminant exposure measures for individuals might be sufficiently precise that the lo cations of the residences could be 'discovered' by searching the map for the exposure level associated with the health data. An attribute perturbation mask adds a random value of the attribute – in this case contaminant exposure value – to each case to mask the ability to recover the location through a systematic search of the exposure contaminant map. The range within which the attribute is perturbed can be computed based on the density of people at particular locations on the map. In this case, the mask would be called a 'spatially adaptive, attribute perturbation mask'.

Pros and Cons of Concealing Details of the Mask

An issue that affects the effectiveness of a mask in ac complishing its confidentiality objective is the infor mation divulged to users of a data set about the mask that was used to provide privacy protection. In general, the more details of the mask are divulged, the less effective is the mask at accomplishing its confidentiality objective. Information about a mask is known as 'mask metadata'. Knowing details of the mask is very useful to the user of the data who can use this information to determine whether the conclusions they have reached after ana lyzing a given data set is likely to be valid, given the characteristics of the mask used. The ability to infer such conclusions is often reached after sensitivity analyses on simulated data that show how results of quantitative tests are sensitive to particular values of the mask. When a geographic mask has been used, there are different views about releasing details of the mask. Anyone using spatial information that has been masked would like to know all the characteristics of the mask in order to decide whether the masked data can be used for their purpose. However, knowing the characteristics of the mask helps anyone who would like to find the identities of people in the database because it helps identify those individuals who might be the persons being protected by the mask. Thus, withholding the characteristics of the mask helps protect any given individual.

Effectiveness of Geographical Masks

The effectiveness of a mask can be measured in two ways. The first is the degree to which the mask protects the privacy of an individual. The second is the degree to which the results of spatial analyses on the masked data are the same or, at least, are consistent with results of analyses on the unmasked data. A common way of measuring the first is by the Spruill coefficient which measures the proportion of records that are located closer to their 'parent' record than to any other record. The larger the proportion, the greater the risk of disclosure for the person at the arbitrary location chosen. Related conceptually to the Spruill measure is the '*k* anonymity' principle which specifies that a mask should supply at least *k* other individuals with whom a given individual would be identified by a given mask. The privacy of the given individual is assured by the fact that the characteristics of all the *k* other individuals could be attributed to the given individual. The effectiveness of such a mask depends on the fact that no other characteristics of the individual are known. If other characteristics of the individual are known, the risk of disclosure increases since all members of the *k* anonymous set who do not have the known characteristic would then be excluded from the set which could decline drastically in number. Armstrong *et al.* describe the information preserving qualities of various geographic masks.

Alternatives to Geographic Masking

Because of the risk of disclosure of individual data through data linkages, some organizations that archive or produce data attempt to reduce the risk of disclosure by restricting access to data to individuals whose research objectives have been approved and who agree to work in a monitored environment where their access to data is supervised and any data they take from the protected environment is examined to ensure that future disclosure is impossible. These facilities are called 'enclaves'.

Data sharing agreements are a second alternative to geographic masking. Such contractual agreements are made between willing partners where researchers promise not to engage in data linkages that might lead to the identification of records. The data user generally agrees to restricted terms of data use, including accepting controls by the data provider over the release of geographical details of the data. An example of a data use agreement (DUA) can be seen at website of Centre for Medicare and Medicaid Services.

A third alternative is to process the geocoded data to produce a 'de identified data set' prior to release to researchers or to the public. In the USA, the HIPAA law provides explicit rules for producing a de identified data set. Holders of data that follow these rules are said to be invoking the 'safe harbor' clause of the HIPAA privacy law and are safe from prosecution. Current rules in the USA stipulate that in a de identified data set, the following identifiers of any individual are removed:

> All geographic subdivisions smaller than a State, including street address, city, county, precinct, zip code, and their equivalent geocodes, except for the initial three digits of a zip code if, according to the current publicly available data from the Burea of the Census; (1) the geographic unit formed by combining all zipcodes with the same three initial digits contains more than 20 000 people; and (2) the initial three digits of a zip code for all such geographic units containing 20 000 or fewer people is changed to 000. (http://www.hipaaps.com/legislation/45CFR 164.514.html)

A fourth alternative is for the holder of individual level information associated with exact or near exact information on their locations make available, usually on a website, a secure server which is capable of providing selected spatial analyses of their data. The server is organized not to allow any original data to be returned to the enquirer; instead, it returns results of analyses which do not include any individually identifiable information either directly or by inference by connecting the results to any other information in the possession of the enquirer. In such cases, as Boulos *et al.* note, only the software agent 'sees' the locational detail in the data and it is the software, associated with control over the server that functions as a geographic mask. This approach to geographic masking is likely to prevail in many socially sensitive areas of human geography in the future.

Conclusions

Given the vast amounts of information in spatial databases held by many organizations, interest in the ethical uses of such information and in preventing the disclosure of personally identifiable information, interest in the responsible release of data will undoubtedly grow. We have seen that there are many different ways to mask geographical detail to fulfill these ethical purposes. No single mask is more effective than all others. Masks limit detail, add noise, and search for the myriad ways, and attempt to thwart other data that might be linked to those released in order to identify individuals. Work in this area of human geography will continue. Research on the ef ficacy of masking methods is in its infancy.

See also: Ethical Issues in Research.

Further Reading

Armstrong, M. P. and Ruggles, A. J. (2005). Geographic information technologies and personal privacy. *Cartographica* 40(4), 63 73.

Armstrong, M. P., Rushton, G. and Zimmerman, D. L. (1999). Geographically masking health data to preserve confidentiality. *Statistics in Medicine* 18, 497 525.

Boulos, K. M. N., Cai, Q., Padget, J. A. and Rushton, G. (2006). Using software agents to preserve individual health data confidentiality in micro scale geographical analyses. *Journal of Biomedical Informatics* 39, 160 170.

Brownstein, J. S., Cassa, C. A. and Mandl, K. D. (2006). No place to hide   reverse identification of patients from published maps. *New England Journal of Medicine* 355(16), 1741 1742.

Kwan, M. P., Casas, I. and Schmitz, B. C. (2004). Protection of geoprivacy and accuracy of spatial information: How effective are geographical masks? *Cartographica* 39, 15 28.

Leitner, M. and Curtis, A. (2004). Cartographic guidelines for geographically masking the locations of confidential point data. *Cartographic Perspectives* 49, 8 25.

Monmonier, M. (2002). *Spying with Maps: Surveillance Technologies and the Future of Privacy.* Chicago, IL: University of Chicago Press.

National Research Council (2007). In Gutmann, M. P. & Stern, P. C. (eds.) *Putting People on the Map: Protecting Confidentiality with Linked Social Spatial Data. Panel on Confidentiality Issues Arising from the Integration of Remotely Sensed and Self Identifying Data*. Washington, DC: The National Academies Press.

Ratcliffe, J. H. (2002). Damned if you don't, damned if you do: Crime mapping and its implications in the real world. *Policing and Society* 12(3), 211 225.

VanWey, L. K., Rindfuss, R. R., Gutmann, M. P., Entwisle, B. and Balk, D. L. (2005). Confidentiality and spatially explicit data: Concerns and challenges. *Proceedings of the National Academy of Sciences of United States* 102(43), 15337 15342.

Waldo, J., Lin, H. S. and Millett, L. I. (eds.) (2007). *Engaging Privacy and Information Technology in a Digital Age*. Washington, DC: National Research Council, The National Academies Press.

Zimmerman, D. L., Armstrong, M. P. and Rushton, G. (2007). Alternative techniques for masking geographic detail to protect privacy. In Rushton, G., Armstrong, M. P., Gittler, J. *et al*. (eds.) *The Use of Geographic Codes in Cancer Prevention and Control, Research and Practice*, pp 127 137. New York: CRC Press. ch. 7.

Relevant Websites

http://www.cms.hhs.gov
Centre for Medicare and Medicaid Services.

http://www.hipaaps.com
HIPPAs Law.

Geographically Weighted Regression

A. Páez, McMaster University, Hamilton, ON, Canada
D. C. Wheeler, Emory University, Atlanta, GA, USA

Glossary

Bayesian Statistics A branch of statistics that incorporates prior information about phenomena with current data in a statistical model to perform inferences on model parameters.

Bonferroni Adjustment A technique that decreases the statistical significance value of individual tests in a situation where multiple significance tests are carried out simultaneously.

Collinearity The presence of correlation between explanatory variables in a statistical model that may interfere with statistical inference.

Cross-Validation A model parameter estimation technique that systematically removes observations from a dataset to minimize prediction error.

Distance Decay A way of describing the diminished interaction or correlation of locations as distance between them increases.

Expansion Method A method for estimating spatially varying parameters defined as deterministic functions of the coordinates of observations.

Hierarchical Bayesian Model A statistical model that uses unknown random parameters that are in turn specified with other parameters.

Kernel Function A mathematical function that accepts an input of distance between observations and returns a value that represents a measure of association.

Kriging A method for spatial interpolation that models the covariance of observations at certain distances over space.

Spatial Autocorrelation A dependent relationship between the values of a variable due to the geographical arrangement of units on a map.

Spatial Heterogeneity The variation in a condition or relationship from one place to another.

Spatial Interpolation A method for prediction of an unknown value of a spatially continuous variable at a location that uses the values of neighboring locations.

Spatial Nonstationarity A condition where the relationship between variables is not constant across a study area.

Stationarity The presence of a constant mean and a covariance function that depends only on distance between locations in a spatial process.

Background

Geographically weighted regression (GWR) is a local form of spatial analysis introduced into the geographical literature drawing from statistical approaches for curve fitting and smoothing applications. The method works based on the simple yet powerful idea of estimating local models using subsets of observations centered on a focal point. By placing GWR in the context of analytical geography, this modeling approach has appealed to the geographical imagination by implementing, within a locally linear regression framework, the distance decay effect popularized by Tobler with the first law of geography.

GWR fits into the tradition of analytical geography. Following important developments in human geography in the 1950s and 1960s as part of the so called quantitative revolution, this technique is part of established practice in geography of using quantitative techniques and models as tools to assist discovery. Recognizing the fact that models are abstractions and simplifications of reality, statistical models were adopted in early times for their ability to describe or explain observed effects in terms of a reduced set of relationships with explanatory factors thought to be relevant for a specific process. In particular, the following model became a workhorse of geographical analysis:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i \\ &= \beta_0 + \sum_K \beta_k x_{ki} + \varepsilon_i \end{aligned} \qquad [1]$$

This model is a simplified and stylized representation of reality, developed to describe or explain the systematic variation of a phenomenon, while allowing for uncertainty. The model relates a number (k) of relevant explanatory (or independent) variables x to a response (or dependent variable) y for unit of observation i ($i=1, 2, \ldots, n$). The relationships between response and explanatory variables are given by coefficients β_k which represent the mean response. The model is called a regression, since it 'regresses' the variables toward the mean. In addition to this mean structure, uncertainty enters through a set of terms ε_i, variously called residuals, errors, or disturbances. These terms, estimable from the data once a model has been specified, measure the difference between the systematic (mean) description of the process, and the observed outcome.

The usual assumptions regarding the residuals ε in regression analysis are randomness (i.e., all systematic variation is assumed to be accounted for in the model), an expected value of zero, independence, and constant variance. Soon geographers realized that some of these assumptions, in particular independence, were often untenable when the location of the observations was introduced. Technical developments initially focused on inferential analysis of spatial patterns (e.g., Moran's I and Geary's c statistics of spatial autocorrelation), and later tools to model spatial pattern (spatial autoregression, and regression with spatially autocorrelated errors). The techniques matured, and by the mid 1990s, thanks in part to new software, spatial analysis had found application in many areas of geography, and had begun to transcend the discipline to become routinely used elsewhere, notably in civil engineering, real estate economics, and regional and urban studies. But some analysts realized that, despite their ability to capture map patterns, the techniques in use (autocorrelation statistics, regression analysis, spatial regression) were still insufficiently geographic. Auto correlation statistics were able to detect map patterns, but did not produce mappable results. In other words, they did not identify locations where spatial effects were relevant or strong. The regression model in eqn [1] il lustrates this situation by assuming that relationships (the coefficients β_k) are stationary or constant across space.

Subsequently, spatial statistical analysis and modeling came to be seen as an exercise in the discovery or val idation of global trends or averages – that is, in the study of stationary relationships assumed to represent equally well, anywhere within the system, the phenomenon being examined. It has been suggested, however, that the search for general trends and rules overlooks contexts, and especially geographical variations. There has since emerged an interest in forms of spatial analysis that focus on what Michael Goodchild has termed the second law of geography, or the principle of spatial heterogeneity. These methods yield more accurate and sophisticated representations of the world.

Geographically Weighted Regression

Among the various exploratory and explanatory methods for local spatial analysis, GWR has attracted considerable attention. This method was proposed in 1996 by Bruns don, Fotheringham, and Charlton to study spatial para metric nonstationarity – the variable nature of relationships in geographical space. GWR is related to the nonparametric technique known as locally weighted regression, developed in statistics for smoothing and curve fitting applications based on the simple yet powerful idea of estimating statistics using subsamples of data around specific points. The innovation in the case of GWR, consisted of the selection of subsamples of data based on distance in geographical space, and the notion that, unlike a majority of applications in statistics whose main concern is with curve fitting (in a predictive sense), nonstationary relationships could be investigated as well, in an exploratory or explanatory sense.

GWR is defined as a model that is locally linear in the mean:

$$y_i = \beta_{0i} + \sum_K \beta_{ki} x_{ki} + \varepsilon_i \qquad [2]$$

The elements of the model are as defined above, except that now the coefficients β_{ki} are not stationary but cor respond to location i, termed the focal point, which is the point selected for local model estimation. In general, the number of local models m is not fixed and an arbitrary number of them ($m \geq 1$) can in principle be estimated for a region. In many applications, the focal point for each local estimation is made to coincide with the location of an observation i, although this need not be the case. In this sense, coefficients can be estimated for any location within the study area, regardless of whether data have been col lected at that point – say, a generic focal point o. Given covariate data, the response can be estimated at unsampled locations as well, a feature of the model that makes it attractive for spatial interpolation purposes that closely resemble the origins of GWR. The primary feature of GWR of interest in the geographical literature concerns the coefficients in eqn [2], which are specific to location i.

By obtaining location specific coefficients, GWR provides a tool to investigate the issue of geographical non stationarity. An example of GWR output is shown in **Figure 1**. A model of proportion of bachelor degree holders (dependent variable) is estimated using different socioeconomic and demographic measures as covariates, including proportion of the population that is rural, proportion of population below the poverty line, etc., using the counties in Georgia, US as units of analysis. The results of the regression suggest a negative rela tionship between proportion of rural population in a county, and the proportion of degree holders in the state of Georgia. However, this relationship appears to be considerably stronger (i.e., more negative) in the south east parts of the state. Similar output can be obtained for all coefficients of the model.

Estimation of the Model Parameters

The GWR model is relatively simple, and involves weighting observations depending on their relative lo cation with respect to a focal point. The weights w_{ij} re lating focal point i to each observation j in the system are given by an expression, called a kernel function, that implements the concept of distance decay (or the first law of geography): other things being equal, events closer in

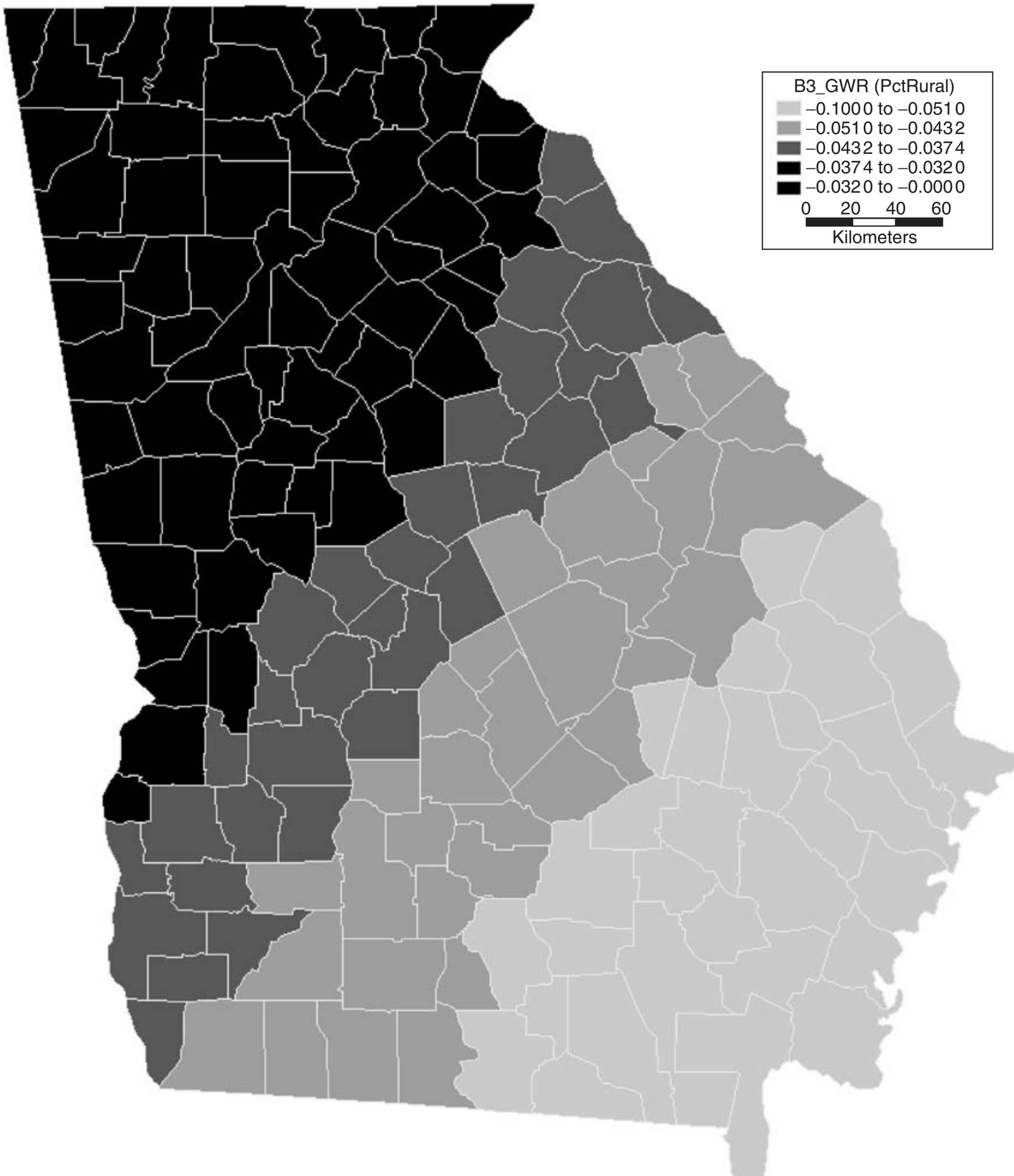

Figure 1 Example of a spatially varying coefficient estimated using the GWR approach.

space are more related than other events. Essentially, the kernel function places a window over the observations, thus selecting a subsample of points. The window can subsequently be displaced for local reestimation of the model. A number of different kernel functions have been proposed and used in practice. In perhaps the simplest case, a binary weighting scheme could be used as follows:

$$w_{ij} = \begin{cases} 1, & \text{if } d_{ij} \leq d^* \\ 0, & \text{otherwise} \end{cases} \qquad [3]$$

where d_{ij} is the distance between observations i and j, and d^* is a cutoff distance that defines the size of the window. Despite its simplicity, this kernel function has not been widely used. Instead, most applications of GWR have favored continuous functions that produce monotonically decreasing weights, such as the negative exponential:

$$w_{ij} = \exp\left(\frac{d_{ij}^2}{2\gamma^2} \right) \qquad [4]$$

or the bi square kernel:

$$w_{ij} = \left(1 \quad \frac{d_{ij}^2}{\gamma^2}\right)^2 \qquad [5]$$

In these formulations, instead of a binary scheme, the weight attached to each observation j with respect to i varies as a function of the intervening distance d_{io} be tween the two locations mediated by γ, a parameter that controls the rate of change of the function, that is, the size of the window. Modified kernel functions have been proposed that adapt to the density of observations within a region by using ranks instead of distance:

$$w_{ij} = \exp\left(\quad \frac{R_{ij}}{\gamma}\right) \qquad [6]$$

with R_{ij} as the rank of j when observations are sorted by increasing distance from focal point i. Alternatively, it is also possible to define the kernel function in terms of nearest neighbors to location i:

$$w_{ij} = \begin{cases} \left[1 \quad \left(d_{ij}/d\right)^2\right]^2, & \text{if } j \text{ is one of the } N\text{th nearest neighbours of } i \\ 0, & \text{otherwise} \end{cases} \qquad [7]$$

where d is the distance to the Nth nearest neighbor.

Once a kernel function has been chosen, the GWR local regression coefficients are estimated using the fol lowing expression (in matrix form):

$$\boldsymbol{\beta}_i = \left(\mathrm{X}^{\mathrm{T}}\mathrm{W}_i\mathrm{X}\right)^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{W}_i Y \qquad [8]$$

where $\mathbf{W}_i$ is a diagonal matrix of size $(n \times n)$, specific to location i, with the ith element in the diagonal (w_{ij}) given by the value of the kernel function for observation pair i and j.

A point to note is that the estimators in eqn [8] are a function of known data (the x and y observations), as well as the weight matrix, which depends on known distance between observations, but also an unknown parameter that defines the size of the window. In general, it is ac cepted that, in terms of estimation, selection of a func tional form for the kernel is less critical than selecting a window size. Therefore, there is a need for criteria to guide the determination of critical distance d^* (in the case of eqn [3]), the value of γ (in eqns [4]–[6]), or the number of nearest neighbors (in the case of eqn [7]) in specific applications.

Cross-Validation

Two general methods (direct assignment of window size or validation of window size) have been proposed for determining an appropriate window size for estimation.

In direct assignment of window size, the analyst decides the percentage of all observations in the sample to include in each local regression, or alternatively a cutoff distance may be adopted based on practical or theoretical reasons (e.g., flight range of mosquitoes). More commonly, the size of the window is validated through procedures that involve a measure of goodness of fit of the local regressions to the data. The following cross validation procedure remains the most widely used calibration technique:

$$\min_{\Delta}\ \mathrm{CVS}(\Delta) = \sum_i \left[y_i \quad \hat{y}_{-i}(\Delta)\right]^2 \qquad [9]$$

CVS is a cross validation score, in essence a sum of squared errors between the observed value of y at i, and the predicted value of y at i after removing observation i for estimation ($\hat{y}\ \ _i$). The latter value is a function of window size Δ, which depends on either cutoff distance d^*, par ameter γ, or number of nearest neighbors. Simply put, cross validation attempts to find the size of the window that gives the best correspondence (minimizes the differ ence) between predicted and observed values of y. The leave one out protocol (removing observation i in cross validation) is followed to impede the local estimators from wrapping around a single data point, that is, estimating one data point perfectly. Once the window size has been val idated local coefficients can be estimated using eqn [8].

More recently, modifications of the Akaike infor mation criterion (AIC) have been proposed as a window size validation tool. Despite the potential advantages of adopting an information theoretical approach for valid ation, this alternative remains underutilized. It is im portant to note that neither the use of CVSs nor AIC have been systematically investigated in the geographical literature from the perspective of their properties and behavior. As a consequence, it remains unclear whether, and if so, under what conditions, one of these validation procedures should be preferred.

Extensions

In addition to the linear regression framework, the idea of applying geographical weights has been extended to a variety of other techniques, to produce geographically weighted descriptive statistics, discriminant analysis, logistic regression, Poisson regression, and geographi cally weighted probit models. With regards to model specification and estimation, mixed GWR models have been proposed that combine spatially stationary and nonstationary coefficients, while more recent research has introduced local linear expansion of the coefficients.

Issues

A number of issues regarding the development and use of GWR merit mention. The first concerns the topic of inference regarding the local regression coefficients. One approach to developing a testing framework has been to adopt the nonparametric formulation of the method, and the allied concepts, developed mainly in the statistics literature, of effective number of observations and effective number of coefficients. These concepts are analogies of the global regression counterparts, and provide the basis for obtaining confidence intervals and other statistics. A second approach to develop tests has been to see the method as an ensemble of local regression models. This gives rise to a different issue whenever the number of local models to be estimated exceeds one ($m > 1$), since m hypotheses need to be tested simultaneously. The problem is that, as the number of tests increases, there is an increasing probability that some will turn up significant, even if only by chance. There is thus a need to give protection to the group of statements through group error rates as opposed to individual error rates – in other words, the individual level of significance needs to be adjusted for multiple comparisons. Several procedures have been proposed to provide the kind of protection needed when conducting simultaneous inference. A simple procedure is based on the Bonferroni probability inequality, a simple procedure that divides the overall significance level α (group error tolerance) by the number of tests m to obtain an adjusted significance level α/m (individual hypothesis error tolerance). Although this fast and rough procedure does not require distributional assumptions and is simple to use, the resulting tests lack power and are overconservative when the tests are highly correlated (i.e., not independent), which is certainly the case in GWR due to the use of overlapping subsets of data. Other methods exist that improve the power of the tests by introducing multiple step rejective schemes that adjust the size of the penalty to the level of significance in a sequential way.

A second issue with GWR concerns the high levels of spatial variability encountered by practitioners, a situation that arouses concerns that results may be misleading if not interpreted with caution. One source of regression coefficient variability in GWR comes from collinearity, or correlation in the model explanatory variables. Collinearity has been found in empirical work to be an issue in GWR models at the local level even in cases when it is not an issue in the traditional, or global, linear regression model with the same data. In addition to high variability of estimated regression coefficients, there can be strong covariation in the GWR coefficients for different parameters (covariate effects or the intercept), at least partly due to collinearity. The inflated regression coefficient variability associated with local collinearity can lead to coefficient sign reversals and overestimates of effect magnitudes, both of which could lead to incorrect interpretations of regression relationships. Collinearity diagnostic tools have been developed for the GWR framework that can detect when and where there are high levels of regression coefficient correlations due to collinearity. Estimated regression coefficients from calibration locations that are diagnosed as problematic should be interpreted with caution and additional analysis may be required in these areas to understand the nature of the relationships under study.

A more practical issue with GWR is with computational runtime of implementations of the technique. The estimation of GWR model parameters via cross validation can be computationally intensive for relatively modest datasets of thousands of model calibration locations and several covariates. However, the nature of the GWR estimation algorithm lends itself well to parallel computing solutions, where different computers can perform calculations for model predictions at different locations and the answers can be combined to evaluate the objective function for a given window size. Conveniently, there are multiple implementations of the GWR method. While not free, GWR 3.0 is available from Fotheringham and his coinvestigators. A free, open source implementation of GWR written by Roger Bivand is available as an R package. In addition, the basic elements of the GWR technique are relatively easily programmed in other programming languages and software, such as Matlab, as in James P. LeSage's spatial econometrics toolbox.

Debates

While GWR has offered the potential of understanding relationships between variables over space, there have been several concerns voiced about the methodology on both theoretical and practical bases. One concern says that GWR is not a unified, formal statistical model that provides a clear cut approach for inference on relationships. GWR is instead a collection of local spatial regressions where the dependence between regression coefficients at different data locations is not specified in the model, which results in a fixed effects model with no pooling in estimates. Some statisticians prefer the unified framework of hierarchical Bayesian models. In addition, the reuse of data to estimate regression coefficients at different locations creates a multiple comparisons situation that invalidates the usual confidence intervals for the regression coefficients. The debates surrounding development and use of GWR have been complicated by the fact that models with GWR style varying coefficients can be derived parting from a weighted regression framework, as a nonparametric model, or as a model with nonconstant variance. While all these alternatives produce exactly the same expressions for the estimators, they differ conceptually and technically in how and to what extent inference can be conducted. This feature of the regression coefficients, known as equifinality, has made it difficult for some researchers to agree on the properties of the model.

In relation to the above point, one issue of concern surrounding GWR for some researchers is the matter of degrees of freedom, which is directly related to model complexity. Although the concept of effective number of degrees of freedom is an accepted one in statistical analysis (e.g., for nearest neighbor and regularization methods, among others), and the concept has been adopted in some presentations of GWR, some researchers remain uncomfortable with the issue of degrees of freedom of the residuals in GWR, and the potential results from sharing data through spatial weighting to estimate regression coefficients at different locations.

Yet another debate about GWR is in the nature of the application of the technique itself. It has been suggested, for example, that GWR is well suited for estimation and prediction of the response variable but is perhaps less useful for the explanation of, or formal statistical inference on, potentially spatially varying regression effects. Given the roots of GWR in local linear regression, which was developed to estimate a response variable locally, this debate is not entirely surprising. When interpolation of a response variable over space is of interest, estimation issues such as collinearity are no longer a major concern.

Alternatives

GWR has highlighted the existence of potentially complex, nonstationary, relational patterns in geographical modeling. On the other hand, the approach is not unique in its ability to model spatially varying relationships, as there are some earlier and also later developments that accomplish the same objective, although in different ways, both technically and conceptually.

An early example of a local form of spatial analysis is the expansion method, in which the coefficients of a model are allowed to vary as functions of analyst defined expansion variables, typically using the coordinates of the units of analysis. The expansion variables are introduced using polynomial formulations (linear, quadratic, etc.), such that the expanded coefficients incorporate contextual variation across locations to produce a set of local, mappable results useful to detect spatial drift. The expansion method is a simple and intuitive way of building models, but is otherwise limited in that expansions are deterministic and, depending on the form of the expansion, might arguably miss more complex spatial variations. While GWR has been called a natural evolution of the expansion method, the two techniques are in fact very different in how they approach the problem of spatial coefficient nonstationarity, with GWR essentially being a moving windows approach and the expansion method a form of analysis of covariance.

A second development in local spatial analysis that predates the popularization of GWR in geography is the use of hierarchical models. This is an analysis technique similar to the expansion method that allows coefficients to vary by location, while introducing random variation of the coefficients. The method is attractive because models can be simultaneously defined at several levels that explicitly incorporate geographical and other (e.g., social) contexts into the analysis. In multilevel models, a type of hierarchical model, regions (or classes) at each level must be defined, so that individual observations can be 'nested' according to some type of categorization (e.g., geographical zones, age ranges, type of job, etc.). In addition to the usual concerns with modifiable areal units, a consequence of the multilevel model approach is that representation of geographical and attribute space becomes discontinuous or 'fragmented': coefficients are local in the sense that they vary between regions or classes, but are global (i.e., constant) within a given region or class.

Related to the hierarchical model is the hierarchical Bayesian model, in which information about relationships can be incorporated into the model through prior distributions of the parameters. If uninformative (or vague) priors are used for the model parameters and the regression coefficients are specified as independent in the model, the estimated regression coefficients derived from samples from the joint parameter posterior distribution will be, in theory, the same as in the non Bayesian hierarchical model. There are hierarchical Bayesian models with random effects for both the intercept and slopes for covariates, where the random effects may be independent *a priori* and borrow strength across observations globally or defined to have spatial structure and borrow strength locally. The primary alternative within this class of models is the Bayesian spatially varying coefficient model, which defines spatial structure in the regression coefficients through a prior conditional specification of the coefficients that uses neighboring observations only or through a prior joint specification of the coefficients that models the coefficients as a spatial process (SVCP). In terms of the way the dependence in coefficients is modeled, the conditional specification of the Bayesian model is similar to a GWR model with a nearest neighbor kernel with a small bandwidth of adjacent neighbors and the joint specification of the Bayesian model is similar to a GWR model with a distance decay, or geostatistical, kernel function. In contrast to the fitting of spatially local regression models in GWR, the Bayesian hierarchical model is a formal, unified, statistical hierarchical model where the coefficients are modeled as a multivariate spatial process and the correlation between regression coefficients is explicitly modeled. Simulation based comparisons of the accuracy of estimated regression coefficients from both GWR and the Bayesian SVCP model suggest that the Bayesian model may be preferable to GWR for inference on the effects under study, in terms of the proportion of the estimated coefficient confidence intervals that contain the true values. One drawback of the Bayesian model is increased computation time and more difficult implementation for inference. However, more general purpose software for Bayesian inference is becoming available through open source packages such as R and WinBUGS.

Concluding Remarks

The main professed benefit of GWR in geography and other disciplines that have adopted this technique is the ability of the method to investigate spatially varying processes through coefficient variability based on the geographical concept of distance decay. GWR provides an intuitive and technically accessible approach to this end. Computationally, estimation of the coefficients can be intensive, but the problem is highly solvable using parallel computing solutions that are becoming more widely available as mainstream technology.

Historically, GWR is based on the transposition of attribute space used in the method of locally weighted regression in statistics for curve fitting, to geographical space in a locally weighted regression used for modeling potentially spatially varying relationships. While the number of applications of GWR continues to grow, there have been relatively few attempts to assess the extent to which the transposition has been successful in providing a tool capable of retrieving spatial processes in a technically accessible, computationally economic, and ultimately accurate fashion. The evidence available in dicates that GWR is a good tool for estimating and predicting a spatially varying dependent variable, within a technical framework that is less demanding than, for example, variographic analysis and the use of kriging equations. Based on this it may be concluded that GWR is a useful tool for spatial interpolation, the analog of curve fitting in statistics.

With regards to the estimation of potentially spatially varying effects, the evidence indicates that GWR is capable of detecting, and to some extent retrieving underlying coefficient surfaces. GWR provides estima tion of spatially varying coefficients, but not a widely agreed upon inferential framework. As such, it seems prudent at this point to approach the use of GWR for inferential purposes with some caution. While the evi dence available indicates that the approach is useful for detecting nonstationary coefficient patterns, the degree of accuracy may not be extremely high, and therefore its use for drawing conclusions regarding substantive spatial processes should whenever possible be supported with other forms of analysis, such as the expansion method, and/or multilevel or hierarchical Bayesian models. This view of the method would align it more closely with other tools used for exploratory data analysis. Further support provided by other techniques would help to in crease the level of confidence with respect to any con clusions drawn about the character, nature, and possible origins of spatially varying relationships between variables.

See also: First Law of Geography; Kriging and Variogram Models; Modifiable Areal Unit Problem; Quantitative Methodologies; Quantitative Revolution; Regression, Linear and Nonlinear; Spatial Autocorrelation; Spatial Expansion Method; Spatial Filtering/Kernel Density Estimation; Spatial Interpolation; Spatially Autoregressive Models; Statistics, Spatial.

Further Reading

Atkinson, P. M., German, S. E., Sear, D. A. and Clark, M. J. (2003). Exploring the relations between riverbank erosion and geomorphological controls using geographically weighted logistic regression. *Geographical Analysis* 35(1), 58 82.

Brunsdon, C., Fotheringham, A. S. and Charlton, M. E. (1996). Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis* 28(4), 281 298.

Cleveland, W. S. and Devlin, S. J. (1988). Locally weighted regression an approach to regression analysis by local fitting. *Journal of the American Statistical Association* 83(403), 596 610.

Fotheringham, A. S. and Brunsdon, C. (1999). Local forms of spatial analysis. *Geographical Analysis* 31(4), 340 358.

Fotheringham, A. S., Brunsdon, C. and Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: Wiley.

Goodchild, M. F. (2004). The validity and usefulness of laws in geographic information science and geography. *Annals of the Association of American Geographers* 94(2), 300 303.

LeSage, J. P. (2004). A family of geographically weighted regression models. In Anselin, L., Florax, R. J. G. M. & Rey, S. J. (eds.) *Advances in Spatial Econometrics*, pp 241 264. Berlin: Springer.

Leung, Y., Mei, C. L. and Zhang, W. X. (2000). Testing for spatial autocorrelation among the residuals of the geographically weighted regression. *Environment and Planning A* 32(5), 871 890.

McMillen, D. P. (1996). One hundred fifty years of land values in Chicago: A nonparametric approach. *Journal of Urban Economics* 40(1), 100 124.

Mei, C. L., He, S. Y. and Fang, K. T. (2004). A note on the mixed geographically weighted regression model. *Journal of Regional Science* 44(1), 143 157.

Páez, A. (2006). Exploring contextual variations in land use and transport analysis using a probit model with geographical weights. *Journal of Transport Geography* 14(3), 167 176.

Páez, A., Long, F. and Farber, S. (2008). Moving window approaches for hedonic price estimation: An empirical comparison of four techniques. *Urban Studies* 45(8), 1565 1581.

Páez, A., Uchida, T. and Miyamoto, K. (2002). A general framework for estimation and inference of geographically weighted regression models: 1. Location specific kernel bandwidths and a test for locational heterogeneity. *Environment and Planning A* 34(4), 733 754.

Wheeler, D. and Tiefelsdorf, M. (2005). Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems* 7(2), 161 187.
Wheeler, D. C. and Calder, C. A. (2007). An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. *Journal of Geographical Systems* 9(2), 145 166.

Relevant Websites

http://www.mrc bsu.cam.ac.uk/bugs/winbugs/contents.shtml
WinBUGS website.

Geography, History of

A. Maddrell, University of the West of England, Bristol, UK

What Is the History of Geography and Why Do It?

The history of geography is more than a list of dates of the foundation of geographical societies, or even a recitation of key texts and ideas. The two main elements of a history of geography are to do with the institutional frameworks of the discipline (e.g., geographical societies, university departments, state legislation on education) and the geographical ideas and practices developed within (and beyond) those institutions. Histories of the discipline which were dominated by institutional frameworks have been criticized for being too inward looking, giving insufficient attention to the development of geographical ideas; and those dominated by geographical ideas and their associated philosophies and methodologies have sometimes failed to recognize the role played by those institutions in shaping ideas and practices. The emphasis here will be on the development of ideas and practices in the history of geography, but with reference to those institutional frameworks where relevant, as well as recognizing that the subject has been influenced by ideas and practices from beyond what might be recognized as disciplinary boundaries – and that forms of geographical knowledge and practice exist beyond what might be formally defined as geography.

Studying the history of geography illustrates ways in which geographical knowledges have been accrued, validated, challenged; how they have been shaped by national or regional scholarly traditions and manipulated to fit the ideological needs of other agents such as the state or commerce. While there are those who suggest the only way to truly understand the present discipline is to look to socioeconomic, political, and cultural forces at work today, others believe understanding past processes help today's geographers to understand not only how the subject was shaped and used historically, but also sheds light on present forms of geographical knowledge and practice, with all their strengths and weaknesses.

How Do We Write a History of Geography?

Histories of geography appeared early in the life of academic geography as part of the search for disciplinary identity and status when the subject was professionalized within the international university sector in the nineteenth and early twentieth century. New accounts of the history of geographical thought and practice appear at regular intervals, often focusing on a particular country's tradition. Historians of geographical inquiry employ the insights of historiography to analyze and evaluate different histories of geography, and the approach taken to the history of geography reflects the way(s) in which geography as a subject is seen, interpreted, and represented. One of the key analytical tools is to consider the extent to which a particular history might be considered as internalist or externalist. Internalist accounts are those which look within the discipline, to its personnel, its theories and methods in order to explain geographical thought at a particular time and place, rather than placing it in a wider context. These histories of geography have often focused on leading individuals and their followers (typically represented in gendered terms as founding fathers), schools of thought, and institutions. These celebratory accounts often result in a hagiographic rendering of the subject and people, resulting in what has been called the 'textbook chronicles' of the subject. They have also been categorized as 'whiggish' history, creating accounts which give the impression of an ongoing natural progression in geographical knowledge.

Nineteenth and early twentieth century accounts of the history of discipline, often emphasized geography's lineage to the work of European classical scholars and/or focused on the history of exploration; others celebrated anniversaries of foundations (notably of geographical societies). However, these accounts have been criticized for being too internalist and more 'critical' approaches to historiography were called for in the 1970s and 1980s, particularly for histories of what happened within geography to be placed in the context of wider external factors. This externalist approach, as the name implies, stresses the explanatory role of external factors framing the practice of geographical inquiry at a given time and place. These external factors include: ideas from beyond academia, including natural science, religion, politics, and circulating ideologies; they also include what can be classified as socioeconomic and cultural factors, in both their materialist expression such as through war or capitalist processes, and the impact of related discourses such as prevailing views on class, race, gender, for example, on geographical inquiry. In recent years this externalist contextual approach has dominated work on the history of geography and in turn has resulted in what are described as revisionist histories, which use novel data and/or perspectives to shed new light on the nature and practice of geography, revising previous received histories. This includes more 'inclusionary' approaches, which consciously set out to incorporate a wider scope for the definition of geography (such as vernacular or children's geographies) or individuals or groups previously neglected or omitted from geography's histories (such as colonial subjects, women, or disabled people). Clearly there are many potential alternative perspectives on the history of geography, and in turn this gives sense of the rich ground which constitutes the history of geography.

Skinner's important 1969 paper 'Meaning and understanding' in the history of ideas has been particularly influential within interdisciplinary historiography, including that of geography. Within geography, attention was particularly drawn to the relationship between text and context, discipline and wider context: including disciplinary, institutional, political, and economic cycles. In recognizing the influence of wider contexts in the way in which a body of knowledge is shaped, we can see how a discipline – in this case geography – reflects current concerns and agendas at a given time and place. This is known as the social construction of knowledge, which allows for knowledge being ideologically shaped and open to varied interpretation. While few scholars would see a subject as a pure social construct, most recognize that disciplines do not develop in isolation and it is important to acknowledge and examine these wider influences. Related to this was a recognition of the influence of the interests of a given historian in constructing a particular history, whether by choice of subject matter, theoretical and analytical framework, or methodological scope. Many scholars, especially those working from a feminist and/or post structuralist perspective believe this recognition of the author's own positionality in relation to their subject is more honest than claims to neutral objectivity. This awareness has given rise to what is referred to as 'situated knowledge' and can be applied to all forms of geographical inquiry including historiography. Inevitably this means there will be different views of what constitutes geography *per se* and how it developed and changed, and these will vary over time and space and between different practitioners: that is, the history of the subject is contested (this is a theme which will be returned to below).

An important way in which historians of the discipline differ is in identifying the impetus for change within geographical inquiry. Those working from an internalist model tend to see changes emanating from forces within the discipline or wider academia, tracing the exchange of ideas and methods through publications, field work, conferences, and other intellectual networks. Thomas Kuhn's *The Structure of Scientific Revolutions* brought an alternative explanation for change within academia. Looking at academics from a sociological perspective, he saw a competitive drive for status driving change within scientific communities. This competition caused significant shifts between different frameworks or paradigms which successively defined the appropriate subject matter and methodologies for enquiry. The dramatic nature of these shifts in practice caused them to be described as revolutions and the idea of revolutionary relative to evolutionary change dominated historical surveys of geographical inquiry in the 1970s and 1980s. In contrast, externalist accounts looked to the wider socioeconomic and cultural milieu to explain changes within the discipline, for example, the impact of war, social planning or existentialism on geographical knowledge and the status of the discipline. One strand within these externalist approaches was a recognition of the social construction of knowledge (see below) which challenged the view of objective knowledge or the possibility of a single history of the discipline.

Contextual and Discursive Approaches to the History of Geography

The contextual approach has increasingly been normalized within the history of geography, replacing whiggish and exclusionary accounts of the development of the subject; the role of ideas and practices external to academic geography in shaping geographical thought being identified as the antidote to previous shortcomings. David Livingstone's work *The Geographical Tradition* has been pivotal in raising questions about social context, metaphysical assumptions, professional aspirations, and ideological allegiances, asking questions such as: what role did geography play in past society? Was geography used by particular groups for political, religious, or economic purposes? Who benefited and who lost out by the introduction of new theory? Why were particular theories generated, welcomed, or outlawed? Anne Godlewska similarly argues that conceptual history rests on asking basic questions about individuals' lives and work, such as: what were their key ideas? Which were the decisive influences shaping not only their ideas but the method and presentational form of their work? What was, and what has been, the impact of their ideas and methods?

The merits of the contextual approach to the history of geography have been expounded widely on numerous grounds. First, as the name implies it allows historical events, movements, and changes to be placed within wider socioeconomic, political, and cultural contexts – for geographical knowledge at any given time to be recognized as 'situated knowledge' even if it is complicated. Second, it avoids a sense of linear progression and 'good' and 'bad' individual geographers. Third, taking a contextual approach allows for the continuity of some ideas while new systems of thought emerge in parallel; it also avoids an artificial dichotomy between internal and external factors of scientific change as does the discursive approach, which identifies sets of ideas forming and jostling for influence over the way a given society thinks and acts (including academic communities). However, even proponents recognize that there are limitations to the contextual approach which can be presented as a cure all for historiographical work. Contextual historians can be overly influenced by current agenda and fall into the trap of presentism – judging preceding scholars and society by present criteria. Using a contextual approach also raises the question of 'which context?' For example, it has been argued both that there is much to be learned about geography by looking beyond the academy, and that too little is known about geography's place within the academy. Thus, there have been recent calls to study 'minor' as well as 'major' figures, small scale as well as larger schema, to consider ordinary or vernacular his tories, and to look for hidden, obscured, or omitted his tories. Also internalist institutional accounts of geography departments and societies could become obscured in the face of externalist history, at best relatively neglected and at worst dismissed as introspective records of little interest to anyone beyond those associated with the institution.

However, recent work on departmental archives has shed significant light on localized geographical practices and how these relate to geography in national and international terms. These internalist accounts can also be rich data sources on individuals and practices omitted from more wide ranging accounts and should not be dismissed sight unseen.

The frequent characterization of histories as either 'internal, cognitive history' or 'external, contextual his tory', has also been challenged and an acknowledgment of the creation of geographical knowledge through dis course allows an alternative to this conventional dichot omy, thereby avoiding a caricature of knowledge as 'pure' or 'corrupt', when in reality all knowledge has been subject to external and internal influences. The discursive approach allows abridge between different explanatory approaches. If knowledge is discursively constructed then an understanding of institutions and practices is vital to the historical project. Discourses work in social contexts with material consequences: this is both the site and content of 'new' self critical discursive historiography of geography.

The production and reception of geographical and other texts is crucial to understanding the development of a discipline. Debates within contextual history and literary theory concerning the relationship between texts and contexts, knowledge and power, authorial intention and alternative readings can make surveys of geo graphical literature complex but nonetheless rewarding. The French philosopher Michel Foucault's discussion of power and knowledge has been central to much of the debate on the significance of texts, particularly his central argument about the reciprocal nature of these two: power produces knowledge and knowledge presupposes and constitutes the relations of power in a transformative process. Largely influenced by this Foucauldian view, it is widely accepted that texts are not mere reflectors of the material world but are relations of power in themselves. Foucault argues the question (and therefore the strategy) is to consider what it means for a given statement or body of knowledge to appear when and where it did, and we can ask this question about geography as a whole or a given geographical text: what is its location in the cultural complexity of the specific moment of its historical pro duction? However, just as it is important to think about how a text or subject is produced, we also need to think about its reception – how is it received by different audiences in different places and times? Reading differ ent reviews of books can reveal a lot about the different ways in which geographical knowledges and practices varied – as well as the relative merits of which are in cluded or excluded in disciplinary histories.

In the 1990s the postmodern suspicion of transcen dental themes and an awareness of knowledge as a situ ated social practice raised questions regarding the mechanistic interpretation of the relationship between text and context, including geographical texts and related contexts. Geographical knowledge is seen not merely as the data and theory that is contained under the heading of 'geography', but also as a discursive formation. This approach is shared by those who affiliate themselves to, or make use of, the post structuralist work of Derrida and Foucault as a powerful analytical tool because of their recognition that cultural concepts lack transparent and shared meanings. The recognition of the instability and contestation of meanings facilitate and demand an examination of the politics behind the conflictual pro cesses that establish those meanings. The creation of the 'Other' is of particular importance when studying con structions of those beyond the dominant group, as shown in the case of race and gender, as well as other axes of difference such as disability or sexuality. It is therefore crucial to understanding the representation of people and places in geography generally and in geographical texts in particular. Textual means for creating the 'Other' in clude the type of narrative voice adopted, the type of structure the author builds, and the kind of images and motifs circulated in the text. Edward Said's *Orientalism* played a crucial role in understanding how the 'Other' is constructed. Recent critiques of the history of geography have pointed to the marginalization or omission of some groups within the historiography and current practices of geography – including those using an externalist con textual approach. What is omitted from histories can be just as revealing as what is included. For example, many histories of geography give the impression that pre 1970s geography was a male preserve. This could indeed be the impression given by the historical canon of geography, however, as Jan Monk and Mona Domosh show, a more complex underlying picture emerges with more detailed (archaeological in Foucault's terms) analysis of publications, obituaries, institutional, and departmental records. This leads us to a consideration of sources used in constructing histories of the discipline and how these relate to representing difference within the discipline.

Historical Sources

Historical sources are generally divided into two categories: primary and secondary. Primary sources include archives, such as personal and professional papers, letters, institutional records, manuscripts, original published texts, and oral histories. Secondary sources include existing histories, related literature, reviews, obituaries, etc. This immediately raises issues about the number and quality of sources available to any historian of the subject, accessibility of those sources and how those sources relate to one another. Archives, for example, can be experienced as sites of partiality, control, and regulation of knowledge, they may have limited records on a given group such as postcolonial subjects; but for other groups or individuals they can also be potential sites of access and liberation of knowledge. Changes in technologies and cultural attitudes to access to information have combined to increase the accessibility of many archives. Many archive catalogs are available online and a growing number of entire archive collections can be accessed and downloaded from the Internet; other collections, including national archives, increasingly offer customized remote ordering systems for electronic copies of documents (for a fee). Increased public access to private archive collections has also been a condition of state funding for updating archive collections (as seen at the Royal Geographical Society – with the Institute of British Geographers (RGS–IBG)).

Traditionally, the history of geography has relied upon textual sources, but reflecting developments in contemporary history, oral accounts, including autobiographical interviews, are increasingly being sought and integrated into disciplinary histories. This approach was notably undertaken in the 10 year Dialogue Project in Sweden, headed by geographers Anne Buttimer and Torsten Hägerstrand during the 1970s and 1980s. This multidisciplinary project, explored autobiographical interviews as a means of promoting understanding and integration of knowledge, interpreting personal accounts in terms of the three headings of meaning, metaphor, and milieu. Oral histories or autobiographical accounts may be considered to lack objectivity, but their very subjectivity can be a rich source of information. They allow an individual to communicate in a more informal way – an alternative voice – compared to written work and provide personal and institutional contexts (such as personal ideologies, intellectual inspirations, and career rivalries) which may not be apparent in print form, making for a richer and more nuanced set of sources to be combined with textual material. Oral accounts can be of particular value in relation to geographers working before publications were considered as the main academic purpose and for whom few textual records exist of their work and lives. While limited to what might be described as recent history, the value of this approach can be seen in initiatives in various countries to collect autobiographical interviews and oral histories relating to geographers, their geographical practice, and careers (e.g., the American Association of Geographers' interview archive).

The combination of a greater variety of sources, perspectives on, and questions about geography in the past, has resulted in more varied approaches to the history of the discipline. A core element is the recognition of its multiple strands rather than an accepted account or canon: there are histories rather than a single history of geography. Key figures and texts continue to be recognized, but there is a wider appreciation that much can also be gleaned from less well remembered sources and biographies; universities and geographical societies are important windows to geography's past practice, but so too are schools, teacher training colleges, local field study groups, ramblers' associations, and everyday sources such as personal diaries. Turning to these more varied historical sources in conjunction with appropriate historiographical methods sheds light on a more varied body of those we might describe as geographers and a richer variety in the construction and practice of geographical knowledge. This does not make for simple histories (it is for good reason that David Livingstone described his theoretical and methodological approach to the history of geography as 'situated messiness'), it makes for more complex but arguably more interesting histories of geography.

Trends and Contestation in Geographical Thought and Practice

While maps and representations of place can be found in prehistoric cave and rock paintings and vernacular geographies have always been a part of everyday lives, written accounts of geographical knowledge can be found ranging from the classical writings of Herodotus, Strabo, Ptolemy, and Varenius; to the Norse sagas of Eric the Red and Leif; and the writings of Arab scholars such as al Maqdisi, Ibn Batuta, and Ibn Khaldun, with instruments such as the compass and concepts such as decimalization circulating from China and the Hindu world. European documentation of what was known as the 'New World' had been growing since the sixteenth century, largely as a result of seafaring trade and colonization on the part of the Spanish, Portuguese, Dutch, and British. With a rich variety of sources, it is not surprising that geography had been part of European university curricula for hundreds of years, notably within the disciplines of classics and history, however, geography only emerged as a modern independent discipline in the nineteenth and early twentieth century, spreading from the geographical societies founded across Europe and North America into universities and teacher training institutions and schools. Although often represented as of secondary importance, geographical societies not only provided funding for expeditions and a public forum for geographical reports and debate, many also lobbied and provided strong support for the establishment of geography in universities, especially in France and Britain.

Filling in and Laying Claim to the Blanks on the Map

Increased demand for geographical knowledge in the nineteenth century was crucial to the discipline's establishment and popularity. Geographical societies were founded across Europe throughout the nineteenth century: starting with Paris (1821) and Berlin (1828), London (1830), St Petersburg (1845), numbers growing to 16 by 1865, 94 societies worldwide by 1885; 80 of these were in Europe (over half in France and Germany) but those beyond Europe represented the global spread, for example, the American Geographical and Statistical Society (1851), the Geographical Society of Quebec (1877), the Tokyo Geographical Society (1879), the Swedish Society for Anthropology and Geography (1887), and the International Geographical Congress which met for the first time in London in 1895. While the number of societies continued to grow between 1890 and 1920 (including the Association of American Geographers (1904)), the pace of this expansion was slower and the number of societies declined after 1930 (several societies closing during the 1914–18 and 1939–45 wars). In the UK the foundation of the RGS could be seen in the context of the general trend of the foundation of learned literary, scientific, and philosophical societies in the early nineteenth century, but by the 1880s there was a clear sense of the national need to catch up in relation to the geographical knowledge of competing European imperial powers, especially with the Berlin Conference taking place in 1884 when European powers negotiated territorial claims and spheres of influence in Africa. Consequently, the Royal Scottish (1884), Manchester (1884), Tyneside (1887), Liverpool (1891), Southampton (1897), and Hull (1910) geographical societies were founded around Britain in the last two decades of the nineteenth century. This reflected more than a simple thirst for knowledge about the world and scientifically motivated exploration, it also reflected popular zeal for the empire and related political and economic interests. Thus it was not just about the 'heroic' search for geographical knowledge and the filling in the blank areas on maps, but in many cases was about laying political and economic claim to those areas as well: the disciplinary drive for intrinsic geographical knowledge and the external impetus of national geopolitical and commercial interests.

These varied interests were often seen as complementary and often presented as such by geography's relatively young interest groups lobbying governments and universities for recognition. These overlapping interests could also be compressed in the geographical inquiry of one individual, such as British explorer David Livingstone, who was a representative of the British government as a commercial envoy, a Christian missionary society, and two geographical societies, as he journeyed in central Africa. In common with the RGS, the nonmetropolitan societies all attested to their ability to serve the practical needs of numerous different interest groups in their founding documents. However, two other geographical societies were founded in Britain which were oriented to teaching and researching geography: the Geographical Association (1893), principally concerned with teaching geography in schools, and, some 40 years later, the Institute of British Geographers was founded in 1933 as a consciously academic forum for British geography.

The relationship between the emerging discipline of geography and empire has been widely researched and documented, especially in the 1990s. These studies unpicked the ideological, geopolitical, and materialist motives which underpinned many of the popular and 'scientific' drives for geographical knowledge beyond Europe. Interests at play included the globalizing structures of capitalist trade, state territorial expansion and colonization, population management and emigration, racialized control, and cultural assimilation. Indeed, the interests of empire continued to influence the content of European geography into the 1950s and arguably persist in many nations' framing of geography in state school curricula. In turn, geographers from Europe and North America also influenced the diffusion of geography within the colonial network in the first half of the twentieth century, with a significant number teaching abroad and publishing related regional studies. However, it should be noted that this synergy between geographical knowledge and empire applies not only to Europe, but also to periods of expansionism in the Arab world in the Middle Ages, the USA in the eighteenth and nineteenth centuries and Russia/USSR in the twentieth century.

It should also be noted that some geographers actively resisted being complicit with the imperial discourse in their work and alternative discourses emerged within nineteenth century geography. Although not widespread, anarchism had some influence on nineteenth century European geography and has some threads of continuity through to contemporary radical geographical practices. Anarchism eschews authority and takes various forms, and within geography is particularly associated with what has been described as the 'gentle anarchism' of Peter Kropotkin and Elisée Reclus. Kropotkin, born to an aristocratic family in Russia, found both his political views transformed and his interest in geography ignited while posted with the Cossacks in Siberia after which he returned to St Petersburg to work in the Ministry of Internal Affairs and the Imperial Geographical Society. Other nineteenth century Russian geographers had also expressed concern for the plight of the peasants such as Arsenyev and Semenov (Kropotkin worked for the latter at the Ministry) but Kropotkin's anarchist beliefs, writings on the abolition of the state, and support of political movements such as the Narodniks caused him to be imprisoned for 3 years in St Petersburg,. After his escape and exile, he resided principally in London, but was thrown out of Switzerland and imprisoned in France.

While his geographical work was largely in the sphere of physical geography, especially glaciation, two of his philosophical articles continue to have an impact on the discipline. 'Mutual aid. A factor in evolution', serialized in 1902, stressed cooperation rather than the competition propounded by social Darwinists, and his article on 'What geography ought to be' in which he stresses the importance of travel and education, are still cited today and probably represent his main linking of geographical and anarchist concerns. It is interesting to note that the RGS was among those petitioning for his release when imprisoned in France, and that both the Russian Imperial and the British Royal geographical societies approached Kropotkin to become a post holder, which undercuts their traditional uniformly conservative image (although Kropotkin declined these invitations on political grounds). Kropotkin went to Switzerland in the 1870s, where he met and was greatly influenced by French anarchist Elisée Reclus, whose political views were shaped by his travels in Ireland and America and participation in the Paris Commune of 1871, after which he was imprisoned and exiled. Before the Commune he had similarly combined political radicalism (in the First International and Bakunin's 'Brotherhood') and a key role in the Paris Geographical Society. Like Kropotkin, Reclus produced anarchist and geographical publications, his most famous geographical work being his 19 volume *Nouvelle Geographie Universelle* published between 1876 and 1894, in which he combined observations from his travels with secondary material, although his later 1905–08 volume *L'Homme et la Terre* articulated more of his beliefs about society, including a combination of the unique qualities of each area or region with a holistic approach to understanding the equilibrium of people and environment in a given region, as well as an emphasis on industrial and urban spaces. Not surprisingly he also made a critique of imperialism. Kropotkin returned to the new revolutionary Russia during his final few years and Reclus spent his last years in Brussels where he continued to write and was influential in the founding of the New University and the Geographical Institute there. Socialists, Marxists, and the Situationists have all drawn attention to the work of these early radical geographers, especially Kropotkin.

While geographical societies represent the institutionalization of the subject across Europe, it was the opportunity to gain accredited qualifications at university level which professionalized the discipline. The work of Carl Ritter (author of the 19 volume 1822–59 *Erdkunde*) who held the first chair in geography at the University of Berlin from 1820 and Alexander von Humboldt (author of the 5 volume 1845–62 *Cosmos*) have been identified as both the end of the period of universal or renaissance geographies and the rebirth of modern geography. Alternatively, the founding of geography as a discipline in nineteenth century Europe can be attributed to an alternative external social explanation which may account for the rapid establishment of geography in Europe in the second half of the nineteenth century.

Modern geography emerged as a university discipline as much in response to external goals of European society as the result of the growth of specialist geographical knowledge: namely, the demand for education of teachers after the popularization of elementary education; the demands of exploration and colonization and for public knowledge of the latter; and the role given to geography in fostering nationalism. It was the utility of the subject in an age of domestic unification, of war with France and wider geopolitical interests of empire, which inspired the Prussian/German government to found chairs in geography in the 1870s. The Prussian's victory in the Franco Prussian War 1870–71 was attributed to superior geographical knowledge and understanding on the part of the Prussian militia. This stimulated geographical study in France and other European countries, with British geography being much influenced by the publications in Germany and France. Thus, in terms of a contextual or discursive approach to the history of geography, the European subject was shaped very much by the interests of state and commerce – as much as by scientific inquiry and cartographic endeavor. Similarly, in Japan geographical knowledge was strategically useful in the Sino Japanese and Russo Japanese wars of the 1890s and 1900s, and in North America the subject was associated with state interest in expansion and consolidation of territories and national identity.

Geography developed unevenly between countries, which has been attributed in some cases to the slower development of state funded popular education; for example, Britain was slow to establish free popular education (1870) compared with France, Prussia/Germany, Switzerland, and Sweden (1842). By 1875 there were university chairs in Geography in Germany (7), France (7), and Switzerland (3) followed by St Petersburg in 1885 and Lund in 1897. After a short lived appointment at University College London in the 1830s, Britain saw its first permanent university appointment in geography in 1887 and first degrees offered from 1917. In Britain some universities resisted the establishment of geography as an academic discipline, disliking its association with empire and/or considered it lacking intellectual rigor; in France geography met opposition from other sciences and from those on the political left who held internationalist ideas. In Russia expansionism fueled the demand for geographical knowledge and maps from the rule of Peter the Great in the late seventeenth and early eighteenth century; this led to the founding of a geography department in the Russian Academy of Sciences in 1758, but it was almost 120 years before the first university professorship in geography was created.

In common with many other scientists in the early to mid nineteenth century, geographical authors often wrote about the world from a God centered or teleological point of view, as can be seen in the influential writings of Ritter in Germany and Mary Somerville in the UK. While the publication of Darwin's *Origin of Species*… is characterized as the lynchpin of the science versus religion debate in the nineteenth century, and there were philosophical purists on both sides of the debate, many of the issues had been circulating for years and many involved occupied the muddy mire of the middle ground. What can be said about geography is that its voice was increasingly secularized in the second half of the nineteenth century. At the same time geography was increasingly separated from physiography and geology as part of the specialization within natural sciences. It was this specialization coupled with the emergence of the social sciences at the end of the nineteenth century that prompted calls for a separate identity for geography in universities where the subject was not yet established. Britain was a case in point and geography's supporters had not only to define its purpose and intellectual credentials but also prove its utility, thereby justifying its status within the university sector. This view was strongly represented in Britain by Halford Mackinder in his 1887 address to the RGS 'On the scope and methods of geography', in which, reflecting earlier pronouncements, he argued that the new geography would serve the interests of the statesman, merchant, historian, scientist, and teacher. Perhaps in an attempt to emphasize the professional status of the discipline Mackinder did not want to refer to the traveler, novelist, journalist, or painter – and with the exception of 'the teacher' his list of interested parties represented a near exclusively masculine domain at the time of his writing. He also mentioned the regional approach to geography in his address, which would be central to defining the subject and practice of geography internationally in the first half of the twentieth century.

University geographers were of necessity drawn from other disciplines, many attracted by the opportunity for professional advancement in a new discipline, for example, Vidal de la Blache who moved from classics to geography in France and Eva Taylor who switched from chemistry and became the first woman to be appointed as professor of geography in the UK in 1930. Initially, geography was strongly grounded in a holistic approach which founded the understanding of people's relationship to place very much on the physical characteristics of that place: topography, geomorphology, climate, etc. This which led on to the regional approach to geography being the first methodological blueprint in Anglo American and European geography. At the turn of the nineteenth century 'political geography' was rejected as a heading for all nonphysical geography, 'anthropogeography' was adopted for a while and then replaced by 'human geography' in the 1900s. People–environment relationships and the region were the key foci of early twentieth century geography, both of which drew on and linked physical and human geographies.

Unpicking the Philosophy of Geography in the History of Geography

Philosophy at its most basic is the study of what we know and how we know it. Ontologies are theories of what exists, what is real in the world, sometimes referred to as metatheories; epistemologies are concerned with the theory of knowledge, what we know and how we can know it: what constitutes valid knowledge – which has obvious implications for the methodologies adopted. If we look at intellectual developments within the history of geography we can see where geographers have agreed on the ontologies and epistemologies applied to geographical enquiry and where these differed and were contested and how changes came about within geographical thought and practice, as well as identifying continuities and discontinuities in that geographical theory and practice. Two approaches to geography will be discussed in some detail in the light of different ways of interpreting the history of geography, including the shift from one dominant discourse to another; this will be followed by a brief analysis of the alternative approaches to geographical inquiry which have de veloped since 1970.

The Region

The region as a concept and methodology of geographical enquiry was derived principally from Ritter's work and served to unify the identity of the subject internationally – but it was known by different names and conceptualized in different ways both between and within countries. In Britain geographers studied 'regions', in France 'pays', in Germany 'landschaft', and in Russia 'zones', each having the common focus on studying particular places. In Germany the regional approach was characterized by Hettner's *Natural and Land schaft* and was methodologically based on the Landerkunde schema: geology, relief, climate, natural resources, and settlement. The cultural landscape or human imprint on a region and local study (Heimatkunde) was popular in schools. Friedrich Ratzel, seen as the founder of modern human geography, was strongly influenced by Darwinian and associated ideas, seeing a direct relationship between cultural forms and natural conditions. While this may have been based on a deterministic approach to geography, Ratzel was innovative in subjecting cultural phenomena to sys tematic study and his 1881 and 1891 two volume *Anthro pogeographie* demonstrate his belief in environmental influence, but also how he modified his views in the time between volumes one and two. Alfred Hettner, who re mained a champion of the regional chorological approach, similarly rejected simplistic determinism. In his 1897 *Political Geography* Ratzel represented states as organic entities which thrive or decline: Karl Haushofer built on Ratzel's ideas to rationalize German need for space (Lebensraum) and this was infamously used to justify the domination of weaker neighbors, especially when coupled with the 1930s German concept of Volk (folk) which linked blood and soil in terms of race. The Nazi state thereby extracted from geography and related disciplines a form of theoretical justification for aggressive expansionism. However, it has been pointed out that these and similar ideas also underpinned much of the earlier rationalization of wider European colonial expansion. In post war years regional geography continued to evolve in Germany including Carl Troll's landscape ecology approach which studied the relationship between biological com munities and environment in a specific delineated landscape. The breadth of Hettner's chorological geography was criti cized as too all encompassing methodologically, causing Otto Schluter to introduce the more focused approach of landscape morphology, but Hettner's approach was con tinued, notably championed by Richard Hartshorne who was to have international influence, especially in the USA.

Swedish geography was influenced by German thinking from the outset and Hembydsforskning (home area studies) were popular in Sweden's Folkhogskolor (people's or folk colleges). Helge Nelson promoted re gional studies at Lund from the foundation of the geography department in 1916 onwards, stressing links between home area and nation at this time of consoli dation of national identity. However, by the 1920s there was tension between those geographers favoring a focus on people's relationship to the environment and those who favored the chorological approach – analysis of spatial distributions. Ultimately it was the latter which dominated twentieth century geography in Sweden, not least because of the emphasis on applied geography and close links between geography and government planning. Rudolf Kjellen, a pupil of Ratzel's, developed geopolitics on the basis of 'survival of the fittest'. This association with determinism, expansion, and Nazi aggression (as noted above) caused political geography and particularly geopolitics to come into disrepute, but the Cold War caused a revival of interest in the subject (especially in America) and its more recent reorientation to critical political geographies, engaging with issues of environ ment, gender, and contemporary security issues have made for a thriving subdiscipline. A strand of post war Swedish geography also developed a strong thematic and ideological orientation to social equality and develop ment issues, reflecting the domestic social concern which grew out of the 1917–18 famine and building on the work of Alva and Gunnar Myrdal who had been politically radicalized by a visit to the USA in the late 1920s (an experience to be replicated by David Harvey in the early 1970s).

In France early twentieth century *le regionalisme* was based on detailed local studies of pays, resulting in small scale regional monographs and promoted by Jean Bruhnes and Vidal de la Blache. These studies were often based on substantial doctoral theses and Roger Dion's *Le Val de Loire* is considered one of the great French regional monographs. Vidal argued for the essential unity of a region, the inseparability of nature and culture which was interwoven in a complex localized expression as the re gion's *genre de vie*. While grounded in the physical land scape, Vidal, influenced by Ratzel's later work, and took a view of the physical environment providing the oppor tunity for a range of human responses, which people could choose between, what is known as possibilism (although Vidal never used the term himself), and that these particular sets of responses to particular environ ments gave rise to unique regions or pays. He favored a chorological approach as can be seen in his 1903 *Tableau de la Geographie de la France*. Given the regional political structure in France, there were opportunities for these regional monographs to be utilized by administrators and planners within France, but they were criticized for being too based on agriculture and assumed historical con tinuity of practice, which did not always apply, especially in rapidly industrialized areas. Vidal's 1917 *L'Oest*, which studied the industrialized East of France, acknowledged the role of wider economic influences on pays and credited as his most innovative work, was published in the year of his death. Vidal's model was notably de veloped by his student Lucien Gallois who demarcated the regions of France; his son in law, Emmanueal de Martonne, who completed Vidal's book *Human Geography* which was published posthumously; and Jean Brunhes, who is credited with simplifying and popularizing Vidal's work. Vidal influenced many geographers internationally, for example, Hilda Ormsby and Marion Newbigin in the UK and Jovan Cvijic at Belgrade, who linked a region's *genre de vie* with land productivity and economic activities.

In the UK Patrick Geddes, deeply influenced by Frederic Le Play, argued that all education had to be based on regional survey. Although not a geographer *per se*, he deeply influenced a whole generation of the emerging university and school teachers of geography (as well as other disciplines) through his promotion of re gional field work through the UK Le Play Society (fo cusing on place–work–folk relationship) and the summer schools he organized initially in his Outlook Tower in Edinburgh. A. J. Herbertson, who had been an assistant to Geddes at Dundee before being appointed as Lecturer in Regional Geography at Oxford in 1899, published on natural regions 1905 and 1913. Herbertson, who had studied in Germany and whose doctoral work had been on global rainfall, used a larger scale of regions, being based largely on physical factors and particularly climate, with cultural and political aspects a secondary factor – although he did revise his views and by the time of his premature death in 1915 he was writing of the genius loci or 'spirit of place', derived from people's relationship with place. Herbertson's approach has been described as 'regionalization' rather than regional geography as such and was widely disseminated to teachers and through texts in Britain, as well as in Europe, but was criticized as a form of geographical determinism by H. J. Fleure.

Fleure, influenced by Geddes, sought to understand regions as much as the interaction between people and people, as people and environment. He also characterized global regions, but ones defined by human well being: regions of hunger, debilitation, increment, effort, dif ficulty, wandering, and industrialized regions – with the latter being able to replace any of the former. Both Fleure (trained in anthropology) and Newbigin (trained in biology) tried to establish empirical generalizations from regional studies, despite the limitations of available data. Another regional approach influenced by Geddes was that of Charles Fawcett whose idea of functional regions led him to argue for a federal political system for England.

The extent to which the regional discourse or episteme, in Foucault's terms dominated British geography can be seen not only in the number of regional monographs written by academics, but also the fact they were a re quirement in many of the new geography university courses into the 1950s. The methods of regional survey were also to underpin the Land Utilization Survey dir ected by Dudley Stamp, involving over 20 000 school children, as well as most academic geographers and other volunteers nationwide, whose mapping of land use in local areas culminated in a national map which was of strategic importance to agricultural production during the 1939–45 war.

The regional approach in the USA is associated with environmental determinism, notably in the work of Ellsworth Huntingdon and Ellen Churchill Semple. In 1915 Huntingdon used climate to explain the distri bution of civilization across the globe, and closely as sociated this with race. In her 1911 *Influences of Geographic Environment* Semple, influenced by Ratzel, stressed the role of field observation and causal relations and wrote that people and their way of life in a given area were influenced by the earth's surface in their locality. Semple stressed environment rather than race and emphasized the influence rather than determining character of en vironment, but her work was often read as environ mentally deterministic. Although it has been argued that the strength of Semple's argument was influenced by the wider context of the disciplinary need to establish a justified niche for geography within academia in the USA (just as geographers had strategically promoted geography in Europe earlier), it was this combination of observation based methodology and environmental in fluence/determinant which can be traced as a dominant discourse in early twentieth century American geog raphy, as well as having international influence within what was still a relatively small international community of professional geographers. In Australia, Griffith Taylor argued against government policy to populate the out back on the grounds of scientific information about the environment, and felt he was vindicated in the long term after being treated as a political pariah in Australia. Environmental determinism has been identified at best as geography's contribution to social Darwinism and at worst providing justification for imperialist/expansionist racism; it has been further criticized not only for its ideological but also for its methodological shortcomings as well, particularly by materialists favoring technolo gical innovation as a solution to environmental con straints. The geographical work of Isaiah Bowman and Carl Sauer was contrary to the environmental influence/determinist trend in American geography, the latter's chorological regional studies were conceptually framed by possibilism, recognizing the interdependence of people and natural environment, which combine to produce cultural landscapes – an approach which was a precursor to more explicit humanist and cultural geography to follow. In the USSR the economic definition of regions dominated, originating with Nikolai Baranskiy, a leading Bolshevik and friend of Lenin who promoted the subject in schools.

Although expressed differently by the varying national and theoretical schools, the region provided newly professionalized geographers in the early twentieth century with a *raison d'être* which unified physical and political/human studies and provided a framework for local and international studies at a time when geography had to define its disciplinary content and its place in academia. There were clearly weaknesses in the various regional epistemes, not least, in their scope as encyclopedic studies could be overwhelming with detail making analysis difficult; many (but not all) were descriptive rather than analytical; and the focus was on the unique limited comparative studies and theories. Ironically it was the sense of unique characteristics, the idiographic element of regional studies, and their often humanistic/cultural basis which was revived later in late twentieth century cultural and postmodern approaches to geography, if not the region as geographical unit. Despite this, regional geography has suffered from historical homogenization and caricature, and its significance underestimated in the history of the discipline. Of course, regional studies have never completely disappeared, some geographers are known as 'regional specialists' through their work on a particular part of the globe, for example, Latin America or Eastern Europe, but regional studies have often been subsumed in historical or economic studies such as regional development since the 1960s, and it is only in recent years that a more critical regional approach has begun to emerge, perhaps not surprisingly within economic geographies.

Examination of the move to spatial science and quantitative techniques sheds some light on both the internal disciplinary and external contextual factors influencing the dramatic shift from regional studies in the 1960s and 1970s.

Spatial Science and the Quantitative Revolution

The work of Walter Christaller is a central element in the theoretical and methodological shifts of the 1960s, although its significance was only appreciated long after its original publication (which can be interpreted either as an example of the nonlinear development of geographical thought or as a historical antecedent of later work – or both). Christaller's 1933 doctoral thesis *Central Places* in Southern Germany and his Central Place Theory created a geometric economic hierarchical spatial model, was based on his training in economics and was a departure within geography which opened up new horizons. Christaller's work was undertaken within a growing field of locational analysis in geography, stemming from the regional science trend within Germany in particular. His location theory was novel in including nonmaterial goods such as services, but he did not achieve the hoped for international recognition for central place theory at the IGC Congress in Amsterdam in 1938. Keen to see his ideas applied, he sought to influence planning in Germany, especially of new settlement networks. Christaller, who had returned from serving in World War I espousing pacifism and socialism, worked with National Socialist Regional Planning in occupied Eastern Europe 1940–44. This resulted in wariness about his work in Germany in post war years, and his ideas were more readily accepted and developed in Sweden (by Edgar Kant and Torsten Hägerstrand), the USA (notably by Harris and Ullman), Canada and Switzerland in the 1960s and 1970s. Most undergraduates will be familiar with Central Place Theory, but Christaller's other legacies to the discipline are the understanding of the links between theoretical and applied work and between theoretical analysis and empirical field research – his model was seen as representing a training in a scientific approach to geography. Christaller's work gained a wider international audience when translated from German into English in 1966 (followed by Italian) and has been applied in regional planning across the globe, from the polders of the Netherlands to Japan.

Geographical work and skills had been employed by all parties in World War II, geographers were employed in the armed services and in the production of intelligence materials (notably the allied forces Intelligence Handbooks which covered actual and potential theaters of war), meteorology, cartography, map reading skills, and interpretation of aerial photographs, as well as specialist applications for predicting localized tide and beach conditions for landing forces. Male and female geographers were employed for their geographical skills and the allied forces drew on volunteers from the USA and British Commonwealth. Although World War II itself did not produce immediate change in the practice and theory of geography, American geographers were criticized for their focus on limited regions, and a more synthetic spatial science emerged to dominate geography internationally, and the role of the natural environment was pushed into the background. It has been suggested that just as the regional approach offered an alternative to environmental determinism, so spatial science and the quantitative revolution offered an ideologically neutral and scientifically objective approach to knowledge in post war years. Despite various attempts to reformulate the regional method in geography, such as Dudley Stamp's application of small scale studies to address larger questions and problems and R. E. Dickinson's work on functional regions in the 1930s and 1940s, the approach was still widely criticized. The detailed regional approach promoted by Hettner was described by Hägerstrand as 'mission impossible' in the face of growing complexities of lived geographies and Hägerstrand's work on innovation diffusion broke away from the regional mode by focusing on process rather than place. While there was rejection of the lack of theory and explanation in the regional approach, its merits were not completely overlooked and regional science and statistical recognition of regions featured in new approaches to geographical knowledge in what was described as spatial science. This was closely allied to what is known as the 'quantitative revolution' in geography and wider social sciences and allied closely to positivism, with its emphasis on a unitary scientific method based on induction and verification. It replaced comprehensive studies of unique areas with a desire for observable laws which could be established and tested through the application of the scientific method and has been characterized as a battle between descriptive idiographic and generalizing or nomothetic systematic approaches. Although, Schafer had argued as early as 1953 that the subject was still too grounded in exceptionalism and that geographers should be more interested in similarities and causal explanation rather than descriptions of difference.

The origins of the quantitative revolution is attributed to the USA/geography departments of Iowa, Evanston, Seattle, and Madison, and was disseminated in the UK through the Study Group in Quantitative Methods in the Institute of British Geographers. The IGU meeting at Lund in 1960 was pivotal in recognizing Christaller's earlier contribution and offering an international forum for related work. In Sweden, Lund University, where Hägerstrand worked, was the center for spatial science and quantitative techniques, with developments in the field published in the internationally influential Lund Studies in Geography series, including William Bunge's 1960 *Theoretical Geography* where he described spatial patterns in terms of 'morphological laws'. In the USA William Garrison (University of Seattle, Washington) played a pivotal role in promoting quantitative techniques, through courses in location theory, urban and economic geography, and the graduate seminar in mathematical statistics offered from 1955 onwards, which trained many of those to lead the quantitative movement in American geography in the 1960s, e.g., Brian Berry, William Bunge, and Richard Morrill. Key figures in the new movement in Britain also visited Seattle and other similarly oriented departments and Hägerstrand's work gained a much wider audience after visiting the USA.

In the UK S. Gregory's *Statistical Methods and the Geographer* and John Cole and Cuchlaine King's 1968 *Quantitative Geography* were influential undergraduate texts, but theoretically the way was led by Richard Chorley and Peter Haggett's publications, especially their 1967 *Models in Geography*, which was considered to be something of a methodological watershed. David Harvey's 1969 *Explanation in Geography* was very influential as the most developed exploration of the philosophy of positivism in geography. These books went into reprint, appeared in paperback and were translated into numerous languages; they also sparked heated debates within the discipline concerning the nature and purpose of geography. As has already been hinted at, this methodological shift did not spring fully formed from within geography but reflected other changes and pressures within and beyond the discipline.

Just as the foundation of geography was set within wider contextual and discursive frameworks, so changes within the discipline are similarly framed. The influence of the war and the attraction of 'ideologically neutral' epistemologies have already been noted, as well as the desire to construct geography as a science (social and/or physical), but there were also other international and national trends which played their part in geography's disposition to change in the 1950s and 1960s. Technological change was a significant factor, not least in the increased access to computers able to deal with large data sets, inferential statistics and predictive models. All natural and social sciences were undergoing a quantitative phase in the 1960s. The increasing trend to systematic specialization within subdisciplines in geography from the 1950s onwards, encouraged and necessitated geographers to work with colleagues in cognate disciplines, seen in the fruitful interdisciplinary work with economics, following Christaller's lead and the development of 'social physics' in the USA, but this simultaneously represented the spectre of rupture to geographers schooled in a unified view of geography.

The quantitative revolution can be seen as part of post war technological approaches to management and planning in countries developing welfare policies and/or undertaking reconstruction after war damage. There was also a wider sense that universities were expected to produce problem solvers to meet the needs of increasingly complex society. Contextual factors also explain the late acceptance of the quantitative revolution in Germany: wariness of Christaller's work noted above, coupled with a relatively unitary approach to geography in which international rather than home based studies were considered the pinnacle of knowledge and achievement: the 'new' ideas were introduced to German speakers by Steiner's 1965 study of Swiss cantons which used factor analysis. It has been argued that the use of the positivist hypothetico deductive method and inferential statistics brought physical and human geographers together methodologically and epistemologically, as well as gaining increased respectability and funding through association with the methods of the natural sciences and maths – but it has also been suggested that in many cases positivist epistemologies followed quantitative methods in geography rather than vice versa.

However, while the so called quantitative revolution spread in receptive departments, it was far from un contested and met resistance in some established jour nals, as well as particular departments and was criticized for its language and methodologies which seemed in decipherable to the uninitiated. Journals such as *Progress in Geography* and *Geographical Analysis* were founded to provide outlets for the innovative work, but critiques continued with warnings of disciplinary civil war and fragmentation, the limitations of quantitative techniques in some fields of research, the neglect of the environment as a reality, and the Occidental nature of work centered on existing large data sets. Peter Taylor argued from the sociological perspective on disciplinary change that the role of mathematics in spatial science was not merely to facilitate quantification and verification, but also to fa cilitate the establishment of a new generation of geog raphers who defined themselves against what went before and established a new research agenda and methodology, accompanied by its own specialist language which ex cluded traditionalists.

By the early 1970s even proponents such as David Harvey and Bill Bunge were recognizing the limitations of spatial science and quantitative techniques in the face of social inequality. Geometry had become the language of geography, but it was a language and practice that relied on representing and creating fixed patterns which did not necessarily reflect the dynamic socioeconomic geography contained therein, or, more basically, did not incorporate the people whose lives were modeled. Problems also emerged with the application of some statistical methods to spatial phenomena with multiple influences (notably spatial autocorrelation). While spatial science built on regional science, it attempted to apply methods from the natural sciences which it was thought would achieve sci entific rigor; the focus was shifted from regions to ques tions of spatial order and to real applied issues: quantitative techniques were central to achieving these aims. These epistemological and methodological changes created a community, which exchanged ideas, methods, and personnel, but claims that it was revolutionary need to be set in a wider context.

By the late 1980s, the positivist, if not always the quantitative, turn in geography had come under serious criticism. It was challenged by those economic, social, and urban geographers following a political economy approach; one of the most influential critics was David Harvey, who found the previous attractions of quantifi cation inadequate to the task of understanding and the orizing the historical geography of capitalism and the associated inequalities of uneven development and pov erty locally and globally. However the challenge to quantitative techniques did not only arise from materialist critiques, another principal strand was ex pressed in behavioral and humanistic geography. More recently postmodern approaches have articulated a dis trust of a single privileged path to knowledge and today geographers are taught to choose the methods appro priate to their subject of enquiry. Despite these critiques, the value of quantitative techniques persists within the discipline, although not in the logical positivist sense as the only way to ascertain truth. In the USA there seems to be more of a dichotomy between spatial–analytical and social–cultural geography, whereas in the UK quantita tive methods, hypothesis testing and verification are now usually seen as one means of measuring and analyzing human phenomena and are often combined with field study and small scale intensive studies using qualitative techniques such as interviews and participant obser vation. Similarly, qualitative studies can incorporate quantitative data but interpret findings within a qualita tive framework.

The Last Thirty Years

The idea of a quantitative revolution (influenced by Thomas Kuhn's work) was popularized within the history of geography at a time when dramatic social changes were occurring in terms of race, class, and colonial emancipation. However, as has been suggested above, while representing a significant shift in geography's epistemology and methodologies, the spatial science of the 1960s and 1970s had antecedents in earlier work, failed to extinguish existing approaches, and spawned competing epistemes. The character of successive alter native approaches to geographical enquiry are revealing in terms of their impact on the discipline epistemo logically and methodologically. Two very philosophically different responses to the limitations of spatial science and positivist epistemologies were those of Marxist/so cialist and humanist geographies (both representing broad coalitions). Marxist/socialist and other radical geographies are ideologically related to earlier nine teenth century anarchist and socialist geographies; they are grounded in a materialist view of the world which centers on the power relations of capital, notably seen in the work of David Harvey, Neil Smith, Doreen Massey, and Richard Peet for example. While quantitative tech niques may be useful to demonstrate these power re lations they are explanatorily inadequate in the face of economic power distributions. At once highly theoretical and seeking application to social problems, this school of geographers has been very influential within social, urban and economic geography in particular, normalizing the consideration of socioeconomic class and power in geo graphical studies. In contrast, humanist geographies drew inspiration from the Vidalian school as well as phenomenology or existentialism which made human experience central to geographical questions and methods, as articulated by Anne Buttimer and applied by David Lowenthal and Yi Fu Tuan for example. In turn these approaches were found at least partially wanting by others working from different theoretical perspectives: post structuralists who stressed the multiplicity and instability of the meaning of texts; postmodernists who emphasized difference and the inadequacy of grand metatheory and notions of objectivity; postcolonialists who exposed ways in which Western epistemologies can be embedded in racialized colonial power relations; feminists who identified both the gender bias and absence of gender in geographical epistemologies and methodologies; lesbian, gay, and Queer theorists who stress sexuality as a crucial factor in understanding spatial phenomena. These epistemological approaches have produced related methodologies (including an important range of qualitative techniques) and are associated with particular subdisciplines, but many have been creatively combined as can be seen in cultural geography which has widened the ontological base of what can be considered as geographical enquiry and the related epistemological questions and methods, both of which draw theoretically on numerous of these recent (as well as past) theoretical approaches within human geography, creating hybrid geographies.

Just as spatial science and quantitative techniques were found to be of limited application and explanatory value, so other approaches offering universal analytical frameworks or metatheories have been found to have their limitations and human geographers increasingly draw from more than one epistemology, for example, many development studies have combined postcolonial, socialist, and feminist theories. However, not all geographers are happy with this creative mix approach to geographical enquiry which has been characterized as alphabet soup, and some hold a fixed ontological position for philosophical, ideological, political, and/or strategic reasons. Similarly, this subject and philosophical diversity has had an impact on the scope and character of the history of geography, as can be seen in the cases of gender, imperial, and postcolonial geographies.

While we should be wary of histories that stress the continuity and unity of the discipline, for example, Hartshorne's 1939 *The Nature of Geography*, which infamously addressed deviations from the mainstream development of geography, we need to be sensitive to the evolution and diffusion of ideas in the face of the representation of attractive 'heroic revolutions'; we also need to be sensitive to the multiplicity, synchronicity and often circular loops in the development of knowledge and associated methods and theories. After Foucault, changes in Western thought have been seen more in terms of shifts and discontinuities, discursive rivalries, and combinations rather than the binary divide between change by evolution or revolution. Arguably both development and environmental agenda have gone some way to bringing physical and human geographers into conversation with one another after the post war divide between physical and human geographies. Technological changes have brought GIS to geography and GPS to the public; email and the Internet make for near instant communication which eases collaborative work and both temporary and permanent global interest networks such as specialist research networks and the Critical Geography Forum. External political forces impact on sources of funding and the means of evaluating geographical output. Like academia as a whole, human geography is a much more diverse field than it was in the 1950s. However, while some see that as a threat to the unified definition and practice of geography (which can be important for a secure place within the broader academy), it can also be seen as a strength that human geography has become much more sensitized to and expressive of a wide range of social, cultural, economic, and political relations. In turn, approaches to the history of the discipline have to be sufficiently sensitive to that range, to difference within history, in order not only to represent that variety but gain historical insights from different sources and different epistemologies, resulting in more inclusive, perspectives on the history of geography.

See also: Geohistory; Possibilism; Regional Geography I.

Further Reading

Bell, M., Butlin, R. A. and Heffernan, M. J. (eds.) (1995). *Geography and Imperialism, 1820 1940*. Manchester: Manchester University Press.

Blunt, A. and McEwan, C. (eds.) (2002). *Postcolonial Geographies*. London: Continuum.

Bunge, W. (1962). *Theoretical Geography*. Lund: Gleerup.

Buttimer, A. (ed.) (1983). *The Practice of Geography*. New York: Longman.

Dickinson, R. E. (ed.) (1976). *The Regional Concept. The Anglo American Leaders*. London: Routledge and Kegan Paul.

Domosh, M. (1991). Towards a feminist historiography of geography. *Transactions of the Institute of British Geographers* 16, 95 104.

Driver, F. (2001). *Geography Militant. Cultures of Exploration and Empire*. Oxford: Blackwell.

Freeman, T. W. (1961). *A Hundred Years of Geography*. London: Gerald Duckworth.

Godlewska, A. (1999). *Geography Unbound*. Chicago, IL: University of Chicago Press.

Godlewska, A. and Smith, N. (eds.) (1994). *Geography and Empire*. Cambridge, MA: Blackwell.

Hartshorne, R. (1939). *The Nature of Geography*. Washington, DC: Association of American Geographers.

Harvey, D. (1969). *Explanation in Geography*. London: Edward Arnold.

Harvey, D. (1973). *Social Justice in the City*. London: Edward Arnold.

Johnston, R. and Williams, M. (eds.) (2003). *A Century of British Geography*. Oxford: The British Academy/Oxford University Press.

Keighren, I. M. (2006). Bringing geography to the book: Charting the reception of influences of geographic environment. *Transactions of the Institute of British Geographers* 31, 525 540.

Kimble, G. H. T. (1951). The inadequacy of the regional concept. In Stamp, L. D. & Wooldridge, S. W. (eds.) *London Essays in Geography. Rodwell Jones Memorial Volume,* pp 151 174. London: Longmans, Green.
Kwan, M. P. (2004). Beyond difference: From canonical geography to hybrid geographies. *Annals of the Association of American Geographers* 94, 756 763.
Livingstone, D. N. (1992). *The Geographical Tradition*. Oxford: Blackwell.
Lorimer, H. (2003). Talking small stories: Spaces of knowledge and the practice of geography. *Transactions of the Institute of British Geographers* 28, 197 217.
Maddrell, A. (2009). *Complex Locations. The Production and Reception of Women's Geographical Work and the Geographical Canon 1850 1970.* Oxford: RGS IBG/ Blackwell.
Martin, G. J. (2005). *All Possible Worlds* (4th edn.). New York: Oxford University Press.
Monk, J. (2004). Women, gender and the histories of American geography. *Annals of the Association of American Geographers* 94, 1 22.
Peet, R. (1985). The social origins of environmental determinism. *Annals of the Association of American Geographers* 75, 309 333.
Rose, G. (1993). *Feminism and Geography*. Cambridge: Polity.
Stoddart, D. R. (ed.) (1981). *Geography, Ideology and Social Concern*. Oxford: Basil Blackwell.
Withers, C. W. J. (2002). Constructing the geographical archive. *Area* 34, 303 311.
Withers, C. J. and Mayhew, R. J. (2002). Re thinking 'disciplinary' history: Geography in British universities, *c*. 1580 1887. *Transactions of the Institute of British Geographers* NS 27, 1 19.

Geohistory

P. Claval, Université de Paris-Sorbonne, Paris, France

Glossary

Fernand Braudel (1902–85) Historian; Braudel taught in Algeria and worked on the history of the Western Mediterranean during the Renaissance. His doctoral dissertation, *La Méditerranée et le monde méditerranéen à l'époque de Philippe II* (1949), made him famous. He held important positions in the French academic system as a professor at the Collège de France and the Ecole des Hautes Etudes en Sciences Sociales, and as chairman of the Maison des Sciences de l'Homme. He was a historian of the slow evolution of economic and social structures. His influence was important in Italy, Spain, Latin America, Poland, and, after the translation of his major books into English, in Britain and the United States.

Longue durée "Traditional history, with the attention it devotes to short time, got us used, since a long time, to its brief time, its rapid, dramatic, short breathed narrative [...] There is also, besides this narrative, a recitative of conjuncture, which involves the past by large slices: ten, twenty or fifty years. "Much beyond this second recitative is situated a history with a wider breath, of secular length this time: long duration history. (Braudel, 1958, 1969)

"Good or bad, [the term structure] dominates the problems of long duration. By structure, the observers of social realities understand an organization, a coherence of relatively stable relations between realities and social masses. For us, historians, a structure is certainly an assembling, an architecture, but is much more a reality that time does not wear away and which persists for very long periods." (*ibid.*)

The Annales School The *Annales School* of history was a way to conceive history developed from 1929 by Lucien Febvre and Marc Bloch through a new journal: *Annales. Sociétés, économies, civilisations*. Instead of focusing on the events of political life, it analyzed the demographic, social, and economic transformations of the world as well as the symbols which weighed upon them. The *Annales School* dominated French history until the 1980s and had a strong international innfluence. It provided a new interest in economic realities, geographic constraints, social organizations, and representations.

The French School of Geography Developed in France at the end of the nineteenth century, geography blended the lessons of German geographers (mainly Humboldt, Ritter, and Ratzel) with the French tradition of historical approach and regional analysis. Shaped by Paul Vidal de la Blache, this conception dominated French geography for two generations. Its impact was strong abroad, particularly in Spain, Italy, the Netherlands, and Britain. For the first time, geography focused on regional analysis – or, in a simplified and often caricatural way, regional description.

Introduction

The term 'geohistory' was coined by Fernand Braudel in the conclusion of the first part of his doctoral dissertation on *La Méditerranée et le monde méditerranéen à l'époque de Philippe II*:

> Geohistory? [...] when speaking of geohistory, we wish to designate [...] something more historical and at the same time broader [than geopolitics], which would not be only the application to the present and future situation of States, of a schematized spatial history [...].
>
> In fact, a double approach. To put human problems as they may be observed, spread over space and as far as possible, mapped, a human intelligent geography: yes, certainly, but to put them not only for the present and in the present, but in the past [...]; to build a really retrospective human geography: [...] to pay more attention to time and [...] to mind more about space, what it supports, what it generates, what it impedes in a word, to [...] take sufficiently into account its formidable permanency. (Braudel, 1949)

Such was the program illustrated by the first part of the book: a wonderful analysis of the Mediterranean, its environments, the sea and its borders; an evocation, also, of their transformation:

> A history slow to run and to change, often made of unceasing move backs, of starting anew cycles. I wished neither to neglect this history, almost out of time, in relation with inanimate things, nor content myself, concerning it, of these traditional geographical introductions to history [...] with their mineral landscapes, ploughed fields, and flowers which are rapidly presented, and from which there is nevermore question, as if the flowers did not bloom again each spring [...]. (Braudel, 1949)

Such are the main characteristics of Braudel's geohistory: its foundation in Vidalian human geography ("a human intelligent geography"), its curiosity for the links between social life and 'inanimate things', and its relation with long duration, the other major concept Braudel invented at that time.

The idea of geohistory, as developed by Braudel, was rooted in the history of French geography and history since the beginning of the twentieth century. His con ceptions greatly influenced historians and geographers.

Environments and History as Conceived by Geographers and Historians from the Early Twentieth Century

The History Practiced by Vidalian Geographers

Vidalian geography was shaped by Vidal de la Blache and his students, who were responsible for the development of a new 'genre': the regional monograph. From the be ginning of the twentieth century, doctoral dissertations in geography were devoted to the study of a particular re gion. These dissertations stressed geography's physical features, the nature of its environments and the present distribution of its population, settlements, and activities. An emphasis was put on the historical dimension of re gional structures: the various milieux were conceived as active factors in local development. In Jules Sion's *Les Paysans de la Normandie orientale*, for instance, the econ omy of the fertile Pays de Caux plateau was initially based on subsistance farming. Since its social structure was deeply unequal, poor peasants also became weavers from the seventeenth century onwards: merchants pro vided them with cotton and commercialized their pro duction. Pays de Caux produced enough food for this proto industrial population.

From the beginning of the nineteenth century, the significance of the environment changed: the well watered rivers dissecting the plateau were harnessed in order to move the new weaving machines imported from England, hence the development of new industrial vil lages in the valleys. Later, these factories relied increas ingly on coal imported from England through Rouen, which allowed for the further growth of the textile in dustry in the lower valleys converging towards this port.

In most regional dissertations, geography was not a neutral framework; it participated in the history of the studied areas.

Vidal de la Blache stressed more systematically the role played by the environment in the articles he pub lished on the genesis of "the great human agglomer ations" in the *Annales de Géographie*: the success of the moderately fertile Mediterranean was linked to the close juxtaposition of plains, hills, plateaus, mountains, and deeply indented coasts, since they created comple mentarities, explained the early development of trade, and the construction of polities stretching over these various milieux and controlling them.

Pierre Denis used a similar historical approach to explain the specificities of a national state: Argentina. The country was made of the Andes cordillera, their piedmonts, low plains, and plateaux including the tem perate Pampa. During the Inca period, it was organized along the route that linked the Bolivian highlands to central Chile through the Mendoza oasis and Cumbra Pass. The significance of this road was reinforced during the colonial period by the silver mines of Potosi: the piedmonts oases provided the latter with food, thanks to long caravans of pack mules moving up and down the mountain slopes. A side road was opened during the sixteenth century to the Paraguay River and further down to the Plata River, Buenos Aires, and Europe. A complete recentering of the country on Buenos Aires occurred in the nineteenth century with the conquest of the Pampa and the development, in that part of the country, of commercial of agriculture based on export ation. A few years later, Emile Félix Gautier proposed a fantastic geohistory of North Africa.

About the same time, Albert Demangeon explored the British Empire with a similar ambition: to study its evolution through the way British colonizers invented new methods to organize and control space.

To use the analysis of regional differentiation, the nature of different environments, and the techniques of spatial organization as key factors in the building of territorial units was thus a tradition in French geography.

The Geography Practiced by Historians

Historians and geographers received the same training in France until the 1940s. Both groups shared a similar interest in regional realities. Influenced by Vidal de la Blache, Henri Berr, a leading historian of the time, launched a series of monographs in 1903 on *Les régions de la France*. The aim of these monographs was to analyze the historical individualities that people had built in each of these areas. Lucien Febvre – a close friend of Jules Sion in the *Fondation Thiers* – was preparing a doctoral dissertation on *Philippe II et la Franche Comté*. He accepted to compose the volume on Franche Comté: *Les régions de la France. IV La Franche Comté* was published in 1905.

Febvre first stressed the differences between various components of this province:

> Such are, in their deep diversity, the constituting parts of the *comtoise* province. What do they share? Very few things, apparently: neither their origin nor their geo logical formation [...]. And in their natural and agri cultural flora, the distribution of their populations, the conditions nature offered to their settlements, how many contrasts [...]: it was an endless variety of materials, plans, agglomeration, and aspects (Febvre, 1905: 19).

This variety appeared, however, as an advantage for building a strong territorial unit: "It is a great surprise to see all these contrasts, all these dissimilarities melted together in the political and moral unit of a province, associated for centuries for the same destinies." Why? Because of the complementary nature of all these elements.

> [Franche Comté] remained the privileged country which was self sufficient and absorbed all its production. For the meat of his cattle, the *montagnon* received the wheat of the low country; for the wine of the vineyards, he offered his cheese: old secular practices of trade, observed over time by travellers; they kept on uniting indissolubly the Burgundy of pastures to that of open fields (Febvre, 1905: 66).

The *Annales* school of history, as founded by Lucien Febvre and Marc Bloch at the end of the 1920s, relied on the idea that history had much to gain in cooperating with other social sciences, economics, sociology, anthropology, or geography. For them, the latter offered rich perspectives on the long term evolution of countries and the slow rythm of their transformations.

Braudel's Geohistory

Philippe II and the Mediterranean

Braudel was well aware of the possibilities geography offered for explaining the construction of territorial units. When preparing his doctoral dissertation, he developed a rich personal experience of the Mediterranean: he lived in Algeria and spent a few months every year exploring Spain, Italy, the former Yugoslavia, and their archives. He was perfectly aware of the diversity of Mediterranean environments and relied on the categories proposed by Vidal de la Blache to analyze how they were used by human groups, and what they offered to them, at the end of the Middle Ages and during the Renaissance. Mountains, plateaus (and hills), and plains (either inland basins, piedmonts alluvial fans, or sea level marshy lowlands) were the basic elements of the Mediterranean reality. These regions were linked by the sea, which explained the similarity of their climates. It was a factor of human unity when navigation was safe, and a cause of isolation when insecurity prevailed – hence the ups and downs experienced by the islands. The *rivieras* played a major role in the development of trade. The construction of a unity out of this mosaic of humanized environments was achieved through the development of commercial routes and the initiation of city formation.

For Braudel, the environment was not given once and for all. It changed because of trade and economic relations, or the building of roads and ports, and thanks to the initiative of urban centers and political rulers. The unity of the Mediterranean countries was real, but its human half was a product of human initiative as much as the expression of environmental conditions. Hence the refutation of environmental determinism which concluded the first part of *Philippe II et la Méditerranée* and was more clearly formulated in the second edition of the book.

Exploring Geohistory Through the Relations Between Material Civilization, Economy, and Capitalism

As conceived by Braudel, geohistory explored the role of environmental features in history, with the idea that they evolved, even slowly, through time, and were permanently and actively reinterpreted by social groups through commerce and the building of political organizations: hence the role of sea and land routes and the growing influence of cities; hence also the growing significance of those who lived in these cities, that is, from the second half of the Middle Ages on, of their *bourgeoisies*. In this way, geohistory was central in the Braudel's study of Western capitalism.

From Vidal de la Blache and Lucien Febvre, geographers and historians were conscious of the role of technical innovations in the changing relationhips between social groups and their environment: it was the basis of possibilism. In this perspective, Braudel explored first the technical dimension of man–milieu relationships: it was the theme of *Civilisation matérielle and capitalism*. The book delved not only into the tools and methods of farmers or cattle raisers, but also into the exploration and exploitation of mines, and the machines built for metallurgy or textile industries. The chapters on the techniques used for building houses or roads, and on cities as geographical devices, were really new. With the development of urban *bourgeoisies* and capitalism, the relations between human groups and their environment were increasingly mediated through markets.

The second volume of the trilogy on *Civilisation matérielle, économie et capitalisme, Les Jeux de l'échange*, analyzed the way some Western European big cities, and especially harbors (like Genoa and Antwerp, and later on Amsterdam and London), enlarged trade realms and dominated them through the monopolies they developed – with the complicity of the new sovereign states.

Geohistory ceased to be focused on the evolution of particular areas, like the Mediterranean. The last volume of the trilogy, *Le Temps du monde*, covered this aspect of geohistory: Western countries managed to build for the first time a World Economy, with its core area in Western Europe, and its impact all over remote and related outer spaces.

The Last Phase: *l'Identité de la France*

In this way, Braudelian geohistory had turned its major focus towards economy, trade, money, and power. It offered an original interpretation of the changing world scene – later exploited by Immanuel Wallerstein and geographers like Peter Taylor. But such an approach forgot one aspect of the evolution of geographical real ities through time: the remarkable permanence of some of their features. Braudel faced this problem in his last and uncompleted book, *L'Identité de la France*. In the first volume, he analyzed the physical, or 'geographic', en vironment and concluded that it was unable to explain the stability of French social and spatial structures over time. Was the reason for their permanence to be re searched in economy and social organization? No: the only significant factor in this field was the weight of peasantry in French society, which contributed to the slow rythm of change characteristic of French history during the last century.

The reason for the historical solidity of France was certainly to be sought in the cultural field. Braudel died before completing the last part of *L'Identité de la France*. As a result, his vision of geohistory remained incomplete.

Geohistory: An Overall Vision of History

Braudelian geohistory was not an easily defined part of history. His ambition may be really assessed in his text book, *Grammaire des civilisations*: geohistory was for him a way to combine synthetic historical and geographical approaches to build an overall picture, both spatial and temporal, of human civilizations and destiny.

Braudel first made known this conception of geohis tory in 1949, but to measure its breadth, people had to explore all his works: it was not a limited research pro gram which was easy to apply. That explained why 'geohistory' did not attract much comment and was not transformed into a new historical genre. Just as in the early twentieth century, it was a way of combining history and geography as practiced by both historians and geographers.

Braudel's geohistory changed along his career, since for him space was increasingly made of networks, infra structures, and mental schemes: natural constraints, still very significant in 1949, were losing their weight.

Historians' Geohistory

Regional Geohistory

French historians often chose to study the history of regional entities in their doctoral dissertations. Lucien Febvre's *Philippe II et la Méditerranée* was the first example of this orientation. He was mainly imitated from the 1940s onward. René Duby proposed his new and very fecund conception of feudal times through his analysis of Mâconnais. Emmanuel Le Roi Ladurie followed the transformations of Languedoc peasantry from the Middle Ages to the eighteenth century. Goubert showed how the social hierarchy functioned in the seventeenth century rural areas of Beauvaisis. After 1970, there was a decline in this type of doctoral dissertations.

At a Larger Scale

The most interesting examples of geohistory were de veloped at another scale: instead of dealing with the Mediterranean, like Braudel, René Grousset, a well known essayist of the inter war period had produced a thought provoking book on *L'Empire des steppes*, in which he explored the impact of long range nomadism on sedentary civilizations from the fifth to the sixteenth century.

Maurice Lombard was a specialist on Islam. In *L'Islam dans sa première grandeur*, he proposed a geographical in terpretation of a well known fact: "out of the successive phases experienced by Islam, [...], the first was that of its widest extension, the apogee of its power and civilisatory expansion" (Lombard, 1971). His interpretation was a simple one: "The Islamic moment in world history was explained by the construction of a wide economic net work, putting in touch until then distinct realms" (Lombard, 1971).

Denys Lombard summarized his knowledge of Java in a wonderful book: *Le Carrefour javanais. Essai d'histoire globale*. The reality in Indonesia was a complex one. In order to decipher it, the best way was to start with the influence of the West since it was so strong from the sixteenth century onwards. Slightly deeper lay a layer of urban and trading communities, whose mental con ceptions and mental maps were shaped by Islam and the Chinese influence. The third and deepest layer was built at the time when the agrarian kingdoms of Central and Eastern Java were shaped by Indian culture, with its concentric cosmographies. In this way, Denys Lombard interpreted the history of Java in terms of spaces and networks, agrarian inland areas, and trading seashores, and renewed significance of crossroads.

The work of Pierre Chaunu, a historian of the Spanish colonization of Latin America and Atlantic economy in the modern times, paralleled in many ways that of Fernand Braudel.

Geohistory as the Study of Networks

The orientation of the *Annales* school of history changed deeply in the late 1970s and 1980s. Economy lost its prominence. There was a growing interest in everyday life and mental representations. History also became more spatial. The space it explored had nothing to do with the natural environments which appeared so im portant in the analyses of early twentieth century his torians. Space was constantly reshaped and restructured by human agency. It did not appear as a surface, or a set of discrete surfaces: it was increasingly made of nodes and networks.

Bernard Lepetit, who was very influential in this evolution, was a specialist on urban history: he practiced geohistory of an abstract and systematic brand: in a way, he had just moved a step further than Braudel – but his geohistory sounded completely different, without any sense of concreteness. Environments had almost com pletely disappeared from his studies, and with them, the sense of atmosphere.

Is it for that reason that geohistory is less popular among historians than a generation ago?

Geographers' Geohistory

A More Dynamic Historical Geography

At the time when doctoral dissertations in geography dealt with regional entities, French geographers were also historians. With the development of more systematic orientations, the form of geohistory involved in the preparation of regional theses regressed. During the 1960s, some of the followers of Pierre George advocated a complete separation between the two disciplines. His torical geography became a specialized field, practiced only by a few geographers, some from the older gener ation, like Roger Dion, and some from the younger, like Xavier de Planhol.

Geographers were, however, fascinated by Braudel: hence the renewed interest in the historical dimension of spatial organizations, either at the regional, national, or international scale.

Geohistory and Chorems

For Roger Brunet, regional analysis had always been the core area of geography. It was mainly among his followers that geohistory developed as a geographical enterprise. For Roger Brunet, linguistic procedures have to be ap plied to geography: just as a sentence may be divided into a set of elementary phonetic units, 'phonems', or a set of elementary significant units, 'sems', regional organization results from the combination of elementary forms of organizations, or 'chorems'.

In such a perspective, geohistory analyzes chorems and the way their combination evolves, or remains stable, through time. It means that strategic and political factors play an essential part – it was already clear in the first book based on this paradigm, Reynaud's *Une Géohistoire. La Chine des Printemps et des Automnes*.

Christian Grataloup is today the main exponent of this new form of geohistory. In his book *Lieux d'histoire. Essai de géohistoire systématique*, he shows how territorial constructions can be reduced to a few types: the mono concentric model, the polycentric one, and the pro tected angle one. To simplify maps in order to demonstrate the main relations they express is an ex cellent idea. As soon as these simplified elements are used to build complex pictures, the advantage of simplicity is lost. The emphasis on topological features is conducive to a weakening of the social and economic dimensions so important in Braudelian geohistory, as well as its cultural significance.

Conclusion

Geohistory was closely linked to the Vidalian conception of geographical analysis, with his emphasis on the chan ging significance of environment depending on techni ques. It was present from the beginning of the twentieth century, both among geographers and historians.

A divorce occurred between French history and geography at the mid century. Most geographers in creasingly considered themselves specialists of the pre sent time. Thanks to Fernand Braudel, the Vidalian paradigm was not lost to historians: his ambition was to write neither pure history nor pure geography, but to combine them in a dynamic apprehension of spatial or ganization. Braudel started with the Mediterranean and later tried to provide a similar approach for the world economy which began to be built from the time of the Discoveries. At the same time, he tried to capture the reasons that explained the remarkable permanence of some geographical features.

Geohistory was not conceived by Braudel as a sub discipline. Some historians and geographers proposed, as Braudel did, synthetic views of places and times. From the 1980s, a more restrictive conception of geohistory de veloped: historians focused increasingly on the role of networks in structuring time; among geographers, the followers of Roger Brunet tried to develop a systematic topology of elementary forms of spatial organizations or chorems. These new developments are interesting, but do not present the fascinating quality of Braudel's geohistory.

Braudelian ideas had a strong influence outside France: Braudel drew a part of his inspiration from German historians; he has very close personal links with Italian, Polish, and Portuguese historians. His ideas about capitalism were diffused in the anglophone world by Immanuel Wallerstein.

The Braudelian theme of geohistory was endorsed by some British geographers, such as Alan Baker, and more deeply, by Irish ones (Estyn Evans, for instance). A general discussion of Braudel's influence is presented by Marino.

See also: Historical Geography; Historical Geography, Evolution of.

Further Reading

Boulanger, P. and Trochet, J. R. (eds.) (2005). *Où en est la géographie historique? Entre économie et culture*. Paris: l'Harmattan.

Braudel, F. (1949). *La Méditerranée et le monde méditerranéen à l'époque de Philippe II*. Paris: A. Colin, (*Eng. trans. The Mediterranean and the Mediterranean World in the Age of Philip II*, vol. 2, New York: Harper and Row.).

Braudel, F. (1958). Histoire et sciences sociales. La 'longue durée'. *Annales ESC* 13, 725 745.

Braudel, F. (1967). *Civilisation matérielle et capitalisme*. Paris: Armand Colin.

Braudel, F. (1969). *Ecrits sur l'histoire*. Paris: Flammarion.

Braudel, F. (1979). *Civilisation matérielle, économie et capitalisme, 3 vols.: 1 Civilisation matérielle et capitalisme; 2 Les Jeux de l'échange; 3 Le Temps du monde*. Paris: Armand Colin.

Braudel, F. (1986). *L'Identité de la France, 3 vols.: 1 Espace et histoire; 2 et 3, Les Hommes et les choses*. Paris: Arthaud Flammarion.

Braudel, F. (1987). *Grammaire des civilisations.* Paris, Arthaud; 2ème éd. de la contribution de Braudel à : Baille, S., Braudel, F., Philippe, R., 1963, *Le Monde actuel. Histoire et civilisation*, Paris, Belin.

Brunet, R. (1980). La combinaison des modèles dans l'analyse spatiale. *L'Espace Géographique* 9(4), 253 265.

Burguière, A. and Revel, J. (eds.) (1989). *L'Espace français*. Paris: le Seuil.

Chaunu, P. (1969). *L'Expansion européenne*. Paris: PUF.

Chanu, P. (1974). *L'Histoire, science sociale*. Paris: SEDES.

Claval, P. (1984). The historical dimension of French Geography. *Journal of Historical Geography* 10(3), 229 245.

Claval, P. (1989). *La Conquête de l'espace américain*. Paris: Flammarion.

Claval, P. (2003). Fernand Braudel 1902 1985. In Armstrong, P. H. & Martin, G. J. (eds.) *Geographers. Biobibliographical studies,* vol. 22, pp 28 42. New York: Continuum.

Claval, P. (2004). *La Fabrication du Brésil*. Paris: Belin.

Denis, P. (1920). *La République argentine*. Paris: A. Colin.

Demangeon, A. (1923). *L'Empire britannique. Etude de géographie coloniale*. Paris: A. Colin.

Dion, R. (1990). *Le Paysage et la vigne. Essais de géographie historique*. Paris: Payot.

Duby, G. (1941). *La Société au XIe et XIIe siècles dans la région mâconnaise*. Mâcon: Annales d'Igé.

Dunford, M. (1998). Economies in space and time: Economic geographies of modernization. In Graham, B. (ed.) *Modern Europe*, pp 53 88. London: Arnold.

Febvre, L. (1905). *Les régions de la France. IV La Franche Comté*, Publications de la *Revue de Synthèse historique*, 77p. Paris: Cerf.

Febvre, L. (1912a). *Philippe II et la Franche Comté*. Paris: Honoré Champion.

Febvre, L. (1912b). *Histoire de la Franche Comté*. Paris: Boivin.

Febvre, L. (1922). *La Terre et l'évolution de l'humanité. Introduction géographique à l'histoire*. Paris: La Renaissance du Livre.

Friedman, S. W. (1996). *Marc bloch, sociology and geography. Encountering changing disciplines*. Cambridge: Cambridge University Press.

Gautier, E. F. (1927). *Les Siècles obscurs du Maghreb*. Paris: Payot.

Grataloup, C. (1996). *Lieux d'histoire. Essai de géohistoire systématique*. Paris Montpellier: Reclus.

Goubert, P. (1960). *Beauvais et le Beauvaisis de 1600 à 1730*. Paris: SEVPEN.

Graham, B. (ed.) (1998). *Modern Europe*. London: Arnold.

Grousset, R. (1939). *L'Empire des steppes*. Paris: Payot.

Kisser, S. (1981). The geohistorical structuralism of Fernand Braudel. *American Historical Review* 86, 63 105.

Lepetit, B. (1988). *Les Villes dans la France moderne (1740 1840)*. Paris: Albin Michel.

Le Roy Ladurie, E. (1966). *Les Paysans du Languedoc*. Paris: SEVPEN.

Lombard, D. (1990). *Le Carrefour javanais. Essai d'histoire globale, 3 vol*. Paris: Editions de l'Ecole des Hautes Etudes en Sciences Sociales.

Lombard, M. (1971). *L'Islam dans sa première grandeur*. Paris: Flammarion.

Lombard, M. (1972). *Espaces et réseaux au Haut Moyen Age*. Paris: Masson.

Marino, J. A. (ed.) (2002). *Early modern history and the social sciences: Testing the limits of Braudel's mediterranean*. Kirksville, MO: Mich.

Ozouf Marignier, M. V. (1992). Géographie et histoire. In Bailly, A., Ferras, R. & et Pumain, D. (eds.) *Encyclopédie de Géographie*, pp 93 107. Paris: Economica.

Planhol, Xavier de, (1988), *Géographie historique de la France*, Paris, Fayard. (eng. trans., (1994). *An Historical Geography of France*. Cambridge: Cambridge University Press.).

Revel, J. (ed.) (1999). *Fernand Braudel et l'histoire*. Paris: Hachette.

Reynaud, A. (1992). *Une Géohistoire. La Chine des Printemps et des Automnes*. Montpellier: Reclus.

Roncayolo, M. (1989). Histoire et géographie: les fondements d'une complémentarité. *Annales E.S.C.* 44(6), 1427 1434.

Sion, J. (1908). *Les Paysans de la Normandie orientale*. Paris: A. Colin.

Taylor, P. (1985). *Political geography. World economy, nation state and locality*. London: Longman.

Vidal de la Blache, P. (1918). Les grandes agglomérations humaines. *Régions méditeranéennes. Annales de géographie* 27, 174 187.

Vidal de la Blache, P. (1922). *Principes de géographie humaine*. Paris: Armand Colin.

Wallerstein, I. (1974 1988). *The modern world system*, 3 vol. 3, New York: Academic Press.

Geomatics

M. Duckham, A. Kealy, and J. Wallace, University of Melbourne, Melbourne, VIC, Australia

Glossary

Cadastre An authoritative, large-scale parcel map representing the existing uses of land.
Geoinformation Geographic information; information that is referenced in some way to locations on the Earth.
Geosensor Network A wireless sensor network used to monitor geographic environments.
Land Administration The institutions and processes involved in organization of land uses and delivery of security of tenure.
Location-Aware Computing The integration of positioning information and mobile computing devices to provide information to a user related to that user's current location.
Location Privacy The right of an individual to determine to what extent information about the location of that individual is communicated to others.
Ubiquitous Positioning System A positioning system that can provide positioning information everywhere required by a specific application (e.g., globally or across a wide range of indoor and outdoor environments).
Uncertainty An endemic feature of spatial data arising from imperfection in that data, including imprecision, inaccuracy, and vagueness.
Wireless Sensor Network Networks of miniature sensor-enabled computing devices connected using short-range wireless communication.

Fundamentals

The study of geomatics has its roots in the disciplines of land surveying and mapping. As related topics, such as GIS and visualization, have grown and converged, the use of the term 'geomatics' has grown correspondingly, reflecting the increased breadth of topics under investigation. The term geomatics first emerged in the early 1970s as the French term 'géomatique'. The similarity between the French and English translation made it especially popular in Canada, the country first to adopt the term. The term has now gained widespread acceptance and recognition, and is found in the names of university departments and degrees, industrial companies, and international standards.

Numerous different definitions of geomatics have been proposed. For example, the Oxford English Dictionary (OED) defines geomatics as:

> The mathematics of the earth; spec. the science of the collection, analysis, and interpretation of data, esp. instrumental data, relating to the earth's surface (OED, 2002.)

A common theme to most existing definitions is that geomatics encompasses all scientific studies of 'geoinformation' (information that is referenced some way to locations on the Earth). A definition of geomatics that highlights this fundamental theme, synthesizing elements of many of the existing definitions, can be stated as follows:

> Geomatics is the scientific study of the structure, properties, acquisition, analysis, distribution, communication, and use of geoinformation.

What distinguishes geomatics from the individual sciences that comprise it, is that geomatics is also concerned with the integrated processes through which geographic data is collected, transformed into information, disseminated, and manipulated to meet user requirements within specific application domains. In short, geomatics is concerned with the interaction and convergence of traditionally separate disciplines: geomatics is more than the sum of its components.

Although geomatics is an area of scientific investigation, its growth and development is inextricably linked with the growth and development of enabling technologies, in particular global positioning systems (GPSs), geographic information systems (GISs), mobile computing devices (e.g., personal digital assistants, PDAs), and the Internet. These enabling technologies are not only encouraging the growth of a broad range of geomatics applications and services, such as in car navigation systems and location based services (LBS) but also their rapid uptake by society. Consequently, many people today routinely use geomatics without being aware. Geomatics has evolved to not only encompass a broad spectrum of fundamental concepts, but also to include the scientific exploration of the processes through which geoinformation is captured, used, and transformed in these integrated applications.

Many of the fundamental concepts in geomatics are historically based around those of land surveying and mapping. These include coordinate reference frames, geodetic datums and map projections, measurement adjustment, and error analysis. As geomatics has grown to encompass not just the acquisition but also the use of geoinformation, so the fundamentals upon which geo matics is based has expanded to include those of spatial cognition, scientific visualization, and human–computer interaction, as well as more traditional cartography and map design. Bridging the gap between spatial data ac quisition and spatial data use, other fundamental con cepts in geomatics are drawn from geographic information science and spatial computation, including spatial data structures, spatial databases, spatial analysis, and web based and distributed mapping. **Figure 1** sum marizes the relationship between geomatics, its defin ition, and its fundamental concepts.

Many systems in common use in society now inte grate across the spectrum of geoinformation acquisition, analysis, and application. For example, an in car navi gation system must integrate information from a number of positioning systems, such as GPS and inertial pos itioning, to generate real time information about the location of a vehicle. This location information must then be analyzed and combined with geoinformation about the road network to generate routing information, using a spatial database or GIS. Finally, the combined infor mation must be displayed to the vehicle driver in a concise form that supports his or her spatial decision making without distracting attention from the hazardous task of operating a vehicle.

Crucially, the different stages of this integrated process may constrain and influence one another. In the in car navigation example, the quantitative coordinate information produced by the positioning system can present problems to the user of that information. For humans, qualitative representations of geographic space are often more usable than precise quantitative repre sentations (e.g., 'turn right' instead of 'turn on a bearing of 84°'). Consequently, the in car navigation system must encode mechanisms for converting between quantitative coordinate systems and salient qualitative geoinforma tion. Indeed, geomatics is not restricted to measurement based representations of space. The scientific examin ation of the structure, properties, and use of purely qualitative geoinformation, such as proximity (e.g., 'nearness'), are also important topics within geomatics.

Integrative Topics in Geomatics

As discussed previously, geomatics concerns the inte grated processes by which geoinformation is collected, transformed, disseminated, and used. A variety of topics emerge directly from this integrated process, including ubiquitous positioning systems, digital cadastral data bases, location aware computing, and wireless geosensor networks. In addition, long standing topics, such as lo cation privacy and uncertainty, have gained new rele vance and urgency as a consequence of this integration.

Ubiquitous Positioning Systems

A ubiquitous positioning system aims to provide pos itioning information everywhere that is required by a specific application. While many sophisticated pos itioning technologies currently exist, for many appli cations no single solution is independently able to meet

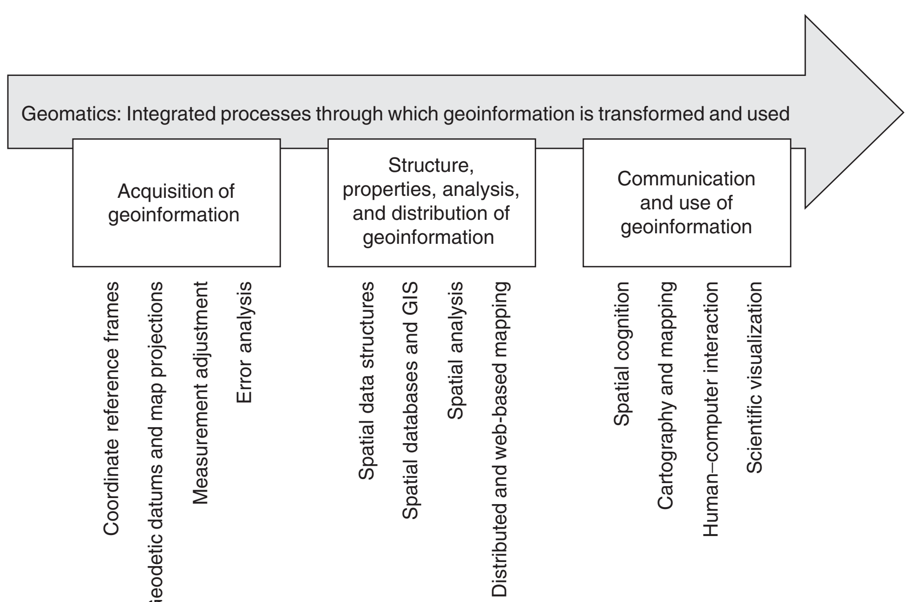

Figure 1 Geomatics as the integrated process of geoinformation transformation and use, and the fundamental topics and concepts that comprise geomatics.

all these requirements. For example, in outdoor positioning domains where there is a clear view of the sky, GPS can offer high precision and accuracy positioning at a global scale. The availability of additional satellite infrastructures, such as the European Galileo and Russian Global Navigation Satellite System (GLONASS), is helping to extend the coverage and robustness of GPS based positioning. However, the physical properties of GPS signals means that signal attenuation and multipath errors inevitably limit the reliability and availability of GPS in locations where the sky is partially or completely obscured, as in urban canyons or indoor environments.

Other technologies that do not suffer from these limitations do exist. However, each of these possesses its own inherent weaknesses. In populated areas with cellular phone network coverage, the cell ID of mobile phones and other devices provides positioning information, albeit at low levels of precision (typically to within 200 m at best). Enhanced cell phone networks can provide greater precision, but cannot offer as extensive coverage as conventional cell phone networks. Inertial navigation systems (INS) suffer from accumulations of errors over time, making them unreliable if used in isolation. Although some studies have shown that positioning using wireless fidelity (WiFi) and other short range wireless radio frequency networks can match or even outperform the accuracy and precision of GPS in urban environments, these techniques suffer from limited and inhomogeneous coverage. For many applications, the ability to provide a seamless positioning capability between outdoor and indoor environments is a problem that can only be solved through the integration of individual technologies.

Uncertainty

Uncertainty in geoinformation is a topic that cuts across the entire spectrum of different domains within geomatics. The measurement sciences, including geodesy, surveying, and photogrammetry, have developed sophisticated stochastic techniques for reducing uncertainty through the combination of different information sources. Storing and managing the uncertainty associated with geographic information has also been extensively studied in GIS. Similarly, cartographers have studied the communication of that uncertainty to users using maps and computer visualization.

Uncertainties can be introduced at every stage of the integrated process of data collection, transformation, manipulation, and dissemination. As a result, understanding and allowing for uncertainty in geomatics applications is especially important. Three fundamental types of uncertainty are commonly identified in geomatics: inaccuracy (a lack of correctness in information); imprecision (a lack of detail in information); and vagueness (the existence of boundary cases in information). Vagueness in geoinformation is exemplified by geographic concepts and relations like 'mountain', 'forest', 'bay', and 'near'. Many of the approaches to representing and computing with uncertainty in geomatics are statistical (e.g., using standard deviation to measure imprecision in the location of a point). However, nonstatistical approaches, such as fuzzy sets, rough sets, and other nonclassical logics, are also widely used, particularly in the representation and analysis of vagueness in geographic information (**Figure 2**).

Another topic closely aligned with uncertainty in geomatics is that of quality control (QC). This is particularly important given the pervasiveness of geomatics technologies across society. For example, users of GPS rely on the positions determined by low cost receivers for a number of leisure applications. Naïve users rely on these solutions and often have no method for determining their validity. This is even more significant for applications that are safety critical such as aircraft landing systems. QC relies on the ability of the measurement system to not only maintain the accuracy and precision of the solution but also the integrity, that is, the ability of the system to inform a user of a potential failure before it occurs. Integrity is enhanced through the

Figure 2 The fundamental types of uncertainty are inaccuracy, imprecision, and vagueness. The statement 'the city of Exeter is located at point b' is inaccurate: in actuality the location of Exeter is more correctly represented by point 'a'. The statement 'the city of Exeter is located in region R' (the county of Devon) is accurate, but less precise giving less information about exactly where Exeter is located. The statement 'the city of Exeter is located in the south of England' is accurate, but vague since the 'south of England' has no precisely defined boundary. All vague information is also imprecise, but imprecise information is not necessarily vague (e.g., Devon has precisely defined boundaries, so 'Exeter is in Devon' is imprecise but not vague).

availability of redundant measurements which can be increased through the robust integration of technologies.

Digital Cadastral Databases

Land administration systems fundamentally rely on the identification of land parcels to support legal, tradeable property rights. Identification relies on surveying techniques to determine the boundaries of parcels on the ground, in buildings, underground, or even in marine environments. When the parcels are systematically described in a reliable parcel map, a cadastre is the result. Digitization of the cadastre results in a digital cadastral database (DCDB). The key characteristics of a modern DCDB include:

- the large scales of its data (typically 1:500–1:1000 in urban areas with 1:2000–1:10 000 in rural areas) covering very small areas of land, such as street blocks or local government areas;
- the seamless horizontal or 2 D merger of the areas to reflect actual urban and rural land uses of a country, road networks, and other built features; and
- a data collection process which relies on rigorous methods to measure and mark the earth used by surveyors who establish the legal boundaries of parcels.

A DCDB is vital for modern land administration and sustainable land use. DCDBs form the major land information component in a country's spatial data infrastructure (SDI). The digital capture, storage, and use of cadastral information is therefore reliant on the integrated technologies and processes within geomatics. DCDBs must be integrated with other mapping and topographic information held in GIS, and with updates to that information, for example, resulting from subdivision or resurvey of parcels. Information about land rights, land values, and land uses held in a DCDB must be accessible in a form that nontechnical users, from land administrators to the general public, can comprehend. In the future, the representation and manipulation of parcels in three dimensions (width, depth, and height) will be critical to the development of the next generation of '3 D cadastre', suitable for management of modern cities.

Location-Aware Computing

The integration of portable, real time positioning systems (such as GPS) and low cost, powerful mobile computing technology (such as PDAs) has led to the development of location aware computing. Location aware computing systems use a mobile user's current position to provide more relevant information services to that user. The primary application for location aware computing is LBS, such as navigation systems and location based games.

Location aware computing systems must overcome many technical challenges, for example, performing the complex calculations necessary for positioning and geoinformation processing using low power, limited capacity mobile computing devices. But, in addition to the technical challenges, location aware computing has exposed many new and fundamental scientific problems in geomatics connected with the integration of positioning and mobile computing.

For example, understanding and allowing for uncertainty in geoinformation is critical in location aware computing. Unlike conventional GIS, where users remotely manipulate information, users of location aware computing systems are typically interacting with the geographic environment about which they are receiving information. Uncertainties of positioning systems (such as inaccurate locations), in the stored data (such as imprecise map data), and in the user's interaction with the system (such as in a vague query for 'nearby' points of interest) can result in location based services being delivered to the user that are clearly contradicted by the user's own interaction with the geographic environment.

The interaction itself also demands new ways of presenting information to the user. Conventional map based visual interfaces are often inappropriate in location aware computing environments, where small screens and low power devices place severe limitations on conventional cartographic techniques for representing geographic space. In some circumstances, such visual interfaces may even be hazardous (for example, where map based interfaces distract a driver's visual attention away from the task of avoiding collisions). Consequently, investigating new ways to communicate salient geoinformation to location based service users is an important focus for geomatics research.

Developing new ways of communicating salient geoinformation to human users demands an understanding of the entire process of capture, transformation, and use. Textual, iconic, and verbal communication modes are often more suited to location aware computing environments than traditional visual or cartographic communication modes. Qualitative and vague spatial concepts (like 'north' and 'near') are common in human textual and verbal communication. Location based services interfaces increasingly aim to be more usable by presenting the user with information that is 'cognitively ergonomic' in the sense that it is closer to human spatial concepts and so requires less cognitive effort to process. As already noted, such aims often necessitate the definition of processes for transforming the quantitative coordinate information generated by most positioning systems into qualitative geoinformation about a user's relative location in geographic space.

Location Privacy

Location privacy concerns the right of an individual to determine to what extent information about the location of that individual is communicated to others. The inte gration of positioning systems with mobile computing devices enables frequently updated, high precision, and accurate information about the location of an individual to be generated and processed. The development of wireless communication technology and distributed, web based GIS architectures means that personal infor mation about an individual's location can be disseminated more rapidly and widely than ever before possible. Understanding and protecting an individual's location privacy is therefore of critical importance in society, as location area computing systems are becoming more and more familiar.

The topic of location privacy is very broad and en compasses many aspects that extend beyond the domain of geomatics, such as the legal status of location privacy or the extent to which societal attitudes to location privacy are changing. However, there are still ways in which geomatics is contributing to the science of location privacy. For example, one approach to the problem of protecting an individual's location privacy is to delib erately degrade the precision and accuracy of location information, termed 'obfuscation' or 'spatial cloaking'. The individual can control how much information about their location they choose to reveal in order to receive some useful information service. **Figure 3** gives a simple example of a situation where an individual can receive high quality information about the nearest point of interest based on low quality information about her lo cation. The process of degrading real time location in formation and making the best use of low quality information about a user's location presents some im portant theoretical and technical challenges for geomatics.

Wireless Sensor Networks

Advances in sensor technology, coupled with the ad vances in positioning technologies, mobile computing, and wireless communication discussed above, have led to the development of miniaturized, sensor enabled com puter systems. A wireless sensor network (WSN) is a collection of these miniature sensor enabled computing devices connected using short range wireless communi cation network. WSNs used to monitor information in geographic space are sometimes referred to as 'geosensor networks'.

Although technological advances have made WSNs possible, these technological advances expose many new and fundamental scientific challenges. Plans for dense WSNs with thousands or millions of nodes will provide geoinformation at much greater spatial and temporal granularity than has ever previously been possible. Making sense of this overwhelming volume of spatial data is motivating the development of new information processing paradigms. For example, the huge volume of readily accessible real time data generated by WSNs means that storing all this data becomes impractical. One approach to this problem is 'stream processing', where

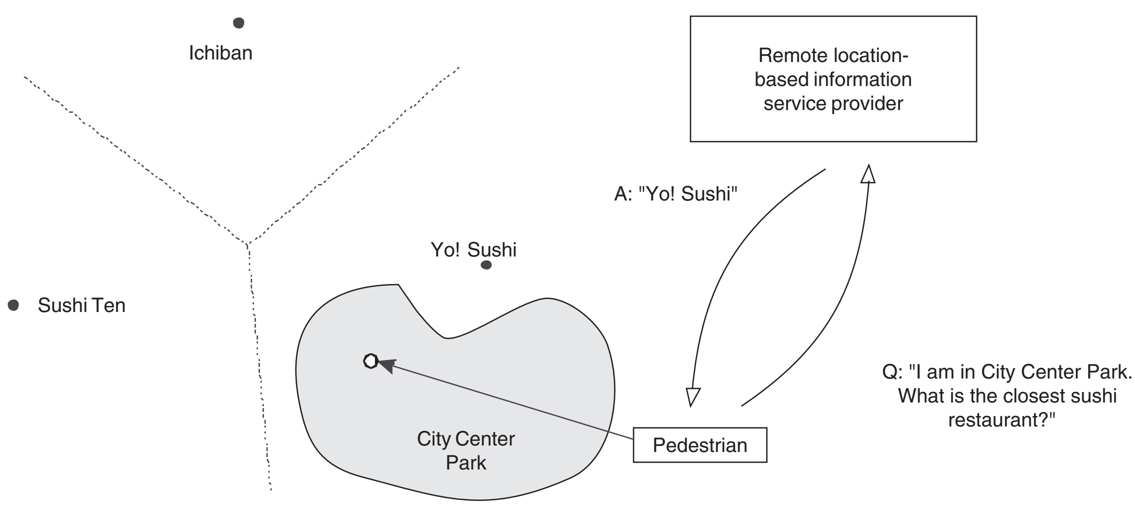

Figure 3 To protect her location privacy, a pedestrian might deliberately degrade the quality of information about her location generated by her location-aware computing device, termed 'obfuscation'. In the example above, the pedestrian has degraded her personal location information from precise GPS coordinates to the lower-precision information that she is in the 'City Center Park'. However, such low-quality positional information can still potentially be used to receive high-quality information services, such as finding the closest sushi restaurant from a remote location-based service provider. Modified from Duckham, M. and Kulik, L. (2006). Location privacy and location-aware computing. In Drummond, J., Billen, R., Joao, E. & Forrest, D. (eds.) *Dynamic & mobile & GIS: Investigating change in space and time*, pp 35 51. Boca Raton, FL: CRC Press.

information generated is automatically searched for sali ent patterns and summarized. Only the salient patterns are stored and reported to users, while the raw data is simply discarded. Another approach is to perform partial processing and aggregation of information within the WSN. Thus, each node in the network locally processes the information generated by itself and its neighbors, before passing this partially processed information on to other nodes. Such new modes of spatial computation, with partial processing distributed throughout the network, are at odds with the traditional geoinformation processing paradigm in GIS, where information is collated within a single system before being processed.

Outlook

New technologies mean that society's access to geoin formation is growing across a broad range of applications. Correspondingly, the need for integrated processes to support these applications means that geomatics is also growing. The definition of geomatics provided at the beginning of this entry relates exclusively to geoinfor mation (information that is referenced in some way to locations on the Earth). However, we conclude by ac knowledging the increasing range of applications that apply the techniques within geomatics to spatial, but non geographic information. For example, the creation of spatially accurate virtual worlds found in many computer games, architectural systems, and archaeological re constructions, utilize many of the integrated techniques found in geomatics. Similarly, accurate 3 D measurement, modeling, and visualization of objects like teeth (used in forensic dentistry to identify criminals in violent crimes) and motor vehicles (used in crash analysis to determine the cause of a road accident) also rely on geomatics based techniques applied to non geographic spatial domains.

See also: Geovisualization; Integrated Spatial Data Infrastructure; Photogrammetry/Aerial Photography; Spatial Databases; Surveying.

Further Reading

Duckham, M. and Kulik, L. (2006). Location privacy and location aware computing. In Drummond, J., Billen, R., Joao, E. & Forrest, D. (eds.) *Dynamic & mobile GIS: Investigating change in space and time*, pp 35 51. Boca Raton, FL: CRC Press.

Economic Commission for Europe Geneva, Land Administration in the UNECE Region Development Trends and Main Principles. http://www.unece.org/env/documents/2005/wpla/ECE HBP 140 e.pdf (accessed on 18 July 2007).

Retscher, G. and Kealy, A. (2005). Ubiquitous positioning technologies for modern intelligent navigation systems. *Journal of Navigation* 59(1), 91 103.

Schiller, J. and Voisard, A. (eds.) (2004). *Location based services*. San Francisco: Morgan Kaufmann.

Shi, W., Fisher, P. F. and Goodchild, M. (eds.) (2002). *Spatial data quality*. Boca Raton, FL: CRC Press.

Stefanidis, A. and Nittel, S. (eds.) (2004). *GeoSensor networks*. Boca Raton, FL: CRC Press.

Stoter, J. E. and van Oosterom, P. (2006). *3D Cadastre in an international context: Legal, organizational, and technological aspects*. Boca Raton, FL: CRC Press.

Wallace, J. and Williamson, I. (2006). Building land markets. *Journal of Land Use Policy* 23, 123 135.

Williamson, I. P., Rajabifard, A. and Feeney, M A. E. F. (eds.) (2003). *Developing spatial data infrastructures: From concept to reality*. London: Taylor & Francis.

Worboys, M. F. (2001). Nearness relations in environmental space. *International Journal of Geographical Information Science* 15(7), 633 651.

Worboys, M. F. and Duckham, M. (2004). *GIS: A computing perspective* (2nd edn). Boca Raton, FL: CRC Press.

Zhao, F. and Guibas, L. (2004). *Wireless sensor networks: An information processing approach*. San Francisco: Morgan Kaufmann.

Relevant Websites

http://www.cig acsg.ca
Canadian Institute of Geomatics.

http://www.fig.net
Cadastre 2014, FIG.

http://www.gps.gov
Global Positioning, System Serving the World.

http://tycho.usno.navy.mil/
GPS, Current GPS Constellation, US Naval Observatory.

http://www.esa.int
Navigation, ESA.

Geopolitics

P. Reuber, University of Münster, Münster, Germany

Glossary

Critical Geopolitics Rooted in post-structuralism, critical geopolitical inquiry is, at its core, concerned with the operation, interaction, and contestation of geopolitical discourses.

Discourse The setting or domain in which words are used and take on specific meanings.

Geopolitics The study of the impact of geographical distributions and divisions on the conduct of world politics.

Nation-State A state in which the territorial boundaries also enclose a claimed nation which the state represents.

National Socialism German political movement initiated in 1920 with the organization of the National Socialist German Workers' Party. The movement culminated in the establishment of the Third Reich, the totalitarian German state, led by the dictator Adolf Hitler from 1933 to 1945.

Post-Structuralism Influenced by Heidegger and Nietzsche, French philosophers such as Derrida and Foucault suggest that all text has ambiguity and because of this the possibility of a final and complete interpretation is impossible. This means first that language and society are shaped by rule-governed systems and second that it is impossible to step outside of discourse and survey the situation objectively.

Introduction

Geopolitics – that is nothing 'real'. Nothing 'real' from an essentialistic point of view. Geopolitics is a social con struct, a sort of social practice, a discursive formation depending on what theoretical perspective you take on this phenomenon. No wonder that there is no solid core of knowledge in this respect, geopolitics seems to be fluid, runs through your fingers.

As intangible as it might be, a significant impact it still has. This phenomenon is neither harmless nor innocent. It is highly politically relevant. Wars are led in the name of geopolitics. People are driven out and murdered to serve geopolitics. It is a clear proof for how discursive constructions turn into social practice, how represen tations and imaginations grounded in discourse turn into powerful instruments of political actions.

What exactly is geopolitics, though? What is so fas cinating about it? Why does it have such an impact? In order to answer such questions the line of argumentation follows two consecutive sections:

In the first section, 'Geopolitics as a discursive con struction – some theoretical remarks' the nature of geo politics is considered from a conceptual perspective. The goal is to refine the nexus – created in terminology and practice of geopolitics – between society, space, and power; the nexus which is not essentially preexistent but that forms a fragment of the hegemonic discourses of modernity and has a powerful impact on social practice. These details are considered in subsections 'The nexus between society and space as a basic element of the geopolitical imagination' 'The nation state as the fundamental "territorial trap" of the geopolitical im agination', and 'Post structuralism and discourse analy sis as a basis for the deconstruction of geopolitical representations'.

Against this background the next section, 'The tidal waves of discourse: geopolitics and geopolitical repre sentations in different historical periods', presents ex amples from the history and development of geopolitical discourses to show how and based on what historically situated framings geopolitical representations develop and grow their specific impact. The historical comparison shows how little the geopolitical regionalizations have in common even when they frequently choose to justify their social legitimacy and authority with reference to an ultimate 'scientific objectivity'. From the beginning of the last century, when the term 'geopolitics' itself was first used by the Swedish political scientist Rudolf Kjellén, the range of explanations widely extends. The subsections 'Geopolitics as a discourse in the historical context of biological determinism and geodeterminism at the end of the nineteenth century', 'Karl Haushofer and geopolitics in national socialism', 'Cold War geopolitics', 'The pluralization of geopolitical representation after the end of the Cold War', and 'The new spatial organization of power in geopolitical representations after 9/11' provide an overview of the subject.

Geopolitics as a Discursive Construction–Some Theoretical Remarks

From a political–geographical perspective, geopolitics can be seen as all those elements of discourse and/or social practice that deal with the construction of large scale regional up to global worldviews. The expression geopolitics is semantically plurivalent and according to theoretical perspective comprises:

1. geopolitical discourses as representations of space and power;
2. geo policymaking of the political actors;
3. geo policymaking of scientists, consultants (in inter national relations (IR), political geography, 'geos trategic analysts' in scientific think tanks, etc.), and the media; and
4. the scientific deconstruction of geopolitics and geopolitical imaginations from a post structuralist perspective.

These four aspects, which overlap and unify in the word geopolitics, show not only the broad range of contents that is represented by this expression, but also points out the different social groups that are engaged in the field of geopolitics – from politics over political advisors and the media to scientific consultants and critical scientists.

This heterogeneity also expresses in the various the oretical approaches that form the conceptual basis of geopolitics. These approaches partly vary to such a great extent that they lead to diametrically opposed forms of scientific analysis, practical recommendations, and pol itical positioning. What scientists publish under the label of geopolitics today reaches from latently geodeterminist over positivist and structuralist to post structuralist approaches.

This variety of concepts has made geopolitics a topic in various scientific disciplines (mainly in the fields of IR and political geography). Due to this variety geopolitics has quite heterogeneous options of political influence – according to the conceptual orientation. Considering the historical development these options reach from phrasing hawky, demagogic Darwinian policy of living space (Lebensraum) at the beginning and middle of the last century to nowadays' emancipated criticism of the pol itics of geopolitics.

The Nexus between Society and Space as a Basic Motive of the Geopolitical Imagination

From a theoretical point of view, geopolitics is the dis cursive coupling of social and spatial categories. The decisive point in linking social and territorially defined elements lies in the resulting 'purification of space'. By this kind of discursive practice the 'self' and the 'other' is framed along spatial categories and provided with bor ders. This results in the construction of culturally/so cially homogeneous territorial groups. Such constru ctions can be found as geopolitical representations in all historical epochs (see below). As much as their contents may vary according to the hegemonic discourses of their time, (e.g., if they exploit expressions like race, class, or cultural diversity when constructing putative homogeneous spaces), as uniform their decisive mode of distinction is in the construction of homogenous social spaces that separate from the other.

In this way, the politics of geopolitics is a structuring element in social negotiation processes and conflicts. Geopolitical imaginations offer the option to society to find something like orientation, straightforwardness, and assumed safety in the contingent diversity of being. They are historically alterable narratives and – apart from stable phases in which they may appear to be virtually objective formations – are subject to constant negotiation and renegotiation. From the geographical perspective space and place are of crucial importance in the con struction of geopolitical representations.

The Nation-State as the Fundamental 'Territorial Trap' of the Geopolitical Imagination

Concerning this matter in the epoch of modernity, the nation state advanced to be the territorial masterframe for countless political strategies of containment and ex clusion. The nation state was the basis of Friedrich Ratzel's biologistic *Staatsorganizismus*. He represented the territorial container of the nation state as the putative logical outcome of a corresponding development of human history. Even though Benedict Anderson's long standing conceptualization of nations as 'imagined com munities' has once more characterized them as social constructions, the nation state order has become such a persistently successful framing that it is taken for granted as a kind of 'realistic' entity in the discourses on identity over the last two centuries. Even in the times of increasing globalization its effect appears to hold its ground – despite the undeniable and partly grave power loss to the transnational networks of 'global governance'.

John Agnew calls this stiff fixation on the space con tainer of the nation state a 'territorial trap'. This ex pression characterizes the unreflected taking over of the concept of nation states, or more precisely, a virtual ontologization of the territorial principle as a decisive framing of social, especially political structuring. Ac cording to Agnew, the territorial trap is based on three main assumptions that must be analytically separated but overlap in social practice. The territorial 'logic' of mod ernity expresses in the fact that

1. the sovereignty of the modern state requires defined regional authorities;
2. the internal and foreign policy are strictly separated; and
3. the territorial state serves as the spatial container of national societies.

Based on these assumptions the geopolitical order ap pears as a patchwork of separated, secluded, and in dependent entities, implicitly marginalizing the fact that all individual elements of such a territorial structure are socially constructed and – when looking at the longe durée of historic eras – are constantly renegotiated.

Post-Structuralism and Discourse Analysis as a Basis for the Deconstruction of Geopolitical Representations

Though scientific methods of analyzing geopolitical representations are discussed in further detail in other contributions of this handbook (compare keywords such as critical geopolitics, political geography, post colonialism, geographical imaginations, postfeminism, etc.), it may be helpful for understanding the following reconstructions of historic geopolitical narratives to mention at least some core aspects and basic principles out of the wider field of theoretical and methodological reflections concerning this matter.

Most of the theoretical framework can be found in post structuralist theories, especially in discourse analysis and semiotic analysis. The deconstruction of geopolitical representations in particular works against the reemergence of strongly traditional, sometimes practically natural determinist ideas, for which geopolitics is the doctrine of the influence of the physical space on the politics of a state. Post structuralist reasoning in political geography show how such a conceptualization of geopolitics leads to an ideological as well as an epistemological *cul de sac*. Post structuralist approaches are characterized by a general shift of perspective. They take the geopolitical discourses themselves as the object of their analysis. The issues to address from this point of view are: (1) the construction of discursive and cartographic representations and (2) how these representations intersect with political and social imaginations.

Beyond such lines of research more strictly oriented to post structuralist discourse analysis, we can find a broad range of approaches that consider geopolitical narratives and imaginations more as a discursive strategy of intellectuals of statecraft serving toward the legitimization and achievement of their aims (critical geopolitics).

Both approaches follow a constructivist perspective and a linguistic (respectively semiotic) turn which regards geopolitical representations as constructions using geographical semantics, metaphors, symbols, etc. Such a perspective sharpens the awareness of the normative character of competing imaginations. The greater the extent to which their relativity and their subtle role become apparent, the less they can fulfill their occasionally problematic role in the geopolitical debate on power and space – both in political discourse itself and in the polarization and instrumentalization of public opinion.

The Tidal Waves of Discourse: Geopolitics and Geopolitical Representations in Different Historical Periods

The following analysis of selected geopolitical representations follows a more post structuralist approach which directs its interest toward the discursive production of space. Echoing Said's *Orientalism* and Gregory's *Geographical imaginations* geopolitical discourses are regarded in Foucault's sense as linguistic formations of power and space.

Against this background, it becomes apparent how much geopolitical imaginations and representations are situated within their social and/or discursive context. This means that geopolitics is subject to change that – from a discourse theoretic perspective – takes place along the slow oscillating of hegemonic discursive formations. In order to show this, some basic elements of the geopolitical framings of different historical eras are shortly and exemplarily discussed in the following. Even in this shortened form it becomes clear to what extent geopolitical representations – from Ratzel's bio and geodeterminism to Huntington's cultural deterministic *Clashing Civilizations* – are interwoven with the hegemonic discourses of their times in order to construct the corresponding historical specific context of society, space, and power. To illustrate these correlations the following phases (periods/eras) will be discussed:

1. discursive establishment of the sciences of political geography and geopolitics in the historical context of biological determinism and geodeterminism at the end of the nineteenth century;
2. political use of the (scientific) geopolitical discourse in context with imperialism, colonialism, nationalism, and national socialism;
3. discursive framing of the foreign policy and IR in the post war phase (Cold War); and
4. development of new geopolitical representations after the end of the Cold War.

Geopolitics as a Discourse in the Historical Context of Biological Determinism and Geodeterminism at the End of the Nineteenth Century

Generally, it was argued that geopolitical representations have always been part of the hegemonic discourse of their time. Nevertheless, the kind of geopolitics that developed at the end of the nineteenth century is special in this regard for two reasons: first, this period is characterized by powerful discursive framings of territorial order. This expressed in political everyday life where the concept of the nation state emerged to be one of the most important structuring principles of society. The nation state – following Foucault's considerations about the development of governmentality – had successively formed to be a crucial element in replacing the feudalistically organized society by a civil society. In the following period of a state centered imperialism and colonialism, a geopolitics of nation states and their global areas of influence established.

Second, this period lies in a phase in which naturalist and biologistic narratives in European societies for the first time in history widely diffused into the discourse of the whole society. At the same time they worked as key concepts in science and thus greatly influenced newly developing disciplines, for example, sociology and – as the science of space – geography.

For the development of geopolitical framings it was essential that spatial and biologistic fragments of discourse began to interact. Some examples can demonstrate this. As Schultz states geopolitics is founded on the traditional natural geographical concept, which, for example, is reflected in Herder's construction of nations as largely influenced by climatic geography. "Nature produces families; the most natural state therefore is a people with a national character. … A people is a plant of nature as well as a family, but one with many branches". (Herder cited by Schultz, 1998: 87)

Ritter, one of the founders of modern geography, shows a similar understanding in combining fragments of spatial and naturalistic discursive fragments. In general, ideas concerning the power of space began to enter man's conception of the world at the end of the nineteenth century. As Windelbrand so aptly sums it up in 1809: "No matter what part of the earth we observe, we find religion fighting against religion, denomination against denomination, race against race, people against people in a tougher conflict than ever before. As the last century rationalised, ours wants to nationalize."

It is not surprising that such a construction of geopolitics in the political contexts of its time fell on fertile ground. As a scientific support to imperialism geopolitics gained sovereign tasks rather quickly. Consequently, the global space was segmented, conquered, and measured. Quantifying the earth according to exact space measures such as circles of latitude and degrees of longitude was only possible from the imagination of the world "... as a differentiated, integrated, hierarchically ordered whole" (Gregory, 1998). Geopolitical constructions served this imagination by dichotomizing along and with the help of spatial demarcations. The developing parts of the world became the excluded Other: West against East, North against South, Orient against Occident, and sea power against land power – these are only a small part of the broad range of such dichotomizations.

Some of these binary imaginations of the nineteenth century achieved prominence that partly can still be found in geopolitical discourses today. For example, the construction of an antagonism between sea and land, between sea power and land power, formed the basis for Mackinder's (1861–1947) model that becomes an important part of Anglo American geopolitics (**Figure 1**). In 1904, he presented a dichotomy of power to the Royal Geographical Society that separated the world in states of sea power and land power and thus formed the basis of many geopolitical imaginations still valid today. In this case, Russia is seen as the power center of the continental empire – the heartland ('pivot area') – without sufficient waterways to the ocean. All around it there is a seam of areas that do have access to the seas. They stand out due to their conflict causing middle positions and are subject to both oceanic and continental impacts. Around this seam the rest of the world is arranged being purely under oceanic influence (rimland). Among these we find Japan, Great Britain, and the United States. The 'balance of power' within this dual world system is the main drive of any development in the field of foreign policy. The conflicts are mainly carried out in the rupture zones. That is why Mackinder considers Eurasia to be the pivot of world politics. For him, controlling this area means reigning the world. From the perspective of the sea powers he sees the danger for them to be reigned by the land power center sooner or later. He confronted the sea based British Empire with this horror scenario. The success of his representation is therefore mainly due to the construction of a discourse of threat and security understandable for the British Empire.

Nicholas Spykman (1893–1943) used this threat and security scenario for the American administration in the run up to World War II. In contrast to Mackinder's attention to the heartland Spykman emphasized the rimlands. In order to prevent predominance of an individual power in the area of the rimlands he thought it was necessary to pursue an interventionist foreign policy. Otherwise the security of the USA would be at stake. This imagination was to become a central element of the geopolitical discourse during the Cold War.

The geopolitical representations in the field of German geopolitics in the late 19th cnetury were completely different. They were based on the determinist concept of an exceptional geographical position of the German Reich. This thinking led Sombart to unanimous rejection of anything that was somehow close to English or Western European thinking or feeling. The aversion expressed here was directed toward liberalism, positivism, and subjectivism. These concepts were opposed by the ideas of community, entireness, and organism. In this nationalist oriented climate the German protagonists of geopolitics – Friedrich Ratzel and Karl Haushofer – developed their geopolitical representations. They were so successful because they fit to the prevailing discourse, and so the way from their imaginations to Hitler's geopolitics of living space (Lebensraum) was not very far.

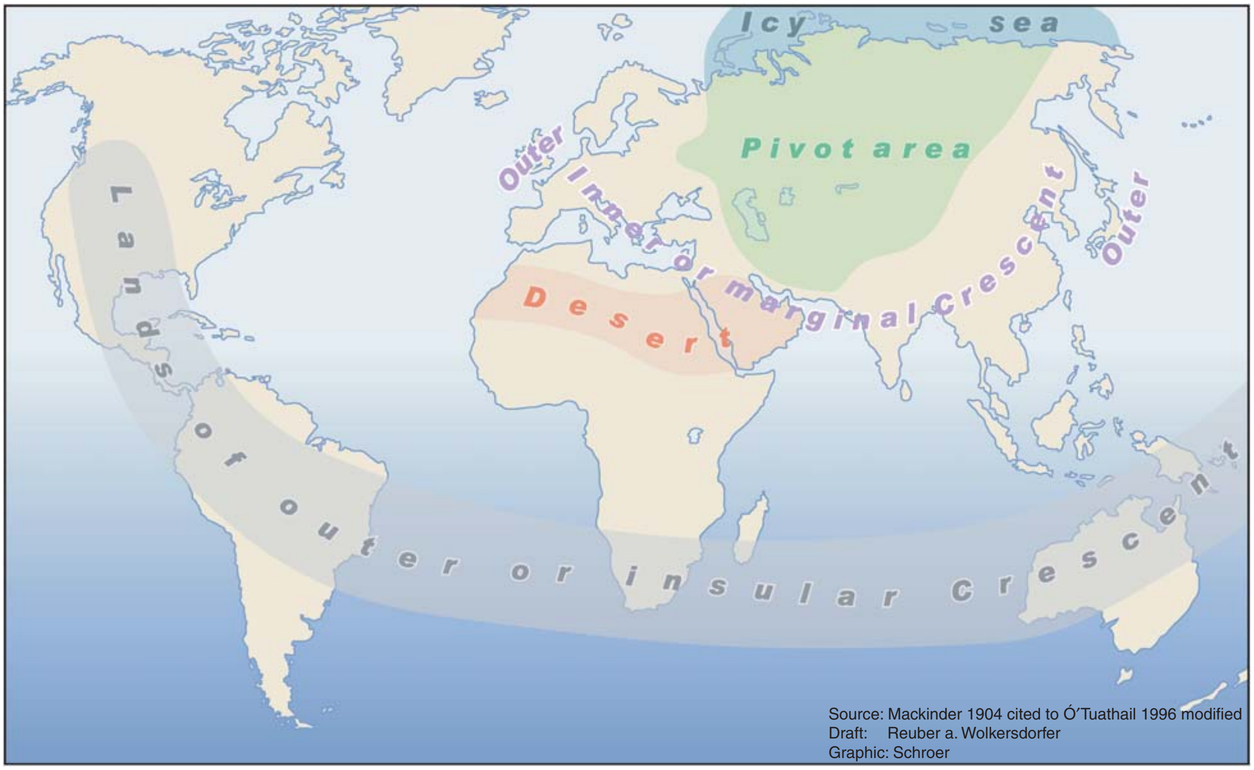

Figure 1 Geopolitical representations of intellectuals of statecraft: Mackinder's natural seats of power. Modified from Reuber & Wolkersdorfer, 2004.

The State as an Organism – Ratzel's Biodeterministic Construction of German Geopolitics

The basic element of Ratzel's mode of construction consisted of the nation states' dependency on nature dynamic, namely, the biological plan of natural selection. Ratzel (1844–1904), who had moved from zoology to geography in his scientific career, built his political geography entirely on a positivistic natural scientific conception of the world. His construction of the relationship between politics and space is determined by the physical basis of the nation state. Societies, culture, and economics and their effects on the nation state are hardly dealt with in his discussions. Another crucial point in his construction is his biological analogy of the nation state as an 'indigenous organism'. The nation state in this respect is attributed with the traits of a living being, an organism that only displays health and strength when it is capable of growth, in other words, territorial expansion. This form of geopolitics according to Darwin's hypothesis legitimates every form of imperialism and expansionism. Following Ratzel, as a result of a growing and healthy population, a nation state needs more space to continue the development of its civilization. In this manner, a struggle for more space automatically develops between nation states.

Repeatedly Ratzel makes a close connection between the growing population and the growing space and puts this claim into practice regarding the German Reich in the run up to World War I:

> Wherever you look space is won and space is lost. Set back and progress everywhere; there will always be reigning and serving peoples. Even the peoples must be ambos and hammer. In no case Germany must limit itself to Europe; as a world power among world powers it can only hope to protect the ground for its people that it needs to grow. (translated from Ratzel, 1906: 377)

That is how Ratzel delivered with arguments of scientific repute the geopolitical basis for the colonial policy and fleetpolitics of the German Empire. Even after World War I the connection of politics and scientific geopolitical representations carried out here led to a fast expansion of political geography at the universities. With his discourse of the 'logic' of growing space Ratzel gave the essential impulse for the development of the ideology of living space (Lebensraum).

> Thus Ratzel's theory could not only be connected to the classic concept of geography but also to the ideology of living space of the Third Reich. The reorientation to race as the decisive power in history was already started there." (translated from Schultz, 1998: 217)

The term 'geopolitics' itself was first used by the Swedish political scientist Rudolf Kjellén. For him, geopolitics implied teaching about the state as a geographical organism or as an appearance in space. His geopolitics builds upon Ratzel's ideas and forms the basis for establishing geopolitical discourses both in science and politics.

Karl Haushofer and Geopolitics in National Socialism

Karl Haushofer (1869–1946) is of particular interest because in his person, as well as in the history of his family, the discursive significance of the geopolitical concept can be clearly portrayed. In this respect, Haushofer and his family are of interest within the line of argumentation in this article because of the effect on the development of the geopolitical discourse in Germany as well as in the USA.

Karl Haushofer, who considered Ratzel his mentor, shaped the German geopolitical discourse more than any other individual in that time. During World War I he was an officer, rising to the rank of major general. After release from the military after World War I, his career as a professor of geography at the University of Munich (1921–39) began. He published many books and papers, was editor of the *Zeitschrift für Geopolitik* (*Journal of Geopolitics*, 1924–44), and president of the German Academy. Haushofer's position in the Third Reich was shaped by his close relationship with Rudolf Hess and by the political standing of his sons Heinz and Albrecht in the Nazi regime. Through these contacts the geopolitical representations of Karl Haushofer had direct influence on the policy of the Third Reich and became eminently powerful. So in the years between World War I and II geopolitical representations achieved, at least in Germany, a hegemonic position in discourse.

For this development, Haushofer's chief pupil, Rudolf Hess, had special significance. Through the extensive contact between Hess and his mentor Haushofer, which among other things led to Haushofer hiding Hess for a few months at his place after the Putsch in Munich had failed, Haushofer's geopolitical imaginations became part of the Nazi ideology. During Hitler's and Hess' period of arrest in Landsberg, the ideological basis of national socialism originated with Haushofer's active participation: The myth of the Führer and the demand for more Lebensraum (living space) got into the writings in *Mein Kampf*. That Haushofer did not appear more visibly within the national socialist government was, according to many historians, a result of seeing himself as an eminence rise in the background.

The historical evaluation of Haushofer's role, and the surviving myth surrounding him, will surely continue. Interestingly enough, the allegorical figure of Haushofer offered an opposite picture in German and Anglo American historical reconstructions. Especially in American reconstructions, all geopolitics in the German speaking sphere seemed to be developed through Haushofer. As a consequence, part of the American geopolitics of the post war era originated from the myth around the Geopolitical Institute and its director, Haushofer. At least here, Haushofer was considered the 'scientific brain' behind Hitler. His discursive excesses even got as far as Hollywood, where in an American propaganda film Haushofer and his Geopolitical Institute are represented as the headquarters of national socialism. As Ó Tuathail shows in his *Critical Geopolitics*, Haushofer and his imaginary Geopolitical Institute in Munich were considered the beginning of the conviction that "... geopolitics was something that America had to know." It was assumed that there were hundreds of scientists occupied with forming strategic designs of national socialist expansion. When the Americans liberated Munich in 1945 their special mission was to locate this institute at the University of Munich. Reality remained well below the level of the visitors' expectations. Instead of the Geopolitical Institute, they only found a professor's office.

The American geopolitician Edmund A. Walsh, who was assigned to participate in the Nuremberg Trials, could not conceal his admiration for Haushofer and hence did not criticize the mode of geopolitical representation itself, but accused Haushofer of pursuing the wrong type of geopolitics. His demand for a 'true instead of a false' geopolitics locates the discipline within the democratic structure of Western society and can be explained at the level of modernizing basic theoretical assumptions.

Cold War Geopolitics

The years after 1945 were characterized by new forms of geopolitical representations that developed to a much greater extent than before in the scientific and political think tanks of the USA. Their goal was mainly a reframing of the American relation to the Eurasian continent. From the perspective of discourse theory, it is interesting to see to what extent these new representations rely on long standing geopolitical imaginations like Mackinder and Spykman. They incorporated the old argument of a 'natural' line of conflict between a Western sea power (now the USA) and a central land power (now the Russian Federation). The practical consequence of such a geopolitical representation was the assumption that the political situation after 1945 did not allow for a stable order on the Eurasian continent without the permanent American presence. This was to prevent the expansion of the Soviet land power. The geopolitical construction of Mackinder's heartland theory is more than apparent here – it is the core element of the new geopolitical doctrine of the Cold War.

The prevailing geopolitical representation led to a clear change or realignment of American foreign policy. You can distinguish here:

1. the post war period with the Truman doctrine that brought the intervention decisions in favor of the outer margins of the Eurasian continent. They became apparent, for example, in the American engagement in the conflicts in Germany, Turkey, and Vietnam and
2. the relaxation of tensions phase under Nixon/Kissinger with its policy of opening toward China as the beginning of a genuine American geopolitics of 'realism'.

The geopolitical imagination of an upcoming 'Cold War' was first formulated by the American journalist Walter Lippmann in 1947. The discursive construction of geopolitical enmity was charged with ideological opposites. Seen from the perspective of the USA it was a fight between liberty, democracy, and market economy on the one hand and communism, totalitarian dictatorship, and command economy on the other. From the perspective of the Soviet Union, the framing of good and bad went just the other way round.

On 12 March 1947 the American president Harry S. Truman announced his new geopolitical representation in the American congress. With its dualistic approach and its universalistic claim (liberty vs. totalitarism) it polarized even the parts of the world that so far had not been influenced by post war geopolitics. After the Truman doctrine it became a basic principle of the foreign policy of the USA to deliver active support to all peoples whose putative liberty seemed to be endangered by communist or socialist states or by inner, left wing, and militant minorities. The first examples were Turkey and Greece, but with hindsight also (Western) Germany. For decades the Truman doctrine became the discursive legitimation of the American containment policy and the basic leitmotif of the Cold War. A further milestone in this geopolitical representation was the domino theory that was announced by US president Eisenhower on 7 April 1954. It said that if one state falls, that is, it becomes communist, the neighbor states will also fall – like domino stones. Regarding, for example, the inner conflicts of Southeast Asia that meant: if South Vietnam becomes communist the neighboring states will become communist too. "Like apples in a barrel infected by one rotten one, the corruption of Greece would infect Iran and all to the east. It would also carry infection to Africa through Asia Minor and Egypt, and to Europe through Italy and France, already threatened by the strongest domestic Communist parties in Western Europe."

The politics of containment of the immediate post war years formed, however, no consistent structure of argumentation. It much more oscillated between different axes of argumentation. One scenario for the conflict considered the power struggle between the USA and the Soviet Union to be ended only at the level of international confrontation. The other argumentation assumed that sooner or later the Soviet system would erode internally so that in the political power struggles in Moscow those would win whose will for peace would lead to a quasi self containment of the Soviet expansionism. The perspective mentioned first implies that the own security and the security of the West as a whole depends completely on the USA, decisively and actively containing the Soviet Union mainly within the NATO alliance; the second perspective prefers solutions that give global security over to collective institutions like the United Nations in order to prevent history based fears of Soviet threat. These two argumentations might be quite different contentwise; at the level of geopolitical representation, however, they both support the underlying binary worldview of ideological bloc confrontation and the inherent blaming of the 'other'.

From a post structuralist perspective it can be said that the geopolitical representation of the Cold War was successful because it could be connected to from both ends.

> Historically ... the Cold War served the interests of both the USSR and the United States. For this reason neither sought to alter the nature of the relationship once it had been established. Their goal, therefore, was not so much victory over the other as the maintenance of balance. In this sense the Cold War was ... a carefully controlled game. (Cox 1990, cit. in Ó Tuathail/Dalby, 1998: S. 51f.)

The 1970s marked in many ways the turning point in international relations. Nixon's and Kissinger's geopolitical imaginations of the 'grand design' was the conceptual consequence of the developments since the beginning of the 1960s (mainly the war in Vietnam) which had not necessarily diminished the power of the US but had made its application in a more and more diversified world much more difficult. Kissinger called it the return to politics of realism and thus a policy closely oriented to the putative political facts and realities. Along with that the ideological discourses of American–Soviet power struggle were reduced whereas the geodeterminist framings were enhanced (referring to e.g., natural resources, economic potentials, and the geographical position).

The Pluralization of Geopolitical Representation after the End of the Cold War

With the end of the Cold War the geopolitical representation of a world of ideological blocs and iron curtains fell apart. Facing new processes of regional fragmentation within a global network society, this imagination faded away within just a couple of years. New patterns of power on a global scale demanded for new interpretations, new geopolitical representations to deliver a frame for the ongoing transformations, worldwide conflicts and wars of the 1990s, and the new millennium (see **Figure 2**). Against this backdrop it is not surprising that, already in the early 1990s, a series of new geopolitical narratives began to appear. All of them claimed to offer plausible discursive framings for changing global constellations of power. They did not only arise from the classical field of IR sciences (see below). Among them were, for example, geoecological narratives, which the politicians of Western industrial nations and transnational organizations some times used to frame neocolonial policy. However, those kinds of arguments still did not lead to a stand alone geopolitical regionalization on a global scale. This par ticular discourse rather goes along with established di chotomies with roots in colonial history and the debate on developing countries, that is, spatial dichotomies such as North versus South or developed versus less developed countries, etc. A similarly implicit kind of global geopolitics came at the beginning of the 1990s in the form of an effort to substitute 'geoeconomics' for geopolitics. According to Luttwak, political conflicts were more and more fought with economic weapons in this era of increasing (mainly economic) globalization. Visions of a network society, including new forms of global gov ernance, were challenging the traditional relationship of politics and space, especially the role of the nation state in international relations.

The deployment of such alternative discourses chan ged due to the conflicts and wars of the last decade. Since then new geopolitical framings developing during the beginning of the 1990s have been on an upswing in both politics and the media. To bring the recent master frames of these 'politics of geopolitics' into some kind of per spective, one might work out two lines of argumentation that can be separated only for academic purposes and that interact in the current geopolitical discourses in various ways: (1) the geopolitics of cultural difference, fragmentation, and conflict; and (2) the geopolitics of universalism and hegemonic superpower.

The Geopolitics of Cultural Difference, Fragmentation, and Conflict

In the aftermath of September 11 the discourse of a 'clash of civilizations', which had already been promulgated by Huntington in 1993 and immediately used by himself to discursively frame the Balkan Wars, appeared ubiqui tously in the media and public opinion. Even though many politicians sought to emphasize repeatedly that the events did not represent a clash of civilizations, they still reified the basic idea of a geopolitics of cultural differ ence. Huntington was not solely responsible for this. He only gave new life to an underlying discourse of cultural difference that had developed over more than two cen turies. Huntington updated the discursive mapping of the self and the other from the perspective of the 'modern' West. These representations had their roots in historic geographical imaginations of power such as Ritter's 'Gang der Kulturen über die Erde', Kolb's 'Kulture rdteile', and the Orientalism discourse, which has been a central focus of Western thought at least since the era of colonialism. As Ó Tuathail (1996) so aptly sums it up: "Huntington's thesis is not about the clash of

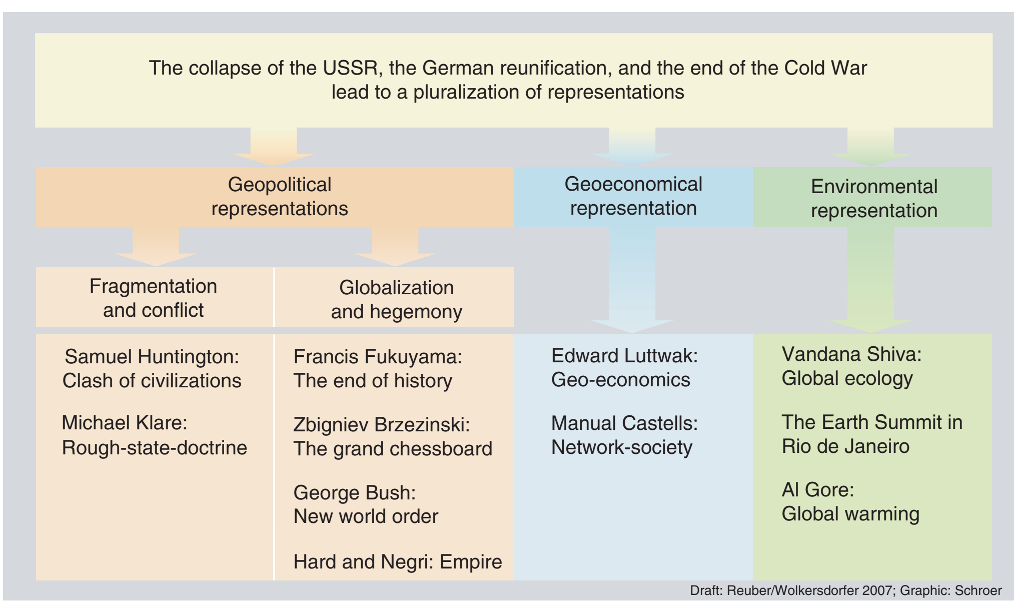

Figure 2 Different geopolitical discourses after the end of the Cold War. Modified from Reuber & Wolkersdorfer, 2004.

civilizations. It is about making global politics a clash of civilizations."

The Geopolitics of Universalism and Hegemonic Superpower

The second major geopolitical narrative of the post Cold War era serves the imagination of a new hegemon, the 'only superpower USA': Francis Fukuyama's thesis of *The End of History* at first glance does not seem to aim for a geopolitical segmentation of the world after the Cold War. Nonetheless, it still has distinct geopolitical implications. Fukuyama believes that we are entering an era characterized by "the end point of mankind's ideological evolution and the universalization of Western liberal democracy as the final form of human government." Fukuyama or 'George W. Bush's Aristotle' (a headline of the *Sueddeutsche Zeitung*) represents the USA as the ultimate democratic state. His ideas are informed by a reductionist interpretation of Hegel's idealistic philosophy of the end of history. This construction, in all its simplicity, marks the beginning of a geopolitical narrative that helps to pave the way for a new discourse of American hegemony after the end of the Cold War. Even if the 'end of history' thesis does not mark the beginning of geography in geopolitical discourse, it still contains an underlying geopolitical narrative which is only a short step away from the Bush (Sr.) doctrine in the first Gulf War, and still continues in the geopolitical representations and mappings of US foreign policy after 9/11 (**Figures 3** and **4**).

The New Spatial Organization of Power in Geopolitical Representations after 9/11

The attacks on New York and Washington formed the 'discursive event' in the face of which the emerging geopolitical representations rose to prominence. At the same time these events may be regarded as a shift in the geopolitical representations of power. For the first time the traditional order of 'modernity' (the Westphalian system of territorially defined nation states) was challenged by an enemy, who was organized (and acting) according to the principles of network societies. In response to September 11, politics and the media have impressively shown how difficult (if not impossible) it was to comment on and to react to the attacks beyond the spatial framings and representations established in traditional geopolitical discourses. Thus, the narration about the conflict had to be (and was) partly reframed as a 'war between nation states'. This was indispensable since, to quote Dalby, (2003) "the modern geopolitical reasoning ... operate(s) in terms of a political ontology of states." In this respect, soon and once again the nation state became the most important symbol framing

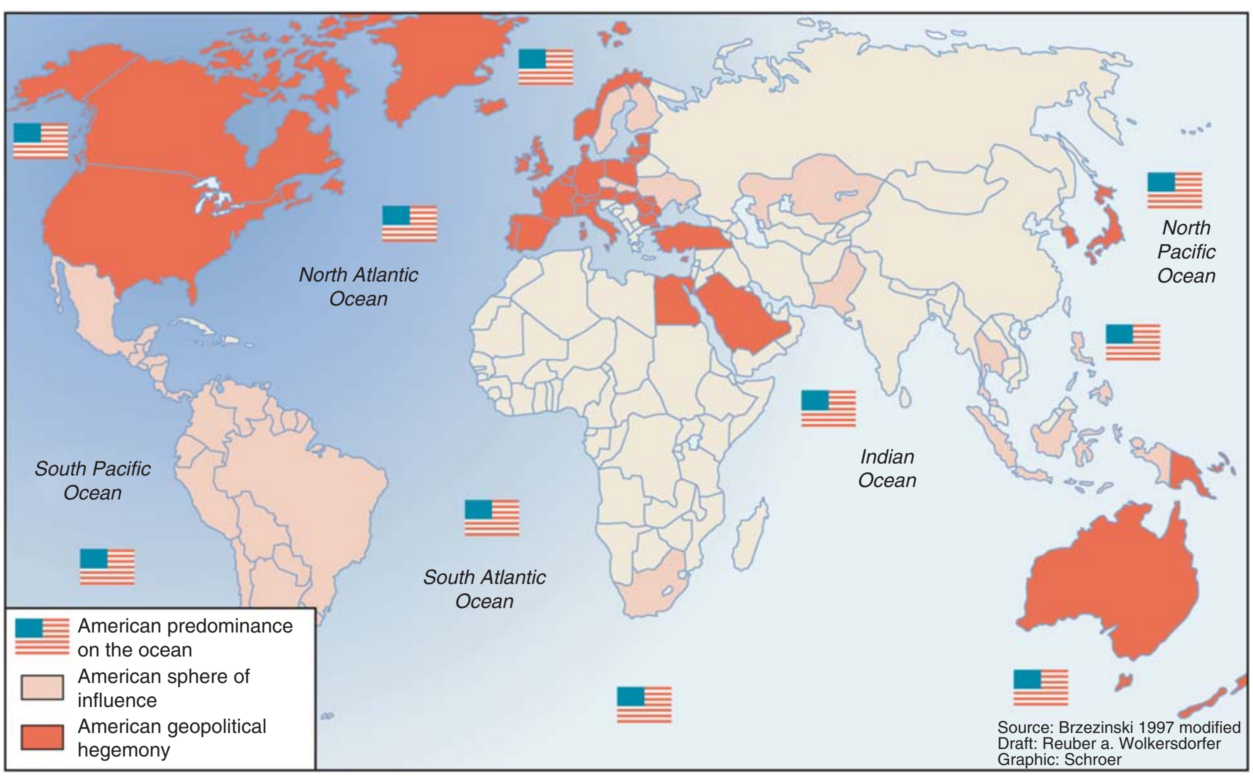

Figure 3 Geopolitical representations of intellectuals of statecraft: Brzezinski's only superpower. Modified from Reuber & Wolkersdorfer, 2004.

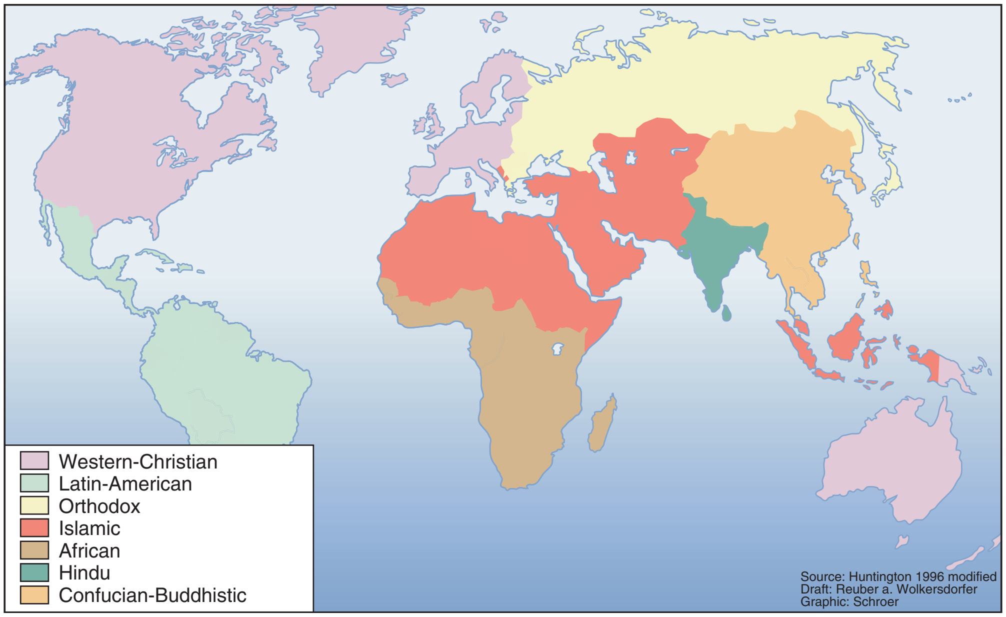

Figure 4 Geopolitical representations of intellectuals of statecraft: Huntington's clash of civilizations. Modified from Reuber & Wolkersdorfer, 2004.

political reactions as well as the innumerable reports in the media. Since the attacks were understood as 'national' affront and 'attack on America' (as a nation) – and the proposed counter measure was the 'international coali tion against terrorism', this refers to the 'nationalization' of the 'terrorist attacks' on World Trade Center and Pentagon.

The demand to 'localize' the Al Qaeda terrorist net work led to the necessity to attach them to a place re spectively; and Afghanistan was established as a first target hosting the enemy. This framing rests on geo political imaginations, which have been developed al ready during the last decade, for example, by the 'rogue state' doctrine of Michael Klare, George Bush's (Sr.) 'axis of evil', or more general by the discourse about 'failed states' coming from IR think tanks in the USA. Thus, the new storyline of a global 'war against terrorism' turned incredibly fast into one of the most powerful doctrines of international geopolitics.

For this change the subtle interplay of the two above mentioned geopolitical representations of hegemony (the only superpower) and fragmentation (the clash of civil izations) were particularly helpful. Both of them served as framings for the numerous speeches of politicians as well as reports and comments in the media. Most of them reified the underlying concept of a cultural fragmen tation of the world deriving from that the necessity of a strong US leadership within a 'coalition of the willing'. Based on such an imagination the wars against Afghani stan and Iraq could be justified toward the world public more easily. However, Huntington's rhetoric of conflict did not only fit into the discourses of Western military strategists and politicians but also into the saga of the holy war of the opposite side. Thus, the narrative of the clash of civilizations offered a discourse that also an tagonists such as Osama Bin Laden would easily accept. To quote Agnew, "Bin Laden is the Samuel Huntington of the Arab world … He is a prophet and organizer of inter civilizational conflict. Bin Laden is the modern Arab geopolitician par excellence" (Agnew, 2001).

As parts of the currently circulating representations of fragmentation and hegemony even traditional elements of classic American geopolitical imaginations reoccur. This is reflected explicitly, for example, in Brzezinski's remarks on the USA's political and military engagement in Eurasia. He suggests that in order to reduce the danger of the USA being on the geopolitical rim their troops always have to be stationed on the borders of the Eur asian continent to have a massive impact on the countries in the Eurasian hinterland in case of a conflict. Brzezinski calls Eurasia the grand chessboard and thus the region where a potential competitor of the USA for the world power might grow up. In his opinion, the priority of an American geo strategy must be to support their geo political interests in Eurasia. Brzezinski's reasonings contain elements of geopolitical imaginations that were developed in the middle of the last century. In particular, Nicholas Spykman argued already in 1944 that the coastal areas of Eurasia or the 'rimland' (as he called it) are the key to control the center. With such a concept he updated the British geographer and geopolitician Halford Mackinder. To some extent, therefore, the geopolitical representation of Mackinder serves as a persistent blue print of the foreign policy of the USA: Who rules Eastern Europe – so this imagination – rules Central Asia, and who rules Central Asia rules Eurasia, and who rules Eurasia rules the world.

The Pentagon's New Map

Geopolitical representations like the ones mentioned above are, of course, not the end of all storylines. Due to the changing circumstances in IR at the beginning of the new millennium another couple of them entered the scene. To pick out just one of them, Thomas Barnett's geopolitics of 'core and gap' shall be discussed just in its very general terms in order to show some similarities and differences to the framings above. Thomas P.M. Barnett, professor at the US Naval War College and after September 2001 a close advisor of US Defense Secretary Donald Rumsfeld, has designed a 'new world map of the Pentagon' (**Figure 5**). It has been published in the *Esquire*, March 2003, and in different journals of political science and has called a lot of attention among advisors. In Barnett's representation, the war against Iraq marks geopolitically a kind of historical turning point – the moment when Washington 'really' takes over its role as a guardian of the strategic security in the age of globalization. Barnett's simplistic geopolitical world order is: regions with stable governments and a high standard of living often are, at the same time, well connected within the globalized network society where democracy, liberal media, civil security, and human rights protection are part of the basic normative assumptions of society. These parts of the world he calls the 'functioning core'.

In opposite, regions that are hardly touched by globalization and suffer from repressive regimes with widespread poverty and diseases and – most important in Barnett's opinion – with chronic conflicts in which the coming generation of global terrorists is brought up, he calls the 'arc of instability'.

Barnett isolates his geopolitical representation in both a simple and radical way: he draws a line around the majority of US military interventions abroad, and by this he maps the 'nonintegrating gap'. Obviously, there are outliers excluded geographically by this simple approach, such as Israel isolated in the 'gap', a North Korea adrift within the 'core', or a Philippines straddling the line. Looking at this mode of representation, it is hard to deny the hidden logic behind the picture: If a country is either losing out to globalization or rejecting much of the content flows associated with its advance, so the implicit message of Barnett's suggestions, there is a far greater chance that the US will end up sending forces at some point. Conversely, if a country is largely functioning within globalization, it tends not to have US forces sent

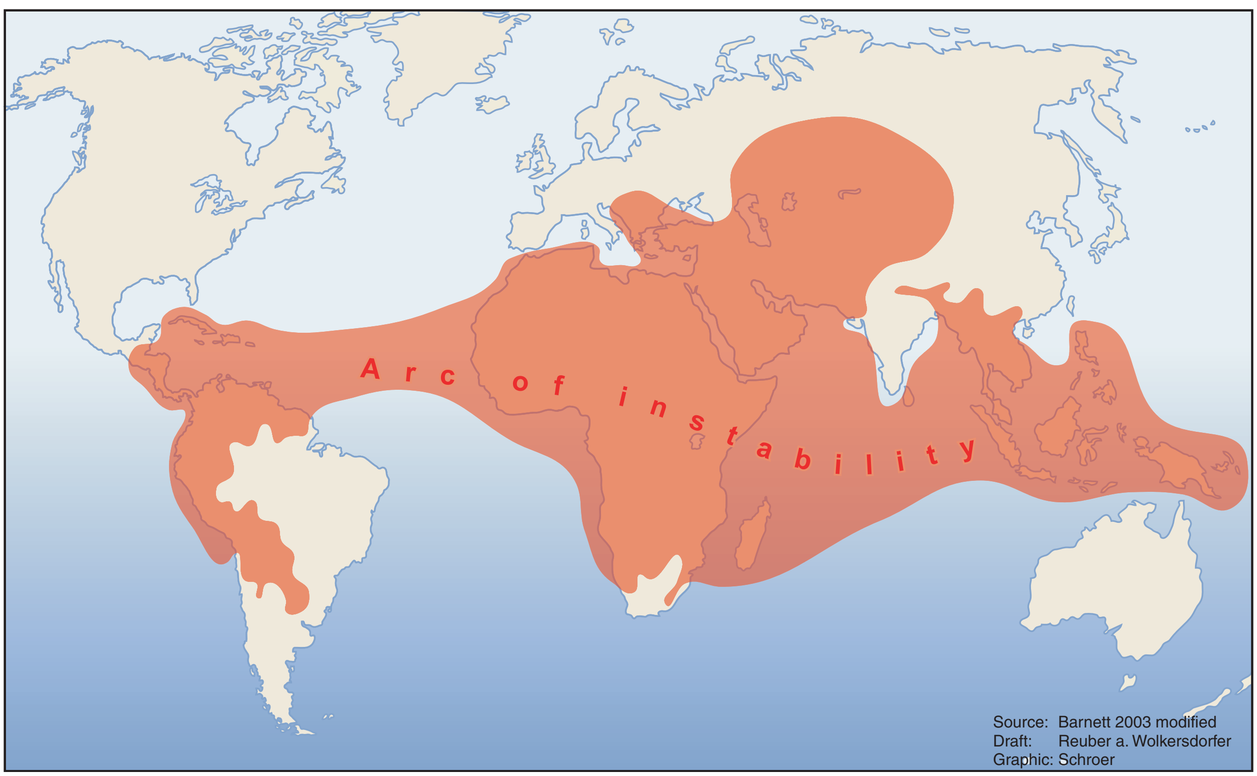

Figure 5 Geopolitical representations of intellectuals of statecraft: Barnett's Pentagon's new map. Modified from Reuber & Wolkersdorfer, 2004.

there to restore order and to eradicate threats. The sci entist was selected as the 'strategist of the year' in 2003 by the magazine *Esquire*.

See also: Cold War; Critical Geopolitics; Discourse Analysis; Foucauldianism; Hegemony; Imperialistic Geographies; Political Geography; Poststructuralism/ Poststructuralist Geographies.

Further Reading

Agnew, J. A. (1998). *Geopolitics. Re Visioning World Politics*. London: Routledge.

Agnew, J. A. (2001). Not the wretched of the earth. Osama Bin Laden and the 'clash of civilizations'. *The Arab World Geographer* 4, 112 117.

Agnew, J. A., Mitchell, K. and Ó Tuathail, G. (2002). *A Companion to Political Geography*. Malden: Blackwell.

Albert, M., Reuber, P. and Wolkersdorfer, G. (2006). Geopolitik. In Schieder, F. & Spindler, M. (eds.) *Theorien der internationalen Beziehungen. Eine Einführung*, pp 501 535. Opladen: UTB.

Anderson, B. (1983). *Imagined Communities. Reflections on the Origins and Spread of Nationalism*. London: Verso Ed.

Barnett, T. P. M. (2005). *The Pentagon's New Map: War and Peace in the Twenty First Century*. Berkley: Berkley Trade.

Brzezinski, Z. (1998). *The Grand Chessboard: American Primacy and Its Geostrategic Imperatives*. New York: Basic Books.

Dalby, S. (2003). Calling 911. Geopolitics, security and America's new war. *Geopolitics* 8(3), 61 68.

Dodds, K. J. and Sidaway, J. D. (1994). Locating critical Geopolitics. *Environment and Planning D: Society and Space* 12, 515 524.

Fukuyama, F. (1992). *The End of the History and the Last Man*. London: Perennial.

Gregory, D. (1998). The geographical discourse of modernity. In Gebhardt, H., Meusburger, P. (eds.) *Explorations in Critical Human Geography*: Hettner Lectures 1, pp 45 70. Heidelberg: University of Heidelberg.

Huntington, S. P. (1996). *The Clash of Civilizations and the Remaking of World Order*. New York: Touchstone.

Mackinder, H. J. (1904). The geographical pivot of history. *The Geographical Journal*, 23, 421 444.

Ó Tuathail, G. (1996). *Critical Geopolitics. The Politics of Writing Global Space*. Minneapolis, MN: Routledge.

Ó Tuathail, G. and Dalby, S. (eds.) (1998). *Re Thinking Geopolitics. Towards a Critical Geopolitics*. London: Routledge.

Ratzel, F. (1906). *Kleine Schriften*. Munchen: Oldenbourg.

Schultz, H. D. (1998). Herder und Ratzel. Zwei Extreme, ein Paradigma? *Erdkunde* 52(2), 127 143.

Sprengel, R. (1996). *Kritik der Geopolitik*. Berlin: Akademie.

Walker, R. (1993). *Inside /Outside. International Relations as Political Theory*. Cambridge, MA: Cambridge University Press.

Geopolitics and Religion

G. Dijkink, University of Amsterdam, Amsterdam, The Netherlands

Glossary

Chosen (Elected) Nation The idea of having a special mission in the world as nation (sanctified by God).
Civil Millennialism Identification of earthly powers or events in terms of religious prophecy.
Civil Religion The application of a religious attitude to political and public affairs.
Clash of Civilizations Theory predicting that war will increasingly occur between cultural groups, particularly those that are religiously different.
Critical Geopolitics Study of beliefs or implicit knowledge in statesmen, intellectuals of statecraft, or the public about geographical power arrangements in the world.
Cuius Regio, Eius Religio Whose the region is, his religion: the sixteenth-century principle that the official religion in a country will be the religion of the ruler. It was the first acknowledgment of the legitimacy of Protestantism.
Ethnoreligious Conflict Conflict between a religious minority and the state in which religion is an issue.
Jihad Struggle for religious purity in one's own life and in one's environment, religious 'war'.
Millennialism Utopian expectation of a new world order in which 'the faith' will triumph.
Sacred Something that transcends the individual and the present; something related to the supernatural world.
Secular, Secularization Loss of religiosity and disappearance of religious norms in daily life.

Introduction

Geopolitics is traditionally the study of how political power is reinforced or undermined by geographical arrangements (boundaries, coalitions, spatial networks, natural resources, etc.). Geopolitics also denotes the practical decisions of political and social leaders with respect to geographical arrangements. 'Critical geopolitics' deconstructs the assumptions about the world behind such decisions. Religion is a way to explain the world (just like geopolitics) and give moral meaning to events and relations between people and their environment on the basis of supernatural forces or deities. Religions create a bond of solidarity between people and therefore can have an impact in the sphere of power and politics. A religion may also offer its own normative arguments about geographical arrangements and may in this way act as a kind of geopolitical 'theory' based on supernatural rather than natural forces. Both aspects, religious solidarity and religious theory about the world, are relevant in contemporary geopolitics.

The study of geopolitics has always pivoted on the state as the main power agency in the world. Since the (Western) state became strongly defined by secular aims and values in the nineteenth century, religion was not geopolitics' foremost preoccupation. The secular view was dominant when most governments in the world were either European, offspring from Europe (Americas, Australia) or dominated by a European colonizer (Africa, Asia). From the moment that non Western states were able to throw their weight about in the world and had replaced their Westernized political elites, religion could also reappear in world affairs. One factor still smothered the emergence of religious politics on the international scene: communism (or Marxism). Marxism was a tempting ideology in former colonies because it was seen as the logical 'Third World' answer to 'First World' exploitation. The attractiveness of communism in some newly independent states has also been explained from its neat fit to the socioreligious order of society (like Brahmanism in India). Anyhow, the end of the ideological struggle between capitalism and communism created new opportunities for other interpretations of the world. The most significant one in geopolitical terms was the rise of political Islam (Islamism). It touched a large world region that was moreover significant because of its oil resources and antagonistic sentiments issuing in terrorism.

It is not surprising that political scientists in the West became more attentive to religion after 1990. New publications reflected two perspectives. The first perspective focused on the idea that Western (or American) foreign politics had missed chances to avoid conflicts or failures like the alienation of Buddhists during the American intervention in Vietnam. American policy makers did not understand that religious legitimacy can make or break leaders in other nations. The collection of essays *Religion: the missing dimension in statecraft* is particularly characteristic of this perspective. Another perspective boils down to the vision of a new type of world order in which antagonisms between groups are defined by culture rather than ideology. Huntington's *The clash of civilisations* is the most significant publication in this context. The book experienced a new surge of interest after the attacks of September 11 in New York and Washington.

Huntington's is by far the most geopolitical of all studies about the politics of religion appearing after 1990 because it extends the familiar idea of strategic blocs (like those of the Cold War) to the world of civilizations, essentially conceived as religions. These civilizations are portrayed as antagonistic.

Although geographers have since long written on the geography of religion, they were not in the forefront in studying the new geopolitical implications of religion. Reflecting the tainted history of geopolitics (after its popularity in Nazi Germany), much effort went into denouncing the classic geopolitical lapses in such views as Huntington's. Several decades ago the geographer John Kirtland Wright already coined the term 'geopiety': the religious interpretation of landscape elements like the endowment with a fertile or safe environment as sign of God's benevolence or specific geological conditions as witnesses of the Noachian Deluge. Such interpretations may play a role in a geopolitical context but in Wright's essays this extension was never made. Roger Stump's *Boundaries of faith* is a combination of a more traditional study in cultural geography and the new interest in the territorial strategies and geopolitical visions of religions. The study partly covers issues of geopolitical interest.

Recent work of geographers has been for the greater part inspired by the critical geopolitics perspective. That means a predominant orientation on the (implicit) geopolitical content of texts and messages of religious leaders and movements particularly among Christians and Muslims. This may be more specifically applied to the question how politicoterritorial conflicts and strategies are supported by religious ideas or movements: for example, the Palestine–Israeli conflict or the changing conceptions of the nation in a de and reterritorializing world.

The pivotal role of the state in geopolitical writing suggests that nation building and its reverse, transnationalism, cannot be ignored. So, an extensive discussion of religion in the context of geopolitics deals with three issues: religion and nationalism; religion as transnational force; and religion and territorial conflict. These constitute theoretical chapters rather than a sound classification of political events in the world.

Religion and Nationalism

While Kings of late medieval states and early modern states would claim to have received their authority from God, Protestant (Republican) polities rising in the sixteenth century more explicitly evoked the idea of a 'covenant' between God and people in analogy to the Biblical narrative of the Jewish wanderings to the Promised Land. The idea of a 'chosen' people easily rose in situations where a nation felt threatened or marginalized by such dominant powers as the medieval German Empire (feared by the thirteenth century French Capetians who, as a reaction, appropriated the status of 'chosen' country), the Habsburgs (English and Dutch self image as chosen people in the sixteenth century), or the British Empire (the Boers in South Africa in the nineteenth century). The voluntary exile of early Protestant settlers in America, aspiring to a religiously more pure community than was possible in the Old World, also raised the expectation that the new millennium with the Second Coming of Christ would start in America. The kind of state that developed in the nineteenth century transformed the old religious connotation into a more secular 'exceptionalism', that means the idea of having a special mission in the world like guiding other people (Britain), propagating democracy (the US), or economic development of poor nations (Europe).

Nationalism scholars like Anthony Smith have pointed out that the nation has a 'sacred dimension'. 'Ethnic election' is one manifestation of this dimension. There are many national myths that express the chosenness of the nation like the appearance of an angel to the (alleged) founding father of the nation or the presence of relics that express special divine benevolence (the shrine of Jasna Gora – the Black Madonna – in Poland). More important from the viewpoint of geopolitics is the idea of 'sacred territory'. It implies that, according to tradition, God has promised the nation a specific territory that not necessarily corresponds with the current territory of a nation. The occurrence of specific holy events in the past (like the appearance of an angel on the battlefield of Kosovo Polje in the Serbian myth about the historic defeat in the war against the Ottoman Turks) or the presence of sanctuaries that are essential to a religion (Jerusalem) may be an important source for claims on territory and motive for waging war. Other aspects of the sacred dimension of nations that are mentioned are 'ethnohistory', particularly in the shape of a celebrated period in national history, and the idea of 'national sacrifice', blood spilled for the nation as commemorated by numerous war monuments. All these aspects of the sacred dimension testify to the nation as an immaterial and eternal entity but involve different ways of applying the words 'religious' or 'sacred'. While the ideas of 'ethnic election' or sacred territory may appeal to a specific religion or Holy Scripture in order to legitimate such claims, ethnohistory and national sacrifice rather suggest 'deification' of the nation itself. The latter is compatible with a strongly secular society like post Revolutionary France, a society highly aware of administering a very special heritage entailing personal obligations and worship.

Whereas nationalist movements may easily absorb religion, the above also suggests that nationalism itself constitutes a (pseudo) religion. This is particularly strong in longer established states. Religion in the proper sense, however, is rather visible in mobilization against state power. The Evangelical Church was active in coordinating protest against the communist regime in the last days of the German Democratic Republic. The role of the Roman Catholic Church in keeping alive a critical attitude to international communism in Poland during the Cold War is generally acknowledged. Although it would carry the influence of religion too far to attribute the fall of the Soviet Empire to the Polish pope and the religiously inspired Polish resistance (as some commentators insist), religion certainly made Poland an awkward member of the international socialist family. The overthrow of the Shah regime in Iran in 1979 also depended on the strong religious leadership of the exiled Ayatollah Khomeini. The subsequent development of international religious leadership in the shape of a 'theocratic' state in Iran may have damaged rather than helped the cause of Shiite minorities in other states. The failure to see the potential weakness of the Shah regime is one of the notorious cases in the story of lacking religious sensitivity of Western (US) foreign policy and intelligence elites.

Ethnonationalist struggles for autonomy or independence are obviously reinforced by the feeling of being religiously different, but one may doubt whether in such conflicts religion is any different from other cultural markers. Researchers into the causes of ethnonationalist conflict usually reject the idea that religion itself causes ethnonationalist mobilization and religious issues constitute a minor issue in such conflicts. This usually concerns limitations on religious freedom of a minority and in some cases a conflict on the 'ownership' of a sacred site (Hindus and Muslims disputing the religious origin of the Ayodhya Temple). The fact that religion implies a comprehensive worldview that is incompatible with other religious views does not completely fit the observation that in many countries different religions coexist peacefully.

Religion as Transnational Force

The era of 'globalization' has brought back the phenomenon of religion as an influence competing with state power. This is particularly relevant in states that do not have a strong legitimacy rooted in historic freedom struggles and centuries of defense against outside or domestic threats. It is the world where Islam has stretched its wings: Asia and Africa. Although Islam can create quick transnational solidarities, particularly when Muslim minorities are in danger (Bosnia) or 'the faith' is offended (cartoons of the Prophet published in a European newspaper), it does not aim at replacing the state. It only wants to replace the moral superiority that the state has assumed in the West. Islam (including so called fundamentalist Islam) has surged into an important political movement for three reasons. First, it is the most accessible means of emancipation analogous to nationalism in Western countries. According to certain nationalism specialists, the adoption of a strict variant of Islam (fundamentalism) by rural people is comparable with the embracing of a 'high (urban) culture' by peasants in the course of the nineteenth century nationalist revolutions in Europe. Second, Islam has adopted a welfare role where the state fails to provide such care. The building of mosques goes along with the establishment of schools, care for the poor, and administration of justice. Even in utterly failing states in Africa, some prosperous native will see it as his task to build a mosque and set up the concomitant social institutions. Third, the Muslim world, and particularly its Arab branch, has a strong feeling of deprivation toward the West which can be partly attributed to the legacy of unfulfilled promises by colonial powers in the early twentieth century, partly to the unsolved Palestinian problem, and partly to the perception that the materially advanced Western world discriminates Muslims. These factors explain why Islam is the most obvious identity for millions of people in Asia and Africa who try to make sense of the world. They expect that their states adopt this identity, not that they amalgamate in one big political unit.

The real transnational manifestation of Islam as 'neofundamentalism', however, derives from the fact that migration has turned Muslims into minorities in a cultural alien world, the West. This 'deterritorialization' has disengaged traditional (territorialized) Muslim culture and religion. Challenged by the universalistic pretensions of the West, second generation Muslim migrants cling to Islam as a superior identity marker. However, this involves a rejection of all traditional authorities, dogma, and cultural impurities ending in a strongly individualized form of religiosity. While this attitude facilitates transnational networks like global terrorism, according to Olivier Roy, networks like Al Qaeda are not true underground movements with a geostrategic perspective. They are just as powerless against the world of states as Islamism is against the dominant nationalism in the Muslim world.

In spite of Islamic terrorism, there is no contemporary example of a religion directly acting as transnational political force like the Catholic Church in medieval times or the early Islamic empire. In the world of today, religious viewpoints on worldly affairs are mediated by states or intergovernmental organizations like the Organization of the Islamic Conference. The Roman Catholic Church has an authoritative voice on policy matters like war or health ethics, but does not mobilize people into a strong transnational political movement. Of course, we may understand 'geopolitics of religion' as designation of the strategies that religious institutions employ to extend their influence in the world by means of missionary activities or work in the social and health sphere. The expansion of Islam or the surprising success of Evangelical movements (like Pentecostalism) in Latin America would fall under this definition. Converting people is not politics in the usual sense, but since military interventions in other countries nowadays operate under the flag of 'winning the hearts and minds of people', we should admit that cultural and political influence are difficult to separate.

Religion and Territorial Conflict

The role of religion in ethnonational conflict has been mentioned. Religion is one of the cultural markers that creates a distinction between 'us' and 'them', religious leaders can have a decisive influence on social mobil ization and a religious worldview implies certain prac tices and norms in daily life that may conflict with the rules of a dominant ethnic group. The religious motiv ation behind the resistance against the dominant political group in Chechnya, the Armenian enclave Nagorno Karabakh, Eastern Timor or Northern Ireland cannot be denied but does not permit any conclusion about the incompatibility of these religions in other settings. While the theory of 'the clash of civilizations' offers a (severely criticized) thesis on the inevitability of such conflicts and the prospect of enduring territorial separation between civilizations and religions, the study of (critical) geo politics can more directly address the potency for conflict by looking into religious doctrines in Holy Scriptures or in the interpretations of influential preachers, ayatollahs, imams, priests, or 'holy men'. These sources may reveal three themes that can cause territorial conflict: holy land, holy war, and millennialism.

Holy Land

Apart from its nationalist meaning as 'sacred territory' (that means the worshipped homeland of a nation), a holy land may have significance for those living outside its territory, most commonly in the peaceful sense of a place of pilgrimage. This is usually based on artifacts reflecting the divine manifestation on Earth or the fact that crucial events of the religious narrative have occurred on that place. This condition may also produce conflict. The presence of American troops in the holy land Saudi Arabia was the reason for attacks on American interests; at least in the declarations of Osama bin Laden and his Al Qaeda companions. Conversely, American support for Israel seems to exceed what a balanced or 'realist' Mid Eastern policy would require. The spread in the United States of a Christian doctrine saying that nations will be judged in the Final Judgment according to their attitudes to Israel is well documented. The question about its in fluence on the White House, however, is controversial.

A country may also be treated as a holy land because it is considered as 'belonging' to a certain religion. This perspective is rather strong among Muslims who have a tradition of making a sharp distinction between the 'Land of Islam' and the 'Land of War or Unbelief'. Would a country fall into the hands of unbelievers then a strong reaction is called for. This happened when the Soviet Union installed a pro communist government in Kabul in 1979. The events in Afghanistan mobilized young Arab idealists to revive the idea of *jihad* as a territorial struggle and participate in the Afghan guerrilla against the infidel occupier.

Holy War

Holy war, most commonly associated with the medieval Christian Crusades and the Muslim rule of *Jihad*, is quite ambiguously related with the Holy Scriptures of the two religions. Both the Bible and the Koran contain stories of divinely sanctioned war that suggest a certain exegesis, but these texts do not unequivocally call to propagating the faith with the sword. *Jihad* basically means that the faith requires a personal struggle with evil temptations. The step to imagining a territorial struggle is obviously tempting but it is certainly not the dominant view among Muslims. The current world in which Muslim migrants are allowed to practice their faith in the midst of another civilization is completely different from the world in which Islam had to be defended militarily. Some ideas in religions reflect conditions in earlier historical periods. Both Christians and Muslims recognize the need to propagate the faith. The course of history shows an al ternation between more extravert and introvert inter pretations of this assignment. Sometimes believers discover that the faith requires 'acting in the world', sometimes the emphasis is more on personal devotion. Such modes of introversion or extraversion are induced by the international political context rather than ex plainable by a religion itself.

Millennialism

Millennialism concerns religious expectations on the end of the current world and the start of a new (earthly) 'millennium' in which the world will be reordered ac cording to God's will. In the Books of Revelation and Daniel, cues can be found on these events involving a final struggle with evil in the shape of a Beast or Devil (the Antichrist). Details about events and places may suggest similarity with current conditions. This may in duce politicians who believe that we are nearing the Great Reconstruction to refrain from solving conflicts and problems or even to accelerate fateful events. Al though the Koran does not provide similar apocalyptic details, millennialist movements also occur among Muslims. The portraying of America or the West as the 'Great Satan' with which a final battle has started made many Muslim's hearts quicken particularly when hearing about the attacks of September 11.

Conclusion

Geopolitics has been particularly engaged with territorial conflict and typically applies a zero sum perspective on power that does not fit in very well with expectations of world peace. Religion appeared to be surprisingly com patible with a geopolitical perspective after the events of September 11. This obfuscated the fact that religions or religious leaders may also act as peace promoters on the international scene. The more typical geopolitical ap proach would anyhow be to focus on the universalistic pretensions of believers and see if these lead to territorial exclusionary behavior. There are three collision fronts in which religion plays a role: (1) between ethnic minorities and the state, (2) between states or blocs representing different 'civilizations', and (3) between states and transnational religious movements or terrorists. Such collisions are never a 'logical' result of the contact be tween different religions. The important thing that geo political analysis can help to make clear is that a so called 'clash of civilizations (or religions)' is the consequence of religious bonds of solidarity that have been forged by political events (rather than the reverse). The supposed *jihad* of Muslims against the West cannot be understood outside the context of what happened in Afghanistan in the 1980s. The analysis of texts such as produced by Osama bin Laden does not reveal a deep justification of terrorist attacks by Islamic codices; they are rather *ad hoc* talks with the air of piety. Religion is important as one of the 'possible' frameworks for mobilization but collision lines may also run through a religion as the wars between Shiite and Sunni factions in Iraq and the religious wars in early modern Europe show. The interaction between type of political event, type of religion, and territorial aims is the main challenge for geopolitical analysis.

See also: Ethnic Conflict; Ethnicity; Modernity; Nation; Nationalism; North-South; Terrorism; War.

Further Reading

Agnew, J. (ed.) (2006). Special issue on religion and geopolitics *Geopolitics* 11, 183 191.

Fox, J. (2002). *Ethnoreligious Conflict in the Late Twentieth Century*. Lanham: Lexington Books.

Gellner, E. (1992). *Postmodernism, Reason and Religion*. London: Routledge.

Goh, R. B. H. (2004). Asian Christian networks: transnational structures and geopolitical mappings. *Journal of Religion and Society* 6, 11 13.

Hervieu Léger, D. (2002). Space and religion: New approaches to religious spatiality in modernity. *International Journal of Urban and Regional Research* 26, 99 105.

Huntington, S. P. (1996). *The Clash of Civilisations and the Remaking of World Order*. New York: Simon and Schuster.

Johnston, D. and Sampson, C. (eds.) (1994). *Religion: The missing Dimension in Statecraft*. New York: Oxford University Press.

Nyroos, L. (2001). Religeopolitics: dissident geopolitics and the fundamentalism of Hamas and Kach. *Geopolitics* 6, 135 157.

O'Brien, C. C. (1988). *God Land: Reflections on Religion and Nationalism*. Cambridge, MA: Harvard University Press.

Roy, O. (2004). *Globalized Islam. The Search for a New Ummah*. New York: Columbia University Press.

Smith, A. D. (2000). The sacred dimension of nationalism. *Millennium* 29, 791 814.

Stump, R. W. (2000). *Boundaries of Faith. Geographical Perspectives on Religious Fundamentalism*. Lanham: Rowman and Littlefield.

Relevant Websites

http://www.hiik.de
Heidelberg Institute of International Conflict Research, conflict data.

http://www.adherents.com
National and World Religion Statistics Church Statistics World Religions.

http://www.oic oci.org
The Organization of the Islamic Conference.

http://www.cidcm.umd.edu
University of Maryland, MD.

http://www.worldvaluessurvey.org
World Values Survey Organization.

Georeferencing, Geocoding

X. Yao, University of Georgia, Atlanta, GA, USA

Glossary

Attribute Data used to describe characteristics of things (e.g., spatial features) in a computer database.

Coordinate System A mathematical system for recording the location of a point in an n-dimensional space by a set of n values.

Frame of Reference The context, perspective, and/or a set of presumptions or rules within which instances are observed.

Geocoding The process of associating addresses, place names, or other referents with geographic locations. It is also referred to as discrete georeferencing.

Geographic Information System A system to capture, store, display, communicate, transform, analyze, and archive georeferenced information.

Projection A systematic transformation of coordinates in a three-dimensional coordinate system to a two-dimensional coordinate system.

Referent A description that is used to refer to a geographic location.

Introduction to Georeferencing

Georeferencing is fundamental to geographic information system (GIS). It refers to the process of assigning locations to geographical objects within a geographic frame of reference. Depending on the spatial resolution in effect, georeferencing mechanisms can be generally classified into metric georeferencing and indirect georeferencing. A few variations of indirect georeferencing methods exist, among which discrete georeferencing (also called geocoding) is commonly used in many applications.

Definition and Properties of Georeferencing

The usage of the term 'georeferencing' has not always been unambiguous in literature. Some definitions strictly refer to metric georeferencing and/or coordinate transformation, while some others solely imply the process of discrete georeferencing. A broad definition of georeferencing can be simply put as the process of assigning locations to geographical objects within a certain frame of reference such as a coordinate system.

The atom of geographic information is a three dimensional (3 D) unit, represented as 'location, time, attribute'. In order to communicate effectively, we need standard ways to handle the three dimensions. For example, the Gregorian calendar is the system being used worldwide for specifying time. Any moment in the 'time' dimension can be uniquely represented in this time system. The calendar serves as a frame of reference to specify time. There are also other calendars that can uniquely assign a time stamp to any moment in the time dimension. In other words, there are multiple possible frames of reference for time specification. Georeferencing mechanisms serve a similar purpose in the location dimension as the calendar systems do in the time dimension. A referent can be expressed as coordinates, street address, or other forms of measurements in an appropriate geographical frame of reference. Corresponding to the variety of geographical frames of reference, there are multiple georeferencing methods. Two basic properties are fundamental to all georeferencing methods, as discussed below.

Uniqueness

A georeferencing method must assign a single reference per location, and every location gets a unique reference. It is noteworthy that the uniqueness is required only over a certain spatial extent in a defined time frame. Spatially, a georeferencing method may be (applied to) the entire world, a country, or any other defined area. Temporally, it could be valid forever, for a specified time period, or until changes occur.

Spatial resolution

The spatial resolution of a georeferencing method is the size of the area covered by a single referent. A referent expressed as coordinate (x, y) is a point. Another referent '123 Main Street' pinpoints down to the footprint of a building, while 'metropolitan Atlanta' refers to a geographic area over 15 000 km^2 in size. Each georeferencing method comes with an inherent level of spatial resolution. Particularly, metric (or coordinate based) georeferencing can represent any location with no limitation on precision, which means it imposes no limitation on spatial resolution. Other georeferencing methods have a spatial resolution of various sizes depending on the specific referent in use.

Taxonomy of Georeferencing Methods

The International Organization for Standardization (ISO) categorizes georeferencing mechanisms into coordinate based and non coordinate based georeferencing. These are also called metric georeferencing and indirect geor eferencing, respectively, as revealed in **Figure 1**. The taxonomy virtually organizes georeferencing methods by the level of spatial resolution in effect. Metric georefer encing, with which location measures are taken on a continuous scale, has no limit on spatial resolution. In direct georeferencing, on the other hand, locates discrete units on the Earth's surface such as street addresses, post code areas, and statistical areas, etc. The spatial resolution of the indirect georeferencing methods are constrained by the size of the georeferenced spatial feature.

Many other terminologies are seen in literature, connoting the same meaning as that of the above men tioned terms of georeferencing types. Coordinate based georeferencing is more commonly referred to as 'con tinuous' georeferencing or 'metric' georeferencing. The non coordinate based georeferencing is also called 'in direct' georeferencing. Indirect georeferencing includes, but is not limited to, the popular discrete georeferencing methods and linear referencing system (LRS). Discrete georeferencing is more commonly referred to as geo coding. **Figure 1** provides a taxonomy of georeferencing methods and synonymous terminologies.

Georeferencing Methods

Metric Georeferencing

Metric georeferencing, also called continuous georefer encing, represents locations using their coordinates in the geographic coordinate system or a projection of it. The geographic coordinate system is a 3 D coordinate system for the Earth. It represents every location on the Earth's surface with a pair of longitude and latitude. **Figure 2a** shows the geographic coordinate system, in which the prime latitude (0° latitude) is the equator and the prime meridian (0° longitude) is set as the meridian passing through Greenwich, England. Through a projection of the geographic coordinate system, a 2 D coordinate system can be created to approximate the representation of the 3 D Earth surface in a 2 D coordinate system. **Figure 2b** shows an example of it. There are a variety of projected 2 D coordinate systems for different parts of the world or different purposes of representation. How ever, in each of all these coordinate systems, a location is uniquely represented by a coordinate.

Coordinates in one coordinate system can be trans formed into any of the other coordinate systems. Par ticularly, for example, georeferencing is named for an important procedure performed to remote sensing im ages. This procedure transforms coordinates in a local coordinate system of an image to the geographic co ordinate system (or a projection of it).

Indirect Georeferencing

Indirect georeferencing is a broad umbrella term for all non coordinate based georeferencing methods. To ex plain the nature of indirect georeferencing, ISO states that indirect spatial referencing methods are based on a relationship with a location defined by a geographic feature or features. The relationship of the position to a spatial feature could be any of the following:

1. containment, where the position is within the geo graphic feature, for example, in a country;
2. based on local measurements, where the position is defined relative to a fixed point or points in the geo graphic feature or features; for example, a given dis tance along a street from fixed point on the street;
3. loosely related, where the position has a qualitative spatial relation with the geographic feature or fea tures, for example, adjacent to a building.

The important message from this statement is that in direct georeference methods locate positions according to some kind of relations between the location and spatial feature(s). The necessary information for indirect geo referencing is usually stored as attribute(s) data related to the location. In a nutshell, because indirect georeferen cing methods make use of easily obtainable and

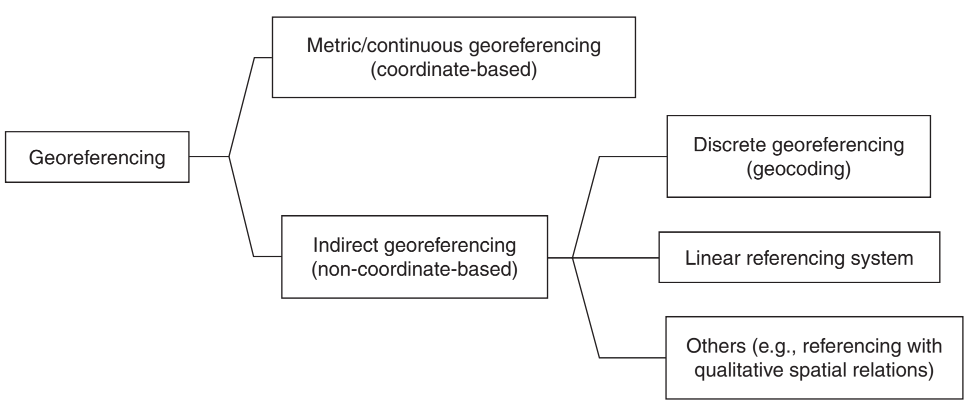

Figure 1 Taxonomy of georeferencing methods.

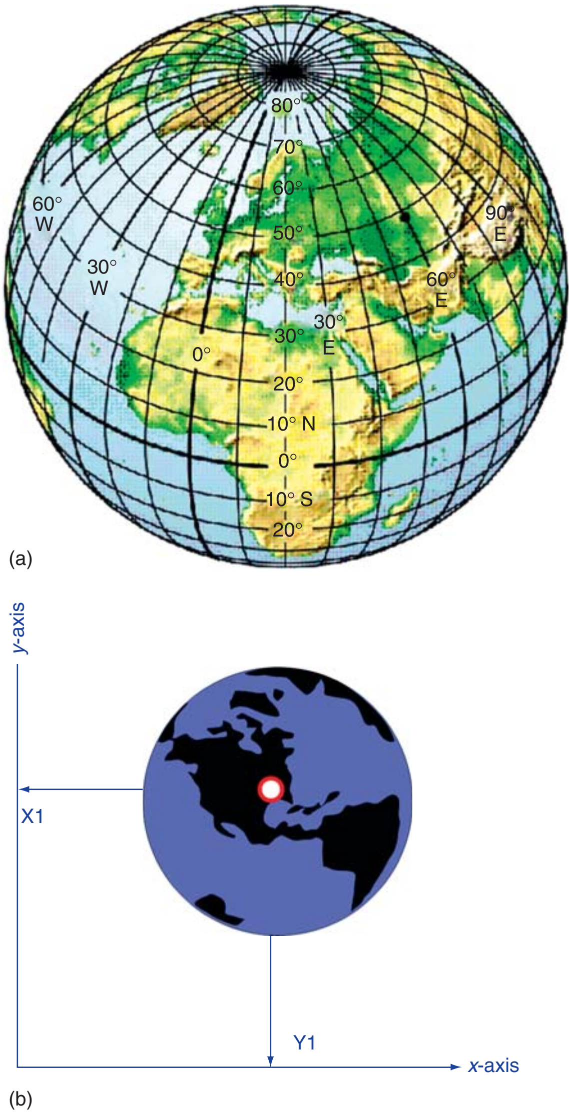

Figure 2 Metric georeferencing. Panal (a) shows the geographic coordinate system. Panal (b) shows an example of it.

meaningful attribute information and tie them to spatial locations, they are very commonly used in human geography, as well as many other fields. According to different types of attribute data and spatial relations used, there are a few different types of indirect georeferencing mechanisms as discussed below.

Discrete georeferencing

The most popular indirect georeferencing nowadays is probably the discrete georeferencing, also referred to as geocoding. Discrete georeferencing identifies discrete geographic units on the Earth's surface. The referents are usually names, indices (e.g., street addresses), or codes (e.g., zip codes). In order to be uniquely identified for the corresponding discrete geographic unit, the referent must have a clearly defined meaning in the underlying geographical frame of reference. For georeferencing in GIS, a geographical frame of reference is typically surrogated by a GIS data set consisting of spatial units, each of which is identified by a unique name or code. As illustrated in **Figure 3a**, a country name is used as a referent for the geographic territory of the named country. In such a discrete georeferencing, the geographical frame of reference consists of the political boundaries of world countries. There are other types of frame of references. Typical examples include those for geocoding street addresses, postal codes, administrative or statistical regions, and telephone area codes, etc. Discussed below are some of the examples.

Street addresses

A well known geocoding system is to identify the spatial locations for street addresses, and thus it is also called 'address matching'. This address matching method interpolates the location of a street address on the associated segment for the named street. The geographical frame of reference used for address matching is a specially designed geographic database of a street network.

Street address can be expressed in a variety of ways. In the United States, for example, a street address contains street number, street name, city, and the state (or just zip code). In all street address expressions, a part of it serves to narrow down the searching area in which the particular street name is uniquely identifiable. Illustrated in **Figure 3b** is an example of geocoding a US street address '2175 W Broad Street' in Athens, Georgia. The line segment in red indicates a road segment with the street name of interest. The street addresses of starting and ending points are labeled for both sides of the road segment. Because 2175 is an odd number, it means the discrete unit (a land parcel) to be represented is on the side of the street with odd number starting and ending addresses (2101 and 2199, respectively, in this example). The footprint of this parcel that has this address is highlighted in yellow in **Figure 3b**. It should be noted that an address matching result is a point location, interpolated between the starting point and ending point of the segment. This point location is shown with a red symbol in **Figure 3b**. The address matching result is a point location serving as the spatial proxy for a discrete spatial unit.

The example in **Figure 3b** also reveals that address matching imposes special requirements on the underlying street database. It requires that each street line segment in the database has accompanying records of its starting and ending street addresses. These addresses will be used for location interpolation. In the United States, the first generation of this type of street database design was called Dual Independent Map Encoding (DIME) files. It had been adopted by the US Census Bureau since the 1960s to build street databases. Since 1990, these

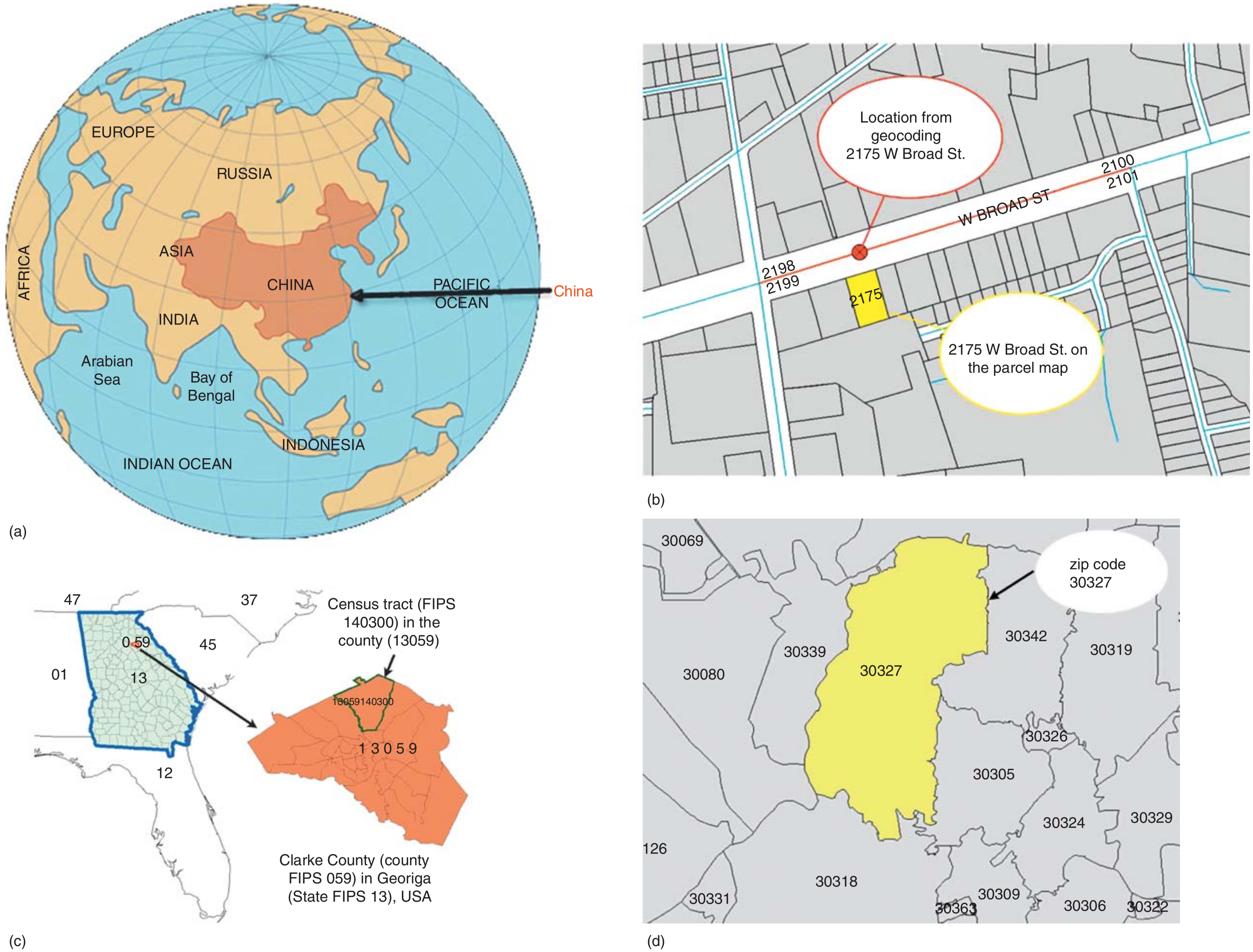

Figure 3 (a–d) Examp es of types of nformat on used for d screte georeferenc ng.

DIME files have been updated, extended, and renamed as Topologically Integrated Geographic Encoding and Referencing (TIGER) files. In the database design of TIGER files, a road segment is considered to have two sets of street addresses, one for the left side and one for the right side of the road, as shown in **Figure 3b**. Today, many digital geographic data vendors (e.g., NAVTEQ, Geographic Data Technology Inc., etc.) provide high quality street database based on the same design.

Administrative or statistical areas

Discrete georeferencing can also be applied to geocode administrative or statistical areas in a country. Each country has a hierarchical system of administrative and statistical areas. For instance, there are state, county, census tract, block group, and more, in the United States. At each level of the hierarchy, an area has a unique name or a uniquely assigned code. The names or codes can be used as referents for geocoding. The geographical frame of reference in use is the GIS database of ad ministrative or statistical areas at each level of the hier archy. In the United States, each administrative or census statistical area is assigned a unique federal information processing system (FIPS) code. **Figure 3c** shows an ex ample of three levels of administrative/statistical areas in Southeast United States. The two digit codes in blue are the FIPS codes for states. In a state, there is a three digit FIPS code for each county. The county with FIPS code '059' is displayed in an enlarged view for demonstration. Although county code '059' is only unique in the speci fied state, it becomes nationally unique if it is prefixed with the state code. Thus the longer county FIPS code '13059' can be used nationwide – it refers to Clarke County (059) in Georgia State (13). The same principle applies to FIPS code in any subsequent levels in the hierarchy. **Figure 3c** shows an example of a census tract in Clarke County. Such a coding system allows the FIPS codes to be used as referents for geocoding.

Postal code

Georeferencing postal codes work in a very similar way. Here the geographic frame of reference in use is the GIS database of geographical divisions associated with postal codes. The referents are the postal codes. Although each country has its own postal code system, a code in any postal system represents a uniquely identifiable geo graphical area in the country of interest. **Figure 3d** il lustrates such an example of five digit zip codes in the United States. The referent in the figure is the postal code 30307, which is georeferenced as the spatial areas highlighted in yellow in **Figure 3d**.

Linear referencing and dynamic segmentation

Although every spatial location can be uniquely repre sented in a coordinate system, metric georeferencing does not necessarily ensure faithful spatial relations. LRS was proposed in response to the need of retaining this topological relation, point 'on' a linear feature, when referencing locations within a linear network. For in stance, an accident occurred on a highway. The site, represented as coordinates (x, y) in a coordinate system, may not fall exactly on the highway due to measurement inaccuracies. In response to these kinds of problems, a linear referencing method identifies such a location by recording the linear feature (here the highway) and the network distance of the referenced location to the starting point of the linear feature. This way the linear referencing method warrants the existence of the location on the linear network. Specifically, a typical LRS uses road (or route) name and milepost. The road (route) name is used to identify the linear feature. The milepost measures the network distance of the referenced location to the origin of the road (or route). **Figure 4a** is an example of an accident location on Highway 80. The location is referenced as (Highway 80, 145.7) in an LRS.

LRS also provides the basis for dynamic segmentation that is used to represent any line segment on a linear feature. Dynamic segmentation records the starting and ending location of the segment using LRS. **Figure 4b** shows an example of dynamic segmentation. The geor eferenced is a road segment of a specified speed limit. The advantage of dynamic segmentation is at least threefold. First, the line segment can be dynamically referenced or modified while the underlying linear feature remains in tact. Second, the topological relationship can be kept be cause the line segment in dynamic segmentation is by definition on the underlying linear feature. Third,

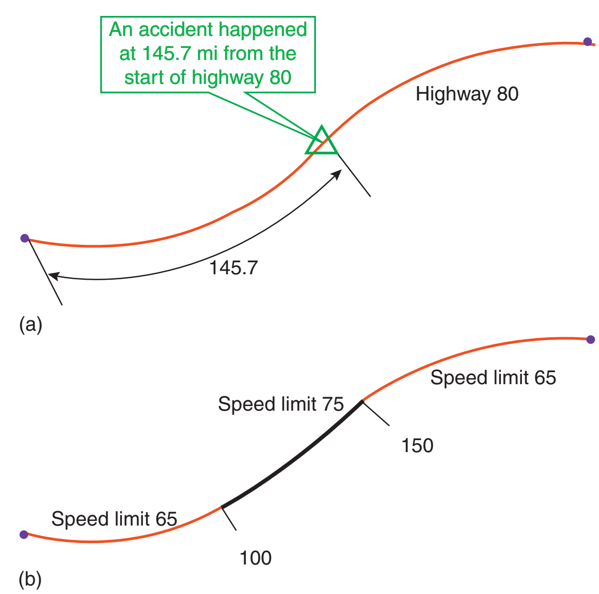

Figure 4 (a, b) Linear referencing system and dynamic segmentation.

multiple line segments, even spatially overlapping ones, can be referenced along the same linear feature.

Georeferencing with qualitative spatial relations

Spatial relations play an important role in indirect georeferencing. Discrete georeferencing and LRS are both based on two specific types of relations. Yet, there are many other types of spatial relations that can be used to describe locations. Indeed, it is quite common in people's daily life to refer to locations using spatial relations between the referents and other spatial features. For example, one may describe the location of a post office as "on Main street, north of the gas station at the intersection of Main and Second street." Such a description provides enough information to locate the post office, thus theoretically it could be georeferenced through the interpretation of such qualitative spatial relations. This could be particularly useful when discrete georeferencing is not applicable, for instance, when a place has no street address or when the street address is unknown to us.

In practice, current GIS technology and online location finding services provide some limited georeferencing capabilities by the use of qualitative spatial relations. These limited capabilities fall in two types: that by the use of topological relation 'intersection' and that by the use of proximity spatial relation 'near'. It is possible to find road intersections if the names of the two roads are provided. An example of using 'intersection' relation for georeferencing is shown in **Figure 5a**. In addition, some online map search engines (e.g., Google maps) also allow qualitative inputs of 'near' in spatial query. **Figure 5b** illustrates such an example. It finds all hotels 'near' a specific location. However, what is done in these GIS services is to simply translate 'near' into a certain geographical buffer distance and return all found locations within the distance. It is an oversimplified interpretation of this spatial relation because 'near' actually means different distances in different contexts.

To summarize, current geospatial technology only has very limited capability to support georeferencing with qualitative spatial relations. This remains to be an area for further research.

Using Georeferencing

Application of Georeferencing

Metric georeferencing lays the foundation of GIS, location based services, and other geospatial applications. The coordinate based georeferencing is used to build GIS databases, some of which can provide the underlying frame of reference for indirect georeferencing. Based on the metrically georeferenced databases, more sophisticated spatial analysis can be performed. In addition, a state of the practice example is the in vehicle navigation systems using global positioning technology to do real time metric georeferencing. The positioning technology obtains coordinates of the current location, which is then georeferenced in the street network GIS database on the fly. Metric georeferencing is also being applied in many other location based services. For example, Global Positioning System (GPS) is the georeferencing technology for geocaching, the popular web based loation finding adventure.

Indirect georeferencing links the commonly available text based place names, codes, and addresses with metrically georeferenced locations or features. It provides a bridge between textual information and spatial location information. After georeferencing the text based attributes, more spatial analysis can be conducted based on the locations. Thus, indirect georeferencing is very useful and indeed widely used. Example applications include georeferencing reported addresses through emergency responses, crime records, real estates, package delivery, market analysis, etc.

Among indirect georeferencing, address matching is particularly popular. However, despite its popularity, there are several noteworthy common issues with it. First of all, sometimes imperfect matching rates are observed. Two possible problems account for low matching rates. One problem relates to possible imprecision of the underlying street databases, and the other is due to inaccuracy of the recorded addresses (referents) for geocoding. If a street goes by several alternative names, the adoption of different names in the database and in the referents could cause unsuccessful matching result. Second, discrete georeferencing is sensitive to spatial and temporal accuracies of the underlying geographic frame of reference. The frame of reference, such as street networks in GIS, may change over time. This can be a particularly critical problem for applications that require such information for planning and emergency response. Accidents that occur on new building sites or new roads cannot be accurately reported and geocoded if the new street information is not available in the database yet.

How to Do It

Continuous georeferencing can be done through various techniques such as field survey, GPS technology, remote sensing, and photogrammetry, obtaining location information from existing georeferenced maps, or obtaining georeferenced digital data directly from public or private sources. Particularly, in many countries, available to the general public are a variety of digital files of georeferenced locations of cultural and physical features. In the United States, for example, it is the federal mandate that each state hosts a website of so called 'GIS data clearing house' for the public to freely access some commonly used georeferenced GIS data. There are also

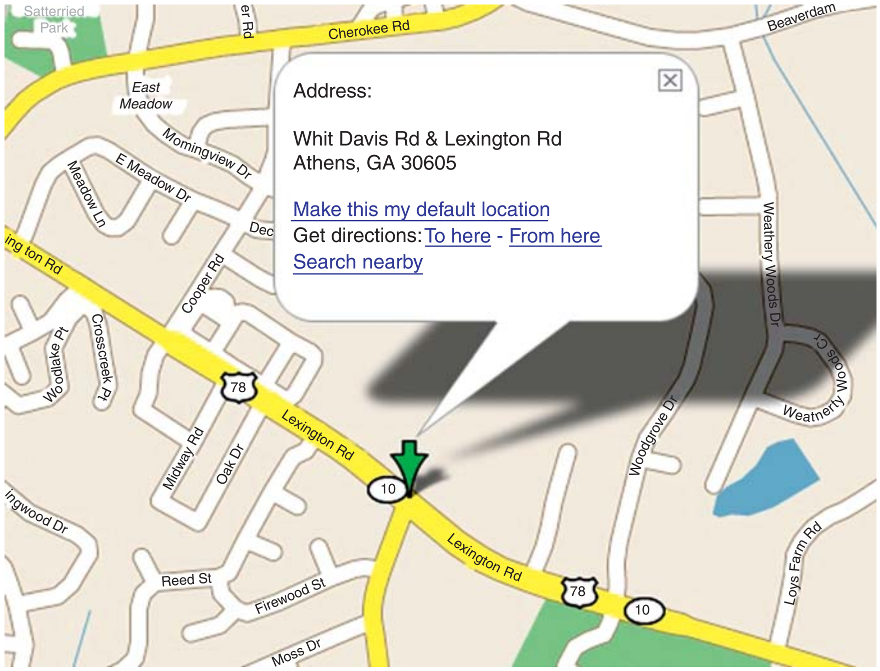

(a)

(b)

Figure 5 (a, b) Georeferencing using qualitative spatial relations.

modern personal digital devices equipped with global positioning capability allowing for instant metric geo referencing.

Discrete georeferencing can be done in a GIS pack age. Conventional geocoding GIS functions are available in most GIS packages. There are also web based location finding services that can be used for discrete georeferencing purposes. Popular services can be found at Google maps, MapQuest, Map on us, and Yahoo maps, to name a few examples. In addition, there are also many online yellow page services for business listing and georeferencing.

See also: Geomatics; GIS, Mobile and Locational Based Services; Global Positioning/GPS; Integrated Spatial Data Infrastructure; Projections.

Further Reading

Cowen, D. J. (1997). Discrete georeferencing. NCGIA core curriculum in geographic information science, http://www.ncgia.ucsb.edu/giscc/units/u016/u016.html (accessed January 2008).

Hill, L. L. (2006). *Georeferencing: The Geographic Associations of Information*. Cambridge, MA: MIT Press.

ISO (2003). *Geographic information Spatial Referencing by Geographic Identifiers. ISO 19112*. Geneva, Switzerland: International Organization for Standardization.

Linear Referencing System (2002). Linear referencing system. *Report of a Pilot Project Jointed Conducted by Natural Resources Canada and Land Information Ontario*, http://www.geoconnections.org/publications/reports/LRS 04 2002 E.pdf (accessed January 2008).

Longley, P. A., Goodchild, M. F., Maguire, D. J. and Rhind, D. W. (2001). *Geographic Information Systems and Science*. West Sussex: Wiley.

Relevant Websites

http://www.maps.google.com
Google Maps (there are many other similar location finding, direction finding, or business finding websites).

http://www.mapquest.com
MapQuest.

http://www.maponus.com
The Easy Way to Find Businesses and People.

http://www.census.gov
US Census Tiger Files, US Census Bureau.

http://www.maps.yahoo.com
Yahoo Maps.

Geospatial Intelligence

K. C. Clarke, University of California at Santa Barbara, Santa Barbara, CA, USA

Glossary

NGA United States National Geospatial Intelligence Agency, the federal agency charged with providing and using geospatial intelligence products to serve a variety of military, civil, and international needs.

NIMA National Imagery and Mapping Agency, Created in 1996, Reformed in 2003 as NGA.

NPIC United States National Photographic Interpretation Center, was formerly a component of the CIA Directorate of Science and Technology and produced imagery interpretation reports.

NRO National Reconnaissance Office. Agency that handled overhead reconnaissance imagery. Merged into NGA in 2003.

TPED Intelligence acronym for standardized procedures of imagery acquisition and use: tasking, processing, exploitation, and dissemination.

The collection of information of use in intelligence for the military and for spying goes back to prehistory. Necessarily, the information includes maps and geographical facts. While in the past, a large variety of media and communications mechanisms were used in military intelligence, after World War II the collection, assimilation, and distribution of intelligence information became highly controlled and increasingly mechanized. Large amounts of geographical data were collected by routine and open means, such as the CIA's World Fact Book. Collection of map information was initially supported by the creation of maps series such as the US Defense Mapping Agency's Joint Operations Graphics charts and the topographic coverage of countries around the world. Activities varied by nation, but in most cases the principal mapping agency in any given country is closely affiliated with the National mapping effort, and these agencies also mapped enemy or potential enemy territory. In the predigital era, this entailed photogrammetry, geodesy, mapping, and charting. After the Cold War, it increasingly included use of data from overhead reconnaissance from aircraft and satellites. As automation of the collection of map and image data became common, greater emphasis was placed on integration of the data, and its interpretation for decision makers. At first, separate agencies collected and interpreted geospatial information, but especially since the terrorist attacks of 11 September 2001 there was an increasing recognition that there were disadvantages to the separate management of intelligence operations.

The term Geospatial Intelligence was introduced in 2003 and formalized in legal code creating and tasking the United States National Geospatial Intelligence Agency (NGA). The goal of the designation was to create a new form of INT or Intelligence, on par with the existing human intelligence or spying (HUMINT), signals intervention and decoding (SIGINT), and other forms of intelligence analysis. NGA defines geospatial intelligence, or GEOINT, as "earth referenced information about natural and man made objects or events with National Security implications", and also as "the exploitation and analysis of imagery and geospatial information to describe, assess and visually depict physical features and geographically referenced activities on the Earth that have national security implications". These were the motives behind the creation of NGA from the former National Imagery and Mapping Agency (NIMA), itself created to absorb in entirety the former Defense Mapping Agency, the Central Imagery Office, the Defense Dissemination Program Office, and the National Photographic Interpretation Center. NIMA also incorporated elements of the Defense Airborne Reconnaissance Office, the National Reconnaissance Office, the Defense Intelligence Agency, and the Central Intelligence Agency. A consequence was the merging of two 'cultures', a group that had trained and worked with overhead satellite imagery and another group trained as traditional paper cartographers for maps and nautical charts.

The creation of GEOINT and the NGA was a conscious effort to remove the historical divisions, a necessary step toward moving forward with an all digital geographic information system and cartographic environment within the United States intelligence agencies. The term was adopted in the NGA's name both to recognize the fundamental property of geographical location associated with the data that NGA and the intelligence community produces and analyzes, and to emphasize the value added analyses that the agency performs to create GEOINT as a distinct type of actionable intelligence. A similar emphasis can be seen at Australia's Defence Imagery and Geospatial Organisation and the United Kingdom's National Imagery Exploitation Centre.

The prior model of geospatial data, rather than intelligence, centered on map layers or themes, the logical equivalent to map series. From the historical perspective, the goal was to create and supply information largely independently for maps and charts, geodesy and geophysics (measurement and modeling of the Earth's shape and gravity), precision location (usually with global positioning systems), and photogrammetry. The merging and overlay of the data required advanced geoprocessing, remote sensing and analytics, including time–space an alysis, and analysis of the behavior and interaction of natural and human phenomena at the Earth's surface. Ensuring an ongoing supply of these themes required efficient TPED, that is, tasking, processing, exploitation, and dissemination of data, over vast geographic areas and at the time intervals of interest. Tasking implies sched uling of data collection and data acquisition, processing includes data integration, georectification, and regis tration, exploitation means analysis, and dissemination is routine, that is, to all who require the data or the ana lytical results, and to the databases where the data are to be archived for potential future reuse, for example, in change detection.

This model sufficed for geospatial intelligence de mands during the Cold War. However, recent changes in the demands for intelligence have meant that the tradi tional division of labor works against the contemporary need for speed, accuracy, and synthesis, hence 'action able' intelligence. It was this gap that geospatial intelli gence as a new form of intelligence, and not as a set of data flows and standard analytical reports, was intended to fill. Consequently, the driving force behind the cre ation of GEOINT is the increased need for integration. Geospatial intelligence has been called "a powerful new analytic discipline – the product of increasingly complex sources which, together, are greater than the sum of the parts. GEOINT is emerging as the most valuable tool for envisioning and predicting activity around the world. It serves everyone – from the White House to the pilot house, from the Situation Room to the ready room" (Schultz, 2004). The director of NGA has stated, "GEOINT is about more than pictures. GEOINT makes possible in depth assessments and judgments based on the information that is gleaned from visual depictions. In short, GEOINT is more than imagery, maps, charts and digital displays showing where the bad guys are. GEOINT at its best is the analysis that results from the blending of all of the above into a dynamic, composite view of features or activities – natural or manmade – on Earth" (Clapper, 2004).

The call for synthesis has been echoed in academic geography. Many geographical approaches use scale as an integrating factor, arguing that all scales impact all oth ers. In Geographic Information Science (GIS), standards based interoperable systems are becoming commonplace, usually linked to powerful databases via the Internet. Within information science, new methods and theory of data fusion, knowledge discovery, and data mining allow search tasks with little human intervention. As the geo graphical sciences move beyond the analytical and compartmentalized approach inherent in most early GIS analysis, such as features, layers, themes, overlays, etc., the geospatial intelligence vision becomes more feasible and more useful for geography. In its narrow sense, as a way to do synthesis driven intelligence work, GEOINT is already the focus of considerable research and funding. A challenge will be adopting the broader views of the geographical tradition to the new generation of methods, tools, and approaches that GEOINT is already beginning to create and nurture.

See also: Cold War; Maps; Mapping, Topographic; Remote Sensing.

Further Reading

Clapper, Jr., (Ret.) Lt. Gen. James R. (2004). Imagine the Power of GEOINT, Pathfinder, vol. 2, no. 1, May June 2004. National Geospatial Intelligence Agency.

Clapper, Jr., (Ret.) Lt. Gen. James R. (2006). Progress in Defining GEOINT, Pathfinder, vol. 4, no. 2, March/April 2006. National Geospatial Intelligence Agency.

National Research Council (2006). *Priorities for GEOINT Research at the National Geospatial Intelligence Agency, Board on Earth Sciences and Resources*. Washington, DC: National Academies Press.

Schultz, M. (2004). Letter to Our Readers, Pathfinder, vol. 2(1), May June, 2004. National Geospatial Intelligence Agency.

Relevant Websites

http://www.nga.mil
National Geospatial Intelligence Agency.

http://books.nap.edu/catalog.php?record id=11601
National Research Council. Priorities for GEOINT Research at the National Geospatial Intelligence Agency.

http://www.usgif.org
United States Geospatial Intelligence Foundation.

Geovisualization

M. J. Kraak, International Institute of Geo-Information Science and Earth Observation, Enschede, The Netherlands

Glossary

Cartography The art, science, and technology of making and using maps.

Geovisualization A loosely bounded domain that addresses the visual exploration, analysis, synthesis, and presentation of geospatial data by integrating approaches from various disciplines including cartography with those from scientific visualization, image analysis, information visualization, exploratory data analysis, visual analytics, and GI Science.

Map A symbolized representation of a geographical reality, representing selected features and characteristics, resulting from the creative effort of its author's execution of choices, and is designed for use when spatial relationships are of primary relevance.

Visual Thinking The thinking process that results from interacting with graphic representations such as maps and diagrams.

Introduction

Maps are familiar to all of us. Everyday decisions are made based on maps and the geographic information they represent. This varies from simple route finding from A to B, to locating an event from the news, or studying a weather map. In a professional environment the examples are more complex and might include planning a new railroad, assisting during an earthquake disaster, or understanding the global spread of the bird flu. Maps offer us insight in our geoenvironment, because they have the ability to present, synthesize, analyze, and explore the real world. Maps do this well because they only present a selection of the complexity of reality and visualize it in an abstract way. The cartographic discipline has developed a whole set of design guidelines to realize the most suitable map that offers insight in spatial patterns and relations. In general, professional geographers don't have to be convinced of the unique qualities of maps to express their ideas, to make a point, to obtain new knowledge, communicate with colleagues, and of course to orientate and navigate. Maps are also appreciated outside the professional community. Maps have been used throughout history and across all cultures in day to day activities.

Recent technological and societal developments have acted as a positive stimulus for map creation and use, both for professional and for nonprofessional users. The recent success of products like Google Maps and Google Earth has demonstrated this. Both products combine an intuitive interface with a good map design, a wide availability of data, and the possibility of linking your own data to the maps and satellite imagery. It has resulted in a tremendous increase in the number of maps produced and disseminated, and allows the wider public to participate in the mapping process. In addition, maps that used to be complicated to produce can now be created to present many alternative views by the single click of the mouse.

However, this last aspect is not always realized because the view on maps is often limited to its presentation function. Even the professional geographer might not be fully aware of the possibilities. Although it has to be realized that the traditional role of a map to 'present' is still very much valid, today's challenge is to see the map as a flexible interface to geospatial data, since it offers interaction with the data behind the visual representation. The interactive and dynamic environments in which maps are used, often in combination with other graphic representations, encourage exploration. As such they are used to stimulate (visual) thinking about geospatial patterns, relationships, and trends. The contexts where maps like this operate have led to an extended view on cartography, and can be deemed the world of geovisualization. This can be described as a loosely bounded domain that addresses the visual exploration, analysis, synthesis, and presentation of geospatial data by integrating approaches from disciplines including cartography with those from scientific visualization, image analysis, information visualization, exploratory data analysis, visual analytics, and GI Science.

Nature of Geovisualization

The above developments in the integration of methods and techniques used by different disciplines is a logical step, especially because many challenges being faced in the natural and social sciences have an inherent complexity and interdisciplinary nature. Examples include: how to assess the vulnerability of regions and their human populations to global environmental change; how to measure and sustain biodiversity; how to predict and cope with changing disease incidence patterns; and how to manage the increasing traffic flows of cities more effectively. For each of these problems, georeferencing provides a fundamental mechanism for linking the diverse forms of data needed to attack these problems. Currently no fully developed, integrated models of complex human and environmental systems exist. The international research community tries to come to solu tions from different perspectives. The geovisualization perspective has the potential to provide 'windows' into the complexity of phenomena and processes involved, through innovative scene construction, virtual environ ments, and collaboration, thus prompting insight into the structures and relationships contained within these complex, linked data sets.

The nature of geovisualization research is wide ran ging and in the research agenda established by the inter national cartographic community, particular areas have been identified. These focus areas are closely interrelated and should be seen in connection with each other. They include representation, visualization–computation inte gration, interfaces, and cognitive/usability issues.

Representation

In the past, cartography had a more or less fixed set of options available for a correct graphic representation of any type of geographic data. Today technology offers many new extensive opportunities for which the appli cation of existing methods are less certain. Examples are animation, hyperlinking, immersive environments, and multimodal interfaces. Despite these advanced technol ogy driven options, the available software does not cater for all possible map types available, because these require typical human creativity to create. Interaction and dy namics are today a default component of the represen tation tool box. In addition, the amount and complexity of the data to represent is increasing. Large multivariate spatiotemporal data set, including the third dimension as well as time, are available and have to be understood. Based on these observations, there is a need to consider conceptual and technological issues:

- to develop a theory for georepresentation and for malizing representation methods,
- to develop new forms of representation that support the understanding of geospatial phenomena and space–time processes,
- to adapt representation methods to meet the changing nature of data to be represented,
- to adapt representation methods to the increasing range in kinds of tasks that visual geospatial repre sentations must support, and
- to take advantage of recent (and anticipated) techno logical advances in both hardware and data formats.

Interaction Forms and Interface Design

In a visual environment the interface is most important. The users should intuitively be able to execute the tasks they have in mind. This is not trivial, because the geo visualization environments are often composed of mul tiple coordinated views, each of which give a different visual perspective on the data at hand. Changes in one view will result in corresponding changes in other views. Next to the use of the large screens that support these coordinated views, the trend of small screens on all kind of mobile devices can be witnessed as well. To make it all work one has to

- extend our understanding of interface design to take advantage of the potential of virtual environments,
- develop and assess formalizations for specifying interface operations appropriate to geovisualization environments,
- develop a comprehensive user centered design ap proach to geovisualization usability,
- develop the understanding and mechanisms for cap italizing on the potential of geovisualization to prompt creative thinking,
- extend our understanding of metaphor for geovisua lization, and
- investigate interfaces to support the digital earth and the concept of digital earth as an interface.

Integration of Visualization and Geocomputation

Although a visual approach to problem solving is very strong, it cannot do without computational support. Most graphic actions that are initiated require computational support, and again graphics are needed to get an insight in computational results. What needs to be done is to

- address the engineering problems of bringing together disparate technologies,
- explicitly incorporate the location and time com ponents of multivariate data within visual and ana lytical methods,
- represent geographic knowledge,
- incorporate geographic meaning within visualization environments,
- develop visual approaches to geospatial data mining, and
- integrate visual and computational tools that enable humans and machines to collaborate in knowledge construction.

Cognition and Usability Issues

It is good to see how new tools can be used in one's discipline, but it is even better to know if those tools work – in other words, to consider the implications of new forms of representation and interaction. Usability studies have to be developed to see when and why the tools do work and for whom. Here the geovisualization community can draw upon experiences in the field of human computer interaction. This requires:

- development of cognitive theory to support and assess geovisualization methods in VE & dynamic display,
- development of an integrated understanding of metaphors and knowledge schemata for geovisualiza tion interface design,
- understanding individual and group differences re lated to use and usability of geovisualization,
- extending our perspective on cognitive and usability issues associated with geovisualization to contexts involving group work,
- determining contexts in which geovisualization is successful, and
- developing methods and tools that will enable the kinds of cognitive and usability research called for.

The four focus areas are interdependent, and are all re lated to integrated technologies that take advantage of the potential offered by increasingly experiential repre sentation technologies (multimodal maps), to the meth ods and tools that enable understanding of and insight into the increasingly larger and more complex geospatial data sets, and to a new generation of geovisualization methods and tools that support group work, and human centered approaches to geovisualization.

Alternative Visual Representations

In a geovisualization environment, maps are used to stimulate (visual) thinking about geospatial patterns, re lationships, and trends. One important approach here is to view geospatial data sets in a number of alternative ways, for example, using multiple representations with out constraints set by traditional techniques or rules. This should avoid the trap that most researchers tend to rely on, that is, well worn procedures and paradigms. Instead they should realize that creative discoveries, in both art and science, often occur in unusual situations, where one is forced to think unconventionally. In these circum stances it is better to remove mental roadblocks and create some distance from the discipline in order to re duce the effects of traditional constraints. One can choose from many different alternative mapping methods. However, it is not just maps one can use. Other inter active graphics like diagrams and georeferenced imagery should be considered, especially because of the large data set one has to deal with.

Many alternative mapping methods are familiar to frequent map users. When it comes to traveling, most of us use public transport maps. Their task is to guide us from A to B, and since one does not have to worry about driving, lots of (traffic) information can be omitted. This can lead to interestingly alternative map designs. The best known example is the map of the London Under ground originally created by Beck in the early 1930s. The concept of this map has been copied by many other cities, although not always successfully. While designing the map, Beck decided to schematize the route network, as the map image of the city center became too cluttered. The traveler would, according to his opinion, not be hindered by this approach since it was dark underground anyway. The London Underground website has many examples, and one can in website allows one to interact with the network and request the real time situation regarding delays, plan a tour, and buy tickets.

Travelers to mountainous areas will be familiar with the panorama map. Most tourist areas can no longer do without them. Often a summer and winter version of the map, each highlighting the characteristics of the re spective seasons, is available. The winter version high lights the magnificent sky slopes while the summer version concentrates on the fine overall view.. These maps are highly appreciated, and despite the fact that the images are geometrically distorted, they give most vis itors a better impression of the area than topographic maps would. Google Earth can be considered an inter esting alternative, since it allows one to combine different graphics representations on top of the satellite imagery draped over the terrain. This might help in the inter pretation of the landscape.

Revealing Patterns: Understanding History

To illustrate the geovisualization approach, concen trating on representational aspects, Charles Minard's well known map of Napoleon's 1812 campaign into Russia (see **Figure 1**) will be discussed. The map por trays the dramatic losses of Napoleon's army during his Russian campaign. The author's purpose of both maps was to stress the senselessness of war. This section illus trates the argument that if one is able to look at the data from different perspectives, for instance via alternative map views, sometimes in combination with other graphics such as diagrams, graphs, or even photographs and videos, one will better appreciate the nature of the data at hand. The map is rather well known and has been mentioned by many researchers earlier. It has been claimed that "it may well be the best statistical graphic ever drawn," and "a narrative graphic of time and space which illustrates how multivariate complexity can be subtly integrated … so gentle and unobtrusively that viewers are hardly aware that they are looking into a world of four or five dimensions."

The Minard map shows several 'variables'. As with any map, there is location. Next to the major paths, some minor ones are depicted as well. Linked to the retreat path (the black solid band) is a diagram indicating tem perature. Additionally, the map shows the size of the

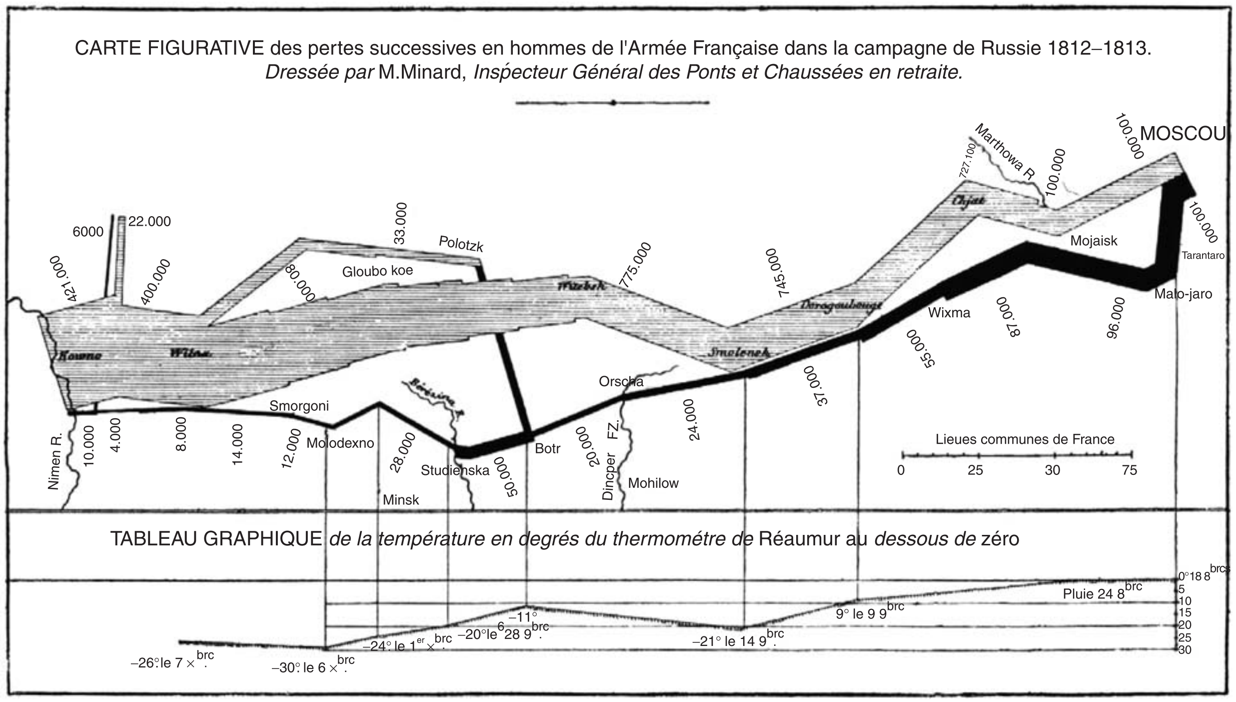

Figure 1 Minard's map from 1861 'Carte figurative des pertes successives en hommes de l'Armee Française dans la campagne de Russie 1812 1813', or 'Napoleon's March on Moscow'.

army by numbers and by the width of the advance and retreat bands. Names indicate major battles and important geographic features. Time is inherent in the map (it shows a clear distinction between the advance (going east) and retreat path (going west)), but absolute indications are given only in the temperature diagram.

Traditionally, cartographers have three options to display geospatial data with a temporal component. These are a single (static) map, in which specific graphic variables and symbols are used to show change and represent events. Minard's map is a good example of this category. The second option is a series of (static) maps, sometimes called a small multiple (**Figure 2a**). The single maps represent snapshots in time, and together the maps make up an event. Change is perceived by the succession of the individual maps depicting the event in successive snapshots. It could be said that the temporal sequence is represented by spatial sequence, which the user has to follow, in order to perceive temporal variation. Finally, one can create an animated map. Here change is perceived to happen in a single image by displaying several snapshots one after the other. The difference with the animation from the static series of maps is that the variations introduced to represent an event do not have to be deduced from a spatial sequence, but from real movement on the map itself (**Figure 2b**).

Obviously, Minard's map includes a temporal component; however, if one were to ask "what is the situation on August 24?", the answer is not straightforward. Even if one could locate that moment in time, all other campaign details remain visible and may cause confusion. Alternatively, adding a slider to the map would introduce time more explicitly, and allow the user to define the progress of the campaign. However, in this process there is no link between world time and display time. Also, moving the slider with regular intervals would not result in a regular passing of time because Napoleon remained at certain places for a longer period. For instance, in Moscow he paused for more than a month – that is, information that cannot be derived from the original static map. Some of the animations have similar problems because frames exist only for the moment in time that the change is registered. The use of a series of maps (or a small multiple), each representing a particular moment in time could be the alternative. In **Figure 3**, two alternatives are given, each showing a different (traditional) cartographic solution. The maps depict the position of individual French troops on July 24 and August 24. The viewer is not distracted by previous or future moments in time. Additionally, it provides (just as in the original map) an overview of the whole campaign. The path of the campaign passed up to the particular date has been highlighted.

Interesting alternative visualizations have been produced by other researchers. They used Minard's map to demonstrate the capabilities of a scientific and information visualization software. Their products are examples of visualizations not influenced by traditional cartographic rules. **Figure 4** presents two snapshots of the visualization produced by the SAGE software. In the snapshots the map is linked to diagrams, a principle Minard also followed in 1861.

The maps in both snapshots are oriented north, with the longitude indication along the horizontal axis. In both diagrams the horizontal axis represents time. In the upper snapshot of the diagram, the vertical axis represents longitude, and in the lower snapshot it represents the number of troops. These combinations reveal some interesting facts when compared with the original map in **Figure 1**. The diagram in the upper snapshot reveals two battles took place at Pollock (located at the northern path in the map and circled in the diagram) instead of one. The gap in the diagrams in both snapshots shows that Napoleon stayed for a month in Moscow before returning west, information not found in the original map. These examples show that alternative views can be revealing and clarifying.

What if another dimension is added? **Figure 5** presents a 3D view in which the height of the columns of the path segments represents the number of troops. The crossing of the River Berezina is highlighted. Napoleon lost half of his remaining troops during this crossing. This information is also found in the original map, but the 3D view presents this much more dramatically. Other variables such as temperature can be represented in 3D view as well. The columns can be colored depending on the temperature, applying color schemes with blue for cold and red for warm. Links with diagrams such as shown in **Figure 4** are also possible, including the interaction options presented. The ability to manipulate the 3D scene in space to change the perspective view is a prerequisite, since many interesting facts might be hidden behind objects. If one would have looked at the River Berezina crossing from the north, it would have been hidden by the troops columns representing the advance. Additional layers with other information can be stacked below or even above the campaign information.

Based on terrain heights, a fly through can be generated demonstrating the impact of the terrain on the campaign. In this particular case, it does not result in a spectacular view since the landscape Napoleon passed by is relatively flat. Google Earth would be an apt program

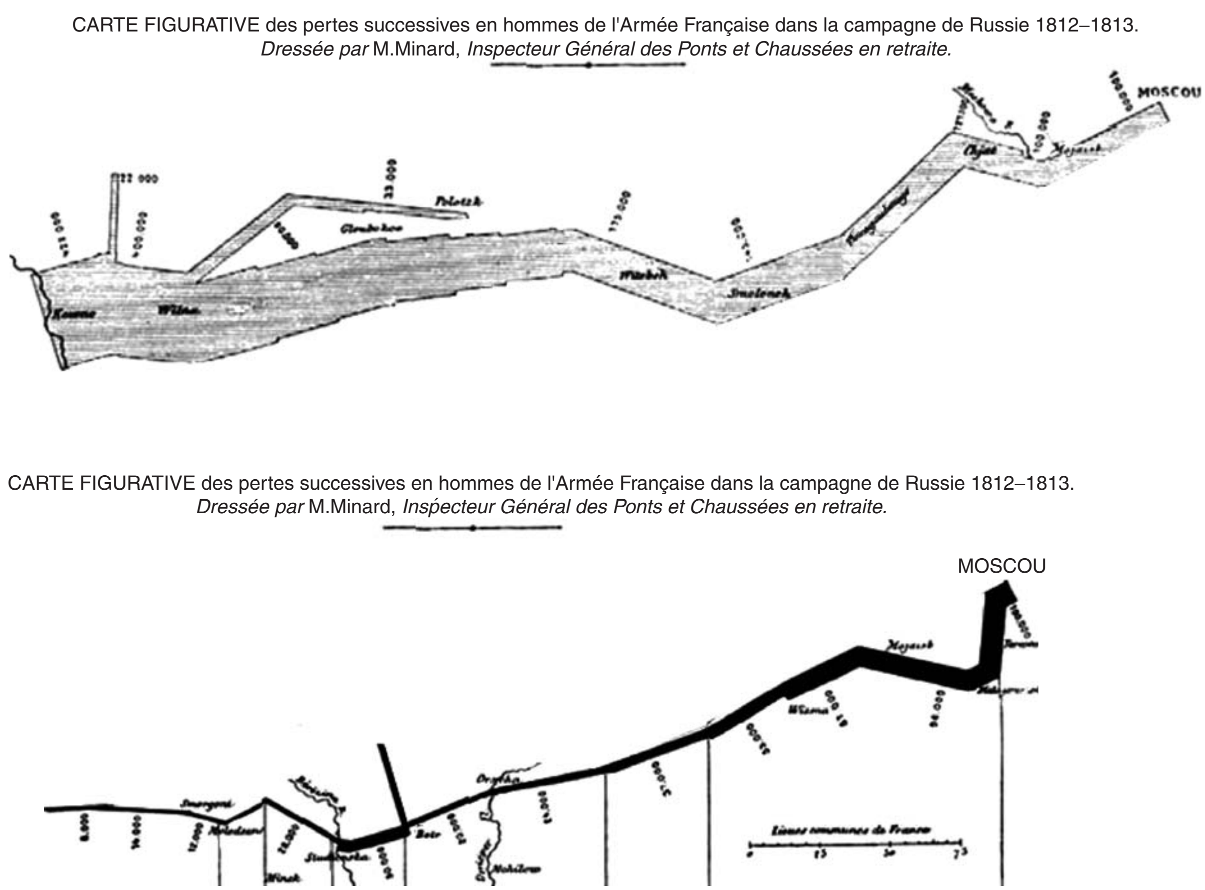

(a)

Figure 2 (a) Minard's map as a set of static maps; (b) two snap shots from an animated Minard map.

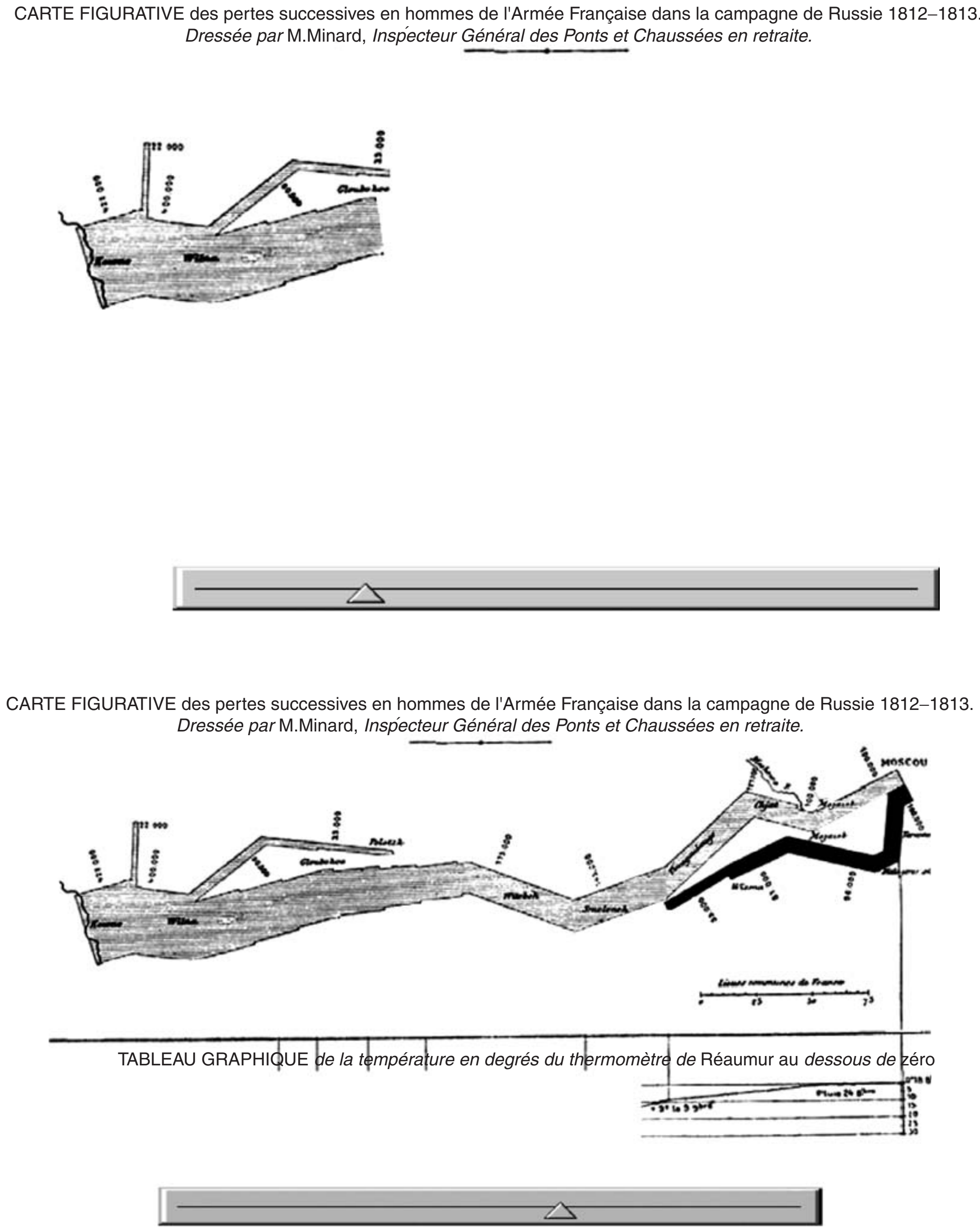

(b)

Figure 2 (Continued)

to use in this case. Of course, one has to realize that land use patterns would have changed considerably in some places over the last 200 years. **Figure 6** displays the path some of Napoleon's troops took, projected on today's satellite imagery.

Alternatively, software exists which allows one to dress the landscape randomly and apply light conditions, based on season and time of the day. Even a snow cover can be simulated. One has to realize the application of the rules of nature results in a fictional world, even if real topographic data has been used. Application can be found in virtual reality environments. For example, in a two dimensional animation one can change the background of the landscape depending on the temperature from blue in winter to green in summer. **Figure 7** shows how the background color of the map has been used to indicate the temperature during the campaign.

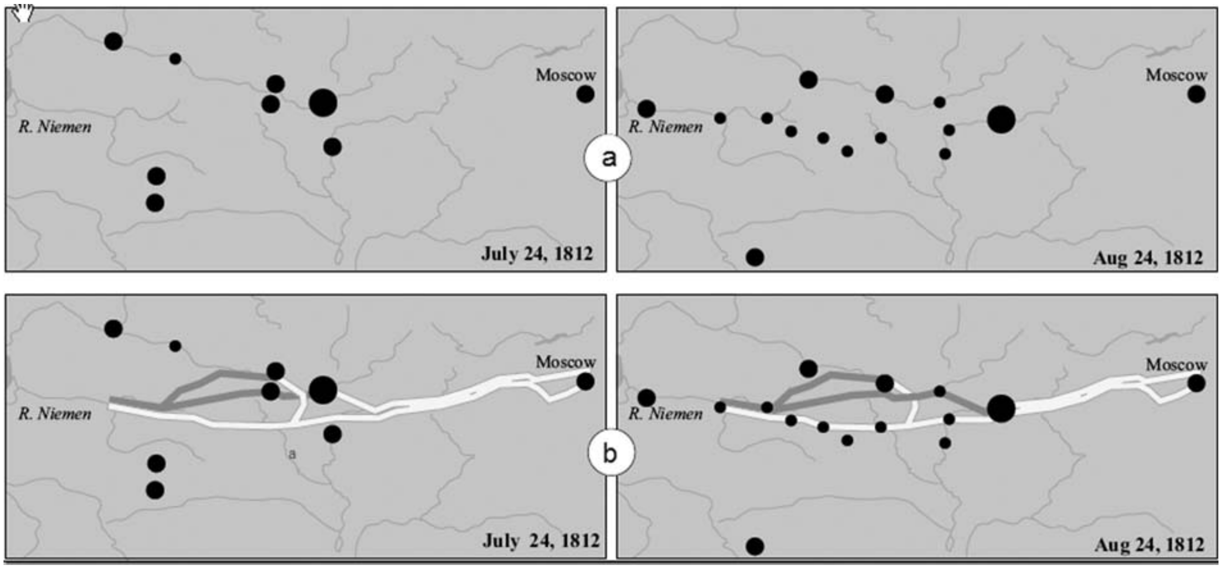

Figure 3 The campaign in small multiples. Here two pairs of two snapshots with the location of the French troops on 24 July and 24 August 1812 are depicted. Each pair shows an alternative cartographic representation of the data.

The third dimension can be applied from a completely different perspective. **Figure 8** presents a so called space–time cube in which the x and y axes represent the geography and the z axis represents time. Again, this solution would benefit greatly from interactive options to manipulate the viewer's perspective of the cube. An additional option could be slider planes along each of the cube's axes. The user can move them through the cube and as such can highlight a time period or location. One could, of course, also change the type of data represented along the axes and, for instance, create temperature versus troops versus time.

Revealing Patterns: Predicting the Future

For the study of dynamic phenomena, such as the weather, and specifically cloud developments, satellite images are the main source. These image repositories are becoming the fastest growing archives of spatiotemporal information and the users are confronted with this flow of data that need to be explored. Here visual exploratory tools are also useful. To predict storms and severe weather conditions, images are searched to look for convective clouds. How are the dynamics of cloud development displayed? An immediate response might be interactive animation. However, despite being interactive, animations have some drawbacks and can easily lead to information overload, limiting their exploratory usage. An additional alternative view that allows the user to reduce the information load and focus attention, could enhance understanding.

Next to the animation the visualizations of clouds can be presented in an abstract way, based on quantitative data, enable essential attributes to be mapped, and have dedicated interactivity. The visual exploration is improved by coupling computational methods – here used to detect and track clouds – with abstract and selective visualizations of the tracked information in a single environment (**Figure 9**). Tracking reduces the complexity of time series data sets: each cloud feature is described in terms of its attributes (position, size, image intensity, etc.) and its lifetime. This offers quantitative, abstract, and richer graphical representations of each cloud's evolution. In particular, the proposed multiple views in combination with the traced cloud paths can show the essence of the object evolution and history. Abstract representations help the user to search for objects of interest (clouds with negative temperature gradients are most likely to be convective and selective interactivity assists the user in focusing the attention on the particular clouds). With the presented functionality, the lower level visual task of detecting and tracking dynamic clouds is performed computationally. Vision can now be completely devoted to a higher level visual task. So, the meteorological scientists can pay more attention to exploration of convective clouds and generating hypotheses about their dynamics and evolution.

Challenges

The examples in the previous section demonstrate that visual support in any phase of the geospatial data handling process can be realized and is helpful. However, they are mostly individual solutions.

For a better integration, mechanisms must be provided that can visualize the connections between the

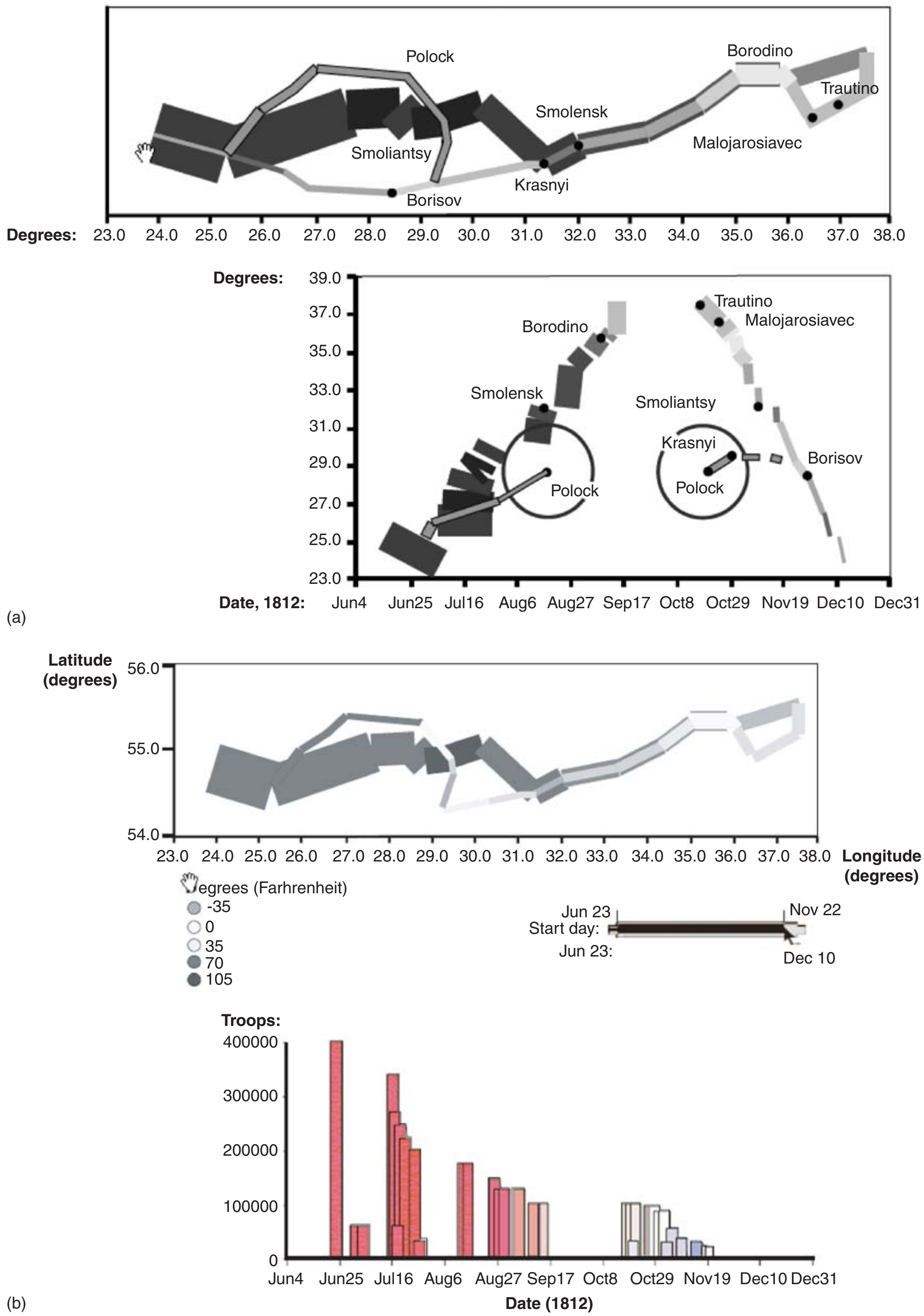

Figure 4 The campaign data in multiple linked views. (a) Linking the map with a longitude-time diagram, and (b) linking the map with an attribute-time diagram. (Based on Roth, S. F., Chuah, M. C., Kerpedjiev, S., Kolojejchick, J. A. and Lucas, P. (1997). Towards an information visualization workspace: Combining multiple means of expression. *Human-Computer Interaction Journal* 12(1&2), 131 185.)

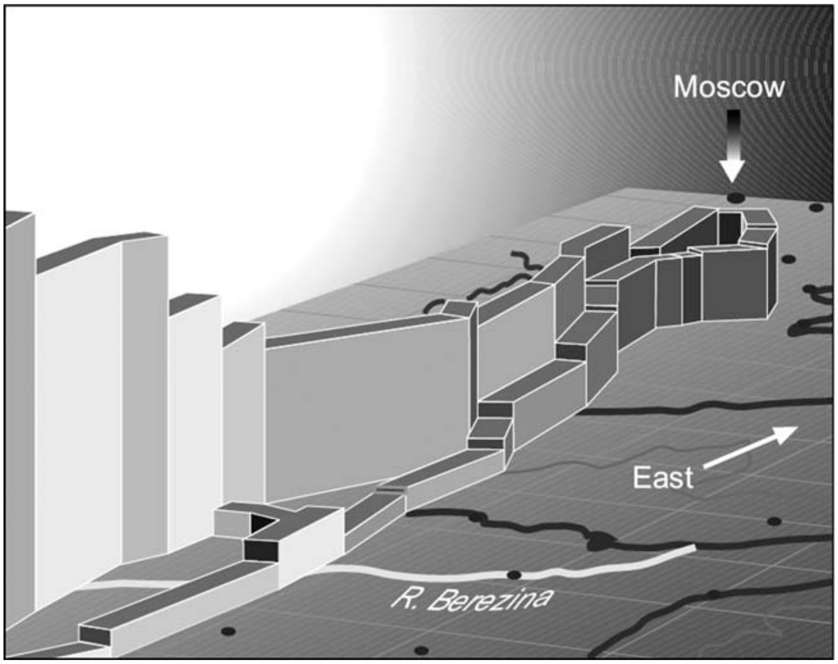

Figure 5 Three-dimensional view of the size of Napoleon's troops during his Russian campaign in 1812.

(a)

(b)

Figure 6 Napoleon's campaign mapped on Google Earth imagery.

Figure 7 The temperature during the campaign: (a) a green map background indicating relatively warm weather; (b) the map with a white background suggesting very low temperatures.

various stages of analysis and show how concepts relate to data, how models relate to concepts, and so forth. Today, one cannot find a single program to execute all the visualization operations discussed. Such functionality is needed, but what is debatable is if we really require this in a single package. With geodata infrastructure approach in the future and its distributed geoservices, it's more important that one has access to (visualization) functions and data to create the representations required. An important precondition is, of course, interoperability, for example, so that the smooth exchange of data and functions is possible. Currently, these requirements are not yet fulfilled. Standardization initiatives from the Open Geospatial Consortium and work executed following the principles of open source software development will bring these requirements closer to realization. Reaching this goal will facilitate collaboration in solving geoproblems. This approach has become more common in academia; however, many interesting and advanced exploratory geovisualization tools are found in experimental lab environments. These solutions are often open for others to access and apply, but interested researchers as well as other geoprofessionals might find they aren't always easily adopted. What are the future concerns to be addressed?

Stabilization

It is needed to incorporate the new methods and techniques in existing (cartographic) theories. This process is always far behind and problematic. Those active at the frontier will make use of new trends and the gap between established practices might widen, while technological progress is faster as ever. Examples are efforts focused around portable and wearable PDAs and other communication devices that draw upon real time personal global positioning. Such exciting developments are likely to continue to attract significant attention in the near future. This location based personalization re emphasizes one of the original defining themes of visualization as a form of map use – the map devoted to particular needs and designed based upon these requirements. Advances in communication technologies mean that users of such maps can now be loosely connected through phone and wireless networks providing universal access to customizable, map based interfaces to geospatial information that may be location and context sensitive. It is interesting to conjecture if geovisualization developments are motivated by known requirements and the need for solutions, or developing possibilities and the desire to explore and apply them. This can be considered a cause of concern between approaches that are initiated by user needs and technological opportunities respectively and draws attention to geovisualization research initiated by 'demand' as opposed to 'supply'.

Outreach

Next to an orientation toward 'technology oriented disciplines' one has to strive to apply the geovisualization

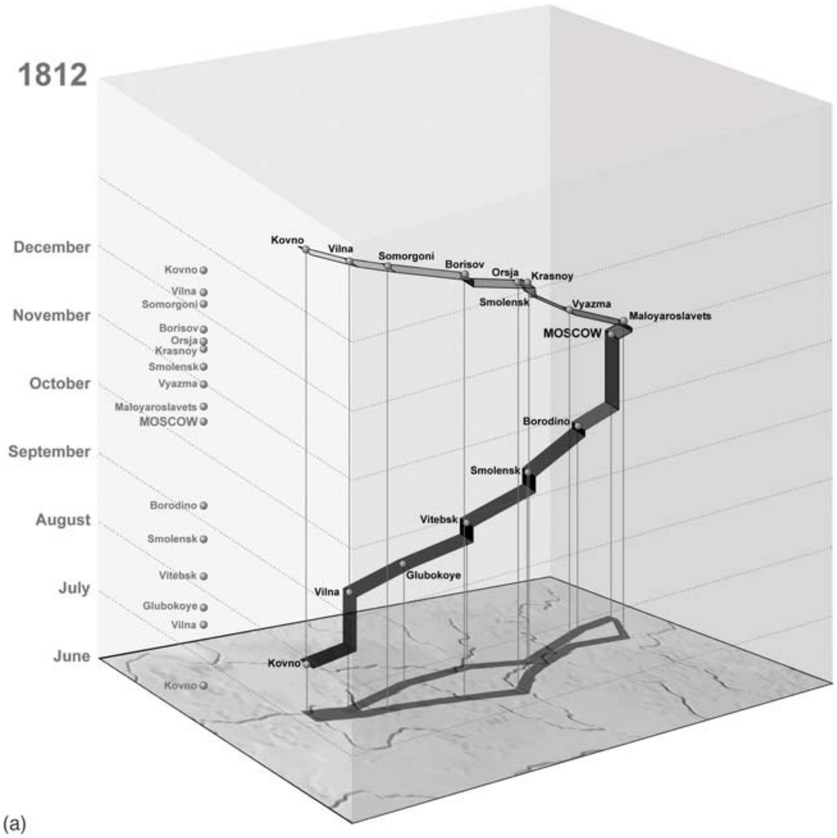

(a)

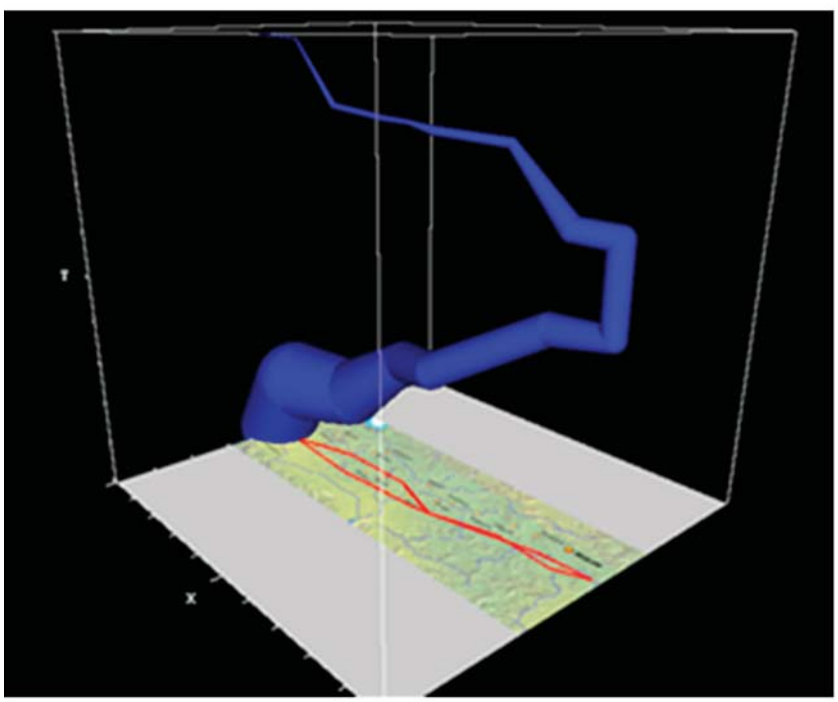

(b)

Figure 8 A space time cube of Napoleon's march in Russia: (a) the path; (b) the path and number of troops.

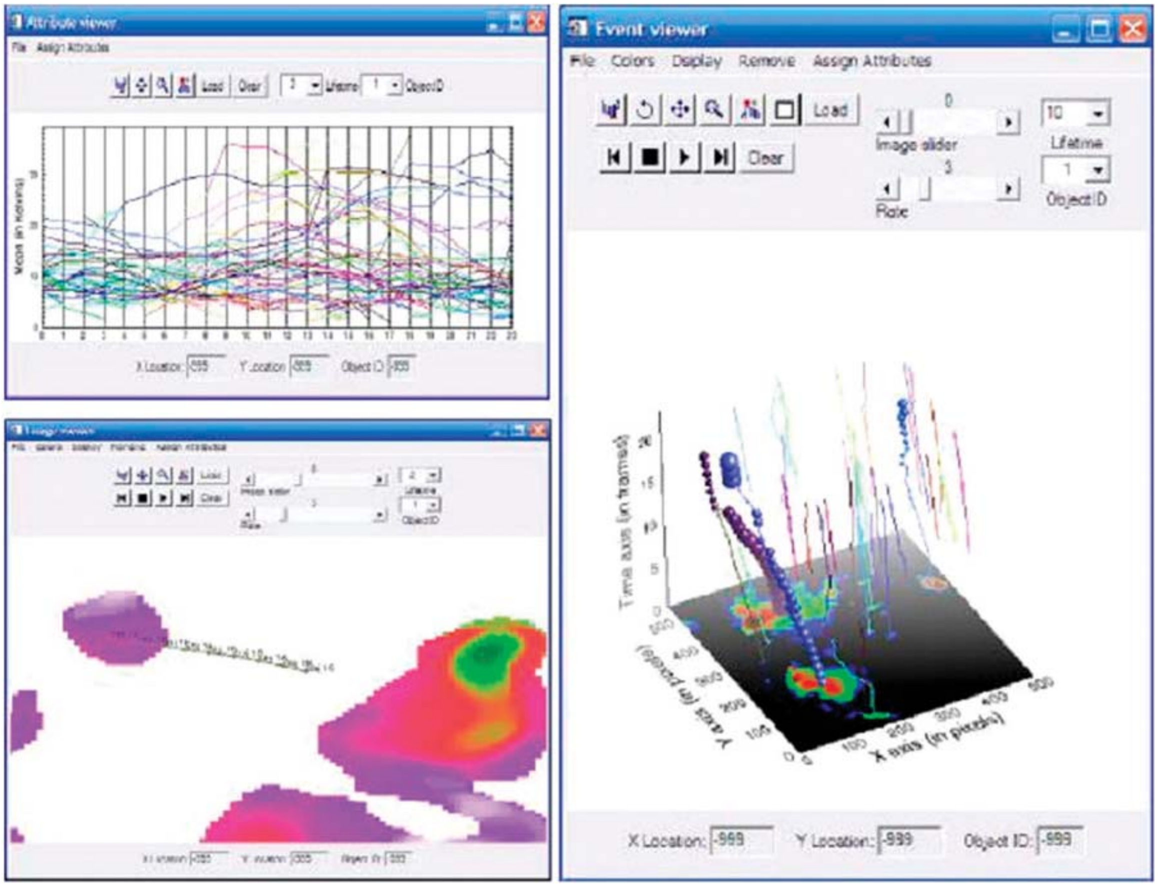

Figure 9 Multiple linked views to study the spatiotemporal characteristics of dynamic phenomena. Right: a view of the space time-cube based event graph showing the evolution of the tracked objects. The lower left view is the (animated) image viewer with graphics to focus the user's attention to the object(s) of interest. In the upper left view the attribute graph displays various characteristics of the tracked objects such as the rate of change of area in time.

knowledge in application fields. A useful example is disaster management, where people might benefit from the tools during their activities before, during, and after disaster management. In other words, geovisualization tools must be used by people solving the large geopro blems. User expectations also rise, and not only by 'Google Earth type' of developments. This widespread adoption of techniques initially used to further scientific advancement to meet the information needs of a far wider range of map users, means that the design process must consider and cater for a far more extensive and heterogeneous user group than the 'maps for experts', typical of early geovisualization efforts. It is interesting to note that geovisualization techniques are also applied by other disciplines to nongeographic data. The resulting graphics are called spatialization. In these situations the map is used as metaphor to get insight in the data.

Incorporation

The geodata infrastructure plays a more prominent role in offering data and services. How about specific visualization services where the user could select representation to execute a process as displayed in **Figure 9**? In addition, one can witness real time data collection and processing of large volumes of data via remote sensing or sensor web enabled collecting devices in the field. These can generate new and fresh ideas about the role of the map. In this light the incorporation of new technology in geovisualization remains important; incorporation of geovisualization in the geodata infrastructure is important as well. With the trend of geovisualization moving from the realm of experts well into the public domain, it has progressed from a focus on the individual (user, tool, data set) to support collaborative and distributed geovisualization which is facilitated by geoservices available via the geodata infrastrucuture. With distributed geovisualization, the inputs can come from many sources; geovisualization resources can be built from distributed components; and multiple individuals can collaborate with and through the resulting geovisualization tools. Geocollaboration faces some pragmatic problems. Collaboration could be in the same place at the same time or at the same place at different times. However, it is also possible to work at different places at the same time or at different places at different times. This will put specific constraints on the software, and data and version man agement. Problems like which of the collaborators is in charge, and how to deal with access right to the data etc. might surface.

Collaboration

This topic follows the current policy of the Commission on Visualization and Virtual Environments to import and export ideas from other disciplines. In the new field of visual analytics, one is trying to improve insights inherent in large volumes of data, which might sound familiar in geovisualization. The research agenda of this field is stimulating analytical reasoning, the creation of new visual representations and interaction techniques, as well as new data representations and transformations, production, presentation, and dissemination techniques. As in infor mation visualization, the data at hand is not necessarily geographic in nature, but methods and techniques can be exchanged. Geovisualization, which see the maps as an instrument of visual thinking, can contribute the geo graphical component in this field of visual analytics.

Conclusions

The geovisualization approach offers a fresh look on geographic data, because it allows maps and other graphics to operate in an interactive and dynamic environment.

Maps retain their traditional roles such as presenting geographic facts, but they should also be seen as flexible interfaces to geospatial data; they do offer interaction with the data behind the representation. This makes them instruments that encourage exploration and are used to stimulate (visual) thinking about geospatial pat terns, relationships, and trends.

See also: Georeferencing, Geocoding; Geospatial Intelligence; Maps.

Further Reading

Andrienko, N. and Andrienko, G. (2006). *Exploratory analysis of spatial and temporal data A systematic approach*. Berlin: Springer Verlag.
Dykes, J., MacEachren, A. M. and Kraak, M. J. (eds.) (2005). *Exploring geovisualization*. Amsterdam: Elsevier.
Hearnshaw, H. M. and Unwin, D. J. (eds.) (1994). *Visualization in geographical information systems*. London: John Wiley and Sons.
Kraak, M. J. (2003). Geovisualization illustrated. *ISPRS Journal of Photogrammetry and Remote Sensing* 57(1), 1 10.
Kraak, M. J. and Ormeling, F. J. (2003). *Cartography, visualization of geospatial data*. London: Prentice Hall.
MacEachren, A. M. (1994). *How maps work: Representation, visualization, and design*. New York: Guilford Press.
MacEachren, A. M. and Kraak, M. J. (2001). Research challenges in geovisualization. *Cartography and Geographic Information Systems* 28(1), 3 12.
MacEachren, A. M. and Taylor, D. R. F. (eds.) (1994). *Visualization in modern cartography*. London: Pergamon Press.
Peuquet, D. J. (2002). *Representations of space and time*. New York: Guilford Press.
Spence, R. (2001). *Information visualization*. Harlow: Addison Wesley/ ACM Press Books.
Tufte, E. R. (1997). *Visual explanations*. Cheshire, CT: Graphics Press.
Ware, C. (2004). *Information visualization: Perception for design*. San Francisco: Morgan Kaufmann Publishers.

Relevant Websites

http://www.itc.nl
International Institute for Geo Information Science and Earth Observation, Geovisualization and Minard's map.
http://www.opengeospatial.org
Open Geospatial Consortium, Inc.
http://www.cs.cmu.edu
The SAGE Visualization Group, SAGE, Samples.
http://www.berann.com
The world of H.C. Berann, The Cartography, Panoramas.
http://www.tfl.gov.uk
Transport for London, Tube and DLR.

German-Language Geography

M. Hess, University of Manchester, Manchester, UK

Introduction

Since Immanuel Kant gave the first geography lecture on physical geography at the University of Königsberg (now Kaliningrad) in 1784, German language geography has come a long way. It developed as an academic discipline in Germany, Austria, and Switzerland during the nineteenth century, with the first chair for geography established at the University of Berlin in 1825 held by Carl Ritter, following his appointment as adjunct professor in 1820. Friedrich Ratzel's two volume *Anthropogeographie*, published in 1882 and 1891, is widely regarded as a decisive work and foundation of human geography. In addition, Ratzel is also considered to be the founder of (modern) political geography. During the first half of the twentieth century, German language geography continued to develop strongly and to influence the scientific community beyond the language boundaries. The work of geographers such as Otto Schlüter and Siegfried Passarge introducing the concept of landscape (*Landschaft*), but in particular the work of Alfred Hettner, then probably the leading methodologist of geography and best known for his concept of regional geography (*Länderkunde*), has had a profound impact on colleagues abroad. Their ideas have influenced scholars such as cultural geographer Carl O. Sauer of the 'Berkeley School' and Richard Hartshorne, one of the most eminent US geographers of his time. This period of intellectual development and international exchange of ideas came to an end with the rise of National Socialism and World War II, after which German language geography never regained the leading edge it once had. However, geography reestablished itself as an academic discipline of scientific and societal relevance and rebuilt its links to the non German speaking world, albeit slowly and gradually. The following sections are not meant to provide a full history of the subdiscipline, but will rather attempt to sketch the continuities and changes that shaped the current intellectual landscape of human geography.

Developments and International Dialogs

As indicated above, German language geography has for decades been what Geoffrey Martin and Preston James called an 'export product', whereas post 1945 it was mainly importing ideas, according to the same authors. There are two aspects of this observation that deserve attention, however: the development and/or adoption of ideas, concepts, and theories themselves and the ways in which these are communicated through language and received through publications in specific languages. In other words, post war geography in Germany, Austria, and Switzerland might either have failed to develop new ideas and rather adopted them from 'outside' – mostly from the US and UK geographers – or it has produced new concepts and theories but not communicated them to a non German speaking audience. There are indications – although not uncontested – that both readings contain some truth, and the reasons for this perceived 'loneliness of German geography' (and, by implication, Austrian, and to a lesser extent, Swiss geography) are manifold and complex.

It is safe to assume that political and institutional change during the twentieth century had a major impact on the way German language geography and its role for global academic dialogs developed. The influence that German language geography had early on and which is manifest, for example, in the well known exchanges between Alfred Hettner and Richard Hartshorne, or the deep and lasting impact of location theorists, such as Johann Heinrich von Thünen, Alfred Weber, Walter Christaller, and August Lösch, can partly be attributed to the appreciation of German science and philosophy by academics worldwide. With the rise of nationalism and the Nazis coming to power, geography as an academic discipline and how it was linked to the state apparatus came under increasing scrutiny, and ultimately led to an international isolation that would have lasting impacts – despite attempts immediately after the war by Carl Troll and other leading German language geographers to restore the lost reputation of the discipline. This legacy was reflected in the efforts and struggles to formally reorganize the institutions of higher education geography in the form of various associations, but maybe more profoundly the impacts were felt in the way human geography, in particular political geography, developed. In West Germany, political geography – let alone geopolitics – was basically absent as a research area until the 1980s, when the first major publications began to emerge, grounded mostly in Anglo American critical political geography. An important contribution in this context was Jürgen Oßenbrügge's PhD thesis, published in 1983 under the title *Politische Geographie als räumliche Konfliktforschung* (*Political Geography as the Study of Spatial Conflicts*). Meanwhile, human geography in East Germany – in line with the state doctrine – had long ago adopted a socialist–Marxist approach, and was organized in a single, unified Geographical Society which was answerable to the state office of the German Democratic Republic.

Alongside political and institutional changes, the role of language and the related publication system is important in understanding the development of human geography. While the work of early German language theorists was either read in the original language or saw translations into other languages – especially English – this was hardly the case over the last 50 years. Although the publication output increased dramatically, little of this output was published in English, thus minimizing the impact of human geography research in a global academic environment where the German language (along with others like French) had become peripheral. As a consequence, the scientific community especially in Germany and Austria was increasingly talking to itself, reinforcing and possibly prolonging its 'loneliness'. Under such circumstances, it is easy to see why it would be difficult for any theories and concepts developed in German language geography, as well as the results from substantive empirical research, to travel beyond the language community. This has been thematized as early as in the 1960s, for instance, in the work of Dietrich Bartels, and triggered sometimes heated debates about the state and future of German language geography in a globalizing world of science, to paraphrase the title of an article by Ehlers. The number of critical reflections about this issue has grown in the last two decades, and this in turn reflects changes in the practice of human geography and its academic environment. It has become increasingly common for German academics to aim for publications in English, the *de facto lingua franca* of science (and all this entails in terms of the much debated, Anglo American 'linguistic imperialism'). Most of the younger scholars now have spent time as academics in foreign countries, taking up new ideas from abroad, while at the same time helping to communicate research from the German speaking world to a wider audience. The biannual German Geographical Congress, usually bringing together up to 2000 and more academics and students, offers an increasing number of English language sessions, allowing more foreign colleagues to participate in a meaningful way. Some international publishers have realized the market potential of a large German speaking scientific community and now are present at these conferences, which offers additional opportunities for contacts between German, Austrian, or Swiss scholars and international publishing houses. At the same time, the number of academics from these countries attending international conferences like the annual meetings of the Association of American Geographers or the Royal Geographical Society (with the Institute of British Geographers) is growing. For example, at the 2007 Global Conference on Economic Geography in Beijing, German participants made up the fourth largest group after academics from the US, Great Britain, and Canada. Arguably, such developments are encouraging as they can help to address the 'trade imbalances' in exchanging ideas, theories, and concepts. Some observers however are skeptical about what this means for geography as a discipline.

A Late Discipline? Changing Paradigms and Perspectives

German language geography has not only been described as 'lonely', it has also been considered to be a 'late' science, lagging behind developments in other parts of the world and cognate disciplines – typically by about 10 years. Indeed, it is possible to identify a succinct time lag in the adoption and use of different approaches and paradigms in German language human geography. While the quantitative revolution and the related spatial approach took hold in the US and UK human geography, much of the work in Germany, Austria, and Switzerland was still heavily influenced by 'classic' concepts based on notions of landscape research (*Landschaftskunde*) and regional geography (*Länderkunde*). By the time the spatial approach began to find its way into German language geography (amidst much controversy about the future of human geography, especially during the 1969 German Geographical Congress at Kiel University), the late US geographer Allan Pred had published his work on the behavioral approach in human geography. Again, this was to become a widely used paradigm in Germany about 10 years later. Similar time lags could be observed regarding other approaches and paradigms like radical geography or feminist approaches.

As in other scientific communities around the globe, human geography in the German speaking world today is characterized by a plurality of coexisting heuristic frameworks and research perspectives. In a recent article, Austrian geographer Peter Weichhart identified 12 paradigms currently underpinning research, including post structuralist approaches, new regional geography, and new cultural geography. The latter has made major inroads into German language geography only very recently, with an overview published in a textbook edited by Hans Gebhardt, Paul Reuber, and Günter Wolkersdorfer. One of its aims is to provide a picture of scientific plurality and difference, rather than producing a single, dominant conceptual core. Again, while this is not necessarily specific to German language geography, the timing and the arguments made in various debates – often with a missionary zealousness – are a particular feature of its discourse. In other words, the current pluralism of approaches and paradigms is still quite often considered to be a major weakness which jeopardizes the standing of human geography.

In English language geography, by contrast, this plurality has been embraced since the early 1980s a major (and generally positive) element of disciplinary identity. One example of these debates about multiperspectivity and paradigmatic dominance can be found in social geography. Swiss born and educated geographer Benno Werlen challenged established hitherto widely accepted paradigms of German language geography with his major multiple volume publication *Sozialgeographie alltä glicher Regionalisierungen* (*Social Geography of Everyday Regionalisations, 1995 and 1997*), in which he elaborated on a concept of social geography based on action and structuration theory. His work has attracted substantial criticism, not least accusing him to create a geography without space (*Geographie ohne Raum*), but also found in creasing recognition within the scientific community. Indeed, his reconceptualization of social geography is one of the few post 1945 German language contributions to have made its way into Anglo American textbooks. Two other recent German language scholars to have contributed substantively to international theoretical advance are Harald Bathelt on the development of a relational geography and Gernot Grabher on the socio economics of space.

Like social geography, other fields of human geog raphy have seen similar struggles over competing para digms, for instance, in economic geography or the field of cultural geography. As outlined above, this has not, however, led to 'Kuhnian' shifts or revolutions in the sense that one dominating paradigm is gradually super seded and replaced by a new one; the coexistence of multiple approaches is rather something that remains a fundamental characteristic. In the context of German language geography this could be of particular relevance, as the current and prevalent multiperspectivity can be regarded not only as some postmodern 'irritation' indi cating an arbitrary understanding of science, as critics would argue, but has the potential to stimulate debates and speed up theoretical developments by way of chal lenging conceptual inertia and standstill. The same ar gument can be made as far as the relationship of human geography in German speaking countries with its neighboring disciplines is concerned. Contemporary human geography in Germany, Austria, and Switzerland has made major progress here, and is clearly engaging in interdisciplinary cooperation as well as recognizing in tellectual contributions in other fields, in spite of ongoing debates about the future of geography between discip linary unity and fragmentation or dissolution. Geo graphy's engagement in interdisciplinary research is important not least because many German research funding bodies such as the German Research Council put increasing emphasis on national and international col laboration as a precondition for granting awards. Another example where geography features prominently are the various frameworks of interdisciplinary research set up and funded by the state of Bavaria.

Organization and Research Areas

Geographical teaching and research in Austria, Germany, and Switzerland today is organized in different associ ations which all operate under one national umbrella organization. For Austria, this is the Austrian Geo graphical Society (*Österreichische Geographische Gesellschaft*) with about 1500 individual and institutional members, among them the five existing university geography de partments. In Switzerland, the main national organization is the Swiss Association for Geography (*Verband der Geographie Schweiz*), the members of which are the eight university geography departments, four specialized as sociations (teaching, applied geography, geomorphology, and young geographers), as well as the regional associ ations. In Germany, the umbrella organization is the German Society for Geography (*Deutsche Gesellschaft für Geographie*, DGfG) with four subsidiary organizations: The Association of Geographers in German Universities, German Association for Applied Geography, University Association for Geographical Education in Germany, and the Association of German Geography Teachers. As in Switzerland and Austria, there are a number of regional geographical associations that among other things act as a vehicle to communicate geographical research to a wider public audience. In terms of scientific exchange, the bi annual geographical congress – organized by varying university geography departments on behalf of the Ger man Society for Geography – is the main platform to bring together not only scholars and students from Austria, Germany, and Switzerland, but increasingly also participants from other countries. The 2007 congress at the University of Bayreuth was the first to bring together all subsidiary associations for a common event in an ef fort to improve the dialog between universities, school teachers, and practitioners in applied geography.

At university level, geography is a major discipline that is represented at six Austrian universities, eight in Switzerland, and more than 60 in Germany. Taken toge ther, there are currently about 30 000 students enrolled in bachelor, masters, and teacher training diploma courses and degrees. The latter will be ceasing in the near future, as all universities are changing to the bachelor/masters model established in the Anglo American system and introduced in Europe through the Bologna process, which aims to create a European Higher Education Area with comparable institutional models. Geography departments are the main sites for German language geography research, but there are some important non university research centers that play a vital part in developing geographical knowledge, most notably the Leibniz

Institute for Regional Geography (*Leibniz Institut für Länderkunde*) in Leipzig. Funded by the German federal government and the state of Saxony, this institute has a rich and eventful history and was refounded in 1992. Today, the institute's more than 30 academic members of staff conduct research in the fields of theoretical geography, regional geography with a focus on Central and Eastern Europe, and engage in numerous activities of knowledge transfer.

In terms of fundamental research, German language geography encompasses all areas of human geography, and thus is well positioned to add to the existing inter national knowledge base – both in terms of generic re search subjects and the spaces and places on which research is conducted. To list all the topical research areas in detail is beyond the scope of this article and there is literature available for a more comprehensive overview of German geography development and trends from the 1950s to the early 1990s. It is notable, however, that while the traditional strength of research on particular regions continues to be a feature of German language geography, as documented in analyses of eco nomic geography and transport geography by Rolf Sternberg and Andreas Kagermeier, formerly neglected but internationally important research topics are now increasingly the subject of novel contributions by aca demics from Austria, Switzerland, and Germany. This is maybe most obvious not only in the field of (critical) political geography, but is also evident in developments of urban geography, social geography, and other areas of human geography. What is more, the traditional bound aries between the various established subdisciplines have begun to blur as research is focusing on substantive issues beyond classic subjects. The comparative strength and early work of German language geography on issues of sustainability is a good indicator for this change, as are the changes we currently witness in the context of DGfG related and independent working groups. New groups are being established now that do not have one of the traditional sectoral or regional foci, but rather center around specific issues, for instance, housing markets, or critically engage with mainstream geography, for in stance, the Critical Geography association (*Verein 'Kri tische Geographie'*) in Austria or the working group on political geography (*AK Politische Geographie*) in Germany.

While German language geography has been de scribed as a 'late' and 'lonely discipline' as far as theory building and international visibility is concerned, it has been and still is strong as an applied science. Whether in urban geography, economic geography, tourism, or any other field of geography, many academics and, of course, the practitioners trained as geographers engage in con sultancy work for political, economic, and social actors at different scales. In a sense, the tradition of regional geography arguably is an advantage when it comes to informing development policies on the local, regional, and national level, which forms an integral part of geo graphical consultancy (*Politikberatung*). Applied geography also plays a major part in advising societal stakeholders other than elected politicians and political bodies, for instance, in working with companies, labor unions, or social and environmental nongovernmental organizations (NGOs) on a wide range of different issues. The current academic and public discourse of human–environment relations, on the global as well as the local level, should certainly be something through which German language geography can continue to develop as an applied science as well as fostering conceptual work.

Conclusions: Talking to Whom? Language and User Communities

For any research to be relevant and have an impact – scientifically as well as in the societies that support it – the results have to be communicated to different audi ences. The audiences human geography is addressing (or ought to address) are not only the national and inter national scientific communities, important as this is to generate and maintain academic reputation. University based geography also has a responsibility to talk to users outside the 'ivory tower', and this also means teachers and students in secondary and higher education, policy and decision makers within societies, and not least the wider public and society at large. Language therefore becomes crucial in reaching these different audiences, and therefore it is important that geography continues to publish its conceptual and empirical research in German language. First, there are about 20 000 geography tea chers in German schools alone, and communicating the state of the art geographical knowledge between uni versities, school teachers, and students is maybe more important than ever in a complex, globalizing world fa cing tremendous social, economic, and environmental challenges. Second, the discipline needs to reach the audiences living their lives in various localities, and has to inform politics governing the development of places and spaces at various scales if it wants to continue to be both an academic endeavor and an applied science dealing with real world issues and making sense of the world.

At the same time, the discipline must continue to im prove its publication output in English language as well as international conference activities (the 2012 International Geographical Union (IGU) conference will be held in Cologne/Germany), both of which are crucial for the exchange of ideas and research results. A good example of this are the Hettner Lectures at the University of Heidelberg, which started in 1997 and since then bring together distinguished international geographers as guest speakers, graduate students, and early career researchers. According to the organizers, the annual lecture series is devoted to research frontiers in the borderlands between geography, economics, the social sciences, and the hu manities, and aims to further the exchange of methodo logical and theoretical developments across disciplinary and national boundaries, and to transmit the relevance of geographical thought to a wider audience. German language geography today is well placed to actively con tribute to this exchange, and thus to the development of the discipline both 'at home' and internationally. Having a voice both at home and abroad should also be a feature of other linguistic geography communities, like French, Spanish, or Mandarin, in order to further move forward geographical research, even if the (currently?) dominating language of global communication and international ex change is English, as in this brief article.

See also: Anglo-American/Anglophone Hegemony; Citation Geography; Geographical Journals; Historical Geography; Human Geography; Interdisciplinarity; Landscape; National Schools of Geography.

Further Reading

Burk, T. (2006). *Raumtheoretische Positionen in angloamerikanischen und deutschsprachigen sozial und kulturwissenschaftlichen Publikationen seit 1997*. *Überarbeitete Fassung*. Berlin. http://raumsoz.ifs.tu darmstadt.de/forschung/fo05 literatur/lit raumtheorie.pdf. (accessed in Aug. 2008).

Ehlers, E. (ed.) (1992). *40 Years After: German Geography. Development, Trends and Prospects 1952 1992. A Report to the International Geographical Union*. Bonn and Tubingen.

Ehlers, E. (1998). German geography in a globalizing world of science. *A documentation and personal interpretation. GeoJournal* 45, 57 68.

Ehlers, E. (2005). *Deutsche Geographie Geographie in Deutschland: wohin des Weges? Geographische Rundschau* 57(9), 51 56.

Fassmann, H. (2004). Geography in Austria. *BELGEO* 5(1), 1732.

Hartshorne, R. (1958). The concept of geography as a science of space, from Kant and Humboldt to Hettner. *Annals of the Association of American Geographers* 48(2), 97 108.

Harvey, F. and Wardenga, U. (1998). The Hettner Hartshorne connection: Reconsidering the process of reception and transformation of a geographic concept. *Finisterra XXXIII* 65, 131 140.

Kagermeier, A. (2000). German research in transport geography: Life in the space between objective analysis and political advice. *GeoJournal* 50, 17 24.

Kulke, E., Lentz, S. and Wardenga, U. (2004). Geography in Germany. *BELGEO* 5(1), 81 93.

Miggelbrink, J. (2002). *Der gezähmte Blick. Zum Wandel des Diskurses über "Raum" und "Region" in humangeographischen Forschungsansätzen des ausgehenden 20*. *Jahrhunderts. Leipzig: Institut für Länderkunde* (*Beiträge zur Regionalen Geographie 55*).

Monheim, H., Schwarte, M. and Winkelkotter, C. (1999). *Die deutsche Geographie dreißig Jahre nach Kiel*. *STANDORT Zeitschrift für Angewandte Geographie* 3, 46 49.

Oßenbrugge, J. (1983). Recent developments in political geography in West Germany. *Political Geography Quarterly* 2, 71 80.

Reuber, P. (2000). Conflict studies and critical geopolitics theoretical concepts and recent research in political geography. *GeoJournal* 50, 37 43.

Schultz, H. D. (1980). *Die deutschsprachige Geographie von 1800 bis 1970. Ein Beitrag zur Geschichte ihrer Methodologie*. Berlin.

Sternberg, R. (2000). State of the art and perspectives of economic geography taking stock from a German point of view. *GeoJournal* 50, 25 36.

Weichhart, P. (2001). *Humangeographische Forschungsansätze*. In Sitte, W. & Wohlschlaegl, H. (eds.) *Beiträge zur Didaktik des* "*Geographie und Wirtschaftskunde*" *Unterrichts*, pp 182 198. Vienna (*Materialien zur Didaktik der Geographie und Wirtschaftskunde*, Band 16).

Weichhart, P. and Arnreiter, G. (1998). *Rivalisierende Paradigmen im Fach Geographie*. In Schurz, G. & Weingartner, P. (eds.) *Koexistenz rivalisierender Paradigmen. Eine post kuhnsche Bestandsaufnahme gegenwärtiger Wissenschaft*, pp 53 85. Wiesbaden: Opladen.

Gerrymandering

R. Morrill, University of Washington, Seattle, WA, USA

Glossary

Gerrymander Territorial manipulation of electoral districts to discriminate against parties, persons, or political interests.

Malapportionment Unequal population between districts.

Introduction

Gerrymandering refers to the deliberate manipulation of territory to gain disproportionate power in elections. The term 'gerrymandering' immortalizes the name of the Massachusetts governor, and the name of the cartoonist, Tisdale, who in 1821 likened the shape of one manipulated district to a salamander (**Figure 1**). The purpose of gerrymandering is to help 'your side' and to hurt the 'other side', whether this refers to political party, a racial or ethnic group, kinds of territory (urban, rural, suburban, rich, poor), or individuals (incumbents or challengers), and thereby to obtain a higher share of seats than of votes.

While the most publicized examples of gerrymandering recently have bizarrely shaped congressional districts (see examples below), gerrymandering is practiced at all levels of government in which there are electoral districts with single members – legislatures, county and city councils, and many others. And clever map drawers can gerrymander – that is successfully discriminate – without resorting to extreme irregularity.

How to Gerrymander

There are various means of gerrymandering electoral districts (see **Figure 2**):

1. Splitting or dilution of the concentrations of the other side, so as to leave them a minority in as many districts as possible; cracking is a form in which the district of an incumbent of the other party is split.
2. Packing or concentrating the other side in as few districts as possible, so that many of their votes are 'wasted', while also creating many districts with moderate margins for your side.
3. Placing incumbents of the other side in the same revised districts.
4. Creating multimember districts where your party is in the majority.

Dilution is effective where the other party has enough support in one area to elect one representative, but splitting can prevent any representation. Packing is the ideal tool where the other party has enough votes – in a region or metropolis – to elect two or more representatives. Carving a district out of the densest central concentration will pack minority party voters so that they receive 85–90% of the vote, thereby 'wasting' 35–40%.

Placing two incumbents of the other party in revised districts was a key component of the infamous DeLay Texas redistricting. Multimember districts in which one party has a majority was historically an effective way of preventing minority representation, for example, Blacks, from representation in local and state government, mainly in the south, but it also works for partisan discrimination, as in the Indiana legislature (Indianapolis).

Most gerrymanders are executed by the very legislative bodies they control, the state legislatures, city or county councils, an obvious but accepted conflict of interest, but even bipartisan commissions or state or federal courts are not immune from the temptation and practice. At the very least, courts routinely approve gerrymanders, no matter how extreme.

Examples of Gerrymandering

Many wonderful and egregious examples of gerrymandering are widely viewable on websites (please see the section entitled 'Relevant website'). Monmonier's book, *Bushmanders and Bullwinkles*, shows several. Also look at the US Congressional District Atlas. Two famous examples of rather irregular gerrymandering are reproduced here. **Figure 3** shows part of the recent and court approved Texas redistricting plan and **Figure 4** the famous original 1992 North Carolina congressional district plan, and its notorious first and twelfth districts – racial gerrymanders.

While gerrymandering often does take the form of extremely irregular districts, subtler gerrymandering can appear reasonably compact, while using any one of these four tools. Changes to the congressional districts of Mississippi illustrate the transition from traditional communities to a historic discriminatory gerrymander against black voters, by splitting the north to south trending community via east–west districts (**Figure 5**). Similarly, Republican districters have sometimes been able to reduce democratic representation to a single 'packed' core seat, while splitting the remaining democratic vote area into several surrounding

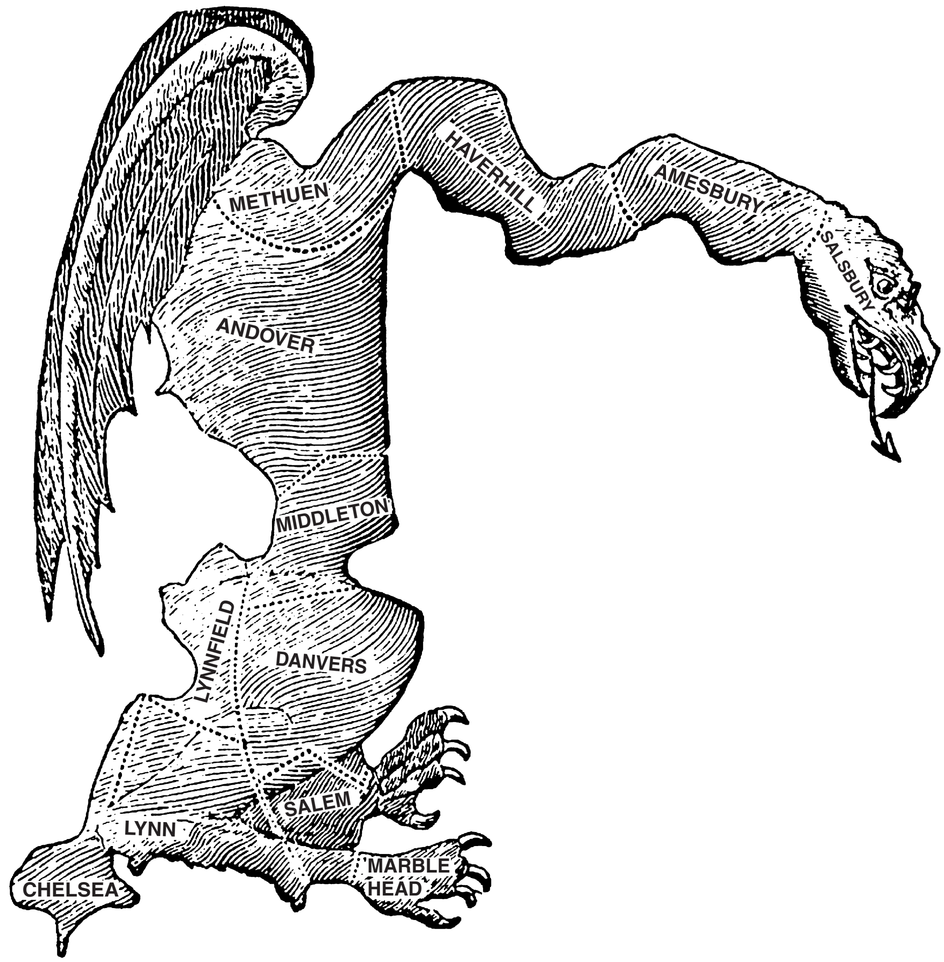

Figure 1 The original Gerrymander (1812).

districts, each with a majority of suburban Republicans (e.g., Minneapolis, Atlanta, Dallas, Houston).

Gerrymandering and the Courts

Gerrymandering was rarely challenged before the late 1960s, in part because it was simply accepted as 'politics as usual', and also because it was not as outrageous or as necessary. Malapportionment, or unequal population among districts was by far the more obviously unfair form of electoral discrimination, and even this was not struck down by the US Supreme Court until 1962 (Baker vs. Carr).

US Courts have both condemned and endorsed ger rymandering, but only with respect to discrimination against a racial or linguistic minority. How can this be? Gerrymandering was considered 'just politics' until critical decisions on the basis of the Voting Rights Act (1965, and subsequent amendments), which outlawed racial dilution, disallowed retrogression, and required 'preclearance' of plans by the Department of Justice in states with a history of racial discrimination. The most definitive decision was Thornton versus Gingles (1986), which established the standard that race can be one but not the only factor in drawing a district, but that it is reasonable to design 'majority minority' districts where practicable. The basic problem with race conscious dis tricting was the risk of packing in the large urban ghet toes and of extreme irregularity in assembling sufficient minority numbers in rural and small city parts of the south. The one criterion of racial representation was elevated at the expense of others, such as compactness or nonracial communities of interest.

The Department of Justice, many legislatures, and courts began the1990 round of redistricting with the expectation of striving for proportionality, even if heroic gerrymandering was required. This led to creation of well publicized irregularities in many redistricting plans, with disparate parts connected by strips of interstate highway, as in the North Carolina 12th District (**Fig ure 4**). The gerrymandered districts did elect a record

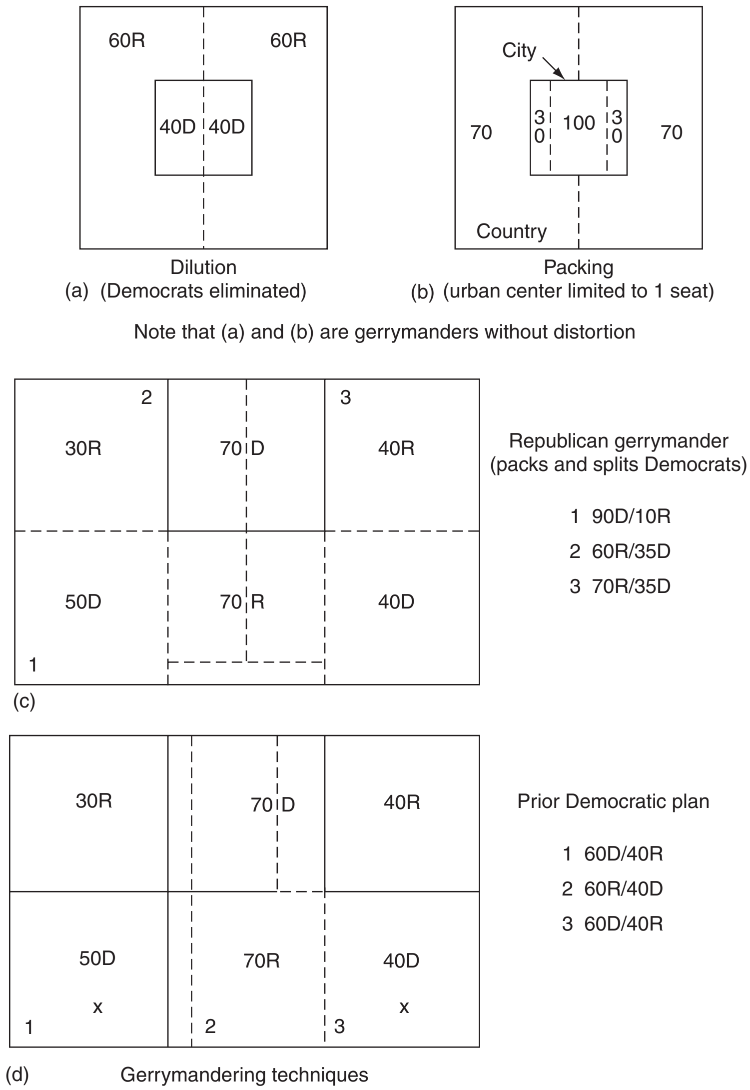

Figure 2 How to gerrymander. (a, b) Gerrymanders without distortion; (c) Republican gerrymander; and (d) prior democratic plan.

number of minority members in the 1992 elections, but at the expense of any geographic semblance of community. Voters in several states filed actions against plans. In the majority of such cases the Supreme Court did reject the plans as unconstitutional racial gerrymanders, and required less draconian plans (Turner vs. State of Arkansas). Yet several other infamously irregular plans were found to be acceptable.

Unresolved Issues

This leaves unresolved two aspects of gerrymandering and electoral fairness: those of party and of territory (e.g., urban, suburban, and rural small town), but there is no judicial consensus to invalidate manipulation against a political party or recognized place or community. Compactness and community of interest are viewed as 'desirable' but not as essential. Courts have ruled against recognition of any kind of community except race or language, and have accepted blatant and intentional political gerrymanders. Although the courts reference inexcusable irregularity in striking down a racial gerry mander, they totally fail to be concerned with convoluted districts that are political (Badham vs. Eu, Bandemer vs. Davis).

After both 1990 and 2000, Republicans, when in control, ingeniously pursued creating racial majority–minority districts, as this frequently eliminated liberal white democrats. The most extreme case after 2000 was in Texas, where the Republicans replaced a court drawn plan, leading to a massive shift of seats to the Repub licans. The plan was upheld at all levels, revealing that the US is not ready to tackle the unfairness of partisan gerrymandering, no matter how seriously it dis enfranchises voters, in a winner take all electoral system.

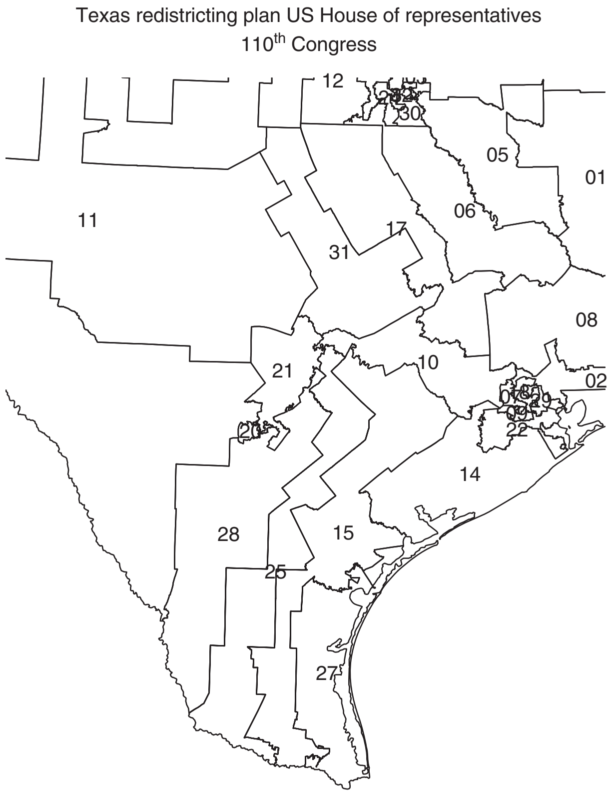

Figure 3 Racial gerrymandering in Mississippi.

What is fairness and why does it matter? Fair representation means that citizens' votes are meaningful, that over time and over a small set of districts, the shares of seats won is proportional to the share of votes received – that some peoples' votes are not weightier than others. The problems with gerrymandered districts which violate this sense of fair representation include distrust of government, apathy, not voting and withdrawal from the political process, lack of competition, and lack of change (shift of voter preferences does not translate to a shift in representation, a consequence of the almost automatic reelection of incumbents). Lack of competition results in many districts with no opponents even willing to run. The further risk is nonresponsiveness and corruption of government.

Gerrymandering, Computers, and GIS

Party workers, in presumably smoke filled rooms, have always used a combination of maps and knowledge of returns by voter precincts to accomplish their gerrymandering. The geographic information system (GIS) revolution in geography, in two respects, has also created a revolution in redistricting and therefore in gerrymandering technology. First was the development of computer cartography, in turn dependent on the creation of digital line (boundary) files and second was the extension of the census

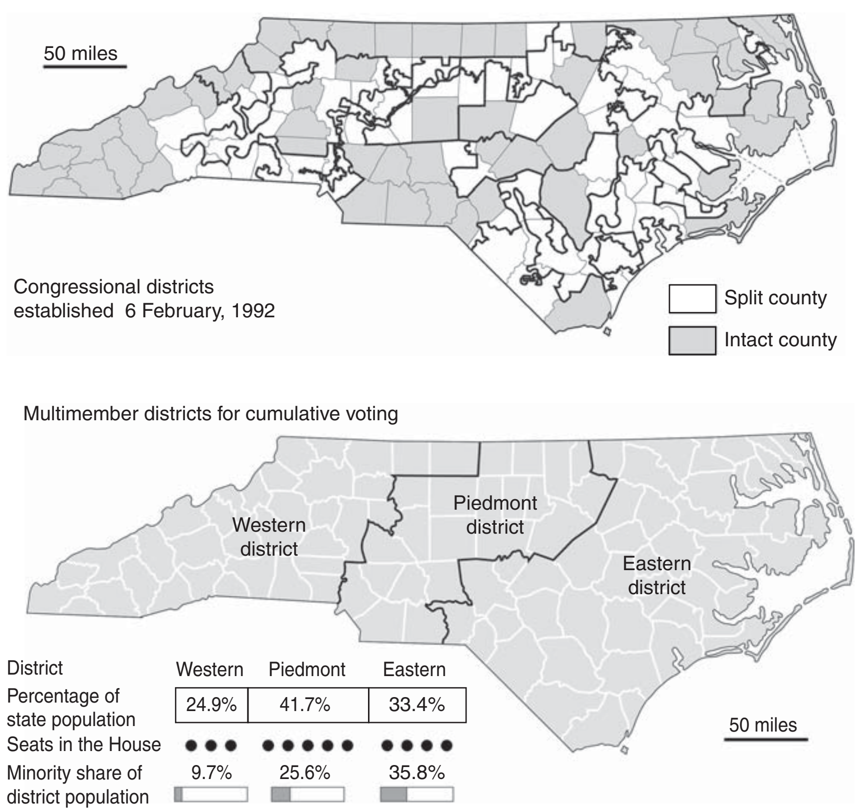

Figure 4 The famous North Carolina 12th District.

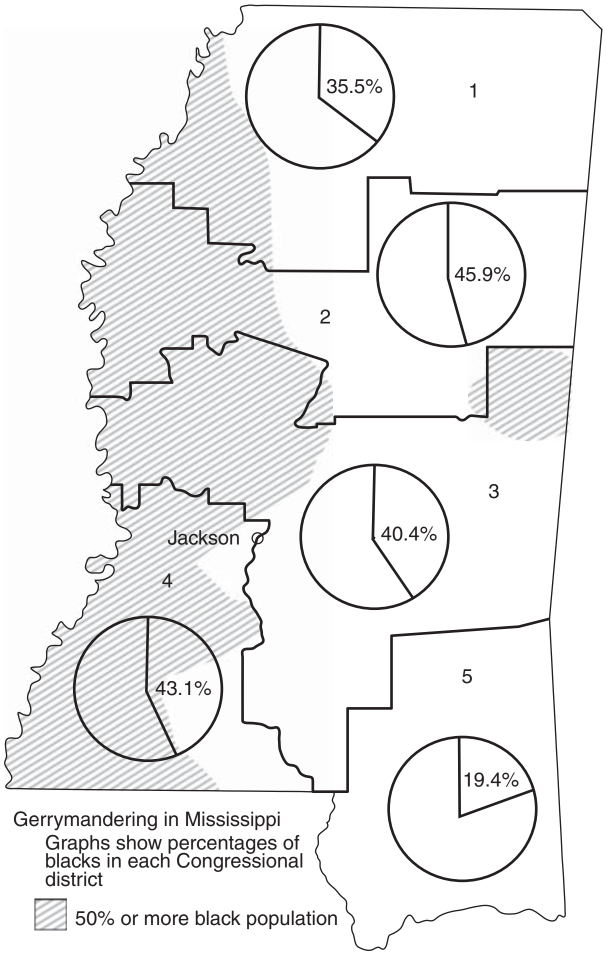

Figure 5 DeLay Texas redistricting.

down to the block level. The availability of data at the block level, together with efficient and interactive GIS packages ironically made it far easier to carry out ever more egregious gerrymanders, for example, by using stretches of highway to connect disparate parts of a district. Yet, excellent programming models exist which could draw fair and compact districts. The best models not only achieve political fairness between the parties, but also compactness and respect for communities of interest and the integrity of political boundaries; these are almost never used.

Redistricting Reform

Does it matter who does the districting? Plans drawn by truly independent commissions, as in Iowa, or the courts will be the least gerrymandered. Bipartisan commissions or divided legislatures tend to gerrymander for max imum protection of incumbents, although the plans may be reasonably balanced for the major parties. Legislatures controlled by one party, with a governor of the same party, usually results in the most extreme gerrymanders, one that creates open seats for them, while placing the other party's incumbents against each other, packing the other party's voters, and creating many marginal districts for themselves, although this can be risky if there is a very large shift in voter allegiance.

Reforms

Bipartisan redistricting commissions (usually still ap pointed by the parties, which still have to approve the plan) and the adoption of formal criteria for districting are the two reforms easiest to achieve. Several states have such commissions, which do at least give lip service to good government criteria such as compactness, com munity of interest, and avoidance of breaking up political jurisdictions. Only very few, such as Iowa and Hawaii, have technical commissions completely independent of parties. Increased political fairness and competitiveness would be truly possible only where the location of in cumbents is ignored – this is probably impossible.

Multimember districts with some form of pro portional representation is difficult to introduce in the United States, with its long history of single member districts, but they are widespread, even the norm in much of the world. Guinier proposed multimember (3) districts as a solution to the racial and fair representation dilemma raised by the North Carolina redistricting case (Shaw vs. Reno) – wise and clever, but utterly rejected. A variety of techniques is available, such as concentration of votes on one of the three candidates, or the 'single transferable vote', in which an individual's vote is allocated to the first preference, and then after candidates are elected or eliminated, then unneeded or unused votes transferred in accordance with voter's preferences.

Finally, the US Supreme Court could decide that political gerrymandering is an unconstitutional in fringement of fair representation and could reject a plan that disproportionately discriminates against one party. This is the most likely eventual solution, but one that is not imminent.

In gerrymandered election districts
The voters don't choose their politicians
The politicians choose their voters!

See also: Electoral Cartography; Electoral Districts; Electoral Geography.

Further Reading

Archer, J. C. and Shelley, F. (1986). *American Electoral Mosaics*. Washington, DC: Association of American Geographers.

Grofman, B. (ed.) (1990). *Political Gerrymandering and the Courts*. New York: Agathon Press.

Grofman, B. (ed.) (1998). *Race and Redistricting in the 1990s*. New York: Agathon Press.

Guinier, L. (1994). *Tyranny of the Majority*. New York: Free Press.
Johnston, R., Shelley, F. and Taylor, P. (1990). *Developments in Electoral Geography*. New York: Routledge.
Monmonier, M. (2001). *Bushmanders % Bullwinkles*. Chicago: University of Chicago Press.
Morrill, R. (1981). *Political Redistricting and Geographic Theory*. Washington, DC: Association of American Geographers.
Morrill, R. (1987). Redistricting, regions, and representation. *Political Geography* 6, 241 260.
Shelley, F., Archer, J.C., Davidson, F., and Brunn, S. (eds.) (1996). Representation and gerrymandering. *Political Geography of the United States*, ch. 6, pp 155 192. New York: Guilford Press.
US Bureau of the Census (1993). *Eongressional District Atlas, 103rd Congress*. Washington, DC: US Government Printing Office.

Relevant Websites

http://www.westmiller.com
Fair Vote 2 K Initiative (view from the right), Westmiller.
http://www.fraudfactor.com
Fraud Factor: Gerrymander and Redistricting Fraud, Fraud Factor.
http://www.gerrymanderingmovie.com
Gerrymandering: A New Document.
http://www.fairvote.org
Gerrymandering, FairVote.
http://geography.about.com
Gerrymandering, Geography at About.com.
http://www.motherjones.com
Gerrymandering: Why Your Vote Doesn't Count, Mother Jones.

Ghettos

T. Slater, University of Edinburgh, Edinburgh, UK

Glossary

Apartheid A term meaning 'separation' in Afrikaans, referring to a period in South Africa from 1948 to 93 when the ruling white supremacist National Party divided the country into four racial groups and created separate residential areas for each group along a distinct racial hierarchy.

Assimilation The process of an ethnic minority group acquiring the characteristics and sharing the resources of the 'host' or majority group through residential settlement patterns and/or behavior.

Banlieue The French word for 'suburb', it is now most commonly applied to the declining working-class zones of the French urban periphery housing relatively high concentrations of ethnic minority residents.

Blockbusting A profiteering tactic deployed by realtors or developers to convince white homeowners to sell their houses by telling them that black people are moving into their neighborhood, exploiting fears of lowered property values. The price of the house is then raised by the realtors when sold to black people.

Cités The French word to describe large estates of cheaply built high-rises situated in the *banlieues*.

Ethnic Enclaves An urban settlement where the dense clustering of one or many ethnic groups occurs, a consequence of group solidarity, cultural affinity, kinship ties, and occupational concentration, and usually accompanied by economic and social activity tied to the ethnicity of the group(s) in question.

Ethnoracial Group A group classified along the lines of skin color, caste, culture, and/or religion.

Ghetto A space deployed by discriminatory authorities to isolate, contain, and exploit a single ethnoracial group; a place to cast out (permanently) a group outcast from society.

Index of Isolation A measure of how often individuals from a particular group are likely to meet other individuals from their group – a group is 'isolated' if it rarely comes into contact with different groups.

Index of Segregation A measure of the percentage of a minority population which would have to change its location in order to replicate the distribution of the rest of the population of a city. It has a range of 0 (no segregation) to 100 (total segregation).

Jim Crow A system of legally enforced racial segregation (separate and inferior schools and public spaces, along with separate churches and clubs) in the southern states of the USA lasting from roughly 1865–1965. The term comes from a minstrel show song-and-dance routine entitled Jumping Jim Crow.

Redlining A term coined in the late 1960s by community activists in Chicago to describe the government-initiated (in 1935) practice of marking a red line on a map to delineate the area where banks would not invest (especially through rejecting mortgage loan applications). During the heyday of redlining these areas were almost always black inner-city neighborhoods.

Segregation Either voluntary or involuntary, the exclusion of a social group from the rest of society, usually used in reference to spatial residential patterns, but it can also relate to economic, social, and political spheres.

Introduction

'Ghetto' is without question one of the most misunderstood and inappropriately deployed terms both within and beyond geography and the social sciences. It is regularly used (or abused) as a metaphor to describe the spatial clustering of any social group (student ghetto, immigrant ghetto), or indeed simply a dilapidated and/or poverty stricken urban area, rather than correctly used to account for and analyze the selected few areas on the planet that exhibit the involuntary spatial confinement of a single ethnoracial group (in line with the history of the term). It is also something of a surprise, perhaps even an embarrassment to those who patrol the disciplinary walls of geography, that the majority of the most widely read and cited scholarship on ghettos – inherently geographical expressions of extreme racial oppression – has come not from geographers but from sociologists. Furthermore, arguably the strongest and most famous studies of ghettos tend to have in common their anchoring in ethnographic methodologies – very few geographers researching the formation and social conditions of ghettos have contributed to this tradition pioneered by the Chicago School. The paucity of geographical research that is accurately tuned in to the history of the term, together with the qualitative geographical disengagement and increasing confusion emanating from countless (and rather numbing) indices of segregation/isolation, means that analytical and policy errors toward marginal neighborhoods, whether ghettos or not, should come as little surprise.

Origins: The Jewish Ghetto

Perhaps due to the power of stereotypical representations of ghettos in the United States, and more recently the rise of the 'ghetto fabulous' discourse, it is seldom acknowledged that the term 'ghetto' in fact comes from Renaissance (sixteenth century) Venice. Following the defeat of the Venetians by the French at Agnadello (northern Italy) in 1509, Jews fled into Venice in increasing numbers, and quickly gained in social status via the professions of, *inter alia*, moneylending and medicine. Fledgling fear and hatred of Jews by the ruling Christians was in part a consequence of a campaign for the moral reform of the city following the weakening of its regional power at the time. Jews were viewed by the Venetian authorities to be the sole source of moral corruption, as well as carriers of disease (the spread of syphilis was incorrectly attributed to 'overly sensual' Jewish religious practices, rather than to the growth of prostitution; and, ironically, Jews making profits through usury was seen as tantamount to prostitution, and thus disease). There were widespread calls for the expulsion of all Jews to save the city from further moral degeneracy and contamination, but that would have been economic suicide for Venice, which desperately needed the high taxes, rents, and skills that the Jewish population provided. The solution of the Senate of Venice was thus complete spatial segregation, not expulsion: in 1516 all Jews were rounded up into an abandoned foundry called *ghetto nuovo* ('ghetto' originally meant 'foundry' in Italian) (see **Figure 1**). The ghetto was surrounded on all sides by water, not unusual in Venice, yet its walls were the highest in Venice, its outer windows and doors permanently sealed, its two bridges to the rest of the city guarded, and its surrounding canals patrolled by police boat. Jews were allowed out into the city during daylight hours only for economic reasons, but they had to wear special yellow badges – the same color as the scarves prostitutes were forced to wear. At dusk, Jews had to return to the ghetto or face serious consequences. (It remains something of a mystery why Shakespeare's *Merchant of Venice*, Shylock, moved with so much freedom around the city!)

While oppression and violence toward Jews in Europe did not begin in Renaissance Venice (ghettoization was in fact a huge improvement on massacre, which happened to Jews in medieval York under Henry II), the Venetian Ghetto became a blueprint for cities throughout Europe. The same fears about disease and moral upheaval, in tandem with the same needs to retain and exploit a Jewish population for its economic function, saw the involuntary segregation of Jews in cities across the continent, as geographically distant as Rome, Frankfurt, Prague, and Amsterdam. While they are often referred to today as 'Jewish Quarters', historically they were ghettos in both form and function. Yet while characterized by intense overcrowding, squalor, disease, and deprivation, they were also zones of collective consciousness, solidarity, and social organization. Because ghetto Jews were subject to territorial confinement by night and overt racial ostracization during the day, Jewish culture and community had no option but to develop in the densely packed ghetto. The peculiar combination of spatial entrapment and social cohesion saw the emergence of a 'city within a city', with its own institutions and administrative priorities. Everything from synagogues to markets, from schools to charities formed and developed behind ghetto walls, amplifying the rigid segregation of the Jews from the rest of society, yet strengthening bonds and kinship among ghetto dwellers, to the extent that strong divisions between Ashkenazic and Shephardic Jews collapsed in city after city under the experience of complete and involuntary ethnoracial containment. If we adhere to this historical fact – important if the ghetto is to be correctly distinguished from other marginal spaces that emerge from a different set of structural constraints (requiring a different set of analytical lenses) – then the ghetto can be defined as a space deployed by discriminatory authorities to isolate, contain, and exploit a single ethnoracial group; a place to cast out (permanently) a group outcast from society.

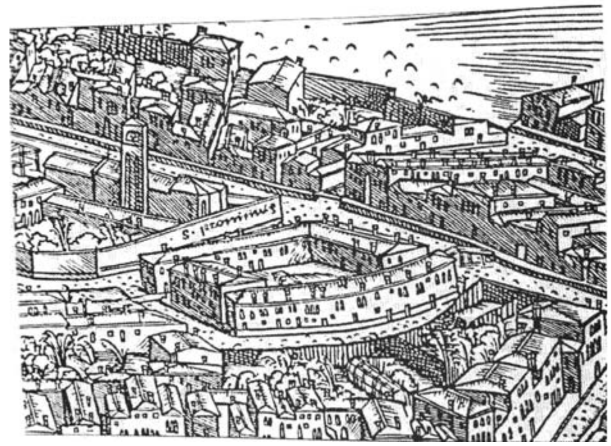

Figure 1 Jacopo dei Barbari's woodcut drawing of the Venetian Ghetto, 1500. Reprinted from Sennett, R. (1994). *Flesh and Stone: The Body and the City in Western Civilization.* New York: Norton & Co.

Transatlantic Connections: The African-American Ghetto

From the late nineteenth century onward, as urbanization accelerated in the cities of the north and eastern United States (partly fuelled by immigration from Europe), ethnic 'clustering' intensified as a consequence of group solidarity, cultural affinity, kinship ties, and occupational concentration. Three crucial factors distinguish the resultant 'ethnic enclaves' (Little Italies, Little Polands, Little Irelands, etc.) from ghettos – first, they were not products of ethnoracial control and containment by external overarching powers (one residential block in an ethnic enclave often contained people of many different nationalities); second, segregation within and between them was neither extreme nor rigid; and third, they were 'temporary' spaces for the many thousands of European immigrants residing in them, acting as springboards for an assimilative leap into the American 'melting pot' once plentiful opportunities for upward social mobility and class ascension were captured (many ethnic enclaves dissolved within two generations). A grave and terribly influential error of the Chicago School of Sociology, an extraordinary flowering of urban scholarship that drew their inspiration from the natural sciences in order to document the human urban equivalent of animals adapting to and competing in their natural habitats, was to equate ghetto with ethnic enclave and predict that all ethnic groups would assimilate into American society after initial clustering in their 'natural areas'. This was not the experience for one group, African Americans, who from the early twentieth century migrated in massive numbers from the southern states to northern cities for the twin reasons of fleeing Jim Crow racism and abject rural poverty, and for employment in the booming factory based industries of the north (where unskilled labor was desperately needed to grease the engines of Fordist capitalism).

The arrival of substantial numbers of blacks in northern cities, just like the arrival of Jews in Venice four centuries earlier, was seen by whites as a disturbing threat to racial purity, the moral order of the time, and the livelihoods of working class whites employed in the same industries. Blacks encountered harassment, hostility, and violence from whites, strictly on the basis of the color of their skins, together with a brutal tightening of previously porous systems of segregation in employment, education, and especially housing. For a black person, to cross the residential color line that emerged was to risk serious physical harm inflicted upon them by whites who had been socialized *en masse* into regarding blacks as subhuman (a legacy of slavery). Furthermore, neighborhood organizations lobbied successfully to prevent blacks from moving into white neighborhoods, while realtors set in motion a profiteering process of 'blockbusting', thus trapping African Americans in overcrowded residential areas where they quickly had to develop their own institutions to ensure the delivery of basic needs. This is another clear parallel with the Jewish ghetto, another city within a city, this time black within white, bounded by the walls set up by racism smouldering at the core of American society, and filtered through private and public institutions that ensnared a group that was nonetheless necessary for capitalist gain. Without question the most famous (and most studied) example is the 'Black Belt' of Chicago's South Side (see **Figure 2**), the Bronzeville of Drake and Cayton's magisterial 1945 study, *Black Metropolis*, and an area that housed an almost all black population, and 92% of Chicago's entire black population by 1930. Bronzeville is often the 'dual definition' marker used by many analysts to define a ghetto: almost the entire population in the area was comprised of one ethnic group, and almost every member of that ethnic group in the entire city lived in that one area.

It must be noted here that, first, African Americans are the only group ever to experience urban ghettoization in the United States (a legacy of entrenched systems of racial oppression, and of the white supremacy so ingrained in the power geometry of American capitalism); and second, that the African American ghetto has not been consigned to history as a feature of the Fordist metropolis, but is still a feature of the urban landscape in city after city across the post Fordist Rust Belt of the United States. Following World War II, the color line hardened even further due to the influence of federal

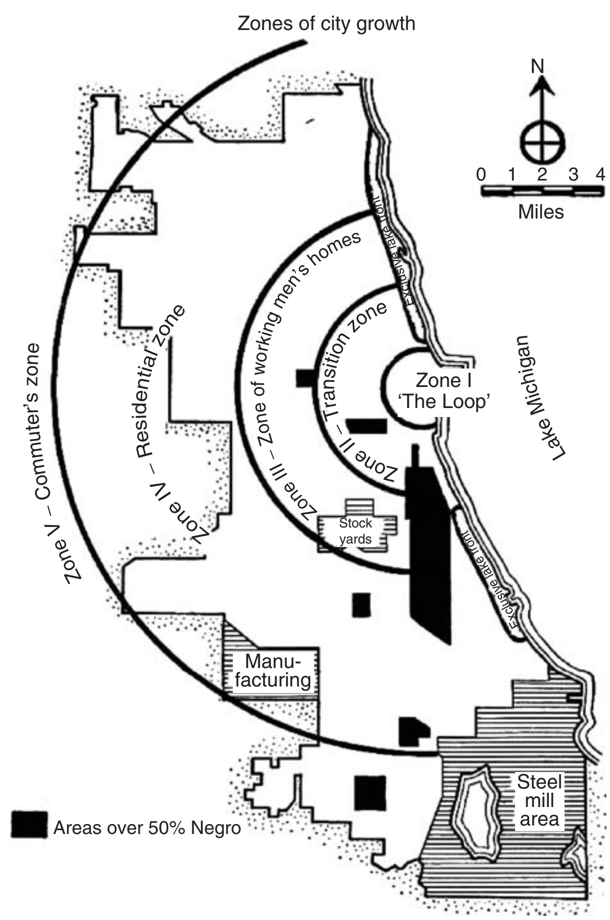

Figure 2 The location of Chicago's 'Black Belt', 1945. Reprinted from St. Drake, C. and Cayton, HR. (1993). *Black Metropolis: A Study of Negro Life in a Northern City* (first publ. 1945). Chicago, IL: University of Chicago Press.

government initiatives in spurring rapid suburban expansion. Mortgage loans for white families led to what became known as 'white flight' to the suburbs, and the construction of highways enabled transportation between downtown employment centers and the suburbs. The problem of ghetto overcrowding was addressed by a federally sponsored program of urban 'renewal', whereby freeway construction ploughed through existing ghettos and residents were relocated into vast public housing projects invariably built alongside the freeways, effectively turning public housing into reservations for the black urban poor, often termed 'second ghettos'. For example, between the early 1950s and early 1960s, the Chicago Housing Authority built 33 housing projects comprising 21 000 units, 98% of which were built in neighborhoods almost all black. Combined with the discriminatory practice of 'redlining', which made it impossible for black homeowners to sell up and leave ghettos, systematic capital disinvestment amplified entrenched mechanisms of racial exclusion to ensure that ghettos expanded and persisted, rather than dissolved and eroded. It is therefore no surprise that the Kerner Commission report of 1968, following the widespread black uprisings or 'ghetto riots' of the 1960s (described by Dr. Martin Luther King Jr. as the "language of the unheard"), concluded that one factor above all underlay all social and economic malaise – racial segregation.

One could be forgiven for assuming that the gains of the Civil Rights Movement in the US led to the gradual dismantling of the structural apparatus that had confined African Americans to ghettos, but ghettos persisted through the 1970s and onward due to massive economic restructuring and its effects on the poorest African Americans confined to ghettos. The major spatial impact of Civil Rights and the abolition of overt racism within housing markets (redlining was officially outlawed in 1977) was to facilitate the departure only of middle and lower middle class blacks from ghettos, leaving behind a subproletariat lacking the political clout, the institutions, the education, and the services required for upward mobility. The advent of neoliberalism triggered the rapid decline of manufacturing (and thus decent wage blue collar employment), the suburbanization of alternative employment, and the growth of the de regulated low wage service sector labor market (which offered no social protection, unionization, or opportunities for promotion). The outcome was the removal of the ability of young black men able to support a family, so the number of female headed families grew exponentially at a time when welfare payments were cut back, leading to deepening 'feminized' poverty on a staggering scale in economically traumatized cities such as Chicago, Cleveland, Detroit, and Buffalo.

The contemporary post Fordist African American ghetto is thus a different animal from its Fordist predecessor in terms of social organization and structural constitution. The neoliberal state has systematically dismantled the welfare state that allowed people to survive outside the formal economy, and has replaced it with a punitive 'workfare' state which coerces people into the de regulated low wage labor market. Young black men who either reject or are rejected by this labor market often turn to the criminal economy in order to survive, and thus come face to face with the increasingly punitive arm of the neoliberal state (police, courts, jails, prisons), which easily justifies its actions (mass incarceration) by blaming the 'underclass' poor for not exercising 'personal responsibility'. This has led some scholars to argue very convincingly that ghettos and prisons have become a single, interconnected system for the punitive spatial entrapment and management of the black urban poor.

Other Examples

Ghettos in the United States, then, while unique urban phenomena in the Western Hemisphere, are not unprecedented – they are closely tied to the Jewish ghettos of Europe in exhibiting the involuntary containment, exploitation, and control of a single demonized group. While ghettos in the accurate geographical, historical, and sociological sense of the term are rare urban forms, there are other zones of oppression and exclusion scattered throughout the world that can be termed ghettos. The areas that contain Japan's taboo Burakumin population, descendants of the country's rural outcasts, are a case in point. The premodern Tokugawa (or Edo) period (1603–1868) saw the organization of Japanese society into the four class categories of samurai, peasant, artisan, and merchant. Below the merchant class were the outcast groups of *hinin*, usually town guards, street cleaners, and executioners; and *eta*, usually butchers, tanners, leather makers, and undertakers. *Hinin* literally translates to 'nonhuman', while *eta* literally translates to 'filthy' – reflections of the attitudes held toward them (*hinin/eta* villages were rarely printed on official maps). Unlike Jews in European cities and blacks in American cities, the descendant Burakumin were never an immigrant group, yet they were widely perceived as a different race and officially labeled outcasts by the Tokugawa government. Burakumin were only permitted certain types of clothing, not allowed to leave their *burakus* (rural hamlets) to enter towns at night, and banned from entering Buddhist and Shinto sites of meditation and worship. Rigid and total segregation was enforced via strict rules controlling the size of their houses and settlements, and even laws prohibiting their windows from facing tracks and roads.

As the Tokugowa era gave way to the Meiji period, in 1871 the Burakumin were emancipated, but akin to the African American experience this did not lead to simple acceptance and assimilation into Japanese society. The Burakumin became an increasingly urbanized population and experienced new patterns of segregation through the mechanisms of housing (Burakumin populations were invariably forced into the poorest neighborhoods of Japanese cities with the worst public services – Kyoto in particular contains many segregated, poor Burakumin), marriage (there was strict monitoring of family registration records to prevent marrying into other groups), education (separate schools), and employment (they were informally barred from industrial employment, and forced to work traditional, low paying, 'unclean' jobs). Yet, in line with a chief characteristic of ghettos, to meet their own needs the Burakumin developed their own institutions replicating those of the society which had excluded them, and organized efficiently to lobby for urgent action to counter discrimination. While there have been considerable gains in Burakumin rights since the 1960s, and evidence of desegregation in many cities and villages, Burakumin ghettos still exist in tandem with the discrimination that created them – while the latter may be more subtle than in the past, continuing prejudice is reflected in the resistance non Burakumin children often encounter from their parents when they wish to marry Burakumin.

Another example of ghettoization which departs from the textbook model but still exhibits the same organizing principles is the townships of apartheid South Africa. The white supremacist National Party came to power there in 1948, viewing itself as 'God's chosen race' to 'civilize' the southern tip of black Africa, despite the fact that whites were massively outnumbered by black Africans in that country. Hendrik F. Verwoerd, Minister of Native Affairs in the early 1950s and widely regarded as the architect of apartheid (a word literally meaning 'separation' in Afrikaans), viewed complete and total racial separation as the 'answer' to conflicting material interests and cultures, to perpetuating white rule (via economic control of vast mineral resources), and to reversing the decline of the white population. Shortly after it came to power the government enforced the classification of South Africa's population into four casually defined but tightly policed racial groups, with a distinct hierarchy – whites, Indians, 'coloureds' (mixed race), and Africans (later 'blacks'). In keeping with the accurate meaning of the ghetto term, the National Party realized that South Africa's economy needed a vast supply of cheap non white labor, but also felt that under no circumstances should that labor live in city centers or anywhere near desirable (white) areas. Therefore, they legislated in *de jure* racial segregation to intensify, extend, and cement in place the *de facto* segregation that already existed, via the Group Areas Act (1950), called "one of the cornerstones for preserving a white South Africa" by Minister of the Interior, T. E. Donges that same year.

Between 1950 and 1993 it is estimated that over 1.5 million people were forcibly removed from their homes under the Group Areas Act, and relocated to other parts of South African cities, away from whites. 'Group Areas' (both business and residential) were 'proclaimed' by the government for occupation/ownership by a specific racial group, and persons of any other racial group became 'disqualified' persons who had to move out before a government deadline (usually 3 months). This vast, brutal system of social engineering through spatial manipulation led to the creation of vast township ghettos involuntarily confining, exploiting, and controlling subordinated racialized groups, with devastating psychological, economic, and social consequences for displaced persons. The legacy of the Group Areas Act is a highly segregated urban landscape with millions of non white township dwellers living in grinding poverty, now amplified by the neoliberalization of South Africa, extreme overcrowding via mass migration from all over sub Saharan Africa, and the spreading AIDS pandemic. Townships, then, can accurately be termed ghettos, but they are unusual in that they are relics of racist 'minority' rule that sought to enclose a 'majority' population, and are located inside cities systematically and 'entirely partitioned' along racial fault lines.

Case Study: The Myth of Contemporary European Ghettos

It should be clear from the discussion above that not all segregated areas are ghettos. Ghettos should be viewed as a peculiar, extreme type of urban form rooted in discrimination toward a particular group of people loathed by their oppressors, to the point where the oppressed are forced to develop their own institutions behind the parameters of their involuntary confinement to necessitate their survival. To designate as ghettos any other segregated areas, such as gated communities of affluent white suburbanites, residential concentrations of students in neighborhoods adjacent to universities, or even the diverse neighborhoods of Western cities themed along ethnic lines ('Little Italy' in Toronto; 'Japantown' in San Francisco; 'Banglatown' in London), is to erase over 500 years of the term's history and to begin any analytical focus from an entirely inappropriate conceptual point of departure. Yet time and time again, 'ghetto', a morally loaded term carrying with it a set of pernicious representations, is used incorrectly by scholars and especially by policymakers. Two recent examples from Britain and France, reinforcing the myth that ghettos on a par with African American ghettos exist in European cities, serve to illustrate problems of misuse.

The fear of nascent ghettos is a recurrent one in Britain, but went into overdrive in September 2005, following a declamatory yet high profile speech entitled

'After 7/7: Sleepwalking to Segregation' by Trevor Phillips, then Chair of the Commission for Racial Equality (now the Commission for Equality and Human Rights). Phillips argued that Britain has some residential districts (in Bradford and Leicester) that are nearly as segregated as Chicago's South Side, and "are on their way to becoming fully fledged ghettos – black holes into which no one goes without fear and trepidation, and from which no one ever escapes undamaged." Not surprisingly the national media picked up on this part of the speech and massaged it into a colossal moral panic about the failures of multiculturalism, integration, community cohesion, and immigration policy (**Figure 3**). However, the speech was written without careful attention to the 2001 census, which revealed that there is not a single administrative ward, let alone neighborhood, in Britain in which the population is 90–100% ethnic minority, whereas tracts with these percentages are commonplace in the US. Furthermore, for all ethnic minority groups identified by the census, indices of segregation in fact fell between 1991 and 2001. If he did do much preliminary reading, Phillips was most likely misled by research on segregation using less reliable indices of isolation, which did imply that Britain's ethnic minorities were becoming more isolated.

Regardless of what measures were consulted, however, Phillips was painfully unaware that geographers have been at the forefront of published research confirming that Britain does not have ghettos. In Leicester, for example, the only city in which a segregation index rose between 1991 and 2001 (from 30 to 39 for people of Black Caribbean descent), local residential concentrations of ethnic minorities are most definitely not a consequence of the involuntary spatial containment and control of an outcast single group that has since developed its own parallel institutions for survival. All ethnic minority groups in that city and elsewhere remain heavily dependent on the welfare state to supply basic needs, and even if discrimination is present both in public and private institutions (such as in housing and employment), Britain's often ethnically diverse neighborhoods offer no material for sensible comparison with the hostility and systematic racial exclusion that historically and geographically characterize rigid ghettoization. There are two dangers of labeling as ghettos those British neighborhoods that house a greater proportion of ethnic minority groups than others. First, it creates reputational ghettos, bogus territories that are nonetheless avoided through fear and damaged through stigma, perpetuating the harmful discourse of 'self segregation' and stirring up another myth, that of ethnic withdrawal from society. Second, it obscures the increasing segregation that is happening in Britain – between rich and poor (not between white and non white, or between Christian and Muslim, for example) – and deflects social policy attention away from class stratification/inequality and inappropriately toward fictitiously broad racial divisions.

Figure 3 Trevor Phillips, along with some of the headlines his speech created. Reprinted from The Observer, 25 September 2005.

Britain is not the only country to experience ill informed policy led 'ghetto blasting'. French urban policy toward its poorest neighborhoods has for some years now been guided by the myth that the *cités* (housing projects) in the working class *banlieues* (suburbs) of its largest cities are very similar, if not identical, to African American ghettos in the United States. A few weeks after Trevor Phillips' speech was funnelled through hysterical national media in the UK, the legend of structural and geographical equivalence between the black American ghetto and the French urban periphery was reinforced by the policy and media reaction to the major civil disturbances that occurred in several *banlieues* in city after city across France for a period of nearly 1 month. The uprisings were a reaction to many months of severe police oppression and harassment of marginalized first and second generation immigrant communities in those areas, mostly from North and West African nations. On 27 October 2005, these 'zero tolerance' surveillance tactics of then Interior Minister (and now President), Nicolas Sarkozy had culminated in the death by accidental electrocution of two teenagers trying to escape 'routine' police questioning in the *banlieue* of Clichy sous Bois. In the weeks that followed, as cars and buildings burned on the streets of urban France, commentators and pundits quickly resuscitated long running debates over immigration policy and the failures of assimilation ideology, the tensions between Islam and the secular policies of the French state, and in particular the dilapidation and deepening misery of the *cités*, which in demographic composition (large ethnic minority population), socioeconomic condition (crushing poverty and widespread unemployment), and behavioral features (pandemic despair cum violence) were viewed to be the

European example of the American ghetto, as they have been viewed for decades.

While the deterioration of living conditions and declining life chances in the French urban periphery since the advent of the intertwined processes of post-industrialism and neoliberalism cannot be disputed, to designate as race ghettos the *cités* and/or *banlieues* robs these areas of an appropriate conceptual and analytical framework that otherwise may go some way toward explaining their plight. Why, then, are they not ghettos, as the respected source of BBC News repeatedly claims that they are (see **Figure 4**)? Two crucial reasons (among many) begin to explain why. First, a brief glimpse at censuses of population, or even a few hours spent there, will reveal that the population residing within them, just like in many other 'notorious neighborhoods' in European cities, is characterized by remarkable ethnic heterogeneity, as opposed to the strictly enforced (from without) ethnic homogeneity of ghettos. What unites the marginalized urban population of France is (working) class position, not racial subjugation via spatial entrapment. Second, residents in housing projects of suburban France have regular contact with and dependency on (through employment and consumption practices) residents of adjacent neighborhoods or those in the central city. This was not the case in recent memory in Chicago's West Side ghetto, for instance, where so segregated and contained were the residents of one housing project, the Henry Horner Homes, that a young African American boy repeatedly asked his mother if all black people lived in housing projects, and if all black people were poor. With the possible exception of religious institutions, in no *cité* or *banlieue* of (sub)urban France will a rigorous analyst find evidence of a set of parallel institutions forming to serve the basic needs of the excluded population – by contrast with what is common to all ghettos, the same analyst will quickly discover that these are areas where the tentacles of the welfare state, while shorter than in previous decades, still stretch into and attempt to assist lives of those for whom employment is precarious or absent and opportunities for upward mobility limited by language and skills.

Conclusion

Contrary to the powerful imagery associated with the 'ghetto fabulous' discourse, an attempt by hip hop industry moguls in the United States to celebrate the triumph of those who have escaped ghetto poverty through

Figure 4 Urban legends are reinforced through inappropriate misuse of 'ghetto'. Reprinted from the BBC News Website.

music and fashion (and now live lives of lavish con sumption amidst 'bling bling' Cristal champagne, ex pensive jewellery, and stretch limousines), ghettos are far from fabulous, and instead should be viewed as tragic urban forms that result from fear, oppression, and hatred. If there is anything fabulous about ghetto life, it is the extraordinary ability of their residents to form insti tutions of their own in the absence of external concern, and to form strong communities that, contra the Chicago School, are not 'socially disorganized' around the de monic imagery of the criminal 'underclass', but are or ganized in collective response to brutal confining forces that seek to isolate yet exploit, and contain yet retain. Reclaiming the ghetto term from inappropriate usage and restoring it through direct linkage to its historical/geo gaphical roots is vital if policies toward marginal neigh borhoods are to be accurately formed and sensitively implemented.

See also: Chicago School; Gay Geographies; Multicultural City; Planning, Urban; Race; Redlining; Segregation; Segregation Indices; Segregation, Urban; Slums; Suburbanization; Underclass; Urban Morphology.

Further Reading

Anderson, E. (1978). *A Place on the Corner*. Chicago, IL: University of Chicago Press.
Commission for Racial Equality (2005). Trevor Phillips' speech, 'After 7/ 7: Sleepwalking to segregation'. http://83.137.212.42/sitearchive/ cre/Default.aspx.LocID 0hgnew07s.RefLocID 0hg00900c002.Lang EN.htm (accessed Jul. 2008).
Drake, S. and Cayton, H. (1993). *Black Metropolis: A Study of Negro Life in a Northern City* (first publ. 1945). Chicago, IL: University of Chicago Press.
Hane, M. (1982). *Peasants, Rebels and Outcastes: The Underside of Modern Japan*. New York: Pantheon Books.
Hannerz, U. (1969). *Soulside: Inquiries into Ghetto Culture and Community*. New York: Columbia University Press.
Hart, D. (1988). Political manipulation of urban space: The razing of District Six, Cape Town. *Urban Geography* 9, 603 628.
Hirsch, A. (1998). *Making the Second Ghetto: Race and Housing in Chicago 1940 1960* (2nd edn.). Chicago, IL: University of Chicago Press.
Johnston, R., Forrest, J. and Poulsen, M. (2002). Are there ethnic enclaves/ghettos in English cities? *Urban Studies* 39, 591 618.
Kotlowitz, A. (1991). *There are No Children Here: The Story of Two Boys Growing up in the Other America*. New York: Doubleday.
Ley, D. (1974). *Association of American Geographers Monograph Series, 7: The Black Inner City as Frontier Outpost: Images and Behaviour in a Philadelphia Neighbourhood*. Washington, DC: Association of American Geographers.
Massey, D. and Denton, N. (1993). *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, MA: Harvard University Press.
Peach, C. (1996). Does Britain have ghettos? *Transactions of the Institute of British Geographers NS* 21, 216 235.
Philpott, T. (1978). *The Slum and the Ghetto: Neighbourhood Deterioration and Middle Class Reform*. Oxford: Oxford University Press.
Sennett, R. (1994). *Flesh and Stone: The Body and the City in Western Civilization*. New York: W.W. Norton & Co.
Wacquant, L. (1997). Three pernicious premises in the study of the American ghetto. *International Journal of Urban and Regional Research* 21, 341 353.
Wacquant, L. (2001). Deadly symbiosis: When ghetto and prison meet and mesh. *Punishment and Society* 3, 95 134.
Wacquant, L. (2007). *Urban Outcasts: A Comparative Sociology of Advanced Marginality*. Cambridge: Polity Press.
Western, J. (1996). *Outcast Cape Town* (2nd edn.). Berkeley, CA: University of California Press.
Wilson, W. J. (1987). *The Truly Disadvantaged: The Inner City, the Underclass and Public Policy*. Chicago, IL: University of Chicago Press.

Relevant Websites

http://www.blhrri.org
Buraku Liberation and Human Rights Research Institute.
http://www.districtsix.co.za
District Six Museum, Cape Town: Documents and Explains the Group Areas Act.
http://www.fcit.usf.edu
Excellent Photographs of the Venetian Ghetto, A Teachers Guide to Holocaust, Florida Center for Instructional Technology.
http://www.urbandecay.ca
Remarkable Photographs of Chicago's Ghetto Neighborhoods and Public Housing Projects (Before They Were Knocked Down by the Federal HOPE VI Program).
http://www.detroityes.com
The Fabulous Ruins of Detroit, DetroitYES.

GIS and Cartography

M. F. Goodchild, University of California, Santa Barbara, CA, USA

Glossary

Canada Geographic Information System A computer-based system constructed by the Government of Canada in the 1960s, believed by many to be the first GIS.

Choropleth Map A map depicting statistical data for irregularly shaped reporting zones such as countries, using variations in shading or color.

Class Interval A range of values associated with a single shade or color on a choropleth map.

Deconstruction The scholarly practice of inferring hidden agendas from the content of published material.

Geovisualization Presentation of geographic information in visual form.

Google Earth An Internet service that allows the user to manipulate a virtual image of the Earth.

Map Scanner Device for converting the contents of a map into a digital image.

Mercator Projection One of the most popular ways of distorting the Earth so that it can be shown on a flat paper medium; the Earth is stretched at high latitudes.

Object-Oriented Database Design A technique for representing phenomena in a digital database.

Introduction

Map making is an ancient field dating from the first efforts by humans to create pictorial representations of the world around them, by drawing on the walls or mud floors of caves. Cartography, the modern discipline of map design, compilation, and publication, is most often associated with the paper map, a flattened representation of the curved surface of the Earth, but it may also refer to the creation of globes, and increasingly it refers to the use of digital computers to manage the acquisition, manipulation, and eventual display of geographic information on the screens of computers. The advent of digital technology in the 1960s began a transition in the field of cartography that continues today.

Geographic information systems (GISs) are computer applications concerned with the manipulation of geographic information, and today these software packages are capable of the representation, analysis, and visualization of virtually any form of information about the distribution of features and phenomena on the surface of the Earth. The initial developments of GIS occurred in the 1950s and 1960s, as computers became powerful enough to manage the large volumes of data being collected from satellites or digitized from paper maps, and today virtually all geographic information is in digital form, stored in a computer, at some point in its life.

These definitions clearly overlap substantially, and while there might have been a period in the 1960s and 1970s when it was possible to distinguish between a cartography focused on the paper medium and a GIS focused on digital systems, today the distinction is hopelessly blurred. Many cartographers now prefer the term geovisualization, and many GIS professionals consider themselves expert in the design of maps. The following sections explore the many dimensions of the relationship between cartography and GIS, and end with a brief speculation on the relationship's future.

The Power of the Digital Medium

In the 1960s, it was not at all obvious that there would be advantages to representing geographic information in digital form. Computers were still regarded as fast calculators, capable of finding numerical solutions to differential equations. They had little memory capacity, and the devices needed to digitize or draw paper maps did not yet exist. When the Canada Geographic Information System (CGIS) was developed, it was to serve one specific purpose: to calculate the areas of irregularly shaped features, and to present the results in tabular, numeric form. To do this, a number of historic breakthroughs had to be made. A map scanner had to be built, software had to be designed to perform tasks that computers had never performed before, and methods had to be found to represent geographic features as linear strings of binary digits on magnetic tape.

By the 1970s these breakthroughs had paved the way for the management and manipulation of many forms of geographic data, and for the cartographic production process to begin a slow migration to digital media. Just as word processors allow easy editing of digital text, cartographers also welcomed the ability of digital media to support rapid editing, and to perform the complex numeric operations involved in map projections. Within a decade the entire cartographic process had migrated to digital media, and today it is rare to find any part of the process being conducted manually or in analog form. Geographic data are increasingly digital throughout their life history.

The digital transition, implicit in the development of GIS, had profound effects also on the nature of cartography. No longer were maps condemned to be static once printed, since digital maps could be edited, updated, and corrected at any time. As the graphics capabilities of personal computers grew, it became possible to represent dynamic phenomena, and cartographers began to investigate the potential of animated maps. The acoustic channel, now supported on personal computers with sophisticated sound systems, was explored as a means of augmenting maps, and systems were built that allowed users to interact with maps in both visual and auditory senses. Thus one could explore a digital representation of a wilderness area, zooming to greater detail or accessing textual information associated with specific features, and even hearing the sounds associated with places and their animal inhabitants. Many researchers found the traditional term cartography too limiting as a description of this new, much richer world, and began to describe their field as geovisualization.

Today the effects of this transition can perhaps be seen most clearly in the various virtual globe services that are now available, typified by Google Earth. These services provide access to vast stores of imagery and digital map data through a simple interface that a child of ten can learn in a few minutes. The user is able to visualize geography as if looking at and manipulating a globe, with none of the distortions that are introduced when the Earth is flattened for a paper map. The virtual globe can be manipulated by rotating and zooming, from resolutions of 10 km to less than a meter. Information can be merged from a vast range of sources using the techniques now popularly known as mashups, and these extend to three dimensional representations of buildings and structures. The user's viewpoint can be shifted from vertically above to oblique, and it is possible to simulate a flight over the surface.

Google Earth is in many ways more akin to the traditions of map making than to GIS. Little of the analytic power of the latter is present; instead the emphasis is on visual representation of the Earth's surface, and an important measure of the service is the degree to which its results resemble the real thing. By emphasizing imagery, and allowing the user to overlay other information on an imagery base, one ensures that the result is familiar, recognizable, and readily understood. By contrast, displays generated using GIS tend to be highly abstracted, and while the information they present will likely be precise and in most cases accurate, the GIS lens clearly provides a more analytic and less intuitive view of the world. Indeed, it is common for researchers to combine the two technologies, using GIS to analyze patterns, search for anomalies, test hypotheses, and compute numerical results, and using Google Earth to provide additional contextual information that may not be present in the highly abstracted GIS database.

GIS as Threat

Google Earth abandons many of the cherished icons of cartography such as projections, rendering them largely irrelevant, and in this sense can be seen as threatening to a discipline's cherished expertise. Similarly, GIS enables anyone equipped with data and a few simple tools to produce a map that previously would have required the expertise of a trained cartographer. While great efforts have been made by developers of the more elaborate GIS packages to include support for sophisticated cartographic techniques, such as methods of cartographic generalization and alternative techniques for assigning class intervals to choropleth maps, nevertheless it is easy to find examples of the misuse of GIS. For example, several recent news stories of the missile threat of North Korea have included maps purporting to show the areas reachable by missiles of a given range launched from Pyongyang – by drawing concentric circles on a Mercator projection. Any cartographer, and hopefully most trained GIS users, would know that the scale of the Mercator projection changes rapidly at high latitudes, and that on this projection the locus of equal distance from Pyongyang is only a circle when the distance is vanishingly small. Unfortunately, the result is a severe underestimate of the areas that can be reached (**Figure 1**).

It is sometimes argued, therefore, that GIS, and the popularization of digital geographic information technologies in general, represents a threat to the field of cartography – that 'GIS is killing cartography'. Map making, it is argued, is a sophisticated pursuit that is best left in the hands of experts. Maps are often persuasive, capable of influencing opinion and policy; and just as the playing of concert pianos is limited to a few experts, so also the creation of maps should be limited to trained professionals. Attempts have been made in several jurisdictions to restrict mapping practice to accredited professionals, in some cases with success.

Another dimension to this argument surfaced in the early 1990s as part of a rapidly emerging social critique of GIS. In the late 1980s, Brian Harley had introduced the concepts of deconstruction to cartography, arguing that all maps were social constructions that could be read as texts for evidence about the agendas of their makers. These arguments found ready acceptance, since maps have always been important as tools of power, and mapmakers have often followed quickly on the heels of conquerors. The selection of features for maps, and the styles in which selected features are rendered, are part of the process of cartographic design, and clearly open to conscious or subconscious manipulation for purposes that range from visual clarity to more sinister coercion.

At the time, cartographers saw GIS as a new and wildly popular tool that appeared insensitive to these ethical arguments. GIS seemed to be grounded in the naïve assumption that one could achieve a scientifically

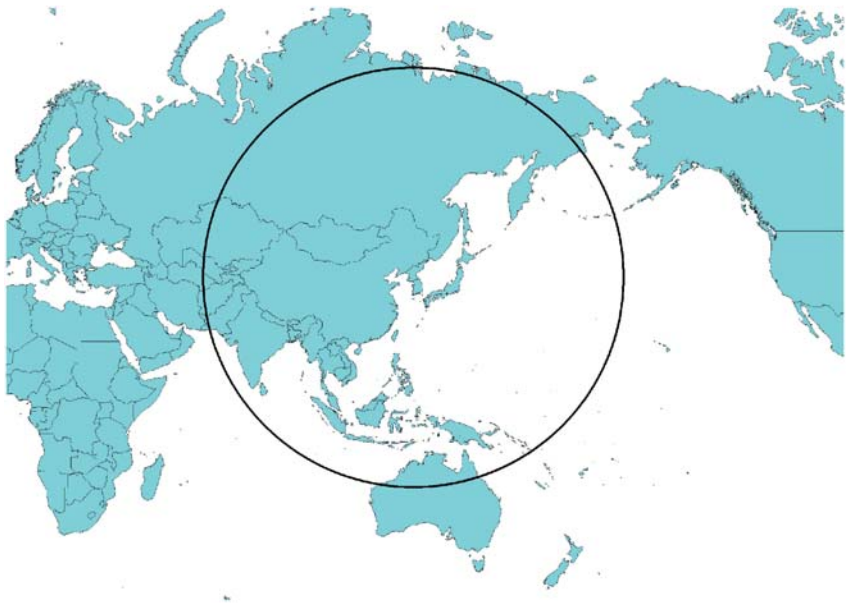

Figure 1 A circle centered on Pyongyang and superimposed on a Mercator projection. Similar maps appeared in several newspapers and magazines to illustrate the threat of North Korean missiles with 10 000 km range. However the locus of points that is 10 000 km from Pyongyang is not a circle on the Mercator projection, and includes the North Pole and much of the continental US.

rigorous description of the world, and store it in precise form in a digital computer – that the contents of a GIS database represented the results of objective, replicable scientific measurement. The ability to compute measures such as area to large numbers of decimal places served to reinforce this view, and it was clear that GIS was being marketed by commercial software developers as a scientifically rigorous approach to geographic problems. To cartographers influenced by Harley, GIS users were the barbarians, sensitized to none of the nuances of mapping practice or to the tension between cartography as science and cartography as art.

These critiques came to a head in the early 1990s, and led to a series of meetings in which each side slowly achieved an understanding of the other's position. Today, the social context of GIS is one of the major themes of GIS research, and many developments have addressed the issues raised by Harley and others. The notion of objective truth, reflected in the use of such terms as accuracy and error, has been replaced by concepts of uncertainty and of the relationship between truth and power, in the realization that many aspects of GIS practice are not replicable, that many of the key definitions are inherently vague, and that seemingly objective technologies can be molded to the agendas of their owners and sponsors.

GIS as Science

GIS is an attractive computer application, combining the visual and the numeric, and it is not surprising therefore that it has captured the attention of many professionals with backgrounds in science and engineering. Cartography may indeed be part art and part science, but to these scientifically minded GIS users the new world of GIS offers the opportunity to escape the constraints and vagueness of the past, by developing new ways of describing the world that are indeed replicable and objective. Science has always attempted to sweep away the vague and subjective, replacing terms such as cold, warm, and hot with replicable measurements on standard scales – why not, therefore, replace the vague and partially subjective classifications of a soil map with measures of soil properties that are both replicable and useful to farmers? From this perspective the power of GIS lies in its ability to improve on past practice, rather than to replicate it in a digital environment.

The transition to digital technology had both technical and economic impacts on map making. On the technical side, it allowed entirely new approaches to map design that escaped the constraints of manual cartography. For example, line widths could be varied continuously, rather than being constrained by the width of a nib or scribe tool. Shading and color could also be varied continuously, leading to many more possibilities in the portrayal of spatially continuous phenomena such as topography or temperature. In short, the new technology helped to underscore for the first time the degree to which cartographic practice had been constrained by the technology of manual map making – by the marks a human could make with a pen and ink.

Nowhere is this more evident than in the debates over interval free choropleth maps. These maps are used to portray statistical information, such as population density

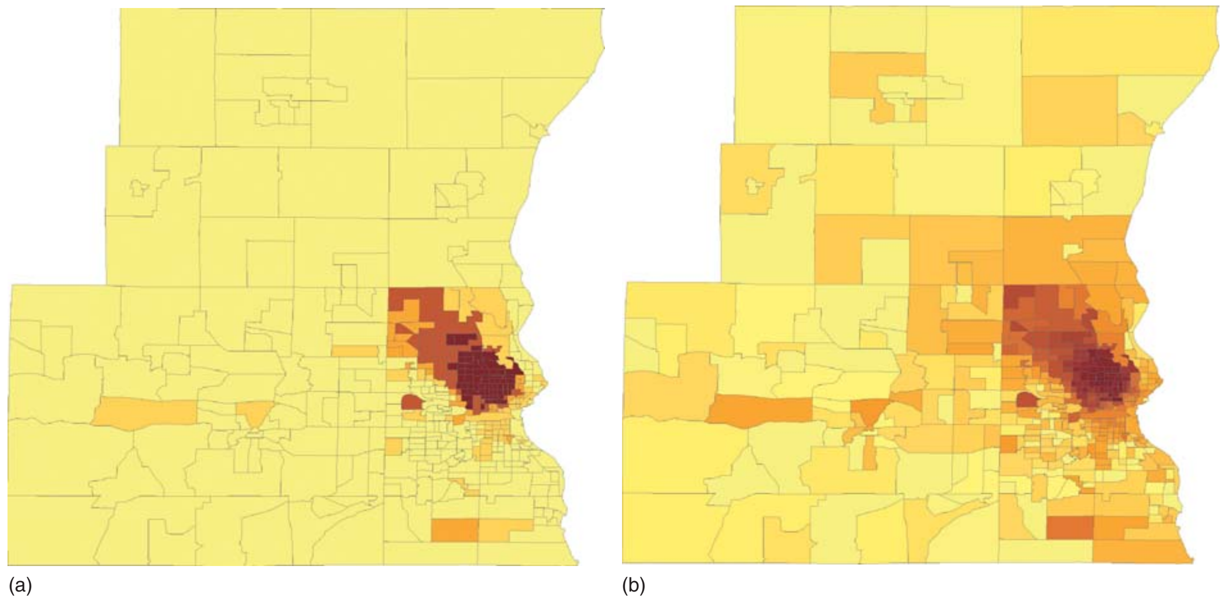

Figure 2 Two choropleth maps of Milwaukee census tracts, showing percent black. (a) Uses five classes following the Jenks natural breaks method of class interval selection. (b) Uses 32 classes, and approaches the Tobler interval-free technique.

or average income, for irregularly shaped reporting zones such as countries or postal code areas. The attributes of each area are shown using color or shading, each color or shade corresponding to a defined part of the range of observed values. The relationship between value and color or shade is usually shown in a legend on the map, and the set of possible shades or colors is usually limited to less than ten, reflecting the difficulty of generating many different shades or colors with traditional tools. There is a vast literature in cartography on alternative methods of partitioning the range of values – of defining what are often known as class intervals.

In the late 1970s Waldo Tobler had suggested that in the digital world this use of a finite set was no longer necessary, and that one could instead establish a continuous relationship between observed value and color or shading density in what he termed interval free choropleth mapping. This would have the advantage that the information lost during classification would be preserved, since it would no longer be necessary to lump a range of values into each class, and as a result the map would show a more accurate picture of the data (note, however, that choropleth maps also distort the spatial aspects of data by averaging or lumping spatially continuous phenomena within discrete, irregularly shaped areas). Accuracy and the avoidance of information loss are clearly close to the heart of scientifically trained professionals.

The response from the cartographic community was overwhelmingly negative, and today it is difficult to find mention of this technique in textbooks, or implementations in software (ESRI's ArcGIS, the leading GIS software, supports many strategies for defining class intervals but does not support interval free choropleth mapping, or allow the number of classes to exceed 32). Many cartographers argued that the assignment of class intervals was a vital part of the cartographer's art, and an essential tool in using the map to tell his or her chosen story about the data (**Figure 2**).

Today there continues to be a strong association between GIS and science. The term 'geographic information science' was coined in 1992, and serves two somewhat distinct purposes: first, to describe the set of fundamental scientific questions and principles that underlie GIS, and second, to describe the use of GIS in a scientifically rigorous context. Cartography as a discipline clearly overlaps strongly with both interpretations of GIScience, but while there have been efforts to emphasize the scientific aspects of cartography, most notably in the analytic cartography pioneered by Waldo Tobler, the artistic underpinnings of the field remain as important as ever.

Current Research

Both cartography and GIScience have developed and published strongly overlapping research agendas in recent years, and a comparison makes interesting reading, for both the similarities and the differences. Both emphasize representation, and the need for continuing research on how to extend existing methods to capture phenomena that are complex, dynamic, and three dimensional. Automation of the map making process, and particularly the thorny problem of automatic generalization, remains a

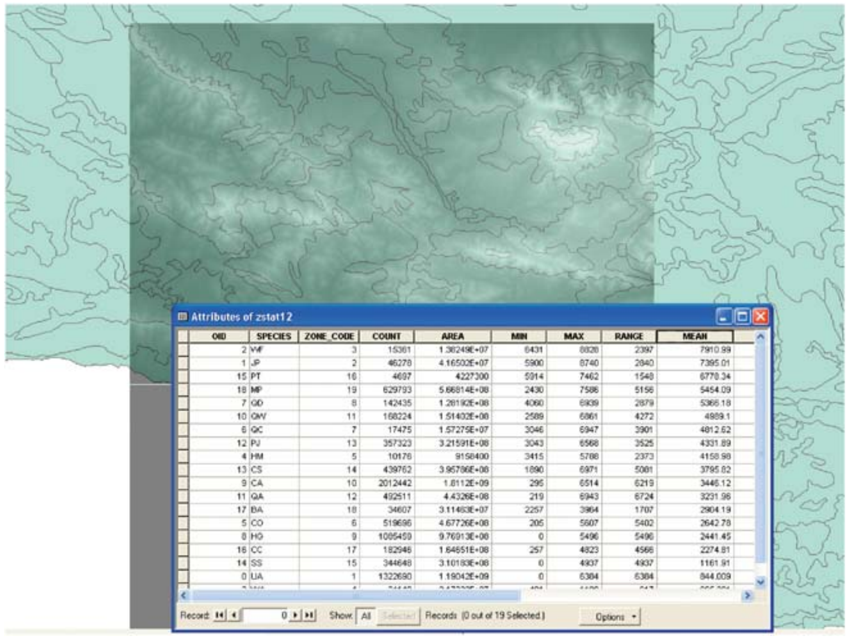

Figure 3 An analysis of the relationship between vegetation cover (each green area is a different vegetation cover type) and elevation (the black and white image). The results of such a GIS analysis are more likely to be presented in numeric form in a table (the column headed 'MEAN' shows the average elevation for each SPECIES).

concern of cartographers, as does the design of user interfaces, particularly in devices such as mobile phones that offer only limited area for visual display, and the role that cognitive science can play in helping to address such issues. Both fields are concerned about uncertainty, and how to convey what is known about it to the user; and both fields are concerned about the rapidity of technological change, and the abilities of researchers and developers to exploit innovation.

The Power of the Visual

In the mid 1960s design of the CGIS there was no support for visual output or for map making, and although the database included full representations of the contents of maps, the only output was tabular and numeric, in the form of summary measures of area (**Figure 3**). Even today a tension continues between visual and numeric output in GIS, and one still finds the occasional pejorative use of phrases such as 'merely making a map', implying that numerical analysis is the true purpose of GIS, whether it is the inductive search for pattern and anomaly or the deductive testing of hypotheses. Recent advances in data modeling, particularly the adoption of object oriented techniques of database design and CASE (computer assisted software engineering) tools, have made it possible to represent phenomena in GIS that were never suitable topics for map making, including events and transactions, flows, and fine resolution time series.

At the same time maps and visual displays remain a central part of all geographic information technologies, and provide much of the attraction that draws new recruits to the field of GIS. Maps are in many ways visual trophies that can be incorporated to great effect in presentations and posters, and hung on walls for purposes that go far beyond the objective representation of information. GIS is still explained to the uninitiated using the metaphor of maps, and students are still judged in GIS classes in part on the visual quality of the maps they produce. Students are trained to be sensitive to the degree to which maps can mislead and even lie, and to the importance of careful visual design in telling a desired story.

The Relationship Today

The past few decades have seen a steady convergence of cartography and GIS, as the former has become more embedded in the digital age, and as GIS professionals have grown more sensitive to the issues that cartographers address. Both fields have made strong efforts to emphasize the principles underlying their techniques, and to show that dealing with spatial information, particularly information about the surface of the Earth, requires a special perspective based on special concepts. The question "What is special about spatial?" has been asked many times, and has been addressed very effectively by recent research.

Nevertheless significant differences remain, and it is possible to identify different flavors in the ways geographic information technologies are used. To a cartographer the primary means of communicating the results of investigation is through the map. To GIS users, however, the map is only one way of communicating information, which may also be presented in the form of text, numbers, sounds, and images or illustrations that would not qualify as maps in the normal sense of the term. Maps are compelling, but other media may be more appropriate in a given context. The definition of a map is changing also, and there is an increasing interest in maps that distort the geometric world in specific ways; in animated and three dimensional maps; and in maps of spaces other than the geographic.

Cartography has much older roots than GIS, even though analog precursors of many GIS functions can be found dating from hundreds of years ago. This tends to give cartography a more scholarly flavor, and it will be many years before society accords the same value and significance to historic applications of GIS as it does to historic maps. Maps are accorded places in libraries that GIS databases are not, and often used as decoration; and the English language assigns a single word to a practitioner of cartography but not yet one to a practitioner of GIS (GISer can be found in some of the critical writing of the early 1990s but was not widely adopted).

The legacy of Brian Harley lives on in the general perception that cartography is the more artistic, subjective side of geographic information technology, and GIS the more scientific, objective side. Scientists using computers to make maps are more likely to call what they do GIS than cartography, and more likely to assume that the information they work with is the result of scientific observation and measurement.

To the outside world, however, these nuances are largely invisible. By now the general public is well aware that important things are happening in geography, whether it is as a result of a GPS enabled navigation system in a car, or exploring the world through Google Earth, or geotagging a photograph. In this world, the concepts and principles in which both cartographers and GIS professionals are trained are largely invisible, and it is possible for entire books to be written on the subject of map making with Google Maps without a single reference to the literature of either field. Formal training in what it means to think and express oneself spatially remains the preserve of a few, in contrast to formal training in thinking numerically or in the use of language.

See also: Art and Cartography; Critical Cartography; Critical GIS; GIScience and Systems; Geovisualization.

Further Reading

Clarke, K. C. (1990). *Analytical and Computer Cartography*. Englewood Cliffs, NJ: Prentice Hall.
Clarke, K. C. (2001). *Getting Started with Geographic Information Systems*. Upper Saddle River, NJ: Prentice Hall.
Dykes, J., MacEachren, A. M. and Kraak, M. J. (2005). *Exploring Geovisualization*. Amsterdam: Elsevier.
Erle, S., Gibson, R. and Walsh, J. (2005). *Mapping Hacks: Tips and Tools for Electronic Cartography*. Sebastopol, CA: O'Reilly Media.
Goodchild, M. F. (1988). Stepping over the line: Technological constraints and the new cartography. *American Cartographer* 15, 311 319.
Jones, C. B. (1997). *Geographical Information Systems and Computer Cartography*. Harlow: Longman.
MacEachren, A. M. and Kraak, M. J. (2001). Research challenges in geovisualization. *Cartography and Geographic Information Science* 28(1), 3 12.
McMaster, R. B. and Usery, E. L. (eds.) (2005). *A Research Agenda for Geographic Information Science*. Boca Raton, FL: CRC Press.
Slocum, T. (2003). *Thematic Cartography and Geographic Visualization*. Upper Saddle River, NJ: Prentice Hall.

GIS and Society

T. Nyerges, University of Washington, Seattle, WA, USA

Glossary

Alternative Representations Different approaches to portraying geospatial information to address a particular task.

Critical GIS Social theoretic issues that underlie geographic information science and systems.

Geographic Information and Society A relationship between geographic information use and a context within a society.

Geographic Information System (GIS) A combination of database management capabilities for collecting and storing large amounts of geospatial data, together with spatial analysis capabilities to investigate geospatial relationships among the entities represented by that data, plus map display capabilities to portray the geospatial relationships in two- and three-dimensional map form.

Participatory GIS Members of community group(s) directing the use of GIS to address community outcomes.

Public Participation GIS Members of a community responding to local government request to participate in the governance process of that community.

Social Construction Influences from society directing development and use of geographic information systems.

Society A community, nation, or broad grouping of people having common traditions, institutions, and collective activities and interests, or more specifically an enduring and cooperating social group whose members have developed organized patterns of relationships through interaction with one another.

Spatially Integrated Social Science Social science explicitly recognizing a spatial component that improves understanding in the science.

Scoping GIS and Society Activity

Geographic information system (GIS) technology has emerged in the latter part of the twentieth century as an integrative information technology. In its most basic form, a GIS includes database management capabilities for collecting and storing large amounts of geospatial data, together with spatial analysis capabilities to investigate geospatial relationships among the entities represented by that data, plus map display capabilities to portray the geospatial relationships in two and three dimensional map form. In the 1980s, a collection of academic researchers, software industry, and US federal agency personnel participated in a group process to define the term GIS as a combination of hardware, software, data, people, procedures, and institutional arrangements for collecting, storing, manipulating, analyzing, and displaying information about spatially distributed phenomena for the purpose of inventory, decision making and/or problem solving within operations, management, and strategic contexts. Three perspectives are represented within that definition. The first perspective involves components of a system: a combination of hardware, software, data, people, procedures, and institutional arrangements. The second perspective includes processes utilized within a system: for collecting, storing, manipulating, analyzing, and displaying information about spatially distributed phenomena. The third perspective provides motivation for system use: for the purpose of inventory, decision making and/or problem solving within operations, management, strategic contexts; such contexts include use by governments, businesses, research and not for profit organizations, as well as by community groups. The definition implies a relationship to society because it contains references to people, organizations, and purposes for using information technology.

During the 1980s and 1990s, GIS was adopted by subunits within organizations like governments and businesses. As such, GIS technology is pervasive in larger organizations across the world at the present time. GIS applications exist in dozens of sectors including but not limited to:

- agriculture/farming, libraries
- weather – research and news media
- ecosystem management
- census enumeration – federal, state, and local
- defense intelligence
- public utilities – gas, cable, electricity, and water
- exploration – oil and gas
- environmental management
- forest and timber management
- mining
- ground and surface water resources
- public health
- historic preservation and archaeology
- national security
- land management
- local, state, and federal government planning
- marine and coastal resources

- petroleum refining
- telecommunications
- highway–rail transit transportation
- and many more

In the early to mid 1990s, many researchers began to question the overall intent of the technology and its seeming bias toward larger organizations, in contrast to individuals and community groups. It was about then that concerns about development and use of information in societies across the world began to surface to look behind some of the technological assumptions of systems, showing that it privileged certain groups of people over other groups in society. In the 2007 online version of the Merriam Webster's dictionary, the term 'society' is taken to mean a community, nation, or broad grouping of people having common traditions, institutions, and collective activities and interests, or more specifically an enduring and cooperating social group whose members have developed organized patterns of relationships through interaction with one another. The overarching concern in the societal use of GIS is how this technology will influence – and be influenced by – events, processes, mechanisms, and structures of society. What influences GIS adoption, development, and use within different societal contexts? In what ways is GIS shaping society and, in turn, in what ways is GIS being shaped by society, and by which segments of society? Research addressing the interrelationships between GIS and society will, of necessity, often be an aspect of other GIS research, programmatically and at the project level, and also of more general research into the relationship between new information technologies and society. Nonetheless, examining relationships among GIS and society as a focus of study in itself is worthwhile.

During the mid 1990s, much of the motivation for GIS and society research was coming from critics of GIS technology resulting in a seemingly polarized view about GIS–society relations. There were enthusiasts of GIS and related technologies expounding on the positive forces 'for good', as GIS was aiding in the expansion of human capacities for knowing accurately and intervening creatively in their 'worlds'. In contrast to that view were those who were/are ultimately highly critical of GIS, seeing the technology as yet another form of technology leading to the 'colonization' of everyday lives, a technology with enormous potential to render people much more visible, locatable, traceable – to become 'objects' of a new 'governmentality'. Still others worried about everybody being 'geo coded', hence the term 'geoslavery' was coined to imply that people can be tracked everywhere and at anytime. GIS and its related technology of geospatial positioning devices provides for extended visibility, surveillance, and traceable data trails. The latter activities are part of research that claims GIS and related geospatial technological activities fall within the broader critiques of Western philosophy, capitalist and neoliberal political economies, changing forms of governmentality, and other topics, all of which seemingly conspire to demand a more 'visible' social (and sometimes 'natural') world – in which everything can, in principle, be appropriately 'geo coded'. The fear that 'big brother is watching us' now exists in everyday life. The hundreds of video cameras in downtown London, and in other major cities around the world, wherein the cameras installed by police departments watch at all hours the everyday moves of traffic and pedestrians on the streets. Of course, the cameras did capture images of the subway bombers in the summer of 2005, and the cameras continue to patrol for crime everyday, perhaps a benefit by original intent.

The debate about the 'good' side and the 'evil' side of geographic information technologies continues, but in a much more nuanced form. The simple polarization of 'good and evil' is a naïve perspective. The simplicities of such bipolar thinking begin to unravel quite readily once one begins to more fully research, deconstruct, and reconstruct the relationships at play. Thus, the naïve perspective has given way to a 'critical GIS' perspective. Agents, whether as industrial/commercial, governmental, or not for profit oriented have their particular interests to pursue in a world that continues to change, with the fears of 'information good' and 'information evil' still before us. To unravel simple perspectives, all information intents need be situated to understand the particularities about GIS technology development and use. This more systematic pursuit of geographic information and society relations is part of a growing interest in wanting to create a foundation for how we see the development and use of geographic information.

While GIS and society studies were developing during the mid 1990s, many researchers of the GIS and society effort, as well as many other GIS researchers from various research niches, were becoming increasingly dissatisfied with the conceptual underpinnings of GIS software. Those researchers wanted to explore underlying GIS concepts, and to help articulate a science if one could, and thus the S in GIS became associated with the term science – a GIScience. An international conference by that name occurs every two years. Many GIS researchers see the phrase GIScience and Society as an expanded perspective for GIS, particularly researchers who view GIScience as part of the broader field of information science, and those who adopt a critical GIScience perspective.

While all that was occurring, other researchers in the late 1990s were seeing new developments of hardware and software associated with the Internet, and thus some researchers started to use the term services for the S in GIS. The technical activity spawned still further interest in societal aspects of GIS and thus, some now see all four words – science, systems, services, and society – being associated with the S, while the GI (geographic information) part has been broadened to 'geospatial' information. Consequently, the S in GIS has become a rather expansive idea as the technology continues to spread worldwide in more dimensions than can be counted in any practical sense. Additional relationships between GIS and society can be glimpsed through topics such as Critical GIS; Feminism, Maps and GIS; Geographic Information Science and Systems; GIS and Cartography (Relationship); Mobile GIS/Mobile Mapping/Location based Services; Public Participation GIS; Qualitative GIS; and to some extant by Social and Cultural Geography, Society; Development, Civil Society; and Philosophy, Society space.

With all of the above activity, it is not difficult to believe that GIS as science, systems, services, society, research, and practice is very alive and well across the world. However, although the topic of GIS and society is quite vibrant, there are many tensions at play. The research and practice aspects of GIS and society are in full blossom year round. Together and separately, research and practice involved with GIS and society are responsible for creating major interests in the still emerging interest of geographic information. GIS and society research is different than GIS and society practice, but both are very important in their separate and combined ways. There is a link between the two – as inabilities, inequities, and overall challenges with using geographic information in practice has motivated research to improve practice. Research, as the systematic pursuit of knowledge to further societal aims, and practice, as the pursuit of action to further societal aims are blooming. The material herein tends to address more of the research and educational aspects, because it is that process that is better documented than the practice. Nonetheless, the practice of GIS is very important, and it continues to grow by leaps and bounds as GIS applications diffuse to topics and situations around the world. Researchers as well as practitioners, and particularly many that are not geographers, around the world find map based expression to be quite powerful and engaging.

A Chronology of GIS and Society Research Events

Since the early 1990s there have been several research events – workshops, conferences, special issue journals – contributing to the development of GIS and society research activity. Much of the research activity reports on aspects of practice, but the emphasis is on research as systematic investigation. The US National Science Foundation funded the National Center for Geographic Information and Analysis (NCGIA) in the late 1980s and early to mid 1990s to synergize the development of the basic concepts – toward a basic science of geographic information. In 1993, NCGIA sponsored a GIS and Society Research Workshop at Friday Harbor Washington, and from that a special issue of *Cartography and GIS* was published in 1995 to present many of the basic issues relating GIS to society. These two activities are often cited as the major motivation for GIS and society research.

A few years later, in 1996, the NCGIA sponsored a research workshop about public participation GIS (PPGIS) at the University of Maine that was reported in a special issue of *Cartography and GIS* in 1998. Only a few months later in 1996, the NCGIA Varenius project sponsored a broader workshop about GIS and society, the results of which were reported in a technical report. Then, in 1998, the NCGIA Varenius project sponsored a research workshop titled Empowerment, Marginalization, and PPGIS, and a research report was produced to document the results. In the same year, the NCGIA Varenius project sponsored a research workshop about Geographies of the Information Society, after which a book titled *Information, Place, and Cyberspace* was published in 2000 to document the findings. To add more momentum to the effort, the University of Minnesota in cooperation with NCGIA sponsored the International Conference on GIS and Society; a proceeding was published to document the effort in 1998. In 1999, several of the researchers from the NCGIA Varenius project authored a summary and prospective article about geographic information and society. The GIS and society research effort started to focus a bit more with the NCGIA Varenius project sponsored workshop titled Community Participation and Geographic Information Systems, with many of the contributions being published in the form of a book in 2002.

In 2003 to follow, the Association of Geographic Information and Laboratory in Europe and the US University Consortium for Geographic Information Science cosponsored a workshop about access to geographic information and participatory processes that was followed with a special issue of the *Urban Regional Information Systems Association Journal.* Moving into the topic of critical GIS, a special issue on GIS and Social Science was published in *Cartographica* in 2004. The NCGIA research activity moved in the direction of a workshop reported later in a book titled *Spatially Integrated Social Science.* This was followed by a long term and more general perspective on the GIS and society research agenda developed by several co authors who were members of the University Consortium on Geographic Information Science and published by Taylor & Francis in 2005.

Continuing as a subset of the GIS and society research activity, there were six PPGIS conferences 2002–07 sponsored by the Urban and Regional Information Sys tems Association (URISA). The first few were separate conferences, but PPGIS has now been included as a one day co convened event with the annual URISA confer ence. Several of the documents – reports, journal issues, and books – created from those events are listed in the further readings section.

Research Themes and Topics

Based on a review and synthesis of the topics contained in the above documents, seven themes appear to be fairly representative of the research activity underway.

- Fundamentals of Geographic Information and Society
- Spatially Integrated Social Science
- Alternative Representations of Geographic Infor mation and Society
- Organizations and Institutions
- Ethical and Legal Issues
- Group, Public, Participation, and Community Issues
- GIS/Society Impacts and the Social Construction of GIS

An exhaustive coverage of the themes and topics would take considerable space; therefore, a sample presented below presents the reader with a basic appreciation for the breadth of the research. We will start with the fun damentals of geographic information and end with the social construction of GIS. In many respects, one might say this sequence starts with topics that are more related to deep seated longer term issues and ends with more everyday connections. This is not to say that the former are more associated with basic research and the latter more applied, but it is suggestive that the former themes are more 'academic society' and the latter are more 'everyday society'.

Within theme 1, fundamentals of geographic infor mation and society, among the fundamental concepts are ontological and critical social theoretic perspectives on geographic information creation and use. Studies of ontology in geographic information science have been a major activity over the past 10 years, involving how in dividuals and organizations understand the world around them. These studies focus on interpretations of what exists within the realm of reality for an individual and/or organization. Of course, groups of people reenforce each others' ontological perspectives but this perspective has not yet received much attention in the research. Critical social perspectives about how geospatial information is created and interpreted are part of a growing research activity. The meaning of information deals with identity of place as relates to power relations of place, and of course the two issues are related. Regardless of how people understand their connections to place in the world, new technological methods, such as location based services in mobile devices, continue to get de ployed in everyday reality. Furthermore, applications about basic issues such as hunger, employment, and housing also continue to be investigated.

Within theme 2, spatially integrated social science deals with a latent understanding about how many of the social sciences connect to geospatial perspectives, but without explicitly recognizing those connections. Once again, an ontological perspective is important, and per haps even more important as scale matters in social sci ence applications. Research is underway to explore the changing significance of key geographical concepts in the information age as applied to nonspatial social sciences and reflected in the character of GIS. Social networks are being given a spatial dimension (call spatialization), and studies of the World Wide Web (WWW) are taking on an explicit spatial dimension to understand relations among rather abstract information concepts. For example, re searchers are using GIS to better understand communi cation breakdowns that occurred during hurricane Katrina in 2005 on the US Gulf Coast. Lending a spatial dimension to communications networks fosters deeper understanding of the social connections and gaps about information flow.

Alternative representations of geographic information and society within theme 3 focus on geovisualizations depicting the nature of geospatial knowledge, for ex ample, local knowledge about place versus technical knowledge developed from data measurements. Often times alternative cartographic map representations can portray realities from different perspectives and thus provide a sense that multiple realities exist, for example, about how to develop natural resources in a community. Some researches focus on the knowledge and its differ ences and other researchers focus on the cartographic design and geovisualization methods. Applications relating to Native American (indigenous) Indian or Australian Aboriginal issues are part of this theme. Geovisualizations of spatial equity and their connections to urban environmental justice issues are also quite important.

Organization and institutional issues associated with GIS have been investigated for quite some time. Part of this research, in theme 4, has examined the roles of GIS in the public, private, and not for profit sectors, and the relationship of organizations and institutions to social change – as an external influence, while another part of the research has looked at the influence of GIS in or ganizational change – an internal influence. One of the long term methodological topics has been about system integration, and this leads to integration of various geospatial technologies like remote sensing, GIS, and global positioning systems. Along these lines, different levels of local, national, and global data infrastructures are emerging, making it easier to acquire spatial data in a variety of data formats; but at the same time, these readily available datasets influence (bias) what we know about the world. Studies about the appropriateness of data use can be counted among these studies, but they are best contextualized to specific organizations. As such, studies about the implementation of GIS within and across organizations have been very important types of studies. Data fitness for use is a major part of those kinds of studies.

Ethical and legal issues as the basis of theme 5 address ethics and legality in geographic information use. It might be legal to use certain kinds of data for certain applications, as for example, housing location from some of the readily available mapping programs from Google and Yahoo, but is it ethical to use such data when one highlights certain characteristics about households. Place and identity are very important issues for those who fear a loss of such identity in the digital age. Although we can promote easier public access to information, at the same time security leaks associated with private information present a privacy and confidentiality dilemma. Therefore, these studies must focus specifically on the types of information to be made available, about whom, and at what aggregate level of resolution. Another important topic associated with this theme involves error, risk, and liability of data use. Making decisions with poor quality data can involve both ethical and legal challenges, particularly when such data are originally known to be of poor quality, rather than discovered later to be of poor quality. What it means to be of good and poor quality is all part of this research area.

Theme 6 has perhaps the largest number of researchers of any subarea in the GIS and society realm. Groups, participation, public, and community are all issues that require rather nuanced interpretations to understand how collections of people make use of geographic information. Whether the public is constituted of average individuals or collections of stakeholder groups, and how they might work with GIS data either in face to face settings or across the WWW are issues of concern for researchers in this subarea. Among some of the subtopics are concerns about empowerment and marginalization of community groups. What kind of participatory processes, for example deliberative democratic or representative democratic, are better for what kinds of issues, like involvement in hazardous waste cleanup in local communities, is a major concern. There might be educational, social, political, and economic barriers to access to information, and perhaps there are exemplary ways that communities can overcome such barriers. Much of this deals with political power relationships that influence who has access to GIS information for community (or for that matter local government) decision making. Many studies have empowered community groups, making use of existing GIS technology, to address community problems. Some researchers have been more successful than others, despite the fact that most GIS software designs were not intended for casual users. Enabling community groups in decision processes has been more or less successful, given the current state of software usability by diverse groups. Consequently, software development for decision support of diverse groups is an important topic associated with this theme. New software designs that foster alternative representations of information from participatory perspectives can be considered part of the effort to develop second generation GIS – or GIS2. Modeling and implementation of new participatory processes are part of the GIS2 activity. WWW based approaches to participation are counted among these activities.

The final theme provides, perhaps, some of the best examples of GIS and society research. How GIS influences changes in society and how society influences changes in GIS are among the most prevalent of topics. Power relations associated with information within organizations and societies are among the topics. Societies across the world are moving toward increased digitalization. However, the gap in access to digital resources – called the digital divide – might be growing wider in some countries than others. Studies, in some cases, are not about access to computers, but who has access to training. Online communities are growing. The phenomenon called social computing, mostly for the young, is becoming ever more pervasive. How the social aspects of computing influence the development of computing systems, and how the computing systems change the dynamics of people interacting with people, and virtual worlds continue to grow in importance as more and more people are affected by these developments. A major issue in this research theme is the social reflexivity that occurs when the technology is used to create 'social class', wherein it is supposedly being used to identify social class to be used by various geodemographic marketing organizations. A problem arises when the act of identifying groups together with marketing pizzazz is actually creating consumption patterns, rather than responding to the whim of a consuming public. Clearly, this kind of research links to other themes such as the institutional and organizational theme 4 and the ethical issues of theme 5.

In 2005, the University Consortium for Geographic Information Science completed development of a nearly 10 year effort to articulate a *Research Agenda for Geographic Information Science*. Chapter 10 of that book describes five of the themes listed above and provides an array of research questions that will likely help guide GIS and society research for some time to come. The themes for integrated social science and alternative representations are two additional themes drawn out in this new synthesis.

Ongoing Debates

As GIS and society research has emerged, several critiques were levied against GIS technology, and those critiques remain part of the rich debate about various topics – clarifying, questioning, and rethinking – relationships between GIS and society. The debates provided below are sequenced roughly in the order in which they appeared in literature. Some are more vibrant than others at the time of publication of this encyclopedia.

A foundational debate is about the representational character of a Cartesian geometry. Cartesian geometry, like Cartesian thinking in general, is formed by identifying dimensions that are mutually exclusive when it comes to measuring data values. This same fundamental bias of exclusivity is questioned when it comes to describing phenomena of everyday reality. It is hard to break the world into finite categories and expect to provide full insight, when the world is always in flux.

A second debate is about ethics and privacy as it concerns the use of geographic information. Many researchers would like to provide very fine resolution locations for the whereabouts of delivery drivers, mountain climbers, children, and pets among other cherished humans and animals. However, other researchers claim that such high resolution knowledge is an invasion of privacy, and it is unethical. Many of the concerns depend, of course, upon the situation. But writing laws and regulations cannot always take circumstances into account. So, the debate continues about what to do depending on the situation.

A third major debate is about the nature of participatory processes that involve geospatial/geographic information interpretation, understanding, and decision making. Participatory GIS adherents would like the term 'participatory' to refer to a bottom up, grass roots effort by community members to engage in discussion about their world, while making use of geospatial/geographic information. A frequent process involves community resource mapping. In that view, a community is not the same as a public. PPGIS researchers agreed that such a view is very important, but do not bother to distinguish that view from searching for a better way to characterize publics as stakeholder groups, and support all groups with geospatial information technologies appropriate to the needs of those who wish to use it. The discussion has been ongoing for years. Only with additional systematic research results that address all of the issues can researchers sort through the important aspects of the debate and provide informed perspectives that aid understanding.

A fourth debate focuses on the level of detail needed in evaluations of peoples' use of information. This debate is about whether sophisticated research designs are necessary for building knowledge about GIS use. The debate is perhaps more about how much work is necessary and sufficient to build knowledge that is valid – when validity has a variety of interpretations and meanings. Nonetheless, without research designs it is impossible to know how information and findings can be triangulated to build knowledge. Some proponents contend that more research using qualitative research approaches are necessary to provide nuanced insight into the character of GIS. Other proponents contend that both qualitative and quantitative based (or mixed methods as they are commonly called) research methods are necessary to provide comprehensive insights. Both qualitative and quantitative have a place.

However, the real debate is perhaps over whether the term qualitative GIS is actually communicating something new, as all studies start out with forming categories of something, whether in database design categories or studies about politics and information use, then from those categories we refine our observation (measurement) into more nuanced levels of specification – the quantitative measurements needed to make finer distinctions.

Other discussions among many topics exist, but these four debates provide the reader with a sense that there is much more to be done, as GIS and society topics continue to emerge for investigation.

See also: Civil Society; Critical GIS; Ethical Issues in Research; Feminism, Maps and GIS; GIScience and Systems; GIS and Cartography; GIS, Mobile and Locational Based Services; GIS, Public Participation; Qualitative Geographic Information Systems; Society-Space.

Further Reading

Burrows, R. and Ellison, N. (2004). Sorting places out? Towards a social politics of neighborhood informatization. *Information, Communication and Society* 7(3), 321 336.
Craig, W. J., Harris, T. M. and Weiner, D. (eds.) (2002). *Community participation and geographic information systems*. London: Taylor and Francis.
Goodchild, M. F. and Janelle, D. G. (eds.) (2004). *Spatially integrated social science*. Oxford: Oxford University Press.
Harris, T. and Weiner, D. (eds.) (1996). GIS and Society: The Social Implications of How People, Space, and Environment Are Represented in GIS, Specialist Meeting 19, http://www.ncgia.ucsb.edu/Publications/Tech Reports/96/96 7.PDF
Janelle, D. and Hodge, D. (eds.) (2000). *Information, place, and cyberspace*. Berlin: Springer.
Longley, P., Goodchild, M., Maguire, D. and Rhind, D. (2005). *Geographic information systems and science*. Chichester: Wiley.
McMaster, R. and Usery, E. L. (eds.) (2005). *A research agenda for geographic information science*. Boca Raton, FL: CRC Press.
Obermeyer, N. J. (1998). The evolution of public participation GIS. *Cartography and Geographic Information Systems* 25(2), 65 66.
Pickles, J. (ed.) (1995). *Ground truth: The social implications of geographic information systems*. New York: The Guilford Press.
Poiker, T. and Sheppard, E. (ed.) (1995). *GIS and society*, special issue of *Cartography and Geographic Information Systems* vol. 22(1).

Schuurman, N. (2006). Formalization matters: Critical GIS and ontology research. *Annals of the Association of American Geographers* 96(4), 726 739.
Schuurman, N. and Kwan, M. P. (2004). GIS and Social Science. *Cartographica* 39(1).
Sheppard, E., Couclelis, H., Graham, S., Harrington, J. W. and Onsrud, H. (1999). Geographies of the information society. *Internatinal Journal Geographical Information Science* 13(8), 797 823.

Relevant Website

http://www.urisa.org
PPGIS Conference topics 2005, Urban and Regional Information Systems Association.

GIS, Mobile and Locational Based Services

J. Raper, City University, London, UK

Glossary

Application Logic They are programming statements that execute computations and transformations on data in line with some technical or business purpose.
Content Management Systems (CMS) They manage large collections of multimedia data and metadata.
Creative Commons It is a body designing and promoting licences for the exploitation of copyright on a wide range of discretionary terms.
Differential GPS It is the practice of correcting GPS code signals at a survey site using known errors in the GPS code signals occurring at a nearby known location.
Dijkstra Algorithm It is an algorithm that computes the shortest path between two points in a network.
General Packet Radio System (GPRS) It is a method of transmitting data to and from the Internet over a digital cellular phone service, usually at speeds of up to 64 kbps.
Geocaching It is a treasure hunt using mobile devices to find fixed caches with log books.
Geodetic Control It is a physical monumentation linking survey observations and known earth dimensions.
Geopositioning It is a technology to find current location in global or local coordinates and supply it to ubiquitous computing services.
Georeferencing It is the assignment of coordinate values to a feature recorded in a spatial database.
Global Positioning System (GPS) It is the US government-funded satellite positioning system available on a wide range of mobile devices.
Ground Control It is the physical monumentation linking aerial survey and earth-surface coordinates.
Intelligent Transport Systems (ITS) It is an automated self-adapting transport infrastructure.
Java Midlet It is a java application program that can execute independently on a computing platform in either signed (approved) or unsigned (nonapproved) modes.
Light Detection and Ranging (LIDAR) It is a system for measuring the time taken for the pulse of a survey laser to be emitted, reflected, and returned to an instrument, which can be converted to a distance measurement.
Micro-Electro-Mechanical Systems (MEMS) It is an integrated system on a silicon chip containing microelectronics and micromechanical systems.
Ortho-Rectified It is the rotation, scaling, and warping of an image until it fits the shape and configuration of the land surface depicted in an image.
Photogrammetry It is the technique of making precise measurements on photographs, whether aerial or terrestrial.
Primary Triangulation They are the precisely known connections between the survey stations known most accurately in a triangulation network.
Personal Digital Assistants (PDAs) They are handheld computers controlled using a stylus and touch screen.
Real Time Kinematic GPS It is the practice of correcting GPS signals at a survey site using GPS phase signals at a nearby known location and measuring the precise difference in position from the known location to the survey site.
Ubiquitous Computing It is a model of computing in which devices interconnect and share services pervasively through wired and wireless communications.
Unmanned Aerial Vehicles (UAVs) They are small remote-controlled planes designed to carry monitoring or military equipment such as aerial imaging technology.
Wireless Application Protocol (WAP) Browser WAP is a protocol for the display of marked-up (self-describing) web content that is suitable for display on small screens, such as those on mobile devices.

Introduction

The technologies of mobile global information systems (GIS)/mobile mapping/location based services (LBS) first emerged in the mid 1990s when it became possible to integrate global positioning systems (GPS), mobile data connections, and geographic information into a portable device with a screen. The early solutions (e.g., the Bosch TravelPilot in 1995) were in car navigation systems with GPS receivers, maps on CD, a screen showing directions, and traffic reports broadcast on radio frequencies. However, the release of personal digital assistants (PDAs) such as the Palm (1996) and iPAQ (in 2000) allowed the creation of battery operated systems with maps downloaded into PDA memory and GPS receivers connected by cable. With the widespread implementation of the general packet radio system (GPRS) on mobile devices after 2000, it became possible to link mobile devices to the Internet wirelessly, thus adding real time information to the system.

These technologies remained restricted to professional surveyors or in car navigation in high end cars until the first mobile devices with GPS navigation, Internet access, and onboard GIS software arrived in consumer form in 2003. These integrated systems have been reduced in expense and form factor since 2003, and the increasing bandwidth of mobile communications has allowed greater use of the Internet to provide real time information based on current location in LBS. The technology for mobile GIS/mobile mapping/LBS has advanced to the level where it is both in use in professional and consumer settings, and adaptations of working practices and lifestyles are underway.

Mobile GIS and mobile mapping are the professional implementations of this technology and require training and system integration to use. In contrast, LBS are designed for use on consumer devices with no prior knowledge. The prospect of such capabilities becoming ubiquitous has profound implications: for example, legislation has dictated that all new third generation (3G) mobile phones in Japan must have a GPS capability so that emergency services can pinpoint callers and avoid sending duplicate responses to the same incident.

Mobile GIS

The development of mobile GIS has allowed the technology of spatial representation to move out of the laboratory and into the field, allowing a move toward a more egocentric type of representation. Using a portable mobile device in the field has brought GIS closer to the data and to the settings where the data are produced and consumed. This has changed the way decision making and data capture with GIS can be carried out.

Defining Mobile GIS

The defining characteristics of mobile GIS are:

- geopositioning integrated into a mobile device,
- GIS data capture tools,
- a set of GIS analysis/data management functions that can be used in the field, and
- access to networks from the field.

Mobile GIS based on this design are now used widely in industry and public services, sometimes based on ruggedized mobile devices. Mobility permits ubiquitous availability, true portability, autonomous use, and data reevaluation, all of which save time and increase efficiency in dealing with spatial distributions. The range of tasks for which they are well suited includes field mapping, navigation, geocoding, and ground truthing.

Implementing a mobile GIS is not simply a case of transferring the office system into the field, as the field use of GIS is a different kind of activity. Field use of mobile GIS involves detachment from services such as power and fixed networks, and involves immersion in the environment, which raises questions about the ability of the user to make appropriate judgments on matters such as classification schemes. Field use of mobile GIS is also a social process involving teamwork and engagement with people: hence, an exercise to identify and record vacant lots might uncover a property being occupied without consent. Poorly trained users and badly designed data capture schemes can be compromised by the realities of information systems use 'on the street' where there are distractions and dangers.

Using Mobile GIS

Field mapping with mobile GIS allows the recording of point, line, and area features directly into the system. Using a stream of positions from (say) a GPS receiver, allows the user to occupy the target locations and record them into the appropriate feature category, for example, a set of positions defining a parcel boundary. As the locations are occupied, the user can add attributes to tag the observation with an identity. Precise surveying at sub centimeter accuracy to mark out new developments or to establish geodetic control are beyond the scope of mobile GIS and require heavyweight survey equipment and deeper skills. Note that in some countries surveying must be carried out by certified surveyors, whose work will be covered by guarantees of professional competence.

Mobile GIS can also be used for navigation. Navigation involves following a route (e.g., to find a given address) or moving directly toward a target location (e.g., a previously used environmental sample point). Mobile GIS can be used for route finding in conjunction with road network data: this is a key function in the transport logistics industry. Most shortest route calculations are based on the Dijkstra algorithm, suitably weighted with constraints (e.g., low bridges) and additional information (e.g., congestion). Congestion information can be acquired from roadside speed monitors or from vehicles moving on the road network. Mobile GIS can be used to navigate directly to specific points over land or sea using direction indicating displays (such as compasses) or audio (using sound volume or pitch). Mobile GIS cannot be used to support high precision navigation of large marine vessels or the maneuvering of structures into position.

Mobile GIS can be used to geocode incidents, fixed assets, or mobile 'agents'. Using geopositioning and data capture tools, the operator can record position and attributes, and integrate the observations with existing databases of both fixed and mobile agents. Mobile GIS are designed to be used as primary data capture tools so that both geographic context and previous data are available to data collectors.

Ground truthing with mobile GIS allows existing geographic information to be checked for accuracy and updated directly. Being in the field with a mobile GIS allows the shape, neighborhood, and identity of features to be updated; and it also allows imagery to be checked or interpreted through a direct field assessment.

Case Study

In the last few years, a new class of handheld ruggedized devices with high accuracy GPS, large internal memory, 'working day' battery lives, cameras, and embedded GIS software has emerged from survey manufacturers at moderate cost. These 'all purpose' mobile GIS devices have become pervasive in environmental sampling, rapid response incident mapping, remote area survey, and asset inventory. City and Kingston Universities in London have been involved in a two decade long coastal monitoring project, on the North Norfolk coast, England, that has now migrated from traditional surveying using a theodolite to standalone mobile GIS tools consisting simply of a handheld device and a correction service (**Figure 1**).

Figure 1 The author with three levels of mobile GIS attached to the survey pole (from bottom to top): mobile phone with ArcPad GIS and standalone Bluetooth GPS, Trimble GeoExplorer with ArcPad GIS and differential GPS, and Leica GPS 1200 with kinematic GPS. Photograph by J. Raper.

Mobile Mapping

The development of new digital measuring and pos itioning technologies has created a wide range of new mobile mapping techniques where previously heavy and expensive equipment was needed. The arrival of mobile mapping means that mapping can become more oppor tunistic, less labor intensive, and more regularly updated. These technologies are primarily used by professionals as the cost of the new tools is still not reduced to the consumer level, and the new technologies are mostly still quite complex to use.

Defining Mobile Mapping

Mobile mapping is not one coherent field of integrated technologies, but a technological frontier where indi vidual technologies are being used to enhance specific mapping tasks. The first modern maps were made with theodolites in the late eighteenth century by finding absolute positions of fixed, typically high points from solar observations, then measuring out the edges of large triangles over the landscape to find the positions of other high points. From this 'primary triangulation', finer resolution triangulations were constructed and points of interest (POI) were positioned within this framework until a full landscape model was constructed with topo graphy and urban structure. This mapping infrastructure was refined incrementally by national surveys in the nineteenth century until it became possible to update large areas simultaneously from aerial photography in the early twentieth century. Map revision is still con ducted through a combination of aerial photography and ground survey with modern theodolites, though now based on GPS rather than solar observations.

Mobile mapping technologies have reduced the cost and logistics of all these mapping processes by the use of digital technologies. While most framework maps are still made by national surveys, the availability of these tech nologies has had the effect of democratizing the mapping process and allowing maps to be made for purposes ranging from planning to pressure group campaigns. Participatory mapping groups have begun to make copyright free maps by capturing GPS tracks from large numbers of individuals.

Using Mobile Mapping

There are five groups of technologies being used in mobile mapping:

1. The first technology group includes the range of geopositioning tools available, led by the US Global Positioning System (GPS) (and being followed by the

Russian GLONASS and European Galileo systems). GPS positioning can be carried out at a range of ac curacies from centimeters with real time phase based kinematic GPS, to 1–2 m with differential code based GPS and to 5–30 m by standalone systems. Raw GPS data are presented in latitude/longitude using the WGS84 geodetic datum, and need to be transformed into the coordinates used in national mapping. A variety of other geopositioning tools are available based on terrestrial wave based systems such as mo bile phone signals, radio or TV signals, and a host of short range communications signals such as WiFi, Bluetooth, Zigbee, Ultra Wide Band (UWB), and radio frequency identity (RFID).

2. The second technology group includes aerial or ter restrial digital image capture. Aerial photography with film cameras has been the standard for mapping data capture for almost a century, but has now largely been replaced by digital equipment. At the high end this is a very capital intensive undertaking requiring special planes, gyro stabilized platforms, and large format cameras. However, the introduction of low cost digital cameras and video camcorders has made it possible to carry out *ad hoc* aerial imaging from light planes with detachable camera mounts or model planes and un manned aerial vehicles (UAVs). In order to produce mapping that can be integrated into GIS, the captured imagery must be ortho rectified and georeferenced from GPS collected on the plane or from ground control taken from maps. This can be done in many GIS, though the full toolset for these operations can be found in photogrammetric systems, especially for stereo imagery handling. The same technologies can be used to undertake terrestrial imaging from a suit ably stabilized vehicle, though the processing of the imagery to permit accurate measurement of objects within the image or their positioning in map co ordinates requires complex transformations that are limited to specialized software and research. Aerial and terrestrial imagery produced is often highly de tailed and can be collected 'to order' at the time it is needed, for example, when the foliage is off the trees.
3. The third technology group includes a wide range of portable and affordable remote sensing techniques. LIDAR data can now be collected by relatively small and light units that can be mounted on vehicles or even trolleys. Aerial LIDAR data collection needs specially adapted planes to realize the full accuracy of the technique; however, there is a growing range of terrestrial applications in which the LIDAR scanner is moved slowly through the environment capturing precisely positioned data points in three dimensions, several hundred times per second. This has made it possible to image buildings, structures, and even crime scenes to a resolution of less than 1 cm and to import the three dimensional model created into a GIS. This model can be combined with digital photography to create photorealistic models. This technique has also been used to detect human movement in crowded places such as railway stations by setting the LIDAR sensor at ankle level and integrating between imaging cycles to identify movement within the field of view.
4. The fourth technology group includes a range of advanced measuring technologies that can be used in tandem with geopositioning. The most pervasive is the inertial navigation system (INS), which uses acceler ometers and gyroscopes to measure movement and distance directly. These are important where GPS cannot get a signal, for example, tunnels. INS have to be calibrated and maintained and without correction, for example, from a GPS; their measurements drift out of synchronization. INS can be implemented using 'micro electro mechanical systems' (MEMS) allowing them to be small and integrated with GPS to fill any gaps in coverage. Measurement of distances up to a kilometer can also be carried out using handheld laser rangefinders with the data recorded in a mobile GIS. Laser rangefinders can be used to measure accessible distances, for example, two sides of a triangle, so that inaccessible distances can be computed using Py thagoras' theorem.
5. The fifth technology group includes the range of wireless networking protocols, viz. WiFi, Bluetooth, and cellular networks. Mobile mapping is greatly enhanced by access to geopositioning correction ser vices (e.g., differential GPS) and by the ability to read and write data from remote servers from the field.

Case Study

The OpenStreetMap organization is one of the largest and most active participatory mapping movements in the world. OpenStreetMap use mobile mapping techniques based on GPS to gather road centerlines recorded by many different users and integrate them into a single map. The maps created are licensed using a Creative Commons 'attribution share alike' method so that the data can be used in any way, so long as derivative products also allow sharing. Other groups use GPS to create art on the Earth's surface either using 'free drawing' on open areas like beaches, or using the shapes of roads to spell out words, for example, gpsdrawing.com (**Figure 2**).

Location-Based Services

In contrast to mobile GIS and mobile mapping, which are professional tools, LBS are aimed at consumers. LBS are information services whose content depends on location:

Figure 2 A collaboratively produced OpenStreetMap of the Isle of Wight, UK. http://wiki.openstreetmap.org/index.php/Image:low 2006-06-06.png. (Collaboratively produced mapping by OpenStreetMap) From promotional images offered on site.

they require a mobile device, a geopositioning service, and access to networks. LBS take a wide variety of forms ranging from integrated systems like a personal navigation device (PND) (also called a 'sat nav' in the UK) right across to information delivered on standard mobile phones. This diversity reflects the range of user needs that exist for location information in context. Gartner group estimated that the global market revenues for LBS would grow from $485m in 2007 to $8bn by 2011, a truly startling rate of growth.

Defining LBS

LBS can be defined by technology, by information type or by business models. Each of these approaches will be reviewed in turn.

Using a technological framework, LBS can first be divided into integrated systems and second, into services running on open hardware platforms. In the 'integrated' case, hardware suitable for the end use is designed and software is written specifically for this. PNDs designed for a vehicle are the most widely used example of this approach with millions in circulation, although small network connected tags without screens for child tracking are growing fast. In the 'open' case LBS are designed to run within the computing context provided by a suitable mobile device, which may range from a Wireless Application Protocol (WAP) browser, through a Java Midlet to a proprietary LBS program running on the device. In these 'open' cases the primary challenge is integrating the wide range of elements (e.g., geopositioning, application logic, user interface, and payment systems) into a service that can operate in real time and also degrade gracefully if any element is missing, for example, geopositioning. The computing, networking, and interface demands of LBS and their users make them one of the most challenging applications to create, and only a few fully commercial services are running in 2008.

LBS can also be differentiated by information provision. Some LBS are built around a 'mobile search' paradigm where the LBS provide a mobile interface to georeferenced information resources. Google Maps for Mobile and Microsoft Virtual Earth Mobile offer this kind of LBS, which are close analogs of their desktop search services. Other LBS are tied to specific locations and are built on geospatial content management systems (geo CMS) so that the services deliver content derived from or designed in association with the locations of operation. Some LBS are designed to generate and exchange location based data (e.g., location based gaming) or to facilitate location based collaboration. Finally, some LBS are dedicated navigation devices with information about points of interest.

Novel business models are needed to support LBS as the value that can be placed on a service that is geared to location is difficult to model, especially given the extended nature of the LBS value chain (**Figure 3**). The value chain connects satellite services, location technology (e.g., GPS), chipset integration, handset design and the operator infrastructure to the network middleware, service portal, authentication module, mapping, and the LBS application. Who should get what share of the revenue is a complex business problem made more difficult by the free at the point of use model that dominates Internet services. The most successful solutions have involved LBS that integrated the whole value chain into a single proposition, for example, personal navigation systems in which everything is included in a single package. The largest such services in the world are based in the Far East, for example, Navitime pedestrian

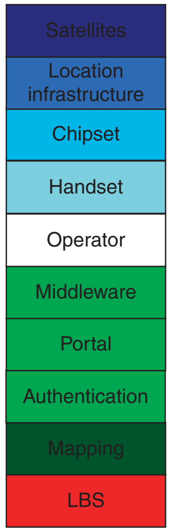

Figure 3 An LBS value chain. Diagram by J. Raper.

navigation service in Japan. However, some service providers also use a rental business model to simplify and assure the user experience.

Using LBS

The range of potential uses of LBS is very wide. Since people and goods are perpetually in motion in the world today, it is inevitable that systems to manage and leverage location on mass market mobile devices will be widely used. Early successes (using the classification in the previous section) are technically integrated, navigation focused devices delivered in all in one packages. In future, devices will have more powerful processors; there will be better device positioning, lower data costs, and a move to a web based platform. This will encourage the development of new LBS that do not need to create their own integrated hardware/software system but will be able to use standardized interfaces and services to create innovative offerings.

LBS applications already span a very wide area. The most widely used applications are personal navigation systems offering routing to POI; however, these systems are already offering traffic updates and friend finder functions. Mobile guides are also popular, often based on a rental model in a place, such as a museum, historic site, or protected area, where the use of the LBS is a powerful proposition. Location based gaming is also growing in popularity through geocaching and the type of gaming that uses real space as a dimension of a gaming application. LBS are being developed for mobility assistance and personal transportation planning, and also for vehicle to vehicle LBS using mobile *ad hoc* networks to form part of intelligent transport systems (ITS). LBS are also developing in personal security for older people or those with cognitive disabilities, for offender tracking, and for health management, for example, organ donation networks. Further diversity in applications can be expected as the infrastructure and services mature (**Figure 4**).

Figure 4 An older user of an LBS mobility support system. Photograph by J. Raper.

LBS clearly have the potential to discipline and to emancipate by virtue of their universality and power. Tracking records are likely to become a battleground of human rights and privacy over the next few years as LBS become more widely used. However, tracking is not always an intrusion into privacy, as there are occasions when mobile phone positioning has saved lives, for example, in mountain and sea rescues. LBS have been used to create art installations based on real time mobile transactions, for example, Wiki City in Rome, or by the use of the tracking record as a drawing tool. Child tracking devices have also been used to monitor the amount of exercise taken by children as part of health promotion initiatives. Location based social networking allows social connections to be defined on the basis of location or spatial behavior. Research has also been undertaken into location 'cloaking' by which users can anonymize their tracking data and obscure their location.

Case Study

LBS have also been shown to be useful as mobility support tools for blind and older people, especially to identify buildings or POI and to provide local transpor tation information. The LBS4all project developed an application called the Address Finder that used a GPS and digital compass to identify buildings when pointed at by users. The Address Finder used a screen reader to read out the descriptions and the distances to addresses, which were played through standard earphones plugged into a consumer PDA.

See also: Geodesy; GIScience and Systems; Global Positioning/GPS; GIS, Public Participation; Mapping, Topographic; Photogrammetry/Aerial Photography; Surveying.

Further Reading

Dearman, D., Hawkey, K. and Inkpen, K. (2005). Rendezvousing with location aware devices: Enhancing social coordination. *Interacting with Computers* 17, 542 566.

Dobson, J. and Fisher, P. (2003). Geoslavery. *IEEE Technology and Society* 22, 47 52.

Jiang, B. and Yao, X. (2006). Location based services and GIS in perspective. *Computers, Environment and Urban Systems* 30, 712 725.

Laurier, E. (2000). Why people say where they are during mobile phone calls. *Environment and Planning D* 19, 485 504.

Mountain, D. and Macfarlane, A. (2007). Meeting the needs of mobile users of information systems. *Journal of Information Science* 33.

Raper, J. (2008). *ArcPad: The Field User's Guide*. Redlands. CA: ESRI Press.

Raper, J. F., Gartner, G., Karimi, H. and Rizos, C. (2007). A critical evaluation of location based services and their potential. *Journal of Location Based Services* 1(1), 5 45.

Relevant Websites

http://www.esri.com
Arcpad Mobile GIS.

http://www.informaworld.com
Informaworld (Journal of Location Based Services).

http://www.placelab.org
Intel Place Lab.

http://www.lbs4all.org
LBS4all project.

http://www.mrl.nott.ac.uk
Mixed Reality Lab, Nottingham.

http://www.navitime.com
Navitime pedestrian navigation.

http://mobilesearch.nokia.com
Nokia Mobile Search.

http://www.opengeospatial.org
Open Geospatial Consortium Inc.

http://www.openmobilealliance.org
Open Mobile Alliance.

http://www.openstreetmap.org
OpenStreetMap.

GIS, Public Participation

S. Elwood, University of Washington, Seattle, WA, USA

Glossary

GIScience Basic research field that studies fundamental issues arising from geographic information, especially its representation, manipulation, and analysis in digital form.

GIServices Geographic information services, typically delivered online, such as mapping, route selection, query and retrieval of spatial data, or spatial data analysis.

Markup Languages Computer language systems that typically use tags to designate the structure, format, or presentation of enclosed text, images, or files, as in the case of HTML, a markup language for visualizing content on the World Wide Web.

Participatory Action Research A research design and method in which individuals and communities, sometimes in partnership with academic researchers, collaborate to identify major issues, concerns, and problems; design and carry out research; and apply research results to address their needs and concerns.

Spatial Data Infrastructure Local, national, or global system that enables geospatial data to be compiled, shared, and made available to users, often over the Internet. Includes data, technologies, standards, and institutions.

Introduction

Public participation geographic information system (PPGIS) emerged in response to a number of issues raised in social and political critiques of GIS in the mid 1990s. These critiques focused on the epistemologies and ontology of GIS, the significance of its intellectual and labor histories for its use and impacts, and its implications for participation and power in decision making. Other concerns included its potential for surveillance and invasions of privacy, and the social and political implications of how people and place could or could not be represented in a GIS. PPGIS coalesced around several issues raised in these 'GIS and society' debates. Critics had argued that GIS in its then current form was not able to include and represent diverse forms of spatial knowledge, tending to best support the inclusion of quantitative or cartographic forms of information. Thus, they argued, GIS was unlikely to include the knowledge, needs, and priorities of a diverse range of social groups. Additionally, critics were concerned with barriers posed by the substantial hardware, software, data, and skills requirements of GIS, noting that these limitations would impinge upon participation in GIS based decision making processes. Rather than reject GIS completely, a number of researchers and practitioners began exploring ways of reworking GIS software and data structures, as well as different ways of incorporating GIS into participatory processes, to try to address these concerns. These efforts focused particularly on a small but growing number of nontraditional GIS users such as activist groups, nongovernmental agencies, community based organizations, and participatory research collaborations, and other institutions experiencing unique needs and constraints in gaining access to and using GIS.

PPGIS has come to include a diverse set of social, political, and technological practices that seek alternative approaches to creating and using GIS in a wide variety of application contexts. These have included urban community development, community resource management, land reform movements, and environmental justice activism. This diversity of applications, as well as the continuing growth of initiatives referred to as PPGIS, make it difficult to provide an exhaustive or fixed definition of its characteristics or specific practices. PPGIS initiatives tend to involve spatial data collection and GIS use by, with, and for grassroots groups, local residents, or nongovernmental institutions, to support their research, collaborative decision making, advocacy or activism, or efforts to change local conditions. Many are characterized by a commitment to including multiple forms of spatial knowledge in a GIS and also in the decision making processes in which it is used. PPGIS projects seek to understand and mitigate differences in knowledge and power that may give some individuals and social groups greater role or influence in planning and decision making processes.

Research on PPGIS explores the intersecting technological and sociopolitical processes that shape and are produced through these kinds of GIS initiatives. Working from an early proposition that the social and political implications are strongly contingent, researchers have identified institutional, political, and other factors that shape the sustainability and effectiveness of PPGIS initiatives. Other PPGIS projects have used digital sketch boards to incorporate residents' own maps, or had participants take photos of important community sites, assets, or problems. Other research has sought to identify characteristics of PPGIS processes that are likely to foster meaningful participation by diverse individuals and social groups, and inclusion of their knowledge and priorities. PPGIS research also examines how GIS and spatial data are used in collaborative decision making processes, to understand their role at different stages of these processes, especially in contentious decision making or conflict resolution. Finally, this research also engages the technologies of GIS, developing techniques for adapting GIS software and data structures to incorporate qualitative and noncartographic representations of spatial knowledge. The following sections expand this discussion of PPGIS as research and application areas, discussing key conceptual debates within PPGIS research in its first decade, examining its practices in more detail, and exploring new technological and methodological developments affecting future PPGIS research and practice.

Conceptual Debates in PPGIS

PPGIS research has focused on understanding how spatial technologies, spatial data, and social and political processes together shape knowledge, participation, and power in PPGIS initiatives, as well as in the social situations they seek to influence. The theoretical traditions informing PPGIS research are diverse, including political and social theory, social studies of technology, organizational theory, and feminist theory and method.

Some of these conceptual discussions have been definitional in nature, seeking to delineate key elements of PPGIS such as participation, public, or access. With respect to participation, a major point of debate has been what sort of involvement or actions should be considered participation in a PPGIS initiative. Is participation defined by direct manipulation of GIS software for analysis and mapping? Or might involvement in defining project goals, contributing data, or directing others in use of the GIS be sufficient to count as participation? As PPGIS has grown and developed, there is growing agreement that participation need not be solely defined as direct use of the GIS software, and that the collaborative decision making processes of PPGIS might feature a range of different types of participation. Researchers have developed a number of frameworks for conceptualizing different types and modes of participation in PPGIS. Many of these recognize different forms of participation with reference to Arnstein's ladder of participation, a framework developed to describe different degrees of public involvement in planning and policy decisions.

Agreement on how to define the public in PPGIS has proved similarly challenging. Does the public consist of any interested person, the residents of an area directly affected by a planning or decision making process, or organizations charged with representing the views and needs of a particular group? Drawing heavily on planning theory and urban and political geography, most contributions to these debates propose that there are multiple publics who might participate, depending on issues or problems it engages, spatial knowledge being gathered, geographical areas affected, or policies requiring involvement by particular individuals or groups. Because of these shifting and multiple definitions of the public, some researchers have proposed removing the problematic term altogether, in favor of the moniker 'participatory GIS'. However, some PPGIS researchers and practitioners informed by critical development theory argue that participatory GIS is not an appropriate characterization for many initiatives undertaken under the auspices of PPGIS. Specifically, the wish to differentiate between forums that are initiated and controlled by grassroots groups, activists, and local residents, and those that are not bottom up, perhaps being initiated and controlled by government institutions as part of a collaborative planning program. The term 'participatory', they contend, should only be used to describe the former type of initiative, to avoid glossing over critical differences in autonomy and power that exist within PPGIS.

One of the central goals of PPGIS has been articulated as empowering marginalized communities and social groups by fostering greater access to GIS, spatial data, and the decision making processes in which they are used. Thus the notion of access has also received considerable attention. Early on, PPGIS researchers noted that ability to procure hardware, software, and digital spatial data for GIS use was an important part of access, but was not a sufficiently expansive conceptualization. Many grassroots groups, even when they had achieved this sort of access, still reaped few benefits from their GIS use. Some had difficulty obtaining spatial data appropriate to their unique needs. Others did not have an understanding of cartography and spatial information to be able to apply their GIS resources to their activities and problem solving efforts. Access in PPGIS is now typically conceptualized more broadly to include notions of hardware and software provision, appropriateness of existing spatial data, and conceptual or cognitive access – meant to characterize participants' ability to understand and use results.

Beyond these definitional debates about key tenets of PPGIS, researchers have sought to theorize the mechanisms and process that structure participation, spatial data and technology access, and the social, political, and material impacts of PPGIS. Researchers have focused on explaining how participation and power are negotiated in PPGIS and in GIS based decision making processes more broadly, and also on understanding how GIS based spatial knowledge and visual representations in PPGIS may be used for particular social and political goals. Most of the conceptual contributions in these arenas have been developed from inductive research conducted with individuals and institutions involved in PPGIS initiatives.

Research designs vary, including participatory action research, single and comparative case studies, experimental designs, ethnographic studies and survey research; and researchers have relied on qualitative, quantitative, and mixed methods.

Research on the negotiation of participation and power in PPGIS demonstrates that axes of social difference such as race, class, gender, age, and status or role in a community affect the knowledge that participants contribute to a PPGIS initiative. For instance, in a community resource management project, local women may have key knowledge about forest resources, but gender roles may inhibit their involvement in community meetings at which resource management decisions are made. As well, power differences along these same lines have been shown to structure the degree to which individual and social groups may be able to influence priorities, research practices, implementation strategies, or outcomes. Other research has emphasized the need to better understand how and when participants use spatial information and GIS in processes of deliberation and conflict resolution. That is, participants use spatial knowledge for different purposes at different moments in collaborative decision making – at some moments to gather information for defining a problem, or at other times to establish and defend particular positions.

Another contribution emerging from PPGIS research has been the proposition that the definition, validation, and use of spatial knowledge play a significant role in determining participation and power. Early research showed that in many planning and decision making situations, quantitative and cartographic forms of knowledge are granted great weight than qualitative or experiential accounts of spatial knowledge. These 'knowledge politics' can have several conflicting implications for the negotiation of participation and power in PPGIS. Some grassroots groups are able to represent their knowledge, needs, and priorities in these terms, gaining greater influence with key decision makers (though sometimes in direct competition with and to the detriment of other grassroots groups). But because participants may have different forms of knowledge to contribute, there is also the potential for exclusion of those whose knowledge is less easily incorporated in a GIS. Additionally, PPGIS research has explored how participation and power are negotiated through visualization of spatial knowledge. The strong appeal of GIS for many grassroots groups lies in its capacity to enable creation of visual representations of people and place. Recent research examines maps generated in PPGIS initiatives as visual texts that produce narratives about places and their needs and characteristics, to better understand how participating individuals and institutions use these narratives to advance particular political projects.

Another major strand of research has explored the contingent nature of PPGIS projects, as part of efforts to understand why these initiatives may have particular social and political impacts and to identify factors affecting their sustainability and effectiveness for grassroots groups. A multitude of case studies in the PPGIS literature have shown that these impacts are not primarily technologically determined, nor can they be attributed to singular factors such as a grassroots group's available resources. Some of these studies have documented the importance of public spatial data availability for many PPGIS initiatives, showing how this access is structured by local and national policies, legal structures, as well as cultural traditions or citizens' expectations around access to public data. For example, the United States policy and tradition suggests that government data be made readily available to citizens, usually on an 'at cost' basis, whereas in many European countries, access to these same national level data are restricted by policy or in practice through high costs. Not all PPGIS initiatives need or use public or governmental data, but many do, even if only to provide their own critique or reinterpretation of these data. Because of the unique capacities and needs of the grassroots institutions that tend to be involved in PPGIS, a number of institutional or organizational factors also play an important role. These factors include stability of organization's staffing and funding, as well as the strength of their collaborative relationships with other organizations and individuals who might be able to provide needed resources such as data and training. Grassroots groups that can creatively share resources and ideas for data access and GIS application are generally much more able to sustain their GIS activities and use them to advance their social and political goals. Institutional knowledge is another important determining factor, involving not just knowledge about GIS and spatial data, but also about ways of acquiring data, sources of technical assistance, or successful strategies for using spatial data and maps to influence political processes.

In sum, conceptual debates and contributions of PPGIS have focused upon defining its central tenets of public participation and access; theorizing its sociopolitical and organizational/institutional processes; and understanding its unique knowledge politics and visualization practices. These contributions have informed efforts to facilitate sustainable and inclusive PPGIS initiatives. As well, these contributions from PPGIS research are significant for GIS and society research on the social, political, and economic implications of GIS and digital spatial data development; urban and political geography research on collaborative governance and participatory democratic practice; and GIScience research on qualitative data representation, visualization, and spatial data infrastructures (SDIs).

As PPGIS research has developed, so too have internal and external critiques of its conceptual and applied contributions. Some researchers have expressed concern about what they perceive as an overemphasis on facili tating grassroots GIS access, at the expense of continued efforts to better understand its impacts. This critique bears careful consideration on its own merits but also because of the extremely rapid growth of GIS initiatives that are self identified by sponsors or participants as PPGIS. These initiatives vary tremendously with respect to the nature or degree of participation that occurs; with some purportedly PPGIS initiatives being initiated and directed in a top down manner, with quite limited con tributions from supposed participants. While the growth of participatory spatial technology initiatives around the world is exciting, it is essential to continue critically examining their impacts and participatory processes. Other critics have charged that PPGIS research remains theory poor, having yet to formulate a largely case based body of research into theoretical propositions that could serve as a conceptual foundation guiding and informing PPGIS research. Some scholars have argued that future research in PPGIS needs to move toward a PPGIScience, by which they mean a more systematic approach to in vestigating fundamental questions about the use and impacts of GIS in grassroots contexts or public partici pation forums.

PPGIS in Practice

There exists a wide range of technological, methodo logical, and political practices that are considered part of PPGIS. Some focus specifically on software or spatial data, while others have addressed the processes in which GIS is applied. PPGISs have involved intervention in both the technologies of GIS as well as the participatory processes in which it is used. These practices include changes to GIS at the level of software or data structures, and efforts to alter collaborative or participatory pro cesses in which GIS is used.

Software and data interventions in PPGIS have sought ways of diversifying the range of spatial knowledge, perceptions, and experiences that may be included and represented in a GIS. Multimedia approaches to GIS, for instance, are oriented toward including sounds, narra tives, pictures, animations, or sketches in a GIS. Using this approach, a GIS might include oral history narratives describing a place, its important characteristics, or the needs of people living in it. A planner's land use map of an urban neighborhood might be supplemented with children's sketches of their perceptions and experiences in that neighborhood. Some multimedia GIS use the Internet to hyperlink these alternative forms of spatial data to specific locations on a map. Other software modifications enable three dimensional representations of space and use this third dimension to include non Cartesian spatial knowledge such as time, emotion, or perception of experiences in a place. While these modifications focus on visualization and representation of diverse spatial knowledge, other software level inter ventions have focused on incorporating qualitative or noncartographic data directly into GIS databases. Tech niques for extending existing structures for metadata to include more information about the social, political, and institutional contexts in which it was created (and which have bearing on how it can and should be used) are one such approach. Using markup languages such as UML or XML to represent attribute data in a GIS has been considered as a means to diversify the types of 'attributes' of spatial locations that may be incorporated.

Alongside these data and software developments in PPGIS, researchers and practitioners have also de veloped participatory approaches that offer alternative ways of using GIS in research, planning, problem solving, and decision making. Participatory sketch mapping and other techniques for collaborative spatial data develop ment are commonly used in PPGIS initiatives. Some projects use focus groups for planning and carrying out a GIS application and implementation of its results, while others incorporate GIS into a broader participatory action research framework. Other projects incorporate additional audio or video technologies to try to create new ways for potential participants to become involved and to contribute their knowledge and concerns. Trying to overcome time or mobility barriers that might exclude participants, some PPGIS initiatives have developed web based forums that allow participants to contribute information, respond to and revise data and maps being developed, or weigh in on decisions or actions being considered. Electronic polling systems and digital sketch boards have been used in some cases to try to facilitate involvement of large groups of people in a PPGIS pro cess, or to more readily incorporate participants' con tributions (in the form of sketch maps or markup of an existing map) into a GIS. Because grassroots groups, activist organizations, and small nongovernmental or ganizations often have difficulty assembling the necessary hardware, software, and specialized skills for GIS, PPGIS projects tend to rely on a range of collaborative strategies for initiating and sustaining GIS use. Some PPGIS pro jects involve university–community partnerships, work ing with university students and faculty. Some rely upon intermediary organizations (usually nonprofit groups) that provide technical expertise for grassroots groups. Others collaborate closely with other grassroots GIS users to share data, technology, and expertise.

The practices that characterize PPGIS are closely related to other alternative and participatory approaches to cartography or information technology use, includ ing countermapping, community informatics, and com munity mapping. Countermapping, a term emerging from social and political geography and critical cartog raphy, involves spatial data collection and mapping to support activist and resistance efforts. Some counter mapping efforts use GIS and others do not. Counter mapping typically creates alternative or reinterpreted representations of people and place, to try to disrupt hegemonic representations that silence or oppress marginalized groups, and has been used especially effectively in indigenous groups' political and social movements. Community informatics involves the use of information and communication technologies in community development efforts, often initiated by and for the use of local level community development organizations and their residents. While most community informatics initiatives involve the Internet or wireless technologies, many also involve the collection and dissemination of spatial data with these technologies, and some incorporate maps or GIS. For example, one such community informatics initiative, Neighborhood Knowledge California, disseminates demographic, socioeconomic, and environmental spatial data and maps for neighborhood activism and revitalization to its members through a web portal (http://nkca.ucla.edu).

Finally, community mapping involves spatial data collection and mapping efforts by local residents of place, often to research concerns and problems, or gather information about local resources and assets. Increasingly, the term community mapping is used to identify projects that are carried out in collaboration with primary and secondary school students, as part of community service projects or active learning curricula. Countermapping, community informatics, and community mapping, even when they do not specifically employ GIS software, share with PPGIS an emphasis on participation and self representation in spatial data creation and use and on including diverse spatial knowledge for self empowerment by marginalized individuals, communities, and social groups.

New Developments in PPGIS and Future Directions

The previous two sections on PPGIS and its conceptual debates and contributions reflect upon a research and application area that is constantly changing. Ongoing technological and methodological developments in GIScience are changing PPGIS, opening a number of new questions about how these developments will affect the negotiation of participation, power, and spatial knowledge. In recent years, some PPGIS initiatives have begun to incorporate and rely on Internet GIS, mobile and handheld technologies, and freeware mapping and analysis softwares, many written using open source programming languages. Here, the term Internet GIS is used to refer to online GIServices, GIS based Internet mapping systems, web portals for acquiring and sharing spatial data, and a range of other distributed GIS practices that use the Internet as a platform. In particular, the emerging new forms of user customized web services known as Web 2.0 are an important development and new resource for PPGIS. These developments include open source applications that allow users to add mapping functionality to their own web services, or create interfaces that allow their website users to add their own new spatial data or narrative information about locations on a map.

These developments open new possibilities for PPGIS, and some help overcome some of the unique barriers faced by nontraditional or grassroots GIS users. Web based geoportals have made existing spatial data far easier to obtain, thought with the caveat that many user interfaces are not designed for users with limited computer or GIS experience. Internet mapping systems make it possible for individuals and grassroots groups to produce their own maps without having their own GIS software. Though a common criticism of these sites is the limited ability for users to tailor data and maps to meet their own needs, the interactivity of Internet map systems continues to expand. Some sites now enable users to sketch objects on the screen and have these objects included as a temporary layer in their map display; or to carry out more powerful spatial queries and download resulting data sets. These expanded functionalities may help PPGIS participants expand their access to existing spatial data or gain access to GIS capabilities without needing to have GIS software on a local computer. A few well resourced PPGIS initiatives use the Internet to publish their own data and maps, or host their own Internet mapping sites.

The rising functionality and diminishing costs for mobile technologies such as Global Positioning System (GPS) units or handheld devices for spatial data entry are altering local data collection in many PPGIS initiatives. Local data collection by participants has long been a hallmark of PPGIS, but the growing accessibility of wireless and handheld devices can expand local participants' ability to create their own digital spatial data, or edit these data in the field. On the software side, a growing number of PPGIS projects are beginning to use free or low cost mapping and analysis software, usually downloaded from the Internet. The obvious benefit of freeware in PPGIS is the ability of under resourced communities and organizations to gain access to GIS without purchasing expensive software. But also, for freeware that uses open source code, benefits may include adapting software to particular needs of a PPGIS initiative, its participants, and its constituent community.

These new technological developments in GIS are changing available ways that PPGIS initiatives can collect, obtain, visualize, and disseminate spatial knowledge. But it is important that future research in PPGIS not overstate their benefits or overlook ways they introduce new barriers or challenges for grassroots GIS users. For all the promises of wireless and mobile technologies in GIS, their relative costs continue to be high and wireless service limited to a small number of places around the world. Most Internet GIS sites and online GIServices are quite difficult to use without high speed Internet access, and here too, the global extent of such services are quite limited. Many online GIS resources continue to be de signed in ways that are difficult for novice users to navigate. Finally, relatively little is known about the na ture or impacts of internet mediated participation in PPGIS initiatives, so this must be included as part of future PPGIS research.

Another new development in GIScience that holds promise for future PPGIS research and practice is an emerging body of work on qualitative GIS. Qualitative GIS focuses on (but is not limited to) developing ways of incorporating qualitative data or carrying out qualitative analysis of spatial knowledge in a GIS. Software level interventions in qualitative GIS are particularly in triguing with respect to their future potential for PPGIS. Some of these techniques modify existing spatial data structures so that qualitative spatial knowledge such as sketch maps or images can be directly incorporated into a GIS database. Other scholars are working on ways of adapting GIS software to facilitate qualitative analysis of these data. As with Internet GIS, it is important not to overstate the immediate benefit of such developments to PPGIS initiatives and to remain attentive to ways in which they may worsen existing barriers to participation and access or introduce new ones. Nonetheless, qualita tive GIS is likely to be a fruitful source of new ap proaches that directly speak to one of the central goals of PPGIS: diversifying the forms of spatial knowledge that may be included, represented, and analyzed in a GIS.

Over the past decade, PPGIS has developed into a diverse set of technological, social, and political practices, as well as a sub area of GIScience research. PPGIS has significantly influenced GIS through its software and data structure interventions, as well as through efforts to facilitate GIS based decision making processes that are more inclusive and robustly collaborative. As well, PPGIS has contributed to a reconceptualization of GIS as a sociopolitical process, pushing beyond more tech nology focused definitions. There remains a great deal of work to be done. PPGIS practitioners and the institutions that support them will continue to face challenges posed by unequal access to hardware, software, and digital spatial data, as well as by ever emerging new technolo gies in GIS. On the conceptual side, PPGIS research needs to continue efforts to connect its conceptual con tributions to GIScience research more broadly. This diverse interdisciplinary research area will continue to face the challenge of developing theoretical frameworks that are sufficiently nuanced in their contributions to GIScience, social, and political theory.

See also: Counter-Mapping; Critical Cartography; Critical GIS; Digital Divide; Feminism, Maps and GIS; GIScience and Systems; GIS and Society; Indigenous Knowledges; Indigenous Mapping; Internet/Web Mapping; Maps and Protest; Mapping, Non-Western; Mental Maps; Nongovernmental Organizations; Participation; Participatory Action Research; Qualitative Geographic Information Systems.

Further Reading

Craig, W., Harris, T. and Weiner, D. (eds.) (2002). *Community Participation and Geographic Information Systems*. London: Taylor and Francis.
Elwood, S. (2006). Negotiating knowledge production: The everyday inclusions, exclusions, and contradictions of participatory GIS research. *The Professional Geographer* 58(2), 197 208.
Ghose, R. (2007). Politics of scale and networks of association in public participation GIS. *Environment and Planning A* 39(8), 1961 1980.
Knigge, L. and Kwan, M. (eds.) (2006). *Special Issue: Qualitative GIS. Environment and Planning A* 38.
Kwan, M. (ed.) (2002). *Special Issue: Feminist Geography and GIS. Gender, Place and Culture* 9, 3.
Kyem, P. (2004). Of intractable conflicts and participatory GIS applications: The search for consensus amidst competing claims and institutional demands. *Annals of the Association of American Geographers* 94(1), 37 57.
Obermeyer, N. (ed.) (1998). *Special Issue: The Evolution of Public Participation GIS. Cartography and Geographic Information Systems* 25, 2.
Onsrud, H. and Craglia, M. (eds.) (2003). *Special Issue: Access and Participatory Approaches in Using Geographic Information. The URISA Journal* 15, APAI & APAII.
Pickles, J. (1995). *Ground Truth: The Social Implications of Geographic Information Systems*. New York: Guilford.
Ramasubramanian, L. (1999). GIS implementation in developing countries: Learning from organizational theory and reflective practice. *Transactions in GIS* 3(4), 359 380.
Sheppard, E. and Poiker, T. (eds.) (1995). *Special Issue: GIS and Society. Cartography and Geographic Information Systems* 22, 1.
Sieber, R. (2000). GIS implementation in the grassroots. *URISA Journal* 12(1), 15 51.
Sieber, R. (ed.) (2001). *Special Issue: A PPGIScience? Cartographica* 38(3/4), 1 4.
Talen, E. (2000). Bottom up GIS: A new tool for individual and group expression in participatory planning. *APA Journal* 66(3), 279 294.
Wong, S. and Chua, Y. (2004). Data intermediation and beyond: Issues for web based PPGIS. *Cartographica* 38(3/4), 63 79.

Relevant Websites

http://www.iapad.org
Integrated approaches to participatory development.
http://www.urisa.org
The Urban Regional Information Systems Association.

GIScience and Systems

M. F. Goodchild, University of California, Santa Barbara, CA, USA

Glossary

Attribute Table A table with one row allocated to each feature and one column allocated to each defining characteristic.
Buffer Dilation of point, line, or area features by a defined distance.
Datum A mathematical reference model for global coordinates.
Geodesy The science of precise Earth measurement.
Geoportal A single point of entry to a distributed collection of data sets.
Hybrid Model A data model in which feature attributes and topological relationships are stored in a relational database, while point coordinates are stored elsewhere.
Join Combination of attribute tables based on common keys.
Map Algebra A language for manipulation of raster layers using symbolic representation.
Metadata Description of the properties, format, quality, and lineage of a data set.
Object Orientation An approach to data modeling that recognizes, among other principles, that classes of features can inherit properties from more general classes.
Overlay Calculation of a new topologically rich layer from two or more input layers.
Polygon An area represented as a sequence of points connected by straight lines.
Polyline A line represented as a sequence of points connected by straight lines.
Projection A mathematical transformation of a curved surface to a plane.
Raster Representation of space as a collection of equal, rectangular areas ordered in a defined sequence.
Relational Database A data model in which all information is represented in tables linked by common keys.
Topology In mathematics, properties that are invariant under distortion of the containing space; in GIS, relationships between features such as adjacency and connectivity.
Vector Representation of space by identification of features contained in the space, and use of coordinates to specify locations of features.

Introduction

Geographic information can be defined as information linking places on or near the Earth's surface to the properties present at those places. In essence, any item of geographic information must possess both a specification of location, in the form of a place name, geographic coordinates, street address, or some other suitable format; and a specification of such properties as atmospheric temperature, average income of residents, or elevation above mean sea level. Thus both an informal statement such as 'it is cold today in Santa Barbara' and a more formal one such as 'at 12 noon local time on January 25 at latitude 34 degrees 26 minutes 41 seconds North, longitude 119 degrees 48 minutes 26 seconds West the air temperature in a Stevenson screen was 12.6 Celsius' are examples of statements meeting this definition.

Traditionally humans have handled such statements in a variety of ways: for example, as elements of the contents of maps, as observations recorded on paper, or as statements printed in text or communicated through speech. Over the decades many conventions have been adopted to codify practice, as for example in the standardization of longitude following the International Meridian Conference of 1884, or the adoption of the World Geodetic System of 1984. Increasingly, however, and in line with trends in society generally, geographic information has been converted to digital form, and has become just one of a myriad of different forms of data that pass along the Internet and are stored on the hard drives and CDs of contemporary digital technology.

The term geographic information system (GIS) was coined in the mid 1960s to describe a computer system for handling geographic information in digital form. Motivations for the adoption of GIS take many forms. One of the earliest was the need to analyze, to produce summaries of the amounts of area devoted to different uses on the Earth's surface. Digital processing appeared to offer enormous advantages in the form of precise computation, replacing manual processes that were frequently tedious, labor intensive, and inaccurate. Another motivation arose in statistical agencies such as the Bureau of the Census, where digital processing had the potential to automate various operations, from the management of door to door interviews to the aggregation of statistics for multiple reporting zones. Still another arose in map producing agencies, since digital processing offered the possibility of easy editing, rapid calculation of map projection equations, and automated plotting. By the late 1970s these applications had converged, as it became clear that the commonalities were sufficient to support a distinct type of computer application. A substantial software industry emerged, major efforts were made to standardize terms and methods, and education programs evolved in many universities and colleges.

The term geographic information science (GIScience) was coined in 1992. As with many scientific tools, including the telescope and the microscope, the development and adoption of GIS provoked a series of fundamental questions: what theory of the geographic world did GIS implement; could the traditional processes of manual cartography be formalized in computational systems; how could one measure the uncertainties inherent in any representation of the world? GIScience is variously defined as the science behind the systems; the set of fundamental questions raised by GIS; the research field that will define the next generation of GIS; the body of knowledge that GIS implements; and the use of GIS as a tool for scientific research. As a science of geographic information it represents a subset of information science; while its relationship to the discipline of geography, and to other disciplines that deal with the surface and near surface of the Earth, is a subject of continuing debate.

Applications of GIS

A vast number of applications have been identified since GIS was first defined. Virtually all human activities deal in some sense with the surface of the Earth, and 'everything that happens, happens somewhere'. Yet rates of adoption vary dramatically, and are determined by a number of factors.

The first set of adoptions occurred in activities that have relied strongly on the use of maps. Many of the first sales of GIS software occurred in the early 1980s to agencies concerned with the management of natural resources, notably the forest industry and its regulatory agencies. By the late 1980s virtually all such agencies had adopted GIS, making investments that were massive even by today's standards. Mapmaking agencies have already been mentioned, and the advantages of computerized map production are so dramatic that virtually all such agencies had adopted some form of GIS by the late 1980s.

Another strong group of early adopters were corporations and agencies with large quantities of geographically distributed assets, who required GIS to maintain inventories, manage assets, and schedule maintenance. The utility companies and agencies – water, gas, electric, telecom – found the case for GIS compelling, and by the 1990s had become one of the mainstays of the GIS industry.

It has been argued that the academic GIS community has consciously or subconsciously ignored its military and intelligence roots. Many of the original technical developments were made in this domain, in remote sensing, missile guidance, positioning, and many other areas. Today digital geographic information technologies are indispensable to military operations and intelligence gathering.

Among academic disciplines, geography was understandably an early adopter of GIS. By the early 1990s any discipline concerned with physical or biological dimensions of the Earth's surface had adopted GIS as a standard tool, but within the social sciences adoption was rather slower, due in part to the debates that still rage in these disciplines over the appropriate roles of scientific method, data, and experimentation. Adoption has been strong in the more empirically based disciplines, such as criminology and archeology, and rather slower in the more theoretically and conceptually oriented disciplines of economics or sociology.

The twenty first century has brought a new awareness of the vulnerability of modern society, and several major disasters have shown the importance of geographic information technologies in all aspects of planning, response, and recovery. GIS played a significant role in the recovery effort following the attacks of 11 September 2001. After Hurricane Katrina it was painfully evident that anyone with a broadband connection anywhere in the world could view the events as they unfolded and at high resolution – whereas those close to the disaster had no access to the necessary power, computers, and Internet connections.

Concepts of GIS

The Layer

The layer is perhaps the most conspicuous and iconic concept of GIS, having its roots very early in GIS history and capturing one of its strongest motivations. It proposes that the geographic world can be represented as a series of thematic layers, each carrying information relevant to one particular thematic domain. To mapping agencies, each layer might correspond to the information portrayed in one color of ink on a topographic map: contours of topographic elevation in brown, urban areas in pink, or wooded areas in green. To Ian McHarg, a landscape architect seeking to develop a new model for a department at the University of Pennsylvania in the late 1950s, each layer represented the perspective of one discipline; the geologic layer captured the factors believed by geologists to be important in planning; the ecological layer captured factors originating in concern for biological conservation; and various social layers captured factors related to the economy and human populations.

By stacking the layers, one could combine factors in various ways. A power line, for example, might require one distinct set of weights to be applied to each dis cipline's layer, while a new shopping center might require a different set of weights.

This notion of partitioning geographic variation into a number of thematic layers underlies much traditional practice in cartography, where each layer might be the subject of a different printed map, or might be depicted on a single map in a distinctive color. From another perspective, it captures the ability of GIS to relate seemingly unrelated information through common geo graphic location. Only in a GIS, it is argued, can one combine detailed information on the ethnicity of a city's neighborhoods with detailed information on patterns of atmospheric pollution, to address the central question of environmental equity: do minorities bear a dis proportionate impact from industrial pollution? Only in a GIS can one combine a map showing the distribution of schools with one showing the distribution of liquor out lets. While all of these tasks could in principle be completed manually, the effort needed to redraft onto transparent media and to correct for differences of scale and map projection is often prohibitive.

Raster and Vector

Digitization is a form of coding, and in the binary rep resentations of computers it implements a coding scheme capable of reducing the phenomena of a domain to a linear sequence of two values: 0 and 1, or 'on' and 'off'. Most domains have by now developed standard coding schemes to do this with their particular phenomena. In music, for example, the dominant standards are MP3 and MIDI. Digital representation of geographic information similarly requires a suite of coding schemes, and in es sence much of the development effort in GIS over the past few decades can be summed up as 'finding the equivalent of MP3 for maps'.

Two classes of coding schemes emerged very quickly, based in part on discoveries already made in computer graphics. Raster schemes represent the map as a regular array of square or rectangular picture elements of fixed size, and record the contents of each element as a simple value in some binary code, proceeding in a standard sequence that is often row by row from the top left (northwest corner). Suppose, for example, that the first four cells of the top row contain the values 5, 5, 4, 4, denoting the presence of land cover categories 5 and 4. Then a simple coding scheme would allocate 4 bits to each cell, and record the binary equivalent of each cell's integer value as the sequence 0101010101000100. All detail smaller than the cell size is necessarily lost in this ap proach, unless special rules are adopted to recognize the presence of very small features in each cell's coded value.

Vector schemes, on the other hand, first identify each feature on the map. Features might include patches of homogeneous land cover, roads and rivers, contours, or point like phenomena such as oil wells. Each type of feature is first categorized as point like, line like, or area like, and the location of each feature is then recorded in a suitable coordinate system such as latitude/longitude. Points will be given a simple pair of coordinates; lines will be represented as sequences of points connected by straight line segments (termed polylines), and areas will be represented as similar sequences connected in loops (polygons).

In the early days of GIS development each system was predominantly one or the other, and raster GIS and vector GIS competed for a while. Before long, however, every major GIS package developed tools for handling both, and today one expects to be able to input, store, manage, and analyze data in either form, transforming from one to the other as necessary. **Figure 1** shows an example in which both raster and vector data are displayed.

There are many dimensions to the choice between raster and vector, some of them more subtle than others. Simple epithets such as 'raster is vaster but vector is correcter' are superficially true but largely unhelpful on closer examination. In reality the following issues are important in deciding which to use in a particular situation:

- *Volume.* Representing an area such as a county in vector requires the coding of a number of coordinates, and this can be done to a high precision with the use of many numerical digits. To achieve the same pre cision in raster would require the identification of each cell lying inside the area, and matching the precision of vector coordinates could require a vast number of cells. However many elegant methods of data compression can be applied to raster data. Moreover allocating many numerical digits to a co ordinate will increase precision but will not neces sarily improve accuracy, and in practice the positional accuracy of vector coordinates is often disappointingly low.
- *Data source.* Imagery collected from Earth observing satellites is an increasingly important source of data for GIS. Such data invariably arrive in raster form, with a cell size determined by the sensor's designer. Raster formats also dominate data on topographic elevation.
- *Processing.* Although many operations can be carried out on either raster or vector representations, some are either faster or only possible on one. For example, the calculation of a viewshed, or the area visible from a point on a topographic surface, is exclusively raster, while routing tasks such as determining the shortest

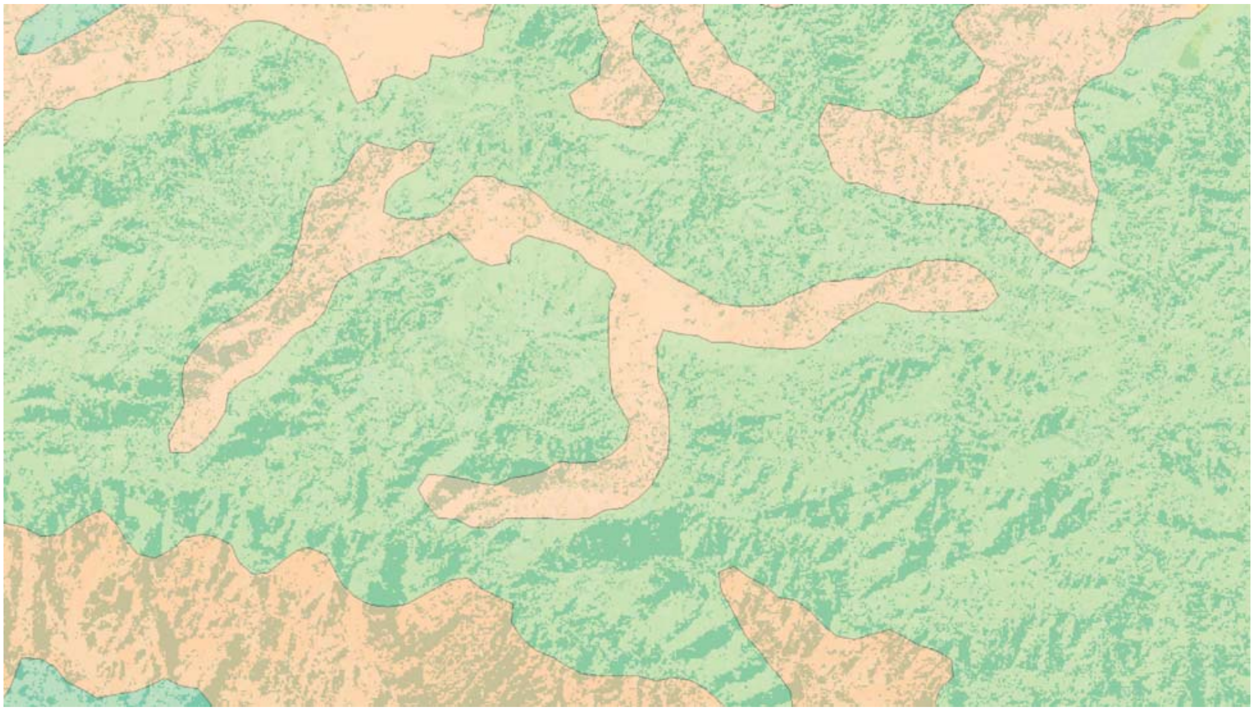

Figure 1 A GIS display of land cover in an area north of Santa Barbara, California. The map covers approximately 21 km from left to right. The California vegetation map of 1977 provides the coarse vector polygons (note the straight edges) superimposed on the raster National Land Cover Dataset with a spatial resolution of approximately 30 m. The polygons are colored by dominant species, while the raster cells are colored according to a land-cover classification.

path through a network are exclusively vector. Al though most software packages are able to handle both raster and vector, some are still primarily identified with one or the other.

- *Decision context.* GIS analysis is often used to support decisions of a spatial nature, such as the development of land use plans. Implementation of any plan is likely to involve existing spatial units of some kind, such as land parcels or agricultural fields. It makes sense, then, to tie the choice of raster or vector to the geometry of the decision units – for example, a plan for parcels of land would likely be developed using vector repre sentations in which each parcel appears as a polygon.

Entries in attribute tables may take many forms, and the number of forms has increased as the technology for handling tables has evolved. Entries may be numeric, ei ther whole numbers (integers) or decimal numbers (reals), or alphanumeric, including names, text descriptions, or dates. They may also include images, such as photographs of a site, or sounds, such as interview responses. Finally, entries may be keys pointing to entries in other tables – for example, a point feature denoting the home location of a patient might point to a record in another table describing the attributes of the hospital at which the patient was treated. Such keys are a central concept of the relational model, which was introduced to GIS in the 1970s and dominated database thinking until recently.

Attribute Table

In the vector world, the characteristics of points, lines, and areas are termed attributes, and represented in the form of attribute tables (**Figure 2**). Each table stores the attributes of one class of features, all of which have the same topological dimension (all points, all lines, or all areas), and all of which have the same set of defined attributes. For example, a class of land parcels might have numerous associated attributes: owner, area, assessed value, identification number, etc. Each feature occupies one row of the table, and each attribute occupies one column.

Topology

In mathematics, a property is said to be topological if it is invariant under stretching and distortion of the con taining space. If a loop is drawn on a rubber sheet, for example, it is impossible to break open the loop by stretching the rubber, and thus the distinction between area, line, and point is a topological distinction. Topo logical properties are important in GIS, since they include such useful spatial properties as adjacency and connectivity, properties that are important in many forms of analysis. For example, a system to plan optimum routes through a road network must include

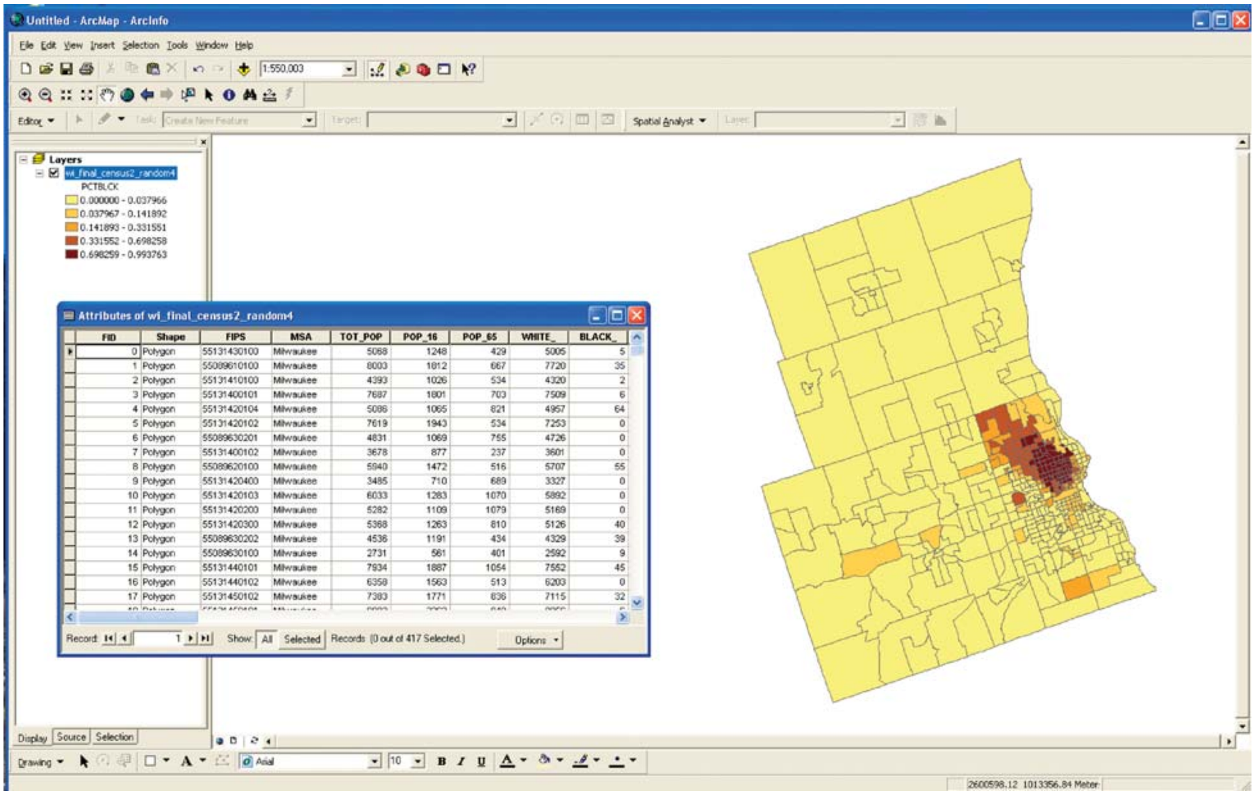

Figure 2 A screen shot of ESRI's ArcMap showing census data by tract for the city of Milwaukee. The screen simultaneously displays part of the tract attribute table and a map in which each tract is classified by its percent black population.

representation of the connections between network links, and important properties of such connections including turn restrictions or grade separations at intersections.

The need to represent topological properties emerged very early in the history of GIS, and by the 1970s was regarded as essential to success, at least in vector systems. The vegetation polygons of **Figure 1** illustrates a type of map – often termed an area class map – that is very common in land management applications. If each area on the map was digitized separately as a polygon, then each internal boundary would be represented twice, and it might be difficult to ensure that the two versions were identical. Moreover, it would be possible to edit each polygon in dependent of its neighbors, possibly producing gaps and overlaps that in principle cannot occur in such maps. These problems can be overcome by the simple expedient of treating the common boundary between two areas as the basic record. In the Canada Geographic Information Sys tem (CGIS), widely regarded as the first GIS, area class maps were stored in a linear sequence on magnetic tape, each common boundary being recorded as a coded rep resentation of (1) the attributes of the area on its left, (2) the attributes of the area on its right, (3) the number of points needed to capture the boundary's geometry as a polyline, and (4) a sequence of coordinate pairs. CGIS had no table of polygon attributes, but obtained all of its results, including measurement of polygon areas, by processing this linear sequence of boundaries.

By 1980, a consensus had emerged among GIS de velopers on the representation of area class maps in topologically rich form. Each internal boundary between two areas would be termed an arc, and its attributes would include pointers to the areas on its left and right, and to the boundary network junctions or nodes at its start and end. Each area would have an attribute table defining its properties, and perhaps pointers to the arcs forming its boundary. The points defining the polyline representation of each arc would be stored separately, and because the number of points per arc was variable the file containing these points would not have the normal tabular structure – for this reason this approach was termed hybrid. Simple extensions to this basic model allowed for the storage of street networks, administrative boundary maps, and many other forms of geographic information, and it survives today in the form of Environmental Systems Research Institute's (ESRI's) coverage model.

Object Orientation

By the 1990s, dramatic advances in computing power, storage capacity, and software had made many of these arguments for topology superfluous. Moreover the concept of the common boundary as a fundamental component of geographic information seemed to conflict with the intuitive ways in which humans view the world and its representations in maps. Object orientation emerged in the late 1990s as a new, more powerful and at the same time simpler way of representing the world that treated topological relationships as optional extras rather than as essential properties.

ESRI's shapefile model provides an appropriate example. In this approach, introduced in the late 1980s, each feature is stored as a point, polyline, or polygon. Where boundaries are shared, they occur in the database twice, and powerful routines exist to compare and correct differences between them. Similarly topological properties such as connectivity can be computed by analyzing the geometry of features, and automatically ignoring small gaps or overshoots when these fall below established distance thresholds. Features can be linked during edit to ensure that moving one feature also moves any other features tied geometrically to it.

The set of relationships allowed in object oriented databases is much richer than in the earlier relational approach. Links between features, such as between a patient and a hospital, are termed associations. The concepts of aggregation and composition allow complex features to be formed as collections of other, simpler features, matching a common property of geographic information. For example, an airport, which might be depicted at a coarse scale as a point, might be linked to its constituent runway, hangar, terminal, and other features, each of which might be depicted in fine scale visualizations. A state might be modeled as an aggregation of counties, and a county as an aggregation of cities. None of this was possible in the earlier approach.

Finally, object orientation adds the concept of inheritance. Many classes of features can be regarded as specialized versions of other, more general classes. For example, a county is a specialized form of polygon in GIS representation; a bicycle path is a specialized form of transportation link, which is itself a specialized form of polyline. Specialized classes inherit all of the properties of more general classes, and add new properties of their own. For example, all polylines have the properties of length and number of defining points; all transportation links have these properties, plus such other properties as number of lanes, travel speed, and degree of congestion; and bicycle paths have these properties plus others more relevant to cyclists, such as whether the path is on or off road.

The adoption of object oriented concepts is a major step in the evolution of GIS. While previous approaches were intimately tied to the map as the source of geographic information, object orientation has been applied to the successful representation of phenomena that have never been associated with maps, including events, transactions, flows, and interactions. In this sense, then, GIS has been able to move beyond the map metaphor in recent years. Nevertheless, the metaphor remains a powerful way of describing GIS, as 'a computer containing maps'. But today one might also describe a GIS without using the map metaphor, as 'a computer containing descriptions of features on the surface of the Earth, in which the geographic location of each feature is identified as part of its record'.

Figure 3 shows the UNETRANS model, an application of object oriented principles to the representation of features relevant to transportation applications of GIS. It supports representation of the basic transportation network; of the physical assets associated with the network, such as traffic lights and signage; of features important to navigation on the network, such as connectivity and turn restrictions; of vehicles on the network; and of the routes and schedules of public transit.

Overlays and Joins

The set of possible topological relationships between features is very rich, and while topology may no longer be as important in the representation of geographic data, it plays a very important role in the kinds of analysis needed to support GIS operations. In network analysis, for example, the connections that exist between network links provide the essential basis for planning routes, determining least cost and least time paths, and many more sophisticated forms of analysis that fall under the broad headings of logistics and location analysis.

Some of the most notable advances in GIS research have concerned the enumeration of possible relationships between features, and the identification of these relationships is a basic function of all GIS software. Suppose that two layers exist for the same area, one depicting current land use as a collection of nonoverlapping, space exhausting area features, each showing an area of uniform land use; and a second layer showing a similar depiction of land capability for agriculture. Suppose further that the analyst wishes to determine the area of land that is both (1) currently not used for agriculture and (2) of high potential for agriculture. Intuitively one would think of overlaying the two maps, assuming that one was transparent, and measuring the appropriate areas of overlap (**Figure 4**). In vector GIS terms, the task is known as the polygon overlay problem, and results in a new layer in which each polygon is homogeneous with respect to its characteristics on both of the input layers. Operationally this requires the calculation of every intersection between the two sets of polygon boundaries, plus the assembly of new and generally shorter boundary segments into new polygons. In practice the task is often plagued by the existence of large numbers of slivers, which result

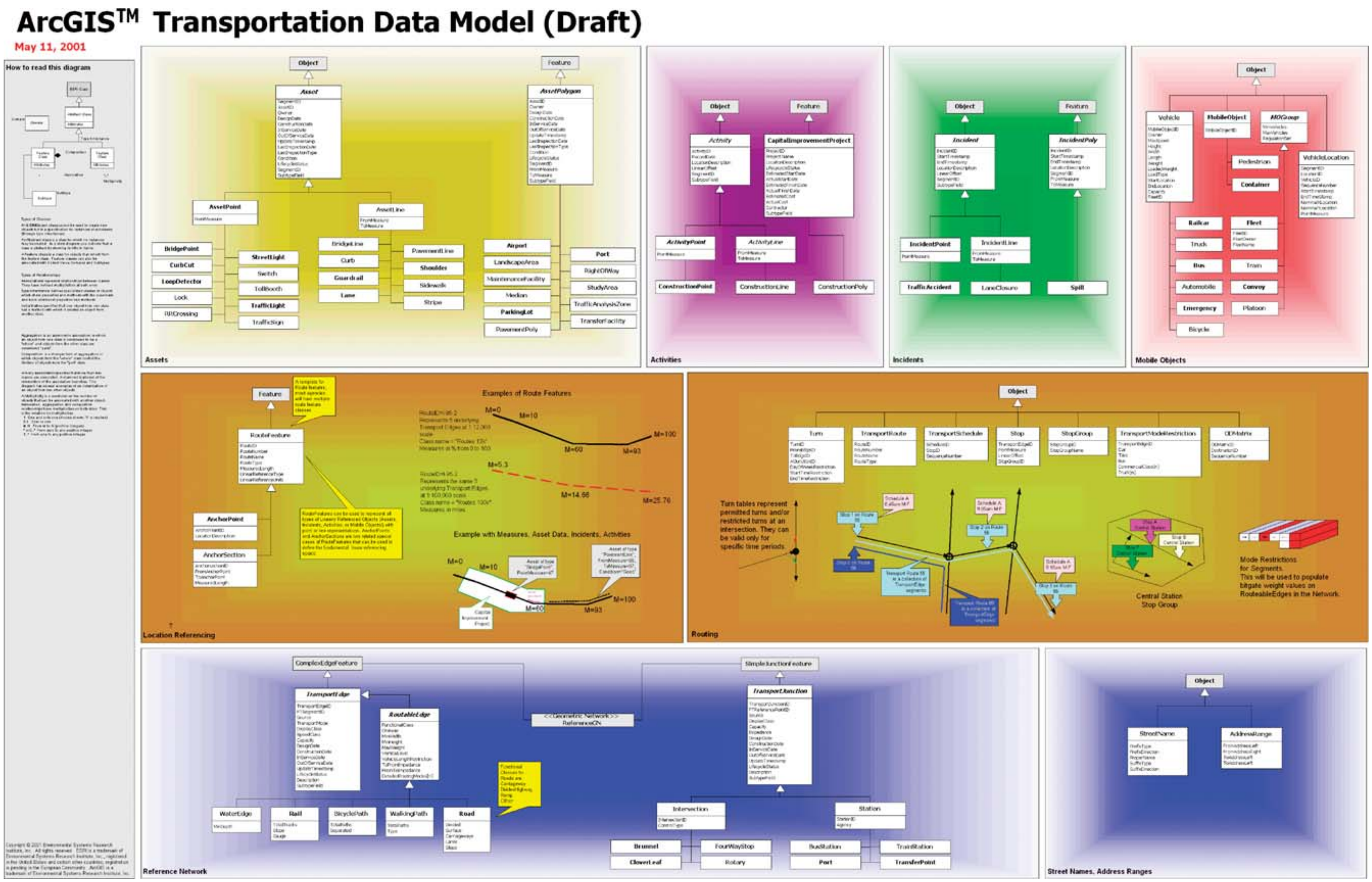

Figure 3 The UNETRANS model, an object-oriented data model for transportation applications of GIS. Each box represents one class of features, and the links between the boxes represent various types of relationship.

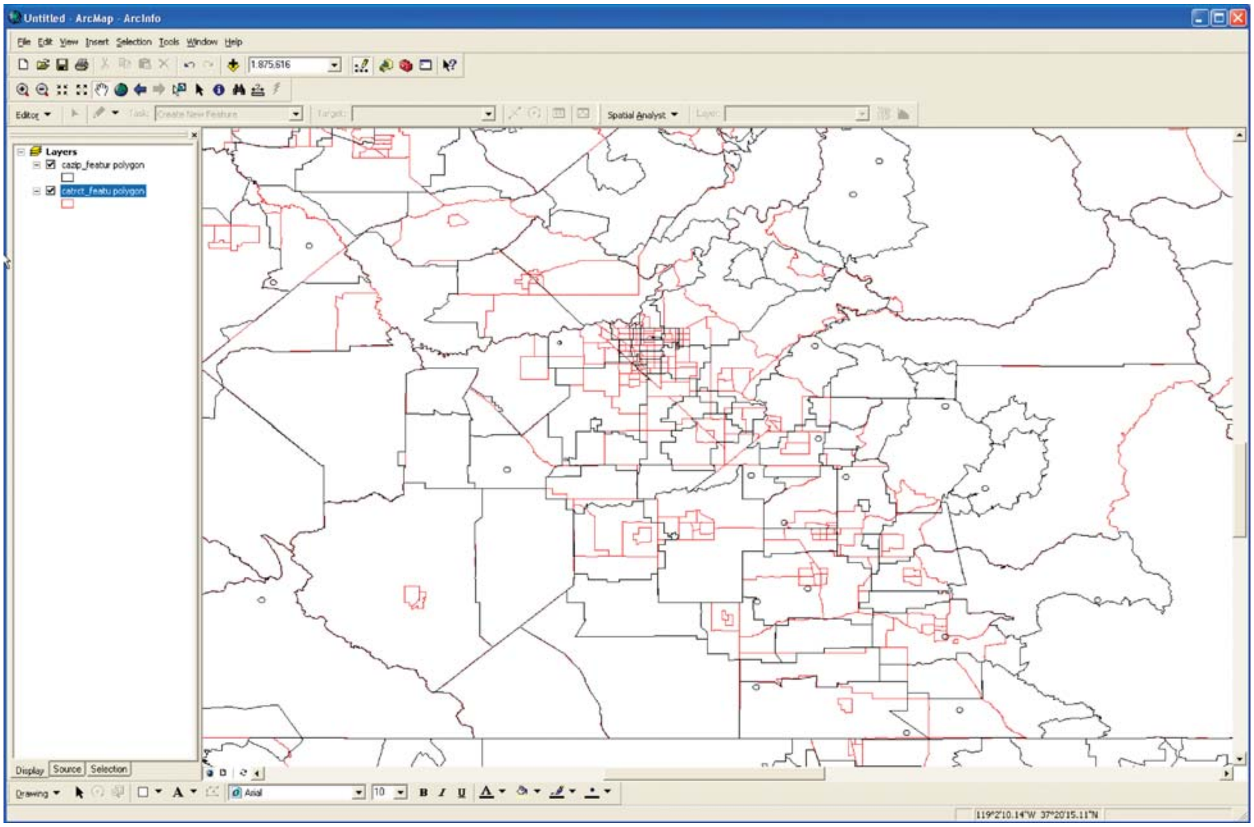

Figure 4 A screen shot of ESRI's ArcMap showing census tracts (red) overlaid on postal ZIP code boundaries (red) for an area of central California. Polygon overlay might be used in such a case to estimate missing attributes of ZIP codes, such as population, based on areas of overlap with tracts.

when boundaries which should be identical on both maps in fact differ by small amounts. In raster the problem is formulated a little differently, since the input attributes must be combined in some way for each pixel, and the spurious polygon problem does not apply because of the raster's finite resolution.

Overlays can be defined for any combination of point, line, or area features. The point in polygon problem arises when a layer of points is overlaid on a layer of polygons, in order to determine the containing polygon of each point. It is frequently applied in the analysis of point data to examine relationships between the density, count, or attributes of points and the attributes of containing areas. For example, one might use it to analyze the relationship between the number of instances of a disease in an area and the socioeconomic attributes of the area, or the area's average levels of atmospheric pollution.

A join occurs when the results of overlay are used to combine the attributes of the relevant features. For example, a point in polygon overlay might be used to add the attributes of each containing polygon to those of each point. In this way the name of the containing county might be added to each record in a set of points denoting accidents; or the name of the nearest street might be added to each record in a set of points representing tourist attractions and other landmarks.

Buffers

Another commonly employed function in GIS is the computation of a buffer. Any feature can be dilated by a specified distance to create a new feature known as a buffer. For example, a width of 100 m on either side of a stream might be designated as off limits for logging, or requiring special kinds of agricultural practice, to limit erosion and contamination. In some jurisdictions it is common for buffers to be used in notifying residents of proposed new developments, or to ban certain kinds of activity within a specified distance of schools (**Figure 5**). In vector GIS, buffering points result in polygonal approximations to circles; buffering lines result in polygonal bands on either side of each line and around its ends; and buffering areas result in new dilated polygonal areas. Buffer widths can be varied based on the attributes of individual source features.

The raster form of the buffer operation is notably different. In its simplest form, a layer is first defined as the source, and selected cells of that layer are labeled

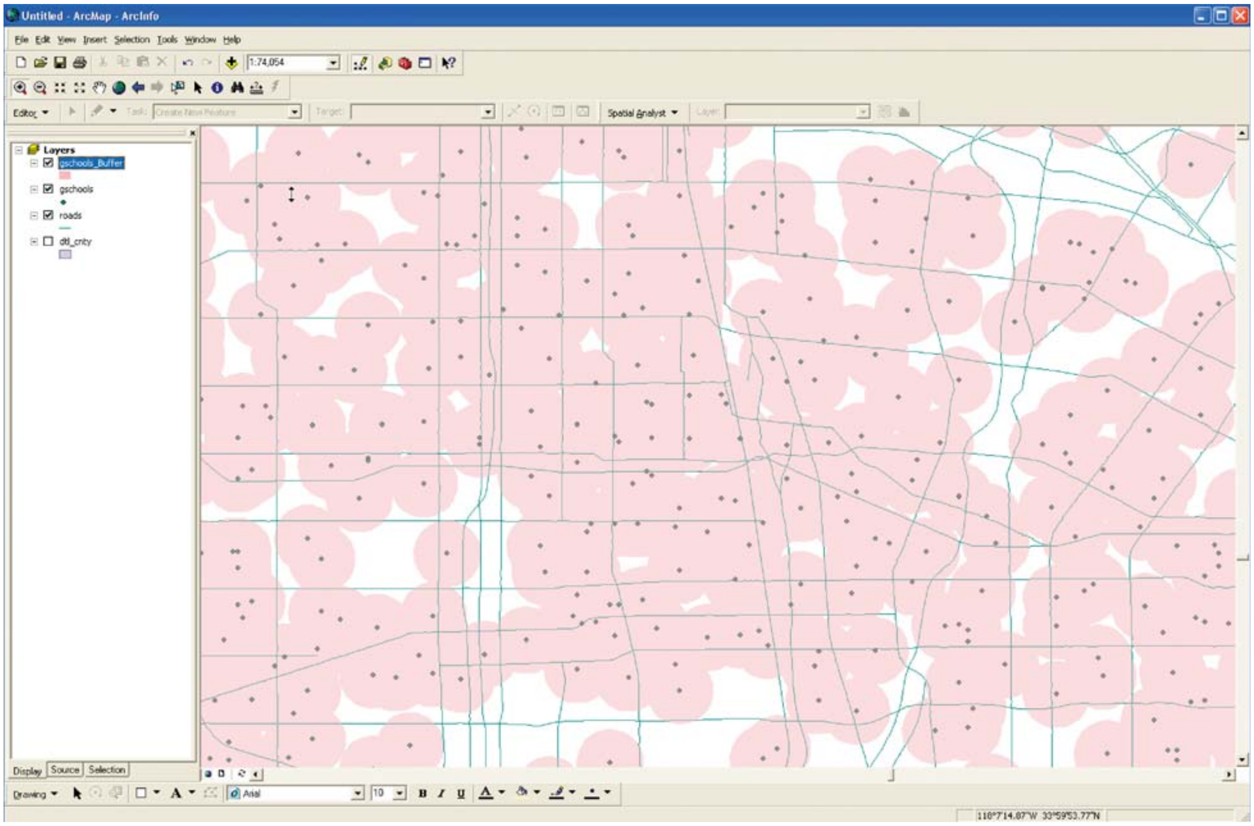

Figure 5 A screen shot of ESRI's ArcMap showing half-mile exclusion buffers drawn around every school in South-Central Los Angeles.

appropriately. The buffer operation then results in a new layer in which cells are labeled if they fall within the prescribed distance of a labeled cell on the source layer. The task can be generalized so that every output cell receives a measure of distance from the nearest cell on the source layer, rather than a binary indicator of whether distance is below a specified value. It can be further generalized if a layer of travel speed or travel cost or friction is available, in which case the buffers can be made to spread at variable rates, an operation that is virtually impossible in vector GIS.

Map Algebra

Provided the various layers in a raster GIS are precisely co registered, the operations of a raster GIS are in many ways simpler to organize and execute than their vector equivalents. Some of the most successful early GIS, particularly GIS designed to run on small computers, were based on raster representations, and much greater progress has been made in systematizing and standard izing raster operations. The most successful of these is map algebra, introduced in the late 1980s. Suppose that each layer in a raster GIS is addressed symbolically, just as scalars, vectors, and matrices are addressed in mathematics or in programming languages. Then the instruction '$C = A + B$' might be interpreted to mean: 'take the values stored in each pixel of layer A, add them cell by cell to the corresponding values in layer B, and store the result in a new layer C.' The need to loop through every cell of the raster is implicit.

In map algebra virtually all raster GIS functions are arranged in four categories. Functions that compare layers cell by cell, such as the previous example, are termed local. Focal functions compare or convolve a cell with its neighbors, and are commonly used to smooth rasters by averaging, or to predict the direction of water flow on a raster of elevations. A zone is defined as a contiguous patch of cells of the same value, and zonal operations address such properties as area, for example by creating a new layer in which each value is the number of cells in the containing zone. Finally, global operations are performed on an entire raster layer, for example by computing the mean value of the layer or other summary statistics.

Projections and Geodesy

Any analysis of the Earth's surface is inevitably compli cated by the fact that the surface is not flat. Curvature matters least in the analysis of small areas, but grows in importance as the extent of the analysis approaches global dimensions. While one might argue that all computer based analysis should respect the reality of curvature, in practice there are several strong arguments for projection, in other words for flattening the surface prior to its representation and analysis, despite the in evitable distortions that ensue. First, the paper medium is flat, and much human activity still revolves around the paper map, despite the popularity of digital media. Second, flattening is essential if one is to be able to visualize the entire Earth's surface at once. Third, it is impossible to rasterize a curved surface, so any raster must involve some form of projection. Finally, many of the operations of GIS were devised for the analysis of small areas and assume a flat surface.

To project the Earth's surface it is first necessary to find a mathematical form that approximates what is in reality a very complicated shape. Many such math ematical forms are in use in various parts of the world, though they are increasingly being replaced by a single universal form known as the World Geodetic System of 1984 (WGS84); though the latter does not provide such a good fit to any specific part of the Earth, its universality provides certain advantages. Nevertheless GIS users must confront the fact that in most areas of the Earth at least two such mathematical forms or datums are in common use; in North America WGS84 competes with the earlier North American Datum of 1927 (NAD27).

Several map projections based on these mathematical forms are commonly encountered in GIS. High precision projections such as the Universal Transverse Mercator (UTM) divide the surface into a number of zones to minimize distortion, ensuring that distances are distorted by only a few parts in ten thousand, and that directions and small shapes are correctly preserved. When estimation of area is important then a projection such as Albers, which preserves area correctly while distorting distances and shapes, is commonly used. Any projection implies a planar coordinate system and a set of equations that transform to and from latitude and longitude. Finally, many GIS applications use the so called 'unprojected' or cylindrical equidistant projection, which simply maps latitude to y and longitude to x. Its exaggeration of longitude relative to latitude increases rapidly at high latitudes, it distorts shapes and areas, but its simplicity is often attractive.

Metadata

Since the advent and popularization of the Internet vast investments have been made in sharing geographic data sets, particularly data sets representing commonly used features such as roads, rivers, and political boundaries. Whenever a data set is acquired from some remote source it is essential that the prospective user understand the data's basic properties, such as its contents, quality, projection, datum, and format. Metadata are defined as data about data, and have become an indispensable part of the entire data sharing enterprise. They can be as sembled into catalogs, and used to search for data meeting specific requirements. They also allow the user to determine the suitability of any data set for a specific use, to determine the data set's format and the meaning of its various components, such as its attributes, and the conditions under which the data were compiled.

The first widely accepted standard for metadata was promulgated by the US Federal Geographic Data Com mittee in the early 1990s. Since then it has been refined, and adopted after suitable modification by many national and international agencies. Its description of quality is based on five components: positional accuracy, attribute accuracy, logical consistency, completeness, and lineage. Nine other components are described, including details of the projection and datum, the date of validity, and the definitions of attributes.

Geoportals

A geoportal can be defined as a single point of access to geographic data sets located on servers distributed over the Internet. By providing a single point of access, geo portals allow users to search for data that might exist on any one of thousands of servers in data warehouses and digital libraries. A geoportal includes a catalog, which is compiled either manually from information sent by data custodians, or automatically by harvesting robots that scan the Internet much as do the operators of search services such as Google. Thus a geoportal might contain metadata records describing tens of thousands of data sets, all of which are actually held on remote servers. The user is able to search the catalog, identify suitable data sets, and retrieve them from their host servers using appropriate protocols, and after addressing such access issues as licensing or use fees. Geoportals are increas ingly important elements of the evolving spatial data infrastructure, or the system of institutions, standards, and protocols by which geographic data are assembled and distributed for use in many parts of the world.

The Agenda of GIScience

The first attempts to define a research agenda for GIScience, and to clarify the relationship between GIScience and the technology of GIS, occurred in the late 1980s. In the UK, the Department of the Envir onment's Committee of Enquiry into the Handling of Geographic Information (the Chorley Committee) saw three specific stimuli: the rapidly falling costs of hard ware, which had reduced the cost of entry into GIS and related activities from \$500 000 at the beginning of the decade to \$10 000 at the end; the advent of COTS (commercial, off the shelf) software to perform the basic operations of GIS; and rapid growth in the availability of spatially referenced digital data. In the US, the National Science Foundation announced a competition for a National Center for Geographic Information and Analysis (NCGIA), to advance the theory and methods of GIS, to promote the use of GIS across the sciences, and to increase the nation's supply of experts in GIS.

The first research agenda was published in 1988 by Rhind, who identified problems in what he termed the handling of geographic data: the volumes of data involved, the numerous types of queries that might be addressed, the prevalence of uncertainty in geographic data; the need for integration of data among organizations; and the lack of awareness of such issues as scale. He recognized that the solution of the more generic of these issues would come with time from mainstream information technology; but that issues that were more specific to the geographic case would have to be solved by an active research community focused on GIS. He saw a substantial role for knowledge based or expert systems in the automated extraction of features from images; the integration of disparate data sets; the development of intelligent search procedures; the automation of cartographic generalization; the development of machine based tutors; and the elicitation of knowledge from data. He also recognized the importance of research into better methods of visualization for geospatial data; the role of organizations; the legal issues of liability and intellectual property; and the costs and benefits of GIS.

The NCGIA research agenda, published in 1989, has much in common with Rhind's, but already shows signs of a search for the more fundamental issues of GIScience, in contrast to the practical issues of GIS. Five major research areas are identified:

- 'Spatial analysis and spatial statistics', the techniques used to model uncertainty in geospatial data, to mine data for patterns and anomalies, and to test theories by comparison with reality;
- 'Spatial relationships and database structures', addressing the representation of real geographic phenomena in digital form, and the interface between digital structures and human reasoning;
- 'Artificial intelligence and expert systems', reflecting Rhind's concern for the role of advanced machine intelligence in GIS operations;
- 'Visualization', and the need to advance traditional cartography to reflect the vastly greater potential of digital systems for display of geographic data; and
- 'Social, economic and institutional issues', the host of social issues surrounding GIS.

The NCGIA went on to propose 12 specific research initiatives within this general framework:

- 'Accuracy of spatial databases', focusing on error models for geographic data with strong links to the discipline of statistics, and attempting to characterize the errors inherent in both raster and vector representations;
- 'Languages of spatial relations', including principles of spatial cognition and linguistics;
- 'Multiple representations', the need to integrate different representations of the same phenomena on the Earth's surface, which might be raster or vector, or generalized at different scales;
- 'Use and value of geographic information in decision making';
- 'Architecture of very large GIS databases';
- 'Spatial decision support systems', the design of systems to support decision making by groups of stakeholders;
- 'Visualization of the quality of geographic information', through methods that explicitly display information about the uncertainty associated with data;
- 'Expert systems for cartographic design', using intelligent systems to augment the skill of cartographers;
- 'Institutions sharing geographic information', including research on the impediments to sharing between agencies;
- 'Temporal relations in GIS', the extension of GIS data models to include time;
- 'Space–time statistical models in GIS', the extension of spatial analysis to include time; and
- 'Remote sensing and GIS', researching the issues involved in the integration of data acquired by remote sensing with data from other sources.

Eventually, NCGIA sponsored a total of 21 research initiatives between 1988 and 1996.

Very substantial progress was made on most of these topics in the years following their publication. In addition, four factors contributed to the evolution of these research agendas in the 1990s: first, the continued arrival of new technologies, including most notably the World Wide Web, the Global Positioning System, object orientation, and mobile computing; second, the broadening of the research community, to include active participation by new disciplines, including linguistics and cognitive science; third, the trend away from technical issues of systems to fundamental issues of science; and fourth, the recognition that certain topics were in effect dead ends. This last perhaps accounts for the virtual disappearance of expert systems, despite their prominence in Rhind's 1988 agenda.

In 1996 the recently formed University Consortium for Geographic Information Science published the first edition of its research agenda, the result of a successful consensus building exercise amongst the thirty or so research institutions that were then members (the number has since risen to more than 80). The agenda had ten topics:

- ‘Spatial data acquisition and integration’, including new sources of remote sensing, ground based sensor networks, and fusion and conflation of data from dif ferent sources;
- ‘Distributed computing’, and the issues of integrating data and software over large heterogeneous networks;
- ‘Extensions to geographic representations’, addressing particularly the third spatial dimension and time;
- ‘Cognition of geographic information’, including studies of the processes by which people learn and reason with geographic data, and interact with GIS;
- ‘Interoperability of geographic information’, including research to overcome the difficulties of different for mats and lack of shared understanding of meaning;
- ‘Scale’, and the complex issues surrounding repre sentations at different levels of detail;
- ‘Spatial analysis in a GIS environment’, advancing the analytic capabilities of GIS;
- ‘The future of the spatial information infrastructure’ and the institutional arrangements that provide the context for GIS;
- ‘Uncertainty in geographic data and GIS based analy sis’, advancing understanding of the nature of differ ences between raster and vector databases and the real phenomena they attempt to represent, and including the modeling and visualization of data quality; and
- ‘GIS and society’, the study of the impacts of GIS on society, and the societal context in which the tech nology is used; these issues are increasingly recog nized under the more general heading of critical GIS.

UCGIS later added four emerging themes to the list:

- ‘Geospatial data mining and knowledge discovery’, the development of methods for extracting patterns and knowledge from very large data sources;
- ‘Ontological foundations of geographic information science’, addressing the fundamental components on which our knowledge of the Earth’s surface is based;
- ‘Geographic visualization’; and
- ‘Remotely acquired data and information in GIScience’.

A somewhat different approach to framing the research agenda was taken by NCGIA’s Project Varenius, a re search effort begun in 1996 to advance the fundamentals of GIScience. In this strikingly simple model, GIScience was anchored by three concepts – the individual, the computer, and society – represented by a triangle, with GIScience at the core. Research about the individual would be dominated by cognitive science, and its concern for understanding of spatial concepts, learning and rea soning about geographic data, and interaction with the computer. Research about the computer would be dom inated by issues of representation, the adaptation of new technologies, computation, and visualization. Finally, research about society would address issues of impacts and societal context. Many research issues would involve the interaction between the three corners of the triangle.

While all of these various attempts to define the re search agenda of GIScience have intellectual merit and show a definite emergence of consensus, they all lack the compelling appeal of such scientific mega projects as Martian exploration or the mapping of the human ge nome. But the concept of Digital Earth perhaps has the ability to capture popular imagination. The term was coined by then Senator Al Gore in 1992 and elaborated in a much quoted 1998 speech:

> Imagine, for example, a young child going to a Digital Earth exhibit at a local museum. After donning a head mounted display, she sees Earth as it appears from space. Using a data glove, she zooms in, using higher and higher levels of resolution, to see continents, then regions, countries, cities, and finally individual houses, trees, and other natural and man made objects. Having found an area of the planet she is interested in exploring, she takes the equivalent of a ‘magic carpet ride’ through a 3D visualization of the terrain. Of course, terrain is only one of the numerous kinds of data with which she can interact. Using the system’s voice recognition capabil ities, she is able to request information on land cover, distribution of plant and animal species, real time wea ther, roads, political boundaries, and population. She can also visualize the environmental information that she and other students all over the world have collected as part of the GLOBE project. This information can be seamlessly fused with the digital map or terrain data. She can get more information on many of the objects she sees by using her data glove to click on a hyperlink. To prepare for her family’s vacation to Yellowstone National Park, for example, she plans the perfect hike to the geysers, bison, and bighorn sheep that she has just read about. In fact, she can follow the trail visually from start to finish before she ever leaves the museum in her hometown. She is not limited to moving through space, but can also travel through time. After taking a virtual field trip to Paris to visit the Louvre, she moves backward in time to learn about French history, perusing digitized maps overlaid on the surface of the Digital Earth, newsreel footage, oral history, newspapers and other primary sources. She sends some of this information to her per sonal e mail address to study later. The time line, which stretches off in the distance, can be set for days, years, centuries, or even geological epochs, for those occasions when she wants to learn more about dinosaurs.

Almost 10 years later it is possible to recognize many but by no means all of the elements of this vision in such services as Google Earth, and research is underway both to realize the remaining elements of the concept and to examine its deeper social meanings.

Google Earth has had a massive impact since its launch in 2005, in introducing many of these concepts of GIS to tens of millions of users. Google Maps, Microsoft Virtual Earth, and many other Web services have only increased the impact. Today, the term neogeography is being used to describe a renewed interest in geography on the part of users of these services. Moreover, nu merous Web sites such as Wikimapia, Flickr, and Open StreetMap are allowing millions of individuals to create and integrate their own geographic information in large repositories, potentially bypassing the traditional means of geographic information production, which has been dominated by the national mapping agencies.

Other technologies, such as remote sensing, GPS, and RFID (radio frequency identification), are challenging traditional notions of locational privacy, as it becomes increasingly possible to track the geographic positions of individuals and vehicles, and to obtain information from them in real time. The term spatial web is being used to describe a future in which many features in the geo graphic world will be responsive in the way that aircraft are responsive to air traffic controllers: able to report their positions and identify themselves automatically. This future already exists in many retail stores where items for sale carry RFID tags, in many areas of livestock farming, where every animal is tagged, and in modern construction, where the components of a building are able to identify themselves.

Despite the longstanding nature of many of the con cepts reviewed in earlier sections, the world of GIS is evolving more rapidly than ever. The number of people directly engaged in some aspect of its use has increased by many orders of magnitude since the early days of GIS in the 1980s, and seems set to continue increasing as services become easier to use, and as people recognize the im portance of information about location in their daily lives.

See also: Geomatics; GIS and Cartography; GIS, Mobile and Locational Based Services; GIS, Public Participation.

Further Reading

Arctur, D. and Zeiler, M. (2004). *Designing Geodatabases: Case Studies in GIS Data Modeling*. Redlands, CA: ESRI Press.

Bugayevskiy, L. M. and Snyder, J. P. (1995). *Map Projections: A Reference Manual*. London: Taylor and Francis.

Cho, G. (1998). *Geographic Information Systems and the Law: Mapping the Legal Frontiers*. Chichester, UK: John Wiley & Sons.

Department of the Environment (1987). Handling geographic information. The Report of the Committee of Enquiry Chaired by Lord Chorley. London: Her Majesty's Stationery Office.

Duckham, M., Goodchild, M. F. and Worboys, M. F. (eds.) (2003). *Foundations of Geographic Information Science*. New York: Taylor and Francis.

Goodchild, M. F. (1992). Geographical information science. *International Journal of Geographical Information Systems* 6(1), 31 45.

Goodchild, M. F., Egenhofer, M. J., Kemp, K. K., Mark, D. M. and Sheppard, E. S. (1999). Introduction to the Varenius project. *International Journal of Geographical Information Science* 13(8), 731 746.

Gore, A. (1992). *Earth in the Balance: Ecology and the Human Spirit*. Boston: Houghton Mifflin.

Knowles, A. K. (ed.) (2002). *Past Time, Past Place: GIS for History*. Redlands, CA: ESRI Press.

Longley, P. A., Goodchild, M. F., Maguire, D. J. and Rhind, D. W. (1999). *Geographical Information Systems: Principles, Techniques, Management and Applications*. New York: Wiley.

Longley, P. A., Goodchild, M. F., Maguire, D. J. and Rhind, D. W. (2005). *Geographic Information Systems and Science* (2nd edn.). New York: Wiley.

National Research Council (2003). *IT Roadmap to a Geospatial Future*. Washington, DC: National Academy Press.

National Science Foundation (1999). Geographic information science: Critical issues in an emerging cross disciplinary research domain. *Report of the National Science Foundation*. Washington, DC: National Science Foundation. http://www.geog.buffalo.edu/ncgia/workshopreport.html (last accessed April 2008).

Pickles, J. (ed.) (1995). *Ground Truth: The Social Implications of Geographic Information Systems*. New York: Guilford Press.

Rhind, D. W. (1988). A GIS research agenda. *International Journal of Geographical Information Systems* 2(1), 23 28.

University Consortium for Geographic Information Science (1996). Research priorities for geographic information science. *Cartography and Geographic Information Science* 23(3).

Zeiler, M. (1999). *Modeling Our World: The ESRI Guide to Geodatabase Design*. Redlands, CA: ESRI Press.

Global Commodity Chains

L. van Grunsven, Utrecht University, Utrecht, The Netherlands

Glossary

Arm's Length Trade The buying and selling of standardized commodities based on price.

Codification The transfer of complex tacit information into standard sets of written instructions that can be generally understood.

Commodity Chain The full range of activities that are required to bring a product from its conception to its end use and beyond.

Externalization The transfer of operations to outside the boundaries of the firm (via *inter alia* outsourcing).

Fragmentation The subdivision of the value chain (operations) of a commodity in individual operations to be carried out either inside the firm or by different firms.

Governance The exercise of coordination and control over the structuring of value chain operations, organizationally and geographically.

Industrial Upgrading Improvement of the ability of a firm to move to more profitable and/or technologically sophisticated capital- and skill-intensive economic niches.

Lead Firm A firm that through its position in the value chain and ownership of specific assets is able to coordinate and control the structuring of the value chain, from which it derives power.

Vertical Integration The organization of all operations of the value chain of products within a single firm.

Global Commodity Chains: Key Ideas and Concepts

Globalization has led to increasingly diverse, complex, and dynamic forms of industrial organization, divisions of labor, and geographies of production and consumption. The architecture of cross border economic activity and its meanings being far from straightforward, the search has been for devices enabling a grasp of this diversity, complexity, dynamics, and geographies of the patterning of production versus consumption. One of these devices is the global commodity chain (GCC).

Coined in the 1980s, in various aspects thinking surrounding this concept has evolved over the past decade. The understanding of the implications of a GCC view on the functioning of the global economy has improved as well. In part, this is a result of substantial empirical research conducted firstly into the – dynamics of – organizational and territorial structures of production and consumption, and the links between these, as to a range of commodities in various sectors; and secondly, into the implications of these structures for specific actors operating in them.

Central in commodity chain thinking is the idea that moving a commodity from initial conception to final use can be thought of as a process consisting of a number of sequential operations linked in a chain (each of which add value). Organizationally, thus a core idea is segmentation and separation of a process in discrete activities, each of which however is necessary. Roughly, main operations are design/product development, procurement of inputs, production, distribution, marketing, selling/retailing, and servicing consumers (see also **Figure 1**). Deep segmentation involves the additional separation of main operations in more specific processes. Isolated operations may be identified as 'nodes', both organizationally and territorially. (Global) commodity chain analysis focuses on describing and explaining the organizational and territorial construct of individual nodes within the entire sequence, the linkages (conceived as transactions) between nodes in the overall process, how these are shaped, as well as the implications for individual nodes and actors in these of the construct of the overall process.

The above is reflected in definitions of commodity chain. Prior to its popularization in the GCC literature during the 1990s, it was described as 'a network of labor and production processes whose end result is a finished commodity'. Since then, a terminological jungle has come to characterize the literature. The table below (**Table 1**) gives an overview of the different – to some extent overlapping – terms and definitions found. Frequently, the term (global) value chain is used as substitute for (global) commodity chain. A more recent definition is: 'the full range of activities that are required to bring a product from its conception to its end use and beyond'. A definition like 'a set of networks (nodes) clustered around one final product or service and linking firms, industries, and communities to one another across the world economy', expresses the 'global' element brought into it by globalization discourse. Over the last decade or so, conceptual elaboration of GCC departed from four main dimensions:

1. An 'input–output structure', or the sequence of interrelated value adding activities, as depicted above, as well as the configuration of the actors carrying out main (and sub) operations. The latter are seen to range

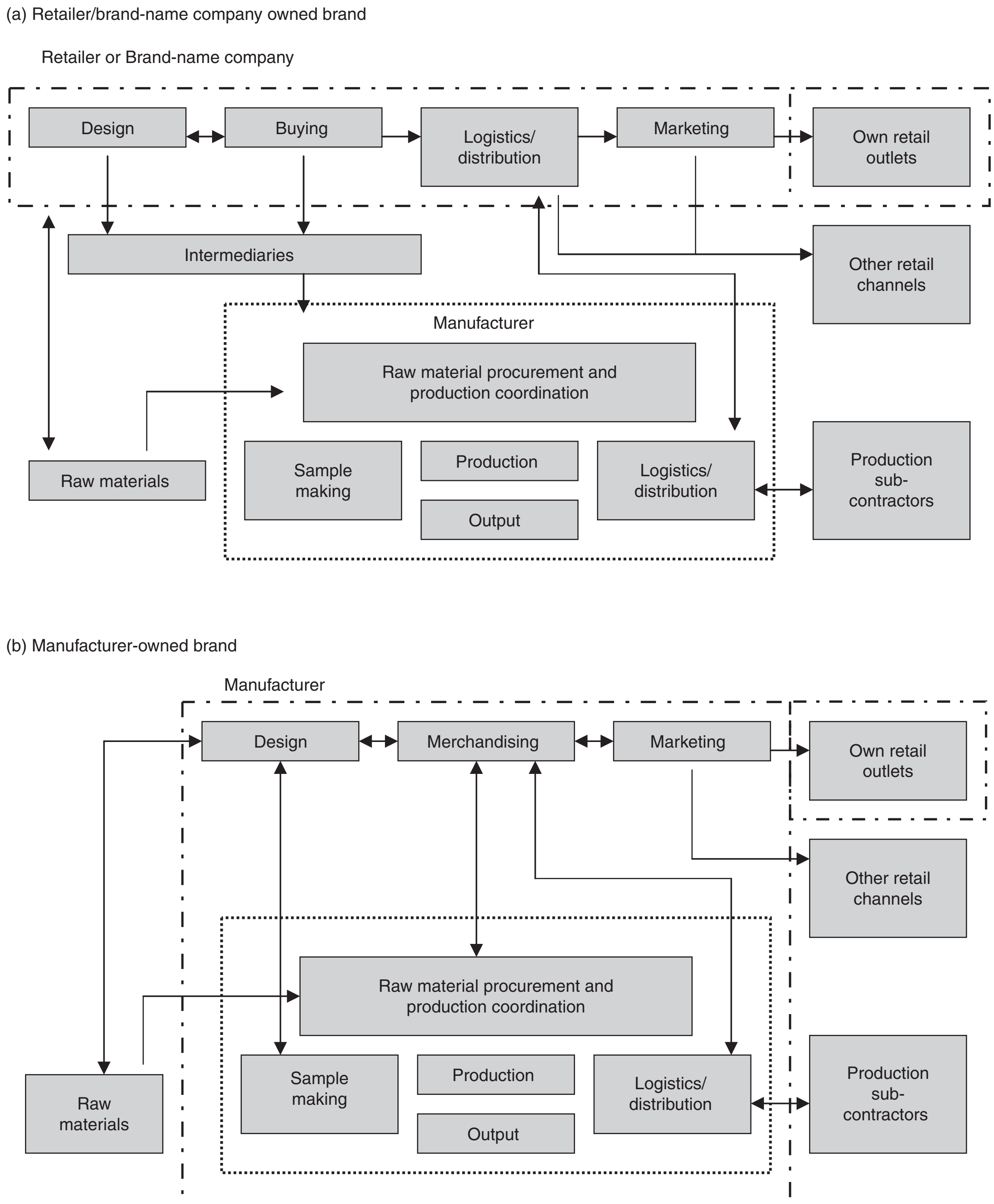

Figure 1 Clothing intra-industry commodity chain configurations according to brand-name ownership (simplified). Source: Adapted from Hassler, M. (2003). The global clothing production system: Commodity chains and business networks. *Global Networks* 3, 513 531.

from a single firm that internalizes the entire chain to an array of firms (brought together in a network) performing individual processes.

2. A 'governance structure', or the coordination and control of transactions in the process and the power relations that determine how economic surplus is distributed within the chain. A distinction here could be made between two, not quite independent, aspects: who governs, in terms of the nodes, and the forms coordination assume. Some actors in the chain (lead firms) have power over others, setting the terms under which others in the chain operate. The latter are seen

Table 1 Networks and chains: terminology and definitions

Concept/term (synonyms)	*Definition*	*Focus/what it highlights*
Supply chain	Generic label for input output structure of value-adding activities, beginning with raw materials and ending with finished product	• Input output structure
Porter's value chain	The different stages of the process of supply (inbound logistics, operations, out-bound logistics, marketing and sales, and after-sales) and the support services the firm marshals to accomplish this task	• Draws away attention from the physical transformation to where value is added in the process of bringing a product or service to the end user • Gaining competitiveness through management of the value chain as a system
Value system (Commodity/production value chain, (French) Filière[a])	A sequence of productive (i.e., value-adding) activities across industries (inter-industry linkages) leading to/supporting, end use	
Production chain (Commodity chain value system/ chain (French) Filière)	A transactionally linked sequence of functions in which each stage adds value to the process of production of goods or services	• Input output structure • Coordination and regulation of chains • Geographical configuration
Commodity chain (Value system/chain, (French) Filière, production chain)	A network of labor and production processes whose end result is a finished commodity	• Input output structure • Governance structure
Value chain (Commodity/production chain, value system (French) Filière)	A sequence of productive (i.e., value added) activities often across industries (interindustry linkages) leading to, and supporting, an end use (product or service)	• Relative value of activities required to bring a product or service from conception through phases of production to end use and disposal • Value chains as repositories for rent • Governance/systemic efficiency required for effective functioning of chain
International/global production network (Supply base)	A set of interfirm relationships that bind a group of firms into a larger economic unit and in which actors coordinate activities across countries and even continents	
Global commodity chain (Global value chain)	A set of networks (nodes) clustered around one final product or service, linking firms, industries, and communities to one another across the world economy	• Geographical configuration, international dimension • Territorial embeddedness/social and institutional context • Networks as strategic assets • Distribution of wealth within chain is outcome of relative intensity of competition within different nodes; nodes rather than national economies are locus where surplus accrues

[a] The French Filière approach is not so much centered around a distinct school of thought or a unified theoretical framework. Instead it is a loosely knit set of studies with the common characteristics that they use the filière or chain of activities and exchanges as a tool and to delimit the scope of their analysis. It is thus less a theory than a 'meso level' of analysis.
Source: Own compilation from various sources.

to range between two extremes, namely arm's length trade and vertical integration.

3. A 'geographical configuration', referring to the spatial structure or territoriality of nodes. As each node can by itself be conceived as a network connected to other nodes with related activities, not only the geographical location of nodes is relevant but also the geographical structure of internal and external linkages. The ter ritoriality of chains ranges from high geographical concentration to high geographical dispersion, from the local scale to the global scale. This is influenced by technologies of transport and communication, by international trading arrangements, and by a range of other factors that will be referred to later in this article.
4. A 'social and institutional context', referring to the norms, values, and regulatory frameworks in various fields and at different scale levels within which nodes are embedded, to which actors are subjected and that impact options as to other dimensions.

An often made observation is that – when these dimensions are combined and observed through time – there appears to be a trend for commodity chain activities to be divided among (many) different firms and spread globally over many locations rather than contained within a single firm and a single location. Increasingly the GCC framework embodies the notion of organizational, transactional, and geographical stretching whereby commodities are produced (in externalized networks) by firms located in less developed countries, to be sold to consumers in countries at the other side of the globe by production or retailing firms that orchestrate the cross border production and flows of these commodities. Hence the term global commodity chains.

Reflecting the main direction of academic work, below we elaborate on the insights that have evolved as to the governance dimension.

Types and Governance of GCCs

Within the GCC framework, increasing diversity, complexity, and dynamics of industrial organization, divisions of labor, and geographies of production and consumption are captured through the idea that GCCs are heterogeneous in terms of the firms or actors that drive the chain, how the chain is configured, and how linkages and transactions are managed. These have rendered governance a key point of attention.

Governance is considered key also as it determines not only how and on what terms producers are inserted into the global division of labor and the returns to them, but also the capacity to exercise power over other parts of the chain. This capacity exists in gradations reflected in governance within a chain perspective taking on different forms. The term lead firm is associated with assuming critical governing roles. By positioning themselves strategically in the chains, firms can – by exercising some form of power – diminish transaction costs, enhance their capacity to raise entry barriers, and increase economic returns.

Over the past decade, gradually more sophisticated typologies of governance have been developed and a rather refined 'theory of governance' has been advanced. Initially, the focus was on 'who governs' and thus on the dominant actor in the chain. **Table 2** depicts current typology of firm actors in commodity chains. Predominantly, in the literature a distinction was made between 'buyer driven commodity chains (BDCCs)' and 'producer driven commodity chains (PDCCs)'. In the former, governance is in the hands of large retailers or branded marketers and relationships between actors are trade based. Such chains were held to be typical for consumer nondurable industries such as garment, footwear, and toys. In the latter, coordination is in the hands of large producers and relationships between actors are investment based. These were held to be typical of consumer durable industries such as automobiles, computers, and aircrafts.

This simple dichotomy has later on been reworked into a typology based on brand name ownership (as a main driving force) that also to some extent conveys how governance can be configured as different forms of brand name ownership create different coordination systems of intra and interfirm relationships resulting in characteristic ways in which chains operate. While avoiding the identification of specific drivers (or main actors) with specific sectors or industries, by recognizing that there are various possible intrasector or intra industry forms of brand name ownership, this typology still distinguishes only two types of governance, according to whether a brand name is owned by a retailer/brand name company or by a manufacturer.

In case of the former, chain configuration tends toward sourcing of products whereby the retailing or brand name company concentrates on the decisive nodes of product design, marketing, and coordination of an external manufacturer network. However, intermediaries like representative offices or agents may take over, in part or wholly, producer network coordination. In case of the latter, chain configuration appears to tend toward maintaining product development, production, and distribution in the firm's internal organizational structure (vertical integration), leading to an intrafirm network of wholly or partly owned production facilities. The two configurations are depicted in **Figure 1**, applied to the clothing sector.

A recently proposed theory of governance, besides outlining a more complete typology, postulates that substantial variations in governance result from the interplay of three main variables.

The complexity of transactions. More complex transactions require more interaction among actors and thus stronger forms of governance.

The 'codifiability' of transactions. In some industries schemes have been worked out to transfer complex information into standard sets of written instructions (codification) in a manner that can be handed off between partners with relative ease. If suppliers have the competence to act upon such instructions, and if the schemes are widely known and used, then outsourcing with little monitoring is likely to emerge. If not, then lead firms might either keep operations in house, leading to more vertical integration, or exert tight control over and monitor suppliers, or have a dense relationship with suppliers.

The competences of suppliers. The ability to receive and act upon complex instructions from lead firms requires a high degree of competence on the part of suppliers. The transfer of such instructions can be achieved by widely known standards and protocols or through intense

Table 2 Commodity chain actors

Actor	*Scope of activity*	*Other names*	*Firm examples (USA)*
1. Integrated firm	• Product strategy • Product definition • Design • Manufacturing • Subassembly • Component manufacturing • Marketing, sales, and distribution	• Modern corporation • Dinosaur	• Phillips • The 'old' IBM • The 'old' Ford
2. Retailer	• Sales • Marketing • Value-added packaging and system integration	• Marketer • Distribution • Reseller • Value-added reseller	• Amazon.com • Sears • Gap • Banana Republic
3. Lead firm	• Product strategy • Product definition • Product design • Manufacturing • End-user sales • End-user marketing	• Brand-name firm • OEM • Anchor firm	• Dell • Nike • Daimler • The 'new' Ford • The 'new' IBM
4. Turn-key supplier	• Complex parts and services • Process R&D	• System supplier • OEM supplier • First-tier supplier • Contract manufacturer • Full-package supplier • Global supplier	• Flextronics • Celestica • Solectron • TSMC, UMC • Dana, Delphi • UPS, Fedex • Arthur Anderson
5. Component supplier	• Discrete elements (component parts and services)	• Lower-tier supplier • Specialized supplier • Sub-contractor • Commodity producer	• Intel • Microsoft • BF Goodrich

Source: Gereffi, G. and Kaplinski, R. (eds.) (2001). The value of value chains: Spreading the gains from globalization. Sussex: Institute of Development Studies. IDS Bulletin 32., No. 3.

interaction. Where competent suppliers do not exist, lead firms must either keep the function in house or outsource it to suppliers that they tightly monitor and control.

The more complete typology of chain governance posits arm's length trade among firms and vertical integration at opposite ends of a spectrum of coordination, with network relationships constituting an intermediate mode. As to the network category three distinct types are identified: modular, relational, and captive. This resulting typology is as follows.

Arm's length trade. This entails the buying and selling of products among firms and individuals with interaction limited to exchanging goods and services for money. The governance mechanism is price. Transactions may be repeated. Transactional dependence (i.e., the costs of switching to new partners) is low for both parties. As the information that needs to be exchanged and knowledge that needs to be shared are straightforward, linkages between chain activities are 'thin'. Drivers are: easy codification of transactions, simple product specifications, and producers having the capability to make the product without involvement from buyers.

Modular commodity chains. Suppliers in these chains make products or provide services to a customer's specifications. Irrespective of the detail of instructions, modular suppliers are highly competent, with an ability to provide 'turn key' or 'full package' services. They can deliver complete products, often using generic machinery that spreads investments across a wide customer base and keeps switching costs low. As buyer–supplier interactions can be complex (with high volume of information flowing across), linkages are necessarily thicker than with buying/selling. At the same time codification schemes and the embodying of much of the knowledge in modules can keep interactions between partners from being very dense. Thus, drivers are the ability to codify complex product specifications, technical standards reducing variation, high competence of suppliers reducing the need for direct monitoring and control.

Relational commodity chains. This governance pattern is driven by the inability to codify product specifications necessitating the exchange of knowledge in person, complex transactions, and supplier capabilities that involve complementary competences. To support dense interactions and knowledge sharing, partners have deep understanding of one another. There is mutual dependence regulated through reputation, social and spatial proximity, family and ethnic ties, contracts that impose high costs on breakage, and the like. As trust and mutual dependence take a long time to build up, are difficult and time consuming to reestablish with new partners, and proximity is limited to a small set of firms, the costs of switching to new partners tend to be high.

Captive commodity chains. This last network style governance pattern is driven by substantial ability to codify high complexity of product specifications accompanied by low supplier capabilities. Small suppliers are dependent on larger, dominant lead firms that provide a large share of the business and process technology. This raises switching costs for suppliers who are therefore 'captive'. There is a high degree of monitoring and control by the lead firm, which also creates dependence for the lead firm on its 'captive' supply base as a whole. Given power relationships, suppliers have to conform to the modes of interaction specified by the dominant customer, leading to 'thick' linkages and high switching costs.

Vertical integration. This pattern is characterized by transactions taking place inside a single firm, coordinated by managerial control. This occurs when product specifications cannot be codified, products are complex, highly competent suppliers cannot be found, there is a need to manage complex webs of inputs and outputs, and to control resources like intellectual property.

Table 3 summarizes how the 'values' of the variables discussed earlier combine into the different governance types.

Apart from the above typology and interpretation, the theory also postulates that governance is dynamic. The form of coordination tends to change (in a predictable way) as and when there is a change in any (or more) of the variables listed above. Obviously, such changes do frequently occur. For example, the theory states that rising supplier competence might result in captive networks moving toward the relational type while better codification schemes enable modular networks. Conversely, when a new technology renders an existing codification scheme obsolete and effectively lowers the competence level in the supply base, one might expect modular coordination to become more relational, and if competent suppliers cannot be found, then captive coordination and vertical integration will become more prevalent. One illustration in the clothing sector is the case of the Spanish company Inditex, owner of the Zara and other middle/higher segment brand names, which has adopted a unique and complex fast fashion business model. The demands imposed by the model have resulted in Inditex adopting a vertical integration form of coordination, supplemented with a captive network of small scale producers in close proximity to its own production hub. **Table 3** also summarizes the main types of transition that one is likely to encounter.

It may be noted that the theory has been set up from a dominant lead firm perspective. Apart from stating that competences of suppliers may change, it generally pays little regard to how governance is influenced by strategies that may be adopted by nonlead firms (suppliers), rendering more power to these firms than the theory seems to allow for. For example, many suppliers in the clothing sector explicitly pursue a multiple client base, in order to avoid lock in into captive transactions with high dependency. Second, in network type chain the theory assumes direct linkage between Western lead firm and supplier. Recently, it has been argued that non Western intermediaries are starting to figure more prominently in such chains, not only increasing the variation of governance types but also significantly influencing power relations, as these intermediaries assume a more extensive chain coordination role.

Integration in GCCs and Industrial Upgrading: The Role of Governance

GCC discourse has substantially contributed to the debate on globalization and local industry development in

Table 3 Key determinants of global commodity chain governance

Governance type	*Complexity of transactions*	*Ability to codify transactions*	*Capabilities in the supply base*	*Degree of explicit coordination and power asymmetry*
Market	Low	High	High	Low
Modular	High	High	High	
Relational	High	Low	High	
Captive	High	High	Low	
Hierarchy	High	Low	Low	High

1. Increasing complexity of transactions also reducing supplier competence in relation to new demands: from modular to captive.
2. Decreasing complexity of transactions and greater ease of codification: from captive/relational to modular.
3. Better codification of transactions: from relational to modular.
4. De codification of transactions: from modular to relational.
5. Increasing supplier competence: from captive to relational/modular.
6. Decreasing supplier competence: from modular/relational to captive.

Source: Gereffi, G., Humphrey, J. and Sturgeon, T. (2005). The governance of global value chains. *Review of International Political Economy* 12, 78 104.

developing countries by arguing that incorporation into GCCs driven by Western lead firms constitutes the best option for development of the capabilities of local firms thus creating the conditions for upgrading. Contract manufacturing in BDCCs is seen to enable 'latecomer' firms in developing countries to overcome the dis advantages of being far removed from lead user markets and sources of technology. Achieving partici pation in global chains requires local firms and industries to engage in initial learning and upgrading of quality to meet the minimum conditions of lead firms. It is sub sequently seen as putting firms in developing countries on a fast track to growth, upgrading and improvement of competitiveness as it provides access to resources, thus speeding up, and bringing down the cost of, learning processes.

Industrial upgrading is defined as an improvement of the ability of a firm to move to more profitable and/or technologically sophisticated capital and skill intensive economic niches. Integration in global commodity chains is perceived in first instance to allow four different shifts: (1) 'process upgrading', transforming inputs to outputs more efficiently by reorganizing the production system or introducing better technology; (2) 'product up grading', implying firms move into more sophisticated, higher unit value, product lines; (3) 'functional up grading', implying firms acquire new functions or abandon existing ones so that they increase the overall skill content of their activities (e.g., complement pro duction with design or marketing); and (4) 'intersectoral upgrading', where firms apply the competence acquired in a particular function of a chain to move into a new sector. Particularly, functional and intersectoral up grading is seen to possibly lead to a repositioning of (groups of) firms in global markets.

Linked to this categorization of shifts is a specific firm development model depicting the trajectory firms are able, are even 'supposed', to follow according to the upgrading proponents. The so called organizational succession and export learning model suggests a trajec tory of moving from simple assembly to original equip ment manufacturing (OEM) on the basis of third party designs and brands, to own design manufacturing (still third party brands but significant design input), to eventually own brand manufacturing (OBM). The latter necessitates functional upgrading. **Figure 2** portrays process, product, and functional upgrading in an inter related way, triggered by profit opportunities provided by fast growing export markets, access to which is secured through integration in GCCs.

With new insights produced by more empirical re search, the above portrayal of industrial upgrading has increasingly become contested. First, criticisms concern the failure to properly define and problematize the up grading concept, that is, to distinguish between the 'process' of upgrading and upgrading as an 'outcome'. Second, it is often ignored that upgrading means to do things differently, and/or to do different things – com pared to competitors! Third, viewing upgrading from an internal perspective, external processes, for example, in the global and local social and institutional environment, are often overlooked.

Most importantly however, much of the earlier dis course presents a 'conveyer belt' picture of industrial upgrading, conceiving longer term development trajec tories as unilinear and unproblematic. More recent thinking has acknowledged the role of governance in shaping upgrading opportunities, producing alternative trajectories. Thus it is suggested that firms in captive networks are likely to experience rapid product and process upgrading, but opportunities for functional upgrading are severely curtailed. The former can be understood from factors discussed already in the above, the latter meets with obstacles such as lead firm power and resource requirements. The limited interest of lead firms in networked firms broadening their roles in GCCs and high switching costs lead networked firms to con tinue in their existing role (path dependency) producing 'lock in'. The point of switching costs (associated with the necessity to invest in learning beyond production to acquire sufficient competences in other nodes of the chain) also often leads firms in relational and even modular chains to voluntarily stick to their existing position and emphasize on outperforming other firms in the same node. On the other hand, functional upgrading is much more common among local firms in developing countries that 'choose' to operate outside global com modity chains and/or pursue other routes toward building export (capability). This brings us to consider a few of the shortcomings of GCC thinking identified in the literature.

Global Commodity Chain Thinking: Some Shortcomings and Notes

The Typology of Commodity Chains Revisited: Away from Western Centrism

Global commodity chain thinking has increasingly been criticized for a rather narrow interpretation of how firms from developing countries can integrate in global in dustries and markets. By giving a central position to Western lead firms, by arguing that firms from developing countries start by exporting as suppliers, and by arguing the crucial role of this in building competences, GCC thinking is marked by Western centrism, as a result ig noring a larger variety of possible development paths of firms and routes to export. Recent research has unveiled growth trajectories of local firms in a range of industries that start with producing own brand products for the

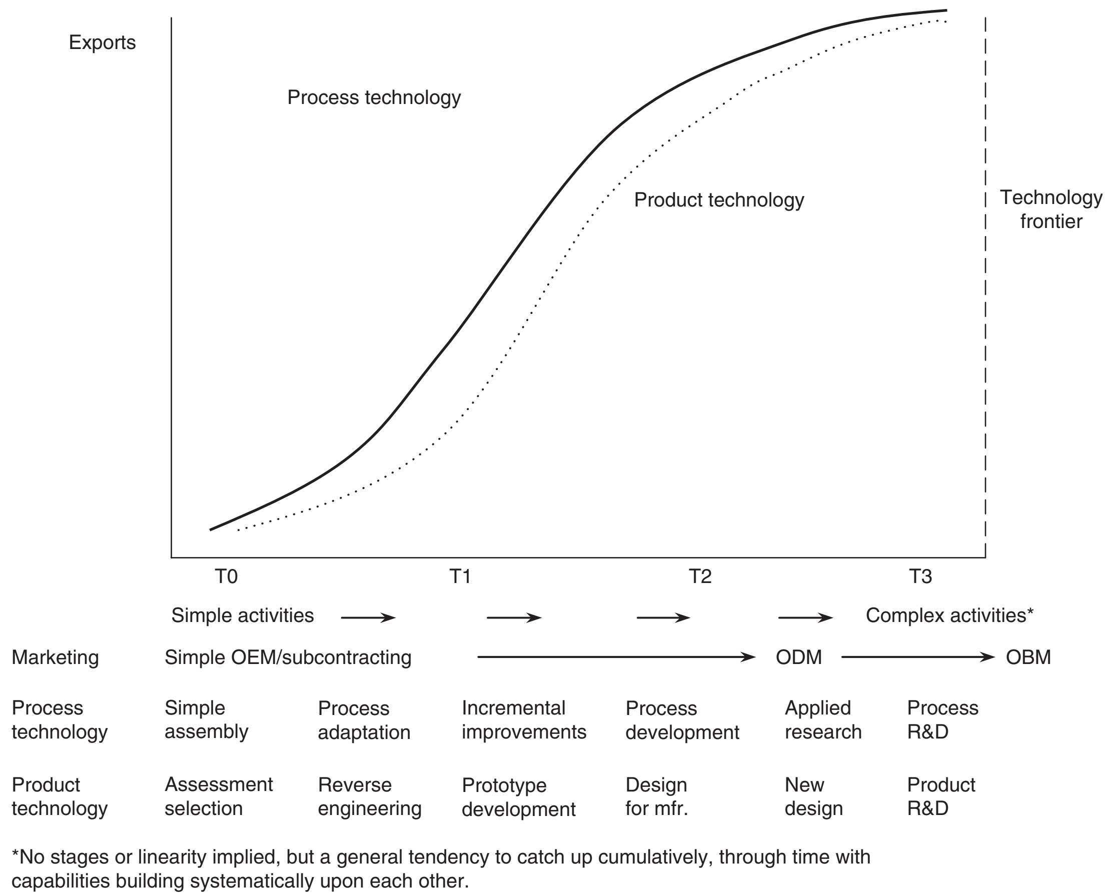

Figure 2 Product and process upgrading over time. Source: Adapted from Hobday, M. (1995). East Asian latecomer firms: Learning the technology of electronics. *World Development* 23(7), 1171 1193.

domestic market, followed by a gradual broadening and deepening of functional roles and competences through the coordination of domestic (national) chains. This is subsequently used as a springboard for penetration in foreign markets through exports. Thus, reversing the order, functional diversification preceeds exports and from own brand, manufacturing firms may diversify into OEM for third parties. It has been established that the most successful local firms in terms of functional diver sity and range of competences are often marked by never having been incorporated in global chains serving Western lead firms. Underlying are decisions taken early in the life course of the firm, in large part determining their subsequent development.

Not only is there a lack of recognition of different chain strategies of firms in developing countries, but also extra global chain processes have been viewed mostly in a negative way and extra chain possibilities for develop ment have been largely ignored. Instead, the disadvan tages of exclusion from global chains have been pointed out at length. Although not without legitimacy, it over emphasizes the link between connections to global chains and possibilities for local development.

Underdiscussed Aspects

It has been argued that the omissions observed above stem largely from an under researching of other di mensions of commodity chains besides input–output and governance structure. Not only has there been little work specifically on dynamic geographies, but also too much emphasis on within chain factors and too little on extra chain aspects, influencing configurations. First, the regulatory environment that refers to the conditions under which control over market access and information are exercised is perceived to have remained hitherto under researched. Second, it has also been argued that the local industry environment, in terms of (national) institutional arrangements, has not been duly considered, although 'national business systems' may be highly in fluential. The fact that similar global processes have produced differential outcomes at the local level under scores the importance of placing commodity chains in the wider context of the institutional environment and re gional economy in which they are embedded.

Also, the role of consumption in shaping the archi tecture of chains has remained under researched.

Research has privileged sites of production over sites of consumption. However, retailers are significant as actors holding knowledge about consumption. Trends in con sumer tastes, desires, and behaviors are in fact a key consideration and with it differential cultures and cul tural change that underlie many facets of consumption by influencing meanings attached to specific goods in a dynamic way. It has been argued that more explicit consideration of the consumption aspect would produce much richer geographies of commodity systems.

The Value of a Chain Perspective

GCC thinking has met with reservation as to the idea of sequentiality and unidirectional linearity in the con ception of how goods are moved. An alternative 'school' has emerged arguing that the morphology of relationships between production, distribution, and consumption is better depicted by the network metaphor. This metaphor perceives the connections between actors (nodes) as complex webs of interdependence, connecting firms through the exchange of commodities and binding toge ther additional agents through supporting multidirectional flows of information, knowledge, and materials. Such a conception avoids privileging one node over another. An evaluation of the merits of a network perspective rather than a chain perspective is outside the scope of this article. No matter the outcome, GCC analysis has contributed much to deepening understanding of the diverse archi tectures and implications of economic globalization pro cesses and linked international trade dynamics. These have helped to provide rationale for new development policy initiatives that will be part of the agendas of multilateral agencies for some time to come.

See also: Consumption; Firms; Global Production Networks; Globalization and Transnational Corporations; Governance, Corporate; Industrial Organization; Spatial Division of Labor.

Further Reading

Dicken, P., Kelly, P. F., Olds, K. and W Ch Yeung, H. (2001). Chains and networks, territories and scales: Towards a relational framework for analysing the global economy. *Global Networks* 1, 89 112.

Gereffi, G. and Kaplinsky, R. (eds.) (2001). *The Value of Value Chains: Spreading the Gains from Globalization*. IDS Bulletin 32, No. 3. Sussex: Institute of Development Studies.

Gereffi, G. and Korzeniewicz, M. (eds.) (1994). *Commodity chains and global capitalism*. Westport, CT: Praeger.

Gereffi, G., Humphrey, J. and Sturgeon, T. (2005). The governance of global value chains. *Review of International Political Economy* 12, 78 104.

Gibbon, P. (2001). Upgrading primary production: A global commodity chain approach. *World Development* 29, 345 363.

Hassler, M. (2003). Crisis, coincidences and strategic market behaviour: The internationalization of Indonesian clothing brand owners. *Area* 35, 241 250.

Hassler, M. (2003). The global clothing production system: Commodity chains and business networks. *Global Networks* 3, 513 531.

Henderson, J., Dicken, P., Hess, M., Coe, N. and W Ch Yeung, H. (2002). Global production networks and the analysis of economic development. *Review of International Political Economy* 9, 436 464.

Hobday, M. (1995). East Asian latecomer firms: Learning the technology of electronics. *World Development* 23(7), 1171 1193.

Hopkins, T. K. and Wallerstein, I. (1986). Commodity chains in the world economy prior to 1800. *Review* 10, 157 170.

Hughes, A. (2000). Retailers, knowledges and changing commodity networks: The case of the cut flower trade. *Geoforum* 31, 175 190.

Humphrey, J. (2005). *Shaping Value Chains for Development: Global Value Chains in Agribusiness*. Eschbor: Deutsche Gesellschaft fur Technische Zusammenarbeit (GTZ) Gmbh.

Leslie, D. and Reimer, S. (1999). Spatializing commodity chains. *Progress in Human Geography* 23, 401 420.

Schmitz, H. (ed.) (2004). *Local Enterprises in the Global Economy: Issues of Governance and Upgrading*. Cheltenham: Edward Elgar.

Sturgeon, T. (2002). Modular production networks: A new American model of industrial organization. *Industrial and Corporate Change* 11, 451 495.

Relevant Websites

http://www.globalvaluechains.org
Global Value Chains Initiative.

Global Positioning/GPS

J. Lee, The University of Seoul, Seoul, South Korea

Glossary

Datum Basis for measurement of position on the Earth's surface.

Geoid The surface mapped by all of the local variation from the ellipsoid.

Geostationary Satellite Broadcast-corrected GPS signals provided by WAAS control stations for use by aviators or civilians equipped with GPS/WAAS receivers.

Geographic Information Systems It is designed to capture, store, display, communicate, transform, analyze, and archive georeferenced information.

Location-Based Services It provides personalized services to the subscriber of the system based on their current position.

Orthometric Height The height of a point above the geoid. The height obtained by GPS is referred to the reference ellipsoid (ellipsoidal height) and should be converted to the orthometric height using geoid models built in GPS receivers.

Selective Availability To degrade the autonomous GPS accuracy for security reasons, intentionally implemented by the US DoD, but terminated on 1 May 2000.

Overview of GPS

GPS is a space based navigation system nominally consisting of a 24 satellite constellation that provides positioning and timing information to military and civilian users worldwide. GPS satellites circling the Earth every 12 hours broadcast continuous GPS signals on two different L band frequencies, L1 and L2. In addition to the GPS satellites, the system consists of a worldwide satellite control network and GPS receiver units that acquire the GPS signals and translate them into positioning and timing information. GPS provides several services including highly accurate and three dimensional location information, precision velocity and timing services, and accessibility to an unlimited number of global military, civilian, and commercial users.

The basic idea behind GPS to determine a user's location is relatively simple. If the distances from a user (a GPS receiver) to three GPS satellites and the satellite locations are known, the location of the user can be determined by applying a well known triangulation survey mapping procedure. The receiver will be located at the intersection of three spheres; each has a radius of one receiver–satellite distance and is centered on that particular satellite. Theoretically, three distances to three simultaneously tracked satellites along with their coordinates are needed. The positioning information is the partial outcome of GPS signal processing. The accuracy of positioning was limited to 100 m for the horizontal component, 156 m for the vertical component, and 340 ns for the time component under selective availability (SA) policy. With terminating the SA on 1 May 2000, the obtained horizontal accuracy is expected to improve to about 22 m (95% probability level). To further improve the GPS positioning accuracy, the GPS relative positioning methods are developed and the modernized GPS signals (L2C, L5, and L1C) are designed and will be utilized in the near future. In this case, the level of positioning accuracy will be improved to few meters or a subcentimeter.

The fact that the topographic surface of the Earth is highly irregular makes it difficult to determine the user's location. To overcome this problem, geodesists adopted a smooth mathematical surface, called the reference surface, to approximate the irregular shape of the Earth. One such surface is a biaxial reference ellipsoid, known as the geodetic datum. A geodetic datum is a mathematical surface, or a reference surface, with a well defined origin (center) and orientation.

The satellite coordinates as given in the broadcast ephemeris refers to the World Geodetic System of 1984 (WGS 84) reference system. The WGS 84 is a 3 D, Earth centered reference system developed by the former U.S. Defense Mapping Agency now incorporated into a new agency, National Imagery and Mapping Agency (NIMA). It is the official GPS reference system. In addition to the geodetic datum, the so called vertical datum is used as a reference surface to which the heights (elevations) of points are referred. The geoid is often selected to be the vertical datum. The height of a point above the geoid is known as the orthometric height, which is usually found plotted on topographic maps. However, in the case of GPS, the obtained heights are references to the reference ellipsoid, not the geoid, known as the ellipsoidal heights. Many GPS receivers and software package have built in geoid models for automatic conversion between orthometric and ellipsoidal heights.

Time plays a very important role in positioning with GPS. The GPS signal is controlled by accurate timing system devices, the atomic satellite clocks. In addition, measuring the distances from the receiver to the satellites is based on both the receiver and the satellite clocks. Two time systems, the coordinated universal time (UTC) and the GPS time, are very important to GPS users. UTC is an atomic time scale based on the international atomic time (TAI). TAI is a uniform time scale, which is com puted based on independent time scales generated by atomic clocks located at various timing laboratories throughout the world. GPS time is the time scale used for referencing the GPS signals. It is computed based on the time scales generated by the atomic clocks at the monitor stations and onboard GPS satellites. The GPS time scale was set to UTC on 6 January 1980. However, due to the leap seconds introduced into the UTC time scale, GPS time moved ahead of the UTC by 13 seconds on 1 January 1999. The difference between GPS and UTC time scale is given in the GPS navigation message.

Global Navigation Satellite System (GNSS)

The GNSS was declared by the International Civil Aviation Organization (ICAO) consisting of the satellite navigation systems including interoperable regional and local augmentation systems. Currently, one of the satel lite navigation systems is fully operational and widely available to civilian users, which is NAVSTAR GPS (Navigation System Timing and Ranging Global Pos itioning System) owned by the US government. The GPS was developed by the US Department of Defense (DoD) in the early 1970s with other technical resources such as the Defense Mapping Agency and NATO nations. The initial operational capability (IOC), a constellation of 24 operational satellites, was completed in July 1993. The GPS was officially declared to have achieved full oper ational capability (FOC) on 17 July 1995. Currently, the NAVSTAR GPS is managed for DoD by the NAVSTAR Joint Program Office (JPO) located at Los Angeles Air Force Base. The JPO oversees the development and production of GPS satellites and conducts field tests for all GPS equipment and techniques used by the US military. The civilian point of contact for NAVSTAR information is the U.S. Coast Guard's Navigation Center (NAVCEN). The NAVCEN provides system status in formation in the form of a document called a Notice Advisory to NAVSTAR Users (NANU). Civil users have input to NAVSTAR system issues through the Civil GPS Service Interface Committee (CGSIC), which is organ ized by the Coast Guard.

Other satellite navigation systems, which are partially operated or under construction, are Russian GLONASS (Global Navigation Satellite System) program and European Galileo program. GLONASS is Russia's ver sion of GPS. The first GLONASS satellite was launched in 1982, but the system remains partially operational due to political and economic issues. The system is used to augment GPS for applications using real time kinematic (RTK). As of 25 September 2006, there were ten func tional GLONASS satellites. The country intends to make GLONASS fully operational (24 satellites) by 2010. Galileo is the European counterpart to GPS. Unlike GPS and GLONASS, which are financed and controlled by their countries' respective military organizations, Galileo is controlled by civilians. The system is designed to provide five levels of services from open (simple pos itioning available to everyone) to public regulated (en crypted, integrity, continuous availability) services. In December 2005, the first Galileo test satellite (GIOVE A) was launched, and two further experimental satellites are planned by the second half of 2008. The full con stellation (27 satellites plus three spares) will be deployed between 2007 and 2010, according to a report by the European Space Agency.

Perhaps, the greatest advantage of Galileo will be its interoperability with NAVSTAR GPS and GLONASS. The improved geometry of having an average of 20 sat ellites in view (more than twice than what we have) will improve the level of accuracy, reliability, and availability. These satellite systems are important components to develop and implement the GNSS. The GNSS improves efficiency for all users, and benefits all modes of trans portation as well as multiple nontransportation related applications. As the number of GNSS increases, GNSS is slowly but most certainly replacing GPS as the acronym used when discussing satellite navigation.

GPS Segments

The NAVSTAR GPS consists of three major segments: space segment, control segment, and user segment. The current GPS constellation (as of May 2006) consists of 29 satellites: one Block II satellite, fifteen Block IIA satellites, twelve Block IIR satellites, and one Block IIR M satellite. This is the nominal 24 satellite constellation and five spare satellites. The first Block II satellite was launched in February 1989; the most recent Block IIR M satellite (**Figure 1**) was launched on 26 September 2005, which is the first modernized GPS satellite. The GPS satellites are placed in six orbits labeled A through F, which are inclined by 55^{o} with respect to the equator. GPS satellite orbits are nearly circular (an elliptical shape with a maximum eccentricity of about 0.01), and the semimajor axis is about 26 560 km above the Earth's surface. The corresponding GPS orbital period is about 12 sidereal hours. Since more satellites are currently available than the nominal 24 satellite constellation, an orbital plane contains four or five satellites (**Figure 2**). The satellites are identified by the space vehicle number (SVN) and the pseudorandom noise (PRN).

Figure 1 GPS Block IIR-M satellite. Source: http://pnt.gov/public/images/IIR-M 2.jpg

Figure 2 Navstar constellation: GPS Block IIA, IIR/M, IIF. Source: http://www.defenseindustrydaily.com/images/SPAC GPS NAVSTAR IIA IIR IIF Constellation lg.gif

Each GPS satellite (except Block IIR M) transmits a navigation signal including two carrier frequencies (L1 and L2), two digital codes (P code and C/A code), and a navigation message. The L1 frequency is for either mili tary or civilian users and the precise L2 frequency is for military only. The Block IIR M satellite is designed to broadcast two new military signals and a second civil signal (L2C). The carriers carry the codes and the navi gation message as binary biphase modulations. The car riers and the codes are used to calculate the distance from the user's receiver to the GPS satellites, while the navi gation message contains the location of the satellites as a function of time along with other information. The transmitted signals are controlled by highly accurate atomic clocks onboard the satellites. The three dimen sional location of the receiver can be determined by the calculated distances and the positions of satellites at the moment of transmission.

The control segment of the GPS consists of a worldwide network of tracking stations with a master control center (MCS) and five ground control stations. The MCS, located in Colorado Springs, CO, USA, is the central processing facility of the control segment and is manned at all times. The five monitor stations are located in Colorado Springs, Hawaii, Kwajalein, Diego Garcia, and Ascension Island. The primary task of the oper ational control segment is tracking the GPS satellites in order to determine and predict satellite locations, system integrity, behavior of the satellite atomic clocks, atmos pheric data, the satellite almanac, and other consider ations. The information is packed and uploaded into the GPS satellites through the S band link from three of the monitor stations (Kwajalein, Diego Garcia, and Ascension Island). All of the monitor stations and the ground con trol stations are unmanned and operated remotely from the MCS. Monitoring the GPS integrity is also one of the tasks of the MCS. The status of a satellite is set to an unhealthy condition by the MCS during satellite main tenance. This satellite health condition appears as a part of the satellite navigation message on a near real time basis. Scheduled satellite maintenance is reported in a message called NANU, which is available to the public through the US NAVCEN.

The user segment includes all military and civilian users. With a GPS receiver connected to a GPS antenna, a user can receive the GPS signals, which can be used to determine a user's position anywhere in the world. The NAVSTAR GPS is currently available to all users worldwide at no direct cost, while the Galileo system will provide two levels of services: free direct charge service and chargeable service that offers additional features.

In 1980, only one commercial GPS receiver was available on the market, at a price of several hundred thousand US dollars. Recently, more than 500 different GPS receivers are available and the price varies from about US$100 for the simple handheld units to about US$15 000 for the sophisticated survey purpose units. Commercial GPS receivers may be divided into four types, according to their receiving capabilities: single frequency code receivers, single frequency carrier smoothed code receivers, single frequency code and carrier receivers, and dual frequency receivers. Single frequency receivers access the L1 frequency only, while dual frequency receivers access both the L1 and the L2 frequencies. GPS receivers can also be categorized based on their number of tracking channels. Presently most GPS receivers have 9–12 independent (or parallel) channels, with each channel dedicated to continuously tracking a particular satellite. When selecting a GPS receiver, many features should be considered: cost, ease of use, power consumption, size and weight, internal and/or external data storage capabilities, interfacing capabilities, and multipath mitigation.

GPS Position and Travel Time Measurements

As a dual use positioning system, the US DoD provides two levels of GPS positioning and timing services: the precise positioning service (PPS) and the standard positioning service (SPS). PPS is the most precise autonomous positioning and timing service. It uses precision code (P code), which is accessible by authorized users, including military forces. The expected positioning accuracy provided by PPS is 16 m for the horizontal component and 23 m for the vertical component (95% probability level). SPS, which is less precise than PPS, uses coarse acquisition code (C/A code), which is available free of charge to all users worldwide. After discontinuing the SA, the SPS autonomous positioning accuracy is presently at a comparable level to that of the PPS.

The GPS codes, C/A code and P code, are carried by two carrier frequencies (or sine waves), which are the L1 carrier frequency generated at 1575.42 MHz and the L2 carrier frequency generated at 1227.60 MHz. All GPS satellites transmit the same L1 and L2 carrier frequencies. The GPS navigation message is a data stream added to both the L1 and L2 carrier as binary biphase modulation. The navigation message contains, along with other information, the coordinates of the GPS satellites as a function of time, the satellite health status, the satellite clock correction, the satellite almanac, and atmospheric data. Each satellite transmits its own navigation message with information on the other satellites, such as the approximate location and health status.

The current GPS signals structure was designed in the early 1970s. To meet the future requirements, a GPS modernization program has been undertaken to provide signal redundancy and to improve positioning accuracy, signal availability, and system integrity. The modernization program includes a new GPS civil signal known as L2C on the L2 frequency, the third frequency (1176 MHz) primarily for the civil aviation community, another proposed civil frequency (L1C) for the survey/mapping community, and two new military codes (M codes) on both the L1 and the L2 frequencies. Two major benefits of the modernized GPS signals are to enhance the ionospheric correction and to track better in tough GPS conditions (such as under tree foliage). The Block IIR M and IIF satellites are designed to broadcast the new GPS civil signal (L2C), and the first IIR M satellite was launched in September 2005. Eight IIR M and ten IIF satellites are scheduled for creation and launch by 2010. The Block IIF satellites are also designed to broadcast the third GPS civil frequency known as L5, which will be available by 2014. GPS III is a long range program expected to satisfy civil and military requirements through the year 2030. The modernized GPS signals are very similar to the proposed Galileo signals (L1, E5a, and E5b). The system interoperability will improve the level of accuracy and signal availability.

In order to determine the GPS receiver's position, as discussed above, the distances from the receiver to three simultaneously tracked satellites and the satellites' coordinates are needed. Because the coordinates of GPS satellites can be obtained from the navigation message transmitted by GPS satellites, the distances between the satellites and the receiver need to be measured by using two measurement methods: a pseudorange measurement and a carrier–phase measurement. The pseudorange is a measure of the range, or distance, between the GPS receiver's antenna and the GPS satellite's antenna, using either the P code or the C/A code. When the PRN code is transmitted from the satellite, the receiver generates an exact replica of that code at the same time. After some time, the transmitted code will be picked up by the receiver. By comparing the transmitted code and its replica, the receiver can calculate the signal travel time in space. Multiplying the travel time by the velocity of the signal gives the distance between the satellite and the receiver. The velocity of the signal is equal to the speed of light (299 729 458 m s ^{1}). Another way of measuring the range to the satellite is a carrier–phase measurement using either the L1 or the L2 frequencies. The range would simply be the sum of the total number of full carrier cycles plus fractional cycles at the receiver and satellite, multiplied by the carrier wavelength. The ranges determined with the carriers are far more accurate than those obtained with the pseudoranges. From the measurements three spheres can be obtained: each has a radius of a receiver–satellite range and centered on that particular satellite. The receiver will be located where three spheres intersect (**Figure 3**).

GPS Error Sources

Both GPS pseudorange and carrier–phase measurements are affected by several types of systematic or random errors and biases. The errors may be classified as errors originating at the GPS satellites, errors originating at the GPS receiver, and errors due to signal propagation (atmospheric refraction). In addition, the accuracy of the computed GPS position is affected by the geometric distribution of the GPS satellites in view.

The errors from the satellites include GPS ephemeris errors, satellite clock errors, and the effect of SA. The SA

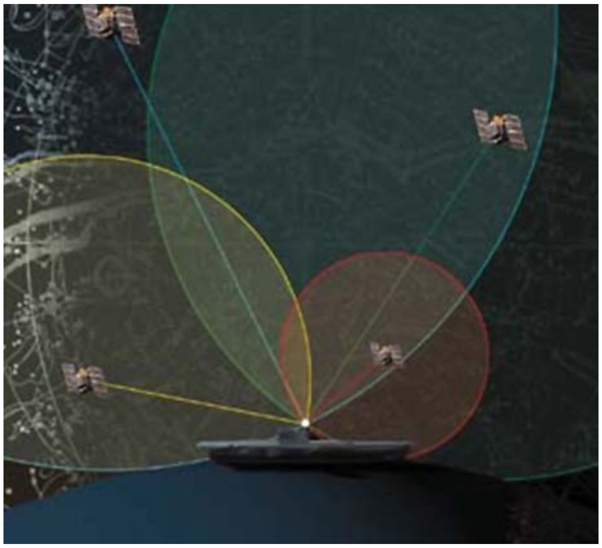

Figure 3 Determining position by three spheres intersect. Source: http://www.nasm.si.edu/exhibitions/gps/work.html Copyright Owner: Smithsonian Institution, http://photo2.si.edu/copyright.html

was intentionally implemented by the US DoD to de grade the autonomous GPS accuracy for security rea sons, but was terminated on 1 May 2000. The ephemeris errors are involved in the estimated satellite positions, because the satellite positions as a function of time are predicted from previous GPS observations at the ground control stations. The error is usually on the order of 2–5 m, and can be corrected in the postprocessing mode. Each GPS Block II and IIA satellite contains cesium and rubidium atomic clocks. The Block IIR satellites carry rubidium clocks to provide the frequency and the timing requirements for generating the GPS signals. The GPS satellite clocks, although highly accurate, are not perfect. The clock errors cause additional errors to the GPS measurement because of the assumption that the receiver and satellite clocks are synchronized with each other. GPS receivers use inexpensive crystal clocks, which are much less accurate than the satellite clocks.

The errors originating at the receiver include receiver clock errors, multipath error, receiver noise, and antenna phase center variations. Multipath is a major error source for both the carrier–phase and pseudorange measure ments. Multipath error occurs when the same GPS signal from a satellite arrives at the receiver antenna several times as a direct signal and reflected signals through different paths. Multipath distorts the original signal through interference with the reflected signals at the GPS antenna. The error in the pseudorange measure ment is much larger than that in the carrier–phase. With new advances in multipath mitigation techniques, the pseudorange multipath error is reduced to several meters or submeters, even in a highly reflective environment.

The signal propagation errors include the delays of the GPS signal as it passes through the ionospheric and tropospheric layers of the atmosphere. It is assumed that the GPS signal travels at the speed of light in a vacuum (free space). The ionosphere is a dispersive medium, which bends the GPS signal and changes its speed as it passes through the various ionospheric layers to reach a GPS receiver. The inability of L2 only receivers to handle the ionospheric propagation delays is one of the greatest sources of error in GPS positioning. The new modernized GPS signals (including L2C and L5) are designed to correct the ionospheric errors accurately. The tropospheric delay depends on the temperature, pressure, and humidity along the signal path through the troposphere. The delay results in positioning errors on the order of 2.3–28 m. Several mathematical models are used to predict surface meteorological measurements to reduce the tropospheric error.

The last error source affecting the positioning accuracy is based on the satellite geometry, which represents the geo metric locations of the GPS satellites as seen by the receiver. Good satellite geometry is obtained when the satellites are spread out in the sky. In order to examine the effect of the satellite geometry on the quality of the measured 3 D position of a receiver, the value of the position dilution of precision (PDOP) can be measured. The PDOP represents the contribution of the satellite geometry to the 3 D pos itioning accuracy. To ensure high precision GPS pos itioning, a suitable observation time should be selected to obtain the highest possible accuracy using GPS mission planning software. A PDOP of five or less is usually recommended.

GPS Data and Correction Services and Techniques

Positioning with GPS can be performed in two ways: point positioning or relative positioning. GPS point positioning employs one GPS receiver that measures the pseudor anges to determine the user's position instantaneously, as long as at least four satellites are in view. Although the positioning mode is used mainly with relatively low ac curacy required, GPS decision makers and scientists have developed technologies to improve the accuracy of GPS in many ways: (1) modifying the GPS signal structures (L2C, L5, and L1C), (2) performing its interoperability with GLONASS or Galileo system, and (3) developing ad vanced techniques in GPS receiver to reduce clock errors and multipath errors.

GPS relative positioning, called differential positioning, employs two GPS receivers simultaneously tracking the same satellites to determine their relative coordinates. One receiver is selected as a base receiver, which remains sta tionary at a site with precisely known coordinates. The other receiver, called a rover receiver, is set up over a point for which coordinates are sought (referred to as the un known point). In principle, because the measurements of two receivers simultaneously tracking a particular satellite contain more or less the same errors and biases, if we take the difference between the measurements of the two re ceivers (called differential positioning), the similar error will be removed or reduced. Carrier–phase and/or pseudorange measurements can be used in relative positioning. A variety of the relative positioning techniques are used to provide a post processing or real time solution, which include real time kinematic (RTK) GPSs, state wide multisite base station network systems, differential GPS (DGPS) and the wide area augmentation system (WAAS). The RTK sur veying is a carrier phase based relative positioning techni que, and is suitable for determining the coordinates of the rover's location within about 5–10 km from the base receiver in real time. As a survey purpose system, the ex pected positioning accuracy is on the order of 2–5 cm. The RTK with a single reference station is limited to the rela tively small area surrounding the reference station. In order to overcome this problem, a system based on a number of widely separated reference stations has been developed at state and federal levels in the United States. Most state governments (departments of transportation) have developed a state wide multisite base station network sys tem. The RTK network systems determine the differential errors between the reference stations within the network. Once the rover receiver provides its approximate position to the control station, the raw measurements of the rover's position are calculated based on the differential errors. The error corrections can be processed by post processing or real time modes.

The US government has designed and implemented two different augmentation systems to improve the pre cision of C/A code, which are the DGPS radio beacon system and the WAAS. The DGPS developed and managed by the NAVCEN is a code based relative pos itioning technique. The Coast Guard DGPS designed with the surface (maritime and terrestrial) user is fully operational and meets all the standards for the harbor entrance and approach phases of navigation. In the DGPS, a reference station creates the real time DGPS corrections in a standard format called Radio Technical Commission for Maritime Service (RTCM). These cor rections are digitally modulated using minimum shift keying (MSK), and transmitted from the radio beacon. A user equipped with an MSK beacon receiver can receive the transmitted DGPS corrections within the coverage area of a particular beacon station. DGPS is designed to provide better than 10 m navigation service (95%), but typically provides better than 1 m horizontal positioning accuracy (95%) when the user is less than 100 nautical miles from the DGPS transmitting site. Accuracy then degrades at a rate of approximately 1 m per hundred nautical miles as the user moves away from the trans mitting site. A total of 56 maritime DGPS sites provide coastal coverage of the continental United States, the Great Lakes, Puerto Rico, portions of Alaska and Hawaii, and portions of the Mississippi River Basin.

WAAS developed by the Federal Aviation Adminis tration (FAA) provides augmentation information to GPS receivers to enhance the accuracy and reliability of position estimates for civil aviation. The signals from GPS satellites are received at 25 widely spaced wide area reference sta tions (WRS) sites. The WRS locations are precisely sur veyed so that any errors in the received GPS signals can be detected. The GPS information collected by the WRS sites is forwarded to the WAAS master station (WMS) via a terrestrial communications network. At the WMS, the WAAS augmentation messages are generated. These mes sages contain information that allows GPS receivers to re move errors in the GPS signal. The differential message is broadcast from two geostationary satellites. The GPS/ WAAS receiver processes the WAAS augmentation message to calculate the user's position by assessing in factoring the signal delay caused by various error sources. WAAS is currently available only in the United States. On the average, WAAS and DGPS accuracy are virtually the same. The WAAS architecture is designed to provide uniform 7 m accuracy (95%) regardless of the location of the receiver – within the WAAS service area.

GPS Applications

GPS has been available for civil and military uses since 1995. Especially, after terminating the SA in 2000, nu merous new GPS applications have been developed in many areas. Because GPS provides high accuracy pos itioning, global availability, and cost effectiveness, the information technology has found its way into many applications replacing conventional methods in most cases, including surveying and mapping, land, sea, and air navigation, search and rescue operations, spacecraft op erations, military applications, recreational uses, and other specialized uses. The GPS also has been used a means of gathering data on human activity data for social science research. With the positioning and tracking abilities of GPS and geocomputational capabilities of geographic information systems (GIS), it is now more feasible than ever before to study human activities and movements in space and time. It covers a wide range of topics such as migration, residential mobility, shopping, travel, and commuting behavior based on Torsten Hagerstrand's time geography model. The time geo graphic perspective has inspired generations of social scientists, especially geographers and transportation re searchers, in the description and analysis of human ac tivities in space–time.

In addition, these locational devices are equipped with mobile GIS software (such as ArcPad) and can generate on screen georeferenced maps to support users' operations onsite (e.g., providing navigational guidance). The GPS/GIS system provides a cost effective, efficient, and accurate tool for creating utility maps, for managing forestry and natural resources, for helping retail business in delivery and real time inventory monitoring, and for providing vehicle drivers with route guidance. Using vehicle navigation sys tems (VNS) seen in **Figure 4**, the GPS determined vehicle location is superimposed on an electronic digital road map containing in its geographic information such as street names and directions, business listings, attraction points, and other relevant information. Once the drive inputs a des tination, the VNS finds the best route to reach the destin ation with considerations of traffic impedance factors such as shortest distance or travel time distance, one way roads, turns, traffic volumes, and rush hour restrictions. The driver usually gets directional information to the destination with audio and/or visual indications. Some manufacturers add cellular systems to provide weather and traffic information and to locate the vehicles in case of emergency. Recent advances in wireless communication make it possible for drivers to remotely access the Internet from their vehicles, and for control centers to track and monitor criminal of fenders or vehicles such as rental cars, police cars, or utility vehicles. Recently, nearly all new cellular phones, called GPS phones have some GPS receiving capability built in. This allows the phones to transmit a user's location to a public safety answering point (PSAP) after dialing 911 ac cording to the United States Federal Communications Commission's (FCC) Enhanced 911 (E911) program. Also, the GPS phones are used for providing location and tracking services including vehicle or employee's location tracking, turn by turn direction services, location based services (LBS) for recreations and business activities (**Figure 5**).

Issues

Although GPS provides high accuracy positioning in a cost effective manner, it requires, at all times, a direct line of vision between the receiver's antenna and the satellites. Otherwise, the receiver cannot pick up the signal. The signal loss is caused by obstruction of the GPS satellite signal due to buildings, bridges, and other objects. The signal loss must be identified and corrected to avoid large errors in computing coordinates. This signal ob struction problem was successfully overcome by integrating GPS with other positioning systems. The integrated system is better than either system alone. GPS/LRF (laser range finders) integration was developed to overcome the signal loss in areas with heavy tree canopy, and the integrated GPS/DR (dead reckoning) system is widely used in auto matic vehicle location (AVL) applications. In order to ob tain high accuracy positioning in obstructed areas and/or under dynamic conditions, GPS can be integrated with a relatively environment independent system, the inertial navigation system (INS).

Although the temporal GPS signal loss problems were resolved by integrating with other positioning system as mentioned above, GPS is not able to provide positioning information inside buildings or built environments. It is still a challenging problem in most indoor LBS in urban areas because 70% of LBS are being initiated indoors where GPS

Figure 4 Vehicle navigation systems. Source: http://news.cnet.com/i/bto/20070417/AVIC-S2 Map 4in 300dpi 270x241.jpg Copyright Owner: Pioneer Electronics Inc.

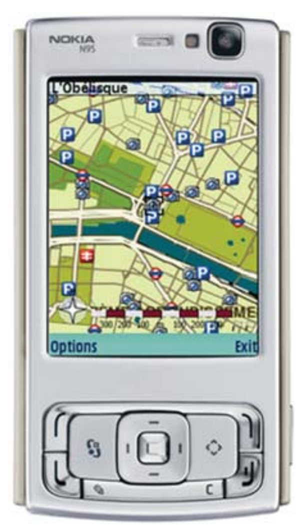

Figure 5 GPS phone: Nokia N95. Source: http://www.smh.com.au/ffximage/2007/05/03/n95 high gps 05 narrowweb 300x496,0.jpg Copyright Owner: Nokia Inc.

fails to deliver an acceptable end user performance. In re cent years, 3 D indoor positioning techniques have been developed to improve the quality and quantity of these data and to reduce their cost using location aware devices such as cellular phones, wireless PDAs, GPS receivers, and radiolocation methods. The positioning within mobile networks using only the information related to the base network transmitter is a very cost effective method, but it is very inaccurate and not practically applicable for LBS. The only advantage, compared to GPS positioning, is the pos sibility of working inside buildings. None of techniques increases the accuracy by more than 50 m. An Indoor GPS or Assisted GPS (A GPS) has been used to detect highly attenuated GPS signals indoors. A GPS, however, requires the use of modified handsets that receive the GPS signals and then send those readings to a network server. Wireless LANs (WLANs), radio frequency identification (RFID), WiFi, and Bluetooth are used to track mobile users in closed spaces such as buildings and tunnels. These systems for location positioning run on a standalone server and give the *x*, *y* position and the floor of the mobile unit. Despite providing accurate positioning for indoors, however, the positioning technologies have problems with implemen tation because they require a reference database for average signal measurements at fixed points throughout a building. Although there are currently developments in the LBS that seeks to provide better coverage of location information inside buildings, much further research is still needed to address this problem.

The positioning technologies including GPS and others prove to be a great enhancement in civil or military applications, but the technology has recently created privacy concerns among civilian applications. General privacy concerns relate to the collection, storage, and release of personal information – the nature and level of detailed data and the accessibility to stored data. In order to deal with the geosurveillance challenges, both technological and legal solutions should be proposed to prevent access to the data by unauthorized users, and to protect individual human rights and personal privacy.

See also: GIScience and Systems; Remote Sensing; Spatial Databases; Time Geography.

Further Reading

El Rabbany, A. (2002). *Introduction to GPS: The Global Positioning System*. Boston, MA: Artech House.

Gakstatter, E. and Flick, J., (2006). Navigating the world of GNSS. *Geospatial Solutions* Apr/06, 22 25.

Goodchild, M. F. and Janelle, D. G. (eds.) (2004). *Spatially Integrated Social Science*. New York: Oxford Press.

Kennedy, M. (2002). *The Global Positioning System and GIS: An Introduction* (2nd edn.). New York: Taylor & Francis.

Kwan, M. P. and Lee, J. (2005). Emergency response after 9/11: The potential of real time 3D GIS for quick emergency response in micro spatial environments. *Computers, Environment and Urban Systems* 29, 93 113.

Leick, A. (2004). *GPS Satellite Surveying* (3rd edn.). New Jersey: John Wiley.

Tsui, J. B. Y. (2000). *Fundamentals of Global Positioning System Receivers A Software Approach*. New York: John Wiley.

Relevant Websites

http://www.esa.int
European Space Agency home page (Galileo system).

http://gps.faa.gov
Federal Aviation Administration Satellite Navigation home page (WASS).

http://gps.losangeles.af.mil
NAVSTAR Global Positioning System Joint Program Office Home page.

http://www.glonass ianc.rsa.ru
Russian Space Force Home Page (GLONASS system).

http://www.navcen.uscg.gov
US Coast Guard Navigation Center Home Page.

Global Production Networks

N. M. Coe, The University of Manchester, Manchester, UK

Introduction

As economic globalization has accelerated, a consider able body of literature has evolved that attempts to ex plain how global industries are organized into systems of chains or networks, and how, in turn, the nature of those systems affects the opportunities for economic develop ment and upgrading for the various regions and firms involved. One increasingly important strand of this work concerns the geographical analysis of global production networks (GPNs), using a framework developed pri marily by researchers in Manchester and their collabor ators. In short, GPN analysis seeks to reveal the multiactor and multiscalar characteristics of transnational production systems – and their developmental impli cations – through exploring the intersecting notions of power, value, and embeddedness.

This article seeks to provide an initial synopsis and evaluation of this emerging and fast developing area of research. It will proceed in four sections: first, the theo retical background and context for the GPN approach are reviewed; second, the main conceptual building blocks of GPN analysis are introduced; third, the im plications of the framework for understanding (subna tional) regional development are considered; and fourth, the most significant challenges facing ongoing research in this area are profiled.

Theoretical Antecedents

Before introducing the notion of GPN in more detail, it is important to situate the approach within the wider literatures on economic globalization. Over the past 20 years or so, a plethora of research avenues have emerged using notions of 'chains' or 'networks' to conceptualize production systems. Each of these approaches emphasizes the interconnected structures of different economic ac tivities, with each link or element in the chain adding value in some way to the production process under consideration. Some of this work has emerged from the business/managerial literature, such as that on *value chains*, and some from the sociological literature, such as that exploring the *social embeddedness* of economic net works. The GPN approach has particularly developed, however, from the critical evaluation of two network/ chain lines of enquiry (see **Table 1**).

First, emerging from the development studies tradi tion and world systems theory is the notion of the *global commodity chain* (GCC). These have been defined as interorganizational networks linking households, firms, and states together across the global economy. Each GCC has four interrelated dimensions: an input–output struc ture, a territoriality, a governance regime, and an insti tutional framework. While an impressive body of work has been produced since the concept's initial formulation, several clear limitations have become apparent. Despite identifying four dimensions analytically, in reality gov ernance structures – and in particular rather crude dis tinctions between producer and buyer driven chains – have dominated discussion and research in the field. Moreover, in terms of exploring the territoriality or geography of commodity chains GCC analyses have, until recently at least, tended to focus on the national scale with less attention paid to the processes of local and regional economic development *within* countries. The approach has also proven to be relatively weak at re vealing the ways in which firm networks link together, and are constitutive of societies with different social and institutional frameworks, and relatedly, there has been little consideration of firm ownership patterns, and in particular how firms of different nationalities may adopt varying organizational structures and strategies.

More recently, the notion of the GCC has been partly superseded by the work that conceptualises *global value chains* (GVCs). GVC research has made significant ad vances with respect to two of the criticisms of GCC re search raised above. Through looking at supplier capabilities and different forms of knowledge within value chains, a more sophisticated typology of govern ance regimes has been developed, with captive, modular, and relational forms being identified alongside traditional market and hierarchy forms. Equally, GVC studies have started to profile the different ways in which subnational clusters of economic activity are 'plugged into' GVCs.

Second, emerging from investigations into the soci ology of scientific knowledge, *actor network theory* (ANT) is an approach that emphasizes how entities in networks are shaped by, and can only be understood through, their relationships to other entities. A key implication for study of transnational production systems is that space and distance have to be seen not in absolute terms, but as interrelations of influence, power, and connectivity. ANT also involves a rejection of traditional binaries such as structure agency, and recognizes that nonhuman forms (especially technologies) are an integral component of networks. Through its conceptualization of social systems as complex networks of power relations with emergent

Table 1 Theoretical antecedents to the GPN approach

	Global commodity/value chains (GCC/GVC)	*Actor-network theory (ANT)*	*Varieties of capitalism (VoC)*	*Global production networks (GPNs)*
Disciplinary background	Economic sociology Development studies	Sociology of science	Political science	Economic geography
Object of enquiry	Interfirm networks in global industries	Heterogeneous networks of association between human and nonhuman actors	Variations in national institutions and systems of capitalism	Global network configurations and regional development
Orienting concepts	Value-adding chains Governance models Organizational learning	Relational networks Human and nonhuman actors	Organization of production regimes	Value creation, enhancement, and capture
	Industrial upgrading and rents	Immutable mobiles Topological surfaces	Market-related institutions Modes of coordination	Corporate, collective, and institutional power Societal, network, and territorial embeddedness
Intellectual influences	MNC literature International business Trade economics	Sociology of science Post-structuralism	Classical political economy Institutionalism	Relational economic geography GCC/GVC, ANT, VoC

properties, ANT has a lot to offer the study of GPNs in both theoretical and methodological terms. However, the utility of ANT to the study of global economic networks is somewhat constrained by its downplaying of the structural preconditions and power relations – that is, power 'beyond' the network – that in effect provide the 'rules of the game' for the formation and operation of production networks.

In developing a GPN perspective, approaches looking at production chains and networks need to be augmented with insights from work that emphasizes the particularities of what are variously called 'varieties of capitalism', 'national business systems', or 'welfare state regimes' (see **Table 1**). The argument here is that different societies exhibit significant social and institutional variation – due to differences in culture, polity, education system, financial regulations, and so on – and therefore embody different production and welfare regimes, labor market practices, and capacities for state economic management. These variations will leave a distinctive imprint on the elements of production networks that originate from, and/or are located in, particular national territories. The downsides of such an approach include an inclination toward developing rather crude typologies of national formations on the one hand, and a tendency to neglect intranational variation in political, institutional, and economic formations on the other.

Ultimately, however, when undertaking research into GPN, it is necessary to steer a delicate path between overemphasizing the transformative effects of the networks in the various economies where they are situated, and overstressing the extent to which national conditions shape their operations in particular countries. Instead, the aim should be to explore the *mutual transformation* of both the firms and the places in which they are embedded. Let us now move on to look in more detail at how the GPN approach conceptualizes these complex interactions.

Outlining the GPN Approach

Building upon elements of the cognate approaches introduced above, the GPN framework emphasizes the complex intra , inter and extrafirm networks involved in any economic activity, and how these are structured both organizationally and geographically. A GPN can thus be broadly defined as the globally organized nexus of interconnected functions and operations of firms and nonfirm institutions through which goods and services are produced, distributed, and consumed.

The GPN approach can usefully be distinguished from its antecedents in five respects:

- First, through the explicit consideration of *extrafirm networks*, it necessarily brings into view the broad range of nonfirm organizations – for example, supranational organizations, government agencies, trade unions, employer associations, NGOs, and consumer groups – that will (or may) shape firm activities in the particular locations absorbed into GPNs.
- Second, GPN analysis is innately *multiscalar*, and considers the interactions and mutual constitution of all spatial scales from the local to the global.

- Third, this is an avowedly *network* approach that seeks to move beyond the analytical limitations of the 'chain' notion. Production systems are seen as networked 'meshes' of intersecting vertical and horizontal connections in order to avoid deterministic linear interpretations of how production systems operate and generate value.
- Fourth, the *governance* characteristics of GPNs are taken to be much more complex, contingent, and variable over time than is suggested in GCC/GVC analysis in particular.
- Fifth, and finally, a central concern of GPN analysis is not to consider the networks in an abstracted manner for their own sake, but to reveal the dynamic *developmental impacts* that result for both the firms and territories that they interconnect.

The operationalization of the framework depends on the analysis of three interrelated variables (see **Figure 1**). First, processes of *value creation, enhancement, and capture* are scrutinized. Second, the distribution and operation of *power* of different forms within GPNs is considered. Third, the *embeddedness* of GPNs – or how they constitute and are reconstituted by the economic, social, and political arrangements of the places they inhabit – is investigated. The consideration of embeddedness is a distinctive characteristic of the GPN perspective, and is reflective of a desire to highlight the important sociocultural and institutional contexts of all economic activities. Before going any further, it is important to consider the operationalization of these key variables in more detail.

First, the term value is used to refer to the various forms of economic rent that can be realized through market, as well as nonmarket transactions within GPNs. Rent is created in a situation where a firm has access to scarce resources that can insulate them from competition by creating barriers to entry for competitor firms. Firms in GPNs may be able to generate rents in a number of ways: from asymmetric access to key product and process technologies (*technological* rents), from the particular talents of their labor force (*human resource* rents), from particular organizational skills such as 'just in time' production techniques (*organizational* rents), from various interfirm relationships that may involve the management of production linkages with other firms (*relational* rents), or from establishing brand name prominence in major markets (*brand* rents). In certain sectors and circumstances additional 'exogenous' rents may accrue to some firms as a consequence of preferential access to natural resources (*resource* rents), the impacts of government policies (*policy* rents), the uneven availability of infrastructure (*infrastructure* rents), and the nature of the financial system (*financial* rents).

At least three implications flow from this broad interpretation of value and its realization. First, while rents provide a source of sustainable income, they require considerable investment over time, meaning that they are both cumulative and dynamic. Second, particular firms will not be able to create all of these different forms of rent, but rather will specialize to a degree on certain kinds of inter and extrafirm network relations. Third, value may take on different forms as it is transferred through GPNs. Once the value characteristics of a given GPN have been identified, it is then possible to think dynamically about the potential for *value enhancement* through firm to firm processes of knowledge and technology transfer, and industrial upgrading. Ultimately, it is important to consider which firms, and which locations, in the GPN succeed in *capturing value*, which leads directly to the second key GPN variable, power.

Power in a production network context can be thought of as the ability of one actor to affect the behavior of

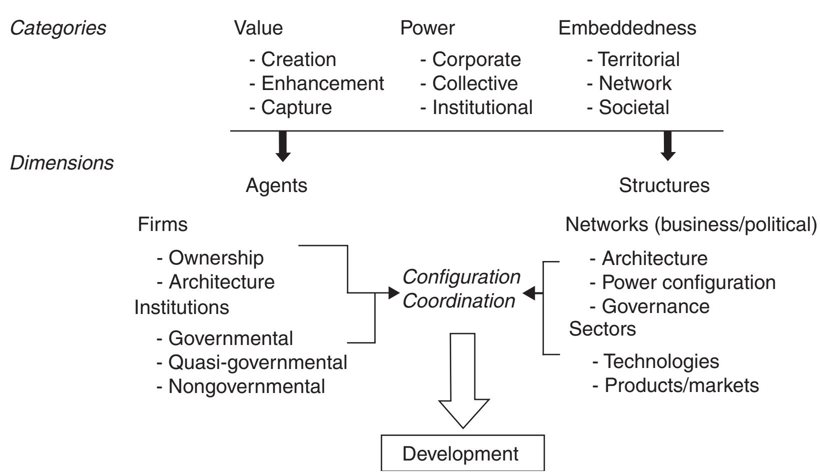

Figure 1 The global production network framework a schematic representation. Source: Henderson, J., Dicken, P., Hess, M., Coe, N. M. and Yeung, H. W.-C. (2002). Global production networks and the analysis of economic development. *Review of International Political Economy* 9, 436 464. Figure 1. Reproduced with permission of Taylor and Francis Ltd.

another actor in a manner contrary to the second actor's interests. This interpretation of power is based upon three assumptions. First, that power is seen as *relational*, meaning that it is not a commodity that can be accrued like money or land, but rather it varies according to the actors involved in the network, the structural and informational resources (i.e., rents) that they have at their disposal, and the skill with which they are mobilized. Second, and relatedly, power structures at a given point in a network will influence and be influenced by power structures at other stages of the network. Power relations in supply networks are therefore *transaction specific*. A GPN can be seen as a series of exchange relationships, and variations in the power balance along the network will affect the ability of its members to capture value. Third, any given interfirm relation cannot be purely about power as there is always a measure of *mutual interest* and dependency involved. While the relationships among participants are rarely symmetrical, firms in a production network to some degree depend on each other and work together for mutual benefit. These understandings of *corporate* power need to be augmented, however, by the recognition of other sources of power relations – most notably the *institutional* power held by the state and various suprastate organizations, and the *collective* power of labor unions, trade associations, nongovernmental organizations (NGOs), and the like – that may also shape the structure and nature of GPNs.

Third, three specific yet interrelated forms of embeddedness are employed within the GPN framework. *Societal* embeddedness connotes the importance for economic action of the cultural, institutional, and historical origins of the economic actor in question. As such, it is the one that is most frequently mobilized for explanatory purposes in the varieties of capitalism literatures introduced earlier. For example, when a company invests overseas it takes with it some of the social and cultural attributes that it has acquired in the process of its evolution within the context of its home base. These can include attitudes toward labor management relations, working conditions and welfare benefits, how supplier networks should be organized, and the appropriate role for host country governments in the business environment. *Network* embeddedness refers to the network structure, the degree of functional and social connectivity within a GPN, the stability of its agents' relations, and the importance of the network for its participants. In addition to interfirm relations, network embeddedness also takes account of the broader institutional networks including nonbusiness agents (e.g., government and nongovernment organizations such as trade unions) that are often involved. It highlights the connections between heterogeneous actors that constitute a GPN (i.e., organizations and individuals), regardless of their location, and is therefore not restricted to one geographic scale.

Territorial embeddedness captures how firms and institutions are 'anchored' in different places. GPNs do not merely locate in particular places; they may become embedded there in the sense that they absorb, and in some cases become constrained by, the economic activities and social dynamics that already exist in those places. This 'anchoring' will reflect a firm's dependence on the particular resources, labor markets, state policies, and so on found in particular places. A key element of territorial embeddedness is the extent and nature of the relationships formed between transnational corporations and local firms. These forms of embeddedness, both on their own and together, will clearly affect the prospects for economic and social development in given locations.

GPNs and Regional Development

With these 'building blocks' in place, we can now move on to consider how these ideas can help us to understand (subnational) regional development in a globalizing world. Importantly, analytical attention needs to be paid to *both* endogenous growth factors within specific regions *and* also to the strategic needs of translocal actors coordinating GPNs. Regional development can thus be conceptualized as a dynamic outcome of the complex interaction between region specific networks and GPNs within the context of changing regional governance structures. We argue that it is these *interactive effects* that contribute to regional development, not inherent regional advantages or rigid configurations of globalization processes. Despite certain path dependent trajectories, regional development remains a highly contingent process that cannot be predicted *a priori*.

In this view, endogenous factors are necessary, but insufficient, to generate regional growth in an era in which competition is increasingly global. There is no doubt that, for development to take place, a region must benefit from economies of scale and scope derived from the local human, technological, and institutional resource base. In **Figure 2**, the term 'regional assets' is used to describe this necessary precondition for regional development. In general, these assets can produce two types of economies. First, economies of *scale* can be achieved in certain regions through highly localized concentrations of specific knowledge, skills, and expertise. Second, economies of *scope* can exist if these regions are able to reap the intangible benefits of learning and the cooperative atmosphere embedded in these agglomerations. GPN analysis suggests that economies of scale and scope embedded within specific regions are only advantageous to those regions – and bring about regional development – insofar as such region specific economies can complement the *strategic needs* of translocal actors situated within GPNs. As shown in **Figure 2**, when such a

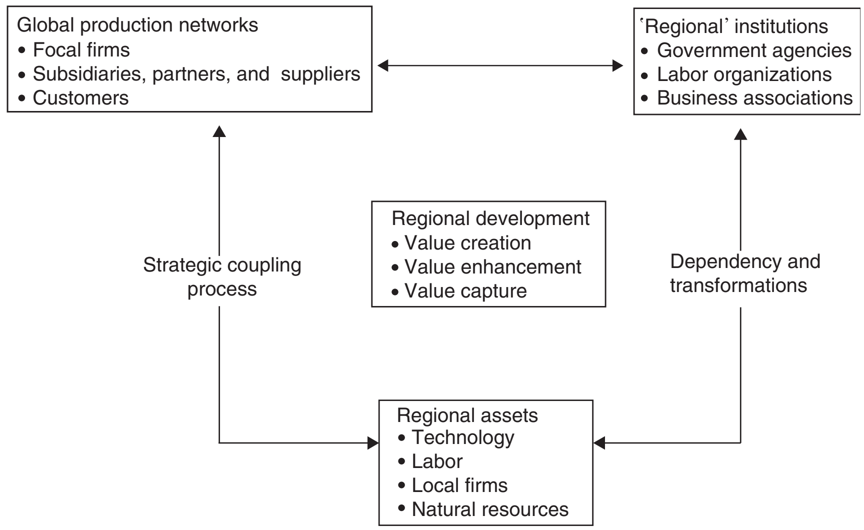

Figure 2 Global production networks and regional development. Source: Coe, N.M., Hess, M., Yeung, H.W.-C., Dicken, P. and Henderson, J. (2004) Globalizing regional development: a global production networks perspective. *Transactions of the Institute of British Geographers* 29, 468-484, Figure 1. Reproduced with permission of Blackwell Publishing.

complementary effect exists between regions and GPNs, a *coupling process* will take place through which the advantages of regions interact with the strategic needs of actors in these GPNs. Regional development thus depends on such a coupling process that evolves over time in relation to the rapidly changing strategic needs of GPNs and the rather slow transformations in regional economies of scale and scope.

In **Figure 2**, the coupling process is seen to work through the processes of value creation, enhancement, and capture identified earlier. The fact that a region is 'plugged into' a GPN does not automatically guarantee its positive developmental outcome because local actors in this region may be creating value that does not maximize the region's economic potential. Hence, regional assets can become an advantage for regional development *only if* they fit the strategic needs of GPNs. The process of 'fitting' regional assets with strategic needs of GPNs requires the presence of appropriate *institutional structures* that simultaneously promote regional advantages and enhance the region's articulation into GPNs. It is crucial here that the notion of 'regional' institutions includes not only regionally specific institutions, but also local arms of national/supranational bodies (e.g., a trade union's 'local' chapters), and extra local institutions that affect activities within the region without necessarily having a presence (e.g., a national tax authority). These multiscalar regional institutions are important because they can provide the 'glue' that ties down GPN in particular localities.

Three dimensions of such institutional structures are crucial to regional development. The first dimension involves the *creation of value* through the efforts of regional institutions in attracting the location of value added activities, for example, training and educating the local workforce, promoting start up firms and supplier networks, facilitating venture capital formation, and encouraging entrepreneurial activities. Secondly, *value enhancement* essentially involves knowledge and technology transfer and processes of industrial upgrading. The influence of regional institutions – government agencies, trade unions, employer associations, and so on – can be significant here. The third dimension of regional institutions in promoting regional development rests with their capacity to ensure *value capture*. It is one thing for value to be created and enhanced in some regions, but it may be quite another for it to be captured for the benefit of these regions.

Issues of *power* and *control* are critical in the analysis of value capture and the distributional aspects of regional development. Generally, the more a region is articulated into GPNs, the more likely it is able to reap the benefits of economies of scale and scope in these networks, but the less likely it is able to control its own fate. However, in certain circumstances, regional institutions may mobilize their region specific assets to bargain with transnational firms such that their power relations are not necessarily one way in favor of the latter. The bargaining position of these regional institutions is particularly high when their region specific assets are highly complementary to the strategic needs of transnational firms.

Future Challenges and Research Priorities

The GPN approach – broadly conceived – is currently gathering momentum. A wide range of studies have now been undertaken across a number of sectors – for example, retailing, temporary staffing, software, financial services, telecommunications, clothing, and automobiles – and across a broad range of (interconnected) geographies. It is fair to say, however, that thus far the theoretical development of GPN analysis has tended to outstrip the empirical elucidation of the dynamics being considered. Undertaking research in this vein is challenging both theoretically and methodologically.

Theoretically, GPN analysis seeks to tread a difficult path along the intersection of two important binaries – namely those of structure/agency and economy/culture. In the context of a broader globalization literature that tends to emphasize the structure/economy dimensions, GPN analysis must take care not to overstress the agency of economic actors, and the way in which their actions are shaped by sociocultural formations, when trying to correct the overall balance. Methodologically, GPN research has tended to proceed through the collection of sector specific, multisite, qualitative interview materials. Accordingly, the results of many studies have tended to be indicative rather than comprehensive. Such research arguably needs to be combined with more quantitative and extensive research techniques: this is particularly important if the aggregate effects of different yet overlapping GPNs for territorial development are to be deduced.

We can also point to a range of substantive issues that merit further elucidation under the broad umbrella of GPN research. Five will be outlined here. First, research is needed that unpacks the complex, multiscalar territoriality of GPNs. This creates complex dialectical relationships within and between firms engaged in the material processes of production, and the political and social actors and institutions embedded at different territorial scales. In turn, these relationships will determine the extent to which the strategic coupling of GPN and regional assets drives economic development in particular localities.

Second, research in this vein has tended to focus primarily on interfirm relationships, and in particular, the governance of those relationships. A significant contribution of GPN research thus far has been to draw attention to the wide range of extrafirm connections – to a wide variety of state and nonstate organizations – that are an integral part of a GPN. What remains left out, however, is what goes on *inside* firms; the firm still remains something of a 'black box' in this kind of research. Internal investment decisions within large firms, for example, will have a significant influence on the structure of GPNs and their developmental impacts.

Third, we need to know more about how governance structures within GPNs are increasingly being shaped by technological and organizational innovations that are codified in a series of *global standards*. Such standards vary greatly in terms of the aspects of GPN they apply to – for example, technological or labor conditions – and they may take 'strong' or 'weak' forms, for example, enforceable rules or less binding codes of conduct. Standards are both produced through actions within complex governance structures and are a means to subsequently exert power through those governance structures. These dynamics are innately geographical on at least three levels. It is important to recognize the multiple scales at which global standards are *negotiated*. Many of the private and public actors involved – firms, NGOs, consumer groups, trade unions, etc. – are distinct at the local, regional, national, and international level – thereby creating new institutional arrangements that link across geographical scales. We can also think about the various territories or networks *covered* by the jurisdiction of different forms of standards. Finally, we can think about the *impacts* of standards initiatives which may be felt at geographically disparate points of production networks.

Fourth, there is a significant gap in the GPN literature with respect to *labor*. Where they are discussed, workers are often seen in economistic terms as an asset underlying the structure of production networks. Equally, the labor process within the firm is rarely explored in relation to dynamics of value creation and enhancement. Links could productively be forged to the small but growing body of work that has emerged over the last decade or so under the banner of 'labor geographies'. Such work has a desire to open up analytical space for the agency of workers, and worker groups, to shape the geographies of capitalism. The labor geographies literature is not without its problems: it tends to be driven by case studies of successful actions; it tends to focus primarily on unions and union strategies, manufacturing sectors, and developed world examples; and the notion of 'agency' is oddly rather underdeveloped and thinly conceptualized. Nonetheless, this is a potentially analytically and politically rich line of enquiry that can, in certain contexts at least, serve to reveal the active and constituent role of workers within the value dynamics of GPNs.

Fifth, there appears to be a neglect of *final consumption* in GPN and cognate analyses, reflective, perhaps, of a productionist legacy in work of this kind. Even where the whole network, from production to consumption, is identified conceptually, there is seemingly a tendency in empirical terms to focus on the production 'stage', and a resulting assumption that the key dynamics shaping the system as a whole emanate outward from there. There is a need to more fully integrate the role of final consumption in our understandings of these systems. Most encouraging perhaps is the growing range of studies looking at the impacts of different kinds of consumer politics and initiatives (e.g., the UK's Ethical Trading Initiative, Fairtrade labelling, and the US campus anti sweatshop campaigns). The potential of these approaches lies in the way that they seek to integrate shifting notions of responsibility and ethical consumerism with the changing governance of production networks.

To conclude, GPN research is a new, ambitious, and still developing economic–geographical perspective on the processes of economic globalization and their impacts. While considerable progress has been made in conceptualizing the complex organizational and geo graphical structures of GPN, much remains to be done in theoretical, methodological, and empirical terms.

See also: Embeddedness; Global Commodity Chains; Globalization and Transnational Corporations; Networks; Regional Production Networks.

Further Reading

Bair, J. (2005). Global capitalism and commodity chains: Looking back, going forward. *Competition and Change* 9, 153 180.

Coe, N. M. and Hess, M. (2006). Global production networks: Debates and challenges. http://www.sed.man.ac.uk/geography/research/gpe/gpnworkshop.htm (accessed on 28 July 2007).

Coe, N. M., Hess, M., Yeung, H. W. C., Dicken, P. and Henderson, J. (2004). Globalizing regional development: A global production networks perspective. *Transactions of the Institute of British Geographers* 29, 468 484.

Dicken, P. (2005). Tangled webs: Transnational production networks and regional integration, *SPACES 2005 04.* Marburg: Department of Geography, University of Marburg.

Dicken, P. (2007). *Global shift: Mapping the changing contours of the world economy* (5th edn). London: Sage.

Dicken, P., Kelly, P. F., Olds, K. and Yeung, H. W. C. (2001). Chains and networks, territories and scales: Towards a relational framework for analysing the global economy. *Global Networks* 1, 89 112.

Gereffi, G., Humphrey, J. and Sturgeon, T. (2005). The governance of global value chains. *Review of International Political Economy* 12, 78 104.

Henderson, J., Dicken, P., Hess, M., Coe, N. M. and Yeung, H. W. C. (2002). Global production networks and the analysis of economic development. *Review of International Political Economy* 9, 436 464.

Hess, M. (2004). 'Spatial' relationships? Towards a re conceptualization of embeddedness. *Progress in Human Geography* 28, 165 186.

Hess, M. and Yeung, H. W. C. (2006). Whither global production networks in economic geography? Past, present, and future. *Environment and Planning A* 38, 1193 1204.

Kaplinsky, R. (2005). *Globalization, poverty and inequality*. Cambridge: Polity.

Globalization and Transnational Corporations

P. Dicken, University of Manchester, Manchester, UK

Glossary

Foreign Direct Investment (FDI) Investment by a transnational corporation in productive assets outside its home country aimed at controlling their use.

Transnational Corporation (TNC) A business organization (firm) that has the power to coordinate and control operations in more than one country, even if it does not own them.

Transnational Production Network An interconnected set of functions of production, distribution, and consumption organized across national boundaries by a transnational corporation.

Introduction

'Globalization' has become one of the most dominant and controversial issues of the early twenty first century. The transnational corporation (TNC) is one of its major, some would say its most important, agents. Where does geography fit in all this? Does globalization, as is so often alleged, render geography irrelevant? Are TNCs placeless institutions with less and less connection with places and communities? These are the key questions addressed in this article.

Geographical Perspectives on Globalization

Globalization Is Geography

The myths continue to circulate. 'Globalization' means the 'end of geography', the 'death of distance', the emergence of a 'borderless world', of 'a level playing field', of – in the words of one recent best selling author – 'a flat world'. These are powerful discursive images that play a very important role in the ways in which influential economic and political actors make decisions that affect all our lives. But they are wrong; they are, indeed, myths. Globalization does not mean the end of geography. In a very real sense, globalization is geography. It consists of a whole syndrome of processes (economic, social, cultural, political) that occur in highly uneven and geographically differentiated forms. It involves both a stretching and an intensification of processes and relationships across a spectrum of spatial scales which are, themselves, being continuously reconfigured through processes of spatial switching (at the same geographical scale) and scalar switching (as processes are switched between different spatial scales). Globalization, in fact, is both a product and a creator of multiple geographies of immensely complex and dynamic scalarity.

Geographies of Globalization

The nature of all social processes (in the broadest sense), and the actors or agents involved, is that they are literally 'grounded' or 'embedded' in specific places. The nature of that grounding – the contexts in which they are created – is highly significant in influencing (though not determining) how such processes develop and how actors behave. At the same time, social processes and actors create new geographies, especially through the connections they make within and between spaces. In the case of economic activities, during the past century and, especially, during the past 40 years, there have been enormous shifts in the contours of the global economic map. Such changes, however, have not occurred along a simple unidirectional trajectory.

Processes of localization of economic activities remain very powerful; indeed, the clustering or geographical concentration of activities not only persists but is the norm. At the same time, there are different forms of geographical dispersal of economic activities: internationalization (simple extension across national boundaries but with low levels of functional integration); globalization (extensive geographical spread with a high degree of functional integration); and regionalization (the operation of 'globalizing' processes at a more geographically limited (but supra national) scale, as in the case of the European Union (EU) or other less developed regional groups). Quite clearly, globalization is not a universalizing tendency.

The global economic map is continuously being reconfigured. Old geographies of production, distribution, and consumption are continuously being disrupted; new geographies are continuously being created. In that sense, the global economic map is always in a state of 'becoming'; it is never finished. But the new does not simply obliterate the old. On the contrary, there are complex processes of path dependency at work. What already exists constitutes the pre conditions on which the new develops.

Geographically, the global economy has become increasingly multipolar. New centers of production – new geographical divisions of labor – have emerged in parts of what had been, historically, the periphery and semiperiphery of the world economy. Most significant of all has been the resurgence of Asia (especially East Asia, and including, most recently China) to a position of global significance commensurate with its importance before ‘the west’ overtook it in the nineteenth century. At the same time, many parts of the world remain, to a greater or lesser degree, disarticulated from the engines of economic growth. The contours of the global economic map reflect a landscape of staggeringly high peaks of affluence and deep troughs of deprivation interspersed with plains of greater or lesser degrees of prosperity.

A Synoptic Geographical Perspective on Globalization

Such dynamic geographies of globalization are the outcome of extremely complex actions and relationships between economic, political, and social institutions and actors all of which are deeply embedded in specific geographical structures. **Figure 1** provides a synoptic perspective. It is an approach that emphasizes the essentially networked nature of the global economy, one that conceives of economic processes (production, distribution, consumption) in terms of connections of activities, linked through flows of both material and nonmaterial phenomena (like services) into circuits and networks. Such circuits and networks constitute relational structures and processes in which the power relationships between the key ‘actors – firms, states, individuals, social groups – are uneven. Networks are always in flux; they are always in a process of becoming. They also, of course, do not exist in isolation. In particular, they are embedded within the broader macrostructures of the global economy as well as grounded in the prevailing geographical structures of the real world.

Within the geographically differentiated macrostructural framework, it is primarily the actions of, and especially, the interactions between, the five actor centered networks shown in the central section of **Figure 1** that shape the changing geographical configuration of the global economy at different spatial scales. Such interactions are sometimes conflictual, sometimes collaborative but, overall, the system is one in which power relationships tend to be asymmetrical. Some actors are, without doubt, more influential and powerful than others. The remainder of this article focuses on one such powerful economic actor – the TNC.

Transnational Corporations

The Myth of the Global Corporation

A parallel myth to that of the ‘flat world’ of globalization is that of the ‘global corporation’ – placeless behemoths whose gargantuan footprints trample unhindered across national boundaries, emasculating the autonomy of nation states, rendering policy makers powerless, and obliterating local cultures. Again, this is a myth that must be refuted, although there is absolutely no doubt that TNCs have become major shapers of the global economy and its geography. But they are far more varied in size, in structure, and in their geographies than the global corporation stereotype suggests. Indeed, very few of even the world’s 100 leading TNCs can be regarded as ‘global’ corporations in terms of their geographical extent. The vast majority of the top 100 TNCs still retain more than half of their activities in their home country. In fact, TNCs come in all shapes and sizes, from the so called global corporations operating in scores of countries to TNCs operating in only one or two countries outside their home base. What they all have in common is that they operate in different political, social, and cultural environments. In so doing, however, they retain many of the characteristics of the places/countries from which they originate. The nature and characteristics of a TNC’s ‘home base’ continue to exert a significant influence on its behavior.

Transnational Corporations and States

A TNC is a firm that has the power to coordinate and control operations in more than one country, even if it does not own them. By the very fact of operating across national boundaries – whether in just two, or in scores of, countries – a TNC ‘incorporates’ part of a state’s territory into its own network. It is this interpenetration of autonomous national spaces that is at the heart of the potential conflict between TNCs and states. The basis of a TNC’s power lies in its potential ability to take advantage of geographical differences in the availability and cost of resources and in state policies and to switch and reswitch operations between locations. However, this does not necessarily mean that TNCs always have the advantage. It can be argued that TNCs and states exist in symbiosis, each needing the other. For example, states need firms to generate material wealth and provide jobs for their citizens; conversely, firms need states to provide the infrastructural basis for their continued existence, both physical (the built environment) and social (a legal system to protect private property and to regulate business transactions).

In fact, TNCs and states are continuously engaged in intricately choreographed negotiating and bargaining processes. On the one hand, TNCs attempt to take advantage of national differences in regulatory regimes (such as taxation or performance requirements, like local content). On the other hand, states strive to minimize such ‘regulatory arbitrage’ and to entice mobile investment through competitive bidding against other states. The situation is especially complex because there is a

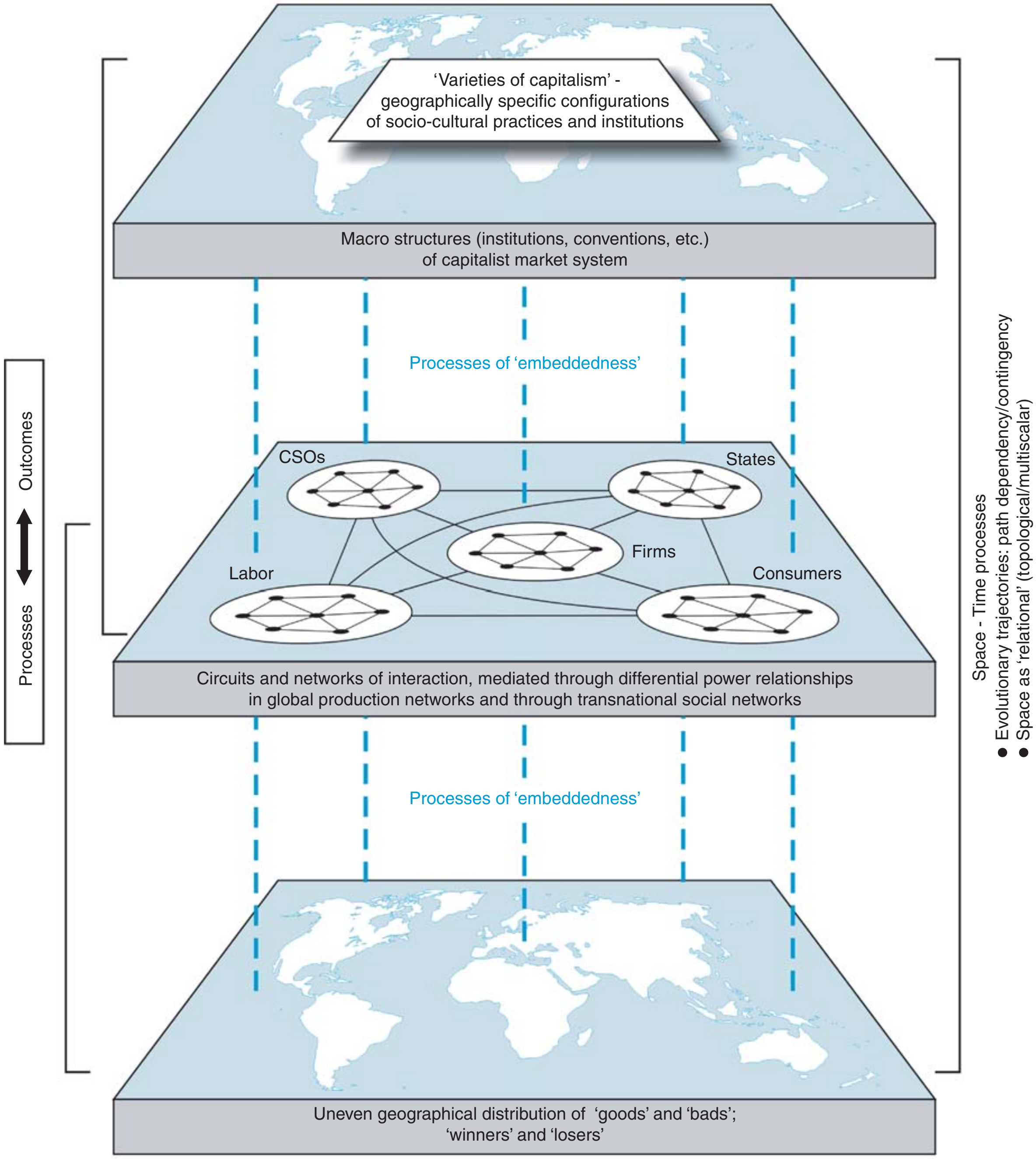

Figure 1 A synoptic geographical perspective on globalization. Source: Dicken, P. (2007). *Global Shift: Mapping the Changing Contours of the World Economy*. 5th edn. London: Sage Publications, London, Figure 1.3.

territorial asymmetry between the continuous territories of states and the discontinuous territories of TNCs and this translates into complex bargaining processes in which, contrary to much conventional wisdom, there is no unambiguous and totally predictable outcome. TNCs do not always possess the power to get their own way. In the complex relationships between TNCs and states – as well as with other institutions – the outcome of a specific bargaining process is highly contingent. States still have significant power vis à vis TNCs, for example, to control access to their territories and to define rules of operation. In collaboration with other states, that power is increased (the EU is an example of this). So, the claim that states are universally powerless in the face of the supposedly unstoppable juggernaut of the 'global corporation' is nonsense; the question is an empirical one.

Geographies of Transnational Corporations

Geographical origins and destinations

Probably the earliest, genuine TNCs were the colonial and merchant trading companies that emerged in Europe from the fifteenth century onward – companies like the Hudson's Bay Company and the English and Dutch East India Companies. Like the later manufacturing TNCs, which developed in the last quarter of the nineteenth century, such companies originated from the dominant economic nations of the day, notably Britain, the United States, and Germany. Today, most of the world's TNCs still originate from the developed economies – around 95 of the world's largest 100 nonfinancial TNCs and some 90% of the world's foreign direct investment (FDI). Nevertheless, changes are occurring. An increasing number of firms from newly industrializing economies (NIEs) is beginning to transnationalize, notably firms from the fast growing economies of East Asia (Korea, Taiwan, Singapore, and, most recently, China), some Latin American countries (notably Brazil), and some Indian firms (especially in the IT sector). The geo graphical origins of TNCs are clearly becoming more diverse, albeit slowly.

In contrast, the destinations of TNC activities remain very strongly concentrated geographically. Three quarters of the world's FDI goes to developed countries in a complex network of cross flows of investment. Of the one quarter going to developing countries, around three quarters of that investment is located in just ten countries – the fastest growing NIEs of East Asia, in particular. **Figure 2** maps the geographical pattern of FDI and shows its extraordinary unevenness. Outside the 'triad' of North America, Europe, and East Asia, the activities of TNCs, at least as measured by FDI, are spread very thinly indeed.

Why firms transnationalize

The most basic motivation for firms to establish trans national operations is, of course, to enhance their prof itablility. Developments in the enabling technologies of transportation, communication and product and process technologies, together with the trend toward greater political liberalization of trade and investment regimes, have both increased the possibilities for transnational operations and also intensified inter firm competition at much enlarged geographical scales. Operating across borders, whether directly or indirectly, has become the norm, even for some relatively small firms. The days when only large firms that had established a strong market position in their home countries could move abroad are long gone. In some cases, new firms operate transnationally from the very beginning.

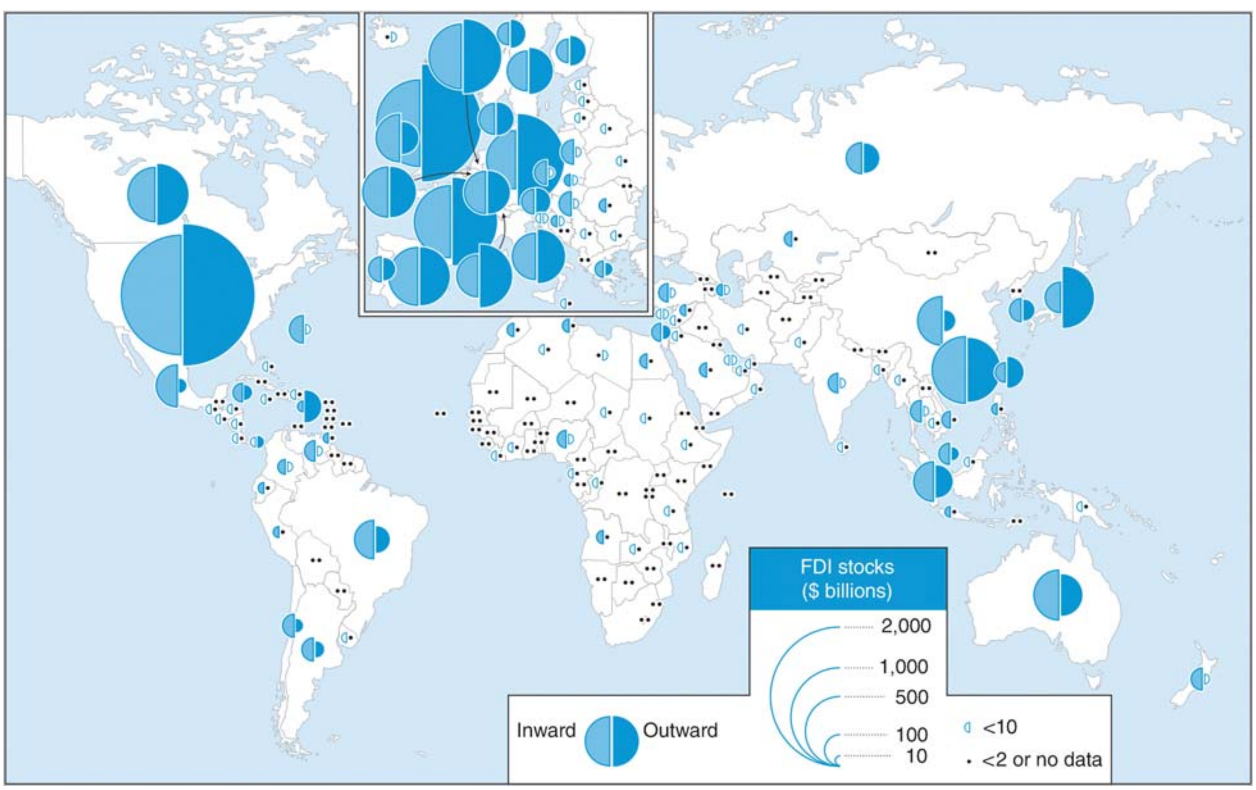

Figure 2 The global geography of foreign direct investment. Source: Dicken, P. (2007). *Global Shift: Mapping the Changing Contours of the World Economy*. 5th edn. London: Sage Publications, London, Figure 2.22.

There are two major sets of reasons why firms transnationalize their activities: (1) the attraction of new markets and (2) the differential geographical availability of assets and resources.

Market orientation

Much TNC investment continues to be market oriented. A firm may have reached saturation point in its domestic market so that increasing profitability may depend on expanding its market beyond its home territory. It may have identified new markets that require a direct presence in order to serve them efficiently; for example, transportation costs may be excessive (although this is less and less the case in the case of many products) to make exports uneconomic. Access to the market may be restricted because of political regulatory structures (such as import tariffs). The idiosyncratic nature of a particular market may necessitate a direct presence in order to understand, and to cater to, such specific circumstances. Both for political, as well as cultural reasons, it may be desirable for a TNC to appear to be strongly embedded in a local market. In other words, both the size and the particular characteristics of markets continue to influence the locational decisions of TNCs.

Asset-oriented investment

The 'assets' that firms need to produce and sell their products and services are geographically very unevenly distributed, and, therefore, may need to be exploited *in situ*. Traditionally, of course, it was the geographical localization of many natural resources that drove much of the early development of TNCs. It is no coincidence that many of the early leading TNCs were in the natural resource sectors, including energy and industrial resources as well as in agriculture. However, developments in transportation and communications technologies, as well as in production process technologies, have increased the ability of firms to access other unevenly distributed assets on increasingly wide geographical scales. This is most notably the situation in the case of human assets – the skills and knowledge embodied in people in specific local settings, with their particular assemblages of social and cultural institutions and practices. Such assets are, again, very unevenly distributed geographically. It is really only in the past 50 years or so that these kinds of assets have come to play a significant role in transnational investment.

Initially, it was the attraction of cheap – and usually unorganized – labor that was the primary attraction for firms in certain industries, such as textiles, garments, footwear, toys, and consumer electronics, to establish production abroad. The so called 'new international division of labor' that sprang into prominence in the literature of the 1970s and 1980s was based upon the claims that firms in the western industrialized countrie were fleeing the constraints of high cost, militant labor to tap cheap and malleable labor in developing countries Without doubt, some types of transnational investment have been – and remain – very sensitive to geographical variations in labor costs.

But it is not only labor costs that are the major driving force. Except in those industries where unskilled – and, therefore, cheap – labor is important, it is other attributes of human capital that have become more significant. In particular, it is increasingly the availability of well educated, highly skilled and strongly motivated workers located in 'quality' communities that are exerting a very strong influence on TNCs. Of course, where these attributes are combined with relatively low costs, the attraction is much reinforced. Such circumstances underpin much of the current transnational investment in East Asia, in the emerging market economies of Eastern and Central Europe, and in some cities in India (e.g., the much quoted example of the IT cluster in Bangalore).

How firms transnationalize

There are two major ways in which firms develop transnational activities: one is through what is known as 'greenfield' investment; the other is through engagement with other firms, through either merger and acquisition or some form of strategic collaboration.

Greenfield investment is simply the building of totally new facilities (an administrative office, a factory, a research and development facility, a sales and distribution centere, and so on). By definition, it adds to the productive stock of both the firm itself and the country/community in which it occurs. For that reason, it is generally the type of investment most favored by host countries. However, greenfield investment is far from being the most common mode of overseas expansion. Building a totally new facility, especially one of a significant scale, is a risky venture. For that reason, a firm may prefer to establish a presence in an overseas location through an involvement with an existing firm. Many firms have preferred to merge with, or to acquire, another firm to establish, or to expand, their presence in a particular overseas location.

Another widely used mode of TNC expansion is to enter into a strategic collaboration with one or more other firms. Many companies are forming not just single alliances but networks of alliances, in which relationships between partner firms are increasingly multilateral, rather than bilateral. Unlike mergers and acquisitions, in which the identities of the merging partners are completely subsumed, in an alliance, only some of the participants' business operations are involved; in every other respect the firms remain not only separate but also usually competitors.

'Webs of Enterprise': Geographies of Transnational Production Networks

TNCs can best be envisaged as 'networks within networks', with highly complex and dynamic internal and external geographies reflecting the ways these networks are configured and connected. By the very nature of their dispersed geographical spread across different political, cultural, and social environments, TNCs are far more difficult to coordinate and control than firms whose activities are confined to a single national space. They require, in other words, a more sophisticated organizational architecture. However, there is no single solution adopted by all TNCs but, rather, a diversity of different ways of coordinating TNC networks. Some of this diversity reflects the different geographical origins of TNCs, with a tendency for TNCs from different countries to prefer a particular organizational form.

However, a few generalizations can be made. Characteristically, the corporate headquarters of TNCs invariably remain in the firm's home country (often in the community in which the firm originated). Some kinds of headquarters functions may well be dispersed to key locations within the firm's transnational network, usually in key cities. These may, for example, be regional headquarters functions or possibly specialized control functions for specific lines of business. Core research and development (R&D) facilities also still tend to remain in the home country although some kinds of R&D have become increasingly dispersed, particularly to tap into localized sources of scientific and technological expertise, both institutional (as in the case of universities or other research institutions) and human (pools of key scientific and technical workers), or to adapt new products to local conditions. Conversely, sales and marketing functions tend to be dispersed to locations in key markets, while production functions are sensitive to the technical needs of the specific sector in question.

Compared with corporate headquarters and R&D facilities, production activities have become more dispersed geographically in the search for key assets and/or proximity to markets. A number of geographical configurations of TNC production activities are apparent. One option is to concentrate production at a single location. Such globally concentrated production generates economies of scale in production but increases transportation costs and lessens the firm's knowledge of distant markets. A second option is to produce specifically for a local/national market. Here economies of scale are limited by the size of the market. A third option is to create a structure of specialized production for a regional market (such as the EU, North America, or East Asia). A fourth possibility is to segment the production process and to locate each part in different locations – a form of transnational vertical integration of production.

Of course, TNCs never work with a blank surface. They have to deal with a complex mix of facilities that have been built up, or acquired, over a period of time and whose structure, at any one time, may no longer be appropriate for changing circumstances. For that reason, TNCs are constantly engaged in processes of restructuring, reorganization, and rationalization. TNC networks are always in a perpetual state of flux. At any one time, some parts may be growing rapidly, others may be stagnating, and others may be in decline. The functions performed by the component parts and the relationships between them will alter. Such restructuring and rationalization inevitably causes tensions between TNCs and other 'stakeholders', notably governments and labor.

Changes to a firm's geographical configuration often occur as a result of the firm's decision on what to produce for itself, in house, and what to externalize to independent suppliers. The 'make or buy' decision has become particularly critical as competition has intensified as firms strive to increase their efficiency to enhance or maintain profitability. TNCs, therefore, are highly dependent on other firms for many of their needs. The specific relationships between TNCs as customers and other firms (including other TNCs) as suppliers, both organizationally and geographically, are changing rapidly. Pressures on suppliers to deliver 'just in time'; pressures on them to reduce prices; pressures on them to take on more responsibility and risk have come to characterize most global industries.

The geographical extent of such transnational production networks is highly variable. In fact, few such networks can be described as being truly 'global'. A marked recent trend is for such networks to have a strong regional dimension, that is, networks organized on a multinational scale of groups of contiguous markets. In some instances, such a tendency is reinforced by regional political structures but this is not a prerequisite. Simple geographical proximity is in itself a powerful stimulus for integrating operations. Transnational production networks organized at the regional scale are evident in most parts of the world, but most especially in the three 'triad regions' of North America, Europe, and East Asia. In North America, the establishment of the NAFTA is leading to a reconfiguration of corporate activities (particularly involving Mexico) to meet the opportunities and constraints of the new regional system.

In Europe, the increasing integration and enlargement of the EU has led to substantial reorganization of existing corporate networks and the establishment of pan EU systems by existing and new TNCs. Such regional reconfigurations increasingly incorporate the transitional economies of Eastern Europe, especially those that have recently become members of the EU. However, the process is complicated. On the one hand, supply side forces encourage a pan EU structure of operations to take advantage of scale efficiencies. On the other hand, demand side forces are still articulated primarily at the country specific level, where linguistic and cultural dif ferences play a major role in the demands for goods and services. In effect, the strategic tensions between global integration and local responsiveness are played out at the EU regional level.

Although East Asia does not have the same kind of regional political framework as the EU or NAFTA, there is very strong evidence of the existence of regional production networks organized primarily by Japanese firms, although non Asian as well as some other Asian firms (e.g., from Korea, Hong Kong, and Taiwan) also tend to organize their production networks regionally. Within East Asia, a clear intraregional division of labor has developed consisting of four tiers of countries: Japan; the so called 'four tigers' of Hong Kong, Korea, Singa pore, and Taiwan; the South East Asian 'later indus trializers' – Malaysia, Thailand, Indonesia, the Philippines; China, together with, at least potentially, countries such as Vietnam.

The Ties That Bind

Transnational corporations are, undoubtedly, one of the – arguably the most important – primary shapers of the contemporary global economy. There is no doubt that their significance is increasing as more companies, from a widening range of home bases, become transnational at an earlier stage of their development. But TNCs are far more diverse than is often claimed. Not all are 'global' corpor ations. Indeed, very few are. TNCs come in a whole variety of shapes and sizes and there remain significant differences between TNCs from different countries of origin. Diversity, rather than uniformity, rules.

Both the organization and the geographies of large TNCs, and of their transnational production networks, are immensely complex and dynamic. In a very real sense, the global economy can be pictured as intricately connected localized clusters of activity embedded in various ways into different forms of transnational cor porate network that, in turn, vary greatly in their geo graphical extent. Some TNCs are globally – or at least regionally – extensive; others are more restricted geo graphically. In all cases, however, firms in specific places – and, therefore, the places themselves – are increasingly connected into transnational networks. Inevitably this creates tensions between TNCs and other significant actors in the global economy: states, local communities, labor, consumers, and civil society organizations. TNCs not only have their own geographies, but they also create new geographies in which places across the world be come incorporated. Whether this is good or bad is a matter of heated debate.

See also: Agglomeration; Embeddedness; Foreign Direct Investment; Global Commodity Chains; Global Production Networks; Globalization, Economic; Networks; Regional Production Networks; Transnational Corporations in Developing Countries.

Further Reading

Amin, A. (2001). Globalization: Geographical aspects. In Smelser, N. J. & Baltes, P. B. (eds.) *International Encyclopaedia of the Social and Behavioural Sciences*, pp 571 587. Amsterdam: Elsevier.

Bridge, G. (2002). Grounding globalization: The prospects and perils of linking economic processes of globalization to environmental outcomes. *Economic Geography* 78, 361 386.

Coe, N. M. and Yeung, H. W c. (2001). Geographical perspectives on mapping globalization. *Journal of Economic Geography* 1, 367 456.

Dicken, P. (2000). Places and flows: Situating international investment. In Clark, G. L., Feldman, M. P. & Gertler, M. S. (eds.) *The Oxford Handbook of Economic Geography*, pp 275 291. Oxford: Oxford University Press.

Dicken, P. (2003). 'Placing' firms: Grounding the debate on the 'global' corporation. In Peck, J. A. & Yeung, H. W c. (eds.) *Remaking the Global Economy: Economic Geographical Perspectives*, pp 27 44. London: Sage Publications.

Dicken, P. (2004). Geographers and 'globalization': (Yet) another missed boat? *Transactions of the Institute of British Geographers* 29, 5 26.

Dicken P. (2006). Economic globalization II: Corporations. In Ritzer G. *Blackwell Companion to Globalization*. ch. 16. Malden, MA: Blackwell.

Dicken, P. (2007). *Global Shift: Mapping the Changing Contours of the World Economy* (5th edn.). London: Sage Publications. New York: The Guilford Press.

Dicken, P., Kelly, P. F., Olds, K. and Yeung, H. W c. (2001). Chains and networks, territories and scales. *Global Networks* 1, 89 112.

Dicken, P. and Malmberg, A. (2001). Firms in territories: A relational perspective. *Economic Geography* 77, 345 363.

Kelly, P. F. (1999). The geographies and politics of globalization. *Progress in Human Geography* 23, 379 400.

Peck, J. A. and Yeung, H. W c. (eds.) (2003). *Remaking the Global Economy: Economic Geographical Perspectives*. London: Sage Publications.

UNCTAD (annual). World investment report. *Report of the United Nations Conference on Trade and Development.* New York: United Nations.

Yeung, H. W c. (1998). Capital, state and space: Contesting the borderless world. *Transactions of the Institute of British Geographers* 23, 291 309.

Yeung, H. W c. (2002). The limits to globalization theory: A geographic perspective on global economic change. *Economic Geography* 78, 285 305.

Relevant Websites

www.unctad.org
United Nations Conference on Trade and Development.

Globalization, Cultural

K. N. Rankin, University of Toronto, Toronto, ON, Canada

Glossary

Articulation A joining together of diverse elements in the constitution of societies structured in dominance; a Gramscian interpretation of articulation is engaged here, which involves, as Stuart Hall has noted, a focus on the production of meaning through practice such that the connection or link must be positively sustained by specific processes and can under some circumstances be overthrown, leading to new connections – rearticulations – being forged.

Conjuncture The notion of conjuncture comes from Antonio Gramsci who sought to develop a method of moving from abstract Marxian laws of economic development to concrete analyses of 'current relations of force'. As Stuart Hall has argued in bringing Gramsci's work to bear on contemporary issues, a conjunctural analysis involves both a materialist premise, that the analysis of political and economic structures must be grounded in their material conditions of existence; as well as a historical premise, that the specific forms of these relations cannot be deduced *a priori* from abstract economic laws, but must rather be made historically and spatially specific.

Flexible Accumulation The regime of accumulation characteristic of post-Fordism, involving methods of production that are considered flexible, such as flexibly deployed labor, easily programmable machines, and just-in-time production.

Frankfurt School A group of radical scholars who established critical theory as a central trajectory of 'Western Marxism', and who were associated with the Institute of Social Research founded in Frankfurt, Germany in 1923.

Historical Materialism An analytical method associated with Marxism rooted in a premise that social relations are grounded in their material conditions of existence and emerge out of concrete historical developments.

Post-Fordism A collection of workplace practices, modes of industrial organization, and institutional forms identified with period since the mid-1970s, following the era referred to as Fordist – and commonly understood to extend to associated modes of social regulation.

Postmodernism Emerging first in architecture, a theoretical orientation that expresses skepticism toward modernist universalisms, especially metanarratives that privilege the vantage point of the artist, theorist, or observer, and favors instead a popular eclecticism in theory, art, philosophy, and social science.

Post-Structuralism A theoretical orientation that rejects the formalism of structuralist linguistics and its epistemological subject.

Practice Theory A field of anthropological inquiry which consolidated in the 1970s and 1980s around a common interest in exploring the 'processes' through which cultural systems are produced and reproduced, and how they change – and which is associated with the work of Pierre Bourdieu and Marshall Sahlins.

Introduction

The term 'cultural globalization' evokes images of malnourished children drinking coca cola in impoverished agrarian settings, of satellite television broadcasting MTV and American sitcoms in remote Himalayan villages, and of the consumption of 'exotic' novelties – Thai food, flamenco dance, African music – in metropolitan centers all over the world. The study of cultural globalization is motivated variously by a worry about the colonization of cultural life by capitalism, a celebration of the cultural hybrids that emerge as people, images, ideas and money traverse the globe, and an expectation that 'culture' itself could become a site of resistance to prevailing ideologies and conditions. As such, cultural globalization as an object of academic study emerged in the late 1980s and early 1990s, during a period of rapid functional integration of economic activities across the globe, but in which 'the economic' was perceived to be an inadequate lens for assessing experiences of disjuncture and entanglement, as much as the symbolic, ideological and discursive dimensions of global interconnectedness. It was taken up with particular vigor in cultural studies, where the emphasis has been on representational politics, as well as in anthropology and geography, which have tended rather to explore some of the contradictions between the production and experience of cultural globalization.

'Circulation' and 'flow' are the fundamentally geographic metaphors engaged in academic and popular venues alike to capture the movement – of people, ideas, money, cultures, and information – that constitutes cultural globalization. As anthropologist Anna Tsing has noted, circulation and flow replace the metaphor of 'penetration' once used on the critical left to characterize the incorporation of peripheral regions into the capitalist world system; while penetration evokes the violence of rape, circulation and flow evoke more benign images like blood coursing through the body. Their use signals a shift in focus from production to exchange and consumption as constitutive domains of globalization in an era of post Fordist flexible accumu lation. It also signals the turn in human geography and across the social sciences to 'culture' as a category of analysis, as circuits and flows index interconnections among peoples, places, and ideas – relationships forged at least in part through negotiations over mean ings, values and beliefs, as well as through social networks and associations. These metaphors have served a range of contradictory claims about the politics of globaliza tion – that it frees people and capital from restrictive regulation; that the rapid expansion of communications technology gives more people access to information; and that the cultural disjunctures associated with movement ultimately present a challenge to American cultural hegemony.

Despite this apparently indiscriminate celebration across the ideological spectrum, it is possible to dis tinguish a range of epistemological framings and moral valuations in the geographic and related literatures on cultural globalization. This article aims to tease out these differences by presenting a typology of approaches to the circulation metaphor and their attendant representations of 'the cultural' and 'globalization'. One approach em phasizes the flow itself – what is being circulated, where, how, and to what effect. Within this epistemological ru bric, we can find moral valuations that are both positive (in the form of postmodern celebrations of difference) and negative (in Frankfurt School type alarm about the Americanization or standardization of culture and the destruction of local differences). A second approach considers instead the political–economic conditions of circulation. Here the emphasis is on closing the space between the economic and the cultural by considering the economies of cultural globalization and the cultures of economic globalization. The prevailing normative in terpretation in this scenario is critical – for example, lamenting the commodification of the aesthetic or chronicling the alternatives to capitalism in people's everyday practices around the world. But considerable normative ambiguity remains, insofar as this epistemo logical framing can also be deployed in the service of neoliberal capitalism – by, for example, injunctions to 'get the institutions right', promote social capital, strengthen civil society, and so on.

The article argues that more sound normative footing can be found in a third approach, which challenges the unity of the circulation metaphor and the implication of 'the global' and 'the cultural' in a reductive binary logic. Drawing on the explicitly Gramscian concepts of the conjunctural and the hegemonic, the articulation ap proach traces historically and geographically specific processes of spatial interconnection. In so doing, it rejects the abstract determinisms and inexorable forces so fre quently conjured up by notions of 'cultural globalization', to chronicle the distinctive cultural forms of capitalism taking shape in different regions of the world. It adopts an explicitly critical stance along the Marxian lines of seeking to understand the world in order to change it. The article is organized into three sections, which discuss each of the three approaches, followed by a concluding section that considers the implications of the favored 'articulation' approach for methodology and political praxis in human geography.

Flows-and-Impacts Approach

The crudest versions of the flows and impacts approach express a worry about the export of American mass consumer culture around the world. Framed in this way, the research problem becomes one of specifying the 'cultures of globalization' as well as determining the impacts of cultural globalization on various peoples and places. Attempts to specify an essential global culture generally focus on the diffusion of ideologies of con sumption through communications and information media, and an attendant convergence of cultural styles as Western 'values' of consumerism are embraced around the world. Thus, geographers have speculated on how increased access to mass media has set in motion a trend toward individualization, rooted in the growing import ance of material consumption for expressions of indi vidual identity. Anthropologists have likewise attempted to specify a 'culture of neoliberalism' featuring con sumption as a primary determinant of identity, the eclipse of social class as a relevant social category, and the emergence of race, gender, and age as key principles of identification and mobilization. Such cultural homogen ization theses also find expression in much of the lit erature on the cultural industry which laments the incorporation of increasing domains of cultural life into webs of capitalist organization. Not even the aesthetic remains as a sanctuary from the economic colonization of culture. A corollary to this line of thinking is to speculate about the dilution of traditional 'local' cultures, for example, the decline in significance of territorially based group identity in the wake of processes of individualization.

Now, it is certainly reasonable and necessary to pay attention to the ways in which corporate capital aims to diffuse certain consumerist practices as elements of mass culture – eating fast food, dressing fashionably, pursuing the exotic, and so on. But, as many cultural geographers have pointed out, cultural homogenization theses too easily elide the production of dominant cultural ideology – consumerism – with historically and geographically contingent cultural practices – consumption as a process of meaning making and social reproduction situated in relation to class processes and associated power relations. It is one thing to say that the neoliberal project of the 1980s sought to make private, market based choice central to social life, quite another to claim that it succeeded in producing economically 'rational' subjects. Perhaps the most significant danger of the elision between the production and practice of culture lies in the way it echoes, even fuels, some extremely regressive cultural nationalisms prevalent in many parts of the world today. Claims being made in many postcolonial states about the cultural essence of the nation rest on similar arguments about the threats to local cultural distinctiveness posed by the colonizing force of Western culture. Much violence has been committed in the name of resistance to Western cultural imperialism, not least the rolling back of core liberal rights and protections against (particularly gender) discrimination. The cultural homogenization thesis also perpetuates the familiar common sense binary of the global local, whereby the 'local' is not only opposed to the 'global' in an essentialist manner, but also contained within it. As such, it denies agency to the people who are assumed to passively and uncritically absorb the ideologies and lifestyle positions of corporate controlled media, along with fetishizing information technology by attributing it with transformative powers.

The more interesting geographic work in the flows and impacts genre asks about how dominant cultural representations are 'indigenized' in particular times and places. Anthropologist Arjun Appadurai has been particularly influential among cultural geographers, for his formulation of the 'disjunctive global economy of culture' constituted by nonisomorphic flows of people, technology, images, money, and ideas. The landscapes formed by the intersection of these flows – which he designates as ethnoscapes, technoscapes, mediascapes, financescapes and ideoscapes – in turn constitute the 'imagined worlds' for historically and geographically situated peoples and groups. Global flows connect people in far flung lands, but not in a linear or predictable fashion; they are fragmented, irregular, move in many different directions, and look different from different locations and perspectives. In so doing they constitute multiple, distinctively 'non Western' capitalist and postsocialist modernities. Appadurai thus moves beyond static center periphery models to emphasize relationality and interconnectedness and poses a challenge to the notion of globalization as unfolding teleology. In this way his work resonates with, and in fact informs, much of the work on transnationality in cultural geography, which has similarly engaged relational theorizing to consider how transnational cultural flows forge connections across and between entities formerly theorized as discrete and autonomous – thus dispensing with rigid state centric narratives of globalization processes. Studies of consumption in cultural geography and related disciplines have also taken their cue from Appadurai, to examine 'the social life of things' – how the commodities that are flowing change physically and symbolically as they move through global space, to what ends they are used, and how they create new meanings wherever they go. In this way, commodities themselves provide a lens for exploring interconnections between locations and movements, as well as the politics of consumption.

The focus here on how people find meaning in and respond to globally circulating forms of material culture (as well as migration, technology, finance, and ideology) suggests that the 'impact' of cultural globalization is not homogenization, but rather fragmentation, not smooth transmittal but disjuncture. An emphasis on indigenization would seem to suggest that the deepening of local particularities is in fact a defining characteristic of cultural globalization – and indeed many of the contributions to Frederic Jameson and Masao Miyoshi's *Cultures of Globalization* are committed to documenting the vitality of indigenous languages, peasant political movements, ritual revitalizations, and other local emergences of difference as constitutive dimensions of globalization. Paradoxically, these emergences may be motivated by elite consumption in global cities of the North and South, as much as by articulated resistance to cultural nationalisms or cultural imperialisms.

A related, more explicitly post structuralist interpretation of fragmentation attributes a subversive quality to the cultural mixing that transpires particularly in the 'liminal spaces of in between' (as in, in between cultures, communities, or nations). 'Hybridity' is the designation that cultural critic Homi Bhabha attaches to such processes – suggesting that they entail a progressive appropriation and rearticulation of dominant cultures. Processes of transculturation, creolization, and hybridization that transpire most vigorously at the margins or in liminal spaces of cultural contact, that is, have the potential to destabilize cultural purities, subvert categorical oppositions (e.g., local/global), and create the conditions for cultural reflexivity. A related set of arguments posits cultural globalization as entailing eclectic contacts and borrowings that encourage the proliferation of new cultures – representing culture as impure and disorderly combinations, in contrast to its hegemonic representations as isolated and regulated tradition. We have here a view of the world as fundamentally decentered and a claim that cultural hybrids themselves constitute a critique and subversion of cultural imperialism and attendant forms of domination.

Reservations about accounts of fragmentation, indigenization, hybridization, and the like have centered on their generally positive moral valuation and political naivitée. In an effort to overcome essentialist interpretations of culture as tradition or barrier to change, such accounts end up celebrating flow itself. Geographer Katharyne Mitchell has been among the most cogent critics of post structuralist interpretations of transnationalism and cultural globalization – suggesting that the tendency to attribute a progressive politics to cultural expressions of in betweenness results in neglect of the oppressive socio economic forces that often underlie conditions of hybridity and transnational flow. As she puts it, "[i]n an era of global capitalism, the heralding of subject positions 'at the margins' too often neglects the actual marginalization of subjects" – the poverty, under employment, ambivalent citizen status, cultural misrecognition, and social ruptures that are common to the 'experience' of migration and other conditions of in betweenness. Imputing a progressive politics of intervention to the subjects of cultural hybridity, moreover, overlooks the ways in which those 'on the margins', themselves can play a role in reproducing both socioeconomic and cultural forms of injustice in the margins.

In fact, it can be said that the flows and impacts approach in general suffers from too narrow interpretations of culture and geographic scale that ultimately leave unchallenged neoliberal assertions of market triumphalism at 'the end of history'. For however much it seeks to deterritorialize culture, the proposition that once immobile 'local' places are now transcended by 'global' flow suggests that there are still, as Anna Tsing puts it, some local people stuck inside local situations – and belonging to stagnant 'local cultures'. The more nuanced interpretations of flow do of course move beyond popular representations of the local as the stopping point of global circulations, where global flows are unquestioningly consumed and incorporated. But, as Tsing has argued, even the concept of hybridity relies on an understanding of the local as the place where global flows fragment and become something place bound and particular. Such invocations of the 'global' and the 'local' produce an understanding of geographic scale as essentially fixed and hierarchal domains in which particular processes are consigned to specific levels. As geographer Gillian Hart points out in a series of important essays on development, they also play into disabling discourses of globalization that reinscribe a series of related hegemonic binaries. In addition to active/passive, male/female, colonized/colonizer, production/consumption, public/private, economy/culture, these include dichotomous understandings of time and space in which time is treated as a primary analytical category while space appears as a passive container. The effect is to naturalize hegemonic forms of economic globalization that currently prevail and are taken for granted as producing outcomes in particular places. Within this epistemological framing, the fundamental analytical move is limited to chronicling the 'impacts' of globalization on places and people. In the cultures of economies approach, to which we now turn, the case is made instead for a more nuanced interpretation of agency and causality by deconstructing in particular the economy/culture dichotomy.

Cultures-of-Economies Approach

Some of the most persuasive and interesting remedies to the limitations of the flow approach to cultural globalization can be found in the subfield of economic geography. Formerly narrow representations of the economic in terms of modes and forces of production (determining the organization and experience of social life) gave way in the 1990s, to an appreciation of the social embeddedness of economic processes. At that conjuncture, geographers were responding not just to post structuralist critiques of reductive economisms within academia (some of which, it must be said, reflect a serious misreading of Marxist political economy, and on the contrary galvanized some persuasive illustrations of historical materialism as a method for cultural interpretation, such as David Harvey's *Condition of Postmodernity*). They were also taking their cue from nonacademic sources: from, as geographer Phil Crang has suggested, the shift in popular radical politics away from concerns with socioeconomic redistribution to a focus on representational politics aimed at remedying cultural forms of injustice; and also from the ways in which the capitalist economy began to define itself through culture – seeking, for example, to create the institutional conditions for innovation, learning, knowledge sharing, and so on. The turn to 'the cultural' in economic geography thus marked a new interest in discursive constructions and processes of meaning making as constitutive of economic practice. 'The cultural' moved from a derivative and secondary category of analysis to take center stage alongside 'the economic' in a more relational approach to understanding economic life. For our purposes of reviewing geographic approaches to cultural globalization, these developments are significant because they brought economic geographers into the business of documenting not only the political–economic conditions of global cultural flows, but also the ways in which culture, construed in this processual sense, contours global economic processes.

One formulation of these relationships seeks in effect to specify the intersection of 'the cultural' with 'the economic', the two way mutually constitutive interrelations between constructed economic and cultural spheres. Thus, for example, in a seminal study on the *Spatial Divisions of Labor*, Doreen Massey chronicles the reciprocal relationship between public and private investment in South Wales and key cultural characteristics – namely, geographically and historically specific forms of patriarchy, masculinism, Methodism, trade unionism, and corporatist community life. Rather than merely chronicle the impact of direct investment on localities, she explores how changes in industrial geography are contoured by embedded social and political struggles. Studies of global commodity chains have similarly foregrounded the significance of consumption for capitalist expansion and for tracing links among spatially dispersed histories. This work has antecedents in anthropology where there is a longer history of documenting the ways in which 'local' cultural practices contour particular moments of integration into global economic processes; an exemplary study in this regard is Sidney Mintz' *Sweetness and Power*, a historical ethnography of the emergence of a taste for sugar among the working class in Britain and its relationship to the dramatic rise of sugar cane production in the Caribbean colonies. Even the globalizing finance sector, often taken to be a purely 'economic' domain, has been subject to this sort of cultural analysis. Given the significance of personal interaction, discourse and interpretive skills in the conduct of financial intermediation, geographers such as Linda McDowell began in the 1990s to analyze the cultural practices of employees – how they perform gender, class, ethnicity, and so on – as well as the places of doing business and the discursive strategies engaged.

Others, most notably David Harvey, have maintained that the intersection between the economic and the cultural is rather more unidirectional. While not denying the significance of culture for the expansion of global capitalism, the argument in *The Condition of Postmodernity* – arguably one of the most widely influential books published in geography – is that in the last instance, the economic is determining: the postmodern (cultural) condition can be traced to the spatial and temporal dynamics of capitalism. This polemical tract was in fact written in response to Jean Francois Lyotard's *The Postmodern Condition* (1979), which aimed to dispense with modernist metanarratives of determination and ontology in favor of a populist eclecticism and deconstructivist stance. Harvey argued that the expansionist requirements of capital led in the 1980s to the dismantling of the Fordist economy and introduction of more flexible regimes of accumulation relying on global telecommunication and other means to reduce barriers of space by the speeding up of time. These increasingly polarizing and debilitating forms of capitalism have produced the cultural experiences of hybridity, pastiche, and disorientation so often characterized, and celebrated, in postmodern theory as the 'cultures of globalization'. Indeed, postmodern epistemology itself can and must be theorized in relation to these economic conditions of possibility – and the injustices of our times can only be confronted through this sort of historical materialism.

Within cultural studies these commitments find expression in work that takes as its problematic the export of American consumerism as a mode of daily life, but goes beyond simple homogenization arguments to consider the politics of cultural globalization. Frederic Jameson has been at the forefront of efforts to trace how the free movement of American culture (construed here as cultural industry, sectors of cultural production within which value depends on symbolic rather than utilitarian attributes – movies, music, fashion, television, and so on) subverts artistic production and spells the commodification of the aesthetic everywhere (including in America). In a contribution to *Cultures of Globalization*, he suggests that the problem with cultural homogenization theses is that they conflate 'the economic' and 'the cultural' as the cultural, and thus overlook the fact that American mass culture is fundamentally an economic enterprise. By way of example Jameson chronicles the concrete provisions worked out in global and regional trade agreements that secure American dominance and hegemony, such as the removal of subsidies to national film as 'unfair competition' or the opening of quota barriers on film in countries around the world.

An analogous investigation could be made of the political–economic circumstances surrounding the publication of this encyclopedia, with an eye to the ways in which academic publishing, too, is controlled by corporate interests. For the producers and consumers of academic knowledge it is common to view such publications as platforms for critical debate – and indeed critique about the polarizing modalities of capitalism. Rarely does one pause to consider the ways in which academic publishing (and thus scholarship) is imbricated in these very processes. The recent activity of a subsidiary of Reed Elsevier – publisher of this encyclopedia along with several key geography journals – has provided grist for such reflection within the community of critical and political geographers. That subsidiary, Reed Exhibitions, had included within its portfolio exhibitions of Defence Systems & Equipment International (DSEI), the world's biggest weapons fair. The stark proximity of critical geographical scholarship to a company with links to the arms trade prompted a heated debate on the Critical Geography listserv and in the pages of *Political Geography* in the first half of 2007 about the moral perils entailed in the commodification of academic knowledge. Those discussions raised the spectre of the many complex and contradictory ways that academia, and indeed Geography, is linked to the military and defense industries. They also revealed the political force academics can wield when they galvanize public attention. In a June 2007 news release, Elsevier announced its intentions to exit the defence exhibitions sector, in response to the 'very real concerns' of 'growing numbers of important customers and authors'.

In the case of the film industry, Jameson argues, this sort of morally ambiguous, if not repugnant, corporate control has substantively constrained content – drastically compromising posibilities for filmic experiments in imagining radically different social alternatives to this one we now live under. Thus, the triumph of Hollywood is not merely an economic triumph, it is also a political one. Unfortunately, Jameson moves too reductively from lamenting the role of American mass culture in quelling political innovation to lamenting its role in breaking up entire cultural systems in other parts of the world with the allure of Hollywood film and the consumerist way of life it represents. Like Harvey, he is certainly on the firmest ground when he focuses on the economic basis of cultural production, and here the work of both Jameson and Harvey offers some clear direction for radical praxis and counter hegemonic strategy. For consumerism and postmodernism are revealed in their work not as free floating abstract values, but as ideologies – views of the world – that are grounded in capitalist relations of production and concretely produced in particular historical and geographical conjunctures. The task then becomes one of identifying, and challenging, the political–economic conditions for their production – the trade agreements, capital market structures, legal systems, barriers to local ownership and innovation, and so on. Once these relationships are established, moreover, key hegemonic formulations can be refuted, such as the imputed causal relationship between free markets and political democracy. If consumerism is really a form of manufactured consent, 'free choice' in the market might in fact be shown to repress rather than foster political participation.

Of course, the standard critique leveled against materialist accounts of cultural globalization such as Jameson's or Harvey's is that they posit a too easy chain of determination from capitalist political economy to cultural forms and experience. Limiting one's reading of cultural forms to only those aspects that might be economically determined means that other bases of subject and cultural formation, such as ideologies and experiences of race and gender, may be missed. Another tack within the cultures of economies approach has been to focus instead on the economic rationality of culture (as opposed to its determination). This is a move which forces a productive reconceptualization of culture not as fixed signs and symbols with assigned meanings and values, but in terms of what people do. This interpretation may be located within the anthropological field of practice theory, which consolidated in the 1970s and 1980s around a common interest in exploring the 'processes' through which cultural systems are produced and reproduced, and how they change. Particularly influential in geography has been the work of Pierre Bourdieu, who in his writing on aesthetics and the 'economics of practice' is concerned not so much with the colonization of people's hearts and minds by corporate interests, as with engaging economic concepts to explore the role of cultural practice in reproducing (or transforming) existing social structures. Among those of unequal status, for example, gifting and other modes of reciprocity generate affective bonds that obfuscate – and cement – social hierarchy, and as such they constitute a kind of symbolic capital, if not symbolic violence. Likewise, practices of aesthetic appreciation articulate acts of self expression with processes of social reproduction. Crucially, the articulation of cultural practice with processes of social reproduction extends across transnational and global space. In their formulation of the 'translocal village', for example, Selvaraj Velayutham and Amanda Wise consider how South Indian Tamils residing in Singapore maintain strong social ties with fellow caste members in their village of origin. Through returns, remittances, flows of obligations and the replication of village practices in Singapore, a moral economy of obligations is expanded spatially across national borders. This transnational social field is perpetuated through regimes of affect – guilt and shame are mobilized, coercively, to reproduce, in a spatially expanded form, the endogamous caste practices of cross cousin marriage, kin networks, and ritual obligations. Without abandoning a concern with social justice, then, such representations give greater emphasis to questions of agency and subjectivity in global cultural processes than we find in the more structuralist accounts of Jameson and Harvey.

Others have sought to move beyond the structural coupling of distinct economic and cultural spheres, to consider how economic activity itself is always culturally constituted and spatially placed. Karl Polanyi is perhaps the seminal thinker in this regard. The key intervention in *The Great Transformation* was to dispute the premise in liberal economic theory that the self regulating market is some naturally occurring 'pure' form of the market that can be achieved by removing all forms of political intervention. Rather, it is the product of historical experiments in creating markets for land, labor, and money – beginning, Polanyi argues, with the Enclosure Movement in the 1500s – and an impossible utopian project. If anything can be said to be 'naturally occurring', the argument continues, it is the countermovements for social protection that spontaneously arise in reaction to the profound havoc visited upon people when these domains of social life are subject to market forces. The empirical task then becomes one of examining how 'the economic' is instituted in different times and places, and, specifically within capitalist contexts, how countermovements for social protection are contained through particular constructions of market society. Within comparative political economy, these commitments have found expression in attempts to document the 'varieties of capitalism', the continued institutional divergence that can be found among advanced capitalist countries in the face of global pressures for economic convergence. The emphasis here is on the 'institutional field' within which firms operate, including not only market competition, but also inter actions among firms, institutions, labor, and market rules – a conjunction widely summarized as introducing 'cul ture' into political–economic analyses. A corollary ar gument, coming from sociology and in particular the work of Mark Granovetter, considers the extent to which firms are 'embedded' in cultural environments that are suitable (or not) for technological and organizational innovation – such as social networks through which tacit knowledge is transferred.

Economic geographers have weighed in on this dis cussion by expanding the scalar focus of the analysis and moving beyond an understanding of culture as 'local environmental milieu' to recognize the significance of meaning making for constituting economic practice. For example, as Meric Gertler's work has consistently shown, national scale regulation institutionalizes certain ways of viewing the economy which in turn structure firm practices, what counts as economically rational behavior and even what appear to be cultural traits, like co operativeness or competition. At the scale of the firm, economic geographers have examined how such norms are institutionalized in particular constitutions of eco nomic identity and space, such as through recruitment policies, disciplining discourses of organizational change, or the construction of gendered, raced, and classed time spaces of production. The dynamics of (and obstacles to) change, about which the 'varieties of capitalism' approach is largely silent, must then be assessed in relation to the complex interactions among such scaled cultural articulations.

Before proceeding in a final section to elaborate the concept of articulation, one other direction economic geographers have taken the notion that economies are culturally constructed must be noted. Under the rubric of the 'diverse economies' project, J.K. Gibson Graham has advocated engaging post structuralist epistemologies to deconstruct dominant 'capitalocentric' representations of the economy in order to make way for counter hege monic reconstructions. *The End of Capitalism* (*as we knew it*) argues that the structuralist epistemologies underlying political economy reinforce capitalist hegemony by taking it for granted. Seeing capitalist hegemony as a dominant discourse rather than a social structure or systemic unity presents possibilities for creating com peting discourses, a language of economic alterity to reflect both capitalist and noncapitalist practices that constitute diverse economies. What is perhaps most re markable about this initiative has been the extent to which it has polarized economic geographers into camps of protagonists and critics. On the one hand, it has spawned conferences, edited volumes, special journal issues, and a diverse, all committed to the cause of economic alterity. On the other hand, it has provoked a vitriolic and defensive response by those who read Gibson Graham to be naively and arrogantly seeking to bring down capitalism through discursive and theoretical finesse.

As her later book, *Postcapitalist Politics*, makes clear, the desired intervention is more modest and tentative than such critiques would suggest, and centers on an import ant insight about the role of economic subjectivity in political change processes. The point is that economic subjectivity is more complex than conventional class analysis allows; depictions of class as a group of unified economic subjects fail to grasp the multiple class pro cesses in which individual subjects are always engaged – working in a capitalist firm, sure, but also likely doing unpaid housework, maybe participating in a babysitting club, hunting or picking berries to supplement food supply, and so on. Renaming class homogeneity as class heterogeneity makes it possible to take seriously the af fective and emotional stances that attach to different subject positions, to appreciate, for example, how the feelings of despair and vulnerability associated with some forms of labor pose a real obstacle for generating eco nomic activism. The point of documenting economic diversity through participatory means is in large part to change such debilitating modes of subjectivity and build a constituency for social change. Once people recognize the innovation and dynamism that already exist in their everyday economic practices, that is, they can undergo the 'ethical self transformation' necessary to participate in collective processes of building 'intentional economic communities'. Freed in this way from hegemonic ac counts and positionings, they can begin to view the economy not as an autonomous zone that impacts their lives, but as an 'ethical space for decision', within which all kinds of possibilities arise for engaging social surplus to replenish and expand the commons and the productive base. The hope that is vested in creating a language of economic diversity, then, has not so much to do with challenging capitalism by discursive fiat, as with in terpellating alternative political subjectivities and en abling new forms of economic activism.

Articulation Approach

Approaches discussed in the previous section are all significant for their attempts to view 'the economic' in relation to 'the cultural', and thus for attempting to challenge the hegemonic binaries so commonly associ ated with discourses of globalization. Yet in practice they often foreground one domain over the other. Political economic analyses contribute crucially to our understanding of the material conditions of cultural flow, but in subsuming 'the cultural' to 'the entertainment business' or a universal experience of time space com pression, such formulations evoke 'culture' in a secondary role to assist in the explanation of phenomena that are seen as primarily economic. Jameson, for example, ad dresses the production of consumerist ideology, but his analysis is limited to political–economic conditions of possibility and conflates too easily the production of ideology with the everyday practice of culture. This work thus overlooks a whole set of important questions about the way consumerist ideologies articulate social for mations in particular time space contexts. Likewise, ex panded notions of economic subjectivity (à la Gibson Graham) or the economics of culture (à la Bourdieu) go a long way toward revealing how power operates through culture, right down to the way people carry their own bodies, all of which certainly has important implications for the possibility of economic activism. But they, too, leave aside that point of intersection between the eco nomics of practice and the politics of economies by es chewing political–economic questions about control over the means of production and the structures that support systemic inequalities. The point in reviewing these omissions is not to suggest that one might somehow add up the two sides of the coin (the economic and the cul tural, the global and the local) in order to see the whole picture. It is rather to suggest that an entirely different epistemological stance is required, one that would seek to view the world more relationally from the outset, or, as geographer Gillian Hart puts it, 'to clarify diverse and interrelated trajectories of sociospatial change'. As this section elaborates, such epistemological distinctions matter, not just for the sake of theoretical precision, but also for how politics can be imagined and practiced.

The articulation approach thus begins, but only be gins, with the premise that thinking about 'cultural globalization' requires thinking about how products whose value depends on symbolic attributes (like fashion, film, or music) have been commodified under regimes of flexible accumulation; and that 'the cultural industry' is deeply implicated in American imperialism and capitalist expansion today, along with its attendant exclusions, in equalities, and class processes. But that is not to assert a singular and inexorable transnational capitalist force encompassing farther and farther flung corners of the globe into its fold. An articulation approach points rather to more conjunctural questions about how conditions for capital accumulation are actively created and constantly reworked in particular historical and geographical con texts. The notion of conjuncture comes from Antonio Gramsci who sought to develop a method of moving from abstract Marxian laws of economic development to concrete analyses of 'current relations of force'. As Stuart Hall has so carefully argued in bringing Gramsci's work to bear on contemporary cultural–political dilemmas, it involves a materialist premise, that the analysis of political and economic structures, such as the cultural industry, must be grounded in their material conditions of existence; as well as a historical premise, that the specific forms of these relations cannot be deduced *a priori* from abstract economic laws, but must rather be made historically and spatially specific. From here we can pick up again Polanyi's formulation about the social and political constructedness of capitalist 'free' markets, to ask how are markets in cultural products constituted in particular space–time conjunctures – for example, through, global trade agreements, as Jameson elaborates; how and who are the winners and losers selected; and how, crucially, are countermovements for social pro tection anticipated, managed, and contained?

To understand fully such processes of capitalist market formation, however, requires achieving not only an economic and political unity, but also, as Gramsci argued, a moral and intellectual unity, involving the coordination of the interests of (some) subaltern groups with those of (some) dominant groups. In his reformu lation of Lenin's idea of hegemony, Gramsci was con cerned largely with the cultural processes by which such alliances among differently positioned social groups are built to form the political constituency for a particular market formation. The death of local film industries around the world cannot be achieved singlehandedly by Hollywood; it has of course required a range of state, multilateral, corporate, and civil society actors heralding from different national political contexts and seeking to promote a particular configuration of film production and exchange; and to understand how it could be sustained politically requires understanding the enormous appeal of Hollywood film and television to viewing audiences occupying widely varying class, ethno cultural and spa tial positions. Their addictions constitute a significant dimension of the sort of American cultural hegemony that concerns Frederic Jameson, and illustrate the role of strategic alliances in forging hegemony. As Hall suggests, the question then becomes, how are such alliances ac tively constructed and positively maintained in historic ally specific (and temporary) 'moments' in the life of a society? What sort of cultural work is entailed in winning over the hearts and minds of subordinate groups to particular capitalist formations so that relations of dom ination can be secured by consent, not coercion?

Cultural industries lend themselves particularly well to exploring the ways in which processes of meaning making conjoin with capitalist social relations to consti tute the various capitalisms that have gone into the making of globalization. For these sectors make especially clear the connection between the production of symbolic value and the creation of the popular habits and spaces necessary to sustain profitable commodity exchange. Economic geographers have gone a long way in showing through studies of fashion, home furnishings, film, and so on, how capital takes seriously and makes explicit the cultural work entailed in commodifying distinction and difference as a basis for economic profit. But this sort of relationality between economic practice and meaning making is not unique to cultural industries. The production and exchange of goods and services is always intimately bound up in the production of ways of doing and thinking. These ways of doing and thinking may contribute to the formation of historically and spatially specific hegemonic alliances insofar as they get taken up as practical common sense within particular social groups in particular space time conjunctures (or they may not). And they will often have economic effects, in terms of patterns of consumer demand or the competitiveness of particular firms, regions, commodities, or services.

The limitations of a Gramscian framework lie in the way that the notion of hegemony construes subjectivity, by suggesting that people either consent to the exercise of power or they resist it. An analytic grounded in the work of Michel Foucault shares an interpretation of capitalist relations as a vehicle not just for extraction, but also for imparting ways of doing and thinking. And it too is concerned with how the exercise of power is achieved in the absence of coercion or force. But, as anthropologist Tania Li has argued in a study on the *Will to Improve*, within the field of power that he termed 'government' Foucault understood subjects to be formed by practices of which they may not be aware, and to which their consent is neither given nor withheld. And he highlighted the ways in which power enables as much as it constrains—through, as he succinctly put it, 'the conduct of conduct', the rationalities of rule, and assemblages of techniques that cultivate the conditions in which subjects live their lives. Thus it is not only modern markets that are socially constructed, but also modern subjects, and these are relational processes. One could explore the relationship between television advertisements aired during cartoons, or the practice of contracting soft drink vending machines in public venues such as schools or zoos, or even cultural products themselves, with the constitution of the consuming subject. Or how the idea and practice of the living wage was coproduced with the masculinist laboring subject within particular Fordist regimes of accumulation. Or the role of microfinance programs in community development programs all over the world in constituting the entrepreneurial subject. The Foucauldian notion of governmentality thus encompasses but decenters the role of the state in the practice of government, making room for other sorts of actors such as corporations, nongovernmental organizations, or labor unions to play a role in projects of rule. And it expresses well the temporary and provisional nature of such projects, which must be actively created and constantly reworked, and are always subject to critique and resistance by those whose conduct is being conducted.

Thus far we have been focusing on the production of globalization through hegemonic alliances and regimes of rule. And, as Hart has pointed out, it is easier to recognize the contingency and situatedness of globalization from the standpoint of its production than from the standpoint of its experience or reception. The articulation approach to cultural globalization also takes up the question of 'reception', but rejects top down impact models of change by focusing instead on the interrelations between wide scale macroeconomic and associated ideological processes with cultural practices and social formations (understood themselves as interrelated with state and economy) in particular regions of the world. Here again it may be useful to return to Bourdieu's understanding of culture as practices of meaning making that are inextricably linked to the material circumstances and dynamics of power within which lives are lived. Anthropologist Mark Liechty's study of consumer culture in Kathmandu offers a good example of how such an interpretation of culture might be engaged to investigate the articulation of globally circulating images and commodities with a particular economics of practice. In *Suitably Modern*, he describes how middle class Kathmandu residents have since the lowering of import bans in Nepal in the 1980s been engaging commodity consumption as a marker of class distinction by participating in a rapidly expanding commoditized regime of value, in which personal prestige and identity are related to the control and display of possessions. Though many people describe the practices of 'doing fashion' or hosting extremely costly 'buffet style' parties as intensely stressful, they feel they have no choice but to participate for fear of losing their place in a consumption based prestige economy. For others such as low castes the new social currency of consumer goods presents opportunities for transcending low status positions associated with previously salient forms of prestige such as caste.

Here we see how a global cultural economy of consumer goods gets harnessed to the imperatives of place based cultural projects (in this case, projects of subject formation through class practice). The global mobility of cultural forms and products may have entailed processes of deterritorialization, the dislocation of culture from centers of cultural production (which, it must be emphasized, are multiple – Asian, European, Latin American, African, Middle Eastern as much as American). But, as anthropologists Inda and Renato Rosaldo have argued, consumer goods and other cultural forms are always reterritorialized through distinctive time space contexts of consumption. Returning now to a Gramscian analytic, the question is, to what extent do such reterritorializations reinscribe existing hegemonic formations or create the conditions for moments of critical consciousness? In the Kathmandu case noted above, for example, we find a simultaneous deepening of class hierarchies and loosening of caste based modes of distinction. The feelings of frustration, even violation, that accompany middle class peoples' embrace of consumer lifestyles provide grist for critical insight into the way their subjectivity as consumers is being constituted. But this critical consciousness has for the most part not catalyzed individual tactics of subversion or more potent and collective forms of resistance to the new commoditized regime of value. A Gramscian analytic thus demands a focus on how and why particular situated subjects do or do not mobilize to contest their oppression. With respect to these questions, the primary methodological act then becomes learning to read the modes of representation that convey contradictory consciousness and to identify the social and political forces through which countermovements for social protection are constituted in any particular historical–spatial conjuncture.

The articulation approach thus requires a processual and relational understanding of cultural globalization that explores how multiple forces come together in practice to produce particular dynamics or trajectories. The focus here on interconnections among multiple historical/geographical determinations in fact displaces 'globalization' as a suitable metaphor for understanding the historically specific collaborations that create distinctive cultural forms of capitalism. The term 'globalization' reinscribes local/global dichotomies and remains mired in an interpretation of history as teleological abstraction. In an articulation approach, all regions of the world are seen as both local and global. The emphasis, as Doreen Massy put it in her influential and important book, *Space, Power and Gender*, is on how places are set in the context of forces and relations that lie both within them and beyond them, and how therefore the particularity of any place must be recognized precisely through the specificity of the mix of interconnections to that 'beyond'. In this view, places are imagined not as bounded 'local' entities impacted by outside 'global' processes, but as open and porous networks of social relations, a large proportion of which are constructed not 'inside' that place but on a larger scale. The 'global' and the 'local' are just names we give to different aspects of these intersecting relations.

Conclusion: The Cultural Politics of Globalization

The goal in these concluding remarks does not summarize the argument presented thus far, but focuses specifically on the political implications of the three approaches to cultural globalization, and in particular on the political implications of the meanings they convey about the nature of globalization and of the cultural. The limiting and ambiguous politics of the flows and impacts and cultures of economies approaches will be noted briefly, with the remainder of the conclusion elaborating methodologies and political praxis associated with the articulation approach.

Within the flows and impacts approach we have noted that globalization is viewed as a process of flow from centers of economic and cultural production to diverse sites of reception around the world. And 'the cultural' is construed in rather essentialist and static terms as cultural signs and symbols, and the meanings and values associated with them; even if new cultural hybrids are produced at the margins and in liminal spaces, we are still talking here about culture as text. This approach lends itself to a range of moral valuations – the celebration of cultural difference, the fear of cultural homogenization, the embrace of opportunities entailed in the expansion of capitalist markets, or the critique of the ways in which globalization forces the periphery to integrate into the capitalist world system. The stance within the cultures of economy approach expresses a more consistent critique of capitalist relations of production and other systems of domination, and some of the work within it introduces an analytically more useful interpretation of culture as processes of meaning making. But it encompasses a range of views on how the economic articulates the cultural – as an intersection, a determination, a mutual embeddedness – with competing interpretations of the ideal sites and strategies for political change. And in its imputed separation of the economic from the cultural this epistemological framing is open to appropriation by more regressive normative agendas that would seek to claim culture as a 'resource' for deepening capitalist markets and subjectivities.

The articulation approach rests on firmer normative ground. It also takes up an interpretation of culture as processes of meaning making, but consistently seeks to ground such processes in situated material practices and their associated power relations. This commitment informs an understanding of globalization not as inexorable force, but as webs of interconnection among concrete places in the world, which are themselves constituted through social relations operating at multiple spatial scales. It also demands attention to the constellation of social relations and technologies of government through which dominant cultural productions are secured, as well as to the conditions in which particular situated subjects mobilize to contest their oppression. And it demands research methodologies suited to tracing how the playing out of social relations across scales constitutes, as Hart puts it, diverse and interrelated trajectories of social change, for the explicit purpose of drawing inferences about directions for political praxis.

Two recent critical ethnographies of global connection are worth mentioning in this regard. Cindi Katz' *Growing Up Global* (2004) draws a contour line that links the fate of children growing up in the Sudan and the United States, as wide scale capitalist processes articulate practices of social reproduction. While attending to the specificities of the two historical geographies, the suggestion here is that unexpected similarities in experiences as well as consciousness of economic restructuring offer common grounds for political response across scale and space – for, as she puts it, a countertopography of global capitalism. James Ferguson's *Global Shadows* (2006) offers a more disjunctive view of Africa's global interconnectedness: as he puts it, networks of political–economic connectedness span the globe but do not cover it; rather they hop over large (African) territories and populations to land in spatially dispersed centers of capitalist extraction. Such highly selective and nodal forms of global connection leave most Africans with only a tenuous connection to 'the global economy'. Both books exhibit a relational and conjunctural approach to geography and history. If the trajectories of connection and change they describe are unique to particular political–economic and cultural–political conditions, then alternative trajectories are always possible. But they must, as Hart suggests, be produced on a terrain given by concrete conjunctures of multiple scaled social relations.

Several considerations suggest themselves for moving from an analysis of practice to a praxis of politics. First, as Ferguson persuasively argues, is to recognize how cultural practice always indexes membership in different and unequal social groups. Thus measures to 'respect difference' that are the typical focus of a politics of recognition all too often contribute to exacerbating the lines of socioeconomic inequality. At the same time, the contours of inequality in particular space time conjunctures are different; they are born out of distinct historical geographies. A second consideration then has to do with the challenge of finding commonalities within the differences, as Hart, Harvey, and others have argued, in order to develop a politics that is collective in its orientation, but that also remains sensitive to such irreducible distinctions. This involves, as Katz' notion of countertopography so powerfully evokes, thinking about how struggles in disparate places might unexpectedly link with one another. A final consideration must then attend to how subjectivity, sentiment, and critical consciousness are constituted under prevailing conditions – how a constituency for such a collective politics might form. As a Gramscian analytic makes clear, while capitalist processes may by definition create winners and losers, it is precisely the role of hegemony to make such arrangements appear natural and inevitable to adequate numbers of losers. The question then becomes, under what circumstances do those who are oppressed in particular socio spatial arenas come to recognize the arbitrary foundations of the prevailing system and mobilize to change their situation? And how might those circumstances be facilitated? Such considerations are crucial for the project of imagining and building alternative political economies based on solidarities across interconnected historical geographies.

See also: Cultural Economy; Cultural Politics; Cultural Studies and Human Geography; Globalization and Transnational Corporations; Globalization, Economic.

Further Reading

Appadurai, A. (1996). *Modernity at Large: Cultural Dimensions of Globalization*. Minneapolis, MN: University of Minnesota Press.
Crang, P. (1997). Introduction: Cultural turns and the (re)constitution of economic geography. In Lee, R. & Wills, J. (eds.) *Geographies of Economies*, pp 3 15. London: Arnold.
Ferguson, J. (2006). *Global Shadows: Africa in the New World Order*. Durham, NC: Duke University Press.
Gibson Graham, J. K. (2006). *A Postcapitalist Politics*. Minneapolis, MN: University of Minnesota Press.
Katz, C. (2004). *Growing up Global: Economic Restructuring and Children's Everyday Lives*. Minneapolis, MN: University of Minnesota Press.
Hall, S. (1980). Race, articulation and societies structured in dominance. *Sociological Theories: Race and Colonialism*, pp 305 345. Paris: UNESCO.
Hart, G. (2002). Geography and development: Development/s beyond neoliberalism? Power, culture, political economy. *Progress in Human Geography* 26, 812 822. (See also Hart's two other Progress Report contributions in 2001 and 2004.)
Harvey, D. (1990). *The Condition of Postmodernity: An Enquiry into the Origins of Cultural Change*. Oxford: Blackwell.
Inda, J. X. and Rosaldo, R. (2002). *The Anthropology of Globalization: A Reader*. Oxford: Blackwell.
Jameson, F. and Masao, M. (eds.) (1998). *Cultures of Globalization*. Durham, NC: Duke University Press.
Li, T. M. (2007). *The Will to Improve: Governmentality, Development and the Practice of Politics*. Durham, NC: Duke University Press.
Liechty, M. (2003). *Suitably Modern: Making Middle Class Culture in a New Consumer Society*. Princeton, NJ: Princeton University Press.
Massey, D. (1994). *Space, Place and Gender*. Minneapolis, MN: University of Minnesota Press.
Mitchell, K. (2003). Cultural geographies of transnationality. In Anderson, K., Domosh, M. & Thrift, N. (eds.) *Handbook of Cultural Geography*, pp 74 88. London: Sage.
Tsing, A. L. (2000). The global situation. *Cultural Anthropology* 15, 327 360.
Velayathum, S. and Wise, A. (2004). Moral economies of a translocal village: Obligation and shame among South Indian transnational migrants. *Global Networks* 5(1), 27 47.

Relevant Websites

http://www.communityeconomies.org
Community Economies.
http://www.darkmatter101.org/site/2007/02/28/stuart hall globalization cartographies of power/
Darkmatter Post colonial Futures.
http://www.boell meo.org/en/web/73.htm
Heinrich Boll Foundation.
http://www.globalization101.org/issue/culture
The Levin Institute.

Globalization, Economic

H. W.-c. Yeung, National University of Singapore, Singapore

Introduction

"Geography is still important. Globalization has not diminished the economic significance of location" – John Kay, one of Britain's leading economists, proudly declares in the headline of his fortnightly column in the world's famous pinkish newspaper, *The Financial Times (FT)*. The fact that business gurus like John Kay write about globalization is nothing new; indeed, prominent geographers, such as Peter Taylor, Kelvin Cox, and Peter Dicken, have associated the dreaming up of 'globalization' as the lexicon in the new century with such business gurus (as well as chief executive officers (CEOs), global financiers, business schools, and international organizations). What is particularly interesting in Kay's FT column, however, is his unreserved defense for the importance of place and location in today's globalization. Coming from outside geography as an academic discipline, his views on globalization provide significant legitimacy for the claim that geographers might have some very useful things to say about globalization.

The missing link in Kay's rather polemical acknowledgment of the importance of geography in globalization, however, is what constitutes this importance. Does it refer to geography as preexisting spatial configurations that simply do not wither away in the midst of globalization processes? Or perhaps more significantly, what is the explanatory relationship between geography and globalization? Is geography simply an outcome of globalization that poses itself as the cause? How does geography matter in our explanations of globalization? In their laudable efforts to relate geography to globalization, Taylor *et al.*, in 2001, noted that "[w]hatever your own opinion may be, any intellectual engagement with social change in the twenty first century has to address this concept [globalization] seriously, and assess its capacity to explain the world we currently inhabit." Attempting to assess the capacity of globalization as a concept and a set of tendencies to explain the geographical world is the primary focus of this article.

Economic globalization, by its very nature, demands serious attention from geographers. The processes that have, over the last few decades, led to dramatic increases in the intensity and extensity of international economic interdependence and integration are inherently and unavoidably spatial in character. Any international economic relationship – be it associated with trade, investment, finance, labor, or technology – is a reflection of, and partly constituted by, spatial difference. Geographers have thus coined the metaphor 'mapping globalization' to highlight what we see as a pressing need for globalization studies that analyze the phenomenon from theoretically and empirically rigorous geographical perspectives. As argued passionately by Amin, in 2001, "the distinctive contribution of this discipline [Geography] within the congested study of globalisation, has lied in the study of the spatiality – social, economic, cultural and political – of what is increasingly being seen as a single and inter dependent world."

Indeed, the first step toward undertaking effective mappings of globalization is to be explicit about how the phenomenon is being conceptualized and delineated. In one of the earliest geographical contributions to globalization debates, Dicken *et al.*, in 1997, argued that "definitions are not mere semantic peccadilloes; they remain crucial not least because they caution against the kind of caricaturing of globalization which has become all too common." The tendency in the media, and certain political and academic circles, to label simply any kind of 'international' economic relation as indicative of economic globalization quite simply destroys its usefulness as an analytical category. In other words, we have to be clear about what we mean by economic globalization and how it relates to, and contrasts with, other forms and processes of international economic restructuring. This is, of course, not a new argument – albeit, it is still not made forcefully enough – and there is now a wealth of critical literature that can help in formulating effective and usable definitions of globalizing processes.

Rather than formulating a definitive framework on geography and economic globalization, this article seeks to open up spaces for debates among geographers and social scientists interested in studies of global economic change. It contributes to the current efforts in critical geography to move beyond 'first wave' analyses of globalization as a geographical phenomenon. Cognizant of the enormous amount of globalization studies, this article is not intended to be a critical assessment of the globalization literature that is extremely diverse and, perhaps, chaotic. Instead, this author aims to develop a geographical perspective on globalization and argues that globalization is an inherently geographical phenomenon, and therefore, its alleged causality and outcomes might be better understood and accurately assessed from this perspective. Globalization processes are conceived as spatial tendencies contingent on, and accounted for by, certain stringent and necessary requirements, some of which are geographical while others are political, economic, and/or technological.

Contrary to the efforts of what sociologists Mike Featherstone and Scott Lash called "the spatialization of social theory," this geographical perspective attempts to destabilize the hegemonic mobilization of globalization as the universal explanation of empirical outcomes. Our analytical concern should not be about explaining how globalization is reshaping our lives – a key observation made by most social/globalization theorists. Rather, geographers should be interested in what geographical preconditions enable globalization to take place in tandem with this reshaping of social life – a geographical theory of globalization. This geographical perspective views the spatiality of globalization as an outcome of social constructions of space that are mediated through historically specific political, economic, and technological forces. As a description of spatial tendencies, globalization does not have an independent existence outside these forces, and thus cannot be mobilized to explain empirical outcomes in the absence of these forces. This perspective can make a potential contribution to demystifying, or at least qualifying, the alleged explanatory power of globalization and space. Before we proceed further, it is useful to recap geographers' participation in the study of globalization.

Globalization as Both Material and Discursive Processes

Anyone working on globalization these days immediately runs into a seemingly intractable problem of defining globalization. The definitions of globalization in the literature unfortunately remain largely elusive, vague, and, in sociologist Bob Jessop's view, 'chaotic'. It is now generally agreed among human geographers that globalization should be viewed as both a set of material processes of transformation and resistance, and a set of contested ideologies and discourses that operate across a variety of spatial scales. Globalization is certainly not akin to a mega trend 'out there' to be grappled objectively by social scientists. We cannot define globalization as concrete processes of socioeconomic and political restructuring on a global scale without underscoring the ideological foundations on the basis of which these processes operate. This definitional problem makes it incumbent upon us to present a very specific and precise definition of globalization that hopefully avoids the problem of 'chaotic concepts'.

This article focuses more specifically on 'economic' globalization, defined as the rapid proliferation of cross border production, trade, and investment activities spearheaded by global corporations and international financial institutions that facilitate the emergence of an increasingly integrated and interdependent global economy. There are three specificities to this definition of economic globalization. First, it includes only material processes of economic transactions and linkages across borders, and excludes other globalization imperatives that may have economic consequences (e.g., migrations and civil society networks). Second, the definition encapsulates an actor oriented approach to understanding globalization by focusing specifically on the transformative role of global corporations and international financial institutions. This specificity avoids the flawed structural view of globalization as some sort of mega trends 'out there'. It allows us to bring back the Prince of Denmark (social actors) in the discussion of *Hamlet* (globalization). We need to identify the 'captors' and 'captives' of globalization. Third, it explicitly refers to the global scale as critical in defining 'global'ization. Whatever its constituents, contestations, and consequences at different geographical scales, globalization is necessarily an integrating set of tendencies that operate on the global scale and intensify connections and flows across territorial borders and regions. This narrower definition of globalization avoids the dangerous tendency to use 'globalization' as a catchall phrase for describing all kinds of contemporary processes of change. Despite its causal links with the contemporary capitalist imperative, globalization is certainly not just about the latest phase of uneven geographical development, de/reterritorialization, or crisis induced capitalist restructuring. Globalization, indeed, has much deeper historical roots in the founding of the modern nation state during the seventeenth century.

Although *Global Shift*, the now influential text on globalization by an economic geographer Peter Dicken, was published as early as 1986, geographers had not been very evident in the globalization debate until the early 1990s when 'the end of geography' thesis became increasingly popularized in the media, policy circles, and even academic worlds. In its essence, this thesis argues that the juggernaut of globalization, as a planetary force, is capable of penetrating all kinds of national boundaries and eroding any geographical differences therein. The convergence effects of globalization are so strong that there is no place for geography in the processes of global change. The critique of this strong convergence thesis of globalization has reverberated seriously in the social sciences, and notably, in human geography. Globalization has seemingly led to significant material transformations in the global economy, most appropriately termed 'global shift'. These global transformations entail the rapid proliferation of cross border trade and investments by global corporations and financial institutions, the ruthless penetration of global cultures epitomized by McDonald's, Hollywood movies, MTV, and the Internet, and the reluctant power shift from nation states to global governance.

These global transformations, nevertheless, are not geographically even and without their resistance at different spatial scales and in different countries/regions. The global reach of corporate activities has failed to transform the world economy into a singular global production factory. In fact, what appears to be more convincing, according to Allen Scott and Michael Storper, is the phenomenon of regionalization through which regions have emerged as the major motor of the global economy. Global production activity seems to be taking place in high tech regions such as Silicon Valley (northern California), Third Italy (Emilia Romagna), and Baden Württemberg (southern Germany) and low cost regions such as Dongguan (China), Penang (Malaysia), and Bangalore (India). Global finance remains highly rooted in existing global finance centers such as the City of London, New York, and Tokyo, and new international financial centers as Hong Kong and Singapore. Geographers have, therefore, argued for an examination of specific regions and territorial ensembles in order to appreciate the inherently geographical nature of globalization processes. Other geographical studies have shown that while globalization has been spearheaded by the cross border operations of transnational corporations, the spatial transfer of business and industrial practices is by no means unproblematic. There remain significant place based institutional limits to the globalization of business cultures and economic practices. For example, while capital can be transferred almost effortlessly across space, labor remains highly place bound and locally embedded. In other words, globalization encounters most 'friction' over space at the local and regional scales. It is clearly unable to erase spatial differentiation even in the realm of economic processes and governance.

The arrival of global cultures has further accentuated the awareness of local differences and cultural responses. While the global flows of information, images, and artifacts are greatly facilitated by the advancement in telecommunications and transport technologies, the recipients and consumption of these flows remain highly territorialized and embedded in specific geographical boundaries. Although few people in developed countries do not have access to global cultures in their embodied and commodified forms, a large segment of the population in developing countries is still highly dependent on localized products and cultural services. Most Filipinos and Brazilians, for example, are still drinking their local beverages instead of Coca Cola and Pepsi. Furthermore, global cultures are equally resisted in many countries that view these cultural inflows as threatening to their national interests and political legitimacy. There are countermovements to the globalization of cultures through the reassertion of local/national cultures, religions, and other forms of social identities. Despite the dominance of American movies in the English speaking world, the local movie industry in India and Hong Kong continues to thrive from strength to strength. These countermovements, enmeshed in the global web of cultural flows, have created a world of cultural mosaics and differences. We are certainly still far away from a culturally homogeneous world of a global village.

The political economy of globalization does not necessarily indicate the demise of nation states as the primary locus of political governance. Some geographers have argued that the world is not becoming more unruly and disorderly in the governance sense. Rather, there is a reconfiguration of political governance and institutional power through which the nation state is joined by supranational institutions and others institutions (e.g., nongovernmental organizations) to govern globalization. Just witness the immensely complex organization behind those antiglobalization protests in Seattle (USA), Melbourne (Australia), and Genoa (Italy), and the rapidly evolving World Social Forum. Nation states, worldwide, are experiencing a relative decline in their capacities to control the whereabouts of global corporations, to provide social welfare for their citizens, and to contain the post Cold War world threats. These states, however, are not fixed entities to be overwhelmed by the unruly processes of globalization. Indeed, some nation states are active agents in promoting globalization, for example, through the liberalization of financial markets. Still, other nation states have globalized themselves by building interstate consensus and contributing to trans state sources of power and authority. It is, therefore, too premature to proclaim the death of nation states.

Globalization is also as much a set of material processes as a set of contested ideologies and discourses. The ways through which globalization is represented can have equally significant impact on material processes. While we cannot equate globalization with neoliberalism *per se* (i.e., defined as a political–economic ideology in favor of market mechanisms in *lieu* of state intervention), it is true that neoliberal programs of policies have greatly facilitated the advent of globalization. Just think of how liberalization and market based economic policies have enabled certain giant corporations to emerge and dominate the global economy. In other words, globalization cannot proceed smoothly without its supporters and champions. As expected, business gurus, media pundits, and policymakers (and even academics) are often the strongest supporters of globalization. In championing for globalization and making it appear a natural force, these people have 'naturalized' globalization processes and portrayed them as 'necessary', 'inevitable', and 'beyond our control'. Powerless citizens have to choose either wholesale embracing of globalization or economic decline and social exclusion.

Toward a Conceptual Remapping of Economic Globalization

What geographers have plainly shown in their recent work is that this either/or ideological choice of globalization should be seriously questioned. In fact, localities can reassert their power in the global economy through certain discursive practices and/or constructing alternative globalizations. These practices include the building up of social capital, the experimentation with nonmarket forms of economic exchanges, the reorganization of local work practices, and so on. To challenge the ideological supremacy of neoliberalism and its much broader processes of globalization, some localities and communities have actively demystified globalization as a necessary and inevitable phenomenon. The rise of local civil societies in many parts of Western Europe, for example, is a direct answer by localities to regulate the apparently unruly effects of globalization.

To sum up this discussion so far, globalization is understood in geography as a set of material processes and discursive practices that operate across different spatial scales. The differences that these spatial scales make are that globalization is about changing relationships between geographical scales and that these scales are socially constructed through ideologies and discourses. Geographers' greatest contribution to the globalization debate is the recognition of how spatial scales matter in our understanding of the complex operations of globalization processes. Globalization is sometimes highly localized, whereas in other times, it is a region that witnesses most of the effects of globalization. Viewed in this geographical perspective, globalization is 'not' necessarily an essential global force that homogenizes national and local differences. It has as many local and regional dimensions as being a global force.

What we discern to be the most effective characterization of globalization can be summarized by the following set of interrelated assertions. This conceptual mapping of globalization places us unequivocally in David Held's 'transformationist' segment of the literature that sees globalization as an ongoing transformative process, rather than either an unattainable end state of universal interconnection ('the skeptics'), or an inexorable global force that will inevitably eradicate local differences and put paid to the nation state ('the hyperglobalists').

- Globalization is constituted by a set of processes that are increasing in extensity and intensity, but are not even approaching any kind of homogenized end state of universal global interconnection. As a concept, it represents a 'qualitative' change in the nature of the global system rather than an epochal shift to a global mode of organization.
- Globalization is best represented as a relative increase in the functional integration of economic activity across national borders (i.e., characterized by complex interdependencies) and is, therefore distinctive from simple internationalization. Transnational corporations are playing a relatively more important, but by no means hegemonic, role in these processes that are uneven both across and within economic sectors.
- Globalization is not spatially homogenizing, but instead depends upon, and contributes to, uneven geographic development at different scales. The spatial economic pattern that results can perhaps be best described as a mosaic of interconnected national and subnational clusters of activities. Different territories are integrated into the global system to greater or lesser extents, and some are currently excluded almost entirely.
- Globalization is both leading to, and partly constituted by, a mutating of the role of the nation state, which is not declining appreciably in importance. However, there has been a degree of 'hollowing out' of functions to both supranational and subnational organizations.
- Globalization is, however, resulting in a reworking of the different scales of economic activity. Two scales, in particular, have seen a relative rise in importance as a result: (1) the global city region is argued by many to be an increasingly important scale for the coordination, regulation, and promotion of economic activity; (2) the three triadic regions – North America, Western Europe, and Japan/East Asia – are also becoming clearly visible, although more perhaps as scales of political and corporate integration rather than cohesive trade and investment blocs.
- Globalization is as much a neoliberal discourse as a set of material processes. The two domains interact in complex but often mutually reinforcing ways. For example, certain neoliberal discourses 'naturalize' processes of globalization, thereby constructing notions of economic space that lead to a deferral of political options to the global scale.
- Globalization is not some kind of immutable inevitability, but a set of processes that is socially constructed, and therefore, can be encouraged or resisted by actors/institutions at various scales. Globalization is, therefore, a set of complex and conflicting tendencies, the outcomes of which often cannot be predicted *a priori*.
- Globalization is not significantly eroding the importance of national business systems, understood as distinctive configurations of institutionalized economic activities. Rather than being replaced or eroded by globalization tendencies, national business systems are experiencing highly variable processes of reconfiguration as they are integrated at the international scale in a highly uneven manner.
- Globalization is constituted and reflected by qualitative changes in nature of the organization of economic

activity, often associated with a rise in business net working, in particular, through intra firm (exploiting core competencies) and inter firm networks (equity and nonequity arrangements, such as subcontracting, strategic alliances, etc.).

Bringing Geography Back Into the Globalization Agenda

What then are the implications of this article for future geographical studies of global economic change? There are important geographical preconditions of globalization worthy of further theorization. Ash Amin has developed a relational concept of place in his theorization of the re lationship between geography and globalization. He argues that "a relational or global sense of place will interpret the spatiality of globalisation – however defined – rather differently from a territorial perspective." This relational perspective on the spatiality of globalization views locales, such as cities, as "a meeting place for di verse geographies of linkage" that clearly go beyond a territorial and scalar reading of the position and function of urban economies in globalization. Other geographers have recently developed the geographical concepts of global city regions as "points at which globalization processes crystallize out on the geographic landscape and as active agents in shaping globalization itself" (Scott, 2001: 7) and global cities as relational networks.

Whether these concepts herald the sociospatial con struction of new geographical scales is a moot point here. But what is clear in these conceptual interventions is the explanation of globalization tendencies in relation to their various sociospatial formations, not the other way around as so often done in globalization theory where globalization is pretended to explain those very for mations. Space does play an analytical role in explaining globalization tendencies, although this role is not ex clusive to geography, or other social sciences for the same matter. Sociologist Saskia Sassen, in 2007, noted that "Two disciplines more than any others have contribu ted significantly to the study of the global as it gets constituted subnationally. They are geography and anthropology, specifically, particular branches of each. Economic and political geography have done so espe cially through a critical development of scale and scaling. This work recognizes the historicity of scales and thus resists the reification and naturalization of the national scale as present in most of social science."

Another important implication is related to an agency focused approach to globalization research that places more significance on actor transformation and resistance. Much of the existing globalization literature takes on an abstract macro and structural view of the global economy. What is missing in this globalization agenda is the attention to social actors who are not only constructing globalization in their various capacities, but also experiencing significant transformations in their own everyday social life. By focusing on these actors in globalization processes, we can make a much better connection between the apparently nonhuman processes of globalization and their manifestations in all sorts of human activities, practices, and struggles. This approach also allows us to study the contested sociospatial pro duction of scales and globalization discourses. In short, we have to pay attention to globalization as a 'living' experience, which entails both transformations of social practices and resistance from social actors. The empirical orientation of geography as an intellectual discipline allows us to bridge the still largely unfilled gap between overwhelmingly stylized abstractions in most global ization studies and the contradictory experiences of most people who are living with economic globalization as an everyday fact.

Now that we have a better appreciation of the geo graphical specificity of globalization, it is perhaps time to move beyond the simplistic notion of globalization as merely a set of end state or 'placeless' phenomena. There are many geographical questions that should inform fu ture research on globalization and four are highlighted here. First, we need more balanced, empirical assess ments of globalization processes. At both research and policy levels, it is really not helpful just to embrace globalization wholeheartedly or to condemn it un reservedly. We understand from previous studies that the Janus face of globalization may simultaneously benefit as well as harm localities. It is imperative for future researchers to continue to evaluate the uneven geo graphical outcomes of globalization processes in order to arrive at better informed development policies.

Second, while there is a wealth of 'top down' studies of globalization, particularly those measuring the extent of globalization via specific economic indicators, there is clearly room for further 'bottom up' studies of global ization that take a more agency oriented approach. We certainly need to continue to understand and critique the strategies of global corporations and international or ganizations. Equally important, however, is the pressing need – both intellectual and social – to appreciate how globalization is contested at various scales by social actors, such as cocoa farmers in West Africa and auto mobile component suppliers in Germany and Detroit, whether at the level of the individual household decision, the workplace, or the international mobilization of labor unions.

Third, we need to think creatively about the new kinds of spaces that are being created by globalization. For example, anthropologist Aihwa Ong talks in concrete terms about areas of 'graduated sovereignty' – such as export processing zones – where being connected to global networks is predicated upon the existence of customized regulatory and disciplinary regimes. At a more abstract level, John Beaverstock, Richard Smith, and Peter Taylor describe the emergence of 'translocalities' within major cities, spaces that are relatively more connected into transnational networks than 'local' relations within the urban area. Relational, network based views of space will need to be developed to conceptualize how globalization is carving out new kinds of regulatory and lived spaces.

Finally, future researchers need globally coordinated research that is well executed, locally. In other words, globalization research requires the construction of international networks that parallel, overlie, and criss cross the very global networks we are concerned to investigate. Only through such cross national/regional efforts can we improve our chances of coming to terms with globalization as a complex set of phenomena that are increasingly impinging upon everyday lives in most, if not all, parts of the world.

See also: Global Commodity Chains; Global Production Networks; Globalization and Transnational Corporations; International Organizations; Neoliberal Economic Strategies.

Further Reading

Amin, A. (2001). Globalization: Geographical aspects. In Smelser, N. & Baltes, P. B. (eds.) *International Encyclopedia of the Social and Behavioural Sciences*, pp 6271 6277. Amsterdam: Elsevier Science.

Amin, A. (2002). Spatialities of globalisation. *Environment and Planning A* 34(3), 385 399.

Amin, A. (2004). Regulating economic globalization. *Transactions of the Institute of British Geographers* 29(2), 217 233.

Beaverstock, J. V., Smith, R. G. and Taylor, P. J. (2000). World city network: A new metageography? *Annals of the Association of American Geographers* 90(1), 123 134.

Cameron, A. and Palan, R. (2004). *The Imagined Economies of Globalization*. London: Sage.

Coe, N. M. and Yeung, H. W. c. (2001). Geographical perspectives on mapping globalization. *Journal of Economic Geography* 1(4), 367 380.

Cox, K. R. (2004). Globalization and the politics of local and regional development: The question of convergence. *Transactions of the Institute of British Geographers* 29(2), 179 194.

Dicken, P. (2004). Geographers and 'globalization': (Yet) another missed boat? *Transactions of the Institute of British Geographers* 29(1), 5 26.

Dicken, P. (2007). *Global Shift: Mapping the Changing Contours of the World Economy* (5th edn.). London: Sage.

Dicken, P., Peck, J. and Tickell, A. (1997). Unpacking the global. In Lee, R. & Wills, J. (eds.) *Geographies of Economies*, pp 158 166. London: Arnold.

Economic Geography (2002). *Special Issue on Global Economic Change. Economic Geography* 78(3), 253 386.

Elden, S. (2005). Missing the point: Globalization, deterritorialization and the space of the world. *Transactions of the Institute of British Geographers* 30(1), 8 19.

Held, D., McGrew, A., Goldblatt, D. and Perraton, J. (1999). *Global Transformations: Politics, Economics and Culture*. Cambridge: Polity.

Kelly, P. F. (1999). The geographies and politics of globalization. *Progress in Human Geography* 23(3), 379 400.

Ong, A. (2007). Neoliberalism as a mobile technology. *Transactions of the Institute of British Geographers* 32(1), 3 8.

Peck, J. and Yeung, H. W. c. (eds.) (2003). *Remaking the Global Economy: Economic Geographical Perspectives*. London: Sage.

Sassen, S. (2007). Introduction: Deciphering the global. In Sassen, S. (ed.) *Deciphering the Global: Its Scales, Spaces and Subjects*, pp 1 18. New York: Routledge.

Scott, A. J. (ed.) (2001). *Global City Regions: Trends, Theory, Policy.* Oxford: Oxford University Press.

Scott, A. J. and Storper, M. (2003). Regions, globalization, development. *Regional Studies* 37(6/7), 579 593.

Smith, N. (2004). *The Endgame of Globalization*. London: Routledge.

Taylor, P. J. (2007). Problematizing city/state relations: Towards a geohistorical understanding of contemporary globalization. *Transactions of the Institute of British Geographers* 32(2), 133 150.

Taylor, P. J., Watts, M. J. and Johnston, R. J. (2001). Geography/globalization. *GaWC Research Bulletin*, No.41. Department of Geography, Loughborough University.

Walker, R. (1999). Putting capital in its place: Globalization and the prospects for labor. *Geoforum* 30(3), 263 284.

Warwick, M. (2006). *Geographies of Globalization*. London: Routledge.

Yeung, H. W. c. (1998). Capital, state and space: Contesting the borderless world. *Transactions of the Institute of British Geographers* 23(3), 291 309.

Yeung, H. W. c. (2002). The limits to globalization theory: A geographic perspective on global economic change. *Economic Geography* 78(3), 285 305.

Yeung, H. W. c. (2003). Globalization. In Rogers, A. & Viles, H. A. (eds.) *The Student's Companion to Geography* (2nd edn.), pp 1103 1107. Oxford: Blackwell.

Relevant Websites

http://www.johnkay.com
Jonkay.

Golledge, R.

G. Rushton, The University of Iowa, Iowa City, IA, USA

Glossary

Analytic Behavioral Geography The investigation of human activities in a time–space framework, focusing on the way that spatial decision making is guided by cognitive processes.

Cognitive or Mental Maps The internalized representations of space which guide everyday wayfinding and decision making.

Consumer Spatial Behavior The rules by which people select one place over another in which to consume products.

Nonmetric Multidimensional Scaling A method for inferring locations of points from independent, ordinal, measures of relationships between them.

Positivism A philosophy of science in which individual observations are seen as instances of universal laws which are developed by induction and shown to be related by deductive, usually mathematical, principles.

Spatial Cognition The process of acquiring, coding, using, and storing information from the multitude of environments external to mind.

Spatial Structures The geometrical relationships between elements in time and space.

Wayfinding The process by which an individual processes information about their current location in relation to a second location to which they wish to reach in an economical way.

Born in Dungog, New South Wales, Golledge (**Figure 1**) received his BA (honors in geography) at the University of New England, Armidale, NSW in 1959 and MA, geography, at the University of New England in 1961. He received his PhD at the University of Iowa in 1966, subsequently becoming a leader in the field of analytic behavioral geography for 40 years, defining its links with the associated disciplines of psychology and, more recently, computational science. As a result, he received an honorary PhD in psychology at Goteborg University, Sweden, in 2001, and an honorary LLD from Simon Fraser University in Canada in 2001.

Golledge was a senior demonstrator in geography at the University of New England in Australia when Harold McCarty of the University of Iowa met him and persuaded him to go to Iowa to pursue a PhD. At Iowa, Golledge studied how human spatial behavior contributed to the organization of spatial structures. His early focus was on how spatial choices supported specific characteristics of urban settlement patterns and intraurban spatial structures. He noted that spatial structures are maintained by the spatial choices people make as well as the locations of facilities selected by businesses, governors, and the public. These structures in turn constrain the spatial choices that people make.

As a PhD student at Iowa in the summer of 1964, Reg Golledge and Gerard Rushton, along with W. A. V. Clark (recently graduated from the University of Illinois), worked together on a research project under the supervision of Dr. Ron Boyce, who gave them the flexibility to pursue ideas on how human spatial behavior contributes to urban settlement patterns, based on the idea that people make decisions according to their perception of the world, not its reality. They found considerable empirical support for the idea that there are reciprocal relationships between behaviors and spatial structures. This became a lifelong theme that gave theoretical coherence to Golledge's work.

In collaboration with Kevin Cox, he subsequently brought widespread professional recognition to the work of behavioral geographers who worked in the analytic mode by organizing sessions at the Association of

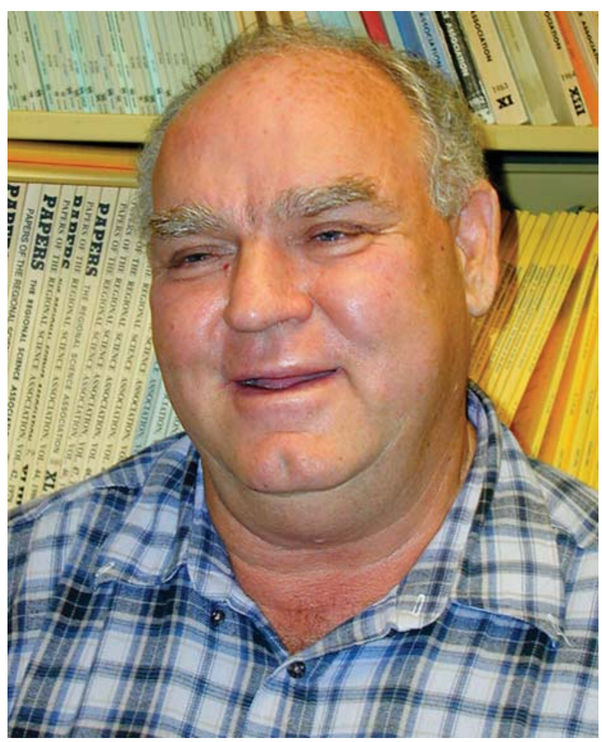

Figure 1 Reginald G. Golledge.

American Geographers meetings in the late 1960s and producing a widely cited collection *Problems in Behavioral Geography* that was republished in 1981.

In the 1970s, as an associate professor at Ohio State University, his work focused on decision making processes and locational behaviors. He became interested in cognitive images of environments and in developing theories for how these knowledge structures are formed and changed as people gain more experience of their environments. There clearly was a measurement problem with gaining knowledge of cognitive environments and representing this knowledge in ways that were helpful to interpretations of observed spatial behaviors. He introduced nonmetric, multidimensional scaling, developed in psychometrics, as a method for inferring spatial structures from survey data from people about their knowledge of their local environments.

Over time, his interests broadened beyond consumer spatial behaviors to migration and mobility research as he noted that cognitive images appeared to be fundamental to the understanding of many kinds of human spatial behavior. During this time, he became the second editor of the journal, *Geographical Analysis*, which introduced much of the new, quantitative, and theoretical research that characterized human geography in the 1970s.

By the late 1970s, analytic behavioral geography was under fire from a group of scholars who were critical, to say the least, of positivism and its role in research in human geography. With Helen Couclelis, Golledge defended analytic human geography as it had evolved, and argued that much of behavioral geography, as practiced by the late 1970s, had transcended its logical positivist roots. While acknowledging that "with positivism came a model of explanation that subsumed particular phenomena to universal law, a formula – the scientific method – for obtaining valid knowledge, and a cult of quantification and measurement that changed current perceptions of what empirical reference may mean," Couclelis and Golledge (1983, p. 332) argued that there were important ideas in analytic behavioral geography that had enduring value. These ideas included reliance on logico mathematical structures and an "insistence upon open, public intersubjective tests of knowledge by continuous reference to experience" (Golledge, 1983, p. 334). What is left, they argued, is a mode of discourse, which, when pursued with honesty and humility, gives us fundamental principles from which to study human spatial behavior with the objective of finding valid general statements about human behavior in space. A summary of many of Golledge's ideas in this area came together in his book with Stimson, which takes the reader through the tenets of behavioral approaches to human geography, principles of spatial cognition, theories of learning, daily activity patterns, spatial choice models, migration in a behavioral context, and residential location decision making.

David Simonett, the new chairman of the department of geography at the University of California, Santa Barbara, decided that he wanted Reg Golledge to head up his quantitative human geography program. After an intense recruitment effort, Simonett succeeded in 1977. Golledge, in collaboration with Larry Hubert, a professor of statistics at University of California, Santa Barbara, produced a series of papers that searched for the principles of measuring order in spatial data. Their classic paper, published in *Geographical Analysis* in 1981, focused on spatial autocorrelation. They used randomization procedures for producing reference distributions describing the relationships that would be expected to be found in data if spatial order were not present. These ideas were seminal in moving the geographic literature from its reliance on distributional assumptions in measuring statistical significance. Their explanation and illustration of measuring spatial association and spatial autocorrelation coincided with the availability of increased computing power in universities, which made the implementation of reference distributions practical. The random relabeling ideas first described in the papers with Hubert have become foundations for spatial analysis methods for understanding many geographical distributions where events occur with defined likelihoods and outcomes vary stochastically but are consistent with the priors that are consistent with theoretical norms. For these contributions to geography in particular, Golledge received the honors award for scholarship from the Association of American Geographers in 1981.

After becoming legally blind in the mid 1980s, Golledge turned his interests and knowledge of spatial learning and cognition to problems faced by the visually handicapped. Soon he was writing on 'wayfinding' and route selection, and developing haptic soundscapes for blind users to access computer screen imagery. He spoke at international conferences on 'Navigation without sight', 'Navigation aids for the blind', 'Analysis of navigation without sight', 'Spatial cognition of the blind and visually impaired', and 'Auditory maps as alternatives to tactual maps'.

10 years later he was writing professionally about the problems academics confront when facing blindness. When nominating Golledge for a lifetime career honors award to the Honors Committee in 2006, Professor Grant Thrall noted that "Reg turned a misfortune of health into a laboratory to improve the quality of life for the visually handicapped."

Emeritus Professor Les King said of Golledge: "He not only learned to live with his blindness but made it a subject of first class research and fount of inspirational instruction and leadership."

Besides receiving many honors, he was elected fellow of American Association for the Advancement of Science in 1990; vice president and president of the Association of American Geographers from 1998 to 2000; and fellow of the American Academy of Arts and Sciences in 2005.

In his post presidential address to the Association of American Geographers in 2002, Golledge said, "Finding an answer to questions of how and why geographers think the way they do has stimulated much of my research over the last 35 years." Indeed, in his estimation the changes of the last century have moved geography from an inventory dominated and descriptive activity (finding 'what' and 'where') to emphasizing cognitive demands (finding 'why' and 'how').

He also said that perhaps the greatest changes in the nature of geographic knowledge over the last half century have been:

1. Recognition of the difference between the acts of accumulating geographic facts and representing the spatial form embedded in these facts, and under standing the processes involved in understanding and analyzing those facts to produce new information and knowledge that is not directly observed during data gathering.
2. The development of spatially relevant theories about the location, arrangement, and distribution of geo graphic phenomena and the spatial interactions among both physical and human components of those phenomena.

"I believe," he said, "that geographic knowledge is con cept rich, has a substantial theoretical base, is replete with distinct analytic forms (both qualitative and quan titative), lends itself to graphic, cartographic, and other forms of geo representation, and is imbued with know ledge derived from using place specific reasoning to in tegrate components of its physical and human domains."

Golledge has influenced geography through his many professional contacts, graduate students, international and national presentations, and the offices he has held at the University of California, Santa Barbara, where he chaired the department of geography for 4 years in the early 1980s during a period of consolidation and growth.

See also: Behavioral Geography; Mental Maps; Multidimensional Scaling; Positivism/Positivist Geography; Spatial Autocorrelation.

Further Reading

Couclelis, H. and Golledge, R. (1983). Analytic research, positivism, and behavioral geography. *Annals of the Association of American Geographers* 73, 331 339.

Cox, K. R. and Golledge, R. G. (eds.) (1981). *Behavioral Problems in Geography Revisited*. New York: Methuen.

Egenhofer, M. J. and Golledge, R. G. (eds.) (1998). *Spatial and Temporal Reasoning in Geographic Information Systems*. New York: Oxford University Press.

Golledge, R. G. (1979). The development of geographical analysis. *Annals of the Association of American Geographers* 69, 151 154.

Golledge, R. G. (1997). On reassembling one's life: Overcoming disability in the academic environment. *Environment and Planning D: Society and Space* 15, 391 409.

Golledge, R. G. (1999). *Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes*. Baltimore, MD: The Johns Hopkins University Press.

Golledge, R. G. and Stimson, R. J. (1987). *Analytical Behavioural Geography*. London: Croom Helm.

Golledge, R. G. and Stimson, R. J. (1997). *Spatial Behavior*. New York: The Guilford Press.

Hubert, L. J., Golledge, R. G. and Costanzo, C. M. (1981). Generalized procedures for evaluating spatial autocorrelation. *Geographical Analysis* 13, 224 233.

Rice, M., Jacobson, R. D., Golledge, R. G. and Jones, D. (2005). Design considerations for haptic and auditory map interfaces. *Cartography and Geographic Information Science* 32, 381 391.

Relevant Websites

http://www.geog.ucsb.edu
Golledge take, a lively chronicle 1963 2000, Department of Geography, University of California, Santa Barbara.

Gottmann, J.

L. Muscarà, Università del Molise, Isernia, Campobasso, Italy

Glossary

Circulation (Movement) The displacement of people, goods, armies, money, information, ideas, etc.

Crossroads A multiscalar concept defining a position in space where flows of circulation meet, mix, and irradiate their orbit to the surrounding space.

Geographical Space Space accessible to people and their activities, defined by Gottmann as continuous, limited, expanding, diversified, organized, and partitioned.

Hinge The interface function of a city or region between two or more geographical scales.

Iconography Set of beliefs and abstract symbols relating a community to its territory at any given scale. Its contents are less important than its action in affecting movement in space.

Megalopolis A polynuclear urban region, with a nebula structure, inhabited by at least 25 million people, which functions as a 'hinge' between the local, regional, national, and global scales. Originally developed to describe the Northeastern urbanized seaboard of the US, it was then applied to other parts of the world by other scholars (Great Lakes megalopolis, Tokaido megalopolis, Northwestern Europe megalopolis, etc.). It differs from Mumford's megalopolis.

Political Partitioning of Space The process of dividing geographical space into political units.

Regional Monograph Classic, analytical tool developed by the French School of Geography.

Jean Gottmann's (1915–94) contribution to human geography has long been associated with urban geography and, in particular, *Megalopolis*, his study of the post World War II urbanized Northeastern seaboard of the US. Sometimes controversially, this became a paradigm subsequently applied to different regions of the world by other scholars. The opening of the *Fonds Gottmann* at the National Library in Paris, coupled with an international and intergenerational effort, has allowed for reassessment of his contribution to urban, regional, economic, historical, and political geography, reconciling the profound fragmentation of awareness and recognition of his work, which brought some to observe that: "each geographer had its own Gottmann."

Twice a political refugee, an academic and linguistic nomad, Gottmann was active for six decades, with numerous appointments and a record of approximately 400 publications, mostly in English and French (though not respectively translated). A French citizen from 1939 onwards, he is considered the first cosmopolitan geographer, who merged the French and Anglo American methodologies.

Born in Kharkov to a Jewish family, active in railroad development in the Ukraine, then part of the multinational Czar imperium, Jean Gottmann lost both parents during the Russian revolution and fled, with three surviving relatives, to Constantinople in 1919. Once in Paris (1921–40), he studied under the French system: *Lauréat du Concours Général des Lycées et Collèges de France* in geography (1931), he enrolled at the Sorbonne in the faculties of law, history, and geography (1932). *Docteur ès Lettres* in 1934, in his diploma thesis on irrigated agriculture in Palestine, he merged physical, economic, and cultural geography, with a touch of Zionist irredentism injected in him by his stepfather Michael Berchin.

As research assistant to Demangeon (1936–40), through the Rockefeller Foundation's Social Science Program at the University of Paris, he worked on the structure of agriculture in France and its Mediterranean colonies, and on France raw materials supply. His knowledge of the Russian language granted him the opportunity to review, for the *Annales de Géographie*, different aspects of Soviet economic geography. On theoretical grounds, he criticized Huntington's environmental determinism in the debate then opposing the French and German schools, and was attracted by Isaiah Bowman's 'pioneer fringe'. However, his French career, interrupted by the Nazi invasion, was completed years later.

A refugee scholar, he fled to Montpellier and reached New York in December 1941, where he joined De Gaulle's 'France Libre'. His Mediterranean knowledge and French regional approach attracted strong interest in the US war effort. Member of the Institute for Advanced Study in Princeton, NJ (1942–44), he consulted in Washington, DC (1942–43), and taught at Bowman's Johns Hopkins University, Baltimore, MD (1943–48). In 1945, he served the French ministry of economy in Paris and Washington. Director of studies and research at the UN Secretariat (1946–47), in a seminal methodological essay on human geography (1947), he proposed an analysis of network flows ('circulation') through multiscalar nodes ('crossroads chains'), and the use of psychology to study the cultural geography of consumption patterns. And yet, despite his close ties with both Oppenheimer and Princeton physicists, he only foresaw the quantitative developments of Anglo American geography. In his post war unsettlement, he set out on an academic transhumance across the Atlantic and on a cross cultural effort to explain post war Europe to America and vice versa (1949, 1950).

While resuming his French academic career at *Centre National de la Recherché Scientifique* (CNRS) (1948–51) to study the Boston–Washington urban corridor – where he had spent most of his time during the war – Gottmann began teaching at the *Institute d'Etudes Politiques* (1949–55), where he developed a general theory of political geog raphy in several articles, published in America and France, and in a book (1952). He insisted on the 'hinge' between the regional and the global, and explained the political partitioning of geographical space as the result of the interaction of network flows ('circulation') and territorial identities ('iconographies') of communities, through the key role played by nodes (crossroads chains), and their orbits in shaping human regions, beyond the framework of the nation state. Political territoriality was seen as historically contingent and – as he elaborated later (1973) – dependent on the community's search for opportunities or security. Without a permanent ap pointment, he found support from private foundations to direct a study of Virginia (1952–55) – the first attempt at a regional monograph in the French geographic tradition of a US region – and the research on *Megalopolis*, re spectively thanks to Flexner and Oppenheimer.

Mistakenly understood by many as a regional study, *Megalopolis* was viewed as the incubating laboratory of major trends in post war urbanization: the polynuclear structure, the then unmatched density of cities, the 'hinge' of American economy, serving 'both as an oceanic façade for relations abroad and as a springboard for the settlement and development of the continent inland', 'the symbiosis of rural and urban' in land use, and the coming dominance of a 'quaternary' sector of labor force, due to the advent of an information economy. Recent studies explained how the genesis of *Megalopolis* stems from his earlier political geography.

Between 1961 and 1983, he became *Directeur d'Etudes* at the *Ecole des Hautes Etudes en Sciences Sociales* in Paris; in 1968 he was offered the chair of the school of geography at Oxford University (Hertford College), which he viewed as a synthesis of France and America, the close of his yearly Atlantic transhumance, and a springboard to Japan (**Figure 1**). Urban geography remained pre dominant in his theoretical and applied work from 1956 to 1994. In his last two decades, he elaborated further on the concepts of territory, centre periphery, the political aspects of urbanization, capital cities, telecommunica tions and settlements, and diasporas. Free from any de terminism, Gottmann's general contribution to human geography focused on the organization of space, in a historical perspective, and with a particular interest for how psychology shapes the material forms and structures

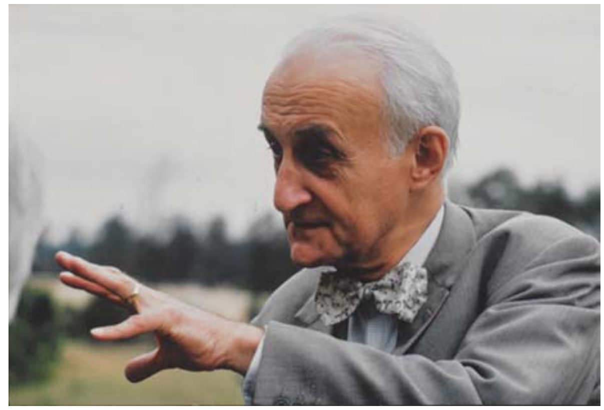

Figure 1 Jean Gottmann in the 1970's.

of cohabitation. He retired from Oxford in 1983 as emeritus fellow, and in 1993 was nominated honorary fellow. He died in Oxford in 1994.

In his long career, Gottmann chaired the International Geographical Union (IGU) Commission on the *Amén agement Regional* (1949–52), *Centre d'Etudes Prospective* Committee on French Urbanization in Paris (1962–66), *Fondation Européenne de la Culture* Committee on Urban ization (Amsterdam, 1969–72), World Society for Ekistics (Athens, 1971–73), Permanent Research Committee on Political Geography of the International Political Science Association (1976–94). Honorary fellow of the American Academy of Arts and Sciences (1972); fellow of the British Academy (1977); honorary member of the American, Dutch, Italian, and Lièges geographical soci eties; vice president of the Royal Geographical Society; *Chevalier de la Légion d'Honneur et des Palmes Académiques*; doctor *Honoris Causa* at the University of Wisconsin, Southern Illinois, and Liverpool; honorary citizen of Guadalajara (Mexico) and Yokohama (Japan).

See also: City-Region; Core-Periphery Models; Cosmopolitanism; Growth Poles, Growth Centers; Historical Geography; Mega-Cities; Network Regions; Polycentricity; Regionalism; Sovereignty; Territory and Territoriality; World/Global Cities.

Further Reading

Agnew, J. (2007). Territoire, souveraineté et le corps politique. *La Géographie* 1523bis, 91 103.

Baigent, E. (2004). Patrick Geddes, Lewis Mumford and Jean Gottmann: Divisions over 'megalopolis'. *Progress in Human Geography* 28, 687 700.

Beaujeu Gamier, J. (1986). A la remise de la grande medaille d'argent au Professor Jean Gottmann, Oxford, 9 Octobre 1986. *Acta Geographica* 67, 6.

Clout, H. (2006). Review: La strada di Gottmann. *Progress in Human Geography* 30, 696 698.

Corey, K. E. (1995). In memoriam: Jean Gottmann, 1915 1994. *Annals of the Association of American Geographers* 85, 356 365.

Gottmann, J. (1947). De la méthode d'analyse en géographie humaine. *Annales de Géographie* 56, 1 12.
Gottmann, J. (1949). *L'Amérique*. Paris: Hachette.
Gottmann, J. (1950). *A Geography of Europe*. New York: Henry Holt.
Gottmann, J. (1952). *La politique des Etats et leur géographie*. Paris: A. Colin. [Reprinted 2007, Paris: CTHS.]
Gottmann, J. (1955). *Virginia at Mid Century*. New York: Henry Holt.
Gottmann, J. (1957). *Les Marchés des Matières Premières*. Paris: A. Colin.
Gottmann, J. (1959). *Etudes sur l'Etat d'Israel et le Moyen Orient*. Paris: A. Colin.
Gottmann, J. (1961). *Megalopolis, the Urbanized Northeastern Seaboard of the United States*. New York: The Twentieth Century Fund.
Gottmann, J. (1966). *Essais sur l'aménagement de l'espace habité*. Paris: Laye, Mouton.
Gottmann, J. (1973). *The Significance of Territory*. Charlottesville, VA: The University Press of Virginia.
Gottmann, J. (ed.) (1980). *Centre and Periphery, Spatial Variation in Politics*. Beverly Hills, CA: Sage.
Gottmann, J. (1983). *La Città Invincibile, una Confutazione dell'Urbanistica Negativa*. Milan: Angeli.
Gottmann, J. (1994). *Beyond Megalopolis, towards a World Community?* Tokyo: The Community Study Foundation.
Gottmann, J. and Harper, R. (eds.) (1990). *Since Megalopolis, the Urban Writings of Jean Gottmann*. Baltimore, MD: Johns Hopkins University Press.
Gottmann, J. Muscarà, C. (eds.) (1991). *La città prossima ventura*, Roma Bari: Laterza.
Hooson, D. (1983). National cultures and academic geography in an urbanizing age. In Patten, J. (ed.) *The Expanding City*, pp 157 178. London: Academic Press.
Hubert, J. P. (2007). L'art de réinventer la région. *La Géographie* 1523bis, 104 115.
Johnston, R. (1996). Jean Gottmann: French regional and political geographer extraordinaire. *Progress in Human Geography* 20, 183 193.
Muscarà, C. (2003). In the steps of Jean Gottmann. *Ekistics* 70, 418 423.
Muscarà, L. (2005). *La Strada di Gottmann*. Rome: Nexta.
Muscarà, L. (2005). Territory as a psychosomatic device: Gottmann's kynetic political geography. *Geopolitics* 10, 26 49.
Muscarà, L. (ed.) (2007). *Une Géographie dans L'histoire, Actes du Colloque Internationale: L'orbite de la Géographie de Jean Gottmann*. Paris: Société de Géographie.
Sanguin, A. L. (1996). Jean Gottmann et la géographie politique. In Claval, P. & Sanguin, A. L. (eds.) *La Géographie Française à L'epoque Classique (1918 1968)*, pp 229 240. Paris: L'Harmattan.
Sanguin, A. L. and Prévélakis, G. (1996). Jean Gottmann, un pionnier de la géographie politique. *Annales de Géographie* 587, 73 78.

Relevant Websites

http://catalogue.bnf.fr/
Fonds Gottmann at the Bibliothèque Nationale de France.

Governance

M. Goodwin, University of Exeter, Exeter, UK

Glossary

Keynesian Welfare State A political and economic system, which balanced mass production with mass consumption, underpinned by state welfare provision.

Metagovernance The coordination and organization of governance structures – sometimes referred to as the governance of governance.

Multilevel Governance A political system in which different territorial levels of governance share decision-making competencies across a variety of scales, linking local governance to regional, national, and international political networks.

Partnership An arrangement which deliberately draws together the resources of specified partners in order to achieve a defined objective.

Introduction

The 1990s saw the widespread adoption of the term governance in academic, policy, and practitioner circles alike, and by the end of that decade governance had gained a widespread currency across many of the social sciences. To those interested, for instance, in the geographies of public policy, political institutions, organizational change, and social networks, the term governance opened up new avenues of enquiry into a wide range of social and political systems. Where we once spoke almost automatically of the government of such systems, we now routinely talk of their governance. This article traces the reasons behind this shift, and examines the usage of the term governance within human geography.

The Turn to Governance

It was no accident that by the 1990s the term governance had begun to appear regularly in the geographical literature. Indeed, like many other terms used within the social sciences it is date and place specific, and its ascendancy can be linked to a series of social and political changes which swept across the Anglophone world in the 1980s and 1990s and together resulted in a changed meaning to the notion of the government. This changed meaning emerged through three interconnected processes, which individually and collectively served to question the established ways of governing society. First, there had been the fiscal crisis of the post war Keynesian welfare state, and its associated modes of economic and political coordination. Although the precise nature of the welfare state varied from country to country, by the 1980s governments across Western Europe, North America, and Australasia were reducing their spending and cutting their welfare budgets. Second, in many instances – especially the UK, USA, Australia, and New Zealand – neoliberal governments were swept to power on a wave of new right ideology which provided a politically effective critique of state activity, government bureaucracy, and 'red tape'. Once in power, however, the ideology turned from a 'rolling back' of the state to a search for new forms of intervention and control. These new state forms were constituted by a raft of new agencies and institutions, some based in the civil service, some operating at 'arms length' as quangos (quasi autonomous nongovernmental organizations), and others occupying a somewhat indeterminate sphere as nondepartmental public bodies. This was reinforced by the third process, whereby political actors at the local level have themselves formed a range of coalitions and partnerships which have also transformed the institutional structure of government. Some of these have been funded and promoted by higher tiers of government, but others are the result of local decisions and arrangements to involve nonstate actors from both the private and voluntary sectors.

The end result was a transformation in the institutional map of government, and a change in the way society was governed. Where public services had once been the sole province of central and local governments, these changes to the role of the state transformed this system into one which instead involved a wide range of agencies and institutions drawn from the public, private, and voluntary sectors. These still include the institutions of elected government, but also involve a range of nonelected organizations of the state, as well as institutional and individual actors from outside the formal political arena, such as voluntary organizations, private businesses, and corporations and supranational institutions such as the European Union. Reflecting these changes, the term governance is thus used to refer "to the development of governing styles in which boundaries between and within public and private sectors have become blurred" (Stoker, 1998: 17). Hence the concept is not simply an academic synonym for government. Its increasing use signifies a concern with a change in both the meaning and content of government. As Rhodes puts it, "governance signifies a change in the meaning of government, referring to a *new* process of governing; or a *changed* condition of ordered rule; or the *new* method by which society is governed" (Rhodes, 1996: 652–653). Where government signals a concern with the formal institutions and structures of the state, the concept of governance is broader and draws attention to the ways in which governmental and nongovernmental organizations work together, and to the ways in which political power is distributed, both internal and external to the state. In the words of Jessop, the term governance signals a shift to "a broad concern with a wide range of governance mechanisms with no presumption that these are anchored primarily in the sovereign state" (Jessop, 1995: 310–311).

This 'broad concern' has been picked up across the social sciences, and the parallel emergence of the term 'governance' across economics, international relations, organizational studies, sociology, public administration, and political science owes something to the fact that each was struggling to understand these broad sets of changes in the relationship between elements of the state, market, and civil. In institutional economics the notion of governance led to a growing interest in forms of economic and corporate coordination which bridge the divide between state intervention, pure markets, and corporate activity. In international relations, the focus is on the coordination of nation states through international agencies such as the United Nations (UN) and through international agreements such as the Kyoto Treaty on Climate Change. In organizational studies, the focus brought by a concern with governance has again been coordination, this time of interorganizational relationships, but there has also been a concern with the self organization of institutions, especially through the study of organizational intelligence and learning. In sociology, theorists have drawn on Foucaultian inspired notions of governmentality and discipline to examine the intersection between the technologies of government and the technologies of the self, introducing an individual and subjective stance into the debates on governance. In all these disciplines, one will see the term governance used slightly differently – although sharing a common root.

In human geography, it is perhaps the explicit policy focus of the political science literature that has had most influence, as political geographers have sought to understand the changing ways in which different places and different territories are governed. The concept of governance explicitly focuses attention on the relations between different actors, and on their coordination – and crucially from our perspective draws attention to the complex geographies now involved in the act of governing any particular area. The governing of any particular space – whether that is a city, a region, or a rural area – is now a complex, differentiated and multiscaled process, involving a range of different actors and agencies. Indeed, this complexity has been recognized by the development and use of the term 'multilevel governance' within both political science and geography. The term signals a concern with the way in which different territorial levels are linked within a dispersed decision making system, and it has been especially used as the focus of empirical work on governance structures within the European Union. This has meant that governance studies have extended well beyond the Anglophonic world and empirical studies of governance cover a wide range of different countries. We now turn to the ways in which the concept of governance has opened up a set of interlinked research questions.

Governance as a Research Perspective

Importantly, the term governance does not just refer to a new set of institutions and actors involved in the process of governing. It also signals a change in the nature of the relationships between governing organizations. Taken together, these uses of the term provide us with a new framework for interrogating and understanding the changing ways in which the society is governed. We can analyze this framework by using five propositions set out by Stoker, and interested readers are referred to his article for further details. The five propositions are given in the following:

1. Governance refers to a set of institutions and actors that are not only drawn from but also beyond government.
2. Governance identifies the blurring of boundaries and responsibilities for tackling social and economic issues.
3. Governance identifies the power dependence involved in the relationships between institutions involved in collective action.
4. Governance is about autonomous self governing networks of actors.
5. Governance recognizes the capacity to get things done, which does not rest on the power of government to command or use its authority. It sees government as able to use new tools and techniques to steer and guide.

These should not be seen as statements that can be proved or disproved, but rather as propositions which open up a number of research issues which would not have been investigated under the old systems of government. As such, each unfolds new avenues of enquiry for human geographers.

Institutions Beyond Government

The first proposition challenges understandings of systems of government that focus heavily on the formal structures of elected government, at both local and central levels. But by drawing our attention to the complexities of modern governance, the first proposition also points to the highly differentiated polity within most contemporary nations. It also leads us to ask questions about how this polity is differentiated spatially and functionally. Governing organizations are now linked across local, regional, national, and supranational scales, and they are drawn from the public, private, and volun tary sectors. **Figure 1** shows something of this complexity.

It is drawn from the work of Winter and shows the governing networks operating around one small re generation project in the County of Devon, in the UK. The connections in the network reach from Hatherleigh, a small market town, to regional and national tiers of government, and they involve a host of partnership or ganizations, which themselves have member organiza tions drawn from different sectors. Although this figure simply refers to one service area in one small town, it shows the nature and range of organizations involved in the new rural governance. Under the old structures of the welfare state, which predated the turn to governance, this diagram would have had only four organizations on it – the Hatherleigh Area Project (HAP) itself, the Borough Council, the County Council and central government. Now, after the shift to governance it has over 25, all with differing relations to each other. The diagram de ploys different arrow thicknesses in order to capture these different types of organizational relationship. Thin lines represent relationships of praxis where HAP works with another local institution because of mutual interest or complementarity. Dotted lines mark strategic rela tionships, where HAP is guided in some way by the policy or strategy of a higher level organization. Financial relationships are represented by thick lines, and rela tionships which are both strategic and financial are marked by double lines.

The messiness and scope of this complexity raises fur ther questions concerning legitimacy and power. Research has uncovered widespread concerns about the operation of this new structure of governance – many components of which are unelected and unaccountable. This is more than a normative concern about where power should and should not be held. It also raises the very practical problem that to be effective in the long run, holders of power should have legitimacy. Yet it is very difficult to trace lines of ac countability within such a system.

Partnerships and the Blurring of Boundaries

The second proposition points not just to the complexities of modern governance, but to the ways in which these shift responsibilities away from the institutions of elected structures of government. This shift can be summed up by ongoing concerns about the active citizen and the

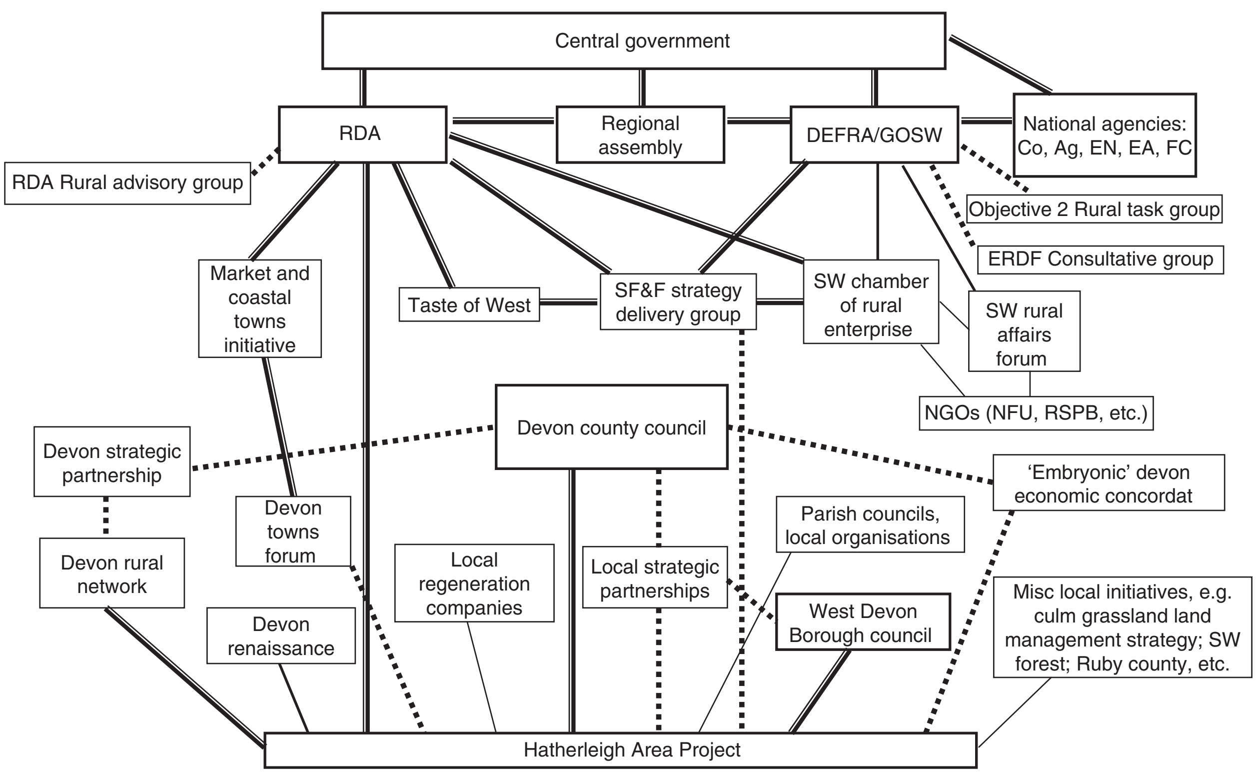

Figure 1 Rural regeneration: the governance network for Hatherleigh in West Devon. Reproduced from Winter, M. (2006). Rescaling rurality: Multi-level governance of the agro-food sector. *Political Geography* 27, 748, with permission from Elsevier.

responsible corporation – now both engaged in the partnerships which deliver public services. **Table 1** is taken from the work of Edwards *et al.* and shows the extent of the agencies involved in the new structures of governance.

A number of issues arise from this table. The first is the sheer number of partnerships involved in the new structures of governance. In many ways partnerships are the emblematic form of governance, as they seek to draw together the resources of a number of different partners, and replace the role and functions of single, elected government authorities. These 'resources' may be financial, practical, material, or symbolic, while the partners may be drawn from different sectors or from the same sector. The rationale for partnership creation is that the blending of resources is necessary in order for a desired outcome to be achieved. These partnerships have grown almost exponentially in recent years, and communities of all sizes now usually boast a plethora of groups focused on community development, civic refurbishment, business development, marketing, sustainability, and service delivery in policy areas such as health, social services, childcare, education, policing, transport, and tourism – all constituted as some form of partnership, bringing together a range of organizations, often from across the public, private, and voluntary sectors.

Edwards *et al.* looked purely at regeneration partnerships, and only examined three rural counties in the UK, but they found over 150 partnerships in operation. Had they examined other service areas, such as health, education, and housing, and other places, then the numbers would have been multiplied accordingly. Aside from the sheer number, the complexity of their structure stands out. Some partnerships had over 50 members, and many had 10–15. The involvement of so many partners leaves considerable scope for blame avoidance and scapegoating. There is within the new governance what at times seems like an endless circle of responsibility which can never quite be closed. Each agency or actor blames another, and the public can never quite manage to pinpoint which of the partners is responsible if things go wrong. Although one of the advantages of partnership working emphasized within policy discourse is the capacity of partnerships to integrate different sectors, in practice few partnerships meet this tripartite ideal. Research has consistently confirmed that a minority of partnerships do actually have representation from all three sectors, and for a significant proportion the partnerships exist solely between agencies in the public sector. Research has also found that both the voluntary and private sectors are more strongly represented in partnerships scaled at a local level. This raises questions of access to power, as community and private sector actors are more commonly enrolled into partnerships focused on delivery, than into partnerships focused on strategic discussion and decision making.

Power Dependence and Collective Action

The third proposition concerns the power dependence involved when a range of agencies come together to act collectively. Put simply there is a divergence between the wish for effective and swift action and the lack of co-ordination which inevitably follows from dependence on others. Governance as an interactive process between a range of organizations draws attention to the difficulties of negotiating shared goals and agendas. Moreover, the contribution of partners drawn from different sectors to partnerships is not equal. Research has shown that public agencies and local authorities are frequently responsible for financing and initiating partnerships, and play the major role in resourcing them on a day to day basis, especially in the provision of office space and personnel. Indeed, the concept of governance as an interactive process between a range of organizations draws attention to the difficulties of negotiating shared goals and agendas, and to the acceptance that many intended actions will not match eventual outcomes. Most studies have found that despite a rhetoric of inclusion and partnership, there is a considerable imbalance of power between those involved in the new rural governance, and many point to the difficulties of involving both private and community sectors, and have also identified considerable gender imbalances in the new governance structures which tend to be male dominated.

Self-Governing Networks

The fourth proposition focuses our attention on self-governing networks, and much research has been carried out on urban regimes. A regime is formed when a variety of local interests mesh together to form a relatively stable coalition, and involves a number of groups cooperating behind a certain set of policies to achieve their own ends – which may well vary from group to group. Research pays attention to the more informal arrangements that surround and support the official workings of government, and indeed such informal coalitions often outlast formally elected administrations. Often regimes are directed at securing competitive advantages for particular urban localities, as local actors seek collective arrangements that will promote their locality in an increasingly competitive global economy. Such regimes have been given the label of urban growth coalition. The idea of growth coalitions was first developed in the late 1970s and early 1980s to account for the pro growth nature of much urban politics in the USA. Research argued that a key feature of US urban government was the formation of coalitions of local business leaders seeking to promote growth. The key players in such coalitions (also known as growth machines) were local property owners and property developers who stood

Table 1 The structure of regenerat on partnersh ps n m d Wa es and Shropsh re

Scale of territory	Number of initiatives	Focus of partnership		Funding of partnership					Initiation of partnership			Duration of partnership		Sector mix of partners				Aver. nos. of partner
		Facilitative	*Delivery*	*EU*	*Q*	*LA*	*P*	*TEC*	*Prior to 1993*	*1994–96*	*1997–2000*	*Ongoing*	*Limited life*	*Q*	*LA*	*V*	*P*	
Reg ona and above	73	47	26	28	43	27	14	10	9	35	29	38	35	64	51	30	36	10.38
County	50	13	37	20	14	26	5	26	3	25	20	32	18	43	40	19	32	7.48
Subcounty *Area*	14	1	13	7	10	9	0	0	0	8	6	2	12					7.29
															25	30	22	19
Subcounty *Place*	17	3	14	4	10	8	1	1	0	10	8	7	10					4.35
Tota *N* (%)	154	64	90	59	77	70	20	37	12	78	63	79	75	132	121	71	87	7.39
	100	42	58	38	50	45	13	24	8	51	41	51	49	86	76	46	56	

Key EU = European Union Q = Quango LA = County or District Council P = Private Sector TEC = Training and Enterprise Council V = Voluntary/Community
Source Edwards *et al.* (2000) *Partnership Working in Rural Regeneration* Bristol Policy Press 8

to gain the most from a rise in property values, due to economic growth which would fuel intensification of land use and demand. They were joined by other pro growth interests – financiers, regional banks, construction companies, estate agents, retailers, etc. – who wished to mobilize the powers of local government in order to create a business climate conducive to economic growth. They were acting out of self interest, but the key to success was to present this to local politicians and the local electorate as value free development which was self evidently good for the locality. This model of local politics seeks to explain how economic agents in the city realize their vested interests through influencing local politics – the key point here is that the new mechanisms of governance are giving much more scope for this influence to be exercised.

Of course, not all regimes are concerned with the pursuit of growth, and much of the work in this area has focused on the ways in which different types of regimes are established in different urban contexts. Many simply wish to protect the status quo, while others are concerned with long term transformative change. It is interesting to note that one of the very first (and still most influential) geographical papers to explicitly use the term governance – that by Harvey – focuses on the different strategies employed by different types of entrepreneurial urban governance.

New Tools of Government

The fifth area of research prompted by a governance perspective actually focuses on the institutions of government, in recognizing that the capacity to get things done does not rest with government alone. In short, it recognizes that government itself is now operating in a changed environment, and consequently is concerned with analyzing different modes of intervention. Indeed, there is a strong sense in the governance literature that government itself has acknowledged this, and is busy reinventing itself in a variety of guises – as an enabler, commissioner, manager, and shaper. The role for government shifts from direct intervention to one of coordinating and integrating others, hence the famous remark about governments now being more involved in steering rather than rowing. It is seen more as one of identifying appropriate stakeholders and then developing the relevant opportunities and linkages for them to be brought together. However, even where government does succeed in carving out a new role for itself, there is no guarantee of success. Indeed, all kinds of tensions and difficulties among partners and between different institutions may well cause governance failure. Ironically, while many theorists see the emergence of governance as a response to market failure and state failure, we still need to allow for the possibility of governance failure. Governance is not necessarily a more efficient solution than the state in providing public goods and services.

Beyond the Dichotomy of Government and Governance

This final proposition also reminds us that government and governance are not two separate processes, and that there is not simply a linear path from one to the other. Indeed a review of research in this area raises a number of thoughts over how we might conceptualize the rise of the new governance. Initially, when researchers found evidence of the rise of partnerships, quangos and other new governing structures, these were conceptualized in terms of a shift from (local) government to (local) governance. However, recent research has uncovered attempts to coordinate the profusion of governance structures at local and regional levels. In the UK, for instance, Local Strategic Partnerships and Local Area Agreements have been introduced in an explicit effort to rationalize the ever increasing complexity of governance. This might be better conceptualized not as a continuation of governance but as a response to the problematics of governance. In succinct terms, what we are seeing is an attempt to govern the structures of governance. Thus governance still operates in the shadow of government, and a critical conceptual task is to understand how the two systems are linked and interrelated. One idea which can be drawn on to investigate the relationship between governance and government is that of metagovernance, a concept which focuses explicitly on the practices and procedures which secure governmental influence, command, and control within governance regimes.

The picture which emerges from recent empirical work is indeed one where the structures of government, including local government, are attempting to secure 'influence, command, and control'. For instance, formal institutions of elected government provide the ground rules for the operation and funding of partnerships, they provide much of the resources for their day to day operation, they deploy information and intelligence on partnership activities and they resolve disputes and shape partnership behavior through an increasingly complex system of guidance, monitoring, assessment, and discipline. When the partnership structures grow too unwieldy, the state steps in to replace them with a more rationalized and refined system. What we are witnessing is the development of new constellations of government and governance.

This does not mean that we can replace concepts of governance with those of metagovernance. Indeed, metagovernance is premised on the existence of governance regimes. Rather we should be looking to extend work on governance by connecting it to an understanding of changes in government control and authority. In this way we can look to move beyond the increasingly sterile 'government versus governance' debates which can des cend into either proving or disproving a shift from one to the other. Metagovernance offers us a way out of this impasse by offering us the opportunity to connect changing governance structures on the ground to chan ging political strategies and changing articulations of state power, in examining new constellations of both government and governance.

See also: Entrepreneurship; Governance, Urban; New Urbanism; State; Urban Growth Machine; Urban Policy; Urban Regimes.

Further Reading

Edwards, B., Goodwin, M., Pemberton, S. and Woods, M. (2000). *Partnership Working in Rural Regeneration*. Bristol: Policy Press.

Goodwin, M. and Painter, J. (1996). Local governance, the crisis of Fordism and the changing geographies of regulation. *Transaction of the Institute of British Geographers* 21, 635 648.

Gualini, E. (ed.) (2006). *Governance rescaling in Europe: Analytical and empirical explorations*. *European Planning Studies* 14(7 Special Issue), 881 995.

Harvey, D. (1989). From managerialism to entrepreneurialism: The transformation in urban governance in late capitalism. *Geografiska Annaler* 71B, 3 17.

Imrie, R. and Raco, M. (1999). How new is the new local government; lessons from the UK. *Transactions of the Institute of British Geographers* 24, 45 63.

Jessop, B. (1995). The regulation approach, governance and post Fordism: Alternative perspectives on economic and political change? *Economy and Society* 24(3), 307 333.

Pierre, J. (ed.) (1997). *Public Private Partnerships in Europe and the United States*. London: Macmillan.

Rhodes, R. (1996). The new governance: Governing without government. *Political Studies* XLIV, 652 667.

Stoker, G. (1998). Governance as theory: Five propositions. *International Social Science Journal* 155, 17 28.

Ward, K. (2000). A critique in search of a corpus; re visiting governance and re interpreting urban politics. *Transactions of the Institute of British Geographers* 25, 169 185.

Winter, M. (2006). Rescaling rurality: Multi level governance of the agro food sector. *Political Geography* 27, 735 751.

Governance, Corporate

D. Wójcik, Oxford University Centre for the Environment, Oxford, UK

Glossary

Board of Directors A group of people who oversee the affairs of a corporation, usually headed by a chairman.

Co-determination Practice whereby employees have a role in management of a company.

Corporate Stakeholders People and organizations affected by any action taken by a corporation.

Equity In finance referred to as the residual value of a company's assets after all liabilities have been allowed for.

External Capital Capital obtained from sources other than a company's retained earnings, mainly from equity investors or lenders.

Hostile Takeover Occurs when one company (acquirer) attempts to buy out another (target) company whether the management of the target likes it or not.

Shares One of a number of equal portions in the capital of a company, which entitles the owner to a proportion of distributed profits (dividends) and a part of residual value if the company goes into liquidation.

Definitions and Origins

The distinction between narrow and broad definition of corporate governance is illustrated in **Figure 1**. In a narrow sense, corporate governance concerns the relationships between providers of finance to companies and corporate managers, and focuses on legal prescriptions, business practices, social norms, and other ways in which providers of finance can obtain returns on their investment. In principle, investors in companies should be able to count on the willingness of their management to maintain a good reputation and obtain further finance in the future. Such mechanisms alone, however, would be insufficient to maintain investors' confidence and assure a steady flow of finance to companies. In terms of the narrow definition, the objective of corporate governance is therefore to enhance firms' access to external capital and facilitate investors' access to investment opportunities.

Among providers of external capital to companies, we can distinguish between owners as providers of equity finance, and lenders as providers of debt finance, especially banks. In contrast to lenders, which through debt contracts are entitled to a predetermined flow of repayments, owners' returns on investment consist of dividends, that is, the part of net corporate profits divided among owners, and capital gains, that is, the difference between the price at which owners can sell their ownership titles and the price at which they acquired them. Lenders are interested primarily in the company's ability to meet the promised repayments, while owners are concerned with its overall profitability and corporate value assessed by potential buyers. Lenders also typically hold collateral in the form of part of a borrower's assets, which they can sell if the company defaults on its repayments. If a company goes bankrupt and is liquidated, lenders also have priority over owners in sharing the proceeds from liquidation. Because of the character of owners as ultimate or residual risk takers in companies, corporate governance research focuses on owners, and is concerned with the mechanisms with which equity holders can obtain returns on their investment in companies. Corporate governance may refer to a variety of corporate forms. At one extreme, there are enterprises in which all managers are owners, and all owners are managers. At the other, there are corporations in which thousands of investors, whether individuals or institutions, hold shares, which are traded on a daily basis on stock markets. The complexity of corporate governance is obviously higher in the latter cases, and consequently the primary focus of corporate governance research is further narrowed down to relationships between corporate managers and shareholders.

Broad definitions of corporate governance extend the subject to other stakeholders of the firm: employees, suppliers, customers, local communities, and the government. As such, corporate governance concerns relationships of each of these groups with corporate managers, and also their relationships with each other, to the extent to which they influence how companies allocate resources and distribute returns. In terms of broad definition the objective of corporate governance is to contribute to economic growth and its fair distribution. The distinction between the narrow and the broad definition has a positive and normative aspect. On a positive level, narrowly defined corporate governance can be viewed as a subset of wider corporate governance issues. The subject of corporate governance research can thus be seen as a spectrum, from the relationships between shareholders and managers, to those among all possible corporate stakeholders, with a wide range of intermediate forms. These include the frequently studied mutual relationships between owners, managers, and

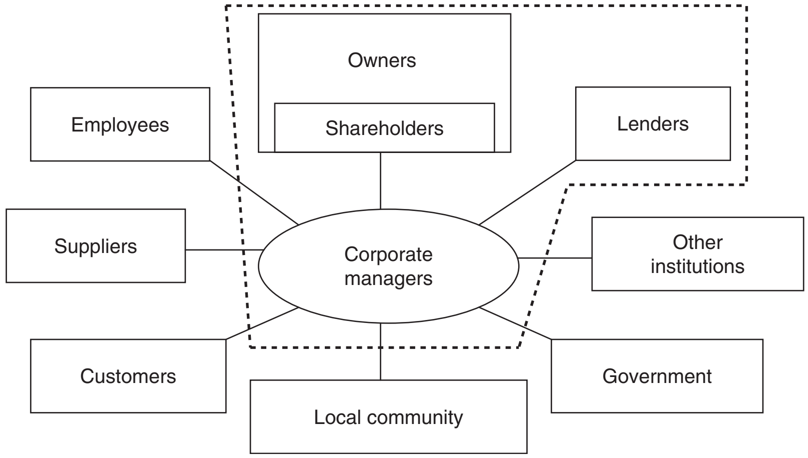

Figure 1 Broad and narrow definition of corporate governance.

employees. On a normative level, however, many proponents of the broad definition claim that narrow definitions promote the attitude that corporations should be run solely in the interest of owners, thus unfairly justifying the disempowerment of other stakeholders, in particular employees. This tension is an important source of dynamics in corporate governance research, highlighting its relevance for policy making, as well as for understanding the contemporary world.

With the very subject area of corporate governance being contested, corporate governance research is best seen as an interdisciplinary debate, rather than a focused academic discipline. The disciplines with the most direct stakes in corporate governance research are finance, economics, law, accounting, business, and management. With time, however, other social sciences including sociology and human geography should develop a growing stake in the debate.

Historically, the roots of corporate governance as a socioeconomic phenomenon can be traced back to the emergence of joint stock companies in the seventeenth century, when numerous cases of managers abusing shareholders on a mass scale highlighted governance problems. The implosion of the Mississippi Company in France and the South Sea Company in Great Britain are two infamous examples. Classical economists, including Adam Smith, highlighted potential problems arising from the separation of ownership from company management. Throughout the nineteenth and the twentieth centuries interest in corporate governance issues grew with the rise of joint stock companies, and theorizing on the nature of the firm. Research addressing corporate governance explicitly, however, started in earnest in the US only in late 1970s. By the late 1980s it had spread to other developed economies including Germany, Japan, and the UK. Only in late 1990s did it become worldwide, including applications in emerging and developing economies.

The Context of Contemporary Corporate Governance Debate

A number of developments contributed to the rise of corporate governance policy, research, and media debate in 1990s and early 2000s.

The privatizations of public assets, starting in the UK in 1980s and later unfolding in Western Europe, Asia, Latin America, and the former Soviet block, could be seen as an experiment in corporate governance on a mass scale, opening up questions of who the owners of companies should be and how they should relate to other stakeholders. In some cases it was seen as desirable to find large strategic investors for the newly privatized companies. In other cases, the preferred option was to sell shares to thousands of small investors on public stock markets. Both routes encountered problems and failures. Some companies were sold at low prices to strategic investors that turned out to be controlled by former managements, leaving nobody able to undertake the painful process of restructuring the company. In other companies managements were able to exploit the inability of small investors to act collectively. In addition to highlighting difficulties in establishing appropriate ownership and control structures in companies, privatizations also contributed to the creation of a large class of small shareholders – millions of individuals relying on the corporate sector as a source of investment income.

Another event that fuelled the debate was the economic turmoil in East Asia, led by speculation in 1997 against the Thai and Malaysian currencies. For some, this was caused by the flaws of international financial markets, including harmful speculation and the lack of coordination between foreign investors withdrawing their money from East Asia. For others, a major reason behind the crisis was the failure of the East Asian corporate sectors to create shareholder value. A spectacular version of the latter view appeared in the Western financial press under the heading "crony state capitalism" – a phrase that invokes the ex istence of strong and often hidden reciprocal relationships among industry leaders, their government patrons, and complicit bankers aimed at circumventing market mech anisms. Later events, however, demonstrated that the US and Western Europe are not immune to severe corporate governance problems. In 2001, one of the largest US companies, Enron, until then considered a model of cor porate governance, admitted to fraudulent activity by concealing gigantic losses and went bankrupt bringing demise also to their fraud assisting auditors Arthur Andersen. In 2003, the Dutch corporation Ahold, the third largest food retailer in the world, also announced huge previously concealed losses and found themselves on the verge of bankruptcy. During the periods preceding fraud detection, in both Enron and Ahold top managers' and directors' salaries were spiraling upward.

Corporate governance problems around the world have necessitated responses by national governments and international organizations. In the US, the Sarbanes Oxley Act of 2002 has introduced stricter requirements with regard to corporate accounting, internal control systems, directors' responsibilities, and investor pro tection. At the international level the most influential initiative was the 1999 publication of Principles of Cor porate Governance by the Organization for Economic Co operation and Development (OECD), promoting the protection and equitable treatment of shareholders, responsible boards of directors, timely and accurate disclosure, and recognition of the rights of other stake holders. Both the Sarbanes Oxley Act and the OECD Principles have influenced the development of corporate governance codes and standards in many countries. The interest of national governments and international eco nomic organizations in corporate governance can be seen as a part of a shift from national and international eco nomic coordination by monetary and fiscal policy, to a greater emphasis on institutional development, including the rule of law. By undermining investors' confidence, corporate governance problems can unsettle the financial systems of whole countries.

Another set of interests shaping the development of corporate governance has been institutional portfolio investors, mainly pension funds, mutual funds, and in surance companies, which gained prominence in inter national finance in the 1990s largely as a result of a shift from pay as you go to funded pension systems. This began in high income countries but has spread to the rest of the world. These investors are large, but aim to hold highly diversified portfolios, so that their individual in vestments typically do not exceed several percent of any company's equity. Institutional portfolio investors are therefore corporate outsiders, but larger and more powerful than individual shareholders. Those skeptical of the positive influence of institutional portfolio investors argue that, managing other people's money, they struggle with their own internal governance problems, and may collude with corporate managers at the expense of small shareholders and other stakeholders. On the other hand, they have demonstrated a growing interest in the cor porate governance of companies and even whole coun tries, and have thus sustained private sector corporate governance services.

Such services range from legal assistance for share holders trying to influence corporate decision making or intending to sue a company for governance misconduct; through advice to companies on investor relations or on the design of internal control procedures and systems; to corporate governance ratings, whereby individual com panies are rated according to standards of good govern ance. They include law firms, management consulting companies, large accountancy firms, credit rating agen cies, as well as specialized corporate governance service providers. In addition, stock exchanges have responded to the prominence of corporate governance debate by cre ating stock market indices based on the stock prices of companies that satisfy certain standards of corporate governance. This gives qualifying companies additional visibility in the market, and promotes the stock market as a place for trustworthy investment by individual and institutional investors. In summary, there is a growing market for corporate governance services and standards, as corporate governance is seen as an important source of risk by investors, and a significant risk category for managers trying to attract external finance.

Theoretical Underpinnings

With the significance of corporate governance subject to interdisciplinary debate it is difficult to identify its the oretical underpinnings clearly. Nevertheless, the theories that have so far had most influence on the debate in corporate the interrelated concepts of agency, contracts, and collective action.

At a basic level, relationships between managers and stakeholders in a company can be conceptualized as those between agents and their principals, so that the central question becomes how the former act in the interests of the latter. This raises several problems. First, even if managers work with the best intentions, they have to use their business judgment and make timely decisions in reacting to the market, without procrastinating on how each decision may affect stakeholders. Second, and more serious, in reality we cannot expect all managers to be so devoted to stakeholder interests. By the nature of their positions, managers have superior access to corporate information compared to stakeholders. They may there fore make decisions to enrich themselves at the expense of stakeholders, for example, by overpaying for transac tions with firms controlled by relatives, or by engaging in projects that bring them social prestige without maxi mizing the interests of the company.

One way for principals to make agents act in their interest is by entering into legally enforceable contracts, such as employment contracts, loan contracts, contracts with suppliers, or service and sale contracts with customers. The existence of enforceable contracts is highly significant for the proper functioning of corporate governance, but contracts do not present an ultimate solution to the agency problem. The relationships with managers set in the con tracts may be affected by the relative power of different stakeholders. For example, employees may have employ ment contracts with the company but the conditions offered may still be poor. Contracts may also be in complete, and may not fully protect the principals. So, employees cannot be guaranteed lifetime employment in a particular company, and owners' contracts do not give ex plicit guarantees of return on their investment. The view that corporations should be run solely in the interest of owners is primarily based on the argument that they put everything they invest at risk, while other stakeholders have contracts which, although never perfect, guarantee some predetermined rewards.

Two other generic problems complicate the relation ships between corporate managers and stakeholders. First, different groups of stakeholders have different, often con flicting interests. An obvious example is that corporate profit can be used either to increase salaries, to pay bonuses to managers, or to pay dividends to owners. Managers may be able to take advantage of such conflicting stakeholder interests. For example, spending excessive corporate re sources on local cultural and sporting events may earn them social prestige. Second, there may be many entities or individuals within a group of stakeholders. Despite some commonality of interests this can therefore lead to col lective action problems. Even if it is highly beneficial for shareholders as a group to exert pressure on corporate management, the costs for individual shareholders may exceed their expected benefits. In other words, shareholders may wait for others to take the initiative in the hope of free riding on the potential shared benefits. Collective action and free riding problems also apply to other groups of stakeholders including customers and employees, within consumer protection associations, and trade unions.

Major Mechanisms of Corporate Governance

Boards of Directors and Executive Compensation

Boards of directors (boards), typically elected and re moved by shareholders, and themselves appointing and removing senior managers, are meant to act as inter mediaries by representing the interests of shareholders to managers. In some countries the boards also include representatives of other groups of stakeholders, for ex ample, employees. The main problem with boards acting as representatives of shareholders is that they are often dominated by the management or by large owners at the expense of small shareholders. Their members have limited access to information compared to managers and lack incentives to confront managers. One commonly proposed governance standard for the functioning of boards is their degree of independence from manage ment. Specifically, it is recommended that at least half of the directors have no links with management or large owners, and that the company chief executive officer (CEO) should not be the same person as the chairman of the board. Independent directors are believed to be more likely to defend shareholders' interests, although research shows no clear cut evidence for the benefits of in dependent boards. One of any board's tasks is also to determine the level and structure of compensation for top managers. In theory, one way to align the interests of managers with those of shareholders would be to base some part of executive compensation on corporate fi nancial performance and especially stock market prices. Skeptics nevertheless point out that managers may ma nipulate both of these.

Concentrated Ownership

The existence of large owners is the most common mechanism to counterbalance the power of managers in corporations. The concept of concentrated ownership is relative. It might refer to a majority stake exceeding 50%, or one of only 10% if the company otherwise has diluted ownership. However defined, large owners have more incentives than small owners to monitor management. The main disadvantage of concentrated ownership is that large owners may treat themselves preferentially at the expense of small investors, as well as employees and other stakeholders. Another problem with large shareholders involves the trade off between control and liquidity. Large blocks of shares offer control but the number of potential buyers for them is much smaller than that for small amounts of shares. Thus, the prevalence of companies with concentrated ownership depletes the li quidity of the stock market, and hinders the development of the latter as a source of external finance for companies.

Hostile Takeovers

In company takeovers the bidder makes an offer to share holders of the target firm to buy all or a fraction of shares at a stated tender price. In hostile takeovers, the objective is to gain majority representation on the board, and to appoint new board members and a new CEO against the wishes of the current management. The bidder gains control if it acquires more than 50% of the voting shares. In practice, the threat of a hostile takeover often disciplines management and, even if the threat itself is insufficient, poorly performing management may be removed after a takeover bid. Most corporate governance codes support the process of hostile takeovers, recommending restrictions on the use of takeover defenses, such as subjection of share transfers to board approval. The latter may reduce the threat of hostile takeovers, but can also strengthen the bargaining position of the target, often for the benefit of target's shareholders. To further complicate the picture, there is no conclusive research evidence that hostile takeovers are directed mainly at poorly performing firms, nor that the efficiency of companies increases after a takeover. Hostile takeovers may also be disruptive for companies, where human capital and the creation of informal relations of trust between management and other stakeholders, including employees, is important. If these parties know that current management can be so easily replaced, they may be reluctant to commit themselves to such trust based relations and developing firm specific human capital.

Disclosure

Disclosure enhances shareholder and other stakeholders' access to information about corporate affairs that would otherwise be restricted to management. A distinction needs to be made between disclosure of financial statements and corporate governance disclosure. Financial disclosure contributes to corporate governance but strictly speaking is a subject of research primarily in accounting, not corporate governance. With regard to corporate governance it is expected that companies should disclose and make publicly available information on the rights and duties of shareholders, antitakeover protection, and board structure and functioning, including details of board members' and executives' remuneration. Disclosure is stressed by all corporate governance standards, ratings, and codes, and it is probably the least controversial mechanism of corporate governance. While a corporation may have a sensible explanation for particular antitakeover protection or for the lack of independent board members, there seems to be no justification to keep these elements of corporate governance secret from the public. To be sure, transparency of corporate governance does not interfere with sensible commercial confidentiality protecting sensitive information concerning products, markets, or innovations.

International Corporate Governance

Corporate governance mechanisms vary greatly between countries. While dispersed ownership structures prevail in large US and UK corporations, concentrated ownership is dominant in the rest of the world. Germany, Austria, Switzerland, and the Netherlands, for example, have a two board system in which a separate management board is overseen by a supervisory board, with members of the latter barred from performing management functions. In some countries including Germany, Austria, Netherlands, Denmark, Sweden, Luxembourg, and France companies satisfying certain criteria, mostly with respect to size, are required to have employee representatives on boards, thus involving what is called co determination. Hostile takeovers are very uncommon outside the US and the UK.

The combinations of corporate governance mechanisms in different countries are referred to as national systems or regimes of corporate governance. There are two main typologies suggested by international comparisons of corporate governance regimes. At a global level a distinction is made between countries with more or less effective corporate governance and those where corporate governance is ineffective. The allocation of countries between these groups corresponds roughly with the division between developed and developing countries, with transition and emerging economies on the path to achieve effective corporate governance. Developing countries are considered to be struggling with basic corporate governance problems including the underdevelopment of independent and efficient courts that could enforce contracts between stakeholders, and low levels of corporate transparency. Within developed economies a further distinction is made between countries where corporate governance is based on stock markets, and those where it is centered on large owners, often with a relatively active role played by employees, banks, and the government. A stock market based regime prevails in Anglo American countries, with a preponderance of dispersed ownership, institutional portfolio investors, stock market based compensation, and relatively frequent takeovers. Continental Europe, Japan, and the developmental states of South Korea, Taiwan, and Singapore are considered main bastions of corporate governance based on concentrated ownership.

Such typologies of national corporate governance systems tend to be based on very broad generalizations. Corporate governance systems are embedded in the wider socioeconomic systems of respective countries, and are affected by multiple historically and geographically contingent factors. These include the corporate scandals that have made public opinion suspicious of the operations of stock markets, cosy trust based relations between stakeholders, an emphasis on cooperation and solidarity as social priorities, rather than competition and individualism, as well as the often covert political and ideological orientations of business. Moreover, corporate governance systems can differ within countries, between regions, sectors, and individual companies. Industries relying on human capital may require different corporate governance from those depending more on financial capital. Companies with long research, development, and implementation periods may require different solutions from those with short project cycles. Different corporate governance arrangements may also suit different regions depending on their inherited socioeconomic structures.

Corporate governance systems around the world are not set in stone. During the 1990s and 2000s, stock market based executive compensation, shareholder suits, and hostile takeovers have entered the corporate stage in Continental Europe and Japan. Many companies, from Europe, Latin America, and Asia cross list their shares on Anglo American stock markets, thus subjecting themselves at least partially to Anglo American corporate governance regimes. Institutional portfolio investors, led by pension funds, have established themselves internationally. In terms of regulations, the OECD Principles, and the 'Cadbury Code' in the UK, have been emulated to varying degrees in many countries. The dynamics of corporate governance trigger the question of whether there is a 'best' system of corporate governance, which the world might be moving toward. Some commentators welcome the shift toward the principles of Anglo American corporate governance developed in 1990s and early 2000s, claiming that such convergence is driven by internationalization of product and capital markets. Others argue that diversity of socioeconomic circumstances, systems of value, and traditions will continue to resist the imperatives of global finance, recalling that in the 1980s, before Anglo American corporate governance 'came into fashion', Japanese and German corporate governance regimes were highly praised for being conducive to long term cooperative investment in research and development, and close relationships with suppliers.

Given the multifaceted, embedded, as well as historically and geographically contingent nature of corporate governance, there are ample opportunities for cross fertilization between human geography and other disciplines contributing to the debate. On the one hand, familiarity with corporate governance issues can enhance human geographers' understanding of the firm as a nexus of contracts and often conflicting relationships between multiple constituencies, and also of the influence of owners, including shareholders, on the direction of corporate change. On the other hand, the analysis of the spatial diversity and appropriate differentiation of corporate governance is an area where human geography can significantly contribute to the corporate governance debate. Whatever the direction taken by corporate governance across the world, a significant part of change involves growing spatial mobility of managers and employees, as well as the activities of capital owners. It seems reasonable to expect corporate governance, understood as a complex set of relationships between different groups of stakeholders, to become more adaptable to their varying identities across widely differing commercial global environments. If human geography offers an insight into such mobility and adaptability, it may also offer key insights into the future dynamics of corporate governance.

See also: Business Services; Capital and Space; Corporate Responsibilities; Financial Risks and Management; Firms; Globalization and Transnational Corporations; Privatization.

Further Reading

Becht, M., Bolton, P. and Roell, A. A. (2003). Corporate governance and control. In Constantinides, G. M., Harris, M. & Stulz, R. (eds.) *Handbook of the economics of finance* (1 edn.). pp 1 109. Amsterdam: Elsevier/North Holland.

Berle, A. A. and Means, G. (1932). *The modern corporation and private property*. New York: Macmillan.

Christopherson, S. (2002). Why do national labor market practices continue to diverge in the global economy? The 'missing link' of investment rules. *Economic Geography* 78, 1 21.

Clark, G. L. and Wójcik, D. (2007). *The geography of finance: Corporate governance in a global marketplace*. Oxford: Oxford University Press.

Denis, D. K. and McConnell, J. J. (2003). International corporate governance. *Journal of Financial and Quantitative Analysis* 38, 1 36.

Jensen, M. C. (2000). *Theory of the firm: Governance, residual claims, and organizational forms*. Cambridge, MA: Harvard University Press.

La Porta, R., Lopez de Silanes, F., Shleifer, A. and Vishny, R. (1998). Law and finance. *Journal of Political Economy* 106, 1113 1155.

Mallin, C. A. (2004). *Corporate governance*. Oxford: Oxford University Press.

Monks, R. A. G. and Minow, N. (2004). *Corporate governance*. Oxford: Blackwell Publishing.

Morck, R. K. (ed.) (2005). *A history of corporate governance around the world: Family business groups to professional managers*. Cambridge, MA: National Bureau of Economic Research.

OECD (1999). *Principles of corporate governance*. Paris: OECD.

Roe, M. J. (2002). *Political determinants of corporate governance: Political context, corporate impact*. Oxford: Oxford University Press.

Shleifer, A. and Vishny, R. W. (1997). A survey of corporate governance. *Journal of Financial Economics* 52, 737 783.

Wójcik, D. (2003). Change in the German model of corporate governance: Evidence from blockholdings 1997 2001. *Environment and Planning A* 35, 1431 1458.

Relevant Websites

http://www.ecgi.org
European Corporate Governance Institute (ECGI).

http://www.dcgn.de
German Corporate Governance Network.

http://www.gcgf.org
Global Corporate Governance Forum, International Finance Corporation.

http://www.icgn.org
International Corporate Governance Network (ICGN).

http://www.oecd.org
Organisation for Economic Co operation and Development (OCED).

http://www.ssrn.com
Social Science Research Network (SSRN).

http://www.thecorporatelibrary.com
The Corporate Library.

http://www.worldbank.org
The World Bank.

Governance, Good

G. Williams, University of Sheffield, Sheffield, UK

Glossary

Clientalism The structuring of political power through informal networks of 'patron–client' relationships, in which powerful patrons provide their dependent clients with resources in return for their loyalty.

(Third World) Debt Crisis The default of many countries in the global South on rapidly increasing debt service payments in the early 1980s, which threatened not only the viability of their own economies, but also sections of the international banking industry (commonly dated to Mexico's default on its international debt in 1982).

HIPC 'The Heavily Indebted Poor Countries' initiative was launched by the World Bank and the IMF in 1996. To qualify as heavily indebted, a country must have annual per capita income below the threshold for eligibility for concessional borrowing from both the World Bank and the IMF, and external public debt that exceeds 150% of its annual exports.

Liberal Democracy A form of government characterized by periodic elections in which a plurality of political viewpoints vie for a popular mandate, and underpinned by constitutional restrictions on state power and the protection of individual rights.

Millennium Challenge Corporation Established in 2004 by President George Bush, the MCC is the body that administers the Millennium Challenge Account, a bilateral development fund directed at 'well-governed' developing countries; the fund received more than US$ 2bn over the first 2 years of its operation.

Neoliberalism A capitalist political philosophy, initially set up in opposition to Keynesian ideas of economic management, that stresses the importance of a 'free-market' approach to economic growth.

PRSP Poverty Reduction Strategy Papers are national programs for poverty reduction. PRSPs should present a policy framework and agenda that has been produced by a national government for tackling poverty and has that government's full commitment. Completing a PRSP – which then must be approved by the Bank and the IMF – is a critical criterion in qualifying for HIPC debt relief.

Rent Seeking The extraction of value from others without making any contribution to productivity – particularly used to refer to the use of government power to extract such 'rents' by regulations which restrict free-market competition.

Introduction: 'Good' Governance and Its Evaluation

Governance, as noted elsewhere in this encyclopedia, is a concept that encompasses forms of rule and the exercise of power in ways that extend far beyond the institutions and actions of government alone. Definitions of govern ance are often all encompassing as a result (**Box 1**); but despite this breadth, good governance has come to take on a very specific meaning within international devel opment. While all states and regimes have always used their own norms and values to monitor the quality of their rule to some extent, the explosion of interest in governance among international development agencies since the late 1980s represents something rather different. It has aimed to produce a degree of global consensus over what constitutes 'good governance', and to make gov ernance reform a legitimate and central issue within their development agendas. The World Bank's own definition of good governance (**Box 2**) is thus intended not only to be an international standard against which all countries can be measured, but delivering good governance has also become a major concern of its own development pro grams, where it is seen as a causal variable in the delivery of economic growth and poverty alleviation. The ability to define norms of good governance is clearly an im portant exercise of power in itself – the attempt to en force them through development interventions, still more

Box 1 Defining governance

UNDP. "Governance is viewed as the exercise of economic, political and administrative authority to manage a country's affairs at all levels. It comprises mechanisms, processes and institutions through which citizens and groups articulate their interests, exercise their legal rights, meet their obligations and mediate their differences".

Institute of Governance, Ottawa. "Governance comprises the institutions, process and conventions in a society which determine how power is exercised, how importance decisions affecting society are made and how various interests are accorded a place in such decisions".

International Institute of Administrative Sciences. "Governance refers to the process whereby elements in society wield power and authority, and influence and enact policies and decisions concerning public life, and economic and social development".

From Weiss, T. G. (2000). Governance, good governance and global governance: Conceptual and actual challenges. *Third World Quarterly* 21(5), 795 814.

so. A concern with the quality of governance, an in herently political issue, also represents a new departure for bodies such as the World Bank and the International Monetary Fund (IMF), who are prohibited in their art icles of agreement from interfering in the internal pol itics of nation states.

Underlying the seemingly reasonable measures in **Box 2** are a set of values and assumptions that both define good governance in very particular ways – as a feature of liberal, democratic, and capitalist regimes that resemble those in the West – and claim legitimacy in projecting these values on to others. The emergence and spread of the 'good governance agenda' thus raises questions that are central to the concerns of political and development geography: about the nature of contemporary inter national power relations, the degree to which the sover eignty of individual developing countries is being challenged, and about the evolving roles of international development agencies in the post Cold War era. To ad dress these questions, it is necessary to first look at how and why development institutions became concerned with governance, to see how the good governance agenda has evolved over time, and finally to briefly review the practical mechanisms through which it has been enforced.

Box 2 The World Bank's six dimensions of good governance

1. 'Voice and accountability', the extent to which a country's citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and free media.
2. 'Political stability and absence of violence', perceptions of the likelihood that the government will be destabilized or overthrown by unconstitutional or violent means, including political violence and terrorism.
3. 'Government effectiveness', the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies.
4. 'Regulatory quality', the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development.
5. 'Rule of law', the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, the police, and the courts, as well as the likelihood of crime and violence.
6. 'Control of corruption', the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as 'capture' of the state by elites and private interests.

From Kaufmann, D., Kraay, A. and Mastruzzi, M. (2006). Governance Matters V: Aggregate and Individual Governance Indicators for 1996 2005. Washington: The World Bank.

The Emergence of 'Good Governance' as an Aim of International Development

Until the 1980s, international development agencies showed little interest in the quality of governance, or even the nature of the state itself, in the countries to which they provided assistance in the global South. Following the tenets of modernization theory, it was expected that developmental states would need to be proactive and strongly interventionist in transforming 'traditional' economies and cultures, and international agencies such as the World Bank and UNDP primarily treated governance issues as an 'internal' matter. Where multilateral or bilateral aid relationships did seek to exert pressure on individual Southern countries, this was dir ected more at the international geopolitical concerns of the Cold War and the ending of European colonialism than it was with the nature of rule within countries.

This pattern began to change over the 1980s for a number of reasons. First, the World Bank and the IMF had begun to push a more aggressively neoliberal agenda that directly defined 'bad' governance in terms of too much state intervention in the economy. The Bank's 1981 report 'Accelerated Development in Sub Saharan Africa' (com monly called 'the Berg Report', after its editor Elliot Berg) was a landmark document here, arguing that poor agri cultural performance in the region was a result of poor government policy. State procurement mechanisms pro vided cheap food for urban dwellers, but Berg argued that doing so they undermined farmers' incentives to raise production. The recommendations of this report (dis mantling state run Marketing Boards, and devaluing local currencies to encourage export led agricultural growth) were indicative of a wider policy change in the Bank that saw opening up of Southern economies to the global market as essential for their development. This new policy direction gained real teeth as the 1980s 'debt crisis' spread across the global South; the World Bank and the IMF engineered a series of 'structural adjustment programs' (SAPs), bailing countries out of debt, but only on the condition that they accepted significant international oversight of domestic economic policy. SAPs were thus important in extending the 'legitimate' intervention of international development agencies into domestic affairs.

Second, the clear political illegitimacy of some Southern regimes, from white only rule in Rhodesia and South Africa to the extremes of Pol Pot's Kampuchea or Jean Claude Duvalier's Haiti, suggested that domestic governance structures should be a subject of international concern. Indigenous pro democracy movements in Latin America, Eastern Europe, and elsewhere over the 1980s suggested that there was some popular legitimacy for calls for 'good governance' that emphasized the import ance of citizens' voice and governments' accountability, expressed through multiparty democratic states.

Finally, the collapse of Eastern European communist regimes and the USSR itself by the end of the decade cemented these changes. In practical terms, the absence of a global counterweight to the US was important in en abling international financial institutions to play a more interventionist role in the internal affairs of national governments. There was no alternative source of economic support for countries facing SAPs, and far fewer geo political constraints on the financial institutions' areas of operation, as shown by the setting up in 1991 of the new European Bank for Reconstruction and Development for the former Soviet economies. In theoretical and ideo logical terms, the 'triumph' of the West meant that neo liberal economic policies and liberal democratic ideas of the role of the state were to play a major role in the restructuring of the post Cold War world.

By the start of the 1990s, the stage was thus set for development agencies to take a greater interest in the nature and quality of governance in countries of the global South. Although the 'good governance agenda' was (and still is) shaped by the geopolitical context in which it emerged, there have been a number of important changes within this agenda, and it is these that we turn to now.

Good Governance – Key Moments in a Changing Agenda

As Adrian Leftwich observed, good governance can be defined at a number of different levels. At a fundamental or systemic level, debates around the term have assumed that good governance will take place in the context of democratic states operating within a global capitalist system. At the level of political structures, Southern nations have been expected to follow a western inspired model of the state. This includes a clear division of powers between the legislature, executive, and judiciary, and a political system in which parties are allowed to openly compete for a democratic mandate. At a more surface and practical level, good governance is also supposed to involve significant administrative and managerial reform of the public sector to produce a state with an efficient, accountable, and professional bureaucracy.

Given the World Bank's articles of agreement, it has been this final and more limited sense of administrative reform which has been the explicit focus of its own good governance agenda, although the other two elements have also been implicit within its approach. The Bank's 1989 report 'Sub Saharan Africa: From Crisis to Sus tainable Growth' is credited with popularizing the term 'good governance' in this sense, and was itself a response to the criticism of the effects of SAPs unleashed on the continent in the wake of the Berg Report. Whereas Berg had seen 'bad' governance as being primarily the result of an oversized state and problematic macroeconomic pol icy, the 1989 report tempered this view slightly. The 'rolling back' of the state by widespread privatization of public assets continued, but alongside this was a recom mendation for increased expenditure on human resource development. This recognition of the need for 'adjust ment with a human face' did not, however, leave the existing structure of the public sector unchallenged, and instead ushered in widespread administrative reforms. Thus the report advocated reshaping the delivery of basic public services such as primary education, primary health care, and water supply in accordance with market prin ciples, and the use of user charges as a means of (partial) cost recovery. More generally the Bank policy of the early 1990s continued the 1980s theme of governance reform through two major policy directions. The first was restructuring the internal rules of government: key concerns here were improving external transparency (through reform of budgeting and public auditing rules) and strengthening institutional checks and balances on executive power (some of these – such as the call for independent judiciaries and central banks – represented a fundamental reshaping of the overall architecture of the state in line with idealized Western norms). Second, the state was to open its operations up to competitive pres sure; here, alongside the outright privatization of state functions, there was an increased emphasis on principles of new public management to produce internal markets for government services and performance based in centives for government personnel.

In 'World Development Report 1997: the State in a Changing World', the Bank expanded on the underlying thinking behind the individual reform programs it had advocated over the 1990s. The role of the reformed state (**Table 1**) was envisaged as providing the supporting conditions by which markets could deliver sustainable

Table 1 Matching the state's roles to its capacity: a World Bank view

Functions of the state	Addressing market failure			Improving equity
Minimal functions	Providing pure public goods			Protecting the poor
Intermediate functions	Addressing externalities	Regulate monopoly	Overcoming imperfect information	Providing social insurance
Activist functions	Coordinating private activity			Redistribution

From World Bank (1997). World Development Report, 1997: The state in a changing world. Washington, DC: The World Bank.

growth (in particular, correcting those areas where markets were absent or failing) and providing some degree of social justice (improving equity between rich and poor). There was a call in this report for governments to 'get the basics right first' – to focus on fulfilling the minimal functions of the state – until they had the capacity to progress to more technically demanding aspects of promoting growth and equity. Alongside this agenda of mapping the state's roles to its capacity, the report recognized that after almost 2 decades of neoliberalism, a degree of state 'strengthening' was also necessary; public sector reform was about processes of institution building and not merely cutting back the state. The Bank's earlier strategies for achieving reform by changing the internal rules of government and introducing competitive pressures were now enhanced by a third mechanism: enhancing voice and partnerships. It is here that the good governance agenda really began to move beyond the actions of government alone – with NGO support, public–private partnerships, and community participation all being seen as important parts of governance networks alongside a restructured relationship between the state and the market.

From the turn of the millennium, the debates on what constitutes good governance within the Bank moved on still further. Given the wider focusing of international development intervention around the 'Millennium Development Goals', good governance was presented as integral to poverty alleviation, and 'World Development Report 2000/1' contained detailed discussion of how governance reform could be made 'pro poor'. In part, this was a recasting of its existing governance agenda through the lens of poverty (e.g., the report presented curbing corruption not merely as an end in itself, but as instrumentally necessary in attacking poverty because the costs of corruption are disproportionately borne by the poor). In addition, this report also recognized the social and political barriers marginalizing poorer groups from engaging with the state, and suggested a number of overtly political strategies to overcome these. Distancing itself still further from its pronouncements on public service delivery in the 1980s, the Bank has also recently acknowledged that sound management of public services needs to be responsive to a range of local conditions (**Figure 1**).

Since the late 1980s, the World Bank's good governance agenda has thus evolved considerably in both its intentions and its content. From being primarily concerned with 'rolling back' the state, it has moved to increasingly sophisticated forms of public sector reform conducted in partnership with both the private sector and the civil society. Its ultimate aims have also expanded reflecting a broadened set of mainstream development objectives that have now (re)instated an explicit focus on poverty alleviation and citizen participation alongside the narrower market led growth agenda of the Thatcher–Reagan years. This broadened agenda has brought the Bank into making policy recommendations that are explicitly political in their content (such as building alliances to strengthen the 'voice' of poor people within government), a move which sits uneasily next to its repeated attempts to present reform as a technical/administrative exercise designed to implement an allegedly 'universal' and 'neutral' vision of what constitutes good governance. The evolving agenda has, however, largely left the Bank's underlying assumptions of this vision untouched. Its aim remains the creation of an efficient, but fairly minimal, state that can support a dynamic, market led economy. The means of reaching this end remains reforms that are primarily 'internal' to Southern countries, and operate by 'getting the incentives right' such that the self interested action of government employees and agencies support the greater common good.

There are, of course, conceptions of good governance within the international development community that differ greatly from those of the Bank. The UN based institutions (such as UNICEF and UNDP) have always placed a greater emphasis on the human costs and consequences of development, and within good governance debates, this has meant that they have stressed questions of human rights and democratic accountability more than administrative efficiency or service delivery. A more overtly political agenda is also present within the growing community of international NGOs (such as Human Rights Watch) that have an interest in governance issues. In addition, international intervention in 'failing' states – actively promoted by UN Secretaries General Boutros Boutros Ghali and Kofi Annan, over the 1990s – may be uneven in its implementation and effectiveness, but its justification on humanitarian grounds represents a significant development. It places limits on national sovereignty in extreme cases of bad governance and also supports an alternative conceptualization of good governance that places at its center, states as responsible political actors rather than mere facilitators of economic growth.

A focus on the activities of the World Bank is, however, justified because it remains such an important international development institution, and has emphasized good governance so strongly within its own lending. According to Nanda, there has been a proliferation of activity from the Bank – between 1998 and 2000, it ran over 600 governance related programs in 95 countries and it initiated significant governance reform in the public sector in at least fifty countries. To really understand how the good governance agenda is reshaping the relationships between states and their citizens in the South, it is therefore necessary to look at the concrete mechanisms through which this wave of reform has been implemented.

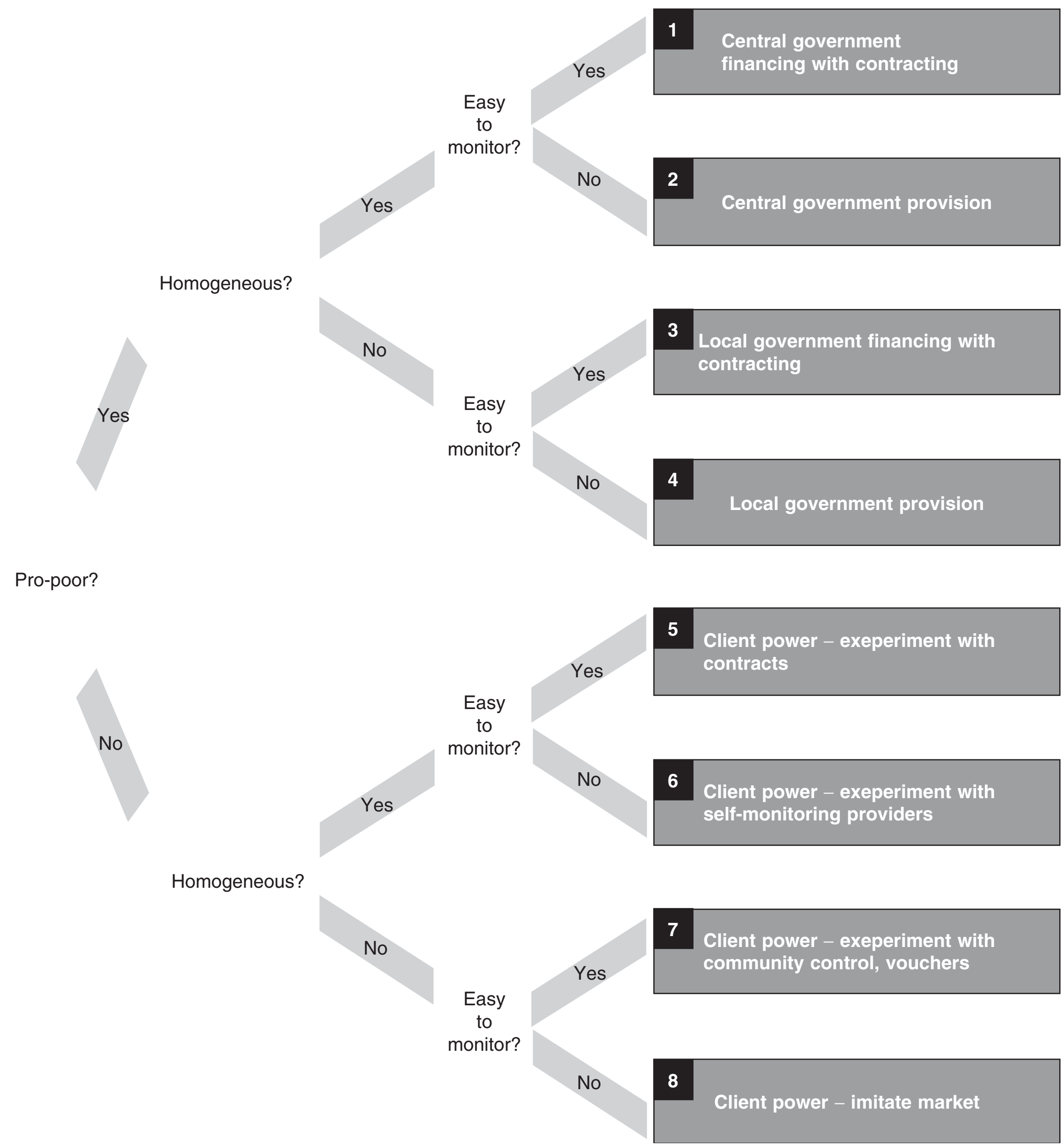

Figure 1 Eight sizes fit all? Rethinking public service delivery in World Development Report 2004. In its recent thinking, the Bank has tried to recognize the range of political and other conditions that might affect the delivery of public services. Different models of delivery (numbers 1 8) are recommended depending on whether the overall political regime is pro-poor (or clientalist), whether or not the population the service aims to reach is relatively homogeneous, and whether or not the quality of the service delivered is easily monitored. Importantly, market competition at the point of delivery is now seen as only one of a range of possible institutional arrangements. From Devarajan, S. and Shah, S. (2004). Making services work for India's poor. *Economic and Political Weekly* 36, p. 916.

Aid Conditionality and the Implementation of 'Good Governance'

As with the earlier generation of SAPs, the conditions placed on development finance have been a major mechanism through which countries of the South have been required to commit to a good governance agenda. For some, this has meant involvement in multiple rounds of new aid programs aimed at improving governance structures and building state capacity (**Table 2**). For those already facing high levels of international debt, the achievement of debt relief, through the recent 'heavily indebted poor countries' (HIPC) initiatives, has also meant meeting a series of governance related conditions. In both, similar techniques have been used to 'discipline' government, such as improving information management systems to make public sector spending more transpar ent, and enforcing centralized budgetary management

Table 2 Governance reform initiatives in Tanzania and Uganda

Year	*Uganda*	*Tanzania*
1989	Public Service Review and Reorganisation Commission	Public sector reform as part of economic and social adjustment program
1990		
1991	Creation of capacity building Secretariat	
1992	Civil Service Reform Programme	
1993	Economic and Financial Management Project I	Parastatal and Public Sector Reform Project (PPSRP) Financial and Legal Reform Project
1994	*Workshop on Ethics and Transparency [Economic Development Institute; now World Bank Institute]*[a]	
1995	Institutional Capacity Building Project	*Service Delivery Survey [Economic Development Institute; now World Bank Institute]*[a]
1996		End of first phase of PPSRP Tanzanian Revenue Authority created
1997		
1998		Second phase of PPSRP
1999	Public Sector Management Project Economic and Financial Management Project II	
2000		Launch of PSRP (2000 2011) Privatisation and Private Sector Reform Project
2001	Privatisation and Utility Sector Reform *Public Expenditure Review*	Accountability, transparency, and Integrity Project *Public Expenditure Review*
2002	Capacity Building/Civil Service Reform Capacity Building II	
2003	Capacity and Performance Enhancement Program Local Government Development Project II	

[a] Note: Italics denote World Bank support for research attached to reform initiatives. From Harrison, G. (2005). The World Bank, governance and theories of political action in Africa. *British Journal of Politics and International Relations* 7(2), 240 270.

within which there is donor oversight. With regard to the latter, countries required to sign up to a 'medium term expenditure framework' effectively fix budget allocations in conjunction with the Bank's advice, enforcing both overall financial restraint and tying individual ministries' spending to their performance. This extensive external intervention in countries' internal affairs is justified in two ways. First, there is intended to be explicit national 'ownership' of the reform packages, whether through identifying internal 'champions' or 'change agents' who will drive forward civil service reform, or through the government's own drafting of the 'poverty reduction strategy papers' (PRSPs) which are central to HIPC based debt relief. Second, and as part of the wider move from 'government to governance' from the 1990s, part nership and civil society involvement in governance re form has also been stressed. Thus the production of PRSPs is explicitly intended to involve the participation of the population at large, both through local assessments of poverty itself, and through civil society input to the policy making process.

The net result of these mechanisms in delivering 'good' governance is open to debate. Certainly, they are making formal government decisions (especially those over budget allocation) more visible and predictable for donor agencies. To this extent, they are ensuring a greater degree of government accountability, and pos sibly limiting some aspects of corruption. But questions have to be asked about the wider impacts of donor driven accountability: if governments are repeatedly to look 'upward' to donors for approval of their actions, their sense of responsibility 'downward' toward their own citizens may actually be eroded. There certainly seems to be evidence for this in the states featured in **Table 2**, despite the fact that both are considered as success stories by the Bank: in Tanzania, 'civil society' engagement in the PRSP has been largely orchestrated through a select group of international NGOs, and in Uganda it has undercut earlier reforms aimed at improving local government's financial autonomy and responsiveness to local conditions and 'voice'.

To reinforce its own claims to objectivity, transparency, and accountability, the World Bank has been involved in an ongoing program of research to find reliable statistical measures for each of the six aspects of good governance highlighted in **Box 2**. Its series of 'Governance Matters' reports, dating back to 1997, have outlined the evolving methodology by which several hundred individual perception measures – that draw data from a range of sources such as global opinion polls, the country analysts of major multilateral development agencies, NGOs such as Amnesty International, and commercial risk rating agencies – have been amalgamated to produce their 'worldwide governance indicators'. These allow the investigation of trends in the quality of governance over time, and inter country comparisons are used to offer statistical evidence to support the Bank's claims that improving governance will foster economic growth and alleviate poverty (**Figure 2**). Given the Bank's championing of 'transparency', it is somewhat ironic that virtually the only elements of the data set that make up the indicators that have not been made publicly available are its own Country Policy and Institutional Assessment data.

It is almost impossible for any aggregate measure to capture what will inevitably be very diverse experiences of governance, but these indicators have become powerful pieces of information in themselves regardless of whether they tell us anything significant about the ground realities within individual countries. As the international donor community seeks to maximize the impact of a limited aid budget, governance indicators are playing a new role in aid conditionality. Rather than commitments to 'future improvements' in governance being the condition of lending or debt relief, within USAID and the USA's Millennium Challenge Corporation (MCC) a state's 'established track record' of good governance is becoming a qualifying condition for future lending. The MCC uses the Bank's world governance indicators within a set of sixteen measures of good governance, and to qualify for its support a country must score better than the median in at least half these measures, and surpass the median on controlling corruption. This explicit decision to focus the USA's bilateral development aid on those who are already 'good performers' is a new and worrying trend; if the Bank's understanding of the importance of governance reform as a driver of economic growth and poverty alleviation is correct, this amounts to writing off precisely those countries that are most in need of improvements in government capacity, and consigning their citizens to a future of poor economic and state performance.

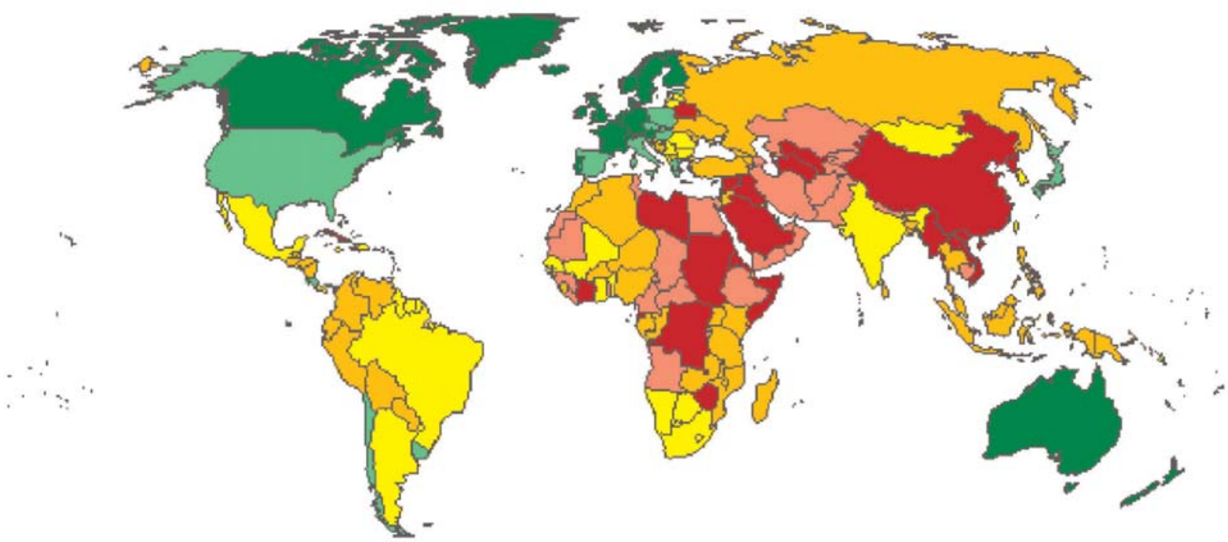

Figure 2 Mapping good governance. The World Bank's governance and anti-corruption web pages allow access to the 'world governance indicators' dataset. The website allows users to make a number of geographical, historical, and indicator-specific comparisons of the quality of governance using this data, such as this map showing the relative rankings of different countries in terms of 'voice and accountability' one of the Bank's six dimensions of good governance. In maps such as these, the complex (and contested) methodologies that link a myriad of data sources to the Bank's own definitions of good governance largely recede into the background, and a powerful and deceptively simple image of a world divided into 'strong performers' and 'problem areas' remains. Key: 90th 100th percentile (dark green); 75th 90th percentile (light green); 50th 75th percentile (yellow); 25th 50th percentile (orange); 25th 50th percentile (pink); 10th 25th percentile (red). Source: http://info.worldbank.org/governance/kkz2005/worldmap.asp#map.

Questioning the Good Governance Agenda

'Good governance' is thus very far from being a neutral term, and its implementation raises important questions about the exercise of power by international development agencies, and as such it has been subject to a range of criticisms from those opposed both to dominant con ceptualizations of the problem of 'bad governance' and to the actual practices of 'reform'. One fairly limited form of critique is to accept that the good governance agenda may be desirable in theory, but, to argue that, in practice it is 'setting the bar too high' for developing countries. The list of areas in which governments 'could' improve their practice, and therefore reform is to be applied, has lengthened considerably over the last decade. As a result, Merilee Grindle has asked whether searching for 'good enough governance' – analyzing where and why examples of relatively good performance emerge within a country specific context, and from this looking to spread their effects to other areas of state activity – is therefore a more appropriate policy response than insisting that all countries respond simultaneously to all areas of an ever growing global agenda. More broadly, historical analysis shows that democratic and publicly accountable insti tutions took centuries to evolve in the West (and were often the result of economic development, rather than one of its pre conditions); for the international donor community to suggest that Southern countries should effect similar changes within a decade or so of intentional reforms is thus somewhat hypocritical according to Chang.

A more radical critique, informed by debates in crit ical development geography and political geography, would question the theoretical underpinnings of the good governance agenda. From a geographical perspective, its neoliberal conceptualization of the reform process can be seen offering an analysis of power that is spatially de ficient in three important respects. First, the Bank's thinking on political agency and change cause it to ignore the specificity of the places in which reform plays out. Viewing governance reform through the lens of rational choice offers a hopeful prognosis; once externally mo tivated 'change agents' and 'champions' have freed bur eaucracies from the shackles of perverse incentive structures (such as rent seeking or of clientalist politics), individual self interested behavior will generate its own momentum for reform in any and all contexts. This is, however, an analysis that ignores the historical and geo graphical conditions which produced 'bad governance' in the first place. Both Marxist and post colonial theories of the African state offer their own explanations of why political and economic power are frequently conflated in Africa, and why a neoliberal vision of the mutually supporting roles of states and markets is often untenable there. As Graham Harrison has argued, by ignoring a contextual understanding of how political elites operate, the Bank risks both mistaking the 'performance' of reform for significant change, and overlooking how aid given in the name of 'good governance' may be reproducing their power.

Second, implicit within the Bank's approach is the idea that the conditions of 'good' and 'bad' governance are primarily located 'within' Southern countries them selves. By framing questions of governance through a national lens, this mode of analysis crucially underplays the power of donor agencies and other international players as part of the 'problem' as well as the solution. For many Southern countries, earlier rounds of SAPs played a direct role in eroding and undermining state capacity and the World Bank's failure to recognize this history is at best a highly convenient piece of institutional amnesia. Other donor agencies offer a more realistic analysis: the DFID 2006 White Paper, 'Making Governance Work for the Poor', contains welcome recognition of the fact that it is often the actions of powerful agents in a globalized economy such as transnational corporations and 'first world' states that create the conditions under which bad governance is reproduced in the South. This analysis is taken to its logical conclusion, and the White Paper ar gues that any credible good governance agenda must address current failures in 'global' governance (by actions as varied as tackling international money laundering to enforcing transparency on international public and pri vate sector transactions) instead of merely pointing the finger at 'inefficient' public bureaucracies within the South.

Finally, there needs to be more critical consideration of the global geopolitical conditions under which a good governance agenda has developed. Debates around the nature of good governance have taken place in an international context that has itself undergone dramatic changes: the world at the end of the first decade of the twenty first century looks far different from the late Cold War period in which governance debates initially emerged. It is difficult to see how arguments for 'good governance' could have expanded so rapidly over the 1990s, and been so closely related to liberal democratic ideas of the role of the state, if an ideological and geo political counterweight to the West had existed during this period. But if a Western derived conceptualization of good governance has enjoyed a free reign since the col lapse of the USSR, it may not do so in future. As emergent nations in the global South, and China in particular, begin to flex their economic and political muscles internationally, the World Bank, the IMF, and individual donor agencies such as USAID or DFID may well find that they have a declining ability to enforce *any* notion of good governance through aid conditionality. Foreign investment conditional on a string of measures to ensure transparency, accountability, or public sector reform may seem like an irrelevant affront to national sovereignty for those governments lucky enough to have natural resources valuable to the new economic powerhouses of the twenty first century. Where countries are resource rich but governance poor, the challenges of molding governance structures in the image of the West may well be set to escalate rather than diminish, despite the growing sophistication of aid agencies' agendas.

See also: Aid; Civil Society; Debt; Governance; Modernization Theory; Neoliberalism; Neoliberalism and Development; Nongovernmental Organizations; Participation; Poverty; State; Structural Adjustment.

Further Reading

Chang, H. J. (2002). *Kicking Away the Ladder: Policies and Institutions for Economic Development in Historical Perspective*. London: Anthem Press.

Craig, D. and Porter, D. (2003). Poverty reduction strategy papers: A new convergence. *World Development* 31(1), 53 69.

Devarajan, S. and Shah, S. (2004). Making services work for India's poor. *Economic and Political Weekly* 36, 907 919.

DFID (2006). *Eliminating World Poverty: Making Governance Work for the Poor A White Paper on International Development*. London: HMSO.

Grindle, M. S. (2004). Good enough governance: Poverty reduction and reform in developing countries. *Governance: An International Journal of Policy, Administration and Institutions* 17(4), 525 548.

Harrison, G. (2005). The World Bank, governance and theories of political action in Africa. *British Journal of Politics and International Relations* 7(2), 240 270.

Jenkins, R. (2001). Mistaking governance for politics: Foreign aid, democracy and the construction of civil society. In Kaviraj, S. & Khilnani, S. (eds.) *Civil Society: Histories and Possibilities*, pp 250 268. Cambridge: Cambridge University Press.

Kaufmann, D., Kraay, A. and Mastruzzi, M. (2006). *Governance Matters V: Aggregate and Individual Governance Indicators for 1996 2005*. Washington: The World Bank.

Leftwich, A. (1993). Governance, democracy, and development in the Third World. *Third World Quarterly* 14(3), 605 624.

Mercer, C. (2003). Performing partnership: Civil society and the illusion of good governance in Tanzania. *Political Geography* 22, 741 763.

Nanda, V. (2006). The 'good governance' concept revisited. *Annals of the American Academy of Political and Social Science* 603, 269 283.

Weiss, T. G. (2000). Governance, good governance and global governance: Conceptual and actual challenges. *Third World Quarterly* 21(5), 795 814.

World Bank (2003). World Development Report, 2004: Making services work for poor people. Washington, DC: The World Bank.

Relevant Websites

http://www.dfid.gov.uk
The Department for International Development (UK Government).

http://www.mcc.gov
The Millennium Challenge Corporation (USA).

http://www.worldbank.org
The World Bank.

Governance, Transport

I. Docherty, University of Glasgow, Glasgow, UK
J. Shaw, University of Plymouth, Plymouth, UK

Glossary

Accessibility The degree to which goods and services can be reached. Although transport is clearly of importance here, it is not the only factor which determines accessibility: Internet access, proximity of services to markets, and structural factors such as age, gender, poverty, and so on need also to be taken into account.

Filling In The introduction of new and/or different state functions to replace old ones being discontinued. An example is the rolling out of a system of contract-based management to govern the private sector provision of public services, to replace a command-and-control approach when the service provider was in government ownership.

Governance A more inclusive notion than that of traditional government, acknowledging that public policy is delivered by a range of organizations, from the state through to the private and voluntary sectors. The term also refers to the nature of the relationships between these organizations.

Government The formal institutions of the state such as ministries, departments, municipalities, and so on.

Hollowing Out The ceding by a government of its powers to other organizations. These might be 'above' the administration in question – such as a sovereign government transferring powers over an area(s) of its jurisdiction to a supranational body such as the European Union – or 'below' – such as to a regional assembly or a private sector contractor.

Mobility The ability of individuals or groups to move around.

Natural Monopoly The situation where a good or service is most efficiently provided by one supplier. Examples are those industries involving a highly capital-intensive network such as water and electricity and, in the transport sector, roads and railway infrastructure.

Introduction

Transport has long been recognized as fundamental to the functioning of societies and economies, but the extent to which governments attempt to influence transport activities varies considerably. It will also depend upon a range of factors from the practical, such as the availability of finance, to the more abstract, such as changing political philosophies and theories of governance. The nature and scope of state intervention can also produce different outcomes affecting the geography of not only transport but also economic, environmental, and social activities. This article explores the rationale for state involvement in transport, and in so doing explains how trends in the governance and ownership of transport systems and operations have changed over time.

Governments all over the world are involved in transport by providing, or regulating the provision of, infrastructure and services. Particular interventions in a transport market will be determined by the policies developed and adopted to suit prevailing circumstances. One definition of transport policy is "the position that some level of government takes on a particular transport issue," and Rod Tolley and Brian Turton refer to "the process of regulating and controlling the provision of transport." Although transport policy is both a public and a private sector endeavor – private companies will themselves have policies to govern their operations – governments are frequently the most involved in making and administering policies because many components of the transport system are owned or managed by the state.

Policies for transport are derived not just in relation to particular modes, but will also relate to different spatial scales. Some, such as European Union (EU) regulations, will apply supranationally; others will be national or regional in scope. Local authorities will generally be concerned with policies relating to a neighborhood, town, or city. The originating body/bodies and relevant spatial scale of a policy will depend on the arrangements for government in existence in the territory to which the policy applies. In the USA, for example, policies might flow from federal, state, county, and city governments, in addition to some multijurisdictional districts in metropolitan areas. We note later in the article that the recent shift in many countries from a traditional type of 'government' toward a more complex system of 'governance' can result in more joint working between layers of government and other stakeholders.

It is not surprising that changes in transport policy can have significant impacts upon the transport geography of the area to which it applies. It might, for example, alter a rail network by providing for the construction of new high speed lines bringing places very much closer together in terms of travel time (such as the TGV network in France – **Figure 1**) or the closure of a relatively

Figure 1 TGV network in France.

little used passenger route (such as the infamous 'Beeching cuts' of the 1960s in Great Britain). Shifts in transport policy will not just alter the geography of transport activities: depending upon the type of change involved, the economic, social, and/or environmental geographies of an area will also be affected. For example, the TGV reduces greenhouse emissions by abstracting passengers from the airlines (an environmental impact at the global scale), and by reducing journey times influ ences the (re)location of businesses and residential de velopments (economic and social impacts on a local scale, and potential environmental benefits at the local and global scales if car usage is reduced as a result of more walking and public transport usage). Likewise, broader policies such as a rise in interest rates or the general level of taxation might impact upon the geography of transport activity within a given area by, say, cutting the demand for rail or air services and prompting a reduction in the levels of service provided to certain destinations.

Why Is the State Involved in the Transport Sector?

Transport is often described as a universal good, since most human activities depend on it to some extent. Markets depend on raw materials and finished products being transported from producers to consumers, and the myriad social and economic interactions we undertake in our daily lives – from visiting friends, to traveling to school, work, business meetings, and the shops – require movement between places. Managing this complex pattern of mobility and the infrastructure and services that it requires has long been at the heart of why governments seek to involve themselves in transport. A better understanding has also been gained of the need to manage transport's 'negative externalities' or, in the words of Bill Black, the "events [such as car pollution] that result in significant disbenefits for persons who had little or no role in the decision making that led to those events." A further factor, and one which has recently increased in prominence following terrorist activity in Western countries, is the influence of military and national security considerations. In addition to bespoke facilities supporting, for example, air force and army bases, large scale ostensibly civilian networks such as the Interstate system in the USA have been developed with national security in mind. Equally, decommissioned military installations, especially airports, are often reclassified for use as civilian transport hubs.

At the most basic level, the state is involved in the regulation and management of transport activities because the conditions rarely exist for a genuinely free market to function. Throughout much of the twentieth century, the narrative of transport governance was dominated by economic and social imperatives. On the one hand there was the need to address natural monopoly (i.e., where a good is most efficiently supplied in monopoly conditions) and market failure within transport systems, and on the other there was a desire to promote social equity in the transport opportunities available to individuals. These economic and social imperatives are, of course, interlinked.

Monopoly in Infrastructure Provision

The street is a good example of how certain aspects of transport activity tend toward natural monopoly. Streets do not just provide for mobility, they are also spaces that enable the exchange and sharing of goods and knowledge on which civic, economic, and cultural life depends. It makes no sense to have competing sets of streets; as a public good, the benefit of the street is maximized when it is part of a single, coherent geography to which access is open to all. Conceiving of the street in this way brings the role of governance squarely into focus because access to streets needs to be managed if the levels of security, social welfare, environmental protection, and economic interaction demanded by liberal societies are to be realized.

Economic theory emphasizes the potential of carefully targeted state intervention to maximize a range of benefits to society unlikely to be expressed through private profit in the market. In transport terms, this means that public provision of key infrastructure and services can be more efficient and equitable than the market alternative. Indeed, it is often essential, since the cost and risks involved mean private companies are frequently reluctant to provide them without some form of government underwriting. State intervention can also avoid wasteful competition and/or duplication of assets, such as separately owned, parallel streets and railway lines; this can also ensure that good value is obtained from the limited capital resources available within an economy if the planning and implementation of transport investment is carefully managed.

The power of this argument is reflected in the fact that it became orthodoxy for governments of all political persuasions for much of the twentieth century. Moving away from the model of competing private sector railway companies (many of whom had gone bankrupt or were forced to merge) and eliminating private toll highways had obvious, immediate appeal to the political Left. Interestingly, though, it was also attractive to many on the Right. Innovations such as expressways and passenger aviation provided the opportunity for the state to seek a step change in the efficiency of the market economy by transforming the level of mobility available to individuals and firms so that trade would increase substantially and competition would spread.

Social Equity

The second key goal of transport governance has been to intervene in the provision of transport infrastructure and services so that a range of social policy goals can be achieved. Controls and systems of subsidy disbursement built up from the nineteenth century in order to promote customer, employee, and vehicle safety and to prevent the emergence of significant social and financial disparities between areas and groups. Examples of regulatory interventions are those designed to: ensure affordable rather than monopoly fares; guarantee concessionary fares for the young, old, and disabled; promote a high quality of service provision; and stipulate the quantity of services provided to ensure (so far as was reasonable) a comprehensive transport network, especially in thinly populated areas. Social democratic and socialist governments alike nationalized various land transport modes, and air transport was publicly owned from an early stage in many countries.

A significant element in the notion of transport equity is the development and support of public transport to (among other things) allow for trips to be made by disadvantaged individuals. Overall social equity can be maximized through the provision of a transport system which enhances the accessibility of a wide range of services to as many people as possible. Indeed, the promotion of transport equity is often thought of as a key aspiration within a comprehensive statement of policy objectives for the enhancement of quality of life. Moreover, widening access to life opportunities for disadvantaged groups can also serve to promote broader economic benefits by, for example, increasing the mobility of a potentially neglected part of the labor force.

It is also important to note that the capital cost of providing any new transport initiative should preferably be at the lowest social cost. This social cost will include the cost of the negative externalities associated with the transport development. These externalities were until relatively recently thought of as being primarily at the local scale, such as land take for new infrastructure developments and adverse impacts on air quality. In addition, the building of a piece of transport infrastructure between two places reduces the available resources to do the same elsewhere. Investing in roads reduces the potential to enhance the railway network; subsidizing public transport fares reduces the capacity to improve infrastructure; and so on; it can be difficult to get the decision between infrastructure projects 'right', and there are many who would argue that the state is generally a poor judge.

The Neoliberal Age

Compelling as it was for decades, wide ranging public sector involvement in the transport sector was in the end undermined as an approach to transport governance. Stagnating economies and a sharp increase in oil prices in the 1970s prompted governments to find ways of cutting spending across the public sector. At the same time, a new generation of national leaders including Margaret Thatcher and Ronald Reagan also saw political and philosophical merit in reducing state expenditure and the size of the public sector more generally. Administrations in countries such as the UK, the USA, and New Zealand started to adopt policies informed by neoliberal thinkers such as Milton Friedman, Friedrich von Hayek, William Niskanen, and Charles Tiebout.

As a political and economic philosophy, neoliberalism countered the prevailing assumptions underpinning the 'welfarist' economies of much of the developed world by stressing the need to reduce state intervention in the economy and promote market forces which encourage competition, enterprise, and individual self reliance. Neoliberal ideas developed and spread through a range of institutions and political networks throughout the globe, including the International Monetary Fund (IMF) and the World Bank, and were adopted by governments across Central and Eastern Europe following the collapse of the Berlin Wall. Adam Tickell and Jamie Peck suggest that neoliberalism has provided "a kind of operating framework or 'ideological software' for competitive globalisation, inspiring and imposing far reaching programmes of state restructuring and rescaling across a wide range of national and local contexts."

The transport sector has been affected significantly by this shift in political philosophy. Notwithstanding its peculiarities, from the neoliberal perspective transport is the same as any other good, subject to market forces and the rigors of competition. The private car became seen as a symbol of freedom and progress – occupants can within reason go anywhere at any time – and governments deemed sustained increases in personal mobility facilitated by the car essential for economic growth. Large road building programs were often implemented to facilitate increased traffic volumes as a result. In respect of those modes in competition with the car, successive governments championed a range of reforms including the removal of competition restrictions – the process of deregulation – in an attempt to open up a market for services previously provided only by the public sector, and the downsizing, administrative fragmentation and in many cases outright privatization of state owned transport undertakings. These measures were deemed necessary to reduce the perceived oversupply of transport services, improve transport providers' responsiveness to changing customer demand (often a function of rapidly increasing car ownership and use), and limit the need for public sector ownership, management, and subsidy. The procurement and financing of infrastructure also became more dependent on the private sector as governments looked to reduce their spending and debt liabilities.

In continental European states, the 'welfare' economic model generally retained more influence, and the restructuring and liberalization of the transport sector only really gathered pace in the 1990s. Compared to what is (inaccurately) regarded as the 'Anglo Saxon' world, political discourse was far less influenced by faith in deregulation as the means to economic efficiency. Instead, reform proceeded at a slower pace, with the objective being to deliver a true European common market for major transport services such as road haulage, aviation, and the railways dominating policy shifts. In some countries, this process was accompanied by the privatization of state owned companies in an effort to strengthen their competitiveness as well as to limit public subsidy.

There is little doubt that some of these market reforms succeeded in realizing their intended cost savings, improvements in process efficiency, and reductions in bureaucracy. Equally, others did not bring about most of the positive effects forecast in terms of improving the scope and quality of transport services. In many instances, reform was sufficient to significantly alter the pattern and nature of service provision. This brought clear advantages, but there were also adverse effects such as service contraction, reductions in public transport patronage, and increasing traffic congestion.

Redefining the Role of the State

Just as the welfarist model of wide ranging state provision of transport services was superseded by neoliberal ideas, so policies in the first decades of the twenty first century seem likely to have a different emphasis to those in the 1980s and 1990s. In many countries, something of a retreat from – or at least a redefinition of – neoliberalism is evident, as politicians seek to reengage the state with many areas of public policy. Governments have begun to adopt a more interventionist stance toward transport infrastructure and service provision. In those countries where the pace of the neoliberal project was greatest, politicians are reflecting more on the experiences of other countries where reform proceeded more slowly, and where the desire to achieve economic efficiency never attained preeminent status.

Hollowing Out and Filling In

In one sense, the bold rhetoric of neoliberalism reducing the role of the state was always misleading. It is true that a 'hollowing out' of the state's existing functions was taking place, as privatization and deregulation significantly curtailed the public provision of services. Yet, as Clive Charlton and Richard Gibb point out, despite the "undeniable prominence of liberalised transport systems," there was "an apparent contradiction between the broad policy support for deregulation and privatisation and the continued, and even enhanced, regulatory environment affecting many transport systems." This observation refers to the fact that, in reality, it was never possible to privatize or deregulate many transport markets to the point where the state could simply withdraw from them. This is not least for the reasons discussed earlier in this article – monopolies could not be broken up, consumers had to be protected, and social policy goals could not be ignored – and progress toward neoliberal goals had to be tempered by a degree of *realpolitik*.

In other words, the process of 'filling in' was always in evidence. In the transport sector (as elsewhere), this has often been expressed by the introduction of contractual agreements – as a substitute for the command and control management style possible under direct ownership – to ensure that a range of outcomes specified by the state is actually delivered by the private sector. The British government's decision to protect unprofitable rural bus services following deregulation is one example (the same industry might soon become subject to tighter regulation in order to address some of the ongoing shortcomings in service provision).

Rail privatization in Great Britain has involved the tight regulation of a monopoly infrastructure owner and a plethora of contractual stipulations to which the passenger train operators have to adhere. In large part, this was to guard against the significant network contraction – effectively a complete recasting of Britain's railway geography – that would have ensued if private sector operators had sought to maximize revenue by closing unprofitable routes. The near collapse of the privatized rail industry in 2000–01 following a fatal crash at Hatfield, near London, saw ministers rescue the bankrupt infrastructure company and provided a stark illustration of the state's continuing role as guarantor of the last resort in many areas of transport endeavor, even in a supposedly privatized system of provision.

Governments are facing the reality that while market based transport policies work well in some sectors, they function less effectively in others and in some cases tend not to function at all. The same applies in terms of the geographical implications of such policies: at the simplest level, the market will provide services in areas where they are profitable, but without appropriate regulation will precipitate their withdrawal or downscaling where money cannot be made. All of this means that privatization and deregulation should not be interpreted as a simple reduction of state intervention and control. Rather, they are part of a broader process of state restructuring which incorporates elements of both deregulation and reregulation as established modes of organization are dismantled and new ones established.

Public and Private Interdependence

At the same time, as this process of 'filling in' of the state's functions is becoming better understood, parallel developments generating increased interaction between the state and private transport interests are coming increasingly to the fore. At least three important trends are apparent here: the 'crisis of mobility' brought about by worsening congestion; the recognition of the need to pursue environmental policy goals through intervention in transport markets; and the general shift to a more complex system of governance.

The enormous growth in car (and air) traffic levels in recent decades has left a legacy of severe congestion in much of the developed world. Despite significant road construction programs, governments (or private sector holders of building concessions) are unable to build themselves out of congestion. Partly this is because the financial, political, and practical limitations on building new infrastructure are real – roads are expensive, attract a lot of opposition, and fill up very quickly after they are built (the phenomenon of 'induced traffic') – but it also dawned on politicians from the Right that continuing to accommodate such vast increases in mobility was increasingly paradoxical as it implies very high levels of government spending.

The resulting crisis of mobility has been an important driver of the state's re engagement in the regulation of the transport sector. Essentially, some form of demand management needs to be put in place once the level of congestion reaches a certain threshold, otherwise the costs of this congestion jeopardize economic efficiency by creating delay and uncertainty in the transport system. The London congestion charge is a good working example of demand management – congestion has fallen by 26–30% following the introduction of a daily charge – and the Dutch government is introducing a nationwide charging scheme early in the next decade.

The second significant factor is the scale of transport's contribution to global environmental damage. A critical introductory passage from the Brundtland Report noted the need for concerted government action because of "a growing realization… that it is impossible to separate economic development issues from environment issues… and environmental degradation can undermine economic development." Not only has an extensive literature developed in the field of environmental economics that attempts to formalize and quantify some of these relationships, but also the case for government intervention in the transport market is much stronger since the international scale of transport related pollution and emissions has become apparent. States and other government bodies are widely seen as the only institutions with the power to make any meaningful impact on such strategic processes, even if ultimately they rely on market mechanisms such as road user charging – derived from the work of environmental economists among others – to ascribe the costs of transport more appropriately to users.

The more explicit recognition by many recent governments of the importance of environmental as well as economic and social policy aspirations underscores the contemporary 'three legged stool' conception of sustainability as applied to transport policy, which seeks to balance these often competing demands. It is also interesting to note that the emergence of congestion and pollution as significant multiscalar issues repositions the state firmly in the role of arbitrator between public and private interests, as envisaged by those economists who identified nearly 100 years ago the potential for carefully targeted state intervention to benefit society. In order to address congestion and pollution, it is becoming necessary for the state to take a leading role in restricting individual mobility.

The third factor to consider is changes to the structures of governance witnessed in many countries as state institutions have sought to adapt to changing socioeconomic contexts. The emergence of supranational institutions (such as the EU) and other international treaties and organizations has challenged existing states from 'above', and in some countries there have also been challenges from 'below' in the form of (re)emerging nationalist movements and/or more vocal desires for increased regional autonomy and discretion in policymaking. The establishment of devolved governments within the UK in Scotland, Wales, and Northern Ireland illustrates this well. Such 'spatial' hollowing out – powers transferring up or down to institutions with different jurisdictional territories – combines with the processes of functional hollowing out and filling in to produce complex, multi layered systems of governance often involving stakeholders from the voluntary and private sectors. Understanding how this impacts upon transport regulation and control – as for other areas of public policy – can be complicated and requires an analysis of the changing ways the state seeks to achieve its goals.

Bob Jessop's work on the processes underlying the changing character of state reorganization is useful in explaining how the regulation of transport might begin to look in a 'post' neoliberal world. The first such process, de nationalization, encapsulates the apparent spatial hollowing out of transport regulation. From 'above' comes, for example, the EU's drive for the harmonization of transport markets between states in areas such as air traffic control and open access to rail markets. From below, subnational governments take on responsibility for the specification and franchising of transport activities such as local rail services in France, Germany, and Sweden. The second process, de statisation, reflects the increasing role played by nonstate actors in regulation and service management, such as the control of some regional airports in Europe by Chambers of Commerce or wide ranging public–private partnership consortia. Finally, the third process, the interstate transfer of policy, is clearly evident across the transport sector, be it the lessons learned from the privatization of British Rail (generally regarded as an overly complex and highly costly experiment), or the influence of the German *Verkehrsverbund* tariff management concept in the franchise models of bus provision in many major metropolitan areas across the globe.

In making sense of these developments, states are increasingly seeking to follow what Anthony Giddens characterizes as a 'third way', between neoliberal and welfarist thinking, by combining elements of market discipline in transport procurement and delivery with stronger state regulation for economic, social, and environmental reasons. Governments are both reacting to new threats such as widespread congestion and the environmental crisis, and addressing less immediate but equally important trends, such as aging demographic profiles that place new demands on systems of mobility and connectivity. They also attempt to replicate the old consistency of state provision of services with complex regulatory systems based on contractual relationships. And in doing all of this they routinely – or at least frequently – share or cede/delegate responsibility to a far greater range of public, private, and voluntary sector stakeholders than has historically been the case. Although the neoliberal experiment was designed to facilitate a shift to the market provision of transport services, the need for state intervention in the system, albeit in a different form, has very clearly reasserted itself.

Conclusion

This article has explored transport governance and ownership from a geographical perspective. It is important to understand the role of the state in transport because territorial governments are heavily involved in the management and ownership of transport activities; the strategies and policies they adopt over time will shape the development of transport infrastructure and services in a variety of ways at a variety of spatial scales. In addition to military and national security considerations, taken as a whole, state interventions in the transport sector seek to promote economic development and efficiency while pursuing social objectives such as widespread mobility and accessibility, and addressing environmental imperatives like global warming and local air quality.

The level and scope of state intervention in transport activity will depend upon a range of factors. The finances available to politicians at any given time are clearly important, but more abstract factors such as political ideologies must also be taken into account. A major refocusing of state involvement in the transport sector was attempted in response to the rising influence of neoliberal ideology, although in reality the extent to which states were able to withdraw from the transport sector is debatable; in many territories direct ownership and operation of components within the transport market was replaced with management through contract and other regulatory apparatus. Most recently, transport systems around the world have witnessed something of a resurgence in state intervention arising from what might be described as a 'crisis of mobility', heightened concern about the global environmental impacts of transport, and the increasing complexity of the state sector. Looking to the future, it is inevitable that the relationship between transport and the state will continue to evolve in response to broader social processes.

See also: Transport and Deregulation; Transport and Sustainability.

Further Reading

Black, W. (2003). *Transportation: A Geographical Analysis*. New York: The Guildford Press.
Charlton, C. and Gibb, R. (1998). Transport deregulation and privatisation. *Journal of Transport Geography* 6(2), 85.
Docherty, I. and Shaw, J. (2008). *Traffic Jam: 10 Years of 'Sustainable' Transport in the UK*. Bristol: The Policy Press.
Giddens, A. (2000). *The Third Way and Its Critics*. London: Polity Press.
Harvey, D. (2005). *A Brief History of Neoliberalism*. Oxford: University of Oxford Press.
Jessop, B. (2004). Hollowing out the 'nation state' and multi level governance. In Kennett, P. (ed.) *A Handbook of Comparative Social Policy*, pp 11 25. Cheltenham: Edward Elgar Publishing.
Knowles, R., Shaw, J. and Docherty, I. (2008). *Transport Geographies: Mobilities, Flows and Spaces*. Oxford: Blackwell.
Marsden, G. and May, A. (2006). Do institutional arrangements make a difference to transport policy and implementation? *Environment and Planning C: Government and Policy* 24(5), 771 789.
Niskanen, W. (1973). *Bureaucracy: Servant or Master?* London: Institute of Economic Affairs.
Rodrigue, J. P., Comtois, C. and Slack, B. (2006). *The Geography of Transport Systems*. London: Routledge.
Tolley, R. and Turton, B. (eds.) (2001). *Global Transport Issues*. London: I.B. Tauris.

Relevant Websites

http://www.internationaltransportforum.org/
International Transport Forum; Passenger Transport Executive Group.
http://www.stif.info/IMG/pdf/plaquette anglais.pdf
Introducing the STIF (Paris region transport authority).
http://www.tfl.gov.uk/corporate/about tfl/2881.aspx
Transport for London legislative framework.

Governance, Urban

M. Raco, King's College London, London, UK

What Is Urban Governance?

Urban governance is primarily concerned with the processes through which government is organized and delivered in towns and cities and the relationships between state agencies and civil society (a term that is used to include citizens, communities, private sector actors, and voluntary organizations). Governance has multiple meanings. For some, it represents a concept and an analytical approach that opens up new ways of thinking about processes of government, urban politics, accountability, and democracy. For others, governance is a more descriptive term that focuses attention on concrete institutions and their financing, roles, and responsibilities. Key questions for the analysis of urban governance include:

- Who is it that makes decisions about the organization and implementation of policy in urban areas?
- How are these decisions made?
- Who controls agendas and how?
- What policy making processes exist?
- Which institutions/interests have the power and resources to shape policy agendas?
- What control do local people and individual citizens have over the way their cities are governed?

A focus on governance, therefore, draws attention to those interests that have the 'power' to make decisions about policy in cities and how processes of decision making operate. Effective policy development and implementation depends on how systems of governance are organized, shaped, and structured.

These processes are particularly important in urban areas, for it is in cities that the full effects of economic restructuring and de industrialization have been most keenly felt. Cities suffer from high levels of unemployment, crime, and social problems, and often possess deprived districts alongside areas of relative affluence. It is also in cities that, arguably, some of the most difficult contexts for governing can be found. Modern urban areas often possess a greater diversity of social groups than nonurban areas and have a greater concentration of problems to be tackled. Populations are also increasingly urbanized and what happens in towns and cities has broader social, political, and economic impacts beyond the urban scale. Indeed, it could be argued that urban governance is at the heart of broader changes to the organization of the state and political debates over its appropriate roles and functions. In short governance is not simply a matter of technical and/or bureaucratic procedures but consists of a set of processes and structures that shape and constitute broader relations of power, domination, and authority.

Urban governance is principally concerned with the following topics:

- Urban politics: Urban politics relates to the complex relationships between different groups of people and interests in society, especially, those relationships involving the acquisition of power and authority. It is closely related to broader conceptualizations of identity and representation and the decision making processes through which representations and decisions are made. Urban politics is both formal and informal in character. The former consists of elections, political parties, representative systems, and the like whereas the latter relates to the softer social relationships and networks that exist between different groups, actors, and interests.
- Changing modes of democratic representation: Governance is also about the structure of political systems and broader concerns with questions of political representation, democratic legitimacy, and accountability. Well functioning systems of governance both help to create, and are the consequence of, healthy democracies. When systems of governance function poorly, the legitimacy and relevance of democracy to citizens and communities is reduced, and this leads to a spiral of declining participation and legitimacy. The movement towards urban governance has also been elided with broader shifts in the nature of democratic representation. Old fashioned systems of government were built upon 'representative' modes of accountability through regular elections, and the principle that elected representatives carry a democratic mandate to represent those who have elected them. The new urban governance, by contrast, is presented as a system in which individual citizens, and the communities of which they are a part, engage in direct, 'participative' modes of governance.
- Citizenship: This shift to participative governance has significant implications for citizenship and the relationships between citizens and the institutions that govern them. It involves the establishment of a new balance of rights and responsibilities and the redrawing of boundaries of state action and regulation. Participative governance requires a greater focus on

what has been termed 'active' citizenship. In simple terms, active citizens are characterized as being pol itically, socially, and economically independent. This independence enables them to take on more responsibility for their own welfare so that they do not have to rely on state bureaucracies and/or welfare services. Passive citizens, on the other hand, are often portrayed as 'dependent citizens'. Their passivity is increasingly seen, by (neo liberal) governments, as a problem to be cured through government programs and actions. In many countries governments have, therefore, introduced measures to foster new types of citizenship to underpin wider policy initiatives and governance reforms.

- Economic competitiveness: In recent decades, there has also been growing interest in the relationships between governance and economic development. With the growth of the so called 'new economy', economic activity is increasingly dependent on the networks established between firms and the support structures that exist with institutions of government and civil society. Much of the rhetoric that accom panies the emergence of the new urban governance is directly linked to the imagined role that decision making structures can play in creating entrepreneurial and attractive environments for competitive busi nesses and individual entrepreneurs. Decision making and governance, it is argued, needs to be flexible and able to respond to new challenges and opportunities in the global economy. The governance of economic development has become an increasingly important issue in urban politics as industrial cities across the Western world have sought to manage the negative consequences of economic decline.
- Urban sustainability: Governance is one of the key pillars of the sustainability agendas promoted by agencies such as the World Bank, national govern ments, and nongovernmental organizations (NGOs). It is argued that economic development projects will only promote new forms of equity if decision making systems and structures are broken open and made responsive and accountable to a broad section of civil society. There has, therefore, been a renewed interest in the mechanisms through which decisions are made, particularly at the urban scale where, the impacts of development projects on different social groups have been particularly marked. Urban politics, it is argued, can only become truly sustainable if new modes of participative governance are introduced alongside economic and environmental policy programmes.
- The 'hollowing out of the state': For many writers, particularly those writing from a regulationist per spective, the emergence of the new local governance reflects and reproduces the broader transfer of powers away from the nation state to other actors, working at a variety of scales. This process of 'hollowing out' represents a reaction to the expansion of government, local and national, in the post war era. In ideological terms, this represents part of a wider trend towards the creation of what neoliberal writers, such as Osborne and Gaebler, have termed 'leaner and fitter' govern ment. Battles over the form and character of govern ance are at the heart of contemporary political debates over the organization of the state and its capacities and priorities.
- Social control: Governance is also bound up with broader questions of social control. Historically, the ruling classes have seen cities as disordered places that require careful management, regulation, and control. Establishing effective forms of governance is one of the best mechanisms for the quelling of urban unrest whilst enabling certain powerful groups to ensure that their visions, needs, and priorities are articulated in policy programs.

The Rolling-Out of the Governance Agenda

The term governance is clearly not a 'neutral' or apolitical term. In large part, it represents a way of thinking about how politics at a variety of levels should be organized and managed. During the 1990s and 2000s, global institutions, such as the World Bank and United Nations, have taken it up with relish and national governments are increasingly required to reform existing new modes of government in return for assistance. Considerable efforts have been made to define 'good governance', with the United Nations Development Program, for example, claiming that:

> good governance is among other things participatory, transparent and accountable. It is also effective and equitable and it promotes the rule of law. Good gov ernance assures that political, social and economic pri orities are based on broad consensus in society and that the voices of the poorest and the most vulnerable are heard in decision making over the allocation of resources.

Similarly, the self appointed Commission on Global Governance are keen to stress that,

> there is no alternative to working together and using collective power to create a better world. Governance is the sum of the many ways individuals and institutions, public and private, manage their common affairs. It is a continuing process through which conflicting or diverse interests may be accommodated and co operative action may be taken. It includes formal institutions and regimes empowered to enforce compliance, as well as informal

> arrangements that people and institutions either have agreed or perceive to be in their interest.

Bad governance is, conversely, elided with post war welfare states and 'inflexible' bureaucratic procedures and power structures. Indices of good and bad governance, such as those produced by the World Bank's 'anticorruption' unit, are increasingly used to compile comparative 'league tables' of different government systems and the extent to which they can be characterised as 'good' and 'bad'.

Institutions at other scales have been pursuing similar agendas. For the European Union, effective and efficient systems of governance provide the mechanisms in and through which the twin objectives of development policy – economic competitiveness and social cohesion – can be delivered. Its spatial development framework, *Agenda 2000*, specifically calls for new modes of governance to be introduced as a prerequisite for other elements of development policy. Urban policy programs, in particular, have made governance reform a core objective. National governments have also been rolling out governance reforms. For those of a less interventionist, more liberal persuasion, the new agendas reflect and reproduce a new set of philosophies about what the state should be doing to support its citizens and, more importantly, what those citizens should be doing to support themselves. The UK has led the way in Europe, but similar models of reform are being adopted by governments across the EU and beyond.

Conceptualizing Urban Governance

Academic work has focused on both the conceptual and empirical dimensions of urban governance. As with any disciplinary area, very different approaches have been applied. Much of the writing on urban governance has been developed in the United States where there has been a long tradition of strong territorial and urban politics. In the 1950s and 1960s, writers such as Robert Dahl argued that urban politics was, essentially, characterized by open and accountable systems of democracy. Power was dispersed between a plurality of interests and state agencies, particularly urban governments, acted as neutral arbiters, weighing up the competing claims of different interests and making just and inclusive decisions in the name of a 'common good'. This pluralist urban democracy was the embodiment of the American democratic system. Other groups of writers, however, developed much more critical approaches. C. Wright Mills famously wrote of urban 'elites' or groups in whom power was concentrated. Far from representing a plural and open system, urban governance, it was argued, was often dominated by a small number of groups whose power and decision making practices were 'hidden' and systemic rather than open, transparent, and accountable. Elites include economic groups, political cliques, social organizations, private sector actors, and technical–bureaucratic agents. Debates became focused on the extent to which urban politics represented an 'open' or 'closed' democratic system and the relationships between institutions of government and the citizens and communities that they (ostensibly) served.

Following on from these classic accounts, a range of different theories have been put forward that seek to explain how urban politics and governance works. In the 1980s and 1990s, a group of American writers, including Clarence Stone, developed a set of ideas that became known as 'regime theory'. Using examples such as Atlanta, regime writers set out to show that urban governance was dominated by identifiable coalitions of interests working to tightly focused agendas. In different cities, different groups came to dominate these coalitions or regimes. Other writers, however, have argued that such definitions of urban politics are too narrow. In order to capture the processes and practices of governance and decision making at the urban level, authors such as Rod Rhodes suggest that the focus should be on fluid and changeable 'policy networks' that coalesce around particular problems of government such as 'community safety' or 'town center management'. The two approaches do, however, have similarities in focusing on the horizontal relationships of power at the urban level and giving less attention to the vertical or structural relationships of power in which regimes or networks are operating.

Other writers have approached urban politics and governance in a more structured way. Mike Davis, John Logan, and Harvey Molotch, for example, have identified what they term 'urban growth machines' as dominant players in the development of many US cities. Drawing on cases such as Los Angeles and Las Vegas, they argue that urban politics has become a vehicle for powerful coalitions of developers, locally dependent companies, and leading politicians to take control of profit making urban development programs. Developers are structurally privileged actors in these new, relatively closed arrangements. Neo Marxist authors, such as David Harvey, go further and argue that urban governance, primarily, serves to reproduce the needs of capitalists at the expense of other interests. Place competition has encouraged new processes of 'urban entrepreneurialism' to emerge in which towns and cities act like private sector businesses, speculating in order to accumulate. For Harvey, urban politics has become trapped in a zero sum game in which places compete vigorously for increasingly mobile forms of investment and what Richard Florida terms a 'creative class' of entrepreneurial workers. Whilst such approaches have significant explanatory power they have also been criticized for being reductionist and for failing to acknowledge the complexities of governance relation ships that emerge in different places.

Other groups of writers have approached urban governance from a rather different direction. Neo Foucaudian authors, such as Nikolas Rose and Barbra Cruickshank, argue that governance is explicitly con cerned with the control of populations and the shaping of what Foucault termed 'governmentalities'. Such work emphasizes the ways in which structures of governance become infused with relations of power and control so that, actors' ways of thinking and approaching politics and political questions are shaped by those in power. Giving individuals and communities new powers and responsibilities, it is argued, represents a form of aban donment in which the legitimacy and political support given to collective welfare provision is eroded. Citizens have to 'go it alone' with a minimum of support. The ultimately pessimistic and all encompassing nature of such accounts is both their intellectual strength and their greatest weakness.

Urban governance is also closely related to geo graphical questions. In many urban political systems the democratic legitimacy of institutions is articulated on a geographical basis. Elected representatives, often, repre sent particular localities and it is assumed that all those present in a particular place will have shared interests and needs that can be represented politically. However, there are clearly problems with this type of approach. Critics such as Geoff Mulgan, for example, argue that governance should be organized at the 'community' level, rather than at the traditional urban or local government level – as communities represent the key scale of human interactions and imaginations. Peter Taylor has tried to capture these dynamics by focusing on what he terms 'place–space tensions' in the organization of governance. Individuals and communities often associate themselves with particular places, or life worlds. Where a high de gree of congruence exists between these imaginations and the spatial boundaries of governance institutions, the more representative, legitimate, and effective organiza tions will be. Conversely, where there is a mismatch in place–space relationships, tensions are enhanced and systems of governance are weakened and undermined.

Within some strands of geography and political sci ence, many writers now talk of 'multilevel governance' and focus on the ways in which urban governance has to be conceptualized as part of a nested hierarchy of scales, as clearly bounded administrative and political spaces that play an important role in the wider political system. This nesting is most effective where place–space tensions are reduced, and least effective where they are enhanced. However, a number of critical geographers, such as Ash Amin and Doreen Massey, have recently argued that urban politics should be seen in more 'relational' terms. Rather than the conceptualizing of urban governance as a spatially bounded and defined process that relates to particular places, relationalists argue that the focus should, instead, be on the interconnections between places and people. This focus on interconnections makes the category of 'urban governance' problematic in that, what happens in towns and cities is always relationally bound up with external with what takes place elsewhere.

Overall, then there have been competing and con trasting conceptualizations of urban governance. The term's growing political significance has ensured that it remains a key area of debate across the social sciences.

From Government to Governance – The Practices of Governance

The changing nature of urban governance has had sig nificant practical implications for the ways in which urban areas are governed and managed. Since the early 1980s, the rationalities, structures, and mechanisms of urban governance in many countries have undergone major reform. In the post war period, the role and powers of local governments expanded with the growth of Keynesian welfare state systems. Local government played a key role in the delivery of welfare services such as education, housing, planning, and healthcare. In most cases, they were (and still are) run by elected councillors and their decisions had a significant impact on the everyday lives of local residents and businesses.

However, during the 1970s and 1980s, with the rise of neoliberalism the values and rationalities of local gov ernment and the wider welfare state came under attack. In many countries, local governments' powers and re sponsibilities were transferred to a range of other actors drawn from the private and voluntary sectors. Non elected quango agencies (quasi autonomous nongovern mental organizations) also took on many of the activities formally carried out by local government. For example, in the UK in 2006, there were 883 government quangos responsible for spending a total of £128 billion. As un elected agencies, they were managed and run by ap pointees (there were approximately 21 000 appointees in 2006) and their agendas were set through a combination of central government targets and internal management processes. Such organizations raise important questions over accountability and democracy as their formal rela tionships to local civil society are not founded on democratic elections. In addition, this fragmentation of responsibilities has forced governments to try to develop new forms of institutional cooperation at the urban level. There has been a new emphasis on 'joined up working' between organizations, and 'partnerships' have become commonplace.

However, there are also serious deficiencies in the new model. The growing fragmentation of institutional responsibilities for different aspects of governance has made the job of governing an increasingly complex and opaque process. In a context where partnerships have become an increasingly dominant vehicle of decision making, identifying who it is that is responsible and accountable for urban governance has become more and more difficult. The new agendas are also built on a critique of local government that, for political reasons, both exaggerates the 'bureaucratic' nature of earlier periods of governance and underplays the central role that local government continues to have in organizing and managing these new partnerships. One further irony of the 'new' urban governance is that it has gone hand in hand with the extension of central state power through the introduction of new forms of top down 'public management'. Local governments and other agencies that are in receipt of government funds now have to meet centrally directed targets, milestones, and other 'value for money' criteria. In countries such as the UK, such processes are well advanced, and this has an enormous impact on the ability of local actors to act independently and flexibly in response to local circumstances and political pressure.

Future Directions and Trends

The political, social, and economic forces that have underpinned the 'new' urban governance are likely to become more powerful in the coming years. The neoliberal clarion call for fitter, leaner, and less interventionist forms of government is being taken up by governments and agencies operating in very different contexts. The new discourses preach that it is not up to governments alone to govern. Instead, it is up to individuals and communities to take more responsibility for their own lives and to tackle the problems that they face; it is up to city leaders to identify and mobilize those who live and work in their cities and develop political mechanisms of inclusion; it should be the responsibility of business leaders to play a greater role in the development and delivery of welfare services; and it is essential that new forms of government are flexible and light touch and do not interfere with the operation of market systems and business entrepreneurialism. In addition, these new forms of governance are becoming increasingly elided with visions of sustainability and the cross national sustainable city building agenda. There seems little prospect of a reemergence of well funded local 'government' state agencies playing the same significant role that they played in the post war decades.

Advocates of the new urban governance present it as a vehicle of empowerment for communities and citizens. Old vertical hierarchies are broken down, with power and responsibility transferred to a range of new actors. A virtuous democratic circle emerges in which active citizenship generates new forms of engagement and influence which in turn generates a greater desire for more participation, activation, and inclusion. The power of established professional and bureaucratic elites is undermined and the process of governance becomes geared up to the needs of citizens as consumers, who are able to exercise their 'choice' over services and participate directly in their own decision making and governance.

However, critics argue that at the same time as these discourses of choice based governance have become more widespread, socioeconomic inequalities between and within countries have grown inexorably. The emphasis on active citizenship has effectively shifted the responsibility for socioeconomic problems onto those who are often least able to challenge relations of inequality and influence agendas. At the same time, the new focus on partnership has a tendency to cloud or obscure relations of power. The term partnership is often used as a vehicle for 'including' a range of political stakeholders, whilst nullifying opposition and protest. Systems of local governance are, in many ways, becoming 'less flexible' as powers and responsibilities are taken away from elected local authorities and handed to private companies and other quangos. In key policy areas such as transport and healthcare, there is much less scope for changing policy agendas now then there was during the post war decades when utilities and service providers were under the control of elected local government.

Overall, the study of urban governance has taken on a new significance not only within urban geography but also across the social sciences. Its popularity as a term has grown inexorably during the 1990s and 2000s, and it is now commonly used as a frame of reference by state actors and those in the voluntary, private, and social sectors. There remain significant disagreements over the term's meaning and its practical implications for urban politics and government. It is also becoming an overused and increasingly politicized term, making it all the more important that its core concerns with political power, democracy, representation, and accountability remain at the forefront of critical social research.

See also: Entrepreneurship; Neoliberalism, Urban; Social Justice, Urban; Sustainability, Urban; Urban Growth Machine.

Further Reading

Buck, N., Gordon, I., Harding, A. and Turok, I. (eds.) (2005). *Changing cities Rethinking urban competitiveness, cohesion and governance*. London: Palgrave.

Harvey, D. (1989). From managerialism to entrepreneurialism: The transformation in urban governance in late capitalism. *Geografiska Annaler B* 71, 3 17.

Healey, P. (2006). *Collaborative planning: Shaping places in fragmented societies*. New York: Palgrave.
Imrie, R. and Raco, M. (1999). How new is the new local governance? Lessons from the UK. *Transactions of the Institute of British Geographers* 24, 45 63.
Judge, D., Stoker, G. and Wolman, H. (eds.) (1995). *Theories of urban politics*. London: Sage.
Kearns, A. and Paddison, R. (2000). New challenges for urban governance: Introduction to the review issue. *Urban Studies* 38, 845 850.
Kooiman, J. (2003). *Government as governance*. London: Sage.
Osbourne, D. and Gaebler, T. (1992) *Reinventing Government*. London: Penguin.
Rhodes, R. (1998). *Understanding governance: Policy networks, governance, reflexivity and accountability*. Buckinghamshire: Open University Press.

Relevant Websites

http://www.nlgn.org.uk/public/
The New Local Government Network webpage.
http://www.gdrc.org/u gov/ugov define.html
The U Gov website on ‘urban governance’.
http://www.unhabitat.org/categories.asp?catid=25
The UN HABITAT webpage.

Governmentality

J. Häkli, University of Tampere, Finland

Glossary

Biopolitics A term coined by Michel Foucault, meaning the style of government that regulates entire populations by fostering life in all its aspects, including its biological basis, health, reproductive practices, habits, customs, and well-being.

Geopolitics Analysis of the strategic relationship between spatial context and politics at various scales ranging from individual body, home, street, city, region, state to international.

Neoliberalism Political rationality that espouses liberty as a means of promoting social, economic, and political development. Historically refers to social-scientific and political sentiments that manifest themselves in policies supporting market as a form of regulation.

Subjectification Process through which individuals become subjects in the dual sense of forming identity through conscience and self-knowledge, and being subject to dependence of and control by others.

Territorial Sovereignty Principle of international law according to which a government has the exclusive right to political authority within its territory.

Introduction

The term governmentality was first used by the French historian of ideas and philosopher Michel Foucault in the context of his studies on governmental rationality in the late 1970s. The idea of governmentality reached the English speaking world largely through an English translation of Foucault's lecture on governmentality given at the Collège de France in February 1978. Published first in 1979 in journal *Ideology and Consciousness*, and again in 1991 in an edited volume *The Foucault Effect*, the lecture gradually introduced governmentality to a widening audience of scholars in a range of social and cultural studies disciplines.

The concept of governmentality is key to several important aspects of Michel Foucault's thought. Foucault coined the term in the later years of his career, which ended prematurely to his death in 1984. Since the late 1970s, Foucault had been working on issues related to individualization, subjectification, rationality, and government in attempting to clarify the regimes of power that work through the expansion of governmental rationality in modern societies. In particular, he wanted to reveal the ways in which social power is constituted by the regulation and self regulation of embodied action. He insisted that what best characterizes the contemporary mode of government is not the state's encroachment on the liberties of society, but rather the governmentalization of the state. The latter refers to a mode of scientific state government that takes the welfare of the population as its end as well as its means.

While Foucault mainly discusses state government, the concept of governmentality does not point merely at the reality of the state, or even the population as a whole. Of equal importance are the subjectivities that governmental rationality impinges on individuals. Foucault insisted that the modern individual is not a given, but constituted as a domain of initiatives and self government that together operate the dispersed social power typical of modern Western societies. Instead of direct forms of repression and constraint, social control is achieved by means of subtler strategies of normalization, which produce self regulating individuals. This power is operative not only on the individual level but in the context of the state government of population that Foucault calls biopolitics. Above all, Foucault claims that modern regimes of power operate to produce subjects who are both the objects and vehicles of power. Thus, the modern individual is not the opposite of power but rather one of its prime effects.

Hence, in analyzing governmentality, Foucault does not see the state as a singular all powerful apparatus that exists in opposition to the civil society. On the contrary, he is critical of state theory that approaches society as external to the state, or seeks to reduce the state to one of its functions, such as the provision of security or the reproduction of relations of production. By illuminating how the modern state and the modern society are co constructed, the concept of governmentality works to bridge the gap between macro and micro levels of analysis of power.

'Gouvernementalité' – The Birth of a Concept

In the early 1970s, Foucault was preoccupied with the task of understanding forms of disciplinary power based on the disciplining of individual bodies. This mode of power was epitomized by the military rationality he analyzed in *Discipline and Punish*. He gradually extended his analysis of power from disciplinary to biopolitical regimes and the phenomenon of governmentality in two annual courses of lectures at the Collège de France titled *Security, Territory and Population* and *The Birth of Biopolitics*. The term gov ernmentality (in French, *gouvernementalité*) appeared first in the concluding part of a lecture given in February 1978. Foucault summed up what he had described as historical movement in the art of government that brought about population as target of knowledge and governmental intervention. In this context, governmentality refers to three things: (1) the institutional ensemble that allows the exercise of a mode of governmental power that is not re ducible to sovereignty or discipline; (2) the long term trend toward the preeminence of this type of power over the other modes (sovereignty and discipline) resulting in the formation of specific governmental apparatuses and re spective rationalities throughout the West; and (3) the process of the governmentalization of the modern state.

The term governmentality consists of two parts (*gou verner* and *mentalité*), which underlines a semantic link between the acts of governing and modes of thought underpinning them. Understanding this link requires a wider conception of government not restricted to state's politics alone. Foucault shows that before the nineteenth century, the question of government was discussed not only in relation to state government, but also in a broad array of contexts ranging from philosophical to religious, medical, and pedagogic texts. The art of government thus refers to the general reflection of conduct, or the conduct of conduct, seeking guidance in various aspects of life ranging from individual self control or the management of family all the way to the government of population. In revealing that self government of individuals (technolo gies of self) and the government of population (tech nologies of domination) are two sides of the same coin, the concept of governmentality offers novel insight into the nature of social power.

The originality of the term lies in its capacity to move beyond the dichotomy of freedom and constraint, or consensus and coercion, as two opposite forms of power. Governmentality points to a power that operates not simply by forcing people to obey the will of the governor, but rather by balancing between techniques that assure coercion and those processes through which individuals themselves modify or construct self. In governmentality, the direction of the conduct of individuals and a popu lation is based on forms of power that are exercised over subjects who are free. With the history of govern mentality, Foucault purports to show that instead of being opposites, the modern sovereign state and the modern autonomous individual are historically co con structed, reflecting the expansion of governmental ra tionality throughout the West. Hence, what Foucault calls the governmentalization of the state refers to a devel opment by which state power gradually came to rest not merely on sovereignty and discipline but also on gov ernment, that is, the link between the autonomous individual's capacity for self government and rational political objectives.

Mentality is key in the formation of this link in so far as it is through the internalization of particular political rationalities that individuals come to govern themselves in ways that serve politically convenient ends. For ex ample, a person's concern for her or his health and all actions to improve it may first seem choices made by an autonomous individual. However, at the same time matters pertaining to personal health are part of the state's interest in governing public health. This attempt to manage health on the level of the whole population produces a wealth of knowledge providing guidance for individuals on how to lead a healthy life: how to care for oneself, how to eat, exercise, work, engage in sexual conduct, and so on. In Foucault's thought this notion is captured in the concept of biopolitics, which refers to state actions to govern population and the underlying rationality that links, for instance, individual's care of self with the state's care for the public health.

The Rise of Governmentality

From the Middle Ages to the sixteenth century political theory was predominantly concerned with the relationship between the ruler and the sovereign territory, seen in juridical terms. For Foucault the emergence of doctrines concerning the purpose of state in sixteenth century Europe designates the starting point of modern govern mental rationality. At that time the notion of state is dis tinguished from and set above religious considerations, and the new secularized state is given immanent autono mous principles that it should necessarily follow. To know how to govern would then require knowledge of the state and its interests, neither of which are self evident. Hence, governmental rationality starts to build on specific forms of knowledge that observe both the praxis of government and the nature of what is governed.

The first body of knowledge that fully articulates the purpose of the state as the science of government is the eighteenth century Germanic *Polizeiwissenshaft*, or science of policing (i.e., policy). This corpus of theory aspired to maximize the strength of the state by nurturing the real basis of this strength, the population, and life in general. This objective resulted in endless lists of matters per taining to people's happiness and prosperity that the state should be concerned with, ranging from policing health, foods, and highways to policies dealing with commerce, poverty, religion, etc.

While the science of policing now seems too indefinite and chaotic to really serve governmental policy making, it nevertheless paved the way for the increasingly de tailed state intervention into various fields of social life. The multitude of foci that still characterize the science of policing in the seventeenth century were gradually replaced by a more coherent system of thought centered on political economy by the middle of the eighteenth century. Foucault distinguishes a turning point in the breakup of the conceptual link between family and economy, whereby the latter no longer means the management of family property. The new concept of political economy pertains to a totally different, much larger plane, that of the population as a whole. According to Foucault, it was the science of political economy that enabled the identification of problems that are specific to a state's population, and thereby fostered reflection on the problem of government outside the juridical framework of sovereignty.

This gave rise to a new and progressively more intimate relationship between the state and society. In contrast to premodern government confined mainly to aspects of territorial sovereignty, modern government had the general welfare of the population as its purpose, which grew into an ever extending field of state interventions. These governmental policies not only targeted the population but also used techniques immanent to the population as their means, including the stimulation of birth rates, the prevention of diseases, the promotion of healthy diet, the direction of the population into certain occupations and industries, and so on. This productive governmental rationality Foucault calls biopolitics, that is, the attempt to secure growth by means of designing interventions in the field of economy and population.

Summing up this development, Michel Foucault states that the modern state became fully governmentalized roughly by the end of the eighteenth century. While the term governmentality is not restricted to any one historical period or power regime, it is often used by Foucault, as well as many of his followers, in referring to advanced Western liberal democracies. Hence, the notion of governmentality applies particularly to neoliberal societies characterized by dispersed power structures and emphasis on the freedom of choice based on the predominance of market mechanisms. In the context of neoliberalism, individuals have an active role in determining the course of their lives and the state appears as the provider of social welfare services and various possibilities for the fulfillment of each individual's full potential. Neoliberal governmentality nourishes the population by influencing the practices and rationalities of autonomous individuals' self regulation.

Studies in Governmentality

In the English speaking world, interest in the concept of governmentality arose after the publication of an English translation of Foucault's lecture on governmentality in the early 1990s. Since then, governmentality studies have expanded across the whole spectrum of humanities and social sciences, ranging from anthropology to critical criminology; from nursing science to social psychology; and from sociology to political science. Arguably, governmentality is one of the major legacies of Foucauldian thought in contemporary critical social science research.

However, as is the case with Foucault's corpus throughout, there exists no single common way of appropriating his conceptual tool box. The interdisciplinary field of inquiry into modern governmental rationality has brought about a number of different approaches to governmentality. Some of these weigh on the analysis of knowledge production in relation to the rationalities of rule (how regimes of truth constitute the social world as governable and administrable), while some emphasize more on the technologies and practices of government (the operationalization of governmental rationalities in material processes and acts). In the former approach attention is focused on the mentalities and rationalities that underlie rule – the ways in which problems are defined, expertise recognized and enacted, information construed and codified, participants identified and organized, and codes of conduct arranged and circulated. The second approach focuses more on how techniques of the self interact with structures of domination. Attention is paid to government seen as an ensemble of institutions, procedures, analyses, and reflections that facilitate the exercise of this complex form of social power through the activities of actually existing governments.

While there is no single one paradigm, some major themes can be distinguished in the governmentality approach. First, scholars in this area have charted the changes in the state/civil society relationship brought about by the reshaping of government under neoliberalism. This work has shown that political power is exercised today through a multitude of agencies and techniques, many of which are not associated with the formal organs of state. Hence, what in neoliberalism is frequently seen as the withdrawal of the state from its regulatory function vis à vis the civil society, can actually be understood as a new technique of government based on reorganization and transferring the state's regulatory task to responsible individuals. Seen as a specific form of governmentality, neoliberalism is revealed as a rationality that encourages individuals to demand an increasing scope for self determination and autonomy – that is, self regulation – in areas which had hitherto been the domain of state agencies specifically responsible for regulation. Yet, insofar as the state continues to have an interest in the outcomes of this market like regulation (e.g., its citizens' health, education, socioeconomic well being, productivity, innovativeness, etc.), the neoliberal development marks not the end but a transformation in societal regulation and a fundamental change in the relation between state and civil society actors.

Second, studies in governmentality have worked to show that neoliberal political economy operates on multiple political scales, demanding responsiveness not only from the individual body but also from various collective bodies and institutions (e.g., universities, cor porations, non governmental organizations, and public institutions). All these bodies may be called upon to be in shape, flexible, productive, and self organizing by a ra tionality that conceals its multiscalar power geometries. Critical of the idea of scale as a fixed arena of politics, the governmentality approach facilitates the analysis of micro level and macro level politics as implicated in each other through the internalization of a particular mentality of rule (such as neoliberalism) by individuals and collectives alike. The construction of new domains of action in society, including codes of conduct and new roles and identities, often cuts across various scales of action. For instance, new forms of regulation in health care may pass directly from the national medical regu lation into the practices of clinical work.

A central avenue for the emerging interlinkages be tween different scales of bodily self regulation is the apprehension of particular knowledge as true. Studies in governmentality purport to reveal the politics of truth that underlie neoliberal rationalities, such as the need to be attractive in order to succeed in competition, be it over a good job or foreign direct investments in a city. Moreover, the concept of governmentality has the cap acity of highlighting the intimate relationship between representations and agency, thought and action, both in the governmental and nongovernmental realms of social life. This enables a sharper analysis of the effects of neoliberal government not simply in terms of the re production of existing social asymmetries, but also through the invention of new social mechanisms of domination on the basis of emerging political rational ities and forms of society.

Third, the analysis of governmentality has dis tinguished subjectification as a technology that tends to conceal the power of self regulation from the autono mous individuals themselves. Rather than government simply dominating individuals, in neoliberal govern mentality subjects actively participate in their own sub jectification. This is primarily because the technologies of self – that subjects, as autonomous members of liberal democratic societies, are called upon to practice – are imbued with governmental aspirations to shape indi viduals' conduct so as to produce desired effects. Hence, there is a double bind between subjection (the relations of governmental power) and subjectification (the con struction of the individual subject), a bind that operates through the naturalization of certain forms of identity and agency, such as the consumer, manager, investor, customer, or entrepreneur, as a fulfillment of inherent qualities of the human nature and society.

Foucault addressed this duality inherent in the con struction of subjects using the French term *assujettissement* (translated in English variously as subjectification, sub jection, or subjugation). In liberal democratic societies individual and collective subjectivity is expanded, in corporated, and organized as an integral part of govern ment, so that subjection and subjectification are layered upon each other in complex ways. The institution and proliferation of self reflexive subjects with the capacity to act channels autonomous human agency into a resource for, rather than an obstacle to, governing. In a neoliberal society the reflexive subject emerges as a strategy in the promotion of freedom, understood in terms of the capability of an autonomous individual to establish an identity. Freedom is seen as the capacity to realize one's desires, to fulfill one's potential, and to shape a mean ingful everyday life through acts of choice. This entails becoming a subject whose freedom is a condition of subjection, and at the same time subjection, a condition of freedom. Hence, subjectification and subjection inter twine and condition each other. In a sense this intimate relationship culminates in the discourses and practices of personal self assessment. As an awareness of how to evaluate and act upon ourselves to remedy various per sonal and social problems, self esteem is a technology of self par excellence.

Governmentality in Human Geography

Interest in the concept of governmentality within human geography research has grown steadily since the mid 1990s. The concept started to gain ground first in the context of political geography, where it was used within wider debates about the role of knowledge production in the constitution of the geopolitical world. While critical of the conventional understanding of the state as the single most important locus and scale of politics, much of this research took the state government as its starting point. The critical analysis of geopolitics has shown that the practices of statecraft depend on and produce particular depictions, categorizations, analyses, and hierarchies, portraying the sociospatial world as a patchwork of naturalized and mutually exclusive terri torial entities: states. The knowledge thus produced in forms practices of statecraft and foreign policymaking, which have been subjected to much critical scrutiny by workers in the area.

The concept of governmentality has figured into this work in at least two important senses. First, it has shifted the emphasis from the 'what' to the 'how' in the analysis of states' geopolitical practices. Consequently, there is a growing understanding of the ways in which the states not only represent and intervene in domains they seek to govern, but also constitute the geopolitical reality as governable, thus fashioning the conditions for any possibility of knowing the world more generally as well. Despite the emphasis on state practices, attention has also been paid to the production of geopolitical world descriptions by media and other nonstate actors, examining their role in the constitution of popular geopolitics. This analysis has helped in rethinking the scales of geopolitics, and how the mentalities of geopolitical rule are constituted, disseminated, and internalized through popular cultural products such as movies, newspapers, cartoons, and internet websites.

Second, the concept of governmentality has helped political geography analysis to broaden its conception of the organization and politics of space. Foucault inspired studies have brought into focus new forms or aspects of spatiality, such as mobility, fluidity, distance shrinking networks, and the circulatory spaces of security. By analyzing processes such as the geography of statistic production or the ordering of space through cartographic techniques or urban house numbering, this work has shown that governmentality is a technology of power that is used to govern contingency and chance inherent in the complex relationships between population and territory. Political geography's traditional concern with territoriality has not been pushed aside by the emphasis on governmentality; the research rather has purported to show that only by means of a detailed and consistent knowledge of the territory and everything in it has the state been able to develop systems of government that can be characterized by the term governmentality. Security, in this context, means enabling normalization and requires that the state is capable of opening up and liberating the use of space in the name of the circulation of goods, people, and wealth, rather than restricting and enclosing territory. Although circulation and passage require some regulation, in liberal society this should be minimal and largely dependent on self regulation on the level of autonomous individuals and bodies.

Besides political geography, the concept of governmentality has surfaced in human geographic research in areas such as the regulation of biological risks and disorders in biological life through biotechnology, and the social production of nature through the intersection of scientific, governmental, and capitalist production. New areas of research also include urban government and planning practices. Specific attention has been paid to issues such as the role of the institutional spaces of statistics and strategic interdependencies in the construction of relations between local and national government.

Growing interest has also been directed to the spatialities involved in managing large scale sociospatial processes (e.g., traffic systems or the public health). This work has shown that practices of governmentality may change in pace with the introduction of new information technologies. For example, governmentality based on information capture integrates data gathering mechanisms with systems that they seek to monitor and regulate, and thus actively influence the experience of practitioners involved. Automated management has the capacity to produce sociospatial processes and formations that stretch across various scales from the micro setting of components on to operative environments, broader technological infrastructure, systemic assemblage, and so on. This research has also charted the ways in which expertise inextricably linked to governmentality gets institutionalized, produced and embodied, and moreover, what spatial configurations the processes involve. Attention has been paid especially to the spaces in and through which health is problematized either in terms of knowledge production, education, therapy, or treatment. The governmentality framework has led to consider, for example, of how therapeutic sites shape individuals' health related conduct by turning them into governable subjects.

Critiques

The growing interest in governmentality studies shows that the analytical potential of the concept has not been exhausted. However, arising from heterogeneous works by Michel Foucault and subsequent scholarship, the relevance of the concept of governmentality has been challenged in the social sciences and human geography. Among the key issues of debate is the historical sequence by which governmentality emerged as part of the democratic liberalization of Western societies. While Foucault contends that governmentality is a contingent phenomenon that belongs to Western modernity, authors such as Giorgio Agamben see it rather as an ontological grounding of modernity, which, thus, is much more deeply implicated in every aspect of the origin and development of Western political rationality as such.

A related issue of contention is the balance in Foucault's thought between territory and population as the nexus of modern governmental rationality. While some authors accept that population rather than territory is the driving force in the governmentalization of the state, others argue that this represents a misleading ontological distinction between territory and population.

Yet another question that has provoked criticism is the difficulty that studies of governmentality have in accounting for resistance to the norms and rules of neoliberalism. It is claimed that an emphasis on technologies of self leads to the early dismissal of a proactive role that socio cultural change may play in shaping formal practices of politics, policy, and administration. Yet freedom to govern and control may also afford subjects possibilities to resist and modify those norms and rules that are generated by governmental technologies of power. Critique notwithstanding, it seems that the val idity of the concept of governmentality for understanding power relations in contemporary neoliberal societies has gained broad acceptance.

See also: Biopolitics; Body, The; Care/Caregiving; Citizenship and Governmentality, Rural; Critical Geography; Critical Geopolitics; Foucauldianism; Identity Politics; Neoliberalism; Regulation; State Theory.

Further Reading

Agamben, G. (1998). *Homo Sacer: Sovereign Power and Bare Life*. Stanford: Stanford University Press.

Dean, M. (1999). *Governmentality: Power and Rule in Modern Society*. London: Sage.

Foucault, M. (1979). *The History of Sexuality, Volume 1. An Introduction*. London: Allen Lane.

Foucault, M. (1991). Governmentality. In Burchell, G., Gordon, C. & Miller, P. (eds.) *The Foucault Effect: Studies in Governmentality*, pp 87 104. Chicago, IL: University of Chicago Press.

Foucault, M. (2007). *Security, Territory, Population*. London: Palgrave Macmillan.

Hakli, J. (1998). Discourse in the production of political space: Decolonizing the symbolism of provinces in Finland. *Political Geography* 18, 331 363.

Hannah, M. (2000). *Governmentality and the Mastery of Territory in Nineteenth Century America*. New York: Cambridge University Press.

Jessop, B. (2007). From micro powers to governmentality: Foucault's work on statehood, state formation, statecraft and state power. *Political Geography* 26, 34 40.

Lemke, T. (2001). The birth of bio politics. Michel Foucault's lecture at the Collège de France on neo liberal governmentality. *Economy and Society* 30, 190 207.

Ó Tuathail, G. (1996). *Critical Geopolitics: The Politics of Writing Global Space*. London: Routledge.

Rose, N. (1999). *Powers of Freedom: Reframing Political Thought*. Cambridge: Cambridge University Press.

Sinnerbrink, R. (2005). From Machenschaft to biopolitics: A genealogical critique of biopower. *Critical Horizons* 6, 239 265.

Green Revolution

J. Briggs, University of Glasgow, Glasgow, UK

Glossary

CGIAR Consultative Group on International Agricultural Research. This influential group consists of 15 member research centers spread across Asia, Africa, Europe, and the Americas.

CIMMYT Centro Internacional de Mejoramiento de Maiz y Trigo (the International Center for the Improvement of Maize and Wheat) in Mexico, where the pioneering work by Nobel Prize winner, Norman Borlaug, was undertaken on developing high-yielding, dwarf varieties of wheat.

Genetically Modified Crops GM crops are currently being developed mainly by large corporations using various biotechnology methods. Whereas traditional cross-breeding methods involve different varieties within the same crop species, GM crops can involve genetic engineering between entirely different and unrelated species. An example of this is the use of fish genes in breeding new strains of tomatoes.

Green Revolution The name given to the dramatic increases in crop yields, especially of wheat and rice, that occurred in the 1960s, based on the work led by Norman Borlaug at the Centro Internacional de Mejoramiento de Maiz y Trigo in Mexico.

High-Yielding Varieties (HYVs) The HYVs of wheat and rice formed the basis of the Green Revolution. These were essentially dwarf varieties which were very receptive to the application of both a regulated water supply, usually by irrigation, and fertilizers, especially nitrogen.

Malthusian Specter Thomas Malthus, writing in 1798, suggested that whereas population numbers tended to grow geometrically, agricultural output grew only arithmetically. Hence, there comes a point when agricultural production can no longer supply the population with enough food, and starvation and famine follow. This is the Malthusian specter.

Introduction

The Green Revolution is the term applied to the major advances in crop breeding genetics made in the 1960s which significantly raised the yields of some grain crops. Although crop experimentation is something that takes place continually, and has done so since the earliest settled agriculture about 8000 years ago, there are nonetheless step like advances made from time to time in producing new seed varieties, and the Green Revolution can be seen as one of these. However, rather than the full range of crops experiencing rises in yields per hectare, it was only really wheat in particular, and rice to a lesser extent, which experienced the dramatic yield rises characterizing the Green Revolution (**Tables 1** and **2**). Between 1961 and 1971, the advances of the Green Revolution can be clearly seen, as, in that decade, yields per hectare of wheat increased by about two thirds, and those of rice by about one third. In all the selected regions shown in **Tables 1** and **2**, it can be seen that yields of both wheat and rice have continued to rise in each subsequent decade, including in Africa, such that between 1961, before the benefits of the Green Revolution, and 2005, the last year for which figures were available at the time of writing, wheat yields per hectare have increased by over 200% in Asia, Africa, and India, although the overall increase in rice yields over the same time period has been rather more modest.

It can be argued that the term 'Green Revolution' was very much a product of its time. Although the term can be seen as referring to the greenness of crops, it can also be interpreted as being a Green Revolution as opposed to a Red Revolution. During the 1950s and 1960s, at the height of the Cold War, socialism envisaged progress being made by changing the relations of production primarily through the communal or state ownership of the means of production. Capitalism, as championed by the United States, saw progress as being achieved through the transfer of science and technology. In this case, traditional agriculture would be transformed by the adoption of a new, imported technology, forming the basis of the Green Revolution. The political consequences of the Green Revolution were seen, particularly by the United States, to be very important as a solution to food shortages and famine in Asia in particular, and therefore a bulwark to the spread of socialism in that continent at the height of the Cold War.

The research and development of the new high yielding varieties (HYVs) of wheat were undertaken at Centro Internacional de Mejoramiento de Maiz y Trigo – the International Center for the Improvement of Maize and Wheat (CIMMYT) in Mexico, and that of rice HYVs at the International Rice Research Institute (IRRI) in the Philippines, the latter using plant materials originally from Taiwan and Japan. Both these centers were members of the Consultative Group for International Agricultural Research (CGIAR), a network of 15 research centers across the world, each with its own major

Table 1 Yields of wheat by selected regions (tonnes per hectare)

	India	*Asia*	*Africa*	*EU*	*N. America*	*World*
1961	0.85	0.74	0.69	1.96	1.33	1.09
1971	1.31	1.25	0.98	3.02	2.15	1.64
1981	1.63	1.77	1.13	3.80	2.23	1.89
1991	2.28	2.44	1.90	5.41	2.29	2.47
2001	2.71	2.65	2.03	5.52	2.44	2.78
2005	2.74	2.82	2.11	5.74	2.75	2.94
Growth 1961 2005	222%	281%	206%	193%	107%	170%

Source: FAO (http://economics.cimmyt.org)

Table 2 Yields of rice by selected regions (tonnes per hectare)

	India	*Asia*	*Africa*	*EU*	*N. America*	*World*
1961	1.54	1.86	1.48	5.34	3.82	1.87
1971	1.71	2.38	1.79	4.98	5.29	2.36
1981	1.96	2.88	1.69	5.47	5.40	2.82
1991	2.63	3.61	2.07	5.93	6.42	3.54
2001	3.12	4.00	2.19	6.40	7.28	3.94
2005	3.00	4.12	2.04	6.44	7.40	4.02
Growth 1961 2005	95%	122%	38%	21%	94%	115%

Source: FAO (http://economics.cimmyt.org)

research responsibilities and priorities in the field of agricultural development. The pioneering work on wheat was undertaken by Norman Borlaug who subsequently received a Nobel Prize in 1970 for this work. The basis of the work was to raise agricultural productivity through technical progress, but without any meaningful economic and/or social reform of the agricultural sector. The challenge of development, therefore, was seen as one of technology, and Western technology at that, and not one of socioeconomic reform. The assumption was made that the new technology would be scale neutral, that is, the new varieties would be just as viable on small farms as on large ones. Through a process of hybridization and development, the so called 'miracle' seed, IR8, was developed. This was a dwarf variety of wheat which was especially responsive to a combination of a carefully controlled water supply and fertilizer application, par ticularly of nitrogen. The dwarf varieties had some clear advantages over traditional varieties. There was a greater proportion of grain head to stalk; there was greater strength in the shorter stalk so that there was less bending or snapping of the stalk caused by a heavy grain head; and the growing season was shorter which created the possibility of growing two crops per year, this double cropping therefore raising cropping intensity. This, of course, reinforced the need for irrigation to underpin the adoption of Green Revolution HYVs.

Impacts in India

Through the 1950s and 1960s, India had been increasingly facing food shortages, and even the Malthusian specter of famine. That the Indian Government should embrace the new technologies of the Green Revolution, and should promote the new 'miracle' seeds, was therefore of little surprise. Indeed, the government promoted its own re search stations to continue the improvement and devel opment of grain crop yields, but with a greater sensitivity to local conditions. **Table 3** shows that the total production of wheat in India, as a consequence of the Green Revolution, was much more dramatic than that of rice. Since 1961, the total production of wheat has grown by 555% and that of rice by a more modest 141%. Nonetheless, the Green Revolution dealt with the food shortage threats of the 1950s and 1960s in gross terms.

Table 3 Total wheat and rice production in India, 1961 2005 (000 tonnes)

	Wheat	*Rice*
1961	10 997	53 495
1971	23 833	64 602
1981	36 313	79 883
1991	55 135	112 042
2001	69 681	139 900
2005	72 000	129 000
Growth 1961 2005	555%	141%

Source: FAO (http://economics.cimmyt.org)

However, the geographic impacts of the Green Revolution have been far from even. Three states in particular – Punjab, Haryana, and Uttar Pradesh in northern India – have benefited considerably due to the relatively ready availability of irrigation, especially from tube wells, and, arguably, the more positive attitudes of farmers in these states in dealing with risk and uncertainty in agriculture. Wheat HYVs now account for about 50% of all the production in these three northern states and nearly all of the government's grain purchases. The gains in the rice growing river and delta areas of India have been far less impressive. This not only reflects the fact that the improved yields of the new rice seed varieties were significantly lower than those for wheat in the first place, but also that there were greater water management problems and more serious pest problems affecting the new seed varieties of rice in these areas. Ironically, since the mid 1970s, rice has become an important second crop in the three big wheat area states of Punjab, Haryana, and Uttar Pradesh. Perhaps even more ironically, the rice from these areas is produced primarily as a commercial crop for the market, with about 90% of production being sold, compared to only about 30–40% from the traditional rice growing areas of India. There is also evidence to suggest that, with the increased total production levels of both wheat and rice through the use of HYVs, there has been a reduced rate of growth of grain prices for consumers.

An Evaluation of the Green Revolution

There has been a considerable amount of debate over the success or otherwise of the Green Revolution. Whereas it cannot be denied that the adoption of HYVs staved off the threat of famine in India, and that the continued rise of yields per hectare through until the present day has reduced (but not always eliminated) food insecurity in many other parts of the world, there have been a number of environmental, economic, and social consequences of the Green Revolution, some of which have been less than positive. The greater use of irrigation water has frequently resulted in the general lowering of the water table in many areas, while, ironically, leading to greater salinization in those areas where water tables have risen locally or where excess water has not been successfully drained off the land. Some studies have shown that this has disproportionately affected poorer people because of their greater reliance on shallow surface wells, whereas richer farmers generally have the capital to construct deeper tube wells which offer some insurance against these problems. The new HYV seeds respond very well to the combination of controlled water and fertilizer use. However, if there are any problems with water application, the use of inorganic fertilizer, especially nitrogen, can have a deleterious effect on the soil as it can result in 'burning' of the soil. Traditionally, farmers used a range of seed varieties and often intercropped grain crops with other crops. However, the use of the new HYV seed varieties has resulted in a loss of biodiversity in many of the rural areas, as monocropping of the same HYV variety has become the norm. This, in turn, has impacted on pest control, as these traditional rotations and intercropping helped to control pest infestation, which has now subsequently increased in many areas. Consequently, farmers have had to purchase greater amounts of pesticides and fungicides to guard against crop losses. This has meant not only an increase in input costs for farmers, but has also resulted in an increase in toxic pesticides being left behind as a residual in the environment.

The reduction in intercropping has had other consequences. For example, there has been an impact on household diets. In many areas, it was common in the past to intercrop grain with pulses. With pure stands of HYVs, this now rarely happens, and hence there has been an increased dietary imbalance for many rural people. Intercropping was also used as an insurance policy against the failure of one (or more) crops as a result of drought, flooding, or waterlogging, but with HYVs this insurance has disappeared. Significantly, if irrigation water (or rain for that matter) is well timed, dwarf wheat does well. However, yields are very sensitive to both drought and waterlogging; without regular water application, diminishing returns of HYVs can set in quite quickly, such as has happened in parts of Tamil Nadu. Ironically, the important advantage of HYV dwarf varieties having a larger grain head in relation to stalk length has had an unfortunate consequence for those farmers who also keep livestock. Stalks have long been used as a source of animal feed and the amount of animal feed material from the dwarf HYVs is much less than from traditional wheat varieties. To make up this shortfall, many farmers now have to buy in more feed for livestock than they did previously, so adding costs to the household budget.

A contested area of debate has revolved around the economic and social consequences of the Green Revolution. Although there is evidence to suggest that neither farm size nor tenure arrangements have had a significant impact in constraining the adoption of HYVs in the first place – that is, HYVs do indeed appear to be scale neutral – there is strong evidence to support the view that the subsequent gains from cultivating HYVs are much greater for those larger, more capital intensive farmers with larger farm units. In other words, there has developed a disparity in income distribution related to farm size and consequently the Green Revolution has increased rural inequalities. This is hardly surprising, as larger farmers, with easier access to capital resources, are in a stronger position to invest in better irrigation facilities, more reliable supplies of fertilizers, pesticides, and fungicides, to invest in machinery to displace labor, and generally to benefit from economies of scale in production. Indeed, it would seem that many of the larger, more capital intensive farmers in the 1960s and 1970s took advantage of subsidized credit and the overvalued exchange rate in India at that time to import machinery from overseas. Interestingly, it has been suggested that many of the investments in machinery have an important gender dimension, as much of the invest ment in labor saving/displacing machinery has been in male labor areas such as land preparation and harvesting. Rather less has been invested in traditional areas of women's labor input such as weeding.

As there was a link between farm size and likely monetary gains and increased margins, it was clearly in the interests of landowners to take over cultivation of the land themselves and employ labor as needed. Hence, a con sequence of the Green Revolution has been that many share croppers and tenant farmers have lost their access to land, with a subsequent increase in landlessness and un employment. For some other farmers, who owned their own land, but had only small farms of less than a hectare in many cases, it proved to be very difficult to achieve the full production gains of the Green Revolution, and, for these, the sale of their land to bigger farmers proved to be irresistible. However, this only led to further increases in landlessness. Between 1971 and 1981 in the Punjab, the number of smallholdings below 2 hectares fell and land concentration increased. There is, however, some evidence to suggest that, despite the increases in landlessness and poverty levels, as parts of the countryside, especially in the Punjab, grew wealthier as a consequence of the Green Revolution, rural employment opportunities also grew, particularly outside the agricultural sector.

The Green Revolution and Africa

Although the Green Revolution has clearly been effective in changing the food production fortunes of India and elsewhere, it has had only a limited impact on Africa. Indeed, since 1990, the number of malnourished people in Africa has increased by 20% and food production per capita since 1981 has either been stagnant, or has even declined in the cases of some African countries. Most of the productive agricultural land is already being farmed and so Africa cannot rely on converting new land to agriculture as a solution. Hence, it is argued by many that an agricultural transformation, involving the development and adoption of new seed varieties, is urgently needed.

The reasons for Africa so far missing out on the gains of the Green Revolution are clear. First, the main staple crops in Africa are not those which have seen the greatest yield gains from the Green Revolution, that is, wheat and rice. **Table 4** shows that, although wheat and rice are grown in Africa, they are relatively minor crops, in terms of area, in comparison to maize, sorghum, and millet, and, to a lesser extent, cassava. Whereas Africa grows less than 5% of the world's wheat and 6% of the world's rice by area, it grows well over half the world's sorghum, millet, and cassava, and a good proportion of the world's maize. However, the yield gains made by these four African staples since 1961 have been relatively modest

Table 4 Areas harvested of selected grain crops in 2005 (000 hectares)

	Africa	*World*	*Africa as % of world*
Wheat	9645	211 862	4.6
Rice	9130	153 546	6.0
Maize	27 890	146 741	19.0
Sorghum	26 013	44 261	58.8
Millet	20 490	35 687	57.4
Cassava	12 345	18 689	66.1

Source: FAO (http://economics.cimmyt.org)

(**Table 5**), and, for three out of the four crops, the yield growth in Africa has been below that for the world as a whole. Hence, not only have African staples not shared the yield improvements of wheat and rice, where yield improvements have been made, they have generally been very modest. A significant part of the problem is that African agriculture is overwhelmingly rain fed, rather than irrigated, and so it is very dependent on the vagaries of the annual climate. In addition, one of the key drivers in the success of the dwarf wheat and rice varieties is the controlled application of irrigation water, a factor largely absent from African agriculture as a whole. Indeed, this issue of an inadequate infrastructure relates to more than just access to irrigation, important as this is. The sheer physical size of Africa, the inadequacy and unreliability of the transport system, poor but improving information channels, inadequate credit facilities, and problems of seed and fertilizer development, manufacture, and dis tribution all currently militate against a potentially suc cessful Green Revolution in that continent.

Genetically Modified Crops

As we have seen, although genetic advances in seed breeding and development frequently occur in step like progressions, there is also generally a steady rise of yields over time between these steps (see **Tables 1** and **2**). Al though CGIAR was the dominant force in crop genetics in the 1960s and 1970s, and CGIAR very much drove the Green Revolution, its relative influence over the last couple of decades has declined, to be replaced by large, and in some cases very large, corporations such as Mon santo. The attraction for these corporations is the poten tially large margins and profits that can be made. New crop breeding programs are being promoted, some using well established methods of cross breeding different var ieties within species, but others using new biotechnologies and genetic engineering which introduce genes across species barriers. This latter has become a major concern more widely, as the long term consequences of such interspecies cross breeding are unknown.

However, arguments are made that contemporary bio technologies can nonetheless be very beneficial. A common

Table 5 Yields of African staples (tonnes per hectare)

	Maize		*Cassava*		*Sorghum*		*Millet*	
	Africa	*World*	*Africa*	*World*	*Africa*	*World*	*Africa*	*World*
1961	1.05	1.95	5.66	7.40	0.82	0.89	0.58	0.59
1971	1.35	2.66	6.00	8.36	0.79	1.24	0.57	0.68
1981	1.86	3.50	7.06	9.26	1.00	1.61	0.73	0.72
1991	1.48	3.70	8.08	9.84	0.82	1.31	0.66	0.68
2001	1.59	4.43	9.00	10.75	0.87	1.37	0.65	0.79
2005	1.70	4.73	8.88	10.88	0.86	1.32	0.69	0.77
Growth 1961 2005	62%	143%	57%	47%	5%	48%	19%	31%

Source: FAO (http://economics.cimmyt.org)

example quoted relates to the development of a cyanide free cassava which would eliminate the need for its careful, time consuming preparation before cooking. As this task is usually undertaken by women, this would reduce this part of women's work burdens, as well as using less precious water, further reducing women's burdens, as water col lection in most parts of Africa is also women's work. However, it has to be recognized that the corporations' targets are generally not small scale farmers in low income countries, but rather capital intensive farmers in the West, where significant profits are more likely to be made.

Conclusions

There is little doubt that the Green Revolution had a profound effect on the production of wheat and rice during the 1960s and 1970s, and especially so in India. It was, however, a very technocratic solution to the issue of agricultural transformation, and it left unaddressed issues related to the economic and social conditions of pro duction. As such, the Green Revolution was very much a product of its time, when it was thought that the problems of the so called Third World could be addressed by the transfer of science and technology, usually quite uncritic ally, from the West. Today, the question of development in the low income countries of the South is seen as a much more nuanced and complex set of related issues, of which Western science and technology make up simply one part of the jigsaw, rather than being the full jigsaw itself.

There is also an irony that we can see today in the use of the term 'green' that was not apparent in the 1960s and 1970s. In the early part of the twenty first century, the use of the term 'green' implied sustainability, a strong environmental empathy, and the use of sympathetic and locally appropriate technologies. In this sense, the Green Revolution appears to be anything but 'green' in the sense of the word today, given that it is highly techno logical and depends for its success on expensive inputs, including inorganic fertilizers and pesticides. Because of this, some have called it a 'high external input agri culture', and thus not one which would easily meet the requirements of the green movement today.

See also: Agriculture, Sustainable; Development I; Famine; Modernization Theory; Socialism; Third World.

Further Reading

Bayliss Smith, T. and Wanmali, S. (eds.) (1984). *Understanding green revolutions*. Cambridge: Cambridge University Press.
Conway, G. R. and Barbier, E. B. (1990). *After the green revolution: Sustainable agriculture for development*. London: Earthscan.
Ghatak, S. (2003). *An introduction to development economics* (2nd edn.), ch. 9. London: Routledge.
Glaeser, B. (ed.) (1987). *The green revolution revisited: Critique and alternatives*. London: Allen and Unwin.
International Food Policy Research Institute. http://www.ifpri.org/pubs/ib/ib11.pdf. (accessed on 25 August 2007).
Hazell, P. and Ramasamy, C. (eds.) (1991). *The green revolution reconsidered*. Baltimore: Johns Hopkins University.
Jirstrom, M. (1996). *In the wake of the green revolution*. Lund: Lund University Press.
Lipton, M. and Longhurst, R. (1989). *New seeds and poor people*. London: Unwin Hyman.
Perkins, J. (1997). *Geopolitics and the green revolution*. Oxford: Oxford University Press.
Rigg, B. (1989). *The green revolution*. Cambridge: Cambridge University Press.
Scoones, I. (2002). *Agricultural biotechnology and food security: Exploring the debate*. Working Paper 145. Brighton: Institute of Development Studies, University of Sussex.
Seshia, S. and Scoones, I. (2003). *Tracing policy connections: The politics of knowledge in the Green Revolution and biotechnology eras in India*. Working Paper 188. Brighton: Institute of Development Studies, University of Sussex.

Relevant Websites

http://www.actionbioscience.org
A very useful source of statistical information relating to agricultural production data.
http://impact.cgiar.org
CGIAR Impact, Consultative Group on International Agricultural Research.
http://www.cgiar.org
Consultative Group on International Agricultural Research.
http://www.fao.org
Food and Agriculture Organization of the United Nations.
http://www.foodfirst.org
Food First, Institute for Food and Development Policy.
http://www.monsanto.com
Monsanto is an Agricultural Company.
http://www.actionbioscience.org
Promoting Bioscience Literacy.

Greenfield Development

P. Breathnach, National University of Ireland, Maynooth, Republic of Ireland

Glossary

Agglomeration Economies Savings made jointly by firms through clustering together in particular localities.

Foreign Direct Investment (FDI) Investment in operations abroad by firms which confers a degree of direct control over the foreign operations in question by the investing firms.

Greenfield Development A business investment by a firm which involves setting up an operation in a new location.

Industrial District/Cluster A localized concentration of industrial activities characterized by a high level of interaction between the firms carrying out these activities.

Multiplier Effects The stimulation of additional economic activity in a region arising from the establishment of a new business in that region.

Polarized Growth Tendency for economic growth to be spatially concentrated in growth poles/centers.

Spatial Division of Labor (Intrafirm) Location of different segments of a firm's production chain in different places/regions.

Technology Transfer The transfer of technological and/or organizational know-how between regions/ countries.

Introduction: Defining Greenfield Development

The term greenfield development originally referred to economic activities (usually manufacturing plants) which were established on sites which had not previously been developed (except for agricultural purposes). Such sites contrasted with so called brownfield sites which were, or had been, previously developed for industrial purposes, particularly involving old generation industrial activities, such as coal mining, heavy industry, and dockland activities. As the term itself suggests, greenfield development in this context generally referred to establishing economic activities either in rural areas or, more typically, on the outskirts of towns and cities.

Over time, the use of the term was broadened to encompass any business investment by a firm which involves setting up an operation in a new location (whether urban or rural) rather than, for example, expanding or acquiring existing operations. This new activity may involve a net addition to the firm's productive capacity, or simply a relocation of an existing activity from another site, or sometimes a rationalization of a number of existing activities at different sites. These different possibilities carry with them very different implications in terms of their local, regional, and national economic development impacts.

Analysis of patterns of greenfield development bring together two important areas of concern in economic geography – industrial location and regional development. Geographers have long been interested in analyzing patterns of industrial location – including the establishment of new industrial plants on greenfield sites – and the factors which influence how these patterns are shaped. Interregional variations in forms and levels of economic development and the processes producing these variations have also been a major focus of geographical enquiry. The stimulation of greenfield investments in manufacturing (and, increasingly, services) activities in less developed regions has been an important feature of public policy in most advanced economies, and geographers have made an important contribution in terms of examining the nature, location, and economic impact of such investments.

Factors Influencing Greenfield Investment

Firms take a range of considerations into account in deciding whether to remain in and/or expand existing industrial plants or set up new plants in greenfield locations. Factors which may favor expanding production at existing locations include the prospects of obtaining economies of scale; the fact that it is usually cheaper to add capacity to an existing plant rather than build a new one from scratch; and the simpler management structures involved in keeping production under one roof.

The factors favoring establishment of new plants comprise a combination of the push and pull factors normally associated with migration behavior, the former prompting firms to move from existing sites and the latter attracting firms to new sites. Firms may be stimulated to relocate existing production where existing plants are characterized by high costs, labor relations problems, and obsolescent technology. Alternatively, firms may choose to install additional production capacity at new sites where existing plants are constrained by lack of space, or where plant enlargement may create production control difficulties, or problems in maintaining effective communications with an enlarged workforce. Firms may also be attracted to new locations in order to exploit local market opportunities, to reduce operating costs, or to avail of favorable financial incentives, although set against these will be the costs of identifying appropriate locations for new investments and of operating in unfamiliar environments. The establishment of new branch plants in greenfield locations, accordingly, tends to be mainly confined to larger firms with appropriate resources and experience, especially where the new plants are outside the home region of the parent firms.

Greenfield Development and Regional Policy

Much of the early impetus behind greenfield development emanated from the desire of industrial firms to move out of the inner city locations where most manufacturing industry established in the nineteenth and early twentieth centuries had been concentrated. Such relocation became increasingly common from the 1950s onward and was prompted by a variety of mutually supporting considerations:

- the growth of automobile transport meant that city centers were becoming increasingly congested and polluted;
- space for expansion was at a premium, especially in the light of changing technologies which favored single story over more traditional multistory industrial buildings;
- much of the traditional workforce had been removed from inner city areas to new suburban housing projects;
- alternatively, new technologies created a desire to access new sources of labor; and
- substantial capital gains could be realized by selling off valuable central city sites and moving elsewhere.

Much of this relocation simply involved a move to nearby suburban areas where adequate supplies of land and labor were available and which were much more accessible to the motorized transport which was increasingly being used by industrial firms for deliveries of both inputs and outputs. However, in many advanced economies, these relocations involved interregional movements, particularly to declining industrial regions or peripheral rural regions where cheap labor and financial and other incentives were available. In Britain, by 1966, 10% of the manufacturing workforce were employed in plants which had moved from one region to another, or in entirely new plants owned by firms headquartered outside the regions in which the plants were located. In France, between 1954 and 1982, industrial employment grew by 368 000 in the underdeveloped west and southwest regions, while falling by 275 000 in the Paris region. These forms of industrial movement were one element of the new spatial divisions of labor which were being created within large manufacturing firms, whereby different segments of these firms' production chains came to be located in different places (depending on such factors as access to resources and level of production complexity), with administrative and research and development functions being located at different places again.

Relocations of existing manufacturing activities were particularly common in the clothing industry which, facing increasing competitive pressures from new technologies (e.g., synthetic fibers) and new production locations overseas, sought to cut costs by moving production to cheap labor sources. Thus, there was a spectacular movement of the US clothing industry from its original base in New England to the southern states; while in the UK, there was also a significant movement of the cotton and woollen textile industries from their original bases in Lancashire and West Yorkshire to peripheral rural areas.

This interregional movement of branch plants was not confined to declining traditional sectors. Many of the new consumer industries of the twentieth century (such as electrical and electronic products) also from an early stage began to set up production plants – mainly in final assembly – in areas where cheap (and mainly female) labor was available, particularly where this was accompanied by government location incentives, such as financial grants, tax concessions, labor subsidies, and the availability of serviced sites. Indeed, the attraction of branch plants through such incentives became the standard formula for the new regional development policies which were commonly introduced in advanced economies during the recessionary 1930s and then extended in the decades after World War II.

Thus, in Ireland, substantial grants were introduced in 1952 to encourage the establishment of manufacturing plants in the underdeveloped western part of the country; while in Italy, substantial funding was made available in the 1960s to encourage industrial firms based in the prosperous northwest to set up branch plants in the very poor southern part of the country.

Virtually from the beginning, therefore, regional development policies in advanced economies became strongly associated with the attraction of inward branch plant investment to greenfield sites within lagging regions. The logic involved in these policies was simple (if not, indeed, simplistic): manufacturing industry was seen as the essential driver of economic development in advanced industrial economies; in regions where traditional industries were in decline, the formula was to replace these industries with modern industries which were emerging elsewhere. In peripheral rural regions which had no previous experience of industrialization, this was seen as evidence of a lack of indigenous natural or entrepreneurial resources, in which case the introduction of manufacturing employment from outside the region was seen as the most direct route to raising average living standards to the levels prevailing elsewhere.

Transnational Corporations and Greenfield Development

Up to the 1960s, regional development policy was largely focused on inducing interregional industrial movement within national economies. With the rapid growth in foreign direct investment (FDI) by transnational corporations (TNCs) from that decade on, regional policy became increasingly focused on attracting the branch plants being set up by these corporations in various parts of the world. Such investments are attractive for a number of reasons. Driven by the forces of globalization, the absolute volume of mobile FDI has expanded enormously over the last 30 years. Being inherently footloose, such investment is amenable to being attracted to locations which can provide an appropriate package of attractions. TNCs tend to be to the forefront of technological innovation, so their production plants are in little danger of suffering obsolescence and might, indeed, function as sources of technology transfer to local businesses. Also, some TNC branch plants are very substantial operations which can bring substantial employment benefits to host locations. In addition, new production and communications technologies and new organizational structures have enhanced the ability of TNC branch plants to perform effectively in noncore locations, and have widened the range of possible functions which can be moved to such locations.

An important consequence of the latter point is the growing tendency on the part of TNCs to include service activities in the portfolio of projects potentially available for location in lagging regions. These include data processing, telephone call centers, and back office administrative functions. In the aftermath of the creation of the Single European Market in 1993, many TNCs have been restructuring their European operations, one of the results of which is the rationalization of back office functions into fewer 'pan European' centers which, thanks to the proliferation of new information and communications technologies, have in many cases been set up in peripheral regions where cheap labor and investment incentives are available.

Locational Patterns of Greenfield Investments

The locational patterns associated with greenfield investments in lagging regions have been quite diverse. When regional policy first became a major issue in the advanced economies in the 1960s, it was strongly influenced by theories of polarized growth which were prevalent at that time, and which argued for the concentration of new developments in what was hoped would become self sustaining growth poles or centers. However, political problems quickly fell in the way of attempts to implement such policies. The main problem was related to the idea that, in order to achieve sustainability, growth centers would have to reach a certain minimum size in order to generate substantial agglomeration economies. This in turn implied that the number of centers selected would have to be small. This predictably generated a negative reaction from the majority of urban centers in lagging regions which were not designated as growth centers under this policy. Thus, in southern Italy, a growth center policy introduced in the 1960s became increasingly diluted under local political pressure until eventually so many towns had been designated for development that the policy became meaningless.

In Ireland, by contrast, a proposed growth center policy was rejected by the government in the late 1960s (again, mainly in response to political protest) and, instead, a very ambitious program of industrial dispersal was introduced in the 1970s. Ireland's accession to the European Economic Community in 1973 was accompanied by a major acceleration of inward investment, and the government set out to distribute this investment as widely as possible throughout the country. The main methods used to achieve this objective were an extensive program of advance (ready built) factory construction combined with spatially variable grants to guide industrial projects to particular locations. This policy worked quite well for a while, and by 1980, virtually every town in Ireland of any size had acquired at least one significant industrial project, most of them foreign owned. However, the policy ran into difficulties in the recessionary 1980s when industrial contractions and closures became quite common without an adequate flow of replacement projects.

It is possible that widespread dispersal could have materialized even without the government's advance factory and variable grant measures. There was growing evidence in the 1970s of corporate branch plants actively seeking out locations in previously unindustrialized areas rather than in established urban centers. Apart from lower site and service costs, a key factor in this locational avoidance was the desire to access labor with no previous industrial experience and, in particular, no contact with trade unions. This had been made possible by the deskilling of manufacturing work (especially through the development of assembly line technology), which meant it could be performed effectively with minimal training.

Subsequently, a desire to avoid competition for workers became another important reason for firms to set up greenfield operations in small rural towns. This has been a key feature of the clusters of component suppliers which have developed around the set of Japanese auto mobile assembly plants which were established in the American Midwest in the 1980s and 1990s. These sup pliers are mainly located within 4 h driving time from their focal assembly plants, and in most cases are also located individually in small towns about 100 km apart so they do not form overlapping commuting zones.

In declining urban industrial regions with high population densities, spatial concentration of new in dustrial plants was the norm more or less from the be ginning of regional industrial policy. In former coal mining and heavy industrial areas, with their derelict and blighted landscapes, the typical approach was to provide serviced sites on industrial estates or parks located in previously undeveloped green areas, frequently adjacent to main road routes. In Britain, this was frequently done in conjunction with the many new towns built after World War II to resettle people who had been living close to the now closed or greatly reduced mines and smelters. The idea was that these towns would be self contained units, with adequate employment provided in adjacent industrial estates, so that inhabitants would not have to commute long distances to find work.

Economic Impact of Greenfield Development

In general it can be said that the economic development impacts of greenfield investments in lagging regions have been disappointing. Frequently, the benefits of such in vestments are of the static variety (involving a once off increase in output, employment, and wage payments) with few long term dynamic effects, such as technology and know how transfer, creation of local linkages, and spawning of spin off firms. There has been limited creation of local linkages as branch plants have typically sourced their in puts from outside the region, frequently from affiliate units of the same corporation. Technology transfer has been constrained by the low skill levels in branch plants, the fact that, typically, they involve only a single stage of the overall production chain, and their low use of local suppliers. Surpluses generated have been mostly transferred out of the region rather than being reinvested locally.

Even consumer multiplier effects have been constrained by the low pay levels in many branch plants, which in turn is linked to the high levels of female employment in these plants. Policies which were designed to attract replacement jobs for those lost by the menfolk in the coal mines, the steel mills, and the shipyards instead had the effect of at tracting jobs in such areas as clothing and electrical as sembly targetted at women workers whose inferior social position meant they could be paid much lower wages.

Greenfield developments which have little economic interaction with the wider regional economy in which they are located are termed enclave developments. Even where branch plants do develop linkages with local firms or institutions (which becomes more likely the longer the branch plant is located in the region), sometimes the firms or institutions (e.g., training agencies) in question can become heavily dependent on the branch plant, due to the absence of alternative clients in the region. This may lead to the creation of an extended enclave em bracing the initial branch plant and the local suppliers who are largely or entirely dependent on that branch plant. Ultimately, the entire region may become dan gerously dependent for its welfare on a single major branch plant whose long term stability will always be in doubt (a situation referred to as regional capture).

Regional Implications of New Trends in Transnational Investment

Recent changes in the nature and organization of transna tional investment carry with them varying implications for the prospects of TNC branch plants developing increasing linkages with the regional economies in which they are located. There has been a growing trend for many TNCs to confer greater functional autonomy on subsidiaries, in order to enhance their flexibility and learning capabilities. In theory, this should allow subsidiaries to make greater use of local suppliers and other local resources such as research institutions. However, the fact is that many peripheral re gions lack the technological capability to support the so phisticated supplier or research base which TNC subsidiaries require, and the general view is that this type of subsidiary development is only likely to occur where sub sidiaries are located in already advanced regions. Indeed, it has been suggested that low tech subsidiaries may be closed down altogether (especially in the light of increasing out sourcing by TNCs of routine forms of production) or re located to central regions for the precise reason of enhancing their learning capabilities.

A second major recent trend as regards transnational investment has particularly serious implications for those regions which seen, in inward greenfield investment, their route to a prosperous future. This refers to the fact that the great acceleration in global FDI, which has occurred since the early 1990s, has mainly involved mergers and acquisitions (M&As) rather than establishment of new activities in greenfield locations. M&As rarely entail a net addition to production and employment and, in fact, are more likely to lead to job losses as they normally involve some degree of rationalization of the operations which have been merged or acquired. Furthermore, M&As are more likely to involve activities which are already located or headquartered in central regions.

Even when the focus is confined to new investments in productive facilities (as opposed to mergers and takeovers of existing facilities), recent trends do not favor greenfield investments in new locations. This is because a growing proportion of such investments (up to one half in some West European countries at the end of the 1990s) consists of repeat investments in existing facilities. This reflects the fact that, after up to 50 years of transnational investment inflows, the stock of foreign subsidiaries has both built up and matured in many regions, such as Ireland, Central Scotland, and South Wales. As a result, a growing proportion of new FDI is being devoted to the replacement of equipment in established branch plants, the expansion of existing lines of production at these plants, and/or additions to their product lines or range of functions. Of particular importance in this respect is an increasing tendency toward co location of process/ product development with manufacturing production in order to facilitate quicker and more hands on feedback (previously these two functions tended to be spatially separated, with process/product development carried out in specialized research units). This trend could spell danger for less sophisticated production activities in low tech countries/regions as production may be moved to where development staff are available.

Alternative Scenarios

Despite these trends, which, for the most part, tend to favor investment in subsidiaries located in more advanced regions, international competition for inward investment continues to grow as more and more regions come forward seeking to attract mobile FDI as a means of achieving economic development. China has achieved particular success in this respect, due to both the large and rapidly growing size of the domestic market and the availability of very cheap labor for export oriented branch plants. In Central and Eastern Europe (the so called CEE coun tries), various states, as part of their transition to private enterprise following the collapse of central planning in the late 1980s, have been seeking to maximize inflows of capital from abroad. Initially, the main focus was on at tracting investment to take over existing brownfield plants (mainly inefficient state owned enterprises undergoing privatization), but in the late 1990s the emphasis switched to the attraction of new greenfield plants. The increasing intensity of competition for inward investment has meant a growing tendency for regions to outbid each other in making concessions and offers to potential investors. This has created fears of a race to the bottom as localities sacrifice tax revenues, proper planning, environmental concerns, workers' rights, and working standards to entice desperately needed projects, but at the cost of an erosion of local living standards or life quality.

The low operating costs which are the main attraction for potential inward investors offered by China and Eastern Europe are considerably offset by lower prod uctivity levels and, in any case, provide a poor founda tion for sustainable long term investment. The type of branch plant industrialization stimulated by low costs and financial incentives rarely generates the kinds of interfirm synergies upon which self sustaining growth is based. The economic success of prosperous central re gions tends to be built around complexes of inter dependent firms in particular sectors, with some of these firms involved in export markets while drawing on inputs of equipment, materials, and services from specialized local suppliers. The broad base of such industrial clusters and the inherent dynamics arising from interaction be tween exporting firms and their suppliers can provide a much more secure foundation for sustainability than the narrow base normally associated with branch plant in dustrialization, which typically involves the creation of few linkages with the remainder of the regional economy.

See also: Agglomeration; Economies, Branch Plant; Embeddedness; Foreign Direct Investment; Industrial Districts; Industrial Location; Industrial Restructuring; Regional Development Models; Spatial Division of Labor.

Further Reading

Breathnach, P. (1982). The demise of growth centre policy: The case of the Republic of Ireland. In Hudson, R. & Lewis, J. (eds.) *Regional Planning in Europe*, pp 35 56. London: Pion Ltd.
Brown, A. and Burrows, E. (1977). *Regional Economic Problems*. London: George Allen & Unwin Ltd.
Dicken, P. (2003). *Global Shift: Reshaping the Global Economic Map in the 21st Century.* London: Sage Publications.
Dunford, M. and Perrons, D. (1986). The restructuring of the post war British space economy. In Martin, R. & Rowthorn, B. (eds.) *The Geography of De Industrialisation*, pp 53 105. London: Macmillan.
Hayter, R. (1997). *The Dynamics of Industrial Location*. Chichester: Wiley.
Helsinka Hughes, E. and Hughes, M. (2003). Joining the competition: Central and Eastern European challenge to established FDI destinations? In Phelps, N. & Raines, P. (eds.) *The New Competition for Inward Investment: Companies, Institutions and Territorial Development*, pp 155 172. Cheltenham: Edward Elgar.
Keeble, D. (1977). *Industrial Movement*. Milton Keynes: Open University Press.
Mair, A., Florida, R. and Kenney, M. (1988). The new geography of automobile production: Japanese transplants in North America. *Economic Geography* 64, 352 373.
Massey, D. (1984). *Spatial Divisions of Labour*. London: Macmillan.
Parr, J. B. (1999). Growth pole strategies in regional economic planning: A retrospective view. Part 1: Origins and advocacy. *Urban Studies* 36, 1195 1215.
Parr, J. B. (1999). Growth pole strategies in regional economic planning: A retrospective view. Part 2: Implementation and outcome. *Urban Studies* 36, 1247 1268.
Phelps, N. and Fuller, C. (2000). Multinationals, intracorporate competition and regional development. *Economic Geography* 76, 224 243.
Porter, M. E. (1990). *The Competitive Advantage of Nations*. London: Macmillan.
Raines, P. (2003). Flows and territories: The new geography of competition for mobile investment in Europe. In Phelps, N. & Raines, P. (eds.) *The New Competition for Inward Investment: Companies, Institutions and Territorial Development*, pp 119 135. Cheltenham: Edward Elgar.

Gregory, D.

B. Warf, University of Kansas, Lawrence, KS, USA

Glossary

Theoretician One who engages in the production of new forms of theory, particularly social theory.

Gregory, Derek (1951)

Derek Gregory is one of contemporary human geography's leading theoreticians (**Figure 1**). Born in Kent, England in 1951, he became a Fellow at Sidney Sussex College, Cambridge, in 1969. He earned his BA, MA, and PhD (1981) at Cambridge and subsequently became university lecturer there. In 1989, he moved to the University of British Columbia in Vancouver, Canada, where he became professor of geography. He is the recipient of University of British Columbia, Killam Research Prize and the Killam Teaching Prize, 2005, and has received the Founder's Medal of the Royal Geographical Society.

For Gregory, language occupies a central position in the interpretation and representation of people and places. His works have been widely noted for their elegant rhetorical style, intriguing metaphors, and seductive phrasing, helping to legitimate a broader shift in academic prose from sterile, cut and dried style long characteristic of most social sciences to a self consciously poetic model evocative of the richly symbolic worlds inhabited by real human beings.

Gregory's numerous works have been very influential. In 1978, his MA thesis, *Ideology, Science, and Human Geography*, was an elegant critique of emerging post positivist lines of thought. He offered a contextual summary of the rise of positivism in *Recollections of a Revolution*, which portrayed that paradigm as the product of scholars laboring to refashion the discipline in the face of intense political opposition rather than the unstoppable triumph of rational logic over empiricism. Gregory also wrote devastating critiques of location theory, systems theory, diffusion theory, and humanistic geography.

Figure 1 D. Gregory.

Simultaneously, his works introduced geographers to Anthony Giddens' highly influential theory of structuration, which effectively overcame the long standing divide between views that emphasized social structures at the expense of individual consciousness on the one hand and ahistorical idealist views such as phenomenology that celebrated individual behavior but ignored social contexts. Structuration theory sought to recover the thoughts, desires, and meanings of actors but simultaneously avoided collapsing notions of social structure into those same motivations, with sociospatial relations always escaping the intentions of their creators. The construction of geographies is thus described by Gregory as a "sensuous swirl of contingency and determination." Gregory's works helped to integrate structuration theory with existing lines of thought such as time geography and humanistic approaches. This view became increasingly accepted throughout the discipline in the 1980s and 1990s, with the volume *Social Relations and Spatial Structures*, a particularly important milestone in this process.

Gregory's works injected social theory into historical geography, which had long been characterized by a straightforward empiricism. For example, he wrote about the uneven spatiality of class struggle in England during the Industrial Revolution with his doctoral dissertation applying structuration theory to the Yorkshire woolen industry. He also examined information circulation in light of the massive time–space compression generated by modernity. These excavations of the origins of modernity enticed geographers and others to unveil its historically specific nature, demonstrating that Enlightenment claims of the universal applicability of rationality (e.g., in Marxism and positivism) representing a particular configuration of power/knowledge.

By the late 1980s, Gregory had begun to explore the vast implications of postmodernism, including its general distrust of broad, overarching theoretical paradigms and the manner in which it opened the door for a renewed regionalism. In 'Solid geometry', he lambasted structuralists who attempted to theorize space, time, and society independently of one another. Geographies and landscapes, he insisted, are not simply passive reflections of some aspatial social structure; rather, space is inevitably an active participant in the formation of social relations and their changes over time. Thus, Gregory's view advocated a much more aggressive role for space without retreating to the frozen spatial geometries of positivism.

The book *Geographical Imaginations* marked the initiation of a new phase in Gregory's career. Drawing upon a wide array of philosophers and intellectual historians, he elaborated upon the ways in which seemingly different theoretical approaches reflected historically specific 'scopic regimes' – ways of seeing and knowing the world that appear consistent and complete within their own frames of reference. Since the Renaissance and particularly the Enlightenment, Western social science, including geography, has subscribed to the notion of a Cartesian observer, who is detached, all knowing, and objective. Because this observer in this model of knowledge has no social context, it is sometimes called the 'view from nowhere', what Gregory calls the "world as exhibition." This notion underlies many seemingly disparate perspectives, including Marxism and positivist conceptions, and confers the power to know and define the world – both social and natural – on a rational, and presumably male, all knowing ego. Drawing upon the seminal work of Martin Jay, Gregory's world as exhibition is a form of 'ocularcentrism', one that arose within the contours of modernity and facilitated the historical process of commodifying social life and space. Within geography, it was reflected in the triumph of abstract space over lived experience, a position made possible only if representations of the world are held to be detached from the world they reflect. The rise of postmodernism and post structuralism, which legitimated relativist points of view, increasingly rendered this position indefensible.

In contrast to the world as exhibition, Gregory argues that every theoretical position is necessarily incomplete and situated, linked to a power interest, and reflective of its utility in social life. Putting aside this scopic regime opens up numerous epistemological standpoints that take power and social context seriously, and in particular give voice to marginalized subject positions, including feminism and subalternity. Avoiding a purely idealist history of social and spatial thought, Gregory links this shift to the massive waves of restructuring and time–space compression characteristic of global capitalism in the late twentieth century. Gregory's work thus paved the way for much post structuralist theorizing within geography, and, unsurprisingly, generated considerable debate.Gregory has also co edited multiple editions of *The Dictionary of Human Geography*, which provide concise summaries of the discipline's terminology and concepts, making clear the meanings and implications of terminology often lost in the jargon of obfuscationist rhetoric.

More recently, drawing upon the concept of Orientalism initiated by Edward Said, Gregory has focused on European colonial representations of non Western spaces. As a power/knowledge relation, the active 'geographing' of various parts of the globe was an integral part of Western administrative control of colonized regions. Not simply a list of use values, geography was also central to the ideological legitimation of colonial power relations. The colonial world as exhibition was thus closely linked to the naturalization of Western dominance and ideologies of non Western inferiority. The colonial geographical imagination demarcated space and brought alien territories into Western frames of understanding in a manner that drew critical boundaries between identities, self and other, and whites and non whites. Colonialism was hence every bit as much a cultural and ideological project as an economic and political one.

Empirically, Gregory's postcolonial work has focused largely on the Middle East, particularly European and American representations of Egypt in the late nineteenth and early twentieth centuries. In the nineteenth century, as Egyptology became fashionable, Egypt occupied a central ideological position in the evolving self conception of the West as both proximate and distant, foreign and yet familiar, ancient yet stagnant, and in dire need of being rendered sensible through the 'empire of the gaze'. Reworking discourses of colonial travelers and administrators, Gregory reveals how Western imaginative geographies were laced with patriarchical, sexualized, and often racist imagery that effectively 'Othered' Arabs in their own lands as Egypt was thoroughly 'geo graphed' by a panoptic foreign authority. Gregory's postcolonial turn also led him to a creative encounter between Orientalism and recent conceptions concerning the social construction of nature. During the colonial era, for example, Europeans often depicted tropical environments as either fecund Gardens of Eden inhabited by childlike innocents in need of conquest or as disease ridden swamps populated by dark savages. Finally, in *The Colonial Present*, he extended postcolonialism into a lucid critique of the 'war on terror' in Iraq, Palestine, and Afghanistan, revealing it to be a war of terror against people rendered invisible by Western discourses that marginalize Muslims.

Gregory's contributions to geography are both conceptual and methodological. For close to 30 years he has repeatedly and emphatically stressed the importance of historical context in academic work, a move that thwarted the aspirations of Enlightenment modernists to construct universal interpretations independent of time and space. He has remained steadfast in his insistence that geographies are peopled, that time and space are made by ordinary folk, and that the outcomes are always contingent and never predetermined, even if, 'contra' overly structuralist interpretations, they are likewise never random or chaotic. His works have decisively demonstrated the need to understand capitalism in dis cursive terms, and he has been an acute observer of the roles played by language in the process of representation. In so doing, he has helped geographers to problematize the question of intellectual inquiry, wrenching it from its perch as a passive exercise in interpretation and recasting it as an active, often muddy, process of constructing meaning. He has remained sensitive to the politics of representation, the ways in which all forms of under standing are inescapably entwined with temporally and spatially varying regimes of power–knowledge: dis courses simultaneously reflect and constitute the re lations they depict and that give rise to them. Finally, his turn to postcolonialism has helped to awaken the dis cipline to the profound but often invisible ways in which geography as discourse was complicit in the systematic Othering of non European peoples.

See also: Historical Geography; Human Geography; Industrialization; Orientalism; Postcolonialism/Postcolonial Geographies; Postmodernism/Postmodern Geography; Structuralism/Structuralist Geography; Structuration Theory; Structurationist Geography; Time Geography.

Further Reading

Duncan, J. and Gregory, D. (eds.) (1999). *Writes of Passage: Reading Travel Writing*. London: Routledge.
Gregory, D. (1976). Rethinking historical geography. *Area* 8, 295 299.
Gregory, D. (1978). The discourse of the past: Phenomenology, structuralism and historical geography. *Journal of Historical Geography* 4, 161 173.
Gregory, D. (1978). *Ideology, Science and Human Geography*. London: Hutchinson.
Gregory, D. (1980). The ideology of control: Systems theory and geography. *Tijdschrift voor Economische en Sociale Geographie* 71, 327 342.
Gregory, D. (1981). Human agency and human geography. *Transactions of the Institute of British Geographers* 6, 1 18.
Gregory, D. (1984). Contours of crisis? Sketches for a geography of class struggle in the early Industrial Revolution in England. In Baker, A. & Gregory, D. (eds.) *Explorations in Historical Geography*. Cambridge, MA: Cambridge University Press.
Gregory, D. (1985). Suspended animation: The stasis of diffusion theory. In Gregory, D. & Urry, J. (eds.) *Social Relations and Spatial Structures*. New York: St. Martin's.
Gregory, D. (1987). Postmodernism and the politics of social theory. *Environment and Planning D: Society and Space* 5, 245 248.
Gregory, D. (1988). Solid geometry: Notes on the recovery of spatial structure. In Golledge, R., Couclelis, H. & Gould, P. (eds.) *A Ground for Common Search*. Santa Barbara, CA: Santa Barbara Geographical Press.
Gregory, D. (1989). Areal differentiation and post modern human geography. In Gregory, D. & Alford, R. (eds.) *Horizons in Human Geography*. Totowa, NJ: Barnes and Noble.
Gregory, D. (1994). *Geographical Imaginations*. Cambridge, MA: Blackwell.
Gregory, D. (1995). Between the book and the lamp: Imaginative geographies of Egypt, 1849 50. *Transactions of the Institute of British Geographers* 20, 29 57.
Gregory, D. (1995). Imaginative geographies. *Progress in Human Geography* 19, 447 485.
Gregory, D. (2001). (Post)colonialism and the production of nature. In Castree, N. & Braun, B. (eds.) *Social Nature: Theory, Practice, and Politics*. Oxford: Blackwell.
Gregory, D. (2002). Orientalism and the cultural politics of nature. In Wynn, G. & Gregory, D. (eds.) *Nature, Culture, Colonialism*. Toronto, ON: University of Toronto Press.
Gregory, D. (2004). *The Colonial Present: Afghanistan, Palestine, Iraq*. Oxford: Blackwell.
Gregory, D. and Urry, J. (eds.) (1985). *Social Relations and Spatial Structures*. New York: St. Martin's.

Relevant Websites

http://openanthropology.wordpress.com/2008/10/24/derek gregory the cultural turn in late modern war and the rush to the intimate/
http://web.mac.com/derekgregory/iWeb/Site/Welcome.html

Grounded Theory

M. Cope, University of Vermont, Burlington, VT, USA

Glossary

CAQDAS (Computer-Aided Qualitative Data Analysis Software) A generic name for a wide range of computer programs that aid in storing, organizing, coding, and analyzing qualitative data.

Coding The process of identifying, assigning, and analyzing codes based on inductive review of transcripts and other data for the purpose of revealing commonalities and disjunctures, investigating patterns and themes, and producing new knowledge and insight. Coding is both a data reduction process and a recursive, fluid, powerful analytical technique. Different types of coding include *in vivo*, axial, and open coding; these represent a spectrum from descriptive to analytical.

Constant Comparison A central characteristic of grounded theory; a process in which researchers use multiple forms of comparison to evaluate and reevaluate their data, codes, themes, and theories.

Recursive Analysis The process of going back cyclically in multiple iterations to reevaluate data and previous analysis in light of new insights.

Social Construction of Knowledge The perspective that knowledge is produced iteratively and recursively through active processes in the social world, including through the research process.

Theory Building In grounded theory, the process of developing themes, concepts, and theories inductively from repeated, iterative data collection, analysis, and identification of core themes.

Introduction

It could be said that geographers rarely 'call' their analytical practices 'grounded theory', but they do often practice the strategies, use the tools, and produce grounded theories without invoking the term. Other disciplines have been much more transparent in their dealings with this set of inductive guidelines for analyzing data, including sociology (where grounded theory was born), anthropology, and nursing. However, with the 'cultural turn' and the revival of qualitative research of the past 15 years, geographers have increasingly sought out methods for conducting rigorous research and grounded theory has much to offer in this regard.

Grounded theory has a relatively short history, being only about 40 years old, but it is one full of changing collaborations, rivalries, disavowals, new developments, divergent trajectories, and colorful personalities. The details of this history are thoroughly covered elsewhere so this article focuses primarily on the form and practice of grounded theory, as well as some of the ways that geographic perspectives can both benefit from grounded theory and contribute to its future. There are many streams of practice of grounded theory, as well as many subtleties in their strategies and methods, but the general coverage offered here is based on those principles closest to current practice in critical human geography. To this end, this article focuses primarily on 'constructivist' grounded theory, that is, the stream taking the view that social realities and meanings are perpetually socially constructed, lived within and against, and reconstructed, and also that research itself is another such dialectical process whereby data are not 'collected' but, rather, 'created' in the course of research interactions. Constructivist grounded theory also fits well with much of postmodern and post structuralist geography because it allows and even encourages the production of knowledge through the recognition of a multiplicity of truths and takes seriously the question of the context and positionality of both the researcher and the participants.

What Is It?

Grounded theory involves, at its heart, a set of strategies, tools, and central principles that aid researchers in doing inductive, reflexive, and rigorous analysis of data. The goal in adopting grounded theory is to produce new insights that are thoroughly 'grounded' in the empirical findings, but also which have strong conceptual threads tying them together. Grounded theory is typically considered an analysis technique, but it actually informs (even shapes) all areas of research – identification of the research questions, data collection, analysis, theory building, writing, and sharing of results. Indeed, one of the central tenets of grounded theory is that these areas of research are not discrete and sequential but instead are iterative and recursive. Data can be collected – or, more accurately, 'created' – at any point during the project and indeed, researchers are expected to go back and forth between multiple data sources, through various analytical techniques and perhaps even writing, frequently returning to do more data collection as new themes emerge, new questions arise, and new concepts need clarification or validation. Thus, in grounded theory much of the linear formula of 'the' research process (hyper legitimated by its roots in the 'scientific method') many of us are trained in is turned inside out and scrambled around, but with the purpose of more rigorous and defensible results. Of course, despite the linear way scholarly articles present research, in reality most social research does not really follow a rigid step by step process of data collection analysis writing up, but already involves the back and forth steps intrinsic to grounded theory. Therefore, grounded theory offers us a chance to formalize and treat more systematically the inherent piecemeal practice of social research.

Strategies and Tools

Grounded theory involves several specific strategies for analysis, though they are not all necessarily essential or foundational to the approach. One of the most central tools of grounded theory is the use of coding. Coding can be practiced in different ways, but the central goal is to organize, categorize, and draw insights from data by identifying core themes that emerge inductively. For example, a geographer might notice that participants mention playing up – or hiding – certain aspects of their identity (age, gender, race, sexuality, etc.) over others depending on which local setting they are in, but they may do so in many different ways. Codes could be used to organize these data so that important meanings, constructions of social space, power, and forms of local knowledge can be pulled out. Coding would allow the researcher to preserve important differences between participants, but also to pull out commonalities in, for instance, people's strategies to avoid confrontations on the street. Coding is in part a process of categorizing, and therefore is a data reduction technique, but it is also an important step in analysis and concept building. So in this example, while there might be a code for the street behavior 'walking tall', there is also a set of concepts that go with this code based on power, fear and other emotions, construction of social space, and identity that needs to be developed.

This is where diagramming and 'memoing' become useful. Diagramming involves mapping the networked relationships between research elements (including codes, themes, quotes or sections of dialog, photos, etc.) in order to see new ways that they condition each other, are contingent upon or consequences of each other, or are otherwise connected. Creating memos is also an important tool of grounded theory because memos about data, about analytical operations, about the meaning or origin of codes, or about emerging ideas of the researcher are an ongoing reflexive type of metadata for the research, which is a rich source of knowledge production. Memos also help to keep track of decisions that are made or concepts that are in their infancy, which in turn boost the rigor of the research by allowing the scholar to be more consistent in coding and to document analytical developments.

Modern computer aided qualitative data analysis software (CAQDAS – a generic term for a host of software programs) are generally well integrated with grounded theory, and thus typically have excellent capacity for constructing, refining, displaying, and analyzing the codes, code families, diagrams, memos, networks, hierarchies, conditional relationships, and metadata that are intrinsic to qualitative research projects. While these programs are very powerful for organizing, storing, and analyzing data, having an understanding of the grounded theory helps researchers both to use the programs more effectively and to maximize their return on investing in the time to learn and use them. Of course, the programs will not 'do' the analysis but they can greatly enhance a project with their functions.

Another important element of grounded theory involves the process of 'constant comparison'. This means that data, codes, and emerging conceptual themes are continuously held up for scrutiny, for reevaluation in light of new understandings, and for examination of their validity. Comparison allows researchers to be more confident in confirming the trends or concepts they have identified in the data, but it also forces researchers to dig deeper to understand mismatches and divergences in the data. To continue with the above example, one study found that elderly white women in an urban neighborhood sometimes used 'walking tall' as a strategy on the street to avoid confrontations or getting mugged, but sometimes some of the participants used the opposite strategy – stopping to look carefully at a street sign in pretend confusion while a group of young people walked by – to avoid attention in public space. These kinds of empirical inconsistencies can actually provoke new inquiries and stimulate new conceptual insights so the goal of 'constant comparison' is not merely to confirm, but also to complicate. Comparison is also important between multiple sources of data (e.g., semistructured interviews as well as travel diaries kept by the elderly women) and helps to tease out subtleties and emerging themes. At the more abstract level of conceptual development, comparison of codes and emerging themes enables researchers to refine their concepts, reach a point of 'saturation' (wherein further analysis of the data stops revealing new themes), and ultimately to build theory that is thoroughly grounded in data.

All of these practices serve several purposes. First, they encourage rigorous and reflexive research through their demands for the careful, critical attention of the scholar. This in turn serves an important legitimating function, which has been especially important for qualitative work. Finally, grounded theory done well produces theory that has been actively and mindfully 'built' that remains sensitive to context, positionality, the social production of knowledge, and the contingencies of the social world. The resulting theory will be both durable – it holds up because it incorporates variations – and flexible – it accommodates researchers' future refine ments as more comparison is completed.

Geography's Contributions

The discipline of nursing has used grounded theory ex tensively and made contributions to it by breaking down barriers between the clinical side of medical practice and the more holistic side of patient care and advocacy. Sociologists have sought to envision 'constructivist' grounded theory that fosters understanding of multiple realities and multiple viewpoints, along with the under standing that the research process constructs 'data' and renders participants' stories in dialogic ways. How might geography or taking a spatial perspective contribute to grounded theory? Issues of context, multiple/intersecting processes and identities, and the social construction of knowledge are familiar terrain for contemporary critical human geographers so it should not be a stretch to envision some potential areas of contribution that the discipline and practice of geography could make to grounded theory. The three points in the following represent a beginning in this process.

Settings and Stories

Geographers are finely attuned to the contextual factors that influence and are influenced by human social action. In some cases location information will include objective site based measures such as addresses and Global Pos itioning System (GPS) points (and these have their uses – see below), but there are also many other characteristics of the settings of the social world that have relevance for grounded theory researchers. Political processes of both formal and informal power relations; economic processes of trade, investment, exchange, and labor; cultural pro cesses shaped by language, history, religion, and change; and social relations of identity, social regulation, social reproduction, and institutions are constitutive of the situations of human agents, who in turn act upon these processes in complicated ways. Geographers are skilled in identifying and analyzing these multiple forces in an integrated way by, for instance, examining the con struction of 'place' as both a product and the medium of everyday life. A place based approach not only benefits from the analytical rigor of grounded theory, but the integrative, comparative, and recursive strategies used in critical geography could also serve to push toward new practices of grounded theory. Discrete boundaries be tween quantitative and qualitative research, between so cial/economic/political processes, between the physical and social world, and between 'researcher' and 'subject' have all been challenged and reshuffled in geography, with varying effects. While much of the work done in other disciplines using grounded theory takes a 'slice of time' approach, place based perspectives offer some fresh opportunities for grounded theory practice. For example, among grounded theorists there is burgeoning concern for subjectivities and how to avoid overemphasizing the individual, which could benefit from geography's con siderations of how people construct place through rou tine performance, actions and verbalizations, expressions of identity and power(lessness), bodily comportment, etc., and simultaneously how people react to place, live in place, resist place, experience place, and 'be' place(d). Geography's focus on context, multiple and intersecting processes, and inherent comparative facilities could help in this reworking of grounded theory, and the mutual benefits would be invigorating.

Qualitative Data Analysis and Geographic Information Systems (GISs)

An area of significant potential contribution from geog raphy to grounded theory is through the integration (or at least communication) of the digital tools we use for data handling, storage, analysis, and representation. Some initial steps along the lines of constructing a 'qualitative geographic information system (QualGIS)' have used 'Atlas.ti' (a CAQDAS that uses nonhierarchical net working of codes and data) and ArcGIS. Because Atlas was developed as a tool for grounded theory analysis, its central characteristics and principles are embedded in the system in various ways. Using the 'codes' that emerge from grounded theory analysis in Atlas as a bridge a connection is then made to a database in the GIS, allowing social and spatial data to blend. An example of this is one researcher's ability to analyze the concept of 'community identity' based on children's artwork and child led neighborhood tours in part by embedding their representations within a map of the city that also shows quantitative data, such as race and poverty distributions, and qualitative data, such as photographs of the location.

Another example of the integration of social and spatial data, 'mash ups' (e.g., using Google Maps), cur rently have no particular relationship to grounded theory but there is the potential to make good use of them. Mash ups are the products of laypeople who incorporate various sorts of qualitative data into existing web based mapping applications. While some of these are purely for fun (e.g., tracking foul smells people encounter at various New York City subway stations), others represent the cutting edge of real time, useful public information, such as mapping and reporting the spatial extent of wildfires in California along with evacuation orders and sites for safe refuge. Mash ups could be especially useful for re searchers needing a spatial element of their 'constant comparison', and would also be very useful for the metadata functions of 'memos' because they are quick and easy to do but can convey a lot of information about spatially referenced data in a simple visual form.

Both qualitative GIS and mash ups are techniques in their infancy, but their future is open to creative new approaches. There is no reason grounded theorists could not employ such integrative methods in all areas of re search: data production, analysis, and representation.

Grounded Visualization

Extending from 'qualitative GIS' and the tool based discussion above, a final realm for productive exchange between geographers and grounded theorists is in the area of grounded visualization. Geographers, particularly those using GIS and 'exploratory data analysis' (EDA), are quite familiar with visualization, which involves looking at spatially referenced data in different ways in order to discern new patterns, trends, clusters, absences, or co occurrences. For example, to take one of the sim plest techniques, zooming in or out changes the scale of the map, which may reveal visual patterns that were not noticeable at the original scale of analysis. In this process, the researcher is exploring the data visually by changing the resolution, filtering out or highlighting certain data, or incorporating statistical techniques to identify clusters or other important patterns. This process is inductive, iterative, and reflexive, just as grounded theory is. These commonalities were identified and built upon by a re search team of two geographers who were interested in the complexities of how community gardens played a role in the community identity of a marginalized social group. They developed the practice of grounded visualization, which fosters rigorous grounded theory analysis using qualitative and quantitative data, spatially and non spatially referenced data, and a set of recursive, reflexive processes for understanding intricate social worlds. In this example, the potential impact of geography on grounded theory is particularly apt as it represents the symbiosis between two methods of analysis that had many common roots but had never been combined.

Concluding Thoughts

Grounded theory has stood the test of time in part through its composition of strategies and characteristics that are flexible enough to adapt to changing epis temologies and ontologies. Early grounded theory (and indeed, much work still done today) was decidedly objectivist and based on the positivist's belief that there existed an independent reality 'out there' that awaited discovery. More recent developments have reflected the postmodern turn of much of social science (and other disciplines, such as nursing) in which there is acknow ledgement that stories told through grounded theory may indeed be rigorously researched but they still represent the mutual and contingent actions, thoughts, and inter pretations of the researcher and the participants at a given moment and in a given place. This reflects a growing understanding that the social world itself is conditioned by multiple processes, informed by multiple 'truths', saturated with power relations, and always contextually situated. Geographers are well poised con ceptually, analytically, and technologically to contribute to the future of grounded theory.

See also: Ethnography; Qualitative Geographic Information Systems; Transcripts (Coding and Analysis); Visualization, Feminist.

Further Reading

Bringer, J., Johnston, L. and Brackenridge, C. (2006). Using computer aided qualitative data analysis software to develop a grounded theory project. *Field Methods* 18(3), 245 266.

Charmaz, K. (2000). Grounded theory: Objectivist and constructivist methods. In Denzin, N. & Lincoln, Y. (eds.) *Handbook of Qualitative Research* (2nd edn.), pp 509 535. Thousand Oaks, CA: Sage Publications.

Cope, M. (2003). Coding transcripts and diaries. In Clifford, N. & Valentine, G. (eds.) *Key Methods in Geography*, pp 445 459. Thousand Oaks, CA: Sage Publications.

Jung, J. K. (in press). Computer aided qualitative GIS (CAQ GIS): A software level integration of qualitative research and GIS. In Cope, M. & Elwood, S. (eds.) *Qualitative GIS: Mixed Methods in Practice and Theory*. Thousand Oaks, CA: Sage Publications.

Knigge, L. and Cope, M. (2006). Grounded visualization: Integrating the analysis of qualitative and quantitative data through grounded theory and visualization. *Environment and Planning A* 38(11), 2021 2037.

Miller, C. (2006). A beast in the field: The Google maps mashups as GIS/2. *Cartographica* 41(3), 187 199.

Mills, J., Bonner, A. and Francis, K. (2006). The development of constructivist grounded theory. *International Journal of Qualitative Methods* 5(1), 1 10.

Strauss, A. and Corbin, J. (1998). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* (2nd edn.). Thousand Oaks, CA: Sage Publications.

Relevant Websites

http://www.qualitative research.net/index.php/fqs
Forum of Qualitative Research, a peer reviewed, multi lingual journal.
http://www.qualitativeresearch.uga.edu/QualPage/index.htm
Qual Page, lots of links to helpful resources & reviews.
http://www.nova.edu/ssss/QR/web.html
The Qualitative Report, journal and links.

Growth Poles, Growth Centers

U. Rossi, University of Cagliari, Cagliari, Italy

Introduction

It can be argued that the theory of what came to be known as 'growth poles' (*pôles de croissance*) – a 'magic label' or a 'mythic catchword' as some commentators put it at the time – symbolized in the 1950s and the 1960s what industrial districts theory and discourse represented in the 1980s and the 1990s: a pathbreaking strand of research and theoretical elaboration, intimately linked to the leading approaches of industrial policy being adopted at the local level, particularly in the disadvantaged regions of both developed and developing countries. It is thereby interesting to look back again to the theory of growth poles and centers and to its translation into a concrete strategy of regional policy, also because it testifies to the ways in which theories and concepts of regional development cyclically appear and disappear within this field of inquiry and its related policy sphere. This circulation of ideas sheds light on the highly historicized and contingent character of regional economic development policies and theories. Regional policies witness the spirit of the times in which they appear and take form and all of them are, in a way or in another, subjected to cycles of rise, consolidation, decline, and then more or less rapid dismissal.

While the place where growth pole theories and strategies appeared and gained their academic reputation was France, for industrial districts theory and some other more recent regional development theories and policy recipes this role has been played by Italy and most notably by the Third Italy. In both cases, these theories and the related policy instruments have acquired an influence transcending their respective contexts of origin and have been applied in a growing number of regions in both the developed and developing countries. Variants of growth poles and centers policies have been implemented all over the world, as this article documents in more detail by referring to cases in Southern Europe, Latin America, and Northern America. The variety of policy applications shows how growth pole theory has become a truly 'cross national' instrument and conceptual framework of regional policy during the post war decades. This is a demonstration of the high level of homogenization at a world scale level experienced by regional policies during the Fordist–Keynesian era.

A difference between Fordist and post Fordist regional concepts and related models of economic and spatial policy lies in the fact that the former, such as growth pole theory, have been adopted most frequently within the context of lagging behind regions, while the latter, such as industrial districts theory and other more recent conceptual tools in the field, have been primarily formulated with reference to more prosperous regions of Western Europe (Third Italy, Baden Württemberg, Catalonia, etc.). This shift testifies to the changing perspective that has gained ground in regional development studies and policies in recent years as compared to the situation in the post war decades, with attention being diverted from the structural conditions of less favored regions and a greater emphasis being placed on the conditions of economic success rather than on the recovery from economic backwardness.

In light of this shift in regional development literature and policy action, this article presents and discusses the scholarly literature dealing with growth pole theory and strategies. The article is organized as follows: first, it reconsiders the conceptual value of growth pole theory; second, it analyzes the implementation of growth pole strategies in a number of regional contexts at an international level; finally, some conclusions are drawn reflecting on the intellectual legacy of this regional development theory and policy experience.

The Theories of Growth Poles and Centers: Potentials and Pitfalls

As originally presented in the mid 1950s and onward by French economist François Perroux, growth pole theory was a largely abstract conceptualization. The theory implied the idea of a complex of industries mutually linked by functional relations and dominated by a propulsive industry (the so called *industrie motrice*), the latter being the engine of the development dynamic, thanks to its intrinsic capacity to innovate and stimulate economic growth as well as to nurture the formation of other economic activities and industries (the so called *industries mues*, or 'mute industries').

Conceptually, Perroux premised his theory upon a neo Schumpeterian understanding of the mechanics of development within capitalist economies. Development proceeds, in Schumpeter's view, by the direct and the indirect effects of innovations which are able to take an economy away from a stationary equilibrium (both sectoral and spatial). This means that the newer and more efficient industries in which innovations take place grow at a faster pace compared to the older and more static industries. Regional economic development, therefore, implies cumulative sectoral and spatial differentiation in impact and thus deviates from a stationary conception of equilibrium growth. Building on this conceptual argumentation, which had many points of convergence with subsequent theories of unbalanced and cumulative growth (notably those of leading economists and planners such as Albert Hirschmann, Gunnar Myrdal, and John Friedmann), Perroux thought that in order to act as a pole the propulsive industry should satisfy the three criteria of: (1) large size, (2) a potential of economic leadership, and (3) a rate of growth faster than that of the local and regional economy in which it becomes embedded. The existence of these conditions allowed the deployment of a mechanic of polarization, which in Perroux's view could take place in two specific respects: first, the leading firm can make anticipation of demand, both correct and incorrect, affecting smaller firms; second, the effects of the leading firm is able to change the balance of factor inputs in other firms.

In formulating the growth pole theory, critics have noted that Perroux referred to a conventionally abstract, topological space of Euclidean kind. It was only at a subsequent stage of the debate on growth poles that some scholars, especially in France, sought to apply Perrouxian theory to a geographic dimension of enquiry and theoretical reflection. Francophone human geographers were in the forefront of this attempt. At that time, human geographers and most notably economic geographers had become persuaded about the necessity to rethink conventional categories and visions of regional growth and change, particularly those inherited from the tradition of regional geography established by Vidal de la Blache and his followers in the first half of the twentieth century. This emergent body of geographic thinking and research advocated the embracement of a more proactive approach to regional enquiry, which was presented under the banner of *géographie active*. In doing so, these geographers asserted the aim of going beyond the objective and prefixed identity and status of regional economic spaces while paying greater emphasis on the role of the state and other planning entities in giving form and shape to regional economies and societies. While the homogenous region was the favorite regional form of Vidalian geographers, during post war decades and most notably during the 1960s the success of polarization theories led to the fore debates about the so called 'polarized region'. According to those proposing this concept such as Jacques Boudeville, a polarized region exists only as a product of a regional economic planning and policy initiative.

Somehow consistently with the aim of spatializing growth pole theory, other authors – most notably urban scholars – referred to a location, usually a city region, having a polarizing effect and potential in terms of 'growth center'. This latter field of enquiry offered an expanded interpretation of the Perrouxian theory of growth poles, trying to show how the role played by growth centers in regional development is a particular case of the general process of innovation diffusion. In doing so, these scholars drew on functional urban theory: in particular, Brian Berry's conceptualization of urban systems theorizing a system of cities arranged in a hierarchy according to the functions performed by each. Discussions about the critical size of the city for being a 'growth center' were central to this line of enquiry, with the 250 000 population figure being the conventional threshold. Theorizers of growth centers related the differential of demographic growth displayed by cities and metropolitan areas to the local economic structure and particularly to the sectoral division of labor at the urban level. In advanced countries – these authors argued – the greater the earnings derived from primary activities, the lower the growth. As earnings from manufacturing increase, the growth rate tends to stabilize around the national average. The greater the share of earnings from the tertiary sector, involving innovative activities such as research and development and education, the greater the population growth rate.

But even in those contributions that retained a closer linkage to Perrouxian theory, while attempting to offer a spatialization of the theory itself, the conceptual mismatch was evident. In general terms, the translation of the Perrouxian theory of the growth pole into a theory of the 'growth center' (or growth areas, development nuclei, core areas, according to the varying definitions) entailed a translation of an economic concept into a geographic concept. This act of translation from the realm of economic space to that of geographic space has been among the most controversial and critiqued aspects in this literature. While Perrouxian theory itself did not remain untouched by critiques stressing its incompleteness and vagueness, an even much stronger and more severe criticism has been directed against the geographic applications and reformulations of his theory. Geographic translations of the Perrouxian theory have been widely criticized for absence of theoretical rigor and sophistication, for lack of a satisfactory empirical evidence and at the same time for a tendency toward empiricism. According to critics, the concept of 'growth center' has exceeded the original conceptualization of the Perrouxian notion of 'growth pole', but at the same time has also developed an ambiguous relationship with it. In particular, three weaknesses have been identified within those theoretical efforts attempting to spatialize growth pole theory and to apply it to specific geographic and policy contexts. First, critics have questioned the basic assumption that the growth pole in geographic space (the 'growth center') was simply a specific variant of the growth pole in economic space, producing the semantic confusion of attributing to a location the growth characteristics of the pole (industry); second, critics expressed skepticism about the main assumption underlying the geographic translation of growth pole theory, that is, the idea that the 'natural' growth pole existing in geographic space could be automatically replicated in the form of a planned growth pole; third, they lamented the loss of the original neo Schumpeterian attention toward economic innovation and signaled, on the contrary, a prevailing emphasis being laid on firm size with its multiplier potentials as a distinguishing feature of growth centers and poles.

However, despite this conceptual looseness as regards the relationship to the Perrouxian formulation, the geographic understanding of growth pole theory provided a bridge with another and older set of theories investigating the organization of human activity in geographic space: most notably, August Losch's economics of industrial location and Walter Christaller's central place theory. Albeit these theories and the geographic reformulations of growth pole theory differ in their way of reasoning, the former embracing a distinctively deductive method and the latter an inductive method of analysis, it can be nonetheless argued that they are complementary in their contribution to regional science. In fact, while Jacques Boudeville's theory provides an explanation of the developmental impact of localized poles of growth in geographic space, it is not by itself a theory of location explaining where the functional growth poles are or where they will be localized in the future. To clarify this, some authors have argued that the growth pole theory should rely on theories of location, such as those of Losch and Christaller. In fact, central place and industrial location theories, despite their obsession with spatial regularities and their blindness toward the understanding of the dynamics of economic growth, are to be regarded as useful points of departure for analyzing the impact of development in a given center on the other centers and, in doing so, appear to well supplement theories of growth poles and centers, which on their part concentrate their analytical efforts on the functioning of the economic process.

The translation of the growth pole conceptualization into a geographic theory of regional development has proved to be problematic not only at a theoretical level but also at a practical level. The experiments of regional policy that have been undertaken in different regions of the world are illustrative of the potential as well as the limitations of this controversial theory and policy tool.

Using Growth Pole Theory as a Regional Development Policy

The formulation of growth pole theory and the related application to concrete regional contexts arose in the wake of the search for policy and planning tools to solve problems of imbalance among regions in both industrialized and so called 'underdeveloped' countries during the post war decades. Despite its conceptual vagueness and elusiveness, as highlighted in the previous section, growth pole theory proved to be a highly popular tool of regional policy in a number of different geographic and economic contexts. Its popularity largely lies in the fact that it was viewed by planners and policymakers as a sort of 'complete theory' of economic growth, capable of integrating all interrelated aspects of development in a large variety of geographic settings.

In Southern Europe growth pole policies have been pursued in Italy and Spain in the 1950s and the 1960s. At the time these countries strongly differed in their political, economic, and geographic situations. Italy was an emerging capitalist country, experiencing an intense process of industrialization in the years between 1958 and 1963, albeit its economic geography was strongly characterized by a deep North–South divide. On the other hand, Spain was politically isolated, the economy strongly nationalized and heavily regulated from the center, while the dominant features of its economic geography were not those of dualism as in Italy but of fragmentation and dispersal: Spain's main centers of activity were far apart from each other and were linked by an imperfectly developed transportation system. In Italy, a growth pole strategy was devised in the southern regions of the country. The growth pole strategy took the form of an industrialization plan which was organized by concentrating public investments in a relatively large number of industrial poles, in which the mechanic of growth was to be generated by the settling of a propulsive industry specializing in a capital intensive and relatively technologically advanced sector of the economy. A law approved by the Italian Parliament in 1957 established about 100 industrial poles, even though those that proved to be effective were much fewer and – even more importantly – the propulsive effect of many of them proved to be weaker than that envisaged by Perrouxian theoretical framework. Several poles of industrial growth demonstrated to have been created with goals of consensus building, consistently with the functioning of the Italian system of political clientelism. In the end, only a limited number of regions had benefited from the growth pole strategy: the Campania region with some large industrial complexes established in the broader metropolitan area of Naples; the Puglia region with a large chemical industrial settlement in Brindisi and a big steel plant in Taranto; and then Sicily with the industrial sites of Gela and Syracuse specializing in chemicals and petrochemicals. Geographically the policy strategy was therefore implemented mainly in the coastal regions, while the inland areas became only marginally involved in the industrialization plan.

In Spain, the growth pole policy was much more selective than the one implemented in Italy: in 1964, seven poles were established (Saragoza, Sevilla, Valladolid, La Coruña, Vigo, Burgos, and Huelva) and, in 1969, five more poles were created by the national government. One of the most interesting facts in the establishment of these poles was that they were not concentrated in the most backward regions. Thus, in the Spanish experience the growth pole strategy was not used as an instrument for developing the poorest regions, but was approached as a way to find suitable centers for expansion along the major (present or intended) transport and development axes. This means that considerations of regional equity were not central to the pursuit of the policy strategy. In general, the strategy suffered from a lack of interinstitutional coordination between the national and the regional levels of government, and its effects were limited by the scant attention that was devoted to crucial issues in regional economic development such as the enlargement of the firm size (Spanish industry was dominated by very small firms at the time) and the linkage with environmental and social issues such as the provision of social housing.

One of the best known growth pole strategies has been the one pursued in the Appalachian Region in the United States. This was an unusual example of lagging region, as it was located between two of the most highly industrialized and urbanized regions of the world, the Atlantic megalopolis and the industrial Midwest. The Appalachian Regional Commission viewed the lack of urban centers capable of providing producer services, trained labor, and other external economies as the factors lying behind the backwardness of this region. It is in light of this context that the Appalachian Regional Development Act which was approved in 1965 has to be understood. In contrast to the scattering of public investments that characterized earlier attempts to aid depressed areas, the Act concentrated the public investments in areas where there was the greatest potential for future growth. The Commission identified 30 'growth areas' as principal recipients of the public investments. Each area was organized around a 'growth center', which was defined as "a complex consisting of one or more communities or places, which, taken together, should provide, or were likely to provide, a range of cultural, social, employment, trade and service functions for itself and its associated rural hinterland." A recent assessment of this policy experiment has pointed to some critical issues that resemble those raised with reference to the Italian and the Spanish experiences (especially the former): an inconsistency between theory and implementation; the low level of place selectivity and the identification of too many areas, many of which could not reasonably qualify as growth centers; the policy's inherent exposure to political pressures. Especially the latter has proven to be a decisive factor in growth pole strategies' failures or only partial success: the identifications of only a few growth centers amid political demands and pressures to designate many have been the main contradiction in the implementation of this policy.

In Latin America growth pole strategies were devised in almost all countries, even though concrete results were poor. In Chile in the late 1960s a growth pole strategy was established in each of 11 regions (typically the provincial capital), a few of which were selected as functional poles: for example, Arica for automobiles and electronics, Santiago for electronics, Concepción for steel and petrochemicals, and Punta Arena for petrochemicals. However, despite some promising signs of regional resurgence, polarization strategy was adopted too passively on the regional level and was abandoned by the national government too early to be effective. In Peru the 1971–75 plan attempted to alleviate the economic dominance of Lima–Callao through a decentralization policy which identified two novel axes of national growth where secondary cities were located. However, the industrial aspects of the pole growth strategy were overlooked and the political will behind the strategy itself proved to be weak. The Venezuelan case has been relatively well known, particularly for the Ciudad Guayana project which, on the one hand, transformed a frontier region into an integral part of the national economy but on the other hand was not exempt of enclave effects. In Colombia the 1969–72 Development Plan aimed to counteract Bogotá's dominance by developing the next three largest cities (Medellín, Cali, and Barranquilla) as equilibrium metropolises. However, the strategy never got off the ground as the Borrero administration that came into power in 1970s subjugated polarization to sectoral priorities. In Bolivia, the 1971–91 Plan published in 1970 included a growth pole strategy with Cochabamba designated as a service pole, Santa Cruz and Oruro as industrial poles, and the much smaller centers of Sucre and Tarija as agro industrial poles, but was never implemented and remained a 'paper plan'. In Argentina the 1970–75 Plan identified several regional growth poles, but the spatial choices were poor and orientated by political motives rather than by a policy rationality: for instance, three of the poles were in Patagonia, which has only 2% of the population, while the high density areas of the North were relatively neglected. In Brazil the most successful attempts at polarization strategy have been the promotion of small metropolitan poles within the São Paolo state and the transfer of the administrative capital to Brasilia, but overall the regional policies and programs in this country have been too diffuse to be described as a coherent growth pole approach. Generally speaking, Latin American experiments of growth pole policies have suffered from similar limitations that have been detected with reference to other geographic contexts: politicization of localization choices, lack of integration with wider national development strategies, and neglect of institutional factors. Moreover, these experiments have encountered a number of regionally specific problems: most notably, the weakness of demo cratic political structures and the prevailing of authori tarian political regimes; the importance of monopolistic industrial structures and associated autarchic economic policies in these countries; the relative absence of middle class elites, managerial talent and associationist leaders, particularly in secondary cities; and the lack of an inter dependent inter urban network and the weak economic and functional base of most Latin American cities, except for coastal cities and capital cities.

Conclusion

Growth centers and poles strategies have been imple mented in a tremendous variety of geographic settings and socioeconomic contexts: not only in those illustrated of the Italian Mezzogiorno, of Spain, Latin America, and the Appalachian Region in the United States, but also in other Western countries (Canada and the UK) as well as in developing countries such as India and Tanzania. While these development strategies have occasionally achieved success, on the whole final verdict is un questionably negative: growth poles have not met original expectations within academia, policymakers and the wider public alike. Policy's advocates lament the sus picious *volte face* that has led scholars of regional eco nomic development first to celebrate the virtues and potentials of growth pole theories and policy recipes then to quickly dismiss them as something *passé* or even as a sort of relics of post war top down planning approach. However, as a matter of fact, the theory originally pro posed by François Perroux and subsequently amended by other scholars of regional economic planning has sub stantially failed to generate self sustaining regional de velopment processes and then has disappeared along with the general decline of Keynesian policies and of related state centered modes of local economic governance.

Even so, the intellectual legacy and influence of growth pole theory are still remarkable. In particular, some of the fundamental tenets and concepts in this literature survive within contemporary strands of re search commonly labeled as 'new geographical eco nomics' (the work, among others, of Krugman, Glaeser, and Porter) and 'endogenous growth theory' (led by economists such as Lucas and Romer), which are highly influential in today's mainstream regional and urban economics. Most notably, ideas about agglomeration and localization economies and a number of neighboring concepts and catchwords such as 'industrial clusters', 'cities and regions as engines of growth', 'learning re gions', 'industrial districts', and the like are reviving the classical, post war tradition of regional development studies of which growth pole theory has been an important part. Some overly simplistic versions of growth pole theories have become clearly obsolete, particularly those associating economic and spatial externalities with the size of a city (or a region). Other aspects of growth pole theory are certainly outdated (the main are probably the rigidly top down approach to economic spatial planning, and the neglect of the sociocultural factors in regional development processes), but of course it would be ungenerous to expect from a theory that was formu lated in the 1950s to express views and conceptual sensibilities that at the time had not appeared yet.

What has to be emphasized is the geographic focus of the theory itself and its significance in the light of the current scenario of economic development and related regional development studies. In the 'Introduction' to this article it has been noted how growth pole theory re flected the spirit of the times in which it appeared in the ways in which it gave prominence to the intervention in the regions lagging behind both in developed and in developing economies. Today, on the contrary, impera tives imposed by neoliberal ideologies are shifting at tention from policy strategies tackling the structural conditions of economic backwardness to those aiming at enhancing the competitiveness of urban and regional economies in an increasingly globalized economic and societal realm. While the main body of economic theory is shared by both contemporary and past strands of re search, at least within mainstream urban and regional economics, it is this shifting attention from the less favored to the prosperous regions and cities as subjects and objects of theorization and intervention that makes the difference between the past and the current state of affairs within this field of enquiry.

See also: Agglomeration; Cassa per il Mezzogiorno; Central Place Theory; City-Region; Corridor and Axis Development; Firms; Fordism; Fordism, Post-Fordism and Flexible Specialization; Functionalism (Including Structural Functionalism); Industrial City; Industrial Districts; Industrial Location; Industrialization; Latin American Structuralist School; Local Economic Development; Métropole d'équilibre; Neoliberal Economic Strategies; Neoliberalism and Development; Region; Regional Development Theory; Regional Inequalities; Regional Planning and Development Theories; Regional Science; Transport and Accessibility; Uneven Development.

Further Reading

Berry, B. J. L. (1972). Hierarchical diffusion: The basis of developmental filtering and spread in a system of growth centers. In Hansen, N. M. (ed.) *Growth Centers in Regional Economic Development*. New York: The Free Press.

Boudeville, J. R. (ed.) (1968). *L'espace et les pôles de croissance. Recherches et textes fondamentaux*. Paris: Presses Universitaires de France.

Darwent, D. F. (1969). Growth poles and growth centers in regional planning a review. *Environment and Planning* 1, 5 32.

George, P., Guglielmo, R. and Kayser, B. (1964). *La géographie active*. Paris: Presses Universitaires de France.

Hall, P. and Hay, D. (eds.) (1980). *Growth Centres in the European Urban System*. London: Heineman.

Hansen, N. M. (ed.) (1972). *Growth Centers in Regional Economic Development*. New York: The Free Press.

Hermansen, T. (1972). Development poles and development centres in national and regional development. In Kuklinsky, A. (ed.) *Growth Poles and Growth Centres in Regional Planning*. Paris: Mouton.

Lasuen, J. R. (1969). On growth poles. *Urban Studies* 6(2), 137 161.

Muscarà, C. (1967). *La geografia dello sviluppo. Sviluppo industriale e politica geografica nell'Italia del secondo dopoguerra*. Turin: Edizioni di Comunità.

Nichols, V. (1969). Growth poles: An evaluation of their propulsive effect. *Environment and Planning* 1, 193 208.

Parr, J. B. (1999). Growth pole strategies in regional economic planning: A retrospective view. Part 1. Origins and advocacy. *Urban Studies* 36, 1195 1215.

Perroux, F. (1955). Notes sur la notion de pôle de croissance. *Economie Appliquée* 8, 307 320.

Perroux, F. (1964). *L'économie du XX siècle*. Paris: Presses Universitaires de France.

Richardson, H. W. (1971). Regional development policy in Spain. *Urban Studies* 8(1), 39 53.

Richardson, H. W. and Richardson, M. (1975). The relevance of growth center strategies to Latin America. *Economic Geography* 51(2), 163 178.

Wood, L. E. (2001). From theory to implementation: An analysis of the Appalachian Regional Commission's growth center policy. *Environment and Planning A* 33, 551 566.

H

Habitus

G. Setten, Department of Geography, Norwegian University of Science and Technology, Trondheim, Norway

Glossary

Capital The resources that actors take to the field. The meaning of capital is not limited to the conventional economic field, but is also applicable to resources such as status, power, knowledge, and social networks. In addition to economic capital, Bourdieu identified three forms of capital: social, cultural, and symbolic.

Field The analytic definition of field is a network, or a configuration, of objective relations between positions. More simply, a field (e.g., the economic/scientific/sport field) is a particular social space of institutions and forces. Our practical sense is specific to the field in which we are acting at any moment. A field is a site of tension, contradiction, and social struggle, and hence the site where different forms of capital are produced and reproduced.

Practice The act of doing something or the ongoing mix of human's everyday activities. For Bourdieu, practices are an effect of the temporal and spatial operation of habitus and capital within different fields.

Habitus

Habitus, stemming from Latin, refers to habitual or typical condition, state, or appearance. In human geography, habitus is known as one of French sociologist Pierre Bourdieu's (1930–2002) key terms. In *The logic of practice*, Bourdieu defines habitus as:

> systems of durable, transposable dispositions, structured structures predisposed to function as structuring structures, that is, as principles which generate and organize practices and representations that can be objectively adapted to their outcomes without presupposing a conscious aiming at ends or an express mastery of the operations necessary in order to attain them. Objectively 'regulated' and 'regular' without in any way the product of obedience to rules, they can be collectively orchestrated without being the product of the organizing action of a conductor. (Bourdieu, 1990)

This is an abstract and formal definition of what is ultimately aimed at describing a set of embodied dispositions that are constituted in practice and are always oriented toward practical rather than cognitive functions.

Bourdieu's use of 'open concepts', as well as the complex nature of his writing, makes it very difficult to pin down a precise meaning of habitus. His sense of meticulous empirical work suggests that a concept can only prove its use by being put to use empirically. To get a grip on what habitus might mean, then, it is useful to see it as a theoretical viewpoint of looking at and interpreting the empirical world, not limited to a conceptual map of the world that can be read off from the application of the notion. Habitus is thus a theoretical attitude or belief system, which can be understood as that which links objective structures and individual action. It refers to how systems of social norms, understandings, and behavior are embodied in individual actors. Habitus does not determine action, but it ensures that individuals are more inclined to act in some ways rather than others.

The Genesis of Habitus

Underlying Bourdieu's writing was the need to escape the unsustainable, yet widespread, Cartesian dichotomy of objectivism and subjectivism (or structure/agency, mind/body). In *The logic of practice*, he examines the traditions of objectivism and subjectivism, and positions himself against both structuralist (i.e., objectivist) approaches that fall into mechanical determinism, and subjectivist approaches that presuppose calculation, rationality, or continuous conscious decisions by actors. As a replacement Bourdieu puts forward the creative, active, and inventive capacities of agents, that is, the active side of practical knowledge, and proposes habitus as that which practices operate from.

Habitus: Accounting for the Logic of Embodied Practices

Habitus is a socialized subjectivity and provides a sophisticated and important perspective on complex processes of embodiment in that it describes our customary relationship to the world shown through our embodied expressions, understandings, and actions present at any time. Habitus therefore seeks to understand and explain individuals' sense of how to act in specific situations while not making fully conscious decisions, that is, habitus accounts for people's practical sense or their 'feel for the game'.

Habitus is a useful tool for thinking about how social relations are internalized and experienced as 'natural', even necessary. Habitus, being the embodiment of the social, is 'at home' in the field it inhabits. Individuals, then, bring to any field their embodied dispositions, as well as their capital or accumulated resources. Moreover an individual's social position becomes expressed through his or her capital. Habitus thus becomes the product and the producer of the division of society into groups and classes, thus also representing the resource for social struggle, unevenness, and contestation within and across different fields.

Bourdieu argues that our being in the world is largely a practical mastery of the implicit principles of the social world in general and different fields in particular. Thus habitus accounts for the logic of practice without having logic as its principle. Habitus is consequently not mere routine, it is improvisational and open to innovation. It has a mode of development through time, both individually and from one generation to the next. Habitus is thus a flexible and negotiable representation of oneself through practice adjusting to the demands inscribed as objective potentialities in the situation. It follows that habitus is relational and it is thus possible to state that habitus is individual, yet social; it is stable, but also dynamic; and it is action through time, being something we continually put into action when producing, constructing, and holding onto differing situations. When cultural and social situations are produced through embodied practices, we have them in our actional purview, that is, in our range of embodied understanding. Socially and culturally contingent situations are therefore held through body and memory, and this is how the durability of habitus is expressed: by my active holding on to a situation so as to prolong what I experience beyond the present moment.

Problems with Habitus

Scholars have in particular identified three interlinked problems with habitus. First, claims are made that habitus has a tendency to become deterministic. Even though Bourdieu emphasizes the generative capacities of habitus, he makes clear that it is its structured and reproductive nature that is crucial for its formation and operation. Bourdieu's interest in habitus as a theoretical tool lies in its abilities to explain the reproduction of class relations, rather than the dynamics of organisms.

The second problem following on from the first, concerns whether habitus is essentially static or whether it is able to change according to changing conditions and circumstances. Because Bourdieu emphasizes and explains the inertia of habitus, and to a lesser extent its dynamism, the temporality of habitus tends to be downplayed in his writing.

The third problem concerns the question of consciousness. The whole notion of habitus was proposed in order to reject positivism and to overcome subjectivist/objectivist, structure/agency, mind/body relationships. It is thus ironic that Bourdieu tends to take a reductionistic view on consciousness by arguing that practical belief is not a state of mind, but a state of the body. He thus reproduces the very idea he sets out to criticize.

Habitus and Human Geography

Although human geographers' engagement with habitus has been relatively limited, three different uses can be identified: first, as a bridge between structure and agency; second, in relation to class, cultural capital, choice, and constraint; third, as a perspective on the body, embodied knowledge, practice, and performance. The use of habitus by human geographers is thus thematically wide ranging, as demonstrated by the following works:

Thrift's development of a 'nonrepresentational' social theory draws on the notion of habitus. According to Thrift, habitus frames practice and time, but not practice and space. In 'nonrepresentational' theory, habitus' qualities in explaining practice, practical action, and time is coupled with a theory sensitive to spatial flows of conduct that are embodied, social, and contextually specific.

Gerber [Tsouvalis] discusses the duality of culture and nature in Western thought. She argues that in order to work across the boundaries between the physical, the mental, and the social, there is a need for a new language. Habitus is proposed as being crucial for this language: it connects the physical and the mental because of its embodied character, and because it generates and is generated through practice it is also social.

Setten puts habitus to work as a 'third term' in order to explain how farmers, that is, bodies, and landscapes are tied together. In order to grasp the landscape practices of a farming community in southwestern Norway, she argues that habitus, as betweenness, is what ties farmers, their embodied practices, and the landscape together.

Harvey and Jones have used habitus to argue that continuity and custom are seen as important themes in the construction of institutional legitimacy, communal identity, and memory. Examining founding legends and hagiographies of Middle Ages Wales and Cornwall, they state that at the heart of change is a concept of continuity. Habitus mediates such permanence in change.

As Painter, another human geographer arguing for the usefulness of habitus, has noted, Bourdieu's work has in (human) geography for the most part been referred to in passing. However, as demonstrated by the above works, the flexibility of habitus as a notion offers considerable scope for human geographers to engage habitus in more depth.

Habitus in Future Human Geography

For human geographers pursuing a social theory that is concerned with the contextual conditions for human behavior, investigating the operation of habitus, capital, field, and practice represent a potentially valuable ap proach. Good social theory can provide critical insight into and explanation of the nature of the conditions that work to produce and reproduce social lives. Embedded in the very notion of habitus are clues as to how certain conditions work to cement certain social lives, that is, how habitus conceals and reproduces unevenness, dom ination, and hegemony. There are also, on the other hand, clues as to how habitus can work to change such con ditions, that is, how contesting, changing, and evolving habitus(es) might lead to transformed social conditions. The nature of production and reproduction of habitus needs, however, to be subject to further scrutiny by human geographers, not least because the dual nature of habitus needs to be made spatially sensitive in order to be a relevant disciplinary concern.

See also: Body, The; Embodied Knowing; Human Geography; Non-Representational Theory/Non-Representational Geographies; Performance, Research as; Performativity; Structuralism/Structuralist Geography.

Further Reading

Bourdieu, P. (1977). *Outline of a theory of practice*. Cambridge, UK: Cambridge University Press.

Bourdieu, P. (1985). The genesis of the concepts of habitus and of field. *Sociocriticism* 1, 11 24.

Bourdieu, P. (1990). *The logic of practice*. Cambridge, UK: Polity Press.

Bourdieu, P. and Wacquant, L. J. D. (1992). *An invitation to reflexive sociology*. Cambridge, Oxford, UK: Polity Press.

Gerber [Tsouvalis], J. (1997). Beyond dualism the social construction of nature and the social construction of human beings. *Progress in Human Geography* 21, 1 17.

Harvey, D. C. and Jones, R. (1999). Custom and habit(us): The meaning of traditions and legends in early Medieval Western Britain. *Geografiska Annaler* 81B, 223 233.

Hillier, J. and Rooksby, E. (eds.) (2005). *Habitus: A sense of place* (2nd edn.). Aldershot: Ashgate.

King, A. (2000). Thinking with Bourdieu against Bourdieu: A 'practical' critique of the habitus. *Sociological Theory* 18, 417 433.

Noble, G. and Watkins, M. (2003). So, how did Bourdieu learn to play tennis? Habitus, consciousness and habituation. *Cultural Studies* 17, 520 538.

Painter, J. (1997). Regulation, regime and practice in urban politics. In Lauria, M. (ed.) *Reconstructing urban regime theory. Regulating urban politics in a global economy*, pp 122 143. London: Sage Publications.

Painter, J. (2000). Pierre Bourdieu. In Crang, M. & Thrift, N. (eds.) *Thinking space*, pp 239 259. London: Routledge.

Setten, G. (2004). The habitus, the rule and the moral landscape. *Cultural Geographies* 11, 389 415.

Thrift, N. (1996). *Spatial formations*. London: Sage Publications.

Relevant Websites

http://www.csc.ehess.fr
Centre de Sociologic européenne.

Hagerstrand, T.

R. Flowerdew, University of St Andrews, St Andrews, Fife, UK

Hägerstrand, Torsten (1916–2004)

Stig Torsten Erik Hägerstrand (**Figure 1**) was born in 1916, the son of a village schoolteacher, and grew up in the Swedish countryside, which he continued to love throughout his life. It is relatively easy to find out more about the development of Hägerstrand's ideas because several of his papers contain autobiographical material, described and discussed in some detail to illustrate theoretical concepts. Most useful, perhaps, is an auto biographical essay (1983), which is accompanied by a career profile chart (**Figure 2**). This shows that Hägerstrand's ideas were taking shape well before he matriculated at Lund University in Southern Sweden in 1937, studying geography and the history of art. In 1944 he obtained his BA, and he continued to a *licentiate* degree (approximately similar to a PhD), which he obtained from Lund in 1953. The length of time taken reflects interruption by military service, as well as the mammoth self imposed task of collecting data on everybody who had lived in the small settlement of Asby from 1840 to 1940.

Figure 1 Torsten Hägerstrand. From Buttimer (1983, p 238).

In 1953, he was given a staff appointment in the Human Geography Department at Lund. He stayed at Lund, serving as Head of Department from 1957 to 1971, being awarded a personal research chair in 1971. He at tracted many international visitors to Lund, developing links with the leaders of the quantitative movement in the United States, visiting the University of Washington in 1959. In addition to his academic work, Hägerstrand became heavily involved in practical planning work for the Swedish government through the 1960s, although he was able to return to his own research after 1971. From the 1970s, he assembled an impressive group of col leagues sharing his interest in time geography. He died in Lund in May 2004.

Hägerstrand was initially not very keen on geography, especially regional geography, which he found boring, repetitive, and lacking intellectual challenge. His interest in the subject developed, however, especially through working with Edgar Kant, an Estonian geographer who moved to Sweden as a refugee in 1944. Kant was inter ested in the daily life of ordinary people as a valid topic for geographical research. He also had interests in lo cation theory, and introduced Hägerstrand to the work of von Thunen and Christaller.

In 1940, Hägerstrand was asked to work on a regional monograph. He decided to concentrate on population development, noting that there had been large scale emigration to the US in the later nineteenth century. In order to investigate this, he decided to try to trace the life of every individual within the small settlement of Asby from 1840 to 1940 (10 000 lives in all). This was done using a mixture of historical research, talking to people, and environmental investigations, producing a general picture. "The arduous acquisition of this image," said Hägerstrand (1957), "is my single most essential experi ence as an adult geographer." He continued throughout his career to use these data to illustrate a variety of concepts as his distinctive world view developed.

Hägerstrand's PhD thesis was concerned with innova tion diffusion as a spatial process. He studied the dif fusion of several innovations in central Sweden, including agricultural innovations (e.g., state subsidized pastures, the control of bovine tuberculosis and soil mapping) as well as general innovations (e.g., postal checking services, the car). His work started with detailed original data collection, and proceeded to construct simulation models

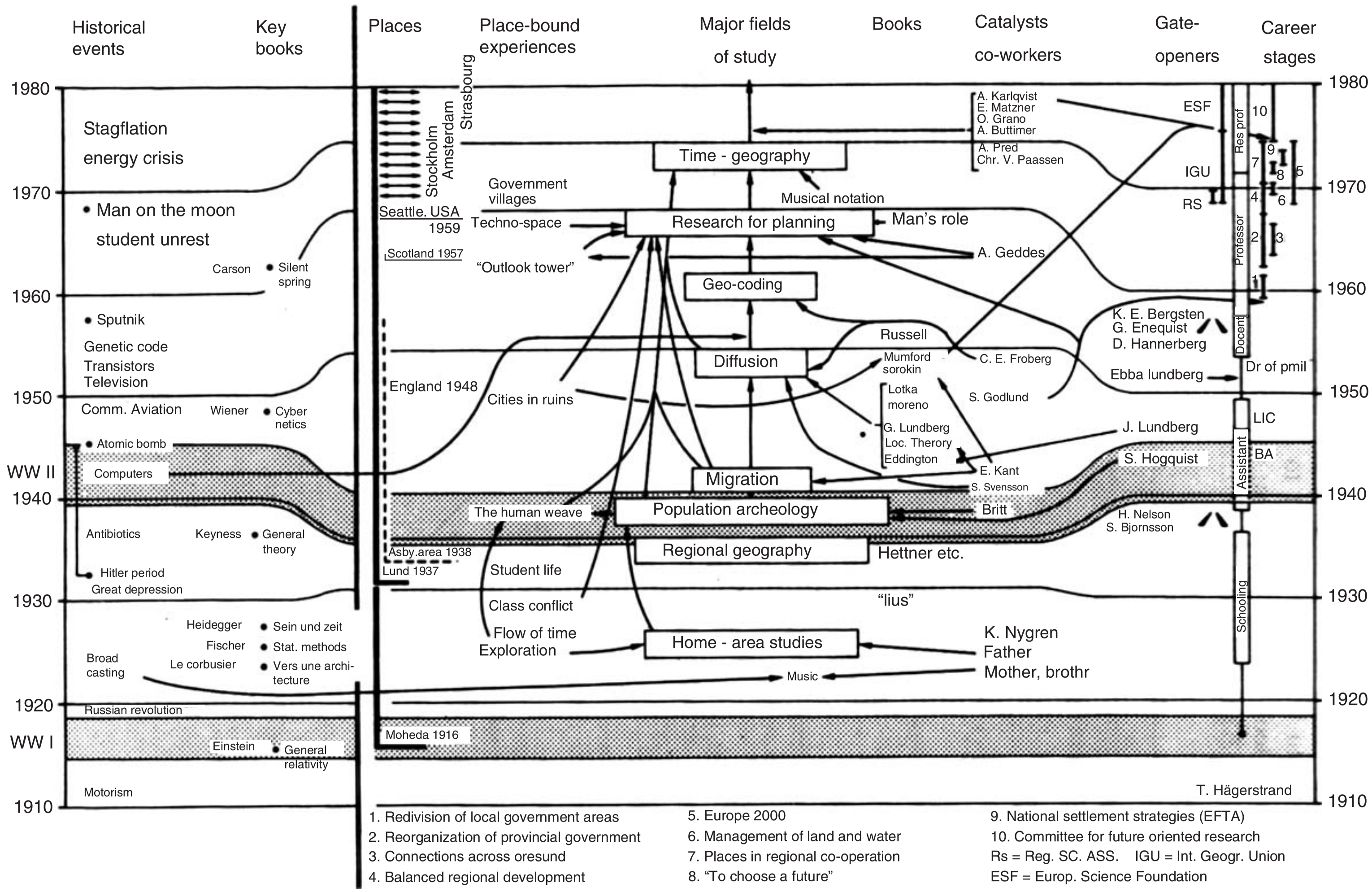

Figure 2 Career profi e: Torsten Hägerstrand. From Hägerstrand, T. (1983). n search for the sources of concepts. n Butt mer, A. (ed.) *The Practice of Geography*, pp 238–256. London: Longman.

of the spatial diffusion of the innovation. The thesis was published in Swedish in 1953, and translated into English by Allen Pred in 1967, accompanied by a postscript by Pred discussing the significance of Hägerstrand's work. This thesis proved tremendously influential in general izing the concept of spatial diffusion, and in encouraging the use of simulation methods, based on the concept of the 'mean information field'. Hägerstrand's innovative use of Monte Carlo simulation techniques was based on drawing random numbers to model the spread of the innovation.

Hägerstrand's approach to innovation diffusion in volved modeling the behavior of individual people, and his work was part of a reorientation of the discipline toward the study of individual rather than aggregate behavior. Some of his work explicitly argued for geog raphy to pay more attention to the individual level of analysis, asking the question "What about people in re gional science?" He criticized the restriction of regional science models to the study of mass aggregate effects, and complained that "we treat a population as a mass of particles, almost freely interchangeable and divisible" (1970: 9). Hägerstrand hence insisted that people are individuals, that they retain their identity over time, and that the life of an individual is his or her foremost project. This led to an emphasis on biography as a key concept, both as method and subject matter.

At the same time, Hägerstrand began to outline some of the basic principles of time geography, which he and his students developed through the 1970s. He developed the idea of time–space prisms, derived from diagrams illustrating the spatial extent of the area that can be reached in free time between two fixed commitments. Time geography could depict an individual's path through space using three dimensional diagrams, where time is the vertical axis (**Figure 3** shows some of the basic ideas). This type of analysis can be especially useful in studying the temporal and spatial constraints everybody faces in carrying out their various projects. Hägerstrand identified three main types of constraint: 'capability constraints' (referring to restrictions on distance of travel, including the necessity of sleep and of eating); 'coupling constraints' (concerning where and when two or more people have to meet); and 'authority constraints' (where a segment of time–space is under somebody's control and hence inaccessible to other people). In a paper published

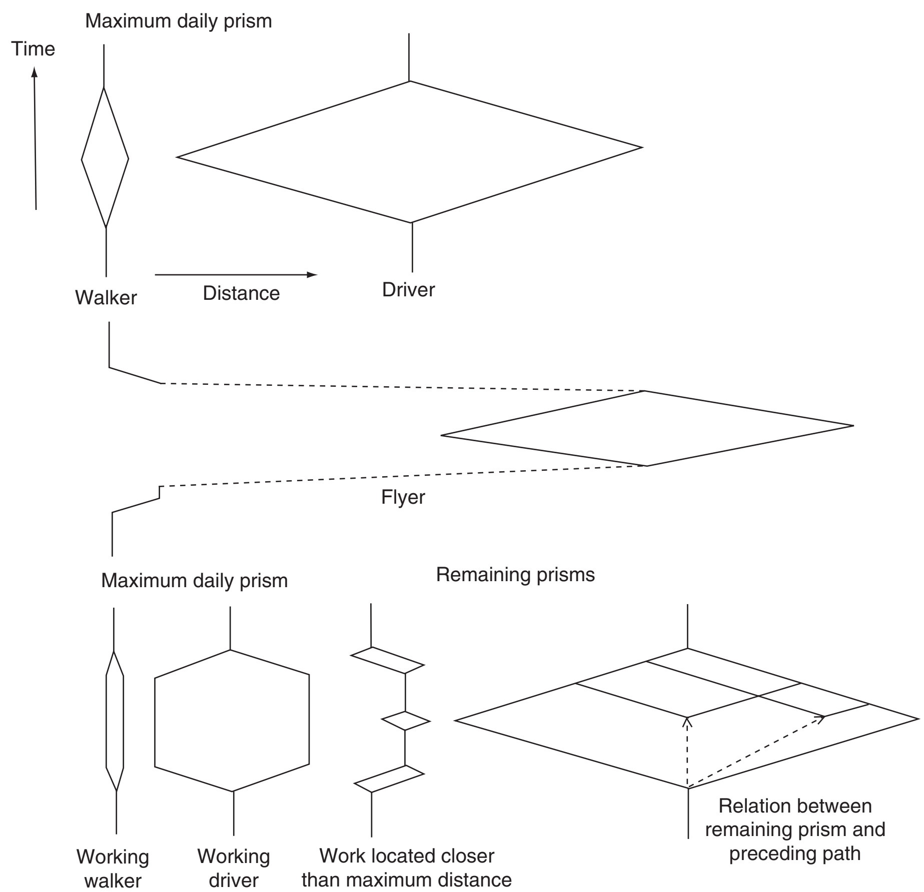

Figure 3 Daily prisms. From Hägerstrand, T. (1970). What about people in regional science? *Papers of the Regional Science Association* 24, 7 21.

in 1982, he also advocated the idea of diorama as a way of depicting the many different activities going on in a place, because it allows analysis of the interactions be tween the activities. He illustrates his point with a de tailed discussion of the area he grew up in, including memories of his own childhood activities.

In an autobiographical essay, Hägerstrand in 1983 talked of his "lifelong marriage to geography" although he also admitted to "a state of dissatisfaction with my chosen field of learning." He liked geography partly be cause it has "very few limiting fences in its structure and content." His work displayed linkages to many different approaches to human geography, not just the spatial an alysis approach taken in his spatial diffusion work. In much of Hägerstrand's later work, he discussed some wider issues – for example, the relationship of people and the landscapes they inhabit. He consistently stressed the notion of people as purposeful actors who are trying to carry out some project within the context of the human and natural environments. He also acknowledged the impact that environmental factors may have on people, based on his own experiences growing up in rural Swe den, and was willing to recognize the importance of the regional concept as a way of synthesizing individual biographies (though he was unimpressed with the idea of areal differentiation).

Like Peter Haggett and Peter Gould, Hägerstrand de ployed a strong graphic imagination. Throughout his work, maps and diagrams were employed to illustrate the points he was making, and many of his graphics represent original ways of conveying ideas to the reader. This is particularly so for his time geography work, where the field as a whole has owed a great debt to Hägerstrand's development of novel and effective diagrams which foster new ways of thinking about movement through time and space.

In summarizing Hägerstrand's geographies, Anne Buttimer in 1983 drew attention to the breadth of Hägerstrand's encounters with different traditions of scholarship, arguing he was a scholar who saw "no in trinsic conflict between art and science, between poetry and mathematics, who can be just as thrilled by the aesthetics of landscape and music as by the rigor and power of scientific models." Despite the innovativeness of his work on diffusion and on time geography, Häger strand's basic interests were perhaps more traditional: he claimed his ultimate aim was to try to turn human geography into a study of the conditions of life in a re gional setting.

Hägerstrand's reputation was thus based originally on his painstaking and ground breaking historical study in which he plotted individual migrations for the settlement of Asby, but his subsequent innovative use of cartographic and probabilistic techniques was influential in en couraging quantitatively inclined human geographers around the world. Hägerstrand's use of an azimuthal logarithmic projection centered on Asby (see **Figure 4**) not only conveyed a striking image of Asby's relationship to the rest of the world, but also encouraged the in novative use of cartography to illustrate patterns of spatial interaction. Peter Haggett hence described

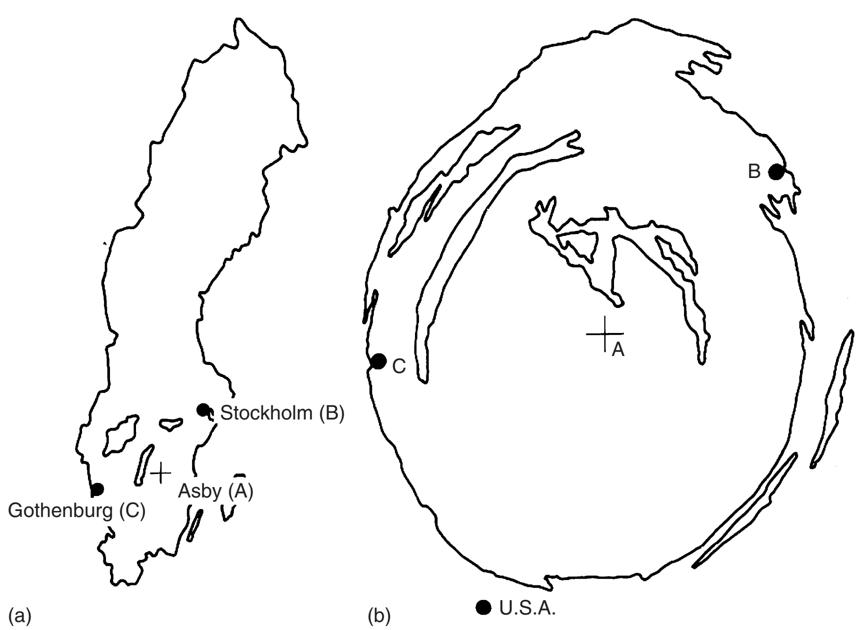

Figure 4 (a) Conventional map of Sweden. (b) Transformed to an azimuth, logarithmic-distance map based on Asby. Hägerstrand, T. (1957). Migration and area. In Hannerberg, D., Hägerstrand, T. & Odeving, B. (eds.) *Migration in Sweden: A Symposium.* Lund Studies in Geography, Series B, No. 13.

Hägerstrand as "the geographer who particularly influ enced my own research."

The legacy of Hägerstrand's work on the diffusion of innovations was a major rise in interest in the subject and its inclusion in human geography courses and textbooks. It was developed through the application of diffusion models to other topics, such as the growth of the Seattle ghetto (by Richard Morrill) and studies of the spread of contagious diseases (especially by Peter Haggett). However, some (notably Derek Gregory) criticized Hägerstrand's approach to diffusion for ignoring struc tural constraints to the adoption of innovations, such as social class, ethnicity, and gender, and for failing to in vestigate the processes by which individuals decide whether to adopt on the basis of their social context rather than just their location.

Hägerstrand's ideas on time geography have been followed up by his students and colleagues, including Bo Lenntorp, Solveig Martensson, Tommy Carlstein, and Lars Olof Olander, a group referred to as the Lund School. Their work has been taken up and applied by other geographers, including Australian geographer Don Parkes and British geographer Nigel Thrift. It has been influential too in sociology, where Anthony Giddens was to prove an enthusiastic advocate: one of the rare ex amples where ideas developed in geography have been taken on board by other social sciences. Yet Hägerstrand's time geography has also had its critics, mainly on the basis that it has little to say about attitudes, motivations, and values, and that it tends to treat people as un differentiated units, shorn of their cultural identities. Hägerstrand himself has admitted the validity of some of these arguments, recognizing that his work neglected the internal realms of experience and meaning.

See also: Diffusion; Innovation; Migration; Simulation; Time Geography.

Further Reading

Giddens, A. (1984). *The Constitution of Society: Outline of the Theory of Structuration*. Berkeley, CA: University of California Press.

Gregory, D. (1985). Suspended animation: The stasis of diffusion theory. In Gregory, D. & Urry, J. (eds.) *Social Relations and Spatial Structures*, pp 296 336. London: Macmillan.

Hagerstrand, T. (1957). Migration and area. In Hannerberg, D., Hagerstrand, T. & Odeving, B. (eds.) *Migration in Sweden: A Symposium*. Lund Studies in Geography, Series B, No. 13.

Hagerstrand, T. (1967). *Innovation Diffusion as a Spatial Process* (translated by Pred, A.). Chicago: University of Chicago Press.

Hagerstrand, T. (1970). What about people in regional science? *Papers of the Regional Science Association* 24, 7 21.

Hagerstrand, T. (1976). Geography and the study of interaction between society and nature. *Geoforum* 7, 329 334.

Hagerstrand, T. (1978). Survival and arena: On the life history of individuals in relation to their geographical environment. In Carlstein, T., Parkes, D. & Thrift, N. (eds.) *Human Activity and Time Geography*, pp 122 145. London: Arnold.

Hagerstrand, T. (1982). Diorama, path and project. *Tijdschrift voor Economische en sociale geografie* 73, 323 339.

Hagerstrand, T. (1983). In search for the sources of concepts. In Buttimer, A. (ed.) *The Practice of Geography*, pp 238 256. London: Longman.

Hagerstrand, T. (1995). Action in the everyday physical environment. In Cliff, A. D., Gould, P. R., Hoare, A. G. & Thrift, N. J. (eds.) *Diffusing Geography: Essays for Peter Haggett*, pp 35 45. Oxford: Blackwell.

Haggett, P. (2000). *The Geographical Structure of Epidemics*. Oxford: Clarendon Press.

Morrill, R. (1965). The negro ghetto: Problems and alternatives. *Geographical Review* 55, 339 361.

Haggett, P.

R. Flowerdew, University of St Andrews, St Andrews, UK

Haggett, Peter (1933–)

Haggett was born in 1933 and grew up in rural Somerset before going to St. Catherine's College, Cambridge. He graduated in 1954 and went on to a PhD at Cambridge, completed in 1969. His thesis was concerned with forestry in southeastern Brazil, and his subsequent publications show that he was already applying quantitative techniques as varied as central place theory and trend surface analysis to his Brazilian data. He taught at University College London for a brief time, followed by a junior post back at Cambridge, moving up to a lectureship there in 1962. In 1966, he became a Professor of Urban and Regional Geography at the University of Bristol, joining David Harvey to create what Jeremy Whitehand later termed the 'Cambridge–Bristol axis' of quantitative geography in the late 1960s. He has stayed at Bristol ever since, serving as acting vice chancellor in 1984 and 1985. As well as receiving many notable honors (e.g., *Prix International de Geographie* in 1991; gold medal of the Royal Geographical Society in 1986), he is one of the few professional geographers to have been awarded the title Commander of the Order of the British Empire (CBE) in 1993 for services to geography. Since retirement, he has maintained his activity in geographical research and publication, retaining the title of Emeritus Professor at Bristol (**Figure 1**).

Figure 1 Peter Haggett (1933).

Human geography at the start of Haggett's career was, in the United Kingdom, a low status discipline which kept itself to itself and learnt little from other subjects. It was still recovering from the determinism/possibilism debate and still thought of itself as having to focus on how environment affected human activity. Most human geographers were not prepared to cope with generalization beyond the descriptive. While geographers did use numbers to firm up population counts or measures of energy production, few were able to test relationships for significance. In essence, it was an atheoretical discipline concentrating on description, with little serious attempt to measure the patterns observed and even less attempt to explain them. There were exceptions, of course, such as Garrison and his students at the University of Washington and Hägerstrand in Sweden, and Haggett avidly read their work and fitted it into his own framework.

Haggett was one of the geographers most responsible for turning human geography around (perhaps the most important single influence). The changes concern the nature of human geography, its methods, its theories, and even its subject matter. They occurred, to a considerable degree, through Haggett's attempts to restyle geography as a modern, outward looking, scientific, and theoretically informed subject. This involved several distinct changes. Haggett argued for a more open outlook toward other disciplines, recognizing where they could provide useful material for geographers. His focus on spatial patterns and processes led some of his followers to define geography as a spatial science, though Haggett himself defines geography more broadly as "the study of the Earth's surface as the space within which the human population live." Regarding human geography as a science, or even as a social science, brings up the question of scientific method. Haggett was interested in the role of theory, models, and paradigms, introducing geographers to these concepts in the first chapter of *Models in Geography*. Geography had traditionally been idiographic, concerned with individual cases; he argued that it should be nomothetic, concerned with generalization.

Given his role in effecting geography's 'theoretical turn', it is hard to overestimate Peter Haggett's contribution to human geography from the 1960s onward, both in Britain and internationally. One of the first signs of his dissatisfaction with established descriptive and idiographic modes of enquiry was the series of symposia he organized with the physical geographer Richard Chorley at Madingley Hall, outside Cambridge, resulting in the very influential coedited *Frontiers in Geographical Teaching* and the even more influential *Models in Geography*. Both books consisted of a series of chapters discussing new approaches in the different branches of geography, many of them written by the editors and their Cambridge colleagues. Not all the chapters in these books repre sented major departures from the established ways of studying the subject, but most – especially Haggett's own – were influenced by a new and powerful set of ideas taking root in Anglo American geography, which sub sequently became known as the 'quantitative revolution'.

The term quantitative revolution, however, does scant justice to the changes that Haggett and his colleagues tried to bring about (and largely succeeded) in the way geography was studied. Certainly geographers made in creasing use of quantitative methods, measurements, and computers, rather than rely on verbal descriptions, but the changes went way beyond this. The use of statistical analysis, largely unknown in geography beforehand, be came more and more important. Geographers began to think in terms of hypotheses and how to test them, many adopting the hypothetico deductive approach regarded as typifying the scientific method. Geography also de veloped the idea of constructing theory, sometimes based on working out the consequences of a set of assumptions. Haggett did much to popularize the work of von Thunen and Christaller, little known previously in geography but highly relevant to Haggett's concepts of spatial analysis as applied to agricultural land use and urban location, respectively.

Practically, one of the most important aspects of his work was to align human geography with the social sci ences rather than the arts, and to increase its respectability as an intellectual discipline. He also related changes in the discipline to Kuhn's concept of paradigms and paradigm shift. Most fundamentally of all, he introduced ideas, concepts, and readings from a wide range of other dis ciplines, starting the outward looking tradition that is ar guably one of geography's greatest strengths today.

For example, another very influential book, *Locational Analysis in Human Geography*, developed a completely new approach to human geography, organized not by region or by systematic subfield, but by geometric form, typified by the chapter headings – 'Movement', 'Networks', 'Nodes', 'Hierarchies', and 'Surfaces'. It introduced to geography a whole range of new topics and new methods, including the work of a few hitherto little known geographers, such as Christaller and Hägerstrand, and many economists, such as Weber, Losch, and Isard, together with engineers, systems theorists, and philosophers of science. An ex panded second edition was produced in 1977.

Network Analysis in Geography, another publication with Chorley, discussed networks in both human and physical geography. Types of networks studied included transport networks and river systems, though the book used a classification based on geometric form, similar to the one in *Locational Analysis*. Part of the agenda for this book was an attempt to find a common ground between human and physical geography. At a time when it was no longer viable to argue that the natural environment controlled the human response, Haggett hoped that the use of similar spatial structures and models would give human and physical geography some common methodological concerns.

In all these books, Haggett adopted an abstract ap proach, using mathematical, statistical, and cartographic methods to tackle a wide range of examples from all over the world. His role as a contributing editor to *Geographical Analysis* and *Environment and Planning A* also allowed him to oversee and encourage the development of quantita tive and applied work in the discipline, while his role as co founder of *Progress in Geography* (later *Progress in Human Geography* and *Progress in Physical Geography*) underlines his enthusiasm to develop an integrative ap proach to geography.

In 1972, Haggett reached out to a new audience by publishing an introductory textbook *Geography: A Modern Synthesis*. Widely adopted in Britain, North America, and elsewhere (being translated into six languages, including Russian in 1979 – a rarity in Soviet times), it was less revolutionary than his other books, though some of the models discussed in *Locational Analysis* make another ap pearance. Haggett was keen to uphold the unity of geography at a time when many human geographers were turning their backs on physical geography (and vice versa). Even so, the physical geography in the book only extends to sections on 'The environmental challenge' and 'Man's ecological response', with no geomorphology. Many of Haggett's ideas filtered through to school geography, especially in Britain, through this text. His use of the beach as a focus for explaining many of what he regarded as basic geographical concepts gave the book a thought provoking start, although the attempt to introduce the G scale as a way of expressing a variety of spatial scales has been unsuccessful in the longer term.

In his later career, Haggett became interested in medical geography. He had long been interested in Hägerstrand's work on geographical diffusion, and in modeling economic changes at the local level, and he began to think about applying these ideas to the diffusion of infectious diseases. Much of his work on this topic has been done with colleagues, especially Andrew Cliff. On the basis of models of spatial diffusion, he developed new models appropriate for particular diseases and particular study areas, involving considerable conceptual and mathematical development. Much of his most interesting work has been based on the study of relatively small and isolated areas, such as Cornwall and Iceland. By careful choice of disease and study area, he was able to find out a great deal about the spread of infectious diseases, in cluding the ability to predict the return time and pro portion of population likely to be infected. With collaborators, he produced an *Atlas of Disease Distributions* and more recently *The International Atlas of AIDS*. He has also acted as a consultant to many scientific and medical bodies, including the World Health Organization. In *The Geographical Structure of Epidemics*, Haggett provides an excellent and well written discussion of the major themes of his work in this field. Although much of his work has established the spatiotemporal characteristics of epi demics in small and relatively closed geographical areas, he has developed an interest in processes operating at a global scale over a long time period. He has also studied ways in which the spread of epidemics can be controlled.

With Cliff *et al.*, Haggett also studied space–time diffusion in economic systems, such as spatiotemporal changes in unemployment. His work thus stressed the practical uses of such models, with his work on the de velopment and application of spatial forecasting proving particularly influential in this regard. His applied work on spatial forecasting models, epidemiology, and rural ser vices underlined the promise and potential of spatial scientific modeling in human geography. Nevertheless, he maintained interests in many branches of the subject; indeed, Chorley described him as "a quantitative, re gional, historical, and economic geographer with bio geographical interests."

Haggett has been a great innovator in British human geography. His work was one of the main foundations of the definition of geography as spatial science. His adop tion of the scientific method, later pilloried as an example of positivist reductionism, was also seminal for a gener ation of geographers in Britain and elsewhere. Both no tions were poorly received by establishment geography at the time (e.g., David Stoddart provides an account of the hostile reception given to Haggett's 1964 Royal Geo graphical Society lecture presenting a statistical model for the distribution of forest cover in Brazil). However, Haggett's ideas were ultimately to achieve acceptance, especially among the younger generation of scholars (an unusually numerous group because the British university system was expanding rapidly at the time). Peter Taylor latterly provided an interesting perspective on the quantitative revolution, arguing that the shift of paradigm was aided by the use of language which was difficult for an older generation to come to terms with (certainly true of statistical arguments) but not too difficult for the younger ones to master. Similar claims may subsequently be lodged with respect to Marxist, phenomenologist, and postmodern approaches to geography, all of which have been presented in language that quantifiers, now the older generation, have found difficult to follow.

How did Peter Haggett manage to transform his dis cipline so quickly and so effectively? Partly because the need for change was increasingly obvious, especially to younger geographers. But the speed and success of the quantitative revolution, particularly in Britain, owed much to Haggett personally. One contributory factor may have been the breadth of his reading. Reference lists for his books and journal articles in the 1960s are full of up to date material from several other academic subjects, showing how good he was at keeping up an intelligent interest in many areas and at working out how their theories, concepts, and methods could be applied to human geography. These reference lists might also yield treasures from little known geographical journals and working papers, which Haggett found, explained, and popularized.

Second, Haggett is a great synthesizer. *Locational Anal ysis in Human Geography* is a brilliant synthesis of lots of other studies, from geography and elsewhere, all fitted into the geometric classification of movement, networks, nodes, hierarchies, and surfaces – so the whole becomes much more than the sum of the parts. *The Geographer's Art* is also a kind of synthesis, bringing together some of the same material organized in a more personal way. Of course, his textbook presents a synthesis of geography, whether 'modern' (in earlier editions) or 'global' (in later editions).

Third, Haggett is an excellent communicator, both on paper and in the lecture room. Throughout his work, he has managed to make complicated or technical material accessible to students and colleagues. He makes extensive and effective use of diagrams and maps, which do not just illustrate his argument but also are a massive aid to understanding. As a speaker, he also makes copious use of visual aids. With or without their help, he has the ability to make difficult concepts and analysis crystal clear. His own enthusiasm for his subject was transmitted to gen erations of students.

Lastly, Haggett's own personality may have been relevant. He is quiet and unassuming; at conference dinners, he is as happy chatting to unknown graduate students as much as to senior professors. Although early presentations of his quantitative work annoyed and threatened the traditionalists in his audience, he stuck to his guns and persevered until he had convinced a critical mass of colleagues and students. His calm, clear, and nonconfrontational manner must have helped in gaining acceptance or at least tolerance for his work. Even those who disliked his work must have found it hard to dislike the man.

Yet, while positivism and spatial science dominated Anglo American human geography in the 1970s, criti cisms of positivism and quantification began to emerge before too long. Such criticisms came from two main quarters. First, an increasingly vocal group of Marxist geographers, augmented by some of Haggett's earlier followers, such as David Harvey, Allen Scott, and Peter Taylor, gained impetus from the wider critique of capitalism, the military–industrial complex, and the Vietnam War. Their belief that unobservable structures provide the basis of social and spatial formations, and their dedication to dialectic reasoning rather than scientific method, illustrates that they were significantly opposed to the particular positivist tradition that Haggett had brought into geography.

A second group of humanistic geographers also reacted against the abstract models Haggett had made fashionable in geography, stressing the importance of artistic, cultural, and historical insights into the nature of place, showing up the limitations of idealized models (such as the notion of 'economic man') which underpinned some attempts at modeling in human geography (e.g., in housing models, retailing catchment analyses, and migration analyses).

The definition of geography as spatial science also came under attack at the same period. One element of this attack was from geographers whose interests did not lie in this area. These included people interested in the traditional geographical concern of the relationships between people and environment, and others interested in development studies grounded in specific places and their experiences. For some, the concerns of spatial science seemed unimportant compared to other disciplines. More philosophically grounded objections appeared, such as Sack's rejection of the idea that space could be a subject of study in itself. Both Marxists and humanists, not to mention feminists and postmodernists, tended to reject or at least play down the spatial perspective, either because it was unimportant in their view to the major social trends or because abstract models of spatial structure did not apply to the real people and events they were interested in.

Haggett's ideas, however, are far from dead. Geography has opened itself up to other subjects and is much the better for it. Thus, Haggett may have managed to get us to engage with work in economics and operations research, while other geographers introduce us to themes in history, philosophy, and literary theory. There have also been cases where work in geography has itself had a major impact on other subjects, such as the interest across the social sciences in David Harvey's work and Giddens' interest in Hägerstrand's time geography.

Despite the unfashionability of modeling, quantification, and positivism for the last two decades, work of this kind continues to flourish, arguably being returned to prominence by the growth of Geographic Information Systems (GIS), geodemographics, and spatial computation in the 1990s. Similarly, locational analysis remains as an important interest in contemporary human geography, if only because there are many applications to commercial and public location problems. The role of space as a concept in geography has also revived in recent years, based on the attention given to how space is constructed by social groups, and how space affects or mediates social relations and processes. Haggett himself, while unfailingly polite and respectful of the work done under new paradigms, remains consistent in his interests and approaches, and continues to be an enthusiastic and effective advocate of his ideas.

Given the range and scope of Haggett's academic and professional achievements, a short article like this can do little more than give a flavor of Haggett's contributions. Fortunately, two books have appeared which give far more information about his biography and geographical concerns. Haggett himself produced a memoir, *The Geographer's Art*, intended to convey his approach to, and passion for, the subject. A few years later, Cliff *et al.*, produced a *Festschrift*, which includes papers by leading geographers on most of Haggett's major interests, and more personal memoirs by Chorley and Thrift. Both testify to the range of Haggett's contributions to the discipline, his role as leader and mentor, and his enduring legacy as a leader of geography's 'quantitative revolution'.

See also: Quantitative Methodologies; Quantitative Revolution; Spatial Science.

Further Reading

Chorley, R. J. (1995). Haggett's Cambridge: 1957 1966. In Cliff, A. D., Gould, P. R., Hoare, A. G. & Thrift, N. J. (eds.) *Diffusing Geography: Essays for Peter Haggett*, pp 355 374. Oxford: Blackwell.

Chorley, R. J. and Haggett, P. (eds.) (1965). *Frontiers in Geographical Teaching*. London: Methuen.

Chorley, R. J. and Haggett, P. (eds.) (1967). *Models in Geography*. London: Methuen.

Cliff, A. D., Gould, P. R., Hoare, A. G. and Thrift, N. J. (eds.) (1995). *Diffusing Geography: Essays for Peter Haggett*. Oxford: Blackwell.

Haggett, P. (1965). *Locational Analysis in Human Geography*. London: Arnold.

Haggett, P. (1990). *The Geographer's Art*. Oxford: Blackwell.

Haggett, P. (2000). *The Geographical Structure of Epidemics*. Oxford: Clarendon.

Haggett, P. (2001). *Geography: A Global Synthesis*. Harlow: Prentice Hall.

Haggett, P. and Chorley, R. J. (1969). *Network Analysis in Geography*. London: Arnold.

Thrift, N. J. (1995). Peter Haggett's life in geography. In Cliff, A. D., Gould, P. R., Hoare, A. G. & Thrift, N. J. (eds.) *Diffusing Geography: Essays for Peter Haggett*, pp 375 395. Oxford: Blackwell.

Haptic or Touch-Based Knowledge

R. D. Jacobson, University of Calgary, Calgary, AB, Canada

Glossary

Cognition The decision-making process of interpreting sensory information that involves remembering, recognition, association, and other high-order mental processes.

Cutaneous Pertaining to the skin itself or the skin as a sensory organ, including sensations of pressure, temperature, and pain.

Force Feedback The mechanical generation of haptic information used and sensed by the human kinesthic system, specifically in human–computer interaction.

Haptics Relating to the sense of touch, in all forms and in its widest context.

Kinesthesia It is the sense of movement through space, relating to the sensations originating in muscles, tendons, and joints.

Perception The reception of information as mediated by the senses, facilitating a relationship with the world, and subsequent decision-making processes (cognition).

Proprioception It is the awareness of our body's position in space, sensory information about the state of the body, including cutaneous, vestibular, and kinesthic sensations.

Tactile Relating to touch and the cutaneous skin sense, more specifically pressure rather than temperature or pain.

Tactile Map A cartographic representation that is designed and constructed to be read via the sense of touch, requiring the display to be elevated in the third dimension.

Introduction

Haptics is a term relating to touch, and active touch in its widest context, and how we are able to gain information about objects by manipulating them. Haptic perception involves the sensing of the movement and position of joints, limbs, and fingers, and also the sensing of information through the skin. The word 'haptics' is derived from the Greek term *haptikos*, from *haptesthai*, meaning 'to grasp, touch, or perceive', which is equivalent to hap(tein) to grasp, sense, or perceive.

Touch is generally regarded as the most proximal of all senses available for acquiring geographic knowledge, the only exception being the sensation of taste. Touch is a binary sense, in the manner that you are either in contact with an object or not, on or off. However, it is possible to feel a local spatial extension to this requirement for contact, to feel radiant heat, or the deep bass of heavy rock music. Just as there are extensions to other senses, to enhance them, for example, the telescope for vision, there are similar analogies of an extension in touch based sensation. These would include feeling the surface of paper through a pencil, or the undulations of a sidewalk via pushing a child in a stroller, and the proxy sensations of the road surface experienced when driving. Representations of geographic information can also be presented in analog and digital interfaces to be explored via touch.

Touch can be regarded as a mechanism to gaining geospatial knowledge about the environment from proximal information, things that we are able to touch directly, via these local extensions to touch, and through models and representations of geographic space that are designed to be experienced and interacted with through touch. This would include tactile maps, three dimensional models, and touch based representations in virtual reconstructions that can be interacted with through touch and through force feedback in haptic based virtual reality simulations.

Touch is unique among the senses and that is both an input and output modality. We are able to receive information such as heat and pressure, and also to communicate information. Although the spatial realm of touch may appear to be small, confined to areas around the body, it is a key sensory modality we use when generating spatial knowledge. To paraphrase the philosopher Bertrand Russell, it is touch that provides us with our sense of reality, our whole conception of what exists outside us, is based upon our sense of touch.

Touch is used in environmental navigation; we are able to make use of multiple kinds of touch based information, involved in walking, surface texture underfoot, and the sensations of inclines when walking uphill. Our stride changes in response to the surface of the environment, when it is icy, or snow covered, soft, or hard. Touch is a particularly resilient sense; in spite of its central importance, it is often taken for granted due to being superseded by vision, its role is often unnoticed. Although recently, touch has been used to perceive or represent reality, which previously was thought to be only accessible via vision or audition, through technological innovation in presenting information via vibration, facilitating reading, speech signals, pictures, maps, and music.

The Role of Touch-Based Knowledge in Geography

We all generate internal constructs of our everyday and foreign geographic environments through a multitude of information sources. The role of touch with regard to generating geographic understanding requires a broad overview with details from physiology to matters of communication, cognition, and representation. Our perception of the world is mediated through cognition, prior experience, and cultural constraints. This internal construct is a mix of location based information, and the beliefs, attitudes, thoughts, and feelings associated with those locations. Sight is a privileged sense in gaining this information. Perception occurs only when a chain of events is complete. Seeing starts with an object in an environment and runs through the actions of an observer in the environment, and the object providing light to the retina reflected from a few luminous sources in the environment. With touch, the object provides resistance to the senses under manipulation, such as compression. The actions of the observer, in the case of vision, include adjustments of the eye and changes in the vantage point to gain more information about the object. The actions of the observer, in the case of touch, include adjustments of the hands to explore the object tactually. Touch also integrates movement as an exploration strategy. As the individual continues to explore the object, gaining more information about the object, and through movement between the object and the observers' actions, come many sensory links to complete this chain of events.

Geographic Learning

The senses are geographical in the way that they facilitate and contribute to an understanding and orientation in space, either from direct environmental perception or from representations such as maps and models. The senses provide a common framework through which space and time are experienced and understood, that is, made sense of. Each sensory modality and combination of modalities have their own geographic scale range, and also the multisensory nature of interacting within a geographic environment means that experiences vary not only across space and time, between different individuals and groups, but also between cultures. Certain contexts may then lead to the role of the sense and sensation in culturally defining and establishing geographic meaning.

Haptic relationships are established between an individual and one's environment. The relationship alters as distance is introduced, moving from communication in the most proximal to exploration, simple contact, contact via a proxy, the vibrations of a car to sense the underlying road surface, and then to separation. Much geographic learning takes place through representations of environments, maps, and models; here, touch is able to mediate information that we could not normally grasp, such as tactile maps of the world, just as vision is used to represent information beyond our immediate sensory scope.

Geographic Knowledge Construction

Acquiring information from the environment or a representation such as a map vision is the spatial sense par excellence, yet touch has a pivotal role in formulating our understanding of the world. Cognitive scientists, linguists, and developmental psychologists have stressed the central and important foundations that touch and object manipulation play a role in facilitating our understanding of complex spatial arrangements. This early manipulation allows us to generate the cognitive foundations that are later described in language, and formally in logic and Boolean mathematics. This understanding is at the core of data manipulation in geographic information systems (GISs). The importance of touch and its centrality to our ways of thinking are illustrated in language: 'having a good grasp' of the subject, or 'having a good feel' for a complex dataset, we are able 'to hold an idea' hint at the power of our haptic senses for comprehending the real world. They feed our consciousness and subconscious in ways that are deep and not fully understood.

Key Components of Touch

Touch is superior in some aspects of perception and functions very differently from the visual system. It is able to discern very fine surface texture changes, grit in surfaces, hardness or softness, and many of these functions are complimented via hearing or seeing information about the same objects. The potential properties of the surroundings accessed through touch do not shout at us like colors in the visual system, they are muted until we interact with them. The information gained via touch, such as viscosity, slipperiness, softness, texture, and elasticity, are so complex that they defy any ordinary language categorizing schemes.

Touching is a combination of information gained from the tactile sense through the skin (cutaneous), through movement (kinesthesia), and an awareness of our body's position in space (proprioception). During passive touch, neither the skin nor the object moves; at other times, we remain stationary as objects move past us, and during active touch an observer moves and manipulates an object to gain additional insight into the properties of an object. There is intentionality in movement; different strategies are adopted, such as scanning left to right, then top to bottom, when investigating a tactile map. With new technology, objects can be moved with fingers, or guided to a location in a touch interface.

A classic illustration of the efficacy of touch is the Braille system, originally developed as a system for military communication in darkness, now internationally adopted as a touch based reading mechanism for blind individuals. Raised dots in a 2×3 array are used to represent the alphabet and contractions of words.

As objects are explored, we are able to explore properties through examination of the surface, the object volume, and immersion inside the object. Through this, we able to obtain information on geometry, material, location, energy, texture, spatial location, shape, extent, orientation, curvature, proportions, projections, size, symmetry, complexity, and ridges.

Social and Cultural Constructs of Touch

Touch has a significant affective and sensuous component, such as pain and pleasure; we are able to readily feel the important attributes of an object, such as hot or cold, soft or hard, rough or smooth, and many spatial attributes. Components of touch are filtered via social and cultural processes, affecting our interpretations at the fundamental levels of cognition and perception. The legitimacy and importance we ascribe to this information is too culturally mediated, as are societal norms of tactile interaction, such as shaking hands, kissing on the cheek, when greeting, and notions of necessary and acceptable personal space.

Touch is mediated by psychophysics – there are certain properties, thresholds, distances, beyond which we cannot sense, individual differences in our perceptual abilities (psychophysics), and of our memory and cognitive interpretations. There is a negotiated reality where an individual has a common or similar sensory representations of the world that we all share, such as social and cultural norms, of acceptable differences, and sensations. Perception is shared, social, and our senses or responses to them may be learnt and provide a sensory filter to reality. Seeing may be believing, but feeling provides the truth. We are able to get to know things in the world through touch, rather than encountering them at a distance through an abstracted visual system.

How Touch Relates to Other Sensory Modalities

The mechanisms by which touch relates to and mediates other senses are not entirely clear. Due to the ocularcentric bias in the making and framing of knowledge through science, where 'seeing' is equated with 'knowing' touch is often viewed as more primitive and less important than vision. Historically, touch was at times regarded as dominant over vision and audition, the role of sculptors in art and the recreation of scenes, for example. Touch remains dominant in the proof of the existence of an object, for example, a dream, mirage, or hallucination is often sought validation via touch. During sensory illusions, touch is often used to calibrate and verify other senses. While rowing, as an oar enters the water the refraction of light leads to the appearance of a bent oar, yet touch confirms that it is straight.

Multisensory Integration

There has been a long historical debate as to relationships between different sensory modalities. Are we able to obtain the same information about geographic objects and events from vision and via touch? How does the nature of the information differ? Once we have received the information through perception and processed it through cognition, does it reside in modality specific memory that can be translated to different senses for communication? Or are originally separate, modality specific pieces of information retained in a central amodal core, independent of their sensory derivation?

There are several frameworks proposed as to how the senses are able to combine, interact, and integrate. These are becoming increasingly important with the advent of novel technologies to move information from one sensory method of representation to another in multimodal computer interfaces. If sensory modalities are able to cooperate, here information is combined to be greater than the sum of its parts, substituted from one sense to another, augmented, and conflicts between the senses resolved. In a hierarchical framework, not all senses are equal, vision often dominates, for example. However, this varies by task, object manipulation through touch may exceed the information provided by visually surveying the object. Sequencing or an ordering of the senses offers an alternative framework, developmentally; in early infancy, the importance of sensory information is ordered as tactile, auditory, and then visual; by adulthood, this sequence is reversed. Clearly, there are thresholds not only of stimulation for different senses based upon the psychophysics of the sense, but also of individual differences, habituation, and the meaning and importance we ascribe, for example, esthetically to the sensory information.

Certain types of information from the geographic environment are sensory specific, listening to ambient sounds, traffic noise, and smelling pollution. When translating from one medium or modality to another, there is a degree of information loss (and selection classification, etc., akin to cartography). The information gained from viewing a scene is different from using spatial language to describe the scene.

Touch-Based Representations

Multiple geographic representations are accessed through touch. There are touch based (tactile and relief) maps designed to be read by touch, and complex virtual displays of geographic information where haptic interfaces are used to communicate spatial information. In conventional human–computer interactions with geospatial information, in a GIS, for example, employing the windows, icons, menus, and pointers (WIMPs) interface touch remains a pivotal mechanism through our manipulation of keyboards and computer mice.

When touch is used in the absence of vision, there are notable differences in gathering information from a map like display. Typically, touch based information is ob tained sequentially, with a more limited 'field of view'. Vision is able to provide a gestalt like view of infor mation, allowing the perception of the synoptic whole and almost simultaneously its constituent parts. In static, analog, touch based systems such as a tactile map, while the fingertips provide a good sensitivity to information, its shape, and spatial context in the scene, it remains a challenge to communicate contextual information. A tactile map reader may well be able to discern the outline shape of a circle, however, the context may remain un clear. Is the circle a globe, a symbol for a city, or a geometry diagram?

Cartographers have developed mapping methods that incorporate single as well as multisensory input. Most commonly, the sensory input channels include visual, auditory, and haptic. Tactile maps are representations of space and are designed to be interpreted through touch. Tactile maps may be produced for the single sensory input method (touch) or may include a combination: tactile/visual such as a raised relief map, or tactile/audio and touch augmented with sound and speech.

Tangible Models

These are three dimensional models used for repre senting geographic environments. Simple versions in clude those using *papier mache*, blocks, cardboard, and so forth to represent buildings, educational campuses, cities, and other spatial concepts. Complex three dimensional models are generated to illustrate the terrain of an area and other geographic processes such as geo logic activity.

Tactile Maps

Tactile maps are the geographer's traditional way of providing information to blind or visually impaired people and to all those who may have visual difficulties. The map is designed and constructed to be read by the sense of touch, requiring portions be elevated in the third dimension. Tactile maps and graphics are produced in a variety of forms. Reproducible tactile maps may be cre ated from molds, or by a pressing procedure using ma chines that might otherwise be used to make molds or metal plates used in the printing industry. Computer production of tactile map has also made it possible to design, produce, and reproduce tactile maps. Maps can be printed directly from the computer through specially designed printers, such as an embosser or an inkjet that deposits polymer inks to controlled and varying heights onto a plastic substrate.

Audio-Tactile Maps

A combined tactile/auditory representation of geo graphic space is created by overlaying a tactile repre sentation on a touch sensitive pad attached to a personal computer (PC) or laptop.

For any particular tactile graphic, key information about places on the graphic were located by keyboard input to the grid location at which it could be found when the tactile representation was placed on the touchpad. This facilitates access to tactile maps and graphics in a way not previously possible. Through fur ther manipulation, spatial relations such as distance, direction, connectivity, and configuration could be ob tained by exploring the totality of the constructed system. This methodology allows an individual to get a direct touching experience of the phenomena being explored, while simultaneously accessing localized auditory infor mation tied to specific locations and activated when these locations are touched.

Virtual and Haptic Interfaces to Geographic Information

In human–computer interaction, a haptic device gives people a sense of touch with computer generated environments, so that when virtual objects are touched, they seem real and tangible. An example is a medical training simulator in which a user can feel a needle push through virtual tissue. Haptics is applicable across nearly all areas of computing, including video games, medical training, scientific and geographic visualization, engin eering design, remote vehicle and robot control, art, medical rehabilitation, and interfaces for the blind.

The majority of haptic interfaces utilize force feed back, a user can feel forces applied to a user's body by the movements of a haptic device, sensed by the user pri marily through musculoskeletal forces, and also through the skin that touches the physical interface to a haptic device. This is often accomplished through a user's hand grasping a handle connected within the device to motors that generate the force and 2D haptic devices such as force feedback steering wheels and force feedback joy sticks. Vibrating motors are applied within something that is held (e.g., a game controller or a force feedback mouse). Tactile feedback is used to apply forces directly to the skin to provide a sensation of texture, or a change in temperature.

Haptic devices have varying complexities, levels of engineering, fidelity and cost, and can move in different ways. Force feedback devices are often described by their degrees of freedom (DOF). A DOF refers to a direction of movement. Common DOFs include right–left move ment (*X*), up–down movement (*Y*), forward–backward movement (*Z*), roll (rotation about the Z axis), pitch (rotation about the X axis), and yaw (rotation about the Y axis).

The utility of haptic interfaces with geographic in formation has been explored to mitigate the effects of distance; remote users are able to interact; telepresence is used to manipulate objects at a remote or safe distance, such as nuclear operations, and undersea exploration. Several international groups of researchers are de veloping haptic interfaces to geographic information for blind and vision impaired users.

The development of multisensory interfaces of par allel and complimentary information provides and pro motes 'universal access' through combinations of speech, sound maps, and haptic (force feedback) computer based interfaces. This facilitates access for those without vision, who need ways of exploring spatial information through maps, diagrams, and graphs that are central to education; promotes novel ways of augmenting visual displays; en ables the exploring and mining of spatial data, and other hidden properties such as uncertainty; and enhances navigation through spatial information where vision is restricted, such as within vehicle navigation systems, and hazardous environments encountered while firefighting. This multisensory approach can be directed to develop solutions to provide access to digital, spatial based in formation for novice and expert users, for the young and old, for people with sensory impairments, and in situ ations where vision based displays may not be optimal, due to limited screen size on mobile telephones and portable data assistants.

Future of Haptic Geographies

The role of touch and haptics is increasingly important to human geography for its power to represent infor mation, connect people separated by distance, and im prove the quality of interaction and immersion with media used to represent geographic information. In sensory combinations in multimodal interfaces, the role of haptics continues to grow. It has expanded from in dustrial, military, medical and research labs to the desktop, and is now being widely disseminated in gaming systems and telephones.

Researchers have been striving to develop map and information systems that are accessible to a wide range of individuals, irrespective of their skill levels, age, or sensory impairments.

Geography has historically been predominantly visual in nature. With the emergence of haptic technologies, opportunities for exploring the role and importance of touch are increasing. The blending of the virtual and real world through representations, immersive scenarios, online communities, social networking, web mapping, and location based services are likely to lead to an in creased focus on the role of touch and presence in geography for many years to come.

See also: Intermodality; Landscape Perception; Learning Regions; Map Interactivity; Map Perception and Cognition; Mental Maps; Representation and Re-presentation; Simulation.

Further Reading

Burdea, G. C. (1996). *Force and Touch Feedback for Virtual Reality*. Chichester: Wiley.
Fisher, P. and Unwin, D. (eds.) (2002). *Virtual Reality in Geography*. London: Taylor and Francis.
Goldstein, E. B. (2007). *Sensation and Perception*. Belmont, CA: Thomson Wadsworth.
Golledge, R. G., Rice, M. and Jacobson, R. D. (2005). A commentary on the use of touch for accessing on screen spatial representations: The process of experiencing haptic maps and graphics. *The Professional Geographer* 57(3), 339 349.
Golledge, R. G., Rice, M. and Jacobson, R. D. (2006). Multimodal interfaces for representing and accessing geospatial information. In Rana, S. & Sharma, J. (eds.) *Frontiers of Geographic Information Technology*, pp 181 208. Berlin: Springer Verlag.
Hatwell, Y., Steri, A. and Gentaz, E. (eds.) (2003). *Touching for Knowing: Cognitive Psychology of Haptic Manual Perception*. Amsterdam: John Benjamins.
Heller, M. A. and Schiff, W. (1991). *The Psychology of Touch*. Hillsdale, NJ: Erlbaum.
Jacobson, R. D., Kitchin, R. M. and Golledge, R. G. (2002). Multimodal virtual reality for presenting geographic information. In Fisher, P. & Unwin, D. (eds.) *Virtual Reality in Geography*, pp 382 400. London: Taylor and Francis.
Katz, D. and Krueger, L. E. (1989). *The World of Touch*. Hillsdale, NJ: Erlbaum.
Kennedy, J. M. (1993). *Drawing and the Blind: Pictures to Touch*. New Haven, CT: Yale University Press.
Nathaniel, I. and Durlach, A. S. (eds.) (1995). *Virtual Reality: Scientific and Technological Challenges*. Washington, DC: National Academies Press.
Rice, M., Jacobson, R. D., Golledge, R. G. and Jones, D. (2005). Design considerations for haptic and auditory map interfaces. *Cartography and Geographic Information Science* 32(4), 381 391.
Rodaway, P. (1994). *Sensuous Geographies: Body, Sense and Place*. London: Routledge.
Schiff, W. and Foulke, E. (1982). *Tactual Perception: A Sourcebook*. Cambridge: Cambridge University Press.

Relevant Websites

http://www.ggy.bris.ac.uk
Haptic Spaces: How spatial cognition, the way we process and negotiate space, is mediated by touch, University of Bristol.
http://haptic.mech.northwestern.edu
Haptics Community website.

http://www.haptics e.org
Haptics e: The electronic journal of haptics research.
http://www.immerse.ucalgary.ca
Investigating Multimodal Representations of Spatial Environments (Immerse) Lab, Ongoing research and developments in touch and haptic interfaces to geographic information.
http://touchlab.mit.edu
Massachusetts Institute of Technology (MIT) touch lab.
http://www.cim.mcgill.ca
McGill University Haptics Lab.
http://www.sensable.com
Sensable Technologies: Haptic technology hardware and software.

Harley, J. B.

N. Collins-Kreiner, University of Haifa, Haifa, Israel

Glossary

Cartography It is the science that deals with the principles, construction, and use of maps (in Greek chartis=map and graphein=write).

Harley was one of the most influential and remarkable figures in the recent history of cartography. Harley began research in the Department of Geography at the University of Birmingham in 1956 under the late Professor Harry Thorpe. The subject (predictably, at that point in Thorpe's own work) was the historical geography of medieval Warwickshire, with its contrasting regions of Arden and Feldon. It was from this that he developed his interest in maps, mapmaking, the meaning of maps, and their cultural significance. Later on in his career he taught and conducted research at the Universities of Liverpool and Exeter where he was promoted to the Montefiore Readership in geography. In 1986 Harley moved to a chair in the University of Wisconsin at Milwaukee and became director of the American Geographical Society's map collection (Figure 1).

Brian Harley produced approximately 150 publications, embracing many aspects of mapping. But above all he is known for a series of important methodological and epistemological papers on the nature of maps. He co edited (with David Woodward) in 1987 and 1992 the first two volumes of *The History of Cartography* www.geography.wisc.edu/histcart, an unprecedented survey of the development of cartography both as a science and an art, which demonstrated the value of maps for societies worldwide.

However, perhaps the most famous of Harley's theoretical statements is "Deconstructing the Map" (*Cartographica* 26, no. 2, 1989: 1–20). It establishes modern maps as representations of life, with Harley claiming that studying the meaning of maps should go beyond the geographical and topographic features to excavate "an alternative meaning" which has been neglected in current literature, since maps are a product not only of "the laws of alternative order" but also of "the norms and values of social tradition and order" (Harley, 1992: 232). His research dealt with a new vocabulary in cartography and in the use of concepts which until then were rarely heard in this context, such as status, gender, race, ideology, power, knowledge, myths, images, and rites.

Harley's particular interest was thus to criticize modern academic cartography more than to reflect on cartographic history. A crucial component of this article is the issue of how a mapmaker might act ethically, given the power relations inherent in the profession. One of his objectives was to suggest that an alternative epistemology, rooted in social theory rather than in scientific positivism, was more required if geographers were to reestablish cartography within a critical human geography.

He therefore 'deconstructed' and critiqued many of the fundamental assumptions underpinning cartography, and reconceptualized cartography as implicated in processes of social and cultural reproduction. He also discussed the dualisms that have structured our understanding of the history of cartography, so that, for example, maps are described as either reflecting 'ground truth' or are dismissed as subjective, distorted, and biased. This 'discourse of opposites' is based on a series of assumed dualisms, such as art/science, inaccurate/accurate, and subjective/objective. Harley argued that the uncritical acceptance of such dualisms is detrimental to a viable history of cartography, and that the historical origins of such dualisms should be carefully examined. Uncritical reliance on the simple dualism of true or false, or art and science may thus limit our ability to make sense of what cartography has been and what it is becoming. This was necessary in order to sever the scientific connection between 'reality' and its presentation, which dominated cartographic thought, and to present a new epistemology in which maps are seen as socially constructed.

The field of the history of cartography has been transformed in the past decades, largely because of Harley's research. A map has been traditionally defined by geographers as implying a certain level of physical correspondence. A new definition was formulated by Harley: maps were no longer a "representation of things in space" but "graphic representations that facilitate a spatial understanding of things, concepts, conditions, processes, or events in the human world" (Harley and Woodward, 1992). This definition changes one's view of maps as objective representations of physical space to human documents with all their attendant biases and failings. Hence, such artifacts as diagrams of imaginary cosmographies, landscape paintings, and 'mental maps' may now be considered maps, amenable to deconstruction and critical interrogation.

Harley died when his career was still in its ascendancy but his publications will continue to be influential for many decades.

See also: Critical Cartography; Deconstruction; Historical Geography; Maps; Mapping, Philosophy.

Further Reading

Harley, J. B. (1986). *Imago Mundi*: The first fifty years and the next ten. *Cartographica* 23(3), 1 15.
Harley, J. B. (1987). Maps, knowledge, and power. In Cosgrove, D. & Daniels, S. (eds.) *The Iconography of the Landscape: Essays on the Symbolic Representation, Design and Use of Past Environments*, pp 277 312. Cambridge: Cambridge University Press.
Harley, J. B. (1988). Silences and secrecy: The hidden agenda of cartography in early modern Europe. *Imago Mundi* 40, 57 76.
Harley, J. B. (1989). Deconstructing the map. *Cartographica* 26(2), 1 20.
Harley, J. B. (1989). Historical geography and the cartographic illusion. *Journal of Historical Geography* 15, 80 91.
Harley, J. B. (1989). The myth of the divide: Art, science, and text in the history of cartography. Paper presented at the 13th International Conference on the History of Cartography, Amsterdam.
Harley, J. B. (1992). Deconstructing the map. In Barnes, T. J. & Duncan, J. S. (eds.) *Writing Worlds: Discourse, Text and Metaphor in the Representation of Landscape*, pp 231 247. London: P. Routledge.
Harley, J. B. (1997). Power and legitimation in the English geographical atlases of the eighteenth century. In Wolter, J. A. & Grim, R. E. (eds.) *Images of the World: The Atlas through History*, pp 161 204. New York: McGraw Hill for the Library of Congress.
Harley, J. B. and Woodward, D. (eds.) (1987). *The History of Cartography, Vol. 1: Cartography in Prehistoric, Ancient, and Medieval Europe and the Mediterranean.* Chicago, IL: University of Chicago Press.
Harley, J. B. and Woodward, D. (1989). Why cartography needs its history. *The American Cartographer* 16, 5 15.
Harley, J. B. and Woodward, D. (1992). *The History of Cartography, Vol. 2 (pt 1): Cartography in the Traditional Islamic and South Asian Societies*. Chicago, IL: University of Chicago Press.

Relevant Websites

http://www.geography.wisc.edu/histcart
The History of Cartography Project.

Hartshorne, R.

F. Harvey, University of Minnesota, Minneapolis, MN, USA

Hartshorne, Richard (1899–1992)

Richard Hartshorne was renowned as one of the key Anglo American geographers of the twentieth century. Following publication in 1939 of the classical *The Nature of Geography*, and having served as Chief of the geography division in the US Office of Strategic Services during World War II, Hartshorne acquired great influence on postwar American geography, especially through his significant impact on the development of Area Studies. His work hence became the prevalent reference for many geographers in the post–World War II era. His work remains an important milestone of twentieth century geography although widely ignored over the last 20 years.

Hartshorne graduated from high school in Phoenixville, Pennsylvania, moving to Princeton where he studied mathematics and entered into a graduate program in mathematics, graduating in 1920. After becoming interested in the ideas of Ellsworth Huntington (1876–1947) he contacted Huntington to get his advice on pursuing the study of geography. Huntington recommended studying at Columbia, Pennsylvania, or Chicago, which had viable geography programs at the time. In 1921 he entered the University of Chicago's geography program. At Chicago, his first advisor was Harlan Barrows, but the impact of Charles Colby and Wellington Jones were greater influence. With Wellington Jones his exchanges developed and Jones became his thesis advisor. His dissertation was titled 'The Lake Traffic of Chicago', for which he was awarded a doctorate (*magna cum laude*) in August 1924.

After completing his doctoral studies, he took a position at the University of Minnesota in September 1924 and remained there until 1940 when he moved to the University of Wisconsin, Madison, where he remained for the rest of his life. As a professor at both universities, Hartshorne sought to consolidate the discipline of geography in the changing academic landscape. Fulfilling his desire for significant and stimulating dialog in 1925, he took part in a field conference attended by William Morris Davis, Nevin Fenneman, Mark Jefferson, Ellsworth Huntington, Vernor Finch, and other geographers. He took on an early role in Midwestern geographic field conferences held between 1915 and 1940, both as a regular participant and organizer of field study projects and themes, and became a primary progenitor of the Midwestern School of Geography. Delighted with the vibrant exchanges with peers, he continued his statistically grounded economic geographic work and started to engage with ideas in political geography. His reading of Isiah Bowman and James Fairgrieve's work paved the way for his crucial engagement with German geographic literature.

With the support of the Social Science Research Council, he was able to pursue these interests and travel to Europe in 1931–32. This stay allowed him to complete research and publish political geography papers on German/Polish boundary issues as well as economic geographic papers examining upper Silesian industry and coal and iron mining. After returning to Minnesota, his interest in political geography is evident in publications on political geography for geographers and political scientists. Key works he published on political geography included 'Recent developments in political geography' (1935); 'Suggestions on the terminology of political boundaries' (1936); and 'The Polish corridor' (1937). These publications emphasized literature reviews of German research and were seen as more orientated to political scientists then geographers.

As his concerns with the perceived lack of theory and methods in geography grew, Hartshorne began to work on *The Nature of Geography* while on sabbatical in Europe in 1937 and 1938, returning with a draft manuscript. Using Harvard and Clark University libraries he concluded the work, which Derwent Whittlesey had reviewed by Russell Whitaker. Whittlesey proposed that the manuscript be published in the *Annals of the Association of American Geographers* to members of the AAG's council, which agreed. It was published in 1939.

The Nature of Geography is a complex work. It began as a response to Carl Sauer and John Leighly (both at the University of California, Berkeley) and grew rapidly to reflect Hartshorne's sense of uncertainty regarding the place of geography among the sciences. Carl Sauer had published his views, influenced significantly by the German von Richthofen's work with an emphasis on environmental change and natural process, which diverged from Hartshorne's chorological and anthropogenic understanding of geography. John Leighly's (1937) article claimed that no theory could be found to "bring into rational order all or a large fraction of the heterogeneous elements of the landscape." Hartshorne reportedly was also spurred on by Carl Sauer's purported remark, "if you can't see it, it isn't geography." This incendiary comment threatened to discredit the economic and political geography Hartshorne had learned at the

University of Chicago and developed in the Midwestern School's activities. Turning to German geographic lit erature, as Sauer had done, Hartshorne sought to refine and purify geographic theory and methods.

A few key points about Hartshorne's system of geo graphic thought are evident in the *Nature of Geography* and belie significant differences with the German geog raphers, especially Alfred Hettner:

1. Hartshorne accepted Kantian philosophy without any examination, hence with all of the Kantian assump tions; particularly that geography was seen solely as a preparatory (propadeutic) study. Hartshorne develops the chorological ordering of phenomena as the prime function of geography, which are only a small part of Kantian geography.
2. For Hartshorne, all geographic phenomena and rela tionships are causally connected.
3. The functional parts of a single whole (world) can be understood as a system, according to Hartshorne, based on the arrangement of objects in classes.
4. Hartshorne states that Ritter's first principle is: geography must be based on rational principles – from philosophy – or from *a priori* theories of general geography.

However, in the *Nature of Geography* Hettner and other German geographers are critical to Hartshorne as icons. John Agnew (1989) subsequently suggested Hartshorne himself used German 'authorities' (in particular, Hettner) to justify his arguments about the nature of geography. Through a recitation of German geographer's arguments, Hartshorne sought to quell a perceived methodological crisis and intellectual fragmentation in American geog raphy. The copious literature review presents a treatise on geography. Hartshorne presented *The Nature of Geog raphy* as both a distinctly American geographic contri bution and simultaneously as a re rendering of Hettner's publications and system of geography. As Hartshorne wrote in a letter to Hettner "My conviction that it [Nature of Geography] may be of major importance for the development of geographical thought in America is based simply on the fact that it is primarily an ex position–in considerable part by literal tradition–of your studies in this field" (letter sent 26 Nov 1938 from Hartshorne in Vienna to Hettner).

Hartshorne hence asserted the importance of de veloping laws and principles in sciences. Windelband and Rickert, German social scientists, distinguished between nomothetical and ideographic sciences, which, for Hartshorne, conformed approximately to the division between natural sciences and social sciences. Hartshorne saw a need for both in geography, but emphasizes the importance of elaborating general principles in geog raphy, thus, "… geography may be called a generalizing or nomothetical science."

Hartshorne accordingly stands out as a protagonist of the functionalist and utilitarian orientation of the Mid western School's geography. Hartshorne sought to go be yond the application of geography and formulate its essence to assure its place among the sciences. A key part of Hartshorne's strategy rested on his translation of German geographical discourse, particularly the work of Alfred Hettner (1859–1942); however, Hartshorne's pre sentation remains American in its orientation and out comes. Going to Austria and Germany in the 1930s, Hartshorne already had a distinct American process orientated understanding of landscape established through the works of William Morris Davis. The influence of Wellington D. Jones at the University of Chicago is also clear in *The Nature of Geography.*

Hartshorne was successful in his ambition to establish *The Nature of Geography* as the key theoretical and meth odological reference for a generation of American geographers. Hartshorne was later quite clear about the intentions of his publications. Reflecting on the motiv ations in 1959, he wrote that he intended to establish "logically sound structure of the nature of geography" and "determine the actual character of geographic work." The monolithic stature of this work became the center of a number of critiques, most trenchantly formulated by Neil Smith (1989) as the 'Museum' of geographic thought that was out of touch with later intellectual develop ments. As other critics noted, Hartshorne produced a de historicized history. As Bernie Bauer *et al*., write: "Sherman severely chides Hartshorne for his prescriptive proclamations about the nature of Geography, although many took comfort in his unshakeable vision. There is a tendency for people to listen to the purveyor of fine shoes and to sport the latest fashions" (1999: p. 776). The continuing challenge for understanding Hartshorne's geographical thought and impact on the geography re mains, however, relating the vision to the context in which it was developed and became so prevalent.

Hartshorne followed the publication of *The Nature of Geography* with a move to the University of Wisconsin, Madison in 1940, motivated by professional interests, conflicts with the chair of the department at Minnesota, and a prescription against hiring of family members at the University of Minnesota that prohibited his wife from following her academic career. In 1941, he moved to Washington, DC to work as chief of the Geography Division in what later would become the Office of Stra tegic Surveys (OSS) under William Donovan. After hir ing the necessary personnel and acquiring equipment, he took up the work of coordinating OSS research studies. Hartshorne held a key administrative position and was responsible for overseeing the publications of inter disciplinary groups' research activities.

Although he continued his work related to the OSS after 1945, he returned to the University of Wisconsin and served as Chair of the department from 1950–54, retiring in 1970. During the 30 years there, he supervised 32 masters degree students and ten PhD degree students. During the 1950s he was engaged in defending academic freedom from McCarthyism and during the 1960s con tinued his strong support for academic freedom.

Following his retirement in 1970 he continued to publish, travel, and participate in meetings. The publi cations included essays on William Morris Davis and a co authored essay on Carl Ritter. He traveled on lecture tours to the United Kingdom in 1973 and Japan in 1974. Until 1989 he attended annual meetings of the Associ ation of American Geographers. Hartshorne received several awards including the 'Outstanding Achievement' award in 1960 from the American Geographical Society, the Charles Daly Medal (1960) award from Clark Uni versity, and the Victoria Medal from the Royal Geo graphical Society in 1984.

After the death of his first wife in 1972, he married again in 1978. He passed away at home in Madison, Wisconsin in 1992.

See also: Historical Geography; Ivy League and Geography in the US.

Further Reading

Agnew, J. A. (1989). Sameness and difference: Hartshorne's, The Nature of Geography and Geography as Areal Variation. In Entrikin, J. N. & Brunn, S. D. (eds.) *Reflections on Richard Hartshorne's The Nature of Geography*, pp 121 139. Washington, DC: Association of American Geographers.

Barnes, T. (2006). Geographical intelligence: American geographers and research and analysis in the Office of Strategic Services 1941 1945. *Journal of Historical Geography* 32, 149 168.

Hartshorne, R. (1935). Recent developments in political geography. *American Political Science View*.

Hartshorne, R. (1935). Suggestions on the terminology of political boundaries. *Mitteilungen des Vereins der Geographen an der Universitat Leipzig*.

Hartshorne, R. (1937). The Polish corridor. *Journal of Geography*.

Hartshorne, R. (1950). The functional approach in political geography. *Annals of the Association of American Geographers* 40(2), 95 130.

Hartshorne, R. (1954). Comment on 'exceptionalism in geography'. *Annals of the American Association of Geographers* 44, 108 109.

Hartshorne, R. (1956). *The Nature of Geography: A Critical Survey of Current Thought in the Light of the Past* (2nd edn.). Lancaster, PA: Association of American Geographers (reprinted with corrections, 1961).

Hartshorne, R. (1958). The concept of geography as a science of space, from Kant and Humboldt to Hettner. *Annals of the Association of American Geographers* 48(2), 97 108.

Hartshorne, R. (1959). *Perspective on the Nature of Geography*. Chicago, IL: Rand McNalley.

Harvey, F. and Wardenga, U. (2006). Richard Hartshorne's adaptation of Alfred Hettner's system of geography. *Journal of Historical Geography* 32, 422 440.

Martin, G. J. (1994). Memorial to Richard Hartshorne, 1899 1992. *Annals of the Association of American Geographers* 84(3), 480 492.

Smith, N. (1989). Geography as museum: Private history and conservative idealism in *The Nature of Geography*. In Entrikin, J. N. & Brunn, S. D. (eds.) *Reflections on Richard Hartshorne's The Nature of Geography*, pp 91 120. Washington, DC: Association of American Geographers.

Harvey, D.

K. Woodward, University of Exeter, Exeter, UK
J. P. Jones III, University of Arizona, Tucson, AZ, USA

Glossary

Accumulation by Dispossession Processes by which capitalist expansion is accomplished through the privatization of social spaces and services.

Dialectics An analytic, ontological, historical, and political system wherein two terms interactively and processually co-determine each other, constituting a third term marking the continuous, unfolding interaction of the former two.

Flexible Accumulation A post-Fordist style of accumulation that is extremely malleable in terms of: (1) the labor process (utilizing a more contingent and hence cheaper work force) and (2) economic organization (with firms that are multisectoral, more dispersed and spatially nimble, and more vertically disintegrated).

Militant Particularism A term that describes disconnected and narrowly focused political mobilizations that are limited by virtue of being grounded only in the localized and relatively unique experiences or characteristics of individuals.

Space–Time Compression The apparent global shrinkage of distance and speed-up of temporal relations resulting from capitalist transformations in production, distribution, and consumption, all driven by ever more far flung and rapidly organized accumulation processes.

Spatial Fix The tendency of capitalism to remake spaces – either in new places or in the transformation of existing ones – so as to resolve contradictions and crises generated in production and overaccumulation.

Harvey, David (1935–)

Born in Gillingham, Kent, 1935, David Harvey (**Figure 1**) grew up during and immediately following World War II, and was profoundly influenced by the colliding geographies that war brought to the shores of Britain, and by the alternating mood of patriotism and loss of empire that followed in its wake. Harvey recounts (2001, 2002) an early fascination with contemporary and historic maps and a youthful curiosity for geology, vegetation, land use, and local architecture. His training – with all three degrees coming from Cambridge – was largely historical geographic in focus, culminating in a dissertation on changes in production in Kent's nineteenth century hops industry. That early 1960s work set the stage for a long held interest in the relations between economic processes of production and the particularities through which they unfold in local contexts and at particular points in time, the first glimmerings of a long career devoted to the development of a historical–geographical materialism.

During the 1960s, however, geography was in the throes of a so called 'quantitative revolution', a period characterized by the search for trans contextual spatial theories and laws, and Harvey's major work during that period, *Explanation in Geography* (1969), was fully immersed in the discipline's recent philosophical transformation. *Explanation* represented an extended philosophico methodological exploration of geographic approaches developed in spatial science. While generally sympathetic to the movement, *Explanation* presented both internal and external critiques of the systems that emerge through quantitative approaches. Employing analytics borrowed from positivist philosophy of science, Harvey married a concern for mathematic complexity and exactitude with the need to develop clear ontological and epistemological conceptions of the world we interrogate. In this regard, *Explanation* represented an attempt to combine the 1960s dream of quantitative analysis in studies of spatial differentiation with a focus upon the theoretical underpinnings that make any methodology – but especially those of spatial science – coherent and

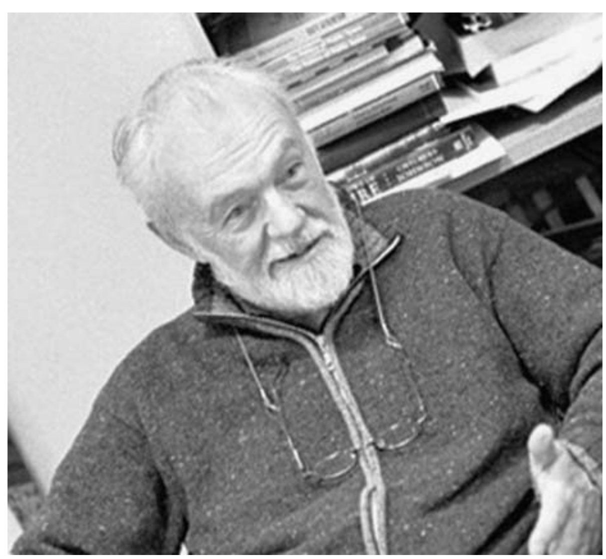

Figure 1 David Harvey.

rigorous. Although dated, *Explanation* is profoundly rigorous and analytic, and still serves as the germinal text within which the seeds of Harvey's mature thought can clearly be witnessed. Central among these are (1) the concern for the theoretical communication between science and philosophy, particularly as it pertains to epistemology and ontology, and (2) an early inclination toward the aesthetics of precision and solvability celebrated by positivist analytics. As he was to remark to an earlier generation trained in the largely descriptive field of regional geography: "By our theories you shall know us".

While *Explanation* appears to be exhaustive in terms of its philosophical attention to the methodological questions associated with positivism, at no point does it attempt to address questions regarding the political implications of geographic research. And yet, the human geography of the late 1960s and early 1970s, with its broad turn to issues of social relevance associated with the Vietnam War, imperialism, environmental degradation, women's rights, and racism and poverty, leveled this as a major flaw inherent in quantitative studies of spatial science in general. These issues were to consume Harvey after the publication of *Explanation*. He moved to Johns Hopkins University in Baltimore, Maryland, in 1969, arriving in wake of intense protests in that city following the 1968 assassination of Martin Luther King. Harvey found himself in a city with chronic impoverishment that jarringly contrasted with the country's status as the world's richest and most developed nation. He became involved in civil rights and anti war movements, conducted studies for the city of Baltimore on housing inequality, and, in effort to better understand the conditions underlying urban deprivation, began to read Marx.

In *Social Justice and the City* (1973), his first major work following *Explanation*, Harvey interrogated several versions of moral and social philosophy in an effort to develop a program for geography that brings to the forefront issues of spatial inequality and social injustice. Evolving from standard liberalist analytics to a discovery and celebration of a Marxism that retains a strongly positivist bent, this volume looks to socioeconomic difference in the urban setting as a device for illustrating the inseparability of social processes and spatial forms. Thus, the practices involved in the production of specific social relations (e.g., the property relation) serve to constitute specific spatialities (e.g., particular built environments and associated social welfare outcomes). At the same time, Harvey follows Karl Marx (and turns away from liberalism) by affirming the collapse of the distinction between production and distribution, allowing for a conceptualization of social justice that turns upon examination of actual social processes, rather than upon the invocation of transcendental notions of justice that supercede the specificities of the situation.

The Limits to Capital (1982) represents Harvey's most explicit and direct immersion in Marxist theory. A lengthy and difficult meditation on Marx's *Capital*, Harvey spends much of the book examining Marx's major theories regarding the production, distribution, and consumption of commodities, the processes of accumulation, and the inherent contradictions and crises in capitalism – notions that will recur throughout the remainder of Harvey's oeuvre. Harvey's major contribution to Marxist literature comes in the closing chapters of *Limits*, where he initiates the first steps toward a spatialization of Marxist production, accumulation, and imperialism. He also contributes to Marxist theory – and presages work by others on economic globalization – by highlighting the important role of financial capital in the capitalist accumulation process. The final chapters of *Limits* serve very much as a theoretical roadmap for almost all of the work that will follow. In particular, it is here that Harvey first describes the processes by which contemporary capitalism finds a 'spatial fix' for its own self produced crises through the exportation of devaluation to other spaces. Harvey suggests that such maneuvers ultimately tend toward the development of expanding crises in the form of imperialism and global war. It would be almost 20 years before Harvey would turn to any extended empirical explication of the nightmare visions that haunt the closing pages of *Limits*. Latterly, however, he has argued at length that the current moment of US imperialism, the hegemony of neoliberalism, and the initiation of global warfare (the war on terror, the wars in Iraq and Afghanistan, etc.) are illustrative of the processes first described in *Limits*.

Alongside the theoretical project of describing Marx for geographers and developing a spatialized Marxist theory for leftists of all persuasions, Harvey continued to write essays in the 1970s and 1980s elaborating a framework for urban processes under capitalism. These works, together with original materials, were assembled as a pair of twin books, both published in 1985 – *The Urbanization of Capital* and *Consciousness and the Urban Experience*. Gone was the tentativeness of Harvey that peppered parts of *Social Justice and the City*. As assembled in book form, the articles trace an exceptionally rigorous and urban specific analysis of capitalism, redressing the historicism in traditional Marxism on two counts: (1) by showing how the space of the city is the key site for accumulation, and (2) by demonstrating beyond *Limits* just how much spatial transformations in built environments are the heart and soul of capitalist accumulation: "Capitalist development has therefore to negotiate a knife edge path between preserving the exchange values of past capital investments in the built environment and destroying the value of these investments in order to open up fresh room for accumulation" (*Urbanization*, p 25). Harvey covers the production of urban infrastructures (or 'second natures' – roads, houses factories, schools, and shops) that arise from capitalist switches among its primary, secondary, and tertiary circuits, as well as the complexity of several types of rents and finance markets on housing provision. The separation of home and work is shown to have implications for not only production but also social reproduction and, finally, for the uneven distribution of class in urban space. Chapters on urban politics and planning close out *Urbanization.* They presage later work on urban entrepreneurialism that was to become one leg of his next major book, *The Condition of Postmodernity.*

As another leg of that next project, *Urbanization's* counterpart text, *Consciousness and the Urban Experience* shifts away from capital's concrete circulation through, and objectification in, the built environment to what he calls the more 'speculative' investigation of political consciousness. The urbanization of capital, then, has two sides, the second of which is a profound reorganization of social, cultural, and political relations, experiences, and meanings. In essence, this book is a description of urban modernism – as cultural phenomena, political life, and epistemology – read through the dialectic of space and society and as imprinted on the urban built environment and the minds and actions of its inhabitants. It is here, then, where one sees the cultural epistemological seeds of Harvey's next book, *The Condition of Postmodernity* (discussed below). Especially noteworthy in *Consciousness* is Harvey's early historical work in Paris, where he lived for a time, both researching the city's history and witnessing the ferment of French philosophical debate (the research would become part of a much later book, *Paris, Capital of Modernity,* 2003). *Urbanization* reprints a famous Harvey study on the Basilica of Sacré Coeur, where he shows a command for interpreting symbolic built environments that is on par with his insightful and rigorous analyses of the cold movements of capital.

As mentioned above, *Urbanization* and *Consciousness* provided Harvey with the tools for his next and most well known project, *The Condition of Postmodernity* (1989), written while he was at Oxford University. He uses them to address post Fordist shifts in economic organization and to assess the transformation of cultural and political consciousness that had emerged under the popularization of postmodernism, an art and architecture movement that celebrated difference, depthlessness, and fragmentation. In *Condition*, Harvey explains that this movement was not merely the product of the celebration of culturally driven difference or changes in perspective (though it was surely this as well). In the final analysis, the postmodern era was brought on by a series of adjustments by capitalism in the second half of the twentieth century in response to increasingly intense competition and rapid technological change. The response – known as flexible accumulation – was to speed up production and delivery by changing labor processes (e.g., outsourcing) and reworking the firm's technical and organizational arrangements (e.g., just in time production, branch plants, and relocations). These conditions effectively did the work of compressing space–time as the ever more rapid processes of capital accumulation – and the risks attendant to over accumulation – began to take on pervasive, global characteristics. While postmodernity was, according to Harvey, principally the product of economic processes, the cultural/artistic, academic, and political impacts of these shifts (i.e., postmodernism) were mediated, if not produced, through the epistemological shifts – the loss of foundations – resulting from space–time compression. The most disconcerting of these changes was the proliferation of perspectives married to a devaluation of the grounds for political mobilization.

> The experience of time and space has changed, the confidence in the association between scientific and moral judgements has collapsed, esthetics has triumphed over ethics as a prime focus of social and intellectual concern, images dominate narratives, ephemerality and fragmentation take precedence over eternal truths and unified politics, and explanations have shifted from the realm of material and political economic groundings toward a consideration of autonomous cultural and political practices. (Harvey, 1989, p 328)

Harvey's take on post Fordism in *Condition* was widely celebrated and critiqued, and in addition to responding to important reviews of the book, in the early 1990s he also turned some of his attention to organizing around and writing about a car plant closure in the Oxfordshire town of Cowley, where a Rover Group factory was located. Subsequently, he returned to Hopkins where he was to write a new book, *Justice, Nature and the Geography of Difference* (1996), an extended attempt to develop a Marxist onto ethical system as a response to the challenges that arose from various ethics of reflexivity and positionality in feminism, critical race theory, post colonial theory, and from what was often characterized as an attendant nihilism or relativism in post structuralist theories. He also attempts to address a longstanding failure of Marxist theory to effectively engage environmental issues. *Justice* made three key contributions to geographic thought in general and to the political debates of the 1990s in particular: (1) it offers a dialectically articulated ontology focused upon interrelating, co constitutive processes; (2) it puts forward a political vision that draws upon his ontological project of collective interrelations and thus rallies against fragmented political movements; and (3) it portends the beginnings of a turn toward considerations of utopianism – something that will come to fruition in *Spaces of Hope* (2000) – as a political strategy for thinking about collective political change.

Militant particularism, one of the key concepts Harvey tarries with in *Justice*, highlights the localized, specific situations from which political movements emerge. There – and more specifically in *Spaces of Hope* – Harvey's concern is to explicate a continuum running from the politicized specificities of the spaces of the body to the collectivized political generality of the global. At its heart, *Spaces* is an extended attempt to theorize routes to political subjectivities and progressive actions that culminate in large aggregates of individuals mobilized against the spread of globalization. While Harvey clearly begins to reposition earlier dismissals of studies of the body and of the politics of positionality through the development of an analytics of the 'body as accumulation strategy', he nevertheless cautions against 'body reductionism' that would establish the individualist perspectivalisms as the limit of political conceptualization. By contrast, Harvey proposes that projects of alliance engaged in alternative economies and collective action can be aided by returning to utopian concepts as a strategy of hope against the cynicism of global capitalism.

Latterly, Harvey has dedicated himself to the development of extended analyses of the rise and spread of globalization through a series of critical histories of the present that might be called the 'neoliberalism trilogy'. While not the first time Harvey has engaged the crises attendant to neoliberalism and globalization, nowhere else do these receive such relentless, surgical examination as in *The New Imperialism* (2003), *A Brief History of Neoliberalism* (2005), and *Spaces of Global Uneven Development* (2006). Together, these works look to the collusion of states (particularly the US and UK, but also notably China), the military industrial complex, and the withdrawal of the left and simultaneous rise of the right, to theorize the emerging socio economico political phenomenon of neoliberalist 'accumulation by dispossession', a process of creating new spaces for capitalism by dispossessing people of spaces and services to which they have certain rights. This is witnessed particularly within disenfranchisements accomplished through: (1) recent US led imperial warfare as a strategy of coercion over consent for participation in neoliberalist capitalism; (2) the withering and privatization of state based social services; (3) the spreading myth that crises in capital equal crises in the state; and (4) a growing gap between the rich and poor that is accompanied – not coincidentally – by an equally large gap between those who hold state power and those who are subject to it.

Written in the current, post 9/11 era of global warfare and new Western imperialist expansion, gone are many of the utopian ideals Harvey had toyed with only a few years before. Here, he reads the continuing devolution of the present situation as dire enough that a more immediate fix, a (re)turn to a 'New New Deal', sits at the forefront of his recommendations for contemporary socio political change. Over his lifetime, the US has fallen from its former stature as the world's producer, and is no longer in a position to claim the mantle of the world's financier; it can now only rely on its military strength and domestic surveillance to bear up its position in world affairs. Thus Harvey, while not denigrating the efforts of the many distinct local movements dedicated to inventing new ways to respond to these crises, nonetheless retains the view that widespread social transformation will not be achieved unless resistance is predicated upon and articulated through the common ground of class based solidarity.

See also: Critical Theory (After Habermas); Entrepreneurship; Geopolitics; Historical-Geographical Materialism; Human Geography; Nature; Neoliberalism, Urban; Postmodernism/Postmodern Geography; Radical Geography; Uneven Development.

Further Reading

Castree, N. and Gregory, D. (eds.) (2006). *David Harvey: A Critical Reader*. Oxford: Blackwell.

Harvey, D. (1969). *Explanation in Geography*. London: Edward Arnold.

Harvey, D. (1973). *Social Justice and the City*. London: Edward Arnold.

Harvey, D. (1982). *The Limits to Capital*. Oxford: Blackwell.

Harvey, D. (1985). *Consciousness and the Urban Experience*. Baltimore, MD: Johns Hopkins University.

Harvey, D. (1985). *The Urbanization of Capital*. Baltimore, MD: Johns Hopkins University.

Harvey, D. (1989). *The Condition of Postmodernity: An Enquiry into the Origins of Cultural Change*. Oxford: Blackwell.

Harvey, D. (1996). *Justice, Nature and the Geography of Difference*. Oxford: Blackwell.

Harvey, D. (2000). *Spaces of Hope*. Berkeley, CA: University of California.

Harvey, D. (2001). Reinventing geography: An interview with the editors of *New Left Review*. *Spaces of Capital: Towards a Critical Geography*, pp 3 24. London: Routledge.

Harvey, D. (2002). Memories and desires. In Gould, P. & Pitts, F. (eds.) *Geographical Voices: Fourteen Autobiographical Essays*, pp 149 188. Syracuse, NY: Syracuse University Press.

Harvey, D. (2003). *Paris, Capital of Modernity*. New York: Routledge.

Harvey, D. (2003). *The New Imperialism*. Oxford: Oxford University.

Harvey, D. (2004). A geographer's perspective on the new American imperialism: Conversation with David Harvey. Berkeley, CA: University of California. http://globetrotter.berkeley.edu/people4/Harvey/harvey con0.html (accessed jun. 2008).

Harvey, D. (2005). *A Brief History of Neoliberalism*. Oxford: Oxford University.

Harvey, D. (2006). *Spaces of Global Capitalism: A Theory of Uneven Development*. London: Verso.

Woodward, K. and Jones III, J. P. (2008). The condition of postmodernity: David Harvey. In Hubbard, P., Kitchin, R. & Valentine, G. (eds.) *Key Texts in Human Geography*, (pp.125 134). London: Sage.

Health and Development

S. Asthana, University of Plymouth, Plymouth, UK

Glossary

Determinants of Health A wide range of factors influence individual health outcomes, although the relative importance of different factors varies according to context. In situations of absolute poverty, poor diet and exposure to pathogens (e.g., due to lack of safe water) are key determinants of high rates of infectious disease among children. In societies where children are no longer dying from infectious disease, other factors take on greater significance, including genetic inheritance, early-life exposures, lifestyle factors, and psychosocial stress.

Development Traditionally seen as the means by which a traditional, low-technology society is changed into a modern, high-technology society, primarily through economic growth. More recently, it has been conceived in terms of the broader concept of 'human development' (comprising social, cultural, emotional, education, and health as well as economic dimensions). Many also argue that development should include improvements in social justice (declines in 'relative poverty', improvements in women's rights and status, and so on).

Globalization A complex term that captures growing global connectivity – in trade, the flow of capital, the movement of people, and the diffusion of social, cultural, and political values throughout the world. Due to the global dominance of western international organizations (such as the World Bank and the International Monetary Fund), market liberalization – and associated social and political values – has been the dominant theme of globalization over the past 30 years.

Polarization The process by which the gap between the socially advantaged and disadvantaged becomes wider with respect to key measures of human development (such as income, health, and education outcomes). Social polarization tends to be greater in market-liberal societies than in countries where strong state intervention is used to redistribute wealth and other key resources.

Primary Healthcare A movement initiated in the late 1970s to improve health status in the poorest countries of the world by shifting the emphasis from cure to prevention (including bringing about improvements in nutrition, education, and sanitation) and achieving universal and equitable coverage in 'basic' health services. It is generally agreed that, due to a change in international social and economic policies, the potential of primary healthcare was never widely tested.

Relative Poverty A concept that is closely related to income inequality. Here, poverty is not conceived in absolute terms (and measured around a basic subsistence threshold) but relates to the prevailing income levels or living standards of a particular society. Thus, even if the incomes of the poor rise, there can still be an increase in relative poverty if the incomes of the better off rise even faster and the distribution of income becomes more unequal.

Introduction

The relationship between health and economic devel opment is well established. Life expectancy is lowest in the lowest income countries where a high proportion of deaths occur among children and are due to infectious diseases associated with poor sanitation and nutrition. Due to the strong association between poverty, poor sanitation, inadequate hygiene, and unsafe drinking water, children living in the poorest communities are more likely to be exposed to infectious agents, such as the fecal–oral pathogens responsible for diarrheal disease. They are also more likely to be malnourished and suffer from micronutrient deficiencies. Poor nutrition weakens their immune systems making them more susceptible to and less able to fight infections than well nourished children. Illness is thus more likely to be prolonged or severe, increasing risk of death. Unfortunately, mal nutrition and infection are joined in a self reinforcing cycle, infection depressing nutritional status still further and increasing the risk of future and more prolonged or severe illness episodes. Thus, child mortality in low income countries is often the result of a long series of recurrent infections and nutritional deficiency.

As the average wealth of a country increases, there tends to be an improvement in living standards and a subsequent decline in diseases that are associated with both poverty and child mortality. Life expectancy in creases, as does the relative contribution made to the mortality burden by noncommunicable diseases such as heart disease and cancers that develop in later life. The process by which the disease profile of a society changes from being dominated by infectious diseases to being characterized by noncommunicable chronic or de generative diseases is formally known as 'epidemiological transition', a concept that provides a useful descriptive device for contrasting the health status of countries/

regions at differing levels of development. According to World Health Organization (WHO) estimates for 2005, an estimated 70% of deaths in the poorest region, Africa, were due to communicable diseases and only 22% to noncommunicable diseases. This pattern reflects both the sizeable proportion of child mortality in the region and the impact of HIV/AIDS. In Europe, by contrast, only 6% of deaths were due to communicable diseases and 86% due to noncommunicable diseases. Here, the vast majority of deaths occurs among older people. Other regions are more economically diverse and this is reflected in their mortality profiles. In southeast Asia, for example, 35% and 54% of deaths in 2005 were attributed respectively to communicable and noncommunicable diseases.

Like early development theories, epidemiological transition is a rather simplistic and linear model. It assumes that poorer countries will simply repeat the development experience of the West and undergo an associated shift in disease burden. In the past few decades, however, globalization has impacted upon the development trajectories of poorer countries in a way that is neither linear nor uniform or necessarily progressive. Production, consumption, trade, finance, communications, travel, social relations, environmental changes, political values, and systems of governance are all subject to growing global connectivity and this has important consequences for the web of factors – both local and global – that impact upon health. Understanding this connectivity – in its economic, social, environmental, and policy dimensions – is key to understanding recent trends in health and healthcare in low and low to middle income countries.

Health, Development, and Globalization: Economic Dimensions

Trends in Absolute Poverty

The principal way of addressing the high burden of infectious and parasitic diseases in low income countries is to address poverty itself. Thus, it is important to examine recent trends in income and living standards. Since the early 1980s, neoliberalism has dominated macroeconomic policy on a global scale, reflecting the belief that, the more open economies become, and the more integrated with world trade, the more that they benefit from economic progress. The World Bank, which has been a major proponent of the 'openness is good for you' school, suggests that this strategy has contributed to a significant decline in the numbers living in absolute poverty. Using an income measure adjusted for purchasing power parity (PPP), an approach that captures differences in domestic purchasing power, the proportion of the world's population living on less than $1 a day fell from 28% in 1990 to 21% in 2001.

Some care should be taken in interpreting the World Bank's figures on poverty trends. Globally, poverty reduction is largely accounted for by the success of East Asia, particularly China. By contrast, poverty levels in Latin America and the Middle East remained fairly static between 1990 and 2001. Meanwhile, there were dramatic increases in the number of people living on less than $1 a day in sub Saharan Africa and on less than $2 a day (a more realistic threshold for extreme poverty in this region) in the Commonwealth of Independent States (CIS) that makes up the former Soviet Union. Both of these regions have become poorer in real terms during the period. Indeed, by 2001, real gross domestic product (GDP) per capita in sub Saharan Africa (£306) and the CIS ($1312) was lower than in 1971. By raising the relative incomes of poorer countries, the World Bank's income adjustments do not fully reveal the extent to which these regions have experienced economic stagnation or indeed disintegration.

These economic trends are reflected in health indicators. Sub Saharan Africa remains the only region in the world where the absolute number of child deaths is rising. Although there have been improvements in immunization coverage, access to safe water and, to a lesser extent, sanitation (demonstrating some of the positive impacts that globalization can have with respect to the role played by supranational organizations and the diffusion of health technologies), high levels of poverty, and little change in the prevalence of childhood malnutrition have contributed to a very slow decline in the under five mortality rate (from 187 per 1000 in 1990 to 171 per 1000 in 2003). In the same period, life expectancy at birth has fallen from 50 to 46 years (a similar level to that of England in the 1840s). This unprecedented decline can be attributed in large part to the high prevalence of HIV/AIDS in the region.

Reversals in health status have also been observed in the CIS where a rapid transition from centrally planned socialist to market economies resulted in economic collapse, increases in unemployment and income poverty, and the disintegration of a comprehensive system of social protection. Child poverty rates are particularly high in the Central Asian Republics and the Caucasus. In addition to increasing rates of child malnutrition, several countries in this part of the region have seen rises in infant and child mortality. Meanwhile, in the Russian Federation, life expectancy for males has dropped from 70 years in the late 1980s to 59 years in 2003, a decline that can be attributed to the very high incidence of cardiovascular disease, a resurgence of infectious diseases, and a rise in alcohol related problems, including homicide, suicide, and injuries. The growth of absolute poverty in this region has played a role in these health reversals. Trends in income inequality are signifi cant too.

Trends in Relative Poverty

A distinction is generally made between the indirect effects of income inequality on health and the direct effects. In highly unequal societies, where wealth is concentrated in the hands of relatively few, there will be relatively large numbers of people with low incomes whose health is also poor. Thus, income inequality pro duces poorer average health outcomes through the ag gregate effects of individual poverty. In addition, it has been proposed that income inequality has a direct effect on health (over and above the aggregate effects of indi vidual incomes) due to the psychosocial effects of having a low social status in an unequal society. In addition to promoting health damaging behaviors, stress has been linked to metabolic effects relating to increased risk for coronary heart disease and to other effects such as de pression and poorer immune function.

Unfortunately, globalization has been associated with an increase in income inequality. In 1960, the richest fifth of the world's population had 30 times more income than the poorest fifth. Today, the average income of the top 20% is about 50 times the average income of the bottom 20%. Were it not for high growth rates in India and China, global inequality would be even higher. Thus, the gap between most developing regions and rich countries is still widening. Polarization is occurring within coun tries as well as between them. In China, inequality is now greater than before the Communists won power in 1949 (although income inequality measures for the country remain relatively low). India also remains subject to large regional, gender, and income based inequalities. Thus, although both countries have experienced significant economic growth, this has not translated into more rapid declines in infant and child mortality.

High levels of income inequality can produce a pat tern of 'epidemiological polarization' in which there is a double burden of pretransitional diseases among the poor and post transitional diseases among the richer sections of society. This pattern is illustrated by several Latin American countries where, although average incomes are in the middle range, income distribution is particularly skewed. Urban poverty has become a large scale problem and, like the rural poor, the residents of slum and squatter communities are at increased risk of poor nutrition and exposure to environmental health hazards. As a result, overall levels of health are low, relative to national average income. Brazil provides a good example of the health damaging effects of inequality. At PPP US$7790 per capita ($2788 in real terms), income in Brazil is considerably higher than in China and Sri Lanka. Yet, Brazil's average life expectancy is lower than in both of these countries. Indeed, it is equivalent to that of Vietnam which has a per capita income of PPP US$2490 ($482 in real terms). Countries such as Sri Lanka and Vietnam, where income inequality measures are significantly lower than the Latin American average, demonstrate that high levels of health can be achieved at relatively low levels of wealth if the benefits of economic development (in cluding social infrastructure and welfare provision) are distributed more evenly throughout the population.

The extent to which observed associations between income inequality and health outcomes can be attributed to indirect or direct effects has been strongly debated. Perhaps the best evidence for the direct role played by psychosocial stress is provided by the CIS. Here, the difficulties people have had in adjusting materially and psychologically to the collapse of communism have been linked to rising stress and associated risk behaviors, a phenomenon described as 'transition trauma'. In Russia, there has an increase in alcoholism and smoking. There has also been an explosion in the numbers of people injecting drugs. Part of this reflects the removal of tra ditional controls over drug misuse and the ease with which drugs can be trafficked across the Russian Feder ation. However, increased psychological stress has also been implicated in risk behaviors. For young people, economic malaise has limited opportunities for economic stability. At the same time, increased travel and access to Western media via satellite television has raised their expectations of consumer lifestyles. In addition to re sponding to feelings of discontent by turning to drugs, there has been a rise in the numbers of young women and men who view commercial sex work as one of their few opportunities to make money. The rise of both injecting drug use and commercial sex is, of course, related to the rapid increase in HIV prevalence in this part of the region.

Health, Development, and Globalization: Social Dimensions

Although some believe that globalization is a force for poverty reduction, others believe that its economic im pacts have been negative, with important consequences for health. Views are also polarized with respect to the social domain. On the positive side, a substantial human rights regime has emerged, supported by alliances of governmental, nongovernmental, international, and supranational organizations. Networks devoted to the promotion of women's rights have been particularly active, giving rise to a universal understanding of the links that exist between women's social, economic, legal, and health status. Practical progress has been in the le gislation passed by many countries to raise minimum legal age of marriage, prevent female genital mutilation, promote equal legal status (e.g., with respect to marriage and divorce, child custody, and inheritance laws), and so on. Similarly, global campaigns to promote social justice for children have led to the ratification of the Convention on the Rights of the Child (CRC) and the passing of supportive legislation by many state governments. Such developments, which have introduced universal stand ards against which the treatment of women and children can (theoretically at least) be measured, suggest that globalization is a force for social good.

Against this, questions have been raised as to whether supportive legislation can result in major improvements in the health and well being of vulnerable groups in the absence of wider socioeconomic change. For example, women's lack of social, economic, and reproductive au tonomy is strongly shaped by poverty. The lowest rates of female school enrollment and the highest rates of child marriage and fertility are found in the poorest region, sub Saharan Africa. Here, a number of factors make high fertility desirable, including the continued importance of smallholder farming (which is reliant on family labor); high rates of infant and child mortality (thus, many births are needed to ensure an optimum family size); and the need, in the absence of wider state social security, for adult children to care for aged parents. Although many African countries have declared 18 as the legal minimum age of marriage, nearly 50% of girls in sub Saharan Africa marry before this age and, as first time mothers, young girls face high risks in their pregnancies (girls aged l0–14 are five times more likely and girls aged 15–19 twice as likely to die in pregnancy or childbirth than women aged 20–24). In 2000, there were an estimated 940 maternal deaths per 100 000 live births in the region. For every girl or woman who dies, many times that number suffer long term damage to their health as a result of childbearing. For example, the prevalence of obstetric fistula is high in areas of high maternal mor tality and the social impact of this distressing condition can be catastrophic for young women who may be os tracized or even abandoned by their families. Infertility is another long term side effect of birth complications, a tragic outcome for women in societies where a high premium continues to be placed on motherhood.

In addition to pointing out the need to recognize the economic context of 'social' health determinants, critics of globalization note that the process does not always result in the liberalization of social attitudes or in the diffusion of values that support health and well being. Within the former Soviet Union, for example, the col lapse of a state system which, in theory at least, upheld equality between the sexes, has heralded a return to more traditional gender relationships. In Central Asia, patri archal religious traditions are reasserting themselves, social control being openly exercised over young women with respect to arranged marriage and early childbearing. In many parts of the region, violence against women – once condemned if largely ignored by the state – has become institutionalized. There is evidence of state complicity in the trafficking of women from several transitional countries.

In other parts of the world, globalization has resulted in growing cultural homogenization. The increasing internationalization and commercialization of the food trade is giving rise to a shift from simple diets with little variation to a greater reliance on processed foods, fats, and sugars. Thus, while undernutrition remains an im portant problem, poor countries are also experiencing rising rates of obesity. There are concerns, too, about the way in which cultural values and norms regarding sex are being universalized. The globalization of pornography, the increasingly sexualized imagery of global capitalism, and the rapid expansion of international travel have been linked to the growing commercialization of sex and, in turn, the transmission of sexually transmitted diseases, including HIV. As with women's status, however, the 'geography' of the sex trade is not shaped by sociocultural factors alone.

Political and economic changes associated with glob alization are also implicated in the rise of commercial sex. These include the removal of strong state controls and the breaking down of national borders (reflected in the growth of trafficking of women from Russia, Ukraine, Mainland China, and Vietnam); the globalization of military activities (reflected in an increase in both sex work and HIV transmission in peacekeeping areas); and economic polarization. Many argue that export oriented growth has threatened rural livelihoods in sub Saharan Africa, leading to increases in labor migration and ur banization. There has been a growth in 'survival sex' among impoverished women. At the same time, male migrants who are away from home for extended periods of time and who are housed in crowded, all male hostels in mining settlements or commercial estates are more likely to engage in casual, unprotected sex. In Carle tonville, South Africa, the largest gold mining complex in the world, HIV rates have been found to be as high as 78% among sampled commercial sex workers and 36% among miners. When they return to their families, migrant workers carry the virus into their rural communities.

Tensions thus exist between the potential of global networks to promote the social status of vulnerable groups and the practical impact of globalization on social inequalities. This extends to the way in which global ization has been associated with a retreat from state welfarism. In keeping with the wider neoliberal program of privatization and the liberalization of trade, there has been a rolling back of the state with respect to the pro vision of welfare and other public goods, arguably lead ing to an increase in women's domestic responsibilities.

Older people, who have increased needs for health and social care, have been particularly vulnerable to reductions in state expenditure. Thus, although supranational organizations have played an important role in highlighting the needs of the growing numbers of older people in poorer countries, the support systems available to older people – statutory and traditional – are being steadily undermined. There is a worldwide trend toward nuclear rather than extended multigenerational families. Women's participation in the paid labor force is growing and an ever increasing proportion of the world's population lives in urban centers, removing support for older people in rural areas. Globalization may also be promoting cultural beliefs that undermine the security and prestige of older people within their families and communities.

Health, Development, and Globalization: Environmental Dimensions

While most accounts of health, development, and globalization focus on the implications of socioeconomic reforms, fundamental changes are also occurring to the planet's ecosystems, with important consequences for human health. Climate change, with respect to both temperature and patterns of precipitation, has been linked to a range of potential health effects, for example in the distribution of malnutrition (through effects on yields of agricultural crops); the incidence and distribution of waterborne, food borne, vector borne, and pollution related diseases; and deaths and injuries from natural disasters. Accelerated losses in biodiversity have important implications for global health, by threatening prospects for developing new pharmacological products. Overall, however, the effects of global climate change are likely to be heavily concentrated in poorer populations where the most important climate sensitive health outcomes (malnutrition, diarrhea, and malaria) are already common and where vulnerability to climate effects is greatest.

Urbanization, an integral feature of globalization, is also associated with a rise in environmental health hazards. Air pollution is concentrated in cities due to industrial emissions and increased motorized transport and, in many cities, air quality falls well below international health standards. In Beijing, for example, more than 400 000 people die each year from pollution related illnesses. The speed at which urbanization is occurring in poorer countries also contributes to environmental health problems as the supply of housing, infrastructure, and other services is not keeping pace with demand. Nearly 1 million people live in slum or squatter settlements, many of which are located in environmentally unsafe areas – steep hillsides, flood plains, or near waste dumps and polluting industries – exposing residents to natural and man made disasters. Overcrowded and poorly ventilated housing helps the spread of diseases by direct contact. Sanitation remains a major problem. Municipal supplies of safe drinking water do not tend to reach slum populations and communal latrines may be shared by thousands of people. Wastewater and overflow from septic tanks may run down open drains and sewers which, during the rains, are subject to flooding. Such factors are associated with a very high environmental load of fecal–oral pathogens. The poor environmental conditions of slum communities also provide ideal conditions for increased transmission of rodent and vector borne diseases, pools of standing water, water drums, discarded automobile tires, and nonbiodegradable packaging providing breeding sites for mosquitoes. Living conditions in urban slums present significant health risks, even for children from higher income families who have been found to have higher rates of diarrhea than children from poor rural families because they are exposed to contaminated water and food. Increased exposure to environmental health hazards, together with high rates of child malnutrition (particularly in sub Saharan and South Asian cities) means that, with respect to health status, slum dwellers are as badly if not worse off than their rural counterparts.

Health, Development, and Globalization: Policy Dimensions

In addition to factors impacting upon health status, globalization is affecting healthcare policy through the global diffusion of health sector reform (HSR). Reflecting the dominance of neoliberal thinking in international policymaking, HSR emphasizes the need to move away from state financed and state run healthcare systems by widening the range of financing options and institutional players who provide healthcare. With respect to financing, user fees, cost recovery schemes, and private health insurance schemes have been encouraged in order to increase resources available for healthcare. Increasing emphasis is also placed on the role of the private sector in delivering healthcare services, reflecting a belief that this promotes higher quality.

The rise of a market based approach represents a significant shift in healthcare philosophy for both low income and former communist countries. In the late 1970s, international health policy drew upon examples of socialist style health systems to propose that low income countries embraced the Primary Health Care (PHC) movement. This called for universal and equitable coverage of basic health services and a shift in emphasis from intensive and sophisticated curative care in urban centers to prevention within the community. Underpinning the movement was an understanding of the need to redistribute resources – not only for healthcare, but for nutrition, education, and sanitation – to drive up standards of health. However, PHC had hardly been launched, before a market philosophy came to dominate economic and public policy throughout the world. In addition to rejecting the need for redistribution, it sought to reduce the role of the state in healthcare provision. During the 1980s, structural adjustment policies sponsored by the IMF and World Bank were associated with significant cuts in healthcare expenditure, particularly in sub Saharan Africa; and since the 1990s, low income countries have been called upon to embark on HSR. The potential of PHC was thus never tested, although the impact of replacing socialist style health systems with a market based approach can be assessed with respect to former communist countries.

In prereform China and the former Soviet Union, strong central planning had created a hierarchy of health services with regard to the skills of practitioners and the degree of specialist care provided. Innovative use of paramedical workers such as feldshers in the USSR and barefoot doctors in Chinese rural communes played an important role in extending basic healthcare to small rural communities. The aim was thus to guarantee universal access to care with opportunities for referral to more specialist centers. Strong emphasis was also placed on public health (e.g., sanitation, immunizations, and other preventive measures). The combination of an emphasis on prevention, the use of paramedical workers, the basic simplicity of management under a highly unified system, and the relatively low salaries of healthcare workers meant that, in both countries, a highly accessible system of healthcare was achieved at relatively low cost. There were, however, important differences between the two systems, the former Soviet Union relying on state and collective funding to a far greater degree than China, which drew funds from a combination of sources, including central government, local communes, work related health insurance schemes, and out of pocket payments.

Such differences, together with China's rising prosperity and the fact that political reform has taken place more gradually than in the CIS may account for differences in the impact of HSR in the two areas. While the shift away from state and collective funding has caused health systems in several parts of the CIS to deteriorate close to the point of collapse, there has been a significant rise in health expenditure in China and a rapid development of private specialist centers, particularly in the cities. However, both countries demonstrate some of the problems that can result from HSR. Direct payments from patients have become the most significant source of healthcare financing in China and the western states of the CIS. There is evidence that this prevents the poor from gaining access to healthcare; while those who do seek medical assistance may spend less on food and other basic household essentials, sell livestock, or borrow money, triggering a process of increasing debts and the loss of entitlements. There is also growing geographical inequality with respect to access to services. Rural services have been particularly affected, with the deterioration or closure of healthcare facilities and the loss of skilled personnel. Preventive services, which are significantly less profitable than curative care, have also suffered from under investment, reversing previous successes in the control of infectious diseases, such as tuberculosis (TB), childhood infections and, in China, schistosomiasis. The experience of HSR also raises questions regarding the quality of private sector care. Because drug sellers and private practitioners have a financial incentive to overprescribe, private sector delivery has long been associated with excessive medication in India, where a significant proportion of expenditure on drugs is unnecessary, irrational, or indeed hazardous. The fact that the overuse of drugs has also become a serious problem in China suggests that some care should be taken in assuming that the encouragement of a private sector in low income countries (where regulation remains a problem) automatically yields quality improvements.

Issues

An understanding of the relationship between health and development must today take account of the impact of globalization on the many factors that influence human health. Diffusion studies have focused directly on the diminished importance of spatial boundaries, conceiving how negative health risks – from HIV and drug resistant forms of TB to avian flu – can diffuse at speed across our shrinking world. However, many of the key aspects of globalization transcend territorial space, requiring a different understanding of the ways in which economic, social, cultural, and political relations intersect with health. With respect to the differential impacts of globalization, territorialism remains important, the historical conditions of particular states and regions shaping their response to different globalization processes. The 'map' of global health has thus become a good deal more complex in recent years.

This complexity has fostered considerable debate as to whether globalization is associated with continuity or change, convergence or divergence, and social progress or injustice. Some, for example, argue that globalization is a postcolonial imperialism which has heightened inequalities between the North and the South and between the rich and the poor of low income countries with deeply negative consequences for health. Others, however, point out that countries previously described as 'developing' are doing very well in the global economy, enjoying greater wealth, a decline in inequality, and rapid improvements in health status. Perhaps ultimately, the question of whether globalization has been a force for good or a negative influence on global health depends upon one's political point of view.

See also: Disease Diffusion; Disease Mapping; Epidemiological Transition; Gender and Health; Healthcare Accessibility; Health Inequalities; HIV/AIDS in Developing Countries.

Further Reading

Barnett, T. and Whiteside, A. (2002). *AIDS in the Twenty First Century: Disease and Globalisation*. Basingstoke: Palgrave.

Cockerham, W. (1999). *Health and Social Change in Russia and Eastern Europe*. London: Routledge.

Doyal, L. (1995). *What Makes Women Sick: Gender and the Political Economy of Health*. Chapel Hill, NC: Rutgers University Press.

Evans, T. (2002). A human right to health? *Third World Quarterly* 23, 197 215.

Kawachi, I. and Kennedy, B. P. (eds.) (2002). *The Health of Nations: Why Inequality/Harmful to Your Health*. New York: The New Press.

Leon, D. and Walt, G. (eds.) (2001). *Poverty, Inequality and Health*. Oxford: Oxford University Press.

McMurray, C. and Smith, R. (2001). *Diseases of Globalisation: Socio Economic Transitions and Health*. London: Earthscan.

Saker, L., Lee, K., Cannito, B., Gilmore, A. and Campbell Lendrum, D. (2004). *Globalisation and Infectious Diseases: A Review of the Linkages*. Geneva: WHO.

Sen, G., George, A. and Ostin, P. (eds.) (2002). *Engendering International Health: The Challenge of Equity*. Cambridge, MA: MIT Press.

Stiglitz, J. (2000). *Globalisation and Its Discontents*. London: Penguin.

United Nations Children Fund (2001). A decade of transition. *Regional Monitoring Report No 8.* United Nations. http://www.unicef irc.org/cgi bin/unicef/download insert.sql?ProductID = 313 (accessed in April 2008).

United Nations Development Programme (2005). *Human Development Report: International Cooperation at a Crossroads: Aid, Trade and Security in an Unequal World*. New York: United Nations Development Programme.

UNHABITAT (2006). *The State of the World's Cities 2006 7*. Nairobi, Kenya: United Nations Human Settlements Programme.

Health Geography

G. Moon, University of Southampton, Highfield, UK

Glossary

Biomedical Model of Health An approach to the study of health that focuses on death and disease and sees these events resulting from disturbances to the internal pathology of the human body.

De-Medicalization The process of challenging the hegemony of the biomedical model of health; also associated with the rise of healthism.

Healthism A shorthand term for society's fascination with health and the emerging tendency to center human activity around health improvement.

Social Model of Health An approach to the study of health that sees ill-health resulting from root causes that are located within society, particularly behavioral and structural causes.

Introduction

Health geography (or the geography of health) effectively emerged in the 1980s as a terminology describing the encounter between geography and death, disease, and related aspects of service provision. It has since coexisted with the longer established field of medical geography with which it shares concerns about spatial variations in its subject matter and the nature of the relationship between health and place. In practice there is often little to distinguish work in health geography from that in medical geography and it sometimes seems that the choice of nomenclature is a matter of semantic preference or academic tribal loyalty rather than any substantive ontological difference.

Despite this fuzzy background, the dictionary of human geography has featured separate entries for health and medical geography. Programmatic statements and reviews of the field have also continued to dwell on claimed, anticipated, and actual differences. Central to the possible distancing of health from medical geography is generally the idea that the former in some way breaks away from a concern with biomedical aspects of its subject matter, embracing instead a broader, more social, model of health. Further differences are often associated with a greater propensity within health geography toward theorization and critical engagement. It is thus in the social model of health, contemporary social scientific theory, and the notion of criticality that the essence of health geography can be found. These ideas in turn resonate with wider academic and popular engagement with health as an experiential and analytic construct.

The Turn to Health

A traditional view of the subject matter of health/medical geography would lay great emphasis on a biomedical model of health. Death and disease and its treatment are, in this model, best understood through reference to the biomedical workings of the human body. Relevant knowledge is rooted in the scientific disciplines of biology and chemistry as applied to human pathology and requires extensive training in the mechanisms of the body and the pharmaceutical, surgical, and other ways in which challenges to the smooth running of that mechanism can be treated. In essence, the biomedical model focuses on bodily malfunction and body repair. Indeed, its gaze can be more than simply embodied; it is often inside the body via a concern with cells and genes.

Few could deny the long term effectiveness of the biomedical model. Standard medical histories are replete with examples of great discoveries that have undoubtedly played their part in combating disease. The vast majority of people reach out for and gratefully embrace the biomedical model when confronted with disease or ill health. Medical doctors, as the key practitioners of the biomedical model, and hospitals as the most visible site of biomedical practice are accorded significant importance in society. The biomedical model can be said to have achieved a hegemonic position in both popular and scientific constructions of health. Yet, alongside this dominance exist critical perspectives drawn from both 'inside' and 'outside' of medicine.

Internal critics of the medical model, practising doctors, have questioned the evidence base underpinning some medical interventions and examined the temporal coincidence between improvements in human health and medical advances. Cochrane noted how significant, well established treatments were often based on very little evidence from soundly designed research studies. His views were influential in the emergence of the concept of evidence based medicine. McKeown, in contrast, pointed out how improvements in mortality often began well before the major medical discoveries popularly linked with such improvements; he stressed the often more significant role of population level public health interventions, such as clean water, and, more generally, improvements in nutrition.

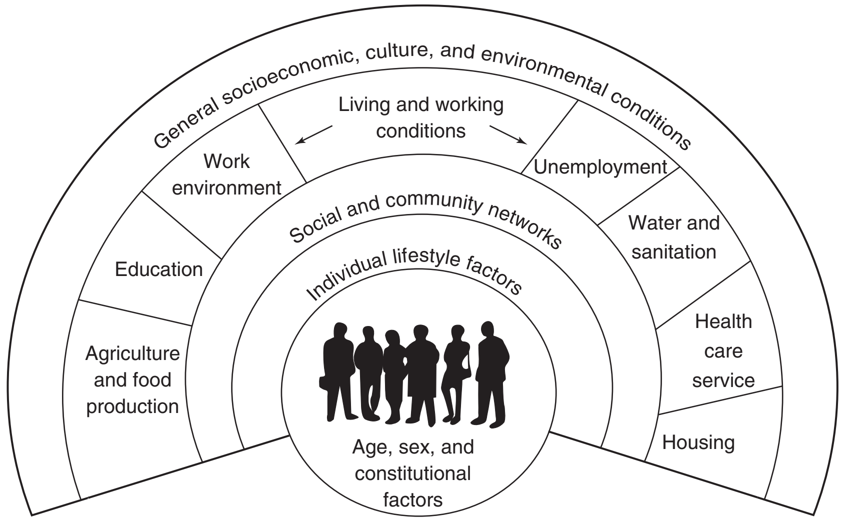

Figure 1 The Dahlgren Whitehead health model. From Dahlgren, G. and Whitehead, M. (1991). *Policies and Strategies to Promote Social Equity in Health*. Stockholm: Institute of Futures Studies.

Other critics of the biomedical model have worked outside medicine, sometimes in the sciences but more often from a social perspective. Dubos exemplified the science critique, placing medicine and medical endeavors in an environmentalist context that resonates significantly with the work of Jacques May, an individual often cited as central to the development of medical geography. More socially driven critiques include those of Illich, Navarro, and Foucault, all working in different ways on the consequences of the powerful position of the biomedical model for an understanding of health and disease.

In addition to challenging the claims of the biomedical model, critiques of the biomedical model also point to an alternative social model of health in which the focus of attention is outside the human body, on the social (and also geographical) context in which ill health is to be found and through which it can be understood, some would argue better understood. In this social model, the locus of interest is resolutely 'upstream'. This fluvial analogy equates the medical model with proximate, even superficial, analyses of ill health; real understanding looks upstream beyond the immediacy of the diseased body to the more distant but more fundamental social causes of ill health. To exemplify, coronary heart disease is proximally caused (among other factors) by fat deposits in the arteries of the heart. The upstream causes cited by a social model would include diet and exercise and, yet further upstream, the poverty that may constrain choices regarding diet and exercise. Dahlgren and Whitehead cleverly captured the essence of this upstream gaze in their mandala model of the determinants of health (**Figure 1**).

This broadening of concern, out from the body, its cells and its genes, to a more societally focussed perspective draws strength not only from the critics of the biomedical model with their various pointers to public health interventions, environmental factors, and the embeddedness of medical power, but also from the more general social shift in attitudes to health that has taken place in recent decades. The reasons for this shift are numerous. They include reducing trust of medicine, in part a public reflection of the more academic critique, greater popular awareness of health matters, the growing take up of alternative and complementary medicine, and, perhaps most importantly, the increasing centrality of health in everyday life. Sometimes termed healthism, this centrality is evidenced by the way society produces and constructs human actions in terms of real or imagined effects on/of health. Manifestations of healthism include an increasing interest in body maintenance and adornment, the rise of the health club and associated cultures of improvement linked to diet, exercise and other health related behaviors, and the use of health as a motif in discussions on areas as diverse as urban design, workplace practices, and family relationships.

In policy terms, the broadening of the health agenda away from the biomedical model toward a more social model and the accompanying rise of healthism has been reflected in an equivalent broadening of the networks of actors involved in the management of ill health. This network has shifted from a biomedical focus on curative treatment, largely by professionals schooled in the medical model, to one of far greater diversity embracing social care, housing, urban planning, education, and transportation among other fields. The idea of joined up policy is central, in this case with health at the centre. Another key shift has been from a reactive to a proactive position. Prevention, positive health, and health promotion have become important means toward the objective of shifting the focus away from disease and its treatment, and toward the avoidance or at least limitation of ill health through social intervention. In Canada, the population health movement has been particularly influential in placing these ideas on the policy agenda while, more generally, the notions of healthy (public) policy and healthy environments have played a part.

The transition from medical to health geography is indicative of geography's participation in these wider epistemic shifts and its disciplinary response. In this respect, it shares similarity with other social sciences. Sociology and psychology have experienced analogous changes, as have those aspects of medicine that align most clearly with the social sciences – witness the emergence of nonclinical public health and critical 'upstream' work in health promotion. In all cases, the 'cultural turn' of the 1990s further fueled the transition as the evident flexibility and changeable nature of health was progressively deconstructed.

The Emergence of Health Geography

How then did health geography emerge as a terminology? Alongside the general emergence of health discourse addressed above, it is possible to identify three further necessary contexts that aided its emergence. All these center to a greater or lesser extent on the practices and concerns of medical geography in the late 1970s and early 1980s.

Contexts

At that time, the established paradigm within the subdiscipline was that of disease ecology. This important perspective had itself emerged in the period between the two world wars and their immediate aftermath. Its focus was very much on disease and death, particularly infectious disease. It saw disease geographies in terms of the ecologies of the locations where they were prevalent, emphasizing physical as well as social factors. Thus, climate and geology played a part in unraveling disease etiology alongside social organization and cultural practices. Specific environmental circumstances were associated with individual diseases and with the organisms implicated in disease causation. Given the focus on infectious diseases, such as malaria, and the relative decline at the time in the prevalence of such diseases in the more developed world, the locus of practice of medical geography tended to be in less developed countries.

Undoubtedly disease ecology was and remains an effective approach to the study of disease. It places medical geography in an interesting position at the intersection of physical and human geography. With hindsight it has much to offer regarding insight into reemergent infectious disease in more developed countries while cementing an important commitment to health development in the global south. Recent epistemological developments in political ecology and more generally in the study of the interaction of society and nature have provided grounds for integration into mainstream geography. Yet 20 or more years ago disease ecology might arguably be said to have provided one of the three contexts necessary for the emergence of health geography.

The rationale for this assertion lies in the practice of disease ecology. First, though it would be unjust, a simplistic analysis of disease ecology could easily equate it with a naive environmental determinism and an interesting but rather repetitive mapping and cataloguing of yet more specific ecologies. Second, when practised effectively, disease ecology requires an increasingly unlikely facility with advanced knowledge across physical and human geography and into such complex areas as parasitology; its synthetic imperative is both academically challenging for the individual and challenging to the increasingly specialist academy. Third, it had limited appeal for researchers conducting work on noninfectious disease; this tended in turn to limit its relevance to researchers working in less developed settings, though there were important exceptions. Fourth, the concern with disease located the perspective clearly within the medical model despite its considerable concern with the contexts within which disease occurred. Together, these concerns meant that researchers working in human geography contexts in more developed countries, looking to conduct research more locally, and concerned to distance themselves from an unfashionable (and unfair) environmental determinism, sought actively to reframe the subdiscipline away from what had long been its dominant paradigm.

The second necessary context for the emergence of health geography concerns the rediscovery of inequality in health. This was particularly associated in the UK with a recognition that the state funded comprehensive national health service (NHS), made available free at the point of use as part of the post World War II welfare settlement had not resulted in reductions in past variations in health status by occupation or other social divisions. Elsewhere the rediscovery was associated with similar crises in confidence in healthcare systems and their health outcomes. Two aspects of the rediscovery of inequality are relevant to the emergence of health geography. First, explanations for inequality were and remain highly contested. At the risk of oversimplification, two candidate explanations stressed individual level explanations: a behavioral thesis linking inequality to the raised prevalence of factors like smoking and poor diet among poorer people, and a more medicalized argument emphasizing processes of social and natural selection. For health geography however, alternative structural explanations of inequality were more important. These sought to explain health inequalities by way of structural divisions in society, notably the constraining effects of differential experiences of poverty on the wider upstream determinants of health. These structural arguments, espoused by most if not all health geographers, fostered an interest in the wider determinants of health; the behavioral arguments led to an interest in how health related behaviors can be affected by structural factors. The interest in health related behaviors steered the focus of interest away from a strict concentration on disease and death. This shift was also evidenced by a growing interest in the nature of health inequality in relation to self rated subjective health and other concepts less directly associated with particular diseases or causes of death.

The third factor in the initial metamorphosis of medical geography into health geography was the growing body of work on healthcare. Interest in the geography of care provision had largely but not exclusively emerged in the period following World War II. The momentum of this research had increased in the late 1960s with the emerging crisis of the welfare state and geography's radical turn. By the mid 1980s the general turn to health identified earlier was leading healthcare geographers away from a focus on the traditional medical elements in healthcare provision. They turned, on the one hand, toward a more generalized interest in health policy and its spatial aspects, engaging with political geography. Examples would include work on the broad sweep of English NHS policy, its increasing engagement with the private sector, and its implications for the regionalization of a national public service. Other geographers developed an increasing concern with nonmedical aspects of care provision. Examples would include work on care provision by volunteers and studies of the geography of complementary/alternative provision.

Moments

In the decade from the mid 1980s to the mid 1990s a series of publications signalled the emergence of health geography and the purported transformation of medical geography into a newer, more contemporary, subdiscipline aligned better with mainstream geography and social science. These 'moments' built upon the contexts noted above and articulated transformative claims regarding the development of health geography.

Early syntheses and developments of the contextual background were evident in two important works from the mid 1980s: Eyles and Woods, *The Social Geography of Medicine and Health* (1983) and Jones and Moon, *Health, Disease and Society* (1987). The first, essentially an extended monograph, offered a critique of existing approaches to medical geography and sought to move beyond the biomedical to a social model and thus, in accordance with its title, integrate the subdiscipline with social geography. It presented health and particularly healthcare as subject to a series of constraints: historical, social, and reflective of the norms and ambitions of the medical profession. Social justice and the inequities consequent upon differential power positions in the health encounter run motifs throughout the work. Alongside these welfarist notions there was also a strong commitment to a more individualized understanding drawing on experiential, behavioral, and phenomenological approaches.

Health, Disease and Society addressed a broader audience and was also more direct and fundamentalist in its position. It brought new ideas to a student audience as well as to researchers. Drawing often on a sometimes rather crude Marxist position, it drew attention to the social production and construction of ill health and articulated a critical perspective on both established epidemiological approaches and on the purpose and direction of health care provision and policy. In the latter case, it too espoused social justice. The key objective was the need to look outside the body and develop an alternative social and environmental perspective on health. Its reception was mixed although it subsequently sold well and continues to be a point of reference. For some it was naïve in its radical political stance; for others it was overly selective in its deconstruction of established perspectives. Overall however it served to introduce novel concepts to a wide audience; putting health rather than medicine at the start of its title was deliberately symbolic. With the book by Eyles and Woods, it codified the changing nature of the subdiscipline.

Early texts represent the first moment in the emergence of health geography. The second moment was a debate in 1994 initiated by Robin Kearns in the journal *The Professional Geographer*. He advocated a refocusing of the subdiscipline building on a growing interest in the impact of place on health. Work at the time has sought to identify the sociological meaning of place and apply it to health matters through notions such as community or through case studies of the specifics of health in particular places. Kearns' plea was to take these ideas further through a quantum shift in the focus of medical geography toward a cultural/humanistic standpoint. In this respect he echoed earlier work by behavioral geographers and also linked the geography of health into the developments that were at the time beginning to engulf cultural geography.

Kearns saw his cultural and humanistic perspective as a route that would lead to a 'post medical geography of health' focusing on a broad definition of health encompassing particularly the ideas of well being and wellness. His post medical challenge drew on behavioral and humanistic geographies and, for some, did not really grasp the more fundamental challenges being laid down by the cultural turn in the social sciences. It lacked a thorough going theorization in the post structuralist sense and missed the opportunity to engage with the then emerging literature on the sociology of the body. Its humanistic position was certainly a potential direction for study but it carried with it its own limitations. Other critiques were more concerned that Kearns was suggesting too radical a shift from the still prevailing orthodoxy of medical geography; the notion of a post medical geography signalled a separation from the bio medical model and a definitive relocation of health geography in the social sciences. For those coming from this more conservative position, Kearns was seeking to move the subdiscipline dangerously far from its traditional ground. In particular, it was suggested that his critique could be challenged by drawing on new developments in disease ecology and by reference to the applied utility of traditional perspectives for the understanding of disease in less developed settings.

A third moment or watershed in the emergence of health geography was the launch of the journal *Health and Place* in 1995. Though it was always planned as an interdisciplinary journal, drawing work from public health and sociology as well as geography, geographers played the lead role in the development of the journal. The central concern of the new journal was to be studies where place matters with regard to health, healthcare, and health policy. It sought papers that embraced a wider definition of health, including ideas of wellness and non medicalized forms of healthcare. It also anticipated papers exhibiting both a greater theorization of its subject matter and also a commitment to methodological innovation. In many ways, its launch represented an explicit distancing from the concerns that had been traditionally the major outlet for research by medical geographers: *Social Science and 'Medicine*.'

In practice the changing nature of research in health geography and in allied disciplines meant that the distinctiveness of *Health and Place* with regard to *Social Science and Medicine* was never as clear as might have appeared possible in the mid 1990s. From the outset there were strong links between the two journals and, in effect, they came to provide alternative outlets for similar material. *Social Science and Medicin*e opened up to more theorized and less biomedical work while *Health and Place,* from the outset, tempered its radical aspirations by announcing an intention to publish papers employing more traditional approaches alongside a commitment to newer perspectives. Papers published over the succeeding decade and more have thus matched a more medical, albeit public health, emphasis, with work on wellness.

The Practice of Health Geography

Commentaries prepared for special editions focusing on health geography that have appeared in journals over the past 20 years have been unequivocal in their claims that the nature of the subdiscipline has changed and continues to evidence a process of ongoing development. These sentiments are backed by the content of the annual reviews of work in the subdiscipline that appear in *Progress in Human Geography*. The former output has stressed a move away from the separation of health and healthcare geographies and an increased awareness that place matters in understanding health outcomes. They have also indicated an enhanced awareness of the impact of difference on health, most clearly in bringing in an increased consideration of gender and impairment. The latter have tended to be framed as agendas for further change rather than statements of record of research conducted over the past year. They have looked outside the subdiscipline, and indeed outside geography itself, at areas in which health geography might frame knowledge or where such a framing had already begun. Examples would include early reviews of work on the epidemiological transition noting its space–time specificity, its political implications, and its neglected complexity, and later moves to introduce health geography more effectively to hitherto little used ideas of embodiment in order to understand better the psycho social context of disability and mental ill health and the emerging use of psycho analytic ideas in geography.

Place-Awareness

Beyond its concern with a broad post medical conception of health, any assessment of the practice of health geography in the aftermath of the critical moments identified in the previous section would almost certainly emphasize a focus on the construction of place. Geographies of health have demonstrated that a nuanced understanding of the complex nature of place is central to an effective understanding of health and its geographical variation. Place has been seen as an experiential construct that is negotiated, flexible, contingent, and socially constructed; research has sought to move away from seeing place as simply the location where something happens and gets recorded: an unproblematized activity container. This enhanced focus on place has sought to show that 'places matter' when it comes to understanding health and healthcare.

Three rather different approaches to this objective can be identified. First, a set of studies have worked with a landscape metaphor. This research, borrowing from both economic and cultural geography, has demonstrated how the production of place specific landscapes of healthcare can result from an interplay of cultural and politico economic factors. There have been landscapes of despair, landscapes of restructuring, and, most notably, therapeutic landscapes. These ideas generally refer both to literal localities and also to theorized notions of place constrained and place enabled interactions between history, social structure, and built environment. A second form of place sensitivity is evident in studies using multilevel modeling methods. This work serves as a reminder that health geography need not be qualitative. Broad ideas of health are equally amenable to quantitative analysis. Multilevel perspectives have been particularly influential in the development of the debate over composition and context; they have recognized the hierarchical organization of much of the social background to health and enabled assessment of the relative importance of people and place characteristics in bringing about health outcomes. Critics argue however that the 'places' that multilevel models capture have little sociological significance and tell us more about data collection strategies than the realities of place effects. A final set of place sensitive studies are those grounded in the specifics of particular localities. These locality studies include work on community responses to threats to health and studies of the place specific aspects of health service restructuring. They have places ranging in scale from the home to the region as their focus and exhibit epistemological underpinnings drawing on positivist, political–economic, and humanist traditions. Arguably the latter has allowed the most effective contributions, invoking consideration of the experience of both literal place and perceived place in the world.

Landscape, multilevel, and locality approaches to place are very different. Landscape approaches come closest to the core ideas of health geography. They are cultural, theoretical, and de medicalized. In Wilbert Gesler's therapeutic landscapes, health geography has a reasonable claim to have developed an original extension of the landscape metaphor. Multilevel perspectives are perhaps best seen as being an effective but limited quantitative methodology for addressing questions of place. Yet they too provide a clear example of an approach to place sensitivity where a development within health geography has had an impact on other parts of the geographical discipline. In locality studies, it is the place where the study is undertaken that is itself important. The extent to which place matters can vary yet is often limited. The evidence thus points to a situation where place, though undoubtedly a focal concept in health geography, is neither unproblematic nor coherently applied. Furthermore, the nature of its application has varied over time and space. Landscape approaches came late to the United Kingdom in comparison with multilevel perspectives. The reverse happened in New Zealand. A strong tradition of locality studies is evident in Canada.

Critical Awareness

Alongside place and health, there has also been an assumption that health geography is in some way more critical and/or radical than medical geography. A critical perspective might be expected to involve opposition to unequal and oppressive power relations, and a commitment to social justice. In fact such ideas are often also evident in medical geography, though perhaps less overtly so. In contemporary health geography, the oppositional stance is most clearly present in the continuing stream of work on health inequality. Surveys have suggested that around 20% of papers in *Social Science and Medicine* and *Health and Place* by health geographers over a 15 year period were on this theme. It continues as a key research theme, not least through the extension of work on relative inequalities into consideration of health related behavior and subjective health indicators. The subdiscipline's commitment to social justice and transformative politics is evident in the reconstruction of healthcare as consumption and patients as consumers. Work on problems of access to healthy food, and on the restructuring of healthcare in rural areas provides examples. Nor need a commitment to social justice be evidenced solely at a local scale. The work on health inequalities again provides exemplification: critical analyses of secondary data have been used to challenge discourses of progress and development by national health services.

Closely connected to critical awareness has been the gradual shift in the construction of 'the field' for health geographers. In this sense a subject specific criticality has been paralleled by an emerging critical methodological stance. To this end there has been a move away from seeing the field as a place of study with the researcher as an outsider expert collecting evidence and uninvolved with the lives of those being researched. Feminist scholarship in particular has led some health geographers to recognize their positionality within in the environments that are being studied. Work on mental health and on disability has been particularly impacted by this turn. As a consequence, recent cultural geographies of health have less frequently been outsider accounts of the exotic and more often critical interpretations of experiential worlds.

A third theme worthy of consideration under the heading of critical awareness concerns the interface of the subdiscipline with medicine. In its early days, the then Medical Geography Study Group of the IBG co opted a medical doctor to assist in building links with clinical medicine. Nowadays, the interest is no longer solely with medicine. Links with other disciplines demonstrating interests in health are seen as equally important. This can be seen as part of the process of de medicalizing health geography. Medicine now sits alongside other disciplines as an equal, not a superior. This process of change is however as much about changes in medicine as it is about changes in geography. Geographers have tended to have a rather outdated view of medicine; the social model of health has had a significant impact on medicine itself. Nor however has the distancing from the medical profession been complete; few health geographers would deny that the continuing power of the medical profession can make an element of collaboration a necessity. Significant numbers of papers continue to feature co authors drawn from both medicine and health geography.

Perhaps the most appropriate verdict on the critical aspect of health geography is that it is critical lite. There is little evidence of fundamental radical work. In part this can be linked to the close association between health geography and policy. In part it also reflects a subversion of the original ambitions for health geography by neoliberal imperatives. The extensive policy changes of the past 20 years have been fertile ground for research and particularly for funded research at a time when public funding for curiosity driven cultural research has not always been high. The need to raise funding for research has, to an extent, driven agendas toward less critical but policy relevant areas. The continuing though reduced links between health geography and medicine can also be seen in terms of ensuring better access to research funds. In competing, often successfully, for such resources, empirical health geography has grown at the expense of criticality. Medical geography always had a strongly utilitarian orientation and the increased profile of health geography in recent years is thus, to an extent a consequence of maintaining that tradition.

Theoretically Aware

Disease ecology was and remains a theoretical standpoint that examines the relationship between health and environment. Social epidemiology is replete with theoretical assumptions on the nature of causation. There is a long established application of ideas such as central place theory in spatial analytic medical geography and much work in the geography of healthcare pursues welfarist or even Marxist theory. Yet health geography explicitly presents itself as more theoretically aware than medical geography. This claim is in large part a reflection of a desire to move away from established theoretical positions and embrace new ideas in cultural–social theory as they emerged in the early 1990s. The aim was to position health geography alongside developments elsewhere in geography and in other (social) sciences; geographies of health were, in part at least, medical geography's cultural turn.

A move to sociocultural theory was an issue in *The Professional Geographer* debate alluded to above and other contemporary work also commented on the relative poverty of social–theoretical engagement in the geography of health. There was a need to recognize the social–theoretic context of health and of health related subject matter. Work taking place on the interface between cultural and social geography using the landscape metaphor was particularly drawn to an engagement with theory. For example, research on the landscapes of private medicine sought to theorize consumption and create new understandings of the consumers of commodified healthcare. Similar links were also evident in research on mental health with the Foucauldian theory being particularly prominent. Other work variously used perspectives derived from interactionist sociology, illness behavior studies, ethics, and queer theory. Recently, writing by geographers working at the permeable edges of health and social geography has remedied a relative lack of engagement with the literature on the body, although it might have been expected that a shift toward geographies of health would have been accompanied by a greater interest in healthy bodies. Such work is largely conducted outside health geography although there is now an emerging corpus of research from within the subdiscipline on the specific consequences for health of body adornment, maintenance, inscription, and management. Finally there has been considerable work drawing on theories of risk. This has been evident in research on mental health and environmental health as well as in studies of health related behavior.

The espousal of new directions in sociocultural theory was accompanied by substantial theoretical continuities in rather different directions. First, some health geography retained a strong allegiance to positivism. A search for law like regularities was evident in work on place sensitivity using multilevel modeling as well as work on environment–health interactions. It was fostered in the UK by the impact of the Economic and Social Research Council (ESRC) program of research on health variations, which encouraged empirical if not necessarily positivist research. Second, particularly in the healthcare field, there was a continued emphasis on welfare theory and notions of equality/inequality and exclusion/inclusion.

In summary, health geography's engagement with theory was typically diverse but underpinned by a post positivist approach that sought to avoid universal truths. Arguably the key framework employed by health geographers to link together these diverse perspectives has been the structuration theory. This has the capacity to integrate people and places as well as the local and the global, and facilitates generalization outward from microlevel case studies. It can also be used to confront perhaps the key challenge facing future theorized health geographies – the incorporation of time. To confront this challenge health geography is now moving beyond snapshot studies of single periods of time toward repeated cross sections and genuinely longitudinal studies. Though the terminology is quantitative, the methodology need not be: while multilevel modeling has been used to examine self reported health over time, there have also been studies of individual's trajectories through care systems and their journeys to care.

Case Studies

Two case studies serve to exemplify current work in health geography. The first takes a quantitative approach to a topic central to the broad de medicalized approach to health – the geography of smoking. The second is more qualitative and is arguably the focal concept within the geography of health – therapeutic landscapes.

Geographies of smoking

Smoking is a health related behavior. Few would dispute its effects on health. It exemplifies the move within health geography away from the study of disease and death. Its reduction is of central importance to health promotion and to the promotion of enhanced population health. It is measurable in a quantitative sense, albeit in many different ways and with variable precision and accuracy, and is thus amenable to statistical modeling. A significant body of work has used smoking as an outcome variable in multilevel models that have sought to unpack the relative importance of composition (individual characteristics) and context (place characteristics) as factors influencing levels of smoking and the likelihood of smoking. Other work has explored the effect of ethnic segregation on smoking.

Turning first to the multilevel work, debate has centered on whether the character of places plays an independent role in shaping individual smoking behavior. Smoking behavior has to be understood not only in terms of individual characteristics, but also in relation to local cultures. To operationalize the structuration theory in a quantitative framework it is essential to recognize that places matter, and the context in which behavior takes place is important both for understanding and for subsequent policy development. At the small area scale, particular attention has focused on whether neighborhood deprivation can predict an individual's smoking status independent of that person's individual demographic characteristics (age, sex, etc.). Results suggest that measures of neighborhood deprivation have an independent effect on individual smoking status even when account is taken of a large range of individual characteristics, both as main effects and interactions. This suggests that the character of the local neighborhood plays a role in shaping smoking behavior. Early multilevel work in both the UK and Sweden challenged this conclusion but subsequent studies demonstrated that place effects, as least as revealed within an extended multilevel framework, are complex, operating at multiple scales and reflecting interactions between people and place characteristics. Further work extending such analyses has enabled the calculation of small area estimates of smoking (and drinking) behavior (**Figure 2**).

Work on smoking and ethnic segregation in New Zealand has focused on understanding the considerable differences between Pakeha (New Zealanders of European origin) and Maori smoking prevalences. Initial work tested is a generalization of the thesis that the greater a nation's income inequality, the poorer the average national health status. Analysis of Maori and Pakeha smoking rates from the 1996 New Zealand census disaggregated by gender and rural–urban location revealed that relative social inequality between Maori and Pakeha had an independent effect on Maori smoking rates after controlling for absolute levels of inequality. Pakeha smoking rates in contrast were more sensitive to variations in absolute deprivation. The effect of inequality was greatest for Maori women, especially among urban residents. By contrast, among Maori men the effects were greatest in rural areas. A multilevel analysis of the same data confirmed the importance of ethnicity and deprivation but challenged crude stereotypes about the apparent geography of smoking, suggesting that, places with high levels of smoking were less high than might be expected given the socio demographic composition of the areas concerned. Subsequent work incorporated a temporal dimension. While Maori smoking rates declined between 1981 and 1996, the gap in smoking levels between Maori and Pakeha increased; Pakeha were more inclined to quit smoking. This suggests that levels of social inequality between Maori and Pakeha during what was a period of considerable economic decline, had an independent effect on Maori smoking rates and that communities which experienced increased social inequality during both the 1980s and 1990s were more likely to have higher Maori smoking rates.

Therapeutic landscapes

The initial development of the idea of the therapeutic landscape by Wil Gesler in 1992 sought to expand the meaning of the concept of landscape as employed at the time in cultural geography. Certain places were perceived to be beneficial to health. What were in effect experiential, even spiritual, place effects built on humanist and psychological conceptions of sense of place and the symbolic value of particular environments (**Table 1**). Research on therapeutic landscapes has made a significant contribution to unpacking the role of place in health and is central to the development of health geography in its focus on positive aspects of health such as wellness and well being. A fuller entry is available elsewhere in this work. Here the basic contribution of the idea is outlined.

Early work explored some of the more obvious settings where health and place are perceived to interact in this therapeutic way. Examples would include Gesler's

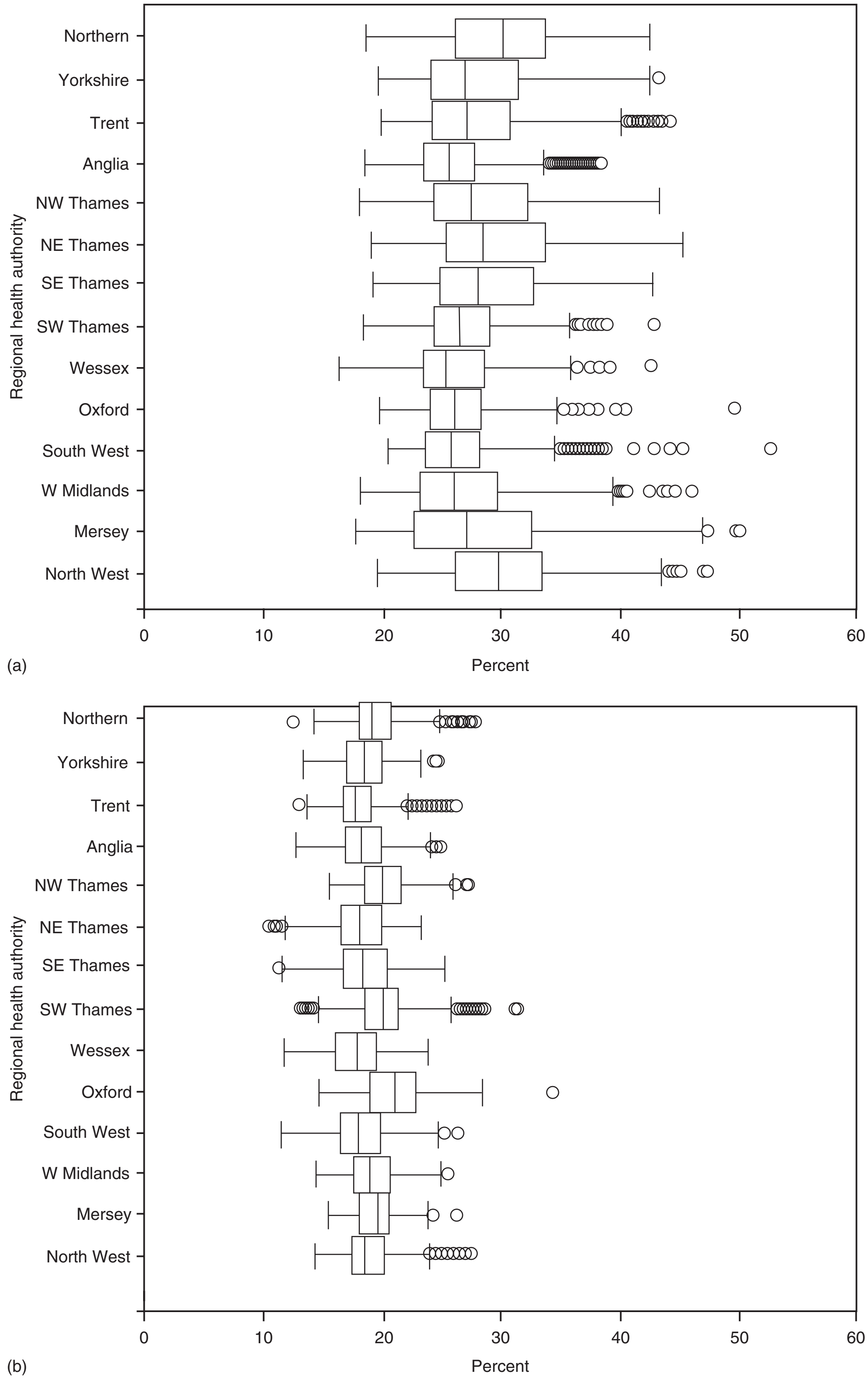

Figure 2 Multilevel estimates of (a) smoking and (b) drinking prevalences in England. From Twigg, L., Moon, G. and Jones, K. (2000). Predicting small-area health-related behaviour: A comparison of smoking and drinking indicators. *Social Science and Medicine* 50, 1109 1120.

Table 1 Exemplifying the humanistic foundation of therapeutic landscapes: The case of holistic medicine

Concept	*Health applications*	*Association with holistic medicine*
Symbolic landscapes	Medical semiotics	*Basic element in therapy, for example, the aboriginal medicine wheel
Importance of meaning, value, and experience	Medical beliefs	*Incorporation of individual and/or cultural belief systems, for example, new age practices
Sense of place	Psychological rootedness	*Belief that environment holds meaning, significance, and felt value, for example, planned home-birthing
Authentic and unauthentic landscapes	Caring environments	*Formation of environments through human networks of care, for example, aboriginal hospitals and birthing centers
Landscapes of the mind	Fields of care	*Imagery used in accessing healing environments, for example, visualization
Hermeneutics/landscapes as text	Tapping into the sensory experience	*Interpretation of health conditions, for example, art therapy

From Williams, A. (1998). Therapeutic landscapes in holistic medicine. *Social Science and Medicine* 46, 1193 1203.

own works on Epidauros (Greece), Bath (UK), and Lourdes (France), sites of secular or spiritual pilgrimage associated with healing springs. Later work significantly extended this focus. The home/dwelling has been pre sented as not only a dwelling space but also a place with multiple meanings: shelter, belonging, security, privacy, and so on. Informal care giving is performed in this spatial context and enhanced by it under certain con ditions. Further increases in the diversity of subject matter being constructed as a therapeutic landscape have seen studies ranging on wilderness and healthcare facil ities and reflection on the symbolic importance of ther apeutic landscapes, and their relevance for different social groups in terms of access, behavioral expectations, and normative definitions of therapeutic. Foucault's no tion of governmentality has been applied in a study of an alcohol and drug addiction recovery centre, understood simultaneously as a site for recovery and for the sur veillance and governance of health related conduct. Other work has seen gardens framed as therapeutic landscapes. Gardening, it is argued, functions both as a leisure activity and as a way of harnessing the healing power of nature for emotional, physical, and spiritual renewal. Particular attention has been paid to communal gardening and gardening by older people. At the other end of the age spectrum, children's health camps in New Zealand have been deconstructed as therapeutic land scapes where child and family welfare is enhanced but where changing understandings of health and children have been closely linked with changing perceptions of what is therapeutic. The key lesson from these last two examples is that settings are not intrinsically therapeutic; they are experienced in very different ways by different people at different stages in the life cycle and reflect both shared interpretations of place and personal con structions of self in place.

Nor is research on therapeutic landscapes limited to developed, western settings. Research in Gambia has considered the way in which indigenous medical beliefs and practices highlight the importance of place, and cultural symbiosis between indigenous health and the local environment is presented as a particular type of therapeutic landscape. Elsewhere the idea of the ther apeutic landscape has also been used to demonstrate the importance of places for maintaining physical, emotional, mental, and spiritual health among First Nations people in North America. Finally, a small body of research has stepped completely out of the real world and traced the therapeutic landscapes evident in fiction, in the tuber culosis sanatoria of Thomas Mann's *Magic Mountain*, in the anti therapeutic wanderings of Holden Caulfield in Salinger's *The Catcher in the Rye*, and the allegorical re demptive landscapes of Balzac's *Country Doctor*.

Conclusion

What's in a name? Without dwelling overly on what health/medical geographers have actually done, an en during debate has been that the name does not really matter. Health geography and medical geography are virtually indistinguishable. Some geographers prefer one name, some the other; many interchange at will without much concern for the type of work that they are doing, and several influential geographers study health (and disease) and are unimpressed by either label. Yet, as discussed above, there were clear differences in intent when health geography first emerged as a terminology and, though it is by no means clear cut, there is a real sense in which health geography when practised in its purist form has broken away from disease and death and focused upstream on broader definitions of health and wellness. The diversity of interests within contemporary health geography is marked in comparison with the medical geography of the past with its stress on disease ecology and, more lately, disease mapping and healthcare provision.

Of course medical geography has also changed and continues sitting alongside the novelties in the ongoing joint project to reinvigorate geographical studies of health and healthcare. It is tempting to conclude by seeing health geography as a project in the cultural sense of the term. While some define themselves as within the project, others do not. The endeavors of all, however, contribute to the reinvention of the subdiscipline. Even those opposed to the tenets of the project, those effectively within the old project of medical geography, benefit from the enhanced visibility that the articulation of the new health geography project brings. Through innovative developments in method (e.g., multilevel modeling) and theory (e.g. therapeutic landscapes), health geography has metaphorically put itself on the map within both geography and the wider health social sciences with regard to the understanding of place/health relations.

What directions might future health geographies take? As yet it is clear that health geography is a work in progress. Much of the early ambition is yet to be realized and concepts require further clarification, development, and replication. The commitment to a broad upstream approach to health is the cornerstone for further work. Thus it is reasonable to expect a continued focus on health related behavior, self assessed health, and the de medicalization of care, particularly with regard to the emotions and psyche. Beyond this a small selection of issues needing further work includes:

1. the incorporation of a temporal dimension to analyses of place effects on health;
2. a thoroughgoing deconstruction of ideas of social capital and their impact on health;
3. placing health in a critical economic context, considering, for example, the role of the food industry in the promotion of health; and
4. seeking further sophistication in multilevel models while paying heed to data limitations and addressing the need for enhanced sociological meaning.

See also: Body, The; Care/Caregiving; Disability and Chronic Illness; Disease Diffusion; Emotional Geographies; Environmental Hazards; Health Inequalities; Housing, Neighborhoods and Health; Medical Geography; Mental Health; Social Capital, Place and Health; Therapeutic Landscapes; Wellbeing.

Further Reading

Dorn, M. and Laws, G. (1994). Social theory, body politics, and medical geography: Extending Kearns' invitation. *The Professional Geographer* 46, 106 110.

Duncan, C., Jones, K. and Moon, G. (1998). Context, composition and heterogeneity: Using multilevel models in health research. *Social Science and Medicine* 46, 97 118.

Eyles, J. and Woods, K. (1983). *The Social Geography of Medicine and Health*. London: Croom Helm.

Gesler, W. M. (1992). Therapeutic landscapes: Medical issues in light of the new cultural geography. *Social Science and Medicine* 34, 735 746.

Gesler, W. M. and Kearns, R. A. (2002). *Culture/Place/Health*. London and New York: Routledge.

Jones, K. and Moon, G. (1987). *Health, Disease and Society: A Critical Medical Geography*. London: RKP.

Kearns, R. A. (1993). Place and health: Towards a reformed medical geography. *The Professional Geographer* 45, 139 147.

Kearns, R. A. and Moon, G. (2002). From medical to health geography: Novelty, place and theory after a decade of change. *Progress in Human Geography* 26, 605 625.

Mayer, J. D. and Meade, M. S. (1994). A reformed medical geography reconsidered. *The Professional Geographer* 46, 103 106.

McIntyre, S., McIver, S. and Sooman, A. (1993). Area, class and health: Should we be focussing on people or places? *Journal of Social Policy* 22, 213 234.

McKeown, T. (1979). *The Role of Medicine: Dream, Mirage or Nemesis* (2nd edn.). Princeton, NJ: Princeton University Press.

Moon, G. (1995). (Re)placing research on health and health care. *Health and Place* 1, 1 4.

Parr, H. (2002). Medical geography: Diagnosing the body in medical and health geography, 1999 2000. *Progress in Human Geography* 26, 240 251.

Smith, S. and Easterlow, D. (2005). The strange geography of health inequalities. *Transactions of the Institute of British Geographers* 30, 173 190.

Williams, A. (ed.) (1999). *Therapeutic Landscapes: The Dynamic between Wellness and Place*. Lanham, MD: University Press of America.

Health Inequalities

D. Dorling, University of Sheffield, Sheffield, UK
R. Mitchell, University of Glasgow, Glasgow, UK
S. Orford, Cardiff University, Cardiff, UK
M. Shaw, Bristol University, Bristol, UK
H. Tunstall, University of York, York, UK

Glossary

Common Identity It refers to similar residential location, occupation status, income, or similar trait of social significance.
Equality The quality of being the same in quantity, or measure, or value, or status.
Health Inequalities The lack of equality between the health status of social groups of people.
Home Counties Some of the most areas in Britain around Greater London where colonial officers in the past often returned home to.
Social Group A set of people that have some common identity.
Social Significance It is of importance to the organization of people that is not expressly stated but can be inferred – for instance in relation to health inequalities.

Introduction

Health inequalities, as one government minister in Brit ain recently stated, are 'the most important inequalities of all'. In the United States of America there remains great confusion and the rather more innocuous term 'variation' is often used. For instance typing 'health in equalities definition' into Google in May 2006 produces:

> This is the term commonly used in Europe to indicate the virtually universal phenomenon of variation of health by socioeconomic status, that is poorer people have poorer health. In the US, there is no single such term, and instead it is referred to as the socioeconomic status and health relationship. (http://depts.washington.edu/eqhlth/pages/academic resources/glossary.html; accessed April 2009)

Health inequalities are essentially a lack of equality between the health status of social groups of people, where social groups have a common identity such as similar residential location, occupational status, income, ethnic group, or other similar trait of social significance. These inequalities can be found in many types of health status ranging from the chances of dying at a particular age to differences in the degrees of well being measured in social surveys.

Measurement

To understand them, health inequalities need to be quantified or else they will remain a purely abstract notion. There are many ways in which health inequalities can be measured numerically. Researchers are beginning to conclude that the best way is to use the Slope Index of Inequality (SII) for measuring absolute differences be tween the status of groups and the Relative Index of Inequality (RII) for measuring relative differences. Low and Low provide more details on these measures. Note that the difference between absolute and relative is not simple when measuring health inequalities because some measures – such as life expectancy – do not lend them selves to be easily categorized. Furthermore, what is worth more: 10 years of life from age 80, or 5 years from age 40?

Social Inequalities

Health inequalities are partly a reflection of social in equalities more widely defined. They also greatly con tribute to those inequalities. Illness prevents millions of working age people from working in affluent countries and is often the precursor to absolute destitution and premature mortality in poorer countries. Nowhere can such inequalities be justified as emphasized by the dead body of a child at the foot of **Figure 1**.

The health inequalities suggested by this poster per sist and are found across the social hierarchy. Health inequalities are not only a reflection of the poor health of the most disadvantaged people but also the apparently limitless health benefits associated with rising socio economic status. Most analysis of the associations be tween health and socioeconomic status suggests there is no threshold at which greater individual socioeconomic status is no longer associated with greater health. Even when people within high socioeconomic status groups in developed countries are compared, those with higher relative socioeconomic status have better health. For example, a long running analysis of British civil servants,

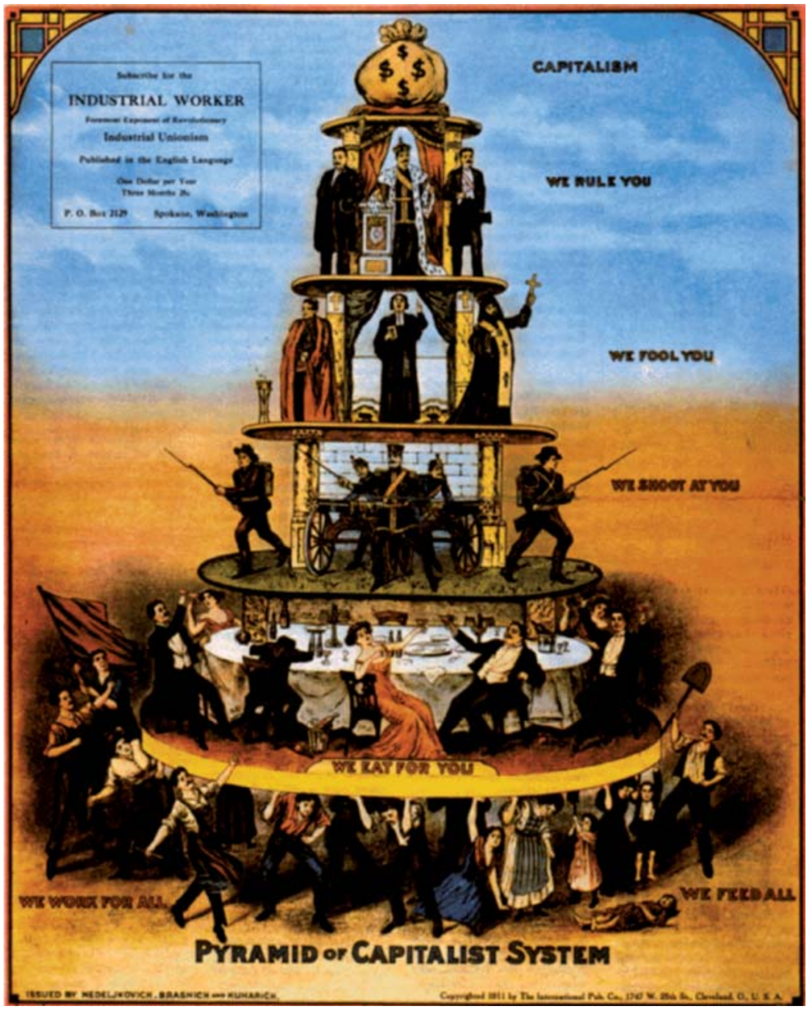

Figure 1 Pyramid of the capitalist system.

the 'Whitehall Studies', has found that employment grade among these professionals is a strong predictor of health outcomes across all levels of their hierarchy.

Income Inequalities

One current debate concerning the causes of health inequalities in rich nations is the extent to which such inequalities are a reflection of income inequalities. It is argued that health inequalities are not simply reducible to income inequalities and that they may also be further exacerbated by them. (See the article by Wilkinson and Pickett in the 'Further reading' section and **Figure 2** for an example of the possible contextual effects of income inequality.)

In **Figure 2**, metropolitan areas of the United States are categorized by their average per capita income into four quartile groups; the relationship with mortality is such that higher income is associated with lower mortality rates, although the relationship is not linear as the rates for the lowest income group (green) are actually higher than those for the 2nd quartile (blue). When income inequality is also considered, we can see that within all income quartiles higher income inequality is associated with higher death rates. Hence, it is not only absolute income that matters for health, but also the distribution of income within society.

Area Inequalities

Stark inequalities are found when comparing health between different countries. See the 'Further reading' section and especially the WHO *The World Health Report* and its annex for many of the statistics of inequality worldwide.

A significant amount of research from geography, epidemiology, and public health shows that where people live greatly affects their health outcomes. This work also includes comparisons of inequalities within regions of the

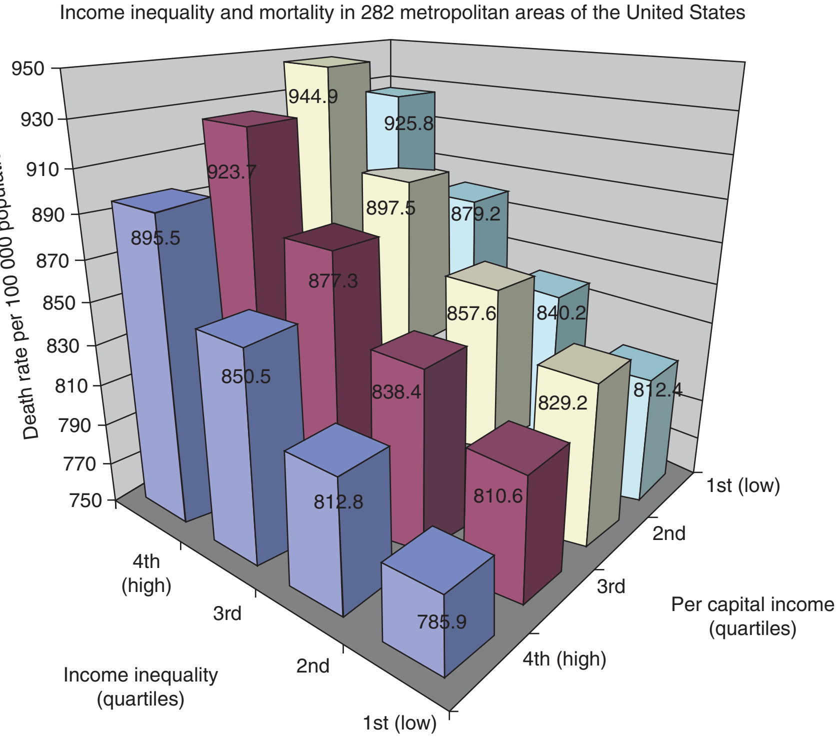

Figure 2 Income inequality and mortality in the United States. Source: Lynch, J. W., Kaplan, G .A., Pamuk, E. R. *et al.* (1998) Income inequality and mortality in metropolitan areas of the United States. *American Journal of Public Health* 88(7), 1074 1080.

world, reports of within country inequalities in health outcomes, and inequalities in health outcomes within specific localities.

Geographical studies have historically concerned themselves with the spread of infectious diseases or the possibly harmful effects of proximity to potentially health damaging sites of pollution exposure. However, in recent years more research has focused upon chronic diseases. This is mainly because chronic diseases are the main cause of premature death within affluent countries and affluent countries fund almost all research on chronic diseases. The key exception has been HIV/AIDS re search upon which funding was concentrated until it was realized that the pandemic was unlikely to have its worse consequences within rich nations.

Geographical analysis of health has to deal with methodological uncertainties as well as social and political priorities. Methodological uncertainties are caused by issues of ecological fallacy, scale, the modifiable areal unit problem (MAUP), and spatial autocorrelation. The eco logical fallacy refers to problems inherent in the inference of group or area characteristics as individual. An example of this would be to assume that in an area of high levels of illness containing many teenage mothers, that teenage mothers in that area will have high levels of illness.

There are also issues of scale that need to be born in mind when considering health inequalities. These are related to both the size of the units of analysis – whether they are local, regional, or national – and to aspects of the MAUP referring to the choice of such units and how this reflects the relations observed. Spatial autocorrelation simply refers to the fact that many phenomena are spa tially dependent – ill people tend to be located near other ill people. All these issues can be summarized as (1) making assumptions about people given their locality, (2a) being concerned about the size of places in studies, or (2b) how places are constructed, and (3) how they are interrelated.

Recently, methodological debate has revolved around research seeking to separate 'area effects' from those 'compositional' effects of simply aggregate population characteristics. The debate over the importance of in come inequalities referred to above is one example of such research. Often the statistical technique of multi level modeling has been used. Unfortunately, charac teristics of places are typically distilled in this kind of analysis to simply a few limited variables; 'area effects' are sometimes not obvious and when found, tend to be small. This analysis of 'area effects' frequently fails to conceptualize 'area' – is it home, street, neighborhood, workspace, society, or where? Also, what is meant by 'effects' is often left ambiguous – what are the causal pathways by which place effects health? Place is often considered a black box (of varying sizes and shapes) in which unidentifiable 'nonindividual' processes take place.

Spatial Sorting

Geographical inequalities in health can be exacerbated by spatial sorting. That is, if the population migrates in such a way that people with better health move toward areas where better health is enjoyed and those with poor health remain, then health inequalities between areas will grow, especially over the course of many years. **Figure 3** shows an image of the propensity of people to migrate by social class between different parts of Britain a generation ago.

The map in **Figure 3** shows all the main migratory flows that occurred between the wards of England and Wales between the years 1980 and 1981. A line is drawn between any wards between which individuals and families moved and the width of that line is drawn in proportion to the number of people who moved. A population cartogram is used as the base map and thus the amount of color is proportional to the numbers migrating. The lines are colored according to the main social class of those moving: blue for affluent toward red for poor. The blue lines tend to be longest – for instance, from around the Home Counties down to the Southwest of England as generally healthier affluent retirement migration flowed. Often the red and orange lines are constrained within the boundaries of what are local authorities as many people moved then within what was called council housing – and did not leave their borough upon retirement if they could not afford to.

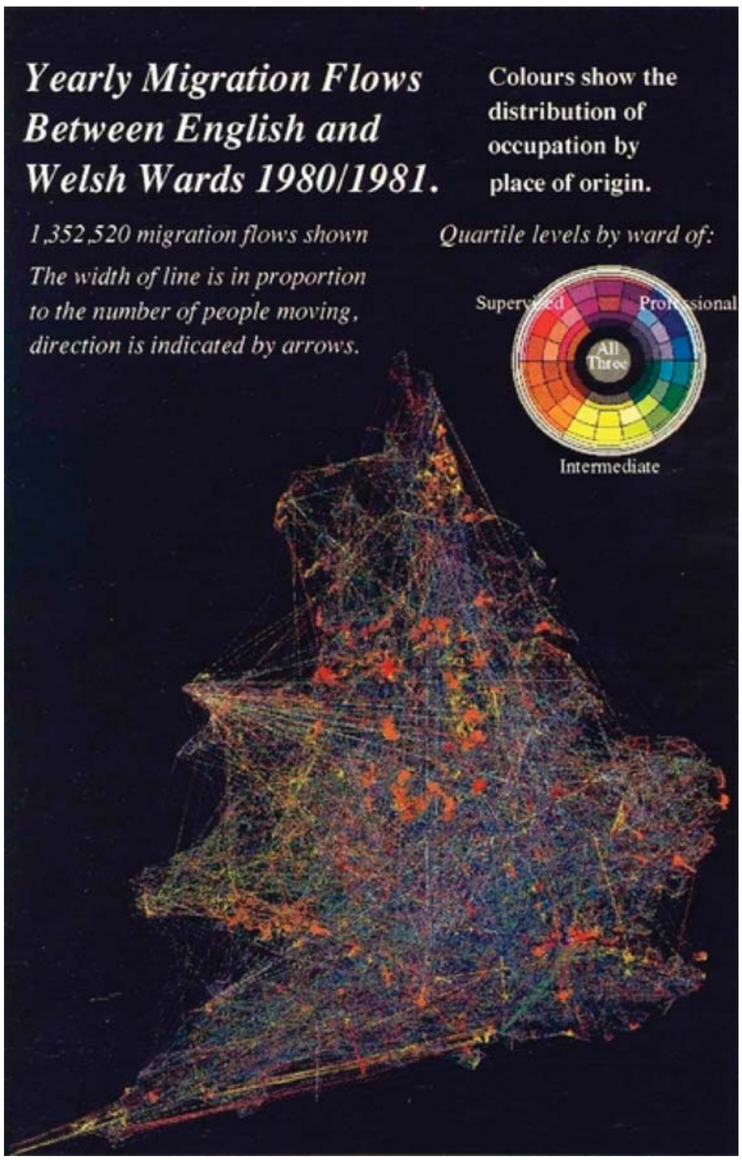

Figure 3 Yearly migration flows in population space by occupation, England and Wales 1981 (color, wax crayon original, cut by computer).

Conclusion

Health inequalities are the most important inequalities of all. Inequalities are best studied through numerical measurement. Discussion of inequalities in the abstract is an interesting area of philosophy to which geographers can contribute little. A few researchers with geographical backgrounds contribute to the debate on health in equalities. Despite its importance, the vast majority of well referenced work on geographical health inequalities is not produced under the auspices of the study of geography. Thus, if you want to learn more about the geography of inequalities in health, it is important not to constrain yourself to searching the geographical literature.

See also: Health and Development; Health Geography; Migration.

Further Reading

Curtis, S. (2004). *Health and Inequality: Geographical Perspectives*. London: Sage.

Davey, S. G., Dorling, D., Mitchell, R. and Shaw, M. (2002). Health inequalities in Britain: Continuing increases up to the end of the 20th century. *Journal of Epidemiology Community Health* 56, 434 435.

Davey, S. G., Dorling, D. and Shaw, M. (2001). *Poverty, Inequality and health in Britain: 1800 2000   A Reader*. Bristol: Policy Press.

Donkin, A., Goldblatt, P. and Lynch, K. (2002). Inequalities in life expectancy by social class, 1972 1999. *Health Statistics Quarterly* 15(autumn), 5 15.

Dorling, D. (ed.) (2007). Health, in Compass Closer to equality? Assessing New Labour's record on equality after 10 years in government. London: Compass. http://clients.squareeye.com/uploads/compass/closer%20to%20equality.pdf

Dorling, D., Shaw, M. and Davey, S. G. (2006). HIV and global health: Global inequality of life expectancy due to AIDS. *BMJ* 332, 662 664.

Low, A. and Low, A. (2004). Measuring the gap: Quantifying and comparing local inequalities. *Journal of Public Health* 26, 388 395.

Lynch, J. W., Smith, G. D., Kaplan, G. A. and House, J. S. (2000). Income inequality and mortality: Importance to health of individual income, psychosocial environment, or material conditions. *BMJ* 320, 1200 1204.

Mackenback, J. P. (2006) Health Inequalities: Europe in Profile, Independent report commissioned by the UK Presidency of the EU (http://www.fco.gov.uk/Files/kfile/HI EU Profile,0.pdf).

Moser, K., Shkolnikov, V. and Leon, D. (2005). World mortality 1950 2000: Divergence replaces convergence from the late 1980s. *Bulletin of the World Health Organisation* 83, 202 209.

Nakaya, T. and Dorling, D. (2005). Geographical inequalities of mortality by income in two developed island countries: A cross national comparison of Britain and Japan. *Social Science and Medicine* 60, 12, 2865 2875.

Pearce, J. and Dorling, D. (2006). Increasing geographical inequalities in health in New Zealand, 1980 2001. *International Journal of Epidemiology* 35, 597 603.

Shaw, M., Davey, S. G. and Dorling, D. (2005). Health inequalities and new labour: How the promises compare with real progress. *BMJ* 330, 1016 1021.

Shaw, M., Dorling, D., Gordon, D. and Davey, S. G. (1999). *The Widening Gap: Health Inequalities and Policy in Britain*. Bristol: Policy Press.

Shaw, M., Orford, S., Brimblecombe, N. *et al*. (2000). Widening inequality in mortality between 160 regions of 15 countries of the European Union. *Social Science and Medicine* 50, 1047 1058.

Singh, G. K. and Siahpush, M. (2006). Widening socioeconomic inequalities in US life expectancy, 1980 2000. *International Journal Epidemiology* 35, 969 979.

Tunstall, H. V. Z., Shaw, M. and Dorling, D. (2004). Places and health. *Journal of Epidemiology and Community Health* 58, 6 10.

Wagstaff, A., Paci, P. and Van Doorslaer, E. (1991). On the measurement of inequalities health. *Social Science and Medicine* 33, 545 557.

Wilkinson, R. (2005). *The Impact of Inequality: How to Make Sick Societies Healthier*. New York: The New Press.

Wilkinson, R. G. and Pickett, K. E. (2006). Income inequality and population health: A review and explanation of the evidence. *Social Science and Medicine* 62, 1768 1784.

World Health Organization (1985). *Targets for Health for All*. Copenhagen: World Health Organization.

World Health Organization (1998). Health in Europe 1997. Copenhagen: World Health Organization (www.euro.who.int/document/e60594.pdf).

World Health Organization (2006). The World Health Report 2006 Working Together for Health, Copenhagen: World Health Organization (http://www.who.int/whr/2006/whr06 en.pdf).

Relevant Websites

www.euro.who.int
The WHO; health in Europe (1997).

http://www.who.int
The WHO The World Health Report and its annex the statistics of inequality worldwide.

depts.washington.edu
UW Departments Web Server.

www.worldmapper.org
Worldmapper.

Health Services Restructuring

T. Brown, Queen Mary University of London, London, UK

Glossary

Consumerism It refers to the commodification of the healthcare landscape and to the recasting of patients as active consumers able to make informed choices.
Determinism The idea that health service restructuring was/is determined by economic necessity, advances in bureaucratic philosophy, and/or by technological developments.
Developmentalism The belief that health service restructuring is an inevitable process that occurs in evolutionary waves.
Keynesianism It dominated the post-war model of welfare, offering a basic level of service provision to those unable to purchase services privately.
Neoliberalism It refers to a political–economic philosophy that rejects government intervention and emphasizes the role of the individual.
Sense of Place It refers both to the intrinsic characteristics of a place and to the emotional attachment/detachment that people experience in relation to a specific place.
Voluntarism It refers to the role played by voluntary organizations in the provision of health and social care services.

Introduction

During the twentieth century, there were several phases of health system restructuring. The first saw the establishing of national healthcare and social insurance systems, throughout the 1940s and 1950s in the most advanced industrialized societies and somewhat later, if at all, in poorer countries. Following this was a phase of exceptional growth, both in terms of the range of healthcare services provided and the percentage of the total population covered by social insurance schemes. However, it is generally accepted that health services restructuring refers to the widespread reexamination of, and reform to, such healthcare provision following the global economic slowdown of the mid 1970s. Framed by a broader ideological debate about the relationship between the State, the market, and civil society, this reexamination raised questions about the sustainability of healthcare systems which were characterized by their ever increasing cost to the national economies.

The latter understanding is reflected in the World Health Organization's (WHO) definition of restructuring as a process that involves a "purposive effort to improve the performance of the health care system." According to this characterization, restructuring is presented as a technical response to system inefficiencies, the most notable of which was the inability to contain costs. While this may be the case, such a definition gives very little indication of the complexity that surrounds the question of health services restructuring. In particular, it does not account for the prevailing conditions that influence the organization, structure, and subsequent reform of health care systems across the globe. Given this, a much broader definition is preferred, one which recognizes restructuring as a range of economic, political, and social changes that vary across nations and have transformative effects on the people and places within which they occur.

Explanations

As indicated, health services restructuring can be viewed as a response to the increased financial pressure being placed upon national economies. A good indicator of this pressure is provided by the amount of a nation's gross domestic product (GDP) that is consumed by healthcare. In 1948, the average total expenditure on healthcare was 3% of GDP; by 1970 many of the Organization of Economic Cooperation and Development (OECD) countries were spending nearly 5% of GDP, and by 1980 this figure had risen to over 7% (see **Table 1**). When set against a backdrop of global financial slowdown in the 1970s and early 1980s, it is perhaps understandable that many nations began to question the affordability of their health care systems. Of particular concern was the recognition that this financial pressure was unlikely to relent: technological and pharmaceutical innovation, demographic changes, most notably an aging population, and the perceived inefficiency of the service sector economy all figured in the governmental calculations.

For those employing a deterministic logic, the model of healthcare restructuring that has followed is shaped by this apparent crisis in welfare provision. In this understanding, the introduction of market oriented reforms, about which more will be said later in this article, are presented as a technical response to the problems presented by the rising costs of healthcare systems. Other commentators employ a developmentalist approach, suggesting that the changes that can be witnessed, not just within OECD countries but in countries as diverse as Argentina, Brazil, Russia, Tanzania, and Zambia, are more akin to a convergence of ideas, attitudes, and values

Table 1 Total expenditure on healthcare in OECD countries as a percent of trend GDP, 1970 2000

OECD countries	*1970*	*1980*	*1990*	*1992*	*1997*	*2000*
Australia		7.5	7.7	7.9	8.7	9.2
Austria	4.1	7.7	7.2	7.6	7.5	7.9
Belgium	3.6	6.6	7.6	7.9	8.3	8.7
Canada	6.3		9.0	9.4	8.6	9.2
Czech Republic			5.0	5.4	7.1	7.1
Denmark	8.0	9.1	8.4	8.3	8.3	8.4
Finland		6.4	8.1	8.3	7.1	6.8
France			8.7	9.1	9.1	9.4
Germany	6.4	8.8	8.8	10.1	10.5	10.7
Greece		7.0	7.4	7.9	9.3	9.3
Hungary			7.1	7.7	7.0	6.7
Iceland	3.7	6.5	8.0	7.7	8.2	9.6
Ireland		8.7	6.4	6.9	6.5	6.8
Italy			8.1	8.3	7.7	8.2
Japan	4.6	6.5	6.1	6.3	7.0	7.7
Korea			4.8	4.7	5.0	5.9
Luxembourg	3.6	5.9	6.1	6.2	5.9	5.6
Mexico			4.5	5.4	5.4	5.6
Netherlands	6.9	7.5	8.2	8.5	8.2	8.9
New Zealand		6.1	6.7	7.2	7.5	8.0
Norway	4.7	8.7	8.7	9.3	9.7	10.3
Poland			5.3	6.6	6.1	6.0
Portugal	2.5	5.6	6.4	7.2	8.6	9.2
Slovak Republic					5.9	5.7
Spain		5.1	6.8	7.2	7.2	7.5
Sweden	7.2	8.9	8.5	8.0	8.0	8.6
Switzerland		7.8	8.9	9.5	10.2	10.7
Turkey	2.4	3.3	3.6	3.8	4.2	2.0
United Kingdom	4.5	5.5	6.2	6.6	6.8	7.3
United States	6.8	8.5	11.8	12.6	13.0	13.2
OECD point average	5.1	7.1	7.4	7.8	7.9	8.2

Source: OECD (2004). *Towards High Performing Health Systems: Policy Studies*. Paris: OECD.

related to the most efficient and effective way of delivering and paying for healthcare. While both approaches reveal general patterns within countries and demonstrate the associations that exist between them, they acknowledge neither the importance of broader ideological debates nor the specific health policy arena within which change occurs.

A useful illustration of the latter point can be found in the Clinton administration's attempts to reform the US healthcare system in the early 1990s. The need for reform is demonstrated both by the proportion of the population that was uninsured, which by 1990 had reached approximately 13% or 33.4 million people, and by the total expenditure on healthcare as a percentage of GDP, which was 12.1% or $666 billion only a year later. Clinton's Health care plan sought to remedy both issues simultaneously, by implementing a national health insurance scheme and by introducing government structured market competition. Yet, the plan for a national insurance scheme floundered in Congress and the ensuing introduction of managed care, which has been characterized as a form of 'unmanaged competition', was implemented by employers rather than the State. What this example highlights, is the need to recognize that restructuring is not simply a technical response but one that is influenced by an array of actors operating within a particular healthcare landscape.

The above illustration also reveals that such change is ideologically and politically mediated. For many commentators, contemporary restructuring debates have been influenced by the neoliberal ideology of the New Right; according to which the once dominant Keynesian model of welfare provision is regarded as overly bureaucratic, inflexible, and inefficient because of its reliance on the State. Fostered under the Reaganite and Thatcherite administrations of the late 1970s and 1980s, the anticollectivist principles of neoliberalism have resulted in quite a substantial change across the healthcare sector. For example, in countries such as the United Kingdom, New Zealand, and Russia, there has been a shift toward the principles of market competition as the most effective mechanism for controlling costs and delivering efficient and flexible healthcare services. This shift is not limited to advanced, industrial economies, for such is the dominance of neoliberalist ideology that international organizations (such as the WHO) measure the performance of national healthcare systems according to the extent to which they have adopted a market oriented approach.

Key Features

Generic explanations are limited with regard to their ability to account for change taking place within individual national contexts. The same argument could, of course, be made for explanations that focus entirely on ideological matters. Yet there are key features that are shared in some way by countries experiencing restructuring of their healthcare systems. This section highlights three of these: decentralization, neoliberal market reform, and consumerism.

Decentralization

Decentralization is defined by Rondinelli as the "transfer of authority and responsibility for public functions from the central government to subordinate or quasi independent government organizations or the private sector." As might be anticipated, the rationale for such a process varies both in time and space. For example, in many Latin American countries decentralization is regarded as an essential component of democratization. This might be contrasted to the situation in the countries of the former Soviet Union where decentralization has occurred on an *ad hoc* basis due to the rapid decline of the previous centralized regime or to some South East Asian countries where decentralization is to some extent motivated by the need to improve service delivery to large populations. Whatever the rationale, the concept of decentralization has become a key feature of healthcare restructuring.

Initially, decentralization was associated with the reorganization of the centralized administrative functions found in many healthcare systems. In opposition to the principal tenets of Weberian bureaucracy, this form of decentralization sought to replace the apparent rigidity of such top down systems with more flexible ones that foster a managerial or entrepreneurial ethos and allow for greater operational authority and responsibility at differing hierarchical levels. One framework for characterizing the various features of this process is Rodinelli's, which identifies four different types of decentralization: de concentration – the shift of administrative authority to regional or district offices; delegation – the transfer of responsibility to semiautonomous agencies; devolution – the creation or strengthening of subnational units of government; and, privatization – the transfer of responsibility to the voluntary or private sectors.

As noted above, such attempts to remove the administrative distance that exists between the various tiers of a healthcare system and the population that it serves are also closely allied to a broader concern with democratization. The argument here is that locally accountable administrative structures, especially those which encourage individual and community participation, provide more appropriate and better quality care. Although there is some international evidence that decentralization can be closely tied to democratization, particularly in Bossert's analysis of the 'developing' world, there is less evidence that it leads to better quality healthcare. Moreover, the commitment to decentralization as a means of countering the democratic deficit and bringing the decision making process 'closer to the people' has, in some instances, resulted in the loss of other forms of democratic accountability; most notably of locally elected individuals or bodies.

Neoliberal Market Reform

The question of neoliberal market reform emerged as an important feature of international healthcare restructuring debates in the 1990s. Ideologically driven, the suggestion was that the introduction of healthcare markets would result in greater effectiveness and efficiency because of the downward economic pressures generated through competition. Before exploring this in more detail, it is important to note that the market being referred to here is not the classic free market; in part, because it was widely acknowledged that the free market is not necessarily conducive to a strategy of cost containment. Rather, what emerged as an alternative to the free market was the quasi market, often used synonymously with the internal and planned markets. Such markets were created through the exercise of state power, were carefully managed, and were closely aligned with the process of decentralization.

There were antecedents to the quasi market. However, it was the ideological commitment to neoliberalism, alongside the reforms to the management and structure of healthcare systems, which provided a platform for the more extensive introduction of the internal market in the mid 1990s. These experiments in market competition, which drew heavily on Enthoven's notion of managed care being developed in the US, were characterized by the attempt to exert pressure on costs by introducing contracts between healthcare providers and purchasers. In New Zealand, as in the UK, this was achieved through the separation of the purchaser and provider functions, where newly established purchasing bodies contracted for care from public, private, and voluntary sector providers. The significance of this maneuver is highlighted in **Figures 1a** and **1b**, which illustrate the close association between decentralization and the implementation of quasi market reforms in this period.

Figure 1a reveals the organizational structure of the New Zealand healthcare system in 1989. This model was established in the mid 1980s and reflects the desire of the newly elected Labor government to decentralize health services through the introduction of area health boards (AHB), which were organized around at least one large district hospital and were responsible for secondary and

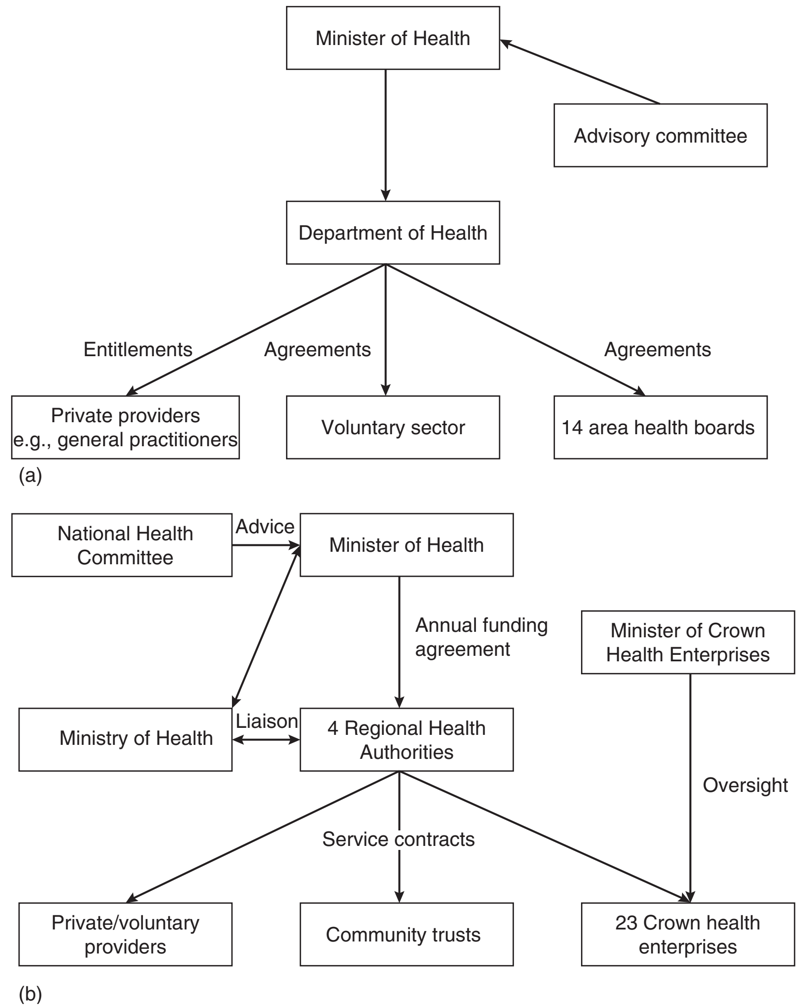

Figure 1 (a) Organizational chart of New Zealand healthcare system, 1989 and (b) organizational chart of New Zealand healthcare system, 1993.

tertiary healthcare. Though the achievements of the AHBs were acknowledged, particularly in terms of better management and tighter contracting, they were dis banded following the implementation of the Health and Disability Services Act 1993. It was at this juncture that decentralization and neoliberal market ideology came together. As **Figure 1b** reveals, state funds were chan neled through four newly established regional authorities who were responsible for purchasing services from pri vate and/or public providers on the basis of competitive contracting. In addition, and in contrast to the UK, existing public hospitals were renamed as Crown Health Enterprises (CHE) and were expected to function on a commercial basis.

Such attempts to introduce active market competition to the healthcare landscape were relatively short lived; lasting between 1992 and 1997 in the UK, between 1992 and 1996 in New Zealand and Sweden, and less than a year in the Czech Republic. The explanations for this vary. For example, in New Zealand the introduction of the quasi market was associated with considerable ad ministrative costs, most notably an operating deficit of NZ$788 million returned by the CHEs in 1996–97, and in the United Kingdom there was a perception that the market reinforced already existing inequalities in access to care. Whatever the explanation, this does not mean that the quasi market, or at least some of its principal features, has disappeared altogether: in many countries, the purchaser/provider separation remains and health care services are still arranged on a contractual basis. Moreover, the experience of the UK especially, despite being largely negative, constituted a reference point for market based reforms throughout Southern Europe and the countries of the former Soviet Union.

Consumerism

A key feature of both decentralization and neoliberal market reform is the belief that each enables a healthcare system to become more responsive to the needs and demands of its patients; whether through a more localized and inclusive decision making process or through the provision of services that are sensitive to patient or consumer demands. The importance of this responsiveness is reflected in the value that many governments attach to surveys that measure patient or consumer satisfaction with healthcare practitioners, with hospitals and other institutional providers and with the health system more generally. However, as a recent report by the OECD suggests, such surveys have revealed that, despite over two decades of healthcare restructuring, many countries have problems with patient or consumer satisfaction. For example, a 2001 survey of five countries – Australia, Canada, New Zealand, the United Kingdom, and the United States – found that the vast majority of those surveyed agreed that fundamental changes to their health system were necessary or that it needed rebuilding entirely.

Surveys such as the above provide some indication of how difficult it is for governments to respond to the ever increasing demands of consumer oriented societies. This does not stop countries from designing initiatives that are shaped by this consumerist ethos. For example, recent reforms to acute sector services in the UK are framed in terms of their ability to provide patients with greater choice when it comes to both the time and space within which healthcare is delivered. To this end, the traditional model of hospital care is being replaced by a more flexible system of delivery which includes online advice and diagnosis services (e.g., NHS Direct), the development of alternative spaces of care (e.g., treatment centers and walk in clinics), and the increased use of technologies that allow for care to be delivered at a distance from the patient (e.g., telemedicine). It is possible to argue that such changes allow patients to act as 'smart consumers', providing them with the information they need to exercise independent and informed choices regarding their healthcare options.

However, there are counter arguments to the development of such consumer oriented services. As has been noted elsewhere, this shift toward consumerism is closely linked with the commodification of healthcare and with the greater use of independent or private sector providers. While this should not be considered as necessarily problematic, there is some suggestion that it might lead to compromises in the clinical and ethical choices made by healthcare organizations operating within a commercial environment. For example, the sociologist George Ritzer has commented on the so called 'McDonaldization' of the healthcare sector, arguing that the emphasis on efficiency and speed can lead to decisions characterized by predictability and haste. Furthermore, questions have been raised about the tendency to locate fast food restaurants within or near to hospitals. Such issues sit alongside broader concerns regarding the ability of a consumer oriented model to deliver health care equitably and in a way that reflects the needs of patients rather than the demands of healthcare organizations.

Geographical Engagement

While geographers have made a substantial contribution to the general debates surrounding health services restructuring, they have also sought to develop quite specific areas of knowledge and understanding. This section draws on examples to explore some of these areas in a little more detail.

Symbolism and Sense of Place

Health geographers have utilized the interrelated notions of symbolism and sense of place to develop an understanding of the ways in which healthcare restructuring affects the everyday attachments that people form with place. Drawing on theoretical insights from humanistic geography, this research recognizes that individuals and communities often form strong emotional attachments with commonplace elements within their local built environment, such as hospitals and schools. Put differently, these elements come to symbolize place and have a positive impact upon the sense of belonging and well being that individuals feel. However, this research does not rely on the promotion of a naïve reading of humanism; rather it positions these emotional attachments in relation to broader social and cultural processes. For example, in their study of restructuring in Tokanui and Hokianga, New Zealand, Kearns and Joseph locate their analysis within the context of rurality and biculturalism.

The importance of this attention to context is that it provides a further explanation for the emergence of local resistance to restructuring. Returning to the example of Tokanui and Hokianga, Kearns and Joseph identify two very different responses to restructuring, with the closure of a psychiatric unit in Tokanui facing little opposition and the threat to a network of outpatient clinics in Hokianga meeting prolonged and bicultural resistance. In order to explain this variance, Kearns and Joseph point to differences in the type of institution and to the impact that restructuring would have on the locality as a whole. In particular, they highlight the fact that the outpatient clinics in Hokianga were closely interlinked with local tradition and identity and were regarded as spaces of social interaction. Therefore, by combining humanistic notions with broader social and cultural forms of explanation, this body of research promotes a complex understanding of the everyday responses to healthcare restructuring.

Inequality and Rural Communities

Geographers have displayed a longstanding interest in the dual questions of healthcare provision and access. While a broad area of activity, research on these topics has produced critical insight into inequalities within and between rural and urban areas and, perhaps most importantly, it has pointed to the negative impact that this has on the healthcare experiences of vulnerable groups, including the elderly, minority ethnic communities, and those with limited resources (i.e., economic, transport, etc.). Of particular relevance here is research which identified the considerable effect that the market oriented reforms of the 1990s had on these patterns of inequality. For example, the introduction of general practitioner fundholding in England, following the 1991 reforms to the National Health Service, was associated with the development of a 'two tier' health system. As Asthana and colleagues demonstrated in their study of fundholding in southwest England, eligibility criteria discriminated against patients living in highly deprived urban areas and communities served by small rural practices.

Such studies are matched by research that explores the impact of restructuring on acute sector services. Much of this research has examined the interrelated themes of rationalization (i.e., the closure of capacity), spatial relocation (i.e., shifts in the location of acute sector provision), and centralization (i.e., the concentration provision within large regional centers). In terms of geographical focus, some of the work in this area has a distinctly urban focus, for example Sara McLafferty's use of spatial interaction models to examine the effects of hospital closure in New York city in the 1980s. However, given the spatial characteristics of these various processes, there is also considerable interest in their impact on isolated and vulnerable communities living within rural areas. In addition to focusing on issues of access, provision, and utilization, research conducted by health geographers such as Robin Kearns and Ross Barnett has provided an important insight into the emergence of health based activism through an analysis of the capacity and efficacy of community resistance to acute sector reforms in rural areas.

New Spaces of Care

As noted earlier, gains in life expectancy and the associated trend toward an aging population are key factors behind the so called crisis in welfare. While there is some uncertainty about the effect that this particular demographic trend will have – although the OECD predicts health costs will increase by two percentage points between 2000 and 2050 as a result of it – many governments have responded to the increased demands on their health and social care systems by implementing reforms to them. It is the resultant changes to the pattern of formal and informal care that has been of interest to an increasing number of scholars. With regards the former, health geographers have contributed to this understanding in a number of ways, including analysis of the expansion of health and social care provision, of the spatial variation in the privatization of the sector, and of the role governments have played in shifting the financial burden of care from the State to the individual.

Perhaps a much more distinctive disciplinary contribution has been made by health geographers interested in the informal and voluntary sectors. Interest in this area builds upon research by Jennifer Wolch which identified the changing relationship between the voluntary sector and the State as a form of 'shadow state'. More specifically, scholars such as Christine Milligan in the UK and Allison Williams in Canada have explored how neoliberal market reforms have changed the pattern of informal and voluntary care provision. For example, Milligan points to the geographically uneven distribution of voluntary services, to a reduction in the services that they provide, and to the greater burden being placed upon informal carers in the space of the home. These findings are borne out by Williams, who suggests that the majority of western countries have turned to this home based model of care as a response to financial pressure and the ideological commitment to neoliberalism.

Summary

As this chapter demonstrates, the processes associated with health services restructuring have had a considerable impact upon the organization of healthcare systems and upon the services that they provide. This impact is not uniform and is, to some extent at least, reliant upon spatial and temporal context. Clearly such differences make it difficult, although not impossible, to make comparisons either within or between nations. However, while this may be the case, it is also important to recognize that contemporary health service restructuring has been framed by the hegemonic position achieved by neoliberalism in the late 1980s and early 1990s. Many of the principles upon which restructuring is based, within western and nonwestern countries, are linked to neoliberal philosophy; whether, consumerism, decentralization, market competition, or privatization. The research that health geographers conduct has contributed to interdisciplinary debates through an analysis of the spatial impact of restructuring and through an awareness of its impact on people's everyday experiences of their healthcare systems.

See also: Healthcare Accessibility; Health Geography; Sense of Place; Voluntary Sector; Welfare Reform.

Further Reading

Atkinson, S. (1995). Restructuring health care: Tracking the decentralization debate. *Progress in Human Geography* 19, 486 503.
Barnett, J. R. (1999). 'Hollowing out' the state? Some observations on the restructuring of hospital services in New Zealand. *Area* 31, 259 270.
Barnett, J. R. and Barnett, P. (2003). Back to the future? Reflections on past reforms and future prospects for health services in New Zealand. *Geojournal* 59, 137 147.
Hanlon, N. T. and Rosenberg, M. W. (1998). Not so new public management and the denial of geography: Ontario health care reform in the 1990s. *Environment and Planning C: Government and Policy* 16, 559 572.
Joseph, A. E. and Kearns, R. A. (1996). Deinstitutionalization meets restructuring: The closure of a psychiatric hospital in New Zealand. *Health & Place* 2, 179 189.
Joseph, A. E. and Knight, D. B. (1999). Social sciences and public policy in restructuring societies. In Knight, D. B. & Joseph, A. E. (eds.) *Restructuring Societies: Insights from the Social Sciences*, pp 1 24. Ottawa: Carleton University Press.
Kearns, R. A. and Barnett, J. R. (2000). Happy meals in the Starship Enterprise: Interpreting a moral geography of health care consumption. *Health and Place* 6, 81 94.
Light, D. (2001). Cost containment and the backdraft of competition policies. *Politics of Health Policy* 31, 681 708.
Litvack, J. and Seddon, S. (eds.) (1999). *Decentralization Briefing Notes*. Washington, DC: The International Bank for Reconstruction and Development/The World Bank.
Milligan, C. (1998). Pathways of dependence: The impact of health and social care restructuring the voluntary experience. *Social Science and Medicine* 46, 743 753.
Mohan, J. (1995). *A National Health Service? The Restructuring of Health Care in Britain Since 1979*. New York: St. Martin's Press.
Mohan, J. (1998). Explaining geographies of health care: A critique. *Health and Place* 4, 113 124.
OECD (2004). *Towards High Performing Health Systems: Policy Studies*. Paris: OECD.
Panelli, R., Gallagher, L. and Kearns, R. (2006). Access to rural health services: Research as community action and policy critique. *Social Science and Medicine* 62, 1103 1114.
Peckham, S., Exworthy, M., Greener, I. and Powell, M. (2005). Decentralizing health services: More local accountability or just more central control? *Public Money & Management* 25, 221 228.

Health Systems and Health Services

R. Barnett and P. Barnett, University of Canterbury, Christchurch, New Zealand, and University of Otago, Christchurch, New Zealand

Glossary

Accessibility A condition which describes the extent of barriers, locational, financial, or psychological to a service.

Corporatization The introduction of for-profit goals into public sector organizations or the presence of for-profit investor-owned private institutions.

Devolution The transferral of government responsibilities from a higher to a lower level of government.

Health System An interrelated system of public, private, and voluntary sector providers concerned with the delivery of health services.

Health Services Primary, secondary, and tertiary providers who provide medical care to the population.

Iatrogenesis The harmful effects of medical treatment.

Internal Market Publicly established quasi-markets characterized by a funder–provider split and involving contracting relationships between funders and providers.

Inverse Care Law Describes a situation where populations in greatest need for health care have the poorest access to health services.

Managed Care Integrated health care organizations which provide both primary and secondary care and which contract with government or private sector funders.

Neoliberalism A political philosophy which emphasises the role of individualism and market forces in preference to collective state intervention in the economy and society.

Privatization The withdrawal of state involvement in the provision, funding, and regulation of services.

Public Health The focus of health interventions upon populations in contrast to the individual focus of the biomedical model.

Regulation Theory A metatheory which stresses the development of new organizational relationships in the economy, polity, and society as a way of responding to capitalist crises.

Restructuring A process of change whereby public sector organizations alter their goals and priorities and the ways they deliver services.

Roemer's Law A situation which describes the positive relationship between the availability of hospital beds and rates of hospitalization.

Social Exclusion A situation where certain groups in the population are discriminated against in terms of the provision of services.

Supplier-Induced Demand The primary care equivalent of Roemer's Law.

Therapeutic Landscapes Particular contexts in the natural or social environment which de-medicalize space and aid the process of healing.

Third Sector A term which describes voluntary sector organizations involved in the provision of health care.

Third Way A new political philosophy which incorporates elements of both neoliberalism and socialism.

Utilization The extent to which health services are used by different groups.

Introduction

Geography as a discipline is well suited to understanding health systems and services. Geography has a 'whole system' way of thinking, concerned with the interactions of populations with their environment and how this varies from place to place, a perspective well suited to the analysis of health systems which tend to be hierarchical entities with diverse sets of providers, and with functions and patterns of provision that vary across space.

According to the World Health Organization a health system comprises "all of the activities whose primary purpose is to promote, restore and maintain health" (World Health Organization, 2000). This definition has limitations since it fails to deal with the integration of the different components (e.g., public and private or primary and secondary) of the system and gives no indication of whether there is oversight of the system and the degree of public regulation. In most countries the state has a role in regulating access to, financing, and organizing the delivery of health care to the population. In fact, in most developed countries, at least since World War II, national governments have assumed the primary responsibility for ensuring the funding, provision, and regulation of health services, although in recent years this pattern has been subject to change.

Modern health systems developed from the nine teenth century when pressures of urbanization and in dustrialization resulted in the first systematic responses by the state and voluntary agencies to emerging health issues. These responses were pragmatic, arising from concerns over the economic impact of infectious diseases, and it was not until the 1930s that social justice concerns came to the fore, with the rise of organized labor. Only after World War II did genuinely 'national' health sys tems begin to emerge as a result of the evolution of the Keynesian welfare state. In developed countries early examples were the provision of free hospital care in New Zealand (1938), the formation of the UK National Health Service (NHS) (1948), or the development of Medicare and Medicaid government programs in the USA (1965). More recent examples include the development of the Canada Health Act (1974) and Medicare in Australia (1975), both of which extended universal care to their populations.

In recent years, health systems have come under in creasing pressure due to growth in demand arising from changes in the physical and social environment and evolving patterns of disease. This has resulted in a con tinuing search for new organizational structures to deal more effectively with the health needs of the community. An important part of health services geography is to understand not only the spatial organization of health services, but also how, where, and why such systems have changed. In the light of this, the focus of this article is twofold. First, we provide an overview of health systems and health services by focusing on the goals of modern health systems, the problems they face, their key com ponents, and some of the ways of evaluating their per formance. In the second part we focus on four issues of the geography of health systems: (1) geographical dif ferences in types of health care system, (2) geographical differences in health system restructuring, (3) how geography is important in assessing the performance of health systems, and finally (4) whether health systems matter in terms of health outcomes.

Overview of Health Systems

Health System Definition and Purpose

What is a health system?

In the introduction we argued that the World Health Organization's view of a health system is too narrow because it fails to acknowledge the important role of governments in modern health systems. Also, the concept of health systems must recognize both the contribution made to health and health services by private and vol untary organizations and that health is influenced by factors beyond the boundaries of formal health organ izations or governments.

Usually the concept of a 'system' includes a sense of purpose, relationships between the component parts, exchange of information, and an integrated approach. One way in which a definition of a health system can be expanded to take these factors into account is to adopt a population or public health approach. This places the health of the population and the influences on it at the center of a health system, so that the contributions of particular policies, services, or other activities can be coordinated in the interests of improving population health status. Drawing on a classic definition of public health, therefore, we can suggest that a health system comprises all the 'organized efforts of society' that are designed to promote, protect, maintain, and improve health. This reference to 'organized efforts' suggests that health arrangements require a certain sense of purpose and a level of coordination to achieve designation of a 'system', implying a significant role for the state.

Goals of health systems

Health systems have two broad sets of goals:

1. The first goal is to improve the health status of the community. This can be done by treating illness in individuals or helping prevent disease. It can also be achieved by working with other sectors that influence determinants of health, such as education, welfare, and employment. Often this goal specifically targets in equalities in health status.
2. The second goal of health systems is to ensure the provision of appropriate, accessible, and high quality services to the community. These services may be provided by public, private, or nonprofit organizations and be funded through a variety of mechanisms. If they are to be directed toward an overall goal of im proving health status and reducing inequalities, they must be accessible on the basis of cost and distance and be of high quality for all.

Problems Facing Health Systems

Patterns of need and demand

The patterns of need and demand for health services change according to the economic development and political and social values of a country. In rich countries many of the so called diseases of poverty have been minimized through well developed sanitation systems, environmental legislation, good nutrition, universal education, immunization, and access to medicines. More affluent countries have, however, experienced an increase in other sorts of diseases, the diseases of aging and of an affluent society. Diseases of aging include many cancers and degenerative diseases such as arthritis and dementia, some leading to long term disability. Diseases of afflu ence are those conditions caused by excessive con sumption and a Western lifestyle, including smoking and obesity related disorders, heart disease, and diabetes.

Poorer countries generally experience poorer health status and different patterns of disease from those found in richer countries. They are more likely to experience high rates of child and maternal mortality, infectious diseases, and injury. However, these countries are also afflicted by increasingly high rates of so called 'western diseases', including those related to smoking and poor nutrition. In addition, a growing rate of HIV/AIDS places excessive demands for treatment and care on struggling health and social services, particularly in Africa. Injury rates and the lesser access to treatment also mean that long term disability rates are also high. These trends suggest that the poorest countries experience a burden of both communicable and noncommunicable disease that is well beyond their capacity to address. Within both rich and poor countries there are inequalities in income, education, and other determinants of health that are reflected in inequalities in health status. Where governments intervene to moderate these inequalities, the narrower they tend to be.

Increasing costs

Health expenditure in any country depends on a mix of social, economic, and political factors and there is no 'right' amount to spend on health. Among western countries, the United States spends approximately 15% of GDP (from all sources) on health, compared with 9–10% in Western European countries and somewhat less (7–8%) in poorer OECD states such as the former Soviet states and satellites. These amounts have grown significantly over the last two decades and a number of commentators question the sustainability of these trends.

Growth in health expenditure is attributed to two main factors. First, there is the development of increasingly complex technology and the ability of health services to intervene more effectively (and expensively) for conditions that in the past may have remained untreated. Aging populations provide greater opportunities for such interventions. Certainly, the investment potential of health technology, particularly pharmaceuticals which is one of the most profitable industries worldwide, encourages the marketing and promotion of new interventions. A second factor is the increasing expectations of the community for more and better services, the product of a more informed population used to searching for the best product and willing to pay for it, either individually or collectively.

Issues of distribution/rationing

With both need and demand increasing and the costs of health interventions rising, many countries with publicly funded systems have been forced to address the question of rationing more explicitly than in the past. Rationing experiments have occurred in a number of countries, most notably in the US state of Oregon where both cost effectiveness analysis and public consultation were used to identify those health interventions that would be made available to publicly insured patients. Other countries (e.g., New Zealand or the Netherlands) attempted to develop comprehensive rationing approaches but more recently health technology assessment (HTA) has been used to assess the performance of individual interventions so that health authorities have better information for decision making. This has been matched by better efforts to assess clinical need as a basis for rationing individual access (e.g., via waiting times for surgery, or establishing criteria for receipt of high cost pharmaceuticals) or community need as a basis for prioritizing funding decisions.

Components of a Health System

A health system comprises three broad components: the inputs to the system, the services and activities of the system, and the outputs, or results.

Inputs and structure

The 'inputs' or 'structure' are the features that determine the overall design and philosophy of the system and the way in which resources are distributed. The first of these is 'governance' which determines the extent to which governments are involved in setting direction, funding, or providing services, or whether this is left to the private sector or the medical profession. These arrangements are closely related to those for 'financing and resource management.' There are four broad models of governance and resource management that reflect both historical trends and modern experience.

The first broad model is the centrally planned and controlled health system, characteristic of socialist states such as Cuba and North Korea. The former Soviet Union and its East European satellites were also examples of this, but since 1990 many such states have moved to more mixed, particularly social insurance, systems.

The second approach is represented by predominantly tax funded systems. These are characteristic of some western countries with a strong welfare state tradition and a commitment to equity. These countries have moved to more decentralized systems, while still maintaining a strong government policy framework, as in the UK, New Zealand, and Norway, but with more decision making devolved to lower levels of government (local or health authorities). The funding and provision of services tends to remain predominantly through public agencies but with increasing involvement of private providers.

A third model is the social insurance system characteristic of Germany, France, and other parts of continental Europe, reflecting the strong welfare tradition of those countries. Funded through social insurance, often employer contributions, rather than taxation, resources may be managed through nongovernment agencies or funds, with the government acting as regulator. Traditionally, there is much greater involvement of private facilities and providers, with payments on a 'reimbursement' or insurance model of funding. In these countries there tends to be less direct involvement by government, with the medical profession and provider organizations exerting significant influence.

The fourth broad model can be termed the entrepreneurial or market model, which is characterized by the United States. There is a strong component of private funding, largely through private insurance programs with the government acting as residual or 'safety net' insurer for high need members of the community and as regulator of the system. Other countries such as Thailand and the Philippines that do not have a strong welfare tradition have also adopted a similar model.

While these models represent certain ideas about how health systems work, in practice, few systems fit the models exactly. For example, the move to social insurance arrangements in some former communist countries was also accompanied by efforts to strengthen policy leadership by governments. Similarly, the introduction of competitive contracting has been a feature of some tax funded systems in attempts to increase efficiency. All models rely on adequate resourcing and some government regulation for the development of a coherent system. However, there are many poorer countries that do not have the resources, governance arrangements, or stability to provide the frameworks necessary for effective system development. These emergent systems may be dependent on international aid, but also forced to address priorities such as the consequences of conflict and HIV/AIDS without an established infrastructure.

Activities and services

Provider organizations may be either publicly or privately owned, and if privately owned then either profit or nonprofit. Despite differing ownership models, contractual relationships with funders often permit full participation of organizations in an integrated system. The overall framework of a system and the degree of government regulation are likely to determine the extent of integration and thus the overall effectiveness and responsiveness of services. For a health system to be considered comprehensive it will include a number of core activities and services: public health, primary health care, specialist (secondary and tertiary) services, and community support services.

Public health. Public health is concerned with the overall health status of the population rather than the health care of individuals, and has two important roles. The first is a strategic role which requires understanding disparities in health status, contributing to policy, and monitoring the impact of interventions. The second involves regulation for and provision of services to protect the health of the community. This may include placing requirements on local government for sanitation and environmental management, border control, health and safety legislation, and the funding or management of preventive, screening, or health promotion programs.

Primary care. Primary care has been defined as the point of first contract with health services and a source of family centered, comprehensive care. In western countries this has traditionally involved access to a primary care doctor although in recent decades the concept of the primary care team, including doctors, nurses, and other health professionals has taken hold. Internationally, thinking about primary care has evolved significantly since the World Health Organization's Declaration of Alma Ata in 1977. This outlined an expanded scope for primary care and its role in health systems, and identified its significance for health improvement and development, particularly in poorer countries. The Declaration emphasizes the importance of training a local workforce, access to clean water, a healthy environment, and preventive services. It also promotes the involvement of communities and the central role that primary care should play in a health system. More affluent countries initially paid little attention to the Alma Ata Declaration, but in recent years the importance of primary care has been more widely acknowledged, not least because it offers a less expensive approach than hospital services.

Specialist secondary and tertiary services. Problems that cannot be addressed by primary care may need referral to more specialist services, normally provided through hospitals or specialist centers. Secondary care provides diagnostic and treatment services for common conditions, whereas tertiary services represent the highly specialized care for more rare or complex conditions that requires a large population base to support it and is often located in larger town or cities. The boundary between secondary and tertiary services is contested, and depends on the expectations of the community, the wealth of a country, and the influence of interest groups. What in one context is seen as a tertiary level service in a more affluent or less regulated environment may be accessible as a secondary service.

Some countries have a network of publicly owned and funded hospitals, others, notably the countries of continental Europe, have a strong tradition of religious and charitable hospitals, now funded largely via social insurance funds. The United States is unusual, given the high proportion of hospitals owned by private corporations. There are challenges to the role of hospitals. The costs of hospital care and the possibility of alternative, cheaper settings for both acute and long term care have led to efforts to relocate some services in the community.

Community support services. Providing services in the community for rehabilitation and enabling people to live outside hospital has been a major trend of the late twentieth century. The provision of home care and support, medical and home nursing has meant that people can live at home rather than be in hospital or a nursing home. This has been of great significance for older people, including those requiring specialist pallia tive or hospice care, and for people with a long term physical or mental disability. This trend, often called de institutionalization, has not been without its critics as often people were assigned to community care without due attention being given to the level of need, provision of resources, or the issues that might arise for families.

Outputs: how do we evaluate the performance of health systems?

There is a well established tradition of service evaluation where health services research contributes to an assess ment of whether the resources (inputs) and service processes make a difference to performance and out comes of an intervention. This can be applied to indi vidual surgical procedures, immunization programs, education initiatives, for example, and provides decision makers with information on how to use resources or improve performance. At a higher level of organization, such as a hospital, primary care center, or laboratory, it is possible to develop performance indicators that are ac cepted as a way of measuring and comparing perform ance, such as re admission rates, cervical smears read correctly, or waiting times for consultations.

Evaluations at a higher level, that of the system es pecially if comparing the performance of different countries, are infinitely more challenging. A method ology for this developed by the World Health Organ ization in 2003 was subject to criticism both because of choice of criteria and lack of comparable data. More restricted evaluations, such as those of the Common wealth Fund in the US, that compare similar countries on a smaller number of dimensions have been more widely accepted, although there is continuing debate about the precedence given to 'expert' or 'consumer' views of health systems. While most of the international com parisons have focused on questions of affordability and access, at the national level a number of explicitly geo graphical criteria, relating to key characteristics associ ated with space and place, are also important (see section titled 'Geographical factors in assessing health system performance').

Geography and Health Systems

Geographical Differences in Types of Health Care Systems

Variations in health systems reflect differences in levels of economic development between places and different political cultures across countries at similar stages of development. Among western countries three broad political traditions have influenced the development of health policy. The first, the social democratic tradition is characteristic of the Scandinavian countries and Austria where there has been extensive welfare involvement and a commitment to universal social policies. These coun tries generally have strong labor movements, high female labor force participation, and a commitment to high taxation and income redistribution. On the other hand, the health systems in many continental European coun tries reflect a Christian Democratic tradition. These countries, although having a strong welfare tradition tend to be more conservative with less government involve ment and a stronger reliance on the family for the pro vision of social services particularly to the elderly, disabled people, and children. These countries tend to have weaker labor movements, lower rates of female workforce participation, more fragmented political coa litions, and lower levels of health expenditure than social democratic countries. Third, there is a group of liberal Anglo Saxon countries such as the UK, Ireland, USA, Canada, Australia, and New Zealand which have been characterized by more residual welfare states and a greater reliance on market mechanisms in the delivery of health services. It has been in this group of countries where recent health sector restructuring has been greatest.

These international differences are also reflected within countries. In the United States, for example, there are substantial differences in health policy between states, with southern and western states having a much greater focus on privatization and less state involvement in health service funding and provision than have the northeastern seaboard states. Consequently, it is in these states where rates of the uninsured are highest. Similarly in other federal systems, such as Australia and Canada, different political cultures are evident in the contrasting health policy outcomes of states and provinces.

Geographical Differences in Health Sector Restructuring

Searching for the right solution: 'revolving doors'

The above typology assumes that health care systems are constant but in reality they are in a state of constant change. Many systems, particularly in Liberal countries, have undergone repeated restructurings ('revolving doors') since the 1970s. These shifts in health policy can be related to broader changes in the welfare state, with shifts to new modes of regulation as described by regu lation theory. Health sector restructuring can be de scribed as a new mode of regulation as key relationships change between the state, key providers, and patients.

Main goals and forms of restructuring

Generally, there have been three main goals of health service restructuring: equity of outputs, efficiency, and effectiveness. The first has focused upon allocating resources more equitably between regions, the second on reducing costs via service integration and improved provider accountability, and the third by improving the impact of expenditure on health. Attempts to achieve these goals have occurred through three main forms of restructuring which may loosely be described as involving redistributive, managerial, and market mechanisms. Redistributive forms of restructuring derive their origins from the ideas of Marx and Titmus and focus on targeting resources more effectively in response to population need. Managerial forms of restructuring involve the transformation of public sector institutions and are heavily influenced by the new public management, focusing upon accountability and performance contracts to improve overall sector efficiency. Spatial components have been evident in particular managerial policies, such as devolution (the establishment of independent service providers with their own budgets), rationalization (closure/centralization of services), intensification (increasing labor productivity in certain sectors/regions), flexibilization (including more flexible work practices and service delivery regimes), technical change (substituting capital for labor inputs particularly in certain less skilled sectors), and proletarianization (reducing the autonomy of more skilled providers and making them more subservient to state or private funders). Health service restructuring has also increasingly used market mechanisms both by increasing the corporatization of the public sector and through increased privatization by greater use of contracting services out to both the private and voluntary sectors. In many western countries this form of restructuring, relying on competition and consumerism, has been heavily influenced by public choice theory and the rise of neoliberal political philosophies.

Different times, different forms of restructuring

The goals of restructuring have varied in importance depending upon the type of government in power and the extent of its commitment to universal or more residual/market based policies. In the early years of reform many governments sought to provide equal access to care, as in the UK following the introduction of the NHS in 1948 and in New Zealand after the enactment of the Social Security legislation in 1938. However, such reforms did not necessarily improve the distribution of medical resources, with governments, by the 1970s and 1980s, becoming increasingly concerned with resource equity issues. This concern saw the introduction of population based funding for hospital care in many countries. Given that these changes often involved increased government expenditure, they can be seen as part of a Keynesian/Fordist regime. By contrast, in the 1980s and 1990s, reforms reflected a 'post Fordist' regime with a more flexible approach to economic management. In the public sector this meant managerialist approaches to contracting, with welfare subordinated to the needs of the economy and the devolution of government powers and responsibilities to lower levels of government or private agencies. Public sector reforms, however, were deemed insufficient in themselves to produce adequate efficiency gains and performance improvements and the focus of reform soon shifted to the use of more market mechanisms.

The adoption of neoliberal health policy in many countries and the focus on market solutions undermined the role of place and community input. This led to public disenchantment, a subsequent retreat from such policies, and the development of alternative approaches, most notably in the UK and New Zealand. Sometimes called 'The Third Way', this approach permits governments to steer an intermediate course between neoliberal or market policies and those dominated by centralization, public provision, and intervention. As a consequence, cooperation has increasingly replaced competition and more balanced health care goals involving efficiency, equity, and an increased focus on effectiveness of health expenditure have predominated.

Different places, different forms of restructuring – three case studies

Although health service restructuring has been influenced by globalization and the diffusion of reform ideologies, it nevertheless has taken on different forms in different places. The explanations of these differences are complex, and relate not only to differences in political ideology, but also to the influence of other factors such as structural economic change or the power of the medical profession in influencing the process and outcomes of restructuring. The following section outlines three different types of restructuring which have occurred to varying degrees in western countries: public hospital restructuring, hospital privatization, and the emergence of new primary care led health care systems.

Public hospital restructuring. In the public hospital sector perhaps the most profound policy change in the 1990s in some countries, particularly in Britain and New Zealand, was the introduction of the 'internal market.' Sometimes referred to as quasi markets, because they attempted to mimic private sector processes, their goal was to increase public sector performance by reducing provider capture and providing incentives for efficiency through competition between services. They had four fundamental characteristics: (1) a funder–provider split whereby regional funders or purchasers contracted with public, private, and voluntary sector providers; (2) integration of funding – one budget now covered all aspects of care;

(3) corporatization or for profit goals; (4) new forms of governance involving little or no citizen input and increased central control.

Internal markets had a number of potential advantages: first, integrated budgets could reduce cost shifting between different health sectors; second, there was a greater ability to emphasize health service goals and consequently a considerable degree of local innovation in contracting with different community organizations; third, ideally more coherent area wide planning was possible under larger regional funding organizations; and fourth, there was the promise of greater integration between primary and secondary services as a result of new contracting arrangements.

However, internal markets suffered from a number of disadvantages: first, true competition was not possible in small hospital 'markets'; second, competition tended to increase rather than reduce health care costs, with efficiency gains less than expected; third, competition created a climate of distrust between different providers and a 'clash of cultures' (commercial vs. professional) resulted in workforce unrest; fourth, despite central government control, the result of devolution increased regional inequalities in the performance of the health system; fifth, devolution resulted in a 'hollowing out' of the state whereby more powerful 'agents' (both funders and providers) devolved responsibility for both the funding and provision of hospital care. Devolution inevitably resulted in service rationalization as providers sought to cope with fiscal deficits. It was also a means by which central governments could avoid blame for local difficulties.

Faced with financial deficits and possible closure, public institutions sought ways to sustain themselves. Some marketed themselves to attract paying patients as a way of generating additional income. Others, such as the St Boniface Hospital in Winnipeg, Canada and the Starship Children's Hospital in Auckland, New Zealand both permitted in house fast food franchises. Such 'place marketing' reflected the new commercial ethos of these public institutions but was at odds with their health goals. Other responses to threats of closure can be characterized as 'geographies of resistance', whereby communities and other affected groups resisted such changes. Some strategies involved the creative use of discourse as was the case in St Bartholomews in London, where the hospital was alternatively portrayed as a community resource, as a site of expertise, as a heritage symbol, and as a site pertinent to the identities of Londoners. The potency of this health care facility as a symbol both of London and of medical tradition was the crucial factor in its reprieve. Similarly, in New Zealand many small rural hospitals averted closure both through community inventiveness in developing local funding bases and through the critical role of community agents in producing legislative change which increased labor flexibility in these institutions.

Hospital privatization. In addition to public sector restructuring, the privatization of hospital services has been another state response in reforming health care in developed countries. Privatization has proceeded furthest in countries such as the United States, Greece, Switzerland, and Australia but has taken on a variety of different forms. Privatization involves changes in state involvement in the provision, funding, or regulation of services and, like public sector restructuring, is designed to improve efficiency by encouraging greater competition among providers and greater consumerism among patients. Given increased hybridity in health care provision where public–private–voluntary sector partnerships are increasingly becoming the norm in the contract state, the terms public and private have less meaning than before.

Two contrasting examples of hospital privatization are corporatization and co location. The former is predominant in the United States, whereas both forms are increasingly common in Australia. In both countries the for profit sector has been the fastest growing source of care. The growth of for profit hospitals can be linked to changing sources of capital. Unlike public or voluntary hospitals which relied upon philanthropy or government grants, corporate institutions rely upon debt (loan) or equity financing (sale of stocks). Economic downturns in many western countries during the 1970s deprived governments of the necessary financing to fund public institutions but provided for profit hospitals with ready sources of capital and an incentive to acquire or build their own hospital chains. These shifts of capital into the health sector have provided the opportunity for profits far exceeding returns from traditional forms of investment in manufacturing and have resulted in a significant transformation in the structure of the private hospital sector from a local industry to a significant global enterprise.

Capital focused explanations, however, ignore the fact that the growth of corporate involvement has been influenced by two other factors: (1) the desire to constrain clinical freedom as a way of containing health care costs and (2) variations in state regulatory activity which may have encouraged or impeded corporate involvement. Variations in local state regulatory mechanisms have been important in affecting the regional development of the for profit sector in both the United States and Australia. In Australia, for example, these have included the relaxation of licensing controls for new private hospitals, the encouragement of public–private partnerships through the co location of public and private hospitals, and, as in the case of the Private Finance Initiative in Britain, the contracting of the private sector to build, own, and operate (BOO) public hospitals. These developments have been particularly evident in New South

Wales and Victoria, the two most populous states with the greatest fiscal burdens. This is particularly so since state governments are responsible for hospital construction. However, in other countries, such as Britain, the concentration of investor owned hospitals has little to do with local fiscal conditions and instead is more of a reflection of the concentration of doctors and affluent consumers in the largest cities, particularly London.

The growth of primary care led health systems. Until the second half of the twentieth century, primary care was a 'cottage industry' in which generalist doctors, working largely alone, provided a service 'on demand' for people in times of sickness. This model pervaded western health care, although the method of paying practitioners was diverse, often made by 'third parties', either governments or through social or private insurance, and on either by fee for service, an annual capitation fee, or by salary. The defining features of this model were the economic and clinical autonomy of doctors who, while often practicing in groups, were not well integrated with other parts of the health system.

This model began to break down toward the end of the twentieth century under pressure from funders, both public and private, to control payments and improve the accountability for use of funds. This involved restrictions on both economic and clinical autonomy through both private and public mechanisms. In the private sector, during the1970 s, corporate investment in health organizations led to greater levels of control, in the United States, for example, through health maintenance organizations (HMOs). This economic pressure forced primary care doctors into even larger groups where they were subject to increased controls and collective consumer pressure. In publicly funded systems, both government regulation (through payment mechanisms and fee setting, or through control of supply of doctors) and managerialist approaches (such as treatment protocols, contracting, or fundholding) were seen to undermine professional roles of doctors.

In the late twentieth century, the Alma Ata Declaration on Primary Health Care (1977) promoted an alternative to the 'independent practitioner' model. The Alma Ata model, for which substantial research support emerged over the ensuing decades, suggested that the most effective health systems were those that had well resourced primary care services at the center of an integrated national health system. It defined the scope of primary care service and made it clear that effective primary care is delivered by a multidisciplinary professional team rather than doctors working alone, and that community involvement is essential to ensure good services. There is growing evidence that, particularly in disadvantaged populations, community involvement is an effective way to ensure the provision of services. In many developing countries, primary care is strongly community based, and in some western countries there is evidence of effective community participation in primary care. For example, in New Zealand, a small but important movement has emerged of community owned and run primary care in areas of high deprivation or in rural areas where local services were under threat. Although this is a small sector, it has been influential on public policy so that a national program of primary health organizations has been established to experiment with greater community involvement.

The transformation of primary care from cottage industry to regulated or corporatized service, with efforts in some places to strengthen both community orientation and policy leadership, indicates the way in which local contexts combine with international trends to provide different solutions in different places. In the UK, for example, policy since the 1990s has been directed toward 'a primary care led NHS' and in New Zealand the New Zealand Primary Health Care Strategy has reorganized primary care along lines consistent with Alma Ata principles. Many of the post Soviet states are developing stronger primary care services as the cornerstone of their health systems, which were previously dominated by hospitals.

Geographical Factors in Assessing Health System Performance

Geographical Bases of Service Organization

Decentralization is an inherent characteristic of most federal type health systems but even here, national governments in Australia and Canada, for example, have attempted to reduce geographical variations in health system outputs by implementing national legislation, which has aimed to provide benchmarks for minimum standards of care. However, in other countries, such as in the United States, where federal control of the health system is less strong, there are marked disparities in the provision of care between states. In particular, many southern and western states have poorer Medicare and Medicaid programs and generally higher rates of uninsured people.

In more centralist countries, such as New Zealand and the United Kingdom, in the early part of the twentieth century, decentralization, especially of responsibility for the provision of hospital services to local authorities and voluntary providers, also meant the evolution of wide geographical variations in the provision of care with wealthier regions inevitably having greater access to such services. However after World War II, governments assumed greater control over the funding, provision, and regulation of services, including the development of more rational funding systems.

However, in the 1980s and 1990s tensions between centralization and decentralization increased again with the development of internal markets and the consequent 'hollowing out' of responsibilities for the provision of hospital services. Despite more rational allocation systems, the implementation of the funder–provider split resulted in a decentralization of responsibility for providing services and the emergence of considerable inequalities between different places in access to services. While there were clear benefits of decentralization in terms of the emergence of new sets of providers and, in some cases greater community responsiveness, the result was increased fragmentation of services, reminiscent of the early twentieth century. As a consequence, there has been yet further restructuring resulting in increased centralization and control yet retaining some of the more beneficial features of the internal market.

Resource Allocation in Relation to Need

Most modern health systems have struggled to develop systems of allocation which target scarce resources to those most in need. Historically, the allocation of health funding was largely determined not by need but by the presence of existing medical resources, local patterns of wealth, or political factors including electoral marginality or professional resistance to change. In the United States, for example, the distribution of federal funds to cities involved in the 'maternal and infant health care' program in the 1960s simply reflected the existing distribution of medical resources and was unrelated to need. Similar patterns were evident in New Zealand where the allocation of hospital funding was unrelated to need, reflecting the existing supply of hospital beds. However, policies for universal access to care resulted in an increased focus both on equality of access and the 'equitable' access to health resources based on need, in both the primary and secondary sectors. As a consequence need based allocation formulae have been developed in most health care systems, the implementation of the 'Resource Allocation Working Party' (RAWP) hospital allocation formula and the Jarman index in primary care in the United Kingdom being two significant examples.

The development of need based allocation systems has posed problems, most importantly the difficulty of defining needs for care. Need can be defined in four ways: revealed needs (utilization), felt needs, objective assessments of need using direct (mortality) and indirect (deprivation) measures, and via 'expert' provider assessments of need. All measures pose problems, in part, because need is an infinite concept, but also because each measure focuses on different aspects of need. Using utilization data is valid if there are no barriers to the use of health services, but this is seldom the case, with care often unavailable, inaccessible, or unaffordable. Felt needs or 'wants', involving individual perceptions of need, suffer from over and underestimation of the true level of need particularly where high levels of disadvantage result in low expectations of the health system. The use of mortality data is also flawed because the links between morbidity and mortality are often weak. As a consequence, the use of indirect census derived indicators of need is often the preferred alternative. However, there are problems with the scale at which deprivation is measured, the timeliness of the data and the assumption that high levels of deprivation also mean that health services, if available, will be used. Finally 'expert' provider assessments of need, while useful in identifying at risk groups, are of limited use in practice because of the cost of developing robust measures and the high variability of individual estimates.

The limitations of specific approaches to measuring need have led to countries developing composite measures involving a mixture of deprivation and mortality measures together with aspects of population structure including ethnicity and rurality. Nevertheless, there continues to be disagreement over the implementation of needs based formula especially since slight changes in the composition of such formula, can have important geographical outcomes in terms of changes in funding levels and consequent changes in access to health services.

Locational Variations in the Provision of Health Services

The classic geographic approach has been to describe and explain spatial patterns of health resources in terms of political, economic, and cultural processes. These are important in both the primary and secondary sectors and shape distributional outcomes. Definition of an equitable distribution of resources is often contested, but is usually demonstrated through calculating the number of doctors or hospital beds relative to the population of an area. Geographers have used a range of methods including simple population/resource ratios, location quotients (which show the extent to which an area or set of areas departs from some norm), and gravity formulations which include both population and the number of providers weighted by distance from a given location. GIS derived measures of travel time to the nearest surgery or hospital, weighted by road characteristics, have also been used.

Using different measures reveal that inequalities in resource distribution occur at different scales, including geographic differences in the provision of resources between rural and urban areas and within cities themselves. These inequalities frequently reflect the 'inverse care law' whereby areas high in need have poor access to health resources. The extent to which the inverse care law persists reflects the extent to which governments have intervened to influence the distribution of resources. For instance, in the case of hospital care, some countries have traditionally been characterized by historical concentrations, particularly of teaching hospitals, in core regions (for example, London in the UK). Only in recent years have governments implemented more rational hospital funding systems with the aim of redistributing such facilities and encouraging other types of care. The presence of regional concentrations of hospitals was often characterized by 'Roemer's Law', where a 'built bed was a filled bed' and, as a result, hospital use largely reflected inequalities in the bed supply across localities rather than any measure of population need. Different patterns of ownership also resulted in inequities between regions. In the United States, for example, initially public and nonprofit hospitals were the dominant providers, particularly in the industrial northeast, but in recent years corporate providers have become more prominent, particularly in Sun Belt states.

In the primary sector the inverse care law has been more apparent, due both to the individualized nature of the service and resistance to any loss of professional autonomy on the part of medical practitioners. Consequently, the inverse care law is most evident in countries characterized by fee for service systems and which have less regulation of primary care providers. In fee for service systems the assumption has been that market forces will produce a more equitable distribution of doctors as increased provider competition leads to the migration of some existing and new practitioners to shortage areas where competition is less. In fact, given the presence of 'supplier induced demand' (SID) such outcomes are unlikely to occur or, if they do, will be the result of other factors such as the in migration of international medical graduates. In Australia, for example, despite the presence of more doctors, urban–rural differences in access to care have increased. By contrast, the opposite has occurred in countries with more regulation, such as the United Kingdom where there is less evidence of such 'patterned' inequality.

Explanations of inequalities have been threefold. Neoclassical explanations have focused on supply and demand and see policy outcomes in terms of these factors. By contrast, political economy approaches focus on broader structural changes in society and in medicine (such as the influence of uneven development, the effects of specialization, the 'white collarization' of the medical workforce, or the increased scale of medical practice) and promote policy options consistent with this view. Thus in the United States the development of family practitioners was seen as a way of increasing health care integration and ameliorating the decline of health practitioners in rural areas. Third, more individualized approaches have focused on the idea of 'place integration' and the important factors such as professional security, freedom, and identity that bond people to particular places. In this model problems of access and retention rates can be ameliorated by policy makers addressing such factors.

Variations in Service Use

While inequalities in service provision are one indicator of health system performance, perhaps more important is the actual utilization of services, which can be influenced by location. Many studies have shown a 'distance decay' relationship between utilization and distance from health facilities, both primary and secondary, with people who live further away from doctors or hospitals less likely to use these facilities. This relationship holds even after accounting for differences in population need. Such outcomes may result from the fact that distance, whether in terms of time or cost, deters patients from consulting their doctors. Alternatively, people may make a trade off between the costs and benefits of such consultation; where the benefits are perceived to be low then service use is less likely. This is particularly true for certain groups, such as the elderly or the poor for whom even small distances may be a significant barrier. This may occur because of childbearing responsibilities, lack of access to public transport, especially for sole parents with small children, or because of insufficient time due to other responsibilities. In view of such barriers, there is a need to adopt more sophisticated measures of access which incorporate the space–time constraints which may limit the use of services.

However, the impact of distance on consultation rates depends upon a number of other factors. For example, patients may not always use the closest doctor. In rural areas where there is a significant turnover of general practitioners patients may travel longer distances to urban centers for treatment by a regular doctor. In this case the benefits of personal contact will tend to outweigh the costs of distance. In urban contexts patients who change residence are often likely to continue consulting their regular doctor in preference to surgeries closer by. Alternatively, minority patients may seek out particular doctors who they feel are sympathetic to their needs.

The effects of distance decay on utilization are also evident in hospital care. In the United States in the nineteenth century, in what became known as 'Jarvis's Law,' patients in 'lunatic hospitals' were far more likely to have been committed from places near the hospital than from places further away. Such effects have also been evident in Britain in more recent times, with both deprivation and travel time having a significant effect on hospitalization rates generally. Unlike primary care, the explanation for distance decay effects in hospitalization is less clear because doctors rather than patients, at least in the case of public hospitals, normally choose the place of treatment. It may simply be the case that patients in remoter areas with fewer doctors and fewer consultations are less likely to be diagnosed and referred for hospital treatment. Or, it may be that certain groups who live further away from hospitals may not be referred by their doctors if the difficulties involved in travelling long distances outweigh the benefits of treatment.

While distance to providers has an effect upon levels of use, so does the actual concentration of doctors or hospitals. In fee for service systems an increased concentration and greater competition between providers leads to 'supplier induced demand.' This occurs because among doctors the incentive is to treat more not fewer patients in order to maintain income levels, and consequently may encourage repeat visits or raising of fees. Similar trends can be seen in the private hospital sector and even in the public sector. In the latter case, given incentives to maintain high rates of occupancy, more hospital beds inevitably means higher rates of hospitalization.

The role of geographic factors affecting service use, therefore, must be considered in conjunction with compositional factors. Changing levels of social deprivation may affect the ability to use services and, through social exclusion, certain groups may be disadvantaged due to gender, ethnic, and social biases operating within the health system. Important here is an understanding of how particular groups of patients cope with disadvantage and how, in turn, medical practitioners aid them or not in this coping process.

The Role of Context in Health Service Provision

While health services have a distinctly spatial component, attention must also be focused on the role of place or contextual factors and how these influence patterns of health service delivery. Three brief examples will be considered: therapeutic landscapes, landscapes of power and social exclusion, and landscapes of inequality.

Therapeutic landscapes

Geographical analysis has focused on the ways in which the places of delivery of health services can be designed according to the principles of therapeutic landscapes. The concept of therapeutic landscapes has been used to present geographical ideas about factors important for well being. It applies perspectives from cultural geography to explore environmental, individual, and societal factors that come together in the healing process in both traditional and nontraditional landscapes. While traditional landscapes refer primarily to the beneficial effects of healing processes in certain natural environments, nontraditional landscapes are situations in which healing might take place within the health care setting such as a physician's office or hospital ward. These landscapes have particular meaning and symbolism for people and affect their health and social well being. Some places have attributes which make them beneficial to psychological and physical health while other environments may be detrimental to health. For example, a place may influence well being because it is a place of refuge and relaxation. On the other hand, it may be symbolically important by being linked to personal identity or it may encourage a sense of belonging through the use of certain design cues and visual symbols.

Traditionally, doctors' surgeries have been medicalized spaces with little citizen input. More recently, with the development of third sector service models and more public participation in the health system, in some instances they have been transformed into more community focused spaces. This has led to a blurring of the conventional distinctions between surgery spaces and other public spaces while, at the same time, creating a sense of community ownership and control. Similarly, some hospitals, which have conventionally been places of medicalized care and highly functional elements in the urban landscape, have adopted strategies for creating therapeutic landscapes. Childrens' hospitals, in particular, have developed more playful and relaxed interior spaces reminiscent of theme parks with the aim of providing a more appropriate environment for sick children. While such ideas are in their infancy, they represent a shift away from modernist hospital forms to postmodern hospital spaces which, while not abandoning the biomedical model, seek to allow other perspectives to be included. Such developments also represent hospitals' attempts to create a greater uniqueness as a way of responding to funding problems created by the restructuring of public hospital systems.

Social exclusion in health systems

While geographic studies of utilization have examined inequalities in access to health care, less attention has been devoted to the 'quality' of health services received by groups in the community. This is an important question given the need to evaluate the effectiveness of health expenditure in improving population health. There is a growing body of evidence which suggests that, although lower income groups may be heavier users of services, higher income patients are more likely to receive more comprehensive care from their doctors. Researchers have found that time spent with the doctor, or the likelihood of the request or access to diagnostic tests, including follow up, varies inversely with socioeconomic status. Similar trends have been observed in the hospital system with lower income groups having shorter lengths of stay and in some cases certain minorities reported as less likely to be referred for surgery than other groups with comparable needs. The fact that there are both socioeconomic and ethnic biases in the way such patients are treated suggests ingrained forms of exclusion within different health systems. Whether such findings reflect communication barriers or the existence of active discrimination by health professionals is uncertain, but the fact that such biases arise in a number of different areas of treatment (general practice, mental health care, cardiac surgery) in different health care systems indicates that they are pervasive.

Landscapes of inequality and challenges for public health

Health promotion is an important part of health systems and public health initiatives have developed in most western countries to combat lifestyle diseases particularly those associated with smoking and obesity. These are most common among more disadvantaged groups with high levels of deprivation and are particularly pronounced among certain minority populations even after differences in socioeconomic status have been taken into account.

However, the effectiveness of such interventions has been questioned especially given their limited success in changing health behaviors and reaching those most in need. For example, while smoking rates have shown a significant decline among the middle class, the decline is less for lower income groups, thus leading to widening health gaps. Part of this lack of effectiveness in inducing behavior change relates to the vehicles through which public health promotion messages are transmitted and to the influence of poverty and inequality as constraints on behavior change.

With respect to health promotion, traditionally the voluntary sector has taken the initiative in developing health promotion campaigns. However, a number of studies have indicated a discordance between the spatial distribution of voluntary sector resources and welfare needs and this also appears to be the case in terms of the membership and health impacts of such organizations. Thus while social mechanisms, such as collective socialization or transmitting health promotion messages more effectively through community organizations may all be health promoting they can also be highly exclusionary, especially if benefits are not evenly shared among different social or ethnic groups.

It is clear that poverty and growing levels of social inequality act as a constraint to behavior change, although this is likely to be context specific. This may reflect not only the level of poverty and inequality and the extent to which such concerns outweigh the impact of any public health messages, but also differences in organizational structure and the ways in which such messages reach the community. For example, rural, more decentralized voluntary organizations may be more effective in reaching deprived groups than larger more centralized and corporate voluntary sector groups.

Do Health Systems Matter? Service Provision and Health Outcomes

At the most basic level health systems matter because of the contribution they make to meeting needs for care and the positive benefits to health. However, this has not always been true. In the nineteenth century medical misadventure (or the idea of 'iatrogenesis') was common in hospitals and today some groups are more likely to experience adverse effects from treatment. In recent times the question of the contribution that health care makes to health outcomes has become more important as governments and international agencies have increasingly formulated health outcome goals. Such moves have focused not only on improving life expectancy but also on the goal of reducing health, inequalities between different groups. However, the contribution that health systems make to health is difficult to judge as different kinds of evidence seem to give conflicting answers.

At the level of interventions against particular diseases there is substantial evidence that large improvements in health can be achieved at reasonable cost such as anti malaria campaigns in many developing countries which have involved early case detection and treatment. At the level of overall progress in health, the 1999 World Health Organization World Health Report cited evidence that across a sample of 115 low and middle income countries scientific and technical progress in health care explained almost half the reduction in mortality between 1960 and 1990 in these countries.

However, an important question is whether geographic differences in access and use of service have any effect on health outcomes. In terms of access to emergency care, for conditions such as road accidents, there is evidence that when access is poor and response times are long then health outcomes are adversely affected. Similarly, there is evidence from the United States which suggests that levels of infant mortality and non normal births are related to geographic differences in maternity care, particularly in rural areas. Such trends possibly reflect the fact that mothers have to wait longer to see their obstetrician and, without good access, patients with complications may experience delays in access to neonatal intensive care services. The declining number of doctors providing obstetrical care has also accentuated such trends.

With respect to nonemergency care there is evidence that for some cancers survival rates vary geographically with patients in more deprived areas having worse survival rates. This may reflect delays in diagnosis as, for example, in the case of melanoma. Unlike most cancers, more affluent groups are more likely to contact mela noma but are less likely to die from it due to the greater likelihood of earlier detection of the disease. Similar trends are evident for other lifestyle diseases such as diabetes where middle class patients are more likely to have been diagnosed earlier and to be involved in edu cational programs to improve their condition.

While, there is evidence that geographic differences in access to and variations in the quality of services have an impact upon population health, this is countered by other evidence which suggests that health systems make little or no difference and that most of the improvements in health status can be linked to positive changes in the physical and social environment. This was certainly the case during the nineteenth century when public health initiatives, such as improved sanitation, accounted for most of the decline in mortality for conditions such as tuberculosis. This occurred before significant medical interventions for the disease. Furthermore, comparisons across countries, while showing that per capita income is strongly related to variations in health status, have also found that there is little independent connection with specific inputs such as number of doctors or hospital beds.

This conflicting evidence can perhaps be reconciled in two ways. First, many studies are cross sectional in na ture and therefore do not adequately test the relationship between health service inputs and health outputs because of lag effects. Second, the population health benefits of particular forms of health expenditure are highly variable both between and within countries and therefore the impact of health resources upon health is far from uni form. For example, it is increasingly recognized that countries (and regions) with higher levels of primary health care expenditure tend to have better health. In the United States, for example, research has shown that among the health care variables, primary care is the most significant variable related to better health status and correlated with both lower overall mortality and higher levels of self reported health. Similarly, countries char acterized by more universal primary health care coverage tend to have higher rates of service use and lower socioeconomic status differences in mortality. Such fac tors need to be borne in mind when assessing the impacts of health services upon health outcomes.

See also: Care/Caregiving; Healthcare Accessibility; Health Geography; Health Inequalities; Health Services Restructuring; Mental Health; Neoliberalism; Voluntary Sector; Welfare Reform.

Further Reading

Barnett, J. R. and Brown, L. J. (2006). 'Getting into hospitals in a big way': The corporate transformation of hospital care in Australia. *Environment and Planning D (Society and Space)* 24, 283 310.
Curtis, S. (2004). *Health and Inequality: Geographical Perspectives*. London: Sage.
Gatrell, A. C. (2002). *Geographies of Health: An Introduction*. Oxford: Blackwell.
Gesler, W. (1992). Therapeutic landscapes: Medical geographic research in the light of the new cultural geography. *Social Science and Medicine* 34, 735 746.
Gesler, W. and Ricketts, T. (eds.) (1992). *Health in Rural North America*. New York: Rutgers University Press.
Joseph, A. E. and Phillips, D. R. (1984). *Access and Utilisation: Geographical Perspectives in Health Care Delivery*. London: Harper and Row.
Milligan, C. and Conradson, D. (eds.) (2006). *Landscapes of Voluntarism: New Spaces of Health, Welfare and Governance.* London: Policy Press.
Mohan, J. (2002). *Planning, Markets, and Hospitals*. London: Routledge.
Moon, G. and North, N. (2000). *Policy and Practice: General Medical Practice in the UK*. London: Macmillan.
Nolte, E., McKee, M. and Wait, S. (2005). Describing and evaluating health systems. In Bowling, A. & Ebrahim, S. (eds.) *Health Research Methods*, pp 12 43. Maidenhead: Open University Press.
Pinch, S. (1997). *Worlds of Welfare: Understanding the Changing Geographies of Social Welfare Provision*. London: Routledge.
Ranade, W. (1997). *Markets and Health Care: A Comparative Analysis*. Harlow: Longman.
Ricketts, T. C., Savitz, L. A., Gesler, W. M. and Osborne, D. N. (1994). *Geographic Methods for Health Services Research. A Focus on the Rural Urban Continuum*. Lanham, MD: University Press of America.
Starfield, B., Shi, L. and Macinko, J. (2005). Contribution of primary care to health systems and population health. *Milbank Quarterly* 83, 457 502.
World Health Organization. (2000). World Health Report 2000. Health systems; improving performance. *Report of the World Health Organization*. Geneva: WHO.

Relevant Websites

http://www.commonwealthfund.org
Commonwealth Fund: International Health Policy.
http://www.euro.who.int/observatory
European Observatory on Health Systems and Policies.
http://www.who.int/topics/health systems/en/
WHO health systems website.

Healthcare Accessibility

J. S. Humphreys and K. B. Smith, Monash University, Bendigo, VIC, Australia

Glossary

Access It refers to the capacity of people to obtain healthcare at the right place and right time regardless of location, socioeconomic factors, or cultural background.

Availability The overall extent to which healthcare systems provide facilities and services to meet the needs of individuals and populations.

Distance Decay It occurs when the rates of utilization of a particular facility diminish with increasing distance from that facility.

Healthcare Accessibility The extent to which individuals and communities can easily obtain health services when needed.

Primary Care It refers to the provision of first contact, person-focused ongoing care over time that meets the health-related needs of people.

Primary Healthcare (PHC) Primary care applied on a population level.

Range It refers to the maximum distance people will travel to obtain healthcare services.

Threshold It refers to the minimum number of people needed to support the existence of a service.

Introduction

Accessibility to healthcare is recognized as a basic right by international bodies such as the United Nations and World Health Organization (WHO), and remains a guiding principle underpinning the policies for the delivery of health services throughout developed and developing countries. In particular, accessibility to primary healthcare services is a health priority for most countries throughout the world. Primary healthcare is usually the first level of care for patients in terms of health protection, promotion, and treatment, and providers of primary care act as 'gatekeepers' to other more specialized health services. Primary healthcare is also the level of care that can be most readily and cost effectively extended to the greater part of the population to improve the effectiveness, efficiency, and equity of healthcare delivery.

Defining Access and Accessibility

Throughout the social sciences, medicine, and other disciplines, the concept of accessibility is often used interchangeably with the broader concept of access. The significance of the distinction between these concepts is often not acknowledged or fully appreciated.

Access refers to the capacity of people to obtain healthcare at the right place and right time regardless of location, socioeconomic factors, or cultural background. According to Penchansky and Thomas, five interrelated dimensions are contained within the broad concept of access – availability, accessibility, accommodation, affordability, and acceptability. Availability refers to the overall extent to which healthcare systems provide facilities and services to meet the needs of individuals and populations. The availability of services is crucial in considering accessibility since services must first exist for them to be accessible. Accessibility is the relationship between the location of supply and the location of clients. The appropriateness and manner in which services are organized to meet patients' needs come under the concept of accommodation. Affordability is the association between cost of services and the ability of clients to pay, while acceptability refers to the attitudes of clients toward service providers and provider attitudes about characteristics of their patients. Factors affecting 'access' to health services thus include the physical and temporal availability of services (including waiting times, choice, and appropriateness), the numerical and professional adequacy of the healthcare workforce, and the cost of service to the individual.

Healthcare accessibility refers to the extent to which individuals and communities can easily obtain health services when needed. The most basic dimension of healthcare accessibility is geographical reflecting the spatial separation of healthcare consumers (the demand side of the concept) and the healthcare facilities and services themselves (the supply side). Measures of accessibility therefore quantify the relationship between obstacles arising from the geographic location of services (travel time, cost, distance) and the capacities of potential healthcare users to overcome these obstacles. Factors affecting 'accessibility' to health services include remoteness, travel time to services, transport (including mode, road conditions, patient mobility), socioeconomic barriers such as income, language and ethnicity, health literacy, and health status (including type or urgency of health problem, disability, and support required).

A distinction can also be made between 'locational' and 'effective' accessibility, the former referring to the level of geographical proximity of services and the latter taking account of when a facility is open, whether it is socially and financially available, and whether potential clients have the time to use the service.

The formulation of a precise definition for healthcare accessibility is inevitably context dependent, with the concept viewed differently by providers, consumers, governments, and funding bodies. For this reason, and given the diversity of local environmental, social, cul tural, political, and organizational factors that must be taken into account, it is impossible to apply the same models of accessibility to every country.

Nevertheless, throughout the world, the definition and measurement of access and accessibility is critical to the planning, funding, and implementation of programs to improve healthcare delivery, in assessing equity in healthcare provision and service utilization patterns, and in evaluating health outcomes, since spatial variations in levels of morbidity and mortality will reflect, among other things, differences in healthcare accessibility.

Measuring Accessibility

Since the advent of computers and the 'quantitative revolution' of the 1960s and 1970s geographers have been able to undertake detailed spatial analysis and modeling of service locations and accessibility differences at a variety of scales from the national to the local or intra urban. In the 1980s and 1990s greater emphasis was placed on the social justice, political, economic, and cultural aspects of accessibility. Most recently geographic information systems (GISs) have transformed the way accessibility can be analyzed and displayed through the management of larger data sets.

Internationally, the importance of accessibility to health services remains one of the primary concerns of medical geography, exemplified in the work of such geographers as Paul Knox, David Smith, Robert Earickson, Gary Shannon, Alun Joseph, David Phillips, John Eyles, Robin Kearns, Graham Moon, Ross Barnett, Sara McLafferty, Jonathan Mayer, Sarah Curtis, and Ellen Cromley among many others.

The core spatial dimension of accessibility is distance. Common measures include physical or linear distance (sometimes distinguishing between road conditions), mode of transport, travel time, travel cost, distance to nearest provider, and individual mobility. The universal tendency to use a measure of physical distance rather than travel time to measure spatial accessibility reflects the difficulties, and underplays the added complexities associated with, measuring the varying methods of jour neys, transport availability and scheduling, road con ditions, terrain, and any other natural barriers people confront when accessing services.

Within any population, accessibility to and the use of health services will also differ across a range of 'nonspatial' considerations – financial, behavioral, cul tural, organizational, and socioeconomic. Age, gender, education, socioeconomic status, ethnicity, and health status, for example, variously act as barriers or facilitators when accessing services. Financial barriers such as the costs of transport, childcare, or earnings foregone may affect attendance at health services. Ability to pay for services is especially challenging for the poor and may well be a critical determinant of whether care is sought.

Since different measures of accessibility emphasize different aspects of the concept, the issue or question under examination should determine the choice of measure. Objective measures, such as service availability, population, socioeconomic factors, health status, distance, and geography are the most obvious factors influencing consumers' experience of access and accessibility. How ever, measures must also be sensitive to cognitive effects on how consumers view their access to health services.

Assumptions that improved accessibility results from closer or the presence of more services, that accessibility is the same for all people in a location, that the fastest and/or cheapest mode of transport will necessarily be used to access services, or that the nearest facility will always be used do not necessarily hold. For example, in Kenya it was found that where health services were closely spaced, patients were more likely to attend the facility offering higher order services. Similarly, in Australia it has been shown that choice is an important consideration in the study of geographical accessibility, with people bypassing nearer services, if they are perceived to be inappropriate or of inferior quality to those further afield, despite the higher monetary, time, and other costs involved.

Because social, economic, and cultural factors, often difficult to measure, may override simple proximity in determining use of services, none of the commonly used measures of accessibility is without its limitations. While simple linear distance measures and provider to popu lation ratios are commonly used in making gross com parisons of accessibility, neither allows for individual behaviors and constraints or permeations across regions or boundaries. Some attempts have been made to develop more complex version of these ratios, such as 'floating' catchment areas that provide provider population ratios within designated travel time catchments that can be moved around.

Role of GISs

The use of GISs to measure accessibility has greatly improved the accuracy and sophistication of analyses. As with any technology, value of output depends on the quality of data regarding service availability, distance measures, travel modes, road conditions, and actual utilization of services.

Commonly, GIS has been used to measure a variety of accessibility aspects, for example, proximity to healthcare facilities, or distance to nearest treatment point, to quantify relationships between road accessibility and disease prevalence, to define areas for the management of primary healthcare, to calculate population per service ratios, to measure impact of travel time, and to analyze spatial variations in clinic use patterns.

Two recent examples of GIS technology for measuring accessibility to health services are WHO's AccessMod and the Pan American Health Organisation's (PAHO) GIS Epi (SIG Epi). These applications use travel times and distances as the geographical indicators of accessibility and incorporate data on population, landcover, transport network, elevation, administrative boundaries, and health facility location to model accessibility to health services. The former measures physical accessibility and population coverage capacity of an existing health facility network to generate catchment areas. Nevertheless, even these two applications are limited by the assumption that access is to the nearest health service rather than incorporating actual utilization data to more accurately reflect patient behavior.

As an example of its contribution to the analysis of accessibility in this field, GIS technology was employed in a study of Kenyan malarial programs to define physical accessibility and utilization of health services in order to assist with healthcare planning and resource allocation based on spatial data relating to health service location and population (see **Figure 1**). Distance to health services was measured for within each of four districts of Kenya using actual transport networks and taking into account topography, natural barriers, mode of transport (in this case primarily pedestrian), and actual use of services.

Advances in GIS geocomputing capabilities enable the combination of spatial, nonspatial and temporal accessibility information in one system which, pre GIS, would have been heavily time consuming, expensive, and probably too complex. Although GIS does not change the conceptual bases of accessibility analysis, it enhances the accuracy and flexibility of accessibility measures by facilitating the storage, management, and integration of very large data sets for large geographical areas. GIS also allows for the layering of information such as population distribution, location of health services and actual usage of facilities, and the mapping and visual display of the outcomes. For example, a study in South Africa utilized GIS to georeference 23 000 homesteads to analyze modal usage patterns, constructing polygons for each fixed clinic and then superimposing these on to the homesteads. The value of GIS lies in its ability to present a high degree of spatial and temporal resolution of a given environment, for example, representing traffic congestion at various times of day, overcoming

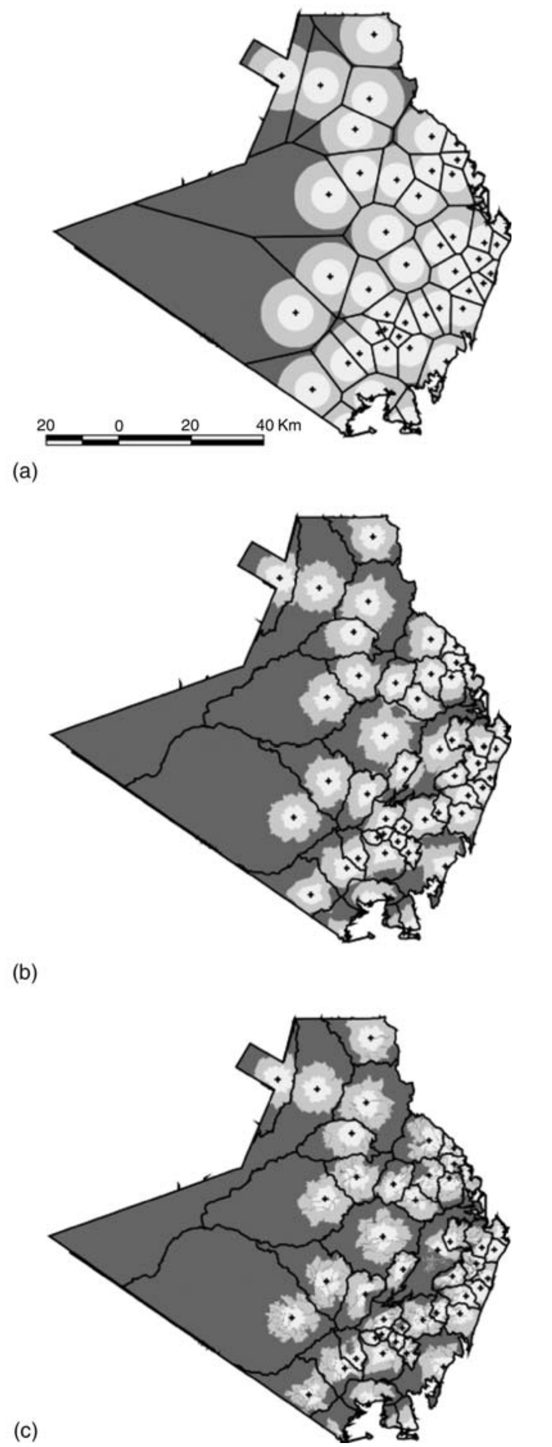

Figure 1 Modeling distances to government health services in Kwale district in Kenya: Maps of Kwale showing access to government health services based on travel time (hours) for (a) the Euclidean, (b) the transport network, and (c) the adjusted transport network models. + government health facility; catchment area; 0 0.5 h; >0.5 1 h; >1 h. All maps are at the same scale. The North direction is toward the top of the page. Reprinted with permission from Noor, A. M., Amin, A. A., Gething, P. W. *et al.* (2006). Modelling distances travelled to government health services in Kenya. *Tropical Medicine & International Health* 11(2), 188 196.

the assumption of constant travel speed or time for journeys.

Accessibility and the Spatial Organization of Health Services

Accessibility to health services is dependent on the way healthcare is delivered and the resultant spatial organization of health services, the location of people in relation to available facilities, physical and topographical barriers, road networks, modes of transport, as well as funding arrangements (e.g., user pays, free public service, private insurance).

Clearly, there are enormous differences between developed and developing countries in the nature of health systems and the provision of health services. Despite these differences the concepts of service threshold and range have universal applicability when considering the delivery of and accessibility to health services.

The threshold of a service refers to the minimum number of people needed to support the existence of that service. The range for a service outlet refers to the maximum distance people will travel to obtain it. Threshold and range vary according to the type of service. Usually the threshold and range for primary healthcare activities such as health promotion, early intervention, immunization, screening, and some domiciliary, rehabilitative, and supportive care, are small. In contrast, the threshold and range of specialized (higher order) and infrequently used services (such as surgery) extend appreciably, since they require significantly more people to sustain the service.

For these reasons it is often impossible for sparsely populated areas to support anything but low order services locally. Nevertheless, even the smallest rural community will need some specialized healthcare or treatment (e.g., for farm injury or road accidents). It is therefore desirable that accessibility to the appropriate health service be facilitated by other means, for example, by the availability of emergency evacuation and transport facilities to take patients to the appropriate service (e.g., trauma center or hospital) as soon as possible and desirably within the first hour after injury. Nor should individuals requiring referral away from their community for specialist treatment be disadvantaged through barriers to accessibility such as lack of public transport or ownership of a vehicle, impassable roads, excessive cost, or poor scheduling of appointments, but this goal remains to be achieved in many parts of the developing world, as in rural and remote areas of some parts of the developed world, like Australia.

In general, where the threshold of any service cannot be met within its range, that service will not normally exist locally. One way of expanding the threshold (or catchment area) is by the use of mobile services which take the service to the people. In this way, communities (each too small by itself to sustain a full time service locally) can be serviced collectively, albeit on a periodic basis (e.g., in the case of the Flying Doctor Service clinics and outreach programs in Australia, South Africa, and Malaysia). One problem for communities that are served only by periodic mobile services is that the occurrence of health problems often cannot be timetabled to meet the service schedule. Under such circumstances the availability of appropriate transport to the appropriate healthcare service at the time of need is therefore critical.

Patterns of Healthcare Accessibility

Organizational and institutional arrangements underpinning the provision of health services, such as health policies, methods of payment for services, and cross border arrangements in the case of federal countries like Canada and Australia impact profoundly on healthcare accessibility. Nonetheless, distance, and the density and dispersal of population, remain key elements in the analysis of health service accessibility. In countries such as Australia, Canada, Brazil, and Russia, with large land masses and dispersed populations, distance is particularly salient for residents of rural and remote areas not only due to the higher costs of providing services to them but also because of difficulties associated with recruiting qualified staff.

In smaller countries distances to services tend to be less so that accessibility is generally perceived as less problematic. However, even in countries such as Scotland, Norway, and New Zealand, problems of accessibility are significant for some services and populations, particularly where the impact of extreme climatic conditions impact upon transport. Other accessibility problems can arise in unstable environments such as situations of political conflict which restrict population movement or alternatively generate mass migration, or lead to the loss of facilities providing care. This is evident in places such as Palestine, Sudan, and Iraq, where violent political conflict severely limits the physical accessibility of the population to primary healthcare.

In all societies significant differences in accessibility to health services exist between urban and rural areas (see **Figure 2**). In general, accessibility issues are most acute in rural and remote areas characterized by low population densities and shortage or dispersal of health services. Shortages of doctors and other healthcare workers and the uneven distribution of providers contribute to differing levels of accessibility in rural areas. For example, in Australia, people in rural and remote areas

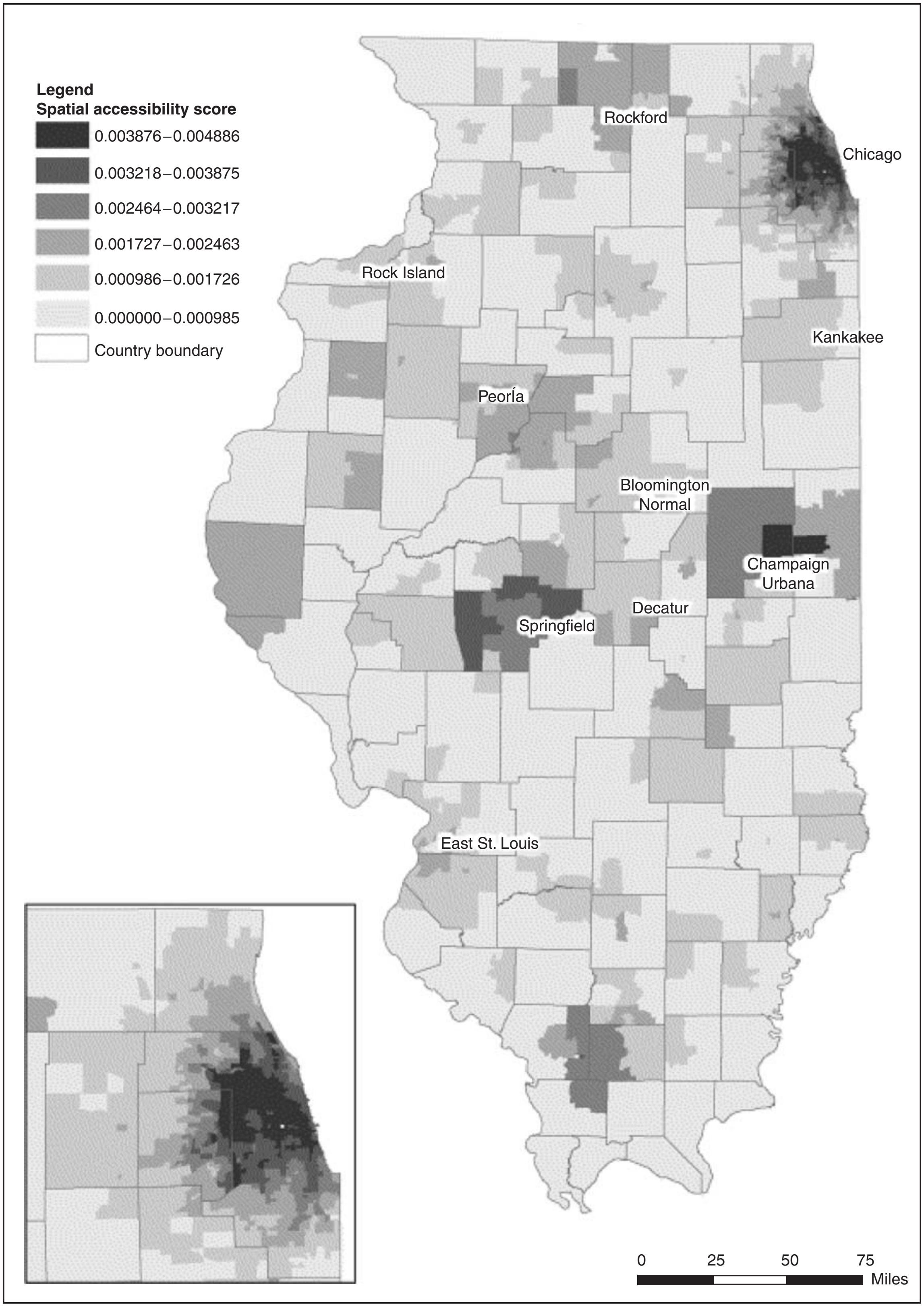

Figure 2 Rural urban differences in accessibility to primary healthcare in Chicago and Illinois, USA using two-step floating catchment methodology. Reprinted with permission from Wang, F. and Luo, W. (2005). Assessing spatial and nonspatial factors for healthcare access: Towards an integrated approach to defining health professional shortage areas. *Health & Place* 11(2), 131 146.

often experience greater difficulty in accessing healthcare than those living in metropolitan or other major urban areas (see **Figure 3**).

Because urban populations are typically less dispersed than rural populations it is usually easier to provide readily accessible services. Nevertheless, intra urban accessibility differentials can still be significant for marginalized population groups characterized, for example, by high numbers of transport poor elderly or disabled citizens, or by low socioeconomic status. Other access disadvantaged groups commonly include young families living in outer metropolitan areas or urban fringes where there is often minimal public transport, or those without access to private transport (see **Figure 4**).

The Changing Patterns of Accessibility and Health Services

The impact of globalization may be seen in changing patterns of healthcare accessibility. In Western countries fiscal concerns with the escalating costs of providing health services to aging populations, problems of workforce supply and recruitment, increased use of technology and consumer expectations have all led to the rationalization or closure of many existing local services in rural areas. Economies of scale coupled with increasingly expensive modes of treatment have resulted in increasing centralization of services in larger centers. Centralization of health services in larger centers has further exacerbated the difficulties confronting rural and remote populations in accessing health services, a case in point being the current reduction in obstetric services in rural regions of many Western countries.

The nature of transportation systems also impacts on accessibility to health services. Private transport availability can significantly increase accessibility to services relative to public transport. Road networks, road type, and conditions may all affect the time taken to access primary healthcare services; mobile services and home based care may improve accessibility by bringing services to those who have difficulty accessing fixed clinics, as may the new telehealth technology. The perception of telemedicine as the panacea for overcoming the tyranny of distance is yet to be realized, however, with its greatest success to date being in information transfer, education, and training rather than in the sustained delivery of clinical services to remote communities.

It should by now be clear that accessibility is a relative and certainly not static concept for it changes according to individual circumstances, such as the resource availability, health status, the level of support needed, stage in the lifecycle, or the availability of transport.

Accessibility and the Utilization of Health Services

Evaluation of healthcare accessibility is often achieved by examining service utilization rates. Utilization is seen as a record of 'revealed' accessibility. Some caution is required here, however, since service utilization will not necessarily occur even under optimal service distribution and proximity conditions. Accessibility should not be conflated with the utilization of health services, because patterns of utilization alone will not reflect the extent to which services are not accessed at times of need due to their nonavailability or other intervening barriers. In Nigeria, Africa's most populous country, for example, the WHO found that the overall availability, accessibility, quality, and utilization of health services has been stagnating or even decreasing significantly during the past decade due to being poorly equipped, and lacking in essential supplies and qualified staff.

Nonetheless, in aggregate terms there does appear to be a direct relationship between accessibility, health

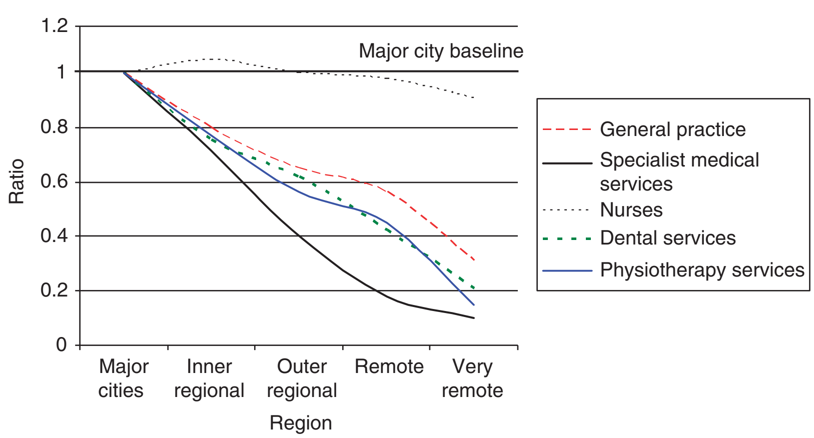

Figure 3 Practitioner to population ratios according to rurality in Australia. Reprinted with permission from Productivity Commission (2005). *Australia's Health Workforce.* Canberra: Productivity Commission.

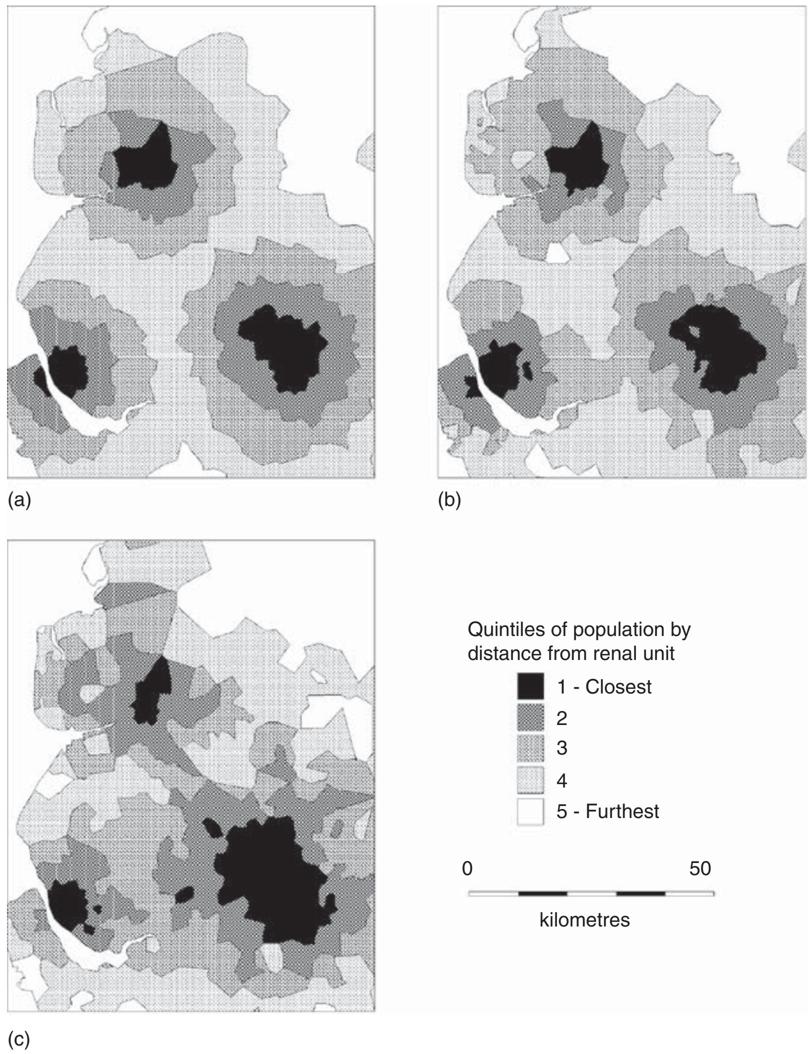

Figure 4 Intra-urban differences in accessibility to renal replacement therapy in England. (a) Crow-fly distance, (b) road travel distance, and (c) road travel time. Reprinted with permission from Martin, D., Roderick, P., Diamond, I., Clements, S. and Stone, N. (1998). Geographical aspects of the uptake of renal replacement therapy in England. *International Journal of Population Geography* 4(3), 227 242.

service utilization, and health outcomes. Distance to a health service impacts on utilization, with people closer to a health provider more likely to seek care when ill than those at a greater distance. This 'distance decay' phenomenon, where the rates of utilization of a particular facility diminish with increasing distance from that fa cility, is evident in both developed and developing countries (see **Figure 5**).

In South African rural communities this distance decay is evident at just under 5 km, while in Papua New Guinea the point at which 50% of potential attendees were lost was only 3.5 km compared with 3.2 km in Uganda and 3.4 km in Nigeria. In Kenya, it was found that usage peaked at 2–3 km from facilities. In Zambia how ever, the impact of distance on healthcare utilization was more pronounced for those living more than 40 km from a health provider.

Differences between developing and developed countries in their burdens of disease and resource ability generate enormous differences in their capacities to provide health services. In developing countries with large proportions of their population living below the poverty line, where the infrastructure is lacking and distances to health facilities can be large, financial ac cessibility is likely to be a critical determinant of whether care is sought. In such circumstances public health fa cilities are often concentrated in hospitals in urban centers, to the disadvantage of accessibility among and thus the health status of rural populations. In 2004, for example, it was noted by the Director of the Pan American Health Organization that 107 million people in Latin America and the Caribbean lacked access to healthcare for geographic reasons and a further 120 million for economic reasons.

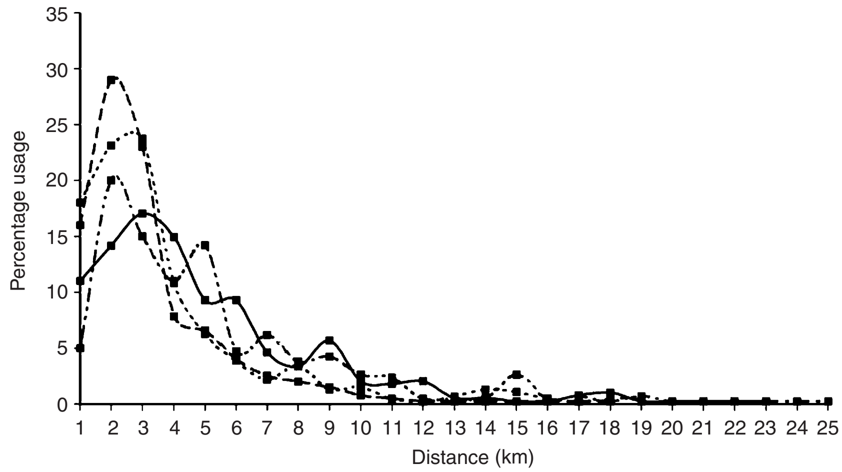

Figure 5 Distance-decay when accessing clinical services in four districts of Kenya. Cumulative health services use against traveled to government health facility where treatment was sought for the four districts. The line ·····■····· represents Greater Kisii, - - -■- - - represents Bondo, –··–■–··– represents Kwale, and ——■—— represents Makueni. The *y*-axis shows the cumulative percentage of patients in each district who traveled to a health facility at the distance (km) value shown on the *x*-axis. The cumulative percentages were computed by summing all patients who went in a district and used a facility at each distance on an 11-point scale (0 1, 1 2, 2 3, 3 4, 4 5, 5 6, 6 7, 7 8, 8 9, 9 10, and >10 km). Eighty-four percent in Bondo and 80, 64, and 62% of patients in Greater Kissi, Kwale, and Makueni used health facilities within 5 km, respectively. Reprinted with permission from Noor, A. M., Zurovac, D., Hay, I. S., Ochola, S. A. and Snow, R. W. (2003). Defining equity in physical access to clinical services using geographical information systems as part of malaria planning and monitoring in Kenya. *Tropical Medicine & International Health* 8(10), 917 926.

Accessibility and Health Outcomes

Measures of health status and improvement in health outcomes are contingent upon access to health services and interventions. Accessibility to healthcare services impacts both on life chances (in terms of promotion of health and on the burden of disease) and also survival because of the delays involved in obtaining care. In Western countries, studies have shown that the variations in accessibility to healthcare services between rural and remote areas are reflected in lower rates of cancer survival, significantly higher levels of displacement and social disruption for pregnant women, and higher death rates from coronary heart disease.

Accessibility is also an important indicator of equity in health systems. While a rural location does not always translate into health disadvantage, equity differences do tend to parallel accessibility differentials. End stage renal disease in Indigenous Australians is higher in remote areas than urban regions, highlighting inequitable accessibility to treatment facilities for remote patients. It has also been shown that centralization of facilities has impeded rural people's accessibility to radiotherapy services, generated significant geographical inequities in rural women's accessibility to bulk billing and obstetric services, and increased their out of pocket expenses relative to women in urban areas. Where healthcare funding is tied to the provision of health services, inequity can result from lack of accessibility to available services and is usually reflected in underutilization of health services at times of need.

Conclusion

The concept of healthcare accessibility is central to ensuring that the provision of healthcare services meets the health needs of individuals and populations. The core dimension of accessibility to healthcare services is undoubtedly geographical. However, the interrelationships between the spatial dimensions and other aspects of societal organization ensure that the existence of, or proximity to, healthcare services within a geographical location does not of itself ensure accessibility to those services.

While the physical accessibility of services may play a major role in influencing patient choice of health facility and therefore the use of a service, health consumers do not necessarily use the nearest health service. Regardless of location, accessibility differs from individual to individual since choice of facility reflects a host of other socioeconomic, cultural, perceptual, and organization factors in addition to the importance of ease of accessibility. Overcoming the friction of distance also has different implications for individuals, for example, those with mobility problems or lacking private transport. Moreover, the nearest facilities may not be accessible to poor people if payment for services is required. In general, however, geographical distance is closely related to travel time and transport costs, with longer distances involving greater time and costs.

Accessibility to healthcare has a direct impact on the burden of disease and is an important indicator of the performance of any national health system. Geo graphical, financial, sociocultural, and temporal accessi bility are key dimensions of effective, integrated, and equitable health systems.

See also: Health Services Restructuring; Regional Inequalities; Rural Populations.

Further Reading

Campbell, S., Roland, M. and Buetow, S. (2000). Defining quality of care. *Social Science & Medicine* 51(11), 1611 1625.

Guagliardo, M. (2004). Spatial accessibility of primary care: Concepts, methods and challenges. *International Journal of Health Geographics* 3(1), 3 16.

Humphreys, J. S., Mathews Cowey, S. and Weinand, H. C. (1997). Factors in accessibility of general practice in rural Australia. *Medical Journal of Australia* 166(11), 577 580.

Joseph, A. E. and Phillips, D. R. (1984). *Accessibility and utilization: Geographical perspectives on health care delivery*. London: Harper & Row.

Noor A. M., Amin A. A., Gething P. W. *et al*. (2006). Modelling distances travelled to government health services in Kenya. *Tropical Medicine & International Health* 11(2), 188 196.

Noor, A. M., Zurovac, D., Hay, I. S., Ochola, S. A. and Snow, R. W. (2003). Defining equity in physical access to clinical services using geographical information systems as part of malaria planning and monitoring in Kenya. *Tropical Medicine & International Health* 8(10), 917 926.

Penchansky, R. and Thomas, J. (1981). The concept of access: Definition and relationship to consumer satisfaction. *Medical Care* 14(2), 127 140.

Productivity Commission (2005). *Australia's Health Workforce*. Canberra: Productivity Commission.

Starfield, B. (2006). State of the art in research on equity in health. *Journal of Health Politics, Policy and Law* 31(1), 11 32.

Tanser, F. and le Sueur, D. (2002). The application of geographical information systems to important public health problems in Africa. *International Journal of Health Geographics* 1(1), 4.

Wang, F. and Luo, W. (2005). Assessing spatial and nonspatial factors for healthcare access: Towards an integrated approach to defining health professional shortage areas. *Health & Place* 11(2), 131 146.

Relevant Websites

GIS websites

http://www.csiss.org/GISPopSci/
Centre for Socially Integrated Social Science. GIS and Population Science.

http://www.amro.who.int/English/sha/be v22n3 SIGEpi.htm
Pan American Health Organization (PAHO): PAHO SIGEpi. Geographic Information System in Epidemiology and Public Health.

http://www.who.int/kms/initiatives/accessmod/en/index.html
WHO Modelling Physical Accessibility to Health Care (AccessMod).

Rural websites

http://www.amref.org/
African Medical and Research Foundation (AMREF) Flying Doctors.

http://www.aihw.gov.au/indigenous/health/access.cfm
Australian Institute of Health and Welfare (AIHW), Indigenous Australians access to health services.

http://www.abdn.ac.uk/crh/about.shtml
Centre for Rural Health, University of Aberdeen.

http://www.ruralhealthresearch.org/
Rural Health Research Gateway.

Hegemony

J. Glassman, University of British Columbia, Vancouver, BC, Canada

Glossary

Articulation A term that refers simultaneously to connections between two entities or processes and to the ability of one entity or process to give expression to another.

Authoritarian Populism A political project that has evolved in a context where there is intensive state control over all spheres of socioeconomic life, in which the populist dimension of the political project deflects criticism from major forms of class power and redirects them to an abstracted notion of the state itself.

Geopolitical Realism An approach to international relations that regards states, for analytical purposes, as individual actors that pursue their own interests (typically wealth and power), within an anarchic global arena.

National-Popular It is a term referring to processes and social struggles that transcend a simple working-class base, and involve alliances between a variety of social groups, organized on a national basis, and typically involving peasants and other agrarian groups.

Neoliberalism It is a doctrine, based loosely on an updating of classic nineteenth-century liberalism and emerging in the 1970s and 1980s, according to which the best political and economic outcomes can be procured through minimizing the role of states and allowing market activities considerable scope, including through encouragement of global free trade.

Organic Intellectuals In the writings of Antonio Gramsci, a term referring to specific groups of intellectuals whose activities are associated with the interests of a particular class or social group, and who help that class or social group to form a stronger sense of cohesion, group identity, and strategic purpose.

World-Systems Analysis A form of analysis developed by historical sociologist Immanuel Wallerstein in the 1970s, built around the claim that contemporary societies can only be understood if they are analyzed as part of a global capitalist system, one that has developed through very long-term historical processes, and is inherently hierarchical.

Introduction

Hegemony is a word with a long history, deriving from an ancient Greek term referring to domination or leadership, especially the leadership of a state or nation in a confederation. In this usage, the degree to which hegemony implies leadership by coercion, consent, or some combination of the two is most often left unspecified. In geopolitical literature up to the twentieth century, where the term hegemony occurred, it was typically used to refer to relations between states, in this underspecified sense, and in some contemporary discussions similar notions of hegemony are still common. The concept of hegemony was substantially transformed, however, by the writings of Antonio Gramsci, becoming a term used to refer to features of class relations and to specify a particular relationship between domination and leadership. This article examines Gramsci's adaptation of hegemony, surveys varied interpretations of Gramsci in both social theory and international relations/international political economy, then notes specific geographical themes related to Gramscian notions of hegemony, and concludes with a discussion of the variety of both Gramscian and non Gramsican conceptions of hegemony, noting the significance of the concept in contemporary political debates.

Gramsci's Adaptation of the Notion of Hegemony

With the publication in 1971 of the first English language translation of extensive selections from Gramsci's *Prison Notebooks*, less differentiated uses of the term hegemony were increasingly displaced in Marxist discourse by uses associated with this text. Gramsci, the leader of the Italian Communist Party in the 1920s before he was imprisoned by Mussolini's regime, wrote prolifically from his prison cell about the development of political struggles in Italy, including the rise of fascism. In elaborating an account of the limits and failures of working class struggles in Italy, Gramsci revised and redeployed the notion of hegemony, developing it as a concept for analysis of class relations within capitalist societies of the West.

Gramsci did not simply invent this conception of hegemony from whole cloth or effect by himself the transformation of the concept into one centered on class. Much of the groundwork for the conceptualization Gramsci developed had been laid by previous debates within the Russian Left. However, Gramsci deployed hegemony in a comparatively unique fashion, using it to analyze the relationship between capitalists and workers within a developed capitalist society. Precisely what Gramsci meant in his various uses of the term has been a topic of considerable debate, at least since the publication of the 1971 English translation of excerpts from the

Prison Notebooks. Nonetheless, virtually all commentators agree that Gramsci gave special impetus to certain aspects of the conception of hegemony, including, in particular, a focus on political and cultural dimensions of class rule and revolutionary struggle.

Class and International Relations in Gramsci

Gramsci's use of the term hegemony as a concept centered on class rather than on relations between states followed directly from his conception of social power and the basis of international relations. In 'The modern prince', one of the writings included within the 1971 excerpts from the *Prison Notebooks*, he asked whether international relations logically precede or follow from domestic social relations, and answered that they follow. Changes in the social structure, expressed militarily, modify international relations, including the geographical positions of national states.

Gramsci did allow that international relations reacted back on domestic social structures and political relations, both passively and actively, but he was clearly articulating a position antagonistic to the geopolitical realist construct of national states as unified individual actors in an anarchic global arena. In contrast, Gramsci derived the international behavior of states, at least to a significant degree, from the social relations and struggles within those states – conceived, it should be added, in largely national–territorial terms. This allowed Gramsci to maintain a focus on class relations within societies – predominantly though not exclusively, in his analyses, Italy – and, thus, on the political struggles in which he was a direct participant.

Gramsci's Nonreductionist Approach to Class

While Gramsci thus reasserted the primacy of class relations, he did not do so in a way that was economically reductionist. As for many Marxist social theorists, class for Gramsci was simultaneously an economic, political, and sociocultural phenomenon. While this view was not unique to Gramsci, the strength of his writings lies in the ways they draw out the political and sociocultural (including ideological) dimensions of class relations.

Here, the concept of hegemony is central. In Gramsci's writings, hegemony refers to a specific relationship between domination and leadership, and coercion and consent. Gramsci differentiated between these terms, and identified hegemony more with leadership/consent than with domination/coercion. Domination and leadership do not, however, exist in a mutually exclusive relationship. As Gramsci put it in his 'Notes on Italian history', another text included in the 1971 collection, a social group dominates antagonistic groups, eliminating them or subjugating them by armed force if necessary, while it leads kindred and allied groups. Hegemony is central to leadership, in that it refers to the process by which varied social groups consent to the leading role played by a particular class, class fraction, or political group. This consent is not based narrowly on ideological power or political persuasion, but rather on the broader powers which enable certain groups to lead. As Gramsci puts it in his notes on 'The intellectuals', in the 1971 collection, the apparently spontaneous consent given by the majority of the population to the trajectories imposed on social life by the dominant group is generated by the prestige this group enjoys because of its power within the world of production. At the same time, as Gramsci noted, the coercive power of the state is still crucial, and imposes discipline on those who do not consent, either actively or passively, to the dominant group.

As such, just as with domination and leadership, consent and coercion do not exist for Gramsci in a mutually exclusive relationship. Instead, as he put it in the 'Notes on Italian history', the 'normal' exercise of hegemony within parliamentary systems is characterized by a combination of force and consent. These tend to balance each other 'reciprocally', without force predominating excessively. Moreover, Gramsci insisted, attempts are always made to ensure that any use of force will appear to be based on the consent of the majority, as articulated in organs of public opinion, such as newspapers.

Hegemony, then, does not refer to a situation in which leadership is maintained solely by consent, but rather to a situation in which both consent and coercion are at work, albeit without force predominating and with the exercise of force made to appear as having popular legitimacy.

This balancing of consent and coercion, for Gramsci, hinges on a combination of economic, political, and sociocultural practices. Gramsci's extended discussion of the role of intellectuals in the construction of hegemony (including via newspapers and the like) pointed in the direction of politics and ideology, but at the same time Gramsci grounded this discussion in an understanding of the relationship of intellectuals to class power and political economic processes.

In discussing intellectuals, Gramsci rarely focused narrowly on academics, but instead showed special interest in what he called the 'organic intellectuals' of a particular social group. As he put it in the writings on 'The intellectuals', every social group, existing on the basis of its relationship to economic production, creates along with itself various kinds of intellectuals, who give the group a greater sense of identity and awareness of its economic, social, and political functions. As an example, Gramsci suggested that capitalist entrepreneurs create alongside themselves such organic intellectuals as industrial technicians, political economists, cultural producers, and legal experts. Since hegemony did not imply, for Gramsci, the absence of opposition, organic intellectuals are not only those who represent the interests of capitalists and ruling groups. Working class parties also develop their own organic intellectuals, and indeed Gramsci went so far as to characterize all members of political parties, whether working class or not, as intellectuals, reflecting the role of parties in consciously organizing and elaborating the struggles of particular groups.

This view corresponded to Gramsci's more general perspective on human intellectual activity. Gramsci saw all humans as being intellectuals, even if not all humans have the social function of working as intellectuals. Intellectual and ideological activity, for Gramsci, is characteristic of all humans, and is intimately bound up with labor and other forms of social activity. In 'The intellectuals', Gramsci put it that the creation of a new stratum of intellectuals involves critical elaboration of the intellectual activity "that exists in everyone at a certain degree of development," while modifying its relationship with human muscular nervous activity. This occurs in such a way as to ensure that muscular nervous activity, which is central to the vital practical efforts through which humans alter the physical and social worlds in which they live, becomes "the foundation of a new and integral conception of the world." The formation of organic intellectuals, and indeed intellectual activity in general, is thus a dimension of the broader social process that involves biophysical activities such as production. From this perspective, Gramsci articulated a notion of hegemony – and the activity of intellectuals in helping generate it – that was not merely cultural or based on intellectual persuasion but based on the relationship of intellectual and cultural activity to specific forms of class power.

The Geography of Hegemony in Gramsci

Though Gramsci was not formally a geographer – indeed, he was not a traditional, institutional intellectual of any sort – there is a very important geography to his arguments. First, Gramsci's arguments were built around an insistence on the specificity – geographical and historical – of situations like those in Italy. Rather than simply adopt existing Marxist conventions, Gramsci adapted and transformed them in relationship to the realities he encountered in the political struggles of early twentieth century Italy. These realities included, crucially, the failure of the Italian working class to launch a sustained, effective, and transformative attack on the Italian capitalist state, and much of his discussion of hegemony is in fact designed to explain how, in such a context, some workers could end up consenting to capitalist leadership.

In addition, there is a more fine tuned geography in Gramsci, one focused on uneven development within Italy and the differing roles of intellectuals in different regions of the country. Gramsci, himself from Sardinia, had begun to understand the potential of working class organizations after moving to Turin for studies. He became keenly aware of the political differences between the more industrialized, working class centers of northern Italy, like Turin, and rural, agricultural regions, particularly in the south – the so called Southern Question. To these differences in forms of political economic development, corresponded differences in the kinds of intellectuals who helped organize hegemonic projects. In the south, and in rural areas more generally, the predominant organic intellectual strata were comprised of actors such as lawyers and priests, while in the north a larger role was played by the factory technician. Much of Gramsci's nuanced analysis of Italian politics and the Southern Question is built around an awareness of these regional differences, the roles of differing kinds of intellectuals and social groups, and the efforts of leading political actors and intellectuals (Bendetto Croce is especially important for Gramsci) to create an effective national leadership in this context. For Gramsci's own political project, this backdrop of uneven development required an adaptation of Marxist theory, so that rather than being based narrowly on working class struggle it could accommodate 'national popular' struggle, inclusive of both working class and peasant politics.

Interpretations and Uses of Gramsci in Social Theory

Gramsci's thought has been of interest to a broad range of social theorists with quite diverse perspectives on his work. The philosophically broad fashion in which Gramsci couched certain of his political arguments has left them open to a wide range of interpretations, including their reassimilation to the major ontological categories of liberal social science, in which politics is constructed as a realm autonomous from economics.

Such a reading has been enabled, to some extent, by the form in which Gramsci's writings emerged. The various tracts that were included in the 1971 translation of selections from the *Prison Notebooks* tended to be those that focus very directly on issues of political strategy and ideology. In 1995, however, a larger collection of writings from Gramsci's notebooks was published in English translation, including a significant number that deal more directly with economic issues. These were not entirely separate writings from those comprising the tracts published in the 1971 selections from the *Prison Notebooks* and would thus call into question any notion that Gramsci was an analyst of politics, culture, or ideology in abstraction from economics.

But even within the 1971 selections from the *Prison Notebooks*, there are numerous passages that would argue against such a reading. For example, in a discussion of political reform and national popular projects in 'The modern prince', Gramsci argues that intellectual and moral reforms have to be linked with a program of economic reforms – indeed, he states, "the program of economic reform is precisely the concrete form in which every intellectual and moral reform presents itself." In an especially pertinent passage from the same text, Gramsci argues that the ideas of the free trade movement are based on a theoretical error that involves turning a purely methodological distinction between political and civil society into an 'organic' (ontological) distinction. Based on this error, the partisans of free trade assert that "economic activity belongs to civil society, and that the State must not intervene to regulate it." But, for Gramsci, civil society and the state are in reality two faces of the same entity, and so *laissez faire* is itself a form of state 'regulation', which is "introduced and maintained by legislative and coercive means." As such, *laissez faire* or free trade "is a deliberate policy, conscious of its own ends, and not the spontaneous, automatic expression of economic facts." It is, moreover, a political program designed to change a state's leading personnel, its economic policies, and outcomes, such as the distribution of the national income. Here again, Gramsci clearly saw economics, politics, and ideology as being of one piece, integrated through class relations.

Yet, even where there has been agreement among Gramscian scholars regarding this general issue, a fairly substantial amount of disagreement has remained regarding more specific implications of the concept of hegemony and other aspects of Gramscian thought. Three debates that developed in the 1980s and 1990s can be used to provide a sense of the range of interpretations of Gramsci and the different ways in which the concept of hegemony has evolved.

Hall's Use of Gramci

In the 1970s and 1980s, Stuart Hall developed a critique of Thatcherism, the neoliberal policies implemented in Great Britain with the rise of Prime Minister Margaret Thatcher in the late 1970s. In this critique, reviewed concisely and further articulated in Hall's 1988 collection, *The Hard Road to Renewal*, Hall borrowed from Gramsci to develop an analysis that could help explain the ability of neoliberals to gain some degree of consent for their policies from various British working class groups. Hall suggested that Thatcherism constituted a variety of 'authoritarian populism' – a term he adopted from Nicos Poulantzas – and that its populist dimension reflected the conscious deployment of anti statist ideology by Thatcherites. This anti statism played to popular frustrations with various aspects of state inertia and trade union privilege to garner popular support for rolling back the Keynesian welfare state. Thatcherism, as Hall put it, was able to stigmatize "statism, bureaucracy, social democracy, and 'creeping collectivism'," counterposing to these such positive traits as "possessive individualism, personal initiative, 'Thatcherism' and freedom." Moreover, Thatcherism developed its populist ideological agenda in part through moralistic appeals to working class insecurity over crime and 'law and order'. As such, it was able to construct a new popular 'common sense' that pulled a certain amount of labor support away from social democracy.

This analysis was contested by other British scholars, including Bob Jessop, Kevin Bonnett, Simon Bromley, and Tom Ling. Writing in *New Left Review*, Jessop, Bonnett, Bromley, and Ling not only challenged a number of Hall's specific contentions regarding Thatcherism, but they also challenged the fashion in which Hall used the concept of hegemony, accusing him of 'ideologism' for the way in which he foregrounded ideological dimensions of Thatcherism. Both Hall and his critics, as became clear in subsequent exchanges, agreed in principle with the Gramscian notion that economics, politics, and ideology are all interconnected phenomena, and that ideological phenomena cannot be analyzed as if they refer to processes playing out in domains separate from economics. Yet, methodologically, Hall championed the legitimacy of foregrounding the ideological dimensions of Thatcherism. He argued that while hegemony is not an exclusively ideological phenomenon, it was important to focus on the ideological dimensions, because, among other reasons, these constituted a distinct level of analysis and had too frequently been neglected by the Left. In reply, Jessop, Bonnett, Bromley, and Ling argued that the abstraction of ideological moments of Thatcherism from issues such as the specific groups of capitalists whose projects enabled and shaped its political and ideological content led to a misleading view of the fashion in which Thatcherism operated. Indeed, Hall's critics did not exempt Gramsci himself from criticism on this score, arguing that Gramsci focused on the politics and ideology of class leadership, while too frequently neglecting structural dimensions of hegemony. Not surprisingly, this was not a debate that led to any simple resolution. It, nonetheless, highlighted some significant issues regarding the ways in which ideological dimensions of hegemony were to be conceived and analyzed, and these issues were to reemerge in several other debates.

Said's Use of Gramsci

A second example of a debate around the concept of hegemony was occasioned by Edward Said's deployment of Gramsci's thought in works such as *Orientalism*, one of the founding texts of contemporary postcolonial studies. Said indicated early in this work his debt to Gramsci, when he said that it was particularly cultural hegemony that gave Orientalism durability and strength. As Said developed this argument, his point was that the cultural dimensions of European colonizing ventures in the Near East, exemplified by the scholarly projects constituting the field of Orientalism, were a crucial dimension of imperialism, functioning in conjunction with – and even enabling – the deployment of military and economic forces. Moreover, Said's project directly invoked a Gramscian strategy for analyzing this hegemony when he noted that in the *Prison Notebooks*, Gramsci called on critical intellectuals to understand themselves as prod ucts of historical processes, which have "deposited in you an infinity of traces, without leaving an inventory." It was imperative at the outset, then, to compile such an in ventory. The special significance of this passage for Said, as he pointed out, related to his own experiences as a Palestinian growing up in the British colonies of Palestine and Egypt, which inspired him to understand Oriental ism as part of understanding his own life. But the passage also pointed toward Said's approach to the interpretation of Orientalism, which focused on compiling an inventory of the traces left in the perceptions of Westerners by the products of Orientalist discourse – specifically, the major literary texts about the Orient which comprise the cen terpiece of Said's analysis.

Rather like Hall, himself a major figure in postcolonial studies, Said was to come under criticism for specific aspects of this methodological approach. Said's general project in *Orientalism*, and his use of Gramsci in par ticular, was criticized by Aijaz Ahmad, in his book, *In Theory*. Ahmad took Said to task for a number of specific aspects of his interpretive approach, many of these cen tering on the fact that Said was fundamentally non Marxist and theoretically eclectic, mixing Gramscian ideas that originally had a Leftist orientation with ideas that were decidedly more nonradical.

But a more general point of his book was to critique postcolonial studies for, in his view, failing to adequately interrogate phenomena such as internal social divisions within colonized societies. In his critique of Said, he makes this point via direct reference to Gramsci, noting that the original Gramscian idea of an inventory of traces presumes that there is a cultural location upon which those traces are inscribed, including 'other' traces, with all of these traces being historically interwoven in com plex fashion. Ahmad argues that Said neglects those traces that have to do with the social structures of postcolonial societies. *Orientalism*, for Ahmad, examines the history of Western textualities about the non West "in isolation from how these textualities might have been received, accepted, modified, challenged, overthrown or reproduced by the intelligentsias of the colonized countries." These intelligentsias, for Ahmad, need to be analyzed as "situated social agents" impelled by their own "conflicts, contradictions, distinct social and political locations, of class, gender, religious affiliation, and so on." Said's failure to do this, according to Ahmad, marks a major lapse in the analysis of hegemony. This criticism of Said, moreover, informed Ahmad's analysis of con temporary Indian politics. For example, in an essay en titled 'Fascism and national culture', in his book *Lineages of the Present*, Ahmad read Gramsci's analysis of the rise of fascism in Italy against his own analysis of the rise of right wing politics in India, using the concept of he gemony to emphasize cultural dimensions of the Right's class projects.

In Ahmad's critique of Said, as in Jessop, Bonnett, and Bromley's critique of Hall, it became clear that part of the issue was the degree to which it is methodologically le gitimate to investigate hegemony by first abstracting its political ideological dimensions and then reassimilating them to their broader socioeconomic context. This was the approach taken by Hall and Said, as well as a number of other scholars in the postcolonial and literary criticism camps. Authors like Jessop and Ahmad, by contrast, seemed to insist on foregrounding political economic dimensions of hegemony, even within the discussion of issues such as ideology. This methodological disagree ment was overdetermined and impinged on factors ran ging from disciplinary politics and personal differences to tactically competing political affiliations.

In the case of Ahmad's criticism of Said, this kind of disagreement also impinged directly on the issue of how one interprets the role of intellectuals in constructing or challenging particular hegemonies. Ahmad was harshly critical not only of Said but also of postcolonial theorists within the West in general, citing their privileged class positions within Western academies as producing sub stantial limits to their own strategies of intervention, which were based predominantly on critical reading and analysis of academic texts. Ahmad's criticism was not without its self contradictory features, given his own one time position as a visiting scholar at Rutgers Uni versity and his seeming intent to at least partly address a Western intellectual audience with his writings. None theless, Ahmad's criticism did raise a significant issue regarding precisely how one reads the role of intel lectuals in the construction and challenging of hegemony, including whether or not contemporary professional scholars should be seen as involved in the sorts of pro jects that Gramsci identified with organic intellectuals.

Laclau and Mouffe's Adaptation of Gramsci

The book which drew the most attention to the post Marxist possibilities of the concept of hegemony was Ernesto Laclau and Chantal Mouffe's *Hegemony and So cialist Strategy*. More than any other such reading of Gramsci, moreover, Laclau and Mouffe's book made clear the contention that hegemony could be used in ways that enabled an analysis of social struggle that moved beyond Gramsci's focus on class politics in the direction of what they term 'radical democracy'.

For Laclau and Mouffe, hegemony referred to the fact that political strategies can never be simply read off of economic realities. Rather, as they argued in surveying the history of early twentieth century Marxist debates, what emerged through the notion of hegemony was a 'logic of the contingent'. Contingency reflected the fact that European working classes turned out to be more fragmented and less revolutionary, in various contexts, than Marxists had originally envisioned them to be. As such, there was no unified proletariat to carry forward the tasks of the revolution. Instead, difficult choices between political reforms, without revolution, and revolutionary struggles based on heterogeneous political assemblages, were the order of the day for early twentieth century Marxists.

Laclau and Mouffe saw Gramsci as having introduced hegemony, an antireductionist concept, in the context of these dilemmas. They argued that Gramsci pushed the logic of contingency, through the concept of hegemony, as far as it can go – within a Marxist framework. Their own project, then, was to push this logic further – namely, in post Marxist directions – by developing a notion of radical democratic politics that no longer centered revolutionary struggle on class. Hegemony, throughout their text, was thus used to refer to a context in which class subjects are no longer accorded any particular privilege as social actors.

Laclau and Mouffe summarized these post Marxist implications by referring to three consequences of following out the logic of contingency and the implications of hegemony. The first of these is that there is no necessary relation "between socialist objectives and the positions of social agents in the relations of production … their articulation must be regarded as a hegemonic relation." The second, discussed directly in relation to the new social movements seen by some as displacing working class movements, is that "there are no privileged points for the unleashing of a socialist political practice." Instead, socialist political practice involves constructing a collective will by organizing a number of dissimilar groups. The third consequence is what Laclau and Mouffe referred to as a 'decentering' of subject positions, so that rather than being seen as already fully constituted and separate, the different struggles posing the possibility of radical democracy are always being constituted through their various forms of articulation. This opens Laclau and Mouffe's analysis in the direction of the discursive processes through which subjectivities are constituted. From this conceptual ground, then, Laclau and Mouffe constructed an account of processes of articulation, an account designed to illustrate the salience of their particular concept of hegemony and radical contingency.

As with the previously mentioned debates – indeed, perhaps more sharply – Laclau and Mouffe's overt call for a radical democratic politics that transcends traditional class struggles generated substantial criticism. Predictably, some of the most heated exchanges were with authors antagonistic to Laclau and Mouffe's use of Marxist terminology for post Marxist purposes. Even commentators less hostile to post Marxism critiqued Laclau and Mouffe, however, for their abandonment of class and their assertion of the radical character of contingency in hegemonic contexts. Geographers Gillian Hart, Alan Pred, and Michael Watts, for example, argued that Laclau and Mouffe produced an impoverished concept of articulation. As Hart put it in her book *Disabling Globalization*, Laclau and Mouffe's conception "abstracts from historically and geographically specific processes, material conditions, and forms of power that set the conditions for articulatory practices." Hart argued, like Pred and Watts, that instead of the retreat from class signaled by Laclau and Mouffe's use of hegemony what is needed is "to retheorize where class has gone to, and to rethink, and reassert it in nonessentialist ways" (Hart, 2002).

In debates around Laclau and Mouffe's work, as in those around Hall's and Said's works, the positions of intellectuals and the relative autonomy of culture – or, in Laclau and Mouffe's case, discourse – were central. In this sense, such debates came to appear for many as variants on debates between postmodernists or post structuralists and their critics. However, Laclau and Mouffe did not claim allegiance to postmodernism, and such generalizations about their debates with their critics likely obscure the more specific issues regarding hegemony that were being raised. What the three interpretations of Gramsci discussed here – along with the criticisms of them – might better be seen as doing is developing a terrain of ideas surrounding political practice in capitalist societies, and the roles of class forces and varied kinds of intellectuals in maintaining a social order based in part on the consent of the led. If there was no agreement about precisely how this worked, there was nonetheless agreement that political–ideological factors – in some sort of contested relation to class forces and economic processes – were crucial. In this general sense, there clearly came to be a Gramscian notion of hegemony.

Cox and Neo-Gramscian International Relations/International Political Economy

At the same time as social theorists were rediscovering Gramsci's writings, Gramscian notions of hegemony also came to play a significant role within international relations (IR) theory and international political economy (IPE). Some of the major works deploying Gramscian conceptions in these fields was carried out by Robert Cox. In a series of journal articles reflecting on the significance of the concept of hegemony within IR and IPE, and in a 1987 book, *Production, Power, and World Order*, applying the concept of hegemony to a broad analysis of global re lations, Cox helped give birth to what has come to be known as neo Gramscian IPE. Cox reinvoked the notion of hegemony as a relation between states, but he both specified this relation in Gramscian terms and also pro vided a Gramscian inflection by building class analysis directly into the account, thus moving beyond geo political realism's methodological approach to states. Cox foregrounded his particular conception of hegemony at the beginning of *Production, Power, and World Order* by noting that the term 'hegemony' means more than the dominance of a single world power. For Cox, hegemony denotes an arrangement where the dominant state "cre ates an order based ideologically on a broad measure of consent," and where this order functions in ways that ensure the continuing supremacy of the leading state or states, as well as the leading social classes within those states, while at the same time offering "some measure or prospect of satisfaction to the less powerful." This con ception of hegemony, then, involves articulating the no tion of class leadership within a given society and the notion of global leadership in a given world order. Cox referred to the former as 'social hegemony' and made it a precondition for the latter form of hegemony. As he put it, "a world hegemonic order can only be founded by a country in which social hegemony has been or is being achieved" (Cox, 1987). In this approach, Cox hewed closely to Gramsci's original conception of class as the basis for hegemony, while building outward from this to international relations. This conforms to Gramsci's own antigeopolitical realist conception of the nature of international relations, cited above, but Cox also engaged fully with international relations rather than leaving them outside the scope of his inquiry.

Cox on Productive Power, World Hegemony, and Historic Blocs

While Cox, like Gramsci, emphasized the role of ideol ogy in consolidating hegemony, he nonetheless built his account of hegemony on productive power. The leading global hegemons, in Cox's account, were those who marshaled the greatest productive capacity, and much of their ability to offer "some measure or prospect of sat isfaction to the less powerful" (Cox, 1987) was grounded in this capacity. Here, Cox emphasized phenomena such as the changes in production processes and social or ganization that underpinned the transition from a nine teenth century period of British hegemony, through an early twentieth century period of imperial rivalry, to a post World War II period of US hegemony. What prevented this from becoming a very conventional geo political realist account of competition between leading states in an anarchic global arena is the way Cox linked this transition to transformations in class relations and production processes.

Cox built this synthesis of class processes and inter national relations by analyzing the role of leading forces in hegemonic states in consolidating 'historic blocs' – Gramsci's term for, as Cox put it in *Production, Power, and World Order*, "the particular configuration of social classes and ideology that gives content to a historical state." These historic blocs are grounded in class relations within societies – and thus in social hegemony – but extend outward to encompass leading groups within states around the world. Crucially, Cox's account of this process emphasized the internally variegated character of states, rather than seeing them as unified actors. This allowed for particular classes and class fractions to be active within the state, pursuing hegemonic projects, without the state being reduced to a terrain of unified, classless, national interests.

This kind of historic bloc, for Cox, made struggles around hegemony inherently transnational, even if global hegemony had to be secured in the first instance through social hegemony in the globally hegemonic state. In his discussion of post World War II US hegemony, for ex ample, Cox noted how this hegemony developed in the international arena through a process of consensus for mation in the leading states. In this process, the central agencies of these states, such as prime ministers offices, treasuries, and central banks, interact with each other, sometimes through international financial institutions, sometimes through various kinds of multilateral forums, and sometimes in a complex of bilateral relationships. Cox argued that, within this process, agencies of the US state have a dominant but not necessarily determining role.

Cox on First and Third World Relations and the Role of Intellectuals

This kind of hegemonic consensus building made clear the terrain on which organic intellectuals and ideology have power, for Cox. The role of ideology and organic intellectuals was spelled out even more directly, in re lation to First World–Third World relations within the historic bloc. As Cox saw it, the 'top level countries' jointly limit the developmental options of the late in dustrializing countries. Elites from Third World coun tries participate in the historic bloc, but they "do not participate with the same effective status as top level elites in the formation of the consensus." Nonetheless, this consensus gains ideological support within the Third World, and the process leads to 'ideologically con ditioned agents' being placed in key positions within

Third World states. International finance, for example, flows into Third World countries via institutions (e.g., central banks and finance ministries) that are staffed "by people who have been socialized to the norms of the consensus and of its professional cadres." Such Third World participants in the historic bloc are often "graduates of major advanced capitalist country universities," and many have done stints at places like the IMF institute, which "bring Third World technical financial personnel into personal contact with the milieu of international finance" (Cox, 1987).

Even while emphasizing consensus building, however, Cox pursued a very Gramscian line in articulating the relationship between consensus and coercion within the historic bloc. In line with Gramsci's claim that a ruling group dominates antagonistic groups while exercising hegemony vis à vis its allies, Cox argued that the transformation in Third World states required as part of their integration into the historic bloc "is more openly induced by external pressures" than the internationalizing of advanced capitalist states. It thus generates more resentment, so that organizations like the IMF have come to be seen as political enemies by nationalist and popular forces in the Third World. As a consequence, relations between First and Third World countries remain marked by significant levels of coercion, as Cox noted throughout his study.

Themes in Neo-Gramscian IR/IPE

Cox's writings became part of a broad tendency within IR and IPE known as neo Gramscian. As with Gramscian social and political theory, there is no one, singular Gramscian line of analysis that has unequivocally emerged in neo Gramscian IPE. Here too, though debates have seemingly been less heated, there have developed a variety of lines that build from a broadly shared Gramscian core. Two projects in neo Gramscian IPE can be cited to indicate the directions that discussions of hegemony have taken.

Rupert on Popular 'Common Sense'

Mark Rupert, in his book *Producing Hegemony*, followed Cox's lead in emphasizing the importance of productive power to hegemony. Rupert developed a very specific account of the social hegemony of US capitalists in the post World War II period, deploying a Gramscian conception of popular 'common sense' to explain how these capitalists were able to exercise leadership in relation to major elements of the US working class.

The productive power of US capitalism was such that it enabled substantial material gains for US workers, under the condition of limited trade union power. The US workers, support for the projects of leading US capitalists, for Rupert, was consolidated – on these foundations – through the development of a conception of what it was to be an American, a conception that linked this very tightly to worker prosperity and participation in consumerism. In a sense, the organic intellectuals who helped to forge this popular common sense 'Americanism' included conservative union leaders themselves, though Rupert does not use this terminology to describe them. Their success in this regard helped to produce a US working class – or at least a segment of that class, comprised especially of white, male, industrial workers – for whom Americanism came to be associated with support for a capitalist system that enabled substantial consumption possibilities at home while being aggressively anticommunist abroad. In this sense, the popular common sense 'Americanism' pulled US workers into a historic bloc through which leading US elites combated communism throughout the Cold War period.

Rupert carried this sort of Gramscian analysis further in a subsequent book, *Ideologies of Globalization*, which addressed the different forms taken by antiglobalization politics, particularly within the United States. Recognizing that antiglobalization politics are not necessarily progressive, Rupert interrogated how in specific contexts the kind of popular common sense Americanism that prevailed in the United States during the Cold War era has evolved into an antagonism toward neoliberal globalization that is sometimes racist and xenophobic. Some of the foundations of this phenomenon have to do with the basic material insecurities of workers, who are facing downward mobility as a result of new global economic forces. But the phenomenon also expresses the conscious construction of a right wing nationalist ideology by forces intent on politically tapping this discontent. Rupert's study thus showed how shifting global patterns of productive power have been politically and ideologically worked through by concrete social actors, within a specific national context.

Taylor on Prime Modernities

Peter J. Taylor, in his books *The Way the Modern World Works and Modernities*, developed some of the ideas of neo Gramscian IPE within a world systems framework. World systems approaches, in the work of Taylor, Immanuel Wallerstein, and others, conceived the global political economy to be comprised of global commodity chains and national, territorial states. Class relations are central to the functioning of both commodity chains and states, but in Taylor's analyses of states he focused especially on interstate dynamics. As such, he deployed the notion of hegemony in a rather classic sense, but this did not devolve into a pre Gramscian usage insofar as he foregrounded the Gramscian claim that hegemonic relations are based more on consensus than on coercion, while still allowing, as per Cox's argument, that relations between core and peripheral countries are mediated to a greater extent by force.

Taylor discussed world hegemony specifically in relation to the global dominance of Great Britain and the United States in the nineteenth and twentieth centuries, respectively. In *The Way the Modern World Works*, he argued that both Britain and the United States can be seen as having provided "intellectual and moral leadership," in a Gramsican sense, "based upon their respective economic, political, and cultural preeminence." He argued, moreover, that the concept of hegemony could easily be translated from Gramsci's usage into one appropriate for discussions of international relations by regarding states as subjects and power within the wider world as their objective, rather than focusing on social classes and their attempts to control power within states. Whether or not the translation is indeed quite so straightforward could be contested, given that from the kind of perspective developed by Cox and others opposed to geopolitical realism, states should not be seen as unified actors. Nonetheless, Taylor did deploy a basically Gramscian notion of the exercise of hegemony through ideological and cultural projects.

Taylor elaborated this notion of cultural hegemony in *Modernities*, where he argued that three hegemonic powers, the seventeenth century Dutch state, the nineteenth century British state, and the twentieth century US state, each produced some of the consensus for the leadership that they achieved through the spread of ideas about modernity. Taylor identified three specific ideas with these three hegemonies: mercantilism with the Dutch, industrialism with the British, and consumerism with the United States, and called these 'prime modernities'. Taylor argued that the hegemonic states achieved world leadership during the periods in question in part because their power and prosperity led others to try to emulate them.

Geographies of Hegemony

The development of Gramscian and neo Gramscian conceptions of hegemony has led to the development of varied senses of the spatiality of hegemony, ranging from those focused on uneven national development and the particularity of specific national contexts to those focused on global power relations. These varied geographies of hegemony have been drawn out explicitly by a number of Gramscian and neo Gramscian scholars. Two examples can be used to illustrate some of the spatial complexity this has lent to contemporary understandings of hegemony.

Hart on Globalization

In her study of post apartheid South Africa, *Disabling Globalization*, Hart discussed geographies of hegemony in a fashion that draws on Gramsci, Hall, Henri Lefebvre, and others, to emphasize interconnections between processes at scales ranging from the local to the global. She discussed these interconnections as hegemonic articulations and drew them out in complex and creative ways, including examining both the roles of Taiwanese capitalists in South African rural industrialization and the significance in this context of South Africa and East Asia's very different histories of agrarian transformation.

Hart's objective in analyzing these articulations was to produce a middle ground between the radical contingency of hegemony suggested by Laclau and Mouffe, and accounts of social power that, in contrast, seem to allow little space for local agency. Toward this end, she discussed specific histories of hegemony in different local contexts as being central to an account of the differing political trajectories of two neighboring South African localities. At the same time, she showed how these local histories are linked to and productive of broader realities, including those referred to under the headings of 'globalization' and 'neoliberalism'. This sort of analysis was meant to disrupt globalization narratives that see neoliberal globalization as being constructed over the heads of local actors and developing in such a way as to compel their acquiescence. As such, it suggested local activities that might alter or reconstruct certain features of globalization and enable the production of different hegemonies.

Agnew on Globalization

John Agnew used a Gramscian notion of hegemony in a broad attack on what he called the 'territorial trap' – the tendency of social scientists and policymakers to think of political economic processes as if they take place within territorial containers (states) and only act on each other externally through international relations. Neoliberal globalization, Agnew pointed out, has eroded the conception of states as political economic containers. In his book *Hegemony*, Agnew argued, however, that neoliberal globalization has also eroded the position of the US state as a leading hegemon, creating complexities and instabilities in the global system.

Agnew suggested that as the hegemonic power most responsible for contemporary globalization the United States has given a particular form to globalization – that of a marketplace society in which consumerism is a central ideal. Yet, the very success of this project has depended upon the development of broad, deterritorialized, global networks into which people around the world are enrolled, and as they have become so enrolled the shared ideals of the marketplace society have taken on a global life that transcends control by the US state. Thus, the global political economy is increasingly globalized without there existing any center of imperial control – US government attempts to exercise such control notwithstanding.

At the same time, Agnew argued that while US hegemonic decline might have been expected to be associated with increasing international disorder, this has not necessarily occurred. As he put it in his book *Geo politics*, "a transnational liberalism sponsored by governing elites around the world seems to offer an emerging ideological hegemony without a single hegemon (or dominant state)" (Angew, 1998). Rather, this hegemony is based to a greater extent on consensus about principles of free trade and comparative advantage, including the sense that they will work to the benefit of global transnational economic elites. The existence of this hegemony, Agnew argued, "calls into question many of the established tenets of the modern geopolitical imagination," (Angew, 1998), but this does not signal the "end of geopolitics." Instead, it signals an increased focus on other transnationally interconnected entities such as firms and cities, to which states must now cede more power and control. In short, the hegemony of global neoliberalism is an ideological phenomenon in a truly Gramscian sense, but one in which the leading classes are no longer territorially bound in the fashion of the Italian leadership that Gramsci analyzed.

A Family of Gramscian Uses?

In spite of the varied uses to which Gramscian notions of hegemony have been put, a certain family resemblance has come into view with most such uses, whether those of Marxist, post Marxist, or non Marxist scholars drawing on Gramsci. These resemblances center around the importance of consent and leadership, as opposed to mere reliance on coercion and domination, and on the fact that such consent is seen as established via cultural, political, and ideological practices, even insofar as it is seen to be grounded in political economic processes. Disagreement has not centered on these basics but on issues such as the degree to which class forces need to be foregrounded in discussing hegemony (whether as a methodological or a theoretical matter), the degree to which ideology and culture can be usefully abstracted from political economic forces as a site of hegemonic practice, and the precise roles of intellectuals in constructing or deconstructing hegemony (including the kinds of intellectuals whose activities are emphasized). There has also evolved a creative tension regarding the different geographical scales at which hegemony can be seen to operate and the degree to which there exists an overarching global hegemony of neoliberalism.

Non-Marxist Conceptions of Hegemony

Gramscian conceptions of hegemony became preeminent in much academic discourse by the end of the twentieth century, but pre and non Gramscian uses of the term, that largely fail to differentiate dominance and leadership, and consent and coercion, have remained prevalent outside of academic debates, as well as within some fields of academic inquiry. This has especially been the case in geopolitical discourse, whether in mainstream IR theory or in public policy discussions. Hegemony in such analyses has largely been conflated with dominance, and this has been the case whether the analyses in question are largely critical or largely supportive of such dominance. Thus, for example, Noam Chomsky's critique of Bush II administration's foreign policy was entitled *Hegemony or Survival: America's Quest for Global Dominance*. The terms hegemony and dominance are not discussed in the book, and it is clear that the two are equated in Chomsky's broad critique of US imperialism.

Similarly, but from a conservative geopolitical realist perspective, Zbigniew Brzezinski's reflections on the US foreign policy issues in the post Cold War era, in his book *The Grand Chessboard*, addressed what he called, in the first chapter of the book, "hegemony of a new type." Brzezinski did in fact discuss different aspects of the US leadership, including what he called its 'cultural appeal' and the 'imitation of American ways', thus calling attention to some of the phenomena highlighted by Taylor. But in referring to this as hegemony of a new type, Brzezinski contrasted it with other hegemonies – those of Rome, China, Great Britain, and so on – that did not all have the combined military, technological, political, economic, and cultural dimensions of leadership characteristic of the United States. Thus, hegemony was still for Brzezinski a more generic term, that may or may not include cultural leadership of the sort he saw the United States as exercising.

Such uses of hegemony are perfectly intelligible on their own terms, and are likely to remain part of the theoretical landscape, if for no other reason than the fact that a number of non Marxist analysts are likely to bridle at the prospect of using Gramscian terminology to describe contemporary political situations. This practice has certain costs, however. For example, with the resurgent US militarism of the Bush II regime, many analysts of different stripes began to more intensively discuss empire, US imperialism, US militarism, and US hegemony. These terms have often been run together rather indiscriminately, as Agnew noted in his book *Hegemony*. Such imprecision has the consequence that it becomes more difficult to consider important matters such as whether the increased US militarism and even global military dominance could actually be exercised in ways that lead to declining US hegemony, as Agnew suggested is the case. The fact that this matter itself is an important topic of analysis suggests that Gramscian notions of hegemony are likely to show increased utility, even if they themselves remain in part a terrain of interpretive struggle.

See also: Ideology; Taylor, P.

Further Reading

Agnew, J. (1998). *Geopolitics: Re visioning World Politics*. London: Routledge.
Agnew, J. (2005). *Hegemony: The New Shape of Global Power.* Philadelphia, PA: Temple University Press.
Ahmad, A. (1991). *In Theory: Classes, Nations, Literatures*. London: Verso.
Ahmad, A. (2000). *Lineages of the Present: Ideology and Politics in Contemporary South Asia*. London: Verso.
Anderson, P. (1977). The antinomies of Antonio Gramsci. *New Left Review* I/100, 5 78.
Boggs, C. (1984). *The Two Revolutions: Antonio Gramsci and the Dilemmas of Western Marxism*. Boston, MA: South End Press.
Brzezinski, Z. (1997). *The Grand Chessboard: American Primacy and Its Geostrategic Imperatives*. New York: Basic Books.
Chomsky, N. (2003). *Hegemony or Survival: America's Quest for Global Dominance*. New York: Metropolitan Books.
Cox, R. (1983). Gramsci, hegemony, and international relations: An essay in method. *Millenium: Journal of International Studies* 12, 162 175.
Cox, R. (1987). *Production, Power, and World Order: Social Forces in the Making of History.* New York: Columbia University Press.
Geras, N. (1987). Post Marxism? *New Left Review* I/163, 40 82.
Gill, S. (1993). *Gramsci, Historical Materialism and International Relations*. Cambridge: Cambridge University Press.
Gramsci, A. (1971). In House, Q. & Smith, G. N. (eds. and transl.). *Selections from the Prison Notebook.* New York: International Publishers.
Gramsci, A. (1995). In Boothman, D. (ed. and transl.). *Further Selections from the Prison Notebooks.* Minneapolis, MN: University of Minnesota Press.
Hall, S. (1988). *The Hard Road to Renewal: Thatcherism and the Crisis of the Left*. London: Verso.
Hart, G. (2002). *Disabling Globalization: Places of Power in Post Apartheid South Africa*. Berekely: University of California Press.
Jessop, B., Bonnett, K., Bromley, S. and Ling, T. (1984). Authoritarian populism, two nations, and Thatcherism. *New Left Review* I/147, 32 60.
Jessop, B., Bonnett, K., Bromley, S. and Ling, T. (1985). Thatche rism and the politics of hegemony. *New Left Review* I/153, 87 101.
Laclau, E. and Mouffe, C. (1985). *Hegemony and Socialist Strategy: Towards a Radical Democratic Politics*. London: Verso.
Laclau, E. and Mouffe, C. (1987). Post Marxism without apologies. *New Left Review* I/166, 79 106.
Rupert, M. (1995). *Producing Hegemony: The Politics of Mass Production and American Global Power*. Cambridge: Cambridge University Press.
Rupert, M. (2000). *Ideologies of Globalization*. New York: Routledge.
Said, E. (1978). *Orientalism*. New York: Random House.
Taylor, P. J. (1996). *The Way the Modern World Works: World Hegemony to World Impasse*. Chichester: Wiley.
Taylor, P. J. (1999). *Modernities: A Geo Historical Interpretation*. Minneapolis, MN: University of Minnesota Press.

Relevant Websites

http://www.marxists.org
Antonio Gramsci page, Marxist Internet Archive.
http://www.internationalgramscisociety.org
International Gramsci Society web site.

Heritage

C. Kelly, University of Greenwich, London, UK

Glossary

Authenticity The extent to which a heritage item, artifact, or event is seen to be valid, genuine, real, or legitimate by an individual or an organization.

Commodification The manner in which heritage raw-materials are utilized by other industries, for example tourism, for economic purposes.

Contestation Negotiation, often between groups of people in local, national, or international spaces, in search of legitimate versions of authentic heritage.

Governance The state structures at local, national, and/or international levels that organize, legislate, or fund heritage-related activities.

Heritage Buildings, artifacts, landscapes, cultures, and traditions deemed to be of importance to past, present, and future populations.

Intangible That which cannot be physically touched; in a heritage context, this often refers to a cultural tradition such as a song, a language-dialect, or superstition particular to a local area.

Introduction

Heritage is an elusive concept, to both define and discuss, as it means many things to most people. Indeed, in basic chronological terms, it is often exclusively regarded as an idea that is set in the past. This notion has been challenged in recent years as various economic, environmental, and sociocultural activities have become concerned with heritage in a contemporary context. Many national and international organizations have incorporated the term heritage into departmental titles and functions. The scope of heritage is broad insofar as it may singularly or simultaneously be taken to refer to aspects of people and culture, the built environment, or natural landscapes. In this respect, it can therefore be manipulated by power structures in society to create subjective or partial narratives of the past (or present) for particular cultural, economic, environmental, and/or political purposes.

Defining Heritage

Semantically speaking, the word 'heritage' has existed for centuries, with most dictionaries unanimously defining it as something inherited at birth, handed down, or a legacy of some kind which links the past, through the present, to the future, at an individual or collective level. This legacy involves people, events, cultural landscapes, and objects deemed to be important at local, national, and/or international geographical scales. There is a multitude of ideas, both tangible and intangible, that the word has been applied to within the cultural sector and outside of it. European Architectural Heritage Year, 1975, is regarded by some as a benchmark for the inception of 'heritage' as a public forum concept. Thus, 'heritage' was a term that came to the fore in the 1970s, in Europe, and throughout the 1980s expanded increasingly to encompass other developments such as the establishment of heritage visitor centers by both the public and private sectors.

Amalgamations of definitional discussions from the literature suggest that heritage includes (but not exclusively): buildings, archaeology, cultural artifacts, natural landscapes and landforms, wildlife, historical events or characters, cultural traditions, dress, language, and customs.

Arguably, an object or event can become heritage because a person or organization deems it to be of importance, and therein lies the crux of the situation – heritage can be anything that is valued – from a family photograph to a national monument. Therefore, if individuals construct their own framework within which to define heritage, seeing it as a plural concept is of little value or consequence. Conversely, meanings can be deliberately constructed by the state, or economically driven organizations who select aspects of heritage for mass consumption, which can play a role in identity construction, the tourism industry, or creating a collective sense of belonging.

The United Nations Educational, Scientific, and Cultural Organization (UNESCO) also categorizes heritage into three main subdivisions: built heritage, natural heritage, and intangible heritage, and these groupings affect the way in which global heritage sites are designated and conserved at an international scale (see section on 'Policy and Governance' later in this article). The United Kingdom, for example, has specially designated 'heritage coasts', representing valued natural landscapes. The European Union has also designated 'natural heritage areas' to protect special natural environments in certain member states. The United States designates 'national heritage areas' through its national parks service, representing both built and natural heritage in its aim to 'tell America's story'. As shown in **Table 1** above, UNESCO's

Table 1 Classifying heritage

	'Built' heritage	*'Natural' heritage*	*'Living' heritage*
Nation/Region	Buildings/cities	Landscapes/environment	Communities/people
Area	Historic towns	National parks	Traditional food and costumes
	Seaside resorts	Coastlines	Festivals
	Conservation areas	Designated/protected landscapes	Markets
Site/Activities	Museums Art galleries	Geological features	Folklore, folklife houses/cafes
	Historic buildings	Town/country parks	Craft centers
	Historic sites/monuments	Botanic gardens	Farms
	Heritage centers	Historic gardens	Ecomuseums
	Heritage theme parks	Nature reserves	Everyday spaces
	Archaeological sites Transport	Countryside centers	Communities
		Country parks	Historic events and characters
		Forest	Language
		Ecomuseums	Song and dance
		Archaeological sites	Religious practices/beliefs

groupings reflect a wide variety of spaces, objects, and activities at various scales. Intangible heritage – that which cannot be touched – interestingly, was only added recently in response to criticisms that 'real people' felt more affinity to everyday aspects of heritage such as language, customs, oral histories, and so on than to grand stately homes or ancient monuments, for example. For many urban dwellers, there is often little opportunity to engage with natural heritage on any kind of a regular basis, which can lead to a disassociation with or disregard for it.

Policy and Governance

Heritage can take a variety of official (state validated) and unofficial forms; the latter often conflicts with, or is resistant to, the former. To consider this further, we can reiterate that 'heritage' is an enormously diverse concept – there are multiple heritages, the contents and meanings of which change through time and across space. Interestingly, heritage policy has mainly focused on the built heritage in many nations, followed to a lesser extent by natural heritage (usually the domain of separate, environmental government agencies), and lastly, intangible heritage. This emphasis is accounted for through the obvious visibility of the built environment, consisting of townscapes, buildings, monuments (see **Table 1**), and other man made infrastructures, which is easier to identify, quantify, and therefore monitor. Many countries have policies and organizations that are concerned with creating inventories of buildings of particular types of architecture, quantifying and classifying ancient monuments and archaeological ruins, and managing or conserving such heritage. At an international level, the International Council on Monuments and Sites (ICOMOS, a subcommittee of UNESCO) collates and monitors national registers. Effects of the tangibility of this type of heritage leads to greater attention in policy terms, and therefore, often also in funding terms.

Landscape based natural heritage holds a fond place in the hearts and imaginations of many people. Environmental legislation usually serves to protect and manage nonurban spaces, and global, European, and national regulations are in place to enforce a plethora of policies. Global Earth Summits which discuss international aspects of the natural environment have generated extreme responses as the planet faces a variety of environmental threats. Our global heritage could therefore be considered at a very basic level, in terms of natural heritage and all that it entails. Values related to the natural are also important at an individual level, where a person's sense of identity or heritage can be heightened as they gaze at the Grand Canyon, the white cliffs of Dover, the lavender fields of Provence, Uluru, or other iconic landscapes. Policy provision for this is, however, much more difficult to formulate. Such cultural landscapes have long been the focus of geographers who attempt to discover the uniqueness of relationships between people and their environments.

UNESCO has now recognized the importance of intangible heritage, such as endangered languages and oral heritage; placing as much importance and protection on these aspects as the tangible heritage they also safeguard. Thus, values of cultural expressions and practice and the recognition of communities/groups as those who identify, enact, re create, and transmit the intangible or living heritage are also part of UNESCO's policy remit. At national and local levels, many groups are engaged in cultural forms which can be classified as intangible heritage. The difficulty in creating policies and funding structures around these, lie in the complexity of deciding how to place value and merit on cultural, essentially qualitative issues? Tackling the question of how to fund local heritage sites and events, it can be argued that governments cannot avoid making value judgements when allocating funding but the requirement to be accessible to every possible community group has meant that many institutions are now facing enormous difficulties. One impact has been that many community organizations, for example, follow funding led heritage agendas and themes, for financial survival, which results in a heritage dictated by state policy structures rather than being directed by the people to whom, arguably, it matters most.

Many heritage organizations emphasize the need to measure cultural value, often to justify state funding mechanisms. However, many do not have any 'robust indicators' to measure the benefits of heritage to society or how people ascribe values to their local heritage in return. This may be because such a process requires in depth qualitative research – something that many gov ernment agencies have shied away from in favor of quantitative or 'hard' data, which is immensely easier to collect and analyze. Organizations such as the American Museums Association devise performance indicators, for their members, which allow the progress of such formal heritage spaces to be monitored. Festivals and other transitory cultural events gauge visitor numbers as measures of success. A major challenge for the heritage sector is to find ways in which to equally account for, manage, and value built, natural, and intangible heritage.

Commodification

Commodification is a term used to describe how heritage is utilized as a commercial product and mass produced, often at the cost of authenticity. If we see heritage in terms of a spectrum – at one end heritage has its own intrinsic cultural value, where it matters to individuals, communities, and nations, in its own right, or for its own sake; at the other end, heritage can be seen to have a utilitarian or economic use value. In other words, it can be manipulated and selected somehow as a 'product' which can be presented, interpreted, and/or sold. The main sector that engages somehow with this commodi fication is the tourism industry (although some reference has been made in the literature to the 'heritage industry' itself, referring to the widest possible scope of agencies and activities involved in commodification). The ex ploitation of particular resources for the creation of heritage products inevitably affects the nature of these resources, leading to fears that they will be subject to damage, distortion, or just depletion.

There are, however, both positive and negative aspects to tourism's engagement with heritage.

Heritage Tourism

It could be argued that there are few better ways than tourism to promote understanding between peoples – by inspiring admiration for shared natural and cultural heritages. Heritage sites have always attracted travellers. Masterpieces of humankind and nature prompt in us a sense of wonder that is in itself a unique form of trans portation. Heritage tourism attractions are diverse, re flecting earlier discussions on definitional classifications (see **Table 2**).

Heritage assets allow for national and local areas to capitalize on tourism development and this can be enormously important in places where other forms of economic activity are scarce. However, uncontrolled tourism and ill planned development can cause ir reversible physical and social damage not only to heri tage sites but also to the communities surrounding them. The physical impacts of tourism on natural and built heritage landscapes are the easiest to see, and therefore, to manage. Assessing and monitoring the way in which tourism affects intangible heritage and culture is rather more complex. Concepts of sustainability are important in the management of tourism, since heritage itself forms the very resource upon which the sector depends. The market does not always care for the finer nuances of cultural integrity, requiring state interventionism for conservation purposes.

Heritage Interpretation and Representation

Given the breadth of meaning and conceptualization involved in the term 'heritage', the task of bringing heritage to an audience, be it domestic or overseas, is enormously challenging. Not all heritage, not all stories, and not all perspectives can be adequately displayed or told. Where heritage is engaged with in public spaces, such as tourism sites, heritage centers, museums, and galleries, a process is undertaken by those charged with it, to select, interpret, and present that heritage. This pro cess involves a commodification of the past, of objects and events, in such a way that interest is raised, and in some cases, entertainment provided. In some instances, that 'interest' has profit driven objectives (e.g., the economic viability of a local heritage site), and as such, must ensure that visitors are engaged and satisfied. In this very pro cess, heritage becomes a product which can be used by consumers/visitors, and therefore, conveys particular messages. These messages come from the deliberate se lection of resources, products, and packaging. These, in turn, are chosen on the basis of sets of subjective values (consciously or not) of those exercising these choices and the functional imperative for the selection (economic, educational, entertainment).

The ways in which heritage objects are selected, put together, and written or spoken about have particular effects. These effects are not those of the objects *per se*; it is the use made of these objects and interpretive frame works that can open up or close down historical, social,

Table 2 Heritage tourism attractions

Heritage tourism category	*Examples and sample sites*
Built-heritage attractions	Historic townscapes, architecture, archaeological sites, monuments, historic buildings *Sites:* Dubrovnik, Croatia; The Winter Palace, St. Petersburgh; Stonehenge, England
Natural-heritage attractions	National parks, cultural landscapes, coastlines, caves, geological features *Sites:* Kakadu National Park, Australia; Mount Fujiama, Japan; the Grand Canyon, USA; the Amazon Rainforest, Brazil; Okavango Delta, Botswana
Cultural-heritage attractions	Arts, crafts, festivals, traditional events, folk-history museums *Sites:* Carnival, Rio de Janeiro; Cannes Film Festival, France; Chinese New Year Festivals; Robben Island, South Africa
Industrial heritage attractions	Mines, factories, breweries, shipyards, *Sites*: Ironbridge Gorge, England; New Lanark, Scotland; Rhine Valley, Germany
Religious sites and attractions	Cathedrals, abbeys, mosques, shrines, pilgrimage routes, and cities *Sites:* Taj Mahal, India; Angkor Wat, Cambodia; Sistene Chapel, the Vatican; Uluru, Australia; Santiago De Compostela, Spain
Military heritage attractions	Castles, battlefields, concentration camps, military museums *Sites:* Normandy, France; Hiroshima, Japan; Auschwitz, Poland
Literary or artistic heritage attractions	Houses, gardens, or landscapes associated with artists and writers *Sites:* Wordsworth: The Lake District, England; Dickens: London; Van Gogh: French Landscapes, African Cave Art; Aboriginal Nature Art: Australia

and cultural possibilities. The importance of insightful and inclusive representation–interpretation that acknowledges such legitimization of difference is crucial in spaces of contested identity and multiple heritage(s). There are many such spaces in an increasing number of global zones of conflict where concepts of 'heritage and national cultural identity' have become fragmented, blurred, and often violently challenged. Most First Nations territories, former colonies, and countries with contested borders face challenging negotiations of heritage representation and interpretation. Museums in Jerusalem for example, must assess Jewish and Palestinian claims to different histories in shifting territorial space. Similar obstacles are being faced by heritage professionals narrating the ongoing histories of Republican Catholics–Unionist Protestants in Northern Ireland, Maori–Pakehas in New Zealand, and Shiites–Sunis–Christians in Iraq, to name a few. As tourism increasingly seeks out new products with the growth of cultural heritage tourism, the politics of representation 'for self' (local/national communities as part of identity building) versus 'for others' (tourists, as part of entertainment/visitation) becomes much more loaded. It is important to acknowledge the importance of heritage as an economic resource, but it can also be used to help examine the meanings of culture and power, thereby giving it a vital sociopolitical function.

On a micro level, the politics of representation becomes crucial for the practicalities of heritage site / museum exhibition design and layout. The process of heritage interpretation should be simply about moving knowledge from specialists to the general public in a clear and effective way. What is and what is not said, in complex settings, becomes important, and (more importantly) 'how' messages are portrayed takes on added significance. The task – for curators, educators, and exhibition developers – is to provide experiences that invite visitors to interpret meaning through deploying and extending their existing interpretive strategies and repertoires, using their prior knowledge and their preferred learning styles, and testing their own hypotheses against those of others, including experts!

Authenticity

Much debate exists as to whether tourists or indeed, people in general are concerned with engaging in an authentic heritage experience, or more simply, 'whether the truth matters'. For some, a visit to a heritage tourism site provides a sense of pleasant escapism, for others, it may be a quest for knowledge, or a complex matrix of many other social, cultural, or leisure oriented motivations. Nonprofit, public spaces concerned with the representation of heritage (e.g., free admission national museums) have arguably more leeway in their quest for authenticity. Other economically oriented tourism sites may allow for more nostalgic, stereotyped, or partial versions of the past, events, histories, traditions, and so on, in a bid to excite, entertain, or simply to keep visitors within their comfort zones.

Heritage spaces are sometimes seen as safe places, where visitors can indulge themselves looking for the lost meanings of the present in an idealized past. Yesterday is presented as a time where there is no responsibility or discord. Such nostalgic and sentimental views of the past as a 'golden age' are what makes it such a dangerous place for the escapist to inhabit. Much of our 'heritage' is interpreted to stimulate nostalgia, idealize the past, and leads to a selective understanding of the past that has more to do with fantasy than truth. Many tourism oriented sites construct heritage interpretation very selectively and portray only the lighter side of history in an attempt to exploit the public's appetite for reconstructed, comforting, and nostalgic images of the past. For some, however, heritage interpretation is about education and challenge while encouraging visitors to question their values and attitudes.

The idea of authenticity as a sociological discourse was initiated in the 1970s where it was discussed within the broader framework of a debate about the relationship between tourism and modernity. Commentators saw modern people as alienated from their own society, and therefore, reality and authenticity were thought to be elsewhere: in other historical periods and other cultures, in purer, simpler lifestyles. This builds upon the notion of nostalgia for the past representing dissatisfaction with the present. As the gaps between truth and fiction become more blurred, tourism engages in what has been termed 'staged authenticity' – where the sector knowingly presents heritage almost as theme park like parodies of itself, and in some cases, tourists are happily complicit in such a process.

The Politics of Heritage and Identity

In any nation, a multiplicity of quite different ideologies exist and can be conveyed through the same heritage, rather than any specific, coherent political program intended to support a distinctive prevailing view of society. Whether national conflicts are reconcilable through proactive, inclusive cultural heritage policy is arguable and remains to be seen in many places. Questions of identity and difference are common in the landscape of contemporary political theory. Problems of diverse heritage identities and representation are not exclusive to divided states such as Israel, Palestine, and Ireland, for example. Indeed, this notion of an internal national homogeneity (which draws inevitably upon a particular representation of heritage and a mythology of the past for its coherency and legitimization) – has conditioned western conceptualizations of political space for more than two centuries.

Many writers have examined the interconnection between heritage, identity, and place. Some argue that without being connected to a place, people begin to lose their identity and their heritage. They suggest that the past is integral to our identity and that staying connected to natal or long inhabited locales is a natural inclination. This view is increasingly being challenged as globalization grows in momentum and the role of virtual, cyber based communities who share common interest bases, rather than shared place identities, rises.

A common theme of some heritage literature is the concentration on the local nature of identity and the role of heritage, in times of uncertainty or upheaval. The urge to preserve the past can be viewed as an effort by people to preserve a part of themselves, and those feelings of nostalgia are often strongest when national identity is somehow threatened. Confusion about identity and loss of confidence is often seen in immigrant communities who have no natural or historical connection to their new place of residence. The mass movement of Central and Eastern European migrants in the first decade of the new millennium has seen new cultural compositions emerge in the West of the continent. Poland in particular, has experienced extreme population shift. America has a long history of multiculturalism and its specific heritage challenges range from the authentic representation of native American Indian peoples to the proposed primacy of Spanish as a legal language in California (to reflect the large Mexican/central American population composition). However, if it is culture and heritage, above all else, that are (as research suggests) fundamental to identity, their inclusive representation remains a charged political issue.

Where heritage is displaced (or even misplaced), questions of ownership arise and the heritage of the indigenous population can no longer be promoted as unambiguously universal or necessarily dominant. Concepts of hegemony (supremacy) and ownership are relevant here insofar as the creation of any heritage actively (or potentially) disinherits or excludes those who do not subscribe to the terms of meaning defining that heritage. The term 'dissonant heritage' is used to describe heritage which is embedded in discord and conflicting versions of the past or present. Contestation refers to negotiation, often between groups of people in local, national, or international spaces, in the search for legitimate versions of their own authentic heritage. Heritage is therefore simultaneously: knowledge; a cultural product; and a political resource. Key questions include why a particular interpretation of heritage is promoted, and whose interests are advanced or retarded by such versions of the past, or present? If heritage knowledge is situated in particular contexts of understanding or power relations, it is time specific and thus its meaning(s) can be altered as sociocultural and political situations change. Consequently, it is inevitable that heritage is contested and challenged in a dynamic manner. The question 'whose heritage?' is an increasingly complex one for many nations and localities.

Heritage also plays a crucial social function in terms of identity construction, at personal, local, regional, and national scales where it can be seen as the primary instrument in the discovery, creation, and subsequent nurturing of a national identity. Heritage can be framed as activities and events that establish or symbolize social cohesion. Therefore, heritage undoubtedly has cohesive as well as separatist potential. The psychological continuity and identity of such cohesive groups is a primary ingredient which acts as a cultural frame of reference. This function of heritage is important in periods of social transition or major national reorientation, for preserving hegemonic equilibrium and momentum.

World Heritage Conservation and Management

Attempts to construct a concept of heritage at a global level are conceptually problematic. However, given increasing levels of interdependency between people and places, practical attempts have been made to inventorize value and apply international conservation strategies.

The World Heritage Convention of 1972 was a landmark global gathering where it was stated for the first time, internationally, that parts of our cultural and natural heritage are of outstanding interest, and need to be preserved for humanity as a whole.

It established criteria and procedures for the inclusion, on a single World Heritage List, of properties constituting cultural and natural heritage worthy of preservation.

UNESCO's World Heritage Mission is to encourage countries to sign the Convention and ensure the protection of their natural and cultural heritage through the nomination of sites within their national territories for inclusion in the world heritage list. At the time of writing, over 700 world heritage sites (WHSs) existed in more than 120 countries, and the number increases annually. There are three possible categories for inscription onto the WHS list – (1) cultural heritage, which includes monuments, groups of buildings, and sites, and this is the largest category representing almost two thirds of the entire list; (2) natural heritage, which includes natural features, geological and physiographical formations; natural sites; and (3) mixed properties or cultural landscapes, representing combinations of the above and includes aspects of intangible heritage, which is by far the smallest WHS list category, reflecting classificational and definitional debates in previous sections above.

No two WHSs are (by definition) alike but all share common problems such as the need for a delicate balance between visitation and conservation. All are national flag carriers, symbols in some way of national culture and character. Most are major cultural tourism attractions of their country and some (such as Machu Pichu, the Eiffel Tower, the Great Wall of China) are powerfully evocative symbols of national identity and are universally recognized. World Heritage is a fragile nonrenewable resource which needs to be safeguarded, both to maintain its authenticity and to preserve it for future generations. Visiting a WHS should be a special experience, on a different scale from visiting other tourist sites. The majority of visitors to such places are, predictably, motivated by an interest in heritage and culture, although this motivation may not be matched by any prior knowledge of the site concerned making the provision of appropriate and meaningful interpretation even more important.

In terms of management and conservation, UNESCO aims to preserve for the present and the future the monuments that bear witness to the creative nature of humanity, and to make these treasures available to the widest possible public. In this light, there are specific WHS operational guidelines which designated sites must adhere to. Monitoring is conducted by specific committees, and WHS listing may be withdrawn if guidelines are not adhered to. General guiding principles include: (1) sites should be seen as integral parts of local and regional conservation, tourist and development plans, and not as 'intrusive islands of conservation', (2) good integration of site preservation, conservation, and presentation strategies, (3) the need for the development of suitable management plans, and (4) advanced recognition of threats to sites (e.g., pollution, tourism, farming, building development). Naturally, these issues differ depending on the 'type' of site (e.g., natural or cultural), the 'location' (e.g., urban or rural, developed or developing world), the organizations who own and/or manage the site (e.g., private, public, or voluntary bodies; single or multiple ownership/management). All WHSs must have effective visitor management strategies, thereby ensuring the sustainability of the sites themselves for future generations' appreciation.

Heritage not considered to be World Heritage in UNSECO's classification terms, is nonetheless of enormous value at multiple scales, and worthy of conservation for its own sake. In a changing world, new constructs of heritage, identity, authenticity, and representation will continue to emerge as meanings are constantly renegotiated over time and space.

See also: Memorials and Monuments; National Spatialities; Tourism.

Further Reading

Ashworth, G. J. and Larkham, P. J. (1994). *Building a New Heritage: Tourism, Culture and Identity in the New Europe*. London: Routledge.
Cohen, E. (2002). Authenticity, equity and sustainability in tourism. *Journal of Sustainable Tourism* 10(4), 267 276.
Cosgrove, D. (1998). *Social Formation and the Symbolic Landscape* (2nd edn.). London: Croom Helm.
Graham, B., Ashworth, G. J. and Tunbridge, J. E. (2000). *A Geography of Heritage: Power, Culture and Economy*. London: Arnold.
Graham, B., Ashworth, G. J. and Tunbridge, J. E. (2005). The uses and abuses of Heritage. In Corsane, G. (ed.) *Heritage, Museums and Galleries: An Introductory Reader*, pp 26 37. London: Routledge.

Hardy, D. (1988). Historical geography and heritage studies. *Area* 20, 333 338.
Harvey, D. C. (2001). Heritage pasts and heritage presents: Temporality, meaning and the scope of heritage studies. *International Journal of Heritage Studies* 7(4), 319 338.
Herbert, D. T. (ed.) (1991). *Heritage, Tourism and Society*. London: Mansell.
Hewison, R. (1987). *The Heritage Industry: Britain in a Climate of Decline*. London: Methuen.
Hewison, R. (1989). Heritage: An interpretation. In Uzzell, D. (ed.) *Heritage Interpretation* vol. 1, London: Belhaven Press.
Howard, P. (2003). *Heritage: Management, Interpretation, Identity*. London: Continuum.
Johnson, N. (1999). Framing the past: Time, space and politics of heritage tourism in Ireland. *Political Geography* 19, 187 207.
Lowenthal, D. (1985). *The Past is a Foreign Country*. Cambridge: Cambridge University Press.
Palang, H. and Fry, G. (2003). *Landscape Interfaces: Cultural Heritage in Changeing Landscapes*. Dordrecht: Kluwer.
Shaw, B. J. and Jones, R. (eds.) (1997). *Contested Urban Heritage: Voices from the Periphery*. Aldershot, UK: Ashgate.
Smith, L. (2006). *The Uses of Heritage*. London: Routledge.
Tilden, F. (1957). *Interpreting Our Heritage*. North Carolina: Chapel Hill Books.
Uzzell, D. (1998). Interpreting our heritage: A theoretical interpretation. In Uzzell, D. & Ballantyne, R. (eds.) *Contemporary Issues in Heritage and Environmental Interpretation*, pp 11 25. London: The Stationery Office.

Relevant Websites

http://www.unesco.org
United Nations Educational, Scientific and Cultural Organization (UNESCO) website this lists World Heritage Sites and explains the tangible/intangible dimensions to heritage.

Heritage and Culture

R. Jones, Curtin University of Technology, Perth, WA, Australia

Glossary

Cultural Landscape A physical landscape modified, over time, by human action.
Dissonant or Contested Heritage Objects, places, or behaviors which are regarded and valued in different ways by different cultural groups.
Environmental Determinism The belief that the local nature of the physical environment is the key factor controlling human behavior.
Iconography The study of landscapes, artworks, and other heritage objects in order to understand their symbolic meanings.
Social Darwinism The application of Darwin's views on biological evolution and natural selection to socioeconomic and political affairs.
Time–Space Compression The growing ease of interaction across space as a result of advances in transport and communications technology.
Universal Religion A faith which its adherents believe to be applicable to all human beings.

Introduction

One of the central concerns of historical geography is the study of the physical and social manifestations of culture in time and space. Cultures are manifested physically as people transform their landscapes in distinctive ways through, for example: agricultural techniques, such as terraced rice paddies; building styles, such as Alpine chalets; and even through language, since this dominates the myriad signs and advertisements that are so prominent in contemporary landscapes. They are expressed socially in the ways in which people behave and interact, for example, in the manner by and extent to which the different gender, ethnic, or religious groups are segregated – or segregate themselves – particularly in public spaces.

When certain cultural landscapes and cultural behavior patterns become valued, particularly by powerful and/or majority groups within a society they can come to be seen as part of that society's heritage (literally, as something that has been inherited from that society's ancestors) and thus as something that needs to be preserved, maintained, and revered. In landscape terms, this valorization is characteristically applied to things created by humans, such as temples or patterns of fields and hedgerows. However, particularly for societies with a recent history of hunting and gathering, the (predominantly) natural landscapes which supplied their basic needs are also viewed in heritage and, indeed, spiritual terms as something to which their society owes a stewardship responsibility. In the modern world too, there is a growing concern for natural heritage, for the preservation of, frequently familiar, animal and plant species and, therefore, of at least remnants of the natural habitats and landscapes in which these species can survive.

Heritage/inherited patterns of social behavior are perhaps most clearly associated with the success of powerful groups in imposing and perpetuating these norms. This is most obvious in the context of those religions which have become locally dominant and where behavioral norms of entire societies are (at least claimed to be) governed by texts that have been handed down from previous millennia or through even more ancient oral traditions. Even though nation states are largely a creation of the last few centuries, their governments often invoke more ancient heritages to justify some of their policies and thus to influence the behavior (and thereby the cultural practices) of their citizens, be this by reference to glories of the monarchs in contemporary republican France, of the Romans in modern Italy, or the precolonial Asian values praised by the former Prime Minister of Singapore.

These characteristics of heritage and culture present historical geographers and the wider society with challenges in temporal, spatial, and social terms. Temporally, appeals for the maintenance of heritage(s) and culture(s) would appear to inhibit changes to landscapes and behavior patterns. Yet technological, economic, and demographic changes are among the major driving forces of human history. The results of these changes include the transformations of landscapes (e.g., the disappearance of the open (strip farmed) fields of pre enclosure rural England, or the growth of industrial cities worldwide) and of behavior patterns (e.g., the disappearance of most hunter gatherer lifestyles and of many so called tribal languages). Neither cultures nor heritages are stable and, like the technologies and economies to which they are linked, they are therefore becoming increasingly dynamic.

Spatially, landscapes, whether cultural or natural, are inherently local and unique. For much of human history the same could be said of cultures. Until the urban revolution some 6000 years ago, humans lived in small subsistence groups and had limited outside social contact. Their religions, languages, economies, and behavior patterns were therefore highly localized. However, since then, the economic, demographic, and technological changes referred to above have also transformed the geographies of cultural behaviors as religions, languages, and other cultural traits, such as music, dress, and gastronomy, have become national, continental, and even global in their distributions. An idealized view of a culture and of its accompanying heritage, particularly from the viewpoints of those who see themselves as the guardians of these cultures and heritages, would be of a set of behaviors and artefacts built up in a specific place and remaining relatively unchanged over a long period of time. However, cultures and heritages are now becoming less rooted in place but more fixed, indeed fossilized, in specific time periods. In Australia, before European conquest, there were many indigenous cultural groups, speaking several hundred languages and practicing local cultures that had remained largely unchanged for millennia. For a brief period in the late nineteenth and early twentieth centuries a hegemonic Anglo Celtic Australian national culture appeared to be developing but, in the post war era, migration and other economic and cultural influences from Asia, America, and continental Europe have created a much more complex and dynamic society with far more diverse heritages.

The interrelated phenomena of change and globalization present us with the third challenge to the study, and the practice and preservation, of culture and heritage. Even in a dynamic and interconnected world, some landscape features and some behavioral traits endure and, indeed, are being actively preserved. The social challenge therefore relates to what is being preserved (in both concrete and social terms), to who decides what will be preserved, and to which aspects of whose heritages and cultures are being privileged in the course of these decisions. Contestations over whose cultures and heritages prevail lead to history wars which are being fought out on the ground, in the courts, and through political systems as various pressure groups (for example) oppose freeway construction, protect historic buildings, obtain indigenous land rights, and defend or attack linguistic or religious pluralism – or even foxhunting or bullfighting – in the name of heritage and/or culture.

Traditional Geographical Approaches to the Study of Heritage and Culture

In the late nineteenth and early twentieth centuries several intellectual and political trends and fashions led most Western geographers to focus on the study of those heritages and cultures which they saw as being localized, relatively unchanging, and unique. First, European geography and anthropology (disciplines which were linked in several universities) had a political and economic value in that their practitioners could provide the imperial authorities with useful information on the places and peoples of their empires. Most cultural geographers and anthropologists therefore conducted their research on peoples whose cultures were localized and slow to change – factors which may well have contributed to their incorporation into colonial empires in the first place.

Second, a central concern within geography was to integrate the studies of natural and human phenomena. While there was considerable disagreement between the determinists, who sought to explain human, cultural, and economic development in terms of (social) Darwinian adaptations to the physical environment, and the possibilists of the French Annales and the American Berkeley Schools, who depicted how different social groups developed distinctive cultural landscapes over time in specific places, both groups found local, long settled rural areas most suited to their researches.

Finally, as geography sought to define a distinctive identity for itself, the region became an important intellectual focus for many of its scholars. This 'highest form of the geographer's art' was seen to be the synthesis of all the available information – physiographic, climatic, biological, historical, demographic, social, economic, and political – on a single area. The areas studied in this way tended to be small, given the diversity of information required, and, once again, to be rural, since it was here that the relationships between these various types of information could be most readily demonstrated.

In the early years of geography's modern period, therefore, the discipline's practitioners tended to focus on the study of rural and traditional cultures and to concentrate on the nature and the evolution of distinctive cultural landscapes. Frequently, they would describe processes of sequent occupance whereby several different cultural groups (e.g., Celtic, Roman, Saxon, Scandinavian, and Norman in Britain) had occupied the same area over a long time period. However, at least where these changeovers had occurred in the distant past, the current occupants could often see all these components as part of a heritage landscape with which they could identify because it was 'familiar and well loved'.

These early scholars therefore tended to describe areas and peoples where the economies and cultures were uniform and where change was slow, focusing on rural, rather than urbanizing, areas in the developed world and on cultural groups as yet little affected by processes of colonization and modernization in the Americas, Africa, and the Asia Pacific region. For many of these earlier scholars in cultural geography and its related disciplines any change to these relatively stable and localized phenomena presented a threat to both the cultures and the heritages of their objects of study. As late as 1955 this view was summed up by England's first professor of local history as "since the year 1914 every single change in the English landscape has either uglified it or destroyed its meaning or both." Since that time, however, historical geographers and their academic colleagues have increasingly turned their attention to the study of the forces of change acting upon the world's cultures and the increasingly complex challenges for these cultures' heritages that have resulted from these forces.

Processes of Cultural and Heritage Change

While change is virtually the only constant in human history, the rate of change has accelerated significantly over the last half millennium. The global population, which had barely doubled over the previous 1500 years, increased more than tenfold over the last 500. The factors that made this population growth possible are primarily technological, but they have had a profound impact on both cultures *per se* and on the geographies of these cultures and, in the process, they have added immensely to the complexities of heritage creation, obliteration, and transformation.

On a global scale, improvements in ship design and navigation, in weapons technology, and in administrative and political organizational systems between 1500 and 1900 enabled a range of European powers to colonize most of the rest of the planet and, notably in the case of China, to dominate in economic and trading terms, those areas which they did not control politically and militarily. In the Americas and Australasia, European emigrants also came to dominate these areas demographically and, in the process, to marginalize and subordinate a vast number of indigenous cultures and heritages. While most Asian, African, and Pacific colonies were administered, rather than settled, by Europeans, the colonizing powers nevertheless sought to impose many of their cultural values on the colonial populations. Not only did conquest in itself convey a sense of cultural superiority to the European colonizers who could thereby regard the colonized populations as savages to be civilized to European ways but, as the colonial era progressed, pseudoscientific rationalizations of European/white so called racial superiority and notions of a social Darwinism which favored the survival of the fittest were used to provide justifications for the near extermination of native peoples, the suppression of their languages, religions, and other cultural practices, and their co option into indentured labor contracts or a burgeoning slave trade on an intercontinental scale.

A further rationalization, if not motivation, for both political and cultural imperialism over the last millennium and a half has been the development and spread of the universal, proselytizing religions of Christianity and Islam. Just as the imperialist powers could view the inhabitants of weaker countries as potential subjects and workers and, eventually, as speakers of imperial languages, so could the adherents of these universal religions see the same people as potential converts and co religionists. The Islamization of an area from North Africa to Southeast Asia and the Christianization of most of Europe and then of the Americas, Australasia, and the Pacific and Sub Saharan Africa remain some of the most significant cultural geographic shifts of the last 2000 years.

Following and overlapping with the Age of Discovery (i.e., the period in which Europe discovered the rest of the planet), the Industrial Revolution also impacted on cultural geographies and heritages worldwide. Pre industrial societies were largely rural, with the majority of the workforce being locally focused and involved in primary production. From the early nineteenth century onward, the Industrial Revolution and the resultant mechanization of agriculture, fishing, forestry, and mining displaced an increasing number of former rural workers and obliged them to seek employment in the growing towns and cities. Insofar as culture and heritage are bound up with a way of life, this urbanization process (which has recently seen urban dwellers form the majority of the world's population and shows few signs of abating in the foreseeable future) is a cultural, as well as a physical uprooting of the individuals involved. Cities became, as Barthes put it, "the place of our meeting with the other." People from different rural areas moved into cities looking for work or a better life. As the Industrial Revolution progressed, people also moved from town to town and even country to country as industries and businesses rose and fell in increasingly dynamic industrial economies. From the Irish in Boston to the Greeks in Melbourne or the Chinese (in themselves culturally diverse) in San Francisco, migrants to the world's urban areas have brought and continue to bring a range of different cultures and heritages to the cities that they populate.

A final impact of technological change on the geography of cultures results from some of the most recent advances in transport and communications. Transcontinental rail and road systems, super tankers and container ships, and jet airliners have facilitated long distance migration and trade, contributed to a globalization of the economy, and allowed the development of global brands, such as Toyota, Guinness, and Hilton Hotels. However, the globalization of culture (insofar as it exists) requires that these global brands are globally demanded and consumed and it is here that the globalization of the media and recent advances in telecommunications technology come into play. Major sporting events, such as the (Soccer) World Cup, can be televised globally, major brands can be advertised globally, films and television programs can be distributed and shown globally. 'Mac Donaldization' may not be as traumatic as invasion, religious conversion at sword point, or economically forced migration, but it does change behaviors, causing people in different parts of the world to consume the same (physical or media) products and to aspire to the same lifestyles. It can thus produce 'nonplace urban realm' landscapes where the relative lack of any cultural or heritage markers in a city center or shopping mall are such that they could be almost anywhere in the world.

These relatively recent processes of technological, economic, and demographic change have produced two, apparently divergent cultural and heritage effects. On the one hand, they have produced greater cultural uniformity as local languages (a central component of local heritages) are replaced by imperial ones or, in the case of English, potentially by a global language, as local and tribal religions are replaced by universal ones and as a more standardized urban way of life replaces a variety of predominantly rural folkways. But on the other hand, modernity has created much greater cultural complexity as migration processes have mixed up cultural groups to an increasing extent and as advances in transport and communications have allowed the members of these cultural groups to remain in contact with one another and thus to maintain their heritages and cultural identities to a degree that was formerly impossible.

Forces for the Maintenance of Heritage(s) and Culture(s)

While the forces of modernization discussed above have wrought profound changes on the geographies of the world's heritages and cultures, this has by no means led to the creation of heritage or culture free zones. A whole range of social, economic, and political forces continue to differentiate, sustain, and even invent heritages and cultures worldwide.

An inevitable corollary to the diffusion of cultural traits, such as the global spread of religions and languages mentioned above, is their subsequent adaptation to local conditions. The universal religions are perhaps doubly prone to fragmentation, either for specifically doctrinal reasons (e.g., the debates over transubstantiation and predestination that led to the separation of the Lutheran and Calvinist churches from Roman Catholicism) or for more local ones (e.g., Henry VIII's desire for a divorce/a male heir and the resultant establishment of the Church of England or the development of a Dutch Reformed Church which sought to justify South African apartheid in a Christian context). In a wired modern world, it is unlikely that languages will split in the way that French, Italian, and Spanish evolved as local variants of Latin, though English is still evolving on slightly different lines in Britain, North America, the Caribbean, South Asia, Africa, and Austral(as)ia. Even minor cultural traits may bring about some differentiation. Singapore airport's first terminal was built to a global set of specifications. However, when it was noted that more family members than had been anticipated were accompanying passengers, more meeting space was added in the designs for terminal two. Any attempt to establish cultural uniformity nationally or globally is likely to be met by local resistance as people seek to retain dialects, recipes, and iconic buildings and landscapes to which they are attached and thereby to at least retain, if not create, heritages from below.

Just as an icon of globalization, such as an airport terminal can be modified locally for social reasons so, for economic reasons, do global firms modify their products for different markets, for example, using *halal* or kosher products in their food chains and conducting market research on local consumer preferences. But perhaps the main economic driver for the preservation of local cultures and heritages is the world's largest industry, tourism. While countries, regions, cities, and industries will seek to emphasize the international standard of their service, their tourism marketing is fundamentally place marketing. It stresses what is unique, different, and special about their localities and their peoples. Both natural and cultural heritages are stressed in this context and it is therefore in the interests of the localities which are dependant on this industry to see that both are preserved, though this may involve some compromises. It is hoped that highly regulated and highly priced big game hunting in Africa will bring economic gains and limit poaching in a way that would be impossible if this activity were totally banned. Many traditional dance and music performances are shortened and modified for tourist consumption, but equally the cultural groups involved acknowledge that many of these traditional cultural practices only survive because of tourist interest. It is through such compromises that a heritage industry has developed.

The heritage industry seeks to preserve, replicate, and even fabricate aspects of local and national history and culture at stately homes, historical theme parks, and shops attached to major religious buildings, archaeological sites, museums, and art galleries. The primary aim of such ventures is economic, though the profits may be used for the preservation of cultural landscapes and artefacts. Interestingly, with the accelerating pace of change, the heritage industry has diversified to include the preservation of relatively recent (but disappearing, or at least changing beyond recognition) phenomena. These can include industrial sites, sporting venues, and examples of superseded architectural styles.

However, just as the marketers of tourist destinations seek to emphasize the distinctiveness of their heritages and cultures so too do political leaders. In premodern times, it was taken for granted that these leaders could and would impose their will on their subjects (a telling term in itself). In the modern world, however, citizens of nation states need to be persuaded to support their leaders and their countries. Contemporary leaders therefore invoke the distinctiveness and, implicitly, the superiority of their nations' cultures and heritages in order to obtain popular support. Governments designate national languages and, frequently, national religions; they regulate school curricula and fund the preservation of national 'treasures', such as buildings, monuments, and national parks. They also develop national iconographies consisting of flags, anthems, and notable holidays and ceremonies. In doing so, they frequently contribute to what Hobsbawm and Ranger term the invention of tradition whereby (particularly new) nations can use these symbols to develop a national identity. The speed with which the Anzac legend and its associated rituals became a central part of the national heritages of Australia and New Zealand following World War I is a reflection of the power of this process.

Inhabitants of the contemporary world therefore perform their cultures and valorize their heritages in highly complex ways. Within their communities, they conform to local cultural norms and seek to protect what they see as heritage features of their local landscapes. At the national level they are frequently exhorted to respect their nation's heritage as this is perceived by government and required to adhere to the cultural frameworks set down for them in legal and constitutional terms. At the same time they can experience an infinite number of other cultures and heritages, directly through travel or indirectly through the media, not only on a global scale but often within their own countries and cities.

Conclusion: Contemporary Heritage and Cultural Challenges

Dissonant or contested heritage eventuates when different cultural groups occupying the same locality view the landscapes and social practices which result from that locality's history in different ways. Colonial buildings, monuments, and rituals may resonate positively for a settler population, but negatively for a country's indigenous inhabitants for whom they might symbolize dispossession and oppression. More generally, in any region or country containing diverse cultural groups, the cultural practices of one group may tend to be favored over the others. The restrictions on the use of Chinese signage in Jakarta or English signage in Montreal were put in place when the Indonesian and French Canadian groups, which had hitherto been subordinate, gained greater political power.

Most of the world's countries and cities now contain complex cultural mixes within their populations. For those outside Europe, these consist of varying proportions of the indigenous peoples (often with varying cultural backgrounds since imperial boundaries rarely reflected preexisting cultural divisions), settlers from one or more or Europe's imperial powers, and other migrants imported forcibly (i.e., slaves from Africa), induced (indentured laborers, such as the Indian populations of Fiji and the Caribbean) or attracted (e.g., the Chinese in the Pacific rim) to countries whose economies and societies were experiencing transformations. Within Europe, both the frequent shifting of political boundaries over centuries of conflict and more recent migration from their former colonies to the ex-imperial powers have also resulted in growing cultural complexity.

In such circumstances countries such as Japan, with a relatively uniform culture and heritage, are acknowledged as exceptional. Most nations face the task of dealing with both cultural pluralism and cultural dynamism, both of which can result in heritage conflicts. In addition to the rise in both legal and illegal international migration, many indigenous groups, such as the Maori, who were thought to be in danger of dying out a century ago are now experiencing rapid growth rates and make up a growing proportion of their countries' populations. In democratic states at least, national governments characteristically seek a politically acceptable balance between a range of cultural groups who may see themselves as deserving of special consideration by reason of: prior occupance (e.g., indigenous groups); historical control (e.g., settler groups); population numbers (e.g., the Chinese majority in Singapore); prior injustice (e.g., the Black population of the US); or any number of more idiosyncratic, local factors. Policy responses may include positive discrimination (e.g., for Bumiputras (Malays) in Malaysia), deliberate reconciliation (as in South Africa), antiracism moves (as legally enacted by most European governments), and culturally specific land rights policies (e.g., for the Sami (Lapps), Maori, and Native (North) Americans).

Many of these policies have had beneficial results, but in many cases the granting of cultural rights or the acknowledgment of the cultural heritages of one group may conflict with the cultural heritages of others. The practices of some strict adherents of Islam conflict with the principles of gender equity espoused by many Western nations. Genital mutilation of the kind practiced in parts of Africa is illegal in most countries outside that region. Even though Christianity can be said to have taken over many components of pagan winter solstice celebrations, many Christians see contemporary attempts to rename Christmas as Winterval or to change the 'Happy Christmas' greeting to 'Happy Holidays' as an attack on their Christian heritage.

In these circumstances national governments are obliged to act as intermediaries in two ways. First, in spatial and hierarchical terms, between global bodies, who proclaim a Universal Declaration of Human Rights or designate World Heritage sites, and local groupings, who disagree either with these decrees or with their government's response to them. Indigenous groups in Norway and the Northern Territory of Australia have appealed to the United Nations on heritage grounds over the actions of their national governments. Second, national governments have to deal with different cultural groups within their borders who may find themselves unwilling or unable to conform to national norms, as in the dispute over the wearing of the *hijab* in French state schools or even, as with nationalist right wing parties in Europe or fundamentalist parties in the Islamic world, to tolerate cultural difference or to acknowledge the heritages of others.

The question of who decides which (or whose) heritages and cultures prevail or survive and in which ways they do so is therefore highly vexed. Clearly those with the power to make and to enforce decisions on issues of religion, language, and nationality play a major role. However, many decisions made in the contexts of planning, health, environmental protection, censorship, labor conditions, tourism, sport, and recreation also have significant cultural and heritage related impacts. The study of heritage and culture within geography is therefore no longer conducted in a rural and historical backwater. It is central to a much wider range of contemporary political, environmental, and economic issues.

See also: Diaspora; Globalization and Transnational Corporations; Heritage; Heritage and Economy; Heritage and Identity; Identity Politics; Imperialism, Cultural; Imperialistic Geographies; Modernity; Trade, Transport and Communications, Historical Geographies of.

Further Reading

Dewailly, J. M. (1997). Images of heritage in rural regions. In Butler, R., Hall, C. M. & Jenkins, J. (eds.) *Tourism and recreation in rural regions*, pp 123 137. Chichester and New York: Wiley.
Graham, B. and Howard, P. (eds.) (2008). *The Ashgate research companion to heritage and identity*. Aldershot: Ashgate.
Graham, B., Ashworth, G. J. and Tunbridge, J. E. (2000). *A geography of heritage: Power, culture and economy*. London: Arnold.
Hewison, R. (1987). *The heritage industry: Britain in a climate of decline*. London: Methuen.
Hobsbawm, E. and Ranger, T. (eds.) (1983). *The invention of tradition*. Cambridge: Cambridge University Press.
Hoskins, W. G. (1970). *The making of the English landscape*. Harmondsworth: Penguin.
Johnston, R. J., Taylor, P. J. and Watts, M. J. (eds.) (2002). *Geographies of global change* (2nd edn.). Oxford: Blackwell.
Schama, S. (1995). *Landscape and memory*. London: HarperCollins.
Short, J. R. (2005). *Imagined country; environment, culture, and society*. Syracuse: Syracuse University Press.
Taylor, P. J. and Flint, C. (2000). *Political geography; world economy, nation state and locality* (4th edn.). Harlow and New York: Prentice Hall.
Tatham, G. (1957). Environmentalism and possibilism. In Taylor, G. (ed.) *Geography in the 20th century* (3rd edn.). pp 128 162. London: Methuen.
Tunbridge, J. E. and Ashworth, G. J. (1996). *Dissonant heritage; the management of the past as a resource in conflict*. Chichester and New York: Wiley.

Relevant Websites

http://www.icomos.org
ICOMOS International Council on Monuments and Sites.
http://www.tandf.co.uk/journals/routledge/13527258.html
International Journal of Heritage Studies.
http://www.multilingual matters.com/multi/journals jht.asp?TAG=&CID=
Journal of Heritage Tourism.
http://whc.unesco.org
World Heritage Centre.

Heritage and Economy

G. J. Ashworth, University of Groningen, Groningen, The Netherlands

Glossary

Bequest Value The value placed on the continued existence of a resource for the user of future generations.
Contingent Pricing The price of a nonmarket resource in a hypothetical market in which individuals are questioned as to their willingness to pay to maintain the resource.
Existence Value The value placed on the continued existence rather than the immediate use of a resource.
Free-Riding The benefits gained from a resource whose costs are paid by another.
Hedonic Pricing The use of a parallel market to determine the value of a nonmarket product.
Option Value The value an individual attaches to keeping open the option to use.
Opportunity Cost The benefits forgone when resources are devoted to another alternative purpose.
Public Good A good that is nonexcludable and nonrival: it is available without individual cost to all without excluding anyone.
Travel-Cost Pricing The use of the willingness to pay the costs of travel to a resource in determining its price.
Total Economic Value The total of direct and indirect use values together with other nonuse values.
Use Value The value placed on a site or resources by its users.

On Heritage and Economies: An Inevitable Ambivalence

Heritage, as the contemporary use of pasts, has been created to serve many cultural, social, political, and psychological uses – both individual and collective. As such it has become endowed with many policy expectations. Economic uses have generally been subsequent and secondary. They have often been a barely tolerated extra use for monuments, sites, collections, and places that have been identified, selected, designated, maintained, and interpreted for quite other purposes. Some marginal income or political justification may be obtained from visitors but economic considerations were not decisive in the creation of most existing heritage and, if the economic function disappeared, most would continue to exist to serve other purposes. Indeed many institutions and individuals involved in the heritage creation and management process regard the economics of their endeavors as, at best a distasteful necessity to be tolerated, and, at worst, a distracting and even degrading commercialization of their endeavors.

However, the simple reality is that heritage costs money but also earns money. Although, to many heritage consumers public statuary, museums, conserved cityscapes, and landscapes appear to be freely accessible and with zero, or near zero, priced facilities, to their producers they are costly to create and maintain. A protective heritage designation may appear to cost nothing: it is in reality a moral commitment to restore and maintain into infinity. As important as the direct costs of maintenance, are the indirect costs of the development opportunities forgone, often in urban settings on the most valuable spaces, and the associated restrictions placed on the adaption of buildings, neighborhoods, and landscapes to changing circumstances. This locking up of economic resources and acquisition of financial commitment into the future, the so called 'heritage time bomb', can be seen as a form of economic mortgage upon future generations in the service of the perceived heritage needs of the present. However, in compensation, heritage may earn money in numerous ways directly and indirectly as discussed below. Indeed heritage is being increasingly relied upon as an instrument for the attainment of economic objectives at all spatial scales. Whereas traditionally heritage, like culture in general, was seen as a 'merit good' to be acquired from the surplus of economic development once more essential needs had been met, now the position has been reversed as heritage is enlisted to serve economic objectives alongside its many other uses. Once a result and beneficiary of wealth, heritage is now expected to be a wealth generator.

Given the economic contexts in which heritage must inevitably exist, the failure to develop an economics of heritage is remarkable. With a few notable exceptions, such as the pioneering work of Lichfield and the founding of the *Journal of Cultural Economics* in 1973, there has been a reluctance among heritage 'professionals' to think in economic terms, which are often unfamiliar to them. Similarly among economists, the intrinsic difficulties in applying economic concepts to the heritage experience seem to have dissuaded many of them from the attempt. The shaping of an economics of heritage is, however, a necessity if the multiple uses of heritage are to be reconciled and heritage is to compete for resources with other functions.

The Economics of Heritage

The construction of such an economics of heritage would need to consider three main elements.

The Components of the Heritage Process

The four main components of any economic production system – namely products, producers, consumers, and markets – each have distinctive characteristics when applied to heritage. These differences, from most other economic activities or industries, complicate the analysis but do not render it impossible.

The heritage product is not only highly diverse; it is also vaguely demarcated. Almost any natural or human creation, whether tangible or intangible, can become heritage. The product is the packaged interpretation, and not the resources from which it may be composed: the same resources may contribute to many different heritage products. It may even be a geographical location, site, or place, giving rise to place products that remain places as much as products.

Producers of heritage products are as diverse as the products themselves. It may be a public sector agency, operating in a certain collective public interest, or a commercial profit seeking enterprise. The 'industry' is highly 'vertically disintegrated' in that those who create and manage the resources, package and interpret the products, transmit it to the market and manage its consumption, are likely to be quite different agencies, with different approaches, working methods, and, most important, objectives. Not only do those marketing and managing the consumption often have little control over the resources they are using, many of those concerned with the resources have little knowledge, or even interest, in how they are ultimately consumed. This fragmented production system – in which resource creation, product shaping, transmission, and consumption are performed by separate agencies – is a potential source of not only economic inefficiencies but also conflict between the different parts of the system.

The consumer is equally difficult to isolate and identify. It is, often, not clear who the heritage consumers are and indeed what they are actually consuming. If the heritage product is ultimately an experience, whether educational, aesthetic, spiritual, or just recreational, then each individual consumer can be assumed to be consuming a separate product. Similarly, we all as citizens, residents, and tourists consume heritage in some way, to some extent. The difficulties of demarcating, let alone segmenting and targeting, the market are compounded by the idea of each individual constructing a necessarily individual product from what is made available. The fundamental economic distinction between producer and consumer becomes blurred in the concept of 'prosumer'.

Finally, the market in which producers and consumers come together to exchange values according to their different strategies of profit maximization is as difficult to define and as varied as heritage itself.

The Valuation and Pricing of Heritage

If heritage is to be considered as neither valueless nor priceless, then methods must be found for valuing heritage products and resources and applying some form of pricing to their consumption. Such thinking may be anathema to many who cannot or will not put a price on art, beauty, or historicity, but if heritage is regarded as priceless then it is also valueless in economic arenas when a heritage use must compete with other commercial users for scarce resources or for its very continued existence. Only values and prices allow markets to operate, and even heritage created in the public sector exists within a market context of alternative use of resources.

Heritage acquires economic value in a number of ways which are more diverse than with most products, a condition which in turn determines how a price can be calculated. Heritage demand may be for its immediate use or only for its continued existence. There is also a demand for uses that may be deferred into the future, for an option to use, which will never be realized or for bequest to an imagined future user. This compound evaluation of the value of heritage can be expressed in the equation:

$$\text{Use value (direct} + \text{indirect)} + \text{existence value} + \text{option value} + \text{bequest value} = \text{Total value}$$

The logic is both correct and useful, insofar as it establishes that heritage has values in addition to those of its immediate use, but the task of calibrating such an equation is formidable. However, heritage is not unique in this respect and many methods have been determined for estimating values in addition to use values for commodities not, generally, directly traded, such as environmental quality and safety assessments. There is 'hedonic pricing' which attempts to place a real estate value, through comparison with similar cases, on the additional 'hedonic' value above its use value that is conferred by the heritage status of buildings or landscapes. There is 'travel cost pricing' which implies that the costs of visiting a heritage site represents its added value, which can be aggregated for all visitors. 'Contingent pricing' that imagines a demolition auction in which individuals may bid to save a heritage site or structure that has become popular in planning participation in environmental impact assessments effecting large numbers of individuals. All these, and many more, such as the 'Delphi methods' for consulting experts or stakeholders, produce incomplete and disputable results. They are, however, a considerable improvement on the alternative of inaction, which leaves heritage resources to compete in market places with much of its value unexpressed.

Investment and the Allocation of Profits

Investment in heritage might be viewed as any other economic investment – namely a building, area, or land scape is preserved and renovated by an owner who then reaps the profits from the enhanced value of the property. However, this straightforward relationship between in vestment and return is complicated, in heritage, by three considerations. First, both costs and benefits accrue more widely than just to those who own or use heritage. The economic externalities accruing to a wider society are often more important, and indeed, may be the justification for the investment, rather than the internalities. Second, much heritage is produced and consumed as a public merit good. Consumption cannot be confined to those who pay for it, in addition, it may actually be compulsory in the sense that the individual cannot opt out of ex periencing heritage environments. Third, the costs of preservation, renovation, and restoration are 'front loa ded', that is, the investment must be made substantially long before any returns are obtained and may only be repaid over many years or even centuries. Incremental investment to partially save or restore is rarely possible.

These characteristics are evident in the rational pro cess of deciding to invest illustrated in the 'neighbor's dilemma' model. Here two neighbors, A and B, can choose to invest in renovation. If one does and the other does not, then the external benefits and costs determine that the investor fails to receive a sufficient return, bearing the external cost of a neighboring nonrestored property, while the noninvestor receives a free benefit from the restored neighboring property. The third party, the wider community, receives a positive externality from either neighbor investing. Although a profitable strategy for the individual must always be 'not to invest but hope your neighbor will', return on investment is maximized if both invest simultaneously. Also, the returns to the community are highest if both invest as this justifies community encouragement, and even investment, to initiate the process. Public intervention is thus needed to redress market failure and combine private and public economic interests and to distinguish between individual and collective costs, benefits, and responsibilities. This dilemma lies at the core of area development and re generation issues and there are numerous ways in which it is manifested and managed at the local level.

Heritage in Economics

Having considered the economics of heritage, the various roles of heritage in economies must now be briefly outlined. This can be approached through a number of overlapping models.

Windfall Gain Model

There is a widespread assumption that heritage has been created and is maintained in the service of noneconomic purposes. Its economic use can, therefore, be viewed as a windfall gain. This presupposes that the additional eco nomic demands can be accommodated alongside other demands without competing with, displacing, or reducing the value to other consumers. It also assumes that the additional marginal use does not diminish or alter the resource upon which other uses depend. Third, it is in ferred that the addition of the economic use is con ditional upon the acceptance of other priority uses, and may be monitored and regulated. This model also as sumes free ridership, although who is riding free depends upon how the model is interpreted in a given situation. An economic enterprise may freely use a heritage facility created and maintained for other purposes: alternatively, a heritage enterprise may generate returns that maintain a facility in existence for free use by others.

Turnstile Model

At the level of the individual heritage facility, it may not appear to matter much as to who consumes or why they are consuming the experience. Visitor numbers or ex penditures become the critical variables. However it is dangerous to assume that it is the same product that is being consumed by each customer and much public heritage is a freely accessible public good which thus has no turnstiles.

Industrial Assembly Model

Commodification – the process for creating a tradable product from what was previously not tradable – can be viewed as a form of industrial assembly in which re sources, products, and markets are combined. There are three processes: namely, resource activation and main tenance, product assembly through selection and pack aging, and delivery to consumer markets. These processes are undertaken by three groups of actors: the resource caretakers, the product assemblers, and the market managers. Such a model reveals some of the significant characteristics of heritage, such as the disintegration of different stages in the process, and also that the model is essentially market rather than resource driven. It re mains, however, something of an abstraction for the same reasons. There is in reality no such integrated assembly system: those operating at different stages of the process often have little interest in, or even awareness of, each other. The demand side of the model is certainly market driven but the supply side is usually not, with the re source providers using intrinsic object based values rather than ascribed market values. From this stems much of the lack of understanding, and often friction, between the two sides of the model.

Economic Systems Model

The distinction between internal and external pro duction systems allows an examination of internalities/ externalities and of 'sustainability' as an economic policy option. This, as in the management of many similar en vironmental systems, will need to take into consideration the following variables. The methods of resource use valuation; output equity (whether spatial, sectoral, and especially, generational); carrying capacity (for whatever objective); and the devising of self balancing economic systems. The application of such a model depends upon the understanding and manipulation of the characteristic components of the heritage process, as described earlier.

Heritage in Economic Development Strategies

Heritage is called upon to play diverse instrumental roles within development strategies for economic goals at different spatial scales. This can be approached from a variety of directions.

Heritage as Culture Industry

Heritage can be treated as an economic enterprise, as part of a wider set of cultural industries, with a distinctive geography of production. It has identifiable spatial lo cation factors, distinctive industrial structures of pro duction, patterns of employment and traceable interfirm linkages, and networks. Heritage enterprises can thus be planned and managed similar to other cultural enterprises in the public interest by using its distinctive industrial characteristics of small firms, vertical disinte gration, spatial clustering, a dependence on specialized labor pools and the like, all of which have strong geo graphical dimensions.

Heritage as Tourism Industry

Heritage tourism is not only a special case of a heritage industry but its numerical and economic importance has generated special management and scientific and political attention. Tourism treats much heritage as a zero cost, freely accessible, flexible, and inexhaustible resource. The resource providers are likely to see it as costly, multiused and in danger of damage or depletion. The ubiquity and flexibility of the heritage resource renders it ideal for the creation of tourism products, many of which are linked specifically to places. The urban geography of heritage tourism is considered elsewhere, in the concept and management of the 'tourist historic city'. These place products are consumed selectively, rapidly, and fashionably by tourists, which necessitates a constant extension and differentiation of the product line. Suc cessful heritage tourism places are in a process of con tinuous reinvention of their imagined pasts in response to their changing markets. The appreciation of the issue of tourism as support and justification for heritage or, con versely, as parasitical exploiter and potential despoiler of resources they have not created, do not manage, and are not the sole or even most important user, depends largely on the viewpoint of the observer and its resolution de pends mostly on the priorities set by governments and place managers.

Heritage as Location Factor

Heritage may be a variable, whether positive or negative, in the location decision of other economic activities. The existence of conserved districts and landscapes, of mu seums and heritage centers, and even just the idea that a place is historic may be attractive to residents as an en vironmental amenity or to economic enterprises that can capitalize upon the place association with historicity. Equally, the restrictions on use and adaption imposed by conservational designations upon structures and areas are a cost to residents and enterprises and the heritage place associations may not be appropriate and profitably transferable to the firm.

Heritage and Place-Product Promotion

Related to the above is the practice of using heritage within wider place marketing strategies or as an instru ment of place branding. The essentially unique nature of local pasts, as well as the flexibility of heritage as an imagined construction, renders it ideal for place–product differentiation in search of unique selling point, or unique associations, of the place product or place brand. Historic events, personalities, mythologies, folk lore, and customs are thus widely used in the construction of place images for place product marketing, imaging, and branding as a means of differentiating one place from another.

Heritage in Area Economic Regeneration

Finally in this list of strategic economic uses of heritage, there is economic regeneration policy at various spatial scales. There are a few cases where heritage has been used to fundamentally reshape the economy of a place through becoming in itself the dominant economic sector. The Oberammergau 'Passion Plays', the Chemainus, British Columbia, heritage murals or the Stratford, On tario, Shakespeare festival are notable for both their success and their rarity. Much more widespread is the use of heritage indirectly, through its contribution to a broad based amenity and environmental quality.

Heritage in the form of conserved and adapted buildings, promoted associations with historic events and personalities, and major museums and galleries, which themselves may be new but which house heritage art and artifacts, may all contribute to area regeneration even when economic development has little directly to do with heritage. Waterfront regeneration strategies (from Boston to Bristol), in particular, have used heritage in this way but always in combination with other place management instruments, such as land use measures, area development investments, and place management more generally.

Limits and Possibilities

Both the economics of heritage and the roles of heritage in economies offer many possibilities, but are also constrained by many limits. Only rarely are economic considerations paramount in the selection and management of heritage and heritage the main support or justification for economic development. Heritage possesses many economic dimensions and plays many roles in development strategies. It offers ubiquitous, flexible, and mutable resources and is, therefore, in many ways an ideal instrument of development, most usually, in combination with other management instruments. However, the Hewison school of thought argues, conversely, that heritage is intrinsically an obstacle to development though the restrictions on change it imposes, the resources it consumes and fossilizes, and above all, the backward looking mentality it engenders amongst those who seek refuge in the past from the challenges of the present and the future. Heritage is inevitably about choice: economics provides some guidance in the making of these choices. Equally, economic development and the regeneration of areas involves choices of objectives and instruments. Heritage may, but equally does not have to, be chosen to be among the goals or the instruments.

See also: Cultural Economy; Heritage and Culture; Tourism.

Further Reading

Ashworth, G. J. (2001). Conservation designation and the revaluation of property: The risk of heritage innovation. *International Journal of Heritage Studies* 8(1), 9 23.

Baer, W. C. (1995). When old buildings ripen for historic preservation: A predictive approach to planning. *Journal of American Planning Association* 61(6), 82 94.

Bedate, A., Herrero, L. C. and Sanz, J. A. (2004). Economic valuation of the cultural heritage: Application to four case studies in Spain. *Journal of Cultural Heritage* 4, 1101 1111.

Bianchini, F. and Parkinson, M. (1994). *Cultural policy and urban regeneration*. Manchester: Manchester University Press.

Evans, G. (2001). *Cultural Planning: An urban renaissance?* London: Routledge.

Hewison, R. (1987). *The heritage industry: Britain in a climate of decline*. London: Methuen.

Lichfield, N. (1988). *Economics in urban conservation*. Cambridge: Cambridge University Press.

Lichfield, N. (1997). Achieving the benefits of conservation. *Built Environment* 22, 103 110.

McLoughlin, J., Sodagar, B. and Kaminski, J. (2006). *Heritage impact 2005: Proceedings of the first international symposium on socio economic impact of cultural heritage*. Budapest: Archaeolingua.

Mason, R. (1999). *Economics and heritage conservation: Concepts, values and agenda for research*. Los Angeles, CA: Getty Conservation Institute.

Navrud, S. and Ready, R. C. (eds.) (2002). *Valuing cultural heritage: Applying environmental evaluation techniques to historic building, monuments and artefacts*. Cheltenham, UK: Edward Elgar.

Stabler, M. (1996). Are heritage conservation and tourism compatible? An economic evaluation of their role in urban regeneration. In Robinson, M. (ed.) *Tourism and culture*, pp 417 446. Newcastle, UK: Centre for Travel and Tourism University of Northumberland.

Timothy, D. J. and Boyd, S. (2002). *Heritage tourism*. London: Prentice Hall.

Weiler, J. (1998). Dollars and sense: The economics of conserving built heritage. *Heritage* 1, 17 20.

Wynne, D. (ed.) (1992). *The culture industry: The arts in urban regeneration*. Avebury: Aldershot.

Heritage and Identity

J. D. Bohland, and E. Hague, Hollins University, Roanoke, VA, USA, and DePaul University, Chicago, IL, USA

Glossary

Collective Memory A collection of highly constructed and often mythologized beliefs about the past held in common by a particular group of people.

Diaspora An ethnic or national population dispersed over multiple locations, most commonly associated with Jewish people.

Ethnicity A personal identity derived from membership in a national or subnational group believed to share a common culture.

Hate Crime A premeditated crime, such as an assault, pursued by the proponent on the basis of a victim's identity, most commonly her or his race, sexuality, or ethnicity.

Identity Politics A practice whereby an individual or group pursues an agenda based on their identity (e.g., sexuality, gender, ethnicity) rather than on political party or socioeconomic class.

Multiculturalism The recognition that members of a society exhibit a diverse range of personal and group identities, and advocacy that these should be respected without prejudice.

Nationalism The political belief that a nation has a right to self-determination, governance, and independent territory.

Nativism The political belief that the interests of the native population should be primary and take priority over those of immigrants.

Roots to Routes An understanding that identities and cultures are not unchanging and deterministic (roots) but are multiple, fluid, and continuously being remade as people travel through life (routes).

The concepts of heritage and identity have come to the fore in the current era of globalization, a period which has seen the growth of identity politics, multiculturalism, immigrant and diaspora populations, and significant increases in the number of heritage tourist sites. Although much scholarship on these topics focuses on examples from Western society, this is a burgeoning global phenomenon with many people in postcolonial societies now reinterpreting their pasts and articulating their identities. These processes are complicating traditional notions of the relationship between nation and identity that often simplistically imagined ethnically homogeneous national populations and assumed that an ethnic or national group occupied a distinct territory.

Introducing Heritage

The term 'heritage' gained new resonance in the late twentieth century. It has become a crucial element in the construction of personal and group identities as people seek to locate themselves within specific historical trajectories and understandings of the past. Heritage is most commonly understood to mean that which is inherited from one's ancestors and is most often invoked on a national scale, typically referring to the presentation and articulation of a national past in the present. Other uses of the term include references to agricultural heritage, industrial heritage, and architectural heritage, demonstrating the breadth of the concept and its application. Often nostalgic, historically partial, and simplistically rendered, many invocations of heritage are necessarily spatial, appealing to cultural and historical geographies to engender authenticity and legitimacy.

In general, heritage is understood as something to be protected, preserved, and celebrated. Although many assertions of heritage assume that such practices are both beneficial and politically neutral, conflicts over the meaning of heritage and what or who should and should not be commemorated demonstrate that heritages are contested and often one sided depictions of the past. In the late twentieth and early twenty first centuries, appeals to heritage have increasingly become a focus for political organizations seeking to advance the fortunes of specific ethnic and cultural groups. Through these processes, heritage becomes an important aspect of individual and group identities, having definitional power and providing a political focus for garnering support. These post–World War II incarnations of heritage and identity are concomitant with changing population patterns that have resulted from events such as the decline of the European empires, the collapse of the Soviet Union, and the advancement of civil rights in the United States. The arrival of Afro Caribbean immigrants in the United Kingdom in significant numbers after 1948 ushered in a new era of postcolonial migration in Europe with people moving from former colonies to the colonial core. Similar processes were observed in locations such as France and the Netherlands. Coinciding with economic changes and processes of industrial restructuring which decreased working class employment in traditional manufacturing jobs, the residence of non white immigrants in Europe helped to generate a reassessment of the meaning of white, national, and European heritages and identities. One reaction was the rise of far right political parties such as the National Front in France and the British National Party (BNP) in the UK whose leaders claimed to represent an authentic national heritage and identity, one that was considered to be under threat from demographic and economic change. In the United States, the emergence of identity politics based on gender, race, and sexuality in the 1960s, most notably the African American struggle for civil rights, saw the concept of an assimilative melting pot challenged by a doctrine of multiculturalism. Simultaneously, white people in the southern United States opposed to racial desegregation began to use the Confederate flag as a 'heritage symbol' as part of their efforts to reassert white hegemony in the region.

By the 1970s and 1980s two trends could be observed that belied traditional notions of nation and identity. First, greater public display of and awareness about the multiple identities that people exhibit, as evidenced at ethnic festivals, in gay pride parades, and by immigrants' rights organizations and through antidiscrimination legislation. Second, a reaction against these trends and their accommodation through campaigns and policies promoting a single national language, mandating traditional school curricula and celebrations of a singular national identity. Some reactions were and are violent, for example, race riots, gay bashing, the intimidation of immigrant groups, and hate crimes.

As these demographic changes challenged extant meanings of nations and identities, some locations became heritage sites; sacred places revered by people as repositories of a group's essential identity and understood to be critical to the past and future of a certain population. Such heritage sites both construct and are constructed by collective memories and identities. Although these occurrences are most commonly seen at the national scale, local and global invocations of heritage and identity at certain iconic locations are also common. On a global scale, many places have gained international and institutional recognition through the United Nations Educational, Scientific and Cultural Organization (UNESCO) and its establishment of world heritage sites. Under UNESCO guidelines established in 1972 both cultural heritage (buildings, landscapes, and historic sites) and natural heritage (areas of unique fauna and flora) sites are identified as of importance to everyone and requiring protection and preservation. UNESCO world heritage sites include locations such as Australia's Great Barrier Reef, areas of architectural innovation including Brasilia in Brazil and the old and new towns of central Edinburgh in Scotland, and places of archaeological interest like ancient Troy in modern day Turkey and the Egyptian pyramids.

At a national scale, museums, memorials, monuments, historic buildings, and tourist sites such as castles, palaces, and battlefields often act as warehouses and exemplars of national heritage and identity, places where we learn about ourselves from particular rubrics regarding the national past. This has often resulted in museums presenting 'happy history' exhibits, positive and partial accounts of the past that forward nationalist versions of heritage. These exclusive narratives omit uncomfortable events and avoid competing interpretations of history. When museum displays challenge revered national myths, curators and exhibits can be vilified as betraying the national heritage. In the United States, for example, public controversy erupted in 1995 when veterans' groups complained that a Smithsonian Institute World War II exhibition had been stripped of its patriotic message as a result of external pressure from Japanese Americans and peace activists. Elsewhere, museum curators in the American South who have devised exhibitions that contest the Anglo Southern nationalist myth of the nineteenth century Confederacy or have publicly criticized the apologetic Lost Cause interpretation of the 1861–65 US Civil War have faced pressure and intimidation from regional heritage groups such as the United Daughters of the Confederacy (UDC) and the Sons of Confederate Veterans (SCV) to revise their museum displays and interpretations to present a pro Confederate version. George Ewert, Director of the Museum of Mobile (Alabama), was nearly dismissed from his position by the city council in 2003 after writing a negative review of the feature film *Gods and Generals*, a movie that presented a pro Confederate nationalist narrative of US history. People such as Ewert are often denounced on neo Confederate websites, message boards, and in group publications as 'heritage violators' for their efforts to challenge sacred myths of Confederate heritage.

Memorials and monuments are among the most studied of heritage landscapes, due in no small part to the important symbolic and political meanings associated with these sites. By erecting a public monument, groups seek to permanently instill their own particular version of the past in the cultural landscape, and thus embed their sense of identity and heritage as the authoritative version. The erection of monuments and memorials began during the classical period in Western Europe where Greeks and Romans built ornate edifices to celebrate victories and pay respect to the Gods that granted them success. Monument building enjoyed a resurgence following the Renaissance, as newly formed nations used monuments as powerful public tools to establish their political legitimacy. Monuments became national semiotic markers, symbolically justifying new nationalist discourses with seemingly permanent reminders of a supposedly real and eternal national past.

National monuments nearly always present a highly militaristic version of heritage and identity, most typically honoring heroic male figures from the past. Soldiers or political leaders are commonly depicted as national archetypes and these usually male figures often stand alongside females who figuratively represent the nation and its virtues. Real women and less martial versions of masculinity are rendered invisible and silenced by such memorial landscapes. Monuments dedicated to the anonymous male soldier, for example, began to appear in the United States in the aftermath of the Civil War and in Europe following World War I. These monuments were often funded by subscriptions organized by local women's or veterans' groups and dedicated by these heritage organizations. In the US, for example, the UDC erected hundreds of statues to the Confederate dead between 1894 and 1914, peaking with around 35 unveilings in 1910 alone. These local efforts to commemorate the heroic sacrifice of soldiers present future generations with a model of heroic masculinity and nationalist politics to emulate, as nearly every Southern town features one of these statues outside their courthouse.

Another commemorative effort that has become a way to entrench as particular version of heritage in a cultural landscape is through street naming. In the United States, many cities renamed streets after the civil rights leader Martin Luther King Jr. following his 1968 assassination. In Eastern Europe, Soviet era street names have also been changed and often replaced with monikers that invoke a national heritage that refers to the pre-communist period. The issue of whose heritage is commemorated by street naming is often controversial. In 2006 the City of Chicago stalled an effort to name a section of road in honor of Black Panthers leader Fred Hampton (1948–69) yet the city's landscape still contains Balbo Drive, formerly Seventh Street, which was renamed in the 1930s to honor the Italian fascist leader Italo Balbo (1896–1940). Another controversial street name in Chicago is Pulaski Road. A source of controversy between the 1930s and 1950s, Polish Americans lobbied to change the name of Crawford Avenue to Pulaski in part to celebrate their heritage and honor a Polish hero of the American Revolution, Casimir Pulaski (1747–79). Other residents did not want to be associated with Poland, believing that such an ethnic identity for the street would damage local business and the image of the city.

In the United States since the mid 1970s, recognition of distinct ethnic heritages has largely replaced the assimilationist idea of the melting pot. Magazines, festivals, catalogs, and days of celebration have become commonplace among different ethnic groups in the United States as a way to celebrate heritage. Such assertions of heritage are one manner through which members of the diverse US population publicly announce the presence of their ethnic identities. The 1990s, for example, saw an upsurge in interest in Scottish heritage and ancestry in the United States. Stimulated by movies such as *Braveheart* (1995) and political support from prominent senators such as Trent Lott, National Tartan Day (6 April) was established in the United States in 1998 as a date to celebrate Scottish identity and heritage among the Scottish American diaspora. During the decade, around 100 new Scottish heritage festivals were inaugurated in the United States and numerous Scottish-themed magazines, retailers, and Internet websites catered to an audience keen to engage with its Scottish heritage. In Scotland, these developments were treated with a degree of disdain and derision by commentators who understood American invocations of Scottish heritage to be replete with nostalgia for a mythical and somewhat kitsch old country, stereotypically bedecked in tartan and awash in purple blooming heather. These developments echoed similar celebrations of Irish heritage which in the United States culminate in an annual party on St. Patrick's Day (17 March) which is frequently an occasion for drunken revelry, often to the frustration of more recent Irish émigrés who lament that their Irish identities are overwhelmed by what they consider inappropriate depictions of Irish heritage by Irish Americans. Yet, such festivities of Irish heritage in the United States are not without contention among the celebrants themselves. In 1994 in south Boston, an area of Irish American population, the St. Patrick's Day parade was canceled due to a dispute between the organizers and a group of gay and lesbian participants. In nearby New York City, the Ancient Order of Hibernians has repeatedly denied gay and lesbian participation in their city parade. After the US Supreme Court unanimously ruled in 1995 that the free speech of parade organizers would be impinged upon by the presence of gay and lesbian participants, the Irish heritage on display at both St. Patrick's Day parades has been defined as much by sexuality as national heritage or ethnic identity.

Heritage and National Identity

Heritage and identity are closely connected to places, landscapes, and collective memories. These connections can occur at all scales, although they are most frequently seen at the national level. National heritage is often framed as a story of progress and struggle replete with patriotic heroes and epic battles fought for freedom and territorial sovereignty. Accounts of national heritage that can be used to press political claims to power and territory are crucial components in the development of nationalism. As individuals and communities seek respite from what can seem the chaos of the globalizing, postcolonial world, many do so by looking to the past, imagining a simpler time of secure and homogeneous national identities. Accompanying this vision is an understanding that national identities can be stable and unified, rooted in a specific place, and having consistent characteristics. Rural landscapes are often revered as resonant with the authentic meanings of a nation and national identity. Battlefields, for example, are heritage sites where national invocations of blood and soil are particularly powerful. They become sacred spaces, landscapes for future generations to visit in order to learn of the sacrifices made by past generations and to inculcate future potential soldiers into the importance of fighting for nation, heritage, and identity. Battlefields themselves are monuments of a type, landscapes that serve to honor the fallen that paradoxically provide salient lessons regarding the cost of warfare and its importance for the sustained protection of national values and homeland territory. Many nations have established heritage sites at locations where a nation has seemingly been forged through warfare, for example, the Israeli site of Masada which commemorates the mass suicide of Jews when surrounded by the Romans in AD 73 or the Culloden battlefield in Scotland where British government troops defeated the Jacobites in 1746. In the United States, civil war battlefields like Gettysburg and Shiloh are remembered and represented to visitors as bloody sites where Americans from both armies fought gallantly in a heroic struggle that ultimately covered all participants with honor and valor. The preservation of battlefield sites in the United States is a particularly contentious issue as local real estate developers and heritage groups often clash when past battlefields become threatened by modern development plans.

Perhaps the most infamous battlefield site in recent history is Kosovo Polje. In Yugoslavia in the 1980s and 1990s Serbian nationalists envisioned a past in which the region of Kosovo was central to the definition of the Serbian nation. The Kosovo area had been part of a medieval Serbian kingdom before occupation by Ottomans. Twentieth century Serbian nationalists argued that this region was integral to Serbian heritage, territory, and identity. To emphasize this connection, in 1989 future Serbian leader Slobodan Milosevic held a mass rally on the Kosovo Polje site of a 1389 battle between Serbs and Ottomans. It was at this battle that invading Ottoman forces defeated the army of the Serbian King Lazar. This loss ensured Ottoman hegemony in the region for over 500 years and it brought about great political and cultural change to the region, most notably the introduction of Islam. Milosevic's speech was designed to appeal to the affective national identity of the Serbian Orthodox community, and he called for the Serbian people to be reunified in order to restore their glorious national heritage after centuries of outside domination. During that same period, Croat nationalist leader Franjo Tudjman made similar appeals for Croats to reclaim their cultural heritage, an effort that included the state invention of new 'Croatian' words as part of an effort to claim a cultural and linguistic distinction from their Serb and Bosniak neighbors. Ultimately, the use of heritage to fan the flames of ethnonationalism in the region served to underpin the genocidal wars fought in the former Yugoslavia during the 1990s, a conflict which left the country politically divided and the Bosniak community devastated.

Historical descriptions of events at heritage sites such as battlefields vary considerably in their accuracy and partisanship. Heritage sites usually present 'official history', leading to omissions and forgetting of particular past events which do not fit in the dominant versions of memory. In each of the above cases, visitors to the battlefield sites are encouraged to imagine the sacrifices made by the mythical heroes of the past. Heritage travelers may also seek to visit memory sites out of a desire to learn about the past at the place where it happened, as the 'sacred soil' of a particular memory site affords it a particular normative authority and power for tourists to gaze upon and soak in at the places where 'history really happened'.

Appeals to a distinctive and often monolithic national heritage have been utilized by conservative political leaders to argue against current processes of migration and the establishment of multicultural diaspora populations, particularly in the West. The political recognition of multiculturalism has led to a change in how both immigrant and native populations envision their genealogical ancestries, heritages, and identities. Communities have often seen a resurgence of fundamentalism and belief in the immutability of ethnic identity, in this sense understood to be genetic and inherited rather than constructed and performed. This identity is often linked to discredited biological notions, such as those once pertaining to race, in contrast to the perception that identities are in a continuous process of formation, socially constructed, and able to be manipulated based on context and expediency. These polarized notions of identity often lead to very different beliefs about nationality, heritage, and ethnicity. For some, identity is rooted in ancient ancestries and cannot change and thus an immigrant and her or his descendents can never be fully a part of the nation in which they reside. For others, identity is flexible and people can simultaneously celebrate their ethnicity and ascribe to new national and cultural circumstances. These competing imaginations of national identity and heritage are often understood to be 'culture wars' in which challenges to often simplistic renderings are decried as disturbing long held and cherished traditional versions of heritage.

In reality, many of the most cherished heritage traditions are 'invented', devised in the last 200 years as a way of indoctrinating citizens into highly specific, sanitized, and sanctified versions of the past. Since the start of the twentieth century, television programs and feature films have increasingly played a significant role in the establishment of normative versions of national heritage and identity. Neo Confederate forms of nationalism in the American South have, for example, been engendered by Hollywood renderings of the American Civil War. Sweeping romantic epics such as *Gone with the Wind* (1939) imagined the South as an idyllic fantasy world of moonlight and magnolias, hoop skirts and ballroom dances, while conveniently excluding the harsh realities of African American slavery and racism of the period. In India, the Hindu nationalist Bharatiya Janata Party (BJP), once a small and rather insignificant political party, rose to national prominence following a revival of Hindu nationalism which was in part sparked by the airing of a serialized version of the Hindu epic *Ramayana* on Indian state television in 1987.

Lastly, educational institutions are crucial sites for the dissemination of normative versions of national heritage and identity to children. Elementary and secondary school textbooks often explain history in an overly sim plistic fashion, focusing on heroic stories and leaders at the expense of more diverse narratives. In Northern Ireland, a majority of students continue to attend primary and secondary school in religiously segregated settings. As such, these young students may not learn about 'The Troubles' and the history of violence and segregation in Northern Ireland from any perspective outside of their own religious group. Students in many American schools begin their school day with the Pledge of Allegiance, an oath first written in 1892 by a Baptist minister as a way of marking the 400th anniversary of Christopher Columbus' landing in the Western Hemisphere. The Pledge is a ritualistic promise of allegiance by each school child to the nation, a daily affirmation of nationalism, and a re minder of past sacrifices. Following pressure from the Catholic group The Knights of Columbus, the words 'under God' were added to the Pledge in 1951 as part of an effort to separate the United States from their 'God less' communist enemies.

Heritage Dissonance and Contested Identities

The past is always open to a variety of interpretations which means it is rarely possible to contain all possible understandings within the dominant versions of heritage. These are most often forwarded by government agencies or historical societies at specific sites and as a result, certain groups of people and their pasts are omitted from heritage narratives. In addition, current processes of globalization are increasingly producing hybrid and diasporic heritages as people seek to explore the possi bilities of multiple identities that transcend borders and stand in contradistinction to narrow conceptualizations of monolithic national identities and heritage. Groups marginalized by or excluded from powerful narratives about national and local history due to their race, class, gender, sexuality, or other identities, may offer competing interpretations or countermemories that delegitimize and complicate accounts of a glorious and progressive past. The issue thus becomes whose heritage is worthy of commemoration. When official recognition is commonly accompanied by financial and other resources which enable specific heritages and identities to be preserved and promoted, marginalized groups often struggle to assert their own heritages and identities, be this in pol itics or through monument building.

Many national memorial and heritage landscapes are distinctively masculine, and in the United States and Europe they are often overwhelmingly white, ignoring the contributions to national society of non white groups or denigrating non whites as inferior. Similar processes en sure that bourgeois hegemony is sustained and working class heritages and identities remain marginalized and the sexual identities of those who do not conform to hetero sexual norms are obscured. To challenge these narratives, groups as diverse as New York City gay rights activists, Welsh coal miners, and indigenous people in Australia have erected their own museums and monuments and engaged in their own heritage projects. In doing so, they reveal histories of resistance or struggle that are excluded by dominant nationalist versions of heritage.

The trend of marginalized groups reclaiming their heritage narratives has also enabled white people to press their claims for recognition of their ancestries and identities. In many instances this is largely benign, such as the growing interest in Scottish ancestry among North Americans and Irish bars in the United States, but other examples have seen white Americans and Europeans calling for recognition of their heritage, using the language of multiculturalism and respect for diversity, while pursuing a nativist agenda that positions whites as a nation's true citizens. For example, late 1990s and early 2000s neo Confederates in the USA have claimed that people of Anglo Celtic ancestry are the guardians of a 'true' Southern culture, one developed by their white ancestors since the slave holdings days of the eighteenth and nineteenth centuries. This supposedly authentic Celtic heritage, one that easily conflates in neo Confederate literature with a celebration of white southern nationalism, is understood by advocates to be under attack by federal authorities and other groups whose actions have promulgated the end to legalized racial segregation and continue to challenge the cele bration of the slave holding Confederacy as expressed through monuments, memorials and, most potently, dis play of the Confederate Battle Flag. Aware of the po tential to be criticized as racists and white supremacists, many neo Confederates in heritage organizations such as the Sons of Confederate Veterans have adopted the slogan "Heritage Not Hate" to articulate that their celebration is one of ancestry and genealogy rather than the white supremacist political beliefs of the nineteenth century Confederacy. Such assertions were thin when proponents launch vitriolic campaigns to reassert their Confederate heritage, for example, denouncing those who proposed adding a statue of African American tennis star and humanitarian Arthur Ashe to Monument Avenue in Richmond (Virginia) in the mid 1990s, a road on which previously only white, Confederate leaders had been commemorated.

See also: Citizenship; Cultural Politics; Difference/Politics of Difference; Emigration; Empire; Ethnic Conflict; Ethnicity; Ethnicity and Resistance, Historical Geographies of; Globalization, Economic; Heritage; Heritage and Culture; Identity Politics; Immigration I; Immigration II; Imperialistic Geographies; Lowenthal, D.; Memorials and Monuments; Memory; Multiculturalism; Nation; National Spatialities; Nationalism; Other/ Otherness; Race; Sovereignty; Transnational Ethnic Networks; Whiteness.

Further Reading

Anderson, B. (1991). *Imagined communities: Reflections on the origin and spread of nationalism* (rev. edn). London: Verso.
Billig, M. (1995). *Banal nationalism*. London: Sage.
Gilroy, P. (1987). *'There ain't no black in the Union Jack': The cultural politics of race and nation*. London: Hutchinson.
Graham, B., Ashworth, G. J. and Tunbridge, J. E. (2000). *A geography of heritage: Power, culture and economy*. London: Hodder Arnold.
Foucault, M. (1977). *Language, counter memory, practice: Selected essays and interviews*, Bouchard, D.F. and Simon, S. (trans.), Ithaca, NY: Cornell University Press.
Hall, S. (1991). The local and the global: Globalization and ethnicity, and, Old and new identities, old and new ethnicities. In King, A. D. (ed.) *Culture, globalization and the world system: Contemporary conditions for the representation of identity*, pp 19 39 and 40 68. Basingstoke: Macmillan.
Hague, E., Giordano, B. and Sebesta, E. H. (2005). Whiteness, multiculturalism and nationalist appropriation of Celtic culture: The case of the League of the South and the Lega Nord. *Cultural Geographies* 12, 151 173.
Hewison, R. (1987). *The heritage industry: Britain in a climate of decline*. London: Methuen.
Hobsbawm, E. and Ranger, T. (eds.) (1983). *The invention of tradition*. New York: Cambridge University Press.
Ingram, G. B., Bouthilette, A. M. and Retter, Y. (eds.) (1997). *Queers in space: Communities, public places, sites of resistance*. Seattle, PNW: Bay Press.
Lowenthal, D. (1985). *The Past is a foreign country*. Cambridge: Cambridge University Press.
Schein, R. H. (ed.) (2006). *Landscape and race in the United States*. New York: Routledge.
Seligman, A. I. (2001). *The street formerly known as Crawford*, pp 36 51. *Chicago History Magazine* Spring.
Southern Poverty Law Center (2004). Confederates in the Museum. *Intelligence Report* 116, 32 39.
Winberry, J. J. (1983). 'Lest we forget': The confederate monument and the Southern town landscape. *Southeastern Geographer* 23, 107 121.

Relevant Websites

http://www.splcenter.org
Southern Poverty Law Center.
http://whc.unesco.org
UNESCO World Heritage and Identity Site.
http://whc.unesco.org
United Nations Educational, Scientific and Cultural Organization: World Heritage site.

Heteronormativity

D. Bell, University of Leeds, Leeds, UK

Glossary

Cruising The practice of giving and receiving often fleeting indications of sexual attraction between passers-by, using nonverbal communication, such as eye contact or posing. Cruising has a long history in gay male culture, and is widely practiced in public space as well as in commercial venues. There have been debates about whether similar practices between lesbians, and between heterosexual people, can also be regarded as cruising.

Quest Romance A common narrative form utilized in literary texts and other popular culture forms, such as films and pop songs, the quest romance has a story arc focused on the (usually ultimately successful) search for (heterosexual) love, often using fantasy themes, emphasizing the centrality of romantic (and sexual) fulfillment as a powerful cultural norm.

Introduction

Heteronormativity is a portmanteau word which borrows the 'hetero ' from 'heterosexuality' and bolts this onto 'normativity', derived from 'normative', meaning the establishment of a norm, that is, a standard or type. The term is used in discussions of sexuality, in gender studies and queer studies, where it describes the ways in which social institutions and forms of behavior reinforce – or indeed produce – belief that certain things are normal; in this case, obviously, that heterosexuality is the norm. This normalizing of heterosexuality implies, of course, that anything different from this is not normal, even abnormal. Heteronormativity, then, is a powerful but often unmarked set of assumptions, practices, and beliefs that constantly reinforce the normalness and naturalness of heterosexuality as the only normal, natural form of sexuality. So, to understand heteronormativity, we need first to understand heterosexuality.

Heterosexuality or, perhaps more appropriately, heterosexualities combine a multitude of behaviors, beliefs, rituals, and institutions. At one level, of course, heterosexuality is about sex: about sex between two bodies (or two people) defined as 'opposite' on the grounds of certain characteristics, such as their genital anatomy, their chromosomes, or their genders. But heterosexuality is not just sexual practice between two opposite sexed subjects. It invokes what is sometimes referred to as the 'heterosexual matrix' – and in contemporary Western societies this matrix equates gender, sex, and desire, and binds these together through rituals, assumptions, and sociospatial practices. So the male sex maps onto masculine gender, and maps onto desire for female sex and feminine gender. The condensation of this matrix is the heterosexual couple – perhaps most notably the married couple. This is a romantic, sexual, (potentially) reproductive, monogamous, lifelong union of two oppositely sexed subjects, ritualized through religious or civil ceremony and supported by countless institutions (government, religion, family, etc.) and practices (displaying affection, having children, commemorating anniversaries, etc.). So heterosexuality is at once a sexual orientation or identity – embedded socially as the only legitimate, normal, natural form – and also much more than this. Such is the weight of 'normativizing' that surrounds heterosexuality, that some feminist and queer scholars talk of 'compulsory heterosexuality'; it is the sexuality which you cannot be.

Rather than seeking to expose irrational prejudices, such as homophobia, work on heteronormativity focuses instead on the prescription of certain identities, practices, and institutions, and the concomitant prohibition of others. The term as it is currently utilized was allegedly first coined by US queer critic Michael Warner in 1991, and since then it has come to be increasingly common in discussions about the politics of sexuality – though the term is not uncontested. Part of the task of the term is to remind readers of the sheer weight of this 'normativizing', and therefore the difficulty of living a life outside of heterosexual norms. Its association with queer theory means it is often discussed in terms that draw on psychoanalysis, but it is also deployed in more sociological discussions, such as those which seek to expose the embedding and reproduction of heteronormativity in social structures and social relations. Relatedly, geographers have also explored how space and place are implicated in heteronormativity.

Spaces of Heteronormativity

Work on geographies of sexuality has developed significantly in human geography, mainly since the late 1980s and 1990s, yet the overwhelming focus of research has been on nonheterosexual identities and spaces, from early work on gay bars and gay gentrification, to work on lesbian home making or transgender activism. As we shall see further, the trajectory of this work is revealing of the heteronormativity of geography as a site of knowledge production. This key point aside, much work on geographies of sexuality has implicitly (and sometimes explicitly) discussed heteronormativity, through its focus on the spatial practices on non normative sexual cultures (most usually those lumped together using the acronym LGBT – lesbian, gay, bisexual, and transgender – though we should note non normative heterosexualities, too). This work explores how space and place are part of the heterosexual matrix in all kinds of ways, and how they are deeply implicated in heteronormativizing processes. Here the focus is on five spaces of heteronormativity: home space, work space, social space, public space, and nation space.

Home Space

'Home' as a geographical concept has been a relative newcomer, though the infrastructures and institutions of home – such as housing – have long been a core interest of geographers. More recently, home has been researched not just as bricks and mortar, but also as ideology and discourse, as metaphor and as experience: what does it mean to be at home, to feel at home? Where is home? Even if we do take 'home' and 'house' as synonymous, we can see ways in which home is a powerful space of het eronormativity, for example, through the emphasis in housing design and building on the 'family home'. 'Home' is a key site where the idea of family is produced, then home space is family space. But the idea (or ideal) of family here is very particular: in many white, Western cultures it means the nuclear family, the family made up of two opposite sexed adults, who have a romantic and sexual relationship (and are also bonded through insti tutions such as marriage, taxation, legal rights, etc.) and who, through sexual reproduction, have had children, themselves assumed to be heterosexual, who are in turn nurtured into heterosexual adulthood in the space of the home (and related spaces such as school).

Anyone who lives outside of this 'norm' can be viewed with suspicion, or pity – or at least as having 'not yet settled down'. So being at home is implicated in a distinct set of intimate behaviors, too – 'settling down', or 'living together' – and these are markers of maturity, of adult hood. The ubiquitous sociosexual story of our culture, the quest romance, is a parable of this process, em phasizing the compulsory nature of making 'home' the space of the romantic heterosexual couple or family unit. Children growing up with nonheterosexual identities or desires may therefore experience home as a site of op pression and even exclusion, as countless 'coming out' narratives attest. (At the same time, one of the ways in which homosexuality is rendered abnormal is due to its decoupling of sex from reproduction.)

The idea of home also contains within it particular notions of privacy – the space of the home is private, intimate space, in distinction to public spaces such as streets (see ahead). Home is therefore coded as the 'proper' space for certain kinds of private, intimate be havior, including sex acts (often tellingly euphemized as 'making love'). In the legislation of sexual practice, for example, sex which takes place in the home is often af forded greater protection than sex which takes place elsewhere, whether in a commercial sex venue or in public space. But this idea of what counts as private – and what counts as sex – is also overwritten by hetero normativity, for example, in the legal definition of 'priv acy' meaning that there can be no more than two people present when a private sex act is being engaged in, even in the home. None of this is to say that other ways of living are not accommodated under the idea of home, nor that other life courses are impossible. It is rather to re member the many ways that the idea of home is powerfully coded, closely bound to the idea of family, and to reiterate that home and family are arguably the most concentrated spaces of heteronormativity.

Work Space

Geographies of work have explored how issues of identity – whether individual or collective – relate to experiences of and paths to and through paid employment. Work is seen as an important site of identity formation, but also as a key institution through which norms are produced and reinforced, for example, in relation to gender identities and ideas about 'men's work' and 'women's work', or in terms of class identities and occupational hierarchies. Workplaces are also sites of heteronormativity, whether through practices of discrimination, hostile workplace cultures, or sexualized assumptions about appropriately gendered work roles (which presume that, for instance, any man entering what is coded as a 'female' occupation must be homosexual, given the matrixial equating of male effeminacy and male homosexuality).

One key and often reported response to the hetero normativity of workplaces is nondisclosure of non normative sexual identities, sometimes referred to as 'closeting'. Work functions as a 'closet space', to use Michael Brown's term, for those workers unwilling or unable to disclose non normative identities, practices, or desires, who choose (or are forced) to maintain a silence around their sexuality. Given the implicit and explicit heterosexualizing of many workplaces – heterosexualiz ing endlessly produced and reinforced through formal and informal practices, ranging from sexualized banter between co workers to the display of photos of partners and children at work stations – closeting is for many workers an unavoidable strategy, in spite of anti homophobic legislation in many countries. Stereotypical and commonly held views about 'male' and 'female' occupational sectors also reproduce work as a space of heteronormativity, such that male florists are coded as gay while women plumbers are coded as lesbian – codings which affect individual career choices, among other things. (Debates in recent years about 'gays in the military' provided a particularly vivid example of this.) The relationship between the gendering of work and heteronormativity also plays out through workplace policies around maternity and paternity leave, through dress codes, and through work based socializing, all reminding us of the bleeding between home family space and work space.

Social Space

As noted, some contexts for socializing, such as those revolving around work relationships, or indeed family relationships, may import work based or home based heteronormativity into social space: a drink after work with colleagues is ambivalently at once work and not work, and so may not afford 'de closeting' possibilities. Yet the possibility of separating work life and family life from social life might also mean that social spaces offer opportunities to act differently. Certainly, in the more recent history of gay cultures in the West, social spaces have been seen to provide certain kinds of opportunity to escape from prevailing 'heteronorms' – the development of bars, discos, clubs, commercial sex venues, and so on provides plentiful evidence of the potential for commercial social spaces as spaces apart from heteronormativity. Of course, there are costs to producing this space apart: not just the entry costs of commercial spaces, but the cost of bracketing off such spaces from all other spaces. The production of 'gay space', to use the most accessible example, does little to challenge the status of everywhere else as 'straight space'. Yet using spending power to assert the right to claim space has been a very significant strategy – though not an uncontested one – used by 'non normative' sexual cultures excluded from other kinds of space. In capitalist societies, this possibility of 'buying space to be different' has been manifest in the development of 'gay villages', notably in larger metropolitan centers.

But, as already noted, bounding one part of the city as 'gay space' has ambivalent effects. No city feels the need to designate a 'straight village', since everywhere outside the bounded gay space is unreflexively coded as heterosexualized. Social space is therefore heavily heteronormative, perhaps especially since social space acknowledges and accommodates 'sexual difference'. Selling social space to non normative sexual cultures represents a strategy of containment, too – the remaining space can be safely assumed to be straight, since all the gays are in their village together. And it also commodifies 'gayness' as a spectacle of multiculturalism, something to be displayed as a sign of the tolerant climate of a place. Tolerance is of course a key tactic of heteronormativity, because tolerance reinstates a hierarchy of difference: 'we' (the normal) tolerate 'you' (the abnormal).

Social space extends beyond bars and clubs, of course, to encompass a whole range of sites of leisure and recreation. The list is potentially endless: sports venues, holiday destinations, performance venues (cinemas, theatres, concert halls), consumption spaces of all kinds. Most of these involve a monetary transaction, and here is a key issue: consumer capitalism is also heteronormative, based on (even exploiting) the heterosexual matrix. While much has been made about the 'pink economy' – about how the spending power of gay men has made them a key target market for lifestyle commodities – the volume of discussion around this phenomenon reminds us of a much louder silence: that myriad routine consumption practices reinforce the heterosexual couple/family as the norm. The weekly 'family shop' to the supermarket, piling the trolley with 'family packs' of goods, is a good illustration of this.

Public Space

Of course, a social space like a gay village is not made up of just bars and clubs; it also includes the public spaces around those venues – the streets, alleyways, car parks, taxi ranks, etc. Public space has come under increasing scrutiny from geographers concerned with its perceived privatization by commercial interests. At the same time, there has been considerable interest in exploring what goes on in public spaces – how members of the 'public' interact there. In histories and geographies of non normative sexual cultures, this focus on public space has been echoed, for example, in discussions of the use of streets, parks, and other public spaces as sites of interpersonal intimacy, as in the (mainly) gay practice known as 'cruising'. Accounts of the generative role of cruising in the formation of gay culture emphasize that, in the telling words of historian George Chauncey, privacy could only be had in public, given the heteronormativity of nominally 'private' spaces such as the home.

Of course, public space is heavily regulated, and also subject to the informal social 'policing' that permits some acts while prohibiting others. A key concern in this regard is the possibility of displaying affection or attraction in public. Taken for granted actions like holding hands or kissing have therefore taken on political meanings for non normative sexual cultures excluded from their routine practice in public space. Queer activist groups in the 1990s, for example, staged 'kiss ins' to bring attention to the iniquities around a seemingly mundane act. Not feeling able to engage in public displays of affection due to the pressures of heteronormativity, therefore, marks public space as another potential closet space.

But public space is also political space, the site for protest, hence the kiss ins. Since the birth of the modern gay rights movements in the 1960s, heavy stress has been placed on claiming the right to occupy public space and to resist heteronormativity there. The development of Pride marches or Mardi Gras parades in many countries is the most visible and longstanding manifestation of the politicized use of public space, with the idea of 'pride' itself serving to challenge the supposed 'shame' of non normative sexuality. Again, of course, there are tensions here, as public space is routinely, mundanely heterosexualized, a fact arguably reinforced rather than undermined by any challenges. Nevertheless, this pol iticized use of public space remains a central tactic in antiheteronormative struggles.

Nation Space

The nation is embedded in and productive of hetero normativity in many ways, not least since family and nation are ideologically wedded, with the family the bedrock of the nation and the nation imagined as one big (happy) family. So, for example, defense of so called 'family values' is a key project of governments of all political orientations, since they could hardly stand as antifamily; while the state maintains a regulative role in issues around reproduction, such as assisted conception or parenting roles and responsibilities. At the national scale, too, the footloose cosmopolite can be labeled as potentially traitorous, or morally suspect, for their failure to express fidelity to nationhood – such 'promiscuity' has, in times of geopolitical tension, been ideologically con nected to sexual impropriety, for example, during the Cold War. National citizenship – with its bundle of rights and responsibilities – is also a key site of heteronormativity, as debates over the right of non heterosexuals to join the armed forces illustrate. Homo sexuality is here coded as a threat to national defense, just as the cosmopolite's lack of national(ist) allegiance equaled a threat to national security. Lastly, the rela tionship between the national and the international scale can also bring to light the heteronormativity of nation space, as seen in the contest between European states and the European Union over the legal recognition of LGBT rights.

Challenging Heteronormativity

The discussion of public space earlier raises a bigger issue, therefore: the issue of how to 'challenge' hetero normativity. Once the reality of heteronorms is ac knowledged and made visible, how best can those who dis identify with those norms effect social and political change? One key strategy is to make claims for equal rights – this has been, in fact, a popular and effective platform for challenging the inequalities that hetero normativity promotes. Arguably the most notable rights claim at the present moment surrounds the contentious issue of marriage. In a number of countries around the world, mostly those where marriage has long been constructed as the romantic, reproductive union of two opposite sexed adult subjects and also as the only le gitimate mode of adult living, the issue of 'gay marriage' has become a hot political topic, and in some countries variations of civil partnership (a 'marriage like' con tract) have been legalized. While there have been vo ciferous protests to this, from groups representing some organized religions for instance, civil partnerships are becoming recognized, bringing state defined benefits and also, in some cases at least, a degree of recognition for same sex relationships.

However, critics argue that civil partnerships reinforce heteronormativity by granting same sex couples some of the same rights as opposite sex couples, on condition that they live in the same couple formats. This leaves no room for alternative forms of intimate relationship, effectively reinstating marriage as the only legitimate bonding (even while actually denying marriage *per se*, granting only something that is 'marriage like'). The discourse of 'families of choice', used by some LGBT groups, simi larly maintains the centrality of the idea of family, even as it attempts to rewrite kinship away from biological and affinal lines. The harmonizing of normative and non normative relationship forms in this way has been discussed by critics in terms of promoting a new 'homonormativity' that replicates and therefore reproduces heteronormativity. (Of course, this leaves unquestioned any other non normative sexualities and relationship forms, such as nondyadic relationships, bisexuality, etc.)

Pitched against this new homonormativity are various forms of 'queer' sexual cultures, where queer is used to signal opposition to norms and normalizing. Rejection of heteronormativity is performed by queer subjects and groups through, among other things, political tactics which disrupt the 'naturalness' of heterosexuality, for example, by parodying it (as in forms of cross dressing). Critics of queer theory and politics argue that its sup posed transgressive quality is often little more than empty posing or postmodern posturing, yet there is a rich vein of work exploring how tactics like parody can be utilized in heteronormative spaces as a political provo cation, and this provocation has been widely used by activists to bring to visibility the many ways that het eronormativity does violence to non normative subjects. Whether such performances represent 'better progress' in sexual politics than achieving civil partnership rights is certainly a contentious question. To end, one more key site of heteronormativity is considered: the discipline of geography itself.

The Heteronormativity of Geography

Heteronormativity also frames knowledge, shaping the questions that get asked and the ideas that are taken seriously. As such, it is produced and reproduced by academics who, in their role as 'experts', have (more or less) unchallenged authority to think and talk intellectually about the world around them. But academics are not insulated from that world, and their cultures are also part of the broader cultures of which they find themselves a part. The same assumptions and norms shape the production of academic knowledge – even as academia is sometimes a key site from which challenges to norms are mounted (the development of queer theory, for instance). But such challenges are often marginal to the mainstream operations of academic disciplines, operations performed in the classroom and the common room, in the pages of text books and in conference presentations – and let us not forget that academia is a workplace, too, so the comments made earlier about heternormativity at work apply here.

Geography as a discipline, a site of knowledge production and dissemination, has wittingly and unwittingly perpetuated heteronormative knowledges, despite intellectual challenges from, among others, feminist and queer geographers. Taken for granted assumptions about the family, for example, or about gender roles and relations, have been resoundingly critiqued in some circles, but persist in more recalcitrant parts of the discipline. As noted earlier, for example, the tag 'geographies of sexuality' has, until comparatively recently, been seen to indicate studies of non normative sexualities (mainly gay and lesbian). Leaving heterosexuality unmarked, un researched, is a clear representation of the weight of heteronormativity in the discipline: heterosexuality simply isn't legible as sexuality. Recent work with a 'critical heterosexuality studies' angle is beginning to redress this – a good example is work by Mary Thomas on how teenage girls perform straight identities and spaces – but that work has yet to filter through to the heartland of geography. Working harder to expose and counter the heteronormativity of geography is a challenge that still lies ahead.

See also: Feminism/Feminist Geography; Gay Geographies; Gender, Historical Geographies of; Lesbian Geographies; Patriarchy; Queer Theory/Queer Geographies.

Further Reading

Ahmed, S. (2004). *The cultural politics of emotion*. Edinburgh: Edinburgh University Press.

Bell, D., Binnie, J., Cream, J. and Valentine, G. (1994). All hyped up and no place to go. *Gender, Place and Culture* 1, 31 47.

Binnie, J. (1997). Coming out of geography: Notes towards a queer epistemology. *Environment and Planning D: Society and Space* 15, 223 238.

Binnie, J. and Valentine, G. (1999). Geographies of sexuality A review of progress. *Progress in Human Geography* 23, 175 187.

Brown, M. (2000). *Closet space*. London: Routledge.

Chauncey, G. (1996). Privacy could only be had in public: Gay uses of the streets. In Sanders, J. (ed.) *Stud: Architectures of masculinity*, pp 224 267. New York: Princeton Architectural Press.

Cowen, D. (2004). From the American lebensraum to the American living room: Class, sexuality, and the scaled production of 'domestic' intimacy. *Environment and Planning D: Society and Space* 22, 755 771.

Kitchin, R. and Lysaght, K. (2003). Heterosexism and the geographies of everyday life in Belfast, Northern Ireland. *Environment and Planning A* 35, 489 510.

Richardson, D. (1996). Heterosexuality and social theory. In Richardson, D. (ed.) *Theorising heterosexuality*, pp 1 20. Buckingham: Open University Press.

Valentine, G. (1996). (Re)negotiating the 'heterosexual street': Lesbian productions of space. In Duncan, N. (ed.) *Body/space: Destabilizing geographies of gender and sexuality*, pp 146 155. London: Routledge.

Warner, M. (1991). Introduction: Fear of a queer planet. *Social Text* 9, 3 17.

Warner, M. (1999). *The trouble with normal: Sex, politics, and the ethics of queer life*. Cambridge, MA: Harvard University Press.

Relevant Websites

http://www.leeds.ac.uk/gender studies/index.htm
Centre for Interdisciplinary Gender Studies.

http://www.uky.edu/StudentOrgs/QueerInfo/qstudy.htm
Resources and links for queer studies.

http://www.queertheory.com
Resources and links on queer theory.

Hettner, A.

M. Hoyler, Loughborough University, Loughborough, UK

Glossary

Geographische Zeitschrift Leading German geography journal founded in 1895 by Alfred Hettner, who, as its sole editor for 40 years, exerted considerable influence on the development of the discipline through his editorial policy.

Länderkunde German term for regional geography.

Hettner, Alfred (1859–1941)

Alfred Hettner was born in Dresden on 6 August 1859 to Hermann Hettner, art and literary historian and his second wife, Anna Grahl, daughter of a Dresden art dealer (**Figure 1**). As one of ten children Hettner grew up in an economically and intellectually privileged middle class environment. He attended the Vitzthumsche Gymnasium in Dresden between 1868 and 1877, a private grammar school for the Protestant nobility and wealthy middle classes that provided him with a classical education rooted in humanist values. His early geographical education was primarily topographical and rather uninspiring, but Hettner soon developed an enthusiasm for geography and exploration through an engagement with travel writings, the attendance of scholarly lectures, and a field trip to the Ore Mountains. He took up the study of geography at the University of Halle in 1877 under Alfred Kirchhoff but left a year later to continue his studies in Bonn, disappointed by the focus on teacher training rather than research that he had experienced in Halle. After a year in Bonn, Hettner moved to the University of Strassburg in 1879 to study under the leading geographer Georg Gerland. Hettner gained his doctorate in Strassburg in 1881 with a thesis on the climate of Chile and West Patagonia and returned to Bonn in the same year to work with Ferdinand von Richthofen, the eminent geomorphologist and expert on China. However, when offered the opportunity to work as a private tutor in Colombia after his father's death in 1882, the young Hettner embarked on the first of two major journeys through South America. Between 1882 and 1884, Hettner extensively studied the morphology of the Cordilleras of Bogotá, before returning to Leipzig, the new workplace of von Richthofen. Rather than developing the material he had collected in Colombia for his habilitation thesis, Hettner undertook a detailed study of the geomorphology of Saxon Switzerland, which he completed in 1887 under the supervision of Friedrich Ratzel. A second journey to South America, from 1888 to 1890, was again devoted primarily to geomorphological observations, but Hettner's encounter with the varied landscapes of South America, and his engagement with Alexander von Humboldt's writings, led him to increasingly focus his research on the methodological foundation of the discipline, an early interest since his student days. Hettner resumed his teaching position at Leipzig on his return and was made associate professor in 1894. Only a year later, he became the founding editor of the new journal *Geographische Zeitschrift*, which he was to steer as sole editor for the next 40 years, a role that gave Hettner considerable influence over the development of the subject in Germany. His career took off quickly from then on; in 1897 Hettner accepted a position as associate professor in Tübingen before finally moving to the University of Heidelberg in 1899 as its first professor of

Figure 1 Alfred Hettner. Source: Department of Geography, University of Heidelburg.

geography, initially as associate and from 1906 as full professor.

From early on in his career, Hettner showed a keen interest in the methodological foundation of geography. He was taught by a generation of scholars who had come to geography from a variety of disciplines when the subject became institutionalized at German uni versities during the 1870s, and who lacked a unified methodological and theoretical approach. Hettner was critical of the prevalent notion of geography as general earth science that had emerged in response to Carl Ritter's earlier historical and teleological perspective. In order to move beyond the fragmentation of the general earth science approach and with the intention to de velop a credible scientific identity for the subject, Hettner argued for a chorological foundation of geog raphy and placed regional synthesis at the heart of a unified methodological framework. He distinguished between comparative regional geography (*Vergleichende Länderkunde*), the study of individual geographical fac tors across the globe, and regional geography (*Länd erkunde*), the study of the interplay of various geographical factors in a given location. Central to his conceptualization of geography as chorology was the analysis of regional differences at various scales. For Hettner, regions were analytical constructs produced by the geographer rather than pregiven real world entities. By turning geography into a science of regions, Hettner aimed to retain the unity of the subject in terms of its physical and human aspects through the identification of a common epistemological core.

Hettner attempted to apply his ideas in two major research initiatives during the 1890s, a 50 volume handbook of regional geography, and a global population map, both of which failed to generate sufficient support. From the early 1900s, he focused on his methodological writings (summarized in *Die Geographie, ihre Geschichte, ihr Wesen und ihre Methoden*, 1927) and on the production of a series of popular textbooks, all of which appeared in several editions (e.g., *Das europäische Rußland*, 1905; *Eng lands Weltherrschaft und der Krieg*, 1915; *Grundzüge der Länderkunde, Vol. 1 Europa*, 1907; *Vol. 2 Die außereuropäischen Erdteile*, 1924). In this context, Hettner's system of geog raphy, including his use of the established *länderkundliche Schema* (a template for the sequential presentation of geographical information that proceeded from the physical environment to aspects of human geography) has to be understood as a means for organizing geo graphical knowledge rather than a framework for em pirical research.

Hettner's universalistic approach to geography led him into a series of controversies with those who did not subscribe to his vision of the discipline, notably with Otto Schlüter, who rejected Hettner's all encompassing chorological geography and instead placed the visible landscape at the center of geographical research in his landscape morphology. Similarly, Hettner bitterly op posed the deductive reasoning of William Morris Davis' cycle of erosion model (arguments collated in *Die Ober flächenformen des Festlandes, ihre Untersuchung und Dar stellung*, 1921). Hettner's ideas were increasingly challenged in his later years, especially by proponents of a holistic landscape concept that gained influence in the 1920s and 1930s. He was forced into a defensive position and became intellectually isolated after the takeover of power by the National Socialist (Nazi) party. After his retirement in 1928, Hettner's long planned comparative regional geography was finally published in four volumes in the early 1930s (*Vergleichende Länderkunde*, 1933–35), but he failed to complete his general human geography (published posthumously). Alfred Hettner died in Heidelberg on 31 August 1941.

Both in Germany and internationally, Hettner's work continued to be discussed and adapted (most notably by Richard Hartshorne in the US) but a serious historio graphical evaluation of his ideas only started in the 1980s. In 1997 the University of Heidelberg established an an nual international lecture series in his memory.

See also: Anthropogeography (After Ratzel); German-Language Geography; Hartshorne, R.; Regional Geography I; Regional Geography II; Ritter, C.

Further Reading

Harvey, F. and Wardenga, U. (2006). Richard Hartshorne's adaptation of Alfred Hettner's system of geography. *Journal of Historical Geography* 32, 422 440.

Livingstone, D. N. (1992). *The geographical tradition: Episodes in the history of a contested enterprise*. Oxford: Blackwell.

Plewe, E. (1982). Alfred Hettner 1859 1941. *Geographers: Biobibliographical Studies* 6, 55 63.

Wardenga, U. (1995). *Geographie als Chorologie. Zur Genese und Struktur von Alfred Hettners Konstrukt der Geographie*. Stuttgart: Steiner.

Wardenga, U. (2000). Defining geography The structure and development of Alfred Hettner's methodological construct. In Watts, M. *Struggles over geography: Violence, freedom and development at the millennium (Hettner Lectures, 3)*, pp 113 126 Heidelberg: Department of Geography, University of Heidelberg.

Wardenga, U. (2005). Die Erde im Buch: Geographische Landerkunde um 1900. In Schroder, I. & Hohler, S. (eds.) *Welt Räume: Geschichte, Geographie und Globalisierung seit 1900*, pp 120 144. Frankfurt: Campus.

High-Tech Industry

Y. Zhou, Vassar College, New York, NY, USA

Glossary

Core Nations Typically indicates developed countries of Europe, North America, and Japan.

Newly Industrialized Economies Typically refer to some countries in East and Southeast Asia, and Latin America that experienced rapid economic growth in the last 20 years.

Peripheral Nations Indicate less-developed countries in the periphery of the global economy, such as some countries in Africa, Central America, and Central and South Asia.

Post-Fordism The term refers to the industrial landscape constituted by mostly vertically disintegrated corporations with company headquarters, marketing, and research and development functions located in the core countries and subcontracting production functions in semiperipheral or peripheral countries.

Semiperipheral Nations Indicate newly industrialized countries such as some countries in East and Southeast Asia, and Latin America.

The high tech industry has become an area of intensive geographical inquiry since the late 1970s. There is no clear academic consensus on what constitutes the high tech industry. However, the conventional definition includes those sectors that are spurred by new innovation or enabled by highly complex and rapidly changing technology. The most commonly studied high tech sectors are information technology, telecommunications, biotechnology, the medical and pharmaceutical industries, aerospace, and digital media **Table 1**. These industries are viewed as different from traditional resource dependent industries such as oil, or long standing manufacturing industries like automobile production. Contrast is also drawn with low tech industries like textile manufacturing, which depend on an ample supply of unskilled labor. Differentiation can be rather ambiguous, however, given that traditional or low tech industries are also affected by technological changes. Certain segments of these industries may show dynamism similar to high tech sectors, for instance, the changes driven by new materials in the textile industry and the use of complex electronic equipment in traditional manufacturing. Long standing industries such as auto manufacturing, while not seen as the technological frontier in developed countries, are considered high tech industries in developing countries. Generally speaking, high tech sectors are distinguished by their reliance on a highly skilled professional workforce and a strong emphasis on research and development (R&D) as the central element of corporate competitiveness. In the industrialized countries, high tech industries are among the main sources of economic growth. They also are the primary sponsors of, or fields for, technological innovation.

Background: Impetuses of Geographical Research into the High-Tech Industry

Information technology and biotechnology have only started to attract academic attention since the late 1970s, so the interest in the high tech industry as a distinct analytical category is relatively recent in economic geography. The surge of interest is closely linked to the economic restructuring process in industrialized countries characterized by the global diffusion of manufacturing activities and corporate restructuring toward a post Fordist production regime. It also reflects the discipline wide search for new theories of firm location, spatial organization, and regional development in the context of globalization.

Table 1 High-tech industrial sectors

SIC	*Industry definition*
High-tech manufacturing industries	
283	Drugs
357	Computer and office equipment
366	Communications equipment
367	Electronic components and accessories
372	Aircraft and parts
376	Guided missiles, space vehicles, and parts
381	Search, detection, navigation, guidance, aeronautical nautical systems, instruments, and equipment
382	Laboratory apparatus and analytical, optical, measuring, and controlling instruments
384	Surgical, medical, and dental instruments and supplies
High-tech service industries	
481	Telephone communications services
737	Computer programming, data processing, and other computer-related services
781	Motion picture production and allied services
871	Engineering, architectural, and surveying services
873	Research, development, and testing services

Source: DeVol, R.C. (1999). America's high tech economy: Growth, development, and risks for metropolitan areas. *Miken Institute Report*, 13 July 1999, http://www.milkeninstitute.org/publications/publications.taf?function = detail&ID = 15&cat = ResRep (accessed on Nov 2007).

Starting in the 1960s, labor intensive manufacturing sectors such as textile and low end electronics production have gradually relocated from industrialized nations to developing nations in Southeast Asia and Latin America. This global shift of the manufacturing core away from Western Europe and North America has progessively intensified and has spread to many sectors in manufacturing and services. The resultant de industrialization of North America, Europe, and (to a lesser degree) Japan has generated mounting problems in these regions. They include unemployment, wage depression, and the contraction of public services due to a long term decline in manufacturing job opportunities and tax revenues in once booming industrial regions. In the meantime, however, rapid and occasionally explosive job growth and business activity have been found in services and high tech industries that depend on market proximity and highly skilled workforces. A Miken Institute report published in 1999 found that after the recession in 1990–91, the growth in high tech sectors in the United States was four times as large as the aggregated economy. A thriving high tech industry is thus viewed as crucial for the continued prosperity and economic competitiveness of industrialized nations.

Influential theories of globalization have also provided theoretical support for this particular pattern of late capitalism industrialization. New International Division of Labor theory, for example, conceptualizes the global economy as a two tiered system. The global core, referring mostly to industrialized nations, is the source of advanced technology, organizational leadership, and the main consumer markets. In contrast, peripheral and semiperipheral nations, referring mostly to developing nations, participate by providing low cost human resources (in the form of both skilled and unskilled labor) in a system that is designed to benefit the core nations disproportionately. Other influential theories, including the world system theory of Immanuel Wallerstein and the product cycle theory of Vernon, also predict the continued specialization of high tech industries in developed regions.

Concurrent with the global shift of manufacturing industries is the corporate reorganization in advanced nations, made possible in part by technological progress. Since the 1960s, the world has seen an amazing string of advances in microelectronics, digital telecommunication, robotics, biotechnology, and others. This new technology has spurred the rise of new industries like the information technology (IT) industry. It has also transformed many existing industries through the adoption of new technologies, and has helped to reinvent organizational and production protocols of corporations in diverse sectors. The use of the Internet, for example, has touched upon almost every manufacturing and service sector. While new information and communication technologies have facilitated the spatial dispersion of corporate functions, they also have created new possibilities for convergence and concentration. In particular, some scholars argue that the Fordist model of production, characterized by vertically integrated companies, mass production, and deskilled labor, has evolved into a post Fordist regime of production with vertically disintegrated companies, small batch production, flexible specialization, programmable technology, and variable labor. Doubts have been raised, however, about how recent and prevailing these transitions are. Pollert has argued that mass production has long coexisted with flexible production and will continue to do so. Despite this argument, the prevailing wisdom in both the academic and business worlds seems to be that industrial organization in advanced economies is changing or will have to change in order to remain competitive within the global economy. Companies in the post Fordist era are striving for a lean model that emphasizes focused core competency and out sources associated activities to related companies or abroad. As a result of corporate restructuring, the economic activities remaining in developed nations are increasingly concentrated in and around high value added sectors such as R&D and marketing, as well as services that require substantial knowledge accumulation, long professional training, intimate connections to the market, and/or emphasis in innovation.

Post Fordist evolution can be found in both high and low tech industries. Labor intensive sectors such as the garment industry, have displayed some of the key characteristics of the post Fordist process such as detailed division of labor, intense interfirm networks, and a flexible labor force from the very beginning. The automobile industry, the archetype of mass production under Fordism, also adopted the just in time system that is characteristic of post Fordist organization. However, it is the high tech industries that best exemplify the post Fordist regime of production, by combining rapid technological changes with a professional labor force, extensive R&D, and flexible organization. Given both the global economic shift and the rise of post Fordist production, it is not surprising that the high tech industry has received considerable academic attention.

Beyond societal dynamics, another impetus for high tech industry studies in geography is the disciplinary search for contemporary theories of industrial location and regional development. In economic geography, the 1960s was the period of ‘quantitative revolution’. Considerable studies were performed to apply or test quantitative models such as Weber’s theory of industrial location and the Central Place Theory for retail centers. Most of the classic location theories, however, rely on transportation cost as the key variable in the analysis of geographical decisions. As the transportation cost declines precipitously with new technology, and with global diffusion of the manufacturing and service industries accelerating, it is clear that geography needs to search for a different set of variables to explain the location choices of enterprises. The development of high tech industry, with its emphasis on human resources and R&D, proved to be a rich empirical ground on which to search for strategies and rationales for location choices that were unaccounted for in the classic transportation based theories. The analysis of spatial patterns of high tech industry therefore became a springboard for new ideas and theories in economic geography under a post Fordist industrial landscape.

Initially, interests in high tech industry were mainly confined to the developed world. Since the 1980s, however, the high tech fever has spread to developing countries, especially in newly industrialized economies (NIEs). This is in part because some of the most successful NIEs such as South Korea, Singapore, and Taiwan have made significant advances in industrial and technological sectors and are now prominent players in the high tech industries. Their experiences have become textbook cases of how technological upgrade occurs in developing contexts. The NIEs also serve as inspirations for less developed nations to explore technology related paths of economic expansion. Large developing nations such as India, China, and Brazil have vast human resources and a large domestic market, and have long aspired to achieve technological independence from the Western nations. Since the 1990s, India has become a prominent player in computer software and tech services exporting to developed nations. China has experienced rapid economic development since the 1980s and is currently ranked as one of the world's largest high tech producers and markets. And Brazil has long had some of the most technologically sophisticated manufacturing industries in the developing world, such as automobile manufacturing and aerospace. Many in developing nations believe that although they may have little chance to catch up with the West in traditional industries where the West is dominant, the newer high tech industries may offer opportunities to make major leaps in technology and gain a more level playing field with developed nations. Regardless of whether this belief is true, developing nations are investing increasingly in high tech areas. As a result, although most contemporary research in high tech industry concentrates on North America, Western Europe, and/or Japan, NIEs such as Taiwan, South Korea, Singapore, India, and China also provide a growing body of empirical cases.

Locations and Spatial Patterns of High-Tech Industry

With the rise of the information and telecommunication industry, there have been popular claims that geography will become increasingly irrelevant. However, this has not proven to be the case. On the contrary, it has become clear to geographers, economists, and business observers that both 'centripetal' and 'centrifugal' forces are operating simultaneously in high tech industries, leading not only to a general diffusion at the global level, but also concentrations at the regional level. Certain regions, such as Silicon Valley in California, have become more prominent nodes of R&D and innovation in the world economy, even though globally, more regions have participated in the high tech industries. Traditional location theories are too static and restrictive to explain such patterns. Thus, new approaches have been developed regarding the spatial distribution of high tech industries.

The first approach emphasizes the presence of human resources: given that high tech industry requires a professional labor force, it can be expected that regions with good universities, corporate R&D centers, or features attractive to a highly skilled workforce would be favored locations for high tech industry. Kotkin, for example, argues that the thriving cities of today are no longer sites of traditional manufacturing activities or corporate headquarters, but are magnets for intelligence and talents. Geographical studies have documented that high tech centers tend to be close to large metropolitan areas that have both higher educational opportunities and urban amenities. Florida used the 'Bohemian Index' to measure urban creativity and found a strong positive correlation between the concentration of high tech industry and the bohemian population. Florida reasons that the bohemian concentration signals an environment open to innovation and creativity, and is thus attractive to high tech industry as a location choice.

While these studies have highlighted some important rationales for high tech concentrations, their explanatory power and capacity for prediction are limited. It is true that successful technology centers typically have universities nearby, such as Stanford in Silicon Valley and Harvard and MIT in Boston. It is not true, however, that places with excellent universities will naturally become technological centers. Large and creative metropolitan areas do attract high tech industry. However, each location seems to attract different sets of high tech industry. In short, the presence of human resources is necessary for high tech industry but does not offer sufficient causal explanation for the spatial patterns of such industries. Other geographical processes must be taken into account.

The second approach links high tech development with the clustering effect of industrial sectors. The literature has moved beyond looking at the presence of major factors such as human resources, universities, and corporate R&D facilities to appreciating the unique regional institutional environments that foster the social, cultural, and organizational synergy underlying sustained technological dynamism. Prolific literature can be found in this area, ranging from economic analysis by Krugman, transaction cost theory by Scott, regional innovation system by Brazyk, to the innovative milieu theory by the Groupe de Recherche Europeen sur les Milieux Inno vateurs (GREMI) group. Overall, this approach em phasizes the role of highly developed interorganizational linkages, as well as the social values and institutions that facilitate the proliferation of such linkages. Exemplary scholarship in this area includes the new industrial space thesis, originally proposed by Allen J. Scott, who regards networks of "extremely malleable external linkages and labor market relations" as an efficient organizational structure that reduces transaction costs; Saxenian's comparative study of regional network based industrial systems in Silicon Valley versus vertically integrated corporations along Route 128 in Boston suggests that the decentralized interfirm networks of Silicon Valley pro vide a superior form of organization for sustaining in novation. Castells and Hall stress the importance of networks and synergy between government, R&D insti tutions, and enterprises to the success of technopoles: metropolitan areas with successful high tech sectors. Studies of knowledge based economies also stress the role of tacit knowledge in innovation. Tacit knowledge is the knowledge that cannot be easily codified and it is transmitted by training, or personal communication and experience. Given that communication of tacit know ledge depends to a great extent on the local context and intensity of interactions, innovation is likely to occur only in selected regions. Storper succinctly states that inte grated territorial economies should be redefined as "stocks of relational assets." It is these 'relational assets' that form the institutional base for regional success. In short, this second approach argues not only that high tech industries and innovation have a tendency to create spatial clusters, but also that spatial concentration may promote synergy and technological dynamism.

Some writers of the above approach depend heavily on analysis of firm and market activities, giving limited weight to public policy decisions. In contrast, the third approach toward explaining growth patterns of high tech industry stresses the role of government. This approach focuses on the development of science parks, or tech nopoles, and examines how public policies and state investment may trigger or induce high tech development. Ann Markusen, for example, emphasizes the roles of defense industry in creating high tech regions in the United States. Other studies focus on regions successfully combine universities and private ventures to provide magnets for growth within large regions. Considerable studies have been done on technopoles around the world; these studies have shown that public policy can indeed play a role in planting the seeds of high tech growth. However, they also have made it clear that public sector involvement does not always lead to the creation of successful high tech clusters.

High-Tech Industry and Regional Development

Rapid growth, high salaries, and the tendency toward clustering make high tech industry sought after sectors for anchoring regional development. This attraction holds for both developed and developing nations. Global efforts to clone California's Silicon Valley have been seen in many different settings: Silicon Alley in New York City, Silicon Glen in Scotland, Hsinchu in Taiwan, Zhongguancun in China, Bangalore in India, etc. In all of these places, governments provide various incentives for high tech development both to attract existing high tech firms and to cultivate new ones. These measures can go far beyond providing infrastructure. Enterprising local governments have fought one another to create prefer ential policies such as land and tax deals, cash incentives to attract talents and companies locations, educational grants for local universities, targeted industrial policies, tariff reduction, environmental regulation exemptions, etc. in the hope of generating favorable conditions for high tech industry. Furious interregional competition no doubt sweetens the prospects for high tech companies, but it is not clear whether these policies can achieve the desired effect of leading the development of a larger region. In less favored regions in both developed and developing nations, attempts to attract high tech com panies often fail. Those that succeed have too often amounted to attracting low tech assembly operations that take advantage of low land and labor costs but generate little spillover effect in technological learning for other enterprises in the region. In some cases, regions also suffer negative environmental consequences as a result of industrial location. Moreover, it can be argued that by channeling public funds to private high tech companies, many of which are nonlocal, governments ignore or marginalize the capital needs of local enter prises. This is very common for the developing regions. Such neglect is unfortunate in that most successful high tech regions tend to be those that are anchored by locally grown companies. External companies can certainly contribute to local growth, but they can also move away once the temporary local benefits were fully exploited or exhausted. While it is undeniable that, in selected cases, the development of high tech com panies by multinational corporations has provided role models for aspiring local firms and a fertile training ground for experienced managers and engineers, regions where this is the case tend to be those with already developed human resources and enterprising cultures that can benefit more immediately from technological transfer by multinational corporations. Other regions that do not meet these preconditions may benefit much less from, or even be negatively affected by, high tech industry.

High-Tech Industry and Internationalization of Intellectual Property Rights

In economics and economic geography, it is common for the data on patents and investment in R&D to be used to identify regions of innovation. However, the geographical literature has said little thus far on the relationship between high tech industry and intellectual property rights (IPR). In other disciplines, there has been extensive discussion of the implications of IPR and national competitiveness of the high tech sector. Over the last 20 years, we have seen a proliferation of IPR and its increasingly internationalized enforcement, represented for example, by the Trade Related Aspects of Intellectual Property Rights (TRIPS) agreement within the World Trade Organization (WTO) framework. The step up of enforcement of IPR globally has raised many social, economical, political, and ethical questions. On the one hand, IPR is designed and justified on the ground that protection of IPR will encourage innovation as knowledge creators respond to economic incentives. On the other hand, the sharp increase in IPR claims by large corporations in advanced nations can also be said to serve as a tool for technological monopoly and may hinder the global diffusion of knowledge and technology, particularly from developed to developing nations. Some scholars argue that the international IPR regime was promoted to benefit established corporate giants in the advanced nations. In the areas of biotechnology and healthcare, prominent ethical questions have also been raised about the enforcement of IPR. In some cases, biotech companies in developed nations have patented strains of species found in developing nations. As a result, farmers in developing nations who raise these species traditionally have been required to pay IPR royalties to biotech companies in rich nations. In other cases, IPR has made treatments for global epidemics such as AIDS unaffordable in poor nations. Overall, the effect of the rather recent internationalized IPR regime is profound but not well understood. It seems likely that it reinforces the technology and knowledge dominance of advanced nations over the developing nations. A balance must be achieved that will provide incentives for innovation while also encouraging the diffusion of technology to benefit the global population.

Critique of Literature on High-Tech Industry

The high tech industry has emerged as a hotbed for empirical research and theory formation in geography since the 1970s. However, the tendency toward overgeneralization in high tech studies is common and can be misleading. Many scholars argue that flexible production systems and innovative milieu are too rare so far to have broad theoretical implications. More nuanced studies of high tech industry have also pointed out that spatial dynamics in high tech sectors can be rather diverse. Kenney and Patton, for example, found that the geography of the biotechnology supporting network is far less spatially concentrated than the semiconductor industry and telecommunication equipment industry; thus the pattern of new biotechnology firms can be more dispersed than it is in other industries. In another example, high tech clusters are often thought to be more stable and sustainable than traditional industrial cities in that they have the ability to regenerate via innovation. However, recent evidence suggests that they are as vulnerable to the fluctuation of the new economy as traditional industrial cities were to the 'old economy'. In fact, high tech firms' heavy reliance on venture capital may even exacerbate the fluctuation. All of these studies suggest that careful examination of specific industries' dynamics over a longer historical time span will be necessary to test many of the celebrated theories of high tech industry.

In contrast to the prolific literature on high tech industry in the developed world, empirical studies of high tech industry in developing nations are few and far between. Scholars have suggested that insights derived from industrialized nations can also be useful for understanding technological change in developing nations. However, in the developing world context, the role of external players is extremely important, so it is critical that such studies incorporate both regional and external players.

Last but not least, discussion of high tech industry has largely been limited to the manufacturing sector, although many of its components, such as the Internet and digital media, are increasingly defined as a part of the service industry. Even within manufacturing, the activities that are the most prominent in developed nations tend to be in the areas of marketing, management, finance, and design; all of which have a strong service orientation. Miken Institute's report explicitly defined services as a part of high tech sectors. However, empirical research targeting the service sectors is still rare, especially in comparison to manufacturing industries. Geographical research thus needs to explicitly bring the service sectors into the center of high tech studies, to enable us to have a more comprehensive theorization on how the development of technology and the changes of global economy are reshaping the new geography of high tech industries.

See also: Fordism; Global Production Networks; Industrial Location; Industrial Organization; Information Technology; Innovation; Knowledge Economy; Spatial Division of Labor; Technological Change; Technology and

Regional Development; Transnationalism and Technological Transfer.

Further Reading

Bell, M. and Albu, M. (1999). Knowledge systems and technological dynamism in industrial clusters in developing nations. *World Development* 27, 1715 1734.

Benko, G. (2000). Technopoles, high tech industries and regional development: A critical review. *GeoJournal* 51, 157 167.

Braczyk, H. J., Cooke, P. and Heidenreich, M. (eds.) (1998). *Regional innovation systems*. London: UCL Press.

Cairncross, F. (1997). *The death of distance*. Boston: HBS Press.

Castells, M. and Hall, P. (1994). *Technopoles of the world: The making of 21 century industrial complexes*. New York: Routledge.

Cook, I. and Joseph, R. (2001). Rethinking Silicon Valley: New perspectives on regional development. *Prometheus* 19, 377 393.

Corptech (2000). *Corptech directory of technology companies* (15th US edn). Woburn, MN: Corporate technology information services.

Crevoisier, O. and Maillat, D. (1991). Milieu, industrial organization and territorial production system: Towards a new theory of spatial development. In Camagni, R. (ed.) *Innovation networks: Spatial perspectives*, pp 13 34. London: Belhaven Press.

DeVol, R. C. (1999). America's high tech economy: Growth, development and risks for metropolitan areas. *Miken Institute Report*, 13 July 1999, http://www.milkeninstitute.org/publications/publications.taf?function = detail&ID = 15&cat = ResRep (accessed on Nov 2007).

Ernst, D. (2006). Innovation offshoring: Asia's emerging role in global innovation networks. *East West Center Special Reports* 10.

Feldman, M. (1994). *The geography of innovation*. Amsterdam: Kluwer.

Florida, R. (2002). Bohemia and economic geography. *Journal of Economic Geography* 2, 55 69.

Foray, D. and Lundvall, B. A. (1996). The knowledge based economy: From the economics of knowledge to the learning. In OECD (ed.) *Employment and growth in the knowledge based economy*, pp 11 32. Paris: OECD.

Fromhold Eisebith, M. (2002). Regional cycles of learning: Foreign multinationals as agents of technological upgrading in less developed nations. *Environment and Planning A* 34, 2155 2173.

Gittell, R. and Sohl, J. (2005). Technology centres during the economic downturn: What have we learned? *Entrepreneurship and Regional Development* 17(4), 293 312.

Graham, S. (1998). The end of geography or the explosion of place? Conceptualizing space, place and information technology. *Progress in Human Geography* 22(2), 165 186.

Harrison, B. (1994). *Lean & mean: Why large corporations will continue to dominate the global economy*. New York: The Guilford Press.

Hart, J. A. and Kim, S. (2002). Explaining the resurgence of US competitiveness: The rise of Wintelism. *The Information Society* 18, 1 12.

Henry, N. (1995). The new industrial spaces: Locational logic of a new production era? *International Journal of Urban and Regional Research* 16(3), 375 396.

Kenney, M. and Patton, D. (2005). Entrepreneurial geographies: Support networks in three high technology industries. *Economic Geography* 81(2), 201 228.

Knox, P. J. and McCarthy, L. (2003). *The geography of the world economy*. New York: Arnold.

Koo, J. (2005). Knowledge based industry clusters: Evidenced by geographical patterns of patents in manufacturing. *Urban Studies* 42(9), 1487 1505.

Kotkin, J. (2001). *The new geography: How the digital revolution is reshaping the American landscape*. New York: Random House.

Krugman, P. (1991). *Geography and trade*. Cambridge, MA: MIT Press.

Malmberg, A. (1997). Industrial geography: Location and learning. *Progress in Human Geography* 21, 573 582.

Markusen, A., Hall, P., Campbell, S. and Deitrick, S. (1991). *The rise of the gunbelt: The military remapping of industrial America*. New York: Oxford University Press.

Martin, P. (1996). The death of geography. *Financial Times*, 22 February.

Morgan, K. (2004). The exaggerated death of geography: Learning, proximity and territorial innovation systems. *Journal of Economic Geography* 4(1), 3 21.

Morgan, K. and Nauwelaers, C. (eds.) (1999). *Regional innovation strategies: The challenge for less favoured regions*. London: Routledge.

Nonaka, I. and Takeuchi, H. (1995). *The knowledge creating company*. Oxford: Oxford University Press.

Pavitt, K. and Patel, P. (1991). Large firms in the production of the world's technology: An important case of non globalisation. *Journal of International Business Studies* 22, 1 21.

Piore, M. J. and Sabel, C. F. (1984). *The second industrial divide: Possibility for prosperity*. New York: Basic Books.

Polanyi, M. (1966). *The tacit dimension*. London: Routledge.

Pollert, A. (1999). Dismantling flexibility. In Bryson, J., Henry, N., Keeble, D. & Martin, R. (eds.) *The economic geography reader*, pp 349 356. New York: John Wiley & Sons, Ltd.

Porter, M. (1998). Clusters and the new economics of competition. *Harvard Business Review* November December, 77 90.

Preer, R. W. (1992). *The emergence of technopoles*. New York: Praeger.

Rees, J. (1991). Regional development and policy. *Progress in Human Geography* 15, 576 588.

Sayer, A. and Walker, R. (1992). *The new social economy: Reworking the division of labor*. Oxford: Blackwell.

Schmitz, H. and Nadvi, K. (1999). Clustering and industrialization: Introduction. *World Development* 27(9), 1503 1514.

Schoenberger, E. (1988). From Fordism to flexible accumulation: Technology, competitive strategies, and international location. *Environment and Planning D: Society and Space* 6, 245 262.

Scott, A. J. (1988). *New industrial spaces: Flexible production organization and regional development in North America and Western Europe*. London: Pion.

Shiva, V. (2000). *Stolen harvest: The hijacking of the global food supply*. Cambridge, MN: South end Press.

Sternberg, R. (1996). Regional growth theories and high tech regions. *International Journal of Urban and Regional Research* 20(3), 518 538.

Teece, D. (1981). The market for know how and the efficient international transfer of technology. *Annals of the American Academy of Political and Social Sciences* 458, 81 96.

Thompson, C. (1989). High technology theories and public policy. *Environment and Planning C* 7, 121 152.

Vernon, R. (1979). The product cycle hypothesis in a new international environment. *Oxford Bulletin of Economic and Statistics* 41, 255 268.

Wallerstein, I. (1974). The rise and future demise of the world capitalist system: Concepts for comparative analysis. *Comparative Studies in Society and History* 16, 390 392; 399 406.

Zhou, Y. (2005). The making of an innovative region from a centrally planned economy: Institutional evolution in Zhongguancun Science Park in Beijing. *Environment and Planning A* 37, 1113 1134.

Zhou, Y. and Xin, T. (2003). Interaction between multinational corporations and domestic firms in a high tech service cluster in Beijing. *Economic Geography* 79(2), 129 152.

Hinterland Development

J. Timár, Center for Regional Studies, Békéscsaba, Hungary
Z. Kovács, University of Szeged, Szeged, Hungary

Glossary

Exurb A loose zone lying beyond the suburbs and within the commuting zone of a metropolitan area. Exurbs are usually residential in character, presenting a diverse mixture of land uses.

Gravity Model A mathematical model to predict flows of people, commodity etc. The initial application of the model in human geography drew on the relationship formalized by Newton in his law of universal gravitation. Later the model was fitted by regression methods or exponential model. It is widely used in planning, among others, for measuring the relative attractiveness of settlements as centers.

Rural-Urban Fringe A zone of transition at the edge of a continuously built-up urban area. It is characterized by a changing mix of rural and urban land uses and often seen as a zone of potential social conflicts.

Definitions The Origin of the Term

Hinterland is a German word, originally used in colonial law that constituted part of international public law in the nineteenth and early twentieth centuries to denote such areas over which the conquering country had already been exercising power, where, however, its own laws had not yet been enforced. Hinterland usually meant inland areas behind the coastal zones which were the first to be conquered. The origin of the term is closely related to the fact that expeditions to discover remote continents set off from river estuaries; therefore, for a long time, it was used to refer to the areas, stretching from the occupied estuary to inland regions, to which the conquering country laid claim. For instance, Portugal, which, under international public law, possessed the estuary of the River Congo, laid claim to all the areas along the river even in the nineteenth century. Pursuant to the contemporary interpretation of international public law, only where a state was, in effect, able to exercise its power (via public offices, the military, etc.) could we speak of colonial territories. In contrast, areas which are although not under state control but in the sphere of the conquerors' influence were officially considered to be hinterland.

The term also reflects the fact that in the late fifteenth century, that is, the era of discoveries, the exploration and conquest of continents, islands, or peninsulas were a gradual process, often with a place, for example, a port, taken into possession and serving as a starting point. Such ports were often hubs of colonial commerce: they were where ships arrived and set sail in the course of trading with the homeland. They were where the neighboring areas, strongly dependent on the home port, started to become settled. The term was later also used to refer to relationships between settlements, hinterland came to mean rural areas which surrounded and which were economically strongly dependent on larger cities/towns.

The use of the German word hinterland gradually gained currency in non German speaking countries as well. In Northern Italy, for instance, as a relic of the Austro Hungarian Monarchy, the rural areas surrounding Milan are still referred to as hinterland. In Australia, it refers to regions a long way away from the coastal areas as opposed to the vast expanses of land inside the country (outback). In the English language literature on human geography, it means a tributary area of a port, where exports are collected and which is also the destination of imports. Hinterland is also used in a much broader sense, meaning the sphere of influence of any (mainly urban) settlement.

Several synonyms of the word hinterland in this boarder sense exist in technical literature, for example, catchment area, exurb, urban–rural fringe, and peri urban zone. Hinterlands can be interpreted at several different scales and represent a wide variety of geographic places. In addition, they are also linked to different complementary areas. Generally referred to as foreland or heartland, these strategically, economically, or militarily vital complementary areas are central regions, ports, metropolises cities, or small towns. It is no coincidence that adjectives, for example, rural, urban, or commuting, are often used to clarify (i.e., narrow) the meaning of hinterland.

Hinterlands in a Context of Sociospatial Relations

The structural and functional characteristics of hinterlands are generally described in the context of sociospatial relations. The land use pattern, settlement structure, or the distribution of population of a city's hinterland, each used frequently to characterize the peculiarities of the hinterland, is defined as different from that of the city. So much so that, for example, theories arguing for the cyclicity of urbanization differentiate between the individual phases of urbanization mostly on the basis of trends in the population change in the ring, delimited as a commuting hinterland in the functional urban region, relative to the population change in the core. The evaluation/perception in human geographical literature of the relationship between the city/town/heartland and the hinterland, embodied in production, consumption, exchange, and communication, reflects a variety of social theoretical approaches. What follows highlights a few major components of the differences between them.

Hinterlands in the Central Place Theory

The relationship between the city/town and the hinterland is a major component in central place theories. The basis for these theories is the simple realization according to which settlements develop in space in a manner that there are strong ties between them. Some settlements – central places – stand out in the settlement network because they have such functions (e.g., commerce, education, health care, and public administration) that neighboring areas do not. As neighboring settlements – constituting the hinterland – lack the institutions that perform central functions or if they do not, the standard of the functions performed is not satisfactory, these settlements depend on the central place, of which they become the supplied areas. Based on the weight of the presence of central functions, a peculiar settlement hierarchy evolves.

A classic theoretician of this issue is Christaller, who – studying the spatial location of the central places supplying their neighborhood, while performing empirical observations in southern Germany – found that, ideally, central places of differing ranks were in the centers of hexagons linked to each other. Central functions are concentrated here, in the centers of hexagons, which fill the space perfectly, while subordinate settlements, at a level lower in the settlement hierarchy, are the apexes of hexagons, constituting the hinterland of the central place.

Lösch analyzed the regularity of the spatial arrangements of various economic activities and the manner in which economic supply areas had evolved in an idealized perfectly flat space. He did research in the USA, where the agricultural settlements that surround goods manufacturing cities had a uniform and the highest possible density. Lösch concluded that there emerged six city rich and six city poor corridors or sectors in the hinterland of any given city as a result of the existence of major radial transport routes or corridors.

Both Christaller and Lösch agree that the hexagonal arrangements of market areas around a central place represent the optimal spatial organization for a single good, under the assumption of uniform densities on an unbounded plain with equal access in all directions. These theories, though they agree that the central place needs the functionally linked hinterland for its growth, assign the latter the role of a mere consumer in a closed, hierarchically structured system without external trade linkages or dialectics.

Nevertheless, these two theories featured prominently in geographical studies conducted within the framework of quantitative revolution in several countries. Christaller himself participated in the preparation of putting his theory into practice. As an employee at Himmler's Planning and Soil Office in the Nazi Germany, he worked out integration and public administration related reform plans for the areas occupied during World War II. Focusing primarily on Poland, he planned a system of inferior and superior spatial units, cities, and their respective hinterlands, which was envisaged to be an 'efficient' part of the Third Reich. The theory of the central place – though not necessarily linked to Christaller due to the above political role he had – was applied in several countries. Thus, it can be detected in the planning

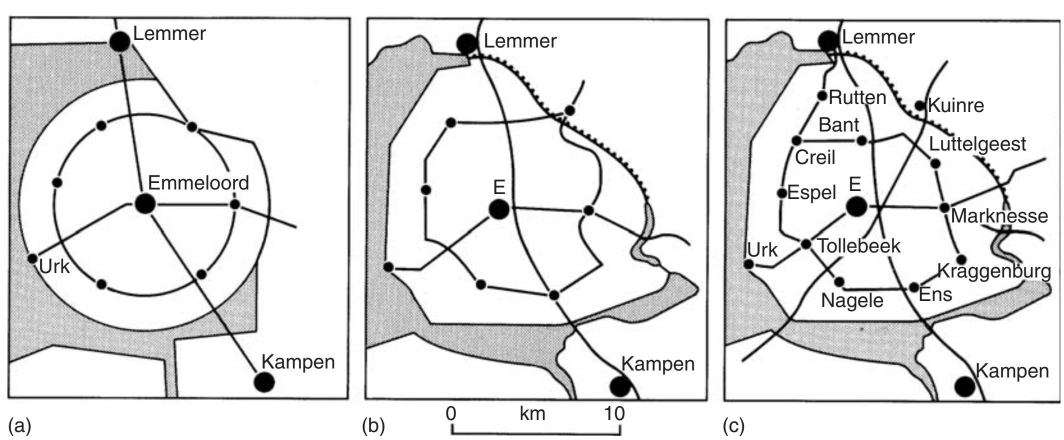

Figure 1 'Christaller' planning of settlement in the North East Polder: (a) geometrical diagram of the settlement pattern, (b) plan with five new villages, (c) the revised plan as carried out. Source: Daniel, P. and Hopkinson, M. (1989). *The Geography of Settlement.* Conceptual Framework sin Geography (W.E. Marsden, General Editor), Figure 8.3, p 272. Edinburgh: Oliver & Boyd.

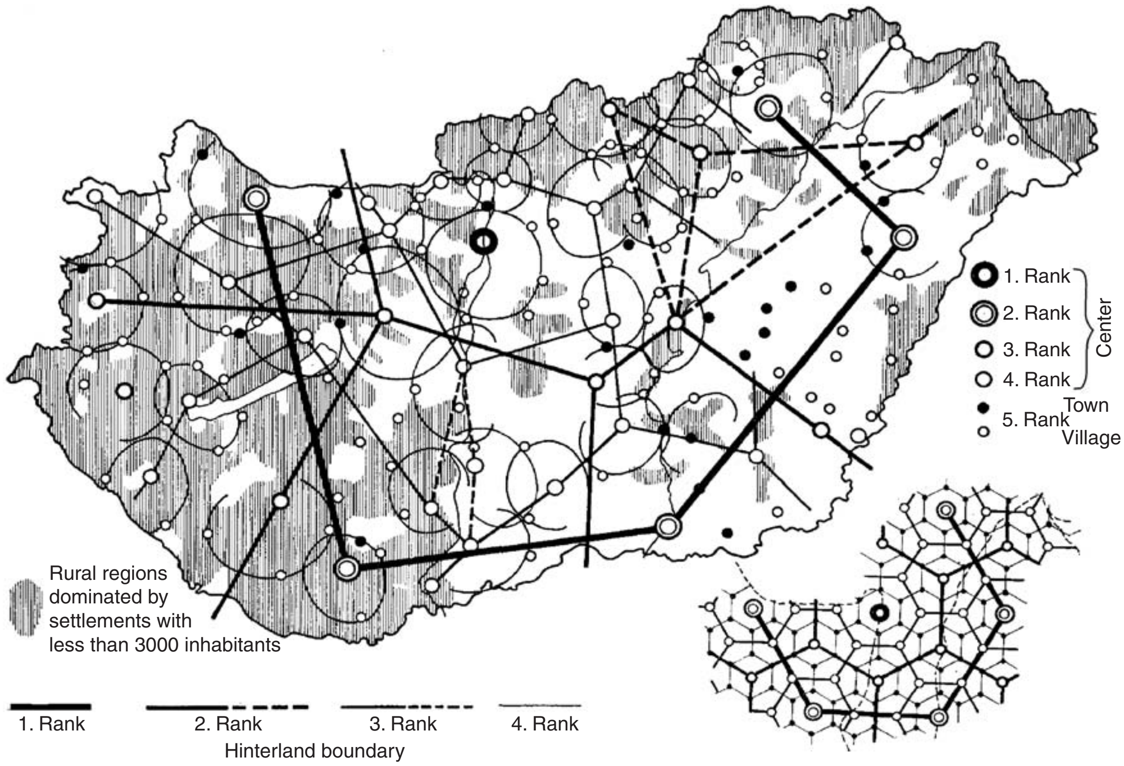

Figure 2 The hierarchical order of the urban network in Hungary. The hierarchical levels were classified according to the annual turnover of the specialist shops in 1961. At the right bottom there is the Christaller model applied to the territory of Hungary. Source: Major, J. (1964). A magyar városhálózatról. *Településtudományi Közlemények* 16, 32 65, p 57.

of settlements in the North East Polder in Holland (**Figure 1**). It was easy to adjust to the policy of centrally planned settlements and the practice of settlement de velopment in state socialist countries. For instance, in Hungary, it featured, in addition to geographical theories (**Figure 2**), in settlement development plans in as early as the late 1940s. Under the public administration reform in 1984, the 'zones of influence' of cities became official categories of spatial planning. In several countries, the strengthening of the economy and supplying institutions of the central place is still considered to be a *sine qua non* of (or even a key to) the regional development of the rural hinterland, especially in the case of hinterlands around small settlements.

'Interdependency' Approaches to Hinterland Development

In studying the relationship between the hinterland and the town/city/metropolis, several researchers focus on the interaction between them, considering mutual de pendence and support to be the essence of such rela tionship. One of the reasons for such a standpoint is an ecological approach represented, for example, by Gras's *oeuvre*, inspired by the Chicago School of Sociology in the pre World War II era. According to his analysis of eco nomic history, 'metropolitan economy' is the last stage of the development that has been going on since prehistory, during which society as a whole and individual com munities developed into more complex organisms. At that stage the urban system was characterized with high interdependence, which materialized through special ization and the division of labor between communities. Metropolises became the nodal points of basic urban functions (commerce, industry, transportation, and fi nance) which had evolved gradually. As a result, me tropolises came to dominate their regions. Gras, however, used the word 'dominance' in accordance with the eco logical concept of the Chicago School of Sociology, that is, it did not mean power relation; rather, it was used to denote coordination function. In this context metro politan dominance meant the function of organizing business for a wide metropolitan area as well as the role of an intermediary in respect of the external relations of the individual hinterland towns. That is to say, the rela tionship between a metropolitan city and its hinterland can be compared to that between the human brain and other integrated parts of the body, where inter dependence is evident.

Emphasis on mutual dependence can be attributed to empirical studies, carried out in the Third World, of the relationship between small towns and their rural hinterlands. This approach is adopted by Titus and Hinderink who – in agreement with a number of researchers – argue that the diversity and quality of the functions of small centers depend on the developments in their hinterland rather than the other way round. They claim that a fundamentally unequal rural–urban relationship can be made more equal. Toward this end, rural economy should be integrated into national and international market economy in a manner that this leads to some net increase in rural surplus production, part of which should remain in the region; in a manner that there is an emerging modern middle class which reinforces small town functions, while creating more efficient hinterland relations. Such conditions may enable rural producers to achieve certain bargaining power in both the private and the public spheres, while small towns may act as intermediaries in addressing any conflict that may arise. In an increasing competition for the purchasing power of rural areas with more efficient higher order centers, small towns 'cannot afford' extremely exploitative relations with their hinterlands.

Any rural policy relying on the foregoing and aiming at selective support for small regional centers means a fundamentally different attitude and brings different results in terms of hinterland development from an urban biased growth center policy based on the central place theory.

'Dependency' and 'Exploitation' Approaches to Hinterland Development

Several approaches perceive the development of hinterlands, interpreted at various geographical scales, as dependent on and controlled, subordinated and exploited by the heartland/metropolis/city. Some views lack strong theoretical foundation. Others are complex bodies of thoughts, such as the dependency theory, the core–periphery model, or the theory of uneven development.

In contrast to the exchange based core–periphery model, some consider the geography of production to be the engine of the increasing development gap between the core and its hinterland. According to the dependency theory, deeply rooted in Latin America, 'metropole' uses economic surplus from its satellite (hinterland) for its own development. The hinterland (i.e., in this context, the Third World) supplied the First World with primary products and low technology manufactures in exchange for high value added goods. The economic dependency that had thus evolved was also reflected in political and cultural neocolonialism.

While some claim that concentration in the metropolis of the power facilitating the exploitation of the hinterland is a universal and timeless phenomenon, others blame centralized control and bureaucratic structures, not specific to any particular political and economic order, for the exploitation of the hinterland, which is considered to be the place of a local action and self reliance. By contrast, dependence theorist Frank (supported by Marxist theoreticians of the uneven development) claims that the underdevelopment of peripheral satellites or hinterlands – similarly to the development of the metropolitan center – is generated by the development of capitalism itself. It is no coincidence, therefore, that under this approach, neither the decentralization of planning and decision making nor the reinforcement of grass roots participation at the local level, recommended by the advocates of the former view, can be considered to be 'a final solution'. Only a radical change in the existing political and economic order is likely to bring true successes.

Geography of Hinterland Development

Depending on the geographical scale, the place and the political economic and cultural context, etc., hinterlands have differing characteristics, with their relationship with the core/heartland/central place changing in space and time. Such changes led to the emergence of an increasing number of versions of the above theoretical approaches. For instance, Vance did not find the rather closed system modeled by Christaller and Lösch suitable for studying such settlement patterns where the emergence and the development of the cities and their hinterlands are driven mainly by external forces. Studying the history of US settlements, he tried adapting the Christaller model to the early (preindustrial) settlement system of the continent. Going beyond Christaller's theses, he sought to set up a new theory which was able to provide an explanation for the development of settlement patterns not only in Northern America, but also in colonies outside Europe in general as well as the evolvement of an early settlement hierarchy. Under Vance's 'mercantile' model, the settlement hierarchy in North America evolved through trading between the motherland and the colonies. The settlement network in North America evolved into a hierarchical system, with evolution starting at the ports along the Eastern Coast and settlements gradually diffusing into the inland. First, the highest levels of the settlement hierarchy evolved, then gradually – through the remaining four stages – the settlements at the lower levels of the hierarchy did, which, thus, was just the opposite of how the settlement network in Europe evolved. The towns established on the seaside in the first phase became the most important trade centers. The extension of their hinterland along the waterways (rivers, lakes, and canals) made easier and more efficient the collection of inland staples for export. In his model, Vance cites examples of a settlement network that was brought into existence by external forces (trading across the Atlantic) rather than one that came into being as a result of endemic bottom up development (**Figure 3**).

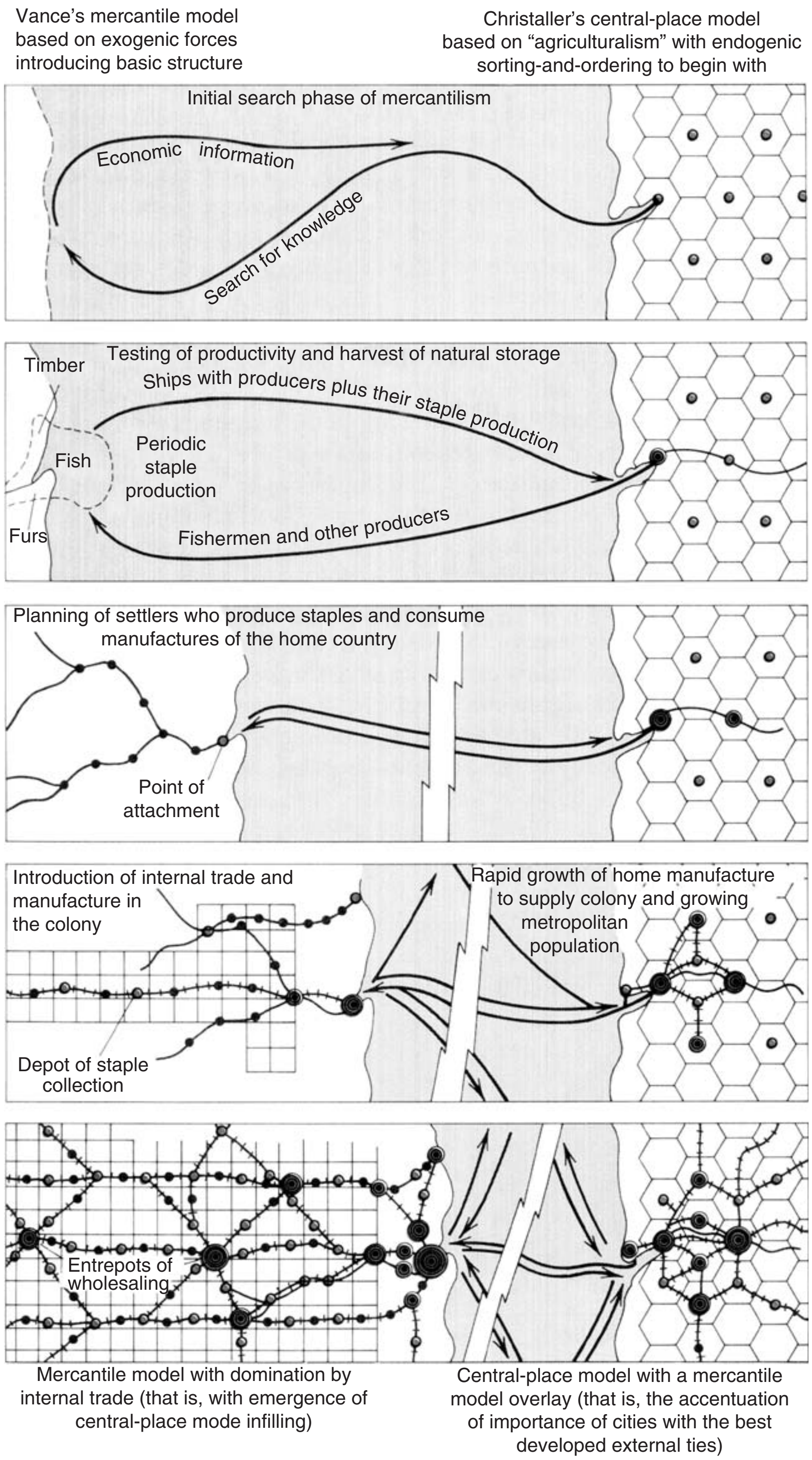

Figure 3 The Vance mercantile model of settlement evolution. Five stages in the evolution of settlement under the mercantile model (left) and the central-place model (right). The term 'exogenic' refers to external, and 'endogenic' to internal forces. The term 'entrepot' is the French word for storehouse and means a commercial center for the import and export, collection and distribution of goods. From Hagett, P. (1983). *Geography. A Modern Synthesis* (Revised 3rd edn.). HarperCollins Series in Geography. D.W. Meining Advisor., Figure 15 15, p 379. New York: HarperCollins.

Current empirical studies also highlight the importance of external relations in developing countries, though in a different context. Series of case studies attest that rural hinterland development around small towns derives its impetus and resources from macro policies and macroeconomic processes at the national level, rather than from impulses from their own local centers. In Latin America, externally induced forms of rural hinterland incorporation in national and international market economies are often related to agricultural colonization areas. Their coming into existence owes nothing to the intermediation of local service centers similarly to what the case is in the densely populated and commercialized rural areas of South and Southeast Asia where increasing marketing and employment opportunities in higher order centers and other regions – by means of improved communications and labor mobility – led to hinterland development. Although rural–urban relations at several places in sub Saharan Africa are much less developed (due to higher degrees of rural self sufficiency and lower level of productivity), research provides evidence for cases of externally induced rural hinterland development.

The rural–urban relations also lie at the heart of recent studies on hinterland development in Europe. According to some research carried out in the EU, generating development, while also exerting pressure, new demands for rural areas (i.e., for quality food production, public amenity space, space for housing, and environment protection) can be experienced. These studies claim that – after a relationship based on the consumption of agricultural produce, followed, after the era of the industrial revolution, by the dependence of rural areas on urban economies – today hinterlands are shaped by more complex rural–urban relationships which can be characterized with increasing interdependencies. The reality of some principles of the EU policies aimed at supporting such a relationship is called into question by the comparative studies conducted by Hoggart et al. in Germany, France, Spain, and England. Rather than promising hinterland functional heterogeneity and improved city region environmental sustainability, current trends point toward more homogeneity (mainly of residential functions) and increasing travel distances for work, leisure, and services. However, instead of the expected social diversity, through the 'invasion of new population', social heterogeneity and segregation prevail; a change in social composition which creates a distinctive social profile in hinterland settlements.

In addition, related to both central place and core periphery theories, the concept of hinterland is still present in regional policies across the world. It is perhaps the development pole theory, first evolved and conceived in an abstract economic space by French economist Perroux, that has gained the widest currency. Credit for transforming it into an operational tool for regional planning goes to Boudeville. This concept, linked to theories on modernization, diffusion, and innovation, was put to practice in some developed countries mainly in the 1960s and in the developing ones in the 1970s. Development was interpreted as economic growth to which innovation, technological progress, and industrialization were keys. Development (growth) poles with economies of scale and agglomeration as their leverage were expected to counterbalance congested metropolitan areas, support, the integration of backward areas, and generate beneficial spread effects to the hinterlands in their immediate vicinity. The fulfillment of the objectives was, however, challenged, due to, among other things, the rather limited beneficial impact on hinterland areas.

Critiques, Challenges

The chaotic usage of the term calls the usability of the concept of hinterland into question. Its meaning ranges from bounded spaces and separate territories to a symbolic meaning of urban influence sphere in literature on human geography. What makes orientation rather difficult is that the meaning of word 'influence' under the functional geographic approach differs from, for example, that under the cultural geographic approach. The issue of the trickiness of the hinterland concept is mainly raised by researchers conducting international empirical studies. For them, the fact that the interpretation of this term varies from one country to the next makes comparison difficult. Furthermore, very often, the hinterland is, at best, the site rather than the subject of research.

No 'comprehensive' history of research in the evolvement of hinterlands is available either. Therefore, it is especially important that critique leveled at hinterland concepts should be studied and evaluated in the given social, political, cultural, and institutional context. The hinterland concept has been and is still used in connection with the widest possible use (and even testing) of the central place theory. Though with differing intensity, the boundaries of hinterlands as discrete regions are still being drawn around cities, in the course of which the mapping of the spatial changes in these regions describes the development of hinterlands. Cities can serve as a starting point for the surveys needed for such mapping. These surveys can identify the settlements that use, for example, certain institutions in the city or the targets of supply. In contrast, rural areas can also serve as starting points for surveys that can identify the places where rural residents use urban services. For delimitation, quantitative methods and models (e.g., the gravity model) are applied. However, while the approach of locational analysis, a school of thought under the aegis of which hinterland studies are also conducted, were severely criticized from various directions (e.g., humanistic and radical geography) in the USA and Great Britain in the 1970s for its positivism, and because it was thought to serve the interest of the capital, this approach and similar studies were perceived as progressive in East Central Europe. The geographic studies that employed this approach revealed that during state socialism centrally designated catchment areas (e.g., those for health care services and schools), which were mandatory places for the provision of services for the population existed side by side with hinterlands (e.g., the retail hinterland) spontaneously evolving as a result of free choice of population. Using this realization as their starting point, many researchers managed to point out the difficulties that the differences in the various hinterlands and their location from mass transportation networks caused in people's daily life.

In contrast to the above, the studies that apply the political economy approach to hinterland development in order to understand the nature of the heartland/city/town–hinterland relationship focus on economic relations, the flow of people, products, capital, technology, and information. However, for this purpose they often find the core–periphery or the urban–rural concept more suitable. Applying the global–local framework also raised further issues. Early studies pointed out that hinterlands overlapped each other (**Figure 4**). However, under conditions of contemporary globalization in the 'world city network formation', every city's hinterland overlaps with every other city's hinterland. The service provision of world cities is of global proportions; therefore, as Taylor puts it, these cities already have 'hinterworlds'. Today the traditional concept of hinterland, he claims, can only be interpreted at the level of local catchment areas (e.g., the zones of the circulation area of local newspapers or those where rural people do their shopping in local cities).

Research focusing on power relations rather than adopting the functional approach drew attention to, for

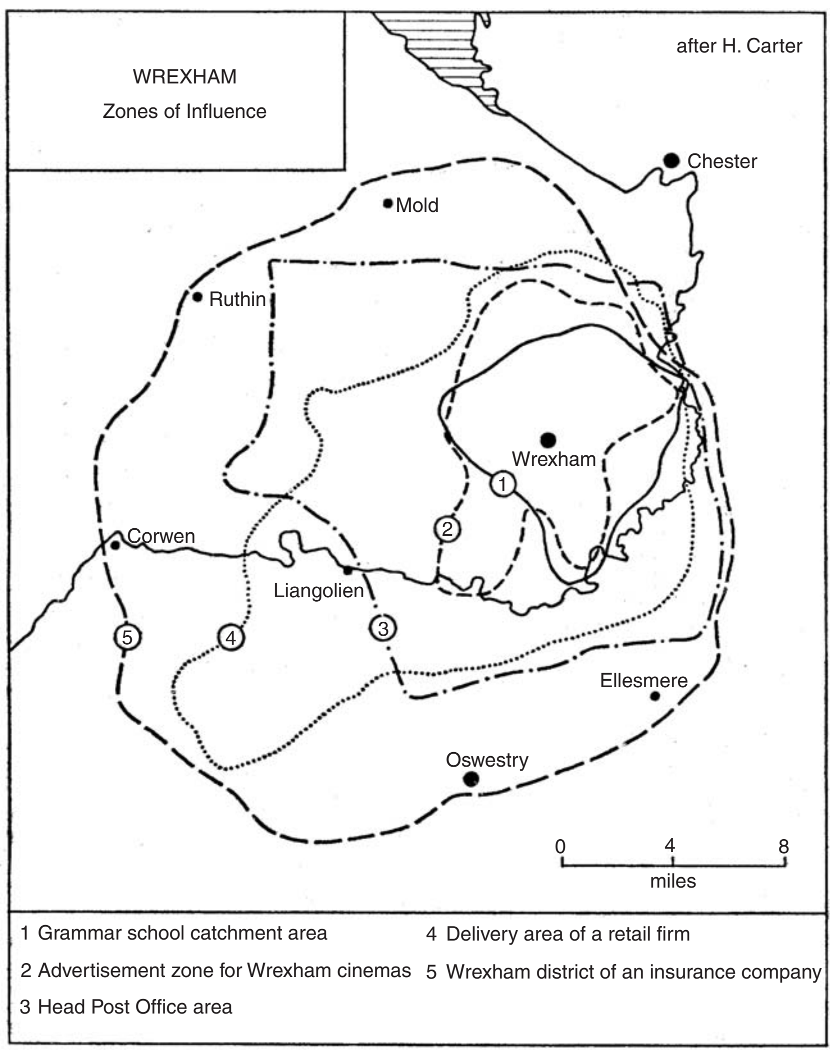

Figure 4 The zones of influence of Wrexham. Source: Johnson, J. H. (1968). *Urban Geography. An Introductory Analysis*. Fig. 22, p 82. Oxford, London, Edinburgh, New York, Toronto, Sydney, Paris, Braunschweig: Pergamon Press.

example, the fight that hinterland communities fought with administrative centers at higher levels of the hierarchy. That having said, however, advocates of this approach (like the ones of the functional concept) confine themselves to the heartland/city–hinterland dichotomy, which may easily divert attention from power relations within hinterlands and paper over social–cultural differences and inequalities. Criticism regarding certain applications of the hinterland concept in regional policies point to similar concerns. It highlights the problems arising from the top down nature of the development pole strategies, among them to the way political regimes can use these strategies to serve their economic interests to the detriment of the poorest social strata in the hinterlands and at the expense of environ mental development.

See also: Capitalism; Central Place Theory; Chicago School; Christaller, W.; Colonialism I; Colonialism II; Core-Periphery Models; Cultural Geography; Diffusion; First World; Growth Poles, Growth Centers; Humanism/Humanistic Geography; Innovation; Marxism/Marxist Geography II; Modernization Theory; Neocolonialism; Political Economy, Geographical; Positivism/Positivist Geography; Quantitative Revolution; Radical Geography; Segregation; Space II; Sustainability; Third World; Uneven Development; Urbanization.

Further Reading

Beluszky, P. (1971). Central places developing in county Szabolcs Szatmár. In Sárfalvi, B. (ed.) *The Changing Face of the Great Hungarian Plain*, pp 165 181. Budapest: Akadémiai Kiadó.

Berry, B. J. L. (1967). *Geography of Market Centers and Retail Distribution*. Englewood Cliffs, NJ: Prentice Hall.

Boudeville, J. R. (1966). *Problems of Regional Economic Planning*. Edinburgh: University Press.

Christaller, W. (1933). *Die zentralen Orte in Süddeutschland. Eine Ökonomisch Geographische Untersuchung Über die Gesetzmäßigkeit der Verbreitung und Entwicklung der Siedlungen Mit Städtischen Funktionen* (English transl. 1967, *Central Places in Southern Germany*. Englewood Cliffs, NJ: Prentice Hall). Jena: G. Fischer.

Davis, D. F. (1985). The 'metropolitan thesis' and the writing of Canadian urban history. *Urban History Review* 14(2), 95 113.

Frank, A. G. (1967). *Capitalism and Underdevelopment in Latin America*. New York: Monthly Review Press.

Gras, N. S. B. (1922). *An Introduction to Economic History*. New York: Harper and Brothers.

Hoggart, K. (ed.) (2005). *The City's Hinterland Dynamism and Divergence in Europe's Peri Urban Territories*. Aldershot: Ashgate.

Losch, A. (1944). *Die Räumliche Ordnung der Wirtschaft* (English transl. 1954, *The Economics of Location*. New Haven, CT: Yale University Press). Jena: G. Fischer.

Perroux, F. (1955). Note sur la notion de pôle de croissance. *Économie Appliquée* 7, 307 320.

Serra, M. A. (2003). Development pole theory and the Brazilian Amazon. *Estudos Econômicos* 33(1), 71 113.

Taylor, P. (2001). Urban hinterworlds. *Geography* 86(1), 51 60.

Titus, M. and Hinderink, J. (eds.) (1998). *Town and Hinterland in Developing Countries*. Amsterdam: Thela Thesis.

Vance, J. E. (1970). *The Merchant's World: The Geography of Wholesaling*. Englewood Cliffs, NJ: Prentice Hall.

Relevant Websites

http://www.espon.eu
ESPON.

Historical Geographies, Rural

P. J. Duffy, National University of Ireland, Maynooth, Maynooth, Republic of Ireland

Glossary

Geographical Inertia Resistance to change and dominance of continuity in the material characteristics of rural settlements.

Land Reform Spatial reorganization, consolidation, or redistribution of inherited farm structures, usually also associated also with migration of farm families.

Morphogenesis Early studies concerned with the historical origins of rural settlement as evidenced in the form and morphology of the settlement.

Open-Field Landscape Unenclosed agricultural lands usually associated with nucleated villages and, historically in Europe, with crop rotation by fields.

Rundale A term for traditional farming of open fields in western parts of the British Isles where individual portions of land were periodically redistributed among families of the kingroup.

Territorialization The development of a hierarchy of rural territorial units usually reflecting landholding histories, or administrative efficiency in the modern era.

Toponymy The study of place-names and landscape history.

Vernacular Buildings Erected in traditional manner using local construction materials, thus closely reflecting local environments and cultures.

Volkerwanderung Migrations of Germanic tribes following the collapse of the Roman empire.

Introduction

Much of historical geography has traditionally been concerned with rural areas and rural societies. Rural historical studies have revolved around the historical evolution of the landscape, the nature of change in rural landscapes and societies, and the role of continuity in the emergence of modern ruralities. Studies have been dominated by a Eurocentric discourse in which European rural society and environment have been the benchmarks for historical studies in other parts of the world. Until recently, the principal legacy of this scholarship has been a preoccupation with empirical geographies of settlement morphologies and material practice in land holding, agricultural husbandry, social and demographic change.

Approaches

Rural historical geographies have in the main been concerned with the facts and artifacts of land, landscape and settlement, and associated socioeconomic and cul tural contexts. Rural landscapes comprise the most ex tensive material record of humanity–environment relations. Past generations have inscribed the rural with the marks of their passage leaving geographers with a richly layered landscape legacy which has often been characterized as a palimpsest, or series of horizons of settlement extending back in time.

Historical continuity in rural landscapes and societies has been one of the traditional themes in rural geo graphical studies. They have also often been infected with a modicum of nostalgia for a disappearing, often idyllic rural, of small scale largely local communities, in opposition to more rapidly changing larger scale urban society. In a rural context, there has been significant overlap between historical and cultural geography. Cul tural geography originally had a greater historical em phasis, especially evident in the works of early cultural geographers like Carl Sauer, though cultural geography today is more focused on issues in contemporary rural societies (**Figure 1**).

Two principal approaches to rural historical geog raphies may be discerned. The earliest and most en during approach was concerned with the ‘making of the rural’, the nuts and bolts of landscape history. This largely empirical approach emphasizes the material and morphological characteristics of rural landscapes and culture and the transition from natural environments to humanized landscapes. Early settlement studies in Eur ope were concerned with morphogenesis which focused on pattern and process in the evolution of settlement forms, for example, the roles of continuity, change, and modernization, and the impact of geographical inertia in rural landscape and social morphologies. The con straining or inhibiting influence on subsequent develop ments of inherited shapes, boundaries, and structures, as well as the conservative impact of institutional structures in society and culture (such as family, kinship, and tra dition) have been important. Interrelations between culture, landscape, and place were manifested in material practices of landownership and landholding structures, farming and husbandry regimes, community and kinship patterns, migration and industrial transitions.

A more recent approach reflecting broader changes in cultural geography focuses on the ‘meaning of the rural’

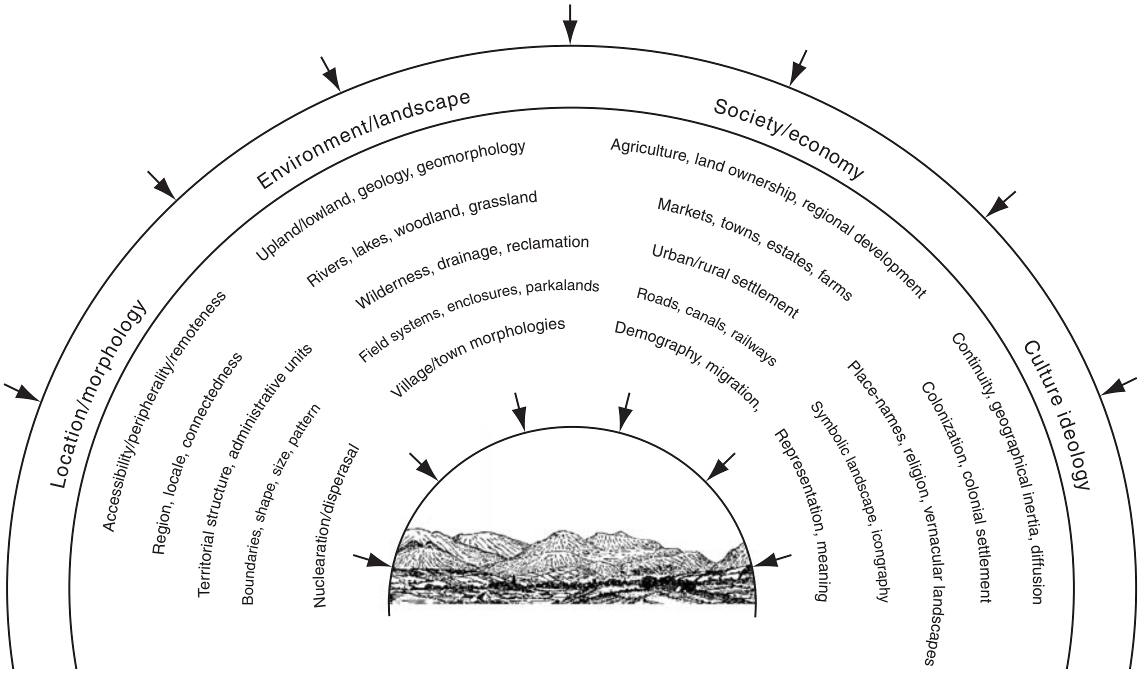

Figure 1 Scope of rural historical geography.

and how it was perceived and imagined. This sees a greater emphasis on theorization through the use of metaphor and allegory, and how we approach past societies and landscapes in terms of ideology and representation. It also engages a more questioning assessment of narratives of societal dominance of nature and the transformation of rural environments and ecologies. The most frequent application of this approach has been in former European colonized lands which have been interpreted through the prism of European traditions of scholarship seeking to examine colonial rural landscapes in terms of their similarity to the rurals of Europe. In the postcolonial world, however, such singular interpretation is being reappraised. Both approaches are interwoven throughout the following discussion.

Themes in Rural Historical Geographies

Claiming/Reclaiming the Land

Deforestation has been the great symbol of expansion of human settlement – the retreat of forest darkness before the light of (European) civilization. One of the earliest interests in rural historical geographies related to the evolution and modernization of the landscape, and anthropogenic impacts and reclamation of natural wilderness. A decidedly Eurocentric emphasis on 'taming' nature was highlighted in the removal of forest and woodlands; in reclamation of marshes and wetlands, moors and peatlands by fuel harvesting and grazing; expansion of cultivated agriculture and grasslands; enclosure of fields with hedges and fencing; and accompanying socioeconomic transitions from tribal to chiefdom, to feudal and modern state structures.

North America, New Zealand, Australia, and other colonized lands beyond Europe followed similar perspectives, except that the process was foreshortened in time. The focus of study was the more recent and rapid reclamation of savannah, forest, and bush – 'ecological shocks' which saw the 'breaking in' of wild places for commercial crop or livestock production, in themes not only of material transformation (often recorded by camera in the nineteenth century) but of colonial dispossession and erasure of indigenous cultural and landscape signatures. Some of these new world environments have more recently provided a platform for a revision in attitudes which de center society from its dominant cultural position in rural landscape history and restore parity with Nature and native societies.

Colonization and Settlement

A major concern of European rural geographies has been the reconstruction of the colonization of lowland, forest, wilderness, moors and heath as well as the abandonment of settlement landscapes from the neolithic and bronze ages. The discovery of extensive landscapes of fields and settlement under the peatlands in the west of Ireland is typical of many such features also found throughout Northern Europe. Very early settlement is most commonly marked by burial monuments which have outlived most of the other marks of human habitation.

The settlement history of rural Europe shows legacies from the Roman empire reflected in landholding tenure and landscape morphologies in southern France and the Mediterranean, and road networks and settlement sites throughout Western Europe. The *Volkerwanderung* which followed the collapse of Roman power from the fifth century initially slotted into the settlement landscapes of Roman Gaul. From the early medieval period there were waves of colonization usually by communal tribal soci eties of the wilderness in Northern and Western Euro pean lands. Scandinavians from Denmark, Sweden, and Norway moved across Northern Europe in all directions, remembered most as raiders and pillagers. While they colonized the islands and lands of the Atlantic fringe, they also expanded rural settlements within Scandinavia itself.

In the high Middle Ages population expansion drove the clearance of woodland and establishment of culti vated fields in more marginal and upland regions in a process of internal and frontier expansion: in Germanic lands eastward beyond regions of the earlier Roman empire, and westward in the British Isles by Anglo Normans and southward in the reconquest of Iberia from Moors. Planned settlements from the twelfth to the fourteenth centuries into Slav lands beyond the Oder were laid down in surveyed landscapes of regular land holding units of *hufen*, overseen by a 'locator', who managed the migrations of settlers and the clearance of woodland to accommodate *Waldhufendorf* planned vil lages, for example. Much of this colonization was undertaken under the aegis of powerful lords, or wealthy institutions such as monasteries which operated within the feudal world. The onset of bubonic plague in 1348 and subsequent crises in fourteenth century agriculture led to contraction of rural settlement throughout Europe, still marked today by many deserted medieval sites.

Much of this settlement process in the later Middle Ages set precedents for later colonization and expansion into what became the 'New World': there were parallels, for instance, with colonial settlement schemes in Ireland and America in the sixteenth and seventeenth centuries. In contrast to medieval precedents, however, newly modernizing states and agrarian and merchant capital provided the impetus and structural contexts for the early modern incorporation of new rural worlds overseas.

Traditional empirical settlement studies in general have been replaced by more recent interest in the structural/political/cultural contexts of settlement ex pansion, such as the roles of lords, kings, and states and the transition from feudalism to capitalism.

Settlement Patterns and Typologies

Historic rural settlement patterns occur in many differ ent forms, from single dispersed farmsteads, through a range of agglomerations of simple farm clusters and hamlets to much more complex nucleations around greens, churches, or castles. Village settlements reflected widely variable plan structures, which German geog raphers in the nineteenth and early twentieth centuries attempted to classify on ethnic or environmental lines. There are innumerable geometric possibilities relating to local terrain and location (such as road, canal, riverbank, or spring line settlements), political conditions, or gen esis of the settlements: colonial villages often had de fensive functions expressed in linear or circular forms (**Figure 2**). The simpler hamlet clusters which charac terized settlement in poorer more difficult agricultural environments were often associated with kinship groups, organic growth of settlements over long periods of time, as well as tribal roots of landownership in the early Middle Ages.

Buildings form, plan, and structure have also com prised a traditional aspect of rural historical settlement landscapes. As a consequence of the greater resistance to change in rural communities, vernacular settlements are characteristic. Vernacular building traditions have sur vived throughout many parts of rural Europe, for in stance, which in combination with morphological layouts of settlements lent distinctiveness to much of the rural landscape and provided the initial stimulus to French rural geography in the late nineteenth and early twen tieth centuries. Vernacular buildings mirror local culture and environment in terms of building traditions, styles, and construction materials. Rural villages in lowland England and France contain many substantial farmhouses and ancillary buildings extending back to the 1600s and earlier. The long houses and courtyard farmsteads of Northern Europe are typical examples. In the poorer margins of northwestern Europe, however, many of the traditional houses have been replaced by more modern structures.

Landscapes of Field Systems

Historically the arrangement of rural fieldscapes had an enormous impact on settlement landscape. Agrarian landscapes existed in a symbiotic relationship with the habitations of the rural settlements. While the principal characteristics of rural landscape typologies have long been defined in European terms, the geographies and socioeconomic processes underlying the management of the land are universal, though there is a considerable degree of local diversity within Europe which adds to the landscape variety (**Figure 2**).

The historic operation, management, and territorial expression of agricultural land was influenced by the nature of agriculture (in terms of the balance between crops and livestock), modes of cultivation, landholding tenures, demographic trends, and cultural practices such

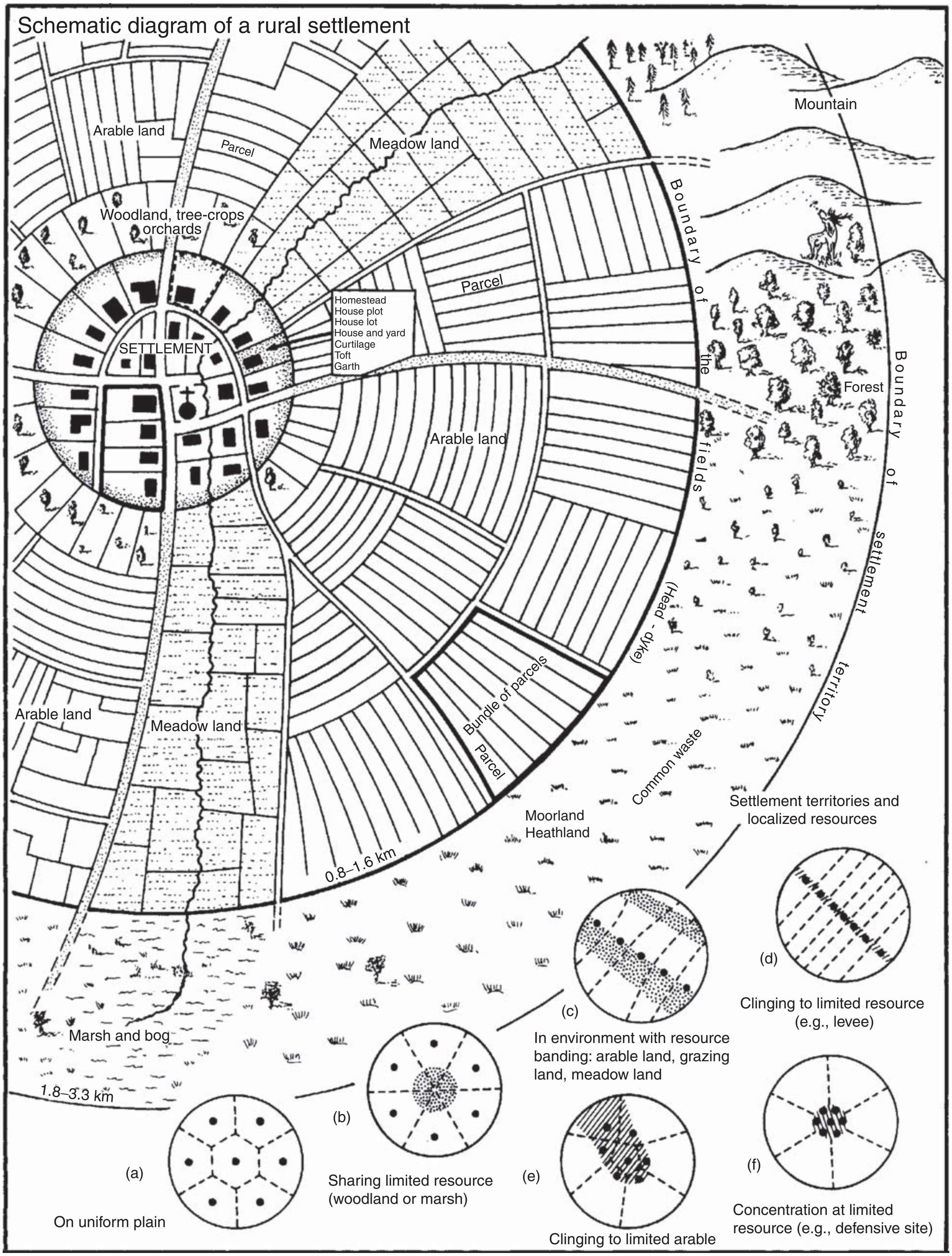

Figure 2 Schematic rural settlement map. Source: Roberts, B. K. (1996). *Landscapes of Settlement: Prehistroy to the Present.* London: Routledge.

as communal cooperation. Throughout extensive parts of medieval lowland Europe, open field agrarian landscapes were the norm, where rotations ranged from one field, through three field to four and five field rotations of crops. These represented increasingly intensive uses of land (with from half to a quarter or a fifth of the land being left fallow in any one year). 'Fields' in this sense were the lands associated with the clustered or nucleated settlements, often permanently or temporarily enclosed with a fence. In the oceanic margins of Europe where extensive tillage gave way to more pastoral agricultural regimes, a simpler one field rundale (or runrig) system was identified in which the 'infield' nearest the settlement cluster was continuously cultivated, and the 'outfield' was grazed and periodically cultivated. Apart from such open field systems, early instances of single dispersed farm steads, often in outlying parts of the community's land, on the edge of parishes or in uplands, were associated with early enclosure of the land into smaller individually worked fields. The more comprehensive 'enclosure movement' which reached its zenith in the eighteenth century, resulted in the systematic fencing of the open fields into smaller, often hedged enclosures which gives much of Britain and Ireland their distinctive patchwork quilt like landscapes today.

Open field systems were also associated with other communal strategies to manage economic resources – such as common meadow lands, marsh, fishing, forest and woodland, moor, bogland, and mountain. In the western, northern, and Mediterranean peripheral regions of Eur ope seasonal transhumance was significant, where village livestock were moved *en masse* for grazing to temporary settlements on summer pastures.

Open fields in medieval Europe were spatially allo cated to tenants or farmers in a variety of ways: in well endowed agricultural lowland soils such as the north European plain, usually in regular strips ploughed by draught animals; in more western pastoral regimes smaller cultivated irregular blocks in the infield were hand worked and subdivided as population increased. The planned colonizations in Eastern Europe usually had regular geometric land divisions arranged in large open field rotations or long linear lots extending into the forest.

Although field and settlement typologies are pre dominantly the result of European scholarship and classification systems, in many cases the historic patterns and processes at work are repeated in African and Asian rural landscapes. In many of these overseas regions, however, European colonial settlement from the early modern period resulted in the insertion of European patterns which often obliterated preexisting native structures. Because most of the colonial settlements took place in the past couple of centuries, they exhibit many of the influences of the agricultural age of improvement and modernization. Combined with a largely individualized settlement strategy in government surveyed lands, this resulted in suites of dispersed farmsteads in geometric fieldscapes, most dramatically exemplified in the west and mid west of the US and Canada.

Newer geographies have attempted to break away from the intellectual aridity of settlement classifications of fixed and immutable explanation to focus more on the ways in which rural landscapes and buildings were experienced and invested with different meanings by contemporary societies. The texts of landscape can be re read for insights into the lives and social practice of individual communities. Rural places were socially con structed, not simply regulated space and geometric order within generic models of feudal or plantation economies.

Estate Landscapes and Land Reform

Much of the morphology and practice of rural landscapes occurred within the parameters of estate landscapes. Across Europe such landscapes have their roots mainly in the medieval feudal period whereby manorial lands were held under the lord by a range of tenantry. Under standing this landownership system is fundamental to explaining many of the features of the historical geog raphies of rural settlement: the delegation of ownership to powerful local landowners resulted in some degree of homogeneity in many of the settlement structures which emerged across Europe. This landowning gentry system continued with regional modifications into the modern era but, except for Spanish America's latifundia, was never replicated in the rural landscapes of the new world.

Tenant and peasant populations and their rural settlements of buildings and fields were controlled to a greater or lesser extent by tenurial conditions. These varied from classical large estates of small fragmented peasant plots of the Mediterranean regions, which con tinued in many cases into the twentieth century, to the gentry estates of Northern Europe. These had been largely transformed by merchant and agrarian capitalism and overseas colonialism in the sixteenth and seventeenth centuries. The earliest and classic example of these changing estate landscapes was in England where im provements from the sixteenth and seventeenth centuries were the drivers for a transformation of rural society and landscape – leading ultimately to emergence of a strong yeoman farmer class and enclosure of medieval fields (**Figure 3**).

By the eighteenth century, a common culture of landowning elites throughout Europe, grounded in social institutions like the Grand Tour, displayed its power and privilege in great houses and mansions, elaborate or namental gardens and parklands, and in many cases regulated farmscapes, fieldscapes, and regimented populations of tenantry.

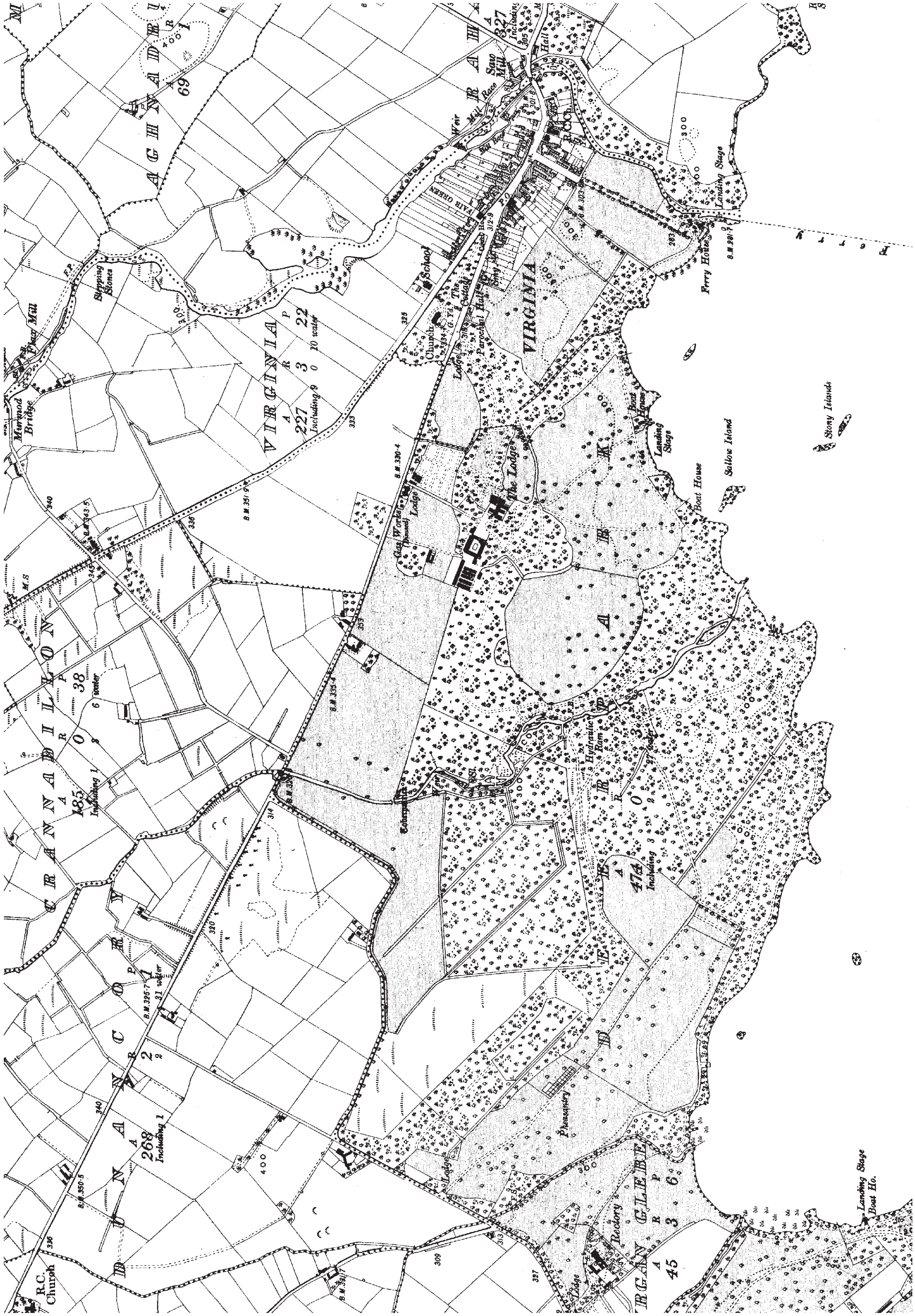

Figure 3 Estate andscape, V rg n a, re and.

The characterization of a revolution in agriculture reverberating across Northern Europe from the eight eenth century as an 'age of improvement' is particularly appropriate to discussion of rural geographical change. It might be seen as the working out of early modern cap italism in various landscapes of Northern Europe through the agency of landed estates. Much of what happened might be characterized in modern parlance as land reform. The emphasis was on 'improvement' in technical, infrastructural, agricultural, and social senses, where the processes at work in the countryside mirrored the economic transformation of the Industrial Revolution.

One major consequence was the modification of older medieval rural landscapes which became increasingly obsolete in the new commercial agrarianism. The most significant 'improvement' was the obliteration of open field countrysides, to be replaced in a couple of gener ations by carpets of rectilinear hedged enclosures. Wri ters have been important agents in the interpretation of changes taking place in rural (and urban) landscapes and society. Poets John Clare and Oliver Goldsmith com mented on the revolutionary effect in eighteenth century England of parliamentary enclosure of fields and com mons and the creation of large parks. Goldsmith's *Des erted Village* railed against the obliteration of ancient fields and communities by powerful elites in the eighteenth century: "…. If to some common's fenceless limits strayed,/ He drives his flock to pick the scanty blade,/ Those fenceless fields the sons of wealth divide,/ And even the bare worn common is denied." Demographic expansion in eighteenth and nineteenth century Europe also witnessed reclamation of wasteland and expansion of settlement into former uninhabited lands, classically re flected in the reclamation of heathlands in Denmark and north Germany, and breakup of many older village settlements to be replaced by more efficient single farm dispersal (**Figure 4**).

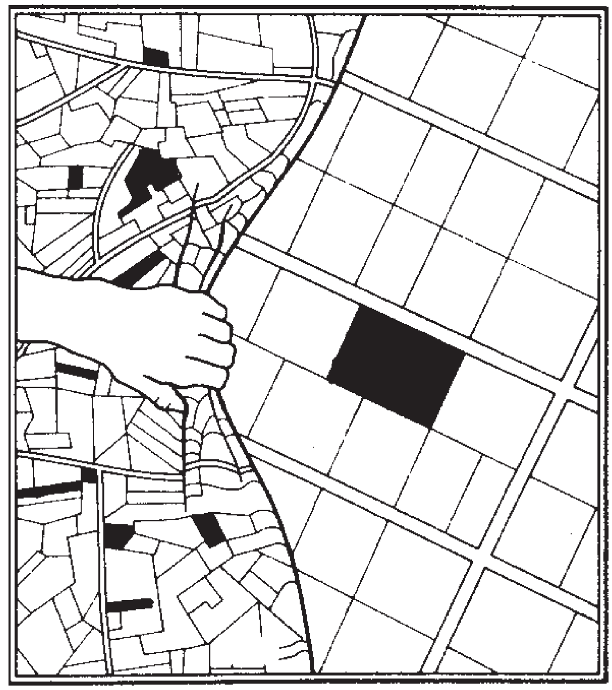

Figure 4 Land reform: replacement of the older medieval land-scape. From King, R. and Burton, S. (1983). Structural change in agriculture: The geography of land consolidation. *Progress in Human Geography* 7, 474.

Accompanying this process was the gradual elimin ation of the peasant farmer in the face of expanding agrarian capitalism, first in England, and later in the peripheral and poorer regions of Europe. The progress of the Industrial Revolution had fundamental implications for rural society and landscape. From the late eighteenth century, there was a fairly radical decline in the agri cultural labor force. In Ireland, as well as other parts of Northern Europe, voluntary and assisted emigration from rural areas in the nineteenth century was often associated with reconstruction and modernization of outmoded farmscapes.

Proto industrialization reflected in domestic crafts, textile spinning and weaving was a feature of much of Northern Europe and a precursor to urban based in dustrialization. Often encouraged by improving land owning elites, it was driven mainly by town merchants subcontracting production to the rural population. Sub sequent demographic expansion was unsustainable in the family farm matrix in the north of England and Ireland as factory made textiles undercut the rural craft economy, impelling migration overseas and to towns.

Ireland witnessed one of the earliest state inter ventions in land reform when the Congested Districts Board and the Land Commission were established in the late nineteenth century to solve a landholding crisis in western regions. Soon afterward, Ireland also experienced a radical elimination of landowning gentry and estab lishment of tenant proprietorship in one of the first state sponsored 'nationalizations' of landownership in Europe. Subsequently in the first half of the twentieth century land reform programs were undertaken in France, Italy, Greece, and Germany to faciltitate a rationalization of the legacy of antiquated landholding structures.

Boundaries and Territorialization

Territorialization and the evolution of spatial order in the landscape was a fundamental phase in the social and economic morphology of rural space. This embraced organic development of territorial structures from a very early period reflecting exigencies of local culture, society, and landholding. Administrative divisions were delib erately created from the early modern period in a process of social and political order – exemplified in the shiring of older territories in Wales and Ireland by the expanding British state in the sixteenth century, the communes and cantons of revolutionary France, or the Poor Law Unions of nineteenth century Ireland. Such territorial geographies are an important component in the vocabulary of rural settlement landscapes.

The role of historical continuity is best demonstrated by the manner in which rural landscape development was constrained by earlier boundaries and territorial morphologies. Phases of colonization and settlement throughout Europe often fitted into preexisting territorial parameters: in some cases the Germanic colonizations in Eastern Europe adapted to earlier Slav geographies; British colonizations in sixteenth and seventeenth century Ireland slotted into earlier Gaelic structures. Some farms in the midlands and southeast of England can be traced back to the eleventh century Domesday Book.

Apart from some Roman legacies, in general, these older structures grew out of medieval tribal and lordship regimes. Much of Europe was fissured with a variety of local territorial units to do with landholding and land assessment, most universally based on tribal assessments of hundreds (lands which could notionally provide one hundred soldiers): in Ireland the *tricha cét* (or thirty hundreds) was an ancient territorial designation arising out of this system which ultimately formed the basis of baronies. Local hundred courts endured in England into the Middle Ages. The most ubiquitous landholding structures were reflected, for example, in Irish townlands, Scottish and English townships (**Figure 5**). Throughout Christian Europe, parishes were the most universal administrative territories, providing pastoral care, collecting taxes from the faithful, and usually exhibiting a spatial synchronicity with local secular territories of tribes, lords, and kings.

The process of administrative control of rural space was replicated in what was perceived as a *tabula rasa* in the New World. Most of its rural territorial designations were products of the modern age, with very little continuity from preexisting native landscapes. These units were deliberately planned spatial constructions contrasting with the more organic antiquity of territorial divisions in the 'Old World', though the nomenclature often reflected European origins, in counties, parishes, and townships of the US and Canada. Rural settlement landscapes in America, Australia, and New Zealand, reflect the consequences of land surveys which rectangularized the landscape with new lines and rational geometries that were discordant with anything that went before.

Rural Cultural Geographies

Carl Sauer in the University of California, Berkeley, encapsulated the scope of culture in the totality of rural historical studies when he asserted in the 1920s that "Culture is the agent, the natural area is the medium, the cultural landscape is the result." Patterns, morphologies, and material legacies of rural settlement are frequently a manifestation of local cultural practice, custom, and tradition which in premodern societies often overrode economic or environmental constraints. Field layouts and management or redistribution of allotments were often linked to seniority of landholders in the kingroup. Rights to land shares in the commons, or fertile or sunny locations, or the disposition of houses, as well as the actual construction and size of houses, also reflected local considerations of family, kin, custom, and communal working arrangements. House construction in parts of rural Ireland or Scandinavia or the Mediterranean region, in Huron and Iroqouian communities of Ontario, or in Zulu kraals, all shared in a variety of cultural imperatives influencing their material expression in the landscape.

Place names as illuminators of rural landscape evolution are one of the earliest preoccupations of rural cultural geography. Toponymy and microtoponymy are the ultimate languages of landscape which have been used to track the expansion of rural settlement in Europe from the early medieval period. Danish settlement in England, Germanic colonizations in Central Europe, Anglo Norman settlements in Ireland, have all resulted in territorial naming patterns which commemorate the language of the colonist. In the absence of documentary records, place names can also illuminate settlement sequences and settlement processes: names with *wald* in central Europe, or *weald* in England, or *cill* in Ireland reflect previously forested lands. Names from the Middle Ages such as Brentwood, or Burntfurze, or Blackditch clearly refer to a pioneering stage in rural settlement. Ireland has many tens of thousands of place names, recorded by colonial settlement schemes from the sixteenth and seventeenth centuries, which assist in reconstruction of past rural geographies. Naming practices also played an important role in New World colonies as symbols of hegemony over native cultural landscapes.

Postcolonialism and the Rural Past

European influence and heritage have been especially strong in relation to rural historical geography – containing the oldest places and landscapes to be recorded and described but more importantly written about by historical geographers, and many of these rural landscapes were replicated in the new world – as in New England's townships, or the seigneuries of Quebec.

Much rural historical research elsewhere in the world has been strongly influenced by a Eurocentric scholarship imbued with a naive empiricism which represented rural landscape and communities as passive receptors of social

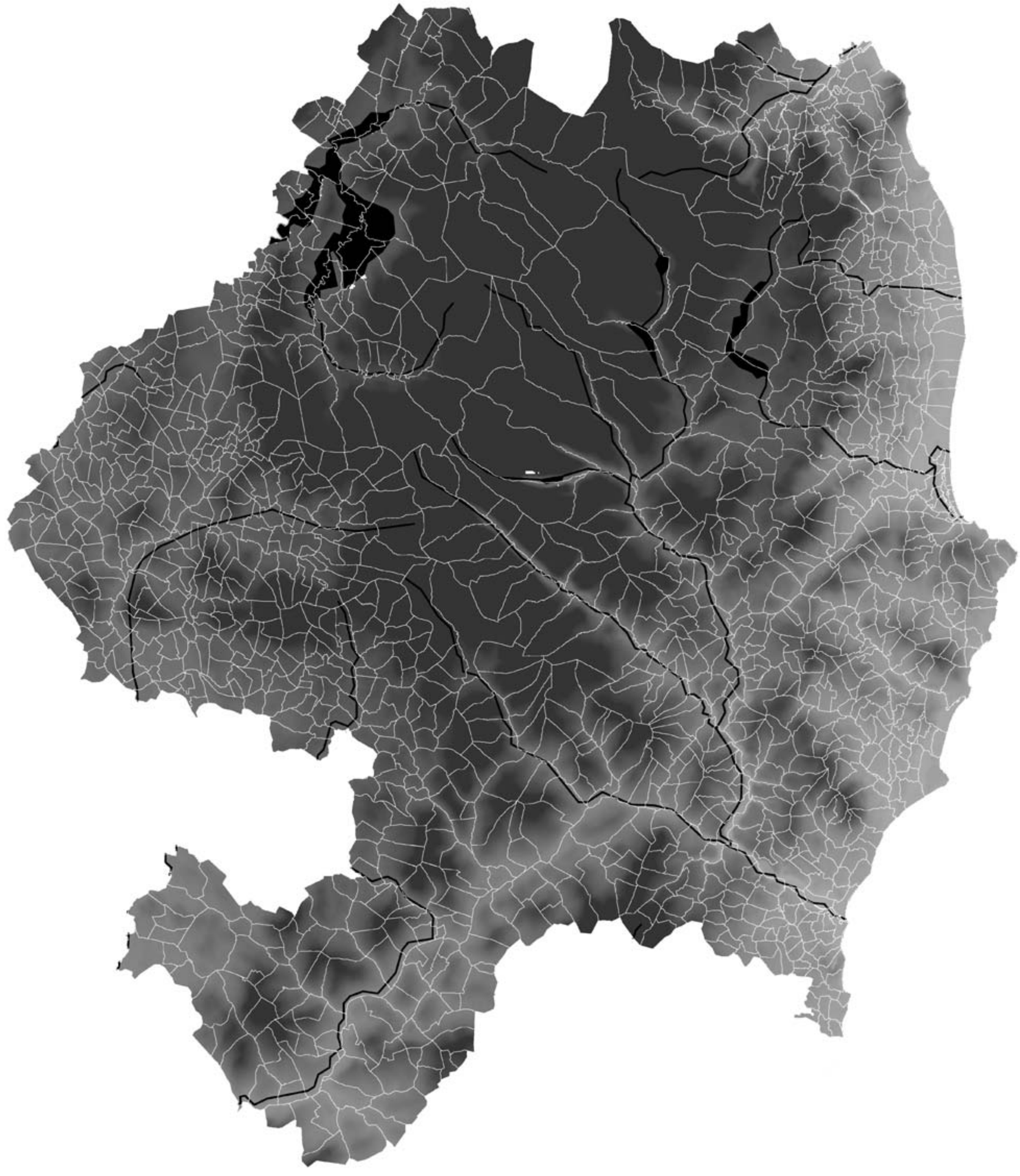

Figure 5 Townlands in county Wicklow, Ireland.

change. More recent cultural geographies of the rural have highlighted the fact that perspectives on past landscapes have been deeply embedded in relations of power and knowledge. Enclosure, one of the iconic re flections of the age of improvement in the eighteenth century was driven by agrarian capitalism to produce a transformed and regulated land. Social relations which stood in the way of modernization at home and abroad were at times ruthlessly obliterated. Studies of such landscapes cannot be described in terms of an innocent transformation but as assertions of power over nature and society – ordering, taming, even torturing a disordered nature or displaying wealth and regulating behavior. Reading past landscapes produces different perspectives on the consumption of landscape: to see and classify countryside, as in a painting (picturesque), is only possible for the privileged perspective of the well off élite: described by Donald Meinig as "the difference between looking at a landscape and living in it."

In many ways the past rurals of many colonized regions in the world have been obliterated or radically modified by colonial European settlement. Native societies were often pushed to the margins beyond settler frontiers. Previously overlooked in classic studies of rural societies, postcolonial studies have focused attention on ways in which earlier Eurocentric scholarship was complicit in the dispossession and silencing of the indigenous societies and cultures and on the material and symbolic representation of colonial power, control, and regulation on the colonized. The symbolism of the march of British/European place names across the rural landscapes of many parts of the world cannot be overlooked.

More recent studies also highlight the previously unacknowledged role of resistance by native peasantry to colonial regulation and surveillance of their lives, spaces, traditional rights, and resources – in many ways paralleling earlier resistances by European peasantry to the modernization of their lands and societies. They also focus on the role of colonial societies in the ecological destruction of environments which resulted from contested evaluations of the meaning of nature and wilderness in these overseas lands, where indigenous peoples and natures were collectively characterized as inferior, in need of 'discovery', civilization, and moral order. Some of these trends see a return to an earlier cultural geographical interest in ordinary, peasant landscapes reflected in ethnographic studies by Sauer, Estyn Evans (in Ireland), de la Blache (in France), and Scandinavian geographers working in the early twentieth century.

Our understanding of past ruralities is shaped to a great extent by extant documentary evidence which embed particular meanings in our reading. These represent records of elites in the main, the powerful in communities or civil society: bureaucrats, writers, artists, mapmakers. These minorities have shaped the historical past as we come to know it and so it is necessary to critically evaluate the record in recovering past rural geographies. Archaeological field evidence, as well as folklore, songs, and ballads, can sometimes help to illuminate the lives and landscapes of our rural predecessors. In looking back at their settlement landscapes, it is important to realize that our perspectives are also shaped not only by the scholarship of more than a century, but by technologies such as aerial photography, Geographic Information Systems (GIS), and Remote Sensing which are especially influential in rural geographies. We can bring to bear perspectives on patterns and morphologies which impose meanings and explanations that were beyond the range of knowledge of our ancestors. In this sense, therefore, the past is very much a foreign country.

See also: Cultural Geography; Evans, E. E.; Field Systems and Enclosure; Landscape; Postcolonialism/Postcolonial Geographies; Sauer, C.

Further Reading

Aalen, F. H., Whelan, K. and Stout, M. (1997). *Atlas of the Irish Rural Landscape*. Cork: Cork University Press.
Atkins, P., Simmons, I. and Roberts, B. (1998). *People, Land and Time: An Historical Introduction to the Relations between Landscape, Culture and Environment*. London: Arnold.
Butlin, R. A. (1993). *Historical Geography: Through the Gates of Space and Time*. London: Edward Arnold.
Clarke, H. B., Prunty, J. and Hennessy, M. (eds.) (2004). *Surveying Ireland's Past*. Dublin: Geography Publications.
Dodgshon, R. A. (1987). *The European Past: Social Evolution and Spatial Order*. London: Macmillan.
Dodgshon, R. A. and Butlin, R. A. (1990). *An Historical Geography of England and Wales*. London: Academic Press.
Duffy, P. J. (2007). *Exploring the History and Heritage of Irish Landscapes*. Dublin: Four Courts.
Harris, R. C. (ed.) (1987). *Historical Atlas of Canada*, vol. 1. Toronto: University of Toronto Press.
Helmfrid, S. (ed.) (1994). *Landscape and Settlements*. Stockholm: National Atlas of Sweden.
Hoskins, W. G. (1970). *The Making of the English Landscape*. London: Hodder and Stoughton.
King, R. and Burton, S. (1983). Structural change in agriculture: The geography of land consolidation. *Progress in Human Geography* 7, 474.
Muir, R. (2000). *The New Reading the Landscape: Fieldwork in Landscape History*. Exeter: University of Exeter Press.
Proudfoot, L. J. and Roche, M. (eds.) (2005). *(Dis)placing Empire: Renegotiating British Colonial Geographies*. Aldershot: Ashgate.
Roberts, B. K. (ed.) (1983). *Villages, Fields and Frontiers: Studies in European Settlement in the Medieval and Early Modern Periods*. Oxford: BAR.
Roberts, B. K. (1996). *Landscapes of Settlement: Prehistory to the Present*. London: Routledge.
Smith, C. T. (1978). *An Historical Geography of Western Europe before 1800*. London: Longman.

Relevant Websites

http://www.suffolk.gov.uk
http://www.enfo.ie
Rural Settlement in Medieval Ireland.

Historical Geographies, Urban

A. Golan, University of Haifa, Haifa, Israel

Glossary

Chicago School A group of Chicago-based sociologists whose model of the ecological theory of the city had initiated during the 1920s the development of urban studies.

Cultural Turn in Geography The reemergence of cultural geography in the 1980s that marked the emergence of a postmodern geography.

Histoire totale Total history. Formed by the school of the *annale*, a post-World-War-I group of French historians that adhered to a research of all aspects, activities, and strata of past human societies.

Liberal Eclecticism in Historical Geography A point of view suggesting that no particular dogma about the nature of historical geography was to be promoted: the study of no problem, period, or place in the past was to be prohibited.

Regionalism A fundamental concept in geography regarding the core of the discipline as the study of areas in their total composition or complexity. Regional geography dominated among geographers in the second half of the nineteenth and first half of the twentieth centuries.

Synthetic Approach in Urban Historical Geography An approach intended at the integration of different concepts, theories, and methodologies into a coherent method of research of urban past geographies.

Urban History A subdiscipline in history concentrating on different social, economic, cultural, and political aspects of urban areas in the past.

Urban Morphology A concept developed by the British geographers, Michael Conzen and Jeremy Whitehand, that approached urban settlements as the outcome of an unconscious production process of built-up areas that took place over long periods.

Urban Studies An integrated discipline including researchers from all fields of social sciences that focus on the research of the urban phenomenon.

Introduction

"Urban historical geography is no longer a fanciful label on an empty shelf in the geographer's library." This was Peter Goheen's opening of a review published in 1984 on Harold Carter's book titled *An Introduction to Urban Historical Geography* that was published about a year before. Following more than five decades of modern historical–geographical research, such an assertion implies a fault of a subdiscipline that intends to study, as defined by Robin Butlin, "…the geographies of past times."

If that was true, it raises the question whether there was no research previous to the 1980s taken on the past of urban areas. Actually much has been written on different aspects of past urban areas by scholars from different academic disciplines. First among them were urban historians, as well as economic and social historians that were also involved in the research of past urban areas. While history is the default discipline for researching the past, scholars from other disciplines were also involved in the research of past urban areas such as sociologists, architects, and urban planners, as well as scholars from several subdisciplines of geography: urban geographers, cultural geographers, social geographers, and economic geographers.

Historical geographers avoided the research of urban areas for a long time. This was due to the dominance of the regionalism in geography during the time of the formation of modern historical geography in the 1920s and 1930s. The founding fathers of the subdiscipline concentrated on the research of rural areas especially of the pre Industrial Revolution British Isles, in which towns were considered as part of the regional system while their inner structure was not yet considered a subject of research.

Urban studies were inaugurated in the United States in the 1910s and 1920s. This was mostly the endeavor of sociologists whose research concentrated on the evolution and development of Midwestern American cities, especially Chicago. The growing interest in urban areas at that time followed the large scale nineteenth century Industrial Revolution that formed the modern world. Urban centers formed the hub of modernity and the urbanization process has fundamentally changed human geographies of Europe, North America, and on a smaller scale was transforming those of South America, Asia, and Africa.

Robert Park, a Chicago based sociologist, was the leading figure among the founders of modern urban studies. He and his associates endeavored to form a model of city structure, epitomized by the well known zonal diagram presenting the concentric zone model of urban development and structure presented by Ernest Burgess in his introduction to the 1925 collection of essays titled *The City*, edited by Park himself and Roderick McKenzie. Renowned as the model of the ecological theory of the city, it suggested a format for urban growth and nature based on the experience of the American Midwest. The work of the so called Chicago school and followers in the 1930s and 1940s, most of which have been trained as sociologists and economists, developed different theoretical and methodological frames and insights for the apprehension of the urban phenomenon. These effected the formation of a new and integrated scholarship of urban studies that concentrated on different aspects of urban forms: physical, social, economic, and others.

Their main interest was recent Western (i.e., European and North American) cities and towns. The question of the origins of cities seemed less significant within the research of urbanity although not totally neglected. The research of the urban past: the history of cities, towns, and urban systems, emerged in Europe. European urban tradition was much longer than that of the Americas. Ancient and medieval towns attracted the interest of late nineteenth and early twentieth century scholars, mainly historians and social scientists researching a long time span urban evolution process. Different from the American urban ecology theory, Europeans developed an institutional theory of the city based on peculiar order and historical primacy of urban institutions. Max Weber, one of the founding fathers of sociology, took the effort in his work titled *The City*, first published in 1921, to integrate theories related to different institutions into one theory. His theory of urban community intended to form an integrated frame for conceiving the development of urban areas since early historical times.

Both American and European sociologies played a major role in the formation of a research of urban areas in general and of urban history in particular. Nevertheless, sociologists' aims and methods were not historical in essence. Research of different aspects of past urban areas took place in the first decades of the twentieth century. The study of different aspects of classical cities had taken place already in the nineteenth century. Pierre Lavendan's work on history of urban planning in the classical world and Henri Pirenne's on medieval cities were published in parallel with the emergence of Park, Burgess, and McKenzie's ecological theory of the city. Yet, urban history did not form a subdiscipline until the early 1960s. Josef Konvitz regards the publication of four seminal works: Jean Gottmann's *Megalopolis*, Lewis Mumford's *The City in History*, Jane Jacobs's *The Death and Life of American Cities*, and Daniel Burstin's *The Image*, as a landmark in the formation of urban history in America. All four books proclaimed the relevance of urban history for the understanding of contemporaneous trends and planning a better future for cities. In the other side of the Atlantic it was the British historian, James Dyos, that initiated the formation of urban history as an academic subdiscipline. Dyos's approach to urban history was in the tradition of the French school of *histoire totale*. He regarded urban history's totality in terms of an integrative scholarship embracing all aspects of urbanity: political, economic, cultural, social, spatial, and demographic. The inauguration of two journals in the mid 1970s: the *Journal of Urban History* in the United States and *Urban History* in Great Britain mark the growing interest in different aspects of the past of urban areas.

The research of history of urban areas was also the interest of scholars from other fields such as urban sociologists, economic historians and geographers, urban geographers, historians of architecture, and historians of town planning. The effort to form an integrated scholarship such as urban history attracted much criticism from other scholars interested in the urban past that ranged from defining urban historians as no more than a clique based on Dyos and disciples to more evident criticism such as the lack of use of theory, over fragmentation, and being no more than a form of social history. Such intellectual debates had been persistent as geographers turned in the 1960s and 1970s to the research of the history of urban areas.

Urban Historical Geography, the Beginnings

In his seminal work on historical geography Robin Butlin detects the origins of historical geographical study to the early eighteenth century. Since then and until the 1950s, including the time of formation of modern historical geography, between the 1920s and the 1950s, urban settlements did not get much attention from historical geographers. Geography in those years adhered to the regional imperative that formed the outline and frame of research. Historical geographers' main concern was regional settlement history in the regional rather than in local or global scales. The research of urban areas was left to social and economic geographers and to a growing group of urban geographers. The post World War II Quantitative Revolution and rise of applied geography during the 1950s and 1960s brought about an increasing interest in contemporaneous geographies in which the urban phenomenon dominated. Geographers studying the evolution and growth of urban areas turned to the use of theories and methods developed mainly by social scientists.

Historical geographers could not have avoided this turn in geography for long. Paul Wheatley, a student of Clifford Darby – the founding father of British modern historical geography, was among the first that turned already in the 1960s to the research of historical urban geography. His work was based on a comparative and meticulous research of the historical geography of urbanism in East and Southeast Asia. Wheatley considered all early urban traditions as sharing an astrobiological paradigm of cosmological speculation that sought to demonstrate the unity of existence by invoking the regularity of celestial understanding with the reality of biological irregularity on Earth. James Vance was another among the progenitors of urban historical geography. His doctoral thesis submitted in 1952 was on the suburbanization of Boston, and during the 1960s and 1970s, he turned to a broader research of the development of urbanization in Western civilization. He asserted the development of modern urbanization as based on the link between capitalist development and formation of new transportation systems. Another group of geographers that researched the urban past were the scholars from the materialist school of urban morphology led by the British geographers, Michael Conzen and Jeremy Whitehand, that approached urban settlements as the outcome of an unconscious production process of built up areas that took place over long periods. They considered the morphology of urban areas as the product of the work of subsequent building activity of successive generations that leave their traces in the urban environment. Tracing and analyzing these layers reveals the origins and development of urban areas.

Scope and interests of historical and other geographers researching the urban past can be found in the *Journal of Historical Geography* that was inaugurated in 1975. Among the first articles published in the journal was David Ward's discussion of modernity and urban development in Victorian Britain. Ward had previously edited a volume discussing different aspects of nineteenth century American cities. Most articles published in the first 5 years of the journal on urban subjects were reviews of books on urban history and morphology. American and British cities and towns have been the main foci of interest, although other areas were considered such as in the article of Kent Mathewson's work on the evolution of Mayan cities of South America reflecting the Saurian tradition of cultural and historical American geography or in that of Yehoshua Ben Arieh, the founding father of Israeli historical geography, on the role of Christian missions in the development of nineteenth century Jerusalem.

Definitions, Methodology, and Theory: Forming the Contours of Urban Historical Geography

Growing number of publications on urban historical geography during the 1970s and early 1980s resulted in a debate over defining and dividing historical geographers from other scholars specializing in the history of urban areas such as urban historians; social, economic, and urban geographers; and historians of architecture.

Richard Dennis's introduction to his book on nineteenth century English cities published in 1984 provides an analysis of different views on the research of the past of urban areas. At the center was the debate between social scientists: urban sociologists, geographers, and economists that used either quantitative or structuralist theories and methods on the one hand, and the more eclectic urban, social, and economic historians relying on the research of case studies for drawing general conclusions, on the other hand. The synthesis adopted by Dennis for his research of the urban historical geography of English nineteenth century towns was based on methods used by social historians and urban geographers. In the middle of his research stands the human agent and different aspects of the formation of urban space by human societies, rendering a thorough and inclusive analysis of urban life in the Victorian period.

Different from Denis's interpretative approach, Harold Carter presented in 1983 a positivist deductive method for forming a synthetic approach to urban historical geography. The first part of his research is dedicated to the definition of cities, review of theories of urbanization, and to the history of the urban phenomenon since the beginning of history and its diffusion in ancient and medieval times. The second concentrates on urban population on functions of cities and on the evolution of urban systems. The third deals with the evolution of town planning and the last, with the internal functional and social structures of the city. Different from Dennis, who wished to form a method for urban historical–geographical research depicted in nineteenth century British cities, Carter presented a historical overview of urbanism from which he aspired to draw a methodological inclusive framework. Nevertheless, the scope of his work, spare the first part, is also limited to the European and North American cities.

The rise of critical Marxist theory among geographers in the 1970s, whose main protagonist, David Harvey, concentrated on the urbanization process of the nineteenth and twentieth centuries, presented a major challenge to urban historical–geographical research. The ambitious attempt to use Marxist theory to form a holistic view of the city and rapid transformation of modern urban areas seems to be a major challenge for both Dennis and Carter in forming an alternative view and method for the research of urban historical geography.

The interpretative approach such as that of Richard Dennis attempted at a study of contemporaneous concepts of cities on the one hand, and modern researchers' concepts on these cities, on the other hand. In this sense the interpretative approach was a predecessor of the postmodern turn to social and cultural–historical geographies. Harold Carter seems to preserve a more positivist–conservative tradition of urban historical–geographical research that follows the school of materialist urban morphology, concentrating on the evolution of the morphology and functions of cities and towns from antiquity to the nineteenth century.

The widening interest in urban historical geography in 1980s resulted in the evolution of an Anglo German group of urban historical geographers, following two seminars held in (West) Germany and Britain successively in 1982 and 1983. The publication in 1988 of a volume edited by Dietrich Denecke and Gareth Shaw titled *Urban Historical Geography: Recent Progress in Britain and Germany* was the outcome of both seminars. According to the editors' introduction, the main aim of the seminars was to initiate the formation of international cooperation in urban geographical–historical research and an international terminology of urban history and urban historical geography. Despite this, it seems as the forming of this venture was not intended to form a subdiscipline within historical geography but to form an interdiscipline group of urban historical geographers and urban historians.

It seems rather obvious that the initiators of both seminars did not wish to form a methodological framework for a school of historical geography. The volume edited by Denecke and Shaw provided a cross section of the work of contemporaneous urban historical geographers, at least those located in Britain and Germany. It ranged from more traditional research of urban morphology, structure, and functions to social–historical geography of cities and towns, from evolution of European towns in the Middle Ages to the formation of modern cities in the nineteenth century. Preservation of historical cites was also in the scope of this volume. The multifaceted quality of urban historical geography was also evident in the chapter titled 'Historical geographies of urbanization' included in Butlin's work on historical geography that was published in 1993.

The aim to form a supranational discussion group of urban historical geographers seemed to correspond to the views presented in 1992 by Richard Rodger, a British economic and urban historian and the editor of *Urban History*. Rodger considered the aim of the journal as embracing methodological, historiographical, and analytical contributions that would enable intertemporal and interdisciplinary interests to coalesce, united by a common interest in the town or city. Accordingly, urban history was the subdiscipline in which urban historical geographers should be operating among sociologists; political scientists; art, cultural, economic, and social historians; and architectural and planning historians interested in the study of urban history. Submerging into urban history and becoming partners in a wide interdisciplinary group may have been luring, but it also meant relinquishing the identity of geographers, which was not in the interest of urban historical geographers.

Urban Historical Geography and the Cultural Turn in Geography

The rise of a new cultural geography afforded historical geographers as a whole the opportunity to redesign their place within the discipline of geography.

The cultural turn in geography was simultaneous to the endeavor to formulate a thematic group of urban historical geographers that took place in the 1980s. Similar to historical geography, cultural geography seemed also to be marginalized following the rise of quantitative methods followed by that of Marxist theories that placed economic and social geographies at the center of the discipline.

Through the 1970s and 1980s historical geographers including those interested in urban areas strove to locate themselves, in the theoretical and methodological realms, among historical research, on the one hand, and theories and methods employed by main currents of geography, on the other hand. The subdiscipline was situated as regarded by Carville Earle "…in midstride, between empirical tradition and epistemological innovation, between the rock ribbed world of materiality and the elusive worlds of representations" (Earle, 1995: 455–459). No wonder he considered "…historical geography is anything but unified…" and historical geographers as those whose research interests "…span the spectrum of geographical research…" (Earle, 1995: 455–459).

These efforts hampered the attempt to form a well distinct group of urban historical geographers, forming a multifaced group among which some inclined to follow conservative methodologies while others tended to adopt theories and methods of social and economic geographies.

Different from the invigoration of historical geography, the Sauerian tradition that dominated cultural geography for decades seemed to be losing its appeal in the 1960s and 1970s. The rise of the humanistic school in geography and domination of Marxist theories, both as responses to dominant quantitative and systematic approaches, gave birth in the 1980s to a new cultural geography. By the late 1980s geography underwent a cultural turn that reformulated not only cultural geography but other subdisciplines as well, including historical geography. Moreover, it was not geography alone that was affected by the cultural turn. All human sciences had to consider postcolonial critique that questioned the dominance of Western political and cultural concepts, as well as post structural critique raised by Continental philosophy on the relevance of Western models of society. In short, the turn from modernism to a postmodernism formed the fundamentals of the cultural turn in geography. Postmodernism has contested intellectual hierarchies and afforded the legitimization of plural views of the world. Many geographers adopted the use of the plural form ‘geographies’, rather than singular ‘geography’, to present a world of multiple identities and views that form different concepts of space and especially of places.

The cultural turn among historical geographers in general and urban historical geographers in particular is evident through scrutinizing reports of historical geog raphy sessions that took place in different conferences since the early 1990s. Representations, identities, icon ographies, and discourses were frequently present in titles of presentations, side by side with older historical geographical traditions that concentrated on assimilation and settlement of immigrants in North America and development of industries and rural areas.

In such an intellectual climate of epistemological, theoretical, and methodological transformations that blurred interdiscipline and subdiscipline lines of de marcation, the effort to define and articulate a group of urban historical geographers seemed irrelevant. This is reflected by Miles Ogborn’s discussion in *Progress in Human Geography* on the work of historical geographers in 1997. Relating to different categories of research he re gards the debate on relations between history and geography, in themes such as geography and empire, environmental history, memory, and heritage, and social and cultural histories. Works on cities and towns are mentioned in both latter categories, but not as a separate theme of research.

Michael Heffernan, the former editor of the *Journal of Historical Geography* defined the journal in his 1997 edi torial note as “…a haven of intelligent eclecticism….” He followed the spirit of his predecessor, Alan Baker’s expansive vision of historical geography as “eclectic and liberal: no particular dogma about the nature of historical geography was to be promoted: the study of no problem, period or place in the past was to be prohibited…the concern of the journal ought not to be the precise def inition and rigorous policing of historical geography’s borders.” The liberal and eclectic spirit of the journal reflected and inspired an intellectual climate that did not encourage the formation of subgroups concerned on one facet of historical–geographical research such as urban historical geography.

The liberal and eclectic nature of historical geography is evident, sustained in the variety of publications on the subdiscipline and those published in both its stages: *Journal of Historical Geography* and *Historical Geography*. This is also true concerning edited volumes on historical geography such as that edited by Brian Graham and Catherine Nash titled *Modern Historical Geographies* published in 2000. Their use of the plural form ‘geog raphies’ presents the multifaced character of historical–geographical research and its liberal and eclectic nature. This is well articulated by Deryck Holdsworth in a paper titled ‘Historical geography: the ancients and the moderns – generational vitality’ published in 2002 in *Progress in Human Geography*. He quotes Graham and Nash’s introduction regarding modern historical geog raphies “that signal new directions informed by femi nism, post structuralism, anti racism and post colonial perspectives sharing concerns about questions of power and meaning with other researchers more readily located within the traditional subdisciplines of economic, cul tural, political and social geography” (Holdsworth, 2002: 671–678). And notice the use of singular while referring to other subdisciplines of geography. The celebration of a modern historical geography does not imply the for mation of a ‘new’ historical geography but rather main tenance of the tradition of liberal eclecticism.

This is also true regarding geographical–historical research of urban areas. A scrutiny of papers published on the urban phenomenon in *The Journal of Historical Geography* since the mid 1990s reveals a diversity of re search methods and subjects. Some trends of research could be, of course noted, as the almost total avoidance of ancient, medieval, and early modern cities and towns in favor of the research of modern urbanization that is of the nineteenth and twentieth centuries. Nevertheless, the effort to define borders and limits of urban historical geography seems to be neglected. Liberal eclecticism affords the recent urban historical geographer to easily mingle, in the academic sense, with urban historians and other practitioners of the urban past, becoming a part of a postmodern interdisciplinary scholarship whose interests are the search of the history of urbanism.

See also: Industry, Historical Geographies of; Street Names and Iconography; Urban Morphologies, Historical.

Further Reading

Basset, K. (2005). Marxian theory and the writing of urban history: On David Harvey’s Paris, capital of modernity. *Journal of Historical Geography* 31, 568 582.

Butlin, R. A. (1993). *Historical Geography: Through the Gates of Time and Space*. London: Edward Arnold.

Carter, H. (1983). *An Introduction to Urban Historical Geography*. London: Edward Arnold.

Denecke, D. and Shaw, G. (eds.) (1988). *Urban Historical Geography*. Cambridge: Cambridge University Press.

Dennis, R. (1984). *English Industrial Cities of the Nineteenth Century*. Cambridge: Cambridge University Press.

Earle, C. (1995). Historical geography in extremis? Splitting personalities on the postmodern turn. *Journal of Historical Geography* 21, 455 459.

Goheen, P. G. (1984). Book review on an introduction to urban historical geography by Harold Carter. *The Geographical Review* 74, 376 377.

Graham, B. and Nash, C. (eds.) (2000). *Modern Historical Geographies*. London: Longman.

Heffernan, M. (1997). Editorial: The future of historical geography. *Journal of Historical Geography* 23, 1 2.

Holdsworth, D. W. (2002). Historical geography: The ancients and the moderns generational vitality. *Progress in Human Geography* 26, 671 678.

Konvitz, J. W. (1993). Introduction. *Journal of Urban History* 19, 3 10.

Ogborn, M. (1999). The realities between geography and history: Work in historical geography in 1997. *Progress in Human Geography* 23, 97 108.

Rodger, R. (1992). Urban history: Prospect and retrospect. *Urban History* 19, 1 22.

Vance, J. (1990). *The Continuing City*. Baltimore, MD: Johns Hopkins University Press.

Wheatley, P. (1977). The suspended pelt. Reflections on a discarded model of spatial structure. In Deskins, D. R., Kish, G., Nyusten, J. D. & Olsson, G. (eds.) *Geographic Humanism, Analysis and Social Action: A Half Century of Geography at Michigan*, pp 47 108. Ann Arbor MI: Department of Geography, University of Michigan.

Historical Geography

G. M. Winder, Ludwig-Maximilians-Universität, Munich, Germany

Glossary

Cross-Sectional Analysis In parallel with the pattern of explanation in geomorphology current before 1950, Derwent Whittlesey organized his research around a series of reconstructions of past geographies. While this approach might highlight change over time, the explanatory power of his ‘sequent occupance’ approach remained at best weak and at worst mystifying. Detailed reconstructions of past geographies became associated with the work of many historical geographers, including H. C. Derby, and remain an important tool, as is illustrated in the many historical atlases published toward the end of the twentieth century.

Geographical History This term has been used in different ways within debates over the scope and content of historical geography. Generally, the term was used either as a synonym for historical geography, or as a descriptor of the special case of the use of geographic factors to explain historical phenomenon. Thus, for example, Carville Earle argued for a revised historical geography, one that focused on geographical explanation of historical problems.

Introduction

Geography and history overlap and the relations between them are complex and made more so by the many projects that expand and contract within and around the two fields as social science has been refashioned. As we work to understand and explain spatial distributions, landscape features, human–environment relationships, and ideas about space, place, and territory, geographers often require historical explanations, and need to theorize processes and to establish which events, patterns, and changes were significant. Similarly, historians often seek to understand the geographical contexts for the action they study, and to write histories of cities, regions, nations, environments, environmental knowledge, and landscape features. Historians, like social scientists, have embraced new theorizations of place and space, mobility, and identity. As in all subject areas, approaches, methods, questions, and ideas arise and are abandoned, so that there is a history of changing ideas and projects. Particularly in the last two decades, researchers in the fields of history and geography have experienced an intense and dynamic period of exchange and development. However, while the academic field of historical geography continues to expand and has become more and more a site for interdisciplinary research, the institutional support for the subdiscipline has narrowed, as geography departments favor appointments of geographers with thematic specializations, albeit with historical research interests. This situation could be read as a sign of the long term success of historical geography: it has helped to establish a permanent place for history within geography.

The origins of historical geography lie in efforts to incorporate historical research at the regional scale on landscape change and past geographies into geography. The scope of this early work is outlined below. Regional historical geographies and work to reconstruct past geographies continue to be important matters of research, but historical geography now embraces many other research foci. Studies of capitalism and its transformative effects, research on environmental change and human–environment relations, and work interpreting the meanings of historical landscapes and landscape change are three other main constitutive areas of study. However, the field of historical geography is rapidly expanding around new themes, such as postcolonial reinterpretations of the history of empires, historical geographies of the colonized world, the connections between identity formation, and the spaces and places of nationalism, ethnicity, and gender. These and other new themes are outlined below. Historical geography is now a broad interdisciplinary field of inquiry and not a narrowly defined subdiscipline of geography. This is a cause of some adjustment within geography, and the concerns of historical geographers for their endeavor are discussed.

Setting a Place for History within Geography

During the mid twentieth century, historical geographers struggled to assert historical methodologies and research questions ‘within’ geography, which at the time tended to emphasize spatial relationships in a geography of regions. Historical geographers such as H. C. Darby in the UK and Derwent Whittlesey in the USA focused on landscape change and process. Both cultivated the core practice of researching past geographies using forms of cross sectional analysis. Similarly, Carl Sauer in Los Angeles promoted studies of the diffusion of cultural objects and practices, borrowing from German geography as well as biology and anthropology. At Wisconsin, Andrew Hill Clark fostered historical geographic work inside North America emphasizing regions and economies. After World War II, a new generation of historical geographers confronted the challenges of the quantitative revolution – is history needed for geography if geography is spatial science? – and the conservative politics of the Cold War decades – can historical explanation be informed by Marxism? Cole Harris argued for geography as a project of regional–historical synthesis, as he and other geographers struggled against the growth of spatial science at the expense of humanistic geographies. Growing numbers of historical geographers met in an internationalizing but loosely organized conference.

Human geography in general began to embrace and discuss historical explanation in new ways in the 1980s and 1990s. Spurred on by debates in the social sciences and history over historical materialism and structuralism, and by feminist theory and phenomenology, many human geographers looked for alternatives to positivist social science. Historical geographers drew inspiration from a wide range of social theory, and humanistic and feminist approaches. They researched diverse archival sources and wrestled with how these could be interpreted and what it meant to work in the archives. They brought new concepts, frames of reference, subjectivities, questions, and topics as well as new quantitative methods, cartographic technologies, and forms of textual analysis to their desks, as they pursued new and alternative forms of historical explanation.

At the same time, other geographers called for critical inquiry into the subjectivities of geography and their histories. Already by 1990, David Harvey confidently argued that analysis of the production of space, the geographical imagination, and the historical geography of concepts of space and time were central to any critical enquiry into the contemporary world. He pushed for both a critical historical geography of capitalist transformations of the world and a critical inquiry into the ways in which geographers framed time, change, and development. Simultaneously, Brian Harley challenged conventional histories of cartography, and mapping practices within historical geography, by drawing out the ways in which cartography had been structured to serve various state purposes. Dennis Cosgrove interrogated the concept of landscape in its historical European uses and found similar issues with geographers' studies of landscape change. It seemed that historical geography would be central to a human geography that was moving beyond positivist framings of knowledge, but the distinctive approaches that emerged cut across earlier framings of historical geography as a subdiscipline of geography organized around inquiries into regional and landscape change and the analysis of past geographies, even as such research reached a high point of sophistication. The result has been a prolific growth of historical geography but in a skein of divergent but entangled strands of research rather than a subdiscipline narrowly focused on landscape change and past geographies at the regional scale.

Research Foci within the Broad Field of Historical Geography

Regional geographies and histories have continued as one thick braid of research. In North America, Donald Meinig and the many students of Andrew Clark were very active over several decades in research and commentary in and around this approach. Their projects emphasized the dilemmas of (re)settlement and human–environment relations. Perhaps their efforts reached an apogee with the publication of Donald Meinig's four volume *The Shaping of America*, Carville Earle's *Geographical Inquiry and American Historical Problems*, Cole Harris' *The Resettlement of British Columbia*, and the three volume *Historical Atlas of Canada*. Thematic mapping has an important place in this research, which drew on and contributed to theory and debates in economic, agricultural, and urban history, and the work of the Annales School. North American regional historical geographers have contributed two key organizing ideas: the geographical personality of particular regions and the formation of culture areas. This regional or area studies approach was not confined to rural landscape change, but included important work on the appropriation of 'new' lands through colonialism, on urban systems and industry. Nor was it confined to North America or to 'national' projects. Historical geographers have written of a kaleidoscopic array of regions and areas. Generally, these are synthetic works that draw together research on diverse phenomena and at many scales. Such research is now of considerable interest to those historians who are rescaling their research either to world, empire, regional, or even micro scales.

A second braid of research brings together work on the historical geography of capitalism and its transformative effects through history. This historical geography research is now more substantial than the regional studies stream. It involves interest not only in how capital works and acts to construct and destroy specific landscapes, work rules, economies, and social identities, but also in the everyday life worlds, responses, resistances, and life chances of those living in a transforming, modernizing world. It ranges from the complex situated montages of the spaces and experiences of world's fairs and dock work by Alan Pred, through Robert Lewis' analyses of the remaking of North America's metropolitan industrial spaces, Richard Harris' unmasking of North American suburban development processes and experiences, Richard Dennis' survey of English industrial cities, and M. Overton's study of agrarian transformation in England, to Richard Walker's work on the capitalist framings and constructions of nature. In Montreal, Sheri Olson and Pat Thorndon led research into a wide range of urban experiences of the industrial city, including re search designed to identify and interpret the adaptive strategies and experiences of individuals in demographic and household economy terms and by class, ethnicity, gender, age, and family status. The construction of new spaces of consumption and segregated ethnic residential neighborhoods, and the contestation of public space have been other important areas for historical geographic re search, but there is a fantastic array of related studies. For example, in *The Ride to Modernity*, economic geographer Glen Norcliffe explored Canadians' peculiar encounter with the bicycle and its spatial ramifications and social meanings as one of the first mass produced consumer luxuries. Historical geographers working in this area have developed synergies and alliances with economic geog raphers; social and urban historians; demographers; labor, economic, and business historians; and cultural geog raphers; among others. Their research has broken with long held historical narratives summarizing spatial pro cesses under industrial capitalism. Such research also challenges many accepted ideas, theories, and foun dational concepts used to understand and represent the significance of present reconfigurations of uneven de velopment under capitalist processes.

Together, environmental transformations and human–environment relations comprise another plait within historical geography. This is a research area with an illustrious pedigree: there are ancient scholarly works on these subjects; modern scholarship is associated with Alexander von Humboldt, George Perkins Marsh, Vidal de la Blache, and Élisée Reclus. Alan Baker identifies an entanglement of roots with historical ecology as practiced in anthropology and biogeography. Pierre Gourou's groundbreaking research into laterization in tropical environments is also part of this legacy. Midway through the twentieth century, *Man's Role in Changing the Face of the Earth*, an encyclopedic survey on the subject edited by William Thomas, drew together an interdisciplinary who's who of the field. This feat was echoed in 1990 with another edited volume *The Earth as Transformed by Human Action*. These themes remain important foci for research within historical geography. One consistent interest has been in the reconstruction of past natural environments. Another focus of attention is on human impact on natural environments. Here, historical geographers such as Donald Meinig who worked out the historical environ mental limits of the American Great Plains and the South Australian wheat belt, J. M. Powell who evaluated water management in Victoria, Australia, and Michael Wil liams, Graeme Wynn, and Michael Roche who each wrote on forests (mis)management, have made sub stantive contributions. Andrew Goudie has penned an introduction to the subject that has served the purpose for a long time now. This is a subject that requires con tributions from physical geographers, and some, such as Stanley Trimble, have obliged. There is also a need for further research on another related strand: the history of perceptions of natural environments.

These efforts of historical geographers are now aug mented by a burgeoning literature in environmental history from the pens of such as Donald Worster, Richard White, Alfred Crosby, William Cronon, Caroline Mer chant, and Richard Grove. Environmental history sprouted in the USA but has now been set to work in many lands. Its practitioners have tended to embrace environmentalist values and write critically of past human interactions with environments. Their research projects are pitched at diverse scales from global patterns of species diffusion to local studies of dams and wetlands. In particular, environmental historians have picked up on the study of urban environments, on animal histories, and on the history of perceptions of the environment and of science networks related to environmental knowledge. There is a growing traffic in ideas and practices between historical geographers researching environmental trans formations and human–environment relations and en vironmental historians. However, there are multiple, entangled strands of research in this thriving field of interdisciplinary research.

Landscape interpretation and landscape change are further key themes for historical geographers. Generally, the foundational work in historical geography focused on the making of landscapes. Carl Sauer, W. G. Hoskins, and J. B. Jackson (founding editor of the American journal *Landscape*) were major mid twentieth century figures writing to understand the making of national and re gional landscapes. Their interests have been taken up inside historical geography in the subsequent works of Wilbur Zelinsky, Pierce Lewis, Fred Kniffen, Yi Fu Tuan, Michael Conzen, Derek Holdsworth, and many others in North America, and of Della Hooke, Hugh Prince, and others in the UK. This tradition emphasizes landscape design, environmental awareness, and their manifestation in, especially, vernacular landscapes. Nevertheless, its subject matter ranges from urban and suburban street scapes, through public and religious buildings and monuments, factories, office towers and banks, to back yards, public gardens, cemeteries, and national parks. Particular attention has been paid to non Western land scape design as, for example, in Hong key Yoon's publi cations on geomancy. Historical geographers share their interests in these topics with professionals in archi tectural history, folk studies, anthropology, and archae ology and in museum studies among other specializations.

The meanings of landscape constitute a second main, but not necessarily separate, theme of landscape studies. Indeed, historical and cultural geographers writing on the making of landscapes were equally interested in the meanings of landscape, especially in the structures and practices of landscape design and also in terms of histories of landscape styles, and cultural interpretations of specific landscapes. Nevertheless, the emergence in the 1980s of the new cultural geography transformed writing in historical geography on the meanings of landscape. By asserting that a landscape is a material object that manifests many discourses, the new cultural geography fostered a burgeoning, critical literature on the meanings of landscape. This literature, heralded by Denis Cosgrove and Stephen Daniels' edited collection *The Iconography of Landscape*, emphasizes symbolic landscape representation in the past, the power relations manifest in such representations, and their work in forming identity, memory, and heritage. Attention has turned to many forms of landscape representation that were not the main foci for earlier research, including landscape painting, travel writing, literature, money, and exhibitions. Cultural geographers such as Denis Cosgrove, Stephen Daniels, Brian Osborne, James Duncan, Jeanne Kay Guelke, Pyrs Gruffudd, Emily Gilbert, Mona Domosh, and Suzanne Seymour have delighted in studying the representational systems and meanings of the past, so that their work has added multiple new strands of research to the broad field of historical geography.

There are many other subjects that could be profitably included in this review, but a full account of historical geography's tangled web is really beyond the scope of this article. Geographers are interrogating imperialism as part of postcolonial reinterpretations and critiques of the history of empires and are writing historical geographies of the colonized world. Historical geographies of identity probe the connections between nationalism, transnationalism, ethnicity, identity, globalization, and place. Geography as a scientific and educational practice has a historiography that is being critically reviewed. There are collaborative projects in geographic information system (GIS) and remote sensing to recreate and reinterpret past geographies. The roles of geographers in past imperialism, cartography, land survey, and Orientalism are properly the subjects of critical enquiry. There are new historical geographies of the present which analyze the use of aids to memory and nostalgia as active forces in today's landscapes. Brian Graham and Catherine Nash identified one major interest in the study of modernity, but even the title of their reader, *Modern Historical Geographies*, asserts an array of approaches, concepts, and perspectives. Attention continues to be directed to many other eras, and historical geographers write from many perspectives and pursue an array of approaches.

Different contexts have inflected specific 'national' foci for historical geography research, and this means that it is problematic to read off the content of historical geography from trends in the UK and USA alone. In New Zealand and Australia, historical geography has long been identified with research into environmental management, land use change, colonial land alienation and resettlement, and now with interpretation of colonial cultural landscapes. By contrast, in Canada, urban, industrial, and cultural landscape foci are each equally as important as environmental history or colonialism. In Japan, pre nineteenth century geographies have been a main focus of interest, but there is new research on more recent periods, and especially in Japan's engagement with modernity and imperialism. In both Singapore and Israel, historical geographers have been directly confronting the colonial and postcolonial policies that have shaped their countries. In Germany and Eastern Europe, environmental and land use change has emerged as the main spring of historical geography, but in each case, it is problematic to make more than these cursory comments, since the field is in flux and 'national' styles of historical geography are inherently problematic.

Fluidity is also true of the institutional frameworks for historical geography. As governments transform their research funding opportunities, geographical research is increasingly organized in research groups and teams and the institutional contexts for historical geography morph. As individual geography departments refocus their programs, there is often no room for a dedicated 'historical geographer', and few such positions are advertised, but it is also expected that geographers of all kinds will be historically informed and inquiring. This means that many departments are homes to excellent historical geographic research but that this is seldom organized under the framework of a department of geography with a chair in the historical geography of a particular region. Colleagues appointed in GIS, physical, cultural, economic, social, or urban geography turn their attentions to historical problems and research. Some historians are housed in geography departments, and there is also movement the other way. In these changed circumstances, knowledge of regional geographies and histories, the historical geography of capitalism, environmental transformations, human–environment relations, past landscapes, archival resources, and historical research methods remain crucial to much research.

Historical Geography in Review

The ways in which historical geographers conceive of their field of study have changed considerably over the last 50 years. Contrasting three reviews of historical geography illustrates the point. In his 1962 classification, H. C. Darby framed historical geography as a subdiscipline of geography which comprised research into geographies of the past, changing landscapes, the role of the past in the present, and the role of geographical factors and explanations in historical re search, or geographical history. Darby's seemingly programmatic statement reflected the struggles of mid twentieth century historical geographers within geography and has been criticized from many quarters in subsequent years. Nevertheless, his review was meant to open a general field of inquiry rather than to constrain research.

When defining historical geography for *The Dictionary of Human Geography* in 2000, Daniel Clayton declared that "geographical work on the past has become eclectic and no new orthodoxy for historical geography is likely to emerge." Such work is self consciously interdisciplinary or without disciplinary label. Clayton leaves the im pression that 'historical geography' is a label attached to a modernist project of academic disciplines that has less and less purchase as critiques of the foundational narra tives of modernity have intensified. However, for Clay ton, the legacy of the subdiscipline continues to make its flagship *Journal of Historical Geography* a valuable arena for research and debate.

By way of contrast, in a review of the relations be tween history and geography published in 2003, Alan Baker acknowledges no old orthodoxy for historical geography and instead identifies three general ap proaches within a broad interdisciplinary field of his torical geography – locational geographies and histories, environmental histories and geographies, and landscape studies – and a fourth, integrative approach that em phasizes region and area. He argues that regional geog raphies and histories have a claim to be at the center of historical geographic research because they are integra tive in an explicitly geographic way and might draw from any of the other three approaches. Nevertheless, Baker is at pains to highlight "the diversity, the coherence and the relevance of historical geography" and "to reconcile some of the oppositions perceived within it and also between historical geography and history" (Baker, 2003). While he unabashedly seeks to discern order in the de velopment of historical geography and rules of engage ment in discourse, he welcomes experiment, dialog and encounter and denies that there ever was a narrowly defined historical geography concerned only with the reconstruction of past geographies and their changes over time.

Historical geography is, as ever, evolving. There are discernible trajectories but just how these are evaluated depends a great deal upon one's point of view. Many historical geographers express concern over the prospects for their subdiscipline and discipline. Their concern seldom arises from any sense of seeing their subject area 'invaded' by other interests, whether by cultural geog raphers or environmental historians. Instead, they express the belief that critical geographers represent historical geography as stuck in positivist, cross sectional, and regional practices, while ignoring most of the substantive research in the field. These historical geographers con tend that, as other interesting and productive foci have emerged and present challenging and inspirational find ings, there is now too little attention to region, urban system, ecosystem, or watershed in (historical) geo graphic research. They perceive a danger that the sweep of regional and systems change will be missed and his torical explanation eschewed.

At the same time, many historical geographers inter pret the same developments in a different way: they celebrate the continued growth and expansion of a dy namic field of interdisciplinary research with many new and interesting concepts, theories, and frameworks for research. Historical geographers are in dialog with his torians of all kinds: economic, social and cultural geog raphers, geographical information scientists, even physical geographers. Theorizing space, place, landscape, maps, populations, cities, and environment is increasingly important in history. This is seen in the many col laborative projects and teams of researchers that have emerged in environmental history, urban history, social science history, history of cartography, landscape and cultural studies, and historical demography. For example, some new works in New Zealand history have featured teams of geographers, historians, and other social scien tists. These include not only the *Bateman Historical Atlas of New Zealand* but also Tom Brooking and Eric Pawson's *Environmental Histories of New Zealand*, and Barbara Brookes, Anne Cooper, and Robin Law's *Sites of Gender*. Indeed, the authors publishing in the pages of *The Journal of Historical Geography* have increasingly diverse backgrounds. Within geography, time, memory, and the past have proved to be of continuing interest as geographers seek to research uneven development in space just as the world seems more intimately wired and networked than ever before. In each and every the matic subdiscipline, geographers need to develop suitable historical geographies relevant to their subject matter.

Historical geography is open to new approaches and is currently expanding as critical geographical work on the past proliferates. The subdiscipline's borders are open, and bridges across the 'divide' with history are multi plying. Historical geography is now more eclectic and more experimental than ever before. Simultaneously, historical geographical research has never been more relevant as social scientists and historians try to work out what aspects of the twenty first century world are new, and how they are significant: after all, the twentieth century is history. It is most certainly the best time yet to be doing historical geography.

See also: Landscape; Phenomenology/Phenomenological Geography; Regional Geography II.

Further Reading

Baker, A. R. H. (2003). *Geography and History: Bridging the Divide*. Cambridge: Cambridge University Press.

Baker, A. R. H. (2007). Classifying geographical history. *The Professional Geographer* 59(3), 344 356.

Brookes, B., Cooper, A. and Law, R. (eds.) (2003). *Sites of Gender: Women, Men and Modernity in Southern Dunedin, 1890 1939*. Auckland: Auckland University Press.

Butlin, R. A. (ed.) (1993). *Historical Geography: Through the Gates of Space and Time*. London: Edward Arnold.

Conzen, M. (ed.) (1990). *The Making of the American Landscape*. Oxford: Routledge.

Cosgrove, D. and Daniels, S. (eds.) (1988). *The Iconography of Landscape: Essays on the Symbolic Representation, Design and Use of Past Environments*. Cambridge: Cambridge University Press.

Gentilcore, R. L., Measner, D., Walder, R. H., Matthews, G. J. and Moldofsky, B. (eds.) (1993). *Historical Atlas of Canada, Vol. II: The Land Transformed.* Toronto, ON: University of Toronto Press.

Goudie, A. (1981). *The Human Impact on the Natural Environment*. Oxford: Blackwell.

Graham, B. and Nash, C. (2000). *Modern Historical Geographies*. Harlow: Pearson Education.

Harris, R. C. (1971). Theory and synthesis in historical geography. *Canadian Geographer* 15, 157 172.

Harris, R. C. (1997). *The Resettlement of British Columbia: Essays on Colonialism and Geographical Change*. Vancouver, BC: UBC Press.

Harris, R. C. and Matthews, G. J. (eds.) (1987). *Historical Atlas of Canada, Vol. I: From the Beginning to 1800.* Toronto, ON: University of Toronto Press.

Harvey, D. (1990). Between space and time: Reflections on the geographical imagination. *Annals of the Association of American Geographers* 80, 418 434.

Hoskins, W. G. (1995). *The Making of the English Landscape*. Harmondsworth: Penguin Books.

Kerr, D., Holdsworth, D. W., Laskin, S. L. and Matthews, G. J. (eds.) (1990). *Historical Atlas of Canada, Vol III: Addressing the Twentieth Century 1891 1961.* Toronto, ON: University of Toronto Press.

McKinnon, M., Bradley, B. and Kirkpatrick, R. (1997). *Bateman New Zealand Historical. Atlas Ko Papatuanuku e Takoto Nei*. Albany: David Bateman and Department of Internal Affairs.

Meinig, D. (1988 2004). *The Shaping of America: A Geographical Perspective on 500 Years of History*. New Haven, CT: Yale University Press.

Norcliffe, G. (2001). *The Ride to Modernity: The Bicycle in Canada, 1869 1900*. Toronto, ON: University of Toronto Press.

Pawson, E. and Brooking, T. (eds.) (2002). *Environmental Histories of New Zealand*. Oxford: Oxford University Press.

Philo, C. (1994). History, geography and the ‘still greater mystery’ of historical geography. In Gregory, D., Martin, R. & Smith, G. (eds.) *Human Geography: Society, Space and Social Science*, pp 252 281. Basingstoke: MacMillan.

Historical Geography, Evolution of

J. M. Powell, Monash University, Melbourne, VIC, Australia

Glossary

Cultural Landscape Initially, those elements in the landscape which have been fashioned (from a supposedly 'natural' landscape) by human agency, implying potential for change under the impress of succeeding 'cultures'; later, the notion of foundational, wholly pristine blank slates was disputed, but the underlying idealization was not abandoned.

Domesday Book A detailed administrative survey of the resources of the bulk of England, compiled in 1086 on the orders of William the Conqueror; famously employed by H.C. Darby and his associates in interpretative reconstructions of the human geography of the eleventh-century England.

Enlightenment Traditionally, an eighteenth-century philosophical movement animated by a resolute faith in human reason and linked with portentous developments within and beyond Europe; recent reformulations variously reject and qualify aspects of definition, timing, and location.

Historiography In the context of this article, the study of the development over time of research and writing, contributing to the evolution of a discipline or subdiscipline.

Postmodernism Characterized broadly, in the context of this article, by a skepticism toward the status and claims made for earlier theorizing; it lends itself to styles of interpretation that show how the location (in culture, class, etc.) of authors and readers influences how a text (document, data, map) is produced and used.

Nurturing Interface, to World War I

As noted in the several influential overviews on which this summary draws, key Western European origins can be traced to intellectual and vernacular understandings of the interaction between human communities and their physical environments – including humanities style preoccupations and scientific or proto scientific appraisals of the 'classical' worlds of Greece, Rome, and Egypt in English, French, and German writings. During the eighteenth and nineteenth centuries, evolving interrelationships between geography and history favored the analysis of contemporary cartographic evidence and adaptations of mapping techniques in the study of underlying territorial themes. To some extent, that also applies to reflections upon the development of environmental concepts and attitudes, and the influence of environmental conditions on human affairs.

Imaginative responses to the increasing pace and complexity of change underpinned a progressive refinement of scholarship during the nineteenth century. Premiums on 'place' and 'space' were enhanced by direct and vicarious experience of global tumult: collisions between bustling nation states; expanding frontiers of settlement in a so called New World; the spread of European imperialism; political, racial, and social friction; revolutionary innovations in transport and communication; and vastly augmented opportunities for commercial trade. Confronting displacement attended deep seated transformations at every scale. From today's vantage point, the sum of those pressures seems to invite the urgent engagement of specialist academic geographers. But university geography seldom attained freestanding status before the early years of the twentieth century, and that helps to explain why early allusions to 'historical geography' have been so readily dismissed as conventional acknowledgments of the value of environmental and locational 'backgrounding' to mainstream history. In fact, they may also be seen as modest anticipators of an independent initiative.

Distinctive Emergence, to *c.* 1960

There was less equivocation after the foundation of autonomous departments of geography in the universities of the Western world. In that novel setting, neophyte brands of historical geography attracted sojourners and immigrants from other branches of human geography, and from economic and other corners of history. Continuity was similarly apparent in histories of exploration and cartography, and in sorties into the history and philosophy of science. During the first three decades of the twentieth century, a growing number of British practitioners created a robust hub of activity. Impressions of contemporary German work on the genesis of cultural landscapes, and on the potential utility of historical geography's intermediate position between geography and history for interpretative regional analyses, may have done something to prepare the way. Awareness of the interdisciplinary work of French historians may have contributed toward a greater receptivity.

Initially, British geographers seemed generally content with the argument that, insofar as an interest in the antecedents of the present was relatively common to the subject at large, the adjective 'historical' could be safely applied across the board. A minority strove instead to define a new subdisciplinary area characterized by a cul tivation of expertise in the handling of preferred types of evidence. H.C. Darby's edited collection of essays on the historical geography of England suggested a useful pre cedent. Its cross sectional studies of the prehistoric, Roman, and Scandinavian periods employed an enter prising mixture of source materials; depictions of other, later Britains were selectively compiled from the re connaissance accounts of contemporary travelers. These ambitious 'slices' or period samples were complemented by essays on evolutionary ('vertical') themes – develop ments in trade; century long panoramas – and by more concentrated analyses of regional change, both rural and urban. Devised as avowedly methodological compromises, they would be seen nonetheless as exemplars of 'classical' historical geography. Precisely dated reconstructions often facilitated manageable projects and staked out new terri tory – disciplinary allegiance being disclosed in linkages (of familiar content or focus) with interpretative geog raphies of the present. 'Slicing' was seldom rigidly con fined, however, and the narrative or connecting essays built confidence, credibility, and interdisciplinary exposure.

An impressive body of pioneering work was com pleted between the 1930s and 1950s. American ap proaches reflected differences in institutional history, intellectual legacies and perceptions of civic responsi bilities, and perhaps greater susceptibilities to the pre dilections of foundation scholars. To some extent there was also a greater interest in tracing the influence of locational and environmental factors on the course of history, together with an identifiably 'new world' interest in environment–community relationships and the mak ing of regional distinctiveness.

On the other hand, Ralph Brown produced an innova tive cross section of the Eastern seaboard of the United States in 1810; indeed, he tried to reproduce the writing style of an imagined contemporary author, restricting the reconstruction to materials conceivably available to that author. Brown in 1948 followed up with a nationwide historical geography, the structure of which has been de scribed as an American version of Darby's successful ex emplification. In contrast, Carl Sauer's research school focused upon long term human–environment relation ships, particularly in the Americas. Its strongest historical geographical themes promoted regional approaches to the production of 'cultural' (or human modified) landscapes, the early domestication of plant and animal species, and induced or accelerated environmental deterioration. Commentators on this school have discerned a respect for traditional German scholarship and the effects of formative encounters with American frontier settlement. Convenient anthologies underline a markedly liberal outlook cultivated in the company of colleagues in anthropology, archaeology, botany, ecology, and physical geography.

In the light of the rather more conservative inter disciplinarity embraced by his British and European counterparts, Sauer's writings could be made to appear hazardously idiosyncratic, but there is no doubt that he inspired generations of talented researchers. In any event, national contrasts were (and are) exaggerated. For ex ample, Darby's expertise on wetland reclamation con tributed an important linking essay in his first edited collection on England; his later publications included a book on the same theme and several broader pieces on the changing English landscape, the making of English agricultural regions, and the clearing of Europe's wood lands. In the interim, British economic historians had picked up the landscape theme. National differences notwithstanding, by the later 1950s the reservoir of his torical geographical material was already emphatically international and widely accessible: in practice, large numbers of scholars and students were introduced to its full range of approaches and content. Given those de velopments, the contribution of Andrew Clark, one of Sauer's most distinguished students, is not satisfactorily described as 'American', 'European', or 'British'. His major books and appeals for the closer interdisciplinary study of regional and environmental change sketched out an inviting middle ground on both sides of the Atlantic, and indeed in both hemispheres.

Core Elaborations, from *c.* 1960

By the 1960s historical geography's widened purview broadly echoed the expanding project of human geog raphy and drew some of its readership and material from geography and history at large. In the process, the sin gular appeal of classical reconstruction was diluted. Ab sorbed within a vastly augmented international academy, historical geography became caught up in the swirling interchanges that enlivened the burgeoning humanities and social sciences. It remained theoretically feasible, documentation permitting, to shadow any major enter prise within present orientated human geography with an historical counterpart, but it was seldom possible or profitable to establish either points of origin or modes of diffusion. Some historical geographers remained more comfortable with co options of modern geographical techniques to clarify questions posed by historians. As in other fields, however, older scholarly goals proved quite resilient. They repeatedly surfaced in the form of the connoisseurial essay or keynote monograph directed to the cognoscenti, or as an ambitiously 'accessible' book espousing civic engagement.

By 1962 Darby was espousing an elementary tax onomy for the growing field: 'past geographies', 'changing landscapes', 'the past in the present', and 'geographical histories'. In practice, as explained in Baker, 2003, the first of these categories remained the most inclusive because it subsumed all forms of geographical enquiry related to the past. Impressive studies of medieval England by Darby and his associates illustrated a potential for greater sophistication in the special uses made of the celebrated *Domesday Book*, various taxation assessments, and other monumental data banks. Updated essays edited by Darby advertised the progressively expanding scope of geographical investigations into the past and produced sturdier combinations of 'horizontal' and 'vertical' approaches. A seven volume *Domesday* series delivered about 800 maps illustrating aspects of the geography of England in some AD 1086. Cumulatively, these analyses demonstrated the emergence of an area of intellectual inquiry in which hitherto underused technical and statistical materials were being brought into sharper relief alongside continuing resort to older 'qualitative' emphases.

The trend continued into the 1980s and beyond, partly in tandem with developments in 'cultural', 'economic', 'environmental', 'political', and 'social' approaches within human geography at large, but without sacrificing the links with history and certainly not at the expense of an entrenched, humanities style regard for individualistic essays. A prodigious literary output confirmed historical geography's status as an established scholarly field that defied simplistic characterization. Baker's persuasive overview finds four overlapping modern discourses: 'location', 'environment', 'landscape', and 'region'. His classification serves to emphasize that, to date, human geography has generally supplied the critical connections. It also recognizes that varieties of 'overlapping' may apply to the interests of individual practitioners as well as to the content of pertinent undergraduate programs.

Several of the works already mentioned are conveniently accommodated under the 'location' heading, together with later refinements on nineteenth century distributions. Good cases were also made for undeniably 'geographical' quantifications. But the interface was also enriched by discoveries or rediscoveries of 'space' and 'place' by Western historians and social scientists. Under 'environment', most indicative bibliographies include a number of national atlases and Glacken's elegant treatise on early ideas on human–environment relationships. As with every other category, historical geographies of the environment have also been embedded in broader geographical works, and resonate with companion works from other disciplinary homes. Baker's 'landscape' division confirms that observation. Convergences on landscape perception and appreciation, landscape heritage preservation, public memory, and the formation of regional and national identity gave abundant openings for cross disciplinary work. Geography's participation produced numerous books and a host of articles and unpublished theses.

The works of historical geography's early pioneers fit well enough into the main groupings, but continuity is perhaps most evident in Baker's final category, insofar as 'region' designates a core concept for geography as a whole. Yet in this instance the problematical impact of variations in scholarly contexts becomes particularly significant – diverse legacies of formal and informal learning traditions; the changing interplay between international and national research networks; inherent opportunities and constraints in specific archival endowments; anxieties and enthusiasms stimulated by discrete economic, environmental, political, social, and political imperatives; and so on. The category warrants as much art as science, and indeed positively invites strongly expressed opinion as well as methodology and mode of presentation. Similarly, exercises in national and international synthesis have covered a wide spectrum of issues and themes – concerning either urban development, say, or agriculture, industry, transportation and trade, banking, communication, or all of those and more, as outlined in Butlin's *Historical Geography*. Given such embarrassment of choice even the most provocative essays may be tolerated, welcomed, and occasionally prized.

One outcome is better seen as a more refined elucidation of the complexities of large scale geographical differentiation than as a resumption of geography's oldest and boldest ideal of 'comprehensive' regional synthesis: 'selective' being more accurate. Perhaps it reflects dispassionate scholarship, deeply personal commitment, and civic engagement, in equal measures. In the Northern Hemisphere, for instance, Sauer's early identification of 'culture areas' based upon the evolution of patterns of living was adapted between the 1960s and 1990s, when the nature of arguably distinctive American regions was busily revisited and ethnicity and landscape were repeatedly emphasized. For the later years the most prominent overview is Meinig's courageously individualistic perspective on 500 years of American history. It testifies to the author's profoundly geographical conceptualization of his own country – as a special or 'particular' system or structure of places that had been created by people engaged with particular locations, and with particular physical environments, at home and abroad.

Elsewhere, Dodgshon and Butlin edited process orientated alternatives to Darby's synthesis on England and timely accounts for Europe; and geopolitics informed several panoramic narratives, including 'spatial stories' on the legacies of Western opportunism in the Middle East. In the Southern Hemisphere, disputatious dialog between popular and official land appraisals, inspections of trial and error testing on settlement frontiers, and interpretations of evolving government policy for resource management, were central to Australian regional studies. Cultural transactions between Europeans and colonized communities underpinned new historical geographies of South Africa. Trans hemisphere studies included a cul tural historical geography of the influences of political and social idealism, environmental imagery, and utopian notions in regional settlement processes, and a three volume interpretation of the historical geography of the oceanic realm of the Pacific.

Geographers also became more concerned with con temporary formations of national identity, cultural geographical perspectives on the unfolding 'moral geog raphy' of the times, representations of cities as types of performance space, and the changing cultural topography and perceived resource endowments of the British Empire in its representation as a spatial system. These trends were accompanied by an imaginative harnessing of new and revisited sources and finely textured interrogations of more recent periods. As Butlin explains, the increasing penetration of nineteenth and twentieth century con texts called for more sophisticated consultations with census data and other statistical materials; newspapers, personal diaries, and business, school, institutional, and municipal records commanded commensurately higher premiums, but so did contemporary art, literature, music, and oral evidence. Cliff and others produced an enter prising statistical account of global epidemiology; Harley and others warned against casual reliance on con temporary maps; Ryan established that contemporary photography offered a neglected imperial archive. Un fortunately, venerable links with physical geography were commonly allowed to atrophy, and historical geographers with strong environmental interests were drawn toward the burgeoning, late arrived counterpart, 'environmental history'.

Intensifying interest in the modern era also justified an organizing focus on the very question of 'modernity' or modernism. Inevitably, many researchers have felt obliged to re examine the contemporary scholarly and civic engagements of the leaders and followers in their own subject, and where relevant to incorporate an ad justed coming to terms into their reconstructions of their own national territories in past periods. In turn, 'post modernism' reacted against all manner of alleged pre sumption by the modernists, subject specialisms included. Our final section takes the point a little further.

Histories of Geographical Thought

Geographical historiography simultaneously proclaims a stand alone, unfettered pursuit, and nourishment for the very conscience of the discipline at large. At the outset its kinship with historical geography exhibited shared skills and scholarly predilections, and commonalities of at traction to certain past periods. Later, increasing spe cialization threatened the relationship, but it endures. One international journal, *Geographers: Biobibliographical Studies*, has published a large number of succinct studies of the contributions made by individuals to geographical thought. Its rubric does not exclude nonacademic pro files, and to some extent recognizes proto scientific contributions. Furthermore, it is already considered a modest entry point for understandings of the evolution of geographical thought in a few national communities. For example, in the case of notionally 'new' countries like Australia, the journal already offers a first port of call – with prompts on the engagements of a sizable group of explorers, writers, statesmen, independent environmental scientists, and others in governmental agencies, as well as on professionally designated geographers and other university personnel.

The study of the history of geographical thought has been reinvigorated by a surprisingly small number of researchers over recent decades. While the traditional relationship with historical geography is affirmed, several of the group's most noted writings quite properly engage with the larger gatherings convened under the 'history and philosophy of science' banner. Research abstracting services, the *Journal of Historical Geography*, *Progress in Human Geography*, and geography's leading 'mixed' international periodicals, record an increasingly lively presence which challenges not only mainstream historical geography but all branches of academic geography. A few examples must suffice.

'Modern' geography commonly located its origins in a period of famously empirical learning – in an elevated regard for 'observation', 'rationality', and 'order'; in heroic oceanic explorations and advances in cartography; and in the works of towering individuals such as Humboldt and Ritter – in short, in an era known as 'the Enlightenment', routinely confined to the eighteenth century. Recent writings have disputed the assumptions of universality and chronology in this conceptualization. For fellow travelers in historical geography, Livingstone in 1992 built a bridge toward circumspection, and Liv ingstone and Withers in 1999 introduced the rapidly diversifying terrain of science history. Students and tea chers alike became more prepared to reconsider the taken for granted titles applied to academic special izations and long entrenched 'eras' or periods of change: it seemed perfectly feasible to discuss, for instance, 'Enlightenment's geographies' and 'geography's en lightenments'. Subsequent consolidations took British narrators well beyond commonly preferred boundaries, and challenged deep seated prejudices against geography's prescientific lineage. Venturing further still, Livingstone in 2003 pointed out that 'place' has in fact been rather critical in the evolution of science. Histories of scientific practice might therefore supply yet another focus for reconstructions of the geographies of the past.

Issues

As its particular scholarly tradition discloses, historical geography's original roots had been repeatedly acknow ledged before the two decades or so preceding the new millennium, when there were perceptible intensifications of 'interdisciplinary' activity. As with so many other es tablished fields, despite built in recognitions of debts owed to initial allegiances, latent centrifugal tendencies have probably been strengthened. Historiography's most elementary lesson cautions that, when entailed obli gations are readily overlooked, it is at some cost to intergenerational transfers of knowledge and to the very retention of disciplinary identity.

See also: Cultural Geography; Darby, H. C.; Diaspora; Enlightenment Geography; Geohistory; Heritage; Historical Geography; Modernity; Moral Landscapes; Postmodernism/Postmodern Geography; Region; Ritter, C.; Sauer, C.; von Humboldt, A.

Further Reading

Baker, A. R. H. (2003). *Geography and History*. Cambridge: Cambridge University Press.
Bassin, M. (1999). *Imperial Visions. Nationalist Imagination and Geographical Expansion in the Russian Far East, 1840 1865*. Cambridge: Cambridge University Press.
Brown, H. C. (1948). *Historical Geography of the United States*. New York: Harcourt Brace.
Brown, R. H. (1943). *Mirror for Americans: The Likeness of the Eastern Seaboard, 1810*. New York: American Geographical Society.
Butlin, R. A. (1993). *Historical Geography: Through the Gates of Space and Time*. London: Arnold.
Butlin, R. A. and Roberts, N. (eds.) (1995). *Ecological Relations in Historical Times*. Oxford: Blackwell.
Cliff, A., Haggett, P. and Smallman Raynor, M. (1998). *Deciphering Global Epidemics: Analytical Approaches to the Disease Records of World Cities*. Cambridge: Cambridge University Press.
Conzen, M. P. (ed.) (1990). *The Making of the American Landscape*. London: Unwin Hyman.
Conzen, M. P., Rumney, T. A. and Wynn, G. (eds.) (1993). *A Scholar's Guide to Geographical Writing on the American and Canadian Past*. Chicago, IL: University of Chicago Press.
Dodgshon, R. A. and Butlin, R. A. (eds.) (1990). *A Historical Geography of England and Wales*. London: Academic Press.
Earle, C. (2003). *The American Way: A Geographical History of Crisis and Recovery*. Lanham, MD: Rowman and Littlefield.
Gilbert, D., Matless, D. and Shorts, B. (2003). *Geographies of British Modernity: Space and Society in the Twentieth Century*. Oxford: Blackwell.
Glacken, C. J. (1967). *Traces on the Rhodian Shore. Nature and Culture in Western Thought from Ancient Times to the End of the Eighteenth Century*. Berkeley, CA: University of California Press.
Gregory, D. (2004). *The Colonial Present. Afghanistan, Palestine, Iraq*. Oxford: Blackwell.
Harley, B. J. (2001). *The New Nature of Maps: Essays in the History of Cartography*. Baltimore, MD: Johns Hopkins University Press.
Harris, R. C. (2002). *Making Native Space: Colonialism, Resistance, and Reserves in British Columbia*. Vancouver: University of British Columbia Press.
Heathcote, R. L. (1965). *Back of Bourke. A Study of Land Appraisal and Settlement in Semi Arid Australia*. Melbourne: Melbourne University Press.
Leighly, J. (ed.) (1963). *Land and Life: A Selection of the Writings of Carl Ortwin Sauer*. Berkeley, CA: University of California Press.
Lester, A. (2001). *Imperial Networks: Creating Identities in Nineteenth Century South Africa and Britain*. Hoboken, NJ: Taylor and Francis.
Livingstone, D. N. (1992). *The Geographical Tradition: Episodes in the History of a Contested Enterprise*. Oxford: Blackwell.
Livingstone, D. N. (2003). *Putting Science in Its Place: Geographies of Scientific Knowledge*. Chicago, IL: University of Chicago Press.
Livingstone, D. N. and Withers, C. W. J. (eds.) (1999). *Geography and Enlightenment*. Chicago, IL: University of Chicago Press.
Livingstone, D. N. and Withers, C. W. J. (eds.) (2005). *Geography and Revolution*. Chicago, IL: University of Chicago Press.
Meinig, D. W. (1962). *On the Margins of the Good Earth: The South Australian Frontier, 1869 84*. Chicago, IL: Association of American Geographers.
Meinig, D. W. (1986, 1993, 1998, 2004). *The Shaping of America*. New Haven, CT: Yale University Press.
Ogborn, M. (1998). *Spaces of Modernity: London's Geographies 1680 1780*. New York: Guildford Press.
Powell, J. M. (1978). *Mirrors of the New World: Images and Image Makers in the Settlement Process*. Canberra: Australian National University Press.
Powell, J. M. (1991). *A Historical Geography of Modern Australia: The Restive Fringe*. Cambridge: Cambridge University Press.
Ryan, J. R. (1997). *Picturing Empire: Photography and the Visualization of the British Empire*. Chicago, IL: University of Chicago Press.
Spate, O. H. K. (1979, 1983, 1988). *The Pacific Since Magellan*. Canberra: Australian National University Press.
Williams, M. (1974). *The Making of the South Australian Landscape A Study in the Historical Geography of Australia*. London: Academic Press.
Williams, M. (2003). *Deforesting the Earth*. Chicago, IL: University of Chicago Press.
Wishart, D. J. (1994). *An Unspeakable Sadness. The Disposse ssion of the Nebraska Indians*. Lincoln, NE: University of Nebraska Press.
Withers, C. W. J. (2001). *Geography, Science and National Identity: Scotland Since 1520*. Cambridge: Cambridge University Press.

Relevant Websites

http://www.elsevier.com/locate/jhg
Journal of Historical Geography.
http://www.historical geography.net/
http://www.h net.org/histgeog

Historical-Geographical Materialism

S. Kirsch, University of North Carolina, Chapel Hill, NC, USA

Glossary

Capital Value created and reinvested in the production of market commodities. Capital travels through different circuits in processes of production and exchange, and takes form as the ability to purchase or invest in labor power, machinery, and raw materials.

Capitalism A specific mode of production, or economic organization, in which labor becomes a commodity and workers are separated from ownership of the means of production. Surplus value created by labor accrues not to the laborers but to the capitalist class.

Dialectics/Dialectical A mode of thought or inquiry, characterized by mutually constitutive subject/object relationships, which takes contradiction as its starting point. Marx's dialectical materialism focuses on the social contradictions arising from material conditions.

Divisions of Labor How tasks are divided among workers. Spatial divisions of labor describe the ways in which different forms of labor are divided across spaces and regions, creating, for example, a coal mining region or a region in which services predominate. The distribution of work by gender, in households as well as places of industry and commerce, defines the gendered division of labor.

Social Reproduction The everyday maintenance of the formal and informal social institutions that make the material reproduction of a given society possible over time. Struggles over social reproduction serve to reproduce and also to transform the legal, political, social, and economic structures of societies.

Totality A (complex) whole to which different parts (complexly) relate. In Marxist thought the mode of production constitutes a rich totality that is, for now, irreducibly capitalist.

Introduction

Historical geographical materialism describes the intellectual project of creating a Marxist science of geography and, simultaneously, of infusing the broader currents of historical and dialectical materialism with explicitly geographical concepts and sensibilities.

Though historical geographical materialism did not emerge concretely in anglophone human geography until the 1970s and 1980s, the term taps into the deeper roots of materialism in Western intellectual history, an explanatory approach which gives ontological priority to the material conditions of existence – the immediate physical world – over spiritual and metaphysical notions of causation. Grounded in the rejection of nonmaterial prime causes, materialism instead defines its own categories on the basis of demonstrable investigations. Materialist frameworks for engagement with the physical world, often arising in opposition to (or in tension with) spiritualism and theocratic domination, were revolutionary in this sense, and have had profound implications for the making and legitimation of knowledge in modern societies, and for the politics of that authority. From Galileo's sunspots to Darwin's finches to Marx's class formations to contemporary debates over global climate change and the genetic body, strains of materialist thought continue to animate public discourse and the unfolding of human relations with, and as part of, the material world.

In popular usage, materialism (and materialist) have for centuries been mobilized as pejoratives, especially by spiritualists in the effort to disparage those possessing insufficiently religious worldviews. In the humanities and social sciences, materialism has, since the late nineteenth century, been strongly associated with the historical materialism of Karl Marx and scholars working in the Marxist tradition. Historical materialism (as Marx's approach to analyzing historical change later came to be called) proposes that historical events and processes – notably, historical transformations, such as that from feudalism to capitalism in Europe – were, at least in principle, no less understandable in terms of demonstrable material causes than are those that constitute the solar system or evolutionary processes. In this analysis, the roots of social change were to be found in the material conditions of social existence, or being. As Marx described this approach in an oft quoted passage from the preface to *A Contribution to the Critique of Political Economy*:

> The mode of production of material life conditions the general process of social, political, and intellectual life. It is not the consciousness of men that determines their existence, but their social existence that determines their consciousness. (Marx, 1970: 20 21)

Thus it is not the idea or spirit which produced the world, as Hegel's dialectic had it, but the world which produced the idea, and Marx stressed in particular the relationship between the forces and requirements of the production process, on the one hand, and the social relations of production, on the other, including the legal and political arrangements of civil society which make the mode of production possible. Turning Hegel on his head, in this sense, had important political as well as philosophical implications. Freedom from the oppressive conditions of modern life, Marx argued, would require not only a transformation of consciousness, but also revolutionary changes in the material conditions of social existence.

For Marx, these conditions were to be understood as historical, not universal or timeless. Historical material ism thus presents itself, in part, as a general theory of historical development, through which large scale his torical transformations could be understood, in dialectical terms, to have arisen from contradictions embedded – materially – in a mode of production, or in successive modes of production. While capitalism was itself revo lutionary, remarkably so, in its mechanisms for continu ally revolutionizing the forces of production, it also produced new contradictions, always setting the stage for future historical transformations: capitalism produced wage labor and pauperization, but it also produced con ditions that made plausible the formation of the working class. Marx's construction of the working class – as a scientific category, imagined community, and new his torical subject – illustrates the complexity, and rhetorical power, of the dialectic of human agency in historical materialism, that people make history but not under conditions of their own choosing. We confront a world that is already constituted, structured by powerful forces beyond the control of any individual or social group, but it is a world still 'actively' constituted; we enter history as its participants.

Of course, the manner and degree to which the social, political, and intellectual are shaped "in the last instance" by the requirements of production, in a world wherein, as Louis Althusser remarked, "the lonely hour of the 'last instance' never comes," continues to be a matter of re search and debate (1969: 113). The mode of production itself, for example, can be read in narrow economic terms and also much more openly in its dialectical relations with social, political, and cultural forms and processes. For Marx, working out the implications of materialism was a means of challenging, on the one hand, the idealism of Hegelian political philosophy and, on the other, the presumed truths of political economy. Political economy, as Marx's critique illustrated, provided scientific justifi cation for treating the capitalist system, and the specific forms of inequality that it produced, as natural. Marx's trenchant analysis of political economy and its logics in the first volume of *Capital*, which included close attention to then contemporary debates over abstract theories of market equilibrium, as well as struggles over the regu lation of child labor and the working day, strongly reflects this materialist approach. The most radical lessons of historical materialism, however, concerned the future, not the past. For if historical transitions between modes of production could be rationally understood, their causes identified, then perhaps so too, as Marx and Friedrich Engels wrote in the *Communist Manifesto*, could the tran sition to a more just, equitable, and rational society be envisaged for – and realized in – the future. Historical materialism, with its priority of matter over mind, thus presented a host of new epistemological, political, and moral–ethical problems; no longer, Marx insisted, could science itself be seen as an enterprise detached from the earthly world of its investigations. The implications of these developments for geographical thought and prac tice are explored in the next sections.

Historical Materialism and Geography

Geographers share a long history of engagement with materialism as a practical philosophy. For the navigators, natural philosophers, cartographers, engineers, explorers, cosmographers, surveyors, social scientists, social theor ists, and other historic and contemporary authors of geographical knowledge, the work of observing, re cording, mapping, and analyzing the immediate physical world has been a fundamental concern. It is hardly sur prising that the geographical language of discovery, for example, would diffuse so widely into the physical and mathematical sciences, for the ongoing reconfiguration of relations among observation, experience, authority, and truth was indeed evocatively framed in terms of the new planetary geographical knowledge produced amid Europe's colonial and imperial ventures.

From a materialist perspective, the centrality and value of geographical knowledge in the work of empire building, natural resource exploitation, and the expansion of legal, economic, and territorial systems, raises ques tions about the material basis of geographical knowledge itself, that is, the conditions that make it possible. We can look back, in this sense, to the emergence and insti tutionalization of modern scientific geography in im perial Western states (and to some extent, Japan) during the late nineteenth century and discern the strong role of imperial states in the formation of geographical know ledge and the modern discipline of geography. While geography was in some ways a materialist Enlightenment science par excellence, its development, in this milieu, was thus tied closely to expansion of territorial power and global commerce, and the justification of such pro cesses as natural; the new discipline of geography, far from picking up on the radical implications of Marx's materialist philosophy, was largely a conservative one. But to emphasize that geographical knowledge cannot be understood independently of its material conditions is not to reduce scientific thought to the horizons of its social origins. The commingling of geography with anarchism and values of mutual aid and respect in the work of Pyotr Kropotkin and Élisée Reclūs, amidst the flourishing of imperial geography in European capitals, suggested the possibility of a more radical, and politically engaged, geographical materialism.

Geographers' record of engagement with historical materialism has been brief but transformative, both inside and outside the discipline. In anglophone human geography, the rise of radical geography in the late 1960s and 1970s has been strongly associated with the rise of anti (Vietnam) war, anti imperialist, feminist, antiracist and other social movements. Geographers took notice of socially and spatially uneven distributions of poverty, wealth and power, at local, regional, and global scales, and vigorously critiqued their discipline's seeming inability to meaningfully address them, even in the world's wealthiest countries. While the efforts to turn geography into a rigorous (and especially, quantitative) science during the decades following World War II had been animated partly by liberal and progressive impulses, some geographers, best known among them, the once leading theoretician of quantitative geography, David Harvey, would turn to historical materialism as a means of rethinking geography and political economy so as to better understand and address social conditions.

For Harvey, materialism was fundamental to geographical science, or at least it had the potential to be. Geographical knowledge, as Harvey defined it in a 1984 'historical materialist manifesto':

> records, analyzes and stores information about the spatial distribution and organization of those conditions (both naturally occurring and humanly created) that provide the material basis for the reproduction of social life. At the same time it promotes conscious awareness of how such conditions are subject to continuous transformation through human action. (Harvey, 1984: 1)

Harvey's brief manifesto, published in a forum of the Association of American Geographers, thus sought to orient geography around process and dialectics. He also sought to liberate geography from a rising positivism in the field which, for Harvey, undermined "its own virtues of objective materialism by spurious claims to neutrality." In a world unjustly structured by capitalism and its imperialist and geopolitical logics, Harvey argued, "Geographers cannot remain neutral. But they can strive towards scientific rigor, integrity, and honesty. The difference between the two commitments must be understood" (Harvey, 1984: 7).

The challenge of articulating historical materialism with geography – and in some ways the premise of the project – was that Marx, among a long line of social theorists, was inadequately geographical: Marxian theory had yet to overcome its emphasis on time and transformation, deal with the spatiality of the processes it analyzed, develop an adequately spatial vocabulary. Although Marx had closely investigated what we might call local contexts, and a range of locally and nationally specific social conditions, their significance, and the problem of geographical variation more generally, had not been systematically integrated into his theoretical formulations. Geography may have been implicit to Marx's critique of political economy, in other words, but in a world profoundly structured by patterns of uneven geographical development, it had to be made explicit. While the work of Lenin and Rosa Luxemburg on imperialism had expanded knowledge of capitalism's spatial dynamics, there was clearly far more to be learned.

By the early to mid 1980s, building on institutional resources created in the previous decade, including the radical geography journal *Antipode*, geographers had begun to seriously rework elements of Marxist theory through explicitly spatial frames. Key works included Harvey's *Limits to Capital*, Doreen Massey's *Spatial Divisions of Labor*, Neil Smith's *Uneven Development*, and *Social Relations and Spatial Structures*, a multidisciplinary edited volume, among others. Carried out alongside (and as part of) a more general spatial turn in social theory, geographers, drawing on diverse influences, brought historical materialism into contact with a range of disciplinary concerns and subfields, from urban studies of ghetto formation, suburbanization, and gentrification to agriculture and ecology to international development, industrial and economic geography, political geography and geopolitics. Terrain as diverse as landscape and culture, political movements and public space, labor, migration, commodification and consumption, technology, and the body have been investigated using historical geographical materialist frameworks. The next section proceeds by sketching the still emerging contours of historical geographical materialism, but stresses that, in the present, highly eclectic domain of human geography, theoretical frameworks do not tend to be seen as dogmatic or as requiring a singular conceptual agenda. Rather, and befitting a broadly materialist approach to science, different forms of materialism (and Marxism) persist, and they are often integrated with other theoretical resources.

Contours of Historical-Geographical Materialism

In his historical materialist manifesto, Harvey promoted the emergence of a new agenda for geographers which could be organized around the "study of the active construction and transformation of material environments (both physical and social) together with critical reflection on the production and use of geographical knowledge within the context of that activity" (1984: 6). The first part of this agenda matches closely with the Marxist dialectic of labor and nature through which the human species being can be understood as both (1) acting on nature (i.e., the material world), chiefly through variegated forms of labor, to produce the conditions of its own survival and reproduction; and (2) at the same time, is itself a part of nature, and subject to natural forces (such as hunger). Transforming nature through work, then, the nature of the species being itself is transformed, and human societies have been made and remade. A clearer understanding of the spatial relations and processes through which capitalist societies, as historically specific modes of production, have been constituted, and a richer knowledge of the geography of capital's circuitry, and capitalism's requisite demands for human (and machine) labor, could thus be achieved through the development of more adequate geographical concepts, and through empirically engaged historical geographical studies. Rethinking the dynamics of capitalism as they articulate with local, regional, and planetary transformations has been, thus far, the most active area of research with an intellectual basis in historical geographical materialism. Following the work of the French Marxist theorist Henri Lefebvre, geographers have, along related lines, sought to understand how the production and appropriation of nature, space, and landscapes have made possible the reproduction of capitalist societies over time.

The second part of Harvey's agenda – with its emphasis on the production and use of geographical knowledge, broadly conceived – similarly underscores the materialist philosophical basis of the project. For if knowledge, even that which is authorized as scientific, is always shaped in terms of the perspectives and contexts in which it is developed, then surely there are many different kinds of geographical knowledges, for example, those produced by military strategists, multinational industrial site selection analysts, real estate agents, metropolitan police, and so forth, for specific practical purposes, or, along other lines, in the atlases and gazetteers produced for consumption by schoolchildren in different national and regional contexts. Prioritizing matter over mind, again, requires acknowledging that the conditions of material existence shape even our ability to understand and explain these conditions to ourselves. Grasping this positionality can also help to clarify the dialectical relations between the production and use of geographical knowledges and the reshaping of social and physical environments. Harvey grants little in the way of disciplinary privilege to geographers in the making of geographical knowledges, but he does envision a special role for historical geographical materialists – precisely because they are conscious of the problematic of the positioned or situated nature of geographical knowledge – in making sense of the complex, dialectical relations among the worlds we know, the worlds we inhabit, and the ones that we build or hope to build.

Critical to Marxist dialectical methodology is the emphasis on 'process': an object (e.g., a nail, a book, or an ear of corn) is constituted relationally, and thus cannot be understood separately from the processes and relations through which it is produced. Historical geographical materialism, in this sense, "turns all geography into a historical geography … not just of the past but also of the present and future." For Erik Swyngedouw:

> Historical geographical materialism starts from the premise that things exist, but these things are the embodiment of (they interiorise) relationships; things become the outcome of processes that have themselves ontological priority. In other words, everything flows; the flow or the process constitutes the things. (Swyngedouw, 2001: 94)

In both physical and cultural geography, the method of working backwards from the object to the processes which produced it has for long been a favored method in morphological approaches. The dialectical approach of historical geographical materialism was not the first which sought to understand geography – the occupied earth, at least – as a product of human activity, but it did open new vistas. Although, unlike a nail or an ear of corn, the production of geographical environments is never complete, *per se*, but rather an open and contested process, historical geographical materialism nevertheless requires an approach that begins with specific material geographies, past or present, and seeks to understand how they were produced, and under what conditions. By insisting that any geographical arrangement be seen as both a reflection and a constitutive element of an ongoing process (and interrelated, complex, and conflicting sets of processes), we sharpen our understandings of historical geographical change, and open knowledge of these dynamics to alternative possibilities. Seeing geographical forms – for example, the spatial variation of land rent values and housing markets, the location of technological innovation, or the landscape of a migrant labor camp – as part of a process, existing in relation to all manner of other processes and structured by what can be described as a capitalist totality, has significantly renewed and complicated Marxist thought by introducing a range of geographical concerns. These concerns have been pursued in a number of different ways, from Harvey's efforts to spatialize Marx's theory to work developed more explicitly around structuralist, structurationist, regulation theory, and post structuralist Marxist traditions. Geographers, in other words, have adopted a multiplicity of Marxisms and, so not surprisingly, debates persist over the explanatory power granted to capitalism and specific capitalist institutional structures as agents of change in relation to the agency of individuals, social groups, and nature itself.

For Harvey, theorizing the geography of capitalist accumulation – most explicitly in *The Limits to Capital* – meant focusing, in particular, on the tensions between capital's need for spatial fixity, on the one hand (e.g., in production facilities, technological investment, state social expenditures to ensure social reproduction of labor), and its demand for spatial mobility and new 'spatial fixes', on the other, that is, the geographical movement of production processes as a strategy for lowering labor and land costs, in the context of unending competition over relative surplus value. In enunciating a historical geographical materialist theory of crisis, Harvey also develops a materialist spatial ontology rooted in multiple and intervening spatialities: geographical space is mediated abstractly, atomized by markets and policed by state regulation; it is dialectical, a product of social and political contest; and, as articulated in subsequent works like *The Urbanisation of Consciousness* (1985) and *The Condition of Postmodernity* (1989), it is experienced and refracted culturally in complex ways. *Limits to Capital*, is not only still being widely read, it is still being written, published in its third revised edition in 2006. But it may be in Harvey's cultivation of a broad sensibility of space as relative to material processes, as compared to the precise lines of philosophical and economic argumentation worked through in *Limits to Capital*, that he continues to shape the historical geographical materialist imagination most actively. Harvey's more recent works (including *Justice, Nature, and the Geography of Difference*, *Spaces of Hope*; *The New Imperialism*; and *A Brief History of Neo Liberalism*) have expanded the scope of his historical geographical materialism by drawing connections among a range of cultural, political and geopolitical, natural and environmental dynamics. While Harvey's concepts of the spatial fix, creative destruction, the annihilation of space through time, and accumulation by dispossession, among others, have gained considerable traction both inside and outside disciplinary geography, different strains of historical geographical materialism have also contributed substantially to Marxist thought and practice and to the spatial turn in social theory more broadly.

Among the most influential have been Doreen Massey's formulations of spatial and gendered divisions of labor which, along with her work on relational conceptions of place, scale, and space time, have drawn significantly on structural Marxism, while also serving to bring geographical thought into closer contact with interdisciplinary cultural studies. Structural Marxism has too often been caricatured as a structural determinism, and interpreted through crude readings of the base–superstructure relation between the economy and civil society. It is better understood, following the work of theorists such as Althusser, Antonio Gramsci, and, in the context of British cultural studies, Raymond Williams and Stuart Hall, as a means of assessing the significance of culture, politics, ideology, and the state – all related to, but not necessarily determined by, capital – in the reproduction of capitalist societies over time. The work of Massey and her colleagues on spatial divisions of labor, which explores the sedimentation or layering of local social patterns in relation to the changing historical geographies of production in Britain, thus added considerable social and geographical nuance to the historical geographical materialist analysis of capitalism and its modes of production and reproduction, while also raising questions about that which may lie beyond capital's logics.

Another strain of historical materialism enters geographic thought from landscape studies. Denis Cosgrove's reworking of the landscape tradition in cultural geography in *Social Formation and Symbolic Landscape*, for example, imported a more contextualized sense of historical materialism, in particular from the art histories of John Barrell, John Berger, and others, one in which it is the relations between landscape as material and cultural symbolic form, that are given ontological priority. Picking up on the implications of this representational turn for situated studies of social struggle, Don Mitchell's research on the landscapes of migrant agricultural labor in California, and on the spaces of homelessness and free speech across US cities, has thus worked partly by holding material form in tension with representation, allowing for the creation of nuanced critiques of capitalist ideologies in spatial representations and the built environment. For Mitchell, though, it is the question posed, in different ways, by Harvey, Neil Smith, and Henri Lefebvre, about how the production of nature, space, and landscapes contributes to the survival of capitalism over time that continues to frame much of the analysis. To answer these questions, however, as with Cosgrove's immersion in the history of landscape painting and perspective, required a deep engagement with labor history, housing studies, legal studies, and other discourses located at specific contextual conjunctures. In this sense, the historical geographical materialist imagination remains a vibrant one in contemporary geographic writing and research more broadly, its analytical resources widely employed, because its critical engagement remains – as it must – an ongoing process. Recent work on urban nature and the metabolism of cities, though not always Marxist in *stricto sensu*, stands out as one active node of research wherein the interplay of materialism with ecology and urban studies, and also with science studies, feminist theory, art and architecture, to name just a few key areas, has helped to reinvigorate questions around the urbanization of nature and the problems of reification and agency in human–nature–technology relations (see Gandy, 2002; Heynen *et al*., 2006; Kaika, 2005; Kirsch and Mitchell, 2004; Loftus, 2006; Swyngedouw, 2004), and contributed to a more general rethinking of materiality in environmental and resource geography (Bakker and Bridge, 2006).

"Meanwhile, and for a start," asked Williams (1980: 121) a generation ago, "how is Materialism?" The answer, increasingly, is geographical; no longer can the spatiality of modes of production and social reproduction be taken for granted as neutral settings where historical processes occur. Historical geographical materialism is not a sub field or province of geography but really a set of onto logical commitments that have been conjoined with political and scholarly traditions, the uses of which have been anything but uniform. Nor will they be. Although there are dangers, in the eclectic theoretical atmosphere of the contemporary humanities and social sciences, that historical geographical materialism may serve, at times, as a kind of shorthand for providing an economic bottom line, it remains, at its best, an ambitiously open intel lectual tradition. For as Williams argued,

> The special character of materialism, and that which alone gives it value, is its rigorous openness to physical evidence. We cannot know in advance, as a test of fidelity, the changing materialist content of materialism. (Williams, 1980: 122)

As part of an ongoing, dialectical interchange between ideas and the material world, historical geographical materialism is indeed still evolving. A Marxist science of geography, and geographical Marxism, could not aspire to less.

See also: Capitalism and Division of Labor; Critical Geography; Geography, History of; Harvey, D.; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Radical Geography; Radical Political Economy; Regulation; Spatial Division of Labor; Uneven Development.

Further Reading

Althusser, L. (1969). *For Marx*. Brewster, B. (trans.). Middlesex: Penguin Books.
Bakker, K. and Bridge, G. (2006). Material worlds? Resource geographies and the 'matter of nature'. *Progress in Human Geography* 30, 5 27.
Bérubé, M. (2005). Materialism. In Grossberg, L., Morris, M. & Bennett, T. (eds.) *New Keywords* (revised edn.), pp 209 211. Oxford: Blackwell.
Castree, N. and Gregory, D. (eds.) (2006). *David Harvey*. Oxford: Blackwell.
Castree, N., Essletzbichler, J. and Brenner, N. (eds.) (2004). *Symposium: David Harvey's The Limits to Capital*: Two decades on. *Antipode* 36(3), 357 554.
Cosgrove, D. (1998). *Social Formation and Symbolic Landscape* (2nd edn.). Madison, WI: University of Wisconsin Press.
Gandy, M. (2002). *Concrete and Clay: Reworking Nature in New York City*. Cambridge, MA: MIT Press.
Gregory, D. and Urry, J. (eds.) (1985). *Social Relations and Spatial Structures*. London: Macmillan.
Harvey, D. (1984). On the history and present condition of geography: An historical materialist manifesto. *The Professional Geographer* 36, 1 11.
Harvey, D. (1985). *The Urbanization of Consciousness*. Oxford: Blackwell.
Harvey, D. (1989). *The Condition of Postmodernity*. Oxford: Blackwell.
Harvey, D. (1996). *Justice, Nature and the Geography of Difference*. Oxford: Blackwell.
Harvey, D. (2000). *Spaces of Hope*. Berkeley, CA: University of California Press.
Harvey, D. (2003). *The New Imperialism*. New York: Oxford University Press.
Harvey, D. (2005). *A Brief History of Neoliberalism*. New York: Oxford University Press.
Harvey, D. (2006). *Limits to Capital* (3rd revised edn.). London: Verso.
Henderson, G. (1999). *California and the Fictions of Capital*. New York: Oxford University Press.
Heynen, N., Kaika, M. and Swyngedouw, E. (eds.) (2006). *In the Nature of Cities: Urban Political Ecology and the Politics of Metabolism*. London: Routledge.
Kaika, M. (2005). *City of Flows: Modernity, Nature, and the City*. London: Routledge.
Kirsch, S. and Mitchell, D. (2004). The nature of things: Dead labor, nonhuman actors, and the persistence of Marxism. *Antipode* 36, 687 705.
Lefebvre, H. (1991). *The Production of Space*. Nicholson Smith, D. (trans.). Oxford: Blackwell.
Loftus, A. (2006). Reification and the dictatorship of the water meter. *Antipode* 38, 1023 1045.
Marx, K. (1970). *A Contribution to the Critique of Political Economy*. New York: International Publishers.
Marx, K. (1992). *Capital: A Critique of Political Economy,* vol. 1, Fowkes, B. (trans.). New York: Penguin.
Marx, K. and Engels, F. (1967). *Communist Manifesto*. New York: Penguin.
Massey, D. (1984). *Spatial Divisions of Labour*. London: Macmillan.
Mitchell, D. (1996). *The Lie of the Land*. Minneapolis, MN: University of Minnesota Press.
Mitchell, D. (2000). *Cultural Geography: A Critical Introduction*. Oxford: Blackwell.
Mitchell, D. (2003). *The Right to the City: Social Justice and the Fight for Public Space*. New York: Guilford.
Peet, R. (1998). *Modern Geographical Thought*. Oxford: Blackwell.
Smith, N. (1990). *Uneven Development: Nature, Capital and the Production of Space* (2nd edn.). Oxford: Blackwell.
Swyngedouw, E. (2001). Marxism and historical geographical materialism: A spectre is haunting geography. *Scottish Geographical Journal* 115, 91 102.
Swyngedouw, E. (2004). *Social Power and the Urbanization of Water: Flows of Power*. New York: Oxford University Press.
Williams, R. (1980). *Culture and Materialism*. London: Verso.

Relevant Websites

http://www.marxists.org/
Marxists Internet Archive.
http://davidharvey.org/
Reading Marx's Capital with David Harvey.

HIV/AIDS in Developed Countries

M. Sothern, University of St. Andrews, St. Andrews, UK

Glossary

GRID An acronym for Gay-Related Immune Deficiency, an early term used to describe what is now AIDS when it was understood to solely afflict gay men.

AIDS An acronym for Acquired Immune Deficiency Syndrome or acquired immunodeficiency syndrome, a collection of symptoms and infections arising from a compromised immune system as a result of HIV infection. Variations exist in the way AIDS is diagnosed – some countries diagnose AIDS when the $CD4^{+}$ T-cell count drops below an established threshold, whereas others require the co-presence of one or more opportunistic illnesses.

HAART An acronym for Highly Active Antireteroviral Therapy, a combination of antiretroviral drugs used to treat HIV since 1996. HAART has been responsible for prolonging life expectancy and for delaying the onset of AIDS.

HIV Acronym for Human Immunodeficiency Virus, the retrovirus that can cause AIDS, was isolated in 1983. There are two types of the HIV virus: HIV-1 is the most common and is found throughout much of the world; and HIV-2 is largely concentrated in West Africa. HIV-2 is less easily transmitted and the period between infection and illness is also usually longer.

PLWHA An acronym for Person(s) Living with HIV/AIDS.

Introduction

Acquired immune deficiency syndrome (AIDS) is commonly understood as the most tragic of the global pandemics. Metaphors of its global ubiquity, however, disguise its very real geographies. AIDS is neither evenly distributed nor does it take the same form everywhere. Because 95% of the estimated 39.5 million person(s) living with HIV/AIDS (PLWHA) are in developing countries, it is understandable that contemporary geographic research has an emphasis on the tragedy of the pandemic in Africa, Asia, and the countries of the former Soviet block. Geographic accounts emphasize that human immunodeficiency virus (HIV), like many preventable diseases, continues to diffuse along the fault lines of vulnerability: this is demonstrated in the work of people like Paul Farmer and Peter Gould where emphasis is placed on gender inequity, heteronormativity, racism, and the uneven geographies of development and poverty as explanatory variables in the diffusion of HIV. Clearly the geographies of vulnerability are uneven and this is especially significant for AIDS because AIDS is not itself a disease but a syndrome. It is usually diagnosed in an HIV seropositive person marked by the presence of an opportunistic AIDS related illness. Because the HIV virus compromises the ability of the body's immune system to fight off infection, it leaves it vulnerable to bacterium and other viral agents encountered in the environment. Thus, in a developed country, an AIDS sufferer is likely to die of an AIDS related pneumonia or AIDS related cancer whereas in sub Saharan Africa the opportunistic infection is more likely to be tuberculosis. While HIV seropositivity may be the necessary and common prerequisite for AIDS, HIV does not itself determine AIDS either biomedically, in terms of the opportunistic infections that mark AIDS, or socioculturally, in respect of the distribution of vulnerability to HIV infection in the first instance.

Viewed this way, AIDS is fundamentally a question of 'geography'. The centrality of geography to understanding the mechanisms of diffusion of HIV and patterns of the emergence of AIDS has resulted in a rich tradition of scholarship in medical and health geography.

A Brief History of AIDS in Developed Countries

In 1981, five unexplained cases of *Pneumocystis carinii* pneumonia (PCP) were reported to the Centre for Disease Control (CDC) in Atlanta. These cases were identified among previously healthy young gay men in Los Angeles; this report drew attention because while the microorganism that causes PCP is common, a functioning immune system is usually capable of warding off its attack. These reports of PCP in Los Angeles were soon followed by details of some 26 cases of Kaposi's sarcoma (KS) among young gay men in San Francisco and New York. KS is a rare cancer usually associated with elderly men. Of these 26 cases, four were also found to have PCP. While the CDC insisted that the etiology of these cases was yet unexplained, it surmised that the "fact that these patients were all homosexuals suggests an association between some aspect of a homosexual lifestyle or disease acquired through sexual contact."

Thus from these first documented incidences of what was to become known as AIDS the disease was intimately associated with the 'lifestyles' and sexual practices of homosexual men. Media and scientific coverage of the 'gay cancer' and 'GRID' (gay related immune deficiency) firmly established the perceived link between sexual deviance and AIDS. This association became the prin cipal mechanism for organizing risk categories, a heur istic device mobilized in the continued search for the cause of AIDS. While debates about the cause of AIDS raged in the early 1980s between theories of its viral or 'lifestyle' origins, the understanding, common both to popular accounts and the scientific community was that in some way homosexual practice must be to blame. This was because homosexuality was the common factor linking these otherwise disparate cases. Epidemiological studies, by focusing on this specific social group, worked in the broader cultural field both to sensationalize aspects of the so called 'gay lifestyle' and to obfuscate similar signs of immune suppression among the nongay popu lation. Histories reconstructing the early period of the AIDS crisis provided by Paula Treichler, Steven Epstein, Michelle Cochrane, and others argue – in what must now be seen as 'the' example of the normalizing function of the medical gaze – that medical science worked to re inforce unexamined assumptions about sexual normality and actively discounted signs of illness in intravenous (IV) drug users, women, hemophiliacs, and children.

By 1982, the framing of AIDS as a sexually trans mitted disease of gay men allowed for similar immuno suppression related illnesses to be identified among gay communities throughout much of the developed world. By the time the HIV virus was isolated and established as the necessary causal agent of the disease in 1983, AIDS both rhetorically and politically 'belonged' to the gay community. While it was now clear that it was the transmission of the HIV virus not the gay lifestyle that caused AIDS, the association of the disease with polit ically and socially marginal gay men was to prove dev astating. In the United States, for example, by 1986 there were over 16 000 confirmed AIDS cases over half of which had already died and yet, 5 years into the pan demic. President Reagan had not spoken of AIDS in a publicly significant way; when he finally did, it was to assure his television audience that there was no need for panic as it had not entered 'the general population'. The politics of AIDS therefore served to reinforce the dif ferentiation, stigma, and marginalization of infected groups – as epidemiological studies began to expand the focus on 'risk groups' beyond gay men to include IV drug users, commercial sex workers, and new immigrants (notably Haitians in the United States and African im migrants in Western Europe), AIDS continued to be a very visible index of race, class, and sexual preference.

However, this abjection that was attached to AIDS' early association with homosexuality proved an import ant resource for responding to the epidemic. It was in communities of gay men in the context of post 1970s sexual liberation movements that the cultural and political resources existed to agitate for research, or ganize against discrimination, invent new safer sexual practices, and to begin to disseminate these through the first HIV education campaigns. Indeed, while the political situation differed markedly between developed countries in terms of the politics of sexuality and the state pro vision of public health, the roots of many AIDS service organizations is to be located in these early responses by the gay community. AIDS served as a mechanism for constituting a broad based gay identity politics move ment in many developed countries; as Douglas Crimp repeatedly argued, AIDS shone an unprecedented media attention that helped to provide 'legitimation through disaster' for the struggle against homophobia in many parts of the world.

The first treatments for HIV emerged in 1988 but the most spectacular successes at staving off the onset of AIDS were seen with the wide adoption of highly active antiretroviral therapy (HAART) around 1996. For those fortunate enough to have access to these treatments HIV diagnosis no longer promised a swift death but, rather, has begun to be considered by some to be more on the order of a chronic disease (effective HAART has only been around for little more than a decade and thus its 'long term' effectiveness has not been established). As a result, there are now a burgeoning number of PLWHA and thus there is more empirical opportunity for contact with and contraction of HIV; moreover, the geography of the spectacular representations of AIDS deaths has largely shifted from the bodies of the developed onto the bodies of Africa – and thus, for many in developed countries, AIDS has become not just a disease of 'them' rather than 'us' but of 'over there' not 'here'. AIDS, in short, has become less frightening. As a consequence many developed countries including Canada, Germany, Japan, the Netherlands, Australia, the United Kingdom, and New Zealand have reported increases in HIV in fection since 2000. Furthermore, AIDS is no longer ex clusively located in the 'risk groups' of gay men, injecting drug users, and commercial sex workers but has become particularly acute in many immigrant communities and especially in prison populations, populations for which explicit and targeted prevention education are still being developed.

Medical Geography of AIDS

Early medical geography work focused on the spatial diffusion of the virus. As with medical geography ap proaches to other communicable diseases, geographers set about mapping the movement of the HIV virus from urban centers outward and across various global, na tional, and subnational scales. Because HIV is transmitted through the exchange of certain body fluids, the cartographies of HIV diffusion have always been more complex than those studies working with simple disease ecology models. Traditional concerns with distance and contagion models that were developed in response to influenza or measles, for example, have had to be re thought to account both for the long asymptomatic period of HIV infection and to allow for the cultural constructs of sexuality and stigma accompanying any discussion of AIDS. Moreover, because AIDS emerged at a time of increased mobility and ubiquity of transporta tion technologies, as Robin Kearns suggested in one of the few sustained reviews of geographic literature on AIDS "[t]he reality for geographers grappling with the AIDS era is that the most important vector may well be the 747," the geometries of the spread of HIV charted by medical geographers have had to become increasingly complex as is witnessed in the classic work of Peter Gould and Shannon, Pyle and Bashur from the early 1990s. Indeed, new geographic information system (GIS) visualization techniques have allowed for the ever more sophisticated correlations between incidences of HIV and structural issues of poverty, race, gender, sexual in equality, and comorbidity factors. The mapping of the geographies of HIV has provided powerful evidence of the ongoing purchase of health inequalities and testi mony to the importance of medical geography to public health interventions.

Health Geography of AIDS

While medical geography approaches mapped the dis tribution and diffusion of the HIV virus, those working in health geography have interrogated the experiential di mension of living with HIV and have focused explicitly on the cultural and political geographies of AIDS. The starting point for much of the health geography response to AIDS has not been the HIV itself but rather the communities most affected by AIDS. Given the history of the emergence of AIDS in developed countries it is not surprising that a focus on AIDS was a common feature of the emergent geographies of sexuality literature that began in earnest in the mid 1990s. Health geographers working on AIDS, in both developed and developing countries, have stressed that an adequate response to the AIDS crisis requires confronting medical geographies erstwhile disease with sexual practice, and in particular those sexual practices that are most culturally marginal (e.g., commercial sex, public sex, and homo sex). Health geographers, as well as others informed by Queer The ory, have stressed the need to center the experiences of AIDS sufferers not only to deconstruct the stigmatizing categories of risk groups but also to foreground and empower the communities of those most affected. Thus health geography approaches to AIDS have tended to focus not on HIV but on the politics of epidemiological surveillance, the ongoing purchase of structural in equality with respect to access to healthcare and explicit prevention knowledge, and to the new relationships to space and place opened by PLWHA.

Future

These characterizations of the broad differences between medical and heath geography approaches to HIV revolve largely around issues of scale and method. Yet, despite their differences, each tradition confirms that AIDS continues to index social exclusion and political mar ginalization. While AIDS is no longer a disease ex clusively associated with male homosexuality, this early association continues to inform the stigma attached to HIV infection and to hamper HIV prevention efforts and the provision of health services to PLWHA. Further more, the increase in the number of PLWHA means new challenges both for the provision of care and for the ongoing efforts toward HIV prevention. Most AIDS service organizations engage in substantial programs designed to overcome discrimination against PLWHA. These programs aim to integrate PLWHA as productive members of the broader society and as such necessitate the overcoming the stigmatized image of PLWHA as Other. Insofar as this politics of inclusion surrounding PLWHA is new – because it is the outcome of effective HAART – it presents as yet unexamined challenges for ongoing efforts of HIV prevention. HIV prevention strategies have always relied on an Othering of the PLWHA. Within the representational logics of pre vention the PLWHA must be seen as a space of differ ence, fear, and death and thus as a space to be avoided. Indeed, the oscillation between representations of PLWHA as a site around which arguments to inclusion and sameness, 'and' as a site around which exclusion and Othering are mobilized has taken on a particular urgency since the turn of the millennium as is evidenced in the increasing ubiquity of the criminal prosecution of those found to be knowingly spreading HIV. Indeed, one can witness in many developed countries increased biopoli tical regulation and surveillance of PLWHA.

To some extent AIDS has always been about the temporal; on the one hand it has been about the tragic loss of life and on the other hand it has generated new relationships to the future and to the time at hand. As Rob Wilton and Michael Brown's work on living with HIV demonstrated, AIDS portended something like what Queer Theorists Judith Halberstam and Lee Edelman call Queer Time; a new modality of existence between present and past, where the here and the now assumes a fullness of possibility. However, in the context of developed countries these new temporal relationships signal a profound impasse between the politics of in clusion and the politics of HIV prevention.

The future challenge for geographers remains "to find ways to integrate the global, national and state level analyses of viral distribution and diffusion with the local specificity of org[a]nizational response and of individual life experience" (Kearns 1996: 129).

See also: Gay Geographies; Health Geography; Identity Politics; Medical Geography; Queer Theory/Queer Geographies; Sexuality.

Further Reading

Brown, M. (1995). Ironies of distance: An ongoing critique of the geographies of AIDS. *Environment and Planning D: Society and Space* 13, 159 183.

Brown, M. (1997). *Replacing Citizenship: AIDS Activism and Radical Democracy*. New York: Guilford.

Cochrane, M. (2004). *When AIDS Began: San Francisco and the Making of an Epidemic*. New York: Routledge.

Crimp, D. (2002). *Melancholia and Moralism: Essays on AIDS and Queer Politics*. Cambridge, MA: MIT Press.

Edelman, L. (2004). *No Future: Queer Theory and the Death Drive*. Durham, NC: Duke University Press.

Epstein, S. (1996). *Impure Science: AIDS Activism, and the Politics of Knowledge*. Berkeley: University of California Press.

Farmer, P. (2001). *Infections and Inequalities: The Modern Plagues*. Berkeley: University of California Press.

Fee, E. and Fox, D. (eds.) (1992). *AIDS: The Making of a Chronic Disease*. Berkeley: University of California Press.

Gould, P. (1993). *The Slow Plague: Geography of the AIDS Pandemic*. Oxford: Blackwell.

Halberstam, J. (2005). *In a Queer Time and Place: Transgender Bodies, Subcultural Lives*. New York: New York University Press.

Kearns, R. A. (1996). AIDS and medical geography: Embracing the other? *Progress in Human Geography* 20(1), 123 131.

Shannon, G. W., Pyle, G. F. and Bashur, R. L. (1991). *The Geography of AIDS*. New York: Guilford.

Stoller, N. E. (1998). *Lessons from the Damned: Queers, Whores and Junkies Respond to AIDS*. New York: Routledge.

Takahashi, L. M. (1997). Stigmatization, HIV/AIDS and communities of color: Exploring response to human service facilities. *Health and Place* 3(3), 187 199.

Treitcher, P. (1999). *How to Have Theory in an Epidemic: Cultural Chronicles of AIDS*. Durham NC: Duke University Press.

Wilton, R. (1996). Diminished Worlds? The geography of everyday life with HIV/AIDS. *Health and Place* 2, 1 17.

Relevant Websites

www.kff.org
Kaiser Foundation HIV/AIDS timeline.

www.pbs.org/wgbh/pages/frontline/aids
The Age of AIDS.

www.unaids.org
United Nations AIDS Organization. Uniting the world against AIDS.

HIV/AIDS in Developing Countries

W. T. S. Gould, The University of Liverpool, Liverpool, UK

Glossary

Acquired Immunodeficiency Syndrome (AIDS) The 'A' means that the virus is acquired through a person's action rather than being spread through casual or inadvertent contact. 'I' and 'D' (immunodeficiency) identify that the virus attacks a person's immune system and makes it less capable of fighting other infections. 'S' identifies a syndrome, several medical conditions that come about as a result of immunodeficiency, rather than a specific disease.

Antiretroviral Therapies (ARVs) AIDS has no cure, though there is the long-term possibility of creating a vaccine. Current medical therapies involve a drug regime of ARVs to directly attack the virus. Since the virus mutates quickly, a range of ARVs is available to deal with different stages in the progress of the virus. ARVs have proved to be very effective in controlling the effects of the disease.

Demographic and Health Surveys (DHSs) A global program of national population surveys, financed by USAID and begun in the 1980s, to undertake national and comparative studies of demographic change. Since 2002, many national surveys have taken a blood sample to identify the sero-status of the national sample.

Epidemic Occurs when the prevalence of a disease reaches unexpectedly high levels in a short time. A disease is endemic where it is continuously present in a population, and is described as a pandemic where the epidemic assumes global dimensions, as with contemporary HIV/AIDS.

Epidemiology The scientific study of the distribution and spread of disease based on statistical analysis of medical data.

Human Immunodeficiency Virus (HIV) Member of a family of retroviruses, with several substrains, all developing slowly to affect the immune system. It is a retrovirus which destroys CD4 cells that protect humans from a range of pathogens – that is, infections that can be viral, bacterial, or fungal. When the CD4 cell count falls as a result of being HIV+, the body loses immunity and is therefore more prone to infections brought by other diseases.

Mainstreaming The approach in government planning for HIV/AIDS in which the care and cure of those infected, and the educational and support needs of those affected are explicit and become central to all sectors of activity and associated funding.

UNAIDS Founded in 1996 and with its headquarters in Geneva, this is a joint coordinating body for ten UN agencies, providing funding and leadership in tackling HIV/AIDS.

Introduction

HIV/AIDS is a global issue. Not only does it have a global distribution, however uneven, but it also has an entire UN agency devoted to it, UNAIDS. It is the focus of one of the UN's Millennium Development Goals (no. 6: "to halt and reverse the spread of HIV/AIDS by 2015"), and billions of dollars have been allocated to its potential cure and to more immediate alleviation of its effects. HIV/AIDS is a new disease, known to science only since 1981. However, it has now been diagnosed in nearly every country of the world. Its causes are known and effects felt by a large proportion of the world's people, although to a very variable extent. AIDS is a 'syndrome' caused by the presence of the HIV virus in the blood, rather than a specific condition, but the presence of that virus can have a range of primary causes: through heterosexual vaginal sex, the dominant cause, but also through male–male intercourse. The latter had been a dominant cause at the beginning of the epidemic in developed countries. However, even in developed countries, as the epidemic has developed, heterosexual transmission has become much more important. In some countries, particularly in Eastern Europe and the former Soviet Union, the growth in HIV/AIDS is associated with intravenous needle use (but even here the relative importance of needle use origins is falling relative to the growth in sexually transmitted origins). In China, some major outbreaks have been attributed to infected blood products. HIV/AIDS is a very different disease from those diseases that are more familiarly the focus of traditional geographies of disease, where there are major environmental controls on the distribution and prevalence – as in the case of malaria – or there are major poverty controls and correlates – as in the case of cholera. Though the fundamental biomedical etiology of HIV/AIDS is now clear, its distribution, prevalence, and spread are primarily behavioral, related to culture and human relationships as much as, or even rather than, to environment or poverty.

On the global scale, however, there is at first glance a clear relationship with poverty and the disease environment. Africa south of the Sahara, with 11% of the global population, had 63% of the estimated 36.9 million people globally who are HIV+, and accounted for 70% of the 2.6 million AIDS deaths in 2006 (**Table 1**). Africa's estimated adult prevalence rate of 6%, overwhelmingly attributed to heterosexual intercourse, greatly exceeds that of any other region, the next largest being the Caribbean region at 1.2%, but with a very small population in gross terms, and all other world regions have rates of less than 1%. Most rapid recent growth in numbers affected has been in the two demographic giants, China and India, but prevalence rates in these countries remain below 1%, and there is a general consensus that an epidemic on the scale of eastern and southern Africa is unlikely in these countries. India now has an estimated over 5 million people with HIV, the highest number in any country, estimated to have over taken South Africa in 2006, but it remains distinctly urban in its distribution and focused on sex workers, truckers, and other high risk groups, and rural society remains distinctly narrow and confined in its family and sexual relations. Absolute numbers in China have been lower and outbreaks more localized, but there may be greater po tential than in India in such a rapidly changing and ur banizing society for a more generalized epidemic.

In other developing countries where there is relatively high prevalence, the growth in numbers has stabilized. In Brazil, for example, there has been a great effort at control through increased use of condoms and wide spread availability of ARVs. In Thailand and Cambodia, the relatively high rate of over 2% has been associated with high rates of commercial sex and a large number of sex workers. Falling rates to below 2% in both countries by 2005 have been largely associated with greatly in creased use of condoms in commercial sex transactions, led by major public information campaigns.

Not only is Africa very poor, but it also has the highest overall disease burden, partly for environmental reasons of high bacterial growth in hot and at least seasonally moist climates. While the disease burden certainly con tributes to the high rates in Africa, it is clearly not wholly responsible, for within Africa there is a wide range in HIV prevalence (**Figure 1**). Highest rates by far are in southern Africa, rising to over 30% in Botswana and Swaziland and to ~20% in South Africa, countries that are neither the poorest nor the most environmentally hazardous – largely upland plateaux rather than hot/wet coasts. In West Africa, by contrast, where the overall disease burden is very high, HIV prevalence rates are generally below 4% (only Cote d'Ivoire is over 4%), and these rates are below 1% in North Africa.

Within most African countries, urban rates typically tend to be approximately double those of rural rates, and socially AIDS is not particularly a disease of the poor: indeed in the early phases of the epidemic, it had been associated with the elite – the wealthy and mobile were, and to some extent still are, the highest risk groups. The 2003/4 HIV/AIDS Indicator Survey in Tanzania, for example, identified national prevalence rates of 7.7% for women and 6.3% for men. However, the rate for women in the lowest wealth quintile was 2.5% but 11.8% for the richest quintile; rates for men in the two quintiles were 4.1% and 9.4%, respectively. As the disease has general ized from the initial high risk groups to whole popu lations in the worst affected countries, inevitably the poor have become more affected: as a result of their chronic poverty and associated high general disease burden and often open sores, they are less able to resist infections. Once infected, they are unable to access appropriate therapies, either because they are simply not available in their area, or else, and increasingly likely, access is ra tioned by price in impoverished public health systems.

Since spatial patterns of HIV/AIDS infection and prevalence are highly differentiated and strongly skewed, on whatever scale, any overview of the geography of HIV/AIDS in developing countries must start with issues of difference, not only in terms of rates and patterns, but also in terms of causes, management, and implications. This article therefore considers HIV/AIDS in developing countries under four conceptual issues of broad concern to analysts, including geographers: diffusion and spread;

Table 1 HIV and AIDS levels by World Region, 2006

	Adults and children living with HIV	*Adults and children newly infected with HIV*	*Adult (15 49) HIV prevalence (%)*	*Adult and child deaths due to AIDS*
Sub-Saharan Africa	24 700 000	2 800 000	5.9	2 100 000
Middle East and North Africa	460 000	68 000	0.2	36 000
South and SE Asia	7 800 000	860 000	0.6	590 000
East Asia	750 000	100 000	0.1	43 000
Latin America	1 700 000	140 000	0.5	65 000
Caribbean	250 000	27 000	1.2	19 000
E. Europe and Central Asia	1 700 000	270 000	0.9	84 000
Western and Central Europe	740 000	22 000	0.3	12 000
North America	1 400 000	43 000	0.8	18 000
Oceania	81 000	7100	0.4	4000
Total	39 500 000	4 300 000	1.0	2 900 000

Source: UNAIDS (2007).

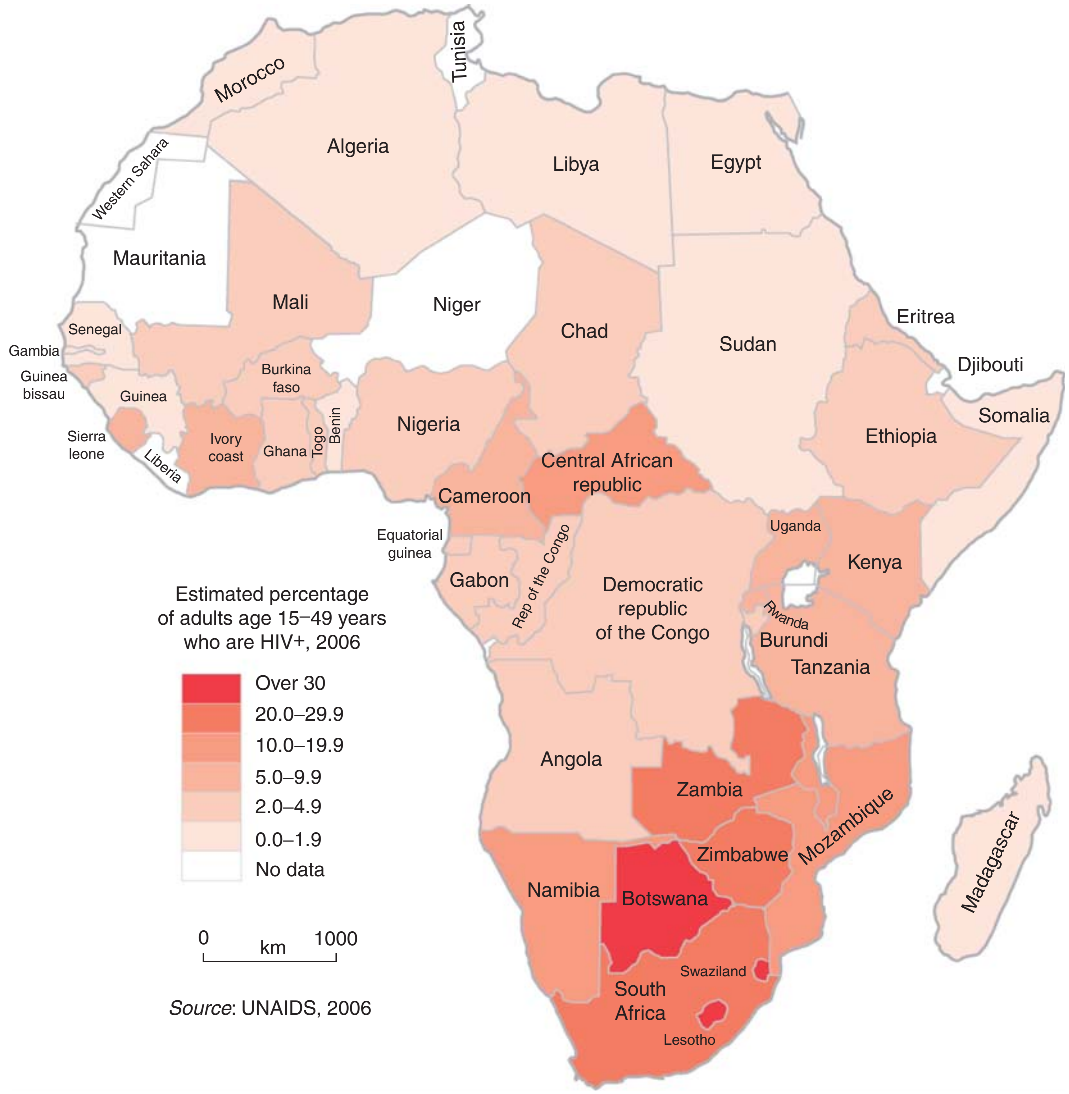

Figure 1 HIV prevalence estimates for Africa, 2006.

exceptionalism and moral panic; development impacts; and management.

An analysis of each of these issues for developing countries needs to be aware of technical problems of data. Being HIV+ is generally asymptomatic – in the early stage of the infection, carriers will not know that they are affected, normally for several years, and not until the onset of AIDS, with its fevers, skin rashes, and weight loss and lethargy, will individuals recognize they are ill and should seek treatment. Even then the stigma of the dis ease is such that, even if healthcare is available, it may not be accessed. Sure case identification is only possible with a blood test. This had in the past been gathered in clinics, with numbers and prevalence estimates based on tests given to pregnant mothers in antenatal clinics and to other clinic patients. Clinic based recording of HIV prevalence was important in the early years of the epi demic in establishing when the epidemic was growing, especially in sub Saharan Africa, and UNAIDS national estimates and comparisons of national prevalence have, in the past, been based on these clinic sources. These are now being superseded in many countries by population based sources which are intrinsically more accurate, based as they are on the whole population (or a systematic representative sample of it) and not a biased sample of those who attend clinics and hospitals. Population based HIV data have been pioneered as part of the Demographic and Health Survey global program, where those in the sample are offered a blood test at the time of interview. Population based data are not only more reliable, but have generally demonstrated that prevalence rates are rather lower than previous clinic based data had estimated. These are now being integrated into UNAIDS national data and projections, and the previously patchy and flawed data systems are now rapidly improving. Nevertheless, UNAIDS still cautions care on reliability of HIV/AIDS estimates and typically provides ranges of possible values for numbers and rates.

Diffusion and Spread: 'Beyond Epidemiology'

A primary analytical tool of established geographies of disease has been epidemiological models rooted in spatial diffusion theory, with major studies of measles, influenza, and cholera, among others, using models that conceptualized the disease in question being spread contagiously, whether by water (as in cholera), by air (as in tuberculosis), or by a vector (such as the anopheles mosquito and malaria or rats and plague). Early work on diffusion of AIDS followed in this epistemological mould with models of the hierarchical spread of the disease in USA, and elsewhere. The work of Peter Gould in the 1990s epitomized this approach. There were certainly constraints to direct contagious diffusion of the virus that could be contracted only through contact with infected blood, normally as a result of sexual intercourse, but such modeling at the time led to grossly inflated predictions of the likely extent and patterns of spread, and particularly in Africa where, with highest current prevalence rates, the probabilities for a general spread, 'like ripples in a pool', were greatest.

We now know that such modeling greatly exaggerated the potential for and patterns of spread even in the highest prevalence countries. The work of the Australian demographer Jack Caldwell and colleagues in Africa clearly showed the patterns of the disease to be not only skewed and focused, but showed them to be much affected by cultural and social variables, including sexual networking and male circumcision. The number of sexual partners, varying culturally and over time within any one culture, will clearly affect patterns and rates of spread, and can be responsible for substantial variation locally between households. High risk groups are those with multiple sexual partners, such as sex workers, truck drivers and migrants generally. These groups have typically been targets both for popular resentment against the carriers of the disease in a culture of blame, but also, and more benignly, for education programs in safe sex, especially condom use. Male circumcision seems, surprisingly, to have the effect of raising immunity to HIV infection, and ethnic differences in prevalence rates often reflect this cultural practice in Africa (e.g., between the higher rates for the noncircumcising Luo men of Western Kenya compared with those of their circumcising Luhya neighbors).

Even in the highest prevalence countries of eastern and southern Africa, from Ethiopia southwards, where the epidemic has spread from a strong focus within high risk groups into the population more generally, that spread remains spatially differentiated. There is a distinctive urban bias that, despite the persistence of circular migration and continuous interaction between rural and urban areas that might have been expected to promote a spread of the virus into rural areas, shows little sign of narrowing. In Uganda, where national prevalence rates famously more than halved during the 1990s, the urban prevalence remains at approximately double the rural. There were declines in both rural and urban areas, associated largely with a sharp reduction in the number of sexual partners.

It is clear that understanding the geography of HIV/AIDS in developing countries needs to go 'beyond epidemiology'. Established epidemiological modeling, rooted as it is in a biomedical model of disease transmission and control, seems inadequate for a disease that is largely behavioral in its incidence – who is affected and who is not? Vulnerability to HIV/AIDS is not merely a matter of exposure and distance, but must relate to broader issues of culture and society, of power relations and historical experience, of entitlements and economic relations within and between households.

Exceptionalism and Moral Panic

It is partly as a result of that failure of established medical practice to deal fully with HIV/AIDS that it has been elevated to the level of a moral panic for global as well as national societies. It is a disease that as yet has no cure, but everyone seems at risk, not only the poor or specifically exposed. Exposure is as a direct result of normal adult behavior – sexual activity, an integral component of human relations in all societies. Absolute avoidance would require lifetime abstinence even within marriage! In the absence of a medical cure, behavior change seems necessary. Current therapies are those of control rather than cure, largely ARVs, that can suppress the effects of the virus to allow people to continue indefinitely with their normal activities in the best case scenario, but they do not destroy it. ARVs also need to be complemented by behavior change for long term survival and to avoid resurgence of the debilitating effects of the virus. The standard behavioral message, associated in particular with USAID rhetoric and associated programs, is for ABC programs – 'abstinence', 'be faithful', and use a 'condom'. Such messages have limited effectiveness, and need to be directed mostly to the adolescent, as yet unmarried population. For them, abstinence is more of an option than it is for married couples, and reducing the number of sexual partners and remaining monogamously faithful certainly helps. Although increasing condom use has become a necessary component of programs funded by USAID, there are problems in using condoms that are not merely about availability and cost – major issues in rural Africa – but also about their social acceptability where their use in the past has been very much associated with prostitutes and transactional sex, and these are not the norm for most people.

Gould and Woods have argued that HIV/AIDS is a "wholly exceptional disease" in medical, demographic, and behavioral terms. Medically it is not a single disease, not a single cause of death or illness, unlike the more familiar diseases of developing countries such as malaria or tuberculosis. People with AIDS do not die of this condition directly, but are likely to die of tuberculosis, pneumonia, or other infections, and would be recorded as such where the death is in the domain of formal medical care. Furthermore, because being HIV+ is asymptomatic for most of its course, typically for several years, it is not normally felt by carriers to be a problem, even if they know their sero status.

Demographically, HIV/AIDS is fundamentally dif ferent in its effects from most other major diseases in developing countries, in that it causes illness and death primarily in adults; in most countries in more women than men; in the sexually active, rather than in children. Since there is an immediate link to the economic and productivity declines as a result of the loss of labor through morbidity (illness) as well as mortality (death), HIV/AIDS has been of much greater concern to gov ernments and international agencies such as the World Bank and the IMF than more established, but less im mediately economically important yet nevertheless so cially devastating, diseases. AIDS has attracted much more direct funding and international attention than any disease previously. However, this demographic effect is felt most immediately at the level of the household, in farm and other economic activities, and in the human and other resources needed for caring for the long term sick.

The exceptional behavioral dimensions of the disease have been identified above. Short of total sexual abstin ence, risk cannot be avoided, as it can with the other main diseases whether directly with vaccinations (e.g., for polio or measles or TB) or in other avoidance behavior (e.g., use of bed nets to prevent mosquito bites for mal aria), or public health measures (e.g., ensuring clean water supplies). Many traditional behaviors (e.g., re peated use of cutting instruments in cleansing rituals, sexual access to widows) seem to encourage infection and spread of HIV/AIDS, and because the condition and its effects are still new, coping mechanisms to mitigate the effect of these behaviors are not yet sufficiently in place. In Malawi, for example, not only does HIV/AIDS ex acerbate rural impoverishment, but rural impoverishment in its turn exacerbates HIV/AIDS through encouraging more risky sexual behavior, for different reasons, among men and women. In Eastern Zimbabwe, by contrast, in a context of sharply deteriorating economic and health conditions, there is evidence of behavioral change re sulting in lower prevalence rates.

Together, these amount to a powerful case for seeing the epidemic as really exceptional rather than yet an other development problem. The most powerful pro ponent of the exceptionalist argument is Peter Piot, Executive Director of UNAIDS. He sees the epidemic essentially as a Malthusian event, one that can lead to major regional or even global collapse of the economic system as a result of production failures and demographic collapse, but one that needs to be reversed as a major priority of the global community. In these terms, it is unprecedented as a disease epidemic, well beyond the effects of the Great Plague of the fourteenth century in Europe, and ranks with climate change and global warming as a medium as well as a long term threat to human survival, but one that can and needs to be man aged to minimize that threat. For others, notably Alan Whiteside, the leading South African social scientist of AIDS, its exceptional importance can be characterized as a long term and Darwinian event – one which will re quire and ill induce a fundamental change in economy and society if it is to be managed at all successfully.

By contrast, however, many of those affected at the household level have perceived HIV/AIDS as yet an other development constraint, but one that needs to be 'mainstreamed' into development thinking and develop ment planning, as 'gender' and 'environment' have been. They are all cross cutting issues that affect all sectors of economy, society, and government activity on all scales. The effects of the disease need to be explicitly con sidered in such matters as demand for schooling and healthcare, as well as in estimating labor force deficien cies and requirements. The impacts of HIV/AIDS have to be managed by the state as well as by households and communities, and this is best done by incorporating them explicitly into the established structures and processes of governance. In these terms, HIV/AIDS is not so ex ceptional that it needs to be treated differently from other major issues and concerns.

Here the contrast between Uganda and South Africa is most instructive. In Uganda, President Museveni and the whole political process from the late 1980s, when Uganda had the highest prevalence rates in the world, gave HIV/AIDS special and overwhelming prominence in govern ment activities and public statements, with frank and explicit public education campaigns on the need for safe sex and to reduce the number of sexual partners. Public knowledge and public discussion were greatly enhanced and the effect was the unprecedented experience of sharp falls in prevalence rates by the mid 1990s. These have remained low. Ugandans, in rural and urban areas, have changed their behavior in response to what was seen and accepted as a severe threat to the established social order, and this is in sharp contrast to rising rates in neighboring Kenya at the same time. In South Africa, by contrast, where the epidemic since 2000 has greatly exceeded anything experienced in Uganda, political support, led by President Mbeki, has been distinctly lukewarm about conceptualizing HIV/AIDS as anything other than yet another problem that is a byproduct of poverty, inequality and a history of oppression and repression, and will be overcome through progress to the elimination of poverty and inequality. Turning moral panic into a learning opportunity, as in Uganda, needs explicit political commitment.

Impact on Development

Cynically, HIV/AIDS might be identified in the poorest countries of Africa as a resource generating event. There has been a massive resource input from international public (multilateral and bilateral aid programs) and private funds (from NGOs and such organizations as the Bill and Melinda Gates Foundation) into AIDS associated research, treatment and care, to the extent that there is some discussion of the level of resources going into the disease causing some macroeconomic instability as a result of imbalance in productive and social allocations. For poor and financially stressed countries, such resource streams represent not only major new lifelines for direct mitigation of the effects of HIV/AIDS, but they are also a source of much needed infrastructural expenditure in health and social services, and often channeled through newly created agencies under direct central governmental control, such as TACAIDS in Tanzania, and then 'mainstreamed' into sectoral ministries. There is an incentive to encourage AIDS related activity as a source of funding, bringing funding into a wider range of activities where there is some AIDS relevant aspect, such as into development of school curriculum and production of learning materials where there is an AIDS awareness program for young people.

There is little doubt that HIV/AIDS has a major adverse effect on levels of economic activity on the national scale. There is a loss of scarce manpower as, disproportionately, it is the educated and skilled workers, including in the key mining sector in southern Africa, that have been most affected. There have also been major impacts on teachers and health service personnel, especially vulnerable in the earliest phases of the epidemic, but these are the groups most likely to be influenced by public education messages to change their behavior. As the epidemic progresses, infection rates for teachers, for example, tend to equal or even fall below the rates for the general population. What is clear is the need for increased training budgets of government and private agencies to replace these workers.

However, it is on the household scale that the economic effects are most immediately apparent with the loss of the labor of men and women, on the farm, in off farm economic activities, and in domestic life, due to prolonged illness as well as death. This loss of labor has had two sorts of effects. One is a commensurate fall in production and income and often immediate impoverishment as crops cannot be planted, weeded, harvested, or marketed, with land then reverting to bush or else taken over by other less affected households with surplus labor. The other, and the more common but less noticed effect, is that there is a reallocation of labor demands within the household, with additional demands for the labor of children and older people who might otherwise have been less directly active. The effects on children have been well documented, and particularly on AIDS orphans, who are more likely than other children to drop out of school in order to provide farm labor or to care for sick adults, to have lower levels of educational attainment, and generally confirm the longer term impoverishment of the household as they are less able to become involved directly in income generating activity.

HIV/AIDS cannot be realistically separated or measured as an independent development force, but it is completely pervasive in exacerbating household and also national economic and social deterioration. The AIDS epidemic will add to the problems of any international efforts to rejuvenate African economies, and will further exacerbate the widening development gap between Africa and other regions of the developing world.

Management

Despite the difficulties associated with the exceptional nature of HIV/AIDS, it does need to be managed directly, on all scales. At the local level, there is a traditional coping mechanism that works – high mortality is not new, though with AIDS its scale is well above normal risk levels – and traditional coping can, as a minimum, mean struggling to making the best of a bad job as the household descends, often very quickly, into poverty as labor is withdrawn, directly due to illness and indirectly to care for those who are ill. Assets, including land, may be sold to pay for the costs of care or for food that might have been previously grown on the farm, and a downward spiral of impoverishment sets in. More positively, coping mechanisms may involve ensuring additional labor is made available to sustain economic activity. This may be by the younger or the older members of the household or else other kin, for there is a tendency for increased household dissolution and increased fostering of fit young adults into affected households. There is evidence that households on average get larger, but that there are fewer of them, as larger households seem more able to withstand the production and income shocks associated with sudden and long term labor loss. As Rugalema pointedly asks, "is this coping or struggling?".

On the national scale, management of the disease as a health problem requires greatly additional resources in the health sector, but in the worst affected countries health systems are overwhelmed with often over 50% of hospital beds occupied by AIDS sufferers. Even the minimum palliative care that can be offered to them imposes high burdens on public provision; little wonder that minimal care at home seems to be the preferable option for many. Since the early 2000s antiretroviral therapies have become increasingly available – both in gross quantities, but, more importantly, with sharply falling prices, as a result of global political action and support eventually through WHO. ARVs offer no cure, and there are medical prob lems about their complexity for those who have access to them, but they can make, and have made, a great differ ence to many people in poor countries, as they have in the global North. In some other richer developing countries with relatively low prevalence, ARVs are available to all AIDS sufferers and have become a prime mechanism of management. Brazil was the first developing country to offer universal access to ARV treatment, as well as pre vention and diagnosis services, to anyone with advanced HIV infection. Only in Botswana, among African coun tries, however, are ARVs available free to those who need them, but even here there are problems of stigma and reporting that prevent their universal uptake by those who need them. In most African countries, there needs to be rationing as supply greatly outstrips demand, and the ra tioning process in an open market becomes one of en titlements where the urban and affluent are more able to command better access than the rural poor.

In the international community, and particularly within UNAIDS and its advocacy, HIV/AIDS has to be seen as a development problem perhaps even more than as a health problem. There is now a direct and systematic concern for scaling up, as in the relatively successful campaign to have 3 million people in Africa with access to treatment by the end of 2005, and more generally to move 'from crisis management to sustained strategic response'. This is evident, for example, in UNAIDS identifying three scenarios for the states of HIV/AIDS in Africa by 2025, and the need for positive and systematic response to minimize the number of infections and deaths. There are clearly three major dimensions of the management system that are needed:

- more resources, in billions rather than millions of dollars, on a scale that is unlikely by itself to manage the disease, whether through increased use and ef fectiveness of ARVs or resources allocated to finding a cure or vaccine;
- political commitment, as in the Ugandan case, raising awareness of the scale of the problem and individual responsibility through behavior change in tackling it; and
- behavioral change, whether from above, from political pressure and the provision of the state, or from below, by changing norms for levels and patterns of sexual activity.

See also: Africa; Aid; Disease Diffusion; Disease Mapping; Health and Development; Poverty; Vulnerability.

Further Reading

Agyei Mensah, S. (2005). The HIV/AIDS epidemic in sub Saharan Africa: Homogeneity or heterogeneity? *Norwegian Journal of Geography* 59, 14 25.
Barnett, T. and Clement, C. (2005). HIV/AIDS impact: So where have we got to and where next? *Progress in Development Studies* 5, 237 247.
Barnett, T. and Whiteside, A. (2002). *AIDS in the twenty first century: Disease and globalization*. Houndmills, UK: Palgrave.
Caldwell, J. C. (2006). Will HIV/AIDS levels in Asia reach the level of sub Saharan Africa? *Asia Pacific Population Journal* 21, 3 9.
Caldwell, J. C. and Caldwell, P. (1994). The nature and limits of the sub Saharan African AIDS epidemic: Evidence from geographic and other patterns. *Population and Development Review* 19, 817 848.
De Waal, A. (2003). How will HIV/AIDS transform African governance? *African Affairs* 102, 1 23.
Dyson, T. (2003). HIV/AIDS and urbanisation. *Population and Development Review* 29, 427 442.
Gould, W. T. S. (2005). Vulnerability and HIV/AIDS in Africa: From demography to development. *Population, Place and Space* 11, 473 484.
Gould, W. T. S. and Woods, R. I. (2003). Population geography and HIV/AIDS: The challenge of a 'wholly exceptional disease'. *Scottish Geographical Journal* 119, 265 281.
Kalipeni, E., Craddock, S., Oppong, J. R. and Ghosh, J. (eds.) (2004). *HIV and AIDS in Africa: Beyond epidemiology*. Oxford: Blackwell.
Piot, P. (2006). AIDS: From crisis management to sustained strategic response. *The Lancet* 368, 526 530.
Rugalema, G. (2000). Coping or struggling? A journey into the impact of HIV/AIDS in southern Africa. *Review of African Political Economy* 86, 537 545.
UNAIDS (2005). *AIDS in Africa: Three scenarios to 2025*. Geneva: UNAIDS.
Zhang, H. (2004). The gathering storm: AIDS policy in China. *Journal of International Development* 16, 1155 1168.

Relevant Websites

http://www.tanzania.go.tz
Tanzanian AIDS Commission (TACAIDS)
http://www.tasouganda.org
The AIDS Support Organisation, Uganda.
http://www.aidsuganda.org
Uganda AIDS Commission.
http://www.unaids.org
UNAIDS.
http://www.liv.ac.uk
Understanding Epidemics.
http://web.worldbank.org
World Bank.
http://www.who.int
World Health Organisation.

Home

T. Peil, Tallinn University, Tallinn, Estonia

Glossary

Dwelling The capacity to achieve spiritual unity between humans and things; dwelling expressed in repeated encounters and complex associations with a place, including practical work; also identified as one way to think spatially about culture.

Family It is defined in various ways from a primary social unit (household) to kinship ties; some have argued that the ties between home and family are so strong that the terms are interchangeable (as in birth family, family of origin). Critics argue that traditional nuclear family is limiting and ideologically laden and other household forms are pertinent to the construction of home.

Habitation It is connected to 'rootedness' and sense of place; demonstrates geographical subject's commitment to place by considering one place as home.

Home A material and an affective space, real or imagined, that is formative of personal and national identity, shaped by everyday practices, lived experiences, social relations, memories, and emotions.

Homelessness It is the condition and societal category of people who lack a regular, safe, and adequate shelter; the term also stands for a condition of abjection, of detachment from society characterized by the lack of affiliative bonds.

Housework The work that ensures the smooth running of the domestic economy.

Private Sphere The space where one has the capacity to establish and control personal boundaries.

Rootedness An unself-conscious feeling of belonging (of being 'at home'); denotes home as a source of identity and meaning in the world.

Topophilia A notion coined by Yi-Fu Tuan indicating 'love of place' or the idea that a place can bring into being deep feelings of value and caring, that is, not only an affective response to place but also a practice that can produce places for people, conscious appreciation of place.

Introduction

Home or its equivalent can be considered one of the most powerful words in any language; the etymology of the word (from the origin of '*kei*' or '*ham*') indicates its common roots in many European languages. A person's home is, in Western culture, usually understood to be situated in space and time, a place to lay one's head, but it becomes 'home' when it is inscribed with (social, psychological, and emotive) meaning for individuals and groups. The meaning is, however, personal, and therefore different people are likely to understand 'home' to mean different things at different times. Home is thus a highly fluid and a contested site of human existence that reflects and reifies identities and values. Researchers have been weary of the use of the term exactly because of the weight of the meaning in everyday life and the ambiguity. Since the home is interconnected with many, if not all, major concerns in many research fields, no one specific definition is agreed upon. In fact, some academics warn against narrowing the usage in the name of intellectual rigor for the fear of losing much of the social meaning. Some have dismissed the home as private, trivial, and mundane compared to the public realm, but increasingly research is dedicated to the critical understanding of the domestic sphere. The wide range of fields includes anthropology, architecture, environmental and gender studies, ethnography, history, human geography, literary study, philosophy, and sociology. Research traditions and main approaches for examining the home include housing studies; home and gender; phenomenological studies of the meanings of home and place attachment; and architectural and environmental design, material culture, considering the notions of being at home, creating or making home, and the ideal home.

Thinking about home can be argued to be basically geographic, highlighting relations between place, space, scale, identity, and power. The home becomes a central concern for geography when the argument is adopted that the primary way in which the geographical subject realizes the commitment to place is by means of habitation, and the home is interpreted as a place where social relations are acted out. Nevertheless, geographers have focused on the home only recently, for instance, while the fourth edition of the *Dictionary of Human Geography* does not include 'home' as a term, the last decade has witnessed an explosion of interest evident in the amount of published literature. Research of the home in geography has been spurred along by increased interest in the everyday, in the production of space in a globalizing world, in the body, as well as by queer geography and feminist understandings of the private and public, and by the analysis of economic and social power relations informed by the postcolonial theory, especially in discussing travel and domestication and the political significance of the home. Domestic space has also interested geographers in an attempt to rematerialize geography and interest in material cultures of objects and their use, display, and meanings. Analysis of the home has turned to include a study of continuous processes rather than isolated actions. In cultural and social geog raphy, the home is usually understood in spatial and political terms with three overlapping components: home as material and imaginative; the nexus between home, power, and identity; and the multiscalarity of home.

Material Home

Traditionally, geographers have been more interested in the notion of the home as the house (residence) or homeplace (dwelling) and the notions of home as one's nation, or place of origin. Home is seen as located in space but not necessarily fixed in space, although the conceptions of home have tended to privilege a physical structure or dwelling, such as a house, flat, institution, or caravan that is bounded in space with certain functional, economic, esthetic, and moral structure. The critics argue that this approach is reductionist, highlighting only one aspect of home.

Research on home ownership and more generally on the economic significance of housing and home econ omies (consumption patterns) emphasizes the inter connected issue of the relationship of home to property and boundaries. Ownership has long been closely iden tified with dominant notions of home as those buying a house are presumed to be properly capable of making a home – a place that is secure, comforting, and welcoming. Patterns in housing tenure, especially the significance of private house ownership, are related to social divisions in society and have cultural purchase in Western capitalist system. The significance of home ownership in personal life and public policy attempts also to bridge the di chotomy between the public and the private sphere.

Economic analysis of labor, including housework, paid domestic work within the home, and the intersections of production and consumption, have also challenged the somewhat narrow view of home as a private, domestic, and female realm. The gendering of both ideals and materialities of home informed by feminist theory has formed a prominent field in the home research since the 1970s focusing on issues of housework, house design, and house as an expression of status. Feminist theory is used to challenge the idea of the home as a bounded place of security and retreat, and criticizes the public–private dichotomy. Lived experiences, social relations, and emotional significance form the home as a site of per sonal history and memory, but the feminists have pointed out that the home also has political implications and agenda within the family and society at large. Arguments about home can have an ideological agenda to resist change or to promote specific views. Home spaces are seen as shaped by inclusions, exclusions, and power re lations. Geographical work in this tradition also covers examining the use of space, for example, the spatial separation of men, women, and children at home, and discussing the implications of gendered spaces, such as the kitchen, or suburbia, or home as work and its impact on national economy.

More recent research therefore views house and home as related but not conflated. In an attempt to elucidate the relationship between house and home, the latter is seen as a spatial and social unit of interaction between the indi vidual and society; a residence inhabited by family, people, things, and belongings; and a space where particular ac tivities and relationships are lived. This spatial and re lational realm is created by family, gender relations, stability, security, and dwelling ownership. Research ap proaches include studies of environmental design, archi tecture, and landscape architecture alongside material cultures of home adopting ethnographic perspective and examining the objects at home, especially paying attention to art and other forms of visual culture. Many researchers have analyzed the ways in which design and artifacts as well as other possessions/objects influence the historical and cultural ideas of home. Ideas about privacy, domes ticity, intimacy, and comfort are asserted to be the or ganizing principles for the design and use of domestic spaces. The notion of the ideal home or house in various cultural contexts and historic periods has gained con siderable attention. It tends to focus on physical structures, but has touched upon the marketing techniques employed by the developers, familial experiences, and social change shaping perceptions. People are increasingly moving away from an experience of unself conscious rootedness and toward an (image conscious) sense of place, which again has economic effects as more actors are prepared to invest in the 'right image' of home and place.

Imaginative Home

The home is the place of greatest personal significa nce and the primary site around which identities are produced and performed. Many see the link between the home and the identity as permeable: a self constitutive practice embedded in a spatial context that is constituted by social practices, while it is equally constitutive of them. Whereas the majority of authors writing about the home tends to view it as a special place; some, especially in feminist approach, argue that it is one space among many. This plurality is, nevertheless, seen to have con sequences for personal identification, and thus indicates that the home has crucial and essential meanings; that it is fundamental to being.

The home is directly experienced, a state of being, and that is argued to give home its power to elicit strong emotional response. The home is identified as an emotional warehouse connected to the individual psychological experiences and the feelings of security and well being. In this view, people are seen to feel at home when what they do has some effect and what they say carries some weight, that is, they have control and power to decide for themselves and influence others. By focusing on the home as a site of constancy, forming a spatial context in which day to day routines are performed, and a site where people feel in control, the emotive aspect of home is closely connected with power: the home starts by bringing some space under control. Phenomenological approach focuses on the individual's feelings of value, caring, and security; and memories and dreams. Here, the home is usually seen as a place that offers security, familiarity, and nurture – a bounded space beyond which lies the unknown into where people venture but hope to return to the safe haven. The social constructionist school claims that home is not just a matter of feelings and lived experiences but also of cognition and intellectual construction. Key signifiers of home are identified as shelter, hearth, heart, privacy, roots (constancy), abode, and (possibly) paradise. The authors writing in these traditions thus highlight the positive experiences and evoked emotions of home.

The critique terms such approach as simplistic and naïve, since it ignores the variety of experiences at home and meanings of home. Simultaneously, the object of the most profound feelings and the subject of the greatest nostalgia, the home is not simply recalled and experienced in positive ways. Issues of gender, ethnic or racial inequality, injustice, displacement, domestic violence, and homelessness are highlighted in pointing out the failure to recognize the politics of the domestic realm or consider the social and discursive fields that frame the experience. The critique is delivered from different angles, such as historical studies focusing on the changes in the meaning of home and feminist ones questioning the adequacy and even accuracy of the representations. Feminists first identified home as a site of oppression, where women are consigned to a life of reproductive and domestic labor with no economic control over its management. More recent research has introduced plurality and examines home as a site of resistance where repressed groups in family (women, girls) or society (minorities) can gain control over certain aspects of their lives, and thus the home is potentially a site for radical subversive activity. This can, for example, be the case in autocratic society where the home becomes a center for opposition. Discussions held at home or activities inside the home, undertaken in private, may have an effect on the public life, thus confusing the division into private and public spheres.

The meanings of home in various cultures and times have been studied with the help of fiction. Textual representations of the home in novels, magazines, and household guides have been important in writings by, for, and about women, and as such tell the story of creative containment and liberation possible within the home.

Homeland

The assumptions of linking the nation and its territory and of creating a nation's imaginary home are connected to the notion of home as homeland. Identities produced in and through places, especially home places, are defined not only in terms of attachment, but also in a large part in contrast to and against an outside world, real or imagined. The idea of the home is thus closely connected to the idea of the foreign and reflects the narratives of nation and empire. For instance, informed by postcolonial theory, novels have been studied to explore the ways in which the home in domestic scale is bound to national and imperial geographies of home.

Interconnectedness of landscape and national identity has been a subject for a rapidly increasing body of academic literature on its representations in text, film, and art, and in political and ideological discourses. In fact, the emergence of geography as a scholarly discipline in many European countries can in itself be interpreted as an expression of Romantic nationalism connecting people with their natural home. Both in the development of the nation and its geography, imaginations turned to the local environments of the homeland, where landscapes were read as indices of cultural belonging. A certain body of stereotypes – the ideal home and the home ideal – thus becomes significant in the collective mind of the nation internally, as well as externally. The imaginary homeland obtains deep significance when left behind, either when growing up or moving away: the landscapes of childhood and home are remembered with nostalgia and idealized. The image may become real in the domestication of the exotic, and may involve the transference of a whole range of objects and ideas, including property laws, renaming space, and material restructuring. The widespread home improvement campaigns of the nation building era aimed at achieving a unified imagery of the homeland. Transferring space into a landscape of home was supposed to make it simultaneously useful and esthetically pleasing, as well as to promote a more efficient use of resources and elimination of waste. In this context, the efficient organization of the home at home, the domestic setting is understood as a model for society that can be replicated especially in nation building or imperial efforts. In contrast to examining ideal homes in a house as home frame, the idealized home in this context refers to the symbolic potency of the homeland and is often discussed in romantic terms and replicated in popular images (the pastoral or the wild as in the Norwegian mountains farm, the German forest, the English lawn, or the cricket field).

Recently, as environmental sustainability has gained more attention, the home is used to explore the domestic entanglements of nature and culture and to find ways to loosen the hold of the wild–domesticated dualism. The relationships between the home and nature are explored through the material and constructed boundaries. Other themes include the ethics of the domesticated environment, private and public gardens, the embodied practice of gardening and landscaping, that is, domesticating nature. Metaphors such as the household of nature are used to discuss people's inhabiting, or, in fact, cohabitation on the planet Earth regarding human and nonhuman relations.

Spatiality and Temporality of Home

The home has a spatial dimension, it is located in space and is meaningful in a way particular to here and now. The home can be seen as unique. At the same time, the home can be studied at many levels. The multiscalarity of homes becomes evident in transnational domestic forms, such as the high rise building or the bungalow. The home is grounded in the activity that occurs in a place, but in recent research the home is rather defined as open and constructed out of movement and communication. Recent geographical research on the home focuses, thus, primarily on practice. The home is seen as the result of a complex, fluid, and contested process, which remains embodied, and is always expressed in literal and material ways. For example, the photographs taken of family members and displayed within the home, or taken of homes for various purposes provide abundant material in the study of identity and belonging. Interest is thus in the dialectical relationship between the self and the object in the intentional production of home. Social relations and attached meanings stretch beyond the physical space into the symbolic and the virtual. Spatially, the home can be understood as a mental state, a house, a neighborhood, a country (or nation), or the world as real or virtual. Examining the meaning of home and being at home in cyberspace and chat rooms, of homepages, and the use of the Internet at home has only started. It remains to be seen what consequences for identification and belonging virtual space will have.

Temporally, the focus in the research about the home has foremost been on the changing meanings of home in history and on personal and public narratives of home. Geographical and societal position, that is, residential segregation and mobility, gentrification, and zoning practices are central themes in current geographical studies, considering the home as residence in time and space. Mobility in space and over time has been a long term concern for geography, including themes such as voluntary and forced migration on temporary or permanent basis. Examples of recent research include writing house biographies that tell the story of a house as home through the lives of its past and present inhabitants. Describing the housing career of an individual or a household over time can be used to discuss physical or social mobility; the latter including the adoption of new values.

Ideas about staying, leaving, and journeying are integrally associated with notions of home. The home can be understood as the place of origin as well as a point of destination, as searching for a home is seen as a basic trait in human nature. The relationship between home and away is often represented as oppositional, but home can encompass strangers and encounters with newcomers and it is not merely the familiar base to start the journeying. It has also been argued that when one moves away from home, the movement occurs in relation to home, and as such is part of the constitution of home. Leaving home is not always about choice, and a wide range of geographical literature focus on the change made under duress, on emplacement and displacement; the loss of home due to domicile or old age that can have equally shattering consequences for a person who has lost control over the organization of their own life. Increasing attention in the global world is therefore given to transnational homes, that incorporate embodied memories of past landscapes and relationships with premigratory lives; to living in several places in a physical sense (commuting to work, second homes, and caravans), or mentally feeling or not feeling at home at your current place of residence, or feeling homeless.

In Conclusion

The home is seen to be at once concrete and abstract; it is a multidimensional concept interpreted in terms of place, feelings and emotions, practices, as a state of being, or sense of place. Rejecting the binary oppositions that were long used to define the home, the home is seen as porous and contested rather than viewed as a fixed, bounded, and static location. Like any term or experience, the home begets its own negation and a self constitutive practice. Contemporary social and cultural geography examines the home critically as produced and recast through a range of homemaking practices that exist over a range of scales. Abandoning simple spatial definitions of home as house or collection of houses and associated social amenities, home research currently includes analysis of lived experiences, social relations, and emotional significance, and traverses scales from the domestic to the global in both material and symbolic ways. The issues of space, place, and residence, virtual home on the Internet, scale and mobility, power and control, identity;

landscapes of home and memory, homeland, and en vironmental sustainability have been examined related to the research on the home.

See also: Belonging; Dwelling; Emotional Geographies; Private/Public Divide.

Further Reading

Benjamin, D. N., Stea, D. and Saile, D. (eds.) (1995). *The Home: Words, Interpretations, Meanings and Environments*. London: Avebury.

Bennett, T. (2002). Home and everyday life. In Bennett, T. & Watson, S. (eds.) *Understanding Everyday Life*, pp 1 50. Oxford: Blackwell.

Blunt, A. (2005). Cultural geography: Cultural geographies of home. *Progress in Human Geography* 29, 505 515.

Blunt, A. and Dowling, R. (2006). *Home*. London: Routledge.

Chapman, T. and Hockey, J. (eds.) (1999). *Ideal Homes? Social Change and Domestic Life*. London: Routledge.

Cieraad, I. (ed.) (1999). *At Home: An Anthropology of Domestic Space*. Syracuse, NY: Syracuse University Press.

Domosh, M. (1998). Geography and gender: Home again? *Progress in Human Geography* 22, 276 282.

Duncan, J. S. and Lambert, D. (2003). Landscapes of home. In Duncan, J. S., Johnson, N. C. & Schein, R. H. (eds.) *A Companion to Cultural Geography*, pp 382 403. Oxford: Blackwell.

Easthope, H. (2004). A place called home. *Housing, Theory and Society* 21, 128 138.

Gabriel, M. and Jacobs, K. (2008). The post social turn: Challenges for housing research. *Housing Studies* 23, 527 540.

Gathorne Hardy, F. (1999). Home. In McDowell, L. & Sharp, J. P. (eds.) *A Feminist Glossary of Human Geography*, pp 124 125. London: Arnold.

Hareven, T. (1993). The home and the family in historical perspective. In Mack, A. (ed.) *Home: A Place in the World*, pp 253 285. New York: New York University Press.

Hinchliffe, S. (2003). 'Inhabiting' landscapes and natures. In Anderson, K., Domosh, M., Pile, S. & Thrift, N. (eds.) *Handbook of Cultural Geography*, pp 207 225. London: Sage.

Massey, D. (1992). A place called home. *New Formations* 17, 3 15.

Mezei, K. and Briganti, C. (2002). Reading the house: A literary perspective. *Signs* 27, 837 846.

Miller, D. (ed.) (2001). *Home Possessions: Material Culture behind Closed Doors*. Oxford: Berg.

Painter, C. (ed.) (2002). *Contemporary Art and the Home*. Oxford: Berg.

Porteous, J. D. and Smith, S. E. (2001). *Domicide: The Global Destruction of Home*. Montreal: McGill Queen's University Press.

Rybczynski, W. (1986). *Home: A Short History of an Idea*. New York: Penguin Books.

Tuan, Y. F. (2004). Home. In Harrison, S., Pile, S. & Thrift, N. (eds.) *Patterned Ground: The Entanglements of Nature and Culture*, pp 164 165. London: Reaktion Books.

Relevant Websites

http://www.enhr.ibf.uu.se
European Network for Housing Research.

http://hearth.library.cornell.edu
Home Economics Archive, Cornell University, USA.

http://www.york.ac.uk
Housing Studies Association, UK.

http://www.jchs.harvard.edu
Joint Center for Housing Studies, Harvard University, USA.

http://www.massobs.org.uk
Mass Observation Archive, University of Sussex, UK.

http://www.indivisible.org
Stories of American community, the Center for Documentary Studies at Duke University in partnership with the Center for Creative Photography, The University of Arizona.

http://www.artscene.org
Stories of home, Lincoln Arts Council.

Homelessness

J. May, Queen Mary, University of London, London, UK

Glossary

Hidden Homeless Those experiencing forms of homelessness that are not immediately evident to others – for example, staying in 'single-room occupancy/bed and breakfast' hotels, in vehicles, or squats, living unwillingly with friends or relatives, 'hot bedding' or 'sofa surfing'.

Homelessness A continuum of accommodation need stretching from a state of absolute rooflessness to severe housing stress or 'proto/incipient' homelessness. It may also usefully be defined along an emotional/experiential register, according to whether a person enjoys the feelings of emotional attachment, safety, and security (more usually) associated with a sense of 'home'.

Poverty Management The organized response by the state and elites to manage and contain potentially disruptive populations, including the mentally disabled, criminals, and homeless people.

Shelter/s Accommodation set aside specifically for homeless people on a temporary basis. The form and quality of shelter accommodation varies widely – ranging from large, dormitory-style establishments to single rooms or family units. In North America and Western Europe, most shelters are provided by voluntary sector/not-for-profit organizations and funded through a combination of state funding and public donation.

Skid Row A distinctive district in which accommodation and services for homeless people are concentrated. Found mainly in North America, where skid row districts have come under severe pressure from central-city redevelopment and gentrification in recent years, such concentrations are less evident in European cities.

Stigma A social process leading to the devaluation of an individual or groups who come to be identified as embodying negative traits. Human geographers are especially interested in the role that spatial metaphors play in demarcating stigmatized groups, and the spatial practices by which such groups come to be separated from mainstream society.

Visible Homelessness Those experiencing absolute rooflessness or making use of accommodation facilities (shelters and hostels) provided specifically for homeless people – and who are hence 'visible' to members of the housed public, social service agencies, and others.

Introduction

Following a dramatic rise in levels of homelessness through the 1980s and 1990s, the last 20 years or so have seen a steady proliferation of work exploring the geographies of homelessness. The majority of this work has focused on the problems of street homelessness, usually in urban settings, almost all of it in an Anglo American context. Early work in this field examined the causes of homelessness, the politics of shelter location, and street survival strategies. More recently, homelessness has emerged as a favored lens through which to examine broader themes in contemporary urban geography. Thus, ever more punitive responses to street homelessness have been used to illustrate wider trends of urban revanchism and the changing politics of public space, for example. Though such work has brought the problems of homelessness to a much wider audience, other responses to homelessness and the experiences of homeless people themselves are not always well specified in such accounts. As a result, despite a growing body of work by geographers 'on' homelessness, the problems of homelessness itself are still not always well understood by geographers.

Understanding Homelessness

Problems of Definition

When people think of homelessness, they tend to think of people living on the streets. Certainly, this rather narrow definition has shaped much of the work by geographers on this topic. In fact, street homelessness is only one form of a much broader problem. When seeking to define homelessness, it is most useful to think in terms of a continuum of accommodation need. Such a continuum stretches from the street homeless population at one end, through other forms of 'visible' homelessness (those staying in emergency shelters and hostels) to a variety of forms of 'hidden homelessness' at the other end (those living in overcrowded and insecure accommodation, or staying unwillingly with friends or relatives). One problem with such definitions is that it reduces homelessness to issues of accommodation and may, in practice, work to reproduce normative understandings that conflate particular forms of accommodation with the experience of 'home' or 'homelessness'. An alternative is to try and capture something of these experiences themselves: defining homelessness in relation to the absence of those feelings of emotional attachment, security, and control (usually) associated with a sense of 'home'. The best work by geographers on these issues has sought to give voice to homeless people's own understandings, tracing the different forms of accommodation (including street encampments, or mobile homes, for example) that for them may equate to a sense of 'home'.

Overlaying these conceptual debates are important legislative constructions that define and divide different homeless populations. In the United States, the services used by homeless people (e.g., emergency shelters, transitional housing, drop in centers, and soup kitchens) are almost wholly provided by voluntary sector organizations supported by a mix of competitive contract funding and indirect state subsidies in the form of welfare benefit transfers. In the UK, the situation is more complex. While the local state has legal responsibility for the emergency accommodation and rehousing of those considered in 'priority need' (mainly families with dependent children, the very elderly, or ill), childless couples and other single, homeless people must rely on the much more basic services provided by a wide variety of voluntary organizations and charities. Because served mainly by voluntary sector organizations these latter groups are often not included in official, government statistics on the 'homeless' population.

Such differences should not be overemphasized. Work by geographers has demonstrated that even in the United States different homeless people are often treated very differently, accessing different services in different parts of the city. Hence, while single, homeless men especially may have few options but to turn to a (declining) stock of 'single room occupancy' (SRO) hotels and traditional dormitory style shelters in 'skid row', for example, homeless women and families are increasingly likely be found in one of a new generation of 'designer' shelters that tend to be located in suburban rather than central city locations. Though physical conditions in such shelters are much improved on the old style shelters of skid row, they often require residents to submit to highly invasive programs of 'welfare and rehabilitation' (budgeting classes, compulsory job searches, and so on) designed to facilitate homeless people's 'reentry' into 'normal' society. Overlaying and complicating this gender divide are divisions of race and ethnicity, with the distribution of different homeless groups tending to mirror the wider distribution of ethnic minority communities across the city, but with a marked overrepresentation of African American men, in particular, in traditional skid row neighborhoods.

As should be clear, differing conceptualizations of homelessness make it extremely difficult to gain an accurate picture of the size of the homeless population, because it must first be determined how 'homelessness' is being defined. Because of the differing legislative contexts shaping definitions of and responses to homelessness in different places, cross national comparisons are even more problematic. What is clear is that street homeless people make up only a small proportion of the wider homeless population in any one place. In the UK, for example, even when at its peak in the early 1990s, the number of people 'sleeping rough' on London's streets never rose above about 3000. In contrast, an estimated 15 000 people are currently housed in London's emergency shelters, with a further 67 000 households (equating to approximately 167 500 people) accepted as homeless and in 'priority need' by London's local authorities. Though it is the US experience that continues to prove of most interest to geographers working in this field, levels of homelessness in US cities, in fact, appear much lower: with recent estimates putting the total number of shelter beds in New York at just 30 000, and with as few as 8 602 in Los Angeles – where levels of street homelessness are, not surprisingly perhaps, much higher.

The Social Production of Stigma

That homeless people must so often continue to rely on the help of voluntary organizations and charities rather than the state is a sign of the stigma still associated with homelessness. Three main roots to this stigma can be traced. First, insofar as they are usually under or unemployed, homeless people tend to be viewed as 'unproductive'. In a society that accords a privileged status to economic productivity, homeless people are at best afforded a marginal position and at worst may be viewed as a drain on collective resources. Indeed, the work that those on the streets sometimes engage in, such as panhandling, are themselves often considered as further undermining normative constructions of both work and welfare. Second, having apparently lost contact with friends and family, homeless people are commonly perceived as 'disaffiliated' – as existing outside the comforts and constraints of mainstream society and its values. Third, whether because of their own habits or because of stereotyping, homeless people often come to be associated with other stigmatized, and threatening, groups: drug addicts, alcoholics, and the mentally ill, for example.

As geographers have shown, the stigma suffered by homeless people is inherently geographical. Most obviously, eating, bathing, and sleeping on the streets contravenes deep rooted and fundamental distinctions between public and private space, and a sense of the appropriate behavior associated with each. This sense of 'rupture' is especially powerful in the case of homeless women, whose presence on the streets undermines traditional understandings of femininity rooted in the private sphere. At the same time, traditional responses to homelessness – and homeless shelters in particular – work to maintain and reproduce this stigma, as people become 'tainted' by their use of services located in poor and marginalized neighborhoods and must take on the label of 'homeless' in order to qualify for such services.

But it is also clear that reactions to homeless people and their plight differs, with some more liable to attract sympathy than others. Important to such differentiation is the extent to which a person might be understood as responsible for the circumstances in which they find themselves, or capable of effecting their own 'escape' from the streets. Hence, while a homeless woman with children might be afforded considerable sympathy, a single, homeless man with an addiction is likely to engender a far less charitable response. Reanimating older divisions between the 'deserving' and the 'undeserving' poor, exactly these divisions can be seen in the distinctions drawn in the British legislative system between those in 'priority need' and others, and in the US between those provided access to a 'designer shelter' and those consigned to skid row.

Explaining Homelessness

Explanations of homelessness vary, and have been driven as much by changes in the homeless population as by intellectual currents within the academy. In the 1960s and 1970s, a period of relatively high or 'full' employment saw the (visibly) homeless population limited to a relatively small number of older, single, homeless men living within the confines of clearly bounded skid row districts. Under such conditions, analysts tended toward individualized explanations. Thus, homelessness came to be understood as a result either of individual pathology (a result of alcoholism, an inherent 'wanderlust', or social disaffiliation, for example) or as an expression of individual vulnerability (relating to a person's ill health, or mental incapacity, for example).

By the late 1970s, however, the growing numbers of mentally ill people on the streets as a result of deinstitutionalization and (the failure of) 'community care' programs, led analysts to begin to examine the various structural changes that seemed to be 'producing' a change in the homeless population. Such a move was given renewed impetus in the 1980s, as the traditional homeless population was first supplemented and finally surpassed by growing numbers of what came to be referred to as the 'new homeless'. At a time when geographers and other social scientists were exploring the relations between 'structure' and 'agency', the rising numbers of younger people, women, and families appearing in drop in centers and shelters led to the emergence of a new academic orthodoxy that remains largely intact to this day. Within this frame, homelessness tends to be explained with reference to the ways in which recent changes to the economy and state welfare regimes have impacted especially hard on exactly those groups most vulnerable to such changes, and who are increasingly to be found among a newly expanded homeless population. The best work in this area by geographers has sought to trace the impact of broader structural changes operating across a variety of scales and at different 'speeds' (from processes of global economic restructuring to changes to state welfare regimes and urban labor markets) on changing levels of homelessness, and on the groups most affected by homelessness, in different locales.

The Geographies of Homelessness

With a better understanding of how homelessness can be defined and of its causes, and of why homeless people continue to be so vilified, we can trace a clearer picture of work by geographers on this issue. Three main bodies of work can be identified.

Shelter Location, Survival Strategies, and the Politics of 'Malign Neglect'

From the late 1970s through to the early 1990s, studies of homelessness focused on three main themes: the causes of homelessness, the politics of shelter location, and homeless people's own survival strategies. Under the first of these, as we have seen, by the mid 1980s, a new orthodoxy had emerged that explained an increase in levels of homelessness, and changes in the homeless population, with regard to the differential impact of broader structural processes on different groups in the city. Under the second, geographers explored a range of issues around shelter location. Hence, studies mapped the distribution of homeless services, for example, or explored land use conflicts and community opposition around the siting of homeless facilities. In the best of such work, scholars sought to deploy the insights of a wider sociospatial dialectic: tracing the ways in which services first followed then reenforced the growing concentration of homeless people and other poor and marginalized groups within traditional skid row districts. Underpinned by a belief that homeless people were ill served by this concentration of services in central city, 'service dependent ghettos', geographers sought to explain why services tended to concentrate in such areas rather than being more evenly dispersed across the city (whether in poor or middle class neighborhoods), and the effects of such concentrations on homeless people themselves. For most, though the 1970s and 1980s were widely understood as an era of 'malign neglect' – with city authorities and the state seemingly happy to leave the 'management' of homeless people and other 'troublesome' populations to the voluntary sector – such concentrations were read as a form of 'containment'. Hence, geographers examined the ways in which urban managers, local business consortiums, and community groups used the urban planning system to restrict (street) homeless people and homeless services to strictly delimited areas of the city, out of the public eye.

Though some of this work was conducted in Britain, it should be noted that the vast majority of these studies were concerned with the geographies of homelessness and homeless service facilities in North American cities. In the absence of a more developed comparative literature, it is therefore very hard to know whether or not the patterns and processes traced in North America were in play in the UK or elsewhere in Europe at this time. Nor were scholars necessarily in agreement as to the impact of such processes. While most argued that the location of services in skid row tended to further stigmatize service users, others suggested that the concentration of SROs, shelters, and other services in tightly defined areas offered an important focus around which homeless people's own support networks could develop.

Indeed, by the early 1990s, geographers were increasingly interested in the survival strategies adopted by homeless people when on the streets and in the role that homeless people's own street networks played in these strategies. Deploying the insights of time geography, workers began to map the daily life paths of street homeless people. Though such pathways were indeed often focused around the institutional spaces (of shelters, drop in centers, and soup kitchens) that provide a focus for homeless people's daily routines, other spaces and other pathways too were recognized as important. Thus, studies began to explore people's earning strategies, for example, or the possibilities that homeless people's own temporary encampments offered for establishing some sense of 'home' among those unable or unwilling to access the strictly regulated spaces of the formal shelter system. Crucial here was the attention paid to gender relations and to the very different roles that homeless men and women often took on when on the streets and in their relationships with one another. Broadly speaking, scholars recognized that such partnerships and wider street networks were often crucial to an individual's survival, and to homeless women in particular, as they enabled a pooling of resources and a more effective allocation of the tasks essential to day to day survival. At the same time, it was clear that as people became more adept at surviving on the streets so too they often internalized a powerfully negative sense of themselves as 'homeless', and experienced a further worsening of their material circumstances.

Urban Revanchism and the Purification of Public Space

In the mid to late 1990s, a new and more punitive response to the problems of street homelessness emerged. Rather than seeking simply to contain homeless people within the confines of skid row, the past decade has seen sustained attempts to clear homeless people from the prime spaces of the city. A key part of such campaigns has been the criminalization of the activities that those on the streets must engage in to survive. Hence, in numerous cities across the United States, it is now a criminal offence to eat, sleep, or bath in public, for example. Elsewhere, there have been increased attempts to hide the evidence of street homelessness through the use of new architectural forms that work either to obscure a city's homeless facilities, or to 'defend' newly redeveloped sites from incursions by homeless people and the poor through the use of hard and soft landscaping, closed circuit television (CCTV), and private security. For some, such a shift signals the end of an era of the 'malign neglect' of homeless people and the birth of a new 'post justice' or 'revanchist' city. Within such a city, the presence of homeless people is no longer tolerated in newly 'purified' public spaces designed to foster upmarket consumption and a revitalized public life. Skid row itself is also under increasing pressure from gentrification: leaving homeless people literally nowhere to go.

Explanations for such a shift differ. For those working within the perspectives of political economy, such developments are best understood with reference to attempts by city managers to attract and retain capital investment and middle class workers, residents, and consumers in an era of intensified interurban competition. Certainly, there would appear to be a distinctive geography to the post justice city, with street clearance campaigns most evident in those cities with a heavy reliance upon the financial and creative industries, tourism, and the convention trade. Within such cities, antihomeless practices seem to be most focused upon sites within or immediately adjacent to the financial district and downtown core. Others have stressed instead the symbolic work of such campaigns, suggesting that attempts to clear homeless people from the streets are best understood as an attempt to constitute and secure a new urban 'public' against a reviled and unproductive 'Other'.

Rather less attention has been paid to the ways in which homeless people have sought to negotiate, or resist, such changes. Where geographers have considered such issues, they have perhaps too often focused upon the high profile demonstrations and legal actions that have occasionally been organized and led by advocacy groups on behalf of homeless people. In fact, it may be more profitable to examine the numerous but more subtle ways in which homeless people themselves seek to resist such pressures: attempting to 'pass' as customers when sheltering in public libraries or making use of restrooms in stations and restaurants, for example, or restricting their movements to the more marginal spaces of the city. Such work is vital to restore a sense of agency to accounts of homelessness in which homeless people's own understanding and experiences are too often ignored.

Beyond Revanchism?

Many of the developments traced in the previous section are indeed increasingly evident in cities across both North America and Western Europe. Nonetheless, not all are convinced by arguments that recent years have wit nessed a wholesale shift from an earlier 'malign neglect' of homeless people to something more punitive, or by the conceptual frame used to explore such a shift. Most ob viously, perhaps, recent work by geographers has made plain that talk of the breakup of skid row is premature. In Los Angeles, for example, skid row still houses the single greatest concentration of shelter beds within the region. More significantly, perhaps, far from declining, the number of emergency shelters has in fact increased in recent years in both the United States and Britain – even in those cities held up as most obviously articulating a 'revanchist' turn. While such an increase many of course be read as a new attempt to better contain the problems of homelessness, it is difficult to understand how it might be seen as articulating an obviously 'vengeful' disavowal of homeless people by city managers and the wider public. Not least, the vast majority of such shelters are provided by voluntary organizations and charities, many of which are reliant upon often substantial state subsidies, public donations, and the work of significant pools of volunteer labor. Rather than any simple or straight forward shift in responses to homelessness in recent years, it therefore seems that homeless people often continue to enjoy considerable public sympathy, with a deep rooted and powerful charitable impulse existing alongside (and frequently undermining) more punitive responses to street homelessness.

This more complex picture is not easily captured by the conceptual apparatus deployed in accounts of urban revanchism or of the 'post justice' city. Not least, with attempts to 'reclaim' the city framed in terms of the needs of city managers to maintain the image of a prosperous and competitive city – that is, by the needs of capital – such accounts leave little room for other actors: most notably, the state. In fact, in both the United States and Britain, the local and central state have taken an increasingly active role in the management of street homelessness: using new systems of funding and regu lation to shape the activities of the voluntary sector organizations working with homeless people, and seeking to influence public responses to street homelessness through various 'educational' and 'diverted giving' campaigns, for example. In an effort to give proper space to the variety of actors shaping responses to homeless ness in the contemporary city, and the complex rela tionships between them, scholars have therefore recently turned to a variety of alternative conceptual frames – ranging from neoliberal welfare restructuring 'poverty management'.

Conclusions and Further Study

The last 20 years or so have seen an increasing body of work by geographers on the problems of homelessness. Accounts of the revanchist or 'post justice' city in par ticular have brought the problems of homelessness to a much wider audience, both within and beyond the dis cipline, and have raised fundamental questions not only about homelessness and responses to it but also about changing constructions of the 'public', public space, and different people's rights to the city.

Nonetheless, both earlier and more recent works by geographers on the problems of homelessness remain curiously delimited. Most obviously, the vast majority of such work has been concerned with questions of street homelessness within an Anglo American urban context. While there is certainly evidence of a broad (neoliberal) consensus with regard to the most effective ways of managing homeless people and other 'troublesome' populations in Britain and the United States, recent re sponses to street homelessness in Britain and the US have by no means been exactly the same. Such differences will be even more marked in other countries, where quite different traditions of social welfare provision and urban policy are evident. One area where geographers can obviously make a contribution to debates on homeless ness is therefore through more comparative work that seeks to think beyond existing Anglo American metrics of knowledge.

But geographers interested in homelessness must also step beyond the city. A rich vein of work is now emerging on the geographies of rural homelessness, with scholars exploring the quite different problems faced by those who disturb deep rooted constructions of the rural idyll. Yet, there remains little work exploring the geographies of homelessness beyond these two extremes. Rather than continue to limit a consideration of the geographies of homelessness to either rural areas or metropolitan cen ters, there is a need to consider issues pertaining to homelessness in the 'ordinary city' where both the ex periences of homelessness, and responses to it, may be quite different.

In light of the quite different ways in which home lessness can be defined, and the variety of responses that different groups may experience from policymakers, service providers, and the public, geographers must also move beyond their obsession with the problems of visible homelessness. Instead, scholars should seek to build upon the small but growing body of work by geographers ex ploring other spaces – family shelters and women's ref uges, for example – to give proper consideration to other forms of homelessness and the experiences of different groups.

Such a call is also, in part at least, a call to pay more attention to the ways in which homelessness is imagined and defined, whether by policymakers, the public, or geographers themselves. As studies of the discursive construction of homelessness make clear, the ways in which the problems of homelessness, homeless people, and the geographies of homelessness are represented in newspapers, novels and films, textbooks, census materials, and legislative systems is of vital importance to the ways in which problems of homelessness are addressed.

See also: Public Space.

Further Reading

Cloke, P., Milbourne, P. and Widdowfield, R. (2002). *Rural Homelessness: Issues, Experiences and Policy Responses*. Bristol: Policy Press.

Deverteuil, G., May, J. and Von Mahs, J. (forthcoming) 'Complexity not collapse: reframing geographies of homelessness in a punitive age', Progress in Human Geography.

Johnsen, S., Cloke, P. and May, J. (2005). Day centres for homeless people: Spaces of care or fear? *Social and Cultural Geography* 6(6), 787 811.

May, J., Johnsen, S. and Cloke, P. (2005). Re phasing neo liberalism: New labour and Britain's crisis of street homelessness. *Antipode* 37(4), 703 730.

Mitchell, D. (2003). *The Right to the City: Social Justice and the Fight for Public Space*. New York: The Guilford Press.

Rowe, S. and Wolch, J. (1990). Social networks in time and space: Homeless women in skid row, Los Angeles. *Annals of the Association of American Geographers* 80, 184 204.

Ruddick, S. (1996). *Young and Homeless in Hollywood*. New York: Routledge.

Takahashi, L. (1996). A decade of understanding homelessness in the USA: From characterization to representation. *Progress in Human Geography* 20(3), 291 310.

Veness, A. R. (1994). Designer shelters as models and makers of home: New responses to homelessness in urban America. *Urban Geography* 15(2), 150 167.

Wolch, J. (1996). From global to local: The rise of homelessness in Los Angeles during the 1980s. In Scott, A. & Soja, E. (eds.) *The City: Los Angeles and Urban Theory at the End of the Twentieth Century, pp 390 425*. Berkeley, CA: University of California Press.

Wolch, J. and DeVerteuil, G. (2001). New landscapes of urban poverty management. In May, J. & Thrift, N. (eds.) *TimeSpace: Geographies of Temporality*, pp 149 168. London: Routledge.

Relevant Websites

http://www.crisis.org.uk
Crisis.

http://www.homeless.org.uk
Homeless Link.

http://www.lacehh.org
Los Angeles Coalition to End Hunger and Homelessness.

http://www.ris.org.uk
Resource Information Service.

Homelessness, Rural

P. Milbourne, Cardiff University, Cardiff, UK

Introduction

Homelessness is generally assumed to be an urban problem in most 'developed' countries. Academic, policy, and media discussions of homelessness tend to focus on its occurrence in particular parts of the city, with its presence in other places often understated and sometimes denied all together. This is particularly the case with small town and rural spaces where relatively little research or policy attention has been directed toward homelessness. Given that homelessness represents a social phenomenon, that is, something that impacts on (particular) social groups regardless of where they live, the spatialized nature of discussions about it raises some important questions for human geographers. Perhaps the most significant of these is why 'the urban' should have captured so much of the coverage of homelessness.

The importance attached to homelessness as a research and policy issue largely relates to its increased visibilities in certain places and spaces at particular moments in time. In some cases, these visibilities concern direct observations of homeless groups in specific spaces; in other cases, they result from media discussions of homelessness. In recent decades, it has been the increased presence and reporting of people sleeping rough, and particularly young rough sleepers, in public spaces within major cities that has led to academic, policy, and media attention being directed toward homelessness. Instead of acting as a springboard for broader discussions of the sociospatial features of contemporary homelessness; however, these discussions have remained largely fixed on these particular forms of literal homelessness present in urban places.

As well as being the most visible form of homelessness, rough sleeping is the type of homelessness that is most obviously associated with the city. Not only are those sleeping rough most numerous in urban places, they are also highly visible in particular spaces within these places, such as shop doorways and public parks, which are passed by many thousands of workers and shoppers on a daily basis. It is also the case that a broad range of support services for the homeless – hostels, shelters, soup kitchens, drop in centers and so on – exist in the city, which act as important sites, sights, and symbols of homelessness in these spaces. Geographers and other social scientists interested in homelessness have, therefore, been drawn to these urban spaces in order to research different aspects of homelessness. Consequently, the key research and writings on homelessness have utilized the city as a research laboratory.

While this focus on the city as the dominant space for homelessness research has produced important evidence on the sociospatial dimensions of homelessness, it has also directed critical attention away from other forms and other spaces of homelessness. Homelessness is not just associated with a lack of physical shelter; it also involves people living in insecure or inadequate forms of housing and in properties that do not constitute 'home'. These broader definitions of homelessness, however, are subject to debate among academic researchers and policymakers, not least because there exists little agreement about what constitutes housing security and home. Nevertheless, it is appropriate to construct homelessness as a continuum of situations of housing need, ranging from a lack of physical shelter, through institutional forms of living, involving hostels and shelters, to precarious housing situations that jeopardize the physical and emotional well being of occupants. Examples of this last category of homelessness include people living in unfit, crowded, or unaffordable accommodation, and those experiencing domestic violence. Considered in these terms, rough sleeping constitutes an extremely small proportion of homeless cases, with most homelessness occurring in the largely hidden spaces of the 'home'.

The domination of city based homelessness research has also downplayed the significance of homelessness in small town and rural spaces. Homelessness has long been a feature of living in rural areas, linked to processes of urbanization, agricultural depressions, land clearances, inadequate housing conditions and poorly developed local welfare infrastructures, and passing references to tramps, hobos, and wayfarers can be found in various publications on historical rural welfare issues. It has only been during the last couple of decades, though, that researchers have turned their attention specifically to problems of 'rural' homelessness. Despite this late start, rural researchers have done much to highlight the significant and, in several countries, increasing presence of homelessness in rural areas. This work has pointed to important similarities between the causes and types of homelessness in rural and urban places; it has also indicated some distinctive elements of homelessness in rural areas, particularly the reduced visibilities of homeless people and the restricted provision of welfare services for homeless people.

Researching Rural Homelessness

Rural homelessness first began to be studied in the United States and the United Kingdom in the late 1980s and early 1990s. Initial work sought to develop statistical profiles of rural homelessness and to compare these with the situations in urban areas. Early attention focused on the numerical significance, social profiles, and causes of homelessness in rural and urban spaces, with some attention also given to the spatial distribution of homelessness within rural areas. Research in the United States drew on rough sleeper counts undertaken as part of the population census, as well as specially commissioned surveys of the homeless population in particular rural states. In the United Kingdom, official statistical data on the changing incidence of homelessness was analyzed at the local level to compare the shifting rural and urban situations.

The volume of research undertaken on rural homelessness in these two countries has increased since the early 1990s, as has the breadth of methodological approaches utilized to examine the issue. Statistical accounts of the (changing) scales, geographies, and features of homelessness in rural areas have continued, based on analyses of official data and also a growing body of evidence from surveys of the rural homeless. Such analyses have played an important role in demonstrating the numerical significance of homelessness in rural areas, as well as some of its distinctive features, to the research and policy communities. Qualitative methodologies have also been employed to explore the nature and experiences of homelessness in particular rural places. Utilizing depth interviews with local welfare agencies and homeless or recently homeless people, and ethnographic techniques, these studies have highlighted the previously hidden dimensions of homelessness and the complex coping strategies employed by homeless people to negotiate rural spaces. Significant within this body of work has been a qualitative study of homelessness in two case study counties in rural England undertaken in the mid 1990s. Most recently, the publication of the first international book on rural homelessness in 2006 has not only provided discussions of the academic and policy neglects of rural homelessness in other 'developed' countries – Australia, New Zealand, Spain, Ireland, Finland, and Canada – but also offered valuable narratives of the extent, experiences, and welfare contexts of rural homelessness in a broad range of rural places.

Restricted Visibilities of Rural Homelessness

Homelessness remains largely hidden within the physical, sociocultural, and political fabric of rural space. Rural areas are characterized by low population densities, meaning that homeless people tend to be dispersed across quite large spaces. In addition, the most obvious sites and sights of homelessness, such as retail centers, public parks, and welfare support facilities, tend to be absent from most rural areas, denying rural homelessness a similar level of visibility and legitimacy to that present in urban spaces. Consequently, homelessness is more difficult to research in rural areas. Literal homelessness is both difficult to practice and when it does occur is found in less obvious spaces, such as woods and abandoned agricultural buildings, while the bulk of homelessness remains hidden within private households and a range of precarious housing situations. It is extremely problematic, then, to count homeless people in rural areas; it is also far from easy to undertake place based depth studies when the rural homeless are not concentrated in particular spaces.

Such difficulties have led to calls for the development of broader conceptions of rural homelessness, which are capable of revealing the wider spectrum of homeless situations in rural areas. In addition to its literal and institutional forms, it is claimed that rural homelessness should be understood as a series of precarious housing situations, involving people living 'on the edge of homelessness' or at risk of homelessness. Such situations may include people pushed into properties deemed to be inadequate on account of their poor physical conditions, unaffordable housing costs or insecurity of tenure, as well as those forced to live in the homes of family members or friends in order to fend off situations of rooflessness. These broader forms of homelessness tend to involve long periods of housing insecurity and 'near homelessness' and relatively short periods of literal or institutional homelessness, which means that they rarely come to the attention of welfare agencies and may not be recognized as homelessness by the people experiencing these situations. It is also the case that these broader forms of homelessness represent difficult research subjects as they are characterized by considerable spatial and temporal dynamism.

The presence of homelessness in rural spaces is also downplayed by important sociocultural processes. Dominant discourses of rurality in most developed countries are bound up with notions of community cohesion, self sufficiency, privacy, and the ownership of property. Given that homelessness transgresses these 'norms' of rural living, it tends to be constructed as 'out of place' or recognized in other ways, as poverty or housing needs, within rural space. In some countries, such as the UK, discourses of rurality are also bound up with notions of idyll, which construct rural areas as places devoid of social disadvantage and thus deny the presence of homelessness within them. Such national discourses of rurality are actively imposed on to local rural space by ex urban middle class groups who continue to move to rural areas in search of idyllized lifestyles.

The restricted provision of welfare services and fa cilities for homeless groups in rural areas compounds the out of place nature of rural homelessness and accentuates the otherness experienced by homeless people. The limited presence of these important signs of homelessness within the rural landscape results from a series of prac tical difficulties associated with identifying and meeting the needs of homeless people in rural areas. However, it also relates to the negative attitudes and (non)actions of local government and other elite groups within rural society toward homelessness. For sets of political, fiscal, and economic reasons, it has been in the interests of these groups to ignore the needs of the rural homeless, allowing them to deal with their own problems or to move to other areas, rather than to provide the types of support services and facilities that have been developed in urban spaces. While such attitudes and actions were mostly associated with the past, the consequences of these historical political and policy denials of local homelessness have, as we will be shown later, carried forward into the present.

Statistical Assessments of Rural Homelessness

Quantifying the extent of homelessness in rural areas is fraught with difficulty. Homelessness data in most coun tries are generally restricted to counts of rough sleepers. These counts are generally based on snapshot surveys of sleepers in the known sites of the homeless, which tend to be located in the cities and large towns, and, as such, provide a rather partial and geographically biased indi cation of the extent of the problem. It is hardly surprising to learn then that rough sleeper counts highlight that this type of homelessness is most significant in metropolitan areas. When more sensitive counting methods are em ployed, involving surveys in a broader range of spaces and over longer time periods, the gap between the size of the rough sleeper population in urban and rural places re duces but generally remains, for reasons discussed earlier.

In a small number of countries, levels of homelessness are routinely recorded by the state as a part of their welfare systems. In the United Kingdom, for example, the government has published local data on the incidence, profile, and causes of homelessness since the 1970s. While these data are based on a rather narrow definition of homelessness, they do provide a useful indication of the changing scales, geographies, and characteristics of homelessness in rural areas. Analyses of these official statistics indicate an increased significance of homeless ness in rural areas since the early 1990s, although standardized rates of rural homelessness, which take into account population totals, have been broadly static across this period. In addition, the standardized level of rural homelessness remains below that recorded for urban areas. This urban–rural differential is also evident from specially commissioned surveys of homelessness in se lected states in the United States and Finland.

Efforts have also been made to examine the spatial distribution of homelessness within rural areas, with these analyses revealing that rural homelessness is characterized by a great deal of spatial unevenness. Studies in the United States and Australia, for example, point to highest rates of rural homelessness in those re mote areas which also record high levels of persistent poverty. By contrast, recent research in England high lights that while homeless rates in the remoter rural areas have grown at a faster rate than all other areas since the early 1990s, it is in the more accessible rural places, lo cated close to major metropolitan centers and subject to acute housing pressures from in migrants, where highest levels of homelessness exist.

While homelessness is something that potentially af fects all groups within rural society, research reveals that particular types of households are at greater risk of be coming homeless than others. Not surprisingly, house holds living in poverty are shown to be particularly prone to entering periods of homelessness. Indeed, much of the literature on rural homelessness in the United States highlights strong causal linkages between poverty and homelessness. Indigenous peoples are also identified as an important group of the rural homeless population in certain countries – the United States and New Zealand, for example – while the growing number of migrant workers engaged in agriculture and food processing are shown to be particularly vulnerable to homelessness on an episodic basis. Beyond these specific groups, the social profile of homelessness in rural and urban areas appears to be broadly similar, although research in the United States suggests that the rural homeless are more likely to be female, white, and in employment than their urban counterparts.

The factors that lead individuals and households to become homeless in rural areas are extremely diverse, and concern a complex mixture of structural and per sonal problems. In terms of structural problems, low in comes, unemployment, poverty, and inequitable housing systems associated with particular rural places are shown to be important causal factors. In relation to personal factors, relationship breakdowns, illness, and alcohol or drugs dependencies are identified as significant. Of course, these structural and personal factors interconnect and it is often difficult to isolate one specific factor that leads to homelessness. That said, research evidence from the United States and the United Kingdom would seem to indicate an increased significance of housing as a causal factor in rural areas.

Coping with Homelessness in Rural Spaces

Perhaps the most interesting material to emerge on the geographies of rural homelessness has been that derived from in depth qualitative studies, which have explored the experiences of homelessness among different groups of the homeless population and in different rural places. Up until quite recently, the small number of studies examining local experiences of homelessness in rural areas were restricted to the United Kingdom and the United States. However, new experiential accounts of rural homelessness in other countries are beginning to emerge. Based on extended periods of fieldwork, these studies have sought to develop accounts of homelessness from the perspectives of local welfare agencies working with homeless groups, and homeless or recently homeless people themselves. Attention has been given to the experiences of different homeless groups, in particular, rural places and spaces, as well as the experiences of specific groups, such as young people, indigenous groups (Native Americans and Maori in New Zealand), seasonal or migrant workers, and trailer park residents.

These studies illustrate a series of everyday difficulties that homeless people face living in rural places which are largely devoid of support services and the coping strategies that they are forced to adopt in order to remain sheltered in these places. One of the most significant problems faced by homeless people in rural areas is the lack of local recognition of their problems. Their reduced number, spatial dispersal, and restricted access to places of support and congregation deny homeless people any sense of legitimacy within rural space. In addition, many of the precarious housing situations occupied by homeless people are discussed by local agencies, communities, and the homeless themselves in rural areas with reference to terms other than homelessness. In certain circumstances, though, homelessness can constitute a highly visible phenomenon in rural areas as the small scale and close knit nature of rural living leads to unwanted public exposure of private problems and the accentuation of feelings of otherness among the homeless.

Rural space thus presents a series of physical and sociocultural obstacles to the existence of homelessness that require the development of coping strategies on the part of homeless people. Those facing homelessness are often required to travel considerable distances across rural spaces, and between rural and urban spaces, in order to access welfare advice and support services. The fact that welfare support provision and emergency accommodation for the homeless in rural areas is mainly located in the larger settlements leaves homeless people the 'choice' of remaining local and dealing with their own situations or moving out of their communities to be assisted by welfare agencies. Both options create additional problems for the rural homeless. Moving away creates new forms of sociospatial isolation at times of considerable stress for homeless households, as well as reinforcing the invisibility of homelessness in the smaller rural settlements. Remaining local necessitates the employment of complex coping strategies, which often involve frequent movements between different precarious housing situations, including sharing temporarily with family or friends and moving into property that is inadequate in terms of its physical condition, size, location, or cost. While these housing situations may be interspersed by brief spells of literal homelessness, they are largely played out in the private sphere and thus rarely come to the attention of welfare officials or others in the rural community.

Recent research in England suggests that rural homelessness also involves nonlocal people. For some homeless or near homeless people, rural areas may provide increased or improved housing options. For example, rural housing costs in most countries tend to be lower than in the cities and some rural places offer particular types of low cost housing options, such as in trailer parks. Others may be seeking the increased privacy and environmental benefits associated with rural living and an escape from problems linked to homelessness in the city. Others still may be stopping off in rural spaces as they travel to and from other places, visit friends, or have a break from city living. Not only are these nonlocal homeless people more visible than their local counterparts in rural settlements, they tend to be treated with less compassion, invariably viewed as threatening community cohesion and undeserving of local welfare assistance. Their presence may also allow homelessness to be constructed by some local groups as a nonlocal and nonrural problem, reinforcing other denials of rural homelessness.

Rural Homelessness and Welfare

Homelessness is generally addressed indirectly as part of broader national welfare, antipoverty, and housing policies. This makes the task of identifying approaches to dealing with homelessness in different countries extremely difficult. The exception is the United Kingdom, where homelessness is recognized within national legislation as a welfare problem deserving of specific policy intervention. Here, national guidance exists to ensure that particular categories of homelessness are dealt with consistently at the local level. While these national systems of provision are intended to deal with homelessness in urban and rural areas, it is clear that they have placed much more emphasis on urban problems, with recent homelessness initiatives in different countries largely geared toward tackling problems of rough sleeping in major cities.

By contrast, very little policy attention has been given to the rural dimensions of homelessness. In some coun tries, such as the United Kingdom and the United States, there is now at least recognition by the government of the presence of homelessness in rural areas, but it is difficult to claim that any specific policy initiatives have been developed to deal with its features. The hiddenness of rural homelessness again plays a significant role in this policy neglect. Without robust evidence on its statistical significance and the particular problems faced by homeless people in rural areas, it is difficult to convince government and other welfare agencies to develop spe cific initiatives to deal with homelessness in rural areas. Unfortunately, the absence of advice and support services for homeless groups in rural areas, which act as physical, statistical, and symbolic markers of homelessness, means that the task of making known homelessness in these spaces becomes extremely problematic. Thus, rural areas are faced with a vicious circle of nonintervention and invisibility in relation to welfare and homelessness.

Much of the welfare support for homeless groups in rural areas is therefore provided by the voluntary sector, with projects generally small scale and locally based. Recent studies undertaken in different countries indicate that these types of initiatives play a valuable role in meeting the needs of specific homeless groups in par ticular rural places. However, a series of practical dif ficulties encountered by welfare agencies working in rural areas are highlighted, including identifying home less people, resistance from local elites and communities to the provision of welfare assistance to the homeless, and additional financial costs associated with providing such assistance in rural areas. In addition, the small scale na ture of many of these welfare projects means that they are better able to deal with the personal consequences of homelessness, for example, securing short term accom modation for homeless people, than the broader struc tural factors, such as low wage economies, poverty, and housing shortages, which result in people becoming homeless in rural areas.

Welfare provision for the rural homeless is also characterized by spatial unevenness in two senses. First, welfare projects tend to be based in the larger rural settlements, which not only creates access difficulties for homeless people in the rural hinterlands but also awards an increased visibility to homelessness in these small town urban spaces. Second, as individual projects are designed and implemented by different constellations of welfare actors to deal with particular aspects of home lessness in specific rural places, the provision of welfare support varies according to the nature of homelessness and the presence of different mixes of welfare agencies in different places. This means that people's experiences of homelessness and their ability to gain assistance for their problems can vary considerably from place to place. Where welfare facilities and services do exist locally it is likely that homeless people will make use of them; in other places, more informal and less satisfactory local support systems will need to be relied upon.

See also: Homelessness.

Further Reading

Cloke, P., Milbourne, P. and Widdowfield, R. (2000). Homelessness and rurality: 'Out of place' in purified space. *Environment and Planning D: Society and Space* 18(6), 715 735.

Cloke, P., Milbourne, P. and Widdowfield, R. (2000). The hidden and emerging spaces of rural homelessness. *Environment and Planning A* 32, 77 90.

Cloke, P., Milbourne, P. and Widdowfield, R. (2001). The local spaces of welfare provision: Responding to homelessness in rural England. *Political Geography* 20(4), 493 512.

Cloke, P., Milbourne, P. and Widdowfield, R. (2002). *Rural Homelessness: Issues, Experiences and Policy Responses*. Bristol: The Policy Press.

Fitchen, J. (1991). Homelessness in rural places: Perspectives from upstate New York. *Urban Anthropology* 20(2), 177 210.

Lawrence, M. (1995). Rural homelessness: A geography without a geography. *Journal of Rural Studies* 11, 297 308.

Milbourne, P. and Cloke, P. (eds.) (2006). *International Perspectives on Rural Homelessness*. London: Routledge.

Housing

M. Pacione, University of Strathclyde, Glasgow, UK

Glossary

Affordable Housing Dwellings built specifically for those on incomes that deny them the opportunity to purchase or rent on the open market.

Gentrification The process of neighborhood upgrading by relatively affluent incomers who move into a poorer neighborhood in sufficient numbers to displace lower-income groups and transform its social identity.

Housing Association An independent, nonprofit making organization funded primarily by government to build, improve, and manage affordable housing for sale or rent.

Negative Equity A situation in which the value represented by the market price of a house is insufficient to cover the cost of repayment of the mortgage taken on to purchase the property.

Socialist City A city that is planned and designed according to the principles of socialism and that operates as part of the command economy.

Slum An area of overcrowded and dilapidated, usually old, housing occupied by people who can afford only the cheapest dwellings available in the urban area, generally in or close to the inner city.

Squatter Settlement An illegal urban development, usually predominantly residential, on land neither owned nor rented by its occupants.

Housing Markets

In market societies private housing is a commodity produced and exchanged for profit. Outside the social or public housing sector where housing may be allocated on the basis of some criterion of need or at a below market price, the residential choices available to urban house holds depend primarily on their ability to compete in the housing market which comprises housing providers, fa cilitators, and consumers. While the wealthy can compete in the market for the most attractive and expensive housing in the best areas, those with low incomes or the long term unemployed are unlikely to be able to gain access to the private housing market except at the lowest levels. Those unable to buy must rent from either a private landlord or in the public social rented sector. In British cities generally, a major distinction is between the housing markets for owner occupied dwellings and for public social housing. In the US city, ethnicity may be substituted for tenure class, with the housing market for African Americans commonly representing a separate submarket with its own distinctive spatial arrangement. In addition to, and often overlapping with, class and ethnicity based housing markets, submarkets also operate within cities to accommodate the needs and preferences of particular groups differentiated on the basis of age, religion, or lifestyle.

Housing Tenure

Over the course of the twentieth century a major trans formation was effected in the tenure balance of housing in the UK. Prior to World War I almost 90% of British households rented their accommodation from private landlords, 2% rented from local authorities and the re mainder were owner occupiers. By 2001 the position had been largely reversed, with only 11% of households still renting privately compared with 74% owner occupation and 15% who were council tenants. Most of the trans formation occurred during the post World War II period (**Table 1**).

The Growth of Owner-Occupation

During the 1920s, owner occupation was not the pre ferred housing option of most people. Most households occupied privately rented accommodation, and council houses produced under the 1923 Housing Act (Cham berlain Act). The latter, because of the high standards of construction and consequently of rent, were reserved primarily for better off members of the working class. As **Table 1** shows, during the postwar period owner occu pation grew to become the largest sector of the UK housing stock. This was encouraged by several factors, including government policy; expansion of specialist fi nancial institutions such as building societies and banks into the mortgage business; an investment climate that has generally been favorable to property development; and households' increasing difficulty in finding alter native accommodation due to the decline in the privately rented and local authority sectors. The growth of owner occupation has been particularly marked since 1979, following the election of a Conservative government committed to the creation of a 'property owning democracy'.

In addition to the stimulus provided by government housing policies, the growth of home ownership in Britain reflected a belief that ownership of a house rep resents a sound financial investment. This assumes that

Table 1 Housing tenure and dwelling stock (*in millions*) in Great Britain, 1914 2001

Year	*Rented from local authority or New Town*		*Privately rented*		*Owner occupied*		*Total number of dwellings*
	No.	*%*	*No.*	*%*	*No.*	*%*	
1914	0.1	1.8	7.5	88.2	0.9	10.0	8.5
1944	1.6	12.4	8.0	62.0	3.3	25.6	12.9
1951	2.5	17.9	7.3	52.5	4.1	29.6	13.9
1961	4.4	26.8	5.0	30.5	7.0	42.7	16.4
1971	5.8	30.6	3.6	18.9	9.6	50.5	19.0
1981	6.6	31.0	2.7	12.6	11.9	56.4	21.2
1991	5.7	24.8	1.7	7.4	15.6	67.7	23.1
2001	3.6	15.4	2.4	10.6	17.1	74.0	24.7

over the long term, house prices will rise at least as fast as and probably faster than general inflation. Accordingly, it is better to pay mortgage interest on a loan of fixed capital value that results in the acquisition of an asset at diminishing real cost than to pay rent, probably at a similar level, but escalating over time and with no asset to show for it. This argument has proved to be generally correct. However, it is also predicated on the continuing ability of the mortgagee to service repayments. In a period of recession, unemployment combined with high debt/equity ratios can result in people's homes being repossessed for nonpayment of loan installments. In the early 1990s in the UK, rising unemployment with the onset of recession (following a government rise in interest rates to control inflation) meant that 1.8 million people who had taken on large mortgages to purchase houses during the boom of the late 1980s found themselves in a position of negative equity, where the exchange value of their house had fallen below the amount of their outstanding loan. In general, however, the advantages of owner occupation have ensured that, since 1971, the majority of the UK housing stock is under this tenure.

The Decline of Private Renting

As **Table 1** shows, from a dominant position in the housing stock prior to World War I that the privately rented sector has diminished in importance. Although it accounted for half of UK housing in 1951, by 1991 its share had fallen to less than 10% of the total stock, although there was a small recovery by 2001. The principal factors underlying this decline are slum clearance of the oldest and poorest dwellings, many of which were in the private rental sector; housing rehabilitation, aided by improvement grants, was often followed by tenant displacement and the sale of properties for owner occupation, particularly in gentrifying neighborhoods; the comparatively poor rate of return from investment in private rented housing; and the ethic of home ownership.

The Development of Public-Sector Housing

The decline of private rented housing, the traditional source of shelter for those denied access to the owner occupied sector, led to direct government intervention in the provision of housing. By 1939, 1.18 million houses had been constructed by the public sector, amounting to 10.3% of total stock. The growth of municipal housing also stemmed from the failure of nineteenth century housing associations to demonstrate that it was possible to provide decent dwellings in sufficient numbers at a rent affordable by the low paid; increased working class militancy over housing conditions, vividly displayed in the Glasgow rent strikes of 1915; the shortage of housing in the post World War I era; and a gradual shift away from the Victorian ideology of self help and the pervading belief in the efficiency of the market mechanism.

By the 1970s the crude housing shortage had been resolved, at least in national terms, although regional shortages remained. Nevertheless, despite significant progress towards the goal of 'a decent home for every family at a price within their means', by the late 1970s a number of criticisms were directed at the UK housing stock. These including the view that the social housing sector was too large and expensive to support; the subsidized rents were not sufficiently well targeted and benefited many who did not need this level of support; the below market level of rents contributed to low levels of maintenance undertaken by many local authorities, which stored up problems for the future; novel building methods employed in the 1960s and 1970s had given rise to structural problems in some estates; the large, homogeneous council estates lacked a mix of social classes and were often constructed without social amenities; and as unemployment rose, problems relating to poverty and deprivation began to concentrate in many estates.

All these factors contributed to a reexamination of national housing policy in Britain. The greatest force for change, however, was the ideological conviction of an incoming Conservative government committed to reducing public expenditure. With reference to local authority housing, a major divide opened between the government's view of council housing as a 'welfare service', assisting households unable to afford other types of housing, and the Labour Party's belief that the public sector should supply housing for 'general needs' (i.e., should satisfy the demand from all households who wish to rent rather than buy). The new government's housing policy set out to increase owner occupation; provide a wider choice of landlords in the social rented sector; revive the privately rented sector; and reduce the role of local authorities as direct housing providers. A key component of the Conservatives' housing policy, introduced in the 1980 Housing Act, was the right of council tenants to buy their homes at a discount price that reflected length of tenancy

Between 1980 and 1991, 1.46 million sales had been completed and the Treasury had received over £20 billion, making the 'right to buy' (RTB) scheme one of the largest privatization initiatives of the Thatcher era. The Conservative view of local authorities as 'enablers' facilitating action by other bodies, rather than direct providers of social housing, is evident in the reducing role of council house building and the growing importance of the housing association sector as an alternative source of rented accommodation. Housing associations obtain their loan finance for investment in new house construction and rehabilitation from the Housing Corporation (which oversees their activities) and, since the 1988 Housing Act, from private sector institutions. By 1998 housing associations accounted for 1.2 million dwellings or 5% of total housing stock. As a consequence of RTB and estate transfers, the public rented sector in the UK is becoming increasingly residualized and there is an increasing possibility that the future role of council housing in Britain will be to provide shelter of last resort to those unable to go elsewhere.

Public Housing in the USA

The restricted role envisaged for council housing by the political Right in Britain finds an echo in the history of US attempts to enact a program of public housing. A fundamental reason for the failure of the US public housing initiative was that it did not have widespread political support, being alien to the dominant market philosophy. Even the initial rationale behind the program was multifaceted: it was seen as a way of simultaneously reducing the high levels of unemployment and assisting the housing industry in the New Deal era of the 1930s, as well as a means of eliminating slums and increasing the supply of cheap, decent housing to the poor. Weak federal controls over development costs and housing quality inevitably led to major difficulties. In particular, local housing authorities received no operating subsidy from federal government, their sole source of income being rental payments. As costs rose, minimum and average rents increased steadily. Prior to the 1974 Housing and Urban Development Act, which introduced a degree of control, rent payments equivalent to half a tenant's gross income were common. The 1974 Act marked a major switch of emphasis away from conventional public housing towards the use of leased housing. However, signs of the increased privatization of public housing had emerged during the 1960s with the 1965 Housing Act authorization of turnkey housing, the sale of public housing to tenants, and the contracting out of management services to private firms. By the mid 1970s, construction of public housing projects had fallen significantly and some of the most deteriorated schemes had been demolished.

The switch of emphasis in social housing policy from supply side to demand side intervention was reflected in Section 8 of the 1974 Housing and Urban Development Act, which introduced a housing allowance program for the lower income households. Rather than making payments directly to needy tenants, the program consisted of long term contracts between the Department of Housing and Urban Development and private developers, landlords or public housing agencies designating specific apartments for Section 8 subsidies. Approved apartments could be existing units or units to be constructed or rehabilitated. Rental payments for tenants were on a par with those for public housing, amounting to between 15% and 25% of their total income. Housing owners received, from Section 8 funds, the difference between the tenant's contribution and a 'fair market' rent based on prevailing private rents in the locality. Section 8 subsidies have been successful in that tenants have a degree of choice of housing unit, and the scheme helped to reduce the over concentration of low income tenants in 'projects', many of which have been razed. The residualization of public housing is well advanced in the USA, with the sector catering for a dependent population characterized by low income, unemployment, reliance on public assistance, and a high concentration of the very old, very young, and minorities, often occupying a deteriorating physical environment.

Housing Affordability

Housing affordability is measured by the proportion of household income spent on obtaining housing. An expenditure of more than 30% of income in the USA and over 20% in the UK has been taken to indicate an affordability problem. Such measures are rather crude, however, since the same proportion of household income has a different impact for those at different ends of the social scale. Neither does it take account of the subjectivity of affordability. Nevertheless, despite definitional problems there is evidence of a growing affordability problem in both rented and owner occupied housing, particularly in overheated housing markets.

In the UK evidence of the growing unaffordability of housing for some social groups is found in the 400 000 households in mortgage arrears in 1995 and the repossession of 250 000 houses between 1991 and 1994. In the USA between 1997–99 while inflation rose by 6.1% house rents increased by 9.9% and house prices by 16.0%. In 1997, 5.4 million very low income families, (those with incomes below 50% of the local MSA median), paid more than half their income for housing or lived in severely inadequate housing, a situation classified by the Department of Housing and Urban Development as 'worst case needs'. This represented an increase of 12% in worst case needs housing since 1991. A growing proportion of these families are working households. In 2002, in no jurisdiction in the USA would a minimum wage job (at $5.15 per hour) provide enough income for a household to afford the 'fair market rent' (FMR) for a two bedroom home. Many low income earners must work two or three jobs to pay their rent. Between 1991–97, the number of housing units affordable to extremely low income families, (those with incomes below 30% of median MSA income), fell by 5% or 370 000 units. In New York City an extremely low income household could afford a monthly rent of not more than $471, while the FMR for a two bedroom unit was $1031.

Housing affordability is a problem for places as well as for people. Many people cannot find housing near their workplace or find work at a reasonable distance from where they can afford to live. In many areas workers critical to the local economy, such as teachers and police cannot afford to live in the communities they serve. The problem is particularly severe in cities with overheated housing markets. These include high tech hot spots in places like Boston, Denver, and San Francisco where rents increased by more than 20% in the late 1990s, and global cities such as London. The starkest indication of an affordability crisis, however, is represented by the growing incidence of homelessness, which is most visible in the inner areas of large cities, where the sight of people sleeping in doorways and living in cardboard shelters is redolent of conditions in the Third World city.

Homelessness

Homelessness is an extreme form of social exclusion, caused by a combination of personal and structural factors. Most homeless people leave their home either because parents or friends are no longer willing to accommodate them or because of the breakdown of a relationship. Structural factors underlying homelessness include insufficient construction of affordable housing, gentrification, cutbacks in welfare budgets, stagnating and falling real incomes, and the rise of part time and insecure employment. Discriminatory practices can also contribute to the problem for some social groups. Those most at risk of homelessness include unemployed people, single mothers, disabled people and frail elderly individuals, runaway youths, battered women and children, immigrants and refugees, substance abusers, and deinstitutionalized mental patients. While many of the homeless are highly visible, (e.g., beggars and rough sleepers), most are not noticeably different from other citizens. The invisible homeless have to cope with the day to day strain of living in temporary accommodation, hostels, bed and breakfast hotels, or in cramped conditions with friends and relatives

Housing in Socialist Societies

Under state socialism, urban land and housing markets did not exist. Before the transition to capitalism almost all housing was state owned and city governments restricted residential mobility as a way to tackle housing shortages. People lived in the same place for long periods and neighborhood change was slow. Planners located stores and services according to the demographic profiles of neighborhoods and once located, urban functions remained in place for decades.

In the transition period since 1989, privatization and replacement of central planning with the free market, together with postindustrial restructuring has catapulted cities such as Moscow from industrial socialism to post industrial capitalism in a short period. In the post Soviet city, capitalist tendencies such as suburbanization and counterurbanization are in evidence, and social differentiation in housing and quality of life are increasingly apparent. In Moscow between 1990 and 1994 the level of private ownership of housing increased from 9.3% to 49.1%. This tenure change and the reintroduction of market mechanisms have been accompanied by increased residential mobility that has enabled some individual citizens to improve their living conditions. As well as suburbanization, gentrification has appeared in central city redevelopment schemes while, more generally, there is a trend towards a 'European' urban residential structure characterized by an affluent inner core and poor periphery. Nevertheless, while individual choice has been enhanced for those able to participate in the housing market, state planning control of zoning regulations and real estate taxes (coupled with a limited private mortgage system and declining real incomes for the majority) ensure that in the short to medium term at least, the social geography of Russian cities will continue to be influenced strongly by the local government. As in the former Soviet

Union, in China the declining role of state enterprises in the economy, the introduction of land and housing markets, and the opening up of cities to foreign invest ment has meant that the state and centrally planned economy have a much reduced influence on urban de velopment. Just as in the post Soviet city, 'Western' processes of urban sprawl, spatial segregation, and social polarization are being reproduced in Chinese cities as well as in other 'transitional societies' such as Vietnam.

Housing in Third World Cities

In most large cities of the Third World much of the population occupies the most rudimentary forms of shelter. The failure of government housing programs to provide affordable housing has forced the mass of the Third World urban population into cheaper alternative forms of shelter, which range from inner city slum tenements and peripheral squatter settlements to the pavements (sidewalks) of major cities. The shortage of adequate housing is exacerbated by high rates of popu lation growth.

The majority of the poor live in slums or squatter settlements. Squatting refers to either the occupation of land without the permission of the owner or the erection or occupation of a building in contravention of existing legislation. The illegal status of squatter settlements provides government with complete legal justification for any 'remedial' actions against squatters. It also generates insecurity among squatters which inhibits their partici pation in city life, including access to services and amenities to which they are legally entitled, and may encourage clandestine occupations that further compli cate the squatters' relationship with the urban authority. Slums are defined as legal permanent dwellings that have become substandard through age, neglect, and/or sub division into smaller occupational units such as rooms or cubicles. Slums can develop either as a result of in adequate maintenance by a landlord, often prompted by rent control legislation, or by the internal subdivision of buildings by residents in order to accommodate new ar rivals or newly created households. Slum dwellings shelter large numbers of the urban poor in the Third World.

See also: Chinese Urbanism; Edge Cities; Gated Communities/Privatopias; Gentrification; Ghettos; Homelessness; Housing, Rural; Slums; Squatter Settlements.

Further Reading

Andrusz, G., Harloe, M. and Szelenyi, I. (1996). *Cities After Socialism*. Oxford: Blackwell.
Harriot, S. and Matthews, L. (1998). *Social Housing: An Introduction*. London: Longman.
Keivania, R. and Werna, E. (2001). Modes of Housing Provision in Developing Countries. *Progress in Planning* 55, 65 118.
Kennett, P. and Marsh, A. (1999). *Homelessness*. Bristol: Policy Press.
Lowe, S. (2004). *Housing Policy Analysis*. Basingstoke: Palgrave Macmillan.
Ma, L. (2002). Urban Transformation in China, 1949 2000. *Environment and Planning A* 34, 1545 1569.
Malpass, P. (2001). The Restructuring of Social Cented Housing in Britain. *European Journal of Housing Policy* 1(1), 1 16.
McCrone, G. and Stephens, M. (1995). *Housing Policy in Britain and Europe*. London: UCL Press.
McDonnell, J. (1997). The Role of Race in the Likelihood of City Participation in the United States Public and Section 8 Existing Housing Programmes. *Housing Studies* 12(2), 231 245.
Ravetz, A. (2001). *Council Housing and Culture*. London: Routledge.
Smith, J. (1999). Youth Homelessness in the UK: A European Perspective. *Habitat International* 23(1), 63 77.
Smith, D. and Scarpaci, J. (2000). Urbanization in Transitional Societies. *Urban Geography* 21(8), 745 757.
Takahashi, L. (1996). A Decade of Understanding Homelessness in the USA. *Progress in Human Geography* 20(3), 291 310.
van Vliet, W. (1990). *International Handbook of Housing Policies and Practices*. New York: Greenwood.
Varady, D., Presier, W. and Russell, F. (1998). *New Directions in Urban Public Housing*. New Brunswick: CUPR Press.

Housing, Neighbourhoods and Health

J. Dunn, St. Michael's Hospital, Toronto, ON, Canada

Glossary

Community Group of people having common interests, who may also share a common geographical area of residence.

Deprivation Index A statistic calculated from widely available data about small geographic areas (usually census data) to represent the level of social and economic deprivation of a given area.

Home An environment, often a house, which people invest with meaning; often it offers security, refuge, and happiness. Home may also refer to a place of origin, either for one's daily activities or one's life.

House A fairly permanent structure serving as a regular dwelling or residence for one or more persons; these persons form a household.

Multi-Level Modeling A statistical method for analyzing data that has a nested structure, for example, individuals within geographical regions. It allows inferences to be drawn about the effects of region-level factors on individual outcomes, while controlling for individual-level effects.

Neighborhood A geographically bounded area, at least partially residential, whose inhabitants may share a common set of identity and daily activity patterns and resources.

Introduction

Over the past several years there has been an increasing interest in the so called 'social determinants of health', and a significant amount of research on the effects that experiences of everyday environments have on health and inequalities in health between social and economic groups. Although initially much of this research focused on characteristics of the workplace and its effect on health, there is a growing interest in the effect of housing on health and now quite a large literature on the effect of neighborhood environments on health. These developments have their roots in research that emerged in the 1970s and 1980s, showing large inequalities in health status by socioeconomic status (SES). It is widely recognized that these inequalities in health status are the result of the cumulative, life long impact of differences in everyday living conditions, which manifest in health differences through identifiable biological processes related to maladaptive stress responses. In geography and a number of other social sciences, a great deal of emphasis has been placed on understanding the important dimensions of housing and neighborhoods for health in recent years.

The Social Gradient in Health and Everyday Living Conditions

There is an increasing recognition in research in the social and health sciences that the everyday environments in which we live exert a significant influence on human health and well being. This perspective, widely known as the 'social determinants of health', has its roots in research that shows large and persistent inequalities in health status by socioeconomic status. Although the term 'determinants of health' is widely used, it is an imprecise name for the phenomenon, which might be more accurately called 'identifiable patterns of factors that influence health'.

Numerous studies, largely in the discipline of epidemiology, have demonstrated that there is a so called 'social gradient in health' whereby individuals of greater socioeconomic status have better health status. **Figure 1** shows an example of a social gradient in health, in this instance showing the relationship between occupational grade in the UK's civil service and the 10 year risk of dying of coronary heart disease. It shows that people in the lowest occupational grade ('Other') have a 4 times greater chance of dying of coronary heart disease than people in the highest occupational grade ('Administrative') and that these differences cannot be explained away by known behavioral risk factors for coronary heart disease. In addition, the relationship between socioeconomic status and health is one that spans all social strata; it is not merely a function of extreme deprivation, the differences exist even among people who live well above subsistence and affluent people. Social gradients exist for most major causes of illness and death in the affluent societies of the world and have persisted for more than a century.

Many researchers believe that the social gradient in health reflects the fact that SES is a marker for significantly different conditions of everyday life experienced across the life course. Some suggest that the more important of these everyday living conditions are 'psycho social' in nature, that is, they concern the meanings attached to their material circumstances and the negative health effects occur as a result of social comparisons with others who have more enviable situations. Many other researchers, however, focus on how material conditions of everyday life cause differential distribution of everyday stress by SES. In other words, the lower one's

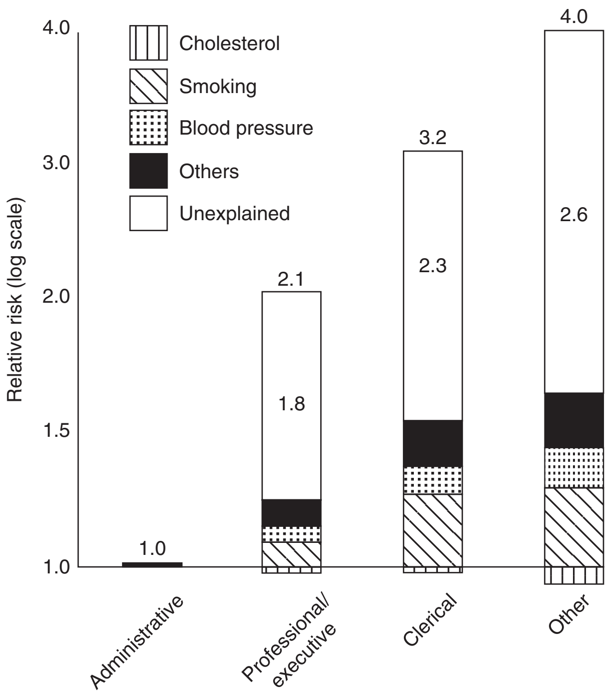

Figure 1 Social gradient in health: as example.

socioeconomic status, the greater the stress experienced and the fewer resources (e.g., economic, personal, social) to respond to the experience of relatively low socio economic status and its attendant stresses. Because housing and neighborhood conditions are such important spaces in the geographies of everyday life, and can be conceptualized both as sources of stress and as resources, it follows that attributes of housing and neighborhoods should be important influences on health.

Housing as a Determinant of Health

Given that housing is a social phenomenon with multiple dimensions it behoves us to identify the attributes of housing that have the potential to influence health. This kind of knowledge is necessary to inform empirical re search investigating the relationship between housing and health, and ultimately to inform successful housing interventions, like subsidized housing and public housing redevelopment. Theoretical and empirical research has shown that the following dimensions of housing may impact upon health either directly or indirectly: physical hazards, physical design, psychological dimensions, social dimensions, financial dimensions, and location of hous ing. Each is discussed in turn.

Physical Hazards

There are clearly well founded concerns about housing disrepair and exposure to toxins and hazards (molds, dampness, dust mites, falls, etc.) that substandard housing may entail. **Figure 2** shows the wide variety of biological, chemical, and physical hazards within the home that may threaten health. Molds, bacteria, dust mites, and pet dander are examples of indoor pollutants that can cause respiratory symptoms. Radon is a naturally occurring gas that is released from the Earth and can enter into houses through the basement. It has a strong link with lung cancer. There are a number of chemicals used in the home for cleaning and other purposes, as well as chem icals that are released from carpets and furniture. Phys ical hazards include improper or dangerous stairs, handrails, and other physical aspects of the home. Not withstanding the importance of bio physical hazards on health, there is evidence that social and economic factors may mediate the experience of health impacts even in the presence of empirically verifiable exposures. One study has shown that the experience of respiratory symptoms among households exposed to molds was significantly mediated by worry about mold.

Physical Design

Aspects of physical design can contribute to health status in a variety of ways. Disrepair of stairs, floors, safety equipment, etc., may create the possibility of falls and other accidents, particularly for people who are frail. Similarly, the absence of specialized adaptive equipment (e.g., handrails) or inadequate fire escape routes may also be threats. The second way in which physical design may affect health is through a psychological pathway. Archi tectural design may provide high levels of complexity, for example, which may result in an overabundance of stimulation and cause stress. For example, design features, like inadequate differentiation of space or long, narrow corridors, may enhance the user's sense of control (mani fested as privacy or the inability to regulate exposure to one's surroundings). Design features such as the amount of available space, visual exposure, structural depth, openness of the perimeter, and extent of view have all been shown to moderate effects of crowding on human behavior.

Psychological Dimensions

The key psychological dimensions of housing, namely the importance of the home as a site for the expression of identity, social status, and prestige, on the one hand, and control on the other hand, have emerged from the dis ciplines of environmental psychology and geography. Analyses of real estate advertising show that adverts are selling an identity as much as they are selling the physical structure of the house. A recent advertisement for a housing subdivision in Canada tells the reader that they can "live life perfectly" in the "Picket Fences" sub division. There is no mention of the quality of the building materials or the builder. Instead, the sales pitch

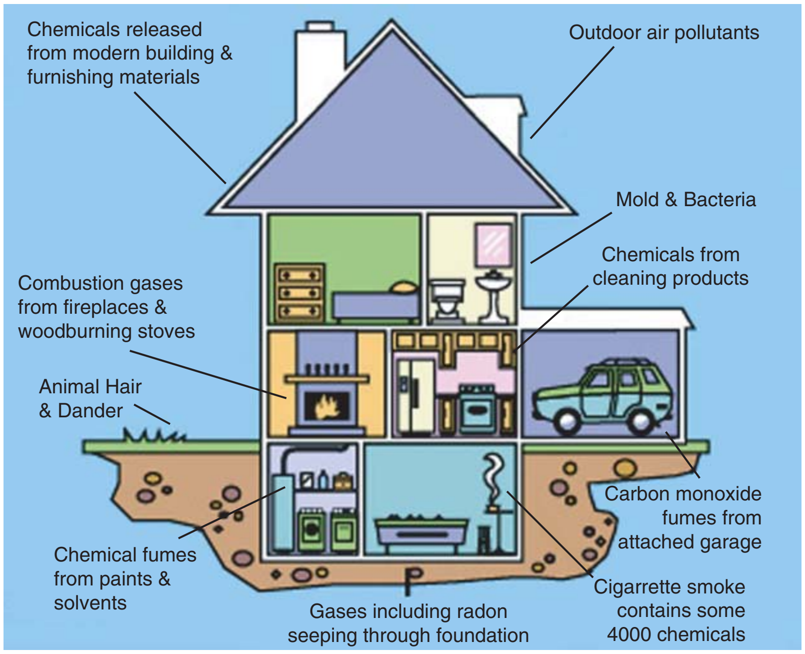

Figure 2 Biological, chemical, and physical hazards of housing.

is focused entirely on enticing the reader to identify with the description of the development. The advertisement describes how the owner will feel as a resident in the community of "fully built" condominiums, urging readers to get on the "priority" buyers list. The identification of the meaning invested in the home as a determinant of health is consistent with arguments that the social gradient in health is partly a product of the meanings people attach to their material circumstances.

Regarding housing and control, the literature on job strain and health indicates that lack of control at work can be influential on the development of heart disease risk factors and can undermine mental health. It follows control is also important outside of the workplace and within the home. There is some empirical evidence to support the proposition that control at home is associated with health. As described above, physical design is one way in which control is manifested in housing, but financial aspects of housing are also important for providing control, as described below.

Social Dimensions

There is great sociological significance attributed to the boundary between the inside of the home and the outside world and this is common to almost all cultures. To invite someone into your home as a guest is a symbol of escalating a relationship, making it closer. The home is an important site for the development and maintenance of social relationships, both with household members and others. The importance of social support is already very well established in the health literature as a significant determinant of health. It follows that the adequacy of one's home both for making and maintaining social ties is an important health issue, although the direct research on this question is still underdeveloped.

Financial Dimensions

Housing costs are the single largest item in most households' monthly budgets and one of the largest assets in their investment portfolio. It follows that financial dimensions of housing interact strongly with socioeconomic status. There are several ways in which housing markets work to redistribute wealth and income in a regressive fashion (i.e., to the disadvantage of lower income households). In Anglo American society, much of this regressive redistribution hinges on the distinction between owners and renters. In most Anglo American countries, the main tax advantage that fuels the regressive redistribution from owners to renters is the capital gains tax exemption for the primary residence. In such countries, capital gains (i.e., the difference between the price one pays for a property for and the price one sells it for) are tax exempt, subject to certain caps. Some countries have a lifetime maximum of exempt capital gains, some have a per transaction cap, but nevertheless, in call cases, this offers owner occupier the opportunity to accrue a great deal of wealth, tax free. This represents foregone revenue for the governments of these countries, and some have questioned the fairness of offering such significant indirect subsidies to homeowners at the same time that governments are disinvesting in publicly funded housing for low income households.

While numerous studies point to the relationship between income and health, it is likely that wealth is also an important indicator of a household's socioeconomic status, and if so, this would underscore the importance of housing given that it is commonly the single most im portant asset of most households. There is some evidence that inequalities in housing wealth are associated with health inequalities.

Location

The locational dimensions of housing are potentially important to health because the home acts as a focal point for everyday activity and offers relative proximity to both health risks and resources, as well as being the site for relationships with neighbors and members of a geographically defined community, who arguably have a number of common interests. Also important to the lo cational dimensions of housing are the meanings that are attributed to the neighborhood of residence, either by people living there, or by outsiders. The locational di mensions of housing, in other words, provide the gateway to an examination of the influence of neighborhoods on health.

Neighborhoods and Health

Over the last several years there has been an increased interest in the effect that attributes of the residential nei ghborhood in which one lives may have an impact upon health over and above the individual's socioeconomic status. By neighborhood, most scholars refer to a geo graphically bounded area around a person's home that carries some common meaning to its residents. The effect of neighborhood attributes on health is particularly rele vant to understanding health inequalities, because as Sally Macintyre, a leading scholar in this area, has argued, the added effect of neighborhoods over and above individual SES may represent a case of 'deprivation amplification'. Another line of reasoning in the neighborhoods and health literature is that health behaviors are significantly con ditioned by attributes of the local environment, or the 'health opportunity structures' a term also coined by Macintyre. A person living in a neighborhood without sidewalks or recreational facilities, for example, lives in an area that structurally lacks opportunities for physical activity.

The implications of these perspectives on neighbor hoods and health are that one's home and its immediate environment is likely to be the setting for exposure to a mix of positive and negative influences on health. One example of the importance of the locational dimensions of housing can be seen by considering the location of the home relative to services and amenities such as schools, public recreation facilities, health services, and job opportunities. This may also explain part of the so called neighborhood effects on health and human development seen in the research literature. Another important aspect of the spatial dimensions of home is the social environ ment it places one in, particularly with respect to social norms. Moreover, housing market dynamics confer sig nificant locational advantages upon households in ways that systematically disadvantage households of lower SES. A location in space, a place to base one's activities and existence, a place that provides access to goods, services, work, and recreation are important functions of housing – location matters.

Roots of Neighborhoods and Health Research

Research on neighborhoods dates back to the early work of the Chicago School of Sociology working in Chicago in the 1920s and 1930s. Some of the notable members of the School included Robert Park, Ernest Burgess, George Herbert Mead, and Roderick McKenzie. These scholars studied the rapid changes going on around them in Chicago during this time, which was a period of rapid urban growth and high levels of immigration. They de veloped a concentric zone model to explain the typical residential location and relocation pattern of households new to the city and to explain the spatial differentiation of social conditions of the time. Several years later, Social Area Analysis became a popular approach to the study of neighborhoods and other geographic regions.

After this initial popularity, however, research on neighborhoods diminished in popularity. A significant milestone in the decline in interest in neighborhoods was work that suggested that the territorial basis for com munity life had been eclipsed by social networks. Key figures here include Barry Wellman and Melvin Webber. No longer important for people's social relationships and their 'community', the argument goes, was the immediate geographical space around one's residence, but instead, it was their social networks that mattered. What mattered, in the words of Webber, was 'community without pro pinquity'. This proposition opened up whole new areas of social inquiry, but led to the temporary disappearance of neighborhoods research.

In the 1990s, however, a series of factors led to the re birth of neighborhoods research in geography and other fields, and research on neighborhoods and health has been a leader in these developments. The increasing availability of geo coded data on small areas from census and other sources as well as geo coded survey data from a number of sources made it possible to download enough data to conduct sophisticated statistical analyses on the 'effect' of neighborhoods on health or other outcomes. In geography, the first users of a new class of statistical techniques, multilevel models, were scholars studying the effects of area of residence on voting behaviors. Much of this research that people's voting behavior is very similar to that of the people they live near. This method has been adopted widely in social epidemiology, geography, and sociology to study the effects of neighborhood factors on health, while controlling for individual attributes. The technique requires quite large samples of individuals and neighborhoods, combined in a database where individuals are 'nested' in their neighborhood of residence. Coupled with the widespread availability of powerful desktop computers, these data and sophisticated statistical techniques led to a proliferation of studies examining the effect of neighborhood on health.

Conceptualizing Neighborhoods and Health

One way of conceptualizing the specific health opportunity structures that underlie neighborhood factors that have been demonstrated or hypothesized to affect health is by categorizing them as one of four 'R's': risks, resources, relationships, and representations. These four R's tend to demonstrate significant spatial variation, meaning that people living in different neighborhoods have different levels or types of exposure.

There are a wide variety of health risks that can vary by neighborhood and have an effect on health. Examples include environmental contaminants, crime, violence, noise, and dangerous roadways or pedestrian pathways or crossings. Resources that may affect health and also vary by neighborhood include health and social services, commercial services like stores that sell healthy foods, schools, recreational opportunities, libraries, public transportation, and a built environment that allows for physical activity. A large number of studies of neighborhood effects on health have investigated neighborhood relationships. Analogous to the large literature demonstrating a relationship between an individual's level of social support and their health (i.e., with close friends and family), research has demonstrated relationships between neighborhood social relationships and health. One of the most popular concepts to be employed in such studies is 'social capital', which refers to actual or potential resources one can access as a result of having relatively durable network of relationships with others. In a health context, this research expands upon the evidence on social support and health to consider the effects of 'weak ties' on health, a term famously coined by sociologist Mark Granovetter. Others have shown a relationship between 'collective efficacy' and health. This is a concept somewhat similar to social capital, but is more focused on how residents of the same community can (or cannot) work together to solve problems and confront threats to their community, especially with respect to youth deviance and crime. An in depth study of the reasons why one poor neighborhood had more deaths in the Chicago heat wave of 1996 than a comparable neighborhood showed that neighborhood social capital was a major factor in avoiding deaths from the heat. Similar results have been shown for a number of health outcomes. Finally, representation of neighborhoods is concerned with the way the neighborhood is represented both by its residents and by outsiders. There is evidence that the reputation of a neighborhood is associated with health status, and this may be the result of both direct and indirect effects. Residents may internalize the poor reputation of their neighborhood resulting in poor self worth and/or mental health. Indirectly, the reputation of a neighborhood may have a self fulfilling effect, whereby services, infrastructure, investment, and migration reinforce and reproduce the negative features of a neighborhood.

These developments clearly signal an exciting future for neighborhoods and health research and policies informed by such research. Decision makers in several jurisdictions around the world have already embraced place based policies and area based initiatives, although these have not been fully informed by research evidence. One challenge to providing such evidence is that there is no clear answer to the question 'do neighborhoods affect health?'. The answer is that some attributes of neighborhoods may affect some health outcomes for some population subgroups in some places some of the time. This implies the need for better theory about the specific pathways linking neighborhood attributes and health, and more empirical research to investigate those theories.

See also: Health Inequalities; Wellbeing.

Further Reading

Badcock, B. (1984). *Unfairly Structured Cities*. Oxford: Basil Blackwell.
Curtis, S. E. (2004). *Health and Inequality: Geographical Perspectives*. London: Sage Publications.
Despres, C. (1991). The meaning of home: Literature review and directions for future research and theoretical development. *Journal of Architectural and Planning Research* 8(2), 96 115.
Dunn, J. R., Frohlich, K., Ross, N. A., Curtis, L. and Sanmartin, C. (2005). Role of geography in inequalities in health and human development. In Heymann, J., Hertzman, C., Barer, M. & Evans R. G. (eds.) *Healthier Societies: From Analysis to Action*, pp 237 266. New York: Oxford University Press.
Evans, G. W. (2003). The built environment and mental health. *Journal of Urban Health* 80(4), 536 555.
Evans, G. W. and McCoy, J. M. (1998). When buildings don't work: The role of architecture in human health. *Journal of Environmental Psychology* 18, 85 94.
Evans, R. G. (2002). *Interpreting and Addressing Inequalities in Health: From Black to Acheson to Blair to…?* London: Office of Health Economics.
Gatrell, A. C. (2001). *Geographies of Health: An Introduction*. London: Blackwell.
Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology* 78(6), 1360 1380.
Klinenberg, E. (2003). *Heat Wave: A Social Autopsy of Disaster*. Chicago, IL: University of Chicago Press.
Macintyre, S., Ellaway, A. and Cummins, S. (2002). Place effects on health: How can we conceptualise, operationalise and measure them? *Social Science and Medicine* 55, 125 139.

Marmot, M. (2004). *The Status Syndrome: How Social Standing Affects Our Health and Longevity*. New York: Times Books.

Park, R. E., Burgess, E. and McKenzie, R. (1925). *The City*. Chicago, IL: University of Chicago Press.

Sapolsky, R. M. (1998). *Why Zebras Don't Get Ulcers*. New York: W.H. Freeman.

Wilkinson, R. G. (1996). *Unhealthy Societies: The Afflictions of Inequality*. London: Routledge.

Wilkinson, R. G. (2005). *The Impact of Inequality: How to Make Sick Societies Healthier*. New York: The New Press.

Relevant Websites

http://secure.cihi.ca/cihiweb/dispPage.jsp ?cw page= PG 284 E &cw topic=284&cw rel=AR 1129 E

http://www03.cmhc schl.gc.ca/b2c/b2c/init.do?language=en&shop=Z01EN&areaID=0000000043&productID=00000000430000000025

www.euro.who.int/Housing

www.ncceh.ca/en/non ncceh reviews/other/housing health

Housing, Rural

N. Gallent, University College London, London, UK

Introduction

At first glance, the housing problems confronting people and communities in the countryside – at least in the developed world – seem to pale against those afflicting urban areas and especially inner cities. The association of housing shortages and low housing quality with crime and safety, unemployment, and a variety of forms of social exclusion has often been viewed as a largely urban phenomenon. This was certainly the case, across Europe and North America, at least until the late 1970s, with this view being driven by academic interest and assumed political priority. But the last 20–30 years have seen an awakening across these continents to the social, economic, and political challenges posed within the field of rural housing.

In the UK, government statistics in the 1980s revealed startling evidence of sometimes atrocious housing conditions facing the residents of rural areas, compounded by endemic housing shortages, low wages and unemployment, and geographical inaccessibility. Similar observations were made concurrently or subsequently around the developed world: the countryside was not immune from the economic and social pressures bearing down on Western society, nor was it a safe haven from crime and social exclusion. The 'rural idyll' – a previously popular image of a countryside inhabited by contented workers living in thatched cottages or the vernacular equivalent – was discredited, and a new generation of researchers found a rich vein of social malaise and inequality to explore outside of major towns and cities.

New pressures on rural society had been highlighted by scholars since the 1930s. The economic crisis of 1929–32 hit rural America particularly hard, leading to a sudden depopulation in many agricultural areas and the disintegration of communities, leaving abandonment and dereliction in its wake. Social issues flowing from less dramatic, but no less difficult, 'rural restructuring' elsewhere once again attracted the attention of rural sociologists and policymakers in the 1950s and 1960s. In many parts of Europe, the depopulation attendant to changing economic circumstances was accelerated in the post war period by new investment in mechanized farming which aimed to achieve 'food security' and protect European agriculture from global competition. At the same time, an emergent nostalgia for the countryside in many developed countries contributed to a process of counterurbanization. Researchers began to highlight the decline of traditional 'farming communities', arguing that the social structure of rural areas was now subject to revolutionary change: communities comprising people sharing the common experience of a land based occupation were being supplanted by mixed communities dominated by middle class incomers who hoped to find respite from urban living, or who simply took advantage of cheaper housing opportunities in the villages cum commuter towns that were now being sucked into larger urban conurbations across North America and Western Europe through motorway, autobahn, or freeway development.

Across the developed world, studies charted the arrival of commuters and the subsequent growth of once isolated villages and market towns. At the same time, post war economic prosperity fueled demand for second and retirement homes within both urban hinterlands and more remote rural locations. In the UK, commentators began to talk about the 'consumption' of the countryside and planning regulations were tightened to preserve the country's 'rural resource'. Elsewhere, this investment in rural areas was often welcomed and encouraged, being seen as the antidote to rural decline with consumption bringing much needed investment to areas that had suffered decades of abandonment. The historic practice of maintaining a rural second home was rediscovered in many parts of Europe, especially in the south and in Scandinavia, bringing an economic lifeline to many areas. In Canada and the US, rural resort development was actively encouraged, with second home and 'vacation cottage' ownership becoming a feature of local economic strategies.

But these social and economic changes – expressed in a restructuring of rural society – have placed new pressures on housing markets. The arrival of affluent newcomers has also concealed a concomitant rise in rural poverty. The rich have often moved in and bought up land and homes, but remnants of the old farming communities remain: these have often been labeled 'local households', who suffer disadvantage relative to the economic power of newcomers. The new rural affluence associated with counterurbanization (both domestic and international) has been balanced by a new rural poverty affecting some sections of this restructured society. In some areas, commentators have argued that this is because the arrival of newcomers has triggered intolerance to land use change as planners try to protect the countryside from further development. Hence, the sup ply of homes has decreased while demand has acceler ated. Affluent incomers compete ever more aggressively for land and property, while poorer sections of rural communities struggle to secure affordable homes, held back by low rural wages and frustrated by the way in which some planning systems seem to prioritize en vironmental goals over social ones. These pressures have been most acute in countries where nostalgia for the countryside – expressed in strong demand for rural homes – has prompted the development of ever more restrictive planning policies: Great Britain is the classic example. On the other hand, restrictions on rural de velopment have been less rigid where there is a culture of tolerance toward rural families meeting their own needs through self build. This tolerance has been manifest in sporadic rural development in southern Europe (espe cially in Italy) and in the Republic of Ireland. Elsewhere, development pressure has been less acute either because of the inaccessibility of rural areas (as in the case of Norway or Denmark) or because the stock of vacant rural housing is so great that it has not yet been exhausted, as in France.

Four themes have, to varying degrees, shaped the nature of the 'rural housing question' in different coun tries. These have all been alluded to in the above intro duction, but are now examined systematically. The first is counterurbanization and the degree of interest in rural housing markets; the second is the extent to which local populations can face this incursion on 'equal terms' with incomers, having secure incomes that enable them to compete in stronger markets; the third (and consequen tial theme) is the social exclusions that may result from economic change and counterurbanization; and the fourth is the planning or general policy response and whether this accommodates the 'rhythms of rural re structuring' or simply creates unnecessary difficulties for some sections of rural society.

Increased Interest in Rural Markets

Second homes symbolize the growing interest in rural housing markets since 1945. Though not a new phe nomenon, there has been a 'social democratization' of second home ownership over the last 60 years: more people own them and they are no longer the preserve of the conspicuously wealthy. In Europe, levels of second home ownership among the general population vary from around 5% in Britain and Ireland to more than 20% in Scandinavia. They are indicative of the fact that housing in rural areas is not only consumed by people who live and work in these areas, but also by those who recognize a recreational potential in the countryside. The rise of second home ownership since the war, particularly from the 1960s onward, sits in marked contrast to the flight from many rural areas in the first half of the twentieth century. In most countries, second home investment is seen as an opportunity rather than a problem, with many rural areas dependent on the money spent by seasonal residents on rural goods and services. But second homes represent the 'luxury' component of a much wider in vestment in rural housing. Road building and increased mobility, across Western societies, have opened up the countryside to populations who were previously far more constrained in their residential location choices. Families, who were once tied to inner urban locations because of work commitments, have snapped up the opportunity to move to growing commuter towns in large numbers. This has sometimes had a 'suburbanizing' effect on near urban countryside (as in the case of the Netherlands) or has resulted in a car dependent sprawl (as in the US). Else where, this growth has been tightly contained within key or 'holding' settlements, which have swelled with the arrival of post war commuters, but which have retained fixed planning boundaries, often being contained within some form of statutory green belt.

These two broadly different experiences have each produced their own problems. In Europe and North America, countries and states have grappled with the basic 'problem' of more people moving to rural areas: second home buyers, holiday home investors, retiring households, commuters, teleworkers, or lifestyle 'down shifters'. Some responses build on the premise that the countryside can accommodate more people, that people have the right to live where they wish, and that devel opment is an intrinsically good thing. Elsewhere, policy development reflects a different attitude: that the coun tryside is an 'environmental good', that openness needs to be preserved, that development is not intrinsically posi tive, and that there should be a broad presumption against housing and associated land use change in rural areas. Different 'representations' of the countryside tend to steer policy development: competing views on what the countryside is 'for' – a context for peaceful enjoyment versus a place for people to work and live – is reflected in national planning frameworks, in attitudes toward de velopment, and in local plan making and implementation.

Two countries that stand in stark contrast to one an other in terms of the treatment of their respective coun trysides, and in terms of their consequent rural housing issues, are the Republic of Ireland and Great Britain. Since breaking away from the United Kingdom, Irish planning policy has come to reflect a very different attitude to land and property rights and to the 'function' of rural areas. A 'working countryside' view has prevailed, with housing built where it is needed or where there is a demand. Local landowners have a right to profit from their landholdings, which has resulted in a proliferation of single dwelling developments across the Irish countryside. Many commentators have decried the environmental cost of this attitude toward Ireland's rural areas and the development free for all that has ensued, ostensibly designed not only to serve an economic agenda but also grounded in a belief in local development rights. In Great Britain, on the other hand, these rights were nationalized in 1947 and both pre war planning debate and post war legislation focused on the inappropriateness of development in open countryside. An 'environmental' view has prevailed, and since the 1960s, some villages (with tight settlement boundaries) have seen almost no growth whatsoever, while the growth that has been permitted elsewhere has been confined to 'key settlements' or market towns. The upland landscapes of Scotland and Wales have been largely preserved, while the agricultural landscapes of lowland England have been transformed by factory farming rather than sporadic housing development. But the number of disgruntled environmentalists in Ireland is at least matched in Great Britain by the army of researchers pointing to the unaffordability of housing in many rural areas. Official statistics have shown that, taken as a ratio of median local incomes to mean local house prices, housing is less affordable in English villages and hamlets than in cities. While a majority of people could afford to purchase a home in rural England if they so wished, a significant rural minority could not.

Hence, these two attitudes – and two policy responses – underpin two different outcomes: development with a heavy environmental cost or protection at the price of social exclusion. Some middle road approach would seem sensible, but in neither the Republic of Ireland or in Great Britain has this been possible owing to the dogged defense of entrenched positions (i.e., a fundamentalist view on property rights versus an equally fundamentalist view on 'rural conservation'). Elsewhere, responses have been more balanced, but wherever planning plays a part in controlling development there remains a possibility that the cost of buying or renting a home will move beyond the reach of some households. Planning can create a scarcity in land supply that causes an elevation in property prices, which may be problematic for some local households because of the weakness of rural economies relative to the urban economies from which new residents derive their income. This problem is often associated with Britain and the housing supply and demand imbalances that can arise from development constraint, and which concurrently affects rural economic development. But many other countries have laggard rural economies, creating a rural underclass locked out of formal rural housing markets, and forced to occupy substandard housing or engage in illegal building. In areas of high 'recreational' demand – and subsequent planning constraints which aim to maintain the environmental qualities which attract tourists and investors to the countryside – low wages in land based or service sector occupations may not provide local people with the means to access local housing.

Rural Economies and Ineffective Demand

'Ineffective demand' is a phrase that has been coined to refer to a situation where the power of local households to compete in a housing market is either nullified by external competition or eroded by a weakening rural economy, which may result in farm wages – in particular – being unable to keep pace with wage levels elsewhere. In the early 1980s, it was observed that the "the essence of the housing problem in rural areas is that those who work there tend to receive low incomes, and are thus unable to compete with more affluent 'adventitious' purchasers from elsewhere in a market where supply is restricted." This observation was borne out with evidence from the English Lake District, but it can apply in any context where these three basic ingredients – planning constraint, low rural wages, and strong migration pressure – are combined. The poorer and more affluent sections of rural society occupy different 'domestic property classes'. In many rural areas, in Europe and North America, the rural poor have historically been tenant farmers or farm laborers. Because they tended to rent their homes, they had little accumulated wealth, and because they were generally unskilled, their incomes were limited. These two characteristics – limited wealth and limited income – placed them in a disadvantaged domestic property class relative to many incomers, who had both wealth and substantial income. This is of course a generalization as the expansion of the European Union in the 2000s has generated a wave of low wealth and low income migrants who have moved westward to take up unskilled laboring jobs. But this in itself has compounded the plight of many existing rural residents who have seen their incomes fall in response to labor oversupply. The same economic disadvantage is experienced by those in the service sector, especially unskilled workers who serve the tourist economy. Like the farm workers in the previous generation, they rent their homes, receive low incomes, and because they occupy the bottom rung of the labor market ladder, may find themselves in direct competition for jobs with migrants from poorer countries. In the southern US, Mexican migrants – legal and illegal – compete with US citizens for unskilled farming and service jobs. In Western Europe, migrants from the A8 accession states have arrived in large numbers. The majority have headed to the major cities, but hundreds of thousands have taken up rural jobs in farming and in construction.

But it is the combination of low wages with housing demand from external purchasers and planning constraints that may lock sections of rural communities out of the market for rural homes, causing a range of housing stresses: overcrowding, homelessness, or the occupation of unfit dwellings. In an open economy, no market is free from external interest: people move in and move out, producing a 'churn' which, alongside land and housing supply, affects house prices. The imbalances that occur in rural areas are not unique. In an 'urban system', those with less income may be forced to move to a less expensive neighborhood, resulting in patterns of socioeconomic segregation. But in the countryside, more limited housing opportunities (and a critical difference in scale) are more likely to cause wholesale displacement 'from' some settlements (especially villages) than segregation 'within' settlements. There are many well known and well documented rural 'hotspots' across Europe where the gentrification of villages and small towns has been spectacular, causing a near total displacement of existing residents in favor of both domestic and international migrants: established examples include the Dordogne, Province, and Courchevel in France; Tuscany and the Costa Smeralda in Italy; parts of the Austrian Alps, especially around Kitzbuhel; and Cornwall, the Cotswolds, and the Lake District in England; emerging examples can be found in Bulgaria and Romania. Further afield, there are now emergent second home markets across Central America, though these tend to be concentrated in the Pacific resorts (including Puerto Vallarta in Mexico) rather than in the rural interior. But the principles driving social change are the same: the arrival of affluent incomers who wield considerable economic power relative to local residents, and a planning system that prioritizes incoming consumption interests above the broader economic needs of the existing population.

New Social Exclusions

Social change has created new tensions and strains in the countryside, particularly between the 'new communities' of commuters, retiring households, holiday makers, down shifters, and investors and the 'residual communities' comprising the remnants of what might be considered the 'traditional' land based class. It has been argued that today's rural communities are 'encapsulated' in the sense that a residual group (what's left of the old land based occupational class) is encapsulated within an increasingly dominant community of 'newcomers' and that the underlying threat of conflict is now a key feature of rural life. The potential for new social exclusions (from services and from wider opportunities – especially education and employment opportunities – as well as homes) is central to this conflict, often because the needs of newcomers and locals diverge, and communities can no longer build consensus around any shared vision of future development.

For example, incoming groups may believe that their own needs are best served by preserving the community as it is – by 'freezing it in aspic' – and advance an 'environmental' representation of the countryside. They have no need for new economic development or additional housing and consequently oppose new development proposals. Investors, second home buyers, retiring households, and commuters will wish to preserve the rural setting of their homes, maintaining open countryside and ensuring that the quintessential French, English (or New England), Italian, or Austrian village remains quintessential. This is easier to achieve where incomers are able to influence local planning decisions directly and in contexts which already strongly favor the preservation of 'rural character'; hence, Great Britain is often held up as the country with one of the seemingly most intractable rural housing questions.

Differences in the needs of 'local' and 'newcomer' groups may account for mixed reactions to development proposals, or the weight of support behind particular community projects. The desire to maintain a local bus service will be greater where more residents rely on that service and weaker where the majority are commuters or second home owners. Support for local schools is more likely to be found in villages where many people have young children, but will diminish where communities have become dominated by elderly and seasonal residents. Local shops, surgeries, leisure facilities, or postal services will only survive where they are regularly patronized, but cease operating when a high enough proportion of villagers shop on the way back from work (or via the Internet), prefer to open a bottle of wine at home, or no longer use post offices. As needs change, the residual community may find itself increasingly marginalized and disadvantaged: unable to access employment without a car (and unable to buy and service a car because of low wages), unable to send children to the now closed local school, and unable to find secure and affordable accommodation in the local area. This community may be squeezed on several fronts and ultimately displaced, with villages becoming dominated by more affluent groups. Relative priorities determine the importance attached to particular services, including new housing provision, though economies of scale (concentrating housing or school provision in key settlements), lower population densities (creating less profitable markets for bus companies or retailers), and a more restrictive approach to planning all of these contribute to social disadvantage in the countryside.

But housing is often seen as the key battleground: the reconfiguration of communities which is central to social exclusion is the product of a snowballing interest in rural housing markets since the 1960s. This form of housing demand is transforming many parts of the European and North American countryside and presenting policymakers with an apparently simple choice. Should this pattern of change be allowed to continue, or should stronger measures be taken to ensure more balanced communities in rural areas? Planning systems and different approaches to housing supply are likely to play a major part in shaping rural housing issues in the future.

Planning Systems and the Supply of Land for Development

Intervention is a key component of rural housing debate, playing a part in the formation of particular tensions, possibly by releasing too much land for development or too little and causing a range of environmental or socioeconomic stresses. Planning systems, underpinned by different values and attitudes toward the countryside, vary greatly, making it extremely difficult to identify a single mode of intervention that might ease tension or produce ostensibly better results (however these are defined). For some observers, the restrictive planning systems of the UK produce broadly desirable outcomes: the preservation of rural character at a tolerable social and economic cost. For others, these systems preside over a disaster, bleeding the countryside of its economic vitality and denying thousands of families the chance of a decent home. In Ireland, what some see as an environmental catastrophe, others view as simple common sense: allowing landowners and communities to develop freely without undue hindrance from government or from local bureaucrats. These two contrasting approaches operate under broadly similar conditions: the rural populations are growing in absolute terms, though interest in particular rural markets varies depending on environmental qualities, accessibility, and proximity to major centers of population. Other countries have different or more variable experiences: the US, for example, has its attractive rural communities in the New England states which have seen a steady influx of commuters during recent decades. But these areas contrast markedly with the mid west and the southern states, which have expansive rural areas that continue to experience depopulation. Likewise, many less accessible parts of Europe – for example, in northern Scandinavia, parts of the Spanish interior, or Italy's Mezzogiorno – have seen little of the investment interest experienced elsewhere in rural Europe, and when it does occur, it is actively encouraged.

Therefore, it is impossible to highlight a single mode of intervention that would suit all circumstances, but it is possible to note approaches that might be applicable depending on context. Building stronger and more resilient rural economies is regularly touted as favorite, and this can mean many things: improving regional infrastructure in order to link rural hinterlands to urban economies; relaxing planning constraints (where they exist) to encourage investment and enterprise (as well as diversification away from farming); and actively encouraging consumption investment, of which holiday home development and second home buying might be seen as a part. Next to the economic dimension, there has been inevitable interest in specific policies designed to increase housing access where this is limited. The most obvious is concerned with mounting a challenge to the culture of rural protectionism, opening up more land for development and establishing a new balance between housing demand and supply: basically, building more homes in an attempt to bring down prices, making homes affordable to low income, low wealth groups.

Countries with rigid constraint policies tend to continually debate this option. Recent analysis for the UK has shown that house price increases over the last three decades have been more than double those experienced in the rest of Europe, resulting in a call to relax planning constraint, build more homes, and seek 'improved market affordability'. The counter position is to argue that more house building will generate more investment long before it helps poorer households: therefore, far too high an environmental price would have to be paid in order to achieve this market affordability. Where more housing is needed, it should be contained within existing settlement boundaries. This very British view – grounded in a context of nationalized development rights – maintains that the countryside is simply not the right context for house building and if more homes are needed, they must be achieved through more efficient use of previously developed land: hence development densities should be increased, not only in urban areas but also in villages and market towns where development is permitted.

Of course, this view is not universally shared: rather than developing 'rabbit hutches on postage stamps', rural development elsewhere in Europe is characterized by single dwellings on large plots (Spain, Italy, and Ireland all display this tendency) and, moreover, by a freedom in building design and style that does not exist in the UK. Again, the contrast hinges on rights and on culture: the rights that individuals have to use and profit from private land holdings, and the culture of representation; of a countryside to be protected (arguably for the benefit of a select few) versus a countryside that comprises exploitable private and community assets. Protectionism and development intolerance in the UK have meant that attempts to increase the affordability of rural homes have been 'piecemeal' or 'innovative' depending on one's perspective: requiring landowners and developers to include low cost housing in developments where these are permitted, granting very exceptional planning permissions on land where housing would not normally be allowed, and working with quasi public bodies to provide cheaper homes for key workers or households with an accepted 'local need'. Such interventions have been necessary because of the extent to which planning constraint forces up land values and subsequent housing costs: the market price of a developable plot of land in southern England, for example, can be many hundreds of times the price of an agricultural plot of equivalent size. While rural households elsewhere in Europe and North America worry about the rising cost of labor and materials (and moot the benefits of flat packed, prefabricated, or timber framed homes as solutions to their rural housing problems), these issues are of less concern in parts of England where build costs are dwarfed by the cost of buying land for housing.

Concluding Remarks

There are many facets to the rural housing problems facing Western societies, and many contexts – geographical, environmental, sociocultural, economic, and political – in which these are played out. Yet a number of themes are conspicuous: the strongest is that of 'man versus nature', with rural housing debate in many countries centering on the balance that should be struck between development in the countryside and the protection of 'landscape resources'. But in recent years, there has been a move away – at least in reasoned debate – from presenting the rural housing question as one of stark choices: homes or landscape quality; jobs or openness; development or nature; concrete or trees. Sustainable development has been viewed, for a long time now, as a means of avoiding these dichotomous choices. Unfortunately, it can also become a banner for interests wishing to halt development in largely rural areas, or for the promoters of discredited development forms rebranded as 'eco towns' or 'eco settlements', being 'eco' in name only. However, consensus is now mustering around the view that planning systems must deliver for 'man and the biosphere', and avoid prioritizing one over the other. This means not only finding a place for development and change in rural landscapes, avoiding irreparable ecological damage, but also accepting the need to deliver living and working countrysides, with viable economies, decent services, and sufficient homes to meet need and demand. Delivering green development strategies, and green housing at the highest environmental standards, is seen as the way forward.

See also: Counterurbanization; Gentrification, Rural; Homelessness, Rural; Rural Communities; Second Homes.

Further Reading

Belden, J. N. and Weiner, R. J. (eds.) (1999). *Housing in Rural America*. Thousand Oaks, CA: Sage.

Evans, A. W. (1991). 'Rabbit hutches on postage stamps': Planning, development and political economy. *Urban Studies* 28(6), 853–870.

Gallent, N., Shucksmith, M. and Tewdwr Jones, M. (2003). *Housing in the European Countryside*. London: Routledge.

Murdoch, J. and Abram, S. (2002). *Rationalities of Planning: Development Versus Environment in Planning for Housing*. Ashgate: Aldershot.

Scott, M. (2007). Rural housing: Politics, public policy and planning. In Norris, M. & Redmond, D. (eds.) *Housing Contemporary Ireland: Policy, Society and Shelter*, pp 344–363. London: Springer.

Shucksmith, M. (1981). *No Homes for Locals*. Farnborough: Gower Publishing.

Relevant Websites

http://www.ruralcommunities.gov.uk
Commission for Rural Communities (United Kingdom).

http://www.rurdev.usda.gov
USDA Rural Development (United States).

Hub Network Location

M. E. O'Kelly, Ohio State University, Columbus, OH, USA

Glossary

Flow Economies-of-Scale A reduction in costs of operation per unit in response to higher levels of flow activity.

Hub A central facility serving as a switching point in a network; characteristic of the hub is a set of flows from and to origins and destinations and these connect via the switching points.

Hub-and-Spoke System In contrast to a directly connected point-to-point interaction system, a hub-and-spoke system channels flows from origins to destinations via central switching locations; typical journeys involve one or more intermediate stops.

Median A centrally located facility that serves as a distribution center or a warehouse without necessarily permitting transshipment or throughput at the central facility.

Network Optimization Specially structured optimization problems that aim to find the lowest costs of operations in complex network flow situations, often subject to numerous constraints that limit the flexibility of the decision maker.

Nonconvexity A type of objective function shape that tends to occur in hub models, and which does not readily yield a globally minimal transport cost.

NP-Hard A technical measure of the computational complexity of a problem.

Origin and Destination (OD) Flows occur between origins and destinations in a set of interacting systems; examples of such flows can include passengers, information, packets, and supplies. Most origins may have spoke-like links to one or more hubs.

Spoke A link joining an origin or a destination to a hub.

Background

Hubs are facilities that serve as switching points for flows in interaction systems with many origins and destinations; they are common in passenger and freight transportation systems as well as in telecommunication networks. Hubs are commonly used in distribution sys tems as a means of spatial organization. In such net works, many types of optimization problems arise: some that minimize communication costs (p hub medians) and some that optimize access and coverage (p hub centers). While these models bear a close resemblance to well studied classical problems in location theory, the special feature of the hub network design is that interactions between pairs of nodes give rise to interdependence in the systems of facilities. In the hub and spoke systems considered here, facilities must be placed not just with regard to single flows (as in flows to or from a factory or warehouse) but with reference to throughputs that de rive from the flows between origins and destinations. It is worth noting that in the absence of some kind of facility throughput, the analysis collapses to the comparatively simple 'median' family of problems. This is an important distinction because the term hub is occasionally em ployed rather loosely as a synonym for a central depot of some kind; such cases are adequately and more properly handled as medians. The focus in this article is on the hub location and network design problem where the pre defined flows between origins (I) and destinations (J) are routed through one or more central switching points known as 'hubs'. It is clear that the number, location, and routing protocols employed have significant consequences for the cost of operation of a particular proposed network. In turn, these decisions become imbedded in optimization formulations that attempt to generate efficient networks within the framework of some particular network protocol.

One such formulation is to pick facilities (denoted say by selecting some subset of p locations from n candidates) so as to minimize the sum of costs of routing the flows; clearly, the objective must add up the routing cost over all origin–destination pairs (i.e., over all I and J pairs) as shown in the following computational formula:

$$T = \sum_i \sum_j W_{ij} \sum_k \sum_m R_{ijkm}[C_{ik} + a \ C_{km} + C_{jm}]$$

where W_{ij} is the flow between i and j, and R_{ijkm} is the variable that denotes the routing and the C_{ik}, C_{km}, and C_{jm} terms (which are constants or 'coefficients') measure the link costs per unit of flow. A notable idea is that the interhub portion of the trip might operate at some dis counted fraction ($a<1$) of the prevailing costs, due say to some extra capacity infrastructure. For example, wide body aircrafts can be used on busy interhub routes. The correct formulation of this effect is the topic of some ongoing research into economies of scale, as discussed further below. The decision variable R_{ijkm} (typically an integer variable that is set to either one or zero) is clearly governed by the decision about which hub facilities to open, as well as the allowable flexibility in allocating nodes to facilities. One particular version of the objective function and constraints leading to a quadratic integer programming formulation was set out in a seminal paper by O'Kelly.

A characteristic feature of hub and spoke networks is the bundling of flows on the interhub links. The agglomeration of flows leads to reduced average travel costs across the interhub links. Providers of transportation services may reduce their average unit costs by bundling flows and channeling them between hubs (also known as concentrators or routers). This aspect of the problem has a number of consequences: introducing flow related discounts into transportation costs, for example, and focusing on the entire system of flows. This feature contrasts with the traditional 'transportation problem' of linear programming where the flows are charged a rate that is independent of the amount of flow. The hub location problem has been a fruitful one for current methodological research (e.g., in operations management) precisely because of the inherent programming difficulties posed by this interdependence.

As a result of this interdependence, the objective function and/or the number of variables in the models need to be modified from the comparable standard location models. Specifically, some hub location models have quadratic rather than linear objective functions and others retain linear objective functions but require large numbers of variables (compare, for example, using R_{ijkm} vs. use of the product of X_{ik} and X_{jm}). Initial research focused on the special case where each origin/destination is allocated to a 'single' one of the fully interconnected hubs. An important insight into this hub location problem is that in contrast to 'medians', which possess a 'nearest is best' allocation property, the hub problem cannot assume that allocation to the nearest hub is best because the bulk of a node's interactions may actually lean toward relatively faraway subsets of the destinations.

As an extension, observe that a broad range of connection protocols can be envisaged including cases where the nodes interact with more than one of the hub locations. A tension occurs between the desire to provide optimal linkages for every pair (leading to a tendency for nodes to have many direct connections), trading off against the resultant dilution of the flows. In other words, the more heavily the network can concentrate flow, the more the network flows on the hub to hub links can benefit from economies of scale. A model might employ a concave cost function to favor the assembly of flows. Such a formulation, apart from being a more realistic representation of the costs of flow, produces some useful mathematical properties by providing an incentive to agglomerate transportation infrastructure.

Optimization and Solutions

The strategic optimization problem for hub and spoke networks involves a host of interrelated decisions: the location of the facilities; the assignment or allocation of the demand nodes to these facilities; the modes to be used on the links (e.g., truck vs. air); and the protocols for routing the flows. Such high level strategic decisions are sufficiently complex that they are often broken down into simple components. For major sorting centers, the 'location decision' is probably the most critical because of the high cost of fixed infrastructure; such decisions also occur rather infrequently in a company's life cycle. Once the system location is selected, however, there remain a large number of operational or routine daily decisions: handling exceptional conditions, rerouting flows around broken links, and even substitution of backup coverage. Eventually hub systems may outgrow their initial location configuration and either choose to close facilities, or to designate satellite minihubs to skim off some of the regional interactions. Many aspects of these decisions depend on the underlying geographical regionalization, as hubs tend to serve particular clusters of interactions. Scale may also be seen to play a role, in that hub like phenomena occur in global, regional, and local configurations. One way to see this is to note the special centrality of Anchorage, Alaska, in terms of connections between US, Europe, and Asia. Dubai also has notable accessibility advantages in terms of its ability to serve as a base for nonstop connections to virtually all parts of the globe.

A single hub can be optimally located by using a variant on a single facility median; the weights used for the problem in that case are simply the sums of inflows and outflows to each of the basic nodes. The problem becomes more complex once there are two hubs to be located, because the nodes need to be divided into service clusters. The two hub system, nevertheless, still has 'polynomial time' algorithms. However, Sohn and Park show that the single allocation problem is NP hard as soon as the number of hubs is three. The general location problem of p hubs may be set up in planar and discrete versions, in common with the general classification of location modeling tasks.

The full suite of modern computational techniques has been directed to many aspects of the hub location program and its variants. All standard techniques have found useful application and are frequently the topic of new research papers in operations research; the following example titles are indicative: genetic algorithms, tabu search, Lagrangian relaxation, neural networks, Benders decomposition, and finally, new mixed integer formulations based on shortest paths. The fundamental problem in location models (whether hubs or regular medians) is obviously that there is an indivisibility of the decisions (a hub is either located or not, requiring a one–zero decision variable). Because of the peculiarities of the hub location objective, one can find that a nonconvex objective function might lead to fractional facilities being located in the absence of special care to enforce in divisibility (this is related to a fundamental, nonconvexity result in optimization devised by Koopmans and Beck man). Much of the challenge in hub facility location is that of incorporating both indivisible facilities, in the presence of a nonconvex objective function.

The work of Klincewicz, in particular, should be mentioned as it exploits the structure and draws insights from the plant location literature and the network design problem. The analytical approach is to use various ex change heuristics; these work with an existing set of hubs and systematically substitute other nodes for the in cumbents based on local improvement measures (so called dual ascent). Again, such a strategy is common in the broad location literature where it might be seen as an example of an exchange heuristic, but the particularities for the hub model require some careful adaptation, in part because the models tend to produce very large in teger programs. As well as insights arising from plant location and so called dual ascent, one can also see that the complexity of the hub models calls for improved linear programming relaxations (in other words, tem porarily dropping some particular integrality constraints, solving the relaxed problem, and then adding back carefully chosen conditions to logically isolate the best solution). In cases with fractional solutions, integrality was achieved by adding a small set of integrality con straints. The best practices point to either tabu search or specially structured models that exploit insights to im bedded routing or assignment problems. Promising work exploiting polyhedral results have also begun to appear.

Hub locations have high relevance for the provision of infrastructure and network design such as might occur in telecommunications, freight, logistics, and air passenger systems. They are also seen as a variant on a classic transshipment node, and these nodes have special rele vance for supply chain management. The problem also has applications in the design of postal delivery networks, particularly in the location of capacitated mail sorting and distribution centers. These will be discussed further below.

Telecommunications Applications

Many of the variants on network design problem occur in the telecommunications field. Space does not permit more than a provisional listing of applications, but it is clear that hub and spoke networks play a key role in the decision support for planning telecommunication net works. The assignment of cells to switches in cellular networks also opens a host of network configuration, layout, and connection problems, many of which have the same important ingredients of ‘interdependence’ alluded to in the introduction to hub systems. One typical design in a two level network has a fully meshed backbone subnet, and each of the nodes is connected to two dif ferent backbone nodes. Typically, these models describe cost minimization over three types of decisions: locating hub facilities, positioning conduits, and installing com munication lines. The problem is usually formulated as a variant on a network design problem, using the dual ascent solution strategy. The size and complexity of these applications is such that one necessarily needs to deploy heuristics or other special decompositions.

Freight Applications

The application of hub and spoke systems in freight transportation and logistics opens up interesting issues of network design and operational planning, including the extent to which some direct transportation can take place outside the hub network. In other words, freight systems are likely to combine elements from hub and spoke de sign, and also reveal additional side arrangements that might handle some freight ‘offline’ on special multimodal carriers, or involve considerations such as assembling less than truckload flows into standard transportation containers. Geographers, especially those with training in operations research and optimization, have made im portant linkages between classical descriptive models of freight systems (tending to emphasize transshipment and optimal capacity) to develop planning applications of the hub network model to provide efficient multimodal sys tems for regional economies. The location of such transshipment facilities is critical to the ‘competitiveness’ of a regional economy.

Perhaps the best example of this marriage of opti mization and regional transportation organization is in the intermodal coordination involved in distribution and supply chain management. A particularly interesting variant on this problem occurs in less than truckload transportation whereby the collection of many small freight flows occurs through a series of intermediate transshipment locations. Similarly, the recent explosion of demand for timely delivery of materials ordered from re mote retailers has prompted a close coupling of the air and ground systems operated by freight carriers such as UPS and FedEx. The many trade offs between time sensitivity, customers’ price sensitivity, and the potential to make on time deliveries using cheaper modes (truck) have pro duced a major application for optimal network planners.

Applications Air Passenger Systems

The hub and spoke network is much maligned in the sphere of air passenger transportation, as it gives rise to necessary plane changes and stopovers, and the apparent increased possibility of missed connections and mis handled baggage. In fact, these troubles were sufficiently well documented for hub carriers that the air industry took note of the success of regional carriers such as Southwest Air in the US, a system that uses a point to point network. Southwest, compared to the major carriers, achieves higher percentages for on time de partures and has fewer problems with missed connections and lost baggage, leading some to speculate that its net work structure gives it an advantage. Clearly this ad vantage has somewhat of a spatial angle, in that the ability to run a point to point system effectively hinges on the regional foci and specialization that are not pos sible in carriers determined to provide trunk, local, and international service. So while nonhub networks have prospered, hub systems persevere, and attempt to operate more satisfactorily by developing better spacing and 'choreography' between waves of interconnecting flights at major peaks. The option of better connections and the possibility of a convenient stopover at a major retail hub (as some airports have become) exploits the connection step of air passenger travel.

Descriptive work that defines the scope of hubbing in networks examines the choice of particular city con figurations by airlines, the emergence of hubs in sec ondary airports near to main markets, and the potential for air carriers to change their configurations, including closure of hubs. The network chosen has a notable im pact on the ultimate success or failure of the carriers. Research in this area has examined network configuration in the presence of competitive and regulatory forces with an emphasis on the ability of one carrier to gain advan tage, or indeed stymie the efforts of its competitors through macrospatial strategy. Adler uses a game theo retic approach (a two stage Nash best response game), to evaluate the most profitable hub and spoke network for an airline to survive in a de regulated environment. Among the class of related problems, one finds concern with selection of international gateway airports, and se lection of equipment and routing strategies across various scales and lengths of haul. Such problems are increas ingly relevant as the structures from 'developed' air networks grow toward integration of newly emerging air markets. The selection of a major international gateway or indeed a regional hub frequently makes news in na tional press and knowledge of these networks spreads rapidly through internet trade and technology news groups. Related model issues are reviewed in Bryan and O'Kelly. Such hub network ramifications have also spread to the manufacturing sector. One of the major airframe manufacturers (Airbus) has made its fleet offerings mesh well with the high scale flows on interhub connections, recognizing that as global travel expands the interaction between major international gateways is likely to require higher capacity long haul jets. The increasing set of internationally identified major hubs is likely to reinforce these choices.

Extensions

Significant variants on the standard model include the hub network design problem with stopovers and feeders. In the network used by the Federal Express Co., some flights to and from the hub make one or more stopovers, and many smaller cities are served by feeder aircraft which connect to other nonhub cities. The paper by Kuby and Gray explores the trade offs and savings involved with stopovers and feeders, and develops a mixed integer program to design the least cost, single hub air network for a fixed hub location.

Variants on the basic model are evident in the active research literature, including consideration of virtually the entire gamut of design decisions that arise in the hub system: macrolocation and structural organization (a high level strategic decision); the division of the network into levels; the operational or daily optimization of the network (operational decision); and of course, the methodological angles that arise as extensions of the hub concept to other nontransport environs. Examples in clude: the routing protocols (e.g., how many stops are allowed?); topological design of a hierarchical two level network; capacity, thresholds, and flow dependent costs; cost allocation to impute the cost of a shared hub to hub journey to the original origin to destination path; hub and cluster assignment models; and, last but not least, the impact of queues on the effective handling of such large systems of flows.

Finally, no review of the hub problem would be complete without mention of the myriad ways in which 'hub' like terminology pervades other transport sectors. Hubs are used as a means of contrasting the typically major airline carrier's network with a more flexible point to point network that has performed so well for those niche carriers exploiting gaps in hub network coverage. Precisely because there are fewer intermediate inter change points, a point to point carrier such as Southwest Airlines in the US is capable of keeping its fleet in the air for a higher proportion of the working day and typically has better on time and baggage handling results than those airlines that use an interchange at a busy hub. (Denver, for example, is home to a much criticized ex periment in highly automated baggage handling necessi tated in part by the major hub interchange traffic at that location.) The hub idea is a major point of national pride in the organization of airline systems and frequently the importance of the primary hub in a system is such that it becomes a rallying point for both political and economic development causes. The hub is also the basis for the idea that there are certain locational 'centrality' benefits that arise in the organization of networks. These become places that gain insurmountable advantages in some air line networks (the 'too big to fail' argument that was once heard in the context of manufacturing industry).

Periodically however, these arrangements are destabilized by the acquisition or merger of networks, and it is a measure of the importance of hubs that certain mergers or acquisition are deemed more or less likely to succeed because of the complementarities of their hub network systems.

See also: Location Analysis; Network Analysis; Spatial Interaction Models.

Further Reading

Adler, N. (2001). Competition in a deregulated air transportation market. *European Journal of Operational Research* 129(2), 337 345.

Bryan, D. L. and O'Kelly, M. E. (1999). Hub and spoke networks in air transportation: An analytical review. *Journal of Regional Science* 39(2), 275 295.

Campbell, J. F. (1996). Hub location and the p hub median problem. *Operations Research* 44(6), 923 935.

Ernst, A. T. and Krishnamoorthy, M. (1998). Exact and heuristic algorithms for the uncapacitated multiple allocation p hub median problem. *European Journal of Operational Research* 104(1), 100 112.

Fortz, B. and Labbe, M. (2002). Polyhedral results for two connected networks with bounded rings. *Mathematical Programming* 93(1), 27 54.

Groothedde, B. (2005). *Collaborative Logistics and Transportation Networks: A Modeling Approach to Hub Network Design*. The Netherlands: TRAIL Research School.

Klincewicz, J. G. (1991). Heuristics for the p hub location problem. *European Journal of Operational Research* 53(1), 25 37.

Klincewicz, J. G. (1998). Hub location in backbone/tributary network design: A review. *Location Science* 6(1), 307 335.

Koopmans, T. C. and Beckmann, M. (1957). Assignment problems and the location of economic activities. *Econometrica* 25, 53 76.

Kuby, M. J. and Gray, R. (1993). The hub network design problem with stopovers and feeders the case of Federal Express. *Transportation Research Part A Policy and Practice* 27(1), 1 12.

O'Kelly, M. E. (1987). A quadratic integer program for the location of interacting hub facilities. *European Journal of Operational Research* 32, 393 404.

O'Kelly, M. E. (1998). A geographer's analysis of hub and spoke networks. *Journal of Transport Geography* 6(3), 171 186.

O'Kelly, M. E. and Bryan, D. (1998). Hub location with flow economies of scale. *Transportation Research Part B Methodological* 32(8), 605 616.

O'Kelly, M. E. and Miller, H. (1994). The hub network design problem: A review and synthesis. *Journal of Transport Geography* 2(1), 31 4034.

Reynolds Feighan, A. (2001). Traffic distribution in low cost and full service carrier networks in the US air transportation market. *Journal of Air Transport Management* 7(5), 265 275.

Skorin Kapov, D. (2001). On cost allocation in hub like networks. *Annals of Operations Research* 106, 63 78.

Skorin Kapov, D. and Skorin Kapov, J. (1994). On tabu search for the location of interacting hub facilities. *European Journal of Operational Research* 73(3), 501 509.

Sohn, J. and Park, S. (2000). The single allocation problem in the interacting three hub network. *Networks* 35(1), 17 25.

Human Geography

C. Gibson, University of Wollongong, Wollongong, NSW, Australia

Human geography is a major subdiscipline within the wider subject field of geography. Traditionally, geography is considered the study of the Earth's environments and peoples, and the interactions between them. 'Geography' comes from ancient Greek origins (Eratosthenes was the first to use it), literally translating as 'to write or describe the world'. In classical and Enlightenment geography, humans and the 'natural' world were usually described in conjunction, often in a regional fashion, as Europeans encountered unfamiliar places in exploration and empire. Since the late nineteenth century, this conjoint understanding of geography – as describing the natural and human world, region by region – has gradually been augmented by more precise subdisciplinary pursuits and identities. The most basic of these describes geography as consisting two fundamental halves: physical and human geography. Physical geography generally means the science of the Earth's surface, while human geography usually refers to the study of its peoples, and geographical interpretations of economies, cultural identities, political territories, and societies. Physical geographers classify and analyze landforms and ecosystems, explain hydrological, geomorphological, and coastal processes, and examine problems such as erosion, pollution, and climatic variability. Human geographers analyze population trends, theorize social and cultural change, interpret geopolitical conflict, and seek to explain the geography of human economic activities around the world. How exactly this division of labor came to be is a most pivotal story of contemporary geography. It is a story about twentieth century scientific fragmentation, and about different theories on the status of humans vis à vis nonhuman nature. It is also a more slippery and difficult story about how academic knowledges are produced, mutate, and travel (and how this happens in particular places), how knowledges find popularity, fade away, or are challenged in time and across space. The central division of labor in geography – produced by these means over more than a hundred years – has established and defined the space within which most human geographical practice now occurs. This article is an introductory overview of contemporary human geography and the stories underlying it.

An outsider could be forgiven for thinking that human geography was the study of the existence and distribution of humans on Earth – of *Homo sapiens* as a distinct species. Literally, human geography could be interpreted as the study of the geography of humans: when, where, and how humans evolved, developed strategies for survival, and dispersed to other parts of the world. Some within geography would indeed consider such themes important to geography. They invite analysis of how humans inhabited and related to physical environments, how humans used (and abused) resources, adapted to different climates, and developed distinct regional cultures. When prominent geographer Halfred Mackinder presented his ideas on the scope and purpose of geography to the Royal Geographical Society in 1887, this idea of geography as "trac[ing] the interaction of man [sic] in society and so much of his environment as varies locally" proved immensely popular and would shape future interpretations. However, for most human geographers in the current era such questions have fallen from favor because of ethical and intellectual concerns (see discussion below), while more contemporaneous themes have grown popular, firmly on the human side of geography. The science of human evolution and geographical dispersal has instead become the purview of other disciplines such as archeology, paleontology, and anthropology.

Most contemporary academic geographers hold some nominal allegiance to either of geography's 'halves' (human or physical), although for important intellectual and philosophical reasons (discussed below) some do resist this division and instead prefer to regard geography as a disciplinary whole, or insist on troubling the conceptual distinction between 'the human' and 'nonhuman' parts of the world. Some commentators have criticized a perceived widening of the gulf between geography's two halves. Others see human geography as merely a convenient badge for its diverse contents, while still adhering to the principle of a wider, umbrella discipline of geography (including physical geography). Regardless, contemporary human geography has taken on a particular character. Human geographers tend to explore social, economic, cultural, political, and demographic dimensions of human existence, and situate analysis in geographical space (conceptualized across and between scales from the body to the city, nation, and globe). While diversity defines contemporary human geography, there are common questions of geographical scale, causality, agency and structure, interrelationships and networks, place and movements. Human geographers are concerned with observed distributions and analytical explanation. They invariably focus on the spatial and, whether implicit or explicit, have a great deal to say about the moral and political dimensions of human activity.

This article is structured in anticipation of the more pertinent questions an outsider might ask of a human geographer: 'what is human geography?'; 'when did human geography emerge as a discipline and how has it changed over time?'; 'what is the character of contemporary human geographical practice?'; and 'how is it practiced in different places?'. At the outset it is worth clarifying the approach taken here. Reflecting current thinking, a single, linear story about canonical development of ideas in human geography is sought to be avoided. Such stories only reinforce the dominant narrative at the expense of less powerful ones, and are invariably Eurocentric. This article also tries, where possible, to show that geographical knowledges are produced by a range of actors in different ways and for different audiences.

However, the manner in which such knowledges are produced, and what those knowledges mean, is contingent on the degree of formal academic practice, on the political, institutional, and cultural circumstances whence they came, and on the perspective of the onlooker or interpreter. Certain actors and institutions benefited (sometimes enormously) from the production of particular kinds of human geographical knowledges, while others suffered. In this article, the author hopes to give some sense of the 'effect' of human geographical knowledges for people, institutions, and places.

In short, this article tries to situate human geography in the times and places in which it has been produced. The author is cognizant that this is one academic's perspective. This article seeks not to define core agendas, but to provide explanation and context. It necessarily cannot do adequate justice to all subthemes, topics, and perspectives. Part of the problem is that most of the disciplinary histories written in English have come from Britain and America. This has subsequently given the impression of an Anglo American axis of academics 'creating' the history of the discipline – a history to which others from elsewhere occasionally contribute, but not to shape in any fundamental way. Alas, probably more than a whiff of such a bias seeps into this story too. There are attempts below to decenter the sense of an Anglo American dominance, but finding and fully surveying the mass of non Anglo American explanations of human geography is beyond my capacity. As Sidaway (1997: 74) argued, "no single author could ever do justice to all traditions of geography, everywhere in the world." Human geography exists as a formal academic discipline in tens, if not over a hundred, different countries and languages, with diverse meanings and contexts. Only a fraction of those stories have been properly documented. Instead, the reader is recommended to treat this article as an entry point, and to explore the subdiscipline's various trails from it. The author hopes it will become apparent as the reader progresses that the development and diffusion of human geography has been dynamic and context dependent – and amidst significant advances, has often been far from straightforward.

What Is 'Human Geography'?

Defining human geography is especially difficult because of complicating factors like the relationship between human geography and geography (the former to many considered a mere subdiscipline of the latter); the rather late professionalization of the discipline; variations in human geography written in different languages; and the difficulty (indeed, sheer dubiousness) of being able to identify definitive research questions, sequential paradigms, or key thinkers. It is tempting to define a common ground for human geography's intellectual core (as Hartshorne attempted), and wish to enforce this. Such a common ground might provide human geography with a sense of unity. But the reality of how human geography is practiced simply cannot sustain this. As David Livingstone so powerfully put it in *The Geographical Tradition* (1992: 28), "The idea that there is some eternal metaphysical core to geography independent of historical circumstances will simply have to go."

Indeed, human geography has not had a linear trajectory of intellectual advances accompanied by more or less parallel diffusions as the subject was established and pursued around the world. Human geography has been energized and replenished in a more decentered manner by multiple authors asking very different questions within the same subdisciplinary space. Some, like Alexander Von Humboldt, are seen as forbearers, but others, such as Friedrich Ratzel and Ellsworth Huntingdon are remembered more for how subsequent geographers disowned their ideas. Somewhat an exception to this is Paul Vidal de la Blache, who took up the chair of geography at the Sorbonne in 1898 and whose students became professors of geography at most of France's (then) 16 other universities. In this instance one could argue that a single leading figure established the terrain and direction of a national geographical tradition – at least for several subsequent generations.

Of course there have been highly influential theorists and practitioners in every era, but human geography has been a remarkably open field, particularly since the 1960s, when especially diverse ideas and political practices found traction, and when links were made to anthropology, sociology, cultural studies, economics, and philosophy. This perspective helps explain how contemporary human geography has emerged as a complex and differentiated grouping of scholars and students, institutions and analytical sites, technologies and methodologies, journals and conferences, arguments and audiences – evidenced by the very diversity of entries covered in this encyclopedia.

For what it is worth, and before discussing some rather more entangled stories about the discipline, it is worth briefly surveying current human geography – at least in a surface scan of its various themes as they are represented in English language academic publishing. **Table 1** summarizes the contents of one leading journal in the field, *Progress in Human Geography*, sorted by the themes of human geography written within its pages. *Progress* was chosen for this exercise not because it was deemed to be superior to other journals – it is not a representative sample because *Progress* has been Anglo American dominated – but because above all other journals its aims most suited this exercise: to 'offer a genuinely comprehensive survey of geographical studies', covering all human geography (rather than just parts of it), with regional reviews and commissioned progress reports on areas of research concentration. Unlike older journals (often published by national geography associations) it is exclusively centered on human geography (with its sister publication, *Progress in Physical Geography*, covering the other 'half' of the discipline). Every article between 1978 and 2007 has been counted in **Table 1**. What it reveals is that contemporary human geography is indeed diverse. Some subdisciplines have had a constant presence (e.g.,

Table 1 Contents of *Progress in Human Geography*, 1978 2007; numbers of articles by subdisciplinary theme[a]

	1978 1982	*1983 1987*	*1988 1992*	*1993 1997*	*1998 2002*	*2003 2007*[b]	*Total*
Economic geography	10	14	16	16	26	22	104
Social geography	9	10	10	15	18	22	84
Cultural geography	6	8	9	18	18	25	84
Spatial science	19	12	17	6	11	16	81
Political geography	4	11	13	8	15	19	70
Philosophy and the practice of human geography	5	2	16	16	15	15	69
Urban geography	15	13	6	4	10	9	57
Regional geography	16	18	9	2	5	6	56
Historical geography	7	11	11	6	6	13	54
Population geography	12	8	10	8	4	5	47
Environmentalism	12	10	7	3	5	0	37
Rural geography	10	5	1	4	6	7	33
Development geography	6	4	9	3	4	3	29
Human/environment relations	7	4	3	2	6	6	28
Agriculture and food	4	7	4	2	3	8	28
Statistical techniques	6	6	3	2	5	4	26
Industrial geography	6	4	4	4	4	2	24
Behavioral geography	10	3	5	2	2	0	22
Health geography	3	2	7	3	4	3	22
Feminist geography	1	0	2	6	6	1	16
Moral and ethical geographies	0	0	1	2	5	7	15
Methodology in geography	0	0	0	1	6	8	15
Transport geography	8	1	1	0	1	1	12
Indigenous and postcolonial geographies	1	2	0	0	2	2	7
Tourism geographies	0	0	1	2	1	0	4
Cultural economy	0	0	1	0	0	3	4
Humanism	0	2	1	0	0	0	3

[a] Commonly accepted monikers were used for themes (e.g., economic, urban). More were added when articles did not seem to be adequately covered otherwise (e.g., health geographies). A subtree of themes was also used. For example, Marxist geographies were included in economic geography. Some theoretical perspectives that human geographers would consider a distinct theme were included, even though strictly they are a particular perspective that could apply more broadly across subjects (e.g., feminist geography). Others were not considered distinct themes because their theories, though important, are not normally associated with separate subdisciplinary identities (e.g., post structuralism, actor network theory). In these cases articles were classified according to the theme of the case study or other subdisciplinary identity (e.g., post structuralist geopolitics within political geography; actor network interpretation of cities in urban geography). Major themes such as geographical scale, GIS, and globalization were also subsumed within categories to aid in reducing the number of categories (in those instances, articles on scale were classified as 'philosophy and practice of geography'; GIS as spatial science, and globalization as economic or cultural (or other), depending on subject matter and theoretical perspective). Where articles clearly covered two listed themes, they were counted twice. Therefore totals do not exactly equate the total numbers of articles actually published in *PiHG* from 1978 2007. Articles on the history of geography were counted in historical geography unless their abstract overtly mentioned philosophy and geographical practice (in which case the article would be included in both categories). Articles on sexuality and gender identities were counted in cultural geography, unless feminist theories were discussed in the text (in which case they were counted in both cultural and feminist geographies). Articles on statistical methods were counted separately from those providing progress reports on methodologies in geography. Some may quibble at the chosen categories (any act of classification involves at least some arbitrariness), but the author proceeded in using these estimations regardless, for they illustrate at least reasonably well an overall sense of English language geographical publishing in the past three decades.

[b] At the time of writing, only two issues for 2007 had been published, and were thus used in these statistics.

urban, economic), some have seen declining number of articles published since the 1990s (e.g., regional geog raphy), while others came to prominence (cultural). Some have ebbed and flowed (e.g., rural) and others have remained constant even though according to some commentators and historians of the discipline their fall from theoretical favor had been presumed complete (e.g., spatial science, statistical techniques). Meanwhile, a more self reflexive element is evident especially since the mid 1990s, focused on the philosophy and practice of human geography, ethics, methods, and links to nature.

When Did Human Geography Emerge as a Discipline?

Geography's 'early modern period' is thought to have begun with the Renaissance in the 1600s and ended with the formalization of geography as a school and university discipline in the late nineteenth century. In this period, "geography was 'precisely defined' as a scale of inquiry, that of the earth as a whole" (Withers, 2006: 712), which distinguished it from cosmography at the larger scale of the universe, and topography and chorography which dealt with regional and national scales. Geographers dealt with locations upon the Earth and described the phenomena to be found in those locations. Human geography was a constituent part of this, rather than a discrete endeavor. Geography was "a coherent body of knowledge about a clearly defined object, namely the situation of places on the earth and the content of those places in natural and human terms" (Mayhew, 1998: 391).

That human geography might be something separate from or a distinct part of geography is a recent conceit. Most professional 'human geographers' work in uni versity geography departments that cover the whole discipline, and most would probably pledge some alle giance to a conception of geography as an entirety that needs both human and physical parts. This nominal unity explains why in many forums – from journal publishing to conferences – the discipline of geography as a whole has been sustained. On the other hand, there is much evidence that human and physical geography are now effectively separate. Both have specialist journals (with subdisciplinary aims); definitive publications that suggest whole of discipline integrity (like this one); and confer ences. Joint teaching programs across human and phys ical geography are more common than genuinely collaborative research projects.

Tracing when and how human geography became recognized as distinct is no easy task. One can trace the term itself, its use, and the manner in which 'human' came to preface 'geography'. According to Livingstone (1992: 198), the term owes its existence to German evolutionary geographer, Friedrich Ratzel, who in *Anthropogeographie* "sought to lay out the conceptual foundations of a new discipline – human geography." In Britain, The Royal Geographical Society was established in 1831, and started *The Geographical Journal* in 1893, but it was not until 1909 that the first article appeared with 'human geography' in its title – a review essay by H. J. Fleure, of Elisée Réclus' six volume *L'Homme et La Terre*. In French geography the term first appeared in the title of Paul Vidal de la Blache's *Principes de géographie humaine*. The discipline defining journal *Progress in Human Geog raphy* was itself only established in 1977.

Standard geography textbooks used in the English speaking world at the turn of the twentieth century did not use the term at all. In 1893, Hughes divided *A Class Book of Modern Geography* into 'mathematical or astronomical geography', 'physical geography', and 'political and com mercial geography'. The latter category most neatly fits today's human geography – although Hughes' further description of it, as that part of geography "that treats of the political divisions of the earth, and the condition and industrial pursuits of mankind [sic]" is hardly com prehensive. Similarly, Lionel Lyde, in his 1895 textbook *Man on the Earth: A Course in Geography*, dealt separately with 'population', 'commercial', 'political', and 'historical' geography – without using the term 'human geography'.

In contrast, the formal use of 'physical geography' can be traced back to examples such as Mary Somerville's *Physical Geography*, Guyot's *The Earth and Man: Lectures on Comparative Physical Geography*, and Emmanuel Kant's *Lec tures in Physical Geography* (1756–96) – although none of these fits perfectly with contemporary meanings either: for Somerville humans were considered part of physical geography; for Guyot (who was heavily influenced by the theological geography of German Carl Ritter), "the whole universe is a thought of God"; and in the case of Kant, not only were humans included, but physical geography was considered the foundation "for 'other geographies' in cluding political, commercial, moral (in the sense of mores), and theological geography" (Hartshorne, 1939: 36).

In English, 'human geography' only really started to be used in articles and textbooks in the 1920s and 1930s, usually when describing environmental determinist anthropogeography (after Ratzel), or when used as a container term for related debates about nature/human influence and causality. The term did not generally apply to political, historical, or commercial (economic) geog raphy. Human geography as an umbrella term for the half of the discipline that dealt with human affairs was not commonly used until well into the 1950s and 1960s.

'The Human' in 'Human Geography'

At this point – and in lieu of a neat story of disciplinary beginnings – it helps to consider how the separation of geography into its 'human' and 'physical' parts reflected a particularly European intellectual heritage – with Biblical origins but gaining pace in the Enlightenment – of separating out concepts of 'the human' and 'nonhuman nature'. While 'human geography' has its particular etymology (spanning a century at best), the intellectual conditions from which separate agendas for 'human' and 'physical' geography could manifest were already apparent – and had been building consistently because of the influence of key thinkers both inside and outside geography – well before the twentieth century.

As Anderson argued, in the eighteenth and nineteenth centuries European intellectuals struggled to come to terms with what it meant to be 'human' in contrast to that part of the world deemed 'nonhuman'. Geographers were important public intellectuals developing theories on this problem. Geographers variously argued that humans were distinct and unique from nonhuman nature; conversely, that humans were an innate part of nature, subject to its universal laws; that humans could be classified into different types (e.g., into 'races') and their 'humanity' ranked (as 'less human' savages, or 'children' in a stage of civilizing development); and that nonhuman nature (including the distribution of soil, landform, climate, and geological types) influenced the geography and manner of human social development.

Before Charles Darwin's *Origin of the Species*, teleological theories dominated. These argued for some kind of grand design at work (usually at the hand of God), or emphasized that nature and humans changed in stages or movements toward preordained destinies. 'Stadial' theories, typically influenced by Locke's famous theory of ascent from a 'state of nature', saw humans as "shar[ing] a potential to lift themselves above other life forms which – so it was assumed – *merely lived*" (Anderson, 2005: 273). For premodern geographers, "imagined humanity sat on the scale of human variation… savagery was something that could be and, on the continent of America, had been surpassed. Only the circumstances of people's mode of living, adapted to specific environments of soil and climate, held them back" (Anderson, 2005: 273). Europeans saw themselves as having successfully arisen from nature, as civilized and rational. Encounters of explorers and geographers on colonial voyages seemed to confirm that other humans were 'savage', 'closer to nature', or in need of civilizing through missionary Christianization.

Darwin's theories of natural selection and evolution would shape the intellectual climate of all the sciences, including geography. So called 'social Darwinists' sought to directly apply his laws of evolution and 'survival of the fittest' to human populations, attempting to more thoroughly theorize the relationship between humans and the environment, rather than just catalog and describe places. It is worth noting that many so called 'social Darwinist' theories were more accurately neo Lamarckian, because they drew inspiration from Jean Baptiste Lamarck's earlier evolutionary theories. These emphasized more speedy, direct evolutionary adaptations by successive generations, and accounted for distinctions in a population "attributable to environment and will in mutual cooperation rather than to the vagaries of some capricious variation" (Livingstone, 1992: 187–189). Whatever the source of inspiration, social Darwinists commonly sought to explain human traits and behaviors as a result of environmental variation. In an illustrative example, in his 1923 address to the AAG, William Morris Davis – a major American proponent – showered warm praise on studies of the Seri Indians, which revealed "the many ways in which physiographic factors have influenced the human inhabitants of the region… by reason of the dry climate, the open surface, the small food supply and the scanty population, the Seri have become exceptionally fleet and enduring as runners; they can capture deer by pursuing them on foot and exhausting them; one of the tribe overtook a running horse." Such ideas were not entirely new – both Hippocrates and Aristotle had philosophized on the links between climate, habitability, and characteristics of people, and before Darwin, Montesquieu's ideas of how climate governed cultural traits and the 'degeneration' of humans were highly influential. But in this era, such ideas found a renewed vigor, frequently resulting in presumptuous and judgmental typecasting of racial types. Many geographers bought into it, particularly in English speaking geography: notably James Bryce, Ellsworth Huntingdon, and Griffith Taylor. Fantastical leaps of (il)logic were common. As late as 1931 Austin Miller argued in *Climatology* that "The enervating monotonous climates of much of the tropical zone, together with the abundant and easily obtained food supply, produce a lazy and indolent people, indisposed to labour for hire and therefore in the past subjected to coercion culminating in slavery." Huntingdon and Taylor both made maps of climatic and racial distributions and sought to schematize races of peoples further along spectrums of civilization and mental development: tropical climates, for instance, produced 'inertias' that eroded the will to work and degraded morals. In craniology the skull sizes of humans from the different races were measured, and a more fully 'innatist' – and thoroughly racist – view of the geography of humanity was proffered, based on the "preposterous contrivance that the world's people could actually be differentiated and calibrated on a scale of distance' from nature" (Anderson, 2005: 276), with white, temperate Europe at its center, and the other continents further from it (and thus closer to the tropics, and to nature).

Darwin's ideas of evolutionary survival would be appropriated and applied to theories about nations and territorial expansion, in turn informing statecraft and practical foreign policy. In Mackinder's 1904 famous geopolitical treatise, 'The geographical pivot of history', one zone in Eastern Europe – the Heartland – was seen to occupy a location key to ultimate global territorial control. The national or racial group that emerged from the struggle for survival to control that space thence controlled Europe, the wider landmasses connected to it, and effectively then the whole world. In *Politische Geographie*, Ratzel would argue that as populations grow, states, like biological organisms, must expand their territorial interests to create *lebensraum*, or 'living space', in which to endure. He provided a convenient rationale for European colonial expansion (then well underway in Africa) as a quest for living space and, literally, an outcome of the struggle of peoples to survive. Moreover, the ideal aim of states was to progress toward *grossraum* (large space) as a reflection of higher levels of civility. Here was, according to Livingstone (1992: 201), "a naturalistic theodicy that justified the imperial order in the language of scientific geography."

In hindsight it is easy to argue, as Johnston (1987: 36) did, that "the environmental influences adduced were gross… it is hard to believe that they could have been written and taken seriously." Certainly there were critics and reactions. The arguments of 'possibilist' authors such as Vidal de la Blache gave agency to individual humans to select from a range of uses of the environment. In another of the more well known reactions to environmental determinism, Carl Sauer, at the University of California, Berkeley, sought to reinstate humans as active agents in the production of landscapes. In his *Morphology of Landscape*, it was humans "who possessed the uniquely controlling capacity to modify landscapes for better and worse" (Anderson, 2005: 269), as 'world altering beings' impacting on pre human, natural places. The subsequent 'Berkeley tradition' of cultural geography, led by Sauer and his disciples, became more interested in the diffusion of culture traits (rather than evolutionist generalizations of their origins), the identification of culture regions (in both material and nonmaterial terms), and the role of culture in conditioning human perceptions of nature and the environment. Sauer was influenced by the German Otto Schlüter, who distinguished natural landscape (*naturlandschaft*) from cultural landscape (*kulturlandschaft*); a distinction underpinning the conceptual separation of human and nonhuman nature, and by extension the formal interests of human and physical geography.

In some parts, debates about environmental determinism lingered on into the mid twentieth century. Griffith Taylor scoffed at suggestions that the Australian landmass could support white populations in the hundreds of millions because of unsuitable climate, lack of rainfall, and productive soil capacity. Against the grain of the then white only immigration policy, he promoted Chinese immigration – because according to him they were of the brachycephalic race and thus 'more suited' to the prevailing environment. So much was the Australian government horrified by Taylor's ideas that he was denied promotion, his texts were banned by education authorities and university senates (because of his pessimistic use of words like 'arid' and 'desert' rather than any racist assumptions), and he eventually left in 1928 to take up the chair in geography at the University of Chicago.

By and large, geographers had by the 1940s felt a loss of face from their detour into debates about human–environment relations. Such efforts had been largely discredited, or in the case of Taylor, were so against imperial and nationalist sentiments as to fall from institutional favor. That which had defined 'human geography' in the first decades of the century – questions of the causal relationship between humans and nonhuman nature – faded into obscurity. Human geography would 'let go' of nature from its theoretical purview, leaving behind an era that harbored inherent racism, excused colonial violence, lacked theoretical muscle, and simply suffered the inaccuracy of overgeneralization.

And so, human geography in the second half of the twentieth century became further defined against nonhuman nature, which physical geographers would come to examine using an increasingly scientific positivist paradigm (quickly developing in parallel). Further separation of human and physical geography was perhaps inevitable. Advancement of the discipline was thought to come from segmentation into fractions that could be manageably examined and theorized to more sustainable and defendable ends.

Although enabling subsequent generations to evade the dangers of environmental determinism, post war human geography did not overturn the human/nonhuman dualism on which environmental determinism so firmly rested. It merely shifted focus, to concern itself with just one side of that dualism. Indeed, in retrospect it has only become 'common sense' to talk about human geography as different to 'physical geography' because geographers have come to accept that division. This relatively recent separation has been both useful and unhelpful. On the one hand it established intellectual grounds within geography where engagements with nonhuman nature were merely optional – in turn enabling human geographers to lead theoretical debate in the wider social sciences (without needing to somehow 'bridge' back to the physical side). On the other hand, many would argue that a wedge was driven between great strands of knowledge, artificially dividing the human and nonhuman parts of our world. For some, this divide created practical, institutional, and political risks. With common purpose, human and physical geographers would be otherwise more strategically positioned to contribute to debates about sustainability, inequality, and our ecological future. For others the divide was analytically poor, because it left assumptions about the distinctions between humans and nonhuman nature in tact and untested: “Geography, like history, becomes the story of exclusively human activity and invention played out over, and through, an inert bedrock of matter and objects made up of everything else” (Whatmore, 1999: 4).

Only in the last decade or so have geographers sought to properly question the very category of ‘the human’ on which ‘human geography’ rests. This has seen inward flows of influence from gender and science and technology studies, and new theories in which human and nonhuman entities are all seen as active agents in wider technosocial assemblages. The point is to emphasize that

> The ‘humanness’ of human beings is not pre given in any absolute contrast with animals… The species specificity of people is not disputed… there is no shrill naturalism at work here which wants to remind us, after Darwin, that essentially ‘we are all animals’… Instead, post humanism’s point of departure… is to refuse the humanist account of living things which *begins* by assuming a radical or pure break between humanity and animality. (Anderson, 2005: 271)

In this way – at least for some – the very distinction that the identity of human geography rests upon has been brought into fundamental question.

Modern Geography, Human Geographies

Between 1950 and 2000, human geography became more diverse, and spread across university systems and languages. It also more wholly engaged with theoretical and epistemological turns across the humanities, sciences, and social sciences. There are different ways to portray this development and diversification. One explanation is that growth and further specialization merely continued to unfurl as ever more themes, subjects, and places were incorporated into geographical analysis. Further parts were simply added as subdisciplinary specialisms developed and became normalized. This rings true for certain places and subdisciplines: urban geography, rural geography, and population geography became fully coalesced specialisms as critical masses of researchers and teachers were reached, as university subjects were proposed and approved, and as academic journals gained publishers and readerships.

This doesn’t however reflect the way other subdisciplinary areas emerged from political impetuses, or in reaction to the perceived intellectual poverty of existing orthodoxies. This was certainly how the so called ‘quantitative revolution’ transformed academic geography in the 1960s. According to Allen Scott, it was

> a period of great intellectual and professional struggle in geography between traditionalists and reformers, with the latter seeking to push geography out of its perceived idiographic torpor and on the basis of quantitative methodologies and formal modelling into a more forthright engagement with theoretical ideas. (2004: 481)

While pre war geographers speculated about environmental determinism, or retreated to regionalism and areal differentiation, other competing academic disciplines had proceeded ahead with conceptual and statistical advances. This triggered reactions in a new generation of geographers with a shared desire to develop a thorough, mathematically justifiable geographical science that could be taken seriously by other disciplines. Geographers created the field of spatial science in the 1960s – with new journals and readerships – attempting to develop general theories of space and human activity using inferential statistical techniques and abstract modeling. The emphasis was on techniques of locational analysis – often inspired by other disciplines such as physics, mathematics, economics, and sociology – but also on developing more defendable theories. Common themes included the geography of land rent in cities; geographical diffusion over time; networks (particularly transport); and the locational decision making behavior and distribution of economic activities.

In turn, radical geography would emerge from the 1970s when geographers would become dissatisfied with quantitative geography, which was perceived as socially irrelevant and apolitical. Quantitative geographers appeared to have much to say about how human economic and social activities were structured across hypothetical, abstract space, but were silent on the worrying problems of poverty, war, and racism afflicting the world at that time. Indeed, radical geographers became concerned that (whether willingly or not) the discipline may in fact contribute these problems rather than provide explanations or solutions. As Harvey explained,

> the radical and Marxist thrust in geography in the late 1960s concentrated on a critique of ideology and practice within the positivism that then reigned supreme. It sought to penetrate the positivist shield and uncover the hidden assumptions and class biases that lurked therein. It increasingly viewed positivism as a manifestation of bourgeois managerial consciousness given over at worst to manipulation and control of people as objects and at best capable of expressing a paternalistic benevolence. It attacked the role of geographers in imperial endeavors, in urban and regional planning procedures designed to facilitate social control and capital accumulation. It called into question the racism, sexism, ethnocentrism, and plain political prejudice communicated in many geographical texts. (Harvey, 1984: 5)

Various liberation movements of the 1960s and 1970s would exert a political influence on the development of new strands of human geography concerned with injustice: the women's movement would inform a new critique of the embedded masculinist and patriarchal power axes in geography (and in society more generally); the civil rights and union movements would inspire young radical social geographers concerned with issues of race, class, exclusion, and inequality in the city; international movements for the recognition of the treatment and rights of indigenous peoples would lay the foundation for new postcolonial geographical critiques of empire and European cartography; and poverty, capitalism, and uneven development were examined in more critical light. Economic geography in particular would be transformed, with new interests in the role of transnational corporations, the politics of trade and foreign direct investment, and international and gendered divisions of labor. New journals emerged (notably *Antipode*, in 1969, and *Gender, Place and Culture*, in 1994) and study groups emerged from within geographical associations to provide more focused arenas for radical discourse. These reactionary voices would have a substantial legacy for contemporary human geography. In addition to the shifting of its topics and themes, they opened up the discipline to political exigency, and to broad critical shifts energizing the social sciences and humanities. In time, postmodern and poststructuralist theoretical perspectives (as well as those from more specific fields such as queer theory, governmentality, science and technology studies, and psychoanalysis) would become incorporated into human geographical praxis and be debated for their merits in advancing meaningful, theoretically robust and relevant research. Engagements with philosophy would take human geography even further into metaphysical and ontological ruminations on the nature of place and space.

Intellectual and political 'turns' were thus vital in replenishing the discipline. However, linguistic and national differences complicated this: even between English and American geography the uptake of new specialisms and influences varied. Lingering legacies of older practitioners were present in one country but not another, as were channels of influence on government and the wider social sciences. The enthusiasm for paradigmatic revolutions from the 1960s onwards also belied historical continuities with earlier approaches. Quantifiers, for instance, did reject description and regional parochialism in a revolutionary manner, but they also reconnected geography with a much earlier legacy of Renaissance mathematical geography. Quantitative geography also secured a more formal role for theory in human geography, which subsequent phases of the discipline would rely upon – even if their theoretical and political influences were different. Certain types of human geography developed consistently without necessarily falling from favor, as in the case of social geography. Similarly, although cultural geography bloomed in the 1980s after the 'postmodern turn' (to the chagrin of some commentators), certain continuities would exist with much earlier traditions (e.g., its focus on landscape). Political geography too would be written for over a hundred years – even though it would have its own internal radical and post structuralist 'turns'. In addition, new areas with their own journals would emerge not from paradigm shifts but in relation to real world cultural and economic changes (e.g., tourism geography), from observations of research gaps (e.g., children's geography), or from the continued needs of the state for certain kinds of geographical knowledge. It is because of this mix of continuities, overturnings and absorptions that it is now possible – perhaps even more appropriate – to talk about human geographies, rather than a single disciplinary corpus. Pluralizing the term human geography is both a linguistic convenience (one which more accurately depicts current practice) and an intellectual imperative for the discipline aimed at celebrating difference.

How Are Human Geographical Knowledges Produced?

Human geography is "an irredeemably situated, positioned system of knowledge" (Gregory, 1994: 76). How knowledge has been produced, distributed, and consumed matters a great deal in how we interpret its impact and meaning. As already mentioned, human geography existed well before it became a formal academic enterprise. In addition to at universities, human geographical knowledges have been produced in a multitude of contexts. "All societies," wrote David Harvey (1984: 2) "possess a distinctive 'geographical lore', a working knowledge of their territory, of the spatial configuration of use values relevant to them." Human geographical knowledges are found on university campuses beyond geography departments (for example, researchers working from history departments on colonial mapping), and where knowledges are produced about people and place in nonacademic, populist ways. Cultural expressions like art, film, music, or literature all in one way or another 'write or describe the world' (as do postcards, television nature documentaries, and tourist brochures) and in that way form legitimate and rich sources of geographical knowledge. In cultural and political geography in particular, popular culture has become an important source of material for analysis of geographical imagery and meaning.

Conceptualizing human geographical knowledges as the manner in which peoples and places have been imagined, written, or spoken about enables one to reveal a plural, diffuse, and ingrained vintage for the discipline. In a very basic way, human geography has been practiced throughout human existence on Earth – or at least as long as humans have made cartographic representations of space and their habitation within them. In Africa rock art depicts topographical features and animals, and people and dwellings are arranged in relation to them; in Australia, aboriginal rock art is part of a living and dynamic history of human geographical representation spanning thousands of years. In Native American art across the southwest, and in varied traditional cultures elsewhere, oral traditions were recorded and transmitted via cartographic techniques, where knowledge of topographical and cultural sites and routes of travel were necessary. Such knowledge was, and still is, at its most fundamental, a human geographical representation of a defined social space.

Before the development of universities, scholarly societies, or printing presses, geographers were employed by royal courts to discover parts of the world and report back on them. Exploration and discovery funded premodern geography. Geographical knowledges dispersed via the commercial trade in maps and globes. This was not a set of activities purely confined to what we now call Europe, indeed it preexisted the age of European exploration. Trade enabled exchanges between 'the West' and various parts of the Ottoman Empire, the Indian Ocean, and Asia that had already been mapped. Additionally, as Sidaway argued, indigenous human geographical knowledges provided the basis for European geographical knowledges, when Europeans, during the Renaissance and later centuries, expanded empires and influence across the globe. Europeans gained advice, traveled over already worn paths and routes, exchanged maps and charts, and built a geographical tradition that hybridized – rather than ignored or replaced – already existing indigenous knowledges.

Prior to the 1400s, written geographical records were hardly scientific – they were often "travel accounts of the weird and wonderful" designed to "titillate readers" (Livingstone, 1992: 34). Writing only became more idealistic and methodical from the late 1400s onwards, after the Renaissance had begun. But the format and popularity of 'travel' geographies persisted. Most people writing geography in the 1700s and 1800s were not 'geographers' in the sense that we understand it, instead being historians, botanists, or 'Grub Street journalists' – the name given to hack writers associated with cheap publications, stemming from the London street of the same name (Mayhew, 1998: 402). These were compilers and collators of others' works, not field scientists, and they published populist books rich with fanciful description that went into numerous successive editions: "Truths were not demonstrated as would be demanded by scientific method; they were simply copied and repeated" (Mayhew, 1998: 405). In a similar manner, for over 400 years a genre of book publishing called 'special geography' sought to describe every country of the world. Written for the US and Britain and most popular from the 1780s to the 1820s, they were prose works rather than encyclopedia, containing secondhand information gleaned from reports of exploration. Their intention was for "pleasure and utility" (Withers, 2006: 713); their readers engaging in "voluntary self enlightenment" (Sitwell, 1993: 9). Not coincidentally, they were most popular at exactly the time that reports of the travels of Cook, Flinders, La Pérouse, and other explorers reached Europe. Consumers not only read books. In Britain, the Royal Geographical Society's lectures were heavily patronized by 'the dining club set', frequently people of military affiliation, for whom geography "centred on regional description and reportage, and the advancement of Britain's colonial exploits overseas" (Livingstone, 1992: 175).

Nowadays academic geographers distance themselves greatly from early precedents, although it would be naïve in the extreme to suggest that the output of current research efforts is somehow more independent or more intellectually 'pure' (in the sense of ideas written for their own sake, irrespective of expectations, pressures, or perceptions of their various audiences). Alongside other kinds of knowledge production, academic human geography is complicit in the continued expansion and power of corporate dominated publishing, events, and intellectual property industries. Geographical research output is increasingly reimagined as 'content' for corporate entities that have bought and amalgamated publishers, and oligopolized academic print outlets. Indeed, only recently are implications of the nesting of geographical pursuits in the corporate knowledge economy becoming more fully apparent. A recent example was the debate surrounding connections between Elsevier, the publisher of this encyclopedia, and a subsidiary company who among other activities were the organizers of defense industry fairs where military related arms contracts and deals were brokered. Geographers raised a range of direct and related concerns about how academic publishing is enmeshed with capitalist social relations; how close academia might be to global trade in the machines of warfare and torture (not just through Elsevier, but through the substantial shares that universities such as Cambridge and Oxford own in arms exporting corporations) and about the tactics academics could use in protest. After pressure from researchers Elsevier would eventually sever links through subsidiaries to the defense industry in order to improve its image of independence. Such dialogs have refreshingly opened up debate about research and the embeddedness of academic pursuits within and beyond geography in the apparatuses of capitalism and state control.

Who Has Benefited and Suffered from the Production of Human Geographical Knowledges?

Recent debates about academic publishing and links to corporations and the military echo earlier critiques of the connections between geography and the imperial interests of the state. They share a focus on who has benefited and suffered from the production of geographical knowledges. Geography was "the science of imperialism *par excellence*" (Livingstone, 1992: 170). From the 1500s up until World War II, geographers mapped places and environments that would become part of European colonial empires, they identified resources that would be plundered, they described cultural practices that were to be 'civilized', races to be 'saved'. But colonial geography was also a means through which European nationalism was cemented, because through it conceptions of European nationhood were forged: expeditions and subsequent geographical writings described other places, environments, and cultures in the eye of the beholder, and reinforced what it meant by contrast to be European, to be Western, to be British, or German. Cultural representations also had material effects. Colonial geographical maps and writings would support actual imperial expansions that made and remade those territories.

In Jeffersonian America, geography proved central to nation building, buttressing attempts to foster patriotism and a sense of independence in a newly colonizing nation. Jefferson himself was a vocal advocate for geography, funding the inland explorations of Lewis and Clarke, and even penning his own geographical (and rather boosterish) text on his home state, *Notes on the State of Virginia*. France's geography was reorganized after the Revolution, not just to change administrative boundaries and the like, but to create "locales wherein a revolutionary political culture was made, challenged and remade after 1789" (Heffernan, 2005: 277). One key exponent, Edme Mentelle, wrote explicitly Republican geographies "to mould new citizens" (Withers, 2006: 716). Elsewhere in Europe, geographical characterization of national populations increasingly showed the imprint of contemporary political concerns. In Greece, "Meletios, Archbishop of Athens, sketched a geography of European civility in his *Geography, Old and New*, praising the French for their culture, the Swiss for their struggles for political freedom and the English for being the 'most civil and tame'" (Withers, 2006: 716), while in Nikiphoros' *Elements of Geography* "Holland is singled out for its religious tolerance, the Swiss and the Swedes for their liberty, the English for their seriousness of character and scientific disposition." These human geographical writings were not born of scientific impulse; instead they intended to diplomatically "reassociate Greece, an ancient centre of classical learning, with modern centres of learning in western Europe" (Withers, 2006: 716).

Other interests have been furthered by human geographical praxis – some subtly contributing to the maintenance of systems of oppression; others contributing more destructively to outright tyranny. Religious maps of the medieval era were deliberately moralizing, and often depicted a world with Jerusalem or Rome as the rightful centers of religious power. The first Greek map of the world, dated at around 500 BC, was 'a diplomatic device', displayed in Miletus in an attempt to persuade Greek states to take up arms against Persia. It showed a world dominated by a vast Europe, towering over diminutive Asia and Africa. Writing about the transition from feudal to capitalist society in Europe, David Harvey described geographical knowledge production as pivotal. Geographers contributed exploration, cartographic survey, regional inventory, geopolitical taxonomy, and resource compilation – all central to imperial manipulation, management, and exploitation. Indeed, mapping in the contemporary world of geographic information systems (GISs) has not ceased to be a technology used in the governing and control of populations.

Motivations for geographers have included the idealistic, the selfish (fame and fortune), and the evangelical. Geographical societies had close links to missionaries and produced knowledges of the savage 'Other' that supported intentions to convert them. Such knowledges enabled immense and often destructive cultural change, although this was often complex in moments of colonial contact. Ratzel, like Mackinder, produced geopolitical knowledges that were meant to be read by foreign policymakers and shape state military actions. Although it is strictly incorrect to claim that Ratzel provided intellectual justification for Nazism, as is often thought the case (his philosophy promoted racial mixing as survival strategy – more accurately it was Karl Haushofer's work that was most used by the Nazis), Ratzel nonetheless justified German colonial activities, just as Mackinder supported Britain's fragmenting of post war Eastern Europe in order to create a buffer between Germany and Russia, thus averting their control of the Heartland, and protecting Britain's dominance.

In recent decades, geographers concerned with these pasts, and with ethics and justice, have thought more deeply about self reflexivity and positionality in research, and about how dubious agendas or questionable goals can be smuggled into academic enterprises. Many have redefined themes and research questions, and the manner in which research is carried out. In contrast to geography's deeply imperial past, research can now be a catalyst for addressing difficult moral and ethical questions, or a vehicle for empowerment and liberation. Research can give voice to marginalized groups, can relate to activist causes, encourage participation of subjects in research design and management, or be imagined ‘for’, rather than ‘on’, the poor or marginalized. Even geographical information systems (critiqued by some as throwbacks to positivism or as technologies of state control) have been increasingly explored as a tool for critical and participatory research.

Where Have Human Geographical Knowledges Been Produced?

Human geography has its own differentiated geography. It has addressed different intents, and has taken place in diverse institutional contexts that all shape its character, orientation, and reception. In some academic disciplines a strong sense prevails of truly international debates and paradigms. This is not necessarily the case in human geography. Ironically, even though human geography has global aspirations and an international presence, a sense of global uniformity or consensus is remarkably absent. Perhaps it is a marker of the respect geographers pay to the crucial roles of place, context, and geographical specificity that the very notion of global consensus is unpopular.

Even by the late 1700s the idea of a coherent and universal explanation of world geography had been challenged. Jedidiah Morse was so incensed by the Anglocentrism of British geography books that he inverted the usual practice of describing Europe first in his *The American Geography*, and dedicated 90% of its 536 pages to the New World. Similarly, William Guthrie, a Scottish writer, sought to expose how English assumptions had pervaded eighteenth century geography, and rewrote it using concepts from the Scottish Enlightenment instead.

Most recently, and perhaps most fundamentally, any sense of a canonical, universal story of human geography has been unsettled by critical geographers both within and beyond the English language sphere. Concern has been raised about the Anglo American and English language dominance of the academic publishing industry (see **Table 2**), and the assumption of universality in much Anglo American geography. American and British writers, it is argued, often assume global relevance for their work, yet are silent on their own particularity and inward gaze, reluctant to engage in comparative research in a truly international sense. Referees and editors of English language journals published from Britain or the United States have demanded that papers from ‘the periphery’ justify their interest to northern audiences or else face rejection. Such attitudes valorize Anglo American geographies as ‘unlimited’, and depict other geographies as ‘limited’. For Minca (2000: 287), “there is a widespread conviction both among many Italian geographers as well as among many of my European colleagues that these journals’ implicit claims of being ‘international’ … are patently absurd if not downright pretentious.” It is perhaps no coincidence that this has occurred in an era when English increasingly became a global lingua franca of business, science and the internet.

But just as Livingstone (1992: 142) cautioned against interpreting American geography in the 1800s as “a mere epiphenomenon, a pale shadow, of its European counterpart,” so too it is similarly naïve – nay offensive – to presume that the dynamic centers of geographical development since the 1960s have been England or the United States. Most English language geographers would not even be aware there are over 270 geography journals in Spanish and Portuguese, published from Spain and throughout Latin America. Meanwhile, in Japan quantitative geographers adopted and debated Western ideas quickly, and critical geographers were as early as the 1920s incorporating German Marxist ideas into their work on poverty (some 50 years before Marxism rose to prominence in English language geography). If Anglo American myopia is true, it runs counter to a history of international exchanges in past phases of geography. Italian geographical scholarship was not just an “implicitly peripheral national variation” but instead spoke “to a *fundamental political moment* in the genesis of European geography” (Minca, 2007: 181). During the quantitative revolution of the 1960s origins and inspiration came from geographers in Germany, Sweden, and Finland. In the case of the concept of cultural landscape, it was trafficked from German (*Landschaft*, in Humboldt) into French (*pays*, in Vidal de la Blache), and into English (Sauer). British and American geography has never only just been British or American. The “suggestion of a self generating (western) geographical tradition smacks of a fetishisation of the west… western geographical knowledge did not develop in some kind of vacuum, away from prior non western geographies. It *depended* upon them” (Sidaway, 1997: 74–75).

In light of these contrasting stories of myopia and cross cultural diffusion, it is fruitful to recast our image of the geography of human geographical knowledge production and dissemination not as some simple linear diffusion model – of centers and diffusions into margins – but as a more complex and hybrid set of flows. Geographers are sometimes connected with one another (and at other times not) across different countries and languages. Human geographical knowledge production and distribution occurs through ‘scenes’, made up of researchers and readers, much as one might imagine music scenes, artistic scenes, or political communities being organized within and across cities and nations. Fall (2007: 195) argued that the content of geography in languages differs less than the manner in which their texts circulate, shaped by “institutional structure, the nature of the academy and styles of debate.”

Table 2 Nationalities of authors in *Progress in Human Geography*, 1978 2007; nation of institutional affiliation of each author on all articles

	1978 1982	*1983 1987*	*1988 1992*	*1993 1997*	*1998 2002*	*2003 2007*[a]	*Total*	*% of Total*
UK	100	79	70	64	98	121	532	53.2
USA	28	29	53	38	47	65	260	26
Canada	13	10	8	14	10	15	70	7
Australia	5	6	2	6	2	6	27	2.7
Aotearoa New Zealand	2	1	5	5	7	4	24	2.4
Singapore	0	0	0	2	8	1	11	1.1
Ireland	0	1	1	0	3	4	9	0.9
Germany	0	0	0	1	0	8	9	0.9
Netherlands	0	3	3	1	1	0	8	0.8
France	3	0	1	0	1	2	7	0.7
Israel	0	3	2	0	1	1	7	0.7
Sweden	0	0	0	4	0	2	6	0.6
Finland	0	0	0	0	4	2	6	0.6
Japan	1	0	0	2	0	1	4	0.4
China	1	0	2	0	0	0	3	0.3
South Africa	1	1	0	0	0	0	2	0.2
Poland	1	0	1	0	0	0	2	0.2
India	1	0	0	0	0	0	1	0.1
Austria	0	1	0	0	0	0	1	0.1
Nigeria	1	0	0	0	0	0	1	0.1
Portugal	0	1	0	0	0	0	1	0.1
Brazil	0	1	0	0	0	0	1	0.1
Italy	0	1	0	0	0	0	1	0.1
Argentina	1	0	0	0	0	0	1	0.1
Swaziland	0	1	0	0	0	0	1	0.1
Hungary	0	1	0	0	0	0	1	0.1
Palestine	0	0	1	0	0	0	1	0.1
Spain	0	0	0	0	1	0	1	0.1
Denmark	0	0	0	0	0	1	1	0.1
Taiwan	0	0	0	0	0	1	1	0.1

[a] At the time of writing, only two issues for 2007 had been published, and were thus used in these statistics.

Human geography must therefore be understood in place and the institutional context within which it is produced. Geography departments can be very large, such as at Universitas Gadja Mada in Indonesia, but house only some of the subdisciplinary areas expected in Britain. Indeed, Gadja Mada has probably the largest academic geography unit in the world, constituting an entire faculty of the university, separated further into departments of physical geography and environment, human geography, cartography and remote sensing, and regional development. Little known outside Southeast Asia, Gadja Mada's teaching and research is focused on what is necessary and required by government, and by local communities with most pressing humanitarian needs (hence foci on agricultural development and rural poverty). In that institution, too, subscription to Western geographical journals is virtually impossible because of funding and departmental space limitations. Their understandable preference to teach and research in Bahasa Indonesia rather than English (which only some staff speak) mediates adoption of outside vernaculars. Even in the Global North, conditions of academic labor vary enormously, as do pressures to produce certain kinds of research and to conform to research impact assessment exercises; all these translate into variegated geographies of human geographical research across universities and cities.

More accurately, the geography of human geography has become hybrid. Editorial boards of most geography journals are now "genuinely committed to trying to deal with the politics of publishing" (Laurie, 2004: 142), although for many this is still a 'work in progress'. Strategies to internationalize and democratize human geography have made some headway, and debates have intensified about the pluralism of geography, the hegemony of English as a *lingua franca* in international publishing, and methods that could be explored to break this down or to use English for cross cultural dialog. Language barriers are never absolute. Some English language journals such as *Social and Cultural Geography* and *Tourism Geographies* translate all abstracts into at least two other languages, and many geographers (particularly from the non English speaking world) publish and work in more than one. Some journals publish in two or more languages (such as the *Canadian Geographer*, the Finnish *Nordia Geographical Publications*, the *Rivista Geografica Italiana*, and *Geographische*

Zeitschrift). Others (such as *Geografiska Annaler* and *Tijdschrift voor Economische en Sociale Geografie*) are from countries and regions that have national languages other than English, but nonetheless publish in English and have become "much more inclusive of alternative traditions" (Rodríguez Pose, 2004: 2). Changing political economic circumstances as well as growing diasporic academic networks have broadened and diversified Chinese human geographical output, even while China is woefully neglected in Anglo American publishing. In the case of *Arab World Geographer*, it has become a journal for a genuinely international mix of scholars from a range of political and epistemological perspectives. In has proved an important arena for debate and dissent over recent global conflicts, having commissioned symposia following the 2000 al Aqsa *intifada*, the 11 September 2001 attacks on the World Trade Center towers in Manhattan, and the 2003 US invasion of Iraq. Although efforts to speak across linguistic, cultural, and epistemological traditions always entail much hard work for those involved, journals that attempt it are more likely to publish genuinely international viewpoints. Greater reflexivity over research production is thus gradually altering the geography of human geographical activity.

Contemporary Human Geographies

Whereas this article began with an uncomplicated description of what is commonly understood to be human geography, it concludes here by arguing that the maintenance of the subdiscipline as discrete and internally coherent is far from straightforward. When most people talk about human geography, they refer to a type of academic pursuit undertaken in the modern era, and particularly from the latter half of the twentieth century onwards, which encompasses a diversity of ways of examining the presence and actions of humans in geographical space. But because of this, human geography is necessarily an interdisciplinary pursuit, one that evades precise definition and intellectual closure. How human geographies are done also varies from place to place and in different linguistic traditions, although traffic of ideas and methods across and between these has also been apparent. Contemporary human geographies are hybrid formations in which multiple paradigmatic viewpoints, drawn from multiple historical and spatial contexts, coexist and jostle within the same institutional and subdisciplinary spaces.

See also: Cartography, History of; Colonialism I; Colonialism II; Cultural Geography; Darwinism (and Social Darwinism); Ethnicity and Resistance, Historical Geographies of; Feminism/Feminist Geography; Feminist Geography, Prehistory of; Geographical Journals; Geography, History of; Geopolitics; Human-Nonhuman; Indigenous Geographies; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Physical Geography and Human Geography; Postmodernism/Postmodern Geography; Poststructuralism/Poststructuralist Geographies; Quantitative Revolution; Radical Geography; Taylor, G.

Further Reading

Anderson, K. (2005). Griffith Taylor lecture, Geographical Society of New South Wales, 2004: Australia and the 'state of nature/native'. *Australian Geographer* 36, 267 282.
Anderson, K. (2007). *Race and the Crisis of Humanism*. New York: Routledge.
Fall, J. (2007). Lost geographers: Power games and the circulation of ideas within Francophone political geographies. *Progress in Human Geography* 31, 195 216.
Gibson, C. and Klocker, N. (2004). Academic publishing as 'creative' industry, and recent discourses of 'creative economies': Some critical reflections. *Area* 36, 423 434.
Gregory, D. (1994). *Geographical Imaginations*. Oxford: Blackwell.
Gutiérrez, J. and López Nieva, P. (2001). Are international journals of human geography really international? *Progress in Human Geography* 25, 53 69.
Hartshorne, R. (1939). *The Nature of Geography: A Critical Survey of Current Thought in the Light of the Past*. Lancaster, PA: Association of American Geographers.
Harvey, D. (1984). On the history and present condition of geography: An historical materialist manifesto. *Professional Geographer* 36, 1 10.
Heffernan, M. (2005). Edme Mentelle's geographies and the French Revolution. In Livingstone, D. N. & Withers, C. W. J. (eds.) *Geography and Revolution*, pp 273 303. Chicago, IL: University of Chicago Press.
Johnston, R. J. (1987). *Geography and Geographers: Anglo American Human Geography since 1945*. London: Edward Arnold.
Laurie, N. (2004). The politics of publishing and print. *Geoforum* 35, 141 142.
Livingstone, D. (1992). *The Geographical Tradition*. Oxford: Blackwell.
Mayhew, R. (1998). The character of English geography, *c*. 1660 1800: A textual approach. *Journal of Historical Geography* 24, 385 412.
Minca, C. (2000). Venetian geographical praxis. *Environment and Planning D: Society and Space* 18, 285 289.
Minca, C. (2007). Humboldt's compromise, or the forgotten geographies of landscape. *Progress in Human Geography* 31, 179 193.
Mizuoka, F., Mizuuchi, T., Hisatake, T., Tsutsumi, K. and Fujita, T. (2005). The critical heritage of Japanese geography: Its tortured trajectory for eight decades. *Environment and Planning D: Society and Space* 23, 453 473.
Nishioka, H. (1975). Location theory in Japan. *Progress in Geography* 7, 133 199.
Rodríguez Pose, A. (2004). On English as a vehicle to preserve geographical diversity. *Progress in Human Geography* 28, 1 4.
Rose, G. (1993). *Feminism and Geography*. Minneapolis, MN: University of Minnesota Press.
Scott, A. J. (2004). A perspective of economic geography. *Journal of Economic Geography* 4, 479 499.
Sidaway, J. (1997). The (re)making of the western 'geographical tradition': Some missing links. *Area* 29, 72 80.
Sitwell, O. F. G. (1993). *Four Centuries of Special Geography*. Vancouver, BC: UBC Press.
Tapiador, F. J. and Martí Henneberg, J. (2007). Best of times, worst of times: A tale of two (Spanish) geographies. *Journal of Geography in Higher Education* 31, 81 96.

Whatmore, S. (1999). Culture nature. In Cloke, P., Crang, P. & Goodwin, M. (eds.) *Introducing Human Geographies*, pp 4 11. London: Arnold.
Withers, C. W. J. (2006). Eighteenth century geography: texts, practices, sites. *Progress in Human Geography* 30, 711 729.

Relevant Websites

http://www.antipode online.net
Antipode online forum.
http://age.ieg.csic.es
Asociación de Geógrafos Españoles.
http://www.aag.org
Association of American Geographers.
http://www.soc.nii.ac.jp
Association of Japanese Geographers.
http://www.cag acg.ca
Canadian Association of Geographers/L'*Association Canadienne des Géographes*.
http://cnfg.univ paris1.fr
Comité National Français de Géographie.
http://www.geographie.de
Deutsche Gesellschaft für Geographie.
http://www.ugm.ac.id
Gadja Mada University, Yogyakarta, Indonesia.
http://www.helsinki.fi
Geographical Society of Finland.
http://www.ucd.ie
Geographical Society of Ireland.
http://www.library.uu.nl
Geosource web directory.
http://www.iag.org.au
Institute of Australian Geographers.
http://www.igu net.org
International Geographical Union.
http://www.nzgs.co.nz
New Zealand Geographical Society.
http://www.ipgh.org
Pan American Institute of Geography and History.
http://www.geogr.ku.dk
Royal Danish Geographical Society.
http://www.geography.nl
Royal Dutch Geographical Society/*Koninklijk Nederlands Aardrijkskundig Genootschap*.
http://www.rgs.org
Royal Geographical Society with the Institute of British Geographers.
http://www.societageografica.it
Società Geografica Italiana.
http://www.ssag.co.za
Society of South African Geographers.
http://www.ssag.se
Swedish Society for Anthropology and Geography.
http://wwwsoc.nii.ac.jp
The Human Geographical Society of Japan.

Human Rights

R. Honey, The University of Iowa, Iowa City, IA, USA

Glossary

Civil Rights Rights to participate in the civil life of a society, without restriction by race, religion, or other contestable categories.

Cultural Rights Right to express one's culture, including language, religion, and patterns of social engagement.

Cultural Struggles Struggles over the norms of a culture or over the adjudication of different norms when the norms of different cultures are in conflict.

Economic Rights The right to participate in an economy, as well as to have basic needs met.

Human Rights The protections and privileges one has as part of the shared heritage of humanity.

Political Rights The right to participate in the political system, including such rights as to expression as freedom of speech, freedom of the press, and the right to assemble.

Social Rights The right to participate in the daily life of a society, as in the right to marry and form a family.

Universality The applicability of all rights to all people under all circumstances.

Introduction

Human rights, in the time since World War II, have become widely recognized as protections for people against the oppressive actions of states and other actors. Human rights are important in several ways, most significantly in that they have come to take on constitutional status as protections for people and as claims to having basic needs met. In theory, human rights are those rights held by all individuals simply as a consequence of their status as people, part of the common heritage and endowment of human kind. Human rights have come to be regarded as inalienable in that they can neither be taken away nor given way. They are not subject to the powers of class or on the basis of classifications by gender, race, religion, or political ideology. In theory, human rights are universal, applying to all people, independent of where those people are.

Human rights reality, however, differs from human rights theory. Though several international agreements hail human rights and affirm their status before the law, the protection of rights on the one hand and their violation on the other, vary over space and through time as people struggle to define and redefine rights, struggle over interpretations of rights, and struggle to rein in human rights abuses.

Though the global human rights movement claims that rights codified in international agreements are 'universal' in that they apply to all people in all places, human rights are the product of political processes and hence have their own political geography. The 'political geography' of human rights involves:

- Where various human rights are protected and, conversely, violated;
- The extent to which societies and polities have bought into competing conceptualizations of human rights; and
- Cultural struggles over human rights at three discernable scales:
  - struggles with a culture over what should be considered rights;
  - struggles between cultures over whose conception of rights should be enforced or how differing conceptions should be adjudicated and practiced; and
  - struggles between a culture and global forces, either resisting widely accepted human rights or resisting global forces that violate rights deemed as important by the culture.

An explanation of the meaning and history of human rights is followed by a discussion of essential debates about human rights, and then an explication of the various political geographies of human rights.

History

From the *Magna Carta* to the twenty first century, rights have been construed as: individual privileges recognized by the state; protections from state interference in a list of individual or group activities; and claims that individuals or groups could make for gaining the necessities of life. Toward the end of the second Millennium, human rights came to be regarded as protections from any potential oppressor, not just a state.

Though the emergence of a global human rights regime only transpired in the twentieth century, the claim for rights, whether logically or legally, can be recognized as far back as the ancient Greeks. The idea of rights gained major momentum with the American and French Revolutions late in the eighteenth century. Most significantly, they challenged the traditional concept of sovereignty, changing it from something held by a monarch (or other ruler) and thus giving that ruler unquestioned, unchallengeable power over the people, to a concept in which the basic root of power was the sovereignty of the people who were empowered to delegate control over the machinery of governance to states or other entities that would then act in accordance with the interests of the people. Neither of these revolutions of course succeeded in establishing a human rights regime. The French Revolution succumbed to its own excesses and the military might of those benefiting from the previous conception of sovereignty. The American Revolution codified the rights of white men of property, not all people. Indeed, the constitution of the United States thrice mentions slavery, the epitome of a political and economic system with the most fundamental violations of what later came to be construed as human rights. Slaves were to be counted as three fifths of a person in two ways, one for taxation purposes, the other when allocating by population the number of seats a state would have in the House of Representatives. As part of the historic compromise between northern and southern delegates, the constitution also precluded prohibition of the slave trade for 40 years after the ratification of the constitution.

Legalized systems of chattel slavery were of course largely abolished in the nineteenth century, in some ways, setting the stage for the attainment of a fully fledged human rights regime in the twentieth century. That regime, though, came to the fore only after the excesses of two world wars and the global economic upheaval of the inter war period. Heretofore, states so jealously protected their sovereign rights – particularly the right to do anything within their own borders with impunity from external sanction – that they rebuffed any efforts to interfere with the power of states in the territories they ruled. The emerging body of international law had not penetrated the control of a sovereign state over the people residing within its recognized boundaries. Woodrow Wilson and others tried to create global institutions, or at least institutions representing the major global powers, in the years immediately following World War I. Their failure to do so, along with the burdens placed on defeated Germany, eventually came to be seen as significant factors in two ensuing disasters: first, the Great Depression that devastated the global economy in the 1920s and 1930s; and second, the rise of fascism, particularly in Germany, that not only led to World War II but also to the most egregious violations of human rights by any state in the history of the world – the Holocaust.

After World War II, the newly formed United Nations set into motion a process that resulted on 10 December 1948 in the adoption of the Universal Declaration of Human Rights (UDHR). The UDHR is a remarkable document not only in its delimitation of a set of 26 articles defining rights but also given the highly politicized nature of the process leading to its adoption. The process almost foundered on a Cold War confrontation between the emphasis of the United States and its western European allies on political and civil rights and the emphasis of the Soviet Union and its eastern European allies on economic rights. Furthermore, people from much of the world had no voice in the deliberations over the UDHR. Living in colonies without UN membership, they were not directly represented in the deliberations defining what their fundamental rights were to be. The adoption of the UDHR, however contentious its generation, marked the first in a series of achievements that created a global human rights regime. Other agreements followed, though with considerable struggle, even rancor. They included two major agreements adopted in 1966, the Convention on Economic, Social, and Cultural Rights and the Convention on Political and Civil Rights. Subsequent global agreements include: the Convention on the Elimination of All Discrimination Against Women in 1979, the Declaration of the Right to Development in 1986, and the Convention on the Rights of the Child in 1989. The United Nations human rights apparatus also provided for a High Commission on Human Rights and regional entities. In time, the regional entities produced their own human rights declarations, among them the European Convention for the Protection of Human Rights and Fundamental Freedoms in 1950, the American Convention on Human Rights in 1969, and the African Charter on Human and Peoples' Rights in 1981.

Significantly, the claims of human rights have been accompanied by arguments that they should be protected against violations by any actor, certainly states but not just states. Otherwise major twenty first century forces such as large corporations, nongovernmental political entities (including military organizations challenging states), international governing bodies (including those consisting of representatives of states, e.g., The World Trade Organization), and nongovernment organizations would be able to violate the rights of people without challenge.

Universality

The logic of human rights requires that they apply to everyone, that is, 'universally'. The argument basically is that only when a right is so broadly recognized and accepted, without exceptions, can it be protective against the onslaught of recalcitrant states or other potential oppressors. 'Universality' in terms of applicability and enforcement, however, does not correspond with 'universality' in terms of agreement. This is true both for the rights adopted by international agreement and for campaigns to get additional 'rights' recognized. Arguments for universality typically call upon some notion of 'natural law', whether from the philosopher Christopher

Brown (in Patman) or the geographer David Harvey. Both Brown and Harvey, after reviewing other ways out of the logical dilemma, conclude that some human protections and claims are so fundamental – so “natural” – that they constitute universal human rights claims based on the common humanity of all people.

Universality does not, however, connote universal acceptance. African and Asian critics have argued for ‘exceptions’ based on their cultures. The Banjul Charter, as the African Charter on Human and Peoples’ Rights is known because it was agreed in the Gambian capital city of that name, calls for two very significant caveats. One limits rights unless a person has met responsibilities. Rather than simply recognizing an inherent right to food, for example, the Banjul Charter conditions the right to food on one’s contribution of labor, more generally conditioning “. . . the enjoyment of the rights and freedoms . . . (on) the performance of duties . . . .” The other protects the rights of families and “peoples” from claims of individual rights. This call for African exceptions to what is seen as Western overemphasis on individual rights is matched by similar claims for Asian exceptions.

Even with the challenges to universality, the concept is extremely important in human rights law because once human rights documents gain international approval, they become more important than laws and even constitutions, ‘trumping’ other legal standards and other legal positions. As alluded to above, gaining international agreements in the years immediately after World War II turned on competition over the preeminence of civil and political rights on the one hand versus economic, social, and cultural rights on the other. This distinction is often explained as a generational distinction among types of rights, as explained next.

Generations of Rights

As suffrage expanded over the centuries, and as the rights of citizens to challenge the state progressed, rights became identified with what now are known as political and civil rights. These rights are designed to assure the freedom of individuals to petition their government and to express themselves even in ways unpopular, whether to the government, in particular, or the society at large, more generally. They include most fundamentally the right of the people to choose their governments through regularly scheduled, fair elections. They also include the communication freedoms – speech, press, assembly – as well as freedom to practice the religion of one’s choice. The United States and its western European allies emphasized first generation rights when the UDHR was being crafted. The rights central to them were clarified in the Covenant on Political and Civil Rights.

The second generation of rights involves not just freedoms but more. These rights include several codified in the UDHR. These and some additional ones are defined more precisely in the Covenant on Economic, Social, and Cultural Rights. In the economic arena, they include the right to have one’s basic needs met, including education, healthcare, housing, and nutrition, as well as the right to employment. In the social arena, they include the right to be treated fairly without regard to one’s race or ethnic background, class, or caste. In the cultural arena, they include the right to a culture, including one’s language. The Soviet Union emphasized these rights in the UDHR proceedings, setting up a position distinctive from that of the Western advocates of civil and political rights.

By the end of the twentieth century, human rights activists and scholars identified a third set of rights with characteristics somewhat different from the other two. These are survival rights, including one’s right to stop others from threatening one’s survival, whether through behavior that damages health, the environment, or the sustainability of the global system. Given the traditional focus of geography on human–environment interactions, these rights are inherently geographical.

Cultural Struggles Over Human Rights

Arguments for universality of rights aside, human rights are, in truth, socially constructed and the consequence of profound struggle. Indeed, there are three readily identifiable scales of these struggles: struggles within a culture over the definition of a just society; struggles between cultures as each tries to dominate the other or at minimum achieve a resolution recognizing the legitimacy of that culture’s values and preferences; and struggles of a culture against global forces, both defense against what is perceived as interference by the global human rights movement and struggles of a culture against the oppression of global forces, whether economic or political. A consideration of struggles at each scale follows.

Human Rights Struggles within a Culture

Human rights struggles have been central to the debates of most societies in the years since adoption of the UDHR as those societies have reconsidered their own conceptualizations of right and wrong, their own senses of how society should be organized, and their own senses of justice. Some of those struggles have involved implementation of the UDHR. The American civil rights movement exemplifies struggle over political and civil rights in the Global North. The United States’ criticism of lack of freedoms in the USSR appeared hollow given the use of Jim Crow Laws to disenfranchise the African American community of the United States. These laws limited the basic freedoms of African Americans in 34 of the then 48 states, the exceptions mainly being Midwestern and Northeastern states. A major element of the American debate over civil rights had to do with whether each of the member states of the Union could determine for itself issues of fundamental rights or whether these should be defended by the federal government over the objection of individual states. All three branches of the US federal government were theaters of struggle as the people of the country reconsidered the historic justifications for *de jure*, or legal, discrimination on the basis of race. The struggle soon swung to *de facto* discrimination based on social and economic distinctions.

A Global South example of struggle over political and civil rights is the struggle over ending the caste system in India. Though the caste system is illegal, it remains the social norm within the country. Mahatma Gandhi, whose philosophy of resistance through nonviolent civil disobedience not only hastened an end to British colonialism in India but also provided the foundation for a similar movement led by Martin Luther King to promote civil rights in the United States, argued strenuously against maintenance of India's caste system. As in the US, *de jure* segregation by caste has largely been shunted to history. Even more than in the US, however, *de facto* discrimination permeates daily life in India.

These two examples – the civil rights movement in the United States and the similar movement against the discrimination inherent in the caste system in India – exemplify struggles within a culture over recognition of rights in the UDHR. Struggles also occur within cultures over the codification of new rights – rights not widely recognized in the ratified international agreements – as those cultures reconsider for themselves what it means to be a just society. Two areas of considerable debate over the recognition of new rights in recent decades have been the campaigns against the death penalty and for gay rights. When the UN adopted the UDHR, almost all countries had the death penalty and almost none protected gays, lesbians, and transgendered people from discrimination.

The tide against the death penalty has been strong in recent years. From 1985 through 2006, 50 countries abolished the death penalty with only four adopting it, reducing the number of countries with legal executions to 68. Latin America has long eschewed the death penalty, the major exception being Cuba. Almost all European countries have moved away from the death penalty, mainly because of a consensus that it is inhumane but also because it is a requirement for joining the European Union. Most Muslim and African countries retain the death penalty, but China is easily the champion of the death penalty. Of the 1591 people officially executed in 2006, China executed 1010, followed in order by Iran with 177, Pakistan 82, Iraq and Sudan each 65, and the United States 53, leaving 139 executions in the other 19 countries. Though 38 of the US's 50 states retain the death penalty, in 2006 only 14 executed anybody, Texas being far and away the leader with 23 executions. Clearly in the 60 years since adoption of the UDHR, the death penalty has become exceptional rather than routine and the prohibition of the death penalty has become a broadly accepted human right not only by such advocacy groups as Amnesty International but widely though, not unanimously, across the globe. In many cultures the struggle against the death penalty has basically been won. In others, with the United States as the major example, the struggle is ongoing. Elsewhere, particularly in Asia and Africa, the struggle has hardly begun with the death penalty widely accepted.

Though not nearly as successful as the antideath penalty campaign, the movement for gay rights has in historical terms made dramatic progress since adoption of the UDHR. With arguments based on scripture, many Muslims and Christians continue to reject gay rights out of hand. Elsewhere, including much of Europe and North America, the struggle is on. In a few places, among them New Zealand, the Netherlands, and the Scandinavian countries, national legislation provides protection against discrimination for those who are gay, a reality hardly conceivable when the UDHR was adopted in 1948.

Human Rights Struggles between Cultures

Clearly cultures struggle over human rights as they consider and reconsider how to define a just society and how they want their own lives to be structured. These struggles are compounded when people with quite different cultures live together in the some country.

A poignant example of human rights struggles between cultures is the ongoing struggle for political power in Fiji, a country fairly evenly divided between indigenous Fijians and Fijians of Indian ancestry whose ancestors moved from South Asia to Fiji when Britain ruled both during colonial times. Not only is there the predictable conflict for political advantage between people with two quite different cultures, there is also a claim that the indigenous Fijians, by right, should maintain control over their ancestral lands, their claims trumping the one person/one vote claims of normal democracy. Briefly put, the argument is that an external power (Britain) usurped the power of indigenous Fijians by inserting an alien population. Justice, in accordance with this argument, requires that the indigenous Fijians be allowed to maintain control over their territory. The counter argument is that a person is a person so that everyone should have an opportunity to participate equally in the political process. Furthermore, the Indo-Fijian population had no voice in going to Fiji, rather being descendants of people who moved there with questionable degrees of influence over their own migration. These people should not be punished because their ancestors chose to move or were moved over their own objections. Since 1987, Fiji has had three military coups, each time the key issue being the continued dominance of the indigenous Fijians. The first time New Zealand considered a military response to return power to the duly elected government. In addition, members of the British Commonwealth expelled Fiji. This is an example with a clear conflict between principles: the rights of indigenous peoples versus free and fair elections based on genuine democracy. The indigenous Fijians have maintained power in the face of external pressure because that pressure has never risen to the level of military intervention or sufficiently serious economic sanctions to threaten the Fijian economy. The struggle continues.

A second kind of human rights struggle between cultures stems from colonial histories that created twentieth century sovereign states from areas that had been convenient for the colonial masters rather than a consequence of a political process that included agreements of people to be part of the same country. This problem is most serious in Africa, though not limited to that continent. Nigeria provides an instructive example. The internally generated difference is that before colonialism the area that became Nigeria had, literally, hundreds of cultures, the great majority having their own polities. The largest was the Sokoto Caliphate which had gained control over most of the north in an Islamic jihad that had begun in the early years of the nineteenth century. The smallest included several very small language groupings of a few thousand people each in the southeast. In addition to the ethnic divide, the Christianity that came with colonialism created a religious divide – with missionaries moving inland from the coast and Islam spreading from the Sokoto Caliphate in the north. Two hundred years after Dan Fodio began the jihad, the modern sovereign state of Nigeria is divided both by language and religion. These divisions permeate all realms of Nigerian life, including cultural struggles over human rights. One struggle has been over the implementation of *shari'a* law in several northern states, though southern Christians interpret the federal constitution as banning it. Accommodations to recognize differences in family structure and tradition even among people living as neighbors have been successful. Family laws, for example, are enforced on a sectarian basis with Muslims following Muslim tradition, Christians their own traditions, and those practicing only traditional religions whatever their traditions call for. Violent clashes among competing groups are all too common, however, with the clashes stemming from such disparate origins as land claims among ethnic/language groups to perceived slights along the religious divide. Nigeria continues to struggle to find mechanisms to allow its very different cultures to live peacefully in each other's midst.

A third kind of human rights struggle between cultures is of much more recent vintage with massive migrations of people from the Global South to work in the Global North, whether Latin Americans in North America or Africans or Asians in Europe. A number of European countries, for example, have to deal with the paradoxes of maintaining their cultures and being sensitive to the need of their immigrant populations, particularly when the immigrants practice different religions and have significantly different marriage practices. Similar tensions exist in the oil countries of the Middle East when workers with very different cultures outnumber the citizens of several sovereign states.

Human Rights Struggles between a Culture and Global Forces

Two kinds of cultural struggles over human rights exist at the global scale. One is the resistance of a culture, however small or large, to the global human rights movement. The other is resistance to global forces that violate human rights.

Before the UDHR and subsequent human rights instruments became available as part of the global human rights movement, oppressors could stand behind the shield of sovereignty to prevent external intervention when the oppressors violated the rights of the people under their control. Oppressors – whether Pinochet, Kim, Mugabe, or any of a number of others, recent and current – still claim immunity, and the world may be either insufficiently organized or committed to do anything about it. At minimum, a significant share of world opinion brings opprobrium down upon recalcitrant tyrants. In addition, some human rights claims to universal applicability can be challenged if for no other reason than that rights can sometimes be in conflict, however much advocates of rights want to claim otherwise. Consider, for example, the inherent conflict between freedom of the press and the right to a fair trial. Britain and New Zealand tend to fall on the side of the right to a fair trial, preventing disclosure by the media of certain facts, including in some cases the identity of the accused, until a decision is rendered. The argument is that identifying a person charged with a crime would tar that person's reputation permanently, even if the person were eventually acquitted. In the United States, the bias falls the other way with very little restriction on coverage in hopes that a fair jury can be found and then be quarantined from the media for the duration of the trial. The fear on the US side, that a nefarious government could control the press inappropriately if given the opportunity to limit coverage – including, for example, preventing the disclosure of information that could be damaging to public officials. Rulers may have some room to maneuver as they resist the global human rights movement.

The African and Asian claims for exceptions to global human rights standards were introduced above. Problems occur when these are merely smokescreens, calling upon tradition to justify female genital cutting or child brides, for example, or to justify inhumane living or working conditions. Whenever an exception to global standards is sought, the onus should be on those seeking the exception to show that it will not subject others to violations of their dignity, rights, and freedoms, to paraphrase the Banjul Charter.

As clear as the potential for exploitation is in the case of cultures asking to be relieved from the pressures of the global human rights movement, the evidence is overwhelming when cultures ask for protection so their human rights are not violated by the operation of the global political economy.

The case of Kil Soo Lee is instructive. A South Korean businessman working in cahoots with local functionaries, Lee was able to establish a shirt manufacturing factory in American Samoa. Lee and his organization attracted Vietnamese workers to his facility. He was convicted of human trafficking for basically holding the workers as slaves in his sweatshop and was sentenced to 30 years in a US prison for his offenses. The Lee case is an example of what Kevin Bales calls the new slavery which in some ways is worse than chattel slavery because the enslaver has no reason to maintain the health of the slave. People are trafficked, meaning moved surreptitiously from one place to another for the purpose of illegal employment under the control of another, on all the inhabited continents. Critics such as Bales berate the global political economy for doing little to trace the trafficked people or hold the perpetrators responsible for their actions. This is part of why the Lee case stands out, given its successful prosecution.

Other areas of local jeopardy – and human rights violations – given the operation of the global political economy include the diminution of education in Africa as a consequence of Structural Adjustment Programs, movement of toxic materials from the Global North to the Global South, and the proliferation of arms through regions with weak governments and people with the ability to pay.

The Political Geography of Human Rights

The introduction above distinguished among three political geographies of human rights: Where various human rights are protected and, conversely, violated; the extent to which societies and polities have bought into competing conceptualizations of human rights; and cultural struggles over human rights at three discernable scales: The third having been addressed at length, focus turns to the first two – the distribution of rights and conceptualizations of rights.

Establishing whether places are better or worse in terms of human rights violations can be addressed in several important ways. One is to take as a gauge the reports of such organizations as Amnesty International and Human Rights Watch, two of the most respected human rights organizations. Each, annually, reports to the world about the status of human rights across the globe, paying particular attention to the behavior of national governments. Rare is the country found to be committing no human rights violations. Sadly, all too common are the countries that heavily violate the rights of their citizens and others under their influence. These tend to be the less democratic and less open countries, particularly those of Africa and both Southeast and Southwest Asia. The United States and, to a lesser extent, Britain have been cited for violations encompassing the war in Iraq. China, continuing its traditional opaque administration even while opening up its economy, is also found wanting. The limitations, particularly with Amnesty, stem from an overwhelming emphasis on first generation rights.

Second generation rights are amenable to other types of analysis. Sometimes very direct measures can be met. The UDHR says that all children of appropriate age have a right to a free education. Is it so? Are all children of appropriate age in school, and are they required to pay? Being in school does not prove that anything was learned, so surrogate measures may be employed to infer the degree to which rights are met. If, for example, all children in the last generation were indeed educated, then a country should have a high literacy rate among its young adults. Furthermore, if girls are being educated as well as boys, the literacy rates for men and women should be similar. The data on these issues are troubling. Thanks to the reduction in educational expenditures forced on poor countries by the International Monetary Fund and World Bank from the 1970s through the 1990s, a lower percentage of African children are in school in the first decade of the Third Millennium than was the case 30 years ago. Likewise, female literacy rates lag woefully in the countries of the Middle East.

The provision of healthcare can also be measured both directly and indirectly. Direct measures include such things as the percentage of children receiving immunizations or the percentage of pregnant women receiving prenatal care. Indirect or outcome measures include such things as rates of diseases or infant mortality. An interesting trend is that though the gap in infant mortality remains, almost all countries have achieved long term reductions in infant mortality. Northern and Western European countries have the lowest infant mortality rates, as they did a generation ago. A number of countries with midlevel economic development actually have infant mortality rates comparable to those of the best countries a generation ago. This good news is offset by increasing problems with infectious diseases, particularly HIV/AIDS and tuberculosis, and such vec tor borne diseases as malaria and encephalitis.

The Political Geography of Human Rights in the Twenty-First Century

Examining the distribution of human rights shows that the degree of human rights violations is strongly related to the political geography of commitment to rights. The globalization that has dominated the world over the last generation not only brought with it economic pene tration that allowed the development of Bales' new slavery, it also brought with it a commitment to demo cratically elected governments and transparency. This is not to argue that governments are truly democratic in their electoral system or, in fact, transparent in their operations. It is to argue that democracy and transpar ency have been global expectations. The places with the strongest human rights records are the places with commitments to open government on the one hand and protection of those who need protection on the other. They are the places in effect with commitments to both first and second generation human rights.

See also: Civil Society; Cultural Studies and Human Geography; Population Geography.

Further Reading

Bales, K. (2000). *Disposable people: New slavery in the global economy*. Berkeley: University of California Press.
Brysk, A. (2004). *Globalization and human rights*. Berkeley: University of California Press.
Harvey, D. (1996). *Justice, nature and the geography of difference*. Cambridge: Blackwell Publishers.
Lauren, P. G. (1998). *The evolution of international human rights*. Philadelphia: University of Pennsylvania Press.
Lee, R. and David, M. S. (2004). *Geographies and moralities*. Cambridge: Blackwell Publishing.
Patman, R. (ed.) (2000). *Universal human rights?* New York: St. Martin's Press.
Santoro, M. A. (2000). *Profits and principles: Global capitalism and human rights in China*. Ithaca, NY: Cornell University Press.

Humanism/Humanistic Geography

J. M. Smith, Texas A&M University, College Station, TX, USA

Glossary

Humanities Forms of scholarship that seek to achieve or enable understanding of the world as it appears to another person, particularly persons of other periods, places, and social roles.

Intention The aim or purpose that gives shape and meaning to a human action; unlike predictions of human behavior, which are supported by evidence, intended acts are justified by reasons.

Interpretation The act of understanding of another person's intentions, whether practiced spontaneously as a natural aspect of social psychology or self-consciously in conformity with an explicit hermeneutic.

Place (*Lebenswelt*, *Habitus*) A locality psychically and materially ordered after a worldview; such a locality understood as intelligible and reasonable (offers reasons to act); such a locality understood with strong feelings of personal affinity.

Understanding (*Verstehen*) To imaginatively grasp the intended object of another person's actions and the assumptions within which those actions are a reasonable means to that end.

Worldview An interpretation of reality, originating in practices, which organizes experience into natural kinds and functional categories, assigns these intentional objects value, and sees in them reasons to act, think, and feel in a particular way.

Introduction

Humanistic geography is the conventional title of a loosely organized school of thought that rose to prominence after 1970, flourished for nearly two decades, and then combined with neo Marxism and postmodernism to form the 'new cultural geography'. This movement was important, and properly styled humanistic, but should not be taken as having begun, or completed, humanistic though in the field. Genuine humanistic geographers have been at work in the field for a very long time, still are at work, and hopefully always will be. This is because some geographers will always find their fuglemen in the humanities, not in the sciences, and because some geographers will always see the point of trying to understand the world as it appears to particular human actors.

Before detailing the history and doctrine of humanistic geography, broadly or narrowly conceived, it is necessary to consider its source and wider context, the doctrine of humanism. Humanism presupposes that humanity possesses a natural dignity, and that this dignity is evident in human individuals, human interests, and human works. Humanism elaborates this dogma of human dignity in four important ways: metaphysical humanism, topical humanism, methodological humanism, and philanthropic humanism.

When humanism first appeared in the fourteenth century, it was in the context of a Christian culture that made similar claims for human dignity and transcendence over the natural order. Christianity taught that men and women were made in the image of God, and set apart from other creatures in a unique spiritual predicament. Christianity also looked upon the human world as meaningful, beautiful, and good, and set its face against nihilism and Gnostic *contemptus mundi*. At the same time, however, Christianity qualified these affirmations with the aspersion that humans and the human world are, for the present, bent and perverted by original sin: that we are fallen men and women living in a fallen world, with no prospect of redemption apart from that afforded by the atoning self sacrifice of Jesus Christ. Humanism may, therefore, be reasonably described as Christianity shorn of the fall, original sin, and the need for supernatural redemption. (Christian humanism is possible because, to many Christians, humanism appears true, but incomplete.)

The first humanists were medieval scholars who studied classical texts of Roman, and later Greek, philosophy, poetry, history, and rhetoric. They were by these secular studies distinguished from divines, who read Holy Scripture and theology. Petrarch is often named as the first humanist, remembered today for his love of the *belles lettres* of the classical world, and as the first man to climb a hill simply to admire the view. Although undoubtedly Christian, Petrarch thereby marks the beginning of a shift in the attention of the best minds in Europe, from Holy Scripture and heaven to secular literature and the visible face of the mundane world. It was not impious or futile, humanist scholars believed, to set aside the lamp of revelation and examine humans and human affairs with natural reason alone. Humans were, after all, possessed of a dignity that, however inferior to the dignity of God, nevertheless merited such study, and as an aspect of this dignity, humanist scholars were possessed of rational powers sufficient to undertake this study unaided.

Begun as a palliation of the Christian doctrine of fallen humans in a fallen world, humanism became in time an increasingly assertive repudiation of these unflattering imputations. Thus, although the earliest humanists had seen secular knowledge as a supplement to theology, some of their successors were epistemological chauvinists hostile to theology. These humanists pronounced natural reason and the example of nature sufficient to human needs, and theology not only dubious to natural reason, but also demeaning to human dignity. Consequently, the title humanism came in some instances to denote metaphysical deism, agnosticism, or even dogmatic atheism. In the last instance the title humanism was intended to emphasize the positive content of the atheistic doctrine (unwilling to surrender the term, Christian humanists describe this metaphysic as secular humanism). Not all geographers are infidels, of course, but almost all of modern geography is written as if metaphysical humanism is true. (Curiously, however, 'humanistic' geographers have written every testimony to religious faith the author has encountered in the geographic literature.)

Whatever their metaphysical and religious prejudices, the early humanists were fascinated by the thoughts of Roman authors, particularly when these thoughts were expressed with unusual eloquence. A writer like Cicero had for them an attractive mind and an attractive voice. Along with other classical authors like Livy, Seneca, Virgil, Ovid, and Catullus, his way of 'seeing' or 'imagining' the world had, for the humanists, a life affirming human quality that, in their opinion, contrasted favorably with the grim and 'gothic' imagination of their late medieval contemporaries. Topical humanism is this deep interest in the way that people, especially extraordinary people, 'see' the world. It is delighted by the human imagination, drawn to examine the inside of human experience, the subjective ethical and esthetic lives of individuals — particularly those with a gift to convey in writing some sense of that inner life. The modern humanities are, of course, far more democratic than the early humanists, but nevertheless continue this interest in imagination, the inner life of a person, and the curious connection of imagination to language and other media of expression. Topical humanism, as we shall see, is central to humanistic geography.

The medieval humanist employed a particular methodology in his research. He mastered ancient languages, compared and edited surviving texts, and wrote interpretive commentaries. In this he differed little, methodologically speaking, from theologians and biblical scholars, who were also engaged in the recovery, preservation, and exegesis of texts. The humanist became methodologically distinctive only in the seventeenth century, when experimental science appeared with its novel method and definition of research. Humanists were thereafter confronted with a powerful epistemological rival, and were before long subject to contumely not unlike that which some humanists had heaped on theologians. Some scientists have always asserted what the geographer Fred Schaefer called 'methodological monism', or the opinion that the scientific method is the only road to knowledge. Humanists have, however, generally succeeded in denouncing these epistemological pretensions as 'scientism' or 'positivism', and affirming the reliability of humanistic methodologies. Humanists argued, plausibly, that the experimental scientist may be able to explain the objective world, and even human consciousness, but that only the humanist, using the humanistic tools of empathy, interpretation, and critical judgment, can hope to 'understand' these things as they appear, from the inside, to human actors. And the dignity of human consciousness and the human world demands a recognition that comes only by its being understood. This methodological dualism between the experimental sciences and the interpretive humanities runs, as we shall see, right through geography.

Given that a methodological humanist is antipositivist, it is ironic to find that in its fourth sense the word humanist is very often closely related to positivism and science. Indeed, positivist geographers very likely think of themselves as humanists in what we call the philanthropic sense. Philanthropic humanism is primarily a product of the eighteenth century Enlightenment, when metaphysical humanism is combined with experimental science. Human dignity seemed to demand an improved standard of living, or 'progress', and the natural and social sciences promised to deliver these improvements. This is why, throughout the nineteenth century, the word humanist was often synonymous with humanitarian, and denoted anyone who worked to ameliorate the human condition. Under this construction the experimental scientist was arguably much more of a humanist than the scholar who labored over old books.

'Humanistic' Geography before 1970

As part of the intellectual project of the modern age, all of modern geography has probably understood itself to be humanistic in the philanthropic sense of contributing to the advancement of human welfare. The author knows some misanthropic geographers, but no intentionally misanthropic geography. However, until the 1970s, when the title took on much narrower connotations, the title humanistic geography was often understood to denote that portion of geography that pursued applied rather than pure research. This is what V. C. Finch meant when he described regional geography as a "humanistic earth science" in 1939, and what Donald W. Fryer and William Bunge meant three decades later when they wrote that a humanist is simply a "partisan for *Homo Sapiens*." Philanthropic humanism has by no means disappeared from geography, but since the 1970s one could not describe all such geography as humanistic with any hope of being understood.

Because overtly teleological and theistic geography largely disappeared in the nineteenth century, there has been no need for geographers to use the phrase hu manistic in its metaphysical sense. Whatever the private religious sensibilities of individual geographers, the great bulk of geographic discourse has, for more than a century, wedded an astringent naturalism to a sometimes bump tious anthropocentrism. The virtual absence in geog raphy of a recognized minority tradition of theistic or specifically Christian scholarship is probably due to the origins of the modern field in geology and biology. Such minority traditions are a recognized presence in all of the humanities, most notably philosophy, literary studies, and history. Yet in geography metaphysical humanism is so much taken for granted that the virtual absence of a vocal minority of metaphysical dissidents is seldom if ever noticed.

Beginning in the late nineteenth century geographers were also using the words humanist, humanism, and hu manistic more narrowly, to indicate human geography. For instance, the 1893 'Report of the committee of ten on secondary school studies' described geography as con sisting of natural and "humanistic geography," meaning by this later term anthropogenic landscape features such as "streets, railways, wharves, harbors, parks, plots, wards, etc." All of human geography was part of what H. J. Fleure in 1944 called the "new humanism," a phrase that indi cated all scientific and historical investigations into the nature and development of humans and human society.

As early as 1906 geographers were using the term humanistic in an even narrower sense still, to describe that portion of geography that was not only topically different from physical geography, but also methodo logically different from science. J. K. Wright, who had bruited the idea since the 1920s, may have written the first fully developed description of this difference in his 1946 presidential address to the Association of American Geographers. Here Wright called for geographers who were "scholars in the humanistic sense," by which he meant "widely read," "masters of [literary] style," and hungry to explore "the most fascinating *terrae incognitae* of all . . . those that lie within the minds and hearts of men." Here we see very clearly expressed the themes of topical and methodological humanism. Provided that these 'geosophers' did not grow too numerous or assertive, Wright concluded, there was no reason to fear that they would "retard the advancement or diminish the prestige of scientific geography."

Several geographers were, indeed, already working on geosophy and the geographic imagination, or were about to begin doing so. Ralph Brown's *Mirror for Americans: Likeness of the American Seaboard, 1810* and *His torical geography of the United States* embodied all of the elements of Wright's scholarship in the humanistic sense. In his 1941 address to the Association of American Geographers, Carl Sauer defined culture as very largely consisting of a learned way of imagining or interpreting environment. By the early 1960s David Lowenthal was developing Wright's basic themes, with far less deference to science. In 1967 Clarence Glacken published *Traces on the Rhodian Shore*, perhaps the greatest example of topical and methodological humanism ever written by a geog rapher. Subtitled 'Nature and culture in Western thought from ancient times to the end of the eighteenth century', Glacken's book was decisive proof that the progress and prestige of geography need not be embarrassed by hu manistic scholarship.

Not long after Wright delivered his address, there appeared in geography harbingers of a new mode of inquiry. This was quantitative, nomothetic, highly for malized, and concerned above all else with scientific explanation of spatial location and covariance, as well as spatial structure, interaction, and organization. These spatial scientists assertively promoted a new, 'positivist' geography, and in so doing frightened a large number of regional, historical, and cultural geographers, as well as some old style economic geographers, into self conscious awareness of their identity as methodological humanists. Thus, a regional geographer like Robert S. Platt was by 1957 drawing a methodological division between geog raphers who undertook "scientific analysis" and geog raphers who undertook "artistic synthesis" using "humanistic tools of rational inference and intuitive judgment." The same methodological division was ac knowledged in J. M. Broek's *Geography: It's Scope and Spirit.* At about the same time Gordon Lewthwaite and O. H. K. Spate distinguished humanistic geography from the new, quantitative geography, objecting that the later, due to its nomothetic aspirations, obscured the significance of the idiographic and the radical indeterminacy occasioned by human freedom.

Some of these antipositivists were older men, but others like D. W. Meinig, John Frasier Hart, Cole Harris, and Wilbur Zelinsky were rising scholars whose best work was yet to come. It would not be correct to describe the somewhat motley assembly that gathered under the banner of humanistic geography at this time as anti positivist, if positivism is taken simply as science by the hypothetico deductive method. What they opposed was complete usurpation of the field of geography by posi tivists waving the red flag of what Schaefer, their hero, had called "methodological monism." Geography was, therefore, well before 1970 split between what C. P. Snow had called "two cultures." An impartial spectator at the time would likely have bet on the eventual victory of the positivists, who were young, coordinated, and zealous; but a change in the zeitgeist was about to favor humanistic geography for close to 20 years.

The New Humanistic Geography, 1970–89

James Parsons was among the first to recognize that the social unrest of the late 1960s would bring much needed reinforcements to the ranks of the humanistic geog raphers. "The generation we are teaching," he wrote from the University of California, Berkeley, in 1969, "gives indication of being more concerned with esthetics, human values, and something they call 'lifestyle' than with social 'laws' or mechanistic analysis of make believe or model worlds." Thus, a "humanistic approach" "may prove more congenial to the coming generation," and "the slide rule and the computer may well be loosing ground to the heart" in the struggle for the soul of geography.

Students of the sort Parsons describes did eventually mature and come to dominate humanistic geography, but through the 1970s most prominent proponents of the approach were middle aged men, not long haired rebels. Wilbur Zelinsky is an outstanding example of this gen eration, and in the fiercely polemical essays that he wrote between 1970 and 1975 we find many of the major themes of the new humanistic movement. The first of these themes is Zelinsky's millenarianism. The end of the world seemed to be at hand, the new horsemen of the apocalypse being militarism, atomic weapons, environ mental degradation, and the 'population explosion'. Ac cording to Zelinsky, each of these perils issued from the logic of civilization itself, a logic that had for thousands of years demanded incessant growth in human popu lation, power, and productivity. The world that the agricultural revolution had brought into being was now about to collapse under the weight of these endless engrossments. Grim as this prospect was, however, the time for despair was not at hand: for the death of civil ization opened the way for a new age that would differ from the disastrous order of civilization no less than civilization had differed from the prehistoric world of hunter gatherers.

Zelinsky's second theme was, therefore, utopianism. Beyond the time of troubles in which he unquestionably lived, Zelinsky descried a new, desirable, and attainable world in which the contradictions of civilization were resolved and the crimes of civilization in some sense redeemed. Following the economist Kenneth Boulding, he interpreted the violence and upheavals of the twen tieth century as part of the "Great Transition" from civilization to post civilization. Boulding had seen post civilization in conventional utopian terms, as an era of peace, stable populations, economic and social justice, and environmental conservation. Likewise, Zelinsky envisioned a "Great Leap Forward from our current malaise and confusion into the promised land of post industrial harmony and plenty." And not only would the objective problems of politics and economics be resolved in this new Age of Aquarius; men and women would be emancipated from the self alienation engendered by the repressive modern superego, and thereby freed to "get on with the serious business of becoming completely human and explore the full range of our emotional, moral, in tellectual, and spiritual capacities."

Zelinsky's third theme was that this Great Leap would be accomplished, not by science or revolutionary politics as conventionally understood, but by "radical reappraisal of our value systems," a "radical reordering of priorities," and a "radically altered vision of the world." In his 1975 presidential address to the Association of American Geographers, Zelinsky argued that science, while not without value, was but one of the sources of "insights into the nature of the real world." And it manifestly delivered much that humanity did not desire, while failing to de liver much that humanity did desire. Taken alone, for instance, "the spiritual vacuity of the scientific creed" seemed to ratify only nihilism and existential despair. Science was, moreover, vitiated by philosophical, meth odological, and institutional defects that were especially damaging when an "outmoded mechanistic natural sci ence model" was applied to questions of social science.

This potent mix of foreboding and hope animated the new humanistic geography, and removed from it the appearance of a stance that was, at bottom, reactionary and obstructionist. Humanistic geography was no longer in the position of a threatened minority defensively pleading for epistemological pluralism: it now assailed positivism with accusations of complicity, if not collusion, in producing the general crisis. It not only affirmed topical and methodological humanism as models of inquiry that deserved toleration because they were un likely to bring geography into disrepute; it trumpeted them as proper and effective instruments of philanthropic humanism.

Positivist science, according to the new humanistic geographers, dehumanized humans by representing them as passive agents of social, psychological, and economic forces. These forces, according to science, caused humans to behave in predictable ways, and a progressive and rational society would seek to control human behavior indirectly, through modulation of these forces. Against this view humanistic geographers raised an alternative image of humans as deliberately acting on the basis of reasons that they considered in the light of their own intentions, interests, and values. A progressive and ra tional society (and social science) would, therefore, rec ognize and respect the dignity of the individual actor, and induce change by offering the individual reasons to undertake what Zelinsky described as a "radical re ordering of priorities" and a "radical reappraisal of our value systems."

The concept of a reason has three characteristics that would greatly affect the new humanistic geography. First of all, a reason comes to us as a symbol, a signifier as sociated in our minds with some signified. If a bull chases us across a field and, eventually, catches and tosses us over the fence with his horns, our forward motion has two explanations. So long as we are running, we are running from a reason, a symbol of danger. Once the bull has hooked us with his horns, we are mechanically caused to fly through the air. Zelinsky and the other humanistic geographers believed that humans were as often running away from bulls as they were tossed by horns. Second, there is in our response to a reason a presumption of freedom and choice. As the room grows warm we are caused to sweat, and we are given a reason, which we may not under the circumstances judge to be an adequate reason, to remove our jacket. Humanistic geographers insisted that humans are free, decision making beings, in part because this fit more closely to the first person perspective of the humanities, but also because it gave them reason to hope that people could choose to change. Finally, if we wish to make sense of the behavior of a human responding to reasons, we must understand how those reasons appear to that person. Here, the inquiry is not a scientific inquiry into causal explanation, but an interpretive or hermeneutic inquiry that seeks what philosopher Wilhelm Dilthey (1833–1911) and soci ologist Max Weber (1864–1920) had called *Verstehen*, understanding. Thus Zelinsky ended his presidential address calling upon geographers to set aside their sci entific prejudices and attempt "unconventional modes of attaining genuine understanding." And this is precisely what a growing number of geographers began to do.

In their attempt to understand the reasons for human behavior, the new humanistic geography continued in the geosophical tradition that J. K. Wright had begun; but more than a decade of positivism had left geographers far more methodologically self conscious than they had been before. The humanistic geographers therefore turned to philosophy for arguments to justify and enlarge their humanistic intuitions. Most were drawn to a species of speculation that philosopher Alvin Plantinga calls creative antirealism. Whatever their differences, the neo Kantian philosophers that fall under this heading agree that humans do not simply see the real world as it is, but rather participate in the creation or construction of the world by their sensory, cognitive, and noetic activity. Most humanistic geographers were drawn to continental phenomenology, although many mixed this with ideal ism, pragmatism, existentialism, and the sociological theory of symbolic interaction. Geographers' readings of these philosophers were often idiosyncratic, and some times arguably erroneous, but their labors yielded some general concepts, ideas, and methods of lasting import ance for the field. Below the article discusses three es pecially important contributions of the new humanistic geography, which roughly align with the three aspects of the concept of the reason given above: the concept of life world, the idea of human agency and action, and the method of interpretation.

Life-World and Place

In 1976 Anne Buttimer and Nicholas Entrikin separately introduced geographers to the concept of life world (the German *Lebenswelt* and French *habitus*). They understood life world in two ways. First, a life world is the world as it appears to a person unreflectively engaged in a par ticular form or way of life; it is the world as it is inter preted, given meaning and value, by what German sociologists called a *relativ natürlich Weltanschauung* (a worldview natural to that way of life). As Buttimer put it, life world is "the prereflective, taken for granted di mensions of experience, the unquestioned meanings, and routinized determinants of behavior." A hiker who in tends to reach a certain campsite before nightfall, and who has an interest in doing so with the least possible effort, will, for instance, see in the terrain before him meanings that are in fact reasons to take one trail rather than another. Unlike the geomorphologist, he is not concerned with explaining the ridges and valleys. He is concerned with understanding whether the nature of the trail before him is a reason to turn left or right, walk fast or slow, and feel anxious or assured. This sort of fully interpreted life world has sometimes been criticized as alien to the true meaning of phenomenology, but it has nevertheless had the greatest impact on humanistic geography and its successors.

Drawing on Edmund Husserl, Buttimer and Entrikin also presented life world as a means to criticize positivist science. The way that the world actually appears to an impartial observer, so the argument goes, is nothing like the world reported as objective and real by empirical science. One does not, for instance, see 'the Sun'. One encounters, rather, a bright and burning disk, and then, if one has learned to do so, conceptualizes or theorizes that this phenomenon is something called the Sun. The phenomenal world is not, in other words, or at least not originally, the array of objects imagined by physical sci ence; nor is it the world spontaneously interpreted by a form of life. To that most impartial of observers, the infant, it is, rather, what William James in *The Principles of Psychology* described as "one great blooming, buzzing confusion" that assails the infant's "eyes, ears, nose, skin, and entrails at once . . . coalesced together in one and the same space." Taken empirically, as 'pure sensation', the universe appears as 'one confused fact'. The upshot of this for critics of positivist science is that scientists do not begin with experience or objective reality; they begin with a representation that is, as deconstructionists like to say, 'always already' theorized and interpreted. Or as Alfred North Whitehead put it in *The Organization of*

Thought, "neither common sense nor science can proceed without departing from the strict consideration of what is actual in experience."

The apparent difference between these two views of the phenomenological life world has often confused geographers. By phenomenological life world do we mean the meaningful, interpreted world of the first description, or the "one great blooming, buzzing confusion" of the second? The answer is that we mean both, for the former is simply a development of the latter. As William James stated it, making the point of greatest importance for humanistic geographers, "infants must go through a long education of the eye and ear before they can perceive the realities which adults perceive." This long education comes through experience and the imposition of culture by the child's family, political community, and peers. In this education the growing child learns 'ideas which the sensation suggests', or 'farther facts associated with the object of the sensation'. Thus, by the time she reaches maturity, her 'every perception is an acquired perception'. The humanistic geographer Yi fu Tuan made much the same point throughout his long and prolific career, and stated the matter succinctly in a 1979 essay. "Landscape [i.e. interpreted lifeworld] is not a given, a piece of reality that is simply there," he wrote. "What is given is an environment to which we respond in automatic and subconscious ways" that we have learned through habit, indoctrination, and experience. "Landscape," on the other hand, "appears to us through an effort of the imagination" and is "an achievement of the mature mind."

The movement from environment to landscape, from the 'one confused fact' of infant sensation to the temporally ordered, spatially extensive, objectively discrete life world of adult perception, is, of course, complex as human life itself. But perhaps the most basic aspect of this movement is carving the one confused fact into discrete categories of things connected to concepts. This is why phenomenology is also called analysis of concepts: for what appears to a mature mind is a world ordered according to concepts. Some of these concepts denote natural kinds, objects we group together because they share the same nature. Cats, for instance, are a natural kind. Other concepts denote practical kinds or functional categories, objects we group together because they serve the same purpose. Pets do not have a common nature, for instance, but a common relation to human beings, and so are a functional category or practical kind.

Practical kinds or functional categories are the most interesting to humanistic geographers (some of whom dispute the existence of natural kinds). This is because the functional categories that order our life world originate in what phenomenologists call our intentions, literally our aims. We know this because the appearance of our life world changes when our aims change. To give a simple example, when we play hide and seek with our small children, our intention is to hide, or find one who is hidden. Consequently we 'see' in the back yard instances of a kind of location that will serve as a hiding place. These hiding places are 'intentional objects', which is to say objects constituted by our state of mind when we form the intention to hide. As such they are not 'given', which is to say apparent as phenomena, to persons who do not aim to hide or to discover another person in hiding.

Intentions not only divide the world, they also evaluate the world. There are, after all, good hiding places and poor hiding places. They also cause the world to appear to a human as reasons, which is to say as possible motives to action. To stay with the above simple example: in the context of a game of hide and seek, the absence of bushes from this side of the house is a reason for us to run round to the other side.

Our most important intentions work in much the same way to order an environment into a landscape, to carve our sensations into functional categories, assign value, and find in the resulting interpretation of the world reasons to act, feel, and think. We are less likely to be conscious of this *relativ natürlich Weltanschauung*, or of the constructed quality of the life world it makes apparent, because it endures for prolonged periods, and so comes to seem absolutely natural. The intention of the child is to play, for instance, and so for years and years the world appears to her largely as divided into the playworthy, the unplayworthy, and the play forbidding. As this intention is replaced in adolescence by, say, an intention to compete and succeed in the world, the landscape and life world of the child disappears. Parenthood brings new intentions, and an altogether new life world in which partly revived functional categories of childhood are supplemented by functional categories unique to parenthood, such as dangerous things and dangerous places, both of which are made apparent by the state of mind known as anxiety and the intention to protect. The disappearance of long established functional categories and life world at the end of a protracted project, such as raising children, a marriage, or a career, is ordinarily experienced as a 'crisis' and is sometimes aptly described with the phrase, "my world fell to pieces."

If we can find significantly different life worlds in the different stages of an individual life, we must expect to find even greater differences in the life worlds of persons engaged in highly dissimilar forms of life. Thus, when humanistic geographers undertook fieldwork, a major goal was to understand the worldview and life world of their subjects. This is why the humanistic geographer John Western made such extensive use of interviews in his books *Outcast Capetown* and *Passage to England*: he hoped to discover how the world appeared to his subjects. Others used participant observation, taking up the roles and daily practices and intentions of their subjects. Sev eral humanistic geographers, for instance, worked as street vendors in South American cities in the hope of obtaining first hand experience of what it was like to view the world through the eyes of a poor street vendor.

The term 'place' very soon replaced the term 'life world' among humanistic geographers, place being more clearly geographical, and life world, perhaps, bringing with it too much philosophical baggage. But place was for humanistic geographers a technical term with a meaning that was closer to the meaning of life world than to lo cation. A 'place' was, rather, a location rendered mean ingful by human imagination.

A location is rendered meaningful when it appears to say something in addition to itself. It says something in addition to itself because it is filled with symbols, or perhaps is itself a symbol, and it is the nature of a symbol always to points to something other than itself. A stone is, for example, only a stone; but a tombstone is a symbol that points beyond itself, to a man, a memory, a life, and an idea. This is the way with every symbol, for the function of a symbol is to take something that is absent and make it present. It may be absent because in conspicuous, intangible, imaginary, or far removed in time and space. Humans represent absent things sym bolically to make themselves mindful of those things, and thereby give themselves motives and reasons to act, think, or feel in a particular way. A location therefore becomes, for me, a place when I see in it something more than the objects of which it is composed, when these objects ap pear to point beyond themselves, perhaps to an in conspicuous hazard, an intangible boundary, a popular fantasy, or an event that happened long ago or far away. It becomes a place as its buildings and streets are trans formed from mere buildings and streets into reminders of people, events, emotions, and acts.

Humanistic geographers studied this concept of place in two ways. The first asked what it was like to be inside an intelligible place, to have, as was often said, a 'sense of place'. To have a sense of a place is simply to find it intelligible: to understand what the symbols point to and what it is 'like' to be in among them. But the slightly different phrase 'sense of place' implies, on top of this, an existential affinity with the place, a feeling that, in 'this' place, one truly is 'in place' or where one belongs. Geographers such as Yi fu Tuan, Edward Relph, David Seamon, and Robert Mugerauer wrote extensively on this topic. It is, perhaps, best understood as the way that geographers have framed classic questions of topical humanism about modern alienation (what Relph called placelessness), our ambivalent feelings about home (which Tuan described as an oscillation between cosmos and hearth), and the appalling loneliness to which humans are subject in the face of a seemingly mute and unsympathetic cosmos.

The second way humanistic geographers studied place looked not at the subjective experience of place, but at the symbols themselves, and the messages they conveyed. Humanistic geographers who asked how the symbols worked turned first to semiotics and the sociological theory of symbolic interactionism. Kenneth Foote's *Sha dowed Ground*, for instance, examined landscape symbols that point back to deeply disturbing acts of gruesome violence, and the decisions communities make whether or not to install such symbols, whether or not to sustain such a memory. Patrick McGreevy's *Imagining Niagara* described how Victorian tourists projected their pre occupations onto the waterfall and transformed it into a symbol pointing beyond the here and now to primeval nature, the afterlife, or a future of limitless productive power. Finally, in *The Past Is a Foreign Country* David Lowenthal exhaustively documents symbols that point to the past, the uses to which they are put, and the various degrees to which they imply that the past is radically discontinuous with the present.

Action and Agency

Science describes the universe as a great, unbroken chain of causes and effects, each cause following on its effect with a very high rate of probability, or even with absolute certainty. In such a universe there appears to be no room for freedom of action, no possibility that a human being could cause something to happen without having been caused to do so. A human might transform an impulse, as when, say, hunger is transformed into riot; but such transformations are predictable, automatic, perhaps de terministic, and not fundamentally different from the transformation of sunlight into sugar by photosynthesis. In the eyes of science, in other words, our behavior is our genetically determined response to a perhaps contingent environment, possibly as modified by our education (culture) and experience of operant conditioning.

This description of human behavior was well estab lished in human geography from its inception. Around the turn of the nineteenth century Darwinian environ mentalists taught a severe form, in which human behavior was not even trained through positive reinforcement. Behavior was simply selected by an environment that eliminated every individual who was physically, psychic ally, socially, or technically unfit. Carl Sauer's famous and influential criticism of Darwinian environmentalism merely allowed that humans respond to the positive re inforcement of successful behavior, and imitate the be havior of other, successful humans. Guided by economics, spatial science and location analysis described humans as passive utility maximizers whose decisions are limited to calculation of the most efficient means to satisfy unchosen ends. This automatism was taken to its logical extreme in the spatial mechanics of social physics.

It was this sort of scientific geographic explanation that Anne Buttimer had in mind when she expostulated with human geographers to describe in their work something "more than the *danse macabre* of materialistically motivated robots." Her statement is hyperbolic, since all geographers recognize that humans possess a capacity for change that robots do not, but correct about unchanging materialistic motivation. The capacity for change that these geographers recognized was a capacity for technical change—a change in how utilitarian goals of comfort and security were realized. Conforming to David Hume's 'subjectivist' or 'noncognitivist' account of practical reason, they assumed that human goals remained constant. Each of us is motivated solely by our environmentally triggered passions, and exercise reason and choice only to calculate and choose the most efficient means to these unchosen ends. This sort of instrumental rationality was not what humanistic geographers meant by 'getting on with the serious business of becoming completely human'.

David Ley and James Duncan were, perhaps, the humanistic geographers most acutely critical of this tendency to describe individual humans as "agents" blindly executing the orders of some supra individual or superorganic entity. The environmentalists had made individuals mere agents of their 'race'. Sauerian geographers moved directive power to 'culture', and attributed to this hypostatized abstraction, as Duncan famously argued, a superorganic will that animated all individuals who bore that culture. Regional geographers and spatial scientists, likewise, depicted individuals as instruments of 'economic forces'. In the 1970s structural Marxism arose as a new sort of philosophic holism in geography, and like its predecessors, structural Marxism reduced humans to passive agents of reified theoretical entities like capitalism, the mode of production, or the profit motive.

In an important, controversial, but highly influential 1981 paper, Duncan and Ley exposed the logical and empirical fallacies of structural Marxism and, by implication, all other social science that pretended to explain change as the result of actions by theoretical structures. They did not dispute the existence of structures, but insisted (as argued in *Humanistic Geography: Problems and Prospects*, which Ley edited with Marwyn Samuels in 1978) that these were properly understood as the "context" within which 'humans' acted. A structure such as social class may forbid some acts, and encourage others, but is itself inert and does not act. To say that it does is akin to saying, not that I walk to the bottom of the hill, but that the hill walks me to the bottom. Duncan and Ley concluded that geographers ought to conduct their research much more in the manner of historians, with far more attention to individuals, context, choice, and contingency, and far less deference to the logical demands of theoretical and historical necessity. They were not proposing abandonment of theory, as we can see in their empirical studies of gentrification and elite suburbs, but insisting that theory should be used to illuminate facts, and facts should not be used only to illustrate theory.

In Duncan and Ley's account, as in all of the voluminous literature that followed (especially important work by Derek Gregory), human agency is opposed to "structure" (a confusing locution, since structures direct "agents"). This way of thinking about agency was useful at the time, for many geographers were then invoking structures; but in our 'post structural' present it may well obscure the central questions of human 'action'. As we have seen, a structure was something theorized as a component in a macro historical process, itself theorized as an 'engine of history'. All talk of structures therefore implied historicism, the idea that history is governed by some fundamental 'logic' or plan. Hegelianism, social Darwinism, Marxism, Comtean positivism: all of these are structural and historicist–secular reworkings, in fact, of the Christian concept of 'providence'. If we suppose, as these theories do, that history is in the hands of God, or *Geist*, or the mode of production, or natural selection, or progress, it will make little difference if any of us chooses to act. Talk of human action arose, in part, because geographers lost faith in the automatic operation of macro historical processes, or what Jean François Lyotard called grand historical metanarratives. But it also arose for two additional reasons that remain important to humanistic geography.

First, humanistic geographers recognized that structural, historicist theories encourage fatalism and political passivity. Although such theories permit social scientists to predict, or imagine they can predict, future events, just like natural scientists, they also affront human dignity by seeing the human future, individual or collective, purely under the aspect of prediction. Yet human beings do not normally, or properly, limit themselves to prediction of their future behavior, or simply offering evidence in support of these predictions. Humans, unless utterly demoralized, express intentions, which are in fact choices about what future 'they' (individually or as a group) will work to bring about. And they justify these intentions with reasons, not evidence. Structural, historicist theories are dehumanizing, and antihumanist, because, when popularized, they discourage action. To the political right this has the undesirable effect of undermining personal responsibility; to the political left it has the undesirable effect of discouraging political activism.

The second reason humanistic geographers wished to represent human as actors who deliberately carry out intentions is that this permitted them to ask ethical questions, for moral judgments are possible only when it is granted that a human has the dignity of acting. An act is a special kind of behavior or event. It is the kind of event that appears properly attributable, not to some prior event, but to a person. If it is not a natural kind, it is surely a functional kind that 'appears' to almost anyone not rendered insensible by structural thinking. If we say, "He hit me!" we assign responsibility for the blow to a person, and in that assignment of responsibility define the blow as an act. We also judge the blow unjustified and demand that the agent answer for his action. The man we accuse might answer, "I'm sorry, *I* did not hit you on purpose. My hand hit you because my wife screamed and I turned suddenly." With this answer he refuses responsibility, describes the blow as predictable behavior caused by environmental factors (the scream) and the centrifugal force of a twisting body. The blow was not an act. Or, imagine him accepting responsibility, and in so doing agreeing to my description of the blow as an act. He might say, "Yes, I did hit you," and would then justify his act, not explain it. He would offer a reason, let us say: "I saw you ogling my wife and thought a punch in the eye would make you stop." Notice that his reason consists of a motive (my ogling), his choice to respond to this motive, his intention (to make me stop), and a calculation that a blow would be warranted and effective. I might then, of course, protest that I had not been ogling his wife with lustful longing, but only staring with concerned alarm because a large spider….

To take this picaresque scenario one step further, imagine that the jealous and pugnacious husband, upon consulting with his wife, discovers that a spider had indeed invaded her undergarments. Being just and honest (as well as temperamental), he would thereupon recant his reason, retract his accusation, and accept my reason for my action. Now recognizing and acknowledging his blow as an unjust injury, for which 'he' is responsible, he might offer me the symbolic restitution of a conciliatory drink. Accepting this atonement, I might shake his hand.

Such is the character of interpersonal relations between beings who act. This sort of morality (honesty) and sense of justice make no sense unless events are taken out from under the description of causality and behavior and placed under the description of intentionality and action. When a scientist asks "Why?," he wishes to learn the cause, and he finds this in some prior event, often in the 'environment'. When a person asks another person "Why?," more often than not she wishes to know the reason. Humanistic geography recognized that interpersonal relations require us to place some human behavior under the description of intentionality, not causation, as it is only under this description that a vast range of functional categories appear: categories like person, responsibility, blame, praise, shame, indignation, answerability, reasons, motives, and intentions. And only under this description are we capable of feeling in any meaningful way social emotions like gratitude, love, hate, contempt, or even equanimity. Until we recognize the functional category of acts, in other words, we cannot begin to understand humans or the human life world.

Interpretation

When we say that we understand an action, we mean that we see its end, or purpose, as well as why the actor, intending that end, has judged the means she is employing likely to succeed. When we say that we do not understand an action, we may mean that we do not understand the end intended – we do not 'see the point' – or that we do not see why the actor expects the means she has chosen to succeed. Unless the actor is utterly deranged, and perhaps even then, our inability to understand is, in other words, an inability to imaginatively enter into the life world of the actor. For to understand another person is quite literally to 'stand' in their place and see the world from their point of view.

Any point of view is what we have already called a *relativ natürlich Weltanschauung*, or a worldview natural to a form of life, and, as we have seen, a worldview carves the world up into the life world (*lebenswelt*, *habitus*) of that form of life. This point of view, or worldview, is an interpretation, a way of conceptually organizing what Yi fu Tuan called "environment." As we saw in the discussion of life world, every child comes to understand the world by seeing it from the point of view of her educators and peers. Thus, if we say, "I now understand that the world is a dangerous place," we really mean "I now stand in the position (actually, or vicariously through imagination) of a being for whom the world is dangerous." This sort of interpretation we might call primary interpretation (the poet and critic Samuel Taylor Coleridge (1772–34) called it primary imagination). Primary interpretation makes things (intentional objects) by carving environment into functional categories, producing a life world. (Some humanistic geographers describe this construction of reality as 'representation', others call it 'imagination'.) Efforts to vicariously enter into the life world of another person and imagine what the world looks like from their point of view, can be called secondary interpretation.

Humanistic geographers have sometimes reflected on their own primary imagination. Autobiographical work such as Yi fu Tuan's *Who Am I?*, Richard Francaviglia's *Believing in Place*, and Edmund Bunkse's *Geography and the Art of Life* are good examples. But humanistic geographers far more frequently employed secondary interpretation (or imagination) to enter the life worlds of other people.

The first thing to observe about secondary interpretation is that it is not an esoteric skill. As social creatures, humans naturally seek to understand other humans; and as psychologist Michael Tomasello has argued, their ability to understand is the basis of human culture. Children at a very young age begin to interpret the behavior of other humans by imaginatively entering into the intention that guides the action. My baby girl, for instance, does not simply observe me doing something, such as picking up toys and putting them in a basket; she observes me trying to accomplish something, in this case trying to clean up the room. As Michael Tomasello observes, we know this because she doesn't simply mimic my behavior; she tries to help me reach the end I have in mind (a clean room), which she has interpreted from my behavior (which she correctly describes as an action). Throughout her life she will continue to interpret the external signs of other people's mental states and intentions. This is why, although one may speak of the invention of the scientific method, and say with truth that there are peoples without science, one cannot say the same of the interpretive method.

Interpretation that follows formal rules is hermeneutics, a method of understanding first developed to permit systematic exegesis of the Bible and later extended to cover other verbal and nonverbal artifacts. Humanistic geography developed in the hermeneutic tradition begun by Friedrich Schleiermacher (1768 1834). Unlike earlier hermeneutics, where the goal had been to discover the objective meaning of words in ancient texts (written in dead languages and dialects), Schleiermacher's hermeneutics said that interpreters must seek to understand the context in which the word was originally written and had its original sense. As Wilhelm Dilthey put it in 1900, understanding is the "process by which an inside is conferred on a complex of external sensory signs," this "inside" being "that psychic reality" of which "the signs given to our senses" are "the expression."

This idea of secondary interpretation as a means to understand, or vicariously inhabit, an alien life world is at the heart of the work of the great culture historian, Jacob Burckhardt (1818–97); of Dilthey's *Geisteswissenschaften*, or science of the spirit; and of Weber's interpretive sociology. Humanistic geographers followed these thinkers in the belief that scholarship consists of what the anthropologist Clifford Geertz famously described as "our constructions of other people's constructions."

There are two reasons why humanistic scholars believe that we should exercise secondary interpretation and attempt imaginative entry into the life world of alien people. The first is that understanding is a prerequisite to explaining 'human' behavior. Because humans act 'with something in mind', it behooves the scholar to find out just what it is they have in mind. (Some scientists of course believe behavior can be more than adequately explained without reference to unobservable entities like 'belief' and 'mind'.) Among humanistic geographers, Leonard Guelke may have made this point most clearly in his argument for 'idealist' explanations of rational human behavior. Guelke's idealism was epistemological, not ontological, for his idealist approach proposed a method to explain human actions, and to choose from among all possible interpretations of an act the interpretation that is most probably correct. Guelke proposed that we have explained an action "when the thought behind the action has been understood." And this thought behind the action has been understood when we see "the sense of an action," when we grasp "the intention of the agent" and "the theoretical ideas that were employed by the agent in the interpretation of his situation." This is secondary interpretation, as described above.

But this alone is not enough, for any discrete action will admit to multiple interpretations. When confronted with more than one interpretation of an act, Guelke stated that we should choose the one that is most plausible because most consistent with the general character and known overall intentions of the actor. Unless there is strong objective evidence indicating that the actor, in this instance, acted out of character or contrary to his normal interests, the interpretation consistent with 'a large number of actions' is to be preferred. For example, in a 1988 article Wilbur Zelinsky interpreted the welcome signs that stand beside the highway on the outskirts of so many American small towns as gestures intended to affirm the individuality and significance of the town in the face of homogenizing mass society. He then argued that this interpretation was plausible because consistent with other contemporary legible artifacts, such tee shirts, pins, and bumper stickers, all of which made sense as expressions of fear and resentment in the face of dehumanizing anonymity.

The second reason humanistic scholars believe that one should exercise secondary interpretation, and attempt imaginative entry into the life world of alien people, is that this is a humanizing exercise that quite literally 'broadens the mind'. The basic notion is this. Any particular person represents a possible way of being human, one possible primary interpretation of what it means to be human. This interpretation may, in unusual cases, be unique to the individual, but is in most cases characteristic of the individual's period, place, and social role. In any such interpretation certain human possibilities, of activity, character, emotion, and intellect, are realized in a very high degree, while others are atrophied or absent. Thus, every way of human being (every period, every culture, etc.) has its excellences and its defects. Some are energetic, some contemplative, some courageous, some benevolent, etc. By vicariously entering into the life world of another person, particularly another person from another period, place, or social role, an individual becomes, as Dilthey put it, more 'rounded', more conscious of his or her full 'humanity'. Continuing with the theme of humanistic education as a means to soften the excesses and supply the defects of one's own age (place, social role, personality), Dilthey continued:

"inasmuch as the exegete [interpreter] projects his own sense of life into another historical milieu, he is able within that perspective to strengthen and emphasize certain spiritual processes in himself and to minimize others." The classic defense of a humanities education is this broadening of the mind by imaginative entry into alien life worlds, by way of history, literature, or humanistic anthropology and geography. Such an education is also called liberal because it promises to liberate the individual from the narrow confines of one life world, and grant her something closer to the full range of human experience.

Dilthey stated that we approach the inside of an alien life world by way of 'external sensory signs' that are 'given to our senses'. These signs are, like any symbols, pointers, in this case pointing to an invisible 'psychic reality'. Humanistic geographers first found such signs in the cultural landscape. As Pierce Lewis stated it in a famous attempt at a landscape hermeneutic, written in 1979, everything in the human landscape has 'cultural meaning' (culture here equivalent to *relativ natürlich Weltanschauung*). Landscape is, indeed, "our unwitting autobiography, reflecting our tastes, our values, our aspirations, and even our fears in tangible, visible form." To interpret a landscape was imaginatively to enter this psychic reality, this life world of tastes, values, aspirations, and fears.

Humanistic geographers also interpreted representations of landscapes and places that they found in literature, paintings, and other media, once again seeking imaginative entry into the world of the author or artist who produced the work. As Douglas Pocock expressed it in 1981, "the novelist has the gift of articulating our own articulations," and of deepening our "insight into place." They are particularly skilled, not so much in vivid world paintings of the visible scene, as in evocation of place in 'experiential terms' such as 'security, stability, and belonging' and 'boredom, drudgery, and entrapment'.

Humanistic Geography Today

The purpose of the humanities is to aid us in what has been called secondary interpretation, so that we may become more fully human by imagining what it is like to see the world through alien eyes. The humanities are our means to know what it is like to be someone other than ourselves, to look out upon the world from a time, place, or social position other than our own. When this is the purpose of geographic teaching and writing, geography is properly a humanity, or humanistic geography.

Such an exercise is worth our effort if we presume that humans are possessed of a certain incorrigible dignity, since it is this dignity that makes the endlessly varied life worlds that humans have experienced worthy of imaginative re experience. The dignity of humanity makes this sort of secondary interpretation intrinsically good, whether or not is good for anything else. Understanding of this sort is not a powerful instrument like scientific knowledge. Being able to imagine (however imperfectly) what it was like to be a Franciscan Friar in the thirteenth century no doubt remedies a defect in most modern personalities, and is conducive to a more 'rounded' human being, but it does not help one to manipulate genetic material, computer code, election returns, public opinion, or anything else that people wish to manipulate in the world today. Humanistic study is not, in other words, a means to power; and as any student of humanity knows, the appetite for power is much stronger than the appetite for understanding.

This is why humanistic geography will normally be a minor theme in the field, secondary to instrumental geographies that work, philanthropically or misanthropically, to enhance the power of some segment of humanity. And humanistic geography itself will, for this same reason, tend to devolve to mere social science, theorizing, explaining, and manipulating the life worlds that the humanistic geographer hopes only to understand. By the end of the 1990s the interpretive methods of humanistic geography began to loose ground to the instrumental rationality of applied geography and geoinformatics. Humanistic geography itself was touched by this will to (political) power, and evolved into the 'new cultural geography'. The legacy of humanistic geography remains, nevertheless, everywhere apparent, in a widespread recognition of the importance of meaning, the individual, the particular, and interpretive methodologies. And there are more than a few humanistic geographers at work today, as can be seen in books such as *Textures of Place: Exploring Humanist Geographies*.

Post-Humanism? Post-Humanistic Geography?

Humanism places humanity in the position an earlier age (and contemporary theists) reserve for God. It asserts that humans should praise, study, and serve humanity, not God. Christian critics of secular humanism have always condemned this apotheosis as idolatry (however much they valued and served human beings on other grounds), but with rare exception the major currents of Western through for the past two centuries have honored and glorified the human spirit, or labored to ameliorate the human condition. Recently, however, there has been talk of post humanism. As is the case with most titles that begin with the prefix 'post', post humanism is not a unified movement but a set of movements that repudiate humanism, albeit for different reasons and without agreeing on much else. Three such movements are especially important.

The first might be called neo Augustinian thought, for like the bishop of Hippo it has a very low opinion of the natural moral character of human beings. Pointing to the horrific barbarity of the twentieth century, as well as contemporary examples, it observes that there is nothing particularly praiseworthy or dignified in these exhibitions of 'humanity' and 'the human spirit'. Humanism, they assert, is a delusion of the liberal's imagination. Humans are worse than half beasts; they are also, evidently, half devils.

Deep ecology and radical environmentalism offer a different reason to repudiate humanism. They condemn humanism for grossly exaggerating the difference between humans and other species, and for privileging human needs, to the detriment of other species and the environment as a whole. Just as humanists replaced the theocentric universe with an anthropocentric universe, so these environmentalists propose to replace the anthropocentric universe with an ecocentric universe.

Antiessentialism is the third major position from which humanism is today criticized. This is because humanism sought to overcome the particularities of creed and culture with an appeal to universal human values. Antiessentialists condemn this cosmopolitanism as cultural imperialism, and dispute the premise of a universal human nature on which it rests. The concept of the human is, for these critics, something imposed on the autonomous and self creating members of our protean species as a discipline, and not some essence or remainder that persists after all accidents of history and culture have been removed.

Obviously, the character of any post humanistic geography will depend on the reason the author of that geography has repudiated humanism. Neo Augustinian geography would, presumably, jettison all talk of moral progress as historicist myth, and peer pessimistically into a bleak future where human half devils prey upon one another with ever more lethal technology. Environmentalist geography of a post humanist sort will have to rationalize abandonment of amelioration of the human condition as the overriding goal of human inquiry, and develop misanthropic doctrines to undermine anthropocentrism. Antiessentialists are already at work in geography, but have ahead of them the formidable challenge of explaining why, if difference is truly radical, anyone is obliged to respect or tolerate that difference (how, in other words, if there is no human there can be a human right). Post humanism is not flattering or attractive, but this should not surprise us: it is hardly pleasant to be pulled from the throne of God just as we were growing accustomed to it.

See also: Affect; Christian Geography; Cosgrove, D.; Cultural Studies and Human Geography; Culture; Diaries (Video, Audio or Written); Dwelling; Emotional Knowing; Existentialism/Existential Geography; Film; Habitus; Home; Jackson, P.; Landscape; Landscape Iconography; Landscape Perception; Literature; Meinig, D.; Memorials and Monuments; Memory; Movies and Films, Analysis of; Oral History; Participant Observation; Philosophy and Human Geography; Place; Poverty; Semiotics; Space I; Structurationist Geography; Subjectivity; Symbolic Interactionism; Text and Textuality; Tuan, Y.-F.; Urban Representation/Imagination; Wright, J. K.

Further Reading

Adams, P. C., Hoelscher, S. and Till, K. E. (eds.) (2001). *Textures of Place: Exploring Humanist Geographies*. Minneapolis, MN: University of Minnesota Press.

Buttimer, A. (1991). Geography, humanism, and global concern. *Annals of the Association of American Geographers* 80(1), 1 33.

Buttimer, A. (1993). *Geography and the Human Spirit*. Baltimore, MD: The Johns Hopkins University Press.

Cosgrove, D. and Daniels, S. (eds.) (1988). *The Iconography of Landscape: Essays on the Symbolic Representation, Design, and Use of Past Environments*. New York: Cambridge University Press.

Cresswell, T. (2004). *Place: A Short Introduction*. Malden, MA: Blackwell.

Guelke, L. (1982). *Historical Understanding in Geography*. Cambridge: Cambridge University Press.

Harris, C. (1971). Theory and synthesis in historical geography. *Canadian Geographer* 15, 157 172.

Ley, D. and Samuels, M. (eds.) (1978). *Humanistic Geography: Prospects and Problems*. Chicago, IL: Maaroufa Press.

Livingstone, D. N. and Harrison, R. T. (1981). Immanuel Kant, subjectivism, and human geography: A preliminary investigation. *Transactions of the Institute of British Geographers* 6, 374 395.

Meinig, D. W. (ed.) (1979). *The Interpretation of Ordinary Landscapes: Geographical Essays*. New York: Oxford University Press.

Meinig, D. W. (1983). Geography as art. *Transactions of the Institute of British Geographers* 8, 314 328.

Mugerauer, R. (1995). *Interpreting Environments: Tradition, Deconstruction, Hermeneutics*. Austin, TX: University of Texas Press.

Pocock, D. C. D. (1981). *Humanistic Geography and Literature*. Totowa, NJ: Barnes and Noble.

Tuan, Y. (1977). *Space and Place: The Perspective of Experience*. Minneapolis, MN: University of Minnesota Press.

Western, J. (1992). *A Passage to England: Barbadian Londoners Speak of Home*. Minneapolis, MN: University of Minnesota Press.

Wright, J. K. (1966). *Human Nature in Geography*. Cambridge, MA: Harvard University Press.

Zelinsky, W. (1970). Beyond the exponentials: The role of geography in the great transition. *Economic Geography* 46, 498 535.

Zelinsky, W. (1975). The demigod's dilemma. *Annals of the Association of American Geographers* 65, 123 143.

Human-Nonhuman

E. J. Roe, Southampton University, Southampton, UK

Glossary

Assemblages The entanglement of relations that knit together the human and nonhuman.
Hybridity It is to recognize that entities straddle modernist divisions of nature and society.
Nonhuman It is everything we know to be not human.
Ontology It is a term philosophers use to consider the nature of reality.
Post-Structural Theory The theories that critique a structural understanding of society, and the categories that were established by structural thought.
Relational Ontology It is understanding the world as produced through a state of relations.

Introducing the Human-Nonhuman in 'Human' Geography

Human nonhuman: an awkward term but in its 'hyphen' contains the theoretical and empirical challenge to think about how humans are dependent in numerous ways upon the group of 'things' we know to be 'not' human or nonhuman. So this term is not about an opposition between human and nonhuman but rather about human and nonhuman relations. What is the relationship between human and nonhumans like? In thinking about this relationship we start to recognize that complex connections between humans and nonhumans exist; these connections necessitate a need to recognize the formation of hybrid entities, for these connections are fundamental to how an entity is known and what is done with it. Hybrid entities in this human geography context straddle any division between human and nonhuman, like the politically significant entity produced when a human is driving a car, the ethically significant entity produced when an African elephant lives in a British zoo, and the elephant is no longer 'in the wild' but living in a human fashioned world. These hybrids are working collectives, or the term we will use – assemblages. Human nonhuman assemblages co constitute each other. Consequently, these human nonhuman assemblages indicate a critical rethinking of the category 'human'.

In human geography the study of 'human nonhuman' attends to how living with nonhumans configures the human, but equally about how living with humans configures the nonhuman. Nonhumans include every nonhuman 'thing' in the spectrum from nature through to technology – from carrots, chickens, and trees to prosthetic limbs, autopoetic processes (self creating processes), and microbes. The nonhuman category is thus very broad, which means the potential for conceiving of human nonhuman assemblages is equally broad and expansive. The term 'assemblages' is shorthand for the entity conceived when one takes seriously the nonhuman challenge to the category of human. For what results is an assemblage of human and nonhuman things that collectively perform a role. Underpinning the studies of human nonhuman assemblages is relational thinking, or what can be termed a relational ontology. Ontology is a term philosophers use to consider the nature of reality; here it is used to understand the world as the production of a set of relations. Thus, the critical analysis of the human within a relational ontology goes as far as to recognize the agency of nonhumans in the world, challenging human supremacy and relinquishing human autonomy, and consequently human rationality as a range of nonhuman actors, such as mobile phones, adverts, and white line road markings appear intrinsic for how we communicate with our family and friends; seduce us to buy one top designer brand over another; and order our driving behavior to the extent that our embeddedness within this world of nonhuman also defines who we are and how we perform ourselves. The world without these nonhuman objects and signs would be quite different, arguably could never exist; and yet for a long time, the significance to the relationships humans make with them has been overlooked, leaving them out of social analysis.

Authors within human geography have drawn upon a range of philosophical thinkers to help them think about human relations with nonhuman things. A particularly popular approach has been to conceive the politics of how realities are made by working off from a non hierarchical, thus flattened world of socionatural messiness, as opposed to studying things in themselves or the social construction of nature. This flattened world puts nonhumans and humans on the same plane, that is, not conceiving humans as dominating nonhumans or vice versa. Importantly, this relational thinking has consequences for how we understand the process of knowledge creation and activities that go with this. The interest is in how, if one assumes no hierarchy, one can aware of how through practice relations humans and nonhumans become 'made and stabilized'; it offers a way to explain how 'hierarchies' become guidelines about how to act in particular contexts. Human nonhuman assemblages co produce knowledge of the world within all spheres of life from the scientific to the everyday; in which case there is always more to know and understand about the world, but that process of human knowledge creation moves with humans', animals', and insects' investigative poten tial and inert nonhumans changing abilities to facilitate this process as human techniques and nonhuman cap acities co evolve.

Human geography's prime interest in studying human nonhuman assemblages is driven by three con cerns: first, attention to human/environment relations; second, to human/technical relations, including that of the biotechnological; and third (which relates to the other two), post structuralist theorizations of nature/society relations that breakdown the modernist division between the stuff of nature and the stuff of society and associated divisions, such as the mind and the body. The article discusses each of these areas in further detail.

Geography has historically always been attentive to human/environment relations, both in physical geog raphy and human geography. Until recently, the Berkeley School of cultural geography is perhaps the exception to Western geography's exclusion of the study of nonhu mans in human geography. Carl Sauer's 1927 work on the cultural production of landscapes, positioned among the disciplines of archaeology, anthropology, and traditional cultural geography, was interested in the domestication of plants and animals. Although these studies adopted a very different tack to the study of nonhumans than the one on which this article is focused upon. In short, the current attention to human nonhuman can be seen as a consequence of a three stage shift of focus to Western geographical studies of relationships humans make with the natural environment, from initially, one based on the study of humans in different environments to secondly, one of a study of the human impact on the environment to thirdly, one of a study of how human's co constitute environments (more involved than physical geography's interest in interactions). Environments consist of non human things such as trees, tower blocks, and ants, and processes such as soil and water erosion, ageing bodies, and metabolism. So, for example, if we consider livestock or food animals as nonhumans, the question has shifted away from how do humans farm food animals in different places? The question instead becomes how do humans breed food animals (or in other words genetically select along a technological spectrum of increasing genetic precision) to be suitable for different environments within commercial requirements?

This question leads us to consider how this question overlaps with another geographical interest pertinent to the study of human nonhuman assemblages, the study of humans' relationship with technology, including the biotechnological. Sociologists of science and technology have been particularly influential in turning to human nonhuman assemblages as explanatory devices. Arguably, the most influential have been the actor network theo rists (see section below for further details) who have argued that the capacity for a human to perform a role, for example, like being an academic, is reliant upon being part of a network of a range of human and nonhuman actors. If you took away colleagues, office, computer, telephone, books, and desk, an academic wouldn't be an academic writing papers, delivering lectures, and pro ducing 'knowledge'. In the current age, our dependence on the mobile phone is perhaps a reminder that what we do and how we feel and communicate is mediated through technology, which appears increasingly embed ded into how we know others and ourselves. Also, there are those whose mobility is facilitated through having a prosthetic limb or whose life can be understood as a consequence of a human egg being fertilized *in vitro*, where being a human nonhuman is perhaps more starkly apparent.

The third area of geographical concern in human nonhuman is in the development of post structuralist theorizations of the stuff of nature and the stuff of society which works to break down modernist binary divisions, strongly influenced by the Enlightenment values that underpinned what was known by placing them in strict nature and society categories. Margaret Fitzsimmon's argument for the 'matter of nature' to be taken seriously has been very influential on the various directions non human geographies of nature have taken onboard and developed post structural thinking. Her work initiated greater analytic attention to the material qualities of the nonhuman, and the consequent material relations and material connectivities generated with humans, and has since become one of the more popular approaches for understanding human nonhuman assemblages.

Some materialists taking what has been termed a neo Marxist perspective have approached nature as (re) constituted by humans, consequently casting nonhumans as social nature. This has allowed them to differentiate the category of nonhumans into those modernity would categorize as natural, such as plants and animals, and those deemed never to have been of nature, such as computers. However, this approach perhaps runs into problems as technological advances produce unfathom able assemblages that seem so unnatural, yet also re semble the 'natural' at some level: for example, Petri dish grown meat – is it natural or unnatural? Maybe it de pends upon the context in which you are being shown the meat. Defining this meat, as social nature, does not help to explain the implications of what has been created; so how can it be named and categorized?

It is these types of problems that have driven another group of materialists to argue that the term 'nature' has become analytically outdated as it is appreciated that what is nature is culturally specific and this has always been so. The term 'naturecultures' has been developed in recognition of this insight. This group has grappled with the 'materiality' of nonhumans to conceive of human nonhuman material connectivities in what has been termed a 'post natural' world; this has included a revival of biogeography as a subdiscipline that potentially could engage both cultural human and physical geographies. Yet, a criticism of this group is perhaps that it delves and pokes its way through layers of materialities and assem blages, and yet will only declare specificities. For it stepping back, leaving the event, and presenting a general picture will always be an impossibility – the world is too messy for reporting such imprecision. Without doubt for both neo Marxist writers and naturecultures writers, the contribution of a symmetrical relation between the nat ural and the social, by actor network theorists, has held much promise in reconceiving the world in nonhuman, materialist terms, and has offered thought provoking analytical takes on human nonhuman assemblages.

Key Theorists of the Human-Nonhuman

The following section summarizes some important the orists for human geography's study of human nonhuman assemblages.

Sociologists of science, Bruno Latour, Michel Callon, and John Law, developed actor network theory (ANT), a set of methodological habits, a toolkit rather than a theory, with the intention to write the world anew by rewriting the 'constitution' of Western knowledge. For geographers, this theory has been particularly interesting because it breaks down the modernist distinctions be tween humans and nonhumans, and nature and culture, through conceiving materially heterogeneous associ ations between human and nonhuman entities to produce a network constituted through the links rather than the nodes of its making. Agency in the world is operated through the durability of these actor networks found in the links, the associations rather than the nodes. The durability of these actor networks is strengthened in the circulation of 'immutable mobiles' – devices, types of people, animals, money, and so on, which make these networks last. ANT has enabled geographers to study the nonhuman without falling back on its representations, instead by studying the practices that mobilize it, that enact it. For actor network theorists', space is conceived topologically as opposed to topographically, or in other words, geometric space is collapsed and folded by the agency of technology. Technologies such as mobile phones, televisions, and satellites bring the previously distant places near; consequently, a topological map would look very different from a topographical map.

Science studies provoked much of the development of ANT. Bruno Latour begun by discussing the work of Louis Pasteur to show how scientists build heterogeneous (the opposite of homogenous) networks made up of different entities and resources that combine in a way to spread scientific facts and artifacts. Latour indicated how Pasteur became the spokesperson for a crowd of invisible, dangerous forces able to strike anywhere – namely, the microbes – through interpreting their behavior and knowing how to control them. He goes on to consider in his book *We have never been modern* the impact of scientific methods and how they have rationalized and purified how we understand the world through the dominance of science, consequently separating culture from nature and humans from other things (see **Figure 1**). Indeed, Latour observed that alongside these purifying processes is the creation of hybrids where science and policy, and nature and culture are mixed together; the presence of hybrids, make it increasingly difficult to keep up the purifying demands of modernity. He moves to look toward how we can accept the connections between nature and culture in his book *The Politics of Nature*, where he explains how to bring the sciences into democracy, or in other words for the 'authority' of science to be challenged by arguing for cultural specificity, for the hybrid entities produced, and thus for the crisis of objectivity that ensues in scientific practice. He sees this as a way for social scientists to be better equipped for critically engaging with scientific practice, and more to the point asserting the view 'we have never been modern', through indicating how, behind the scientific purification machine, there are many con nections between humans and nonhumans that illustrate a range of cultures within scientific practice.

These ideas have been extremely influential in sup porting the development of ideas in geography around human nonhuman, initiating studies of scientific practice, and a range of other research areas such as food and agriculture to acknowledge an entanglement between nature and society. The clumsy and imprecise attempts to diagnose bovine tuberculosis in cattle is one such ex ample of a nonhuman geography where the playfulness of nonhuman materialities in the form of the *Myco bacterium bovis* virus continues to object to scientists' at tempts to develop a conclusive test for the disease; not forgetting the governments, industry, farmers, cattle, and badgers who are held in a dance of trade restrictions, compensation payments, mass culls, and increased public concern in the meantime.

Donna Haraways's conceptualization of human non human is located within embodied, human subjectivities that lead her to ask 'what counts as human?': this is a question examined in her work on the cyborg and on companion species. Notably, as a feminist science writer, she engages more fully with the bodily dimensions than Bruno Latour; she infuses her reading of technoscience – her term that politicizes the techniques of making science – with an eye for how the borders of bodies are being challenged. For example, Donna Haraway's cyborg is a cybernetic organism, a fusion of the organic and the

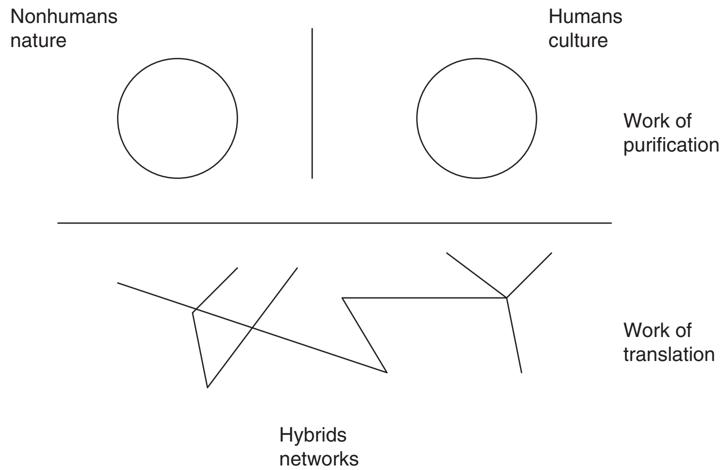

Figure 1 Latour's 'purification and translation'. Redrawn from Latour, B. (1993). *We Have Never Been Modern* (trans. Porter, C.), 11p. Cambridge: Harvard University Press.

machinic forged within particular historical and cultural practices, for example, working on a computer or talking to a friend through a mobile phone. What emerges from Haraway's work is a close working of the pathways that enmesh humans and nonhumans together through the technobiological, boundaries are poorly defined, and this challenges the assumptions of what is a subject and what is an object. She has a vision of connection for everything. More recently, she has turned her attention to companion species, namely, the dog, as she has conceived of a fleshy, biological entanglement of connections between human and nonhuman animals. This point moves the discussion on to consider other authors who have been influential in how to understand embodied human nonhumans in less scientific and rather more everyday engagements.

von Uexküll's '*umwelt*' still continues to be a resource for conceiving ethologically the relations between lively, animate nonhumans and the environment, or more specifically the world of the animal. Reacting against the mechanistic biology of the day, von Uexküll argued that to treat the animal as a mere assemblage of sensory and motor organs is to leave out the subject who uses these organs as tools, respectively, of perception and action. The animal is conceived as a living being in active relations with their lifeworld; or in other words, each living thing inhabits sensorially a world constituted by the specific life of that animal. Tim Ingold has elaborated von Uexküll's work to apply this thinking to studies of human worlds, which following this line of thinking become human nonhuman assemblages, as humans through sensation and movement perceive and act in relation to what is meaningful around them. Consequently, Ingold's environment is alive, interactive, and world making; embodied life forms with specific capacities enmesh with features of the environment to form specific worlds. This work has given geography the opportunity to explore how the specificities of embodiment in particular places enact subjectivities in particular ways; a variety of embodied performances equaling multiple worlds can be witnessed as you move between the activities of people, animals, plants, and insects in a given space. Each body is in the process of making and responding to what is significant for it in that place. von Uexküll and Ingold are not so interested in materiality but more focused on sensate, moving bodies. In contrast, there are other writers who have focused on materiality as the metaphysical bind between humans and nonhumans. This work appreciates materiality as the stuff that animates and enchants, dismissing the purification and rationalization processes of science, like the materialities affected and affecting the *M. Bovis* virus which still evades rational scientistic techniques for capturing it.

The political theorist Jane Bennett, one of these writers, writes of a matter that has a liveliness, resilience, unpredictability, or recalcitrance that in itself is a source of wonder for us. This enchanted materialism finds its ancient counterpart in the atomism of Epicurus and Lucretius. Lucretian physics described a universe of atoms forever falling in a void of space; these primordial materialities, at unpredictable times and in unpredictable directions, 'swerve' slightly from their downward paths and bump into each other. The Lucretian 'swerve' or *clinamen*, in contemporary writing, stands as a metaphysical figure for the unruly, surprising, exciting unpredictability of the world of materialities. This rainfall of atoms weaves unusual connections and molds unusual events in and among humans and nonhumans. Michel Serres offers new prepositions and motifs like – the angel, the parasite – that conjure up the liminal spacings between human and nonhuman – to grasp these complex configurations and alliances between humans and nonhumans. Jane Bennett's work gives us the opportunity to become familiar again with a nonmodern world, a world which bursts with creativity and unlikely alliances through the chance of the swerve.

Jane Bennett draws on the philosophies of Gilles Deleuze and Felix Guattari to sketch out the processes that draw humans and nonhumans to make new futures, new becomings. The vitalist philosophy of Deleuze and Guattari has offered not only a way to grasp the processes that bind humans and nonhumans together, but also the significance of spatiality and temporality to the mode of interaction that they conceptualize as a 'becoming'. Their concept of 'becoming' facilitates an engagement with this world ever in a state of flux, flows, and insubstantial borders. These flimsy, less tenuous, and unstable connections between things in the world and humans become more accessible through their philosophy; attention can then be paid to how human nonhuman assemblages appear to have a degree of permanence, for example, a building. For, in fact, a building is always on a trajectory of becoming derelict as overtime a building materially decays from changing weather conditions and day to day use, but human intervention in the form of constant repair and maintenance enables an appearance of stability and constancy, by intervening in the tendency to decay. In a different expression of becoming, human gestures can take on a 'becoming animal' as the sensibility of a human improvises dog friendliness by letting a dog lick their face. Lastly, Deleuze and Guattari's contribution of the rhizome metaphor is an antidote to the arboreal enlightenment figure. The rhizome is not as strictly flattening as the Latourian 'flat ontology', yet still it does not support a vertiginous hierarchy but instead emphasizing horizontal depths between humans and humans, nonhumans and nonhumans, and nonhumans and humans, thereby offering a way to giving a rich texture to what lies between these entities. One way to explain this is: the rhizome like a potato has nodules growing out on all sides that enables it to spread rapidly horizontally, not vertically (like the enlightenment 'tree' motif).

The Human-Nonhuman in Recent Geographical Writing

Both human geographical and physical geographical research programs often straddle the modernist worlds of nature and society. The animal and other nonhuman geography research that has emerged lately has worked off from these critical intellectual post structuralist impulses (discussed above) and the broader shifts in the social and political landscape associated with postmodernity. A consistent theme through this work is the greater consideration for the ethics and politics of the treatment of the human nonhuman in everyday, regulatory, and commercial contexts. The autonomous human self in various contexts becomes increasingly tenuous as human nonhuman assemblages or entities take on new political and ethical meaning. Indeed, the new realities, or constellations, of humans and nonhumans force new ways of thinking in a range of spheres of everyday life. Geographers have offered theoretical approaches for reconfiguring politics and ethics in the entanglement of human, environment, and animal relations to help understand and support these challenges. With this has come the innovation of methodologies to support the study of the nonhuman, with the aim to remain faithful to a relational world.

Geographers have made a significant contribution to the theories previously discussed by subjecting them to spatial analysis. Jonathan Murdoch outlined the implications for geography's conceptualization of space and time if one studied space through drawing upon ANT. For him, space and spacings become a networked achievement through heterogeneous associations between humans and nonhumans, whether technology or life forms. Geography thus becomes the study of associations or networks since ANT is less interested in time and space, but more in the unique acts of 'spacing' and 'timing', acts that are conducted through associations of various kinds. The challenge for geographers becomes to not only take a relational view of space but also that spatial relations are seen as network relations. Within the network domain, spatial scale is reconceptualized as a network achievement, while networks are reconceptualized as 'heterogeneous engineering' – that is, processes of network building in which entities of various kinds are assembled in ways that allow networks to undertake certain functions. Geographical analysis responds through taking on board the following methodological points: follow the actors as they stitch networks together, observe what is linked to what, and assess how power flows up and down from centers to peripheries and back again.

Another geographer, Sarah Whatmore, in her book *Hybrid Geographies* fleshes out precisely what the 'ethical political' might mean in a world of actor networks. Criticizing anthropocentric traditions of political reasoning, she argues for how prions, cows, microchips, prosthetics, growth hormones, and elephants and other 'actants' might figure as political subjects. In effect, she seeks to recast bordered distinctions as the contingent outcome of networks of relations among the heterogeneous actors of a more than human world. Moving beyond a Latourian ANT framing, she draws upon conceptual resources including, of those already discussed, the biophilosophical resources of Deleuze and Guattari and Donna Haraway to try to energize through bodily material fashioning human nonhuman assemblages. Through this conceptual work, 'lively geographies' are mapped out that bring organic life into human geographical study at such a level that a new meaning of the term ‘biogeography’ starts to find new currency in geographical thought. Bio geography reaches into the realm of physical geography study and indicates how the subject of physical geography and human geography can overlap, if not quite in meth odological terms.

Bruce Braun, David Demeritt, and Noel Castree’s work seeks to understand nature as inescapably and in trinsically social. These authors write extensively about the epistemological conundrum of the term ‘nature’, but give much less attention to the human nonhuman, as their studies have focused more on human relations with social nature, whether in terms of temperate rainforests, climate, or following Marxist nature as a resource. Interestingly, human nonhuman has been mobilized conceptually more by geographical writers who have faced methodological and ontological crises as they have tried to study non human geographies. When faced with one’s own glaring humanness, it has become a way to engage more satisfy ingly with nonhumans, by recognizing that we humans are always with them.

The conclusion mentions some examples of empirical studies in which theories of human nonhuman have been pursued in response to methodological challenges. The nonhumans discussed below do not reside only in what would be known as social nature, but equally include nonhuman technology such as cars.

Geographical studies of nonhuman animals have ranged through different types of animals, from do mesticated pets to animals reared for commercial reasons to animals living in the world. Chris Philo and Chris Wilbert’s book *Animal Spaces, Beastly Places* seeks to treat animals as another social group. This has been ap proached in various ways through a study of the social practices and meanings which shape relationships with animals, the wider social concern about the treatment of animals as new political movements have developed, and the inclusion of animals within the wider set of social actors in the practical business of everyday social life. Animal geography has become another force to critically open up the human in human geography for critical scrutiny.

Jamie Lorimer’s work on the charisma of nonhuman animals (from the stag beetle to the Corncrake bird) configures in different ways the cultural practices of modern conservation networks. Steve Hinchliffe asks what is the space for nonhumans? What is the nonhuman space? His argument for an ecologized politics (as op posed to a politicized ecology which he has argued has representational credentials) provides scope for a cos mopolitics, which moves away from an undifferentiated flow between humans and nonhumans, and instead seeks to look for entanglements between humans and nonhu mans to make the nonhuman multiple in an ecologized politics of difference. He examines how the water vole is enacted into the world through vole writing – how we know the water vole – and the struggle for the world of water vole to be included politically in conservation debates. This is part of a larger argument about cosmo politics. Cosmopolitics means directly letting in the dynamism of political subjectivities rather than having organizational bodies do the representing for them. This requires an ability to listen to the vagueness of the epi stemic thing (whether the water vole or the *M. Bovis* virus) in science and politics: Hinchliffe following phi losopher, Isabelle Stengers, argues that this is what makes good science. Through listening to the nonhuman as colleagues in the process of producing knowledge, new knowledges are made possible. If we do this, Hinchliffe argues, it becomes possible to learn how things matter to humans and nonhumans by working with an ontological politics that enables them to object to our ways of knowing them, to intervene in our processes as much as we intervene in their processes.

Nick Bingham’s work focuses on the plurality of nonhumans. He calls on us to focus on being bodies that are in the midst of things like bees, butterflies, and bac teria, and that humans and nonhumans collectively have a capacity to be threatened by events like genetic modifi cation. He criticizes those who recognize entities as linked, and instead calls for conceiving human and non human beings as open to and with others. For Bingham, the human nonhuman becomes a motif for a cosmopoli tical position of being with others and asking questions like ‘how many are we?’

Moving onto non animal nonhumans, other geog raphers are engaging more closely with the material and the bodily in apprehending human nonhuman assem blages. Emma Roe has explored human nonhuman in the context of embodied food consumption practices. Like Sarah Whatmore, by engaging in the material flows that move through human and nonhuman bodies in the agro food network, she has found a way to engage with the metabolic, intimacy that is intrinsic to shared, corporeal material flows. This has methodologically led to a focus on embodied practices rather than discourse analysis, and to grasp the spacings and timings of the material flows of nonhuman foodstuff as it is prepared, consumed, and digested by humans. Russell Hitching reveals the agency of plants in the garden as forces of enchantment, bringing about a shift to a different kind of gardening practice and ethical attachment to nature.

Moving toward the technological, Gail Davies has developed hybrid geographies as a concept for investi gating and intervening in the cultures of nature, exper tise, and ethical debate, which accompany changing practices in the life sciences. Her work on xeno transplantation has illustrated how human nonhumans entities, produced by scientific knowledges, provoke a range of responses from humor to alarm in the general public. She argues that these lay or nonexpert knowl edges engage with the implications of emerging bio technologies at a sophisticated and complex level. How ever, human nonhuman entities in contemporary times do not necessarily take such immediately alarming forms. Pete Merriman has shown how different nonhuman media and technical devices articulate particular programs of governance through the agency they hold within human nonhuman networks. Cars, the contents of the Highway Code, and road markings and signs are just some non humans with whom humans are very frequently forming assemblages. Today, humans have become so used to forming human nonhuman assemblages with cars that there is no humor or horror about the prospect of car travel anymore.

Prospects

This overview of how the human nonhuman has become a significant area of study for human geographers indi cates its wider relevance both to understandings of human/environment relations and human/technological relations, and approaches for theorizing what happens when the borders between the natural and the social are removed and human nonhuman assemblages loom into view. As commented at the outset, tackling the category of the human nonhuman assemblage is not only a way for thinking critically about studying the nonhuman but also equally challenges ways of studying humans, thus it continually demands theoretical and empirical con sideration in a number of research arenas, and not only those in human geography but also equally for physical geographers and the wider realm of natural science and engineering. Indeed, examples are increasing of natural scientists acknowledging how their accounts of the 'nat ural' world can be democratized by including the methodologies and insights from human geographers specializing in human/environment relations to find a new way of doing science together. For example, flood risk modeling research is being tackled using cosmopo litical apparatus as the narratives of nonscientists who experience flooding are heard by hydrologists; the etho logical study of egg laying hen welfare problems is turning to social scientific approaches to reflect appre ciation that hen–stockperson interaction is just as rele vant as hen–hen activity in understanding how welfare problems are manifested; and there is interest in com paring humans and animals as experimental research subjects as a way to raise questions about human welfare in clinical drug trials. Both of these examples illustrate the different questions that are being asked as the borders of nature and society falter, and interdisciplinary ap proaches appear necessary to study the politics, ethics, and existence of human nonhuman assemblages, whether the nonhuman is an animal, an alien gene, a vegetable, a human organ, a robot, or an environmental process.

See also: Actor-Network Theory/Network Geographies; Animal Geographies.

Further Reading

Bennett, J. (2001). *The Enchantment of Modern Life. Attachments, Crossings, Ethics*. Chicago, IL: Princeton Press.

Bingham, N. (2006). Bees, butterflies, and bacteria: Biotechnology and the politics of nonhuman friendship. *Environment and Planning A* 38, 483 498.

Braun, B. and Castree, N. (1997). *Remaking Reality*. Routledge: London.

Castree, N. (2003). Environmental issues: Relational ontologies and hybrid politics. *Progress in Human Geography* 27, 203 211.

Davies, G. (2006). The scared and the profane: Biotechnology, rationality and public debate. *Environment and Planning A* 38(3), 423 444.

Deleuze, G. and Guattari, F. (1986). *A Thousand Plateaus: Capitalism and Schizophrenia* (Translation and foreword by Massumi, B.). London: Athlone Press.

Fitzsimmons, M. (1989). The matter of nature. *Antipode* 21, 106 120.

Haraway, D. (1996). *Simians, Cyborgs and Women: The Reinvention of Nature*. London: Free Association Books.

Haraway, D. (2003). *The Companion Species Manifesto: Dogs, People and Significant Otherness*. Chicago, IL: Prickly Paradigm Press.

Hinchliffe, S. (2007). *Geographies of Nature*. Sage: London.

Hinchliffe, S., Degen, M., Kearnes, M. and Whatmore, S. (2005). Urban wild things: A cosmopolitical experiment. *Environment and Planning D* 23, 643 658.

Ingold, T. (2000). *The Perception of the Environment: Essays in Livelihood, Dwelling and Skill*. London: Routledge.

Latour, B. (1993). *We Have Never Been Modern (trans. Porter, C.)*. Cambridge: Harvard University Press.

Laurier, E. and Philo, C. (1999). X morphising: Review essay of Bruno Latour's Aramis, or the love of technology. *Environment and Planning A* 31(6), 1047 1107.

Law, J. (2004). *After Method. Mess in Social Science Research*. London: Routledge.

Lorimer, J. (2007). Nonhuman charisma. *Environment and Planning D* 25, 911 932.

Merriman, P. (2006). 'Mirror, signal, manoeuvre': Assembling and governing the motorway driver in late fifties Britain. *The Sociological Review* 54(supplement 1), 75 92.

Murdoch, J. (2006). *Post Structuralist Geographies. A Guide to Relational Space*. London: Sage.

Philo, C. and Wilbert, C. (eds.) (2000). *Animal Spaces, Beastly Places*. London: Routledge.

Roe, E. (2006). Things becoming food and the embodied, material practices of an organic food consumer. *Sociologia Ruralis* 46, 104 121.

Whatmore, S. (1997). Dissecting the autonomous self: Hybrid cartographies for a relational ethics. *Environment and Planning D: Society and Space* 15, 37 53.

Whatmore, S. (2002). *Hybrid Geographies*. London: Sage.

Relevant Websites

http://www.lancs.ac.uk/fass/sociology/research/respapers.htm
Lancaster University, Sociology Department, Online Papers.

http://www.open.ac.uk/socialsciences/habitable cities
Open University, Faculty of Social Sciences.

Hybridity

D. P. Tolia-Kelly, Durham University, Durham, UK

Glossary

Cosmopolitanism It is a way of thinking about people and places, cultural identities, and differences without deferring to 'national' categories of cultural identity. Mobility, transnationalism, and translocalness are accepted as part of modern notions of cosmopolitan identity. Cosmopolitanism, as a theory, seeks to disrupt hegemonic 'nationalisms' within social science thinking, aiming for an internationalist or postnationalist approach to theory and research.

Culture Systems of values and practices which reflect collective ways of living. Culture reflects relations between humans and nature, economy, and land. Culture is a contested realm of reference in geography, for Marxist geographies culture reflects the values of dominant groups within society and for poststructuralists culture is unbounded temporally, geographically, and economically. It operates as a body of thought, performance, or as a singular moment or event.

Ethnicity Ethnicity is a way of thinking about cultural difference without deferring to singular notions of culture. Ethnicity is a way of framing categories of human cultures without racial reference, but with an acknowledgement of difference. Ethnicity recognizes geographical, social, and cultural difference as developed through geographical, historical, and economic contexts.

Phenotype These are different 'species' within colonial taxonomies of racial difference. Each phenotype recorded has a related system of measurements and definitions, such as shape of skull, skin tone, and sometimes psychological characteristics linked to geography.

Race It is a set of culturally signified categories of human beings differentiated by skin color, bone structure, geographical origin, and more recently genetic science. Categories of racial types have changed in human history, largely reflecting the dominant theories of human culture based in the 'West'.

Introduction

The origin of the term hybridity lies in historical scholarship on race and evolutionary theory. It has been used in identifying racial categories and used to assert differentiation between categories of human species. Biological notions of hybrid species have led to fears of racial mixing and species miscegenation (namely, the notion of a pure racial group being defiled or diluted through interracial intercourse); these fears continue in contemporary society. The assertion of race hierarchy and difference has enabled the exploitation of racial differences for social, cultural, and economic gain, within societies and within frameworks of colonial and imperial governance. The notion of 'race' has been challenged in science and cultural theory as being a culturally determined category and not a scientific one. The species of 'man' is now regarded as one species, one race, and differentiation in skin tone, skeletal, or other epidemiological factors are secondary to the acceptance of a singular category of 'human'. This biological fact however was disputed in medical text books right up until the late 1960s, again as a sign of cultural values rather than scientific truth. In contemporary society, although science and evolutionary theory are reconciled, the cultural sphere is loaded with notions of biological difference as being evidence of hierarchies of cultural and social value; a belief in the hierarchies of race is embedded in much economic, social, and cultural laws of governance in both the 'West' and 'other' regimes of rule, at various scales. Because of the endurance of racism and a belief in hierarchies of difference in the value of differently formed bodies, 'hybridity' continues to have resonance for both racist thought and the counter narratives that form part of antiracist strategies, and postcolonial politics, including those of scholar Homi Bhabha.

'Hybrid geographies' has come to incorporate a raft of new theories and practices within the social and cultural realms, beyond questions of representation and racial identities. These involve new frameworks of thinking through human–nature relationships, in landscape, geopolitics, and the evolution of geographical epistemologies. The section entitled 'Productive hybridity' outlines the ways in which Bhabha's notions of hybridity have influenced the conceptual frameworks of thinking. It also considers the subjects that have been engaged with 'hybridity' through attention to questions of overlap between nature and culture divide (including posthuman accounts).

The Roots of Hybridity

Hybridity has historical roots in the legacy of colonial science; it has had a significant effect on cultural theory and postcolonial theory; however, its meaning has shifted over time. Hybridity has been positioned within recent cultural theory as a conceptual device that seeks to disrupt and critically implode the bases upon which 'difference' has been articulated in colonial times as well as in the current political, economic, social, and cultural narratives dominant in the everyday. This article's aim is to trace the genealogy of the term hybridity and work through its value in modern social science thinking. As with much modern geography the textual bases for thinking through hybridity lay in disciplines beyond geography, including the foundations of anthropology, Darwin's evolutionary science, and cultural studies. However, the theories of hybridity since Enlightenment thinking, as of now, have been operative within the technoscientific geographical projects of the Royal Geographical Society and its militant past.

Human Categorizations of Man and Others

In material terms, the effects of the schema employed for the categorization of man and others have been manifold. In the first instance, the distinction between white from his other enabled the justification of the slave economy. The reasoning here was that if humans were not once singular species of Biblical origins, those others could not be saved. This account was in conflict with a Godly explanation of the world and the species of man and has continued to be a source of contention within scientific circles. Subsequent processes of dehumanization were faced by these other men racialized and supported by scientific theory and the technologies of measuring phenotype and cultural savagery. Theodor Waitz, a key anthropologist, claimed in 1859 that "Wherever the lower races prove useless for the service of white man, they must be abandoned … All wars of extermination, whenever the lower species are in the way of the white man, are not only excusable, but fully justifiable." Waitz's argument exemplifies the question of the singularity of the species of man.

Hybridity in the eighteenth century was a mode of conceptualizing the biological hybrid, a mixture of two human types (termed species); this is where the notion of a human hybrid emerged. The 'hybrid' test was one used to determine whether different race groups were indeed separate species of man. Cultural historian, Robert Young (1995), has argued that "The use of the term 'hybridity' to describe the offspring of humans of different races implied, by contrast, that different races were different species." In biological terms an infertile coupling between two 'races' was evidence of the incompatibility of different races of White' and Black human beings. Any evidence or anomalies to this 'hybridity' test evidenced that all the species of man was a singular group; as a result, race discourse routed in science was weakened. To overcome this, scientists acknowledged that, if there was successful reproduction over several generations, then these defined as separate 'races' or 'species' were not hybrid at all, but variations of man.

In Darwin's *Origins of Species*, published in 1859, his chapter on 'Hybridism' challenged the infertility rule as not being conclusive, thus reasserting the fact that humanity may indeed be one single species. This scientific claim supported the values and beliefs of the antislavery movement. Also central to Darwin's thesis was a notion of the mobility and dynamism of species as being at the core of evolutionary theory; species evolution rested on organisms, including humans being geographically mobile and biologically dynamic in ecological timescales. These findings challenge much of today's thinking about bounded identities located in territories that form national identities and cultures. In these realms of thought, 'hybridity' is a disturbing occurrence, biologically, socially, and culturally.

Psychoanalytical Theories of Cultural Hybridity

In the early twentieth century, theories of the hybrid in Europe and the US could no longer 'prove' any difference between 'types' or categories of race; therefore, arguments about 'difference' moved into the realms of racial intellectual and psychological characteristics. Black African writers, such as Frantz Fanon, considered the psychological and sociological consequences of scientific categories that led to circumstances of colonization and the psyche of the colonized. Both Fanon and W.E.B. Du Bois argue that the colonial stratification of humankind into the notion of species in variation effectively disrupted the psyche and psychological welfare of those classified as Black or Other. Fanon wanted acknowledgement and recognition of the deliberate alienation and suppression of self conscious autonomy of the colonized. These factors were believed by Fanon to be instrumental to the aims of racial, scientific, cultural, and political facets of the colonial project. This suppression was through a subjugation of physical, intellectual, and psychological processes of positive identification and expression. As a consequence of his theorization, Fanon advocated the celebration of the essence of other identity as a positive enfranchising and liberating sense of being. This was a political project of Blackness. Fanon fought against Black disenfranchisement from a universal human consciousness. Imperialism had ensured a rupture in the Black subjective and expressive consciousness. The 'other' was conceptualized as 'inferior', and the hybrid 'inferior' still continues to be positioned as 'lower' in pseudoscientific categories whether through biological science or cultural prejudice.

In the twenty first century, post Imperial geographies have resulted in migration of peoples across continents, thus creating new communities of sociality. At the heart of Empire, Britain, the questions of ‘hybridity’ have been relocated within the cultural realm; for writers such as Paul Gilroy and Stuart Hall, this shift represents a transposing of racist values of difference into the cultural sphere. It is here that notions of ‘essential’ cultural characteristics are linked to cultural categories, usurping those formerly based around biological categorizations. In this sphere, ‘cultural hybridity’ has been used to contemplate the merging of cultures to create new hybrid musical forms, dance, and visual cultures. Within racialized communities ‘hybridity’ has been embraced both positively and as a necessity. For Black youth experiencing racism these expressive cultures have been the source of expression of alienation from British society and a simultaneous alienation from their cultural roots in the Caribbean, Africa, and South Asia. As postcolonial bodies that represent ‘hybrid’ cultures through their language, and cultural practices, their presence reflects the productivity and difficulties of being postcolonial in a world still embroiled in notions of essential identities and national cultures.

The body of work entitled *Postcolonial Theory* emerged in the 1970s in the writings of postcolonial academics, and a body of cultural theory has emerged. The aim is to consider identity, beyond essential understandings of the concepts of self and other located in concepts of ‘East’ and ‘West’. One trope of this challenge is linked to understanding the psychological complexities experienced by the oppressed, marginalized, other rooted in the writings of Fanon, Wright, Douglas, and W.E. Dubois. The most important scholarship in the psychoanalytical theories of cultural hybridity is by Homi Bhabha. He challenges a universal notion of culture and cultural essentialism by presenting a theory based on the very particular psychosis of the ‘other’ as situated in the ‘West’ and the identity of ‘other’ within modernity, against earlier notions of delineating the ‘other’ as savage, primitive, and thus outside modernity. Bhabha in *The Location of Culture* argues that “The representation of difference must not be hastily read as a reflection of a *pre given* ethnic or cultural traits set in a fixed tablet of tradition. The social articulation of difference, from the minority perspective, is a complex, on going negotiation that seeks to authorise cultural hybridities that emerge in moments of historical transformation. The ‘right’ to signify from the periphery of authorised power and privilege does not depend on the persistence of tradition; it is resourced by the power of tradition to be reinscribed through the conditions of contingency and contradictoriness that attend upon the lives of those who are ‘in the minority’.”

Here, Bhabha defines the nature of cultural hybridities which for him are a moment of resistance to essential understandings of ‘identity’, ‘ethnicity’, ‘race’, and ‘creed’. There are three points to be made here. First, cultural hybridities are temporally situated; the ‘hybrid’ challenges a notion of a linear historical development of cultures and their structures of formation. For example, a notion of English national culture for Bhabha is one that is ‘relational’ to those deemed ‘not English’, including those cultures of the colonized under Empire within and beyond the now British Isles. Second, cultures are not fixed or in stasis, they fold through past and present (dystopian and utopian, imagined identities) and are continually dynamic. Third, Bhabha clearly states that “‘hybrid’ cultures rely on tradition as a cultural source of tension from which they form; cultural hybridity is formed and indeed made possible through a persistence of discourses of domination, power, and tradition”. Tradition for Bhabha is the articulation of the dominant narratives of identity and culture, which continues to be, in E.P. Thompson’s terms, “reinvented and reformed”. This philosophy of the means of producing culture is embedded in a notion that cultures are in continual emergence and can only emerge from the position of being in between, emergent, and formed by a ‘collective’ minority.

Understanding culture for Bhabha is achieved through a psychoanalytical process, where political movements for ethnically pure nations and identity movements are evidence of the ‘psychosis of patriotic fervor’. This analysis is made through diagnosis of psychological states of ‘folk’, and not through biological definitions through skin, phenotype and matters of land ownership, and land cultures (as within the disciplines of archaeology and anthropology). The emergence of cultures in Bhabha’s thesis is based on psychological alienation from a sense of ‘homelyness’. However, Bhabha is keen not to define, describe, or to affirm fixed delineations for cultural identities; instead, Bhabha considers the process of the enunciation of group identity as one which is reductive (temporally and spatially) and thus one which misrepresents these cultures which are continually in formation. A notion of ‘third space’ is introduced here as a space where the misrepresentation of cultural identities can occur and be engaged with, without risk of (mis)representation. The productivity of third space has been phenomenal for postcolonial writers; it posits the possibilities that free those constrained by the forces of ‘tradition’, the colonial frameworks of the West and other dominant figurations of identity. *Third Space* allows the possibility of the emergence of unalienated free expressionism of people with ‘different’ cultural identities and practices: unbounded, dynamic, and truly postcolonial. *Third Space* allows the tools of enunciation to be controlled by those representing ‘hybrid’ cultures and is not a space where definitions are imposed. By exploring *Third Space*, Bhabha hoped that we may elude the politics of polarity and emerge as unconstrained, without the constraints of taxonomies of biology and culture that had gone before. Thus, for Bhabha, 'hybridity' is a productive space for political action. Bhabha's 'hybridity' is a disruptive category, and thus offers a progressive conceptual space; an opportunity to 'be' and 'belong' within modernity – biologically, psychologically, and geographically.

The Limitations of Cultural Hybridity

Bhabha's embracing of 'hybridity' as a productive and opportunistic theory is not always supported. For postcolonial theorist such as Gayatri Spivak, it is difficult to understand cultural identity, or accept new hybrid cultural formations of those previously colonized, as the power to speak and the 'voice' of those peoples themselves has been so ruptured that now these peoples can only speak of their experience in the voice of the colonized. The inscriptions of imperialism have scarred the very processes of expressive cultures because of this legacy and regimes of subjugation. Thus, expressive cultures and their 'producers' are locked into colonial frameworks, grammars, and vocabularies of 'modern' culture; racial identity is continually employed in these. 'Hybridity' rests on notions of 'purity' and also on notions of singular bounded and recognizable categories, how can these be achieved if the psychosociological, linguistic, and the very 'structures of feeling' of a social group have been erased and reworked within the tools, grammars, and vocabularies of the colonizer?

Power is also an issue in this conceptualization. Edward Said in his thesis on *Orientalism* figures identity, knowledge, and representation as being bound up with colonial regimes of power and exploitation. These regimes of colonialism justified slavery. These 'regimes of truth' have shaped the nature of knowledge and its use in the disempowerment and subjugation not only of colonial bodies, but also of the development of psychological consciousness and creative practice.

These 'regimes of truth' fix other subjectivities, bodies, and cultural expression as something beyond (usually before) modernity, and spatially outside of the occidental space of Western Europe and North America. A consciousness of what it is to be a modern subject has evolved with these prejudicial histories and scientific discourses embedded within them. The influence of modern consciousness has extended beyond definitions of peoples, but its 'regime of truth' has ensured a 'process of subjectification'. These regimes of truth continue to be enlivened in contemporary discourses on 'first peoples', for example, and 'aboriginal' communities. The power to dissolve the links between 'racial types' and the threat of 'hybridization' continues to be differently accessible between humans identified as modern and 'other' in modern cultural taxonomies of being.

'Hybridity' and Geography

Cultural Identity

In contemporary geographical thought, cultural identity and racial theory remain important to understanding 'man' and environment relations. This postmodern theorization of subjectivity accurately describes the experience of many postcolonial peoples and has been taken up as a point from which to critique modernity and its fissured taxonomies which have framed and informed questions of humankind, national identity, and cultural expression. However, the freedom from essential identities and partial anthropological histories offers only a temporary consolation from a definition of identity as essential and bounded.

National Identity

National identity is a means by which culture is defined through these bounded, essentialized notions of 'being'. Being is linked to 'belonging' in notions of citizenship. For example, the position of Black identity in relation to belonging in the North (Australia, US, UK, and Europe) often results in a mutual exclusivity between being Black, and being English. Cultural theorist, Iain Chambers, has argued that even radical critics and historians such as Raymond Williams, E. P. Thompson, and Eric Hobsbawm have "in their appeals to the continuities of *native* traditions and experiences, perhaps inadvertently conceded the ethical and racial pretensions of a national (ist) mythology." The Black English have been considered as 'non native' to the English landscape as a result of the envisioning of England and 'Englishness' – morally, biologically, culturally, and textually being an exclusionary discourse of cultural identity. Historically, this is not limited to the nation state of England alone, historian Simon Schama has argued that "*Arcadia* as the site of protected indigenous culture, roots and biological store of authenticity is part of a longer historical tradition in England and other European states." This argument has resonance in North America, New Zealand, Europe, and Australia.

Hybridity, Time, and Nature

The existence of racialized peoples in England has been recorded since Roman times (first century AD). Despite this, however, there is a continual articulation of cultural 'difference' which locates them outside of Englishness itself, threatening to make Englishness 'hybrid' and 'defiled'. The 'newness' of their presence, their urban settlement, and their assumed racial category is dis cordant with the Englishness of the pastoral scene. These resonate with ecological theories of 'native' landscape being invaded by the 'non native'. These arguments can be traced back to the practices of Imperial governance and the cultural logic of colonial taxonomies and defin itions of aboriginal peoples of every continent. In eco logical terms, arguments are raging about which plants contribute to a notion of 'native' and 'non native' species. Central to these definitions are questions of time. At which point do circulatory and mobile plant species be come 'native'? A well known example of 'nativization' is the walnut tree, which is now defined by the Royal Horticultural Society as 'native' to Britain, but historic ally its 'origin' was the Americas, arriving in Britain in the 1700s.

Despite Darwin's thesis on the nature of nature to be dynamic and mobile, ecologists, botanists, and landscape historians have all contributed to the debates about how 'non native' species are invading, exhausting resources, and utilizing space traditionally occupied by 'native' plants and thus native cultural landscapes. Paul Gilroy describes this renewed currency of discourses of 'alien' and 'native' as evidence of the reemergence of biological essentialism that is embedded in genetic science and which feeds into cultural racism. For Gilroy, this is termed an age of the 'rebirth of biologism', which fosters biologically located evidence of race difference and no tions of purity of genetic stock. These moves ultimately serve to bolster political discourses that seek to define and reassert notions of 'native peoples' and 'native cul tural values' and thus conflate with issues of 'cultural belonging'. These arguments are troublesome, especially because of "the assumption that identity derives from deep rooted association with territory has come to be seen as the key pathology of the modern nation state; the root of its intolerance to foreign bodies. This also signifies a return to pre Darwinian hierarchies of 'species."

Cosmopolitanism

Cultural narratives that define 'native', 'non native', and 'hybrid' are contrary to our embracing of notions of 'cosmopolitanism'. There is a recognizable discordancy between our acceptances of the 'hybrid' spaces of glob alization produced through the mobility of bodies as a result of global communication networks. Nigel Clark further challenges these categories from a posthuman perspective. Reflecting on the new acceptance of 'cosmopolitanism', that is, that globalization has led to the creation of spaces and cultures that reflect the mixing, merging, and easing proximities allowable on the global scale as a result of networks of 'globalization' of humans. He has recently argued that within cultural theory there is segregation between 'nature' and 'culture'. He argues that what is important is that the social and cultural di mensions of the global process cannot be severed from the 'nonhuman' and that ultimately, we need to recognize that hybridization is endemic to the living world. We should not demonize 'non natives' as this distinction is an anthropocentric cultural classification imposed onto na tures culture. Thus, "if it is the nature of 'life' to stick to it's own 'turf' then why are so many species from the taxonomic spectrum so eager for relocation and so well disposed to it?"

The 'hybrid' for Clarke is not a threat to national purity of genetic stock or to 'native' natures, but simply a reflection of what has always occurred in nature – mo bility, hybridization, relocation, and newly emergent biological structures of species and variants – nature is hybrid in Clark's embracing of a taxonomy of species based on 'cosmopolitanism'. For theorists working in this realm, cosmopolitanism challenges the need for bounded notions of 'native' natures and 'native' citizens, disabling the racialized arguments against 'hybridization' located in racial prejudice. The call for a 'cosmopolitan turn' in the social sciences is a new way of thinking beyond 'hybridity'. For postcolonial populations and in the geog raphy of the 'occidental' world, mobility and hybridiza tion have been constant; 'cosmopolitanism' acknowledges that racial purity is a mythological notion, for man and nature. 'Cosmopolitanism' as a discourse and definition is an appropriate term for defining the political positioning of racialized diasporic populations internationally, with out being reductive to their identity – biologically, cul turally, and socially. Embedded in this conceptualization is a notion of mobility and temporal flexibility.

Against Hybridity: For 'Ecological' Thinking

Much of the literature on hybridity has been about the struggles against racialized taxonomies, and the sub sequent struggles against oppression by the racialized other, both within and without colonial space time. These taxonomies reveal the cultural and historical 'ways of seeing' the world and its peoples. As social scientists, it is important to learn the lessons of these universalist ways of thinking and to embrace new forms, which are sensi tive to 'difference' in an epoch where hybridity has been critical to postcolonial thinking. The philosopher, Lorraine Code, has argued for ecological thinking in respect of our ethical and social approaches to research with the other. In particular, it is important to acknow ledge the historical and ethnocentric nature of the epis temologies we use in social sciences research and research dissemination. In advocating responsible taxo nomies, there is an emphasis on the importance of the researcher to acknowledge the historical and ethnocentric nature of the epistemologies that we use in social sciences research and research dissemination. An approach advocated by Code, within the discourses of 'hybridity, would have long ago debunked the notion of taxonomies being linked to racial phenotypes, cultural, or national characteristics and 'native and non native' species as well as developing categories of thought which are truly against 'race'.

By using taxonomies reflexively, and in a situated framework, Bhabha's quest for 'homelyness' would be satiated; responsible taxonomies rely on acknowledging difference and ethnicity, without hierarchy. This 'ecology' of thought, within frames of space and time, delivers possibilities for engaging with ethnic differences across the globe in the terms that, Stuart Hall has long stated, are necessary for a politically enlivened and sensitive cultural theory of race. The term 'ethnicity' acknowledges the place of history, language, and culture in the construction of subjectivity and identity, as well as the fact that all discourse is placed, positioned, situated, and all knowledge is contextual. The effect of using responsible taxonomies is that there is a displacement of the discourses of the West from the Center, and puts into question its Universalist claims. This new frame relieves us of the need for thinking along notions of identity that are singular, and without geographies or indeed a merging of two distinct sociogeographical identities in the form of 'hybrids'.

See also: Darwinism (and Social Darwinism); Race.

Further Reading

Agyeman, J. (1990). Black people in a white landscape: Social and environmental Justice. *Built Environment* 16, 232 236.

Ahmed, S. (1999). She'll wake up one of these days and find she's turned into a nigger: Passing through hybridity. *Theory Culture Society* 16, 87.

Anderson, K. (2007). *Race and the Crisis of Humanism*. London: Routledge.

Anthias, F. (2001). New hybridities, old concepts: The limits of 'culture'. *Ethnic and Racial Studies* 24, 619 641.

Appadurai, A. (1996). *Modernity at Large: Cultural Dimensions of Globalisation*. Minneapolis, MN: University of Minnesota Press.

Back, L. and Solomos, J. (eds.) (2000). *Theories of Race and Racism*. London: Routledge.

Baumann, Z. (1998). *Globalisation: The Human Consequences*. Cambridge: Cambridge University Press.

Castree, N. (2003). Environmental issues: Relational ontologies and hybrid politics. *Progress in Human Geography* 27, 203 211.

Chambers, I. (1993). Narratives in nationalism: Being British. In Carter, E., Donald, J. & Squires, J. (eds.) *Space and Place: Theories of Identity and Location*, pp 145 164. London: Lawrence and Wishart.

Clark, N. (2002). The Demon seed: Bio invasion as the unsettling of environmental cosmopolitanism. *Theory, Culture and Society* 19, 101 125.

Clifford, J. (1997). *Routes: Travel and Translation in the Late Twentieth Century*. London: Harvard University Press.

Code, L. (2006). *Ecological Thinking*. Oxford: Oxford University Press.

Crosby, A. (2004). *Ecological Imperialism*. Cambridge: Cambridge University Press.

Demeritt, D. (2005). Hybrid geographies, related ontologies and situated knowledges. *Antipode* 37, 818 823.

Fanon, F. (1992). The fact of blackness. In Donald, J. & Rattasani, A. (eds.) *'Race', Culture and Difference*, pp 220 240. London: Sage.

Gilroy, P. (1987). *There Ain't No Black in the Union Jack*. London: Routledge.

Gilroy, P. (2000). *Against Race*. Cambridge, MA: Harvard University Press.

Hall, S. (1996). New Ethnicities. In Morley, D. & Chen, K. H. (eds.) *Stuart Hall: Critical Dialogues in Cultural Studies*, pp 441 449. London: Routledge.

Hutnyk, J. (2005). Hybridity. *Ethnic and Racial Studies* 28, 79 102.

Kwan, M. P. (2004). Beyond difference: From canonical geography to hybrid geographies. *Annals of the AAG* 94, 756 763.

Lemire, E. (2002). *Miscegination: Making race in America*. Pennsylvania, PA: University of Pennsylvania Press.

Livingstone, D. (1992). *The Geographical Tradition: Episodes in the History of a Contested Enterprise*. London: Blackwell.

Mitchell, K. (1997). Different diasporas and the hype of hybridity. *Environment and Planning D: Society and Space* 15, 532 553.

Nash, C. (2002). Genealogical identities. *Environment and Planning D: Society and Space* 20, 27 52.

Nash, C. (2003). Cultural geography: Anti racist geographies. *Progress in Human Geography* 27, 637 648.

Nava, M. (2002). Cosmopolitan modernity: Everyday imaginaries and the register of difference. *Theory, Culture and Society* 19, 81 99.

Swyngedouw, E. (1999). Modernity and hybridity: Nature, Regeneracionismo, and the production of the Spanish waterscape, 1890 1930. *Annals of the AAG* 89, 443 469.

Turner, B. (2002). Cosmopolitan virtue, globalization and patriotism. *Theory, Culture and Society* 4, 45 63.

Werbner, P. (2001). The limits of cultural hybridity: On ritual monsters, poetic license and contested postcolonial purifications. *Journal of the Royal Anthropological Institute* 7, 133 152.

Whatmore, S. (2002). *Hybrid Geographies: Natures, Cultures, Spaces*. London: Sage.

Young, R. J. C. (1995). *Colonial Desire: Hybridity in Theory, Culture and Race*. London: Routledge.

Young, R. J. C. (2007). *The Idea of English Ethnicity*. London: Blackwell.

Hypothesis Testing

R. Haines-Young, University of Nottingham, Nottingham, UK
R. Fish, University of Exeter, Exeter, UK

Glossary

Asymmetry between Proof and Refutation The proposition that while no number of observations that are consistent with a theory or hypothesis prove that the theory is true, one counter observation can show the theory is false.

Epistemology The methods by which we are able to 'gain' knowledge about something; epistemology is the study of the principles and practices of enquiry.

Hypothesis A logical consequence of a theory that can be tested in some way.

Ideographic The term used to describe the research tradition in geography that emphasized the study of areal differentiation and the uniqueness of place, as in regional geography.

Nomothetic The term used to describe the research tradition in geography that emphasized the need to search for regularities and generalities that are expressed in theories that link geographical phenomena.

Null Hypothesis A device used in statistical hypothesis testing. The null hypothesis is set up as a competing hypothesis to the one the research seeks to explore. The null hypothesis usually asserts that there is no difference between observed and expected outcomes in some sampling process.

Ontology The basis upon which we recognize something as knowledge; ontology is the study of what kinds of things exist and therefore what kinds of things can be known about.

Theory A set of beliefs or assumptions that we hold about the structure of the world and the way it behaves.

Type I and Type II Errors The distinction made in statistics between the mistakes made by rejecting a hypothesis when it is true, compared to accepting a hypothesis when it is false.

Introduction

To test a hypothesis is one of the most important things a researcher can do. Indeed so important is it that many would claim that it is the thing that distinguishes scientific research from other forms of academic inquiry. Certainly hypothesis testing is one of the ways that human geographers have sought to understand their discipline as a science. Perhaps not surprisingly then, it is a technique that is not without controversy, for its use registers with wider debates about the nature and purposes of geographical knowledge. This article discusses the logical framework in which all types of hypothesis testing within human geography can be viewed and, through this, reviews some of the statistical tools available to those working with quantitative data. We then inspect some of the wider philosophical and practical complexities surrounding the process of hypothesis testing in geographical research, and what these issues share with more qualitative approaches to the production of geographical knowledge.

Historical Context

Hypothesis testing emerged in the 1950s as part of debates about the scientific bases of geographical research. To be understood as a scientific enterprise, it was claimed that geography should create knowledge applicable to a range of conditions and circumstances. At that time, 'anglophone geography' had been wedded to the idea of regional analysis, a process of inquiry that emphasized areal differentiation and uniqueness. For some geographers, one of the consequences of this interest in specificity – in idiographic study – was that geography was weak when it came to the generation of models about how the world in general operated. Although the rich accounts of regional character and distinctiveness were interesting, it was argued that the explanatory potential of geography should rest on uncovering regularities and communalities in how social processes came to function across space. In contrast, hypothesis testing emerged as part of an approach to geographical research expressly designed to inform, and be informed by, theory; one that would seek to put its work on a more scientific footing through a concern with generalization – or nomothetic study.

The process by which this argument about generalization was accepted by many geographers and with it the development of hypothesis testing as a key feature of human geography, has been termed the quantitative revolution. This transformation in approach was the basis for understanding geographical endeavor as a spatial science. The development was one whose possibilities and limits have since been subject to great critical comment by human geographers. Hypothesis testing is more than simply a research 'technique' in this respect. It is part of the changing history of ideas concerning what geographical knowledge should look like, and the claims we should seek to make of it.

Hypothesis Testing: The Logical Framework

Put simply, to test a hypothesis is to confront a theory with some kind of data. Theories and hypotheses are often regarded the same kind of thing – they represent our beliefs or guesses about the way the world works. However, it is useful to make a distinction between them. It is a difference that is clearly brought out through the act of testing.

To confront a theory with data means to take some logical consequence of an idea and try to find out whether it is supported by the observations that we make about the real world. A logical consequence is something that must be the case if the theory is true. For example, it is widely accepted that human activities are fundamentally affecting the atmospheric system, and that this will lead to long term changes in climate. If this is indeed the case, then a number of things follow logically from this theory. Detectable changes to the concentration of atmospheric carbon should track growth in use of fossil fuels by people. Future human action to reduce our carbon footprint should result in a reversal of the warming trend. These logical consequences are best regarded as hypotheses. They are specific propositions that we can in principle check out by collecting some evidence. This checking process is what hypothesis testing is all about.

As the twentieth century philosopher Karl Popper pointed out, the important thing to note about the situation we face when evidence conflicts with our hypothesis is that, providing we judge the evidence to be correct, we must reject the hypothesis and the theory from which it was derived. Of course, evidence can be ambiguous or difficult to interpret, and so we must not be too quick to refute the theory. Nevertheless, if we ultimately feel that the data are reliable, then the hypothesis must be considered wrong. As researchers we are not prepared to accept theories and data about the world that conflict with each other, and still assert that we understand or have explained something. Finding ourselves in this situation we must redesign our theories so that different logical consequences follow which are supported by the data. Alternatively, we must try to think up altogether different theories to explain the world around us which are better supported by the evidence.

Popper placed particular emphasis on the act of conjecturing theories or hypotheses and then attempting to refute them because of what he called the "asymmetry between proof and refutation." This is the logical principle that while no number of observations consistent with a theory or hypothesis can prove it to be true, just one counter observation can show it to be false. He used this idea to describe what he thought was the essential characteristic that distinguished science from all other activities – namely, the act of developing theories that are testable in the sense that they have consequences that can be checked out by observations, and the willingness to reject them if we find them wanting. Whether we regard ourselves as scientists or not, the idea is also relevant to other areas of research, in that we try to resolve the conflicts between our general ideas about the way we think the world works and the evidence we gain from our experience about it. Either our general ideas are mistaken, or the evidence is misleading. Hypothesis testing is one of the ways we try to eliminate such ambiguity.

Hypothesis Testing and Quantification

The desire to measure or to quantify is not something that is unique to science, but it is one of the characteristics most often associated with it. The reason why scientists and indeed other researchers try to measure things is that it sometimes allows them to test hypotheses more easily by expressing ideas precisely. This is perhaps best illustrated by looking at some of the tools that are available for testing hypotheses using statistical methods.

Hypothesis testing belongs to that part of statistics called 'inferential statistics', which is concerned with the grounds on which one might accept or reject a proposition. Most statistical tests operate in the same way, in that they try, in the simplest situations, to pose two opposing or counter ideas against each other and look for evidence that might suggest that one should be rejected. The procedure is rather like 'trial by jury', in which we begin by assuming the innocence of the accused. The prosecution makes its case, and we are only prepared to reject our assumption of innocence if the evidence against them is 'beyond reasonable doubt'. In hypothesis testing in statistics, the initial proposition is called the null hypothesis and the strategy is to see if there are any good grounds for rejecting it in favor of some alternative idea.

For example, suppose that we are interested in the knowledge farmers have about diffuse pollution problems and the contribution that their farms might make. We might be interested in this because as environmental managers or policy makers, we want to decide if there is a case for making better information available to farmers to help them minimize their environmental impacts. Dairy farming is often thought to be a major source of pollution and so as decision makers we may want to investigate if the views and beliefs of dairy farmers are different to those of other farmers.

Box 1 shows the analysis of some data collected from 75 farms in South West England, in an area of high conservation importance that is dominated by dairy, beef, and mixed farms. The data were collected by interviewing farmers and coding up some of the questionnaire material

Box 1 Analysis of knowledge of pollution sources among farmers using cross tabulation

These data were collected through interviews with farmers from 75 farms in South West England. The area is of high conservation importance and is dominated by dairy, beef, and mixed farms.

A contingency test was used to examine the frequency (counts) with which different types of farmers recognize different types of pollution as potentially arising from their farm. These tables show the responses for dairy farms versus nondairy units for the pollution sources 'sediment' and 'nitrates'.

In both cases our null hypothesis is that there is no difference between dairy farmers and other farmers in the extent to which they recognize the different types of pollutants potentially arising from their land.

Data for sediment issue:

Sediment Issue		What sort of farm are you?		Total
		Nondairy	Dairy	
Farmers do not recognize problem	Observed Count	**43**	**8**	**51**
	Expected Count	*38.8*	*12.2*	*51.0*
Farmers recognize problem	Observed Count	**14**	**10**	**24**
	Expected Count	*18.2*	*5.8*	*24.0*
Total	Observed Count	**57**	**18**	**75**
	Expected Count	*57.0*	*18.0*	*75.0*

Hypothesis test:

The expected counts are those that would arise if dairy farmers had the same responses as other types of farmers in the sample. The Pearson Chi-Square statistic (χ^2) which compares observed and expected frequencies has a value of 6.04, with one degree of freedom, the probability of exceeding this value by chance is less than 0.014 (i.e., 1.4 times out of 100). Thus, we reject the null hypothesis that there is no difference between dairy farmers and other farmers in the extent to which they recognize that sediment runoff might potentially be arising from their land.

Data for nitrate issue:

Nitrate Issue		What sort of farm are you?		Total
		Nondairy	Dairy	
Farmers do not recognize problem	Observed Count	**45**	**12**	**57**
	Expected Count	*43.3*	*13.7*	*57.0*
Farmers recognize problem	Observed Count	**12**	**6**	**18**
	Expected Count	*13.7*	*4.3*	*18.0*
Total	Observed Count	**57**	**18**	**75**
	Expected Count	*57.0*	*18.0*	*75.0*

Hypothesis test:

In this case the null hypothesis is that there is no difference between dairy farmers and other farmers in the extent to which they recognize nitrates as a potential pollutant arising from their land. For these data, the differences between the observed and expected counts are much smaller. The Pearson Chi-Square statistic (χ^2) is equal to 1.131, and with one degree of freedom, we would exceed this value to be exceeded by chance with a probability of about 0.29 (i.e., 29 times out of 100). Thus, we cannot reject the null hypothesis. Dairy farmers appear to be no different to other farmers with respect to nitrates.

so that it could be analyzed quantitatively. The farmers were asked about the types of pollution they recognized might be coming from their farm; these included sedi ment, nitrates, phosphates, pesticides, herbicides, and plastics. In areas where dairy and beef farming is wide spread, nutrient problems may arise through poor man agement of slurry and manures, or as a result of dirty water running off from farm buildings and yards; it is particularly a problem in dairy units where animals are gathered frequently. Disinfectants in dairy units may also be a source of phosphorus. Sediment runoff may be associated with dairy and beef farming but we might expect it to be less likely to be a problem in areas such as these where grassland is the main cover type.

In **Box 1** we have tabulated the frequency with which dairy farmers thought that the different types of pollution might be generated from their farm compared to all the other types of farmers interviewed. Our initial or null hypothesis is that dairy farmers are no different to other farmers. As decision makers we want to see if there is any evidence that might cause us to reject this idea in favor of one that suggests that the dairy farmers have a different level of awareness about the pollution that might be arising from their farms. In **Box 1**, the statistical test that is being used (Chi squared, or χ^2) simply compares for each pollutant category, whether the difference between the number of times (frequency) with which the dairy farmers recognized that they possibly caused pollution was different to that recorded for the other farm types. The logic behind the test is that we are trying to see if a difference between the 18 dairy farms and the others was likely to be due to chance alone, that is, the kind of variation that would arise if we drew a subsample of 18 farms from the group of 75 at random. If the difference in frequencies is large enough, then we might suspect that the dairy farmers are behaving differently.

The analysis in **Box 1** suggests that dairy farmers are no different from other types of farmers in the extent to which they believe their operations generate the types of pollution they were asked about, such as nitrate, except in the case of sediment, where dairy farmers more frequently cited this possibility than other types of farmers. In the case of sediment, the difference between the two groups was much larger than we might expect by chance alone.

The statistical tools used for hypothesis testing normally give us an estimate of the chance or probability associated with any difference we observe between groups. In the case of the responses to the question about sediment, the size of the difference observed is only likely to have come about 1.4 times in 100 if we had repeatedly sampled sets of 18 farms at random from the group of 75. On the grounds that such a difference in frequency of response as that observed is unlikely, we might be prepared to reject our null hypothesis that in their response to the sediment question, dairy farmers are no different to other farmers, in favor of the alternative hypothesis that they are not responding in the same way.

If we are to fully understand the nature of hypothesis testing in statistics, we need to consider further the logic that underlies the tools being used. What would be better for a decision maker to do – to reject a null hypothesis when it is true, or to accept it when it is false? Statisticians refer to these two possibilities in terms of making 'type I' and 'type II' errors, respectively. In terms of our analogy between hypothesis testing and trial by jury, the first type of mistake is equivalent to finding an innocent person guilty, while the second represents the situation in which a guilty person goes free. Just as we think it better to risk letting guilty people go free in order to minimize the chances of an innocent person being wrongly convicted, so statisticians argue that we should exercise caution in our hypothesis testing. They suggest we should try to avoid making a type I error, by only rejecting a null hypothesis if there are 'good grounds' for doing so.

In our example concerning dairy farmers, we can minimize the chances of making a type I error if we are only prepared to reject the idea that they are no different to others if the chances of finding such a large difference between frequencies are small. Often people use the convention that we should be more than 95% certain, that is, the difference should arise no more than 5 times in 100. However, if it is really important that we avoid this type of error, then there is no reason why we cannot make the decision criteria much tougher – by being prepared to reject the null hypothesis if we are 99%, or even 99.9%, certain.

The example of statistical hypothesis testing presented here is a very simple one, and there are many more sophisticated tools available to test hypothesis using other types of data. For example, in our farm study, if we had measured sediment loss to rivers on the different types of farm, then a *t* test could be used to investigate whether there was a difference in the mean levels observed between types (**Box 2**). Alternatively, if we had data on sediment loss over time following some awareness raising program, we might ask whether there was any evidence of a decline in the rate of loss over time (**Box 3**). In each case, however, the approach is the same, the null hypothesis is set up to challenge the idea that we are putting forward, and we only reject it if there are good grounds for doing so. Moreover, when we do reject the null hypothesis we never claim that the alternative is true. Rather, we simply judge that there is reasonable evidence to investigate that alternative further.

Contextualizing the Hypothesis-Testing Process

As the example shows, the process of hypothesis testing allows the researcher to start making some tentative claims about how the world operates. It also provides a framework in which the logic and assumptions underlying the reasoning process can be seen. Yet the act of hypothesis testing carries with it a wider set of

Box 2 Testing the differences between potential levels of sediment runoff between farm types

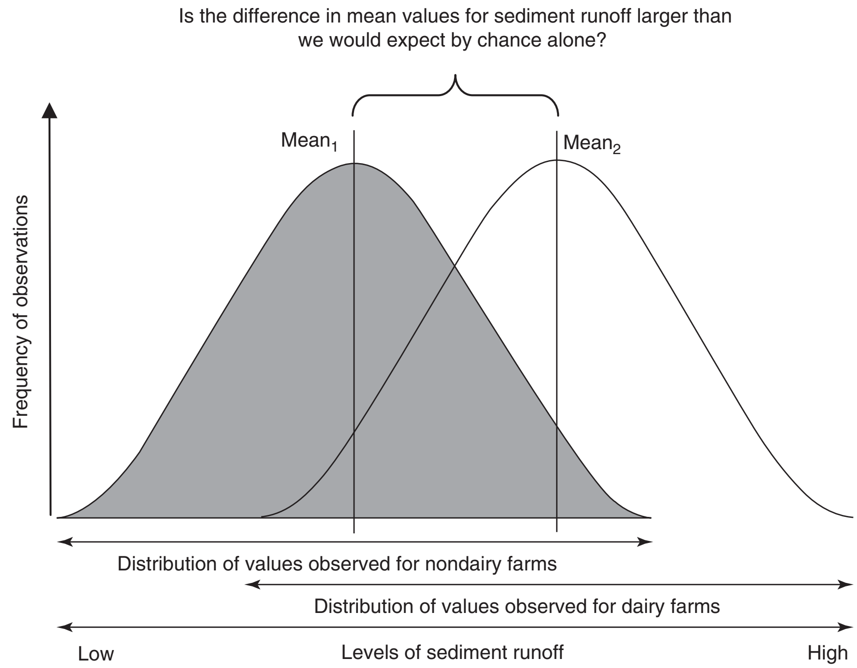

In this situation, a statistical technique called a *t*-test could be used to test the null hypothesis that there was no difference in the means of the daily sediment runoff values observed over, say one year, between dairy and nondairy farms. Only if the difference is larger than we would expect by chance alone would we reject the null hypothesis.

philosophical, practical, and ethical issues that affect how we evaluate it as a way of doing human geography research.

A good way of distinguishing between the issues at stake is to examine the process of hypothesis testing through the ideas of epistemology and ontology. By epistemology we mean the methods by which we are able to 'gain' knowledge about something. In this sense the use of hypothesis testing is an epistemological strategy. It is a technique by which the human geographer can begin to render the world meaningful in a certain kind of way, such as by exploring the association between farming sectors and attitudes toward diffuse pollution. In contrast, by ontology we mean the basis upon which we recognize something as 'knowledge' in the first place. In this sense hypothesis testing is not just a technique to produce knowledge about the world, but a means by which we explore 'what can be known'. This is true of any research process, whether based on hypothesis testing or not. Whichever epistemological strategy researchers have chosen to employ, they have already accepted, by implication, an ontological position. For those employing hypothesis testing, a very specific ontological claim is being made: namely that there are underlying patterns to the world that exist independently of our research into them, such as a potential relationship between farming sector and understandings of diffuse pollution. Hypothesis testing therefore works on the promise that our efforts may begin to uncover these hidden regularities.

Confronting the act of hypothesis testing with the idea of epistemology raises interesting questions about how and why we go about researching the world in particular ways, and to what extent researching social processes can provide us with the type of information demanded by hypothesis testing techniques. Central to these is the epistemological assumption of objectivity. It is suggested that the creation of knowledge under hypothesis testing involves the researcher performing his or her task in a neutral or disinterested fashion. After all, if the world has regularities and patterns that exist independently of human observation, then it might be argued that the assumptions or values of the researcher are largely unimportant. Arguably anyone could arrive at identical conclusions if the same hypothesis testing procedure is followed.

The point we wish to emphasize here is that, as an epistemological strategy, hypothesis testing is part of an active human process involving judgments about the selection and framing of hypotheses, the assigning of significance to them, and the drawing of conclusions from results. Indeed, hypothesis testing must be understood as an instrument of research through which a range of

Box 3 Testing hypotheses about changes in the levels of sediment runoff over time

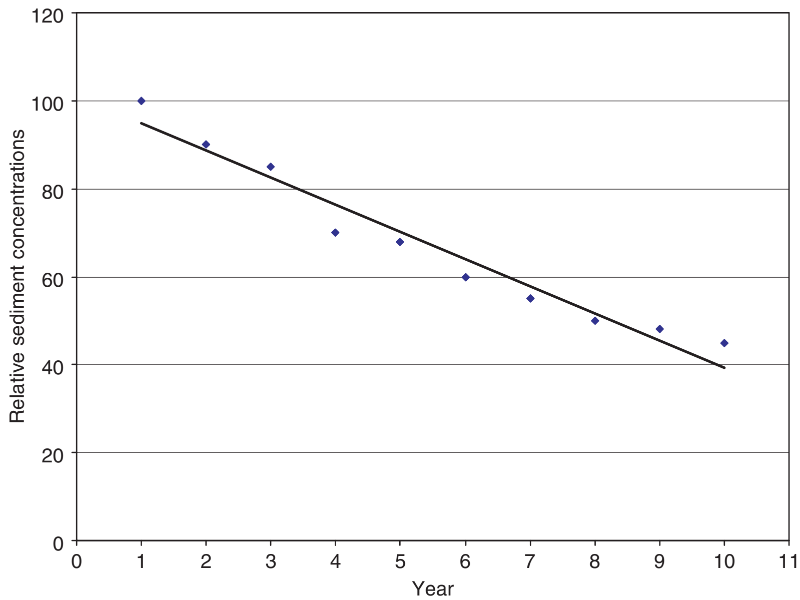

In this situation, a statistical technique called regression could be used to test the null hypothesis that there was no association between the yearly means of daily sediment runoff generated by dairy farms. Only if the slope of the line differed significantly from zero (that is the slope that we would expect if there is no association between the two variables) would we reject the null hypothesis of no change over time.

assumptions about the world can flow. This is apparent in a number of ways.

At one level, judgments arise when a researcher is faced with formulating a hypothesis. If it is one thing to set up a hypothesis that tests the relationship between farmer sectors and understandings of diffuse pollution, it is quite another to explain why the researcher would choose to pursue this hypothesis in the first instance. One answer to this is to say that these choices are logical consequences of a wider universe of hypotheses that contextualize the aims of a given piece of research (e.g., dairy farmers have different beliefs about pollution than other sectors). However, such a rationale does not explain why we choose to assign importance to diffuse pollution and agricultural sectors as a research issue.

Nor is the way we frame a hypothesis 'given' within a particular problem context. Even if the researcher could make a case for exploring the relationship between farming practices and pollution, this does not settle why the researcher would choose to examine attitudes as an important aspect of that debate. Researchers may choose to create and test hypotheses around a variety of equally plausible concerns, such as the market organization of different agricultural sectors, or the level of state assist ance afforded to them. While it may be argued that the process of selecting and pursuing a given hypothesis is precisely designed to start discriminating between contrary, yet plausible, understandings of given problem (i.e., to work with multiple hypotheses), it is nonetheless the case that by choosing one hypothesis over another a particular problem context tends to be emphasized as the way in which we think about research issues (e.g., 'atti tudes rather than money are important to overcoming diffuse pollution'). Realistically, not all hypotheses will be explored and not all associations between relevant vari ables will be discerned. Like any other research pro cesses, then, the researcher who employs hypotheses has to be very explicit about why and on what terms the questions are being framed. There is nothing objective about hypothesis testing in this respect. Even in the de cisions about the level of statistical significance that we will apply in our hypothesis testing we make a moral judgment: how much certainty do we need before we start regulating the dairy farming sector?

Issues of epistemology and hypothesis testing also concern the terms on which data are collected and used to make claims about the world with some degree of confidence. The example we have used in this article is instructive for it uses a data gathering technique widely employed by human geographers: that of the standard ized, closed, and face to face interview. Thus, before we even start to look at associations between variables in the data set, we have to accept the notion that data we re quire will be both robust and collectable. Principally, when conducting a survey like this we assume that interviewees understand what is meant by a particular question; answer a question when it is addressed to them; say what they mean when they respond to a question; say what they believe when offering responses; and respond in the terms that are expected. Researching humans as the basis for hypothesis testing is like any other research strategy in this respect: it comes with few guarantees. In the example data set we draw upon, we were fortunate to have 75 complete survey returns. But these returns in no way carry with them the certainty that we have com prehension, clarity, and frankness when making sense of the data. For example, why is it that dairy farmers rec ognized sediment as a problem but not the other types of pollution? Is it simply that it is more obvious or is sediment really an issue? Alternatively are farmers really telling us what they know about the other sources? Thinking about such issues is least one of the reasons why geographers researching human attitudes and values often emphasize the need to employ mixed method ologies, that link hypothesis testing to more probing and qualitative forms of investigation. What the example of dairy farmers shows, is that while superficially we appear to be testing only some simple, quantitative hypothesis, we are in fact confronting a whole network of hypotheses and assumptions that all need to be looked at critically before any final judgment can be made.

Conclusion

Although many human geographers do not consider discussions of scientific methods to be relevant, or the use of statistical methods as being appropriate (to the kind of geography they do), whenever we seek to examine a case for something by reference to evidence, we are involved in the act of hypothesis testing – whether we formally acknowledge it or not. It can be argued that hypothesis testing in its broadest sense is therefore relevant to a very large part of what human geography does and we ought to be aware of what it involves. At the same time, at the epistemological level, those who test hypotheses as part of the research process can learn usefully from other, more qualitative and interpretative, forms of human geography inquiry that have encouraged researchers to consider more explicitly their positionality when choosing, applying, and drawing conclusions from hy potheses, and in developing research strategies that rec ognize the limits of the data collection process. One way to address these limits, and therefore build confidence in the research process, is to use hypothesis testing and quantitative inquiry as part of a larger, more pluralistic, research methodology.

Even so, it must be recognized that dealing with these epistemological dimensions of hypothesis testing the ontological assumptions that guide its use *per se* are not being questioned. From the perspective of ontology, hy pothesis testing has often been set apart from other arenas of human geography. Indeed, in light of its en trance into human geography under the quantitative revolution, many have since argued that the search for underlying patterns is problematical in the context of social systems, for human attitudes and behavior are like a moving target. That is to say, there is no amount of available data that can make these kinds of systems stable to the degree that hypothesis testing seems to aspire to. One response to this has been to suggest that the onto logical claims underlying hypothesis testing need, at the very least, to be remodified to suit the human conditions they seek to account and speak for. Instead of regarding the world as a set of social processes whose rules can be progressively tamed and revealed, insights should instead aim to be 'soft' assessments, based on a more iterative understanding of what appears to 'work' in any given moment or circumstance. The extent to which such a situation precludes any requirement of researchers to test their social understandings in a critical way is a debate that must be had elsewhere.

See also: Critical Rationalism (After Popper); Philosophy and Human Geography; Physical Geography and Human Geography; Redbrick University Geography in Britain; Scientific Method.

Further Reading

Chalmers, A. F. (2004). *What's This Thing Called Science?* (3rd edn.). Buckingham: Open University Press.
Harvey, D. (1969). *Explanation in Geography*. London: Edward Arnold.
Johnston, R. and Sidaway, J. (2004). *Geography and Geographers Anglo American Geography Since 1945*. London: Hodder.
McGee, B. (1973). *Karl Popper (Modern Masters)*. New York: Penguin (Non Classics).
Peet, R. (1998). *Modern Geographical Thought*. Oxford: Blackwell.
Policy Hub (2008). How research and evaluation evidence contributes to policy making. http://www.policyhub.gov.uk/evaluating policy/ how res eval evid.asp#philosophical. (updated 12/2/2008)
Popper, K. R. (1972). *The Logic of Scientific Enquiry (Sixth Impression)*. London: Hutchinson.
Rogerson, P. A. (2001). *Statistical Methods for Geography*. London: Sage.
Sayer, A. (1992). *Method in Social Science: A Realist Approach*. London: Routledge.

Relevant Websites

http://ccnmtl.columbia.edu
Quantitative methods in the social sciences, QMSS e lessons, Columbia Center for New Media Teaching and Learning.
http://www.eeng.dcu.ie/~tkpw
The Karl Popper Web.

I

Idealism/Idealist Human Geography

L. Guelke, University of Waterloo, Waterloo, ON, Canada

Glossary

Dualism A philosophical position in which the world is seen as encompassing both mind and matter.
Humanistic Geography A component of human geography emphasizing the centrality of human consciousness for an understanding of human geographical activities.
Idealism A philosophy of knowledge in which preeminence is given to mind in the understanding of the human condition.
Inside (of an Event) The Collingwoodian notion that one understands the observable or physical dimensions of a human event by getting inside the minds of the participants or actors involved in the event under study.
Mind A philosophical term for the locus of human thought and the center of consciousness.
Monism A philosophical position in which the world is seen as being comprised of a single entity or substance, usually mind or matter.
Nature A general term for the world of phenomena not created or influenced by human minds.
Outside (of an Event) The Collingwoodian term for the observable or physical dimensions of a human event (see the entry 'inside').
Rethinking or Reenacting Thought – the key method adopted by Collingwoodian idealists of getting inside the minds of human agents.

Philosophical Introduction

There have been several great philosophers, among them Plato, Berkeley, Kant, and Hegel, who have propounded distinctive idealist philosophies. At the most general level a philosophy merits the designation 'idealist' if it emphasizes the centrality of mind or spirit as the basis of reality and the foundation of human existence. Idealism can be contrasted with materialism, which holds that all things have a physical or material basis. On this position the mind and spirit have no independent existence and are merely manifestations of physical processes. The materialist, one might think, has an advantage here in adopting a single substance or monistic philosophy, but many idealists such as Hegel are monists, claiming that ideas have a life of their own with the power to change and transform society. This is not the variety of idealism advocated here.

There have also been a number of philosophers in the idealist tradition, who have advanced middle of the road or dualistic positions. The philosophers such as Kant do not deny the reality of a material world that exists independently of mind, but argue that this world is unknowable until it has been interpreted by mind. The Italian philosopher, Giambattista Vico, presented an early case for an idealist sociology and history arguing that one could only understand things one had made oneself. On this position, nature was knowable only to God (His creation), but one could understand human society as the creation of human beings just like ourselves. Later, philosophers, including Benedetto Croce and R. G. Collingwood, sought to build on the work of Vico and others in their efforts to produce philosophies of human action that were founded on the idea that people have created their own societies as thinking beings.

Idealism in the Collingwood Tradition

The English philosopher and historian R. G. Collingwood had a particular interpretation of idealism, which he developed in several major scholarly works. Among these works was the *Idea of History*, in which Collingwood set out a comprehensive philosophy of history in the idealist tradition. This work has been widely studied by historians and philosophers of history, and is also relevant for human and cultural geographers, who are concerned with the distinctive landscapes and regions of the earth. In *The Idea of History* Collingwood developed an approach to understanding people as historical agents by focusing on their thought. The fundamental assumption here is that people have the capacity to think, and, in fact, have used this capacity to construct intellectual models of themselves and the world around them, which become, in turn, the basis on which they devise strategies to manipulate situations for whatever objectives they might happen to value in specific historical contexts.

The human geographer in adopting a Collingwoodian position is committed in similar vein to the historian to discovering why people acted as they did by probing the thought that is assumed to be imbedded in all deliberate human actions. In human geography there is no need to pay much attention to the irrational (conceived here as random, incoherent thought), because the overwhelming majority of human actions having to do with spatial movement or the use of the Earth are in fact the result of the deliberate thoughts of millions of people seeking to support themselves and/or to live their lives in some culturally meaningful way.

In establishing a model that emphasized thought as the basis on which human societies were built, Collingwood did not ignore the physical dimensions of people and their surroundings. He was quite favorably disposed to science which sought to understand the workings of nature by developing classification systems and to discover laws and theories that connected the elements of the systems it had defined. The scientific approach was also applicable to human beings to the extent that human beings could be treated as natural phenomena. Collingwood used the term 'outside' to refer to the physical dimensions of human existence, which were fully part of nature and controlled by natural laws. The human being, according to Collingwood, was unique among natural organisms in having the ability to think and reason to the point of creating a historical life for itself. This life was not a response to conditioning stimuli, but a creation of human thought, which took over from biological evolution as the driving mechanism of human transformation. The study of thought involved, in Collingwood's terminology, the 'inside' of human actions, which were not amenable to study using scientific methods. The mental achievements of human beings were governed by its own rules of thought and reasoning that transcended the physical foundations of mind and body that made such achievements possible.

In Collingwood we have a dualistic system in which the appropriateness of the approach one adopts depends on the domain of its application. Thus, for Collingwood, there is an appropriate domain of science where its methods are applicable and the tested knowledge gained using such methods is valid. In acknowledging the value of science, Collingwood is able to use the full arsenal of scientific tools in seeking to discover what happened in specific historical situations. In geography, this position means that all available scientific knowledge of physical phenomena can be incorporated in the study of human activity in geographical space. The objective of an idealist geographer would be to nail down what happened in terms of bodies and their movements as carefully as possible, using every scientific tool and dataset available. It is in the domain of explanation that involves human activity that the idealist diverges sharply from the approach of the scientist. This point is at the core of idealism and needs full elaboration.

The idealist constructs the human being as a part of nature subject, as a physical organism to nature's laws. People are trapped in their bodies and have no power to escape the materiality and impulses of their bodily existence. However, people also have minds, which the idealist argues, are capable of controlling the agenda of the physical bodies in which they are lodged. The embodied mind is situated in a specific time and place and develops its power in a growing organism in the context of a specific society or social milieu. The individual mind, however, is not simply filled up with knowledge as a glass is with water, individuals must grasp and understand the ideas presented to them. They can no more escape the social context of their intellectual developments than they can escape their physical bodies, but once learning has been achieved within a specific context, the individual, depending on his or her mental resources and intellectual disposition, can build on, reject, or indeed transcend the knowledge acquired from the social environment in which he or she is located.

In affirming that individual human beings are physical entities, one also acknowledges that they have an existence as living organisms subject to all the natural and biological forces that might apply. Most obviously, people have built in needs for food, water, and shelter, and they go through predictable life stages from birth to death. They are also social and emotional beings with a variety of psychological attributes and needs. The embodied thinking mind, however, is not the mind of passion or emotion, although often in their service, but the mind of self conscious reflection, seeking to make sense of the world, in all its apparent irrationality, in which it finds itself. This mind is the historical mind that has evolved over generations. Each generation is confronted with specific problems it seeks to solve, leaving a legacy of its accomplishments for the next generation to adopt, modify, or reject as it seeks to solve problems of its own times.

The central point about the reasoning mind is that it is the author of all things intellectual from culture to constitutions. In the British American colonies of 1776 there were many angry colonists, who felt that Great Britain had not governed them well, but anger did not write the Declaration of Independence. This document was written for political purposes and involved specific individuals with specific aims setting down in writing a reasoned political agenda. The modern scholar understands this document by putting it in its historical context and understanding the various influences that shaped the way its authors had gone about their work. Was the case well made? Were the grievances it alleged supported by the colonial realities it purported to represent? Was there a hidden agenda here that masked more dubious motivations? Any or all of these questions would be part of an analysis of this particular document. However, the key point is that the analysis is not in the realm of psychology, but in the realm of reason and intelligibility because it is the realm that has governed human societies ever since people established historical societies based on intellectual principles. The term 'reason' is not used in any absolute sense here, but rather as a synonym for a reasoning process that people employ in order to justify their actions to themselves and others in particular historical situations.

The above paragraphs focus on the critical importance of understanding human activity as an expression of reasoning and intellectual activity. Emotions might drive people to do things, but strategies for achieving particular goals are the products of analysis and reasoning. Thus, an individual might be driven to succeed, but success must be earned in the context of a social order, and the ambitious individual needs a strategic plan to realize what might be emotionally desired. Precisely because emotions need not be articulated and are often disguised, the question that arises is how can we know what a person's reasons for doing something or saying something really are? At this point we are in difficult terrain, and the idealist position requires careful and logical explanation. The way one understands another person or group involves, according to Collingwood, rethinking or reenacting their thought in one's own mind. Before this assertion is dismissed out of hand, some explanation is necessary, because the claim is not as absurd as it might seem.

Thought and Action

If one asks the question: "Why did Christopher Columbus sail west across the Atlantic Ocean in 1492?" the answer in a very real sense involves understanding what was in his mind. Columbus believed the Earth to be round, but greatly underestimated its size. He also believed that there were good trade prospects in China. Putting these two pieces of information together, he devised a plan to sail west across the Atlantic with the intention of reaching China. An idealist explanation establishes that there was a logic connecting what Columbus thought with what Columbus did. This procedure involves getting inside the mind of Columbus, and rethinking the thoughts that motivated him to set out on his historic voyage.

In the above example some critical observations need to be made. First, the focus of attention is only on those ideas that are relevant to an understanding of why Columbus sailed west in search of China. The idealist has no power and no interest in rethinking every thought Columbus ever had. The focus is on those thoughts that help us understand specific actions that are deemed important in the context of the particular study being undertaken. Second, the idealist is not given carte blanche to invent thought willy nilly to explain a particular action or set of actions. When a specific action is explained with reference to an individual's ideas, these ideas must be based on evidence that is open to scholarly inspection. The idea that Columbus believed the world to be round and much smaller than it is needs to be verified by careful historical research of his geographical knowledge. One would try to establish the intellectual climate in which Columbus lived and endeavor to establish his beliefs from his journals, letters, and any other written sources that might be available. Third, in seeking to establish what a person or group might have thought, the idealist will adopt a critical, probing approach, in the knowledge that people will often mask their intentions in what they say, but reveal their true objectives in their actions.

The scholar involved in discovering what an individual might have been thinking in taking a particular, or set of, actions does not relinquish the independence of his or her own mind. The quest to enter the mindset of individuals of another time or culture is an imaginative, critical undertaking requiring both empathy and detachment. The goal is not only to discover what people thought, but also to gain some sense of why they thought as they did within the mental universe of which they were part. For example, the mindsets of slaveholders in the antebellum American south are far removed from our contemporary ideas of social justice, human rights, and racial equality, yet the historian of this period needs to have some sense of the slaveholder's mentality if he or she is to make any sense of how such a regime was possible. This task is accomplished by reading what the slaveholders had to say for themselves and gathering whatever evidence is available that could provide clues about how they understood what they were doing and justified their actions to their own satisfaction.

In presenting an interpretation of human activities that is grounded in the ideas of the individuals under study, the idealist aims at making such action intelligible by connecting thought and action in a sequence of logical or understandable steps. The outcome of any specific human episode as it unfolds as a spectacle of human activities is a function of the thought that is imbedded in the actions and reactions of the participants. Collingwood famously stated: "All history is the history of thought." He was not denying the existence of a physical or emotional world, but merely insisting that human beings acted in the world in terms of how they perceived and understood each other and the world around them. The universe might not change in its materiality, but because people can only interpret or model it in terms of their intellectual concepts and theories it exists as a human reality subject to change. In modern terminology one would say that reality is a social construction dependent on the ideas people have employed in constructing it.

The Collingwoodian idealist seeks to provide an accurate account of the topics investigated by providing explanations that are supported by appropriate evidence. In the same way a physicist tries to develop theories that are consistent with empirical data, the idealist scholar is at pains to ensure that interpretations dealing with the thought imbedded in human actions are consistent with the available evidence. The idealist model of scholarship is one of openness and debate. Contending interpretations need to be pitted against each other in an arena in which the coherence of argument and the quality of evidence would ideally determine acceptability. Idealist scholars, one could say, have scientific aspirations in the broad sense that they aim to provide objective explanations of human behavior, supported by evidence open to all. However, idealism is not concerned with prediction and relies on the quality of argument, inference, and evidence to convince other scholars of the value of its approach.

Idealism in Geography

The idealist philosophy has not been widely adopted as an explicit philosophy of explanation among geographers. Some time ago, a geographer referred to positivism as the hidden philosophy of geography, because it was implicitly adopted by many geographers with no special interest in philosophy. In some ways idealism has also been the hidden philosophy of many cultural, human, and historical geographers in the sense that they have given primary importance to human ideas. Geographers in these fields are familiar enough with the importance of perceptions, belief systems, and regional and ethnic customs in their efforts to understand the distinctiveness of individual societies. They are far more apt to rely on interpretative descriptions rather than theories or laws in trying to convey the essential attributes of a regional landscape. Yet, notwithstanding such interests, few such geographers would actually describe themselves as idealists.

The lack of a clear commitment to idealism among geographers who rejected a social science approach is to some degree a function of the way the discipline of geography developed. In the period before the 1970s, practically all geographers were wedded to science as the only respectable paradigm for any kind of empirical work. A first response to the failure of early models of spatial behavior to connect closely with empirical evidence was behavioral and perception geography. The notion was that people viewed their situations and surroundings subjectively, and it was assumed this subjectivity could be measured, analyzed, and incorporated in more accurate models of human geographical activity. This approach made more accurate assumptions about human behavior, but it retained a commitment to the scientific method seeking to discover general principles and laws that accounted for particular kinds of spatial and environmental behavior.

The real challenge to the scientific paradigm emerged in the 1970s. A number of philosophically inclined geographers produced an array of alternatives, from Marxism to phenomenology, with a view to providing geographers with new approaches more suited to their discipline than the strict adherence to the scientific method had allowed. Later, feminism and postmodernism and other approaches were added. Many human geographers had no inclination to worry about the specifics of the various philosophies that challenged the appropriateness of scientific methods in the study of human societies and worked within the broad umbrella paradigm of humanistic geography. Humanistic geography incorporated elements of a wide range of philosophical positions that emphasized human agency, and provided a working, if imprecise, home for many geographers seeking alternative approaches to those offered by social science.

Although it would be stretching things to claim many or even most humanist geographers were in fact idealists, such geographers have clearly tilted their work in the direction of idealism to the extent they focused on societies as human creations and sought to make sense of human activities by locating them in their cultural and historical contexts. A great many cultural geographers studied places and landscapes as embodiments of the distinctive peoples and cultures responsible for their creation. The emphasis here was on the uniqueness of place, region, or landscape and how this uniqueness was connected to the values and ideas of the people who had forged a livelihood in applying these values and ideas to specific geographical spaces.

The notion that idealism provides a solid philosophy for humanistic geography needs to be elaborated in a stronger way. First, one needs to show that many humanistic geographers would be in a stronger position to achieve their basic goals by adopting idealist principles. Second, one needs to demonstrate that the philosophy of idealism has more to offer than that of its possible rivals in terms of what it is capable of delivering when it is consistently and rigorously applied. The claims that will be made for idealism as a philosophy of humanistic geography are closely related to my view of the objectives of humanistic geography, which will be clarified before assessing alternative approaches.

It is assumed that humanistic geography is concerned with the worlds people have made for themselves. It is sharply focused on what people have made of their lives in a wide variety of physical and social settings. People are seen as the authors of their lives, not helpless victims of forces they cannot control. Yet, many people do contend with situations that are difficult and ones in which power imbalances are severe. In these situations people can show amazing resilience, and the humanist geographer is at pains to show that even in these cases people can be creative and skillful in making their lives bearable and meaningful. Humanistic geography gives emphasis to the culture, consciousness, and values of the world's peoples, investigates what people have done to and on Earth, at a variety of scales with a view to gaining greater insight into what it means to be human. If such studies are to be about real people in real situations, the humanistic geography must transcend the subjective: it is not competing in the literary domain and its products are not intended to be fiction.

Why is Collingwoodian idealism a strong contender as the foundational philosophy for humanistic geography? In the first place, this philosophy avoids the extremes, unnecessary assumptions and unrealistic goals of contending philosophies, and can be shown to be applicable to the broad objectives of humanistic geography. The idealist philosophy not only provides scope for individual geographers to be creative and imaginative in studying human situations, but also insists that all interpretations should be grounded and supported with empirical evidence. Idealism, it will be argued below, differs from and is more suited to humanistic geography than the following alternative philosophies: social science, phenomenology, Marxism, and postmodernism, and is fully capable of addressing the challenges of feminist geography.

Alternative Approaches

The social scientist like the scientist adopts a model of explanation in which all phenomena are seen as objects to be understood using laws and theories. The scientist categorizes, measures, generalizes, and theorizes using prediction as a means to test the validity of his or her knowledge. In geography this approach has been most successful in general models such as central place theory, but the key problem for those pursuing a scientific approach has been the difficulty in finding laws of human behavior. The complexity and range of individual human societies has not been amenable to the development of the comprehensive laws needed for scientific explanation. This problem did not escape those geographers seeking to construct models of human behavior based on perception. The cultural context of human perceptions has prevented scientific geographers from developing general laws that would support a perception based scientific geography allowing for the accurate modeling of human behavior.

The philosophy of phenomenology has, of all alternative paradigms, the most in common with idealism and certainly provides a viable philosophical grounding for a range of humanistic studies. The phenomenologist focuses on human experience and is concerned to provide a detailed description of how an individual or group might experience a particular environment or physical event. This goal is approached by the individual immersing him or herself in the rigorous reflection and introspection of experience, and the successful outcome is a rich description of the essence of the experience or object being described. This kind of approach has produced some valuable studies on topics dealing with the experiences of a rural landscape and urban neighborhoods. Yet, notwithstanding the disclaimers of phenomenologist themselves, phenomenology as actually practiced veers toward the subjective. The individual descriptions stand on their own and lack clear, operational verification procedures. In a nutshell, phenomenology purports to support a humanistic geography that provides a foundation for the rigorous description of place, but in reality has often sanctioned a highly subjective poetry of experience. Idealism is more focused on human geographical activities, seeking to understand the strategies people have employed to make livelihoods for themselves and to give meaning to their lives as cultural and historical beings.

There is little that connects Collingwoodian idealism and Marxism apart from their common concern with human societies in their historical contexts. Marx was intent on producing a scientific theory of history which gave little freedom to the historical participants in a drama that might have been titled *The Rise and Fall of Capitalism*. The idealist shares with the Marxist an interest in the role of class to the extent that people in a given society might have understood their situations in terms of class categories. The historical agenda, however, is open and there is no attempt at predicting future historical states of society. It is assumed in idealism that the individual is a free agent, but not necessarily a reckless one. People are quite capable of assessing their situations in terms of their life chances, and will normally adapt their actions in terms of what they regard as realistically achievable. The oppressed classes have often participated in their own oppression by accepting class rules that were detrimental to their interests. This situation occurs where people in inferior roles see no way of changing the power structure within which they are embedded. Revolution is also a possibility, but revolutionaries as much as conformists need to map out strategies that will bring about desired changes in the face of vested interests. In both these cases the idealist will seek to understand how people understood their worlds and calculated their futures based on their values and beliefs.

Postmodernists have not only adopted certain key ideas that have long been associated with idealism, but have also added elements that the idealist finds unacceptable. The postmodern notion that human societies are social constructions not amenable to scientific analysis is also a key idealist principle as it is the idea that people behave in the world as they represent it. Idealism begins to diverge from postmodernism on the issues relating to the status of scholarly interpretations. The idealist scholar is intent on providing an accurate account of the topics studied, and believes that people act for reasons that need to be discovered and explicated. There is no authorization for a scholar to interpret an individual's actions, utterances, or writings in any way apart from seeking to provide an accurate elucidation of what the individual actually thought and meant. Any interpretation must be open to scholarly criticism and the grounds on which it is based is available for inspection. There are, in other words, solid rules of scholarly engagement associated with Collingwood's position aimed at separating fiction from nonfiction, and trying to ensure as accurate a portrayal of the intellectual side of human activities as it is possible to achieve. The idealist scholar is also aware that power relations shape human interactions, but is convinced that such relations can be accommodated within a model of explanation that focuses on how people think about and assess their situations, including any considerations of power that might be taken into account.

Idealism has not received much attention from feminist and others who emphasize the psychoanalytical world of unconscious desires and impulses as providing the basis for an understanding human activity. In focusing on what people think, the idealist might appear to neglect the nonrational aspects of human existence, but this is not the case. The idealist deals with all phenomena as they are understood by the individuals of a given society. The key issue is not to find ultimate causes of specific activity, but to understand how individuals express their unconscious, or indeed conscious, desires and needs in concrete circumstances. The attainment of any particular goal in any human society requires a strategy or plan that lies in the realm of thought, and such plans are accessible to the idealist method of rethinking. The goal itself is simply taken as given, and not considered as requiring an explanation by a scholar concerned with what people have made of their lives as the creators of their own social customs and traditions. Thus, the idealist is not concerned with human sexuality as a biological and psychological phenomenon, but focuses on the myriad of ways different societies have understood the phenomenon and created cultural rules to which men and women are expected to conform and need to take into account in finding ways of expressing their sexual identities. The central question for the idealist is how do people understand themselves and others in a social world in which all the phenomena (including a variety of biologically grounded sexual differences) with which they are content are subsumed within intellectual categories and models generated within the minds of the individuals concerned.

See also: Human Geography; Humanism/Humanistic Geography.

Further Reading

Collingwood, R. G. (1916). *Speculum Mentis or the Map of Knowledge*. Oxford: Clarendon Press.
Collingwood, R. G. (1946). *The Idea of History*. Oxford: Oxford University Press.
Dray, W. H. (1995). *History as Re Enactment: R. G. Collingwood's Idea of History*. Oxford: Oxford University Press.
Goldstein, L. J. (1976). *Historical Knowing*. Austin, TX: University of Texas Press.
Guelke, L. T. (1983). *Historical Understanding in Geography*. Cambridge: Cambridge University Press.
Guelke, L. T. (1997). The relations between geography and history reconsidered. *History and Theory* 36, 216 234.
Johnson, P. (1998). *R. G. Collingwood: An Introduction*. Bristol: Thoemmes Press.
Mink, L. O. (1969). *Mind, History and Dialectic: The Philosophy of R. G. Collingwood*. Bloomington, IN: Indiana University Press.
Van der Dussen, W. J. (1981). *History as a Science: The Philosophy of R. G. Collingwood*. The Hague: Martinus Nijhoff Publishers.
Vico, G. (1968). *The New Science* (transl. Bergin, T. G. and Fisch, M. H., 1744). Ithaca, NY: Cornell University Press.

Identity and Otherness, Rural

J. Agyeman, Tufts University, Medford, MA, USA
S. Neal, Open University, Milton Keynes, UK

Introduction

What does 'rural' mean and what meanings have been at tached to the concept of 'the rural' Who does live in rural places and who is assumed to live in rural places? How are rural places changing? Such apparently straightforward questions belie their complexity and their scale. They are questions which apply locally, regionally, nationally, trans nationally, and globally. They are both current questions and historical questions. Any response to these questions sooner or later – and usually sooner – involves a discussion of processes of social inclusion and exclusion.

Let's look at an example of these points. Mike Davis begins *Planet of Slums* – his indictment of neoliberal, global industrial capitalism – with a startling observation:

> In the imminent future the unstoppable trends in human birth rates in slum areas combined with increasing rural to urban migration will mean that 'for the first time the urban population of the earth will outnumber the rural. Indeed this epochal transition has probably already oc curred' (M Davis, 2006: 1).

On a global scale, what Davis is describing here is the continuation of a transition that began in the nineteenth century with the Industrial Revolution and the rise of cities like Manchester in Northern England. This nine teenth century rural to urban transition was driven by industrializing processes. These included the 'enclosure movement' which saw rural populations evicted from their small holdings and plots as the common land on which they were settled was seized and fenced off by local political elites and land owners. The 'enclosure movement' was itself a forerunner of the 'clearance movement' in the Highlands and Islands of Scotland in the eighteenth and nineteenth centuries that saw Scottish crofters often forcibly removed and evicted from their small tenancies as Scottish and English landowners in creasingly turned the land over to sheep grazing.

The scale of urbanization that Davis comments on is however primarily a phenomenon of the poor, de veloping, majority Global South where megacities such as Sao Paulo, Mexico City, and Mumbai exceed 17 mil lion inhabitants. Rural de population is then no longer the dominant internal migration trend in the affluent countries of the Global North. In contemporary Britain, for example, the rural to urban population has grown at a far smaller rate than the urban to rural population reflecting an increase in counterurbanization flows. In the Global North, the counterurbanization phenomenon has tended (but not always) to involve more affluent and older populations. This not only balances any rural de population, it far exceeds the numbers of those who leave rural areas. In the US, the dominant trend has also been away from urban environments but has veered toward suburbanization rather than toward a rurally based, counterurbanization.

More than anything, these historical moments and seismic settlement transitions illuminate the absence of any notion of a settled or fixed rural population, despite the dominant imagery of timeless, unchanging countrysides. Clearly, across time and space, notions of 'who' lives in rural areas have shifted and changed. Of course, these shifts and changes in human settlement patterns represent and in corporate large scale socioeconomic structural processes and transformations, but they also represent micro patterns of individual agency, desires, and motivations. Rural popu lations and places, like urban populations and places, are turbulently constituted, unsettled, and not fixed.

However, in governmental and popular thinking there are very fixed ideas as to 'who is' rural and who 'legit imately' inhabits rural spaces. It is precisely the con testations between these two positions that have provided a focus in debates on rural identity and rural otherness. This focus itself reflects something of a shift in rural studies approaches to the analysis of the 'concept' of rurality and the range of cultural and sociological meanings attached to rurality. These interrogations have involved an emphasis on those populations considered not 'of' the rural or as having contentious claims to a place within rural spaces. It is in the UK and the English context in particular – a nation described as having 'a cult of the countryside' – that rural otherness has become a key arena for research and discussion. This is not to suggest that rural otherness debates are not transnational ones but rather to note that they are unevenly emergent. In the US context, the research of the rural sociologist Sergio Chavez which documents the hostilities and ten sions between Mexican migrants and white North American residents in the rural Californian town of Yodoy offers a glimpse of contested, exclusionary and racialized rural social relations in the contemporary US when he notes that "almost no [white] resident I spoke with considered Mexicans as contributing members of the town [...rather] Mexicans were seen as creating com munity in ways that promoted their ethnic and class differences which some white residents saw as threatening to the agrarian culture of the town and to American values." Again in the US context, it is possible to see that rural otherness debates have coalesced in connected but rather more diverse directions – for example, the environmental justice/racism has received much more attention than rural racism and rural leisure and recreation has been a focus for issues of exclusion.

In many ways, the discussion here does have an Anglocentric bias to it. This can be seen as a reflection of the peculiarly dominant place the rural takes up in the UK's national story and the corresponding theoretical interest in this. However, in acknowledging this bias, the discussion explicitly addresses the question as to the extent to which the rural otherness debates examined below are nationally bounded or whether these debates, or parts of them, are more globally applicable.

British Rural Studies and the Cultural Turn

By the beginning of the 1990s, rural studies began to move away from a focus on positivistic measurements of the countryside. The tendency to approach rural spaces through quantitative, objective indicators and definitions such as population size, types of land use, agricultural economies, and rural policy formations and interventions increasingly began to give way to cultural definitions and subjective and qualitative indicators. In short, the study of 'the rural' moved from an easily definable and measurable premise and emphasized the rural as a much messier, more slippery, socially and culturally constructed phenomenon. This shift was inflected by and reflective of a series of relational changes in the British countryside – a rural economy based less and less on agricultural production and an emergence of what has been termed a postproductivist countryside; an increasing commodification of the countryside by tourist and heritage industries; significant middle class urban to rural migration; a less distinct set of distinctions between the rural and urban binary; and, in the 1980s and 1990s, the high profile insertions of problematized, 'nonrural' figures in rural spaces, most notably, for example, 'New Age' travelers'.

These types of changes and the opportunities that they have afforded in rethinking rurality have given rise to questions such as 'what is the rural; what types of representational labor does it do and for whom, and who constitutes rural populations'? Questions such as these emerged as central not only to academic, but also popular and governmental agendas. This is not to say that such questions had been previously absent – they had not. The 'right to roam' protest movements of the 1920s and 1930s and even the Diggers movement in the seventeenth century testify to the longevity of struggles as to who could be in the rural (see below). Rather, it is to note how, in the early 1990s, as the cultural turn in geography and the social sciences more broadly, took hold, it became a moment for retheorizing rural spaces and their populations. In particular, it was the hegemonic discourse of rurality and within this the obscured heterogeneity of rural populations and the cultural and representational work that occupied academic and research approaches to rurality. This hegemonic discourse was one that represented rural spaces as picturesque, unchanging sites of social order and deference, community sameness and familiarity, and part of this was a representation of the countryside as a 'white space'. In this way, the countryside was collapsed into a discursive tool of social reassurance. It is the antithesis of the unruly city with its plethora of problems, conflicts, and tensions, and its multiculture. Within the rural reassurance discourse, it was of course necessary to marginalize, invisibilize, and subordinate those populations and issues that did not sustain it. In this way, dissident and alternative rural dwellers and a range of social problems from poverty, homelessness, racism, unemployment, drug abuse, isolation, and inadequate social welfare resources were unaccounted for in notions of what the rural was. Such populations and social problems were effectively hidden and/or neglected in academic, popular, and governmental approaches to the rural and the countryside.

Rural studies scholars' focus on this 'hiddenness' represented a conceptual and methodological embrace of conversations taking place across the social sciences. In particular, the concept of 'otherness' has been used as an appropriate framing device for understanding some of the contestation over rural identity and meaning. In uncovering the more concealed aspects of rurality and the countryside, rural studies scholars were arguing for the recognition of those other countrysides and other rural identities.

Developed out of psychoanalysis and the work of Lacan, postcolonial theorists such as Edward Said, Gayatri Spivak, and Homi Bhabba have used the concepts of others (those threatening, outcast subjects defined in relation to the center/the Empire/the West) and the dialectical process of othering (the discursive and material ways in which othered subjects become defined by and simultaneously define the center or dominant identity) to analyze colonial and postcolonial racisms and racialized difference, subordination, and exclusion. It is possible to identify notions of others and processes of othering in the debates about rural identity. Marginalized, subaltern, silenced, undesirable rural figures – gypsies, travelers, tinkers, tramps, those leading alternative lifestyles, the poor, can be understood as rural others. The social and cultural construction of problem rural populations and 'difficult' and unruly urban rural figures who do not belong can be understood as part of a process of othering which simultaneously defines what/who is defined as rurally desirable and conventional and what/who is not.

Filtering concerns about rural identity through no tions of otherness meant that rural studies was able to avoid a simplistic focus on hidden rurals and invisibilized or problematized rural presences through a relational scrutiny of hegemonic discourses and narratives of the rural and the countryside. In short, issues of power, politics, and representation have shaped rural study inquiry and theorization. This turn was variously cap tured in a number of edited collections and monographs published in Britain in the mid 1990s which addressed a range of neglected and marginalized rural identities and the cultural and social meanings attached to the category 'rural'. These edited collections included chapters on radical environmental protestors; women and gendered experiences of the rural; lesbian communes; children and young people; black and minority ethnic rural residents, workers, and visitors; race and ethnicity; migrant agri cultural workers, rural poverty, and gentrification. Ethnographic research explored New Age travelers' identities, lifestyles, rural attachments, and their (often violently) contested and problematized presence in the English countryside.

The theoretical deliberations emphasized the di vergent, interdisciplinary routes through which much more 'liminal' rural geographies could be conceived. In the postmodern de coupling of the concept of rurality from the conventional measures and markers of what the countryside is and does, the central theorization was the way in which socially and culturally constructed notions of rurality created the rural as a purified and sanctified space requiring (governmental) protection from perceived threats. It is arguments about geographies of purification and exclusion which resonate with a number of the con testations over claims to be in rural spaces. Recent ex amples of such contestations in the British context have been young people and the free party movement and New Age travelers; but these problematized presences are im printed with the ghosts of earlier struggles of urban working class claims to the countryside and the 'right to roam' movement in England in the 1920s and 1930s.

Between the two World Wars there was a vibrant and popular movement that demanded access to the coun tryside for leisure pursuits, especially for the growing activity of rambling, for the urban based working classes. As access to rural areas was regularly and often forcibly denied by landowners despite ancient rights of way and footpaths, the right to roam became increasingly politi cized as ramblers used the tactic of mass trespass to pro test. The first and best known of these mass trespasses was on Kinder Scout in the Derbyshire Peak District during the 1920s and 1930s. It has been observed that the Peak District's location in Northern England, surrounded by large industrial towns and cities such as Manchester, Sheffield, and Derby, helps explain it as a site of intense political activity. In 1928, 3000 people took part in a trespass and in April 1932, 400 trespassing ramblers were confronted by gamekeepers and county police. Six trespass leaders were arrested and charged and five were jailed. Anti Semitism (a number of the protestors were Jewish), fused with accusations of communist plots, featured in the trail that received widespread newspaper coverage.

Despite the entrenchment of the landowning elites, the mass trespasses and the Kinder Scout protests played a partial role in the eventual legislative reform and the creation of national parks in the immediate post war period of 'high' social citizenship in Britain. The National Parks and Access to the Countryside Act (1949) oversaw the establishment of ten national parks during the 1950s including the Peak District National Park.

Rural Identities and National Identity

The right to roam movement is one example of an older struggle for entitlements to exclusionary rural spaces. There are other, older stories too of course – the en closure and clearance movements mentioned earlier and of course the protests of the Diggers in the seventeenth century; but of these, the right to roam protests capture the particular sanctification of rural space. The protests and their legislative settlement also demonstrated the place of the rural in narratives of national identity. As was noted above, the intimacy between notions of the rural and the popular and political, national imaginary can be understood as a particularly English phenomenon. However, rural spaces *do* occupy a central place in other narratives of other nations in Europe, in North America, and in other Anglophone countries. However, it is im portant to note that first, these countryside/nation en tanglements are differently constructed when compared to the English situation and second, that the English rural was a colonial export recreated in the rural spaces of the colonized landscapes in Africa and on the Indian sub continent. For example, it is not unusual, in a routine drive through the countryside surrounding the South African cities of Durban and Cape Town, to pass 1930s mock Tudor rural buildings (often pubs and country clubs) and signs advertising Devonshire cream teas. Similar colonial exports of 'the English rural' have been noted in Australia and India too. We return to the question of a more global context in terms of how and in what ways rural spaces relate to national imaginaries.

The relationship between English rural spaces and national identity and sanctification has emerged a key field of theoretical and empirical inquiry within rural identity and otherness debates. The scrutiny of the represen tational work that rurality does in constructing the dom inant (if anodyne) markers of Englishness have tended to cluster around issues of ethnicity and race and the examination of the countryside as a *de racialized* landscape.

The term de racialized is used very purposely to refer to a dual process by which race is denied as a key driver in a discourse that is very much about race. In dominant rural imaginaries and rural spaces, the notion of race is deemed irrelevant as the spaces become hyper whitened as black and other minority ethnic figures and populations are assumed as an absence, as distant, or as uninterested. The attempt to extract or write out race from popular concepts of rurality reveals the extent to which race is not only relevant but actively shapes popular nostalgic narratives as to what the countryside is – an inert, culturally homogenous 'forever England' place. Some of the rural otherness and ethnicity debate has also been concerned with the extent to which black and minority ethnic rural residents and rural visitors experience racism and racialized marginalization and the extent to which rural spaces are underused and perceived as not easily accessible or welcoming to black and minority ethnic populations.

The de racialized rural discourse performs a number of concerning tasks: it denies and invisiblizes the small, scattered but significant and increasing numbers of black and minority ethnic residents, workers, and visitors in rural areas; it severely limits the multicultural sensibility of rural policymakers, service deliverers, and those responsible for rural well being; and it sweeps under the historical carpet the longevity of the colonial–countryside relation – the stately homes built with colonial wealth, their interiors influenced by a 'colonial aesthetic', and their huge grounds planted with exotic trees, flowers, and shrubs brought from colonized countries, and the rural villages and public houses whose names testify to a colonial familarism and, finally it sets up a potent and reiterative narrative in which a dangerous spatialization of ethnicity and multiculture occurs – ethnicity and multiculture 'belong to' and 'are of' urban environments and do not 'belong to' and are 'not of' rural environments. While there are of course constant challenges as well as quantitative and qualitative evidence which counter and disrupt this narrative, it retains a populist and political tenacity in Britain.

There is clearly, from a policy perspective among others, a need to rethink the de racialized rural discourse for two reasons. First, it is a discourse which denies and conceals rural multiculture. Second, the whitening of rural spaces has not been about whiteness *per se*. The whiteness that has been imagined in rural areas is hierarchically constructed as to its social and political desirability. As the discussions above show, whiteness has never simply equated to belonging in racialized narratives. In rural spaces Gypsies, New Age travelers, and the urban working classes have been problematized, outcast, white presences in both historical and contemporary countryside environments and met with local and national hostility which, at times, has culminated in restrictive legislation.

Alongside a privileging of cultural sameness, the English rural discourse has, in particular, been about social order. The countryside idyll and 'village England' represent a highly regulated, socially deferential space. It reflects control: the country house, the immaculate lawn, 'exotic' plants from distant (more dis orderly) places, land ownership, and the country seat as the base of the traditionalist, hierarchical political system of England. Despite the radical and dramatic political changes that have taken place in the governance of rural populations over the centuries, it is the idea of the rural as an orderly and stable space that is called on in times of political crisis and anxiety. During World War II, posters with views across the English South Downs rallied the population to defend the nation; in the early 1980s, the former Conservative government's Home Secretary William Whitelaw, infamously looked to the rural view from his garden for reassurance as urban unrest spread across inner city areas of Britain. Even in the early 1990s, the former Conservative Prime Minister John Major evoked a ruralist set of images as those which would be able to define a Britain in '50 years time'. Conversely, more recent attempts to define Britain through a more urbanist 'cool Britannia' frame have been troubled. In the 2000s, the new Labour Government, with its metropolitan associations, has had a particularly turbulent relationship with the countryside played out, for example, in the hunting ban legislation and the emergence of the conservative Countryside Alliance.

This particular and necessarily narrow imagining of the rural has held a long appeal for the political right – not only for those with conservative tendencies but also for those on the extreme right in Britain. The latter have, with a passing nod to the rurally focused racialized terror activities of the Ku Klux Klan in the United States, increasingly been concerned with attempting to organize around and disseminate via their newsletter 'Land and People', a notion of rural Britain as the last outpost of an imagined Englishness and the need for it to be 'defended' as such. This has been a strategy which has delivered the British National Party some political successes notably winning councilor seats in Worsthorne, a rural village in Northern England, in local elections in the early 2000s.

This de racing of the British countryside presents a set of processes which are unlikely, if not impossible, to transfer to other national settings. In the United States, for example, the history of the Deep South and the history of Native American Indians mean that rural landscapes are deeply entangled with African American culture, racial violence, segregation, and racial inequality. In Australia too, the rural base of Aboriginal settlement, culture, and struggles fundamentally recasts the rural others debates. In postcolonial Zimbabwe and Kenya and post apartheid South Africa, the high profile contestations over white owned farms and land rights again prohibits the sanitization and purification of rural spaces.

Rural Others – A Very British Story?

Much of the rural identity and rural otherness debates have been situated within a very British and sometimes very English context. In many ways, this 'parochialness' is based on the cultural, social, and political cornerstone that the concept of rurality has taken up in the dominant narratives and formations of national identity. The sanc tification, purification, and associated regulation of rural spaces have been about the cultural representations and social constructions of Englishness and, at various mo ments, Britishness. It is both historical and contemporary. The country house was integral to class and political systems in Britain. In Britain, the imagined and material distinction between country and city has been firmly maintained with the help of the Green Belt policy which aimed to protect the countryside and stop the urban and suburban sprawl. At a moment of affluent counter urbanization and rural repopulation, the contestations over rural space in Britain are as intense as ever.

Nevertheless, the specificity of the British case does not mean that rural spaces do not shape and influence other nations' narratives and signifiers of national identity – France may be Paris, but it is also Provence; Scandi navia is particularly understood through a rural esthetic in part derived from the *Allemansrätten* (Swedish) right to roam; the Outback rural spaces of Australia feature highly in its popular representations; and the mid west prairies, the southwestern deserts, and the New England fall foliage are immediate signifiers of the United States. So it is not that other narratives of national identity do not, and have not, incorporated and mobilized rural spaces as representational markers and signifiers. Nor is this to suggest that such narratives do not contain rural folk devils and rural others (as the hostility and animosity toward Gypsies in Eastern Europe and travelers in the Irish Republic evidence). Rather, it is to suggest that those narratives have not constructed sanctified, purified, and regulated rural spaces to the same totalizing extent. Rural poverty (for example in the Dust Bowl of the 1930s; in the photographs of Dorothea Lange; in the John Steinbeck novels *The Grapes of Wrath* and *Of Mice and Men*; and in the Appalachian Mountains) has not been erased from the US national story – indeed to some extent, it is integral to it. However, what does remain absent from the (dominant) transcendentalist 'natural environment' metanarrative is an African American perspective, a nonessentialized/ing identity grounded firmly in the African American rural experience.

Again, reflecting the diversity of rural debates and the politics of rurality, rural researchers have argued that if the New Age traveler controversy has died down in the context of the British countryside it is because many New Age travelers have moved to continental European rural countries. Although these countries are not without trespass laws, policing, and regulation, New Age travelers have not been viewed through the same lens of anxiety nor seen as presenting a level of threat to countrysides that requires legislation and draconian policing.

Conclusion

A focus on rural identity and rural others involves seeking more 'broken' narratives of rural spaces. By this, we mean that the dominant rural imaginaries, hooked around ideas of sameness and stability and conditional inclusion, require theoretical and empirical challenge and disruption. Broken narratives of rurality recognize the fluidity, ambiguity, and heterogeneity of rural spaces. It is important that not only are countrysides of diversity and difference more easily and readily acknowledged, but that this plurality is also appro priately responded to in political and policy terms. It is important to remember that constructions of others and processes of otherness are part of a dialectic defining of not only the different, problem, transgressive subject but also the hegemonic center. A focus on rural identity and otherness involves not only the visibilization of the hidden, contested, and marginalized rural populations and issues, but also the scrutiny of 'uncontested', 'mainstream' rural populations and dominant discourses.

See also: Ethnicity and Resistance, Historical Geographies of; Nation; National Parks; Rural Geography; Rurality and Post-Rurality.

Further Reading

Agyeman, J. and Spooner, R. (1997). 'Ethnicity and the rural environment'. In Cloke, P. & Little, J. (eds.) *Contested Countryside Cultures*, pp 197 217. London: Routledge.
Bell, M. (1994). *Childerley: Nature and Morality in a Country Village*. Chicago, IL: Chicago University Press.
Bunce, M. (1994). *The Countryside Ideal: Anglo American Images of Landscape*. London: Routledge.
Cloke, P., Doel, M., Matless, D., Phillips, M. and Thrift, N. (1994). *Writing the Rural, Five Cultural Geographies*. London: Chapman.
Cloke, P. and Little, J. (eds.) (1997). *Contested Countryside Cultures: Otherness, Marginalization and Rurality*. London: Routledge.
Davis, M. (2006). *Planet of Slums*. London: Verso.
Dwyer, O. J. (1997). Geographical research about African Americans: A survey of journals, 1911 1995. *Professional Geographer* 49(4), 441 451.
Floyd, M. F. (1998). Getting beyond marginality and ethnicity: The challenge for race and ethnic studies in leisure research. *Journal of Leisure Research* 30(1), 3 22.
Hetherington, K. (2000). *New Age Travelers: Vanloads of Uproarious Humanity*. London: Cassell.
Neal, S. and Agyeman, J. (eds.) (2006). *The New Countryside? Ethnicity, Nation and Exclusion in Contemporary Rural Britain*. Bristol: The Policy Press.
Sibley, D. (1996). *Geographies of Exclusion*. London: Routledge.

Identity Politics

A. Kobayashi, Queen's University, Kingston, ON, Canada

Glossary

Identity A set of socially constructed traits around which members of a group organize a sense of belonging.

Identity Politics The political mobilization of a group to achieve recognition and redress of historical oppression on the basis of group identity.

Irredentism A movement on the part of the nation-state to incorporate within its boundaries all territory to which it has once held claim, often involving the attempt to incorporate minority ethnocultural groups who claim a separate national identity.

Multiculturalism State recognition of group identities as a basis for plural, as opposed to monocultural, societies.

Politics of Difference A contested term that for rights-seeking groups refers to political movements that use recognition of shared historical experiences as a basis for social recognition; conversely, assimilationists view the politics of difference as the basis for separatism.

Early Interest in Identity Politics among Geographers

The ideologically laden term 'identity politics' refers to social movements to gain recognition of historically oppressed ethnocultural or racialized groups. Some theorists also refer to such movements as a 'politics of difference'. Geographers' interest in identity politics began with the development of 'radical' geography during the 1960s, at the height of the American Civil Rights Movement. Radical geography was (and is) mainly a critique of established theories of the state and economy; but William Bunge took radical geography to the streets by initiating the Detroit Geographical Expedition in collaboration with local organizers from the African American community. Although the collaboration lasted only about a year, it focused attention on the possibilities for geographers to work with local activists in advocating for social change. Several similar projects developed throughout the US and Canada, although they tended to be much more strongly focused on issues of antipoverty movements and activism around the definition of public space (both important) than on questions of racial and ethnocultural equality. Despite the strong potential of such early work, radical geography approaches influenced by structural Marxism did not materialize into widespread analysis or involvement at the local scale; and humanistic geographers of the 1970s and 1980s tended to emphasize either individual experience or the creation of symbolic cultural landscapes. Neither approach made issues of identity politics a priority, despite the extremely important place of identity politics in the contemporary world. Over the past two decades, however, there has been renewed and enthusiastic interest both for understanding identity politics and for encouraging research and academic activism in local communities.

The Development of Politics of Identity since the Civil Rights Movement

The politics of identity was an issue of major public concern through the 1970s and 1980s, particularly in white settler societies such as Canada and the United States. These decades saw the emergence of movements to obtain redress for historical oppression across a huge range of issues mobilized using many political positions and strategies, but they fall into two major categories: those that seek a radical separation from the dominant society, and those that seek recognition within dominant society. The first group aims to establish alternate forms of nationalism, while the second appeals to a range of meanings of multiculturalism.

The Politics of Radical Separatism

Radical separatism depends on a sense of identity that is fundamentally irreconcilable with the larger nation. While such movements vary (both by group and historically) over the degree of autonomy sought and in the extent to which they are willing to use violence when other means are unsuccessful, they have in common a belief that the only means of overcoming oppression is removal from the hegemony of the oppressor. They occur in historical situations of colonialism, territorial annexation, or wholesale movements of peoples (e.g., through slavery). They also vary in the extent to which separatism is a means to an end or an end in itself, and in the extent to which separatism involves territorial claims. The geographical, spatial context for radical separatist movements, therefore, is crucial to an understanding of the possibilities for addressing their claims.

The geography of radical separatism based on territorial claims is extremely mutable and varied. Some historic movements, such as Quebecois separatism or republicanism in Northern Ireland, have shifted over time from the blunt politics of opposition to a practice of transformation from within. While no less adamant in their claims to national identity, such groups have expanded their strategic actions to use more conventional political means, such as party politics, to achieve their ends. Quebecois separatists, for example, refer to the Quiet Revolution of the 1960s and 1970s as the transformation to a secular nationalist society in which institution building – language, cultural practices, the arts, the emergence of a national political party, the Bloc Quebecois – is the basis for cultural autonomy within Quebec borders.

Such territorially based identities are seldom hegemonic, however, and are often internally contested. For example, in the case of Basque groups in Spain, identity based institution building is the basis for an autonomous region in which a number of Basque political groups now compete for power. The Euskadi Ta Askatasuna (ETA), however, shuns those institutions, still espouses radical separation, and is willing to use violence to achieve it. In other regions, such as the former Yugoslavia, politics of identity taken to their furthest ends resulted in widespread violence and radical separation and dismantling of a former national polity. There are many other examples throughout the globe – the Kurds in Iraq and Turkey, the Tamils in Sri Lanka, Palestinians in Israel, and indigenous groups in both developed and lesser developed nation states – where ongoing violence defines attempts by historically oppressed groups to establish separatist states based on ethnocultural identity, and where official state violence to suppress such movements takes a range of irredentist practices, from genocide and ethnic cleansing to the imposition of strict policing regimes and the denial of human rights.

Geographers have struggled to develop adequate theories for understanding the politics of identity around radical separatism, especially in situations of extreme violence. Historical materialist accounts of the conditions of oppression under which separatist movements have arisen can explain the desperation of nationalist groups in the face of brutal colonial practices and ongoing denial of the most fundamental human rights (which have occurred to one degree or another in all of the cases cited above). They can also explain the liberating power of resistance to oppression, whether through violence or through more conventional political strategies. But they do not condone violence as a means of liberation. Recent attempts to assess the most entrenched and violent irredentisms, by geographers such as David Delaney, Deryk Gregory, and David Harvey, have stressed the importance of going beyond territorial oppression, of recognizing the extent to which contemporary identity based territories are implicated in global formations of capital, global redistribution of resources, and global political coalitions, and of reimagining the meaning of the nation in light of colonial and other hegemonic forms of domination. While several geographers have cautioned about the danger of romanticizing separatist groups 'just because' they have occupied a historic position of marginality, we still face the question of whether it is a radical politics of identity (whether of the dominant or subordinate group), or only a 'violent' politics of identity that needs to be overcome in situations of conflict.

Not all forms of radical identity politics are territorially based. A major example is that of the Black Panthers, who emerged in the 1960s to advocate for African Americans' right to form their own destiny in order to overcome poverty and gain access to education, health, and other basic needs. The movement failed within a decade due, among other things, to the death or imprisonment of its leaders, internal ideological conflicts, and the actions of law enforcement to restrain its activities. Recent resurgence of interest in the Black Panther movement on the part of academics, including geographers, raises questions about the possibility for non territorial nationalist movements to establish a sense of collective identity sufficient to effect change, and about the logistics of scale in organizing identity based claims when traditional notions of territory are not the basis for mobilization. Strategies for political organization, as well as the rationale for claiming human rights, change dramatically when not territorially based.

The memory of the Black Panthers is also linked to emerging forms of radical identity claims. In the US, several social movement groups, seeing a failure over the past four decades to improve the position of African Americans, are claiming a revival of the Black Panther organization premised on peaceful mechanisms for social change. Moreover, similar claims, especially among Islamicist movements, are increasingly diasporic, transnational, and global against the global gathering of strength of Islamophobia, which is also often violent. These movements are by and large premised on a concept of peaceful activism, although some are willing to resort to violence. Diasporic identity politics are considerably more ideologically varied than are movements that lay claim to specific territories. Paradoxically, they are both more potentially powerful because of their global scale, and more difficult to organize because of their spatial diffusion. Surely, there has never been a more important or compelling challenge to political geographers than to understand the implications of emerging diasporic identity politics.

Identity Politics in a Context of Multiculturalism

Identity politics set within a pluralist, or multicultural, context seek recognition of human rights rather than separation. Although they are usually much less likely to occur in violent situations, this does not mean that they are any less fervently sought, or that the acts of oppression against which they seek redress are any less oppressive. They include, for example, most of the American Civil Rights Movement, including the movement for slavery compensation, the movements of Japanese Americans and Japanese Canadians for redress for human rights abuses during the 1940s, organizations against Islamophobia, organizations of urban indigenous peoples, and struggles for the rights of Latino/Latina migrant workers. In almost every case they occur against a historic background of colonialism that involved the migration of people to places where they could not exercise territorial claims. Most racialized minorities in advanced industrial countries, for example, first arrived (and continue to arrive) in those countries as slaves, indentured, or very cheap labor. It is ironic that their presence has created the conditions of pluralism that have led many states to adopt policies of multiculturalism. Moreover, nonterritorial claims of (dis)placed cultural minorities are often set against the territorial claims of nationalist or indigenous communities, especially in the white settler societies such as the US, Canada, Australia, and New Zealand. The challenge of reconciling the liberal notion of individual multicultural rights with group rights in such settings is one that has yet to be met by policymakers, political philosophers, or social movement groups themselves.

As state policy, the idea of multiculturalism first emerged in Canada in the 1960s and became official government policy in that country in 1971. Australia followed suit in 1973, and several European states, such as Sweden and the Netherlands, subsequently adopted similar state policies. The concept is a liberal democratic one of encouraging individual identity as a member of a particular ethnocultural group, and in advancing public institutions to foster equality. Many policymakers have advocated multiculturalism as a means of overcoming the practice of assimilation, particularly of recently arrived immigrant groups, but without thereby hindering the ability of such groups to achieve full and equitable social participation. Critics of multiculturalism have been less sanguine over the power of multiculturalism to achieve social cohesion, but for very different reasons.

Critics have generally taken one of two polarized positions. The illiberal, monoculturalist position holds that nations should maintain a dominant cultural identity to which Others should adapt or assimilate. For them, identity politics around multiculturalism are disparaged as addressing the needs of 'special interests', rather than those of the dominant society. Such critics also fear that regimes of ascriptive primordialism on the part of ethnocultural groups will entrench difference and encourage fundamentalism, extremism, and antimodernism. They maintain that only assimilation to a dominant culture will overcome primordialism. This position is currently gaining strength in a number of advanced industrial countries of Europe and North America.

The contrary view opposes multiculturalism on the grounds that it is by definition separatist and as such prevents marginalized groups from accessing full equality rights enjoyed by members of the dominant group. These critics fear that group identities will be 'Balkanized' in relation to the dominant group. They are skeptical of the willingness or ability of the state to address any but the most general of claims to equal recognition, but they also caution against the coercive power of the group to enforce a singular identity at the expense of other facets of subject positioning such as gender or class.

Both positions, however, rely on an essentialized understanding of ethnocultural identity as Other, and often depict culture as static, immobile, and necessarily conservative. Ethnocultural groups themselves have also resorted to essentialized concepts of identity – whence the term 'identity politics' – in their struggles to attain social justice through recognition. Many writers have pointed out the paradox of identity politics: that making claims based on historical injustice reinscribes the state of difference that was the original basis of injustice, thus preparing the ground for further oppression. Using the liberal democratic concept of multiculturalism as a basis for equality, therefore, runs the risk of reproducing this fundamental tension between identity and difference in irreconcilable ways.

Nowhere is this tension more pronounced than in contemporary debates over the expression of Muslim identity now occurring in various ways throughout traditionally non Islamic world. In some advanced capitalist countries such as France and the Netherlands, secular multiculturalism has provided the ironic and somewhat disingenuous rationale for curtailing certain public Islamic expressions, such as female dress. In other places, such as Canada and Australia, while there has been no official pulling back of the concept of multiculturalism, Muslim identity is under increasing public attack, and discourses of ascriptive premordialism are increasingly dominant. In the province of Quebec, for example, where territorial identity claims to cultural separatism are often at loggerheads with claims for multicultural recognition on the part of racialized and religious minorities, there is currently underway a public commission on religious accommodation. Public discourse around accommodation overwhelmingly asserts a dominant, normative view of Quebecois culture, with debate occurring over how far the dominant group is willing to go to accommodate difference.

Nonetheless, ethnocultural mobilization has often used the concept of multiculturalism to transcend, rather than reinforce, notions of cultural separation, and such movements can involve considerable negotiation of the terms of cultural identification within groups. Most diasporic groups, within which there inevitably operate a range of political views, are themselves sites of lively debate over what their collective identity consists of, over the meaning of cultural change that is often accelerated by the very experience of diasporic movement, and over the means by which human rights, including the rights of women, can be achieved in a wide variety of national contexts. In many Western countries, the minority groups asserting their own identity have pushed for human rights that affect all citizens. In some cases, coalitions of minority group movements that may at other times seemed odd partners have advocated for social change and, in so doing, have established a basis for reimagining multicultural societies in which ethnocultural identities are not silos of isolation, but expressions of vibrant social interaction.

One example of recent trends is the increasing incidence of transnationalism. Whereas transnationalism as a set of practices that retain links between one or more disaporic generations and a country of origin is by no means a new phenomenon, it is widely believed that in the context of globalization, including global communication and transportation, globalized consumption practices, and related increases in international migration, ongoing interactions between migrants and their places of origin have become stronger since the latter part of the twentieth century. There is now a burgeoning literature on the topic that includes a variety of transnational forms among temporary labor migrants, permanent settlers, and subsequent generations. These connections are culturally, demographically, economically, and politically complex. But what they show in part is that the complexities of negotiating individual and family lives across transnational fields are strongly connected to those of negotiating conditions of citizenship. Contemporary transnationals, many of whom are also very cosmopolitan, have the capacity to participate in social movements that address issues of human rights and social identity and, further, the conception of human rights and social identity is subject to considerable change within the fluid social dynamics of transnationalism, transcending static or primordial notions of citizenship rights.

The increasing practice of holding dual or multiple citizenships also affects the ways in which social identities are transformed and emphasizes the extent to which questions of citizenship have become globalized. For example, when violence erupted in Lebanon in 2006, many dual nationals from countries such as Canada and Australia that recognize multiple citizenship were assisted by their governments to evacuate. The evacuations sparked fierce debates in a number of countries about the value of dual citizenship, the rights of citizens, loyalty to more than one national identity, and the obligations of states to serve the needs of their citizens outside national borders. These debates bring out many geographical issues, not only about the changing relationship between cultural/political identities and territory, but also about the ways in which changing ways of being citizens can transcend older notions of multiculturalism. Many dual nationals feel that their allegiance to both countries is strengthened by an expanded concept of place of residence, and that their experiences contribute to a rich cultural diversity in both places, especially when children born with rights to more than one citizenship exercise those rights through their own international activities. Dual citizenship, and attendant transnational activities, thus works to counteract cultural primordialism.

Such citizens are increasingly involved in social movements that transcend national borders. Transnational social movements (TSMs) have been a major feature of globalization, and transnational identity politics are no exception. Diasporic cultural groups, more than ever, connected by electronic communication as well as the ability to move with greater ease between international destinations, address issues of citizenship at a number of scales, linking national and international citizenship issues. Such social movements organized around identity politics range from movements to bring peace to areas torn by violent conflict, to democratic movements to shift autocratic state apparatuses or to protect minority groups subject to state oppression, to smaller scale projects that include international adoption programs or participatory development projects.

Critics will point out that transnational identity politics also include groups mobilized to effect violence, especially since 11 September 2001. It is important to recognize that the power to effect all kinds of social outcomes transcends borders. The geography of social movements is such that the political outcomes are contingent, and very much rooted in the historical conditions that produce in some cases violence and in others positive social change. The suppression of minority groups is also ongoing and involves the mobilization of state apparatuses capable of delivering widespread violence. Furthermore, while armed violence is only one form of suppressing social movements, the violence of normative ideologies, in the form of racialized and colonial ideas and practices, is a major aspect of the politics of difference. We need to understand, therefore, not only the variety of ways in which social movements work in plural societies, but also the variety of ways in which states address their concerns.

Overcoming the Paradox: Imagining Identity Politics Beyond Identity

Geographers began to direct their interests toward questions of identity politics during the early 1990s through a renewed interest in studying the geography of racism through a post structural lens. While notions of a biological basis to the concept of 'race', and primordial definitions of culture, have been increasingly discredited, post structur alism allows an understanding of the ways in which essen tialist beliefs are constructed, or imagined, through the deployment of dominant discourses that normalize the dominant (white) body and invest power in white cultural formations. All identities are thus already constructed through a historical process such as racialization.

Social constructionism portrays the ontologically unstable status of social categories, but it does not in itself provide insight on the 'politics' of recognition.

The study of identity politics in different forms and in different parts of the world has led geographers to a number of insights about understanding political strug gles. The politics of identity almost always originate in conditions of colonialism or territorial conquest and involve the historic displacement of minorities or the subjugation of minorities in place. Contemporary at tempts to overcome historic oppression, whether waged through violent means or through accomodationist practices, such as multiculturalism, invariably run the paradoxical risk of reinscribing difference and the iden tity of the irremediable other.

Overcoming essentialism is not simply an academic exercise. That is, it is not only essentialism as a theoreti cally inadequate – and inaccurate – explanatory frame work that is troubling, it is the political, ideological use of essentialism to mobilize discourses of domination. It is about the imagination of sameness and difference, and about the lengths that human beings are willing to go to reinforce such normative visions. Recently, geographers and others have begun to theorize ways of reimagining human relations so that dominant groups and minority groups alike can get past the trap of essentialized thinking. By doing so, they also theorize the extent to which social change is possible. But if the refutation of essentialism is old news academically, it is certainly very much a part of political movements, not only because ideas are slow to change but also because certain ideas retain huge, indeed hegemonic, political power. It is, above all, against such hegemonic power that identity politics have emerged and social movements of all sorts have resisted.

See also: Activism; Cosmopolitanism; Cultural Politics; Difference/Politics of Difference; Ethnicity; Human Rights; Multiculturalism; People's Geography; Race; Racism and Antiracism; Resistance; Transnationality; Whiteness.

Further Reading

Cohen, R. (1998). Transnational social movements: An assessment. Working Paper 98 10. Oxford: Transnational Communities Research Programme. http://www.transcomm.ox.ac.uk/working%20papers/cohen.pdf (accessed in June 2008).

Fossam, J. E. (n.d.). Identity Politics in the European Union. ARENA Working Papers 01/17. http://www.arena.uio.no/publications/wp01 17.htm (accessed in July 2008).

Harvey, D. (1996). *Justice, Nature, and the Geography of Difference*. Malden, MA: Blackwell.

Heyes, C. (2007). Identity politics. *The Stanford Encyclopedia of Philosophy* (Winter 2007 edn.) Zalta, E. N. (2nd.), forthcoming. http://plato.stanford.edu/archives/win2007/entries/identity politics (accessed in June 2008).

Jordan, G. and Weedon, C. (1995). *Cultural Politics: Class, Gender, Race and the Postmodern World*. Oxford: Blackwell.

Pulido, L. (2006). *Black, Brown, Yellow and Left: Radical Activism in Los Angeles*. Berkeley, CA: University of California Press.

Pecora, V. P. (ed.) (2001). *Nations and Identities*. Oxford: Blackwell.

Young, I. M. (1990). *Justice and the Politics of Difference*. Princeton, NJ: Princeton University Press.

Young, I. M. (2000). *Inclusion and Democracy*. Oxford: Oxford University Press.

Ideology

K. R. Olwig, Swedish University of Agricultural Sciences, Alnarp, Sweden

Glossary

Anarchism Anarchism, like 'ideology' and 'romanticism', often carries negative connotations. In this case anarchy is often understood to be a total lack of control, and anarchists are seen to be bomb-throwing terrorists. Anarchy, however, simply means lack of a formal system of government, such as a state, and anarchism is a political ideology, opposed to state coercion, that advocates the free economic and political cooperation of individuals in communities. It has significantly influenced the development of the cooperative and labor movements. Some adherents of anarchism have been exposed to violence, others as in the case of many other activists, have advocated violence in response, but the basic thrust of anarchist thought is that peaceful cooperation is the normal mode of human organization, and many anarchists have been pacifists. Anarchist geographers have thus advocated education as an alternative to coercion as a means of fostering social development through cooperative endeavors rooted in an understanding of the geographical environment's potentialities.

Ancien Régime It is the name specifically given to France's political and social system before the 1789 revolution. It also refers, by extension, to any outmoded system or administration, and it thus can be applied more generally to regimes similar to that overthrown in France. The connotation of outmodedness that adheres to this term reflects the negative position of the 'moderns' toward the 'ancients' (see entry 'ancient').

Ancient The term 'ancient' was counterposed to the 'modern' in an influential late seventeenth century literary 'quarrel' taking place in France and England between, on the one hand, proponents of the belief that the ideas expressed in ancient Greek and Roman classics were the ideal source of literary inspiration, and, on the other hand, the thinkers who believed that the development of the human mind, as evidenced by the progress of science, made contemporary works superior to those of the ancients. This debate foreshadowed the 'modern' notion of science-based progress that developed during the 'Enlightenment'.

Enlightenment An eighteenth-century Western European intellectual movement emphasizing the ?'light' of reason, rationality, and science in the understanding of human culture and the natural ?world, and in encouraging its progress. Proponents of the Enlightenment were thus critical of Europe's ?*ancien régimes* as dominated by the aristocracy and justified by ideologies rooted in metaphysics and religion.

Modern (See entries 'ancient' and '*ancien régime*'.)

Romantic Movement This was a late eighteenth- and early nineteenth-century movement in philosophy and the arts. It opposed classicism's cultivation of general, formal, rational, and unchanging models derived from the ancient literary classics and the fixed geometry of classical architecture, which it identified with the *ancien régimes* of Europe's aristocratic courts in capital cities. The romantics looked, instead, for inspiration from cultures and societies seen to be more natural because they lived at some distance from the cosmopolitan core, as in the mountains of Switzerland or Scotland, or were distant in historical time, as in medieval Europe. Instead of the rationality of geometric models, the romantics emphasized the esthetics of organic growth and progress, and individual sensibility, identified with nature.

Ideology

Geography might be described either as a discipline en gaged in the constructive production of ideology in the form of geographical ideas, or as a discipline that seeks to promulgate, critique, or even transcend ideology as a politicized form of consciousness. The difference depends upon changing approaches to geography and how one de fines ideology. This article will begin with an examination of the discipline during the formative early stages of its modern history at the beginning of the nineteenth century, when geography was positively engaged in the production of ideology in a broad educational sense. It will then examine how geography, in the course of the nineteenth and early twentieth century became engaged with political ideologies, such as nationalism, liberalism, imperialism, anarchism, and Marxism, giving rise in the process to what is now known as political geography. At this time, pro ponents of a particular ideology, such as Marxism, might be critical of another ideology, such as that of imperialism, but it was not until the mid twentieth century that geographers sought to transcend ideology entirely by replacing it with a supposedly 'objective', 'spatial science'. The late twentieth and early twenty first century, however, has witnessed a return to a broad engagement with ideological issues. Political geography has particularly enjoyed a renaissance rooted in a critical understanding of geography's past and present ideological uses and misuses that is rooted in postcolonial theory, critical Marxism, post structural philosophy, and a renewed understanding of the ideological role of landscape.

The Enlightenment Origin of Ideology and Geography

Modern geography grew out of the same Enlightenment impulses that produced the discipline of 'ideology' as a branch of knowledge concerned both with the origin and nature of ideas and the role of such ideas in transforming society. Destutt de Tracy (1754–1836), and his circle of French Enlightenment philosophers, who pioneered the science of ideology, thus believed in the power of ideas not just to explain the world, but to change it – especially through education. Ideological knowledge is therefore not neutral and this makes it open to controversy, both if it is seen to influence behavior (making it potentially dangerous) and when it is not seen to do so (making it superfluous). The resultant divided opinion of ideology is evident in the contradictory definitions typically given in dictionaries, which shift abruptly from the matter of fact: 'a manner or the content of thinking characteristic of an individual, group, or culture' (such as that of the medical profession or a social class), to the depreciating: 'visionary speculation – idle theorizing; often: an impractical theory or system of theories'. They likewise can shift from the neutral 'integrated assertions, theories, and aims that constitute a sociopolitical program', to damning as 'an extremist sociopolitical program or philosophy constructed wholly or in part on factitious or hypothetical ideational bases'. Finally, one also finds academic definitions, such as 'a systematic scheme or coordinated body of ideas or concepts especially about human life or culture' – a fitting description of a discipline such as geography! Because ideology is defined in so many, contradictory, ways, the meaning of ideology must be understood in the context of the praxis in which it is used.

Geography as Ideological Praxis

The French Enlightenment ideologist philosophers put their knowledge of ideology into practice through education about the time of the 1789 French Revolution, when the *ancien régime* was waning. They were particularly interested in the relationship between ideas and sensory experience, and thus, more broadly in the relationship between the ideational and the material realms, particularly as revealed through science. Modern geography grew out of the Enlightenment's broad concern with ideology and education as pioneered by the ideologists, and geographers shared their overarching interest in the link between ideas, empirical knowledge of the material world, and human progress. At this time, education in the sciences and technology was 'modern', progressive, and controversial, in contrast to traditional conservative Latin school education, which privileged the ideas found in the Bible and the 'ancient' Greek and Roman classics.

Encyclopedic Geography

The foundational geographical works of the early nineteenth century were encyclopedic in character, in the tradition of the eighteenth century French Enlightenment 'philosophes' whose monumental *Encyclopaedia, or Classified Dictionary of Sciences, Arts, and Trades* inspired the ideologist philosophers. The suffix, 'graphy', comes from the Greek *graphein*, meaning 'to write', and these geographers were, by and large, 'writers' more than scientists. Geographers produced geography, as a form of knowledge, through their writing praxis.

An example of a writer who literally helped produce the discipline of geography through his writing would be Conrad Malte Brun (1775–1826). This Danish aristocrat and poet fled in 1799 into exile in France as a consequence of his satiric writing in the spirit of the French Revolution's critique of the *ancien régime*, and because he was linked through his pamphleteering to an early, controversial, strike by craftsmen. In France he worked as a journalist and became one of the literati associated with Madame Germaine de Staël (1766–1817) who helped set in motion the French Romantic Movement. 'Romanticism', like 'ideology', is often identified with the visionary and the impractical, but at this time its revolutionary impetus lay in its almost sociological concern with historical and social reality, and its cultivation of the culture, language, and landscapes of the common people of different regions, as opposed to the language and values of the court and nobility of the capital (**Figure 1**). This movement, in turn, helped provide the ideological foundation for the nation state, whose geography was seen to be rooted in the culture of the land and people, as opposed to the dynastic identities of the monarchs. There was thus an ideological line of development from Malte Brun's literary and social concerns to his pioneering work as writer and editor of the first six volumes of the vividly written encyclopedic *Précis of World Geography* (1810–29) and his engagement as founder and first secretary of the Société de Géographie de Paris (1821).

Ideology as Worldview

The German geographer Alexander von Humboldt (1769–1859) became well known as a natural scientist and explorer. As a geographer, however, his world fame lay in his writing, as exemplified by his five volume *Cosmos: A Sketch of a Physical Description of the Universe*. The title *Cosmos* harks back to a Renaissance word for geography, 'cosmography'. A cosmography is essentially a 'worldview', and Humboldt and his geographer colleagues were literally working to produce a global, scientifically informed

Figure 1 Swiss landscape: This cover from a 1911 Danish elementary school geography textbook shows a landscape with a Swiss village, surrounded by mountains and glaciers, with a woman and man doing farm chores while another man is perched on a cliff with a rifle. Though close to nature, these Swiss also have constructed a modern railroad that climbs daringly up a steep mountain, over a bridge, and through a tunnel. The text tells us that the Swiss are "spirited and daring people, who love their land and their freedom, which they have defended with great bravery." Switzerland had roughly the same population as Denmark, and was a microcosm of Europe with its multiple languages. Europe in general, with its cool, mild climate, is described as 'the most important continent' and 'the center point for active human life' with the greatest concentration of people.

view of the world. The idea of the worldview was related to the notion of picturing, by which the world is viewed as a picture or 'sketch' in which the differing elements of the landscape scene are framed and organized according to a particular perspective, such as that of science. Humboldt took this notion literally by bringing artists as well as cartographers with him on his scientific expeditions, and by using pictorial language.

Another word for worldview is ideology, and picturing is especially tied to ideology through the concept of the 'idea', which is traceable back to the ancient Greek word *idein*, meaning 'to see'. Throughout Europe, at this time, education and picturing were linked, but this was particularly the case in Germany, where the word for picture, *Bild*, and that for education, *Bildung*, resembled one another. Geography was therefore not just engaged in education, it embodied education through its emphasis upon graphic, pictorial languages of representation. Education and language, furthermore, were major concerns of Alexander's brother and confidant, Wilhelm (1767–1835), a linguist and educational reformer who helped found the modern research university. The concern with the foundation of knowledge is particularly notable in the work of another contemporary German geographer, Immanuel Kant (1724–1804), who is better known for his philosophical concern with 'epistemology', or the study of how we know the world.

Geography and the Politics of Ideology

Geography and Anarchist Ideology

The French polymath geographer Elise Reclus (1830–1905) was engaged with ideology as one of the world's leading anarchists. He produced the 19 volume *The Earth and Its Inhabitants* in which he was particularly concerned with how people adapted the environment to their collective needs, thereby transforming both the environment, their spatial horizons, and themselves – for example, by turning riverine barriers into highways for physical and social communication. Reclus focused, therefore, not upon 'nature', with its unchanging laws, but upon what he termed the 'geographical environment' which changed through time as communities learned to work together to develop the environment's potentiality. This idea was of ideological importance because it suggested that the barriers to free community organization, constituted by the contemporary coercive state, might someday be replaced by freer cooperative forms of organization. Reclus' ideology was supplemented by that of the Russian geographer and leading anarchist, Peter Kropotkin (1842–1921), who argued that even in nature, biotic communities tend to cooperate in the use of particular habitats, rather than engage in a Darwinian struggle for the survival of the fittest. Like Malte Brun, both Reclus and Kropotkin spent years in exile due to their political ideologies.

Ideological Delusion

Napoléon Bonaparte (1769–1821) was apparently the first to use the term ideology in the negative senses of impractical theory and an extremist sociopolitical program. Napoléon, a man of action, tired of the ideologists whose philosophizing, in his opinion, favored demands for liberal democracy, crippling the military will in the process. The positive and negative images of ideology, like those of Romanticism, are thus two sides of the same coin, depending upon the point of view. One man who joined Napoléon's cause, after some hesitation, was Malte Brun, who put his skills to use producing a geo graphical basis, as in his work on Poland, for the national political consolidation of Europe within Napoléon's empire, making him a scientist in Napoléon's eyes. Today, Malte Bruun has come to be regarded as Napo léon's ideologue. This Janus like character of ideology, by which one person's science becomes another's ideology (and vice versa), is particularly apparent in the critical approach of Karl Marx (1818–83) and Friedrich Engels (1820–95), the founders of Marxist or communist ideol ogy. It is important to dwell upon their ideas because it is largely via Marxism that the concept of ideology *per se* plays a role in contemporary geography.

Marxism and Ideology

Marx and Engels were, like the French ideologues, both theoreticians and socially engaged activists, and like them, they were also interested in the relationship between the ideational and the material realm in the context of a will to not just study the world, but change it. For them, ideology was not simply an abstract concept, but the product of a political practice. They stamped anarchist ideology as utopian idealism because they be lieved that the state would be necessary to the transition to communism. Some of the most virulent ideologies of the time, to which both anarchism and Marxism were opposed, were those of nationalism, imperialism, and liberalism. The proponents of these ideologies anchored them in apparently timeless principles of nature, binding individuals by blood and race to their nation and its territory, and engaging them and their nation in a ceaseless Darwinian struggle with other individuals and other nations and races. To Marx and Engels, however, these ideologies were the outcome of developments in the material basis of society (e.g., industrial production) that facilitated the triumph of the bourgeoisie capitalists over the nobility. These ideologies were thus reflections of bourgeois interests, forming an ideological super structure to the material base of society that was the ultimate motor of change.

For Marxism, it is not the ideas and values of liber alism, drawing inspiration from a Darwinian nature or a Protestant God, that created the basis for industrial capitalism, but changes in the mode of production. To believe otherwise is to become the victim of 'false con sciousness'. Ideology becomes a particularly insidious form of false consciousness, from the Marxist point of view, when it is not recognized as being an ideology, but rather regarded as some form of objective, natural, sci entific truth, transcending time and history. Practicing Marxists thus tend to be conscious of their ideology, which they hotly debate among themselves, whereas others in their view have an ideology, but are not con scious of it. Marxist ideology is not false consciousness in the Marxist view, because it sees itself as a body of ideas that is dialectically engaged in a historical and material process transforming both the mode of production and its ideas.

Education and 'Natural' Ideology

The ideological character of geography is exemplified by the importance of education for its growth as an aca demic discipline at the universities. The nineteenth century emergence of public primary and secondary education typically included geography as a core subject, and the need for teachers and competent school texts stimulated the growth of university geography. Geog raphy was important to the ideal of creating a general, educated citizenry capable of taking on the responsibility of representative democratic government.

Education is by no means a neutral activity, and the educational systems that developed in the various nation states were intended to foster a loyal, patriotic national citizenry that shared national antipathies to foreign en emies and supported national and European imperial and colonial activities. The geography textbooks of nine teenth and early twentieth century thus often included blatant examples of nationalistic and imperialistic ideol ogy, supported by ideologies arguing for European social and/or racial superiority (**Figure 2**). The hidden ideo logical power of the texts lay especially in their structure, reinforced by maps and landscape drawings. The foun dation of each nation is the theory of 'environmental determinism' in which it is the physical landscape of a particular territory, which then provides the 'natural' framework determining the culture and economy of the nation (thereby making diasporic peoples, like the Romany or Jews, seem unnatural) (**Figure 3**). Textbooks throughout Europe and America often included the baldest of national and racial stereotypes. Dark skinned people from warm countries were described as hot blooded, lazy, and easy to subjugate, whereas pale skin ned people from colder northern, or mountain, climes were depicted as cool minded, diligent, and freedom loving. Like any other natural organism, the nation was presented as requiring *Lebensraum* ('living space'). Indeed, the more vital a nation was, the more space it required in imperial competition with neighboring nations (**Fig ure 4**). Friedrich Ratzel, Rudolf Kjellen, and Halford Mackinder are among the geographers who played an important role in shaping the discourse of political geography at this time.

Carl Sauer (1889–1975), who founded the department of geography at Berkeley, California, in 1923, became a leading critic of environmental determinism. He was inspired, in part, by the American geographer and phil ologist George Perkins Marsh (1801–1882), who argued

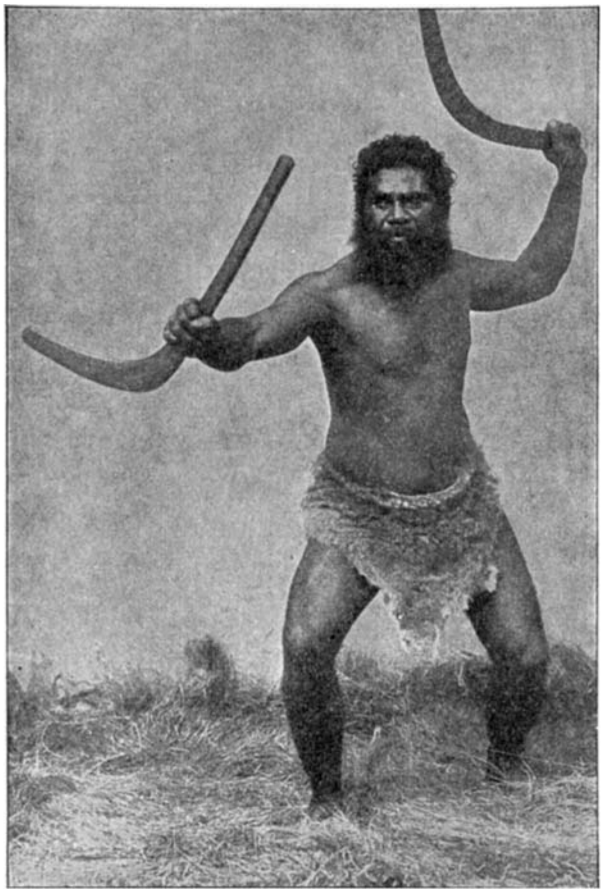

Figure 2 'Savage' with boomerangs: The section on Australia, from the 1911 Danish textbook, informs us that the expansive European population makes up the majority of its 5 million people. The interior, however, is hot, flat, and dry, and here live 60 000 native 'Austral negroes', like the man with boomerangs shown in the photo. According to the text they are "very inferior and raw people. They use neither clothes nor permanent dwellings. The wife is won by predation, and she is the man's slave; and as a symbol of this the man bites a finger joint from her left hand. Many children are born, but only 2 3 of them are allowed to live. The others are killed while they are babies, and devoured." Without mentioning imperialism and colonialism, which was both national and pan-European (Danes were among the settlers in British Australia), the 'facts' and illustrations in the text nevertheless provide ample justification for imperialism's civilizing mission. The quoted text appears to be based largely on an 1885 handbook in geography that was written by the first professor of the geography department at the University of Copenhagen, Ernst Løffler, who espoused environmental determinism.

that society was not determined by nature, but rather both transformed and destroyed nature (**Figure 5**). Marsh helped provide the ideological underpinnings for a 'turn of the century' conservation movement in America and Europe with his 1864 book *Man and Nature: Or, Physical Geography as Modified by Human Action.* By furthering Marsh's concern with the role of culture and ideas in shaping the landscape, Sauer, in turn, helped foster the mid twentieth century environmental movement. Sauer's broad cultural geography, however, was eclipsed to some degree by a post World War II effort to narrow geography to a social science of space and regions (**Figure 6**).

Space and the 'End' of Ideology

World War II was also a war of ideologies, including nationalism, fascism, and communism, and the reaction against war stimulated a reaction against ideology, particularly in America. Geography's theoreticians, such as Richard Hartshorne, became academic discipline builders, advocating the elimination of the historical study of society/nature relations and, with it, the study of landscape, both of which lay at the heart of pre war ideological debate. They argued instead for a geographical science of regions and space. A new generation of scientifically oriented geographers called for a 'quantitative revolution' in which geography would be based on the 'spatial analysis' of 'locations' in space defined as in physics. This new geography was geared to social physics and functionalist spatial planning, and was, hence, highly academic and technocratic in orientation, and based on the 'scientific' principles of positivism.

In this new geography, vivid and pictorial writing for a broad public and for educational purposes was downplayed in favor of mathematical language. These geographers concentrated on making quantified analyses for a narrow circle of academics, technocrats, and planners. This new orientation created a wedge between academic geography and general education that weakened the position of geography in the schools. It also distanced academic geography from the various national geographical societies that cultivated a broad contact with nonacademic segments of society. John K. Wright (1891–1969), at the American Geographical Society, reacted against this tendency, countering specialization with a call to establish a subdiscipline called 'geosophy' that would deal with the geographical knowledge of all manner of people, past and present.

The Return of Ideological Consciousness to Geography

Back to Nature (and to Gender): The Legacy of the Environmental Movement

By the end of the 1960s criticism of geography as spatial science began to mount in part due to the growing environmental movement, which led geographers to widen the constricted discourse of spatial science to encompass the environment more generally. This generated interest in the long history of geographical engagement with ideas concerning the relationship between society and nature, and, hence, with geography as a means to both study and change environmental ideas, including

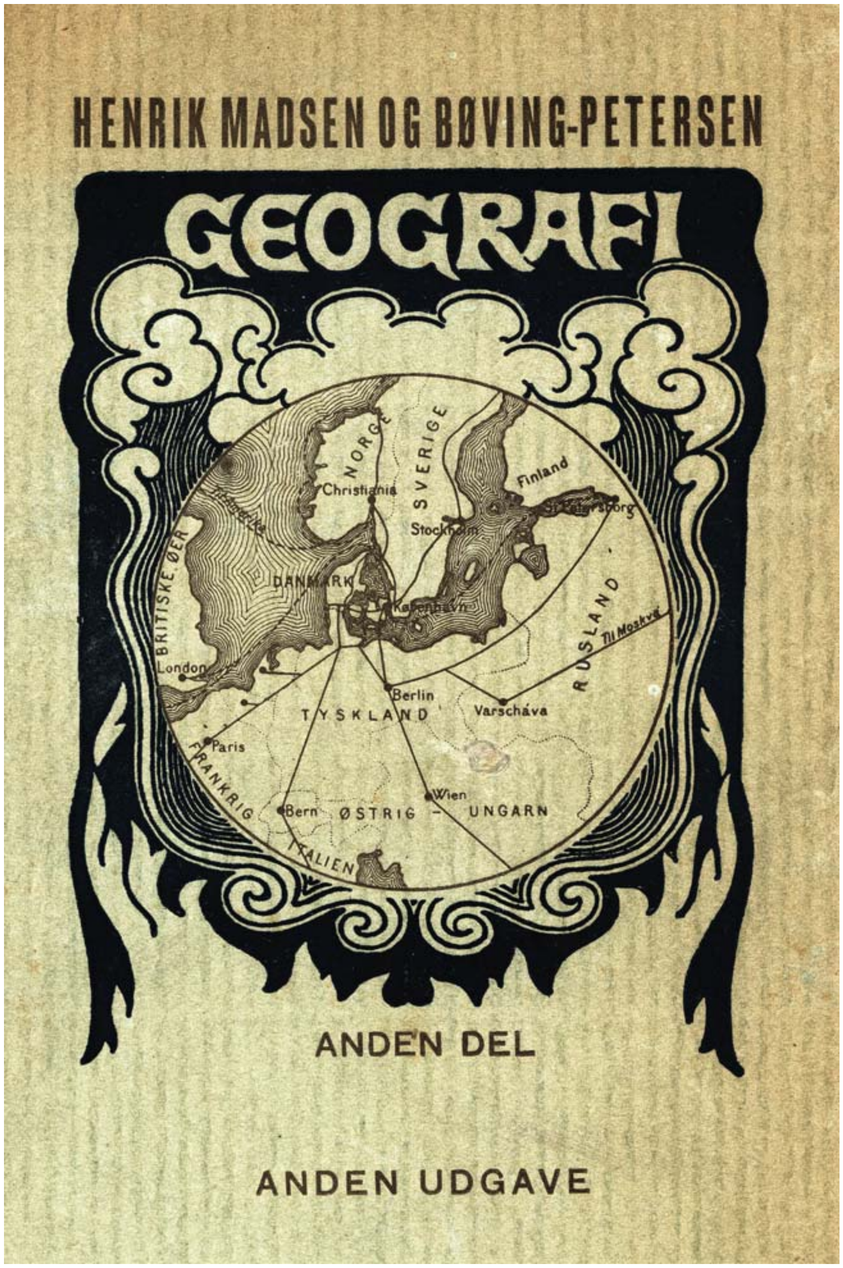

Figure 3 Textbook cover with map: This textbook cover from a 1901 Danish elementary school textbook, shows how railway and shipping links from all over Europe converge on Copenhagen. It places Denmark squarely in the center of the pupil's world-picture. Every nation's school maps, no matter how small and peripheral the nation, tend to place their nation at the center, thereby reinforcing ethnocentricity. The book was published in Copenhagen by *Det Nordiske Forlag* (The Nordic Publisher). The cover map has the effect of making Copenhagen the pivotal point between Norden (Norway, Sweden, and Finland) and the rest of Europe, and the projection makes Norden seem inordinately large. Note that the name of the capital of Norway, then united with Sweden, is called 'Christiania' on the map. This is the name given to it by the seventeenth-century Danish king, Christian IV, when Norway was under Danish rule. After Norwegian independence from Sweden in 1905, the name was changed back to its original Norwegian name, Oslo. A great deal of imperialist and nationalist ideology is expressed through the names on school maps like this helps explain why the Norwegians later changed the name of their capital.

gendered ideas. Wright's geosophy influenced important humanities oriented Berkeley geographers such as Clarence Glacken, Yi Fu Tuan, and David Lowenthal, who studied both the history of geographical ideas and the ideological role of these ideas in shaping the geographic environment and landscape. The writing of geography on environmental topics for a broad, educated public has subsequently gained in importance and polymath writers, such as Kate Sopher, Sarah Whatmore, Jared Diamond, and William Cronon, have continued this tradition to the

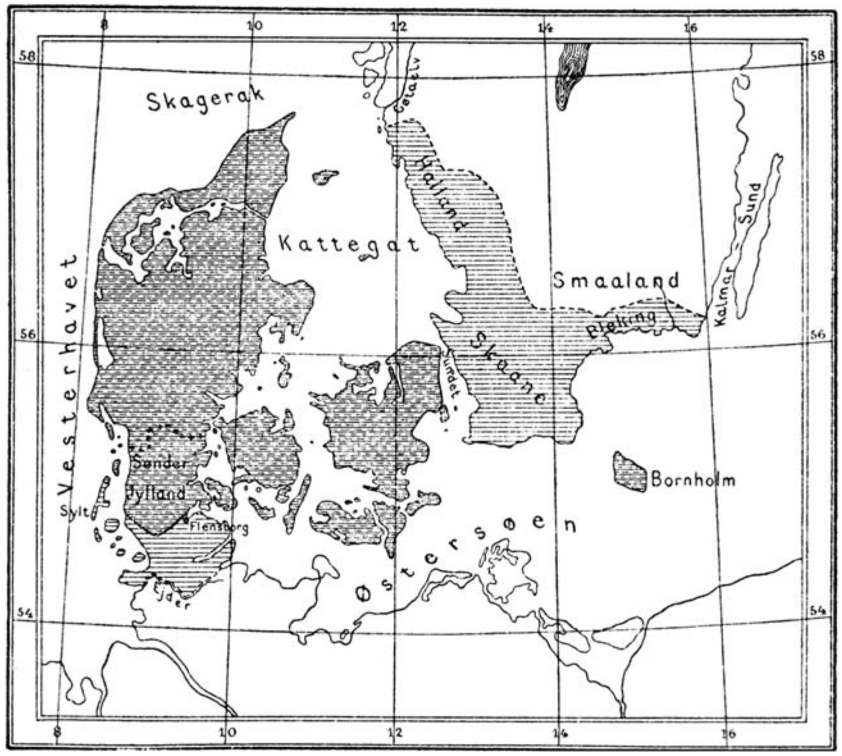

Figure 4 Shaded map of Denmark's historical extent**:** This map from the 1901 textbook shows the area dwelled in by the Danish-speaking 'folk' in times past. The striated area to the east shows Halland, Skaane (Scania), and Bleking; areas lost to the Swedes in 1648. The striated area to the south of Jutland shows the southern part of the Duchy of Schleswig (as it is spelled today in German). The area immediately above it, which is labeled 'Sønder Jylland', and which is shaded the same as the remainder of Denmark, shows the area where Danish is still spoken in the northern part of the Duchy of Schleswig, which was lost in war to the Germans in 1864. Much nationalist ideology, key to several wars and many deaths, is concealed by the apparently innocuous facts on this map. The fact that the map is based on the presence of a 'folk', speaking a specific language, reflects the ideology of the time that language was key to national folk identity, and that such folk naturally belong together in a single nation. The map thus expresses a hidden claim to Danish sovereignty over the northern part of Schleswig, which has been given the Danish name of 'Sønder Jylland' (Southern Jutland) on the map. This is the ancient name of the area that was used prior to the Middle Ages, after which the area gradually was Germanicized, and came to be known as Schleswig. After World War I, in 1920, this area was returned to Denmark after a plebiscite.

present. The humanities oriented tradition also inspired a new generation of ideologically conscious landscape geographers such as Stephen Daniels, Denis Cosgrove, and Kenneth Olwig whose work began to gain attention in the 1980s.

The renewed interest in landscape also drew inspiration from the work of critical Marxists, such as the cultural theoretician Raymond Williams and the work of the historical cartographer J.B. Harley who drew heavily upon the ideas of Michel Foucault. From a Marxist perspective, ideology, cartography, and landscape could be seen to belong to the material base of society because of the role they played in the structuring and planning of land use and ownership. They also, however, played a significant role in shaping consciousness of the land as an object of exploitation, and as a source of social status. The idea of landscape, however, was also important to the ideological superstructure, legitimating the power of private property ownership (an important tenant of liberalism), nationalism, and imperial expansion. The contemporary ideological role of landscape in generating 'false consciousness' has been explored by Gillian Rose in terms of gender relations, Don Mitchell in terms of class relations, and James Duncan and Sharon Zukin in terms of power relations.

Marxism, Education, and Spatial Ideology

The quantitative revolution in geography was rooted in America. It never gained the same foothold in European geography, where geography's educational and, hence, ideological role remained relatively strong, and where political parties tended to have a stronger ideological grounding in, for example, Marxism or liberalism. The 1968 student revolts in the United States and Europe against autocratic university systems, and university science's complicity in the military–industrial complex, raised questions about the objectivity of academic science. It was in this context that 'spatial scientists' also began to question the scientific premises of their endeavor, and that geographers of European background, in

Figure 5 Picture of bombed-out city: The message of the shaded map of the historical spatial extent of the Danish-speaking folk (**Figure 4**), is reinforced by this figure from the same 1901 textbook showing the bombed-out ruins of Sønderborg, Sønderjylland, during the 1864 war. The illustrations from Denmark show thriving, modernizing environments, whereas the image of Sønderjylland is one of unnatural ruin.

particular, began to look to Marxism as a source of critical inspiration.

The Marxist inspired ideological critique of spatial science emanated largely from the work of 'radical' British geographers, many of whom worked in America, but maintained intellectual ties to Europe, especially France. They critiqued spatial models for creating the impression that urban spatial structure was a function of scientific laws, such as the 'gravitational' effect of distance, while they in fact derived from conditions generated by social inequality. The spatial models thus created the false consciousness that social inequality in urban land use was given by space as a form of nature, rather than by social processes. Center–periphery spatial models likewise were seen to literally naturalize uneven development by giving credence to the idea that peripheral areas of the world were underdeveloped because they were 'naturally' wild. The Marxist critique of the ideological role of center–periphery spatial models in legitimating the underdevelopment of peripheral areas in the Third World thereby became a theme in developmental geography. In Scandinavia, where geography long played an instrumental role in regional and spatial planning, geographers were likewise able to show how the language of space and region worked to obscure the technocratic, and even nationalistic, ideologies driving this planning. Spatially oriented geography, however, has tended to prefer to generate new, socially implicated concepts of space such as 'relational space' (deriving from the philosophy of Gottfried Wilhelm Leibniz (1646–1716)) or 'third space' (deriving from the philosophy of Henri Lefebvre), rather than undertake a critique of the ideological history of the concept of space itself, which would parallel the work done by other geographers on the concepts of nature and landscape. Key geographical writers engaged with the ideology of spatial and regional geography are David Harvey, Richard Peet, Neil Smith, Edward Soja, Paul Claval, Anssi Paasi, and Gunnar Olsson.

The continued ideological dimension of geography as a broadly educational praxis is particularly notable in Britain, where geography remains relatively important at all levels of education, and there is a powerful market for textbooks and lexica. Britain thus produced some of the earliest critiques of spatial science, such as that of Derek Gregory, treating the issue of ideology in depth and introducing the neo Marxist 'critical theory' of Jürgen Habermas and others. For critical theory, an important task was to expose the way positive science generated instrumentalist discourses and practices that undermine democratic participation in society's governance. This was relevant both to the ongoing ideological critique of spatial science and to the defense of education's role in fostering civil society. A problem with traditional Marxist approaches is that by privileging the material base of society, they seem to discount the ability of people to consciously change society through their own actions. British geographers, in particular, found a solution to this problem in the ideas of the English sociologist Anthony Giddens. Drawing upon the time geographical theory of the Swedish geographer Torsten Hägerstrand and the sociology of the French theorist, Pierre Bourdieu, Giddens emphasized the role of individual social actors in the 'structuration' of society and space. The ideological potency of Giddens' theories were made manifest when Giddens became a chief ideologist for Britain's 'New Labour' political movement, which gave greater weight to the individual than did traditional socialism. Derek Gregory,

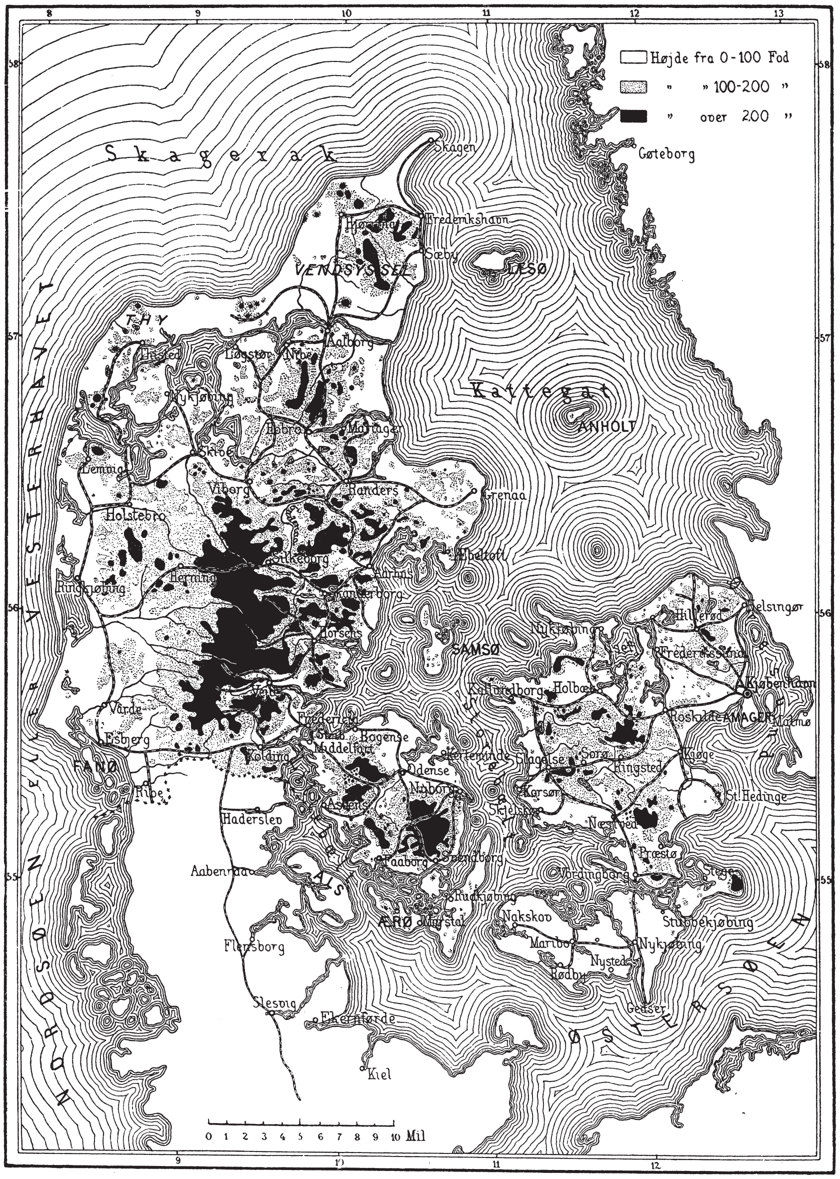

Figure 6 Topographical map of Denmark: A later 1911 Danish school textbook includes a statement and a verse that make clear the nationalistic ideological import of such a map: A 1901 Danish school textbook (see **Figure 3**) includes this map, where the lost Duchy of Schleswig (or Slesvig in Danish), south of the Jutland border, is represented by empty white space.

> Denmark is one of Europe's smallest countries, but it is our fatherland; in it we are born, and it is here we have our home, and we therefore ought to value it higher than other lands. Therefore:
> 'Friends, look at Denmark's map;
> look, so you never forget,
> until every spot has been given voices,
> for this fatherland is ours!
> Measured against the world it is little
> therefore not a spot to waste!
> Measured by the heart it is great
> this is where Denmark's saga is made!'

The source of the verse is not given, but it is taken from a popular 1889 poem and song, *Our Land, a geographical poem*, by Christian Richardt. It concludes with a demand that the provinces Denmark lost in the past be returned to their rightful place on the map no doubt a reference to Sønderjylland, the ancient Danish name for Schleswig, which was lost to the Germans in 1864. The poem's reference to 'voices' may be an oblique way of suggesting that all the land be given a vote, since the Danish word for voice and vote (*stemme*) is the same, and the right to vote was a controversial topic at the time. It may also be a call for the Danish speakers of Sønderjylland to be given a voice in their own national destiny and this actually happened in 1920 when they voted to be united with Denmark. There are thus many potential layers of ideology in a map!

Ron Johnston, Nigel Thrift, and Doreen Massey exemplify the broad British, education driven engagement with ideology.

Conclusion

The concept of ideology provides a useful key to understanding the praxis of geographical writing and graphic communication, especially in relation to edu cation. Through their ideological praxis, geographers have also become involved in the development of specific political ideologies ranging from anarchism and Marxism to nationalism, liberalism, imperialism, and with it, ra cism. Geography has historically made important con tributions to ideological debate in relation to the ideas of nature, environment, landscape, and space. Geography as spatial science sought to be nonideological, but for this reason it may have been all the more ideologically po tent. The concern with ideology, brought about by the environmental and feminist movements, and the critique of spatial science, led to important contributions to the understanding of the ideological dimensions of the con cepts of nature and landscape. The ideological critique of space, on the other hand, has tended to focus on specific concepts of space, rather than the notion of space *per se*. The identity of geography, not the least political geog raphy as a discipline broadly concerned with ideology, has, if anything, been strengthened by the contemporary geographical concern with ideas and concepts, as ex emplified by this encyclopedia. Geography has thereby become an important ideational source for the turn toward landscape, space, and place that has occurred in many other areas of study, including anthropology, archaeology, and sociology.

See also: Determinism/Environmental Determinism; Hagerstrand, T.; Quantitative Revolution; Radical Geography; Spatial Analysis, Critical; Structuration Theory; Structurationist Geography; Time and Historical Geography; Wright, J. K.; von Humboldt, A.

Further Reading

Blunt, A. and Wills, J. (2000). *Dissident Geographies: An Introduction to Radical Ideas and Practice*. Harlow: Prentice Hall.
Cosgrove, D. and Daniels, S. (eds.) (1988). *The Iconography of Landscape*. Cambridge: Cambridge University Press.
Cronon, W. (ed.) (1996). *Uncommon Ground: Rethinking the Human Place in Nature*. New York: Norton.
Duncan, J. S. (1990). *The City as Text: The Politics of Landscape Interpretation in the Kandyan Kingdom*. Cambridge: Cambridge University Press.
Glacken, C. J. (1967). *Traces on the Rhodian Shore: Nature and Culture in Western Thought from Ancient Times to the End of the Eighteenth Century*. Berkeley, CA: University of California Press.
Godlewska, A. and Smith, N. (eds.) (1994). *Geography and Empire*. Oxford: Blackwell.
Gregory, D. (1978). *Ideology, Science and Human Geography*. London: Hutchinson.
Harvey, D. (1973). *Social Justice and the City*. London: Edward Arnold.
Jones, M., Michael, W. and Rhys, J. (2004). *An Introduction to Political Geography: Space, Place and Politics*. London: Routledge.
Mitchell, D. (1996). *The Lie of the Land: Migrant Workers in the Californian Landscape*. Minneapolis, MN: University of Minnesota Press.
Olwig, K. R. (1984). *Nature's Ideological Landscape: A Literary and Geographic Perspective on Its Development and Preservation on Denmark's Jutland Heath*. London: George Allen & Unwin.
Olwig, K. R. (2002). The duplicity of space: Germanic 'Raum' and Swedish 'Rum' in English language geographical discourse. *Geografiska Annaler B* 84(1), 29 45.
Olwig, K. R. (2002). *Landscape, Nature and the Body Politic*. Madison, WI: University of Wisconsin Press.
Peet, R. (1998). *Modern Geographical Thought*. Oxford: Blackwell.
Rose, G. (1993). *Feminism and Geography: The Limits of Geographical Knowledge*. Cambridge: Polity Press.
Smith, N. (1984). *Uneven Development: Nature, Capital and the Production of Space*. Oxford: Basil Blackwell.
Whatmore, S. (2002). *Hybrid Geographies: Natures, Cultures, Spaces*. London: Sage.

Relevant Websites

http://encyclopedia.jrank.org/ALM ANC/ANARCHISM from the Gr av and Ip.html
Entry on Anarchism by Prince Peter Kropotkin in the *Encyclopedia Britannica*, 11th Edition (1911).

Immigration I

J. L. Waters, University of Liverpool, Liverpool, UK

Glossary

Assimilation Based on the assumption that new immigrants are culturally and socially distinct from the host society. Assimilation is an ideology for the incorporation of immigrants. It involves the gradual cultural, social, and economic adaptation and adjustment of immigrants over time so that they become more similar to and indistinguishable from mainstream society.

Diaspora Refers either to particular historically displaced groups of people, or indicates a wide variety of phenomena including the emergence of transnational communities, new kinds of 'deterritorialized' identity formations and representations, the displacement of traditional disciplinary boundaries and categorizations, and the cultural implications of globalization. Both definitions are concerned with representing population groups that are no longer primarily anchored (territorially or ideologically) in any one nation-state, although they maintain the memory or 'myth' of a homeland.

Home Country/Society Usually describes the last place of residence or country of birth of immigrants.

Host Country/Society Describes immigrants' destination country.

Multiculturalism An ideology that recognizes and ascribes value to the cultural and social distinctiveness of immigrant or ethnocultural groups. The concept has its roots in the colonial notion of plural societies.

Nation-State System A political and territorial system for social and economic organization, it denotes a series of legally defined and sovereign territories (countries). Came into being with the French Revolution in 1789, after which an international system of nation-states gradually took shape.

Native Population Commonly used in the literature to describe the nonimmigrant population of a country/society. This may or may not include second- or third-generation immigrants.

OECD Organization for Economic Cooperation and Development, comprising thirty member countries.

1951 Convention on the Status of Refugees The key legal document defining who is a refugee, their rights, and the legal obligations of states toward asylum seekers and refugees.

Introduction

Immigration has a number of key distinctive dimensions. First, it involves the deliberate crossing of nation state borders (usually indicating a long distance movement) undertaken by an individual or group. Second, it denotes the direction of that movement, that is, 'in' to (as opposed to out of) a country. Third, immigration is a political act, and is often discussed in terms of immigration control, and the legal status of migrants. Fourth, immigration is also a social and cultural process, involving adjustment, adap tation, communication, and negotiation. It affects both immigrants and host societies in different ways. Finally, immigration is invariably a highly emotive subject, provoking strong reactions (particularly within Western states), both favorable and negative.

In what follows, some of the more significant dimensions of contemporary immigration are discussed, drawing on a wide range of examples. The article examines, first, the geography of recent patterns of immigration and provides a discussion of different immigrant types. Second, it describes changing perspectives on immigration, starting with early research by the Chicago School on immigrant settlement patterns and integration in the US. It considers the impacts of immigration at the level of the metropolis, recognizing that the vast majority of immigrants locate in a relatively small number of major urban areas. These are discussed in terms of economic, social, cultural, and political impacts as well as the visible impacts on the urban landscape. It then turns to examine more recent ideas about return migration, circular mobility, and transnationalism. The nation state system remains centrally important to understanding con temporary immigration flows; the third section of the article therefore focuses on different aspects of immigration in relation to the nation state. This includes a discussion of: immigration policy; the economy; ideologies of immigrant incorporation (assimilation and multiculturalism); citizen ship; and illegal immigration. The article then considers what can be termed migrant centered perspectives on im migration. Rather than examining large scale flows and impacts of immigration, migrant centered approaches choose to focus on the experiences and objectives of the immigrants themselves and those of their families. Finally, the article offers some tentative conclusions regarding the future importance of immigration for geography and for society.

Contemporary Patterns and Types of Immigration

Geographies of Immigration

There is a distinctive and differentiated geography to patterns of immigration, which can be observed at different spatial scales (global, regional, and metropolitan). Typically, immigrants move from the 'north' to the 'south' and from poorer to wealthier nation states. Countries with the highest inflows of migrants are to be found in North America, Oceania, and Western Europe: Germany, Japan, Spain, Australia, Canada, France, Italy, Korea, Turkey, the UK and the US receive more than 100 000 immigrants annually. The US has by far the largest foreign born population in the world. As well as these broader global patterns, it is also possible to discern regional immigration flows, represented by, for example, a Southeast Asian migration system or an Oceanian migration system. These regional systems have persisted for hundreds of years and are represented by the presence of substantial diaspora populations.

Significantly, most immigrants choose to locate within a relatively small number of major cities in any one host country. In the UK, for example, the vast majority of new immigrants chose a location (at least initially) in and around greater London; in Canada immigrants disproportionably live in one of three major cities (Toronto, Vancouver, and Montreal); while in the US, New York and Los Angeles are highly attractive migrant destinations. This concentration of immigrants in particular cities has a number of implications, both positive and negative. Immigrants' needs – in terms of housing, education, employment, medical and health care, translation services, and language classes – must be met, placing disproportionate strain and burden on the services of a few cities. At the same time, however, immigrants provide cities with much needed labor – many cities (particularly their manufacturing and service sectors) are increasingly dependent upon immigrant labor.

The Age of Migration

Castles and Miller famously described the contemporary period as the 'age of migration'. Since the end of World War II, international migration has grown in volume and changed in character, they argue, in two main stages: 1945 to 1973 (the end of this first period was marked by the global oil crisis) and 1974 to the present. The first stage of global, international migration was characterized by the migration of workers from the European periphery to Western Europe (such as through the 'guest worker' systems described below); the migration of 'colonial workers' to their former colonial powers (such as the migration of Algerians to France and Jamaicans to Britain); and a pattern of permanent migration to North America and Australia from Europe (initially) and after the mid 1960s from Asia and Latin America. The second stage of post war migration has been characterized, in contrast, by a decline of labor migration to Western Europe. This stage has also seen family reunification among former foreign and colonial workers (leading to the creation of new ethnic minority communities within particular cities); the continuation of economically motivated migration to immigration regions within North America and Australasia (but with a significant change in the source countries of these immigrants); mass movements of refugees and asylum seekers; and the international mobility of highly skilled migrants (both temporary and permanent). To this list should be added the significant recent increase in economic migration between European Union (EU) member states, as a consequence of recent EU expansion (such as the substantial migration from Poland to the UK).

Categories of Immigrants

Immigrants can be categorized (both formally and informally) by type or class, depending on the immigration policy within the country in question. Immigration usually falls into one of several different categories (the precise nature of these categories will differ between countries). These include economic, family, and refugee categories. The first of these, known as economic class immigration, describes individuals who are granted entry to a country on the basis of employment skills, qualifications, or experience. These can include entrepreneurs, the self employed, investors, and the highly skilled. It is widely assumed (by the host nation state) that the host country will gain direct economic benefits (including human capital) from this type of immigration. Conversely, immigrants' home countries will lose out (a process known as 'brain drain'). Recently, global flows of skilled immigration have significantly increased, largely as a consequence of the widespread implementation of skilled migration programs throughout the West (the UK, for example, recently introduced a Highly Skilled Migration Program).

Family class immigration refers either to the reuniting of family members within the host country, known as 'family reunification', or to sponsorship, whereby a person already permanently resident in the host country financially sponsors the immigration of another family member. Family reunification usually occurs as a result of an earlier immigration stream (where other family members follow some time after the original migration). Sponsorship requires that a resident of the host country prove that they are able to support, financially, the family member following their immigration.

Refugees represent the third major category of immigrants. Under this class, persons can apply for refugee status (i.e., submit an application for asylum based on the 1951 Convention on the Status of Refugees) either at special offshore processing sites within their home country, or after entering the host country. Their case will then be heard by a committee or board, which will decide whether or not they have a legitimate case for asylum and can therefore be given refugee status. If they are found not to have a case for asylum, then they are (in theory) subsequently deported to the country of origin. If, on the other hand, they are granted refugee status, then in most cases they are given indefinite leave to re main and can apply for citizenship after the required length of time.

Changing Perspectives on Immigration

Understandings of immigration have changed over time, as a result of shifting theoretical approaches. Tradition ally, immigration was seen as a discrete move from one country to another – from the home to the host society. It was generally assumed that immigrants were making a long term commitment (economically, socially, and fi nancially) to the host society, and that this would include permanent relocation. As a consequence of this, immi grants would adjust over time to their new surroundings, assimilating according to various measures (such as economic success, language ability, and intermarriage) to the host society. Milton Gordon's classic work, *Assimi lation in American Life*, captures this perspective on im migration and the process of gradual assimilation among minority groups.

On the whole (and this is particularly true of early work), research has tended to focus on immigration from the perspective of the host society. Two strands of this work can be identified. First, it has examined the rela tionship between immigration and host nation state policy; policy is used to control and direct immigrant flows (this strand of work is described in detail below). Second, research has been concerned with the variable impacts of immigration upon the host society and par ticularly upon urban areas. These include, for example, the consequences of cultural and social diversity, labor market outcomes related to immigration, or the spatial distribution of immigrant groups vis à vis the native population. Both of these strands of scholarship continue to be vibrant, attracting a great deal of new research.

Recently, however, it is possible to identify a change in the way in which immigration is conceived (theoretically) and, related to this, changes in the foci of research (i.e., changes in the kinds of empirical questions that re searchers are asking). Increasingly, immigration is no longer seen as a permanent relocation from one country to another but as a far more fluid and ongoing process that has no definite, identifiable beginning and end point. Many immigrants continue to move between the home and host societies, even after the acquisition of citizenship in the new country. Many do not display the signs of assimilation over time. As a result of this changing per spective on immigration (and what immigration means in practice), research has begun to focus on the immigrants themselves (their experiences and motivations) as well as on the impacts of migration upon immigrants' home so cieties. Perhaps most significantly, new research examines the ongoing relations between the home and host society. In what follows, these different perspectives on, and conceptions of, immigration are discussed.

Immigration and the Metropolis: The Spatial Distribution of Immigrants

Early work on immigration is widely associated with the Chicago school of sociologists, which studied the impact of immigration on the urban form of US cities in the early to mid twentieth century. An important aspect of this work involved an examination of the spatial distri bution of immigrant groups, and related this to levels of assimilation. This perspective is exemplified in Louis Wirth's study of Chicago's Jewish community in *The Ghetto*. Massey and Denton's famous examination of racial segregation in major US cities during the 1980s was also extremely influential for later scholarship on immi gration and the social geography of urban areas.

Particularly since the 1980s, research in geography has used census data to map and describe the spatial distri bution of ethnic minority populations (including first , second , and third generation immigrants) in the UK, North America, and Australasia. Researchers have been particularly concerned with identifying the existence of segregation. Segregation is often measured statistically, applying different indices (such as the index of dis similarity) to census statistics; the distribution of different ethnic groups can be represented through maps and shading, indicating the degree of separation or the density of settlement. The significance of this work lies in the implications of segregation for social cohesion and ex clusion and the assumption that spatial distance is a strong indication of social distance. The spatial separation of groups in the urban landscape can be indicative of more profound social problems (such as isolation, discrimin ation, and the inability to access necessary resources).

Impacts of Immigration on Urban Areas and Communities

While immigration is generally controlled and directed at the level of national governments, arguably its greatest impacts are felt at the urban scale. Immigrants generally choose to live and work in major metropolitan areas and, frequently, congregate within particular districts of these cities – concentrations of immigrants can be observed at several spatial scales. They may also be concentrated in particular employment sectors. As observed by Saskia Sassen in her work on global cities, immigrants are often found working in the low skilled, low waged sectors of the service industry. Put succinctly, immigration has transformed the sociocultural makeup and economies of many cities, including transforming the physical, built environment. In what follows, the most significant im pacts of immigration on urban areas are discussed.

Immigrants bring with them a diverse set of social and cultural traits and practices, knowledge, and under standings. Sometimes these may broadly differ with those of the host society. In reality, host societies are rarely socially and culturally homogenous but are themselves highly diverse spaces, created through, and exhibiting the legacy of, numerous past immigrations. Perhaps one of the biggest and most explicit markers of cultural difference between immigrant and nonimmigrant com munities, however, is language. Many immigrants, on migrating, do not speak the official language(s) of the host society, and this may impact the ability of these indi viduals (and their families) to access resources, including suitable (and desirable) employment, health care and emergency medical treatment, education, transportation, and welfare entitlements. In the most severe cases, this inability to access resources may lead to social exclusion. Immigrants' wages postmigration have been strongly linked to language ability, and social and spatial segre gation among immigrant groups is frequently related to language use within the family. Language difficulties can also result in social isolation and feelings of loneliness among immigrants. Interestingly, age at migration has a strong bearing on the ability and willingness of immi grants to learn a new language. Very often, children will become fluent in the host language before older family members, and act as cultural go betweens for parents and grandparents, helping them with everything from day to day chores to interpreting employment law and even school reports.

There are many other forms of cultural difference and markers of social and cultural diversity. The difference between immigrants and the host society is often inter preted and discussed in terms of race. While racial diversity can have a positive influence on the host com munity, most research and writing on race and immi gration emphasizes the negative, focusing on the stereotyping of immigrant groups, discrimination, and in the worst cases, explicit racism. New theoretical per spectives on race, highlighting the socially constructed nature of racial identities, underline that these identities are frequently imposed on groups by the host society, and are invariably indicative of highly unequal power re lations. Other forms of identity are also socially con structed and imposed on immigrant groups. In the past, immigrants have generally been characterized as poor and associated with a social underclass, resulting in stigmatization and in some cases ghettoization. Often, race and class are seen to intersect in complex ways, confining immigrant groups to marginalized spaces in the city. This is evidenced, for example, by the growth of Chinatowns within cities in Europe and North America during the nineteenth and early twentieth century, which grew out of practices of discrimination and racism by host societies. Although today many Chinatowns play a largely symbolic role in the cultural life of cities, present day versions of confined and restricted immigrant spaces are apparent. In the US, much academic work has fo cused on the contemporary ghetto. Often, such mani festations of segregation in the physical landscape are thought to indicate a lack of integration within main stream society, while also at the same time actually in hibiting integration, by restricting the ability of immigrants to access the resources needed to enhance life chances and enable social mobility. On the other hand, the development of what have been termed ethnic enclaves or ethnic villages can be seen as a positive manifestation, reflecting immigrants' adaptation strat egies and the spatial concentration of social capital often leading to economic success.

Wealth can also be a marker of social difference, separating immigrants from the rest of society. Research on expatriates working in Southeast Asia describes the circumscribed social worlds of professional, wealthy businessmen, involving little interaction with people outside their particular exclusive social spaces. In Can ada, wealthy immigrants from Hong Kong have built what have been locally termed 'monster houses' – large, million dollar boxlike structures that exemplify con spicuous consumption. These houses have provoked hostility among the local population, perceiving them as inhospitable and out of character with the rest of the neighborhood. On a larger scale, wealthy immigrants may invest in real estate and property development, transforming the physical landscape of some cities (such as Vancouver). During the late 1980s and 1990s up to one fifth of what is now the downtown area of Vancouver was purchased and subsequently redeveloped by Hong Kong property developer Li Ka shing. Many condomin iums were marketed and sold exclusively in Hong Kong to prospective immigrants.

Return Migration, Circular Migration, and Transnationality

In a great deal of writing on the topic, immigration is no longer seen as a permanent and discrete movement from one country of residence to another, whereby ties to the country of origin are effectively severed or gradually diminish in importance over time. Instead, many scholars are sensitive to the ways in which immigrants continue to maintain significant, persistent links (emotional, cultural, social, economic, and political) with their country of origin after immigration, greatly facilitated by develop ments in technology and ease of transportation. Import antly, these links appear to be sustained over time. This type of migration has been widely termed transnation alism, and there is now a substantial and continually expanding body of work within geography and the social sciences specifically examining transnational processes (where previously the focus of such research would have been immigration). Empirically, work on transnational migration has focused on a number of themes, such as the importance of remittances, split family formations, transnational identities and diaspora, multiple political and social allegiances, translocal ties, cosmopolitanism, religious affiliations, and dual citizenship. In addition to (and not unrelated to) transnationalism, researchers and governments are increasingly recognizing the circularity of many contemporary international migrations and the high incidence, among some immigrant groups, of return to the country of origin. Both circular migration and return migration challenge traditional conceptions of immigration, which focus on permanent residence and settlement in the new country and the severing of ties with the old.

Temporary Immigration: International Students and Skilled Workers

Increasingly, a proportion of immigration is officially classified as 'temporary,' as seen with the rise in numbers of international students and highly skilled workers. Research is only beginning to acknowledge the important role that students play in contemporary international population flows. Current estimates suggest that there are between two and three million international students in tertiary education worldwide. These figures have doubled between 1980 and 2000 and are predicted to increase substantially over the next decade. Like other forms of migration, international student mobility has a distinctive geography – the most popular destination countries are Australia, France, Germany, Japan, New Zealand, the UK, and the US. The majority of international students derive from East, South, and Southeast Asian countries, particularly the People's Republic of China, South Korea, India, Malaysia, Hong Kong, and Taiwan. While inter national students are expected eventually to return to their country of origin, some remain in the host country (both legally and illegally). Perceiving international stu dents as a valuable asset, countries are increasingly making it easier for foreign graduates, educated locally, to obtain work permits or permanent residency status after the completion of their studies.

Much recent skilled migration is also temporary in nature, and is particularly apparent among IT workers, scientists, researchers and academics, business executives, and managers. The majority of these workers can be found in a relatively small number of countries: Australia, Canada, Japan, the UK, the US, Germany, and France. The OECD has been responsible for commissioning a large survey on skilled migration. According to this survey, an estimated 900 000 skilled workers entered the US between 1990 and 2000 under the so called H 1B visa program, which was set up to facilitate the tem porary entry of skilled professionals. It estimates that temporary workers accounted for one sixth of the total US IT work force during this period. Multinationals have a significant role to play in this process – the transfer of staff within these large organizations has contributed significantly to the increase in numbers of temporary workers. In the mid 1990s, intracompany transfers ac counted for as much as ten percent of total flows of skilled workers from Canada to the US.

Immigration and the Nation-State

Despite recent challenges posed by globalization and transnationalism, the nation state continues to exert a high degree of control over global population mobility. The concept of immigration would, in fact, be meaningless without the continuing salience of nation state boundaries restricting population movements, and a formalized global system of sovereign territorial divisions.

Nation states as a global mode of political and social organization have not always existed, however. Prior to the fifteenth century, feudal societies dominated the Western world. These societies emphasized locality and their populations were relatively immobile. As Yasmin Soysal notes, only certain categories of people – artists, artisans, scholars, clergy, the military, and the merchant class – were permitted to move freely, while all other members of the population had restricted movement. It was not until the late eighteenth century that the concept of the nation state came to prominence, with the French Revolution and the establishment of a republic. The population of France (and then later the same became true for other territories) began, for the first time, to be re garded as an economic and military resource. Con sequently, territorial boundaries were constructed to keep people in (and thereby contain this resource). Initially, then, nation state boundaries (and the early manifestations of immigration policy) were more concerned with con taining people than with excluding them – quite the op posite of the situation in most nation states today.

Immigration Policy

One of the key means by which nation states influence immigration is through government policy. Most states in the world have some sort of immigration policy, designed to control and direct flows of immigrants across their borders. As noted, in the past many states (such as France and China) used policy to contain their populations – to prevent their outward movement (and the consequent loss of human resources). Today, policy is overwhelmingly concerned with controlling inflows of migrants – accepting and embracing some people while actively excluding others.

Different countries adopt different stances when it comes to immigration policy. Some are seen as countries of immigration, displaying a generally favorable attitude toward immigration and the benefits it can bestow. These countries include Canada, the US, New Zealand, and Australia. Other countries have a far more cautious approach to immigration, usually favoring exclusion over active recruitment (such as the UK and, until recently, Japan). Most countries fall somewhere in the middle of these two stances, with variable histories of, and recent trends in, immigration.

Historically, many now liberal democratic states used immigration policy to exclude groups of people on the basis of their country or region of origin. The Commonwealth Immigrants Bill (1962), followed by the Commonwealth Immigrants Act (1968), sought to end immigration to the UK from commonwealth countries. From the early twentieth century to the 1970s, a White Australia Policy restricted the immigration of people of non white (non European) heritage to Australia. Between 1923 and 1947, Canada implemented a Chinese Exclusion Act, prohibiting the immigration of Chinese people. The earliest attempts by the Canadian state to exclude persons of Chinese origin, however, came in 1885, in the form of a head tax. These policies are today seen as inherently racist.

Between 1945 and 1973, a number of Western states implemented guest worker systems, involving the temporary recruitment of labor from overseas. These countries included Austria, Belgium, Germany, France, Luxembourg, the Netherlands, Sweden, and Switzerland. Under these new systems, large numbers of foreign nationals (and their dependents) were admitted as so called guest workers to provide a temporary boost to economies suffering post war labor shortages. The source countries of this labor included Italy, Spain, Portugal, Greece, Yugoslavia, Turkey, and Morocco. Most of these jobs were tied to specific industries, and workers had no right to stay permanently within their host country. Despite this, many workers did remain, as exemplified in the case of Turkish guest workers in Germany.

Outside Europe, other post war changes in immigration policy can be observed. Policy in the US changed significantly in 1965, when a quota system based on the national origins of immigrants was replaced by a system whereby immigrants were selected according to other (nonracist) criteria, such as education and skills and family status. The US has had, as it sees it, a long standing issue of illegal immigration, particularly over the US–Mexican border. In 1986, specially implemented legislation attempted to address this, granting amnesty to three million illegal immigrants that had been living illegally in the US since 1982. Any employers found employing illegal Mexican immigrants after this date would suffer severe penalties. Immigration policy in other immigrant countries, including Canada, Australia, and New Zealand similarly changed in the 1960s, abandoning selection criterion based on country of origin and implementing in its place a points system. The Immigration Act of 1976 is described by Citizenship and Immigration Canada (Canada's federal immigration department) as the cornerstone of present day immigration policy. It came into force in 1978 and represented the first time that the principles and objectives of Canadian immigration policy were explicitly stated. Among the Act's most innovative demands is the requirement that the government 'plan' immigration (i.e., set target numbers for different classifications of immigrants and suggest the ways in which these targets can be met). Planning immigration involved consultation with the provinces about their own particular demographic and economic needs. Refugees are recognized as a special case, selected and admitted separately from other immigrants (another innovation of the Act). The Act recognizes four basic categories of individuals eligible for permanent residency: (1) family class (which includes the immediate family and dependent children, as well as parents and grandparents over 60); (2) humanitarian class (which includes (a) refugees as defined in the 1951 United Nations Convention relating to refugees and (b) persecuted and displaced persons who do not qualify as refugees under the rigid UN definition but who are members of a specially designated humanitarian class); (3) independent class (which comprises applicants selected on the basis of the points system); and (4) assisted relatives, that is, more distant relatives who are sponsored by a family member in Canada and who meet some of the selection criteria of the independent class.

The UK, which has traditionally been seen as a zero immigration country, on the basis of its attempts to prevent all forms of immigration, has recently introduced new categories of desired immigrant under the Highly Skilled Migration Program (HSMP), the Innovator program and the Investor program. For the HSMP, immigrants are admitted to the UK on the basis of a points system. Points are scored in four main areas: qualifications; past earnings; age; and UK experience. Applicants are also required to demonstrate specific proficiency in English. Applicants to the Innovator program need to demonstrate entrepreneurial ability, technical skills, and a good business plan. A points system also applies in this case. Applicants to the Investor program must have the funds to invest at least one million pounds in the UK economy. As these examples suggest, immigration policy is frequently tied to the perceived needs of the national economy. The relationship between immigration and the economy is discussed in more detail in the next section.

Immigration policy is usually implemented by the central or federal government. In some cases, however, regions within nation states may have some say over immigration. In Canada, an agreement was recently signed between the federal government and the prov inces, known as the Provincial Nominee Program, to allow the provinces some measure of control over im migration to their territories (in line with their own perceived needs). Under the Canada–Quebec Accord on Immigration, the province of Quebec is able make its own decisions on immigration. In 1994, voters in Cali fornia (the US state with the largest share of foreign born persons) passed legislation to exclude illegal migrants from the receipt of all public benefits with the exception of emergency health care.

Immigration and the Economy

There is often a strong and direct link between a nation state's immigration policy and its economic policy, and the relationship between immigrants and economic per formance has been widely examined. In many countries, such as Canada and Australia, immigration is perceived as a key means of bolstering a declining working age population and attracting human capital from overseas. These countries set annual minimum immigration targets (usually of several hundred thousand people) and are especially keen to attract business immigrants, with an eye on securing their human and economic capital. Business immigration programs exemplify attempts by the state to align immigration and the economy, and are increasingly widespread among immigrant receiving countries. In Canada, the Business Immigration Program is designed to admit immigrants who fall into one of three categories: entrepreneur, investor, and self em ployed. Prospective immigrants have to fulfill a number of criteria in order to be granted landed immigrant or permanent residency status.

The success of economically motivated immigration policy, however, is equivocal. In the case of Canada's Business Immigration Program, David Ley has shown that the level of achievement of business immigrants admitted under this program is frequently far lower than anticipated, and lower than official government claims suggest. Most of those businesses examined lost money or were barely profitable, and a similar situation has been observed for both Australia and New Zealand. More generally, the issue of accreditation highlights the fact that the transfer of skills from one country to another can be problematic; many skilled immigrants are unable to secure work, following migration, at a level com mensurate with their qualifications and/or experience, and are often forced to take unskilled employment. Employers and employment organizations may require that new immigrants retrain in their chosen occupation, regardless of prior levels of experience/qualifications. The costs of retraining are often prohibitive. Clearly, this situation is in the best interests of neither the immigrant nor the host society. Related to these observations, work in economic geography has also examined the relation ship between immigration and segmented labor markets. It is now well known that many immigrants are in advertently channeled into a very narrow range of oc cupations, working in low skilled and low paid jobs that lack union protection and basic benefits (sick pay, paid holidays, health insurance, and so on).

Changing Ideologies of Immigrant Incorporation: Assimilation and Multiculturalism

Nation states are culturally diverse, and immigration has undoubtedly played a major role in the ongoing trans formation of their ethnocultural makeup. It is no longer possible, if it ever was, to talk about nation states in terms of homogenous cultural spaces. What is interesting, however, is the ways in which nation states – Western liberal democracies in particular – have responded to diversity and continuous change. In various ways, states attempt to incorporate immigrants into the broader na tional community. This incorporation can have a number of different dimensions: social, cultural, political, and economic. Broadly, these dimensions can be classified under one of two dominant ideologies: assimilation and multiculturalism (**Table 1**).

Prior to the 1970s, Western liberal democratic nation states favored the assimilation of immigrants. They up held a Rawlsian notion of individual justice, whereby the state sought to be neutral on matters of culture. Liberal democracies are committed, at least in principle, to representing the needs of 'all' their citizens. Equality could be best achieved, it was believed, through the erasure of social and cultural difference.

Table 1 A comparison of assimilation and multiculturalism

Assimilation	*Multiculturalism*
Tries to erase ethnocultural differences.	Tries to maintain and even enhance ethnocultural differences.
Believes that making everyone 'the same' will promote equality in all aspects of life.	Believes that everyone has the 'right' to maintain their cultural heritage and that depriving them of it negates the possibility of equality.
Builds a nation-state around the concept of cultural and territorial convergence.	Builds a nation-state in a way that conceives the state as a 'mosaic' of different cultures. The nation-state is characterized by difference within its territorial boundaries.

By the 1970s, however, the failure of the assimilationist project (the endurance of sociocultural diversity and the myth of the culturally homogenous nation state) was becoming increasingly apparent. In countries such as Australia, Canada, and the UK, politicians began to talk about cultural mosaics (as opposed to the US notion of a cultural melting pot) and multicultural societies. Multicultural Britain became a widely used phrase. In Canada and Australia, multiculturalism became official government policy.

In Canada, multicultural policy was implemented in October 1971, beginning with a statement by PM Pierre Trudeau. As observed by Audrey Kobayashi, this policy encouraged diversity in the arts and leisure, through the celebration of cultural festivals, of ethnic restaurants, of different national dress, and folk dancing. The first Minister of State for Multiculturalism was appointed in 1972 and in the following year the government spent $2.7 million on multicultural programs. This money included giving grants to cultural or ethnic communities, the development of a 'Canadian identities' program to support the arts, folk festivals and ethnic history programs, multicultural teaching aids as well as money for academic research into multiculturalism and related conferences. The government also established a Canadian Consultative Council on Multiculturalism made up of representatives from various ethnocultural backgrounds, which implemented a program to monitor the multicultural content of Canadian broadcasting. This celebration of cultural diversity – what Kobayashi calls 'symbolic' multiculturalism – was not generally seen, however, as a tool for promoting equality among different cultural groups, placing more emphasis upon what has been described by David Ley and others as 'museum culture' than on anything more meaningful (such as experiences of inequality, discrimination, and racism in society). The policy lacked specific objectives. By the 1980s, calls for a Multiculturalism Act, which would include the legal mechanisms for addressing issues such as access to services, education, employment opportunities, and so on for ethnic minority groups, as well as allowing them to celebrate their cultural heritage, were increasingly widespread. Subsequently, a Department of Multiculturalism was created and the budget for spending on multicultural programs was increased by over US$62 million.

Katharyne Mitchell offers a particular perspective on Canadian multiculturalism. Mitchell argues that the Canadian state's interest in multiculturalism arose from strategic – particularly economic – considerations, making the link between Canada's need for export markets and increased trading opportunities. A more openly multicultural government would, it was assumed, improve relations with Canada's substantial immigrant populations, thereby tapping into their international business networks.

Australia has a similar immigration history to that of Canada, having experienced high levels of immigration from Asia since the late 1960s. Multiculturalism in Australia has been described as a government policy for managing difference. It has evolved in a number of stages since it was first implemented, two years after Canada's policy, in 1973. In 1989, the *National Agenda for a Multicultural Australia* was published, outlining a working definition of Australian multiculturalism. However, the agenda makes it clear that these dimensions will not undermine the unity of the nation state – there are clear cultural 'limits' to multiculturalism. Multiculturalism has not indicated an end to the nation state but rather it has demonstrated how nation states have responded to cultural and social diversity as a consequence of immigration, and attempted to incorporate that diversity. The unity of the nation state is of overriding importance within these multicultural policies.

Reemergence of Assimilation?

Recently, and particularly since the events of 11 September 2001, multiculturalism has lost some of its political appeal. There is a general sense in which government sanctioned multiculturalism has 'failed' as a policy and as a project – it has 'exhausted itself'. Several writers and commentators, including Katharyne Mitchell and chairperson of the UK Commission for Equality and Human Rights Trevor Phillips, have observed this decline. In many Western states, government led multicultural programs have ceased or noticeably declined. At the same time, both the discourses and practices related to immigrant integration and incorporation within liberal nation states are changing. There has been a renewed call for immigrant assimilation – what Rogers Brubaker has called 'the return of assimilation' – and the focus has shifted away from emphasizing and celebrating difference to stressing commonalities and a concern with civic integration. In Australia, the retreat of multiculturalism has been stark – the government Office for Multicultural Affairs was abolished in 1996. In the US, following 11 September 2001, patriotism legislation was passed in many states and in Western Europe (most notably in France) debates have centered on religious dress and other expressions of religious difference.

Citizenship

The relationship between immigrants and the nation state is partly mediated through citizenship. Citizenship may be conferred upon immigrants after a minimum period of residence and after meeting a range of requirements, which vary from country to country (e.g., in the UK an immigrant must be resident for five years, while for Canada the minimum residency requirement is three years). Immigrants usually have to 'apply' for citizenship – it is not conferred automatically. Once immigrants receive citizenship status, they lose (at least officially) their immigrant status.

After the French Revolution, the population became redefined as citizens, and a deliberate effort was made by the authorities to create a nation within the boundaries of the political state – to create a sense of nationhood. As Yasmin Soysal notes, it is no coincidence that the first statistics on international migration were collected after the Napoleonic wars, a period characterized by intense efforts by the French state to construct a national polity.

It was not, however, until the early twentieth century that passports and national identity cards formalized the status of citizen and alien/foreigner. As the nation state system became gradually solidified, so increasing international restrictions on movement were introduced – immigration laws were established throughout the West, designed to police the boundaries and control the borders of the nation state.

In his book *The Invention of the Passport: Surveillance, Citizenship and the State*, John Torpey examines the historical development of passport controls as a way of highlighting the institutionalization of the idea of the nation state as an ostensibly homogenous ethnocultural unit. This historical study has significant implications for contemporary attempts to control the movement of people across borders and within countries – such as recent debates in the UK surrounding the implementation of ID cards. Through the monopolization of what he calls the 'legitimate means of movement' (i.e., acceptable and unacceptable forms of border crossing), the nation state is able to embrace its citizens and extract from them the resources needed to reproduce itself over time. This hinges on the ability of nation states to tell the difference between a citizen and a noncitizen, and this is where the passport enters as an extremely useful tool in the controlling of borders and, thereby, immigrants.

The meaning of citizenship and its relationship to the international movement of people has been challenged, to a certain extent, by the recent extension of dual citizenship/nationality. Seventy nation states, including Canada, Italy, Israel, South Africa, New Zealand, the UK, and the US, allow their emigrants to retain their citizenship after having become citizens of another country. Some countries are even extending citizenship rights to second and third generation *émigrés*. Today, more and more countries around the world perceive the benefits of extending citizenship rights to ex residents. For example, since January 2005, all Indians acquiring citizenship of another state are entitled to retain their Indian citizenship. Not only that, all overseas Indians who migrated from the country after 26 January 1950 will be able to apply for dual citizenship, as long as their host countries permit this. Through the granting of dual citizenship to overseas nationals, countries are changing the meaning of immigration and what it is to be an immigrant. The notion of immigration as a permanent move from one country to another is being redefined to incorporate transnational forms of identity and political allegiance.

Illegal Immigration

One of the biggest and most politically contentious issues facing Western governments today is that of illegal (sometimes called undocumented) immigration. The media are heavily implicated in setting the terms of debate around illegal immigration, often accused by academics of causing a moral panic about its extent and implications. Recent, high profile examples of the dangers of illegal migration (such as the death of 58 Chinese immigrants, found suffocated in a truck entering the port of Dover in 2000 and the death of 18 cockle pickers in Morecambe Bay, England, found to be illegal immigrants working in highly precarious conditions) have highlighted the severity of the issue. Illegal immigrants are frequently blamed (by the media) for a host of social ills, from pressure on welfare services, to unemployment, to increased crime levels, to terrorism. Much of this debate is fueled by speculation – the undocumented nature of illegal immigration means that it is impossible to quantify and very difficult even to estimate. Tackling illegal immigration has become an important political platform for many politicians in Western liberal democracies. The media debate frequently confuses the legal definition of different terms, especially refugee, asylum seeker, and illegal immigrant.

Migrant-Centered Perspectives

Census and other statistical data continue to provide an important insight into immigration and its effects of immigration. In the last few decades, however, more attempts have been made to study immigration from the perspective of immigrants and their families, utilizing qualitative research techniques. These are known as migrant centered perspectives, and a number of themes relating to this research can be identified, including immigrant identities and family relations. These perspectives have led to a significantly increased understanding of the motives for immigration from the migrants' own point of view. These may be very different from official assumptions, involving complex personal negotiations and the balancing of different social interests, underlining the fact that people do not always act in predictable ways.

Immigrant Identities

Immigration clearly impacts the identities of migrants as they are forced to confront and adjust to a new and unfamiliar environment. Immigration is commonly initially associated with culture shock and feelings of homesickness and the desire to return to the country of origin.

Qualitative data incorporating interviews with immigrants reveal the complexity of immigrant identities. In some cases, identification with the home country will fade over time as immigrants become increasingly assimilated into the new culture. Immigrants' identities change through interaction with members of the host society. Some immigrants acquire hybrid identities, which involve elements of both home and host society cultures, and arise as a consequence of movement. This is especially true when immigrants are transnational, maintaining various ties with their home country after migration. Constant engagement with people, ideas, and cultural artifacts from the home country allow immigrants to maintain a strong identification with that country. Immigrants may also actively attempt to renew and reproduce their old identity, through the repetition of cultural acts and rituals (language, dress, food, religious practices, and music), and the consumption of cultural commodities and artifacts. These may be passed down to their children (second and even third generation immigrants).

Immigrant Households and Family Relations

Traditionally, households have been an important unit of analysis in studies of migration – many immigrants do not travel alone, but are accompanied by members of their immediate family. However, the family/household has often been represented as an undifferentiated unit, with little or no reference to relationships within it. In reality, immigration can have significant impacts upon relations within the family unit, as well as affecting family members in very different ways. This omission has been addressed, more recently, through migrant centered perspectives, which have highlighted the need to examine intrafamily relationships in the context of immigration.

Frequently, expected gender roles and expectations are challenged through migration, sometimes leading to conflict between partners. When migration from a more to a less patriarchal society occurs, women are faced with an increased range of opportunities. These opportunities may, in some cases, lead to a greater sense of freedom. On the other hand, they may result in dissatisfaction with existing relationships and family/work roles. Women are often depicted in the literature on migration as the 'trailing spouse' – the assumption being that, during migration, their career is sacrificed for that of their husband. In fact, many men find that they are unable to secure suitable or desirable employment following immigration, and are forced to take a lower skilled, lower waged, and lower status job. Both their authority within the family and their self esteem may suffer as a result.

As well as gender roles and relations, immigration can also impact upon intergenerational relationships within the family – most notably the relationship between parents and children. Children frequently adjust to the postmigration environment far quicker than their parents – through schooling their language ability improves and they make new friends. As a result of exposure to a new cultural environment, children may challenge the authority of older relatives. Parents, finding their own adjustment a difficult process, may be less able to assert their authority over children. Research also shows that older people find the experience of immigration much harder to deal with than their younger relatives, especially if they are retired. Elderly people can find immigration isolating and sometimes unnerving.

Some families are separated through migration, leading to spatially dispersed (transnational) households. Often this is done for strategic reasons; different productive and socially reproductive functions are carried out at different sites (even on different continents). In the case of so called astronaut families (recently observed among immigrants from East Asia in North America and Australasia), the husband/father of the family will return to the country of origin to work, leaving his wife and children behind in the new country. In a different case, through the feminization of some migration streams (where immigration is largely female led), wives and mothers are defined as the principle applicant in immigration. This has been particularly recorded among Filipina women immigrants who live and work overseas as nurses, housekeepers, and nannies. Women must fulfill their own familial caring roles at a distance, employing various strategies in the quest to maintain their intimate social relations from afar.

Conclusion

Immigration is a major feature of the contemporary world and has undoubtedly transformed societies across the globe. Its impacts will continue to be felt into the foreseeable future, as countries with declining birth rates and aging populations turn to immigration as the solution to demographic problems. Temporary forms of labor migration (both skilled and unskilled) are likely to increase, as are numbers of international students, as more and more educational institutions internationalize their student body and curricula, obtaining (through international tuition fees) substantial additional funding.

Immigration is a complex process, implicating immigrants, their families, home societies, and host states in different ways. Traditionally, research has generally focused on the various impacts of immigration upon the host society. In the past, immigration has been conceived as a decisive and invariably permanent move, from one country to another. More recently, new theoretical conceptions have led to a shift in the overall focus of work on migration. While studies of the impact of immigration continue to be important, they have broadened their focus to incorporate an understanding of the home society. Concepts such as transnationalism and circular migration emphasize the fact that many immigrants maintain significant ties with their home country after migration and that some migrants will return to the home country at some point (immigration is no longer seen as necessarily permanent). Despite this, the nation state (and particularly the host country) continues to be of central importance in directing contemporary immigration. Through immigration policies and the granting of citizenship, states retain significant control over who they chose to include and who they chose to exclude. Dominant ideologies of im migrant incorporation (assimilation and multicultural ism) change over time, but continue to reinforce the legitimacy of the nation state. Illegal immigration offers the most potent challenge to this order, and is a major political preoccupation of governments in many Western states.

See also: Multiculturalism; Segregation; Transnationalism.

Further Reading

Baubock, R. (1991). Migration and citizenship. *New Community* 18(1), 27 48.

Boyd, M. (1989). Family and personal networks in international migration: Recent developments and new agendas. *International Migration Review* xxiii(3), 638 670.

Castles, S. and Miller, M. J. (2003). *The Age of Migration: International Population Movements in the Modern World* (2nd edn.). New York: Guilford Press.

Cornelius, W. A., Martin, P. L. and Hollifield, J. H. (1994). *Controlling Immigration: A Global Perspective*. Palo Alto, CA: Stanford University Press.

Frey, W. (1996). Immigration, domestic migration and demographic balkanization in America: New evidence for the 1990s. *Population and Development Review* 22, 741 763.

Jacobson, D. (1997). *Rights Across Borders: Immigration and the Decline of the Citizenship*. Baltimore, MD: John Hopkins University Press.

Joppke, C. (1999). *Immigration and the Nation State: The United States, Germany, and Great Britain*. Oxford: Oxford University Press.

Ley, D. (1999). Myths and meanings of immigration and the metropolis. *Canadian Geographer* 43, 2 19.

Ong, A. (1999). *Flexible Citizenship: The Cultural Logics of Transnationality*. Durham: Duke University Press.

Preston, V., McLafferty, S. and Liu, X. F. (1998). Geographical barriers to employment for American born and immigrant workers. *Urban Studies* 35(3), 529 545.

Sassen, S. (2000). Regulating immigration in a global age: A new policy landscape. *Annals of the American Academy of Political and Social Sciences* 570, 65 77.

Soysal, Y. (1994). *Limits of Citizenship: Migrants and Postnational Membership in Europe*. Chicago, IL: The University of Chicago Press.

Relevant Websites

http://www.migrationpolicy.org/
Migration Policy Institute.

http://www.compas.ox.ac.uk/
The Centre on Migration, Policy and Society.

http://www.canada.metropolis.net/
The Metropolis Project.

Immigration II

I. H. Burnley, University of New South Wales, Sydney, NSW, Australia

Glossary

Acculturation The process whereby an immigrant group adopts many of the social norms and values of the receiving society while maintaining some traditional cultural patterns.

Assimilation The process whereby immigrants adapt to and merge with the receiving society in residence, language acquisition, and social and economic characteristics. As originally used by Park, in 1925, in his influential 'race relations cycle', assimilation also involved the receiving society adjusting to immigrants.

Career Cycle Migration A process wherein persons in the workforce move internationally to further their career prospects by moving up professional and managerial hierarchies, sometimes within transnational corporations or organizations which have branches in several countries.

Ethnoburb A term used by recent authors to designate an outer suburban area of a metropolis in which ehnic entrepreneurship, concentration and cultural mix often occur, and where settlement can be associated with socioeconomic advantage.

Gravitation Migration A form of internal migration within receiving countries when immigrants move interregionally or within a city to be closer to compatriots for communal support.

Migration Cohort Networks Social networks of immigrants who arrived in a country in the span of a few months or years, who may share common experiences of social dislocation and adjustment, and who may form specific in-group networks.

Pull Factors The factors in a receiving country, including economic opportunities, higher living standards, the presence of relatives and friends, and relative political/personal freedom which attract immigrants to settle in a given country rather than another.

Push Factors The factors that influence persons to leave countries of origin, which may include poverty, discrimination, or political oppression.

Reductionist Interpretations Views of migration processes and settlement which oversimplify the complexities involved in community relations and in immigrant geographies.

Replacement Migration A process where international migration takes place and/or is sought by receiving countries because of labor shortages commonly resulting from population fertility decline, and sometimes emigration.

Sojourners A term applied to manual labor migrants – commonly males – who moved internationally in the nineteenth century, usually short term, and who often remitted money to relatives in home countries. Examples were the Chinese gold miners in California, western Canada, Australia, and New Zealand. The term 'new sojourners' has been applied to short- and medium-term skilled and contract workers in contemporary times.

Spatial Assimilation It is the process of, or a measure of, the dispersion of an immigrant origin group and its geographic merging with the wider population distribution.

Transnational Processes The processes whereby immigrants may relate to broader international institutions and economic relations so that a one-way adaptation to a given receiving country or its society is not the dominant issue for the immigrants. The short- or long-term immigrants may relate to each of the sending and receiving societies and other societies and networks at the same time.

Background

Once modern humans emerged in Africa 150 000 years ago, their descendents spread across the world, although constrained by economic, technological, and environmental factors. Few political boundaries existed to deter move ment, for the nation state, with its customs barriers, pass port, and border patrols, did not evolve until the late sixteenth century, and then mostly in Western Europe. Until the late twentieth century, there were notable em pires across which people moved. In the Ottoman and Roman Empires, Istanbul (Constantinople) and Rome be came large cities, as people of many cultures migrated to these imperial capitals for social and economic gain. Later, Paris, London, Madrid, Lisbon, and Vienna attracted mi grants from their imperial lands, and when empires dis solved, cultural connections remained. Examples are Bangladeshi and Pakistani immigrants in London; Al gerians in Paris, Lyon, and Marseilles; and Brazilians in Lisbon (the South American colonies became independent in the nineteenth century; those of Britain and France after 1950). When the great empires began formal colonization, local peoples were pushed back or subjugated and ex sol diers given land in conquered territories to control in digenous communities. Slave migration was involved in several early modern empires and continued in the USA until 1860. Migrants moved within empires, for example, persons from India to Eastern and Southern Africa and Fiji where many became small proprietors.

There were significant migrations of 'sojourners' from China in the nineteenth century – particularly to Canada, the USA, and Australia. There were constraints later in movements from Asia, particularly with the rise of racist sentiments in immigration countries during the late nineteenth century. Legislation limited migrations from India to Australia; Japan to California; and from China to Australia, Canada, New Zealand, and the USA, so that 'mass' migrations from Asia to Western countries did not occur. However, significant minority communities formed in the late nineteenth century, for example, Chinese in Vancouver, San Francisco, and Sydney and Gujaratis from India in London and in Wellington, New Zealand.

Immigration in the Nineteenth and Twentieth Centuries

Had it not been for emigration from Europe to North and South America and Australasia (Australia and New Zealand) – when population growth in Europe was rapid between 1850 and 1920 – population pressures in parts of Europe might have been calamitous. **Table 1** shows immigration in the Americas, Australia, and New Zealand from 1820 to 2006 based on estimates, secondary sources, and arrival and departure statistics where available. These great flows took place as the supposedly unpopulated New Worlds were opened for European exploitation, and transformed from subsistence production to export pastoral and mineral based economies while the urbanization ideal was transplanted. The strong push factors in Europe included crowded towns resulting from industrialization, land clearances, technological change in agriculture, inequality, population pressure, and sometimes oppression. Pull factors included perceived economic opportunities and space in the New World, and as families became established, these social units became migration magnets for others. While much immigration to New World rural areas occurred in the nineteenth century, increasingly, international migration became associated with urbanization and particularly, metropolitanization.

There were relatively few government restrictions on immigration to the USA in the nineteenth and early twentieth centuries. Self selectivity took place: immigrants tended to be healthier young adults, often with children, although males commonly migrated first and sent for their wives and children later. International migration in the nineteenth and early twentieth centuries was a great personal sacrifice – 3 weeks by ship to the USA and Canada, and months in the case of South America, Australia, and New Zealand. Many who did not reestablish families in the countries of immigration never saw their relatives again because of the cost and time involved in returning. Contemporary international migration is mostly by airplane, is more rapid, and less costly, although financial costs are still significant.

Table 1 Estimated immigration in the Americas, Australia, and New Zealand (1820 2006)[a]

Region	*Immigration*
United States	70 000 000
Canada	16 000 000
South America	45 000 000
Australia	11 000 000
New Zealand	2 000 000

[a]Temporary and short term migrants excluded.

Adapted from: Burnley, I. H. (2001). *The Impact of Immigration on Australia*: *Demographic Approaches*. Melbourne: Oxford University Press.

Handlin, O. (1960). *The Uprooted.* Cambridge, MA: Harvard University Press.

Jupp, J. (2000). *Encyclopaedia of the Australian People (Second Edition)*. Melbourne: Cambridge University Press.

SC (Statistics Canada), 2001, Population Census, Canada, Ottawa.

ABS Australian Bureau of Statistics, *2006 census, Birthplace by Duration of Residence Tables.*

USA Bureau of the Census (1971). *Statistical Abstract of the United States, 1971.* Washington: US Government Printing Office.

USA Bureau of the Census (2000). Population Census, USA., Washington.

USA Bureau of the Census (2003). *Data sources on the Foreign Born and International Migration at the US Census Bureau,* Washington, DC., January 1.

UN (United Nations) *Demographic Yearbooks,* 1978; 1990, 2000.

Perhaps the relatively greatest migration flow from any country was the movement of 4 million from Ireland to the USA after the 1840s famine in which 1 million died. Emigration was so strong that today, the population of Northern and Southern Ireland (Eire) is 5 million instead of the almost 9 million at the beginning of the nineteenth century. All told, 15 million migrated from England, Scotland, Wales, and Ireland during the reign of Queen Victoria, 1837–1901. Most preferred America for it was only weeks away and much cheaper than Australasia although over 2 million settled in Australasia and numbers in South Africa. Almost 4 million migrated to the USA from Italy, as many from Germany, hundreds of thousands from each of Switzerland, Sweden, Denmark, and Greece, and several million from Poland, Russia, and other parts of Eastern Europe.

These migrations had major impacts on the growth of large US cities (New York, Boston, Chicago, and Pittsburgh). The Jewish migration to New York was very strong, and there were over 2 million there at the 2000 census; in the 1920s, there had been a million Irish and Italians in New York and large numbers in Boston. More than 100 years on, specific neighborhoods can still be considered as Irish American or Italian American, although American English, with dialect influences is spoken, and many descendents have been residentially and socially mobile.

In 1924, the US government instituted the Quota Acts, which limited immigration from Southern, Eastern Europe, and the Middle East. With the exception of refugee movements, international migration fell in the 1930s and did not revive until after World War II (WWII).

Within Europe, members of many ethnic groups migrated to countries speaking their language after boundary changes in Eastern Europe and Asia Minor as the result of World War I (WWI) – for example, Ger mans moving Westward: at the end of WWI, 13 million Germans were left outside the Reich boundaries and many suffered from national chauvinism in East Europe. In 1923, over a million Greeks were forced out of Ana tolia and Thrace by a renascent Turkey, most having to be absorbed in Greece with a population base of only 8 million. Many Poles moved into former areas of Prussia in the early 1920s as the newly independent Poland ex panded its territories. There were many other related movements and after WWII, over 2 million Germans migrated from Bohemia to Germany and many persons from Greek communities in Bulgaria, Romania, and Yugoslavia moved to Greece or into the wider Greek diaspora around the world. Many of these movements in Europe have had lasting social impacts.

After WWII, America wished to attract immigrants for economic reasons, and Canada and Australia for labor force and population building reasons.

While there were liberal immigration targets, the state in the receiving countries exercised more control over immigration than hitherto. There were refugee waves as the result of WWII itself and the Cold War conflicts: the Korean War with Korean influxes in the western USA, including inner Los Angeles; the Hungarian revolution (1956); the Cuban revolution with immigration impacts on Miami, Florida; the Prague Spring repression (1968); the Vietnam War; the Soviet invasion of Afghanistan (1980s); the Balkan conflicts of the 1990s; oppression of dissidents in the Middle East; civil wars in Africa; and repression in parts of South and Central America. Refugees were taken in adjacent countries and also de veloped countries such as the USA, Canada, Australia, and more latterly, Britain and New Zealand, and north west and northern Europe.

Economically motivated migration from poor, war devastated countries of Europe was stronger to North America, Canada, Australia, and after the fall of com munism, from east to northwest Europe. Over 1.8 million people moved from Britain to Australia and more to Canada. Many Jewish survivors from the Holocaust mi grated to Israel, as well as over a million Jews from the former Soviet Union, augmented from members of the Jewish diaspora in northern Africa, and the Middle East.

Western Europe had labor shortages post war due to the low fertility of the depression era, war casualties, and emigration, and West Germany imported 'guest workers' from Turkey who found difficulty in obtaining citizen ship, more especially when their preferences were to remain rather than to return to more poverty stricken conditions. Many remitted money to relatives at home. Other migrations occurred to France and northwest Europe from Algeria, Morocco, and Tunisia. With the formation of the European Economic Community (EEC), labor flows between member states occurred, especially with the induction of several former Soviet bloc countries into the EEC. Migration from Poland and Russia to the UK, for example, has become substantial. Many of these immigrants have found employment in the service economy.

Globalization influenced the migration of skilled persons between developed countries, between de veloping and developed countries, and more recently, to some developing states, including oil rich Middle East countries. Such migrations have impacted particular 'global cities', notably New York, London, Los Angeles (30% born overseas in 2000); Toronto, Canada (42%); and Sydney, Australia (31%). In these cities, cultural diversity became a criterion for locational decision making and the attraction of entrepreneurs and the highly skilled, although some global cities such as Frankfurt (Germany) and Tokyo (Japan) have had small immigrant populations. Rising real incomes and global economic organizational change not only generated movement of highly skilled persons who diversified middle class areas in many Western cities, but also at tracted less well paid service workers, some from de veloping countries, for example, the migration from the Philippines to Australia, parts of Southeast Asia, oil rich Middle East states, and more recently, to Japan. Many of the women migrants from the Philippines have had ter tiary education but often have not found commensurate employment in the receiving country.

Globalization also gave rise to a large expansion in short term working visa holder movements which might be followed by permanent residence, return migration, or movement onward to another country. Sharp birthrate declines in many Western countries from the 1960s generated contracting workforces later so that 'replace ment migration' (immigration) became crucial to prod uctivity. Several developed countries became competitors for professional, information technology, and skilled trades' workers. Most competing countries, nevertheless, became selective in their immigration intakes, with points accumulated for relative youth (under age 35–45), edu cation, qualifications, occupations needed, relational ties, and health.

Another aspect of economic globalization is growing movement between immigrant receiving developed countries. By 2006, over 1 million Australia born (5%) and 850 000 New Zealand born (18%) resided overseas. Almost 200 000 Australia born lived in East Asia (contract workers) and in rapidly developing Southeast Asia, with more in the USA, continental Europe, and Britain. Some of this involved return migration of second generation members. In 2005, the former emigration country, Eire, through prosperity resulting from its European Commission (EC) membership embarked on an immigration target of 100 000 a year.

Global dispersion of people with common cultural identities – diaspora – has become an important area of enquiry. This includes people who have been economically mobile but who have suffered discrimination, for example, Jews, Armenians, and Romanys, or migratory groups such as the Chinese in Southeast Asia and beyond and persons from India, Ireland, Greece, and Italy. Some studies focus on identities of people in diaspora, or their economic, social, and cultural networks.

A Typology of Immigrant Flows

Table 2 outlines a typology of immigration, focusing on the last two centuries. While it refers more to Western countries, some categories have relevance to non Western countries, notably chain migration, refugee, and some economic aspirational movements. The categories are generic rather than visa types, for visa categories and nomenclatures vary between countries. Some 'permanent' migration was not intended as such by the individuals involved, and some migration intended to be permanent becomes short term because of migrant dissatisfaction or other factors.

Permanent Migration

'Chain migration' occurs when individuals, usually men have migrated first, established themselves, and sent money for *fiancés*, wives, and children to migrate. In the USA before the Quota Acts (1924), and in Australia and Canada post WWII, wider family networks were involved: these partly reflected the close bonded nature of rural sending societies and sometimes as with Greeks and Southern Italians in the USA, Canada, and Australia, substantial components of small village and town populations migrated over decades. Chain migrations strongly influenced group settlement with Italians in inner western Toronto, in Boston, and east side New York, Buenos Aires (Argentina), and Sydney and Melbourne. In Melbourne, for example, a community formed of over 1000 persons from the Italian town of Viggiano in the inner city. Chain migration has recently been important, with migration of Indians from Fiji to Australia and to New Zealand, particularly Sydney and Auckland. Chain migration has been widely involved in migration from former colonies and former League of Nations mandated states such as the Cook Islands and Western Samoa to New Zealand.

Chain migration was historically involved with rural group settlements, as with Swiss, Germans, and Danes in the USA Midwest; Japanese in rural Brazil; Germans in South Australia; and Ukrainians in Alberta and Saskatchewan, Canada. Religion had been a factor when persons were discriminated against at origin – as with Evangelical Lutherans in Prussia and Hutterites and Mennonites in East and Central Europe – and who migrated to the Midwestern USA and Canada to form new rural communities and preserve ways of life under threat.

'Independent (aspirational) migration' typically includes economically motivated skilled migrants, and although some may be part of transnational elites, many migrate as individuals or nuclear families for aspirational reasons and may not form local ethnocultural links. Others may assist relatives to migrate so that chain migration results. Like chain migrants, they may send remittances to home communities – a process which is very important in the Southwest Pacific. Sponsored relatives of independent skilled migrants may have lesser education because family reunion criteria may not emphasize skills.

'Refugee migration', unlike chain migration which continues over a long period, occurs in peaks as conflicts emerge. Great conflicts, such as the world wars resulted in major displacements of population with national boundary changes. Substantial numbers of displaced persons after WWII were settled in France and Britain, and especially the USA, Australia, and Canada. These refugee intakes did not seed chain migrations because the Communist countries restricted emigration.

The Vietnam War resulted in refugee waves of over 150 000 in each of France, Canada, and Australia and many more in the USA. Agreements between the new government of Vietnam and some of these countries allowed subsequent and orderly transfers of relatives of the first refugees. Throughout 1980–2005, there was an annual average of 20 million persons housed in refugee camps globally, mostly in Africa, the Middle East, and Southwest Asia, and the UN goal has been to resettle refugees in their home countries when circumstances normalize. This is often impossible and some countries have annual refugee quotas, for example, Australia with 12 000 a year.

'Illegal immigration' occurs because many refugees cannot be officially settled overseas, some have journeyed to receiving countries as 'the new boat people' and asylum seekers, landing unannounced on remote coasts, for example, in Western Australia. A newer phenomenon has been 'people smuggling' in which asylum seekers under threat from repressive authorities have given their life savings to criminal groups to transport them by boat (or plane) via intervening countries to Western Europe, the USA, and Canada. Illegal migration of Hispanic migrants from the Caribbean and Mexico to the USA has taken this form, and many have died *en route*, while others made their way across the relatively porous border of Mexico. In 2006, Federal legislation restricted the latter by erecting barrier fences but granted amnesties to some of the estimated 15 million illegal immigrants in the USA, many of whom were absorbed into urban localities where other Hispanic persons settled.

'Officially assisted migration' occurred to only a few countries, notably from Britain to Australia and New Zealand. This was necessary because of the cost involved for migrants, in the colonial period and up to the early 1980s (although many migrated independently). Both countries had population building policies, but before the 1970s they preferred 'Anglo Celtic' people; although post WWII, Australia signed bilateral agreements with Malta, West Germany, the Netherlands, Italy, and Greece, and later, Turkey and the former Yugoslavia, the latter being the only European communist country to allow substantial emigration.

'Situational change migration' occurs when onshore visa change takes place and foreign tertiary students obtain permanent residence. Previously, states such as Australia and Canada required overseas students to return to their home country for a period before they could apply to migrate. Since the late 1990s, Australia has allowed student graduates to apply for permanent residence, with additional incentives (lower entry point scores or higher weightings) if they contracted to live outside the main cities for a period. These onshore conversions were allowed because of skill shortages and international competition for professional and technical workers.

Temporary Migration

Increasingly, the distinction between permanent 'settler' migration and 'temporary migration' is blurred and the distinction between **Tables 2** and **3** is somewhat forced. Many post WWII permanent immigrants – once living standards rose in sending and receiving countries – returned to their home countries to live (almost a third of the Greece born in Australia, for example). Canada and Australia signed bilateral agreements, whereby the country of origin and the receiving country contributed to age pensions, provided the immigrant had been in the country of immigration's workforce for 20 or so years. More latterly, many permanent residents have moved on to third countries, particularly if they have qualifications which are internationally recognized. Such movements increased the diaspora of ethnocultural groups in a transnational context.

Guest-worker migration

Post war, West Germany and some other European countries offered 'guest worker' status to migrants from countries with poorer living standards, such as Turkey, for there were low income jobs which the native born would not do. An intrinsic problem was that permanent residence, and particularly, formal citizenship, of the country of residence was difficult to obtain. This raised conflicting issues of assimilation, institutional and structural maintenance of segmented labor forces, and residential segregation in low socioeconomic status areas.

Working holiday makers and backpackers

Several British Commonwealth nations negotiated agreements allowing young adults to move between countries to work and travel. This and the wider 'backpacker' phenomenon have transformed rental neighborhoods in London, Paris, Toronto, Vancouver, Sydney, and Melbourne. In sizeable cities, investors have taken advantage of this cheaper rental housing, often intermixed with overseas born student accommodation. Temporary

Table 2 Typology of 'permanent' immigration

Migration category	*Motivation and circumstances*	*Characteristics*
Slave, indentured	Forced; slave movement abolished in Americas in nineteenth century.	Young persons.
		Indentured persons in some imperial colonies to 1900.
Chain	Economic push factors; movement to join relatives. Remittances.	Men, women, and children meeting relatives in country of settlement.
Independent	Economic push pull factors; aspirational and individualistic motives. Remittances.	Higher occupation status; skilled manual workers. Migration as part of life course.
Officially assisted	Poorer immigrants: state assistance with transport costs. Skilled assisted by private firms.	Families; persons in workforce ages; contractual obligations earlier (Australasia).
Refugee	Political and religious oppression at origin. Willingness of receiving country to accept.	Families, parts of families; individuals.
Illegal	(a) Poverty at origin.	Young adults, accepting low-paid work (Europe; USA).
	(b) Poverty and discrimination at origin; people smuggling.	Individuals and families: Europe, USA; Australia.
Situational change	On-shore visa change; overseas students obtaining residence because of skills.	Young adults with skills and higher education; developed countries.

Table 3 Typology of 'Temporary' Immigration

Migration category	*Motivation/circumstances*	*Characteristics*
Sojourners	Economic push factors; opportunities in nineteenth-century resource industries.	Primarily men; workers on goldfields, agriculture, unskilled urban work. North America, Australia.
Seasonal migration	(a) Transhumance.	Central Asia and parts of Africa. Nomadic males and families following seasonal resources.
	(b) Sports and tourism workers.	Mostly young adults between developed countries.
Guest-worker migration	Poverty at origin. Remittances.	Europe: unskilled; problems of legal citizenship.
Working holiday migrants	Varying affluence.	Young adults moving between developed nations, and from some rapidly developing countries.
Short-term skilled migrants.	(a) Career cycle movers.	Young and mature workforce; some families. Developed nations.
	(b) Contract workers.	Several developing and resource-rich states.

immigrant employment in seasonal harvesting has be come critical because of labor shortages. In 2007, an agreement between the USA and Australia made it pos sible for high school leavers to spend their 'gap year' in the respective countries and obtain temporary work there. Some backpackers and working holiday makers eventually return as 'permanent' immigrants.

Medium-term skilled visa holders

This form of 1–5 year migration involves individuals rather than families and is of growing importance be cause of skilled labor shortages in many Western coun tries. In Australia, between 1997 and 2003, short term movers outnumbered permanent 'settler' migrants by 2.5:1. Some obtain permanent residence later. For some professionals and managers, short medium term visa migration is primarily international career cycle migra tion: persons move within multibranched transnational organization offices in global cities and young scientists, engineers, and academics obtain contract employment with universities, research institutions, and corporations. A problem which has emerged, however, that numbers on skilled visas may be employed in the immigration country in lower skill levels and poorer rates of pay despite the illegality of this practice.

Forms of contract skilled migration occurs from de veloped countries to emerging resource rich states in the Middle East, such as Saudi Arabia and Dubai, notably involving engineers, doctors, nurses, teachers, and tech nicians. This form of short term movement seldom leads to permanent residence or formal citizenship, as in the case of businessmen and professionals and aid workers in postcolonial African states.

Seasonal international migrations

Seasonal international migrations occur across national borders in traditional regions where nomadic, herd keeping, and harvesting movements take place in Central Asia and some parts of Africa. Other seasonal migrations between developed countries in different hemispheres occur in professional and alpine sports and with tourism. Seasonal and other short term migrations occur between poorer Southeast Asian countries and resource rich Middle East countries.

Migration processes influence residential patterns of immigrants and also the cultural geographies of cities and regions. These processes may be more important than group characteristics in adjustment.

Immigration and Residence

In the 1920s, the Chicago School of sociology had a marked influence on social geography, notably the con centric zone model of cities and Park's 'race relations' cycle which incorporated the relationship between spatial residence patterns and assimilation. With the concentric zone model, immigrants settled in zones 'of transition' and 'working mens' homes'. Inner city residence of low income immigrants in US cities took place because cheap housing and industrial employment were located there. Whereas low income immigrant communities tended (and sometimes continue) to settle in the inner city areas of US cities; in Australian cities, more recently, most members of such communities settled in low mixed in come outer suburbs.

The 'Dissimilarist' Approach

In the 1950s and 1960s, quantitative researchers applied indexes of dissimilarity, segregation, and concentration to immigrant groups and African American populations in US cities, and this approach spread elsewhere. The 'index of dissimilarity' summed the percent differences between two populations by spatial units within a metropolis and divided the difference by two. The 'index of segregation' compared the distribution of a given group with that of the rest of the population. An index over 80 implied that a group approached segregation.

There were (and are) several problems with these indexes. First, the reductionist notion that spatial distance is equal to social distance and that spatial concentration means less 'assimilation'. Recent studies have shown that there can be positive benefits to the immigrant community and wider society from ethnic residential concentration. Second, the indexes did not allow direct inferences as to which group was segregated from which. Third, the indexes allow little interpretation of processes of residential adaptation. Fourth, the magnitude of the index values is influenced by the number of spatial units which limits comparative analysis between cities and in the same city at different times. Fifth, residential areas commonly vary by socioeconomic status, and the index value may simply reflect the occupational status profile of the immigrant group (although with appropriate data this can be statistically controlled). Sixth, other factors may influence residence: a population pattern may reflect structural rental housing factors in relation to status and income, rather than cultural choices. Seventh, the index is a summary value: in reality, part of an immigrant group may be strongly segregated, part moderately dissimilar, and the remainder dispersed. Eighth, a conceptual danger is that the observer might conclude that one immigrant group is more concentrated than another and *ipso facto* less assimilated than the other. Negative policy ramifications could result. Ninth, strong spatial dissimilarity does not necessarily mean limited integration: in areas of strongest concentration of persons born in Greece, Italy, Lebanon, and the former Yugoslavia, in Sydney, for example, considerable 'acculturation' to Australian norms took place. Notwithstanding these caveats, the indexes can be useful.

Newer Threshold and Related Approaches

The newer 'threshold' statistical approaches address some of the issues raised above. Threshold approaches use multiple indicators. Proportions of immigrant or ethnic groups comprise determined percentages of area populations within cities or regions. The percentage of given birthplace or ethnocultural groups at given levels of concentration – in areas where they are less than 10%, 10–29%, 30–39%, 40–50%, or higher – is determined. This allows conclusions about the proportions of ethnocultural groups which contribute a majority of the population or significant sections of the population in a metropolis, and permit more balanced perspectives on 'segregation' or 'ghettoization'.

At this point, the term 'ghetto' is reviewed. Adapted from the label applied to areas of Jewish concentration enforced in some European cities, it was applied by Chicago School researchers to areas of US cities where strong majorities of Afro American people settled: descendents of slaves in the south were forced to reside in specific districts because of unconstitutional state segregation laws. Before the 1960s Civil Rights legislation, there were restrictive covenants precluding Afro Americans (and Jews) from residing in particular northern USA city districts.

Taking Peach's dual criteria that most of a group within a metropolis must reside in a particular area and comprise a strong majority of the population there for it to be termed a ghetto, recent studies using threshold approaches have found that few immigrant ghettos exist in British, American, or Australian cities, although immigrant concentrations exist. In Australia, at least, immigrant concentrations seldom persist over more than one generation and much residential mobility occurs within first generations when employment levels are high.

A dynamic variant of the social distance hypothesis is Boal's theory of ethnic concentrations: individuals form residential clusters for social support when they do not speak the host society language, when individuals face discrimination and racism, and where they develop a more organized defense of the group against overt racism, particularly recent arrivals. Four adjustment types were suggested: 'assimilation' – difference reduces and spatial boundaries dissolve; 'pluralism' – group boundaries are maintained although members participate in the wider society; 'segmentation' – much sharper divides occur because of group antagonisms; and 'fragmentation' (extreme segmentation or hypersegregation) in which the social environment is fractured, involving exclusion and ghettos. It has been argued that hypersegregation occurs in several large US cities, particularly among persons of Afro American and Hispanic origin.

Some migration processes can have direct residential and social adjustment outcomes. Chain migration can result in distinctive group settlement as family networks coalesce over time. Examples were Portuguese and Italian settlement in inner western Toronto, Sicilians in inner Boston, and Calabrians in western Sydney (Australia). Often it is networks from a subregion that come to the fore, rather than nation of origin. Indeed, in Australian cities, kinship and family have often much more importance in residence of persons from Asia than ethnicity. Of course, the existence of a multiplicity of kinship groups speaking a common language or dialect attracts institutions, small business, local associations, and churches, temples, mosques, or synagogues.

Refugee migration can also lead to group settlement and as with chain migration, housing costs and socioeconomic status are influential. Immigration hostels can affect later refugee residence because adjacent areas become familiar during hostel residence; migration cohort networks become important, in the absence of previous institutional structures and without longer established familial networks such as with chain migration; gravitation migration to particular cities from dispersed initial employment locations may occur for communal support although dispersion to other residential areas often takes place; entrepreneurial opportunities in ethnoculturally identified areas can occur, and this in turn can attract employees; groupings can also occur in social housing estates for reasons of poverty and special familial cir cumstances, such as in social housing in outer parts of London and Paris.

Factors Influencing Admixture

Persons from cultural and religious groups in diaspora before migration often form mixed communities of resi dence afterward, reinforced by the presence of relatives, services, and religious institutions. These can be skilled, business, or family reunion migrants among whom varying origins, dialects, and housing amenity can be important. Rather than forming homogeneous concentrations, the communities may have diverse origins within the same linguistic group as with Cantonese speaking immigrants in middle and higher income suburbs of Sydney, where persons from Hong Kong, Taiwan, southern China, Ma laysia, Singapore, and Indonesia settled in new apartments or in established houses but often intermixed with Aus tralian born and other immigrants.

Immigrant entrepreneurship can transform poorer areas of immigrant concentration, as with Caribbean and African migrants in Brixton in South London, Chinese and Vietnamese immigrants in Cabramatta in western Sydney, and Chinese immigrants in higher income areas in Rich mond, Vancouver. Terms such as 'ethnoburbs' have been used for instances where new suburban growth, re development, and residential concentration go hand in hand. This raises again the terms used for immigrant concentrations, including mixed local populations. The terms 'ghetto' and 'enclave' are not appropriate in a Western city context. 'Enclave' normally involves a largely culturally distinct and homogeneous population within prescribed areas with separate laws and sometimes ad ministration, as in parts of the Balkans and in some of the former Soviet Central Asian republics. In these cases, the ethnic communities often predated the formation of nation states in which they are situated. In developed Western cities, ethnocultural concentrations abide under the same law as the wider population, so the term is misleading. Other terms such as ethnic 'precincts', 'quar ters', or 'colonies' can also imply homogeneity or entities with barriers even if this is not intentionally the case.

Changing factors within receiving societies alter spatial adjustments. Social networks often spread widely, once immigrants have cars, for they do not have to live in the same local areas to have face to face contact with relatives and friends. 'Heterolocalism' may occur when the ease of mobility in modern cities allows dispersion 'and' the maintenance of contacts within dispersed communities.

Changing expectations within newer streams of highly educated migrants mean that permanent settlement may not be contemplated even if the immigrants have per manent residence or formal citizenship, and they may have no desire for residential cohesion within their origin group. Furthermore, such immigrant networks and the associated social capital are often no longer confined to the same city or society. Indeed, the growth of transna tional communities through migration may mean persons being 'at home' in several places simultaneously without needing to form clusters for support. In addition, trans national social spaces may represent different modes of incorporation in the wider social system than have been interpreted from traditional settlement studies.

See also: Demography; Emigration; Migration; Population Geography; Refugees and Displacement; Transnationalism.

Further Reading

Anderson, K. (1991). *Vancouver's Chinatown, Racial Discourse in Canada 1875 1980*. Montreal, QC: McGill Queens University Press.

Bedford, R. D. (1994). Pacific islanders in New Zealand. *Espace, Population, Societies* 18, 187 200.

Bedford, R. D. and Levick, B. (1998). Recent migration from Fiji to New Zealand. *New Zealand Population Review* 14, 69 81.

Birrell, B. (1992). *The Chains That Bind*. Canberra: Bureau of Immigration and Population Research, Australian government Publishing Service.

Boal, F. W. (1999). From undivided cities to undivided cities: Assimilation to ethnic cleansing. *Housing Studies* 14, 545 600.

Burgess, E. W. (1925). The growth of the city: An introduction to a research project. In Park, R. E., Burgess, E. W. & Mc Kenzie, E. (eds.) *The City*. University of Chicago Press: Chicago.

Burnley, I. H. (2000). Diversity and difference: Immigration and the multicultural city. In Connell, J. (ed.) *Sydney, the Emergence of a World City*. Oxford University Press: Melbourne.

Burnley, I. H. (2005). Generations, mobility and community: Geographies of three generations of Greek and Italian ancestry in Sydney. *Geographical Research* 43(4), 279 392.

Burnley, I. H. and Hiebert, D. (2001). Emerging patterns of immigrant settlement at the metropolitan scale: The need for new concepts and models. *Progress in Planning* 55(3), 127 140.

Castles, S., Booth, H. and Wallace, T. (1984). *Here for Good: Western Europe's New Ethnic Minorities*. London: Pluto Press.

Chen, Y. (2000). *Chinese San Francisco 1850 1943: A Trans Pacific Community*. Stanford: Stanford University Press.

Conradson, D. and Latham, A. (2005). Transnational urbanism: Attending to the everyday practices and mobilities. *Journal of Ethnic and Migration Studies* 32(2), 227 233.

Dinnerstein, L., Nichols, R. L. and Reimers, D. M. (1979). *Natives and Strangers: Ethnic Groups and the Building of America*. New York: Oxford University Press.

Dunn, K. M. (1998). Rethinking ethnic concentrations: The case of Cabramatta, Sydney. *Urban Studies* 35, 503 527.

Dunn, K. M. (2005). A paradigm of transnationalism for migration studies. *New Zealand Population Review* 31(2), 15 31.

Faist, J. (1998). Transnational social spaces of international migration: Evolution, significance and future prospects. *Archive European Sociologique* 39, 213 247.

Ferguson, N. (2006). *The War of The World: Twentieth Century Conflict and the Decline of the West*. London: Penguin.

Friedman, J. (2002). *The Prospect of Cities*. Minneapolis, MN: University of Minnesota Press.

Handlin, O. (1960). *The Uprooted*. Cambridge, MA: Harvard University Press.

Heisler, B. S. (2000). The sociology of migration. In Brettell, C. B. & Hollifield, J. (eds.) *Migration Theory: Talking Across Disciplines*, pp 77 96. Routledge: New York.

Hugo, G. J. (ed.) (1989 98). *Atlas of the Australian People: State Volumes* Canberra: Australian Government Publishing Service.

Hugo, G. J., (2003). Temporary migration to Australia: Trends and implications, *Keynote Paper Presented to the 22nd Conference of the New Zealand Geographical Society*. Auckland, New Zealand, July.

Iredale, R., Hawksley, C. and Castles, S. (eds.) (2003). *Migration in the Asia Pacific: Population, Settlement and Citizenship Issues*. Cheltenham: Edward Elgar.

Johnson, H. B. (1940). German settlement in the mid west of the United States. *Annals of the Association of American Geographers* 16(2), 235 260.

Johnston, R. J., Forrest, J. and Poulsen, M. (2001). Sydney's ethnic geography New approaches in analyzing patterns of residential concentration. *Australian Geographer* 32(2), 149 164.

Johnston, R. J., Forrest, J. and Poulsen, M. (2002). From modern to post modern? Contemporary ethnic residential segregation in four US metropolitan areas. *Cities* 19, 161 172.

Johnston, R. J., Poulsen, M. and Forrest, J. (2002). Are there ethnic enclaves/ghettos in English cities? *Urban Studies* 39, 591 618.

Kaplan, D. and Holloway, S. (1998). *Segregation Geography*. Washington, DC: The Association of American Geographers.

Keely, D. B. (2000). Demography and international migration. In Brettell, C. B. & Hollifield, J. (eds.) *Migration Theory; Talking Across Disciplines*, pp 98 122. Routledge: New York.

Laurence, J. and Vaisse, J. (2006). *Integrating Islam: Political and Religious Challenges in Contemporary France*. New York: Brookings Institution Press.

Lee, E. S. (1966). A theory of migration. *Demography* 3, 47 57.

Ley, D. and Kobayashi, A. (2005). Back to Hong Kong? Return migration or transnational sojourn? *Global Networks* 5(2), 111 127.

Lieberson, S. (1963). *Ethnic Patterns in American Cities*. New York: Glencoe Free Press.

Lochore, R. A. (1949). *From Europe to New Zealand, an Account of Our Continental European Settlers*. Wellington: Department of Internal Affairs.

Mac Donald, P. (2006). *Exiles and Emigrants: Epic Journeys to Australia in the Victorian Era*. Canberra: National Museum of Australia.

Massey, D. S. (1999). *Worlds in Motion: Understanding International Migration at the end of the Millennium*. New York: Oxford University Press.

Massey, D. S. and Denton, N. A. (1989). Hypersegregation in United states metropolitan areas black and Hispanic segregation across five dimensions. *Demography* 26, 373 391.

Peach, C. (1996). Does Britain have ghettos? Transactions of the Institute of British Geographers. *New Series* 22(1), 216 235.

Poulsen, M. and Johnston, R. J. (2000). Ethnic enclaves in New Zealand? *International Journal of Population Geography* 6, 325 347.

Poulsen, M. and Johnston, R. J. (2000). The ghetto model and ethnic concentration in Australian cities. *Urban Geography* 21, 26 44.

Poulsen, M., Johnston, R. J. and Forrest, J. (2000). Intra urban ethnic enclaves; introducing a knowledge based classification method. *Environment and Planning A* 33, 2071 2082.

Price, C. A. (1960). Immigration and group settlement. In Borrie, W. D. (ed.) *The Cultural Integration of Immigrants*, pp 250 273. UNESCO: Paris.

Price, C. A. (1963). *From Southern Europe to Australia*. Melbourne: Melbourne University Press.

Price, C. A. (1974). *The Great White Walls Are Built: Restrictive Immigration to North America and Australasia 1836 1888*. Canberra: Australian Institute of International Affairs and Australian National University Press.

Samers, M. (2002). Immigration and the global city hypothesis: Towards an alternative research agenda. *International Journal of Urban and Regional Research* 26(2), 389 402.

Sassen, S. (1991). *The Global City; New York, London, Tokyo*. Princeton, NJ: Princeton University Press.

Sassen, S. (1995). On concentration and centrality in the global city. In Knox, P. & Taylor, P. (eds.) *World Cities in a World System*. Cambridge University Press: Cambridge.

Smith, M. (2001). *Transnational Urbanism: Locating Globalisation*. Malden, MA: Blackwell Publishers.

Stanwix, C. C. and Connell, J. (1995). To the islands: The remittances of Fijians in Sydney. *Asia and Pacific Migration Journal* 4(1), 69 88.

Voigt Graf, C. (2005). The construction of transnational spaces by Indian migrants in Australia. *Journal of Ethnic and Migration Studies* 35(2), 365 384.

Yeoh, B. and Willis, K. (2005). Singaporean and British transmigrants in China and the cultural politics of 'contact zones'. *Journal of Ethnic and Migration Studies* 31(2), 269 285.

Zang, X. and Hassan, R. (1996). Residential choices of immigrants in Australia. *International Migration* 34, 362 382.

Zelinsky, W. and Lee, B. (1998). Heterolocalism: An alternative model of the socio spatial behaviour of immigrant ethnic communities. *International Journal of Population Geography* 4, 281 298.

Imperial Cities

A. D. King, State University of New York at Binghamton, Binghamton, NY, USA

Definitions

An empire is a form of polity where one state extends its dominion, usually by military power, over the territories and populations of others, generally ethnically and culturally different. In this context, imperial city can refer to both the capital and seat of the emperor and also to other major cities in the empire. Interest in imperial cities (more of a historical than contemporary phenomenon) is likely to have been generated in recent years following the emergence in urban studies of the category of the world or global city and, in cases, of the historical origins of that status in an earlier role as imperial city. An additional explanation may be the formal demise of European colonial empires in the mid twentieth century and the subsequent appearance of postimperial economic, cultural, and especially demographic features in what were previously Europe's imperial cities such as Paris, London, or Lisbon.

Empires, Nations, and Cities

From a long term historical perspective, the dominance of the nation state as a form of polity is relatively recent. Currently, the United Nations Organization recognizes almost 200 independent states, each with its own national capital city. Yet at the end of the eighteenth century, there were only some 20 of what today would be accepted as nation states. The increase has occurred mainly in the nineteenth and especially the twentieth century. Between 1945 and 1990, for example, the number of independent states tripled, from 67 to 186, following the collapse of European imperial regimes, especially in Asia and Africa. As recently as the late nineteenth century, the world had been dominated by some 16 empires, each with their imperial capitals and other important imperial cities. That the six official languages of the United Nations Organization are the languages of what were historically once the world's largest empires (British, Russian, Spanish, Arabic, Qing Chinese, and French) is sufficient comment on the past and continuing importance of previous empires and their principal cities in disseminating their languages, forms of law and governance, and religions and cultures. From the most ancient to the most modern, over 80 different empires have been recognized over the last two millennia, their realms, regimes, and ideologies physically demarcated in both space and stone. For example, in the second century BC, Qin Shi Huang, first Chinese Emperor and founder of the Qing dynasty, united the majority of the Han Chinese under a government based in what is now the modern city of Xi'an, from here introducing a unified legal code, a common currency, and a written language. From this central imperial capital, the spatial extent of the Emperor's rule was marked by the first stages of China's Great Wall, a massive defensive and symbolic edifice, later to be extended under the Ming and other dynasties. Subsequently, in the tenth century, China's imperial capital was moved to Peking (modern Beijing), where the Imperial Palace was situated at the heart of multiple layers of walls and gates, enclosing the Forbidden City, an ingenious spatial mechanism simultaneously projecting and also concealing imperial power.

Simply expressed, towns and cities were historically first created by the dominance of some social groups over others and their ability to appropriate from the less powerful an agricultural surplus; in return the more powerful urban faction might provide different services for the rural populace. If, by the same logic, national capital cities exist by commanding a surplus from the resources and territory of the nation state and providing services in return, imperial capitals grow even larger, richer, and stronger by appropriating the resources and territories of an expanding empire. For these reasons, imperial cities have, at various periods throughout history, and especially in the nineteenth century, frequently become the largest, in terms of population and often the richest, as centers of consumption, of all cities in the world. Moreover, long after the empire itself has disappeared, in some if not every case, the accumulated resources, institutions, infrastructure, transportation networks, as well as cultural attributes of the earlier imperial regime have continued to exist in the one time imperial city. These may include banks, institutions of government and law, universities, museums, libraries, military installations, docks, shipping, road, and in the modern era, rail and airline transport routes, as well as the offices, symbols and practices of empire, including languages, attitudes, and beliefs. To different degrees, therefore, this might be true of the imperial regimes and cities of Rome, Istanbul, Peking, Cairo, Venice, London, Madrid, Lisbon, Amsterdam, Moscow, Vienna, among others. In this way, many one time imperial cities of the late nineteenth century have, at the close of the twentieth, mutated into what urbanists have, since the 1990s, increasingly called world or global cities.

Economically, imperial cities have been able to appropriate and assemble resources, whether grain, precious metals, fuel, labor, building materials, food, looted treasures, and science and knowledge from the far corners of the empire. Ancient empires have assembled slave labor to build massive pyramids, such as those at Luxor (Egypt) or Teotihuacan (Mexico); in modern ones, such as that of Britain, slavery was used to build up vast plantation economies in the Caribbean, generating capital for industrialization and urban development in the metropole. In imperial cities such as Delhi and Agra in the Moghul Empire (1526–1857) or Madrid and Seville in the empire of Spain (1492–1975), spectacular palaces, public works, and exotic gardens have been funded from the proceeds of imperial conquest. In the modern era, in the brief but singularly violent history of the Belgian Empire (1901–62), profits from rubber extraction in the Congo Free State under the rule of Leopold II, a period of gross exploitation and murderous atrocities, were used to build and equip the spectacular Musée du Congo in imperial Brussels (later, the Royal Museum for Central Africa) as well as other monuments. Under the Second French Colonial Empire (1830–1960), territorially the largest after the British Empire, and extending into north and west Africa, the Middle East, Southeast Asia, and Océana, Paris added to its existing imperial features, including Napoleon's tomb in the Invalides and the spectacular Arc de Triomphe celebrating his many campaigns. In the twentieth century, the resplendent tomb of colonial general, Herbert Lyautey, was located close to that of Napoleon. Not far away, an imposing bust of his army colleague, General Joseph Simon Gallieni, rests on the shoulders of four Asian and African women, natives of the various countries where he had served. Following postwar decolonization, the museum established for the 1931 international colonial exhibition in Paris was transformed into the Musée National des Arts d'Afrique et d'Océanie.

Both historically and in the contemporary world, ethnically and culturally different peoples have brought, or have traveled, to what are inherently multicultural cities. Whether defined by the cultural diversity of its peoples or the extent of its trade, the imperial city is, by default, the forerunner of the international city. Imperial trade, imperial wars, and the exploits of imperial armies not only add wealth, knowledge, and economic and political power to the city but generate much of its symbolic capital, the monumental arches, ceremonial ways, obelisks, commemorative parks, equestrian statues, gardens and vistas, war memorials, busts of imperial warriors, and thoroughfares named for faraway places. The imperial city is a site for impressive monuments, arenas, and stadiums to mount spectacular displays, whether exhibitions, sporting contests, or demonstrations of political power.

Imperial cities can be divided into two classes: those designed from scratch and purposely built to function as an imperial capital and others, usually national capitals, which have grown on the basis of centralized regal power and then acquire imperial features as the empire expands over time. Most of the large European capitals fit into the second type. India, while providing historical examples of both, also offers, in the case of New Delhi, a uniquely modern version of the first type. In the early twentieth century, the imperial government, anxious to escape the growing nationalist movement in Calcutta, moved the capital of the British Raj to a less volatile, more geographically central location. The relocated capital was planned on spectacular *beaux arts* principles: vast processional thoroughfares, endless vistas, the viceroy's monumental house of gigantic proportions, the vast iconic statue of the King Emperor sheltered by a soaring canopy, and the residential areas for the imperial bureaucracy laid out according to a rigid social and spatial hierarchy.

Of all ancient imperial cities, however, it is Rome and the cities founded by the Roman Empire that best illustrate the structural characteristics and political processes which distinguish imperial from other types of cities.

Imperial Rome and Imperial Cities

As an imperial city and the center of a vast multiethnic, multicultural empire spreading out from the Mediterranean shores into Northern Europe, the Middle East, and North Africa, Rome and the Roman Empire provide an example of the centrifugal and centripetal movements that influence imperial capitals and the principal cities in its colonies (or provinces). As the empire expanded between 27 BC and AD 330, new cities were founded from which the new provinces were governed and administered. Though influenced by local geographical and cultural conditions, the cities of the empire, by their location on land and seaborne trading routes, were to become part of a new urban network. They carried the Latin language and culture of the metropole to the colony, but also brought back local knowledge and culture from the periphery to the metropolitan core. Carefully planned, Roman Empire cities were generally walled, for both defense and administrative purposes, and in most cases laid out on a grid, a spatial device for colonizing new lands and peoples that was to have immense significance, not only for the Romans but also for later empires, such as that of the Spanish in South and Central America. As Roman rule expanded through Italy, the Eastern Mediterranean, and into Gaul and Spain, cities were established in which the conquering armies were garrisoned and from which Roman rule was extended over the surrounding province. Similar institutions meant that similar spaces and built forms were found in different Roman cities. In Pompei or Trier, where the grid plan dates from the first century, the forum or market place was the center of public activity. Other public buildings included baths, a gymnasium, temples dedicated to different deities, occasionally a circus or race track, and a basilica or audience hall. An amphitheatre provided the setting for gladiatorial displays. As with the massive Porta Nigra at Trier, imposing gateways often marked the entrance.

Founded after the invasion of Britain in CE 43, London was located at an important river crossing. In addition to a forum, basilica, and temples, London, as the capital of a new province, was also the site of a governor's palace. Ephesus, in what is now the southeast corner of Turkey, was the main city of the eastern Aegean and capital of the province of Asia. A busy port, exporting goods back to the imperial capital, in the second century CE it had extensive cultural facilities, including a magnificent library, built to commemorate an earlier governor of the province, Tiberius Julius Celsus. Fronted with huge marble columns and statues dedicated to wisdom, knowledge, intelligence, and virtue, the library had facilities to store up to 12 000 scrolls. In addition, there was a girls' gymnasium, theater, baths, stadium, and temples to the Goddess Artemis, a destination point for pilgrims.

Rome itself was built up especially in the first two centuries CE. At the end of the first, it had a population of some one million people, and under Augustus, emperor for 40 years (BCE 27–CE 14), it became the greatest city in the Western world. Augustus drew on the extensive new wealth coming into Italy from the colonies and spent immense sums on the city. Grain was shipped to Rome from its province of Egypt; olive oil for lamps, cooking, and food came from the Roman settlement in Carthage, North Africa. With his friends and lieutenants, Augustus added some of the city's most spectacular buildings, such as the original Pantheon, the Baths of Agrippa, and the Horologium, an enormous sundial, with an obelisk, imported from Egypt, used as a pointer. The Rome of Augustus was a Golden Age for Latin and the legacy was measured not only in buildings but also in literature and architectural treatises including, during the first century, the *Ten Books on Architecture* by Vitruvius, to become of immense significance for architects in the European Renaissance. The Colosseum, the largest amphitheatre in the Roman world, was dedicated in CE 80. Emperor Trajan (98–117), extending the empire's boundaries with his victories in Dacia (present day Romania) and later conquests in Mesopotamia, returned to Rome with cohorts of slaves and huge spoils of war. His deeds were commemorated in the huge column named after him, erected in a prominent place in the Roman Forum. On this, an intricately sculpted frieze provides a narrative account of his exploits.

Along with the grid plan, the Romans brought peace, trade and governance, street layouts, classical temples, municipal baths, huge aqueducts which brought water into cities, as at Nimes in France and Segovia in Spain. The artificial harbor at Ostia, the port of Rome, was critical in extending trade and shipping facilities. In all the provinces, newly built roads helped to link the provinces together. In CE 122, Emperor Hadrian (117–38) marked the northern limits of the empire in England with the wall named after him. Everywhere, victories abroad were signified by triumphal arches or commemorative columns.

The significance of the Roman Empire for the development of imperial cities is not limited to the founding of many cities, nor to the archaeological ruins which its decline left behind. With interest in the architecture of the ancient world revived by the Renaissance, the Roman Empire was to bequeath to modern Europe and its colonies, including North America, innumerable models of architecture and urban design. In terms of building typologies, architectural styles, or spatial forms, Roman models (incorporating the Greek) were to be spread by European colonialism all over the world.

Imperial London, Empire Cities

In the modern era, the reciprocal nature of the growth of the imperial city and its overseas empire is well illustrated by the symbiotic development of London and the colonial port cities of the British Empire. Beginning with the conquest of Ireland and the earliest settlements in North America in the late sixteenth century, the subsequent expansion of British military, naval, and trading power in the Caribbean, North America, South Asia, Australasia, and Africa, was to have a major impact on the physical growth and economic prosperity of London. As well as the violence, profits from the trade in slavery, and the plantation economies of sugar, tobacco, and indigo in the Caribbean and the American South were behind the setting up of the infrastructure for an overseas trading economy in the late seventeenth century. Merchants in this overseas trade helped in the founding of Lloyd's insurance (1687–88), a stock exchange of sorts (1670s), and the Bank of England (1691), boosting an already thriving European economy. Trade with the East was stimulated by the East India Company (founded in 1600), whose impressive new headquarters were opened near the Bank of England two centuries later. By the mid eighteenth century, colonial conquests in India, North America, and the Caribbean and a boom in the Atlantic economy had made London the largest center of international trade and its merchants the most prosperous in Europe. In terms of population, for over a century and a half, between 1750 and the height of British imperial power in 1910, London remained by far the largest city in the world, the population having grown from almost 1 million to 7 million during the nineteenth century. Reflecting the territorial expansion of this seaborne empire was a completely new system of docks constructed between 1799 and 1828, and their names (West India, East India, and London) linking their distant and local destinations. Collectively, they could hold over 1400 merchant vessels. As the trading infrastructure grew, it greatly increased the importance of the city economy so that, by 1815, London was already a world financial center. During the nineteenth century, the City of London, boosted by Britain's lead in the Industrial Revolution, further enhanced its commercial and financial resources and moved increasingly into overseas investment, in both the British colonial empire and the informal colonies of Latin America. The centrifugal flow of capital was accompanied by the centrifugal flow of people. From 1815 to 1925, over 25 million emigrated from the British Isles, 10 million to British possessions round the world, especially Canada, Australasia, and South Africa, taking institutions and cultural attributes with them, and also retaining economic and social links with people at home. Back in the imperial city, their places in the labor market were, in some cases, filled by the inflow of labor from the oldest colony, Ireland. Even in the early years of the twentieth century, there was a small cosmopolitan population in London, not only from Europe but also from the Middle East, India, China, and Africa.

London, along with other major ports in Britain (Glasgow, Liverpool, and Bristol), was part of a much larger imperial trading system that linked all the major port cities of the empire together. While some of these had developed from indigenous settlements existing prior to colonial contact, many developed especially in the nineteenth and twentieth centuries to become the major cities of the empire. In India, Calcutta was long known as the 'second city of the empire'. With Bombay and Madras, it had been built up and developed from the seventeenth century. Elsewhere, other cities were part of the network – in Southeast Asia, Singapore and Hong Kong; in South Africa, Cape Town, Durban, and East London; in Jamaica, Kingston; in Australia, Sydney, Melbourne, and Perth; and in Canada, Montreal and Halifax. These cities were not only the sites of trade which, in the nineteenth century, in most cases consisted in raw materials shipped to the metropole for processing, in return for industrial products sent back to the colony, but also, as in the Roman Empire, they were the places where new hybrid cultures developed, and as new industrially made products were introduced by the imperial power, indigenous craft production was forced into decline.

Colonial port cities, as well as other inlands also hosted the institutions and buildings of government, administration, education, technology, and the imperial architecture in which they were built. From the mid nineteenth century in India, much of the colonial administration and engineering works were housed by the public works department (PWD), a vast administrative and technical bureaucracy, functioning under the supervision of British imperial engineers. PWD personnel established a network of civil and military stations all over the country, along with barracks, institutional buildings, offices, canals, roads, and railways, linked to the major cities, and enabling the imperial government to maintain control over the country. During the frequent famines in India, brought on by climate failure and the government's policy of replacing subsistence farming by cash crops, such public works were used to help alleviate poverty, but this was never enough to prevent millions of deaths from famine.

As with the Roman Empire centuries before, new ideas about town and city planning, developed in the metropole in response to the urban chaos of industrialization, were exported to the empire's rapidly growing cities in the early twentieth century. In many European colonial cities in Asia, the Middle East, and Africa, ideologies about health, race, and culture had resulted in the founding of 'dual cities', the native town and the European settlement. The new 'science' of town planning was applied principally to the latter. Noting the rapidly growing cities and consumer markets, metropolitan corporate as well as indigenous interests were quick to establish shops, businesses, and factories. Along with the capital and businesses went the urban typologies of the metropole, often drawing for their architecture on Roman models, and also generating hybrid forms. By the twentieth century, the built and spatial environments of these empire cities often had more in common with each other than each had with the culturally and economically different vernacular environments where they existed.

Imperial expansion outside the metropolitan center generated spatial expansion inside it. In the later years of the nineteenth century, Britain extended its empire by a third, with new colonies in Southeast Asia; Southern, Central, and Western Africa; South America; and South Asia and Hong Kong. In London, new buildings for the Foreign Office, Colonial Office, and India Office were established in Whitehall. The Imperial Institute of the United Kingdom, the Colonies and India was set up in Kensington in the 1880s in order to exhibit the Empire's products and raw materials, to disseminate information on emigration, industries, and trades, and to promote further colonization and commercial and technical education in the interests of the Empire. The Imperial College of Science and Technology was established in the early years of the twentieth century. In other parts of the imperial city, colonial agencies provided information to the 8.5 million British people who migrated to British imperial territories in Australia, Canada, New Zealand, South Africa, and the Americas between 1881 and 1910. The growing consciousness of empire in London at this time combined with an increasingly competitive awareness of the new urban improvements being constructed in other European imperial cities such as Paris, Vienna, Seville, and Rome. It became apparent that London lacked a specifically imperial space, providing an imposing vista and site for celebration. Following the death of the Queen Empress Victoria in 1901, a plan for Imperial London was proposed, leading to the building of the Victoria Memorial, fronting Buckingham Palace, "a great imperial and national ideal wrought in marble," according to the *Illustrated London News*. The mall was remodeled as a processional way with a wide vista focused on the new Admiralty Arch in Whitehall.

The expansion of empire involved new peoples, territories, languages, and cultures as well as health hazards, presenting challenges to existing states of knowledge. To meet these challenges, innovative schools and institutions were established at the University of London. Scientific and medical needs were accommodated in the London School of Hygiene and Tropical Medicine, set up in 1899. The School of Oriental Studies, established in 1916, trained British administrators across the empire, offering language instruction and cultural knowledge of older and newer colonies; the School's motto, "Knowledge is power." Africa was later added to its remit. After 1918, the economic significance of the empire increased. In the nationally symbolic space of Trafalgar Square and the adjoining Strand, the most significant imperial territories were soon to be represented by imposing administrative and diplomatic premises: after Australia House (1914), came India House (1924), Canada House (1925), Africa House (1928), and South Africa House (1933). With the British Empire Exhibition taking place at Wembley in 1924, London was provided with its first major national sports stadium. The event also extended imperial toponymy to some of West London's prominent thoroughfares. These visible signs of empire did not of course stand alone but became part of a larger cosmopolitan totality.

Postimperial City

In the twenty first century, postimperial London is as present as was once the Imperial City. Over half a century after the demise of empire, citizens originating in the one time colonies and their descendants have transformed the ethnic composition of the city. Almost half of the country's ethnic minority population lives in London where it forms a substantial fifth of the city's population. While London's cosmopolitan and multicultural society consists of citizens originating from all over the world, speaking over 300 languages, the roots of most of them can be described, according to different perspectives, as either postcolonial or postimperial. What was once the infrastructure for maintaining an empire has become an infrastructure for maintaining, at least in part, a world economy of capitalism: the financial services industry, including banking, insurance, law, and real estate; international institutions and multinational corporate headquarters; cosmopolitan educational and cultural sectors: museums, universities, libraries, theaters, and research institutes. Like others of its kind, this is a postimperial global city.

See also: Urban Architecture; Urban Design; World/Global Cities.

Further Reading

Atkinson, D., Cosgrove, D. and Notaro, A. (1999). Empire in modern Rome: Shaping and remembering an imperial city, 1870 1911. In Driver, F. & Gilbert, D. (eds.) *Imperial Cities: Landscape, Display and Identity*, pp 40 63. Manchester: Manchester University Press.

Chase Dunn, C. K. (1985). The system of world cities, AD 800 1975. In Timberlake, M. (ed.) *Urbanization in the World Economy*, pp 269 292. New York: Academic Press.

Darwin, J. (2007). *After Tamerlane: The Rise and Fall of Global Empires*. London: Allen Lane.

Driver, F. and Gilbert, D. (eds.) (1999). Imperial cities: Overlapping territories, intertwined histories. In *Imperial Cities: Landscape, Display and Identity*, pp 1 20. Manchester: Manchester University Press.

Fuller, M. (2006). *Moderns Abroad: Architecture, Cities and Italian Imperialism*. London: Routledge.

Gillem, M. L. (2007). *America Town: Building the Outposts of Empire*. Minneapolis and London: University of Minnesota Press.

Hall, P. (1993). The changing role of capital cities: Six types of capital city. In Taylor, P., Lengelle, J. G. & Andrew, C. (eds.) *Capital Cities: International Perspectives*, pp 69 84. Ottawa, ON: Carleton University Press.

Hochschild, A. (1998). *King Leopold's Ghost: A Story of Greed, Terror and Heroism in Central Africa*. New York: Houghton Mifflin.

Hopkins, K. (ed.) (2001). *Globalization in World History*. London: Pimlico.

King, A. D. (1990). *Global Cities: Postimperialism and the Internationalization of London*. London: Routledge.

Metcalf, T. (1989). *An Imperial Vision: Indian Architecture and Britain's Raj*. Berkeley, CA: University of California Press.

Said, E. (1993). *Culture and Imperialism*. London: Vintage.

Scarr, C. (1995). *The Penguin Historical Atlas of Ancient Rome*. London: Penguin Books.

Scriver, P. (2007). Empire building and thinking in the Public Works Department of British India. In Scriver, P. & Prakash, V. (eds.) *Colonial Modernities: Building, Dwelling and Architecture in British India and Ceylon*. London: Routledge.

Skinner, G. W. (1977). *The City in Late Imperial China*. Stanford, CA: Stanford University Press.

Smith, T. (1999). 'A grand work of noble conception': The Victoria Memorial and imperial London. In Driver, F. & Gilbert, D. (eds.) *Imperial Cities: Landscape, Display and Identity*, pp 21 39. Manchester: Manchester University Press.

Steinhardt, N. R. S. (1999). *Chinese Imperial City Planning*. Honolulu: University of Hawai'i Press.

Wright, G. (1991). *The Politics of Design in French Colonial Urbanism*. Chicago, IL: University of Chicago Press.

Imperialism, Cultural

S. P. Mains, University of the West Indies-Mona, Kingston, Jamaica

Glossary

All-Inclusive Resort A tourist development (often including a large hotel on extensive grounds), which includes the (relatively) unlimited provision of meals, leisure facilities, and entertainment on-site, as well as taxes and service charges, as part of the total package price.

Cultural Landscape The combination of social and physical characteristics in a specific locale, which help to create a distinctive sense of place.

Discourse A system of representation and discussion through which language is given meaning.

Ethnocentrism A way of thinking and acting in which a specific ethnic group is privileged and believed to be superior over others.

Free Trade Zone A physical area designated for the location of foreign-owned companies who are usually provided with incentives such as tax exemptions, the provision of basic infrastructure, and the availability of inexpensive labor.

Global North Generally, high-income countries located in the Northern Hemisphere.

Global South Generally, low- and middle-income countries located in the Southern Hemisphere.

Remittances Money sent back by international migrants to their countries of origin – often as a regular (and significant) form of financial support to friends and family members.

Introduction

Cultural imperialism is a process of disproportionate influence over social practices and ideologies by one sociopolitical group (quite often a nation state), over a politically weaker and less wealthy group. This is a term that encapsulates a range of complex spatial and temporal relationships. These relationships reflect and influence differential access to power, space, and the ability to produce or challenge dominant ideas about people and places.

It is perhaps helpful to break down the separate cat egories that make up the term cultural imperialism in order to outline central characteristics of this concept (as opposed to the more general term of imperialism). Examining the latter term first helps to narrow the focus: imperialism refers to a process of domination of one social group over another (usually a state authority in fluencing another foreign power), and may be combined with efforts to establish colonial relationships, including territories. Imperialism may occur through the exertion of indirect and/or direct forces, and does not assume a written policy or formal control of foreign territories, such as those experienced as part of colonialism and the establishment of foreign annexes to colonial authorities. Imperialism involves a dominant influence by a particular authority through cultural, economic, and political practices, which facilitate the creation of hegemonic geographic powers. Although imperialism and colonial ism have increasingly been used interchangeably, im perialism can exist without colonialism – although some geographers refer to this newer form of inequality as neocolonialism – whereas all colonial projects have been infused with imperial ideologies. The territorial or ma terial product of imperialism and colonialism has been the establishment of empires and colonies, and more recently these spatial manifestations have been extended to include a broad range of places such as the creation of foreign run free trade zones, sites of military conflict and state oppression (e.g., state boundaries, prisons, schools, refugee camps), and sites of anti imperialism resistance (e.g., street protests, documentary films, community centers).

Geographic research exploring imperialism has in cluded examinations of colonial explorations and the 'age of discovery', globalization, dependency, and economic inequalities, the disproportionate presence of Global North commodities and media representations in Global South states; environmental conflicts linked to Western oriented technological and educational policies; and architectural and planning influences in urban centers.

Studies in human geography examining understand ings of culture have also encompassed a very broad variety of topics. Culture itself is a term that refers to a broad area of social practices and identities and, as such, can be seen as very amorphous and slippery. Culture is everywhere, but attempts to define and list its key char acteristics become very difficult to produce! Despite these difficulties human geographers have explored culture as social practices and relations that are dynamic, and with the ability to travel and transform in different periods and contexts. Culture has been broadly described by human geographers as the contexts in which people exist in space, and the sense of place that these different set tings provide. Consisting of the amalgamation of several factors, culture has been viewed as including: language, arts, economic livelihoods, emotional ties to place, social attitudes, conflicts, religious practices, political milieu, built environments, educational systems, and physical landscapes, to name just a few. In order to understand cultural imperialism, therefore, the inclusion of a broad range of practices and contexts is crucial.

A combination of imperialism and culture provides a concept that integrates some of the processes introduced above: cultural imperialism is a specific form of broader imperialistic practices, and again, refers to the uneven influence of a specific dominant power (or powers) over the social, material, and economic context of another lo cale. This influence is usually regarded as a negative force because it refers to a process through which cultural identities and practices of a dominant social group or authority (such as a particularly powerful state or ethnic group) are reinforced by limiting and/or undermining the existence of other less powerful and politically influential groups (such as smaller and less economically wealthy nation states or ethnic groups). This process of domination is not necessarily accompanied by the use of force – such as military campaigns – but may be part of an indirect process of establishing hegemonic (or overarching) influ ences over a particular social context through implied pressures such as the threat of military intervention, the marginalization of certain cultural practices, the removal of foreign economic support, political sanctions, or wide spread negative media attention. In some contexts this can be referred to as 'soft imperialism', since it refers to a process of dominant outside influences that are not fo cused on the use of violence, but may involve more subtle ways of shaping a cultural landscape.

For example, the US and some EU states are viewed by certain social groups in small Caribbean states as having too large an influence over the types of lifestyle choices that are viewed as desirable by island residents (e.g., through overseas produced advertising and programming that is broadcast in mainstream Caribbean media), as limiting sustainable economic trade opportunities (through pricing, the use of outsourcing to areas of cheap labor, a reliance on foreign tourists as an income generator), and limited political representation in regional and inter national decision making processes (e.g., through limited inclusion of Caribbean based representatives in inter national organizations whose policies affect the region). Each of these concerns combined has resulted in the cri tique that smaller less wealthy states are involved in a very difficult struggle in order to compete culturally and eco nomically with outside influences that turn their territories into landscapes of consumption (particularly in relation to absorbing overseas ideas and goods) rather than production (e.g., nurturing and promoting indigenous ideas and pro jects). While some researchers have suggested that outside influences and the exchange of ideas are inevitable and often positive experiences (and, as such, would not be labeled as culturally imperialist), what earmarks the Caribbean example mentioned above is the unevenness of a process through which smaller island states have a more limited say in the process of exchange and often find themselves in a less favorable dependent relationship.

Through the use of specific case studies, the discussion below briefly examines in more detail some of the key areas through which human geographers have examined cultural imperialism, and the varied ways in which we can contextualize the concept further. The first section is historical in focus, and explores the relationship between cultural imperialism and colonialism, and their links to other ideological perspectives and built environments. This is followed by an examination of more recent eco nomic and social dimensions of cultural imperialism ex plored by human geographers. This second section discusses the examples of multinational corporations and consumption, tourism, and media representations of place.

Cultural Imperialism and Colonialism

Although sometimes used interchangeably, as noted in the previous section, cultural imperialism and colonialism are parallel components of interconnected systems of dom ination. Metropolitan centers of power – for example, elite groups based in places such as London, Paris, Washington, DC, and Madrid – sought to concretize their imperialist (or empire building) aspirations through colonialism and the establishment of colonial territories (or foreign an nexes), over which they aimed to have unchallenged control. The formal establishment of expanded territories through force and coercion, combined with the promotion of certain cultural, economic, and political values and practices in places that were considered to be unclaimed landscapes – that is, not yet under Western control – was an important historical process in terms of manifesting cultural imperialism through space, and in specific locales.

Examining the ways in which colonialist practices relied on the promotion of overgeneralized concepts of place and identity, the postcolonial theorist, Edward Said, illustrates that cultural imperialism was a key part of this process of domination of foreign lands and that it was shaped by a perspective he terms Orientalism. Orientalism reflects a Eurocentric viewpoint – that is, one that views Europeans and European viewpoints as superior to those of other peoples and places – and as such, relies on Western stereotypes of non Western cultures. Indeed, Said points out that despite their pervasiveness in academic discussions and popular culture, even the notion of Western/non Western, or West/East, is problematic. Despite their depictions as natural and neutral, these spatial (and ideological) divisions are subjective representations of the ways in which the world has been arbitrarily divided for academic, political, and economic purposes, usually serving the needs of wealthy nations such as the UK, France, Germany, and the US. In describing European depictions of an Orient, for example, Said notes that this latter region emerges in French and British nineteenth century writings as an exotic culture and region of intrigue. This depiction of a cultural landscape of mystery and curiosity was used to situate the idea of an Orient directly in relation to an Occident (including, for example, France and Britain), and was as much about how Europeans viewed European culture as it was about exploring an Oriental identity. This binary, over generalized, and uneven system of representation became a process – and way of discussing (or discourse) – through which the idea of European superiority was reinforced. The process and practice of cultural imperialism took place (and, many human geographers would argue, continues to do so especially in current debates about terrorism, migration, and conflict in the Middle East) through the construction of non European identities and places (e.g., Egypt or the Arab) that were narrowly bounded social, spatial, and physical stereotypes framed by a European viewpoint. Based on this perspective the Oriental was clever, but unsophisticated; passionate, but overemotional; archetypal and static. As such, this Orientalist discourse – which was intrinsically about intersections of space, representation, and identity – produced through novels, academic institutions, art, foreign policy, and military activities – had a significant impact (both discursively and materially) in the reinforcement and reproduction of European cultural and political activities in non Western contexts.

Cultural imperialism and colonialism have, therefore, been interwoven through a diverse range of spatial representations and built environments, including travel writing, art, religion, education, language, and architecture. Human geographers have pointed out that colonial accounts of overseas territories through exploratory journals or travel writing, which were later published in Europe, usually produced depictions of landscapes and cultures that were framed by a series of class, gender, and racial assumptions and relations, which set these 'newly discovered' places in direct comparison to those of the colonial home, and usually in a manner that reaffirmed the superiority of imperial centers of control. Artistic representations of colonies (as well as domestic rural settings), utilized approaches that idealized marginalized locations (e.g., through nostalgic and utopian representations of rural landscapes), and gendered imagery that depicted primitive and/or passive cultures that required taming, management (and exploitation). Religious, educational, and linguistic dimensions of cultural imperialism can be noted in the present day prevalence of European influenced practices in many locations throughout the world, for example, the prevalence of Catholicism and Protestantism in Latin America and the Anglophonic Caribbean, respectively; Eurocentric educational curricula; and the dominance of forcibly introduced European languages, such as Portuguese in Brazil, English and Afrikaans in South Africa, Dutch in Suriname, and French in Cameroon. Such practices of cultural domination, often accompanied with the outright banning of indigenous languages and religious practices, were undertaken as a means of combining and extending physical coercion and control with the use of social practices that permeated everyday life, and as such became more complicated to challenge. As is noted in the following case studies, the combination of these cultural impositions with changes to the built environment, has produced numerous cultural landscapes imbued with culturally imperialist ideologies.

Cultural Imperialism, Colonialism, and the Built Environment

World's Fairs exhibitions provide an additional case study of cultural imperialism in social and physical landscapes. Preceded by national exhibitions held in France, and initiated in 1851 as the Great Exhibition at The Crystal Palace in London's Hyde park, international World's Fairs provided a showcase of all that was considered to be innovative and productive in relation to the exhibitors' economic (and cultural) gains. Occurring on average every year to 4 years (continuing to the present day, and now every 5 years), and initially rotating between a range of largely European and US cities, the exhibitions provided the opportunity to bolster distinctive urban and national notions of identity, while selling an idea of internationalism that – particularly from the nineteenth to the mid twentieth century – was largely framed by specific European and US interests.

A key attraction of these early exhibitions was the pavilion displays staged by each participant country as part of a project to illustrate the industriousness as well as cultural wealth of Western nations and their empires. These displays – often housed in striking purpose built structures – included examples of inventions related to technology, art and design, and, combined with specialized amusement zones, were designed to entertain the large numbers of visitors who traveled to the exposition sites. As events that were linked to cultural imperialism, the World's Fairs were important because they reinforced and celebrated a notion of culture and commerce that was being imposed in overseas territories (and domestically in factories and schools) – often accompanied by a significant degree of force – in a manner that portrayed Western industrialization and capitalism (including a sanitized portrayal of slavery) as noble, progressive, and inherently egalitarian. For example, ethnocentrism and racism prevailed in anthropological exhibits where many display tableaus included the incorporation of people from respective regions as part of a so called naturalistic system of representation (e.g., while depicting an archetypal African village or Pacific Island dress). Although providing the opportunity for a broader population to virtually travel to distant lands, such displays did little to represent the viewpoints of those they purported to represent: Western perspectives of non Western cultures were viewed as more informed and objective than those provided by inhabitants of the various locales themselves. Colonial territories and a range of locales in the Global South became the objects of a white masculinist Eurocentric gaze that depicted intriguing, but eccentric cultural landscapes set in the past awaiting the catalyst of US or European intervention and innovation to bring them into the modern era. Although many technological developments, including health, transportation, and sanitary innovations, may have been appreciated by the contemporary populace, those involved in anti colonial movements and labor organizations fighting for fair wages, improved working conditions, access to housing, and greater political representation – both in exhibition host countries and colonial territories – disagreed with the triumphalist narratives that were promoted by those official, and usually socially elite, groups administering international Fairs. These challenges, however, were usually marginalized in popular discussions about exhibitions (e.g., in print news media), and caricatures of anti imperialist or socialist activists as extremists or as anti progress were not uncommon.

The impacts of cultural imperialism in the built environment can be extended beyond the World's Fairs to the everyday landscapes of towns and cities globally. In an effort to create satellite versions of London, Paris, and Berlin, during the nineteenth and early twentieth centuries, building designs – for example, those inspired by Victorian or Georgian architecture – were copied and transplanted to railway stations (e.g., in central rail stations such as, the Chhatrapati Shivaji Terminus, formerly the Victoria Terminus, in Mumbai, India), court houses, schools (e.g., Jamaica College in Kingston, Jamaica), and public monuments (e.g., sculptures dedicated to European explorers and military officers). This practice of redesigning and incorporating an imperially inspired aesthetic was another process through which European designs and ideologies were reified by colonialists, while indigenous architecture – with a few exceptions – was frequently viewed as quaint, but inferior and dispensable.

Similar processes of architectural and political privileging (and discrimination) can be seen in earlier periods of imperialist practices, such as those during the Aztec and Spanish settlement and domination of areas within Mexico (and particularly the latter, in various regions of Latin America). Mexico City's main square, or *zócalo* – *Plaza de la Constitución* (Constitution Plaza) – houses perhaps some of the most, and diverse examples of efforts to manifest imperialism physically and symbolically in an urban setting over many centuries. Currently one of the largest main squares in the world, this central location within Mexico's capital emerged during the fourteenth century as the physical and ceremonial center of the indigenous Aztec empire. The city that grew in this site, Tlatelolco – which later became part of the larger Tenochtitlán – was then the site of violence and domination by Spanish colonialists in the sixteenth century. During the seventeenth century, the latter group built the Church of Santiago on top of ceremonial structures constructed by the Aztecs as part of the Templo Mayor and by doing so symbolically and physically reinforced a process of undermining and disempowering of the few remaining indigenous representations in the landscape.

Today the *Plaza de las Tres Culturas* (Plaza of Three Cultures), which partially overlies where the center of Tlatelolco was based, has been officially assigned importance as a site commemorating three key cultures in Mexico City's history: Aztec (or pre Colombian), Spanish (colonial), and Mestizo (a mixture of Spanish and indigenous cultures) – the latter being described as symbolizing the populace of the modern post colonial Mexico. The plaza lies in the vicinity of the excavated ruins of the Templo Mayor (see **Figure 1**) – now part of a museum exploring indigenous cultures and as such providing a resurfacing counter (Spanish) imperialist landscape (with Aztec inspired dancers also attracting many tourists in the plaza area). The seventeenth century Church of Santiago de Tlatelolco, the Ministry of Foreign Affairs, and modern apartments also bound the plaza – depicting shifts in dominant national and international identities over many centuries. Significantly, the Plaza of the Three Cultures has also become an important memorial site for those who were killed during a 1968 massacre, during which government authorities violently halted anti government student protests just prior to the Olympic Games that were held in the city. Following the public availability of previously confidential documents, US intelligence agencies have since been implicated in the dramatic actions taken by the Mexican authorities during 1968. This foreign intervention in a domestic political context could also be viewed as illustrating the ways in which dominant political values, and events that challenge them, from within and beyond nation state boundaries emerge in both discursive and material form.

Mexican painter, Diego Rivera, who was particularly renowned for his large scale murals (or frescoes) – and married to another important painter, Frida Kahlo, who also contributed significantly to a post independence artistic identity – has provided a counter imperialist presence in the nearby landscape, having created several murals depicting pivotal moments in Mexican history on the walls of the key legislative building, the National Palace (facing onto Constitution Plaza). Painted from the

Figure 1 Excavated ruins of the Aztec ceremonial center, the Templo Mayor, in Mexico City.

1920s to the 1950s, Rivera's murals critique the exploitative nature of capitalism and corruption, and attempt to reclaim the importance of indigenous histories and the necessity of an active labor movement. Carrying on this tradition the main *zócalo* itself has been the focal point of numerous protests, rallies, and concerts in efforts to reclaim a social and physical space for large sections of a populace that have often felt marginalized. Human geographers argue that such socio spatial narratives found in places like Mexico City (and many settings throughout the world) illustrate how cultural imperialism can be manifested in various aspects of urban landscapes (from bodies, to plazas, to policies and protests), and that ideologies can (literally) be cemented into the built environment.

Contemporary Cultural Imperialism and Representation: Consumption, Tourism, and Media

During the last 50 years, cultural imperialism has impacted human geographies at the local, national, and international scales. Although now less likely to occur as part of a formal process of domination, such as colonialism, it is viewed as an important means of indirect influence and inequitable control in the social and economic practices shaping places. The economic dimensions of cultural imperialism have attracted a growing body of research in human geography, with particular attention being focused on processes of consumption.

Consumption refers to the acquisition and utilization of goods, ideas, and services. Geographers highlight that practices involving the consumption of products, and even lifestyles, are often reflective of relationships that are politically and economically uneven. This means, for example, that particularly wealthy and publicly visible nation states, private corporations, social groups, or individuals may be able to encourage the selling and purchasing of particular goods (e.g., clothes, cars, or music) or ways of living (e.g., a particular kind of neighborhood, health practice, or voting preference) through the promotion of advertising imagery and public debate that privileges certain ideals over others, and through which material objects become symbolic of broader cultural identities and social values. When the consumption of identifiable goods or trademarks is undertaken in an obvious manner – that is, the consumer is actively making an explicit statement about what kind of material objects and lifestyle practices they prefer – this is referred to as conspicuous consumption. Human geographers, and other cultural theorists and social scientists, point out that problems emerge when the objects and practices that are promoted and desired are insensitive to the cultural context to which they are introduced and – particularly in economically challenged areas – that they may symbolize a tension and inaccessibility between more widespread or indigenous goods and practices, and those that are viewed as exclusive or fashionably foreign. It has been argued, for example, that a shift in diets toward imported and/or fast foods as well as a push for individual rather than collective wealth in countries of the Global South has reduced nutrient diversity in daily food consumption, undermined local food producers, negatively impacted environmental conditions due to large scale agro busi nesses and waste production, increased the homogen ization of restaurant design and the expansion of low paid jobs (with limited skill transfers), and encouraged an emphasis on private automobile transportation, while also undermining cooperative and locally based com munity driven business initiatives (especially if viewed as direct competition). This shift in economic and social practices could be viewed as a demonstration of cultural imperialism by corporations based in the Global North because of the disproportionate economic power and advertising presence that they exhibit, both in domestic markets and in the Global South. Similar criticisms have been made of large retail corporations, such as the US based company Wal Mart, which has expanded inter nationally in recent years and has been described as undercutting the prices of locally owned businesses while benefiting from the production of goods in over seas locations that utilize cheap labor in places such as free trade zones. **Figure 2** illustrates a satirical – and counter hegemonic – comment on the ubiquitous (and almost inevitable) nature of Wal Mart developments. The handmade structure shown in the image is a small Rastafarian business along a main tourist routeway on the Jamaican North Coast; many members of this aforementioned religious community have been critical of the substantial influence of US economic and political practices in the island.

Tourism provides another case study of possible cul tural imperialism – again in relation to inequalities in decision making, economics, and design. The inequalities found in relation to tourist policies and developments, and the dependent nature of the relationship between certain tourist destinations and the investors and tourists they rely on, have led to this practice being described as neo colonial. For example, small island developing states in the Caribbean (also known as SIDS) rely heavily on international tourism and remittances as steady sources of foreign exchange income. Despite – and partly because of – the significance of tourism as an economic con tributor to the region, Caribbean residents and non governmental organizations (as well as geographers) have stated that island residents and governments have been pushed to sell a form of tourism that caters more to the needs and expectations of overseas tourists and investors than to the desires of local residents. In some Caribbean contexts, this has resulted in the commodification of stereotyped ideas of island life, based on non island perceptions, and has often been far removed from the daily, lived experiences of the resident population. Add itionally, large scale tourist developments (such as all inclusive resorts) have been criticized for exacerbating environmental problems along coastlines, promoting lifestyles that may be insensitive to local cultural values, offering limited opportunities for interactions between visitors and locals outside of resort spaces, and en couraging a culture of dependency shadowing that of the former colonial era, in which predominantly black

Figure 2 Resisting cultural imperialism: A locally made Rastafarian structure on the Jamaican North Coast where much of the recent large-scale tourist development has occurred critiquing the ubiquity of large-scale corporations, such as the US-owned business, Wal-Mart.

islanders continue to service the needs of a relatively wealthy, and largely white foreign population.

The role of mainstream media is central to the ex amples of consumption and tourism discussed above, and during the last 30 years has been attracting increasing attention by geographers examining the various forums through which dominant images of place and identity have been (re)produced and challenged. Media formats such as television, radio, film, print newspapers, and online publications can have a considerable influence in the reinforcement of ideologies and the ways in which these become (socially and physically) mapped onto people and places. Geographers suggest that in order to understand media images of places and cultures we have to interrogate who has the power to produce represen tations – or the power to tell stories – and in what way these images can reinforce dominant and/or exclusionary discourses (such as anti immigrant sentiments, racist stereotypes, place misrepresentation, nationalism, and sexist imagery). At the same time, media representations can provide challenges to *status quo* narratives and visual imagery, and thus a growing body of literature has been exploring the use of audio–visual materials and elec tronic formats as a means of producing anticulturally imperialist texts (e.g., through activist based media and independent film).

The shifting and diversifying roles of documentary film provides an example of the way in which the same medium can have varied purposes. During the first half of the twentieth century, documentary films played an im portant pedagogical (and paternalistic) role in informing European audiences about colonial territories and in attempting to socialize newly arrived immigrant popu lations in their destination country (e.g., as can be seen in films produced by the British Colonial Film Unit). In the pre and post World War II period, while still providing a format for traditional images of popular culture, docu mentary (and fictional drama) film increasingly became a forum for providing critically engaged cinema that sought to challenge imperialist and socially conservative representations while attempting to capture the diverse experiences of a broader range of the populace in a variety of locations (e.g., as can be noted in twentieth century British social realist films). Therefore, even within a setting that can be viewed as part of an imperial center, challenges to imperialist representations were present.

More recent – and ongoing – developments in digital video technology and internet based media pose signifi cant areas for future human geography research and explorations into the interrelationships between systems of representation, power, culture, and identity. For, al though cultural imperialism may still exist in a range of media representations, there is always the possibility that those same media may provide a range of alternative visions: hegemony can never be completely sealed up, thus, cultural imperialism remains a sailable, but leaky boat.

See also: Colonialism I; Colonialism II; Dependency; First World; Globalization, Cultural; Hegemony; Imperialistic Geographies; Neocolonialism; Orientalism; Postcolonialism/Postcolonial Geographies; Poststructuralism/Poststructuralist Geographies; Third World.

Further Reading

Blaut, J. (1993). *The Colonizer's Model of the World: Geographical Diffusion and Eurocentric History*. New York: Guilford Press.

Blunt, A. (1994). *Travel, Gender, and Imperialism: Mary Kingsley and West Africa*. New York: The Guilford Press.

Escobar, A. (1995). *Encountering Development: The Making and Unmaking of the Third World*. Princeton, NJ: Princeton University Press.

Gilmartin, M. (2004). Language, education and the new South Africa. *Tijdschrift voor Economische en Sociale Geografie* 95(4), 405 418.

Gregory, D. (2004). *The Colonial Present: Afghanistan, Palestine, and Iraq*. Malden, MA: Blackwell.

Harvey, D. (2003). *The New Imperialism*. Oxford: Oxford University Press.

Jacobs, J. M. (1996). *Edge of Empire: Postcolonialism and the City.* New York: Routledge.

Kincaid, A. (2006). *Postcolonial Dublin: Imperial Legacies and the Built Environment*. Minneapolis, MN: University of Minnesota Press.

Klein, N. (2000). *No Logo*. New York: Picador.

Klein, N. (2007). *The Shock Doctrine: The Rise of Disaster Capitalism*. New York: Metropolitan Books.

Mains, S. (2004). Teaching transnationalism in the Caribbean: Toward an understanding of representation and neo colonialism in human geography. *Journal of Geography in Higher Education* 28(2), 315 330.

Said, E. W. (1978). *Orientalism*. New York: Pantheon Books.

Said, E. W. (1993). *Culture and Imperialism*. New York: Knopf.

Staszak, J. F. (2004). Primitivism and the other: History of art and cultural geography. *GeoJournal* 60(4), 353 364.

Thiong'o, N. W. (1986). *Decolonising the Mind: The Politics of Language in African Literature*. London: James Currey.

Relevant Websites

http://www.bfi.org.uk
British Film Institute: The British Film Institute website provides a range of links to educational and archival material that is preserved and/or held by this national institution dedicated to maintaining the diversity of moving image media available in the UK. Selected sections of the archive can be accessed via the website (particularly through the Screen Online section), and further information regarding accessing archive materials and public screenings is provided.

http://expomuseum.com
ExpoMuseum: ExpoMuseum is a web based museum related to past and upcoming World's Fairs (or World Expositions). This site has a range of historical information and contains links to other sites, such as museum archives, which contain detailed pictorial and textual materials related to the Fairs.

http://www.indymedia.org
Independent Media Center: The Independent Media Center or IndyMedia Center as it is popularly referred to is a website based collective of activists, writers, journalists, and observers who aim to

provide alternative sources of information to mainstream news outlets. The website is available in several languages and provides links to current news items, articles, upcoming events, and related media websites.

http://econgeog.misc.hit u.ac.jp

International Critical Geography Group: The International Critical Geography Group is a network of scholars and activists seeking to actively address issues of inequality and poverty through geographic research and organizing. The website provides information on the goals of the Group, upcoming conferences, and related Internet mailing lists.

http://www.peoplesgeography.org

People's Geography Network: The People's Geography Network is a network of geographers, interdisciplinary researchers, and community activists who seek to broaden the scope of human geography and to increase its accessibility to a cross section of the general public. The website contains materials that can be accessed online, including seminar papers, working papers, upcoming events, and links to related websites.

http://www.forumsocialmundial.org

World Social Forum: The World Social Forum (WSF) is a multilingual website that acts as main clearinghouse and source of information in relation to the activities of community organizations and activists working as part of the WSF. It provides information about upcoming regional and international meetings, organization committees, proposals for challenging indebtedness and the existing process of payments for countries in the Global South, and related news links.

Imperialistic Geographies

G. M. Winder, The University of Auckland, Auckland, New Zealand

Glossary

Border Region Borders are sites in which difference is produced and disrupted. The region on either side of the border is broadly constituted by the presence and contestation of a border, whether between or within states. Border construction, border crossing, and the identities that are formed and dissolved in such acts, seldom make for simple geographies of identity, belonging, or state control. Instead, we might think of border regions (or borderlands) as generating hybrid identities, and weakened or partial state effectiveness.

Colonialism The establishment and maintenance of rule by a sovereign power over subordinate, alien people. The term implies a policy of establishing and maintaining colonies through a series of exploitative practices and relationships, especially with 'backward' or 'colonial' peoples. Such policies may be backed by racist ideologies. The term also implies a subordinate territory geographically separate from the 'home' territory of the ruling power.

Contact Zones These are the sites for contact between peoples. Such zones may be identified at varying scales from border regions to specific locales. Their principal feature is the form of the contact between peoples. The term directs attention to the complex interactions between cultures including the development of hybrid identities and practices, efforts at translation and intermediation, as well as patterns of domination, subordination, and resistance. Postcolonial scholarship points to the ambiguities, mistranslations, hybridities, tensions, and conflicts that manifest in the contact zones of empires.

Cultural Imperialism The process whereby ideas, beliefs, and practices of certain groups come to dominate and thus subordinate the norms of others. It is commonly associated with colonialism and the domination of indigenous peoples by settlers and colonizers with different cultural backgrounds. It is also associated with the colonization of cultures by cultural products emanating from global media.

Domain This term is used by Donald Meinig to refer to the core territory of an empire. Such a territory will benefit from its relations with various (colonial) peripheries and will have a coherent landscape, culture, politics, and economy, most likely characterized by a relatively high density and with the core functions of the empire and its economy. Thus, for much of the period 1870–1970, Meinig identifies the US Manufacturing Belt as the US 'domain'. Curiously, in this case, Canadian geographers regard the manufacturing complexes in southern Ontario and Quebec as extensions of the US Manufacturing Belt, a finding that confirms the economic domination of the North American space economy by US interests.

Enclave A territory lying within a state that does not fall within its jurisdiction. Free cities, and free ports with special treaty rights and tariff regulations, are good examples. The term is also used to name complex localities within a city that have a separate identity and an identifiable zone of activities, such as a ghetto.

Formal and Informal Empire Politically empires expand through various means so that only some territories and peoples are directly governed by imperial authorities as parts of a formal empire. Other territories and peoples may be colonized through various forms of informal empire, including, for example, treaties that establish special rights and limit sovereign power, treaties offering protection and the secondment of military forces giving *de facto* military occupation, declarations of spheres of influence, or direct intervention in the administration of state finances, notably after states get into financial difficulties and are unable to pay back loans. Note that formal empire status need not entail direct rule in all matters from the metropolitan capital. This could be compromised by the election of a settler government or by recognition of a local government or traditional authorities.

Hegemony The domination of one group over another through ideological rather than military means. Here the power to control arises out of the acquiescence of citizens when they accept and reinforce their subordinate status through taken-for-granted practices of everyday life.

Imperialism This term refers to the power relations involved when one group of people occupies the territory of another, subjugating its people. Imperialism implies the existence and perpetuation of an empire and its colonies. It also implies a culture and ideology of empire that transcends the borders of any given territory within the empire. As a term 'imperialism' arose out of Marxist critiques of the economic relationships between corporations, states, and trade and out of Marxist theorization of the reasons for clashes between empires under capitalism.

Internal Colony Since colonialism usually implies a subordinate territory geographically separate from the 'home' territory of the ruling power, the idea of an internal colony extends the sphere over which the practices of colonialism can be exercised to include territories and peoples that are somehow internal to the nation-state or the core territory of the imperial power. Thus, by the late nineteenth century Ireland could be regarded as an internal colony of England, lying as it did within the core territories of the United Kingdom. Similarly, Indian reservations remain internal colonies of the United States.

Locale A setting or context for social interaction. Organizations and institutions commonly create and occupy defined locales of operation: the office for bureaucracies; the classroom or lecture theater for educators; the football field for football matches; the shop for retailers; the factory work room for manufacturers; the house as residence. Such physical settings for activities are usually carefully designed to facilitate typical interactions and thus to make irregular behavior identifiable and to encourage self-policing. Locales may, for example, be highly gendered, as with specific work rooms within the factory. When situated in close proximity, several distinct locales may constitute a locality – such as the office district, school, stadium, retail district, factory, or suburb – which may or may not comprise a neat and orderly ensemble of related activities.

Neocolonialism New forms of colonialism in which a formally independent state is nevertheless subject to domination and exploitation by a foreign nation. Unequal terms of trade, indebtedness, forms of administrative or financial control, reliance on outside experts, and domination of the economy by transnational corporations all play roles in neocolonialism. Key institutions in neocolonial relations include the International Monetary Fund and the World Bank. Often neocolonialism builds upon legacies of former periods of colonialism.

Treaty Port Many ports, for example, Hamburg today, operate under special treaties that facilitate their operations by compromising national government control of the port or part of it. Often the port authorities operate under special tax, customs, or company registration laws that benefit foreign traders. The classic examples are the Chinese ports opened to foreign traders and officials through the ‘unequal treaties’ signed between aggressive foreign powers and China. Under these arrangements, foreign nationals were entitled to live and work in enclaves within the port city administered by local authorities appointed by foreign powers. There they were subject to local administration but were immune from Chinese law.

Introduction

Historical geographers are engaged in critical geographical enquiries into the bounds of empire. At the very heart of empire and imperialism lies the power to dominate people and territory, so the effort to map and remap the historical and geographical forms of imperialism and their effects remains central to any understanding of patterns of domination, terror, enslavement, and resistance, even in the early twenty first century. It is crucial to an understanding of the modes and spaces of liberation and encounter that are bound up with imperialisms. In recent decades the historical geography of imperialism and empire has often been an explicitly postcolonial endeavor but even in the past it has often been critical rather than celebratory or apologist. Historical geographers' accounts of colonial encounter develop understanding of empire and colonialism as complex lived geographical regimes, with colonial authority and the relations between colonizer and colonized often ambiguously muddled by diverse views about difference, superiority, belonging, and identity.

Establishing the bounds of empire is notoriously difficult. It is deceptive to think of any empire as coincident with an easily identifiable territorial unit. The territories that officially comprise an empire are diverse, ranging from metropolitan center and domain, iconic symbolic sites and military compounds, to all manner of peripheral entities: internal colonies, self governing dominions or protectorates, jointly administered territories, enclaves, chiefdoms, and principalities where unequal treaties secure special rights, to ships and aircraft, or the communications equipment atop rocky islets or orbiting satellite stations. Cultures of imperialism extend well beyond borders yet link identities and lives to imperial projects, whether in support, opposition, or some ambivalent accommodation. They do so through the cultural imperialism associated with trade, media, mission, and scientific exploration. Empire is buttressed by geographical imaginaries that fly far and wide. News reports and editorials build empire. Hymns, novels, and ceremony can supply its atmosphere and generate a framework for identification. Images of all kinds portray it, including imaginaries of others who are sometimes not even inside empire but are encountered through it, and imaginaries of the imperialists composed by those subjected to or outside their dominion.

Behind these discursive modes of colonization and promotion lurk brutal and material power relations. Crossing the bounds of empire can be lethal since they are stalked by terror and violence. The locales of empire are constituted not only by modes of domination, hegemony, enslavement, conquest, appropriation, and alienation, but also by cooperation and collaboration as people try to access power. Imperialists project power through violence and the threat of violence, often, but not always, sanctioned by the state. However, even military and police action, survey, regulations, and imprisonment can be opposed through dissidence, flight, reform, resistance, rebellion, and revolt. Empires wax and empires wane. There were premodern empires and non European empires and there have been develop ments in the technologies of empire. Empire did not begin with the expansion of Europe overseas but it is the 500 years of European expansion abroad that has been most closely related to historical geographic research. Officially, the 'age of European empires' ended in the mid twentieth century but imperialism continued on under the Cold War's *Pax Americana* and now under new names including globalization and neocolonialism. In all of these ways the bounds of empire are complex and difficult entities.

Approaches

Historical geographers have written about the geographies created by imperialism and colonialism by using many different research frames at many different scales. Writing of the eighteenth century Atlantic world, Miles Ogborn sketches three such geographical approaches: survey, net work, and trace. Where surveys offer synoptic overviews by drawing places together for comparative and connective purpose (**Figure 1**), work on networks emphasizes the changing webs of relations and material connections be tween people, places, and objects (**Figure 2**). Alternatively, by tracing individual journeys and life journeys historical geographers may capture not only moments of movement, but also the character of encounter and negotiation with their residual effects (**Box 1**). This latter approach em phasises cross cultural negotiation and hybridization, mis understanding, and transformation. Individual traces can be related to time geographies of the everyday experiences of people and their mobility, as well as to the narrative codes and scripts often associated with such personal journeys. Yet, other work focuses attention on contact zones in which the forces and structural properties of imperialism can be apprehended through the contact between people (**Fig ure 3**). Other work concentrates on the analysis of the texts and images used to construct imperialism, its imaginaries, and the species, race, gender, and class identities associated with it (**Figure 4**). Thus historical geographers analyze, among other texts, the geography books of Victorian high schools, the scientific publishing of geographical societies, travel writing, industrialists' advertisements and exhibitions, and commercial atlases to sharpen understanding of the geographical imaginaries, discourses, and identities of past empires.

The approaches outlined here may appear as distinct and incompatible, but historical geographers are aware that combinations and comparisons are crucial if their work is to amount to something more than a series of probes. In what follows, the spaces, texts, and technologies of empire that are foci for research have been surveyed briefly. Donald Meinig's critical survey of the historical geography of the

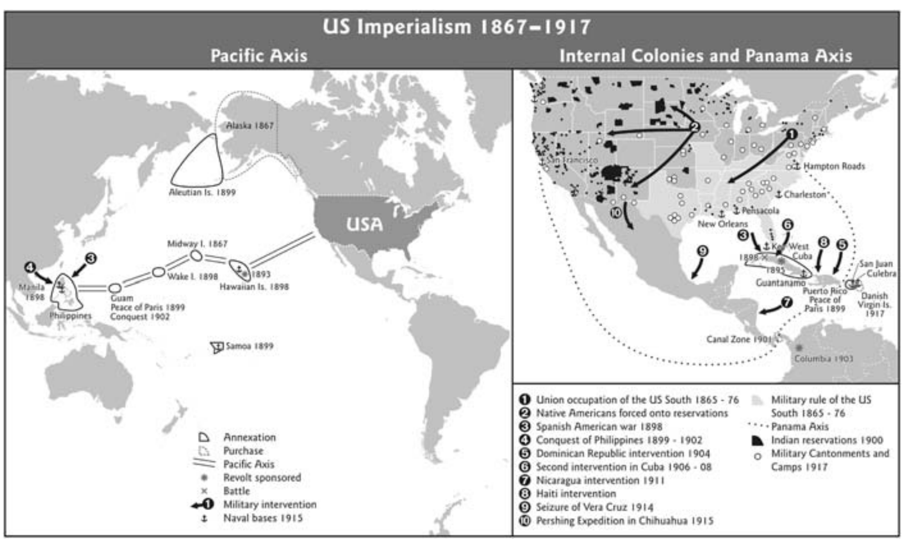

Figure 1 Survey: American Empire 1867 1919. The 'bounds' of the US Empire can be surveyed but only by using a variety of different classes of territories and nodes differentiated by the type of American power inscribed in them. This map surveys forms of political domination exercised from Washington DC from 1867 to 1919. It emphasizes the expansion of the US by military action and purchase from other imperial powers, as well as the policing of internal colonies. Arguably, a survey of New York's economic domination through the 'dollar imperialism' practiced before and after World War I would reveal a much wider and rather different geography of domination, including cultural domination of the world's film screens, direct financial administration of the Dominican Republic, and corporate domination of Columbia, the first 'banana republic', and of Canada, the first 'branch plant economy', through foreign direct investment. Source: Adapted from various figures in Meinig, D. (1988 2004). *The Shaping of America: A Geographical Perspective on 500 Years of History*. New Haven: Yale University Press.

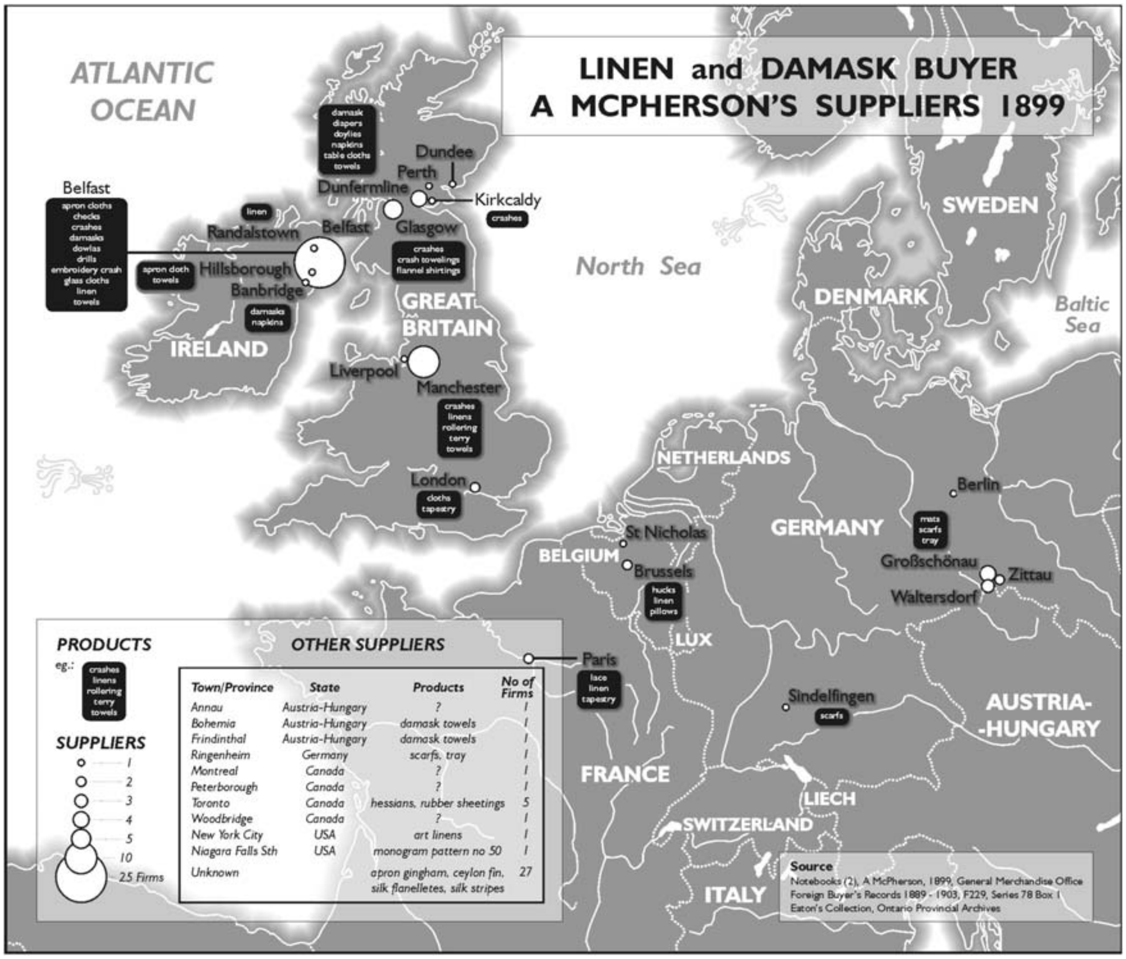

Figure 2 Network: Buyer's address book (1899). McPherson bought linen and damask products for the Toronto department store of T. Eaton and Company. His annual buying trips took him to suppliers in Europe and the US, as he developed an extensive network of suppliers; nevertheless, imperial ties dominated McPherson's commercial network. The patriarch of his firm was Scots-Irish and Presbyterian and gained his early training, experience, and capital in Antrim, Ireland. Eaton and McPherson sought Continental producers to supply damask (silk) products and other fancy goods, while sourcing toweling and linen from factories in Manchester, Lowland Scotland, and Belfast. UK suppliers had advantages in the trade because of 'Empire loyalism' in Canada, the scale of the market established through Empire, restrictions on the production and export of textiles in India, and the cultures of the commercial infrastructure of the imperial linen trade. Source: Eaton's Collection, Ontario Provincial Archives, Toronto, Canada; Photo Winder, G. (2006). Webs of enterprise 1850–1914: Applying a broad definition of FDI. *Annals of the Association of American Geographers* 96(4), 788–806.

United States has developed the survey as an approach to imperialism. Other historical geographers have sought to provincialize Europe, to understand the experiences of those who have been colonized, to understand the historical economies of imperialism, and to understand the spaces of imperial heritage in terms of memory and amnesia. Historical geographers research many spaces of empire and in challenging and critical ways.

Surveying the Spaces of American Imperialism

Donald Meinig's four volume historical geography *The shaping of America* may serve as an entry point to historical geographical research into the spaces of imperialism. The historical spaces of the US as nation state can only be interpreted through the lenses of imperialistic geographies and by working at nested scales from districts within the contiguous states, through regions within the continent, to the Americas as a whole, and to the world scale. As a nation state, the US was forged from territories that were once colonies of Spain, Mexico, France, the Netherlands, Sweden, Russia, or England. The Revolutionary War grew out of smuggling and contraband, and constituted national identities in both the US and Canada through rebellion against and loyalty to the British Crown. A breakaway settler colonialism took hold in the newly independent Thirteen Colonies which morphed into an imperialist project of its own, one

Box 1 Trace: The New Zealand Native Team tours the United Kingdom 1888–1889

The New Zealand Native football representatives undertook a 14-month long, 107-match tour. The team was dogged by controversy, treated as something of a circus, and purposely forgotten by New Zealand rugby authorities. Independent entrepreneurs conceived this as a money-making proposition. They knew that 'natives' were exotic in the UK, that a tour could be a commercial success, and that the standard of UK rugby was not that high. They selected a team of semiprofessional players, only a few of whom were fully 'Maori' or even experienced rugby players. They settled for the name 'New Zealand Native Team' when the New Zealand rugby unions objected to the use of 'Maori'. Their team ran headlong into tensions between the conservative forces of gentlemanly amateurism and working-class pragmatism that subsequently sundered English rugby as 150 northern clubs broke away to form Rugby League. On tour the Native Team were charged with lack of professionalism, unsporting conduct, drunkenness, and not being Maori enough. Crucially, the image of the colonists was at stake, since any poor showing by a purely Maori team could be dismissed as not reflecting the standards of the colony. In the UK the team was treated as a colonial team: there were to be no excuses. The tour, when treated as trace, highlights the renegotiation of national, ethnic, and class identities in key locales of the British Empire: the playing fields of elite schools and clubs. The Native Team's enviable record also reveals the complexities of center periphery relations: UK rugby really was not that good compared with colonial rugby. The tour also reminds us that some contemporary patterns of international sporting encounters update identities and relations established through the transnational projects of imperial sport. Sources: Lewis, N. and Winder, G.M. (2007). Sporting narratives and globalisation: Making links between the All Black tours of 1905 and 2005. *New Zealand Geographer* 63(3), 202 215; Ryan, G. (1993). *Forerunners of the All Blacks: The 1888 89 New Zealand Native Football Team in Britain, Australia and New Zealand*. Christchurch: Canterbury University Press.

emulated by Ottawa. The acquisition and management of internal colonies became Washington's mode of imperi alism and state formation. Peoples in North America experienced imperialist campaigns of violence as First Nations were subjugated, Mexican and Canadian neigh bours were attacked, and the US imperial project temporarily dissolved into civil war. Out of disparate experiences of American imperialism emerged new na tional, regional, class, gender, and ethnic identities. Meinig maps a diverse and fragmented human world coerced through specific imperial projects, many of which were cast as formative of the nation.

Meinig surveys the emerging internal geography of the United States with its capitals, core, domain, colonies, and peripheral enclaves. Washington DC, the political and military capital, and New York, the financial center, net worked the fledgling empire. Compared with analyses of London's role in the British Empire urban system *The shaping of America* only hints at the transnational spaces and circuits of Washington and New York as imperial and commercial centers. For instance, by 1919, films made in and distributed from Los Angeles and New York already dominated world cinema screens in a form of cultural imperialism. Similarly, Meinig perhaps downplays the growing assertion of American imperial power abroad. For example, President Theodore Roosevelt staged a spec tacular projection of US naval power by sending the Great White Fleet on a round the world voyage (**Figure 5**). Meinig turns instead to the edges of the US Empire: to the border regions with Canada and Mexico, the states of the Caribbean dominated by American military and com mercial interventions; and the contact zones of cattle ranching, for instance. Washington DC worked to control other parts of North America, and later parts of Latin America, Oceania, and Southeast Asia. In these ways Meinig surveys the formation of the United States as national entity as a narrative inseparable from aspects of the historical geography of the United States as empire (**Figure 1**).

Provincializing Europe

Historical geographers' tactics of critical enquiry include provincializing Europe: that is, de centering the metro politan centers from the focus of enquiry and breaking up Europe as a category. This is partly done in an effort to critique and step beyond the global imaginaries of imperialisms, with their picturing of the world as for mulated and exhibited in European capitals. It also leads to study of the effects of Empire on Europe. Europeans imposed identities and relations on the rest of the world through discourses about civilization, race, and environment, and developed their own self identity through this mirror. In these ways experience of empire remade Europe yet, intriguingly, much of Europe was subject to imperial rule, and all of it was transformed by it. Europe is the product of conquest, colonization, and attendant cultural transformation. Historical geographers are beginning to provincialize other imperial hegemons – *The shaping of America* is a step in that direction – but this strategy can best be illustrated with Europe in mind.

Historical geographers acknowledge that Europe was in large part made overseas. Colonial categories and discourses helped to frame politics within Europe. The campaign for home rule for Ireland with its related terror incident of the Phoenix Park murders is but one such manifestation. Indeed, the question of Irish nationalism was constitutive of identity politics, migration, labor strife, and as an international phenomenon associated with the Irish diaspora. British politics is hardly under standable without reference to issues of empire and the communications and relations with the colonies.

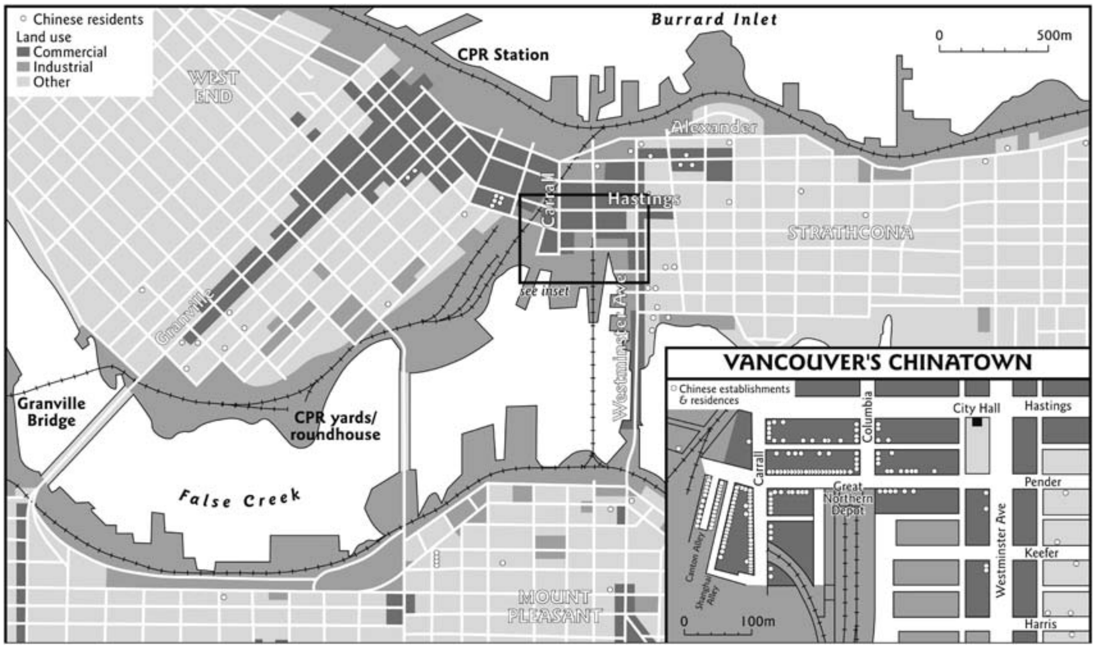

Figure 3 Contact zone: Vancouver's 'Chinatown'. A Chinese-origin community emerged in Vancouver, Canada, in the first decades of the twentieth century. Kay Anderson argues that the 'Chinatown' most Vancouver Chinese occupied was not only a marginal territory from which to launch contested claims to Canadian life but also a racial category constructed by and for European Canadians that was used to define and disempower Chinese in the city. The enclave was overwhelmingly home to men, because of Canada's restrictive immigration laws. These boarding house residents found seasonal work in the city's timber mills and canneries or worked in tailoring and laundry firms, or as cooks, cleaners, pedlars, brick workers, or market gardeners. The enclave was a platform for Chinese businesses (over 230 in 1911) ranging from diversified merchant enterprises through rice millers, silk retailers and labor contractors, to lodging house owners. Although Chinese owned much of the land in this district, 'Chinatown' was above all a racial category imposed from outside. Politicians, police, inspectors, and reformers defined the enclave as an 'unsanitary sink' and a 'nest of vice'. City officials strove to confine Chinese laundries to 'Chinatown', to extinguish licenses to Chinese enterprises, to clear away dwellings under sanitation regulations, to segregate schools, and to apprehend 'gamblers' and 'opium traders' in police raids. Residents fought these intrusions with court action and disrupted the imaginaries forced upon them. This immigrant enclave was a contact zone dominated by the racial stereotypes of Europe and China that framed Vancouver City politics at the time. Source: Adapted from Anderson, K. (1991). *Vancouver's Chinatown*: *Racial Discourse in Canada*, *1875 1980*. Montreal and Kingston: McGill-Queen's University Press, Maps 2 and 3; and McDonald, R. A. J. (1996). *Making Vancouver, 1863 1913*. Vancouver: UBC Press, Map 7.1.

European politics were framed by imperialism in terms of colonial crises, imperial scandals, moral campaigns against slavery or for imperial expansion, and debates over the costs of empire.

Interactions with other cultures and environments remade European culture both in media and on the ground. European literature developed tropes and ima ginaries of the world drawn out of imperial desires. The European imaginary of the South Seas islands was formed through imperialism and its narrative tropes continue to frame inquiries into and reactions to small and distant islands. Other imperial imaginaries – Africa as a 'dark continent', the Middle East as an Oriental other, the Tropics, a transformative 'passage to India' – haunt our ways of thinking about the world. At the level of everyday words, Europe's engagement with empire en tailed a reconstitution of language. Travel writer Jan Morris identifies 'far flung', 'posh', 'sea route', 'sea port', and 'The Lifeline of Empire' (soubriquet for the Suez Canal), not to mention 'bungalow' or 'veranda' as terms conjured up through experience to express geographical aspects of British Empire. By exhibiting exotic others in museums, galleries, lecture halls, and trade shows, Europeans suffused their own cultures with ideas, nar ratives, and images constructed out of imperialistic re lations and in turn confirmed and directed relations through discourses of domination.

On the ground, Europe not only remade environments abroad through 500 years of the export of European plants and animals but was also transformed by the circulation and import of biota from other environments. Botanic and zoological gardens served as key nodes in the redistri bution of plants and animals. European ideas were them selves transformed through encounter with the New World. Indeed, both the Enlightenment and Darwin's Theory of Evolution were charged by exploration of the New World. Historical geographers survey the circulation of ideas and the production of knowledge in Europe.

European geographers themselves have played knowing roles in imperialisms through the geographical

Figure 4 Geographic imaginary of Empire. Imperialist geographical imaginaries were often adopted by companies for advertising purposes, whether they sold at home or abroad. In this lithograph cover illustration on the McCormick Harvesting Machine Company catalog for 1889, the artist shows a Persian horseman looking down upon a caravan of farm machinery with the caption 'Seeking New Fields to Conquer'. Here McCormick advertised machinery to American farmers by wrapping its products in the racist imagery of a peaceful, civilized conquest of backward, foreign lands. Even though he dominates the scene, the 'traditional' Persian warrior's position is perilous: he is outflanked, his rifle useless against this peaceful, modern invading column. McCormick's products, a civilizing force made in Chicago, advance unopposed into the heart of this imagined foreign land. There their work will be revolutionary, 'freeing' peasants from labor, ushering in an age of plenty for some independent, modern farmers. McCormick's reaper is thus portrayed as provident but destined to sweep aside entire ways of life. Such an imaginary bolstered the self-image of American farmers as propertied, productive, and modern, the very backbone of an independent, civilized, and prosperous United States. Sources: Seeking New Fields to Conquer, in McCormick Harvesting Machine Co., 1889, *McCormick 1889*, McCormick Harvesting Machine Company Papers, Advertising leaflets and brochures (printed), McC Mss 5X, Box 3, McCormick Collection, Wisconsin Historical Society, Madison, Image ID: 4412. See also Winder, G. (2007). A trans-national machine on the world stage: Representing McCormick's Reaper through world's fairs, 1851 1902. *Journal of Historical Geography* 33, 352 376.

knowledge we have generated (**Box 2**). This is a matter of continuing enquiry among historical geographers as we seek to further provincialize Europe by acknowledging and understanding the ways in which past geographic practices formed knowledge of the world and aided imperial projects. These range from the exploratory expeditions of members of geographical societies, through geographers as cartographers and compilers of information on other regions for military and spy agencies, to science used to justify and extend imperial missions and ideas. Particular attention has been paid to land surveying, cartography, landscape painting, description and analysis, and climate science as practices framing and enabling imperial projects. Shifts in rhetoric are traced, for example, from biological to cultural racism. In these cases the geographical discourses published by geographers performed the work of empire.

We also acknowledge that Europe was not a cohesive entity. New regional geographies of growth, transformation, and connectivity were generated as some regions within Europe gained more than others from colonization of overseas empires, and different regions were engaged in different ways with Empire. Some parts of Europe (like Ireland) became regions of persistent outmigration. Soldier recruitment dominated parts of Scotland, Brunswick, or Kassel, while industrial production for export stimulated developments in Birmingham, Glasgow, and Manchester. Some cities flourished as metropolitan commercial centers, ports, and financial hubs, but rural areas like Hesse, Lincolnshire, or rural Denmark, struggled to compete with colonial producers of grain, wool, meat, dairy products, timber, or precious metals. Generally western and northern coastal regions of Europe boomed but there was an intricate internal geography of growth, urbanization, and regional decline associated with European imperialism.

Historical geographers devote attention to outposts of empire, their networks, their function as contact zones, and their influence back on the metropolitan centers. An archipelago of forts, naval bases, telegraph stations, post offices, hill stations, work camps, plantations, missions, consular and diplomatic offices, enclaves, and treaty ports grounded imperial projects. These constituted crucial contact points between imperial agents and indigenous peoples. By analyzing the framings of indigenous others, as well as those of their own identities, made and shared by explorers, missionaries, settlers, and other imperial agents, historical geographers demonstrate that the political, journalistic, scientific, and religious texts flying back to imperial centers from the diverse agents located in imperial outposts worked to reform imperial policy.

Experiences of the Colonized

Recovering the experiences of the colonized remains a major problem for historical geographies of imperialism. Colonizers and colonized were all infected (colonized) by imperial practices, discourses, effects, power relations, and oppositions, as well as collaborations. In rebellion

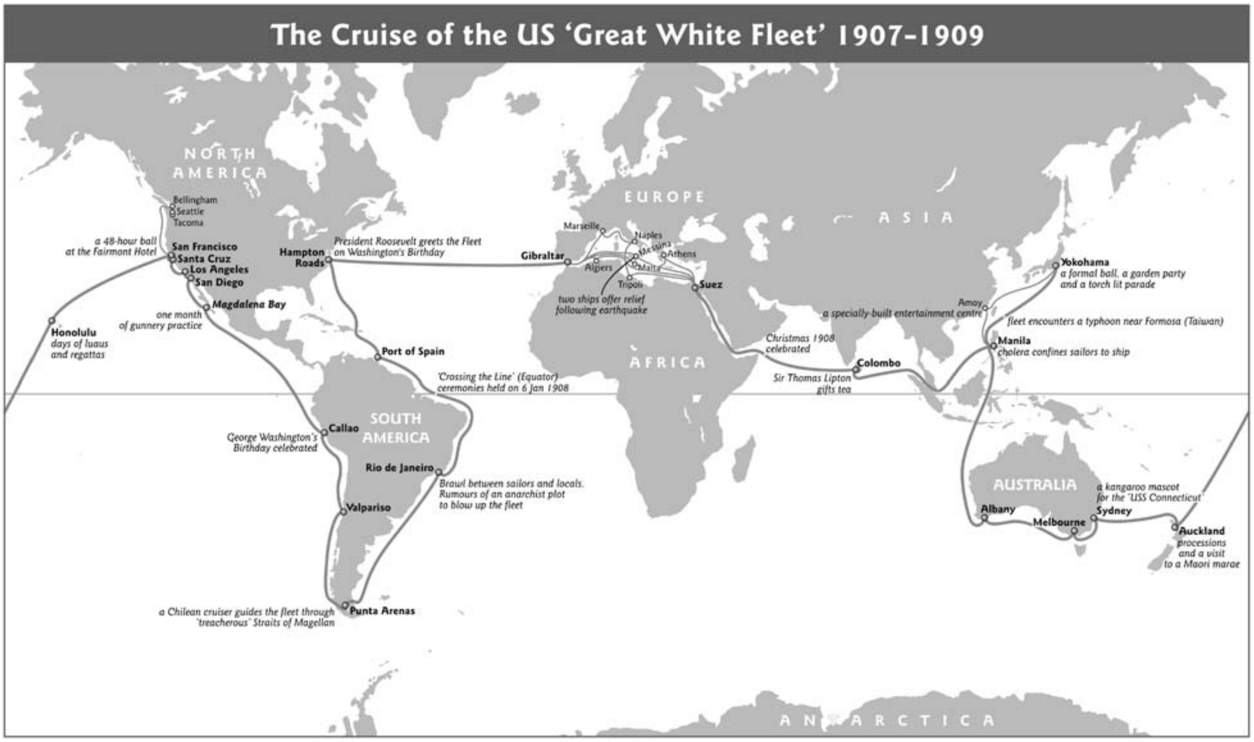

Figure 5 Geopolitics: The cruise of the US' Great White Fleet 1907 08. Brainchild of President Roosevelt, who sought to project US power and prestige abroad, this 14-month round the world cruise of 16 US battleships, their escorts, and 14 000-strong crew, was a promotional coup. Roosevelt had two principal strategic objectives. With the Japanese Navy newly dominant in the Pacific following the Russo-Japanese War of 1905, Roosevelt planned to demonstrate US naval power in the Pacific and thus pave the way for a treaty to curb Japanese imperial expansion. The Root-Takahira Agreement followed hard on the heels of the returning fleet and secured mutual Japanese and US respect for each others' territories, for China's independence and territorial integrity, and for the established 'Open Door' policy in China. Roosevelt also sought to increase the prestige and popularity of the navy at home as he fought for appropriations to pay for four more battleships. Pitched as a successful good-will mission, the cruise revealed a fatal dependence of the US Navy in any effort it might make to project its military power around the globe: it would have to rely on foreign coaling stations. Source: http://www.history.navy.mil/library/online/gwf cruise.htm.

Box 2 In Russian and French prisons

Among past geographers it is perhaps Prince Peter Kropotkin who held the most problematic relation to empire. Having renounced his estates, title, and allegiance to the Russian Tsar following his service in a Cossack regiment, Kropotkin used one hand to print books and articles promoting cooperation, mutualism, and the alleviation of poverty in the names of first nihilism, then anarchism, while the other hand published encouragement to those anarchists seeking to overthrow tyrants by throwing bombs. *In Russian and French prisons*, penned by Kropotkin out of his own experience of imprisonment, might direct our attention to historical geographies of the spatial apparatus of state terror and incarceration, as penal colonies, mental asylums, and concentration camps modernized imperial technologies of domination. Sources: Kropotkin, P. (1887). *In Russian and French Prisons*. London: Ward and Downey; Kropotkin, P. (1927). Prisons and their moral influence on prisoners. In Baldwin, R.N. (ed.) *Kropotkin's Revolutionary Pamphlets: A Collection of Writings by Peter Kropotkin*, pp 219 235. New York: Dover Publications.

and resistance there was often a coupling of colonial and anticolonial discourses. Some of the colonized mimicked the imperialists. The practices of colonization partially erased other ways of knowing. Historical geographers are keenly aware that hybrid spaces and identities emerged in all parts of empires and beyond. Research focuses on the processes forming and curtailing them and on the interactions associated with them.

Many people benefited not only from colonization but from being colonized. They may have been marginal before and could claim power through colonial projects. Colonial governance often required collaboration with local allies, including members of the colonized society. Imperial power was often precarious whether expressed in terms of acquiring territories or holding onto them. How indeed could the puny British military forces in the subcontinent dominate and defeat the much larger ar mies of the Indian princes? As elsewhere the secret lay in collaboration with powerful local allies and not merely in superior weaponry. Spanish imperialism produced dramatic impacts in Latin America not just through the draconian laws of the Indies with their regulations to incorporate colonized peoples into pueblos, missions, and forts and thus to control their lives and relations, but through the Spanish settlers who facilitated colonization. Ironically such settlers could not hold administrative office in the colonies, and this facet of the imperial structure itself fostered rebellion and the sweeping independence wars of 1830. European power was spread thinly and aided and abetted by many settlers and non Europeans.

Historical geographers investigate resistance to imperialism in all its many forms, from disrupting the survey pegs and parties, through peaceful protests and various forms of flight, to revolt, revolution, and terror campaigns, and to the envisioning of utopian communities. Understanding and evaluating these often spatial tactics and strategies of opposition remains a challenging task. Historical geographies of the Toheroa Wars, a century of conflict over shellfish, reveal the New Zealand foreshore as a cultural and political area in which Maori continue to struggle with the buried epistemologies of domination set in place in the colonial past. In another context, the proponents of Irish republicanism have been shown to have mobilized in networks that crisscrossed the Atlantic and featured many of the spatial practices reminiscent of Cold War liberation struggles: fund raising overseas, arms deals, armed struggle, terror campaigns, and espionage. Revolutionary war was a spasmodic feature of ongoing Irish resistance to British imperialism. Other practices earned English names inspired from Irish experience: boycott and moonlighting. A micro politics of cultural affiliation to the spaces of recreation and sports characterized the struggle and, as in other countries, has left a legacy of conflict on the ground in place names, rituals, and segregated spaces. Historical geographers develop a lexicon of the spatial tactics of resistance to imperialism.

Even under direct imperial control, local and indigenous practices and regulations continued in the face of imperial reform campaigns and efforts to impose new regulations. Collaboration led to hybrid and local regulatory frames and to the persistence or perhaps modernization of existing practices and spaces. The Treaty of Waitangi signed by British representatives and Maori chiefs in 1840 made New Zealand part of the British Empire but also guaranteed customary practices, including resource rights, to chiefs. New Zealand's subsequent history is marked to this day by settler efforts to extinguish Maori rights and of Maori to reassert them. By the mid nineteenth century the British Empire featured multiple places, identities, practices, and discourses of prostitution, an array at variance with and often beyond the control of metropolitan discourses against prostitution. Authorities in India modernized the spaces of prostitution by placing brothels in compounds under direct police surveillance. In other jurisdictions British administrators launched raids on brothels in order to police them. The activities of prostitutes, the soldiers, and sailors of the British Army and Royal Navy, and the campaigns of British and colonial reformers shaped a changing geography of locales for prostitution around the Empire. Similarly Britain's parliament declared slavery illegal everywhere in the Empire, but authorities in some colonies, like the Sudan, had neither the resources to eradicate slavery nor the interest: it would simply make labor relations and expenses impossible. Race categories were often used, along with gender and class, to establish varied regulatory regimes and norms of behavior. An extreme case is the formation of segregated spaces and identities in South Africa, elaborated to its fullest extent under apartheid, but similar practices were deployed in other circumstances such as in the delineation of Chinatowns (**Figure 3**) or Indian reserves in Canada. The spatial and discursive practices of such racially based exclusions have been a focus of historical geographic research.

The spaces of women's lives were deeply shaped by imperialisms. Historically, imperialisms have framed women within patriarchal relations so the transformative power of encounter, imperialist desires, and the new spaces of expanding empires offered only limited opportunities for women's agency. In some parts of empire, especially ones where gender ratios were highly skewed, women gained measures of independence and wealth as proprietors of boarding houses, taverns, and other establishments. Empire offered opportunities for some women to travel and thus for travel writing and other modes of expression within the circuits of imperial representation, but few women gained independence through travel from patterns of domination in patriarchal societies. Nevertheless, these are matters of renewed enquiry. Historical geographers have turned to the concept of contact zones as we try to elucidate the ways in which women were constituted through interaction and co presence with others, rather than through separation. Thus the diaries of the Vicereine of India, Lady Curzon, reveal her frustration with a daily routine of confinements and public performances but also her contact with an array of people in diverse settings. Churches might also serve as interethnic contact zones, where middle class women evangelized immigrant Chinese men as they learned English in Sunday school rooms. The cultural dynamics of such contact zones are not straightforward. The experiences of women under empire were varied and hybrid as were the spaces of empires.

Historically, many societies were subject to imperial power but not as formal colonies. Instead they may have experienced imperialism through occasional military incursions, or unequal treaties that promised protection, flags of convenience for trade, or technical advice.

European colonization was often only a passing phase for many peoples and territories. Neither Egyptian nor British imperialists were able to control the territory delineated on maps called the Sudan. At any time between 1850 and 1950 parts of that territory and its peoples lay outside their administrative control. The core of the Ottoman Empire avoided European rule but its vast empire was broken up by European imperialists. The coming of the Europeans was not always a transcendental event, but was often cast as such by European imperialists who sought to substantiate claims to European superiority. In such territories, European imperialism found limited expression: usually as enclaves of European housing and business. For some, it was Japanese or American imperialism, or the aggressive interventions of sub imperialists like New Zealand, Argentina, or the Zulu that proved to have the most impact.

Imperialisms set in train forced migration, resettlement, and increased mobility, forces that worked to build transnational networks, sometimes through diasporic community formation, and to spread their related landscape features. Thus, hybrid work gangs floated on the Atlantic Ocean and washed ashore at ports featuring taverns, seamen's missions, and union halls. Singapore's shop houses and camps featured as locales in one prominent node in the transnational movement of Chinese to Pacific Ocean port cities and their Chinatowns. Relations in Singapore, San Francisco, and Victoria, or in Saint John, Boston, and New York, were made more interesting through the arrivals of multiple others mobilized through the agencies of imperialism and not simply the presence of colonizers. They also made possible new identities and networks within empire. Some colonized subjects acted as factors and agents in imperial trade, others as administrators, soldiers, exhibits, or servants. Intermarriage and a diverse array of sexualized relations formed households that did not conform to discourses of the norm circulated through metropolitan centers. They created specific social strata and mixed identities, for example, as Mestizos and Mulattos in Latin American contexts, Coloureds in South Africa, Métis in Western Canada, or Pākeha Maori in New Zealand. The big houses of Montreal's Protestant Anglo elite might contain an ageing patriarch, his younger wife and offspring, and numerous, especially female and Catholic servants whether of Irish or French Canadian ethnicity. The lived terrain of identity was marked almost everywhere and at every scale by increased migration and mobility.

In this context, 'nation' is a subsequent category to 'empire' and is often formed out of the peculiar circumstances of empire. Independent minded neo Britons and new Spaniards claimed to be progressive and thus justified the formation of new 'nations' out of colonial settlement. Whether we think of Serbs subjugated by the Austro Hungarian Empire or Indians colonized by the British, ethnic groups seized upon ethnic nationalism as a counter to the imperialisms of others. Often the nation states that emerged out of experiences of imperialism were fraught affairs. In Singapore, national identity is produced around conceptions of a multiethnic state and the dominance of Singaporeans of Chinese ethnicity. India's independence produced four nation states and other nationalist movements including those of Sikhs and Tamils whose dreams have not been fulfilled. Nation states, legitimized by the Versailles Peace Treaty of 1919, have fractured existing empires and produced new imperialisms of their own.

Ecological Imperialism

The relations between imperialism, environmental change, and conservation are complex. Work in historical geography has been stimulated by Alfred Crosby's use of the term ecological imperialism to draw attention to the increased trade and transfer of animals, plants, and diseases through European imperialism and mercantilism, and their devastating ecological impacts. The recent work of environmental historian, Robert Grove has highlighted the growing ecological awareness of imperialists, especially those who encountered environmental catastrophes in their experience of encounters with small island ecosystems. Historical geographers have turned their attention to the constructions used by scientists, explorers, and others to frame and understand other environments. Two focal points for this mode of research are tropicality and insularity, which emerged historically as forms of landscape Orientalism, in both cases heavily inflected by gendered associations with a female nature awaiting discovery and encounter. Historical geographers have launched enquiries into the networks of imperial science dedicated to environmental transformation. For example, a team of researchers is currently mapping historical geographies of the imperial networks behind the intensification of production on New Zealand's grasslands. The contradictory intentions and results of wildlife conservation, notably in the national parks systems of African countries, have been the subject of other research. In each case the researchers seek to understand the interplay of imperial and local forces and resistances around imperial projects, as well as the ecological and social impacts of these imperial technologies of environmental management. Historical geographers write environmental histories of land alienation, resource extraction, development, modernization, and their attendant impacts, social injustices, crises, and catastrophes.

Historical Economies of Imperialism

The relations between the commercial world and empire are also important to imperialistic geographies. Although most historical geographies of economy have been efforts to establish or critique narratives of national economic progress, modernization, and development, the growth and changing geography of world trade, telecommuni cations systems, and transnational corporations have been sketched in synoptic terms. Particular attention has been paid to surveying and tracing aspects of the eighteenth century world economy. Other research confirms findings from business history that hybrid organisation – cartels, joint ventures, strategic alliances, arrangements with compradors and indigenous entrepreneurs, networks of merchants or licensed producers, regional business groups – flourished in the historical spaces and interstices of empires in the second half of the nineteenth century (**Figure 2**). Less attention has been paid to the systematic destruction of indigenous economies by imperialists, in the contexts of indigenous incorporation into European dominated resource extraction economies during early phases of colonization, as in Canada or New Zealand. Even less attention has been directed to the destruction of competitive economies by imperialists. Indeed non European economies and commercial networks seldom feature and the networks and traces of smuggling, piracy and illicit production, and trade are underdeveloped within historical geography, as they are within economic geography.

Historical geographers take a long term view and seek to compare and contrast contemporary modes of inte gration within the world economy with those of former, explicitly imperial, times. In this work it is crucial to establish the historical and geographic links between commercial activity and political and military power, if we are to understand the roots of the unequal trading relations under contemporary globalization. Thus, the origins and life of the Manchester free trade doctrine have been sketched within economic geography, but the main accounts of protectionist and market regulations remain those of economic history. Neil Smith argues that, under a rhetoric of 'imperialism without colonies', Washington copied existing and developed new forms of domination after 1918, especially in the realms of com mercial and financial relations, in order to extend its imperial influence while circumscribing the power of other imperialist powers. Economic geographers have rejoined old debates over how best to frame imperialism in economic terms. It has long been thought that Lenin's theory of imperialism as interimperialist commercial and territorial competition lacked relevance in the post colonial, Cold War world. Attention is now redirected to Kautsky's conception of ultra imperialism, marked by cooperation among leading powers to govern the per iphery. Economic geographers point to the emergence of a phantom state of transnational governance in production, finance, and other services, and regulatory structures such as the International Monetary Fund, World Bank, and General Agreement on Tariffs and Trade. Historical economic geographies of empire need to pay more attention to the historical crises of empires and their economic impacts, and to the commercial practices of empires, including the cultures and identities of economic agents.

Empire as Heritage: Remembering and Forgetting

Historical geographers share a focus on interpreting the landscapes of both precolonial and colonial times. It is a mistake to think of imperialism as an annihilating and overwhelming force in all circumstances. Erasure was possible in some places but precolonial landscapes persist in many other places, and palimpsests emerge through the remaking of landscapes. Empires are remembered, erased, and forgotten. For travel writer Jan Morris

> most of us remember [the British Empire] less vividly now as a political organism than as a sort of gigantic exhibition. We have forgotten its statistics, even its ex tent, but we can see the theatre of it still. (Jan Morris)

Empires and their others create and are constituted through monuments and commemorative spaces. Indeed the rituals associated with such memorials continue to be used to shape public behavior. Sites of memorial can be turned to new purposes, protested, or forgotten. Korean symbolic space was remade during and after Japanese imperial rule. One former Japanese prime minister proved highly controversial because of his politically charged visits to Japanese war graves. By replicating or not the imagery of empire, including its heroes and heroines in place names, monuments, chapels, and signs – like those commemorating Queen Victoria or nurse Edith Cavell – colonial subjects embraced or abandoned imperial discourses of morality, belonging, and identity (**Figure 6**). Not all geographical imaginaries of empire were made in the center and circulated outward. Thus the Anchor Butter brand was made in New Zealand and exported to the UK as an icon of the ties that bind the far flung British Empire. Nor are the memorials always sited within imperial territory. Allied war graves and the rituals associated with them are spread around the globe thanks to two world wars. However, the politics of im perial heritage are necessarily tied up with the historical geographies of imperialism.

Critical Review: Historical Geographies of Imperialisms

Historical geographies of imperialism emphasize the complexity of imperialism in space. By exploring colonial encounters, recent research reveals imperialism as a complex lived event and colonial authority as ambiguous. Diverse views about difference, superiority, belonging, and identity complicated encounters and relations among people. By increasing mobility and migration, and thus throwing peoples together in new ways, imperialisms developed hybrid identities, contested spaces, and network connectivity. Imperial projects generated new locales inscribed as expressions of power and domination but often these formed within localities where imperial authorities competed with other forms of governance so that their power was negotiated and challenged. Even the racist and sexist representations that imagined and sustained empires were themselves resisted and reworked, whether in the metropolitan centers or the imperial peripheries. By researching the contact zones, networks,

Figure 6 Ambiguous imperial memorial. This statue of Sir George Grey, successively British Governor General of South Australia, New Zealand, Natal, and New Zealand, was originally erected outside Auckland's Town Hall in the city's main commercial avenue, Queen Street. Grey turned to colonial politics and became a lightning rod for radical dissent against the policies of powerful settler politicians. The widower Grey's domestic arrangements became subject to opprobrium as he took up residence on a coastal island with his niece. His statue became the focal point for political demonstrations, including those of trade union activists at the head of a General Strike in the city in 1908. The City Council subsequently declared Grey's statue a traffic hazard and retired it to a position near the statue of Queen Victoria in sedate Albert Park, a Victorian recreation ground. There, Sir George lost his head in an attack by Maori land activists in the 1990s. Today he has regained his poise but Grey's association with the Empire remains deeply ambivalent. He was seen at various times as an agent of land alienation and a friend to Maori, as a promoter of Empire, but also colonial self-government and radical settler politics.

traces, and surveys of imperial encounters historical geographers are remapping the bounds of empire and challenging synoptic views of empires as homogenous territories.

Nevertheless, historical geographers wrestle with a problem of introversion: European knowledge tends to be at the center of all enquiries and was itself usually framed within imperial and colonial discourses and practices. If we rely on such data sets then it is difficult to apprehend the activities of the colonized or those who lived outside colonial administration and rule. A postcolonial historical geography of imperialism must confront this problem by researching imperialism from the vantage points offered by alternative texts. This also means that researchers must confront the many legacies of imperialism and its discontents that serve as mental blocks or containers: the nation state as frame of reference; the language of the imperialists; settler guilt and the anger of social and environmental injustice. By developing intersubjectivity as a key aspect of their projects historical geographers hope to further postcolonial research into imperialism. Charting the course, problems, and achievements of alternatives to empire and imperialism is thus another issue for historical geographies of empire.

Attention has increasingly shifted to the discursive modes that constituted empires: to imperialistic geographies. This remains a matter of debate in terms of emphasis. While recent attention to travel narratives, spaces of colonial encounter, the exhibition, cartography and exploration of empire, and the geographical imaginaries at play has deepened geographical enquiry there is concern in some quarters that we may be neglecting research into the physical and material forces at work in empires. Any study of imperialism must take power relations seriously. To do so we must pay attention to both the discursive or epistemic violence of colonization, as first formulated by Edward Said, and the modes of conquest, appropriation, domination, and alienation that Cole Harris reminds us are at the bloody edge of imperialism. The geographic technologies of dispossession and domination now stock a formidable arsenal: imaginaries and discourses of imperialism, racism, and difference; the roles of European knowledge in domination, including the work of explorers, cartographers, colonial governors, imperial armies, navies, botanical and zoological gardens, settler governments, missionaries, scientific theories of climate, race, species extinctions, and conservation; the threat of violence and the exercise of bodily domination in many forms from military conquest to state terror. Historical geographers will be stretched if we are to do justice to the spatial practices of imperialism.

What is the current name of empire? If historical geographers are to draw out parallels and comparisons with past imperialisms then we must engage with debates over the contemporary faces of imperialism and their others. Placing the US Empire and its modes of operation in the second half of the twentieth century at the focal point of research must be a priority. Equally, the changing technologies of domination must be charted over the long history of imperialisms. The roles of geographers *vis à vis* empire remain at issue as both the United Nations and military and spy agencies demand up to the minute geographical expertise. Arguably, we now witness intensified colonialisms with perhaps new or reinvented forms of domination, and with revived or new ideologies of imperialism, but we cannot know this without reliable historical geographies of past imperialisms. Without further enquiry into the historical geography of empire, there is a danger that geographers may underestimate the features of classical European imperialism, thinking of it as coloring the world map red, blue, or green. It is equally easy to underestimate the US as empire by reading its projects through the lens of its own rhetoric of an end to colonialism. Caricatures of historical imperialisms will not aid us as we seek to interpret the patterns and implications of the intensifying imperialisms emerging around us.

See also: Critical Geography; Diaspora; Feminism/Feminist Geography; Globalization and Transnational Corporations; Hybridity; Landscape; Orientalism; Utopian Cities.

Further Reading

Anderson, K. (1991). *Vancouver's Chinatown: Racial Discourse in Canada, 1875 1980*. Montreal and Kingston: McGill Queen's University Press.

Blunt, A. and Rose, G. (eds.) (1994). *Writing Women and Space: Colonial and Postcolonial Geographies*. New York: Guilford.

Clayton, D. (2000). *Islands of Truth: The Imperial Fashioning of Vancouver Island*. Vancouver: UBC Press.

Clayton, D. (2003). Critical imperial and colonial geogreaphies. In Anderson, K., Domosh, M., Pile, S. & Thrift, N. (eds.) *Handbook of Cultural Geography*, pp 254 367. London: Sage.

Connell, J. (2003). Island dreaming: The contemplation of Pacific Paradise. *Journal of Historical Geography* 29(4), 554 581.

Driver, F. and Gilbert, D. (eds.) (1999). *Imperial Cities: Landscape, Display and Identity.* Manchester: Manchester University Press.

Driver, F. and Martins, L. (2006). *Tropical Visions in an Age of Empire*. Chicago: Chicago University Press.

Graham, B. and Nash, C. (eds.) (2000). *Modern Historical Geographies*. Edinburgh: Pearson Education.

Gregory, D. (1994). *Geographical Imaginations*. Cambridge, MA: Blackwell Publishing.

Gregory, D. (2004). *The Colonial Present: Afghanistan, Palestine and Iraq*. Malden, MA: Blackwell Publishing.

Harris, R. C. (1997). *The Resettlement of British Columbia: Essays on Colonialism and Geographical Change*. Vancouver: UBC Press.

Harris, R. C. (2004). How did colonialism dispossess? Comments from an edge of empire. *Annals of the Association of American Geographers* 94(1), 165 182.

Kropotkin, P. (1887). *In Russian and French Prisons*. London: Ward and Downey.

Kropotkin, P. (1927). Prisons and their moral influence on prisoners. In Baldwin, R. N. (ed.) *Kropotkin's Revolutionary Pamphlets: A*

Collection of Writings by Peter Kropotkin, pp 219 235. New York: Dover Publications.
Lewis, N. and Winder, G. N. (2007). Sporting narratives and globalization: Making links between the All Black tours of 1905 and 2007. *New Zealand Geographer* 63(3), 202 215.
McDonald, R. A. J. (1996). *Making Vancouver, 1863 1913*. Vancouver: UBC Press.
Meinig, D. (1988 2004). *The Shaping of America: A Geographical Perspective on 500 Years of History*. New Haven: Yale University Press.
Mitchell, T. (2002). *Rule of Experts: Egypt, Techo politics, modernity*. Berkeley: University of California Press.
Morris, J. (1982). *The Spectacle of Empire: Style, Effect and the Pax Britannica*. Garden City, NY: Doubleday and Company.
Ogborn, M. (2005). Atlantic geographies. *Social and Cultural Geographies* 6(3), 379 385.
Pawson, E. and Brooking, T. (eds.) (2002). *Environmental Histories of New Zealand*. Oxford: Oxford University Press.
Proudfoot, L. and Roche, M. (eds.) (2005). *(Dis)placing Empire: Renegotiating British Colonial Geographies*. Aldershot: Ashgate.
Ryan, G. (1993). *Forerunners of the All Blacks: The 1888 89 New Zealand Native Football Team in Britain, Australia, and New Zealand*. Christchurch: Canterbury University Press.
Sheppard, E. (2005). Constructing free trade: From Manchester boosterism to global management. *Transactions of the Institute of British Geographers NS* 30, 151 172.
Sidaway, J. (2000). Postcolonial geographies: An exploratory essay. *Progress in Human Geography* 24(4), 591 612.
Smith, N. (2005). *The Endgame of Globalization*. New York: Routledge.
Smith, N. and Godlewska, A. (1994). *Geography and Empire*. Oxford: Blackwell.
Winder, G. (2006). Webs of enterprise 1850 1914: Applying a broad definition of FDI. *Annals of the Association of American Geographers* 96(4), 788 806.
Winder, G. (2007). A trans national machine on the world stage: Representing McCormick's Reaper through world's fairs, 1851 1902. *Journal of Historical Geography* 33, 352 376.

Relevant Websites

http://www.history.navy.mil
The Cruise of the Great White Fleet, Naval Historical Center.

Indian Ocean

S. Chaturvedi, Panjab University, Chandigarh, India

Introduction

With approximately 74 million sq km and roughly 20% of the global ocean, the Indian Ocean is the third largest ocean after the Pacific and the Atlantic (**Figure 1**). It serves as a natural 'transit lounge' for most of the traffic from the Atlantic to the Pacific and vice versa. Two key characteristics distinguish the Indian Ocean from the Atlantic and the Pacific. First, only one fifth of the total trade is conducted among the countries of the Indian Ocean themselves. About 80% of the trade is extraregional; for example, crude oil to Europe, the USA, and Japan. In the Atlantic and the Pacific, the proportion is exactly vice versa. Second, contrary to the Atlantic and the Pacific as 'open' oceans, the Indian Ocean can only be accessed through several 'choke points': From the West via Cape of Good Hope and the Straits of Madagascar, from the North via the Bab el Mandeb at the end of the Red Sea and the Straits of Hormuz at the exit of the Persian Gulf, from the East via the Straits of Malacca, the Sunda and Lombok Straits, and the Ombai–Wetar Straits (**Figure 2**).

A decision in early 2000 by the International Hydrographic Organization (IHO) to delimit a fifth world ocean from the southern portions of the Atlantic

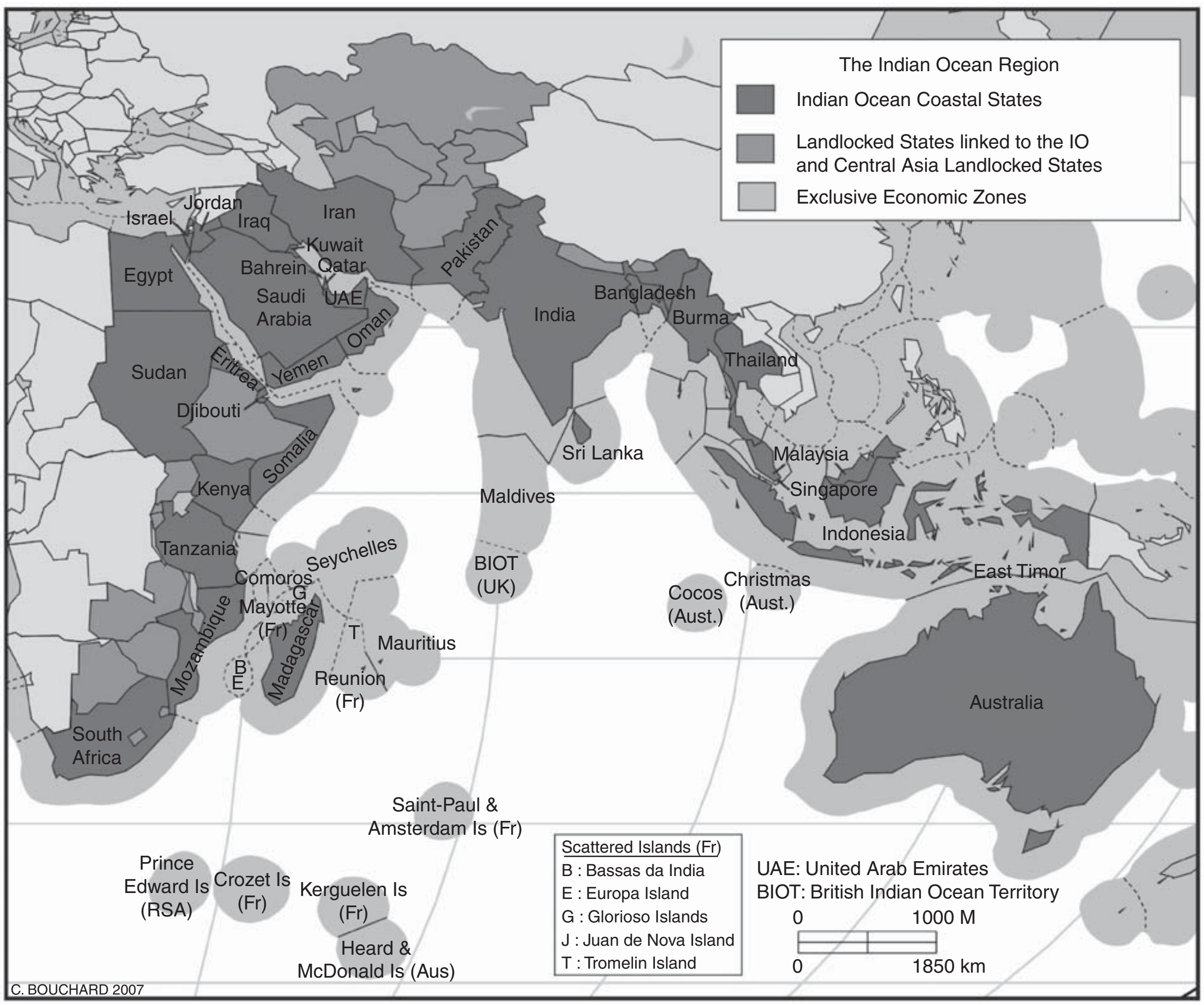

Figure 1 The Indian ocean region. Source: Christian, B. (2007). France and United Kingdom as the 37th and 38th Indian Ocean Coastal States. *Indian Ocean Survey (A Journal of Indian Ocean Research Group)* 3(1&2), 28.

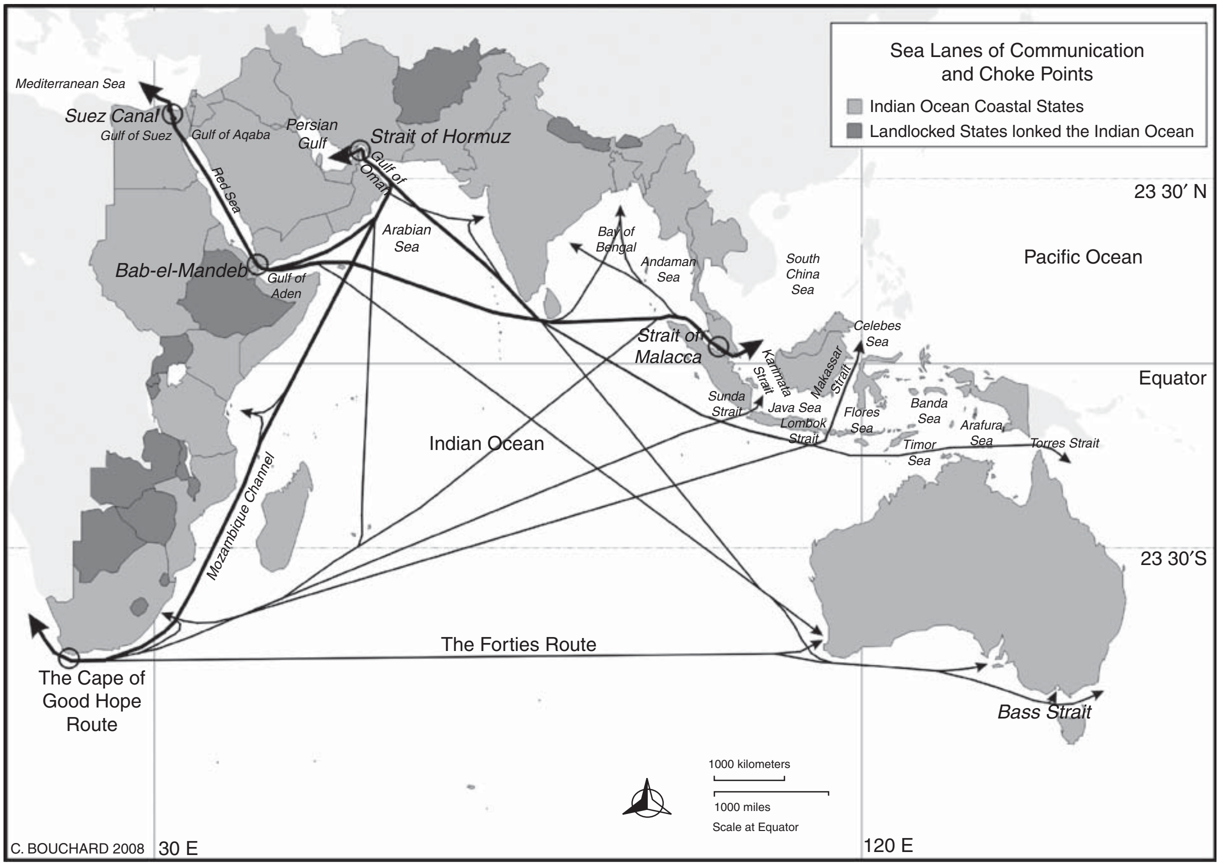

Figure 2 Sea lanes of communication and chocke points.

Ocean, Indian Ocean, and the Pacific Ocean, has reduced the surface area of the third largest of the world's five oceans (after the Pacific and Atlantic Oceans, but larger than the Southern and Arctic Oceans). The revised area of the Indian Ocean basin is about 68 560 million sq km. It includes the Andaman Sea, Arabian Sea, Arafura Sea, Bay of Bengal, Great Australian Bight, Gulf of Aden, Gulf of Oman, Mozambique Channel, Persian Gulf, Red Sea, Timor Sea, and Straits of Malacca and Singapore, and other interconnecting tributary water bodies.

There are at least five interrelated reasons why Western observers generally tend to undermine the strategic importance of the Indian Ocean Region (IOR). First, since some commentators see the Region as pri marily comprising the Third World states, by association the Region has been accorded a lower level of signifi cance compared with other oceans, especially the Atlantic and the Pacific Oceans. Second, the Northern centric primarily English based system of global know ledge contributes to a generalized level of ignorance about the Region and other regional states. Northern–Western resources are unlikely to support Indian Ocean educational programs when better economic returns are perceived to exist elsewhere. Third, since the Region is a former arena of European colonial competition, to some degree, residual core–periphery values of social, eco nomic, and political superiority and security dependency continue to prevail. Fourth, there is relative paucity of reliable writing from students of international affairs on the geostrategic importance of the IOR. Finally, the geopolitical orientation of many Indian Ocean regional states has tended to be away from the Ocean and has been directed internally or toward states and regions of perceived greater geoeconomic significance.

Mapping the 'Indian Ocean': Histories, Categories, and Connections

A rather strict and static delimitation of the Indian Ocean (such as the one attempted above) has been contested by a number of historians and for good reasons. Before the European power projection began in the late fifteenth century, Indian Ocean populations were bonded by large scale maritime trading systems and outside influences were minimal. The so called pre Vasco de Gama period (before 1497–98), was characterized by regional self sufficiency and autonomy. The IOR has been recognized by some researchers as a communication hub for civil izations established well before the rise of Europe. Others would say that in precolonial times the Indian Ocean was a 'crucible' for the first global economy involving Chinese, South Asian, and Middle Eastern trade.

As pointed out by Kenneth McPherson in *The Indian Ocean: A History of People and the Sea*, the IOR was the center of the first sophisticated commercial and maritime activities and the home of the world's first urban civilization. The Indian Ocean enabled the peoples on its shore to share and exchange a range of cultural values and economic activities. What Braudel called, the *longue dure'e* or the long term rhythms of long distance maritime trade formed the basis of the essential unity of the Indian Ocean 'world'. An enormous range of cultural and economic practices ably supported by a unique maritime trading system marked the Indian Ocean space. This web of intimate interactions provided the peoples of that 'world' with an economic unity and certain cultural commonalties, which set them apart from the peoples of contiguous 'worlds' such as the Mediterranean and East Asia. Maritime trade induced linkages facilitated the diffusion of religious and cultural systems such as Hinduism, Buddhism, and Islam across the Ocean and gave rise to cultural commonalities.

Sugata Bose in *A Hundred Horizons: The Indian Ocean in the Age of Global Empire* argues that in order to explore the richness of the Indian Ocean in its rich mega diversity, we need to deterritorialize our imaginations of the Indian Ocean and approach its 'hundred horizons' (or frontiers) in terms of the fuzzy and fluctuating boundaries of its basin and maritime regions or subregions. Kirti N. Chaudhuri reminds us through his *Trade and Civilization in the Indian Ocean: History from the Rise of Islam to 1750* that even though the Indian Ocean did not share a 'common destiny', it reflected a sense of its own unity for certain kinds of analyses – for example, means of travel, movement of peoples, economic exchanges, and climate. Whereas factors such as religion, social systems, and cultural traditions, on the other hand, provided a contrast and challenged its conceptualization as a single unit of space.

This 'Indian Ocean world' with its 'deep structure' (that provided some 'elements of continuity' to its littoral societies) was gradually penetrated (but, according to some historians, not incorporated) by the colonial order, characterized by European economic and political control, which began after 1500 and culminated in the second half of the nineteenth and first half of the twentieth centuries. By and large, the Europeans were compelled to adjust or accommodate themselves to preexisting structures. Michael Pearson in his seminal work *The Indian Ocean* sees this as a turning point from the history 'of' the Indian Ocean to a history 'in' the Indian Ocean, deeply influenced by wider forces emanating from outside its otherwise flexible boundaries.

For the most part of modern history, the Indian Ocean remained a 'British lake', the security of which was guaranteed by the Royal Navy until the termination of all security commitments 'East of Suez' in 1971. The Indian Ocean was a region of huge significance for the British imperial world economy and its strategic integrity for two centuries. It encompassed key sea routes and ocean annexes like the Persian Gulf, the Red Sea, and the Bay of Bengal, and kept the eastern empire in touch with the European heartland. It united Africa, Asia, Australia, and the Middle East and contained within these land borders was a galaxy of islands and atolls. War between Britain and other European powers was a major impetus for an imperial presence, and for imperial conquest, in the Indian Ocean. Victory over European rivals in the wars of the eighteenth and nineteenth centuries left Britain master of the Indian Ocean and the foremost European power in India and the Far East.

Since World War II, the IOR has witnessed tremendous geopolitical change as almost every littoral state gained its independence, and then saw its international relations subordinated by Cold War geopolitical constraints. From the mid 1960s onward, a steady increase of superpower presence, paralleling the decrease of British presence, was noticeable.

Growing Militarization of the Indian Ocean: The Great Base Race

Military strongholds often have played a major role in the shaping of the destiny of the Indian Ocean and its littoral. The Indian Ocean, after nearly a decade long strategic low profile and invisibility is once again becoming an area of geostrategic rivalry among a number of major powers and littoral states. The social construction of the Indian Ocean space as a force field or battlefield is being currently reinforced by the great base race in the region.

The 'Indian Oceanic arc of militarization' stretches from Egypt to Indonesia and Australia. A number of military–naval infrastructure projects of various dimensions are being undertaken by India, Pakistan, Burma, and Malaysia, among others. Many of these new facilities are aiming at power projection and this strategic dynamism reflects the new wealth and confidence of some of these states, besides the increased importance of mobility and connectivity in this era of globalization, the presence in the Indian Ocean of strategic natural resources (especially oil and gas), and the 'rise' of India. According to some analysts, the identity and coherence of the Indian Ocean as a geostrategic region is likely to be deepened due to new naval bases being established in the area. Also, this new 'great base race' will make war, at least among the ocean's littoral states, a more practicable option than previously. Moreover, it may not be easy for Western powers, including the United States, to block or moderate the power of Asian states.

Against the backdrop of the facts that proliferation of weapons of mass destruction (WMD) has considerably enhanced the destructive capacity of the Indian Ocean states, and their strategic space (so far largely confined to the land) is now widening to include the sea, air, and even outer space, regional conflicts might become even more dangerous and further complicate the maneuvers of foreign militaries in the region.

Growing Focus on the Maritime Geopolitical Economy: Development of Ports and the Shipping Sector in the Indian Ocean Rim

As far as the issues of international geopolitical economy are concerned, the IOR clearly appears to be multipolar. The geopolitical economy of this region is essentially characterized by a very large disparity among national economies, the significance of natural resources, and the weakness of intraregional trade. The regional economic powers are India, Australia, Indonesia, Singapore, Thailand, Malaysia, Saudi Arabia, South Africa, and Iran. Together, these nine countries represent 72.9% of the total regional gross national product (GNP) and 78.0% of regional exports. The economic geography of the region is dominated by oil production in the Persian Gulf and the related maritime oil fluxes; the industrial production of the Southeast Asian tigers that is intended to serve world markets; important commercial and financial crossroads such as Singapore; the mineral resources of Southern Africa, Australia (especially uranium), and Indonesia (oil and gas included); the diversified economy of India and the size of its national market; as well as a great variety and the large volume of tropical products exported to the developed countries of the North.

The increasing importance of the Indian Ocean as a global maritime trade facilitator is reinforced by the fact that the bulk of the world's shipping passes through the region. Against the backdrop of trade liberalization, and as a result of the export–import orientation of many trading countries and their energy imports, maritime trade passing through the ocean has dramatically increased. The Indian Ocean provides maritime advantages to several of its littoral states in terms of their strategic location along the Ocean.

The overall growth in global trade has further heightened the importance of this vital ocean – with its strategic waterways and links with major maritime trade routes – in linking the trade of countries and economic regions. The tremendous growth of ports and the shipping sector in the IOR is a testimony to the regional and global trade explosion. Several maritime infrastructures such as ports and shipping facilities have been built and upgraded along the Indian Ocean rim. A number of important harbors and seaports dotted along the Region's shores, stretching from Durban in South Africa to Melbourne in Australia, are critical trade facilitators to the economies of the respective countries and to the IOR.

With an increasing number of Indian Ocean countries engaging in international trade with distant neighbors, increasing volume of the world goods will be transported via the Ocean. In fact, trade volumes have already been increasing steadily in the IOR. Over US$100 billion of China's trade alone passes through the Indian Ocean annually, and this figure is projected to grow rapidly. Bilateral trade between China and India, two of the world's emerging economic powerhouses, stood at US$13 billion in 2004, representing 1% of China's global trade and 9% of India's. There has also been a significant rise in trade volumes moving between post apartheid South Africa and other Indian Ocean states such as India, Malaysia, Singapore, and Australia, and there is substantial potential for future growth. Following the far reaching changes in South Africa's economic, social, and political policies, the country is poised to become a powerful engine for growth among countries in sub Saharan Africa and the island states along the traditional Cape route.

India's emergence as an economic power, with a population of more than a billion, provides another tremendous opportunity for further economic and trade growth in the IOR. It is expected that these developments will have tremendous spillover effects on the growth of the maritime sector across the IOR to support trade expansion. Regional initiatives such as the formation of Association of Shipping and Port Authorities in the Indian Ocean by the Indian Ocean Marine Affairs Cooperation (IOMAC) will further boost efforts to promote port and shipping development in the region.

The demand for new ships, as a result of booming global trade and mandatory phasing out of ageing tankers by the International Maritime Organization (IMO), has resulted in a huge number of building orders. The crisis and collapse of European shipyards and the cost competitiveness of Asian shipyards have resulted in a dramatic shift of the shipbuilding industry from the West to East. This has had a positive effect in the development of shipping in the IOR, to a certain extent. Although the shipbuilding industry is dominated by yards from the 'big three' shipbuilding giants of Asia – China, Japan, and South Korea – several Indian Ocean states have also got into the act. The combined share of the world's orders from yards in India, Pakistan, and Bangladesh stood at nearly 4% in 2004. India has taken a cue from China to embark on a major shipbuilding program by promoting new shipyards and upgrading existing ones.

Looking ahead, much enhanced investment in port facilities and upgrading shipping infrastructure in the IOR would significantly boost regional trade. The projected upward trend in cargo traffic worldwide will require massive investments for developing and improving port and shipping facilities in the region to cope with the growth. Trading states in the Indian Ocean rim will be compelled to upgrade their national lines to meet increasing demand for shipping services and to capitalize on container traffic growth. A more modern maritime sector would provide a catalyst to higher trade volume and brighter economic prospects for many Indian Ocean countries, which are economically dependent on the Ocean.

Security of the Indian Ocean Sea Lanes of Communication

In view of increasing transnational commerce, the shipping lanes of the Indian Ocean (Hormuz, Malacca, Sunda, Lombok, and Suez) have assumed a pivotal significance. A large volume of international long haul maritime cargo from the Persian Gulf, Africa, the Asia Pacific, and Europe transits through its waters. Comprehensive sea lane security requires a shift from governmental, unilateral, and bilateral approaches to system wide arrangements that consider the supply chain in totality.

A steady increase in the number of ships passing through the straits has increased the risk of threat to environmental security. Ship based pollution, although relatively easier to tackle than land based pollution, is more resource intensive. To catch a polluting ship while transiting the straits – carrying out either tank cleaning or ballasting operations – requires much more enforcement capability at sea. Then, there is the ever increasing risk of pollution by discharges of oil or other noxious substances caused by accidents as the straits get more congested. Loss of biodiversity and the damage to ecosystems in the straits will have devastating effects not only on the fishing industry but also the marine tourism industry.

The Strait of Hormuz

Hormuz lies between Oman and Iran. It is strategically important to most countries in the world. The Strait of Hormuz forms a strategic link between the oil fields of the Persian Gulf, the Gulf of Oman, and the Indian Ocean. It has a width between 48 and 80 km, but navigation is limited to two 3 km wide channels, each exclusively used for inbound or outbound traffic. Circulation in and out of the Persian Gulf is thus highly constrained, because the sizable amount of tanker traffic makes navigation difficult along the narrow channels. In addition, islands that insure the control of the Strait are contested by Iran and the United Arab Emirates. Question marks over the security of the Strait have been raised from time to time.

Between 1984 and 1987 the 'Tanker War' took place between Iran and Iraq, where each belligerent (Iran–Iraq War of 1980–88) began firing on tankers, even neutrals, bound for their respective ports. Shipping in the Persian Gulf dropped by 25%, forcing the intervention of the United States to help secure the oil shipping lanes. About 88% of all the petroleum exported from the Persian Gulf transits through the Strait of Hormuz, bound to Asia, Western Europe, and the United States. Its importance in global oil circulation cannot be overstated. For instance, 75% of all Japanese oil imports transit through the Strait. There are thus very few alternative outlets to oil exports from the Persian Gulf if the traffic of about 14 Mb d^{-1} going through Hormuz was compromised. In addition, a no notice closure of the Strait would trap about 7–10% of the global very large crude carrier (VLCC) fleet in the Persian Gulf.

Straits of Malacca

The Strait of Malacca is the longest international navigational route through a strait, where it connects the Indian Ocean via Andaman Sea on the north with South China Sea in the south via the Straits of Singapore. Nevertheless, the Strait provides the shortest route for ships between East Asia and Europe. With an average width between 11 and 200 nautical miles, the seaway in this Strait is not always wide as at certain parts of the Strait, the navigable route is less than 1 nautical mile, and at certain parts its navigable area is less than 30 m deep. At a particular point on the Strait, the maximum draught recommended by the IMO for passing ships is 19.8 m.

Certain 'unfriendly' navigational features notwithstanding, the Strait of Malacca is still an attractive and preferred navigational route for international shippers compared to other alternative routes like Sunda or Lombok–Makassar Straits. The 50 mi long Sunda Strait is 15 mi wide on its northeastern entrance but the strong currents and limited depth has prevented deep draftships of over 100 000 deadweight tonnes (DWT) from transiting this Strait. What makes the Sunda Strait relatively less suitable to oil tankers are features such as a tricky channel, depth limitations, and a live volcano. Moreover, highly irregular bottom topography of the Straits and the presence of rip tides render it further unsuitable for vessels drawing over 18 m of water to use the Straits. In addition, there are hazards posed by the numerous oil drilling platforms to the north of western Java. The Sunda Straits have as yet no available detailed navigation charts.

With a minimum passage width of 11.5 mi and depth of more than 150 m, the Lombok Strait, on the other hand is the safest route for supertankers because it is wider, deeper, and less congested than the Strait of Malacca. Both Lombok and Makassar Straits do not feature serious navigational hazards along the navigational channel but the routes require an additional of 31/2 days navigating at full commercial speed of 14–16 kn, an extra mileage of 1600 nautical miles, which add to transpor tation cost where in the long term operation it is not profitable to the tanker operators.

From the Persian Gulf, the sea lanes transit through the Indian Ocean almost hugging the Indian coast toward the Straits of Malacca. Malacca is the busiest with over 600 vessels transiting through it every day. Virtually all ships destined for East Asia carrying liquefied natural gas (LNG) and liquefied petroleum gas (LPG) pass through Malacca and the issue of safety is likely to grow in im portance as East and Southeast Asia's energy imports grow. It is estimated that the number of tankers transiting through Malacca would increase to 59 per day in 2010 from 45 per day in 2000. Similarly, the LPG tanker traffic is expected to rise to 12 per day from about 8 in 2000.

Because the Malacca–Singapore Straits are territorial seas of the coastal states, these states (Malaysia, Indonesia, and Singapore) have very strong interests in the safety of navigation through them. On this basis, Malaysia and Indonesia have expressed serious reservations about any proposal aiming at the 'internationalization' of these Straits.

Despite increasing optimism in sea based commerce, Indian Ocean sea lanes continue to be vulnerable to disruption, increasingly threatened by piracy, drug traf ficking, gun running, human smuggling, and pollution. The specter of maritime terrorism has acquired addi tional visibility and urgency after 9/11.

Energy Security and Indian Ocean Nuclear Trade

In the future, the IOR will become increasingly im portant to any discussion of energy security, not only because of the critical importance of regional states as energy suppliers, but also because of the essential role of the Ocean as an energy route. In the IOR, while some regional states are 'energy dependent' – for example, India (coal) and Singapore (oil) – on the other hand, several Indian Ocean states are net energy exporters and, on a global scale, the region contains some very signifi cant 'resource holders'. Apart from the Gulf states for oil and natural gas, these include India, Australia, and South Africa for coal, and Australia and South Africa for ura nium. However, in terms of the significance of energy in their overall trade profile, five regional states – Australia, Iran, UAE, Indonesia, and Oman – could be regarded as 'energy niche economies'.

The increasing demand for nuclear energy includes those states which are energy dependent (e.g., India), those which are industrializing (e.g., China), and those involved in the development of nuclear technology (e.g., Iran). Growing regional and global interest in nuclear energy is also being propelled by a belief in its 'green house friendly' status. In 2006, in the IOR, the number of planned or proposed nuclear reactors in five states (India, Indonesia, Iran, Pakistan, and South Africa) represented 39% of the global total. Australia is still debating its position on nuclear energy. Nonetheless, the proposed rapid regional and global growth in the production of nuclear energy in turn places pressure on regional ur anium supply states which are 'uranium rich' but are currently 'underproducing'. The question of the storage and movement of nuclear waste is probably the most difficult and controversial issue in the Indian Ocean nuclear trade and Australia is still seen by many as po tentially the most suitable global waste repository.

Marine Biodiversity, Fisheries, and Climate Change

The IOR contains a remarkably diverse coastal and marine environment that is highly vulnerable to devel opment pressures. Mangroves, as well as sea grass beds, serving as critical habitats for a range of species, are found around the Region. In many parts of the Region, developmental projects on the coastline and shoreline changes (such as loss of arable land; damages to coastal property and infrastructure; and destruction of habitats such as mangroves and coastal forests) have led to de struction of mangroves. Maziwi Island was one of the most important nesting sites in the Western Indian Ocean (WIO) region for three species of marine turtles, namely olive ridley, green, and hawksbill turtle. The Island dis appeared in 1978 apparently due to erosion.

The IOR contains a range of coral reefs that provide an important source of food for local communities and are of major commercial importance for fisheries and tourism. Destructive practices such as excavations of coral and coral sand, sand mining, illegal fishing practices, including the use of explosives and poisons, have damaged these reef systems. It has been estimated that 20% of coral reefs and 5% of sea grass beds have been destroyed; yet there is little documentary evidence of the extent of this damage.

Environmental degradation is exacerbated by rapidly increasing coastal populations, widespread poverty that in turn promotes the unsustainable use of coastal re sources, poorly planned economic development, under resourced government institutions, and weak implemen tation of existing laws and policies.

Coastal development poses challenges even for de veloped countries within the region such as Australia that have, relatively speaking, effective governance arrange ments. As for other developing Indian Ocean states coastal management is clearly even more of a challenge. The Maldives – an archipelago on the North Indian

Ocean – comprises 1190 islands and 26 atolls stretching over 870 km in roughly a north–south orientation. The islands are low in elevation, many rising less than a meter above sea level, making them vulnerable to extreme events – as shown in the case of Indian Ocean tsunami of December 2004. India, in contrast, has a coastline of 8129 km and an exclusive economic zone (EEZ) (declared in 1976) of approximately 2 million sq km, with its coastal zone heavily populated and the location of major industry. On the WIO, Kenya has a coastline of 600 km, with its coastal zone supporting 9% of the country's population. South Africa has a coastline of 3000 km, with the western coast highly productive from an upwelling ecosystem. An estimated 30% of South Africa's population lives within 60 km of the coast, with the additional challenge of the unprecedented growth of informal settlements in the coastal zone.

Fishing is a major activity in the IOR (Indian Ocean has the second largest fishing fleet in the world after the Pacific Ocean) with artisanal or small scale commercial fisheries being important sources of food and protein for communities around the region. Fishing remains critical for both food security and export income for Indian Ocean states. In the case of the Maldives, for example, fishing is a mainstay of the economy, with 70% of export earnings being derived from fishing and related activities. The presence of high numbers of local and foreign fishing vessels have contributed to the steady increase in the total fish catch in the western and eastern IOR which has increased from 1.2 million to 3.9 million t and from 1 million to 4.7 million t, respectively, between 1970 and 2000. This has created tremendous pressure on the fisheries resource with the Food and Agricultural Organization (FAO) expressing concern over the status of tuna resources and fisheries resources in the deep waters of the Indian Ocean. The pressure of overfishing is also felt by noncommercial and endangered marine species. One disturbing feature of the Indian Ocean fishery is the high level of discards amounting to approximately 4.3 million t in 2002.

Within the IOR, the southwest subregion appears to be most adversely affected by overfishing, with 75% of fisheries resources being exploited at the maximum biological productivity level, mostly by external fishing vessels from Spain, Taiwan, Japan, and Uruguay. This has prompted the establishment of the South West Indian Ocean Fisheries Commission (SWIOFC) with the objective of assisting countries to manage the resources in a sustainable manner and to complement the work of existing bodies such as the Indian Ocean Tuna Commission (ICTC). The establishment of such a body, however, needs to be complemented by an overall improvement in the capacity of Indian Ocean states to manage their fisheries resources to ensure sustainability and safeguarding the interests of coastal fishing communities, who depend on the Indian Ocean for sustenance and livelihood.

The Indian Ocean Tsunami caused massive human and economic destruction. Fishing communities along the coastal regions were among the most severely affected. There are certain lessons to be learnt here. In the long run, the local and traditional institutions should be strengthened, after assessing their roles, potentials, and limitations. Protection and restoration of coastal habitats and biodiversity should be undertaken on a priority basis and should not be confined to tsunami affected areas. It is necessary to implement/put in place measures to regulate activities that can pollute, degrade, or otherwise harm the coastal environment and its capacity to protect coastal communities from future natural disasters. These restoration programs in tsunami affected areas should be undertaken in participatory ways.

According to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC), the time for doubt has already passed and there is no denying the fact that our climate system is warming and is directly linked to human activity. Rate of global average sea level rise has risen from 1.8 to 3.1 mm yr^{-1} from 1961 to 1991. Projected sea level rise (due to thermal expansion, melting glaciers and ice caps and the polar ice sheets) at the end of the twenty first century will be 18–59 cm.

Obviously, coastal communities, along with low lying island communities, are also most at risk from climate change. Originally, climate change was considered as a classic environmental security issue in Northern terms. As aforesaid, the North came to comprehend that it had to share the same atmosphere and oceans with the South, and a continued policy of *laissez faire* would be detrimental for the more affluent world. Also, climate change was a classic environmental issue portrayed as a natural 'force', rather than one, which directly related to people. Consequently, climate change has not traditionally been seen as a key issue in the South. In recent times, however, it has assumed greater prominence among majority world environmentalists, due to the fact that some of the biggest state polluters and/or consumers of fossil fuels have not signed the climate change protocols in Kyoto and Johannesburg. Further tensions have emerged, due to the fact that both island states in the Indian Ocean, as well as poorer dwellers and coastal fish workers on coastlines, will be the principal victims in global climate change. In this manner, climate change has metamorphosed from an elite, scientific, Northern issue into one, which can usefully fit into the environmental justice agenda of the South.

Revisiting 'Indian Ocean Studies': Looking Back and Forth

More contemporary factors and forces such as the ones analyzed above demand a critical reassessment of what constitutes (or should constitute) the 'Indian Ocean studies'. As pointed out earlier in this article, a number of studies have interrogated Eurocentric historiography and assumptions that marginalized Asian trading activity as socially insignificant, economically limited, and culturally incapable of resisting the might of Western guns, law, and technology. 'Historicizing the ocean' is a welcome step because it helps bring to attention large scale historical processes of commercial, biological, and cultural exchange that other geographical constructs and imaginations often obscure.

Markus P. M. Vink in his seminal article "Indian Ocean studies and the 'new thalassology'" forcefully argues in favor of acknowledging the imperative to localize and historicize Indian Ocean studies. In his view it is absolutely crucial to focus deterritorialized concepts emphasizing interaction, movement, and migration, as well as the key terms like porousness, permeability, connectedness, flexibility, and openness of spatial, as well as temporal boundaries and borders.

This article has further identified a number of themes of considerable present and futuristic significance that demand and deserve to be brought within the purview of Indian Ocean studies.

Conclusions

Looking to the future, it appears that the 'great void' construction and the perceived (constructed) placelessness of the Indian Ocean, will persist in some form. In such a geoeconomic vision, the Indian Ocean would continue to be idealized as a space of energy and transport flows, devoid of nature, society, or geographical differentiation. As Philip E. Steinberg describes so succinctly in *Social Constructions of Ocean Space* (2001), the great void visualization of ocean space as essentially a transportation space, is likely to be reinforced due to the growth and diversification of container ships and automated, internodal container ports. Once the movement of trade by sea, rail, and truck is constructed as one continuous flow, the unique sociocultural features of the environments through which the containers move are likely to be conveniently pushed into oblivion.

The crucial point here is that beyond the elitist, capitalist, and statist imaginations of maritime life exists an empirically rich world of Indian Ocean communities and individuals (e.g., fishing communities, maritime workers) whose everyday lives, experiences, and worldviews revolve around entirely different encounters with the ocean. In India, for example, there are around 3638 fishing villages all along the Indian coastline and the total population of marine fishers is about 3 million, out of which 1 million are actively involved in the said activity; for example, fishers whose lives revolve around knowledge of distinct fish rich places within the ocean. However, the existence of such communities and their alternate constructions of the ocean as a space of representation are denied by the representations of ocean space that emanate from postmodern capitalism's dominant spatial practices on the one hand, and the ascending geostrategic visualizations focused on warships, choke points, and naval power projections. The future of Indian Ocean studies looks both interesting and challenging.

See also: Military Geographies; Oceans; South Asia.

Further Reading

Berlin, D. (2004). The 'great base race' in the Indian Ocean littoral: Conflict prevention or stimulation. *Contemporary South Asia* 13(3), 239 255.

Bose, S. (2006). *A Hundred Horizons: The Indian Ocean in the Age of Global Empire*. Cambridge: Harvard University Press.

Chandra, S. (ed.) (1987). *The Indian Ocean: Explorations in History, Commerce and Politics*. New Delhi: Sage.

Chaturvedi, S. (1998). Common security? Geopolitics, development, South Asia and the Indian Ocean. *Third World Quarterly* 19(4), 701 724.

Doyle, T. and Risely, M. (eds.) (2008). *Crucible for Survival: Environmental Security and Justice in the Indian Ocean Region*. Chapel Hill, NC: Rutger University Press.

House, J. (1984). War, peace and conflict resolution: Towards an Indian Ocean model. *Transactions of the Institute of British Geographers* 9(1), 3 21.

Ong Webb, G. G. (ed.) (2006). *Piracy, Maritime Terrorism and Securing the Straits of Malacca*. Singapore: Institute of Southeast Asian Studies.

Rumley, D., Chaturvedi, S. and Taib, M. (eds.) (2007). *Security of the Sea Lanes of Communication in the Indian Ocean Region*. Kuala Lumpur: Maritime Institute of Malaysia.

Rumley, D. and Chaturvedi, S. (eds.) (2004). *Geopolitical Orientations, Regionalism and Security in the Indian Ocean*. New Delhi: South Asian Publishers.

Rumley, D. and Chaturvedi, S. (eds.) (2006). *Energy Security in the Indian Ocean Region*. New Delhi: South Asian Publishers.

Steinberg, P. (2001). *Social Constructions of the Ocean*. Cambridge: Cambridge University Press.

McPherson, K. (1993). *The Indian Ocean: A History of People and the Sea*. New York: Oxford University Press.

Pearson, M. (2003). *The Indian Ocean*. London: Routledge.

Pearson, M. (2005). *The World of the Indian Ocean, 1500 1800: Studies in Economic, Social and Cultural History*. Burlington, VT: Ashgate.

Vink, M. P. M. (2007). Indian Ocean studies and the 'new thalassology'. *Journal of Global History* 2, 41 62.

Relevant Websites

http://www.iotc.org
Indian Ocean Tuna Commission.

http://www.sidsnet.org
Small Island Developing States Network.

Indigeneity

E. Cameron, Queen's University, Kingston, ON, Canada
S. de Leeuw and M. Greenwood, University of Northern British Columbia, Prince George, BC, Canada

Glossary

Binary Dualism A view that postulates two things or ideas as opposed and mutually exclusive. For example, the notion that mind and body, man and woman, and nature and culture are separate and opposed realms. In a binary dualism, the meaning of each 'pole' is constructed in reference to its 'other'; the terms have no meaning except in relation to each other.

Constructivism A belief that the world is composed of theories and practices, and is not independent of them. In other words, nothing is knowable outside the categories and imaginings that are placed upon them. Constructivism may be a form of relativism in that it suggests there is no way of judging reality separately from how we theorize it.

Deconstruction A process of exposing the contradictions and contingencies embedded in language, ideas, and concepts. Based on the premise that all knowledge is socially constructed from particular vantage points, and through the operation of power, deconstruction aims to take apart such constructions and to challenge their apparent rationality, stability, and sense.

Environmental Determinism A theory proposing that the physical environment, including climate, elevation, and geology, defines the basic characteristics of human bodies, cultures, and societies.

Essentialism The notion that a person or thing has certain qualities that are essential to it, without which it would no longer 'be' that thing. For example, the belief that Indigenous peoples must live in harmony with nature in order to properly 'be' Indigenous is essentialist.

Fourth World A term employed to characterize highly marginalized Indigenous groups who are dominated by other groups or by state bureaucracies. It signals both the variability of social and economic well-being within the so-called First World countries, as well as the shared political interests of Indigenous peoples who have been dispossessed of the lands that comprise various nation-states.

Naming Indigenous Peoples The nomenclature relating to Indigenous peoples is neither straightforward nor depoliticized. Indigenous peoples are called, and name themselves with, a wide variety of words, many of which are unstable and shift over time. For instance, in both Canada and the United States, the term 'Indian' has been supplanted, respectively, by the terms 'First Nations' and 'Native American'. In Canada, however, 'First Nations' does not include Inuit (formally Eskimo) peoples, nor does the term denote Métis peoples, although all three (First Nations, Inuit, and Métis) are constitutionally recognized under the terms Aboriginal and Indigenous. Other names for Indigenous peoples have also transformed over time (e.g., the Sami people were once widely referred to as Laplanders) and within large groupings of 'Indigenous' peoples there are often distinct and particular tribes, communities, families, or lineages with individual names. Even the collective term 'Indigenous' is unfixed and in some contexts is displaced by terms such as 'First Peoples' and 'Fourth World Peoples'. When writing about indigeneity and Indigenous peoples, most Indigenous peoples recommend the use of specific and local names; for instance, the Zapatistas are the Indigenous Mayans of Mexico, while the Haida are a First Nation of northwest coastal British Columbia in Canada.

Postcolonial A condition of existing after colonization, or in the wake of colonial practices and domination. Also, the long process whereby a dominant colonizing power transfers the power to control cultural, economic, and political institutions to a colonized people.

Post-Structuralism A set of theories positing that language (and therefore knowledge) is given meaning in relation to other words, rather than referencing some extra-linguistic reality, and therefore directing attention to the role of power in constructing and maintaining particular kinds of knowledges.

Power The ability of a person or object to produce an effect upon, to control or to exercise authority over, other persons or objects.

Racialization The processes by which 'race' (a constructed, invented category) is given material and ideological fixity, such that it appears to be a natural and rational means of understanding human differentiation. This often involves the assignation of certain qualities, characteristics, and life chances to particular groups of people.

Realism A belief that there is an absolute and knowable objective reality or truth. The term carries connotations of practicality and factuality in which entities are knowable and separable from the categories and theories that are placed upon them.

Settler Colonialism A kind of colonialism in which settlers from a colonial nation establish permanent settlements in a colony. These are not temporary occupations, but are rather concerted efforts to establish lasting control over a territory and to outnumber Indigenous peoples with a dominant community of settlers. Canada, Australia, and New Zealand are all settler colonies.

Introduction

Within geography, as in other disciplines, ‘indigeneity’ is a protean, contested, and relatively recently considered term. Although geographic engagement with the term is rather limited, the term does have historical roots within the discipline and, however, tentatively, is developing an increasing contemporary purchase among human geographers. Broadly speaking, human geography’s engagement with the concept of indigeneity might be divided into two temporal domains which encompass four streams, or pillars, of thought. Historically, early and mid twentieth century human geography tended to conflate Indigenous peoples with the natural environment and relied on essentialized understandings of culture and identity. By the late twentieth century, and increasingly within the twenty first century, critical human geographers have destabilized the concept of indigeneity. This destabilization takes the form of post structural critiques, postcolonial interventions into the concept, and a politicization of indigeneity both as a descriptor and as a state of being. Notwithstanding the temporal shift, however, there remains within human geography a divide between scholars operating within constructivist parameters and proponents of realist philosophies; it can be argued that research into indigeneity by human geographers is divided along this philosophical line. On the one hand are researchers sympathetic with post structuralist challenges to essential identities and truths and with postcolonial challenges to ‘pure’ Indigenous figures. These scholars tend to challenge and deconstruct invocations of indigeneity. On the other hand are researchers who either identify as engaging in social justice projects or as feeling compelled by the moral rightness and political necessity of claims to ‘being’ Indigenous. Human geographers in this latter grouping are more inclined to address the social, cultural, economic, and political concerns of Indigenous peoples than to devote their energies to more theoretical questions of just what ‘Indigenous’ might mean.

This is a rough and simplifying sketch of geographical research, of course, and few human geographers would argue that theoretical inquiry is distinct from political or ‘everyday’ concerns. Nevertheless, we would argue that geographical research on indigeneity tends to adopt one stance over the other. As geographers are certainly aware, moreover, the breadth and depth of academic work on a given topic does not represent the scope of ‘reality’, and we submit that, to the extent that Indigenous peoples have been excluded from the production of geographical knowledge about indigeneity (and it is to a rather great extent), that knowledge remains partial, both in the sense of being incomplete, and also in the sense of being saturated by the partiality of (largely) non Aboriginal researchers. Accordingly, while we endeavor to provide an overview of past and current geographical research into indigeneity, we also present the words of Indigenous scholars and writers in an effort to sketch out some of the difficulties and gaps inherent in that research. Consequently, in this article we profile the evolution of the term indigeneity in geographic thinking and discuss alternative formulations of what indigeneity might mean, as expressed by Indigenous peoples themselves.

Historical Approaches to Indigeneity

It is not surprising that geographers have taken an interest in the concept of indigeneity, for it is a thoroughly geographic notion. Etymologically, the term ‘indigeneity’ comes from the Latin *indu* meaning ‘in, within’ and *gen* (root of *gignere*) meaning ‘beget’ or ‘produce’. Indigeneity refers, in that sense, to that which is produced or begotten from within a given place. Indeed, as Marie Battiste, a Mi’kmaq educator from the Potlo’tek First Nation, and James Sa’ke’j Youngblood Henderson of the Bear Clan of the Chickasaw Nation and Cheyenne Tribe in Oklahoma, write in *Protecting Indigenous Knowledge: A Global Challenge*, “the ecologies in which Indigenous peoples live are more to us than settings or places: they are more than homelands or promised homelands. Ecologies do not surround Indigenous peoples; we are an integral part of them and we inherently belong to them. As Indigenous peoples…we unfold our structure of Indigenous life and thought from ecologies.” Indigeneity is not only about geographic origins, then, but it also involves an ongoing interdigitation with place and ecology.

Notwithstanding Indigenous peoples’ assertions concerning a deep interconnectivity with the natural environment, such assertions unfortunately resonate with the essentializing and reductionistic conceptualizations of indigeneity made by early human geographers. A number of significant geographic texts, for instance, are anchored in conceptualizations of indigeneity as a foil to modernity and contemporaneity; indigeneity is made to stand for the past, the premodern, and the lost. Yi Fu Tuan’s work on topophilia, homescapes, and connection to place positions Indigenous peoples as existing in harmonious and close connection with nature, in antithesis to the industrial and modernized existences of often urban dwelling people. Tuan offers ‘Indians’, ‘Eskimos’, ‘Mongoloid peoples’, and ‘Aborigines’ as clear models of peoples who, because of their close connection to nature, displayed particular characteristics determined by their environments. This conceptualization of Indigenous peoples is in some ways informed by theories of environmental determinism. Similarly, in *Human Territoriality*, Robert Sack presents the territorial practices of the Chippewa as opposite to, and less developed than, the European territorial practices that superceded them in time and space. Here, indigeneity is a conceptual foil against which theories of territoriality are elaborated. Finally, Carl Sauer's foundational work on cultural landscapes aimed at uncovering traces of past inhabitation in place and posited successive cycles of cultural landscape formation, where one formation would climax and recede as a subsequent culture made its imprint on the land. This model tended to relegate Indigenous presence in the landscape to history and 'tradition', eliding the ongoing presence of Indigenous peoples in contemporary places.

Over the past several decades, geographers have begun to reevaluate their disciplinary heritage and have challenged the assumptions and ideas outlined above. In reworking these problematic approaches to indigeneity within the discipline, however, a tension has arisen between (1) validating and respecting Indigenous peoples' claims of identity and connection with nature and traditions of the past and (2) recognizing that connectivity to nature and tradition, within European discursive constructs, is synonymous with primitivism and tends to overlook the lives of Indigenous peoples in the contemporary world.

Contemporary Approaches to Indigeneity

Conscious of the problems with the work of their predecessors, critical human geographers have set about challenging past ideas about indigeneity and developing new modes of engaging with the concept. We consider in particular the impact of post structuralism, postcolonialism, and political geography in reconfiguring geographic scholarship on indigeneity.

Post-Structuralism

Over the past several decades, the discipline of geography has been heavily influenced by post structuralist theories and modes of inquiry. Post structuralist inquiry often involves the deconstruction of 'essentialist' concepts and claims, and an uncovering of the ways in which knowledge is structured. Thus, geographers interested in troubling and destabilizing concepts of indigeneity are motivated by a desire to understand, if not to overturn, discursive constructs that foreground and privilege some subjects in binary opposition to 'other', marginalized, subjects. To the extent that indigeneity is understood as an essentialist claim – that is, grounded in the belief that certain qualities are inherently and essentially Indigenous – post structuralist geographers are more inclined to challenge and 'trouble' such claims than to accept them as 'true'. Many of the bases upon which indigeneity is asserted have been deconstructed in recent years, as critical cultural geographers have adopted post structuralist, postcolonial, and politicized stances toward concepts such as 'race', identity, and place. Alongside efforts to destabilize essentializing categories, however, some geographers warn that deconstruction can result in the very real and ongoing marginalization of Indigenous peoples through a denial that they have distinct ways of life.

Geographers interested in processes of racialization have pointed out that hierarchies of difference are often articulated through an imagined separation between nature and culture and are perpetuated by buttressing social constructions with claims of objective or scientific authenticity. They explain that the conflation of Indigenous peoples with the ancient past and with nature is a social construction designed to perpetuate European differentiations between civility and savagery and to rationalize the colonization of Indigenous lands. In these binary discourses, it was precisely the imagined closeness of Indigenous peoples with nature that defined them as savage and primitive, whereas European interest in 'transcending' and manipulating nature through technology and reason was held up as a hallmark of a progressive and civilized culture. 'Scientific' theories of race were developed through the lens of European superiority, as were moral imperatives to resettle and 'improve' lands held by Indigenous peoples. A great deal of recent geographic research has been directed toward deconstructing the network of ideas and practices implicated in such racialized constructions.

In an effort to develop new ways of thinking about identity and difference, some critical human geographers have begun to emphasize movement, disconnection, and hybridity over notions of rootedness, belonging, and essential identity. This theoretical and empirical shift in the discipline has led to a certain suspicion of claims to indigeneity, and an interest in troubling and destabilizing the 'truth claims' of any group. Notions of movement, hybridity, and partiality sit uneasily with some Indigenous peoples, and indeed, such notions run the risk of undercutting both the existential and political claims of Indigenous peoples to 'being'. For others, however, this work dismantles restrictive expectations about 'authenticity' and 'tradition' in the constitution of indigeneity and accounts for the multiple and fluid axes of identity through which Indigenous peoples position themselves in the world. It provides an important vantage point from which to query the terms of belonging within Indigenous communities and has also been useful both in reconstructing the continued presence of Indigenous peoples throughout the modern era and in revising histories that have, until now, not acknowledged this presence. Indigenous peoples themselves have clearly asserted their presence in the modern world and insist that no culture is static. Indigenous peoples insist, for example, that their stories, like their orators, change gradually from teller to teller and generation to generation. Similarly, Inuit cultural educator Peter Irniq describes the change and adaptation strategies of Inuit peoples in *A Will to Survive* as follows: "Changes have taken place in our language in recent years. Our language is a living language. New words are being developed to handle modern concepts and possessions.... our culture as well is evolving, [yet] it is still there ... Yes, I am thousands of years old but 54 years in the 21st century." By asserting their dynamic presence in contemporary societies, Indigenous peoples are in some way insisting that the modern world account for them, and particularly for their ongoing marginalization.

Postcolonialism

Human geographers' recent turn toward postcolonialism has had a marked impact on disciplinary conceptualizations of indigeneity. Indeed, the very concept of 'post'colonial in reference to Indigenous peoples is contentious because, as many Indigenous peoples argue, colonial interventions into their lives and lands have not abated nor retreated. This argument is not overlooked by critical human geographers, many of whom observe that postcolonial geographies are not limited to those 'after' colonial rule but include geographies 'in the knowledge' of colonialism. This realization very much reflects Indigenous peoples' position that their lives and lands are not sufficiently conceptualized through a (post)colonial lens. More specifically, however, postcolonial attention to indigeneity and the concerns of Indigenous peoples emphasize expressions and materializations of power, particularly as those expressions and materializations inform relationships between colonial subjects and Indigenous peoples. For postcolonial geographers, many of whom draw on the influential works of Edward Said and his contemporaries Gayatri Spivak and Homi Bhabha, indigeneity is intimately linked with practices and processes of colonization. Colonial control over land, for example, relied upon processes of ideological and discursive 'othering', in which Indigenous peoples were constructed as opposed and inferior to Euro colonial peoples. English language postcolonial geographers conducting research on Indigenous peoples have tended to focus on the former colonies of Britain, specifically New Zealand, Australia, and Canada. Consequently, these nations dominate the literature on indigeneity and postcolonialism, often eclipsing inquiries into other regions with significant populations of peoples who identify themselves as Indigenous.

Most postcolonial geographic concerns with indigeneity identify material and intellectual tools of dispossessing Indigenous peoples from lands and culture. In Canada, postcolonial engagement with indigeneity has resulted in considerable geographic focus on Indigenous peoples of British Columbia, particularly as colonization resulted in the production of spaces ('Indian' reserves) which contained and confined Indigenous peoples. The production of reserves, however, is understood as inseparable from the colonial apparatuses which first imagined, and then legitimated, Euro colonial visions of their right to dominance. Following this logic, material objects like maps, trading posts, missionary schools, and state agents are all implicit to and inextricable from the 'grounded' geographies of 'Indian' reserves; in other words, within colonial practices and processes, the ontological and material cannot be extricated from the epistemological and discursive. The connectivity between culture and power, and an attendant inscription onto land and territory of colonial culture and power, have also been the focus of postcolonial geographers who are investigating indigeneity in Australia and New Zealand. Here geographers have observed that memorialization and the construction of public spaces on landscapes have often resulted in heroic displays of colonial dominance while simultaneously erasing Indigenous precontact presences. Furthermore, there exist ongoing tensions between Indigenous and non Indigenous peoples over boundaries, land claims, and the ownership and allocation of resources. Tensions of representation and display are not confined to inscription on landscape but instead encompass a broad range of nonmaterial spatialities in which Indigenous and non Indigenous peoples are implicated. For instance, postcolonial geographers argue that the very production of knowledge about Indigenous peoples is a practice fraught with constellations of power which in turn affects research methodology, the positionalities of the researcher, and the purpose and outcomes of the research conducted. For postcolonial human geographers interested in the epistemological nuances of research concerning indigeneity, there remains a tension between the knowledge that is produced about Indigenous peoples and those who are engaged in the production of that knowledge. Geographers are increasingly aware of their role in sustaining ideas about Indigenous peoples they had hitherto purported only to mirror, a situation that is inextricable from the fact that the production and dissemination of academic geography remains largely under the control of the non Aboriginal peoples. Addressing these schisms in knowledge production remains an important task of postcolonial human geographers.

It is perhaps not surprising that, just as non Indigenous postcolonial human geographers problematize their roles in producing geographic knowledge about Indigenous peoples, Indigenous peoples themselves are questioning what many suggest is an ongoing colonization of Indigenous knowledges. In other words, while critical human geographers work out the implications of producing knowledge about Indigenous peoples from a (primarily) non Aboriginal perspective, Indigenous peoples themselves are questioning the benefit (or lack there of) of non Aboriginals producing research on 'othered' peoples from an outsider position. For instance, in responding to non Aboriginal researchers' focus on the nuances and fine distinctions inherent to the concept of indigeneity, Sharon Venne of the Cree Nation observes in an essay in *A Will to Survive*, that "colonizers spend so much time trying to determine who is an 'Indigenous' person…[because] it gives a false sense of security for the colonizer trying to control their environment." From a postcolonial perspective then, there is the very real question of how to (de)colonize postcolonial research focused on Indigenous peoples, as opposed to perpetuating ways of thinking and knowing which, in the past and within colonial discourses, often served to legitimize and authenticate non Aboriginal claims to territories and lands occupied by Indigenous peoples.

Indigenous peoples are not the only peoples with connections and claims to place, of course, and postcolonial struggles to reclaim stolen lands inevitably encounter resistance from those who have developed their own connections as settlers. In fact, some argue that settler colonialism requires the generation of a sense of indigeneity among settlers in order both to rationalize the dispossession of Indigenous lands and to develop a sense of belonging and comfort in the settler nation. What separates settler indigeneity from the indigeneity of its first inhabitants is a matter of dispute in many settler colonies, but scholars like Linda Tuhiwai Smith insist that, at base, what distinguishes the claims of Indigenous peoples is the fact that they have no 'other' homeland to return to. Geographers have been particularly attentive to the fraught politics of de colonizing nations with long standing immigrant populations, and have argued that it is not just a desire to maintain control over land and resources that frustrates redistributive processes in de colonizing nations, but also a longing on the part of settlers for the kind of 'ancestral' relationship with place asserted by Indigenous peoples, a longing more broadly associated with modernity. It is in such contexts that a certain desire for the perceived 'benefits' of indigeneity emerges, a desire that is at odds with the ongoing social and economic marginalization of Indigenous peoples and is symptomatic of the continued presence of the colonial imagination in settler colonies.

Political Geography

Indigeneity has also been an abiding theme in political geography, particularly among political ecologists. One of the main premises of political ecological research is that Indigenous peoples experience a disproportionate share of ecological harm wrought by globalized capitalist development. To be Indigenous, in that sense, is to have diminished access to the world's resources, to the political and economic networks of power that manage and distribute those resources, and to be more vulnerable to ecological problems such as drought, erosion, pollution, natural hazards, deforestation, and so on. Similarly, political ecologists have also documented the ways in which Indigenous peoples are excluded from the economic benefits of resource extraction and development, and the ways in which government policy and bureaucracy both facilitate and naturalize unjust divisions of wealth. Underlying this argument is the recognition that social, cultural, and economic well being are structured by thoroughly racialized, gendered, and classed networks of power. Far from naturalizing this inequity, geographers insist that control over land, resources, and capital is an ongoing accomplishment rather than an isolated act, a kind of 'enactment' requiring massive mobilizations of power, aided by discourses and technologies, to sustain it. Political ecologists are particularly sensitive to the ways in which environment, ecology, and nature have been embroiled in processes of racializing, marginalizing, and oppressing Indigenous peoples.

Political geographers have also addressed indigeneity through studies of territorial conflict, law, and geographic scale, considering the plight of Indigenous peoples advocating for national and territorial independence within and across political boundaries that they do not recognize as legitimate, as well as the transnational spaces in which Indigenous identities are produced and reproduced. Indeed, the term 'Indigenous' means nothing to a great number of people who might be described as such. In many cases, moreover, Indigenous peoples have developed a consciousness and identity as 'Indigenous' because of the political necessity of employing this term. The term 'Indigenous peoples' itself emerged from the collaboration of the American Indian Movement (AIM) and the Canadian Indian Brotherhood in the 1970s, and their shared interest in internationalizing the issues and struggles experienced by the world's colonized peoples. Today, multiple, transnational Indigenous movements link specific Indigenous groups with a globalized agenda, movements that are not so much grounded in a single, common, territorialized identity (however much Indigenous peoples might mobilize universal, pan Indigenous discourses) but rather in a shared interest in advancing politically similar causes. Indigeneity 'is' political, according to Roger Maaka and Augie Fleras, because so many Indigenous peoples are engaged in struggles to reclaim their ancestral rights, and because Indigenous peoples continue to live in a world with diminished access to the rights and resources enjoyed by others.

Looking to the Future

If anything, recent geographic engagement with indi geneity and Indigenous peoples has confirmed that knowledge is thoroughly contingent and contested. It is not surprising, then, that Indigenous peoples have placed a great deal of emphasis upon how, why, and by whom knowledge is produced about them, and have insisted that non Indigenous researchers critically evaluate the assumptions, research aims, and methodologies employed in the study of indigeneity. Linda Tuhiwai Smith argues that every step of the research process, from the for mulation of research questions to the interpretation and articulation of findings, must be considered as an im portant venue for de colonization; that is, Indigenous peoples should play an active and directive role in setting research agendas, determining research protocols, carrying out research methodologies, storing and inter preting data, and communicating research findings. In cases where this is not possible, non Indigenous scholars should seriously evaluate the partiality and ethical im plications of proceeding without Indigenous involve ment. To be sure, both Indigenous and non Indigenous peoples have an interest in the places, histories, and so cieties they share, and Indigenous peoples are not calling for a stop to all research relating to them. Instead, they are working to radically alter the terms under which such research proceeds, and to ensure that knowledges pro duced about Indigenous peoples serve, first and foremost, the interests and aspirations of Indigenous communities around the world.

See also: Colonialism I; Colonialism II; Critical Realism/ Critical Realist Geographies; Imperialism, Cultural; Indigenous Geographies; Inequality; Modernity; Nature; Nature-Culture; Postcolonialism/Postcolonial Geographies; Poststructuralism/Poststructuralist Geographies; Racism and Antiracism; Whiteness.

Further Reading

Battiste, M. and Henderson, J. S. Y. (2000). *Protecting Indigenous Knowledge and Heritage: A Global Challenge*. Saskatoon: Purich Publishing.

Bhabha, H. (1990). *Nation and Narration*. London: Routledge.

Castree, N. (2004). Differential geographies: Place, indigenous rights, and 'local' resources. *Political Geography* 23, 133 167.

Castree, N. and Braun, B. (eds.) (2001). *Social Nature: Theory, Practice, and Politics*. Oxford: Blackwell.

Delaney, D. (2005). *Territory: A Short Introduction*. Malden, MA: Blackwell.

Gelder, K. and Jacobs, J. (1998). *Uncanny Australia: Sacredness and Identity in a Postcolonial Nation*. Melbourne, VIC: Melbourne University Press.

Greymorning, S. (ed.) (2004). *A Will to Survive: Indigenous Essays on the Politics of Culture, Language, and Identity*. New York: McGraw Hill Press.

Harris, C. (2002). *Making Native Space: Colonialism, Resistance, and Reserves in British Columbia*. Vancouver, BC: UBC Press.

Maaka, R. and Fleras, A. (2005). *The Politics of Indigeneity: Challenging the State in Canada and Aotearoa New Zealand*. Dunedin: University of Otago Press.

Peet, R. and Watts, M. (eds.) (2004). *Liberation Ecologies: Environment, Development, Social Movements* (2nd edn.). London: Routledge.

Peters, E. (2000). Aboriginal people and Canadian geography: A review of the recent literature. *Canadian Geographer* 44(1), 44 55.

Sack, R. (1986). *Human Territoriality: Its Theory and History.* Cambridge: Cambridge University Press.

Said, E. (1978). *Orientalism*. New York: Pantheon.

Sauer, C. (1975). *Man in Nature*. Berkeley, CA: Turtle Island Foundation.

Spivak, G. (1987). *In Other Worlds: Essays in Cultural Politics*. New York: Methuen.

Smith, L. T. (1999). *Decolonizing Methodologies: Research and Indigenous Peoples*. London: Zed Books.

Tuan, Y. (1974). *Topophilia: A Study of Environmental Perception, Attitudes, and Values*. Englewood Cliffs, NJ: Prentice Hall.

Relevant Websites

http://www.un.org

United Nations Permanent Forum on Indigenous Issues. This international advisory body has a mandate to discuss indigenous issues related to economic and social development, culture, the environment, education, health, and human rights. The Forum has been instrumental in developing a Declaration on the Rights of Indigenous Peoples, also available from the website.

Indigenous Geographies

R. Howitt, S. Muller, and S. Suchet-Pearson, Macquarie University, Sydney, NSW, Australia

Glossary

Colonialism The policy of forced occupation in new territories imposing the practices and assumptions of the colonizing cultures.
Country An aboriginal English term used to define one's relationship to an area (can be land, sea, rivers, or sky) embedded within a cosmological connection through customary law.
Cultural Landscapes A landscapes' perspective that recognizes the influence of people and their cultural practices in shaping and defining the land.
Customary Law Laws that govern Indigenous societies based on cosmologically defined connections between ancestors, country, and people.
Eurocentrism Practices derived from and focused upon European culture to the exclusion of other ontological perspectives.
Indigeneity A controversial term used to describe the state of being Indigenous.
Indigenous Peoples The first peoples to live in an area with customary ties and relationship to particular areas of country.
Native Title Formal recognition of Indigenous people's property rights through their customary connections to the country.
Ontological Pluralism Equitable recognition of multiple ontological perspectives (ways of being and constructing knowledge).
Postcolonialism Movements or studies that investigate, correct, and mitigate the complex impacts of colonialism.

Introduction

Geographical expansion of European and other imperialist economies brought, and continues to bring, the property systems of colonial powers into juxtaposition and often conflicts with Indigenous peoples' ways of knowing and relating to the world. Disruption of Indigenous peoples' familiar cultural landscapes and jurisdictions challenges their ways of relating to social and environmental relationships by imposing new patterns of power, control, and value. Colonizing mobilities and spatialities, and invasive claims and technologies, impose new property rights and grant access to land, resources, and wealth to newcomers and outsiders in ways that subjugate and marginalize Indigenous geographies. Indigenous knowledges and geographies and their characteristic attention to process, relationship, fluidity, context, and specificity contrast with more globally dominant developmentalist, Eurocentric, or Western concerns with property, outcome, intervention, and universalism. For geographers, the key challenges this situation presents are to be more aware of, responsive to, and supportive of cultural diversity and the ontological pluralism that Indigenous peoples bring to contemporary settings.

Over the course of the late twentieth century, Indigenous identity became a powerful, and often contested, political, legal, and cultural issue. Problematizing the term, and interrogating the assumptions which can underlie it, is one of the major challenges of Indigenous geographies. The discussion here examines the idea of Indigenous geographies, not simply in terms of the distribution or location of Indigenous peoples, but more complexly in terms of its challenges for human geography. It focuses on the experiences of Indigenous peoples in the context of colonial (predominantly European) encounters and the postcolonial nation state, but with powerful implications and relevance beyond this and into a range of other relations.

In the context of dominant motifs of contemporary geographical theory such as globalization and terror, it would be easy to relegate experiences of Indigenous peoples to the status of an interesting anachronism, unfortunate collateral damage, or a market opportunity for heritage or eco tourism. The reality, however, is that Indigenous peoples' persistent presence in the cultural landscapes of contemporary postcolonial nation states offers an important double take on the idea of indigenous geographies. These persistent geographies simultaneously require geographers to consider the ways in which historical and contemporary practices of genocide, dispossession, denial, erasure, cooptation, and management have failed to silence or remove Indigenous peoples and to acknowledge the significance of Indigenous peoples' representations of their own geographies. Inverting the dominant representations of Indigeneity by states and scholars to conform with more Indigenous centered ideas of society, culture, and environment provides a powerful Indigenous critique of many key themes in human geography. Not only do Indigenous critiques often contribute actively to postcolonial, post structuralist, feminist, and increasingly postdevelopmentalist theoretical debates, but they often also challenge the terms of such debates, and the implicit assumption of a singular core underpinning the ‘posts’ as if local diversity and specificity were reducible to a single dimension of significant difference, such as class, gender, or relationship to the so called ‘global’. Practical engagement with Indigenous rights, knowledges, and geographies inevitably challenges such reductionist assumptions and also offers valuable lessons for areas of practice, such as natural resource management, geographical education and research, politics, and the future of globalization.

The persistent presence of Indigenous geographies among the ‘a geographical’ level playing fields of globalization, therefore, offers a powerful challenge to geographers’ understanding of and contribution to contemporary social and environmental relations. For example, these diverse geographies often decenter the conventional regional narratives of geographers’ accounts of social change. Indeed, in writing of Maori geographies in Aotearoa, New Zealand, the Maori geographer Dame Evelyn Stokes noted that a Maori analysis of housing would conceptualize housing differently to urban geographers interested in service provision or ethnic segregation. Indeed, for many Indigenous groups marginalized from high cost housing markets, housing becomes a verb rather than a noun – an issue of power and privilege rather than an object for analysis.

In Anglo American geography, the dominance of Eurocentricism persists, despite the critique of scholars such as Jim Blaut and Eric Wolf. The risk of a Eurocentric focus on the colonial, the economic, and the cultural is that it reduces human diversity to a singular ‘Other,’ to be benchmarked only against the imaginary of the dominant culture(s) of white, male, heterosexual, capitalist relations. Such simplistic Eurocentricism, for example, lumps diverse Indigenous interests together as a singularity, as the ‘Indigenous other’. Yet, in an Indigenous centric framework, one necessarily starts with diversity rather than singularity. There is no singular Indigenous people, but a multiplicity of Indigenous peoples. While the colonial encounter and its consequences might provide a common experience within the framework of both Eurocentric and much postcolonial scholarship, the colonial encounter was different in every place, with every people, and in every era. However, it is neither the colonial experiences of Indigenous peoples that are not reducible to a singular representation of Indigenous. Nor are Indigenous peoples’ relationships with place, environmental knowledges, structures of governance, or their particular values reducible to a singular, all encompassing notion of an ‘Indigenous other’. To conflate this diverse cultural experience into a single category that is held against a Eurocentric singular runs the very danger of reinforcing Eurocentric power structures and colonial relationships. It risks pigeonholing Indigenous peoples’ active agency and their fluid, ever changing engagement with their contemporary worlds into naïve and offensive categories such as noble savage or primitive native.

Indigenous diversity and specificity needs to be engaged with in terms of particularities rather than a singularity. This perspective challenges how one might usefully discuss the notion of ‘Indigenous geographies’ without silencing that diversity and without implying a reductionist singularity. In our experience, that is the paradox of coming to terms with the significance of Indigenous geographies in contemporary geographic discourses. The resolution of that paradox needs to be secured through ‘developing relationships’ rather than developing theory, or at least as a basis for theory, and focusing on ‘process’ rather than outcomes. Focusing on relationship and process reorients the discourse toward the ethics of engagement and the specificity of context.

Many geographers engaged with Indigenous geographies have also engaged with Indigenous particularity. While their practice may involve engaging with relationships between specific Indigenous groups and particular administrative landscapes, or the way multiple groups are marginalized in similar ways by structures in a transnational sense, the strength of this ‘Indigenist’ work in geography has been in the framing of their work around ‘particular relationships’ rather than theoretical generalizations. This process of engagement has confronted many geographers involved with questions of ethics, methods, and practices in diverse places from Nepal to Morocco, Costa Rica to Taiwan, and Israel to Malaysia. It has cast academic and applied research into new light as when viewed from Indigenous centered positions, rather than institutionally centered Eurocentric positions, the expectations, experiences, and ambitions regarding research in Indigenous settings are often very different to those formed by dominant cultures and academic disciplines.

Much of the academic discussion of Indigenous geographies based on specificity and relationship building is framed by a series of binaries around notions of presence/absence, possession/dispossession, and rights/charity.

Presence/Absence: The Paradox of Indigenous Geographies

There is a dominant discourse that is theoretical, political, and administrative and represents and imagines Indigenous geographies in terms of absence, placing Indigenous peoples into an ahistorical past or into exotic (and distant, remote) places – anywhere but the contemporary Eurocentric here and now. At the same time, we have a discourse within Indigenous groups which offers an affirmation of ongoing engagement, active agency, and contemporary presence. It is in how the discipline engages with this challenge that we might judge the contribution geographers might make in this field. The founding work of key figures in Indigenous geographies, including Fay Gale, Jan Monk, Evelyn Stokes, Brian Murton, Elspeth Young, Peter Usher, Cole Harris, Bob Galois, Bill Jonas, Ben Wisner, Jonathan Briggs, Garth Cant, and Evelyn Peters, all contributed to research frameworks in which Indigenous people are engaged with as part of the experienced, everyday landscape. For these researchers, unlike those working in more colonial or salvage style frameworks, their Indigenous collaborators and participants were people who were present in diverse cultural settings in rural, urban, and remote locations. In that sense, these geographers could not frame research with Indigenous groups in terms of their absence or their loss of culture, language, well being, or identity. Their work accepted that people were there and worked with them as part of the cultural landscapes that they were interested in as geographers. More recently, Indigenous scholars within geography have begun challenging disciplinary conventions in their applications of geographical tools to community based research, the development of Indigenous methodologies and philosophies, and the orientation of academic accountability to Indigenous values and understanding. The development of this work in geography can be glimpsed in a series of edited collections since the early 1980s, including special issues in *Antipode*, *New Zealand Geographer*, *Geografiska Annaler B*, *Geographical Research*, and *GeoJournal*, the edited collection from the Commonwealth Geographical Bureau, and the establishment of the International Geographical Union (IGU) Indigenous peoples' Knowledges and Rights Commission under the leadership of Jay Johnston in 2006.

Geography was often implicated in the colonizing process by rendering people as absent from the landscape, emptying the landscape of Indigenous presences, and making space for colonial (and postcolonial) claims of ownership, settlement, and domination. In many of the cultural landscapes that geographers work in, the Indigenous presence is persistent and uncanny. This feeling of simultaneously being in and out of place, of belonging and being alienated, of recognizing and denying the presence of others underpins the practical challenges of negotiating the coexistence of Indigenous and settler geographies in the same landscapes. Even where it is asserted that those relationships have been settled by treaty and there is some formal acknowledgement of coexistence, relationships of denial and domination persist. For example, in the United States, where reservations were established and bounded within a long established legal tradition of accepting a level of Native American sovereignty, there has been non Native settlement into those reservations under terms prescribed by the Federal Indian Reorganization Act of 1934 and the provision for sale of reservation lands to dominant society settlers. This creates strange and fragile geographies of coexistence, where neither settler nor Indigenous society is able to deal comfortably with an almost ghostly co presence from the other.

Thus, geographies of Indigenous absence and colonial presence have been literally and metaphorically implicated in the construction of dominant metaphors of dispossession. In Australian landscapes, for example, dominant discourses imagine aboriginal people as being dispossessed by colonial occupation – even where there was no colonial settlement. It is one of the themes of Australian social justice movements to address aboriginal dispossession. Yet, in practice, aboriginal people are often very much in possession of their own country. Often there is no theft or settlement presence other than an administrative presence, and it is this institutional framework that incorporates Indigenous groups into particular sorts of relationships with the state and other bodies. Even where settlement has occurred, Indigenous groups continue to relate to and possess country in multiple ways. They continue to assert their presence and their ways of being and unsettle, challenge, and reconfigure dominant systems, structures, and knowledges. It becomes crucial, therefore, to ask in what sense have people ceased to possess their country?

Possession/Dispossession: Challenging Erasure of Indigenous Geographies

Within many societies, the Indigenous rights movement has often focused more on legal than social relationships, emphasizing the political importance of legal questions that arise within dominant state structures and the legal structures and institutions they authorize. This has meant, for example, that the legal technicalities of land tenure have often been mistaken for the realities of possession or dispossession and the realities of presence or absence. It is such inversions that have misdirected the native title debate in Australia where the courts represent a lack of continuous presence in country as a loss of connection to country, and therefore an extinguishment of Native title – even where people have continued to be engaged with their country and to embed their identity through their relationships with kin and country. Even where the act of disruption was a product of forced relocation, locked gates, frontier violence, or disease, what we see as the end product of this trope of dispossession is an erasure of Indigenous geographies in the administrative geographies of the settler state. This produces a legal imaginary, reinforced by state power, in which the state and those who are the beneficiaries of state presence no longer have to deal with Indigenous people's rights and concerns, even when there is a clear coexistence in time and space. The imaginary suggests that there are no rights attached to the Indigenous presence in the landscape because the dominant society can only understand the idea of rights in terms of the structures, systems, and institutions that are authorized within its own worldview. That is the essence of the dominance that is constructed by the dominant settler society. The continued presence and significance of Indigenous societies, governance institutions, knowledges, and practices simply does not need to be acknowledged, because the dominant society constructs a self referential claim to territorial sovereignty and self authorizing power.

These imaginaries also reinforce dominant discourses which represent Indigenous people as being possessed by, rather than in possession of, their country. It represents Indigenous relationships with country not as a property right but as a (romanticized) dependency. Such misrepresentations not only erase the capacity of the dominant system to recognize coexistence and the significance of Indigenous peoples' prior and persistent rights, but also the capacity of Indigenous people to exercise the rights and responsibilities that are constructed within the jurisdictions of customary law. Perversely, it is precisely that erasure which subsequently provides justification for the dominant society to intervene administratively as the 'protector' of Indigenous interests, or the patronizing senior partner in the process of civilization. The process of domination thus creates a dispossession and absence that the dominant society needs to possess through charity and its own self assured presence. This authorizes settler governments and institutions to define indigeneity in terms of loss and lack.

Engaging with Indigenous geographies and with lived experiences of presence, persistence, possession, and rights challenges this long tradition of defining Indigenous people in terms of absence, erasure, and dispossession. The significant features of Indigenous geographies are not adequately encompassed by focusing on the lack of the dominant language, nor people's lack of the dominant culture, their lack of numeracy or literacy, their loss of their land, or their apparent loss of their own culture. Engaging with Indigenous geographies challenges those processes and their role in constructing Indigenous people as somehow missing from modernity and postmodernity, of always being in an anachronism – of being always out of time and out of place. It challenges the discourses of authenticity which dismiss current Indigenous claims as not being able to reflect an authentic indigeneity because the things that the dominant society authorizes as authentically Indigenous are, by definition, things that can only exist in a past time and other place than the here and now. The authenticity discourse also renders people as objects of charity and sympathy rather than partners in terms of rights.

Rights/Charity: Representation of Indigenous Geographies

The assertion of Indigenous rights that belong to people as collectives as a result of their self defined Indigenous identity is a big issue in human rights discourses that foreground individual rather than collective rights. The debate between collective and individual rights and the ways that different jurisdictions mediate the rights of groups and the rights of individuals has advanced considerably, and the tortuous passage of the United Nations Declaration on the rights of Indigenous peoples has framed a strong response in favor of recognizing and responding to Indigenous rights. This fosters a shift from notions of charity based engagement being the dominant mode of states' and settler institution's intervention in Indigenous communities toward rights based engagements. In terms of policies and practices, however, it remains a very live issue in many jurisdictions. For example, reducing Indigenous state relations to what the Australian Government characterized as an agenda for 'practical reconciliation' or bridging the gap between statistical representations of health, education, employment, and poverty and so on, to bring Indigenous people to a statistical equality, the same life expectancy, and levels of morbidity and mortality, while addressing important issues of equity and justice, misses key issues raised by the rights based claims of Indigenous groups. Such approaches position the foundation of policy of Indigenous state relations in terms of the dispossession discourse, assuming also that the appropriate policy target is to reproduce the characteristics of the dominant society in the Indigenous population.

In terms of representation, those key issues include difficult questions of how people represent themselves, and how they exercise governance, including questions of social control and dissent, and what constitutes the complex geographical totalities of Indigenous peoples' geographies, including Indigenous peoples' multiple, fluid ways of knowing what are categorized as time and space, and nature and environment, in Western worldviews. This also confronts questions of power, autonomy, and sovereignty in relations within Indigenous social formations, and between Indigenous, state, and other jurisdictions. This question of power inevitably frames questions of Indigenous geographies and state policies, including responses to contested claims to territory and rights to dispose of territory and its resources. For example, in the field of resource and environmental management, states claim to be able to dispose of Indigenous territories to advance the national interest and contribute to nation building. This should be seen not only in terms of exploitation of forest, mineral, energy, water, or wildlife resources in Indigenous domains, but also in terms of policies that exploit Indigenous geographies in other ways such as nuclear waste disposal, weapons testing, and tourism based development. In many jurisdictions, state policies on such matters represent an active continuation of frontier settlement and primitive accumulation.

In the case, for example, of the politics of water, we find states insisting on national control of water resources and framing environmental, economic, and development agendas as if Indigenous rights in the water, and the cultural landscapes in which it is embedded, were inconceivable. Even concepts such as 'cultural flows' are treated as problematic and reduced from Indigenous conceptions of complex, sustainable, and just relations in cultural landscapes to engineered amounts of water into intensely 'managed' water delivery systems. Similar issues arise in relation to climate change, where one sees state responses to questions of Indigenous vulnerability in terms of the state's right (even obligation) to intervene in ways that preempt and marginalize Indigenous governance structures. Thus, states assume a right to move people if the state defines them as vulnerable to climate change risks. States assume a right to intervene, rather than constructing the state's relationship with Indigenous peoples as one of partnership and negotiation.

Identity, Scale, Place, and Culture: Lessons and Challenges from Indigenous Geographies

Indigenous geographies present the discipline with some important lessons and challenges. Many Indigenous cultures and their diverse ways of knowing themselves and their worlds offer exemplars of specific relationships with particular environmental circumstances and relationships to place that have proven themselves sustainable and viable. One of the consequences of globalization has been increasing pressure to subsume this human diversity into singular systems of environmental governance, of authoritative state administration, and of predefined or universally defined rights. Yet to lose the human diversity represented by Indigenous experiences and responses to human–environmental relations is not simply unjust and unjustifiable, but it is also to risk reducing the capacity of peoples to continue to live lives that are defined by themselves rather than defined by the state or by the multinational superstate apparatus. In the context of widely perceived environmental and social risk, there is also much to be learned from Indigenous experiences of the ontological risk that has accompanied the threats and damage to the connections and balance between cosmological elements as a result of various intrusions into, claims upon and theft, marginalization and pauperization of Indigenous geographies. The realization of diverse Indigenous geographies; the practices; and experiences of Indigenous place making; Indigenous governance; Indigenous ways of thinking, doing, and being continues to contribute a fundamental element of cultural and human diversity to the challenges of human–environmental relations.

In these discussions of Indigenous experience across the globe, the questions of power and the capacity of people to respond to power are raised again. In exercising their presence through dugong and turtle management in northern Australia, whale and seal hunting in the Canadian arctic, or reindeer herding in northern Europe, and so on, Indigenous groups' social practices confront Eurocentric societies with challenges that reveal not only fundamentally different ways of knowing the world but also the dimensions of power that are contained within dominant administrative structures. Academic and Eurocentric political discourses tend to take for granted the right of states to exercise power in relation to environmental management. They tend to assume (in direct contradiction of the evidence as understood by many Indigenous peoples) that the state is a neutral player acting in the wider national interest or public interest or human interest. Indeed, it is very easy to forget that the state is an agent of power acting on its own behalf. In Indigenous geographies, that power is often laid bare and fierce in response to Indigenous groups trying to effectively challenge that power, just by their presence, by their persistence in activities and practices that do not fit the morays and norms of the dominant culture and the dominant administrative structures. The scales at which Indigenous geographies are embodied, enacted, and governed, for example, are often in contrast to the formal jurisdictions of state power. The intimacy of environmental and social knowledge, for example, often insists on governance structures that combine responsibilities for both oversight and governance of systems and their daily use and management. In contrast, state sponsored and commercially oriented systems often separate research, use, and governance functions into separate structures and fragment, diminish, and marginalize Indigenous interests in integrated systems in the process.

Conclusion

Clearly, Indigenous geographies provide a powerful and challenging window on contemporary issues for geography as a discipline and geographers as practitioners in a variety of fields. For the discipline, engaging with the rights of Indigenous peoples to maintain their identity, maintain their autonomy, and benefit from the new United Nations Convention will reveal some of the hard edges of state administrative structures, private and other institutional frameworks that are too often taken for granted as being neutral or well intentioned. Indigenous geographies offer a window on the state and the exercise of power, which is very different from the window that is offered by more conventional economic and cultural geographies because the state is so much an element of what seeks to overwhelm Indigenous rights.

In responding to these windows on diversity and the challenges of human–environment relationships, there is a mismatch in scales between Indigenous autonomy, which is often exercised at a much smaller geographical scale than state autonomy, and globalization of environ mental governance, economic regulation, and other matters of human affairs. This mismatch risks swamping Indigenous peoples' knowledges, their notions of part nership, of relationships between equals, and of govern ance with powerful and authoritative solutions that are alien to the geographies that allow the world to be. While these matters are challenges equally to the state, to state sanctioned institutions, and to Indigenous groups, many aspects of Indigenous geographies actually challenge the identity of the state itself, and state centric conceptions of human–environment relation.

One of the important insights is that Indigenous peoples are not reducible to a simple equality with other stakeholders. In the fragmentation that often character izes reductionist representations of Indigenous identity, or the essentialist representations of identity which fail to deal with critical issues of connections among people and between people and their environments, we open tremendously challenging questions about how to understand, for example, water in rivers in more than environmental management terms and more in terms of environment and culture. So in Indigenous geographies, a cultural flow is not reducible to an amount of water. It has to be conceptualized as an intervention which cre ates, maintains, or manages connections. It reconnects things that are disconnected through periods of drought or through mismanagement of water. In different land scapes and different geographies, notions of connection will be conceptualized and managed in different ways. But, responding to this insight will involve challenging the dominant binaries that characterize state–Indigenous relations in terms of presence/absence, and possession/ dispossession. Until dominant societies can reframe their relations with both Indigenous and environmental 'oth ers' in terms that accept Indigenous presence, posses sion, and rights, then the assumption of superiority over, and therefore power to govern, will continue to con struct unjust and unsustainable human–environment relations.

In terms of geographical realities, it is not appropriate to assume that states and other Eurocentric institutions will be able to articulate a shift toward new ways of relating to human and environmental diversity. Rather, it is the persistent presence of Indigenous peoples who articulate their continuing presence in and possession of their relationships with their geographies and who insist on recognition within a rights based discourse that re spects the fragile and dynamic geographies of coexistence between Indigenous peoples and others. Even in areas of long settlement, for example, in urban areas, where Eurocentric legal tenure issues have been well and truly settled, Indigenous peoples maintain a presence and a cultural relationship that creates a persistent coexistence of an Indigenous cultural geography with the settled legal tenure geography. Developing formal recognition of that coexistence provides an opportunity to shift re lations from an assumed and enforced absence and dispossession to relations characterized by continuing presence and partnership based on rights. Focusing on presence and partnership encourages a shift to process and relationship, rather than intervention and outcomes, with respect to postcolonial relations. In such a reshaping of our shared geographical imaginaries, we open the prospect for reconfiguring the social, cultural, political, economic, and environmental responses to those geog raphies of coexistence as fundamental contemporary challenges, both to society and to the discipline of geography. This leaves us asking the question of just how the discipline may contribute to both understanding and nurturing those pluralist coexisting geographies. That is, in many ways, the challenge for both the writers and readers of this work.

See also: Indigeneity; Land Rights; Postcolonialism/ Postcolonial Geographies.

Further Reading

Berg, L. D. and Kearns, R. A. (1997). (Re)Formations: Cultural geographies of Aotearoa/New Zealand. *New Zealand Geographer* 53(2), 1 2.

Briggs, J. A. (2005). The use of Indigenous knowledge in development: Problems and challenges. *Progress in Development Studies* 5, 99 114.

Cant, G., Goodall, A. and Inns, J. (eds.) (2005). *Discourses and Silences: Indigenous Peoples, Risks and Resistance*. Christchurch: Department of Geography, University of Canterbury.

Cant, G., Overton, J. and Pawson, E. (eds.) (1993). *Indigenous Land Rights in Commonwealth Countries: Dispossession, Negotiation and Community Action*. Christchurch: Department of Geography, University of Canterbury and the Ngai Tahu Maori Trust Board for the Commonwealth Geographical Bureau.

David, S. and Wisner, B. (1984). The fourth world: A geography of Indigenous struggles. *Antipode* 16(2), 3 13.

GeoJournal (in press) Geography for and with Indigenous peoples: Indigenous geographies as challenge and invitation.

Harris, C. (2002). *Making Native Space: Colonialism, Resistance and Reserves in British Columbia*. Vacouver, BC: UBC Press.

Howitt, R. and Jackson, S. (1998). Some things do change: Indigenous rights, geographers and geography in Australia. *Australian Geographer* 29(2), 155 173.

Johnson, J. T., Cant, G., Howitt, R. and Peters, E. (eds.) (2007). *Special Issue: Indigenous peoples Knowledges and Rights Geographical Research* 45(2).

New Zealand Geographer (2005) 61(2).

Shaw, W. S., Herman, R. D. K. and Dobbs, G. R. (2006). Encountering indigeneity: Re imaging and decolonising geography. *Geografiska Annaler B* 88(3), 267 276.
Starn, O. and de la Cadena, M. (eds.) (2007). *Indigenous Experience Today*. Oxford: Berg.
Stea, D. and Wisner, B. (1984). Introduction to the Fourth World, a geography of Indigenous struggles. *Antipode* 16, 3 12.
Stokes, E. (1987). Maori geography or geography of Maoris? *New Zealand Geographer* 43, 118 123.

Relevant Websites

http://www.unl.edu
IPKRC, The Indigenous Peoples' Knowledges and Rights Commission of the International Geographical Union, Anthropology and Geography Department of University of Nebraska Lincoln.
http://www.iwgia.org
UN declaration of Indigenous rights, IWGIA.

Indigenous Health and Medicine

K. Wilson, University of Toronto, Mississauga, ON, Canada
C. Richmond, The University of Western Ontario, London, ON, Canada

Glossary

Community-Based Research It takes place in community settings and involves community members at all stages of the research program, from the design and implementation of the research projects through to dissemination of the findings. Community-based research builds on and engages in methodologies that empower and legitimize the knowledge, experience, perspectives and voices of communities, and works to make positive change within the communities involved.
Infant Mortality The number of infants who die before the age of 1 (expressed as a rate per 1000 live births).
Life Expectancy The number of years a person is expected to live from birth.
Mother Earth Term that reflects the belief of many Indigenous populations that the Earth provides all things necessary for survival and represents a female entity.
Population Pyramids Graphic representation of the age and sex composition of a population. The proportion of males and females for different age cohorts (usually in 5-year increments) are displayed as a series of horizontally stacked bars. Pyramids can be created for the total population of a country or geographic area or subsets of the total population to demonstrate how the age and sex structure of subsets of the population may differ from the total population.
Self-Determination The right for Indigenous populations to develop and protect their own lifestyles, economies, and sacred traditions. This includes the right to determine what their future and the destiny of their land shall be.

Introduction

The term 'Indigenous peoples' is generally used to refer to those people who descend from the original populations who inhabited a country or region before the period of colonization and the creation of modern states/territories. Oftentimes the term 'Indigenous' is used interchangeably with Aboriginal, First Peoples, or Native, and the term used often reflects the geographic positioning and language of the Indigenous group. For instance in Canada, the Indigenous population is legally defined as 'Aboriginal' peoples of Canada, and they are recognized as three groups: First Nations, Métis, and Inuit. In Australia, the Indigenous population is defined as the Aboriginal and Torres Strait Islanders, while in New Zealand they are defined as Maori. Within Scandinavian countries, the term Sami is often used to refer to Indigenous peoples. Each of these definitions is determined in acute ways by the geographic regions within which these Indigenous peoples have lived in historic and contemporary times.

It is estimated that Indigenous peoples number over 300 million representing 5% of the world's total population. Indigenous peoples can be found in over 70 countries throughout the world ranging from the circumpolar north, South Central, and North America, across regions of Africa, Asia and also Australia, and the Pacific regions. Overall, the world's Indigenous population is much younger, and growing at a much faster rate than the non Indigenous population. In some regions throughout the world, between 30% and 50% of the Indigenous population is less than 15 years of age. Population pyramids for most Indigenous populations are characterized by a very broad base that narrows quite sharply toward the top due to very high levels of fertility and mortality.

Indigenous peoples hold strong ties to their ancestral lands and generally share a worldview characterized by harmonious relationships with the natural environment, and in fact, various creation stories of Indigenous peoples link them to both natural and spiritual worlds, often symbolically understood as 'mother earth.' Indigenous peoples represent a diversity of languages, cultures, and traditions, and were historically governed by sophisticated cultural, economic, social, and political institutions. Among the Kwakiutl people of coastal British Columbia, for instance, the potlatch (i.e., the recognition of social status though commodity exchange) formed the institution through which Kwakiutl social structure was most vividly displayed, reinforced, and modified. The potlatch maintained an effective management of human, environmental, and cultural resources, acting as a fully integrated monitoring and public accountability system.

Precisely because of the important connection between Indigenous people and the land, geographers have long studied the spatial processes tying Indigenous peoples to their environments. However, it is only in the past few decades that health geographers have turned their interests to issues of Indigenous peoples.

Indigenous Concepts of Health

Concepts of health among Indigenous societies around the globe place significant importance on the larger system within which the individual lives, and incorporate concepts including holism, balance, and interconnected ness. In New Zealand, the Maori Public Health Action Plan summarizes three Maori models of health: *Te Pae Mahutonga* (Southern Star Cross constellation), *Whare Tapa Whā* (health as a house), and *Te Wheke* (the octopus). Of these three models, *Whare Tapa Whā* provides a multidimensional concept of Maori health and well being that extends beyond physical health to recognize the dependence of health on a balance of four main di mensions – *taha wairua* (the spiritual side); *taha hinengaro* (thoughts and feelings); *taha tinana* (the physical side); and *taha whanau* (family). Represented by four walls of a house, this metaphor holds that if one of these walls should fall, the house will collapse. From Australia, the National Aboriginal Health Strategy Working Party de fines health as not just the physical well being of the individual, but the social, emotional, and cultural well being of the whole community. This definition in corporates a 'whole life' view, including the cyclical concept of life–death–life. The concept of holism is central to ideas of health and wellness among native Hawaiians, as are dimensions of spirituality and culture. Traditional native Hawaiian concepts of health en compass cultural values of *lokahi* (balance), *pono* (doing the right thing), and *kokua* (working without expecting reward), and aim to strengthen and protect the family (extended family) or *ohara* and larger community, thereby conceptualizing health not as a personal burden, but one that is shared by the whole community.

Among native people of North America, the concept of health is explained through the medicine wheel, which is a wide conceptual understanding of life and the interrelatedness of all its functions (i.e., life, time, seasons, cosmology, birth, and womb). The medicine wheel rep resents a circle divided into four sections that represent the four directions: north, east, south, and west. Ac cording to beliefs, all four elements of life – the physical, emotional, mental, and spiritual – are represented in the four directions of the medicine wheel. If one element within the wheel is neglected or receives too much at tention, then health suffers in all four areas. One of the main teachings of the medicine wheel is that balance between all four elements is essential for maintaining and supporting good health. Balance extends beyond the in dividual such that health and healing also requires an individual to live in harmony with others, their com munity, the environment, and the spirit worlds. Among Inuit peoples of Northern Canada, these concepts are captured within *inuuqatigiittiarniq*, which depends on the balance and harmony of economic, cultural, environ mental, and biological factors. A careful balance of fac tors manifests as *inummarik* – a most genuine person – in a continuous lifelong process of developing interaction with people and animals, community, and environment.

Current Patterns of Health and Well-Being

Through the combined processes of colonialism and neocolonialism, the health and well being of Indigenous populations have fallen 'out of balance' as they have been displaced from their ancestral lands. Though Indigenous peoples are a diverse group, they share a common history of oppression and marginalization as early colonial policies aimed at assimilating Indigenous peoples into the dom inant (i.e., colonizing) population. With colonialism came loss of ancestral land for Indigenous peoples, and forced migration to isolated regions, wherein the truncated lands were often insufficient to support a familiar way of life. For example, Canada's First Nations live on reserved lands that total 0.20% of Canada's total landmass. The loss of land, as well as numerous policies aimed at assimilation (e.g., residential schools, through which the Canadian and Australian governments forcibly removed children from their communities), served to disrupt the sociocultural and economic activities that had tied Indigenous peoples so intimately to their physical environments, thereby upset ting the balance upon which good health rests. These assimilationist policies have contributed to the ongoing legacy of social and health inequalities that we witness among Indigenous peoples today.

A clear distinction can be drawn between biomedical approaches to health and Indigenous approaches to health and well being. Biomedical models constitute linear ap proaches to knowledge in which disease is seen as a discrete (isolated) event that is treatable by the intervention of healthcare. In contrast, Indigenous perspectives of health, as earlier reviewed, constitute a much more holistic approach that emphasizes the interconnectedness of people, com munity, their spirit worlds, and perhaps most importantly, their physical environments. Despite this difference, bio medical indicators are important for highlighting the stark inequalities in health between Indigenous and non In digenous peoples.

Although it is often difficult to collect data on the health of Indigenous populations, available biomedical indicators demonstrate that Indigenous peoples suffer from much higher levels of morbidity and mortality than non In digenous populations throughout the world. While differ ences in life expectancy between Indigenous and non Indigenous populations have decreased over time, gaps still remain with variations across the world. In the United States the gap in life expectancy is 4–5 years, in Canada 5–8 years, in South/Central America 4 years, in New Zealand 8 years, and in Australia 17 years. The gaps in rates of infant mortality are especially large for Indigenous peoples living in Australia and South/Central America.

Indigenous populations have undergone an 'epide miologic transition', characterized by the decline in in cidence of infectious diseases, followed by the rise of chronic, noncommunicable diseases, accidents, and violence. Chronic conditions such as diabetes, cardiovascular disease, and cancer are also increasing at a much higher rate than among the non Indigenous population. Today, Indigenous peoples experience the kinds of health problems most closely associated with poverty, and also suffer from problems linked to a marginal position within their respective nation states. Levels of education and workforce participation of Indigenous peoples lags far behind that of their non Indigenous counterpart, and the paradox for many Indigenous peoples is that their home communities often offer little employment opportunity, thereby reflecting elevated rates of under and unemployment. In Canada, less than 50% of the Indigenous population has completed high school (in comparison to almost 70% of the total population) while in Australia less than 20% of the Indigenous population has completed high school. Given the strong connection between level of education and employment, it is not surprising that unemployment rates are three times higher among the Indigenous populations living in Australia and Canada as compared to their non Indigenous counterparts.

Inadequate and insufficient housing is also a serious problem across Indigenous populations, as remote and rural communities lack basic sanitary infrastructure. Crowded housing, among other health determinants, is responsible for numerous health problems including high rates of infectious diseases such as otitis media and tuberculosis. Among First Nations in Canada, for instance, the incidence rate of tuberculosis is 10 times higher than the national average. Indigenous peoples are also exposed to high rates of other infectious diseases such as pertussis, meningitis, hepatitis, pneumonia, sexually transmitted diseases, and HIV/AIDS, all of which are linked to poverty, under and unemployment, unstable housing, homelessness, sexual/physical abuse, and a concomitant lack of self worth.

Rates of morbidity and mortality due to suicide, injury, accidents, and violence are also much higher in Indigenous populations. Current patterns of health and social suffering reflect the combined effects of colonial oppression, systemic racism, and discrimination as well as unequal access to human, social, and environmental resources. Such inequalities in health and well being have been attributed to the ongoing legacies of colonialism. Indeed, research has shown that among Indigenous populations that have experienced lower rates of assimilation and acculturation, rates of chronic conditions, and the presence of risk factors (e.g., smoking, drinking, and solvent abuse) are also lower.

Indigenous Medicines and Healthcare

There are significant differences in approaches to medicine embraced by Indigenous peoples and those who identify with biomedical models of healthcare. Traditional medicine refers to a diverse range of health and healing practices used by Indigenous peoples that incorporate plants, animals, and the spiritual realm for the diagnosis and treatment of illness, or maintenance of health and well being. A key distinction between traditional medicine and biomedicine is that traditional medicine embraces a holistic approach for the promotion of physical, emotional, mental, and spiritual well being. Unlike biomedicine, traditional medicine does not treat illness in isolation, but rather seeks to restore balance.

Research conducted in various parts of Africa and India demonstrates the use of plants and herbal medicines to treat such afflictions as gastrointestinal disorders, malaria, and weak blood. Food is also a very important component of Indigenous health and medicine. The Navajo rely on four sacred plants (corn, beans, squash, and tobacco) for both food and medicinal purposes. The Inuit of Baffin Island hold the belief that health is achieved when the body is warm, and that sickness occurs when the body becomes cold. Such belief holds that one's body can be 'warmed' through the consumption of country food, in particular seal. The blood of the seal is thought to be warm and therefore it is consumed not just for dietary reasons, but also for treating illnesses. The use of food and plants as medicines for maintaining a balance between hot and cold bodily states has been documented among Indigenous cultures in Latin America, Trinidad and Tobago, the American Southwest, Philippines, Haiti, and Puerto Rico. Today, the impact of traditional medicine reaches far beyond Indigenous populations. In many countries throughout the world, non Indigenous people are adopting the use of traditional medicine (often referred to as alternative or complementary medicine), and in fact, Indigenous peoples often combine conventional healthcare with traditional healing practices.

Access to Contemporary and Traditional Healing Practices

Access to traditional and conventional healthcare is an especially salient issue for Indigenous populations and represents an interesting research topic for health geographers. From a geographic perspective, physical access to conventional health services such as family doctors, emergency care, and specialized technologies is limited for Indigenous peoples living in remote and isolated areas that are a far distance from major urban centers where these services tend to be concentrated. However, even when healthcare services are geographically accessible, other barriers can limit access for Indigenous peoples including language, cultural difference, and discrimination.

Access to traditional healing systems is particularly difficult for Indigenous populations living in an urban context, where they may be far removed from the healing traditions of their families and communities. Decreased access to traditional healing systems is, however, not limited only to Indigenous people in urban areas. In many Indigenous communities, healing traditions have been lost or severely hampered due to colonial policies that forbade the practice of traditional medicine and ceremonies. Colonial governments saw it necessary to prohibit Indigenous peoples from using and practicing traditional medicines and other activities that ran counter to the assimilation effort. Perhaps, the most extreme colonial measure taken resulted in the forced attendance at residential/boarding schools by generations of Indigenous children in Canada, the United States, Australia, and New Zealand, wherein they were deprived of their families, languages, and cultural traditions.

While Indigenous populations experience lower levels of access to healthcare than non Indigenous populations, in many instances access to care can be improved through the process of self determination in healthcare. And in fact, there are a growing number of cases wherein Indigenous communities are implementing traditional healing centers in their own communities, as well as in major metropolitan areas. Such centers provide Indigenous peoples with access to a range of healing practices and services that are geographically and culturally accessible, including traditional healers, doctors, nurses, dieticians and nutritionists, and counselors.

Geography and the Health of Indigenous Peoples

An important issue that needs to be addressed within health geography is the lack of researchers working in the area of Indigenous health – this is so despite changes in the discipline that encourage alternative views of health and place. The relationship between health and place has culturally specific dimensions, yet these tend to be overlooked, particularly in the context of Indigenous health. Given what we know about Indigenous concepts of health, research around the connections between health and place is an important avenue through which health geographers may be able to contribute to a decline in the health and social inequalities endured by Indigenous populations. In **Box 1**, we identify important research directions for geographers. While it is beyond the scope of this article to discuss each component in the list we feel that it is important to acknowledge the potential links between land and health, as land remains at the heart of Indigenous cultures and crosscuts the research directions we have identified.

Given its level of significance in all aspects of Indigenous culture, the land is an important issue to consider in conjunction with Indigenous health. Historically,

Box 1 Research directions for health geographers

- Indigenous peoples' sense of place and culture
- Economic opportunities/limitations for Indigenous peoples
- Indigenous peoples' social institutions
- Healthcare access for Indigenous peoples (traditional and contemporary)
- Access to environmental resources
- Changing social environments of Indigenous communities
- Impacts of environmental degradation on Indigenous communities
- Incidence of chronic/infectious disease among Indigenous peoples
- Indigenous concepts of health
- Traditional Indigenous knowledge
- Everyday living in Indigenous communities
- Changing way of life among Indigenous populations

the land has been a very important component of Indigenous culture. Before colonization, most Indigenous societies could be described as subsistence cultures. Whether by hunting, trapping, planting crops, fishing, or gathering, the diet and daily nourishment of these groups were provided by the land. Indigenous cultures vary widely in their beliefs, traditions, and approaches to health, however, a common theme is a belief in the importance of the land and a life based on stewardship and harmony with the earth. Indigenous people contend that the relationship they have with the land shapes all aspects of their lives: the cultural, spiritual, emotional, physical, and social lives of individuals and communities. In the contemporary context, many Indigenous peoples can no longer live off the land as they once did; however, they assert that access to the land remains an essential component of their culture. Attachment to the land is health promoting, as demonstrated in the ability to practice a traditional way of life, to hunt and fish, to collect berries and fruits, to gather roots and herbs for medicines, and perhaps most importantly, to pass this knowledge and culture to the next generation.

Because of their close ties to the land, environmental degradation can have devastating effects on the well being of Indigenous peoples. Contamination from industrial development and urban growth and sprawl disrupts wildlife habitats and reduces the purity of traditional foods and medicines, all of which impact upon the physical and spiritual health of Indigenous peoples. Numerous studies conducted in circumpolar regions reveal that due to anthropogenic activities, environmental contaminants (e.g., mercury and PCBs) are entering the traditional food systems (fish, game, and plants) of Indigenous populations. Research conducted among populations in warmer climates also reveals that Indigenous peoples are exposed to contaminants through the food chain. Indigenous populations are more likely to be exposed to bioaccu mulating contaminants than non Indigenous populations due to their high consumption of traditional foods (e.g., fish and wildlife). Since the consumption of traditional foods is often an economic necessity, contamination raises issues that extend beyond nutritional and health problems. Due to the poor economic conditions that exist among many Indigenous populations, the consumption of tradi tional foods is an inexpensive and ideal alternative to store bought food. Therefore, within the contemporary context, contamination is not just a physical health issue but also an economic one.

Through colonial processes, Indigenous people have experienced physical displacement from their traditional territories. Culture stress is usually visible in individuals or groups that have experienced massive or forced change in their lives and can manifest itself in both physical and psychological health problems. Hearings from the Royal Commission on Aboriginal Peoples in Canada illustrated that aboriginal peoples forced to relocate experienced psychological health problems, ranging from homesickness to severe depression; these conditions were brought upon by the mourning for lost homelands, culture, and traditional ways of life. Loss of land is also associated with the destruction of traditional ways of life and a lack of access to traditional foods and medicines. In many cases, Indigenous peoples are no longer able to partake in the hunting and gathering of traditional foods thus becoming dependent upon store bought foods. The combined effects of a decrease in hunting and the consumption of traditional foods (and increased consumption of carbohydrates and refined sugars) have resulted in cardiovascular disease, diabetes, and an epidemic of obesity, particularly among In digenous populations in North America.

Decolonizing Methodologies and Community-Based Approaches to Understanding Health

Though geographers have only recently begun to explore Indigenous health issues, there is a long history of research on Indigenous peoples. In the past few decades however, Indigenous communities have raised significant critiques about the way research has been done in their com munities, highlighting significant moral and ethical con cerns about the motivations and implications of such research. Many purport that such approaches serve to perpetuate the colonial agenda of the nation state, which seeks to control and regulate the well being of Indigenous peoples. Indeed, Indigenous communities often speak about being 'researched to death', as they participate in numerous studies, yet see little benefit from the numerous reports, theses, and journal articles that are produced. An important paradox emerges: Indigenous peoples are among the most studied population in the world, yet they continue to bear a disproportionate burden of health and social disparity. Indeed, the history of research from many Indigenous perspectives is so deeply embedded in colon ization that it has been regarded as a tool only of colon ization and not as a potential tool for self determination, development, or autonomy for Indigenous communities.

In the past few decades, there has been a strong movement from within Indigenous populations to engage with, and in come cases, take full control over research done in their communities. Indigenous researchers have been active in seeking ways to dispel such 'colonizing' methods, and focus efforts instead on building and en gaging in methodologies that empower and legitimize the knowledge, experience, perspectives, and voices of In digenous peoples and their communities. This approach has been labeled by some as 'indigenist', as it is formed around three principles of resistance, political integrity, and the privileging of Indigenous voices. Although In digenous approaches to knowledge and learning may challenge Western orthodoxy, there is growing evidence of the value of investing in specific research projects and in capacity building for researchers of Indigenous origin. In fact, in 2002, Australia, Canada, and New Zealand signed a tripartite memorandum of understanding on health research for Indigenous health. All three countries fund research centrally through contestable grant rounds that seeks to develop sectors of Indigenous researchers alongside Western oriented researchers.

Indigenous Health Research: The Way Forward for Geographers

In terms of their health and social conditions, Indigenous populations continue to fare among the worst in the world. Such inequalities undoubtedly reflect the colonial legacy that continues to shadow Indigenous peoples' health. Indeed, the most consequential impact of coloni alism is the broad disconnect between Indigenous peoples and their ancestral place in the world. Perhaps the most pivotal determinant of Indigenous health is place. In the years ahead, geographers can therefore play a central role in exploring some of the key issues explored in this art icle, as well as others that are yet to be explored (see **Box 1**). Ultimately, improving quality of life among Indigenous populations requires research programs and health policy that are community based, and informed by holistic models that recognize the multiple, interacting dimen sions of Indigenous people's health.

See also: Complementary and Alternative Medicine; Environmental Health; Healthcare Accessibility; Health Geography; Health Inequalities; Medical Geography.

Further Reading

Alderete, E. (1999). *The Health of Indigenous Peoples*. Geneva: World Health Organization.

Bartlett, J. G. (2003). Involuntary cultural change, stress phenomenon and aboriginal health status. *Canadian Journal of Public Health* 94, 165 167.

Berry, J. W. (1990). Acculturation and adaptation: Healthy consequences of culture contact among circumpolar peoples. *Arctic Medical Research* 49, 142 150.

Buckley, P. (1997). *Indigenous Peoples' Health: Issues and Care in Rural and Remote Areas of the World. An Annotated Bibliography with International Coverage*. Bendigo: Australian Rural Health Research Institute.

Couzos, S., Lea, T., Murray, R. and Culbong, M. (2005). 'We are not just participants We are in charge': The NACCHO ear trial and the process for aboriginal community controlled health research. *Ethnicity and Health* 10, 91 111.

Cunningham, C., Reading, J. and Eades, S. (2003). Health research and Indigenous peoples. *BMJ* 327, 445 447.

Durie, M. (2004). Understanding health and illness: Research at the interface between science and Indigenous knowledge. *International Journal of Epidemiology* 33, 1138 1143.

Hogg, R. S. (1992). Indigenous mortality: Placing Australian aboriginal mortality within a broader context. *Social Science and Medicine* 35, 335 346.

Smallwood, G., White, C. and Kotiw, M. (1997). The relevance of human rights to health status in Australian aboriginal and Torres Strait islander communities. *Health and Human Rights* 2, 127 136.

Smith, L. T. (1999). *Decolonizing Methodologies: Research and Indigenous Peoples*. Dunedin: University of Otago Press.

Van Oostdam, J., Donaldson, S. G., Feeley, M. *et al*. (2005). Human health implications of environmental contaminants in Arctic Canada: A review. *Science of the Total Environment* 351 352, 165 246.

Waldram, J. (2004). *Revenge of the Windigo: Construction of the Mind and Mental Health of North American Aboriginal Peoples*. Toronto: University of Toronto Press.

Zechetmayr, M. (1997). Native Americans: A neglected health care crisis and a solution. *Journal of Health and Social Policy* 9, 29 47.

Relevant Websites

http://www.healthinfonet.ecu.edu.au/
Australian Indigenous Health web resource.

http://www.maorihealth.govt.nz/
Maori Indigenous Health web resource.

http://www.naho.ca/english/
National Aboriginal Health Organization.

http://www.who.int/hhr/activities/indigenous/en/
WHO Health and Human Rights of Indigenous People.

Indigenous Knowledges

J. Mistry, University of London, Egham, UK

Glossary

Conceptual Framework A theoretical model comprising a set of interrelated concepts or abstract ideas.
Constructivism An approach based on the idea that knowledge is constructed by the knower based on mental activity.
Equilibrium Paradigm Within ecology it is the idea of the 'balance of nature' and stability, which, within an ecosystem, would be internally driven.
Ethnodevelopment This approach to development aims to build on the positive qualities of indigenous cultures and societies, including their sense of ethnic identity, close attachment to ancestral land, and capacity to mobilize resources for shared goals.
Nonequilibrium Paradigm Within ecology this is the idea that ecosystems are driven by external factors such as climate, and as such are constantly responding to changing conditions.
Participatory Action Research The emphasis of participatory action research is on generating knowledge from the perspective of those being researched, not from the perspective of the researcher. It entails collective and self-reflective enquiry with the aim to make some positive change.
Positivism An approach that emphasizes observable facts and scientific knowledge.
Reductionism An approach that asserts that the nature of complex things can always be reduced to (explained by) simpler or more fundamental things.

Introduction: Defining Indigenous Knowledge

Interest in indigenous knowledge has been growing rapidly in recent years. Discourses on development and environment extol the benefits of incorporating indigenous knowledge into policy development, as well as on the ground interventions. There are greater numbers of reports in the popular media of indigenous peoples and their specialized knowledge of their localities, particularly in the face of changing environmental conditions. Whether it be the Inuit people of Canada's High Artic or the San people of Botswana, these microcosmic narratives of the intimate knowledge of these groups on various aspects of natural resource use and management are increasingly portrayed as the possible answers to global environmental problems. This is in contrast to modern Western detachment from nature which is seen as the main cause of these problems.

However, what is indigenous knowledge? This is in no way clear. Various terms are used interchangeably and include 'indigenous knowledge', 'indigenous technical knowledge', 'ethnoecology', 'local knowledge', 'folk knowledge', traditional knowledge', and 'traditional environmental knowledge or ecological knowledge'. Although these all share a common meaning and understanding which can be useful for a mutual space of engagement, there is a fundamental problem in that we are assuming that all people who hold indigenous knowledge are indigenous people. However, defining 'indigenous' is a highly contested issue and the word comes with a multiplex of cultural, moral, and political connotations. For example, people have used criteria such as prior occupancy of land, length of occupancy, and capacity to remain unchanged to identify themselves as indigenous in order to establish rights and to protect their interests. Nongovernmental organizations (NGOs) including Survival International and Cultural Survival have been established to support and lobby for these indigenous groups. However, at the same time, many governments claim that there is nothing indigenous about groups of people whom they view as another minority group under their jurisdiction. With such conflicts over 'indigenousness' it would seem better to find another word to describe a particular kind of knowledge. Of the terms mentioned above, 'local' sounds more neutral but it fails to indicate key qualitative differences in the character of the knowledge usually alluded to, while 'traditional' implicates something 'out of its time' with long term cultural stasis. Debate on the term 'indigenous' will continue, yet it is useful to have some commonly asserted characteristics of indigenous knowledge:

1. Local – indigenous knowledge is context specific in that it has roots in a particular place and in the experiences of the people that live in that location. Transferring that knowledge to another location quite literally makes it meaningless.
2. Oral transmission or through imitation and demonstration – whether it be through stories, myths, songs, or through accompanying other people and observing and learning, indigenous knowledge is rarely written down. In fact, writing it down can change some of the fundamental properties of the knowledge.

3. Adaptive capacity – over time and through everyday life experiences, indigenous knowledge is adapted through repetition, learning, experimentation, and adoption of novel solutions. It is therefore not static but constantly changing.
4. Social memory – indigenous knowledge is shared to a much greater extent compared to other forms of knowledge. This shared understanding and knowledge is important as it helps to establish a long term communal understanding of people's environment and the transmission of pertinent experience. However, the distribution of this knowledge can still be socially differentiated, for example, by age and gender, and preserved in the memories of particular individuals by virtue of their spiritual or political status.
5. Holistic – indigenous knowledge is situated within numerous interlinked facets of people's lives. It is therefore difficult and impractical to separate, for example, the cultural from the ecological, or the rational from the nonrational.

Now that we have a basic checklist of some of the characteristics of indigenous knowledge, we can look at the position of indigenous knowledge within the development field over time.

The Place of Indigenous Knowledge in Development

Up until the mid 1960s, indigenous knowledge was commonly viewed as something backward, nonsensical, superstitious, and un modern. These 'traditional' ideas were treated as something to be undermined and overcome, and their legitimacy was rarely questioned. In many developing countries, development 'experts' and organizations were brought in to deal with natural resource management and extraction issues, yet the conceptual fameworks they brought with them were developed back in their own worlds; a far distance from the realities of the local situations and within a positivist and reductionist framework. In some cases, these scientific models were considered superior to indigenous knowledge, whereas in other cases indigenous knowledge was just not acknowledged to be worth noting. This labeling of indigenous knowledge as 'unscientific' was reinforced by the perception that indigenous peoples often used inappropriate and damaging land use and resource management practices.

A case in point has been widespread belief that in many semiarid lands, wide scale and irreversible land degradation has taken place as a result of overuse of natural resources. Local pastoralists have invariably been regarded as the culprits, and development efforts have been aimed at changing their management strategies, and ultimately their way of life. Little effort at understanding how these pastoralists may have used their knowledge to manage their resources took place. However, recent evidence indicates that in fact many of these semiarid ecosystems are inherently dynamic in response to rainfall, and that so called 'land degradation' may in fact be a temporary transition stage governed by the variable rainfall, as well as other ecosystem 'disturbances', including fire. It has been found that pastoralists living in these environments know exactly how and when to tap different resources in a sustainable and dynamic way through a long term understanding of environmental change.

The process of reversing the marginalization of indigenous knowledge began in earnest in the 1960s. This arose out of the Western environmental movements of the day which were looking for symbolic representations of environmental 'wisdom' to further promote the already decreasing support for modern science and technology. These movements coincided and arose from the publication of books such as Rachel Carson's *Silent Spring*. Indigenous peoples and the romantic notion that they lived in harmony with nature through their specialized knowledge were viewed as the pinnacle principles of what modern society should strive for, although the associated realities of these people, such as poverty, were rarely discussed. Nevertheless, out of this global acknowledgment of the existence of indigenous knowledge emerged the vehicle with which development workers could make indigenous knowledge more acceptable to technocrats and managers. During this same period, new buzz words such as 'participation', 'bottom up,' and 'empowerment' became popular within the development field and the important role of indigenous knowledge began to be disseminated as part of this rhetoric.

However, the way in which indigenous knowledge has been used by Western science in development is not without its critics. When Western trained scientific managers finally began to acknowledge the worth of indigenous knowledge, it was still not considered on par or as 'real' as Western scientific knowledge. As a result, there was little or no legal protection for this knowledge through copyrights or patents. The case in point has been the numerous patents owned by pharmaceutical companies of medicinal plant properties long disseminated through indigenous knowledge.

In addition, Western science has tried to fit indigenous knowledge into boxes within ordered conceptual frameworks that fail to recognize the inherent contextualization of indigenous knowledge. By ignoring or undervaluing the circumstances in which indigenous knowledge is constructed, you take away one of the most important characteristics of indigenous knowledge, namely that it is unique to a locality and group. Even just placing a specific knowledge system under the umbrella term 'indigenous knowledge' removes the context in which that particular knowledge system was established.

Indigenous knowledge is inherently complex – it is embedded within the ecological, cultural, and spiritual realities of indigenous peoples. Trying to put boundaries on this knowledge – namely by using Western scientific knowledge based criteria – is wrought with difficulties and cannot truly represent the intricate nature of that know ledge. In addition, comparing indigenous knowledge with Western scientific based categories may undervalue or ignore certain aspects of indigenous knowledge that do not or cannot have an equivalent Western category.

Therefore, although indigenous knowledge is a major concept within the development discourse, once appropri ated by development experts, some may view it as de contextualized knowledge that has been broken down into parts that can be fitted into Western scientific analytical paradigms. As with other facets of development theory and practice, it has problems of looking for global, general so lutions rather than focusing on local, specific experiences.

Recording Indigenous Knowledge

Realizing the important contribution indigenous know ledge could have to contemporary natural resource management issues, researchers and development work ers have been documenting and recording indigenous knowledge with the aim to produce a bank of knowledge which could be incorporated into development projects. Codifying and documenting indigenous knowledge has involved recording ways in which people use and manage natural resources with little contextual analysis, and presenting results in the form of a 'blueprint' of in digenous knowledge which can be packaged and trans ferred to other localities. Ethnobotanical studies are a good example and the recording of so called 'useful' plants. Many studies did not address the definition of 'useful', the context in which plants were useful and to whom these plants were useful.

However, as has been mentioned before, indigenous knowledge is normally transmitted orally or through imitation and demonstration, and is deeply implicated in people's everyday lives. So can it be recorded through writing, or is this a contradiction in terms? By making it 'portable' are we not falling into the same traps that Western science is criticized for – namely of being nonspecific and unconscious to local needs? Also, we again come back to the question of what is 'local' – if we take knowledge from one group of people and then take this to another group, how will this latter group use the knowledge, considering that they themselves will have their own indigenous knowledge? The holistic nature of indigenous knowledge also means that trying to isolate pieces of knowledge, for example, knowledge of different species, knowledge of different practices, fails to capture the entirety of the situation in which activities occur.

Another important point to consider when recording indigenous knowledge is that it is adaptive over space and time. It is the amalgamation of experiences that occur in different geographical locations within varying points in time. As such, there are many difficulties in applying in digenous knowledge recorded in one place to another, or even the same place in another year. Therefore, when recording indigenous knowledge, it is vital to recognize its constantly interactive and changing nature, as well as the individual and collective skills and practices that produced the knowledge. However, it is also imperative to remember that writing and documenting indigenous knowledge will to some degree transform it and that it will never repre sent the true reality in which it is constructed.

The conclusion we come to is that if the points out lined above are not carefully thought out when recording indigenous knowledge, it leads to the overall de con textualization of knowledge. The subsequent problem of this de contextualization has been the appropriation of indigenous knowledge by nonlocal agents. This has come in the form of 'top down' development which is generally ineffective and contradicts the participatory and em powering aspirations which an indigenous knowledge approach expounds. Further, if indigenous knowledge loses its agency and efficacy through this 'top down' development, indigenous peoples will continue to be regarded as backward and in need of aid by development bodies, ensuring little future political change.

Another way in which indigenous knowledge has been appropriated is through legal means. As the world knows, indigenous people have sophisticated knowledge of a vast number of plant and animal species, their interrelated ecology and their management. This biodiversity in cludes undomesticated, semidomesticated (wild species promoted by human management), and domesticated varieties. However, there is little protection for indigen ous peoples against 'bioprospecting'. For example, in the international Convention for Biological Diversity (1992), rights over resources and associated knowledge are at tributed to nation states, not individuals, communities, or indigenous groups. In addition, within the Convention, intellectual property rights are provided as the only means for 'equitable sharing' and protection of resources. However, intellectual property rights are problematic for indigenous knowledge systems for a number of reasons.

Firstly, they are intended to benefit society by grant ing rights to 'natural' and 'juridical' persons or 'creative individuals', not collectives like indigenous groups. In tellectual property rights can only protect information that results from a specific act of discovery. Indigenous knowledge is an accumulation of generations of experi ences and communication through culturally unique means, is freely and communally shared, and is in the public domain, so cannot be protected. Indigenous sys tems have non Western forms of ownership, tenure, and access that cannot be attributed to a single person or body. Intellectual property rights stimulate commercialization and distribution, which is the antithesis of indigenous knowledge which is communal and free, and they only really recognize the market economic value of knowledge but fail to recognize the cultural, spiritual, and local economic values. Intellectual property rights are expensive, complicated, and time consuming to obtain and defend, but subject to manipulation by those with economic or political power. All these reasons indicate that at present, indigenous knowledge is unprotected once taken out of the system in which it was perceived, and that in fact, intellectual property rights have become a code for the unethical and unsustainable exploitation of local communities and their resources.

Indigenous Knowledge in Development and Conservation Policy

As has been mentioned above, Western science has been increasingly recognizing the importance of indigenous knowledge in areas such as natural resource management, health, nutrition, education and community and social organization. Together with the shift from top down management to participatory, community based management, indigenous knowledge is becoming central to development projects. However, the current popularity of ethnodevelopment, which embraces a wide set of understandings and divergent views about how to make development sensitive to indigenous people's needs, also reflects donor interests in enhancing indigenous social capital under neoliberal visions of development. Many ethnodevelopment programs aim to take indigenous knowledge from the local level to the large scale dissemination. So, for example, the World Bank's 'Indigenous Knowledge for Development' program, United Nations Educational, Scientific and Cultural Organization (UNESCO)'s 'Best Practice of Indigenous Knowledge' and United Nations Development Program (UNDP)'s 'Indigenous Knowledge Program' initiatives all seek to transfer indigenous knowledge for adaptation to other contexts. 'Best practice' that can be moved from one locale to another is sought and in this way, indigenous knowledge is only valued if it has more general importance.

In addition, ethnodevelopment or development with identity responds to the demands of indigenous social movements and key advocates working within NGOs, government, and multilateral institutions. Many indigenous social movements have used the dichotomy between Western science based knowledge and indigenous knowledge as part of a political platform in order to gain control over how their knowledge is represented and valued in development. Some indigenous leaders, for example, have emphasized how indigenous knowledge has global significance and cannot just be seen as empirical and local. As such, indigenous activists are claiming that their knowledge speaks both to, and beyond, their reality, therefore advocating a governing position where indigenous knowledge is central and 'normal', with general rather than empirical development application. However, as has been discussed above, most of the emerging academic literature on indigenous knowledge links it to a specific territory and/or a particular group. Within the development context, policy understandings of indigenous knowledge have been successfully shaped by indigenous movements' emphases on modes of thinking. Such approaches focus on the diverse ways of knowing and processes of learning, rather than on the acquisition of specific knowledge bases themselves. Indigenous knowledge is thus defined in social cultural, relational terms by indigenous movements engaged in political struggle, rather than by spatial parameters that fix particular knowledges in specific places.

Some of the conservation and natural resource management practices based on indigenous knowledge, including monitoring resource abundance and change, total protection of certain species, protection of vulnerable life history stages, protection of specific habitats, and temporal restrictions of harvest, are also practiced within conventional Western science based management systems. However, there are other practices based on indigenous knowledge that have been largely abandoned by conventional resource management systems. These include: multiple species management, for example, through integrated farming and cultivation systems; resource rotation, for example, nomadic herding strategies; and succession management, exemplified by the shifting cultivation system that make multiple use of cleared and regenerating vegetation. In addition, there are some management practices based on indigenous knowledge that are seldom found in conventional resource management. These include management of landscape patchiness, for example, through the use of management tools such as fire watershed based management managing ecological processes at multiple scales, for example, having different sets of practices and rules for larger scale and smaller scale resources; responding to and managing pulses and surprises through reserve areas such as sacred groves; nurturing sources of ecosystem renewal such as fire and forest gap creation.

Indigenous Knowledge and Emerging Approaches to Natural Resource Management

Within the sphere of conservation and natural resource management, theories and approaches to management and practice largely focused on single issues or resources and were based on the equilibrium view of process dy namics, where change was viewed as a 'disturbance' and interactions across time and space were ignored. How ever, with the emergence of the nonequilibrium para digm and the realization that many of the ecosystems and natural resources are vulnerable to change, there has been a growing focus on change itself and mechanisms such as adaptation and resilience to cope with the inevitable change. Many indigenous peoples are still heavily dependent on natural resources for their survival, and have, over time, developed social and cultural mechanisms that reflect the very real dynamics of natural systems. A growing recognition of this has led to an emerging body of academic literature in conservation and natural resource management which is focused on the lessons that can be learnt from indigenous knowledge systems.

Using these indigenous knowledge characteristics, new approaches to conservation and natural resource management are being advocated. Central to these are adaptive management and adaptive governance – this entails holistic and detailed understanding or knowledge of the situation, and learning, so that one is able to re spond constructively to any future unpredictable events. Recent work presents indigenous knowledge systems as holistic 'cultural–ecological' systems – human in nature systems where the human and the ecological are tightly integrated, and where interactions between the two do mains over a range of scales sustain the coupled systems over space and time. Sustainability of these cultural–ecological systems largely depend on the adaptive cap acity of knowledge; an aspect of resilience that reflects learning, flexibility to experiment, and adoption of novel solutions. The collective 'social memory' of knowledge allows long term communal understanding of the dy namics of environmental change, and the transmission of pertinent experience. As the cultural–ecological system evolves over time in response to both biophysical and social changes, so does the social memory, usually re sulting in adaptive and resilient systems. The important practical aspect of this new approach to natural resource management is that it uses the processes through which indigenous knowledge is created as the ways in which natural resources should be managed, for example, responsive to change, through learning, through partici pation, and through action. The methodological impli cation for implementing this new form of natural resource management is that it has to involve what we now call participatory action research. Although these approaches to natural resource management sound ideal, there has as yet been little application of these ideas to practical management on the ground. It is also debatable to what extent techniques such as participatory action research will be taken up by research projects funded by development managers and large donors.

Conclusion

As will be obvious from the discussions above, within the indigenous knowledge debate, de contextualization is viewed as a major problem, namely the separation of knowledge from its human agents and the situations in which it is produced, reproduced, transformed, and is effective. Although the importance of culture in indigen ous societies is widely recognized, within development and conservation, indigenous knowledge is largely separ ated from the cultures in which it originates. Reference might be made to certain rituals or spiritual beliefs, but ultimately indigenous knowledge is placed outside culture so that it becomes more of an 'indigenous science' where what is 'useful' can be separated from what is not. This approach has the effect of defining 'useful' in a non indigenous purposeful way, which in turn externalizes the culture from which that knowledge came and/or labeling the culture as an impediment to successful development.

As media and communications expand, global–local distinctions are becoming ever more blurred. Is it therefore possible to regard local indigenous knowledge as something discrete, or are we trapped by represen tations of 'indigenousness' by the global (Western) media and the indigenous peoples who have learnt to use it to raise awareness of their issues? Should we try to separate local knowledge from global knowledge on the assump tion that one is superior to the other, or give preference to the mixture of local and global knowledge that in digenous peoples now rely on? Many indigenous peoples want to engage with a public beyond their doorsteps, so perhaps a better approach is to embrace the construct ivist processes upon which their knowledge is based ra ther than the context specific nature of that knowledge.

'Indigenous knowledge' may be the buzz phrase to guarantee research, funding, and development projects. Nevertheless, some caution is needed. Although there is a global growing recognition of the cultural–ecological indigenous system and the link between cultural and biological diversity, the law still does not protect in digenous knowledge. Ensuring the rights of indigenous peoples to define themselves, their knowledge, and our access to it is a fundamental challenge for social justice in the field of development today.

See also: Development I; Development II; Empowerment; Participation; Sustainable Development.

Further Reading

Agrawal, A. (1995). Dismantling the divide between indigenous and scientific knowledge. *Development and Change* 26, 413 439.
Berkes, F., Colding, J. and Folke, C. (2000). Rediscovery of traditional ecological knowledge as adaptive management. *Ecological Applications* 10, 1251 1262.

Briggs, J. and Sharp, J. (2004). Indigenous knowledges and development: A postcolonial caution. *Third World Quarterly* 25, 661 676.

Brown, M. F. (1998). Can culture be copyrighted? *Current Anthropology* 39, 193 222.

Ellen, R. and Harris, H. (2000). Introduction. In Ellen, R., Parkes, P. & Bicker, A. (eds.) *Indigenous environmental knowledge and its transformations. Critical anthrological perspectives*, pp 1 33. Amsterdam: Harwood Academic Publishers.

Escobar, A. (1995). *Encountering development: The making and unmaking of the third world*. Princeton, NJ: Princeton University Press.

Laurie, N., Andolina, R. and Radcliffe, S. (2005). Ethnodevelopment: Social movements, creating experts and professionalizing indigenous knowledge in Ecuador. *Antipode* 37, 470 496.

Mauro, F. and Hardison, P. D. (2000). Traditional knowledge of indigenous and local communities: International debate and policy initiatives. *Ecological Applications* 10, 1263 1269.

Posey, D. A. (2000). Ethnobiology and ethnoecology in the context of national laws and international agreements affecting indigenous and local knowledge, traditional resources and intellectual property rights. In Ellen, R., Parkes, P. & Bicker, A. (eds.) *Indigenous environmental knowledge and its transformations. Critical anthrological perspectives*, pp 35 54. Amsterdam: Harwood Academic Publishers.

Sillitoe, P. (1998). The development of indigenous knowledge. A new applied anthropology. *Current Anthropology* 39, 223 252.

Relevant Websites

http://www.cs.org
Cultural Survival A US Based Organization that Fights for the Rights of Indigenous Peoples around the World.

http://www.survival international.org
Survival International A UK Based Organization that Fights for the Rights of Indigenous Peoples around the World.

http://www.cbd.int
The Convention for Biological Diversity This Website Outlines the Protocols, Programs, and Mechanisms of the Convention.

http://www.undp.org
UNDP 'Indigenous Knowledge Program' Follow Links to Indigenous Knowledge for Information on Indigenous Knowledge Programs.

http://www.unesco.org
UNESCO 'Best Practice of Indigenous Knowledge' Follow Links to Indigenous Knowledge for Information on Best Practice on Using Indigenous Knowledge.

http://web.worldbank.org
World Bank 'Indigenous Knowledge for Development' Follow Links to Indigenous Knowledge for Development Approaches and Information Focused on Traditional Knowledge Systems and How They Can Help to Achieve the Millennium Development Goals.

Indigenous Mapping

J. Corbett, University of British Columbia-Okanagan, Kelowna, BC, Canada
M. Chapin, Center for the Support of Native Lands, Arlington, VA, USA
L. Gibson, Land and People Information Sharing Society (LAPIS), Parksville, BC, Canada
G. Rambaldi, Technical Centre for Agricultural and Rural Cooperation ACP-EU, Wageningen, The Netherlands

Glossary

Geographic Information Systems (GISs) A computer-based system designed to collect, store, manage, and analyze digital georeferenced information and associated attribute data.

Global Positioning Systems (GPSs) An array of space-based satellites and ground units that enable a user with a portable receiver to determine their absolute location on the Earth using known coordinates.

Indigenous Mapping The mapping of indigenous lands, by indigenous people, to secure tenure, manage natural resources, and promote cultural awareness.

Map Biography A data collection approach used in indigenous mapping in which individuals are asked to identify areas on a map used for harvesting-related activities.

Participatory Action Research (PAR) Research that involves all relevant parties in actively examining issues in order to change and improve them.

Participatory Development A development approach that supports and promotes community engagement in decision-making through the creation and use of a range of participatory methodologies that gather, analyze, and communicate community information.

Participatory Rural Appraisal (PRA) A development approach that seeks to incorporate the knowledge and opinions of rural people into the planning and management of development initiatives.

Public Participation GIS (PPGIS) It is the intersection of participatory planning and geographic information systems.

Remote Sensing The process of gathering information about the Earth from a distance. Such data are commonly gathered by satellite or air (aerial) photography.

Rapid Rural Appraisal (RRA) A series of swift, cost-effective social survey techniques used by outsiders to learn about rural communities in order that development initiatives can be tailored by those outsiders to meet local issues.

Traditional Land Use and Occupancy Studies An often map-based system commonly used in Canada to identify where indigenous people hunt, fish, and trap. Information contained in a TUS can inform resource management decision-making and identify areas of land-related conflict.

Introduction

Throughout recent history, cartography has been used by governments and elites to apportion and claim land and resources. However, whereas maps have historically been used to delineate and control territories inhabited by indigenous peoples, indigenous peoples are now using Western cartographic methods to reclaim and defend their ancestral lands and resources. Indigenous mapping thus represents a shift in the notion that cartography only reflects and shapes the interests of the powerful. Indigenous groups in Canada and the United States have been mapping their territories since the 1960s, and more recently the practice has been adopted in other parts of the world. The mapping process and products help to communicate indigenous land related knowledge to the state and external agencies, to assist in land use planning and resource management, to strengthen community organization and capacity, and to promote cultural awareness. The power of maps is reflected in the political, legal, economic, and cultural sensitivities evoked by indigenous mapping attempts, especially in states with repressive policies toward indigenous people. This article discusses how indigenous mapping has developed in North America and in other parts of the world. Due to different legal structures, land use traditions, and political and economic realities, approaches and methods vary among projects and geographic regions.

This article discusses how recent attempts to combine indigenous knowledge with spatial technologies have highlighted issues concerning the collection, storage, and representation of data. There is a need to develop clear protocols governing community held data, and to address deeper epistemological issues that arise when Western style maps are used to represent culturally nuanced information. In addition, there can be unforeseen impacts after maps are created and disseminated, which may be either positive or negative for indigenous communities. These issues will have to be addressed as mapping technologies become increasingly accessible and mapping becomes more widely adopted by indigenous communities.

Indigenous People and Indigenous Knowledge

Indigenous communities, peoples, and nations have often been oppressed and marginalized by dominant populations.

This article assumes the definition of indigenous communities, peoples, and nations proposed by the UN Secretariat of the Permanent Forum on Indigenous Issues, which states that the basis of indigenous communities includes the continued connection to and occupation of their ancestral territories. This working definition reflects an increasingly politicized understanding of the term 'indigenous' and is closely linked to issues of power inequality. Many contemporary issues related to indigenous peoples, their knowledge, and future are directly related to their immediate geography, in particular their access to and relationship with their traditional lands. Thus, mapping indigenous spaces has become a valuable tool in contemporary indigenous struggles over land and resources.

Indigenous Mapping and Western Cartography

Maps and diagrams seen through the perspective of the occupiers, in most cases European settlers, have historically represented indigenous spaces. These maps have purveyed social and cultural values through their content, for example, by replacing European place names for indigenous ones, and through their symbolization, for example, by depicting populated indigenous spaces as 'empty'. Western cartography's relationship to indigenous peoples exemplifies Brian Harley's notion of maps as value laden images that reflect and influence social and spatial relations. Maps have played a role as providers of graphic inventories of land and property in colonial agendas of social control and territorial annexation. They are implicated throughout history in the enclosure of common property and indigenous resource regimes, and concurrent transformations of property relationships. Contemporary indigenous mapping is thus an attempt to democratize mapping tools and address the legacy of these colonial agendas.

Indigenous mapping practitioners attempt to make the association between land and local communities visible to outsiders by using the commonly understood and recognized language of cartography. The documentation of place names, land management techniques, rituals, sacred sites, histories, and multiple other features, when recorded on maps, can contribute to the preservation and renewal of languages and culture. Documenting information such as traditional place names is of great cultural importance, as indigenous toponomies often contain and transmit valuable information about specific places. Peter Brosius's research with the Penan Gang of Sarawak showed that in indigenous communities, orally transmitted place names are powerful mnemonics not only for way finding, resource management, and other activities, but also as historical and spatial records of generations of occupancy and attachment to the land. Mapping provides a means for documenting such knowledge, which can be in the form of mental maps and storytelling.

Purposes of Mapping

Indigenous mapping initiatives are undertaken for a number of purposes: to delineate and claim ancestral lands; to articulate and communicate indigenous spatial knowledge to external agencies; to record and archive traditional knowledge for future generations; to assist indigenous communities in land use planning and resource management; to help plan economic development; to increase institutional and organizational capacity within communities; and to promote cultural awareness and cultural identity.

Indigenous Mapping in North America

Studies in the 1950s in the Inupiat region of the North Slope Barrows were the first notable application of indigenous mapping in North America. These studies addressed issues of public policy and conflicting land and resource use. In the late 1960s, mapping was used to portray aboriginal people's land use and occupancy in response to large development proposals on their territories, such as the James Bay Hydroelectric Project in Quebec and the Mackenzie Valley Gas Pipeline in the Northwest Territories. During the 1970s indigenous mapping became a standard component of land use and occupancy studies designed to inform land claim negotiations in the context of new policies that recognized aboriginal rights to land had not been extinguished.

In Canada and Alaska the primary method of documentation has been the 'map biography'. This method charts the subsistence regimen of individuals spatially through time. This method was refined in the 1970s with the Inuit Land Use and Occupancy Project in the Northwest Territories, which documented past and present hunting, fishing, trapping, and gathering patterns from an Inuit perspective. It recorded Inuit perceptions of their relationship to the land, compiling extensive data on history, place names, linguistics, subsistence techniques, campsites, and other cultural information. Aboriginal hunters were asked to map the areas they used for harvesting related activities (such as hunting, fishing, berry picking, and camp locations) during their adult lives. A profile of an entire community was then constructed by aggregating map biography information. The map biography method is tailored to the circumstances of each region and variations have been applied by hundreds of indigenous communities in Canada.

Another noteworthy example of indigenous mapping in Canada is the Delgamuukw versus British Columbia aboriginal title court case (decided by the Supreme Court of Canada on 11 December 1997). Several maps documenting the land use and oral history of Gitksan and We'etsuwet'en people were presented as evidence in this landmark case that ultimately ruled that the concept of aboriginal title to traditional lands existed.

The tribes of the lower 48 United States have been strongly influenced by advanced geospatial tech nologies. A number of tribes are using advanced GIS technologies for (among other things) environmental protection through identifying sites of environmental significance, generating a better understanding of the tribe's holdings and thus guiding tribal development and planning, creating resource inventories to help manage natural and infrastructural resources, as well as facili tating cultural/historical research.

In 1990, the Bureau of Indian Affairs (BIA) established the Geographic Data Service Center (GDSC) in Lake wood, Colorado, with the mission of providing technical resources and training to interested tribes. Later in the 1990s, GIS capacity among tribes began expanding and several were utilizing the digital databases offered by the BIA's GDSC. The Intertribal GIS Council was founded in 1993 with assistance from the First Nations Devel opment Institute, ESRI, NASA, and the USGS. The Hopi, the Blackfeet, the Leech Lake Band of Chippewa, and many others have used GIS, Global Positioning System (GPS), and remote sensing based technologies.

Indigenous Mapping in Asia, Africa, and Latin America

In Asia, Southeast Asia, Africa, and Latin America in digenous mapping has focused primarily on forest dwelling, hunting/fishing, and agricultural communities and groups. Despite geographic differences, these ini tiatives have been used to address similar issues to those facing indigenous peoples in Canada and Alaska. In these regions, other terms have been used interchangeably with 'indigenous mapping', including 'ethnocartography', 'counter mapping', 'self demarcation', a term used in Venezuela, and 'ancestral domain delineation and self determination', used in the Philippines.

Initially, mapping work with indigenous groups in these areas was influenced strongly by Rapid Rural Ap praisal (RRA), Participatory Rural Appraisal (PRA), and Participatory Action Research (PAR) approaches that are embedded in the broader paradigm of 'participatory de velopment'. These mapping initiatives were commonly initiated and undertaken by development workers to in corporate indigenous groups' spatial knowledge into planning and development initiatives. The data were often removed from the community and used to address the agendas of outsiders. The initial focus of these techniques saw practitioners adopting straightforward mapping approaches, such as the use of community sketch maps (freehand sketches drawn on large pieces of paper from memory) and mapping on the ground (often referred to as 'ephemeral mapping', which involves using any available materials, such as soil, plants, rocks, or household tools to construct the map). In some instances, practitioners added survey oriented cartographic tech niques such as transects, compass readings, and modeling, and sought to produce maps that were both rich with local knowledge as well as being georeferenced.

In the 1990s, the focus of indigenous mapping in Asia, Africa, and Latin America dramatically changed as the techniques that had initially been used in an extractive manner were adapted to suit the political agendas of indigenous communities and grassroots nongovernmental organizations (NGOs) representing indigenous peoples. In Southeast Asia, for example, NGO practitioners identify a regional workshop held in Manila by the Philippine Association for Intercultural Development (PAFID) in March 1995 as a key event that was instru mental in spreading indigenous mapping throughout the region. This workshop introduced mapping into Indo nesia where the period from 1995 until 2007 saw 2.5 million hectares of land mapped by indigenous com munities – often with technical assistance from NGOs such as Jaringan Kerja Pemetaan Partisipatif (JKPP) and Pembinaan Pengelolaan Sumber Daya Alam Kemasyar akatan (PPSDAK). Indigenous mapping was thus trans formed from an outsider driven initiative into a political tool used by indigenous groups to delineate traditional territories and ultimately pressure outsiders, including government and corporations, for formal recognition of traditional territories (**Figure 1**).

Since this change of focus in indigenous mapping in Asia, Africa, and Latin America, community control of the map making process, including determination of the information contained on the map, is increasingly em phasized. In addition, indigenous mapping projects have begun to incorporate geo spatial technologies and sys tems such as GPS, GISs, and remote sensing. These combinations have led to indigenous mapping variants such as participatory GIS (PGIS), 'participatory 3 D modeling' (P3DM) and 'participatory photo mapping' (**Figures 2** and **3**).

Issues

GIS and Participatory Approaches

Many indigenous mapping practitioners have attempted to combine participatory approaches and GIS. This has led to discussions regarding the compatibility of tradi tional knowledge and digital technologies. In the late 1980s, when development practitioners introduced PRA approaches such as sketch mapping there was little input from professional cartographers. The primary purpose of

Figure 1 Dayak people making scale maps in East Kalimantan, Indonesia. Photo credit: J. Corbett (2002).

Figure 2 Ogiek people in Kenya working on aerial photographs, Nessuit, Kenya. Photo credit: G. Rambaldi (2005).

this mapping was to document local spatial knowledge and to promote community based capacity building. GIS was introduced when accurate maps were needed for land tenure and legal battles and to communicate with policymakers. The declining cost and increasingly user-friendly interfaces of digital spatial technologies, as well as a greater public availability of data facilitated this movement. One of the goals of the Public Participatory GIS (PPGIS) movement is to determine how spatial technologies can be used by communities, including indigenous ones, in a genuinely participatory manner. However, the combination of high tech and participatory approaches has been uneasy, as GIS and other digital information technologies are viewed by some as capital intensive, complex to use, and expert driven. Furthermore, some GIS practitioners believe that spatial technologies have the potential to represent indigenous views of the environment better than conventional mapping,

Figure 3 Ogiek people using a participatory 3-D model, Nessuit, Kenya. Photo credit: G. Rambaldi (2006).

whereas others suggest that GIS technology is rife with epistemological issues of representation and can lead to assimilation of indigenous views by Western paradigms.

The role of technology intermediaries is a key question, especially if they come from outside of the communities – as, at present, they overwhelmingly do. There has been some concern, especially among indigenous people, that GIS lends primacy to a Western scientific approach at the expense of indigenous worldviews. In addition, GIS data structures, as they are presently configured, tend to prioritize certain types of information over others. Basic implementation is also an issue, as rural areas frequently lack access to electricity, the Internet, and a supportive infrastructure. This can lead to the technological aspects of the projects being controlled outside of the communities and run counter to the goal of true participation.

Unforeseen Impacts of Indigenous Mapping

Indigenous mapping can result in unanticipated consequences. For example, maps cannot always adequately portray the nuances of indigenous resource and property regimes, which can cause conflict when boundaries, resources, and property are delineated. Also, failure to implement clear protocols regarding ownership of information can lead to infringement of the intellectual property rights of indigenous people. Finally, introducing mapping technology can contribute to deepening divisions within communities along gender, economic, or other lines, as some groups prove more capable of adopting the technologies than others.

Conclusion

In recent times indigenous mapping has become a valuable tool for indigenous peoples around the world to delineate and claim their ancestral lands and to promote cultural awareness. It has also been applied to resource management, communication with external agencies, legal struggles, economic planning, capacity building, conflict management, identity building, and cultural awareness. It occurs in a variety of political, economic, and cultural contexts that compel careful thought before, during, and after project implementation, as mapping may serve both to empower and to marginalize indigenous peoples. There are social as well as epistemological issues associated with representing indigenous knowledge in a Western cartographic format, which are further complicated when combined with digital spatial information technologies. Challenges remain in determining how indigenous communities can be supported in adopting mapping and spatial information technology under favorable terms including autonomy over mapping processes and protection of community held data. Finally, the benefits of systematically mapping indigenous lands, whether using participatory methods, spatial technologies, or some combination of these two approaches, are not currently available to all indigenous peoples throughout the world. Discussion needs to continue as to ways to further democratize and disseminate participatory mapping methodologies and appropriate technologies. There is also a pressing need to confront the matter of securing financial resources for carrying out indigenous mapping projects on a large scale.

See also: Counter-Mapping.

Further Reading

Alcorn, J. B. (2001). *Borders, Rules and Governance: Mapping to Catalyze Changes in Policy and Management*. London: International Institute for Environment and Development.
Brealey, K. (1995). Mapping them 'out': Euro Canadian cartography and the appropriation of the Nuxalk and Ts'ilhqot'in First Nations' territories, 1793 1916. *Canadian Geographer* 39(2), 140 156.
Brody, H. (1992). *Maps and Dreams*. Toronto: Douglas and McIntyre.
Brosius, P. (2001). 'Local knowledges, global claims: On the significance of indigenous ecologies in Sarawak, East Malaysia'. In Grim, J. & Sullivan, L. (eds.) *Indigenous Traditions and Ecology.* Cambridge, MA: Harvard University Press and Center for the Study of World Religions.
Chapin, M., Lamb, Z. and Threlkeld, B. (2005). Mapping indigenous lands. *Annual Review of Anthropology* 34, 619 638.
Fox, J., Suryanata, K., Hershock, P. and Pramono, A. H. (2006). Mapping power: Ironic effects of spatial information technology. *Participatory Learning and Action* 54(1), 98 105.
Harley, J. B. (1989). Deconstructing the map. *Cartographica* 26(2), 1 20.
Harley, J. B. (1992). Rereading the maps of the Columbian encounter. *Annals of the Association of American Geographers* 82(3), 522 542.
Harmsworth, G. (1998). Indigenous values and GIS: A method and a framework. *Indigenous Knowledge and Development Monitor* 6(3), 3 7.
Peluso, N. L. (1995). Whose woods are these? Counter mapping forest territories in Kalimantan, Indonesia. *Antipode* 27(4), 383 406.
Poole, P. (1995). Geomatics, who needs it? *Cultural Survival Quarterly* 18(4), 1 77.
Rundstrom, R. A. (1995). GIS, indigenous peoples, and epistemological diversity. *Cartography and Geographic Information Systems* 22(1), 45 57.
Sirait, M., Prasodjo, S., Podger, N., Flavelle, A. and Fox, J. (1994). Mapping customary land in East Kalimantan, Indonesia: A tool for forest management. In Fox, J. (ed.) *Spatial Information and Ethnoecology: Case Studies from Indonesia, Nepal, and Thailand*, pp 1 14. Honolulu: East West Center.
Turnbull, D. (1993). *Maps Are Territories: Science Is an Atlas: A Portfolio of Exhibits*. Chicago, IL: University of Chicago Press.
Wood, D. (1992). *The Power of Maps*. New York: The Guildford Press.

Relevant Websites

http://www.conservationgis.org
ESRI conservation program: native/first nations index.
http://www.iapad.org
Integrated Approaches to Participatory Development (IAPAD).
http://www.nativemaps.org
The Aboriginal Mapping Network in Vancouver, British Columbia.
http://ppgis.iapad.org
The Open Forum on Participatory Information Systems and Technologies.
http://www.pafid.org
The Philippine Association for Intercultural Development (PAFID).

Industrial City

R. Harris, McMaster University, Hamilton, ON, Canada

Glossary

Agglomeration Economies The economic advantages that accrue to manufacturers who cluster in industrial cities.

Downtown A term first coined in the United States to refer to the central business district and, usually, to immediately adjacent industrial or residential areas.

Economic Base The main, or major single, source of local employment. Sometimes, used to refer only to employment in a sector that exports beyond the metropolitan region.

Externalities The uncompensated costs, or benefits, of an economic activity. Commonly exemplified by the polluting factory that imposes health and other costs on those who live nearby, and especially downwind.

Industrial Suburb A suburb that has a substantial base of manufacturing employment, but which is usually integrated into the city-wide labor market.

Telecommuting Using telecommunications, rather than the daily commute, to maintain contact with an employer or client.

Zoning A tool of land-use planning by which municipalities can control the use to which particular parcels of land are put.

Industrial City

An industrial city is an urban place in which manu facturing dominates the local economy and, by extension, the environment and social relations. Such cities have often been associated with high levels of air and/or water pollution. The social structure and class relations of their workplaces are often polarized and this is both reflected and reproduced in a simplified social geography. This was especially true in smaller centers and company towns. The first industrial cities arose, in Western Europe and North America, in the nineteenth century in contrast to their compact, preindustrial predecessors. Large scale production and innovations in transportation technology enabled a growing separation of home from work. Mu nicipal zoning regulations encouraged the formation of extensive, homogeneous land use districts. Because of de industrialization, in Europe and North America, the industrial city has arguably been supplanted by the postindustrial metropolis. In recent years, the most rap idly growing industrial cities have been those in a handful of developing nations. The growth of the industrial city has promoted as well as reflected economic growth, transformed the social and esthetic experience of its residents, encouraged the growth of state regulation as well as voluntary associations and clubs, and created new possibilities for political association.

Definition

Cities are often distinguished in terms of their dominant form of economic organization (e.g., capitalist, socialist), culture (e.g., Islamic), or their technology. These terms imply different views about what forces are the most important in shaping urban growth and change. The concept of the industrial city implies that technology is vital in this regard. According to this view, best articu lated by Gideon Sjoberg, because preindustrial cities depended upon animate sources of energy, production was small scale and relatively inefficient, land transpor tation occurred at a walking pace, and cities were small, compact, and dense. The development of inanimate sources of energy, initially steam power and later elec tricity, increased the scale and efficiency of manu facturing and also of urban transportation. They enabled workers to commute, making possible the large scale separation of home and work and the growth of much larger, industrial cities. Industrial cities, then, can exist and have existed in both capitalist and socialist countries, such as the United States and what was once the Soviet Union, respectively.

Industrial production usually entails a clear separation between mental and manual labor. The class structure of large industrial enterprises tends to be polarized between a large number of relatively unskilled workers and smaller numbers of owners and managers. Industrial cities, then, tend to be socially polarized and with a small middle class, whether professionals or the self employed. The simplest class structures, and the most extreme so cial polarizations, occur in the company town. Most company towns exist to exploit a local resource, for ex ample, a mining deposit or nearby timber stands, or a site of water power, which was a determining influence before the advent of steam engines and then long distance electricity transmission. Even today, the costs of such transmission means that power intensive industries sometimes locate next to major dams, an example being the ALCAN aluminum refinery at Kitimat, British Col umbia. Because the resource must be tapped *in situ*, the company has little or no choice about where to locate its facility and workforce. In some cases, however, manufacturing companies have selected somewhat isolated sites, usually in an effort to control their labor force. Some model industrial communities have worked well, but from the company's point of view the risks are high. A dramatic example was the town of Pullman, Illinois, which was established by a manufacturer of railroad cars in the 1880s. Soon after its foundation, Pullman was the site of a confrontation that grew into the largest industrial strike in US history. Most manufacturers have preferred to locate in larger urban centers, where they are freer to hire and fire at will.

Not all large, industrial cities are equally polarized in social terms. Cities, such as Manchester, England, in the nineteenth century, that contained a small number of large factories are likely to display sharp class contrasts. It is no coincidence that Friedrich Engels, who with Karl Marx developed influential ideas about the importance of class struggle, knew Manchester better than any other British city because he owned a mill there. In contrast cities, such as Birmingham, England, that contained many small workshops where masters and artisans worked together and where social mobility was more common, were less polarized. It is no coincidence that, during the early 1830s, it was in Birmingham that masters and workmen joined together to stage the huge political demonstrations that compelled the British government to broaden the franchise in the Reform Act of 1834.

Another influence upon social relations in the industrial city is the manner in which housing is provided. Tenants move easily and frequently; home owners have a stake not only in their home and neighborhood but in their city. Employers and politicians have usually assumed that home owners are less likely to espouse radical causes, or to go on strike, than tenants. Following the Pullman strike of the 1890s, employers in company towns actively encouraged their workers to acquire homes. In the twentieth century, governments in many countries have encouraged home ownership, in part for its supposed political consequences. Other options are possible. In China after the revolution of 1949, for example, a system was developed whereby state enterprises provided housing (as well as health and education services) for their workers, so that industrial cities in effect became aggregations of work unit compounds, or partly self contained industrial villages. This had the effect of fragmenting the workforce, cutting across the social distinctions of occupation and class, based to some degree on communist party membership. Similarly, paternalist arrangements were developed in the same period by major private sector employers in Japan. Such systems cannot easily coexist with entrepreneurial capitalism, however, and the trend in both these Asian countries has been towards a clearer separation between work and housing relations.

Internal Geography

The class relations of the industrial city are reflected in, and reproduced through, the geography of the urban landscape. Its most characteristic and significant feature is the separation of work and home. In preindustrial cities, the characteristic unit of production was the household, which often contained extended families. Commonly, an artisan lived and worked in the same structure: stairs connected a ground floor workshop to living accommodation, and perhaps, storage rooms on upper floors. Journeymen, apprentices, and, in more affluent households, servants might occupy attic space. In this era, it was largely the merchants' warehouses, used to store goods involved in long distance trade, which defined their own geographical district. Elsewhere, homes, industry, and commercial activity were tightly interwoven.

As the scale of production increased, the corresponding social unit became the company, not the household. At first, as in Coventry in the mid nineteenth century, small manufacturers appended factories to family homes. But proximity to growing numbers of workers, and the noise and pollution from power driven machinery, soon encouraged them to build dedicated factory buildings and to move their homes elsewhere. They became commuters. In the earlier and smaller urban centers, they traveled to work by carriage. Thus Engels described how, during the 1840s, Manchester mill owners commuted from suburban homes to centrally located factories. In larger cities, by the second half of the nineteenth century, they could rely on commuter trains. For the more affluent, among whom women were expected to stay at home, the emerging geography of the industrial city created distinct social worlds for men and women. Workers, including many women, lived close to their jobs. They worked ten or even twelve hour a day, and low wages compelled them to walk to work. Their lives, and those of their families, were audibly regulated by the factory whistle. For many decades, industrial cities displayed a dense and promiscuous intermingling of housing, factories, and the small shops that sold daily essentials. In the era before refrigerators, these included all types of perishable foods. Home may have become separated from workplace but, as a physical as well as an economic fact of life, the latter continued to loom large.

From the late nineteenth century, new methods of transportation transformed the industrial city. Industry in port cities was usually distributed beside or immediately behind waterfront warehouses. In some early industrial cities, for example Birmingham, England, and the adjacent Black Country, canals enabled scattered industrial growth, but in most cities it was the railway that hastened the process of decentralization. Some industries had always been located towards or beyond city limits, especially those like tanning that produced unpleasant smells or those like brick making that required extensive sites. As rail transportation became dominant for freight in the second half of the nineteenth century, however, it became increasingly common for manufacturers to establish or move their operations outwards, along radial rail lines or adjacent to suburban freight yards. By the early twentieth century, land developers laid out and marketed industrial districts, now known as industrial parks which were well served not only by rail but also by water, waste disposal, and electricity. Among the first of these in North America were Chicago's Central and Clearing districts. Rail transportation also fostered the unplanned growth of industrial suburbs, and of satellite towns, communities that lay within the same economic region as a major urban center but beyond commuting distance. This de centralized pattern was reinforced by the advent of truck transportation, which became important during the inter war years for the shipment of goods within urban areas and then, after 1945, for long distance shipments. Since the 1960s, lighter industries have also been attracted to fringe locations by their use of air freight, and all major airports are now large nodes of warehouse and industrial employment.

Among small companies, the removal of factories was accompanied by management and clerical functions. By 1900, however, the growth of large corporations enabled the creation of head offices that remained, or were lo cated, downtown. Especially in North America, where height regulations were permissive, steel frame technol ogy and elevators permitted offices to concentrate in central business districts (CBDs). Since 1945, suburban office parks, especially dedicated to routine functions, have become common.

These changes in manufacturing location have en couraged workers to move away from central residential areas, including those that contemporaries labeled slums. The decentralization of industry was viewed as one of the more positive forces shaping urban areas, around 1900. Near suburban factory sites workers occupied new, sin gle family housing. In Europe, most of this was company or speculatively built. In North America, much was owner built, typically where land was very cheap, be cause they were unserviced and unregulated.

Innovations in transportation technology made it possible for employees to live further from work. Un usually, in London, the government encouraged railway companies to offer cheap fares on workers' trains. In most industrial cities, even those like Melbourne which ac quired a rail network, commuter trains were used by the middle class. Some large cities, such as Paris, Boston, and New York, built subways, where fares became affordable to a wide social cross section. But in most places, the first form of mass transportation was the electric streetcar (tram). During the 1890s, cities across North America, Australia, and Western Europe acquired streetcars that opened up lower density workers' districts that became known as streetcar suburbs. By the 1920s, buses sup ported development in the interstices of rail and car lines. Looser patterns of development became possible with the advent of the automobile. Ford's model 'T' made car ownership affordable to skilled workers in the United States from the 1920s. Elsewhere, cars only shaped the industrial city after 1945, but then at all scales. Houses acquired garages, then prominent double garages; new subdivisions were laid out to accommodate and manage cars, with wide streets and *culs de sac* to prevent through traffic; shopping areas, initially roadside strips and then clustered malls, were given extensive parking; residential development occurred at increasing distances from workplaces; limited access highways became the arteries of the industrial city.

None of these transport technologies determined the exact form of the industrial city. Many municipalities, especially in North America, allowed the car to rule. European cities, notably Stockholm, planned public transit. But each new technology changed the city's geographical possibilities by enlarging them. Together, except for a minority of workers, they removed the sights and sounds of the factory from the home life of workers and their families.

One way that planners and municipalities have ac commodated new methods of transportation has been through land use regulation. In the industrial city, and in early industrial suburbs, homes and industries grew up cheek by jowl. Workers and their families had to put up with the resulting externalities, the uncompensated costs of noise and air pollution. In the early twentieth century, land use zoning was introduced progressively, whereby districts came to be zoned exclusively for housing, in dustry, and commercial activity, and then for sub categories of use – such as light industry or single family dwellings on large lots. The rationale was that zoning could limit externalities. Because of the way this tool was implemented, it inhibited the intermingling even of land uses that are closely complementary (e.g., small stores in residential areas) and mandated a large scale segregation of land uses that many now find monotonous and dys functional. Recently, encouraged by the New Urbanism, more flexible methods of land use regulation have been introduced.

Global Trends and Patterns

Many industrial cities have lasted long enough to ex perience not only a rise but also stagnation and decline. The first industrial centers developed in Britain, Western Europe, and North America. Some, such as New Lanark, Scotland, and settlements in New England, were at tracted to water power and were no more than villages.

As industrial towns developed on the basis of steam power and adjacent to coal fields, larger clusters formed – such as the Black Country, the Ruhr, and the Pittsburgh region. Other regions depended on external economies, notably, the sharing of a skilled, specialized labor force. These served domestic, and sometimes, continental markets. A few, notably the Lancashire mill towns, relied on a captive colonial empire.

Since the 1970s, there has been a globalization of trade and of the markets for manufactured goods. This has led to the decline of many centers of manufacturing in the developed world. Individual industrial cities had experienced decline in the nineteenth and early twentieth centuries, but the extent of recent shifts is unprecedented. Whole industries, such as steel making around Pittsburgh, and spinning in Manchester, have disappeared, and with them the economic base of metropolitan areas. De industrialization has posed a huge challenge to many industrial cities. Some have met it by reinventing themselves. Pittsburgh, whose steel mills have been demolished, is prospering as a center of biotechnology; recycling factories and warehouses as commercial lofts and residential condominiums, Manchester is marketing itself as a cultural center. The esthetic and architectural potential of warehouses and factories have been deployed to revive waterfronts and once industrial areas in many cities, for example, Toronto's Distillery district. But other cities struggle. For largely abandoned places such as downtown Detroit, Michigan, and Camden, New Jersey, loss of factory jobs and the suburbanization of the middle class have produced an urban wasteland. In the work, for example, of Camilo Vergara this has yielded an esthetic of industrial ruin that has a visual appeal but, as yet, no commercial potential. The decline of many industrial cities, and the declining share of manufacturing in the employment base of diversified cities like London and New York, has encouraged some to speak of an emerging postindustrial city. To the extent that manufacturing is no longer the major source of employment in the developed world, this label makes sense. It acquires resonance when nonmanufacturing employees emancipate themselves from the daily commute by telecommuting. As yet, however, the continuities between industrial and postindustrial cities are considerable.

De industrialization in Europe and North America has been accompanied by the rapid growth of industrial urbanization in some parts of the developing, and newly industrialized world. The most dramatic examples are provided by China, which has rapidly become the major supplier of a wide range of goods to the world, from toys, zips, and shoes to refrigerators, steel, and concrete. These goods are produced in a range of places that extend from those that are small and specialized, such as Qiaotou (zips), Sheng Zhou (ties), Wen Ling (shoes), and Yiwu (socks), through larger and heavily industrial places such as Harbin and Chengdu, to diversified metropolitan centers, notably Shanghai, that also contain manufacturing. This range of urban settings was typical of earlier phases of industrial growth and, in that respect, history is repeating itself. A major difference is that, fueled by a large, global consumer market, the pace of growth in modern industrial cities is even faster than that of its predecessors, while its dimensions are larger. Beginning with a single family run business in 1980, for example, entrepreneurs and 50 000 migrant workers have made the town of Qiaotou, Zhejiang province, into the dominant global center for the production of clothing fasteners: production in 2005 included 15 billion buttons and 200 million meters of zips. Manchester was the 'shock city' of the 1840s, Chicago of the late nineteenth century, and Los Angeles – much more a manufacturing center than is commonly realized – of the twentieth century. But what is now happening in a number of Chinese cities goes beyond shock, to awe.

Significance

Our understanding of the nature and significance of the industrial city has been steadily broadened by a succession of intellectual traditions. From the beginning, observers appreciated the significance of urbanization for the growth of industry while recognizing its destabilizing effects. From the 1920s to the 1960s, the dominant Chicago school of urban sociology emphasized the personal and social consequences of urbanism. During the 1970s and 1980s, Marxists and social historians focused upon how the city affected class relations and politics, while feminists showed how industrial urbanism affected men and women differently. Since the 1980s, scholars have explored the cultural effects of cities on our experience of modernity, and upon the way cities themselves, and not just municipalities, govern their occupants. At the same time, the public health issues have been rediscovered, but now within a global and environmental frame of reference.

The industrial city has not only reflected the rise of industrial scale production but also enabled it, while shaping our society and culture. The concentration of enterprises in urban centers yields external economies that include the sharing of a skilled labor force, of specialized suppliers of raw materials and of business services, and of the costs of physical infrastructure. For the individual firm this reduces costs, while in the aggregate it facilitates economic growth. The juxtaposition of many related industries, suppliers, and consumers provides fertile ground for the development, testing, and implementation of new business ideas and products, whether by serendipity or through conscious efforts to outdo competitors. This economic ferment spawns, and is in turn fertilized by, the sorts of social and cultural innovations that have characterized industrial cities. Those, like London, New York, and Shanghai, that are diverse in economic terms have also been important cultural centers, and have generally been better able to innovate and adapt than those, such as Manchester and Detroit, that were more specialized. In various ways, industrial cities have both reflected and hastened economic growth.

The residents of industrial cities have always paid a price in terms of public health. Historically, large scale manufacturing concentrated low income households in poorly serviced slums, and produced polluting chemicals. Until the early twentieth century, mortality rates were higher in urban than in rural areas, and even then, pol lution was accepted as an inevitable correlate of pros perity. In Yorkshire, England, people said "where there's muck there's brass [money]." A growing awareness of public health problems in industrial towns such as Homestead, Pennsylvania, led to improved water and sewer systems, together with public health initiatives and local efforts to limit industrial emissions. Since the 1970s, public awareness of the regional and global consequences of industrial pollution has prompted many national governments to act. In the developed world, legislation has helped improve air and water quality in industrial cities such as Hamilton, Ontario. The absence or lax enforcement of such regulations in the developing world, however, is producing appalling local conditions. Today the most polluted industrial cities are in the developing world, notably China. There is still brass in muck.

The rise of large, industrial cities has also had social consequences that are often known as urbanism. The city dissolves the informal controls of the village or small town. Most urban residents are unknown to one another, and most social interactions in cities occur between people who know each other only in specific roles, such as parking attendant, store clerk, or customer. Individuals become more free to live as they wish, and in ways that deviate from social norms. In response, and because the high density of city living requires the pliant coordin ation of many thousands of people, urban societies, and especially, municipalities have developed a wide range of methods to control urban behavior. These include regulations that control private land use, building con struction and maintenance (to minimize fire risk), and the production of pollution and noise (especially at night). They also regulate public behavior, especially parking and traffic, through prohibitions against jaywalking and speeding. Of necessity, they have established technologies (parking meters, stop lights, remote cameras, traffic calming devices) and armies of people to enforce these regulations, including inspectors and parking wardens as well as the police. To the extent that this regulatory apparatus responded to conditions that were unique to, or which took their most extreme form in, the industrial city, they may properly be regarded as urban in character.

The freedoms, now regulated, of city life have been associated with a common experience of anonymity and alienation in the urban crowd. In response, and because they often could not rely on informal social networks, city dwellers have deliberately sought community, by forming clubs and associations. Because of their specific gender roles, women have been most significant in the more informal and localized of these. City dwellers have also found community by identifying with city wide organizations, notable professional sports teams which, from Aston Villa FC to the Brooklyn Dodgers, were a product of the late nineteenth century city. More gen erally, industrial cities gave birth to new forms of class, ethnic, and political association, many rooted in working class neighborhoods, which shaped municipal and na tional politics.

The most fundamental guideline of all that regulates urban life is the clock, which defines the rhythms of work, shopping, and recreation. These rhythms are faster than those of the country, more predictable in detail, and yet in aggregate more varied. Historically, they created entirely new experiences of speed, fluidity, juxtaposition, syncopation, and change, and then a corresponding esthetic of modernity that has been ex pressed by novelists, composers, and painters. In that sense, industrial cities helped to shape culture in the narrow sense as well as the broad. Arguably, and per haps most important of all, because the rhythm and landscape of the city – of planted parks no less than of parking lots – is artificial, industrial cities have fostered the assumption that humanity is able to control its own destiny. It is only in very recent decades that the ex perience of water and air pollution on a regional scale, and of global warming, has undermined that peculiarly urban assumption.

See also: Capitalism; Central Business District; De-Industrialization; Industrial Parks; Modernity; New Urbanism; Slums; Suburbanization; Urban Planning and Human Geography; Urbanism.

Further Reading

Barth, G. (1980). *City people. The rise of modern city culture in nineteenth century America*. New York: Oxford University Press.
Briggs, A. (1965). *Victorian cities*. New York: Harper and Row.
Buder, S. (1967). *Pullman. An experiment in industrial order and community planning*. New York: Oxford University Press.
Dennis, R. (1984). *English industrial cities of the nineteenth century. A social geography*. Cambridge: Cambridge University Press.
Engels, F. (1952). *The condition of the working class in England in 1844*. London: Allen and Unwin.
Fogelson, R. M. (2001). *Downtown. Its rise and fall, 1850 1950*. New Haven: Yale University Press.
Gillette, H. (2005). *Camden after the fall. Decline and renewal in a postindustrial city*. Philadelphia: University of Pennsylvania Press.
Harris, R. and Lewis, R. (2001). The geography of North American cities and suburbs, 1900 1950. A new synthesis. *Journal of Urban History* 27, 262 293.
Jacobs, J. (1969). *The economy of cities*. New York: Random House.
Joyce, P. (2003). *Rule of freedom. Liberalism and the modern city*. London: Verso.

Lewis, R. (ed.) (2004). *Manufacturing suburbs. Building work and home on the metropolitan fringe*. Philadelphia: Temple University Press.
Masareel, F. (1972). *The city (die stadt)*. New York: Dover (original publication, 1925), http://newpartisan.com/images/masereel/index.html (accessed on Nov 2007).
Melosi, M. (2000). *The sanitary city. Urbna infrastructure in America from colonial times to the present*. Baltimore: Johns Hopkins University Press.
Merrifield, A. (2002). *Metromarxism. A Marxist tale of the city*. New York: Routledge.
Mumford, L. (1961). *The city in history*. New York: Harcourt, Brace & World.
Sjoberg, G. (1960). *The preindustrial city. Past and present*. New York: The Free Press.
Teaford, J. (1993). *Cities of the heartland. The rise and fall of the industrial midwest*. Bloomington: Indiana University Press.
Vergara, C. J. (1999). *American ruins*. New York: Monacelli Press.
Weber, A. F. (1963). *The growth of cities in the nineteenth century. A study in statistics*. Ithaca, NY: Cornell University Press (original pubication, 1899).
Wirth, L. (1969). Urbanism as a way of life. In Sennett, R. (ed.) *Classic essays on the culture of cities*, pp 143 164. New York: Appleton Century Crofts (original Publication, 1939).

Relevant Websites

http://www.makingthemodernworld.org.uk/
A narrative that includes keyword searchable images, London Science Museum, Making the Modern World.
http://digitalgallery.nypl.org
A topical and keyword searchable collection, not exclusively of New York City, New York Public Library Digital Gallery.
http://memory.loc.gov
Browse by topic (e.g., cities, towns), document type (e.g., photos, prints) or keyword searchable, Library of Congress. American Memory.
http://www.artmuseums.harvard.edu/collections
Harvard University Art Museums, Collections.
http://www.library.pitt.edu
Historic Pittsburgh Image Collections, Photographs of Pittsburgh, University of Pittsburgh (University Library SystemDigital Library).
http://invinciblecities.camden.rutgers.edu/intro.html
Howard Gillette and Camilo J Vergara. Invincible Cities.
http://detroityes.com
Lowell Boileau. Fabulous ruins of Detroit, Detroit Yes.
http://www.archive.org
Prelinger Archive, Internet Archive.

Industrial Districts

F. Celata, University of Rome ''La Sapienza'', Rome, Italy
U. Rossi, University of Cagliari, Cagliari, Italy

Introduction

In an essay that published in 1987, economist Giacomo Becattini recalls the time in which the concept of external economies and industrial atmosphere originally appeared. A 1969 study, published in a revised form in 1975 by IRPET (a Tuscany based research institute of which Becattini was the director at the time), offered an interpretation of Tuscany's development path which emphasized the role of the geographic, historic, and societal characteristics of the region. The work touched upon, albeit only in empirical terms, the notion of the industrial district by referring to the flow of external economies directed toward individual firms within a process of productive specialization on a local level. Drawing inspiration from Alfred Marshall's book on industry and trade, the study placed emphasis on the special atmosphere pervading localized agglomerations of small and medium sized firms specializing in leather goods and footwear, textiles, and furniture in Tuscany, such as Santa Croce sull'Arno, Prato, Montecatini Terme, etc. In the essay, Becattini lamented the delayed consideration of these categories by mainstream economists as well as by politicians and policymakers.

In this and in his subsequent work, Becattini – unanimously recognized as the founder of the Italian school on industrial districts – invites the scholarly community to a twofold rehabilitation of geography within the realm of spatial economic theory and applied research. On a disciplinary level, he acknowledges the importance of human geography within the disciplines investigating the dynamics of regional economic development (two contributors to this path breaking publication are geographers themselves: Fabio Sforzi and Maria Tinacci Mossello), and, perhaps more importantly, on a theoretical level, Becattini underlines the importance of a geographic perspective in applied economic research by pointing to 'localized sociospatial systems' and the always unique 'thickening of productive relations' as major units of analysis in industrial economics and in regional economic development studies more generally.

The reassertion of the importance of geography in economic theory and empirical research resonates with the contemporary rehabilitation of space over time by some leading social scientists in the same period with reference to other social issues and fields of investigation. Broadly speaking, it can be argued that in the 1980s the so called 'Tuscan model' came to represent in the field of regional science and local economic development studies something comparable to what the city of Los Angeles represented in the context of postmodern urban development: an allegedly paradigmatic example of sociospatial transformations in post Fordist times. Ironically, the recognition of such a paradigmatic role occurred despite the fact that Tuscany was not a post Fordist region or a 'new' industrial space properly defined, as its industrial development trajectory was much older than those of other notable examples of regional post Fordism (such as those in the American Sun Belt). Or, to put it in a different way, Tuscany symbolized the ways in which the post Fordist mode of production was able to bring together the old and the new, the history and the present of industrial organization in capitalist societies.

In the eyes of industrial districts scholars, Tuscany was indeed a unique and at the same time successful combination of economic and sociocultural, exogenous and endogenous factors in regional development processes. These factors include: first, the national division of labor that prevailed in the Italian economy after the political unification of the country in the nineteenth century with the localization of large factories in the North–West and the other regions of the Center–North, such as Tuscany itself being excluded from mass industrialization; second, the region's long term tradition of craft production in labor and design intensive sectors, such as clothing and the garment industry; third, the share cropping system (*mezzadria*) prevailing in agriculture with its remarkable economic versatility and unchanging social formations (the 'Mediterranean model' of household structure and, most notably, the intense allegiances generated within enlarged family lineages); fourth, the professional ethic of local entrepreneurs, resulting in a widespread recognition of well performed work (epitomized by the enduring belief in the 'mastery of the craft'); fifth, the outward looking character of the regional economy due to the wealth of international economic exchanges (testified by the role played in industrial districts development by brokers acting for foreign firms in channeling international demand toward Tuscany); sixth, the open mindedness of the region linked to the unique legacy of cultural and artistic resources with their potential for attracting foreigners and enlightened visitors.

The Emilian Model and the Institutionalist Turn

Although Becattini referred to the work of Alfred Marshall as the main source of inspiration in the (re)discovery of industrial districts and 'light industrialization' paths in Tuscany, the emphasis he placed on the role of endogenous and sociocultural factors implicitly called for an expansion of industrial district theory itself beyond Marshallian externalities, agglomeration theory, and the conventional boundaries of neoclassical economics. Critics of this classically Marshallian characterization of industrial districts suggested a more explicit engagement with the concept of 'embeddedness' in economic sociology as well as with the emerging strands of neo institutionalist theorizing in regional development studies. The Marshallian framework of analysis, however, has not been abandoned following these critiques but – as we see below in this article – has been updated and amended with other theoretical approaches such as neo Schumpeterian economics and the transaction costs theory, giving rise to a 'neo' Marshallian perspective.

Initially, the 'Tuscan model' of the industrial district provided a fundamental contribution to the making of the geographic imaginary and mythology associated with the rise of the Third Italy as a sociospatial paradigm in post Fordist capitalism. However, following the revisiting of the Marshallian legacy and the influence acquired by neo institutionalist and neo regionalist theorizations, scholars have looked for other regional models of industrial district organization. In this vein, since the mid 1990s Emilia Romagna has gradually taken the role played earlier by Tuscany. Emilia Romagna is a region that has many examples of prosperous industrial districts: ceramics in Sassuolo, knitwear in Modena, motorcycles and shoes in Bologna, and food industry in Parma. Since the beginning, a distinguishing feature of the 'Emilian model' of industrial districts has been related to the specific institutional base of the local economy. Since the early 1980s, economist Sebastiano Brusco had noticed how the artisans and small entrepreneurs of Emilia Romagna cope with customary difficulties associated with the small size of firms by creating associations of mutual help and assistance (such as the National Confederation of Artisans, CNA) and various entities of public–private partnership. These associations and coalitions specialize in the provision of a host of producer services; for instance, they coordinate purchasing and credit negotiations, they set up technical consultancy offices as well as consortia for marketing and purchase of raw and semi fabricated materials, and, most importantly, establish cooperatives providing guarantees for bank loans at the lowest possible rates of interest. This form of interfirm cooperation and interdependence, strongly supported by the local and regional government and other regional actors such as local banks and the trade unions, has created the institutional thickness on whose basis industrial districts have grown up and flourished in the region, compensating the inefficiencies as well as the increasing hollowing out of the state as a regulator of economic activity in post Fordist and post Keynesian societies. The presence of these institutional intermediaries crossing conventional economy/state/civil–society boundaries has allowed industrial districts and similar sociospatial configurations of small firms to be competitive in a world dominated by big firms, global actors, and transnational networks.

The attention that has been devoted to the institutional foundations of the Emilian model has offered a more comprehensive understanding of the role played by sociocultural and endogenous factors in regional economic development, which had been only inductively highlighted in studies on the 'Tuscan model'. Subsequent literature dealing with what came to be known as the 'associational economy', in the mid and late 1990s, has been particularly inspired by studies investigating the role of these institutional intermediaries in the development and reproduction of industrial districts in Emilia Romagna. Elaborating on this example, students of regional economic development have reached a point of consensus over the fact that while networks of technologically advanced firms tend to gain their competitiveness from access to scientific knowledge, codified rules, technical advances, and strategic management, small firm networks tend to rely more on informal, mainly tacit and interactive, knowledge as a source of competitive advantage as well as on individual and collective experience, apprenticeship, incremental innovation, adaptation to changing circumstances, and organizational flexibility. This neo institutionalist theorization has contributed to the building of a theory of industrial districts which is less dependent on the original sociospatial context of the Third Italy (**Figure 1**) and more easily applicable to other geographic contexts.

Traveling Theory: Industrial Districts outside the Third Italy

A crucial contribution to the generalization and the internationalization of industrial districts theory has come from the work of the US scholars Michael Piore and Charles Sabel and particularly from their groundbreaking concept of flexible specialization. Thanks also to their work, industrial districts started to be regarded as concrete manifestations of a generalized shift toward flexible industrial organization in post Fordist times. In the wake of this rising attention at an international level, concentrations of specialized clusters of small and medium sized firms have been found and investigated in many regions of the Western World, including Southern Germany (Baden Württemberg), Southern France (Rhône Alpes), and the US West Coast.

According to these studies, industrial districts take shape not only as a consequence of the thickening of

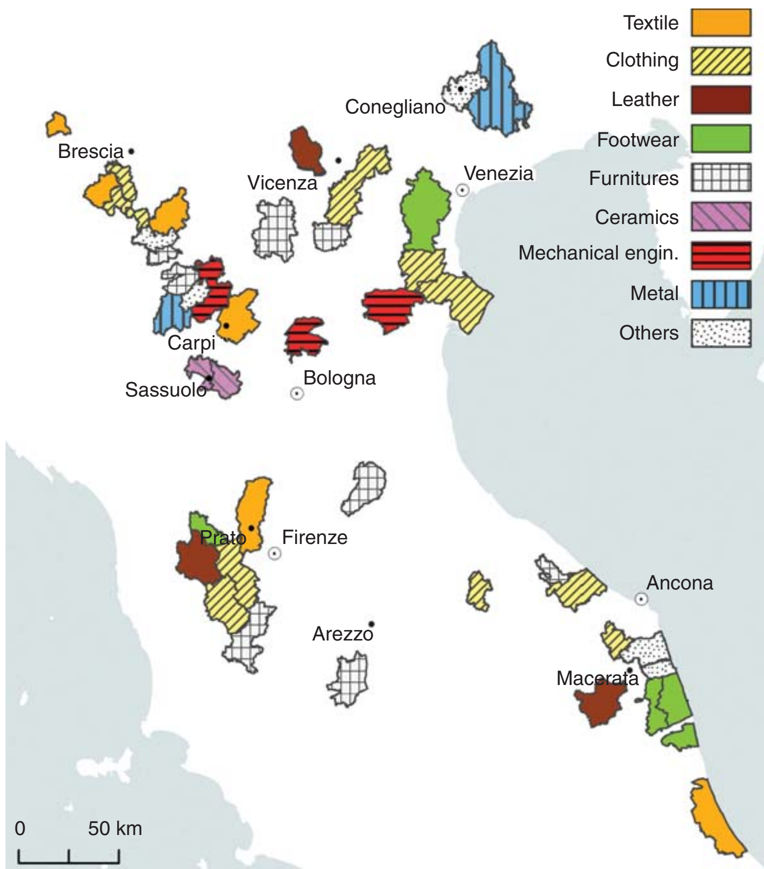

Figure 1 Industrial districts in the Third Italy. Adapted from Sforzi, F. (1990). The quantitative importance of Marshallian industrial districts in the Italian economy. In Pyke, F., Becattini, G. & Sengenberger, W. (eds.) *Industrial Districts and Inter-Firm Co-operation in Italy*, pp 75 107. Geneva: International Institute for Labour Studies.

cooperation ties between autonomous small sized producers, like in the Third Italy, but also as a result of outsourcing strategies and the restructuring of previously vertically integrated production systems or in the wake of state led industrial development. Related dynamics of industrial reorganization have been observed in a wide range of geographical contexts and economic sectors, such as South Korea (electronics, fabricated metal, and textiles), Japan (Toyota city), Seattle (Boeing production system), California (wine industry), Ile de France (fashion clothing and high technology industries), and many others, especially in the so called Global South, as shown in the remainder of this paragraph. A notorious example in this respect is the Hollywood's film and television industry. Similarly to design intensive industries, film production required continuous upgrading and had to deal with increasingly volatile and fragmented markets during the 1950s and 1960s. Responses to this uncertain economic situation have been of a different kind: from vertical disintegration to a widespread reliance on contingent staff, from the decreasing size of studios to their outright abandonment, and from product differentiation to the spread of specialized independent companies. Scholars have observed that these organizational strategies are aimed not only at fostering specialization but also at externalizing and socializing risks and losses.

The rise of industrial districts and the disintegration of Fordist industrial complexes are thus complementary processes highlighting the convergence of both large and small firms toward increasingly networked organizational structures. From this point of view, districts being dominated by locally owned small sized firms can be distinguished from industrial clusters having a more hierarchical organization. The German region of Baden Württemberg, for example, shares many features with the Third Italy, including the role of regional socioeconomic history, local culture, intermediary institutions, and a widespread attitude toward interfirm cooperation. Being specialized in the production of machine tools and automotive components, and similarly to Japanese industrial districts, the innovation process is to a larger extent internalized in leading large companies that are keen to preserve their technological know how and have a key role in coordinating the production process. The shift toward flexible modes of production has been investigated and documented also within old industrialized regions in the context of decentralizing strategies pur sued by large companies. But the majority of flexible production systems has been highlighted, as said in the introduction to this article, mostly with reference either to regions whose local development trajectory has taken shape over the long run, such as those of the Third Italy, or to 'new industrial spaces', such as those of the American Sun Belt, which have developed in more recent times.

The so called Californian School of external econ omies interpreted the rise of flexible production strat egies as a response to the growing uncertainty of contemporary markets. The increasing division of labor between autonomous and at the same time inter dependent firms has been understood as an attempt at maximizing the benefits of specialization and reducing the risks of technological lock in. Disintegration of pro duction and the intensification of interfirm linkages raise transaction costs due to the need for more frequent, less predictable, and more complex interactions. Spatial ag glomeration is an optimal solution for keeping transac tion costs low, allowing a more direct relationship between buyers and suppliers, the adoption of just in time supply systems, mutual trust, and a host of shared interests.

Other contributions in this field have focused on the relationship between agglomeration and innovation, drawing inspiration from evolutionary and neo Schum peterian economic theory and trying to go beyond the abstract idea of an industrial atmosphere as something intangible. In doing so, the analytical focus has shifted from the complex network of interdependencies that keep industrial districts together to the role of these synergies in sustaining the evolution and regeneration of the whole regional innovation system. From this point of view, the economic process has been viewed as aimed at producing knowledge, rather than products or simply technology, and incremental innovation has been inter preted as an interactive learning process. In a world where codified knowledge can be easily transmitted across long distances, noncodified or tacit knowledge becomes increasingly valuable, being difficult to be (re)produced and impossible to be imitated. The notion of tacit knowledge was first introduced by Michael Polanyi to refer to aspects of technological know how which cannot be easily expressed in formal statements or communicated over long distances. Its transmission is based on interpersonal relationships, contextual experi ence, learning by doing or by seeing, and face to face contacts. Information exchanges, as well as market ex changes and the coordination of production, demand physical proximity as they require frequent face to face interaction, rapid feedback, the transmission of non codified information, and multidimensional and per formative communication.

The neo Marshallian conceptualization of external economies flowing through vertical relations with sub contractors along the value chain is complemented with the idea of horizontal relationships taking shape between firms and a wide range of actors. Empirical studies have demonstrated that the role of direct exchanges has been largely overestimated and that input–output relations are not sufficiently dense to account for the concentration of firms in the majority of industrial districts and for their performance. Especially in innovative clusters, the adoption of the industrial district model is not due to the need to reduce transaction costs related to market ex changes, but is due to nonmarket exchanges and a variety of 'untraded interdependencies': firms need to locate in mutual proximity in order to participate in a buzz of communication and interaction that fosters innovation and technological experimentation. This could explain not only the rise of industrial districts but also the ag glomeration of high tech industries, like in Silicon Valley, and even the concentration of financial, advertising, law, and other advanced producer services in global cities such as New York or London. From this perspective, Marshallian industrial districts came to be regarded as a specific manifestation of a more general tendency toward spatial clustering and interfirm cooperation.

A controversy has arisen about whether hyperspe cialized industrial regions are able to provide such a relational and communicative buzz in an efficient way when compared to more diversified and socially dense urban environments. The New York garment district has been presented as a demonstration of the importance of both sources of externalities. The concentration of ap parel manufacturers and their contractors in midtown Manhattan, in proximity with a wide range of fashion related businesses such as retailers and specialized ser vices, design schools, or fashion magazines, has proven to be a crucial source of competitiveness despite the high cost of location that firms have to afford. Similarly, in novative clusters like Silicon Valley are located in the outskirts of metropolitan regions in search of a balance between the benefits of industrial specialization, defined as Marshallian externalities, and the so called Jacobs' externalities (from the work of urbanist Jane Jacobs) as sociated with the economic and social variety of large cities.

Another influential contribution to the generalization of the idea of the industrial district has come from Michael Porter and his theory of clusters. Clusters differ qualita tively from industrial districts, even though the two terms are frequently confused and used interchangeably. Busi ness clusters generally include more varied forms of geographic concentrations of firms compared to those conventionally highlighted under the banner of industrial districts theory. Clusters literature considers, for instance, both hierarchical and nonhierarchical structures in the

Table 1 A typology of industrial districts and related local production systems

Type	*Firm size*	*Specialization*	*Governance*	*Features*	*Regional examples*
(Neo)Marshallian industrial districts	Small	Craft production, design-intensive sectors	Non-hierarchical	Industrial atmosphere	Prato, Santa Croce sull'Arno
The 'Emilian model'	Small and medium	Labor-intensive industries	Semi-hierarchical	Institutional thickness, embeddedness	Modena, Sassuolo
Flexible production systems	Mixed	Manufacturing, technological innovation	Semi-hierarchical	Spatial division of labor, coordination economies	Hollywood, Seattle, Baden-Württemberg, Japan
Learning regions	Small and medium	High-tech	Mixed	Knowledge exchanges, face-to-face contacts	Silicon Valley, Cambridge
Business clusters	Mixed	A wide range of manufacturers and services providers	Semi-hierarchical	Policy-led development, intersectoral linkages	California, Massachusetts, Singapore

governance of contractual interfirm relations, while industrial districts theorists usually privilege the non hierarchical structures developed by locally owned small sized firms (**Table 1**). Moreover, cluster scholarship investigates a set of economic agents which is wider than the one considered in industrial districts studies: for instance, a cluster can be formed by service providers, while industrial districts are customarily indentified with reference to the manufacturing sector. Despite its elusiveness, the theory of clusters has the merit of providing a more simplified and generalized account of the importance of proximity and specialization for industrial competitiveness. From this point of view, dealing with spatial clustering has proven to be a more viable analytical task as compared to the conceptual vagueness and lack of theoretical rigor frequently characterizing the work of geographers and economic sociologists dealing with industrial districts. Consequently, the notion of the cluster has entered economic analysis far more readily than the concept of industrial districts, contributing to a geographic turn in economic theory and to a renewed attempt at integrating spatial agglomeration into formalized economic models.

The theorization of clusters and industrial districts has contributed, moreover, to the updating of the lexicon of local economic development policies. Scholars usually have approached industrial districts not only as viable analytical concepts but also as success stories and as policy tools. These sociospatial entities exemplified the paradoxical image of an increasingly globalized world where competitive advantages increasingly lie in localized knowledge spillovers and relationships. This region centered representation of the world economy is in stark contrast to previous functionalist understanding of the center–periphery relationship and has invited scholars and policymakers to embrace a distinctly endogenous approach to industrial upgrading. Industrial policies, in Europe and elsewhere, have been committed to the enhancing of existing industrial districts (in Italy, for example, with the Law 317/1991) and have more generally emphasized the importance of cooperation, endogenous development, and regional uniqueness. The rediscovery of the region as a strategic unit of coordination and governance has inspired political studies investigating rising forms of new localism and new regionalism.

The industrial district model, originally interpreted as an alternative pathway for the organization of advanced economies, has been exported to the developing world and to the peripheral regions of industrialized countries, such as the Italian Mezzogiorno. This is no surprise, as Marshallian industrial districts are industries mostly located in areas like the Third Italy, which have never undergone a process of Fordist industrialization. The formation of industrial districts is rather the result of the upgrading of local agglomerations of small sized craft producers, which are widespread in the Global South. A well known example is that of the Sinos Valley district, in the Brazilian state of Rio Grande do Sul, with more than 1500 firms involved in shoes production. Unlike Marshallian industrial districts, the cluster has developed as a supplier of the large US companies and many local firms have grown in size and have integrated vertically and horizontally. The co presence of small and large firms, of flexible specialization and Fordist organizational patterns, of localized ties and external linkages has indeed turned out to be a common feature in regions which were customarily regarded as exemplary cases of industrial districts incubators. The two Mexican footwear districts of Guadalajara (women's shoes) and León (men's shoes) have a similarly dualistic industrial organization, with some firms operating in quasi hierarchical chains dominated by the US based buyers and a number of smaller firms producing for the domestic market on a more equal and cooperative basis. Other examples of industrial districts have been found in India, Thailand, Kenya, Zimbabwe, and in many other geographic contexts, in cluding the villages and the township communities of rural China. These countries are thought to hold a comparative advantage in specializing in labor intensive industries which do not require the employment of ad vanced technologies. Focusing on the endogenous evo lution of locally owned small firms, the industrial districts model has offered a regionally embedded pathway to industrialization which diverts from conventional state driven development strategies.

Conclusion

The traveling of the theory of industrial districts through different geographic contexts and theoretical per spectives has contributed to the reassertion of the im portance of space and place in economic analysis and to improve conventional understanding of regional devel opment processes and policies. This not only has en riched the original conceptual model, but has also attracted criticism for a tendency to excessive general ization and empiricism at the same time. Some authors have contended that the generalization of the original model has been mostly instrumental to its translation into a tool of regional and industrial policy, as has happened in other lines of new regionalist enquiry.

The institutionalist rereading of the industrial districts literature has been criticized for its conceptual fuzziness. This literature – it has been argued – provides a so phisticated understanding of the complex links that tie economic processes to specific places, but is unable to explain why this has happened there rather than some where else. The balance between competition and co operation, local and translocal interdependencies, and competitive success and economic decline is inherently precarious, and its repositioning is continuously changing and transforming places.

The spatial clustering of firms is likely to configure itself as a transitional stage in a trajectory of organiza tional change. The standardization of production and an increasing price competition, over the medium run, could favor the integration of production and delocal ization as attempts at reducing costs and enlarging the scale of production, at the expenses of quality and product differentiation, as occurred in both the Third Italy and Silicon Valley during the 1990s.

Being specialized in labor intensive products, (neo) Marshallian industrial districts are particularly exposed to competition from developing countries with lower wage rates. Scholars have long wondered whether industrial districts would survive the increasingly powerful forces of globalization. In the Third Italy, increasingly competitive pressures have led to groundbreaking restructuring pro cesses. Small sized firms have fled high quality niche markets, trying to grow in size, through mergers and ac quisitions. Only a small number of leading medium sized firms have been able to transform themselves into global brands. In many cases, these firms have delocalized the production process in newly industrializing countries and have retained the service functions such as design, mar keting, and retailing. Long distance links with production branches and suppliers located in Eastern Europe (par ticularly in Romania) and in East Asian countries, in many cases, have entirely replaced local interdependen cies or shorter distance networks (with producers located in the South of Italy, for example). The small size of firms, the informality of their mutual relationships, and over specialization, which were once supposed to be strategic assets, have proven to be obstacles put in the way of the necessary internationalization of industrial districts. The institutional foundations of districts, on the other hand, have proven to be difficult to be reproduced over time, being constantly challenged by firms' individualist agency, opportunism, and generational turnover. Case studies conducted in 'paradigmatic regions' such as Tuscany or Baden Württemberg have shown how place based ties are likely to engender forms of institutional lock in: routines and conventions which hinder the district's regeneration process.

Scholars have rarely succeeded in providing an evo lutionary explanation of how industrial districts rise and fall within the contexts of increasingly multiscalar socioeconomic formations. Interscalar production rela tionships lack systematic investigation and have been weakly theorized. The emphasis which has been placed on geographically localized ties overlooks the role played by external relations of coordination linking districts to the global economy. Clusters serve as nodes of dense interfirm relationships which nurture innovations within the global networks of production. These networks are in many cases dominated by large multinational corpor ations needing at the same time to be embedded in specific places and to be stretched across larger geo graphic space. The representation of a locally bounded network of cooperative relations dealing with global markets has thus been amended with the analysis of a wide range of transcalar linkages connecting local pro duction systems to global commodity chains through networks of hierarchical and uneven relationships.

While the emphasis placed on self contained and lo cally bounded interdependencies allowed early pro ponents of industrial districts theory to leave aside conventional sector centered views of the industrial process, embracing instead a place based perspective, more recent insights have reasserted the importance of network representations, approaching spatial clustering as only one among many relational proximities enabling production to be coordinated and the global economy to function. Local production systems are interpreted as entities being directly or indirectly connected to each other across differing geographic scales and persistently rooted in specific places and locales, which are themselves embedded in wider geographic networks. Skepticism toward rigid scalar representations of the spatialities of globalization have opened the way for the embracing of a so called 'flat ontology': connectivity between places and firms is not understood merely in terms of geographic co location, but rather in a topolo gical fashion. Spatial clustering coexists with long distance networking: both local buzz and global networks are important for coordinating production and creating knowledge, for industrial upgrading and adaptation. From the perspective of relational economic geography, this means trying to broaden the reach of production net works and to open up 'black boxes', such as the firm, the district, and the region, in order to understand how networks function and develop connections between actors and structures through power relations.

See also: Agglomeration; Embeddedness; Fordism, Post-Fordism and Flexible Specialization; Global Commodity Chains; Global Production Networks; Institutionalism/ Institutional Geographies; Learning Regions; Local Development; Local Economic Development, Politics of; Local Economic Development; Locality Debates; New Regionalism; Regional Development and Noneconomic Factors; Regional Development, Endogenous; Regional Innovation Systems; Relational Economic Geography; Social Capital; Spatial Clustering, Detection and Analysis of; Spatial Division of Labor.

Further Reading

Amin, A. (1999). The Emilian model: Institutional challenges. *European Planning Studies* 7(4), 389 405.

Amin, A. and Thrift, N. (1992). Neo Marshallian nodes in global networks. *International Journal of Urban and Regional Research* 16, 571 587.

Asheim, B. T. (2000). Industrial districts: The contributions of Marshall and beyond. In Clark, G. L., Feldman, M. A. & Gertler, M. S. (eds.) *The Oxford Handbook of Economic Geography*, pp 413 431. Oxford: Oxford University Press.

Bagnasco, A. (1977). *Tre Italie. La problematica territoriale dello sviluppo italiano*. Bologna: Il Mulino.

Becattini, G. (ed.) (1975). *Lo sviluppo economico della Toscana*. Florence: IRPET.

Becattini, G. (ed.) (1987). *Mercato e forze locali: Il distretto industriale*. Bologna: Il Mulino.

Becattini, G., Bellandi, M., Dei Ottati, G. and Sforzi, F. (2003). *From Industrial Districts to Local Development. An Itinerary of Research*. Cheltenam: Elgar.

Brusco, S. (1982). The Emilian model: Productive decentralisation and social integration. *Cambridge Journal of Economics* 6, 167 184.

Cooke, P. and Morgan, K. (1998). *The Associational Economy. Firms, Regions, and Innovation*. Oxford: Oxford University Press.

Hadjimichalis, C. (2006). The end of Third Italy as we knew it? *Antipode* 37, 82 106.

Harrison, B. (1992). Industrial districts: Old wine in new bottles? *Regional Studies* 26(5), 469 483.

Markusen, A. (1996). Sticky places in slippery space: A typology of industrial districts. *Economic Geography* 72, 293 313.

Markusen, A. (1999). Fuzzy concepts, scanty evidence, policy distance: The case for rigour and policy relevance in critical regional studies. *Regional Studies* 37, 701 717.

Martin, R. and Sunley, P. (2003). Deconstructing clusters: Chaotic concept or policy panacea? *Journal of Economic Geography* 3, 5 35.

Park, S. O. and Markusen, A. (1995). Generalizing new industrial districts: A theoretical agenda and an application from a non Western economy. *Environment and Planning A* 27, 81 104.

Piore, M. J. and Sabel, C. F. (1984). *The Second Industrial Divide: Possibilities for Prosperity*. New York: Basic Books.

Porter, M. (1998). Clusters and the new economics of competition. *Harvard Business Review* 76, 77 90.

Scott, A. J. (1988). *New Industrial Spaces: Flexible Production Organization and Regional Development in North America and Western Europe*. London: Pion.

Sforzi, F. (1990). The quantitative importance of Marshallian industrial districts in the Italian economy. In Pyke, F., Becattini, G. & Sengenberger, W. (eds.) *Industrial Districts and Inter Firm Co Operation in Italy*, pp 75 107. Geneva: International Institute for Labour Studies.

Storper, M. (1997). *The Regional World. Territorial Development in a Global Economy*. New York: Guilford Press.

Tinacci Mossello, M. (1982). Economia e geografia. Dall'analisi delle economie di agglomerazione alla teoria dello sviluppo regionale. *Rivista Geografica Italiana* 90, 304 331.

Trigilia, C. (1986). *Grandi Partiti e Piccole Imprese: Comunisti e Democristiani Nelle Regioni a Economia Diffusa*. Bologna: Il Mulino.

Industrial Location

K. Chapman, University of Aberdeen, Old Aberdeen, UK

Glossary

Agglomeration Economies Economic benefits (especially cost savings) gained by individual firms as a result of their proximal location to other firms.
Behavioral Approach It draws upon insights from psychology to emphasize the role of decision making in human geography.
Cumulative Causation It refers, in a geographical context, to the way in which human actions/decisions through time are interrelated. More specifically, these interrelationships often involve positive feedback effects contributing to uneven development.
Economic Man A concept used in normative (see below) location theory. It assumes that decisionmakers involved in economic activity are motivated exclusively by the objective of profit maximization. Economic man is also assumed to possess perfect information and infallible judgment.
Industrial District A localized region of specialized economic activities gaining agglomeration economies (see above) by the exploitation of close internal (to the district) linkages based on flows of goods and, especially, information.
Locational Hierarchy It refers to the links between the hierarchical structure of organizations, especially large firms, and the geographical distribution of their various activities.
Locational Interdependence It refers to the way in which individual choices of location (by firms) are influenced by the location choices of others.
Mode of Production The structured social relationships involved in the organization of productive activity.
New Industrial Spaces Concentrations of manufacturing, especially associated with newer technologies such as electronics, that have developed in areas, such as Silicon Valley, quite different from the traditional centers of industrial production.
New International Division of Labor A special case of the spatial division of labor (see below) associated with the internationalization of production as reflected in the establishment by transnational corporations of manufacturing facilities in rapidly industrializing countries, notably in Southeast Asia. The phenomenon was characterized as 'new' in an influential book published in 1980.
Normative The normative approach to explanation attempts to establish what 'should' occur in a particular situation defined by a set of carefully specified assumptions. It therefore provides a yardstick or norm with which reality may be compared.
Quantitative Revolution A phase during the 1950s and 1960s when statistical and modeling techniques were widely adopted in Anglo-American geography.
Regional Policy Policy measures addressing the economic, social, and political problems arising from uneven development.
Spatial Division of Labor A concept which describes and seeks to explain the concentration of particular sectors and/or types of production in specific geographical areas.

The study of industrial location is concerned with the spatial distribution of manufacturing activity. It shares this focus with industrial geography and the terms have become virtually synonymous in academic usage. Industrial location is sometimes distinguished from industrial geography by an explicit focus upon the circumstances surrounding the choice of location for new factories rather than a broader concern with the multiple influences (such as factory closures) shaping the distribution of manufacturing. This review embraces both interpretations, attempting to clarify the research questions linking them before proceeding to address: (1) the location of individual factories and (2) the geography of wider systems of production. Finally, attention is drawn to the relevance of industrial location for various strands of public policy.

Factories, Firms, Industries, and Places

Reflecting upon the scope of industrial location/geography, it is apparent that the boundaries are difficult to define. The location of a single factory cannot be understood without reference to its relationships with other factories. These may be part of the same firm or they may be owned by other firms acting as suppliers, customers, or competitors. Thus, the objects of study in industrial location/geography embrace factories and firms as well as aggregations of firms operating within the same industry or within wider systems of production involving different sectors. These objects of study are applied in tackling research questions from two very different, but related perspectives. First, questions may be directed toward solving an optimizing problem – finding the 'best' location for a specific investment. The resolution of such problems, however, clearly has implications for communities chosen (and those not chosen) to receive such investments. Much work in industrial lo cation/geography reflects this reality, representing a second, essentially welfare rather than business oriented approach. Despite ideological differences, the two dif ferent starting points contribute to our understanding of the reciprocal relations between the evolution of indus tries and the evolution of places.

The significance of these reciprocal relations is en hanced by the very uneven distribution of manufacturing activity, evident at a variety of geographical scales. Relatively few countries account for the bulk of world manufacturing output and industrial activity within these countries is similarly concentrated in relatively few regions. The character of these regions, whether tradi tional centers of heavy industry such as the Ruhr in Germany or new industrial spaces such as Silicon Valley in California, is intimately connected with the fortunes of their principal manufacturing sectors. The decline of its coal, steel, and heavy engineering complex has posed a very different set of problems for the Ruhr from those associated with the spectacular growth of the electronics and, more recently, the biotechnology industries in Silicon Valley. Indeed, the pace of industrial change is accelerating, a trend reflected in a substantial research literature concerned with the dynamics of industrial location. This literature focuses upon the relationships between technical change, innovation, and regional de velopment. Biological metaphors such as life cycles have been widely employed in this literature which has re invigorated interest in the study of industrial location.

Location Factors and Factory Location

Different manufacturing processes require different combinations of inputs; iron and steel plants, for example, are major users of energy, whereas labor is much more important in the cost structure of a factory making clothing. A substantial literature has examined the interactions between spatial variations in the availability and cost of the multiple inputs required by industrial processes, the geography of demand for the outputs of these activities, and the choice of location for new fac tories. Attention was initially focused upon these inter actions at the national scale, but they have become even more relevant at the international scale as a result of the cross border operations of large companies (see below). Various typologies of location factors have been pro duced drawing upon essentially empirical, survey based methodologies, but resting upon the foundations pro vided by industrial location theory. The logic of location theory is reflected in **Figure 1** in the distinction between factors primarily influencing the cost of production and those primarily influencing the revenues earned by sell ing manufactured goods.

The work of Weber has, arguably, exerted the greatest influence upon the study of industrial location. He out lines the basic principles governing the choice of location for a new factory, starting from the premise that the best location is the one at which costs are minimized. Con siderable emphasis is placed upon transport costs, al though he acknowledges that spatial variations in labor costs and agglomeration economies (see below) may also be important. He conceptualizes the problem in terms of the outcome of a balance of forces defined by the relative attractiveness of raw material and market locations. Thus, in a world where transport costs are high, the least cost location is, allowing for possible modifications by labor costs and agglomeration economies, defined by the costs of assembling inputs from suppliers relative to the costs of distributing outputs to customers. Although ex tremely simplistic, this view of the industrial location problem was consistent with the prevailing pattern of manufacturing when Weber first published his book in Germany in 1909, as the coalfields attracted heavy, raw material based industries and market oriented industries concentrated in the major centers of population.

Weber's work is generally taken as the point of de parture in normative studies of industrial location, but his ideas were extended by several others following the neoclassical tradition of location theory. Weber un realistically assumed that the manufacturer could sell all s(he) produced and took no account of the potential ef fect of the location and actions of competitors upon sales. This problem was addressed in subsequent refinements which recognized that decisionmakers react to com petitors, giving a degree of locational interdependence as the best location changes according to the strategy of others. A closely related insight, that some factories may be better placed to serve potential customers, shifted the emphasis away from spatially variable costs to variable revenues. Several writers showed how factories may be

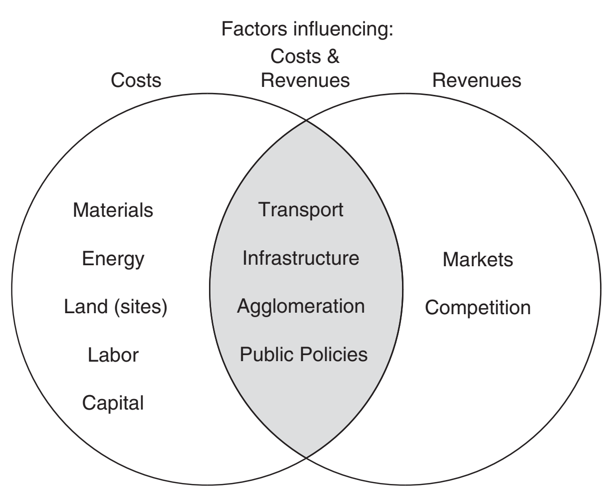

Figure 1 A typology of industrial location factors.

situated to maximize revenues from sales to surrounding market areas, applying similar logic to central place theory.

The least cost and locational interdependence (market area) perspectives each represent partial views of the industrial location problem. Commonsense suggests that the rational business (wo)man should seek a location at which profits (total revenues minus total costs) are maximized. Most contributors to industrial location theory were aware of this and some attempted to derive more ambitious theoretical representations which simultaneously incorporated the variables influencing both cost and revenue sides of the equation. Successive refinements struggled to accommodate the complexities of the problem, including the fact that the cost/revenue variables are not necessarily independent of one another (as represented by the overlapping zone in **Figure 1**). For example, better transport and other infrastructure facilities may not only reduce operating costs in one place relative to another, but also enhance the opportunities for selling products and, therefore, for increasing revenues.

Theoretical approaches to the study of industrial location have proved difficult to apply in the real world of business practice, although they have been influential ideas in the world of academic geography. Weber's approach is both deductive, proceeding from a set of basic propositions regarding the objectives of those responsible for the industrial location decision, and also normative, in the sense that it identifies the optimal outcome for the manufacturer under a clearly specified set of conditions defined by a series of simplifying assumptions. In these respects, industrial location theory both inspired and reflected the prevailing intellectual orthodoxy of the 'quantitative revolution' in the 1950s and 1960s. Although the 'new economic geography' of the 1990s continued to draw upon neoclassical location theory in seeking to explain the uneven distribution of economic activity (including manufacturing), the 'cultural turn' has generated a very different view of the way that business firms reach decisions, including location decisions. Indeed, it could be argued that early efforts to make industrial location theory more 'realistic' anticipated the shift of intellectual direction within human geography represented by the 'cultural turn'.

The assumption of economic man (*sic*) underpinned the models of location theory, including Weber's analysis of the industrial location problem This notional decisionmaker is distinguished by the single minded pursuit of profit maximization and by the possession of complete knowledge of all relevant economic information, including the ability to predict the actions of competitors and future events. Some writers attempted to qualify these unrealistic omniscient powers while retaining the essential assumption that decisionmakers strive to find the optimal factory location. Rawstron introduced the notion of spatial margins to profitability which defined the boundaries of an area, surrounding the optimal location, within which profitable operation can be achieved. This idea was subsequently developed by Smith and Pred, the latter proving influential in the emergence of a behavioral approach applied not only in the study of industrial location, but also in other areas of human geography during the 1960s and 1970s.

The behavioral approach attempted to explain the distribution of business activity, including the location of factories, by focusing attention upon the decision making process. Not surprisingly, this focus highlighted the discordance between the qualities of economic man in the abstract world of location theory and the more complex influences upon the actions of decisionmakers in the real world. More specifically, numerous empirical studies of industrial location decisions involving questionnaire surveys emphasized the importance of somewhat ill defined and ambiguous personal considerations over more rational location factors such as proximity to raw materials and markets. This strand of research in industrial geography proved relatively short lived, not least because the investments associated with the post World War II boom in North America and Western Europe dried up in the 1970s and factory closures became much more frequent events than openings. Nevertheless, location theory and behavioral concepts remain important in the overlapping worlds of management studies and business consultancy as major companies and public/private sector agencies have strong mutual interests in negotiating over the location of large scale, internationally mobile investments such as electronics and motor vehicle assembly plants. These negotiations are often informed by location factor analyses deriving their inspiration from the traditions of location theory and behavioral studies.

Systems of Production

Implicit in industrial location theory is the assumption that the unit of production (the factory) and the unit of organization (the firm) are combined in the form of the single plant firm. Contemporary industrial production is, however, dominated by large, multiplant, multiproduct corporations. The spatial structure of these corporate production systems became a focus of research attention during the 1960s and 1970s, although concern with the role of large firms as agents of geographical change (extending well beyond the location decision narrowly defined) remains an important strand of enquiry. One important consequence of the focus upon large firms has been an appreciation of the fact that the location of factories is only one aspect of the geography of production. The operations of factories are supported by other activities in management, research and development (R&D), and so on. Furthermore, case studies of large firms have emphasized that these activities tend to be found in different kinds of place as hierarchical systems of control are expressed on the ground in locational hierarchies with, for example, head office functions concentrated in major metropolitan centers and manufacturing dispersed to more peripheral locations.

Massey translated these empirical findings into the concept of the spatial division of labor which she subdivided into two basic types. First, a sectoral division of labor associated with regional specialization in particular industries; second, an intrasectoral division of labor based upon the locational strategies of large firms. Sectoral divisions of labor reflect and contribute to the interdependencies (noted above) between the evolution of industries and the evolution of places as the social reproduction of particular skills reinforces the distinctiveness of economic landscapes associated with the production of motor vehicles as compared with those specializing in, for example, the production of chemicals. The identification of intrasectoral spatial divisions of labor was, however, Massey's most influential contribution. This insight provided a link between narrower concerns with the geography of firms and the much broader, Marxist inspired interest in the geography of capitalism during the 1980s. Thus, the impact of industrial restructuring upon different regions was partially explained in terms of intracorporate spatial divisions of labor. In particular, it was noted that manufacturing branch plants in peripheral regions tended to be more vulnerable to job losses and closure in times of recession than facilities situated closer to the centers of corporate control. Intracorporate spatial divisions of labor have not only been identified at the national scale. Frobel *et al.* postulated *The New International Division of Labour* based upon the same link between the organizational and spatial structures of large corporations. Advances in communication and transport technologies have facilitated the coordination and control of corporate systems dispersed between many countries. In these circumstances, global corporations based in developed countries such as the US and Japan have established plants across a wide spectrum of industries, including electronics and clothing, in countries with much lower labor costs such as Mexico and Malaysia. These corporate strategies have contributed to major changes in the location of industry at the global scale. It is important to emphasize that, while intracorporate spatial divisions of labor have exerted an important influence upon the location of industry, the nature of this influence is constantly changing in response to the fluid internal structure and external boundaries of large business organizations. A trend toward flatter networked organizations will create different patterns from those associated with the top down, hierarchical structures which influenced Massey's thinking.

Despite the expanding power and influence of large firms, the distribution of manufacturing cannot be explained exclusively in terms of the operation of intracorporate systems of production. One of the most enduring research themes in the study of industrial location has addressed the phenomenon of agglomeration, defined as the tendency for productive activities to locate in close proximity to one another. This characteristic clustering reflects the existence of regional systems of production based on complex interfirm relationships. Efforts to understand the nature of agglomeration, in particular the benefits gained by participating firms from mutual proximity, have addressed the central problem of explaining the uneven distribution of manufacturing activity. These efforts have extended from Weber's incorporation of agglomeration in his industrial location theory to more recent work on industrial districts, new industrial spaces, and clusters. Simplifying considerably, two general strands may be identified in this substantial literature. The first emphasizes the essentially static, cost based advantages gained by the agglomeration of related firms; the second is concerned mainly with the dynamics of innovation and learning. The pervasive influence of normative industrial location theory and, in particular, the ideas of Weber ensured that the former remained the preferred approach to the analysis of industrial agglomeration at least until the 1980s. This focused on the costs of transactions between interdependent firms. These transactions are sometimes expressed as physical flows or linkages in a production chain concerned with, for example, the manufacture of motor vehicles or chemicals. They may also take the less tangible form of services such as accountancy. Whatever the precise nature of the transaction, the underlying assumption is that the spatial proximity of supplier and buyer reduces costs and enhances the overall efficiency of the production system. Contemporary interest in agglomeration has shifted away from a focus upon transaction costs toward an emphasis upon information based relationships. This partly reflects a greater awareness of the social nature of production. Within this context, the spatial proximity of firms is seen to facilitate the circulation of information and ideas (i.e., knowledge), thereby contributing to the collective learning necessary to sustain the momentum of innovation and improvement upon which the competitive advantage of production systems rests. At first glance, it is surprising that proximity should be so important in a world made smaller by satellite communications and the Internet. Numerous studies have confronted this paradox, noting that social networks, which are typically constrained by distance, are important mechanisms for the circulation of nonroutine, high level information and business intelligence. The tendency for firms involved in relatively new sectors such as biotechnology and nanotechnology to cluster in particular regions has been widely noted and seems to confirm the significance of localized knowledge circulation systems in promoting agglomeration.

Industrial Location and Public Policy

The impact of the uneven and dynamic distribution of manufacturing upon the economic and social character of places has often encouraged states to try to influence the pattern of industrial development. This generalization remains true for very different political systems ranging from the modified state socialism of China to the mixed economies of Europe and the free market capitalism of the US. Many aspects of public policy directly and indirectly affect the manufacturing sector. Thus, all states pursue a wide range of industrial policies incorporating, for example, measures designed to regulate competition or to encourage entrepreneurship. Science and technology policies promoting R&D and trade policies relating to cross border flows of investment also influence the geography of industrial development. All of these various policies are, however, primarily concerned with shaping the commercial environment of manufacturing firms in ways designed to enhance the overall economic performance of the state relative to other states. Any influences upon the location of industry are incidental to this wider objective. By contrast, various strands of regional policy have sought to directly influence industrial location.

Regional policy is concerned with addressing the inequalities of opportunity associated with uneven development which is a basic geographical characteristic of the capitalist mode of production. It is also a basic characteristic of the industrial mode of production regardless of the political system in which it is set. The distribution of manufacturing in the former Soviet Union and in contemporary China both display massive regional disparities which suggest that the mechanisms underpinning agglomeration in space and cumulative causation through time have universal validity. Manufacturing has been and remains central to the operation of these mechanisms even in postindustrial economies where the shift in employment toward services tends to divert attention away from its continuing economic significance in, for example, sustaining a wide range of producer services. Nevertheless, it has usually been the welfare consequences of industrial decline which have encouraged the adoption of regional policy and, more specifically, efforts to attract investment in new, to replace disinvestment in old industries. Regional policy in the UK, for example, had its origins in the depression of the 1930s which was experienced most severely in the industrial heartlands, such as Clydeside and Tyneside, based upon coal, steel, and shipbuilding. The essence of regional policy is that it applies only to certain areas within a country. Defining the boundaries of these areas is, therefore, a prerequisite to policy implementation. Over a period of more than 50 years, various UK governments successively modified and extended the areas qualifying for assistance from those initially identified in 1934. The assistance provided has embraced a wide range of instruments, including grants and loans to encourage capital investment, subsidies to support employment and expenditure on public infrastructure such as transport facilities. Such measures have been aimed at supporting existing firms and industries, but considerable effort has also been devoted to attracting investment into a region from other regions within the UK or from other countries by seeking to influence the location decisions of private sector firms. Similar efforts have been made by national and federal authorities in virtually all developed countries. In the US, for example, most of the poorer southern states have regarded industrial location policies as central to their economic development strategies.

Despite the widespread adoption of regional policy, its effectiveness has always been controversial. Often inspired by short term political expediency rather than any consistent long term objectives, it is difficult to measure success and failure. Most national governments have progressively diminished or abandoned identifiable packages of regional policy measures since the late 1980s in response to these difficulties and to the pervasive influence of the neoliberal orthodoxy favoring market solutions to economic development problems. This influence has been reinforced by the parallel demise of socialist state planning, embracing direct intervention in industrial location, following the fragmentation of the Soviet Union and its empire. Area based development agencies involving loose coalitions of public and private sector interests have partially filled the policy vacuum left by these trends. The areas represented by such agencies extend from local communities through subnational regions to countries and even supranational blocs such as the European Union. Thus, competition to attract industrial investment is no longer restricted to regions within the same country. Large scale, internationally mobile projects such as Japanese owned motor vehicle and electronics assembly plants have, for example, provoked locational tournaments between communities, regions, and countries. Numerous cases have illustrated the continuing importance of efforts to influence industrial location decisions by offering a wide range of incentives, including free sites, investment grants, and various forms of infrastructure provision extending from new roads to education and training facilities for potential employees. Incentive packages offered in support of large projects typically involve a multiplicity of agencies and institutions, emphasizing the complex and sometimes conflicting interdependencies between the various scales of political and economic organization. On a general level, these interdependencies highlight and reflect the role of foreign direct investment as one of the key drivers of globalization. More specifically, they indicate that the location of industry remains an important issue on public policy agendas.

See also: Agglomeration; Firms; Industrial Organization; Industrial Restructuring; Location Theory; Public Policy; Regional Development Models; Regional Development Theory; Spatial Division of Labor; Uneven Development.

Further Reading

Barnes, T. J. (2001). Re theorizing economic geography: From the quantitative revolution to the 'cultural turn'. *Annals, Association of American Geographers* 91, 546 565.
Bathelt, H., Malmberg, A. and Maskell, P. (2004). Clusters and knowledge: Local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography* 28, 31 56.
Chapman, K. and Walker, D. (1991). *Industrial Location: Principles and Policies* (2nd edn.). Oxford: Blackwell.
Commission of the European Communities (1993). New location factors for mobile investment in Europe. *Regional Development Studies* 6,
Dicken, P. (2004). *Global Shift* (4th edn.). London: Sage.
Estall, R. C. and Buchanan, O. (1980). *Industrial Activity and Economic Geography* (4th edn.). London: Hutchinson.
Fisher, P. S. and Peters, A. H. (1998). *Industrial Incentives: Competition among American States and Cities*. Kalamazoo, MI: Upjohn Institute.
Frobel, F., Heinrichs, J. and Kreye, O. (1980). *The New International Division of Labour*. Cambridge: Cambridge University Press.
Gordon, I. R. and McCann, P. (2000). Industrial clusters: Complexes, agglomerations and/or social networks. *Urban Studies* 37, 513 532.
Greenhut, M. L. (1956). *Plant Location in Theory and Practice*. Chapel Hill, NC: University of North Carolina.
Hayter, R. (1997). *The Dynamics of Industrial Location*. Chichester: Wiley.
Hoover, E. M. (1948). *The Location of Economic Activity*. New York: McGraw Hill.
Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy* 99, 483 499.
Losch, A. (1954). *The Economics of Location*. Woglom, W. H. (trans.). New Haven, CT: Yale University.
MacKinnnon, D., Cumbers, A. and Chapman, K. (2002). Learning, innovation and regional development: A critical appraisal of recent debates. *Progress in Human Geography* 26, 293 311.
Malecki, E. J. (1997). *Technology and Economic Development* (2nd edn.). Harlow: Longman.
Malmberg, A. and Maskell, P. (2002). The elusive concept of localisation economies towards a knowledge based theory of spatial clusters. *Environment and Planning A* 34, 429 449.
Massey, D. (1984). *Spatial Divisions of Labour: Social Structures and the Geography of Production*. London: Macmillan.
Moriarty, B. M. (1980). *Industrial Location and Community Development*. Chapel Hill, NC: University of North Carolina.
Nishioka, H. and Krumme, G. (1973). Location conditions, factors and decisions: An evaluation of selected location surveys. *Land Economics* May, 195 205.
Phelps, N. A. and Tewdwr Jones, M. (2001). Globalisation, regions and the state: Exploring the limitations of economic modernisation through inward investment. *Urban Studies* 38, 1253 1272.
Pred, A. (1967). Behaviour and location: Foundations for a geographic and dynamic location theory (part 1). *Lund Studies in Geography Series B* 27.
Rawstron, E. M. (1958). Three principles of industrial location. *Transactions and Papers, Institute of British Geographers* 25, 132 142.
Schmenner, R. W. (1982). *Making Business Location Decisions*. Englewood Cliffs, NJ: Prentice Hall.
Schoenberger, E. (1997). *The Cultural Crisis of the Firm*. Oxford: Blackwell.
Smith, D. M. (1966). A theoretical framework for geographical studies of industrial location. *Economic Geography* 42, 95 113.
Smith, D. M. (1981). *Industrial Location: An Economic Geographical Analysis* (2nd edn.). New York: Wiley.
Storper, M. (1997). *The Regional World*. New York: Guildford.
Storper, M. and Walker, R. (1983). The theory of labour and the theory of location. *International Journal of Urban and Regional Research* 7, 1 43.
Storper, M. and Walker, R. (1989). *The Capitalist Imperative: Territory, Technology and Industrial Growth*. Oxford: Blackwell.
Wannop, U. (1995). *The Regional Imperative: Regional Planning and Governance in Britain, Europe and the United States*. London: Jessica Kingsley.
Watts, H. D. (1980). *The Large Industrial Enterprise*. London: Croom Helm.
Watts, H. D. (1981). *The Branch Plant Economy: A Study of External Control*. London: Longman.
Weber, A. (1929). *Theory of the Location of Industries*. Friedrich, C. J. (trans.). Chicago, IL: University of Chicago.

Relevant Websites

http://www.berr.gov.uk
Department for Business Enterprise & Regulatory Reform.
http://ec.europa.eu
European Commission: enterprise, economy, and regional policy.
http://www.unctad.org
United Nations Conference on Trade and Development.
http://www.unido.org
United Nations Industrial Development Organization.

Industrial Organization

J. Patchell, Hong Kong University of Science and Technology, Kowloon, Hong Kong, China

Glossary

Geography of Enterprise The study of the influence of the policies and structures of multiproduct, multiplant enterprises on changes in industrial location and on process of regional economic development.

Monopolistic Competition A market/industry that allows competition among many firms, but within which each firm differentiates to achieve some market power.

Monopoly A market/industry dominated by one firm.

Oligopoly A market/industry in which a few firms produce a high proportion of total output.

Sunk Costs Investments that cannot be altered or recouped.

Transaction Costs The costs incurred in undertaking transactions (other than production costs).

Untraded Interdependencies The conventions, informal rules, and habits that coordinate market transactions under conditions of uncertainty.

Industrial organization is a conceptual framework for economic and especially industrial geography. Indeed it could be said that industrial geographers study the spatial organization of industry whether that be the way space organizes industry or industry organizes space. Yet industrial geographers do not define their research as the study of industrial organization (IO); this title belongs to one of the strongest subdisciplines of economics. This essay focuses on the interrelations between the two to provide insights into the orientation of industrial geography – its foundations on IO, its distinctions from it, and complementarities between the two approaches.

Sheppard defines IO as an applied branch of microeconomics concerned with effective competition and especially the influence of big business, corporate power, monopoly, and antitrust. For Carlton and Perloff, IO is "the study of the structure of firms and markets and of their interactions." The first of these definitions provides both the traditional and essential contribution, it is concerned with effective competition. IO recognizes that the market rarely takes the shape of the idealized neoclassical model of perfect competition. Rather large firms dominate and have a strategic interest in manipulating the market for their benefit – the greater the degree of monopoly obtained, that is, the greater the degree of market power, the greater the profits. There are in fact several market structures – competitive, monopoly, monopsony, oligopoly, oligopsony, and monopolistic competition. The second definition, while not rejecting the approach of the former, uses price theory to incorporate the operations of the market within and outside of the firm and therefore provides additional explanations for the different types of markets. The overall cohesiveness of the discipline is illustrated in **Figure 1**.

In the effective competition view causality runs strongest from determinants or basic conditions, through structure and behavior onto performance. Technological scale economies, for example, will strongly predetermine whether a market is competitive, monopolistically competitive, oligopolistic, or monopolistic. If the market is monopolistic firm behavior will tend to be opportunistic in the market, while oligopoly fosters collusion. Monopolistic competition, on the other hand, encourages differentiation based competition and advertising. The price theory view (incorporating transaction costs, game theory, and contestable markets) strengthens reverse causality to examine how performance influences behavior, behavior influences structure, and structure influences basic conditions or determinants. In both approaches there are strong causal relationships to public policies and other forms of governance. And crucially for both approaches, evaluation of the system – the performance of the industry/market structure – is the provision of benefits to the consumer.

This unified model while showing a recognized structure does not reflect all the differences within IO. The New Chicago School and the new industrial organization theory, for example, are forgiving of monopoly and oligopoly, claiming such a firm's position reflects efficiency of organization. The unified model of IO has had wider influences in economics and business, however. Prominently, the transaction cost theories of Coase and Williamson intersect IO and the new industrial economics. The objective of fostering pure competition has been reversed to reveal how firms can take advantage of market structure to gain market power. The power of Porter's approach has been to convert his knowledge of the structure of markets to explaining firm strategy. The economics of business and organization benefits from this conversion as well.

The impact of IO on industrial geography has been profound and pervasive. As will be traced out in the following sections, much of industrial geography (or for our purposes geographic industrial organization, GIO) is propelled by the reality of monopoly and oligopoly, explaining its spatial outcomes and searching for alternatives. In the process GIO has incorporated IO explanations such as sunk costs and transaction costs and looked elsewhere to explain organization–space relations. While GIO's spatial

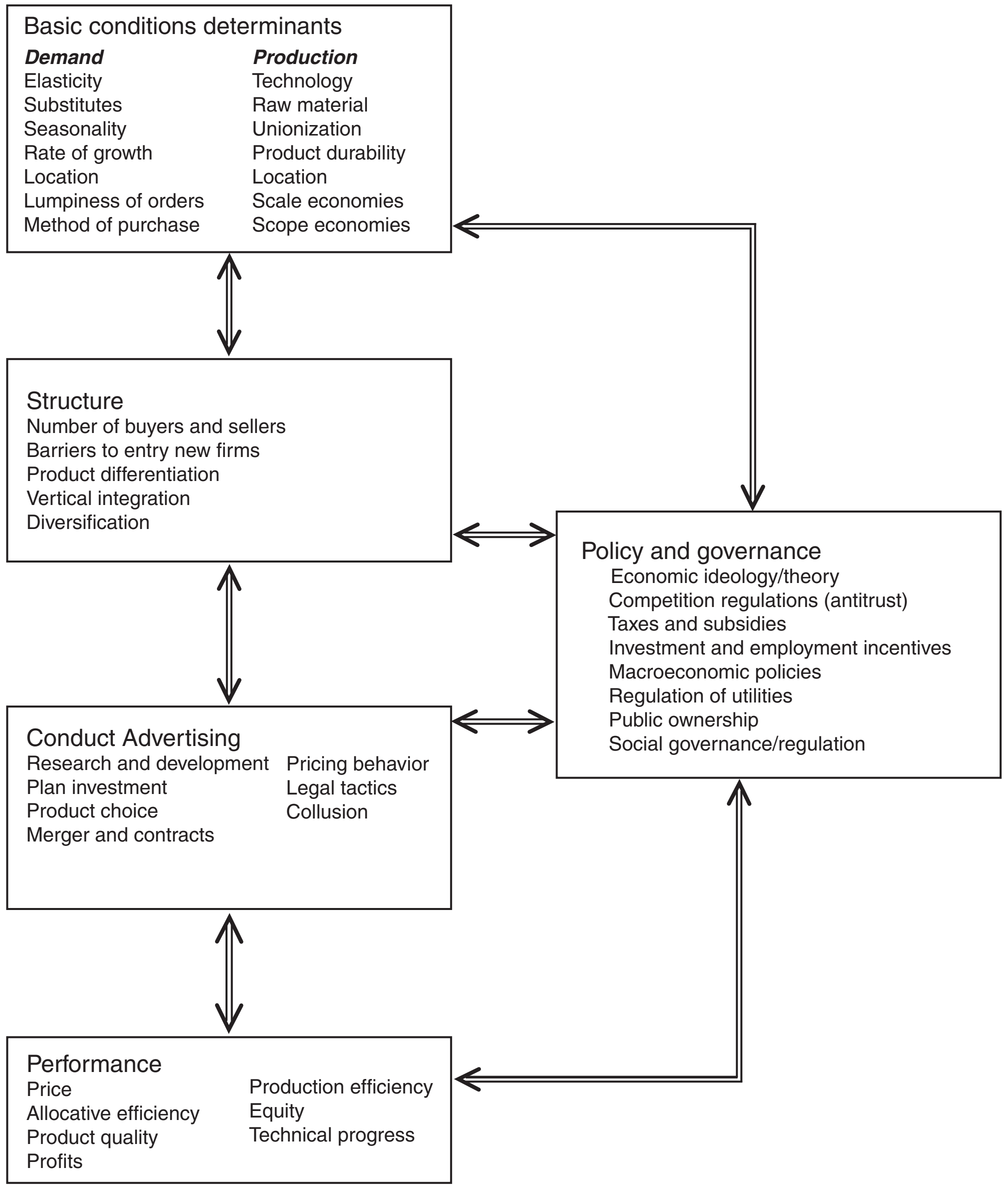

Figure 1 Structure, conduct, and performance. After Carlton, D. W. and Perloff, J. M. (2005). *Modern industrial organization*. Boston: Pearson; Sheppard, W. G. (1997). *The economics of industrial organization*. Upper Saddle River: Prentice-Hall.

focus is an obvious compliment to IO, they differ in a crucial manner. For IO theorists, the bottom line is per formance for the consumer. Geographers, on the other hand, are overwhelmingly concerned with the perform ance of industrial organization on the places of production.

The Large Firm/MNC Model

The IO and geographic industrial organization per spectives have a common origin in recognizing the limi tations of the neoclassical model of perfect competition. The fundamental base of IO is to add monopoly/mon opsony, oligopoly/oligopsony, and monopolistic com petition market structures to the pure competition model. The different structure of these markets results from the benefits of size, and therefore much of IO is concerned with how firms develop and utilize size to their advantage: exploring in particular the possible damage done to con sumers and competitors by the ability of monopolists to set prices above marginal costs or the potential for price collusion by oligopolists. As a reflection, the geography of enterprise or corporate geography (key authors including P. Dicken, R. Hayter, G. Krumme, M. Taylor, N. Thrift, and D. Watts) developed primarily as an analysis of giant multiproduct, multiplant, and multiregion or multi national corporations.

Geography of enterprise, thus, incorporates monopoly/oligopoly as its perspective on industrial organization, with the essential assumptions that these firms are able to achieve economies of scale through vertical integration and economies of scope through horizontal diversification and horizontal differentiation. The economies are obtained not only in production, but also in management, R&D, advertising, and distribution. Incorporating product cycle concepts into this theorization, geography of enterprise examined the location of different corporate activities. Markusen has produced the most comprehensive analysis of how oligopolies evolve to control a national space of production and the consequential reduction of regionally dispersed small firms. Most geography of enterprise research, however, begins with the reality of national or multinational corporations. Investigations of competitive and collusive strategies revealed locational matching by same country companies of branch plants in peripheral regions, competing branch plant location from firms in different regions/countries, and also collusive entry and carving up of regions by oligopolies to create regional monopolies. Geographers extended IO's treatment of barriers and advantages to entry by analyzing multiregional or multinational firms' establishment of new plants and distribution centers. The aspatial theorization of mergers and acquisitions and the role of vertical integration were contrasted against the spatial segregation of particularly corporate head offices and R&D.

Although the geography of enterprise can recall studies of basing point and uniform pricing strategies or point out the administered nature of internal transactions as a nod to the recognition of consumer benefits, for the most part the performance of concern is impact on places of production. Particularly by creating monopsony and oligopsony conditions in local and regional factor markets, large enterprises are capable of manipulating markets and their regulation to their benefit. There have been two overriding and related concerns. One has been to investigate the search for cheaper and more flexible labor and the capacity to influence these conditions.

In the first instance, Clark characterized firms as locating manufacturing production to gain lower costs, control the work force, and to break benefit comparison between blue collar and R&D workers. This focus on manufacturing production intersects with the social relations of production school that offered a broad critique of corporate exploitation of labor conditions in different places but did not link this to market structures. In the second instance corporations bargain with governments to obtain incentives such as infrastructure, regulatory flexibility, tax relief, legalized labor control, and so on. These incentives broach other geographical questions in relation to multipliers and negative externalities, but they also highlight the fact that GIO does not take labor conditions or government policy for granted. Rather it looks at these as malleable by corporate power.

The impacts of large corporations in shaping the structure of competitors and the social division of labor in their industry are also of concern. Fredricksson and Lindmark examined the relative internal efficiencies of large firms and small firms. Taylor and Thrift sketched out the relationships of leading firms, suppliers, and loyal opposition emphasizing the latter pairs role in production buffering and lowering wages. Christopherson and Storper introduced the idea of corporate disintegration of activities to local and distant suppliers. Around the same time the Japanese economists were explaining the quasi vertical integration of their production systems by extending Williamson's transaction cost theory. These concepts were drawn into geographical discussion albeit, focusing on the learning advantages of interfirm relations. The shift from Fordism to flexible specialization, particularly as driven by technological change, but with implications for internal and external restructuring of production was incorporated into these discussions. As another example of GIO following IO's lead in examining the impact of internal cost structures on industry structure, sunk costs have been recognized for proscribing firm organization and location options, and barriers to entry.

The geography of enterprise approach is distinguished by its use of multifaceted empirical case studies, usually incorporating interviews. A case study of the activities of an oligopolist or even of large firms within limited monopolistic competition has obvious advantages to explaining the competitive structure of an industry with few concerns in regard to representativeness. Caution has to be exercised, however, with increasing complexity and numbers of suppliers in a production system. Moreover, as indicated above the openness of the case study approach allows for consideration of the complexities and bargains inherent to labor and government negotiations, the introduction of technologies, and their interrelations. An important facet of the case study approach is to reveal the different strategies used by firms of different countries or among a country's firms.

The Industrial District Alternative

The industrial district was rediscovered in the 1980s as a potential alternative to oligopolies, indeed conceptualized as the 'second industrial divide' it was proffered to reverse this dominance. The importance of small firms and of cooperative–competitive relations in agglomerations gained further attention after Porter recast them as clusters. The ramifications for GIO, however, were to focus attention on explaining these interfirm relations with greater rigor. Geographers did so in a manner that both incorporated IO concepts and sought to overcome their deficiencies.

Scott based his reinterpretation of agglomeration on an inversion of the Coasian transaction cost argument for integrating the functions of the market. He argued that internal economies of scale and scope would likely dis integrate when market instability, varying optimal scales of operation, conditions that preclude market failures, and segmented labor markets prevailed, particularly when agglomeration conditions allowed. Recalling Mar shall, but also citing work in industrial districts in Italy and France, Scott pointed to the knowledge, familiarity, and trust that allowed firms in agglomerations to escape some transaction cost based market failures. Lorenz de veloped a complimentary view citing an industrial dis trict's capacity for providing collective goods and norms of competition. Grabher incorporated Granovetter's weak and strong ties into the analysis. Storper took the syn thesis of external economies and sociocultural per spective further by adding conventions theory to Dosi's concept of 'untraded interdependencies'.

Much of the early industrial district research and theorization focused on production where the majority of inputs and firms of a given industry concentrated with a specific region, and was criticized for the expectation of endogenous heroics. Indeed later research on industrial districts and clusters emphasized the need for 'local buzz' to be fed by 'pipelines' or other forms of interaction with the global economy. On the other hand, in the face of globalization, interest in the industrial district as a comprehensive sector of activity has given way to in vestigating clusters of functional specializations. Pre eminent among these are the networks of producer services such as advertising, finance, music, and R&D functioning through projects and networks. An important aspect of this work is to look at the interrelations of the interfirm networks with the intrafirm organization. This work coalesced with the world cities view of increasing concentration of higher ordered services.

Integrating MNCs, Functions, and Networks across the Global Value Chain

The multinational or transnational view of industrial or ganization was always an integral perspective of the en terprise/corporate geography approach. However, whereas early work had primarily focused on the dispersal of stages in manufacturing, commodity and value chain analyses broaden the GIO perspective to other industries and in corporate market influences more directly into spatial organization. This manufacturing based perspective was challenged by Gereffi's distinction between producer and market driven commodity chains and Porter's directives that value can be produced anywhere in an organizational value chain and anywhere in a spatial value chain.

The burgeoning literature on agricultural commodity chains, revealing industrial production, distribution, and marketing processes, demonstrates compatibility with GIO theorization. This field also proffers alternatives to MNC organization such as short food supply chains and various designations of origin supply chains or filières. Perhaps more than any area oligopoly/oligopsony is most evident in the world of Wal Mart and Carrefour. Wrigley has been at the forefront in drawing this reality in eco nomic geography explaining both the multinational reach of large retailers and how their sourcing and distribution differs from manufactures. In addition the increasingly multinational expansion of producer services firms, es pecially finance and business services, gather attention.

The longstanding interest of enterprise geography in firms being shaped by their environment and shaping their local environments has developed further by incorporating concepts from the industrial districts and networks litera tures into that of global production chains. The impact of multinationals on the various locations of their production chains depends on the positive and negative manners in which they engage the relational assets of a region. These relational assets include external economies of scale and scope, localization economies, and supporting institutions. It is the latter asset where much attention has been focused. Industrial geographers have complemented the IO literature by demonstrating how firms respond not only to one policy and governance regime, but they do so at different scales and in networks of these regimes. Furthermore, a major thrust of research has been to draw out the interrelations among technology, policy, and sociocultural influences.

An important new area of research is to link con sumption with production. Whereas much of GIO re search has investigated production in firms and in industrial districts primarily through the lens of product ive efficiency and without much direct link to demand conditions, researchers are increasingly investigating how demand in place influences production and how com panies influence demand in place. One approach investi gates the interactions of consumer expressivity of demand and corporate manipulation of demand in the home, the streets, and the malls. Importantly, however, this shaping of demand is traced back through the commodity chain to expose the impacts of consumption upon the social and environmental conditions at places of production. This linking of consumption and production further compli cates elucidation of multiscalar policy and governance impacts on the spatial organization of industry as often environmental, social, and economic objectives at either end of the value chain may differ significantly.

Conclusion

The fundamental impulse of economic geography's intersection with IO has been retained over four decades. IO begins with a recognition of the limitations to neo classical competition and the role of big firms in shaping the structure of markets. GIO assesses the location pattern of markets that results from the organizational strategies of big firms and how the firms shape and are shaped by each marketplace. Particularly in the 1970s and 1980s, geography directly or indirectly borrowed many concepts from IO: competition (locational matching), oligopoly collusion, barriers to entry, sunk costs, and transaction costs. The continuing relevance of these approaches is captured by Dicken's iterative appraisals in his popular *Global shift*.

However, since that period geography has not pursued developments in IO. It has not taken up the deeper significance of transaction cost economics – bounded rationality and opportunism are a rejection of the perfect rationality of neoclassical theory. Yet economic geographers generally conflate the transaction cost and neoclassical perspectives and have not attempted to deal with the implications of this radical perspective on its own terms, as for example, Ostrom does with common pool resource theory for distinctly territorial issues. Nor have

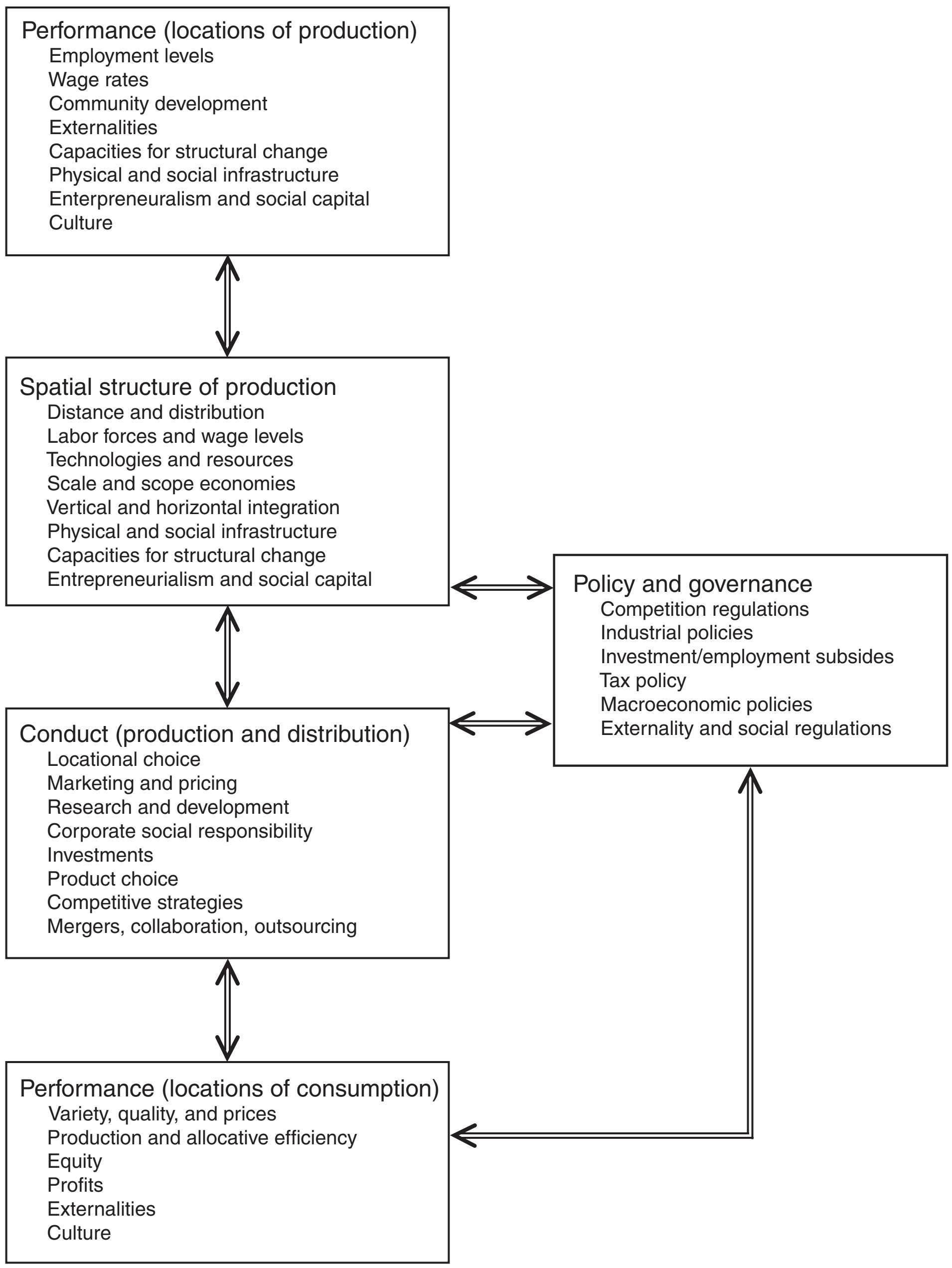

Figure 2 Geography of structure conduct performance. After Carlton, D. W. and Perloff, J. M. (2005). *Modern industrial organization*. Boston: Pearson; Sheppard, W. G. (1997). *The economics of industrial organization*. Upper Saddle River: Prentice-Hall.

geographers made much use of other advances such as Lancasterian differentiation, Akerloff's information theories, or game theory that could illuminate the supply chain perspective that has become the thread for understanding globalization.

Economic geographers have demonstrated a propensity to search for alternatives to the economizing viewpoint inherent to IO. The rediscovery of the industrial district symbolizes this search, and furthermore from the mid 1990s GIO has refocused the differences of regions and contingent institutional and evolutionary processes. This trend puts the divergence from IO in perspective, for GIO has made three notable achievements, which are sketched in **Figure 2**. First, and reaching back to its initial stimulus, GIO has reinterpreted the performance of industrial organization for the people in the places of production. For IO, the people, the resources, the environments, and the institutions of these places are simply inputs for corporate use and consumer satisfaction. Second, GIO demonstrates the diversity of policy and governance, varying vertically in levels of control within nations and horizontally among regions and nations. Furthermore, GIO has fleshed out the diversity of institutions, formal and informal, that make up this governance. Lastly, and at a more formative stage, GIO is revealing the relationship between consumer and production performance by showing the social and environmental consequences of consumption in places of production.

IO and GIO offer much to each other in terms of theory and empirical testing, and in terms of evaluating theoretical benefits against impacts on people in place. Moreover, perhaps both can occasionally revisit the impact of large firms, neither becoming infatuated with the mechanisms of the new industrial economics or the evolutionary processes of geography's institutional turn.

See also: Industrial Districts; Retail Geographies.

Further Reading

Akerlof, G. (1970). The market for 'lemons': Quality uncertainty and the market mechanism. *Quarterly Journal of Economics* 84, 488 500.

Carlton, D. W. and Perloff, J. M. (2005). *Modern industrial organization*. Boston: Pearson.

Christopherson, S. (2002). Why do national labor market practices continue to diverge in a global economy. *Economic Geography* 48, 1 20.

Clark, G. L. (1981). The employment relation and the spatial division of employment. *Annals of the Association of American Geographers* 71, 412 424.

Clark, G. L. and Wrigley, N. (1995). Sunk costs: A framework for economic geography. *Transactions of the Institute of British Geographers* NS 20, 204 223.

Coase, R. H. (1937). The nature of the firm. *Economica* 4, 386 405.

Coe, N. M., Hess, M., Yeung, W. C. H., Dicken, P. and Henderson, J. (2004). Globalizing regional development: A global production networks perspective. *Transactions of the Institute of British Geographers* NS 29, 468 484.

Dicken, P. (1994). Global local tensions: Firms and states in the global space economy. *Economic Geography* 70, 101 119.

Dosi, G., (1987). *Institutions and Markets in a Dynamic World.* Brighton: SPRU Discussion Paper No. 32.

Fredrickson, C. G. and Lindmark, L. G. (1979). From firms to systems of firms: A study of interregional interdependence in a dynamic society. In Hamilton, F. E. I. & Linge, G. J. R. (eds.) *Spatial analysis, industry and the industrial environment 1*, pp 56 77. New York: Wiley.

Gereffi, G. (1994). The organization of buyer driven global commodity chains: How US retailers shape overseas production networks. In Gereffi, G. & Korzeniewicz, M. (eds.) *Commodity chains and global capitalism*, pp 95 122. Westport, CT: Praegar.

Grabher, G. (ed.) (1993). *The embedded firm: The socio economics of industrial networks*. London: Routledge.

Granovetter, M. S. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology* 91, 481 510.

Hayter, R. (1997). *The dynamics of industrial location: The factory, the firm and the production system*. Toronto: John Wiley & Sons.

Hughes, A. and Reimer, S. (eds.) (2004). *Geographies of global commodity chains*. London: Routledge.

Krumme, G. (1969). Towards a geography of enterprise. *Economic Geography* 45, 30 40.

Lancaster, K. (1966). A new approach to consumer theory. *Journal of Political Economy* 74, 132 157.

Lorenz, E. H. (1992). Trust, community, and cooperation: Towards a theory of industrial districts. In Storper, M. & Scott, A. (eds.) *Pathways to industrialization and regional development*. London: Routledge.

Markusen, A. (1985). *Profit cycles: Oligopoly and regional development*. Cambridge, MA: MIT Press.

Ostrom, E. (1990). *Governing the Commons*. Cambridge: Cambridge University Press.

Sheppard, W. G. (1997). *The economics of industrial organization*. Upper Saddle River: Prentice Hall.

Scott, A. (1988). *New industrial spaces*. London: Pion.

Taylor, M. and Thrift, N. (1983). Business organization, segmentation and location. *Regional Studies* 17, 445 465.

Watts, H. D. (1980). *The large industrial enterprise: Some spatial perspectives*. London: Croom Helm.

Williamson, O. E. (1985). *The economic institutions of capitalism*. New York: The Free Press.

Industrial Parks

S. M. Walcott, University of North Carolina, Greensboro, NC, USA

Glossary

Eco-Industrial Park District where businesses cooperate with each other and the local community to efficiently share resources leading to economic and environmental improvements.

Economic Multiples Related financial benefits that flow from a particular economic activity.

Growth Pole Area from which economic benefits radiate out to economically enhance other related businesses in the contiguous region.

Localization Economies Efficiencies that come from shared resources by related firms in a similar industry being in the same location.

Multimodal Various types of interrelated transportation often involved in shipping goods and concentrated in one area, such as ships, trains, and trucks.

Neo-Marshallian District Concentration of firms connected to a community in a particular area; a socioeconomic unit.

Production Chain Firms and activities involved in the production of a particular good.

Social Division of Labor Distinction between different types of labor based on their skill level.

Space of Flows Place and process of transport for ideas, goods, or services.

Tacit Knowledge Information that is acquired and exchanged by verbal, nonwritten transmission.

Introduction

An industrial park consists of a piece of land designed specifically to promote industrial activities through integration with transportation facilities and other supportive infrastructure. The vision that emerged from the 1958 Dartmouth College Conference on Industrial Parks stressed the ideal compatibility of structures supporting industrial activities inside the district with the larger setting of the area outside, favoring a park like character. 'Industrial district' serves as another, broader term for the same designed development, also known as 'industrial estates' outside North America. The relationship of these three terms is based on shared characteristics of a designed concentration of firms who cluster in a particular spatial area in order to derive and create economic benefit from this locational agglomeration. Some industrial districts concentrate on a particular type of product (electronics, leather goods) to cluster production related firms.

Alfred Marshall produced the classic study of an idealized industrial district based on London in 1890. His work stressed the district's benefits of interdependency and external localization economies of scale, particularly for small and medium size enterprises, in a free market setting that otherwise favored large producers with integrated in house functions profiting from internal economies of consolidation. The advantages to firms co locating in a concentrated cluster flowed from a shared social division of labor that brought together employees with a range of skills needed for various tasks at different stages of the production chain. Managers, skilled craftsmen, routine laborers, designers, marketers, and office support staff would be attracted to a location due to the enhanced job opportunities in one place. They would also benefit from mingling together, exchanging tacit knowledge, and creating a 'thick industrial atmosphere' binding businesses with the locale. Reduction of costs and improvement of product were predicted to flow from trust building proximity.

Early industrial parks were located in or proximate to cities in order to benefit from the urbanization economies of a transportation nexus and large pool of appropriately skilled labor. The first US industrial parks began in the late nineteenth and early twentieth centuries (1899 in Chicago, 1905 in Kansas City and New York City). They increased rapidly in the 1960s following development of the interstate highway system which encouraged point to point intermodal transport of goods. Relatively higher city land costs and the movement of workers to the suburbs, along with related service components, pushed parks further to the urban fringe. Whether industrial parks are zoned areas of warehouses, business offices sharing a similarity of form and function, or factories related by production niches, their clustered nature creates a distinct landscape pattern characteristic of a concentrated economic area. Innovations in industrial processes promoted deverticalization – moving away from product assembly in one large factory in favor of parts production in various specialized companies with light assembly of components that flowed from warehouses. Distribution via large trucks needing straight roads and ample off loading space also dictated design of industrial park facilities and prioritized location with easy highway access. As indicated in the model displayed in **Figure 1**, low and long buildings typically predominate in the industrial park mix, with several office towers in an amenity setting of a 'prestige pond' accompanied by tree lined streets. Transport friendly straight roads, linking with a large arterial highway, divide the planned streetscape.

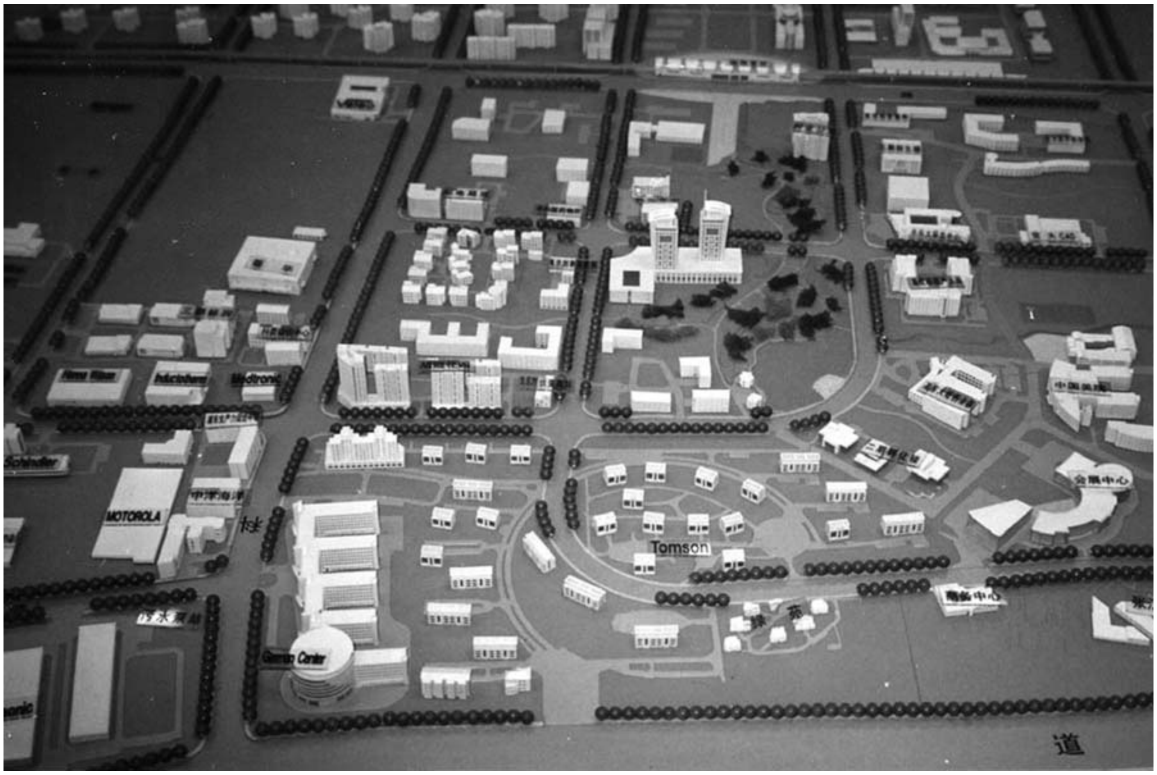

Figure 1 Diorama of Zhangjiang Hi-tech Park layout, Shanghai. The setting resembles a Western suburb with arboreal landscaping around FDI companies.

Locational Attributes: Sticky Place in Space of Flows

Research on industrial parks ranges from consideration of static locational assets such as proximity to major transportation linkages, cost of land, tax relief, access to research facilities, and other business considerations to valuing social factors such as community input and support at least as highly. This cultural turn in economic geography reflects lessons learned from previously successful industrial parks and regions that managed to turn around and advance out of a downturn, as did Silicon Valley in the 1980s. The level of cooperation required can be based on relationship ties or location ties. The 'Italian School' of geographers used case studies based in the more advanced northern Third Italy. They particularly stressed the contribution of local cultural ties to the success of agglomerated firms in certain regions – a factor that industrial parks seek to both benefit from and transcend with a footloose set of globally replicable models.

Particular components of these models vary from place to place, reflecting local strengths, development levels in the surrounding region, and management choices. The flexible specialization of 'new industrial districts' reflected the ability of firms to change products rapidly to meet emerging market opportunities, producing customized goods in small, high profit quantities. Computer aided production, as well as specialized skill sets in the social division of labor enable technically related but diverse outputs that draw varied clusters of companies to a park. Examples include a company that produces both medical devices and car parts, for example, which rely on similar electronic circuitry. Companies that produce by utilizing machine stamping molds can rapidly adapt to custom needs, enabling a variety of goods and markets to be served. Other production process innovations that encourage the growth of industrial parks on the outskirts of cities perched on a major transportation nexus include the popularity of just in time delivery. Parts needed in the production process arrive just as they are needed, reducing stationary warehouse costs in favor of prioritizing transportation flows with trucks as the new warehouses on wheels. Industrial parks cluster firms related in the process, with examples including major car (e.g., General Motors) and electronics (e.g., Motorola) manufacturers. Some concentrations are of a size to constitute a separate semi urban complex (e.g., Toyota City, Japan).

Locations utilize industrial parks to create development 'magnets and glue' – 'sticky places' containing attractive elements for a variety of types of industrial parks. A dichotomous discussion revolves around whether the most important factor in the geography of industrial parks lies in attributes of the park itself, or resides in factors adhering to the location. Is the most critical feature the local manager, university technology transfer, government regulations, trained workforce, embedded cultural attitudes, or a combination of several key in gredients? Why do industrial parks in Sunbelt southern and western locations seem more successful than those in Massachusetts, or the Midwest, for example? Industrial parks in India and China seek to learn from each other, and study the example of Singapore which manages large developments in each country.

Types of Industrial Parks and Features

Traditional industrial parks from the dawn of the Indus trial Revolution – transported by colonial capitalists to occupied territories as industrial 'estates' – lend their name to a new global incarnation as neo Marshallian nodes. These industrial parks encapsulate transnational companies – along with numerous native companies – to draw on clusters of anomalous labor and enhanced infra structure attempting to reproduce conditions at lower costs that prevail in the more developed home country. The growth of industrial parks outside traditional de veloped countries and core originating locations indicates the establishment of new nodes of production sites. These often formerly peripheral locations avoid drawbacks of concentrating in regions with relatively high labor, land, and transport costs. New areas compensate for the initial construction expenses of production and support facilities such as roads, telephone, water, and electrical connections with relatively low costs in other respects.

Other types of industrial concentrations include: (1) hub and spoke districts where companies cluster around a local giant firm that they serve; (2) the satellite in dustrial platform inhabited primarily by branch plants of foreign companies headquartered in their home nation; and (3) a state centered district featuring companies tightly tied to the national government. These globally scattered industrial park variants sustain modernization efforts in their home, as well as headquarter regions.

Another typology of industrial districts considers scalar differences. Within the science based category of technopoles – growth poles of clustered companies whose economic robustness lies in their innovative value added feature – the enterprise complex can be at the regional extent, at an urban level (special technology oriented cities in Japan, Taiwan, China, and Korea), or the inter urban scale. The level of government inter vention varies to reflect overall government involvement in national economic outcomes, or reflects particular scalar priorities such as Research Triangle Park which was initiated by the state government and anchored originally by a national entity.

Research based 'science park suburbs' reflect the in creasing size and importance of agglomerated developments that cluster in an area to take advantage of co located fac tors of production such as research universities and investor funding. Beijing's Zhongguancun, North Carolina's Re search Triangle, Taiwan's Hsinchu, and California's Silicon Valley are examples of burgeoning complexes. Global interconnections between industrial parks in different countries come from personal ties between employees who share similar national and/or school ties. Several features particular to science parks are designed to commodify lab bench breakthroughs into market successful products. These organizations include business incubators providing (ideally) instructions on how to set up and manage a business, as well as networking opportunities for nascent entrepreneurs, shared physical (electricity, telephones, computers, climate control) and human (secretarial, legal, accounting) support and overhead. Large industrial districts can include residences ranging from worker dormitories to management level amenity housing, along with tennis courts, grocery stores, conference centers, parks, and other facilities catering to economic clusters.

Some studies demonstrate that developing the human managerial side exceeds in importance the substitution of early capital outlays for physical investments in a business. Planting does not necessarily lead to growing – nurturing is crucial, and some locations as well as management models perform this function better. The industrial park model provides a spatial and structural shell, planted in various environments and inhabited by various firms and actors. Studies demonstrate that well managed industrial parks level the playing field for firms birthed as university spin offs and those created as corporate spin offs, providing both with similar managerial skill sets and experiences.

Function of Industrial Parks in the Developing World

Industrial parks concentrate scarce fiscal and human capital in the developing world, focusing efforts in a concentrated space, reducing transportation and com munication costs, and hopefully creating a growth pole with economic multiples spilling over to surrounding areas. Spaces designed to contain companies designated to locate within them also contain hopes that they will function as learning districts – places where tacit know ledge is exchanged that will advance the level of value added in products and production processes, leading to increased efficiencies and profits.

Global corporate ties through branch or more removed production chain relationships promote the geographically far flung interdependency of related park occupants. The linkage of firms to an information space of flows both within and outside the firm, district, and region promotes adaptability and thus stability at all scales. This location within a space of flows performs a particularly valuable function for firms located outside their native region. In habiting a familiar environment encourages global lo cations of operations of firms, permitting penetration of new markets designed to fit local needs and production advantages. Regions utilize industrial parks as a com petitive strategy against more advanced regions. In one example an American conglomerate created an industrial park in Vietnam built on a Japanese model but managed by Thai and Vietnamese investors who were interested in outcompeting Chinese locations to attract Western com panies. Singapore's science and industrial park model performed well enough to catapult that city state into an Asian success story, but faced strategic challenges beyond its borders. The Suzhou Industrial Park model, for ex ample, touted its national level ties but suffered setbacks from its failure to cultivate ties to the city and province authorities who started a rival industrial park. Suzhou's ambitious industrial park, spread along the edge of a large lake, included a mini city with shopping in a new down town district. Residences and British style schools served workers from the low lying warehouse style factories in a comprehensive development on the outskirts of the tra ditional Chinese city of Suzhou (**Figure 2**).

Concerns that local and small companies would be dominated by multinational giants, obliterating the ex istence of industrial districts, have not been validated. On the contrary, large companies locate representative branches in developing world industrial parks to learn from each other, group resources to learn about new markets, take advantage of park managers as interfaces with foreign governments, attract labor pools and scarce infrastructural support systems, and serve as mentors for local companies, as well as spin offs in a co pete (co operative competition) environment. The nonpecuniary interdependencies of tacit knowledge exchange and interpersonal network trust building seem especially useful in nonlocal settings, where old structures serve new functions in a globalized business environment.

Eco-Industrial Parks

Another new form of industrial parks responds to a heightened contemporary concern for reducing deleterious impacts of industrial activity on the physical environment. As defined by the President's Council on Sustainable De velopment, an eco industrial park consists of "[a] group of businesses that work together and with the community to efficiently share resources (materials, water, energy, infra structure, natural habitat and information), enhance eco nomic prosperity and improve the environment." Eco industrial parks seek to create attractive natural settings that preserve and enhance the 'green' character of the landscape, minimizing negative impact from production processes

Figure 2 Diorama of China Singapore Suzhou Industrial Park, Suzhou. Comprehensive development plans feature graduated building heights, including a new urban lakeside concentration.

involving chemical, gas, density, and transport intensive effects. General quality of life improvements through networking with the surrounding community increase the scale of park considerations.

Proponents argue that environmentally sensitive development also promotes economic efficiency in several respects. Arboreal settings with winding paths are perceived as conducive to innovative activities in research and development intensive science parks, while clean up costs from pollution are reduced in factory settings by avoiding environmental damage in the production and distribution process from the outset. Conscious cutting edge strategies include energy efficiency, waste management, material reduction, and water treatment schemes.

The variety of forms of industrial parks attests to their utility from the industrial revolution to the increasingly globalized economy and across its varied spaces of production. Shared advantages of proximity may vary within parks, but they clearly reflect the power and attractiveness of clusters by their very prevalence across the economic landscape.

See also: Agglomeration; Foreign Direct Investment; Global Production Networks; Transnational Corporations in Developing Countries.

Further Reading

Amin, A. and Thrift, N. (1992). Neo Marshallian nodes in global networks. *International Journal of Urban and Regional Research* 16, 571 587.

Asheim, B. (2000). Industrial districts: The contributions of Marshall and beyond. In Clark, G., Feldman, M. & Gertler, M. (eds.) *The Oxford Handbook of Economic Geography*, pp 413 431. Oxford: Oxford University Press.

Castells, M. and Hall, P. (1994). *Technopoles of the World: The Making of 21st Century Industrial Complexes*. London: Routledge.

Fosnsa, S., Acharib, G. and Rosso, T. (2004). A fuzzy cognitive mapping analysis of the impacts of an eco industrial park. *Journal of Intelligent and Fuzzy Systems* 15, 75 88.

Hansson, F., Husted, K. and Vestergaard, J. (2005). Second generation science parks: From structural holes jockeys to social capital catalysts of the knowledge society. *Technovation* 25, 1039 1049.

Markusen, A. (1996). Sticky places in slippery space: A typology of industrial districts. *Economic Geography* 72, 294 314.

Park, S. (1996). Networks and embeddedness in the dynamic types of new industrial districts. *Progress in Human Geography* 20, 476 493.

Phillips, S. and Yeung, H. (2003). A place for R&D? The Singapore science park. *Urban Studies* 40, 707 732.

Pratt, A. (1991). Industrial districts and the flexible local economy. *Planning Practice & Research* 6, 4 8.

Saxenian, A. and Hsu, J. (2001). The silicon valley Hsinchu connection: Technical communities and industrial upgrading. *Industrial and Corporate Change* 10, 893 920.

Scott, A. J. (1988). *New Industrial Spaces*. London: Pion.

Storper, M. (1997). *The Regional World: Territorial Development in a Global Economy*. New York: Guilford.

Walcott, S. (2001). Growing global: Learning locations in the life sciences. *Growth and Change* 32, 511 532.

Walcott, S. (2003). *Chinese Science and Technology Industrial Parks*. Aldershot: Ashgate.

Yeung, H. (2000). Organizing 'the firm' in industrial geography I: Networks, institutions and regional development. *Progress in Human Geography* 24, 301 315.

Relevant Websites

http://www.indigodev.com
Eco Industrial Park Handbook for Asian Developing Countries, Indigo Development.

http://www.indigodev.com
Eco Industrial Parks (EIP), Indigo Development.

http://www.answers.com
Industrial Park, Answers.com.

Industrial Restructuring

R. Hudson, University of Durham, Durham, UK

Glossary

Environmental Footprint All economic activity has an impact on the natural environment. This arises because of the transformation of raw materials in production, from the consumption of hydrocarbons in transportation, and from the wastes created as a result of production, transport, and consumption. These environmental impacts are referred to as the environmental footprints of a particular activity or of the aggregate economy.

Political Economy This refers to a variety of approaches to understanding economies and their geographies, including evolutionary, institutional, and Marxian variants. While these differ in important ways, they share a rejection of the assumptions and static-equilibrium approach of mainstream neoclassical economics.

Product Life Cycle Companies engage in research and development to create new products. Once such a product is launched on the market, it has a specific life cycle as the market, for it initially grows, matures, and then declines. The pace of change and the length of the cycle vary with the type of product. For example, for fashion clothing life cycles are short and change rapidly whereas for cans of beans the life cycle is much longer and changes more slowly.

Socially Necessary Labor Time The average amount of human labor needed to produce a given commodity using average technology. Companies that can restructure to use production methods that require less than the socially necessary labor time thus reap competitive advantage and prosper; conversely, those with methods requiring more than the socially necessary labor time encounter problems and if they fail to restructure to address these problems, they may fail and become bankrupt.

Spatial Division of Labor This term describes the ways in which work is divided between places, within and between individual companies, at a range of spatial scales (regional, national, international). Companies can seek to use spatial fixes – that is, searching out locations particularly suited to a given task or stage in the overall production process – as part of their restructuring strategies.

Vertical (Dis)Integration Companies can reduce their unit production costs – for example, the cost of producing an individual automobile – by integrating different stages of production within the boundary of the company and producing in-house. This is referred to as vertical integration. Alternatively, firms may engage in a process of vertical disintegration, choosing to buy goods and services from suppliers rather than producing them in-house – if this cuts production costs.

Introduction

By the mid 1980s the term 'restructuring' had become almost commonplace in economic geography and cognate social science disciplines. It denoted a 'qualitative' change from one state, or pattern of organization, to another. As such, restructuring can involve changes in any, or all, of the how, what, or where of economic activities. However, it is necessary to make a distinction between what Lovering terms "the restructuring approach" and the concept of restructuring as it is more generally deployed in a variety of political–economy approaches, although the two are clearly related. The former had its roots in Marxian political economy and was very much a product of attempts to come to grips with the spatially differentiated impacts of industrial change in the 1970s and 1980s. Lovering suggests that some commentators in the 1980s described the restructuring approach as "the structural approach," indicative of a tendency to read Marxian political economy in a particular structurally inflected way. While this was understandable in the context of contesting most of the approaches of economic geography at the time, it is, in fact, a limiting characterization of both Marxian and restructuring approaches. This change encompassed simultaneous shifts in spatial divisions of labor at spatial scales ranging from the international to the local. The latter refers a variety of approaches to understanding the constant revolutionizing of processes and products within capitalism that certainly includes Marxian approaches, but also includes consideration of evolutionary and institutional approaches to political economy. Bringing these together enables more subtle and nuanced perspectives of the changing how, what, and where of industrial production within capitalism.

A central insight of Marxian political economy is that competition for profits among companies is central to the developmental dynamic of capitalism, providing the impetus for restructuring industry and the economy via revolutionizing the how and what of production. As such, it may also involve restructuring the spatiality of the economy and what is produced where. The fundamental qualitative distinction between different forms of intercorporate competition was famously summarized by

Schumpeter: adaptive forms of competitive strategy take given constraints as binding, whereas creative strategies seek to break these constraints. This dichotomy has been echoed within recent approaches in economic geography under rubrics such as "weak"/"strong" forms, or "structure dependent"/"structure focused" domains, of competition. Although over simplistic, these dichot omous distinctions are nonetheless important in pointing to differing forms and processes of restructuring. Whereas weak strategies revolve around competing on cost for existing products within a given technological–organizational paradigm, strong strategies are based upon what Schumpeter described as market disturbing activity and product differentiation, innovation, and quality ac tivities and attributes that vary between companies, sec tors, times, and places. These distinctions between types of strategy are, however, analytic, with abundant evi dence that in practice corporate decision making cuts across this analytic distinction.

Restructuring, 'Weak' Competition and Changes in Geographies of Production: Opportunities and Limits

'Weak' forms of competition revolve around companies seeking reductions in production cost within the par ameters of an existing technological paradigm, within which all firms producing a particular product operate. These forms of weak competition echo neoclassical economists' views as the 'only' form of competition, predicated on the basis of assumptions of static equi librium and perfect knowledge. Such views of com petition are integral to the locational models of Lösch, Weber, and their adherents. Such competition can occur in many domains, but the most visible and well known direct impacts on geographies of production have re sulted from firms seeking 'spatial fixes' via relocating production, both intranationally and internationally, to places with cheaper (in terms of unit production costs) and more malleable labor to enhance or preserve com petitiveness. There are numerous well documented ex amples of such forms of spatial restructuring, especially in industries such as clothing, garment, and footwear production. Companies in labor intensive sectors exploit variations in labor market conditions as a necessary condition of their competitive survival.

While reliance upon low wages and the transfer of production to more favorable labor market locations is a very sectorally and temporally specific form of inter capitalist competition, such labor cost and labor market differences can nonetheless define the parameters for restructuring labor relations, wages, and working con ditions *in situ* under the threat of relocating unless workers agree to such changes. As part of their restructuring processes, then, companies may openly seek to set factories against one another, playing upon and exploiting the immobility of labor and people's ties to place. The threat of relocation can, in these circum stances, be as effective in meeting corporate goals as actually relocating. This became abundantly clear, for example, in 2006 as Volkswagen set about reshaping its geography of production in Europe. It shifted part of the production of the small Polo passenger car to Brussels in Belgium from Pamplona in Spain as this had become viewed as a problematic factory because of trades union activity. This move was deliberately and precisely cal culated to put pressures on the Spanish trades unions to behave more cooperatively.

However, there are also limits as the extent to which (the threat of) relocation can be used in restructuring strategies, even in activities requiring large amounts of cheap labor, because of other imperatives, such as prox imity to market. For example, many service industries must remain close to markets, despite an explosion of new low wage job categories that require little skill or formal training, but are highly market oriented. These include the armies of cleaners in city center offices, blocks, and hotels, and of private security guards and parking garage attendants. However, in some respects these limits con tinue to be eroded by technological changes. For example, in the new millennium, advertising offices in London are relocating from Soho to the East End, a move made possible by more sophisticated information technology (IT) and communications technologies.

Restructuring, 'Strong' Competition and Changes in Geographies of Production: Opportunities and Limits

Strategies of strong Schumpeterian creative competition revolve around chronic disturbance of markets by con stantly revolutionizing what is produced and how it is produced via innovations that redefine the socio organizational and technological paradigms of pro duction. This constant pressure for innovation, generating a constant flow of new products and pro duction processes, is typically seen as a progressive as pect of capitalist production.

Restructuring via Managerial and Organizational Innovation

Managerial and organizational innovations are ongoing processes. For example, historically they were integral to the emergence of vertically integrated Taylorist and Fordist mass production (discussed below), while in the contemporary economy large companies have devised new 'hybrid' organizational structures to cope with the complex challenges posed by increasingly fluid and dynamic markets. They have developed dual organizational structures to deal with core and noncore activities, respectively. Core activities are vital for the firm and the decisions to sustain them are disconnected from current make or buy decisions relating to routine noncore activities. The emergence of the 'virtual firm' may also signal a significant change in corporate organization. This brings together members of existing companies to execute a specific time limited task, effectively a process of mass customization of the enterprise linked to the emergence of 'agile' manufacturing enterprises, with the ability routinely to form virtual companies.

Restructuring via Process Innovation and New Technological Paradigms of Production

On occasion, companies generate new production processes and ways of organizing production because new products require such changes. More generally however, process innovation denotes technologically new ways of making existing products. Historically, the most significant of these are those referred to as Taylorism and Fordism. There are strong, systemic competitive pressures to find new ways of producing old commodities. Companies seek to restructure and revolutionize production processes to gain competitive advantage over their rivals. This hunt for technological rents allows the labor time needed in production to be reduced below the existing socially necessary amount. Producers who successfully use the new technology can then sell at below average prices, increase market share, and collect surplus (i.e., above average) profits on each unit of the commodity sold. They then retain this competitive edge until the new technology diffuses and becomes generalized across the producers of a given commodity (or sector), establishing new productivity norms for that commodity (or sector). Thus, the new productivity norm reduces the required amount of socially necessary labor time. This, in turn, triggers a fresh round of research and development (R&D) in search of another technological advance and competitive advantage. In this way, the process of restructuring and constantly revolutionizing production technologies proceeds and competition becomes the motor of accumulation.

This conception of restructuring and the dynamics of process innovation is specific to particular approaches to production and market conditions, however: essentially, Fordist mass production of standardized commodities sold in mass markets. In other models of production, process innovation has a differing significance. De coupling product and process technologies in flexible forms of high volume production has both further enhanced and significantly altered the importance of process innovation to competitive strategy, with significant implications for the ways in which companies seek to compete by investing in general purpose processes that are more flexible, responsive and quickly changed, and easily reused across products and product families. Consequently, companies seek to move from the mass customizing of products and services on the basis of flexible process technologies to the mass customization of process uniquely suited to their new market opportunities. There are, however, limits to the decoupling of product and process innovation, which tend to reestablish relationships between product and process innovation strongly reminiscent of those characterizing mass production.

Process innovations can have direct implications as to where production occurs. Conversely, spatial variations in labor market conditions can shape which sort of process technologies and competitive strategies are deployed in particular locations. These locational implications arise for two reasons. First, companies using established and old technologies and/or organizational forms may become bankrupt; or selectively close some plants; or relocate production to places in which labor market (or other) conditions permit the old technology to be competitive with the new one. Second, companies using the new technology may need to find fresh locations in which it can be profitably deployed because of labor resistance to changing working practices and forms of work in long established centers of production. Typically, this could involve moving to areas in which 'green' labor is available. Such labor lacks knowledge and experience of 'existing' norms and practices of production for these commodities and thus will accept the conditions associated with the new technology. Conversely, in other circumstances, the importance of experienced and knowledgeable labor can tie production to particular places as these develop the endogenous capacity to generate new processes (and products) which ensure a competitive edge. Such technology districts are places in which economic growth occurs on the basis of *in situ* restructuring and product based technological learning.

Restructuring via Product Innovation

Restructuring via product innovation as a route to competitive advantage is an ongoing process. Product innovation can take varied forms. It can relate to enhancement of existing, often seemingly mundane, products (e.g., Gillette's creation of new types of razor in the last two decades) or to the creation of qualitatively different products within the same market segment as companies strive for a competitive edge via further segmenting mature markets (e.g., the creation of DVDs to replace CDs). Incremental innovation is increasingly important in allowing changes in response to consumer demand in more differentiated markets for both manufactured goods and services. Creating totally new products is of the greatest significance, however, and involves the con struction of new markets rather than a zero sum game as to the allocation of shares of a given market with, therefore, much greater potential to set in motion radical re structuring. Such product innovation enables companies aim to create new markets in which they can be sole and monopolistic producers. Companies seek to protect their interests and intellectual property through devices such as patents. However, given the pace of technological change, performance superiority is often short lived.

The systemic tendency of capitalist production to gen erate new products, in turn, raises further issues. First, what is the relationship 'between' research and development? Taylorist mass production is based on a linear sequential model of R&D creating breakthrough product innovations and then producing them in long runs of standardized products. In more flexible systems of high volume pro duction, the processes of product innovation and develop ment are structured differently. In particular, there are three distinctive organizational features: self organizing develop ment teams; overlapping development phases, replacing a segmented with a shared division of labor; and a commit ment to continuous learning. Changing the form of product innovation within such flexible forms of production – at the limit producing batch sizes of one, customized to the re quirements of the individual purchaser of – say – an automobile in deeply fragmented markets severely cir cumscribes the utility of the concept of the product life cycle. Competitive success depends upon being the first to identify an emerging market, a segmenting market, or a change in consumer requirements in a previously un changing market. Success depends upon minimizing prod uct development and product changeover times, not upon being the lowest cost or highest throughput producer. This is linked to increasing volatility and ephemerality in fash ions and tastes as product life cycles for many commodities become increasingly brief, even transitory as when the commodity is a performance or spectacle.

The growing emphasis on networked approaches to production organization has also influenced product in novation and leads to a reappraisal of the linear R&D model, in particular, in two respects. First, the emergence of new forms of interfirm network relationships is influ encing the organization of product innovation. Quasi vertical disintegration can help facilitate coordinated de sign and investment. As a result, the product development process, from planning and conceptualization through designing and engineering, to manufacturing, is more re flexive, consisting of overlapping and recursive loops ra ther than discrete linear stages. Second, incorporating network concepts into intrafirm organizational structures has also influenced processes of product development via highlighting the importance of information flows and interaction within and between all sections of a company. Nonetheless, it is important to remember that human creativity remains central to product innovation, however organized and whatever model of production is deployed. This raises important questions about creativity and the creation of new knowledge in production restructuring.

The second issue concerns where new products are produced – in new or old firms, in new or old industrial spaces, in previously unindustrialized areas or in areas with an industrial history and tradition? – and the factors in fluencing locational choice. In some circumstances, there is a close relationship between product innovation and new firm formation in particular locations. In certain respects, the volatility and ephemerality of products is enhancing the significance of particular places as centers of product innovation. Industrial districts, producing a range of com modities, have often been centers of product innovation based on intense linkages and sharing of knowledge be tween firms and other institutions. For example, product innovations in banking and financial services have become closely linked to London, New York, and Tokyo – recog nized centers of innovation, interpretation, and interaction. In the late twentieth century, in new industrial spaces of California, new firm formation was closely tied to product innovation in the IT and computing industries. New product pioneering firms were spun off from existing ones, deepening both horizontal and vertical divisions of labor and the transactional tissue of the local industrial complex.

Restructuring via Market Creation and Marketing Innovation

According to Best, firms can succeed by identifying and "responding to" changes in markets or consumer prefer ences. As such, Best perhaps overestimates the extent to which firms simply respond to exogenously given changes in market demands rather than actively seeking to change consumer tastes and cultivating preferences for new products via advertising, as part of Schumpeterian market disturbing competitive strategies. Because consumers possess – and can 'only' possess – imperfect knowledge, they are susceptible to influence via advertising. However, although advertising has a venerable history, until quite recently it has only been a marginal influence on patterns of sales and production. The development of modern advertising – or put another way, the restructuring of the advertising industry via radical product and process in novation – was a central element in corporate strategies to create, organize, and, where possible, control markets, es pecially for mass produced consumer goods such as automobiles, TVs and other electronic goods, washing machines, refrigerators, and other household white goods. Mass production necessitated mass consumption, and this in turn required a certain homogenization of consumer tastes for final products. The period between 1880 and 1930 saw the full development of an organized system of commercial information and persuasion, as part of the modern distributive system via more conscious and serious attention to the psychology of advertising and changes in advertising media. Indeed, modern capitalism could not function without advertising which offers mass produced visions of individualism.

It is impossible to fully grasp the significance of advertising without realizing that the material object being sold is never enough. Thus, even commodities providing the most mundane necessities of daily life must be imbued with symbolic qualities and culturally endowed meanings: this, indeed, is the crucial cultural quality of advertising's modern forms. Commodities thereby meet both the functional and symbolic interests of the consumer via the 'magic system' of advertising. The competitive pressures of contemporary capitalism have stimulated further refinements to this magic system, further enhancing the importance attached to culturally endowed and symbolic meanings and the identities that people (in part) form through consuming commodities. This has been linked to further product and process innovations in the form of advertisements, as the advertising industry itself has experienced further deep processes of restructuring.

The power to shape demand, to create, disturb, and radically restructure markets, has become an increasingly important part of competitive strategy as product life cycles shorten and/or markets become increasingly fragmented and segmented and/or the boundaries between goods and services become increasingly blurred. The fusion of goods and services, redefining what constitutes the product, and the consequent mutation of markets are undoubtedly the most significant aspects of the new economics of production. But the point to emphasize is that such changes are a result of proactive strategies of strong competition, not simply a weak competitive response to exogenously given changes.

Restructuring Territorial Economies

Changes in the location of activities are clearly an important strand of corporate restructuring. Consequently, restructuring can also relate to the issue of territorial restructuring and the restructuring of spatially defined economies at various scales. Such restructuring is integral to processes of combined and uneven development, the connection between growth in some places and decline in others. The movement of capital as a part of corporate restructuring can open up new developmental possibilities for some places but, equally often, lead to socioeconomic devastation in other places as a result of capital flight and capital investment being suddenly and prematurely written off. One response to this has been the development of state economic developmental strategies to restructure national and subnational economies (regional, urban, etc.) in particular ways in relation to desired developmental outcomes. As one set of economic activities declines *in situ* or takes flight from a place, state agencies seek to shift those territorially defined economies from one developmental trajectory to another, better, and higher one, in response to a range of sociopolitical pressures and demands and to seek to maintain their authority and legitimacy.

Such pressures can be particularly acute when the proximate cause of place specific economic decline is the decision to redefine the boundaries between economy and state via privatization, leading to locationally concentrated capacity and closure and job loss. State strategies to restructure the economies of economically stricken places, such as northeast England or Cape Breton in eastern Canada, involve seeking to entice, via a variety of means, companies that are either establishing themselves for the first time or, more often, already established companies that are seeking to restructure the geographies of their own operations and attract them to particular places. The aim is to create synergies between the company and places – ideally 'win win' outcomes – enhanced profits for the company, new jobs and sources of tax revenue for the place.

The pursuit of private sector investment has often involved a restructuring of the state itself, not least via territorial restructuring and the devolution of power from national to subnational levels, which itself has often involved the construction of new forms of state organization. There are limits to the capacity of states in countering localized economic decline, however. Three reasons are of particular significance. First, there are processes internal to the logic of capitalist development that draw activities to successful centers of growth as a result of agglomeration and urbanization economies. Second, the spatial reach of economic activities typically exceeds the territorial jurisdictions of states, especially as economic activities have become increasingly globalized. And third, economic crisis tendencies are internalized within the state rather than abolished by state involvement, which in turn strongly influences the form that state policies and practices can take and severely limits the capacity of states to do anything other than go with the flow of the logic of capital. This is revealed – *inter alia* – in the shift from plans with definite targets to markets and strategies that respond to market movements as steering mechanisms.

Eco-Restructuring

In recent years, there has been growing recognition that industrial production is predicated upon processes of material transformation that have unavoidable impacts on nature and the natural environment. As concerns with the ecological impacts of industrial production, further heightened by the emergence of China and India as major centers of industrial production in a further twist to the evolving international division of labor, have intensified, there has been a growing interest in the concept of eco restructuring. Eco restructuring involves seeking to move capitalist economies onto more ecologically and environ mentally sustainable trajectories and as such has strong normative and moral connotations. Proponents of this viewpoint argue that via a judicious combination of in novations in product, process, and transport technologies, steered by appropriate state policies, the environmental footprint of industrial production, and in particular the impacts of enhanced carbon dioxide emissions to the at mosphere, can be lessened and kept within sustainable boundaries. It is recognized that this will necessarily in volve significant changes in lifestyle and cultural shifts in attitudes towards production, mobility, and consumption and not just changes in hard technologies. In this sense, this is a move forward from the technological optimism of earlier times but it remains to be seen whether industrial capitalism can be successfully eco restructured to prevent global environmental catastrophe.

Conclusions

Restructuring refers to radical qualitative rather than just incremental quantitative change and innovation in product, process, and organizational forms in economies and economic geographies. It can also refer to changes in territorial economies at a range of spatial scales. Re structuring is endemic to capitalist economies and their practices and will necessarily remain so as long as the economies remain capitalist. Consequently, restructuring remains an important concept in understanding econ omies and their geographies. It also remains important to policy makers seeking to steer economic change and development in particular ways, both in terms of terri torial development and ecological sustainability, and not least in recognizing the limits to their capacity to bend the strategies of capital in the directions that they – and/or those seeking to influence them – see as desirable. Finally, we can note that while the theoretical space exists in which to consider more radical forms of re structuring and moves from capitalism to other modes of political and economic organization, there is little prac tical exploration of this space – although that is not to say that capitalist forms of restructuring are not contested by a variety of social actors.

See also: Industrial City; Industrial Districts; Industrial Location; Industrial Organization; Industrial Parks; Industrialization.

Further Reading

Best, M. (1990). *The new competition: Institutions of industrial restructuring*. Cambridge: Polity Press.
Boddy, M. (1986). Structural approaches to industrial location. In Lever, W. (ed.) *Industrial change in the United Kingdom*, pp 56 66. London: Longman.
Clark, G. (1992). Real regulation: The administrative state. *Environment and Planning A* 24, 615 627.
Donoghue, M. T. and Barff, R. (1990). Nike just did it: International sub contracting, flexibility and athletic footwear production. *Regional Studies* 24, 537 552.
Ewen, S. (1976). *Captains of consciousness: Advertising, the social roots of the consumer culture*. New York: McGraw Hill.
Ewen, S. (1988). *All consuming images*. London: Basic Books.
Frobel, F., Heinrichs, J. and Kreye, O. (1980). *The new international division of labour*. Cambridge: Cambridge University Press.
Harvey, D. (1982). *The limits to capital*. Oxford: Blackwell.
Harvey, D. (1989). *The condition of postmodernity*. Oxford: Blackwell.
Hudson, R. (1986). Nationalised industry policies and regional policies; the role of the state in the deindustrialization and reindustrialisation of regions. *Society and Space* 4, 7 28.
Hudson, R. (2001). *Producing places*. New York: Guilford Press.
Hudson, R. (2005). *Economic geographies: Circuits, flows and spaces*. London: Sage.
Hudson, R. (2006). On what's right and keeping left: Or why geography still needs Marxian political economy. *Antipode* 38, 374 395.
Lajtner, S. and Milne, R. (2006). VW shakes up polo production. *Financial Times* 18 May, pp. 24.
Lash, S. and Urry, J. (1994). *Economies of signs and space*. London: Sage.
Lipietz, A. (1980). The structuration of space, the problem of land and spatial policy. In Carney, J., Hudson, R. & Lewis, J. (eds.) *Regions in crisis*, pp 60 75. London: Croom Helm.
Lovering, J. (1989). The restructuring debate. In Peet, R. & Thrift, N. (eds.) *New models in geography* vol. 1, pp 198 223. London: Unwin Hyman.
Mandel, E. (1975). *Late capitalism*. London: New Left books.
Massey, D. (1995). *Spatial divisions of labour* (2nd edn.). London: Macmillan.
Nonaka, I. and Teece, D. (eds.) (2001). *Managing industrial knowledge*. London: Sage.
Noteboom, B. (1999). Innovation, learning and industrial organisations. *Cambridge Journal of Economics* 23, 127 150.
Pine, J. (1993). *Mass customization: The new frontier in business competition*. Harvard: Harvard University Press.
Sabel, C. and Zeitlin, J. (1997). *Worlds of possibilities*. Cambridge: Cambridge University Press.
Sassen, S. (1991). *The global city: New York, London, Tokyo*. Los Angeles: University of California Press.
Schumpeter, J. (1961). *The theory of economic development*. Cambridge: Harvard University Press.
Storper, M. (1993). Regional 'worlds' of production: Learning and innovation in the technology districts of France, Italy and the USA. *Regional Studies* 27, 433 456.
Storper, M. and Walker, D. (1989). *The capitalist imperative*. Oxford: Blackwell.
Veltz, P. (1991). New models of production organization and trends in spatial development. In Benko, G. & Dunford, M. (eds.) *Industrial change and regional development: The transformation of new industrial spaces*, pp 193 204. London: Belhaven.
Weaver, P. M., Jansen, L., van Grootveld, G., van Spiegel, E. and Vergragt, P. (2000). *Sustainable technology development*. Sheffield: Greenleaf.
Williams, R. (1980). *Problems in materialism and culture*. London: Verso.

Industrialization

D. Simandan, Brock University, St. Catharines, ON, Canada

Glossary

Ideographic Epistemological perspective that focuses the research process on the detection of the uniqueness of each geographical location. It denies the nomothetic belief in the existence of universal patterns, and thus portrays reality in descriptive rather than explanatory terms.

Nomothetic Epistemological perspective that focuses the research process on the uncovering of universal patterns, regularities, or laws. It perceives reality as the interplay of 'surface characteristics' and 'deep structure' and attempts to explain the former in terms of the latter.

Industrialization: Definitions and Measurement

'Industrialization' is a generic name for a set of economic and social processes related to the discovery of more efficient ways for the creation of value. These more efficient ways are lumped together under the label 'industry' or 'the secondary sector' (the primary sector of economic activity referring to agriculture, hunting, fishing, and resource extraction, and the tertiary sector referring to services). Beginning with the late seventeenth century, industrial activity has dramatically enlarged its scope and scale, as machinofacture began to replace manufacture. Historically, industrialization studies have concerned themselves primarily with the period known as the Industrial Revolution, although in geography this area of enquiry has been the focus of many economic geographers interested in the contemporary logic of the global economic landscape.

Using the criterion of the abruptness of change, one can distinguish two types of economic change: events (swift singular changes) and processes (protracted cumulative changes). Industrialization is a process, not an event. A process is an emergent property of a system (country or region) resulting from a collection of events that share a number of similarities and that unfold over a slower timescale than that of its component events. If one entrepreneur opens an industrial plant in an otherwise agrarian region, that singular event cannot be labeled as industrialization. If a collection of events of this same kind achieves sufficient significance for the local economy, scholars and policymakers alike are entitled to refer to it at a higher level of generalization, that is, they can speak of a process of industrialization changing the face of that regional economy. Two further conceptual problems require specification in this context.

First, one must distinguish between quantitative economic growth and qualitative economic change. If an already industrial region witnesses the opening of some new industrial plants, it is inappropriate to label that set of events as industrialization. Instead, we must refer to it as simply continuing industrial growth or economic growth. The concept 'industrialization' should be restricted to the qualitative economic change occurring whenever an agrarian economy becomes to such an extent affected by the opening of new industrial plants that it becomes misleading to continue referring to it as an agrarian economy. In other words, industrialization is an emergent property of an economic system, a qualitative leap resulting from the spatially patterned aggregation of a collection of economic events.

Second, one must pay attention to the measurement of the threshold above which it becomes appropriate to speak of a process of industrialization taking place. If one singular industrial event does not by itself constitute a process of industrialization, when does it make sense to refer to such a process? Geographers, economists, historians, and sociologists of industrialization have been rather casual in their approach to this measurement problem, relying on commonsense just as much as on specific quantitative cutoff points. There are three major ways to decide as to whether one national or regional economy is undertaking industrialization. The first requires a comparison of the relative contribution to the gross domestic product of the secondary sector (manufacturing industry) versus that of the primary sector (agriculture, fisheries, hunting, and extraction of raw materials). The second compares the percentage of the workforce employed in industry versus agriculture. The third is more subjective, but also more geographical, in that it assesses the extent of industrialization by simply observing the landscape of a region. Since industrial activities necessitate a drastic change of the physical landscape (e.g., fixed capital in the form of built environments), they are easier to spot than more subtle social processes such as exploitation, racism, or social stratification.

Explaining and Timing Industrialization

The fundamental question geographers have to ask is whether the concept industrialization has contributed in any substantive way to furthering scientific enquiry into how our social world works. There is probably no easy way to answer in any meaningful form such a question, because the answer given would depend on the level of explanation at which the concept industrialization is being deployed. It has been argued that industrialization has been particularly fruitful in helping geographers and social scientists operate at the higher levels of explanation, or, at what might be termed 'big picture' thinking. Scientists and lay people alike relate to the world by building a more or less detailed and accurate model of what the world consists in and of how it works. At a very general level, it becomes fertile to have an understanding of how the history of the world has unfolded, and such an understanding would need to include the saga of industrialization. As an illustration of 'big picture' thinking, we can consider Alvin Toffler's depiction of the course of history as a succession of three waves. The first wave refers to the shift from hunter gatherer societies to agricultural, sedentary societies. The second wave refers to the relative decline of agriculture and the growth of industrial activities. Finally, the third wave designates the shift from industry based economic growth to service based economic growth, and the relative decline of blue collar workers in favor of white collar workers.

Historians of industrialization have pointed out the fact that the timing of this process is crucial for understanding its nature. In particular, they identify three periods of industrialization: the first refers only to England and pertains to historical contingencies between 1763 and 1846. The second includes countries such as USA, USSR, Germany, and Japan, which became industrialized in the nineteenth century and the beginning of the twentieth century. The third refers to the countries that have started their industrialization after World War II (e.g., the tigers and dragons of Southeast Asia). The important observation in this context is that all other countries except England have had at least some other model of industrialization which they could imitate and emulate. England is unique in that there industrialization appeared spontaneously, unplanned, from scratch, through a set of economic initiatives that only in retrospect have been labeled 'industrialization'. The theories invoked to explain the English Industrial Revolution have not ceased to proliferate and to take into account hitherto ignored factors such as genes. Since for all other cases of industrialization the imitation factor has played a role, it follows that the geographical study of innovation diffusion is a required step in any serious attempt to make sense of this process.

Geography pretends to be a scientific endeavor and the hallmark of a scientific endeavor is the attempt to explain and predict phenomena. To explain something means to uncover the law abiding mechanism that caused it. Scholars of industrialization have fallen short of this task, even though their work has converged on admitting the complexity of this process. There are several interlocked problems that together keep industrialization in the clouds of ambiguity. At the most general level, industrialization is a social process, and epistemologists of the social sciences have cast doubt over the feasibility of explanation in the social realm. The innumerable variables that contribute to social outcomes do not seem to allow the social sciences to aspire to the same level of explanatory rigor as the natural sciences. Therefore, a more modest goal would be to understand rather than explain the process of industrialization. Understanding results from describing and comparing the various historical and contemporary contexts in which industrialization has occurred, without assuming that there is a law abiding mechanism through which industrialization necessarily emerges. The description and comparison of the aforementioned contexts allow researchers to detect both the nomothetic and the ideographic components of industrialization. The nomothetic components refer to those general facets of industrialization shared by all the various contexts in which it has occurred, whereas the ideographic components capture the unique, particular features that have stamped industrialization in a specific context.

A related insurmountable obstacle to the explanation of industrialization is that it is not possible to experimentally test and refute the various theories attempting to account for this phenomenon. Karl Popper made a forceful case for the idea that theories are scientific only to the extent that they are refutable. The problem with the scholarship on industrialization is that one can always invent a plausible 'just so' story and propose it as the explanation for this process, without having to subject it to the risk of experimental refutation.

There is no single cause of industrialization. The process can emerge from a variety of causes. Similarly, the consequences of industrialization vary widely across geographical regions and historical times. In order to grasp these ideas in all their complexity, it is worth disentangling and studying the relations between the often confused concepts of capitalism, modernization, and industrialization.

Industrialization and Capitalism

Capitalism is a mode of production. A mode of production is a particular way of organizing the economy and of allotting the costs and benefits of economic activities. Economic historians have identified modes of production other than capitalism (primitive communism, slavery, feudalism, socialism, and advanced communism) and economic geographers have aptly noted that elements from these other modes of production can coexist, somehow 'etched' in the fabric of the dominant mode of production today – capitalism. There is no relation of logical or causal necessity between capitalism and industrialization. This means that the two concepts do not entail one another either logically or causally. In plain language, capitalism does not necessarily lead to industrialization (although it has often been considered as a favoring factor for industrialization, especially in the scholarship on the first Industrial Revolution in England). Furthermore, industrialization can and has happened in noncapitalist regions (e.g., Stalin's Soviet Union, 1924–53; Mao's China; and Ceausescu's Romania). To look at these relations the other way around, it is worth noticing that industrialization does not necessarily lead to capitalism (see the cases of Cuba, China, or North Korea today) and that industrialization is not a necessary condition for the emergence of capitalism (e.g., Third World countries may have a capitalist economy based on agricultural export oriented monocultures or on tourism). Statisticians' urge to remember that correlation does not imply causation is therefore particularly relevant when studying the relation between capitalism and industrialization: both across historical times, and across geographical spaces the two economic processes tend to go together. At first glance, one could speculate that they are mutually reinforcing processes, although counter arguments to this hasty speculation can also be easily conceived.

To understand the issues involved, note that the most prominent argument for the virtues of capitalism consists in the neoclassical economic theorizing of free markets as best means for the efficient allocation of scarce resources to many needs. That argument, in turn, depends on the assumption of atomistic (innumerable) economic agents forced to coexist and fight with one another in a condition of perfect competition (i.e., none of them is powerful enough to be sheltered from competition). In other words, the alleged virtues of free markets collapse if the assumption of perfect competition is severely put into question by economic realities. The process of industrialization systematically does exactly that: on the one hand, technologies (one type of fixed capital) for the industrial process become yet more expensive (because they embody more and more knowledge), and this need for larger initial investments of capital encourages the concentration of capital in fewer hands (monopolies or oligopolies); on the other hand, the need for economies of scale acts as a catalyst for the further integration and concentration of capital.

However, it is not only the case that the logic of industrialization can subtly move capitalist realities far away from their idealized virtues: as Marxist geographers have amply documented, the logic of capitalism renders industrialization a very fragile achievement. Just as the logic of industrialization favors concentrations of capital, which in turn undermine the free market conditions of healthy capitalism, so too the logic of capitalism favors the geographical relocation of capital, which in turn undermines industrial activities in old industrial regions. In the initial stages of the industrialization of a new region, the prospects of continuous growth seem safe and sure. Nevertheless, as time goes by, there is a tendency for the rate of profit of local capitalists to fall because of factors such as exhaustion of raw materials, new competitors entering the market, saturation of the market, increasing rent, new taxes (e.g., green taxes to internalize environmental externalities), increasing cost of labor because of unionization, etc. Since the logic of capitalism is the making of profits for profits' sake, the local capitalists can choose to close the now unprofitable local plants and reinvest their money elsewhere, in regions where they can make higher rates of return on their investment. These new regions benefit from industrialization, whereas the older ones suffer the costs of the opposite process – deindustrialization.

In a long term perspective, three observations become self evident: the first is that, any apparent beneficiary of capitalist industrialization has its prosperous days counted before turning into one of capitalism's victims. Sometimes, these victims, because they are victims (i.e., high unemployment, therefore oversupply of labor, therefore cheaper labor) might attract new rounds of capitalist investment. Second, from a spatial perspective, industrialization and deindustrialization are processes that together express the historical geography of capitalism, its highly dynamic and uneven nature that so much impressed Marx and his followers. Third, from a political perspective, it becomes clear that the state has a crucial regulatory role to play in deciding the fate of industrialization. Many of the contemporary struggles over economic globalization emerge precisely because of the conflicts of interest between states and capitalists: the former care for economic development, high employment, and national flourishing, whereas the latter obey the different (transnational) logic of capitalist competition.

Industrialization and Modernization

As for the relation between modernization and industrialization, their conceptual knot is more difficult to untie. Modernization is the historical process whereby the social relations governing a human community have shifted from being based on kinship, tradition, collectivism, and magic/religion, to being based on rule of law, rationality, individualism, and scientific knowledge. The trick is to understand not only that modernization can be one of the unplanned effects of spontaneous industrialization, but also that political leaders keen to modernize their countries can deliberately use industrialization as a very potent means to that cherished end.

To give an example, Erving Goffman's account of the moral career of stigmatization and his discussion of 'tribal' (i.e., collective) stigma, together with Karen Horney's work on collective neuroses have recently been deployed by geographers to shed light on the political performance of inferiority complexes and on their role in constituting marginal territories as 'in need of industrialization'. Specifically, Simandan showed how Romanian and Norwegian intellectuals educated in Paris in the nineteenth century, upon their return back home, put their countries on the path of industrialization and modernization by attempting to copy what they saw in France. Industrialization played the role of a difference reducing engine between a deplorable economic backwardness and an intense longing for modernization. The positive stories of industrialization have played a major role in the invention of the modern nations, as an analysis of media and of general education school texts would clearly reveal. They have occupied a privileged position in the 'spaces of constructed visibility' where the imaginary geographies of the nation are fabricated (exhibitions, museums, textbooks, traditional and electronic media, and public ceremonies). Nevertheless, and of paramount significance, these stories have helped fabricating 'the modern fiction' of the nation, and not just some timeless collective identity.

Until recently, the parties concerned with the development of the Third World countries have encouraged their industrialization as a way out of backwardness and poverty and toward civilization, modernity, and prosperity. The problem with this encouragement is that it is value laden: it implies that the values of modernity (which happen to be the values of Western culture) are superior to the values of traditional cultures in the Third World, which is akin to saying that the Westerners are superior to the 'primitives'. Since Western scholars and politicians alike have publicly rejected the older assumptions of Western superiority and have embraced an egalitarian worldview according to which all cultures are equally valuable, the ongoing advocacy of industrialization as a solution for the Third World appears deprived of its most entrenched rationale. To propose a solution implies that there is a problem, and if backwardness is not the problem, than what is it? These thorny questions could be left to the pondering of critical geographers and postdevelopment studies scholars and we could turn instead to the weighting of the overall costs and benefits of industrialization.

Costs and Benefits of Industrialization

Industrialization's legacy of delegation of responsibility for environmental problems has been well documented by geographers and industrial ecologists. The internalization of its negative environmental externalities remains an inconsistent and poorly enforced practice in many parts of the world. Furthermore, the deeper question of the limits of natural resource substitution has received sustained attention only from a few specialists, despite the fact that the fate of industries is written in the answer to that question.

Unlike the developed countries who became industrialized before World War II (England, USA, USSR, Germany, and Japan), the Third World countries who are currently trying to start or speed up the process are confronted with the lack of sufficient local capital. This means that for them industrialization can come only at the cost of increasing dependence on foreign capital. If they choose to specialize in export oriented industrial production instead of import substitution industrial production, this foreign dependence for capital is further amplified by a dependence on volatile and competitive foreign markets. Furthermore, given that current international economic policies set by the World Bank, the International Monetary Fund, and the World Trade Organization implicitly or explicitly support national economic specialization (Ricardo's principle of comparative advantage writ large), the least developed countries in the world are pressured to participate in a global economic gamble in which their odds of winning are very long indeed.

Aside from modernization, the other most frequently invoked benefit of industrialization is economic development. The problem – as some Third World countries have found out – is that industrialization does not necessarily lead to massive economic development. Let us clarify the concepts involved. Economic growth refers to a quantitative increase in the gross domestic product of a country. Economic development refers to a qualitative structural change in a given economy. If a given country or region has some industrial plants specialized in the production of consumer goods and/or is totally export oriented, it runs the risk of witnessing economic growth without economic development. The respective industries are not organically embedded in the regional or national economy and play the role of the cherry on the cake instead of playing the more ambitious role of the yeast that makes the whole cake grow. This latter role usually is performed by capital goods industries, i.e., those industries that produce equipment needed for the development of other industries. The lesson to be gleaned from this brief analysis of economic growth versus economic development is that whether industrialization is beneficial or not critically depends on what kind of industrialization one is speaking about.

Scholars and policymakers have also argued that industrialization is the best way to fight excessive population growth in the Third World. Rural dwellers tend to have very large families partly because they are less educated than urban dwellers, and partly because for them children are a source of wealth and security in old age. The process of industrialization leads to increased urbanization, increased general level of education, and increased income, all of which contribute to changing cultural and demographic patterns in the direction of massively reduced fertility rates. In statistical parlance, the impact of industrialization on fertility rates is mediated by the variables urbanization, education, and income.

Geography and Industrialization

The relation between the study of industrialization and the discipline of geography can be decomposed into: (1) an analysis of how the tools of geography enhance our understanding of industrialization, and (2) an analysis of how the interdisciplinary research of industrialization can add depth and context to the traditional concerns of economic and historical geographers. Geography is a generic name for a set of various scientific practices loosely held together by a common concern for the big themes of 'space' and 'society–nature relations', as well as by the networks generated through its having a distinct position in the academic division of labor. In other words, various texts count as geography to the extent that they emphasize the use of concepts such as 'space', 'place', 'distance', 'region', 'territory', 'landscape', and 'environment' as entry points into the investigation of the social world. To study industrialization from a geographical point of view amounts to embracing a style of thinking that is biased toward the aforementioned spatial categories. Is this bias justified? Instead of arguing that the geographical point of view is indispensable to the study of industrialization, we could make the more modest claim that geography provides a conceptual toolbox for qualifying the sometimes crude accounts of this process. Industrialization unfolds in space and produces space, and so do the related phenomena of deindustrialization, modernization, globalization, dependency, and pollution. The recognition of the spatial dimension of industrialization becomes significant only to the extent that geographers can extract the actual regularities, patterns, or 'laws' of the operation of this process. It is at this level of analysis that old and new challenges keep the geographical conversation open. One of the 'old challenges' comes from the fact that different political economic worldviews force divergent interpretations of the same economic processes. Thus, a geographical perspective indebted to neoclassical economics (e.g., older style industrial location analysis) brings with it a more or less tacit endorsement of the beneficial effects of industrialization, whereas a Marxist perspective carries with it a strong normative baggage that urges sensitivity to issues of social justice. Rather than trying to endorse one view and to discard the remaining views, it might be preferable to think of these different approaches as theoretical resources with complementary roles to play. Each school of thought is a constellation of gains and losses: each take on the spatial logic of industrialization may be particularly insightful in one respect, and appallingly silent in other respects.

Once with the turn to culture in the Anglo American human geography of the 1990s, the geography of industrialization has witnessed, among other revampings, an orientation toward institutional and evolutionary approaches to the analysis of the spatial dynamics of the industrial sector. These new directions have enriched the explanatory power of economic geographies, by showing how path dependency, institutional cultures, and geographical relations complicate the fabric of pure economic logic. Nevertheless, more quantitatively minded geographers lament the lack of clarity, rigor, and empirical support that the new vocabularies of these recent schools brought about.

One of the 'new challenges' that confronts the geography of industrialization comes from post structuralism, feminism, and nonrepresentational theory. These approaches share a reluctance toward grand theories and criticize both neoclassical and Marxist perspectives on industrialization for their illusionary beliefs about an objective economic reality governed by laws about to be 'uncovered'.

This line of criticism alerts us to the limitations of a nomothetic study of industrialization within a presumed global space economy and brings attention, at least indirectly, to the possibility of using other entry points to industrialization research. To illustrate this latter point, Simandan's recent attempt to integrate the field with the help of a master metaphor called 'recursive cartographies' starts with the simple but powerful idea that the world is the result of the interplay and mutual metamorphosis of 'three' elements: rhythms, events, and legacies. A short quote from the industrial geography literature is very helpful for understanding this model:

> …the historical process of industrialisation in North America and Europe is marked by stories of *small accidents* leading to the establishments of one or two *persistent centres of production.* Thereafter *cumulative processes* can generate a geographical structure of production which may be stable for long periods of time (italics added). (Hassink and Dong Ho, 2005: 572)

Three words have been emphasized in the text. The first is 'small accidents' and in recursive cartographies this would be translated as 'events'. Anything that abruptly disrupts the preexisting order of things qualifies as event. The outstanding features of events are that they bring genuine novelty and they perturb the state of affairs, though in largely different extents. The discovery of electricity, political elections, a bankruptcy, a strike, a merger, a foreign investment, etc., are all events.

The second word emphasized is 'persistent centers of production', which in recursive cartographies would be labeled as 'rhythms'. Anything that regulates a place, bringing constancy, predictability, and structural identity to it, constitutes a rhythm. Five decades of communist rule in Cuba, the production of chocolate in Birmingham, the four seasons of the temperate climate, the urban timetabling of the transportation network, and the legal system of a country, are all rhythms. They weave the fabric of a place, while being from time to time wounded by events that challenge their hegemony in processes of place formation. The advent of industrial activity to a rural area is an event; in its wake, that industrial activity becomes a rhythm that begins to reconstitute that area as 'industrial' (workers going to work at 8 am and returning at 5 pm, the schools adapting their curricula to prepare the new workforce, the streams of income in that local community becoming dependent on those industrial plants, etc.).

The third word emphasized is 'cumulative processes', which in recursive cartographies would be classified under the heading 'legacies'. Anything left in the world that is not either 'event' or 'rhythm' qualifies as 'legacy'. Put simply, the legacy of a place is the coagulation of its past events and past rhythms, with the critical observation that this does not mean that legacy is 'dead', lacking agency. Quite on the contrary, it contributes to place formation extensively: our actions are often guided by lessons learned from past actions, the stereotypes that produce the 'image' of a place come from past knowledge (and that image significantly influences present decisions – e.g., to invest or not to invest in a given regional economy), and a rhythm (e.g., the petrochemical industry) might rest on regional legacies (petroleum).

The many types of cartographies of industrialization to be found in contemporary geographical scholarship make research in this area reach a level of vibrancy and sophistication rarely found outside geography and promise to contribute to broader questions and critiques of our present condition.

See also: Capitalism; Development I; Diffusion; Industrial City; Industrial Districts; Industrial Location; Industrial Organization; Industrial Parks; Industrial Restructuring; Innovation; Modernity.

Further Reading

Allen, R. (2007). Engel's Pause: A Pessimist's Guide to the British Industrial Revolution, Economics Series Working Papers 315, University of Oxford, Department of Economics, revised.

Amin, A. (1999). An institutional perspective on regional economic development. *International Journal of Urban and Regional Research* 23, 365 378.

Anderson, B. (1991). *Imagined Communities* (2nd edn.). London: Verso.

Ayres, R. (2007). On the practical limits to substitution. *Ecological Economics* 22, 115 128.

Barnes, T. (2001). Retheorizing economic geography: From the quantitative revolution to the 'Cultural Turn'. *Annals of the Association of American Geographers* 91, 546 565.

Barnes, T. (2003). The place of locational analysis: A selective and interpretive history. *Progress in Human Geography* 27, 69 95.

Bickerstaff, K. and Walker, G. (2003). The place(s) of matter: Matter out of place public understandings of air pollution. *Progress in Human Geography* 27(1), 45 67.

Boschma, R. and Frenken, K. (2006). Why is economic geography not an evolutionary science? Towards an evolutionary economic geography. *Journal of Economic Geography* 6, 273 302.

Bottazzi, G., Dosi, G., Fagiolo, G. and Secchi, A. (2007). Modeling industrial evolution in geographical space. *Journal of Economic Geography* 7, 651 672.

Bunge, M. (2006). *Chasing Reality: Strife over Realism*. Toronto: University of Toronto Press.

Burt, T. (2005). General/particular (key debates in geography). In Castree, N., Rogers, A. & Sherman, D. (eds.) *Questioning Geography: Fundamental Debates*, pp 117 130. Oxford: Blackwell.

Carr, D. (2008). Narrative explanation and its malcontents. *History and Theory* 47(1), 19 30.

Castree, N. (2007). Capitalism, the Left and the new world (dis)order. *Progress in Human Geography* 31(4), 563 570.

Clark, G. (2007). *A Farewell to Alms: A Brief Economic History of the World*. Princeton, NJ: Princeton University Press.

Davies, R. W. (1980). *The Industrialisation of Soviet Russia Vol. 1. The Socialist Offensive*. London: Macmillan.

Gibson Graham, J. K. (ed.) (1996). *The End of Capitalism (As We Knew It): A Feminist Critique of Political Economy*. Cambridge, MA: Blackwell.

Goffman, E. (1990). *Stigma. Notes on the Management of Spoiled Identity*. Harmondsworth: Penguin.

Grabher, G. (1993). The weakness of strong ties; the lock in of regional development in the Ruhr area. In Grabher, G. (ed.) *The Embedded Firm; On the Socioeconomics of Industrial Networks*, pp 255 277. London: Routledge.

Grabher, G. and Stark, D. (1998). Organising diversity: Evolutionary theory, network analysis and post socialism. In Pickles, J. & Smith, A. (eds.) *Theorising Transition. The Political Economy of Post Communist Transformations*, pp 54 75. London: Routledge.

Gregory, D. (1998). *Explorations in critical human geography*. Hettner Lecture 1997 Department of Geography, University of Heidelberg.

Hagerstrand, T. (1952). *The Propagation of Innovation Waves*. Lund: Royal University of Lund.

Harvey, D. (1999). *Limits to Capital* (2nd edn.). London: Verso.

Harvey, D. (2003). *The New Imperialism*. Oxford: Oxford University Press.

Hassink, R. and Dong Ho, S. (eds.) (2005). The restructuring of old industrial areas in Europe and Asia: Editorial. *Environment and Planning A* 37, 571 580.

Horney, K. (1950). *Neurosis and Human Growth. The Struggle for Self Realization*. London: W. W. Norton & Company.

Hudson, R. (2005). Rethinking change in old industrial regions: Reflecting on the experiences of North East England. *Environment and Planning A* 37, 581 596.

Kemp, T. (1978). *Historical Patterns of Industrialisation*. London: Longman.

Kemp, T. (1989). *Industrialisation in the Non Western World* (2nd edn.). London: Longman.

Licht, W. (1995). *Industrializing America: The Nineteenth Century*. Baltimore, MD: Johns Hopkins University Press.

Martin, R. (2001). Geography and public policy: The case of the missing agenda. *Progress in Human Geography* 25, 189 210.

Massey, D. (1995). *Spatial Divisions of Labor: Social Structures and the Geography of Production* (2nd edn.). Basingstoke: Macmillan.

Massey, D. (2005). *For Space*. London: Sage.

Minca, C. (2007). Humboldt's compromise, or the forgotten geographies of landscape. *Progress in Human Geography* 31, 179 193.

Murdoch, J. (2006). *Post Structuralist Geography: A Guide to Relational Space*. London: Sage.

O'Rourke, K. H. and Williamson, J. G. (2005). From Malthus to Ohlin: Trade, industrialisation and distribution since 1500. *Journal of Economic Growth* 10, 5 34.
Painter, J. (2000). State and governance. In Sheppard, E. & Barnes, T. (eds.) *A Companion to Economic Geography*, pp 359 376. Oxford: Blackwell.
Peck, J. (2008). Remaking laissez faire. *Progress in Human Geography* 32(1), 3 43.
Peck, J. and Theodore, N. (2007). Variegated capitalism. *Progress in Human Geography* 31, 731 772.
Popper, K. (1935). *The Logic of Scientific Discovery* (1959 edn). London: Hutchinson.
Plummer, P., Sheppard, E. and Haining, R. (1998). Modeling spatial competition: Marxian versus neoclassical approaches. *Annals of the Association of American Geographers* 84, 575 594.
Rindermann, H. (2008). Relevance of education and intelligence at the national level for the economic welfare of people. *Intelligence* 36(2), 127 142.
Rodríguez Pose, A. and Storper, M. (2006). Better rules or stronger communities? On the social foundations of institutional change and its economic effects. *Economic Geography* 82(1), 1 26.
Rutherford, T. (2004). Convergence, the institutional turn and workplace regimes: The case of lean production. *Progress in Human Geography* 28, 425 446.
Sadler, D. and Thompson, R. (2001). In search of regional industrial culture: The role of labour organisations in old industrial regions. *Antipode* 660 686.
Scott, A. (2000). Economic geography: The great half century. *Cambridge Journal of Economics* 24, 483 504.
Scott, D. (2005). *Conscripts of Modernity: The Tragedy of Colonial Enlightenment*. Durham, NC: Duke University Press.
Simandan, D. (2004). Poverty, resource management, and the cultural politics of industrialization in Central Europe: The case of the Romanian Carpathians. *Romanian Review of Political Sciences and International Relations* I (2), 45 60.
Simandan, D. (2005). *Pragmatic Scepticism and the Possibilities of Knowledge*. Timisoara: West University Press.
Simandan, D. (2006). *Marginally Modern*. Arad: 'Vasile Goldis' University Press.
Sowell, T. (1998). *Conquests and Cultures: An International History*. New York: Basic Books.
St. Andors (1977). *China's Industrial Revolution*. London: Martin Robertson.
Storper, M. and Salais, R. (1997). *Worlds of Production: The Action Frameworks of the Economy*. Cambridge, MA: Harvard University Press.
Swyngedouw, E. (2000). The Marxian alternative historical geographical materialism and the political economy of capitalism. In Barnes, T. & Sheppard, E. (eds.) *A Companion to Economic Geography*, pp 41 59. Oxford: Blackwell.
Thrift, N. (2000). Pandora's box? Cultural geographies of economies. In Clarck, G. L., Feldman, M. P. & Gertler, M. S. (eds.) *The Oxford Handbook of Economic Geography*, pp 689 704. Oxford: Oxford University Press.
Thrift, N. (2007). *Non Representational Theory: Space, Politics, Affect*. London: Routledge.
Toffler, A. (1980). *The Third Wave*. New York: Knopf.
Weeks, J. (2005). *Population: An Introduction to Concepts and Issues* (9th edn.). Belmont, CA: Wadsworth.
Weiss, J. (2002). *Industrialisation and Globalisation: Theory and Evidence from Developing Countries*. London: Routledge.
Yapa, L. (1996). What causes poverty? A postmodern view. *Annals of the Association of American Geographers* 86, 707 728.
Yeung, H. (2005). Rethinking relational economic geography. *Transactions of the Institute of British Geographers* 30(1), 37 51.

Relevant Websites

www.fordham.edu
Industrial Revolution and the Standard of Living Essay.
www.lakemac.infohunt.nsw.au
Industrialization website a good brief overview of history and characteristics of The Victorian Web Site at Brown University. This site is useful for nutshell summaries of the Poor Laws, the Corn Laws, etc., lectures on The Industrial Revolution in England by Arnold Toynbee lectures on The Industrial Revolution.

Industry, Historical Geographies of

R. Lewis, University of Toronto, Toronto, ON, Canada

Glossary

Luddites Early-nineteenth-century British manufacturing workers who resisted the intrusion of new technologies and work practices by the smashing of machines. Short-lived and unsuccessful, they reflect the disruptive power of industrial capitalism on earlier industrial systems.

Underdevelopment The process by which industrialized countries actively blocked the economic development of nonindustrialized countries in most parts of Africa, Asia, and Latin America.

Zaibatsu Japanese commercial groups that became large diversified industrial corporations with interests in many parts of the national economy. Part of the state-led industrialization that occurred under the Meiji period from the late nineteenth century, they continue to dominate Japanese economic and political life.

Uneven Development and the Historical Geographies of Industry

Contemporary geographies of industry look quite different from what they did two or three hundred years earlier. Today, in the first decade of the twenty first century, people are both hypnotized and horrified by China's astonishing industrial development. One hundred years ago, observers were equally alarmed and exhilarated by the rapidly expanding industrial powers of the United States, while 200 years ago many were deeply shocked and excited by the rise of industrial Britain. These stories, stories of the rise and fall of industrial might in the modern period, reflect capitalist uneven development. The systematic process of economic growth and decline over time and space ensures that there are marked differences in the character and effects of industrialization in eighteenth century Manchester, nineteenth century Chicago, and contemporary Shanghai.

The uneven development of industry over time has taken place at different scales, from the global to the local. In all cases, and regardless of time and place, a singular set of processes have, in different combinations and in different ways, shaped the world's industrial history. As a coherent socioeconomic system, capitalism depends on capital accumulation and investment, the subordination of labor, the development of new production, transportation and communication technologies, the expansion of markets, continued access to raw materials, and state intervention and regulation. The changing contours of capitalist industrial growth were associated with major reshuffling of the leading positions within the industrial world order. The European imperial order was refashioned in the late nineteenth century as Germany and the United States took over world leadership from Britain. More recently, the preeminent position of the United States and the major European industrial countries has been challenged by Japan. In the first years of the twenty first century the dramatic expansion of China's state led capitalism heralds a new economic superpower. The shift of industrial capital from place to place over time – the historical geographies of industry – occurred within four overlapping phases.

The first phase dates from 1650 when proto industrialization became commonplace throughout many European rural areas. The development of a large number of small producing units initiated extensive manufacturing throughout France, Germany, Switzerland, Spain, Great Britain, as well as parts of Eastern Europe. Proto industrialization had a dual effect. First, it broke down the local character of trade by incorporating proto industrial areas into large merchant controlled trading regions. Port cities, from Liverpool to St. Malo and Hamburg, drew upon the products of the manufacturing workshops and distributed them throughout the world, and in the process redrew the parameters of the global economy. Second, it transformed everyday life. The search for cheap labor and profits reordered the relations of rural production. While the appearance of large manufacturing units employing machinery and inanimate power would have to wait until the nineteenth century, the social relations of proto industrialization created a proletariat that was dependent on the sale of labor power for its survival.

The second phase involved the rise of industrial capitalism, the de industrialization of older industrial districts, and the ascendancy of Great Britain as the world's preeminent power between 1740 and 1815. Even though proto industrialization was common throughout many parts of Europe, Great Britain was the first country to experience modern industrialization. A unique set of conditions transformed a small industrial producer, creating the world's most powerful industrial state. Increasing incomes and wealth freed up investment capital and generated market demand. A large urban based proletariat supplied cheap and disposable laborers for the mills and factories. The development of scientific knowledge and labor skills promoted incremental technological innovation that was funneled into existing (textile) and new (machine tools) sectors. State support for the creation of modern legal, property, monetary, and tax systems laid a stable base for economic growth. Strong trading links in the countryside, Europe, and the colonies provided resources, markets, and business networks that sustained continent spanning com modity chains. New transportation facilities such as harbors, roads, and canals meshed together the different components of the developing regional industrial division of labor and, in the process, reduced transport costs and delivery times, and facilitated the exploitation of raw materials, especially iron and coal. Together, these conditions fueled rapid in dustrial growth, resulting in the development of Great Britain as the first modern industrial state by the end of the eighteenth century.

The third phase, running from the late eighteenth century to the middle of the twentieth century, involved the slow, nonlinear, and uneven, but nevertheless inexor able emergence of industrial manufacturing in most parts of Europe, as well as the United States, Canada, and Japan. As these regions' industrial base developed, they chal lenged Britain's industrial preeminence. The rise of an Atlantic based industrial complex from the 1870s differed in important ways from the previous round of industrial change (**Figure 1**). Growth was driven by new industries, most notably steel, chemicals, electrical appliances, autos, and new technologies, such as the combustion engine and electrometallurgy. Together, these freed manufacturers from the locational constraints of coal and steam. The large vertical integrated, multiunit corporation replaced the family firm as the central organizational form for capital accumulation and labor control. The banking sys tem became increasingly critical as it permitted indus trialists to assemble capital in a variety of new (and secure) organizational forms. This is not to say that old production methods and organizational forms were redundant; smal ler scale custom, batch, and high volume producers con tinued to coexist with corporations deploying new production methods and organizational forms. Neverthe less, large scale industrial organizations became dominant and reshaped the global economic landscape.

A different process of industrialization took place in Japan. Building on the fledgling manufacturing economy constructed under the Tokugawa regime (1600–1868), Japan experienced extensive industrial growth during the Meiji period (1868–1945). Coordinated by an activist state that invested in the necessary infrastructures (rail roads and energy), Japan's industrial base was built around a tripartite structure by the early twentieth century. The realization of immense economies of scale reduced costs and gave access to international markets. The establishment of zaibatsu allowed a new industrial bourgeoisie to oversee the promotion of new industries, market expansion, and technological change. Large metropolitan complexes, most notably Osaka–Kobe and Tokyo–Yokohama, developed and became the home of the modern industrial state. Economic and political restructuring after World War II refashioned property relations, educational systems, corporations, and capital–labor relations. Cocooned within the sphere of American industrial and political interests, post war Japan experi enced dramatic growth and the rise to international prominence.

Finally, since 1945, industrialization has penetrated most regions of the world, most notably the newly in dustrializing countries of Asia and Latin America, pro ducing a new industrial division of labor and an interlinked and integrated industrial society on a scale never experienced before. Industrial change in these countries has taken a different pathway than the earlier ones. Malaysia, for example, before the 1970s was a typical staples producer, supplying tin, rubber, and vegetable oil to the developed world. Shackled to pri mary commodity production since the nineteenth cen tury, the country was embedded in the European imperial order. In the past 35 years, however, it has re duced this export commodity function and has resituated itself within the new international division of labor. Its industrial base now rests on export oriented manu facturing industries. As in Japan, a centralized state in stituted an industrial policy favorable to foreign direct investment and encouraged fixed capital and social cap ital investments. What is different is that manufacturing expansion is mostly driven by foreign capital, is export led, and heavily relies on low wage industries (textiles, electronics, rubber) that have little ability to generate economic interdependency within Malaysia.

The Spatial Division of Labor and the Historical Geographies of Industry

One of the most dramatic effects of the long term processes of uneven development and capital accumulation has been to divide the world into quite distinct parts, to forge a dis tinctive yet shifting spatial division of labor. Regardless of the labels given to this division, core–periphery, developed–underdeveloped, or First Third–Third World, the primary effect of capitalist industrialization has been to geographic ally divide the global economy by power and profits. This has been a long time in the making. The manufacturing dominance of non European countries more than 250 years ago has been undermined. In the eighteenth century, India and China accounted for 57% of the world's manufacturing output (**Table 1**). The dual processes of modern industrial manufacturing and European empire building, however, destroyed the handicraft character of the early leaders and led to the supremacy of a handful of industrial capitalist countries in Europe and North America. By 1913 more than 60% of all manufacturing output was concentrated in three

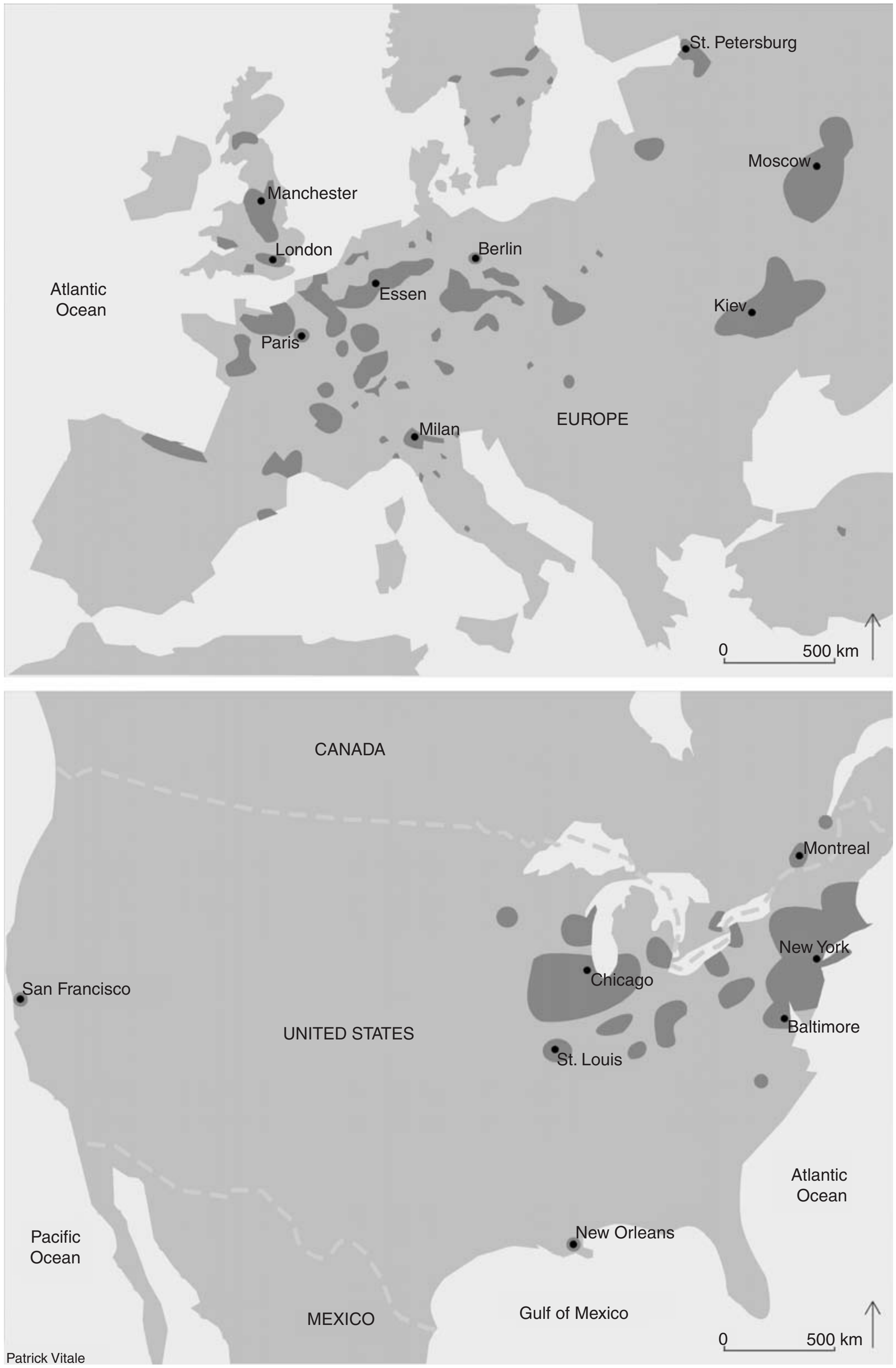

Figure 1 Industrial districts of the US, Canada, and Europe, *c.* 1880. Sources: Based in part on Pollard, S. (1986). *Peaceful conquest: The industrialization of Europe, 1760 1970*. Oxford: Oxford University Press. Meyer, D. (2001). The national integration of regional economies, 1860 1920. In McIlwraith, R. & Muller, E. (eds.) *North America: The historical geography of changing continent* (2nd edn.). Totowa, NJ: Rowman and Littlefield.

Table 1 Manufacturing output of selected countries, 1750 1913

Country	%			
	1750	*1830*	*1880*	*1913*
China	32.8	29.8	12.5	3.6
India	24.5	17.6	2.8	1.4
France	4.0	5.2	7.8	6.1
Japan	3.8	2.8	2.4	2.7
Germany	2.9	3.5	8.5	14.8
United Kingdom	1.9	9.5	22.9	13.6
United States	0.1	2.4	14.7	32.0

Note: Output consists of both handicraft and industrial manufacturing. Source: Tomlinson, B. (1999). Economics: The periphery. In Porter A. (ed.) *The Oxford history of the British Empire, volume III: The nineteenth century*, p 69. Oxford: Oxford University Press.

countries: the United States, Germany, and the United Kingdom.

At the same time, the underdeveloped world has been turned into a huge reservoir of raw materials for the developed world's factories. While rubber from Malaysia, cotton from India, iron ore from Brazil, and coffee from Kenya have fed the industrial machine of the developed countries, little processing has been done in the under developed world. This has several serious consequences. First, most of the value added component of the final product has gone to North American and European manufacturers. Second, the economies of the under developed world are highly unstable because of their reliance on primary products and declining terms of trade which favor the First World. Third, the fluctuating needs of developed countries and the ability of firms to shift capital from place to place in search of ever more better quality and cheaper supplies of raw materials have trapped periphery countries in a vicious cycle of underdevelopment.

Critical to the shifting international division of labor, both in the past and in the present, have been interfirm and intrafirm backward and forward linkages operating across national boundaries. From the earliest days, firms, by deploying their own agents or long distance mer chants, have traded manufacturing inputs. Indeed, to a large extent, Western dominance of global industry over the past 300 years has relied upon the ability to control the commodity chain, from access to raw materials to the distribution and marketing of finished products. Before the end of the nineteenth century, relatively unintegrated industrial firms relied on a wide set of intermediaries to coordinate the flow of manufacturing inputs and outputs. With the merger movements starting at the end of the nineteenth century and the rise of the large vertically integrated corporation an increasing share of trade be tween firms became internal to the firm. A complex system of ownership and intracorporate interaction de veloped. In the early twentieth century, for example, large corporations, such as US Steel controlled raw ma terial sites (iron ore mines), large integrated steel mills, finishing steel works, transportation (railroads and ship ping firms), and wholesalers, were integrated through board membership with financial houses and banks.

Large transnational corporations have the ability to drastically reshape national, regional, and local industrial landscapes. The burgeoning economic power of corpor ations in eighteenth century Britain, nineteenth century United States, and twentieth century Japan has been played out in many other countries. American large scale, direct investment since the 1880s, for example, has led to the development of a branch plant economy in Canada. Seeking the markets of Canada and the British Empire (through Canada's preferential trade agreements), plentiful raw material supplies, and greater profits, American cap italists began venturing into the manufacturing sector of Central Canada and the timber and mineral resources of the hinterland in the last decades of the nineteenth century. This has continued ever since, resulting in limited indus trial development, the flow of profits out of Canada, and massive US business and political influence in Canadian life.

The state has been critical to the formation and un folding of uneven development and the international division of labor. Before World War I, when private entrepreneurs operating through the market were the dominant agents of industrial change, the state's direct role in industrial development was minimal. In Great Britain and many other European countries, however, the developing industrial state did promote and orchestrate economic activities. Various levels of government (from the national to the local) intervened in a variety of ways. Although there were few explicit attempts to build a state led industrial policy before World War I, govern ments, through the redistribution of income, the devel opment on nonindustrial sectors, and support for economic and social institutions, underpinned the for mation of a modern industrial society. In the last third of the nineteenth century in emerging industrial nations, such as Japan, Canada, the United States, and Russia, the state was a more important agent of change and, under the guise of national development, a major vehicle for capital accumulation.

The state has become more involved in supporting, coordinating, and defending industrialization over the past 100 years. It has been involved in drawing up in dustrial policy by, among other things, facilitating in dustrial access to credit, establishing intranational and international trading pacts, and stoking consumer mar kets. The state has become more heavily involved in fi nancing and building necessary infrastructures, from nineteenth century Canadian railroads to twentieth century Chinese ports. The contemporary state has also been extremely active in establishing the ideological, legal, and technological framework of the capitalist and state socialist systems. Finally, the state has formulated and implemented militaristic and territorial expansion in order to drive national economic expansionism. From the British intrusions into its colonies before 1800 to economic colonization by the United States and Japan's territorial ambitions in Asia in more recent times, the state has shaped international, regional, and local divisions of labor.

The recent growth of newly industrializing countries has had little impact on the international division of labor: profits continue to be siphoned off by head offices in Los Angeles, Frankfurt, and Tokyo; technological innovation remains in the older developed world; capital shifts from place to place seeking the best returns; and state policies reinforce existing practices of capital accumulation that favor the developed world. This is evident in the case of the maquiladoras hugging the southern side of the US–Mexican border. Brought in after the closing of the Bracero (guest worker) program in 1965, the purpose of the maquiladora program was to attract foreign assembly plants to Mexico. Starting off quite slowly, the pace of growth quickened after 1980, and by 2000 more than 3500 firms employed 1.3 million workers. Most plants are owned by developed world firms (such as Volkswagen at Puebla and Toyota at Tijuana) which have come seeking low wages, poor environmental regulations, high profits, and a pliable national state. The result is the maintenance of the developed–underdeveloped divide.

Place and the Historical Geographies of Industry

As the Mexican maquiladoras suggest, industrialization has created a complex set of specialized regional and urban conurbations. Regions and urban areas are pools of economic, political, and social resources and offer industrialists accessible and place specific sets of markets, labor, technologies, information, and commodities. From the very beginning, regional agglomerations of particular industries have appeared. In the eighteenth century, production of textiles and other manufactures was concentrated in various regions of India and China. More recently, industry has clustered in Europe and North America: silk in Lyonnais and New Jersey; cotton in Alsace, Lancashire, Belgium, New England, and Saxony; and steel in the West Midlands, the Ruhr, and the American Midwest. At the same time, regions were underdeveloped. The classic case is the American South, which missed out on the first burst of industrialization in the early nineteenth century. As well as supplying raw materials (cotton, lumber, and tobacco) to the US North and Europe, it also had a low wage, low skill, labor intensive industrial economy that remained in thrall to industrialists and financiers in New York, Philadelphia, and London until well after World War II. The capitalist industrial division of labor consists of a regional framework divided by industrial production, raw material supply, and markets.

Urban agglomerations of different sizes and specializations developed as integral parts of the industrialization process. The large metropolis is the home of a diversity of industries. Berlin, London, Paris, New York, Montreal, Tokyo, and Shanghai manufacture a dazzling array of producers' goods (chemicals, machinery, foundry products) and consumer goods (clothing, processed food, books). As major industrial, financial, and political centers, these large metropolitan areas are the primary nodes of their national manufacturing economy and, by extension, the global economy. Further down the urban system, a much larger number of specialized cities turn out particular products. From the earliest days of proto industrialization to the development of new industrial cities in the growth areas of Asia and Latin America, small specialized cities have been essential to the national and international industrial division of labor. From watchmaking in Geneva and Jura to cutlery in Sheffield and rubber in Akron, specialized cities have been part of both corporate and industrial commodity chains that stretch across regions, nations, and the globe.

Industrial change initiated new forms of social differentiation. New class relations developed as proletarianization and the rise of new classes accompanied industrialization. The preindustrial domestic system was replaced by capitalist investment, increased workplace scale, growing mechanization, and elaborate divisions of labor. Blue collar and clerical working classes labored in the factories, offices, docks, building lots, and streets of the capitalist city. In rural areas, new forms of social stratification appeared as industrial farming replaced older types of food production. A well educated, well paid professional class consisting of lawyers, academics, bureaucrats, politicians, and factory managers controlled, managed, and coordinated the functioning of the work and nonwork worlds. The merchant and the lord were replaced by the new overlords, the industrial bourgeoisie of manufacturer, corporate executive, and financier. Although taking different forms at different times in different places, these three classes (workers, professionals, and bourgeoisie) continue to be the basic class structure of a modern industrial society. This has been grafted onto existing forms of social stratification, such as the Indian caste system, and integrated within hierarchies centered on race, ethnicity, and gender.

These social distinctions have transformed the social geographies of the industrial metropolis. When these complicated social differences are combined with the dynamics of the capitalist housing and labor markets, a well defined and highly graded set of neighborhoods are constructed. The social worlds of metropolitan areas are fraught with class, racial, and ethnic based residential differentiation. As Frederick Engels pointed out, the residential geography of nineteenth century Manchester was class based. Similarly, early twentieth century Chi cago exhibited an array of social areas; the exclusive ghettos of the White native born professional and entre preneur were cheek by jowl with the vast working class district, which was differentiated by country of origin, race, and occupation. The physical distance separating the exclusive 'Gold Coast', middle class Hyde Park, immi grant Packingtown, and the African American district south of the Loop (Chicago's central business district) were small, yet the social distances were gigantic. Capit alist industrialization finely calibrates urban social space.

Modern industrialization has triggered large scale population migration. Two movements are particularly important: the shift of people from the countryside to the city, and mass migration from one country to another. Although the reasons and manner in which these oc curred varied across time and space, two basic processes have driven the reshifting of populations within and across national borders. The first has to do with political security and citizenship. From the dismantling of existing socioeconomic systems in the underdeveloped world to the pogroms in Eastern Europe, people have moved to escape political turbulence and oppression and to build a better life. The second has to do with work. Between the seventeenth century and World War II, for example, hundreds of millions of Europeans, uprooted by eco nomic and political changes have sought better working conditions and economic well being. As early as the seventeenth century, growing industrial production and increasing agricultural productivity in Europe freed rural populations from the everyday rhythms of the country side, forcing farmers, peasants, landless laborers, and displaced industrial workers, forcing them to move to the burgeoning industrial cities, such as Manchester, Paris, Berlin, and Prague. More recently, a similar process has taken place in other parts of the world. Today, in one of the largest movements of people ever known, more than 100 million rural migrants live in China's rapidly growing industrial cities.

Migrants have crossed national borders. In some cases, this is legal. Before the 1960s, drawn by industrial pro duction in urban factories and the opportunities of modern agriculture, the British colonies of the United States, Canada, and Australia were settled by large number of Europeans. In the underdeveloped world, other groups, including East Indian and Chinese, have been drawn to trading centers where they have acted as intermediaries in the flow of raw materials and manufactured goods between the developed and underdeveloped worlds. More recent shifts in the global industrial economy have stimulated population shifts on a scale never before experienced. People from Africa, Latin America, and Asia, forced by changing economic and political relations on the land and in the cities, have moved, sometimes as citizens, sometimes as guest workers, to the developed, Westernized countries. In other cases, immigration has been illegal. People from, among other places, Turkey, Mexico, China, and Nigeria, have slipped across the borders of the industrialized West looking for work. It is estimated that the United States may have up to 12 million illegal immigrants.

Modern industrialization has been a painful and violent process. It has dislocated custom, upset routines, and re oriented everyday life. The shift to mechanization, factory work, and intensive forms of social divisions has led to a range of responses: Luddites raged against the machine as a way to defend their future; industrial workers joined unions to protect their rights and improve working con ditions; and political elites and industrial bourgeoisie used the police and military to impose law and order on workers. Less dramatically, but perhaps more insidiously, industrialization has refashioned everyday life: village custom has been replaced by the strictures of urban fac tory time, while older social patterns and structures have been uprooted by modern forms of social stratification. Industrialization in all of its forms requires the sur veillance and rationalization of industrial spaces. From within the factory to the metropolitan landscape, from corporate driven projects to state sponsored continental markets and trade unions, changes to everyday life have been marked, painful, and decisive. Even though these changes have not been always obviously violent or rapid, they have, nevertheless, whittled away at older lifeways, imposed new ones, and reconfigured everyday life to fit the industrial capitalist regime.

A fundamental feature of capitalist uneven develop ment and the changing spatial division of labor is de industrialization. De industrialization has a long history: from the de population of cottage and domestic pro duction in Europe and the state sponsored destruction of colonial industries in India to the more recent dis investment and economic decline in the older industrial regions of Europe (the Ruhr, Northern England, and Scotland) and the United States (the Midwest and New England), the uneven development of capital and power leaves relic and redundant landscapes with their at tendant problems and despair. Two cases, the English Weald and the American Midwest, illustrate elements of the historical geography of de industrialization.

In the early eighteenth century the Weald was a successful proto industrial region built on the production of glass, cloth, and iron goods. By the end of the century, however, these sectors had declined in importance and the region was unable to make the transition to full scale modern industrialization. The reasons for this, although specific to the Weald, also point to the enduring com monalities of capitalist industrialization: dependence of its producers on outside capital and credit, the relegation of local interests to national and international ones, and the absence of interdependent diversity. As a result, the Weald could not compete with the new industrial economies emerging in the North and the Midlands.

The American Midwest was unable to make the transition from an automotive based industrial economy forged before World War I to the electronic one of the late twentieth century. The region's cities and towns, such as Detroit, Akron, Pittsburgh, and Chicago, went into a spiral of economic decline as key industries (automobiles, tires, steel, and electrical appliances) shut down factories and laid off hundreds of thousands of workers. The Midwest's successful production regime was undermined by a combination of declining prod uctivity, technological competition from Europe and Asia, the shifting of capital out of the region, and chan ging industrial policy. As part of a policy of disinvest ment, capital was moved out of the Midwest's industrial centers as manufacturers sought greenfield sites in the American South, Mexico, and Taiwan where they could find higher profits, escape a high wage and unionized workforce, and build new factories where new production processes and technologies could be installed.

In many ways, the fundamental dynamics of capitalist industrialization have not changed much over 400 years. The processes undermining the industrial economies of the English Weald and the American Midwest are the same. Capital continues to transform societies and land scapes. The result has been uneven development, whether in Britain and other parts of Europe during the first round of industrialization or the more recent changes occurring in the United States and Asia. From the small centers in seventeenth century Western Europe to the burgeoning industrial economies of East Asia, accumulation has favored different people and different places at different times. In the process, people have been turned into an industrial proletariat; new technologies have restructured the productive forces; the search for new and expanding markets has produced a cornucopia of commodities that sustain everyday life; and the state has regulated, co ordinated, and planned the industrial economy. Whether it is the French silk manufacturer operating out of eighteenth century Lyon or the automobile executive making decisions in Tokyo's downtown business district, the search for larger profits and markets has driven in dustrial capitalism and created complex spatial divisions of labor over 400 years.

Nevertheless, the differences across time and space are vast and imposing. The manufacturers, merchants, and financiers who gathered in early nineteenth century coffee houses in Manchester, Lyon, and Philadelphia functioned in a different world from the executives of the huge multinational, multiunit conglomerates that dom inate today's global industrial world. The day to day conditions that workers face differ dramatically. Histor ically, the working class elite of Western Europe and North America have reaped some of the benefits of an ever expanding global economy, while the peasants turned proletariat of the industrial South continue to work under the worst of conditions. At the same time, the development of the underclass in Detroit and Los Angeles is just the other side of the coin of the growing size of the wealthy middle class in Shanghai and Mumbai. Large international flows of labor have only complicated this differential pattern of work and everyday sustenance. The specific causes and the effects of modern de indus trialization differ from those that destroyed cottage production in the eighteenth century. Often violent, al ways transformative, mundane, insidious, and everyday, capitalist industrialization has remade the world in its own image time after time.

See also: Capitalism and Division of Labor; Finance, Historical Geographies of; Historical Geographies, Urban; Trade, Transport and Communications, Historical Geographies of.

Further Reading

Barrett, J. (1987). *Work and community in the jungle: Chicago's packinghouse workers, 1894 1922*. Urbana, IL: University of Chicago Press.

Cowie, J. and Heathcott, J. (eds.) (2003). *Beyond the ruins: The meanings of deindustrialization*. Ithaca: Cornel University Press.

Dennis, R. (1984). *English industrial cities of the nineteenth century: A social geography*. Cambridge: Cambridge University Press.

Drabble, J. (2004). *An economic history of Malaysia, c. 1800 1990: The transition to modern economic growth*. London: Macmillan.

Gregory, D. (1990). 'A new and differing face in many places': Three geographies of industrialization. In Dodgson, R. & Butlin, R. (eds.) *An historical geography of England and Wales* (2nd edn.), pp 351 399. London: Academic Press.

Hudson, P. (1992). *The industrial revolution*. London: Edward Arnold.

Kemp, T. (1978). *Historical patterns of industrialization*. London: Longman.

Meyer, D. (2001). The national integration of regional economies. In McIlwraith, R. & Muller, E. (eds.) *North America: The historical geography of a changing continent* (2nd edn.). Totowa, NJ: Rowman and Littlefield.

Meyer, D. (2003). *The roots of American industrialization*. Baltimore: John Hopkins Univerisity Press: John Hopkins.

Mosk, C. (2001). *Japanese industrial history: Technology, urbanization and economic growth*. Armonk, NY: M. E. Sharpe.

Pollard, S. (1986). *Peaceful conquest: The industrialization of Europe, 1760 1970*. Oxford: Oxford University Press.

Tilly, C. (1983). Flows of capital and forms of industry in Europe, 1500 1900. *Theory and Society* 12, 123 142.

Tomlinson, B. (1999). Economics: The periphery. In Porter, A. (ed.) *The Oxford history of the British Empire. vol. III: The nineteenth century*, p 69. Oxford: Oxford University Press.

Inequality

G. DeVerteuil, University of Southampton, Southampton, UK

Glossary

Economic Inequality The unequal distribution of income and wealth within and across societal groupings.
Gini Coefficient Commonly used measure of inequality, expressed in percentage terms between 0 (no inequality) and 100 (total inequality).
Inequality Involves arithmetically unequal distributions that also have a normative/moral dimension.
Inequity Distributions that are unjust (usually determined by need), but not always arithmetically unequal.
Polarization Special subset of inequality, concerned with the top and lower extremes gaining at the expense of the middle within a given distribution (usually income or wealth).
Spatial Inequality The spatial manifestations and implications of inequality.
Welfare Geography Branch of human geography popular in the 1970s that focused on 'who gets what, where, and how'.

Definitions, Expressions, and Measurements

The world in the early twenty first century seems more socially and economically unequal than at any other time in living memory. While the super rich live in mind boggling luxury and privacy, the most destitute – in refugee camps in Sudan, the callejones of Lima, or on the streets of Skid Row, Los Angeles – live in medieval like conditions. Previously socialist nations with moderate levels of inequality, such as China and Russia, are currently undergoing significant growth in the gap between haves and have nots. Most American cities are losing their middle class neighborhoods, bifurcating between privatized enclaves of affluence and concentrated zones of poverty for those left out of the postindustrial, post Fordist urban economy. Los Angeles and New York City experience income gaps that are comparable to cities in the less developed world. In these very cities of the less developed world, almost all growth occurs within peripheral mega slums, disconnected from infrastructure and the victims of social apartheid. Increasing inequality seems to go hand in hand with globalization; while some nations are enjoying unprecedented prosperity, nearly half the world's population subsists on less than US$1000 a year, with some 1 billion citizens living in utter destitution. To make matters worse, many states seem more bent on magnifying and reproducing, rather than countering, these mounting disparities through privatization and drastically reduced ambitions at income redistribution.

Even this cursory examination suggests that inequality must be treated as among the most pressing social and geographical issues of the twenty first century. Witness the recent high profile campaigns to Make Poverty History and reduce debt loads within the less developed world (e.g., Live 8 concerts in 2005). Witness also, worries in the developed world that extreme poverty and inequality may be fertile ground for failed states and ultimately terrorism. Geography is deeply implicated in the concentration of wealth and poverty. As a relative latecomer to the field, human geographical interest in inequality has waxed and waned, but the resurgence of global inequality has once again placed the issue front and center within the discipline.

Organization of Entry

Given the sprawling and multifaceted nature of inequality, this entry is necessarily wide ranging, cursory, and eclectic. Let us begin by defining inequality and its associated terms, then proceed to the concept of spatial inequality and its origins in several key foundational subfields – Marxist urban political economy and welfare geography – that established and formalized geographical interest in issues of inequality in the 1970s. The waning geographical interest in inequality during the 1980s was re energized by the resurgence of global(ized) inequality in the 1990s and beyond. The attention paid to two key contemporary issues – income polarization and concentrated poverty in particular – is necessarily urban focused, because it is in cities, and especially large cities, where inequalities and resistance to it are most apparent, where the rich and poor concentrate and rub shoulders (but mostly ignore each other). The role of the (welfare) state in mediating inequality is also discussed, including the sense that the state is now magnifying, rather than countering or ignoring, inequality. Several examples from Los Angeles are offered. Similar issues are then highlighted in the less developed world, particularly with regard to the urbanization of poverty and the spatial separation of the classes. Here, the case of Sao Paulo is used to flesh out these concerns. Finally, a glimpse is offered into how geographers are currently approaching inequality.

Definitions

Arithmetic inequality refers to a situation in which people are receiving unequal quantities of some attribute within a given distribution. Beyond this relatively basic starting point, defining inequality from a social science perspective is a difficult and daunting task, given its incredibly complex and sprawling usage. Inequality can mean radically different things to different people, and as such, any definition will necessarily be limited. Part of the definitional challenge is also that inequality is related to a great many concepts, including inequity, disadvantage, disparity, polarization, and so forth. The aim here is to derive a parsimonious definition of inequality, while also clarifying how inequality overlaps with some of these associated terms.

David M. Smith is careful to define inequality as a special type of differentiation or variation of character istics among members of a designated group. Certain personal attributes (e.g., name, height, weight, hair color) are different among different populations – this variation is seen as normal and does not provoke any sense of moral indignation or inequality. However, if these superficial personal attributes were suddenly used to discriminate, to make some populations disproportionately better off than others in terms of income, wealth, health, or life chances, then the concept of inequality could be invoked because this is no morally relevant basis to so discriminate.

In other words, there is a moral and normative di mension to inequality, a sense of what ought to be fair or unfair. This moral dimension is captured in the over lapping term of inequity. Inequity refers to distributions that are unjust, but not always arithmetically unequal. Why? Because sometimes (and somewhere) certain groups have greater needs than others, and inequity arises by treating everyone the same – some groups re quire greater quantities in proportion to their greater needs (i.e., more than their arithmetic fair share), and so even inequality may be morally justifiable at times. More often than not, of course, inequity and inequality overlap, as many groups do not even receive their minimum fair share at all (an unjust and unequal outcome). For the purposes of this article, focus is laid on inequality as unjust and unequal distributions and outcomes, thereby capturing the spirit of both inequality and inequity.

The term inequality also brings up issues of social justice, a concept that will be discussed in the final section. Suffice it to say that social justice is concerned with how society's benefits and disbenefits are distributed and why, and ultimately how inequality can be reduced, if not eliminated. So, if 1% of the nation's population, using a particular attribute (such as higher education), was hoarding 99% of the wealth, surely this condition would be deemed one of inequality. By definition, inequality is a relative or comparative concept, spanning standards across the whole of society, regions, or neighborhoods.

Our conceptual understanding of inequality may be further strengthened by defining a few more associated terminologies. Polarization is regarded as a special subset of inequality, occurring when the highest and lowest ex tremes of any social characteristics (however defined –but usually income or earnings) are both growing at the ex pense of the middle, creating a bimodal hourglass distri bution, rather than an egg shaped one. Further, there are many terms associated with the bottom of the inequality spectrum, including disadvantage and deprivation. For instance, socioeconomic analysis in the UK regularly uses six dimensions of deprivation, all ranked by region in a relative fashion: income, employment, health, education/skills/training, housing, access to services. In the less developed world, measures of deprivation are relative and absolute, since there are segments of the population that do not even have access to these basic necessities of life.

Expressions of Economic and Social Inequality

To flesh out the previous conceptualizations, let us briefly illustrate several key expressions of inequality, em phasizing economic inequality – that is, the unequal and inequitable distribution of income and wealth within and across societal groupings. Economic inequality has been focused upon because most of the world's population lives in capitalist societies where access to and quality of various elements of social well being – including nutrition, shelter, health, education, employment opportunities, clean en vironments, leisure, security, social stability, and so forth – are increasingly determined by purchasing ability. Eco nomic inequality inevitably creates social inequality, as some groups are denied access to these basic elements of social well being. As a result, these groups suffer from social exclusion, denial of basic human rights, and a lower quality of life. Economic and social inequalities are tightly linked to class, but interact with a variety of other key social denominators that relate to power, opportunity, and life chances, such as (but not limited to) race, gender, sexual orientation, (dis)ability, religion, and age.

For example, people living long term in deprived areas in the developed world tend to have lower health status and suffer from more violence and self destructive be havior. This is because their relative material deprivation vis a vis the rest of the society is compounded by feelings of chronic stress (due to lack of control over their en vironments), as well as humiliation over their devalued social position, particularly in highly unequal societies such as the United States where race becomes a com pounding factor. In the US case, race and class reinforce each other in particular places – usually inner city ghettos – indicating a deeply stigmatized class position that goes beyond merely the associated material deprivation. An other example involves gender. Through relentless cuts to welfare, single female parents are increasingly consigned to poverty, losing out to dual income earners who are colonizing the most valuable and strategic urban spaces.

Finally, it is important to differentiate among de scribing, understanding/explaining, and addressing (i.e., reducing) inequality. Although these different approaches will crosscut the rest of the entry, suffice it to say here that understanding/explaining inequality goes beyond simple, neutral descriptions (i.e., inequality is or is not occurring) to examining numerous processes that create inequalities. Most agree that inequalities are expressions of the eco nomic and social systems in which they are embedded, and that addressing inequality is important because of the societal breakdowns and costs it engenders. An example of social breakdown is the increased crime and fear that ac company gross inequality, as those left out feel little ob ligation or attachment to a society that excludes them. It should come as no surprise that nations with high levels of inequality also usually have high crime rates. Finally, ad dressing/reducing inequalities involves moving beyond academic research to a more direct engagement with the (unequal) material realities of the world.

Measuring Inequality

There is no single metric (or scale) that can capture all the intricacies of economic inequality. The most popular (and best understood) measure is the Gini index for in come or wealth. The Gini index is represented as a co efficient with a range from 0 to 100: scores near zero display almost no inequality, while those approaching 100 display extreme inequality. Visually, the Gini coefficient is represented as a curve, plotting the cumulative dis tribution of population against the cumulative distri bution of income. A diagonal straight line signifies absolute equality, in which each member of the popu lation has an equal share of income. In reality, it is more likely that distribution is unequal, in which case the line sags below the idealized diagonal. The more the line sags, the greater the inequality. Using the World Income In equality Database, it is possible to obtain Gini index scores for a variety of nations across the world. Nations with very high levels of inequality – such as Brazil – have Gini coefficients over 55, while nations with low levels – such as the Scandinavian countries – have Gini co efficients under 30.

Figure 1 illustrates the Gini index, with the line of equality (Gini = 0), as well as two hypothetical income distributions that depart from perfect equality – one moderately unequal, the other highly unequal.

In the moderately unequal distribution, the poorest half of the population earns 40% of the total income, while in the highly unequal distribution, the poorest half only earns 25% of the total income. Other approaches to measuring inequality include using the ratio between extreme values (usually only the top and bottom deciles), a technique primarily undertaken to measure income polarization. Further, regions can be ranked according to a variety of deprivation and well being measures.

All measures must be approached with some caution, however. The level of inequality measured is complicated by – if not a function of – scale, since different unit aggregations can produce different levels of inequality.

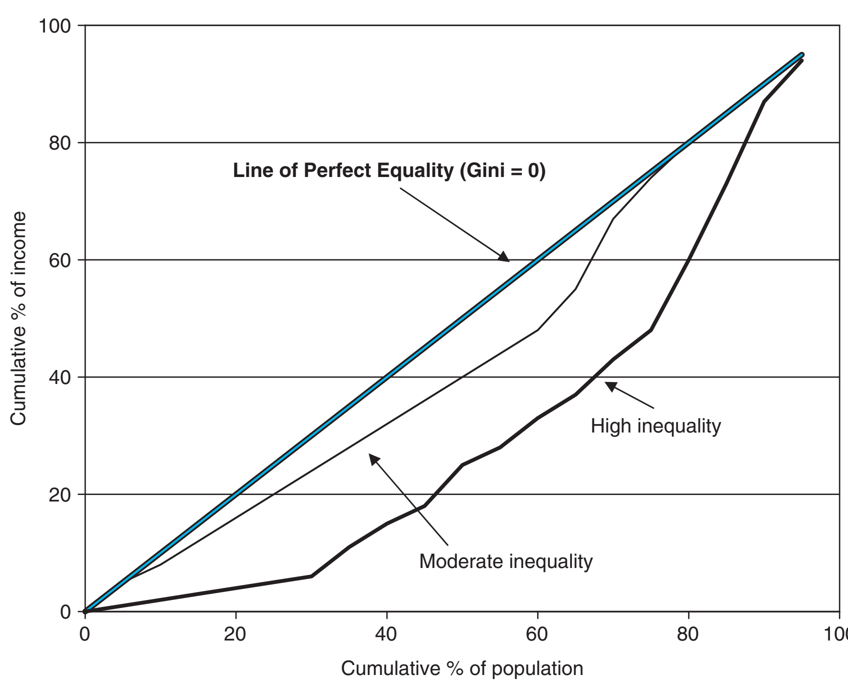

Figure 1 Gini coefficients.

Larger scales – such as the national – can obscure im portant internal variations, a problem of ecological fallacy. For instance, the relatively modest income inequality at the level of the Canadian nation obscures far greater in equalities at the urban level, particularly in the primary city of Toronto, where the poor earn about the same as the poor everywhere else in urban Canada but that the rich are far richer.

The Concept of Spatial Inequality

Up to this point, the discussion of inequality has been largely aspatial, perhaps because the study of inequality has been very much the domain of just two disciplines: economics and sociology. Between them, there is a staggering amount of theoretical and empirical work; economists focus on explaining and measuring income distribution, while sociologists emphasize the role of class stratification, social inequality, and social classes in their research. Perhaps because of the substantial amount of research, the field is highly contested and fragmented. Insights from other disciplines, including geography, have been increasingly contributing to the body of knowledge, particularly with regard to widening the purview of economic models limited to profit maximizing indi viduals, linear time, and aspatial processes.

Before tracing the origins of geography's relatively recent foray into inequality, let us briefly highlight some of the fundamental ways in which space, place, and scale are deeply implicated in the creation and sustaining of inequality, constituting a distinct concept and field of inquiry known as spatial inequality. At its simplest, eco nomic inequality necessarily has a spatial expression, as different places experience different degrees of wealth and poverty. In other words, some places are more im pacted than others, and this realization has a long history. The first systematic mappings of the areal differentiation of inequality began with the Chicago School of Sociology and its concentric zone model. Using a variety of social indicators and measures of social deprivation (disorgan ization), they found an unequal pattern, with most of the problem areas confined to the zone in transition around the central business district (CBD). It was in this zone in transition where a striking pooling of crime, vice, in sanity, or low income could be observed. Since the 1920s, patterns of inequality have been usefully mapped, in creasingly with the tool of geographic information sys tems (GISs) and using exploratory data analysis techniques. However, these mapping exercises tend to be weak in terms of explaining the deep seated causes of inequality.

A deeper approach is to go beyond the mapping of (uneven) patterns to see how place and inequality are mutually dependent and constitutive. As Janet Kodras contended, inequality is a place bound and geographic ally specific interplay between the market that distributes economic resources, and the state, whose role is sec ondary, serving to counter or reinforce disparities. Fur ther, this process is scale dependent, in that broader global and national market and state processes that create inequalities are filtered through specific local contexts, creating place specific variations in patterns of affluence and poverty. The causes and experience of inequality in Detroit are more a product of de industrialization and racism than in Eastern Kentucky, which has always suf fered from isolation and underdevelopment.

Using this more sophisticated approach, place is in extricably implicated in the production and sustaining of inequality, as well as potential solutions. Geographers point to how the place exclusion and separation of the have nots from the mainstream can reduce their social well being; disadvantages are passed on over the gener ations in place through spatial and social exclusion, thus promoting material deprivation and the inability to par ticipate fully in society. Conversely, advantage can be passed on simply by being born in a good place. Geog raphy has an important role in cumulating, reinforcing, and perpetuating inequalities over time. Reducing in equality through state intervention and resistance to op pressive structures is also a deeply geographical process, one which will be highlighted throughout the article.

Emergence of the Geographical Focus on Inequality

Spatial inequality as a legitimate human geography con cept emerged as a result of dissatisfaction with 1960s spatial science. This dissatisfaction stemmed from the latter's lack of policy relevancy, its largely descriptive na ture, and its disinterest in real world social conditions, ideologies, and solutions. By the 1970s geographers were subsequently focusing on questions of social problems – poverty, class, crime, racial discrimination, and inequality – and how public policy, supported by geographically sensitive research, could help alleviate the worst extremes. Nevertheless, certain holdovers from the spatial science era – including the application of the Gini index – have usefully persisted to the present day.

Marxist urban political economy

Marxism had always been attuned to the disparities caused by capitalism. The Marxist approach to spatial inequality was largely set out by David Harvey's *Social justice and the city*. Harvey's book had two parts: the first focused on liberal formulations of urbanism and geog raphy more generally, while the second part exposed some of their very limitations by using a radical, Marxist approach. In Harvey's view, the capitalist city by its very nature reproduced inequality through a regressive redistribution from the poor to the wealthy. This cumulative causation mechanism was related to unfairly structured "location of jobs and housing, the value of property rights, and the price of resources to the consumer." In order to reduce the inequalities inevitably produced by the goals of market efficiency, Harvey put forth the concept of "a just distribution justly arrived at," with the aim of achieving "territorial distributive justice." A "just distribution" must take into account need, contribution to the overall good, and merit. Ultimately, a sense of territorial social justice may occur if the following two sets of conditions are fulfilled:

1. The distribution of income should be such that (1) the needs of the population within each territory are met (needs), (2) resources are so allocated to maximize interterritorial multiplier effects (contribution), and (3) extra resources are allocated to help overcome special difficulties stemming from the physical and social environment (merit).
2. The mechanisms (institutional, organizational, political, and economic) should be such that the prospects of the least advantaged territory are as great as they possibly can be.

The "just distribution justly arrived at" concept would prove influential over the next decade, particularly in welfare geography. In the second part of his book, however, Harvey recast his critique of the capitalist city (and liberal theory) in terms of accumulation and class struggle, the state working to reproduce unequal social relations, exploitation and conflict, of systems not working in balance. Social justice now becomes more contingent on societal processes, rather than some unchanging set of norms. Distribution was more explicitly related to production, and there is more emphasis on solutions that are not based in the capitalist system, as one cannot truly alter the inherently unequal distribution of income without altering the very structure that produced it. Put simply, city structure thwarted all truly redistributive efforts.

Other Marxist geographers would pick up on these themes, particularly the sense that spatial inequality was inherent to capitalism and necessary to its perpetuation. The line of argument revolved around the following: capitalism leads to an unfair allocation of local opportunities and services, based initially on the structure of labor markets. Labor markets determine the level of dependency on the state and the ability to access finance and property markets. Land and housing markets reproduce labor market inequalities, and thus inequality is passed on by the inherited local environment. This endemic inequality at the local level is exacerbated by larger forces of centralization and decentralization that actively produce developed and underdeveloped regions and nations (e.g., uneven development).

Welfare geography

Like Marxist urban political economy, welfare geography emerged as a critique of spatial science. However, welfare geography adopted many of the tools of spatial science and continued the search for generalizable patterns, casting them within a normative and quantitative light. Borrowing liberally from welfare economics, the goal of welfare geography was to shed light on 'who gets what, where, and how'. The role of geographers was to expose spatial inequalities, and perhaps even contribute to better spatial arrangements within society. However, there was no sense of fundamentally changing existing structures; rather, as in the liberal tradition of social justice, the goal was to advantage the worst off and treat the extremes of inequality rather than removing them entirely.

The case of public facility location theory is instructive, as it epitomizes many of the key tendencies of welfare geography in the 1970s. These included the distribution of and access to services, as well as the important normative role of the state in rectifying the capitalist tendency toward unevenness and inequality by seeking the most egalitarian spatial arrangements for society. Public facility location theory aimed at finding the most efficient and equitable locations for public facilities (e.g., libraries, hospitals, incinerators, shelters, etc.). These locational decisions differed from private locational decisions, since the latter does not have to address the irksome concept of equity – what is a fair and adequate distribution of public facilities that maximizes social welfare? While measuring efficiency was straightforward (e.g., staying within a prescribed budget), measuring equity through prescriptive modeling was more difficult, particularly when certain facilities have widespread benefits but localized externalities and spillover costs (e.g., incinerators, airports). Various surrogates for equity were used, including minimizing distance to service facilities for the maximum amount of clients, as well as maximizing the intensity of services to the neediest clients. Through its largely quantitative and neoclassical assumptions about distance, pattern, accessibility, impacts, and externality effects, public facility location theory joined the normative and the spatial, although the latter was weakly developed.

By the 1980s, interest in neoclassical, quantitative modeling of welfare criterion waned, replaced by concerns for broader and more conflictual issues, including the unequal distribution of benefits and burdens across the city and the role of the state in sustaining these distributions. For example, the state sanctioned saturation of de institutionalized mentally ill patients in service rich yet marginalized inner city areas was criticized for being quite convenient for the city as a whole, but unequal for those individuals who had to live under such segregated conditions. As time went on, there was an even more pronounced shift away from the entire idea of modeling and mapping inequality (and issues of distribution, access, and the rectifying state), replaced by more identity based, post structuralist frameworks whereby race and gender (and other key social demarcators) were now central in understanding inequality, along with class. The explicit concept of the geography of inequality became rather muted, if not slowly subsumed into the larger realm of critical geography.

Restructuring and the Resurgence in Inequality

By the 1990s, geographers would once again directly engage issues of (economic) inequality, spurred on by dramatic real world shifts. After a brief post war interlude during which broad social mobility seemed possible, inequality forcefully returned to the scene. The surge in inequality was a result of fundamental economic, political, and social restructuring, including heightened globalized competition that produced a more flexible, lean, and downsized postindustrial production, bifurcation of labor markets and falling wages, welfare state contraction, and the collapse of socialist economies. Geographical space was deeply implicated in and impacted by all of these structural changes. Some examples included: (1) shifts in the place hierarchy, as some places, regions, or neighborhoods lost out and became disadvantaged, thereby producing a more uneven and fragmented world; (2) increasing urbanization of the very poor, along with the pooling of the very rich in urban space; and (3) aggravation of previous inequalities, and the emergence of new ones (skills, immigrants, etc.) in a post Fordist, postindustrial city.

Sociologists were also sensitive to these geographical effects, particular at the urban scale in the developed world. Globalization was seemingly creating greater income polarization in large cities, while simultaneously disadvantage – itself a product of global and local forces (de industrialization, spatial mismatch, racism, etc.) – was becoming more spatially concentrated, with predictably negative effects. These insights on urban inequality were most prominently presented by Saskia Sassen (global city and polarization) and Julius Wilson (underclass and spatial mismatch), respectively. Both focused on the (negative) outcomes of massive restructuring in the developed world after the 1970s, drawing inspiration from the rapidly changing conditions in American cities (e.g., Chicago, Los Angeles, New York), and both were instrumental in promoting a more explicit spatial approach to inequality outside of geography. However, while Sassen was interested in new cleavages formed by globalization – particularly the polarized growth of high end employment versus low end, informal and immigrant economies – Wilson focused on the areas and people left out of (if not left behind by) the process entirely.

Within geography, the renewed interest in inequality was also being informed by trends that emphasized poverty, policy, and social justice. All of these areas of interest once again emphasized matters of exclusion, material realities, class, poverty, and, critically, economic and spatial inequality. Geographers were regaining their role of social critics in matters of inequality, whether in terms of urban exclusion, the geography of neoliberalism, or the declining welfare state.

Polarization and Global Cities

The relationship between globalization and inequality is a massive topic of inquiry in its own right. Many commentators point to how globalization has exacerbated pre existing inequalities by advantaging certain well positioned regions – such as the Pearl River Delta in China – while further isolating regions such as Bihar in India. Incomes in an increasingly globalizing world are diverging considerably, a trend worsened by state reductions in efforts at redistribution (particularly the reduced taxation on the rich). As a result, we have witnessed the spectacular growth of the 'super rich' – the wealthiest 0.25% of the world's population, who now own financial assets equivalent to that of the rest of the world's population.

Fragmentation and divergence are therefore hallmarks of globalization at the regional level, but also at the urban level; globalization of the city seemingly intensifies inequality. By the 1990s, there was a substantial literature on how the bifurcation of the urban labor market, the demise of Fordism, and the retrenchment of the welfare state were creating a dual city in much of the developed world, increasingly polarized between the residual poor and ascendent elites. These trends were at the expense of the middle class, and were purportedly creating an hourglass (rather than an egg shaped) income distribution.

The concept of a bimodal, dual city was most prominently articulated by Saskia Sassen. Her polarized global city thesis became the most discussed theoretical framework for understanding the relationship between globalization and inequality/polarization in the developed world. She reasoned that the formation of large, global metropolitan regions such as New York City and Los Angeles (the two most polarized urban centers in the United States) necessarily entails rising spatial and socioeconomic polarization. Incomes were polarizing because unprecedented increases in the numbers of high level professionals working in transnational industries (their very growth stimulated by globalization) were coinciding with the downgrading of well paying manu facturing jobs and the increasing informalization and casualization of basic economic activities, informalizing being the only way they could still operate in these high costs locales. Unlike the growth of the middle class under Fordism, globalized economic growth at present "is based on an industrial complex that leads not to the expansion of a middle class but to increasing dispersion in the in come structure and in the bidding power of firms and households." Global cities were (and are still) at the ex tremes of this class polarization, as they feature the sharpest concentration of both high and low income groups, the latter of which are increasingly dominated by immigrants from the less developed world. Globalization is not the only culprit in this income polarization, but it has certainly accelerated the tendencies.

Sassen's focus on the local consequences of inequality and globalization has attracted substantial attention and criticism, including its disputable applicability to non American cities where racial and immigrant divisions are not nearly as pronounced, where the middle class is not shrinking, and where the welfare state is more active in countering gross inequalities. For example, polarization is nowhere near as pronounced in most European and Japanese cities. However, there is no doubt that many global cities, from Toronto to Berlin to Auckland, share certain common fault lines, including greater divergence between traditional poverty areas and gentrifying areas housing professional class households. The compart mentalization inherent in this globalized housing market may be seen as the cutting edge, if not the epitome, of the class divides that permeate the twenty first century city.

The Spatial Reorganization of Inequality: Concentrated Poverty

If Sassen was concerned with the polarization of the wealthy and the poor at the expense of the middle class in global cities, Wilson was more interested in the dis turbing fate of those completely disconnected from the global economy and increasingly decoupled from eco nomic cycles entirely. While economic inequality had always been associated with economic segregation, Wil son detailed the spatial reorganization inherent in a qualitatively different kind of urban marginality and in equality of the so called underclass: the permanently unemployed and welfare dependent living in geo graphical and social isolation from mainstream society.

In his 1987 book *The truly disadvantaged*: *The inner city*, *the underclass*, *and public policy*, Wilson emphasized the negative effects of concentrated poverty and spatial en trapment on African American inner city neighborhoods, itself the outcome of economic restructuring (especially de industrialization and the spatial mismatch of jobs as employment opportunities flee to the suburbs), past racism, migration, and demographic shifts. Critically, the spatially contiguous and temporally stable nature of high poverty neighborhoods (traditionally measured as more than 40% of the households are poor) led to social isol ation from mainstream (job) opportunities and norms, which in turn perpetuated high unemployment and a tangle of social pathologies and intergenerational poverty. This is partially due to the fact that poor people are more dependent on and influenced by their immediate milieu than middle class or wealthier groups. The spatial sep aration and concentration of the poor magnified their problems in a myriad of ways. Children growing up in high poverty neighborhoods encountered a hostile and generally negative environment, attended low performing schools that have few resources, lacked networks for durable employment, and were exposed to the tempta tions of street life without strong social buffers.

Despite the significant criticisms leveled against the concentrated poverty thesis – particularly the stigma tization inherent in the term underclass – Wilson did draw important attention to the spatial nature of this new, more magnified urban poverty and inequality. Economic restructuring, geographical space, and race were all im plicated in the formation and sustaining of these new ghettos and their systematic exclusion from mainstream employment and opportunities. This form of 'placism' would be difficult to transcend as it combined deep seated class and racial dimensions.

The effects of this concentrated poverty persist over time, and currently extend to other forms of inequality as well. Living in excluded places usually means dis proportionately higher likelihoods of being exposed to environmental pollutants, being more vulnerable to en vironmental disasters (e.g., earthquakes, heat waves, hurricanes), and lacking access to positive environmental amenities, such as green space. The stressful and de graded social environment also translates into inferior health outcomes. Crime concentrates disproportionately in these same areas, adding further disrepute. Finally, homelessness may be seen as a key demarcator and the tip of the iceberg of extreme and concentrated poverty, as well as being the most visible indicator of inequality in American and other cities. Structurally, the confluence of greater economic vulnerability from the 1970s onward – fewer well paying jobs, cutbacks to welfare – and dwin dling housing affordability in central cities led to the emergence of a large pool of precariously housed indi viduals. For some, these structural constraints intersected with personal vulnerabilities – mental illness, substance abuse, violence, foster care, etc. – to create a surge in homelessness in the 1980s. Concentrated poverty areas are frequently the originating location of many homeless people, as well as the invariable target for homeless services saturation.

The affluent also became more spatially concentrated during the 1980s and 1990s as the classes began to more forcefully pull apart, both socially and spatially. This concentrated affluence was characterized by several urban phenomena: growth of gated and privatized communities, as a means of reasserting some semblance of control over territory; hypergentrification of desirable, amenity rich neighborhoods in older cities; and the erection of bunkered downtown and edge city developments that act as command and control centers for the global city. These class divisions were reinforced and made increasingly rigid through geography, in some ways representing a return to the excesses of the Gilded Age when the poor lived in misery while the rich lived in complacent ignorance.

The Mediating State? Responses to Inequality in the Developed World

All of these trends beg the question of how (or whether) inequality is being mediated by the state within the developed world. As previously noted, the capitalist state's role in shaping and producing inequality is secondary, essentially limited to offsetting or worsening existing market inequalities. Different welfare state interventions will filter the restructuring processes outlined by Sassen and Wilson in different ways, mediating unequal consequences and producing different impacts at national and local levels. The state's mediating role, however, must also be placed within the context of globalization and neoliberalism, both of which justify weakened state interventions in favor of a more unfettered, commodified capitalism. There is a sense that while societal needs and inequalities are mounting, the state's interest in rectifying them is waning. Some commentators go so far as to say that the state is now actively complicit in reinforcing and magnifying inequalities, but this too is geographically variable – not all nations in the developed world have followed the same path in reproducing inequality.

As a key mediator of inequality, the role of the welfare state is to provide a support system for those deemed deserving of being unable to fully participate in the labor market – the disabled, single mothers, and so on. In terms of inequality, the welfare state mediates in several ways: (1) it is a key support system for the lowest strata in society (i.e., those who do not work in the paid labor market), such that any shifts in support will have potentially critical effects on the entire spectrum; and (2) it can mitigate the worst impacts of economic restructuring and market shifts, if there is the political will. In many (but not all) developed nations over the past 30 years – most prominently the United States – the welfare state has largely been dismantled, devolved, and privatized, deteriorating the living conditions of the welfare dependent and exacerbating the differences between the worst off vis a vis the rest of society. The American welfare state was always incomplete and residual, intervening only for market or family failure. Since its apogee with the War on Poverty programs in the 1960s, however, the American welfare state has suffered a lengthy demise with the decline of interventionist ideals and the rise of deregulated neoliberalism. In response to seemingly intractable problems of declining international competitiveness and budget deficits, the Reagan administration began to implement broad cutbacks to the welfare system, eliminating or downsizing many of the War on Poverty programs (e.g., public housing, welfare payments). With the end of the 1980s boom and the advent of the early 1990s recession, caseloads increased dramatically. As America slowly emerged from this last recession, a consensus developed around the idea that the federal welfare state required dramatic change. In 1996, President Clinton signed the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA). The legislation fundamentally altered the federal welfare state, dismissing any entitlement to government assistance and shifting the focus from income maintenance to job training and employment. Moreover, it is based on the 'work first' policy, despite evidence that reveals important spatial and personal barriers to employment among certain segments of the welfare dependent population. By prioritizing 'personal responsibility' over 'work opportunities', PRWORA is a further step toward eliminating welfare, not poverty, and epitomizes the neoliberal, workfare–welfare state. While some of these key provisions have subsequently been revised, the 1996 Act set into motion three irreversible shifts to the federal welfare system: (1) the increased localization of welfare provision and administration to the state and local level, through block grants and devolution of welfare provision responsibilities; (2) emphasis on individual rather than collective responsibility for poverty, in the form of lifetime benefits limits; and (3) the preference of self sufficiency over basic entitlements, embodied in the welfare to work provisions.

Outside of the United States, however, the assault on the welfare state has been less drastic. This is an important point, as many European nations have encompassing and universalistic, rather than residual, welfare states that actively intervene to provide strong supports to not only the nonworking poor but the entire populace. For instance, welfare states in Scandinavia and Western Europe maintain strong monetary and housing supports for the unemployed. As a result, there is far less likelihood of being consigned to live in neighborhoods of last resort, thereby reducing the overall spatial polarization so striking in American cities. Nonetheless, the longstanding stagnation of welfare payments in Europe, Canada, Australia, and New Zealand have had negative impacts on the most vulnerable populations and places. At the very least, there is a sense that even in the most egalitarian of nations, the welfare state is no longer as strong or supportive in rectifying inequality as it once was.

Beginning in the 1990s, there was also a sense that the state – and again, particularly in the US – was moving away from simply cutting back on inequality reducing welfare supports to promoting more explicitly inequality reproducing and inequality magnifying policies. This shift must be seen within a context of globalized and neoliberalized competition that particularly forces the local state to emphasize image and entrepreneurialism over equality. These policies targeted those excluded by previous hollowing out of the welfare state: the poor, unemployed, welfare recipients, immigrants, and low skilled laborers. These groups were now subject to greater vilification, state authoritarianism, and punitive regulation. Examples include the de funding of local housing and welfare programs, the demolition and de concentration of public/social housing, the expressed pro gentrification practices within inner city areas, and the criminalization of the visibly poor. The homeless are particularly targeted under new, more aggressive policies. In many cities worldwide, the need to maintain a positive image for tourists and the middle/upper classes results in the proliferation of antihomeless ordinances. These ordinances make it difficult for homeless individuals to access prime spaces for panhandling, recycling, sleeping, and so on. Concurrently, many marginal spaces in the city are being revalued and thus disappearing.

Skid Row, Los Angeles (**Figure 2**) epitomizes many of these ugly trends to regulate (and spatially relegate) extremes of inequality. The 50 block area east of Downtown Los Angeles acts as a containment zone for the homeless and their services, and is home to over 8000 destitute souls. Skid Row contains a disproportionate amount of shelters, drop in clinics, substance abuse treatment centers, and the like. Social service saturation, along with the sense that Skid Row is a community of last resort, make it difficult for its denizens to actually escape homelessness. Although routinely threatened with displacement, Skid Row has endured because it acts as a convenient container for society's castoffs. As an excluded ghetto, Skid Row represents the very lowest rung of the income strata, no more than a few blocks from the gentrifying Spring Street corridor, the bustling Latino shopping corridor on Broadway, and the corporate high rises of Bunker Hill.

The containment policy of Skid Row is not the only instance of new modes of social and spatial regulation to deal with inequality in the Los Angeles city region. Since its rise to global city status, the preexisting social and spatial cleavages in Los Angeles have intensified to the point where different areas have now become more homogenized, while the social (and geographical) distance between areas has increased. The regulation of gross inequalities has taken other forms that implicate the local state, including privatization, fortification of the built environment, heavy handed policing and surveillance, and the mass incarceration of populations deemed dangerous. For those at the bottom, the state has moved from welfare to outright repression, with occasional resistance in the form of sporadic and sometimes spectacular unrest (i.e., in 1992). In *City of quartz*, Mike Davis detailed the dramatic reconfiguration of public space in Los Angeles to control and contain the deleterious outcomes of deep seated inequalities: "Welcome to post liberal Los Angeles, where the defense of luxury lifestyles is translated into a proliferation of new repressions in space and movement, undergirded by the

Figure 2 Skid Row, Los Angeles, 2000.

Figure 3 East Hollywood, 1999.

ubiquitous 'armed response'. This obsession with physical security systems, and, collaterally, with the architectural policing of social boundaries, has become the zeitgeist of urban restructuring, a master narrative in the emerging built environment of the 1990s." Private citizens have increasingly taken matters into their own hands, fortify ing *in situ* to defend their properties from the fear of crime (**Figure 3**).

The Less-Developed World

In the less developed world, inequality has also surged and urbanized since the 1980s, but this resurgence has been harder to conceal and contain given its greater visibility and scale. There is a sense that globalization is exacerbating inequality through enhanced cumulative causation and increasing returns to certain well situated regions, and the opposite effect for others. Examples of regional divergence include the gaps between North eastern Brazil and the Sao Paulo–Rio de Janeiro corridor, southern versus northern Mexico, coastal China versus the interior, and the high tech poles in southern India versus stagnating provinces in the north. Moreover, many of the key manufacturing centers of the less developed world – Mumbai, Sao Paulo, Johannesburg, and so on – have experienced stagnation coupled with unabated population growth, leading some commentators to speak of over urbanization, whereby city growth occurs with little employment growth or industrialization. Many new arrivals are fleeing economic stagnation in the country side, victims of agricultural de regulation and state in action. Issues of relative inequality are amplified by concerns over absolute deprivation, as many of the poor in the less developed world lack access to even basic necessities and infrastructure (housing, water, health, education, electricity, sanitation, etc.) resulting from extreme social exclusion. As such, poverty in the less developed world is both relative vis a vis the rest of so ciety, as well as absolute in terms of being unable to attain basic survival needs.

In the face of this lopsided reality, the role of the state has been minimal. Indeed, state intervention is more likely to take the form of bulldozing squatter settlements or supporting middle class housing. State abdication in redressing inequality, or even providing taken for granted services, is a symptom of many longstanding problems, namely the failures of developmentalism. Indeed, ac cepting foreign aid usually means following a neoliberal path: opening up protected markets, reducing state interventions and guarantees, and so forth.

There is no doubt that mega cities of the less developed world are at the forefront of the global urbanization of inequality. During the 1970s and 1980s, the majority of the poor living in Latin America moved from rural areas to urban areas, especially to mega cities such as Mexico City, Lima, Bogota, Caracas, Rio de Janeiro, and Sao Paulo. This trend is bound to continue; by the early twenty first cen tury, the majority of the world's poor will likely live in such urban areas. Most of these urbanized poor currently inhabit informal and excluded slums that present a grim future of blocked social and spatial mobility. Moreover, these slums are frequently in precarious natural environments, highly vulnerable to flooding and earthquakes, as well as human made industrial hazards. Slum densities are skyrocketing, in direct contrast to the spacious neighborhoods of the emerging and established middle and upper classes. From Johannesburg to Manilla, these more powerful groups are pushing to create insulated living spaces that further remove and displace the very poor (squatters, itinerant street hawkers, and entrepreneurs) from prime urban spaces. Those displaced, along with arrivals from the countryside, are shunted to peri urban development.

While understandably unique, Sao Paulo also typifies many of these disturbing trends in the less developed world. As Brazil's most populous and most prominent (albeit peripheral) global city, Sao Paulo also ranks among the largest metropolitan conglomerations on Earth. The region produces about one third of Brazil's GDP and over half of its national industrial output, clearly making it the dominant global banking center, urban center, and economic motor. As we shall see, however, being a global city is no guarantee of a good life for the majority of Paulistanos. Sao Paulo is clearly both a First and Third world city – too poor for the former, too rich for the latter. Various contradictory processes – globalization, de industrialization, de concentration and polycentricity, tertiary expansion, polarization, social exclusion, peripheralization of the poor – translate into bewildering and flagrant patterns of economic and spatial inequalities that exceed the local state's fiscal capacity. Twenty percent of the population lives in informal housing (a proportion on the low side for the less developed world), largely in socially excluded peripheral areas, crime rates in some areas approach civil war conditions (20% of homicides can be attributed to the police!), and the wealthy increasingly separate themselves within fortified developments and move about by helicopter. In Sao Paulo, the rich are extraordinarily rich, and the poor are ruinously poor.

One way to approach inequality in Sao Paulo is to examine its internal character and patterns of development, and by extension the role of space and place in reinforcing inequalities. Sao Paulo's intra urban geography of inequality reflects both the degree of exogenous links to the world economy, as well as local forces and legacies. To many, the most prominent feature of Brazilian cities (and the less developed world as a whole) is the stark contrast between rich and poor, to the point where one can talk about a dual built environment – one formal and legal, the other self produced, informal, and illegal. In Sao Paulo, these precarious zones are epitomized in peripheral *favela*s or shantytowns. Arguably disconnected from global flows, *favelas* are very much locally produced and consumed. However, the people who live there are frequently displaced by global forces, especially with regard to agricultural reforms that have pushed farmers all over the less developed world into cities. They are part of the informal and irregular spaces of the city, and frequently built on steep slopes, river banks, and swamps. There are now over 12 000 *favelas* in Sao Paulo. At the opposite end of the global and class spectrum, self proclaimed 'global investment zones' are also emerging, essentially planned edge cities that recreate (and replace) the original CBD, but with more control, less congestion and danger, and designed expressly as head office space for multinational corporations. As a result, Sao Paulo increasingly suffers from the same intense multicenteredness and sprawl that one sees in North American cities, and with the same negative consequences for social cohesion. The image here (**Figure 4**) shows the gleaming skyscrapers of the World Trade Center de Sao Paulo (WTCSP), a business complex complemented by a Hilton Hotel, convention center, and exclusive shopping that is intensely connected to the rest of the world but very much disconnected from its neighboring *favela*, shown here in juxtaposition.

Figure 4 World Trade Center de Sao Paulo (background) Jardim Edith *favela* (foreground), 2006.

Just as the corporate world has sought refuge from the core of the city, so increasingly have the wealthy, particularly in reaction to high crime rates and the perceived lack of spatial separation between the classes. In the most extreme case, an entire gated city has been constructed (Alphaville) that contains homes for 32 000 residents, 1400 businesses, several shopping centers and parks, a private university, and a new Quality Suites hotel. Alphaville is an expression of the increasing concentration and self segregation of wealth and the increasingly divergent lifestyles of urban society under the influence of globalization. The people who live there are consciously following a globalized lifestyle, mimicking a North American privatized and suburbanized approach to everyday life that is not so different from a place like Southern California. Further, the exclusive control of these spaces by private capital can be interpreted as an integral part of development patterns, characterized by de regulation and flexibility, that have emerged in Latin America in the face of reduced state spending. For those in the middle and upper classes who have not fled to gated communities, the alternative is to reside in gated condominium towers or, like Los Angeles, take the matter of protection into their own hands. Similar levels of fortification, fear, and violence are found in the most unequal societies, particularly in places such as Argentina, Venezuela, and South Africa, as well as appearing in places with worrisome rises in inequality, such as Russia, Indonesia, and Trinidad and Tobago. In some of these cases, the significant presence of oil revenue has inevitably sharply widened inequality.

Current Approaches to Inequality in Human Geography

In sum, geographers have been examining the spatial implications of inequality since the 1970s, emerging out of the foundational subfields of Marxist urban political economy and welfare geography. By the 1980s, geographical interest in inequality had ebbed significantly, only to be reawakened in the 1990s. This reawakening was spurred on by the emergence of globalized inequality and its particularly urban manifestations, namely the rise of income polarization and concentrated poverty. In both of these cases, geography was deeply implicated in the divergent, unequal fates of various urban populations. There was also the sense that the state – whether in the more or less developed worlds – was more bent on containing and magnifying, rather than alleviating, growing urban inequality.

At present, geographers are continuing to deepen their interest in issues of economic inequality, particularly as they overlap with major social demarcators such as race and gender. Interest in economic inequality has also widened to include critical geographers of all stripes, not just those influenced by Marxist political economy. The study of inequality is now a fundamental component of critical geography, and approaches include qualitative, case study and even some applied, policy oriented and mixed method/quantitative research. Geographers have grown more comfortable in their role of social critics in matters of inequality, and as such topics range widely. There is continuing work on how neoliberalism is creating increasingly unequal and disparate geographies, to the point where it is now up to individuals to purchase/appropriate their own welfare services, the so called 'DIY welfare state'. Geographers also appreciate that growing inequality is not necessarily an inevitable outcome of present trends. For instance, geographers are articulating Henri Lefebvre's utopian concept of the "rights to the city" to conceptualize strategies to address urban inequality. In this concept, the goal is for all citizens to be able to forge urban life in a mutually beneficial way. Rights to the city can also help frame resistance to oppressive state mediations of inequalities. At the very least, the resurgence in interest in inequality and social justice among geographers promotes a sense that even if full equality is difficult to achieve, providing the greatest benefit for the most disadvantaged segments of society is better than nothing.

The current tendency to invariably castigate the state for magnifying inequality is also being slowly unpacked. It is perhaps not too late to give up on the state's capacity or even willingness in ameliorating the worst inequalities. For instance, the current mayors of the two largest (and most unequal) cities in the United States – New York City and Los Angeles – have been pushing for more equitable urban policies that run against the neoliberalized trend, in terms of affordable housing, the homeless, recent immigrants, and living wage concerns. Indeed, there is probably more ambivalence regarding extreme inequality than outright state punitiveness. More to the point, much of the work being done now to manage the extremes of inequality is in the hands of voluntary sector agencies, many of which focus on providing genuine help to the have nots, even if the sector rarely questions the larger structural conditions that perpetuate inequality. In many European nations, the welfare state and unions remain vital forces in the face of rampant neoliberalism, ensuring that the gaps do not grow too wide.

See also: Equity; Modifiable Areal Unit Problem; Poverty; Social Justice, Urban; Uneven Development; Uneven Regional Development; Welfare Geography.

Further Reading

Badcock, B. (1984). *Unfairly structured cities*. London: Blackwell.
Beaverstock, J., Hubbard, P. and Short, J. R. (2004). Getting away with it? Exposing the geographies of the super rich. *Geoforum* 35, 401 407.

Caldeira, T. (2000). *City of walls: Crime, segregation and citizenship in Sao Paulo*. Berkeley: University of California Press.
Chakravorty, S. (2006). *Fragments of inequality: Social, spatial and evolutionary analyses of income distribution*. New York: Routledge.
Coates, B., Johnston, R. J. and Knox, P. (1977). *Geography and inequality*. London: Oxford University Press.
Davis, M. (1990). *City of quartz*. New York: Vintage Press.
Davis, M. (2006). *Planet of slums*. London: Verso.
DeVerteuil, G. (2006). The local state and homeless shelters: Beyond revanchism? *Cities* 23, 109 120.
Gregory, I., Dorling, D. and Southall, H. (2001). A century of inequality in England and Wales using standardized geographical units. *Area* 33, 297 311.
Harvey, D. (1973). *Social justice and the city*. London: Edward Arnold.
Kodras, J. (1997). The changing map of American poverty in an era of economic restructuring and political realignment. *Economic Geography* 73, 67 93.
Massey, D. S. (1996). The age of extremes: Concentrated affluence and poverty in the twenty first century. *Demography* 33, 395 412.
Sassen, S. (2001). *The global city: New York, London and Tokyo* (2nd ed.). Princeton, NJ: Princeton University Press.
Smith, D. M. (1987). *Geography, inequality and society*. Cambridge, UK: Cambridge University Press.
Smith, D. M. (1994). *Geography and social justice*. London: Blackwell.
Smith, D. M. (2000). Moral progress in human geography: Transcending the place of good fortune. *Progress in Human Geography* 24, 1 18.
Soja, E. (2000). *Postmetropolis: Critical studies of cities and regions*. London: Blackwell.
Wacquant, L. (1999). Urban marginality in the coming millennium. *Urban Studies* 36, 1639 1647.
Wilson, J. (1987). *The truly disadvantaged: The inner city, the underclass, and public policy*. Chicago: University of Chicago Press.
Wolch, J. and Dear, M. (1993). *Malign neglect: Homelessness in an American city*. San Francisco: Jossey Bass Publishers.

Relevant Websites

http://www.inequality.org
America's Growing Economic Divide Resources for Journalists, Teachers, Policymakers, Citizens.
http://www.worldbank.org
Inequality Around the World, World Bank.
http://www.ucatlas.ucsc.edu
The UC Atlas of Global Inequality.
http://www.ucl.ac.uk
Understanding Slums, London's Global University.
http://www.wider.unu.edu
WIDER: The World Income Inequality Database.

Informal Sector

A. G. Aguilar, Universidad Nacional Autónoma de México, Mexico City, Mexico
E. P. Campuzano, Instituto Politécnico Nacional, Mexico City, Mexico

Glossary

Developing Countries The developing countries are those characterized by their low standard of living, poor infrastructure, and generally nonindustrialized. Other common used terms to describe this group include less-developed countries, Third World countries, or nonindustrialized countries.

Introduction

For many years, the informal sector has been a key issue in the literature on economic development, principally in developing countries, and has been a research topic in several disciplines such as geography, anthropology, sociology, and economics. These disciplines have pointed out the importance of studying this complex phenomenon that includes activities as diverse as traditional agri culture, livestock and fisheries, retail, street vendors, home based workers and home production, small work shops, and bricklaying, and even some illegal activities such as the drug trade. Even with a vast body of literature on the theme, the conceptualization of informality is confused due to the lack of theoretical frameworks and appropriate empirical data. The problem of conceptual ization is more common than one might expect and is illustrated by the quantity of adjectives and nouns used to describe the informal sector, such as black, underground, hidden, shadow, clandestine, illegal, unregulated, precar ious, invisible, nonofficial, subterranean, etc.; the nouns used include (informal) economy, sector, activities, and employment/work/jobs. Every characterization de pends on the goals, context and theoretical background of the research. In spite of its fuzzy nature and uses, almost all scholars agree that informality is an important phe nomenon because of its economic and social effects.

Traditionally, the informal sector has been seen as a Third World problem, but for some years, these kinds of problems have increased in the developed world and in the transition countries. Recent evidence in developed countries has shown a growth in the informal economy and labor market, while the transformation of the eco nomic model of the ex socialist countries has been accompanied by a rapid rise of the informal economy – including illegal activities.

Globalization has brought new forms of production, which include new forms of regulation (or de regulation), new types of industrial processes of organization, and the growth of the tertiary sector. These critical changes have encouraged the utilization of unskilled workers, small size enterprises, and family workers in developing countries. Meanwhile, larger cities in the developed world are wit nessing the growth of part time jobs, and the polarization of their labor market. Thus, in different ways, global ization has generated new types of informal jobs and ac tivities (**Table 1**).

Conceptions and Measures of the Informal Sector

Defining the Informal Sector

Characterized by its vagueness, the concept of the in formal sector has been used in different ways, contexts, and for different goals: interpretations range from sur vival activities to illegal procedures. We can distinguish two main strands in the development of the concept: the earlier views and the recent ones.

Keith Hart in 1973 described formal income oppor tunities as salaried jobs and informal ones as self employment. This characterization was adopted by the International Labour Office (ILO) in a mission in Kenya, Africa. In a study in this country, informal employment was conceptualized as labor intensive production, family owned, small scale operations, low entry barriers to entrepreneurship, and unregulated competitive markets.

Following this study, the earlier conceptualizations in the informal sector saw it as encompassing marginal ac tivities which are unregulated and unregistered. These activities include unreported income from the pro duction or trade of legal goods and/or services, either from monetary or from barter transactions. In most of the cases, they comprise small scale production, semilegal, low productivity, and family based production.

Recent conceptualizations of the term define the in formal sector as an important aspect of the economic and social dynamics of any country, but especially in less developed and in transition ones. They include issues like tax evasion (both direct and indirect), social security fraud, and the avoidance of labor legislation. In one strand of the new conceptualizations, the central char acteristic of the sector is the role played by the small entrepreneurs in the economic transformation of their situation and the opportunities created by these actors, including the positive impact on the decrease of unemployment; whereas in other conceptualizations, the informal sector plays a functional role in the

Table 1 Urban and Rural Informal Employment (Absolute and Percentage of the Total Work Force) by Country

Country	*Year*	*Urban/ rural*	*Numbers in 1000s Total*	*Percent of total employment Total*
Informal sector (harmonized definition)				
Ethiopia	1999	Urban	1118.5	49.2
		Rural	2137.7	50.7
Peru	1999	Urban	3606.1	53.8
Georgia	1999	Urban	39	5.6
		Rural	13.5	1.3
India	2000	Urban	29 680	38.1
		Rural	35 880	55
Latvia		Urban	95.4	13.7
		Rural	30.6	18.1
Russian Federation	1999	Urban	2228	4.5
		Rural	560	3.7
Turkey	2000	Urban	1035	10.2
		Rural	388	15
Informal sector (national definition)				
Benin	1999	Urban	275.5	46
Cameroon	1993	Urban	119	57.3
Ethiopia	1999	Urban	1149.5	50.6
		Rural	3665.3	86.9
Ghana	1997	Urban		78.5
		Rural		93.4
Madagascar	1995	Urban	239	57.5
Mali	1996	Urban	370.6	71
		Rural	805.5	
Niger	1995	Urban	302.6	
South Africa	1999	Urban	1549	21.3
		Rural	1157	37.4
Tanzania	1995	Urban	345.9	67
Zimbabwe	1993	Urban	154.1	
Brazil	1997	Urban	12 870.4	27.3
Mexico	1999	Urban	7448.5	29.7
		Rural	1693	47.2
Georgia	1999	Urban	73.4	14.2
		Rural	29.9	3.1
India	2000	Urban	39 970	51.3
		Rural	39 740	61
Kazakhstan	1996	Urban	962.7	17.3
Kyrgyzstan	1999	Urban	139	29.4
		Rural	55.1	18
Nepal	1999	Urban	375	64.8
		Rural	1282	76.2
Pakistan	1997	Urban		61.2
		Rural		68.2
Philippines	1995	Urban	539.3	17.3
Latvia	1999	Urban	127.2	18.2
		Rural	30.6	18.1
Lithuania	1997	Urban	93	8.5
	2000	Urban	7.8	41.3
		Rural	193.8	74.2
Russian Federation	2001	Urban	4525	9.2
		Rural	3654	23.8
Turkey	2000	Urban	962.8	9
		Rural	356.8	13.2
Ukraine	1997	Urban	755.9	4.9
Small or micro enterprises (national definition)				
Senegal	1996	Urban	665	
Uganda	1993	Urban	654.9	83.7
		Rural	11 452	
Argentina	2000	Urban	3067.5	40.6
Bolivia	1999	Urban	1296.3	64.3
Colombia	2000	Urban	3913	60.9
Costa Rica	2000	Urban	132.5	20.2
		Rural	116.4	17.6
Chile	1997	Urban	1368.4	30.3
Ecuador	1997	Urban	1223.6	40
Nicaragua	2000	Urban	305.1	54.7
Paraguay	1999	Urban	554.7	45.9
Peru	1999	Urban	4388.5	62.8
		Rural	3579.7	88.1
Uruguay	2000	Urban		29.7
Venezuela	1997	Urban	3698	46.3
Thailand	2000	Urban	3272.3	47.1
		Rural	20 159.7	77.4
Côte d'Ivoire	1996	Urban	414.2	52.7
Gambia	1993	Urban	100.7	72.4
Bangladesh	1993	Urban	197.5	10
Indonesia	1999	Urban	12 728.3	39.4
		Rural	42 966.7	76.1
Iran,Is.Rep.	1996	Urban	844.2	17.9
Myanmar	1996	Urban	1681.5	54.2
Croatia	1997	Urban	61.6	6.2
Poland	1998	Urban	720	7.6
		Rural	711	11.6

Source: International Labour Organization.

development and reproduction of capitalism by flex ibilization of work, and the reduction of wages. Also, it is recognized that there is not a clear dividing line between formal and informal activities.

Measuring the Informal Sector

As a result of the lack of conceptual consistency, meas uring the informal sector has been a difficult task. Broadly speaking, we can find two ways of measuring the sector: direct and indirect. The first one is used more in anthropological and sociological studies, because it im plies extensive field work and its principal goal is to determine the importance of the informal labor market. The second one is more used in socioeconomic studies and international comparisons.

Direct approaches

Household surveys have become an important source of information on the informal sector. In this kind of survey the key elements are the characteristics of the people engaged in informal sector activities, the participation of women and child labor, the characteristics of the households, and the survival strategies. This kind of measuring method is widespread.

Business surveys are really important in determining the characteristics of the enterprises involved in these kinds of activities. The principal points of inquiry are the number and type of workers, the nature of economic capital used, and the survival strategies of small estab lishments. In this category we can include the surveys taken by fiscal authorities.

There are few examples of household–enterprise surveys. They try to understand the relationship between entrepreneurial and household strategies and how one shapes the other, principally in less developed countries.

Indirect approaches

Despite the complex methods to measure the informal sector, one of the most used is the analysis of discrepancies between national expenditures and income statistics – residual GDP. Briefly, this consists in determining how much GDP is produced and how much money is really expended. The difference represents the shadow economy.

In labor terms, the difference between the official and actual labor force, i.e., the residual labor force, is assumed to be working in the informal sector. The formal labor force can be measured using fiscal records or social se curity records. The difference is assumed to be informal.

Labor Markets, Growth, Urbanization, and the Informal Sector in Developed and Developing Countries

Special attention has been paid to the informal sector in developing countries. Currently, the main conceptions of the sector/activities are related to the African, Latin American, and Asian contexts. The concept, at the beginning, stated the differential access to the urban labor markets in Africa and was characterized by low product ivity, small enterprises, and self employment. In the most recent views, informality is conceptualized as an oppor tunity to challenge the unequal distribution of poverty.

In developed countries the phenomenon has not been studied much, principally due to its small size; however, there is evidence of its growth. As opposed to the studies in developing countries, the principal focus in developed ones is the macroeconomic impact of these activities and state regulations.

Nevertheless, the macroeconomic indicators do not necessarily show an increase in the informal sector in both developed and developing countries. There are other ways to identify these activities, for example, labor data. Studies on global cities reveal a polarization in the labor markets. In global cities, the growth of services creates not only high level employment, but also less qualified jobs. While the former are taken by highly educated and skilled people, the unskilled jobs are taken by recent immigrants, the poorly educated, and mar ginalized populations.

Contrary to common thinking, there is evidence that educated and skilled people also engage in informal ac tivities, mainly due to tax reasons and the desire for economic freedom.

In the following section some empirical evidence is presented to appreciate what is happening with infor mality nowadays.

Current Patterns of Informal Sector Activity

In recent years, world agencies have shown a real interest in determining what informality is and how to measure it. There is no consensus yet, but some institutions such as the World Bank (WB), the International Monetary Fund (IMF), and the ILO have reported on the situation. In this section we present an overview of informal sector trends. We choose the GNP, published by WB, and labor market information, published by ILO, to present the current situation in the world.

In **Figures 1** and **2**, we can see the importance of the informal sector for less developed countries. While the Organization for Economic Co operation and Development

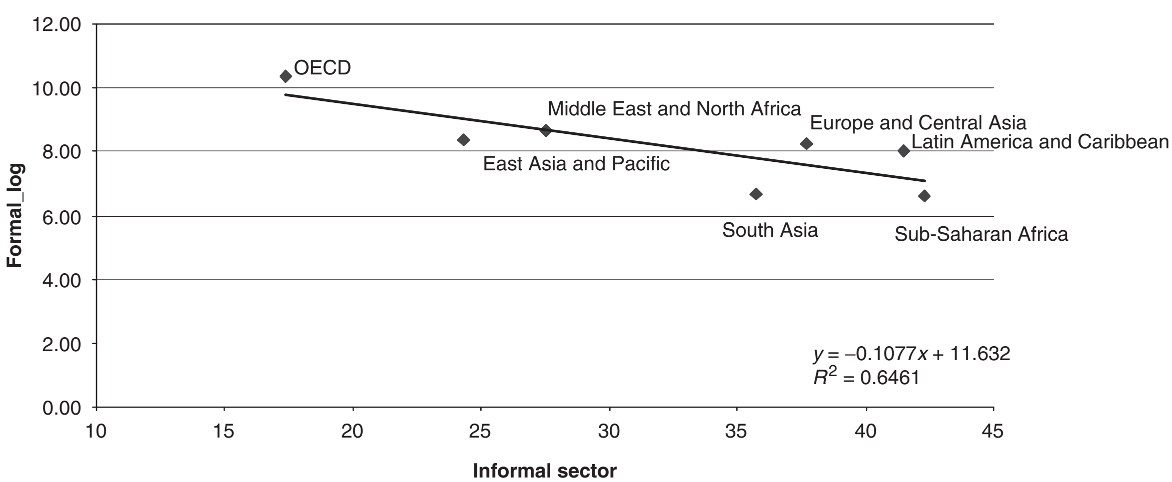

Figure 1 General income per capita and share of informal sector (% GDP). Source: Aguilar and Perez and data from the World Bank and the International Labour Organization.

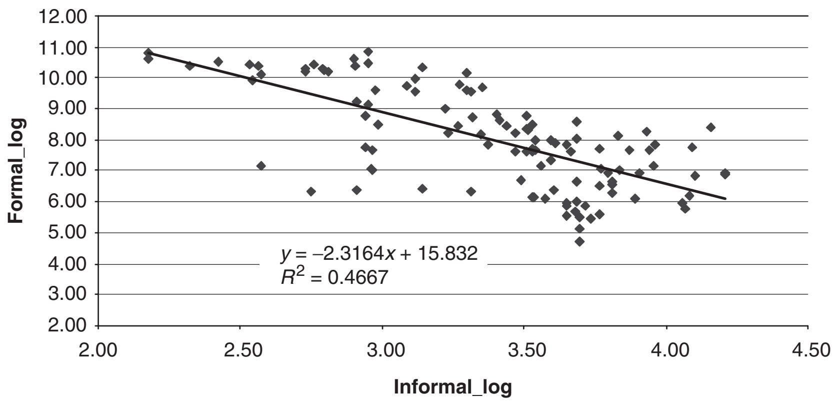

Figure 2 Income per capita and informal sector (% GNP). Source: Aguilar and Perez with data from the World Bank database and the International Labour Organization.

(OECD) countries have the higher per capita income and the lower share of informal economy in the total GNP, Asia, Latin America, and sub Saharan Africa countries have lower income per capita and a higher share of informal sector in their GNP (**Figure 1**). It seems that the higher the income, the lower the contribution of informal sector in the economy ($R^2 = 0.692$). Nevertheless, if we see the same graph at the national level, even when the contribution of the informal sector in developing countries is higher, the relationship is weak ($R^2 = 0.4667$). This means that there are important local factors at play in the shaping of the informal sector. The informal sector is not the same in developed, developing, and in transition countries. Every country has particularities that encourage/discourage informal activities.

Africa

The size of the informal economy in sub Saharan Africa was 42.2% in 2000. In countries like Nigeria, Tanzania, and Zimbabwe, the informal sector contributes more than 50% of the total GDP. In South Africa the share of the informal sector is 28.4%; meanwhile in Ethiopia, Malawi, and Mozambique it is around 40%.

The informal sector makes up around 50% of the total labor market. In Ghana and Uganda more than three quarters of their urban workers are in the informal sector, whereas in Ethiopia, Cameroon, Madagascar, Nepal, Tanzania, Côte d'Ivoire, Gambia, this percentage varies from 50% to 75%.

Latin America and the Caribbean

In Latin America and the Caribbean, the informal sector is around 40% of the total GDP. In Chile and Argentina it was less than 25% in 2000, whereas in Panama and Bolivia it represented around 65%. In the other countries, the share of the informal economy was between 25% and 60% of the GDP.

In terms of the labor market in Brazil, Mexico, Ecuador, Costa Rica, Chile, Argentina, and Uruguay, between 25% and 40% of their urban labor workers were engaged in the informal economy. In Peru, Bolivia, Colombia, Nicaragua, Paraguay, and Venezuela, the share of the urban informal employment represented more than 40%.

Asia and Pacific

In Asia and Pacific, the share of the informal economy was around 28% in the year 1999–2000. In Thailand, Sri Lanka, and Philippines, this percentage was higher than 40% and in China and Singapore this was lower than 15%.

OECD

In the OECD the share of the informal economy was around 17.4% in the year 1999–2000. In Switzerland and USA, this percentage was around 8.8%; meanwhile in Portugal, Spain, Belgium, Italy, Korea, and Greece it was higher than 20% but lower than 29%. In the other countries, the share of the informal economy represented 10–20% of the total GDP.

Central European and Middle Asian Countries

In the Central European and Middle Asian countries, the average of the informal economy was 37.7% in the year 1999–2000. The percentage of the informal sector in Azerbaijan and Georgia was 60.6% and 67.3%, respectively; in Serbia and Montenegro, Poland, Slovenia, Hungary, Czech Republic, and Slovak Republic, this percentage was between 19% and 30%. In the other countries, the share of the informal economy represents more than 25% and less than 50%.

For none of the countries the percentage of the urban informal workers surpasses 40% of total urban employment. Lithuania, Kyrgyzstan, Latvia, Kazakhstan, and the Russian Federation have the larger percentages.

Approaches to the Informal Sector

It is clear that there is no polished theory on the informal sector, but some scholars have pointed out that there are three main approaches: dualistic, structural, and legalistic or institutionalist.

The 'dualistic' view defines the informal sector as marginal or separated from the formal one. The sector is a niche for poor people in periods of crisis and its persistence is due to the lack of sufficient modern labor market opportunities, the faster rate of growth of population in relation to job opportunities, or poor economic growth.

'Structuralism' sees the informal sector as subordinate to the formal one. In this perspective, the informal sector plays a very important role in the development of capitalism via the reduction of labor and input costs. The informal sector provides cheaper labor and input to formal enterprises allowing for the improvement of competitiveness of the latter. From Milton Santos' point of view, there are two main circuits in the economy of developing countries: an 'upper circuit' and a 'lower circuit'. The first one represents the modern economy, is commonly high tech, and consequently has a sophisticated system of production and commerce. On the other hand, the ower circuit is characterized by its small scale, the poor use of technology, and the intensive use of labor. Contrary to the dualistic view, for structuralists – including Milton Santos – the informal sector is not marginal or backward but it is part of capitalism, because the upper circuit dominates and requires the lower circuit to reproduce itself.

The 'legalist', or 'institutionalist', view sees the informal sector as micro entrepreneur's response to the state's over regulation and the lack of clear property rights. For Hernando de Soto – one of the main theorists of this perspective – the main factor in the growth of the sector is the role played by the state, which does not allow the micro or small entrepreneurs to develop their economic potential. These actors prefer to operate outside the regulations to reduce costs, time, and efforts instead of following the complex system of the state's regulations which increases costs and crushes the 'entrepreneurial spirit'.

Although there is no refined theory on informality, there are many recurrent themes. In the next section, the most common are presented.

Principal Themes

In this section our objective is to present a taxonomy of the themes that have been related to the informal sector and appear more frequently in the specialized literature. The division is not static but actually is more complex than we present here; many factors are related to each other in different ways. The section divides themes into three main categories: economic, institutional, and social. The first two have received more attention, whereas the latter one is relatively new. In different ways every vision uses almost all of the factors enumerated below.

Economic themes

Settlement type has been shown to have a strong relationship with the informal sector. In fact, where the informal sector is conceptualized as including waged employment, cities are more likely to have large amounts of workers in the informal sector. Developing countries with large cities have a large quantity of informal workers due to concentration of the labor markets, large amounts of rural out migration, and lack of opportunities in the formal labor market. In addition, urban informality includes a vast diversity of activities and jobs: from craft production or street commerce to companies linked to large transnational companies. Therefore, it is in the cities where the informal sector is more complex and palpable.

Urban labor markets have received much more attention than rural ones. As the informal sector has been conceptualized as an urban employment problem, the main contributions have emerged from this perspective. The increase of unemployment due to the high rates of rural–urban migration, the lack (or decrease) of social benefits, underemployment, subminimum wages, and the presence of unregulated segments of the labor market have been identified as the main contributors to the informal sector. Poor people, in urban scenarios are more likely to be involved in nonformal activities because they cannot access the modern sector of the economy.

Labor markets have been linked to poverty, principally in developing countries. One of the most influential themes is related to the relationship between national economic size and growth and the informal sector. From this point of the view, greater the size and growth of an economy, the less likely it is that people will engage in informal activities. This kind of vision was popular at the beginning of theorization on the informal sector and nowadays it is part of more complex explanations. It is recognized that the rate of growth of one economy is important but does not explain the entire problem. The size of one economy is related to the number of qualified workers, the availability of credit, and the number of small , medium , and large size enterprises. As we have seen, all these factors encourage/inhibit the growth of the informal sector.

Both traditional and contemporary conceptions recognize that the size of the firm is intimately related to the informal sector. Micro and small size enterprises are more likely to engage in the informal sector due to their reduced capital, their incapacity to pay taxes, their flexibility, and greater opportunities to escape state controls. The larger the firm, the lower the possibility of that firm being involved in informal activities because it has enough resources to access credit, a qualified labor force, and to pay taxes.

Furthermore, the industrial structure is also important in the growth of the informal sector. In this sense, the relations between industrial enterprises and the pro ductive chains encourage the growth of informal sector by the de concentration of processes, subcontracting, flexibilization, or the combination of two or more of these phenomena.

In the period of late capitalism, globalization and industrial restructuring have promoted the increase of informal activities through different forms of subcon tracting. Generally, large firms in the developed world subcontract medium and small enterprises in the de veloping world and in the transition countries to produce goods and services. From large companies to home workers and home production there is a complex chain of inter mediaries, contracts (not all legal), and forms of production. At the end of the chain, there are families, small towns, or small firms that are unregulated by labor laws.

In the developed world, the de concentration of production and the seeking of more efficient production methods of production have promoted the increase of the informal sector. Nowadays, practices like outsourcing (buying manufactured products instead of producing them) and subcontracting are common; the result is an increase of the total number of part time jobs and the deterioration of working conditions.

Institutional themes

Generally, the informal sector has been seen as consisting of widely unregulated or unregistered activities. In this case, the principal component is related to the role of the state, state agencies, and the avoidance of many rules.

State regulations are the most utilized issue related to the informal sector. In fact, for many scholars the informal sector could not be defined without its counterpart: the state. The state through legislation and regulations defines what informal activity is and what it is not, and thus its existence; whereas for others, the excessive regulations encourage the growth of the informal sector. Clearly, nowadays, there is a trend to de regulate many markets, including labor ones. Trying to attract Foreign Direct In vestment (FDI), many national states are determined to reduce or cut off many labor rights, such as pensions and social care, and to allow subcontracting or to encourage part time jobs and small scale production. This view has been supported by multilateral organizations such as the IMF and the WB and has been followed by almost all countries.

The argument is more or less clear: the tax system is linked to the growth of the informal sector. The increase of tax burdens encourages the increase of the informal sector, whereas from the point of view of some institu tionalists there is not only the tax burden but the increase of the tax system's complexity and the excessive regu lation over labor markets, licensing requirements, and right to property that de incentivize productivity in creases of productivity. Stronger and more complex le gislation is important in the existence of the sector. Hence, the informal sector is the response of economic actors to the over regulation by the state. For example, in some economic literature the increase of the tax burden influences the growth of the informal sector due to the differential costs of employing or being employed with out taxes. The incentive to avoid taxes is reinforced by the fact that even when many companies or people are engaged in informal activities, they are not excluded from public goods or basic infrastructure. On the other hand, for some institutionalists, the informal sector is a reaction to the inefficiencies in public sector service provision. It has been recognized that if economic actors do not have enough incentives in terms of (cheaper) provision of public services, they are encouraged to pursue other routes, such as the informal sector.

In many cases, state policies tolerate or even promote informal activities at different scales. In developing countries, governments have given incentives to en courage the development of small firms (mostly informal) to decrease unemployment. This situation is common in countries that went through structural adjustment in the 1980s and 1990s.

In many countries, the informal sector has been seen as a reaction to the labor unions in two ways: employers do not want their employees to be unionized and, on the other hand, unions, principally in developing countries, do not have enough credibility. Hence, there are few barriers to the informalization of the workforce.

In developed countries, immigrants, especially illegal ones, are more involved in informal activities. As they do not have permission to work, they prefer to avoid state regulations by working in small and unregulated businesses.

In many developing countries, the informal sector is a reaction to the lack of confidence in state institutions. In these terms, the state is seen as weak and untrustworthy, so the increase of the informal sector is only a part of the lack of confidence in state rules, institutions, or legisla tions. Corruption creates the problem that only a few actors have access to the formal market. The creation of special rules (legal or not) that are on the side of these groups encourages the informal sector in two ways. On the one hand, the protected groups shape the market according to their interests, and, on the other hand, vulnerable groups unleash their energies to avoid these rules, even in the informal sector. In fact, state regu lations support the existence of the informal sector. Work in the informal sector can be an option to circumvent the rules imposed by some unions, such as forbidding certain processes and forms of labor. Also, the implicit high cost of unionization promotes the move of enterprises and workers toward the informal sector.

Social themes

The informal sector is shaped not only by economic and institutional processes, but also by social ones. Since the 1980s, many authors have pointed out the importance of social networks, gender, and local cultures in the growth and relevance of the sector.

Generally, it has been assumed that women are more likely to be involved in the informal sector than men. The informal sector allows them to work fewer hours and combine their income generating activities with domestic household tasks. Also, women usually take this kind of job as a response to a reduction in household income.

The informal sector is more likely to be accepted in some places than others. This issue has been seen as a reflection local and regional traditions. In this case, it is not just that the informal sector is viewed as normal in one place, but that there is a completely history of the labor markets and their social and institutional regulations.

Some scholars have found that social networks, especially those that are dense, are an important factor in promoting the informal sector. These networks are important in at least two senses. First, an easy way to get into the informal sector is to turn to a friend or acquaintance. The spread of the sector generally depends on the social relations between persons in a specific locality. Second, the informal sector requires a communal sense of social participation in order to avoid state regulations.

Many informal workers, in developed and developing countries, state that one of the principal factors that drives them into the sector is the desire for autonomy and flexibility in their work. They find the informal sector more flexible and autonomous than the formal one; hence they have more freedom to do their own business, to establish their schedule, and to combine their work with other activities.

Remarks and Conclusions

Both the informal economy and informal labor market are special subjects in the field of geography, not only as a specific topic but as an area of action. World statistics show the importance of the sector, in both developed and in developing (or transition) countries. Geography can play a very important role in the conceptualization and the determination of policies.

The conceptualization of informality is one of the biggest issues. Despite their divergences, as presented here, many approaches are relatively close to each other; therefore, the next step is to develop a global and comprehensible conceptualization for the understanding of a very complex phenomenon.

Recent evidence has shown that informality grows as a response to the deterioration of living conditions both in the developed and in the developing world and the transformation of global industrial processes. However, there is a problem still: the analysis of the policies. Geography in this field can contribute in many ways: from the study of patterns of localization of informal activities to the analysis of the policies that promote the growth of the informal sector.

See also: Capital and Space; Dependency; Development I; Development II; Economic Crises; Economies, Alternative; Economies, Borderland; Economy, Informal; Global Commodity Chains; Labor Flexibility; Labor Markets, Regional; Neoliberal Economic Strategies; Urbanization; World/Global Cities.

Further Reading

Aguilar, A. G. (1997). The urban labour market in Mexico: Global change, informality, and social polarization. *Urban Geography* 18, 106 134.

Daniels, P. W. (2002). Urban challenges: The formal and informal economies in mega cities. *Cities* 21, 501 511.

De Soto, H. (1989). *The other path. The Invisible Revolution in the Third World*. New York: Harper & Row.

Gerxhani, K. (2004). The informal sector in developed and less developed countries: A literature survey. *Public Choice* 120, 267 300.

ILO (2005). *World Employment Report, 2004 05. Employment, Productivity and Poverty Reduction*. Geneva: ILO.

Laguerre, M. S. (1994). *The Informal City*. New York: St. Martin's.

Mead, D. C. (1996). The informal sector elephant. *World Development* 24, 1611 1619.

OECD (2002). *Measuring the Non Observed Economy. A handbook*. Paris: OECD.

Portes, A. (1995). *En torno a la informalidad; ensayos sobre teoría y medición de la economía no regulada (On informality; essays on theory and measurement of the unregulated economy)*. Mexico City: FLACSO Miguel Angel Porrúa.

Portes, A. (2004). *La Economía Informal (The Informal Economy)*. Santiago de Chile: CEPAL.

Portes, A., Castells, M. and Benton, L. (1989). *The Informal Economy: Studies in Advanced and Less Developed Countries*. Baltimore: John Hopkins.

Santos, M. (1979). *The shared space. The Two Circuits of the Urban Economy in Underdeveloped Countries*. London: Methuen.

Scheider, F. and Enste, D. H. (2000). Shadow economy: Size, causes, and consequences. *Journal of Economic Literature* 38, 77 114.

Williams, C. and Windebank, J. (1998). *Informal Employment in the Advanced Economies*. London: Routledge.

Relevant Websites

http://www.ilo.org
International Labour Organization.

http://www.imf.org
International Monetary Fund.

http://www.unescap.org
United Nations Economic and Social Commission for Asia and the Pacific.

http://www.escwa.org.lb
United Nations Economic and Social Commission for Western Asia.

http://www.uneca.org
United Nations Economic Commission for Africa.
http://www.unece.org
United Nations Economic Commission for Europe.
http://www.eclac.org
United Nations Economic Commission for Latin American and the Caribbean.
http://www.worldbank.org
World Bank.

Informalization

D. Vaiou, National Technical University of Athens, Athens, Greece

The informal sector/economy/activities, after a long standing association with Third World development, has gained currency in the industrialized world as well, particularly since the end of the 1970s. It is to this First World part of the debate that this article will focus. Informalization is proposed here as a reference to trends and processes to do with the restructuring of methods of production and, mainly, of work processes, as well as with deep transformations of the Fordist workplace and cultures of production and work. It can be seen as an emphasis on the implications of the dramatic changes of post war development patterns – changes which modified the geography of production and work in 'fortress rich world' and beyond. Although these trends are worldwide, their intensity and particular content are place specific and linked to local histories of uneven development.

In the mass of publications and policy documents which appeared in the 1980s and 1990s a huge variety of heterogeneous activities and processes were included under the umbrella term 'informal', rendering it a catchall category whose usefulness has at times been questioned. However, its use persisted since it captures, even imperfectly, trends and processes, which would otherwise be overlooked and which play an important role in urban and regional development in many places. Moreover, a new take can be recently identified, associated with international migration flows.

Definitions and Dichotomies

'Informalization', as a term, aims to capture the dynamics of restructuring processes in the pursuit of a neoliberal agenda, rather than describe patterns of production and work characterized as informal. It is associated with a retreat of the state from many areas of public intervention, in the context of strategies which aim to promote market led growth in a globalizing world economy. An important component of such strategies has to do with the de regulation of the labor market and of particular workplaces, as well as of the economy at large, along with an ideological attack on notions of collective organizing and action, all of which are considered as highly problematic rigidities of a Fordist regime. Flexibility is promoted as a remedy for crisis and stagnation and becomes a buzzword glossed over with ideas of innovation, choice, and creativity.

In this context, informalization appears as a process of becoming informal. Becoming presupposes the existence of something formal: a post World War II model of production, based on big firms, developed welfare systems, regular employment relations with full social security rights, and a strongly unionized workforce – all of which constitute the norm. Although this norm was time and place specific, it became a global model, vis à vis which any other pattern of work and production organization ended up being considered informal, a deviation defined in negative terms.

The relevant literature includes not only a lot of controversy and confusion in the use of terms, but also different theoretical and disciplinary perspectives and goals – which, to a certain extent, are also reflected in the terms. Economy, sector, employment (or work), or activities may be characterized informal (sometimes alternating with black, irregular, underground, or shadow), while production patterns and forms of work included in the informal (as part of its definition) vary with the intentions and perspectives of particular writers. A widely accepted and general definition comes from the International Labor Organization (ILO) and, referring mainly to employment, it identifies it as gainful activity without secure contracts, worker benefits, and social protection; it includes both self employment in informal enterprises and paid work in informal jobs.

In this rather slippery context, the informal is defined in terms of a formal/informal dichotomy, which establishes a hierarchy: the first part of the dichotomy is dominant, while the second is secondary, less important, defined in terms of lack. By the same token, those associated with the second part, with the informal, and the places where it predominates are also defined as less important. However, activities/workers/places are informal only from the point of view of debates and interpretations which originate from, and of experiences particular to, the heartlands of Fordism. Such characterizations are rather less helpful in understanding different histories of capitalist development.

Literature originating from Northern Europe and North America initially focused on social security fraud or violation of welfare systems, for example, unemployed people on benefits carrying out work without permission by the social insurance authorities, as well as the unregulated activity of small firms, self employed people, and self provisioning. Informalization, however, emphasizes more on the growing importance, mainly in cities, of the so called creative industries, of individual entrepreneurship, self steering and self promoting strategies, of international networking, and new modes of work. Some of the previously considered villains are then elevated to innovative and creative agents of the new flexible arrangements. Few are still interested in the dark side of such transformations: (almost) unlimited working hours, periods of nonwork (unemployment?), insecurity, and quick devaluation of knowledge and skills.

By contrast, literature from many parts of Southern Europe, where the informal has been, historically, much more pervasive and embedded in local cultures, has emphasized more on the quality and conditions of work, social practices to do with coping and economic survival, tax evasion, and family involvement. Here, the limits between formal and informal are less clear, since the formal model had been only partially applied. In a con text where the productive structure is characterized by small firms and where welfare systems are weak, the di chotomies are neither self evident nor easy to support. One is rather faced with a continuum of activities ranging from purely formal to purely informal – with all inter mediate categories present. Informalization then refers both to transformations in the restricted areas of eco nomic activity and work which were formal and to the changing character and volume of informal activities. Here as well few continue to underline the dire con ditions of work and survival in the less creative and less glossy local versions of flexibility and innovation.

The Specificity of Place

Informalization does not take place in a vacuum, but in specific contexts of formal and place specific regulatory systems. It is therefore important to consider the his torical and spatial origins of what is called formal/in formal at any given time and place, as well as the range of informal activities and forms of work which form part of the specificity of particular places, linked as they are to their development histories. For example, traditional and modern forms of production create different demand for (informal) labor; place specific compositions of employ ment and unemployment lead particular groups of peo ple to informal activities for longer or shorter periods of time, as a first job or in the context of multiple em ployment, while patterns of migration may bring still others to informal activities; traditions of gender relations place women and men in different power positions in formal/informal mixes of activities and work – positions which are in turn reproduced through gender divisions of labor in families and in the labor market; the particu larities of systems of regulation and of their enforcement leave room to devise ways of avoiding legislation and regulation; and relative affluence of certain sections of the population and development of markets for the products contribute to further expansion of informal activities. Such processes and relations constitute the dynamic environments in which successive rounds of informalization may be identified, modifying what counts as formal arrangements at any given place and time. The three cases that are briefly discussed below, although they cannot be considered in any way ‘representative’ or exhaustive, highlight a variety of circumstances, which determine and modify the content of infor malization.

Homeworkers in Local and Global Subcontracting Chains in Valencia

Geographical literature on restructuring has pointed to the importance of subcontracting chains as a response to the problems of large factories and concentrations of workers, not only in backward or disappearing branches of production, but also in innovative industries and firms. Homeworkers, usually women, as the last ring in those chains, thus emerged as an emblematic figure of re structuring processes and informalization – a figure common, albeit hidden, in both the industrial core regions of the First World and in its peripheries. Immigrant women in the garment sweatshops of New York, local and later Chinese women homeworkers in Prato, computer assemblers in Athens, Bangladeshi women machinists or packers in the housing estates of Yorkshire, the ‘wives’ of shoemakers in the area of Valencia, in the Marche, or in central Athens, and many more, unveiled a wealth of human labor, which underlies the constitution of places whose productive structure is rather more complex than their standardized and selective representations.

The area of Valencia in the South of Spain has a long development history of successfully combining export agriculture and manufacturing activities. In its pluriactive households, most men and women have some work ex perience in farming, small scale manufacturing, factory, and service work through their working lives. Household livelihoods thus derive from flexible transformations and adaptations to changing circumstances. Global changes in shoe production since the end of the 1970s greatly af fected the area where there is a considerable concen tration of the sector. Local factories were taking orders from large US or North European customers, in a rela tionship of extreme dependence, which, in any case, is very common in subcontracting. In order to meet orders and tight schedules, local firms mobilized a workforce employed in the factory and a much larger one, involved in the process through putting out batches of work. Putting out was initiated by men factory workers, who took extra work to be done at home, by their wives. Through the wives’ extensive interpersonal networks of relatives, neighbors, and friends, work got distributed to a considerable number of women. Later, small workshops were also formed with one or two employees.

The extreme flexibility and adaptability of this system is well supported by an impressive volume of research in different countries and places. A key element of its effective operation is the informalization of working conditions: working hours adapted to the needs of orders, complete power of distributors and insecurity of workers, lack of contracts, social security payments, and other guarantees, and absolute flexibility in mobilizing and dismissing the necessary workforce, mainly women who seldom gain the status of an industrial worker. This informalized organization of production was made still more informal when competition from low cost countries hit the sector, in the area of Valencia as elsewhere in Europe. However, while registered firms and employment declined, overall shoe production continued. Closing firms started operating as commercial firms and rented their machinery out to workshops, where (undocumented) immigrants from Northern Africa found ways of earning an income. Migrants' ever lower wages, reduced safety, and no benefit payments contribute to yet another round of informalization and to the dynamics of local development.

Selling Kastorian Furs to the Russian *Nouveaux Riches*

International geographical literature since the early 1980s is rich with analyses of local success stories, identified in many countries and regions. Constellations of small and medium enterprises, revitalized artisanal production, revival of untraded interdependencies, to do with local traditions of organizing production, local productive specializations, relations of trust, and so on, have become components of a new, flexible, innovative model, often seen to be replacing inflexible and crisis ridden Fordism. Despite differences, a common denominator of most explanations of local/spatial development patterns remains the almost exclusive emphasis on the firm and its evolution – from large to small and medium, from rigid/hierarchical to flexible, and from paternalistic to learning. Missing from pictures of local success are the social and environmental costs, mirrored in environmental destruction, hyperexploitative working conditions, gender inequalities, law evading on many levels, and generally informal arrangements.

This dark side of success is more than evident in the town of Kastoria, in Northern Greece, a geographically rather isolated town of *c.* 30 000 people, which is a major world center of manufacturing fur clothes from cuttings (miniscule slivers of material). Fur production is the dominant dynamic branch of the local economy, based on historical specialization, which dates back to the seventeenth century. This internationally unique specialization led Kastoria to control, in the early 1980s, 10% of fur trade, in general, worldwide. A key feature of the productive system is its direct coincidence with the routines of everyday life of local people. Dense industrial relations among many small firms cut across local society, are based on family contacts and informal networks of trust, in which deals are verbal, time schedules are tight, raw material is very costly, and economic transactions are done in cash. A conspiracy of silence about the problems of the area is reproduced, along with intense conservative, petit bourgeois, and competitive attitudes, based on a continuous interweaving of enterprise, employment, and family. Spatial diffusion of small firms and the use of space as a means to consolidate and further develop such networks mark the organization of the city and its surroundings as dominant components, positive or negative, in the productive system and in processes of social reproduction.

Working in fur production is an unhealthy job, particularly in workshops and small firms, which try to shrink the cost of production by every possible means, including very small production space, bad lighting and ventilation, badly designed workbenches, and no safety measures for the use of toxic materials. Conditions of work at home may be somewhat better, but fur work upsets the entire household and brings home the tension and pollution of workshops. Fur workers themselves often promote putting out and engage their wives and daughters as homeworkers paid on piece rates. Women are discovered by the local productive system at times of boom and forgotten again at times of crisis. Patriarchal family structures make of them a captive and flexible labor force, on which fur manufacturing depends for the more labor intensive parts of the production process. Women in the area are thus captive in a highly profitable local specialization from which they have minimal benefits. At the same time, to work in fur production is by no means an excuse for their other 'duties' as mothers, housewives, and farm workers.

In the long history of fur processing in Kastoria, many fluctuations in the importance and export orientation of the branch can be identified. Recent difficulties are in many ways linked to ecological campaigns, often violent protests, against killing living animals – which contributed to lower demand for fur clothes. After a short crisis at the end of the 1990s, exports of luxury fur clothes rose again to meet the demand originating from the new market of Russian *nouveaux riches*. This new clientele brought along highly individualized ways of selling the products and yet a new round of informalization and local adaptation.

Creative Industries and the New Precariousness in Postunification Berlin

In many parts of the First World, restructuring of economic activities has brought about tremendous losses of jobs and the spread of poverty among large groups of population. Against this bleak prospect, often a dire crisis, several remedies have been proposed and tried out, including the development of knowledge intensive economic activities, innovation driven restructuring, efforts to attract creative classes, and systematic place marketing, in attempts to activate mechanisms to attract investment, residents, and visitors and initiate new rounds of local economic growth. Geographic debate on these processes has been quite heated and includes a range of arguments, from sophisticated theorizations to selective focus on, and interpretations of, case studies, which are used to support the argument of a change of regime – toward something that has been called 'soft capitalism'.

In postunification Berlin, much of the productive structure associated with traditional/Fordist industries collapsed, leading to job losses and high unemployment, particularly in the east. The reunified capital city saw in the 1990s a boom of real estate ventures, in which the local administration was also involved. An ensuing fiscal crisis led to cuts in a range of social services and in public support to a number of initiatives, including creative ones. At the same time, a major goal of local policy has been the promotion of a new image of the city, a breakaway of its divided past and the separate strategies followed by either of its parts for over 40 years. In this new image of a city of/for talents and tolerance, knowledge intensive industries (including media, design, culture, software, and life sciences) are central. The goal to become attractive as a place of living and working, for the creative classes associated with such industries has led, among other developments, to intensive gentrification of central neighborhoods in East Berlin, like Prenzlauerberg and its Silicon Alley, where clusters of new creative industries are located.

Behind the gloss of knowledge and creativity, researchers have recently started to identify the agents of the new ethos, ambience and lifestyle, and their working conditions. Large numbers of young people, mainly men, are involved in a multitude of micro businesses or in self employment, which is codified as independence, entrepreneurship, and flexibility. The situation in knowledge intensive industries often forces them to interact with competitors and risk losing their cutting edge in innovation. Orders are rarely regular and have to be met swiftly, by mobilizing networks of conjunctural or more permanent collaborators. Periods of excessive working hours, when leisure or personal time disappears, alternate with times of relative or absolute inactivity, thus making survival a rather precarious affair, despite artistic images and self images and allusions to trendy, easygoing lifestyles. The fragmented working lives of young aspiring men do not share the rigidities of the emblematic figure of the collective mass worker. Nor does their informalized, flexible, insecure situation verify a viable alternative to it, despite the rhetoric of choice, creativity, and success.

Winners and Losers

The cases discussed above highlight various processes associated with informalization in different sociospatial contexts. Particular development histories provide useful threads to understand the present positioning of places, as well as why certain activities emerge, thrive, persist, transform, or decline. These histories also indicate how various processes of informalization get mobilized, both historically and actually, integrating or excluding different groups of a potential workforce in a whole range of economic activities, which account for local success/failure. Such processes seem to acquire a new intensity in the context of, and the debate around, globalization. The promised and celebrated impact of new communications technologies underlines accelerating transitions, speeding flows, eliminating frictions, and eroding of borders, identities, and differences – from which particular places and social groups are expected to draw significant benefits. Less prominent in the relevant debate is the underside of such deep material and discursive transformations, that is to say the effects on work and livelihoods of a retreat of the state from many areas of public intervention, in particular from workplace regulation.

Women, migrants, minorities, and youth are highly instrumental to these processes, a kind of natural pool for flexibilization and local adaptation. For a younger generation of people, lifelong professional insecurity is the new norm. Continuous de and re skilling is necessary in order to remain 'employable' and contain the effects of precariousness in what have become highly fragmented and insecure working lives. As far as women are concerned, welfare cuts in the name of efficiency, competition, and rentability have not only removed the protective cushion of provisions, but have also increased the bulk of care work to be done within households. Hence, women, particularly from poorer social groups, who undertake such responsibilities, end up having fewer and more insecure job opportunities and a heavier bulk of unpaid work. A further round of informalization is associated with the post 1989 flows of migrants, many of whom are undocumented and have no choice but to accept poor and insecure working conditions, thereby filling up a labor hierarchy from the bottom. Among migrants, women have become part of the restructuring of reproductive and care work and contributed to revitalize even the job of the housemaid, for many decades vanished at least in Europe. Thus, more affluent local households are in a position to transfer to women migrants, at a low and widely affordable cost, much of the care work which the restructuring of welfare has thrusted upon to them. Migrant men, on the other hand, are ab sorbed in construction, manufacturing, and agriculture, where low pay and precariousness match the demands of cost reduction and survival of many firms.

Informalization, however, has moved beyond young entrants into the labor market, 'invisible' workers, or unpaid care; it has undermined previously regulated areas of employment and privileged groups of workers and affects highly skilled and knowledge intensive labor. The split up of permanent full time jobs and wage cuts, social allowances and social security, subcontracting and homeworking, and the proliferation of mini jobs and precarious activities – to mention only few – create many losers within landscapes of success and winning and within a discourse that celebrates the advantages of new forms of making a living. 'Flexicurity', a new buzzword in European Union (EU) documents, is a case in point here of a discourse that obfuscates informalization processes, which penetrates all areas of production and work but remains generally 'hidden' or unacknowledged.

In order to fully grasp the dimensions of local success, a move beyond the firm and the entrepreneurial spirit of soft capitalism is necessary, both theoretically and em pirically. In order for the Fordist production and work organization to be destroyed, it has taken huge global and local economic restructurings. It has also taken a sweeping assault against important features of the collective mass worker. Routinized drudgery, standardization, rigidity, and de skilling have come under systematic criticism and put against choice, creativity, and flexibility – all of which constitute the emerging new 'norm' of gainful employ ment. Seldom does one see references any more to the security and social guarantees, which were part and parcel of the conditions of work of the collective mass worker and are now defined as comparative disadvantage and as distorting competition. In this context, it is crucial to shift the emphasis to the dark side of success: to the hard and deteriorating conditions of work and living, emanating from informalization and promoted as the sole alternative for local success.

See also: Fordism; Fordism, Post-Fordism and Flexible Specialization; Globalization, Economic; Informal Sector.

Further Reading

Brenner, N. and Theodore, N. (eds.) (2002). *Spaces of Neoliberalism: Urban Restructuring in North America and Western Europe*. Oxford: Blackwell.

Garcia Ramón, M. D. and Baylina Ferré, M. (eds.) (2000). *El Nuevo Papel de las Mujeres en el Desarollo Rural*. Barcelona: Oikos tau.

Kratke, S. (2002). *Medienstadt. Urbane Cluster und Globale Zentren den Kulturproduktion*. Opladen: Leske und Budrich.

Lange, B. (2007). *Die Räume der Kreativszenen. Culturpreneurs und ihre Orte in Berlin*. Bielefeld: Transcriptverlag.

Lipietz, A. (1996). *La Société en Sablier. Le Partage du Travail Contre la Dechirure Sociale*. Paris: La Découverte.

Mingione, E. and Pugliese, E. (2002). *Il Lavoro*. Roma: Carocci.

Paci, M. (2005). *Nuovi Lavori, Nuovo Welfare*. Bologna: il Mulino.

Portes, A., Castells, M. and Benton, L. A. (eds.) (1989). *The Informal Economy*. Baltimore, MD: The Johns Hopkins University Press.

Roberts, B. (ed.) (1994). *International Journal of Urban and Regional Research*, special issue on 'Informal economy and family strategies', 18(1), 1 142.

Vaiou, D. and Hadjimichalis, C. (1997/2003). *With the Sewing Machine in the Kitchen and the Poles in the Fields. Cities, Regions and Informal Work*. Athens: Exandas.

Williams, C. and Windebank, J. (1998). *Informal Employment in the Advanced Economies*. London: Routledge.

Ybarra, J. A., San Miguel, B., Hurtado, J. and Santa Maria, M. J. (2004). *El Calzado en el Vianalopó, Entre la Continuidad y la Ruptura*. Alicante: IDELSA.

Relevant Websites

http://www.info.worldbank.org
Bank Information Center: World Bank (IBRD and IDA).

http://www.uin.org.uk
Union Ideas Network.

http://www.eurosur.org/wide
Women Action Forum for Networking.

Information Graphics

A. L. Griffin, University of New South Wales – ADFA, Canberra, ACT, Australia

Glossary

Brushing An interactive method for highlighting a subset of values in two or more dynamic information graphics.

Cognition The mental processes involved in thinking, such as memory, attention, reasoning, and judgment.

Cognitive Load The level of effort needed for a thinking and reasoning task.

Focusing An interactive method for adjusting a dynamic display to filter out observations (e.g., outliers, or time points in a time series) for improving the clarity of a pattern found in the data.

Hypothesis Testing A set of formal methods in statistical analysis for disproving a hypothesis.

Interaction Change in a graphical display initiated by a user.

Linking An interactive method for coordinating display changes in multiple information graphics that are viewed together.

Preattentive Processing The rapid detection of a select set of visual properties that occurs without the viewer focusing his or her conscious attention on a visual display.

Scientific Visualization Computer-based visualization of real-world, physically-based objects that cannot normally be seen by the naked eye.

Semiotics The science of signs and symbols, including how meaning from signs and symbols is constructed and understood.

Introduction

Information graphics are visual representations of data and information. The term 'information graphics' refers to a diverse collection of graphics, ranging from diagrammatic depictions of physical objects and processes to abstract representations of data that do not have an intrinsic physical interpretation. They include representations such as pictorial instructions, diagrams, charts, graphs, and maps, may be either static or dynamic, and are most commonly used for representing quantitative information. **Figure 1** (itself an information graphic) shows a recent attempt at classifying different types of visualization methods by the purpose for which they are best suited. Within geography, researchers commonly use a combination of information graphics that can help them to understand different aspects of geographical problems. For example, maps and animated maps can reveal spatial (and spatiotemporal) patterns while statistical graphics such as scatterplots or star plots illuminate relationships between different data variables.

The Purpose of Information Graphics: Graphics for Thinking versus Graphics for Communication

Just as graphics can take a number of different forms, they can be constructed for a variety of purposes and for different individuals. One functional model of visual methods that has been used to think about different forms of visual representation compares representations using three characteristics: context of use, knowledge about patterns in the data, and level of reader interaction with the graphic (**Figure 2**). The context of use may range from public to private, where public uses of information graphics support visual communication and private uses are commonly used to support visual thinking. The level of knowledge that the graphic creator has about the dataset is related to the purpose for which the graphic is developed, ranging from exploring unknown patterns and confirming hypotheses to synthesizing or presenting well understood relationships or explanations. The level of reader interaction with the data refers to the extent to which the graphics reader can alter or manipulate any of the characteristics of the graphic (e.g., colors, symbols, data variables represented, axes or scales, and viewpoints).

Visual communication occurs when both the graphic designer and the graphic reader have some common, shared understanding of what the graphical marks that make up a graphic mean. The science that provides a conceptual framework for thinking about this shared understanding is called semiotics. The basic semiotic model is composed of three elements: the sign vehicle, the interpretant, and the referent (**Figure 3**). The sign vehicle is the carrier of meaning that acts as a translator between the graphic designer and the reader. The interpretant is the mental image or conception that the graphic reader has of what is depicted in the graphic, and the referent is the real world object or entity that is being represented in the graphic. Miscommunication can result when the graphic designer's interpretant does not match the interpretant that the graphic reader develops from looking at the graphic.

Visual thinking can be thought of as a cognitive process that produces insight by helping the graphics

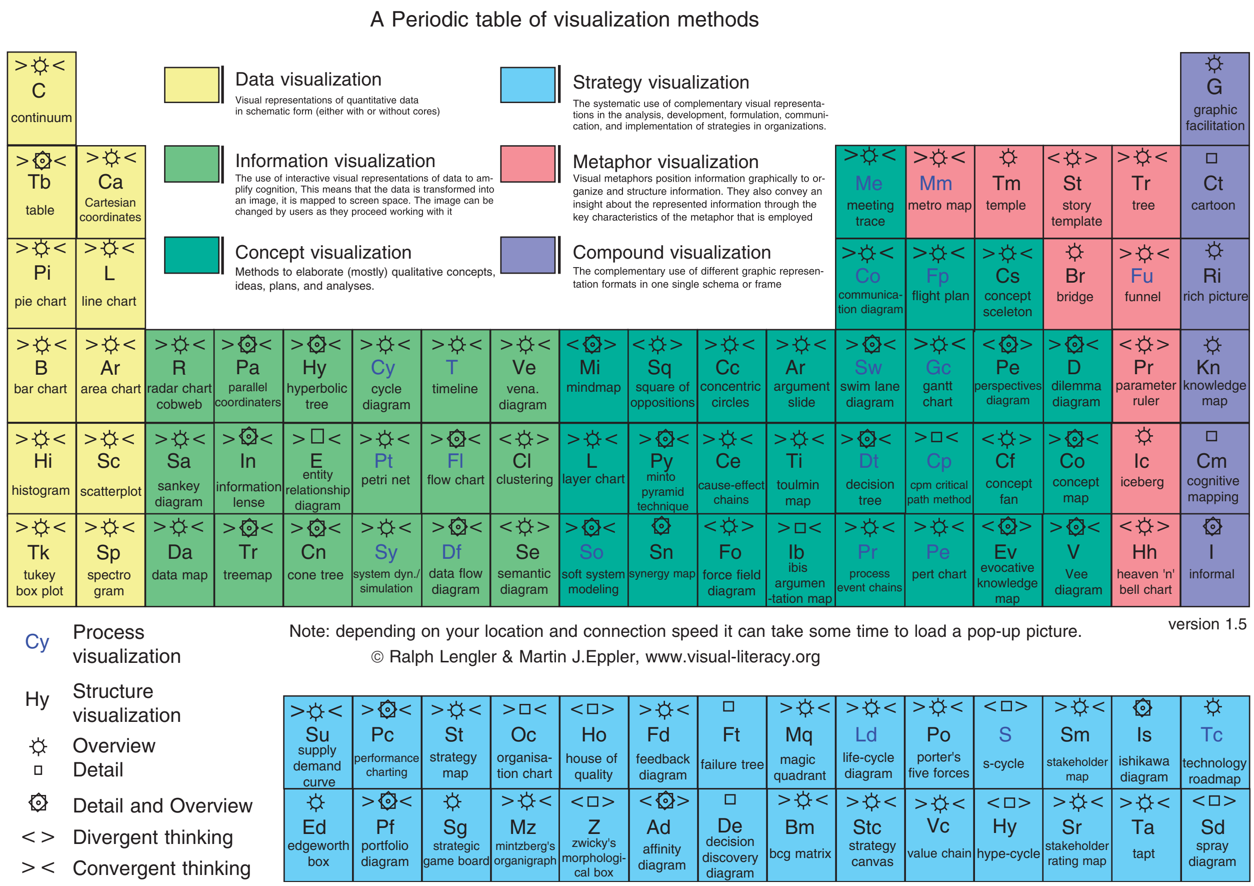

Figure 1 A periodic table of visualization methods that suggests which different information graphical forms are useful for particular goals, ranging from analyzing and communicating the structure of a dataset to capturing the key characteristics of processes. Reprinted by permission of http://www.visual-literacy.org.

reader see patterns, relationships, and anomalies in data. Because the producer and the user of information graphics in the context of visual thinking are often the same person, information graphics that are designed to support visual thinking often take a different form to those designed for visual communication (**Figure 4**). Because there is no need for the graphics reader to interpret the graphical marks on the page (as s/he has designed them), it is more important for the information graphic to show a variety of perspectives on the data of interest. This is often achieved by combining different types of visual representations within a single information graphic. Three main areas of research have been influential in developing knowledge on how to design graphics to support visual thinking: exploratory data analysis, information visualization, and geovisualization.

The first, exploratory data analysis, has its roots in statistics. In contrast to much statistical analysis, however, exploratory data analysis does not make *a priori* assumptions about relationships in the data, as in hypothesis testing. As such, exploratory data analysis involves attempts to identify systematic relationships between variables. Often, this approach involves using many variables and several pattern searching techniques. A basic premise of this approach is that it is important for the graphic designer to find the correct means of graphically presenting the data such that the graphic reader can become part of the pattern recognition process (rather than relying on statistical hypothesis testing). Over time, exploratory data analysis has evolved to include an increased focus on interactivity, incorporating techniques such as brushing, linking, and focusing.

The second, information visualization, is concerned with creating visual aids to support and enhance memory, thought, and reasoning (i.e., cognitive processes). In particular, information visualization focuses on creating computer supported, interactive visual representations of (primarily abstract) data to amplify cognition with the goal of supporting discovery, decision making, and explanation. It places an emphasis on interaction with graphics, because manipulating these visual aids involves a weaving together of internal and external representations of information (i.e., the interpretant and sign vehicle, respectively from the semiotic model in **Figure 3**) that enables the viewer to reason and address 'what if' scenarios in a much more efficient manner than s/he

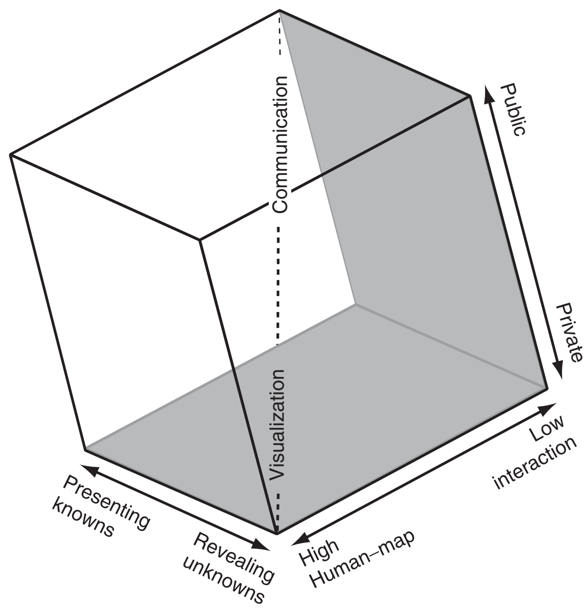

Figure 2 MacEachren's functional model of visual methods. Although this model was originally developed for a subset of information graphics (geovisualizations), it can be applied to thinking about information graphics more generally. Note that within this model, a single information graphic can occupy multiple spaces, depending upon the level of interaction between the reader and the graphic. This is especially applicable to the case of dynamic information graphics, which can be modified through user interaction. Reproduced from MacEachren, A. M. (1994). Visualization in modern cartography: Setting the agenda. In MacEachren, A. M. & Taylor, D. R. F. (eds.) *Visualization in Modern Cartography,* p. 6. Oxford: Pergamon.

would if they had to consider and reason through the constraints of each scenario separately.

The third, geovisualization, has its roots in cartog raphy, scientific visualization, and exploratory data an alysis. This area of work has also viewed visualization as a cognitive process whose goal is to produce scientific in sights and create new knowledge. A major focus of geo visualization has been not only on developing new ways of visually representing information through maps and other types of information graphics, but also on under standing differences between different types of viewers of graphics and how they are able to use the graphics to create new knowledge. Geovisualization has drawn upon perspectives from cognitive psychology in examining how factors such as experience, prior knowledge, and training affect perception, comprehension, and problem solving with maps and map based analysis environments.

A Reference Model for Creating Information Graphics

Information graphics, no matter whether they are con structed to support visual communication or visual thinking, have both a designer and a reader. In some instances, such as when a scientist makes a graphic to try to understand his or her data, the designer and the reader are the same individual. In other cases, such as when a journalist produces a graphic that supports an article s/he has written, the designer and the reader are different individuals. The efficacy and efficiency with which in formation graphics are able to prompt understanding among their readers depends upon decisions made dur ing the design of the graphics and how these decisions interact with the capabilities of the human visual system.

There are three key types of decisions that a designer makes when s/he designs information graphics:

1. how to transform data from their raw state into ordered relations (i.e., data tables);
2. how to create a visual structure for expressing the relations contained in the data tables; and
3. how to provide a particular view of that structure by modifying the characteristics of the visual structure (**Figure 5**).

Organizing the Data

Data, however they are collected, rarely come in a form that allows direct and easy visual representation. For example, data may come as a string of values that needs to be classified and organized into a table through a

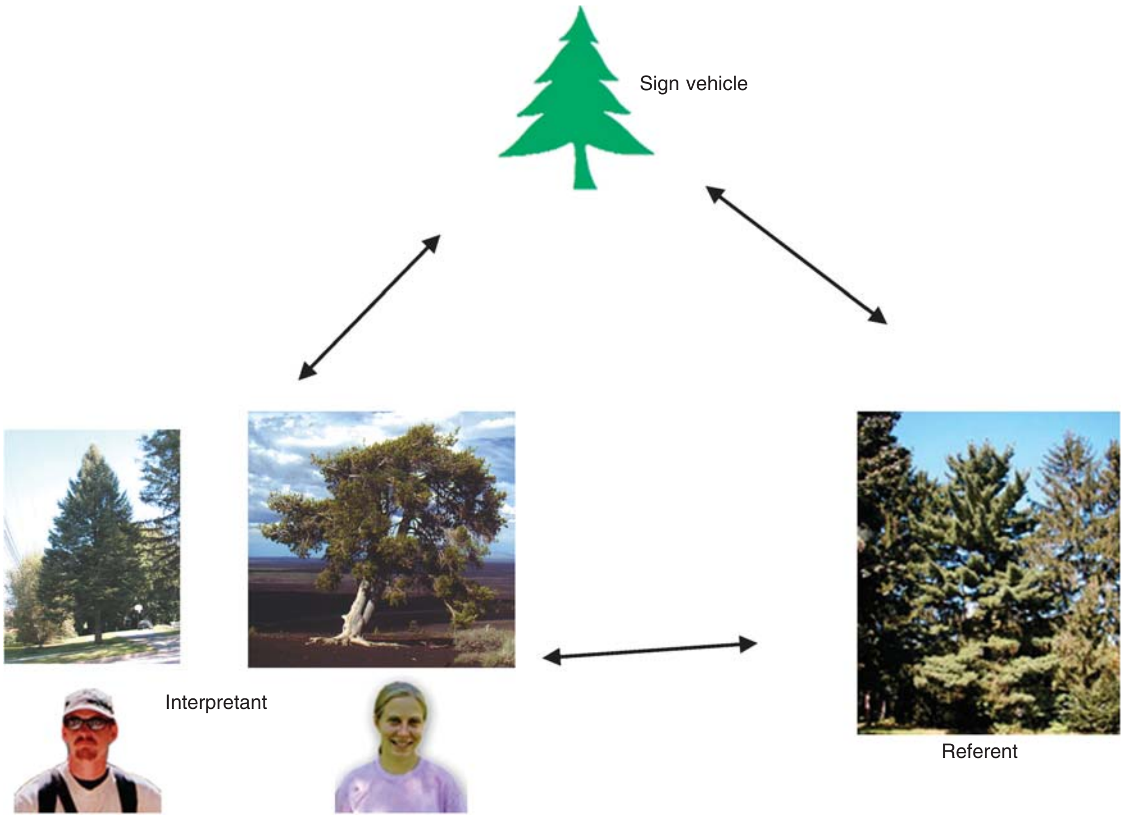

Figure 3 One example of the three components of the semiotic model in an information graphic. In this example, the symbol (upper portion of the diagram) acts as the sign-vehicle, or the carrier of meaning; the information graphic reader's conception or mental image of a tree acts as the interpretant or concept (lower left); and the actual object in the real world (lower right) acts as the referent, or object of reference. Notice that different people may have different mental images of a concept.

cross tabulation. Alternatively, data may consist of raw values that require a statistical operation before they can be used, such as calculating a mean or a median.

The choice of how to first structure the data and then visually structure the information graphic can be critical to the success of the graphic in aiding the reader to make sense of the data. In a well known example of mis specifying which characteristics of the data are important for the problem, engineers from NASA only consulted information graphics that showed the failure of O rings on space shuttle flights structured by date, rather than by temperature, which was the ultimate cause of the O ring failure in the fatal explosion of the shuttle (**Figure 6**).

Choosing a Visual Structure

Once the data are organized in an appropriate configur ation, the designer can decide upon a visual structure for representing the data. Although there are many ways in which the designer can match visual characteristics to data, some choices are better than others. There are two issues to consider when choosing from the many potential options: whether the nature of the visual characteristics matches the nature of the data characteristics, and the limits and parameters of the human visual system.

Characteristics of the data and graphical characteristics

Data variables can be classified as nominal (i.e., qualitative), ordinal (i.e., ranked but not quantified), or quantitative variables. Visual variables (i.e., graphical characteristics) have also been classified according to their suitability for use with nominal, ordinal, or quantitative data. **Figure 7** depicts a set of nine fundamental visual variables: color hue, shape, arrangement, orientation, position in space, color value, color saturation, size, and spacing. An in appropriate matching between a visual variable and the data variable can actually make relationships and patterns in the data harder to see rather than easier (as compared to viewing the values in a table; **Figure 8**).

The human visual system can both be exploited and act as a constraint when designing information graphics. It is therefore important for information graphics de signers to have a basic understanding of how humans process visual information. This processing can be div ided into two major steps: perception of visual stimuli

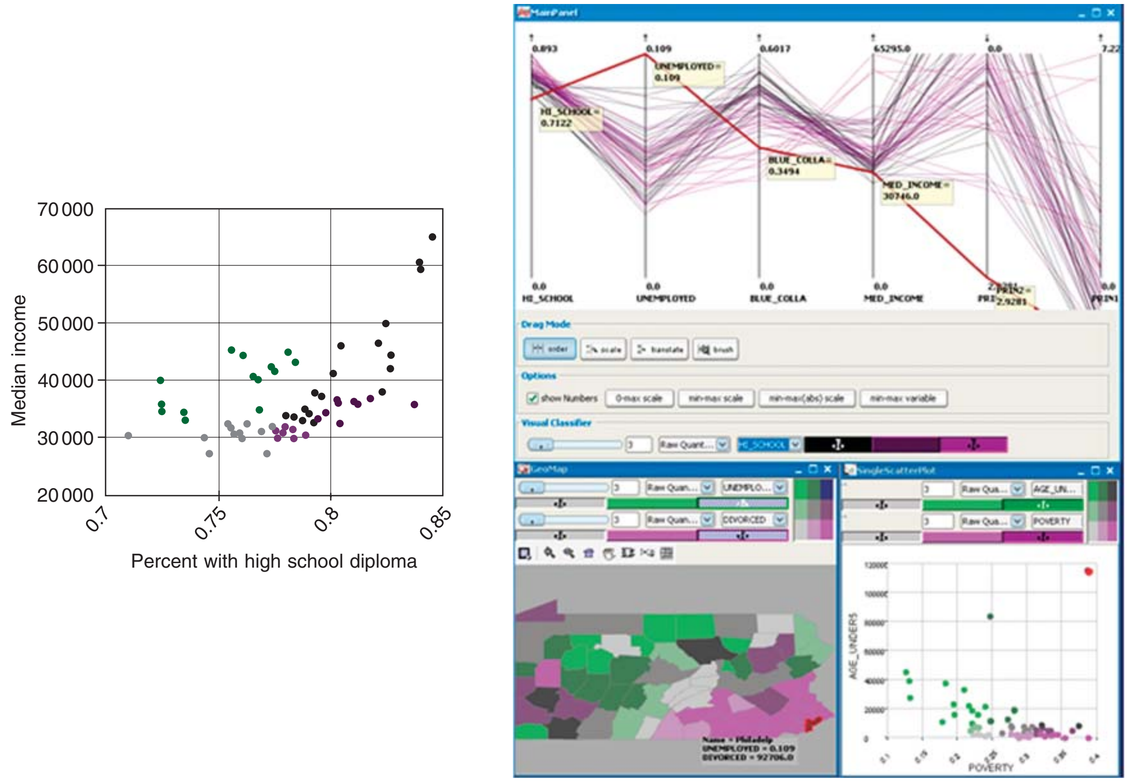

Figure 4 Comparison of an information graphic designed for communication (left) with an example of an information graphic designed for thinking about a problem (right). Notice that the graphic designed for thinking actually uses a combination of different types of graphics to provide complementary viewpoints on the problem at hand, and that it is possible to link these diagrams to explore observations across the different representations. This particular example identifies an outlier in the scatterplot and then uses the map to identify where that outlier is (Philadelphia).

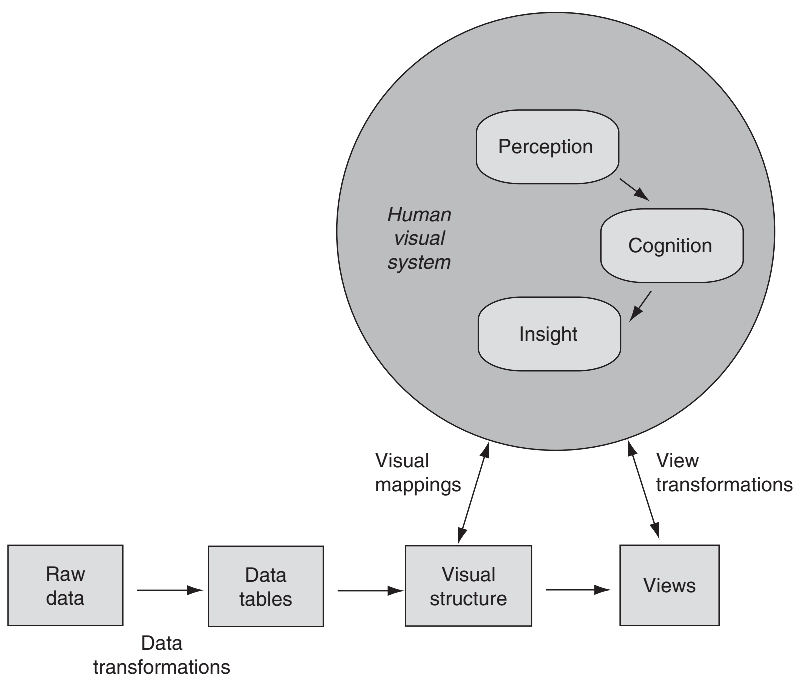

Figure 5 A simplified model of the components of the design process for creating information graphics. There are two key types of features represented in this diagram: physical artifacts (i.e., materials used in different steps of the design process) and types of decisions. From Card, S. K., MacKinlay, J. D. and Shneiderman, B. (1999). *Readings in Information Visualization. Using Vision to Think*, p. 17. San Francisco, CA: Morgan Kaufmann Publishers.

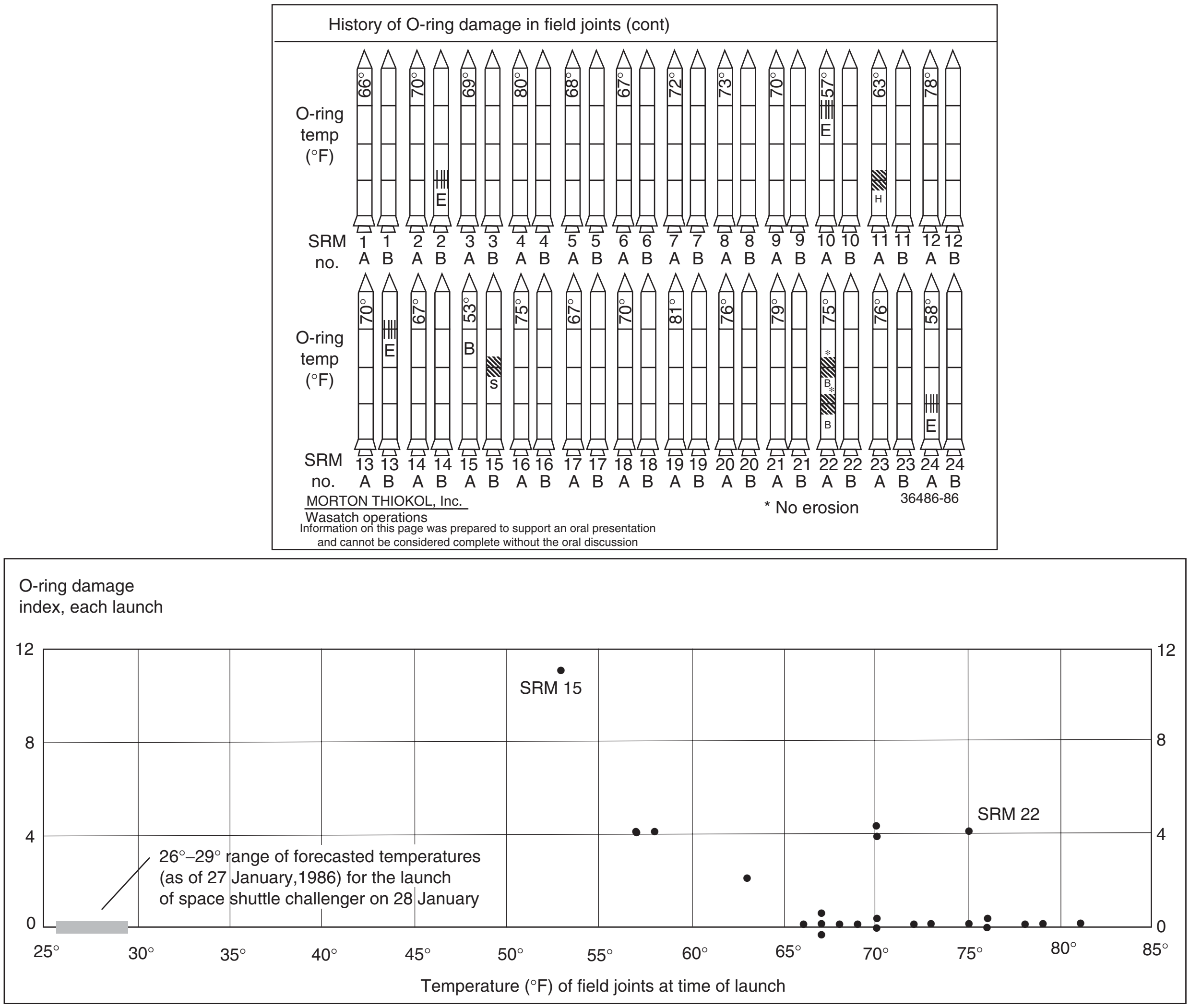

Figure 6 A comparison of an information graphic whose data were poorly structured for the task at hand (top) with a restructured information graphic that more clearly shows that launches with no O-ring damage only occurred at much higher temperatures than those experienced the day of the Columbia shuttle disaster (bottom). Adapted from the Report of the Presidential Commission on the Space Shuttle Challenger Accident. (1986). Washington, DC: Government Printing Office and Tufte, E. R. (1997). *Visual Explanations. Images and Quantities*, *Evidence and Narrative*. p. 45. Cheshire, CT: Graphics Press. Reprinted with permission from Graphics Press.

(i.e., sensing visual information), and cognitive processes that interact with these stimuli and that allow the graphics reader to interpret what s/he sees. It is possible to think of perception as being the 'looking' component of vision, while cognition is the 'seeing'.

Visual perception

All vision requires three basic components: (1) light that is emitted from some source, typically either the Sun, a light bulb, or a computer monitor; (2) the eye, which contains millions of light sensitive cells; and (3) the brain, which interprets the light that enters the eye and makes sense of the sensed patterns of light in a part of the brain called the visual cortex. Although the human visual perceptual system is quite complex and is composed of many subsystems, the process of sensing visual information can be generalized into two major stages: automatic processing and controlled processing. The first stage is characterized by rapid parallel processing of information, sensing of certain kinds of information such as orientation, color, texture, movement and identifying the edges of objects, and storing information in memory in a transitory fashion. The second stage, on the other hand, is characterized by slow serial processing, the use of both short term and long term memory, and direction of attention.

A look at the anatomy of the human eye can explain the roots of the two stages of visual perception (**Figure 9**). The retina is the portion of the eye that contains the photoreceptive cells that respond to light that enters the eye. These cells are not distributed evenly

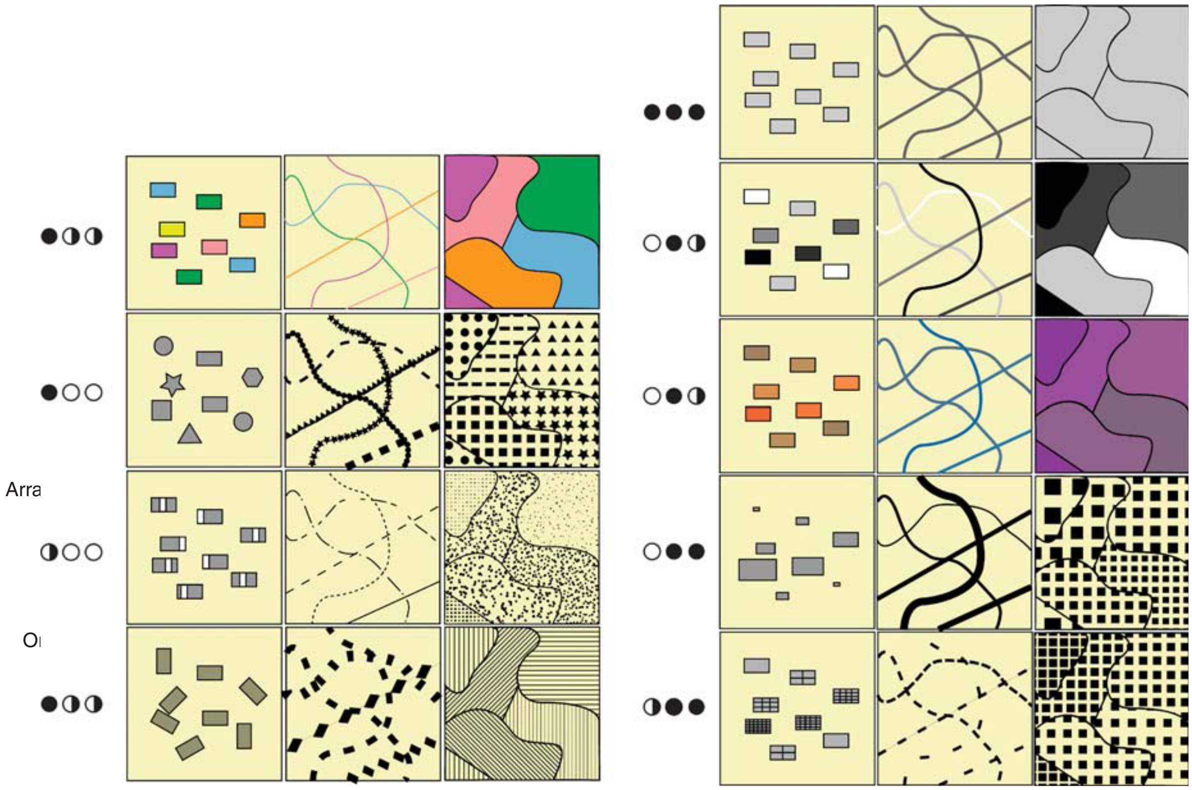

Figure 7 Visual variables and their suitability for representing different types of data variables (nominal, ordinal, and quantitative). This figure shows examples of each visual variable for point, line, and area symbols. Filled symbols indicate that the visual variable works well for representing that type of data. A half-filled circle indicates it may be marginally effective, while an open circle indicates a poor match between the visual variable and the data variable characteristics.

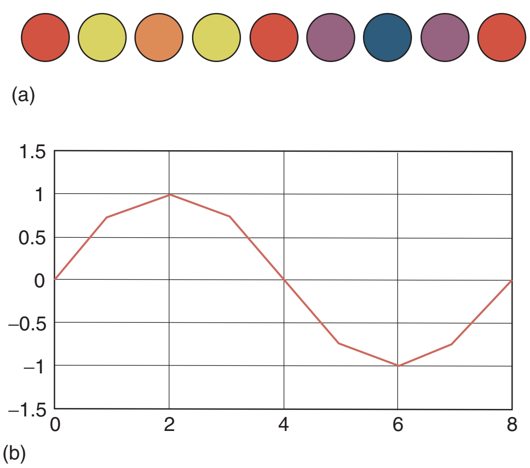

Figure 8 The color hues used in the information graphic at the top (a) are an inappropriate match between a visual variable and the characteristics of the data variable. The sine wave that is being represented in the figure is a quantitative variable, but color hue is best used for representing nominal (i.e., qualitative) information. The information graphic at the bottom (b) uses position in space as its visual variable, and is much more effective in its representation of the character of the sine wave. From Card, S. K., MacKinlay, J. D. and Shneiderman, B. (1999). *Readings in Information Visualization. Using Vision to Think.* San Francisco, CA: Morgan Kaufmann Publishers.

throughout the retina. There is a higher density of cone cells, which detect color, in the center of the visual field, known as the fovea, while there is a higher density of rod cells, which respond to black and white light, in the surrounding area of the retina. The foveal area of the retina is responsible for focal vision, while the nonfoveal area is responsible for peripheral vision. Peripheral vision is particularly good at detecting movement and other changes in the visual field, such as areas of contrast in size, intensity of light, or the color of light. Much of this information processing is preattentive, automatic, and attention shifting, as when movement across a display catches the eye of the viewer, who unconsciously tracks the moving object(s).

While this preattentive surveillance of the eye's field of view is continually updating, a graphic reader can also shift his or her focus (i.e., attention) to specific fea tures in the display, such as text or more complicated symbols. This focused attention, effected by shifting the location of the foveola (the central part of the fovea) within the eye, allows the construction of a more detailed image due to the higher spatial resolution afforded by the higher density of cone cells in the fovea. In contrast to the image produced by the automatic preattentive

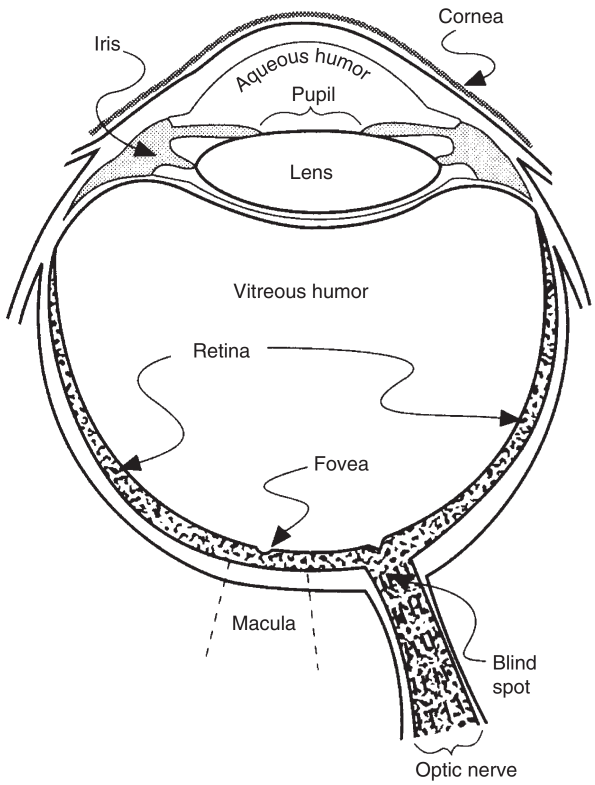

Figure 9 The anatomical structure of the human eye. Reproduced from MacEachren, A. M. (1995). *How Maps Work. Representation, Visualization and Design.* New York: Guilford Press, with permission from Guilford Press.

processing of peripheral vision, the controlled processing of focused attention is slow and serial rather than parallel.

Although perceptual processes have many possible implications for the design of information graphics, two notable issues include the relationship between visual acuity and differentiating between symbols and the establishment of figures and the background within a graphic. Visual acuity, the size of the smallest discriminable feature in a display, like the distribution of photoreceptive cells in the eye, varies spatially across the retina. The eye's highest visual acuity occurs near the fovea, where the density of cells is the highest. The implication of this for the design of information graphics is that symbols with small differences between each other will need to be perceived using the comparatively slower controlled processing of attention as small differences between symbols will not be discriminable through peripheral vision. An example of an information graphic design that could accommodate this feature of perceptual processing might be that differences in symbols need to be exaggerated in the overall display, while on closer inspection (e.g., after zooming in), size differences could be proportionally related (in true scale) to the data values being depicted in the graphic. In the context of the second information design issue related to perception, if position in space is used as a visual variable in an information graphic, a visualization technique that uses an algorithm that causes related objects to cluster tightly in space will lead to the graphic reader seeing that set of objects as a group (and then potentially as a figure), while one that does not cause related objects to cluster tightly may make groups more difficult to identify, which may be undesirable if one of the goals of the information graphic is to facilitate the identification of commonalities between objects.

Visual cognition

Once an image has been constructed in the brain's visual cortex, it becomes understanding when the visual image interacts with the graphic reader's existing knowledge. In other words, the image itself derives its meaning from its relation to other knowledge. In order for a new piece of knowledge to be generated, the graphic reader must access his or her long term memory and compare knowledge that resides there with the visual image, which is itself stored in short term memory (**Figure 10**). This process of comparison between what is perceived and what is already known is often iterative, with the graphic reader further directing his or her visual attention to different areas of the information graphic or to additional information graphic based upon the results of his or her comparison of the visual image with knowledge stored in long term memory.

It is thought that information graphics can aid in cognition because they are external to the brain, and as such are not limited to the capacity constraints of internal cognition. One such well known constraint is that short term memory can only hold 7 ± 2 items at any point in time. Visual representations have a long history of serving as external aids to cognition, such as in writing out the sequence of operations needed to multiply two multidigit numbers together, or Mercator projection maps serving as a store of memory for calculating the shortest route between two locations.

Information graphics have been suggested to improve cognition through a number of pathways. First, because they store information, they serve to increase the quantity of information available for cognitive operations, decreasing the cognitive load because more information can be seen rather than remembered. Second, information graphics are structured (rather than random) representations of information. This structure often means that related information can be grouped together in the graphic, reducing the time required to search for new information that is needed for cognitive activities such as problem solving and making comparisons. Third, this same structure can also allow graphics readers to recognize patterns, which is faster than recalling specific information from memory, and it may enhance pattern

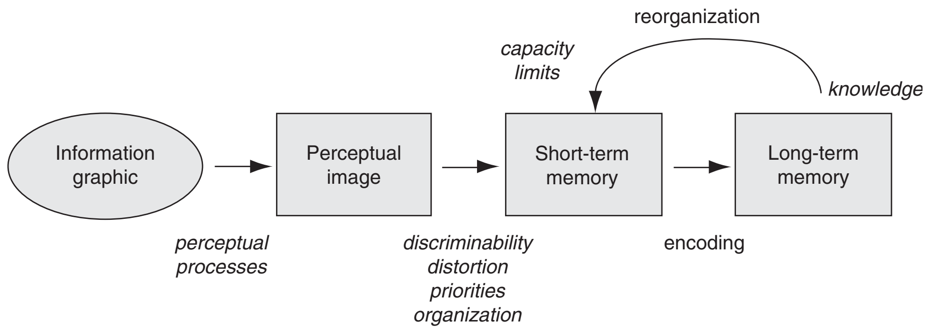

Figure 10 Kosslyn's visual information-processing model. The boxes in this diagram represent three types of visual processing, while the italicized labels describe limits to types of memory or transformations between different types of memory. The plaintext labels indicate processes undertaken in different types of memory. For example, encoding is the process of moving information from short-term to long-term memory, while reorganization is the reshaping or reinterpretation of information stored in short-term memory based upon knowledge that is stored in long-term memory. Reproduced by permission of Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology* 3, 190, © John Wiley & Sons Limited.

identification if the data are organized so that they en hance structural relationships. Finally, information graphics can allow readers to rely on their perceptual capabilities to make inferences more easily. For example, it is easier to infer that there is a north–south gradient in solar radiation by examining a map of solar radiation by country than it would be to examine a table that lists the radiation level for each country. In the graphic, the reader can see a progression in values from north to south that changes more or less in step with latitude. By con trast, when examining the table, the reader would have to decide how to mentally structure his or her comparison of the values, particularly if the table lists the countries alphabetically. This mental structuring, then, requires the reader to put more effort into understanding the rela tionship between values for different locations.

Although using information graphics as external aids to cognition can often augment the reader's cognitive pro cessing abilities, if not carefully designed to account for how cognitive processes operate, information graphics can actually decrease rather than increase the reader's prod uctivity. For example, although animations might seem like a natural choice for representing processes, as they can be used to directly depict change in a system over time, if they are not interactive (i.e., they do not let the graphic reader pause or replay portions of the animation), an animation may actually hinder the graphic reader's development of an understanding of that process if it contains so much in formation that the reader cannot remember all of the critical moments that describe the process.

Deciding upon a Viewpoint

Providing a particular view into an information graphic can be as simple as changing the angle from which a three dimensional scatterplot is displayed in order to minimize the concealment of one group of data obser vations by another group. Other more complicated changes to the viewpoint include taking cross sections of volumes, providing multiple windows that can be viewed at different scales (e.g., detail and overview panes), and applying distortions that highlight relationships on a network, similar to the concept of detail and overview panes, but applied to network distances rather than spatial distances. A final method for providing views into a graphic involves incorporating interactivity into infor mation graphics. With interactivity, readers can dynam ically change both the visual structure that are used to represent information, change the viewpoint presented in the graphic, and link multiple representations of infor mation to simultaneously provide complementary view points on the data being investigated.

See also: Cartographic Animation; Color, Mapping.

Further Reading

Bertin, J. (1981). *Graphics and Graphic Information Processing*, Berg, W. J. and Scott, P. (trans.). Berlin: Walter de Gruyter.
Bertin, J. (1983). *The Semiology of Graphics: Diagrams, Networks, Maps*. Madison, WI: University of Wisconsin Press (French edition, 1967).
Card, S. K., MacKinlay, J. D. and Shneiderman, B. (1999). *Readings in Information Visualization. Using Vision to Think*. San Francisco, CA: Morgan Kaufmann Publishers.
Chen, C. (2006). *Information Visualization: Beyond the Horizon*. Berlin: Springer GmbH.
Harris, R. L. (2000). *Information Graphics: A Comprehensive Illustrated Reference*. New York: Oxford University Press.
Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology* 3, 190.
Kosslyn, S. M. (2006). *Graph Design for the Eye and Mind*. New York: Oxford University Press.
Larkin, J. H. and Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science* 11, 65 99.
MacEachren, A. M. (1994). Visualization in modern cartography: Setting the agenda. In MacEachren, A. M. & Taylor, D. R. F. (eds.) *Visualization in Modern Cartography*, 6p. Oxford: Pergamon.
Tufte, E. R. (1983). *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

Tufte, E. R. (1990). *Envisioning Information*. Cheshire, CT: Graphics Press.
Tufte, E. R. (1997). *Visual Explanations. Images and Quantities, Evidence and Narrative*. Cheshire, CT: Graphics Press.
Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading: Addison Wesley Publishing.
Ware, C. (2004). *Information Visualization. Perception for Design* (2nd edn.). San Francisco, CA: Morgan Kaufmann Publishers.

Relevant Websites

http://utpjournals.metapress.com
Cartographica: The International Journal for Geographic Information and Geovisualization. This is the website for an important journal in the field of geovisualization. It publishes articles on a range of aspects of representing, using, and communicating with geographic information.

http://infoviz.pnl.gov
Information Visualization at Pacific Northwest National Labs: this is the research website for one of the leading groups in information visualization. The site contains links to both research papers and demonstrations of technologies developed in the lab.

http://www.palgrave journals.com
Information Visualization: this is the website for the flagship journal in the discipline of information visualization. The journal publishes a wide array of papers that theorize information visualization, present techniques and methodologies, and evaluate the utility of information visualization techniques.

http://www.infovis.net
InfoVis.net: this website provides an excellent overview of resources that are available on information visualization, including a who's who guide to important researchers, a resources page, and an electronic magazine with short articles.

Information Technology

B. Parthasarathy, International Institute of Information Technology, Bangalore, India

Glossary

Custom Software Software tailor-made for customers to meet their needs efficiently.

High-Level Programming Languages Increasingly human-like languages to communicate with computers, in contrast to communicating with zeros and ones.

Microprocessor A system on a chip; or a silicon device integrating discrete circuit elements such as capacitors, resistors, and transistors, with logic and data-handling capabilities.

Open Source Software Software released in the public domain for anyone to freely use, modify, and redistribute.

Packaged Software A collection of information-processing functions for which there is a mass market.

Software Bottleneck The quality and productivity lag in the development of software vis-à-vis hardware production.

Structured Programming Management of the software development process by segmenting it to permit coordination and control over teams of programmers.

Introduction

Although information technology (IT) has come to be closely identified with microelectronics based computing since the mid twentieth century, it has existed ever since the means to represent information, such as language, graphics (including scripts), and number systems, were invented. Subsequently, the invention of the printing press to record information, as opposed to recording it on palm leaves, or telegraphy, to transmit information more quickly and to further distances than is possible using carrier pi geons, made information available to more people. While the invention of writing or telegraphy in earlier centuries means that microelectronics is just one more in a sequence of forces transforming IT, there are at least two reasons why this transformation is considered significant enough to be the basis of a third industrial revolution.

First, advances in microelectronics make it possible to obtain and deliver information from and to any corner of the globe, in real time, on a scale unparalleled in history. The historical relevance of such possibilities is height ened by concurrent economic and social transformations. The acceptance of economic globalization, for instance, has increased the demand for information about distant markets. Similarly, as the virtues of political democracy are emphasized, access to information is viewed as a fundamental right. Second, IT is changing and becoming increasingly more powerful, affordable, and versatile. As a result, it is no more than a broad and fuzzy term to describe a variety of activities, ranging from issues con nected with the technology, such as computer hardware and software, to the deployment of the technology for information processing in application domains as distinct as design, cartography, and medicine.

The origins of the influence of microelectronics can be traced to the invention of the transistor in December 1947 at the Bell Laboratories of American Telephone and Telegraph (AT&T). The transistor is a device to amplify current, using the properties of a class of elements known as semiconductors, prominently silicon. Transistors are referred to as solid state amplifiers, and the field of study devoted to them as solid state electronics, to distinguish them from the vacuum tubes that they displaced in electrical circuits. While the vacuum tube was a bulky device, the transistor was amenable to miniaturization. Since its invention, the dimensions of the transistor have fallen 10 000 fold and current transistor sizes are down to submicron scales. Thus the term microelectronics.

Mass manufacture of the transistor came in 1951. In 1957, Jack Kilby of Texas Instruments and Robert Noyce, then with Fairchild Semiconductor and who later co founded Intel in 1968, integrated discrete circuit elem ents such as transistors, resistors, capacitors, and the associated wiring on a single block, or chip, of silicon to develop the first integrated circuit (IC). By 1959, Noyce also developed the planar process to reliably manu facture ICs. Integration techniques improved rapidly and, in 1971, Intel embedded enough logic functions and data handling capabilities to produce the first microprocessor, or a monolithic system on a chip. The Intel 4004 microprocessor contained 2400 transistors on a 12 mm^2 area. Since then, a prediction made in 1965 by Gordon Moore, another co founder of Intel, that has come to be known as Moore's law, has defined the technological evolution of the microprocessor. Although the original prediction was that the number of transistors on a chip would double every year, and Moore later altered the forecast to every 2 years, the truism has since been used to refer to the doubling in the performance to cost ratio every 18 months. Thus, Intel's Itanium II processor, re leased in 2003, had 220 million transistors, and the cost per transistor fell from approximately 0.15 cents in 1980 to about 0.0001 cents in 2005.

While advances in microelectronics have physically shrunk, made affordable, and enhanced the functional potential of various information processing devices, their impact would be limited if they could not be instructed to undertake a range of tasks. It is these instructions, also referred to as software programs, or code, that lend versatility to microprocessors and make it possible to exploit their potential in various domains. In other words, although it has no materiality, it is software that ensures the revolutionary character of contemporary IT. This ability originates from the concept of stored program control (SPC), first enunciated by John von Neumann in the 1940s. By proposing the entry and storage of pro grams as numbers, SPC provided the basis for a common digital language using which information in any form can be processed and conveyed by microprocessor based devices. Consequently, as advances in microprocessors have made possible more versatile applications, the technological characteristics of hitherto distinct electrical devices, such as computers and telephone switches, now resemble one another, a phenomenon referred to as convergence.

The Software Bottleneck

Over the years, the technological convergence has been accompanied by a divergence. New design and manu facturing technologies have enabled the mass production of increasingly more reliable and powerful chips and other hardware, in highly automated, capital intensive operations. In contrast, software production is a relatively craft like, labor intensive operation relying more on trial and error to achieve its goals. The result is that software development is prone to defects, or bugs, delays, and cost overruns. According to one estimate from the late 1990s, 40–80% of a typical software project budget goes into fixing bugs created earlier in the project, and 25% of all software projects are cancelled by which time they are 100% over budget. In other words, software productivity and quality have lagged that of hardware, creating a software bottleneck.

The persistence of the bottleneck is not because of a lack of effort to improve quality and productivity. Instead, it must be understood in terms of the characteristics of software. Developing software is about abstracting and representing processes in various domains of human activ ity, since lines of code are meaningless unless they serve a social purpose. Fredrick Brooks refers to the challenges of conceptualization and abstraction as the inherent or essential difficulties of software, and the problems posed by the representation and production of software as the acci dental difficulties. He argues that the bottleneck persists because most technological advances aimed at improving quality and productivity focus on the accidental rather than the essential. He further argues that overcoming the essential difficulties will be harder and slower, with no easy solutions or silver bullets.

Underlying Brooks' skepticism is an acknowledge ment that developing software is as much a social as a technological process. The social dimension of software development is crucial especially because of the need for communication between various stakeholders in the process.

For software to be effective and useful, its developers must possess both an understanding of the domain for which the software is being developed, and a command over the programming language in which that under standing is to be represented. This is just like an author, say, of a book on geography in English, who must min imally possess both a command over the language, and insights into the subject, for the book to impart any knowledge of geography to its readers. But developers rarely possess domain knowledge, whereas clients, users, and domain experts are rarely trained to represent their knowledge in a programming language. Unfortunately, bridging this gap by having experts convey to developers, the practices and processes in their domain, is fraught with miscommunication, misunderstanding, and con ceptual confusion. This problem is magnified as projects get larger and when the development task is partitioned among ever larger teams. Not only must the task be partitioned to facilitate the seamless integration of vari ous subtasks into a functioning whole, but the subtasks must also be explained to team members who may lack domain knowledge.

Even if client or user requirements are clearly understood, software development faces another prob lem. Since software is a collection of functions, users need not invoke the same set of functions or, even if they do, the functions need not be invoked in identical se quence. Thus, as the number of lines of code grows, there are more routes to navigate through the software and the ways in which it can be used grow disproportionately. Foreseeing the different uses, and predicting how soft ware, especially those with hundreds of thousands of lines of code, will behave, when subject to diverse uses, is a challenge. This is exacerbated by the tendency of cli ents and developers to treat software as soft, that is, they take advantage of being able to add new functions or change the code with just a few keystrokes. But in doing so, they unwittingly tend to introduce new ways in which the software can behave, increasing the chances of cost and schedule overruns.

In response to the persistence of the bottleneck, three distinct approaches to software development have emerged: customized development, packaged software development, and open source development. The rest of this article describes how these approaches have emerged by taking advantage of new technologies, and their geographical implications. In doing so, it shows why, because of inherent limitations, none of them is a panacea to the bottleneck. The three approaches coexist as each transforms the bottleneck in particular ways. The question of which approach is preferable depends on the kind of software needed.

Custom Software Development and Software Engineering

Software development spread with the development of computers, as that is where microelectronics first found application and programming was in demand. Since early computers were expensive, customers developed software customized to meet their needs as efficiently as possible. But the bottleneck was recognized early and, in 1968, the North Atlantic Treaty Organization sponsored a conference in Germany to discuss how reliable software could be developed for increasingly complex and mission critical systems. At this conference, the term software engineering was first used, to encourage the nurturing of a discipline analogous to industrial engineering that has helped improve manufacturing productivity. Drawing on similar ideas, software engineering has proceeded along two fronts.

One front is production rationalization using structured programming to manage the software development process by segmenting it to permit coordination and control over programmer teams. The classic conception of the development process is the six stage waterfall model, so called because each stage – requirement analysis, specification, design, coding, testing, and maintenance – must be completed before flowing into the next (**Figure 1**). Recent software management practices have focused on adopting industry wide certification norms so that organizations adhere to the discipline of codified quality procedures in the development process. The norms include the ISO 9001/9000 3 standards prescribed by the International Standards Organization, and the Software Engineering Institute's (SEI) five stage Capability Maturity Model that was originally developed for the US Department of Defense. The stages measure process maturity from an initial stage where the process is *ad hoc* or chaotic, to a fifth and final stage where the process is well documented and optimizes quality based on continuous quantitative feedback. A slightly different approach is the move toward object oriented technologies. The focus here is on developing a library of objects, or modules of code akin to manufactured components, that can be assembled in various ways to create different products, with the ultimate goal of making all users developers.

The second front along which software engineering has advanced is automation. The shift to object oriented programming itself is based on the emergence of newer programming languages. The so called first generation languages of the 1950s required a working knowledge of the machine, as they were no more than the binary instructions, or zeros and ones, which directly controlled the computer. With subsequent generations, however, languages became less machine specific and increasingly high level or human like. High level languages allow faster programming by less skilled programmers, reducing labor time and costs. Further, the more software is intertwined with hardware, the less convenient it becomes in the long term. This is because maintaining and transferring machine specific older software to improved hardware, based on new technologies, is tedious and difficult. The development of Java in the 1990s, a cross platform object oriented language that can run on any computer, irrespective of the specifics of the underlying hardware, is another step in the development

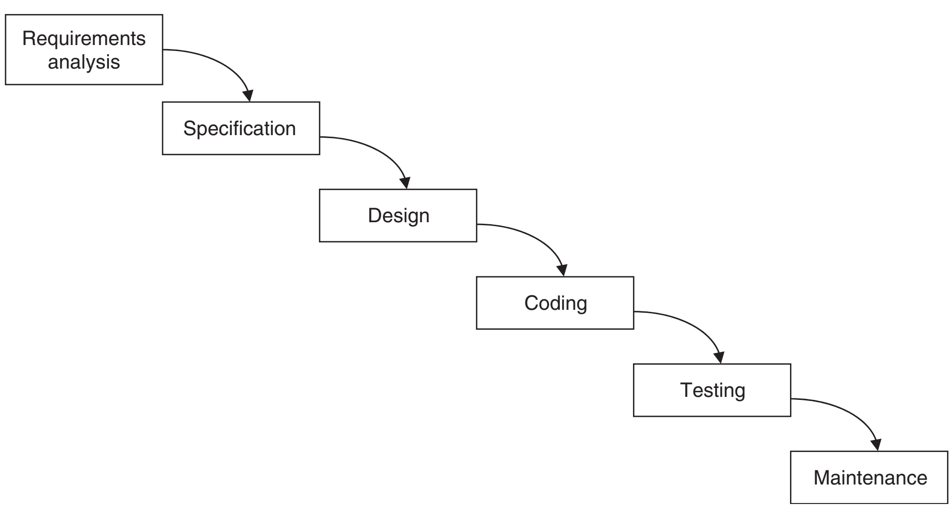

Figure 1 The waterfall model.

of high level languages. This 'write once, run anywhere' universality permits the hiring of programmers on the basis of their domain experience rather than their knowledge of specific technologies.

Although software engineering and management practices have helped improve overall productivity and quality, the software bottleneck is far from overcome, as there are wide variations in programmer productivity and quality. Indeed, since even the issue of whether current productivity metrics are at all appropriate remains unclear, some argue that reports of software productivity are often misleading numerical nonsense.

The waterfall model of development has been discredited for enforcing a rigid top down methodology. It is argued that no generic process model will work for all projects, much less one that is linear and sequential. In practice, software development tends to be iterative; there is learning by doing, with no sharp demarcation between conceptualizing and implementing. Thus, predictions about structured programming leading to a division of labor between mental and manual labor, as in manufacturing, have not been validated. Although the first three stages in the waterfall model are relatively more conceptual and skill intensive, and the last three stages have been amenable to automation through the use of computer aided software engineering tools, such automation demands clear, detailed specifications. As a result, a lot of implementing, testing, and coding is still done manually and the division of labor in software development is better understood as a division between more and less skilled labor, but skilled all the same. Similarly, the extent to which process standards, such as those laid down by the ISO and SEI, will guarantee a quality end product, is contentious.

While a product oriented approach like object oriented programming is popular, it is uncertain whether it has delivered what it promised. This has to do with the difficulty of overcoming the biases in the training of programmers that favors doing everything from scratch. The reliance on virtuoso problem solving highlights another problem: software engineering lacks the institutionalized mechanisms of a mature engineering discipline for recording and disseminating demonstrably good designs and ways to choose among design alternatives. Although a discipline of computer science has evolved to provide a theoretical foundation to, for instance, build formal mathematical models before writing code, packaging these results for operational use is inadequate. The result is that current software practice proceeds independently of this organized knowledge. Thus, what is practiced today is perhaps better described as software management than engineering, and some prefer to describe practitioners as soft crafters rather than software engineers.

The limited improvements in productivity and quality make software development a skill and labor intensive process. As the demand for new software spread, along with the need to maintain older software, skilled labor has become sought after. Unable to find enough people with the right skills, and under pressure to cut costs, customers in the advanced industrial countries of North America and Europe, have been forced to out source their development, triggering the globalization of the software industry since the 1980s. Facilitating this process, ironically, has been the falling cost of data communications. Thus, with custom software production relying on software engineering, the software bottleneck manifests itself as a labor shortage, and countries with the right skills at a low cost, such as India and Ireland, seized this opportunity to become global centers of custom software production.

Packaged Software

While software engineering was the adoption of mass manufacturing techniques for unitary software projects, the mass production of software in the form of packaged software had to await a mass market. The proliferation of computer usage from the late 1970s, with the availability of increasingly more powerful and inexpensive personal computers (PCs), created the market for packaged software and the opportunity for a different means of overcoming the software bottleneck. The market for a package of functions opens up when there are enough users with similar needs, each of whom finds cost advantages in shifting from a custom product. This is because, once software is developed, the cost of producing additional copies, or the marginal cost, is negligible since duplication is an electronic process, with almost no material or labor costs. Thus, the average cost of a package declines with every additional user as the development costs are spread across more users.

For producers, once a package is introduced for a specific set of functions, there are high entry barriers for a similar package. Software packages exhibit network externalities, where any increase in the total number of users increases the value of the package to the individual user. As the user base grows, the larger the number of people with whom a single user shares a standard means of representing, using and exchanging information. This, in turn, also favors monopolization since price is inversely proportional to the number of customers, and the lower the price, the more attractive it becomes to a larger number of users.

But as the numbers of users grow, a different issue arises. For users, the decision to buy a ready solution represents a tradeoff between the cost of the package and the utility of having software tailor made to their needs. While the developer of a package incorporates a general set of functions to reach as many users as possible, there are limits to the extent to which the package can anticipate and address various specific functions that a growing user base will demand. In other words, with an increasing number of users, a functional gap develops between the generality of the package and specific needs of individual users. For the user, there is a point beyond which these specific needs must be addressed by developing customized extensions. Then there is a wider market for a set of specific functions that, of course, opens up an opportunity to develop a new package.

This has been the mechanism and basis for the rapid growth in the use of packaged software. Aided by the proliferation of PCs and copyright laws, packaged software is a structural response to the software bottleneck. While software engineering attempts to lower development costs by relying on methodologies like object orientation to facilitate software re use within organizations, packaged software lowers the unit cost of software by reusing software across organizations, thereby incorporating all the productivity advantages associated with software engineering at the scale of the entire economy. Thus, packaged software transforms the software bottleneck. With custom software production relying on software engineering, overcoming the bottleneck calls for deploying more software professionals; with packaged software, it becomes a question of addressing the functional gap. The different approach to the software bottleneck is evident in the cost structures. Whereas the cost of labor and materials that go into production and distribution typically accounts for two thirds of the sales of custom software services firms, for packaged software firms it is less than one third, although labor is critical for developing any kind of software.

With labor costs playing a less important role, packaged software producers do not seek low cost locations. Instead, critical to meeting the software demands of any domain is an insight into the functional demands of the domain, and the practices that have evolved to address them. That proximity to the domain leaders offers the most potential for acquiring domain specific knowledge, especially intangible knowledge, is evident from the shifting geography of the US packaged software industry, the largest in the world. With the increase in the number of packaged software firms, the industry has moved away from semiconductor and computer manufacturing regions to those with a concentration of sectors where software usage is high. For instance, producers of packaged software for the oil industry are clustered in an application district around Houston, the energy capital of the country.

Open Source Software

When computers were expensive and rare, software that was developed was often shared among various users. But as software usage proliferated, and it became a lucrative product, commercial producers, especially of packaged software, released their products in binary code. This denies users the opportunity to either see or modify the source code, which is written in a high level programming language, even if the software is buggy. Frustrated with buggy proprietary software, users began to voluntarily come together to develop an open source alternative, where the source code is released with the software, for anyone to use, modify, and redistribute. The release relies on the Internet to provide an electronic public domain. The rationale is that with enough eyeballs, all bugs are shallow.

Software in the public domain is more likely to be used in various ways by different users. If functional enhancements are required, or if bugs are noticed, even a technologically ignorant user can bring it to the attention of the project community in the hope that someone capable will find the problem interesting enough to solve, thereby benefiting everyone. Since anybody can potentially provide the solution, open source development is a decentralized process. It is compared to a bazaar, where many voices and agendas are pursued. It is in contrast to development within a firm that is akin to cathedral building using a unique top down master plan. However, when having to choose between solutions offered to a problem, open source project communities typically evolve a decision making structure that decides between alternatives based on a combination of technical rationality and a normatively defined notion of elegant code.

Being a user driven development process, open source software production overcomes the problem of miscommunication between developers and those with domain knowledge. Further, while custom software projects and packaged software have to schedule formal tests with users before product release, the open source process makes every user a tester every time, and draws on their feedback to react nimbly to alter the software whenever a bug is identified or an enhancement is required. This improves quality and productivity, and minimizes maintenance costs.

Although it offers superior quality and productivity, participation in any open source project has two prerequisites. First, in a community defined solely by an interest, and with no face to face contact, shared domain knowledge is vital. While there have been a few high profile projects, such as the Linux operating system, that have formed a community running into the thousands, a 2002 study of 100 open source project found that most projects had a community of less than ten. This is because the larger ones tend to be based on universally shared domain knowledge in computer science and engineering. In contrast, developing an application in, say, the domain of banking, which is influenced by local regulatory requirements, makes it hard for users outside that regulatory jurisdiction to participate in development. Thus, with open source production, the software bottleneck is trans formed to an application gap. The second prerequisite for participation in an open source project is Internet access, which is hardly uniform across the world. Thus, the geography of open source production is dictated by the geography of Internet access and the digital divide.

Conclusion

After explaining the centrality of software to the revo lutionary character of the microelectronics based IT revolution, this article explained the quality and prod uctivity problems that have led to the software bottle neck. Three development approaches have emerged as responses to the bottleneck by taking advantage of technological advances. However, each of these ap proaches has only transformed the bottleneck, with an attendant geography of software production, and none provides a panacea. Thus, there are no silver bullets, and all three approaches coexist despite emerging from dif ferent technological periods.

The question of which approach to use depends on the kind of software required. Customized software de velopment that relies on software engineering emerged when computers were expensive and rare. However, it is still relevant when mission critical software or a unique application is needed. Packaged software emerged with the spread of PCs, and works when there is a mass market for a set of functions and users are willing to prioritize price over a tight fit with their specific needs. The open source approach is appropriate when there are users so frustrated with commercial output that they are willing to work voluntarily with strangers to develop reliable, robust software.

See also: Global Production Networks; High-Tech Industry; Industrial Organization; Innovation; Knowledge Economy; Regional Development and Technology; Technology and Regional Development; Technology Industries.

Further Reading

Abbate, J. (1999). *Inventing the Internet*. Cambridge, MA: MIT Press.
Brooks, F. P. (1999). *The Mythical Man Month: Essays on Software Engineering* (2nd edn.). Singapore: Addison Wesley Longman.
Brown, J. S. and Duguid, P. (2002). *The Social Life of Information*. Boston, MA: Harvard Business School Press.
Campbell Kelly, M. (2003). *From Airline Reservations to Sonic the Hedgehog: A History of the Software Industry*. Cambridge, MA: MIT Press.
Ceruzzi, P. E. (2003). *A History of Modern Computing* (2nd edn.). Cambridge, MA: MIT Press.
Cusumano, M. A. (1991). *Japan's Software Factories: A Challenge to US Management*. New York: Oxford University Press.
McConnell, S. (1999). *After the Gold Rush: Creating a True Profession of Software Engineering*. Redmond, WA: Microsoft Press.
O'Riain, S. (2004). *The Politics of High Tech Growth: Developmental Network States in the Global Economy*. Cambridge: Cambridge University Press.
Parthasarathy, B. (2004). India's Silicon Valley or Silicon Valley's India? Socially embedding the computer software industry in Bangalore. *International Journal of Urban and Regional Research* 28, 664 685.
Shapiro, C. and Varian, H. R. (1998). *Information Rules: A Strategic Guide to the Network Economy*. Boston, MA: Harvard Business School Press.
Weber, S. (2004). *The Success of Open Source*. Cambridge, MA: Harvard University Press.

Informational City

J. Rutherford, LATTS (Laboratoire Techniques, Territoires et Sociétés), Marne-la-Vallée, France

Glossary

Informational Capitalism A techno-economic system held to have emerged in the last thirty years or so in which information generation, processing, and transmission are seen to drive the productive forces and restructuring of capitalism.

Informational Worker/Labor Tertiary and quaternary sector workers whose main tasks involve the production, exchange, interpretation, and/or translation of information and knowledge.

Multimedia Convergence A process involving the merger of data-processing technologies and computer services with telecommunications systems to allow the widespread networking and circulation of information. The integration of broadcasting technologies and services into this combined operating platform is a further recent step.

Network Society A concept covering views that social and technological networks (and the increasing convergence between the two in some cases) now constitute the dominant social form of contemporary societies through their organization of major economic, cultural and political structures, and activities.

Space of Flows Castells' concept to capture the communicational space created by new techno-economic networks and which is seen as now constituting the dominant functional arena (as opposed to the space of places) in which information elites mobilize, connect, and interact to sustain global urban economies.

Symbolic Analysts Elite workers whose main tasks involve interpreting, analyzing, and transforming symbolic (valued) information based on words, numbers, images, etc.

World City A strategically important urban node in which certain activities and operations are carried out which may shape and contribute to economic, political, and cultural globalization.

Preamble

Information is, in spite of its somewhat opaque and elusive nature, the *sine qua non* of the city, essential for creating and sustaining all kinds of intra- and inter-urban social, economic, political, and technical linkages without which the city as a concentration of people, buildings, institutions, firms, capital, etc., would not be able to exist. The mutual dependencies between cities and information have rarely been anything less than crucial, but they seem to have been reinforced even further in recent times by ongoing socioeconomic and technological transformations. Social and economic change has, in particular, redefined urban regions as systems for the production, management, and distribution of information, held to be the key raw material in a postindustrial society. At the same time, technological changes created first new, highly reliable forms of information technologies, and then, through multimedia convergence and digitalization, allowed these to be linked to ever faster telecommunications to promote intensive global networking and exchange. Because of the importance of information circulation to the functioning of urban societies and economies, information and communications networks have been seen by some writers to be as crucial for contemporary cities as the development of the railway was for industrial cities in the nineteenth century. Yet, the networked nature of this information processing has served to loosen the ties that have traditionally bound cities to specific territories. The informational city is the geographical expression of this changing techno-economic paradigm. We should note that it touches on aspects of other urban theories, notably the global city and its concentration of command functions and producer services supporting the global economy, even if the latter more clearly defines a specific set of applicable strategic sites, rather than having tendencies toward a more comprehensive theory of urban development in advanced economies.

Introducing the Informational City

It is argued that one of the most fundamental ongoing societal changes is concerning the shift from an industrial age to an informational age: that is, from an economy and society based upon the production and circulation of physical goods (from raw materials) and the forms of primary and secondary labor this necessitates, to an economy and society based upon the production and circulation of information and knowledge (symbolic goods of a more abstract or virtual form), in a pluralist sense, and the tertiary- and quaternary-sector workers (the informational labor or symbolic analysts) required to interpret or translate this information. Economic productivity is now derived primarily from knowledge generation, information processing, and their technological transmission and exchange. Information is the raw material which requires some level of knowledge to be transformed, with accumulated knowledge and more complex forms of in formation commodities as the outcomes (which are then fed back in as the process begins again).

From this new form of socio technical organization – an informational mode of development – in which the emergence of new information and communications technologies was paramount, the concept of the infor mational city was developed during the 1980s by the urban sociologist Manuel Castells to capture how urban regions in advanced economies were being transformed by parallel dynamics associated with capitalist re structuring and the rise of information intensive indus tries (informational capitalism). Castells sees spatial form as the expression of societal organization, so in his ar gument there are technological and organizational di mensions to the informational mode of development, and it is their intertwining which permits a reconfiguration of production processes and therefore of urban and regional space. Information intensive firms are restructuring and diversifying their production systems and locational strategies according to the means and capacities of new information technologies. This creates new spatial dy namics for the internal organization of these firms, but importantly also for cities and regions up and down the urban regional hierarchy.

In this way, the impact of new technologies upon cities is addressed by exploring the sociospatial complexity of the restructured relations between these technologies, the production process, urban society, and urban space. Gone are the simplistic prophecies and hype which have commonly enveloped the emergence of new urban technologies in the past, and most notably ideas about a dissipation of cities with increasing ubiquity of informa tion. Informational cities have important interconnected infrastructures of information and communications technologies, but it is the basis of their economies upon information which distinguishes their development. The dominant spatial logic associated with this growth of information processing is one of parallel decentralization and concentration, focused on and enabled by a dense web of techno economic networks which Castells has termed the 'space of flows'.

The Economic Basis of the Informational City

Information technologies have transformed the social and spatial organization of information processing activities, which, in turn, are viewed as more and more central to urban economies because economic production, distri bution, and management are increasingly dependent on knowledge generation, information exchange, and infor mation handling (and therefore on skilled informational labor). Yet, the spatial logics resulting from these transfor mations are not homogeneous. Noting that the emerging techno economic paradigm applies to the new industrial space and the new service economy, Castells observed four main spatial dynamics common to large scale information intensive industry firms (which include banking, insurance, business and legal services, central administration, etc.):

- a distinctive spatial division of labor with different parts of the production process requiring different types of work and work arrangements;
- the main function in these industries is that of infor mation production (which may rely either on quan titative data handling or on more tacit information processing) which is located in a small number of innovative milieux;
- decentralization of other functions in the production process, which leads to a spatial hierarchy in parallel with the structural hierarchy of the industries; and
- locational flexibility in production (except for the innovative milieux) as market locations change constantly.

These functional and spatial specialization dynamics represent considerable changes to the internal structures of firms which have become more hierarchical, splintered, and flexible, or even vertically disintegrated. Information technologies have promoted flexible production techni ques, while technological convergence with communi cations networks has allowed greater intrafirm networking to keep operations coherent and efficient. Multinational companies create competitive advantage therefore by information exchange and management across their structures. Nevertheless, the idea of totally footloose en terprises being able, via information and communications technologies, to locate all their units anywhere is some thing of a myth. The decentralization of some activities reinforces the need for a concentration of others, par ticularly tactical, coordinating, or strategic activities which remain heavily dependent on advantages of agglomeration such as face to face interaction or copresence, tacit knowledge, and interfirm synergies (which represent a more intensive, scarcer, and valuable set of information inputs), for example, in financial centers, technology clusters, or industrial districts. In this way, almost in spite of technological advances, the information economy re mains at least partly an economy of presence. As tech nologies facilitate the production of more information, there is a need for increasing numbers of workers and interaction between workers to interpret and to act on the information. Workforce restructuring is nevertheless an important element in the reorganization of firm strategies. This may lead to such processes as streamlining, sub contracting, automation, relocation, or use of temporary or part time labor, all of which have important consequences for work and labor in the informational city. Indeed, Castells suggests that employment structure in advanced economies is now dominated by skilled, highly adaptable information workers supported by a more flexible, pre carious workforce submissive to the needs of markets at particular times.

Organizational hierarchy in information production and processing leads in turn to a spatial hierarchy of urban regions based on these parallel decentralization – concentration logics. Decentralization of specific activ ities may involve a move to either smaller urban regions where operating and workforce costs are lower or sub urban areas of larger cities thus also privileging labor and market accessibility. At the top end of the hierarchy, major cities (and indeed the main business districts within these cities) are reinforced as the sites of infor mation production (most recently including that of the Internet content market) and informational decision making. They become the clearing houses for global information transactions. Crucially, it is the process of new technologies facilitating dispersal of some activities that allows the centralization of other activities.

The global informational economy is thus founded on a network structure, or, more precisely, on a structure of innumerable intertwined networks. These techno or ganizational networks might be specific to individual firms or individual sectors, or they might interconnect multiple groups and organizations. They do tend, nevertheless, to share a common interurban logic, due to the location of economic actors (information elites) pri marily in cities.

The Spatial Logic of the Informational City: The Space of Flows

Command and control functions within firms and within the information economy as a whole are allocated then to central cities. These cities (or the districts within them) can be seen as increasingly disconnected from the rest of the production space and increasingly interconnected between themselves in specialized global information networks. Castells has developed the concept of a space of flows to describe what he perceives to be the now dominant deterritorialized communicational space – of information, people, organizations, capital, etc. – which organizes many socioeconomic processes, and through which cities (or specific zones of cities) are interlinked. This new space of flows is the archetypal functional logic of the informational city. Supported by information and communications technologies, it is held to dominate (but not eliminate) the more traditional, historically con structed space of places. The former overlays the latter like a red carpet on the ground. Although the space of flows permits deterritorialized network connection, which is the new structural logic of the global economy, its central nodes remain at the same time anchored in the traditional space of places. Organizations are located in places, and their workers and materials are place based, but the overall organizational or functional logic is flow based, whether we are talking about a multinational company or an airport.

Global finance firms locate in the City of London or in lower Manhattan both for the planetary connections these nodes attract and for the dense series of place based interactions made possible by agglomerations of similar activities. Nevertheless, information elites and infor mation based firms have been freed up from the local constraints of the latter and now mobilize, connect, and interrelate far more in flows than in places. Indeed, this may even be an ultimate objective of all the techno economic change that urban regions have undergone in recent times. The organization of the global urban economy primarily in a space of flows allows production, management, and markets (economic power) to bypass as much as possible forms of territorial regulation which exist to control them.

Described by Castells, the space of flows has three layers of (mutually constituting) physical supports: a circuit of electronic impulses such as telecommunica tions; a network of place based nodes and hubs; and the spatial manifestations of dominant groups and interests. In this way, while the technological infrastructure of the first layer provides a foundation for contemporary socioeconomic processes, which take place both in and between the key cities of the second layer, it is in the third layer that powerful and wealthy elites control, configure, and articulate their own spatial requirements, frequently at the expense of the disconnected masses. The space of flows does not therefore correspond only to cyberspace or to a network of urban nodes, but works through and is the outcome of dominant social practices. The flows in the space of flows are always both the basis and product of decision making, and are structured and directed in accordance, which helps to explain why access to the space of flows is sociospatially highly uneven.

The Informational City as Dual City

The space of flows, the new architecture of the global economy, is then sociospatially differentiated, in terms both of where and how its informational webs connect to and become accessible by various places and social groups, and of how the space is actually managed and controlled by particular information elites. While the latter mobilize and circulate (physically and virtually) across the globe ever more rapidly and intensively, other social groups remain relatively or completely excluded from this premium net work space. A dual city results – the outcome of perpetual socio technical and instrumental segmentation. It is here that the space of flows–space of places argument extends the economic development focus of the global/world city hypothesis into the domain of social class distinctions.

In an informational economy and network society, those who manage and circulate information (net workers) constitute a socioeconomically affluent group who are fully integrated into all manner of high level networks. By contrast, those whose labor does not meet the informational requirements of the global economy constitute a devalued and socioeconomically precarious group who do not fit in with the logic of the network society and are therefore disconnected from many aspects of social and economic life. Such sociospatial polarization is occurring both within and between urban regions in all parts of the world. Graham and Marvin have examined the sociospatial consequences of combined logics of network disconnection and wider urban fragmentation in a process they call 'splintering urbanism'. While information and communications technologies may have been the driving force behind the emergence of the space of flows, at the same time they are being configured to create and reinforce specific forms of dominant socioeconomic practices with socially divisive repercussions:

- these technologies and the services offered through them are accessible only to the most affluent groups in society, such that urban populations today are digitally divided into information haves and have nots;
- this is increasingly affecting accessibility of less affluent groups to basic essentials such as money as financial systems become informatized and withdrawn into the realm of electronic space; and
- these technologies are also increasingly being used for social control and surveillance, discriminately (or indiscriminately) tracking individuals through all aspects of their daily lives.

The informational city and its dominant spatial form, the space of flows, are therefore both constitutive of and operated by a vast array of asymmetric power relations between highly connected elites and disconnected social groups. We should note, however, that this digital control perspective is somewhat complicated by views of elites holding more essential command over information economies of presence through their management of face work.

On a global level, similar polarizing processes are at work creating and sustaining an informational urban hierarchy atop which commonly are the urban centers of command and control in the Global North, while many of the most populous cities of the Global South struggle to integrate themselves into an informational, network based world economy.

Implications for Urban Theory

The informational city is based on a techno economic reconfiguration of production systems around information processing units, with subsequent joint decentralization and concentration dynamics, in which functions of control and management are located in major urban centers. These centers and the elites responsible for information circulation are highly interconnected within a new spatial form, the space of flows, but at the expense of disconnected and devalued social groups, which leads to new forms of urban polarization. Given the complex nature of these processes and logics, it is unsurprising that the notion of the informational city contests in many ways some of the more traditional definitions and ideas of the city.

Foremost among these is the vision of the city purely as a bounded, territorial entity. In an economy and society dominated by information and knowledge production, circulation and mobility through overlapping networks, and the logic of the space of flows, it is clear for many writers that the city has become more of a dynamic process than a fixed place. The space of flows as the dominant spatial form of the informational city has three layers of physical supports, each of which depends on and brings together processes and practices from different scales. Technological infrastructures are designed and deployed according to local contexts, national regulations, and global markets. Networks of global cities develop under similar multiscalar conditions. The spatial practices of elites are at once constituted by personal micro networks of contacts, spaces, commodities, and lifestyles, and functional macro networks which cast these into the transactional, decision making arena of the space of flows. Cities which are composed primarily by such technical, economic, and social flows can hardly be seen only as territorial units nested within nation states. In an age of almost limitless information, the informational city may also be virtually limitless. At least one reading of the global city argument discounts the idea of multiple global cities interlinked by networks in favor of one truly global city. These visions of the city as a process have opened up a fruitful research avenue into the nature and intensity of inter city information flows. This research, conducted most notably by the Globalization and World Cities (GaWC) network, is based on the argument that the power and control capacities of certain cities are now determined more by the information that flows through them than by any concentrations of information within them.

All this does not mean, however, that the city as a place no longer exists. Indeed, informational cities are diverse because of their varying place based contexts and assets. Castells has himself adapted or qualified some of his analysis regarding the spatial logic of the informational city. He has stressed, for example, how the space of flows can include some places (and the space of places some flows) reaffirming therefore how it is not just a global virtual space of electronic connections. He notes that the informational city is no more likely to continu ously replicate Silicon Valley or Los Angeles than the earlier industrial city was to resemble Manchester. It is just that the specificities of these contemporary cities are outweighed by a common spatial logic: they are domin antly organized socially and economically by network flows. Some of the most exciting recent urban research has therefore been intent on exploring the varying ten sions between network and territorial logics in urban transformations. In particular, there are constantly changing implications for urban governance in the in formational city as the production and locational strat egies of economic firms evolve. The actual or potential influence and action of local and regional authorities, still bounded by spatial and temporal restraints, appears out of step with the unhindered decision making of large information intensive firms. Rejuvenating local govern ments to develop controlling capacities over network based flows would be desirable, but much research indicates that in these entrepreneurial times they are more often pandering their territories to the space of flows than regulating the effect of the latter on people and places.

Critiques and Challenges

A number of critiques have been leveled directly or in directly at the concept of the informational city. The following are the most notable.

First, some commentators have observed that there is little or nothing new about the informational city, the resources upon which it is based, and its spatial, flow based logic. This critique is similar to that concerning the global/world city (or indeed globalization as a whole) where it is argued that the transnational economic system of the seventeenth century was just as focused upon a small interconnected group of functionally important cities as it is now. Information and knowledge are held to have always been determining factors in urban pro duction systems, permitting significant comparative ad vantages for cities throughout history, from the German Hanse of the fifteenth century and in northern England during the Industrial Revolution, right up to modern day Silicon Valley. In all these cases, flows of information are central. The argument that there is now merely more information or more rapidly available information has difficulty fitting with a theory propounding a major re structuring of economies and societies. For neo Marxist critics, there is either not enough focus on how the in formational society and the informational city are still constitutive of capitalism, or there is not enough evidence of how these represent a major new phase in capitalist development.

Second, the information and knowledge resources upon which the techno organizational complex of the informational city depends have perhaps been overly assimilated. While data information is widely available, knowledge (seen as skill or know how) is created by skilled interpretation or manipulation and as such is more rare and valuable. Crucially, contrary to infor mation, it is thus far from clear that the space of flows logic is applicable to knowledge organization, which may be territorially embedded and more reliant on locational proximities and face to face interaction. In this way, the notion of an informational city at the heart of economic production concentrates too much on ubiquitous infor mation flows and devotes too little consideration to the place of knowledge elements which are increasingly central to prosperous urban economies. The tendency to see information or the information economy as almost ubiquitous renders these terms almost meaningless as virtually any form of work can be seen as informational under some guise. The critique of this sees a far smaller percentage of urban labor as information workers which questions, as a result, the whole concept of the infor mational city.

Third, there is an overwhelmingly economic focus to the informational city, with relatively little consideration of the wider range of social groups and viewpoints in volved in its constitution or reshaping. Castells' more recent work has begun to address this, but there is a need to integrate a more diverse set of narratives and voices into this perspective to get beyond the dominant infor mational bourgeoisie (who themselves are undoubtedly more diverse than often portrayed). More focus on the plurality of information has been suggested as one way forward. It is also worth remembering that intensive socio technical connection is not a common objective for everybody in society. Many people complain about in formation overload and become rather blasé faced with multimedia technology systems.

Fourth, there are arguments suggesting a dominant technological or network determinism to aspects of the work of Castells. This derives from the way in which he attributes major importance to the information technol ogy revolution as a kick start to a new informational mode of development. Yet, Castells explicitly rejects any notion of technological determinism and is indeed at pains to show how the new mode of development is as dependent on organizational change as much as tech nology and is also tied into the wider restructuring of capitalism from the 1970s onward. The informational city is therefore an expression of all arrangements of the in formational society. A wider network determinism ar gument suggests that employing a unique network rationale to describe general spatial and organizational processes within an informational paradigm is problematic in the sense that it flattens quite significant distinctions between different forms of network as well as underplaying the specific importance of institutional regulation in the shaping of networks (e.g., technical networks) over time and space.

There are then a number of interesting theoretical issues around informational cities and political challenges for urban managers, but we have the space to highlight only two here.

First, what of the (possibilities for) development of more ordinary informational cities beyond both the core set of global cities and the Global North? Further research might focus, for example, on the degree to which information technologies and information production, processing, and circulation as a whole are actually transforming urban space and urban form in cities of the South. Or within a global political economy of information, are these cities still destined, at least for the time being, to remain bypassed by and unintegrated into the space of flows? If there still are fewer telephone lines in the whole of sub Saharan Africa than in Manhattan, perhaps we need to talk here about diverse and quite subjective forms of information and technology on the ground, rather than assuming that there is one exportable (or desirable) informational city model that can be developed the world over.

Second, and partly following on from this, the main political challenge will be to explore the possibilities for opening up the informational city to a broader set of social groups than is currently the case (which will undoubtedly require a renewed role for all levels of the state), or for urban based resistance to the dominant (economic) informational logic. In this regard, Castells himself has begun to discuss these possibilities for grassrooting the space of flows by illustrating how the nature of the space of flows as produced and managed by dominant groups opens it up to alternative projects of resistance, diversification, and new social and cultural meaning. Furthermore, most recently, he has outlined a notion that it is the switchers who hold the most power in the space of flows by acting as the intermediaries between differing networks. Importantly, this group does not necessarily coincide with the affluent, dominant groups who are most intensively integrated into these flows. In any case these tendencies offer possibilities or opportunities which may ultimately reshape the space of flows and therefore the political economic functioning of societies.

People live in places, while power and domination rules through flows: thus is the overwhelming and slightly contradictory logic of the informational city. It reminds us though that there is nothing at all abstract (or indeed certain) or discarnate about current urban restructuring and development. Cities of the twenty first century will continue to be built and reshaped according to the intertwining logics of territories and networks, the local and the global, and meaningful places and functional flows.

See also: Information Technology; Networks, Urban; Technological Change; Telecommunications; World/Global Cities.

Further Reading

Brunn, S. D. and Leinbach, T. R. (eds.) (1991). *Collapsing Space and Time: Geographic Aspects of Communications and Information*. London: Harper Collins.

Castells, M. (1989). *The Informational City: Information Technology, Economic Restructuring and the Urban Regional Process*. Oxford: Blackwell.

Castells, M. (1996 1998). *The Information Age: Economy, Society, and Culture*, 3 vols. Oxford: Blackwell.

Catterall, B. (2002). Informational cities: Beyond dualism and toward reconstruction. In Bridge, G. & Watson, S. (eds.) *Companion to the City*, pp 192 206. Oxford: Blackwell.

Gottmann, J. (1983). *The Coming of the Transactional City*. College Park, MD: Institute of Urban Studies, University of Maryland.

Graham, S. and Marvin, S. (2001). *Splintering Urbanism: Networked Infrastructures, Technological Mobilities and the Urban Condition*. London: Routledge.

Hall, P. (1998). *Cities in Civilization: Culture, Innovation, and Urban Order*. London: Phoenix Giant.

Hepworth, M. (1989). *The Geography of the Information Economy*. London: Belhaven.

Mitchelson, R. L. and Wheeler, J. O. (1994). The flow of information in a global economy: The role of the American urban system in 1990. *Annals of the Association of American Geographers* 84, 87 107.

Robins, K. (ed.) (1992). *Understanding Information: Business, Technology and Geography*. London: Belhaven.

Sassen, S. (2000). *Cities in a World Economy* (2nd edn.). London: Pine Forge Press.

Storper, M. (1997). *The Regional World: Territorial Development in a Global Economy*. London: Guilford Press.

Thrift, N. (2002). A hyperactive world. In Johnston, R., Taylor, P. & Watts, M. (eds.) *Geographies of Global Change: Remapping the World* (2nd edn.), pp 29 42. Oxford: Blackwell.

Webster, F. (2002). *Theories of the Information Society* (2nd edn.). London: Routledge.

Wilson, M. and Corey, K. (eds.) (2000). *Information Tectonics: Space, Place and Technology in an Electronic Age*. London: Wiley.

Zook, M. (2001). Old hierarchies or new networks of centrality? The global geography of the Internet content market. *American Behavioral Scientist* 44, 1679 1696.

Relevant Websites

http://www.lboro.ac.uk/gawc
Globalization and World Cities (GaWC) network.

Innovation

R. Sternberg, University of Hannover, Hannover, Germany

Glossary

Innovation A new combination of production factors; in a narrower sense the first introduction of an invention in the market.

Innovation Policy All policies which intend to support the innovation activities of (real or potential) innovative actors in a given area; incentives try to reduce the cost and risks of innovation.

Innovation System Comprises all determinants of the innovation process, that is, all important economic, social, political, organizational, institutional, and other factors that influence the development, diffusion, and use of innovations; spatial (e.g., continental, national, regional, and local) as well as aspatial innovations systems (e.g., sectoral) are distinguished.

Knowledge Spillover The shift of newly created and/or accumulated knowledge from the inventing person/organization into productive capacity and the related persons and organization; such spillovers mainly take place within the same locality.

Tacit knowledge Not codified and/or codifiable knowledge which primarily flows through informal networks and which is usually privately owned.

Why Is Innovation Relevant – for Economic Geographers?

Definitions

Innovation as a concept was introduced by the eminent Austrian economist Joseph Schumpeter and means more than just novelty. To him, innovation was a new combination of production factors that set a process of creative destruction (including in regions) in motion. He saw basic innovations, clustered in time, as the triggers of long waves of economic development, which other scholars later applied also to regions. For a long time, the innovation process, defined as the first introduction of an invention in the market, was presented as a linear process consisting of the phases of scientific discovery (invention), product development (innovation in its true sense), market introduction, and diffusion. Today the consensus is that the chain linked model is moving closer to reality with its numerous linkages and feedbacks between discovery, development, and commercialization in times of flexible specialized regional economies; however, it is more relevant to industrial innovation instead of innovation in general (see **Figure 1**). Although the term innovation can also be applied to noneconomic phenomena (social innovations, organizational innovations), the economic causes and effects are by far the most significant for economic geographic innovation research.

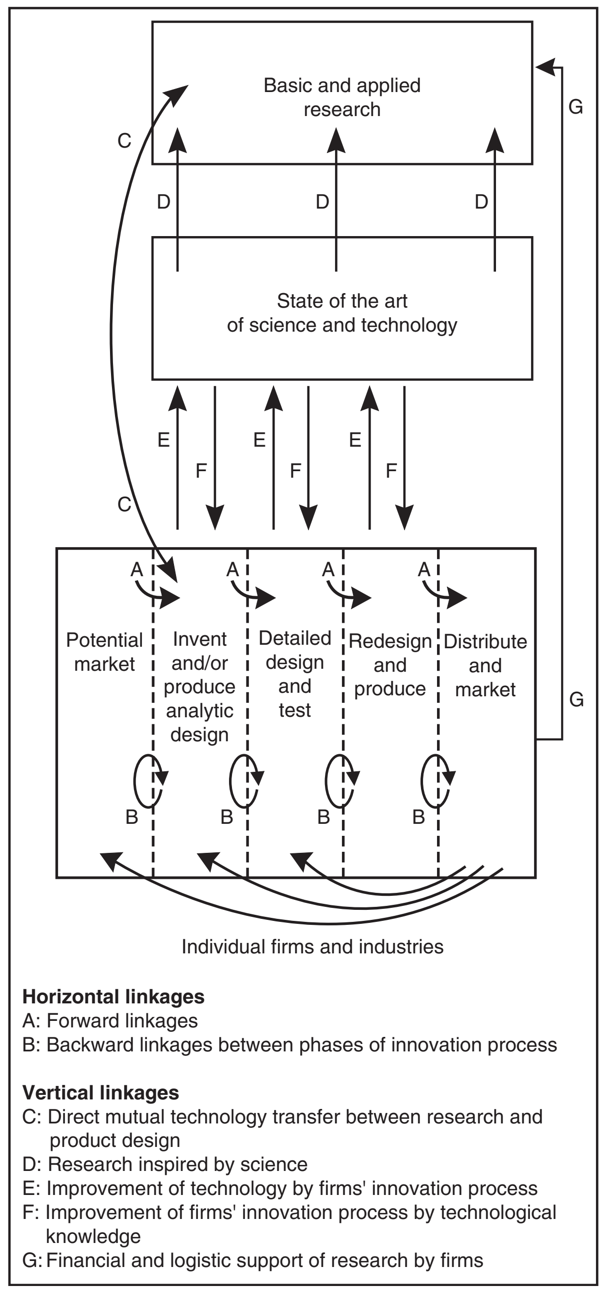

Figure 1 The chain-linked model of the innovation process.

Spatial Implications of the Genesis of Innovation and of Innovation Diffusion in Space

There can be no question that the generation of innovations and their diffusion in space and time are among the main determinants of the economic and technological competitiveness of nations and subnational regions. From an economic geographer's perspective, it is important to emphasize that, first, the genesis of innovation differs considerably from one region to another and, second, the characteristics of regions play a part in shaping this genesis. The geography of innovation is very uneven. Why? Essentially, there are five characteristics of knowledge underlying why location matters for innovative activity: (1) the uncertainty of the innovation process, (2) the reliance on university research, (3) the cumulative character of innovative activity, (4) the importance of learning by doing, and (5) the complexity of the innovation process. Innovation needs new knowledge, which can be either tacit or codified. The transition from tacit to codified knowledge is difficult and distance sensitive.

New knowledge in turn requires the presence of persons who create this new knowledge. In contrast to information, new knowledge is mainly tacit, meaning it is tied to individual persons and cannot be transferred, or at least can only be transferred by face to face contact. Face to face contact, however, entails spatial proximity to the persons generating or in possession of this tacit knowledge. New information and communication technologies do not help reduce the distance sensitivity of tacit knowledge as they are only suitable for the diffusion of already codified knowledge. Since the number of creatives is rather small, and they are also very spatially mobile and selective, there are only a few potential regions in a position to generate genuinely radical innovations with economic potential. Depending on the theoretical perspective, the innovativity of regions depends on very different determinants. From the perspective of the individual, cognitive abilities, entrepreneurial capacity, and attitudes toward risk and competence are important. From an institutional perspective, the culture, educational values, and resistance to change are important. The innovation system approaches emphasize sectoral systems, networks, and regional and national systems of innovation, and from a firm perspective it is crucial how R&D, the intrafirm selection, and diffusion process are organized.

The second important aspect from the perspective of economic geographers is the diffusion of innovation over space. The Swedish geographer Hägerstrand demonstrated empirically based on observations in agriculture that the adoption of innovations outside their region of origin depends on hierarchies (city rankings) and neighborhood effects (as a result of social contacts). Hägerstrand modeled the diffusion processes using Monte Carlo simulations and other stochastic processes, and was able to visualize different phases of the diffusion process in the form of a logistic S curve. More recent works researching diffusion demonstrate that spatial diffusion depends in particular on the willingness of innovators to share their tacit knowledge and on the absorptive capacity of the (potential) adopters. The concept of absorptive capacity can be translated to apply to regions as the varying openness and capability of regional innovation systems for extraregional innovations determines the perception and usage of innovation opportunities within a region. Entrepreneurs play an important role as knowledge filters in this process as they are in a position to and are interested in adopting innovations quickly. As entrepreneurs are additionally an element of the endogenous regional growth potential as a result of their considerable spatial immobility, it is also efficient from the regional perspective to support entrepreneurs for the sake of boosting the innovational strength of the region. Learning processes are necessary for the generation and transfer of new knowledge and therefore innovations, as the generation of innovations is typically a shared process due to the high level of uncertainty and the limited capacity of individuals. Inter and intraregional innovation networks and collective learning therefore play an important role. These processes also have clear spatial implications, as the discussions of intraregional innovation networks as a result of the distance sensitivity of knowledge spillovers and tacit knowledge, and the discussions of learning regions show. Another argument for the significance of collective learning processes permanently fed by intraregional networks is 'ubiquitification'. Here the fact is addressed that formerly locally specific production factors (e.g., specialized labor or infrastructure like business incubators) are more easily accessible worldwide in this period of internationalization and globalization; in other words the source regions tend to lose their competitive advantage. Solely knowledge, which is hardly able to be codified (and, thus, more difficult to transfer), yet well embedded in the regional innovation system, is immune to ubiquitification.

To conclude, the innovation issue is relevant for economic geographers for various reasons. In a globalized world with decreasing lengths of product cycles, the economic growth of nations and regions depends more than ever before from their ability to continuously produce all kind of innovations. As for the distribution of innovations over time and space, it seems to be clear that we have an uneven landscape of innovations and innovation behavior: regions (as individuals) differ in terms of their capacity to absorb new knowledge created elsewhere and to transfer it into growth and development. And finally, the creation of new knowledge as an indispensable precondition of innovation is distance sensitive and depends, at least partially, upon the local/regional embeddedness of actors and institutions. Therefore the importance of innovative (regional) clusters and innovative regions is reinforced by the fact that spatial proximity and the regional embedd edness of innovative actors like firms, individuals (entre preneurs, creatives, researchers), financiers, and policymakers is key to the effective production and sharing of tacit knowledge.

Regional Innovation Theory

The explanation of regional innovation processes has been one of the most popular subjects of human geog raphy and neighboring disciplines in the last two decades. There are many theories and schools of thought com peting with one another – none of which has yet achieved general acceptance. The consensus among economic geographers is that the genesis of innovations and their diffusion over space can be explained to a considerable extent by regional features – a finding that is constantly attracting more supporters also in the most important neighboring discipline within the framework of evo lutionary economics in particular, despite the fact that some scholars correctly point to additional dimensions of proximity (besides the spatial dimension).

Territorially Based Systems of Innovation Approaches

The regional innovation system (RIS) is a concept that has been created as a consequence of the national in novation systems (NISs) concept developed several years before. Both approaches take into consideration the sys temic character of the innovation process. A system of innovation comprises all determinants of the innovation process, that is, all important economic, social, political, organizational, institutional, and other factors that in fluence the development, diffusion, and use of innov ations. The constituents of systems of innovation are components (organizations and institutions) and rela tionships among them. Important from an economic geography perspective is that an RIS (NIS) considers all components and relations among these organizations and institutions relevant in and for a selected region (nation). This does 'not' mean that all relevant organizations have to be permanently domiciled in the region (nation) in question. Rather, interregional (international) con nections and therefore extraregional (extranational) or ganizations may also influence the functioning of an RIS (NIS). Scholars emphasize the significance of intra regional connections within RISs, mainly because of the spatial consequences of tacit knowledge and the ex change/transfer of such knowledge. The most important factors to be mentioned in this respect are knowledge spillovers (e.g., persons or institutions adjacent to persons or institutions producing new knowledge may profit from this neighborhood and they innovate more than if they would be located elsewhere), the similarly distance sensitive transfer of tacit knowledge, and continually significant differences in innovation intensity between regions. Innovative firms (and innovative entrepreneurs running such firms) are the most important players within an RIS. In the work sharing innovation process, they use and need the complementary assets of other innovation actors in the region. The relations between innovative actors within and between RISs can be structured as shown in **Figure 2**.

Cyclical Models of Innovation-Based Regional Development

The basic postulation of the 'product cycle theory' is this: Products and industries undergo an aging process con sisting of a phase of genesis and introduction, a growth phase, and a maturity or standardization phase with no definite time limit. Each phase has characteristic prop erties regarding market structure, competitive position, size of enterprise; and each phase makes specific demands on the location factors (see **Figure 3**). The spatial variant of the product cycle hypothesis explicitly includes technological progress in explaining the course of the phases and thus possesses a dynamic component: during the introduction phase, product innovations with their corresponding spatial demands as to manpower, R&D input etc., determine the locational decision. As the aging process progresses, process innovations and – due to declining unit cost yields – labor cost saving procedures and, at the same time, decentralized locations gain in importance. This approach has repeatedly been used to explain international and interregional locational shifts convincingly, although the empirical results are by no means uniform. The product cycle hypothesis elucidates the major part of the relationship between spatial division of labor and technological change by associating the early phases of the cycle – characterized by human capital intensity and product innovations – with the agglomer ations and the late phases – distinguished by wage cost intensity and process innovations – with the periphery. As the cycles of innovative products become steadily shorter, the standardization phase reduces as well or does not set in at all; this precludes the postulated de centralization during later cycle phases. Criticism of the product cycle theory essentially concentrates on deter minism relative to location shifts from the center to the periphery, the product concept, the applicability to all industrial products, and the insufficient explanation for the location decision during the development and introduction phases.

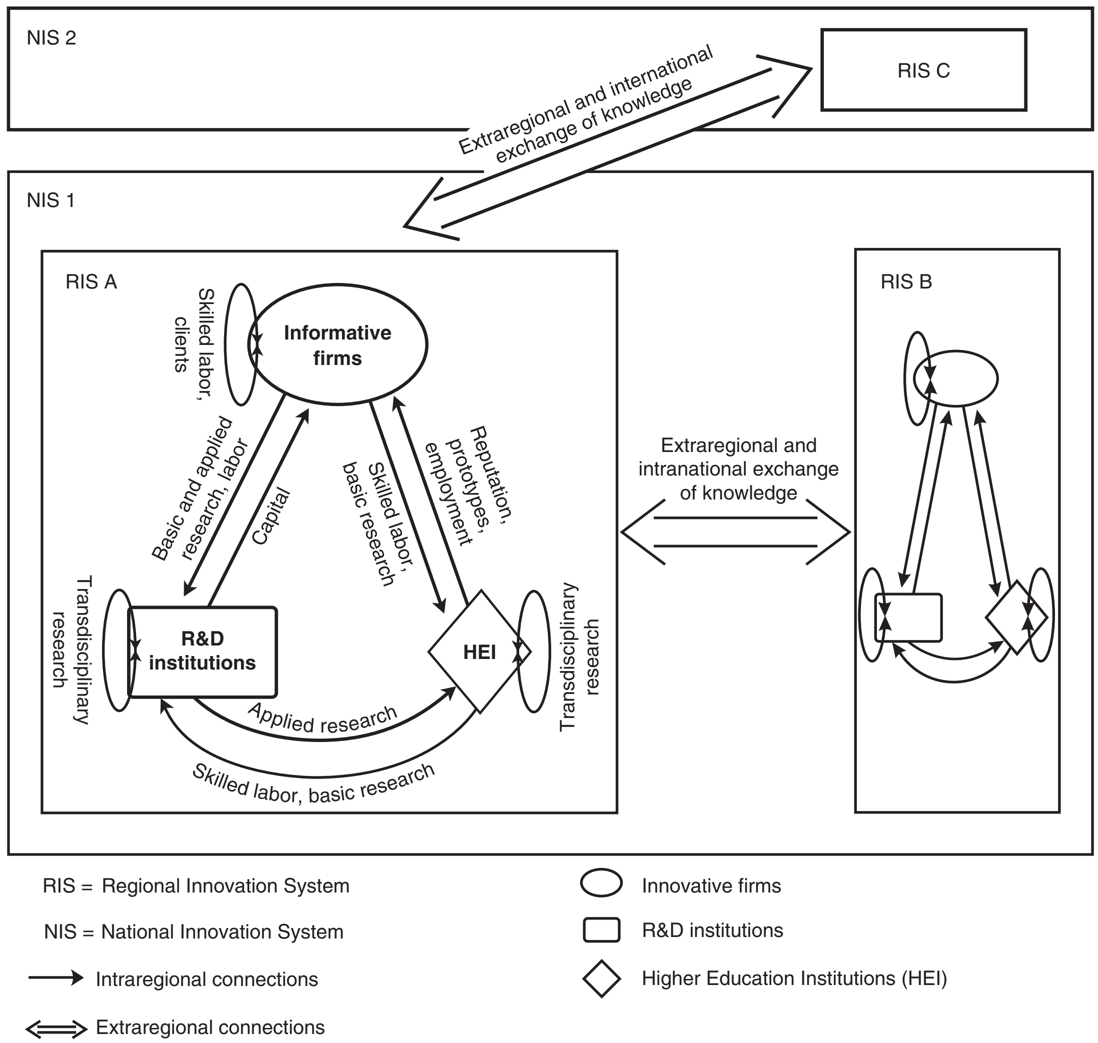

Figure 2 Innovative actors in Regional and National Innovations Systems and its connections.

The theory of the 'long waves of economic development' in the tradition of Schumpeter postulates that long term economic development is subject to cyclical fluctuation. The cycles differ in length (each consisting of economic growth, prosperity, recession, and depression) and are caused by a group of more or less continuous, partially linked basic innovations of individual entrepreneurs in young sectors of technology which, subsequently, cause new industries to emerge. Phases of growth are the expression of prior, phase specific basic innovations and follow a long term cycle. When the old basic innovations are exhausted, meaning that the yield of the corresponding investments dwindles, a phase of economic decline ensues. Since the mid 1970s, economic geographers have voiced an opinion on the 'geography of the Fifth Kondratieff'. They identify regional long waves, that is, regions too develop in cycles dependent on superior economic cycles. In the process, each long wave creates its own 'new' regions which – as the wave recedes – suffer economic recession as well. It is believed that at the root of the growth of young regions induced by long waves is the need of weak capital to make its risky investments in new technologies outside existing centers. Currently, there is disagreement on essential elements of the theory of the long waves, which renders its application to innovative regions difficult. Controversially discussed are the length of the cycles, the innovation concept, and the significance of an enterprise's size and form of organization. Likewise, there is no unanimity regarding the question of whether the phases of the long waves' upward trend are really caused by endogenously generated innovations or rather by such factors as wars and historical coincidences.

The Innovative Milieux Approach (GREMI)

According to the milieu theory bearing the mark of the francophone Groupe de recherche européen sur les

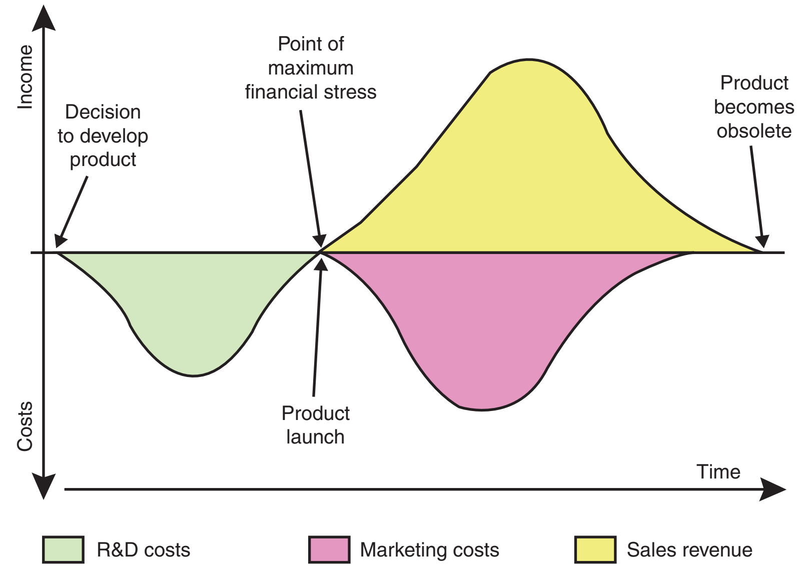

	Invention and innovation	Growth	Maturity	Eclipse
Dominant factor of production	Information, highly qualified labor, capital	Qualified labor	Capital	Unskilled labor
Main location	Agglomerations with a very good R&D infrastructure and advantages in amenities	Intraregional decentralization	Interregional decentralization (peripheral inland areas)	Complete shift of production to low-cost countries

Figure 3 Innovation-driven regional product cycles.

milieux innovateurs (GREMI) group, innovations are the result of a collective, dynamic process set in motion by many individuals and institutions in a region forming a network of synergy producing linkages (see **Figure 4**). Hence, the milieu of a region results from interaction between enterprises, institutions, and labor, all of which, by way of collective, cooperative learning, help reduce the inherent insecurity during a period of changing technological paradigms. This kind of learning from and with one another is especially enhanced by the mobility of labor, by supply linkages, and in the form of face to face contacts which are facilitated by spatial proximity. Small enterprises profit most from integration into regional networks.

GREMI defines the milieu as a regional production system originating from integrated networks, consisting of tangible and intangible resources, characterized by a common, historically matured culture, based on collectively developed and usable, regionally specific levels of knowledge, and founded on a system of social relationships among local players that incorporates competition together with cooperation. The creative or – synonymous with it – the innovative milieu and innovative networks are inextricably linked. The former refers to the regionalized part of an innovative network. Only part of the milieu and its constituent networks are ever innovative at any given time. A milieu need not be restricted to a region. Spatial proximity, due to the spatial determination of numerous elements of the milieu, does however increase the effect of the milieu considerably. These elements include human capital, which is less mobile interregionally than intraregionally, the considerably more intense interrelationships between innovation players in the form of direct, informal contacts, and the common cultural, psychological, and political background of the members of the network. Spatial proximity of the players does not, however, guarantee their interaction.

Milieu theories primarily serve to explain the *genesis* phase of an innovation. The majority of empirical milieu studies is confined to regions characterized by innovative networks. Technological information as a 'raw material', a sufficiently large pool of mobile labor, and the availability of investors willing to finance the risky projects in the field of new technologies have been identified as necessary (though not sufficient) preconditions for the development of an innovative milieu. Criticism of the milieu theory is directed above all at its low level of empirical evidence. Neither the milieu in general nor the innovative milieu are defined sufficiently clearly or at least

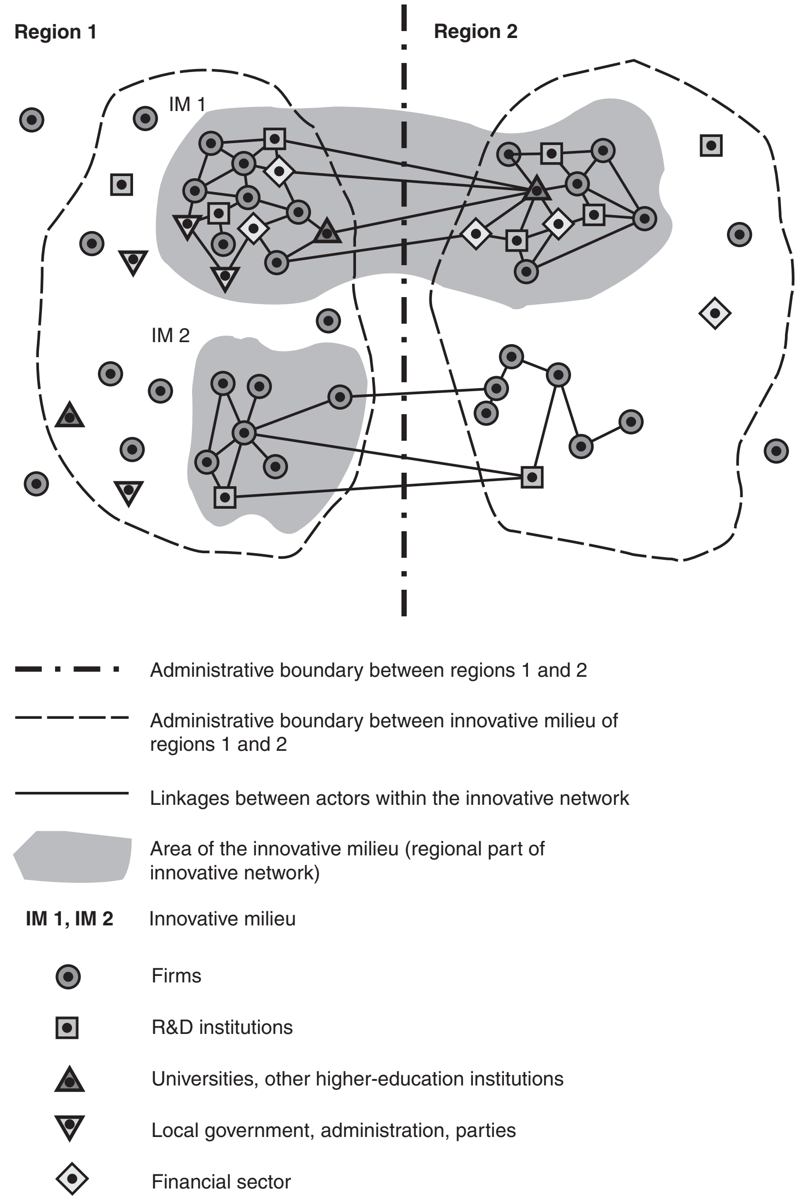

Figure 4 The innovative milieu.

unanimously. Very often the spatial reference remains unclear. The integration of the innovation network idea, which is primarily person oriented, into the region oriented milieu approach does not yet seem fully operational. Since – in addition – the concept is heavily oriented toward small and medium sized enterprises (SMEs) as supporters of the innovations, it obviously neglects several aspects of reality.

Learning Regions and Collective Learning Processes

As the creation of innovations requires the exchange and passing on of tacit knowledge in particular, learning processes are necessary. Individuals who do not yet possess the knowledge necessary for innovations have to have it imparted from those who do possess it; they have to learn. The fact that tacit knowledge in its purest form can only be transferred through face to face contact makes spatial proximity between the teacher and student indispensable. Tacit knowledge does not travel easily – if at all. Spatial proximity alone is not a sufficient condition for learning. Individuals should share some basic commonalities such as the same language, conventions, and norms. Regional embeddedness of both partners supports the probability of such commonalities because trust is facilitated. Former interactions, especially informal ones, help to bring innovative actors (back) together. While other forms of learning also exist, the concept of learning regions emphasizes the relevance of learning by interacting. The more innovative (or potentially innovative) individuals there are available in a region the better – and the greater the probability of intraregional and collective learning processes between them. The more these individuals are socially embedded in the region the more they share the commonalities mentioned above – and the more the learning process takes place within the region (rather than between regions). As a final consequence, scholars claim that learning regions exist in which these processes are optimized for the promotion of the generation and diffusion of innovations. Another argument supports the relevance of intraregional and collective learning processes: to a degree the knowledge and, consequently the resulting innovation itself, are region specific. Each region possesses a region specific capital stock (knowledge) which can be interpreted as a comparative regional advantage. Localized capabilities and intangible assets further support this hypothesis – with the consequence that it is rare for such knowledge and innovations to travel beyond the boundaries of learning regions. From the perspective of the learning region this means that it benefits first from the economic advantages of an innovation. The concept of learning regions has, despite the initially limited level of empirical evidence to date, already been included in numerous policy measures supporting regional innovations.

Empirical Evidence for the Regional Implications of Innovation Processes

Measuring Regional Innovation

Regional innovation research is still searching for the ideal data for measuring innovation. Because innovation has by definition a certain kind of novelty included it is very difficult to use data from secondary, official statistics when measuring the innovativeness of firms or regions. Public industry classifications are always late – that is, they are always lagging behind product and process innovations (e.g., biotechnology). The ideal method would be to measure the innovations themselves as an indicator of output, for example, in terms of the number of new products within a given space and time. However, it is rather difficult to find adequate data that are comparable between regions. Another fundamental issue is what we actually mean by 'new'. Additionally there are significant differences between radical innovations (like the microprocessor) or incremental innovations (from black and white TV to color). In most empirical studies, scholars therefore have to make compromises, some more satisfying, some less so. These include indicators of innovation input (although the intention is to measure output) such as R&D expenditures against turnover, or share of R&D personnel against the total number of personnel in the region. Even patent data, which are frequently used as a parameter, are not an output indicator, but more accurately a throughput indicator, and have certain shortcomings (e.g., distortions as a result of different industries' varying propensities to patent). The availability, including regionally, of such 'auxiliary variables' has improved significantly in many industrialized countries in recent years, however.

The Uneven Geography of Innovation

In times of increasing globalization in particular, the spatial (and sectoral) clustering of innovative activities is a characteristic feature of the economic geographies worldwide. There is even clear empirical evidence that the spatial concentration of such economic activities is greater the more knowledge intensive the economic activity is. In Japan, 41% of all persons employed in high tech industries are concentrated in the Tokyo metropolitan region which is home to 'only' 25% of the country's total population. Empirical studies on the uneven geography of innovation in the US reveals that the strong geographic concentration of new product innovations is the result of a regional technological infrastructure which promotes information transfers and decreases the risks and cost of engaging in innovative activity. The 11 leading states, led by California and New York, account for more than 80% of all product innovations. In France, almost 40% of national employment in high tech industries are to be found in the Ile de France region around Paris. Even in a country such as Germany with a relatively balanced spatial structure, Munich as a high tech region is at the top of the German regions in terms of almost all innovation indicators, in some cases far ahead of the other regions. In developing countries and/or transformation economies spatial concentration is even stronger. For most states there is a considerable consistency between the corresponding regional rankings over time, although the rankings between the individual indicators certainly may vary.

The Spatial Range of Innovation Linkages

Both interregional (especially international) 'and' intraregional linkages are important for firms and regions when innovation is considered. International innovation linkages are necessary to avoid the risk of technological and other lock ins. Intraregional linkages are crucial for the exploitation of local knowledge resources (tacit knowledge). The European Regional Innovation Survey (ERIS) shows that whether intra or interregional cooperation in innovation predominates is dependent upon the type of innovation actor (see **Figure 5**). For public research institutions, intraregional linkages are more common, whereas manufacturing firms attach greater importance to interregional cooperation in innovation. From the perspective of public research institutions, intraregional cooperation with businesses is highly significant; on the other

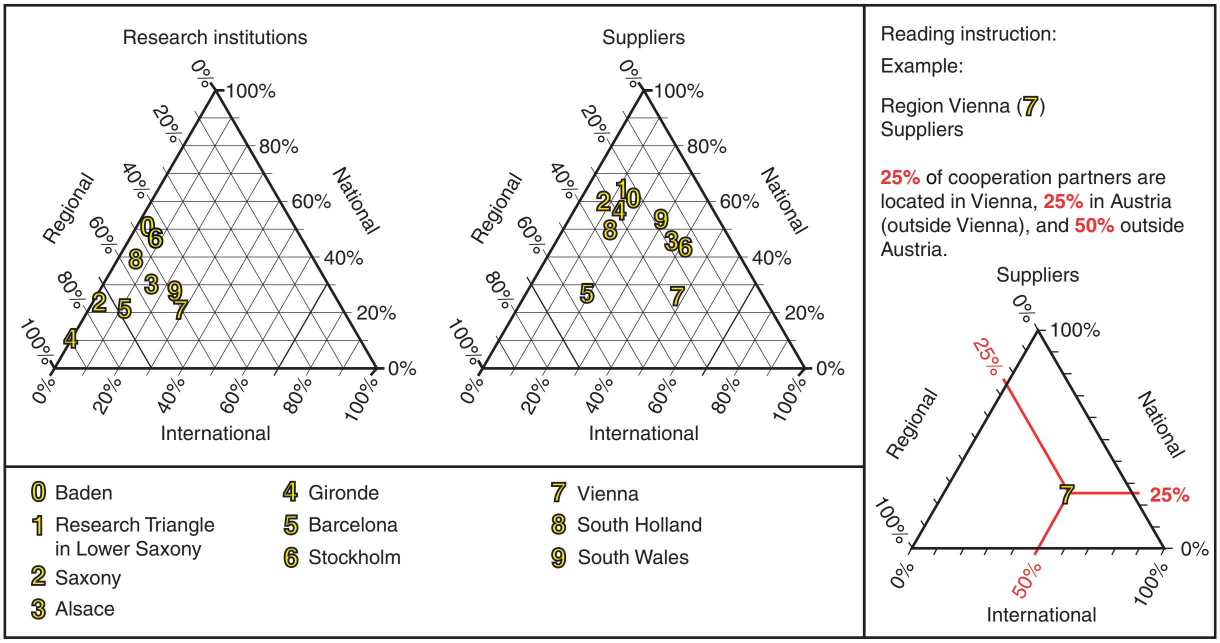

Figure 5 Spatial range of innovative linkages of manufacturing firms Data source: European Regional Innovation Survey (ERIS).

hand, public research institutions cooperate more fre quently with other research facilities (rather than with businesses) across regional boundaries. Innovative firms do tend to place greater stronger emphasis on interregional linkages than less innovative firms.

Another possibility for evaluating the relative role of the various spatial ranges of innovation linkages is the answer to the question of what influences firms' innovation activ ities more: firm characteristics or regional determinants. ERIS data from 12 European regions show that firm spe cific determinants of innovation are more important than either region specific or extraregional factors. However, the distinction between firm level and region level determin ants of innovation is not strictly an 'either/or' matter. Of course, a firm without considerable innovation potential cannot generate notable innovations, even if the innovation conditions in the region where it is located are very fa vorable. However, the regional environment can help firms to realize their existing, but sometimes unexploited, in novation potential. In particular, the R&D environment that the region (and the other firms based there) offers an innovating firm is certainly capable of influencing a firm's innovation behavior, even independently of firm related features. This has an impact on firm behavior as well. The firm and the region, therefore, consist of two groups of determinants of firm innovation activities that influence each other, even if not to the same level of intensity. The regional environment is not an independent determinant of firm innovation activity but is influenced by the charac teristics of local firms. Interfirm innovation cooperation can improve the competitiveness of firms and increase their skills. It appears, however, that although the regional environment and innovation networks strengthen firms' existing development potential, they are not able to gen erate this potential themselves.

Local Geographical Knowledge Spillovers

In view of the at least partially tacit nature of new knowledge, it is more probable that such knowledge tends to spread or be passed on more intraregionally than interregionally. Knowledge spillover as a form of ag glomeration economies can therefore be identified within regions in particular. Empirical studies from US regions in particular show, for example, that patent applicants in manufacturing cite other patents originating in the same city more frequently than they cite patents from other cities. Local geographical knowledge spillovers also exist between university research and firms innovative activity. Empirical results for US regions reveal that the spillovers of university research on innovation extended over a range of 50 miles from the innovating metropolitan area, but not with respect to private R&D. Star scientists at universities can play a key role in this respect in innov ation based regional development and can become an engine of regional growth provided they succeed in generating such intraregional knowledge spillovers.

Regional Innovation Policies

Rationale for/of Regional Innovation Policies

Policy cannot ignore the uneven geography of innovation. It can, however, make use of it to achieve political economic goals and, in doing so, has the fundamental choice between two alternatives. A growth oriented in novation policy would strengthen the strengths, that is, support the already relatively strong innovative agglom erations in order to promote national competitiveness and growth. Regional innovation policy aimed at achieving a balance, on the other hand, would try to support the in novation potential of (economically) peripheral regions with public investment in R&D infrastructure to help them catch up with already strong regions. While regional innovation policies aimed at achieving a balance domin ated for a long time in many countries, a change in thinking in many market economies now seems to be underway toward growth oriented regional innovation policies. Innovation oriented regional policies are in creasingly being replaced by regionalized innovation policies, particularly in member states of the European Union. Programs such as the 'bioregio' program in Germany, for example, which was set up as a competition between regions, tend to support regions which are al ready innovative, as they are in the best position to emerge victorious from such competitively structured programs. Even programs that are actually intended to reduce innovation disparities between regions, such as the Japanese 'Technopolis' program in all of the 47 pre fectures, have shown that the most successful regions are those which enjoy relative proximity to the prospering region around the capital city, Tokyo.

Why Regional Innovation Policies Are Potentially Relevant for Firm's Innovation Behavior

The reason for innovation policy is generally market failure: the risks and uncertainties associated with in novations mean that entrepreneurs would innovate less than is necessary for economic purposes if politics did not intervene. This means incentives are needed. This still leaves the difficulty of selecting the right future oriented technologies and industries that have the potential to grow in future (and are therefore worth supporting). This consideration can also be applied to regional innovation policies: regions where the level of innovation is lower than preferable from an overall economic perspective can be supported with policy instruments. The alternatives described under the section 'Rationale for/of regional innovation policies' then come into play when considering the question of which regions are to be supported and how intensively (and by whom: the regions themselves or central government, see **Figure 6**).

As for innovation networks, empirical research reveals that public innovation policies should promote all types of innovation linkages, because a composite of coopera tive intraregional and interregional linkages promises to have particularly positive effects on the innovation out put. However, local political capacities to promote international linkages are limited. Local and regional

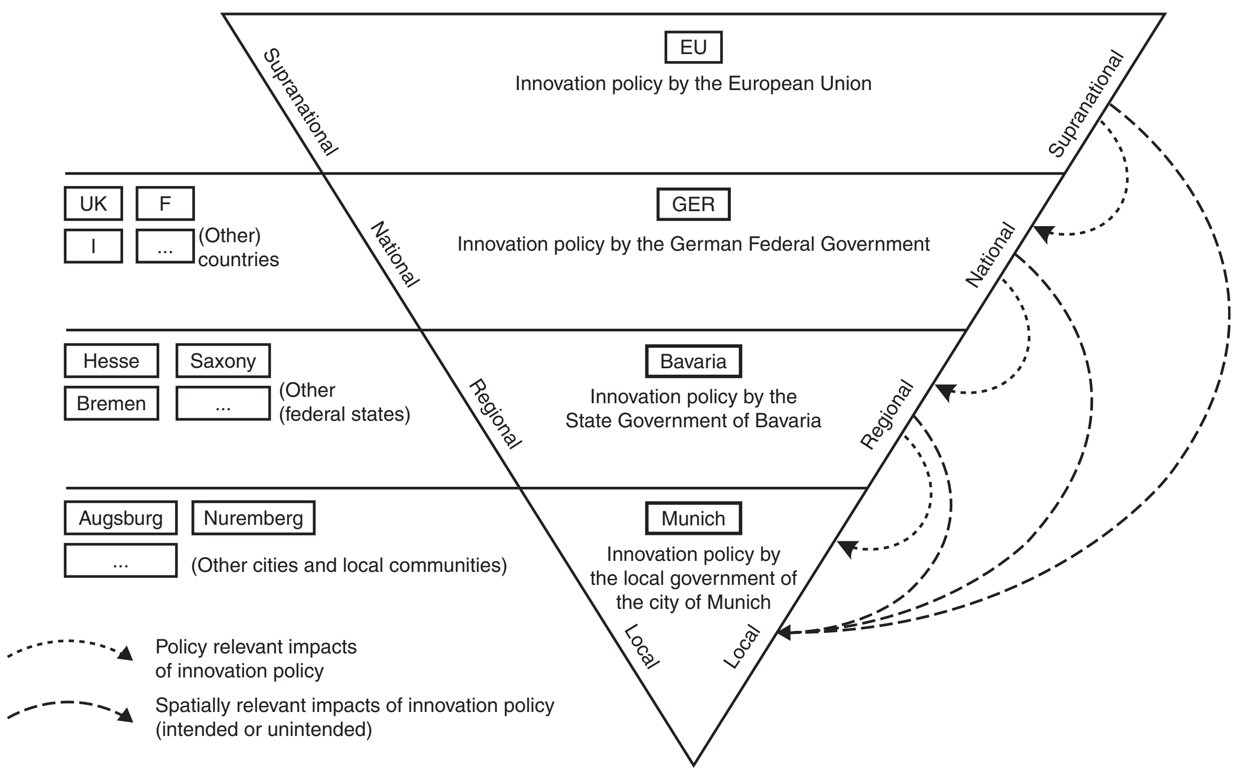

Figure 6 Spatial scales of innovation policies (example of Munich/Germany).

innovation policy must be aimed primarily at the innovating firms. The regional environment does indeed influence the scope and quality of operational innovation activities. This influence depends, however, to a considerable extent on the characteristics of the firms and their entrepreneurs. The firm independent, autonomous influence of the regional environment must not be overestimated. This does not mean, however, that concepts that put the region as a whole in the spotlight are obsolete. The idea of the 'learning region' and the concept of 'regional innovation systems' are complementary and offer concrete recommendations for political action in the regions. It should, however, not be forgotten that a region represents an ensemble of individual innovation players (mostly firms). These firms do indeed consciously or unconsciously take advantage of the regional environment. However, the general focus of their activities is not primarily on the economic welfare of the region but on their own economic welfare. Political programs intended to increase the innovativeness of regions must therefore always begin with the innovation activities of local firms.

Targets and Instruments

Most regions try to achieve two complementary goals with their innovation policy: to improve their regional innovation infrastructures and to promote interaction between innovative actors. The aim is to improve the realization of the value of local resources for economic purposes, in other words to generate more innovations out of existing infrastructure and personal resources and, from innovations, to generate more innovations. Finally, the knowledge transfer from knowledge providers (e.g., universities, public research institutions) to knowledge recipients (private business) needs to be intensified. Potentially this is easier at an intraregional level than interregionally, but many barriers between business and academia hinder this process. The most popular policy instruments include business incubators, science parks, technology transfer agencies, start up programs (especially dedicated to new technology intensive firms). All of these instruments and related strategies need a great deal of time before they (may) become successful. Empirical studies of high tech regions such as Silicon Valley (USA), Greater Boston (USA), the 'western crescent' to the west of London (UK), Munich (Germany), and Shanghai (China) have demonstrated, however, that policy matters. However, this policy dividend is often not the result of explicit regional innovation policy, but far rather of implicit innovation policy (without any regional goals, but regional effects). The effects of military R&D expenditure by Department of Defense and NASA in the US in the 1950s on the genesis of Silicon Valley are clear proof of this.

See also: Agglomeration; Diffusion; Embeddedness; Knowledge Economy; Regional Innovation Systems; Technology and Regional Development.

Further Reading

Asheim, B. and Gertler, M. (2005). The geography of innovation: Regional innovation systems. In Fagerberg, J., Mowery, D. C. & Nelson, R. R. (eds.) *The Oxford handbook of innovation* (1st ed). pp 291 317. Oxford: Oxford University Press.

Camagni, R. (ed.) (1991). *Innovation networks: Spatial perspectives*. London: Belhaven Press.

Castells, M. (1999). *The information age: Economy, society and culture*. Oxford: Blackwell.

Cohen, W. and Levinthal, D. A. (1991). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly* 35, 128 152.

Feldman, M. (1994). *Geography of innovation*. Boston: Kluwer Academic Publishers.

Hagerstrand, T. (1967). *Innovation diffusion as a spatial process*. Chicago: University of Chicago Press.

Jaffe, A. B., Traijtenberg, M. and Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics* 108, 577 598.

Lambooy, J. (2005). Innovation and knowledge: Theory and regional policy. *European Planning Studies* 13, 1137 1152.

Malecki, E. J. (1991). *Technology and economic development: The dynamics of local, regional, and national change*. New York: Wiley.

Malmberg, A. and Maskell, P. (2002). The elusive concept of localization economies: Towards a knowledge based theory of spatial clustering. *Environment and Planning A* 34, 429 449.

Morgan, K. (1997). The learning region: Institutions, innovation and regional renewal. *Regional Studies* 31, 329 343.

Ratti, R., Bramanti, A. and Gordon, R. (eds.) (1997). *The dynamics of innovative regions: The GREMI approach*. Aldershot: Ashgate.

Simmie, J. (2005). Innovation and space. *Regional Studies* 39, 789 804.

Sternberg, R. (1996). Government R&D expenditure and space: Empirical evidence from five advanced industrial economies. *Research Policy* 25, 741 758.

Sternberg, R. (1997). New industrial spaces and national technology policies The case of Kyushu and the Japanese 'Technopolis' Strategy. In Simmie, J. (ed.) *Innovation, networks and learning regions*?, pp 159 177 (Regional Policy and Development 18). London: Jessica Kingsley Publishers.

Sternberg, R. (1998). Technologiepolitik und High Tech Regionen ein internationaler Vergleich (Technology policies and high tech regions an international comparison), 2nd edn (Wirtschaftsgeographie, Vol. 7). Munster, Hamburg: Lit.

Sternberg, R. and Arndt, O. (2001). The firm or the region: What determines the innovation of European firms? *Economic Geography* 77, 364 382.

Sternberg, R. and Tamásy, C. (1999). Munich as Germany's No.1 high technology region Empirical evidence, theoretical explanations and the role of small firm/large firm relationships. *Regional Studies* 33, 367 377.

Relevant Websites

www.druid.dk
Danish Research Unit for Industrial Dynamics.

http://cordis.europa.eu/innovation policy/studies/
European Commission Cordis.

Input–Output Analysis

G. J. D. Hewings, University of Illinois, Urbana, IL, USA
M. Sonis, University of Illinois, Urbana, IL, USA, and Bar Ilan University, Ramat Gan, Israel

Glossary

Backward and Forward Linkages The purchases and sales relationships for a sector.
Computable General Equilibrium More extensive models, incorporating both supply and demand responses and in which input–output systems are embedded.
Economic Landscape Presentation of the Leontief matrix in a hierarchical form.
Interindustry Transactions Sales and purchases relationships between sectors.
Key Sectors Sectors whose change in production would generate an above-average impact on the economy.
Leontief Inverse Matrix The matrix showing the direct and indirect relationships between all sectors.
Multiplier The ripple effect on the whole economy generated by a change in one or more sectors.

Input–Output Analysis

All economies are characterized by a degree of interdependence between firms, households, and government; the input–output system of accounts provides a framework for measuring the strength of these linkages. From these accounting systems, an input–output model can be developed to trace the way in which changes in one part of the system affect the whole economy. Most recent developments have placed the input–output accounts within computable general equilibrium and other more integrated models that enable consideration of both supply and demand changes within the economy. Further research has explored dynamic versions of these models as well as their presentation within multiregional frameworks to capture interdependencies over space.

Introduction

The progenitor of input–output analysis, Wassily Leontief, was able to translate Leon Walras' ideas for a general equilibrium framework for the economy into a formal model. One of the fundamental questions that needs to be answered about an economy centers on its structure. However, there is very little consensus about what is meant by the term structure and how it might be interpreted. Many analysts resort to descriptions based on sectoral shares of total output, drawing on the ideas of Kuznets, Williamson, and others. However, these descriptions provide little guidance to the nature and strength of the interactions between these sectors. Are they connected, in what ways, and how strong are these connections? The input–output table and associated model provide one important set of economic data that can contribute some answers to these questions.

Essentially, the purchases and sales relationships of all establishments (referred to as interindustry transactions) are aggregated into broad sectors based on the nature of the commodities produced in the plants or the services provided (for nonphysical commodities). These transactions are then arranged in a matrix, as shown in **Table 1**. The sales made by each sector to the other sectors are shown in the rows; the purchases made by each sector can be examined from inspection of the columns. In addition to transactions between sectors themselves, sales are also made to a category referred to as final demand; included here would be households, governments, and exports outside the nation or the region being considered. Similarly, to produce their products or provide services, the sectors employ labor and capital as well as import some of the needed items. This double entry accounting system is balanced in that total sales are equal to total purchases for each sector; in the example shown, all sectors have the same level of production but this would not be the case in a real economy.

Table 1 may be considered as the set of accounts for an economy; using this base, it is now possible to create a model of this economy. First, it is assumed that the interindustry transactions are endogenous, while (initially at least) final demand is exogenous. Second, the recipe by which each sector makes its products or services is assumed to be fixed over all levels of production. We assume that each sector produces goods and services according to a fixed recipe (formally known as a production function); hence, to produce \$1 million worth of steel, we assume a fixed proportion will be allocated to inputs such as limestone, iron ore, scrap, energy, coke, labor, and so forth. Inputs are expressed in monetary terms since it would be difficult to combine tons of iron ore with megawatts of electricity, or hours of labor in some consistent fashion. This fixed recipe enables us to express the transactions in proportional form, known as direct coefficients; these are shown in **Table 2**. We used a simple case in which each sector's output was \$100 million; to obtain the direct coefficients, we divide the

Table 1 The input output table

		Interindustry transactions					Final demand				
							Households	Government	Exports	Total final demand	Total sales
	Sector	1	2	3	4	5	6	7	8	9	10
1	1	21	0	9	3	0	30	15	22	67	100
2	2	1	8	7	29	0	25	7	23	55	100
3	3	3	20	0	50	7	5	9	6	20	100
4	4	31	2	38	0	3	12	13	1	26	100
5	5	10	25	26	1	4	9	19	6	34	100
6	Total intermediate	66	55	80	83	14	81	63			
7	Value added	20	40	10	17	40	2	62			
8	Imports	14	5	10	0	46	77	55			
9	Total inputs	100	100	100	100	100	160	180			

Table 2 The recipe for production by sector

Sector		1	2	3	4	5
	1	0.21	0.00	0.09	0.03	0.00
	2	0.01	0.08	0.07	0.29	0.00
	3	0.03	0.20	0.00	0.50	0.07
	4	0.31	0.02	0.38	0.00	0.03
	5	0.10	0.25	0.26	0.01	0.04

entries in the 5×5 part of **Table 1** by 100. We further assume that these proportions are invariant with levels of production (i.e., no economies of scale – see **Figure 1**) and that substitution is not permitted.

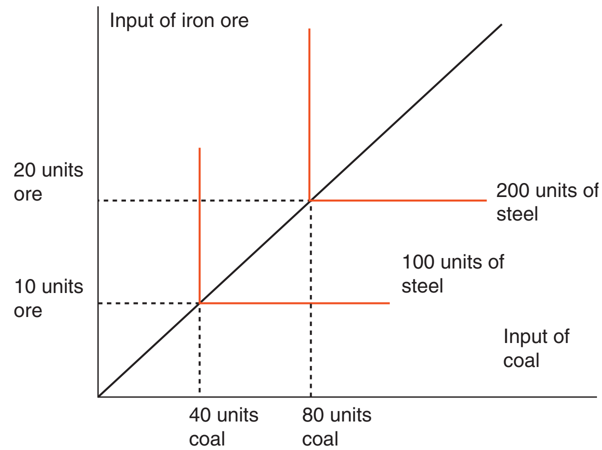

Figure 1 Input combinations are fixed over all levels of production.

The Input–Output Model

Given the structure and assumptions described above, we are now ready to develop the input–output model. Let T describe the 5×5 transactions table from **Table 1**; let x be a 1×5 vector of total production levels by sector and let f be a 1×5 vector of total final demands (column 9 in **Table 1**). The row balance for the table may be pre sented as:

$$T + f = x \qquad [1]$$

Now define A as a 5×5 matrix of input coefficients (essentially the entries in **Table 2**) thus providing the relationship:

$$T = Ax \qquad [2]$$

Substituting eqn [2] into [1], we have

$$Ax + f = x \qquad [3]$$

Further manipulation yields:

$$\begin{aligned} x \quad Ax &= f \\ (I \quad A)x &= f \\ x &= (I \quad A)^{-1} f \end{aligned} \qquad [4]$$

The last entry in eqn [4] may be read as follows: total production is equal to final demand premultiplied by a (multiplier) matrix that essentially captures the direct and indirect linkages of each sector to every other sector. It is referred to as the Leontief inverse matrix and is shown in **Table 3**. The entries reveal the direct and in direct impacts on a sector when final demand in the sector at the top of the column changes by \$1 (or \$1 million or \$100 million). From **Table 2**, sector 3 pur chased \$0.09 for every \$1 of production from sector 1; in **Table 3**, the corresponding entry is \$0.18. When sector 3 buys goods from sector 1, sector 1's production increases and this, in turn, generates additional demands on other sectors and eventually some of these sectors will require further inputs from sector 1. The difference in the cor responding entries in **Tables 2** and **3** are the indirect effects. At the bottom of the multiplier table there is a row labeled multiplier; note that these values vary from 1.37 (sector 5) to 3.23 (sector 4). These entries provide information on the impact on the rest of the economy (including the sector in question) of a unit change in final demand in any sector. The value 2.98 for sector 3 tells us

Table 3 The Leontief inverse matrix

Sector		1	2	3	4	5
	1	1.33	0.05	0.18	0.15	0.02
	2	0.23	1.17	0.30	0.50	0.04
	3	0.40	0.36	1.41	0.82	0.13
	4	0.58	0.19	0.61	1.38	0.09
	5	0.31	0.41	0.48	0.38	1.09
Multiplier		2.85	2.18	2.98	3.23	1.37

that for every increase of $1 in final demand in that sector an additional 1.98 worth of activity is generated for a total value of production of 2.98. Why do these values vary? They reflect the degree to which a sector is dependent on other sectors in the economy for its inputs and as a source of consumption for its products and they also depend on the structure of production (the recipe). Multipliers not only vary by sector they also vary by economy; small, regional economies that are very open would tend to have smaller multipliers since they may be more dependent on input sources outside the region and also sell a significant portion of their production to de manders outside the region.

Using the Input–Output Model

In the Leontief inverse matrix, each sector is directly and indirectly linked with all other sectors. For each sector i, these linkages are divided into two different types: 'backward linkages' describing the direct and indirect economic inputs of other sectors into sector i, and 'for ward linkages' describing the direct and indirect eco nomic inputs of sector i into all other sectors. The backward linkages of sector i are associated with the column i and its forward linkages are associated with the row i of the Leontief inverse. Therefore, the overall intensity of backward and forward linkages of sector i can be measured as sums of the components of the ith column and the ith row of the Leontief inverse. Thus, the column and row multipliers of the Leontief matrix $B = ||b_{ij}||$ can be defined as:

$$B_{\bullet i} = \sum_{s\;\,1}^{n} b_{si};\, B_{i\bullet} = \sum_{s\;\,1}^{n} b_{is} \qquad [5]$$

In vector notation, the row vector of column multi pliers and the column vector of row multipliers may be presented as follows:

$$M_{\bullet c} = \begin{pmatrix} B_{\bullet 1} & B_{\bullet 2} & \ldots & B_{\bullet n} \end{pmatrix};\; M_{r\bullet} = \begin{pmatrix} B_{1\bullet} \\ B_{2\bullet} \\ \vdots \\ B_{n\bullet} \end{pmatrix} \qquad [6]$$

The global intensity of the backward and forward linkages of the Leontief inverse is the sum of all column multipliers or the sum of all row multipliers:

$$V = \sum_{j\;\,1}^{\mathrm{n}} B_{\bullet j} = \sum_{i\;\,1}^{n} B_{i\bullet} = \sum_{i,j\;\,1}^{n} b_{ij} \qquad [7]$$

and the average multiplier is equal to V/n.

These characterizations can be used to identify key sectors, those that create an above average impact on the rest of economy. Two main types of indices have been proposed:

1. *Power of dispersion* for the backward linkages, measur ing the degree to which a change in the sector of reference is greater than the average for all sectors:

$$BL_j = \frac{1}{n}\sum_{i\;\,1}^{n} b_{ij} \Big/ \frac{1}{n^2}\sum_{i,j\;\,1}^{n} b_{ij} = n\;\, B_{\bullet j}/V \qquad [8]$$

2. *Sensitivity of dispersion* for forward linkages, measuring the degree to which a unit change in all sectors' final demand would create an above average increase in output in the sector of reference:

$$FL_i = \frac{1}{n}\sum_{j\;\,1}^{n} b_{ij} \Big/ \frac{1}{n^2}\sum_{i,j\;\,1}^{n} b_{ij} = B_{i\bullet} \Big/ \frac{1}{n} V \qquad [9]$$

Moreover, the sum of all backward linkages and the sum of all forward linkages are equal to n, and the average linkage is equal to 1.

Table 4 presents the Leontief inverse, column and row multipliers and backward and forward linkages cal culated for the regional input–output system. Arranging the row vector of column multipliers according to their size, we have the row vector of the rank size hierarchy of column multipliers where the roman figures $I, \ldots, V$ de note the ranking of these multipliers by size. The cor responding rank size hierarchy of backward linkages is also shown in **Table 4**. A similar organization of the row multipliers can be presented in the same table. The usual interpretation is that a backward linkage index $BL_j > 1$ indicates that the corresponding column multiplier $B_{\bullet j}$ is larger than the average column multiplier and, therefore, a unit change in final demand in sector j will create an above average increase in the economy. Similarly, for the forward linkage index $FL_i > 1$, the corresponding row multiplier FL_i is larger than the average row multiplier, and, thus, a unit change in all sectors' final demand (in cluding sector i) would create an above average increase in sector i.

Using these hierarchies of backward and forward linkages, we can divide sectors into four groups: (1) sector i is considered a 'key sector' if $BL_i > 1$ and $FL_i > 1$; (2) sector i is considered as 'a backward linkages oriented sector' if $BL_i > 1$ and $FL_i < 1$; (3) sector i is defined as 'a forward linkages oriented sector' if $BL_i < 1$ and $FL_i > 1$;

Table 4 Linkages and key sectors

Sectors	*Leontief inverse* 1	2	3	4	5	*Row multipliers*	*Forward linkages*	*Rank-size hierarchy of forward linkages*
1	1.33	0.05	0.18	0.15	0.02	1.73	0.69	V
2	0.23	1.17	0.30	0.50	0.04	2.24	0.89	IV
3	0.40	0.36	1.41	0.82	0.13	3.12	1.24	I
4	0.58	0.19	0.61	1.38	0.09	2.85	1.13	II
5	0.31	0.41	0.48	0.38	1.09	2.67	1.05	III
Column multipliers	2.85	2.18	2.98	3.23	1.37	*V* 12.61		
Backward linkages	1.13	0.86	1.18	1.28	0.55			
Rank-size hierarchy of backward linkages	III	IV	II	I	V			

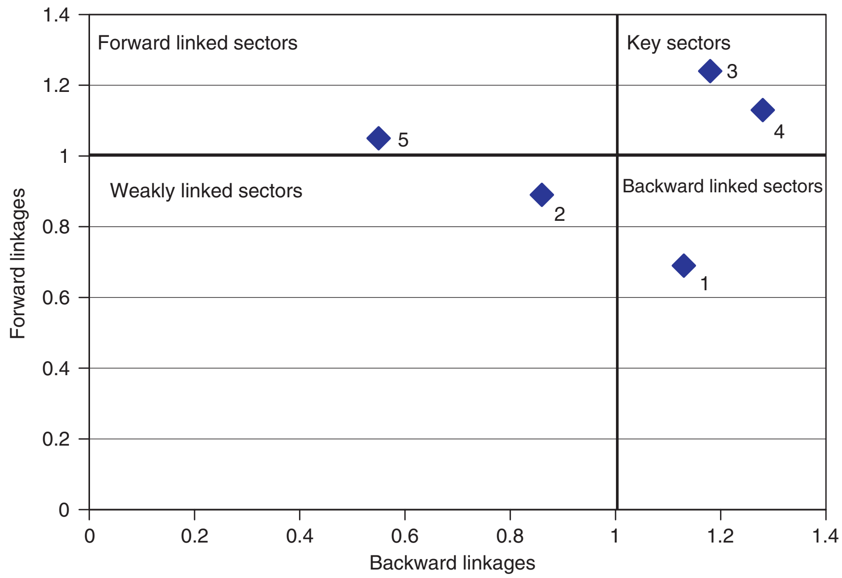

Figure 2 Classification of sectors.

and (4) sector i with both backward and forward linkages <1, is considered as a 'weak linkages oriented sector'.

Figure 2 presents a graphical illustration of the above concepts, where each sector i is represented by the point i with coordinates (BL_i, FL_i). There are two key sectors (3 and 4), the forward oriented sector (5), a backward oriented sector (1), and a weakly oriented sector (2).

It should be noted that the above classification of the economic sectors can be useful for the characterization of economic structure and structural change of the input–output system. For example, how does the classification of sectors in this economy compare with those for other economies? To what degree can these differences be explained by differences in resource endowments, different growth paths, or varying dependencies on internal and external markets? Furthermore, the classification, applied at different points in time for the same economy, will provide one perspective on the evolution of that economy over time.

Multiplier Product Matrix and Structural Economic Landscape

This section provides a new framework for key sector analysis based on the introduction of a new theoretical tool, the 'multiplier product matrix' (MPM). MPM is derived from the products of the corresponding column and row multipliers of the Leontief inverse and provides a presentation of key sector analysis of backward and forward linkages in the visual form of a macroeconomic photograph (economic landscape) of the structure of an economy, providing a basis for the comparison of structures of different economies or the same economy over time.

The components m_{ij} of the multiplier product matrix $M = ||m_{ij}||$ have the form:

$$m_{ij} = \frac{B_{i\bullet} B_{\bullet j}}{V} \qquad [10]$$

It is important to underline that the column and row multipliers of the MPM are the same as those of the Leontief inverse:

$$\sum_{j=1}^{n} m_{ij} = \sum_{j=1}^{n} \frac{B_{i\bullet} B_{\bullet j}}{V} = B_{i\bullet}; \quad \sum_{i=1}^{n} m_{ij} = \sum_{i=1}^{n} \frac{B_{i\bullet} B_{\bullet j}}{V} = B_{\bullet j} \qquad [11]$$

Essentially the components of the MPM are arranged in a hierarchical fashion from the northwest to the southeast corner of **Figure 3**. This ordering also preserves the rank size hierarchy of sectors developed in **Table 4**.

Essentially, MPM is presented as a three dimensional picture of the economy; by careful manipulation of the row and column orderings, it is possible to directly compare the economic structure of several economies. By preserving the ordering of the rows and columns developed over one time period, it is possible to visualize how the structure has changed over time. It is also possible to compare two economies. **Figure 4** compares the structure of two large Chinese metropolitan economies – Shanghai and Beijing. Using the former as the base, it is possible to see how Beijing differs – the economy is much more homogeneous than Shanghai and the rank orderings of the importance of transactions vary.

Input–Output Analysis and Structural Change

Leontief suggested four basic concepts of structural analysis that could be revealed from an examination of input–output tables: dependence and independence, hierarchy, and circularity. Dependence and independence are determined by the presence or absence of interindustry transactions. The more cells that are filled, that is, the more interindustry transactions in a system, the more dependent the industries are on each other. A case of complete dependence occurs when all the cells are greater than zero, generating a system of mutual direct dependence. In this case, an increase in the final demand of any sector will generate interindustry transactions that will affect, directly and indirectly, the output of all other sectors without exception. The chain of effects will be circular, in a sense that the sector that provokes the initial increase will receive back, indirectly, some response to the stimulus that it produces. However, in most cases, the system is characterized by one in which not only some but the majority of the cells are empty. One can rearrange the rows and columns of the original table in an intentionally predefined way, in order to identify the hierarchical pattern of the system. In general, this procedure leads to a triangulated pattern that represents, analytically, a system where the effects of changes in the final demand spread from the upper sectors to the lower ones, whereas the reactions in the opposite direction are very limited. In such a system, a hierarchy is

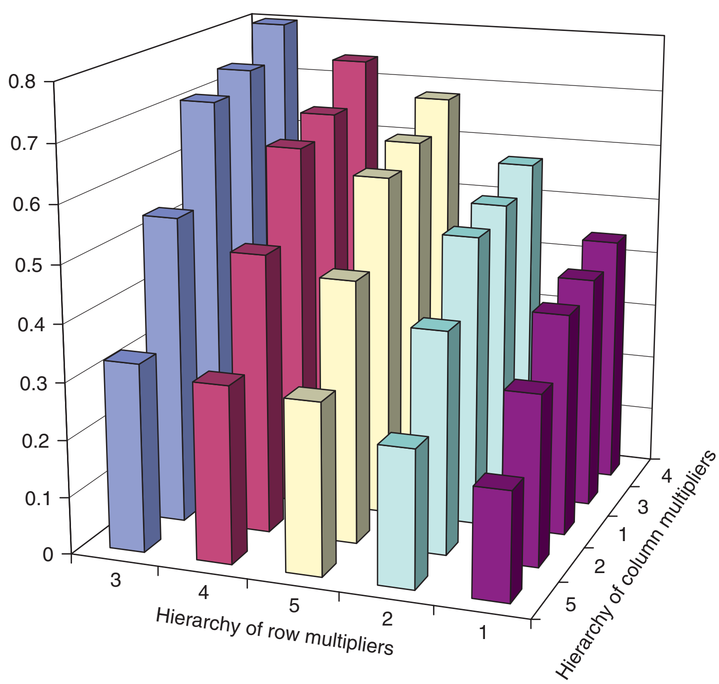

Figure 3 The economic structure interpreted by the multiplier product matrix.

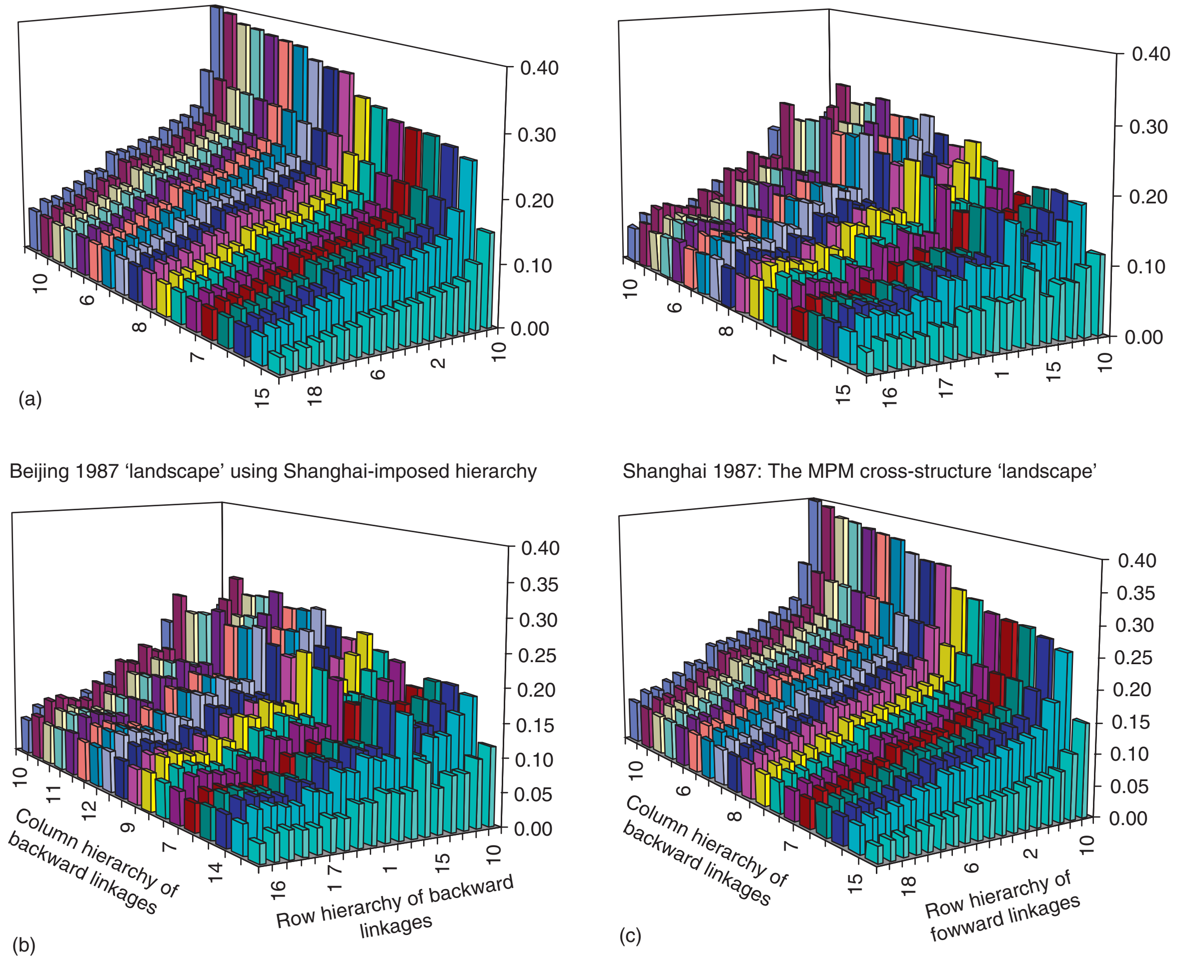

Figure 4 Comparison of two economies, Shanghai and Beijing (using Shanghai's hierarchy), 1987.

established, where the upper sectors are buyers while the lower ones can be characterized as suppliers.

Researchers have used the triangulation process to identify those properties and to compare the internal structure of the interindustry transactions of the US and Western Europe economies. More recent analysis has used the input–output system to unravel the contributions that changes in technology and changes in demand have made to changes in production in economies. In developed economies, it would appear that changes in demand – both in terms of level of demand and allocation across sectors – have been more important than changes in technology. In developing economies, the latter component is much more important, as these economies embrace newer, more sophisticated technology. However, even in these economies, changes in demand are still extremely important.

Even though input–output tables have been produced for a large number of national and regional economies, there has been relatively little work exploring the nature of similarity or differences in structure across these economies. Some researchers have proposed methods to slice the transactions in a hierarchical way to reveal the structure of different levels of interaction; in some very open economies, a small number of slices account for a large percentage of the transactions. This work has also been linked to external linkages, exploring the role of trade in the transformation of economic structure.

Applications of Input–Output Models

The range of applications of input–output models is extensive. They have been used to help economies recover from serious shortages (e.g., the post World War II recovery in the US) and to assist in the articulation of development strategies to simple exercises designed to measure the contribution of a new firm to an economy or the loss of an existing firm. Much of the focus of these exercises centers on the estimation of the multiplier effects; in challenging fiscal conditions, input–output models have often been used to estimate the tax implications of alternative strategies or the contribution that might be made through the reallocation of resources from one program to another. During the 1970s and 1980s when there was a great deal of attention directed to regional development strategy based on growth centers, input–output models were in great demand. More re cently, the models have been employed in the evaluation of competitiveness studies and their associated cluster based initiatives. These analyses move beyond simple consideration of linkages and agglomeration to consider the strategic contributions of supply chains in promoting economic development of targeted industries.

Input–Output in More Extensive Systems

Input–output components are now more often nested within broader modeling frameworks; these would in clude computable general equilibrium models (CGE), econometric input–output models, and linear program ming models. These developments have preserved the advantages of the input–output framework (e.g., the focus on interdependency) while introducing more flexibility into the way in which the production system is modeled. In CGE models, substitution is allowed, scale economies can be introduced, and consideration given to different market structures (imperfect competition or monopol istic competition). Further, some attempts have been made to integrate the developments in the new economic geography with spatial general equilibrium models; one avenue has explored this in an interregional context in which flows of goods between regions are moved over a transportation network rather than in a topology of re gion links that does not differentiate distance, congestion, or capacity constraints.

The econometric input–output framework provides the capability to combine the sectoral detail of the input–output structure within an econometric system, thus providing both impact analysis and forecasting capabil ities. These models have also been used to examine the impacts of alternative visions of the future of the econ omy. As with all forecasting systems, considerable effort is placed on forecasts of the exogenous components of the model but, equally, attention is also directed to the mechanisms of structural change – including changes in technology, the quality, quantity, and variety of goods or services produced and, at the regional level, the location of inputs and markets. These models have been used to examine issues such as the hollowing out of regional and national economies (decrease in dependence on local sources of supply and markets) and the fragmentation of production systems to exploit scale economies as a result of decreases in transportation costs.

A further development is in the integration of input–output systems into extensive models to address the ana lysis of unexpected events, such as climate incidents or earthquakes. The challenges here are often to link the essentially macroeconomic character of the input–output framework with networked structures in which the time dimension assumes a critical role. Recovery from earth quake disruptions, major floods, or other catastrophes often requires careful consideration of the optimal allo cation of resources to ensure restoration of the economy as quickly as possible. The inherent advantage of the input–output system in charting the links between sec tors proves to be of considerable value; recent develop ments have transformed these models from annual to quarterly or even continuous time formulations to make the linkage with continuous time network systems (such as electricity or water supplies) more realistic and valuable.

Development in computer technology now offers the possibility of developing models with hundreds or even thousands of equations. One area under development is the presentation of results in more graphical formats and the use of geographical information systems to provide attractive, flexible formats for data retrieval and specifi cation of the spatial structure of the models.

See also: Industrialization; Quantitative Data; Quantitative Methodologies.

Further Reading

Bacharach, M. (1970). *Biproportional Matrices and Input Output Change*. Cambridge: Cambridge University Press.
Bergman, E. M. and Feser, E. J. (nd). *Industrial and Regional Clusters: Concepts and Comparative Applications*. http://www.rri.wvu.edu/WebBook/Bergman Feser/contents.htm (ac cessed June 2008).
Brocker, J. (1998). Operational spatial computable general equilibrium modeling. *Annals of Regional Science* 32, 367 387.
Cooper, R. J., Donaghy, K. P. and Hewings, G. J. D. (eds.) (2007). *Globalization and Regional Economic Modeling*. Heidelberg: Springer.
Dewhurst, J. J. L. l., Hewings, G. J. D. and Jensen, R. C. (eds.) (1991). *Regional Input Output Modelling: New Developments and Interpretations*. Aldershot: Avebury.
Doi, M. (ed.) (2006). *Computable General Equilibrium Approaches in Urban and Regional Policy Studies*. Hackensack, NJ: World Scientific.
Hewings, G. J. D. and Madden, M. (eds.) (1995). *Social and Demographic Accounting*. Cambridge: Cambridge University Press.
Hewings, G. J. D., Sonis, M. and Boyce, D. (eds.) (2002). *Trade, Networks, and Hierarchies*. Heidelberg: Springer.
Hewings, G. J. D., Sonis, M., Madden, M. and Kimura, Y. (eds.) (1999). *Understanding and Interpreting Economic Structure*. Heidelberg: Springer.
Leontief, W. W. (1986). *Input Output Economics* (2nd edn.). New York: Oxford University Press.
Miller, R. E. and Blair, P. D. (1985). *Input Output Analysis: Foundations and Extensions*. Upper Saddle River, NJ: Prentice Hall.
Miller, R. E., Polenske, K. R. and Rose, A. Z. (eds.) (1999). *Frontiers of Input Output Analysis*. New York: Oxford University Press.
Morishima, M. (2002). *Equilibrium, Stability and Growth A Multi Sectoral Analysis* (2nd edn.). Oxford: Clarendon Press.

Okamoto, N. and Ihara, T. (2005). *Spatial Structure and Regional Development in China: Interregional Input Output Approach*. Tokyo: Institute for Developing Economies.

Roy, J. (2004). *Spatial Interaction Modeling*. Heidelberg: Springer.

Ten Raa, T. (2006). *The Economics of Input Output Analysis*. Cambridge: Cambridge University Press.

Treyz, G., Rickman, D. and Shao, G. (1992). The REMI economic demographic forecasting and simulation model. *International Regional Science Review* 14, 221 254.

Web Book of Regional Science, *Industrial and Regional Clusters*; *Regional Impact Models*; *Keystone Sector Identification* http://www.rri.wvu.edu/regscweb.htm (accessed June 2008).

Institutionalism/Institutional Geographies

D. Mackinnon, University of Aberdeen, Aberdeen, UK

Glossary

Embeddedness A concept which highlights how particular forms of economic activity are grounded or rooted in particular places.

Governmentality An approach, derived from Foucault's later work, that focuses on the techniques and rationalities (mentalities) which governments deploy to govern their subject populations.

Institutional Thickness The notion that the role and quality of institutions in a region is a determining factor in economic development.

Mode of Regulation A concept developed by the French regulationist school of political economy which emphasizes the role of wider processes of social regulation in stabilizing and sustaining capitalism.

Neoliberalism A political and economic framework that seeks to reduce state intervention and embrace the free market, stressing the virtues of enterprise, competition, and individual self-reliance.

Path Dependency A concept adopted from institutional and evolutionary economics which highlights the influence of past decisions and experiences in shaping how economic actors respond to wider processes of economic change.

Post-Structuralism A set of theories derived from philosophy and literary and cultural studies which stress the fractured identities of individual subjects, seeing these as products of broader social categories and discourses.

Regime of Accumulation Another key regulationist term, referring to a relatively stable form of economic organization which structures particular periods of capitalist development, based on a balance between production and consumption.

Structuration Theory A sociological theory developed by Anthony Giddens in the 1980s which sought to integrate an emphasis on wider social structures with an understanding of individual actions.

Introduction

Institutions have become an important focus of interest in contemporary human geography, informed by various strands of institutionalist thought. Institutions can be defined as both formal organizations such as government agencies, corporations and voluntary bodies, and broader sets of habits, rules, and values that structure people's actions. The term tends to emphasize taken for granted behavior and practices that have become routine and normalized. Geographical research generally examines the role of institutions in structuring the management and organization of space. Geographers are particularly interested in the spatial scales or levels of activity at which institutions operate, ranging from the local and regional to the national, international, and global. Institutionalism refers to any strand of thought that is centrally concerned with the role of institutions in shaping social life. In many cases, human geographers have adapted institutionalist theories originally developed in other social sciences and humanities such as economics, sociology, political science, philosophy, and literature. Three different forms of the claim that 'institutions matter' are identified by the sociologist Bob Jessop. First, the notion that institutions are important in a thematic sense as a key focus of research. Second, a methodological claim that institutions provide a useful entry point for researching different aspects of social life. Finally, an ontological claim that institutions are a fundamental part of human existence. Much geographical research on institutions is thematic or methodological in nature, although some of the theories imported from other disciplines make strong ontological claims (e.g., institutional economics, the regulation approach, and structuration theory).

Geographical research on institutions in the 1970s and early 1980s tended to regard institutions as synonymous with organizations. One key expression of this work was an edited book, entitled *Institutions and Geographical Patterns*, published in 1982. Here, institutions covered bodies such as building societies, local authorities, development agencies, banks and pension funds, and planning authorities. The main focus was on how such organizations shaped the geographical landscape through particular decisions and interventions, considering, for example, how private corporations and development agencies shaped patterns of industrial location. Such concerns overlapped with another key strand of 1970s and 1980s 'institutionalism': namely the managerialist school of urban research which assessed the role of urban managers and 'gatekeepers' in shaping urban land and housing markets. The main institutions of interest here were building societies, local authorities, estate agents, and property developers. Interestingly, the adoption of an 'institutional approach' tended to be viewed as a middle or 'third' way between privileging either general theory or particular examples and observations. In contrast to the prevailing philosophies of spatial science and humanistic geography which stressed the importance of general processes and individual actors respectively, institutionalism provided a 'methodological' framework for linking particular case studies to wider concerns.

Institutions and Economic Geography

In contrast to mainstream neoclassical economics, economic geographers tend to emphasize the institutional foundations of markets, recognizing that the economy is shaped by a wide range of institutional forms and practices. Drawing upon the 'old' institutional economics of the late nineteenth and early twentieth centuries (rather than the 'new' (neoclassically oriented) institutionalism developed by mainstream economists), institutions are defined as sets of rules, habits, and values, although distinct organizations such as firms and development agencies have tended to be emphasized in much economic geography research. The focus on institutions represents a key dimension of the shift toward a more culturally informed economic geography, in contrast to the reliance on economic categories derived from neoclassical and Marxist thought in the 1970s and 1980s. Economic geographers are particularly interested in how local and regional economies are shaped by distinctive institutional arrangements. This has encouraged the rise of a 'new regionalism' which examines the effects of social and cultural conditions within regions in helping to promote or hinder economic growth. Successful regions such as Silicon Valley in California, Northeastern and Central Italy, and Baden Wurttemberg in Southern Germany have attracted particular attention, as economic geographers have attempted to identify the social and cultural foundations of economic growth and prosperity.

The regulation approach, developed by French economists in the 1970s, has focused attention on the role of particular institutional 'supports' in stabilizing capitalism, allowing periods of sustained growth to occur. These institutional 'supports' are understood in terms of the mode of regulation, referring to the norms and conventions which shape the process of capitalist development. Key aspects of capitalism upon which regulation is focused include labor and the wage relation, forms of competition and business organization, the monetary system, the state, and the international regime. When these different objects of regulation are linked together in a coherent fashion – achieving a balanced relationship between production and consumption – a period of stable growth, known as a regime of accumulation, ensues. The key regime of accumulation identified by regulationists is Fordism, prevalent between the 1940s and the 1970s in Western Europe and North America primarily. It was based on mass production and mass consumption, linked through rising wages for workers and increased productivity in the workplace. After the breakup of Fordism in the 1970s and early 1980s, elements of a new 'post Fordist' regime of accumulation based on flexible production emerged. Regulationist work in geography in the mid to late 1990s moved away from the Fordist/post Fordist debate, however, recasting regulation as an uneven process grounded in a range of different spaces of regulation and highlighting the role of local state agencies in mediating between local conditions and national regulatory frameworks.

In addition to the theoretical influence of institutional economics and the regulation approach, the emphasis on institutions has been reinforced by studies of industrial districts in the 1980s. Drawing on the earlier writings of the English economist, Alfred Marshall, this work draws attention to the renaissance of industrial districts within developed countries, such as Italy, France, Japan, Denmark, and Spain, involving a revival of local craft traditions and skills. The districts of Central and Northeastern Italy became particularly emblematic (**Figure 1**), becoming known as the 'Third Italy' (in contrast to the rich, industrial North and the poor, agrarian South). A key theme concerned the role of wider institutional and sociocultural factors in supporting networks of small firms and entrepreneurs. A dense network of local authorities, labor unions, industry associations, and chambers of commerce developed a sophisticated reservoir of knowledge, skills, and resources and provided business premises and a range of services to small firms.

A related set of ideas from economic sociology have been utilized to argue that the economy is always grounded in social relations, particularly Karl Polanyi's conception of 'the economy as an instituted process'. Geographers have 'spatialized' the sociological concept of 'embeddedness', emphasizing how particular forms of economic activity are not only grounded in social relations, but also rooted in particular places. Such 'embeddedness' has been linked to a focus on the role of local institutions in mobilizing resources for development. A key contribution here is Ash Amin and Nigel Thrift's concept of 'institutional thickness'. Four levels of 'institutionalization' are identified: the number of institutions present; the degree of interinstitutional interaction; the formation of coalitions; and the development of a common agenda incorporating key institutions and actors. Another strand of research on learning regions stresses the importance of path dependency and the role of institutions in shaping economic development trajectories. Path dependency means that the ways in which economic actors respond to wider processes of economic change are shaped and informed by past decisions and experiences. Reflecting its 'new regionalist' outlook, the

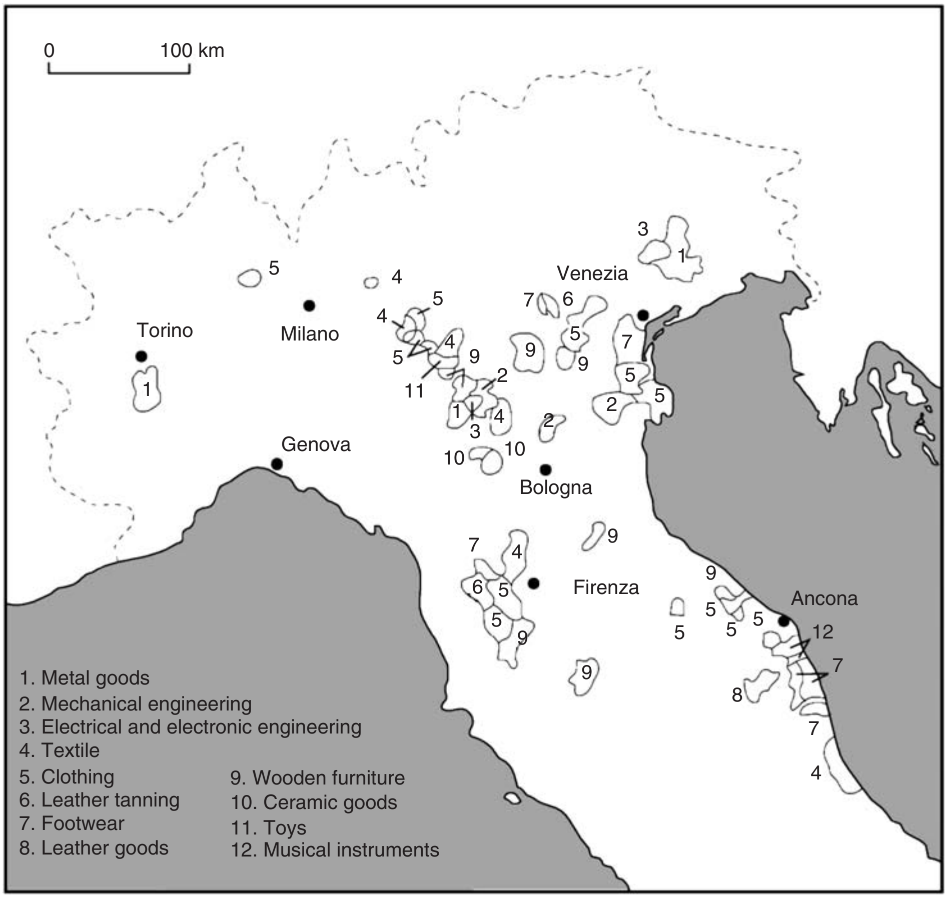

Figure 1 Industrial districts in Italy. Reproduced from Amin, A. (2000). 'Industrial districts'. In Sheppard, E. & Barnes, T. J. (eds.) *A Companion to Economic Geography*, pp 149 168. Oxford: Blackwell, with permission from Blackwell Publishing.

learning regions approach argues that globalization is encouraging the emergence of new forms of the spatial agglomeration or clustering of economic activity based on knowledge creation. Geographical proximity allows firms to transmit tacit forms of knowledge based on direct experience and expertise through face to face com munication, supported by a range of informal conven tions and ties (institutions in the broader sense).

While institutionalism and the 'new regionalism' have been highly influential in economic geography since the late 1980s, they have also attracted some criticism in re cent years. Both the institutional thickness and learning regions' literatures have been criticized for failing to ad equately specify the mechanisms which link insti tutionalization and agglomeration as processes to regional growth and learning as outcomes. More fundamentally perhaps, the underlying focus on regions as bounded spatial units has been called into question by the 're lational turn' in economic geography which emphasizes the importance of the broader networks and relations that link different places in an increasingly globalized world. Accordingly, the focus on local relationships between firms and institutions has been supplemented by a new concern with the role of extraregional networks in linking firms to distant sources of information and support.

Institutional Geographies of Social Welfare and Control

Institutions have also been a major focus of research in social geography since the 1980s, particularly in work on practices of social welfare and control. In particular, such research has concentrated on institutional spaces such as prisons, hospitals, asylums, and workhouses established to house those regarded as different or deviant in relation to mainstream society – 'the mad, the bad, and the sad'. This body of research investigates institutions in a very direct and immediate sense as actual physical buildings and sites which have sought to restrain, control, treat, and cure patients according to prevailing conventions and norms. Much of this work is strongly historical in nature, tracing the establishment and development of institutions in the late eighteenth and nineteenth cen turies as part of the emergence of modern society and the bureaucratic state. The key theoretical inspiration here is the work of the French post structuralist philosopher Michel Foucault, characterized by a general concern with the production or formation of human subjects through wider social and institutional processes. Distinguishing between the categories of 'same' and 'other', Foucault was particularly interested in understanding the operation of modern society through the classification, isolation, and disciplining of those seen as abnormal or deviant.

A key element was the movement away from the physical punishment of the deviant and criminal (a characteristic of premodern times) to an effort to cure and treat such groups in line with modern principles of legality. Interestingly, the need for the geographical separation or seclusion of the deviant became a major theme, in contrast to the premodern era when they tend to be mixed in with the broader community or housed in general purpose prisons. For example, Miles Ogburn assessed the separate confinement of prisoners as a cor nerstone of penal policy in the nineteenth century in the United Kingdom (UK). Similarly, Chris Philo examined the growth of a 'medico moral' discourse of nineteenth century professionals which argued for the establishment of separate asylums to regulate and treat the insane, something which shaped the institutional geography of the so called 'mad business'. A third example, analyzed by Felix Driver, concerns the growth of workhouses in England and Wales following the 1834 Poor Law, both reflecting and contributing to the expansion of the state through the collection and control of information and the standardization of bureaucratic practices. A key element in this process was the replacement of a complex array of private, charitable, and locally run facilities with a ra tional, integrated, and state controlled system of insti tutions (**Figure 2**). Accordingly, geographical uniformity would replace local differences with the legitimacy of the state depending on the consistent application of its pol icies throughout the territory. The role of central state agencies and inspectors was crucial, although standard ization was often resisted by local interests.

A distinction between two types of institutional geographies is made by Chris Philo and Hester Parr: a geography of institutions and a geography in institutions. The former involves an assessment of the actual locations of asylum and other facilities in relation to population, settlement, land use, etc. As outlined above, physical or spatial separation from the rest of society has been a key locational principle, with the dominant 'medico moral' discourse of the nineteenth century, for instance, em phasizing that asylums should be sited in tranquil rural environments where attractive surroundings, coupled with some gentle agricultural labor, would be therapeutic for inmates. Other work has focused on the location of mental health facilities in urban areas in the post war period, examining the impacts on, and attitudes of, surrounding communities. Studies of the internal

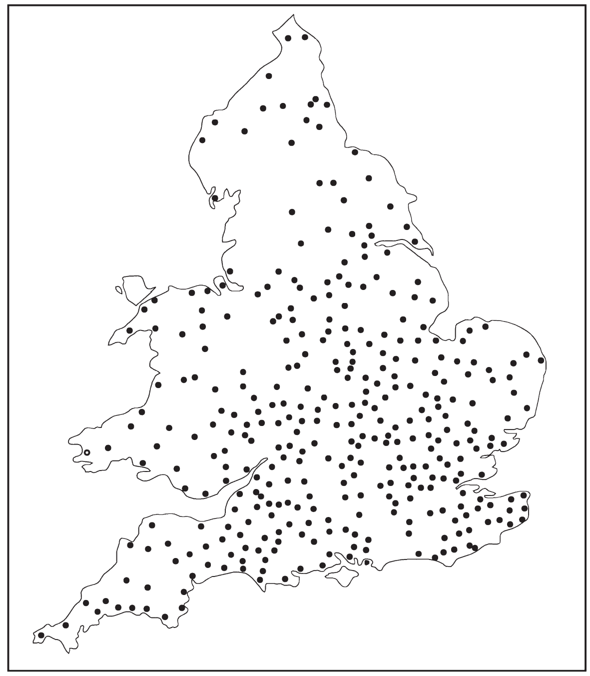

Figure 2 The geography of authorized workhouse construction in England and Wales, 1835 39. From Driver, F. (1987). The historical geography of the workhouse system in England and Wales, 1834 1883. *Journal of Historical Geography* 15(3), 269 286.

micro geography of institutions have, on the other hand, examined the arrangement of inner spaces in order to regulate and rehabilitate patients for their ultimate reintroduction into society. Of central importance here is the image of the panopticon – an institutional design which enables the authorities to observe all prisoners without the latter being able to tell if they are being observed (**Figure 3**). Developed by the English philosopher Jeremy Bentham in the late eighteenth century, the image of the panopticon was deployed by Foucault in his study of prisons and as a broader metaphor for modern society with its orientation toward observation and regulation. Another strand of research has examined the widespread process of 'deinsti tutionalization' in recent decades, something which reflects dissatisfaction with the idea of separation as a solution to deviance. Studies have assessed community and voluntary provision and investigated the identities and attachments to place of people with mental health problems and disabilities, employing qualitative methods such as interviews, focus groups, and participant observation.

Institutions and State Restructuring

Traditionally, the state represented the key institution of interest in political geography. More recently,

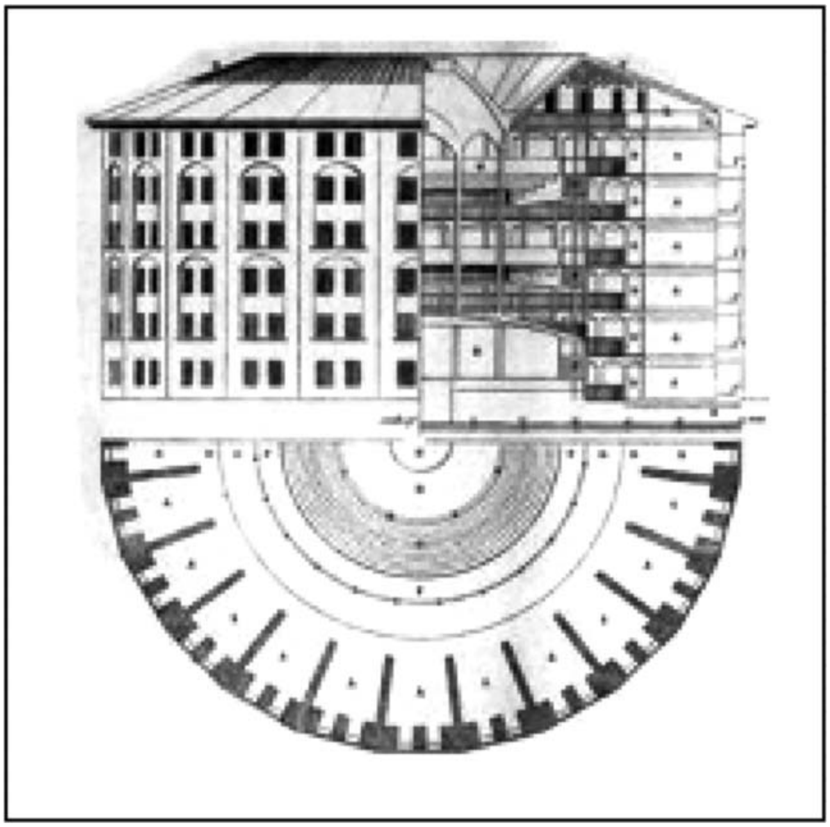

Figure 3 Panopticon blueprint by Jeremy Bentham, 1791. From http://en.wikipedia.org/wiki/Panopticon.

however, work on the state's role in the organization and management of space – something which can seen as formal 'large P' politics – has been supplemented by a focus on more informal 'small p' issues in terms of the politics of identity, gender relations, and the growth of social movements. A wider sense of institutions as sets of social and political norms and conventions is discernible in some of this work. At the same time, the conventional focus on government, referring to formal state institutions and bureaucracies, has given way to the broader notion of governance, incorporating various semi and nonstate actors such as special purpose agencies (often known as 'quangos' in the UK), private sector corporations, and voluntary bodies. The role of such actors in the management and delivery of services has been sub stantially increased in many countries since the 1980s, reflecting, in part, the influence of neoliberal reform programs.

Neoliberalism is a political and economic project which seeks to reduce state intervention in the economy and society and embrace the free market, stressing the virtues of enterprise, competition, and individual self reliance. It was embraced by the United States (US) and the UK in the late 1970s and early 1980s before being adopted by the World Bank and International Monetary Fund (IMF), who often required developing countries to introduce neoliberal reforms in exchange for financial assistance. Neoliberal policy measures in clude privatization (transferring state owned enterprises into private ownership), liberalization (opening up protected sectors to competition), and de regulation (relaxing the rules and laws under which business oper ates). In theory, neoliberalism stresses the need for 'de institutionalization', involving the abolition of existing organizations and rules in favor of market forces. In reality, however, the introduction of neoliberal reforms has required substantial 'reinstitutionalization', relying on the establishment of supporting frameworks and norms.

Two of the most influential 'institutionalist' theories in recent political geography research are the contrasting approaches of the political sociologists Michael Mann and Bob Jessop. While Mann's account of the develop ment of the state in relation to the sources of social power has been deployed largely to examine historical aspects of state formation and growth, Jessop's neo Marxist theories have directly informed political geography re search into contemporary processes of state restructuring. Jessop adopts a broad and inclusive perspective which seeks to place state power in its social, political, and economic context, focusing on the relationship be tween state institutions and social groups. Another influential approach which has inspired recent geographical research is the governmentality theory developed by Foucault. In broad terms, Foucault understood government as the 'conduct of conduct' – the array of techniques, practices, and rationalities (mentalities) that political agencies and authorities have deployed to govern and shape the activities of their subject populations. More recently, neo Foucauldian writers have sought to analyze advanced liberalism (neoliberalism) as a distinct political formation. This broad program of government is articulated and enacted through an array of mechanisms, techniques, and procedures which can be referred to as technologies of government. The role of these technologies – embedded in specific practices such as audit, targeting, financial management, and business planning – is important in giving central government the capacity to shape the practices of local agencies 'from a distance'.

A major theme of recent research has been the so called 'hollowing out' of the national state, referring to the transfer of functions and responsibilities to other scales and sites of governance. In a heavily cited account, Jessop has identified three main processes of 'hollowing out'. First, the denationalization of the state involves the transfer of some responsibilities and functions 'upwards' to supranational institutions such as the European Union (EU) and 'downwards' to local and regional agencies. Second, the 'destatization' of the political system refers to a shift of power 'outwards' to various arms length agencies, private interests, and voluntary bodies. Third, the internationalization of policy regimes involves a growing volume of linkages between national, regional, and local institutions and personnel in different countries. This enables policy transfer as initiatives introduced in one country are taken up and adopted by officials in another. Notable examples include privatization and the transfer of workfare initiatives from the US to the UK.

While the supranational tier of government has become more prominent, the subnational scale has attracted the most geographical interest. Research has examined the creation of 'new institutional spaces' at the local level in the 1980s and 1990s through the establishment of new local state agencies such as Urban Development Corporations, Training and Enterprise Councils (TECs) and National Health Service (NHS) Trusts by the Conservative government in the UK (**Figure 4**). In an important contribution, the geographer Jamie Peck argues that geographies of governance emerge out of a process of interaction between new regulatory tendencies at the national level and the preexisting institutional landscape at the local scale. The influence of preexisting institutional structures in shaping the local implementation of national regulations means that governance takes different forms in different places. Research on the TEC network in England and Wales and local enterprise companies (LECs) in Scotland and Wales deployed notions of institutional 'layers' and 'filters' to examine the shifting geographies of governance within the UK.

In recent years, devolution has become a key focus of research, particularly in the UK. It represents a key process of state restructuring in many countries since the 1970s and new devolved institutions were established in the UK in 1999. In the latter case, it can be seen as the latest round or 'layer' of state regulation, creating new geographies of governance. The notion of the 'filling in' of the state has been developed by a team of British political geographers, arguing that this metaphor can provide a handle on the processes occurring at the regional level under devolution, compared to 'hollowing out' which is focused on the national scale. 'Filling in' provides a framework for assessing the reorganization of governance within particular territories, involving the establishment of new institutions and/or the reconfiguration of existing ones, a process which has accentuated the geographical unevenness of the institutional landscape.

Institutions and the Construction of Regions

Reflecting growing dissatisfaction with both spatial science and structural Marxism, a number of prominent human geographers called for a 'new' regional geography in the late 1970s and 1980s. This can be seen as part of a search for a middle way between a focus on general processes and mechanisms and an emphasis on the particular and unique. The key theoretical inspiration here was the sociologist Anthony Giddens' structuration theory which emphasized the role of regions or 'locales' as settings for social interaction where people are socialized into society through various institutional processes, particularly education. In this context, the term institutions tended to refer to taken for granted rules, habits, and routines through which society is reproduced. Ultimately, however, Giddens tended to take the existence of regions for granted, failing to consider how they were historically produced and transformed, issues that were exciting the attention of geographers. In addition to Giddens' structuration theory, this rediscovery of regional specificity was linked to a Marxist informed understanding of regions as local responses to wider capitalist processes and more humanist view of regions or places as the focus for cultural identification.

One of the most interesting strands of this 'new' regional geography is the Finnish political geographer Anssi Paasi's examination of the construction and institutionalization of regions. Paasi conceptualizes this in terms of four stages: the development of 'territorial shape' as regional boundaries and structures are defined;

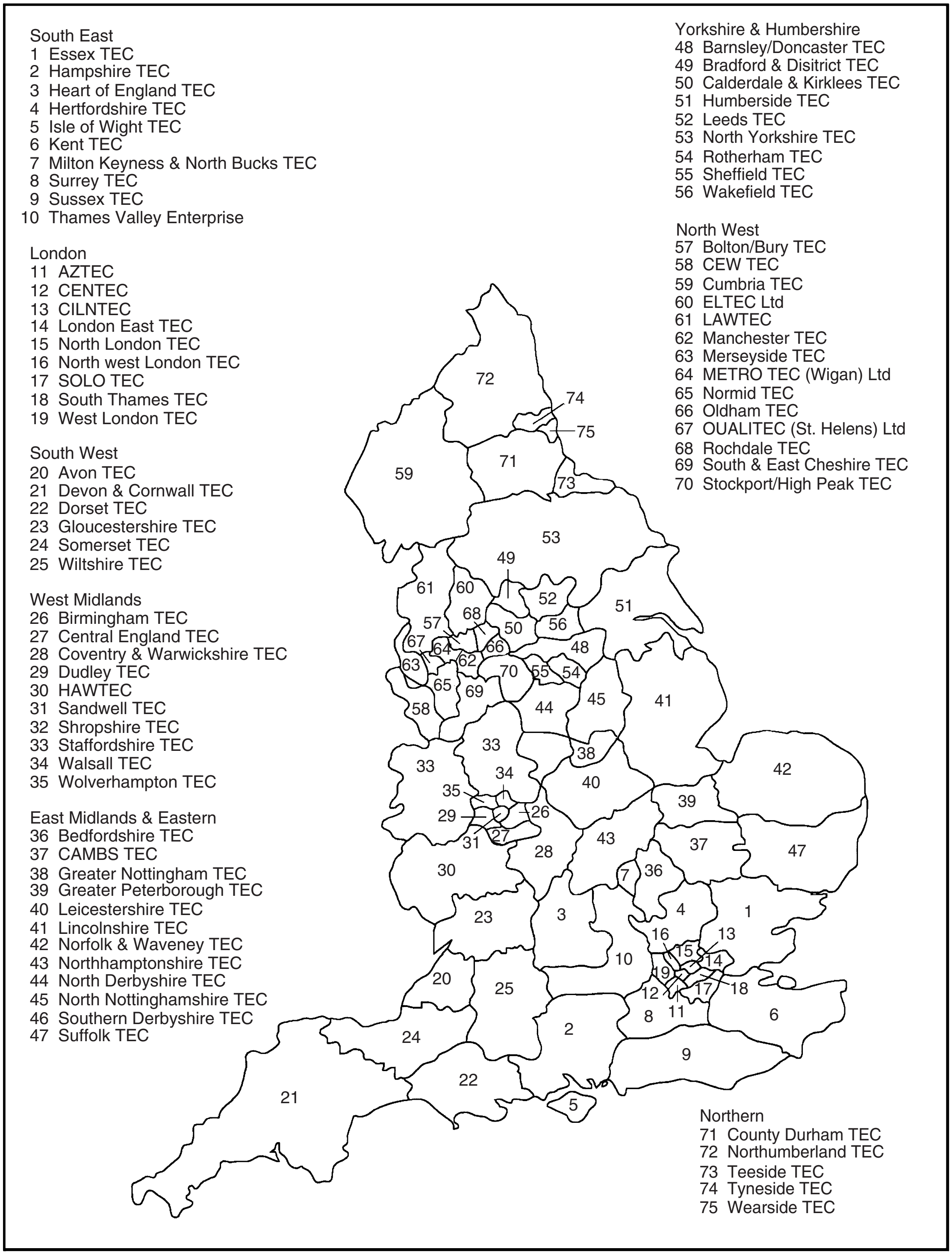

Figure 4 New institutional geographies: the 'original' TECs in England. Note: TECs were replaced by Learning and Skills Council in 2001. From Jones, M. (1999). *New Institutional Spaces: TECs and the Re-Making of Economic Governance*. London: Jessica Kingsley.

the establishment of 'symbolic shape' as regional symbols and images enter into the consciousness of the wider society; the emergence of institutions as key agencies of socialization reproducing regional distinctiveness; and the consolidation of the region within wider social and economic structures as these institutionalization processes persist. He emphasizes the reproduction of regions through collective, institutionally mediated rules and the development of regional consciousness.

Summary

As this article has shown, institutions are an important focus of research in geography. They began to attract significant attention in the late 1970s and early 1980s as certain geographers searched for a middle ground between spatial science and structural Marxism with their emphasis on general theories and processes and humanism which focused on individual consciousness and experience. Interest in structuration theory, the regulation approach, and institutional economics can all be seen as expressions of this search for a middle ground. The influence of post structuralist theorists like Foucault, by contrast, has emphasized difference and subjectivity, linking to long standing interests in the local and unique within geography. As a result of these diverse influences and orientations, institutional work in geography exhibits considerable richness and diversity. Much of it combines a sensitivity to difference and agency with an awareness of the broader 'structural' forces shaping the organization and management of space. Institutional research has attracted some criticism, mainly in economic geography, but institutions seem set to remain a central theme of geographical research. Future developments are likely to place more emphasis on the need for engagement with issues of public policy across different policy fields and geographical scales.

See also: Embeddedness; Fordism; Fordism, Post-Fordism and Flexible Specialization; Foucauldianism; Governance; Governmentality; Neoliberalism.

Further Reading

Amin, A. (2000). Industrial districts. In Sheppard, E. & Barnes, T. J. (eds.) *A Companion to Economic Geography*, pp 149 168. Oxford: Blackwell.

Amin, A. and Thrift, N. (1994). Living in the global. In Amin, A. & Thrift, N. (eds.) *Globalisation, Institutions and Regional Development in Europe*, pp 1 22. Oxford: Oxford University Press.

Driver, F. (1989). The historical geography of the workhouse system in England and Wales, 1834 1883. *Journal of Historical Geography* 15, 269 286.

Flowerdew, R. (1982). *Institutions and Geographical Patterns*. London: Croon Helm.

Jessop, B. (1994). Post Fordism and the state. In Amin, A. (ed.) *Post Fordism: A Reader*, pp 251 279. Oxford: Blackwell.

Jessop, B. (2001). Institutional re(turns) and the strategic relational approach. *Environment and Planning A* 33, 1213 1235.

Jones, M. (1999). *New Institutional Spaces: TECs and the Re Making of Economic Governance*. London: Jessica Kingsley.

MacKinnon, D. (2000). Managerialism, governmentality and the state: A neo Foucauldian approach to local economic governance. *Political Geography* 19, 293 314.

MacLeod, G. (2001). Beyond soft institutionalism: Accumulation, regulation and their geographical fixes. *Environment and Planning A* 33, 1145 1167.

MacLeod, G. and Jones, M. (2001). Renewing the geography of regions. *Environment and Planning D, Society and Space* 19, 669 695.

Massey, D. (1994). A global sense of place. In Massey, D. (ed.) *Place, Space and Gender*, pp 146 156. Cambridge: Polity.

Paasi, A. (1996). *Borders, Territories and Consciousness: The Changing Geographies of the Finnish Russian Border*. Chichester: John Wiley.

Painter, J. and Goodwin, M. (1995). Local governance and concrete research: Investigating the uneven development of regulation. *Economy and Society* 24, 334 356.

Philo, C. (1987). Fit localities for an asylum: The historical geography of the mad business in England as viewed through the pages of the Asylum Journal. *Journal of Historical Geography* 13, 389 415.

Philo, C. and Parr, H. (2000). Institutional geographies: Introductory remarks. *Geoforum* 31, 513 521.

Thrift, N. (1983). On the determination of social action in space and time. *Environment and Planning D, Society and Space* 1, 23 57.

Williams, P. (1982). Restructuring urban managerialism: Towards a political economy of urban allocation. *Environment and Planning A* 14, 95 106.

Relevant Websites

http://www.devolution.ac.uk/
ESRC Research Programme.

http://www.institutions.org.uk/index.html
Rossbret Institutions Website.

http://www.worldbank.org/
The World Bank.

Integrated Spatial Data Infrastructure

R. Foley, NUI Maynooth, Maynooth, Ireland

Glossary

Data Protection The requirement that data about people conforms to rules of nondisclosure and privacy.

Geospatial Information Any information that can be geographically referenced, that is, describing a location or any information that can be linked to a location.

Global Positioning System (GPS) A satellite-based navigational system allowing the determination of a unique point on the earth's surface with a high degree of accuracy given a suitable GPS receiver.

Interoperability It is the ability to link different applications regardless of proprietary software used at the source together so that several web services are available concurrently.

Metadata The data describing the data. Typically this consists of a number of descriptor fields identifying content, quality, currency, ownership, and other characteristics.

Standards An agreed set of minimum requirements, which are to be reached by all data providers.

Introduction and Context

With the development of Geographic Information Sys tems (GISs) from the late 1960s on and with the rapid development of computing technology and power, geo graphical information (GI) and digital spatial data have become key resources in all branches of geography. As cartographic data have become increasingly digital, both in terms of their sourcing and manipulation, the power of spatial data is increasingly recognized across the globe. The recent arrival of cartographic resources into main stream consciousness through the development of Goo gle Earth and other Internet based data has also created an increased awareness of the power of mapping. At a governmental level the increasing use of GI in all bran ches of government and its widespread adoption and use in civil society is well established in the developed world and is growing in the developing world as well. Within society as a whole there is also an increased awareness of the ways in which spatial data underpin society in the form of in car navigation systems, satellite imagery use in the media, and the use of GI in business demographics. One of the principal tasks of GI experts therefore, is to manage and shape this increased potential and awareness and one of the key global models used to try and do this is spatial data infrastructures (SDIs).

One definition of an SDI is:

> The technology, policies, standards, and human re sources necessary to acquire, process, store, distribute, and improve utilization of geospatial data. (FGDC, 1994)

SDIs were first discussed and proposed in the early 1990s in both Europe and the US. The influential report by Lord Chorley published in London in 1987, focused on the potential of spatial data to underpin much of the future informational needs of society. Specifically, Chorley made a number of recommendations around the human dimension of GI and the need for a coordinating body to manage the development of the industry and the data. This recognition of the need for coordination was particularly pertinent to discussions on copyright, standards, interoperability and take up. The coordination role to fully develop and explore the potential of GI is one that lies at the heart of SDI. In the early 1990s a similar initiative in the US, driven by the Federal Geo graphic Data Committee (FGDC), led to the formal endorsement by the then President, Bill Clinton to de velop a national spatial data infrastructure (NSDI). **Figure 1** below outlines the core components envisaged in the US NSDI.

One of the leading proponents of SDI is Ian Masser, who saw spatial data as a national asset, having something in common with other infrastructural elements such as railways and roads. The spatial scale of operation for SDI is important and much attention in recent years has been focused on the need for regional, national, and even global organizations all of which are linked by common standards but which also vary from continent to continent and country to country. Masser also identified the NSDI as the most developed level with other levels identifiable at supernational (Global Spatial Data Infrastructure (GSDI), and Inter Continental versions) and subnational level (typically in federal countries). At the beginning of 2007, there were over 120 countries worldwide either actively developing, or considering the development of, an SDI.

Key Components of an SDI

There are a number of key components that underpin SDI, whatever be the spatial scale at which it operates. These include legislative frameworks, content and data frameworks, metadata and standards, quality issues, geoportals, human factors, and dissemination and

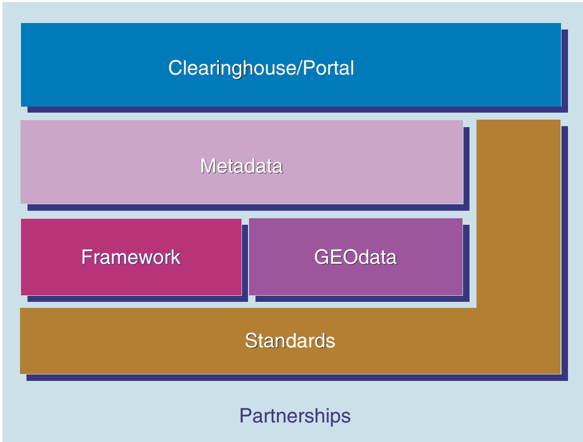

Figure 1 Suggested componets of a Spatial Data Infrastructure (FGDC, 1994)

implementation. By definition, each of these are diffuse and complex areas but to realize a vision of an SDI, they are all essential in teasing out the real life issues that such an initiative faces. They also need to be understood in placing SDI more closely within the wider GIS and geography disciplines.

Legislative Frameworks

The development of any SDI requires a minimum level of legal standing and support. The relationship between GIS and the 'law' is still being worked out and varies considerably across the globe. This depends also on the geographical scale of the framework and can range from the local and regional to the national, international, and global. Initiatives at world (Global SDI) and continental level (e.g., the EU INSPIRE initiative) aim to set these agreed legal frameworks in place. At a spatial data level, agreements on data protection, copyright, and ownership all have the potential to be serious barriers to the de velopment and adoption of SDI and need to be addressed at a very early stage. It is clearly easier to handle national legal requirements than international ones.

Content and Data Frameworks

The range and coverage of spatial data are extensive at the beginning of the twenty first century. Many subjects collect and use spatial data and these range from ad ministrative data routinely collected by government on employment and electoral rights to subject specific data such as health care utilization and land ownership. The level of coverage does however vary considerably both between and even within countries. While SDIs do not aim to suggest an agreed standard minimum data set, they provide instead a framework or template into which sets of digital spatial data may be arranged and organized to an agreed international standard. Agreements on what core data layers should be included in these frameworks is a source of considerable debate within SDI. The US NSDI lists seven framework themes, which include: geodetic control, cadastral, orthoimagery, elevation, hy drography, administrative units, and transportation. These provide a good working model for most SDIs.

Metadata and Standards

As part of the development of spatial data standards, in part driven by the Chorley Report, a series of technical committees, namely 211 and 278, have met at ISO level to develop and agree a set of technical standards and agreed metadata formats for spatial data. Current standards include the ISO19100 Series/TC 211, as well as standards developed by agencies such as the FGDC in the US and the private sector Open GeoSpatial Con sortium (OGC). Metadata which, simply put, is 'data about data', is essential to these standards in that it re quires data holders to use standard descriptors to docu ment the precise nature of the spatial data holdings. A number of international standards such as the Dublin Core are already being used in this area.

Quality Issues

Clearly, it is impossible to manage and standardize spatial data from all over the globe. Quite apart from difficulties with projections and coordinates, national and private mapping agencies produce maps at a bewildering range of scales, few of which are standard. Similarly, the der ivation and accuracy of digital spatial data varies hugely around the world. Even in countries with a long scientific cartographic tradition such as the UK, Germany, and the US, the arrival of new technologies such as satellites and GPS has brought with them full adjustments to the na tional grids. Much of the initial work of global SDI has focused on the development of initial small scale but worldwide layers such as the Digital Chart of the World, a global dataset released in 1993. Other quality issues relate to scale, currency, and the breadth of attribute data.

Geoportals and Distributed Information Systems Architecture

A key aspect of any SDI is that, once data have been collected, collated, and verified to agreed international standards, it must be disseminated to potential users. Quite apart from the complex legal aspects of data dis semination and data sharing, the technology must be capable of allowing spatial data to be distributed equally, correctly, and with an acceptable level of ease for the end user. The growth of the Internet has been a huge factor in making such geoportals feasible for the distribution and online cataloging of spatial data sets. Geoportals are a generic name for the IT service locations which users access to identify, view, and download spatial data sets. As such they are the 'public face' front end to any SDI and are a key outcome of developments in 'Internet GIS'.

Human Factors

As the quantity and quality of spatial data has mushroomed since the 1980s, the value and importance of spatial data has also been recognized. In an increasingly IT literate world, information and access to it brings with it associations with power and financial gain. As such, the ownership of data, who is allowed use of it, and who should benefit from it, are key concerns. Linked to this, the rights of individuals to privacy and the wider importance of data protection offer a real challenge to SDIs worldwide. As the fundamental aim of SDI is to pull together and disseminate spatial data sets, the potential for individuals and agencies to prevent this from happening is enormous for a variety of reasons, including, national interests, legal challenge, work priorities, and commercial concerns.

Dissemination and Implementation

Quite apart from the effective dissemination mentioned above, the real challenge to full SDI development lies in its implementation. Much of that implementation is linked to both data audit (on the collection side) and public access (on the dissemination side). A good SDI should by definition have a minimum core set of data, at an acceptable scale and with an acceptably complete level of coverage. All of these elements are easier said than done.

Current Spatial Data Issues

This section outlines current research around a range of key spatial data infrastructural issues, including standards for both data and metadata, interoperability versus harmonization, data quality, access, intellectual property rights (IPR) and digital rights management, and managing change. All of these issues have an impact on how spatial data is collected, processed, stored, interpreted, used, and disseminated by the owning organization and how it is accounted for by data custodians.

Much of the research on SDI has been driven by technical discussions around standards, interoperability, and dissemination. The work of Burrough and Masser in Europe and by the FGDC in the US had been mirrored by work in Africa, Asia, and Australasia. Much of this work has yet to filter into mainstream geographical literature though the work of Pickles in creating a debate on the human implications of GIS may yet be replicated as the wider world picks up on the potential of GI and its organization within SDIs. In particular, critical debates on data ownership, power, and representation are also appearing in the wider human geography literature but these may be necessary to fully understand the implications of SDI development and to help clarify where SDI can really work for the common good in terms of opening up access to spatial data and promoting its optimal utilization.

The literature on Standards and Interoperability is best summarized in some of the current literature. Masser, in particular, identifies the current state of play around the world in terms of national, regional, and global standards and provides a selective history of their development. He comments in particular on the recent organization of a global initiative, GSDI that is a natural progression from the smaller national level SDIs. Issues of data quality continue to be crucial to the acceptability and success of SDI development. Evidence suggests that public expectations have increased with the IT revolution and such expectations are clearly linked to acceptable standards in terms of the level of accuracy, the number of available datasets, and usability of spatial data sets collected and disseminated within SDI frameworks.

Data access is a subject that has always exercised geographers and despite the obvious efficiency gains in storage and maintenance in a digital age, there are still trade offs taking place between data owners and data users. The dissemination of spatial data within SDI is in part driven by an understanding of public rights to spatial data. This clearly varies internationally in that different states have different attitudes to the release of spatial data, ranging from free public access to heavily protected and costly data sets. Some of these difficulties also reside in debates around IPR and 'added value'. Arguments abound as to the fact that raw spatial data is developed by states but the value added product is the comprehensible spatial map layer which the user needs and the ownership of that is vested in the data creators/owners. Increasingly, privately collected spatial data, enhanced by technologies like GPS and Lidar, are an additional complication in relation to ownership and access.

Current Drivers and Barriers

While much of the literature and discussions around SDI have been embedded in technical discussions and the wider policy and political implications, it is worthwhile to summarize the current state of play in terms of Drivers and Barriers. An implicit aim of SDI is to place in the public arena a set of spatial data layers which can be easily accessed, used, and manipulated. It is almost certain that the Internet will be the medium for this dissemination and that this in turn may allow for potentially free and unrestricted use of those datasets. This outcome is what should ultimately guide how SDI is developed. If there are concerns and difficulties these will determine the nature of the data finally provided and the one im ponderable at this stage is what form the access and utilization of that data will take. Will access to such data create a demand for more? Will the ways in which the public interpret and use that data affect what data is provided? Will the data be exploited commercially by the data users more than the data providers? Will SDI underpin global developments in e government in a positive or negative way?

Arguably, once users understand and use distributed spatial data sets, as envisaged by SDI, the flexibility and sheer ease of access should arguably create an additional demand. The success of geoportals such as the Alexandria Digital Library and the National Statistics website in the UK show just how such access has enabled individuals, community groups, agencies, and other government de partments. With the development of the Internet and more significantly the growing use of spatial data in normal daily life (SatNav, GPS, Weather maps, Google Earth), there is evidence for a growing voluntary spatial data creation process which currently lies parallel to any putative SDI framework. For many data providers, a continuing concern is the way in which their data might be misused or misinterpreted.

Spatial data have enormous potential in maximizing economic, social, and environmental benefits. The com mercial, social, and economic value of such data sets is being recognized daily. A conflict may exist between the raw data and its use for commercial profit or state control. One thing that should be pointed out is that SDI makes no explicit provision for free access to spatial data, though some people may understand it as this. Essentially, there is the scope in SDI to disseminate data in a charging regime that may range from free to full commercial cost. Getting the pricing right is one of the challenges which may affect the success of SDI development.

At a wider level, SDI has the potential to provide a good example of technology transfer, bringing the tech nical knowledge of the spatial data community into the real world for its greater benefit, while at the same time safeguarding the standards of that data. Other potential benefits relate to the development of agreed models and templates to improve data sharing protocols and also provide a form of future proofing as volumes of available data increase. The global initiatives around e government are also very strong drivers for the implementation of SDI as they provide a technical and political solution in disseminating government data.

Barriers to the successful implementation of SDI are manifold. Some of the difficulties encountered in de veloping the US NSDI are well known. Interestingly, standards and interoperability aspects have proved much less difficult than agreements on what data sets should be included and the difficulties of working with multiple stakeholders. The latter aspect, which incorporates legal and human factors associated with data sharing, data access, and both national and international agreements continues to hold up progress in a number of key areas. Additionally, the continuing speed and unpredictability of ICT technological change (ICT – Information and communication technology) through the development of location based services and mobile technologies repre sent both a barrier and an opportunity. Data access and data creation are enabled through such technologies. Linked to this the awareness of spatial data through so cietal advances in GPS/mobile applications may embed the power and value of spatial data even further into human consciousness. Such developments have also the potential to spiral out of control and perhaps a full and concretized development of SDI may be one way to manage this potential for the common good.

The Future of SDI

Finally, there are a number of ongoing discussions and large scale research initiatives about the ways in which SDI may develop. The Australian SDI (ASDI) identifies five priority areas, namely: governance, access to data, data quality, interoperability, and integratability. It identifies the need for clear policy mechanisms, cham pions of the concept, better understanding of governance, and stronger ties with wider technological frameworks. One example from Europe which may exemplify this is the Infrastructure for Spatial Information in Europe (INSPIRE) initiative in the EU. This is at the stage of ratification and the result will be an agreed and fully developed model for public sector SDI across the member states that is to be implemented through an EU Directive before 2010.

At the same time it is important to be aware of real izable targets worldwide. One of the potential dangers of SDI, especially to its proponents and some members of the GI community, is its aspirational aspect. The exist ence of strong legislative backing, which initiatives such as INSPIRE promote, may not be practical in other parts of the world. Progress may be slower in places which are either less technologically developed or where national interests around data protection and national security (such as in the US and China) may hamper that devel opment for the moment. The unpredictability of human response and take up are also issues that will continue to hamper development. It is ultimately important to point out that, quoting the case of the US NSDI:

> the NSDI is not a concrete 'thing' but is more of a vision, a state of mind, a campaign and an enabler for better use of scarce resources. (Longley *et al.*, 2005)

See also: GIS and Society; Global Positioning/GPS.

Further Reading

Burrough, P. and Masser, I. (eds.) (1998). *European geographic information infrastructures: Opportunities and pitfalls, GISDATA 5*. Taylor and Francis Ltd. and European Science Foundation: London.

Cho, G. (2005). *Geographical information science: Mastering the legal issues*. Chichester: Wiley.

Crompvoets, J., Rajabifard, A., Bregt, A. and Williamson, I. (2004). Assessing the worldwide developments of national spatial data clearinghouses. *International Journal of Geographical Information Science* 18(7), 665 689.

Curry, M. R. (1998). *Digital places: Living with geographic information technologies*. London: Routledge.

Department of the Environment (1987) *Handling geographical information: Report of the inquiry chaired by Lord Chorley*. London: HMSO.

Federal Geographic Data Committee (1994) Executive Order 12906, http://www.archives.gov/federal register/executive orders/pdf/12906.pdf (accessed 16 July 2007).

Groot, R. and McLaughlin, J. (eds.) (2000). *Geospatial data infrastructures: Concepts, cases and good practice*. Oxford: Oxford University Press.

Longhorn, R. and O'Riain, G. (2005). *Investigation of Best Practice Principles for a Spatial Data Infrastructure in the Technical Field*. Dublin: Report to the Irish Spatial Data Infrastructure Committee.

Longley, P., Goodchild, M., Maguire, D. and Rhind, D. (2005). *Geographic information systems and science* (2nd ed.). Chichester: John Wiley & Sons.

Martin, D. (1996). *Geographic information systems: Socioeconomic applications* (2nd ed.). London: Routledge.

Masser, I. (2005). *GIS worlds: Creating spatial data infrastructures*. Redlands: ESRI Press.

Nebert, D. (ed.) (2004). *Developing spatial data infrastructures: The SDI cookbook, version 2.0*. Reston: GSDI.

Pickles, J. (ed.) (1995). *Ground truth: The social implications of geographical information systems*. New York: Guilford Press.

Schuurman, N. (2005). *GIS: A short introduction*. Oxford: Blackwell.

Williamson, I, Rajabifard, A. and Feeney, M. E. F. (2003). *Developing SDI: From concept to reality*. London: Taylor and Francis.

Relevant Websites

http://www.agi.org.uk/default.asp
Association for Geographic Information UK.

http://www.anzlic.org.au/infrastructure ASDI.html
Australian Spatial Data Infrastructure.

http://www.geoconnections.org/CGDI.cfm
Canadian GeoSpatial Data Infrastructure.

http://www.eurogi.org/default.asp
European umbrella organisation for geographic information.

http://www.fgdc.gov/components
Federal Geographic Data Committee.

http://www.gsdi.org/
Global Spatial Data Infrastructure Association.

http://www.ec gis.org/inspire/
INSPIRE.

Intensive/Extensive Research

K. Falconer Al-Hindi, University of Nebraska at Omaha, Omaha, NE, USA

Glossary

Abstract In critical realist analysis, the conceptual level at which all but the necessary and contingent properties of an object or social structure have been stripped.

Abstraction A crucial step in critical realist analysis. It takes a form called retroduction in which the scholar begins with an event that takes place in a specific concrete setting. Next, the scholar develops a one-sided abstraction from that event in successive rounds in order to identify the necessary and contingent properties of an object or social structure.

Concrete In critical realist analysis, the conceptual level at which occurrences may be observed and documented (e.g. sale of property, or political rally).

Contingent Relations When two phenomena have a connection to one another that is not required in order for those phenomena to be what they are, the relations are said to be contingent. That is, the association between the two could cease without disturbing the status of the property of interest. For instance, the husband and the baker he patronizes have a contingent relation. Both would be what they are without the latter selling to the former.

Critical Realism The philosophy of science originated by Roy Bhaskar and widely adopted by geographers following Andrew Sayer. A key tenet of critical realism is that what causes something to happen has nothing to do with the frequency of occurrence.

Empiricism Philosophy of science which holds that what is apparent is all that is real, and perception – usually visual perception – is all that is required to know the world, or especially to know a specific phenomenon. Empirical observations are much more important than theory, which is often considered irrelevant by empiricists.

Necessary Relations A connection between two individuals or objects without which the things cannot be what they actually are. For example, the husband–wife relation is a necessary one, for without a wife, a man cannot be a husband (we might say he is single, divorced, a widower, etc., but would not describe him as a husband). The condition of being a husband requires that the man is married to a wife.

Realism A philosophy of science that distinguishes among what can be perceived, what makes things happen, and the level of underlying social structures. Naïve realism is virtually synonymous with empiricism, and insists that what is perceived directly is all that is real (e.g., the 'common sense' geographer). Contrast with critical realism.

Retroduction A 'thought experiment' that identifies objects, mechanisms, necessary and contingent relations to be investigated in intensive research.

Structure and Agency A sociological view of human social life that distinguishes between individuals as agents, who act in ways that either uphold structures or challenge them, and institutions, which, though they are human creations, comprise sets of rules and practices that both constrain human action and offer opportunities.

Definitions

Intensive and extensive have specific meanings within the philosophy of science known as critical realism. Intensive research investigates causal relationships among phe nomena in order to discover the mechanisms that cause an event to take place. Extensive investigations seek to discover how common or widespread a phenomenon is; that is, to uncover empirical regularities. Intensive and extensive researches may complement one another. For example, an intensive study of the gendered division of household labor might examine causal factors such as the gender ideologies of adults in the family. Other important factors might include: ages and any special needs of household members; hours of paid work; commute times; and self assessments of skill and interest in household work and dependent care. An extensive study of the gendered division of household labor would investigate the frequency with which women perform more than two thirds of the work. According to critical realism, how often something happens (in this case, how often women assume the second shift of household work and childcare) has little or nothing to do with why it happens or what causes it to occur. The same event can have different causes under different circumstances; here, one woman might perform the second shift because doing so is consistent with her gender ideology, while another woman might do so because her male partner refuses to do half of the domestic work.

Origins of Intensive/Extensive in Human Geography and the Social Sciences

Intensive and extensive have come with critical realism into geography primarily through the work of Roy

Bhaskar, Rom Harré, and Andrew Sayer. Bhaskar and Harré are philosophers of science and it is with critical realism as such that they are concerned.

Intensive and extensive first appear in Harré's *Social Being* (1979) during his discussion of acceptable research methods in sociology. He makes three important points. First, intensive research studies an individual believed to be typical of the group of interest, or a small number of individuals who belong to the group. In the latter case, the small group's members will share a relatively large number of characteristics. Extensive research, by contrast, involves a large group of individuals who share a small set of characteristics. Second, intensive research seeks to reveal structural relations and interactions among group members; it also permits detailed descriptions of individuals. Harré clearly favors intensive work: he argues for an intensive and idiographic social science. Sayer's *Method in Social Science: A Realist Approach* (1984, 1992) introduced critical realism, and the key concepts intensive and extensive at a time when many geographers were struggling to get past positivism and empiricism, and offered a new conceptualization of the relationship between theory and empirical work alongside a compelling critique of positivism.

Intensive/Extensive in Methodology

Intensive and extensive are central to critical realist philosophy with respect to methodology, method, and ontology. To begin, they are distinguished from each other in each aspect of research methodology and method. These properties are summarized in **Table 1**.

Methodology begins with research questions: what will the study investigate? Intensive and extensive approaches are useful for different kinds of questions. The types of questions that intensive research answers concern causality and process: for instance, what causes an event to take place? The spatial relations in intensive research are substantive; there is a meaningful, necessary association between agents or things in geographic space. Groups are studied in which individuals are causally related to one another and interact in meaningful ways, for example, employers and employees. These two categories of individuals, to extend the example, interact with each other in ways that bear on phenomena that may interest geographers and other social scientists, such as the displacement of workers by technology or the gendering of workplaces. Intensive research produces causal accounts capable of explaining the mechanisms and relationships through which specific events take place. Corroboration of research results affirms the findings of intensive research. Intensive research is conducted using methods such as semi structured and in depth interviews, ethnography, and other qualitative approaches. Intensive accounts are characterized by detailed descriptions of causal relationships among members of the group studied. The concrete patterns and contingent relations discovered are unlikely to be found elsewhere exactly as they are in the specific case studied.

Table 1 Intensive and extensive research: a summary

	Intensive	*Extensive*
Research question	How does a process work in a particular case or small number of cases? What produces a certain change? What did the agents actually do?	What are the regularities, common patterns, distinguishing features of a population? How widely are certain characteristics or processes distributed or represented?
Relations	Substantial relations of connection	Formal relations of similarity
Type of groups studied	Causal groups	Taxonomic groups
Type of account produced	Causal explanation of the production of certain objects or events, though not necessarily representative ones	Descriptive representative generalizations, lacking in explanatory penetration
Typical methods	Study of individual agents in their causal contexts, interactive interviews, ethnography. Qualitative analysis	Large-scale survey of population or representative sample, formal questionnaires, standardized interviews. Statistical analysis
Limitations	Actual concrete patterns and contingent relations are unlikely to be representative, average, or generalizable. Necessary relations discovered will exist wherever their relata are present, e.g., causal powers of objects are generalizable to other contexts as they are necessary features of these objects	Although representative of a whole population, they are unlikely to be generalizable to other populations at different times and places. Problem of ecological fallacy in making inferences about individuals. Limited explanatory power
Appropriate tests	Corroboration	Replication

Reproduced from Sayer, A. (1992). *Method in Social Science: A Realist Approach* (2nd edn.). London: Routledge, with permission from Routledge.

Intensive research is not representative or generalizable in the usual sense, but may be so in the critical realist sense, where conditions that activate causal powers inherent in necessary relations among objects may be found in settings beyond those investigated directly.

Extensive research is interested in spatial patterns and the distribution of a phenomenon: how often does the event occur? Under what conditions? The spatial relations in extensive research are formal, concerning phenomena that fulfill standard criteria. The groups studied belong to the same taxa, or category. Continuing the examples begun above of employers and employees, an extensive study could survey a large sample of employers for information on numbers of employees displaced by technology, or collect data from a large number of employees on their perceptions of gender in the workplace. Extensive research is associated with quantitative measurement and uses standardized instruments for collection of data from large numbers of individuals, such as surveys or standardized interviews. Extensive accounts are characterized by claims to generalizability. Extensive research does not provide an explanation why the phenomenon occurred (here, unemployment due to replacement of workers by technology or the character of gender relations in the work setting): only its frequency and possibly some correlations which, instead of being causal, are only casual. Replication is often used to test research results gained through extensive studies. Results are representative only for the population from which the sample is taken. Extensive realist research may appear more familiar to many (noncritical realist) geographers, although within a critical realist framework it can be distinguished from empiricism and most versions of positivism by its concern with causality: what causes events to take place? Because the individuals studied are related only formally (e.g., they are all employees), the explanatory power of extensive research is much less than that of intensive study, but it may aid in identifying external and contingent factors. The search for causal relations distinguishes extensive research in critical realism from that in positivism.

Critical realism is distinguished by its deep ontology, which identifies three domains: the real, the actual, and the empirical. The real is a realm in which objects with structures and powers exist. Examples of objects include different entities such as the US Census Bureau and natural gas. Each has the capacity to make things happen; the structures that enable such occurrences are called necessary properties of each object. For example, natural gas has the ability to burn. Whether or not it burns depends on a contingent relation: its contact with a spark. In the domain of the actual are occurrences and their effects that arise from the activation of objects' powers. The domain of the empirical is where people experience events (which may or may not be observed; there are other ways of sensing events). To continue with the example of natural gas, inhaling it may cause a person to lose consciousness: an empirical event. To work through an example beginning in the empirical domain, start with the US Census survey newly arrived in the mail. In order for census surveys to be distributed, enormous resources for printing, mailing, and delivery must be activated, which occurs at the level of the actual. At the level of the real we find the US Census Bureau, a division of the United States Federal Government (the State) itself, with its mandate, policies, procedures, budget, ideological underpinnings, and employees. Critical realism's ontological depth permits it to investigate what makes things happen, in contrast to positivism, in which realism's three realms are collapsed into just one: the empirical. With only the realm of experience to study, critical realists argue that positivism's ability to understand phenomena is severely limited.

Intensive and extensive are key to realist ontology. Continuing with the example of the US Census, an intensive investigation of the question "why were few surveys returned?" concerns the three domains of empirical, actual, and real. In the actual realm, perhaps some surveys were not delivered. At the real level we might find that the US Census Bureau did not command sufficient resources to print and distribute all of the surveys. Alternately, we might discover that so many citizens are experiencing a crisis of confidence in their institutions that they decided not to complete the surveys. An extensive investigation would proceed very differently. Rather than posing a 'why' question, a question concerning extent is appropriate: How many surveys were returned, and from where? The addresses from returned surveys could be entered into a geographical information system (GIS) and mapped in order to answer these questions. Depending on the resulting spatial distribution, additional questions could be asked or hypotheses developed. Note that the research question associated with extensive research is one of number and pattern, while the intensive question is concerned with understanding.

The connection between intensive and extensive concrete research is integral to the concrete–abstract relationship in critical realism. While intensive and extensive may start from the same observation, they are concerned with different types of questions and use different methodologies. Intensive research investigates causal processes in one or a small number of cases; extensive research examines common properties and spatial patterns of a population or large sample. Some studies combine the two. Abstraction takes a specific form in critical realism called retroduction. These thought experiments must proceed recursively; that is, the scholar must work through higher to lower levels of abstraction as well as from lower to higher levels in recursive rounds. If the abstraction does not appear to work at each level, then it must be modified. Intensive research investigates the abstraction, and relationships discerned through retroduction, while engaged with all three of critical realism's ontological levels.

Intensive and extensive research can be, but may not be assumed to be, complementary. A number of studies have begun with an extensive investigation and then developed methodologically related research questions based on the findings for intensive investigation. How ever, such an intimate connection between the two ap proaches is required in order to consider the studies complementary of one another. Members of just any pair of intensive and extensive studies of the same phenom enon are unlikely to complement one another in a strong sense. One exploration of contingency (causal mech anisms that may or may not be activated in a specific context) has argued that the difference between intensive and extensive may be collapsed when research methods that account for local variation in explanatory factors and that can process large amounts of quantitative data from diverse locales, such as the expansion method, are employed.

Intensive/Extensive in Space

The spatial properties of intensive and extensive came to geographers' attention, along with the first discussions of critical realism, in the early 1980s, with what became known as the localities debate. Many scholars linked the study of localities with critical realism and sought to identify the necessary aspects of the spatial relationships they observed by conducting intensive case studies. A discussion arose concerning the difference that space makes in such research, with Andrew Sayer maintaining that space mattered at the concrete level but that it was irrelevant at the most abstract level of analysis. Dis agreement over the role of geographic space was attrib uted to confusion between abstract and concrete research, and here intensive and extensive play crucial roles. Abstract research in critical realism relies on thought experiments (painstaking mental work) to dis cern the necessary properties of objects at the level of the real. From concrete situations at the empirical level the scholar is to discard properties of objects that are un necessary to producing the effects these objects have on the ground. As the argument went, the specific spatial properties of objects such as the capital–labor relation or patriarchy are left behind through succeeding rounds of abstraction from the empirical level to the real. Abstract does not have to mean a spatial, but some interpreted it as such.

By the early 1990s, three signal contributions of in tensive and extensive (and critical realism more gener ally) to geography could be identified. First, intensive research and attendant methodological arguments offered insight toward a resolution to the debate between agency and structure. Intensive investigations of indi viduals (not necessarily people) positioned in causal re lationships to one another permit study of how social structures and their powers work out in concrete settings. Neither agency nor structure collapses; rather, agents are shown to operate in complex contexts in which struc tures' mechanisms are at play. Second, the debate over space advanced geographers' conceptions of geographical relationships. Engagement around the notion that there could be aspatial abstractions ultimately produced con sensus: that any abstraction devoid of spatial content is probably too abstract and not useful for social science; that spatial properties may in fact be necessary to causal mechanisms and in fact are central to objects' being what they are and thus contribute to places being what they are; and that all social relations are spatial relations also. Widespread recognition of the political and relational nature of social space among human geographers gave further credence to the idea that space must always be 'in' critical realist analysis. Intensive research has the capacity to reveal that the spatial effects of, for instance, a national policy vary according to local contextual factors (say, gender ideologies). Extensive research lacks the ability to investigate these and so, as in positivism, the researcher may *de facto* assume that the same causal agents at work in one place are at work equally in all. Realists criticize much positivist and empirical geography research, saying that the search for empirical regularities and predictive power utterly confuses intensive and extensive by employing extensive methods of inquiry (e.g., large scale surveys) to answer intensive questions (e.g., what causes industry to locate in a particular area?). Local context may not be reduced to a few spatially variable factors such as employment rates, etc., but must be handled as the complex and combinatory locale that it is. Third, feminist geography's engagement with critical realism, and particularly with retroduction, advanced understandings of methodology and the role of theory in feminist geography. A heated exchange among several feminist geographers in the pages of *Antipode* com plemented, in effect, the concurrent exchanges in main stream outlets about abstraction and the importance of spatiality to realism's objects. Intensive case studies are an antidote against both the risk of over abstraction and the temptation to draw conclusions based on thought ex periments alone without evidence gathered through empirical research.

Both intensive and extensive contribute to con ceptions of space at small and large scales, and for both smaller and larger areas. Intensive and extensive entered the discipline via investigations of regional, or meso scale, inequalities that were emerging in the United Kingdom by the 1980s. Intensive investigations at multiple and often far flung sites generated understand ings of how national policies combined with contingent forces in specific settings to produce change. They were also conducted at single sites. Extensive studies have been less common but also involve small or large scales, smaller or larger areas. The difference between intensive and extensive lies with their frameworks of explanation and not their scale or areal extent.

For geographers who associate critical realism with the localities studies of the 1980s or generally with economic or political geography, it may seem that in keeping with these subfields' historical approaches to study, intensive and extensive researches are solely con cerned with spatial relationships exterior to homes, of fices, and other buildings. To the contrary, for example, peoples' reasons for their activities can be causal forces and so both intensive and extensive research pursue in formation on these from individuals in their routine settings: often, interior spaces. Intensive research in particular has contributed to a relational conception of geographic space that is manifest in homes, offices, and institutional settings by tying observations of concrete events in them to mechanisms and necessary properties of objects such as patriarchy.

Building on the theoretically informed empirical work of geographers studying economic change in the United Kingdom, scholars in the United States and beyond took up critical realism and particularly its intensive research methodology to investigate phenomena such as North west lumber production, women's telecommuting, and transnational corporations in Southeast Asia. Frequently, these studies revealed unexpected explanations for spa tial relationships, as in Yeung's discovery – through conducting intensive and extensive research simul taneously – that personal networks were key to firms' decisions about locating in Southeast Asia. They also often complicated the classic notion of the region in ways that integrated newer insights such as a relational con ception of space.

Places are shown by such studies to be complex rather than static and constitutive of social life rather than simple stages for human interaction. Critical realism's abstractions are meaningless without their effects on the ground; intensive research uncovers and explains these effects while extensive research elucidates spatial rela tionships between such places. Thinking of places as al ways in process is one of the contributions to geography of intensive studies. This idea of the ever becoming na ture of places informed a new approach to regional geography and, for some, reinvigorated what had other wise become an intellectual backwater.

Since the beginning of geography's engagement with critical realism, scholars have worked to compare, con trast, and reconcile critical realism with other approaches. Discussions of the role of intensive and extensive research have often arisen in these exegeses. For these workers, teasing out the implications of intensive and extensive for spatial relations and places has contributed immensely to more sophisticated and nuanced conceptions of spatial relationships and places.

Most recently, intensive and extensive are mentioned in critical realist excursions in the feminist, education, and nursing literatures. The concepts are not developed in much depth however and it could be argued with respect to the last two that these engagements are just beginning. Lack of attention to place and space is notable. Geographers who engage with scholars in these areas will have much to contribute.

See also: Capital and Space; Critical Realism/Critical Realist Geographies; Locality Debates; Mixed and Multiple Methods; Positivism/Positivist Geography; Quantitative Methodologies; Scientific Method; Spatial Expansion Method; Structuralism/Structuralist Geography.

Further Reading

Bhaskar, R. (1975). *A Realist Theory of Science*. Leeds: Leeds Books.

Del Casino, V. J., Grimes, A. J., Hanna, S. P. and Jones, J. P. III (2000). Methodological frameworks for the geography of organizations. *Geoforum* 31, 523 538.

Harré, R. (1979). *Social Being*. Oxford: Blackwell.

Jones, J. P. III and Hanham, R. Q. (1995). Contingency, realism, and the expansion method. *Geographical Analysis* 27(3), 185 207.

Sayer, A. (1992). *Method in Social Science: A Realist Approach* (2nd edn.). London: Routledge.

Sayer, A. (2000). *Realism and Social Science*. Thousand Oaks, CA: Sage.

Sayer, A. and Morgan, K. (1985). A modern industry in a declining region: Links between method, theory and policy. In Massey, D. & Meegan, R. A. (eds.) *The Politics of Method*, pp 147 168. London: Methuen.

Yeung, H. W. C. (1997). Critical realism and realist research in human geography: A method or a philosophy in search of a method? *Progress in Human Geography* 21(1), 51 74.

Relevant Websites

http://www.lancs.ac.uk/fass/sociology/profiles/Andrew Sayer/
Professor Andrew Sayer, Department of Sociology, Lancaster University.

Interdisciplinarity

P. Raento, University of Helsinki, Helsinki, Finland

Glossary

Cultural Geography A subfield of human geography with interest in cultural phenomena.

Cultural Turn A renewed and expanded interest in the study of cultural phenomena in human geography in the 1980s and 1990s.

Diaspora The dispersion of a population that shares a place of origin or cultural characteristics. When written with a capital D, refers to the scattering of Jews after their exile from Babylonia.

Holistic It is concerned with wholes rather than their parts.

Hybrid The offspring of two different species, an outcome of mixed origins.

National Science An academic field of study deemed central for the construction and maintenance of a national identity.

Pandemic A worldwide epidemic of a contagious disease.

Paradigm A predominant viewpoint in academic research, established and accepted by the majority of scholars.

Positivism A science–philosophical way of thinking founded on the belief in objectivity, first-hand observations, and measurability.

Synergy An additional force or energy produced by working together.

Interdisciplinarity commonly refers to the joining of forces by two or more academic fields of knowledge. The word well describes the phenomenon, for the first part (inter) implies togetherness, mutuality, and reciprocity. The second part of the word (discipline) connotes rules, boundaries, and order. Built in the idea of inter disciplinarity is the existence of disciplines, branches of knowledge in academic institutions. Interdisciplinarity therefore is a twentieth century concept, which followed the institutionalization and segmentation of scientific research in the nineteenth century. Technological in novations, growth of population and industrial pro duction, and the consolidation of modern state structures all increased the need for knowledge. The differentiation within disciplines encouraged the crossing of their boundaries. The roots of interdisciplinarity, like those of disciplines such as geography or philosophy, nevertheless reach to classical Greece and Rome.

The motivation for interdisciplinary collaboration often emerges from a specific problem in need of a solution or a significant new development in society. Globally relevant examples include the impact of information technologies or pandemics such as HIV/AIDS and the bird influenza. Collaborating scholars approach their topic of study using the theoretical, methodological, and conceptual toolkits characteristic of their own disciplines, but adjust them to serve a common goal. Their work ideally produces a comprehensive understanding of a particular phenom enon, a region, or its people. Rather than merging entire fields, interdisciplinarity draws in a mutually supporting fashion from particular specialties within disciplines. Success highlights instrumental problem solving and a constructively critical exchange of ideas.

A Response to Major Changes

Interdisciplinarity typically evolves through para digmatic changes which often follow transformations in society. In these situations, predominant views are chal lenged and even replaced, because they no longer can answer emerging questions in a satisfactory manner. In human geography, one recent landmark change was the renewed interest in culture in the 1980s and the 1990s. In this so called cultural turn, the ways in which human geographers had traditionally approached the complex concept of culture were questioned. Critical researchers moved the focus from the study of forms and structures to the interpretation of meaning in the examination of landscapes, from high culture to popular culture, from quantification toward qualitative methods, and from so cietal and general to quotidian and individual foci. Per spectives employed in anthropology and sociology were particularly useful in this remaking of culturally oriented human geography. The debate resulted in a considerable diversification of human and cultural geographical in vestigations and new collaborative networks. The changes supported the visibility and credibility of human geog raphy as a social science and underscored the importance of space in human phenomena. This diminished the relative isolation of human geography, caused by geo graphy's strong roots in the positivist tradition of science. However, the development also worked to widen a gap between human and physical geography, and between approaches within human geography, raising concerns about the future of holistic geography and a com prehensive understanding of regional processes.

The diversification and hybridization of views were related to a broader change in the philosophical foun dations of science and understanding of human phenomena, in the shift from modern to postmodern approaches. This has meant the questioning of previous, historically overarching, and predominantly Western, often masculinist, ways to think and explain the world. A need to take into account a variety of voices and to inspect multiple scales of activity (from the local to the global level) has been highlighted.

Human geographical studies of consumption exemplify the shift toward postmodern and cultural approaches. Early studies about consumption relied heavily on economy in their modeling and evaluation of networks from the perspective of efficient flows, accessibility, and most suitable locations. Along with the challenge to predominant views in architecture and with human geographical interest in behavioral studies came questions such as, "How do spatial designs impact behavior and identity of consumers?" Subsequent inquiries enhanced contacts between semiotics and the geographical interpretation of symbolic meanings (drawing from landscape studies and cartography). Also evident were connections to the sociological study of inclusion and exclusion in public and private space. These changes emphasized qualitative methods and individuals over quantification and systems. Simultaneously, new themes – such as food – entered human geographical studies about consumption, and researchers asked how spaces themselves are being consumed. For example, in the marketing of cities and their neighborhoods, images of cultural and intellectual vivacity, cuisine, economic prosperity, comfort and convenience, and safety are valuable place promotional tools. Recent concerns about the environmental impact of agriculture, the role of food in the development and political stability of societies, and the geopolitical implications of food related risks have shown, however, that knowledge from both human and physical geography is necessary for reaching a comprehensive geographical understanding of complex phenomena like food and cities.

Research Institutions

Interdisciplinary institutions typically take the form of centers, programs, or other special units of a university or a college rather than being entirely independent. Examples of interdisciplinary employers of human geographers include area and cultural studies institutes, migration research centers, and women's studies programs in universities worldwide. The work conducted in these units focuses on a particular globally relevant phenomenon, a cultural sphere, or a territory. The employees represent a variety of academic backgrounds, so that human geographers work together with historians, anthropologists, sociologists, and political scientists.

The history and research emphases of these units often reflect the settlement history and cultural characteristics of the surrounding region. For example, in the USA, the migration history of a university's host region steers academic interests toward population groups that are prominent in the area. In many smaller countries in Europe, interdisciplinary research units may have been designed to improve knowledge of economically and politically important regions because of their historical influence in that country. Quite frequently these connections are evident in the outside funding received by these institutions. Foreign governments or civic organizations may provide grants or other support for particular units in order to promote their own visibility, to maintain active connections with an emigrant population, and to stay updated of the latest relevant research. These funds may be earmarked for researcher exchange, popularization of science, or the promotion of cultural activities. The so called third task of universities – the involvement of academics in societal discussions and activities – typically carries notable weight in interdisciplinary units.

Two Examples

At the University of Nevada, Reno, the Center for Basque Studies brings together scholars who conduct research on the Basque population and society in Spain and France and their history and cultural influence in the New World. The center has a specialized research library, an interdisciplinary publishing program, and an international, multilingual faculty who represent a variety of social sciences. The center's small group of doctoral students works together with scholars from other departments at the University of Nevada to combine their regional–cultural expertise with broader general methodological, conceptual, and theoretical knowledge. One of the collaborating departments within the center's network is the Department of Geography. A student may thus graduate with the interdisciplinary degree of Doctor of Philosophy in Basque Studies (Geography) and pursue an academic career in similar interdisciplinary research centers or in human geography.

The location of the center is no accident, for Northern Nevada lies within the hearth of Basque culture area in the American Southwest. The need for a Basque research program emerged in the 1960s, when scholars working on a larger research project in the surrounding Great Basin recorded multiple layers of Basque history and culture in the area. Over the decades, the center has grown in size and diversified its activities. The faculty and staff work with the local community organizations to broaden knowledge of regional history and culture and to preserve and study these traits. The government of the Basque Autonomous Community in Spain has sponsored international research conferences about Basque society in Europe and the Basque diaspora in the New World.

The Renvall Institute of Area and Cultural Studies at the University of Helsinki, Finland, was founded in the late 1960s to promote interdisciplinary and international exchanges. The institute is part of the College of Arts, but it collaborates with several departments and colleges. The scope of the institute has gradually expanded from regional interests to be global in its reach, reflecting changes in the Finnish society and in universities worldwide. Thus, former foci on the German culture area and on the British Isles have been merged into a broader European Area and Cultural Studies program. The institute currently hosts five other regionally defined cultural and social studies programs, which reflect Finland's political, cultural, and economic history and interests, and academic tradition. These programs are Nordic Studies, Russian and East European Studies, North American Studies, Latin American Studies, and Asia Pacific Studies (African Studies are housed under a different unit at the University of Helsinki). In these six programs, interdisciplinarity is promoted through diverse international teaching and visitor activities. Daily routines are conducted in at least six languages. Human geographers (who represent the College of Science) work at the institute as adjunct faculty members, guest lecturers, board members, and administrative coordinators.

Interdisciplinary Practices

The names of units only tell a partial story, for many interdisciplinary activities take place in informal networks and routines of individual faculty members. Program coordination, guest lecturing, and seminar attendance are all tasks that enhance the role of interdisciplinarity in one's professional profile, but this may be hidden behind disciplinary definitions and titles.

The specialization of disciplines has created new hybrids and new knowledge through the crossing of conventional boundaries. Arguably, many large geography departments across the world are interdisciplinary in character, for they host a variety of specialties across the broad scope of geography. As a result of this broadness, the faculty often consists of human geographers, physical geographers, and, in many countries, urban and regional planners, each of whom represent highly specialized subfields. Human geography alone has numerous subfields, the histories, methodologies, and theoretical foundations of which can be notably different from one another. The growth of diversity continues, containing both opportunities and challenges for collaboration. A specialist in transportation geography may approach space as logistical networks, measurable flows, and location patterns, using sophisticated computer assisted cartographic tools to visualize the findings. In the meanwhile, the cultural geographer next door may investigate the meaning of landscapes, its material constituents, and their representations in visual arts and people's perceptions. In a research project about new transportation arrangements in a city, these approaches, the expertise of urban planners, and physical–geographical knowledge about urban nature may come together in complementary, mutually supporting fashion so that infrastructural solutions accommodate the preferences of local residents while meeting logistical and environmental requirements.

In some geography departments, this potential for synergy and the respect for the old ideals of synthesis and unity in science are used to support research profiling thematic emphases, such as broadly defined urban studies or environmental studies. Under one banner, human and physical geographers and urban planners combine their expertise to produce essentially interdisciplinary perspectives on cities and the environment. A study of waterways in urban areas can thus result in a holistic understanding of a physical–geographical phenomenon and its economic, ecological, and recreational impact on, and meaning to, city dwellers. By promoting these interdisciplinary possibilities within geography and countering the concerns about compartmentalization, individual departments have sought to stand out in the competition for funding and students. Administrative and teaching duties related to these activities are interdisciplinary in character.

Networking

Rather than housing everyone under one roof, participants in loosely arranged networks are scattered across space in separate departments, cities, and countries. Successful maintenance of these networks requires mobility, active communication, and coordination over distance. Technological advancements that have greatly supported this collaboration include electronic mail and advanced audio–visual devices that allow comprehensive and convenient person to person and group communication, irrespective of location. Also important is the evolution of air travel into a relatively inexpensive, efficient, and convenient tool for global access. These technologies are irreplaceable for contemporary academics, for they facilitate communication in person. Several studies show that most scholars still prefer to exchange ideas face to face and appreciate the informal social dimensions of their work. Not surprisingly then, an increasing number of congresses, thematic workshops, and seminars are organized around interdisciplinary research themes all over the world. The meetings of national or other major geographical societies, such as those of the Association of American Geographers (AAG) and the International Geographical Union (IGU), are also interdisciplinary, in that they bring together geographers from all specialties. For many geographers, the opportunity to learn about a new place and its people is an additional incentive for participation. These observations underscore the complexity and informality of the practice of interdisciplinary.

Interdisciplinary Publishing

Researchers and publishers depend on one another, for research provides the raw material for publishers and researchers need to get published to promote their findings, careers, and name recognition. Not surprisingly, academic publishing has followed the trends of growth, specialization, and interdisciplinarity characteristic of academic research activities and institutions. Typical publications with interdisciplinary contents are edited books, in which diverse groups of scholars team up to assess particular themes from multiple angles. Within disciplines, interdisciplinary approaches are promoted in books where one or two authors challenge predominant views with the help of other disciplines. The bulk of interdisciplinary publishing, however, takes place in academic journals.

Journals

Interdisciplinarity takes many forms in journal publishing. The market has grown and diversified greatly since the 1970s, again reflecting changes in academia and society. One result has been the emergence of thematically defined journals, which encourage and invite submissions from a variety of fields. An example is the *Annals of Tourism Research*, published since 1973. It seeks to advance the understanding of tourism by serving as a forum of interaction for multiple disciplines. The growth and specialized segmentation of the global tourism industry especially since the 1990s increased social scientific research about the branches of this industry. This specialization, new themes and processes, and the scholarly pressure to publish encouraged the launching of new, interdisciplinary, and more narrowly defined journals, such as *International Gambling Studies* in 2000. This journal attracts a heterogeneous group of gam(bl)ing and tourism scholars, who represent numerous social sciences (including human geography), economy, and psychology. In addition to scholars and students, the readers are policymakers and casino business representatives. The founding of a new journal indicates that a field of research has gained prestige and moved from the margins of academic interest toward the center.

Several geographical journals are interdisciplinary in character. One example is journals that focus on a particular subfield of human geography from a broadly defined thematic perspective. For example, in *Political Geography*, the focus is on the spatial dimensions of political phenomena. The definition attracts not only political geographers from various research traditions, but also authors and readers from other fields, especially political science. The unifying theme is the relationship between space and politics, which is at the core of, but not exclusive to, political geography. Journals like this resulted from the increased specialization and hybridization within geographical research in the 1970s and 1980s.

Another example is all inclusive journals of geography that seek to represent the entire spectrum of geographical research. These journals are usually older than the specialized ones, reflecting the evolution and foci of academic geography since the late nineteenth and early twentieth century. The publishers are likely to be national geographical societies, the task of which is to promote geographical education and knowledge, to support a professional community of geographers, and to foster holistic collaboration across geography's internal boundaries. The *Annals of the Association of American Geographers* and *The Geographical Review*, by the American Geographical Society, are examples of these journals in North America. Some societies, such as the Royal Geographical Society in Britain and the Institute of British Geographers, use a commercial publisher for the production and distribution of their journals (*Area* and *Transactions of the Institute of British Geographers*). In small, linguistically unique countries, such as Finland, where geography is considered to be a national science, the all inclusive approach is seen as a valuable way to support the training of young scholars and the evolution of the national language(s). *Terra*, a Finnish and Swedish language quarterly journal of the Geographical Society of Finland, serves the entire community of Finnish geographers by informing the readership about the latest trends in geography, about contacts to its sister disciplines and to other countries, and about the evolution of Finnish and Swedish language terminology in geography and other relevant fields. The journal both maintains the professional identity of geographers in Finland and aids in collaboration across (sub)disciplinary boundaries by publishing theme issues designed for this purpose. Confirming the relatively long and well established history of institutional geography in the country, the journal reached its 120th volume in 2008.

However, human and physical geographers, and their peers in other fields, follow different publication strategies and have different preferences regarding audiences, outlets, and styles of presenting their findings. The scene is, therefore, very complex and some tension exists. In Britain, one study has shown that human geographers favor geographical journals, whereas their colleagues in physical geography prefer to publish their work in specialized journals outside of geography. Especially scholars outside of the Anglo American core of academia make strategic and ethical choices of language when searching for outlets for research articles. Some approaches or topics that would not be of interest to an international audience may be nationally or regionally relevant and offer access to valuable information to speakers of a particular language. In some countries, established geographical journals may have prestige that attracts submissions from other disciplines, whereas elsewhere geography is less prominent. Factors affecting the choices include personal career goals, national science policies, evaluation practices, and perceptions of prestige and hierarchies. All this steers the interdisciplinary research and publishing practices of individual scholars in particular settings.

Instruction

In contemporary universities, students have more and more choice regarding what to study. Disciplinary specialization and interdisciplinarity have provided students with an unprecedented opportunity to shop for knowledge. Behind curricular innovations are also the needs of students, the need to train scholars, and the need to offer professionals continuous education opportunities. Sometimes institutional rearrangements, such as downsizing of budgets, produce mergers that enhance interdisciplinarity in academic curricula.

Interdisciplinary programs offer opportunities for joint teaching and guest lecture series. On the one hand, this approach is instructive for students and teachers alike, for it promotes comparative and critical discussion in the classroom. On the other hand, the joining of forces in teaching may underscore one's disciplinary background and subject matter, especially if the perspective is comparative. Furthermore, interdisciplinary units often invite instructors to assess a topic from the perspective of a particular branch of knowledge. In a North American Studies program, the adjunct faculty member from a geography department may teach an introductory course about the regional or human geography of North America, cross listing the class under geography. These courses typically contain elements of both human and physical geography, based on a view that basic knowledge about physical–geographical processes is needed to understand human activity, the formation of regions and cultural spheres, and the continuous change of landscape. The course may form part of a series of specialized first year classes, which supports a social scientific synthesis of human phenomena in North America at a more advanced level of study. Disciplinary specialties of faculty members can be used in this way to promote interdisciplinary and holistic thinking among students and introduce to them a selection of approaches available in the emphasized field.

Interdisciplinary instruction is a challenge to the teachers' teaching philosophy and practices, because these scholars can expect to meet a class of students from diverse academic backgrounds. It also matters where one teaches, for practices vary from one country and institution to another. In North American universities where many courses can be selected freely, one teacher may have students from over half a dozen disciplines, with considerably diverse prior knowledge about the subject matter. In these cases, the challenges are at least twofold, irrespective of whether the class is listed under geography or some interdisciplinary program. One challenge is to guarantee that the applied perspectives, supporting concepts, and the goals of the class are clear to all attendees. Another is to make sure that the students are offered an equal opportunity to participate in the class discussion and exercises. In contrast, in countries where entry to an undergraduate program is based on qualifying examinations or some other form of preselection, the new students will have an even level of skills and relatively standardized background knowledge about their major subject, but know fairly little about approaches in other relevant disciplines. In both contexts, the choice of reading materials must be carefully balanced. The teacher of an interdisciplinary class, therefore, needs to solve how to assess key concepts and methodologies, and how to encourage the students to search and apply information across disciplinary boundaries. Possible solutions are multiple and vary, but, generally, teachers of interdisciplinary themes and groups of students constantly scrutinize their teaching philosophy and practices to adjust flexibly to the demands of each class.

Benefits and Obstacles

Interdisciplinarity means different things to different scholars, and practices vary. The role of individuals is, in any case, crucial in the evolution of interdisciplinarity. This is so because individual scholars make decisions and interact with others in groups, networks, and institutions. Successful collaboration requires careful and conscious attention to communication, for specialized expert communities typically have their own, nuanced vocabularies, and one word can mean different things to different people. Trust and social skills that can accommodate differently sized egos are among the critical constituents of successful interdisciplinary work. The design of interdisciplinary projects, even within geography and one linguistic sphere, demands that the participants fully understand the ways in which they themselves and others construct, order, and communicate knowledge.

At one end of the spectrum of views are those who see interdisciplinarity as an opportunity and strength – as something desirable. In their view, boundary crossing, experimental hybrids, and exploration beyond conventions support creativity and innovation. From this perspective, interdisciplinarity encourages continuous learning, intel lectual stimuli, and freedom of choice. Open mindedness and constructive competition of perspectives are mutually supportive and help to underscore the spatial dimensions of human phenomena. According to those who think of interdisciplinarity as a strength, one outcome is an en hanced and better defined professional identity as geog rapher, human geographer, and, say, political or cultural geographer, because interdisciplinary scholars must understand the specifics of their own contribution. The holders of these views consider all this beneficial to the evolution of academic research, in general, and human geography, in particular. The emphasis here is on the first part of the word (inter): on a confident collaboration of disciplinary bases in order to understand phenomena such as migration, pandemics, or consumption.

Those of a more critical view place the emphasis on the latter part of the word (discipline). From this perspective, caution is recommended, and disciplinary cores and boundaries need to be maintained to protect continuity and clear professional identity. The holders of this view point out that novel ideas are not always welcomed and competition, disciplinary differences, and subsequent dif ficulties in communication are likely to strengthen boundaries rather than lower them. The promotion of new ways of thinking and networking can require an un comfortable degree of risk taking, not least because the pushing of boundaries of the mainstream may lead to defense reactions and resistance. Scholars proposing es tablishment shaking ideas may encounter difficulties in finding research funding, which depends on outside evaluation and trends in national or organizational science policies. Critics point out that the pursuit of inter disciplinarity may include difficulties of getting tenure and feelings of being marginalized or confused about one's professional purpose. Some face difficulties in finding suitable publication channels. All these obstacles may harm individual career development and delay public scrutiny of the proposed ideas. On the other hand, successful overcoming of these difficulties can profile these individuals as path opening pioneers in their re spective fields, adding to the lure of interdisciplinary exploration.

See also: Cultural Turn; Geographical Journals; Human Geography; Knowledge Communities; Los Angeles School of Post-Modern Urbanism; Migration; National Schools of Geography; Research Funding Bodies.

Further Reading

Agnew, J. and Duncan, J. S. (eds.) (1989). *The Power of Place. Bringing Together Geographical and Sociological Imaginations*. Winchester: Unwin Hyman.
Bracken, L. J. and Oughton, E. A. (2006). 'What do you mean?' The importance of language in developing interdisciplinary research. *Transactions of the Institute of British Geographers* 31, 371 382.
Dear, M. and Flusty, S. (eds.) (2002). *The Spaces of Postmodernity. Readings in Human Geography*. Oxford: Blackwell.
Dogan, M. (1997). The new social sciences: Cracks in the disciplinary walls. *International Journal of Social Sciences* 153, 429 443.
Douglass, W. A. and Raento, P. (2004). The tradition of invention. Conceiving Las Vegas. *Annals of Tourism Research* 31, 7 23.
Gaile, G. L. and Willmott, C. J. (eds.) (2006). *Geography in America at the Dawn of the 21st Century*. New York: Oxford University Press.
Gregory, D. (1994). *Geographical Imaginations*. Cambridge: Blackwell.
Johnston, R. (2003). Geography: A different sort of discipline? *Transactions of the Institute of British Geographers* 28, 133 141.
Jones, P. and McDonald, N. (2007). Getting it wrong for the first time: Building an interdisciplinary research relationship. *Area* 39, 490 498.
Klein, J. T. (1990). *Interdisciplinarity. History, Theory and Practice*. Detroit, MI: Wayne State University Press.
Klein, J. T. (1996). *Crossing Boundaries. Knowledge, Disciplinarities, and Interdisciplinarities*. Charlottesville, VA: University Press of Virginia.
Library Trends (1996). Special issue 45 (2).
Murphy, A. B. (2006). Enhancing geography's role in public debate. *Annals of the Association of American Geographers* 96, 1 13.
Philo, C. (ed.) (1991). *New Words, New Worlds. Reconceptualising Social and Cultural Geography*. London: Institute of British Geographers.
Sluyter, A., Augustine, A. D., Bitton, M. C., Sullivan, T. J. and Wang, F. (2006). The recent intellectual structure of geography. *The Geographical Review* 96, 594 608.

Relevant Websites

http://www.ees.elsevier.com/atr
Annals of Tourism Research. A journal by Elsevier.
http://www.aag.org
Association of American Geographers, Washington, DC.
http://basque.unr.edu
Center for Basque Studies, University of Nevada, Reno.
http://www.tandf.co
International Gambling Studies. A journal by Taylor and Francis.
http://www.igu net.org
International Geographical Union.
http://www.ees.elsevier.com/jpgq
Political Geography. A journal by Elsevier.
http://www.helsinki.fi/hum/renvall/english
Renvall Institute for Area and Cultural Studies, University of Helsinki, Finland.
http://www.rgs.org
Royal Geographical Society, London, UK.
http://www.helsinki.fi/maantiede/geofi/terra/
Terra. A journal of the Geographical Society of Finland (in Finnish and Swedish).
http://www.amergeog.org/gr
The Geographical Review. A journal of the American Geographical Society.

Intermediate Technology

M. Parnwell, University of Leeds, Leeds, UK

Introduction

The appropriate role of technology in the process of development lies at the heart of the notion of 'inter mediate technology'. The concept was originally de veloped by E. F. 'Fritz' Schumacher, whose seminal volume *Small is beautiful: A study of economics as if people mattered* (1973) challenged many of the orthodox views, processes, and outcomes associated with Western models of economic development. Although Schumacher's thesis was formulated at a particular moment in the history of modern development, it contributed significantly to an intellectual shift which today can be found rooted in the environmental, democratization, and grassroots devel opment movements, and still has a certain currency and validity today.

Schumacher's Thesis

Fritz Schumacher questioned the paramount position that had been ascribed to technology and innovation in the process of economic development in the influential work of Joseph Schumpeter, most particularly his *The theory of economic development: An inquiry into profits, capital, interest, and the business cycle*. Technological innovation was seen as an essential means of strengthening the competitive position of firms, and the pursuit of technological advancement was positioned as the motor driving capitalist industrialization forward, particularly in very large firms which could support research and development. This 'idolatry of giantism' and 'reification of technology' had provided a cornerstone of Western industrialization, and was widely prescribed as a model with which to bring about economic development in some of the world's poorer regions.

Schumacher profoundly disagreed. Although the transmission of technology led capitalist industrialization had some role to play in many so called less developed countries (LDCs), modern capitalist development had generally taken the form of small dynamic islands sur rounded by massive stagnant seas of poverty, despair, and unemployment. A classical 'dual economy' had emerged, with a technologically advanced modern sector accessible only to a small and privileged minority, and a swollen traditional sector which provided a means of survival but little else for the great mass of the population. Set against a backdrop of a poor and predominantly rural population, large scale distress migration to the cities, mass unemployment and underemployment, limited domestic availability of capital, and a limited indigenous pool of skills and expertise, Schumacher argued that a labor displacing development strategy centered on the import ation of technology and investment resources, and the promotion of capital intensive Fordist mass production systems was wholly inappropriate to the stage of devel opment, the current capabilities, and the social and welfare needs of most LDCs. In his characteristically polemical style, Schumacher wrote of a 'process of mutual poi soning', whereby modern industrial development in the cities destroyed (through competition based on greater efficiency and sophistication) the economic structure of the (rural) hinterland, which 'took its revenge' on large urban areas through mass migration which rendered the cities 'utterly unmanageable'. Schumacher was also pres cient in his view that this modern technology driven de velopment strategy was 'on a collision course with nature'.

Schumacher advocated a people focused development strategy based on immediate needs, capabilities, and circumstances, and appropriate technology choices. He argued for autonomous and sustainable development which built up locally the capacity for education, or ganization, discipline, and ingenuity in an evolutionary manner, rather than being created or borrowed from afar through a process of technological 'leapfrogging'. Schumacher's writing on technology was profoundly in fluenced by Mahatma Gandhi's ideas of *swadeshi* (local ized production using local resources and mainly for local markets), and was further reinforced by a visit to Burma in 1955, which stimulated his thesis on an anti materialist, antiacquisitive, spiritually centered, and nonviolent Buddhist economics. Schumacher cham pioned 'production by the masses' ahead of 'mass pro duction', and advocated a rural, decentralized, and employment generating (labor intensive) bias in devel opment policy as opposed to an urban, industrial , cap ital , and technology driven strategy that relied on an uncertain process of trickle down to deliver doubtful developmental benefits to the mass of the population.

Schumacher was not antitechnology, nor was he ad vocating a regressive development strategy. His notion of 'intermediate technology' represented an appropriate point on a continuum between basic, indigenous tech nology (which he characterized symbolically as a £1 technology in a relative calculus of equipment costs per workplace), and sophisticated, modern technology (£1000 technology). The gap between the two was

Table 1 The main characteristics of orthodox and alternative development

	Orthodox	*Alternative*
Technology	Modern	Alternative
Scale	Large	Small
Cost	Expensive	Inexpensive
Energy input	Large	Small
Ecological impact	Unsound	Sound
Pollution rate	High	Low
Centralization	Centralist	Decentralist
Sophistication	Complex	Simple
Production focus	Capital intensive	Labor intensive
Relation to human needs	Incompatible	Compatible

Adapted from Varma, R. (2003). E. F. Schumacher: Changing the paradigm of bigger is better. *Bulletin of Science, Technology and Society* 23(2), 114 24.

enormous; the introduction of the latter not only tended to kill off the former, and dramatically reduce its employment and income earning capacity, but it also limited the scope for endogenous technological evolution. Intermediate technology was characterized symbolically as a £100 technology – more advanced than traditional but less sophisticated than modern, but in reality very variable according to local conditions – which was more affordable, more in tune with the setting and context, and more likely to raise productivity without reducing employment. Schumacher presented intermediate technology not only as an alternative (i.e., non Western) model of development but also as a philosophical idea about a culture of production. The main features of Schumacher's alternative to mainstream Western development are summarized in **Table 1**.

Intermediate and Appropriate Technology

Schumacher used the phrases 'intermediate technology' and 'appropriate technology' more or less interchangeably. There is no universal definition of 'appropriateness' – there are social and environmental, as well as economic, dimensions to be considered, and these will be weighted variously by different interest groups. At the time of his writing, a form of production technology that was an improvement on indigenous technology but did not have the labor displacing, small firm destroying, and capital consuming character of advanced Western production methods was deemed 'appropriate' in most LDC contexts, although the appropriate level of technological sophistication varied by setting and circumstance. Appropriate technology would bring the optimum economic, social, and environmental returns to the community concerned, set against local resource endowments and prevailing conditions. It would utilize local materials and power sources; would involve local people as much as possible in the development and introduction of improved technologies; use existing or easily developed skills and require

Figure 1 Appropriate technology in Northeast Thailand. © Michael Parnwell.

only limited amounts of investment; would enhance employment opportunities and require limited movement of labor; would minimize local, social, and cultural disruption; cause little environmental damage; and would ensure that the main economic benefits go to local people in greatest need.

Figure 1 provides a good example of appropriate technology: the vehicle depicted, on a street side in Northeast Thailand, is a locally manufactured vehicle which farmers use to take their farm produce to market, and which uses for its motive power an engine which also functions as a water pump and an electricity generator. **Figure 2** contrasts traditional ploughing technology in Burma with the intermediate technology of the power tiller in Indonesia and a tractor in Thailand: each technology may be considered 'appropriate' or 'inappropriate' depending on the local setting and circumstances.

Intermediate technology can derive from a number of sources. It can be taken 'off the technological shelf' by giving poorer or more remote parts of the world access to (perhaps outmoded developed world) technologies that are available elsewhere. It might take the form of improvements or 'upscaling' of traditional technologies, or the blending of traditional and imported technologies (e.g., micro hydropower technologies which generate energy for small communities), or the modification (downscaling) of imported technologies. Or it may center on the local or domestic development of original and new technologies for specific production, livelihood, or lifestyle purposes.

Examples of Intermediate Technology

There are many examples that could be used to illustrate intermediate technology as Schumacher may have

Figure 2 Ploughing technologies in Southeast Asia: traditional, intermediate, and advanced. © Michael Parnwell.

envisaged it (e.g., in agricultural production, manufacturing, energy generation, domestic life, pharmaceuticals, transportation, recycling, etc.), but we will restrict ourselves to just one: the promotion of improved cooking stove design in developing countries. This is chosen because it provided quite an important focus of a group formed by Schumacher to disseminate appropriate forms of technology to the poorer regions of the world – the Intermediate Technology Development Group (formed in 1964, and now called Practical Action) – under its Stoves Project. It allows us to show how any form of technological development must be fundamentally rooted in its prevailing social, cultural, economic, and environmental contexts if it is to be both appropriate and successful.

An improvement in cooking stove technology, while on the face of it quite a small step and a basic process, has considerable potential to make a difference to the health, quality of life, and livelihoods of a great many people in the poorer parts of the developing world. Problems with traditional stove technology (typically the three stone hearth – a fire made between three large stones which are used to support a cooking pot) include: fuel–heat conversion inefficiency, and thus high fuel consumption, which has either cost (purchase) or time (collecting) implications, and may contribute to a decline in biomass resources and associated environmental problems; health hazards from smoke and poisonous gas, as well as fire hazards; inconvenient and time consuming with regard to lighting and attending, and difficult to regulate heat output. There are many potential solutions to these problems – improved stove design, use of different stove construction materials, fuel substitution and fuel processing, ventilation through hoods and chimneys, etc. – many of which may appear quite basic, but without reference to the situation within which they are placed they may prove to be inappropriate to local people's needs, capabilities, and preferences, and thus uptake levels.

Although traditional cooking technology has the advantage of zero or negligible cost, use of widely available fuels, and zero maintenance, it is inefficient, wasteful, difficult to control and use, it has a high risk of scalding and burning, and smoke may cause respiratory problems. Many improvements on the open fire have been developed, ranging from simple charcoal stoves through multipot chimney stoves up to gas burning hobs and ovens. In India a bio residue gasifier stove (BRGS) has been developed locally by the Centre for Appropriate Rural Technologies at the National Institute of Engineering in Mysore, which is a significant improvement on traditional and even improved mud stoves. The traditional stoves have an energy efficiency of only around 10%, meaning that they use a lot of woodfuel, and create a great deal of smoke while generating little heat. In contrast, the BRGS consists of two metal cylinders, one placed inside the other (see **Figure 3**). Agricultural waste such as coconut fronds, bagasse (the waste from cane sugar manufacture), groundnut shells, rice husks are placed in the outer section, and small pieces of fuel wood in the inner section, which is then lit using kerosene. One load of fuel will burn smokelessly for about 3 h at a 40% efficiency rate. The waste agricultural residues in the outer chamber can be then reused as fuel. This stove is a dramatic improvement on the traditional stove, and obviates the need for more expensive higher technology

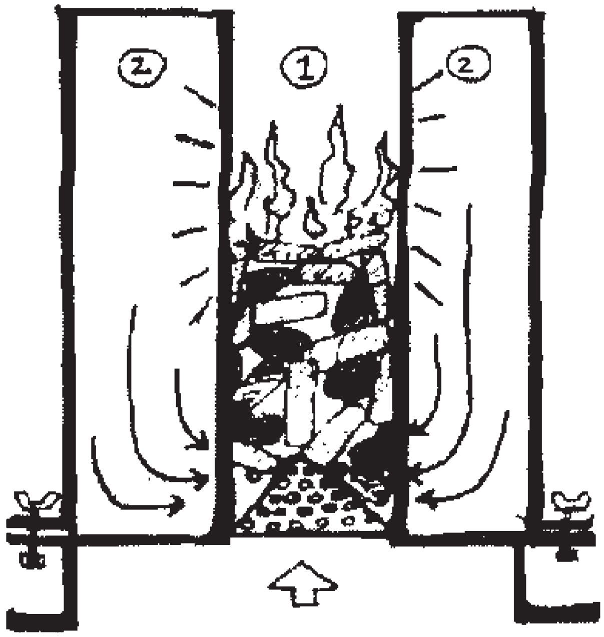

Figure 3 Cross-section of a bio residue gasifier stove. (1) The inner sleeve with small pieces of wood; (2) the outer sleeve with biomass. © *Appropriate Technology*. Source: "Energy Efficient, Smokeless Stoves Launched in India", *Appropriate Technology*, 33, 2, 55 56 (2006).

cooking solutions such as gas burners using liquid petroleum/natural gas cylinders. However, finding a technological solution to the problem of open fire cooking is not simply a matter of substituting one form of technology for another with superior features. The Intermediate Technology Development Group (ITDG) publication *Improved wood, waste and charcoal burning stoves: A practitioner's manual* identifies a wide range of criteria that have to be ascertained by means of an on site 'situation assessment' before a correct technology choice can be made. These include: the fuel types that are used and alternative sources that may be available locally; locally available materials and skills; social activities that may be associated with cooking, including the gender division of labor; user preferences and potential; awareness of and receptiveness to alternative designs and the benefits they can provide; the potential for local manufacture of alternative technologies; the cost of alternative stove technologies and their local affordability; the uses to which stoves are presently put, and the utensils that are used; frequency of use; maintenance potential for improved stove technologies; and so on. It may be necessary to incorporate several aspects of the traditional technology and accommodate various elements of the prevailing social, cultural, and environmental setting to ensure that improved designs are acceptable to the target groups. Meeting needs, matching potential, and juggling costs and benefits are thus just as important as advancing technology. Context counts.

Criticisms of Intermediate Technology

Schumacher's intermediate technology agenda has not been without its critics. It is argued to have created a rather romanticized and blinkered vision of the true situation. Thus while Gandhi and Schumacher were advocating intermediate forms of technology for rural India to protect peasants from the pernicious impact of modernization, farmers themselves were actively seeking to acquire new, improved technologies, such as were associated with the 'Green Revolution', in order to strengthen their livelihoods. In other words, outside experts sometimes presented an ideologically biased perspective on technology that went against the grain of what local people wanted, and perhaps needed. By restricting the kinds of technology that were deemed 'appropriate', local communities were cast in a survivalist role which, through the use of a relatively inferior technology which was conducive neither to efficiency nor growth, denied them the possibility of and opportunity for development. To a degree it also perpetuated the 'technology gap' between the rich and the poor, the advanced and the backward.

A second important criticism is that the advocacy of a gradualist and evolutionary approach to technology development runs against the grain of recent experience. The dramatic development of East and Southeast Asia from the late 1970s onward, much of it predicated upon precisely the processes that Schumacher had been so critical of – foreign capital investment, Western industrial transplants and their (relatively) advanced technology, export oriented industrialization, rapid urbanization, significant rural–urban migration – suggested either that Schumacher got it wrong or that his philosophy was time specific to a stage of serious underdevelopment. Rapid export oriented manufacturing development led to the dramatic transformation of several Asian economies (Japan, Taiwan, South Korea, Hong Kong, Singapore, Malaysia, Thailand, China), and quickly soaked up large pools of hitherto underutilized rural and urban labor, as well as accelerating the demographic transition. While the early stages of these countries' industrialization utilized relatively labor intensive production methods, because cheap labor had drawn foreign firms and foreign investment to East and Southeast Asian shores, subsequent development has increasingly been built on these countries' ability to acquire, assimilate, and learn from imported modern technology, and to innovate and advance their domestic technological capabilities. Their development has resulted from an openness to advanced foreign technological knowledge, not a denial of it as advocated in the intermediate technology approach.

A third criticism is that the focus on technology tended to downplay the influence of other contextual factors (e.g., social, cultural, political, capitalist relations, inequality, power, repression, disenfranchisement) on the developmental outcomes – such as unemployment, in debtedness, dependency – that intermediate technology was intended to address. Many of these problems had little specifically to do with technology choices, and thus suggested the need for social and political change to precede technological change, as the latter is very un likely to bring about the former.

A final criticism is that the advocacy of a gradualist and evolutionary approach to the development of technology denies many parts of the world the opportunity to access the environmental, social and economic benefits that ad vanced technology has the potential to deliver. These technologies are often superior, safer, cleaner, and more efficient than the outmoded technologies they replace, and a strategy of inferiorization thus runs counter to objectives such as reducing the environmental external ities associated with development. Accelerating rather than retarding the technology transition, and finding ways of speeding up and making more complete the process of technology transfer is thus advocated as the way forward. Late developers have the enviable opportunity of 'tech nology leapfrogging', allowing selective use to be made of available technologies without making the same mistakes that may have been encountered in the evolutionary de velopment of technology in economically more advanced regions, and also avoiding the conservative influence of vested interests which may otherwise retard technology development. Developing countries can, for example, leapfrog to greener production technologies.

Technology Leapfrogging

There are many examples of where technology leap frogging has brought significant benefits to recipient communities in some of the poorer regions of the world. These benefits include improved convenience and qual ity of life, standard of living, and livelihood. Two short illustrations will suffice here.

Grameen Telecom (GT) is a nonprofit organization in Bangladesh that has close connections with the world famous Gremeen Bank, a nongovernmental organization (NGO) that specializes in providing micro finance for micro enterprise development among the country's poor (especially women) to enable them to diversify their economic activities and build stronger livelihoods. GT provides mobile telephones to poorer rural people and communities as part of the government's Village Phone Program. This program is designed to give rural people access to communications technology and the infor mation revolution in a country where the landline system is woefully inadequate, not least in more remote rural areas. In addition to providing a source of income to the 'phone ladies' who lease the phones from GT and operate the system by charging fellow villagers to use the service, the program enables villagers to keep up to date with commodity prices and thus avoid being cheated by merchants, and to keep in touch with family members who have migrated to work in other parts of the country. Thus a cutting edge technology has been used to cir cumvent deficiencies in a less advanced technological system to bring definite social and economic benefits to poor rural communities.

One of the problems with the GT system concerns the charging of cell phone batteries in communities that have not yet been connected to the national electricity grid. Another advanced technology – solar photovoltaic (SPV) – has been used to overcome this problem and provides our second illustration. There are still many parts of the developing world which have either not been connected to national electricity grid systems or where the level of energy demand or the dispersed nature of the population does not make such a large scale system economically viable: 1 billion people in Asia do not have access to electricity. As such rural communities are de prived of the convenience and opportunities that is provided by an electricity supply that most areas of the world now take for granted. Instead they have to rely on alternative, often far less convenient sources of energy and lighting, such as biomass, kerosene, candles, battery powered torches, etc., or on diesel powered generators which typically consume significant amounts of imported petroleum. Where – as in many parts of the developing world – sunshine is in plentiful supply, SPV technology represents a promising high technology solution to local energy problems, especially where appliances (e.g., do mestic lighting, radio, TV, refrigerators, water pumps, vaccine storage, telecommunications, etc.) can be run off 12 volt batteries which SPV cells are typically used to charge. SPV technology can be used to power solar lighting kits for community centers, for instance, which have been used to help improve literacy levels and other education and training outcomes. In addition, SPV technology helps to reduce consumption of biomass and fossil fuels, and thus makes a contribution to the ameli oration of global warming.

At present SPV technology is relatively expensive, and cannot compete with commercial energy systems, but in more remote locations it often represents the best, and most cost effective option, especially if it is introduced as part of international or local development assistance programs, or where governments help reduce the cost of the technology by waiving import duties and sales taxes because of the technology's importance for local devel opment. Nonetheless, the very poor may still be excluded from this technological fix as a result of their lack of collateral to obtain the necessary equipment on credit, or to pay the running/electricity costs to the energy pro viders. There is 1.5 MW of installed SPV capacity in

Mali, for example, but much more potential has been identified, and government policy is now giving SPV technology a high priority. Some 1416 photovoltaic bat tery charging stations (PVBCS) had been installed in villages in Thailand by the late 1990s (56% in the upland parts of the northern region), giving people a local source of electricity when previously they had had to take their batteries into far away towns for recharging once a week. In a survey, more than two thirds of respondents con sidered their quality of life to be 'better' or 'much better' as a result of this relatively high technology response to a local development problem.

The Future

The pendulum appears to have swung away from the idea of controlling technological advancement in poorer countries and regions in order to protect jobs, limit capital leakage, and strengthen the capacity for en dogenous technology development, to a strategy of ac celerating the North–South transfer of appropriate forms of technology and the capacity of poorer countries to absorb and make use of it. As such, Schumacher's advo cacy of intermediate technology appears to have been relevant to a particular moment in the history of modern development and a specific segment of the global econ omy. Nonetheless, his philosophy has taken on a much wider resonance in recent years as more and more people have started to question the destination that a technol ogy driven path is taking the world in social, economic and, most particularly, environmental terms. Technology has delivered amazing advancements and improvements, but has also wrought incredible destructive power. Humankind's ability to overconsume natural resources, pollute the biosphere, and threaten future development has been facilitated by the cavalier and often in appropriate application of modern technology.

These problems have stimulated a reaction which is gaining in scale and momentum across the globe, and which has been strongly influenced by Schumacher's philosophical position, *inter alia*, on the scale and capital intensity of production, our complacent attitude and behavior toward resources and the natural environment, the importance of building according to context and capability, and the need for people centered and self reliant forms of development. Described as the 'father of the Green movement', Schumacher's influence can be identified in the current preoccupation with sustainable development, calls for a radical ecological economics, and movements against materialism, overtechnification, developmentalism, and globalism which have much in common with the Buddhist economics that Schumacher espoused.

See also: Development I; Green Revolution; Nongovernmental Organizations; Sustainable Development.

Further Reading

Anon (2006). Energy efficient, smokeless stoves launched in India. *Appropriate Technology* 33(2), 55 56.
Francis, A. J. and Mansell, D. S. (1988). *Appropriate technology for developing countries*. Blackburn: Research Publications.
Grieve, R. H. (2004). Appropriate technology in a globalizing world. *International Journal of Technology Management and Sustainable Development* 3(3), 173 187.
Gullberg, M., Ilskog, E., Katyega, M. and Kjellstrom, B. (2005). Village electrification technologies: An evaluation of photovoltaic cells and compact fluorescent lamps and their applicability in rural villages based on a Tanzanian case study. *Energy Policy* 33, 1287 1298.
Kaplinsky, R. (1990). *The economies of small: Appropriate technology in a changing world*. London: Intermediate Technology Publications.
Schumacher, E. F. (1973). *Small is beautiful: A study of economics as if people mattered*. London: Abacus.
Schumpeter, J. A. (1934). *The theory of economic development: An inquiry into profits, capital, interest, and the business cycle*. Cambridge: Harvard University Press.
Sriuthaisiriwong, Y. and Kumar, S. (2001). Rural electrification using photovoltaic battery charging stations: A performance study in northern Thailand. *Progress in Photovoltaics: Research and Applications* 9, 223 234.
Stewart, B. (1987). *Improved wood, waste and charcoal burning stoves: A practitioner's manual*. London: Intermediate Technology Publications.
Varma, R. (2003). E.F. Schumacher: Changing the paradigm of bigger is better. *Bulletin of Science, Technology and Society* 23(2), 114 124.

Relevant Websites

http://www.grameenphone.com
Grameenphone.
http://www.grameen info.org/grameen/gtelecom/
Grameen Telecom.
http://www.itdg.org/
Intermediate Technology Development Group.
http://www.practicalaction.org/
Practical Action.
http://www.schumachersociety.org/
The E.F. Schumacher Society.
http://www.grameenfoundation.org/
The Grameen Foundation.
http://www.schumacherfoundation.org/
The Schumacher Foundation.

Intermodality

A. R. Goetz, University of Denver, Denver, CO, USA

Glossary

Containerization Process by which steel boxes are used to store, protect, and handle cargo being transported and easily interchanged between trucks, trains, and ships without rehandling contents.

Double Stacking Practice whereby containers are placed on top of one another on a flat or recessed railroad car that improves capacity.

Land Bridge Using railroads to ship containers from one coastal port to another, thus facilitating the maritime movement of containers; most often used in the context of shipping containers across North America from west coast to east coast ports.

Piggyback A truck–rail intermodal service where a truck trailer with wheeled chassis is loaded onto a railroad flatcar for transport; also known as trailer-on-flatcar (TOFC).

Post-Panamax A generation of container ships that are built with dimensions larger than can be accommodated by the Panama Canal.

TEU (Twenty-Foot Equivalent Unit) The standard measure of container traffic based on 20 ft long containers.

Introduction

Precisely defining intermodal transportation can be a tricky endeavor. There are numerous definitions of the term, each with nuances that may be applicable in particular contexts. A basic definition of intermodality is transportation by more than one form of carrier during a single journey. In this sense, virtually all transportation historically could be classified as intermodal since the use of more than one mode is typical, such as when the first sailing ships were loaded with cargo taken from horse drawn carts. Sometimes the terms 'multimodal transport' or 'combined transport' are used to describe transport involving two or more modes. A somewhat broader definition of intermodal transport from Gerhardt Muller of the US Merchant Marine Academy is "the transportation of passengers and freight on two or more different modes in such a way that all parts of the transportation process, including the exchange of information, are efficiently connected and coordinated." This second definition is more descriptive of the larger meaning associated with intermodalism, particularly regarding the efficiencies that accrue as a result of connected and coordinated services. Key differences include the ease of transferability through the use of new technologies and organizational control, and the increased reliability of on time, door to door services geared to the needs of customers. Simple transfers of freight or passengers from one mode of transport to another had come to be viewed largely as separate activities from the perspective of the individual modes of transportation. An intermodal approach sees the whole process of moving people or goods as a connected system working together as opposed to separate modal systems working independently and/or competitively.

Broader definitions of intermodalism have been proffered by practitioners and academics. Former US Federal Highway administrator Gloria Jeff defined intermodalism as "a system that is both safe and efficient, and productive and flexible, in responding to the needs for goods movements and offers people choices and flexibility in their personal movements; the system must also be international, intelligent, and inclusive." An even broader definition was proposed by University of Denver Professor, Joseph Szyliowicz, who incorporated the concepts of sustainability and ethics in his definition: "An intermodal system is one in which the individual modes are linked, governed, and managed in a manner that creates a seamless and sustainable transportation system; such a system should be economically efficient, environmentally sound, safe and secure, and ethically based." Each of these definitions in turn broadens the intermodal concept to embrace a much larger meaning beyond simply 'transportation involving more than one mode'.

Brief History

The intermodal approach to transportation has become more widely recognized and adopted by private sector providers of transportation, governmental agencies, and the larger transportation community. The private sector has led the way in the adoption of intermodal technologies within the freight transportation arena, most notably the pioneering efforts of Malcolm McLean and Sea Land Services in the 1950s in spearheading the container revolution. The era of containerization officially commenced in 1956 when the ship *Ideal X* sailed from New Jersey to Texas with a load of containers that could be transported on ships, rail cars, or trucks. At about the same time, railroads developed specialized 'piggyback' services featuring truck trailers (with wheeled chassis) on flatcars [TOFC] and later, containers on flatcars [COFC]. In 1964, the International Standards

Organization (ISO) adopted a standard for container size of 8 ft width and height, and 20 ft length, thus establishing the twenty foot equivalent unit (TEU) as the standard measure of containerized freight. Today, both 20 ft and 40 ft containers are used extensively.

Starting in the early 1980s, shipping lines and rail roads began to experiment with 'double stacking' con tainers on rail cars to increase volume throughput. In 1984, the shipping giant American President Lines (APL) through its Stacktrain unit pioneered double stacking of containers on rail flatcars that dramatically improved the capacity and productivity of landside container transport. Soon thereafter, other shipping lines, such as Sea Land, K Line, Maersk, and NYK developed their own double stacking operations in conjunction with the major railroads. Many railroad bridges and tunnels had to be rebuilt subsequently to accommodate the increased height requirements of double stacked containers. The success of containerization and double stacking in the railroad industry has led to the development of 48 ft and 53 ft domestic containers, in addition to the standard 20 ft and 40 ft international containers. These larger domestic containers make it easier for railroads to interface with trucks, since these containers conform to more standard truck trailer lengths.

Deregulation in the US and other states from the early 1980s played a major role in facilitating inter modalism by making it easier for transportation com panies to coordinate services across modes. The success of integrated package delivery firms, such as FedEx, UPS, and DHL, has underscored the value of the intermodal concept. More recent innovations in electronic data interchange (EDI) and the development of specialized logistics and supply chain management firms (third party providers) have facilitated the development of intermodalism. Today, intermodal operations are part and parcel of the maritime, railroad, trucking, and air cargo businesses, and have resulted in a much more in tegrated and efficient freight transportation system. An intermodal approach has resulted in increased profit ability, better utilization of existing infrastructure, re duction in need of additional capacity, fewer environmental externalities, lower prices, and better customer service.

While freight intermodality pertains to how goods move, passenger intermodality is focused on moving people within and between the major passenger transport modes: in cars on roads and highways, by airplane, in buses, by rail transit, in water ferries, by bicycle, and as pedestrians. In a strict sense, most personal transportation is already intermodal in character. Whether walking to one's car or to a bus stop, or taking a train to get to the airport, there usually are two or more modes of travel involved in most trips. But again, the concept of pas senger intermodalism goes beyond just involving two or more modes to emphasize the ease of connections be tween the modes that relies on seamless coordination. In North America, passenger intermodalism has not been developed nearly as well as freight intermodalism, due mainly to the overwhelming dominance of the auto mobile in personal transportation. In Europe and Asia, by contrast, passenger intermodalism is more advanced be cause passenger transit systems are more extensive and better coordinated with each other.

Governmental transport agencies in North America, Europe, Asia, and around the world have reacted to the intermodal revolution in freight transportation by in creasingly adopting and promoting it as official policy in both freight and passenger transportation. In the US, the Intermodal Surface Transportation Efficiency Act (ISTEA) of 1991 was the first major piece of federal transportation legislation that did not contain the word 'highway' in its title. It signaled a shift in thinking about transportation away from individual modal per spectives to a more integrated intermodal approach. Since government is responsible for providing much of the infrastructure used in both freight and passenger transportation worldwide, government recognition of intermodalism has been an important step in the devel opment of the larger system, particularly for the pas senger sector.

The Intermodal Concept

As the term intermodal has become more widely used, its exact meaning has become more difficult to pinpoint. The intermodal concept has proven to be somewhat amorphous, as different individuals tend to have widely differing notions of what intermodal really means. Its definition oftentimes hinges on the differentiation be tween 'multimodal' and 'intermodal', or between 'pas senger intermodalism' and 'freight intermodalism'.

Multimodal versus Intermodal

One distinction that is often made is between the terms 'multimodal' and 'intermodal'. To some, multimodal and intermodal are interchangeable terms; while to others, the distinction is important. Strictly speaking, multi modal means 'more than one mode', while intermodal suggests 'connections between modes'. Thus, multimodal transport may involve two or more separate transport systems, thus providing more system choices, while intermodal transport emphasizes the most efficient way of moving from point to point within an integrated trans portation system. Multimodal transport can involve adopting a generic, non mode specific approach to problem definition and problem solving. In intermodal transport, key interactions between modes (including transfers) and also policy and service interactions are paramount. Some feel that intermodal transport, strictly focused on connections between modes, represents a subset of the much larger field of multimodal transport, while others feel that intermodal transport is the broader field, and that multimodal transport is a necessary pre requisite to the more comprehensive and better con nected intermodal system.

This latter conceptualization was expressed more formally by Robert Martinez of the US Department of Transportation (USDOT) Office of Intermodalism who introduced the following three part definition:

- choice among transportation options provided by competing modes, independently or in combination (choice also means that decision makers need to consider alternative systems to address transportation needs before investing in infrastructure);
- connections that provide convenient, rapid, efficient, and safe transfer of people or goods from one mode to another (including end point, pickup, and delivery) during a single journey to provide the highest quality and most comprehensive transportation service for cost; and
- coordination and cooperation among transportation organizations to improve transportation service, quality, safety, and efficiency across all modes or combinations of modes in an environmentally sound manner.

This definition of intermodalism has been used by the USDOT and also by numerous research studies investi gating intermodal planning practices. Other studies prefer to use multimodal as the key word to describe a similar process, while still others use both key words together. In a European context, terms such as 'combined transport', 'integrated transport', or 'interoperability' are used to describe similar processes.

Freight Intermodal versus Passenger Intermodal

Another important distinction is between freight and passenger intermodal transport. Growing interest in both freight and passenger intermodalism in government and the private sector has created some confusion when using the term intermodal. The intermodal concept was ini tially developed within the freight industry. To many in the private sector and some in government, intermodal transport means freight intermodal transport. To others, mostly in local government and planning, intermodal transport means passenger intermodal transport. As a result, it becomes necessary to specify which type of intermodal transport one is discussing. The following sections illustrate the contemporary landscape of both freight and passenger intermodalism.

Contemporary Freight Intermodalism

Intermodalism, and especially containerization, has pro foundly affected all modes of freight transport, but it has probably had the greatest effect on the maritime industry. Loading and unloading ships used to be a very labor intensive activity performed by gangs of dockers at ports. Containerization has facilitated the use of large special ized gantry cranes that have dramatically improved the speed of handling. A 40 000 ton container ship can be loaded or unloaded today using less than 750 man hours, while traditional methods would require 24 000 man hours. This reduction in the amount of workers required on the waterfront has had significant labor ramifications in an industry where unions have been historically strong. Protection of cargo from damage or loss, and overall security has been enhanced through containerization.

The economic success of containerization has led to the construction and utilization of ever larger container ships, the largest of which today have capacities ex ceeding 8000 TEUs. The largest container ships are classified as 'post Panamax' because their large dimen sions do not permit them to cross through the Panama Canal, or sometimes 'Mallacamax' referring to the size limit of ships passing through the Strait of Malacca near Singapore. The economies of scale engendered by con tainerization, as well as by ocean shipping deregulation and economic globalization, have resulted in a more concentrated maritime industry, characterized by a small number of global strategic alliances (**Figure 1**).

Ports have been significantly affected by container ization due to the need for significant investments in gantry cranes and for greater berthing and channel ca pacity to accommodate larger ships and increased con tainer storage in port areas. These economic pressures have also intensified inter port competition. There are several major 'ranges' or geographic regions featuring major ports that container ships serve, such as the Far East (Hong Kong, Singapore, Kaiosiung, Pusan, Kobe, and Yokohama), West Coast North America (Los Angeles/Long Beach, Oakland, Vancouver, Seattle, and Tacoma), East Coast North America (New York–New Jersey, Baltimore, Hampton Roads, Charleston, and Halifax), and Northwest Europe (Rotterdam, Hamburg, Antwerp, Bremen, and Le Havre). Today, the largest container flows are from the Far East to the West Coast of North America, reflecting the huge growth of trade in manufactured goods from China and East Asia to the US. Projections indicate that container traffic across the Pa cific will double over the next 20 years. Intermodal container traffic has been increasing at ports throughout the world, including in the Caribbean and Mediterranean Seas, as well as in ports in Australia and Oceania, Latin America, Africa, the Middle East, and South Asia.

Figure 1 A post-Panamax container ship. Courtesy: NYK Lines.

Integrating land and sea movements of containers as part of a through transport concept is a critical development in freight intermodalism. Previously separate and highly competitive, the railroad and trucking industries have become much more integrated with each other and with the maritime shipping industry. The railroad industry, in particular, especially in North America, has been radically transformed as a result of containerization, double stacking, deregulation, and integrated services (**Figure 2**). Spurred by the rapid rise in the use of containers by the shipping industry, railroads in North America began in the 1980s to dedicate blocks of trains to move double stacked containers from West Coast ports to inland terminals, especially Chicago. The use of railroads to move containers across North America has also created a form of ‘land bridge’ between Asia and the East Coast of North America or Europe. The land bridge concept became more popular once post Panamax container ships began service. Most intermodal containers on rail are unloaded at inland centers or switching yards, and transferred to trucks for final distribution. Today, the intermodal shipment of containers accounts for a large and rapidly growing segment of business for each of the five major North American railroads – the Burlington Northern Santa Fe (BNSF), Union Pacific (UP), CSX, Norfork Southern (NS), and Canadian National (CN) – and has become an important component of some of the largest trucking firms, such as JB Hunt or Schneider Transport. In Europe, intermodal development of railroads has been hindered by the shorter land distances that are better suited to truck transport, and by problems of interoperability between national rail systems. There are, sometimes, loading gauge problems for container trains as well as capacity problems on rail networks which are much more heavily used for passengers in Europe than in the US. Also, there is a greater emphasis on the use of inland barges on the Rhine and other river systems in Europe. Landside intermodalism in Asia is limited by the insular geography of much of the region, but significant investments in rail and road infrastructure in China may lead to a much more integrated system there. Australia and New Zealand are also developing a more coordinated intermodal system. As the intermodal concept spreads through other regions of the world, more efficient landside connections will be improved in conjunction with regional economic development.

Air cargo transport has benefited from containerization through the use of specialized containers that fit within the rounded fuselage of wide bodied aircraft. Integrated carriers, such as Fed Ex, UPS, and DHL rely on the speed of air cargo and the improved transferability of freight to other modes to build highly successful package and express freight delivery operations.

Contemporary Passenger Intermodality

Imagine a future whereby an individual takes a trip from Meridian, Mississippi, USA to Vouray, France, using a combination of modes, including local transit, intercity bus, intercity high speed rail, and international air in which only one through ticket has been purchased and baggage is picked up at the origin and delivered to the final destination. This vision describes the essence of what intermodal passenger transportation is all about: seamless, efficient, environmentally sound, and safe

Figure 2 Double-stacking containers on a railroad flatcar using an overhead gantry crane. Courtesy: Pacer Stacktrain.

transportation for travelers moving from point to point throughout the world.

When Fed Ex, UPS, or DHL delivers a package from Manassa, Colorado, USA to Mandya, India, several modes of transportation are involved, but the customer calls one company and the delivery occurs with a guaranteed time of arrival. But if a passenger wishes to make the same trip, that person would have to contact at least four separate transportation providers to schedule the trip at a relatively high cost with no assurances of actual time of arrival. This example illustrates important differences between the freight and passenger transportation sectors. To be fair, it should be recognized that it is easier in most cases to ship packages rather than people. Packages do not talk back, and do not mind being stacked up in warehouses or on trucks. Still, given the importance of transporting people, intermodal passenger performance can be improved to at least approximate the efficiencies already achieved in the freight sector.

The vision of a seamless, integrated passenger system can be made more focused by considering best practices in places where passenger intermodalism has been more successful. The infrastructure of some European and Asian passenger systems is better suited to intermodalism as the individual systems of intercity rail, bus, and local transit are themselves better developed than in North America or other parts of the world. But Europeans and Asians have also made important strides in linking these systems together so that connections are much more convenient to the traveler. The classic examples are the intercity rail stations with direct access to national high speed networks located in the lower levels of major European airports, including Amsterdam, Copenhagen, Frankfurt, Milan, Oslo, Paris, Stockholm, and Zurich. With the development of the Trans European High Speed Rail Network, an international flight to any of these major airports provides convenient access to numerous cities and towns throughout Europe. Other airports, such as Heathrow and Gatwick in London, provide rail access only to local area rail terminals, whereupon travelers can be connected to a national rail system. Once reaching the desired city, the more ubiquitous local rail transit, streetcar, or bus systems can take travelers close to their final destinations. Good air–rail connections can also be found in Tokyo, Osaka, Seoul, Hong Kong, Singapore, and Shanghai which features the world's first and only operational magnetic levitation (Mag Lev) train line. Improving air–rail connections in North America include New York (JFK, LaGuardia, and Newark Airports), Washington, DC (Reagan National and Baltimore, Washington), Chicago (Midway and O'Hare), Atlanta, Boston, Cleveland, Fort Lauderdale,

Figure 3 Intermodal passenger operation featuring rail service to an airport. Air Train-JFK connects Kennedy Airport's eight terminals with the New York City subway system and the Long Island Railroad. Courtesy: Air Train-JFK.

Miami, Minneapolis, Philadelphia, Portland, Providence, St. Louis, and San Francisco. Most of these US air–rail connections are to local rail systems, not regional or national (**Figure 3**).

Passenger intermodalism is facilitated through the use of intermodal passenger stations and terminals that serve as a focal point for rail, bus, taxi, ferry, bicycle, and pedestrian transport. These intermodal stations and terminals play a critical role in providing enhanced connectivity throughout the entire system by serving as the linchpins of an integrated passenger transport system. In most instances, the central railway terminals serve as the main intermodal stations for cities, but subsidiary intermodal centers at airports, seaports, and other strategic locations on the intermodal network are also quite important. Many outlying rail stations around major cities, for example, have large automobile park and ride lots that serve as an intermodal transfer spot between cars and trains.

Conclusion

The intermodal concept is revolutionizing both freight and passenger transport around the world. Relying on innovations such as containerization, double stacking, and the through transport concept, the speed and reliability of goods movement involving the maritime, railroad, trucking, and air cargo industries has been greatly enhanced. Passenger intermodalism has been slower to develop than on the freight side, but significant progress in establishing more and better air–rail linkages and better connected local and regional transit systems through intermodal passenger terminals and stations herald a brighter future.

See also: Logistics; Railways; Transport and Globalization.

Further Reading

Boske, L. B. (1998). Multimodal/intermodal transportation in the United States, Western Europe, and Latin America: Government policies, plans, and programs. *Lyndon B. Johnson School of Public Affairs, Policy Research Project Report 130*. Austin, TX: University of Texas.

Carmichael, G. (1997). An overview of the 21st century North American intermodal system. *Transportation Law Journal* 25(3), 321.

Coogan, M. A. (1993). Airport access: Case study in intermodalism. In *Intermodal Planning: Concept, Practice, Vision, Special Report 240*, pp 90 91. Washington, DC: Transportation Research Board, National Research Council.

Coyle, J. J., Bardi, E. J. and Novack, R. A. (2006). *Transportation* (6th edn.). Mason, OH: Thomson South Western.

Crain and Associates, Inc. with Pacific Consulting Group. (1996). Institutional barriers to intermodal transportation policies and planning in metropolitan areas. *Transit Cooperative Research Program Report 14. Transportation Research Board, National Research Council*. Washington, DC: National Academy Press.

Goetz, A. R., Szyliowicz, J. S., Vowles, T. M. and Taylor, G. S. (2004). *Assessing Intermodal Transportation Planning at State Departments of Transportation*. Denver, CO: National Center for Intermodal Transportation.

Goetz, A. R. and Vowles, T. M. (2000). Progress in intermodal passenger transportation: Private sector initiatives. *Transportation Law Journal* 27(3), 475 497.

Levinson, M. (2006). *The Box: How the Shipping Container Made the World Smaller and the World Economy Bigger*. Princeton, NJ: Princeton University Press.

Martinez, R. (1993). Perspective from the office of intermodalism. *ISTEA and Intermodal Planning: Concept, Practice, Vision, Special Report*

240, pp 47 54. Washington, DC: Transportation Research Board/ National Academy Press.

Muller, G. (1999). *Intermodal Freight Transportation* (4th edn.). Washington, DC: Eno Transportation Foundation and Intermodal Association of North America.

Peyrebrune, H. L. (2000). Multimodal aspects of statewide transportation planning. *National Cooperative Highway Research Program, Synthesis of Highway Practice 286. Transportation Research Board, National Research Council.* Washington, DC: National Academy Press.

Rodrigue, J. P., Comtois, C. and Slack, B. (2006). *The Geography of Transport Systems*. London: Routledge.

Slack, B. (1998). Intermodal transportation. In Hoyle, B. & Knowles, R. (eds.) *Modern Transport Geography* (2nd edn.) pp 263 289. Chichester: Wiley.

Szyliowicz, J. (2000). Intermodalism: The challenge and the promise. *Transportation Law Journal* 27(3), 299 315.

TransManagement, Inc. (1998). Innovative practices for multimodal transportation planning for freight and passengers. *National Cooperative Highway Research Program Report 404. Transportation Research Board, National Research Council*. Washington, DC: National Academy Press.

US Government Accountability Office (2005). Intermodal transportation: Potential strategies would redefine federal role in developing airport intermodal capabilities. *GAO 05 727*. Washington, DC: US Government Accountability Office.

Relevant Websites

http://www.intermodal.org
Intermodal Association of North America (IANA).

http://www.du.edu/transportation
Intermodal Transportation Institute (ITI), University of Denver.

http://www.iaro.com
International Air Rail Organization (IARO).

International Organizations

R. Peet, Clark University, Worcester, MA, USA

Glossary

Bretton Woods Meeting held in 1944 to set up an international economic regulatory regime for the post-war period.

GATT The General Agreement on Trade and Tariffs that governed international trading relations between 1948 and 1994.

Governance Code of conduct and regulation of the operations of institutions usually situated in civil society.

IMF The International Monetary Fund, an institution set up at Bretton Woods to give short-term loans to member countries experiencing balance of payments and currency crises.

Keynesianism Following the economist J.M. Keynes, this term refers to intervention by states in the running of capitalist economies through monetary and fiscal policies.

Neoliberalism Late-twentieth-century revival of nineteenth-century, free trade liberalism that advocates de-regulation, privatization, and other policies.

Washington Consensus Informal agreement reached by the US Treasury, the IMF, and the World Bank in the 1980s on a set of neoliberal policies prescribed as conditionality on loans made by the international financial institutions.

World Bank An institution set up at Bretton Woods to give long-term development loans initially to countries undergoing reconstruction after World War II, but eventually mainly to Third World and post-Communist member countries.

World Social Forum Meeting of social movements held each year usually in Porto Allegre, Brazil to discuss alternative development.

WTO The World Trade Organization, a more formal successor to GATT with extended powers to regulate intellectual property rights and trade in services, as well as international commodity trade.

Introduction

International organizations are private and public institutions operating in several or many countries. Private international organizations are typified by multinational corporations that operate in several countries, but maintain a definite base in one; and by transnational corporations, operating internationally and without particular national commitment. Public international organizations are typified by the international financial institutions, as with the International Monetary Fund (IMF), the World Bank, and the World Trade Organization (WTO) and international political, cultural, and economic institutions, typically the United Nations (UN) and its various agencies like the World Health Organization (WHO), the UN Development Program (UNDP), and the Food and Agriculture Organization (FAO). There are also several major institutions that have memberships restricted to small groups of countries, yet operate internationally, as with the Organization for European Economic Cooperation and Development (OECD) and the group of seven or eight countries (G7/8). International organizations have assumed increasing significance in the post World War II period as economies, cultures, and polities have globalized – that is, as the time needed to traverse space has decreased and the operations of institutions have increased in scale.

Theoretical Approaches

One way of looking at the relationship between the two types of international organizations, private and public, is that the public institutions regulate a global economy and polity that would be more crisis prone if it were to operate alone. A version of this that is significant in human geography is referred to as regulation theory. Regulation theorists place their conception of economic growth (regime of capital accumulation) in the context of broader sociopolitical institutional forces (mode of regulation). They see economic history as alternating phases of relatively stable capital accumulation, interrupted by sharp, punctuating restructuring crises. That is, regulationists do not see capitalism as self stabilizing, and certainly not as self healing, but see precarious stabilization being achieved by the power of agency – increasingly global agency. That is, stability has to be reestablished by institutional actions that may or may not occur, may not be completely successful, or may have unintended but significant consequences. As part of the regulation school derives from Marxism, agency can be transformative in socialist, ecological, feminist, and other directions – for regulationists, a far better economic world is possible. In this mainly French school of thought, the central concept 'mode of regulation' is understood as a 'situated rationality, illuminated by a dense network of institutions' that influences and directs the trajectory of an economy. That is, with the regulation school, economic growth is directed by an institutionally produced rationality, rather than growth being 'natural', 'endemic'. International organizations try to produce a more rational global economy.

Another aspect of the analysis concerns its level or specificity. International organizations can be looked at via an analysis intermediate between grand theories of world systems or modes of production like capitalism, and empirical details of characteristics and events, as with histories and detailed geographies. An example of this intermediate level analysis deals with institutions. Interest in institutionalism has recently been revived in a number of academic disciplines – institutional economics, for instance. In human geography, the argument is that economic activity is socially and institutionally situated rather than autonomous. Institutional economic geography looks at the shaping of space economies by environments of institutions, with one main strand of institutionalist interest focusing on social regulation, governance, and the effect of policies in shaping global, national, regional, and local economies. The main problem with the term 'institution', however, is that no one is sure what it means – organization, as with a corporation or a global bureaucracy, or values, like efficiency or benevolence? In what follows, institution initially means an organizational entity, located in a space, with a mission and declared purpose, backed by command over some kind of resource (ideas, expertise, money, and connections). But additionally, institution is used in the sense, derived from the social theorist Michel Foucault, of a 'community of experts', an elite group controlling an area of knowledge and expertise, that exerts power over spaces that extend to the global. International organizations are among the most powerful of a system of institutions involved in global governance.

Global Governance

The global capitalist economy is made up of thousands of specialized components tied physically together by flows of products, resources, people, and money, and tied together organizationally by markets governed by governmental, intergovernmental, and governance institutions – at the global scale these are international organizations. In neoliberal theory, governmental intervention into market mechanisms is thought to be unnecessary. Indeed, in extreme neoliberalism, government is said to be harmful to economic efficiency and dangerous to political liberty, because it constrains the freedom of individual action. But in practice, all market systems are embedded within institutional or organizational contexts that include state legislation, governmental intervention, and increasingly governance regulation. When market systems were primarily national in geographic scale, the necessary institutional framework could similarly be limited to the nation state. With regionalization, such as the partial integration of the West European economies, governance expanded to the international scale, as with the European Union. However, the globalization of capitalism has long required and, indeed has been led by, the parallel globalization of institutional regulation by international organizations. In the first globalization of the late nineteenth century, British national institutions precariously managed international economic relations, with sterling as agreed upon currency of account, and London as global, hegemonic center of political–economic power. Pax Britannica was challenged during two 'world wars', with Britain demoted in their aftermath to the status of sub hegemony – although remnants of the previous hegemony remain, for example, in London's strong position in global currency exchange and the international bond market. In the second globalization of the late twentieth century, US institutions have come to precariously manage international economic relations, with the US dollar as currency of account, Wall Street as main center of political–economic power, and Washington DC as center of political power. However, the second globalization learns from the collapse of the first globalization in the Depression of the 1930s that nationally based institutions cannot suffice, no matter how economically and militarily powerful Pax Americana may be. Hence the International organizations such as the IMF and World Bank, as well as international currencies, like the US dollar, the Yen, Deutschmark, Pound and Euro, elements of a spectacular array of regional, supranational, international, and truly global entities, are greatly influenced by, but only partly disciplined by, American political economic and political militaristic power.

The capital invested in global development is controlled by large financial institutions: global governance institutions, such as the IMF, the World Bank, and the G7/G8; and private banks, of various kinds, involved in global capital markets. These institutions supply much of the capital needed to finance development. They also control the financial expertise necessary for investment decisions. International financial organizations of all types, governance and private, consult with each other on a daily basis and form conglomerates to tackle investment related problems. These institutions form complexes of financial power bound together by regular exchanges of personnel and policy ideas. They exercise control, through money and expertise, over billions of people the world over, at the basic level of livelihood. Banks and other financial institutions see global development as an opportunity for earning interest on deposited funds. The global governance institutions like the World Bank, but some private financial institutions like investment banks, try to portray development financing as motivated partly by humanitarian concerns, like global poverty. Yet there remains a widespread, critical belief that global finance capital produces a form of development conducive to its own interest – making money. Global financial organizations thus become a center of controversy, where interests collide over vital issues of global development.

Origins of International Financial Organizations

The idea of establishing international financial organizations to regulate global economic development came as a response to growing problems in the years between the two world wars of the twentieth century. At first, bankers were the main actors rather than governments who, following a liberal ethos, initially tried to avoid direct intervention in explicitly financial international affairs. Far more aware of the effects of national interest rates, for example, in attracting or repelling capital, bankers advocated some kind of international cooperation. For example, the German reparation payments, made under the Dawes Plan of 1924, were negotiated by the chairmen of the boards of the New York Federal Reserve Bank and the Bank of England. There was an international organization in place – the League of Nations – but it remained primarily a 'political' organization. However, it maintained an Economic and Financial Organization, and it sponsored a series of international conferences in various European cities between 1920 and 1933. In these initial moves we can see the beginnings of the idea that international financial organizations should regulate global development.

In the mid 1920s, the League of Nations helped arrange loans to stabilize the economies of several European countries. An international economic conference, convened by the League at Geneva in 1927, and attended by several nonmember countries, such as the US and the Soviet Union, came up with a series of resolutions dealing with trade, cartels, and other issues that were thought to constitute an international code of behavior in policy matters. Discussion of a Bank for International Settlements (BIS) took place in 1930. And since then, regular meetings have been held in Basel, Switzerland, among central bank governors and experts from other financial agencies. The BIS conducts its own research in financial and monetary economics and collects, compiles, and disseminates economic and financial statistics, supports the IMF and World Bank, performs traditional banking functions for national central banks (e.g., gold and foreign exchange transactions), as well as trustee and agency functions. The main international financial institutions, however, were formed by governments through the 1944 Bretton Woods Agreement. This established permanent international organizations to promote international monetary cooperation and provide the machinery through which countries could consult and collaborate – the IMF, the International Bank for Reconstruction and Development (IBRD) (later World Bank), and an International Trade Organization (ITO) that never got off the ground. The boards of governors and executive boards of these institutions would be controlled by the countries with the largest investments (quotas).

There was one issue, later to become highly controversial, that divided the Americans and the Europeans. The European view of the lending operations of the IMF was that resources would be provided to member countries more or less on request, as they were needed. In particular, the British delegates to Bretton Woods thought that members should be free to pursue whatever domestic policies they desired, even if these affected exchange rates, a central international concern of the conference. By contrast, the Americans thought that borrowing foreign currency (dollars as it turned out) from the IMF was not an unqualified right. At a pre-conference, the US delegation proposed that language of the proposed Articles of Agreement (Article V) be changed from a "member shall be entitled to buy another member's currency from the Fund" to a "member may buy the currency of another member from the Fund" (emphasis added). The United Kingdom had the support of virtually all other countries in successfully opposing this change. However, later the US director on the IMF executive board insisted that use of IMF resources should be subject to close scrutiny to assure a country's adherence to its principles and purposes. Indeed, the US executive director challenged several requests to draw on IMF funds in the late 1940s on these grounds and, as a result, in 1950 the managing director specified that countries would have to lay out the specific steps they would take in overcoming balance of payments difficulties. Britain and France abstained in the subsequent vote on this issue, while other countries agreed to the American notion of 'conditionality' only because the US was their main source of credit. This conditionality laid on IMF and World Bank loans has subsequently become the main issue of controversy in global economic governance. Nation states resent having their economic policies scrutinized and controlled by economists at the international governance institutions directly, and by the US Treasury Department indirectly.

The Bretton Woods Institutions were supposed to govern agreed upon principles for the conducting of economic affairs decided at the conference. The main principles governing post war development were as follows:

1. The experience of the competitive depreciation of currencies during the 1930s led to the principle of

international control over exchange rates between national currencies under what can be termed 'man aged flexibility'. Under this, the par values of member country currencies were defined in gold, but allowed to vary around these standards, with par values changed by more than this only under drastic con ditions and by permission of the IMF.

2. A pool of gold and currencies to be drawn on in case of balance of payments difficulties was subscribed by member countries to the IMF, according to a quota system adjusted every five years. The quota deter mined the country's drawing rights and also its vote in the institution. Countries could exchange specified amounts of their currencies for those of other counties under conditions supervised by the IMF.
3. To enable multilateral trade, after a five year transition period, all member countries were to eliminate controls making currencies convertible into one another at the official rates without restrictions or discrimination unless approved by the IMF.
4. Various 'scarce currency' provisions were made to decrease instabilities resulting from shortages due to sustained surpluses in a country's balance of payments – the IMF was authorized to declare the scarcity, ra tion its remaining supply, require a country to sell currency to it, in exchange for gold, or use other measures.
5. The Bretton Woods Agreement established a per manent financial institution to promote international monetary cooperation and provide the machinery through which countries could consult and collabor ate. This institution, called the IMF, a specialized agency of the UN, was part of an envisaged system that also included a bank dealing with long term in vestments, the IBRD, a trade organization (ITO) that eventually became the WTO, and actions to promote full employment under the UN Economic and Social Council. The IMF would have a board of governors, representing all member countries that met annually, an executive board, of which five members would be from countries with the largest quotas, meeting con tinuously, and a managing director, who was not a board member. Voting on both boards was to be ac cording to quotas. Under the original quota system, the US had 27.9% of the vote and the UK 13.3%. Today, the US continues to have the largest bloc of votes and together with the other big developed countries controls policy formation in the inter national financial organizations.

The IMF

As originally conceived at Bretton Woods, the IMF was to be a supranational body essentially doing two things: it would regulate the rates at which currencies were ex changed among member countries; and it would help ensure international stability by making loans at times of crisis in member countries' balances of payments. Though its mission statement remains essentially the same, the IMF has subsequently undergone major changes, including several apparent reversals of fortune, that nevertheless resulted in an overall accumulation of power and influence. Today IMF policies directly affect the economies of 184 countries and influence, sometimes drastically, and some say disastrously, the lives of the vast majority of the world's people. The IMF is probably the single most powerful nonstate (governance) institution in the world. Publicly, governments have to praise the IMF while bickering privately about the policies imposed on them, because they may soon need it. By contrast, thousands of workers and students demonstrate against the IMF, in many cases losing their lives in the process because, they say, its economic policies produce poverty, hardship, and starvation.

The IMF gives short term loans to member countries experiencing balance of payments crises, essentially from previously deposited funds, but with the essential dif ference that the loans are made in 'harder' (more inter nationally acceptable) currencies than the deposits. Originally, under the Bretton Woods Agreement, the conditions for IMF loans to member countries required simply "an effective program for establishing or keeping the stability of the currency of the member country at a realistic exchange rate." Over 50 years this limited 'conditionality' has grown into something far more. Loan conditionality, imposed on what are usually desperate governments, has become a way for the IMF to regulate the entire gamut of a country's economic policies. It is not the actual deliverance of the loan, but the conditions laid on its acceptance, that is contested. Governments are required to adopt a set of economic policies and pre scribed financial measures based on what the IMF thinks will promote economic stability, increasing the govern ment's capacity to service loans – that is, make interest and loan repayments. By what theoretical means does the IMF establish these policies? The institution appeals to the best of neoclassical economic science backed now by 50 years of experience in the loan business. However, a critical examination of that experience suggests instead that it extracts repayment at the expense of the economy, and especially on the backs of the poor people, of the borrowing country. In addition, a critical look at its economics, especially after the mid 1970s, suggests the IMF adheres to a neoliberal version of neoclassicism, codified as the 'Washington Consensus'. The term Washington Consensus refers to the policy reforms im posed when debtor countries borrowing from the IMF and World Bank submit to strong conditionality. By 'Washington' is meant the political Washington of the US

Congress and senior members of the administration, and the technocratic Washington of the international financial institutions, the economic agencies of the US govern ment, the Federal Reserve Board, and the think tanks that cluster in Washington, London, and other centers of international power. By 'policy' is meant policy instru ments rather than more general objectives of policy or the eventual outcomes of policy. In terms of the insti tutional formation of recent neoliberal economic policy, the term Washington Consensus can be used to refer to some of the political interests that brought right wing 'progressive reform' ideals to Washington in the mid 1970s and early 1980s, together with the bureaucratic–technical interests whose professional training in neoclassical economics proved amenable to persuasion by neoliberal ideas. The IMF operates primarily as banker to the central banks of nation states. The policies suggested by the IMF almost always require reducing tariff barriers on imports and thus eliminate jobs. The IMF increases interest rates to cool the economy and reduce inflation, and this too reduces employment. At the same time, the IMF imposes austerity programs that cut back government services and remove state subsidies that have kept food prices low. So, critics argue, IMF con ditionality creates unemployment and poverty while re ducing the national state's power to remedy the resulting social problems. Immediately, people who can least afford it are made to pay for loans to governments whose pre vious policies are deemed mistaken by IMF economists. More than this, the IMF imposes its own economic be liefs on countries that might wish to develop differently. IMF loans then become a point of tension at which social struggles within a society articulate with tensions be tween the society and the global system, with two main institutions – the national state and the IMF – at the center of controversy. From a critical perspective, the IMF extracts conditions that favor repayment at the ex pense of poor and working people.

The World Bank

The second main international organization set up to govern development at Bretton Woods was the IBRD or World Bank as it became known. The World Bank Group actually consists of five specialized institutions:

1. The IBRD makes development loans, guaranteeing loans, and offers analytical and advisory services. The IBRD borrows at low interest rates by selling bonds in private capital markets in First World countries and makes near market interest loans to 'creditworthy countries' in the Third World and elsewhere. It has made about $360 billion in loans over its lifetime and currently lends $10.5 billion a year for some 90 new operations in 36 countries.
2. Countries that are 'usually not creditworthy' in international financial markets can get loans from the second banking institution, the International Devel opment Association (IDA). IDA loans carry no inter est, but a 0.75% administrative charge is made annually. The IDA, averaging $6 billion a year in lending to the poorest countries, is funded from member governments' national budgets.
3. The International Finance Corporation (IFC) is the largest multilateral source of loan and equity financing for private sector projects in the developing world.
4. The Multilateral Investment Guarantee Agency (MIGA) provides investment insurance.
5. The International Centre for Settlement of Invest ment Disputes (ICSID) facilitates the settlement of investment disputes between governments and foreign investors (World Bank website).

The terms 'World Bank' and 'the Bank' refer properly only to the IBRD and the IDA. The president of the World Bank is appointed on recommendation by the US (with the director of the IMF appointed by the Western European countries). Its headquarters is in Washington, DC and it has the same membership as the IMF – 184 countries. Its staff numbers 8000 employees in Wash ington and 2000 in the field. The World Bank is owned by its member countries represented by a board of gov ernors, that meets once a year, and a Washington based board of directors that conducts day to day business.

The purposes of the World Bank, as stated in Article 1 of the original (Bretton Woods) Articles of Agreement, are:

1. To assist in the reconstruction and development of territories of members by facilitating the investment of capital for productive purposes, including the res toration of economies destroyed or disrupted by war, the re conversion of productive facilities to peacetime needs, and the encouragement of the development of productive facilities and resources in less developed countries.
2. To promote private foreign investment by means of guarantees or participations in loans and other in vestments made by private investors; and when private capital is not available on reasonable terms, to sup plement private investment by providing, on suitable conditions, finance for productive purposes out of its own capital, funds raised by it and its other resources.
3. To promote the long range balanced growth of international trade and the maintenance of equi librium in balances of payments by encouraging international investment for the development of the productive resources of members, thereby assisting in raising productivity, the standard of living, and con ditions of labor in their territories.

4. To arrange the loans made or guaranteed by it in relation to international loans through other channels so that the more useful and urgent projects, large small alike, will be dealt with first.
5. To conduct its operations with due regard to the effect of international investment on business conditions in the territories of members and, in the immediate postwar years, to assist in bringing about a smooth transition from a wartime to a peacetime economy (World Bank 1989).

For the first decade of its existence, the IBRD made loans for the post war reconstruction of Europe. As it turned toward the richer Third World countries in the 1950s, the World Bank concentrated on project loans for infra structural development. In the late 1960s and 1970s, the World Bank added a basic needs program aimed at poverty alleviation. In 1979 the World Bank began to stress program loans to induce 'reforms' in recipient, mainly middle income counties; by reform was meant structural adjustment lending to promote export orien tation and trade liberalization. In this, the World Bank followed the lead of its senior partner, the IMF, under a division of labor that allocated 'stabilization programs' (short term adjustment lending) to the IMF and longer term 'structural adjustment lending' aimed at correcting deeper 'structural' problems, to the World Bank. Under this new orientation, poverty took a back seat to new driving forces of macroeconomic policy, stabilization and balance of payments adjustments, all understood within a neoliberal doctrine stressing strict limits to governmental intervention and the virtues of flexible, self adjusting, free markets. Some commentators find the World Bank shifting at the end of the 1980s and in the early 1990s to a revised neoliberal model stressing market friendly state intervention and good governance (political pluralism, accountability, and the rule of law) with a renewed em phasis on social issues like poverty and education and a dedication to debt reduction. Thus, in the 1990s, the various 'World Bank development reports' outlined a new 'holistic approach' to development involving social safety nets, poverty, health, education, environment, rural areas, and gender considerations, in concert with con ventionally neoliberal areas, like increased property rights, trade liberalization, and privatization. Both the IMF and the World Bank now say that they are com mitted to global development through debt relief for poor Third World countries and to poverty eradication.

The WTO

After World War II, trade reassumed the crucial position for the growth of economies it had in the second half of the nineteenth century. Increasingly trade became the single most important part of the expansion of economies linked into a globalizing system. Global institutions governing the conditions of trading relations among countries were, at least potentially, placed into positions of great political–economic power. However, the recog nition by nation states, particularly the US, that the global governance of trade might exert significant control over national economies, and the apprehensions aroused by this in a system riven by international competition, long prevented this power potential from being formal ized in an institution. Instead international trade was intermittently regulated through rounds of bilateral and multinational negotiations and the set of agreements known as GATT, signed by 23 governments in 1947. Governance was intermittent because agreements could be unilaterally abrogated, retaliatory actions could easily be initiated, especially by powerful countries, and there was little in the way of an enforcing mechanism. How ever, the number of countries involved in the GATT rounds grew, the proportion of world trade that came under its aegis increased, and a system of power relations stabilized under the domination of the US, so that finally GATT could be concretized as the WTO.

Established on 1 January 1995, the WTO is a more formal, institutionalized version of the GATT. The GATT/WTO system regulates international trade in goods and services using a system of objectives and rules laid out in articles of agreements among member governments. As an organization, the WTO consists of a director general, a deputy director general, and a sec retariat (or bureaucracy) housed in Geneva, Switzerland, where many UN agencies have their headquarters. Like the IMF and the World Bank, it also exists as a broader institution, in this case consisting of trade representatives sent from member countries to meetings organized at a number of levels, and thousands of specialists, consultants, and lobbyists who exercise considerable power – the intricacies of trade regulations inviting ex pertise. The WTO operates within a discourse that, while changing in emphasis over time, has consistently advocated the 'liberalization of trade' – that is, the freeing of international movements of commodities and (recently) services from governmental restraint. Freeing trade from tariffs and other governmental restrictions, and thereby allowing competition and markets to func tion more freely at the international level, is said to lead to more rapid economic growth that benefits everyone. Trade liberalization is one of the leading aspects of the post war economic regime that led to global neoliberal ism. However, while GATT was relatively uncontro versial, the WTO, as a leading proponent of neoliberal ideas, is at the center of controversy over the direction globalization has taken. Indeed protests against the WTO ministerial meeting in Seattle in 1999 have come to symbolize the entire debate over the future course that globalization might take.

The basic power of GATT and the WTO resides in GATT Article I, the General Most Favored Nation Treatment clause, saying that any advantage, privilege, or immunity granted by any member country to another shall be immediately and unconditionally accorded to all other member countries. That is, lower tariffs granted to a favored nation are automatically extended to all members of GATT/WTO, but are not necessarily granted to nonmembers, unless negotiated separately through bilateral agreements. By joining GATT/WTO, a country can gain far freer access to global markets for its exports, with the WTO making sure that the full range of access rights are granted and exercised. This places the WTO at the center of power in the regulation of the global economy, in a position different in emphasis (in that trade is the focus), but at least equal in importance, to that occupied by the IMF. And given the extensions made to the power arena of GATT during the Uruguay Round of international discussions, into services, intellectual property rights, and the governance of certain capital movements, even as it came into being the WTO was also assuming perhaps the dominant position in the global economic governance structure.

The WTO is an institution formed through the interactions of governments under certain conditions. Governments meet at the WTO through their trade ministers, representatives, and delegates under specific circumstances: (1) to discuss trade issues – with declared immediate objectives in mind; (2) to reduce trade barriers, settle trade related problems; and (3) with an overall purpose – to increase the volume of trade, increase production, and raise incomes. More than an institution, the WTO is a place, or a discursive space, where intergovernmental meetings occur, experts congregate, expertise is employed, and decisions are made within a common understanding expressed in a specific, political–economic language. In other words, the WTO is part of a center of hegemonic power, part of a broader international complex of institutions, and ensconced within a broader geoeconomic regime.

International Organizational Complexes

The World Bank and the IMF cooperate on a daily basis with finance ministries, with central banks, and with the big private banks controlling access to global accumulations of capital through the financial markets. Any conception of the regulation of global development by quasi public financial institutions has to take this connection with the banking world into account. The prominent trade economist Jagdish Bhagwati calls this connection the "Wall Street Treasury Complex." Joseph Stiglitz, formerly chief economist at the World Bank, says that the IMF in his experience follows "an ideology that was broadly consonant with the interests of the financial community." Elsewhere the financial connection is described as the Washington–Wall Street Alliance, an institutional complex centered on the US Treasury Department, the IMF, and the World Bank, with an intellectual offshoot to Harvard University, particularly the Harvard Business School, whose MBA, doctoral, and executive education programs train the corporate and banking elites, but with the leading role being played by Wall Street bankers, especially investment bankers. This alliance of powerful institutions headquartered mainly in the US operates within a broader context of international organizations, such as the OECD representing 30 industrial countries, and through organized, intergovernmental meetings. Since 1975, the heads of state or government of the major industrial countries have met annually to deal with major economic and political issues facing the global capitalist system. At first, France, the US, Britain, Germany, Japan, and Italy were referred to as the Group of Six (G6) countries, and then with Canada in 1976 the Group of Seven, or G7, and in 1998, with the full participation of Russia, the Group of Eight, or G8. The G7/8 Summit deals with macroeconomic management, international trade, and relations with developing countries – as with debt forgiveness. In addition, the G7/8 holds a series of ministerial meetings among finance ministers, foreign ministers, and environmental ministers, among others. Since 1999, the G7/8 has also sometimes met with 'emerging market' countries, such as Argentina, Australia, Brazil, China, India, Mexico, Russia, Saudi Arabia, South Africa, South Korea, and Turkey, referred to as the G20, within the framework of the Bretton Woods institutional system. Even so, a leading issue remains as to why a small number of countries and finance institutions should control development in a world that considers itself to be democratic?

Global Power

Financial power is centered in a few 'global cities' that accumulate the capacity for widespread political–economic control by serving as concentrated command points in the organization of the world economy and as key locations for finance and specialized service firms. Control is centered in the few cities that move from being company and corporate headquarters into being 'capital markets', financial centers for national and international economies. The globalization of the late twentieth century saw increased concentration of capital at the top of this global hierarchy, especially New York, London, Tokyo, and Frankfurt, with one city, New York, consistently presiding over the rest. The dominance of New York comes from the amount of capital the city controls and the technological sophistication with which money can be moved from Wall Street to anywhere in the world. Essentially, this kind of global power is exercised by controlling access to the biggest capital accumulations in the world and directing flows of capital in various forms – as equities, bonds, derivatives, foreign exchange, and direct investment – to uses and places that are approved by Wall Street banks and investment firms. Capital markets are filled with financial experts making money from the uncertainties that stem from the normal oper ation of capitalism – a system whose overall 'rationality' emerges from the collision of millions of self interested actions motivated by profit. Expertise tries to overcome uncertainty by collecting and applying financial and technical information that has to be combined with subjective judgments based on practical experience – expertise is institutionalized experience. As a result, the annual passage, through a global city like New York, of documents controlling trillions of dollars in commercial and financial assets leaves, as its institutional residue, a system of thousands of specialized companies filled with connections, knowledge, and expertise. The combination of expertise, concentrated in specialized institutions, with real control over substantial capital investment, is the basis of the political–economic power exercised by global centers of financial power. This power takes many forms.

One form is the influence of capital markets on the making of global development policy by governmental and governance organizations. The most important actors in international capital markets are corporations directly investing in foreign countries and 'professional' or 'in stitutional' investors, financial institutions that invest the savings of individuals, particularly through pension and mutual funds, and invest for nonfinancial companies, such as insurance companies, in the forms of bonds, equities, and loans. In making these decisions, corporations and investment professionals employ assessments of countries using a range of consultancies and specialized insti tutions, especially national credit ratings made by agen cies like the Economist Intelligence Unit, Business Environment Risk Information, International Country Risk Guide, and Euromoney Institutional Investor. Pro fessional investors in such capital markets look at the development policies of governments as signals of the investment risk involved in buying sovereign debt in the form of government bonds. An assessment that a government bond issue carries a high risk of default or rescheduling, or that the inflation rate will rise, or ex change rates become volatile, means that investors demand a higher interest rate, increasing the cost of debt service, or that bond issues will be under subscribed or, worse still, withdrawn from the market. The capital market, acting through institutional investment decisions, influences development policy through mechanisms like bond interest rates, direct investment, and speculative trading in national currencies. Some theorists of inter national capital markets argue that globalization, ac companied by increased market openness and greater capital mobility, has increased the power of capital over government policies. This is particularly the case for the fiscal and monetary policies of governments that are quickly and efficiently punished by reactions from the international bond and currency markets. During the 1970s, the financial markets punished governments that persisted in following social democratic, Keynesian de velopment policies that investors believed led to state deficits that would impede repayment of borrowed funds. The paradigmatic case was the disciplining of the British Labour Party in 1976. The terminal case was France's Mitterand government committed to economic expan sion combined with income redistribution, but forced by bond interest rates of 17.4% to reverse course in 1982 83. Since then, policies in developed countries have con verged on a neoliberal policy model preferred by the market, consisting of smaller governments, lower deficits, less state provision of social services, lower levels of taxation, less regulation, and smaller unions. Since the mid 1980s international investors have had a high degree of confidence in the policies of what became liberal, ra ther than social, democracies in the developed world. Financial markets retain a strong influence on govern ment policies but now consider only a limited set of policies in assessing risk premia, leaving some room for policy maneuver and choice (within a basically neoliberal regime) in development policy decisions. By comparison, in the case of 'emerging markets' (low or middle income, developing countries), investors retain a strong and broad influence on governmental policy. Although emerging market governments are not always constrained to follow the dictates of global capital markets, the cost of defying them is often high.

Wall Street

New York's leading position in global finance capitalism translates into deep and sustained political–economic influence on development policies. Influence is exerted through a number of power relations linking Wall Street institutions to governments and governance institutions. Standard and Poor's and Moody's, two powerful credit rating organizations, assess the bonds and loans of over 100 sovereign nations, as well as thousands of corpor ations the world over, to ratify their eligibility for pur chase by global investors – this 'rating' includes examining the country's economic policy regime – ac cording to 'objective criteria' established on Wall Street. A low rating sends a signal to governments that devel opment policies should be changed. This kind of indirect influence might be termed Wall Street's setting of the 'investment climate' within which nation states, corporations, municipalities, and other organizations may determine their own policy direction, knowing that some are far more costly than others. Finance capital also directly affects the economic policies of governments through the persuasive influence of Wall Street on the International Financial Institutions involved in global economic governance, especially the IMF and World Bank, organizations that in turn coerce governments to borrow outside the private capital markets.

More exactly, the main line of influence between the New York capital market (Wall Street) and the international financial organizations lies through the investment banks. Investment banking is a specialized part of US finance capital created by the Bank Act of 1933, more commonly known as the Glass–Steagall Act. The Act was designed to separate banking from the securities business, and divide commercial banking, taking deposits, and issuing short term loans, from investment banking, concerned more with corporate finance, mergers, and acquisitions. Glass–Steagall restricted the primary market, involving the 'floating' of publicly traded corporate securities, to investment banks – floating means setting the price and terms on initial public offerings (IPOs) of stocks and bonds that are later traded on secondary markets (such as the stock exchange). In particular, one of a small number of investment banks, the 'special bracket firms' (traditionally Merrill Lynch, Goldman Sachs, Salomon Brothers, First Boston, and Morgan Stanley), takes the lead in forming a larger syndicate of banks that underwrites major issuances of corporate stock and bonds and handles the intricate connections with regulators (as with the Securities and Exchange Commission), the specialized accounting and legal service firms, and the potential buyers in the secondary market. Here, the connection with large blocks of capital is crucial because underwriters have to 'take up' any unsold shares or bonds in any new issue. Since repeal of Glass–Steagall in 1999, commercial banks, like Chase Manhattan Corporation and Citibank, have developed, or acquired, investment banking facilities, as well as a range of insurance and stock brokerage services, to produce financial conglomerates that are the biggest corporations in the world, making profits even larger than the huge productive and retailing corporations. These financial organizations essentially control access to the New York capital market and its accumulated expertise. Through this, they exercise considerable control over global development.

In the post war period, especially as the neoliberal policy regime was established, many of the key posts in the US government and in the global governance institutions involved with the financial aspects of national and global economic policy were occupied by experts from the investment banking industry. An elite corps of policy makers circulates between investment banking and the highest levels of governmental and governance institutions. The change in direction of economic policy, from an essentially Keynesian to a neoliberal position, begun during the US Nixon and Ford Republican administrations, solidified in the 1980s, under the Reagan and (first) Bush Republican administrations, and continued in the Clinton and second Bush administrations, was overseen by investment bankers. Investment banking exercises a range of interconnected powers that reach back into the capital market and forward into the production of policy, with control exercised as a combination of expertise, meaning experts armed with specialized knowledge, connections with the financial, corporate, and governmental worlds, and command over financial resources, especially large blocks of investment capital. Political–economic power is then exercised on development as combinations of ideas, expertise, and finance.

If we re read the neoliberal Washington Consensus development policies, adopted by the financial governance institutions in the late 1980s, and dominant ever since, from the bankers' perspective, we can see that minimizing state spending, increasing competitiveness, securing property rights (including those of foreign companies), and having export orienting economies to produce hard currency, all maximize the loan capacity of developing countries and ensure, to the fullest extent possible, the ability of an economy to repay principal and interest. This does not amount to some kind of cynical conspiracy by bankers and multinational corporations to create puppet economic regimes, despite the evidence that meetings recur among banking and governmental allies. Also, once appointed to bureaucratic positions, bankers have considerable freedom to interpret what remains, however, predominantly a banking point of view. However, an export oriented, privately controlled, market economy 'is' the banker's conception of the 'sound' economy, even when rethought in terms of the more general common good. The problem is that this notion of a sound economy does not appear good to those about to be made unemployed, cut off from public services, and paying exorbitant prices for water, electricity, and other recently privatized services. The Washington–Wall Street Alliance, the OECD, and the G7/8 governments have established, protected, and reinforced a neoliberal policy regime that served to de regulate the world economy (in terms of national state intervention), freeing the way for global, and particularly US corporations, the trading of industrial commodities without interference, and the movement of capital assets across national boundaries that have been reduced in significance. For critics, rather than a sound global system, the result is a wild economy of colliding interests and immanent debt crises beyond the control of any particular interest or institution. It is evident, from bankers' emphasis on financial reform in debtor countries, and in their belief that real economic growth is the only way debtor countries can service debts and repay loans, that they think the financial structure and the quality of financial services play a decisive role in making real growth possible, and that, reciprocally and mutually, investment in production and in long term development projects is crucial to a country's financial status, its creditworthiness with regard to banks, and by extrapolation, to the international financial system and its stability. The IMF and the World Bank were established to play this dual role, of stabilizing the global financial structure and encouraging economic development. The questions remain, however, what kind of development, toward what ends, and in whose interest?

The most difficult question of all in the area of international finance is control over development by 'The Market', meaning the various capital markets centered on Wall Street, New York City. Saying that Wall Street is the most powerful actor in setting the policies used by global governance institutions to develop countries the world over implies a more conscious collective actor than may indeed be the case. The central financial institutions lo cated on Wall Street are the equity and bond markets, and these no one controls. While financial analysts may objectively 'study the numbers' while balancing return with risk in evaluating whether to invest client money in the bonds or shares issued in foreign countries, the de cision is rendered into subjective judgment by un certainty about the future when the bonds will be repaid or the shares sold again. The international financial market, then, is a meeting point where biases and guesses collide. Yet the collision of uncertainties in the market determines the economic fate of nations, states, and the global economy. Global development, directed by inter national financial organizations, private and public, has radical uncertainty at its heart.

Counter-hegemonic Organizations

Every year, social movements opposed to neoliberal globalization meet at the World Social Forum (WSF). There they discuss an alternative 'Southern agenda'. The meeting takes the form of a festival. But the thou sands of people who attend a week of conferences, panels, workshops, seminars, cultural activities, and demon strations have the serious intent of producing real social and economic alternatives to neoliberal globalization. The idea for a WSF came out of the massive protests against the WTO meeting in Seattle in 1999. Three ac tivists met in Paris in January 2000 to discuss alternatives to contemporary globalization: Oded Grajew, founder of the Brazilian Business Association for Citizenship (CIVES), an organization of progressive businesses aligned with the Brazilian Workers Party; Francisco Whitaker, of the Brazilian Justice and Peace Commission (CBJP); and Bernard Cassen, chair of Association for the Taxation of Financial Transactions for the Aid of Citi zens (ATTAC), France. Grajew proposed a Social Forum as an alternative to the World Economic Forum (WEF), held each year in Davos, Switzerland, that convenes leaders of Western states, CEOs, and some NGOs, to discuss the global economy; the WEF supports neoliberal globalization. Grajew, proposed an alternative Social Forum, held at the same time, to discuss alternatives to those supported by business at the WEF. The idea was to hold the Forum in the Global South, preferably the city of Porto Alegre, Brazil, known for its participatory bud get process, and controlled at that time by the Brazilian Workers Party. The city's agreement was secured, and the first WSF was held nine months later with 10 000 people attending. The second Forum in 2002 had 50 000 people show up; the third in 2003 had 100 000 activists attend; the fourth, in India, had 120 000 participants; and the fifth, in 2005, in Brazil, had 150 000 registered. In 2006, a polycentric WSF was held in several places: Bamako, Mali; Caracas, Venezuela; and Karachi, Pakistan. Along with these, many smaller social forums have taken place, including the European Social Forum, the Asian Social Forum, the Mediterranean Social Forum, the Italian Social Forum, Liverpool Social Forum, and the Boston Social Forum. The social movements at the WSF believe in radical, participatory democracy – participatory in the sense that civil society organizations are engaged directly in government. Movements from the Third World in particular want decisions about debt (to the First World) and debt repayment to be made by civil society rather than states – many activists are for debt repudiation on the grounds that only the elites benefited from the loans. While some movements want to reform the state, many see the Third World state as oppressive, and want local control and participatory democracy, with links among more decentralized communities in a 'Southern Front'. Networks of movements and activists have emerged in the last few years; many activists find networks to be a powerful instrument to change the world since they 'anticipate' the features of the new world they want to build – they are decentralized, democratic, coercion free, and radically nonhierarchical. There is a counter hegemonic institutional structure that mobilizes millions of people in social movements and develops a 'counter expertise'. This counter expertise serves social movements by reformulating their wants and needs into policy proposals – within an awareness that this creates power inequalities between people and experts. Counter expertise also opposes the policies and practices of the conventional international organizations. Another world is made possible only by thinking carefully in advance.

Implications for Global Economic Geographies

Global economic geographies can be conceived in terms of institutions, power centers, and flows of influence. In advanced modernity, governance policies and practices are conceived by experts in institutions concentrated in a few power centers, global cities that exercise power in world space. Each power center can be thought, in and of itself, as a place, in the sense of a cluster of inter connected institutions, or what might be termed an in stitutional complex. Furthermore, centers of power can be classified as belonging to three main types, according to the dominant purpose of their leading institutions and the type of power they initiate: (1) economic, meaning that their leading institutions deal primarily in money, as with financial markets, investment banks, or corporate headquarters, and transmit power as control over in vestment and financial expertise; (2) ideological, meaning that institutions deal in ideas produced at the level of theory, as with universities, research institutes, and foundations and transmit power as scientifically justified ideas, rationalities, and discourses; and (3) political, meaning that institutions construct and enforce ideas in practical formats, as with government and governance centers that transmit power as policy. Power, then, takes three forms: ideological, meaning control over rational ities; economic, meaning control over capital; and polit ical, meaning control over practice.

Power centers composed from complexes of insti tutions can also be mapped by looking at power relations or as power transmitted across space – this provides an organizing framework for global economic geographies. To begin with, each power center concentrates resources (capital, ideas, and expertise) from a broader 'field of power'. This field may be physically contiguous, in the sense of a hinterland. It may consist of networks among widely separated centers. Or it might be 'virtual' in the sense of position in a world wide web. Experts clustered in power centers process intellectual, theoretical, and practical resources drawn from fields of power by ap plying their own concentrated knowledge and expertise. As argued above, power centers formed by institutional complexes can be classified as hegemonic, meaning that they produce ideas and policies with sufficient theoretical depth and financial backing that they dominate thought over wide fields of power; sub hegemonic, referring to peripheral centers of power that translate received dis courses, modify and add to ideas, and provide evidence of their validity through regional practice; and counter hegemonic, meaning centers, institutions, and movements founded on opposing political beliefs that contend against the conventional, and advocate policy alternatives. Furthermore, in the modern world, ideas backed by political, cultural, and economic resources are transmitted among power centers in the specific ideological form of discourses that are legitimated as scientific, and backed by resources that are recognized as productive, such as capital investment. At the receiving end, in sub hegemonic centers, discourses release their contents as 'power effects', as with theoretical persuasion. Even so, this explosion of persuasive content should not be seen as a one way movement of uncritically received power but, rather, as an articulation in both senses of the term: as creative combination with ideas and expertise originating in local institutions; and therefore as a process of rewriting and modification, but also independent dis covery and creative adaptation. Peripheral power centers are sub hegemonic in that they basically support, while modifying, the ideological positions taken by central institutions. However, they may also be counter hege monic in that they differ fundamentally by offering al ternative ideas backed by their own varying constellations of experiences, modes of interpretation, and intellectual, if not financial, resources. Formed into alternate dis courses by organic intellectuals, peripheral experiences contend against policy in sub hegemonic, hegemonic, and (though they are rare) counter hegemonic centers of power. Power of even the most apparently solid, in disputable kind, is ever destabilized by class, gender, ethnic, and regional differences in experience and interpretation.

See also: Financial Centers, International; Foreign Direct Investment; Globalization and Transnational Corporations; Trade, International.

Further Reading

Dryzek, J. S. (1996). *Democracy in capitalist times*. New York: Oxford University Press.
Eckes, A. E. Jr. (1975). *A search for solvency: Bretton Woods and the international monetary system, 1941 1971*. Austin: University of Texas Press.
Germain, R. (1997). *The international organization of credit: State and global finance in the world economy.* Cambridge: Cambridge University Press.
Kaufman, G. G. (1986). *The US financial system: Money, markets, and institutions*. Englewood Cliffs: Prentice Hall.
Kidwell, D. S. and Peterson, R. L. (1987). *Financial institutions, markets, and money.* Chicago: The Dryden Press.
Mosley, L. (2003). *Global capital and national governments*. Cambridge: Cambridge University Press.
Obstfeld, M. (1998). The global capital market: Benefactor or menace? *The Journal of Economic Perspectives* 12, 9 30.
Peet, R., Born, B., Davis, M. *et al*. (2003). *Unholy trinity: The IMF, World Bank and WTO*. London: Zed Press.
Peet, R. (2007). *Geography of power: Making global economic policy.* London: Zed Press.
Sassen, S. (1991). *The global city: New York, London, Tokyo*. Princeton: Princeton University Press.
Scammell, W. M. (1980). *The international economy since 1945*. New York: St. Martin's Press.
Sparke, M., Brown, E., Corva, D. *et al*. (2005). The World Social Forum and the lessons for economic geography. *Economic Geography* 81, 359 380.

Relevant Websites

http://www.brettonwoodsproject.org
Bretton Woods Project.
http://www.imf.org
International Monetary Fund.
http://www.multinationalmonitor.org
Multinational monitor.
http://www.worldbank.org
World Bank.
http://www.wsfindia.org
World Social Forum India.

Internationalization of Education

J. L. Waters, University of Liverpool, Liverpool, UK

Introduction

It is widely believed that the growth of mass, standardized education in the nineteenth and early twentieth centuries played a crucial role in the development of the modern nation state and the expansion of national consciousness. This bond between education and society was gradually transformed over the course of the twentieth century, as education systems were increasingly linked to capitalist expansion in a global economic context, which has included the gradual presence of the market in educational funding and delivery. Over the last 20 years, the internationalization of education has marked a deepening and expansion of the relationship between education and transnational interests, allied to the ascendancy of neoliberal forms of governance and the establishment of a global knowledge economy.

The internationalization of education is a multifaceted process, occurring simultaneously at different spatial scales. Most obviously, it involves the international mobility of more than 2 million students pursuing education outside their home country, but it also includes the strategies of supranational organizations, national governments, provincial and regional ministries, and municipal school boards, as well as numerous individual educational institutions, promoting and enabling internationalization. Organizations such as United Nations Educational, Scientific and Cultural Organization (UNESCO), the Organisation for Economic Co operation and Development (OECD), and the World Trade Organization (WTO) have been actively involved in advancing an internationalizing agenda. Governments, particularly those in the English speaking West, have fostered international alliances with overseas academic institutions and developed national education brands, although frequently cross border institutional collaborations have occurred without the direct involvement or guidance of national governments. Many public education institutions have recently engaged in entrepreneurial activities related to the development, marketing, and provision of international products and services, establishing satellite university campuses, distance or e learning programs, international MBA courses, and offering joint degrees with overseas institutions while encouraging student and staff exchanges. Some have been involved in the recruitment of students through overseas education events, while the employment of local private consultants to work on their behalf is increasingly common. Importantly, however, the internationalization of education has not affected all countries, regions, and localities in equal measure or in the same way – there is a distinctive and significant geography to this process. This article identifies the most important dimensions of this: international student mobility, the emergence of a multibillion dollar international education industry, and recent moves toward the global standardization of educational testing and delivery.

The Geography of International Student Mobility

The internationalization of education is most commonly associated with the significant and increasing mobility of international students. According to a recent survey by the British Council, there are more than 2.1 million tertiary level international students – twice the number of 20 years ago. Over this period, the international mobility of students has followed a distinctive geographical pattern: from developing to developed countries, from East to West and from non English speaking to English speaking countries. The British Council has coined the term main English speaking destination countries (MESDCs) to describe the preferred locations of international students. These countries, namely, the United States, Australia, Canada, New Zealand, and the UK have attracted by far the largest share of such students globally (**Figure 1**). Demand for international student places, in contrast, comes principally from Asia, particularly China (PRC), Hong Kong, India, Malaysia, and Singapore (**Figure 2**). In *Vision 2020*, a report commissioned by the British Council in conjunction with International Development Program (IDP) Australia and Universities UK, provides a forecast of international student mobility. By 2020 demand for student places from Asia is expected to reach 1.9 million or 71% of the global total for the MESDCs; in contrast, Europe will account for only 12% of the total figure.

While the majority of international students apply for a student visa on an individual basis, schemes such as European Community Action Scheme for the Mobility of University Students (ERASMUS) have sought to organize international student mobility on a larger scale. ERASMUS is the European Commission's educational program for higher education students, teachers, and institutions. Introduced in 1987, its primary aim has been

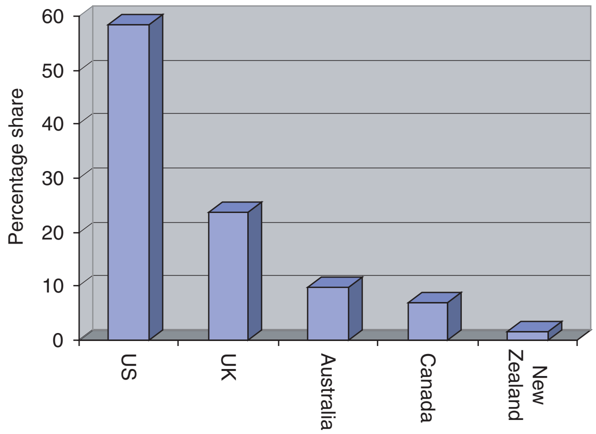

Figure 1 Market share of international students (2002). Source of data: Böhm, A., Follari, M., Hewett, A. *et al.* (2004). *Vision 2020: Forecasting international student mobility - a UK perspective.* British Council, Universities UK and IDP Australia.

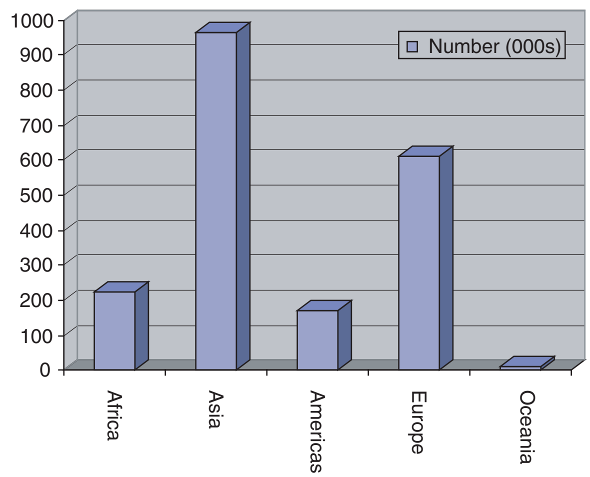

Figure 2 Global demand for international student places (2003). Source of data: Böhm, A., Follari, M., Hewett, A. *et al.* (2004). *Vision 2020: Forecasting international student mobility - a UK perspective.* British Council, Universities UK and IDP Australia.

to increase student mobility within the European Community. Since this time and as a consequence of this program, more than 1 million students have spent a period of their study overseas.

International Education: A Student's Perspective

Students are attracted to overseas study for a number of reasons. First, they may perceive deficiencies or inadequacies in the education available in their home country. At the same time, there is a widespread perception that an overseas education – particularly a 'Western' education – is invariably superior.

The advantages of international schooling include a greater range of specialist subjects, a preferred pedagogic style (promoting creativity over rote learning), as well as providing the opportunity to become proficient in a foreign language. All international students, regardless of their actual academic achievement, acquire the cultural capital associated with 'overseas experience'; immersion in a foreign culture – especially a Western, English speaking culture – is thought to promote a more cosmopolitan outlook and an ability to work in cross cultural settings. International students are seen as part of a global, transnational, and highly skilled workforce. In certain professions, particularly those associated with global industries such as financial services, a cosmopolitan sensibility is regarded as a valuable commodity; there is some evidence to suggest that international students who return to their home country following graduation find it easier to get a job in these preferred professions than graduates educated locally. More controversially, there is some concern among government agencies in host countries that international study is sometimes regarded as a backdoor route to permanent settlement, where illegally overstaying a student visa may be followed by employment, citizenship, and the subsequent migration of other family members.

The Business of International Education: Political and Economic Implications

International education is big business with estimates for the worldwide market in international higher education ranging between $30 and $100 billion. Its relative value varies for different national economies, being especially significant within the MESDCs. For the United States, Australia, Canada, New Zealand, and the UK, international activities of various kinds are an increasingly significant source of revenue for governments and universities. Indeed, in a current climate of neoliberal governance resulting in state funded disinvestment in key services including education, the present and future of state funded higher education is becoming reliant upon the income generated by fee paying international students.

Despite a concomitant dependency on international students, these countries have approached the internationalization of their education systems in quite different ways. The internationalization of Australia's higher education sector is highly centralized, coordinated by one main organization (IDP) with support from Australian Education International (AEI). IDP is currently owned by 38 of Australia's 39 universities; it is a nonprofit organization with an annual turnover of $108.1 million. In 2002, it ranked 353 in Australia's top 500 private companies, blurring the boundaries between public and private initiatives in education.

The UK has also centralized the organization of its internationalization process through the British Council. This is a nonprofit organization, operating as an executive nondepartmental public body and sponsored by the UK Foreign and Commonwealth Office. For 2003/2004, the financial turnover of the Council was estimated at £479 million, including £164 million from a core grant in aid from the Foreign and Commonwealth Office and £154 million from the selling of services such as English language courses and examinations. In 2002–03 the Council worked in 110 countries and taught English to 500 000 students. It has also collaborated closely with IDP in developing the International English Language Testing System (IELTS). In April 2004 these two organizations hosted a one day conference in London resulting in a publication outlining a vision for the future of international student mobility.

International education in New Zealand is coordinated by Education New Zealand – a single nonprofit organization that describes itself as "independent of government, recognized by government, and industry owned." It acts on behalf of both public and private education providers at secondary and tertiary levels and lists its functions as: "advocacy on behalf of New Zealand's education export industry," "assisting the New Zealand government and its agencies with policy development in relation to the education export industry," "promotion of New Zealand as a study destination," and "promotion of the 'Educated in New Zealand' brand." It exemplifies the intragovernment collaborations characteristic of international education, involving the Ministry of Education, the Ministry of Foreign Affairs and Trade, New Zealand Immigration Service, and the Asia 2000 Foundation.

The development of national education brands, promoted globally, has been a key element of these internationalization strategies. New Zealand launched the Educated in New Zealand brand in 2001. The government utilized the services of a private company to devise an international marketing plan, allocating $3.5 million over 5 years to develop the brand, including support for a 'one stop shop' website. In June 2004, control of the international marketing of New Zealand's education was transferred to the 'Promotions and PR Reference Group' and core funding for international education thereby shifted from government to industry. In the UK, efforts to develop a national education brand have been coordinated. As in New Zealand, this process has involved collaboration across several government departments; the Council receives funding from the Department for Education and Employment, the Department of Trade and Industry, the Ministry of Defence, and the Foreign and Commonwealth Office. It also employs several large private companies for research, public relations, and brand development/advertising. 'Education UK' was officially launched in 2000, involving several high profile overseas events, beginning in Hong Kong (British Council 2000). Presently, 'Education UK' offers six 'products' to potential students, including English language courses, degree courses, and MBA and postgraduate courses.

Canada and the US have so far lacked a centralized, government wide policy in respect of international education. In Canada, federal government and provincial government have jurisdiction over different aspects of what could potentially fall under the rubric of 'international education'. Educational policy and practice, through the Ministry of Education, is the responsibility of provincial government, and the provinces differ considerably in their approaches to internationalization. At the same time, other important facets of international education come under the jurisdiction of federal government – foreign policy, human resource development, international trade, and immigration policy are all the responsibility of federal departments. International Trade Canada (ITCan) administers the export of educational services, and Foreign Affairs Canada (FAC) is responsible for international diplomacy, treaties, and the promotion of Canadian culture. Human Resources and Skills Development Canada (HRSDC) advocates international partnerships which foster collaboration in learning while also promoting a positive image of internationalization amongst universities and colleges in Canada. Immigration policy is the responsibility of Citizenship and Immigration Canada (CIC), and this includes the issuing of international study permits (or student visas). It is this complex division of responsibilities, between federal and provincial government, that makes the conception and implementation of a distinct plan for international education problematic.

Educational exports are worth an estimated $12 billion annually to the US economy and yet it still lacks a centralized approach to international education. The events of September 11, 2001 have encumbered the development of a national strategy for international education: the US Department of Homeland Security now manages a national computer database, designed to track foreign students for the duration of their study, while the student rejection rate has also increased significantly over the same period.

As these examples suggest, particular Western countries have been extremely proactive in the commercialization of their education systems, involving private corporate involvement in addition to numerous public agencies. Work by geographers is also beginning to uncover the internationalizing strategies of countries outside this Western 'core' of MESDCs, such as Singapore's self styled development as a global education hub through formal arrangements with a variety of overseas educational institutions.

Regional Initiatives and Transnational Partnerships

Various regional partnerships have been pursued under the internationalization of education such as the Association of Pacific Rim Universities and the European area of higher education. In Canada where, as noted, education is the responsibility of the different provinces, a regional strategy toward the internationalization of education has been necessary and provides a useful illustrative example. In Advancing the Canadian agenda for international education, a Report of the Millennium Consultation on International Education, the Canadian Bureau for International Education (CBIE) describes provincial level approaches. Provinces are autonomous and directly competitive. They are also unambiguously concerned with the revenue generating potential of internationalization.

The CBIE proffers several examples of the active involvement of the provincial government in international education. In 2000, the Department of Education in Manitoba founded the Manitoba Council for International Education, which is described as a consortium of fee paying members that included public schools, higher education institutions, and private educational providers, with the primary aim of coordinating joint marketing activities. In New Brunswick, the provincial government has established and continues to fund the Universities International Centre for recruiting international students and promoting the New Brunswick 'branded' umbrella overseas. Alberta's International Education Strategy has resulted in the Alberta Centre for International Education, comprising partnerships between public and private secondary and postsecondary institutions. According to the CBIE report, it operates as a joint marketing agency, extracting fees from its member institutions. In addition Quebec, also, has recently announced the launch of a plan to encourage international activities – the *Programme pour l'Internationalisation de l'Education Québécoise.* All of these provincial initiatives signify collaborations between 'multiple stakeholders', intersecting public and private interests with an overriding economic motive. The British Columbia Centre for International Education (BCCIE) is the longest running of all provincial initiatives in Canada, established in 1990 as a consortium of postsecondary institutions and funded by the British Columbia Ministry of Education. It is a nonprofit society governed by a board of directors that includes educators and representatives from government and business and its membership comprises all 27 public postsecondary institutions in British Columbia. The province is overtly concerned with the financial potential of internationalization, embracing the concept of marketization. It can be seen how such initiatives further the development of an internationally differentiated market for education: the BCCIE is tasked with creating a distinguished 'brand image' for provincial education, described in terms of 'educational products and services'. It highlights a pervasive concern with the skills and attributes necessary to work in a global market. The evident interest in benchmarking and quality assessment measures is characteristic of the application of neoliberal economic tools to education systems globally. And finally, the BCCIE is attempting to forge a number of different international and translocal connections – sometimes involving the federal government and sometimes bypassing it. According to the BCCIE, international education generates "significant revenue for Canada, British Columbia and the public postsecondary institutions in the province" amounting to CA$2.5 billion per year – the equivalent of 27 500 jobs. For British Columbia, the market is in excess of $200 million boosting it beyond commercial fishing in terms of economic importance.

The internationalization of education also includes some unexpected alliances that often seem to bypass the influence of the nation state, exemplified by Universitas 21 (U21). It describes itself as an international network of leading research intensive universities whose purpose is to facilitate collaboration and cooperation between the member universities and to create entrepreneurial opportunities. The organization was founded in 1997 and has 17 members from 10 countries. (The members of Universitas 21 are: University of Melbourne, University of Queensland, University of New South Wales, University of Auckland, University of Hong Kong, National University of Singapore, Peking University, Fudan University, University of British Columbia, McGill University, University of Virginia, University of Birmingham, University of Nottingham, University of Edinburgh, University of Glasgow, Lund University, and Albert Ludwigs Universitat Freiburg.) In 2001 'Universitas 21' initiated 'Universitas 21 Global' – a result of collaboration between 16 of the member institutions and the Thompson Learning Corporation. U21 Global is concerned with establishing online degree courses, and in 2003 it launched the first completely online MBA program. The Masters course is sold to major markets in Hong Kong, Singapore, and Malaysia. U21 Global exemplifies the common and close relationship between the public and private sectors in international education. 'Thompson Learning' is a large, private publishing company that has provided the technology and expertise behind U21 Global in addition to US$25 million in funds. Universitas 21 is also implicated in the drive to manage modes of assessment and enforce quality control in international education, marketing the U21 brand on the basis of this image; it has been directly involved in the development of international quality assessment tools through U21pedagogica.

On one level, Universitas 21, along with other emergent forms of 'e learning', would seem to be redefining the geography of international education. New technology is undermining the necessity of spatial mobility – 'WebCT' is used to run tutorial groups simultaneously on three continents. U21 is also piloting the use of electronic books. Simply put, the developments represented by U21 Global are allowing students access to an overseas education at home. While this might appear to represent the democratization of access to Western forms of education outside the West, in reality the costs of these courses prohibit all but the wealthiest students from participating.

Overseas Institutions and Offshore Schools

One very recent development in the internationalization of education has been the establishment of 'overseas' institutions (schools and university campuses) outside the MESDCs. In early 2006, the University of Nottingham in the UK became the first foreign higher educational institution to be granted a license to open a campus in China. This was followed closely by The University of Liverpool, which opened a new university in Suzhou west of Shanghai in September 2006. Students of these institutions will graduate with full University of Nottingham and University of Liverpool degrees (respectively). While the practice of establishing an overseas campus or institution is relatively rare, other universities in the MESDCs have been heavily involved in setting up 'joint educational initiatives' with foreign institutions.

These practices have an unexpected precedent at the provincial level in Canada, involving the establishment of certified 'offshore schools'. At present, Ontario, British Columbia, Nova Scotia, and New Brunswick all have offshore schools, enabling students to receive a provincial Ministry of Education certified education 'in China'. The program is taught in English by provincially certified teachers, and its graduates are issued with a provincial (e.g., British Columbia) High School Graduation Certificate. The first school of this kind was 'Dalian Maple Leaf International School' in Liaoning province, which began in 1997 as a pilot project. Student enrolment in the program rose from 55 in 1997 to 1300 in 2003. This trend is likely to continue beyond China into other countries with a high demand for international credentials. It enables students to receive an 'international education' at home.

International Standardization of Education

Through the implementation of principles of standardization, the proliferation of testing, ranking, and benchmarking in education is increasingly international in scope. Large private companies intersect with public higher education institutions as they formulate and administer standardized testing systems, thereby attempting to create global markers of educational distinction. One of the foremost examples of global educational testing is provided by the Educational Testing Service (ETS), a private company that in 1947 procured the testing programs of the American Council on Education, the Carnegie Foundation for the Advancement of Teaching, and the College Entrance Examination Board. The ETS possesses several trademarks associated with international educational testing, including the Graduate Record Examinations (GRE), The Praxis Series: Professional Assessments for Beginning Teachers, Test of English as a Foreign Language (TOEFL), Test of Spoken English (TSE), and Test of Written English (TWE). It describes itself as "the world's largest private educational measurement organisation" and claims that since 1964, nearly 20 million students have taken the TOEFL in more than 105 countries. According to information provided by TOEFL, more than 4500 institutions worldwide use the scores from these tests for international recruitment. In addition, ETS runs the 'Center for Global Assessment' and the 'Global Institute', with the mandate to develop curricula and courses that meet the demands of education officials in government offices, colleges and universities, and testing agencies. The ETS not only administers and designs tests, therefore, it also propagates this system through the training of 'officials' within the education sector.

The International English Language Testing System (IELTS) provides another example of the strategic internationalization (and marketing) of testing systems, and their consequent worldwide adoption. It demonstrates, also, the international alliances forged through these processes, being jointly managed by the University of Cambridge ESOL Examinations, the British Council, and IDP Education Australia. Figures for IELTS Australia claim an annual turnover of $19.8 million, and a growth in turnover from 2001 of 61%. In 2002, IELTS Australia operated in 26 different countries, it administered 92 986 tests and it has 92 test centers worldwide. Perhaps most significantly, IELTS has captured the Chinese market through a 'network partnership' (the China IELTS Network). In 2002, 89 134 tests were administered through this network (a growth of 95% on the previous year). The IELTS is the only test accepted by the Australian Department of Immigration and Multicultural Affairs for student visa purposes (according to Studyinaustralia 2004 – the official government website for international students), suggesting a gradual cornering of the global market in testing and standardization.

There has been a growing interest, also, in the relationship between large supranational organizations and the internationalization of education, particularly the role of the UNESCO, the OECD, and the World Bank. Founded in 1946, UNESCO has 189 member states and is involved in scientific, cultural, and educational projects throughout the world. Its overarching objective is to promote cooperation and understanding among nation states, which consequently facilitates the development of comparable, normalized education systems globally. The organization's mandate includes a role in developing and implementing the practice of 'standard setting' in inter national education through "preparing and enabling the adoption of international instruments and statutory rec ommendations." These forms of standardization involve, principally, the measurement and monitoring of national education systems and their insertion into and consoli dation within a formal ranking of global education.

Several key features of this particular role can be identified. UNESCO is responsible for producing the 'World Education Report', containing statistics on various aspects of educational provision for 190 countries and territories. Since the 1993 report, UNESCO has sought to expand and deepen available data, through the worldwide dissemination of the revised International Standard Classification of Education (ISCED) – a for mula devised by the organization. Moreover, in 1997, and in conjunction with the OECD, UNESCO launched the World Education Indicators (WEI) project, the aim of which was to develop 'policy relevant' education indi cators that enable the international comparison of edu cation systems. The project has produced a comparative data set of education indicators for 1998–2001, including 'detailed mappings of national education systems' and three substantive reports (Investing in Education (1999), Teachers for Tomorrow's Schools (2001), and Financing Education – Investments and Returns (2003)). The or ganization has also developed the Global Forum on International Quality Assurance, Accreditation, and the Recognition of Qualifications. The forum was launched in 2002.

Conclusions

This article has highlighted the most significant dimen sions of an emergent geography of international edu cation, although a comprehensive examination of the multitude of current initiatives is an impossible task. Reflected in the discussion above, recent interest in the internationalization of education has led researchers to examine the process from a number of different per spectives, including the relationship between inter national education and immigration policy, family strategies and configurations, different forms of capital (particularly cultural and social), the impact of inter national students on 'host' societies and communities (such as on housing), highly skilled migration, new forms of institutional power, the rescaling of social repro duction, the creation of cosmopolitan citizens, and global cities and national objectives. The internationalization of education is clearly an emergent and increasingly vibrant field of research.

See also: Knowledge Economy; Labor Flexibility; Labor Market; Migration; Neoliberal Economic Strategies; Transnational Elites; Transnationalism.

Further Reading

Appleton, S., Sives, A. and Morgan, W. J. (2006). The impact pf international teacher migration on schooling in developing countries The case of Southern Africa. *Globalisation, Societies and Education* 4(1), 121 142.

Bohm, A., Follari, M. and Hewett, A. *et al*. (2004). *Vision 2020: Forecasting international student mobility a UK perspective*. British Council, Universities UK and IDP Australia.

British Council (1999). *The Brand Report Building a world class brand for British education*. London: British Council.

Butcher, A. P. (2004). Educate, consolidate, immigrate: Educational immigration in Auckland, New Zealand. *Asia Pacific Viewpoint* 45, 255 278.

ERASMUS. http://ec.europa.eu/education/programmes/socrates/erasmus/erasmus en.html (accessed Oct. 2007).

Farquhar, R. H. (2001). *Advancing the Canadian Agenda for International Education*. Canadian Bureau for International Education.

Huang, S. and Yeoh, B. S. A. (2005). Transnational families and their children's education: China's 'study mothers' in Singapore. *Global Networks* 5(4), 379 400.

Lewis, N. (2005). Code of practice for the pastoral care of international students: Making a globalizing industry in New Zealand. *Globalisation, Societies and Education* 3(1), 5 47.

Mitchell, K. (2006). Neoliberal governmentality in the European Union: Education, training, and technologies of citizenship. *Environment and Planning D: Society and Space* 24(3), 317 474.

Olds, K. (2007). Global assemblage: Singapore, Western universities, and the construction of a global education hub. *World Development* 35(6), 959 975.

Olds, K. and Thrift, N. (2005). Assembling the "global schoolhouse" in Pacific Asia. In Daniels, P., Ho, K. C. & Hutton, T. (eds.) *Service Industries, Cities and Development Trajectories in the Asia Pacific*, pp 199 215. London: Routledge.

Robertson, S. (2003). WTO/GATS and the global education services industry. *Globalisation, Societies and Education* 1(3), 259 266.

Robertson, S. (2005). Re imagining and rescripting the future of education: global knowledge economy discourses and the challenge to education systems. *Comparative Education* 41(2), 151 170.

Waters, J. L. (2006). Emergent geographies of international education and social exclusion. *Antipode* 38(5), 1046 1068.

Waters, J. L. (2006). Geographies of cultural capital: Education, international migration and family strategies between Hong Kong and Canada. *Transactions of the Institute of British Geographers* 31, 179 192.

Waters, J. L. (2005). Transnational family strategies and education in the contemporary Chinese diaspora. *Global Networks* 5(4), 359 378.

Relevant Websites

http://www.bccie.bc.ca
British Columbia Centre for International Education (BCCIE).

http://www.britishcouncil.org
British Council.
http://www.educationnz.org.nz
Education New Zealand.
http://www.ets.org
Educational Testing Service (ETS).
http://www.idp.com
IDP Education Pty Ltd.
http://www.unesco.org
UNESCO 2003a.

Internet, Economic Geography

M. Zook, University of Kentucky, Lexington, KY, USA

Glossary

Domain Names A domain name is a unique identifier (e.g., nytimes.com or manchester.ac.uk) associated with an IP address that allows users to easily access a specific Internet resource such as a website.

E-Commerce E-commerce encompasses a wide range of technologies from placing and receiving simple email orders to inventory tracking systems which automatically place orders based on sales. All e-commerce systems share a common focus on the use, organization, and transmission of electronic data to coordinate the movement of physical objects.

EDI Electronic data interchange refers to a standardized data format that allows for the exchange of structured and complex information between and within firms.

Knowledge (Tacit and Codified) Codified knowledge is knowledge that is possible to record or transmit and therefore lends itself more easily for transmittal across distance. Tacit knowledge is knowledge acquired through observation or interaction and is often thought to be best transmitted face to face.

Point of Sale (POS) Systems POS systems are computer systems that collect data on the sale of items (generally through a bar code reader) and automatically send this data to inventory management software which can automatically place an order to restock items.

RFID Radio frequency identification (RFID) is a system by which items receive small tags which can be read at a distance by specialized RFID readers making inventory and sales easier.

The Internet Changes Everything

A popular meme during the 1990s was "the Internet changes everything." While clearly hyperbole, the idea of technologically driven change was and remains an important element in popular views of the economy. Geographers take a more measured view that strives to keep technological and geographical determinism in check. Rather than generalizable linear relationships, geographers focus on the complex collections of differ ence between places and are acutely cognizant that the creation and use of new technologies is embedded in larger social networks. This is not to deny the transfor mative power of the Internet but to recognize that this power is filtered through existing and evolving spatial, institutional, and cultural aspects of the economy. Firms, governments, workers, consumers, etc., all engage in the dynamic use of the Internet and these actions – in con cert and conflict with the myriad of simultaneous and parallel actions by other actors – shape how the Internet unfolds and is deployed across the spatial economy. This process holds the potential for great change, for example, the nineteenth century technologies of the railroad, tel egraph, and telephone helped usher in the modern corporation and shifted the competitive advantage of regions, but it is prudent, to distinguish between the possible and the evitable.

Toward this goal, this article reviews the interaction between the Internet and economic geography. The term Internet is used here as a catch all term for the use of a range of general purpose information technologies (ITs) and software deployed across public, private, and peered networks rather than a more limited definition. This decision reflects both the diversity of implementations of IT and the Internet and the rapid change associated with it. This allows one to focus on the larger processes at work rather than the specific manifestations tied to par ticular technologies. This article covers three main sub jects. The first examines the role of distance in the age of the Internet. In other words, to what extent has the Internet 'destroyed' distance and what type of economic frictions remain relevant? The article then considers how Internet technologies are used to organize and move the inputs and products of the economy via a range of soft ware and hardware platforms. Encapsulated under the label of electronic commerce (e commerce), this process can engender great change within the organization and location of production and consumption. The final topic focuses on how the uneven use of the Internet across time, place, and organization creates a differentiated geography of change and challenge. Together the review of these topics illuminates the complex economic geog raphy of the Internet and the ways in which it has in sinuated itself into the daily tasks of production and consumption.

Calculating Distance in the Age of the Internet

Tobler's first law of geography, that is, "…everything is related to everything else, but near things are more re lated than distant things" is often cited as a casualty of the Internet era. Pundits of various stripes have been declaring the death of distance, for over a decade.

More recently, Thomas Friedman's bold declaration that the "world is flat" because IT has removed geographic and other barriers to competition has gripped the popular and policy imagination. Despite this prevailing rhetoric, researchers have repeatedly demonstrated the continued uneven spatial pattern of Internet infra structure, use, and content throughout the world. As **Figure 1** illustrates the countries with highest levels of per capita use of the Internet remain largely Western (with the notable exceptions of Japan, Singapore, South Korea, and Taiwan) and wealthy. Even countries within this privileged group with high numbers of Internet users per capita exhibit wide variation in the adoption of do main names – a key indicator of Internet content distri bution – with countries like Taiwan, Korea, and Japan ranking far below other developed economies such as Germany, the United Kingdom, and the United States. These differences are in turn replicated at all scales as access and use of the Internet varies at the subnational and local level.

Thus, the adoption of the Internet is far from uniform and is instead shaped by an inherently messy negotiation between people, polities, and technologies. Despite its ability to organize and shift information easily across vast distances, the Internet is also fundamentally tied to physical places and the people, who create, regulate, distribute, and consume its data. Simply put, the Internet has been, is, and likely will always be clustered in par ticular places. Even the seemingly simple expectation that someday the Internet will be accessible everywhere is belied by the continued pace of change and innovation. Rather than there being one final and fixed goal, new technologies regularly appear so that even if all places had Internet dial up service (a near impossibility in and of itself), better access such as high speed broadband or wireless would appear during the rollout of dial up ser vice replicating the digital divide albeit at a different strata. In short, the Internet will always be differentially distributed across space and likewise, the use of the Internet or Internet based applications will be differen tially distributed across users and firms based on their cognitive and technical skills.

Nevertheless, it is important to recognize that "death of distance" discourse is compelling because it does contain elements of truth. ITs are making it easier to overcome the friction of physical distance and establish and maintain regular and meaningful contact with people around the world. There are, however, two important caveats to this. First, knowledge creation and transfer are increasingly important factors in economic fortunes and these capabilities seem to work best in physical proximity which facilitates tacit knowledge exchange. While access to raw materials and transportation remains as necessary as at the height of Fordism, knowledge embodied in skilled labor is a key input to many activities and one must be cognizant of how creation and use of knowledge takes place in the Internet era. While IT enabled logistics systems make it easier to stretch out just in time supply

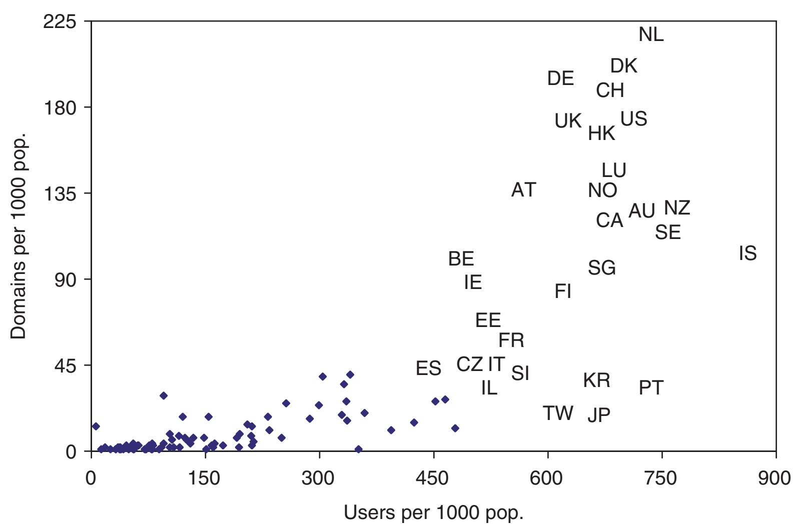

Figure 1 Domains and users per 1000 people, June 2007. Codes based on Internet Assigned Numbers Authority (IANA) top-level domain (TLD) assignments; codes for selected countries: AU Australia; AT Austria; BE Belgium; CA Canada; CZ Czech Republic; DK Denmark; EE Estonia; FI Finland; FR France; DE Germany; HK Hong Kong; IS Iceland; IE Ireland; IL Israel; IT Italy; JP Japan; KR South Korea; LU Luxembourg; NL Netherlands; NZ New Zealand; NO Norway; PT Portugal; SG Singapore; SI Slovenia; ES Spain; SE Sweden; CH Switzerland; TW Taiwan; UK United Kingdom; US United States. Due to legibility issues, countries with lower levels of users and domains are simply represented by diamonds. From http://internetworldstats.com (users); http://www.zooknic.com (domains); includes registrations of biz, com, info, mobi, net, org and ccTLD domains.

chains it remains difficult to replace all the advantages of physical proximity, for example, face to face communication, chance meetings, access to 'buzz' or gossip, etc. There are numerous attempts to use the Internet to increase the quality of knowledge transfer across physical distance, for example, voice over internet protocol (VOIP), email, thread discussion groups, instant messaging, as well as more complex knowledge management systems, but these remain imperfect substitutes. In short, the more complex an interaction is, the more difficult it is to overcome the friction associated with physical distance (see **Figure 2**).

The second caveat is that there are additional dimensions to economic friction beyond simple distance that can impede the flow of knowledge and forestall a successful economic interaction such as those listed in **Figure 2**. These frictions are increasingly coming to the fore as the task of moving objects has become easier. Although the idea of industrial districts where "the mysteries of the trade…were in the air" have been around for over a century, scholars have reinvigorated this field over the past three decades, partially to answer the question of why clusters persist in the Internet age, particularly in the very IT industries that are providing the means to transcend space. Pointing to the advantages of shared culture, networks of social relations, common behavioral norms and expectations, etc., researchers have constructed theoretical justifications backed by empirical evidence as to why this clustering occurs. One particularly relevant example to this article was the clustering of dot com companies during the 1990s, that is, companies developing and using the very technology that was reputed to destroy geography. Given the high complexity of the activities of these companies, combined with the relational, organizational, and labor need frictions of venture funded startups, it is not surprising that these firms clustered despite the popular rhetorical expectations to the contrary. As Leamer and Storper note (p. 653) "…the exchange of uncodifiable ambiguous information that depends on a high level of trust and shared context is likely to continue to require a significant amount of co presence."

In short, the Internet does not make Tobler's law obsolete, but it vastly complicates the types and definition of distance that economic geography must consider. Near things do matter more than distance things but measures of close and far are no longer a one dimensional variable of physical proximity but a composite measure encompassing relational, cultural, technological, and linguistic components in addition to Euclidian distance. Moreover, the component that rises to the top as most crucial varies depending on the frequency, nature, and complication of an interaction. The economy in the Internet era is increasingly structured in complicated networks which connect some places and leave others – even those physically proximate to key nodes – cut off. In a very real sense, people and places may be more closely tied economically with the other side of the planet than what is right down the road. This, however, represents a complication of distance rather than its simple destruction.

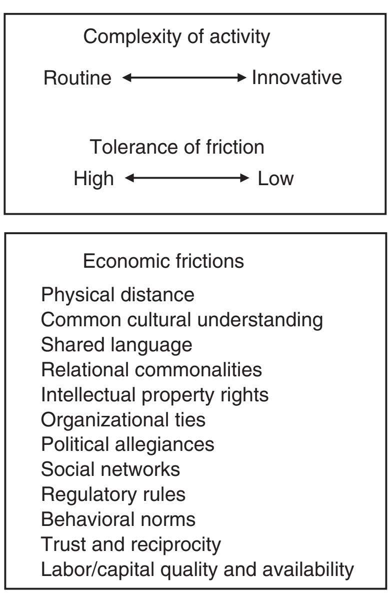

Figure 2 Complexity and frictions.

Moving Stuff: Atoms versus Information

Despite the complex definition of distance and friction operating in the Internet era, it is easy to become overly focused on the movement of information and knowledge versus the movement of atoms (physical products). The fact that information and content can be easily and widely distributed is often mistakenly extended to the movement of things. Instead, particular places are connected in highly specified ways, for example, low cost labor, financial resources, markets, etc., to global networks resulting in a system of production that is both place rooted and networked at the same time. In actuality, there is a small range of informational products and services that are open for instantaneous shipment, for example, music or video downloads, data services, etc., and even these purely informational products still rely upon material reality, that is, recording studies, musical venues, for their production. Moreover, their ease of shipment has made it simple for consumers to redistribute the product for free much to the chagrin of copyright holders. As a result, many dominant powers in these industries, for example, the Recording Industry Association of America (RIAA) are attempting to reinvigorate another friction, that is, copyright law, to maintain their sales.

Nevertheless, the more important role of the Internet in the space economy is the movement of atoms via e commerce. E commerce has existed for some time, for example, electronic data interchange systems (EDI) date back to the 1970s but greatly expanded during the 1990s as the Internet made it feasible for businesses of all sizes; not just large corporations. Although e commerce en compasses a wide range of technologies from placing and receiving simple email orders to inventory tracking sys tems which automatically place orders based on sales, all systems share a common focus on the use, organization, and transmission of electronic data to coordinate the movement of physical objects. Because of this variety, many analysts including the U.S. Census (2007) define e commerce quite broadly on the basis of the use, organ ization, and transmission of electronic data to sell goods and services. This definition focuses on the economic activity, that is, sales, rather than a specific technological platform. Based on this definition it is possible to track the growth of e commerce over time and across sectors via the U.S. Census data.

Figure 3 shows the size and industrial distribution of e commerce within the US economy. Despite the Inter net bubble of the late 1990s and early 2000s, e commerce has steadily expanded to US$2.4 trillion worth of trans actions in 2005. More importantly, these data highlight the concentration of e commerce in manufacturing and wholesale activities, that is, business to business (B2B) transactions, and the very low level of e commerce use within the retail and service sectors, that is, business to consumer (B2C). Thus, despite the visibility of online retailers, the bulk of e commerce takes place between businesses with an ever increasing share of these transactions conducted via Internet based platforms ra ther than EDI. More intriguing, however, than the simple size of e commerce activity is how it allows business and industries to reorganize themselves. Simply installing e commerce software need not create value for a company in some predestined and linear fashion (a fact aptly il lustrated by the dot com bust). Instead, firms can use e commerce to change the organization and geographies of their production systems and in so doing create value (through the reduction in inefficiencies or the rise of new types of interactions) and blend the world of atoms and the world of information.

This promise, however, remains a work in progress as the Organisation for Economic Co operation and Development (OECD) notes that few firms (with the notable exception of larger corporations) have used e commerce to reorganize their production chains and tightly integrate themselves with suppliers and/or cus tomers. This is not surprising as it mirrors the decade long process by which earlier technologies were used to transform existing business systems. A key issue for economic geography is how this process plays out over space. Although the spatial contours of e commerce use is firm and industry specific, there are fault lines emer ging such as urban/rural differences, large/small firm size, etc. Although rural firms are arguably positioned to accrue the greatest benefits from e commerce adoption they are often at a disadvantage because of lower edu cation levels and other capabilities. Additionally, because of the uncertainty around how best to implement e commerce, it is likely that firms located in cities with better human resources and access to knowledge will be earlier adopters. This divide (again based on resources

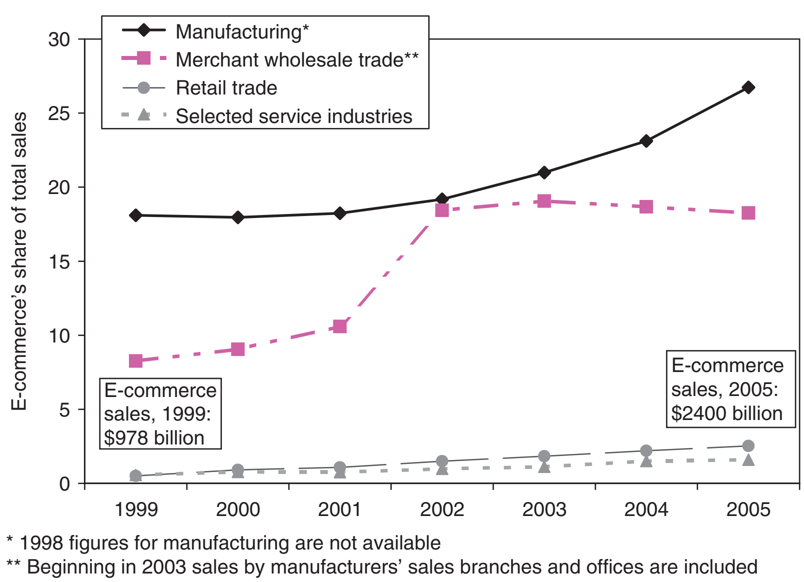

Figure 3 Size and industrial distribution of e-commerce (US data), 1998 2005. E-commerce figures are based on data from the U.S. Census E-stats Program (http://www.census.gov/eos/www/ebusiness614.htm).

and knowledge) is also likely to be evident between large and small firms and independent and branch plants.

Aoyama provides a particularly compelling example of how the variation in e commerce implementation with her study of how consumer based e commerce in Japan took a radically different path than that of the US. Rather than relying upon desktop computers and shipping to one's home, Japan developed a system of convenience store kiosks that acted as site of sales and service. Moreover, a firm's placement and power within an industry has considerable impact on how e commerce is implemented and which firms are at an advantage given a particular design. Wal Mart is well known for the considerable pressure it is able to bring to suppliers to adopt certain technologies and procedures and the US auto industry witnessed the imposition of the Covisint system on auto suppliers by the big three assembly firms. Yet again this shows the complexity and wide range of variables that shape the use of the Internet in the space economy.

An Uneven Geography of Change and Challenge

The combination of new definitions of distance and the reorganization of industries engendered by e commerce brings up important questions of how this evolving geography impacts people and places. Although it is easy to view the Internet as yet another means by which capital can be mobilized to the disadvantage of labor and place, this is an overly simplistic reading. Firms remain constrained by various economic frictions (See **Figure 2**) and some workers (generally the higher skilled) can take advantage of new spatial and temporal flexibility that the Internet offers. Mirroring the trajectory of earlier technologies, the Internet will continue the historical uneven process of economic development, albeit via networks that stretch across vast distances and in configurations that are complex and bewildering. The effect of these changes depends upon the position (both spatial and organizational) one holds in the economy. Leamer and Storper note that technologies like the Internet will "…increase product differentiation in the economy and create new forms of complex transactions, even as it simplifies others and permits further spreading out of routinized behaviors."

Figure 3 illustrates that there is high e commerce use within manufacturing and relatively low use in retailing and services. Indeed many personal services such as hairdressers and medical care and industries such as construction simply require that labor be physically present. On the other hand, information and communication technology (ICT) enabled e commerce has allowed for significant changes within a number of retail jobs. Powerful point of sale (POS) technologies within large retailers increase distribution and stocking efficiency; self check out kiosks and automated toll booths have lowered the number of needed cashiers; and fast food restaurants are using IT to relocate the workers who take drive through orders to centralized back office locations. This shifting of customer service activities has a significant effect in shaping the form and geography of work for these job but it is not the Internet *per se* that changes working conditions. Rather it is the manner in which the Internet is used that reshapes the spatial conditions of workers.

Moreover e commerce systems are helping to de skill or eliminate some jobs, for example, bar scanners and radio frequency identification (RFID) tags lowering the amount of labor needed for inventory taking. This, however, is clearly not a linear process as new jobs related to the creation and maintenance of these IT systems are also created. Although this provides cold comfort to workers who lose their jobs, especially since the location and skills required for this new work need not be anywhere near them, it demonstrates that it is not a linear process but one mediated through layers of economic friction. For example, the standoff between the longshoremen union and the port of Long Beach in the fall of 2002 was less about the introduction of IT *per se* and more about the wish of mangers to use the Internet to spatially shift work inland to sites with nonunion labor. In other words, jobs that had been confined to the port for practical reasons were suddenly open to a spatial shift because of the technology. The decision to do so as well as the resistance to it shows the larger societal and political dimensions to what at first might seem to be a simple technocratic decision.

An important concern, of course, is the extent to which the potential of the Internet and relative power differential between firms versus people and places will lead to the offshoring work or a "flat world." While clearly taking place, for example, the OECD estimates that 20% of jobs in OECD countries could potentially be done from any geographic location, offshoring is tempered by a number of issues (**Figure 2**) ranging from the use of tacit knowledge in production to concerns about privacy and the ability to access skilled labor. In short, the Internet does not lead ineluctably to offshoring, but it has allowed firms to pursue a variety of location strategies that are sometimes contradictory, sometimes complementary, and whose logics vary across space. This means that the Internet alone does not determine the fortune of people and places. Rather, the outcomes will result from sustained political contestations. While there is ready evidence of jobs being eliminated or relocated, there are also a number of factors which will keep work tied to particular locations.

Change Is Real but Geography Remains Relevant

Since its conception as a military communications system during the Cold War, the Internet has permeated across the globe and into many aspects of daily life and business. Indeed, the Internet has enabled a wide range of eco nomic activities that hitherto would not have been possible, for example, eBay's enabling small scale sales between people separated by any number of economic frictions including distance. Moreover, the possibilities are tremendous as users and firms experiment with and innovate around the Internet.

As a result there are a number of promising avenues for future research. At the core is the fundamental task of extending our understanding of how e commerce (based on Internet related technologies) changes the spatial organization of production and consumption. This article sketched the basic contours of existing research but there remains much to do. For example, has e commerce led firms to rely more upon externalized market transactions rather than keeping functions internal? In which areas have firms achieved more efficiency (profit) through the use of e commerce? How have these changes impacted the number and type of jobs within firms? How do changes in structure and efficiency affect a firm's inter action with near and distant markets and suppliers? Has e commerce made certain channels to final and inter mediate consumers possible at lower volumes? And most importantly, what are the geographies of these transfor mations in production, distribution, and consumption?

A related issue is the evolving relationship between the creation, manipulation, and utilization of information in the economy, for example, the entertainment and media industries. The rise of web 2.0 applications – the use of the Internet by online communities and social networking sites to create and collaborate – blurs the division between (and the geographies of) producers and consumers. The applications of web 2.0 include wikis (user generated re sources open to anyone to edit), folksonomies (the ag gregation of users' annotations of online resources into classifications and indices), and mash ups (user created hybridizations of two or more online services). The growth of web 2.0 means that users are increasing co producers of the information and entertainment resources they con sume and gives rise to a number of possible research topics. How will the cooperative spirit of online com munities fare in the face of efforts to commercialize as pects of web 2.0? How will geographically defined regulatory regimes such as copyright interact with this alternative (and network defined) paradigm? How will the delineation between the public and private spheres evolve? Will web 2.0 applications live up to their collaborative/ democratic ideal or be captured by narrow commercial or geographically defined interests?

Another important trend in the relationship between the Internet and economic geography is the growing ubiquity of mobile devices alongside the increasing availability and use of spatial data. Websites such as Google Maps and Microsoft Live Search Maps allow anyone to access and easily search street maps and sat ellite imagery. Meanwhile web 2.0 functionality (such as Google's My Maps) is allowing these same people to make and share spatial annotations to anywhere on the globe. As a result, users of the Internet are increasingly able to inhabit physical environments while simul taneously moving through a dynamic information cloud (accessed and edited through mobile devices) that is in dexed to their spatial location. This prompts a number of questions. How will physical proximity interact with prominence within the information cloud? How will this impact the visibility of local businesses, particularly retail locations? What new information services will emerge to take advantage of this spatialized data (particularly user generated data)?

A final direction for future research lies within the virtual games/worlds such as Worlds of Warcraft or Second Life in which millions of people spend hours a day using. Most relevant for economic geography are the worlds with tradeable currencies and active economies. Enterprising individuals within these worlds are suc cessfully constructing professions and businesses dedi cated to the creation and sale of virtual items, for example, a custom designed house that ones computer self can occupy and enjoy or other consumer goods ranging from a sports car to a customized private island. There are even sweatshops in China where workers conduct repetitive and mind numbing actions in the online worlds for hours on end to collect virtual gold pieces that are in turn sold to US players who covet digital items such as a magic sword and do not want to take the time necessary to earn the gold on their own. Key questions to be explored include, will these synthetic worlds become important economic sites in their own right or remain academic curiosities? How successfully do virtual worlds overcome the economic frictions highlighted in **Figure 2**? For example, can virtual world teleconferencing provide enough of a simulacra of co presence that it will ease the transfer of uncodifiable knowledge?

This short review of but a few possible research agendas highlights the exciting potential of the Internet and illustrates that no particular use or implementation is inevitable. Moreover, it argues against characterizations that the "world is flat" or "everything is changed." The various drags of economic frictions are still present, and are only partially and selectively being overcome. In fact, one of the great ironies of the twenty first century is that as economies and firms become more reliant upon the Internet to transcend space, the particularities and abilities of places are increasingly key points of differ entiation. In short, the Internet doesn't reduce the rele vance of geography. It simply adjusts which aspects of geography and which types of economic friction are most relevant for analysis.

See also: Agglomeration; Cultural Economy; e-Business and e-Commerce; Embeddedness; Information Technology; Internet/Web Mapping; Knowledge Economy; Mapping, Cyberspace.

Further Reading

Aoyama, Y. (2001). The information society, Japanese style: Corner stores as hubs for e commerce access. In Leinbach, T. & Brunn, S. (eds.) *Worlds of Electronic Commerce*, pp 109 128. New York: John Wiley.

Cairncross, F. (1997). *The Death of Distance: How the Communications Revolution Will Change Our Lives*. Boston, MA: Harvard Business School Press.

Castronova, E. (2005). *Synthetic Worlds: The Business and Culture of Online Games*. Chicago, IL: University of Chicago Press.

David, P. A. (1990). The dynamo and the computer: An historical perspective on the modern productivity paradox. *American Economic Review* 80(2), 355 361.

Freidman, T. (2005). *The World Is Flat: A Brief History of the Twenty First Century*. New York: Farrar, Straus and Giroux.

French, S., Crewe, L., Leyshon, A., Webb, P. and Thrift, N. (2004). 'Putting e commerce in its place: Reflections on the impact of the Internet on the cultural industries'. In Power, D. & Scott, A. J. (eds.) *Cultural Industries and the Production of Culture*, pp 54 71. London: Routledge.

Leamer, E. and Storper, M. (2001). The economic geography of the Internet age. *Journal of International Business Studies* 32(4), 641 665.

Leinbach, T. (2001). Emergence of the digital economy and e commerce. In Leinbach, T. & Brunn, S. (eds.) *Worlds of E Commerce*, pp 3 26. New York: John Wiley.

Leyshon, A. (2003). 'Scary Monsters? Software formats, peer to peer networks and the spectre of the gift'. *Environment and Planning D: Society and Space* 21, 533 558.

OECD (2006). *Information Technology Outlook*. Paris: OECD.

Porter, M. (2001). "Strategy and the Internet". *Harvard Business Review* March, 63 78.

Townsend, A. M. (2001). The Internet and the rise of the new network cities, 1969 1999. *Environment and Planning B: Planning & Design* 28(1), 39 58.

Zook, M. (2000). The web of production: The economic geography of commercial Internet content production in the United States. *Environment & Planning A* 32(3), 411 426.

Zook, M. (2001). Old hierarchies or new networks of centrality? The global geography of the Internet content market. *American Behavioral Scientist* 44(10), 1679.

Zook, M. (2005). *The Geography of the Internet Industry: Venture Capital, Dot Coms and Local Knowledge*. Oxford: Blackwell Publishers.

Relevant Websites

http://internetworldstats.com
Internet Usage World Stats.

http://www.oecd.org
OECD's Directorate for Science, Technology and Industry.

http://www.census.gov
U.S. Census Bureau E stats Program.

http://www.zooknic.com
Zooknic Internet Geography Project.

Internet/Web Mapping

M.-J. Kraak, ITC – International Institute of Geo-Information Science and Earth Observation, Enschede, The Netherlands

Glossary

Cartography The art, science, and technology of making and using maps.

Geodata Infrastructure (GDI) Geospatial data infrastructure encompasses the networked geospatial databases and data-handling facilities, the complex of institutional, organizational, technological, human, and economic resources which interact with one another and underpin the design, implementation, and maintenance of mechanisms facilitating the sharing, access to, and responsible use of geospatial data at an affordable cost for a specific application domain or enterprise.

Map A symbolized representation of a geographical reality, representing selected features and characteristics, resulting from the creative effort of its author's execution of choices, and is designed for use when spatial relationships are of primary relevance.

Web Map A map presented in a web browser.

Introduction

The environment in which maps have been produced and used has changed considerably. During the last two decades, cartographers, geoscientists, and computer scientists have witnessed highly dynamic and important developments in the fields of acquiring, managing, analyzing, interacting with, and visualizing large amounts of geospatial data. Especially, the field of visualization has undergone a rapid change. Next to static maps, nowadays, immersive and highly interactive virtual environments can be used to explore and present dynamic geospatial data. The World Wide Web (WWW) is today's prominent medium to disseminate geospatial data and maps. In the past producing maps was the realm of cartographers. Their skills and laborious activities were a guarantee that authorative maps were produced. The rise of geographic information systems (GISs) has increased those involved in making maps. Today's revolution around the WWW has even further increased the number of people involved in making maps. Making maps via GIS still involves geo professionals; however, making maps via the WWW includes potentially everyone having access to this new medium. These new mapmakers have different demands, and force a change from a traditional supply driven map production to demand driven map production. Irrespective of these changes, it is still about maps, which remain the valuable products they used to be. Expertise is still required, but differently, since the role of maps has changed and expanded. In the past (paper) maps were designed to be both database and presentation media. Digital cartography and GIS have split these tasks, but these days the link between the two tasks can be even better preserved. In modern on screen environments, one can store much more data 'behind' the map, than on a single paper map. Linking multimedia elements to the map can enrich them even more, especially on the WWW with its hyperlink options and capability to disseminate interactive three dimensional (3 D) maps. The new mapping environments can be characterized by two keywords (interactivity and dynamics) and as such introduced the word visualization, which also requires a different view on map design, especially with the WWW as the output medium.

Internet Technology and Maps

The Internet originates from the United States Advanced Research Projects Agency Network (ARPAnet), a military funded communication network, which was started in 1969 to allow collaboration among researchers at different locations. With the adoption of transmission control protocol/internet protocol (TCP/IP) by ARPAnet, user network (USEnet), and EUnet, the network expanded globally in 1983. This moment can be seen as the birth of the Internet. The layer structure of TCP/IP allowed other standard protocols like File Transfer Protocol (FTP) for file transfer, telecommunication network (TelNet) for terminal services, and simple mail transfer protocol (SMTP) for e mail to be incorporated. In 1993 with the definition of the hypertext transfer protocol (HTTP) and hypertext markup language (HTML) protocols at European Organization for Nuclear Research (CERN), the WWW was born.

The WWW can been seen as a network of linked documents, pictures, videos, and other information sources. For the layout of individual web pages, HTML is still widely used. For the exchange of the content (documents, etc.), extensible markup language (XML) is more and more used. It encodes the content in a (text) document for storage and transfer purpose, and it also describes the structure of the data. This combination makes XML documents self descriptive. Individual tags can be defined which allows for XML languages for different purposes, like the geographical markup language (GML) for geographic data, and, for instance, also mathematical markup language (MathML) for mathematical formulas.

Until recently, maps presented on the web have been raster based. They had a fixed resolution, and would not allow for interactivity, other than clickable maps. Common file formats are Graphics Interchange Format (GIF), Joint Photographic Experts Group (JPEG), and Portable Net work Graphics (PNG). With the introduction of the Scalable Vector Graphics (SVG), resolution independent vector graphics can be presented. SVG is an XML based language for 2 D vector graphics. It allows for many interactive functions such as zoom, pan, different trans formations, as well as dynamic change of object attributes.

The so called client–server architecture plays a prominent role in how users are able to display a map on a screen. Via a web browser (the client), one can send a request to a web server (the server), which will respond and return the requested item. In a simple setup this item could be a JPEG file that contains the image of a map.

Offering interactivity to the user makes the client–server structure more complex. As a solution, one can work with plug inns on the client side, for example, the user has to download a program to be able to enjoy the interactivity. When most activity takes place on the ser ver, but required data to be sent up and down between the client and server, one deals with so called thin clients. A thick client results in less data traffic and more activity on the user side. There are many different alternative solutions. Popular are the so called distributed systems. Here a set of independent servers each execute their own specific task, but a user will not notice this. An online bookshop is a good example. If one orders this book, Elsevier will send it from its warehouse, the payment will go via a credit card company, but the customer will be lieve he/she only dealt with the online bookshop.

A distributed approach puts demands on the individual components that have to work together. One has to realize seamless subsystem integration. Technology to be used should allow the interoperability of heterogeneous com ponents without knowing their implementation details. The Open GeoSpatial Consortium (OGC) has set up a set of specifications to make this work in the geo world. In this context, one has to specify and set up an infrastructure of interoperable web services. A service can be described as a distinct part of an overall functionality which is accessible via standardized interface. It could, for instance, be an algorithm that calculates the shortest route between two locations. An interface describes the methods that the software component implements and their calling con ventions without specifying how it is implemented. It is obvious that standardization is an important keyword. The OGC and ISO TC/211 are the important standardization organizations. The first has defined rules for the web map service (portrayal of maps), the web coverage service (interface to raster data), and the web feature service (interface to vector data). For the definition of the data a geoversion of XML, GML is used. A GML description includes both spatial and nonspatial properties of geo graphic objects. To be able to draw a map, one has to transform the GML data using a portrayal service, which results in a map graphic, for example, using SVG.

Maps on the Internet

Maps on the WWW can play different roles. Of course, they will function as they did before. They will present a message by showing geospatial patterns and will increase insight in geospatial relations. However, on the WWW the map can function as an interface to other (geo) in formation in cyberspace or will guide the surfer navi gating the WWW. In providing access by clicking objects, the map will bring the surfer to other web pages, which can contain again maps, photographs, or text. In this re spect the map could play a prominent role in a country's geospatial data infrastructure (GDI). In its role of guiding the surfer through parts of cyberspace, maps will help to keep track of paths through, for instance, a single website.

Why is the WWW an interesting medium to present and disseminate maps and/or the data behind those maps? The answer is that information on the web is virtually platform independent. Also, many users can be reached at minimal costs and it is easy to update the maps frequently. Furthermore, the WWW allows for a dynamic and interactive dissemination of geospatial data. This results in new mapping techniques as well as new use possibilities not seen before with traditional printed maps and most on screen maps. While preparing maps for the WWW, one has to remind that its access is geographically skewed, for example, some part of the world still have poor access, such as Africa, Asia, and parts of Latin America. Also, one has to remember that the WWW is a fast medium used by impatient people. Long download times will make users lose their interest. In addition, people expect up to date information, and if a site appears older than several weeks, it will be considered of no value.

Type of Maps

What type of maps, from both provider and user per spective, can be found on the web? **Figure 1** presents a classification. It gives an overview of current possibilities, from the perspective of how the map image is used based on the keywords, interaction and dynamics. The scheme distinguishes between nondynamic and dynamic web maps. Each of these categories is further subdivided into noninteractive and interactive maps. The most common map found on the WWW is the static noninteractive view only map (see **Figure 2**). In most cases, these are scanned existing paper maps. This form of presentation

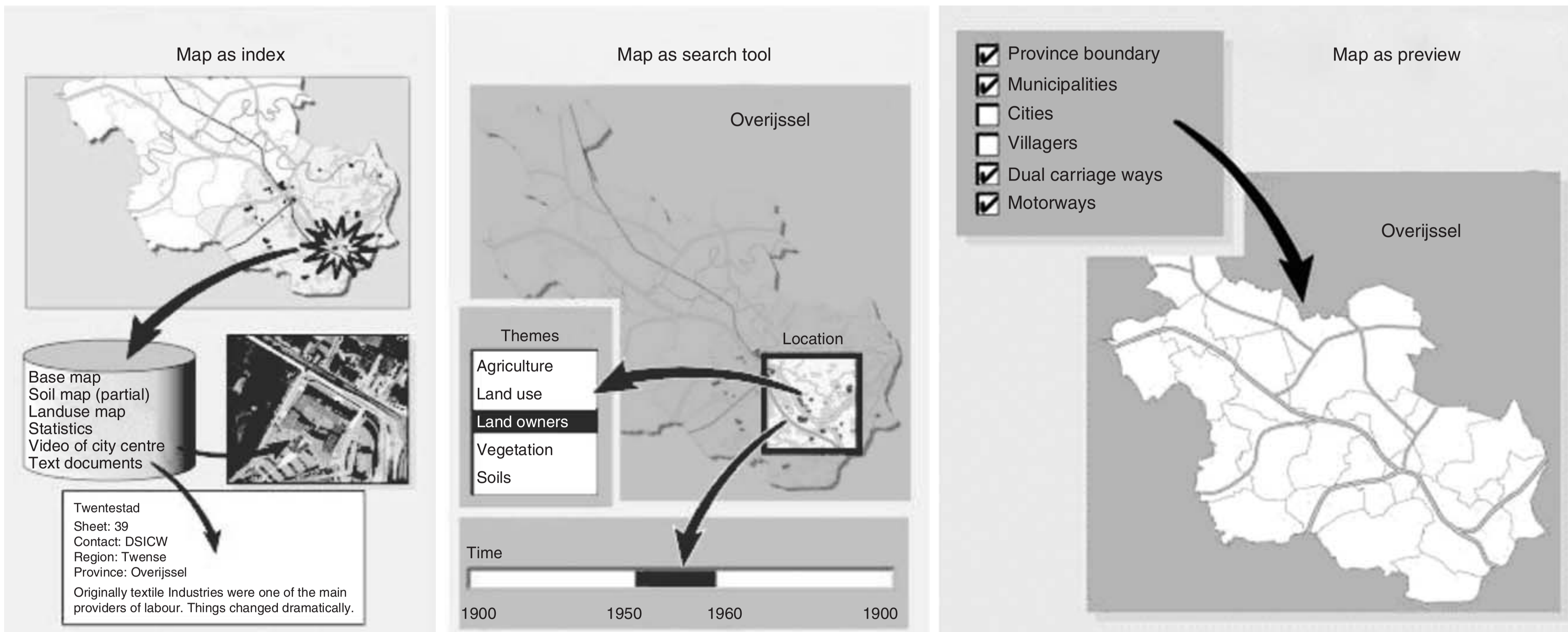

Figure 1 Add t ona ro e of the map on the WWW: funct on as an ndex (eft); search too (center); prev ew (r ght). From Kraak, M.-J. (2000). Access to the G , the funct on of v sua zat on too s. n Groot, D. & McLaugh n, J. (eds.) *Geospatial Data Infrastructure – Concepts, Cases, and Good Practice*, pp 217–232. Oxford: Oxford Un vers ty Press.

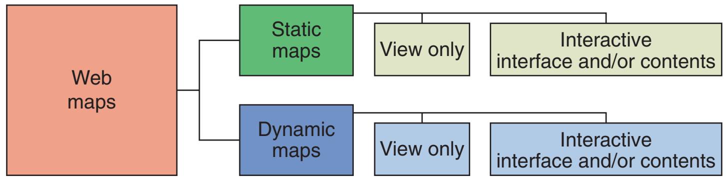

Figure 2 Classification of web maps. From Kraak, M.-J. (2000). Settings and needs for web cartography. In Kraak, M.-J. & Brown, A. (eds.) *Web Cartography Developments and Prospects*, pp 1 8. London: Taylor & Francis.

can be very useful, for instance, to make historical maps more widely accessible. To have such maps on the WWW allows one to study and share (inter) national cartographic heritage. The nondynamic view only maps can serve to give the web surfer an impression of avail able products, and the National Mapping Organizations can show details – teasers – of their map series. A problem with many of the nondynamic view only maps is that they have not been designed for the WWW, and the colors of the original maps will often not reach the user as planned, due to technical (color) constrains. Another factor that can cause problems is the fact that most ori ginals have an information density too high for the WWW.

The nondynamic maps can be interactive. These are the so called 'clickable' maps. In this case, the map functions as an interface to other data. Clicking on a geographic object could lead to other information sources on the web, such as other maps, images, or text. Inter activity could also mean that the user has the option to zoom and pan. Google Maps has made this map type to be the minimum in options expected by users. Alter natively, it could allow the user to define the contents of the web map by switching layers off or on. Sometimes, it is also possible to choose symbology and colors. The WWW is especially suitable to present dynamic maps, like animations that represent dynamic (geo) processes. The so called animated GIF can be seen as the non interactive version of the dynamic maps. Several bitmaps, each representing a frame of the animation, are pos itioned after each other and the WWW browser will continuously repeat the animation. Famous examples are the commonly used spinning globes (URL 6). Slightly more interactive versions of this type of maps are those to be displayed by media players in, for instance, Moving Picture Experts Group (MPEG) format. The WWW browser and its plug inns define the interaction options, which are often limited. More interaction can be created by JAVA, JavaScript, or virtual reality modeling language (VRML). VRML especially offers opportunities to define the travel path through 3 D scenes.

Design of Maps

What makes web maps different? Starting point is that web map design is no different from other map design. That is, while producing the map one will apply the cartographic grammar considering the nature of the data to be represented, and the message the map should tell. Normally, one would include audience in the previous sentence as well. However, for maps presented on the web, the audience is very much unknown. Even though one might aim at a particular user community, one will never be sure of who is looking at the maps. An exception might be if the maps are presented in a 'closed' intranet environment. In some cases the geospatial data provider will not even be sure on how their maps look like at all. This might be partly due to the user's web browser en vironment, but can be due to interaction and symbol ization options given to the user as well.

The fact that most maps have to travel half the globe before they reach the (unknown) user's browser dictates that the web map should not be too large in both image and file size. The image size should be limited because scrolling up and down or left and right is not promoting a sense of map overview, and the file size should be kept to a minimum while it is unlikely the user will be willing to wait for the map to download. The map contents (information density) should be low as well, because a typical user surfing the web has a relatively short span of attention. All this may look disadvantageous but it offers a challenge to produce well designed maps. It is possible to put all kinds of additional (or omitted) information behind the map image, because of the interactive nature of the WWW. Clicking objects or mouse over techniques can make this (additional) information accessible. These techniques could also be used to display text on the map. Presenting text in an on screen map is, from a point of visibility and legibility, rather problematic and could be solved via mouse over techniques. When the user moves the cursor over the map and it hits a town symbol, it could show the name of the town.

Mouse over techniques can also be used to highlight geographic objects based on user interaction. The basic map could be viewed in pastel tints while moving the mouse around will present the object under the pointer in bright colors. Design techniques in relation to clarity (crispness of the symbol's appearance) and transparency can be applied. Shading could be applied in combination to offer a 3 D look to those symbols selected. In summary all these 'special effects' can not just be used to make the maps attractive, but also to provide additional infor mation or help in navigating the web map, and function even as a map legend. Clicking a map object could access a database with additional information related to the object. The application of such an approach depends on what information the provider wants to put on the WWW: present a map or offer geospatial data. In the design of web maps, color is a separate issue. It can be used to realize many of the above effects, but it is also one of the weak points of web browser environments. Although computer screens nowadays can present millions of colors, even more than the human visual system can handle, a safe use of color would limit one to the 246 colors of the so called web safe color palette.

How to finally design or produce the map will depend on the provider's aims. Questions such as which type of web map (nondynamic/dynamic) is applicable have to be answered. When it comes to production, the mapping process will often have to be adapted compared to traditional mapping, if one intends to use the new opportunities offered by the WWW as medium to present maps. Another important question: what (design) action do we leave to the user? The possibilities offered by the WWW have extended the traditional cartographic variables. Web design software enables the application of new variables, like blur, focus, and transparency, while shadow and shading play a prominent role as well. Blur give symbols a fuzzy appearance and can, for instance, be applied to visualize uncertainty, while focus will introduce blinking symbols to attract attention. Both transparency and shading/shadow can be used to simulate a 3 D look. Transparency can be seen as a kind of fogginess, by which part of map content is obscured or faded in favor of other information. For example, it can be applied to subdue the background in a map in order to enhance the main theme in the foreground (for instance a drape of geology over terrain features). In a 3 D 'landscape' environment it can also be used as a depth cue. The use of shadow and shading increase the sense of depth. Shading is commonly used to increase the contrast between 'figure' and 'ground' or, for instance, in relief maps to create a 3 D terrain impression. Shadow, also known as cast or drop shadow, can be applied to give the symbols a 3 D look. In web maps, this 3 D feel of the symbols invites users to click on them to activate hyperlinks or mouse over effects. The visual effect of shadow is casting a shadow of the symbol on to the background. The origin of these new variables is partly found outside cartography; however, it is interesting to note that the cartographic principles have found their way in noncartographic literature (**Figure 3**).

The new variables have been fully accepted in cartography. However, further research on the new variables effectively remain to be done, because despite all technological progress it remains important to wonder if the new representations and interfaces really work. An advantage of those empty maps is that they can be easily used in the world of mobile geocomputing. People will have portable devices such as mobile phones and/or personal digital assistants and call in for location information. Small maps with answers to questions such as how to get to the railway station, or where the nearest bookshop is, will appear on the small screens of these

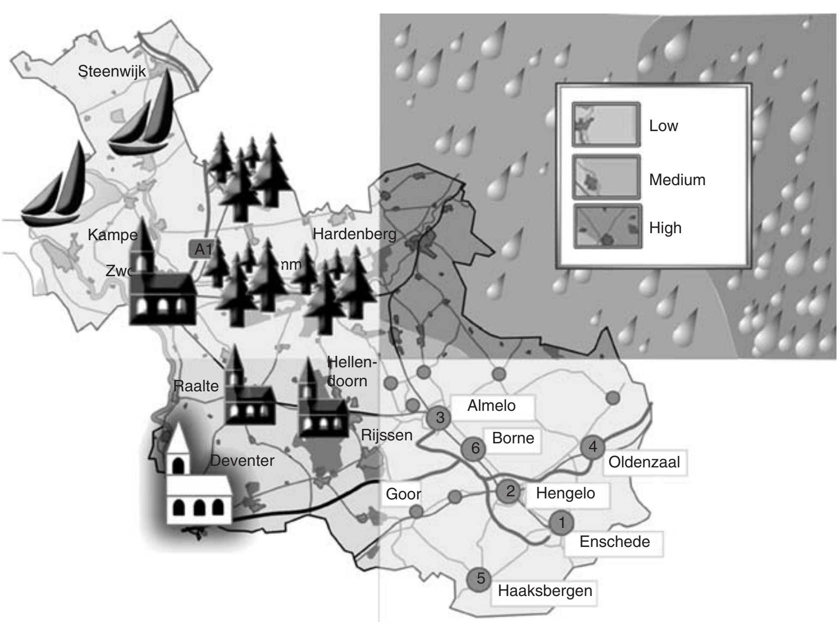

Figure 3 New graphic variables: shadow/shading (upper left); blur (lower left); transparency (upper right); and blinking/focus (lower right). From Kraak, M.-J. and Ormeling, F. J. (1996). *Cartography, the Visualization of Spatial Data.* London: Addison Wesley Longman.

devices which often will have a Global Positioning Sys tem (GPS) receiver included to establish the owner's position. Those small screens require an economized map design. This should not be considered problematic since one can include lots of information behind the map image or individual symbols. Access to this hidden in formation can be obtained via, for instance, mouse over techniques or clicking map symbols. To ensure this approach, one has to make sure the symbols have an appearance that invites clicking them. In case of mouse over techniques, the appearance of a symbol will change or textual information appears on the map. Clicking the symbol might open new windows or activate other web pages.

Maps and the Geodata Infrastructure

To realize the availability of maps over the Internet, not only the technology as described in the introductory section should be available, but the infrastructure to disseminate and access the maps should be available too. The need for such infrastructure is there because the traditional map and data supply driven environment is changing rapidly in a map and data demand driven en vironment. A more flexible approach is required because more data is available from different sources and more non experts are using the data, and the web has become the predominant working environment and dissemination channel. Many data providers are offering their maps and data via the web and a start has been made to create portals, also called clearinghouses, where users can get access to the data in a kind of one stop process. This has been facilitated by a so called GDI. A GDI encompasses the networked geospatial databases and data handling facilities, the complex of institutional, organizational, technological, human, and economic resources which interact with one another and underpin the design, im plementation, and maintenance of mechanisms facili tating the sharing, access to, and responsible use of geospatial data at an affordable cost for a specific appli cation domain or enterprise.

These developments have also introduced distributed geoservices. Based on the request, geoservices can in clude the delivery of data and maps, but they can also offer computational support to convert coordinated be tween differently systems. An example of this last option is a service that converts a GPS track collected by the user in a format that can be used in particular software. Other examples of geoservices are gazetteers where one types a geographic name and the service returns its lo cation in a map or as position in longitude and latitude. A well known service is offered by route planners. One gives a departure and destination, and the geoservice returns a driving advice either in a map or a list with directions. It is also possible to combine different geo services in a single solution, for instance, in a crisis management center, one is in need of different datasets to monitor events. The data are often supplied by different organizations and are traditionally difficult to obtain and combine. The geoservices offer the possibility – again while adhering to Open Geospatial Standards – to combine the distributed data in a single application and keep track of events. If one of the suppliers update their data, this new information can be automatically streamed into the application of the crisis management center. Even field data collected with, for instance, a personal digital assistant (PDA) can be automatically forwarded to the center or the center can supply the field with the latest overall information.

The web map service of National Geographic is also an interesting case. Besides a full fledged web map ser vice where the user can select between different map types and satellite imagery while zooming in from world to street level, the website also offers 'print friendly' maps which provide basic maps of countries and regions. Strength of the website is the possible link between maps and the other resources of the National Geographic Society.

Google offers Google Earth, a mapping product that combines an intuitive interface with a good map design, the wide availability of data, and the possibility to link your own data to the maps and satellite imagery. In fact, Google Earth is a good example of a geoservice because it streams satellite and other map data from external servers, while these graphics can be combined with other data provided by Google, such as, for instance, National Geographic photographs. Users can also add their own data to the imagery. These could be GPS tracks of tours or scanned maps that are draped over the imagery. The user can set the transparency of the maps added and as such compare, for instance, a scanned topographic map with the 3 D Google Earth model. Adding 3 D objects, such as buildings, is also possible. Next to these static examples, it is also possible to create dynamic Google Earth web services in which live data streams are added to the imagery. An example is the Tour de France, Europe's biggest annual cycling event. Users can do a live tracking of the stages because every twenty seconds, the position of the cyclist is updated and shown on the imagery.

Another trend that allows users to participate in web mapping is realized via so called mash ups. Here users can incorporate particular web services in their own web page. For instance one could show part of a map offered by Google and add a symbol to indicate the web page owner's home location. In addition, this location could be the start/end point for route calculation to offer visitors of the web page the fastest/shortest path to that location from the user location.

Conclusion

An important advantage of the Internet is that the geospatial data and maps can be easily reused for different purposes. Somehow, the same data is offered by different web map services to both professional and public user groups. This, for instance, results in road datasets created by national mapping agencies being enriched by commercial vendors who offer the data to the public via online route planners. Tomorrow's picture will not only show a complete merger of both professional and public mapping worlds, but the geodata infrastructure, and as such maps, will be incorporated in a larger information and communication infrastructure. External (technological) influences will keep dictating the future appearance of the map, as has been the case in the past.

See also: Digital Divide; Digital Earth; GIS, Mobile and Locational Based Services; Map Hacking; Map Interactivity; Mapping, Distributed.

Further Reading

Cartwright, W., Peterson, M. M. and Gartner, G. (eds.) (2007). *Multimedia Cartography*. Berlin: Springer.

Erle, S., Gibson, R. and Walsh, J. (2005). *Mapping Hacks: Tips and Tools for Electronic Cartography*. Sebastopol, CA: O'Reilly Media.

Groot, D. and McLaughlin, J. (eds.) (2000). *Geospatial Data Infrastructure Concepts, Cases, and Good Practice (Spatial Information Systems)*. Oxford: Oxford University Press.

Kraak, M. J. (2000). Access to the GII, the function of visualization tools. In Groot, D. & McLaughlin, J. (eds.) *Geospatial Data Infrastructure Concepts, Cases, and Good Practice*, pp 217 232. Oxford: Oxford University Press.

Kraak, M. J. (2000). Settings and needs for web cartography. In Kraak, M. J. & Brown, A. (eds.) *Web Cartography Developments and Prospects*, pp 1 8. London: Taylor & Francis.

Kraak, M. J. and Brown, A. (eds.) (2000). *Web Cartography Developments and Prospects*. London: Taylor and Francis.

Kraak, M. J. and Ormeling, F. J. (1996). *Cartography, the Visualization of Spatial Data*. London: Addison Wesley Longman.

Mitchell, T. (2005). *Web Mapping Illustrated: Using Open Source GIS Toolkit*. Sebastopol, CA: O'Reilly Media.

Plewe, B. (1997). *GIS Online: Information Retrieval, Mapping and the Internet*. Santa Fe, NM: OnWord Press.

Taylor, D. R. F. (ed.) (2006). *Cybercartography: Theory and Practice*. Oxford: Elsevier Science.

Relevant Websites

http://earth.google.com
Google Earth.

http://plasma.nationalgeographic.com
MapMachine, Maps, National Geographic.

http://www.opengeospatial.org
Open Geospatial Consortium.

Internet-Based Measurement

M. Dodge, University of Manchester, Manchester, UK
M. Zook, University of Kentucky, Lexington, KY, USA

Glossary

Domain Name A domain name is a unique identifier (e.g., nytimes.com or manchester.ac.uk) associated with an IP address that allows users to easily access a specific Internet resource such as a website.

IP Address Internet protocol (IP) addresses (e.g., 64.246.60.38) uniquely identify sites on the Internet and are necessary to ensure the delivery of traffic. They are little seen or used by typical users.

Latency The time (measured in milliseconds) that it takes to transmit and receive data between two nodes on the Internet.

Ping A network utility which sends test data to a target Internet site to determine whether it is 'live' and accepting data and reports the latency to the site.

Screen Scraping The use of a computer program to automatically collect online data or data output of Internet-based resources – most often a web page.

Topological Location The location of sites on the Internet in terms of how they are connected to the rest of the network rather than a point defined by latitude and longitude.

Traceroute A network utility which maps out the path that data packets take between two hosts on the Internet, showing all of the intermediate nodes traversed, along with an indication of the speed of travel for each segment of the journey.

Whois A utility to search for the ownership details of a domain name.

Internet based measurement is a set of methods that have been applied to quantitatively describe the structure, workload, and use of the Internet. They provide a practical means of doing a kind of virtual 'fieldwork' on the Internet using online tools and network monitoring techniques to gather fine scale primary data, as opposed to relying on aggregate secondary data sources (such as government statistics). Typically, these methods use freely available software tools and web resources to explore the internal topology of Internet links and/or the external geography of network infrastructure, content production, and use. By measuring the operation of the Internet in terms of where things are produced and consumed, who owns them, and how data travels, researchers are able to critically engage and analyze the key network of the information age. In pedagogic terms, the openness of these techniques can help users of the Internet to transform themselves from passive consumers to more informed and active explorers of their world.

Internet based measurement as a methodology for human geography is important and innovative in several respects. First, the Internet itself has several important geographical dimensions, and quantitative measurement techniques can provide unique data to analyze this. The focus has been primarily on mapping the material geography of the infrastructures of the Internet via automatic surveys of connected hardware and software services. Knowing where things are physically located is useful analytically because variations in spatial patterns provide researchers insight into underlying processes.

Second, 'ordinary' users can explore and measure the structure and operation of the Internet for themselves. This is because the Internet was purposefully designed as an open network that encourages active exploration and experimentation. This methodology does not require a costly investment in expensive, specialized tools to obtain large and representative data samples. Many of the tools and techniques for Internet based measurement already exist, having been created by engineers for the practical purposes of debugging network problems. These tools can be successfully leveraged to generate data useful for the context of social science research questions by providing tactical knowledge of the network that cannot be gained in any other way.

Using the Internet to measure and map itself is particularly useful for studies of the social geography of online interaction or the economic geography of website production where location is a key variable. For example, Internet based measurement can reveal how territorial geographies of regulation and enforcement, particularly obscenity or libel laws, help shape the location of Internet activity. In addition, because the freedom to surf the web is not universal, Internet based measurement methodologies are useful in identifying and evaluating state attempts to censor the dissemination or consumption of information by their inhabitants. Technically, censorship is often performed on behalf of a national government by commercial Internet service providers (ISPs) who block content reaching customers using a list of banned websites or domain names. Conceptually, this is the same process as software filters that parents can install on individual personal computers (PCs) to block access to inappropriate materials by their children. However, given the dynamic nature of the

Internet, this type of censorship is often partial and re quires authorities to continuously update the list of blocked sites. More insidious censorship can also be undertaken by search engine companies, who adapt their algorithms to give different, selective, results to users in particular countries. Crucially, users are not made aware that their results have been altered; for example, Google's agreement with the Chinese government means it alters key search results for Chinese users, including searches on Tiananmen Square, Tibetan independence, and Falon Gong.

Third, the Internet is a repository of a wide range of data which can be collected and cross referenced to allow researchers to create databases that measure offline phenomenon such as fine grained geographies of crime statistics or apartment listings. Geographic location (e.g., postal codes) is one of the most effective means of in dexing data (Internet based or otherwise) since it enables linkages to a vast array of existing secondary data, such as demographic statistics from standard censuses. More fundamental, however, is the ability of empowered users or social movements to aggregate data spread across multiple web sources at relatively low cost to cast new light on long standing problems.

Locating the Nodes of the Internet

The ability to reliably determine the geographic location of the nodes of the Internet's infrastructure is the first step for Internet based measurement. This task is chal lenging as the Internet was designed as a logical network that only 'knows' about topology (i.e., the location of connections capable of exchanging data) which can have little to do with physical location, defined by geographic coordinates. Thus, while the Internet has a robust and scalable system of unique locations (e.g., identifiers like Internet protocol (IP) addresses or domain names), these locations are not fixed to physical points on the ground or any particular position in the network. Moreover, be cause the Internet is a network of networks, rather than a homogeneous entity, the control of these location iden tifiers is decentralized and fluid. In short, 'no one owns the Internet' as a whole, and instead each component part is owned and operated by many different organizations and individuals. Consequently, no one institution has a synoptic view of the whole Internet and no one maintains a register of where all the components are physically located.

Determining the geographical location of components of the Internet is further complicated because different characteristics of a site's operation can be in different places. For example, there are five distinct kinds of geographical location which are important for fully characterizing a website (or other Internet resource). These types of locations are:

1. lexical: a website is where its content refers,
2. hardware: a website is where its hardware server is physically located,
3. production: a website is where the author/maintainer responsible for it is located,
4. ownership: a website is where the legal owner is lo cated, and
5. users: a website is where its users are located.

In some cases, all five locations will be largely coincident geographically (e.g., a university's website). However, it is easy to imagine plausible scenarios in which a web page providing information on vacationing in Lexington, Kentucky, is hosted on a server in London, written by someone in Manchester for a website owner in Miami, which is read by people from across the world. The geographical precision of these different physical lo cations can also vary. Sometimes, location might be de termined as the precise x, y position (e.g., street address of the building containing the web server); other times one might only know city or national jurisdiction. Each type of geographical location of a website is determined via different techniques.

Geography of Content

The first, and most obvious, method for determining the location of a website is based on lexical geography. Here, the content of the website is browsed to try to find an 'about page' or 'contacts page' that provides a postal address or telephone number for the website. Other cultural and linguistic clues (e.g., flags, symbols) in the content of a website might give useful indications of 'real world' location. This method, however, is far from foolproof as many Internet resources do not provide readily identifiable measures of this type or do not allow a researcher access, for example, password protected sites. Additionally, it is an extremely time consuming approach as it requires a human being to visually inspect and categorize each site.

Geography of Hardware

The second, and arguably most straightforward geo graphic measure of the Internet relies upon IP addresses. IP addresses are unique numeric identifiers, for example, 169.229.39.137, assigned to networked computers to ex change data. A variety of private and public databases exist that provide the associated geographic information for a particular IP address (**Figure 1**). While not fool proof, geo coding IP addresses are reasonably accurate (particularly at the national level) and are widely used by companies to track users, guard against credit card fraud, and provide web content tailored for different territories. Moreover, it is possible to automate this process via software scripts in order to locate tens or hundreds of thousands of IP addresses in a very short amount of time.

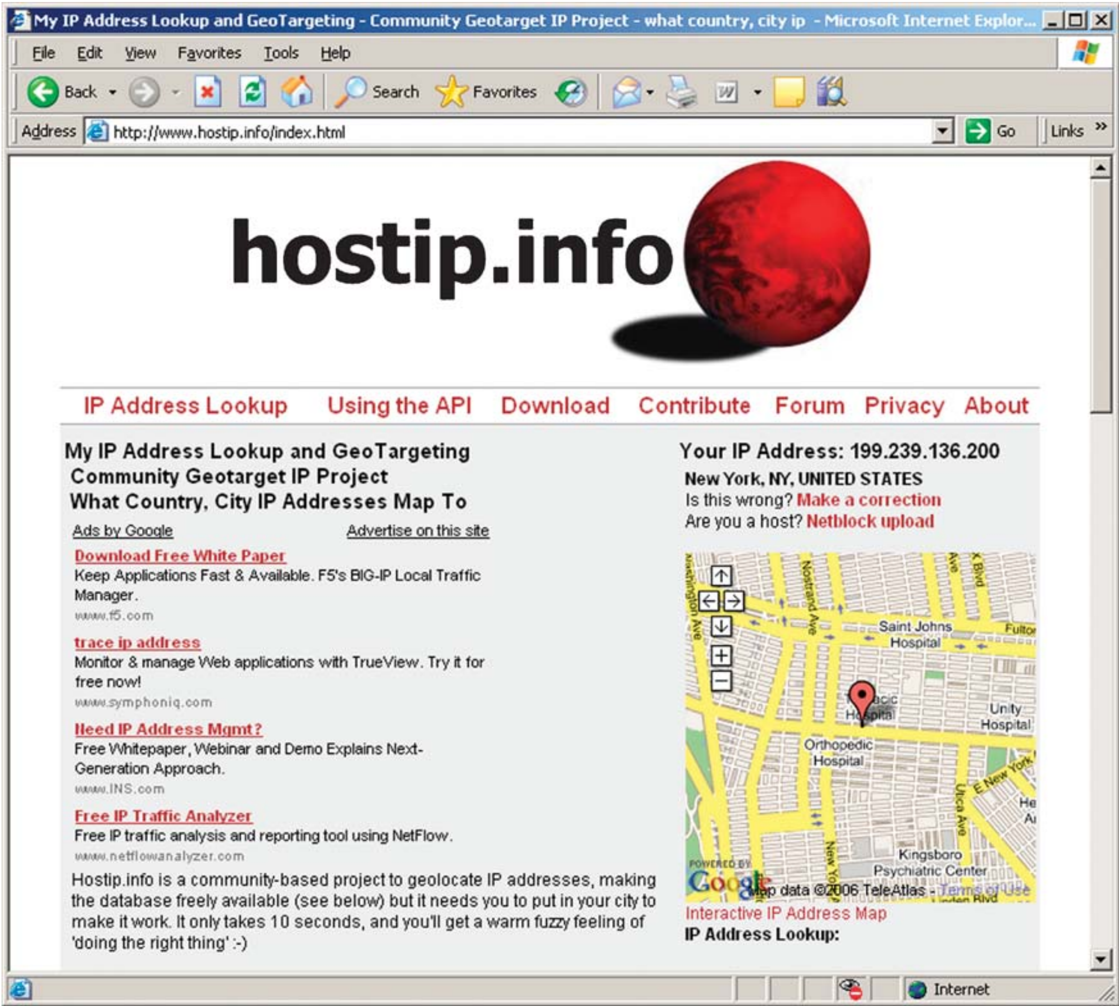

Figure 1 The result of a database lookup on an IP address using the web service offered by hostip.info that gives a geographic address for the registered owner. Source: author screenshot.

The weakness of IP address geo coding is that the use of anonymizers and other techniques can mask a user's actual location. More important for researchers is that most websites are hosted at dedicated server farms that have little to nothing to do with the location where the content for the site is generated or where the owner is located. Thus, IP addresses often highlight Internet infrastructure locations rather than content production centers. The applicability of this, of course, depends upon the research question pursued.

Geography of Production/Ownership

Because IP addresses are awkward for people to use, the domain name system (e.g., nytimes.com or manchester.ac.uk) was introduced in the 1980s and now comprises a key component for Internet navigation and measurement. Domain names are organized according to top level domains (TLDs) consisting of country code TLDs (ccTLDs) associated with domains ending with two letter International Organization for Standardization (ISO) country code (e.g., .ca for Canadian domains, .ie for Irish domains, etc.) and generic TLDs (gTLDs) such as .com, .net, or .org. Approximately 35% of all domains are under ccTLDs and provide a crude measure of geographic location. However, the use of a country code domain name does not guarantee that the website is actually within the country indicated. The ownership, production, hosting, and use of that website could well be in another country or several different countries. Furthermore, approximately 65% of domain names fall under the category of gTLDs and are not related to any country.

Thus, a much more accurate geographic location for domain names is derived via the online utility known as 'whois' which provides the ownership (listed as the registrant) information for a particular domain. Generally, it is possible to freely consult this registration information via a whois query but not all domain registration databases publicly give out the full address details of the owner. A whois query can be done interactively from any number of websites (**Figure 2**) and multiple whois queries can also be automated using software scripts.

While the results of whois queries can be helpful in finding out where the registered owner of a domain name is, they are not always accurate. First, registration details held on a given domain name may be out of date, incorrect, or deliberately false (e.g., spammers try to hide their true geographic location and would be unlikely to complete the registration honestly). Second, registrant information from a whois query only provides one location for a domain and it is not possible to determine whether this is indicative of the site of ownership, production, or both. Third, the registrations for large organizations often give a single postal address (their headquarters) and, thereby, may mask where the content for these individual domain names is actually being produced.

Despite these issues, the technique of using whois information for geo coding websites were vital to the research by Matthew Zook on the geography of Internet content production. Using automated whois queries, he gathered comprehensive data on the location of .com domain name registrations in 1998 and demonstrated that the production of Internet content exhibited a significant degree of clustering in particular cities in the US and globally. Relying upon these techniques (supplemented with the use of IP location data), he has analyzed the geographies of a number of Internet based activities

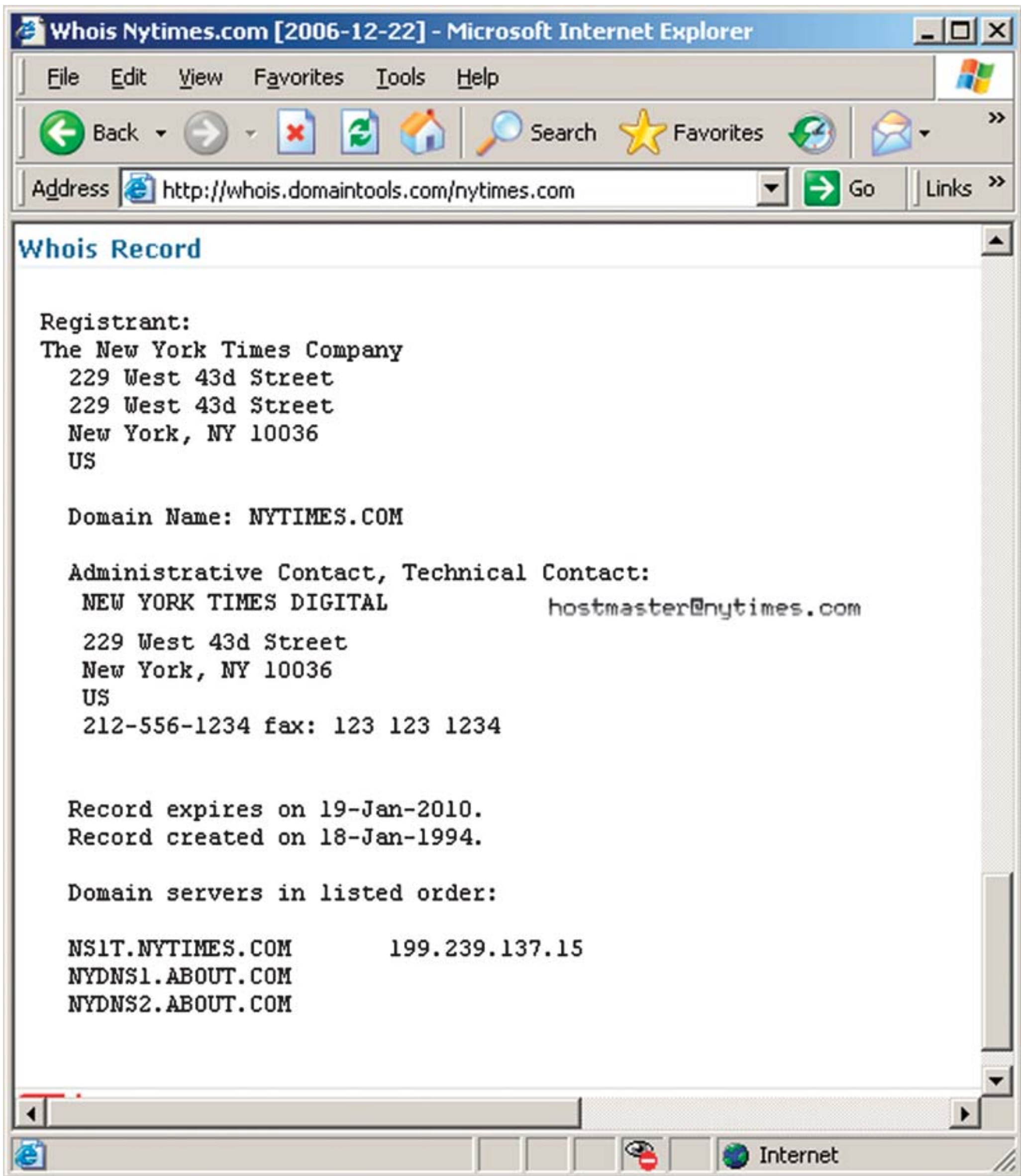

Figure 2 The result of a whois database lookup on the nytimes.com domain name using a free web service called domaintools.com. The output gives registration details including the postal address of the owner. Source: author screenshot.

ranging from the clustering of Internet startup companies during the 1990s to the location of adult oriented web sites (**Figure 3**).

Geography of Users

The location of users of websites is arguably the most difficult to measure as it is the least centrally organized aspect of the Internet. Moreover, it is an ill defined and dynamic variable as new people continuously come online and existing users adapt their online practices. At the level of individual websites, however, it is possible to gather rich data on the number, location, and activities of users (**Figure 4**). While potentially helpful, it can only shed light on the users of a particular website and gaining access to the user logs of leading websites is a difficult

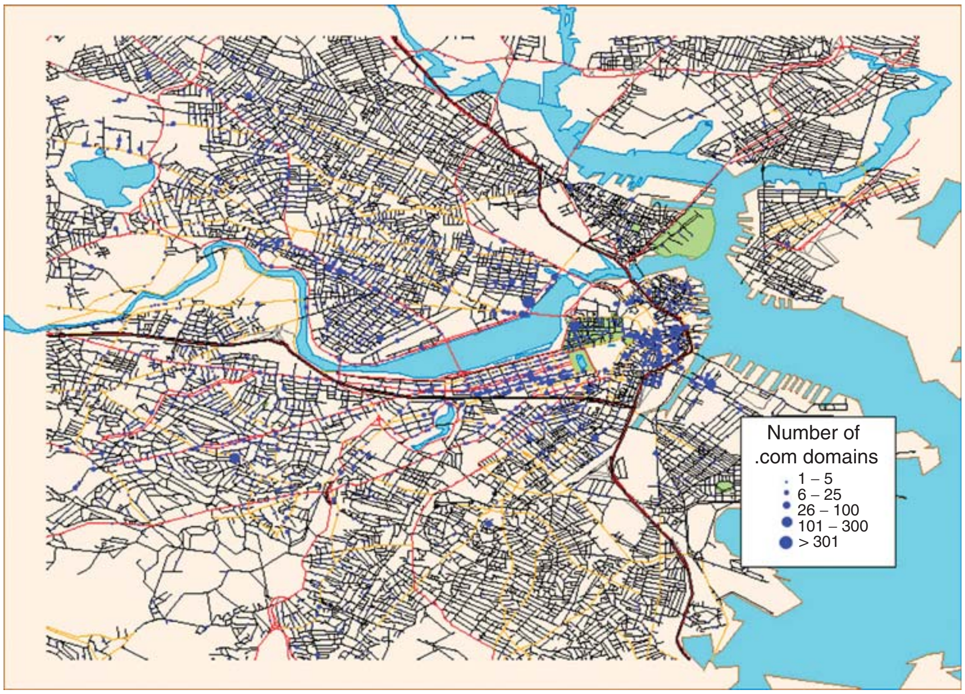

(a)

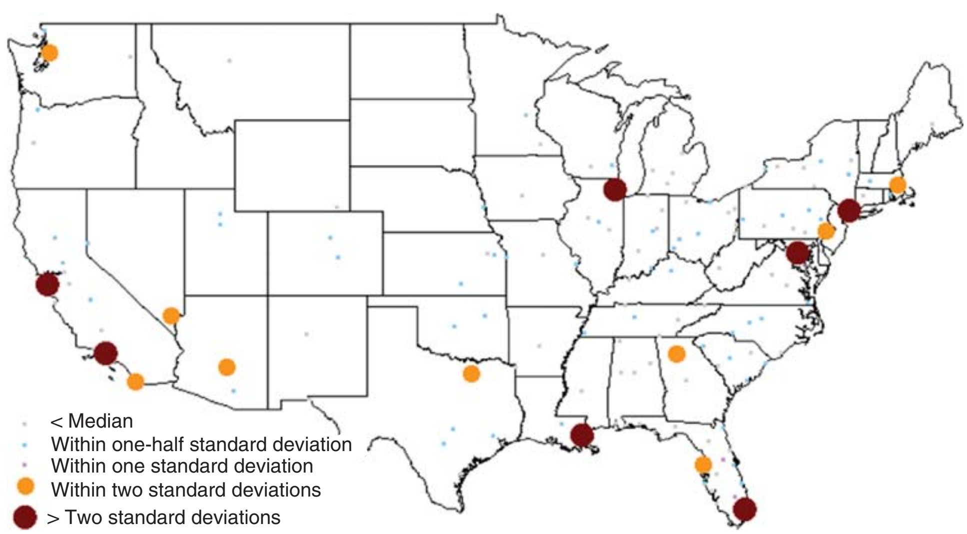

(b)

Figure 3 (a) The ownership pattern of .com domain names in the Boston metropolitan region (July 1998) and (b) the US distribution of adult websites at the metropolitan statistical area (MSA) level (July 2001). The darker and larger circles indicate the MSA are more than two standard deviations above the average number of websites per MSA. Source: Matthew Zook.

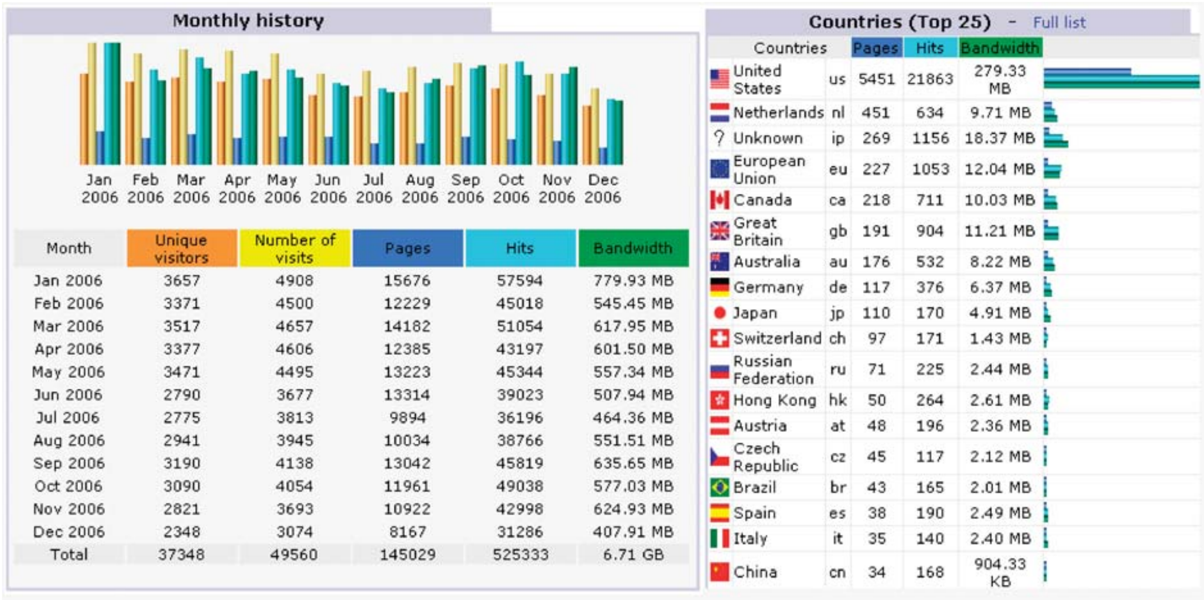

Month	Unique visitors	Number of visits	Pages	Hits	Bandwidth
Jan 2006	3657	4908	15676	57594	779.93 MB
Feb 2006	3371	4500	12229	45018	545.45 MB
Mar 2006	3517	4657	14182	51054	617.95 MB
Apr 2006	3377	4606	12385	43197	601.50 MB
May 2006	3471	4495	13223	45344	557.34 MB
Jun 2006	2790	3677	13314	39023	507.94 MB
Jul 2006	2775	3813	9894	36196	464.36 MB
Aug 2006	2941	3945	10034	38766	551.51 MB
Sep 2006	3190	4138	13042	45819	635.65 MB
Oct 2006	3090	4054	11961	49038	577.03 MB
Nov 2006	2821	3693	10922	42998	624.93 MB
Dec 2006	2348	3074	8167	31286	407.91 MB
Total	37348	49560	145029	525333	6.71 GB

Countries (Top 25) - Full list

Countries		Pages	Hits	Bandwidth
United States	us	5451	21863	279.33 MB
Netherlands	nl	451	634	9.71 MB
Unknown	ip	269	1156	18.37 MB
European Union	eu	227	1053	12.04 MB
Canada	ca	218	711	10.03 MB
Great Britain	gb	191	904	11.21 MB
Australia	au	176	532	8.22 MB
Germany	de	117	376	6.37 MB
Japan	jp	110	170	4.91 MB
Switzerland	ch	97	171	1.43 MB
Russian Federation	ru	71	225	2.44 MB
Hong Kong	hk	50	264	2.61 MB
Austria	at	48	196	2.36 MB
Czech Republic	cz	45	117	2.12 MB
Brazil	br	43	165	2.01 MB
Spain	es	38	190	2.49 MB
Italy	it	35	140	2.40 MB
China	cn	34	168	904.33 KB

Figure 4 An example of typical website usage statistics giving details on the number and nationality of visitors. Source: author screenshot.

undertaking at best. For example, much could be learned by analyzing the geography of users of sites such as Google, Amazon, or eBay but these data are closely guarded as commercially sensitive.

Measures of the number of Internet users at the national level is available but suffers from several drawbacks. First, the scale at which this data is organized prevents analysis at any subnational units such as city or region. Also problematic is that the data on users is generally constructed by combining national sample surveys employing different methodologies and definitions of Internet use. Finally, these data provide at best a measure of the potential demand for Internet resources but say nothing about the types of activity in which users are engaged. Although less comprehensive, researchers have generally conducted their own surveys of users to gain higher granularity and more specificity.

A middle ground between the rich albeit narrowly focused data from individual websites and the shallow yet comprehensive data from counts of users, are rankings that provide measures of the amount of user traffic to all websites. Although there are a number of ranking services, Alexa.com provides a long running and independent measure of the popularity of websites among users. Alexa.com's rankings are based on tracking the surfing activity of a panel of Internet users who have downloaded a web browser tool. Alexa.com relies upon this sample to judge the most popular websites on the Internet and even disaggregates the top websites per country (**Figure 5**). While this represents a reasonable approach, it is unclear whether these users are a representative sample of Internet users (particularly when disaggregated to the country level) and several avenues for bias have been identified by Alexa.com and others. Nonetheless, it provides one of the best publicly available means to compare the location of the users of websites at the country level.

Measuring Distance and Routes Across the Internet

Another important element of Internet based measurement is assessing distance between sites within the network. Given the topological structure of the Internet, physical distance between sites has little meaning. Instead, relative distances are measured using the journey time (i.e., latency) taken to transmit and receive data. Increasing latency implies increasing relative distance between two sites on the Internet. It is important to note, however, that there can be many different technical factors (e.g., types of hardware and network configurations which are shaped by ownership and institutional structures) that effect latency. An interesting point of analysis, both on the Internet and in the 'real' world, is to compute the relationship between distance on the ground and time–distance for different places. This relationship is not always linear because of barriers, lack of connectivity, and poor accessibility. Sometimes, the quickest sites to reach are not the closest physically while locations just down the street are completely offline. Analyzing the variable patterns in time accessibility can provide insight into underlying structural processes.

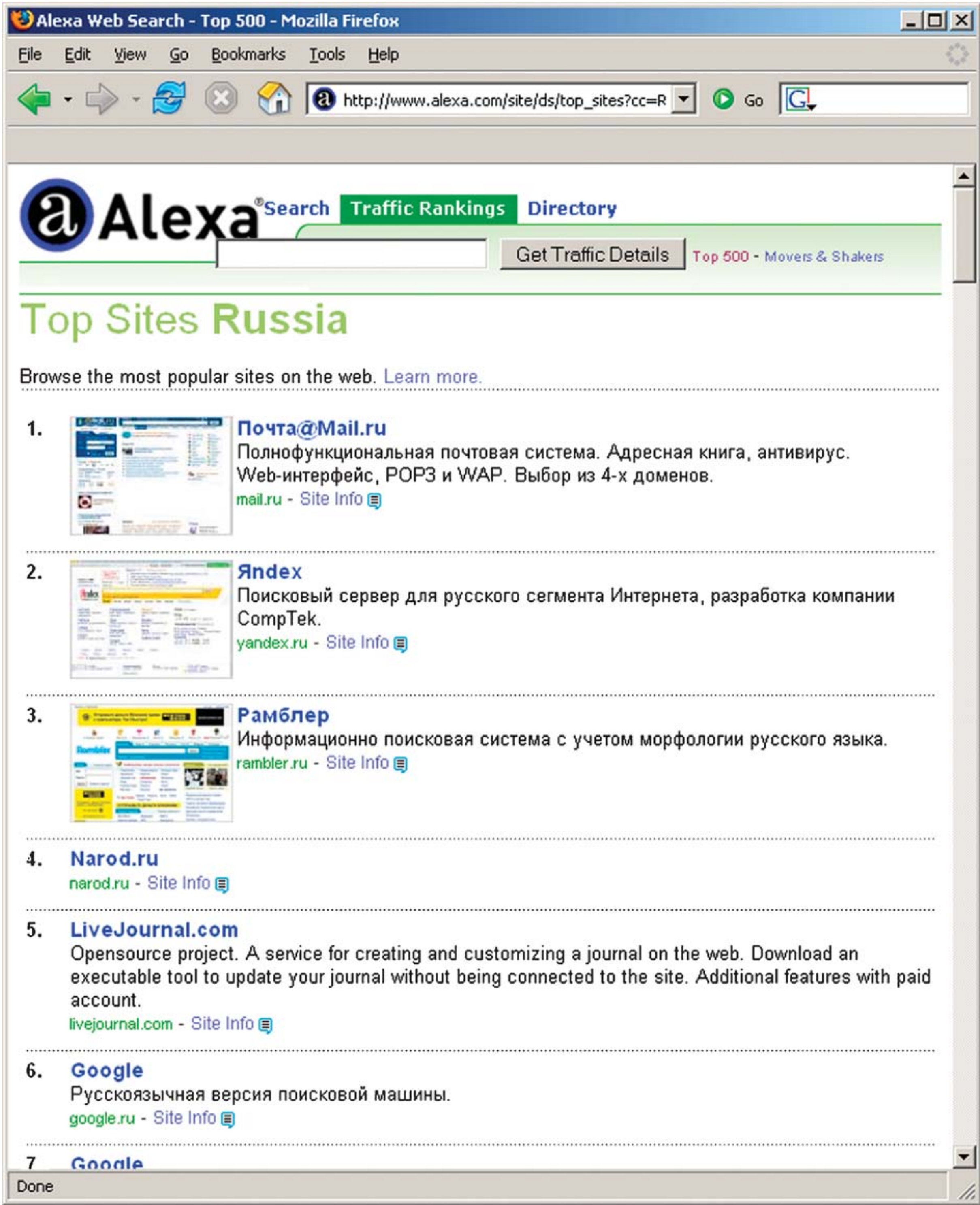

Figure 5 A sample of Alexa.com's data listing of the most popular websites in Russia. Source: author screenshot.

Distance Measurement

The simplest technique to measure latency uses the net work utility 'ping' which reports whether a particular site on the Internet is 'live' and accepting data. It works by sending out test data to a target site and listening for a response. It is useful for distance measurement because it reports the round trip time of data packets. For example, **Figure 6** shows the time (in milliseconds) each packet took to go from Lexington, KY to the web server at the University of Manchester and back again. The last line of

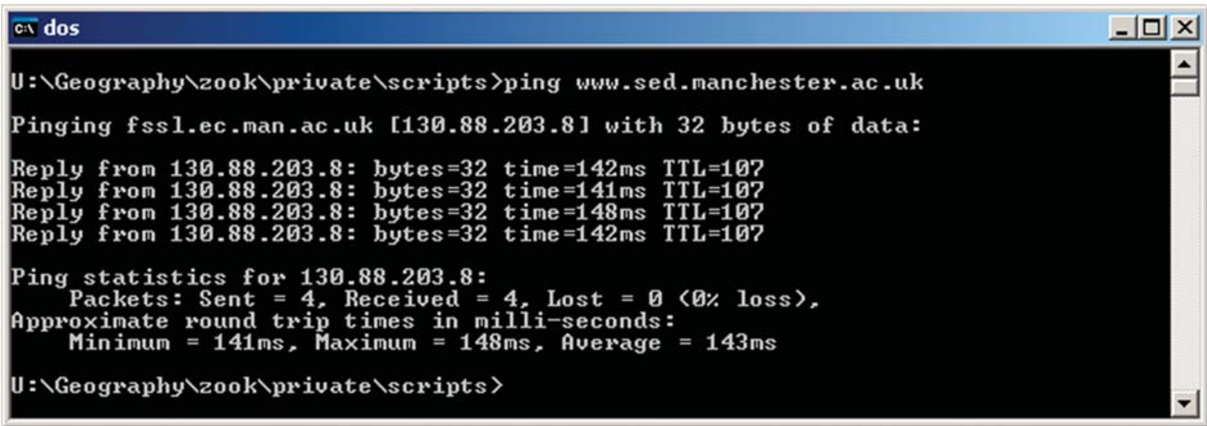

Figure 6 A ping query from the University of Kentucky to a web server at the University of Manchester that provides a measure of latency between origin and target. Source: author screenshot.

the output reports the overall statistics. According to this, the average 'distance' for this particular journey across the Internet as measured by latency was 143 ms. Latency distances are very susceptible to changes in conditions on the Internet and can provide ways of quantifying possible traffic congestion, much like measuring car speeds gives an indication of the level of road congestion.

There also are several ways that 'pinging' latency distances can be used to learn more about structure of the Internet. First, and most obviously, a sequence of pings to the same site at different time periods can be used to build up a comprehensive longitudinal profile of latency. Another useful extension is to take pings from different places on the Internet to triangulate in on a particular site. By triangulating from different points it is possible to get a sense of the relationship between latency and physical distance, assuming that the (approximate) geographic location of the origins and target are known. More importantly, combining the latency and physical distance can provide a measure of whether a place is readily accessible on the Internet or not.

Data Route Measurement

A much more sophisticated means of measuring distance through the Internet than ping is gained via the use of the utility 'traceroute' which reports details on the route data takes through the Internet. Traceroute is used in much the same way as ping but provides greater detail. It effectively maps out the path that data packets take between two sites on the Internet, showing all of the intermediate nodes traversed, along with an indication of the speed of travel for each segment of the journey. Although traceroute is primarily for network engineers debugging routing problems, it has also been used by researchers to expose the political economic structures of the Internet. It reveals the hidden complexity of data flows, showing how many nodes are involved, the seamless crossing of oceans and national borders, and the typically convoluted transfers through separate networks owned and operated by competing companies.

To illustrate how traceroute maps the Internet, it was used to chart the path from a PC at the University of Kentucky in Lexington, KY to a web server at the University of Manchester (**Figure 7**). The output looks rather cryptic at first sight, but it is in fact a kind of one dimensional map, with each node traversed listed on a separate line. It gives a complete linear route listing showing how data packets traveled through the Internet starting in Lexington, traveling via Atlanta, Washington, Amsterdam, London, Reading, and Warrington and ending at Manchester. The three time measurements in milliseconds – such as 142 ms, 149 ms, and 144 ms – are round trip times for that segment and give a useful indication of the speed of each link.

Each node traversed is identified by its domain name and/or IP address. Most nodes have unusually long domain names (e.g., atla.abilene.sox.net) which identify specialized routing computers at the core of the Internet not normally seen by users. With a little bit of decoding, the names of these routers can yield useful information, such as the type of node hardware, the bandwidth of the link, the name of the ISP that owns a node, and often a node's approximate location (usually at the city level). Many large network operators apply consistent naming conventions throughout their infrastructures, as in the machine names of the nodes of geant2.net. For example, the 'lon.uk' portion of the router name for segment 10 could reasonably be taken to mean London, UK (**Figure 7**).

Utility of Traceroute Measurement in Research

Just like ping, the usefulness of single traceroutes can be extended by running them from different sites to triangulate the Internet's structure. Web based traceroutes

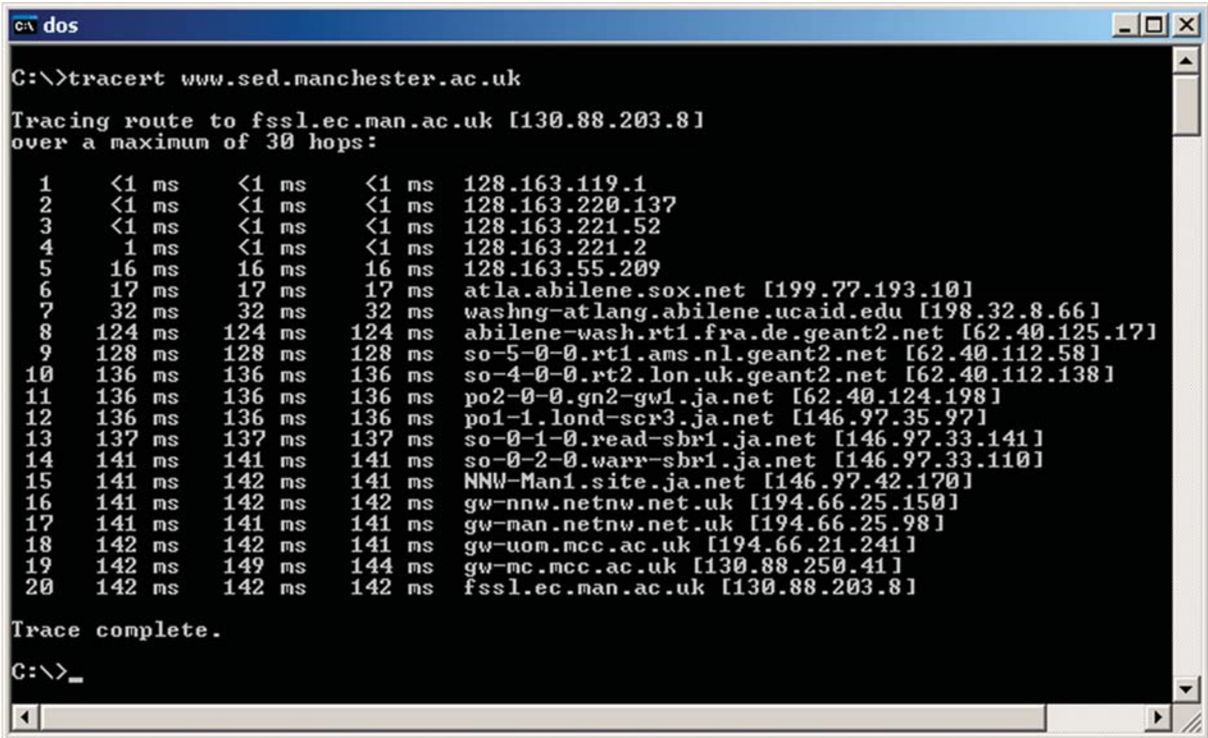

Figure 7 Traceroute measurement from the University of Kentucky to the University of Manchester. Source: author screenshot.

make it possible to run traces from many different starting points, including from different networks and in different countries. Web traceroute gateways are very useful for the active exploration of the Internet's topology from across the globe and illustrate the degree to which routes vary. Running multiple traceroutes to lots of different points across the Internet has also been used by researchers to gather large datasets on the topology of the core of Internet.

Data gathered by traceroute can provide evidence of the Internet's business 'logic' of following the cheapest paths rather than the shortest. Much international Internet traffic is still routed through the US as the cheapest means of transit between regions. This can result in quite anomalous looking, geographically circuitous routes being chosen.

Traceroute data can also be used for forensic analysis of the Internet's structure. For example, it is useful for deducing the approximate location of Internet hosts (such as websites) in terms of 'hardware geography'. The output shows the location and identity/owner of the 'upstream' network provider even if the final destination of the server is unclear. If the data travels into a certain city and does not leave it again, it is probable that the target is located there. Also, the 'upstream' network providers may keep logs for identifying a host that is of interest (this is of particular concern for law enforcement agencies in tracing the source of illegal activities).

Traceroute also reveals something of the hidden political economy of the Internet. The patterns of traffic routing show transit agreements and mutual peering relationships between competing companies. Details on these arrangements are often deemed commercially confidential, but are revealed by necessity in how and where the actual networks interconnect to share data. The routing of traffic reveals the structuring of business relationships in terms of who connects to who and the hierarchy of these connections (from periphery to center to periphery again). It can also show which telecommunications carriers dominate the transfer of traffic between particular countries and continents. These companies are likely to be influential in the structuring of global communications and tracerouting could provide an alternative way to quantify the extent of their power.

Last, the output from traceroute provides a useful way to assess the number of international borders crossed and determine which territories (i.e., separate legal jurisdictions) the data transits. The more 'points of contact' in the flow from origin to target, the more potential there is that Internet traffic could be intercepted and subjected to local regimes of monitoring, filtering, censorship, and data retention. In short, does an e mail message transit through a third party nation that has hostile intentions, for example, does an e mail to someone in Palestine transit through Israel? Particularly in regions of conflict, being able to identifying territories that are transited might be vitally important in terms of the reliability of communication.

Internet-Based Measurement of Other Socioeconomic Phenomena

A final aspect of Internet based measurement is its use as a means to gather data on phenomenon in physical space. One common method is the use of web based surveys. While these surveys are relatively easy to setup and conduct, sampling design and response rates are crucial to ensure statistically meaningful results. Researchers must ensure that their sample accurately represents the population they are studying rather than simply being a convenient way to conduct a survey. A related use of the Internet is to gather data that, particularly for datasets, are either prohibitively expensive to access or simply do not exist. A number of geographers have already lever aged the electronic format of the Internet to construct datasets (ranging from firms receiving venture capital to the location of corporate board members) in a relatively straightforward and, more importantly, cost effective manner. To date, most of these efforts are largely manual retrieval operations driven by the fact that the desired data exists in multiple locations and does not have a standardized format.

An extension of this method known as screen scraping allows access to data that would not otherwise be avail able or would be excessively time consuming to aggre gate into a dataset. Screen scraping is defined in this context as the use of a computer script (generally written by the researcher) to automatically collect the data out put of Internet based resources – most often web pages. These scripts are akin to user written macros to auto mate simple but repetitive tasks in a spreadsheet and are designed to generate automatic queries to web pages and collect and store the data received. While it is possible to conduct the queries by hand, screen scraping automates the process, greatly reducing research drudgery. It does, however, require programming and data cleaning skills as there is no 'off the shelf' program that is readily avail able. Instead, researchers must craft their scripts specif ically to the structure of the data which they are trying to collect.

A ready application of screen scraping is the well established technique of using directories as a method for identifying the locations of businesses. This method has been previously utilized with paper based data within historical urban geography. Although issues of the re liability and accuracy of directories are a concern, the key limiting factor in the use of this method is the amount of work needed to assemble the data. The technique of screen scraping can significantly simplify this task, making the collection of up to date, fine scale spatially referenced data (e.g., polluted sites, retail site locations) an easy task.

Conclusion

As the Internet grows in size, expands in scope, and becomes increasingly embedded as a banal background to everyday living, it becomes ever more important to understand the politics surrounding its production. Understanding the topological structures and geog raphies of the Internet, through quantitative network measurement using the techniques and tools described here, provides one of the most valuable avenues into network politics (e.g., the issue of net neutrality), allow ing researchers to gather information firsthand and crit ically question network operations directly. The medium of communication might be virtual, but the Internet is dependent on physical infrastructure and human labor, most of which is invisible to users. The computers are small in scale and are usually hidden from view in an onymous server rooms and secure, windowless buildings, while the cables are under floors, in ceilings, and in conduits buried under roads. The technical geography of Internet infrastructures are easily overlooked (just like for other essential utilities of water, electricity), but they are not naturally given. The geographical structure and op eration of networks that service modern living can be exposed through Internet based measurement.

Internet based measurement is likely to become easier as new and more powerful software tools for scanning the structure of Internet become available. Also, as com mercially provided search engine tools develop, they are increasingly providing new ways of surveying the infor mation structures of the web. Of course, researchers will continue to have to tread carefully the ethical bounda ries between critical fieldwork and potentially criminal hacking. At the same time, Internet based measurement is also getting harder and riskier to do. Many parts of the Internet are being designed and operated in a much more closed fashion. For example, some networks block ping and traceroutes as a security precaution against malicious scanning. Although this makes Internet based measure ment more difficult, it becomes evermore important to ensure that researchers remain capable of analyzing it independently. In particular, as the Internet intertwines with physical space it is essential that human geography follows this evolution in order to understand the in creasingly hybridized spaces inhabited in the twenty first century.

See also: Cyberspace/Cyberculture; Georeferencing, Geocoding; Internet, Economic Geography; Mapping, Cyberspace.

Further Reading

Barabasi, A. L. (2003). *Linked: The New Science of Networks*. New York: Perseus Books.
Branigan, S., Burch, H., Cheswick, B. and Wojcik, F. (2001). What can you do with traceroute? *Internet Computing* 5(5), 96.
Castells, M. (2001). *The Internet Galaxy*. Oxford: Oxford University Press.
Cukier, K. N. (1999). *Bandwidth Colonialism? The Implications of Internet Infrastructure on International E commerce*. Paper presented at INET'99 Conference San Jose, California. http://www.isoc.org/inet99/proceedings/1e/1e 2.htm (accessed Mar. 2008).
Dodge, M. and Kitchin, R. (2000). *Mapping Cyberspace*. London: Routledge.
Dodge, M. and Kitchin, R. (2006). Net: Geography fieldwork frequently asked questions. In Weiss, J., Nolan, J., Trifonas, P., Nincic, V. & Hunsinger, J. (eds.) *The International Handbook of Virtual Learning Environments*, pp 1143 1172. Netherlands: Springer.
Grubesic, T. H. (2002). Spatial dimensions of internet activity. *Telecommunications Policy* 26(7 8), 363 387.
Hayes, B. (1997). The infrastructure of the information infrastructure. *American Scientist* 85(30), 214 218.
Lakhina, A., Byers, J. W., Crovella, M. and Matta, I. (2002). *On the Geographic Location of Internet Resources*. Technical report 2002 15. Computer Science Department, Boston University. http://www.cs.bu.edu/techreports/pdf/2002 015 internet geography.pdf (accessed Mar. 2008).
Murnion, S. and Healey, R. G. (1998). Modeling distance decay effects in web server information flows. *Geographical Analysis* 30(4), 285 303.
Shiode, N. and Dodge, M. (1999). Visualising the spatial pattern of internet address space in the United Kingdom. In Gittings, B. (ed.) *Innovations in GIS 6: Integrating Information Infrastructure with GI Technology*, pp 105 118. London: Taylor & Francis.
Spring, N., Wetherall, D. and Anderson, T. (2004). Reverse engineering the internet. *ACM SIGCOMM Computer Communication Review* 34(1), 3 8.
Townsend, A. (2001). Network cities and the global structure of the internet. *American Behavioral Scientist* 44(10), 1697 1715.
Zook, M. A. (2000). The web of production: The economic geography of commercial internet content production in the United States. *Environment and Planning A* 32, 411 426.
Zook, M. A. (2005). *The Geography of the Internet Industry*. Oxford: Blackwell.

Relevant Websites

http://www.alexa.com
Alexa the Web Information Company: website popularity survey.
http://www.arin.net
American Registry for Internet Numbers (ARIN): whois query interface.
http://irrepressible.info
Amnesty International.
http://www.cia.gov
Central Intelligence Agency (CIA): World Factbook, Internet Users.
http://www.caida.org
Cooperative Association for Internet Data Analysis.
http://www.domaintools.com
Domain Tools.
http://www.hostip.info
HostIP IP Geo Coding.
http://www.iana.org
Internet Assigned Numbers Authority: IANA's list of all the country code top level domains.
http://www.isc.org
Internet hosts survey, conducted by Internet Systems Consortium/Network Wizards.
http://www.lumeta.com
Internet mapping project.
http://www.internetworldstats.com
Internet Usage World Stats.
http://www.netcraft.com
Netcraft web server survey.
http://opennet.net
OpenNet Initiative.
http://www.rsf.org
Reports without Borders.
http://www.telegeography.com
TeleGeography.
http://www.traceroute.org
Thomas Kernen's web traceroute list.
http://www.visualroute.com
Visual Route.
http://www.zooknic.com
Zooknic Internet Geography Project: Maps of Internet Users, Zooknic Internet Intelligence.

Interviews: In-Depth, Semi-Structured

R. Longhurst, University of Waikato, Hamilton, New Zealand

Glossary

Cold Calling This is calling on people (usually strangers) without any prior approach or contact being made in the first instance to ask if they would be prepared to be interviewed. Interviewers often get a high refusal rate using this approach.

Positionality It describes a person's position within the midst of complex, shifting and overlapping political, economic, cultural, social, sexual, gendered, and racialized processes. How people are positioned in relation to various contexts of power affects the way they understand the world. Some geographers have argued that reflecting carefully on and declaring one's positionality may lead to more sound research because it becomes apparent that all knowledge is partial.

Recruiting on Site It involves making contact with potential participants at a site or location that is relevant in some way to the research, for example, if someone was researching the culture of rugby they might attempt to recruit participants at a rugby club and/or field.

Reflexivity Critical and conscious introspection and analytical scrutiny of oneself as a researcher. Reflexivity involves examining one's own practice in order to gain new insights into research.

Situated Knowledge A notion that challenges the idea that knowledge is simply 'out there' waiting to be discovered. Instead, knowledge is 'embodied', that is, it is made by people who are situated within particular contexts.

Snowballing A technique used by researchers whereby one contact, or participant, is used to help recruit another, who in turn puts the researcher in touch with another. The number of participants soon increases rapidly or 'snowballs'.

Triangulation It involves using a variety of methods to collect data as opposed to relying on one single form of evidence as the basis for findings.

Introduction

In depth, semi structured interviews are verbal interchanges where one person, the interviewer, attempts to obtain information from another person by asking questions. Although the interviewer prepares a list of predetermined questions, in depth, semi structured interviews tend to unfold in a conversational manner. This offers participants the opportunity to explore issues they feel are significant. The interviewer does not keep a tight rein on the interview but instead allows the interviewee, through the use of open ended questions, to explore the subject in as much depth and from as many angles as they please. Interviews are usually conducted individually but can also be conducted in pairs or groups.

In depth, semi structured interviews are probably one of, if not the most commonly used qualitative method in the discipline of human geography. Geographers have used in depth, semi structured interviews to collect data on a huge array of subjects, including gay pride parades, merchant banking, and geography field trips to name just three. They have also reflected in great depth on the use of this method. At the level of research design geographers have considered issues such as who is interviewed, what questions they are asked, how to select and recruit participants, where to meet, how to record and transcribe interviews, and how to analyze data. Geographers have also considered many of the ethical, political, and epistemological issues surrounding the use of in depth, semi structured interviews.

Interviewing takes a number of different forms. At one end of a continuum are structured interviews in which the interviewer asks a list of prescribed questions (these questions are always asked in the same way and the same order). In structured interviews respondents are discouraged from deviating from the questions. At the other end of the continuum are unstructured interviews. In unstructured interviews respondents are encouraged to direct the conversation instead of the interviewer. In depth, semi structured interviews sit somewhere in the middle of this continuum. They involve directing respondents to some degree mainly by talking and listening to people but in ways that are self conscious, orderly, and partially structured. However, in depth, semi structured interviews are more than simply chats to people. It can take days, even weeks, to get prepared to conduct several interviews. Background reading, formulating questions, and contacting interviewees to set up appointments can be demanding and time consuming.

Conducting In-Depth, Semi-Structured Interviews

In depth, semi structured interviews can be used as a stand alone method or in conjunction with another method or several other methods (this is sometimes refereed to as 'triangulation'). The researcher needs to prepare questions, decide who to recruit to participate, contact potential participants, choose a location, carry out the interviews, and transcribe the data, all the while being cognizant of the ethical issues and power relations involved in the research. Perhaps it is not surprising, therefore, that interviewing requires a high level of interpersonal skill. Interviewers need to be able to put people at ease. They also need to be able to ask questions in a straightforward but interesting way. Listening care fully to responses, acting in a way that develops trust and that is supportive, but without swaying participants to a particular position requires considerable skill.

Before formulating interview questions researchers need to brief themselves fully in the subject area. Having done this it is important to create a list of themes or questions to ask participants. People who are very con fident at interviewing sometimes equip themselves with just a list of themes. It is imperative to have a logical sequence of questions or themes which have been pre pared in advance. Questions may be designed to elicit information that is 'factual', descriptive, provides background information, thoughtful and/or emotional. A combination of different types of questions can be effective depending on the research topic. Researchers tend to begin with a question that participants are likely to feel comfortable answering. More difficult, sensitive, or thought provoking questions are best left to the second half of the interview when participants are likely to be feeling more comfortable. It is useful for interviewers to be familiar with the questions so as not to break the flow of the interview by having to read them from a paper all the time.

The next step is to select participants. Usually people are chosen on the basis of their experience related to the research topic. While the aim of many quantitative methods is to choose a random or representative sample, to be 'objective', and to be able to replicate the data, the aim of qualitative methods is to choose respondents who will help the researcher make sense of people's experi ences. Thinking about who to interview often prompts researchers to reflect on their own identity and how this might shape the interactions during the interviews (this point is expanded in the section 'Ethical issues and power relations').

There are many strategies for recruiting participants for interviews. Sometimes researchers carry out a simple questionnaire survey to gather contextual information and include a request at the end asking respondents who are prepared to engage in a follow up interview to pro vide their contact details. Sometimes researchers adver tise for participants in local newspapers or on radio stations requesting interested parties to make contact. Other methods for recruiting participants for interviews are 'cold calling' and 'recruiting on site'. After inter viewing a participant it is common for researchers to ask the interviewee if they know others who might also be interested in being involved in the research. This tech nique is refereed to as 'snowballing'.

Having selected and recruited participants it is then necessary to decide where to conduct the interview. Obviously place can make a significant difference to the way in which an interview unfolds. It is important that both the interviewer and interviewee feel comfortable. For this reason a relatively neutral setting can be good. If the interview is to be taped then it is also important that the interview is conducted somewhere that is relatively quiet, otherwise transcribing the tape is likely to prove difficult, if not impossible. When interviewing business people or officials from organizations or institutions it is common to interview them on their own premises. This is often easier logistically than attempting to meet in another location. An advantage of meeting a respondent at their workplace is that it is often very useful for a researcher to be in the environment that they are studying (assuming 'work' is the focus of the interview). A potential disadvantage is that the interviewee may feel uncomfortable speaking freely especially if they want to say negative things about their particular business, organization, or institution. It is not always possible to carry out in depth, semi structured interviews in an ideal setting but so long as it is relatively neutral, comfortable, quiet, and accessible for both parties then it is likely to prove successful.

Semi structured, in depth interviews tend to last, on average, one h (but this can vary enormously). Busy people are reluctant to give up much more time than this and it is important not to overstay one's welcome. If more time is required, an interviewer can ask the respondent if it is possible to arrange another interview for a later date. Having a second interview rather than one long inter view can also provide the interviewer with an oppor tunity to begin analyzing the material from the first interview thereby returning to the second interview with more pertinent questions.

Some researchers take notes when they conduct in depth, semi structured interviews. Others audio–video record the interview which provides them with both an aural and a visual record. Most researchers, however, simply audio record the proceedings. The advantage of this over note taking is that the researcher can focus fully on the conversation as it unfolds instead of feeling pressure to try and record the participant's comments in a notebook. A disadvantage of audiotaping is that some participants may be uncomfortable being taped. Taping can affect dramatically the way in which people feel they are able to tell stories about themselves and others. Regardless of whether an interviewer decides to tape the conversation or not, it is a good idea after the interview to document the general tone of the conver sation, any important nonverbal cues, the key themes that emerged, and anything that might have been particularly interesting or surprising during the interview. This helps the researcher begin the task of analyzing the data.

If an interviewer does audiotape an interview it is helpful to transcribe the conversation as soon as possible after conducting it. Hearing the taped conversation when it is still foremost in one's mind makes transcription much easier. Sometimes researchers use a variety of different transcription codes, for example, the starts of overlap in talk can be marked by a double oblique //; pauses can be marked with a dot in parenthesis (.); nonverbal actions, gestures, and facial expressions can be noted in square brackets; and loud exclamations can be noted in bold type face.

Once transcripts have been completed the researcher begins the task of analyzing the data – determining the meaning in the information gathered in relation to the purpose of the research. Most researchers look for themes, commonalities, and patterns to try to make sense of the information. It can also be useful to think about the differences that emerge in the data, noting themes that run counter to, or contradict, those that are emerging. Some researchers employ computer based programs (such as NUD*ist) to assist with the analysis of tran scripts. Others use more 'low tech' options such as marking significant words and phrases with highlighter pens. If more questions are raised that need to be clarified in order to serve the purpose of the study, then another in depth, semi structured interview may be warranted to examine the issue more comprehensively.

Strengths and Weaknesses of In-Depth, Semi-Structured Interviews

Having addressed some of the 'nuts and bolts' of organ izing and conducting in depth, semi structured inter views I now discuss some of the strengths and weaknesses of this method. A strength of the method is that it is helpful for delving into and attempting to understand complex behaviors, experiences, and opinions. Other methods such as observation, closed questionnaires, census data, and structured interviews do not allow for much discovery or probing. In depth, semi structured interviews, however, offer interviewers and interviewees time and space to explore issues thoroughly. The method is also useful for collecting a range of opinions on a topic. Sometimes interviews reveal consensus but often they illustrate that people of different ages, ethnicities, gender, and sexualities have diverse opinions and experiences and do not agree on many issues.

Semi structured, in depth interviewing can prove particularly useful for investigating personal, sensitive, or confidential issues which informants might find dif ficult to disclose and discuss in a group interview or focus group. Such issues are also difficult if not impossible to raise in a questionnaire. It is possible that some people might feel under too much scrutiny in a semi structured, in depth interview and be less willing to open up than in the relaxed atmosphere of a group, but a skilled interviewer is usually able to overcome this by establishing a good rapport and trust with the respond ent. Most interviewees are forthcoming and eager to help out.

Another strength of in depth, semi structured inter views is that potentially the method can be used in a way that shows respect for people. Interviewers usually indicate that they value the interviewee's opinions (although complex ethical issues can arise when, for ex ample, an interviewee expresses highly discriminatory opinions in the interview). This does not necessarily imply that researchers empower their interviewees be cause the fact remains that most research benefits the researcher far more than it does his or her interviewees.

A weakness of the method is that it tends to be very time consuming. Formulating a schedule of questions and/or prompts, recruiting participants, organizing times and spaces in which to conduct interviews, and tran scribing and analyzing interviews all add up to many hours of labor.

Another weakness of the method is that interviewing can open up participants to feeling manipulated and betrayed (but also to feeling empowered). Interviewing relies upon intersubjectivity – people engaging with and positioning themselves in relation to each other. It de pends upon human relationships and these relationships can be rewarding but also disappointing. These strengths and weaknesses illustrate that there are numerous ethical issues and power relations involved in conducing in depth, semi structured interviews.

Ethical Issues and Power Relations

Confidentiality and anonymity are two important ethical issues that have been discussed by geographers and others in relation to in depth, semi structured issues. Participants need to be guaranteed that all the data collected will remain secure under lock or on a computer database accessible by password only; that information supplied will remain confidential and participants will remain anonymous (unless they desire otherwise); and that participants have the right to withdraw from the research at any time without explanation. Most re searchers offer to provide participants with a summary of the research results at the completion of the research project and to follow through on this pledge. Summaries can take the form of a hard copy or an electronic copy posted on a website.

Another ethical issue that can arise when conducting in depth, semi structured interviews is that an interviewee may express sexist, racist, or other discriminatory views. It is often argued that researchers ought to listen, pay attention, and be nonjudgmental, but sometimes being nonjudgmental might reproduce and even legitimize an interviewees' discrimination through complicity. Researchers need to think carefully about how to deal with such situations. There are no easy answers as to how to deal with these situations. Feminist, social, and cultural geographers in particular have made a useful contribution in this area. It cannot be assumed that interviewers will always be in a position of power over their interviewees. For example, a relatively young female graduate student might find herself in the position of interviewing a wealthy middle aged businessman who expresses stridently discriminatory opinions. The young graduate at some points during the interview might find it difficult to know how to respond.

Researchers also need to think carefully about how to interview in different cultural contexts. For example, 'First World' researchers investigating 'Third World' 'subjects' need to be highly sensitive to local codes of conduct. During interviews the interviewer and interviewee are likely to keep repositioning themselves with respect to the multiple roles and fractured identities that they take up. How people position themselves in relation to ethnicity, gender, sexuality, class, age, and so on during interviews has implications for the interview overall and for interpreting interview data. Over the past decade, taken for granted notions about who is an insider, who is an outsider, who has access to 'authentic' knowledge, and who can speak on a particular subject have been raised by geographers in relation to interviewing.

Geographers have also discussed and debated the concepts of 'positionality' and 'reflexivity' in relation to interviewing. Positionality describes a person's position within the midst of complex, shifting and overlapping political, economic, cultural, social, sexual, gendered, and racialized processes. How people are positioned within various contexts of power affects the way they understand the world. Reflexivity involves reflecting carefully on one's position as a researcher and declaring one's positionality in the work. The aim of this strategy is to make explicit the researcher's role in knowledge production. It becomes evident in such research that knowledge is 'situated' and partial rather than neutral and universal. Knowledge is not neutral but created by people who occupy particular subject positions. Some argue, however, that acknowledging one's positionality and being reflexive about the research is perhaps easier said than done. Questions have been raised about how possible it is to fully understand ourselves as researchers, the researched, and the research context. While researcher reflexivity may be a useful goal, examining our own practice in order to gain new insights into research practice and knowledge production may be difficult to achieve. Clearly there are many ethical issues and power relations that need to be teased out when conducting in depth, semi structured interviews.

In depth, semi structured interviews now appear to be orthodox in human geography. They are no longer slighted as 'soft methods' (in fact, there is a long tradition in human geography – throughout the twentieth century – of researchers 'talking to people'). There are, however, a number of other qualitative methods that although they share a great deal in common with in depth, semi structured interviews still seem to be reasonably poorly regarded in the discipline. Other qualitative methods such as focus groups, group interviews, ethnography, auto ethnography, autobiography, discourse analysis, and visual analysis still at times struggle for legitimacy in some fields such as economic and political geography. Overall, though, it seems that the discipline may be entering a period of more mature evaluation and reflection on qualitative methods and this has in part been spurred by geographers' extensive use of in depth, semi structured interviews.

Conclusion

In depth, semi structured interviews are a commonly used method in geographical research. The method involves talking with people – it relies on a reasonably informal interaction between interviewees and interviewers. In depth, semi structured interviews are useful for investigating complex behaviors, opinions and emotions, and for collecting a diversity of experiences. The method does not offer researchers a route to 'the truth' but it does offer a route to partial insights into what people do and think. Sometimes researchers rely upon in depth, semi structured interviews as a stand alone method; sometimes they are used in conjunction with other methods. In depth, semi structured interviews make a significant contribution to geographic research especially now that debates about meaning, identity, subjectivity, politics, knowledge, power, and representation are high on many geographers' agendas. Critically examining the construction of geographical knowledge as partial and situated has led to an interest in developing methodological strategies, such as in depth, semi structured interviews, that have potential for raising questions about researcher reflexivity, power relations, and 'different' ways of writing and knowing.

See also: Embodied Knowing; Emotional Knowing; Ethical Issues in Research; Feminist Methodologies; Focus Groups; Mixed and Multiple Methods; Performance, Research as; Quantitative Data; Situated Knowledge, Reflexivity; Subjectivity; Transcripts (Coding and Analysis); Triangulation.

Further Reading

Baxter, J. and Eyles, J. (1999). The utility of in depth interviews for studying the meaning of environmental risk. *Professional Geographer* 51, 307 320.

Bennett, K. (2002). Interviews and focus groups. In Shurmer Smith, P. (ed.) *Doing Cultural Geography*, pp 151 162. London: Sage.

Cloke, P., Cook, I., Crang, P., Goodwin, M., Painter, J. and Philo, C. (2004). *Practising Human Geography*, pp 148 159. London: Sage.

Crang, M. (2002). Qualitative methods: The new orthodoxy? *Progress in Human Geography* 26(5), 647 655.

Dunn, K. (2000). Interviewing. In Hay, I. (ed.) *Qualitative Research Methods in Human Geography* (1st edn.), pp 50 82. Melbourne, VIC: Oxford University Press.

Elwood, S. A. and Martin, D. G. (2000). Placing interviews: Location and scales of power in qualitative research. *The Professional Geographer* 52(4), 649 657.

Kitchin, R. and Tate, N. J. (2000). *Conducting Research into Human Geography*, pp 211 219. Edinburgh Gate: Pearson.

Longhurst, R. (2003). Semi structured interviews and focus groups. In Clifford, N. J. & Valentine, G. (eds.) *Key Methods in Geography*, pp 117 132. London: Sage.

McDowell, L. (1998). Elites in the city of London: Some methodological considerations. *Environment and Planning A* 30, 2133 2146.

Parr, H. (1998). The politics of methodology in 'post medical geography': Mental health research and the interview. *Health and Place* 4(4), 341 353.

Schoenberger, E. (1991). The corporate interview as a research method in economic geography. *The Professional Geographer* 43, 180 189.

Valentine, G. (1997). Tell me about …: Using interviews as a research methodology. In Flowerdew, R. & Martin, D. (eds.) *Methods in Human Geography: A Guide for Students Doing a Research Project*, pp 110 126. Edinburgh Gate: Addison Wesley Longman.

Valentine, G. (1999). Doing household research: Interviewing couples together and apart. *Area* 31, 67 74.

Relevant Websites

http://www.geog.le.ac.uk
Exploring online research methods in a virtual training environment: online interviews (University of Leicester).

http://www.managementhelp.org
General guidelines for conducing interviews, Management Library.

http://www.msu.edu
Interviewing in qualitative research, Michigan State University.

http://www.intute.ac.uk
Intute: social sciences, qualitative methods, web resources for education and research.

http://www.resolutions.co.nz
Research solutions: in depth interviews, innovative marketing research and strategy.

http://www.scu.edu
Resource papers in action research: convergent interviewing, a technique for qualitative data collection, Southern Cross University, Australia.

http://aix1.uottawa.ca
Tips for research interviews, University of Ottawa.

http://www.wpi.edu
WPI Worcester Polytechnic Institute Global Perspective Program Handbook for IPQ (Interactive Qualifying Project) advisors and students. Chapter 11: introduction to interviewing techniques.

Investment Promotion

N. A. Phelps, University College London, London, UK

Introduction

International economic integration – the international flows of money, goods, productive investments, labor – has been neither natural nor inevitable. It has been orchestrated and contested by a wide range of actors including nation states, international organizations (such as the World Bank, International Monetary Fund, the United Nations, etc.), multinational enterprises (MNEs), nongovernmental organizations, and a whole host of what might be termed intermediary professions and organizations. The investment promotion community might be considered as one increasingly important set of intermediaries in the process of 'globalization'. Indeed, the highly diverse set of activities, individuals, organizations, and associations that constitute the investment promotion community might be considered in the vanguard of those promoting international economic integration.

While the promotion of inward investment may have a long history dating back to the sorts of place promotion apparent in now developed countries in 1800s, the rise of the investment promotion community ostensibly has been a post war phenomenon aligned with the major growth in foreign direct investment (FDI) flows since the 1940s. According to the major reference source on the subject:

> investment promotion efforts are the result of competition by governments in the effort to attract foreign direct investment. This competition is not entirely new; what is new is its aggressiveness and intensity.
>
> (Wells and Wint, 1990: 2)

During the 1950s, 1960s, and 1970s many nations – especially developing countries – actually became more restrictive of FDI flows into their territories.

As flows of investment in the form of loans and aid from international organizations have tended to decline in recent decades, private investment flows in the form of FDI have assumed a greater importance and one which has seen the international policy environment surrounding the regulation of FDI liberalize considerably. Since the 1980s, changes in national FDI policy environments overwhelmingly have been in the direction of liberalization, as governments have come to perceive MNEs and their associated FDI as bearers of scarce capital, technology, and knowledge essential to economic development. For example, between 1991 and 1998, a total of 895 regulatory changes were made by nations of which 848 (95%) were more favorable and just 47 (5%) less favorable to FDI.

The intensified competition for FDI has also been exacerbated by changes in the dynamics of competition among MNEs. Not only have the largest MNEs become more familiar with overseas operating environments they have also tended to accelerate decisions to invest in rather than export to overseas markets and to diffuse newer products and technology to these overseas locations. These sorts of practices have become both more generalized and more pronounced as 'time based competition' has resulted in increased actual or potential mobility of productive investments.

It is hardly surprising therefore that investment promotion has emerged as something of a growth industry, though there are very few studies of investment promotion activity let alone estimates of its effectiveness or global significance in terms of employment. This article therefore attempts to sketch some rather vague boundaries around the investment promotion community situating it in a wider context of activities geared toward facilitating FDI flows. It will draw attention to the diversity of actors and practices involved in intermediating the global economy and to the uneven geography of these practices of intermediation.

The Broad Contours of the Investment Promotion Community

As a narrowly defined set of practices investment promotion sits alongside a whole host of other practices that are concerned directly and indirectly with the facilitation of FDI. Wells and Wint exclude practices such as the granting of incentives, the screening of investors, and negotiating with investors (even though these are often conducted within the same organization) let alone the other related practices to which we refer below. Narrowly defined then investment promotion consists of:

> advertising, direct mailing, investment and seminars, investment missions, participation in trade shows and exhibitions, distribution of itineraries for visits of prospective investors, matching prospective investors with local partners, acquiring permits and approvals from various government departments, preparing project proposals, conducting feasibility studies, and providing

> services to the investor after projects have become operational.
>
> (Wells and Wint, 1990: 8)

In some important respects this narrow definition ig nores the intimate connections between promotion and a range of other activities involved with promoting or fa cilitating FDI, and it is this wider sphere of activities which we wish also to examine. So, in developed coun tries, for example, promotion activities defined in the narrow terms above almost always take place in a context in which financial incentives are being offered. Indeed, evidence from the UK indicates that the activities of those involved in promotion have a bearing upon nego tiations with investors (including over grant aid) even when these are conducted in separate organizations. There is a host of activities that, outside of such a narrow definition, involve an element of promotion or with which promotional activities narrowly defined articulate very closely in the facilitation of FDI.

Conceived broadly, the investment promotion com munity is actually composed of an incredibly diverse set of individuals, organizations, and associations whose status in society and economy is rather difficult to define. At the most general level of abstraction, and in light of the criticisms of this notion of global class formation at the global scale, it is hard to conceive of the investment promotion community as part of a 'transnational capit alist class'. At a lesser level of abstraction, neither should, those involved in investment promotion, be considered as a coherent and defined profession – there are no pro fessional credentials that are associated with this activity and indeed one of its defining features is the blurring of private and public credentials. They may instead be most accurately conceived of as cadres or intermediaries whose argument and persuasion shape the material and dis cursive terms on which processes of international eco nomic integration proceed.

Those involved range from individuals acting as consultants to companies and governments, to small companies that facilitate investment into a single host country, to vast multinational companies (such as the large accountancy firms) one element of whose work is concerned with consultancy and investment promotion. Interestingly some of the oldest and largest companies concerned purposely with investment promotion have been formed as specialized units within major multi national companies to evaluate new factory or office lo cations (for example, IBM, Plant Location International). Others are more recent spinout companies from large multinational accounting and consulting firms (for ex ample, OCO Consulting).

The investment promotion community includes these private sector actors but also is composed significantly of state or quasi state organizations charged with promoting investment into their respective national and subnational territories – often referred to as investment promotion agencies (IPAs). There may be good reasons – as Paul Krugman notes – why in economic theory countries do not and ought not to compete like businesses. However, the reality of economic policy and practice is that,

> the new competitive foreign investment environment has prompted analogies between competition among gov ernments for foreign investment and competition among firms for market share.
>
> (Wells and Wint, 2000: 4)

The investment promotion community, including IPAs, is at the heart of this process. As important prac titioner literature elaborates,

> with very few exceptions, countries serious about in vestment promotion depend heavily on the IPA to market and sell the location as a host to investment' and 'similar to the sales and marketing arms of companies they try to attract, IPAs are the cornerstone of a lo cation's effort to attract "clients" …
>
> (MIGA 2000b: 2)

These state and quasi state bodies alone vary con siderably in terms of their size, recruitment policies, funding, and other resources. Their detailed remit and their status within or relationship to their respective state apparatus also varies considerably. Some of the most effective IPAs have a strong advocacy position within their respective governments and have been able to shape national economic priorities beyond the realm of investment promotion *per se*. Singapore's Economic Development Board (EDB) would be an example of an agency which has moved well beyond simple attraction of FDI to articulate with other government bodies in order to, for example, stimulate the use of local suppliers by FDI, stimulate outward FDI by Singaporean com panies, develop overseas industrial and technology parks, and recruit skilled labor from overseas to Singapore. Over the last 30 years or so the number, size, and scope of these IPAs alone has grown steadily. According to one recent estimate, there were 410 national and subnational IPAs worldwide. The growth of these IPAs is symp tomatic of the widespread absorption of such competitive signals into the workings of national gov ernments. There is an assumption here that investment promotion activities are cost effective in generating FDI, and indeed studies find that investment promotion yields positive returns to investment; however, there is also something of a 'development paradox' here to which we return below.

Several of the major international organizations such as the World Bank (in the form of its Multilateral Investment Guarantee Agency, MIGA; and Foreign Investment

Advisory Service, FIAS), the Organisation for Economic Co operation and Development (OECD), and the United Nations (in the form of its United Nations Industrial Development Organization and United Nations Confer ence on Trade and Development) have an active interest and generate earnings from their investment promotion consulting and services. In fact, there is an element of competition among these organizations to provide services to governments wishing to promote their territories to overseas investors while each also tends to have a differ ential presence in different parts of the world. Arms of national governments such as United States Agency for International Development (USAID) and Australian Agency for International Development (AusAID) also are involved with technical assistance to particular subsets of host countries, with involvement seemingly closely linked to geopolitical concerns and domestic business and com mercial interests overseas.

Moreover, the actual technical basis of those involved in investment promotion can vary from marketing, man agement consulting, law, architecture, land use, planning and economic development, engineering, environmental consulting, and political risk analysis to name the main closely related activities. Each of these embodies a distinct set of skills, though not all are practiced under some form of professional accreditation. Marketing and management consulting are sufficiently broad to encompass practices which involve a range of skills and can be predicated on proprietary 'tools' for location appraisals, benchmarking of organizational performance, etc.

Finally, one indication of the growing maturity of the investment promotion industry is the sheer volume of information and materials that relate to all of the activities described immediately above. International organizations such as the World Bank and United Na tions produce annual and occasional reports on the likes of FDI trends and have themselves done much to en courage the competition for FDI through the likes of FDI competitiveness indexes. There are also regular country by country FDI policy reviews produced by the likes of the FIAS of the World Bank and the OECD. The World Bank, the Asian and African Development Banks, and major multinational consulting firms produce regular economic and investment profiles and assess ments and ratings of individual countries. Fledgling web and e mail based networks related to investment monitoring and information exchange have appeared over the past decade. *FDI Xchange* and *IPAnet* are run by the MIGA of the World Bank. Specialist magazines covering corporate location have existed for some time previous to this in major nations like the United States. Finally, associations – such as the World Association of Investment Promotion Agencies – catering to the in vestment promotion industry have also begun to thrive in the last decade.

Some Geographical Dimensions of Investment Promotion

Although we might style the investment promotion community as the promoters of the global economy, it is as well to remember that their role as intermediaries, as well as being temporally uneven, as alluded to above, is also geographically quite differentiated.

First, the various actors we have enumerated above and the various activities in which they are involved have an uneven presence across different parts of the world. For example, investment promotion activities of inter national organizations such as the World Bank, OECD, and United Nations tend to be unevenly focused. Just at present, the World Bank and United Nations are most preoccupied with providing technical assistance with investment promotion to African nations in an attempt to assist their integration into the international economy. The OECD on the other hand has taken a lead in the transition economies of East and Central Europe. In similar vein, the largest multinational accountancy and law firms invariably play a role in facilitating investment in most national settings, although they may well be overshadowed in many of these settings by domestic in vestment promotion companies. Even the large specialist investment location consulting companies that exist (for example, IBM Plant Location International, Buck Con sultants International, and OCO Consulting) tend to have particular national markets and associated networks of clients in which they are embedded.

Second, while Klak and Myers identify a "remarkably rapid unification of policy around a neoliberal develop ment model across the third world," any such general or global investment promotional 'models' are refracted through and actively modified in the process of transla tion to the specifics of local environments. Undoubtedly, stories of the successful capture of an FDI project and the organization and ethos of particular IPAs circulate internationally as models of best practice. Two model IPAs whose practices are imitated widely are Singapore's EDB and the Industrial Development Agency. However, even the literature of the investment promotion com munity is self critical enough to realize that such models need to be adapted to the specifics of local FDI environments.

Third, and following on from this, any convergence on neoliberal policy orthodoxies perceived or desired by commentators is predicated on a rather narrow view of the market for FDI. The investment decisions made by MNEs (the demand side of the market for FDI) are re markably diverse and involve considerations other than cost minimization. They vary according not only to the division of labor associated with a particular product or service (where distinctions are often drawn between fundamental research, development, and design let alone other separate functions such as sales and marketing and the various stages of production) but also according to the likes of product cycle considerations and corporate strategy and culture. Moreover, this differentiated demand for investment locations from MNEs is met and, it stands to reason, sustained by variety in the supply of potential investment locations offered by different nation states in terms of their differing factor endowments, business systems, and their different national modes of social regulation. Here, intermediaries play an important role in shaping opinions regarding particular investment locations as literature from the World Bank's MIGA *Investment Promotion Toolkit* makes clear; because

> no investor has access to all the information needed to make a decision, they fill in the gaps by making assumptions, extrapolating based on known information and relying on the advice and opinions of others (including the media, consultants, and competitors).
>
> (MIGA, 2000a: 3)

Fourth, according to one recent survey, the most important organizations used as sources of information in the international site selection processes of investing companies were in order: investment or merchant banks; the big accountancy firms; national IPAs and economic development agencies; specialized site selection firms; and general management consulting firms. Immediately we can draw an important observation from this: that it is the most internationalized of private sector actors, the investors themselves, as well as the largest merchant banks and accountancy firms that are the prime sources of ideas and practices relating to the competition for that investment. This reflects a more general pattern regarding the international spread of business regulation, in which private sector actors are the prime movers. Most investments by most MNEs actually take place without significant recourse to the investment promotion community as we have tended to look at it. Nevertheless, there is an important geographical paradox in the role and development of the investment promotion community that we should be aware of.

Finally, in global terms there exists a paradox as far as investment promotion, FDI, and development are concerned. Contrary to popular perceptions the vast majority of MNEs are actually rather small in size and often extremely risk averse in their investment decisions. The fact that the majority of the world's FDI flows continue to circulate among the most developed nations is partly a testament to this. This inertia in FDI flows stems from the fact that these countries represent the major markets and sources of knowledge, and it is these countries with which MNEs are already familiar and in which there are relatively complete and secure business and commercial environments. It is these environments in which one would imagine that the need for intermediaries in the investment process is less. Yet, if anything, it is in these countries that the investment promotion community broadly defined is at its most developed!

It seems certain that intermediaries of various kinds are playing a role in facilitating FDI into major but complex emerging markets such as China, although this role may be rather muted in a setting in which investment promotion is firmly entrenched within an extensive state apparatus. Africa has long been largely excluded from the processes of international integration embodied in FDI flows and if anything has become further isolated. It is precisely the sort of risky and uncertain FDI environment in which one might expect to find the investment promotion community playing a significant role. However, one recent survey found that only 7% of FDI inquiries into a selection of African countries came from intermediaries. Nevertheless, intermediaries can have a much wider impact in contexts such as Africa where they can be very influential in shaping investors' impressions of host country business environments. Here, then, a major part of the work of IPAs such as Kenya's Investment Authority is to convince overseas investors of the viability of the operating environment in general terms.

Instead, it is the developed nations which benefit from the largest flows of FDI in which the greatest resources devoted to investment promotion – primarily from taxpayers into state and quasi state IPAs. These ought to represent relatively safe, familiar, and core investment environments to MNEs such that the need for intermediaries is virtually nonexistent. It is IPAs in these countries that have the greatest resources to compete for FDI and the greatest human resources and organizational reach to promote FDI. In the likes of the UK it has become apparent that a 'grants culture' has emerged in which FDI is facilitated by significant and repeated grant aid and in which the investment promotion community is implicated. And so the chicken and egg like questions remain: Does investment promotion stimulate FDI or does FDI support the investment promotion industry? Can investment promotion play a role in reducing economic development disparities between developed and developing countries in a liberalized investment climate or is it symptomatic of significant geographical inequalities?

Conclusions

The investment promotion community broadly defined includes a diversity of practices and actors that have intermediated or facilitated international economic integration via FDI. The boundaries of this community are hard to define and tend to cut across some quite im portant professional distinctions. To date, the commu nity's quantitative significance in stimulating FDI flows may be rather minor. However, the qualitative role played by the constituents of this community in trans mitting myriad signals regarding FDI may be much more significant. It is this qualitative role which positions the community as part of cadres concerned with articulating national and international economies with processes of international economic integration.

While private and public sector organizations with an international reach are part of the investment promotion community and while there are some reasonably com mon aspects to investment promotion, the intermediation of FDI flows is geographically uneven. The comparative youth of this industry and its implication with recent rapid rises in the level of international economic inte gration via FDI flows means that the role of investment promotion remains opaque. Even in those developed countries where the investment promotion community is most developed there is a genuine lack of knowledge regarding the role it plays and the attendant ramifications for policy regarding incentives and rules based com petition for FDI. In addition, it seems clear that such intermediaries will play an important though as yet un studied role in the integration of the major emerging developing country markets of China and India into the international economy.

See also: Foreign Direct Investment; Global Production Networks; Trade, International.

Further Reading

Braithwaite, J. and Drahos, J. (2000). *Global business regulation*. Cambridge: Cambridge University Press.

Klak, T. and Myers, G. (1997). The discursive tactics of neoliberal development in small third world countries. *Geoforum* 28, 133 149.

MacKinnon, D. and Phelps, N. A. (2001). Devolution and the territorial politics of foreign direct investment. *Political Geography* 20(3), 353 379.

MIGA (2000). *Investment promotion toolkit: Strengthening the location's image*. Washington, DC: MIGA.

MIGA (2000). *Investment promotion toolkit: Developing an investment promotion agency*. Washington, DC: MIGA.

MIGA (2002). *Foreign direct investment survey*. Washington, DC: Multilateral Investment Guarantee Agency.

Morrisset, J. and Andrews Johnson, K (2004). *Occasional paper 16: The effectiveness of promotion agencies at attracting foreign direct investment*. Washington, DC: Foreign Investment Advisory service.

NAO (2003). *Regional grants in England: Report by the comptroller and auditor general*. London: National Audit Office, HC 702. Stationery Office.

Phelps, N. A. (2006). Capital's search for order: Foreign direct investment, models and hybrid models of social order in Southeast Asia. In Yeung, H. W. C. (ed.) *Handbook of research on Asian business research*, pp 250 265. Cheltenham: Edward Elgar.

Phelps, N. A., Power, M. and Wanjiru, R. (2007). Learning to compete: The investment promotion community and the spread of neoliberalism. In England, K. & Ward, K. (eds.) *Neo liberalization: States, networks, peoples*, pp 83 109. Oxford: Blackwell.

Phelps, N. A. and Raines, P. (eds.) (2003). *The new competition for inward investment: Companies, institutions and territorial development*. Cheltenham: Edward Elgar.

Phelps, N. A. and Wood, A. (2006). Lost in translation? Local interests, global actors and the multi scalar dynamics of inward investment. *Journal of Economic Geography* 6, 493 515.

Robinson, W. I. and Harris, J. (2000). Towards a global ruling class? Globalization and the transnational capitalist class. *Science and Society* 64, 11 54.

Safarian, A. E. (1999). Host country policies towards inward foreign direct investment in the 1950s and 1990s. *Transnational Corporations* 8, 93 114.

Sklair, L. (2000). *The transnational capitalist class*. Oxford: Blackwell.

UNCTAD (2002). The world of investment promotion at a glance: A survey of investment promotion practices. Advisory Study No. 17.

United Nations (1999). *World investment report: Foreign direct investment and the challenge of development*. Geneva: United Nations.

UNIDO (2003). *Africa foreign investor survey 2003*. Vienna: United Nations Industrial Development Organization.

Vernon, R. (1979). The product cycle hypothesis in a new international environment. *Oxford Bulletin of Economics and Statistics* 41, 255 268.

Wells, L. T. and Wint, A. G. (2000). *Occasional Paper 13: Marketing a country: Promotion as a tool for attracting foreign investment (revised edition)*. Washington, DC: Foreign Investment Advisory Service.

Relevant Websites

http://www.edb.gov.sg
Economic Development Board, Singapore.

http://www.ibm.com
IBM Business Consulting Services: Plant Location International (IBM PLI).

http://www.idaireland.com
Industrial Development Agency, Ireland.

http://www.investmentkenya.com
Kenya Investment Authority, Invest in Kenya.

http://www.waipa.org
WAIPA, World Association of Investment Promotion Agencies.

http://www.miga.org
World Bank Group, Multilateral Investment Guarantee Agency.

Irredentism

G. W. White, Frostburg State University, Frostburg, MD, USA

Glossary

Irredenta Land to which an irredentist claim is made.

Irredentism The desire to annex adjacent land to a state because it historically belonged to the potential annexing state or because it contains people who are culturally akin to those in the potential annexing state.

Introduction

The term 'irredentism' originates from the Italian phrases *terra irredenta* ('unredeemed land') and *Italia Irredenta* ('unredeemed Italy'). Matteo R. Imbriani coined the latter term in 1877 as a reference to Austrian possessed Trieste. Earlier, in 1861, the modern Italian nation state was formed by the welding together of numerous principalities on the Italian peninsula. Seeking to incorporate the entire peninsula into their new Italian nation state, Italian nationalists sought to expand their new state in an Italian unification project known as *Risorgimento* ('Resurgence') (**Figure 1**). Venice (Veneto) was obtained in 1866, and Rome and the Papal States were annexed in 1870. However, several territories still eluded the unification movement in the 1870s, causing great frustration and the emergence of the concept of irredentism. In addition to Trieste, Italian nationalists desired Trentino (Trento) and Istria, both controlled by the Austrian Habsburgs. Some desired Nice (Nizza; Nizzardo) and Savoy (Savoia), both of which had been traded to France in 1860 as part of a bargain that gained Napoleon III's support for driving the Austrian Habsburgs out of the Italian peninsula. The famous Italian nationalist, Giuseppe Garibaldi was outraged at the bargain because he wanted his hometown and province of Nice to become part of the Italian nation state. Though Savoy was relinquished, the House of Savoy provided the modern Italian nation state with all of its modern kings. Other Italian nationalists expressed desire for Corsica, Malta, Rijeka (Fiume), the Dalmatian coast, the Ionian Islands and the Swiss canton of Ticino. Though Italy had been a member of the Triple Alliance with Germany and Austria Hungary, strong Italian irredentist feelings led the Italian government to side with the Triple Entente (France, the United Kingdom, and Russia) during World War I for the promise of Trieste, South Tyrol, and Istria. Allied victory in the war allowed for Italy's annexation of these territories.

Though the term irredentism stems specifically from the Italian unification movement, it is more broadly the product of modern nationalism and its corollary the nation state idea, the belief that every nation should have its own sovereign state. Modern nationalism began with the French Revolution of 1789 and spread until nationalist revolutions erupted throughout Europe in 1848. The French Revolution was inspired by the ideas of the Enlightenment and its emphasis on the individual and civic responsibility. Later, the emergence of Romanticism and its emphasis on the collective group defined by common language influenced nationalist thinking. Thus, replacing dynastic imperial governments with popular governments as had happened in France was not sufficient; Romantic nationalists also desired to unite the speakers of a language into a single nation state even if it required the merging together of numerous dynastic territories. The Italian and German unification movements are two of the earliest and most significant examples. However, the conservative dynastic imperialist forces crushed the nationalist revolutions in 1848 and remained an obstacle to national unification projects through World War I. Frustrated because they were unable to quickly create nation states during revolution, nationalists were forced to slowly build their nation states piece by piece. The term and concept of irredentism emerged in the latter part of the nineteenth century after new nation states emerged but their territorial extents were incomplete. Today, many irredentist projects are labeled with the adjective 'Greater', as in Greater Serbia and Greater Croatia.

Evolving and Varying Definitions of Irredentism

Today, irredentism generally refers to the desire to have a state annex any territories of neighboring states on historical grounds or because such territories contain inhabitants of similar cultural/ethnic characteristics as the people within the potential annexing state. Often irredentism is seen as the attempt to bring all members of a nation and the territories they inhabit into a single state, commonly a nation state. Irredentism is a straightforward concept but it is less clear which of the many conflicts around the world are caused by it. Certainly, the numerous spatial disjunctions between peoples and states make irredentism a potential source of conflict in many places. Yet, resources, ideology, and border disputes also cause wars. In other cases, an ethnic minority may engage in armed resistance against a state but the goal may be secession, not irredentism. To distinguish between conflicts caused by irredentism and those caused by other

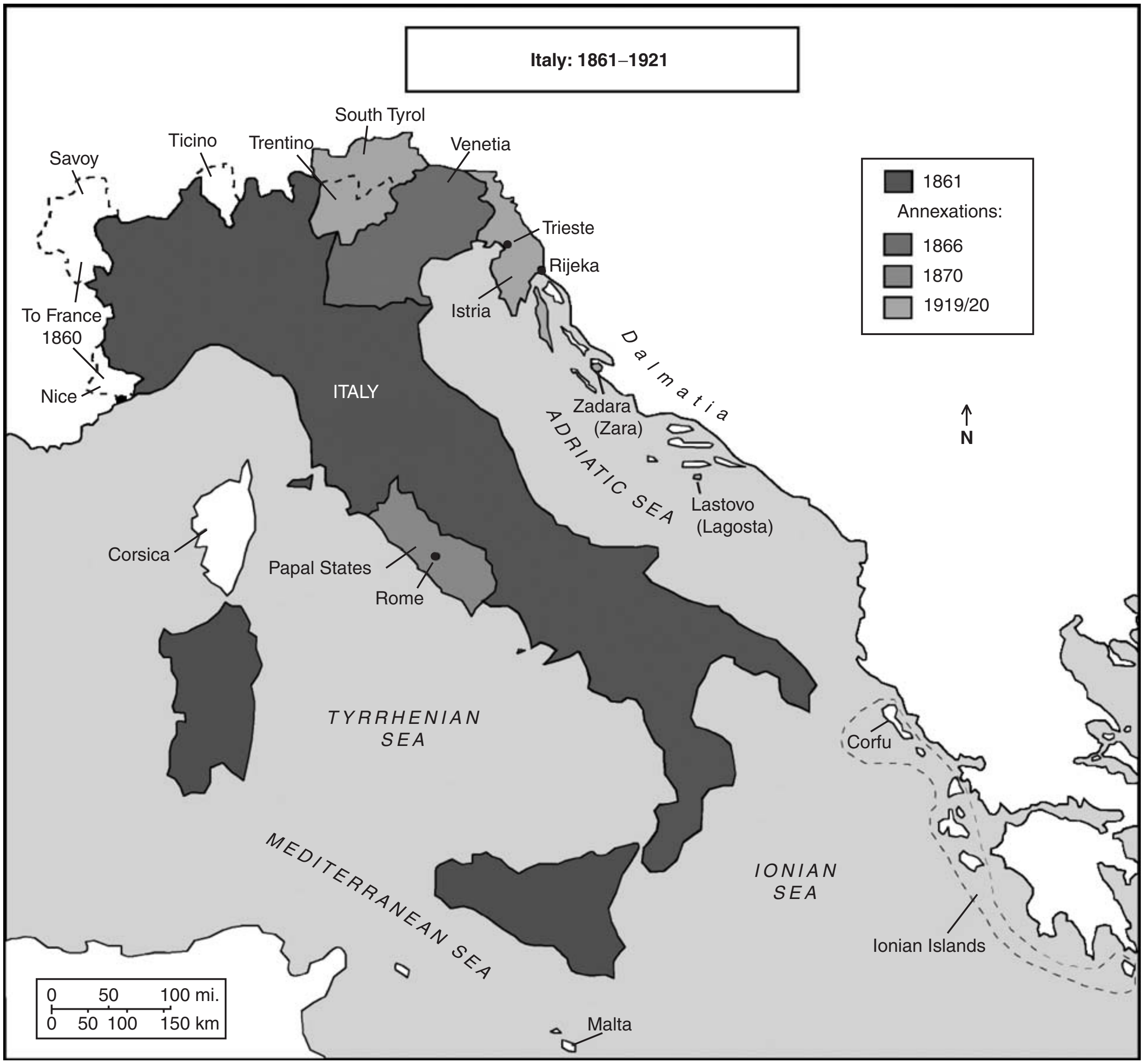

Figure 1 Annexations after unification of Italy.

phenomena, a more nuanced definition of irredentism is frequently sought. However, the attempt to create a more specific definition has resulted in a term with varying definitions.

Noami Chazan and David Horowitz view irredentism as having two types. The first involves the incorporation of co nationals and their territories in neighboring states with their kin into a parent state. Examples are the German minorities in Poland and Czechoslovakia and Greeks in Albania and Turkey as these situations existed between the two 'world wars'. Chazan disqualifies the unification of the two Germanies and the desire to unify the two Chinas and Koreas as irredentism because political differences separated the peoples of these respective countries though they shared culture and history. Chazan's and Horowitz's second type of irredentism refers to cohesive groups that are minorities in multiple countries and seek to be united into parent state. Chazan cites the Kurds of Turkey, Syria, Iraq, and Iran who desire their own Kurdistan. However, if one segment of a people is successful and the others are not, the situation results in secession (a distinct concept) though irredentism was the cause. Such happened with the Slovenes and the Azeris: Slovenes seceded from Yugoslavia but the Slovenes in Austria have not joined them; Azeris seceded from the Soviet Union but the Azeris in Iran have not united with them. Interestingly, these successful secession movements make these groups move from being examples of Chazan's and Horowitz's second type of irredentism to their first type.

Benyamin Neuberger and Thomas Ambrosio do not accept Chazan's and Horowitz's second type of irredentism whereby a stateless people spread across more than one state unite and create their own new nation state. Neu berger points out that in such cases the creation of new states is too much like secession and, thus, would not sufficiently distinguish the two concepts. To maintain a distinction, Neuberger insists that irredentism must in volve a parent state so that peoples and territories across a border can be 'redeemed' as the etymology of the term implies. Moreover, though Neuberger agrees that ir redentism involves annexation of neighboring territories for ethnic reasons, he also identifies three other ways in which territories are identified to be redeemed: (1) pre colonial history (i.e., recreating a past kingdom or empire), (2) colonial history (reestablishing past colonial bound aries), and (3) geography (namely physiography).

Ambrosio also requires a preexisting state and further requires that territorial claims be based solely upon na tional affinity. Viewing irredentism through the prism of international law, Ambrosio identifies three types of irredenta: (1) state–state, (2) decolonization, and (3) state collapse. State–state involves two preexisting states. He points to Somalia's invasion of Ethiopia as the classic example. By including the cases of the two Koreas and the two Germanies of the Cold War in this category, Ambrosio accepts the cases of one nation dominating two separate states as examples of irredentism though Chazan does not. Decolonization refers to states' claims to ter ritories held as colonial possessions and uses Indonesia's annexation of the former Portuguese colony of East Timor as an example. Finally, state collapse refers to the disintegration of states and national groups within them claiming territories beyond those of their existing ad ministrative units in an attempt to unite co nationals into a new nation state. Ambrosio uses the collapse of socialist Yugoslavia and the Croatian and Serbian irredentist projects, as well as the Soviet Union's disintegration and the Armenian project as examples.

Differing definitions of irredentism result in the ac ceptance or rejection of some actual cases as truly irredentist. Nevertheless, considerable overlap in defin itions exists. Context is very important to understanding the nature of irredentism. Indeed, the term was coined to describe the Italian nation state building project. Thus, it was important to describe the context in which irredentism emerged. To gain further appreciation for irredentism's nuances and evolving definitions, it is im portant to examine more closely the situations that are labeled irredentist. German, Armenian, and Somali ir redentism serve here as examples.

Examples of Irredentism

German Irredentism

Much like the Italian situation, no single unified German state existed for much of the nineteenth century. The German state building project, like the Italian project, was inspired by Romantic nationalism with its emphasis on the collective group (i.e., *das Volk*) defined by common language. Indeed, Romantic nationalism traces its origins to early nineteenth century German philosophers. How ever, despite the idea of a common language that united a people, languages themselves were not easily identified because dialects varied considerably over distance and many places contained a mixture of languages. To define the geographic limits of German speech and the so called rightful boundaries of the German nation state, German Romantics used physiographic boundaries. Examples of these boundaries are identified in *Lied der Deutschen* (Song of the Germans) or *Deutschlandslied* – also commonly known as *Deutschland über alles* (Germany above all), ori ginally written in 1841. The very first stanza indicates that Germany should extend "From the Maas to the Memel, From the Etsch to the Belt." The Maas flows deep into the Netherlands, the Memel is found in Lithuania, the Etsch (more commonly known as the Adige) flows in northern Italy and into the Adriatic at a point far south of Venice, and the Belt(s) are waterways well inside Denmark.

These kinds of boundary conceptions provided the framework for the German nation state building project which proceeded by the welding together of numerous German principalities found within the spatial limits of this framework. Prussia and Austria were the driving forces behind this project but were competitors. As early as 1848, the Greater Germany (*Großdeutschland*) and Small Germany (*Kleindeutschland*) proposals emerged. The former would have included most German speaking lands but by including the Austrian Empire, it would have also included a large number of non Germans (e.g., Slavs, Hungarians, etc.). Prussia's triumph over Austria led to the creation of a German nation state in 1871. Though it had non Germans within it, the new German nation state represented the Small Germany conception because it excluded the German speaking areas of the Austrian Empire.

Following World War I, the defeated German Empire was forced to relinquish a number of its territories (**Figure 2**). The Austro Hungarian Empire was dis mantled with many German inhabited areas awarded to the new countries, such the Sudetenland to Czecho slovakia. A newly created German Austria made the idea of unified Germany and Austria more desirable for German nationalists but the Allies prohibited their uni fication. Also, though the German and Austrian Empires lost territories that contained many non Germans, most of these territories were seen by German nationalists as rightfully German lands. Irredentism became strong within Germany.

Adolf Hitler and the Nazis exploited German ir redentist feelings to rise to power. Hitler then satiated irredentist desires by annexing Austria, the Sudetenland,

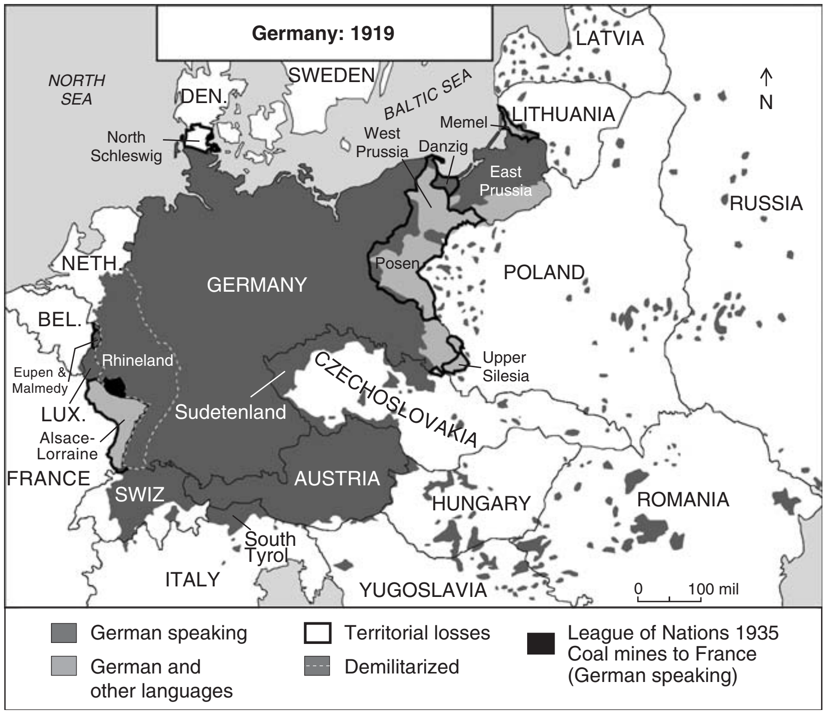

Figure 2 Germans and the German state after 1919. Neth. Netherlands; Bel. Belgium; Lux. Luxembourg.

and many of the other 'lost' territories, as well as ex panding Germany's boundaries to incorporate many scattered German communities of Central and Eastern Europe (**Figure 3**). Harking back to the Greater Ger many idea, the term Greater German Empire (*Groß deutsches Reich*) was employed. More commonly used was Third Reich, a term which proceeded from the Second Reich (1871–1918) and the First Reich (843–1806) (otherwise known as the Holy Roman Empire (of the German Nation)). This term meant to give the Nazi regime historical legitimacy and attempted to legitimize German irredentism.

The destruction of Nazi Germany at the end of World War II resulted not only in the loss of the lands claimed for Germany during the war but also additional lands (**Figure 4**) and the expulsion of Germans from these lands. Re education within Germany has resulted in di minished irredentist feelings except from among the most extreme German nationalists.

Armenian Irredentism

After the dissolution of the Soviet Union in 1991, the Armenian Soviet Republic declared its independence and became the modern Armenian nation state (**Figure 5**). Its boundaries had been determined in 1938 by Soviet leader Joseph Stalin. Modern independent Armenia, however, represents only some of the eastern portions of historic Armenia (**Figure 6**) and does not include all Armenians within it. The western portions of historic Armenia are in Turkey and contain many places, such as Mount Ararat, that are of great significance to the Armenian nation. Additional eastern portions of Armenia are officially part of Azerbaijan, namely Nakhichevan and Nagorno Karabakh.

In the late 1980s, Armenians desiring to unite Nagorno Karabakh with Armenia held mass public demonstrations. In 1989, the Armenian Pan national Movement (APM) emerged and was recognized by the Armenian government. Soviet leader Mikhail Gorbachev proposed greater autonomy for Nagorno Karabakh but both Armenians and Azerbaijanis rejected the plan. The Azerbaijani government began an economic block ade of Armenia. A National Council representing Ar menians in Nagorno Karabakh formed and declared the region's secession from Azerbaijan and its union with Armenia. The Armenian Supreme Soviet declared Nagorno Karabakh's incorporation into Armenia. After Armenia and Azerbaijan became independent in 1991, the conflict over Nagorno Karabakh evolved into full scale armed conflict. Battle lines moved frequently, causing much death and large refugee flows. By 1993, Armenian forces controlled most of Nagorno Karabakh, the land between Armenia and Nagorno Karabakh, and territories south to the Iranian border and eastward into Azerbaijan proper. In 1994, the Russian Federation helped to broker a cease fire. The status of Nagorno Karabakh has not been agreed upon by the Armenian and Azerbaijani governments, though Armenians control the territory.

Figure 3 Hitler's Greater Germany. Switz. Switzerland.

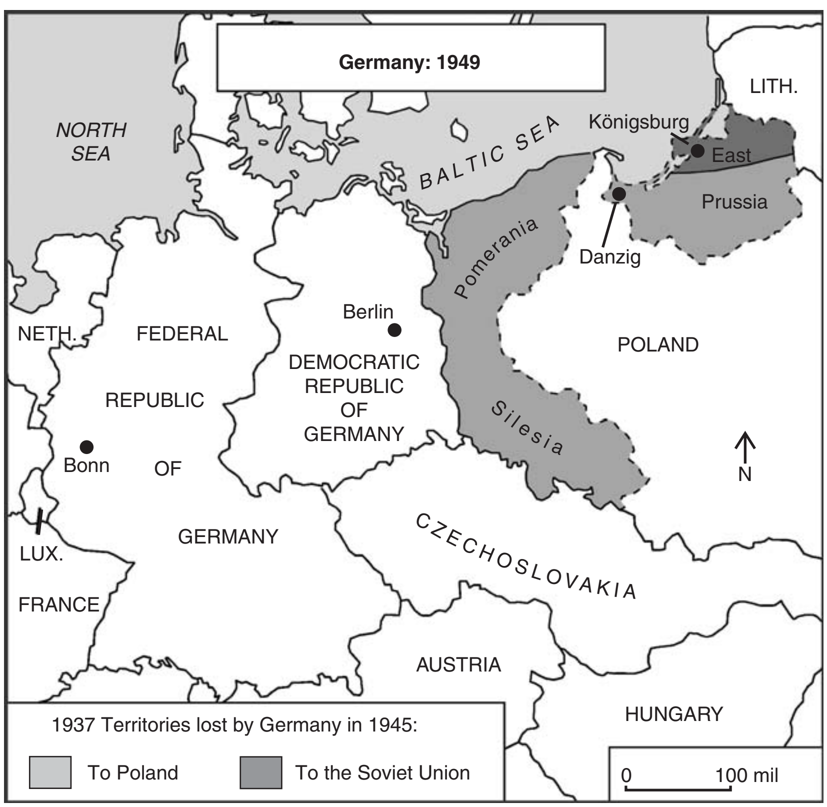

Figure 4 Germany after World War II. Neth. Netherlands; Lux. Luxembourg.

Somali Irredentism

Within Africa, irredentism has been expressed by Somalis, whose distribution extends beyond the modern Somali state into Djibouti, Ethiopia, and Kenya (**Figure 7**). The French, British, and Italians divided the Somalis into different colonies when they carved up the Horn of Africa in the latter part of the nineteenth century. French Somaliland became the independent country of Djibouti in 1977. British Somaliland stretched from the Gulf of Aden in the north to the Ogaden region. The British were forced to yield the Ogaden region to Ethiopian Emperor Menelik II in 1897, effectively creating an Ethiopian Somaliland. Some Somalis were allocated to the British East Africa colony, later to become Uganda and Kenya. Italian Somaliland was in the south along the coast of the Indian Ocean. The modern independent Somali state emerged in 1960 from the unification of British and Italian Somalilands.

Somali nationalists have expressed their irredentism in a number of ways. Somalia's flag features a five pointed star whose points represent the aforementioned Somali inhabited territories. The preamble of Somalia's Constitution, written in 1961, states that the "The Somali Republic promotes by legal and peaceful means, the union of the territories." Somalia's Constitution also grants citizenship to all ethnic Somalis outside the Somali state. Though the Somali government did not claim sovereignty over neighboring territories, it claimed Somalis living in them be granted the right to self determination. For example, as the British government prepared to grant independence to Kenya, which it did in 1963, the Somali government lobbied the British to first relinquish the Somali inhabited Northeast territories of their Kenyan colony. Though a British commission determined that Kenya's Somalis supported annexation, the Kenyan colonial government ignored the findings.

The Somali government and Somali militias supported Somali guerillas in adjacent territories from 1960 to 1964. Somali and Ethiopian armed forces engaged along the border in 1964. The Organization of African Unity (OAU)

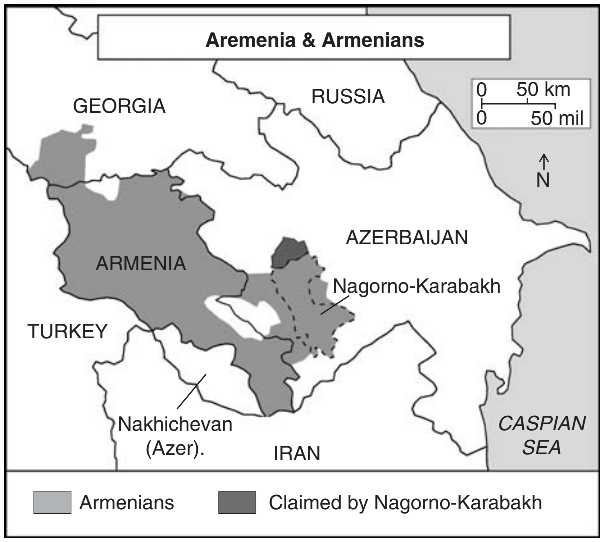

Figure 5 Armenians and the Armenian state in 1991. Azer. Azerbaijan.

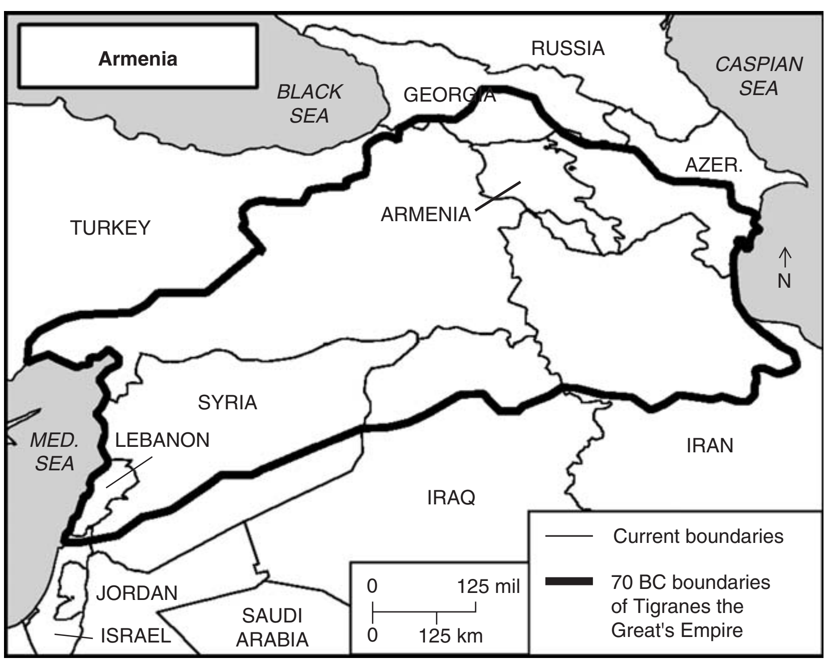

Figure 6 Historic Armenia. Med. Sea Mediterranean Sea.

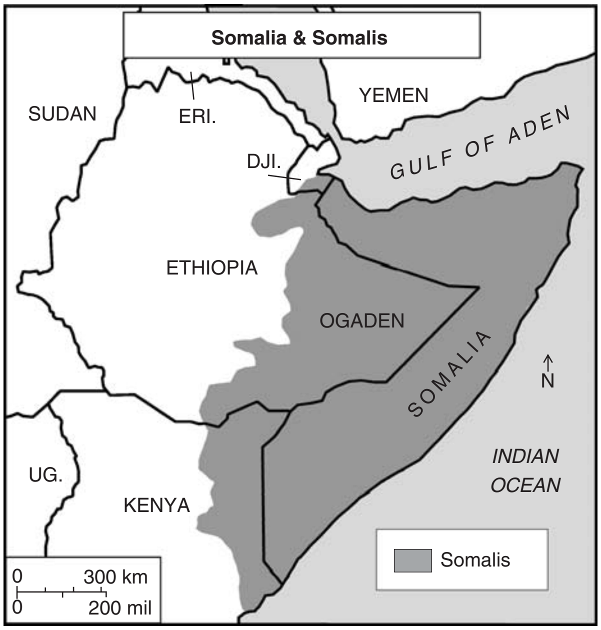

Figure 7 Somalis and the Somali state. Eri. Eritrea; Dji. Djibouti; Ug. Uganda.

brokered a cease fire and established a demilitarized zone on both sides of the border. The armed hostilities prompted Ethiopia and Kenya to create a mutual defense pact that provided for the coordination of both countries' forces in the event of a military attack by Somalia.

After the Ethiopian government weakened in 1974, a Somali separatist group known as the Western Somali Liberation Front (WSLF) attacked government outposts. The Somali government supplied the WSLF and then ordered the Somali National Army (SNA) to invade Ethiopia in 1977. Within a month, more than half of the Ogaden was occupied. Foreign assistance to Ethiopia, including foreign soldiers, defeated the WSLF and SNA in early 1978 and forced the SNA to withdraw from Ethiopia. Though Somali guerillas continued the cam paign, the Ethiopian government firmly reestablished its control over the Ogaden by 1980. With the SNA significantly destroyed, the seriously weakened Somali government has not advanced its irredentist claims. In 1977, 95% of Djibouti's citizens voted for independence rather than union with Somalia though more than 60% of Djibouti's citizens were Somali. In 1981, Somalia's leader ended his claims to Kenyan territory; in 1984, the Somali and Kenyan governments agreed to end hostilities along their countries borders.

Irredentism in the Rest of Africa

With Africa having perhaps the greatest number of spa tial disjunctions between nations and states, irredentism could be a serious issue throughout the continent. However, the number of irredentist claims is small compared to the divided peoples and lost lands. Benya min Neuberger addresses the question as to why more cases of irredentism do not exist in Africa and why the few irredentist projects have failed. He makes several points. Most notably, he argues that Africa has few true nation states – that is, states dominated by one ethnonational group. Many of Africa's partitioned ethnic groups are minorities in all the states in which they reside. Thus, many such groups have no desire for uni fication with their brethren if they also will have minority status in the annexing state and their situation will not improve. Furthermore, with few African states having little more than an ethnic plurality, it means that par titioned groups are too small in number and thus too weak politically to maneuver the state into an irredentist project. Multiethnic state governments themselves do not desire to incorporate more people through land annex ation because such an act would only add to their states' diversity and disrupt their internal balances of ethnic politics. Certainly, the declaration of war likely would only exacerbate internal weakness that in turn would threaten the very existence of the state, as has been the case with Somalia. Moreover, with Africa having so many multinational states, any state's irredentist project only invites counter irredentist claims. It is not surprising then that the African Union (AU) does not support any alteration of country boundaries in Africa.

Conclusion

Irredentism focuses heavily on the idea of uniting people of a shared ethnic affinity and the territories that they inhabit into a single, independent, political state. This idea emerged with the rise of modern nationalism, which began with the French Revolution in 1789. However, ethnic and national identities are not concrete and un problematic but fluid, evolving, and frequently inextric ably intertwined with conceptions of territory. It should not be forgotten that as recently as 1789, only 50% of the French spoke French and only 12–13% of France's population could speak and write French according to the standards of the time. In Italy only 2.5% of the population regularly spoke Italian in 1860. French na tionalists define the French speaking region, and thus the appropriate boundaries for the French, by physical geography: the lands bounded by the Atlantic Ocean, the Pyrenees Mountains, the Mediterranean Sea, the Alps, and the Rhine River. Italian nationalists defined Italy by the peninsula south of the Alps. Early German nation alists defined the German speaking lands by waterways. Other national groups defined their territories by historic kingdoms. Many of these kingdoms, however, had boundaries demarcated along physiographic boundaries because they were easy to identify and defend in premodern times. Interestingly, despite rejecting colonial control, some peoples used historic colonial boundaries to define their national homelands. For example, Cam eroonian irredentists seek to alter the Cameroonian nation state by re establishing the boundaries of the German Kamerun colony as it was before it was divided into a French Cameroun and British Cameroon.

Following World War I at the Paris Peace Conference in 1919, state boundaries were redrawn to large degree according to the principle of 'self determination'. Shared cultural characteristics such as language and religion were emphasized more than ever, ironically helping to fuel irredentism among the defeated who lost territory (e.g., Germans) and helping to bring about another war. After World War II, the international community ceased to accept irredentist claims based on ethnicity though ethnicity became the primary focus for irredentism, es pecially the study of it. Alexander B. Murphy points out that the only legitimate justification for territorial claims since World War II have rested in the argument that a "territory itself had been wrongfully seized." Thus, contemporary irredentist projects may be grounded in the desire to annex adjacent territory across a border inhabited by ethnic brethren; however, in order for the international community to accept such a territorial claim, the claim must have a historical justification.

See also: Political Boundaries; Territory and Territoriality.

Further Reading

Abrahamian, L. H. (1998). Mother tongue: Linguistic nationalism and the cult of translation in postcommunist Armenia. *Institute of Slavic, East European, and Eurasian Studies. Berkeley Program in Soviet and Post Soviet Studies*, http://repositories.cdlib.org/cgi/viewcontent.cgi?article=1014&context=iseees/bps (accessed in May 2008).

Ambrosio, T. (2001). *Irredentism: Ethnic Conflict and International Politics*. Westport, Conn: Praeger.

Ambrosio, T. (2002). Irredentism: Self determination and interstate war. In Beck, R. J. & Ambrosio, T. (eds.) *International Law and the Rise of Nations: The State System and the Challenge of Ethnic Groups*, pp 284 312. New York: Chatnam House Publishers.

Brubaker, R. (1996). *Nationalism Reframed: Nationhood and the National Question in the New Europe*. New York: Cambridge University Press.

Carment, D. and James, P. (1995). Internal constraints and interstate ethnic conflict: Toward a crisis based assessment of irredentism. *Journal of Conflict Resolution* 39, 82 109.

Chazan, N. (ed.) (1991). *Irredentism and international politics*. Boulder, CO: Lynne Rienner Publishers.

Curtis, G. E. (ed.) (1995). *Armenia, Azerbaijan, and Georgia: Country studies*. Washington, DC: Federal Research Division, Library of Congress.

Herb, G. H. (ed.) (1997). *Under the Map of Germany: Nationalism and Propaganda 1918 1945*. New York: Routledge.

Hewsen, R. H. (2001). *Armenia: A Historical Atlas*. Chicago: University of Chicago Press.

Hobsbawn, E. J. (ed.) (1990). *Nations and Nationalism Since 1780: Programme, Myth, Reality* (2nd edn.). Cambridge: Cambridge University Press.

Horowitz, D. (1992). Irredentas and secessions: Adjacent phenomena, neglected connections. *International Journal of Comparative Sociology* 33(1 2), 118 130.

McClellan, G. B. (1915). War stirs Italy to new ideals; irredentism has given way to a larger nationalism, which has united the country as it has never before been united. *The New York Times*, 16 March http://query.nytimes.com/gst/abstract.html?res=9A0CE6D8123FE233A25755C1A9639C946496D6CF (accessed in May 2008).

Metz, H. C. (ed.) (1993). *Somalia: A Country Study* (4th edn.). Washington, DC: Federal Research Division, Library of Congress.

Murphy, A. B. (1990). Historical justifications for territorial claims. *Annals of the Association of American Geographers* 80, 531 548.

Neuberger, B. (1992). Irredentism and politics in Africa. In Chazan, N. (ed.) *Irredentism and International Politics*, pp 97 109. Boulder, CO: Lynne Rienner Publishers.

Schraeder, P. J. From irredentism to succession: The decline of Pan Somali nationalism. *After Independence*, http://www.press.umich.edu/pdf/0472098985 ch5.pdf (accessed in May 2008).

Suganuma, U. (2000). *Sovereign Rights and Territorial Space in Sino Japanese Relations: Irredentism and the Diaoyu/Senkadu Islands*. Honolulu: University of Hawaii Press.

White, G. W. (2004). *Nation, state, and territory. Vol. 1. Origins, Evolutions, and Developments*. Lanham, MD: Rowman and Littlefield.

Relevant Websites

http://repositories.cdlib.org
Mother tongue: Linguistic nationalism and the cult of translation in postcommunist Armenia, eScholarship Repository.

http://www.press.umich.edu
The decline of Pan Somali nationalism, The University of Michigan Press.

http://query.nytimes.com
War stirs Italy to new ideals; irredentism has given way to a larger nationalism, which has united the country as it has never before been united, The New York Times, 16 March, 1915.

Islamic Urbanism

N. AlSayyad, University of California, Berkeley, CA, USA
I. Türeli, Brown University, Providence, RI, USA

Glossary

Caliphal Palace A palace for the caliphs, the leaders of the Muslim World.

Capital City A city that houses the government of its state.

Garrison Town A town developed as a base for an army, and one that features military barracks.

Islamic City An urban form influenced by Islamic practice.

Market (in Arabic, *sūq*) The commercial quarter in a city.

Middle East A loosely defined term used to describe the region between India, Russia, and the Balkans, including the Arabian Peninsula, Turkey, Iran, Iraq, Syria, Palestine, Egypt, and at times part or the whole of North Africa.

Mosque (in Arabic, *masjid*) The place of worship for Muslim males.

Urban Form Physical manifestation of an urban society in the form of buildings, roads, parks, and infrastructure.

Introduction

Initially preached by its Arabian Prophet Muhammad, Islam started as a religion around the beginning of the seventh century AD in the city of Mecca, in what is today Saudi Arabia. By the end of the eighth century, the Islamic empire had extended, reaching India and China in the east and Spain and Morocco in the West. Beginning with the Umayyads (661–750) ruling out of Damascus and later Córdoba, a series of dynastic empires assumed the leadership of Islam each contributing to the development of Islamic urbanism (**Figure 1**). Following nearly 11 centuries of expansion, the Muslim World started falling under European colonial rule during the nineteenth century. It was gradually incorporated into the European dominated world economic system. In the twentieth century, a series of independence movements led to the establishment of new nation states. Despite all the political and economic transformations, these societies continued many Islamic practices under colonial rule. Hence, the concept of Islamic urbanism can incorporate not only historic cities but also cities under conditions of colonization and globalization.

The European invasion of the Muslim World began in Moghul India. Next, Britain, France, and Italy established protectorates over North Africa. Only the Ottoman and the Qajar Empires remained as sovereign but economically dependent Islamic states by the early twentieth century. World War I (1914–18) resulted in the partition of most of the Ottoman Empire's Arab majority provinces by the French and the British. Many Muslim provinces in the Balkans and in Central Asia came under control of the Soviet Union until the end of the Cold War. After the age of empire and colonization, the subsequent nation states formed by Muslim majority societies often claimed Islam as an aspect of national cohesion, as a source of legislation, or as one of the principal sources of constitutional authority.

Islamic urbanism is generally identified with the geographic area designated as the Middle East, but only one fifth of the world's Muslim population lives in this area. Setting aside the popular Western misconception that the Middle East is mainly Arab Muslim land, it is important to note that there is immense diversity within the Middle East with its Coptic, Druze, Maronite, and other religious groups. Indeed, the Muslim World extends well beyond the Middle East. It stretches from Turkey and the Central Asian Republics in the north to Nigeria in the south; and from Mauritania in the west to Indonesia in the east. Hence, the Islamic World is constituted of societies of different economic and political systems; of numerous ethnic and linguistic groups; and of distinct geographic and climatic conditions. A little less than a third of the world's Muslim population lives in the South Asia region. Indonesia is the world's most populous Muslim majority nation with more than 200 million Muslims at the beginning of the twenty first century. It is followed by Pakistan, Bangladesh, and then India, which has the largest Hindu population in the world but whose Muslim minority is still substantial, exceeding 100 million people. In addition, there are significant Muslim populations in China, Russia, and Africa as well as in Europe and the Americas.

Historiography of a Concept

It may be more appropriate to talk about the urbanism of Islamic societies than to talk about Islamic urbanism *per se* although the two narratives are often invoked interchangeably in much of the literature. The 'Islamic' in Islamic urbanism does not only refer to the religion of

Dynasty	Capital cities	Period
Umayyads	Damascus; Córdoba	661–750; 756–1031
Abbasids	Baghdad, Samarra	759–1258
Aghlabids	Kairouan	800–909
Samanids	Bukhara	819–1005
Tulunids	Fustat	868–905
Fatimids	Kairouan, Mahdiya, Cairo	909–1171
Ghaznavids	Ghazna, Lahore	977–1186
Seljuks	Merv, Isfahan; Konya	1038–1194; 1072–1243
Almoravids	Fez, Marrakech	1056–1147
Almohads	Marrakech, Seville	1130–1269
Ghurids	Ghazna	1000–1215
Ayyubids	Damascus, Cairo	1169–1260
Delhi Sultanate	Delhi	1206–1555
Ilkhanids	Tabriz, Sultaniya	1256–1353
Mamluks	Cairo, Damascus	1250–1517
Ottomans	Bursa, Edirne, Istanbul	1281–1922
Timurids	Samarqand, Herat	1370–1506
Mughals	Agra, Shahjahanabad (Old Delhi), Fatehpur Sikri	1526–1858
Safavids	Tabriz, Qazvin, Isfahan	1501–1732
Qajars	Tehran	1779–1924
Fulani	Hausaland, Sokoto	1815–1903

Figure 1 The important dynasties of the Muslim World and their capital cities.

Islam as codified in its sacred texts but also refer to the general culture of Islam as understood by people in different locations. Therefore, it covers the practices of people who may not be strictly following the Islamic faith or who may not be practicing Muslims, nevertheless may be living and working within Islamic societies.

In talking about cities or urbanism, scholars do not invoke notions like Christian urbanism, or the Christian city. So what is it that endows the urbanism of very di verse societies in the Muslim World with that religious title? Some have argued that there may be a special re lationship in these territories between Islam as a system of belief and the practices of Islam on one hand, and the process of urbanization on the other. The religion of Islam came into being in a city, Yathrib, mainly referred to, in Islamic texts, as *Medina* (in Arabic simply meaning the city), suggesting a direct link between Islam and urbanism. Islamic life in *Medina* remains a utopian ideal in Islamic thought until today. Some have even argued that Islam is essentially an urban religion, because only in a city can all of its potential be realized.

As Islam spread, Muslims took over older existing cities as well as established new cities from scratch. Muslim urban order evolved together with the larger culture of Islam. The urbanism of Islamic societies does indeed reflect to a reasonable degree this culture, but it may be difficult to argue that this situation is different from other religiously based civilizations. For example, we no longer speak of Christian cities, like we do of Islamic ones, either as cities simply based in the Christian World, inhabited predominantly by Christians, or formed by a Christian ethic. But it is also true that for a con siderable part of the Middle Ages, medieval European cities could have been labeled Christian cities, particu larly by the then more advanced Muslim Middle East. It is important to note that the 'Islamic' city was originally labeled so by Orientalist scholars.

Islamic urbanism may be an appropriate theme within urban geography because the concept of the Islamic city was fundamental in theorizing the urbanism of the Middle East, the Arab World, and possibly the rest of the Muslim World, particularly in the urban studies litera ture. Early scholars first understood the Islamic city as a self contained entity, made up of a distinct society fundamentally different from its European counterpart. According to the theoretical frameworks and personal politics of these early scholars, discussions on the Islamic city have been carried out under multiple titles such as the 'Muslim city', 'Middle Eastern city', 'near Eastern city', and the 'Arab Muslim city'.

Despite the important contributions of other ethnic groups to Islamic urbanism, studies have focused on Arab cities and the distinct Arab culture that accompanied the rise of Islam. Non Arab Islamic cities were often un studied because it was the Arabs who established the first cities that functioned under an Islamic form of govern ment. The dominance of Arab culture in much of the Middle Ages has also eclipsed the variety of others Is lamic states and regions.

The theorization of the Islamic city began in the service of colonization. After Napoleon's short lasting urban based occupation of Egypt (1798–1801), the study of the Orient by the Europeans evolved into a systematic discipline. In the decades following the French invasion of Algeria in 1830, French scholars were commissioned to describe different aspects of the country in multivolume reports. Hence, the first studies of the Islamic city were produced in this area (Maghrib) by French scholars, and were pursued by the British, German, and later, American scholars in different parts of the Middle East.

Historians and philologists were among the earliest scholars to pay attention to Islamic urbanism. As other scholars such as architectural and art historians and urban geographers participated in the debate, the nature of inquiry changed from a search for a basic Islamic essence to a discussion about the degree of the role of Islam in shaping urban form. This questioning initially generated new debates. Some scholars reverted to the 'essentialists' arguments originally introduced by the Orientalists to revive the concept of the Islamic city. Others attempted to identify unique aspects of Islamic social and legal practices to which they attributed Islamic urbanism.

In terms of method, there have been two approaches to the Islamic city both influenced by the ideas of Max Weber. The art historical approach analyzed the Islamic city formally using morphological and typological analysis. The second, sociological approach examined the structural characteristics of the urban order. Both of these approaches however described a model for the Islamic city that was static and unchanging. Weber theorized the medieval European city in comparison to its non Western, Oriental counterpart. While attempting to explain the origins of modern capitalism, he credited religion for the uniqueness of the European city, and in turn, the European city as its locus. Weber was later criticized for his Orientalism because he claimed the superiority of the European city over its Oriental counterparts in India, China, and the near East. In addition, he also overlooked the impact of European colonialism on these cities. Nevertheless, his thinking continued to inform studies of the city in general, and of Islamic urbanism in particular, throughout the twentieth century.

Weber identified the following five characteristics of the European city: fortification, market, autonomy, association, and autocephaly (independence of the church). Based on his studies of Oriental cities, he argued that several of these characteristics may have existed in the Orient, but they came together only in the European medieval city. While he did not conduct a specific study of the Islamic city or Islamic urbanism as such, he made comments on the negative role of Islam in the development of urban culture. Weber and later his sympathetic critics argued that Islamic cities lacked corporate institutions and a civic life comparable to the medieval European city. While studies based on rigorous archival research into local sources, from Ira Lapidus onward, undermined such assertions, Weber's criteria – religion, kinship, and guilds – remained important issues that influenced Islamic urbanism.

In the Orientalist tradition, Islam was credited with significantly increasing the degree of urbanization in the Middle East. It was assumed that Islam introduced a characteristically Islamic city. The development of this concept was paralleled by the development of descriptions of its physical form. The stereotypical model of the Islamic city consists of a mosque, a palace, a citadel, a market, and residential quarters with inward oriented housing.

The Marçais brothers, writing in the inter war period, first offered the thesis that Islam is an urban religion. They attempted to isolate elements of urban form based on observations of Fez and Marrakech, and theorized the city as a morphological unit. Neither discussed institutional structure and social organization. Their Islamic city consisted of the Friday mosque with a nearby market, and surrounded by a series of public baths. Their model was adopted and developed by other scholars such as Roger Le Tourneau, after World War II, who combined their views with his previous work on the city of Fez. These three important French scholars who studied the Maghrib not only privileged the physical form of cities in their research, but also made assertions about the rest of the Muslim World.

During the same period, Jean Sauvaget identified the market as the central element of his typical Islamic city. The mosque no longer occupied a central place, and a citadel/palace was placed at a high point. These elements appropriated the former urban order, so he argued. In the post war period, Gustav E. von Grunebaum likewise brought together the earlier scholarship on North Africa and that on Syrian cities presenting the 'typical physical form of a city' in terms of institutional structures. He argued that the Islamic city was segregated ethnically, tribally, and occupationally. It lacked a municipal organization as well as a guild like socioeconomic structure. Defining the Islamic city by what it supposedly lacked, he went on to suggest, like Weber before him, that it did not represent a uniform type of civilized life, as did the Greek or Roman town.

It was only in the later 1960s that Ira Lapidus became the first to challenge this model and Weber's assertion that Islamic cities lacked civic culture. Lapidus investigated the forces that established Islamic cities as functional urban entities based on the case studies of Aleppo, Cairo, and Damascus during the Mamluk period. He concluded that Muslim urban society divided essential powers and functions among its different component groups organized in quarters, and urban form was the outcome of the relations among these groups.

Entering the 1970s, Albert Hourani and colleagues followed this line of reasoning in *The Islamic City.*

Referring to Weber's critical components of the city, Hourani suggested that although Islamic cities may have lacked some of his components, they were still able to flourish and maintain a high level of urban activities. Andre Raymond's book *Cairo*, which used court records to identify social and economic communities active in the life of the city during the eighteenth century, set up an example for future studies of Islamic urbanism by introducing a dynamic human element.

It was not until Edward Said's critiques in his seminal book *Orientalism* that scholars started forsaking the Euro centric perspective. Janet Abu Lughod's seminal article 'The Islamic City' critiqued the Orientalist chain of authority that plagued studies of the Islamic city and warned against making generalizations based on singular case studies. She emphasized, instead, the role that the Islamic legal system played in shaping the Islamic City. AlSayyad argued that the Islamic city model was in adequate because as a stereotype it did not take time and urban growth into consideration, and also because it di vorced form from function. He showed that the planned garrison towns, of medieval Egypt, Syria, and Iraq, were shaped mainly by administrative motives and that their growth was neither unstructured nor solely determined by Islam. In the case of taken over towns, the process of their Islamization often involved compromises to deal with predominantly non Muslim populations. Many of the capitals of Muslim dynasties were initially Caliphal palaces whose urban function could not be explained by invoking Islamic urban principles. Urban geographers like Eckart Ehlers contributed to the debate by pointing to the need to include the contemporary city into studies of the Islamic city, and by emphasizing the influence of such institutions as the religious endowment (*waqf*) and the phenomenon of rent capitalism. More recently, Eldem and his colleagues worked on Ottoman cities in cluding the Arab Ottoman city of Aleppo challenging the Arab ideal of the Islamic city generated in part by Euro centrism and in part by modern nationalism.

Urban Form and its Elements

One can identify three historical periods of urbanization and urban development in the Muslim World: a medieval phase, a premodern phase, and a modern phase. During the earlier stage of expansion from the Arabian Pen insula, which we may call, the medieval phase, the Is lamic empires established three kinds of cities: First, they established garrison towns such as Basra (**Figure 2**), Kufa, and Fustat, which they need to maintain the military campaign. Second, they took over older existing cities like Damascus (**Figure 3**). And third, they built palace complexes that developed into imperial capital cities like Baghdad of the Abbasids (**Figure 4**) and Cairo of the Fatimids. In the first type, a variety of Islamic principles calling for modesty and simplicity had an important impact on the allocation of urban functions, the place ment of institutional buildings, and the configuration of urban spaces. These principles were derived from general tradition (*'urf*) and jurisprudence (*fiqh*). Many of the cites that were taken over during the Islamic occupation of the Middle East also underwent a substantial transformation in their form and functions. Planned capital cities, which were initially large palace complexes, experienced a similar process as a result of their being opened up to the public or due to population expansion. All three types of cities seem to have developed the irregular, labyrinthine urban forms (**Figure 5**).

Cities of the Muslim World experienced intermittent periods of instability due to the changing circumstances of trade routes between the eleventh and fifteenth cen turies. The Ottoman Empire expanded its rule to North Africa, the Middle East, and the Balkans between the fifteenth and eighteenth centuries. The security provided by the single state and particularly its institutions such as the religious endowment fostered a vibrant commercial and civic life.

Through time and geography, the central role of the mosque within cities of the Muslim World remained remarkably consistent. Islam did not require a specific place for prayer, nor did it justify an institutionalized clergy. Hence, the mosque evolved as a modest place for the community of (male) believers to come together weekly, regardless of social stature. Traditionally, there were two types of mosques both of which served as spaces of praying, learning, and socializing, albeit at different scales. Residential quarters had small mosques for their own communities. Sultans or high ranking of ficials commissioned much larger congressional (Funer ary or Friday) mosques. Gradually, the urban functions of the congressional mosque was expanded to include formalized spaces for religious schools (*medrese*), Islamic convents (*tekke*), soup kitchens (*imaret*), *khan*s (an urban *caravansary* or inn), public baths (*hamam*), hospices, hos pitals (*darussifa*), and tombs (*türbe*). In particular, Ottoman urbanism came to be characterized by the austere monumentality of these mosque complexes (*külliye*) and the character they gave to the city as exemplified, at its best, in Istanbul.

Islamic empires facilitated trade by providing facilities that served travelers in their cities. They commissioned baths for cleaning, markets for exchange; and accom modation for storage and overnight stays. Markets con sisted of covered streets (*sūq*), monumental strong rooms or buildings for the trade of precious objects (*bedestan*), and warehouses. Islamic empires further contributed to the civic life of the cities under their control by building citadels, monastic complexes, and educational establishments.

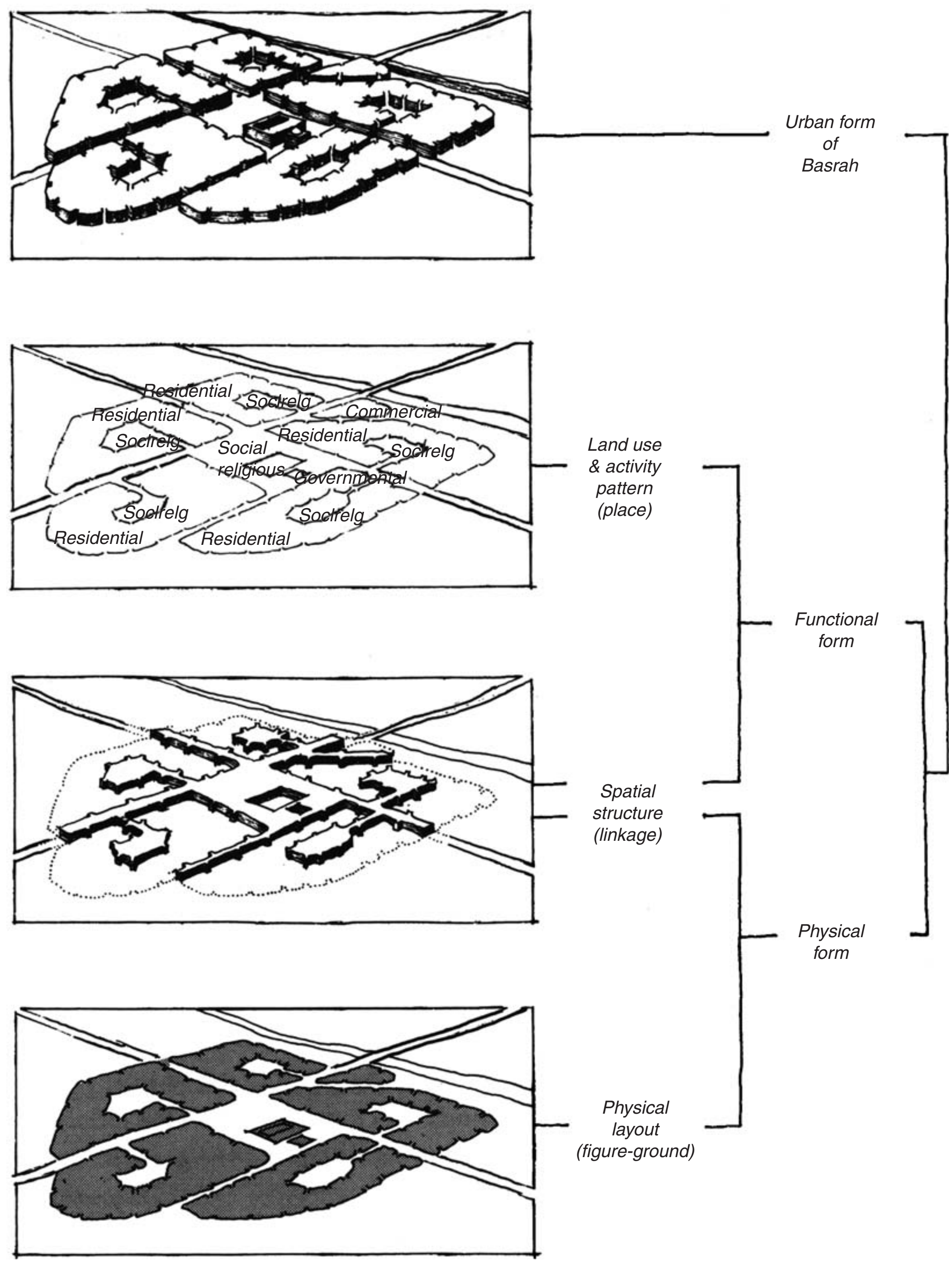

Figure 2 Basrah's urban form analyzed. Source: AlSayyad, N. (1991). *Cities and Caliphs: On the Genesis of Arab Muslim Urbanism*. Westport, CT: Greenwood Press. p 53.

A third element of urban form observable in the cities of the Muslim World was the palace complex. The Ab basids initiated the grand palace city complexes with the round city of Baghdad. The Safavids' capital Isfahan featured a palace complex consisting of a succession of rectangular units of palaces and gardens inspired by the traditional Iranian gardens. The last of the Islamic grand palace complexes was the Ottomans' Topkapi Palace, which manifested the sovereignty of the sultan to his subjects through its location and architecture, looking over the whole city.

The fourth element of urban form in cities of the Muslim World described in various models of the Islamic city was the residential quarter. The quarter was defined according to the tribal, ethnic, or religious character of their occupants. It was a cohesive community the

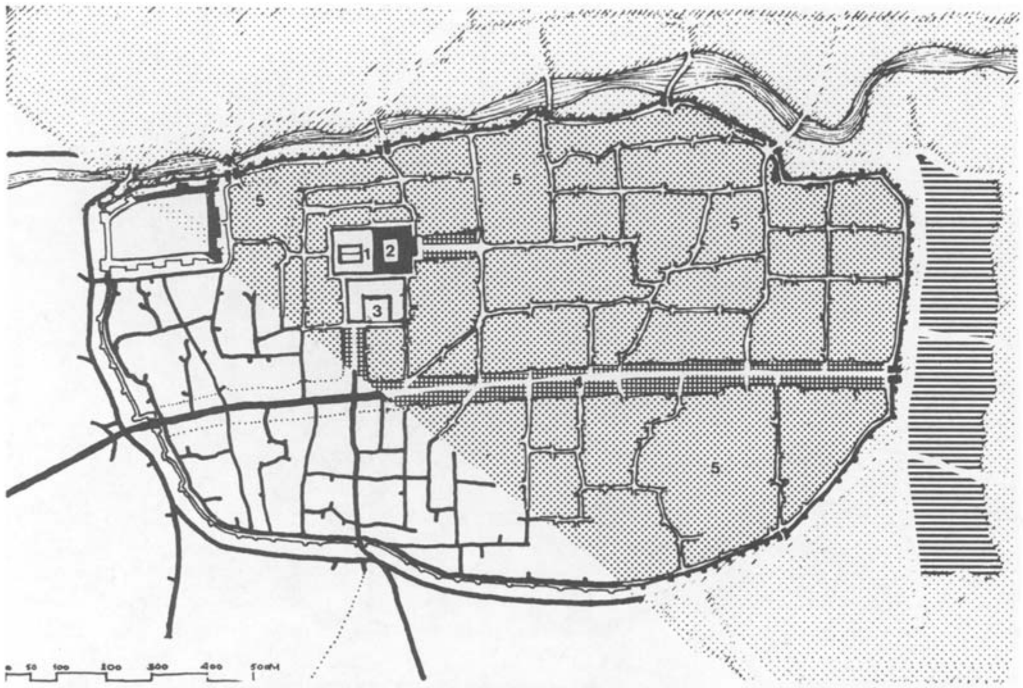

Figure 3 The plan of the Islamic Damascus, based on Al-Munjid. 1. Church of St. John 2. Umayyad Mosque 3. Al-Khadra Palace 4. Market/Colonnade 5. Residential Quarters 6. Cemetery. Source: AlSayyad, N. (1991). *Cities and Caliphs: On the Genesis of Arab Muslim Urbanism*. Westport, CT: Greenwood Press. p 94.

admission to which was regulated. Administratively, the quarter was a taxation unit for the government. In terms of public services, it was responsible for taking care of itself. The houses within the quarter were characterized by extended family living, and inward orientation organized around a courtyard (**Figure 6**).

The Contemporary Islamic City

The Muslim World witnessed a second period of urban restructuring, owing to the rise of modern capitalism and the emergence of organized political and economic colonialism. As Muslim majority societies were incorporated into the new economic system, they had to turn away from manufacturing to agricultural production. They became importers of European products. The mercantile cities of the Muslim World became nodes of transport between their agricultural hinterlands and the industrial port cities of Western Europe.

The cities of the Muslim World were transformed formally by the colonial powers through the addition of dual cities. The development of these planned urban entities, initially for the colonizing elite, is well illustrated in such cities as Algiers and Rabat in North Africa. In planning the foreign districts, colonial administrations followed specific European models, including the use of geometric grids and a hierarchical ordering of urban functions and spaces. In some instances, as in Cairo, the new urban expansion followed the French planning tradition of Haussmann, by which squares were constructed as urban nodes, and then, connected diagonally to penetrate the older, irregular fabric.

Following de colonization and the rise of independence movements from the 1940s through the 1960s, nationalistic attitudes brought about an increasing awareness of the cultural values implicit in the urban systems of Islamic societies. Governments welcomed operational definitions of Islamic urbanism before drawing new planning guidelines. This partially explains the numerous meetings on the Islamic city sponsored by research groups and political organizations throughout the Muslim World and their numerous publications ranging from reactionary treatments by nationalist scholars to the more critical additions to and critiques of the Orientalist approach.

By the early 1960s, most peoples of the Muslim World had regained independence. New Islamic nation states in Africa, Asia, and the Middle East emerged as international deals struck at the close of World War II and from successful revolutionary movements. Following independence, departments of planning, housing, and urban development were set up as cabinet level ministries to implement the agenda of modernization.

Massive migration from rural areas to cities has led to enormous changes in the urban scene. Economic modernization has dramatically transformed the urban form of cities in the Muslim World. The social order of both

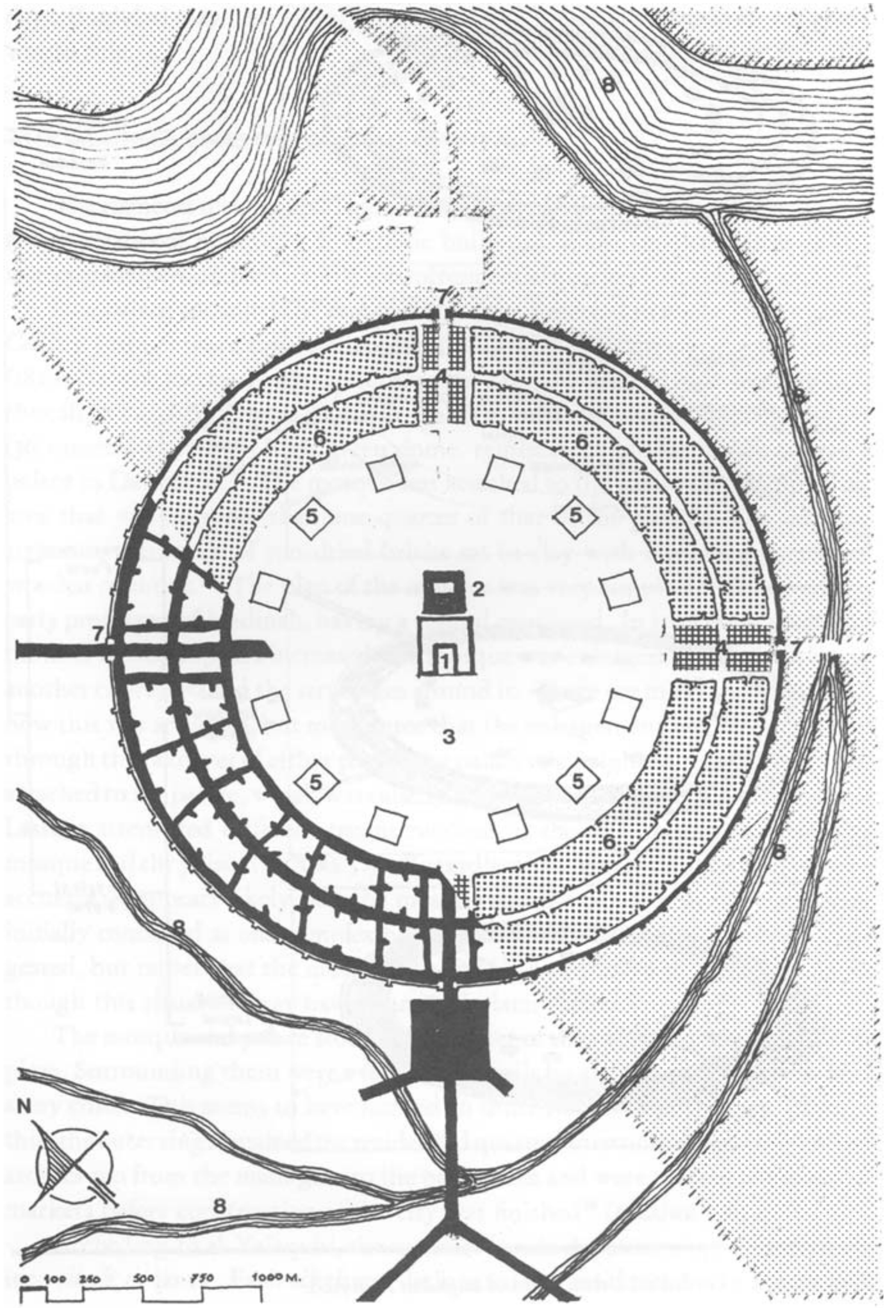

Figure 4 The general site arrangement of Baghdad and its urban elements. 1. Caliphal Palace 2. Mosque 3. Rahbah 4. Markets 5. Palaces of the royal family. Source: AlSayyad, N. (1991). *Cities and Caliphs: On the Genesis of Arab Muslim Urbanism*. Westport, CT: Greenwood Press. p 121.

the extended family and residential quarter transpired along with the urban morphology they engendered. The mosque complex is no longer the heart of the city as many small and medium size mosques were built everywhere in and around cities. In some older cities, the market has turned into a tourist attraction.

In most cases, independence brought with it the nationalization of land and a planned economy, which gave rise to the informal housing sector. In major urban centers, more than half of the urban population lives in the so called informal settlements comprising single and multistory dwellings, which are built on subdivided agricultural land without building permits. The impasse of modern planning, against a vibrant informal sector, led design and planning professionals in the Muslim World by the early 1980s, to question the validity of their

Figure 5 Urban fabric in the stereotypical Islamic city (Isfahan).

professional teachings derived from Western models and to visit traditional Islamic urbanism as a possible guide to the contemporary city.

More recently, with neoliberal urban restructuring, gated enclaves and 'splintering urbanism' started making a presence in the lives of Muslim people. Oil dollars have opened a new chapter in Islamic urbanism. As aspiring national leaders of Islamic countries are seeking to turn their countries into players in the global marketplace, their cities are becoming key to this objective. One of the most ambitious projects of the last decades is Malaysia's Multimedia Super Corridor, a vast urban redevelopment that stretches from the Petronas Twin Towers to the Kuala Lumpur International Airport, designed to jump start Malaysia in the information age. The island cities of Dubai (The Palm Islands, The Dubai Waterfront, and The World) in the United Arab Emirates are another such ambitious project. All these investments have captured worldwide attention and stimulated capital flow to these Islamic countries, but they have also highlighted issues of social injustice.

While one must accept that the model of the Islamic city developed by early scholars has now been debunked, it is also necessary to acknowledge that the concept of Islamic urbanism embodies historical realities that are fundamental in understanding the issues that are specific to the Muslim World. Hence, Islamic urbanism remains a compelling analytical category in the study of urbanization and urbanism of Muslim majority societies today.

Figure 6 Enclosed residential quarters (Riyad).

See also: Medieval Geography; Medieval Historical Geographies; Middle East and North Africa; Urban Architecture; Urban Design.

Further Reading

Abu Lughod, J. L. (1987). The Islamic city: Historic myth, Islamic essence, and contemporary relevance. *International Journal of Middle East Studies* 19(2), 155 176.

AlSayyad, N. (1991). *Cities and Caliphs: On the Genesis of Arab Muslim Urbanism*. Westport, CT: Greenwood Press.

Bonine, M., Ehlers, E., Krafft, T. and Stober, G. (eds.) (1994). *The Middle Eastern City and Islamic Urbanism: An Annotated Bibliography of Western Literature*. Bonn: Ferd. Dummlers Verlag.

Brown, L. C. (ed.) (1992). *From Madina to Metropolis; Heritage and Change in the Near Eastern City*. Princeton, NJ: Darwin Press.

Ehlers, E. (1992). The city of the Islamic Middle East. In Ehlers, E., Bahr, J. & Harris, C. D. (eds.) *Modeling the City: Cross Cultural Perspectives*. Bonn: In Kommission bei Ferd. Dummlers.

Eldem, E., Goffman, D. and Masters, B. A. (1999). *The Ottoman City between East and West: Aleppo, Izmir, and Istanbul, Cambridge Studies in Islamic Civilization*. Cambridge, MA: Cambridge University Press.

Grabar, O. (1973). *The Formation of Islamic Art*. New Haven, CT: Yale University Press.

Haneda, M. and Miura, T. (eds.) (1994). *Islamic Urban Studies: Historical Review and Perspectives*. London: Kegan Paul.

Hourani, A. H. and Stern, S. M. (eds.) (1970). *The Islamic City: A Colloquium*. Oxford: Cassirer, University of Pennsylvania Press.

Lapidus, I. M. (1967). *Muslim Cities in the Later Middle Ages*. Cambridge, MA: Harvard University Press.

Lapidus, I. M. (ed.) (1969). *Middle Eastern Cities*. Berkeley, CA: University of California Press.

Le Tourneau, R. (1957). Les Villes Musulmanes De L'afrique Du Nord. Alger.

Marçais, G. (1940). L'urbanisme Musulmane. *Congres de la Federation des Societes Savantes de L'Afrique du Nord*, Algeria.

Marçais, W. (1928). L'islamisme Et La View Urbaine. *L'Academie des inscriptions et belles lettres: Compts Rendus* 86 100.

Raymond, A. (2000). Wood, W. (trans.). *Cairo*. Cambridge, MA: Harvard University Press.

Sauvaget, J. (1941). *Alep*. Paris: Librairie Orientaliste P. Geuthner.

Von Grunebaum, G. (1955). The structure of the Muslim town. In *Islam: Essays in the Nature and Growth of a Cultural Tradition*, pp 141 158. London: Routledge and Kegan Paul.

Relevant Websites

www.archnet.org

Archnet is an online English spoken community with a focus on art, architecture, and design in the Muslim world.

Island Development

E. Clark, Lund University, Lund, Sweden

Introduction

The contexts of island societies differ in decisive ways from those of continental societies. Differences in size, transportation conditions, diversity of natural resources, and relations to hinterlands are among the most significant characteristics that underlie the rationale for singling out islands as a distinct field of regional development research and policy. Insularity is commonly defined and measured in terms of three key variables: size, distance to a mainland, and extent of contact and exchange with other places.

Island societies account for a disproportionate share (in relation to land area) of global human population. With decolonization they also account for a disproportionate share of sovereign states. Islands furthermore harbor a disproportionate share of global biocultural diversity (biocultural signifying the intrinsic relatedness of biodiversity and cultural diversity). These conditions contribute to the rise in interest in island development issues during recent decades. The interests and perspectives of island societies have been increasingly voiced and heard in international organizations. At the same time, critical research poses questions concerning who is speaking for whom, and in whose interests various regional development models work. The history of development on many inhabited islands today is stamped by colonialism and imperialism, as well as opposing forces. In more recent postcolonial history (not in denial of the colonial present), small island states fare remarkably better than their larger continental counterparts in practicing democracy, with all that implies for development. There are advantages as well as disadvantages associated with islandness, and the great diversity of geographical conditions and historical contexts of development in island societies does not lend itself to simple generalizations.

Islandness and Size

Islands with large land masses, large populations, and/or located adjacent to a continent or other large island share little in common with small remote islands. Honshū and Easter Island are both islands, the former the 11th largest landmass on Earth and home of the global city of Tokyo, the latter topping the United Nations Environment Programme (UNEP) isolation index of inhabited islands (although Tristan da Cunha, Saint Helena, and other islands without airports are arguably more isolated). For this reason it makes analytical sense to subdivide islands into groups rather than treat them all as one broad category. Furthermore, small landlocked continental states have more in common with small island states than do islands with large populations and land areas. For this reason, there is considerable overlap in research and policy organization between small island (including nonsovereign jurisdictions) and small state issues.

Drawing the line around what constitutes an island – upward vis à vis continents and downward vis à vis skerries and sandbars – is notoriously arbitrary. Sweden has, for instance, over 220 000 islands according to Statistics Sweden, but only 24 by European Union standards. The population ceiling for delineating small states or economies varies from less than 10 million to less than 1 million, while lower ceilings of 100 000 or even lower magnitudes are commonly used to identify small islands. The less tangible but no less significant dimension of consciousness is also invoked to narrow in on small islands, namely, those islands where the inhabitants are permanently conscious of being on an island.

Challenges to Island Development

Challenges to development particularly associated with small islands are peripherality in relation to centers of decision making, limited diversity of natural resources and skills base, small domestic markets, income volatility, and high transportation and communication costs. Absence of contiguous hinterland and associated economies of scale coupled with low level of resource diversity impose openness toward world markets and specialization in one or a few income generating sectors. A crucial challenge to the success of these sectors is securing adequate transportation infrastructure, often requiring access to external capital in order to realize such large investments.

On the other hand, the same absence of a rural hinterland has been forwarded as an important explanatory factor behind the rapid growth of island city states, such as Singapore and Hong Kong, for which the slow track of agriculture was not an option. Furthermore, some studies suggest that small size may be associated with greater solidarity and equality, and thereby fewer and less disruptive free rider problems and violent conflicts. In addition, flexibility and capacity to adapt have been argued to be central components of social capital of particular importance in small island societies. Island societies have been innovative in governance and in tackling challenges associated with insularity. Jurisdiction has been effectively utilized to strategically attract rent based incomes, from prison services and hosting film productions to offshore finance centers and the sale of Internet domains. Economies of place – capitalizing on geographical uniqueness that can add value to goods; a form of monopoly rent – can make up for diseconomies of scale. Some recent studies indicate that small island states are both better off and more successful in terms of economic growth than large countries. Also in dealing with challenges of local sustainability, island societies have been highly creative in environmental management and ecological practices, providing rich variation in experience from which to learn.

Transportation

Conditions surrounding transportation are a key aspect of island development, inherently tied to relations to hinterland. Connections between small islands and other places are of necessity via sea or air. Island societies are therefore dependent on sea and air links for exchange with the outer world. Islands with fixed links in the form of bridges or tunnels to a mainland or large island are commonly excluded from the category of island for this very reason: they lose a characteristic that is considered fundamental to islandness.

As transportation technology has developed and relative costs of land, air, and sea transportation have changed, so have the conditions for this key aspect of island development. Shipping and cargo handling have been revolutionized in keeping apace with developments in efficiency of air and land transportation, with consequential peripheralization of islands previously functioning as hubs in transportation networks. Costs for investments in modern transportation infrastructure are prohibitively high for small island societies, exacerbating dependency on exogenous sources of finance. Exogenous interests tend toward investment in external links more than internal network development.

While the constraints imposed by modern transportation technology are widely considered the single most distinguishing constraint of island development today, developing transportation capacities to surmount these constraints commonly entail concessions to exogenous interests tied to the financing of such large infrastructural investments. These interests and the projects they pursue are not necessarily the same as those of the island population, and the material and value flows generated by the projects do not necessarily benefit the island population. Transportation infrastructure facilitates penetration of exogenous economic forces as well as expansion of endogenous economic activities. Transportation facilitates exchange, also ecologically unequal exchange.

Models of Island Development

A model of island development that has been highly influential since its formulation in 1985 is the so called MIRAB model, an acronym for migration, remittances, aid, and bureaucracy. It describes a form of welfare maximizing strategy in late colonial or postcolonial contexts based on remittances from migrants abroad, aid (which is often more easily procured in small economies, especially from a former or vestigial colonial power), and accompanying aid and workfare bureaucracy. Once established, it tends to lock into place and is sometimes seen as an all too widespread paradigm of development failure, to be overcome rather than emulated.

While some small island developing states around the world have been found to fit into this category, there are other development paths and models that differ markedly from MIRAB. One such alternative based on five capacities which characterize the creative political economies of another set of small island societies is the PROFIT model, another acronym standing for people considerations (affecting citizenship and rights of residence and employment), resource management, overseas engagement and recognition, finance, and transportation. In contrast to MIRAB, these island societies are characterized by economically strategic immigration and migration policies, toughness in negotiating exploitation of local resources, control over viable means of transportation, and tax regimes designed to attract foreign investment, all of which are based upon self rule. The PROFIT model emphasizes the use of jurisdiction and paradiplomacy to establish endogenous policy formulation and powers of implementation.

Another model, under yet another acronym, is the SITE model of small island tourist economies. Islands command a powerful resource as imaginary and real sites of retreat and refuge from the stress of work, or escape and hideaway from the mundane humdrum of everyday relations. This source of rentier income has become a major sector of many small island economies. Like the export oriented plantation economies before it, however, the tourist industry often contributes little to sustain local economies in the form of local investments, tourism income flowing, instead, largely out of the economy. The expansion of tourism often involves absentee ownership, displacement, and gentrification of island communities, that is, a form of accumulation by dispossession. As long as islands continue to lure tourists, tourism will continue to lure island ‘developers’, maintaining its position as strategic sector for many island societies, of MIRAB, PROFIT, SITE, or other mold.

Uneven Island Development and Island Gentrification

Islands may be geophysical givens of material reality, but how they develop is determined more by their positions, positioning and relations in an evolving social ecological system than by their physical limits or characteristics, however much the preexisting territorial structure of smaller landmass surrounded by water constitutes a precondition for – and is discursively utilized in – the production of regionality. If, as Torsten Hägerstrand suggested, geography is constituted through struggles for power over the entry of entities and events into space and time, then understanding development on small islands must involve scrutinizing such struggles relevant to that development. Why remittances and aid here, and offshore finance there? Seeking explanations limited to endogenous characteristics of inertia or creativity, resource endowment or policy choices, fails to position struggles over island time–space and resources into broader contexts of historical global capitalist relations.

The above models are commonly presented as descriptions of reality, and also carry a significant element of normative policy recommendation, more or less explicitly with an eye toward economic growth. Here is a good example of how to encourage economic growth, there is an example of poor achievement in economic growth. From the perspective of uneven development, however, these models emerge as coherently connected rather than mutually exclusive ‘paths’ which island societies can ‘choose’ to follow, if only they make the optimal policy decisions. Island societies need to be seen as connected to geographically broader processes of capital accumulation. They are not only vulnerable to forms of exploitation and devaluation inflicted by capital accumulation elsewhere, but also actively engage in processes of capital accumulation. The PROFIT model focuses on island societies that have been particularly successful in engaging in accumulation by dispossession (both at home and abroad), while the MIRAB model applies to islands that have been more the peripheral object of than the designers of capital accumulation. The one is the upside, the other the downside of the same connected processes of uneven development.

The Cayman Islands, now known largely as a center for offshore finance, are an applauded example of success in making the transition from one of the poorest to one of the richest Caribbean island societies in a very short period. This is the PROFIT model par excellence. The perspective of uneven development underscores how the offshore finance of the Cayman Islands not only dispossesses considerable swaths of the population in the unfolding of this success story, but is also directly related to the redlining of and exploitative extraction from other less successful island societies, some perhaps locked into MIRAB conditions, others following (or subjected to) a SITE development path. Characteristics emphasized in analyses of PROFIT islands, namely, flexibility and adaptability to strategic opportunities in the global arena, are the very same characteristics of capitalist, historical geography emphasized by theorists of uneven development. It seems island societies can also play the game, and play it well.

Consequently, gentrification of island communities has become a common conflict ridden process associated with development on many islands. The challenge for these communities is to take control over the development process and strengthen local economies while avoiding displacement in the process.

Conclusion

The specific landform of small islands lures not only imagination and tourism, but also a susceptibility to fall into one of two traps. It is common to see islands as microcosms of the larger world. This holds only in the very general sense that the Earth is an island in the cosmos, and tends to distract from serious considerations of the more specific contexts and issues of small island development. Except in this very general sense, small islands are not microcosms of the larger world, even if some common relations are more readily cast in relief in the context of small island societies – which explains the disproportionate use of islands as laboratories for scientific fieldwork as well as imaginary places in works of fiction.

The opposite susceptibility is to fall into the ‘local trap’, to perceive these places as unique not only in the banal sense that every place is unique, but also indeed essentially different, bounded, self contained communities, outside and apart from their continental other, where ‘global’ processes originate and are orchestrated. The myth of the primitive isolate refuses to be forgotten, in spite of overwhelming evidence that there are and have always been connections everywhere, that the world of humankind, as Eric Wolf argued, constitutes a totality of interconnected processes. Island societies are not isolated from global processes of capital accumulation, or outside the uneven regional development these processes generate. They are, however, unequally positioned to meet and engage in these processes. A major question for future research into island development is how to move from the MIRAB, PROFIT, SITE, and other formulas of unequal development toward more equitable, just, and sustainable paradigms of island development.

Islands are the sites of a disproportionate share of biological and cultural extinctions in the global decline of diversity. As generators and custodians of both bio diversity and cultural diversity, island societies will therefore unavoidably play a key role in any political process aiming to stem this decline. Not least for this reason, island development issues concern us all.

See also: Diaspora; Finance, Offshore; Gentrification; Local Development; Migration; Oceania; Oceans; Regional Development Models; Regional Inequalities; Remittances; Scale; Tourism; Uneven Development.

Further Reading

Baldacchino, G. (2006). Managing the hinterland beyond: Two ideal type strategies of economic development for small island territories. *Asia Pacific Viewpoint* 47, 45 60.

Baldacchino, G. (2006). Innovative development strategies from non sovereign island jurisdictions? A global review of economic policy and governance practices. *World Development* 34, 852 867.

Baldacchino, G. (ed.) (2007). *A World of Islands: An Island Studies Reader*. Charlottetown, Canada: Institute of Island Studies.

Bayliss Smith, T. P., Bedford, R., Brookfield, H. and Latham, M. (1988). *Islands, Islanders and the World: The Colonial and Post Colonial Experience of Eastern Fiji*. Cambridge: Cambridge University Press.

Bertram, G. and Watters, R. F. (1985). The MIRAB economy in South Pacific microstates. *Pacific Viewpoint* 26, 497 520.

Biagini, E. and Hoyle, B. (eds.) (1999). *Insularity and Development: International Perspectives on Islands*. London: Pinter.

Brookfield, H. (1990). An approach to islands. In Beller, W., d'Ayala, P. & Hein, P. (eds.) *Man and the Biosphere Series 5: Sustainable Development and Environmental Management of Small Islands*, pp 23 33. Paris: UNESCO.

Clark, E. (2004). The ballad dance of the Faeroese: Island biocultural geography in an age of globalization. *Tijdschrift voor Economische en Sociale Geographie* 95, 284 297.

Clark, E., Johnsson, K., Lundholm, E. and Malmberg, G. (2007). Island gentrification and space wars. In Baldacchino, G. (ed.) *A World of Islands: An Island Studies Reader*, pp 483 512. Luqa, Malta: Agenda Academic.

Garcia Herrera, L. M., Smith, N. and Mejías Vera, M. Á. (2007). Gentrification, displacement and tourism in Santa Cruz de Tenerife. *Urban Geography* 28, 276 298.

Gossling, S. (ed.) (2003). *Tourism and Development in Tropical Islands: Political Ecology Perspectives*. Cheltenham: Edward Elgar.

Hodgett, S., Johnson, D. and Royle, S. A. (eds.) (2007). *Doing Development Differently: Regional Development in the Atlantic Periphery*. Sydney, NS: University of Cape Breton Press.

Lockhart, D. G., Drakakis Smith, D. W. and Schembri, J. (eds.) (1993). *The Development Process in Small Island States*. London: Routledge.

McElroy, J. L. (2006). Small island tourist economies across the lifecycle. *Asia Pacific Viewpoint* 47, 61 77.

Srebrnik, H. (2004). Small island nations and democratic values. *World Development* 32, 329 341.

Relevant Websites

http://www.globaloceans.org
Global Forum on Oceans, Coasts, and Islands.

http://www.globalislands.net
Global Islands Network (GIN).

http://igu islands.giee.ntnu.edu.tw
IGU Commission on Islands.

http://www.insula.org
International Scientific Council for Island Development (INSULA).

http://www.geol.utas.edu.au
International Small Islands Studies Association (ISISA).

http://www.islandstudies.ca/journal
Island Studies Journal.

http://www.sidsnet.org
Small Island Developing States Network (SIDSnet).

http://islands.unep.ch
UNEP Islands.

http://www.unesco.org
UNESCO Environment and Development in Coastal Regions and in Small Islands (CSI).

http://www.smallislandsvoice.org
UNESCO Small Islands Voice.

Italian Language Geography

C. Minca, Royal Holloway, University of London, Egham, UK

Introduction

Mapping contemporary Italian geography today is a rather difficult task. Despite the discipline's strong reputation in the past, when Italian scholars contributed in important ways to continental debates on key geographical concepts, such as landscape and region, the last few decades have been characterized, by and large, by the lack of a clear theoretical core or neatly delineated schools of thought.

Some historical and institutional context might help in understanding the present situation. Modern Italian geography began taking shape in the last decades of the nineteenth century, but was limited in its influence by the political and cultural context of the newly unified Italian state. In contrast to other European countries, academic geography emerged at a time when the Italian state was still struggling to organize and administer a territory that was extremely diversified and fragmented from a cultural, a political, and an economic point of view. It was not until the inter war period, under the Fascist regime, that the discipline would begin to be consolidated. The fascist state saw geography as a key element in the education of the 'new Italian man', but also as an important tool in the production of the scientific and cultural legitimacy for the Italian colonial enterprise (and, more generally, for the affirmation of Italy as a great power). This was a phase of rapid expansion for the discipline, both at the school and university level, marked by the influence of both German geography and *geopolitik*, but also the French tradition of regional geography. The influence of these two traditions would leave an important mark on Italian geography in the years to follow, privileging the production of regional monographs marked by a largely descriptive and somewhat deterministic flavor. The distinct disciplinary foundations laid down during this period would have a significant impact on the future of Italian geography at an institutional level in the post World War II era as well, as the newly formed Italian Republic grappled with the legacy of the regime.

In the decades following World War II, geography took on a strong humanist tinge, assuring it significant visibility in many faculties of philosophy, literature, and foreign languages, while the discipline remained relatively marginalized within the social sciences, with the exception of some faculties of economics where economic geography created a niche for itself (although this latter suffered a rapid decline in the last decade or so, in some cases disappearing entirely from the curriculum). The reason why geography's positioning within the teaching structure of Italian higher education is particularly important is that its research profile has been largely determined by this very positioning. This explains, for example, why there are so few 'departments of geography' proper in Italy, and why most geographers work and teach – often in rather marginalized positions – in departments or faculties dominated by other disciplines. The first consequence of this academic structure is the relative absence of strong clusters of geographers specialized in one specific subfield and working together under one institutional umbrella, as is common in other national contexts; a second, related consequence, is the seemingly scattered nature of much of the research produced by Italian geographers. A third – and equally important – consequence is the apparent difficulty encountered by the different Italian geographical associations and societies in keeping the national community of geographers compact and visible – this is also due to the fact that individual geographers often struggle with 'local' academic conditions where, to survive the disciplinary competition, they are forced to accept major administrative roles and sacrifice their research objectives. Finally, the lack of any formal (and financial) mechanism for recognizing the quality of research in the Italian university system, coupled with the increasing pressure for more and diversified teaching, has inevitably taken many geographers away from mainstream theoretical debates and from active participation in the international geographical community.

This relatively weak structuring of Italian academic geography has, indeed, produced in the last few decades rather 'flexible' disciplinary borders and some methodological 'looseness', often resulting in internal fragmentation and lack of continuity and communication. But paradoxically, as happens sometimes in less structured fields of knowledge, such relative disengagement and fragmentation have also allowed for the emergence of moments of exceptional creativity and originality on the part of small groups of individuals, together with some episodes of extraordinary scholarship. For this very reason, it is perhaps better to avoid any attempt at a conventional mapping of disciplinary boundaries and achievements, and focus rather on identifying and commenting on a limited but significant series of disciplinary threads that, although in a rather fluid and discontinuous fashion, have been 'making the discipline' in the last decades, or at least have been the source of some of its more interesting production. Although it might be very difficult to trace a coherent (and especially an exhaustive and illustrative) picture from the complex interweavings of such ‘threads’, recalling some of these intriguing connections of people, places, publications, and events does offer a more appropriate account of the field of knowledge that we refer to as ‘Italian geography’.

Geographical Threads

It is useful to start by identifying some key figures and groups, together with some topics and areas of interest that have most strongly influenced recent Italian geography and that have, in some cases, also gained currency beyond the national context. From a theoretical point of view, some of the most interesting contributions of recent Italian geographical thought have been focused on: (1) the metaphorical power of geographical representations; (2) the relationship between cartographic reason, bourgeois geography and modernity; (3) the crisis of geographical representation; and (4) the epistemologies of geographical praxis. During the 1980s, very important work was also published on the concepts of regionalization and territorialization, while the past decade has witnessed pioneering research on local development and the urban milieu, with geographers’ seminal contributions gaining wide recognition also on the part of scholars in competing disciplines.

In the last two decades, Franco Farinelli (at the University of Bologna) has literally rewritten the history of geographical thought, providing a new interpretation of the deterministic approaches that dominated the discipline during a great part of the 1900s and highlighting the (largely forgotten) ontological dimension of the Erdkunde project, while also forcefully condemning the ‘cartographic prison’ to which Western thought has been constrained since Anaximander. Farinelli’s theory of cartographic reason has indeed been an extraordinary contribution to the understanding of the ‘deep’ nature of geography and the genealogies of power that have always accompanied it. His latest work – *Geografia. Un’Introduzione ai Modelli del Mondo* (*Geography. An Introduction to the Models of the World*) – is probably one of the most striking books on the philosophy of geography ever written. Farinelli’s elaboration of the relationship between Western thought and the translation/reduction of reality into the two dimensional plane of the map is a pathbreaking contribution not only to Italian geography but to the history of geographical thought more broadly.

In 1985, Torino geographer Giuseppe Dematteis – possibly the best known Italian geographer, both within and outside the discipline – published *Le Metafore della Terra* (*The Metaphors of the Earth*), a book that would change the course of Italian geography. Anticipating the post structuralist turn in Anglo American geography by several years, Dematteis’ book exposed the ways in which the power of geographical representations does not – and cannot – derive from their ‘scientificity’ but is, rather, the product of their metaphorical nature. *Le Metafore della Terra* is, indeed, a critical and sophisticated meditation on the ambiguous nature of geographical thought, calling for an acceptance of the partial truths with which geographers (as all social scientists) must grapple. But unlike the other social sciences, Dematteis would note, geography has as its object of analysis something quite solid, inevitably ‘material’ – the territory. This is why geography, now as always, functions as a precious cognitive interface between the materiality of the world and its diverse representations. Dematteis’ reflections continue to strongly influence Italian geographical debates and are certainly some of the most original contributions that Italy has given to geography more broadly.

Together with his work on the power of geographical metaphors, Dematteis also inaugurated a new season of theoretical and empirical reflection on the concept of urban milieu as a novel way of understanding local and especially urban development, anticipating some of the debates on the same topic later developed in English speaking and French speaking geography. A vibrant reflection has emerged from his work, with a successful and growing team of economic and urban geographers applying Dematteis’ conceptualizations to numerous empirical cases and influencing contemporary debates on the ‘urban’ also outside of academia. The ‘Torino school’ of economic geography is, in fact, probably the only school recognized as such within Italian geography at present.

The third individual/thread whose contribution deserves to be highlighted is Angelo Turco (at the University of L’Aquila), best known for developing a geographically informed theory of complexity. Turco’s is a decisively ‘cultural’ reading of territoriality and territorialization, closely engaged with new theorizations in semiotics and Luhmann’s work on complexity theory. In two of his best known books – *Geografie della Complessità in Africa* (*The Geographies of Complexity in Africa*) and, especially, *Verso Una Teoria Geografica della Complessità* (*Towards a Geographical Theory of Complexity*) – Turco elaborated a complex theoretical and methodological apparatus, tying it with a wider cultural–intellectual project. Turco’s work and that of the extended group of geographers associated with him throughout the 1990s focused especially on processes of territorialization in Africa, with the group’s work leading to the creation of a new journal (*Terra d’Africa*), the organization of countless international and national seminars and conferences, and the publication of numerous essays and articles; it is another one of the few examples in Italy of a clearly identifiable research ‘strand’.

It is also in these same years that Adalberto Vallega (working at the University of Genova) published a series of volumes on the concepts of region and regionalization, grounding his reflection in an original approach to systems theory. His 'systemic geography' and his books on the history and the theory of the region in geography soon became a must read for most junior (and not so junior) human geographers in Italy. What is more, Vallega's geographical interpretation of systems theory, especially as understood through the concept of 'territorial systems' (also adopted but with a different take by geographers working on complexity theory) gave decisive momentum to the theoretical debate in Italian geography during the 1980s. The national research program on 'regions and regionalization' in which both Vallega and Turco participated (along with many other influential Italian and European geographers) represented in many ways the apex of this theoretical reflection, and has been enormously influential since (as has been the edited volume that came out of the project – *Regione e Regionalizzazione* – *Region and Regionalization*, edited by Turco in 1984).

These four 'threads' are perhaps the lines of thought that have most clearly marked Italian geographical debates in the past three decades, in some cases giving life to truly novel understandings of geographical representations. Needless to say, this list is incomplete and certainly risks marginalizing some other very interesting sources of reflection and equally valuable research endeavors; unfortunately, in the attempt to reconstruct the contours of such a fragmented field as contemporary Italian geography, this risk is probably unavoidable. Nonetheless, and despite the lack of a visibly formalized disciplinary debate, the ideas and people mentioned here have been fundamental in opening new and highly influential research directions, something that is by and large accepted by the majority of Italian geographers today.

Key Sites and Networks

Alongside (and together with) these lines of thought, a distinct series of other places, events, and networks have framed Italian geography more broadly in the last few decades. Within this wider research community, we can certainly identify the impact of the above cited theorists and those with whom they have principally dialoged, but also the influence of a variety of other strands of reflection.

One of these informal networks is a group of (predominantly Italian) geographers that have gathered around the initiatives spurred by the so called 'Palermo School' which, under the guidance of Vincenzo Guarrasi, over the past decade or so has been recognized as a key 'cultural geography' research 'cluster' and has organized a series of very successful events/episodes of cutting edge geography. Among the many contributions of this group of researchers has been that of introducing semiotic approaches to the analysis of the relationships between culture and space, often in close collaboration with other local scholars from different disciplines; approaches that would later be adopted by 'new cultural geographers' as well, and which would have an important influence on the work of successive generations of Italian geographers more widely. This group of cultural geographers has been very prolific through the years, organizing a variety of meetings and conferences, and publishing its contributions within a number of edited volumes. The group's most recent initiative, under the guidance of Giulia de Spuches and Vincenzo Guarrasi, has been the project of an *Atlante Virtuale* (*Virtual Atlas*), culminating in an edited collection and an extraordinary itinerant seminar organized in Palermo in 2000 that, through a series of (virtually) interconnected 'events' and 'moments', interrogated the very idea of the 'performance' of geographies 'in place'. Indeed, a whole series of new and alternative gatherings of Italian and continental geographers have been organized in Palermo in recent years under the umbrella of Guarrasi's research project on the 'cosmopolitan city', gatherings that have become a sort of annual appointment for an informal group of geographers, Italian and not only.

This informal network is partially the legacy of a group of (largely) Italian, Swiss, and French geographers that has been meeting for over 20 years to reflect on the nature and the limits of geographical representations. The group has had the intermittent participation not only of the previously cited Italian geographical theorists, but also a number of well known figures such as Denis Cosgrove, Gunnar Olsson, Claude Raffestin, and Ola Söderström to name those whose work has perhaps had the greatest impact on Italian geography and geographers. It is extremely difficult (if not impossible) to map the wider circle of protagonists, encounters, and collaborations linked with these initiatives – or their outcomes: it has been, nonetheless, the most prolific 'network' of Italian and foreign geographical theorists – and undoubtedly the most influential in its effects on Italian human geography. It has been particularly influential, indeed, in providing alternative directions in an intellectual environment that, up until the 1970s, was still marked by a (confused) Vidalian legacy on the one hand, and an almost unconscious determinism on the other, coupled with an imported version of the quantitative revolution.

Others, perhaps more indirect influences, are visible in the 'Venetian' geography of the past two decades. The 1988 workshop on 'language and representation' organized in Venice by Gabriele Zanetto was an important marker in this regard, as was the two volume edited collection *Les Langages des Représentations Géographiques*

(*The Languages of Geographical Representations*) that followed the meeting; a work that remains one of the finest ex amples of disciplinary dialog between Italian and other European geographers. Following the success of this initiative – and thanks, above all, to the leadership of Zanetto – all through the 1990s, Venice continued to engage with international geographical debates in a variety of ways. For instance, Fabio Lando's 1993 edited collection *Fatto e Finzione* (*Fact and Fiction*), a beautifully crafted anthology of Anglophone humanistic geography, introduced some of the most innovative theories and writings in this tradition to Italian geographers. Lando's book paved the way to new research on geography and literature, inspiring a number of works on this topic. Its most important contribution has been, however, that of introducing an entire body of foreign geographical re search to Italian geography and geographers. More re cently, Venice has also been the locus of reflection on postmodern and post structuralist geography, thanks to the contributions of a younger generation of geographers strongly engaged with Anglo American debates.

Another vital site for dialog among Italian human geographers has always been Milano. Of particular im portance here has been the human geography series published by Unicopli and directed by Giacomo Corna Pellegrini. The series introduced to the Italian market, among other things, the writings of the American be haviorist school, and published numerous monographs as well as translations of important works of human, polit ical, and cultural geography. Milano was also the focus of a number of other publishing initiatives at the end of the 1980s and beginning of the 1990s, key among which two volumes dedicated to landscape research in geography by Maria Clara Zerbi, marking the return of Italian geog raphers to work on this topic. In recent years, the Istituto Orientale in Napoli has also become an important site in the production of cutting edge geographies. Under the leadership of Pasquale Coppola, a brilliant team of young geographers has begun to theorize in truly innovative ways the cultural, political, and urban geographies of the Mediterranean and the Italian South (as well as Latin America). The widespread international networks linked to this group and to what is probably the most successful doctoral program in geography in the country (formally, a degree in development geography) have made of Napoli a very prolific site of geographical production and the intellectual home of a by now quite conspicuous number of young scholars whose work is widely appre ciated both in other Italian universities and abroad.

Finally, we should note the important contribution made by Italian geographers to Marxist geography, in particular by Massimo Quaini who, following in part in the footsteps of historical geographer Lucio Gambi, published what is probably to date the only Italian geo graphical monograph translated into English: *Geografia e Marxismo* (in the English edition: *Marxism and Geography*). Quaini's writings have also focused on landscape, the history of geographical thought, and cartography, all very well established themes in the Italian geographical lit erature, approaching them with a fascinating philological touch, as in his most recent *L'ombra del Paesaggio* (*The Landscape's Shadow*) and *Il Mito di Atlante. Storia della Cartografia Occidentale in età Moderna* (*The Myth of Atlas. A History of Western Cartography in the Modern Age*).

Geographical Societies and Journals

Beyond these initiatives and events that mark some key foci of theoretical reflection in contemporary Italian geography, the role of several institutional bodies rep resenting the discipline should also be recognized. These bodies have been particularly important not only because they constitute real, institutional places that host key events and are the gathering points for geographers (something that is true for any national geographical association), but also because their journals act as the official and more visible fora of an 'Italian geography', although these are not as influential as they could (and probably should) be. Their conferences and workshops are key sites for the production of the identity of the discipline and, in many ways, represent the interests of geography in the governmental bodies that determine the funding and the direction of research and teaching in Italy.

The *Associazione di Geografi Italiani* – AGeI (the Asso ciation of Italian (Academic) Geographers) plays a key role in representing the discipline and in organizing national research groups and key conferences. In par ticular, the AGeI is charged with organizing the annual '*Giornate della Geografia'* (*Geography Days*) and the Congress of Italian Geographers, held every 4 years. The AGeI also publishes a relatively new journal – *Geotema* – with a series of special issues edited by different members of the association.

The *Società di Studi Geografici* (the Society for Geo graphical Studies), founded in 1895 as the *Società di Studi Geografici e Coloniali* and based in Firenze, has not only always been a prestigious home for seminars and work shops, but also publishes the *Rivista Geografica Italiana*, the only peer refereed journal in the field at present, and a journal that has played a very influential role in the shaping of the discipline in years past.

Another key geographical society is the *Società Geo grafica Italiana* (the Italian Geographical Society), foun ded in Firenze in 1867 and since 1872 located in Rome in the prestigious Villa Celimontana, an important symbolic space for Italian geographers. In many ways, the *Società Geografica Italiana* plays a role similar to that played by sister institutions in Britain and other European countries. The *Società*'s Roman location is the site of the most complete geographical library in the country and every year hosts a number of important public gatherings and initiatives. It publishes the *Bollettino della Società Geografica Italiana*, the other key 'historical' journal of Italian geography, with contributions from both geographers and other scholars interested in 'geographical themes'. The *Società* has also recently been quite influential in diffusing and connecting the work of geographers to the concerns of governmental institutions and, more generally, to the intellectual life of the capital, and it has sponsored a series of new editorial initiatives. The role and international prestige of Villa Celimontana further increased since 2000, when it was nominated as the location of the Home of Geography, hosting the archives and study center of the International Geographical Union (IGU). The director of the Home of Geography is currently Italian geographer Giuliano Bellezza. Italian geographers have, indeed, been active members of the IGU, with a number of Italian scholars holding the leadership of various commissions over the past decades. The late Adalberto Vallega was elected IGU President in 2004, an important recognition not only of Vallega's own work but also of the contribution of Italian geography more broadly.

Conclusion

As outlined in the above paragraphs, the absence of a clear mapping of Italian geographical production reflects the status of a discipline which might appear somewhat disarticulated and relatively isolated from international theoretical debates – and yet capable of producing, at times and within particular contexts, works of brilliance which would probably deserve better recognition by the geographical academy. Equally, the important contribution given by leading Italian geographers to the IGU and other key international networks marks an impact that conventional disciplinary mappings often overlook.

See also: Regionalization/Zoning Systems.

Further Reading

Atkinson, D. (2000). Geopolitical imaginations in modern Italy. In Doods, K. & Atkinson, D. (eds.) *Geopolitical Traditions: A Century of Geopolitical Thought*. London: Routledge.

Corna Pellegrini, G. (ed.) (1987). *Aspetti e Problemi della Geografia*. Milan: Marzorati.

Corna Pellegrini, G. and Brusa, C. (eds.) (1990). *La Ricerca Geografica in Italia*. Varese: Ask Edizioni.

Dematteis, G. (1985). *Le Metafore della Terra*. Milan: Feltrinelli.

de Spuches, G. and Guarrasi, V. (eds.) (2002). *Atlante Virtuale*. Palermo: Università degli Studi di Palermo.

Farinelli, F. (1992). *I segni della Terra*. Rome: La Nuova Italia Scientifica.

Farinelli, F. (2003). *Geografia. Un'Introduzione ai Modelli del Mondo*. Turin: Einaudi.

Lando, F. (ed.) (1993). *Fatto e Finzione. Geografia e Letteratura*. Milan: Etaslibri.

Minca, C. (ed.) (2001). *Introduzione alla Geografia Postmoderna*. Padua: Cedam.

Quaini, M. (2006). *Il Mito di Atlante*. Reggio Emilia: Diabasis.

Turco, A. (ed.) (1984). *Regione e Regionalizzazione*. Milan: Franco Angeli.

Turco, A. (1988). *Verso una Teoria Geografica della Complessità*. Milan: Unicopli.

Vallega, A. (1976). *Regione e Territorio*. Milan: Mursia.

Vallega, A. (1982). *Compendio di Geografia Regionale*. Milan: Mursia.

Zanetto, G. (ed.) (1989). *Les Langages des Représentations Géographiques*, 2 vols. Venice: Università di Venezia.

Zerbi, M. C. (1993). *I Paesaggi della Geografia*. Turin: Giappichelli.

Ivy League and Geography in the US

R. Wright and N. Koch, Dartmouth College, Hanover, NH, USA

Glossary

Ivy League Eight prestigious, private universities in the northeastern United States.

The Ivy League formed in 1954, but the study of geography at these eight Northeastern American colleges began much earlier. Although explicitly a sports conference, the Ivy League comprises some of the oldest and most prestigious academic institutions in the United States. Seven members date to the colonial period: Harvard (est. 1636), Yale (est. 1701), University of Pennsylvania (est. 1740), Princeton (est. 1746), Columbia (est. 1754), Brown (est. 1764), and Dartmouth (est. 1769). The eighth, Cornell, was founded in 1865. Each Colonial institution provided instruction in geography in the early days and most subsequently established undergraduate or graduate programs at one time or another – some more than once. Although many Ivies have geographers on their faculties or retain some aspects of the field in the guise of programs such as Urban Planning or Development Studies, Dartmouth is the only one with a Department of Geography. Geography's fate at these institutions has hinged on the activities of individual professors, the attitudes of key administrators, trends within the discipline, and events outside the field itself.

Pre-1830: From University Subject to School Subject

Mirroring instruction at Oxford and Cambridge at the time, the curricula of the seven colonial Ivy Leagues, included geography. Departments as we now know them did not exist and the study of geography (commonly referred to as 'the use of the globes' or 'mathematical geography') was part and parcel of teaching history, mathematics, and astronomy. Globes ('one celestial and one terrestrial') were central tools of instruction. A number of key texts complemented the study of the globes. Just as Isaac Newton did at Cambridge in the 1600s, Columbia, Dartmouth, Harvard, Pennsylvania, Yale used Varenius' *Geographia* (1650) to teach mathematical geography throughout the 1700s. In the mid to late 1700s, Brown and Princeton used Martin's *Use of Both the Globes*, and Gordon's *Geographical Grammar* (1719) held pride of place at Harvard and Princeton. Brown, Dartmouth, Harvard, and Yale incorporated both Guthrie's *Geographical Grammar* (1770) and Morse's *Universal American Geography* (1797) in their curricula of the late eighteenth century. (Grammar references a text organized by nation, as opposed to alphabetically – a gazetteer.) Like their university counterparts in Britain, the Ivies had few instructors dedicated explicitly to the subject; geography was typically embedded in the general study of natural philosophy or mathematics. Nevertheless, in 1784, arguably the first American professional geographer, John Daniel Gross, became a professor at Columbia.

Although the Ivy Leagues included geography in their early curricula, its presence began to wane in the early 1800s. The exact reasons for this remain unclear; perhaps 'mathematical geography' became more deeply ensconced in the intellectual areas we now know as mathematics, physics, and astronomy and a consensus emerged that the remainder – the grammar and gazetteer geography – was best taught in schools rather than at university. The decision to drop geography came first in 1816 at Harvard. That university had included geography on admissions exams since 1803, and continued to do so through most of the nineteenth century. In 1819, Dartmouth followed suit and eliminated the subject from a new curriculum. Both Columbia and Yale dropped instruction in favor of an entrance exam in 1825, and Brown and Princeton did the same soon after. By 1830, the University of Pennsylvania stood as the only college still offering geography and, although it began requiring an admissions exam in this year, it also began a new program to intensify its instruction in geography.

1830–1900: Slow Revivals

The elimination of geography from the curricula dealt a blow to its development and institutionalization in higher education. Moreover, its relegation to the status of 'school subject' may have seeded the perception of geography as an idiographic and nonintellectual subject. In the nineteenth century, universities evolved away from the notion that students had to master a certain knowledge toward the idea that students could have more choice in what they learned. As the course offerings at elite universities broadened and included electives, the Swiss evangelical environmental determinist, Arnold Guyot, sparked the interest of northeastern intellectuals when he gave a series of lectures in the US in the late 1840s (published in 1849 as *Earth and Man*). Ira Young, a professor of natural philosophy at Dartmouth, immediately picked up on these lectures and the associated monograph and in 1850 offered a course based on that text. In 1854, Guyot joined Princeton's faculty, eventually chairing a Geography Department there until 1880. Guyot's crude and theologically informed linking of nature and human activity, for better or for worse, helped leverage geography as an academic discipline, and the use of his text diffused from Princeton and Dartmouth to some other Ivies.

Darwin's theories of evolution stimulated new interest in the environment, furthering geography's recovery as a university subject in the Ivy League. Geography coursework at Dartmouth, Pennsylvania, and Princeton became available, moving away from the early accents on mathematical geography and navigation toward understanding 'man–land' relationships. Harvard formally offered instruction in geography again soon after the Civil War and Yale reintroduced geography with Daniel Gilman's appointment as Professor of Physical and Political Geography. Gilman's appointment, it may be worth noting, occurred two decades before Halford Mackinder's selection as Reader in Geography at Oxford. An active scholar and teacher, Gilman left to take up the presidency of the University of California in 1872 and later headed Johns Hopkins University. His employment at Yale, nevertheless, led to growth in geography at that institution. Francis Walker (Professor of political economy and history) succeeded Gilman and taught regional and statistical geography. William Brewer soon joined Walker on the faculty. A polymath with a specific interest in fluvial geomorphology, Brewer worked on sedimentation.

The emphasis at Harvard in the late nineteenth century also inclined toward physical geography, a trend cemented in 1878 with the appointment of William Morris Davis as an instructor in geology. Davis became the most influential geographer of his time, building the program at Harvard, shifting the center of inquiry in US Geography to theories of physiography and causation, and more generally championing the discipline. In 1885, he oversaw geography's move from the Department of Natural History to the newly created Department of Geography and Geology.

As Harvard's program grew, the University of Pennsylvania's Wharton School of Business developed a rival approach to that of Davis' physiography. Wharton scholars under Emory Johnson focused on economic geography, as the study of human use of natural resources. With the emergence of the social sciences as fields of study in the late nineteenth and early twentieth century, the emphasis on physiography at certain institutions complicated geography's intellectual identity. Some programs, however, maneuvered between physical and human geography. At Dartmouth College, for example, President William Jewett Tucker pushed for the development of the social sciences throughout the 1890s and when the Department of Social Science split into Sociology and Economics, these new entities both sponsored instruction in geography. In addition, Dartmouth's new Tuck School of Business Administration offered geography courses from the start. Brown revived geography around the same time with modest offerings in physical geography. Cornell's curriculum included geography for the first time in the 1890s; both economic and physical geography were offered. At Yale in the late 1890s, Herbert Gregory began strengthening the geography wing of the geology department. Columbia made its Geography Department independent in 1896.

1900–20: The Heyday of Ivy League Geography

By the turn of the century, geography was part of the curricula of all eight Ivy League universities. This period may have been, in some senses, the heyday of Ivy League Geography. The Ivies, and especially Harvard, produced the first generation of professional geographers in the United States. One of its graduates, Wallace Atwood, took a position at Harvard and subsequently moved to Clark, establishing the Graduate School of Geography at that institution. Clark now boasts the oldest sustained program in the US and one that has generated more PhDs in geography than any other. Another Davis student, Herbert Gregory, hired half a dozen geographers at Yale between 1900 and 1910.

The intellectual core of the discipline at this time largely centered on Davisian physical geography. Davis himself came to recognize, however, that the field was broader than this. He tried to get practitioners, especially his mentees, to develop a set of subfields like climatology, anthropogeography, and commercial geography under the rubric of what he called 'ontography'. Some of Davis' students, Isaiah Bowman (PhD Yale and on Yale's faculty 1905–15) and Ellsworth Huntington (also Yale), helped provide meaning for this new term, retaining a connection to Darwinian ideas of natural selection and evolution but assessing human activity in the context of the physical environment. Lurking in all this was, of course, environmental determinism. With roots tracing back to Guyot and some strong tendrils in eugenics, determinism's dark gravitational force continued to drag on the discipline.

The loss of a number of important intellectual leaders at Harvard, Cornell, and Yale checked much of the progress of the previous couple of decades. In 1906, Harvard physical geographer Nathaniel Shaler died and 6 years later, Davis retired from Harvard. His departure was a huge loss; none of his successors could match his energy and vision. With Atwood leaving in 1920, Harvard's program in geography was unquestionably weaker than 20 years earlier. In this period, Cornell's Ralph S. Tarr, a student of Shaler and Davis, died unexpectedly.

He had developed a burgeoning geography program through the Geology Department. Tarr oversaw Cornell's development as a center for research and graduate education, and eventually headed the new Department of Physical Geography in 1906. At Princeton, William Libbey, Guyot's successor, left the university's geography program in a reduced state. In 1915, Herbert Gregory, the champion of geography at Yale, became seriously ill and was forced to relinquish his chairmanship. In the same year, Isaiah Bowman – the Harvard trained scholar diplomat – became the new director of the American Geographical Society. The geography wing of the Geology Department was subsequently terminated. The University of Pennsylvania also suffered a major loss in 1919 when J. Russell Smith left for Columbia. A dedicated organizer, Smith oversaw the unification of geography and industry as a separate subdivision of the Department of Economics and Social Science at Wharton. He was recruited by Columbia's President to organize economic geography at the University's new School of Business, as he had done at Wharton. Three other established scholars also soon left. Given Wharton's previous loss of Walter Tower to the University of Chicago, Pennsylvania geography was substantially diminished. Brown's geography instruction terminated around 1914, leaving only Columbia and Dartmouth well positioned in the inter war period. All this downsizing occurred as many English, French, and German universities institutionalized geography as a discipline, associated with both a new phase of empire building and pedagogic needs (for high school geography teachers). The absence of formal empire probably contributed to the erosion of the place for geography in the US during this period.

1918–45: Theory and Practice

World War I produced new attitudes about geography as a discipline. The large scale employment of professional geographers by government agencies led to a push for the development of the technical aspect of the field over the theoretical. At the same time, its major theoretical foundation – environmental determinism – came under increasing scrutiny. Determinism had little international support as well as a skeptical domestic audience, yet some influential scholars, notably Ellen Semple (Chicago) and Huntington, continued to champion it. Although the late 1920s were marked by a new emphasis on the region, the regional concept failed to provide the unifying framework geography needed to maintain its prestige in the most elite institutions of higher learning. Historians of the discipline note that it was not intellectually strong in the first half of the century. 'Amateur' and 'mediocre' are two of the most damaging terms used to describe the scholarly state of affairs by both commentators at the time and by those writing histories. At elite, influential universities, such frailty would be a particular problem.

Increased demand for practical education and the attack on environmental determinism was devastating for geography in the Ivy League universities; these institutions increasingly emphasized the importance of theory and held technical instruction in low esteem. In contrast, geography departments in the new land grant colleges of the Midwest prospered in this environment. Because these universities were designed in part to support to the Midwestern agricultural economy and serve the broader public, they welcomed the applied elements of geography. Thus, the considerable expansion of geography in the inter war period occurred most significantly in the state universities and comparatively less so in the elite colleges of the Northeast. Ivy League geography programs – which were already vulnerable from major pre war losses that exposed their still unsuccessful institutionalization – began to lose ground to Midwestern geography departments. Ironically, the earliest geography faculties at many land grant universities were drawn largely from students trained at the Ivy Leagues, under the likes of Davis, Gregory, and Tarr.

Despite these shifts, Dartmouth and Columbia, the two institutions that had not suffered major setbacks in the pre war era, fared relatively well during this period. At Columbia, J. Russell Smith managed the geography program and the university averaged a PhD per year in geography in the inter war period. Dartmouth, which has historically focused on undergraduate education in the Arts and Sciences and has never formally offered graduate coursework in geography, continued to develop a program for undergraduates. Dartmouth's administrative leadership looked upon geography with favor, and in 1942 the discipline achieved department status.

1945–Present: And Then There Was One

World War II, like World War I, stimulated institutional interest in geography in the American academy. In the Ivy League, however, the picture was much more mixed. By the 1980s, every Ivy League program with the exception of Dartmouth's had been disbanded. The reasons behind these terminations vary around themes of weak faculty and the discipline's uncertain intellectual terrain. The adverse fiscal context faced by institutions in the aftermath of World War II probably made things worse. Although geographers can be counted among the faculty at Brown and Princeton (notably Julian Wolpert – the Bryant Professor of Geography, Public Affairs, and Urban Planning at the Princeton's Woodrow Wilson School of Public and International Affairs), these two institutions never developed programs in geography to the same extent as their sister schools. The rest of the essay describes events at the remaining six universities.

The closure of Harvard's program in 1948 dealt a severe blow to Ivy League geography and to US geography more generally. Harvard's program had remained modest in the 1920s. Still part of the Department of Geology and Geography, geography had nevertheless begun to show signs of life. Kirk Bryan was appointed in 1926 to bolster physical geography; 2 years later, French geographer Raoul Blanchard and Derwent Whittlesley joined the faculty. These appointments were designed to build up the human side of the program and steer geography toward being a separate field. These appointments appear to have affected PhD production; Harvard granted two PhDs in geography in the 1920s (including the only one to a woman – Millicent Todd Bingham) whereas eight were awarded in the 1930s. The appointments of Edward Ullman (PhD Chicago) and Edward Ackerman (PhD Harvard) in the 1940s augured a new day for geography at Harvard – built on human geography. In May 1947, however, the *ad hoc* promotion committee agreed to promote Edward Ackerman to Associate Professor; conflict ensued when geologist Marland Billings protested the vote. Billings appealed to Provost Paul Buck, asserting Geology's need for the half position they would be losing by Ackerman's full appointment in geography as well as casting doubt on the intellectual merits of human geography as a field.

That year, an *Ad Hoc* Committee on Geography was formed, and included the President of Johns Hopkins, Isaiah Bowman. Bowman knew the President of Harvard well, and thus played an influential role in the decisions regarding the future of geography. Bowman had little respect for Harvard's geography program as it tended more toward the human side of the field and away from the physiography he enjoyed as an undergraduate at Harvard. Whittlesley, the only tenured human geographer, proved to be a poor advocate for the discipline; Edward Ullman did much more in the program's defense. Yet Bowman's lack of support probably sealed the program's fate. His inaction was perhaps a favor to President Conant, who likely already had geography's elimination as a set objective, given budget stresses at the university at the time and his personal disdain for the discipline. While protests from numerous geographers ensued, the Committee used selective evidence from Carl Sauer, who wrote the chairman, saying that the field of geography "is very poorly defined" and that geography has often been done best by nongeographers. Ackerman and Ullman were fired; Richard Logan, an instructor teaching introductory courses on a term appointment, was not rehired. Whittesley was retained as the sole tenured geographer; Ullman departed Harvard for the University of Washington in 1951 to join William Garrison and some other (quantitative) revolutionaries in Smith Hall (*see* Quantitative Revolution). Ironically, 15 years later when Harvard became the home for the Laboratory for Computer Graphics and Spatial Analysis (LCGSA), University of Washington graduates Waldo Tobler and Brian Berry figured among the first scholars to give papers at the new facility. Berry would return (1976–81) as an endowed Chair and Professor of City and Regional Planning and Director of the LCGSA. This laboratory developed SYMAP, the early computer mapping program and precursor to geographic information science (*see* Bowman, I.; Garrison, W.).

While many commentators recall Jean Gottmann's declaration that the closure of this department was "a terrible blow … to American geography," the fallout from the demise of geography at Harvard was not immediate. One year after the events at Harvard, Yale reinstated Geography to departmental status. After the collapse of Yale's geography program in 1915, Ellsworth Huntington had returned as a research associate and the only geographer in the Geology Department. A few dissertations were completed at this time, and almost all of the doctoral work completed in the 1930s was in physical geography. In 1945, however, Stephen Jones, hired originally in International Studies, was appointed associate professor of geography. Although geography gained department status in 1949, and had hired several more geographers on faculty, Jones' efforts to fully institutionalize the subject were unsuccessful. The geographers he employed were not productive scholars and proved incapable of maintaining a coherent independent identity for the program in the face of competition for resources – especially from area studies. The department at Yale awarded one PhD in the 1950s and another one in the 1960s. In addition, Time Magazine's 1963 article on 'easy A's' highlighted the 'gut' geography courses at Yale mocking the political geography course that had no required reading. Yale terminated geography for the last time in 1967.

At the University of Pennsylvania, geography moved along an alternative trajectory with geography being organized through the university's business school, Wharton. Although the inter war period had been marked by the growth of business administration with an emphasis on economic geography, the trend was reversed in the post–World War II years. Declining enrollments in geography resulted in a shift in business programs away from economic geography. In the mid 1950s, Walter Isard moved from MIT to join Penn's economics faculty, with the mandate to establish a regional science PhD program. Within 2 years, Isard was chairing a separate Department of Regional Science, which had links to geography. Michael Dacey, Duane Marble, Julian Wolpert, and Allen Scott were all faculty members at one time in Penn's Department of Regional Science and Regional Science at Penn has produced scholars who now occupy positions in geography. In 1963, the University of Pennsylvania closed its Geography Department and shifted some faculty positions to Regional Science. Isard moved to Cornell in 1979; 14 years later, Penn's Regional Science department closed – having granted about 140 PhDs since its inception. Cornell's City and Regional Planning program contains the only Regional Science track in the Ivies. In the past two decades, Cornell has conferred 31 PhDs in regional science; some of these graduates have also obtained academic positions in geography.

Columbia University's excellence in graduate studies in geography went unmatched by the other Ivy Leagues for much of the twentieth century, but this did not prevent the department's eventual termination. After early success in establishing geography at Columbia, Russell Smith did not do enough to strengthen the program and consolidate its position at the university: he spent little time on campus and dedicated most of his efforts to writing textbooks. In later years, Columbia's graduate enrollment numbers were reasonable, but the best and most promising students were opting instead to study at expanding departments at Michigan, Wisconsin, or Chicago. With roots in Columbia's School of Business, the department also suffered from the decline of economic geography in the post war period. Smith's retirement during this time was followed by the eradication of two tenured positions in the 1970s, appointments filled by junior colleagues on 3 year contracts. Another loss came when the department's chairman, William Hance, retired in 1981. Although graduate enrollments were still around a dozen each year in the early 1980s, the administration claimed that the quality of the students had declined. Undergraduate course requirements in geography were eliminated in 1985 and new courses in remote sensing were not permitted. Given the lack of popularity of geography courses outside the department, administrators felt that it was not fulfilling either its intellectual or its service function. The decision to close the department finally came in 1986.

The establishment of a full blown Department of Geography at Dartmouth coincided with the appointment of a third professor of geography Van English; he joined Al Carlson and Trevor Lloyd on the faculty. All three held Clark PhDs. The curricula offerings in the 1940s reflected the times with a War Course as well as map interpretation as part of Defense Instruction joining the more regular fare of regional and physical geography. With the end of global conflict, war related courses defaulted to more conventional rubrics of air photo interpretation and cartography. Robert Huke (PhD Syracuse) joined the faculty in 1953 and championed the discipline on campus. He was a beloved teacher and an energetic promoter of geography among students and faculty.

The department slowly grew to include six full time lines by the mid 1970s. Laura Conkey became the first woman appointed to a tenure track position in the department and subsequently the first woman tenured in a Geography Department in the Ivy League. Between 2000 and 2005, the department added three new positions. Today the department includes nine full time tenure track faculty; two hold joint appointments (with Women's and Gender Studies and Environmental Studies respectively), and two others now hold endowed chairs – the first such appointments conferred on geographers at Dartmouth in its history. In 2007, women in tenure track appointments in the department outnumbered men for the first time.

Why has this Ivy League department not only survived but also flourished when the others foundered? Without a doubt, Bob Huke's personality and tireless efforts on behalf of geography helped trump doubts about the department's viability and place in a liberal arts environment. The department always cared deeply about undergraduate teaching and developed consistently sound undergraduate enrollments. The department has also hired carefully, emphasizing teaching and research excellence as well as collegiality. In the last couple of decades, the department has only put forward strong candidates for tenure and promotion whose work resonates across campus and the discipline. Geography at Dartmouth has been opportunistic, open to moving in new directions when the chance arose. In addition, for decades, department chairs have steered a course that promotes the unique contribution of geography as a theoretically oriented discipline, and marking it as distinct from potentially competing campus programs in environmental studies, various area studies, and its sister social science departments – Anthropology and Sociology. The department also values and celebrates research collaborations between faculty and undergraduates – this article being one example among many.

The Geography Department at Dartmouth recently developed *ad hoc* postbaccalaureate and postdoctoral programs, which the faculty plan to leverage into a masters degree course sooner rather than later. Hopes for new Ivy League graduate programs in geography, however, do not rest entirely with Dartmouth. In 2003, Brown's initiative in spatial structures in the social sciences came on line; this interdisciplinary research cluster analyzes the impacts of spatial relations and contextual effects on social problems. Harvard's interdisciplinary Center for Geographic Analysis, opened in May 2006, also augments geography's (Ivy League) standing. Whether these new research centers seed new departments at Harvard and Brown or new programs in geography elsewhere in the Ivy League remains to be seen.

See also: Anglo-American/Anglophone Hegemony; Feminist Geography, Prehistory of.

Further Reading

DeBres, K. (1989). An early frost: Geography at Teachers College, Columbia and Columbia University, 1896 1942. *The Geographical Journal* 155(3), 392 402.
Fellmann, J. D. (1986). Myth and reality in the origin of American economic geography. *Annals of the Association of American Geographers* 76(3), 313 330.
Glick, T. F. (1988). Before the revolution: Edward Ullman and the crisis of geography at Harvard, 1949 1950. In Harmon, J. E. & Rickard, T. J. (eds.) *Geography in New England*, pp 49 62. New Britain, CT: New England St. Lawrence Valley Geographical Society.
Koelsch, W. A. (2001). Academic geography, American style: An institutional perspective. In Dunbar, G. S. (ed.) *Geography: Discipline, Profession and Subject Since 1870: An International Survey*, pp 245 279. Norwell, MA: Kluwer Academic Publishers.
Martin, G. J. (1988). Geography, geographers and Yale University, *c*. 1770 1970. In Harmon, J. E. & Rickard, T. J. (eds.) *Geography in New England*, pp 2 9. New Britain, CT: New England/St. Lawrence Valley Geographical Society.
Martin, G. J. (1998). The emergence and development of geographic thought in New England. *Economic Geography* 74, 1 13.
Simpson, R. E. and Huke, R. E. (1988). One view of geography at Dartmouth College. In Harmon, J. E. & Rickard, T. J. (eds.) *Geography in New England*, pp 10 24. New Britain, CT: New England/St. Lawrence Valley Geographical Society.
Smith, N. (1987). Academic war over the field of geography: The elimination of geography at Harvard, 1947 1951. *Annals of the Association of American Geographers* 77(2), 155 172.
Spencer, J. E. (1974). The evolution of the discipline of geography in the twentieth century. *Geographical Perspectives* 33, 20 35.
Warntz, W. (1964). *Geography Now and Then*. New York: American Geographical Society.
Withers, C. W. J. and Mayhew, R. J. (2002). Rethinking 'disciplinary' history: Geography in British universities, c. 1580 1887. *Transactions of the Institute of British Geographers* 27(1), 11 29.

Relevant Websites

http://www.s4.brown.edu
Brown University Spatial Structures in the Social Sciences.
http://www.dartmouth.edu
Department of Geography, Dartmouth College.
http://www.gis.harvard.edu
Harvard University Center for Geographic Analysis.

J

Jackson, P.

C. Dwyer, University College London, London, UK

Peter Jackson (**Figure 1**) has been a leading figure in social and cultural geography, playing a particularly prominent role in the emergence of a 'new cultural geography' in the late 1980s. He forged new directions for work on the social geographies of race and racism and masculinity and, through his research on geographies of consumption, has been central to the reorientation of economic geography. Jackson's theoretical contributions revolve around his interest in the geography of identities, conceived as the sites of cultural and material struggles, and produced by social discourses and practices.

Figure 1 Professor Peter Jackson. Photo courtesy: Graham Allsopp.

Jackson did his first degree and doctorate at the University of Oxford. He was a lecturer and senior lecturer in geography at University College London from 1980–93. He moved to take up a chair in human geography at University of Sheffield in 1993 and was appointed as director of research in social sciences there in 2004. He has been on the Council of the Royal Geographical Society (with the Institute of British Geographers) since 2003 and was elected to the Academy of Social Sciences in 2001.

Jackson's doctoral research on the social geography of Puerto Ricans in New York, completed in 1980, led him to question the dominant focus of social geography at that time, concerned as it was with mapping segregation and measuring spatial 'integration'. Writing in collaboration with Susan Smith, Jackson challenged a social geography of ethnicity framed by the biological metaphors of the Chicago School of sociology, arguing instead for a theory which recognized ideologies of racism and challenged biological understandings of 'race'. As such, ethnicity was understood as a product of social relations, as contingent and negotiated rather than essential. Galvanized by the contested racial politics of Britain in the early 1980s Jackson mapped an agenda for a social geography which prioritized understanding the geographies of racism and the dynamic cultural politics of race. At the heart of this work was a theorization of the social construction and contestation of 'race', and the making of racism and ethnic identities through specific histories and geographies.

Jackson's empirical work ranged from studies of representations of 'race' in the media and advertising, discussions of studies of the racialization of labor in Britain and Canada, and the contested ethnic politics of carnival to a critical engagement with constructions of 'whiteness' in the geographical imagination. This mapped a theoretical agenda for research on geographies of race and racism, which proved highly influential. Jackson's ongoing commitment to the production of antiracist geographies was marked by collaborations with school teachers and a focus on the teaching of geography through his role as a founder member of the Equal Opportunities Working Group at the Institute of British Geographers.

Jackson's work on the geographies of race and racism not only challenged the dominant focus of social geography in the 1980s but, through its re theorizing of the concept of ethnicity as a dynamic social process, critiqued the theories which underpinned contemporary cultural geography. Jackson argued that cultural geography needs to move beyond a narrow, largely North American focus on the study of the visible landscape and urged an engagement with social geography and an analysis of social interaction and social relations. He took inspiration from the burgeoning work of the new cultural studies in Britain, particularly the writings of Raymond Williams, Stuart Hall, and research on subcultures pioneered by the Birmingham Centre for Contemporary Cultural Studies (CCCS). This call for a 'new' cultural geography was first presented in a co authored paper with Denis Cosgrove in 1987, which argued that geographers should expand their field of study to consider popular subcultural forms in their 'specific material contexts'.

In 1989 Peter Jackson published a major book *Maps of Meaning*, which set out a new agenda for cultural geography. Central to the argument of *Maps of Meaning* was a shift in focus from culture, *per se*, to 'the domain of cultural politics'. As Jackson argued in his introduction to the volume:

> the approach adopted here… rejects a unitary view of culture as the artistic and intellectual product of an elite, asserting the value of popular culture both in its own terms and as an implicit challenge to dominant values. Culture emerges as a terrain in which economic and political contradictions are contested and resolved. (Jackson, 1989: p. 1)

In *Maps of Meaning*, Jackson outlines a geographical 'cultural materialism', drawing on the work of Raymond Williams and his understanding of culture as materially produced within specific historical (and geographical) contexts. His arguments are made both through an analysis of the concept of ideology and discourse and exemplified through chapters which map out a new agenda for cultural geography which encompasses gender, sexuality, race and racism, as well as language and popular culture. The book also, if implicitly, produced an expanded methodological repertoire for geographers, emphasizing ethnography as a means to access the contested politics of culture. *Maps of Meaning* provoked important debates about the nature of the cultural in geography, and effectively signaled the beginning of a broader 'cultural turn' in human geography. In the context of this, now widely pervasive, engagement with the cultural in human geography, it is worth recognizing the insistence with which Jackson understood the cultural as social and political. As he emphasized in the conclusion of *Maps of Meaning*:

> It is an impoverished view of culture that stresses text, sign and discourse to the exclusion of contexts, action and structure. Meanings must always be related to the material world from which they derive. (Jackson, 1989: p. 185)

In *Maps of Meaning*, Jackson's discussion of geographies of gender emphasizes the social relations of gender and he urges geographers to focus on "relations between dominant and subordinate groups" (Jackson, 1989: p. 180). His own subsequent empirical work focused on the geographical constitution of masculinity as well as the spatial constitution of 'whiteness'. Work on masculinities and consumption led Jackson to develop a broader program of research on geographies of consumption. This research has focused on practices of shopping and local consumption cultures, on transnational consumption cultures, and on food consumption and commodity chains. This work has been characterized by a multidisciplinary focus and involved collaborations with both anthropologists and sociologists.

Jackson's work on consumption cultures reiterates his interests in both the making of identities and the significance of popular culture. Thus, consumption practices are a means to interrogate how individuals make meaning through materials of the everyday. This work is also an intervention in economic geography which goes beyond the argument that the economy is embedded within cultural framings or that the cultural must be part of any economic explanation to argue for a exploration of the 'contested geographies of contemporary commercial culture'. This approach challenges the opposition between 'the cultural' and 'the economic' through an exploration of the complex circuits and networks within which commodities are embedded. While some critics have argued that Jackson's emphasis on consumption and identity downplays the power relations inherent within geographies of production, Jackson remains insistent on a focus on the appropriation and transformation of the materials of everyday social and cultural life, "the maps of meaning through which the world is made intelligible" (Jackson, 1989: p. 2) as political. Recent writing makes these links explicit as Jackson argues for "empirically grounded studies of the impact of 'globalization' on consumer cultures in different geographical contexts" (Jackson, 2004: p. 166) and the 'rematerialization' of social and cultural geography. With this writing Jackson remains at the heart of debate in social and cultural geography, provoking discussion about the nature of materiality and the conception of commodities as well as raising important questions about writing critical geographies.

See also: Commodity Chains; Consumption; Cosgrove, D.; Cultural Geography; Culture; Ethnicity; Race; Racism and Antiracism; Social Geography; Whiteness.

Further Reading

Cosgrove, D. and Jackson, P. (1987). New directions in cultural geography. *Area* 19, 95 101.
Jackson, P. (ed.) (1987). *Race and Racism: Essays in Social Geography*. London: Allen and Unwin.
Jackson, P. (1989). *Maps of Meaning*. London: Unwin Hyman.
Jackson, P. (1991). The cultural politics of masculinity: Towards a social geography. *Transactions, Institute of British Geographers* 16, 199 213.
Jackson, P. (1998). Constructions of whiteness in the geographical imagination. *Area* 30, 99 106.
Jackson, P. (1999). Commodity cultures: The traffic in things. *Transactions, Institute of British Geographers* 24, 95 108.
Jackson, P. (2000). Rematerializing social and cultural geography. *Social and Cultural Geography* 1, 9 14.
Jackson, P. (2002). Commercial cultures: Transcending the cultural and the economic. *Progress in Human Geography* 26, 3 18.
Jackson, P. and Penrose, J. (eds.) (1993). *Constructions of Race, Place and Nation*. London: UCL Press and University of Minnesota Press.
Jackson, P. and Smith, S. (eds.) (1981). *Social Interaction and Ethnic Segregation*. London: Academic Press.
Jackson, P. and Smith, S. (1984). *Exploring Social Geography*. London: Allen and Unwin.
Jackson, P., Stevenson, N. and Brooks, K. (2001). *Making Sense of Men's Magazines*. Cambridge: Polity.
Miller, D., Jackson, P., Thrift, N., Holbrook, B. and Rowlands, M. (1998). *Shopping, Place and Identity*. London: Routledge.
Mitchell, D. (1995). There's no such thing as culture: Towards a reconceptualisation of the idea of culture in geography. *Transactions of the Institute of British Geographers* 20, 102 116.

Relevant Websites

http://www.bl.uk/learning/histcitizen/foodstories
British Library "Food Stories" Collection.
http://www.sheffield.ac.uk/geography/staff
University of Sheffield, Geography Department, Academic Staff.

Japanese Geography

A. Saito, National University of Singapore, Singapore
F. Mizuoka, Hitotsubashi University, Tokyo, Japan

Glossary

Geopolitik A school of thought derived from geopolitics and developed during the inter-war period in Germany. Geopolitics is a long-standing area of geographic enquiry concerning the significance of space in understanding the nature of international relations and politics. In the twentieth century, however, it was utilized to justify imperialism and military aggression. In Japan, Geopolitik was interpreted and understood under the heavy influence of ethnocentric cultural and intellectual climate at that time.

Shintoism The native religion of Japan which involves the worship of *kami*, *kami* gods. Some *kami* are local and can be regarded as the spiritual beings/spirits or genies of a particular place, but others represent major natural objects and processes: for example, the sun goddess, or Mount Fuji. Shinto is an animistic belief system. The word Shinto is commonly translated as 'the way of the Gods'. Shintoism was the official state religion since Meiji until the end of World War II; but, after the war it lost its status.

The Great East Asian Co-Prosperity Sphere The name that Japanese government created to justify the attempt by Japan to create a sphere of influence over Asian nations in the interimperialist rivalry among the world powers that led to World War II. Under this name, Japan claimed to 'free Asia from Western powers'; in reality, it aimed to extend Japanese frontiers of power. The attempt failed due to Japan's defeat.

Tokyo Higher Normal School A college established in 1902 as one of the national teacher training colleges to educate secondary school teachers. It provided free education and thus attracted many bright but poor students. Tokyo Higher Normal School was one of the most prestigious colleges before World War II, and its academic level was comparable to the university. It evolved into Tsukuba University, one of the most research-oriented Japanese universities, after World War II.

Introduction

This article attempts to introduce the development of geography in Japan to the international audiences. However, it does not aim at presenting a tedious chronology of its development without much conceptual import. Instead, several unique contributions to spatial or environmental conceptions developed by the Japanese geographers are selected from a wide range of human, economic, social, as well as critical geography with variegated political positions and ideological stances.

In spite of the Tokugawa Shogunate (military) government maintaining watertight boundary controls, in 1713 Hakuseki Arai wrote *Chorography of The World* based on the deposition of an Italian missionary who had intruded illegally into Japan. In early nineteenth century, Tadataka Ino surveyed the Japanese coast and produced precise maps, using surveying technology that was developed indigenously.

Japan opened the door to the West in the 1860s, when France made brief and futile attempts to bring the Shogunate government eventually under its protectorate. It was intercepted by the British through supporting fiefdoms in southwest Japan, which led to the Meiji Restoration in 1868. Geography thus began to be established as an academic discipline under the Western influence. Physical geography and geology were the first to be established before the turn of the century at Tokyo Imperial University. The first academic unit of human geography was established in Kyoto Imperial University in 1907, in the Department of History (Ichiro Suizu). It became the core of the Human Geographical Society of Japan when it was established in 1948. Commercial geography was first taught in Tokyo Commercial School in 1886, aiming to achieve a comprehensive understanding of social/economic/political elements of a particular region/locale for trading purposes. In 1930, it developed into the first academic unit of economic geography at a college which is now Hitotsubashi University (Hiroshi Sato). Education in geography derived from a need to raise geography teachers for school education, in order to give pupils awareness of the territory of the nation as well as a psychological affinity to it. Several teachers training colleges and normal schools were created, and research of geography was also conducted there (Keiji Tanaka).

In academic geography, Japanese geographers maintained a strong interest in the development of the discipline in the West, and they learned from the thoughts and key concepts of geographers in Germany, France, UK, and the United States and were often keen to introduce them into Japan. German Geopolitik School became hegemonic in political geography, particularly in the Tokyo area in the 1930s, and some location theories were brought into Japan in the pre World War II years. Yet, some geographers in Japan ventured to develop their own geopolitics (Saneshige Komaki).

As for the political standpoints, it is fair to say that most of Japanese academic geographers remained con ventional, although Western Marxist began to influence the geographer's circle in Japan as early as the 1920s. A notable exception was the post war period until the early 1970s, when, as the contradiction of capitalist society became apparent, a unique contribution was produced in the critical geographers' circle in Japan (Noboru Ueno). However, after the oil crisis and the end of the post war rapid economic growth, critical schools waned and many of them converted into more conservative perspectives, sometimes even supporting the government land development policies. This situation necessitated more politically conscious Japanese geographers to engage in struggle in order to resurrect the critical heritage (Fujio Mizuoka) and to make positive contribution to the worldwide resurgence of critical geography since the 1970s.

Tanaka Keiji and the Concept of 'Chorography'

Keiji Tanaka (1885–1975) was one of the pioneers of modern geography in Japan. He studied in the depart ment of geographic education in Tokyo Higher Normal School, and later studied in the US and Europe under the renowned contemporary geographers such as Davis, Vidal de la Blache, and Penck. His main academic achievement was the establishment of the concept of 'chorography' in Japan.

Drawing upon Hettner and Hartshorne, Tanaka in sisted that chorography should be the foundation of geographical thought. Chorography is one of the oldest traditions of geographical thought and can be dated back to classic Greek civilization. It is derived from one of the fundamental human interests, namely the curiosity about knowing the place never visited; its main focus is to 'describe' the 'parts' of the world. This is quite different from another tradition which attempted to provide 'a view of the whole' by not describing the peculiarity of particular place but 'mapping' the entire globe. Never theless, Tanaka was not content with mere description, but attempted to give the concept of chorography the much desired dynamism that would make it a more systematic subject. Akin to the primitive version of Hägerstrand, he postulated that there are three phases of 'phenomena' through which a geographical object de velops: (1) initial phenomenon or the process of pro liferation, (2) conspicuous phenomenon or the state of domination, and (3) residual phenomenon or the process of disappearance. The actual landscape was regarded as an overlapping multiple stratum of different geographical phenomena at different phases. In 1927, Tanaka also divided the Japanese archipelago into several regions based on natural boundaries and climate, which are in dependent of the administrative boundaries. He tried to understand the characteristics of a particular locality by analyzing the relationship between different human phenomena, and the interaction between man and nature in the local context.

He also emphasized the importance and utility of the interviewing method and field work for data collection. His position as a professor at Tokyo Higher Normal School made him hugely influential among school teachers as well as his contemporary conventional geographers. He encouraged them to organize a com prehensive survey of their localities as a self discovering exercise which should also build a local pride and identity. This is shown in his book *Our Land*, first pub lished in 1929 which became seminal and continued to be read even during the 1950s.

His theory and practice can be seen as parochial be cause locality tended to be treated in isolation and as a self contained entity, not in the dynamic interaction with the surroundings. His study as well as the attempt to bring dynamism into chorography, however, advanced chorography from a mere description of a particular place to the analysis of chronological relations of human phenomena and the human–nature interactions. At its best, by successfully synthesizing the detailed geo graphical knowledge, it could lead to a comprehensive, coherent, and holistic understanding of local character istics. He was instrumental in establishing the Chor ography School in the geographical community in Japan. After retirement from Tokyo Higher Normal School, he took up the position of professorship at Rissho University and contributed to the development of the geography department there to become one of the prestigious ones in Japan.

Saneshige Komaki and Japanese Geopolitics

Born in a family of Shinto priests, Komaki Saneshige (1898–1990) was originally a historical geographer who was educated at Kyoto Imperial University. Komaki be came an advocate of the Japanese imperialism, using geopolitics as his conceptual base as he was promoted to the professorship at the Department of Geography of Kyoto Imperial University in 1938. He was initially in fluenced by the idea of Geopolitik in Germany, which was introduced into Japan in the 1930s. Hitler, who originally had an ambition to expel all the Anglo Saxon powers out of the entire Eurasian continent, planned to bring Stalin's USSR into the group of Axis countries.

Karl Haushofer gave academic glitter to Hitler's ambition with his work *Der Koritinentalblock.* After Stalin rejected Hitler's plot manifested in the Molotov–Ribbentrop agreement in 1939 and Hitler in turn commenced hostilities against the USSR in June 1941, Haushofer lost influence and Japan became isolated in the global politics.

Komaki was insightful enough to discern that this new stage of interimperialist rivalry across the globe necessitated a new ideological base that was independent of Europe. He attempted to establish Japan's own geopolitics to support its colonial venture over the other Asian countries ideologically. Komaki criticized Geopolitik as having a Eurocentric focus and an element of direct colonial powergame. He had a great antipathy toward Western imperialism and uniquely tried to establish Japanese geopolitics based on the Shintoism embodied by the Japanese Emperor. He claimed that indigenous space in Asia, particularly in China, including Manchuria, should be emancipated from the Western colonial hands and recovered to Asia through the leadership of Japan.

His idea won great support from the Japanese military regime, which proclaimed the 'Great East Asian Co-prosperity Sphere' under the paternalistic leadership of the Japanese Emperor. Komaki organized a nationalistic research group called Yoshida no Kai, which prepared the intellectual foundation for the Japanese imperialist advance to China and Southeast Asia. Yoshida no Kai gave, in exchange for affluent financial and material endowments, geomilitaristic advice to Japanese armed forces about actual tactics in attacking major Asian cities (such as Singapore and Hong Kong) under European colonial powers.

After the war, Komaki was severely denounced as having ideologically legitimatized Japanese imperialism and military aggression against its Asian neighbors. Komaki was purged because of his role in supporting the military regime, and he resigned from the Kyoto University. Later, he came back to the public life in 1957 to become the president of Shiga University, a position he held until 1965, although he never regained the leadership he had once enjoyed.

His idea is viewed in both positive and negative lights. On the one hand, his criticism of Western colonialism and the role geography played toward their imperialistic ventures is noteworthy. Particularly, his attempt to deconstruct the seemingly neutral and scientific notion of geography by uncovering its ideological bias based on supremacy of the Western imperialist powers can be noted from today's vantage point. On the other hand, however, his alternative Japanese geopolitics was based on another imperialist notion in favor of the Japanese Empire. It was mobilized to give legitimacy to Japan's military aggression, in the name of emancipation, from the Western imperialist powers to establish 'universal brotherhood' under the Japanese Emperor.

Hiroshi Sato and Dialectical *Wechselwirkung*

Hiroshi Sato (1897–1962) was one of the early Japanese geographers who specialized in economic geography. He graduated from the Department of Geography in Tokyo University and taught in Tokyo College of Commerce (now Hitotsubashi University) throughout his academic life. He served as the first president of the Japan Association of Economic Geographers, which was once the critical institution of geography in Japan. Though his main academic interest was around economic geography, he wrote extensively in other fields, such as political geography and even a popular essay. Sato also got involved in the discussion of the definition and concept of 'Landschaft (landscape)' in the 1920s and 1930s.

Sato's main contribution to Japanese geography is in the area of environmental theory in economic geography. He rejected simple environmental determinism and emphasized dynamic interrelations between society and environment, by combining the concepts of two well-known German geographers, Dietrich and Wittfogel. The critical concept of Wittfogel, then a member of the German Communist Party and critic of geopolitics, had been translated into Japanese before World War II by Seikan Kawanishi and was well known among Japanese geographers. Starting with their concepts, Sato devised a dialectical understanding of environment and economy based on historical materialism. He borrowed the chronological interpretation of *Wechselwirkung* (interaction) between human, nature, and development of culture from Dietrich, who had put forward that human beings try to change the environment and create a higher level of culture. Sato also learned the importance of the production process, from Wittfogel, as a crucial linkage between nature and economy. Particularly, he considered three elements as being the most important: labor force (human resources, skill, knowledge, organization), means of production (land, energy, etc.), and natural resources. On the basis of these two theories, he insisted that in order to understand the economic activity in relation to a particular place, it is necessary to analyze the way the production process is organized in the particular context of historical stages of cultural development and civilization. As an example, he took a case of coal mining in the Kyushu region and explained its development as an industry not just by its environmental factors but as a social process involving human organization, production process, support infrastructure, and cultural institutions. He was also aware of the interdependence between different regions of Japan and even beyond the national borders to understand the nature of regional economy. For instance, the development of coal mining in Kyushu was intrinsically linked to the expansion and concentration of heavy industries around Tokyo, and the phenomenal urbanization of Tokyo in the twentieth century.

His work represented one of the attempts by critical geography to free the Japanese geography from environmental determinism and introduce a greater dimension of socioeconomic activities. In particular, his articulation of the production process as a critical medium between the environment and society is a useful contribution to advance the methodological foundation of economic geography. Unfortunately, he remained firm on the idea that the spatial differences could be understood only through the differences in physical nature, and in a critical perspective he did not attempt to put the location theories, which explain spatial heterogeneity, out of the economic logic in relation to the homogeneous space. This task was left to Toshio Aoki, the successor to the professorship of economic geography at Faculty of Economics, Hitotsubashi University.

Ichiro Suizu's Concepts of 'Basic Region' and 'Regional Unity'

Ichiro Suizu (1923–97) was a social geographer. He taught geography at Kyoto University and Nara University. Born in Yamaguchi, Japan, Suizu completed his undergraduate and graduate studies in geography at Kyoto University. His research consistently considered 'region' as a tangible object, rather than a mental construction, and he formulated the notions of 'basic region' and 'regional unity'.

In criticizing Durkheim's placeless conception of primitive community, Suizu set out from the hypothesis that every community is inevitably place bound, forming a unity of society and place. Here 'place' is not limited to geometry of space, but includes a local system of ecology that is an integral part of a community, which forms a 'regional unity'. Suizu gave special connotation to its smallest endogenously formed regional unity, which he called 'basic region', which roughly corresponds to a feudal village in Japan. Suizu went on to stress the dialectical reciprocity between society and place: 'regional unity' is not merely society projected on ground, but it also works positively to form and change the society. The 'regional unity' also includes the spiritual unity, which works to unify the members of the community, often deploying physical and visible symbols. According to Suizu, these processes make the 'basic region' a tangibly integrated geographical object.

This 'regional unity' was claimed to be found at various spatial scales. For larger scales, Suizu stated that it had some regularity of the nodal system of space as put forward by Christaller, although this system is sometimes distorted by various local factors. What was more important was that it is the socioeconomic structure that supported the spatial system, which in turn regulated the society. Suizu stated that this dialectics of the socioeconomic structure and spatial system manifested itself along various lengths of time, and the historical development of regions was thus nothing but the unfolding of this dialectics.

Suizu tended to see the 'regional unity' as a socially homogenous entity, and his arguments lacked the analysis of the class conflicts and other conflicts within a society. Rather, this shortcoming had a bizarre consequence when Suizu dealt with 'regional unities' at higher order scales. Drawing upon his pre War German mentors, Suizu regarded the 'regional unity' as an organic body, with its own mechanisms of motion and reaction to the society, without regard to the socioeconomic relations that actually regulate its motion within the 'regional unity'. Suizu regarded a nation as well as an organism. It is a short step from this view to the former German and Japanese mystics of geopolitics that supported the wars of aggression.

In spite of this shortcoming, Suizu's conceptions, formulated by the early 1970s, were surprisingly close to such global intellectual currency of social geography in the late twentieth century as sociospatial dialectics. It is deplorable that there were only few Japanese geographers who inherited Suizu's concepts and developed them further. Nevertheless, his conceptions were later developed in effect from a critical perspective by Noboru Ueno, to whom now we turn.

Noboru Ueno's Concept of 'Humanly Created Environment'

Noboru Ueno (1926–) was trained as a Marxian economist in his undergraduate and graduate studies at Kyushu University. Ever since he took up a teaching position (teaching geography education) at Miyazaki University, he got involved in the research on the arsenic poisoning problem in Toroku, a remote mining village in Miyazaki. Minamata, the site of notorious mercury poisoning, was not very far from Miyazaki as well. It is these experiences that led him to the formulation of his concept of 'humanly created environment'.

In the 1960s, a considerable number of critically minded Japanese geographers developed original conceptions of critical geography. It would be fair to say that Ueno's conception, published in 1972, stood at the acme of the critical heritage that Japanese geographers' circle once nourished.

Ueno formulated his conceptions through critical reappraisals of Tetsuro Watsuji, a renowned Japanese phenomenological philosopher who wrote a book on humanistic climate, interpreting climate as an integral part of 'brother like symbioses' of society. Drawing upon him, Ueno stated that local environment had been subsumed as an intimate part of his/her body, to form a 'regional human group' as a dialectical unity of society and environment. Here Ueno's conception has similarity with that of Suizu; yet, Ueno was critical of Watsuji's ignorance of conflicting class relations.

Ueno stated that spatial unfolding of the capitalist labor process produced its own configuration of production space and environment. This produced environment was often degraded, which alienated a part of the body and transformed that part into a hostile entity. The physical, degraded environment having been produced by a capitalist production process thus transformed the interpretation of the environment among the locals intersubjectively. For Ueno, the environment was an integral element of the human body; yet, it was not symbiotic, but contained dialectics and alienation that was generated by the capitalist production process. The locals, having realized the alienation, would set out to struggle against the environmental degradation, which inevitably transformed itself into the struggle against capitalist mode of production. Ueno went on to state that this dialectics had various manifestations depending on places, creating 'localized human groups' with variety, which Marxist chorography was to investigate. Ueno's mentor in formulating this concept was Henri Lefebvre, who saw the physical environment as a base of praxis seeking alternative intersubjectivity.

With Ueno's contribution, critical geography in Japan stood at the forefront of critical geography across the globe in 1972, when it was not yet common among the Anglophone geographers' circle to adopt the concepts of Lefebvre or the phenomenological philosophy into geography.

Unfortunately, Ueno's conception did not find any successors in the geographers' circle in Japan. In the following year, when David Harvey published *Social Justice and the City*, Toshifumi Yada put forward his conception of 'regional structure' (*chiiki kozo*), by condemning Ueno as one of the advocates of 'economic chorography school'. As the faction of Yada picked up in numbers, the critical heritage in Japan dwindled and gave way to more conventional empiricist geography, having a closer association with neoliberalism. Ueno's conceptual achievement thus remains as a lofty peak in the heritage of geography in Japan.

Fujio Mizuoka and the Theory of Space Subsumption

When Fujio Mizuoka (1951–) studied economics as an undergraduate, critical geography in Japan enjoyed its heyday, whereas in the Anglophone world, it was about to emerge. The rapid economic growth had created problems in many localities, while universities still retained embers of student protests. Mizuoka studied, among others, Ueno's book, Yada's conception of *chiiki kozo,* and other works of critical geographers in Japan, as well as the work of David Harvey.

Although Mizuoka initially joined the study group organized by Yada, he became increasingly skeptical of the critical nature of Yada's *chiiki kozo* concept. Disappointed, Mizuoka went to the United States to pursue his graduate study at Clark University, where members of the department were still engaged in *Antipode*, the first radical journal of geography in the Anglophone world.

Mizuoka's concept of space subsumption (**Figure 1**), which first emerged in his doctoral dissertation titled 'Annihilation of space' in 1986, was thus the mixed product of two academic heritages: critical spirit in geography having developed in Japan until early 1970s and rapidly developing critical conceptions in Anglophone geography thereafter.

The idea of space subsumption came out of the concept of the 'formal and real subsumption of labor' by Marx as well as the two attributes of physical space, absolute and relative, quoted in *Explanation of Geography* by Harvey. Space, the indispensable element of society, is subsumed into society initially in its pristine form.

Pristine absolute space has the propensity of the infinite and the contiguous, offering a container of the components of society; yet, it has propensity of equilibrating anything contained within it, which undermines the presupposed independence of subject and social group. This contradiction of the formal subsumption has to be transcended by the process of real subsumption, or the production of a new element of space: bounding. It separates one social component from another, enabling each of social components to enjoy exclusive command over a territory.

Relative space, on the other hand, has the property of position that identifies points of action; yet, it has property of distance, which separates components of society, making interaction of each component difficult. This contradiction of formal subsumption has to be transcended by the process of real subsumption: spatial integration or production of network of communication and transportation. It puts together separated components to create a spatially contiguous and integrated society.

These processes bring about a new set of attributes: contiguity both in pristine absolute space and produced relative space, and separation both in pristine relative space and produced absolute space. Being the relation between absolute and relative attributes as well as pristine and produced spaces, the space with this new set of attributes is called relational space.

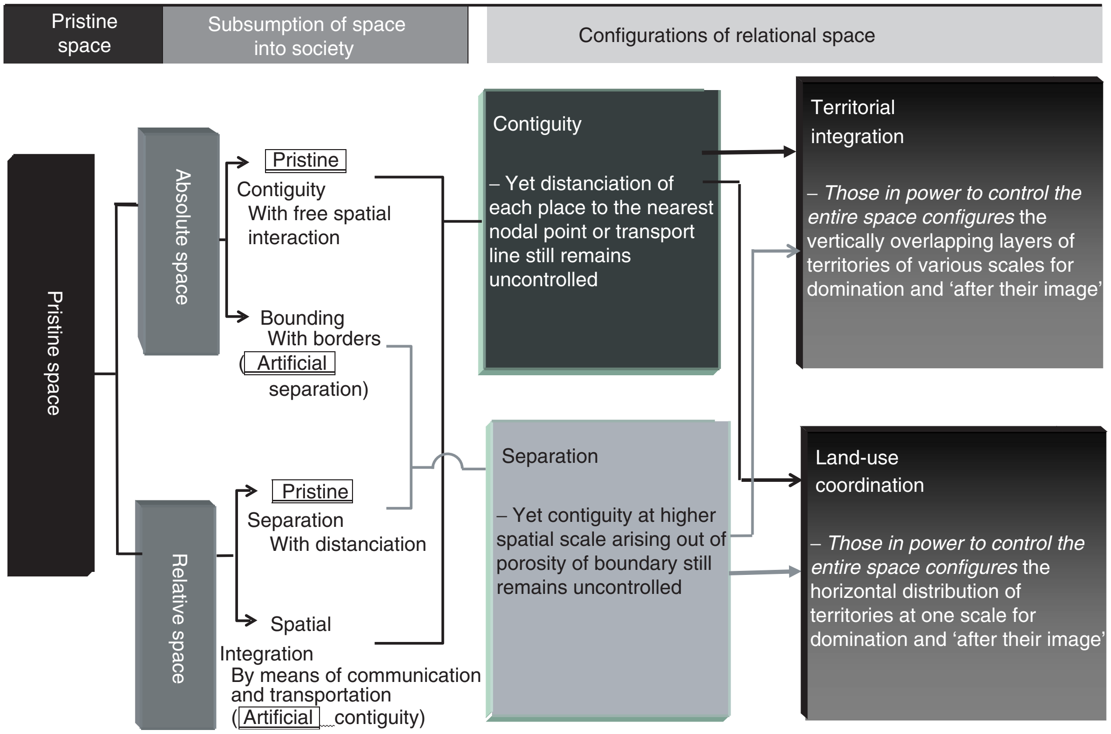

Figure 1 Production of spatial configurations through the subsumption of space into society.

It is relational space in crude form that society sub sumes in creating the configuration that supports the society. The real subsumption of relational space takes dual processes: territorial integration and land use co ordination. The configuration of space is then curved into ground as the built environment, which, as we know, creates obstacles once the social relation that created the built environment transforms itself.

Toward Resurrection

As we have seen, Japanese geographers' circle indeed had the intellectual power to create own conceptions; albeit the fact that the western geographical as well as philo sophical, sociological, and economic thoughts contributed considerably to these conceptions, they were original indeed. Some of them had insights which even stood further forefront to their counterparts in the west. Few of these contributions have become known to the com munity of global geographers, however, mainly because of the language barrier.

After the late 1970s, Japanese social and economic geography became increasingly relegated to backwash position, as compared to the community of the Anglo phone geographers who rapidly developed highly cre ative conceptions of critical geography, which eventually became intellectual currency that is now common globally.

There are some rationales of disciplinary politics behind this drastic dwarfing of the intellectual creativity. First, in the economic geographers' circle in Japan, at tempts to establish fiefdoms within the Japanese geog raphers' circle took momentum by the faction effectively led by Toshifumi Yada. According to him, his own *chiiki kozon ron* (regional structure theory) was one of the leading theories of economic geography in the world. Partly because of this sort of academic arrogance, a ser ious attempt to adopt rapidly emerging new concepts of geography in the Anglophone countries was much delayed in Japan.

Second, there was politics within geography depart ments of Japanese university campuses. In graduate schools of geography, students were asked strongly by professors to engage in empirical field research, whereas the study of philosophy and conceptions of geography was much discouraged. In carrying out the empirical studies, most of the graduate students borrowed the theory or conceptual framework from disciplines other than geography. This circumstance and the practice of 'rehashes' shunted young Japanese geographers away from serious study of critical philosophy, sociology, and radical political economics, which made them inept in appreciating conceptual import of new conceptions of Anglophone geography in its full breadth. As a con sequence, the critical concepts of geography developed by Anglophone geographers have been absorbed in Japan more readily and effectively by critical sociologists and

Marxist economists who are more theoretically oriented than geographers. Some geographers in Japan who un consciously conformed to the common practice even played 'second rehash' – adopting the works of those Japanese sociologists who wrote papers based on the concepts of critical geographers in Anglophone countries in order to conduct their research and classroom teaching.

Third, many Japanese geographers are inherently conservative and prefer to sell their discipline to the ruling government than to the oppressed grassroots. A former critical geographer even sought to join govern ment councils, which are normally controlled by the government bureaucrats.

Fourth, there is also an element in the recent devel opment of Anglophone geography itself that provided these Japanese geographers a nice excuse for this prac tice. This is the advent of postmodernism, which relieved geographers of an often painful study of universal theory of critical social sciences. Few critical geographers in Japan have come to terms with the debates on the grand picture of global economy or new imperialism of their own at a higher spatial scale in the critical circle of Anglophone geographers.

This deplorable trend needs to obviously change. There is a sign of this recovery as we saw in the last section; yet, it is still a long way to go until Japanese geography recovers the potential of intellectual creativity that existed decades ago.

See also: Anglo-American/Anglophone Hegemony; Asia; Geopolitics; Imperialistic Geographies; Locality Debates; Region; Regionalization/Zoning Systems; Territory and Territoriality.

Further Reading

Hall, R. B. (1978). *Japanese Geography: A Guide to Japanese Reference and Research Materials*. (University of Michigan Center for Japanese Studies graphy: No. 6). Greenwood, MI: University of Michigan.

Komaki, S. (1943). *Nihon Chiseigaku Sengen* (Declaration of Japanese geopolitics), (2nd edn.). Tokyo: Hakuyo Sha.

Mizuoka, F. (1996). The disciplinary dialectics that has played eternal pendulum swings: Spatial theories and discontructionism in the history of alternative social and economic geography in Japan. *Geographical Review of Japan, Series B* 69(1), 1 2. http://econgeog.misc.hit u.ac.jp/mizue/altergeog.html (accessed in May 2008).

Mizuoka, F., (ed.) (2002). *Keizai Shakai no Chirigaku* (Geography of economy and society). Tokyo: Yuhikaku.

Mizuoka, F. (2006). The demise of a critical institution of economic geography in Japan. In Mizuuchi, T. (ed.) *Critical and Radical Geographies of the Social*, *the Spatial and the Political*, URP Research Paper No.1, Osaka: Osaka City University. http://www.ur plaza.osaka cu.ac.jp/library/URP1.pdf (accessed in May 2008).

Mizuoka, F., Mizuuchi, T., Hisatake, T., Tsutsumi, K. and Fujita, T. (2005). The critical heritage of Japanese geography: Its tortured trajectory for seven decades. *Society and Space* 23(3), 453 473.

Okada, T. (2000). *Nihon Chirigakushiron* (History of Japanese geography). Tokyo: Kokon Shoin.

Sato, H. (1967). *Keizai Chiri* (Economic geography). Tokyo: Shin Kigen Sha.

Suizu, I. (1972). *Chiiki no Ronri: Sekai to Kokka to Chiho* (The logic of region: World, nation and local). Tokyo: Kokon Shoin.

Suizu, I. (1980). *Shakai Chirigaku no Kihon Mondai: Chiiki Kagaku e no Shiron* (The fundamental issues of social geography: An essay on regional science), (rev. edn.). Tokyo: Taimeido.

Takeuchi, K. (2000). Japanese geopolitics in the 1930s and 1940s. In Dodds, K. & Atkinson, D. (eds.) *Geopolitical Traditions: Century of Geopolitical Thought*. New York: Routledge.

Tanaka, K. (1948). *Warera no Kokudo* (Our land), (new edn.). Tokyo: Kokon Shoin.

Tanaka, K. (1958). *Chiriteki Sogo Kenkyu: Kawasaki shi to Tokyo Koto ku, Tokyo* (Comprehensive research in geography). Tokyo: Kokon Shoin.

Ueno, N. (1972). *Chishigaku no Genten* (The ultimate origin of chorography). Tokyo: Taimeido.

Yada, T. (1973). On economic geography. *Keizai Shirin* (The Hosei University economic review) 41(3, 4), 375 410.

Relevant Websites

http://www.asia arsenic.jp
The Asia Arsenic Network, where Noboru Ueno is the Director.
http://wwwsoc.nii.ac.jp
The Association of Japanese Geographers.
http://wwwsoc.nii.ac.jp
The Human Geographical Society of Japan.

Johnston, R. J.

James D. Sidaway, School of Geography, University of Plymouth, UK

Glossary

Anglo-American Human Geography A term used in the subtitle to a widely read textbook on *Geography and Geographers* by Ron Johnston (first edition 1979, five subsequent editions). It refers to the idea that–especially since 1945–there has been a transatlantic space of geographic scholarship; with nodes in the UK and USA. More often today the term Anglophone geography is utilized to signify a broader English language space of scholarly circulation. In recent years the putative hegemony of Anglophone geography has been a subject of heated debate.

Electoral Geography the study of the spatial relationship between voting, constituencies, and political outcomes. Key questions addressed in electoral geography related to how the size and shape of voting tracts or constituencies might effect the outcome of parliamentary or local elections.

Paradigm A concept associated with the work of Thomas Kuhn (originally published in 1962) on *The Structure of Scientific Revolutions*. In Kuhn's view a paradigm constitutes a widely accepted philosophy and methodology. These are subject to change when anomalies accumulate that cannot be satisfactorily accounted for within the existing paradigm, thus science does not simply proceed in steady, incremental fashion, but through periods of relative paradigmatic stability punctuated by periods of rapid shifts in assumptions; so called scientific revolutions. The applicability of these ideas to human geography including their relationship to generational changes and the distinctiveness of the field has been extensively debated.

Johnston, Ronald John (1941–)

Ron Johnston was born in the English provincial, 'railway' town, of Swindon, Wiltshire; just a few months after the Royal Air Force had repelled the Luftwaffe's attempt to dominate the skies over southern England. Johnston's subsequent scholarly trajectory took him to Manchester University (UK) at the close of the 1950s and into the early 1960s, Monash University (Australia) and Canterbury University (New Zealand) later in that decade, and then back to England in the 1970s. It has involved major administrative roles (he was Pro Vice Chancellor at the University of Sheffield from 1988–1992 and Vice Chancellor of the University of Essex in the 1990s) and has produced a vast range of geographical scholarship in *several* fields. Johnston is undoubtedly the acknowledged authority on the twentieth century history of human geography as epitomized in the textbook *Geography and Geography Geographers: Anglo American Geography since 1945*, (which has gone through six editions since it was first published in 1979, the last co authored with James Sidaway). His scholarship has also made very significant contributions to urban social geography, quantitative methods, and played a leading role in the study of electoral geography.

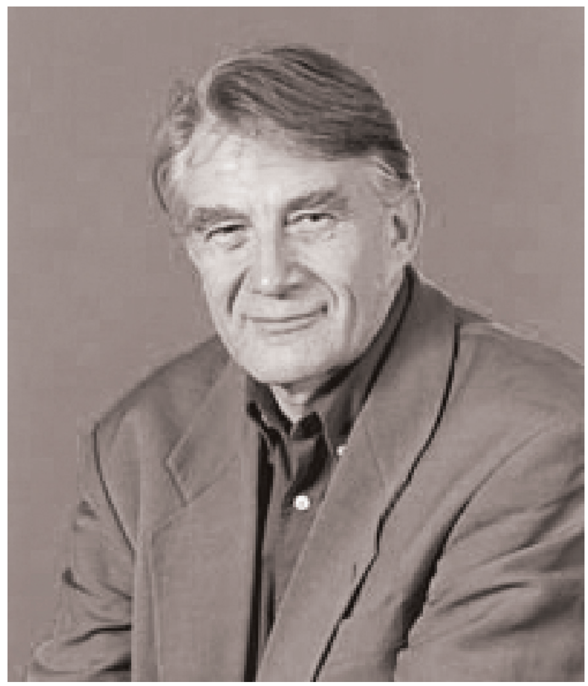

Johnston's broad span of publications is probably unique in extent, especially in an age of increasing sub disciplinary specialization within human geography. Though frequently collaborative, much of the work is sole authored, and he is arguably the most prolific living geographer over the past four decades (of human geographers in the twentieth century perhaps only Richard Hartshorne or Dudley Stamp begins to match the magnitude of his vitae). He has produced an extraordinary range of scholarship ranging from hundreds of refereed papers, through edited, co and sole authored monographs and textbooks (including the *Dictionary of Human Geography,* now in its fifth edition), to many dozens of book reviews and short commentaries. In addition he was editor for over 20 years of the journals *Progress in Human Geography* and *Environment and Planning A* and served on the boards of *The Geographical Magazine* and the *Australian Geographer*. This combination of prolific scholarship, journal editing, and senior academic management positions means that he has often been asked to serve as an advisor/evaluator and has *de facto* acted as a disciplinary gatekeeper/arbiter. This is acknowledged in a series of honors, including two Royal Geographical Society awards (the Murchison Award, 1985 and Victoria Medal, 1990), and an Association of American Geographers' award (1991). For his outstanding contributions to human, political, and economic geography and for his work on the history and nature of geographic thought in 1999 he was elected a Fellow of the British Academy and a Foundation Academician of the Academy of the Learned Societies in the Social Sciences, and was awarded the Prix Vautrin Lud by the Festival Internationale de Géographie.

Johnston began his training as a geographer in a provincial English University at a time when the commitment to *ideographic* (although those who practised it would probably not have recognized that word) regional geography was still strong. He studied geography at the University of Manchester from 1959–1964 where he took Bachelors and Masters Degrees. He then moved to Monash University, Melbourne, Australia where he undertook a PhD, focusing on residential structure and urban morphology. This move to the southern hemisphere coincided with a dramatic shift in Western geographical thought and approaches to geography as the 'quantitative revolution' was (to use some of the language it adopted) diffusing from its principal points of formulation in Lund (Sweden), Seattle (US), Bristol and Cambridge, England. Amongst his first published papers (which appeared during his time based in Melbourne) are works on urban social geography that begin to adopt the then still relatively novel apparatus and approaches of quantitative geography. Johnston then worked as a Lecturer in Christchurch, New Zealand (where he came into contact with the radical, Wellington based, geographer Keith Buchanan) from 1967–74, before returning to England, as Professor in Human Geography at Sheffield University (1974–1992), as Vice Chancellor at Essex (1992–1995), and latterly at the University of Bristol.

Despite the range of administrative responsibilities that these moves have precipitated, Johnston has retained a remarkable publication rate, developing four (frequently overlapping) strands: relating to disciplinary histories, political geography (and electoral geography in particular); urban social geography; and quantitative methods. In relation to the former, Johnston's work on the history of the discipline has focused both on institutional and contextual histories *and* the intellectual biographies of individuals. This has entailed attentiveness to the institutional frameworks (changing university cultures and wider economic and political norms) within which twentieth century human geography evolved in the UK and North America. By interrogating the work of Thomas Kuhn (1962), who had described the history of science as one of the discontinuities characterized by periodic paradigm shifts, and in identifying the pivot of contemporary Anglophone human geography around a transatlantic (UK US space) centred on certain nodes/departments, his early writing on the history of human geography has been regarded as enshrining a particular narrative in which disciplinary change is associated with individuals and networks of influence breaking through constraints of established norms to promote new world views (via paradigm shifts). Subsequent work on the history of the discipline has introduced more nuanced ideas about disciplinary transformations (drawing on a growing literature on the histories, politics, and geographies of science), but Kuhn remains a significant point of departure and Ron Johnston's work has been key in evaluating its limits and utility and introducing its vocabulary to the study of the recent history of human geography.

Urban social geography was an earlier concern in Johnston's work. Registering the wider shift (to use Kuhn's terms, a 'scientific revolution' and 'paradigmatic shift') from descriptive regional geographies and urban morphologies to work informed by statistical procedures that coincided with his early career, this work has systematically interrogated census data, survey data and legal jurisdictions and rulings to understand socio spatial inequality and difference at the urban scale. This social geography research continues in the form of recent work on socio spatial and 'racial' segregation in housing and schooling (focused on the UK case). Both this and the work on elections are relatively distinctive in blending the insights of a human geography influenced by social and political theory (more often associated with *radical geography* and *critical geography*) with a command of and ongoing commitment to statistical analysis (indeed he has written extensively on the use of quantitative methods in human geography). His work in the field of electoral geography used *spatial science* techniques to illuminate debates over the process of political constituency definition (which in North America is termed redistricting and 'gerrymandering') in electoral democracies and the impacts of local electoral campaigns. This expertise now finds practical application in his role as an assistant boundary commissioner in the UK.

Johnston's work has thereby become an essential reference point in electoral geography, particularly in the British context (but also in investigations of how the iniquitous distributions of public goods and favors – what in the US has been called 'pork barrel' politics–affect elections). Yet his influence is also inescapable in any account of the structure and recent history of Anglo phone human geography with his *Geography and Geog raphers* being a reference point against which all other disciplinary histories are judged. Both these and the other strands of his work illustrate Johnston's commitment to the nuanced understanding of place(s) as a key point of departure for broader questions/intersections of power and space, and highlight a range of work that is perhaps unequalled in terms of its breadth and scope.

Further Reading

Jackson, A., Harris, R. J., Hepple, L. W. *et al*. (2006). Geography's changing lexicon: Measuring disciplinary change in Anglophone human geography through journal content analysis. *Geoforum* 37, 447 454.

Johnston, R. (1979). *Geography and Geographers: Anglo American Geography since 1945*. London: Edward Arnold (five subsequent editions published).

Johnston, R. (1980). *City and Society: An Outline for Urban Geography*. Harmondsworth: Penguin.

Johnston, R. (1991). *A Question of Place: Exploring the Practice of Human Geography*. Oxford: Blackwell.

Johnston, R. (2003). Geography: A different sort of discipline. *Transactions of the Institute of British Geographers* 28, 133 141.

Johnston, R. J. (2006). The politics of changing human geography's agenda: Textbooks and the representation of increasing diversity. *Transactions of the Institute of British Geographers* 31, 286 303.

Johnston, R. and Pattie, C. (2006). *Putting Voters in their Place: Geography and Elections in Great Britain*. Oxford: Oxford University Press.

Relevant Websites

http://www.ggy.bris.ac.uk/staff/staff johnston.html
Johnston's homepage at the University of Bristol.

K

Knowledge and Education, Historical Geographies of

R. A. Butlin, University of Leeds, Leeds, UK

Glossary

Chorography The knowledge of geographies of regions and particular places.

Cosmography The study and description of the Earth and its wider relations with the whole universe or 'cosmos'.

Enlightenment A very broad term applied to the complex, and historically and geographically varying, processes of new ('enlightened') understandings of a wide range of knowledges, including science, law, morality, and religion, primarily (but not exclusively) associated with late-eighteenth-century Europe and North America.

Natural Theology That part of Christian faith that is deemed to be understandable through common reason and philosophy, as opposed to special divine revelation.

Introduction

The historical geographies of the production and evaluation of various kinds of geographical knowledge, together with the spatial and chronological unevenness of geographies of other kinds of knowledge and belief and their dissemination by individuals and agencies, form a key part of the understanding of the human and physical geographies of the past, and they in turn inform some of our current knowledge of places and peoples.

Geographical information and related ideas and concepts presented for elite and more general consumption have been strongly bound up in contemporary societies, economies, political and cultural systems, technologies, gender relations, and understandings, together with dominant and minority philosophies and belief systems. Each has produced arrays of ideas, artifacts, and evidences, on the basis of which those systems can be understood and analyzed. Certain aspects of geographical knowledge, such as the history of exploration and (re)discovery, are better understood than its actual dissemination, both to specialists and to those with more general interests.

Approaches to the historical geography of geographical knowledge have changed dramatically during the last decade or so, with the general rejection of the notion of a single pathway or body of geographical knowledge in favor of a pluralistic approach which evaluates, as Teresa Ploszajska has indicated, "diverse sets of knowledges and understandings of the world which are constituted and transmitted through popular social and cultural practices as well as through the printed text." The printed text includes textbooks for schools, which offer further scope for research, but limited space prevents their consideration here, where the emphasis will be on individuals and institutions.

Origins and Development

The varied sources of geographical knowledge go back at least two millenia, and developed within the contexts of the increasing spatial awareness and representational facilities of sophisticated prehistoric and early historical civilizations in Europe, Asia, Africa, and the Americas. New knowledge was produced by transglobal voyages of discovery in the fifteenth, sixteenth, and seventeenth centuries, of which we know most about those originating from Europe, and their incorporation into cultural, belief, and mapping imageries (some of which maintained an imaginative and 'fabulous' character, others of a more authentic and scientific nature). This time of voyages of discovery was facilitated by better ships and navigation methods and motivated by religious, material, political, and scientific considerations, and linked to the intense intellectual and cultural curiosity and achievements of the Renaissance in Europe, the outcome being the increase in accounts of far off places and their flora and fauna, in the number of maps produced, and in ambition for trade in exotic goods such as spices from Asia. There was no abrupt change in geographical knowledge, however, more of a transition from an older body of classical scholarship from Greek, Roman, and Arabic sources, to the new images brought back from the transglobal voyages.

There were further changes in the eighteenth century, with the engagement of geography and geographical knowledge into the Enlightenment. As David Livingstone and Charles Withers have asserted, "geography's perennial fascination with the far away, with mapping the world, with exhibiting and classifying knowledge, and with the imposition of European ways of thinking on global realms are all recognizably Enlightenment preoccupations," and were evidenced by natural theological perspectives of the world, geographical imagery, natural histories of particular places, and the promotion and reception of further voyages of discovery and rediscovery over land and by sea. The historical geographies of the science and philosophy of the seventeenth and eighteenth centuries were however greatly varied and essentially regional in their manifestation, phenomena to be seen from a pluralistic perspective rather than as a single process. Voyages led by experienced naval captains such as Captain James Cook combined scientific enquiry with broader imperial and humanitarian issues. Cook's first voyage to the Pacific (1768–71), promoted by the Royal Society and the Admiralty, had as a primary aim the astronomic observation of the transit of Venus across the Southern sky, but included seaborne exploration to establish and map the location of what was thought to be a great southern continent, and also raised the added humanitarian issue of ways of dealings with indigenous people. In the three voyages (1768–71, 1772–75, and 1766–69) extensive samples of flora and fauna were collected and recorded. Collections and images of plants were made by artists, botanists, and naturalists such as Joseph Banks, and the coastlines of Eastern Australia and the Torres Strait, of the two islands of New Zealand and of the islands of the Pacific were mapped. As David Livingstone has shown, there were problems with the credibility of the visual images of places, peoples, and animals which were made on these voyages, and different artists and botanists produced images in different styles to accommodate the varied requirements of different sections of the population at home.

Cook's voyages were paralleled by similar voyages from other European states, for example, the Trans Pacific voyage by the Frenchman Louis Antoine de Bougainville in 1768–69, as part of a global circumnavigation. He reached Tahiti in 1767 and the Tuamoto Archipelago in 1768, returning to France in March 1769. Before leaving Tahiti, Bougainville formally took possession of it in the name of France. The expedition by the Spaniards Alejandro Malespina and José Bustamente to the Pacific from 1789–94 reflected Spain's political and commercial interest in its possessions in South America and the Philippines.

Overland Expeditions

In addition to the sea based investigations there were also from the mid eighteenth century onwards European overland expeditions to investigate the hitherto largely unknown geographies (unknown to Europeans, that is) of a range of territories. Largely from trading contacts, an increasingly significant focus of investigation was to be Africa. Asian knowledge of Africa was greater than European knowledge. From the fifteenth century onward, European knowledge of the coastlines of Africa increased with voyages of discovery en route to the Indies and the Far East, through trade, including slave trading, and various European coastal settlements so that by the end of the seventeenth century the basic shape of the coastline of Africa was quite well known, but large inland areas were still largely unknown to Europeans. However, by the mid eighteenth century there was increased demand for geographical knowledge, partly through British opposition to the slave trade and partly through aspirations for extending more legitimate forms of commerce with and beyond Africa.

The Association for Promoting the Discovery of the Interior parts of Africa (The African Association) was founded in London in 1788, its aim being to increase the geographical information about inland Africa, and to promote commerce, scientific knowledge, and humanitarian effort. The antislavery campaigner William Wilberforce was a member for one year. The association promoted exploration, focusing on north and West Africa. The explorer Mungo Park was supported by the Association in an expedition in 1794–95 from Gambia to Timbuktu, and among other discoveries established the course and direction of flow of the River Niger, and encouraged further British interest in commerce with West Africa. The Swiss explorer, J. L. Burckhardt, was engaged by the association and traveled in Syria, Egypt, Palestine, Arabia, and Nubia from 1809–15. With the foundation of the Royal Geographical Society (RGS) in 1830, the members of the association decided to merge with the RGS, which itself promoted exploration in Africa and many other regions in the nineteenth and twentieth centuries.

The Nineteenth and Early Twentieth Centuries

The production, representation, and dissemination of wide ranges of geographical knowledge and sources in the nineteenth and twentieth centuries have been effected through a miscellany of agencies, including talented, curious, and sometimes idiosyncratic individuals, and various amateur and professional societies, including the many geographical and related scientific and social societies that sprang up, mainly in the nine teenth and early twentieth centuries, first in Europe and then elsewhere. The dissemination of 'geographical' knowledge has taken, and still takes, many different forms, including public lectures (with different types of illustrations, including maps and lantern slides), publi cations, including the formal proceedings of learned so cieties published in their own journals, popular accounts in journals and newspapers and in book form, the pro duction of maps both as important means for describing the geography of places and for inscribing 'new' facts acquired through exploration, rediscovery and discovery, and many different kinds of display and exhibition.

In addition, the growth of education systems, in cluding those for young pupils in schools and those for students in higher education establishments, incorpor ated changing geographical knowledge. Some of it was new, some of it badly out of date and reflective of rather limited cultural and geopolitical ideas. There were in the nineteenth and early twentieth centuries strong con nections between national confidence and geographical knowledge, especially as deployed successfully or un successfully in major military conflicts, such as the Franco Prussian War (1870–71), at the end of which both sides bemoaned their respectively perceived inadequacies of geographical knowledge, notably accurate maps. The links between geographical knowledge and the promo tion of imperial and colonial ideals and settlements by European states (and later by the United States and Japan) were close but complex.

Cultures of Exploration and the Dissemination of Geographical Knowledge

Exploration and the acquisition and promotion of geo graphical knowledge became part of British and much Continental European and some colonial culture, ad vancing scientific understanding through measurement of such features as altitude, exact location of places and the nature of landforms, vegetation, and of human and ani mal inhabitants of distant places, in addition showing the confidence in a changing national identity closely tied up with the aspirations of imperialism and colonial settle ment. In the second half of the nineteenth century, not withstanding British involvement in India and the settler colonies of Australasia and North America, and increas ing French interest in Indo China, and of most of the European powers in the Pacific territories, Africa became a major focus of interest and ambition. As Robert Stafford has indicated, in his study of the geologist/geographer imperialist, Sir Roderick Murchison, it was in Africa.

> more intensely than anywhere else, [that] the explorer acted out the European longing to be challenged by nature in a wild and exotic setting. In so doing, he simultaneously verified the superiority of European civilisation, opened new frontiers for expansive capital ism, and provided an outlet for emotional impulses sti fled by industrialisation and urbanisation. (Stafford, 1989)

The interest of Europeans in the information brought back by explorers of Africa is evidenced in the attendance at public lectures given by the explorers, by the numbers of books sold that recorded their exploits, and the at tention given to them by newspapers. The image often presented is of heroic white males, countering many environmental and human hazards with the aid of a small amount of scientific equipment, some weapons, and a large number of indigenous 'porters'. This general image requires much qualification. Apart from the wide vari ations in male experience, there were also female tra velers and explorers, such as Mary Kingsley, Gertrude Bell, and Isabella Bird, whose exploits gradually engaged the wider public and the attention of the geographical societies. In addition, the role of indigenous peoples in advancing geographical knowledge, often incorporated within or resulting from contact with European exped itions, has been greatly undervalued.

The agencies for promotion of expeditions and of wider processes of collecting and disseminating geo graphical knowledge included departments of state, notably those with military or naval responsibility, sci entific societies (metropolitan and provincial), and other institutions with particular perspectives such as colonial and emigration societies. In Britain, significant insti tutions were the Admiralty, the India Office, the Colonial Office, the African Association, the RGS, other geo graphical societies in Edinburgh, Glasgow, Dundee, Manchester, Liverpool, Hull, Newcastle upon Tyne, and Southampton. There were similar institutions in France, Germany, Belgium, the Netherlands, Portugal, Spain, and Italy. The question of links with commerce is interesting, with the provincial societies in Britain, through prox imity to the aims of chambers of commerce and major merchants and traders as members, showing more interest than the RGS. The only attempt to establish a British Society of Commercial Geography (in 1884) was unsuccessful, but commercial geographical societies were founded elsewhere in Europe, notably France and Germany.

In Britain a major promoter of expeditions and of geographical research and education at home and abroad was the RGS, founded in 1830. Like other scientific so cieties, its roles in the production and dissemination of geographical knowledge were varied, for the energy and priorities were not those of a monolithic entity but of a complex, often contradictory, and changing mixture of individuals and groups of members. The Society pro vided small grants, loans of instruments, training in surveying and mapping, and advice on preparation for hazardous travel and expedition, and later in the nineteenth century for more general travel, through its booklet *Hints to travellers*, first published in 1854. It also promoted a populist image of exploration through its lecture programs at its various locations in London, with the meetings addressed by famous explorers of Africa such as Stanley and Livingstone attracting large audiences. Its officers were mainly aristocrats, military and naval officers, senior politicians and government administrators, and it played a varied role in the promotion of geography within Britain and more widely through its award of medals and its exchanges of publication with other geographical societies around the world. Unlike many other geographical societies it excluded women as members ('Fellows') until 1913. On the whole it eschewed an overtly political, commercial, or imperial stance, unlike sister societies in France and Germany, for example. It preferred to shelter, at least in a formal sense, behind its constitutional designation as a scientific society, but the affiliations of many of its Fellows meant an inevitable link with all these areas, with the exception of commerce. The Society's officers were active correspondents on geographical aspects of national issues in the major newspapers, and its publications, including its *Proceedings…and Monthly Record of Geography*, the *Geographical Journal* (from 1892), and more popular publications such as its *Ocean Highways* and later *Geographical Magazine* (from 1878). At the beginning of the twentieth century its journal showed a transition between continuing coverage of exploration and hunting in Africa, and scientific surveys of Scottish lochs and of the human geography of individual English parishes.

By the early twentieth century the major remaining areas for exploration were the Arctic and Antarctic, and European expeditions to both, but particularly the Antarctic, attracted a very wide range of (often conflicting) attention and comment. The attention given to explorers such as Robert Scott (1868–1912), Ernest Shackleton (1874–1922), Fridtjof Nansen (1861–1930), and Roald Amundsen (1872–1928) was extensive. The Norwegian explorer Amundsen, who had navigated the infamous Northwest Passage from the Atlantic to the Pacific, was the first to reach the Antarctic South Pole in December 1914, one month before Scott's ill fated expedition, two epic journeys that received massive press attention and caused deep reflection on the efficacy of Scott's logistics and on national pride.

A detailed history of British geography and geographical knowledge after World War I through to the early twentieth century cannot be provided here for reasons of space, but in essence it involved the inter war and post 1945 expansion of university geography departments, the continuing role of the RGS in the advancement of geographical research through exploration, encouragement of research, publication, and debate, the foundation of the independent and highly influential Institute of British Geographers in 1933, the merger of these two institutions in 1995, progress in the teaching of geography in schools, and an increasing – yet still understated – presence of geographers in public affairs and debates. Major changes have occurred in the techniques of measurement and evaluation of a wide range of data, including remote sensing and geographical information systems, and also in the approaches to many old and new geographical issues, such as gender, development, historical change, the philosophy and ideologies of geography, globalization, urbanization, economic change, and many others.

The Production of 'Colonial' Geographical Knowledge

The history and geography of knowledge producing agencies in some European colonies is worth brief examination, reflecting issues that have a spatial distinctiveness and at the same time mirroring broader issues of common interest with 'mother' societies of the metropoles. The relative independence of future dominions throughout this period must, however, be recognized. The existing geographical societies in Europe had been joined, by the late nineteenth century, by many more in Europe and the Americas and by those that were founded in the European overseas empires, including Australia, North Africa, and parts of Asia. Most kept some kind of connection with the metropolitan societies and aired local geostrategic issues, but they also gave attention through their meetings, publications, excursions, and sponsored expeditions to economic development, population migration and settlement, labor forces, and environmental hazards. They exchanged publications with geographical and other scientific societies, thus promoting and receiving a wide array of information. Joe Powell has suggested that the geographical societies founded in Australia in the late nineteenth century "reflected the odd mixture of national and imperial sentiments which permeates so many Australian institutions." This is demonstrated in the interests of branches of the Geographical Society of Australasia (itself founded in 1883, given a 'Royal' prefix in 1886, with largely autonomous branches founded in 1883 in New South Wales and Victoria and in 1885 in Queensland and South Australia), in exploration of Australia itself and of New Guinea and the Antarctic, and indicates in part the significance of webs and nodes of imperial and sub imperial connection, including the roles of colonial governors and senior religious and administrative figures who were often presidents of such societies. It should also be noted that competition from some local scientific 'Royal Societies', notably in exploration, sometimes overshadowed the activities of the geographical societies. There was overlap in membership, so the resolution of the exact roles of these societies is not easy.

The South Australian branch of the Royal Geograhical Society of Australasia (RGSA), based in Adelaide, organized meetings, published its proceedings, and sponsored expeditions, its constitution listing its main purposes as scientific, commercial, educational, and historical (including the collection of historical records of geographical interest), and the compilation of data on the geography of Australia. The first president, from 1885 to 1894, was Sir Samuel Davenport, an English born, Adelaide based businessman and banker, major land owner and politician, who promoted the production of olive oil, tobacco, and silk, and was President of the Chamber of Manufacturers. The society supported the Elder expedition and the Calvert expedition of 1896–97. The Queensland Geographical Society was of similar composition and focus, but was, for obvious reasons of proximity, very much bound up with Australian attempts to persuade Great Britain to annexe the non Dutch part of New Guinea in the face of growing interest from Germany and France, and supported a number of expeditions to that territory whose outcomes were reported in its publications, at its lectures, and in local newspapers.

The surveyor, J. P. Thomson, who had previously worked in Fiji, was an influential officer of the Queensland Society. He published widely on Australian geography, in local and in British and French geographical journals. The Society had 14 sub branches in Queensland in the 1920s and 1930s, though they were mainly short lived. It raised funds for Antarctic exploration and the exploration and conservation of the Great Barrier Reef, and like other geographical societies in Australia hosted visits by famous explorers, such as H. M. Stanley who addressed a meeting of the Society in Brisbane in 1891. One of the issues which was debated and published by the Queensland Society was that of the provision of an adequate agricultural labor force, notably for the sugar cane areas of the north, and this in turn reflected a wide range of opinion on color and race, and the roles and potential of Australian aborigines.

South Africa represents a contrasting source of geographical knowledge in a much older colony. Settled by Europeans since the seventeenth century, it was a region of continuing conflict, substantially over land use and later mineral rights, between Europeans, notably the British and the Dutch based settlers – the Afrikaaners/Boers – and the various groups of indigenous African peoples. The production of maps and written geographical accounts knowledge dated back to the time of European settlement, but the foundation of scientific societies in South Africa came somewhat later than in other colonies, and the first geographical society – The South African Geographical Society – was not established until 1917.

W. S. Barnard has described the context as one in which

> Two issues dominated the political discourse: the tussle between the English and Afrikaans white elites about South Africa's place in the British imperial sun and the relationship between the country's black and white inhabitants. (Barnard (1999))

The new Geographical Society's aims were

> To raise the standard of geographical education, to encourage geographical research in all its branches, and to arouse in the general public some enthusiasm in a subject which has a direct bearing on every day life.

This society had imperial links, and tended toward the British tradition, most of the published papers being in English, though membership was constitutionally open to all. It was not instrumental in the promotion of expeditions and had a small membership, but nonetheless attempted to promote geography professionally at all levels of education and of research. However, there were no real reflections on the life and culture of black Africans, for whom much education, including higher education, was limited, and which became worse after 1948. Afrikaaner culture and politics led to the establishment of an Afrikaans Society for Geography, and the two societies, each of which had published its own journal, merged in 1994 to form the Society of South African Geographers.

Geography had been advanced in France in the eighteenth century, conspicuously in mapping and survey and infrastructural improvements by the Napoleonic *ingénieurs géographes*. As Anne Godlewska has observed, they

> mapped rivers, mountains, canals, towns, cities, roads and ways, monuments, and the movement of armies all over Europe from Italy, Piedmont and Savoy to Elba, Corsica, Bavaria, Austria, Russia, Spain and Egypt … They must have produced thousands of maps in the course of Napoleon's reign. (Anne Godlewska)

This output included the mass of maps and surveys linked to the expedition to Egypt from 1798–1801. Most of these developments were closely linked to the development of national identity and imperial ideology, and to an increasing tightening of the control of scientific knowledge by the French state.

The *Service des Missions*, a government research council established in 1842, promoted French scientific missions within Europe and to overseas countries by the award of grants, and whose records show not only the extent and nature of the individual projects, but also what

Michael Heffernan has described as "the petty rivalries, feuds and political struggles which characterized French scholarship at this period." This was but one among a complex and changing sets of geographical knowledge which could be used in the contested projects of French imperialism in the nineteenth and twentieth centuries. Among the institutions of science were the geographical societies. The Paris Geographical Society, had been founded in 1821, and this was followed by the foundation of 32 additional French geographical societies by 1894. In the French colonies geographical societies had been established in Algiers in 1857 and revived in 1880, Oran in 1878, Constantine in 1883. In 1917 a Moroccan section of the Algiers Society was established in Tangiers, and another society was founded in Casablanca in 1922.

Universities and Schools

The knowledge promoted in teaching of geography in universities, schools, and other educational institutions naturally changed through time, with the earlier focus being on maps, basic geographical descriptions, and links to historical information, for example, the geographies and images of classical Greek and Roman times, and they also reflected strong influences from religious and other cultural systems. The basic science of exploration and navigation typical of the seventeenth century was replaced by a more sophisticated science of exploration, discovery, and recording in charts and maps by the late eighteenth century, as shown in the information brought back by voyages in the Pacific, some of which are described above. The work of these and other important scientists such as the Prussian explorer and scientist Alexander von Humboldt, who had traveled in South America, brought together new analyses of human and environmental data and symbiosis, laid the foundations for a new kind of scientific geography, and gradually informed the kind of geographical knowledge that was taught in various educational institutions.

Keltie, the RGS, and Geographical Education

Geography in many different forms had been a part of the teaching of British universities in the sixteenth, seventeenth, and eighteenth centuries, and included mathematical geography, historical geography, cosmography, descriptive geography, astronomical geography, and chorography, and often linked to theology. There were also geographers in the English and Scottish royal courts, and the subject was taught in schools and promoted by learned societies.

The spread of new ideas and data in geography from the late eighteenth century onward was spatially uneven, reflecting both national and local traditions and preferences. Within Europe, innovative geography advanced at a more rapid pace in France and Germany than it did in Britain, for example. This was clearly shown by the survey of geography teaching undertaken in the late nineteenth century for the RGS, which became increasingly concerned with the minimal interest in geography within the British educational system, including the schools and universities, civil service, and the army and navy. In 1869 the society had instituted the award of 'public schools medals' for the promotion of geography, but the system was ineffective and discontinued in 1884.

In 1884 the society commissioned a review of geographical education by John Keltie, appointed as its Inspector of Geographical Education. His report on 'geographical education' to the society in 1885, published in 1886, marked a major step forward for the future development of the subject. Keltie visited schools, universities, and military academies in Britain and Continental Europe – Germany, Austria, France, Switzerland, Italy, Belgium, Holland, and Sweden – and obtained material from Spain and the United States. His visits in Britain were mainly to public schools which had competed for the society's medals, the data supplemented with material from normal and primary schools, for which the state had responsibility, and in which geography of a basic form had been made compulsory.

His general view of geographical education in Europe was that it was less tainted by a reliance on textbooks than that in Britain. He was particularly impressed with the teaching of geography in the schools, military academies, and universities in Germany – (Germany provided "the model which all the other Continental countries are following, as far as their special circumstances will permit"). Studies commenced with *Heimatskunde*, the geography of the home region, and proceeded to regional, historical, political, physical, and regional geography through middle to higher education and even into PhD level. He wrote about the rapid rate of progress of geographical education in France from 1871, the year in which a report to the Minister of Education had bemoaned the lack of interest in geography. Generally similar progress was reported for Italy.

From his visits and the additional information collected, he concluded that Britain needed much more effort to develop the teaching of geography in universities and schools, conspicuously through stronger representation to government and education authorities, some financial support, the award of prizes, and displays of maps and other kinds of geographical apparatus. These proposals were followed up by the RGS, but less than a decade later, in 1893, a further step forward was made with the foundation of the Geographical Association, specifically intended to promote the improvement of illustration in geography teaching. One of the proponents was Halford J. Mackinder, a talented and energetic geography teacher at Oxford, whose advancement of a 'new' geography was most influential from the late nineteenth and early twentieth centuries. Another was J. S. Keltie, by then Secretary and Editor at the RGS.

In 1885 there were, according to Keltie, 12 chairs of geography in Germany, with two in the process of being added in Münster and Berlin (the second, in physiography), eight in Austria Hungary, eight in France (three jointly with history), 12 in Italy, one in Switzerland, one in Belgium, and none in Sweden.

In Britain, though geography had been taught in Oxford since the mid sixteenth century, significant progress was not made in the institutional development of the subject until after the mid nineteenth century, partly by the prompting by the RGS. H. J. Mackinder was appointed Reader in Geography in 1887, a School of Geography was established in 1899, but a professorial appointment in geography was not made until 1910. There was no honors school there until 1932. At Cambridge, similar representations had been made by the RGS for the development of geography, and an appointment was made in 1888. The first reader was appointed in 1898, and the Geographical Tripos was instituted in 1919. The first honors degrees were awarded in 1921, and the first professor of geography was appointed in 1931. A chair in geography was established at University College, London in 1903, and Britain's first honors school of geography was established at Liverpool in 1917.

From the 1870s, the Royal Scottish Geographical Society (founded in 1884) had pressed for more geography in local universities. In the case of Edinburgh University it was supported by the famous mapping company run by J. G. Bartholomew, and a post in geography was created in 1908; a chair was eventually established in 1930. At Glasgow the first appointment in geography was made in 1909.

The development of various types of geography teaching in schools goes back well into the seventeenth century in Britain, and was in part influenced by the types of geographical material available and by the different forms of school. In England and Wales there was state support for voluntary societies from 1833, including the churches and nonsectarian societies that ran schools, and from 1870 compulsory education was introduced at elementary level, with public provision for secondary education in England introduced in 1902. There are different histories for education provision in Scotland, where there had been an element of compulsory education since the fifteenth and sixteenth centuries, with strong church responsibility; major state provision for education was made from 1872.

The work of the Geographical Association, and other geographical societies, helped geography accelerate as a subject in universities and teacher training colleges from the beginning of the twentieth century, and to become a significant part of the school curriculum. Characterization of a progression in school geography from rote learning of geographical facts and places ('capes and bays') to more interpretative and better illustrated kinds of geographical knowledge probably underestimates the variety of information available from the later nineteenth century. Keltie's report does suggest, however, that rote learning was an undeniable feature of some geography teaching. The contents of textbooks, and other kinds of literature available for young people, including fiction, do reflect European characteristics of modernity, imperialism, nationalism, patriotism, citizenship, and commercial advancement – and also, though more rarely, opposition to such processes. Another significant feature is the extent to which geography texts and materials published for consumption by indigenous pupils in schools in the colonies failed for quite some time to adjust the material to local circumstances and interests. In extreme form it can be quite horrifying, as Teresa Ploszajska has shown in the context of racialist geography textbook coverage in Britain in the late nineteenth and early twentieth centuries of aboriginal people in Australia.

Geographical Education in France

In France, promotion of different aspects of geographical knowledge was, as elsewhere, a complex process, involving a wide range of individuals, institutions, and state policies. Progress was relatively slow in schools and universities. By 1880 chairs of geography had been established at the universities of Lyons, Bordeaux, Caen, and Dijon; there was one at Lille in 1893 and at Rennes in 1899. L. A. Himly held a chair at the Sorbonne, and another chair – of colonial geography – was added in 1892. Geography was, however, also promoted at other higher level educational institutions, including the Collège de France, various teacher training colleges, and the military academy at St. Cyr. Paul Vidal de la Blache was appointed to the chair at the Sorbonne in 1898. The teachings and research of Élisee Reclus (1830–1905), Emile Levasseur (1845–1918), and Paul Vidal de la Blache (1845–1918) were the foundations of the French School of Geography, partly influenced by German geographical ideas, but developing distinct characteristics, including emphases on the regional components in spatial systems, reflected in their different personalities. Their ideas were taken forward by such scholars as Lucien Gallois (1857–1941), coordinator of the famous Géographie Universelle, Emmanuel de Martonne (1873–1955), and Jean Brunhes (1869–1930).

In French schools the provision for the teaching of geography at primary level was quite limited before 1870, in spite of reforms of 1857 in primary education and in

1864 in secondary education, with emphasis on descriptive vocabulary of basic geographical features and some rote learning of names of objects and places. At secondary level the priorities and emphases first laid down by teaching by Jesuits and other religious orders dominated, as Paul Claval has outlined, in the Napoleonic schools and colleges, where classical history required some modest background knowledge of geography. In the universities there was limited geographical presence, with a chair at the Sorbonne, formerly of history and modern geography, becoming a chair of geography in 1812.

From 1871, in response to the energetic attempts to revive French status and national pride, geography moved forward. This is evidenced in the publication in 1877 of the *Tour de France de deux enfants*, a book used as a primary text. Its focus is on two orphans who escape from the recently annexed Lorraine and travel in France looking for their uncle; the by product is a series of geographical descriptions and maps of France and its regions. In the same period there was a revival of regional novels, and wider provision of geographical information through travel and adventure novels and the production of guide books by the publishers Hachette. After 1870 there was reform in French geographical instruction, with priority given to better knowledge of France itself and its place in the changing world, and of the use of maps. Prior to this wall maps as teaching aids had been scarce, and their design and circulation were improved. At primary level children were introduced to the rudimentaries of geographical description and the use of maps, and the basic characteristics of the geography of France and the rest of the world. At secondary level, an improved program for geography was introduced in 1872 and updated in 1890 and 1902, and included material on economic and political geography. A leading proponent for the teaching of new improved geography was the historian Ludovic Drapeyron (1839–1901), who founded the journal *Revue de Géographie*, in 1877.

Geography and Geographical Education in Asia

It is important to bear in mind also the developments in geography in countries and cultures that are not European. Geography had been part of the Japanese epistemology of its early civilizations, but advanced in the nineteenth century with exposure to Western and Chinese ideas after the Meiji restoration in 1868 and the introduction of a compulsory educational system in 1872. Geography was employed in the attempts to create a strong national identity and a modernized state, and posts in geography were created from the late nineteenth century. Chairs of geography were established in the imperial universities at Kyoto in 1907 (within the Institute of History) and Tokyo (1911). A Tokyo Geographical Society was founded in 1879. Geography was also bound up with Japanese imperial ambitions in the first half of the twentieth century, but after the end of World War II expanded to cover a wide range of geographical topics.

Similarly geography in China has deep roots, dating from prehistoric and early historic times, reflecting investigation of the geography of China and of regions explored by the Chinese over land and by sea, including the Mediterranean and the Indian Ocean. Hooson suggests however, that

> when the western world (including the Russians) forced its way into China in the nineteenth century, it found a crumbling, still feudal empire that had forgotten its ancient civilization, including geographical knowledge. (Hooson, 1991: 31 32)

Thereafter, the pattern of development related to interchange with Western ideas and personnel, and the first university department of geography in China was established in 1928. Almost inevitably, Chinese geography at all levels was then influenced by nationalist and communist ideologies. From 1949, with the establishment of the People's Republic of China, it was inextricably linked with an extreme cultural and political ideology, from which it is only just beginning to recover.

In India the development of modern geography was also quite late, and tied in with the increasing intensity of British control from the mid nineteenth centuries, but it was built on much older and varied Indian geographical traditions, dating back to late prehistory, and including in early historical time the geographical material presented in Hindu ashrams and Jain and Buddhist *viharas*, and the map data from Arabic contact in the eighth century. The Bombay Geographical Society was founded in 1832, mainly on the initiative of naval officers of the East India Company – only 2 years after the RGS – of which, for 5 years, it existed as a branch. It published proceedings from 1837, and in 1873 merged with the Bombay branch of the Royal Asiatic Society. Kapur contends that:

> The British were the official couriers of the seeds of a segregated and defined discipline of Geography in India. For over a hundred years the British used the tools of this science to measure, collect, organize and document the geographical facts of India. But the study of geography in India did not find a habitat of teaching and research until the early 20th century. (Kapur, 2002)

The mapping of India from the late eighteenth century by the Great Trigonometrical Survey was one of a number of survey and mapping institutions set up by the British, others including the Geological Survey of India (founded 1851), the Census of India (1861), The

Archaeological Survey of India (1861), and The Indian Meteorological Department (1875).

In spite of this, and the establishment of universities in 1857 in Calcutta, Bombay, and Madras, in 1882 at Lahore and in 1877 at Allahabad, no geography was taught until 1920, in colleges linked to Lahore University, and sub sequently at Aligarh in 1924 and Patna in 1927. In 1925 there were 20 universities in India, none of them in cluding geography teaching, but by 1947 there were 37 universities, of which 17 taught geography. In the same time period the number of geographical societies and geographical journals increased from one to six. By 1970 there were 48 university departments of geography. A National Association of Geographers was established in 1978. The progress of geography in Indian universities does not, however, seem to have been matched by similar progress in schools, where geography is taught as part of a social studies curriculum that combines geography and history, and is often taught by those without specialist geographical knowledge.

Geography in the United States and South Africa

In the United States geography teaching and learning goes back to colonial times, but in its modern form ap peared from the mid nineteenth century, with major publications by George Perkins Marsh, William Henry Brewer, and others. Appointments were made at Princeton and Yale and geographical societies were founded at Philadelphia (1881), and Chicago (1898). Geography departments were introduced at University of California, Berkeley, and at Columbia University Tea chers' College in 1898. There was a strong tradition of field survey which provided substantial foundations for the subject. W. M. Davis, who taught geography at Harvard from 1878–1912, made a major contribution to the investigation of landforms; other noted pioneers of geography included Mark Jefferson, Isaiah Bowman, Ellen Churchill Semple, and Albert Perry Brigham. The Association of American Geographers was founded in 1904. After World War I there was substantial progress in all branches of physical and human geography, and be yond World War II American geographers pioneered developments in a number of fields, including quantita tive geography.

In South Africa, despite the absence of university geography departments and formal geographical soci eties, geography became part of the primary school curricula from 1839. From the mid nineteenth century it emerged at the secondary level, albeit with lowly standing, in the Cape Colony and in Natal, the Orange Free state, and Transvaal. Thereafter, the teaching of empire geography, notably in the Afrikaaner territories, served to promote British imperial ideologies, and that process accelerated after the Anglo Boer War, with geography of the British Empire as a key component in the draft matriculation certificate syllabus of 1913. There is little information at present of the views of the black population on contemporary geographical instruction, though it is clear that the use of geography by the British to justify the Anglo Boer War and its outcome was anathema to Afrikaaners. Wesso and Parnell have sug gested that while the founding of the South African Geographical Society stimulated geography in South Africa, produced its first journal, and generally helped to progress geography in tertiary education, the main in fluence was still British, with a strong incorporation of environmental determinism.

The first chair in geography was created at Stellen bosch in 1920, the holder being Piet Serton (1888–1963) from the Netherlands. The first university geographers in South Africa were British or European trained, and some progress was made in the period 1920–60. Ideo logical differences were reflected in school and university admissions policies and curricula. The Extension of University Act of 1959 required that black students had to have special permission to study at white universities, or they could study at separate black university colleges, where they were often taught by Afrikaaners – so that, as Wesso and Parnell indicate,

> black geography students, a large proportion of which were student teachers, were confronted with a geography with strong ideological and environmental determinist undertones. (Wesso and Parnell, 1992: 186 199)

See also: Genealogy and Family History; Historical Geography, Evolution of; Imperialistic Geographies; Nature, Historical Geographies of; Nature, History of.

Further Reading

Baker, A. R. H. (2003). *Geography and history. Bridging the great divide*. Cambridge: Cambridge University Press.
Barnard, W. S. (1999). Encountering Admastor: South African founding geographers and their regional geographies. In Buttimer, A., Brunn, S. D. & Wardenga, U. (eds.) *Text and image. Social construction of regional knowledges*, pp 192. Leipzig: Institut fur Landerkunde.
Butlin, R. A. (2006). Geography Imperial and local: The work of the Royal Geographical Society of Australasia (Queensland Branch) 1885 1945. In Childs, I. R. W. & Hudson, B. J. (eds.) *Queensland geographical perspectives*, pp 217 241. Brisbane: The Royal Geographical Society of Queensland.
Claval, P. (1998). *Histoire de la géographie Française de 1870 à nos jours*. Paris: Nathan.
Dodgshon, R. A. (1998). *Society in time and space. A geographical perspective on change*. Cambridge: Cambridge University Press.
Driver, F. (2001). *Geography militant. Cultures of exploration and empire*. Oxford: Blackwell.
Godlewska, A. (1994). Napoleon's geographers (1797 1815): Imperialists and soldiers of modernity. In Godlewska, A. & Smith, N. (eds.) *Geography and empire*, pp 31 54. Oxford: Blackwell.

Godlewska, A. (1999). *Geography unbound: French geographic science from Cassini to Humboldt*. Chicago: University of Chicago Press.
Graham, B. and Nash, C. (eds.) (2000). *Modern historical geographies*. Harlow: Pearson Education.
Heffernan, M. J. (1994). A state scholarship: The political geography of French international science during the nineteenth century. *Transactions, Institute of British Geographers, NS* 19, 21 45.
Hooson, D. J. M. (1991). China (People's Republic). In Dunbar, G. S. (ed.) *Modern geography. An encyclopedic survey*, pp 31 32. Chicago and London: St James Press.
Hudson, B. (1977). The new geography and the new imperialism. *Antipode* 9(2), 12 19.
Johnston, R. (2003). The institutionalization of geography as an academic discipline. In Johnston, R. & Williams, M. (eds.) *A century of British geography*, pp 45 90. Oxford: Oxford University Press for the British Academy.
Kapur, A. (2002). *Indian geography, voice of concern*. New Delhi: Concept Publishing Company.
Livingstone, D. N. (1992). *The geographical tradition*. Oxford: Blackwell.
Livingstone, D. N. (2003). British geography, 1500 1900 an imprecise review. In Johnston, R. & Williams, M. (eds.) *A century of British geography*, pp 11 44. Oxford: Oxford University Press for the British Academy.
Livingstone, D. N. (2003). *Putting science in its place. Geographies of scientific knowledge*. Chicago: Chicago University Press.
Mayhew, R. J. (2000). *Enlightenment geography: The political languages of British geography, 1650 1850*. London: Macmillan.
Peake Jones, K. (1985). *The branch without a tree. The centenary history of the Royal Geographical Society of Australasia (South Australian Branch) Incorporated. 1885 to 1985*. Adelaide: Royal Geographical Society of Australasia (Adelaide Branch).
Powell, J. M. (1986). Approaching a dig tree: Reflections on our endangered expedition. *Australian Geographical Studies* 23, 3 25.
Ploszajska, T. (1999). *Historical geography research series no. 35: Geographical education, empire and citizenship: Geographical teaching and learning in English schools, 1870 1944*. London: Institute of British Geographers.
Ploszajska, T. (2000). Historiographies of geography and empire. In Graham, B. & Nash, C. (eds.) *Modern historical geographies*, pp 121 145. Harlow: Pearson Education.
Stafford, R. A. (1989). *Scientist of empire. Sir Roderick Murchison, scientific exploration and Victorian imperialism*. Cambridge: Cambridge University Press.
Wesso, H. (1994). The colonization of geographic thought: The South African experience. In Godlewska, A. & Smith, N. (eds.) *Geography and empire*, pp 316 332. Oxford: Blackwell.
Wesso, H. and Parnell, S. (1992). Geographical education in South Africa: Colonial roots and prospects for change. In Rogerson, C. & McCarthy, J. (eds.) *Geography in a changing South Africa. Progress and prospects*, pp 186 199. Cape Town: Oxford University Press.
Withers, C. W. J. (2001). A partial biography: The formalization and institutionalization of geography in Britain since 1887. In Dunbar, G. S. (ed.) *Geography: Discipline, profession and subject since 1870*, pp 79 119. Dordrecht, The Netherlands: Kluwer Academic Publishers.
Withers, C. W. J. and Livingstone, D. (1999). Introduction: On geography and enlightenment. In Livingstone, D. & Withers, C. W. J. (eds.) *Geography and enlightenment*, pp 1 32. Chicago: University of Chicago Press.

Relevant Websites

http://www.amergeog.org/
American Geographical Society.
http://www.socgeo.org/
la Société de Géographie.
http://www.hakluyt.com
The Hakluyt Society.
http://www.rgs.org/
Royal Geographical Society.
http://www.ssag.co.za/
Society of South African Geographers.

Knowledge Communities

S. Pinch, University of Southampton, Southampton, UK

Glossary

Community A much-used term with little specific meaning but usually refers to a social group characterized by dense networks of social interaction reflecting a common set of cultural values. A community is often, but not necessarily, geographically concentrated.

Community of Practice A group of people who form common values, norms, and understanding through their direct engagement in common tasks, often in the same organization. Similar to a knowledge community but involves direct engagement in common tasks.

Epistemic Communities Groups of people who produce new types of knowledge, legitimated by a commonly accepted procedural authority, in pursuit of particular objectives.

Knowledge Community A group of people, typically a professional, technical, or scientific group, unified by a common set of values, norms, and working practices, who produce knowledge for a given purpose.

Networks of Practice Interfirm networks that can arise through workers with similar occupations sharing common understanding but without direct personal contact.

Nonscalar Approach One that attempts to avoid conceptualizing social processes as bounded within particular spaces or levels.

Sociology of Scientific Knowledge A complex and diverse body of work that considers the production of knowledge as a context-dependent social process resulting from the values, norms, and common understanding of communities of scientists and technicians.

Tacit Knowledge Knowledge that is acquired by practical experience, is difficult to codify in written form, and which therefore often requires face-to-face demonstration to facilitate its transmission. May also include knowledge which the individual or group possesses but is not aware of or able to articulate.

What is a Knowledge Community?

Within geography the term knowledge community has been applied most frequently to workers in particular industries who constitute a group of people, often in separate organizations, but united by a common set of norms, values, and understandings, that define the knowledge and production trajectories of the economic sector to which they belong. However, the term know ledge community, together with numerous closely re lated concepts (i.e., various types of community or collectivity – aesthetic, epistemic, interpretative, learn ing, practice based, technical) is more widely applied in the social sciences to understand the knowledge gener ating activities of a wide range of other social groupings. This section will trace the evolution of the concept of knowledge communities, its main characteristics and uses, and some of the major controversies that it has generated within geography.

We should note at the outset that concept of com munity is one of the most widely used and controversial notions in the social sciences. Most work on the idea of community has followed in the wake of the pioneering sociologist Ferdinand Tonnies, who made a distinction between Gemeinschaft – tight knit social relationships based around family and kin, characteristic of traditional, feudal, agrarian societies, and Gesselleschaft – the loose knit relationships manifest in urbanized capitalist soci eties. There followed much debate on the extent to which urban life really did lead to the destruction of close social bonds, with many citing the counterfactual presence of urban villages. With the growth of globalization and new electronic telecommunication systems, there has been debate over the extent to which close knit ties can be maintained across large distances. Hence, the term community has progressively become divorced from its earlier association with local neighborhoods. In particu lar, community has been used as a metaphor to analyze how various organizations operate.

Origins of the Concept of Knowledge Communities

The concept of a knowledge community, together with related notions, represents the coming together of a number of strands of intellectual thought. The most important of these are the sociology of scientific know ledge (SSK), the social ethnography of work, and studies of the relationships between science, technology, and society. Although the various strands of the SSK vary considerably in approach, they share a common desire to avoid technological determinism – the view that tech nology is some independent, external, force acting upon society. Instead, they adopt a social constructivist ap proach that conceptualizes both scientific advances and technological innovations as intrinsically social processes resulting from the operations of communities of scientists and engineers.

Within a constructivist approach the acceptance or rejection of scientific ideas or technological innovations is to be found within the social rather than the natural world. It is argued that there is no natural logic to the world waiting to reveal itself; hence, there is usually enormous uncertainty surrounding new theories or sci entific developments – what is termed the period of in terpretative flexibility. Following this period, scientific and technical communities gradually form a consensus about the meanings of new developments and social closure takes place. However, scientific studies often fail to produce conclusive evidence and so social closure must be secured by the forging of discourses – shared sets of meanings about the nature of the evidence. Accord ingly, truth is not envisaged as something intrinsic to a statement but rather a form of representation that is forged through power struggles between groups of competing knowledge producers. Hence, the creation of knowledge is regarded as essentially a linguistic project, dependent upon forging a common language, identity, and purpose. This means that knowledge is always con text dependent or as is sometimes termed, socially embedded.

Evidence for this interpretation of science is found in the localized nature of initial advances in knowledge, since these usually take place in particular research laboratories and scientific institutes. This geographic concentration is often ascribed to the need for specialized equipment and common understanding that involve a great deal of tacit knowledge (i.e., knowledge that is ac quired by practical experience, is difficult to codify in written form, and which therefore often requires face to face demonstration to facilitate its transmission – also commonly termed 'know how'). For this reason it is sometimes difficult for scientists or engineers in distant locations to replicate experiments in the early stages of scientific advance because they lack the detailed under standings that are forged by the tight knit, innovating, community. Eventually, however, knowledge becomes codified and embodied in standardized equipment that is used in commonly understood ways and it is therefore widely disseminated. These perspectives have led to knowledge being regarded as not so much as a thing or a stock (as in the case of data or information) but as a social process. These social processes that lead to the formu lation of knowledge (or truth claims) are central to understanding how knowledge communities operate.

Such perspectives on knowledge have been applied to other types of knowledge producers including occu pational groups in industry and commerce. Another powerful influence upon the concept of knowledge communities in this context has been the discipline of economics in the form of New Growth Theory. This approach argues that that knowledge is now crucial in promoting economic growth. In particular, new forms of knowledge that are difficult to copy are a key element for competitive advantage for firms in the so called know ledge based economy. The knowledge based view of the firm argues that, rather than being conceptualized as a set of responses to information and transactions costs, as in neoclassical economics, the firm is better seen as a re pository of skills and knowledge.

Communities of Practice

The immediate forerunner to the concept of the know ledge community is the concept of a community of practice. This idea releases the notion of community from a necessarily geographical background, although propinquity may be an element in its constitution. The concept was coined by Jean Lave and Etienne Wenger following their work on how employees learn skills 'on the job' and published in their book *Situated Learning*. Their analysis was based on five remarkably diverse sets of people who all formed a community of practice through their extensive apprenticeships: Goa tailors, meat cutters, nondrinking alcoholics, Yucatan midwives, and US Navy quartermasters. Wenger subsequently popularized the notion, especially in business studies, by his influential book *Communities of Practice*. Wenger drew attention to the widespread processes of learning and identity formation that are fostered by participation in the practices of various social communities including families, workplaces, professional and voluntary organ izations. The crucial element in the formulation of a community of practice therefore is the mutual engage ment of the participants in common tasks – it is not sufficient to merely belong to the same social category, be part of the same social network, or be a member of a traditional community in the sense of a local neighbor hood. Communities must be engaged in knowledge production, much of which is tacit in character through demonstration, experiment, and trial and error. Wenger notes that communities of practice may be diverse in character (as with family members or work colleagues) and need not be consensual; indeed, they may be riven with tensions and conflict (again as in many families or workplaces). Although powerful external forces inevit ably shape these communities, they do, nevertheless, have to negotiate some shared sense of common purpose, together with a mutual recognition of responsibility and accountability. Underpinning a community of practice therefore is what Wenger terms a shared repertoire that includes routines, words, tools, ways of doing things, stories, gestures, symbols, actions, or concepts that the community has produced (see also **Table 1**).

Table 1 Identifying elements of communities of practice

1. Sustained mutual relationships harmonious or conflictual
2. Shared ways of engaging in doing things together
3. Rapid flows of innovation and propagation of innovation
4. Absence of introductory preambles, as if conversations and interactions were, merely the continuation of an ongoing process
5. Very quick set up of a problem to be discussed
6. Substantial overlap in participants' descriptions of who belongs
7. Knowing what others know, what they can do, and what they can contribute to the enterprise
8. Mutually defining identities
9. The ability to assess the appropriateness of actions and products
10. Specific tools, representations, and artefacts
11. Local lore, shared stories, inside jokes
12. Jargon and shortcuts to communication, as well as the ease of producing new ones
13. Certain styles recognized as displaying membership
14. A shared discourse reflecting a certain perspective on the world

Source: Adapted from Wenger, E. (1998). *Communities of practice: Learning, meaning and identity*. Cambridge: Cambridge University Press.

Wenger illuminates these ideas with insights into the work of insurance claims processors. As with many tasks, processing health insurance claims 'by the book' was unworkable – there were simply too many problems, anomalies, and time constraints. Instead, the personnel engaged in processing claims through various social interactions, devising numerous 'quick fixes' and working practices that made the job manageable. These ideas echo Orr's pioneering work on photocopier technicians. The main problem facing this group was that new versions of photocopiers typically produced error statements that did not correspond to the solutions suggested in the early instruction manuals. Solutions to cope with malfunctioning copiers were eventually devised through informal trial and error procedures, often the result 'water cooler talk' with colleagues. Both these case studies of insurance processors and technicians drew attention to how learning is a social process.

Knowledge Communities in Geography

The concept of a knowledge community as operationalized in geography is very similar to that of a community of practice but it has a slightly different emphasis. In particular, there is less emphasis upon mutual engagement in the same tasks, typically in the same organization, and more stress upon an underlying communality of interest among a professional or occupational group within an industrial sector. Thus, in a knowledge community workers might never meet personally, yet, through various means – such as job transfers between companies, professional journals, gossip, and rumour – may come to share a number of common values, working practices, and aspirations. A knowledge community therefore has much in common with Brown and Duguid's notion of networks of practice. These relate to interfirm networks that can arise through workers with similar occupations sharing common elements, but without direct personal contact. There is currently extensive debate in the social sciences over the extent to which professional and occupational groups within firms constitute either communities of practice or wider forms of knowledge community. For example, it has been argued that while in house product development teams within firms may form a community of practice, increasingly more common temporary project teams can constitute more of a knowledge community. The role of Internet based technologies in sustaining virtual knowledge communities is central to these debates.

The concept of knowledge communities can be related to a number of other popular concepts in contemporary social theory. First, there is Anderson's notion of an imagined community, which he used to explain the processes leading to shared notions of national identity among people who never formally meet. Knowledge communities also have links with notions of purified communities whereby some people may be excluded from membership. Professional knowledge communities, for example, frequently try to appropriate key knowledge assets so that they can command power and resources. Knowledge communities may also be similar to transnational communities – international migrants who maintain linkages with more than one nation. For example, there is evidence from Silicon Valley (Palo Alto in California) that immigrant entrepreneurs in the electronics industry maintain various technical and commercial links with their Taiwanese counterparts. Finally, the concept of knowledge communities is also closely related to the notion of epistemic communities. This concept has been used most extensively within the field of international relations to refer to groups of people who produce new types of knowledge, legitimated by a commonly accepted procedural authority, in pursuit of particular social goals. Examples are the diverse international environmental and poverty lobby groups providing new forms of understanding of problems such as climate change and Third World debt relief with the aim of changing policies in these spheres. The crucial element here is the conscription of people toward attaining a common goal. Again, there is less mutual engagement than in a community of practice but the formulation of new ways of thinking, and new ways of validating this knowledge, are crucial. In this sense the epistemic community is related to Knoor Cetina's notion of epistemic cultures used to understand the working of scientific communities. It has been suggested that a community of practice might be more concerned with the use of tacit localized knowledge, whereas an epistemic community (like a knowledge community) is more likely to be involved with spreading codifiable knowledge. However, like most issues relating to knowledge, this is a questionable distinction.

Knowledge Communities and Industrial Agglomeration

The main use of the concept of knowledge communities within geography has been as an explanation of industrial agglomeration – the tendency for firms (usually small and medium sized) to locate together in relatively close proximity in the same region. Initially, these agglomerations (also termed industrial districts or new industrial spaces) were attributed to the post Fordist striving of firms for greater flexibility of production. This led to both vertical and horizontal disintegration (i.e., firms becoming both smaller and more specialized) as they engaged in more flexible relationships of subcontracting and strategic alliances. It was argued that the increased need for interfirm linkages gave advantages to firms who were clustered together; in this way they could help minimize the transaction costs involved in doing business with each other. However, later explanations for agglomerations gave prominence to the covert, informal, untraded exchanges of knowledge between firms which are seen as the essential glue binding them together in relatively close proximity in region. The activities of knowledge communities are seen as an important element in knowledge exchanges.

An Illustration of the Workings of a Knowledge Community

The operation of a knowledge community can be illustrated with the example of racing car engineers in the UK. About three quarters of the world's advanced racing cars are designed in an industrial cluster in southern England, in what has often been dubbed Motor Sport Valley. Building a competitive racing car requires the product to be at the cutting edge of technology within the rules and regulations permitted by the particular racing formulae involved. The firms wish their products to be superior and the engineers therefore work in great secrecy, yet the knowledge of how to construct competitive cars seems to leak out of these firms quite quickly through a wide range of processes. The knowledge can vary from routine, ostensibly mundane, but nevertheless important practices, to radically new ways of doing things. Such is the competition for skilled engineers that there is a great deal of interfirm labor mobility that amounts to staff poaching. Many engineers will have worked with each other at certain times in their careers so they will have some insights into how former colleagues in other companies think. This interfirm mobility is enhanced by the very high rates of firm failure and new firm formation, often in the form of spin offs. However, knowledge is also spread through gossip and rumour, some of which is spread through diverse social networks. Various conventions, stories, and narratives about the best way to do things get established and manifest in the racing cars. These stories are not just 'made up' they have to work on the racetrack. However, given the many factors that can explain variations in car performance, there is typically great uncertainty about the best solutions at any given time. This means that other racing teams will rapidly copy solutions that seem to work in one team, even if they are not fully clear about the scientific or engineering principles involved. As with many other high technology industries, therefore, racing car engineers constitute a small specialized 'world' of production centered on specific sets of knowledge. Although these engineers are usually not in direct contact with each other, they are engaged in a collective project with identical aims and objectives.

To What Extent are Knowledge Communities Scale Dependent?

In Wenger's analysis, communities of practice are constituted by what people do rather than where they are located, although clearly being located in the same office or factory can be important. In contrast, although involving less direct mutual engagement, the empirical applications of knowledge communities in geography have tended to focus upon their operation among a group of organizations in a regional setting. This difference has inevitably raised the issue of the role of geography in the formation of both types of epistemic community. A key debate in contemporary human geography, therefore, centers on the extent to which knowledge communities are rooted in particular places, such as offices, factories, cities, regions, and nations, or are spread out across the globe. One perspective, the 'regionalist approach', argues that certain forms of knowledge and especially that of a tacit character are 'sticky' and difficult to transfer across distances; hence, knowledge communities tend to be localized. The other approach, what might be termed the 'transnational networks perspective', argues that it is relational rather than geographical proximity that is crucial for knowledge generation and transfer – that is, similarity of attitude, disposition, and values. This latter perspective suggests that knowledge communities can be widely dispersed and reflects the desire for a nonscalar approach, that is, one that sees social processes as not confined within particular spaces.

At first glance, there would seem to be a great deal of evidence indicating that knowledge assets can be located at particular scales. For example, at the level of the firm, many have ascribed the rise of the Toyota Motor Corporation in the last 50 years to the distinctive knowledge assets possessed by this company. It has evolved superior knowledge of how to produce an enormous range of commercially successful vehicles in a fashion that many other car companies seek to acquire. If we look to the level of the region, then there is none better than Motor Sport Valley in the UK. While Toyota has been producing millions of cars the small companies in this region have accumulated superior knowledge of how to build very small numbers of successful racing cars. And if we stick with the automotive theme and move to the level of the nation, while the products of the Italian motor industry might not have the reliability of Toyota's products, there is little doubt that, in terms of styling, Italian car designers have left us with many classics. However, further consideration reveals that the spatial dimensions of knowledge are quite complex. For example, although Toyota is heavily rooted in its home region in Japan, it has successfully exported its working practices and knowledge assets to many branch plants and associated suppliers around the globe. In the case of Motor Sport Valley, as well as being generated internally, the knowledge used in this region is imported and reconstituted from all over the world. Finally, the creations of Italian car designers are used by automotive manufactures in many different countries.

One possible resolution of the local versus global aspects of knowledge exchange is to suggest that local exchanges of knowledge (sometimes termed buzz) are more conducive to the spread of radical innovations compared with global pipelines that tend to deal more with incremental changes. However, there is within geography a growing awareness of the diverse configurations of networks through which knowledge of various types can circulate, and a reaction against such correlations of knowledge types with particular scales.

Future Directions in the Study of Knowledge Communities

The concept of the knowledge community is one of the most important in human geography, helping to account for many features of the emerging space economy and sociospatial structure of contemporary society. However, it has been applied in so many different ways to such a variety of scientific, technical, professional, and occupations groups, that the associated plethora of similar concepts raises a number of issues that require further study, both in terms of conceptual clarification and empirical elucidation:

- What types of knowledge are generated, appropriated, and used by these various communities (e.g., scientific or technical, radical or routine, tacit or codifiable, epistemic or utilitarian, etc.)?
- How important are direct personal contact and mutual engagement in common tasks in the creation of knowledge communities?
- What is the relevance of the various network perspectives in social science for understanding the operation of knowledge communities?
- What is the role of advanced telecommunications systems in facilitating the creation and maintenance of knowledge communities compared with direct face to face contact?
- To what extent do knowledge communities operate within and across organizations (e.g., firms, scientific institutes, professions, occupational groupings, cities, regions, nations)?
- To what extent can knowledge communities, their members, knowledge, and activities, be mapped within geographical boundaries and to what extent is either a multiscale or nonscalar approach more appropriate in analyzing these factors?

See also: Actor-Network Theory/Network Geographies; Agglomeration; Competitiveness; Creativity; Discourse; Embeddedness; Habitus; Haptic or Touch-Based Knowledge; High-Tech Industry; Industrial Districts; Informational City; Innovation; Institutionalism/Institutional Geographies; Knowledge Economy; Learning Regions; Local-Global; Network Regions; Networks, Urban; Regional Actors; Regional Development and Noneconomic Factors; Regional Development and Technology; Regional Development, Endogenous.

Further Reading

Amin, A. and Cohendet, P. (2004). *Architectures of knowledge: Firms, capabilities and communities*. Oxford: Oxford University Press.

Bathelt, H., Maskell, P. and Malmberg, A. (2004). Clusters and knowledge: Local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography* 28, 54–79.

Benner, C. (2003). Learning communities in a learning region: The soft infrastructure of cross firm learning networks in Silicon Valley. *Environment and Planning A* 35, 1809–1830.

Brown, J. S. and Duguid, P. (1991). Organizational learning and communities of practice: Towards a unified view of working, learning and innovation. *Organization Science* 2, 40 57.

Brown, J. S. and Duguid, P. (2000). Mysteries of the region: Knowledge dynamics in Silicon Valley. In Lee, C. M., Miller, W., Hancock, M. & Rowen, H. (eds.) *The silicon valley edge: A habitat for innovation and entrepreneurship*. Palo Alto CA: Stanford University Press.

Coe, M. N. and Bunnell, T. G. (2003). 'Spatializing' knowledge communities: Towards a conceptualization of transnational innovation networks. *Global Networks* 3, 435 456.

Haas, P. M. (1992). Introduction: Epistemic communities and international policy coordination. *International Organization* 46, 1037.

Hakanson, L. (2005). Epistemic communities and cluster dynamics: On the role of knowledge in industrial districts. *Industry and Innovation* 12, 433 463.

Hildreth, P. and Kimble, C. (2006). *Knowledge networks*: *Innovation through communities of practice*. Hershey, PA: Idea Group Publishing.

Knorr Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University Press.

Lave, J. and Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.

Lesser, E. L., Fontaine, M. A. and Slusher, J. A. (eds.) (2000). *Knowledge and communities*. Boston, MA: Butterworth Heinemann.

Lindkvist, L. (2006). Knowledge communities and knowledge collectivities: A typology of knowledge work in groups. *Journal of Management Studies* 42, 1189 1210.

Loasby, B. J. (1998). Industrial districts as knowledge communities. In Bellet, M. & Harmet, C. L. (eds.) *Industry, space and competition: The contribution of the economists of the past*, pp 70 84. Cheltenham: Edward Elgar.

Pinch, S. (1998). Knowledge communities, spatial theory and social policy. *Social Policy and Administration* 32, 556 571.

Pinch, S. and Henry, N. (1999). Discursive aspects of technological innovation: The case of the British motor sport industry. *Environment and Planning A* 31, 665 682.

Saxenian, A. and Hsu, J Y. (2001). The Silicon Valley Hsinchu connection: Technical communities and industrial upgrading. *Industrial and Corporate Change* 10, 893 920.

Wenger, E. (1998). *Communities of practice: Learning, meaning and identity*. Cambridge: Cambridge University Press.

Wenger, E., McDermott, R. and Snyder, W. M. (2002). *Cultivating communities of practice*. Boston, MA: Harvard Business School Press.

Relevant Websites

http://www.knowledgecommunities.org
Knowledge Communities.

Knowledge Economy

A. Kerr and S. Ó. Riain, National University of Ireland, Maynooth, Republic of Ireland

One of the most influential approaches to the under standing of contemporary socio spatial organization is that which emphasizes the emergence of a 'knowledge economy'. The knowledge economy highlights the in creasing importance of the informational content of goods and services, the mobilization of knowledge in the production process, the occupational trends toward pro fessionalization, the commercialization of knowledge, and the sectoral shift from an economy based on the manufacturing of industrial goods by production workers to one based on the design of informational goods and services by knowledge workers.

This informational form of capitalism is also linked to socio spatial change and particularly the rise of a global economy, spanning local and national borders and apparently operating on a planetary scale. Furthermore, this global informational capitalism is also characterized by new organizational forms, typically captured under the loose rubric of 'network' forms of organization, and new patterns of socio spatial inequality, linked in part to the 'digital divide' but reaching more broadly into the restructuring of work, employment, and social reproduction.

We examine a number of key questions regarding this knowledge economy. First, we look at the origin of the concept as well as early attempts to define and map the knowledge economy empirically. Second, we examine a variety of perspectives on the socio spatial organization of the knowledge economy and approaches which link techno economic change and socio spatial organization. Building on a critique of these perspectives, we then go on to develop a view of a knowledge economy that is contested along each stage of the process of the pro duction, use, ownership and transformation of know ledge. We show that these struggles occur both globally and locally and are crucial forces shaping contemporary socio spatial organization. Finally, we briefly discuss the emergent patterns of socio spatial inequality associated with this politically constructed knowledge economy.

Origins of the Knowledge Economy: Empirical and Conceptual

For many people the US was the first information/ knowledge economy. This view has its theoretical and empirical origins in the work done by Fritz Machlup and his study of the economic contribution of the knowledge and information industries to the US economy. Machlup recategorized national statistical data and found that information industries accounted for almost 29% of the adjusted Gross National Product (GNP) in 1958 and that information occupations accounted for 32.4% of total employee compensation.

Machlup has been credited with founding the 'eco nomics of information and knowledge' subdiscipline. For him information and knowledge were intrinsically re lated but nevertheless distinct, for while informing can bring about knowledge not all information is knowledge. He defined the production of knowledge as "any human activity effectively designed to create, alter or confirm in a human mind – one's own or anyone else's – a mean ingful apperception, awareness, cognisance or con sciousness of whatever it may be". His taxonomy of knowledge was broad and included knowledge outside the formal economy, although he admitted that he could not measure the contribution of such knowledge to the economy.

Machlup's approach was replicated by others and these early statistical studies of knowledge production highlight significant shifts in terms of employment and work from the late 1940s in the US and at varying times in other OECD countries. For Machlup and many others the growth of information industries and information type jobs was driven by technology, particularly com puter technology. For others the seeds of these changes were sown further back when the technologies of the industrial revolution, like steam and electricity, created a demand for communication and automated control. Overall early mapping studies of the knowledge economy were highly problematic in terms of their methodology and one must be careful when using their results. They also contained little theoretical or explanatory substance.

More qualitative and theoretical work was done in the 1960s and 1970s by sociologists like Daniel Bell. Bell achieved widespread notoriety for this work on the 'post industrial society', a term he was later to use interchangeably with the information and the knowledge society. His examination of the social structure (defined as the economy, the occupation system, and technology) found that between 1947 and 1968 service employment grew by almost 60% while growth in manufacturing remained below 10% in the US. Within service em ployment Bell suggests that automation and technology were having an impact not only on the sectors in which people were working but also on what they were doing. For Bell these shifts signaled a fundamental change whereby economic activities no longer involved man struggling against nature but rather involved man and information. He also argued that theoretical knowledge was of growing importance as a source of innovation and for policy formation. Thus, Bell highlights the growing importance of one type of knowledge, 'theoretical knowledge' in the economy, and for him knowledge is defined as "what is objectively known, an intellectual property, attached to a name or a group of names and certified by copyright or some other form of social rec ognition (e.g. publication)".

Bell's work stands as one of the first attempts to offer an explanation for the development of a post industrial knowledge economy. His explanation rests on the importance of technology and the crucial role of 'theoretical knowledge'. He has been accused of being a 'technological determinist' whose understanding of social change is too dependent on the role of technology and someone whose work was driven by an implicit assumption that economies would converge on the 'most advanced' model if the right technologies were adopted. Indeed, one of Bell's central, and most criticized, points was his division of societies into preindustrial (Asia, Africa, Latin America), industrial (Western Europe, Soviet Union, Japan), and postindustrial (United States). How ever, he brought a vital focus on the specific importance of knowledge, albeit one particular type of knowledge.

Attempts to take techno economic changes seriously were also evident within attempts such as Freeman's to locate information and communications technologies within broader and more complex histories organiszed around successive 'techno economic paradigms', where new technologies became the basis of new products and cycles of investment and production, prompting eco nomic and social re organization. While there is a lin gering technological determinism in these studies, the historical locating of the knowledge economy and the opportunity for comparison with other techno economic paradigms was of great value in introducing a historical dimension that went beyond simple modernization perspectives.

Increasingly, economists and policy makers today believe that technological change and investment in new knowledge are important sources of economic growth and key aspects of a knowledge economy. Patent data provide one indicator as to the level of new knowledge produced in an economy and according to one source it was not until the early 1980s and the 1990s that patent data in the US started to accelerate, particularly in new sectors like biotechnology and computing. In part, this may be due to a new more proactive and pro patent legal and regulatory regime, which by all accounts is another key feature of the knowledge economy, but in part it may also signal increasing investment in research and devel opment in those sectors.

Many economists have become interested in the specific characteristics of information, and endogenous growth theory has placed technological change at the heart of economic growth. These theorists place invest ments in human capital and research and development at the center of economic growth and suggest that these investments generate increasing returns to the spread of technical and business knowledge. Similarly, economists such as Douglas North have picked up Bell's interest in the underlying logic of social transformation that is driving the growth of a knowledge economy. However, rather than focusing on shifting technologies as Bell did, they located the source of the knowledge economy in growing social and organizational complexity, creating a need for greater organizational integration of knowledge to manage transactions.

Machlup's challenge to take the expansion of a knowledge economy seriously has therefore been taken up, with efforts to identify the characteristics, dynamics, and history of such an economy. However, in the process, Machlup's own emphasis on the location of commercial and bureaucratic knowledge within the broader range of knowledges that are produced everyday in social inter actions and institutions was lost. The 'knowledge econ omy' has come to be seen as unproblematically identified with specific industries, occupations, and technologies, rather than as a social and political construct.

Socio-Spatial Organization of the Knowledge Economy

Bell devotes relatively little space to analyzing the rela tionship between knowledge, technology, and space. In recent years, however, the spatial organization of the knowledge economy has become a focus of much debate in human geography and beyond, with a variety of competing concepts of socio spatial organization.

Time–Space Distanciation

Reich argues that new information and communication technologies make it possible and even necessary to re organize firms into 'global webs' and employees into global telecommuters. For Reich these webs operate smoothly, destroying constraints of space and social structure, moving in conjunction with the ever circling hands of the clocks on the walls of corporate offices around the globe. The global workplace is essentially a 'virtual place' where the constraints of space, social or ganization, and local institutional arrangements have been overcome. The politics of the workplace is replaced by a virtual space of knowledge sharing and creation, particularly for 'knowledge workers' such as software developers. Space is disembedded while control of the labor process is essentially cooperative.

Reich's analysis of the process of globalization is compatible with that of Giddens who argues that glob alization occurs through a process of 'time–space dis tanciation' – where time and space are universalized and 'lifted out' or made independent of their immediate contexts. For Reich the global workplace is disembedded from its temporal and spatial contexts and becomes a 'pure' space for communication and innovation among employees. Giddens' analysis casts light on how this is possible. He argues that communication across distance depends upon the existence of expert systems, or systems of knowledge which actors understand and trust (such as the technical language of high tech industry), and upon symbolic tokens, or media of communication that can serve as coordinating mechanisms for long distance so cial relations where social cues and monitoring are absent or opaque (e.g., money). So for these analysts the knowledge economy and its technologies are central to both the content and circulation of the new economy. However, technological determinism once again rears its head in this work. For Reich, symbolic analysts get re warded fairly for their skills and knowledge, and all types of knowledge are evaluated fairly – the politics of the recognition and rewarding of knowledge are neglected. Giddens' emphasis on symbolic media and expert sys tems opens up the possibility of analyzing how the or ganization of these media shape the knowledge economy itself. However, in Giddens' work these are seen simply as carriers of information rather than as contested fields of socio spatial structuring of the knowledge economy.

Time–Space Compression

Other authors have portrayed a fundamentally different global workplace where 'globalism' – whether it is the ability to create virtual connections or to become in creasingly mobile across the globe – is distributed un equally. Harvey argues that time and space have not become distanciated as Giddens claims, but that mobile capital and corporations have colonized local spaces and time has annihilated space in a process of 'time–space compression'. Corporations assert their control over technological change through shaping the trajectory of research and its goals, and protecting their 'intellectual property rights' and monopolies of the means of com munication. The promise of global communications creating an 'electronic republic' is undermined by commericalization, corporate dominance, and media concentration.

Such a perspective rightly draws our attention to the critical importance of corporate power and capitalist social organization in shaping the knowledge economy, particularly given the growing global media oligopolies, the decline of public service media, and the commer icalization of the Internet. However, it blinds us to the ongoing dependence of capitalists themselves upon the knowledge production of workers and citizens, suggesting that there is a much greater field of struggle and con testation within the knowledge economy than these analyses suggest. An emphasis on time–space compression crucially directs our attention to corporate power in the knowledge economy but overstates the degree to which capitalists can secure control over knowledge and underestimates the various struggles and negotiations that are taking place within the knowledge economy.

Time–Space Embedding

Differ though they may on the degree of capitalist domination of the 'space of flows', both of these theories of time–space distanciation and compression agree that globalization is increasingly disembedding the globalized economy and workplace from their local contexts. Other theorists have argued, however, that the demands for increased flexibility and specialized learning in the knowledge economy actually make embedding the global workplace in local spaces even more critical. Efficient production and constant innovation require the con struction of shared physical spaces where workers can interact and communicate on a face to face basis and where shared goals and meanings can be created and maintained. Global competitiveness is therefore dependent on increased spatial embeddedness. The local spaces which can achieve this happy state of affairs will be able to buffer themselves somewhat from the demands of globalism because they possess territorially specific and unique institutional and cultural capacities – although the importance of social interaction will often promote new forms of global organization, such as transnational technical communities. Global workplaces are deeply embedded in these territorially and tempo rally specific contexts under a process that might be called 'time–space embedding'.

However, while recognizing the spatial embeddedness of the knowledge economy is a vital contribution, these perspectives tend to neglect how this spatial embedding is itself shaped by power structures around innovation and knowledge, and focus on the process of innovation and knowledge creation rather than the politics of real ising the returns to knowledge.

The Knowledge Economy as Embedded and Politically Contested

This section argues for a view of the knowledge economy as organized, politically constructed, and contested through local and global networks and accountabilities. We suggest that it is crucial to examine the struggles along the entire chain of technology production and use and argue that two crucial dynamics of the knowledge economy are (1) the struggle between the 'knowledge economy' and the 'knowledge society' over the scope and boundaries of valuable and valued knowledge and knowledge work; and (2) the social shaping of technology production and use within the knowledge economy itself.

How Does the Knowledge Economy Emerge from the Knowledge Society?

We can identify a relatively distinct sphere of the knowledge economy, based around 'knowledge embodied in commercialized media' including silicon chips, software programs, games, media broadcasts, consultancy reports, and other relatively cosmopolitan forms of knowledge. These commericalized forms and media of knowledge, and the activities associated with their production and use, are the core of contemporary capital accumulation in the knowledge economy. However, they remain embedded in a wider process of knowledge production that is intensified by the movement toward occupations based on working on people, relationships, organizations – work that was often in the past done within the household.

At the core of the creation of the knowledge economy is a deep conflict over property rights – with a contest between what Weber and Bussell identify as a regime based on 'property rights imperialism' and one based on a 'shared global digital infrastructure'. In the first, Peter Evans argues, "the Northern corporations that currently dominate global ownership of intangible assets successfully defend their politically protected monopoly rights while simultaneously extending those rights by transforming into private property an ever larger set of ideas, information, and images previously considered part of nature or a shared cultural heritage. In the other scenario, property rights are restructured along the lines pioneered by the open source software community to create a 'new commons' of productive tools which allows for both a more egalitarian redistribution of intangible assets and a wider, more effective engagement of human ingenuity for creating innovative solutions."

The outcome of the conflict between these regimes is not a technologically determinist matter of whether information wants to be 'free' or 'expensive' – it is a matter of political contestation. A crucial component of analyses of the knowledge economy then will be the analysis of this contested terrain of global technology production and business networks, the growth of high tech regions and transnational technical communities, the role of new forms of state developmentalisms and the negotiation of intellectual property regimes through international institutions such as the World Trade Organisation (WTO), the EU's Lisbon Agenda, and the role of the OECD in promoting the concept of the 'new economy'.

What Are the Politics of Knowledge Production and Use?

Robert Reich's 'symbolic analysts' that produce the knowledge economy are as subject to the politics of the workplace as any other workers. These workers work in a variety of settings and their workplaces are subject to a variety of different conflicts – that may, crucially, shape the outcome of their work. These workers work within informal compromises, forged between the technical communities that develop the technological standards and innovations and the lead technology firms that hold the intellectual property rights to many of those innovations. Such technical communities and cultures have long been central to processes of scientific development. However, despite the variety of studies of knowledge production, we lack an ongoing dialogue between students of work, science and technology design that might articulate this more complex perspective on how the conditions of knowledge production shape knowledge itself.

Similarly, social scientists from Adam Smith to Frederick Taylor and on to Harry Braverman have long viewed technology as a source of automation and deskilling, whether for better or worse. However, more complex understandings of technology use have emerged that emphasize how the politics of technology adoption and use shape and effect workers' skill, knowledge and autonomy – and whether that work is in Zuboff's terminology 'automated' or 'informated'. Even within the knowledge economy itself, it is those clerical workers who do the routine testing and processing of programs and data that are most vulnerable to exploitation, while knowledge producers are more likely to combine employee involvement with performance pay structures. For many, these concerns with the social effects of technology in the workplace and beyond have led back to questions of design – and the crucial importance of user involvement in the knowledge production process.

The Politics of Knowledge: Intensifying Time and Space?

Ó Riain argues that the knowledge economy is characterized not by the disembedding of time and space but by 'time–space intensification'. Interactions between exploitation and embedding 'intensify time' with more rapid rates of change in technologies, product markets, and organizational structures; increased turbulence, turnover, and mobility in the labor market; the importance of deadlines in work organization; and a revolution in expectations of time and a shortening of time horizons. The 'intensification of space' is not just a matter of the increased importance of face to face interaction or of competition between places but of the increasing importance of the ability to mobilize space as a resource in social relations in the face of pressures from the cap italist market. While this mobilization of space can pro mote competitiveness in the market, it must also be protected from the pressures of marketization which will erode the very social relations upon which it is based. Ultimately, the intensification of space raises, over and over again in concrete social relations, the deep tension between market and society and opens up political spaces for contesting and reconfiguring capitalist social relations.

New Patterns of Social and Spatial Inequality

Academics and analysts differ both in the role they at tribute to technology in the process of social change and in their assessment of the socio spatial implications of these changes. For Castells while some people in certain countries are networked and connected, many countries are predominantly disconnected from global networks. The continued dominance of developed countries and cities in the global knowledge economy is most evident in the global governance of the internet and the location of the governing body of domain names, ICANN, and 10 of the 13 top level domain name servers in the US. The fact that this situation remained unchanged following considerable opposition at the Tunis round of the World Summit on the Information Society (WSIS) points to continuing inequalities in the global distribution of in formation and knowledge.

While the distribution and control of knowledge is unevenly distributed, it appears that not all economies conform to the trajectory of development assumed by scholars from developed nations and even where there has been a shift to knowledge based industries and oc cupations the results are not always benign. Certain economies still exist which are largely dependent on agriculture and natural raw materials and in other places growth in information industries has been accompanied by significant growth in unemployment and under employment. Indeed, studies of knowledge work have found little change in the types of work or management of work carried out in these new industries despite the new labels. Further, studies point to the negative en vironmental impact of the production, consumption, and dismantling of knowledge goods and highlight the fact that these goods are extremely toxic and polluting.

Despite such inequalities in terms of the global con trol and distribution of knowledge, the concept of the knowledge economy is widely used by corporations, policy makers, and international organizations to legit imate a particular ideological agenda. From the early statistical studies the knowledge economy has been heralded as a stage of development that all countries will or should attain and knowledge itself has come to be valued only in terms of the degree to which it can be codified, measured, owned, and traded. Further, it tends to place the focus on consumption and consumers rather than on citizens and users, on selling and buying rather than on informing and empowering. This is despite the fact that information networks, especially the internet, can empower receivers and decenter producers.

However, producers have been greatly favored by the politics of the knowledge economy. Even dominant lead firms such as Microsoft and Intel were assisted by the global support of the US government in establishing Wintelism as a global standard. Shadlen *et al.* argue that an extensive and effective international institutional regime has emerged in software where the US, WTO, and other related institutions promote the protection of intellectual property rights in a wide range of countries, with uncertain and worrying consequences for innovation and deeply unequal global effects. Ultimately, an imperialist and economistic approach to knowledge production and con sumption not only limits and narrows our understanding of knowledge but threatens to devalue forms of knowledge and knowledge production which do not conform.

See also: Competitiveness; Creativity; High-Tech Industry; Informational City; Innovation; Learning Regions; Regional Development and Technology.

Further Reading

Arriaga, P. (1985). Toward a critique of the information economy. *Media, Culture and Society* 7, 271 296.
Arthur, B. (1994). *Increasing Returns and Path Dependence in the Economy*. Ann Arbor: University of Michigan Press.
Barley, S. and Orr, J. (1997). *Between Craft and Science*. Ithaca: Cornell University Press.
Barnes, B., Bloor, D. and Henry, J. (1996). *Scientific Knowledge: A Sociological Analysis*. Chicago: Chicago University Press.
Bell, D. (1974). *The Coming of Postindustrial Society: A Venture in Social Forecasting*. London: Heinemann.
Benner, C. (2002). *Work in the New Economy*. Oxford: Blackwell.
Carroll, P. (2006). *Science, Culture and Modern State Formation*. Berkeley: University of California Press.
Castells, M. (1997). *The Rise of the Network Society*. Cambridge Mass: Blackwell.
Dodge, M. and Kitchin, R. (2004). *Atlas of Cyberspace*. London: Addison Wellesley.
Evans, P. (2005). The new commons vs the second enclosure movement. *Studies in Comparative International Development* 40, 86 94.
Freeman, C. and Louca, F. (2002). *As Time Goes by: From the Industrial Revolutions to the Information Revolution*. Oxford: Oxford University Press.
Frenkel, S., Korczynski, M., Shire, K. and Tam, M. (1999). *On the Front Line. Organization of Work in the Information Economy*. Ithaca: Cornell University Press.
Giddens, A. (1991). *The Consequences of Modernity*. Oxford: Blackwell.
Godin, B. (2004). The new economy: What the concept owes to the OECD. *Research Policy* 33, 679 690.
Hart, J. and Kim, S. (2001). The global political economy of Wintelism: A new mode of power and governance in the global computer

industry. In Rosenau, J. N. & Singh, J. P. (eds.) *Information Technologies and Global Politics: The Changing Scope of Power and Governance*. Albany, N.Y: SUNY Press.
Harvey, D. (1989). *The Condition of Postmodernity*. Oxford: Blackwell.
Jessop, B. (2000). The state and the contradictions of the knowledge driven economy. In Bryson, J. R., Daniels, P. W., Henry, N. D. & Pollard, J. (eds.) *Knowledge, Space, Economy*. London: Routledge.
Kerr, A. (2006). *The Business and Culture of Digital Games*. London: Sage.
Kumar, K. (2004). From post industrial to post modern society. In Webster, F. (ed.) *The Information Society Reader*, pp 103 120. London: Routledge.
Machlup, F. (1962). *The Production and Distribution of Knowledge in the United States*. Princeton: Princeton University Press.
McChesney, R., Wood, E. M. and Foster, J. B. (eds.) (1997). *Capitalism and the Information Age*. New York: Monthly Review Press.
North, D. (2005). Capitalism and economic growth. In Nee, V. & Swedberg, R. (eds.) *The Economic Sociology of Capitalism*, pp 41 52. Princeton: Princeton University Press.
Ó Riain, S. (2000). Net working for a living: Irish software developers in the global workplace. In Burawoy, M., Blum, J. A., George, S. & Gille, Z. (eds.) *Global Ethnography*. Berkeley: University of California Press.
Ó Riain, S. (2004). *The Politics of High Tech Growth*. Cambridge: Cambridge University Press.
Ó Riain, S. (2006). Time space intensification: Karl Polanyi, the double movement, and global informational capitalism. *Theory and Society* 35(5 6), 507 528.
Oudshoorn, N. and Pinch, T. (2003). *Users Matter*. Cambridge, MA: MIT Press.
Perez, C. (2002). *Technological Revolutions and Financial Capital: The Dynamics of Bubbles and Golden Ages*. Cheltenham, UK: Edward Elgar.
Porat, M. U. (1977). *The Information Economy: Definition and Measurement*. Washington DC: Superintendent of Documents, U.S. Government Printing Office.
Powell, W. and Snellman, K. (2004). The knowledge economy. *Annual Review of Sociology* 30, 199 220.
Preston, P. (2001). *Reshaping Communications*. London: Sage.
Rantanen, T. (2005). The message is the medium. An interview with Manuel Castells. *Global Media and Communciation* 1(2), 135 147.
Reich, R. (1991). *The Work of Nations*. New York: Vintage Books.
Rossiter, N. (2005). The world summit on the information society and organised networks as new civil society movements. In Servaes, J. & Carpentier, N. (eds.) *Towards a Sustainable Information Society*, pp 97 116. Bristol: Intellect.
Saxenian, A. (1994). *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.
Saxenian, A. (2006). *The New Argonauts*. Cambridge, MA: Harvard University Press.
Shadlen, K., Schrank, A. and Kurtz, M. (2005). The political economy of intellectual property protection: The case of software. *International Studies Quarterly* 49(1), 45 71.
Storper, M. (1997). *The Regional World: Territorial Development in a Global Economy*. London: Guilford Press.
Weber, S. and Bussell, J. (2005). Will Information technology reshape the north south asymmetry of power in the global political economy? *Studies in Comparative International Development* 40, 62 84.
Webster, F. (1995). *Theories of the Information Society*. London: Routledge.
Zehle, S. (2001). Virtual water: On the hydropolitics of IT development. Andere Sinema, 160(Fall/Winter) 36 49 (See also http://lists.brunny.com/pipermail/fibreculture/2005 January/004335.html (accessed Oct. 2007)).
Zook, M. (2005). *The Geography of the Internet Industry*. Oxford: Blackwell.
Zuboff, S. (1984). *In the Age of the Smart Machine*. New York: Basic Books.

Relevant Websites

http://incommunicado.info
Incommunicado (Information Technology for Everyone Else).
http://www.icann.org
Internet Corporation for Assigned Names and Numbers (ICANN).
http://www.itu.int
World Summit on the Information Society (WSIS), International Telecommunication Union (ITU).

Knowledge Intensive Business Services

P. Wood, University College London, London, UK

Glossary

Consultancy Business consultancy is a process that provides organizations with outside expertise and advice about many aspects of management. These may include finance; technology; production efficiency and competitiveness; markets, including international development; advertising and graphic design; the law; human resources, including training and recruitment; construction and engineering; and environmental and other public regulations. To be successful, this advice must be fully integrated into client activities, and requires close exchange, or 'co-production', with the client's own experts. Not all outsourcing therefore necessarily involves consultancy. Its interactive nature also makes assessment of the economic impacts of consultancy difficult.

Financial Intermediaries Private institutions dealing in capital exchange, including banks, mortgage companies, building societies, and other credit-granting agencies; insurance and pension fund managers; unit and investment trusts; securities broking and fund management; and the administration of financial markets.

Global City Regions 'Global cities' are major urban centers, such as London, New York, or Tokyo, whose institutions control worldwide patterns of capital exchange, corporate ownership and trade, as well as aspects of cultural and political affairs. This influence has been particularly associated with the activities of international knowledge intensive business services (KIBSs). Global institutions tend to be concentrated into the centers of global cities, but they draw on the resources of their wider urban regions, for example, for skilled labor, international transport links, and high technology research and development.

Information and Communications Technology (ICT) The dominant component of recent technological innovation, focusing on the design and development of computer-based information systems and electronic communications. It has transformed the volume, speed, and geographical reach of economic transactions.

Multinational Companies (or Corporations) (MNCs) Enterprises that manage production or deliver services in many countries. The most dominant, increasingly including KIBS firms, operate in most continents and many countries.

Offshoring Outsourcing by organizations to overseas centers, either to their own branch plants or other agencies. Such moves depend on modern ICT, but the attractions of lower labor and other costs must be balanced against greater operational complexity and higher risks. KIBSs have supported offshoring trends since the late 1990s, both through their own international activities and in advising clients.

Outsourcing An aspect of the division of specialized labor by organizations in which they go to outside agencies to 'buy in' specialized skills, usually within the same country, rather than using their own employees. The growth of KIBS is an important symptom of the modern expansion of such practices.

Small–Medium Enterprises (SMEs) Companies whose employment or turnover falls below defined levels. These vary between studies, but the European Union adopts a standard of up to 50 employees for small companies, and 250 for medium-sized companies. In the USA, limits of 100 and 500 are generally applied.

KIBS Functions

Knowledge intensive business services (KIBSs) are private sector firms that offer specialist professional, business, or technical expertise to other organizations. During the twentieth century, three KIBS functions expanded their influence, at first in the USA and UK and then internationally. Today many major KIBS companies offer all three:

1. Professional functions: What would now be called KIBS firms have provided regular professional advice to clients since at least the nineteenth century. For example, many of today's major accountancy firms trace their origins back to when company law first required independent financial auditing. Businesses offering other forms of advice have also expanded, including specialists in investment, law, property, marketing, and employment practices. In general, they support established client procedures and goals, rather than seeking significantly to change them.
2. Consultancy functions: From the 1920s, some KIBS firms in the USA developed a more interventionist role, promoting scientific management techniques to monitor and direct their clients' commercial performance. They included such pioneering consultancy firms as AD Little, Booz Allen and Hamilton, McKinsey, and Arthur Andersen. Similar services developed in

Europe, and American firms began to move into international markets after World War II. Management consultancy expanded rapidly during the 1960s, often as a profitable extension of professional service firms, especially accountants (e.g., Ernst and Whinney, Arthur Young, Peat Marwick, Price Waterhouse, Coopers and Lybrand, Deloitte, Haskins and Sells). From the 1970s, these functions underwent further rapid expansion as more clients sought advice in adapting to an acceler ating rate and unpredictability of commercial change. Multinational clients looked initially to multinational consultancies, which progressively merged into fewer, larger firms (**Table 1**). Information and communi cations technology (ICT) companies also expanded into consultancy activities to exploit their specialist skills. Further diversification occurred into other in novative sectors, including the financial services, con struction, architecture and planning, marketing and advertising, research and development, and human re source management. This also encouraged a boom in small–medium consultancies, which provided clients with specialist knowledge tailored to their individual needs, the more successful of which expanded into international markets. The rise of business consultancy thus marked a significant development in the creation, diffusion, and application of economic and technolo gical knowledge and expertise. It also transformed the ways in which local, regional, national, and inter national sources of such expertise interact.

Table 1 Leading 20 management/strategy consultancies (ranked by employee numbers, 2000)

Rank	*Consultancy*
1	Accenture
2	CSC[a]
3	Cap Gemini/Ernst & Young
4	Price Waterhouse Coopers
5	KPMG Consulting
6	William M. Mercer
7	Deloite Consulting
9	Andersen
10	Towers Perrin
11	AMS
12	Booz Allen Hamilton
13	McKinsey & Company
14	Watson Wyatt
15	AT Kearney
16	PA Consulting
17	Cambridge Technology
18	Boston Consulting
19	Arthur D Little
20	Bain & Company

[a] CSC is an IT firm but is involved heavily in consultancy related to management issues.

From Jones, A. (2005). Truly global corporations? Theorizing ‘organ izational globalization’ in advanced business services. *Journal of Eco nomic Geography* 5, 177 200.

3. Outsourcing functions: Since the mid 1990s, many major business and computer consultancy firms have developed new markets, managing clients’ outsourcing of routine, but administratively and technically com plex, functions such as data processing, customer re lations, or payroll management. This allows KIBS firms to supplement high cost, project based consult ancy work with more regular, contract based income, and to become tied into closer and longer term col laboration with clients. From around 2000, some of these functions began to move offshore from de veloped economies to countries with educated, low wage workforces, such as India.

Definitions: The Emergence of KIBS

Definitions of KIBS have evolved and become more re fined since the 1970s, often differing from one study to another. They generally rely on sector based classifi cations of firms, nowadays following the International Standard Industrial Classification (ISIC). The notion of KIBSs as a distinct category of activity dates only from the early 1990s. Before this, they were usually regarded as part of either producer services or business services, and were sometimes synonymous with professional ser vices. **Table 2** summarizes two comparisons of producer service activities defined in this earlier research. In 1988, Marshall reviewed five studies published between 1966 and 1983. The numbers in brackets in the table indicate how many of these studies included each service category as a producer service. At this time, only knowledge based financial, accountancy, legal, real estate, advertising, and research and development could be separately identified, but some studies also included material based transpor tation and wholesaling. Other KIBS tended to be placed in residual categories with quite different types of ser vice. As KIBS grew during the 1980s, new activities be came separately identified, as shown in the similar comparison of 11 producer service studies between 1987 and 1995 by Wernerheim and Sharpe. These now often included computer, engineering, architectural, and management consulting services. About half the studies did not include financial and banking, however, because their profits primarily come from returns on financial assets rather than the quality of expertise alone. A mi nority of studies still included materials handling func tions, and a few even counted any activity with some business clientele, including hotels, restaurants, and tra vel companies.

Producer services had at first simply included any service not wholly directed to consumers, usually in cluding materials handling functions as well as know ledge based financial and professional services. The title also implied that their most important role was to

Table 2 Comparative definitions of producer services

No.	*Producer service activities*
11	Accountancy (5); advertising (5); legal service (5).
10	Computer services.
9	Engineering services.
8	Architectural services; employment agencies; management consulting.
6	Financial services (nonbanking) (5); insurance services (5); real estate services (5).
5	Banking services (5); research and development (5).
4	Cleaning and maintenance; market research; security brokerage and services; transportation services (3); typing and copying services.
3	Equipment rental; miscellaneous business services.
1	Equipment repair; hotel and restaurant; purchasing; storage services; travel and accommodation; wholesaling (3).

Numbers of studies classifying listed activities as producer services, 1987 1995: (total = 11)
Source: based on Wernerheim, M. C. and Sharpe, C. (1999). Producer services and the 'mixed market' problem: Some empirical evidence. *Area* 31, 123 140.
(In brackets) Numbers of studies classifying listed activities as producer services, 1996 1983: (total = 5); Source: Marshall, J. N. (1988). *Services and Uneven Development*. Oxford: Oxford University Press.

support manufacturing production, commonly viewed as the main driver of economic growth. Later, business services were defined as activities serving any form of business, including other services, and even public agencies. These also often included operational services, such as recruitment agencies and cleaning, security, secretarial or packaging services that are not primarily knowledge intensive. Emphasis on their knowledge intensity emerged only in the 1990s, with growing interest in sources of high quality skills, especially to support business and technological change. Although financial services were often separated, their expanding consultancy functions increasingly overlapped with other KIBSs.

Nowadays, when employing modern industrial classifications, KIBS studies should if possible include and distinguish between the following sectors:

1. financial services, such as banking, insurance, and other institutions supporting capital management; (ISIC divisions 65–67);
2. professional services offering real estate (division 70), legal, and accountancy advice (classes 7411–7412);
3. business planning and strategy related services such as business and management consulting, market research (classes 7413–7414) and advertising (group 743);
4. computer and data management services (groups 721–724); and
5. technical services, including architecture and engineering consultancy, and independent technical testing (group 742), and research and development services (division 73).

Problems of KIBS Measurement

In spite of their common use, sector based classifications are an inherently unsatisfactory basis for measuring service functions. They were intended mainly to be applied to manufacturing industries, and reflect the technical basis of production, not the markets they serve. Most of the KIBSs categories above, except business planning and strategy related services, therefore serve substantial consumers as well as business markets. They include banks, insurance companies, lawyers, real estate agents, accountants, and even architects and computer services. Such KIBS firms themselves usually make a clear distinction between business and consumer markets. In large firms, they may be served by separate divisions, and they support quite different types of small–medium firms. Sector based data do not generally allow these distinctions to be made, and so cannot accurately measure the scale of truly business oriented services.

It is, however, possible to distinguish KIBS markets using another type of economic evidence, available from input–output tables, although usually only for whole countries, and for all KIBSs together. These tables provide estimates of the annual shares of each sector's outputs in a country going to other sectors (as intermediate consumption), and to consumers, government, exports, and investment (i.e., final consumption). Such tables are now regularly available for most developed economies. For KIBS they demonstrate the scale of their mixed business/consumer markets. For example, **Table 3** shows the major UK markets for so called financial intermediation in 2003. This designation is misleading, since all the KIBSs listed above are included, plus the renting of machinery. Only 62% of such KIBSs output goes to other business and public sectors (as intermediate consumption), while one quarter goes directly to consumer households. This share would certainly be higher for the financial and professional services listed above (generally estimated at over 40%), and lower for the business and technical services. Also note that only about 10% of the direct market for KIBS is in the material processing activities, including manufacturing, construction, primary production, and the utilities, while over 8% goes to public administration, education, health, and the social services. During the previous decade, KIBS markets had also grown fastest among KIBS themselves, and in transport and distribution and the public sector. **Table 4** shows that these were also the most intensive KIBS users. While such evidence may therefore suggest that sector based measures exaggerate the scale of truly business oriented services, it also demonstrates the dynamic role played by KIBS in the modern information driven economy. Further, **Table 3** shows that this role is extending to their international export contribution. More difficult to assess, of course, are their

Table 3 KIBS, UK Markets in other sectors, 2003 (£'000 million at purchasers' prices)

Intermediate consumption by sector:	*2003*	*% of all KIBS*	*% Change 1993 2003 (Highest bold)*
Agriculture	1.7	0.3	106
Mining and quarrying	2.5	0.4	125
Manufacturing	32.4	5.5	147
Utilities	2.5	0.4	192
Construction	18.9	3.2	220
Wholesale & retail	53.2	9.1	**283**
Transport & communications	26.3	4.5	**296**
KIBS ('Financial intermediation, etc.')[a]	129.5	22.1	**290**
Public administration	16.3	2.8	236
Education, heath, & social work	18.2	3.1	249
Other services	16.4	2.8	198
Not allocated to industries[b]	45.9	7.8	234
Total intermediate consumption	*363.8*	*62.0*	*243*
Exports: EU	22.9	3.9	**309**
Non-EU	37.3	6.4	**301**
Gross capital formation, etc.	20.1	3.4	212
Household consumption	143	24.4	184
Total Consumption of UK KIBS	*587.0*	*100.0*	*228*
Total UK Domestic Output	2061.9		177

[a] Financial intermediation, etc. includes banking and insurance, real estate, renting of machinery and equipment (including for households), computer related, legal, accountancy, market research, business, architectural and engineering consultancy, advertising, research and development.

[b] i.e. 'Financial intermediation services indirectly measured' (FISIM), based on differences between the actual interest payable and a 'reference' base rate of interest. It represents the 'risk premium' earned by financial intermediaries above the pure cost of borrowing.

Data from Office for National Statistics (2005). *United Kingdom, Input Output Analysis, 2005 edition*. London: Her Majesty's Stationery Office. http://www.statistics.gov.uk/STATBASE/Product.asp?vlnk=7640

effects on clients, and their wider economic significance, for example, in supporting innovation and competitiveness. This requires more intensive research on samples of KIBS and their clients.

The Modern Growth of KIBS and the Impacts of Consultancy

The remarkable modern growth of KIBS reflects a much wider transformation in the organization, methods, and goals of capitalist business since the 1970s. These have been associated with the speed of technological, market, and other changes, which have increased the costs and risks of training and permanently employing specialist expert staff. Organizations thus examine much more closely whether they can support functions that are not directly involved in their core manufacturing or service production. Many now tend instead to buy in specialists from outside to perform such tasks when required, on a project by project basis. Individual project costs often appear high, but for the client they remain lower than the long term expense of employing their own specialist staff. KIBS firms also benefit from economies of scale in developing specialist expertise, through staff recruitment and training, investment in information systems, and working with many clients. Competition may also stimulate the quality and innovativeness of their expertise. For clients, if outside expertise is effectively managed, its costs and quality may often be easier to monitor,

Table 4 Share of each sector's consumption spent on KIBS, UK, 2003 (£'000 million at purchasers' prices)

Sectors	*Consumption*		*KIBS % of total*
	Total	*KIBS*	
Agriculture	11.9	1.7	14.3
Mining and quarrying	9.3	2.5	26.9
Manufacturing	256.1	32.4	12.7
Utilities	31.8	2.5	7.9
Construction	98.1	18.9	19.3
Wholesale & retail	143.9	53.2	37.0
Transport & communications	89.6	26.3	29.4
KIBS ('Financial intermediation')	205.6	129.5	63.0
Public administration	56.2	16.3	29.0
Education, heath, & social work	85.0	18.2	21.4
Other services	46.9	16.4	35.0

Data from Office for National Statistics (2005). *United Kingdom, Input Output Analysis, 2005 edition*. London: Her Majesty's Stationery Office. http://www.statistics.gov.uk/STATBASE/Product.asp?vlnk=7640 control, and vary as the need arises, than those of em ployed staff.

At first, some argued that such externalization simply transferred services that remained essentially the same to lower cost providers. This may have been the case in some circumstances, especially for routine professional functions. Some growth after the late 1990s was also due to client outsourcing of routine functions. But, even in such cases, achieving the full cost and efficiency benefits would still require improved service quality and client management procedures. Since the 1980s, however, the fastest growing KIBSs have been engaged directly with the sort of strategic financial, managerial, and technical functions that in the past would have been kept strictly within client organizations, and even within their boardrooms. Consultancy has thus penetrated far deeper than simple efficiency based outsourcing, becoming as sociated with progressive changes in the very nature of management. The interaction of consultants and clients also often supports new functions that might never have been developed by the client alone.

A further source of consultancy growth has come from the public sector (**Tables 3** and **4**). In many countries, led by the UK in the 1980s, neoliberal economic policies have required intensive restructuring of public agencies, including their privatization and its aftermath. Major resources have gone into KIBS support to transform them into private companies and adapt their management to more competitive market and financial conditions. Many other public services have also increased their use of outside contractors and consultants in attempts to control costs, improve quality, and monitor outputs. These rela tionships have more recently been extended in the UK to the development of various forms of public–private partnerships for major public investment projects.

International Variations in KIBS Development

From the 1970s, KIBSs in the USA and UK initially tended to grow fastest in the financial and management services, but during the 1990s their work became in creasingly dominated by ICT developments. These sup ported a huge expansion in hardware, software, and systems work by the established consultancy giants. New multinational ICT consultancies also emerged, while many small–medium enterprises (SMEs) spun off com puter hardware and ICT applications companies, where their specialist staff had gained training and experience. These tended to serve particular markets by adapting specialist software and systems applications to their needs, such as financial data processing, information manage ment, web design, or logistical control. Many consultancy processes themselves have also been transformed by ICT. Some KIBSs employment is thus declining or geo graphically dispersing, while demand increases for more innovative skills, usually in concentrated urban regions. ICT developments have also created entirely new mar kets for KIBS, for example, supporting multimedia based entertainment, and e commerce.

These patterns of growth have varied between major developed countries. This reflects different corporate norms, and the regulation of financial, fiscal, and other business services, especially accountancy and legal advice. In Germany and Japan, the traditional self sufficiency of the dominant production and financial conglomerates, and the continuing importance of manufacturing, has focused independent KIBS growth on technical, rather than business process consultancy. Even here, however, the globalization of markets and technologies has ex panded multinational business consultancy, and also supported the development of small–medium KIBS firms. The opening of new geographical markets has boosted demands for KIBS expertise. This was evident in Europe during the 1990s, as Mediterranean countries such as Spain, Portugal, and Greece liberalized their economies. More recently, similar growth has taken place in Eastern Europe, and India and China are now the fastest growing international KIBS markets. Wherever there is rapid economic change, demand for KIBS grows, and the profile of KIBS supply constantly shifts as a consequence.

Problems of KIBS Access and Evaluation

The high costs and corporate focus of mainstream con sultancies generally deter their use by all but the most innovative SMEs. Many SMEs, valuing their own in dependence, also tend to resist seeking advice from con sultants. They mainly rely on routine accountancy, financial, legal, and other professional advice, often from small–medium KIBS firms, although this may extend to wider business problem solving. In many countries, pub licly supported national or regional business advice agencies have been established to encourage marketing, technological, or training innovation among SMEs, al though their services are often difficult to target effect ively. Relations between public agencies and commercial KIBSs also often encounter problems. Many public pro jects, for example, involving the introduction of ICT to tax collection, defense, health, or education administration, require complex and changing specifications, extending over long time scales. This makes it difficult to sustain the clarity of goals and commercial styles of interaction re quired for successful commercial consultancy.

More generally, these essential processes of inter action with client capabilities make the value and ef fectiveness of KIBS particularly difficult to assess. Their worth can only be judged through their impact on clients, but this also depends on how effectively clients commission and implement their expertise. The contribution of consultancy therefore can seldom be independently measured, and it is often easy for clients to blame consultants for perceived failure. Anecdote and hearsay thus sometimes support a poor image of consultancy. Nevertheless, their extraordinary modern expansion, and increasingly common and repeated use by commercial and public organizations throughout the developed world, suggest that they must be having some net beneficial effect for their clients.

KIBS, Global Cities, and Regional Inequality

The growth of KIBS has reinforced the dominance of information and expertise exchange within and between the major global urban regions, including the major cities of the USA, Europe, and Asia, especially New York, Los Angeles, London, Paris, Tokyo, and Hong Kong. Here they interact most intensively with their corporate clients, innovative SMEs, and public agencies. They also often closely collaborate with other professional and consultancy firms. The primary contribution of KIBS is to adapt generic technical and commercial knowledge, often drawn from international experience, to the requirements of particular clients. Success depends on how effectively their skills are adapted to each new project, and global cities offer arenas within which expertise can be moved most flexibly across conventional corporate and sector divides.

The international office networks of the major KIBS firms now largely define global cities. These include banking, finance and insurance, accountancy, commercial law, management consulting, ICT systems and software, engineering and architectural consultancy, and marketing and advertising. Such KIBS MNCs have also expanded into rapidly growing global market regions, including India, China, Southeast Asia, Eastern Europe, and Latin America, apparently reinforcing the control of the established cities over international expertise exchange. By the nature of their work, however, as they have expanded internationally, multinational KIBS firms have had to adapt to more varied client expectations, and local economic, legal, and cultural norms. Most have therefore devolved key client relations, management, and staff recruitment functions to global regional offices. They have also established offshore, computer based operation centers, many in India, to serve their traditional core markets. The need for close contact with clients, and the benefits of offshoring, seems therefore to be loosening the grip of dominant global centers over international KIBS operations. More varied, devolved, and networked forms of international control are emerging, although within management regimes that still attempt to impose centralized staff training and financial control to preserve the distinctive corporate cultures and the reputations of their global brands.

The great majority of KIBS firms, however, are regionally based SMEs, many with modest ambitions to offer professional support to other SMEs, developing familiarity and mutual trust through repeated projects. Some consultancy SMEs are more innovative, but usually depend on major clients. Their potential for growth depends on the quality of this demand, and the wider competitiveness of their regional economies. This means that the highest concentration, variety, and quality of such KIBSs are found in global city regions. Here, access to large business markets encourages specialization and competitiveness. The primary sources of KIBS entrepreneurship are often the experienced former staff of large firms. Working with corporate clients also enhances the reputations of KIBS SMEs, often promoting their growth in wider markets, including internationally.

Outside these dominant city regions KIBS SMEs also tend to be urban focused, but are often constrained by lower levels of less specialized demand. Any significant growth commonly depends on developing interregional and national markets before undertaking international work. A critical mass of specialized regional demand and competition may encourage local KIBSs to specialize and thus compete in wider markets. This may be associated, for example, with specialist manufacturing, concentrations of research and development, high technology functions, primary production, transport nodes, or other tradable services such as finance and tourism.

KIBS Patterns within Cities

Within cities, patterns of KIBS development follow and reinforce wider property development trends. Their need for close contact and regular knowledge exchange with corporate clients attracts major KIBS companies to prestigious city center sites or to major new office based developments nearby, including reclaimed docklands and rail yards. Their main requirements are for modern offices, with access to expert labor drawn from a wide city hinterland. ICT has given such KIBS firms easy access to commercial information and contacts networks. Routine tasks, such as data processing or outsourcing, tend to be decentralized to more peripheral office parks, lower wage regions elsewhere, or even abroad.

KIBS SMEs are usually found in more complex geographical patterns across city regions. Some, serving major companies, locate in high cost city centers, often collaborating with major KIBS firms by offering contacts and close familiarity with local commercial conditions and business cultures. Many others, however, are primarily oriented to the needs of regional businesses, in manu facturing, transport, trade, retailing, or the public sector, while some may also serve national and even international clienteles. They generally require low cost accommoda tion and easy access to other areas within and outside the city. Some even work from suburban homes. Others have also developed in more remote, nonurban regions, where modern communications enable highly qualified single entrepreneurs to establish KIBS firms, maintain contact with clients, and benefit from a good quality of life.

KIBS and Economic Growth

Traditionally, manufacturing has for long been regarded as the basis for economic growth, with services playing a subsidiary role. The modern version of this argument is that manufacturing has a higher capability to improve labor productivity, and thus competitiveness, than ser vices, especially labor intensive KIBS. As we have seen, however, there are inherent problems in measuring the output and productivity of many KIBSs. Their rapid ex pansion has undoubtedly created circumstances in which they lead growth by providing key information, expertise and advice to clients at critical times. Such circumstances depend on the specific division of labor between them and client specialist functions in any project, and the outcome for client competitiveness. In manufacturing, for example, technological innovation may create new products or processes, but the management of their development and their subsequent marketing requires much wider techni cal, financial, organizational, marketing, or training skills. When the staffs of manufacturing companies lack such skills, success may depend on KIBS support. The in novative success of apparently high productivity manu facturing (e.g., in ICT) therefore necessarily depends increasingly on supposedly low productivity KIBS.

We have also seen (**Table 3**) that manufacturing gen erally occupies only a small share of KIBS markets. Most are in other business, professional, consumer, and public services. In these, competitiveness or greater efficiency are just as likely to benefit from the management, marketing, training, and business skills of KIBS, as from the clients' own expertise. This has been particularly the case for the various changes arising from new ICT, where KIBSs have been highly influential in supporting clients across the economy. KIBSs have, of course, adopted ICT innovations themselves but, with other services such as retailing, their major contribution has been in encouraging the success of these innovations in other sectors.

There can also be little doubt about the contribution to economic growth of KIBSs in the urban regions where they are most concentrated. This cumulative concen tration has widened regional economic divides in many countries, especially in the UK, Southern Europe, and many rapidly developing economies. In others, such as the USA, Germany, and Northern Italy, the urban sys tems remain more polycentric. In France, policies to overcome the dominance of Paris have met some success, and cities such as Lyon, Toulouse, and Lille have de veloped as nodes of tradable KIBS. In general, however, KIBS growth nowhere favors the poorest regions.

The clearest recognition of the economic potential of KIBS exchange is to be found in international negoti ations under the World Trade Organization to establish the General Agreement on Trade in Services (GATS) since the early 1990s. As modern communications de velopments continue to foster the international exchange of expertise, whether measurable or unmeasured, service trade will be an increasingly significant subject of such negotiations and the economic development processes they seek to encourage. Meanwhile, this trade increas ingly supports the dominance of globally oriented cities within their own countries. Powerful KIBS supported processes are therefore making it even more difficult to reduce regional inequalities within countries than are uneven distributions of either manufacturing or high technology functions. KIBS have become key agents in the exchange of business and technological expertise between regional, national, and global sources of in telligence. Their activities augment and reinforce trends toward geographical inequality inherent in corporate globalization both within and between nations.

Future Research Agendas

Geographical studies of KIBSs were stimulated in both the USA and UK during late 1980s and early 1990s by their unexpected rapid growth. The focus was on small–medium KIBS firms and the conditions that favored their success, especially in core urban regions. From the late 1990s, research attention shifted to the activities of multinational KIBS corporations which, as we have seen, also tend to favor such regions. Little is still known, however, about how KIBS firms and their clients interact at various geographical scales, and the consequences for uneven economic development in and between many countries. What is the significance of KIBSs in supporting regional innovation and competitiveness? How influential are they in speeding the pace of economic change, and thus creating problems and opportunities for community adaptation? As major players in the circulation of busi ness expertise, when do KIBSs follow or lead change? Such processes are increasingly important for the ways in which expertise is brought to bear on all forms of eco nomic challenge, from global competition to training labor forces, or adapting to environmental change. The difficulties of assessing the operations of KIBS means that their economic significance remains obscure, and yet their role is likely to grow as the application of special ized expertise becomes everywhere increasingly central to economic development.

See also: Business Services; Finance, Offshore; Financial Centers, International; Financial Knowledge; Globalization and Transnational Corporations; Information Technology; Innovation; Knowledge Communities; Neoliberal Economic Strategies; Networks; Privatization; Services, Professional; Spatial Division of Labor; Telecommunications; Transnationalism and Technological Transfer; Uneven Development.

Further Reading

Allen, J. (1992). Services and the UK space economy: Regionalization and economic dislocation. *Transactions, Institute of British Geographers* NS 17, 292 305.

Castells, M. (1989). *The Informational City: Information Technology, Economic Restructuring, and the Urban Regional Process*. Oxford: Blackwell.

Gadrey, J. and Gallouj, F. (eds.) (2002). *Productivity, Innovation and Knowledge in Services. New Economic and Socio Economic Approaches*. Cheltenham: Edward Elgar.

Gallouj, F. (2002). *Innovation in the Service Economy: The New Wealth of Nations*. Cheltenham: Edward Elgar.

Gluckler, J. (2006). A relational assessment of international market entry in management consulting. *Journal of Economic Geography* 6, 369 393.

Jones, A. (2005). Truly global corporations? Theorizing 'organizational globalization' in advanced business services. *Journal of Economic Geography* 5, 177 200.

Lindahl, D. P. and Beyers, W. B. (1999). The creation of competitive advantage by producer service establishments. *Economic Geography* 75, 1 20.

Marshall, J. N. (1988). *Services and Uneven Development*. Oxford: Oxford University Press.

Riddle, D. (1985). *Service Led Growth: The Role of the Service Sector in World Development*. New York: Praeger.

Sassen, S. (ed.) (2002). *Global Networks, Linked Cities*. London: Routledge.

Taylor, P. (2003). *World City Networks: A global Urban Analysis*. London: Routledge.

Wernerheim, M. C. and Sharpe, C. (1999). Producer services and the 'mixed market' problem: Some empirical evidence. *Area* 31, 123 140.

Wood, P. (ed.) (2002). *Consultancy and Innovation: The Business Service Revolution in Europe*. London: Routledge.

Wood, P. (2006). Urban development and knowledge intensive business services: Too many unanswered questions? *Growth and Change* 37(3), 335 361.

Wood, P. A. (1996). Business services, the management of change, and regional development in the UK: A corporate client perspective. *Transactions Institute of British Geographers* NS 21, 649 665.

Relevant Websites

http://www.amcf.org/index.asp
Association of Management Consulting Firms.

http://www.cric.ac.uk/cric/compprojects/project 6.htm
Centre for Research on Innovation and Competition.

http://www.lboro.ac.uk/gawc/rb/rb5.html
Globalization and World Cities Study Group and Network.

http://www.ibconsulting.org.uk/Landing Page 1.aspx?id=10:2396
Institute of Business Consulting.

http://www.nesta.org.uk/taking services seriously/
National Endowments for Science, Technology, and Arts "Taking Services Seriously".

Kolossov, V.

J. O'Loughlin and E. Holland, University of Colorado, Boulder, CO, USA

Glossary

Eurasianism A political and cultural ideology emerging in Russia in the nineteenth century and popular especially among Russia *émigrés* in the early twentieth century after the communist state was established. It posits that Russia is neither a European nor Asian state but has an ethnos and culture that has emerged from its geographical location and vast extent. It is now closely associated with political movements and ideologies that are anti-Western and especially anti-US, which it views as engaged in a hegemonic action to control Russia.

Limology The study of political borders.

Perestroika Literally 'restructuring' and refers to the attempts by Soviet President Mikhail Gorbachev to reinvigorate the communist economy between 1985 and 1990. The 'perestroika' years constituted the last half decade of the Soviet Union.

Pseudo-States (also Known as Quasi-States) Self-declared independent or autonomous political units that emerge from civil war ceasefires. Controlled usually by a territorially concentrated ethnic minority and supported by a strong external power, these entities have many of the trappings of statehood (flags, armies, police, border controls, public services, etc.) but little international recognition. Tensions with the state from which they are separated are tense and borders heavily controlled. Contemporary examples include Transdniester Moldovan Republic (Moldova), Abkhazia (Georgia), South Ossetia (Georgia), and Turkish Republic of Northern Cyprus (Cyprus).

Slavophilism It is based on a pan-Slavic notion that there exists a singular Slavic ethnoculture with unique values and principles. Though strongest in Russia after its emergence in the mid-nineteenth century, it had adherents in most Slavic states and was seen as the basis for political alliances and opposition to neighboring states (Ottoman Empire and Germany).

Kolossov, Vladimir (1953–)

Vladimir Kolossov has played a leading role in the de velopment of political geography in the international academic community for over a quarter of a century. A versatile and gifted scholar, Kolossov has contributed more than 250 published works addressing broad intel lectual questions within the field of political geography, doing so with specific reference to Soviet/post Soviet territories. Just as importantly, Kolossov has made a sig nificant and lasting contribution to the development of political geography within Russia. In this capacity, he is recognized not only for his scholarly contributions, but also for his work in promoting political geography within domestic academic circles as well as serving a mentor and teacher to a new generation of Russian political geographers.

Kolossov received his academic training at Moscow State University, matriculating into the department of geography in 1970. He completed his doctorate in 1979, writing on the electoral geography of Italy, and earned his *Habilitation*, also in geography, from the Russian Academy of Sciences in 1992. He has held his position as head of the Center for Geopolitical Studies at the Academy of Sciences since earning this postdoctoral degree. He is currently (2008) vice president of the International Geographical Union (IGU).

Kolossov's work is truly international in scope. It has been published in his native Russian, as well as English, French, and Italian. Much of his research in Russia has focused on explicating the principles of political geog raphy for Russia's domestic academic audiences. Kolossov has then applied these principles to geopolitical questions facing contemporary Russia. Books on geopolitics are quite popular in Russia, specifically when they focus on the country's position in the post–Cold War order, and Kolossov has addressed topics similar to those written by more populist authors by primarily writing for the aca demic community. In contrast, his work in other lan guages has been broader in scope, discussing themes both important and emergent themes in political geography. While Russia and its environs serve as the primary spatial context for such research, the works' theoretical signifi cance is equal to the contribution it makes to an em pirical, region specific literature.

Kolossov's contributions as a scholar are not limited to his research. He has contributed to political geography in a variety of forums. In Russia, Kolossov plays a leading role in the field, serving as head of the Center of Geo political Studies of the Institute of Geography for the Russian Academy of Sciences, founded in the 1990s. Functioning as an administrative center in the Academy with a number of specialists and doctoral students working on the political geography of Russia and the near abroad (the countries of the former Soviet Union), the center also acts as a focus for the still new science of political geography in Russia. Kolossov has written or edited six general texts on the post Soviet political geography scene, one published in the UK, two in France, one in Italy, and two in Russia. These works supplement and distill his earlier endeavors to broaden the scope of politico geographic studies and their acceptance within the Russian academic community. As a prominent international academic, Kolossov serves on the editorial boards of a number of political geography journals. Moreover, he has played the role of ambassador for the discipline, holding short term academic positions in the UK, France, Italy, Germany, the Netherlands, and the US. Through these exchanges he has forged strong aca demic partnerships with scholars outside of Russia, contributing to the development of political geography through collaborative research. Kolossov also served two terms as Chair of the Commission on Political Geog raphy of the international Geographical Union (IGU) from 1996–2004, as well as vice president of the IGU.

From the beginning of his career, Kolossov has played an important role in the development of political geo graphy in Russia. His early research, starting during the mid 1980s, was situated in the Soviet context, and reflected the political, ideological, and academic limitations imposed during that period. Despite these constraints, this work laid the foundation for Kolossov's wider project, which has been aimed at promoting and cultivating political geography in the Soviet Union and in Russia after 1991. As its foremost scholar, Kolossov's career mirrors the growth of political geography in this former communist country.

To advance political geography in the Soviet Union, Kolossov's early work specifically focused on situating political geography within the context of the Soviet sys tem. In his 1988 monograph, *Political Geography: Problems and Methods*, Kolossov conceptualized political geography as a scientific and academic discipline of relevance to Soviet scholarship. Before explaining some of the primary principles accepted and studied by political geographers, he set out to establish an explicit definition of political geography that emphasized its position within human geographical studies. He made specific reference to the importance of the spatial organization of political systems and the territorial distribution of political power and the economic means of production. Topics such as the role of national governments in the world system, the geo graphical bases of federal states, and electoral geography were then discussed in greater detail. In evaluating the contribution of this work, it should be emphasized that the book was produced in the last years of the Soviet Union. As such, Kolossov was constrained by party norms and regulations, and thereby focuses on the connection between political geography and Marxism. His other early works, aimed at noncommunist audiences, were also constrained by this closed political environment; for example, writing in the journal *Soviet Geography* in 1985, Kolossov discussed the need for electoral geography to integrate a Marxist approach, yet situated this argument in Marxist theory as interpreted by the communist state. Despite the ideological and academic restrictions, his overarching intellectual focus remained centered on the promotion of political geography in the Soviet Union.

In spite of his academically stifling surroundings, Kolossov grew in stature as a scholar interested in the changing politics of his own society and the changing world political map at the end of the Cold War. With the development of greater political openness in the last dec ade of the Soviet Union, his work distinguished itself from the entrenched paradigms imposed on Soviet scholarship and began to approach more intellectually independent research. The movement toward open elections in the USSR, spurred by nationalism among ethnic minorities and the resultant violence in places such as Nagorno Karabakh (an ethnic Armenian territory within the re public of Azerbaijan), galvanized not only the national communities within the USSR but also the whole tradition of intellectual scholarship in political geography.

As the changing intellectual and political climate of the Soviet Union allowed in depth analysis of political choices for the first time, Kolossov and his colleagues at Moscow State University began work focused on specific political and electoral topics. The March 1989 parlia mentary elections served as the basis for Kolossov's most important early academic project, and, after researching and analyzing the results of the elections, Kolossov and his colleagues Nikolai Petrov and Leonid Smirnyagin collaborated to produce *Vesna '89 (Spring '89)*. This mono graph addressed the electoral geography of the March 1989 Soviet elections, with specific focus on the regional differentiation in electoral results, and with respect to the nascent position of democracy within the Soviet system. For the first time, data and results were available that illustrated the political will of the Soviet populace. These elections represented the first, tentative steps toward a democratic political system (though communists still dominated among the successful candidates).

Spring '89 engendered a number of questions from the politico geographic perspective. Fundamentally, what was the inherent importance of such elections for the Soviet Union? Could the data and results from these elections be used to precisely explicate the spatial vari ation of the *perestroika* political culture? What were the broader ramifications of these elections; were they the first step in the broader development of a democratic capacity within Soviet society? Using politico geographic methodologies, Kolossov and his co editors concluded that there was significant regional variation in electorates across the Soviet space, exhibited in each of the 11 electoral districts. This variation is connected to three implicit trends in voting behavior, associated with geo graphical and spatial differences between north and south and east and west, as well as the urban–rural divide. And while these conclusions are relevant, the historical value of this work is more significant; *Spring '89* was a definitive step toward the legitimization of political geography in the post Soviet academy.

The idea that the March 1989 elections were indica tive of a broader political liberalization was borne out in the following years. For Vladimir Kolossov, the fall of communism opened up a variety of intellectual avenues as the post Soviet space has proved to be fertile ground for political geography research, both in terms of relevant topics and scholarly freedom. Since 1991, Kolossov's *oeuvre* has addressed an array of topics of relevance in current politico geographic studies. This body of work is most clearly divided into two categories: work aimed specifically at the Russian domestic audience and re search on wider intellectual themes within political geography as they pertain to the post Soviet space. Re search in the former category has been centered on questions of significance to both the Russian academic community and the Russian populace more generally; these relate to Russia's position as a geopolitical power within the post–Cold War order. At the same time, this work continues to underscore the importance of political geography by forwarding the academic development of the field in Russia. When examining more theoretical topics, Kolossov has emphasized subjects that are highly topical in post–Cold War political geography. In this literature, aimed at Western European and American academic audiences, Kolossov has discussed themes such as ethnic minorities, boundaries, and continued his work on Russian national identity.

Vladimir Kolossov's work that has been produced domestically for both scholars and average citizens has centered on Russia's post–Cold War geopolitical position and forms of civic and political identity within the state. Kolossov and his co editors first discussed the nature of political identity in *Spring '89*, with specific reference to the Soviet Union. In this work, they argued that the elections provided a medium for exploring how geog raphy affects political processes in the Soviet Union. In particular, the work made a case for significant regional differentiation within the Soviet state, and aggregates these differences into three distinct politico geographic subcultures, the Asian, the Western, and the Russian. In the Asian sphere, there is no historical basis for demo cratic government; there, political power has taken on a formal, functional structure. The Western subculture is small but influential, and is constructed around the pol itical viewpoints of the Baltic republics and the western portions of Belarus and Ukraine, where the traditions of democracy and political independence are more formally developed. These two traditions are, however, peripheral to the Russian political subculture. This outlook, ac cording to Kolossov, occupies a middle ground between the Asian and Western constructs of political society, where personal independence and populist dependence both play important roles. This conceptualization of Soviet political society was transferred over the next two decades to consider Russia in its post–Cold War context, with Kolossov recently extending this project to specif ically address political identification in Russia in two recent works: *Russia's Geopolitical Situation* and *The World through Russian Eyes*. Kolossov's international publications (mostly in French and English with a variety of col leagues) remained centered on post Soviet space, while addressing broader theoretical questions of relevance to the field of political geography. This research can be loosely grouped into three primary areas: boundaries and border studies; identity formation in Russia; and ethnic minorities in changing political environments.

Since the breakup of the Soviet state, Kolossov has approached questions pertaining to boundaries from a number of theoretical perspectives. While holding a visiting scholar position at the University of Durham with the International Boundaries Research Unit in 1992, Kolossov collaborated with Olga Glezer and Nikolai Petrov on an initial study of boundaries and conflict in the post Soviet space. This work investigated the inter play between geography and borders, with specific ref erence to how these boundaries reified territorial based identities and supported ethnic conflict. It also served as a springboard for future research in border studies. A major commitment of his research since the 1992 project has been investigated border identities, border definitions, and border impacts and the interplay between such conceptualizations and state formation in Russia and its adjoining former Soviet neighbors. Collaborating with John O'Loughlin (University of Colorado), he has in vestigated territoriality and borders in the post Soviet spaces of Ukraine and Moldova using a world systems theory framework augmented by field work in these rapidly changing environments. Given that boundaries have been increasingly approached as a theoretical field of investigation in political geography during the last 10 years, Kolossov's work underscores an important trend in the field. In 2005, he continued to promote the study of emerging borders and summarized the state of the field by conceptualizing limology (border studies) as a suc cession of stages characterized by specific content and achievement (historical evolution, typologies of borders, a functional approach to borders, the social represen tation of borders, and state policies and management of border regions).

Kolossov has further extended the study of Russian political identity to the Western European and American academic audiences. This work has centered on Russian approaches to geopolitics in the context of the post 9/11 geopolitical environment, investigating through survey work how Russian citizens see their country on the world stage and how they view the US led 'war on terror' since

2002. His research, in collaboration with John O'Loughlin and Gerard Toal, has also served to more completely develop the ideas of Russian public intellectuals who have addressed the thread of Eurasianism that emerged in the nineteenth century and became evident again in Russian politics during the 1990s, as the country was forced to reposition itself geopolitically on the world stage in the wake of the erosion of its great power status.

Other threads in Kolossov's recent and ongoing research have centered on the position of ethnic minorities in Eastern Europe and the former Soviet Union. This is a broadly scoped research agenda, including projects on the relation between ethnic identity and conflict in Transdniestria (Moldova) and post–civil war identities in Russia, especially in the tumultuous zone of the North Caucasus. The focus of the work is to identify the territorial bases for nationalist mobilization and to probe the conditions for separatist demands. Further, the work describes the 'in between' nature of new political territories that have emerged as a result of the frozen conflicts in the space of the former Soviet Union, neither controlled completely by the states of which they are nominally a part nor recognized by the world community. Based on his fieldwork, a key part of his approach to the study of borders and nationalism, Kolossov has promoted the idea of 'pseudo states' as a new concept in world politics that evolve from the interplay of the local demand for autonomy and geopolitical realities surrounding the aims of the great powers.

Vladimir Kolossov is one of the leading scholars in international political geography. His contributions, both in terms of research and service in a variety of national domains and in the journals and publications of numerous countries with many collaborators, have significantly affected and promoted the development of the field in the post–Cold War decades. But undoubtedly, his most significant role has been to establish and legitimize political geography as an academic discipline in Russia after its disappearance for much of the 70 years of Soviet rule.

See also: Borderlands; Critical Geopolitics; Ethnic Conflict; Political Boundaries.

Further Reading

Brunet, R., Ekert, D. and Kolossov, V. (1995). *Atlas de la Russie et des Pays Proches*. Paris: RECLUS La Documentation Française.

Kolossov, V. (1985). Political geography in the service of the bourgeois state. *Soviet Geography* 26(2), 98 107.

Kolossov, V. (ed.) (2000). *Geopoliticheskoe Polozhenie Rossii: Predstavleniya i Real'nost' (Geopolitical Situation of Russia: Representation and Reality)*. Moscow: Art Kur'er.

Kolossov, V. (ed.) (2001). *Collocazione Geopolitica della Russia: Rappresentazioni e Realtà (The Geopolitical Positioning of Russia: Depictions and Reality)*. Torino: Edizioni della Fondazione Giovanni Agnelli.

Kolossov, V. (ed.) (2003). *Mir Glazami Rossiyan (The World through Russian Eyes)*. Moscow: Institute of the Fund for Public Opinion.

Kolossov, V. (2004). American geography and the rise of political geography in Russia. *GeoJournal* 59(1), 59 62.

Kolossov, V. (2005). Border studies: Changing perspectives and theoretical approaches. *Geopolitics* 10(4), 606 632.

Kolossov, V. and Mironenko, N. (2001). *Geopolitika i Politicheskaya Geografia (Geopolitics and Political Geography)*. Moscow: Aspekt Press.

Kolossov, V., Glezer, O. and Petrov, N. (1992). *Ethnoterritorial Conflicts in the Former USSR*. Durham: IBRU Press.

Kolossov, V. and O'Loughlin, J. (1998). New borders for new world orders: Territorialities at the *fin de siècle*. *Geojournal* 44, 259 273.

Kolossov, V. A. (1988). *Politicheskaya Geografia: Problemy i Metody (Political Geography: Problems and Methods)*. Leningrad: Nauka, Leningradskoe otd nie.

Kolossov, V. A., Petrov, N. V. and Smirniagin, L. V. (eds.) (1990). *Vesna 89: Geografia i anatomia parlamentskikh vyborov (Spring '89: The Geography and the Anatomy of Parliamentary Elections)*. Moscow: Progress.

O'Loughlin, J., Ó Tuathail, G. and Kolossov, V. (2004). Russian geopolitical storylines and public opinion in the wake of 9 11. *Communist and Post Communist Studies* 37, 281 318.

Relevant Websites

http://www.homeofgeography.org
IGU Home of Geography.

http://www.colorado.edu/ibs/waroutcomes/
The Dynamics of Civil War Outcomes in Bosnia and the North Caucasus.

http://www.cas.muohio.edu/igu cpg/
The International Geographical Union Commission on Political Geography.

Kriging and Variogram Models

C. A. Calder and N. Cressie, The Ohio State University, Columbus, OH, USA

Glossary

Cokriging Extension of kriging methods for multivariate geostatistical data.

Geostatistics Class of statistical methods for the analysis of data generated from a continuously-indexed spatial process.

Intrinsically Stationary Process A spatial process with a constant mean function that has the property that the variance of the difference of the process at any two locations is a function of the displacement between the two locations.

Isotropic A property of a covariance function that indicates that the covariance of a process at any two locations is only a function of the distance between the two locations.

Non-Negative-Definite A necessary and sufficient condition for a real-valued function to be a valid covariance function.

Nugget Effect A non-negative quantity consisting of the measurement-error variance and any additional microscale variation that cannot be attributed to measurement error.

Second-Order Stationary Process A spatial process with a constant mean function and stationary covariance function.

Separability A property of a space-time covariance function that indicates that the covariance function can be factored into the product of a spatial covariance function and a temporal covariance function.

Stationarity A property of a covariance function that indicates that the covariance of a process at any two locations is only a function of the displacement between the two location.

Introduction

Kriging and variogram models are a set of widely used statistical tools for spatial prediction and interpolation. These tools are the fundamental components of a class of spatial statistical methods known as geostatistics. Geostatistics has its origins in the mining industry and early methodological developments were motivated by applications such as estimating the amount of metal in an orebody and how to selectively mine the orebody based on a set of observations at known locations. One of the earliest papers in the application of statistics to mining was written in 1951 by a South African mining engineer named D. G. Krige. In the late 1950s and early 1960s, G. Matheron, of the French Ecoles des Mines, developed spatial prediction and named the methodology after Krige. Further methodological advances were motivated by a variety of other areas of application, including analyses of environmental and agricultural data.

Broadly, geostatistics refers to the class of statistical methods for the analysis of data generated from a continuously indexed spatial process $Z(\boldsymbol{s})$, for $\boldsymbol{s} \in \mathscr{D} \subset \mathbb{R}^{d}$ defined on a domain $\mathscr{D} \subset \mathbb{R}^{d}$. Observed observations of the process $Z(\cdot)$ at a set of known locations $\{\boldsymbol{s}_i : i = 1, \ldots, n\}$ are called geostatistical data. The geostatistical framework is appropriate for the analysis of data such as atmospheric concentrations of a pollutant measured at a fixed set of monitoring locations across a city. Such data can be viewed as the observations at a known set of locations of the continuous pollution process, which is defined over the entire city. The primary goal of a geostatistical analysis is often prediction of the spatial process at locations where the process is not observed. Spatial prediction can be performed on a fine grid covering the spatial region of interest and, therefore, can be used to produce an exhaustive map of predicted values. Given the nature of spatially dependent processes, geostatistics based predictions draw more heavily on nearby observations than on observations farther away. In addition, the uncertainty measure associated with the prediction of the process at an unobserved location depends on the proximity of observed values of the process. We note that the probabilistic nature of geostatistical methodology distinguishes it from other more *ad hoc* methods of performing spatial prediction/interpolation.

In the remaining parts of this article, we first introduce the methodology known as kriging, which defines a class of optimal spatial predictors. Then, we review techniques for estimating the covariance structure of a spatial process, which is needed for kriging. Sometimes it is easier to estimate a variogram, which can be used for kriging in a similar manner. Finally, we present an analysis of rainfall data and briefly introduce several extensions of the standard geostatistical methodology.

Kriging

Consider a set of random variables $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathscr{D} \subset \mathbb{R}^{d}\}$, which we call a (random) spatial process. We define the mean function of the process $Z(\cdot)$ to be $E[Z(\boldsymbol{s})] \equiv \mu_Z(\boldsymbol{s})$,

$\boldsymbol{s} \in \mathscr{D}$, and the covariance function of $Z(\cdot)$ to be

$$C_Z(\boldsymbol{s}_i, \boldsymbol{s}_j) \equiv \mathrm{cov}(Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j))$$

for any $\boldsymbol{s}_i$ and $\boldsymbol{s}_j \in \mathscr{D}$. Note that both functions exist if we assume that var $(Z(\boldsymbol{s})) < \infty$, for all $\boldsymbol{s} \in \mathscr{D}$.

Spatial processes of the sort given above form the basis for statistical analyses of geostatistical data. For example, let $\boldsymbol{Z} = (\boldsymbol{Z}(\boldsymbol{s}_1, \ldots, \boldsymbol{Z}(\boldsymbol{s}_n))')$ be an $n \times 1$ vector of observations associated with known locations $\{\boldsymbol{s}_i : i = 1, \ldots, n\}$. A classical geostatistical analysis of such data would seek to characterize the mean, $\mu_Z(\cdot)$, and covariance, $C_Z(\cdot, \cdot)$, of $Z(\cdot)$ in order to obtain optimal predictors of the process at unobserved locations $\{\boldsymbol{s}_0 \in \mathscr{D} : \boldsymbol{s}_0 \neq \boldsymbol{s}_i;\ \ i = 1, \ldots, n\}$. In the following discussion, we develop kriging in the more general situation where the data are values of a spatial process $Z(\cdot)$ that can be represented as the sum of a latent, or unobserved, spatial process of primary interest and a non spatial measurement error process. In this situation, the value of $Z(\cdot)$ at any location $\boldsymbol{s} \in \mathscr{D}$ can be written as

$$Z(\boldsymbol{s}) = Y(\boldsymbol{s}) + \varepsilon(\boldsymbol{s}) \qquad [1]$$

Analogously to the process $Z(\cdot)$, we let $E[Y(\boldsymbol{s})] \equiv \mu_Y(\boldsymbol{s})$, $\boldsymbol{s} \in \mathscr{D}$, and

$$C_Y(\boldsymbol{s}_i, \boldsymbol{s}_j) \equiv \mathrm{cov}\big(Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)\big), \boldsymbol{s}_i, \boldsymbol{s}_j \in \mathscr{D}$$

for the process $Y(\cdot)$. The measurement error process, $\varepsilon(\cdot)$, is assumed to have zero mean so that $E[\varepsilon(\boldsymbol{s})] = 0$, for all $\boldsymbol{s} \in \mathscr{D}$. We also assume that the $\varepsilon(\cdot)$ process is independent across space, implying that for all $\boldsymbol{s}_i, \boldsymbol{s}_j \in \mathscr{D}$,

$$\mathrm{cov}(\varepsilon(\boldsymbol{s}_i), \varepsilon(\boldsymbol{s}_j)) = \begin{cases} 0, & \text{for } \boldsymbol{s}_i \neq \boldsymbol{s}_j \\ \sigma_\varepsilon^2, & \text{for } \boldsymbol{s}_i = \boldsymbol{s}_j \end{cases}$$

where $0 \leq \sigma_\varepsilon^2 < \infty$. In addition, we assume that the processes $Y(\cdot)$ and $\varepsilon(\cdot)$ are independent.

In this more general problem formulation, the methodology known as kriging refers to the optimal linear prediction of the process $Y(\cdot)$ at a location $\boldsymbol{s}_0 \in \mathscr{D}$, given observations $\boldsymbol{Z} = (Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n))'$, under the assumption that $C_Y(\cdot, \cdot)$ and σ_ε^2 are known. In deriving the optimal predictor of $Y(\boldsymbol{s}_0)$, we first consider a special case known as simple kriging, in which $\mu_Y(\cdot)$, the mean of the $Y(\cdot)$ process, is also assumed to be known. Because the process $\varepsilon(\cdot)$ has zero mean, it should be noted that $\mu_Y(\cdot) = \mu_Z(\cdot)$.

The simple kriging predictor of $Y(\boldsymbol{s}_0)$, $\hat{Y}^{(sk)}(\boldsymbol{s}_0)$, is optimal in the sense that it minimizes the mean squared prediction error (MSPE), where

$$\mathrm{MSPE} \equiv E\Big[\big(Y(\boldsymbol{s}_0) \quad \hat{Y}(\boldsymbol{s}_0)\big)^2\Big]$$

over the class of all heterogeneously linear predictors.

$$\left\{ \hat{Y}(\boldsymbol{s}_0) = \lambda_0 + \sum_{i\ 1}^{n} \lambda_i Z(\boldsymbol{s}_i) : \lambda_0, \lambda_1, \ldots, \lambda_n \in \mathbb{R} \right\}$$

Equivalently, we can view the MSPE as a function of λ_0 and the vector $\boldsymbol{\lambda} \equiv (\lambda_1, \ldots, \lambda_n)'$ by writing

$$\mathrm{MSPE}(\lambda_0, \boldsymbol{\lambda}) = E\big[(Y(\boldsymbol{s}_0) \quad \lambda_0 \quad \boldsymbol{\lambda}'\boldsymbol{Z})^2\big] \qquad [2]$$

Then, we can find the values of λ_0 and $\boldsymbol{\lambda}$ that minimize this expression for the MSPE to obtain the simple kriging predictor $\hat{Y}^{(sk)}(\boldsymbol{s}_0)$. This minimization yields the simple kriging predictor

$$\hat{Y}^{(sk)}(\boldsymbol{s}_0) = \mu_Y(\boldsymbol{s}_0) + \tilde{\boldsymbol{c}}_Y(\boldsymbol{s}_0)'\boldsymbol{\Sigma}_Z^{-1}(\boldsymbol{Z} \quad \boldsymbol{\mu}_Y) \qquad [3]$$

where the (i, j) element of the $n \times n$ matrix $\boldsymbol{\Sigma}_Z$ is equal to

$$C_Z(\boldsymbol{s}_i, \boldsymbol{s}_j) \equiv \begin{cases} C_Y(\boldsymbol{s}_i, \boldsymbol{s}_i) + \sigma_\varepsilon^2, & \text{if } \boldsymbol{s}_i = \boldsymbol{s}_j \\ C_Y(\boldsymbol{s}_i, \boldsymbol{s}_j), & \text{if } \boldsymbol{s}_i \neq \boldsymbol{s}_j \end{cases} \qquad [4]$$

and

$$\mathrm{cov}(Y(\boldsymbol{s}_0), \boldsymbol{Z}) = \begin{cases} (C_Y(\boldsymbol{s}_i, \boldsymbol{s}_1), \ldots, C_Y(\boldsymbol{s}_i, \boldsymbol{s}_i) + \sigma_\varepsilon{}^2, \ldots, C_Y(\boldsymbol{s}_i, \boldsymbol{s}_n))', & \text{if } \boldsymbol{s}_0 = \boldsymbol{s}_i, i = 1, \ldots, n \\ (C_Y(\boldsymbol{s}_0, \boldsymbol{s}_1), \ldots, C_Y(\boldsymbol{s}_0, \boldsymbol{s}_n))', & \text{if } \boldsymbol{s}_0 \notin \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\} \end{cases}$$
$$\equiv \tilde{\boldsymbol{c}}_Y(\boldsymbol{s}_0)$$

In addition, the MSPE, or simple kriging variance, is

$$(\sigma^{(sk)}(\boldsymbol{s}_0))^2 = C_Y(\boldsymbol{s}_0, \boldsymbol{s}_0) \quad \tilde{\boldsymbol{c}}_Y(\boldsymbol{s}_0)'\boldsymbol{\Sigma}_Z^{-1}\tilde{\boldsymbol{c}}_Y(\boldsymbol{s}_0)$$

Recall that the simple kriging predictor of $Y(\boldsymbol{s}_0)$ is derived under the assumption that $\mu_Y(\cdot)$ is known. We now consider the more general case, known as universal kriging, where the mean is specified as a linear model. In particular, we assume that

$$\mu_Y(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s})'\boldsymbol{\beta}$$

where $\boldsymbol{x}(\boldsymbol{s}) \equiv (x_1(\boldsymbol{s}), \ldots, x_p(\boldsymbol{s}))'$ is a $p \times 1$ vector of known covariates at location $\boldsymbol{s} \in \mathscr{D}$ and $\boldsymbol{\beta} \equiv (\beta_1, \ldots, \beta_p)'$ is a $p \times 1$ vector of unknown regression coefficients. We note that ordinary kriging corresponds to a special case of universal kriging when $\mu_Y(\boldsymbol{s}) = \mu$, where μ is an unknown constant.

In deriving an optimal predictor of $Y(\boldsymbol{s}_0)$ in the universal kriging setting, we restrict our attention to the class of homogeneously linear predictors that minimize the MSPE. However, now we impose the additional assumption that $E[\hat{Y}(\boldsymbol{s}_0)] = E[Y(\boldsymbol{s}_0)]$. Since the generalized least squares estimator of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$, is unbiased, substituting $\hat{\mu}_Y(\boldsymbol{s}_0) = \boldsymbol{x}(\boldsymbol{s}_0)'\hat{\boldsymbol{\beta}}$ for $\mu_Y(\boldsymbol{s})$ satisfies this additional restriction. It turns out that this substitution is in fact optimal and, hence, it leads to the universal kriging predictor of $Y(\boldsymbol{s}_0)$:

$$\hat{Y}^{(uk)}(\boldsymbol{s}_0) = \boldsymbol{x}(\boldsymbol{s}_0)'\hat{\boldsymbol{\beta}} + \tilde{\boldsymbol{c}}_Y(\boldsymbol{s}_0)'\boldsymbol{\Sigma}_Z^{-1}(\boldsymbol{Z} \quad \boldsymbol{X}\hat{\boldsymbol{\beta}}) \qquad [5]$$

where $\boldsymbol{X}$ is an $n \times p$ matrix with rows $\{\boldsymbol{x}(\boldsymbol{s}_i)' : i = 1, \ldots, n\}$. Finally, it can be shown that the universal kriging variance is

$$\begin{aligned}(\sigma^{(uk)}(\boldsymbol{s}_0))^2 = {} & C_Y(\boldsymbol{s}_0, \boldsymbol{s}_0) \quad \tilde{\boldsymbol{c}}_Y(\boldsymbol{s}_0)'\boldsymbol{\Sigma}_Z^{-1}\tilde{\boldsymbol{c}}_Y(\boldsymbol{s}_0) \\ & + (\boldsymbol{x}(\boldsymbol{s}_0) \quad \boldsymbol{X}\boldsymbol{\Sigma}_Z^{-1}\tilde{\boldsymbol{c}}_Y(\boldsymbol{s}_0))'(\boldsymbol{X}'\boldsymbol{\Sigma}_Z^{-1}\boldsymbol{X})^{-1} \\ & \times (\boldsymbol{x}(\boldsymbol{s}_0) \quad \boldsymbol{X}\boldsymbol{\Sigma}_Z^{-1}\tilde{\boldsymbol{c}}_Y(\boldsymbol{s}_0))\end{aligned}$$

We note that $\sigma^{(uk)}(\boldsymbol{s}_0) \geq \sigma^{(sk)}(\boldsymbol{s}_0)$, as a result of the additional uncertainty due to the estimation of the mean of the $Y(\cdot)$ process.

Covariance Function/Variogram Estimation

The kriging methodology discussed in the previous section was developed under the assumption that the covariance function of the process $Y(\cdot)$, $C_Y(\cdot,\cdot)$, and the variance of the process $\varepsilon(\cdot)$, σ_ε^2, are known. Then from eqn [4], the covariance function of the process $Z(\cdot)$ is known. Typically, however, these quantities will have to be estimated (also using the observed data $\boldsymbol{Z}$) before kriging can be performed.

Consider the covariance function of the process $Z(\cdot), C_Z(\boldsymbol{s}_i, \boldsymbol{s}_j) = \operatorname{cov}(Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j))$. Various simplifying assumptions can be made about the functional form of $C_Z(\boldsymbol{s}_i, \boldsymbol{s}_j)$. For example, the covariance function of $Z(\cdot)$ is said to be stationary if the covariance between the process at any two locations is only a function of the vector connecting the two locations. Therefore, stationarity implies that $C_Z(\boldsymbol{s}_i, \boldsymbol{s}_j) = C_Z(\boldsymbol{s}_i - \boldsymbol{s}_j)$. Further, if the covariance function only depends on the distance between pairs of locations, so that $C_Z(\boldsymbol{s}_i, \boldsymbol{s}_j) = C_Z(||\boldsymbol{s}_i - \boldsymbol{s}_j||)$, the covariance function is said to be isotropic. A spatial process $Z(\cdot)$ is said to be second order stationary if its covariance function is stationary and its mean function is constant (i.e., $\mu_Y(\boldsymbol{s}) = \mu$, for all $\boldsymbol{s} \in \mathscr{D}$).

Rather than estimating the covariance function, it is common in geostatistics to work with the variogram function. Letting $\boldsymbol{h} = \boldsymbol{s}_i - \boldsymbol{s}_j \in \mathbb{R}^d$ be the displacement (or lag) vector from location $\boldsymbol{s}_j$ to $\boldsymbol{s}_i$, the variogram function evaluated at $\boldsymbol{h}$ is defined to be

$$2\gamma_Z(\boldsymbol{h}) = 2\gamma_Z(\boldsymbol{s}_i \quad \boldsymbol{s}_j) \equiv \operatorname{var}[Z(\boldsymbol{s}_i) \quad Z(\boldsymbol{s}_j)]$$

assuming that the right hand side is indeed a function of the vector $\boldsymbol{s}_i - \boldsymbol{s}_j$. The function $\gamma_Z(\boldsymbol{h})$ is called the semivariogram. If, further, the mean function is constant, $Z(\cdot)$ is said to be intrinsically stationary. Likewise, if the mean function is constant and the covariance function of $Z(\cdot)$ satisfies $C_Z(\boldsymbol{h}) = \operatorname{cov}(Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j))$, for $\boldsymbol{h} = \boldsymbol{s}_i - \boldsymbol{s}_j$, then $Z(\cdot)$ is said to be second order stationary, and the covariance function is said to be stationary. Consequently,

$$2\gamma_Z(\boldsymbol{h}) = 2C_Z(\boldsymbol{0}) \quad 2C_Z(\boldsymbol{h}) \qquad [6]$$

for a second order stationary process. There is a mathematical requirement that $C_Z(\boldsymbol{h})$ be non negative definite and that the variogram, $2\gamma_Z(\boldsymbol{h})$, be conditionally non positive definite. A discussion of how this requirement can be guaranteed through the use of a parametric class of covariance/variogram models is provided below.

For an intrinsically stationary process, the expected value of the difference between the process at any two locations is 0 and the variance of the difference is a function of the vector connecting the two locations. This class of processes can be shown to contain the class of second order stationary processes. Thus, intrinsic stationarity is a slightly weaker assumption than second order stationarity. Further, it can be shown that estimation of the variogram is more stable than estimation of the covariance function directly. Therefore, when estimating the covariance function of a spatial process, often the spatial trend is removed so that an assumption of a constant mean is reasonable. Then the process is taken to be intrinsically stationary and both the variogram and σ_ε^2 are estimated from the observed data.

Before discussing methods for estimating the variogram, we describe several of its general properties. First, we note that $\gamma_Z(\boldsymbol{h}) = \gamma_Z(-\boldsymbol{h})$ and $\gamma(\boldsymbol{0}) = 0$. We define the nugget effect to be the quantity, $c_0 = \sigma_\varepsilon^2 + c_{ms} \geq 0$, where $\gamma_Y(\boldsymbol{h}) \to c_{ms}$ as $\boldsymbol{h} \to 0$. The quantity c_{ms} captures microscale variation in the spatial process that cannot be attributed to measurement error. In practice, unless repeated observations at certain locations can be obtained, only the sum, $\sigma_\varepsilon^2 + c_{ms}$, can be estimated, but neither component can be estimated individually. Often, as $||\boldsymbol{h}|| \to \infty$, $\gamma_Z(\boldsymbol{h}) \to c_s$; the quantity c_s is known as the sill. If, in addition, the covariance function is stationary, $C_Z(\boldsymbol{0}) = c_s$. Finally, for an isotropic variogram, the value c_r such that $\gamma_Z(c_r) = c_s$, but $\gamma_Z(c_r(1-\delta)) < c_s$ for any $\delta > 0$, is called the range of the process.

Empirical Variograms

Empirical variograms are nonparametric estimators of the variogram of a spatial process. Assuming that $Z(\cdot)$ is intrinsically stationary, it can easily be shown that $2\gamma_Z(\boldsymbol{h}) = E[Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s})]^2$. Therefore, a reasonable empirical variogram, or moment based estimator of $2\gamma_Z(\boldsymbol{h})$, is

$$2\hat{\gamma}_Z(\boldsymbol{h}) \equiv \frac{1}{|N(\boldsymbol{h})|}\sum_{N(\boldsymbol{h})}[Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j)]^2$$

where $N(\boldsymbol{h}) \equiv \{(\boldsymbol{s}_i, \boldsymbol{s}_j) : \boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h};\ \ i,j = 1, \ldots, n\}$ and $|N(\boldsymbol{h})|$ denotes the number of unique pairs of locations in $N(\boldsymbol{h})$. When $Z(\cdot)$ is isotropic, the empirical variogram can be written as a function of the distance between pairs of observations. In this case, a powerful graphical tool for exploring the spatial dependence structure of $Z(\cdot)$ is a scatter plot of the empirical variogram, $2\hat{\gamma}_Z(||\boldsymbol{h}||)$, against $||\boldsymbol{h}||$.

When data are not located on a regular grid, the number of pairs of locations separated by a particular distance will not be sufficient to provide a precise estimate of the variogram. In this situation, distances between pairs of observations can be placed into bins centered at distances $\{l_i : i = 1, \ldots, m\}$. Then the empirical variogram can be calculated at each l_i by redefining $N(l_i)$ to be the set of all pairs of locations whose distance apart falls into the ith bin.

Finally, we note that the empirical variogram is not robust to outliers. As a result, a robust empirical variogram

$$2\hat{\gamma}_Z^{(R)}(\boldsymbol{h}) \equiv \frac{\left[(1/|N(\boldsymbol{h})|)\sum_{N(\boldsymbol{h})}\left|Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j)\right|^{1/2}\right]^4}{0.457 + (0.494/|N(\boldsymbol{h})|)}$$

has been proposed. This robust estimator is designed to downweight observations that are unusually large or small compared to neighboring observations.

Parametric Variogram Models

One limitation of the nonparametric approach to variogram estimation is that the approach does not guarantee that the empirical variogram is conditionally nonpositive definite. Therefore, the empirical variogram tends to be used primarily as an exploratory tool in geostatistical analyses and is followed by a parametric approach to obtain a valid variogram/covariance function to use in kriging. The idea behind the parametric approach is to consider classes of valid variogram models,

$$\Gamma^{(P)} \equiv \left\{2\gamma : 2\gamma(\cdot) = 2\gamma^{(P)}(\cdot;\boldsymbol{\theta}); \boldsymbol{\theta}\in\Theta\right\}$$

Then, within this class of models, the goal is to find a value of $\boldsymbol{\theta}\in\Theta$, $\hat{\boldsymbol{\theta}}$, for which $2\gamma^{(P)}(\cdot;\hat{\boldsymbol{\theta}})$ is a 'good' estimate of $2\gamma_Z(\cdot)$. Notice that $2\gamma^{(P)}(\cdot;\hat{\boldsymbol{\theta}})$ is a valid variogram, by definition.

Before discussing various methods of estimating $\hat{\boldsymbol{\theta}}$, we introduce a couple of commonly used classes of isotropic and stationary (parametric) variogram models in $\mathbb{R}^d$. For pairs of observations separated by $h \equiv ||\boldsymbol{h}||$, the class of exponential power variogram models has the form

$$\gamma^{(P)}(h; c_0, a, b) = \begin{cases} c_0, & \text{if } \ h = 0 \\ c_0 + a\left(1 - e^{-|h/b|^r}\right), & \text{if } \ h > 0 \end{cases}$$

where $a > 0$, $b > 0$, $c_0 \geq 0$, and $0 < r \leq 2$. This class of variograms is known as the exponential class when $r = 1$ and the Gaussian class when $r = 2$. The unknown parameters can be estimated from the observed data using one of the methods discussed below.

Another popular class of parametric variogram models is the Matérn class. The models in this class have a more complicated analytical form than the models in the exponential power class, but Matérn models have proved to be very flexible in describing the dependence structure of spatial data. The Matérn class is usually described in terms of its isotropic covariance function, although the corresponding parametric variogram model can be obtained using eqn [6] and

$$C^{(P)}(h; c_0, a, b_1, b_2) = \begin{cases} c_0, & \text{if } \ h = 0 \\ c_0 + \frac{a}{2^{b_2-1}\Gamma(b_2)}(||h||/b_1)^{b_2}\mathscr{K}_{b_2}(||h||/b_1), & \text{if } \ h > 0 \end{cases}$$

where $a > 0$, $b_1 > 0$, $b_2 > 0$, $c_0 \geq 0$, $\Gamma(\cdot)$ is the gamma function, and $\mathscr{K}_{b_2}(\cdot)$ is the modified Bessel function of the third kind of order b_2. Again, the unknown parameters can be estimated using the techniques discussed below.

Estimation of the unknown parameters in parametric variogram models is typically performed by one of two types of methods, a least squares based method or a likelihood based method. An advantage of the former is that it only requires mild distributional assumptions; however, it can be difficult to implement properly. On the other hand, a likelihood based method requires the specification of the joint distribution of the data, but estimation is straightforward using standard statistical software packages.

The simplest type of least squares based method is ordinary least squares (OLS) estimation. Consider an empirical variogram $2\hat{\gamma}_Z(l_i)$, where $i = 1, \ldots, m$, and a parametric variogram model $2\gamma^{(P)}(\cdot;\boldsymbol{\theta})$. The OLS estimate of $\boldsymbol{\theta}$ is obtained by minimizing the quantity

$$\sum_{i=1}^{m}\left(2\hat{\gamma}_Z(l_i) - 2\gamma^{(P)}(l_i;\boldsymbol{\theta})\right)^2$$

with respect to $\boldsymbol{\theta} \in \Theta$. A better estimator can be obtained using generalized least squares (GLS), since it takes into account the covariance between the values at the different $\{l_i\}$. In order to obtain the GLS estimate of $\boldsymbol{\theta}$, first define

$$2\hat{\gamma}_Z \equiv (2\hat{\gamma}_Z(l_1), \ldots, 2\hat{\gamma}_Z(l_m))'$$

$$2\gamma^{(P)}(\boldsymbol{\theta}) \equiv (2\gamma^{(P)}(l_1;\boldsymbol{\theta}), \ldots, 2\gamma^{(P)}(l_m;\boldsymbol{\theta}))'$$

and $V \equiv \mathrm{var}(2\hat{\gamma}_z)$. Then the GLS estimator of $\boldsymbol{\theta}$ is the value of $\boldsymbol{\theta} \in \Theta$ that minimizes

$$\left(2\hat{\gamma}_Z \quad 2\gamma^{(P)}(\boldsymbol{\theta})\right)' V^{-1} \left(2\hat{\gamma}_Z \quad 2\gamma^{(P)}(\boldsymbol{\theta})\right) \qquad [7]$$

In all but a few special cases, obtaining V is difficult. Therefore, a compromise between GLS and OLS, known as weighted least squares (WLS), is often used; this estimator is obtained by minimizing eqn [7] with $V = \mathrm{diag}(\mathrm{var}(2\hat{\gamma}))$, a diagonal matrix with possibly nonequal variances along the diagonal. An approximation to this matrix, $V \simeq \mathrm{diag}\{2(2\gamma^{(P)}(l_i;\boldsymbol{\theta}))^2/|N(l_i)| : i = 1, \ldots, m\}$, is often used, resulting in an estimator of $\boldsymbol{\theta} \in \Theta$ that minimizes

$$\sum_{i\ 1}^{m} \left(\frac{\hat{\gamma}(l_i)}{\gamma^{(P)}(l_i;\boldsymbol{\theta})} \quad 1 \right)^2 |N(l_i)|.$$

Maximum likelihood (ML) methods for estimation of $\boldsymbol{\theta}$ usually rely on a Gaussian assumption for the sampling distribution of the data. Suppose that data $\boldsymbol{Z}$ have a multivariate normal distribution with mean $\boldsymbol{X\beta}$, where $\boldsymbol{X}$ is an $n \times p$ matrix with rank $p < n$, and variance $\boldsymbol{\Sigma_Z}(\boldsymbol{\theta})$, where $\boldsymbol{\Sigma_Z}(\boldsymbol{\theta}) \equiv \{\mathrm{cov}(Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j)); \ \ i,j = 1, \ldots, n\}$ depends on $\boldsymbol{\theta}$. Then maximum likelihood estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ can be obtained by minimizing the negative log likelihood function,

$$\ell(\boldsymbol{\beta},\boldsymbol{\theta}) = \frac{n}{2}\log(2\pi) + \frac{1}{2}\log|\boldsymbol{\Sigma_Z}(\boldsymbol{\theta})| + \frac{1}{2}(\boldsymbol{Z} \quad \boldsymbol{X\beta})'\boldsymbol{\Sigma_Z}(\boldsymbol{\theta})^{-1}(z \quad \boldsymbol{X\beta}), \quad \boldsymbol{\beta} \in \mathbb{R}^p, \ \ \boldsymbol{\theta} \in \Theta$$

One drawback of the maximum likelihood approach is that the estimates of $\boldsymbol{\theta}$ will be biased, especially when p is large relative to n. As a result, a popular alternative likelihood based method known as restricted maximum likelihood (REML) can be used to obtain an estimator that is less biased than the ML estimator. A REML estimator of $\boldsymbol{\theta}$ can be obtained by finding a matrix of contrasts $\boldsymbol{A}$ such that $\boldsymbol{A}'\boldsymbol{X} = \boldsymbol{0}$. Then, if we let $\boldsymbol{Z}^* = \boldsymbol{A}'\boldsymbol{Z}$, $\boldsymbol{Z}^*$ will be distributed normally with mean $\boldsymbol{0}$ and covariance $\boldsymbol{A}'\boldsymbol{\Sigma_Z}(\boldsymbol{\theta})\mathbf{A}$. Minimizing the negative log likelihood function of $\boldsymbol{Z}^*$ with respect to $\boldsymbol{\theta} \in \Theta$ is equivalent to minimizing

$$\ell\left(\hat{\beta}(\theta), \theta\right)\left(\frac{1}{2}\right) log\left(|X\boldsymbol{\Sigma_Z}(\boldsymbol{\theta})^{-1}X|\right),$$

with grespect to $\boldsymbol{\theta} \in \Theta$, where $\hat{\beta}(\boldsymbol{\theta}) = (\mathbf{X}'\boldsymbol{\Sigma}_Z(\boldsymbol{\theta})\ \ ^1\mathbf{X})\ \ ^1 (\mathbf{X}'\boldsymbol{\Sigma}_Z(\boldsymbol{\theta})\ \ ^1\mathbf{Z})$. The value of $\boldsymbol{\theta}$ that minimizes this expression is the REML estimate.

Rainfall Example

To illustrate the kriging/variogram methodology, we present an analysis of a spatially referenced dataset performed using the geoR package available in the R statistical software environment. The data consist of total rainfall during the dry season (May–June) averaged over several years at a set of 143 locations in the state of Parana, Brazil. **Figure 1** shows the locations and values of the rainfall data.

Following other statistical analyses of these data, we first removed north–south and east–west trends in the data and then performed a variogram analysis of the residuals. **Figure 2** shows a binned robust empirical semivariogram, as well as a fitted (using GLS) exponential variogram (i.e., the exponential power model with $r = 1$) with estimated parameters: $\hat{\sigma}^2_\varepsilon = 151.5$, $\hat{a} = 1189.1$, and $\hat{b} = 117.2$. Using this fitted variogram model, we predicted the average rainfall during the dry season on a fine grid covering Parana using universal kriging. Kriging estimates (left plot) and kriging standard errors (right plot) are shown in **Figure 3**, along with the

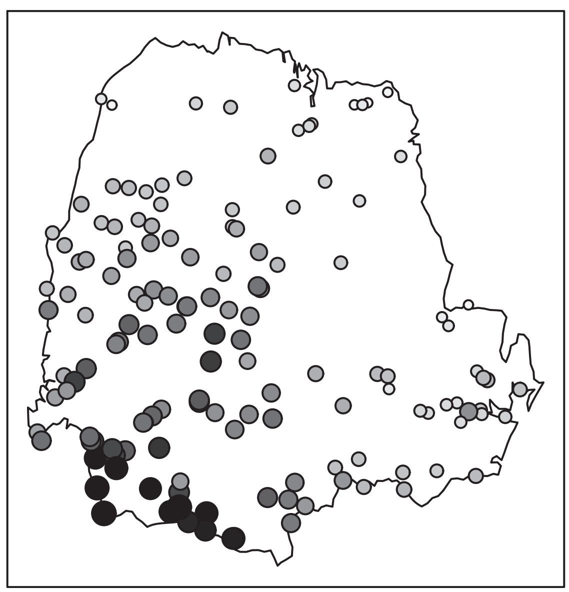

Figure 1 Circles indicate the locations of the average yearly rainfall during the dry season (May June) observations in the state of Parana, Brazil. The size and shade indicate the variation in total rainfall across the region (the larger and darker the circle, the higher the yearly rainfall).

locations of the observations (circles). Notice that in areas with relatively large numbers of observations, the standard errors are smaller than in regions with fewer observations.

Other Topics

Multivariate Methods (Cokriging)

A multivariate spatial (random) process $\underset{\sim}{Z}(\cdot)$ of dimen sion q is defined to be a vector whose entries are the univariate spatial processes $Z_1(\cdot), \ldots, Z_q(\cdot)$, each defined over a region $\mathscr{D}$. For example, a collection of air pollu tant processes, such as the US Environmental Protection Agency's (EPA) six criteria air pollutants (ozone, nitrogen dioxide, particulate matter, sulfur dioxide, carbon mon oxide, and lead), can be viewed as a multivariate spatial process. This multivariate process is monitored by the EPA as a result of amendments made to the Clean Air Act in 1990, which required that monitoring networks be established to record the ambient concentrations of these pollutants over the country. The data collected by these monitors are an example of multivariate geostatistical data, and geostatistical methods can be used to predict the ambient concentration of one or several of these pollutants at unobserved locations. In particular, the methodology known as cokriging provides predictions of one pollutant based on the concentrations of all six pollutants in the surrounding area. In situations where there is substantial dependence among the different pollutants, as is the case with several of the criteria pollutants, predictions made using cokriging are more accurate than if they were made using univariate kriging.

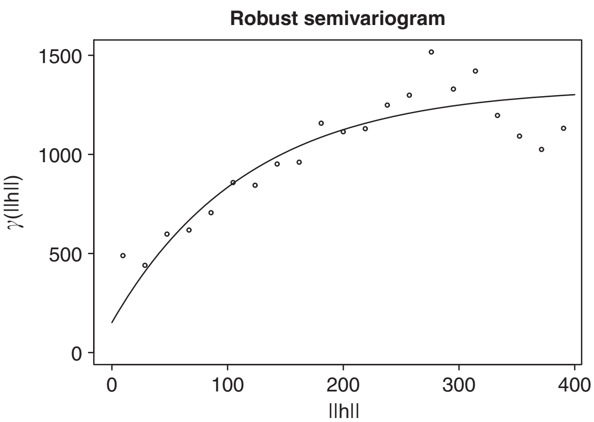

Figure 2 Small circles indicate the value of the binned empirical semivariogram for the residuals (after removing both north south and east west trends) evaluated at various distances. The smooth curve represents the fitted exponential semivariogram estimated from the observed data.

Space–Time Kriging

One straightforward extension of geostatistics in a d di mensional space is to $(d+1)$ dimensional space–time geostatistical data. Geostatistical analyses of such data are based on space–time processes $Z(\boldsymbol{s}^*)$, where $\boldsymbol{s}^* = (\boldsymbol{s}, t)$ such that $\boldsymbol{s} \in \mathbb{R}^d$ represents the spatial dimension of the process and $t \in \mathbb{R}$ represents the temporal dimension.

In practice, the covariance structure of a space–time process is often assumed to be separable. This assumption implies that the covariance function of the process $C_Z((\boldsymbol{s}_i, t_i), (\boldsymbol{s}_j, t_j)) = \operatorname{cov}(Z(\boldsymbol{s}_i, t_i), Z(\boldsymbol{s}_j, t_j))$ can be fac tored into the product of a spatial covariance function, $C_Z^{(S)}(\boldsymbol{s}_i, \boldsymbol{s}_j)$, and a temporal covariance function, $C_Z^{(T)}(t_i, t_j)$. This restriction implies that the covariance between the process at any two locations does not depend on time and that the covariance between the process at two time points does not depend on location. Due to the restrictive nature of the separability assumption, classes of nonseparable

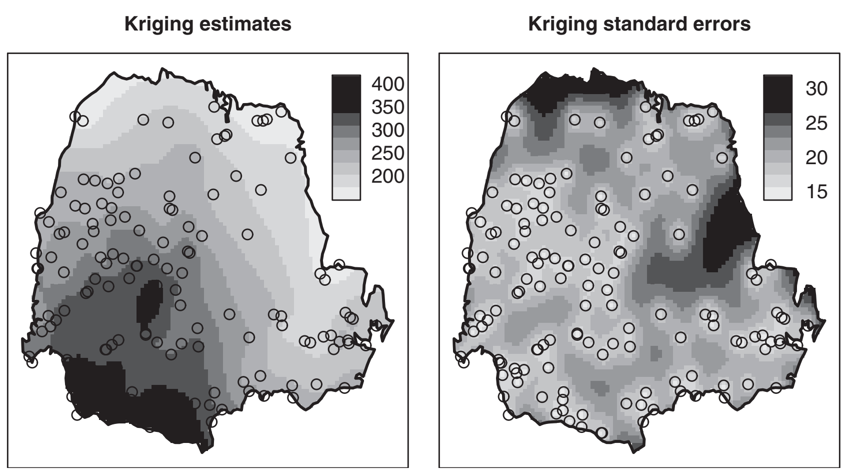

Figure 3 Kriging predictions (left plot) and standard errors (right plot) on a fine grid covering in the state of Parana, Brazil. The black circles indicate the locations of the observed rainfall values.

space–time covariance functions have been proposed re cently in the literature.

Finally, we note that there has also been a movement in the spatial statistics literature to model space–time geostatistical data as a dynamic spatial process, or as a spatial process whose structure changes over time. This flexible alternative to standard space–time kriging is appealing from both a scientific as well as a computa tional perspective in a variety of areas of application.

Nonstationary Models

The assumption that the covariance function of a spatial process is stationary (i.e., $C_Z(\boldsymbol{s}_i, \boldsymbol{s}_j) = C_Z(\boldsymbol{s}_i - \boldsymbol{s}_j)$) is often not an appropriate assumption for the dependence struc ture of spatial data. However, classes of nonstationary covariance functions are not readily available. As an al ternative to directly specifying covariance functions, a class of models known as process convolutions or spatial moving average models have developed. These models differ from traditional geostatistical models in that the covariance structure is defined in a constructive, as op posed to a functional, manner. For example, a Gaussian spatial process can be constructed by convolving a two dimensional kernel with a two dimensional white noise process. Nonstationarity can be introduced by allowing the parameters of the kernel function to vary over space. Another class of methods for modeling nonstationary geostatistical data are spatial deformations. These methods assume that the spatial process is stationary and isotropic only after some nonlinear deformation of the sampling space. This approach is conceptually elegant, but can be difficult to implement in practice.

Bayesian Methods

Bayesian methods have become increasingly popular in analyses of geostatistical data in recent years. The Bayesian paradigm provides a coherent approach for specifying sophisticated hierarchical models for complex data, and recent computational advances have made model fitting in these situations feasible. The drawback to using Bayesian methods is that all joint distributions of processes and parameters have to be specified, often via a collection of conditional distributions. Distributional as sumptions are not necessary in classical geostatistics, which require only the specification of means, variances, and covariances. Despite the need for more restrictive assumptions, a powerful advantage of the Bayesian ap proach is that it allows a deeper scientific understanding of the nature of the process $Y(\cdot)$.

See also: Spatial Interpolation; Statistics, Spatial.

Further Reading

Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2004). *Hierarchical Modeling & Analysis for Spatial Data*. New York: Chapman & Hall/ CRC.

Cressie, N. (1990). The origins of kriging. *Mathematical Geology* 22, 239 252.

Cressie, N. (1993). *Statistics for Spatial Data* (revised edition). New York: Wiley.

Diggle, P. J. and Ribeiro, P. J., Jr. (2007). *Model Based Geostatistics*. New York: Springer.

Journel, A. G. and Huijbregts, C. J. (1978). *Mining Geostatistics*. London: Academic Press.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology* 58, 1246 1266.

Schabenberger, O. and Gotway, C. A. (2005). *Statistical Methods for Spatial Data Analysis*. New York: Chapman & Hall/CRC.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer Verlag.

Stein, M. L. (2005). Space time covariance functions. *Journal of the American Statistical Association* 100, 310 321.

Ver Hoef, J. M. and Cressie, N. (1993). Multivariable spatial prediction. *Mathematical Geology* 252, 219 239.

Waller, L. A. and Gotway, C. A. (2004). *Applied Spatial Statistics for Public Health Data*. New York: Wiley.

Wikle, C. K. and Cressie, N. (1999). A dimension reduced approach to space time Kalman filtering. *Biometrika* 86, 815 829.

Relevant Websites

http://www.esri.com
ArcGIS Geostatistical Analyst.

http://www.mrc bsu.cam.ac.uk
GeoBUGS.

http://www.est.ufpr.br
geoR Package.

Kropótkin, P.

G. Kearns, Virginia Tech, Blacksburg, VA, USA

Peter Kropótkin (1842–1921) (**Figure 1**) was born into a wealthy landowning family in Russia; his father had several thousand serfs on his estate. As a youth, Kropótkin was sent to the court of the Tsar to be groomed for high office in the Russian state. He joined an elite group within the Russian army in much the way his father had hoped. However, Kropótkin's grooming to be a Russian bureaucrat was flawed. His mother had died when he was very young and, being brought up by peasant servants, he retained a deep sympathy for the plight of Russia's poor. In 1861, as Kropótkin embarked upon his military career, the Tsar, Alexander II, abolished the institution of serfdom. Although the peasants continued to be exploited economically, they were no longer subject to the indignities of servility. At this same time, the Russian state was extending itself into Siberia and in this unglamorous, frontier region Kropótkin imagined that he might work as part of an enlightened administration. He went to Siberia in 1862 and hoped to be part of a general process of reform. He worked first on reports about the treatment of prisoners sent in exile to Siberia. These reports did not produce a change in policy from St. Petersburg. In St. Petersburg newspapers, Kropótkin reported on the repression of Polish revolutionaries sent to Siberia, and in 1866 he resigned from the military after the government in Siberia summarily executed five such prisoners. Kropótkin spent increasing amounts of his time on exploration in the mountain regions between Russia and China. Now he turned to a career in the civil service but devoted much of his time to exploration.

Figure 1 Peter Kropótkin (1842 1921).

He worked on Arctic exploration, became a leading figure in the Russian Geographical Society, and through the 1870s published many reports on Siberia and Finland. In 1872 he resigned from the civil service to continue both with his work in geography and also with his new career as a political organizer among the working class. Kropótkin visited Western Europe in 1872 and deeply admired the socialist and anarchist organizations among the workers of Switzerland. He was struck by their principled commitment to socialist anarchism as both a means and an end. Back in St. Petersburg, under the name of Borodin, he worked with discussion groups among workers as he both taught and developed his ideas about socialism, Marxism, and anarchism. In 1874 the authorities connected Borodin to Kropótkin and he was arrested. In 1876 he made a daring escape from a prison hospital and fled, via Finland, to Britain. There, he took on the name of Alexis Levashov and made a scanty living writing reviews and translating articles for the *Times*, for *Nature*, and for the *Proceedings of the Royal Geographical Society*. In the absence of a vibrant anarcho socialist movement, Britain was a lonely place for Kropótkin and, within a year or so, he began to spend most of his time in Switzerland and France, among anarchists, working mainly as a journalist and author. When political repression on the mainland became, periodically, too dangerous for him, he would return to England, but in 1882 he was arrested in France on a false charge of complicity in a series of strikes. Despite a petition signed by dozens of British academics and politicians, he languished in French jails until 1886. It was there that he wrote his passionate account of 'What geography ought to be'.

He came back to Britain and lived there until 1917 although he made brief visits to continental Europe and even to North America. He played a full part in the revival of socialist and anarchist organizations from the late 1880s and he was a prominent intellectual both in geography and the biological sciences. Some academics at

Cambridge even thought he might be suited to a chair in geography there but Kropótkin was unwilling to give up anarchist propaganda for the 'ivory tower' and it was not clear at least to him that he could follow both. When autocracy was relaxed in Russia from the mid 1890s, Kropótkin began to correspond with Russian revolutionaries schooling them away from terrorism toward organizing mass movements. After the revolution of 1905, he was a leader among those in Britain protesting the violence of its repression by the Tsarist authorities. He saw World War I in nationalist terms and took sides against Germany, favoring instead any alliance that included France, the source, to his mind, of progressive and revolutionary principles. Very many contemporary anarchists rejected the nationalism of Kropótkin's reading of the war. Kropótkin went to Russia in 1917 to work for his anarchist ideals in the new socialist state, but the Bolsheviks suppressed the anarchists and, despite a brief audience with Lenin, Kropótkin lived on in Moscow, tolerated but unheeded, until his death in 1921.

Kropótkin's earliest contributions were in physical geography. The Himalayas were mapped from its edges and Kropótkin was among the first to describe correctly its interior structure of folds and peaks. He was also early among the advocates of the claim that, at one time, all of Northern Europe was covered by a massive ice sheet. Finally, he proposed a general theory of the desiccation of Europe, with ice sheets giving way to lakes, and then lakes giving way to marshes. The retreat of lakes within the historic past was part of this drying out, rather than being an index of the progressive warming of Europe over the same period. Yet, he was always a student of nature rather than of physiography. It was in the *Cosmos* of Humboldt and in the poetry of Goethe that he found an echo of the feelings with which he contemplated the landscapes of his youth. In his autobiography, Kropótkin described the epiphany of his contemplation of the place of human labor in transforming the rural landscape of Finland. The labor that shaped the fields, planted the crops, and produced the food was unfree. The beautiful veil of nature obscured the crime of labor exploited and Kropótkin ever after abjured the guilty pleasures of a purely physical geography.

This insight breathes passion into his manifesto, 'What geography ought to be'. In 1886 John Scott Keltie was working to get geography established in British universities. As part of this campaign, Keltie had written a report on the comparative position of geography in the universities of Europe and had followed this up with a series of lectures in London to promote geography as a vital research discipline. Halford Mackinder's 'Scope and methods of geography' was part of this campaign and made a case for the subject as a post exploration science for empire. Keltie sent a copy of his report to Kropótkin, then in a French prison, and from his cell Kropótkin replied with an article supporting the campaign, but with a very different politics to that articulated by Mackinder. For Kropótkin, the study of geography taught knowledge of other peoples, a knowledge that bred mutual respect rather than served imperial control. Indeed, the more we learned about the practices of other peoples, the less we could judge them as inferior and thus the less likely we were to patronize (or exploit) them with our enlightened domination.

Many of Kropótkin's contemporaries might fairly be called social Darwinist. They believed that Charles Darwin had taught that fitter forms survived the struggle for food and resources. They applied this insight not to the struggle between species and their environment or even between species within an environment, but, rather, to the conflict between classes, or between nations, or between races. They were able to persuade themselves that Darwin's theory of evolution implied that the rich were simply fitter than the poor; and that the powerful nations and races had to fight the others to survive. This was indeed a bleak view of society and economy. It seemed to justify *laissez faire* capitalism, aggressive colonialism, and even eugenic racial segregation. Kropótkin had a more optimistic view of nature. He thought nature's bounty was more than enough for all. Scientific agriculture and industrial technology could serve ample portions to all if only fairness governed the distribution of the social product.

Kropótkin believed both that Darwin had been misinterpreted and that his ideas were incomplete. Kropótkin argued that individuals, classes, nations, and races had no biological status in Darwin's theory. Evolution and the survival of the fittest related to species and to nothing else at all. In the second place, Darwin's own ideas were incomplete because he had, by and large, treated species as aggregations of individuals and had ignored the social aspect of their groupings. Competition prevailed between individuals only in the lower forms of life. The social factor was vital both in the development of higher order forms of life and in their ways of living. It was through cooperation that members of a species were fed and sheltered; no single bee could build a hive, as no lone wolf could chase down a bear. Social forces were at the heart of the development of language in animals and were most highly developed among the superior species. Kropótkin proposed that *Mutual Aid* was a central evolutionary force and had direct implications for social as well as biological evolution. In this respect, the antisocial stance of imperialists and free market economists was regressive.

Kropótkin believed that cooperation implied reciprocity and freedom. He was fundamentally opposed to coercive social and economic relations. He had faith that human energies would develop most fully when allowed to flourish most freely. This was the basis of his anarchism. For Kropótkin, freedom was both a means and end to social and economic progress. This meant that he was opposed to centralization and indeed to the institutions of the state. He believed that the medieval commune was probably the largest human community that could be governed by direct democracy and thus it was the apex of social evolution. The development of centralized states thereafter reduced freedom, promoted private property, and resulted in the competitive geopolitical world order that was modern imperialism. His utopian geography was fundamentally cellular. Distinct communities using, in the main, local resources would trade only what is sur plus to local needs. They would not produce things primarily for exchange but for subsistence.

Kropótkin's ideas did not prevail in British or European geography but they were part of an anarchist tradition that remained a significant, if not dominant, presence in geog raphy and related disciplines such as sociology. The sci entific prestige of Kropótkin as a biologist and of Reclus as an encyclopedic synthesizer of geographical knowledge meant that in Britain, in Spain, in Italy, in Russia, in Germany, and in France, there were significant anarchist resources for geographers and significant geographical re sources for anarchists. Regionalism in geography long served as a refuge for ideas about federalism and local autonomy that were anarchist in tendency if not in purpose. Cooperation was important to the sociological thought of urbanists such as Lewis Mumford and Patrick Geddes and sustained developments such as the community colleges of both the United States and the United Kingdom. Within education, Kropótkin's ideals were celebrated by thinkers such as Paul Goodman and Paulo Freire.

Yet, in the main, it must be admitted that anarchist studies in geography remain a hope rather than a reality. With the radical turn in geography after the political and intellectual ferment that followed 1968, scholars began again to look at Kropótkin. He appeared to offer a radical vision more ecological and less state centric than the Marxism that drove the countercultural agenda in Anglo American geography, but the anarchist critique of Marxism was never developed with either the theoretical or empirical force of its radical rival and thus these di mensions are still underdeveloped in modern radical geography. There are few studies of the biogeography of *Mutual Aid*, the historical geography of medieval com munalism, and the economic geography of voluntary association. The legacy of Kropótkin has potential.

See also: Anarchism/Anarchist Geography; Cultural Geography; Darwinism (and Social Darwinism); Eurocentrism; Exploration; Gentrification; Humanism/ Humanistic Geography; Imperialistic Geographies; Lamarck(ian)ism; Mackinder, H. J.; Radical Environmentalism; Radical Geography; Radical Political Economy; Reclus, E.; Segregation, Urban; Socialism; State; Territory and Territoriality; von Humboldt, A.

Further Reading

Campbell, J. A. and Livingstone, D. N. (1983). Neo Lamarckism and the development of geography in the United States and Great Britain. *Transactions of the Institute of British Geographers* 8, 267 294.

Galois, B. (1976). Ideology and the idea of nature: The case of Peter Kropotkin. *Antipode* 8, 1 16.

Hooson, D. J. M. (1968). The development of geography in pre Soviet Russia. *Annals of the Association of American Geographers* 58, 250 272.

Kearns, G. (2004). The political pivot of geography. *Geographical Journal* 170, 337 346.

Kropotkin, P. (1885). *Words of a Rebel*. Montréal, QC: Black Rose Books. [Reprinted 1992, Montréal, QC: Black Rose Books.]

Kropotkin, P. (1899). *Memoirs of a Revolutionist*. London: Creset Library. [Reprinted 1962, London: Creset Library.]

Kropotkin, P. (1904). *Modern Science and Anarchism*. London: Freedom Press. [Reprinted 1912, London: Freedom Press.]

Kropotkin, P. (1922). *Ethics Origin and Development*. New York: Tudor Publishing. [Reprinted 1947, New York: Tudor Publishing.]

Kropotkin, P. (1955). *Mutual Aid: A Factor of Evolution*. Boston, MA: Horizons Books.

Kropotkin, P. (1978). What geography ought to be. *Antipode* 10, 6 15.

MacLaughlin, J. (1986). State centred social science and the anarchist critique: Ideology in political geography. *Antipode* 18, 11 38.

Miller, M. (ed.) (1970). *P. A. Kropotkin: Selected Writings on Anarchism and Revolution*. Cambridge, MA: MIT Press.

Walter, N. and Becker, H. (eds.) (1988). *Act for Yourselves: Articles from 'Freedom' 1886 1907 by Peter Kropotkin*. London: Freedom Press.

Wise, M. J. (1986). The Scott Keltie Report 1885 and the teaching of geography in Great Britain. *Geographical Journal* 152, 367 382.

Woodcock, G. and Avadumovic, I. (1950). *The Anarchist Prince: A Biographical Study of Peter Kropotkin*. London: Boardman.

Relevant Websites

http://dwardmac.pitzer.edu
Dana Ward's homepage: Kropotkin Archive.

http://www.marxists.org
Marxists Internet Archive: Kropotkin Reference.

http://recollectionbooks.com
Recollection books used: Peter Kropotkin page from the Daily Bleed's Anarchist Encyclopedia.

L

Labor Control Regime

A. E. G. Jonas, University of Hull, Hull, UK

Glossary

Capital Mobility The propensity for branches of production to relocate; can also refer to the increasing rate of flow of investment from one place to another. Often associated with economic globalization.

Cultural Hegemony A concept used by the founder of the Italian Communist Party, Antonio Gramsci, to describe a variety of coercive and noncoercive economic strategies and cultural practices deployed by the dominant ruling class to control and regulate the behavior of the working class.

Global Labor Control Regime A tendency is for labor control to stabilize around globally dominant spatial forms and practices. Can refer to the process of McDonaldization, as described by the sociologist Ritzer.

Hegemonic Despotism Refers to a central feature of the contemporary labor control regime which is capital's capacity to enhance spatial mobility as a mechanism to extract concessions from workers.

Historic Bloc A period of sustained cultural hegemony.

Labor Flexibility The intensification of the detailed division of labor around more flexible production processes than those typically found in mass production. More generally, can refer to the contemporary labor control regime.

Labor Geography The study of workers and their organizations from a spatial perspective.

Labor Market The place of recruitment and exchange of labor; can also refer to the institutions regulating the conditions under which workers acquire or develop requisite job skills, such as the labor market inside a given industrial sector.

Labor Process The social and technological organization of labor including the detailed division of work tasks at the point of production.

Labor Regulation The wider institutions of state and economy, which nationally or globally govern wage labor and the workplace.

Local Labor Control Regime The particular labor control practices dominating workers in a place or region.

Spatial Divisions of Labor The variety of ways in which the division of labor in production is organized and unfolds across space.

The labor control regime refers to the social need in capitalism for employers to invest in means of integrating labor into the production system. This need arises from conditions peculiar to capitalism in which workers are ideologically free to sell their labor power to any em ployer for a wage, and yet they neither own for them selves nor exercise direct control over the means of production. If workers did own the means of production, presumably there would not be a need for labor control externally imposed on them by another class. Instead, workers would undertake necessary tasks and regulate the labor process to meet a self conscious material and social need rather than engage in patterns of work dic tated by capitalist conditions of competition and the restless search for profits. However, as things stand there is a considerable investment by capitalists (and state managers) in the socialization and integration of labor. Such investment is not so much a factor cost in pro duction as it occurs on a habitual basis through the de tailed organization of workplace practices governing the quality and quantity of labor available in the labor market and deployed in production.

Over time, managers working in private or state run capitalist enterprises have developed a variety of means (directly and coercively, as well as indirectly and re ciprocally) of integrating the workforce into production systems and labor processes. From the standpoint both of the capitalist class and of the state, the integration of workers into production ideally occurs in a consensual fashion but, for a variety of reasons, this does not ne cessarily happen. Indeed, the process of labor integration is rife with conflict. Workers are potentially resistant to, and in fact often do resist, the time–space disciplines of industrial production. Place based labor control practices have emerged specifically to deal with worker unrest or the possibility thereof. These practices connect labor to the workplace in fundamentally a sociospatial manner. They even extend into the wider labor market and the sphere of labor reproduction. This spatial extension oc curs at times informally through, for example, business involvement in community activities, or formally, for instance by means of state regulation and provision of education and training, eventually producing a local labor control regime.

The Labor Control Regime in Historical Context

Social science research on labor control builds upon the insights afforded by Karl Marx's historical materialism and the writings of the early twentieth century radical thinker and political activist Antonio Gramsci. If Marx argued that wage labor in capitalism was not so much a natural condition as a social necessity, he also recognized that it was historically contingent whether workers were compliant and willing to exchange their labor power. This raised the question of how capitalists successfully integrate workers into production, given workers' ten dency as a class to resist on a collective basis.

Writing in the first part of the twentieth century, Gramsci began to reflect upon the role of economic and cultural barriers to organized working class resistance. During World War I, he was involved in organizing in dustrial workers in Turin, a major center of the auto industry in Italy. In the early 1920s, he became frustrated with the failures of the Italian socialist movement and founded the Italian Communist Party. In 1926, the Fascist national government imprisoned Gramsci for holding radical political views and engaging in activism against the ruling class. He experienced a deterioration of his already fragile health, eventually dying in 1937.

During his time in prison, Gramsci compiled notes on why the working class struggle had failed to develop into a successful movement for socialism, eventually pub lished posthumously as *Selections from the prison notebooks*. In the *Notebooks*, he argued that the state and the ruling capitalist class deployed various coercive and non coercive economic strategies and cultural practices to control and regulate the behavior of the working class. This was Gramsci's theory of 'cultural hegemony' and central to it was the rise of a 'historic bloc'. This referred to a period of sustained dominance of the working class by the particular economic and cultural practices of the ruling class.

In Gramscian terminology, a labor control regime is a particular form of cultural (and economic) hegemony. It is the concrete expression of two fundamental historical–geographical tensions: (1) the latent tendency for the ruling class to develop cultural practices to prevent organized class struggle on a widespread basis, and (2) the particular economic conditions and cultural cir cumstances of worker integration or resistance in the workplace under a given production system.

Gramsci was writing about cultural hegemony during the period of Fordist mass production in northern Italy under Fascism. However, it has been possible to identify historic blocs when other forms of labor control domi nated industrial and cultural landscapes. These forms would have been adapted to the specific organizational, technological, and social demands of workers and capitalists in a given period of industrial development. Modern labor historians and sociologists, such as Burawoy, describe relatively stable periods of industrial production in terms, such as factory regimes. The regimes are periods when particular conditions and methods of labor control and integration are dominant.

For example, the Fordist industrial period incorpo rated various coercive or direct (e.g., surveillance through Taylorism) and noncoercive or reciprocal (e.g., corporate welfarism) labor control practices. It gave rise to a labor control regime, at the national scale built around collective bargaining and a culture of consensus, and locally around the particular management practices and workers' con sumption habits found in numerous Fordist workplaces and communities. To a greater or lesser extent, Fordism dominated social conditions of production and con sumption in communities throughout much of the twen tieth century, dampening organized resistance in the workplace and labor's class consciousness. With its socially extended means of labor control, Fordism supplanted earlier paternalistic factory regimes and direct owner capitalist surveillance and coercion of workers within the workplace. Paternalism was typical of the formative period of machinofacture in the latter part of the nineteenth century. It grew up mainly around company towns and other self contained communities dominated by a single industry or a private (often family owned) company.

Geographers and Labor Control

Although labor scholars like Gordon in the early 1980s suggested there were important relationships between labor control and the spatial organization of production, it was not until the mid 1980s and 1990s that human geographers seriously tackled this question. The wider context was the crisis of mass production in Europe and North America and debates about industrial transi tion. Geographers contributed to these debates in a characteristically empirical fashion, grounding abstract theories of industrial restructuring and transition in the conditions of production and work in particular regions and territorial economies. Out of this emerged some key geographical statements on production, labor control, and territory, including Massey on the spatial division of labor, and Storper and Walker on spatial shifts in growth centers of capitalism.

Massey drew some inspiration from the writings of labor scholars such as Braverman, who she nevertheless criticized for offering a teleological and functionalist analysis of labor control in mass production and the process of de skilling. Massey signaled what became a politically urgent research agenda for economic geog raphers, namely to bring knowledge of place and con tingency into a historical materialist approach to spatial structures of production and labor control. She identified a variety of ways in which the geography of production shapes the politics and culture of different places and regions, and how over time the differently constructed politics of places and regions influences the geography of production and labor integration.

What was central to Massey's 'spatial divisions of labor' approach was the understanding that employers are aware of the different ways in which places contribute to the development and stabilization of production and employment relations. Notably, places and regions comprise particular combinations of production strat egies and labor relations so that it is impossible to impute to local political conditions and outcomes an unmediated role for wider forces of production and cultural hege mony. If the study of cultural hegemony was not a central feature of Massey's approach, the difference place makes to the ways such dominant economic and cultural pro cesses unfold in particular localities and regions was nevertheless of great theoretical importance.

Labor scholars, including the economic geographers Storper and Walker, started to assign space a more central role in the labor control regime. For example, Burawoy argued that a new regime of labor control built around capital's spatial mobility replaced consensus under Fordism. He called this regime 'hegemonic despotism'. Under hegemonic despotism, labor susceptibility to the growing international mobility of capital is the source of capital's power to extract concessions from workers and assert cultural hegemony over the workplace. Recogniz ing this tension between the workplace and capital mo bility is an important development in approaches to the labor control regime.

The Control and Regulation of Labor in Place

In 1996, Jonas offered a spatial theory of labor control in a paper published in a special issue of *Regional studies* devoted to the governance of the local labor market. In the same year, the publication of Peck's *Work place* rep resented the clearest and most comprehensive statement on the changing spatial conditions of labor regulation and control in the industrial core during the transition away from Fordist production and Keynesian state manage ment. The respective contributions of Jonas and Peck offer representative statements on the geography of the labor control regime.

Despite a growing number of locality case studies of workplace politics and culture, there was no explicitly spatial theory of labor control. Jonas sought to address this lacuna by thinking through the spatial logic under pinning local labor control strategies, and mainly from the standpoint of the material interests of different lo cally dependent businesses. For Cox and Mair, businesses become locally dependent when they rely on particular place based social relations and networks for access to markets, resources, labor, and state support. These re lations and networks are not so readily available and reproducible elsewhere. Jonas drew on this insight to argue that when producing, investing, expanding, and competing firms deploy particular labor control strategies in place. In time, these strategies might be replicated across the landscape but only if firms expand spatially in response to problems of local dependence. Spatial ex pansion and replication occurs through the extralocal investment strategies of firms or via mergers and acqui sitions, involving enterprises based elsewhere.

Given that production occurs at specific locations, Jonas argued that there is a tendency for labor control to stabilize around place specific social practices, which affect the social integration of labor inside the workplace but influence conditions outside it as well. When stability occurs, conditions are in place for a 'local labor control regime'. Jonas further argues that some firms are pro foundly dependent upon places and their constituent labor control practices, and to such an extent as to pro duce, in effect, locally specific forms of cultural hege mony. This refers to the ways in which the properties of a local labor control regime infuse other workplaces and many aspects of local community and civic life throughout the locality or region. The resultant local political effects of a local labor control regime might (or might not) activate when the place or region experiences an external threat or undergoes a new round of inward investment, as Jonas demonstrated in the case of a manufacturing company based in the City of Worcester in central Massachusetts. The company was able to ward off an external competitor by drawing upon local social and political networks constructed around its local labor control practices.

Peck's particular contribution to the study of the geography of labor control involves in the first instance an analysis of macro scale modes and forms of labor regulation after Fordism–Keynesianism. Here Peck is not so much interested in the specifics of a local labor control regime as attempting to identify wider spatial structures and processes of labor regulation. In this extent, his work is consistent with the Gramscian intellectual tradition, which has undergone a revival through the work of state regulation theorists. Nonetheless, the primary emphasis in Peck's work is on economic and political (i.e., state orchestrated) forms of labor regulation rather than cultural forms *per se*. Such economic and political forms of labor regulation and control include training and enterprise schemes, temporary and part time forms of work, labor market intermediaries, and in particular (the threat of) capital mobility. These new spatial regulatory forms have emerged in the context of state withdrawal from the direct provision of support to workers and the unemployed in the form of unemployment benefit, workers' compensation, and welfare. The disciplining effect upon workers and those outside paid work only intensifies with globalization.

In considering Peck's contribution to the geography of labor, it is useful to draw a broad distinction between labor control and the 'regulation of labor'. Although both concepts overlap insofar as they each refer to a process of socially integrating labor, they can be treated as heuristically different in that they entail different entry points into the labor market and its attendant institutions, agencies, spatial practices, and territorial scales. Local labor control refers to the particular processes by which workers are integrated into the workplace and become habituated to production and work in a given place or region. By way of comparison, labor regulation refers not so much to the relationship between the labor process and the labor market *in situ* as to the wider (national and international) social conditions (i.e., those external to a workplace or regional industrial cluster) structuring the supply and quality of labor in the labor market.

Whereas the approach to local labor control outlined above extends from the factory or workplace into the wider labor market and the sphere of social reproduction (**Figure 1**), the analysis of labor regulation covers the wider institutions of state and economy, which nationally or globally govern wage labor and the workplace. Seen in this light, one of Peck's major contributions is to give closer scrutiny to the distinctive spatial institutionalization of labor regulation around specific workplaces. In this respect, he recognizes that the appearance of locally distinctive labor control regimes represent contingent responses to regulatory dilemmas of a more enduring, nonspatially specific nature.

From such an analysis of the wider spaces of labor regulation, it is possible to work back into the local or regional territorial production system and examine the local labor market, firms, workers, labor enclaves, and the 'reserve army' of labor (the unemployed, etc.). At this point in analysis, what are otherwise quite separately constructed control and regulatory processes start to converge around particular places, production systems, labor control strategies, and spatial scales. Yet the form of that convergence is a matter of historical struggle and spatial contingency. In this respect, both the mechanism of labor control and the form of labor regulation must be considered when studying the properties of a local labor control regime, albeit that as we will see later on labor integration can often occur outside formal (state) spaces of regulation.

Peck saw the period from the early 1980s to mid or late 1990s as one of regulatory instability and experimentation with a new labor control regime. This regime involved various state strategies to lubricate the capitalist labor market around more flexible yet intensely coercive workplaces. This period also marks the dismantling of centralized (national scale) forms of welfare provision and the substitution of these by territorially devolved (local scale) institutional arrangements, especially around workfare. As Peck forcefully argues, labor is not just regulated but also, in effect, put in its place. This occurs through various strategies of localization, including new economic modes of welfare provision and political forms of labor regulation. He sees this trend as symptomatic of the active construction of neoliberal regulation in capitalism on a global scale. Another way of thinking about this is that the forces of cultural hegemony reconstitute in post Fordism through practices associated with state workfarism, but that this is happening perhaps on a far more global scale than might have possibly been envisioned by the likes of Gramsci.

Labor control

Workplace → Labor market

Labor regulation

Labor market → Workplace

Figure 1 Directions of causality in analyses of labor control and regulation.

Geographies of Labor Control

Human geographical studies of labor control are part of a wider disciplinary development known as 'labor geography', the aim of which is to examine workers and their organizations (e.g., trades unions) in the economic landscape. If there is a difference between studies of labor geography and those of labor control, it is that whereas the former focuses on labor as an economic and political agent 'for itself' the latter thinks more in terms of the structures and constraints under which wage labor exists 'despite itself' (i.e., its subjugation to new rules in capitalism). This makes the study of labor control not only a distinctive theme within labor geography but also quite different to other branches of economic geography.

Detailed empirical studies of after Fordist production systems have identified a number of additional geo graphical characteristics of the new labor control regime. Some have examined the role of 'labor flexibility' and 'capital mobility' in shaping the new regime. Others have questioned whether transition has occurred at all or been limited to the core.

In the industrial core, post Fordism is most commonly associated with the emergence of flexibly specialized production systems and flexible labor markets. Typically, conditions of flexibility in work and production exist well away from traditional centers of Fordist production and flourish in places like the Silicon Valley or the Third Italy. However, labor scholars and feminist geographers such as Pollert, Hanson and Pratt, and Pollard question whether labor flexibility is a peculiar feature of post Fordist spaces of production or instead simply the his torical intensification of the labor process, especially as more and more women have entered the labor market in a variety of emerging sectors and places. This raises questions about the social and cultural means of inte grating a more flexible workforce into new economic sectors and the attendant gender divisions of labor in specific locations.

There is some evidence that flexible labor control regimes thrive in particular labor markets such as those dominated by female, immigrant, casual, and part time labor. Moreover, flexibility exists in localities and regions where workers are less well organized and employers can experiment more freely with alternative management practices, new just in time supply systems, and less de marcated job categories, as well as subcontracting and flexible work schedules. All of these local characteristics of management and labor control might be essential conditions in the establishment of neo Fordist, post Fordist, or just in time spaces of production simul taneously at the local, regional, and global scales. For example, in his detailed study of the automobile industry in and around Paris, the capital of France, Samers shows how a form of labor regulatory 'structured coherence' developed around the recruitment of immigrant workers from the former French colonies. Presumably, aspects of this coherence have dissipated more recently as class conflict has intensified across the working class suburbs of Paris, as well as in other industrial cities with large spatial concentrations of immigrants.

Knowledge of the geography of labor control explains the increasing spatial mobility of certain branches of industry and factions of capital. The rise of the multi locational firm involves the relocation of particular stages in the labor and production process to places where the requisite labor is more malleable and tractable. In the early and middle twentieth century, branch plants of national enterprises based in Europe and the United States tended to relocate to the suburbs and peripheral regions of host countries, where accordingly more rou tine aspects of production could expand. By the end of that century, such activities were increasingly likely to be located offshore to newly specialized centers of routine high technology functions and semiautomated pro duction tasks in the semiperiphery, such as call centers in Bangalore in India, or zones of maquiladora production in northern Mexico.

These places, no less than those in the industrial core, now have had to confront problems of labor regula tion and control associated with rapid industrialization. Typical of such problems are underinvestment in ne cessary social infrastructure and support facilities for workers, failure to recruit and match necessary skills, and migration of skilled workers away to rival centers offering higher wages. Likewise, new social regulatory issues relating to the social integration of workers confront the maturing industrial agglomerations in the industrial core.

Globalization and Labor Control

Yeung suggests that labor control is an important element in understanding how capital embeds within, or con versely becomes disembedded from, particular places. With respect to these disembedding processes, the soci ologist Ritzer argues the case for a tendency for par ticular technologies of labor control to become universal throughout all of society and its places. He suggests that modern society rationalizes under a single dominant labor control regime, a historical process he terms McDonaldization. Ritzer's otherwise neo Weberian theory integrates the Gramscian notion of cultural hege mony into the rationalization of the labor process. He shows how the cultural integration of workers occurs not simply through the labor process but also around their consumption behavior and reproductive habits. This happens insofar as schools, hospitals, and restaurants as sume to operate along the lines of a McDonaldized labor system.

However, it might be stretching the point to argue that there is a tendency toward a single and dominant 'global labor control regime' along the lines of McDo naldization. For a start, geographers might want to question whether place matters no more within such a regime. A problem harking back to Massey's pioneering work is the danger of overgeneralizing from empirical regularity and observed tendencies. Geographers would not overlook the continuing importance of place based labor control practices to global processes of production and investment.

Figure 2 Actors and institutions in the labor control regime. Reproduced from Kelly, P. F. (2001). The political economy of local labor control in the Philippines. *Economic Geography* 77(1), p. 7. Reprinted with permission of Clark University.

Research in the global semiperiphery highlights how social and economic conditions in place are very im portant in creating a labor control regime suitable for multinational corporate investment. Kelly examines the local labor control regime constructed around foreign direct investment in the burgeoning garment, textile, and electronic goods industry sectors in the Philippines. He finds that, far from there being a preexisting global re gime imposed by multinational corporations, a local labor control regime emerges from the activities of a variety of national and local actors. These range from those of the investing firms to labor organizations and recruitment agencies, and to local village and community leaders. Moreover, many of the practices and institutions involved in building the local labor control regime in the Philippines operate outside formal (state) spaces of regulation and global circuits of investment. **Figure 2** is Kelly's schematic representation of local labor control regime in the semiperiphery.

Conclusion

In recent years, the labor control regime has emerged as a useful concept in economic geography. The 1990s was an especially productive period of research on this topic. However, the initial enthusiasm seems to have waned. In terms of pointing a way forward, it is worth noting that the labor control regime links into other branches of human geography including cultural and political geog raphy. For example, there have been relatively few de tailed studies of how Gramscian notions of cultural hegemony apply, both nationally in the context of pro moting flexible and open labor markets, and locally in terms of detailed knowledge of new place based systems of labor control, forms of community integration, gender relations, and modes of consumption. Further research is required on how public, private, and third sector em ployers are engaged in new modes of cultural and eco nomic integration in the workplace, not least given a wider context of growing tensions around work–life balance, environmental and social sustainability, and the withdrawal of private and state supported pension ar rangements. Recent work by economic geographers such as Bauder potentially points to new ways of approaching and thinking through the relationship between culture, labor regulation, and the geography of labor control.

The empirical study of local labor control regimes does present a number of methodological and ethical challenges. Published data on labor productivity and management techniques are unreliable and hard to come by. There are issues of access to corporate information, employers, employees, and private facilities. There are challenges to obtaining informed consent from inter viewees, not least workers subject to coercive manage ment practices. For these reasons, most empirical studies of the labor control regime have resorted to secondary sources, such as company and workforce histories and biographies, discourse analysis of local or national media, and state policy reviews and political discourse analysis. Primary research on labor control uses ethnography, participant observation, and interviews, and there is a focus upon not just formal regulatory processes but also informal relations between workers, recruitment agen cies, and the living place. Researchers have studied companies, recruitment agencies, and day labor sites, using interviews or surveys of managers, employees, and trade union representatives in a variety of industries, sectors, and places. Other important questions for labor control research include the unethical or racist treatment of immigrant workers by employers (state and private), workers' involvement in living wage campaigns, and how in response to growing pressures employers and busi nesses attempt to manage the modern work–life balance. Information about some of these issues is freely available on various company, government, and labor and not for profit organization websites.

In conclusion, analysis of the local labor control re gime has contributed to a shift in explanatory focus in economic geography away from the politics of 'work' places (i.e., firms and industries) to that of work'places' (labor practices cutting across pre given boundaries be tween firms, industries, labor markets, places, and the state). The local labor control regime remains an im portant concept, not just in terms of representing the entry point into the relations between production, work, and territory, but also because of the possibility to examine how workers and those outside paid work en gage in struggle and resistance to socially undesirable forces of social, technological, and cultural change.

See also: Capital and Space; Corporate Responsibilities; Feminism and Work; Fordism; Fordism, Post-Fordism and Flexible Specialization; Industrial Restructuring; Labor Geography; Labor Market; Labor Unionism; Migrant Workers; Spatial Division of Labor; Transnationalism and Labor Geography.

Further Reading

Bauder, H. (2001). Culture in the labor market: Segmentation theory and perspectives of place. *Progress in Human Geography* 25(1), 37 52.

Berberoglu, B. (ed.) (1993). *The labor process and control of labor: The changing nature of work relations in the late 20th century*. Westport, CT: Praeger Publishers.

Braverman, H. (1974). *Labor and monopoly capital: The degradation of work in the twentieth century*. New York: Monthly Review Press.

Burawoy, M. (1985). *The politics of production: Factory regimes under capitalism and socialism*. London: Verso.

Castree, N., Coe, N., Ward, K. and Samers, M. (2004). *Spaces of work: Global capitalism and the geographies of labour*. London: Sage.

Cox, K. R. and Mair, A. (1988). Locality and community in the politics of local economic development. *Annals of the Association of American Geographers* 78, 307 325.

Gordon, D. M. (1984). Capitalist development and the history of American cities. In Tabb, W. K. & Sawers, L. (eds.) *Marxism and the metropolis*, pp 21 53. New York: Oxford University Press.

Gramsci, A. (1971). *Selections from the prison notebooks*. London: Lawrence and Wishart.

Hanson, S. and Pratt, G. (1992). Dynamic dependencies: A geographic investigation of local labor markets. *Economic Geography* 68, 373 405.

Herod, A. (ed.) (1998). *Organizing the landscape*. Minneapolis: University of Minnesota Press.

Jonas, A. E. G. (1992). Corporate takeover and community politics: The case of Norton Company in Worcester. *Economic Geography* 68, 348 372.

Jonas, A. E. G. (1996). Local labor control regimes: Uneven development and the social regulation of production. *Regional Studies* 30, 323 338.

Kelly, P. F. (2001). The political economy of local labor control in the Philippines. *Economic Geography* 77(1), 1 22.

Marx, K. (1976). *Capital*, vol. 1. Harmsworth: Penguin.

Massey, D. (1984). *Spatial divisions of labor: Social structures and the geography of production*. New York: Methuen.

Peck, J. A. (1996). *Work place*. New York: The Guilford Press.

Pollard, J. S. (1995). The contradictions of flexibility: Labor control and resistance in the Los Angeles banking industry. *Geoforum* 26, 121 138.

Pollert, A. (1988). Dismantling flexibility. *Capital and Class* 34, 42 75.

Ritzer, R. (1996). *The McDonaldization of society*. Thousand Oaks, CA: Pine Forge Press.

Samers, M. (1998). 'Structured coherence': Immigration, racism and production in the Paris car industry. *European Planning Studies* 6, 49 72.

Storper, M. and Scott, A. J. (1989). The geographical foundations and social regulation of flexible production complexes. In Wolch, J. & Dear, M. (eds.) *The power of geography: How territory shapes social life*, pp 21 40. Boston: Unwin Hyman.

Storper, M. and Walker, R. (1989). *The capitalist imperative: Territory, technology and industrial growth*. New York: Blackwell.

Warde, A. (1988). Industrial restructuring, local politics and the reproduction of labor power: Some theoretical considerations. *Environment and Planning D: Society and Space* 6, 75 95.

Yeung, H. W. (1998). Capital, state and space: Contesting the borderless world. *Transactions of the Institute of British Geographers* N.S. 23, 291 310.

Relevant Websites

http://www.marxists.org
Antonio Gramsci, Marxists Internet Archive.

http://world.honda.com
Community, Honda Worldwide.

http://www.csreurope.org
CSR Europe.

http://www.employersforwork lifebalance.org.uk
Employers and Work Life Balance.

http://www.ilo.org
International Labour Organization (ILO).

http://www.livingwagecampaign.org
Living Wage Resource Center.

Labor Flexibility

C. Benner, University of California, Davis, Davis, CA, USA

Glossary

Employment Relationship between worker and employer, including management practices and compensation systems.

Flexibility The ability to change or react to change with little penalty in time, effort, cost, or performance.

Flexible Specialization Production system organized around use of skilled labor, computerized technologies, and networked firms in order to meet specialized often niche markets.

Flexicurity Policy strategy that attempts to enhance the flexibility of labor markets, work organization, and labor relations on the one hand, and to enhance employment security and social security on the other hand.

Functional (Internal) Flexibility It involves labor practices, such as multitasking, teamwork, and broad job descriptions, that increase the ability of workers inside the firm to adjust to changing demands.

Mass Production Production system organized around large-scale enterprises, creating standardized products designed to take advantage of economies of scale in serving mass markets.

Numerical (External) Flexibility It involves labor practices, such as outsourcing, subcontracting, and use of temporary workers, that allow firms to easily adjust labor deployment.

Work Actual nature of the activities people do while engaged in the process of production.

Introduction

In speaking about labor flexibility, former US Secretary of Labor Robert Reich once remarked that "rarely in international discourse has [a term] gone so directly from obscurity to meaninglessness without any intervening period of coherence". Labor flexibility itself is not new. Indeed, neoclassical economists argue that the market itself is a pure form of promoting flexibility, while ex periences of fragility and mutability have characterized the experience of economic actors in many sectors and countries throughout the history of capitalism. But the notion of labor flexibility became nearly ubiquitous in the last decades of the twentieth century, as analysts tried to understand the dramatic economic changes that seemed to threaten the relative stability and predict ability that had come to characterize post World War II labor markets in the developed world.

The concept of labor flexibility gained particular salience with the publication of Piore and Sabel's *The Second Industrial Divide.* Their argument, in essence, was that the economic crisis of the 1970s was resulting in changes in the dominant mode of economic organization, with mass production systems increasingly being replaced by 'flexible specialization' systems that de pended on multiskilled labor often utilizing new com puterized technologies. Since then, the issue of flexibility has occupied a central position in debates on economic organization, regional development, new forms of eco nomic competition and capital accumulation, as well as the restructuring of labor markets. The debates have been particularly heated in relation to flexibility in the labor market, given the high stakes involved in shaping people's livelihoods. While at least some forms of labor flexibility seem to be increasingly important for firms to remain competitive in today's rapidly changing marketplace, all too often companies' drive for flexibility is simply a desire for decreased deregulation and the freedom to hire and fire employees at will. Many workers, their advocates, and policy analysts, understandably, challenge the drive for flexibility, showing clear evidence that it results in in security, declining wages, and deteriorating working conditions for a large portion of the workforce.

In addition to the unequal outcomes of labor flexibility, part of the problem also lies in the diversity of types of labor flexibility and the lack of clarity around the term itself. Labor flexibility can be associated with short term, cyclical, and temporary economic circumstances, as well as deeper structural changes, and is influenced by political forces and changing regulatory regimes, along with eco nomic changes. To understand the diversity of processes of labor market restructuring clearly requires a sophisti cated appreciation for differing historical and geographic trajectories, and a sensitivity to the dangers in simplifying complex social phenomenon under a term as 'loaded' as flexibility. Trying to understand these diverse processes solely under the term 'flexibility' is clearly inadequate. Nonetheless, rapid change and unpredictability is a key characteristic of contemporary labor markets, and thus understanding the nature of this labor flexibility has been of central importance in economic geography.

Economic Changes and Flexibility

At the core of debates over labor flexibility are efforts to understand the changes in the global economy since the 1970s that have increased the level of instability or volatility workers can expect in the labor market. The most prominent of these changes include the following:

- *Changing boundaries of the economy.* Though a world economy has existed since at least the sixteenth cen tury, in recent years there has been a rapid acceler ation in processes of 'globalization', with increasing globally integrated production chains, international trade, investment, and capital flows. This increasing globalization is made possible by the infrastructure created with the rapid development and diffusion of information technology. Though it may seem contra dictory, the economy is also typified by increasing 'localization'. This phenomenon results partly from long term sectoral shifts in the economy, in which the share of the workforce in industries that serve only a local area (e.g., retail sales, social and health services, education, and so on) have increased. Economic lo calization is also rooted in the increasing importance, for the competitiveness of globally integrated in dustries, of 'relational assets' often rooted in the social interactions, sharing of tacit knowledge and develop ment of conventions, informal rules and common cultures that are built in region specific clusters of firms. Both globalization and localization are increas ing labor market volatility. Increased global com petition, the frenetic pace and harsh demands of global financial markets, and the mobility of pro duction processes greatly increase the pace of indus trial shifts and economic change. Increasing localization results in more people employed in the many, though not all, locally serving industries that have long had unstable employment patterns, such as retail sales and personal services.
- *Changing industrial organization.* Information technology has also enabled new forms of management and control, both within the firm and between firm boundaries, in office and service industries as well as manufacturing. The digitization of information makes possible the simultaneous, precise coordination of economic activity in multiple economic locations and across firm boundaries. This enabled corporations in the 1980s and 1990s to flatten management hierarchies, eliminating multiple layers of middle management, and to increasingly subcontract or outsource major components of their operations, both locally and around the globe. In addition, firms have restructured internal operations into distinct profit centers that blur firm boundaries by bringing market dynamics inside the firm. The result is the increased importance of constantly shifting networks, rather than individual firms, in the structure of economic organization, which increases changes in labor markets as well.
- *Changing forms of competition.* In the stable, mass pro duction markets that dominated the economy prior to 1970s, firms competed primarily by improving their productivity, decreasing their costs, or trying to expand their market share within existing markets. With the development of an information economy, the ability of firms to effectively adapt to changing market conditions, identifying and capitalizing on new op portunities and successfully responding to new chal lenges, has become increasingly important, not just for competitive advantage but indeed for economic survival of many firms. Researchers from a variety of fields have extensively documented the critical im portance of this innovation in building and sustaining economic competitiveness. This constant drive for developing new products and new production pro cesses also contributes to the increased pace of change in labor market dynamics.

While changes in all three of these areas have resulted in changing labor market dynamics and changing demands for labor, there is widespread disagreement about the best ways of understanding these labor market changes.

Analyzing Flexibility

The most common way of analyzing flexibility in the labor market and in production systems is to start with firms as the unit of analysis. In this framework, firms can pursue both 'internal flexibility' and 'external flexibility' or, as it is sometimes put, 'functional flexibility' and 'numerical flexibility'. Internal or functional flexibility involves a series of labor practices that increase the ability of workers inside the firm to adjust to changing demands. Some of these practices involve the skills and capacities of workers themselves, and may involve hiring people with more polyvalent skills or providing training to give existing workers more flexible capacities. Some of these practices also involve changing management prac tices that promote broader job categories (rather than highly constrained jobs with narrow job descriptions) and the promotion of more teamwork and collaborative labor practices. These practices all relate to internal firm processes.

'External' flexibility or numerical flexibility is a series of practices that allow firms to take advantage of external relations. Some of these practices may include accessing specialized skills and expertise, such as through hiring highly skilled contractors or entering into subcontracted business relations rather than hiring people internally. Other practices relate to quantitative changes or fluctu ations in labor demands, such as high levels of employee turnover, the use of temporary and part time employees, increased levels of subcontracting, or use of home work.

In both internal and external relations, labor flexibility may also relate to linking wage rates and other forms of compensation more directly to output. This can include variable compensation schemes (piece rate, incentive schemes, and so on), stock options, or simply individualized employment contracts.

Using the firm as the key unit of analysis in understanding labor flexibility provides certain important insights. The pressures on employers and their responses shapes much of the dynamics and structure of labor markets. Firms pay wages, determine hiring and firing decisions, and provide the bulk of training and career development opportunities. Public policy is most frequently centered on how to influence the behavior of firms and the consequences of particular policies for the ability of employers to compete.

Nonetheless, changes in the world economy since the 1970s have cast doubt on ideas that use the modern corporation as the central organizing principle in the economy. Within a broad neoclassical framework, an increasing interest in transactions costs has opened up the examination of the firm as a governance structure and the relationship between firms and markets as means of economic coordination. Network theory moves beyond this market hierarchy spectrum, arguing that networks provide an entirely different form of economic governance and coordination. The firm itself is increasingly seen as a socio spatial construction embedded in broader discourses and practices – essentially serving as sometimes temporary platforms for organizing production. Relations of trust, 'untraded interdependencies', milieux of innovation, learning communities, firm culture – these are all analytical categories that have developed in recent years that recognize the fluid boundaries of firms while identifying sources of competitiveness that lie outside of firm activities.

Thus, some analysts have increasingly argued that it is more useful to go beyond only using the firm as the fundamental unit of analysis in studies of labor flexibility. The wide spread use of telecommunications technologies means that many of the very tasks people perform on the job increasingly cut across firm boundaries, requiring workers to communicate more regularly with customers, suppliers, colleagues in other firms, and so on. Skills required on the job are arguably less firm specific and instead require greater knowledge of trends in the sector or among customers. Thus, a framework that assumes strong firm boundaries in the nature of work activities frequently misses these aspects of work. Similarly, employment relations stretch across firm boundaries and across multiple firms over time. Single worksites and production processes may include workers employed by multiple firms, such as when temporary, contract, and 'permanent' workers all are employed side by side. Entire supply chains have now become the object of management's attention. Workers themselves are arguable, moving more frequently from firm to firm. The skills and experiences that workers develop are built across multiple organizational contexts. To understand these trends requires a greater attention to workers' career paths and understanding the relationships that get built across these multiple employment contexts over time. A framework of labor market flexibility that does not account for these trajectories, with the ability to trace networks across firm boundaries, also is limited in its explanatory power.

Finally, the attention to the boundaries of firms leads to a core periphery perspective, in which firms are seen to reduce the size of their core workforce, while accessing 'noncore' employment through the use of various external flexibility mechanisms. This approach, however, does not address the issue that much of the employment that is external to a 'firm' may in fact be quite central or core to an 'industrial complex'. Some contingent workers, for example, have high wages, significant career mobility, and a clear preference for contingent employment arrangements, and many of these roving specialists may provide crucial services to a string of individual firms at particular points in time, but would not be considered core to any single firm. On the other side of the coin, many workers are full time, permanent 'core' employees of certain firms whose activities could be described as 'peripheral' to an industrial complex. This includes activities such as building services, commercial security services, cafeteria services, landscaping services, and so on.

Thus, an alternative approach to analyzing labor market flexibility is to make a distinction between flexibility in work and flexibility in employment relations, recognizing that both categories cut across firm boundaries.

The term 'work' refers to the actual nature of the activities people do while engaged in the process of production. It includes the physical and mental processes required, the tools and technology used, and the relations with other people – customers, co workers, colleagues from other firms, suppliers, and so on – that they engage in during the process of performing those activities. 'Flexible' work – rapid change in the arena of work with little penalty in time, effort, or performance – directly influences the demand for labor in the way often described as numerical flexibility – fluctuating demand in the quantity of work needed in any particular enterprise. Flexible work also refers to changes workers experience on the job, the changing value of their skills and experience in the marketplace, the changing social relationships and patterns of communication required to be effective in performing their work, and so on.

Work is distinct from employment. This term refers to the nature of the relationship with the employer, the processes employers use in directing, motivating, and monitoring workers activities and the nature of compensation provided for the activities performed. Temporary employment, part time employment, self employment, independent contractors, and various compensation schemes that link compensation to performance (stock options, performance evaluations, etc.) are all issues related to flexible employment, but may or may not be directly linked with flexible 'work'. For example, temporary employees at an electronic manufacturing services firm doing routine assembly work may be performing routine work (at least for a period of time) while being engaged in flexible employment.

Geography and Labor Market Dynamics

Geography plays an important role, in particular, in shaping the dynamics of labor flexibility. Labor geographers have illuminated essentially three ways that geography matters for labor relations, providing particular insights into the complex and variable dynamics of labor flexibility. First, labor relations are embedded in spatial relationships that operate at multiple scales, from local to global. The geographical relationships within which labor market actors operate significantly shape the possibilities for social relationships, facilitating certain types of flexibility and constraining others. Thus, something as simple as workers being able to physically reach their workplace depends on a certain degree of territorial proximity, and is shaped by the quality of regional transportation systems, and spatial patterns of residential and industrial/commercial land use. Similarly, global production systems are dependent on the spatial infrastructure of communication, transportation, and electric power distribution systems as well as the characteristics of labor quality and regulatory systems in any particular place. Labor flexibility, like all forms of labor relations, involves labor market actors connecting both 'horizontally' within particular territories, and 'vertically' across multiple sites within complex spatial divisions of labor or migratory labor networks. The character of different places (including skills and demographics of the workforce, regulatory and tax environment, and types and quality of necessary infrastructure) and the spatial dimensions of these labor relationships shape the types of flexibility different labor market actors try to pursue or are willing to tolerate. Furthermore, the strategic choices labor market actors make are not simply a result of an economic landscape, but also in turn shapes that landscape in an ongoing socio spatial dynamic. The economic landscape is a reflection of, as well as constitutive of, labor relations.

Second, these economic landscapes are highly differentiated spatially. The people, firms, and labor market institutions that shape labor relations come together in unique ways in different locations, related to the particular nature of their embeddedness in different horizontal and vertical spatial relationships, ranging from local to global scales. Labor market actors seek to utilize this spatial variation to gain advantages and to minimize their vulnerability to labor restructuring. Again, the variation in these characteristics of particular places enables certain forms of flexibility and constrains others, leading to highly differentiated strategic choices for labor market actors located in different places – choices that further reinforce this spatial variation. This is perhaps most obvious in comparing different national experiences with flexibility. Thus, for example, while short term employment relations, high use of temporary help firms, extensive subcontracting arrangements, and individualized compensation may be widely accepted in the US, such arrangements may be staunchly resisted in Germany or France, which have stronger labor market regulations and higher levels of national level industrial bargaining. Yet such geographical differences also affect patterns of flexibility within countries. For example, regional differences in things such as levels of unionization, workforce characteristics, sectoral structure of the economy, or even cultural norms and expectations shape the forms of flexibility that labor market actors try to pursue.

Finally, labor market actors' decisions enfold over time, and the cumulative effect of their decisions is reflected in their spatial relationships. These spatial relationships maintain significant social importance often long after the time they were initially created. The economic landscapes that get 'laid down' in previous rounds of economic activity continue to shape how new relationships unfold, again shaping the types of flexibility that can be pursued in particular places. Thus, for example, many territories that developed in the context of mass production systems, with factories employing large numbers of people in standardized, regular shift based employment, often have particular spatial arrangements of residential and industrial space, with particular transportation systems connecting them and particular social norms associated with regular work hours and separation between home and work life. Such spatial arrangements and cultural norms often constrain the development of production systems based on many different worksites employing smaller numbers of people (often in office and commercial environments, not industrial), with more varied work hours and constantly shifting employment relationships.

It is important to note that with the development of new technologies along with cheaper global transportation and communications, firms have been generally more able than workers to take advantage of 'spatial fixes' to labor problems they encounter. Thus, in the simultaneously antagonistic and mutually dependent nature of the relationship between workers and employers, employers frequently have more power and are thus able to pursue forms of labor flexibility that meet their interests more readily than many workers.

Flexibility and Labor Market Policy

The distinction between different types of flexibility is especially important when considering labor market policies and regulation. Since the 1980s, many govern ments, particularly in Europe, Asia, and the developing world, have sought to promote economic competitiveness by deregulating labor markets, thereby promoting greater 'labor flexibility'. The arguments put forward are that lightly regulated labor markets, with low unemployment benefits, tight benefit conditionality, weak trade unions, and low coverage of collective bargaining, deliver higher levels of growth and employment than more regulated systems. In these 'flexible' labor markets, workers will arguably be more quickly redeployed from declining sectors or regions to growing sectors or regions. In es sence, the argument is that by enabling the labor market to respond more quickly, the risk of long term structural unemployment and investment in stagnant industries is minimized. The argument suggests that there is a direct tradeoff between promoting labor flexibility and pro moting social welfare, at least as it has been pursued through labor market regulation.

Many studies, however, have challenged these con nections between lightly regulated labor markets and economic performance. Empirically comparing growth rates across Europe and North American, there is little correlation between growth rates and overall measures of labor market flexibility. Theoretically, many analysts point out that light labor market regulation and low so cial benefits may in fact 'hurt' growth. For example, workers with access to low unemployment benefits of limited duration may be forced to take 'any' job that happens to be available, when it would be better for them and the economy in general to wait longer, improve their skills, find a better match, work more productively and earn higher wages.

Ultimately, the relationship between flexibility and security, on the one hand, and flexibility and growth, on the other hand, depends on what type of flexibility is being considered. Here, the distinction between work and em ployment is critical, as the simultaneous goal of promoting flexible work and employment security can be pursued through different means. In Europe, the concept of 'flex icurity' has taken increasing hold. Flexicurity has been defined as "a policy strategy that attempts, synchronically and in a deliberate way, to enhance the flexibility of labor markets, the work organization and labor relations on the one hand, and to enhance security – employment security and social security – notably for weaker groups in and outside the labor market on the other hand." The ap proach has been particularly strong in Denmark and the Netherlands, but has also been taken up by the Inter national Labor Organization (ILO), the European Union, and a number of specific European countries.

The specific mix of policies in efforts to promote 'flexicurity' vary depending on the context. Promoting flexibility in work tasks often includes efforts to pro mote skill development and life long learning, as much as they do more flexible work tasks and mechanisms to improve labor market adjustment to changing com petitive conditions. Efforts to promote security include not just efforts to promote income or job security, but also efforts to improve social wage supports, re em ployment systems, and greater employability in the external labor market.

See also: Fordism, Post-Fordism and Flexible Specialization; Labor Control Regime; Labor Market; Labor Markets, Regional; Spatial Division of Labor; Transnationalism and Labor Geography.

Further Reading

Arthur, M. and Rousseau, D. (eds.) (1996). *The Boundaryless Career: A New Employment Principle for a New Organizational Era*. New York: Oxford University Press.
Auer, P., Cazes, S. and International Labor Office Employment Strategy Department (2003). *Employment Stability in an Age of Flexibility: Evidence from Industrialized Countries*. Geneva: ILO.
Benner, C. (2002). *Work in the New Economy: Flexible Labor Markets in Silicon Valley*. Oxford: Blackwell Press.
Bernhardt, A., Morris, M., Handcock, M. and Scott, M. (2001). *Divergent Paths: Economic Mobility in the New American Labor Market*. New York: Russell Sage Foundation.
Blank, R. (ed.) (1994). *Social Protection versus Economic Flexibility: Is There a Trade Off?* Chicago IL: University of Chicago Press.
Cappelli, P. (1999). *The New Deal at Work: Managing the Market Driven Workforce*. Boston, MA: Harvard Business School Press.
Carnoy, M. (2000). *Sustaining the New Economy: Work, Family, and Community in the Information Age*. New York: Russell Sage Foundation.
Freeman, R. (2005). Labor market institutions without blinders: The debate over flexibility and labor market performance. Working Paper No. 11286. Cambridge, MA: National Bureau for Economic Research.
Herod, A. (2001). *Labor Geographies: Workers and the Landscapes of Capitalism*. New York: Guilford Press.
ILO (2004). World summit for social development. ILO High Level Meeting. Geneva: International Labor Organization.
ILO (2005). Labor market flexibility, stability and security. 7th European Regional Meeting. Geneva: International Labor Organization.
Massey, D. B. (1995). *Spatial Divisions of Labor: Social Structures and the Geography of Production*. New York: Routledge.
Melvin, B. (1994). Labor market flexibility: A changing international perspective. *Monthly Labor Review* 117, 53 60.
Muffels, R. (ed.) (2008). *Flexibility and Employment Security in Europe: Labour Markets in Transition*. Cheltenham: Edward Elgar.
Neumark, D. (ed.) (2000). *On the Job: Is Long term Employment a Thing of the Past?* New York: Russell Sage Foundation.
Ozaki, M. (1999). *Negotiating Flexibility: The Role of the Social Partners and the State*. Geneva: International Labor Office.
Peck, J. (1996). *Work Place: The Social Regulation of Labor Markets*. New York: Guilford Press.
Piore, M. J. and Sabel, C. F. (1984). *The Second Industrial Divide: Possibilities for Prosperity*. New York: Basic Books.
Pollert, A. (1988). Dismantling flexibility. *Capital and Class* 34, 42 75.

Rainnie, A. and Herod, A. (2007). Spatializing industrial relations. *Industrial Relations* 38, 102–118.
Rousseau, D. (1995). *Psychological Contracts in Organizations: Understanding Written and Unwritten Agreements*. Thousand Oaks: Sage Publications.
Scott, A. J. (1988). *New Industrial Spaces: Flexible Production Organization and Regional Development in North America and Western Europe*. London: Pion.
Standing, G. (1999). *Global Labor Flexibility: Seeking Distributive Justice*. New York: St. Martin's Press.
Wilthagen, T., Tros, F. and Lieshout, H. V. (2004). Towards 'flexicurity'? Balancing flexibility and security in EU member states. *European Journal of Social Security* 6, 113–136.

Relevant Websites

http://www.tilburguniversity.nl/faculties/law/research/flexicurity/
Tilburg University: Tilburg Flexicurity Research Programme.

Labor Geography

T. D. Rutherford, Syracuse University, Syracuse, NY, USA

Glossary

Business Unionism A strategy practiced by many North American unions in the post-World War II boom era of narrowly focusing on advancing wages and benefits for their membership in exclusion of making links with wider social movements or engaging in aggressive organizing. This also meant accepting managerial prerogatives over business strategy, investment, and the workplace.

Internationalism A strategy which involves unions and the labor movement organizing or engaging in solidarity across different nations. Dating back to the nineteenth-century internationalism has taken new impetus from globalization and increasingly involves networks of unions and labor activists organizing between advanced and developing economies not only over more traditional workplace issues but also trade, social, and environmental policies.

Labor Market Intermediary Refers to institutions which match labor supply to demand. This can include public and private employment agencies, professional associations, and unions. Such institutions often play a critical role in urban or other local labor markets characterized by employment in turn characterized by high turnover not only in high-skill new-economy industries but also in services such as cleaning and catering.

Spatial Division of Labor An important concept developed by Doreen Massey and Richard Meegan in which a region is viewed as the outcome of successive rounds of investment or accumulation reflecting the role the region has played in the wider national and international division of labor. It thus refers to the way different tasks in production are allocated to particular groups of people in particular locations. Thus, at any point in time the spatial division of labor is both the legacy of past investment and class relations and an influence on future investment and class relations.

Spatial Fix A concept first developed by the Marxian research of David Harvey. It refers to the way in which the process of capital accumulation is spatially uneven and that capital seeks places in which the conditions for profit realization are superior to other locations for reasons of less competition, stronger market growth, and/or lower labor costs. Thus while capitalists may attempt to resolve any crisis of profitability via introducing new products or technology, work reorganization, and intensification, they may also do this by relocating to these places.

Introduction

Labor geography's development from the mid 1980s onwards reflects debates and developments which occurred both outside and within human geography. In social sciences generally and especially history and sociology, researchers such as E. P. Thompson, Herbert Gutman, and others were pioneers in recognizing the role of subaltern and working class agency in wider social change. In a context of increasing labor and social unrest which characterized advanced capitalism during the 1960s and 1970s, researchers in North America and the UK adopted a range of Marxian and radical theoretical frameworks, which challenged elitist or Whiggish interpretations of history.

Within human and especially economic geography, the adoption of structural Marxist perspectives on geographical and economic change lead to the development, by Doreen Massey, of meso level concepts such as the spatial division of labor. Although the spatial division of labor was principally firm focused, as this concept developed, the importance of labor and class relations in determining the form of local change was stressed. Moreover, increasing criticism of the often abstract and deterministic nature of structural Marxism leads to greater attention to the agency of workers and other actors in shaping change, especially in the UK where this shift was facilitated by a number of government funded research programs on the socioeconomic restructuring of localities.

Thus, by the mid 1980s human geographers began a serious examination of the agency of labor in shaping capitalism. In the United States, Peet derived indexes of high and low class struggle states based on patterns of strike activity. Research also reflected concerns over the adverse impacts of economic restructuring and the increasing adoption of neoliberal policies by advanced capitalist states. This emerging approach was best summed up by Herod as a move from a geography of labor which was seen to characterize neoclassical and Marxian perspectives, to a labor geography, in which the construction of place and space reflected the collective capacity of labor to impact change at different scales. The latter perspective recognized that the biological and social reproduction of labor was inevitably place bound. Since labor is relatively immobile in comparison to capital, it often places greater significance on the economic viability of localities in ways that can militate against the development of a wider class consciousness. Nonetheless, labor's struggles means that capital is not the sole agent in the construction of landscape and indeed workers are also capable of developing their own 'spatial fix'. In theoretical terms labor geography became increasingly eclectic. Marxism still influences labor geography but many labor geographers adopt less essentialist and sometimes poststructural positions in which labor is seen as more fragmented and differentiated. Whatever perspective is used, labor geography views the scaling and the landscape of capitalism as the outcome of conflicts between labor and capital's different spatial fixes.

Based on this broad consensus, labor geography grew quickly during the 1990s examining patterns of trade union organization, the role of working class agency in the creation of agricultural landscapes, and the rescaling of worker struggles especially internationally as globalization became more significant. Labor geographers published in all of the important journals in human geography with *Antipode* – the journal of radical geography and a number of key edited collections from this journal, being an especially critical outlet. As labor geography developed, there was also an increasing recognition of the differentiated nature of labor in terms of gender, race, and ethnicity which challenged traditional workplace and hierarchical forms of union organization. This contributed to new forms of organization based on social movement unionism stressing greater links to communities not only within localities and nations but also internationally. Outsourcing and new forms of work organization such as lean production also attracted attention as did worker organization outside of traditional manufacturing in services such as cleaning, catering, and hotels. All these themes will be explored in greater detail below.

Key Themes within Labor Geography

Geographies of Trade Unionism

In most labor geography the most important institution for worker agency in capitalism is the trade union. The trade union is a formal association for the collective representation of employee interests and it is usually organized on the basis of 'locals' representing workers in a given enterprise or firm. Typically, locals are part of unions representing workers in a particular industry or sector at a national and even international scale. Unions can be occupational or craft based, employer specific, organized industrially irrespective of occupation, or general in that they represent workers regardless of industry or sector. Because trade unions negotiate the terms and conditions of work with employers, they can have an important impact on the economic geography of capitalism. Moreover, the ability of unions to be organized and negotiate effectively can be highly uneven and often dependent on local factors. Thus, labor geographers including Herod on US dockers unions negotiating technological change and Mitchell who stressed how struggles over defining union organizers as part of the local community during migrant worker union organizing in California in the 1930s have often explored how differences in local leadership, organizing traditions, and community institutions and economies contribute to an uneven geography of trade union organization.

Many of the conditions of union organization are inevitably local, but labor geographers also stress how national and increasingly global factors play an often determining role on the impact of unions. Researchers in the 1980s and 1990s emphasized the importance of legal frameworks at state or provincial as well as national scales in regulating unions and overall employee–employer relations. More lately, geographers such as Peck, Herod, and Wills have engaged more directly with industrial relations perspectives, while arguing for the significance of geographic differentiation on how industrial relations systems operate.

These understandings of unions assisted labor geographers to analyze how unions were undermined in the 1980s and 1990s by economic restructuring and the increasing adoption of neoliberalism. Thus, research by Clark in the United States argued that a contributing factor to the post 1975 crisis of unions was the way in which the 1930s New Deal defined unions in terms of local interests and how National Labor Relations Board (NLRB) decisions during the 1980s increasingly favored employers. In the UK, Sunley and Blomley also emphasized the role of changes in labor law in the defeat of the 1984–85 miners strike. They also analyzed regional patterns of decline and change, while Massey and Painter as well as Martin, Sunley, and Wills, debated whether unions had been able to maintain their core regional strength in the wake of restructuring and an increasingly hostile regulatory environment.

During the 1990s, labor geographers studied organizational innovations by unions as they responded to the challenges posed by neoliberal restructuring. The traditional workplace centered and often formal hierarchical orientation of unions was seen as vulnerable if not increasingly irrelevant, as firms reorganized work via outsourcing networks (often on a global basis) and the share of unionized employees declined in most industrial nations. The ending of the post war industrial relations consensus in the advanced economies meant that more bureaucratic 'business unionism' with its narrow focus on improving wages and benefits was also questioned. Increasingly, in confronting neoliberalism unions were also making links to wider community movements and coalitions not only to sustain investment and employment but also to protect overall social and environmental programs. While unions often had important links with their communities – for example, in the United States during the organizing drives of the 1930s and 1940s – contemporary community or social unionism challenged union structures and strategies. As documented by Wills and other labor geographers, unions shifted toward a more networked, if not heterarchic, organizational forms in which the traditional focus of unions around workplace wages and conditions was less important than broader coalitions with a range of social groups over a wider set of issues. Social movement unionism meant that while unions could draw on a rich legacy of organizational experience, they were increasingly involved in coalitions with diverse groups in which they might not exercise authority and leadership. These campaigns reflected a blurring of workplace and community interests – especially where this involved those traditionally more marginalized by unions such as women and workers of color. It could also involve international social movements as globalization led to protests against organizations such as the World Trade Organization (WTO) and the International Monetary Fund (IMF).

Internationalism

Internationalism became an increasing important strategy for labor in the 1990s and 2000s. Yet as labor geographers revealed, this trend was marked by tensions and contradictions. Indeed, as Herod has documented unions had a long international tradition dating from the nineteenth century of involvement in various socialist internationals to the role of large US international unions during the Cold War period in organizing non Communist unions in developing nations. With the end of the Cold War and as globalization developed, internationalism took on a new dimension. As key sectors of manufacturing and later services began shifting to newly industrializing economies, workers and their unions in the advanced economies were confronted with dilemmas of how best to react to this challenge. In many instances they pressed their respective governments for protection of their industries from imports. As such, workers in developed nations might practice 'accommodatory solidarity' in which the spatial interests of advanced economy workers were protected although this went against stronger international solidarity with developing nation workers. Yet as research by Johns on US Guatemala union networks showed, workers could also develop 'transformative solidarity' strategies in which protecting class interests internationally prevailed. Despite tensions and overlap between these strategies, the struggle between the United Steel Workers in West Virginia and the Ravenswood Aluminum Company documented by Herod revealed that unions across a wide range of nations could support the cause of workers and win by using the firm's global networks against itself.

Globalization also raised the issue of how labor organized and fought campaigns. For many labor geographers such as Waterman and Wills, globalization undermined classical labor internationalism based on nationally oriented trade unions. Furthermore, as noted earlier, their hierarchical and bureaucratic nature and their narrow focus on a homogenous (i.e., male) working class subject made classical labor internationalism ill suited to participate in many contemporary struggles, whether they are anti globalization, or environmentally and community oriented. However, freed from the dead hand of Cold War politics, unions were seen as reacting to the threats and opportunities presented by globalization by developing a new labor internationalism, or global social movement unionism (GSMU). Thus, globalization actually offers an important opportunity for unions in north and south to find common ground since both are increasingly confronted by the specter of rapid shifts in investment and neoliberalism. Like social movement unionism, GSMU is based less on official bureaucratic ties and more on direct contacts between shop floor workers and community and grass roots organizations and broader campaigns of solidarity between workers in North and South. As such, it is argued that this new international unionism is, like capitalism itself, becoming more network based and is one in which labor does not have a privileged role, but is nonetheless central to a broader struggle.

However, there are great problems which confront any move toward GSMU. Thus, as Johns and Vural documented, clothing worker union's campaigns with consumer groups against sweatshop labor both domestically and in developing nations, while achieving some positive aims, also illustrated conflicts over tactics (boycotts versus organizing, protectionism versus greater trade openness) between these different groups. Yet despite such tensions, attempts to develop international solidarity are not confined to fostering North–South links. Thus, Lambert and Webster show how unions in the Southern Hemisphere have connected democratic and independent unions struggles in South Africa and Australia and have coordinated a 1 day general strike by unions in South Africa, India, and Korea.

Yet thus far continental economic integration such as the EU and NAFTA has advantaged capital rather than labor. Thus, Hyman, Wills, and Sadler have argued that European integration has contributed to the erosion of relatively strong national industrial relations systems, and newly created institutions such as European Works Councils (EWC) have provided more of a forum for an employer dominated corporate monolog to worker representatives than it has encouraged pan European unionism. However, there is some evidence that even though they are relatively weak institutions, EWCs have potential to create greater international collaboration between unions, while other research reveals that NAFTA's North American Agreement on Labor Co operation (NAALC) can be used to expose poor labor standards and anti union practices.

Despite evidence of a shift toward GSMU strategies, this is not simply a linear move from local to national to global scales. Because even global capital must of necessity locate somewhere, locally based union cam paigns can be effective if they are focused at strategic locations of TNCs. Moreover, in some cases national unions have developed over international ones even as globalization leads to increased economic integration. Thus in the mid 1980s, the Canadian Auto Workers (CAW) broke from their US based international union, the United Auto Workers (UAW). As Holmes and Rusonick's study reveals this reflected not only insti tutions and leadership differences but increased uneven development between the US and Canadian automobile industries. Moreover, there is debate among unions and labor geographers that, with unions struggling to maintain themselves in their own national space, whe ther making global campaigns the principal strategy for labor is an effective use of their resources. Thus, Gindin argues that nationally based (as opposed nationally fo cused) unionism did not preclude global solidarity, but could be internationally progressive if by example it could show that labor could confront and defend its interests versus capital.

Gender, Race, and Ethnicity

Initially, most research in labor geography focused rela tively exclusively on unions and often implicitly assumed a homogenous, white male subject. Yet both globally and increasingly in advanced economies not only were the great majority of workers not unionized, but they were mostly workers of color or were women who had tradi tionally not been strongly organized by unions. During the 1970s and 1980s, feminist writers and especially socialist feminists such as Rowbotham drew attention to the gen dering of work and the activism of women to address in justices to them not only as workers but as women and other identities. By the 1990s, labor geographers increas ingly stressed the differentiated and diverse nature of labor and the challenges confronting these workers in organizing who were often highly disadvantaged in the labor market due to their identity. In so doing, labor geographers often adopted feminist and poststructural approaches em phasizing identity and performativity which deepened and complicated the more traditional concerns of labor geog raphers with class. Feminist inspired research by Hanson and Pratt on workplace–domestic sphere relations in Worcester, Massachusetts revealed how women experi enced this differently than men and, along with ethnic diversity, helped determine patterns of occupational seg regation and shape attitudes toward unionization. Wright's studies on the Mexican Maquiladoras revealed that sys tematic devaluation of women's bodies and work by their male counterparts and employers contributed to high levels of violence including a wave of unsolved murders against women. The local community and unions were largely indifferent until as series of campaigns by women began to highlight this violence.

These issues were further highlighted as labor geog raphers investigated workers organizing outside of the traditionally more white male industries of mining, automobiles and steel manufacturing, such as cleaning, catering and retailing, and other services characterized by predominantly female and non white workforces. For many of these groups identity struggles coincided and, in some cases, superceded more traditional union concerns in the workplace. The need to respect difference yet build sustainable coalitions between different groups also led to greater stress on workplace–community links such as the Justice for Janitors campaigns by a principally Latino workforce in Los Angeles documented by Savage, but also by Tufts and Wills respectively on hotel employees in Toronto, Canada, and London, UK. The fact that in many cases employees in these industries might also be illegal immigrants often increased their marginalization. Yet organizing these workers challenged unions in other ways. Often the predominantly white male leadership of unions was unable to connect with the identity, recognition, and in some cases the workplace struggles of their member ship. Unlike the mass workplaces of industries like auto mobiles, in catering and cleaning and other services, employees might be scattered in small groups over a host of workplaces and thus organizing drives required greater time and resources. This could also lead to innovations in union organizing as in the Justice for Janitors campaign in Los Angeles, in which the traditional model of workplace organizing was replaced by regional collective agree ments. Moreover, in developing stronger links with a di verse community unions might devote greater resources to nonwork 'cultural activities' examined by Houston and Palido and Tufts, such as staging plays, skits, sponsoring choirs, and participating in arts festivals. An important question was how to reconcile traditional union concerns with greater equality and redistribution with the need for greater recognition by these groups and whether in creasing the community orientation of unions by recog nizing different identities served to develop solidarity or weaken it.

New Forms of Work Organization

A major theme in labor geography is the organization of work. Initially, labor geographers dealt with issues of work organization and new technology more as one factor influencing trade union organization – thus, for example, Herod's research on the impact of containerization on US east coast unionism during the 1960s. During the 1990s, new forms of work organization and technology became a more central concern. As debates over the impact of Japanese production methods most notably lean production in such industries as automobiles began to intensify, labor geographers became increasingly interested in the labor process, how technology was utilized, and the social relations between workers and managers. This empirical concern coincided with attempts to more deeply theorize work, labor markets, and the reproduction of labor under capitalism by researchers such as Hanson and Pratt noted above and later Peck. Work and labor market organization were recognized as a spatially segmented and uneven process which featured contests for the control of work between labor and capital and institutions which regulated skill formation and overall social reproduction.

In understanding the impact of new forms of work and technology, labor geographers also emphasized the active agency of labor. Capital needs to control labor and also to gain its consent and develop its subjective capacity. Given this, workers and unions can assert influence in shaping and humanizing work, but the extent of this influence has weakened due to intensifying competition and volatility in labor markets as the ability of firms to access a global labor force has increased. Nonetheless, as Herod documented, workers and their unions could take advantage of new forms of production organization such as Just in Time (JIT) systems in the automobile industry which, through its reliance upon low inventories and tight delivery schedules, made firms highly vulnerable to employee industrial action. Similarly, labor geographers studied the adoption of lean production methods in the automobile and other industries which was often associated with an intensification of work and, as Leslie and Butz showed, led to increasing injuries to worker's bodies. Yet other studies also revealed that lean production could be negotiated by unions that modified work practices in ways that go against a simple diffusion of a 'one best practice' as suggested in such management oriented publications as *The machine that changed the world*. As Rutherford and Gertler showed in a study of the Canadian and German auto industries, the ability of union locals to negotiate the adoption of lean production depended especially on national scale union resources and industrial relations institutions.

Labor geographers have also researched the impact of new forms of work organization in nonmanufacturing sectors such as services. Thus, work intensification, outsourcing networks, in addition to increasing shares of contingent (part time/temporary and flex time) employees have become the hallmarks of many sectors such as financial services, the public sector, retailing, and cleaning and catering. For example, Reimer examined the impact of privatization on catering and cleaning employees in the UK and a recent co edited Aguiar and Herod volume based on studies of cleaning workers in North and South documented the impacts of deregulation, and new forms of work design and intensification on their struggles. More generally, as the traditional work domestic sphere divide has become a blurred and information technology (IT) linked 'work nexus', as Allen and Henry drawing on Beck argue, risk (of unemployment, income loss, etc.) is being shifted from employer to employee. Similarly, in emerging cultural and creative industries such as new media, even high skilled employment is project based such that an employee's tenure in any one workplace may not extend beyond an immediate contract. Labor geographers such as Christopherson note how this situation, coupled with a highly educated young workforce sometimes unsympathetic to collective bargaining, makes new media, cultural, and information and communication technology (ICT) industries difficult for unions to organize. Yet as both Christopherson and Benner's research reveals, some unions have adopted an older guild hall strategy to represent contract and other employees as labor market intermediaries by providing training, healthcare, pensions, and other benefits from which they are often excluded. Like in the Justice for Janitors campaigns these shifts coincide with a move from individual workplace toward regional or local bargaining strategies which match shifting patterns of employment. More generally, as national systems of industrial relations have either weakened or decentralized, unions in Germany, Italy, and Scandinavia have developed a stronger regional and local orientation in part to deal with large firms outsourcing to SMEs and other suppliers on this geographical basis. Yet despite this rescaling of industrial relations governance, national systems still matter to employee representation and work organizations – thus even in new media, Swedish and German employees are much more likely to be part of a union or works councils than their North American counterparts. In the former nations, firms in both more traditional manufacturing and knowledge intensive new economy sectors also rely on these institutions for initiating and implementing innovation.

Retrospect and Prospects for Labor Geography

Over the past quarter century, labor geography has made a vital contribution to the development of human geography. In recognizing the agency of labor not only has it generated a wealth of empirical research but in so doing it has developed a greater sense of the dynamic in the development of the scaling and landscape of capitalism as a product of labor's struggles. It has also recognized and provided insights into the differentiated nature of labor both within the workplace and the community in contributing to the uneven nature of labor organization. Despite such an undeniable and rich legacy, several criticisms can be made of the approach. As an over reaction to the earlier narrow focus on capital, at times labor geography has abstracted labor from the broader political economy of restructuring. While it is true workers have both individual and collective agency in creating their own spaces, this occurs within a broader context of capital accumulation. Thus, labor geographers need to be careful not to view labor or the spaces it creates, as the result of a purely autonomous agency. There has also been a general eschewing of the implications of labor's struggles for the political economy of capitalism. The theoretical eclecticism of labor geography has no doubt facilitated its ability to highlight difference and avoid essentialism, yet it may also fetter abstraction and synthesis.

Also understandably, labor geography has focused on the scalar politics of worker struggles and distributional and solidarity issues rather than dealing with labor's agency within production. Thus, despite the existence of a significant institutional and industrial relation's literature on unions and innovation, little attention has been paid in the labor geography literature to the role that unions play in shaping technological change and firm innovation, or with a few exceptions such as DiGiovanna, cluster/industrial district governance. Similarly, while researchers such as Clark and Rutherford have focused on labor's relationship with the state (especially national) and state policy formation, overall this has been a lacuna in labor geography. Furthermore, labor geographers have begun to engage more directly with industrial relations approaches, but because of the more empirical and localist tendencies noted above, they have not addressed labor's role in, for example, the wider institutional political economy of innovation.

Indeed, some might question whether unions deserve the singular attention paid to them by labor geographers given the reality that in most advanced and developing economies only a minority (and a declining minority at that) of employees are unionized. Certainly, despite an increasing interest in wider community movements, there is a relative lack of research by labor geographers on professional and other associations in which workers are also involved. Indeed, for some observers, unions are seen as increasingly irrelevant in an individualized economy, which is no longer considered as characterized by the formal employment relation of capitalism. Post structuralist geographers such as Gibson Graham strongly imply that unions and those in the formal economy are incapable of generating anti capitalist movements and indeed by their very nature both discursively and materially reproduce capitalism. Rather, informal or diverse economies outside of the formal wage economy are seen as generating noncapitalist spaces and the communities around them are seen to challenge capitalism by developing alternatives to wage work.

Yet, if Marx's prediction of a revolutionary proletariat has thus far proved illusive if not illusionary, it can be argued that whatever insights Gibson Graham have produced into the diverse dynamic of informal economies; these often are directly linked via outsourcing and other networks to capitalism. Nor is it clear what strategies this leaves for unions and the great majority of workers who are engaged in formal wage work. Indeed, the last two decades have witnessed the rapid expansion of the global wage labor force, not its decline. Evidence also suggests that rather than the decline of unions being the result of some objective structural shift in the nature of work, its causes are more fundamentally institutional and legal. Where enabling legal frameworks remain in place such as in Scandinavia, unionization levels have remained high if not expanded over the last 25 years – even in new economy sectors. Surveys cited by Boxnall and Hayes in the United States, where unionization has fallen the most precipitously, suggest that the representation gap – the difference between those wishing to be in a union compared to those that are – is at a historical high constituting between one third and one half of the employed workforce. There remain good reasons for labor geographers to focus on unions for although they are not the sole associational activity of employees, they are one of the few independent institutions for representing their interests as workers. Furthermore, Gibson Graham rightly maintain that productive activity is not confined to the formal capitalist economy, yet it is difficult to conceive of any progressive, much less revolutionary change, without some kind of independent representation and organizational activity by wage laborers in capitalism – especially given their strategic position within the formal economy.

Still, whatever strategic position unions and workplace organizing may play in the labor geography analysis, there remain questions about how best these struggles should link to the community and other social movements. Recent commentaries in journals such as *Antipode* involving Amin and Thrift and labor geographers such as Hudson show that there continues to be debate about whether unions and more traditional labor concerns should continue to play a strategic role in a contemporary heterarchical left. Thus while its contribution to human geography is undeniable, for many labor geographers the theoretical and empirical challenge remains as it was articulated by Sheila Rowbotham, Lynn Segal, and Hillary Wainwright over a generation ago – namely how to go beyond the fragments?

See also: Fordism, Post-Fordism and Flexible Specialization; Labor Control Regime; Labor Market; Labor Markets, Regional; Spatial Division of Labor; Transnationalism and Labor Geography.

Further Reading

Aguigar, L. and Herod, A. (eds.) (2006). *The Dirty Work of Neoliberalism: Cleaners in the Global Economy*. Oxford: Blackwell.
Amin, A. and Thrift, N. (2005). What's left: Just the future. *Antipode* 37, 220 238.
Boxnall, P. and Hayes, P. (2005). Employee voice in the Anglo American world: An overview of five national surveys. *Labor and Employment Relations Association Series, Proceedings of the 57th Annual Meeting*, 7 9 January, 1 12.
Castree, N., Coe, N., Ward, K. and Samers, M. (2004). *Spaces of Work: Global Capitalism and Geographies of Labor*. London: Sage.
Christopherson, S. (2002). Project media in context: Regulatory change and the new geography of media. *Environment and Planning A* 34, 2003 2015.
Clark, G. (1989). *Unions and Communities under Siege: American Communities and the Crisis of Organized Labor*. Cambridge, MA: Cambridge University Press.
Gibson Graham, J. K. (1996). *The End of Capitalism (As We Knew It): A Feminist Critique of Political Economy*. Oxford: Blackwell.
Gindin, S. (2004). Globalization and labour: Defining the 'problem' paper presented at Brandeis University 24 April 2004.
Herod, A. (ed.) (1998). *Organizing the Landscape: Geographic Perspectives on Labor Unionism*. Minneapolis, MN: University of Minnesota Press.
Herod, A. (2001). *Labor Geographies: Workers and the Landscapes of Capitalism*. New York: Guilford.
Hudson, R. (2006). On what's right and keeping left: Or why geography still needs Marxian political economy. *Antipode* 38, 374 395.
Johns, R. and Vural, L. (2000). Class, geography and the consumerist turn. *Environment and Planning A* 32, 1193 1214.
Leslie, D. and Butz, D. (1998). GM suicide: Flexibility, space and the injured body. *Economic Geography* 74, 360 374.
Massey, D. (1984). *Spatial Divisions of Labour: Social Structures and the Geography of Production*. London: Macmillan.
Mitchell, D. (1996). *The Lie of the Land: Migrant Workers and the California Landscape*. Minneapolis, MN: Minnesota University Press.
Peck, J. (1996). *Work Place: The Social Regulation of Labor Markets*. London: Guilford.
Rowbotham, S., Segal, L. and Wainwright, H. (1981). *Beyond the Fragments: Feminism and the Making of Socialism*. London: Merlin Press.
Waterman, P. and Wills, J. (2001). *Place, Space and the New Labor Internationalisms*. Oxford: Blackwell.
Wills, J. (2001). Community unionism and union renewal in the UK: Moving beyond the fragments at last? *Transactions of the Institute of British Geographers* 26(4), 465 483.
Wright, M. (2001). A manifesto against femicide. *Antipode* 33(3), 550 566.

Relevant Websites

http://laborstrategies.blogs.com/global labor strategies/global unionism/index.html
Global labor strategies.
http://www.ilo.org/global/lang en/index.htm
International Labor Organization.
http://www.learningwork.ca/csew/wg/lcu
Learning community unionism center for study of education and work University of Toronto.
http://www.seiu.org/property/janitors/
Service Employees International Union Justice for Janitors.
http://www.caw.ca/index.asp
The Canadian Auto Workers.

Labor Market

P. Carmody, Trinity College, Dublin, Republic of Ireland

Glossary

Agglomeration Economies Economic advantages derived from business activities being located close to one another.

Apartheid System of racial segregation and domination operated in South Africa in the latter twentieth century.

Bretton Woods System System of fixed exchange rates globally established after the World War II.

Capital Regime The rules governing movement of money across borders.

Corporatism A system of state–society coordination, based around organized interest groups such as employers, trade unions, and farmers.

Decommodification Removing something from the market sphere of social relations.

Labor Power The ability to do work.

Lumpen Proletariat Unemployed and excluded underclass.

Primary or Primitive Accumulation Process through which communal land or other assets are made private (theft).

Relational Assets Economic advantages derived from links and interactions with other economic actors, such as knowledge exchange.

Social Reproduction Functions State functions undertaken to ensure that society is able to reproduce itself, generally including healthcare and educational provision.

Wage Goods Goods commonly consumed by workers.

Welfare Regime Set of rules governing access to state welfare entitlements and benefits.

Introduction: The History and Social Construction of the Labor Market

The labor market and private property are the two fundamental and interrelated institutions of capitalist economic systems. In conventional economic theory, labor is considered as a factor of production along with capital, land, and knowledge. In this conception, labor is the people employed in the production of goods and services, as well as their skills and effort. A market is a set of transactions where a commodity or service is bought and sold. The labor market is then the set of transactions where the ability to do work is contractually exchanged, either formally or informally. Markets have geographies and while some, such as the stock market, have tended historically to be concentrated in particular locations, given their foundational importance labor markets tend to be geographically wider.

Conventional economic theory treats labor as a commodity: something produced through, and sold on, the market. However, labor power is not a commodity in the same way as apples are, for example, because it is embedded in workers who are sentient human beings. Consequently labor power cannot be produced and traded in the same way as inanimate commodities, as ownership of work is never fully exchanged, and it is consequently subject to ongoing social monitoring and regulation. Labor is not produced through market mechanisms but in the family, and through the social reproduction functions of the state, such as education. As a result, labor power is what some have called a 'fictitious' or 'psuedo commodity', dependent on nonmarket relations for its production. The fact that the supply of labor power is consequently only partially subject to the price mechanism of the market also calls for social regulation.

In the advanced (post)industrial countries most members of society are dependent on wages for their primary income, because they do not own sufficient assets to enable them to live off these alone. Thus the unequal distribution of property impels some people to put their labor power on the market to ensure their own reproduction. Private property was, in turn, partly created through a process of what Karl Marx called primary, or primitive accumulation, and geographer David Harvey calls "accumulation through dispossession." This was where property owned by no one, or by the community as a whole, was made private. The most famous example of this was the enclosure movement in sixteenth century Britain, where communal lands were fenced off by private landlords. The dispossession which accompanied this process compelled those who lost access to land to sell their labor power on the newly emerging labor market. Newly 'freed' labor was then compelled to sell its services to the highest bidder. Free labor, as a social relation, was therefore created through the widening of the institution of private property. Dispossession deepened the market for commodities, rather than people continuing to be self provisioning. Thus the creation of a labor market helped widen the market for commodities more generally, despite being initially associated with widespread increases in poverty.

The ability of employers to adjust the amount of labor power purchased in line with production needs and the (circumscribed) spatial mobility of labor, so that it can migrate in response to changed economic conditions in different regions account, in part, for capitalism's dynamism, flexibility, and longevity. However, the labor market is imbued with power relations, as workers' de pendence on wages and their inability to limit the supply of labor puts them at a disadvantage in relation to busi ness (capital), as does capital's greater spatial mobility. However, workers are also socialized to want to partici pate in the labor market through family traditions and associations, for example, so that the supply of labor power is not accomplished by the compulsion of market forces alone. Capital may be attracted or repelled by the nature and quality of locally available labor.

The rules governing labor markets are subject to contestation by different social forces. Workers may resist the introduction of new technologies and work practices which are labor displacing, or increase managerial con trol. Exposure to business strategies which treat labor power as a commodity encourage workers to cohere and to form associations to promote their rights – trade unions. Thus the fact that labor is a social relation, as defined by lack of property and control of the labor process, also encourages it to become a collective actor.

Trade unions typically lobby and struggle for more state, and less regulation by the market, so that there are better and more remunerative working conditions for their members; that is the so called 'decommodification of labor'. Labor may successfully impose limits on the spatial extent of labor markets to preserve its members' interests, and agglomeration economies, and attendant labor shortages, may give workers leverage in their struggles with business in particular places. This is also the case where there are specific skill shortages.

On the other hand, employer's associations typically argue that state over regulation is contrary to the inter ests of society, because by raising wages above their free market price, employers will reduce their demand for it, leading to unemployment. Under the Bretton Woods System, which restricted capital mobility until the 1970s, higher overall wages created more demand in national economies and so led to increased employment. Wages thus serve a dual role in capitalist societies as a source of demand, and also as a cost of production. This dualism partly explains why labor markets are sites or structures of conflict and accommodation between business and workers.

In the labor market employers tend to favor: (1) numerical flexibility, that is that they have the ability to hire and fire in line with market conditions; (2) wage flexibility, so that they can adjust wages up or down in line with market conditions; and sometimes (3) functional flexibility, so that they can move workers between jobs within the firm if needed. Often employers wish to re duce wage costs, although it may be possible, in some cases, to raise profits by paying workers more, so that they work harder – an effort bargain based on the so called 'efficiency wage'.

Segmentation, Local and National Diversity: Developed and Developing Countries

Labor markets are constituted geographically, and by oc cupation. The rules which govern the buying and selling of labor power, and the area in which these apply can be thought of as a particular labor market – for example, the labor market for software engineers in the US, or in 'Silicon Valley' in California. These rules are created and governed not only by different social coordination mech anisms, particularly the state and the market, but also by civil society organizations, such as trade unions, and by international organizations, such as the International Labour Organization of the United Nations. Labor mar kets are also segmented or divided locally based on the structure of the economy, occupation, race, gender, labor demand, and other forms of social regulation; that is the different sections, regionally, operate according to differ ent rules.

Different countries have different welfare and labor regimes, or sets of rules, which de commodify labor to varying degrees. The welfare regime mediates the extent to which people are impelled to put their labor power on the market. Labor is substantially more de commodified in Western Europe than in the United States. This is sometimes called the 'European social model'. In Japan, until recently, a widespread practice among major com panies was lifetime employment in order to foster worker loyalty and attachment. In developing countries without substantive welfare states, labor markets are freer and more rapacious.

The idealized vision of perfectly competitive labor markets does not operate, even in developed countries. One study in the US in the 1960s found that only one in five workers operated in unstructured, competitive labor markets. Within firms there are internal labor markets, where people are promoted internally. Even in sup posedly free labor markets, such as the California fruit harvest, the supply of labor is regulated by labor con tractors; while employer's associations organize to min imize wage competition between businesses. More recent attempts to establish a guest worker program in the US show that while labor power is desired, the people in which it is embedded are often not, and should return home after yielding up their work. In Southern Africa under apartheid domination, a similar system which sought to displace the costs of social reproduction to women in rural areas was developed. Male workers often migrated to work in mines in urban areas, whereas women were left in rural areas to raise the next generation of labor.

In many parts of the developing world, but particularly in Sub Saharan Africa, Latin America, and South Asia, most people work in either agriculture or the informal sector. Some argue that the retention of communal property ownership in parts of Africa under colonial rule, and the absence of free labor markets accounts for the underdevelopment of the continent, as capitalist social relations were not generalized, and commodity markets remained shallow. The influential dual economy model, of Sir William Arthur Lewis, posed the problem of economic development in formerly colonized countries as one of the intersectoral labor market: moving labor out of low productivity agriculture into high productivity manufacturing. Policies based on this model encouraged younger and more skilled workers to migrate to urban areas, contributing to uneven development. Some also argue that regulation in the formal urban economy (sometimes called the covered sector) deepened dualism in the labor market by creating a so called 'labor aristocracy' of insiders in secure, regulated employment and an excluded 'lumpen proletariat' in the informal sector. For instance, in some African cities today up to 80% of the labor force is employed in the informal sector. What defines informality is lack of state regulation of the labor contract, and the absence of tax collection, although some governments, as in Malawi, have recently attempted to collect taxes from informal street traders. Geographers and others have noted the functional nature of urban dualism, as the formal sector sometimes relies on the informal sector to provide cheap wage goods.

Informal labor markets are often thought of as unregulated; however, there are social institutions, other than the state, which structure them, such as kinship connections, for example. Affective or family ties may blunt the tendency to take unfair advantage of other's labor, although informal labor market institutions may be extremely exploitative. Given the widespread poverty and lack of alternative income earning opportunities, around the shores of Lake Victoria in Uganda there is a system, *Jaboya*, where female fish traders may exchange sexual favors to be allowed to buy stock from fishermen; making it rational for people to expose themselves to the risk of sexually transmitted diseases. In this instance sex work is traded for access to commodities and markets.

Globalization and Labor Market Restructuring

Since the 1970s the global economy has entered a period of accelerated globalization. This has been facilitated by policies of economic liberalization, pioneered initially in Chile under the military dictator (and geographer) Augusto Pinochet, and subsequently adopted by the governments of Margaret Thatcher in the UK and Ronald Reagan in the US, and by the international financial institutions of the World Bank and International Monetary Fund through their Structural Adjustment Programs (SAPs). These SAPs sought to bid wages down to their supposed market clearing level – to make them low enough to encourage employers to mop up unemployment. However, there are physiological limits to treating labor as a commodity, as workers must be able to afford to eat. These limits were often breached under SAPs in Africa, with disastrous social, and sometimes political, consequences.

The Thatcher and Regan revolutions were fundamentally about restructuring the labor market. In their analysis, the sluggishness of the UK and the US economies in the 1970s was the result of overprotection of workers from dismissal and an overgenerous state, which gave them the option of not working if they so chose. In the 'free market' or neoliberal formulation, the excessive power of labor contributed to wage inflation, undermining national economic competitiveness. Thus, the objective was to abolish the existing wage bargain between big business, labor, and government, and replace it with flexible labor markets based on market regulation. The social policy concomitant to this was workfare, which withdraws universal rights to welfare and makes it conditional on work.

Structural changes in the global economy have meant that in some areas, the young, old, and certain disadvantaged groups have become excluded from the labor market. Sometimes this is presented as weak attachment to the labor market, with a resultant so called 'culture of poverty'. The purported solution is coercive attachment to the labor market through workfare.

One of the main aims of neoliberal globalization is to increase insecurity in the labor market' to eliminate the difference between primary (core, secure) and secondary (insecure) labor markets, and to make them operate on a single global market logic where commodification is the norm. Neoliberal globalization can be thought of as an attempt to create a single global market for commodities, including labor. It attempts to reduce the complexity of local social relations (the places of labor) to the spaces of capital. The role of the state in increasing labor market flexibility has been central in this.

Attempts to create a global labor market are achieved through trade and capital regime liberalization, which facilitate strategies such as 'whipsawing', where corporations reduce wages and working conditions by moving their factories to low cost overseas locations, or extract concessions from their workers by threatening to do so. The spatial mobility of multinational corporations in terms of where they locate their plants; where there are low corporate tax rates, for example, enables them to engage in what has been called regulatory arbitrage; creating pressure to bid down corporate tax rates and regulations globally, although there are also counter tendencies at play driven by agglomeration and scale economies. Sometimes these bidding down pressures are called a 'race to the bottom'. For example, some states in the Southern US (the high growth 'Sun Belt') have so called 'right to work' laws, which enable workers not to have to join trade unions if they do not wish to, and not to have to observe strikes. Another way to think of these laws would be as 'right to compete' laws, which may strengthen the power and discretion of employers.

There have also been attempts to introduce the global labor market logic regionally through the controversial European Commission Services Directive which would enable consumers to contract workers under the rules pertaining in the country of origin, rather than in the country in which the work is undertaken. In Europe this is sometimes referred to as the 'Polish plumber' issue, and some fear it will contribute to a regulatory race to the bottom in terms of conditions of labor service. However, even if there is regulatory convergence, local labor markets would still be unique because of differences in labor quality and rates of mobilization.

Labor is also an agent of globalization, and its global organization predates the emergence of transnational corporations. Yet labor's relations with capital are contradictory, as it wants to increase wages, but also maintain employment and hence competitiveness. In the US, what has been called labor's 'go global' policy has often been pursued in collaboration with capital to facilitate the US geoeconomic interests, particularly market opening. In other cases, however, trade unions reject collaboration and a narrow focus on regional labor market issues in favor of attention to broader societal concerns: social movement as opposed to business unionism.

Under conditions of economic globalization, it is the cost properties of wages which come to the forefront of government policy, as they seek to contain them and/or raise labor productivity to achieve, or maintain, international competitiveness.

The greater mobility of capital under globalization has led to pressure toward convergence on the neoliberal Anglo American approach globally. As a result of regulatory arbitrage, there has been a shift in taxation away from capital toward less spatially mobile factors of production: land and particularly labor. However, higher taxes on labor raise the cost of employing workers, encouraging more capital intensive production methods, thereby potentially raising unemployment. It may however be possible to reduce taxes on both labor and capital, thereby attracting investment away from other locations, while raising the absolute amount of tax raised. In a European context this is sometimes referred to as 'social' or 'regional dumping', as locations with higher tax rates have less employment and higher social security expenditure.

In terms of tradable goods and services, almost everywhere is potentially in competition with everywhere else under economic globalization, linking wages in different places through trade and investment. The labor market and wages are also regulated through macroeconomic policy. In order to prevent what is seen as excessive wage inflation, central bankers in the advanced economies ensure that unemployment does not fall below a certain level. Sometimes governments attempt to lower wage costs by encouraging more immigration. Some argue that the state is anxious to maintain a 'reserve army of labor' to ensure continued private sector profitability, thereby displaying its pro business bias. Governments may also liberalize trade to reduce domestic prices and contain pressure for wage increases, using the global market to discipline labor.

The increased skill intensity of production and the decline in demand for semiskilled labor in rich countries is related to changed trading patterns and technological development. In order to retain or develop capital within the national economy, there is pressure on governments to 'vocationalize' the education system; to focus on the production of skills which are of use to business. This is sometimes presented as a process of upskilling to enable movement into more advanced economic activities where there is less competition, and which are consequently more profitable. Private business tends to under invest in skills development, because workers can move to better paying employment after they have been trained. Where labor markets are competitively regulated excessive staff turnover may prove costly for capital. Consequently, it often falls to the state to invest in education to solve this collective action problem for business.

'Active labor market' or 'activation policies', which seek to reduce the mismatch between supply and demand for labor, through training and education, were pioneered in the Scandinavian countries. Government responses to the problem of skill production, in part, determine whether economies follow 'high' or 'low road' paths, where increasing labor productivity through better skills and technology characterize the former, and the bidding down of wages the latter.

Skills are also fungible or 'leaky' across national boundaries. There is a (nested) global labor market for highly skilled labor, with a substantial proportion of engineers in Silicon Valley coming from India, for example; part of the so called brain drain, where tax payers in poor countries subsidize skill development for the high income countries. Business can use restrictions on immigration to its advantage by relocating to lower cost labor markets, as evidenced in the software sector in India, for example, where engineers typically earn roughly a tenth of what they would in the US. Where business cannot relocate, because it provides nontradable services or because of geographically embedded relational assets or agglomeration economies, low cost immigrant labor may be sourced to provide janitorial services, for example. There is a debate about the effects this type of immigration has on wages; particularly whether it results in job displacement of indigenous workers, or raises economic growth and demand, as indigenous workers would not undertake these types of low skill, tedious jobs. The balance of academic research favors the latter interpretation.

Some argue that in order to sidestep social restrictions on its profitability, capital is now trying to detach itself from labor, by investing in globalized financial activities, such as currency speculation. For example, General Motors, the major US car manufacturer, earns much of its profits from its credit business, rather than the sale of cars *per se*. In order to reduce wage and control costs there is also an ongoing substitution of capital for labor in some industries, with robots now often undertaking much of the car assembly, for example. However, as robots do not consume commodities, a question is who will buy the cars produced in such a scenario? This is partly responsible for a shift of work toward the service and information sectors in rich countries.

The Gendered Geography of Labor Markets and Homeworking

For some, the center of the labor market is the relationship between home and work, which is deeply gendered. Women's segregation in the labor market is closely related to what are perceived to be societally appropriate roles for them. Women in paid employment often work closer to home, in more spatially restricted labor markets than men as a result of obligations associated with social reproduction functions in the household. This spatial restriction on job search areas may partially account for pay differentials between genders (the gender pay gap), although lower wages may mean women cannot afford some modes of transport, such as private cars, which might enable them to travel further afield. However, research in the US found that African American and Hispanic women commuted as far to work as their male counterparts and spent more time commuting than Euro Americans, perhaps because they are largely reliant on public transport.

The greater flexibility of labor markets globally has been associated with greater female participation and weaker labor contracts. The number of women at work in the US doubled from 1970 to 1997, from 30 to 60 million as overall wages stagnated or fell, and a second family income became more important. Feminisation also occurred because employers sought out (female) non unionized labor, and because of women's struggles for labor market inclusion.

Greater labor market flexibility has also been associated with the resurgence of the nineteenth century phenomenon of hiring homeworkers to sew garments, for example. Homeworking reduces overhead costs, such as the need for factories and lighting, and monitoring and control costs, as women engage in self exploitation at home in order to eliminate paying childcare. It also has the other advantage for business of fragmenting labor as a collective actor, and is based on piece rates, where workers are only paid for the number of items they produce. Italy has the highest proportion of people self employed in the older EU states as a result of outsourcing to female homeworkers by companies such as the clothes manufacturer Benetton. Benetton combines nineteenth century work practices with twenty first century technology, such as electronic point of sales devices in its stores, which transmit data on purchases instantaneously to headquarters to facilitate individualized restocking.

Alternative Labor Market Responses to Globalization

While most governments around the world have promoted neoliberal market regulation of labor, there have also been alternatives. A European corporatist response to globalization has been social pacting, where unions, government, and employers agree to contain wage increases in return for tax reductions, protection against layoffs, or improved public services.

The socialist government in France in the 1990s introduced a statutory 35 h work per week to redistribute employment. For a time this was compatible with accelerated economic growth and reduced unemployment. However, it has been criticized on the basis of the 'lump of labor fallacy'. According to critics, this fallacy posits that there is a set amount of work to be done in any society, and that this can consequently be redistributed without welfare losses. However, these critics point out that as people differ in their skills, one person will not necessarily be able to do another person's job, or to do it with the same efficiency. Consequently they argue that it is erroneous to assume that work can be redistributed.

Some argue that strong social security systems are needed to cope with the vicissitudes of globalization. One policy based on this theory, pioneered in the Netherlands, is called 'flexicurity'. This aims to combine flexible labor markets, so that the economy is competitive, with strong social security protection. Others argue that there is a need to go further than this and to reconfigure welfare systems around basic income grants which would be given to all members of society to ensure social inclusion. Some also argue for more individualized and tailored interventions to enable people to move into employment with training and job placement, and the importance of ensuring sufficient demand in the econ omy so that jobs are created: 'bridge building programs'. There have also been innovations in the developing world with the Employment Guarantee Scheme, which guarantees public work to workless households in Maharashtra State; recently extended to a national level by the Indian government, for 100 days a year at least. Although taking different forms, struggles to make labor markets less exploitative will continue to be important in the twenty first century, as they were in the twentieth.

See also: Africa; Apartheid/Post-Apartheid; Brain Drain; Civil Society; Competitiveness; Economics and Human Geography; Economy, Informal; Europe; Feminist Geography, Prehistory of; Fordism; Globalization and Transnational Corporations; Human Geography; Inequality; Informal Sector; Informalization; Labor Control Regime; Labor Flexibility; Labor Geography; Labor Markets, Regional; Labor Unionism; Migrant Workers; Neoliberal Economic Strategies; Neoliberalism; Poverty; Regional Competition, Regional Dumping; Relational Economic Geography; Structural Adjustment; Third World Cities; Underclass.

Further Reading

Custers, P. M. (1997). *Capital accumulation and women's labour in Asian economies*. London: Zed.
De Koning, J. and Mosley, H. (eds.) (2001). *Labour market policy and unemployment: Impact and process evaluation in selected European countries*. Cheltenham: Edward Elgar.
Esping Andersen, G. (1999). *Social foundations of postindustrial economies*. Oxford and New York: Oxford University Press.
Glyn, A. (2001). *Social democracy in neoliberal times: The left and economic policy since 1980*. Oxford and New York: Oxford University Press.
Hanson, S. and Pratt, G. (1995). *Gender, work, and space*. London: Routledge.
Herod, A. (2001). *Labor geographies: Workers and the landscapes of capitalism*. London and New York: Guilford.
Paratian, R. and Dasgupta, S. (eds.) (2004). *Confronting economic insecurity in Africa*. Geneva: International Labor Organization.
Peck, J. (1996). *Work place: The social regulation of labor markets*. London and New York: Guilford.
Peck, J. and Theodore, N. (2000). Beyond 'employability'. *Cambridge Journal of Economics* 24, 729 749.
Portes, A., Castells, M. and Benton, L. (eds.) (1989). *The informal economy: Studies in advanced and less developed countries*. Baltimore, MD: Johns Hopkins University Press.
Prugl, E. (1999). *The global construction of gender: Home based work in the political economy of the 20th century*. New York: Columbia University Press.
Santos, M. (1979). *The shared space: The two circuits of the urban economy in underdeveloped countries*. New York: Methuen.
Standing, G. (1999). *Global labour flexibility: Seeking distributive justice*. London: Macmillan.
Standing, G. (2002). *Beyond the new paternalism: Basic security as equality*. London: Verso.
Sunley, P., Martin, R. and Nativel, C. (2006). *Putting workfare in place: Local labour markets and the new deal*. London: Blackwell.

Relevant Websites

http://www.iza.org
Institute for the Study of Labour (IZA).
http://www.ictur.org
International Center for Trade Union Rights (ICTUR).
http://www.icftu.org
International Congress of Free Trade Unions (ICFTU).
http://www.ilo.org
International Labour Organization (ILO).
http://www.sewa.org
Self Employed Women's Association of India (SEWA).
http://www.etes.ucl.ac.be
The Basic Income Earth Network (BLEN), Université catholique de Louvain.
http://www.epinet.org
The Economic Policy Institute (EPI).
http://www.usbig.net
The U.S. Basic Income Guarantee Network.
http://www.wftucentral.org
World Federation of Trade Unions (WFTU).

Labor Markets, Regional

A. E. Green, University of Warwick, Coventry, UK

Glossary

Employment Field The area in which residents of a specified area actually work.

Labor Demand Jobs offered by employers and vacancies.

Labor Market Arena where those who are in need of labor and those who can supply the labor come together.

Labor Market Intermediaries Organizations and individuals which help in matching of labor supply and demand, and so in the functioning of labor markets.

Labor Shed A line enclosing the area which supplies workers to a workplace.

Labor Supply This consists of people who are employed, as well as those people defined as unemployed or economically inactive, who can be considered to be potential workers.

Low-Skill Equilibrium A situation in which employers compete in low value-added markets and demand relatively low skills from employees, which is ultimately reflected in the supply of skills.

Self-Containment The extent to which an area has cross-boundary flows.

Travel-to-Work Area An approximation to a self-contained labor market based on commuting-to-work patterns.

Introduction

Debates on labor market structures, outcomes, and policies focus on national and supranational , national , regional , and local level developments. Drivers operating at all these scales have uneven geographical impacts, such that there are important variations in structures and experience at regional and subregional scales – hence, the importance of a geographical perspective on labor markets. This contribution focuses on the inherently geographical nature of labor markets and their operation and character at regional and local scales. This is in keeping with the emergence of the 'region' (and subregions) as a significant and extensive level of policy design, service delivery, and governance in many advanced economies.

Following an introduction to conceptual issues underpinning spatial labor markets, this article outlines how regional and local labor markets are not merely functional entities or spatial containers, but are also socio institutional constructs. Traditionally, geographers are interested in spatial patterns of employment and nonemployment. Some of the key dimensions of labor market change are highlighted and geographical variations in labor market opportunities are outlined in this article. Policymakers at regional and local levels are concerned about these variations and what they mean for the achievement of economic competitiveness and social cohesion. Moreover, with increased mobility of labor and capital, local, regional, national, and international influences impact on the fortunes of regional and subregional labor markets and there are increasingly important links between labor markets in different parts of the world, although the focus of this article is on regional labor markets and formal (as opposed to informal) working in 'Northern developed countries'. Before advancing further, however, it is appropriate to outline what a 'labor market' is.

Building on the notion of a 'market' where buyers and sellers come together to exchange goods and services, a 'labor market' is an arena where those who are in need of labor (i.e., employers) and those who can supply the labor (i.e., workers) come together. Employers and the jobs they provide represent the 'demand side' of the labor market and workers represent the 'supply side' of the labor market. For the labor market to operate successfully, there needs to be a 'match' between demand and supply. Hence, employers need information on the characteristics and location of workers, and (potential) workers need information on what jobs are available, the requirements of those jobs, and where those jobs are. Labor market intermediaries, such as public employment services, recruitment agencies, and training providers, play an important role in facilitating the operation of labor markets by bringing demand and supply together. Often, however, there is a degree of 'mismatch' between demand and supply. 'Mismatch' can take two main forms. First, 'skills mismatch' occurs when the skills that workers supply do not match those demanded by employers. Second, 'spatial mismatch' occurs when there are 'jobs without workers' in one area and 'workers without jobs' in another area. The influence that the changing geography of employment has had on the access to jobs by less skilled workers and others who are otherwise disadvantaged in the labor market is a primary concern of the spatial mismatch debate, especially in North American cities. Here the decentralization of employment from central locations to urban peripheries has posed particular problems for those left behind, and especially those reliant on public transport, since public transport networks tend to operate along radial routes to central nodes, and so are not often well suited to the economic geography of the twenty first century.

Conceptual Issues: Regional and Local Scales

The references above to ‘the labor market’ are suggestive of a unity absent in practice. In practice, ‘the labor market’ is composed of a multiplicity of submarkets demarcated by various criteria – including occupation, industry, and, of particular relevance here, geographical area. Hence, the labor market is said to be ‘structured’ in accordance with particular labor market subgroups; although these labor markets for particular subgroups are linked because labor moves between occupations, sectors, and geographical areas.

The geographical subdivision of labor markets is largely a consequence of the monetary and psychological costs of extensive daily commuting journeys to work and the often greater costs of migration between different areas. These costs spatially subdivide a labor force that is already stratified occupationally, because by deterring migration and restricting commuting they tend to restrict the labor market to that geographical area which is accessible from a given residence.

Most individuals seek work, and most employers seek most of their workers, within a limited geographical area. The area within which employers seek their labor is known as a ‘labor shed’, while the area in which residents of a particular local area work is termed an ‘employment field’. Here, it is worthnoting that the effective employment field is delimited by the employment opportunities that workers are aware of. These may only be a subset of all opportunities that are objectively within reach, because there may be spatial and social barriers that influence labor market perceptions, causing some workers to search for work in only a small geographical area or to discount some job opportunities that are, in fact, available.

The term ‘local labor market’ is used to describe a relatively self contained area in terms of journey to work flows, within which the majority of residents work and the majority of jobs are filled by local residents. The principle of self containment underlies the definition of functional local labor market areas. In reality, however, the ‘boundaries’ are not ‘fixed frontiers’; rather they may be thought of as resistance points on a scale of mobility or as ‘zones of transition’ between different spheres of influence. Administrative areas (e.g., counties, local districts, and municipalities) vary markedly in the extent to which they are self contained in terms of commuting flows, and so many administrative areas are not functional labor market areas, and therefore, some commentators argue, are not suitable spatial units for local labor market policy. In order to overcome such deficiencies, there have been a number of attempts to define functional local labor market areas. One of the best known is the travel to work areas (TTWAs) in the UK and similar constructs in other ‘developed’ economies. These provide an indication of the catchment areas of labor markets, and of how residency relates to workplaces. Hence, they fulfill the aim set out for them of representing a spatial unit for policymaking that reflects the relationship between labor supply and demand within a local area.

As commuting flows have become longer and more diffuse as a result of increasing car use and lower commuting costs, TTWAs have become very large, taking on a regional, as opposed to a local scale, in some instances. As they increase in size, they are increasingly likely to conceal concentrations of high unemployment. TTWAs have been criticized by some geographers and policy analysts on the grounds that they conceal differences between different subgroups of the population. Here, it is salient to note that the majority of commuting journeys remain relatively short: despite debates about the impacts of globalization and hypermobilities, the majority of labor markets are locally specific – reflecting the fact that labor is less mobile than capital. However, there are important differences in commuting patterns between different subgroups of the population. So, for instance, commuting journeys and so local labor market areas for women tend to be smaller than those for men, those for part time workers tend to be smaller than those for full time workers, and those for unskilled workers tend to be smaller than those for workers in professional occupations. The relative costs of commuting are higher for less skilled and part time workers, leading to shorter travel to work distances and less geographically extensive job search than for more highly paid workers. Consequently, those with higher level skills are likely to restrict their choice of jobs more than those who are less skilled. This means that the more skilled workers become the more likely it is that they will search over a wider geographical area. These differences between groups of workers are important in understanding the labor market behavior of different subgroups and are crucial for policymakers concerned with facilitating the matching of labor supply and demand (as discussed in more detail below).

All of the considerations outlined above underline the fact that, in practice, ‘the local labor market’ is not unambiguously defined; rather, there is a mosaic of overlapping submarkets which are variable in their spatial extent.

Regional and Local Labor Markets as Socio-Institutional Constructs

It is important that an emphasis on the spatial definition of labor markets as ‘spatial containers’ in the tradition of regional science does not lead to the importance of social and institutional factors in the formation and operation of labor markets being overlooked. A body of literature in economic geography adopting an institutional per spective on labor markets emphasizes that local and regional labor markets are socially embedded and con stituted institutional spaces, in which formal and informal customs, norms, and practices underpinning employ ment, working practices, labor relations, and wages are played out. They are the result of the interaction between employers' practices, institutions, state policy, and re gional and local labor market histories. Place specific developments matter in understanding how local and regional labor markets evolve over time and how eco nomic actors, agencies, and individuals behave. So, alongside economic factors, social and institutional fac tors are important in understanding how labor markets operate and evolve over time. This is especially the case in relation to ethnic labor markets and informal working arrangements, which have typically been absent from consideration in regional science approaches to labor markets.

Labor market intermediaries are key players in the social and institutional landscape of regional and local labor markets. They have always played a role in the operation of labor markets, although the nature and ex tent of that role has changed over time. Traditionally, the main role of labor market intermediaries was to obtain job placements for workers at local, regional, and national levels. With increased global competition, regional labor markets have grown more complex. Employers are more likely to engage in outsourcing than was previously the norm, some groups of workers have become more mobile within and across national boundaries, and some labor market intermediaries are acting on the global stage in transnational labor markets. Hence, there has been an increase in the scale and scope of activities of labor market intermediaries. Their traditional 'jobs fair' func tion has expanded to incorporate a 'market maker' role with impacts on both the supply and demand ends of the labor relationship. Labor market intermediaries are concerned to a greater extent now than previously with helping employers with the recruitment process and also with advancement and skill improvement for incumbent workers, thus helping improve business productivity and workforce outcomes.

As labor market intermediaries have come to play a more prominent role in regional labor markets, their manifestations have become more varied. In the private sector, they include temporary help firms – which have long played a role in helping employers meet peak de mands for labor, and also consultant brokerage firms – which are concerned with the recruitment of skilled workers, especially to fill skill shortage vacancies. In the public sector, there are public employment services and associated agencies, professional employers' associations, and trade unions. In recent years, there has been a growth in web based information provision and job sites, and to some extent these are replacing some of the more traditional labor market intermediaries. Finally, in the voluntary sector, there are a range of community based organizations which play an increasing role as labor market intermediaries, particularly for groups who are most distant from the labor market (i.e., those with poor skills and with other barriers to labor market entry) and also for growing numbers of migrant workers.

It is important to note that labor market intermedi aries are not just economic organizations; they also play an important role as social networks. Among lower skilled workers who might otherwise lack 'bridging' social networks which are important in gaining work, they may provide new routes into work. For highly skilled workers, they can help strengthen high quality social networks that help support professional advancement and entre preneurial success. Indeed, by extending these social networks to other regions (sometimes on the other side of the world) highly skilled migrants can transplant the institutions and relationships of entrepreneurship in highly economically successful regions – such as Silicon Valley – to other parts of the world.

Geographical Variations in Labor Market Opportunities

The quantity and quality of labor market opportunities at regional and local scales are the outcome of local, national, and international processes. Quantitative vari ations in numbers of employment opportunities are manifest in geographical differences in employment and nonemployment, whereas qualitative variations are reflected in the nature of employment opportunities available and prospects for labor market advancement. When and where labor markets are 'slack' (i.e., when demand for labor is insufficient to absorb the supply of labor), there is particular interest in 'numbers' of jobs in order to reduce mass worklessness. In circumstances of 'tight' labor markets (i.e., when demand for labor outstrips supply), there is an emphasis on attracting new sources of labor (e.g., older workers, migrant labor, and women returning to work after childbearing/rearing) into employment. Geographers have long been interested in spatial variations in the quantity of employment – which matter especially pronounced for those with poor skills and who are otherwise vulnerable in the labor market. However, over time, they have placed greater focus on the variations in the quality of employment and what these mean for regional and local economic de velopment and for the socioeconomic circumstances of individuals.

For analytical purposes, employment is often disaggregated by sector and by occupation. Traditionally, economic geographers have been most interested in sectoral variations in the composition of employment. Historically, there were marked geographical variations in key sectors, reflecting local concentrations of raw materials (especially in mining and extractive industries) and specialist manufacturing traditions, which in some instances are reflected in place/area names (such as 'The Potteries' for the Stoke on Trent area in England). Over time, however, sectoral differentials in employment structures over space have become less pronounced, such that regional variations in employment composition are less stark than previously. For example, rural economies now have a similar sectoral structure to urban ones.

Since the 1970s, mass manufacturing, mining, and agriculture have undergone large job losses in many Western economies. The impact of deindustrialization, even a few decades ago, can remain relevant to contemporary patterns of spatial nonemployment. Many parts of the service sector – including financial and business services, distribution, hotels and catering, education, and health – have experienced job gains. There has been an important shift in the structure of employment in favor of services.

Transformation in the sectoral composition of jobs has had implications for the occupational and skills profile of employment. Changes in occupational structure within sectors have tended to operate in the same direction. The general picture has been one of rapid growth in numbers and shares of managerial, professional, and professional occupations associated with higher level qualifications in the so called 'knowledge economy'. Employment has grown in personal service and sales and customer service occupations also, while for skilled trades, operatives and elementary occupations (i.e., manual workers) the picture is one of decline.

Geographers have become increasingly interested in the shape, character, and implications of more 'professionalized' employment structures and the so called 'knowledge economy' where such jobs predominate. Large metropolitan labor markets in core regions tend to be characterized by a greater demand for and a greater supply of highly skilled labor. As such, they are favored locations for 'dual career households' with two partners wishing to pursue labor market careers. Conversely, in peripheral rural areas the knowledge economy is more 'shallow', productivity tends to be lower and individuals face fewer and less varied employment opportunities.

Although many regional and local labor markets have witnessed a loss of manual and less skilled jobs, in many areas a substantial number of employment openings remain at the bottom end of the labor market for individuals with limited skill requirements. However, in many instances, these jobs are associated with poor working conditions and low pay. In some parts of the world, such jobs, which may be associated with varying degrees of precariousness, may be the only ones available at local/regional level, and individuals may choose, or be obliged, to work in two or more jobs in order to maximize their pay. Where a broader range of jobs are available, those associated with poor working conditions and low pay may be unattractive to many workers and are often associated with high rates of 'churn' between employment and nonemployment (and often concentrations of low skilled immigrants (whether temporary or permanent)), poor labor market prospects, and low incomes.

Policy Issues and Perspectives

Policymakers at supranational, national, regional, and local levels are concerned with raising and sustaining employment rates and enhancing productivity. Traditionally, the main emphasis of labor market policy has been on challenges of labor supply, encompassing both raising employment rates and ensuring that the (potential) workforce has the necessary skills and attributes to ensure employability in a global economy. The labor market is seen as playing a central role in social inclusion, and a lack of participation in the labor market being a key indicator of social exclusion. People with low or no skills are disproportionately concentrated among the socially excluded, as are older people, those from certain ethnic minority groups, those with disabilities, and those in areas of relatively low labor demand. Spatial variations in worklessness are greater than spatial variations in unemployment, and intraregional differences in worklessness are greater than interregional ones. They are a particular cause for concern because they represent an underutilization of local labor resources and can become 'locked in' to local economies, because even when economic conditions improve the long term workless tend to remain at the back of the queue for jobs. In order to address these disparities, targets have been designed to steer policy in the direction of increasing employment rates among disadvantaged groups and in disadvantaged areas.

Despite agreement on the need to address spatial concentrations of worklessness in order to foster both social inclusion and enhance economic competitiveness, there is considerable debate as to how spatial concentrations of worklessness at metropolitan levels can best be tackled. Proponents of a 'segmented model' of the labor market, who believe that the local labor market consists of a number of spatially defined submarkets, suggest that local concentrations of worklessness arise primarily because of deficiencies in highly localized demands for labor, and therefore conclude that worklessness' rates will remain high in the absence of a renewed supply of

Employer demand	Skills surplus: Low	Skills surplus: High
High	Skills shortage imbalance Mismatch caused by companies demanding higher qualifications than are available in the local workforce	High-skill equilibrium A local/regional economy with a strong demand for high-level skills, which has a positive effect throughout the supply chain on enhancing the aspirations and actions of individuals with respect to participation in education and training
Low	Low-skill equilibrium Employers face few skill shortages in a predominantly low-skilled workforce, where there is no incentive to participate in education and training and raise qualification levels and aspirations	Skills surplus imbalance Mismatch caused by a workforce which cannot find local employment to match their skills and aspirations

Figure 1 The relationship between employer demand and skills availability.

appropriate jobs within short commuting distances of the jobless – that is, they favor an approach of 'taking work to the workers'. Conversely, proponents of a 'seamless model' of the labor market argue that city regions are single markets in which transactions between labor and capital take place regardless of the location of residence and employment sites. They argue that 'taking work to the workers' will only have a short term effect at best, because spatial labor markets are permeable and local residents will be subject to city region wide competition for jobs. Hence, the solution is to raise aggregate demand for labor and to upgrade the skills of the workless in order that they are better able to compete for the jobs available. This is because job growth does not necessarily 'trickle down' to local residents – even if local residents have the required skills, higher skilled workers can 'bump down' in the labor market and compete for lower skilled jobs: processes of mobility and job competition leave behind the weakest. In practice, over the years, a range of people based policies (including New Deal initiatives to up skill individuals and help them find work) and area based initiatives (such as Employment Zones) have been implemented. However, the key emphasis has been on supply side solutions to labor market problems.

There is increasing recognition, however, that a focus on labor supply alone is likely to be insufficient to overcome economic problems of low pay and low productivity in regional labor markets. Indeed, recent regional economic strategies, particularly in regions characterized by large scale economic restructuring with the decline of traditional industries, have pointed to the existence of a 'low skill equilibrium' and have identified the necessity of paying attention to labor demand issues, and to retention and advancement in employment, for regional and local economic development. Typical characteristics of an economy in a 'low skill equilibrium' include a coincidence of relatively low pay levels, a preponderance of poorly qualified workers and an underrepresentation of jobs at the upper end of the occupational hierarchy, reflecting an overreliance on low knowledge intensity sectors and a competitive strategy based on low cost. In the face of a lack of demand from employers for higher level skills, there is little or no incentive for individuals to participate in education and training and raise qualification levels and aspirations. Hence, the demand for relatively low level skills is reflected, ultimately, in the supply of skills.

This situation contrasts with a 'high skill equilibrium' situation in which a local or regional economy is characterized by a strong demand for high level skills, which has a positive effect throughout the supply chain on enhancing the aspirations and actions of individuals with respect to participation in education and training. Hence, raising demand for skills has a key role to play in economic development policies concerned with moving from a situation of 'low skill equilibrium' to one of 'high skill equilibrium' (see **Figure 1**). The policy challenge is to convert this into a 'virtuous circle' in which investment in education and skills helps drive competitive success and economic growth. Labor market intermediaries have an important role to play here in bringing together the demand and supply side of the labor market through improving labor market information, job brokering, and the promotion of effective recruitment and enhancing the way that agencies work together with employers and individuals.

Conclusion

Geographers have long been interested in the inherently geographical nature of labor markets (a topic largely ignored by labor economists) and in regional and local variations in the volumes and characteristics of jobs and workers. This article highlights the importance of recognizing multiple geographies of the labor market and how the markets for different types of labor actually operate and are played out spatially. Looking ahead, geographers are increasingly emphasizing transnational links between regional labor markets. The migration of capital and labor is not a new phenomenon, but is now taking place at unprecedentedly high levels. Regional labor market impacts are especially pronounced in terms of the movement of Mexican labor to the USA, increasing labor flows within the European Union, and from Africa and Asia to many parts of the world. The penetration of immigrant labor beyond traditional migrant 'gateways' to smaller settlements and rural areas links regional and local labor markets transnationally to a greater extent than previously, and has implications for regional labor markets in both origin and destination countries. The increasing mobility of highly skilled workers and information, as well as the fragmentation of production in advanced technology sectors, has led to unprecedented opportunities for formerly peripheral economies to develop and compete in new economic sectors. The fortunes of regional labor markets are interlinked to a greater extent than ever before.

See also: De-Industrialization; Industrial Restructuring; Labor Geography; Labor Market; Migrant Workers; Public Policy.

Further Reading

Benner, C. (2003). Labour flexibility and regional development: The role of labour market intermediaries. *Regional Studies* 37, 621 636.
Castree, N., Coe, N. M., Ward, K. and Samers, M. (2004). *Spaces of Work: Global Capitalism and Geographies of Labour*. London: Sage.
Coombes, M. G., Green, A. E. and Owen, D. W. (1988). Substantive issues in the definition of localities: Evidence from sub group local labour market areas in the West Midlands. *Regional Studies* 22, 303 318.
Finegold, D. (1999). Creating self sustaining, high skill ecosystems. *Oxford Review of Economic Policy* 15(1), 60 81.
Gilloth, R. P. (2004). *Workforce Intermediaries for the Twenty first Century*. Philadelphia, PA: Temple University Press.
Goodman, J. F. B. (1970). The definition and analysis of local labour markets: Some empirical problems. *British Journal of Industrial Relations* 8, 179 196.
Green, A. E. and Owen, D. (2006). *The Geography of Poor Skills and Access to Work*. York: Joseph Rowntree Foundation.
Hanson, S. and Pratt, G. (1992). Dynamic dependencies: A geographic investigation of local labour markets. *Economic Geography* 68, 373 405.
Hepworth, M. and Spencer, G. (2004). *A Regional Perspective on the Knowledge Economy in Great Britain*. Report for the Department of Trade and Industry. London: Local Futures Group.
Herod, A. (2001). *Labor Geographies: Workers and the Landscape of Capitalism*. New York: Guildford Press.
Martin, R. and Morrison, P. (2003). *Geographies of Labour Market Inequality*. London: Routledge.
Massey, D. (1984). *Spatial Divisions of Labour*. London: Macmillan.
Morrison, P. S. (2005). Unemployment and urban labour markets. *Urban Studies* 42, 2261 2288.
Peck, J. (1996). *Workplace: The Social Regulation of Labor Markets*. New York: Guildford Press.
Saxenian, A. L. (2006). *The New Argonauts: Regional Advantage in a Global Economy*. Cambridge, MA: Harvard University Press.
Vance, J. E. (1960). Labor shed, employment field and dynamic analysis in urban geography. *Economic Geography* 36(3), 189 220.

Relevant Websites

http://www.statistics.gov.uk
Map of 2001 TTWAs in the UK.

Labor Unionism

A. Herod, University of Georgia, Athens, GA, USA

Glossary

Fordist Economy Refers to the way in which the economy of the US and a number of other industrial economies were organized during the period from about 1945 to 1970. It is characterized by mass production of standardized commodities, mass consumption (often funded by easy access to credit), and an interventionist national government pursuing Keynsian macro-economic policies. The archetypal business model was one in which, among other things, firms engaged in vertical integration, relied upon a full-time laborforce, and production chains were geographically extensive (with research and development occuring in one region and manufacturing in another).

Politics of Scale Refers to the manner in which the spatial scales at which social life operates are constructed through political and economic conflict. Spatial scales, then, are social products, rather than naturally given ways for dividing up the world into various scalar units ('the national', 'the regional' etc.).

On the Structure of Labor Unionism

It is hard to determine when the first labor unions (sometimes also referred to as trades unions) formed. Some have suggested that if we consider labor unions to be organizations whose members come together for mutual assistance in the face of their employers, then the history of labor unionism is very long, stretching back some two and a half millennia to when Greek artisans banded together to exert some modicum of control over their employment conditions. Likewise, during the Middle Ages in Europe workers like stonemasons joined together to form guilds, again to exert some control over their worklife. Modern unionism, however, really emerged in the late eighteenth century in Europe, be ginning with various 'friendly societies' (also sometimes called 'benevolent societies' or 'mutual aid societies') whose goal was to provide assistance to members, typi cally for burial. Unlike guilds, whose members were confined to particular professions, friendly societies' members often worked in a variety of trades but came together around some commonality, such as a shared religion or political views, or the fact that they lived in the same neighborhood. Despite the fact that they were illegal in many countries – in Britain unions were pro scribed for much of the early nineteenth century under the 1799 Combination Acts, for instance – the coming of the industrial revolution led to the growth of unions in mining, manufacturing, and agriculture.

Historically, unions have generally organized them selves on the basis of particular crafts, particular indus tries, or particular geographical areas. As the name suggests, 'craft unions' are those whose members belong to particular crafts – carpenters, plumbers, bakers, and the like – and it was these kinds of unions which were the earliest to form. In the twentieth century the rise of mass production encouraged 'industrial unionism' to emerge in many economic sectors, with workers in a particular industry – say steel – joining the same union, regardless of which specific job they did in the workplace. Finally, some unions have organized geographically to include all workers in a particular area within their ambit, regardless of the work they do (in the United States the nineteenth century Knights of Labor adopted such a model). Whereas many unions have focused upon workplace issues, others have also focused upon issues impacting workers' lives beyond the workplace – such as the pro vision of union run housing for members.

If unions may be distinguished according to whether they are craft, industrial, or geographically based, they can also be typed by the spatial level at which power is ex ercised – are they highly centralized or decentralized? – and by what functions they perform. With regard to the first matter, in most countries unions are made up of local branches and national level organizations. The key ques tion concerns the political relationships between these two levels. Thus, in the United States the model which has emerged since the 1930s has generally been one in which local union branches retain significant power vis à vis the national body. In contrast, in countries like Germany na tional level union bodies have significant ability to enforce their will upon local union branches and to force them to accept national union policies. With regard to the function performed by unions, one of the most significant dis tinctions relates to whether such unions are 'free' or whether they are state run. 'Free' unions, though they may be connected to political parties which may even end up forming the government, primarily are supposed to fur ther the interests of their members. By way of contrast, state run unions typically function to ensure that gov ernment economic planners' wishes are enforced – hence, the Soviet model of unionism envisaged unions as 'transmission belts' between the government and the shopfloor, with their role being to make certain that production quotas were met in the factory or mine.

Although much recent writing concerning globalization has tended to ignore unions, it is important to recognize that they have a long history of acting internationally, going back to the mid nineteenth century. Such international activities have primarily been conducted through two sets of international organizations: associations of unions working in particular economic sectors and associations of national union centers. The former – historically called International Trade Secretariats but more recently renamed Global Union Federations (GUFs) – cover a number of sectors, including metalworking, mining, journalism, education, and chemicals. Currently, there are ten GUFs, of which the International Metalworkers' Federation, the International Transport Workers' Federation, and the International Federation of Chemical, Energy, and General Workers' Unions are some of the best known.

Of the latter, the most influential in the past few decades has been the International Confederation of Free Trade Unions (ICFTU), a body made up of over 230 national union centers (including the American Federation of Labor Congress of Industrial Organizations (AFL CIO) and the British Trades Union Congress) from 154 countries representing some 155 million members. Although the ICFTU was the largest such organization, there are two others which have played similar roles, these being the World Federation of Trade Unions (WFTU) – primarily made up of Communist and Arab unions during the Cold War and now largely moribund, though still claiming 129 million members in 130 countries (as of 2004) – and the World Confederation of Labor (WCL), which was initially constituted by Catholic trade unions in Europe but subsequently became more secular (as of October 2001 its membership of some 26 million drew from 116 – mainly developing – countries). (In 2006 the ICFTU and WCL merged, creating the International Trade Union Confederation.) Each of these has operated regional organizations (e.g., the ICFTU's Organización Regional Interamericana de Trabajadores) and industry focused entities – the GUFs in the case of the ICFTU, a number of Trade Union Internationals in the case of the WFTU, and several International Trade Federations in the case of the WCL. Finally, in addition to such organizations there are myriad other structures through which unions cooperate internationally, including the International Labor Organisation of the United Nations, the World Social Forum, and entities such as the US National Labor Committee in Support of Democracy and Human Rights in El Salvador.

Geography and Labor Unionism

Within the geographical literature, issues of unionism have been incorporated in a number of ways. For much of the twentieth century unions were largely ignored and workers were only dealt with somewhat tangentially through being viewed as 'factors of production' – for instance, the famous industrial location model of Alfred Weber saw 'labor' only in terms of its cost to industrialists. In response to this lacuna, in the 1970s and 1980s an engagement with Marxist theory led a number of geographers to reexamine the role of labor in the structuring of economic geographies. Thus, David Harvey sought to understand how the incorporation of living labor into the capital circulation process shaped the construction of capitalism's landscapes, whereas Doreen Massey argued that capital produced particular spatial divisions of labor at different times and that these divisions subsequently affected how capitalism developed spatially, given particular locales' geographical 'path dependence'. However, such approaches, while raising the question of how capitalists' efforts to extract surplus labor from workers shaped the economic landscape's genesis, tended to have a rather one dimensional view of labor, viewing it in terms of how its price, skill level, rates of unionization, and the like affected the decision making of capitalists – what has been termed a 'geography of labor' view.

By the early 1990s, however, issues of workers and their organizations (such as unions) began to be reevaluated with the emergence of what some have called 'labor geography'. Principally, a growing number of geographers sought to explore how workers' lives are spatially constituted and how, in turn, workers (through their organizations) are active geographical agents whose practices have an impact upon the production of the geography of capitalism, at scales from the very local to the truly global. As such, labor geography was an attempt to counter both the taxonomic approaches to labor which had dominated in mainstream industrial location analysis and the capital centric approaches of early Marxist geography. Although the theoretical and political intent of labor geographers was to examine the spatialities of workers' lives in toto and how the geography of capitalism (or other social systems) may both enable and constrain workers' social praxis, in practice it has tended to focus upon the activities of unionised workers. Four reasons account for this.

First, many labor geographers have actively worked with and/or have been members of unions, which has undoubtedly impacted their research interests. Second, while appreciating that unions are not the only organizations through which workers may express their economic and political aspirations, many analysts nevertheless see them as being important vehicles for worker empowerment. Third, given that unions usually have highly formalized structures and rules for operating, they tend to be easier to research than are other, less formalized worker organizations – they tend to keep more archives, for instance. Finally, as with capitalists, workers' ability to achieve their goals is usually greater when they act collectively than when they act individually. The latter two points in particular are significant because much of the focus upon unions has been not to understand them *per se* (although that is certainly important) but, rather, to use them as empirical exemplars to make theoretical interventions concerning how economic landscapes are made. Consequently, at the time labor geography was emerging as a subfield focusing upon such formally structured entities was seen as the most productive way of challenging the capital centrism of the then dominant modes of thinking and of persuading skeptics of the conceptual points labor geographers were trying to make – that the geography of capitalism is not the product solely of capital, that economic landscapes' spatiality shapes how workers behave socially (i.e., that space is constitutive of social praxis and not just reflective thereof), and so forth. More recently, however, the focus of labor geography has begun to broaden to examine how the social practices of non unionized workers and those belonging to other types of social movements shape, and are shaped by, their spatiality.

Areas of Focus

In terms of how the study of unions has been articulated within geography there are, at least, four (overlapping) areas of focus: (1) how labor law and state regulation influence the spatial structure of unionism; (2) how unions have shaped economic landscapes; (3) how spatial concerns affect the political geography of labor organizing; and (4) how the particularities of place impact the nature of unionism.

Labor Law and State Regulation

Some of the earliest work within geography relating to unionism concerned how its practice was shaped by the regulatory framework within which unions have to operate. A prominent writer in this regard was Gordon Clark, who was primarily interested in two issues – challenging traditional industrial location theory for its failure to account for how regulatory structures, specifically the law, shape economic geographies, and exploring how labor law influences the geography of unionism. Focusing upon the United States, Clark argued that US labor law privileged 'local' systems of labor management relations and worker representation, in contrast to the more 'nationally' oriented systems in many European nations, and that such localism had made it more difficult for unions to develop national strategies through which to counter nationally and transnationally organized firms. Drawing upon such theoretical arguments, much of Clark's work focused upon examining how the National Labor Relations Board (NLRB) – established as part of the 1930s New Deal and tasked with being an umpire between employers and unions – has helped fashion the economic landscape through its decisions.

Following in these footsteps, others have examined how judicial decisions have shaped the geography of unionism, particularly in the context of multiple unions seeking to represent the same workers and jurisdictional disputes between unions over work. Thus, Andrew Herod detailed how rival dockers' unions along the US East Coast battled each other for members during the 1950s and how the NLRB's judgments determined the geography of union representation in the industry, as well as how unions representing dockers and lorry drivers ('teamsters') engaged in protracted legal battles to determine which was entitled to claim the work of packing and unpacking shipping containers after they began to be widely used to transport cargo. For his part, Christian Berndt has shown how the decentralization of union regulation in Germany in the 1990s has changed the traditional power geometry between capital and labor, in capital's favor.

Shifting focus to the international arena, Jane Wills has investigated how a number of GUFs have pressured transnational corporations to sign International Framework Agreements under which they agree to respect workers' rights and how European Union legislation concerning the establishment of works councils is impacting the geography of unionism. Meanwhile, others have examined how the manner in which governments regulate unions has shaped international flows of capital and how international codes of conduct secured by unions are affecting corporate investment strategies.

Unions and the Shaping of Economic Landscapes

A second area of research has focused upon how unions' struggles over work location and the right of workers to unionise have played active roles in molding economic landscapes. Herod, for instance, has recounted how a global corporate campaign by US steelworkers forced a multi billion dollar corporation to forgo plans to expand operations into Eastern Europe, together with how the AFL CIO worked hand in hand with US corporations and the government in the twentieth century to ensure that Latin American countries would open their markets to US produced goods, thereby actively shaping the geography of the global economy. Seeking to understand how unions are responding to the neoliberalization of the economy, Lydia Savage illustrated how unions in Boston played key roles in shaping the merger of two hospitals and how this allowed them to protect healthcare workers' conditions of employment and remuneration. Others have shown how, for instance, women unionists have challenged new labor regimes in the Australian steel industry and how labor unions in Toronto have quite literally constructed the urban built environment.

In related fashion, a number of geographers have highlighted how unions have indirectly played a role in shaping the geography of capitalism as firms have sought to avoid areas of militancy. Thus, Brian Page demonstrated how meatpackers relocated from Chicago to the Great Plains during the mid twentieth century, not to avoid unions *per se* but, rather, to seek out placid unions in rural areas – an avoidance of militancy which has had significant impacts upon that industry's geographical structure. Likewise, Don Mitchell has illuminated how farmers' efforts to restrict the spread of unionism in 1930s California played a central role in the development of the agricultural landscape. Equally, others have emphasized how the restructuring of the economic landscape shapes the possibilities for unionisation – Altha Cravey, for instance, has argued that the development of Mexico's border region is initiating a new geography of unionism in that country as militant unions emerge in the 'maquiladoras' beyond the traditional industrial regions long dominated by more conservative unions.

In sum, then, such research has sought to explore the dialectic between the social and the spatial and to show that the spatial context within which they find themselves both shapes, and is shaped by, workers' and unions' economic and political practices.

Space and the Political Geography of Labor Organizing

Given that unionization is a process of workers coming together over space, labor geographers have been particularly interested in how workers' spatial contexts shape their organizational strategies. Research in this vein has intersected significantly with debates concerning 'the politics of scale'. An early analysis in this regard was that by Roger Bivand, who looked at how support for the Polish union. Solidarity spread spatially from a number of local disputes to a national movement, a process through which it was constructed as a 'national' union (rather than as a series of loosely connected local branch unions). Likewise, Humphrey Southall illuminated how British workers' migration between towns during the nineteenth century provided a means whereby ideas about unionism could be transmitted across space. This was important, he contended, because the differential geographical portability of work skills helped shape what types of unionism emerged in different industries – because each mine's geology is often dissimilar, in the mining industry skills were not easily transferable and locally orientated union consciousness was encouraged, whereas in other trades requiring less geographically specific knowledges (e.g., engineering), a spatially broader consciousness developed. Meanwhile, Nick Blomley's analysis of the British coal miners' dispute of 1984/85 showed how the Thatcher government set about limiting striking miners' geographical mobility so as to 'localize' opposition to mine closures and prevent the exercise of worker solidarity across space.

With recent transformations in the nature of global capitalism reworking the spatial organization of many firms, a number of researchers have questioned whether the models of union organizing that appeared appropriate within a Fordist economy of highly vertically integrated corporations need to be reexamined. For instance, Savage has appraised how janitors in California have developed new organizing modes, whereas Lee Berman showed how clerical workers at Yale University developed spatially sensitive tactics to challenge the administration. In addition to assessing how unions are responding to the new spatial arrangements of work – both within and between workplaces – geographers have analyzed how unionists are building connections between the workplace and the broader community, especially in response to growing calls for unions to engage in 'social unionism'. Hence, Rebecca Johns and Leyla Vural's investigation of organizing in the US textile industry considered how unions have shifted their focus from pressuring employers directly in the workplace to engaging in public shaming actions designed to encourage patrons to stop buying particular companies' goods at retail outlets – moving the focus from the spaces of production to those of consumption, as it were. Others have explored unions' involvement in campaigns to force employers to pay a 'living wage' – applicable to both unionised and nonunion workers in particular communities – while some have suggested that the rise of new technologies like the internet are heralding new models of labor organizing. Thus, Peter Waterman has contended that the twenty first century is likely to be characterized by 'network activism' in which computer literate workers facilitate workplace to workplace contacts without relying upon their union's official organizing departments. Finally, some writers have articulated how it is crucial to place unions' goals when seeking to develop solidarity within a spatial context. Hence, do workers develop solidarity across space as part of a broader class response to capital or do they engage in such practices merely to help workers in other communities improve their lot, thereby making it less likely that investment from their own communities will relocate elsewhere – that is to say, are they merely seeking to protect their own privileged spaces within the global economy by helping others price themselves out of the market?

Place and the Nature of Unionism

The fourth area of interest has focused upon how the resources of 'place' shape practices of unionism. Primarily this has concerned itself with how local labor markets are regulated in place specific ways and with how workers come to identify with particular places. Arguably, Jamie Peck has provided the most influential treatment of how local labor markets are regulated spa tially, showing how the particularities of locality have impacted labor market restructuring and thus the possi bilities for unions to organize – analyzing, for instance, how Australian garment unions have tried to come to terms with the stretching out of the contracting chain and the rise of industrial homeworking in urban immi grant communities. Others have shown how women's domestic responsibilities constrain their mobility within the urban landscape, a factor which shapes their work choices and the likelihood they will be involved in union activities.

The influence of workers' 'topophilia' ('love of place') upon their political praxis has also been a central element in geographic analysis of unions. Hence, in their analysis of the British steel industry, Ray Hudson and David Sadler have argued that as mills closed during the 1980s workers began to shift from a class based politics to protect the industry as a whole to a place based politics designed to save 'their' particular mill. Similarly, Herod recounts how, in response to the threat of closure, unionists at a US glass manufacturing plant adopted a strategy which encouraged quiescence for fear that 'their' plant would be tainted as militant. The importance of considerations of place for shaping unionists' political praxis has also been revealed by Peter Sunley, who ex plored the remarkable parallel between the geographies of support in British coalfields for the 1926 General Strike and the 1984–85 dispute, suggesting that place specific traditions of industrial relations can have long standing impacts upon unions' activities. Similarly, in her analysis of British banking unionism Wills has high lighted how workers who are employed by the same firms, who belong to the same union, and who toil under the same nationally negotiated terms and conditions of work have adopted quite different practices of unionism, the result of the particular geo historical trajectories experienced by different parts of the country.

As these and other studies have articulated, then, workers' entrapment in particular locales and their feelings of loyalty to certain places can have significant impacts upon practices of unionism. Although such loy alty has often been decried as workers' adoption of 'false consciousness', the fact that for many workers their spatial embeddedness means they feel there is little al ternative illustrates that workers' practices and identities are geographically, as well as socially, constituted.

The Future?

During the past two decades unions in many parts of the industrialized world have experienced hard times, the result of deindustrialization, attacks by government, and a plethora of other factors. At the same time, the field of 'labor geography' has become a vibrant subdiscipline within geography. For the most part, the field's prac titioners have focused upon understanding how labor unions are structured spatially and how they shape the production of economic landscapes through their actions. There have been good reasons for this attention, not least because through understanding how the geography of capitalism is being transformed by processes such as globalization many geographers hope to help unions develop structures and strategies which will better enable them to defend successfully workers' interests. However, as the political environment within which workers seek to pursue their goals has been transformed, so are growing numbers of researchers also beginning to ex plore how workers and others are constructing organ izations to parallel unions (such as nonprofit entities designed to end child labor). As labor geography matures, it will undoubtedly expand its intellectual horizons to include research not just on unions and those workers fortunate enough to belong to unions, but also to the billions of workers worldwide who are denied access to unions and to the alternate organizations through which workers are seeking to satisfy their economic and polit ical aspirations.

Further Reading

Berman, L. L. (1998). In your face, in your space: Spatial strategies in organizing clerical workers at Yale. In Herod, A. (ed.) *Organizing the landscape: Geographical perspectives on labor unionism*, pp 203 224. Minneapolis: University of Minnesota Press.

Berndt, C. (2000). The rescaling of labour regulation in Germany: From national and regional corporatism to intrafirm welfare? *Environment and Planning A* 32(9), 1569 1592.

Bivand, R. (1983). Towards a geography of "Solidarność". *Environment and Planning D: Society and Space* 1, 397 404.

Blomley, N. (1994). *Law, space and the geographies of power*. New York: Guilford Press.

Clark, G. (1989). *Unions and communities under siege: American communities and the crisis of organized labor*. New York: Cambridge University Press.

Cravey, A. (1998). Cowboys and dinosaurs: Mexican labor unionism and the state. In Herod, A. (ed.) *Organizing the landscape: Geographical perspectives on labor unionism*, pp 75 98. Minneapolis: University of Minnesota Press.

Harvey, D. (1982). *The limits to capital*. Oxford: Basil Blackwell.

Herod, A. (1991). Local political practice in response to a manufacturing plant closure: How geography complicates class analysis. *Antipode* 23(4), 385 402.

Herod, A. (1997). From a geography of labor to a labor geography: Labor's spatial fix and the geography of capitalism. *Antipode* 29(1), 1 31.

Herod, A. (ed.) (1998). *Organizing the Landscape: Geographical Perspectives on Labor Unionism*. Minneapolis: University of Minnesta Press.

Herod, A. (2001). *Labor geographies: Workers and the landscapes of capitalism*. New York: Guilford Press.

Hudson, R. and Sadler, D. (1986). Contesting works closures in Western Europe's industrial regions: Defending place or betraying class? In Scott, A. & Storper, M. (eds.) *Production, work, territory*, pp 172 193. Winchester, MA: Allen and Unwin.

Johns, R. and Vural, L. (2000). Class, geography, and the consumerist turn: UNITE and the Stop Sweatshops campaigns. *Environment and Planning A* 32, 1193 1213.

Massey, D. (1984). *Spatial divisions of labour: Social structures and the geography of production*. London: Macmillan.

Mitchell, D. (1998). The scales of justice: Localist ideology, large scale production, and agricultural labor's geography of resistance in 1930s California. In Herod, A. (ed.) *Organizing the landscape: Geographical perspectives on labor unionism*, pp 159 194. Minneapolis: University of Minnesota Press.

Page, B. (1998). Rival unionism and the geography of the meatpacking industry. In Herod, A. (ed.) *Organizing the landscape: Geographical perspectives on labor unionism*, pp 263 296. Minneapolis: University of Minnesota Press.

Peck, J. (1996). *Work place: The social regulation of labor markets*. New York: Guilford.

Savage, L. (2004). Public sector unions shaping hospital privatization: The creation of the Boston Medical Center. *Environment and Planning A* 36, 547 568.

Savage, L. (2006). Justice for Janitors: Scales of organizing and representing workers. *Antipode* 38(3), 645 666.

Southall, H. (1988). Towards a geography of unionization: The spatial organization and distribution of early British trade unions. *Transactions of the Institute of British Geographers, New Series* 13, 466 483.

Sunley, P. (1990). Striking parallels: A comparison of the geographies of the 1926 and 1984 85 coalmining disputes. *Environment and Planning D: Society and Space* 8, 35 52.

Waterman, P. (1993). Internationalism is dead! Long live global solidarity? In Brecher, J., Brown Childs, J. & Cutler, J. (eds.) *Global visions: Beyond the new world order*, pp 257 261. Boston: South End Press.

Wills, J. (1996). Uneven reserves: Geographies of banking trade unionism. *Regional Studies* 30, 359 372.

Wills, J. (2002). Bargaining for the space to organise in the global economy: A review of the Accor IUF trade union rights agreement. *Review of International Political Economy* 9(4), 675 700.

Relevant Websites

http://www.aflcio.org
American Federation of Labor Congress of Industrial Organizations.

http://www.global unions.org
Global Union Federations.

http://www.icftu.org
International Confederation of Free Trade Unions.

http://www.icem.org
International Federation of Chemical, Energy and General Workers' Unions.

http://www.ilo.org
International Labour Organisation.

http://www.imfmetal.org
International Metalworkers' Federation.

http://www.ituc csi.org
International Trade Union Confederation.

http://www.itf.org.uk
International Transport Workers' Federation.

http://http://www.ictur.org
The International Centre for Trade Union Rights.

http://www.tuac.org
Trade Union Advisory Committee to the Organisation for Economic Cooperation and Development (OECD).

http://www.tuc.org.uk
Trades Union Congress (Britain).

http://www.wftucentral.org
World Federation of Trade Unions.

http://www.forumsocialmundial.org.br
World Social Forum.

Lacoste, Y.

F. Eva, Universita' Ca' Foscari, Venezia, Treviso Campus, Italy

Lacoste, Yves (1929–)

Yves Lacoste was born in 1929 in Fès, Morocco. He received his training as a geographer under Pierre George in a group that included, among others, Bernard Kayser, Raymond Guglielmo, and François Durand Dastès. Lacoste left the Communist Party in 1956 when his support of Algerian independence put him at odds with the party line. Taking up lecturing in 1962, he was a pioneer in the study of the dynamics of underdevelopment from the geographical perspective. He came to be recognized as one of the founders of a new democratic conception of geopolitics in Europe. Moreover, it is quite well known that in 1973 his 'geographical reading' of the planned bombing of North Vietnamese dams by the United States Air Force triggered international media coverage of its potentially devastating effects during the monsoon season. This resulted in their suspension.

Significantly, Lacoste was the founder and editor of the magazine *Hérodote*, devoted to socio political geography and geopolitics. Lacoste and the *Hérodote* staff took a critical position against so called traditional geography, which they regarded as the servant of the dominant classes. In a long interview appearing in the first issue of the magazine, Lacoste declared his interest in the ideas of Michel Foucault, who is attributed with having reopened the philosophical debate on 'space'. In the second issue, he was highly critical of 'a spatial' Marxist historicism. *Hérodote* also revived (in 1976, 1981, and 2005) the legacy of nineteenth century French geographer and anarchist Elisée Reclus, who Lacoste once defined (1984) as 'the best French geographer'.

In this respect, Lacoste's conception of a 'comprehensive' (physical and human) geography can be demonstrated to be in continuity with the radical French geography of the past. For example, also taking Reclus' ideas, Lacoste proposed an analysis of spatial contexts as always being unique, for which reason it is necessary to individually decode their representations, identify the players behind their dynamics, and analyze their different levels, gleaning from them the different relationships at play. Rather than the traditional concept of scale, Lacoste proposes theoretically relationships of causality and different levels.

In the 1970s, after the publication of his famous work *La Géographie ça sert d'abord à faire la guerre* (1976), Lacoste – partly through *Hérodote* – supported reintroducing the study of geopolitics after it had been discontinued in continental Europe after World War II because of the suspected connections of geopolitics with Haushofer's and the Nazi's *Geopolitik*. Lacoste became the point of reference of a new school of thought that aimed to spread a conception of geography as applicable to concrete dynamics and not limited to the academic and/or theoretical realms. In the field of geopolitics, Lacoste thus drew a distinction between 'external geopolitics' (as commonly associated with international dynamics) and 'internal geopolitics', that is, the power struggle between different players vying for the control of local territory through territorial planning, which is incorrectly and ideologically presented simply as a bureaucratic procedure.

Since 1984, Lacoste has said that geographers have a "social function," that is, "they envisage space in such a way as to manage it more efficiently." Their task, according to Lacoste, is to put their knowledge at the disposal of society "in the interests of the nation to which they belong." Meanings have since been attached to this reference to 'nation' that have been criticized by many 'critical' geographers (inside and outside *Hérodote*). His theoretical (and personal) shift toward institutional geography and the mass media was likewise the subject of criticism. As a geographer, Lacoste's most important contributions include the concepts of territoriality and representation (i.e., perceptions and the collective imagination). His criticism of 'spectacular geography' (from tourist brochures to the mass media) and the nationalistic ideological approach of school textbook geography foreshadowed later discussion of the issue. It is also thanks to his work that today European geographers regard geopolitics as not merely a by product of geography. In his opinion it is not good geography to consider geographical factors without taking into account their geopolitical context – it is important to consider mass dynamics and the balance of power between social, political, and economic players.

For Lacoste, geography is not a science but a dynamic complex of factors relating to power in the broad sense – and not just economic power. This view places him on a collision course with Marxist ideas and, in particular, at odds with Roger Brunet, the best known and foremost exponent in the 80s of French Marxist geography (which he regards as Marxism's final attempt to steer geography).

Vive la Nation, written in 1998, provoked heated debate since, although complex, his positive concept of nation was seen as self criticism in light of his earlier geo graphical vision. A similar thing occurred when Lacoste supported the usefulness of dams in controlling river flows when the anti globalization movement was sup porting the people's fight against them. Lacoste has also contributed to school textbooks and children's books, reflecting his belief that geography should be taught from an early age. He continues to write and his works remain much in demand by publishers. He is an emeritus pro fessor at University of Paris VIII, Saint Denis and teaches at *Institut Français de Geopolitique*. Unpredictably, Lacoste won the *Prix Vautrin Lud* (a very high level geographical award) at the International Geography Festival in 2000. During the same year he was awarded the title *Chevalier de la Légion d'honneur*. He has two sons, both geographers, but in different fields of geography.

See also: Francophone Geography; Geopolitics; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Political Geography; Reclus, E.; Regional Geography I; State; Territory and Territoriality.

Further Reading

Lacoste, Y. (1959). *Les pays sous developpes*. Paris: Puf.
Lacoste, Y. (1965). *Géographie du sous développement*. Paris: Puf.
Lacoste, Y. (1965). *Ibn Khaldoun Naissance de l'histoire du Tiers Monde*. Paris: Maspero.
Lacoste, Y. (1976). *La géographie ça sert d'abord a faire la guerre*. Paris: Maspero.
Lacoste, Y. (1980). *Unité et diversité du tiers monde*. Paris: La Découverte.
Lacoste, Y. (1985). *Contre les anti tiersmondistes et contre certains tiersmondistes*. Paris: La Découverte.
Lacoste, Y. (1986). *Géopolitique des régions françaises*, vol 1: *la France septentrionale*, vol 2: *la façade atlantique*, vol 3: *la France du sud est*. Paris: Fayard.
Lacoste, Y. (1990). *Paysages politiques*. Paris: Livre de poche.
Lacoste, Y. (1991). *L'état du Maghreb*. Paris: La Découverte.
Lacoste, Y. (1993). *Dictionnaire de geopolitique*. Paris: Flammarion.
Lacoste, Y. (1994). *Dictionnaire geopolitique des etats*. Paris: Flammarion.
Lacoste, Y. (1996). *La légende de la terre*. Paris: Flammarion.
Lacoste, Y. (1997). *Vive la nation Destin d'une idée géopolitique*. Paris: Fayard.
Lacoste, Y. (2000). *La légende de la terre*. Paris: Flammarion.
Lacoste, Y. (2002). *L'eau des hommes*. Paris: Cercle d'art.
Lacoste, Y. (2003). *De la géopolitique aux paysages*. Paris: Armand Colin.
Lacoste, Y. (2004). *Maghreb, peuple et civilisation* (avec Camille Lacoste Dujardin). Paris: La Decouverte.
Lacoste, Y. (2006). *L'eau dans le monde: les batailles pour la vie*. Paris: Collection Petite Encyclopedie Larousse.
Lacoste, Y. (2006). *Geopolitique, la longue histoire d'aujourd'hui*. Paris: Larousse.
Lacoste, Y. (2006). *Geopolitique de la mediterranee*. Paris: Armand Colin.
Lacoste, Y. and Giblin, B. (eds.) (1998). *Geohistoire de l'Europe mediane*. Paris: La Decouverte/Livres Herodote.

Relevant Websites

http://www.denistouret.fr/geopolitique/Lacoste.html
Conference in French language
http://php.bm lyon.fr/video conf/detail.php?id=62
Conference in French language
http://www.omegatv.tv/video environnement actualites environnement yves lacoste l impact des changements climatiques id601.html
In French.
http://www.canalc2.tv/video.asp?idEvenement=86
With others in French language

Lamarck(ian)ism

H. Winlow, Bath Spa University, Bath, UK

Glossary

Inheritance of Acquired Characteristics The idea that characteristics acquired over the lifetime of an individual, in relation to their surrounding environment, could be inherited by their offspring.

Linear Progressionism Idea that species evolve through a progressive hierarchical sequence to ever more complex forms. This idea has been applied to biological, social, moral, and intellectual evolution.

Mutual Aid An ecological theory that stressed the importance of cooperation between species in the struggle for existence.

Organism Analogy The comparison of groups in society to the development of an individual organism at a variety of scales, for example, a city, a state, or the world.

Recapitulation A biological and social evolutionary theory which suggested that the history of a species was reflected in the embryonic stages of development of each organism.

Definition

Lamarckianism is a non Darwinian theory of evolution ary change which originated with Jean Baptiste de Lamarck (1744–1829) in the early nineteenth century and influenced practitioners of science throughout the nineteenth century. In the 1890s, following the wide spread acceptance of the biological evolution of plant and animal species, a neo Lamarckian movement emerged in support of an alternative to Darwinian evolution. The two central elements of this movement, which repre sented only selected elements of Lamarck's original theories, included: a belief in the linear progression of different species and the concept of the inheritance of acquired characteristics. The alternative environmental evolutionary mechanism provided by neo Lamarckian ism was popular with scientists in the US and Europe until the 1920s, after which time modern genetic prin ciples were established and linked to natural selection, and Lamarckian theories of use inheritance were dis proved. In the late nineteenth and early twentieth cen turies, neo Lamarckian principles pervaded the work of biological and social theorists, including geographers. In many instances, an intermingling of both Darwinian and neo Lamarckian theories is evident. Lamarckian ideas were used to support a wide array of political standpoints, including legitimation of imperial expansion and racial hierarchy, and social and political reform.

Historical Development

Lamarck's research into evolutionary theory culminated in the publication of *Philosophie Zoologique* in 1809, an in depth account of his biological evolutionary theories. Two major elements of this work included: first, the theory of acquired characteristics, which incorporated the notion of adaptive evolution; and second, the belief that all organ isms progress to higher forms of organization. In relation to the latter point, Lamarck emphasized the idea of separate lines of descent for different biological species, thus incorporating the preexisting idea of a progressionist hierarchy of biological forms. In Lamarck's system, the stress on progressive evolution meant that species never became extinct but simply progressed to higher (and more complex) stages of development. Lamarckian pro gressionism influenced mid nineteenth century naturalists studying the fossil record. In Lamarckian theory fossils represented the remains of organisms which had been forced, by changing circumstances, to undergo biological modification.

Lamarck argued that progressive species evolution occurred partly as a result of the efforts of individual organisms through spontaneous generation in the early stages of evolution. Within this wider evolutionary se quence, geographic environment was responsible for species diversity, and environmental stimulus could in duce the use or disuse of organs. The Lamarckian system left no room for random chance or stasis, in contrast to natural selection in which environmental and geographic differences allowed for the emergence of unique species. Lamarck's system stressed continuous change and meant that even simple existing lifeforms could only be ex plained by a recent origin.

Lamarckian doctrines were applied not only to bio logical species development but also to social evolution. There were two political social reasons for the popu larity of Lamarckian doctrines and opposition to random variation. First, random variation was perceived as ex cluding the idea of progressive improvement in human heredity and, without this, a substantial argument in favor of social change was lost. Second, Darwin's ma terialist approach left little room for religious and philosophical questions about purpose and direction.

Lamarckian concepts were used in support of social progress in the pre Darwinian and the post Darwinian eras. In 1830s London, a range of evolutionary theories gained support in radical working class circles. Lamarckianism was popular because the inherent notion of progress could be used to support a model of social transformation from below. Radicals perceived society as progressing through cooperation, education, emancipation, and technological advance. Lamarck's emphasis on linear hierarchy, combined with the notion of acquired charac teristics, meant that individual effort could be portrayed as the driving force of evolution. This notion was also at tractive to liberals as it was a good basis for claims for a democratic society, as against the power of the aristocracy. The use of Lamarckian concepts in support of social change was later demonstrated by Peter Kropotkin in the early twentieth century, as discussed below.

Besides this radical support, Lamarckian ideas per vaded the work of many mainstream scientists through out the nineteenth century. For example, Stocking has observed that the concept of acquired characteristics was stated or implied in the work of many different writers, reflecting a widespread popular scientific attitude rooted in Western European culture. In 1880s Lamarckian concepts came under ongoing attack from neo Darwinist August Weismann, who had developed the idea of the germ plasm as an explanation for genetic inheritance. Herbert Spencer – whose alternative gemmule theory asserted that physical adaptation could occur within the individual organism after birth – defended the doctrine of acquired characteristics and unilineal social evolution. He launched a counterattack against Weismann in a series of articles in *Contemporary Review* in 1893. This controversy resulted in the development of a self con sciously neo Lamarckian movement in the 1890s. This movement involved the selective revival of Lamarckian doctrines, which were combined with the principle of natural selection to create a distinctly non Darwinian evolutionary framework. This movement had the stron gest impact in the US, but neo Lamarckian concepts also had considerable influence in France, Britain, and Ger many, and appealed to those who were unhappy with the materialist philosophy of Darwinism.

Key Concepts

Inheritance of Acquired Characteristics

Lamarck's overall impact on the development of evo lutionary science is hard to assess, as neo Lamarckians singled out only a small part of a more complex theory. The elements which scientists incorporated into their work in the late nineteenth century included that of the inheritance of acquired characteristics: the idea that characteristics acquired over the lifetime of an individual could be inherited by their offspring. Crucial here was the concept of use inheritance, with the classic example cited being that of the giraffe who, over generations of stretching to reach leaves, had acquired a long neck. Activities engaged in by adults within their environment would become natural to their offspring, thus producing different behavior patterns through a process of adaptive evolution. The idea of use inheritance could also be squared with doctrines of degeneration, where the atro phy or reduction of organs could be plausibly explained through the disuse of organs. Darwinism offered no ex planatory mechanism for these changes.

Linear Progressionism

In contrast to Darwin's branching tree metaphor of evo lution, Lamarck's writing incorporated the notion of an evolutionary ladder where species developed from inferior to superior forms. Natural selection incorporated the no tion of haphazard, and not progressionist, development, and therefore lacked the driving force that explained why life continually developed to more complex forms. The Lamarckian and Darwinian metaphors of development are compared in **Figure 1**. In *Philosophie Zoologique*, Lamarck had incorporated the Enlightenment ideas of 'The Great Chain of Being', but he presented a branched chain as opposed to a single line, realizing that it was not possible to represent the whole animal kingdom in one linear continuum. Within this system, Lamarck suggested transmutation of humans from apes (previously classified together by Linneaus). Concepts of branching evolution ism were suppressed by the progressionist viewpoint in late nineteenth century race science and influenced the ories of social and mental evolution, not least because there was ideological pressure to justify the expansion of the white races. This idea of linear progressionism is clear in Haeckel's comparison of the skulls of modern humans with older human forms (**Figure 2**).

Biological and Social Evolution

Evolution of Mind

As noted, Lamarckian concepts were applied to both biological and social theory. Two roles for the application of neo Lamarckianism to social evolutionary theory included: first, explaining the origin of human racial differences; and second, the provision of a link between social, intellectual, and organic evolution. Since the combination of progressionism and use inheritance meant that individual effort could influence evolution, the appearance of the human level of mental activity could be seen as an outcome of the laws of nature – just as the development of industrial society is the outcome of social laws. These elements were key to the work of British philosopher and social theorist Herbert Spencer (1820–1903).

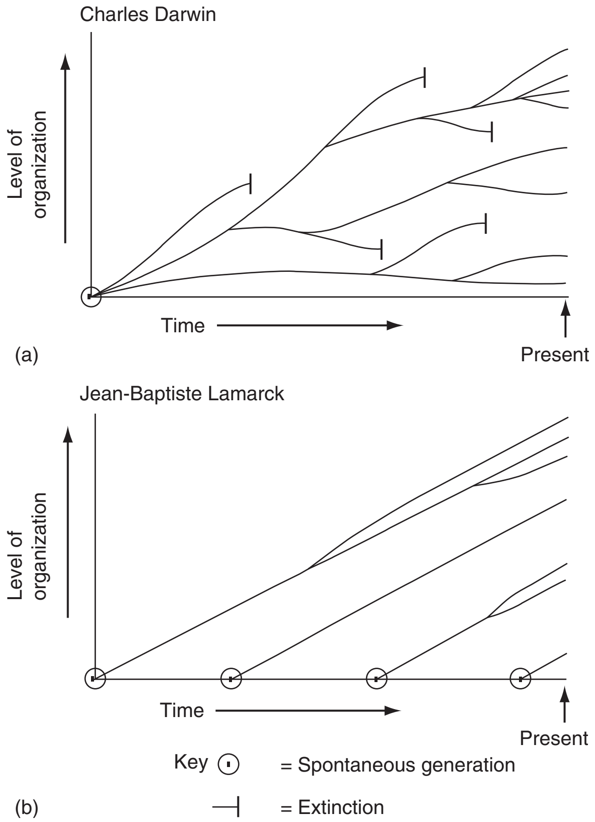

Figure 1 Comparison of Lamarck's and Darwin's explanations of evolutionary progress. Lamarck's model stresses progressive evolution and allows for spontaneous evolution in the present, but does not allow for species extinction. Darwin emphasized random selection and includes species extinction, and, in some cases allowed for reduction in complexity over time (indicated by a downward trend in some branches). Source: Humphreys, J. (1996). Lamarck and the general theory of evolution. *Journal of Biological Education* 30(4), 295 303.

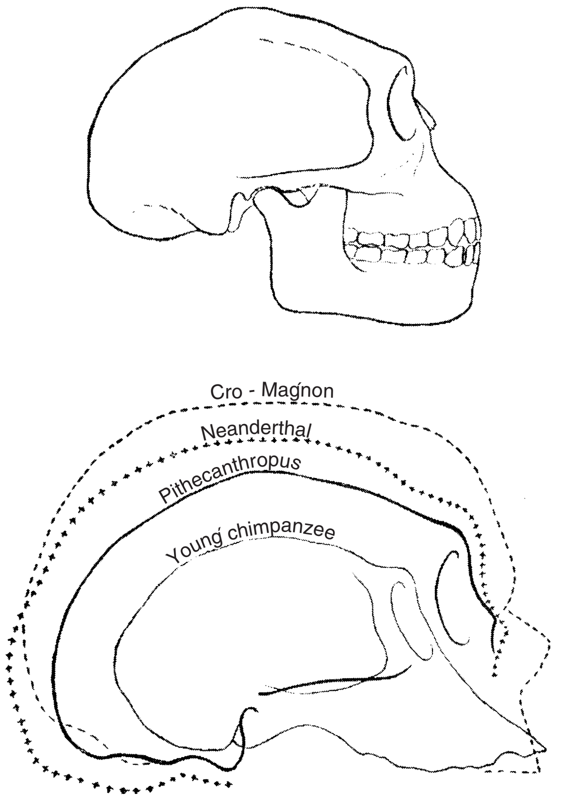

Figure 2 Skull of *Pithecanthropus erectus*, with skulls of a Cro-Magnon, a Neanderthal and chimpanzee superimposed. These are compared with a modern human skull and emphasize the idea of linear progression of ancestral forms. Source: Haeckel, E. (1898). Reproduced in Bowler, P. J. (1989). *The Invention of Progress: The Victorians and the Past*. Oxford: Blackwell.

In Spencer's 1870–72 text *Principles of Psychology*, he argued that national characters had developed as a result of mental peculiarities caused by habit, resulting in different races. He argued that the inheritance of acquired characteristics, as well as explaining racial evolution and instinct, could be applied to the evolution of higher mental faculties. These faculties were associated with more advanced social organisms. In *Principles of Sociology*, published between 1876 and 1896, Spencer argued that only as societies grow and become organized can the powers of thought develop, and that primitive human groups could not evolve these higher faculties in the absence of a fit environment.

Concepts of mental development were also addressed within literature. In an analysis of James Joyce's 1922 novel *Ulysses*, Rickard has argued that conflicting theories of mind were reflected within the text. In particular, elements of innate ideas or racial memory in *Ulysses* reflected the influence of a number of intellectual paradigms, including Freudian, Jungian, and Darwinian theories and theological and Lamarckian ideas. Among Joyce's influences were: Henri Bergson, who developed the notion of *élan vital* which he regarded as a force of creative evolution, and which incorporated the Lamarckian notion that all forms of life move toward greater complexity; Samuel Butler who argued that fetal development and other unconscious processes were the result of an unconscious or racial memory; Carl Gustav Jung, who argued that the unconscious included the personal layer, dating to infancy, and the collective layer, associated with the pre infantile period and relating to innate ancestral knowledge; and George Bernard Shaw, who incorporated elements of Bergson's *élan vital* in his plays, and developed a similar theory of life force. In *Ulysses*, memory plays a central role and the characters have a sense of involuntary motion toward preordained goals, reflecting some wider unconscious purpose. Explicit references are also made to evolutionary theory: for example, the 'Oxen of the Sun' episode is based on the idea of several parallel courses of evolution, including individual, linguistic, and wider biological evolution. This incorporates the idea of recapitulation, and reflects the idea that ontogeny recapitulates multiple phylogenies, concepts explained further below.

Recapitulation

A mixture of Darwinian and Lamarckian elements are evident in the work of Ernst Haeckel. Famous for his recapitulation theory of development, Haeckel published several evolutionary texts, including *Generelle Morphologie* in 1866 and *History of Creation* in 1876. His recapitulation theory included the idealist concept of life unfolding through a progressive sequence, and he coined the term phylogeny to describe hypothetical lines of descent. This theory was summarized by the phrase 'ontogeny re capitulates phylogeny': the idea that the individual re capitulated the growth of the history of the species, passing through stages representing adult ancestral forms – a popular, but erroneous, image.

Haeckel's social embryology was an example of biological determinism that reinforced notions of racial hierarchy, allowing comparisons between adults of supposedly inferior groups with children of superior groups. While incorporating notions of Lamarckian progressionism, Haeckel divided humanity into two major branches at an early stage after ascent from apes (as seen in his evolutionary tree in **Figure 3**). Haeckel ar gued that each branch then subdivided into separate races which (reflecting the influence of physical anthro pometry) could be aligned with hair type: the woolly haired people had collectively retained a lower level of evolution than the straight haired.

Geography and Neo-Lamarckianism

Neo Lamarckian biosocial theories had a wide ranging impact on geography as an emerging academic discip line. Elements of neo Lamarckian discourse influenced geographers alongside other aspects of evolutionary theory, including Darwin's natural selection and Men delian genetics. The influence of neo Lamarckian doc trines varied between geographical locations, and was influenced by wider political concerns. Discussed below are: first, the impact of Lamarckian doctrines on the work of US geographers; second, the development of mutual aid theories in Russia in relation to the work of Kro potkin; third, the influence of Lamarck's doctrines on the regional geography of Vidal de la Blache; and finally, the association between neo Lamarckian doctrines and racial theory in the work of Herbert John Fleure in Britain, and Ellsworth Huntington in the US.

Influences in US Geography

Early modern geography in the US reflected the influence of a powerful neo Lamarckian school of biology, which influenced figures in the public realm, including polit icians and academics. In the political sphere, neo Lamarckian theories were played out in Frederick Jackson Turner's 1893 frontier thesis. Turner reinterpreted the metaphor of the social organism to account for the unique qualities of the American nation as social and political responses to physiographical features. The frontier thesis incorporated the idea of a melting pot of biological and cultural types into a new American type, ideas central to the formation of American national identity and dem ocracy as separate from European heritage.

Although it was possible to view neo Lamarckian theory from a possibilist perspective (in relation to life force and consciousness), US geography took a de terministic stance, drawn from a focus on the inheritance of acquired characteristics. This is evidenced in the ap proaches of William Morris Davis (1850–1934), Ellen Churchill Semple (1863–1932), and Ellsworth Hunting ton (1876–1946).

Davis is famous for his development of the cycle of erosion (first presented in 1889) for physical landscapes, where he interpreted the evolutionary metaphor in in organic terms. This geographical cycle began with an uplift of the land, which would gradually wear back to a peneplain, after going through a number of sequential forms. The theory influenced several generations of geomorphologists because of its simplicity and elegance. Livingstone has noted that Davis spoke of characteristics being persistently inherited wherever the environmental conditions had remained constant, reflecting the theory of acquired characteristics. Davis had studied at Harvard under Nathanial Southgate Shaler, who was a convinced neo Lamarckian. Shaler had worked under Louis Agassiz at Harvard in a transitional period in the development of biological and geological theory. His acceptance of neo Lamarckian doctrines allowed him to combine the tele ological views of Agassiz with more recent evolutionary insights.

Semple's 1911 text *Principles of the Geographic En vironment* preserved the neo Lamarckian element of Ratzel's *Anthropogeographie*, which had stressed that the characteristics of social groups had occurred as local responses to environment. Semple argued that race had developed over time as a result of the different en vironments which had left their mark on the present race in the form of inherited aptitudes and customs. Like other neo Lamarkians, Semple was influenced by 'environmental determinism' in her conviction that geographical isolation was indispensable to organic spe ciation and advance. She argued that the environment most conducive to evolutionary advance was the land mass of Eurasia. These deterministic ideas of racial evolution were strongly reflected in Ellsworth Hunting ton's 1915 book *Climate and Evolution*.

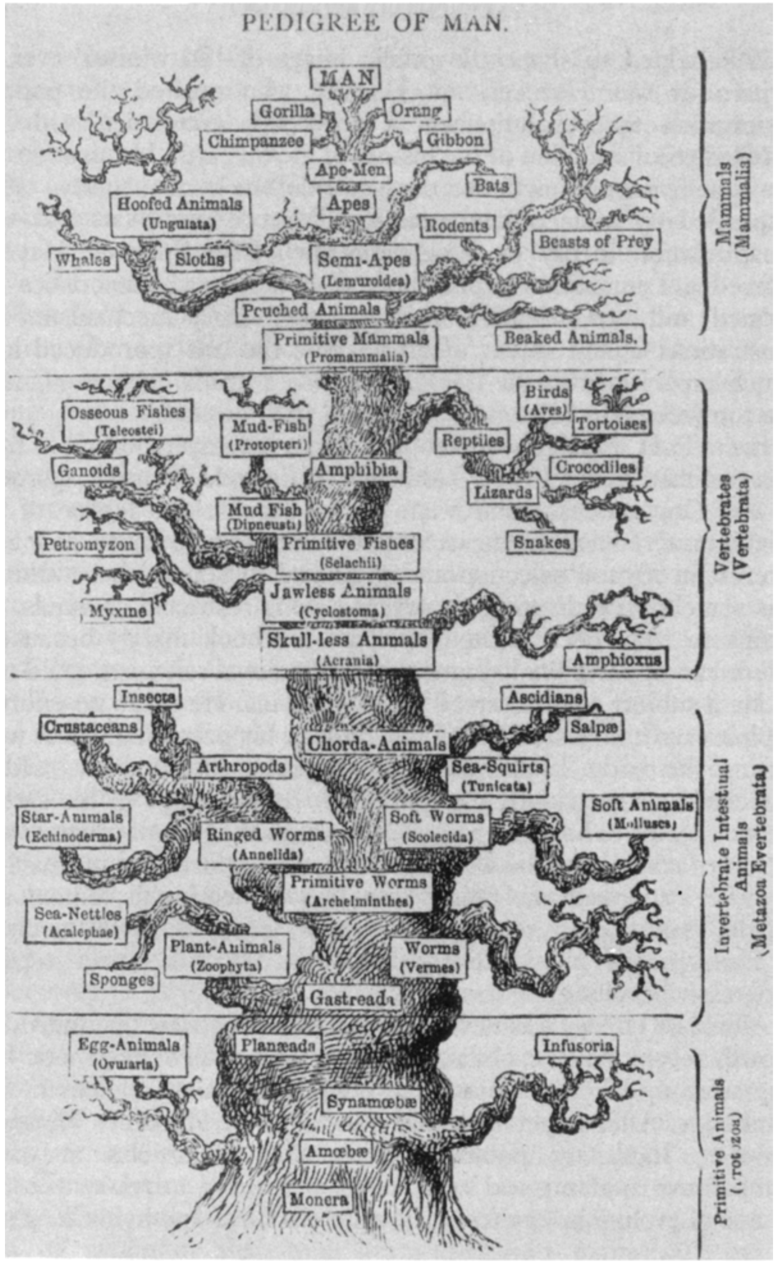

Figure 3 Haeckel's evolutionary tree. Unlike Darwin's tree of life, this represents evolution as an inherently progressive process. Haeckel's tree includes a main trunk running up to humanity at the top. Smaller divisions of this trunk represent the evolution of racial groups. Source: Haeckel, E. (1876). *History of Creation.* Reproduced in Bowler, P. J. (1989). *The Invention of Progress: The Victorians and the Past*. Oxford: Blackwell.

Kropotkin and the Mutual Aid Tradition

Darwin's theory of natural selection and his stress on Malthusian population controls met with resistance in Russia from the outset. Nineteenth century Russian academics and scientists, while divergent in political outlook, had a common response to Malthusianism and either de emphasized or denied its role. This national style resulted from, first, class structure and political traditions which shared a cooperative social ethos and, second the geographic environment. Research under taken by naturalists in tropical environments, which emphasized overpopulation and competition, did not seem applicable to the northern latitudes of Russia. Russian geographer and anarchist Peter Kropotkin (1842–1921) first questioned the struggle for existence while exploring Siberia, to where he accompanied a series of scientific and military expeditions between 1862 and 1867. At the beginning of this period, he was already a celebrated naturalist, and on these expeditions he developed an aptitude as a geologist and geographer. Kropotkin noted of his explorations in Siberia (with his companion Polakoff, a zoologist):

> We both were under the fresh impression of the *Origin of Species*, but we vainly looked for the keen competition between animals of the same species which the reading of Darwin's work had prepared us to expect (Kropotkin, 2006: 7).

Kropotkin's observations of the natural environments of Siberia and Northern Manchuria convinced him that underpopulation and paucity of life were distinctive throughout northern Asia, and that mutual aid was essential to the preservation and evolution of each species. Todes has observed that in the same way that the idea of nature as a jungle had influenced Darwin in the writing of the *Origin*, the idea of nature as a Siberian expanse underpinned Kropotkin's essays on evolution.

For Russians, the seminal mutual aid theorist was Karl Kessler (1815–81), a naturalist, whose 1879 speech 'On the law of mutual aid' transformed widespread sentiment into a coherent intellectual tradition. Kropotkin read this speech in 1882 or 1883 and this motivated him to find further evidence of mutual support in nature. While his speech created a lot of enthusiasm, Kessler died 2 years later without having elaborated his ideas. In his speech, Kessler had argued that mutual aid resulted from the need for organisms to reproduce. He outlined examples of cooperation (and sociability) among many species, including bees, ants, beetles, spiders, fish, reptiles, birds, and mammals, and he viewed mutual aid as central to humanity's material and moral progress. The mutual aid tradition put in place by Kessler offered only an ecological conception (a description of natural relations) and not an evolutionary theory. The concept was useful for describing anthills but not for the appearance of physical characteristics, such as stripes on a zebra (which could be explained through natural selection). When mutual aid theorists joined their ideas to evolutionary theory, they usually emphasized mechanisms other than natural selection, such as direct action of the environment, isolation, and the inheritance of acquired characteristics.

Kropotkin made regular contributions to the British scientific journal *The Nineteenth Century*. His first article on mutual aid was published here in 1890, with others following over the next 4 years. It was Huxley's 1888 Malthusian article on 'The struggle for existence in human society' which prompted Kropotkin into writing his 1902 treatise *Mutual Aid: A Factor of Evolution* in which he provided evidence of mutual aid among animal and human societies. The book advanced a biosocial law based on an analysis of the struggle for existence, and stressed the historical tendency toward cooperation in nature. *Mutual Aid* incorporated a progressionist neo-Lamarckian stance, not least evidenced in the structure of the narrative which began with chapters on mutual aid in animals, before discussing the evolution of human society through savage, barbarian, medieval, and finally modern industrial forms. While Kropotkin argued against the dominance of modern state apparatus in society, he stressed that human morality, expressed largely through mutual aid, had developed progressively over time, with its roots in prehuman societies.

Central to Kropotkin's mutual aid thesis was his concept of sociability. Kropotkin argued that the species at the top of the animal world and most similar to humans in structure and intelligence (such as monkeys and apes) were the most sociable. The concept of sociability was seen as natural and was regarded as an acquired characteristic. *Mutual Aid* focused largely on societal development, with little focus on biological evolution. In later essays Kropotkin attempted to translate mutual aid from an ecological conception into an evolutionary theory and published a number of articles in *The Nineteenth Century* between 1905 and 1919. The central arguments of Kropotkin's alternative Darwinism were: species practice mutual aid during the struggle for existence; individualistic animals decrease in numbers, while those which cooperate increase; cooperative species are more subject to the direct action of environment and variations are passed to the next generation through the inheritance of acquired characteristics. Throughout his writings Kropotkin praised Darwin's theories and drew attention to the fact that Darwin had acknowledged the importance of cooperation in *The Origin*. Kropotkin argued against the strict division of scientists into Darwinian and Lamarckian camps and against the distortion of the principles of their namesakes. Kropotkin's evolutionary theory included a synthesis of French theories and Darwin's ideas on the struggle for existence, while rejecting the Malthusian element. In a similar vein to proponents of Social Darwinism (and imperial expansion) a naturalization of morality, and a biologization of political categories, occurred throughout Kropotkin's work. However, his political convictions led him to use these ideas to argue for freedom and not subjection of human society.

Vidal's Regional Geography

The merger of Lamarckian ideas with the doctrine of possibilism is illustrated in the work of French regional geographer Paul Vidal de la Blache in works, such as his 1903 *Tableau de la Geographie de la France*. Central to Vidal's work was the concept of the social organism. In regional geography, the idea of organic unity served as a unifying theme, which was also dominant in the work of Andrew J. Herbertson and Herbert John Fleure in Britain.

Vidal had a naturalistic approach to geography and believed that the different parts of nature were mutually interdependent and part of a whole, and he regarded the Earth as an organism. Vidal split these interrelationships into two principal cycles, that of the inanimate world (geological and climatic factors) and that of the animate world (relationships between living plant and animal species). For human societies, the social organism included social, political, and industrial factors. To begin with, social relationships between humans had interdependencies with nature – as illustrated by Vidal's localized *genres de vie* – but over time the social organism would begin to grow in areal extent (similar to Ratzel's *Lebensraum*). Transport and communication links would develop and eventually cities would become established. Vidal argued that humans could eventually control the environment – natural features of geography thus influenced, but did not determine, human activities. On a wider scale, these principles could account for the evolution of regional and state organisms. Vidal viewed European imperialism and colonialism in a positive light, as examples of strong organisms reaching greater levels of advancement.

Vidal's theories reflected three Lamarckian elements: the self organization of animate objects; the importance of circulation for organic organization and development; and the general processes of evolution. Archer compares Lamarck's emphasis on subtle fluids which unified parts of a biological organism with Vidal's stress on circulation in unifying human individuals into greater social wholes. Similarly, while Lamarck argued that the internal circulation of fluids was initiated by the organisms' immediate environment and that organisms would eventually stimulate their own evolution, Vidal argued that direct environmental stimuli were most important for the emergence of rudimentary social organisms, after which accelerated circulation between individuals and groups became more important. Vidal's geography stressed the evolutionary process of organic organization, circulation, and evolution, an organicist approach which formed the basis of his general geography.

Environment, Race, and Heredity

H. J. Fleure (1877–1969) is best known for his studies in regional geography and his anthropogeographical work on Welsh racial types. Fleure's undergraduate study and teaching within the natural sciences convinced him of the importance of Darwinian scientism, but his geographical perspective emerged from the interface between Darwinian and Lamarckian elements. The influence of neo Lamarckism on Fleure's work was illustrated as early as 1899, when, in a talk he rejected Weismann's germ plasm theory, arguing instead for Spencer's gemmule theory. Two years later in a talk on 'The factors of organic evolution' Fleure stressed the importance of acquired characteristics, arguing that experiments had revealed the susceptibility of the germ to environmental influence.

Fleure's emphasis on 'possibilism' fitted closely with the Vidalian scheme and Vidal's stress on the relationship between *genres de vie* and place was reflected in Fleure's anthropometric studies in Wales. However, Fleure described his areas of study as human regions, stressing the importance of the social sphere, while Vidal (and Herbertson) stressed the existence of natural regions. In contrast to Vidal, Fleure perceived that overindustrialization lead to a decline of physical and moral race character and the decline of cultural traditions (which he felt were central to the Welsh national character). He applied ideas of biological degeneration to cultural decline which reveal the influence of the Lamarckian biological theory of disuse.

The admixture of Darwinism and neo Lamarckian concepts within Fleure's work was further complicated by the development of genetic theories, such as Mendelism. The influence of hereditarian elements was a feature in his ethnographic studies of rural Wales. The interplay between environment and heredity in defining local race character is clear in his argument that "inherited characters may persist for untold periods" and "many alternative inherited characters may be found side by side in one and the same population, some characters giving clues to immigration of the most varying periods, some probably suggesting to us features of inhabitants of the region in very distant times" (Fleure, 1922: 95). In a number of places Fleure argued that there was evidence for the existence of Neolithic survivals in the moorland areas of rural Wales. Thus, Fleure's approach revealed a synthesis between environmentalism and Mendelian evolutionary ideas.

Like Fleure, Huntington perceived himself as a Darwinist, as evidenced by the introduction to his 1907 book *Pulse of Asia*, where he claimed that the biological sciences stood in utter confusion before Darwin. Although Huntington did not acknowledge it, neo Lamarckian elements were present within his writing. *Pulse of Asia* begins on the premise that the environment had determined the development of moral and mental traits in primitive societies, and that these characteristics were inherited by future generations.

Huntington's deterministic stance is clear in *Civilisation and Climate*, published in 1915, where he argued that the most advanced civilizations had developed in the temperate regions of the globe. However, Huntington's later work includes greater reference to heredity. In Huntington's 1924 text *The Character of Races*, he asserted that races should be defined on the basis of inherited characteristics, although he still accepted the importance of the social and physical environment. In *World Power and Evolution*, Huntington argued:

> For good or ill, each race has acquired certain charac teristics. These may be modified…but they can never be wholly destroyed.…The Chinese, the Negroes, the Anglo Saxons, the Jews, and the Italians cannot be made alike either by the influence of the environment or by education and training. This we must recognise. Mankind is bound to change in the future, but it would be the height of folly for a nation on this account to incorporate into itself elements whose mental and moral aptitudes it does not now approve (Huntington, 1919: 185).

Incorporating neo Lamarckianism and more recent genetic theories Huntington argued that only certain races could, through acquired characteristics, adapt to new climatic conditions, a claim that could be used to restrict US immigration on a racial basis. These argu ments illustrate both a biased attitude toward specific racial groups and a weak understanding of concepts of heredity. This extract from Huntington illustrates how human geography in the early part of the century became a meeting point for evolutionary, environmental, hered itarian, and nationalistic thought, in the quest to interpret the relationship between nature and human society.

Conclusions

These discussions have illustrated the widespread influ ence of neo Lamarckian concepts across national boundaries and in support of a range of biological, pol itical, social, and racial theories in the late nineteenth and early twentieth centuries – influences that were reflected in the work of the first generation of academic geog raphers. Campbell and Livingstone have noted that the eventual achievement of the neo Darwinian synthesis mostly removed the last vestiges of scientific support for neo Lamarckianism, but that this was not eliminated entirely.

Gould and Oldroyd have argued that the area of human culture could still reasonably be viewed as La marckian, because every time a poem is written, or a new tool designed, or a new theory developed, a new char acteristic is acquired, which is transmitted to subsequent generations. Thus, while Lamarckian theory failed to provide a valid explanation for biological evolution, the idea of acquired characteristics can still be considered an applicable metaphor for explaining social and cultural development.

See also: Anthropogeography (After Ratzel); Darwinism (and Social Darwinism); Kropotkin, P.; Mapping, Race and Ethnicity; Possibilism; Vidal de la Blache, P.

Further Reading

Archer, K. (1993). Regions as social organisms: The Lamarckian characteristics of Vidal de la Blache's regional geography. *Annals of the Association of American Geographers* 83(3), 298 514.

Bowler, P. J. (1986). *Theories of Human Evolution*. Oxford: Blackwell.

Bowler, P. J. (1989). *The Invention of Progress: The Victorians and the Past*. Oxford: Blackwell.

Campbell, J. A. and Livingstone, D. N. (1983). Neo Lamarkism and the development of geography in the United States and Great Britain. *Transactions of the Institute of British Geographers, N.S.* 8, 267 294.

Desmond, A. (1989). *The Politics of Evolution; Morphology, Medicine, and Reform in Radical London*. Chicago, IL: University of Chicago Press.

Fleure, H. J. (1922). Some aspects of race study. *Eugenics Review* 14, 93 102.

Gould, S. J. (1979). Another look at Lamarck. *New Scientist* 84, 4 October, 38 40.

Humphreys, J. (1996). Lamarck and the general theory of evolution. *Journal of Biological Education* 30(4), 295 303.

Huntington, E. (1919). *World Power and Evolution*. New Haven, CT: Yale University Press.

Jones, G. (1980). *Social Darwinism and English Thought: The Interaction between Biological and Social Theory*. Brighton: Harvester Press.

Kropotkin, P. (2006). *Mutual Aid: A Factor in Evolution* (reprint of 1902 edn., with preface from 1914 edn.). New York: Dover Books.

Livingstone, D. N. (1984). Natural theology and neo Lamarckism: The changing context of nineteenth century geography in the United States and Great Britain. *Annals of the Association of American Geographers* 74(1), 9 28.

Livingstone, D. N. (1992). *The Geographical Tradition*. London: Blackwell.

Oldroyd, D. R. (1983). *Darwinian Impacts: An Introduction to the Darwinian Revolution*. Milton Keynes: The Open University Press.

Rickard, J. S. (1999). *Joyce's Book of Memory: The Mnemotechnic of Ulysses*. Durham, NC: Duke University Press.

Stocking, G. W., Jr. (1962). Lamarckianism in American social science: 1890 1915. *Journal of the History of Ideas* 23, 239 256.

Todes, D. N. (1980). *Darwin without Malthus: The Struggle for Existence in Russian Evolutionary Thought*. Oxford: Oxford University Press.

Winlow, H. (1999). *Cartographic Representations of Race*: *c. 1850 1930*. Unpublished PhD Thesis, Queens, University Belfast, Belfast.

Winlow, H. (2001). Anthropometric cartography: Constructing Scottish racial identity in the early twentieth century. *Journal of Historical Geography* 27(4), 507 528.

Winlow, H. (2006). Mapping moral geographies: W. Z. Ripley's races of Europe and the US. *Annals of the Association of American Geographers* 96(1), 119 141.

Relevant Websites

http://www.lamarck.cnrs.fr
Oeuvres et rayonnement de Jean Baptiste Lamarck (Jean Baptiste Lamarck: Works and heritage, includes reproductions of publications, and biographical information).

http://evolution.berkeley.edu
Understanding Evolution (includes section on History of Evolutionary Theory).

Land Change Science

B. L. Turner II, Arizona State University, Tempe, AZ, USA

Glossary

Biosphere It is the intersection zone of the Earth's land, ocean, and atmosphere systems that support life.
Coupled Human–Environment System It is a bounded unit comprising a set of human and environmental relationships.
Ecosystem Services These are resources and processes of ecosystems that benefit humankind.
Spatial Econometrics It is the incorporation of spatial effects in quantitative and statistical methods addressing economic principles.
Sustainable Land Architecture The pattern of land uses and covers that provision human uses and maintains a full range of ecosystem services.
Sustainability It is provisioning humankind without threatening the functioning of the biosphere.
Vulnerability It is the degree to which a system is likely to experience harm due to the exposure to a hazard.

Land change science (LCS) is a transdisciplinary field of study that seeks to observe and monitor land cover and land use changes and explain and model these changes as a coupled human–environment (or social–ecological) system. Land cover refers to the biophysical condition (e.g., forest, grassland, and paved surface) of some portion of the terrestrial surface of the Earth, and land use is the human intent or activity associated with that condition (e.g., reserve, pasture, and settlement). In addition to detecting and monitoring changes in land cover, largely from satellite information, LCS examines societal struc tures and individual behavior that determine land uses, the impacts of those uses on the structure and function of the biophysical subsystem, and the environment feed backs, foremost ecosystem (environmental) services, from that subsystem to land uses and the human subsystem. Understanding the coupled system through the medium of land cover and use not only provides insights about outcomes and processes in both subsystems, but also informs synthesis questions about the vulnerability and sustainability of human–environment relationships.

Development of LCS

The question of global warming stimulated various sci ence agendas to improve understanding of the Earth system and the human impacts on it. The resulting research efforts quickly realized that understanding the processes linking the terrestrial surface of the Earth, often addressed in terms of ecosystems, to the Earth system at large was incomplete without understanding the human induced changes on that surface (i.e., as they affect, for example, albedo and sources and sinks of carbon, or methane emissions), including the rates and trajectory of those changes. This need, in turn, required improved understanding of the human drivers (e.g., population or policy change) of land use. This use, however, is also affected by changes in the biophysical subsystem, either directly through land cover changes in the ecosystem services or indirectly through land cover impacts on the Earth system at large, such as global climate change. The importance of these dynamics in creased as concerns about global environmental changes enlarged from climate change to loss of biodiversity and ecosystems, land degradation, and ultimately, sustain ability. Land changes are critical to a range of issues that culminate in the maintenance of the biosphere (that portion of the Earth system sustaining life) and human well being.

Land Change in Global Environmental Change

Given that humankind evolved as a land based species, it is not surprising that human impacts on the terrestrial surface of the Earth are ancient, traced back to the control of fire and perhaps the extinction of Late Pleis tocene and early Holocene megafauna. With each major technological epoch (stone tools fire, domestication, fossil fuels, etc.), the number, magnitude, spatial reach, and pace of land changes wrought by humankind enlarged. Estimates of human induced transformation (radical change) in land cover range as high as 50% of land surface of the Earth. If coadapted landscapes and indirect impacts through climate change are considered, virtually the entire land surface has been affected by human ac tivity. Much of this impact has changed the states of the Earth system, as in the 40% of the land surface in which forest, grasslands, and wetlands have been converted to agriculture. These and other land changes are the sec ond largest source (to fossil fuel burning) of human in duced carbon dioxide, and the significant source of methane, to the atmosphere. Land practices, through the use of fertilizers, make humankind the largest contributor to the global nitrogen cycle, surpassing nature. In add ition, about 85% of annual water withdrawals globally are taken for agriculture, with important consequences for the water cycle.

Clearly, land change is the principal factor to date in changing the land states of the Earth system and, in doing so, has played an important role in changing the bio geochemical cycles that permit that system to function and sustain the biosphere. The magnitude and spatial reach of these changes raises serious questions about the sustainability of human uses of ecosystems, landscapes, and the Earth system.

Observation/Monitoring

Satellite and air based sensors observing the Earth have proliferated over the last quarter century, providing a wealth of new and novel information about the changes in the land surface of the Earth. Improved seamless coverage of global land covers (e.g., forest cover) and high spatiotemporal resolution data on specific locales have helped to document such attributes of the terrestrial surface as increases and losses in the area of temperate and tropical forests, respectively, changes in above ground carbon sources and sinks, losses of ecosystems and biomes of high biotic diversity, urban expansion into agricultural lands, and changing phenology (e.g., timing of the flowering of vegetation) and primary productivity (e.g., photosynthesis). Refocusing on regions and locales, major advances in the development of land cover clas sification provide increasing detail of the land surface, capable of differentiating among forest types and phases of their successional growth as well as detecting invasive species, selective logging, and the buildup of fuel loads in fire prone forests, among other examples. It is precisely such detail that permits tests of various theories and concepts emanating in the social sciences relevant to understanding the human causes and consequences of land change. For example, smallholders throughout the tropical world employ slash and burn cultivation prac tices where land use decisions are affected by the stage of vegetative regrowth in crop–fallow cycles. Distinguishing these stages via remote sensing facilitates tests of theories of agricultural change, a major component of improving understanding of land dynamics in that part of the world.

Drivers/Causes of Land Change

In global environmental change, the term driver denotes a forcing function on some part of the Earth system. Drivers can be both biophysical factors that change land cover (e.g., long term declines in rainfall) and human factors that change land use (e.g., population or policy change). Much social science work, however, treats dri vers as causes of land use and addresses them through theoretical and empirical assessments. The human drivers or causes vary by the scale of analysis employed, and at lower spatial scales, by the context of the case examined. Over the long term and at the global scale, the *PAT* variables of the *IPAT* identity [I (environmental impact) = P (population) x A (affluence or per capita consumption) x T (technology; efficiency of resource use and disposal)] are the only factors to track consistently with land change, perhaps because P and A serve as coarse indicators of the demand for ecosystem goods and services. Below the continental scale or for pan global land covers of specific types (e.g., tropical deforestation), the *PAT* variables need not capture demand, in part be cause economic globalization often disassociates the sources of demand from the extraction or production of resources to fulfill them. In such cases, the *PAT* factors may be operating, but they are not necessarily the most important for land change. The variance in drivers/ causes acting on specific types or cases of land cover change and the variance in land outcomes owing to the synergy of different drivers/causes may be so large that generalizations prove difficult. For specific areas of land change, however, a variety of research has demonstrated the roles of markets, policy, transportation and roads (infrastructure), nongovernmental market interventions, household life cycles, and resource governance, among other causes. Increasingly, attention is paid to nested or linked causes, hierarchically arrayed. For example, international policy affects national and local policies and programs, which, in turn, change resource governance and thus access to land or the income that can be gen erated from it, concomitantly changing land use and cover. Causes combined in explanations that consider both structural (i.e., sociopolitical organization) and be havioral (i.e., individual decision making) factors prove difficult to test, owing to data limitations, but, when possible, provide improved explanation of land changes.

Land–Cover Consequences

LCS pays special attention to ecosystem goods and ser vices (e.g., water filtration, pollination, and food and fiber production) in their roles as resources traded in the market place and thus having a classical economic value (i.e., provisioning services as in food and fiber pro duction) and as factors maintaining the functioning of the biosphere and nature's condition (e.g., regulating and supporting services, such as climate controls and nutrient cycling). The wide range of these goods and services (which also include cultural and preserving services as in, respectively, sacred places and maintenance of biotic diversity) has been examined at different scales of an alysis, demonstrating their importance to the mainten ance of the biosphere, from ecosystem to landscapes to the Earth system, and the consequences on these functions as the goods and services are drawn down by human activity and consumption. The interconnections between human driven changes in land cover and environmental dynamics have drawn considerable attention, demonstrating their implications for such far flung phenomenon and processes as the frequency and magnitude of forest fires, regional scale precipitation changes, and exposing human populations to new disease vectors.

Modeling

Global environmental change and sustainability science are found on experiments, tests, and the need to address the future through quantitative based forecasts. Modeling is essential in either case. Land change modeling has been served by advances in the geographical information sciences (GIScs) and its grounding in the use of and methods for spatially explicit data and products. For many, if not most of the questions addressed by LCS, knowing or forecasting the kind, magnitude, and pace of land change is insufficient; the location of the change is essential as well. The spatial arrangement of ecosystems and landscapes matters in regard to environmental functions, and hence the delivery of ecosystem goods and services that society expects. Likewise, the spatial arrangement of the human uses of the same ecosystems and landscapes determine the consequences for goods and services and their feedbacks on land use. Given the prevalent use of remotely sensed data (Observation/Monitoring above), the pixel in the imagery commonly marks the level of spatial explicitness that the models may take (e.g., about $30\,m^2$, the most prevalent used imagery in LCS). A large range of models are employed, especially as they have emerged from spatial econometrics and spatial analysis. Major advancements have been made in agent based, integrative assessment models – those that attempt to model agents' (individuals') land use decisions in the face of knowledge about changes in environment and institutions (rules of governance). The performance of these models in explaining and projecting the magnitude and location of different land types of change continues to improve, although the uncertainty in projections beyond a decade enlarges substantially.

Synthesis

LCS attempts to integrate the knowledge gained from the various components of its study (**Figure 1**) to inform synthesis issues that typically carry both research and outreach application implications. Examples include various attempts to understand and identify areas vulnerable to or resilient in the face of environmental change (e.g., climate change) and 'natural' hazards (e.g., hurricanes) or sustainable in the face of increasing human uses. Drawing from risk hazard research and critiques of it, both of which focus on the vulnerability of people, and from ecological research focused on the resilience of ecosystem, LCS has developed an emphasis on the vulnerability–sustainability of the coupled human–environment system. This vision recognizes that hazards simultaneously affect both the human and environmental subsystems and that the synergy of the two subsystems determines the total system response to the hazard.

The coupled system approach to vulnerability are illustrated by the recent hazard of Hurricane Katrina, the 2005 tropical storm that devastated coastal Mississippi (state) and the city of New Orleans, Louisiana. The exposure and sensitivity of New Orleans and the wetland habitats between the city and Gulf of Mexico have been altered by hydraulic works designed to facilitate water borne traffic, permitting housing development in low lying terrain, and by constructing access canals to gas and oil pumps distributed across the Mississippi delta. The overall effect of canal and levee construction reduces the buildup of sediments that maintain wetlands in the delta, especially the cypress swamps. These losses reduce the barriers to storm surges surrounding New Orleans, and the canals serve to funnel storm waters into the city, stressing the levees intended to protect the metropolitan area. The vulnerability of New Orleans, therefore, was and remains the product of human–environment interactions, and Katrina's impact on both the city and ecosystems of the delta was profound.

LCS has addressed an array of other synthesis issues and questions. For example, increasing attention is paid to the efficacy of the boundaries of nature reserves and parks, especially in cases where significant land use surrounds or people have the right to cultivate or collect resources within the reserves. In some cases, land changes outside the park, even at considerable distances from the park boundaries, affect biotic diversity within the park, especially where migratory animals are involved, raising the need for and design of biological corridors to facilitate the movement of biota. In other cases, the delimitations of the reserve boundaries and the rules of land and resource uses within the reserve to serve both human and environmental concerns have proven beneficial to the success of the reserve. In another example, LCS addresses the political–economic conditions that facilitate a forest transition – or cases in which economic modernization leads to a return of forest cover on lands once taken from forest. Drawing from lessons in environment development, LCS has also entertained real world application, addressing the co production of knowledge and participatory programs for land management. The resulting approaches can be found in formulizations such as the Drylands Development Paradigm.

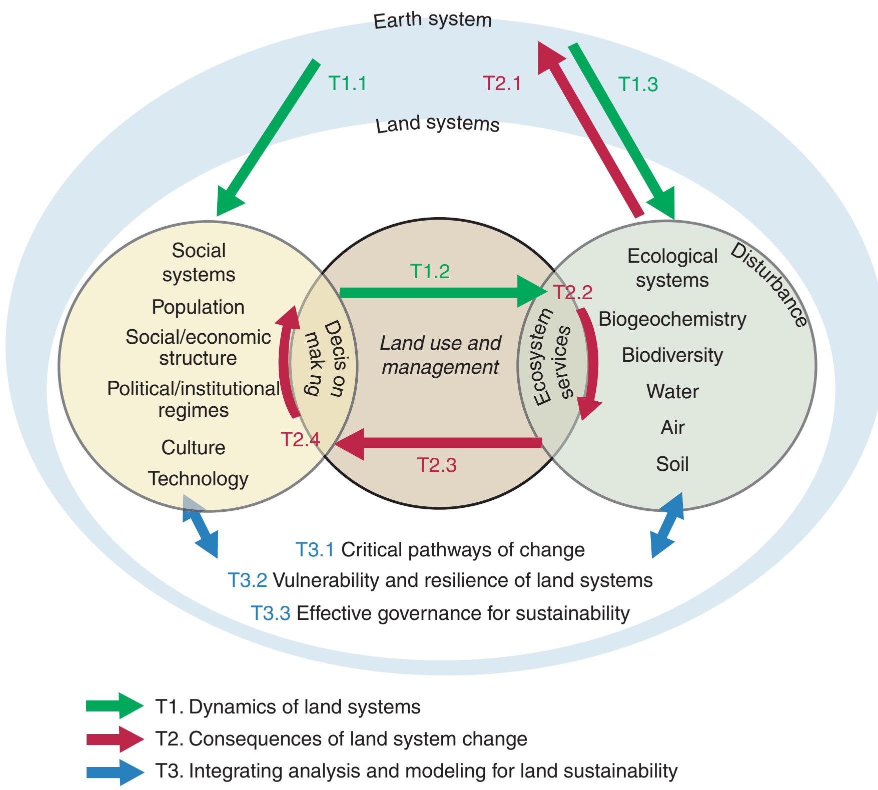

Figure 1 Thematic concerns of land change science Global Land Project (2005). *Science Plan and Implementation Strategy*. Stockholm: International Geosphere-Biosphere Programme Secretariat.

These and other synthesis activities raise concerns about sustainable land architectures. The entire terres trial surface of the Earth will soon be governed, leading to *de facto* or *de jure* designs of land uses and covers. These designs, commonly undertaken at the local or region level, invariably seek to provide a wide range of human uses and ecosystem services. The consequences of these designs for ecosystem services and human well being at ascending spatial scales of concern are largely unknown. The aggregation of local to regional designs to the con tinental and global scales may well produce negative consequences for people and the environment processes operating at the large scales. Understanding the societal and environmental implications of the patterns of land use and land cover at different spatial scales thus beckons as a grand challenge for LCS.

LCS in Geographic Context

Geography has a long tradition of human–environment studies focused on land and landscape, the modern ori gins of which are registered by the nineteenth century German geographic traditions of *landschaft* (landscape), which can be linked to the search for 'the unity in nature' as articulated by Alexander von Humboldt and other students of natural history. The legacies of this vision, some taken to detrimental extremes, range in orientation from the sciences to the humanities as exemplified by the nineteenth century French geography of Vidal de la Blache, and the twentieth century 'geographic factor' of Ellen Churchill Semple and Ellsworth Huntington, the cultural landscape of Carl Sauer, and the human ecology and natural hazards of Harlan Barrows and Gilbert White.

Perhaps the most direct links in contemporary human–environment geography to LCS reside in cultural and political ecology. Indeed, former cultural ecologists from geography and anthropology, steeped in systems based approaches to human–environment problems, helped to spawn LCS as a formal international research program linking with GISc and ecology and environ mental science to address land change in its own right as part of global environmental change and sustainability concerns. With interests in the politics and power re lations of problem framing and answers, political ecology (PE) shares many of the same topical themes with LCS (e.g., causes of land change and vulnerability), but has entertained them foremost through the concerns about environment and development. Global change and sus tainability, and environment development are, of course, intimately intertwined, and the potential for LCS and PE to inform one another is large, as in case of vulnerability research themes, and is beginning to take place across a large spectrum of shared topical interests. Indeed, in an increasing number of cases it is difficult to distinguish the methods used in and the outcomes of land based research themes carried out in either subfield. Examples include the complex connections between international policies, land uses, and drought as they affect desertification in the Sahel of West Africa; Brazilian policy shifts, market va garies, and largeholder–smallholder dynamics affecting the kind, amount, and pace of deforestation in the Brazilian Amazon; and human–environment relation ships that need to be considered in order to establish effective biosphere reserves. Perhaps the main distinction between the two, serving as a modest impediment to more fruitful exchanges, resides in their explanatory foci: LCS anchored in post positive science and PE in struc tural and constructivist explanatory visions.

Regardless of the distinctions among different inter ests in geography, LCS addressed through the coupled human–environment systems appears to be at the van guard of an emerging human–environment science that transcends traditional disciplines. The union of human, environmental, and geographical information/remote sensing research expertise is taking place in different programmatic ways that hold long term consequences for the human–environment traditions in geography. This union has begun to be matched by the newly minted schools and degree programs, especially in North America, that hold the human–environment conditions and the coupled system dynamics that create them, as the objects of their study. This development may be seen as reinventions, with a decidedly science orientation, of the human–environment intellectual traditions championed by Humboldt and of its LCS component as linked to *landschaft*.

See also: Environmental Studies and Human Geography; Physical Geography and Human Geography; Political Ecology; Remote Sensing; Resource and Environmental Economics; Simulation; Vulnerability.

Further Reading

Global Land Project (2005). *Science Plan and Implementation Strategy*. Stockholm: International Geosphere Biosphere Programme Secretariat.

Gutman, G., Janetos, A., Justice, C. *et al*. (eds.) (2004). *Land Change Science: Observing, Monitoring, and Understanding Trajectories of Change on the Earth's Surface*. New York: Kluwer Academic.

Kates, R. W., Clark, W. C., Corell, R. *et al*. (2001). Sustainability science. *Science* 292, 641 642.

Lambin, E. and Geist, H. (eds.) (2006). *Land Use and Land Cover Change: Local Processes to Global Impacts*. Berlin: Springer.

Liverman, D., Moran, E. F., Rindfuss, R. R. and Stern, P. C. (1998). *People and Pixels: Linking Remote Sensing and Social Science*. Washington, DC: National Academy Press.

Millennium Ecosystem Assessment (2005). *Ecosystems and Human Well Being: Synthesis*. Washington, DC: Island Press.

Reynolds, J. F., Stafford Smith, M., Lambin, E. F. *et al*. (2007). Global desertification: Building a science for dryland development. *Science* 316, 847 851.

Turner, B. L. II. (2003). Contested identities: Human environment geography and disciplinary implications in a restructuring academy. *Annals of the Association of American Geographers* 92, 52 74.

Turner, B. L. II, Lambin, E. F. and Reenberg, A. (2008). The emergence of land change science for global environmental change and sustainability. *Proceedings of the National Academy of Sciences* 104, 20666 20671.

Turner, B. L. II, and Robbins, P. (2008). Land change science and political ecology: Similarities, differences, and implications for sustainability science. *Annual Review of Environment and Resources* 33, 295 316.

Veldkamp, T. and Lambin, E. F. (eds.) (2001). Predicting land use change. Special issue. *Agriculture, Ecosystems, and Environment* 85(1 3), 1 6.

Wu, J. (2006). Landscape ecology, cross disciplinarity, and sustainability science. *Landscape Ecology* 21, 1 4.

Relevant Websites

http://www.globallandproject.org/
Global Land Project.

http://www.igbp.kva.se/
International Geosphere Biosphere Programme.

http://www.ihdp.unu.edu/
International Human Dimensions Programme on Global Environmental Change.

http://lcluc.umd.edu/
NASA Land Cover and Land Use Change Program.

Land Rent Theory

J. Jäger, University of Applied Sciences BFI, Vienna, Austria

Glossary

Accumulation The term refers to the growth of the stock of capital. It is commonly assumed to be an important prerequisite for sustained economic growth.
Fictitious Commodity Goods which are not produced directly by a firm but are treated as a commodity such as land, labor, etc.
Surplus It refers to the part of an output which is not used for the reproduction of labor in the form of wages. It is either obtained by capitalists in the form of profit or by landowners in the form of rent.

Definition

Property markets play an important role in structuring geographic phenomena and processes such as the spatial distribution of economic activities or residential segregation. Land is an important factor which has a crucial impact on economic activity and on the distribution of income between individuals and groups in society. The analysis of land markets is important for the understanding of landed property. Therefore, land rent theory is crucial for an understanding of these phenomena. Land rent is not merely about markets, but embraces a systematic analysis of the institutions, societal structures, and interests behind the market which constitute and structure it. In a narrow sense, land rent refers to the payment of a tribute to the land owner for the use of geographical space or natural resources in specific places. The total sum of capitalized future rent payments represents the land price. If the landowner uses the land himself or herself and, hence, no payment takes place, s/he implicitly benefits from the land. This implicit benefit is commonly also considered as rent. Land rent as a theoretical concept had originally been used mainly to explain agricultural rent but later was extended to analyze different spatial phenomena and urban structures and developments such as the location decisions of individuals and enterprises, suburbanization, gentrification, etc., within the urban and regional context. In a classical political economy perspective, prices for a produced commodity depend on the costs of production. As there is no capitalist process of production of land by a single firm, the factors which determine the price of land differ from those of an ordinary commodity. This means that land is a fictitious commodity. Furthermore, besides regulating the access to and the use of land, land rent regulates the distribution of income between different social groups or classes such as capitalists, landed classes, and workers. Therefore, land rent is crucial for explaining spatial developments and their relation with economic and social developments. Although land rent theory cannot build on one single paradigm, two broad lines of theoretical traditions are to be distinguished. The first type of theoretical approach treats land rent as a historical invariant institution which is interpreted as a natural mechanism. The second type of approach interprets land rent as a specific historical institution which changes over time. Furthermore, within both paradigms numerous distinctions are observable. This makes the field of land rent theory a challenging terrain.

Historical Foundations of Land Rent Theory

Although the above definition of land rent as a tribute to a landowner is generally agreed upon, the aim and scope of different approaches varies widely. Diverging aims and scopes explain many important but not all differences between the approaches. In order to provide a sound base for the understanding of land rent different concepts are examined in the context of their broader theoretical, but also historical background.

Land Rent in Political Economy

Rent theory was developed by classical economists. They built their concept of land rent on central categories which had been developed by forerunners such as William Petty in the seventeenth century and James Anderson in the eighteenth century. Starting with Adam Smith and continued by David Ricardo and Karl Marx classical political economists developed different concepts of land rent. While French physiocrats had supposed that land was the origin of surplus, Adam Smith argued that it was the human labor. One of the main concerns of classical political economists was to understand the overall dynamics of growth and to explain the distribution of income among social classes. Land rent was understood as part of the surplus value obtained by the landed classes. In the context of a dominant role of agriculture in society, land rent theory developed by classical economists mainly focused on agricultural rent. Some of the central questions were whether the growth of industrial production was restrained by land rent and how the landed classes were affected by the process of capitalist accumulation. Smith developed a rather har monistic perspective regarding the interests of the landed classes and the capitalists, although at the same time blaming landlords for obtaining unearned income. Tak ing different conceptualizations of rent such as absolute rent and differential rent and having no coherent ex planation for the determinants of profit his statements on land rent resulted in contradictory perspectives. Ricardo had a much more coherent explanation of profit and based his rent theory mainly on the concept of extensive differential rent. In his perspective, land rent was a de duction of profit and determined by the difference in productivity/location between a specific land and the worst land used for production, as the conditions at the latter determine the price of the commodity produced (e.g., corn). Therefore, better land should yield a higher rental payment whereas no rent is obtained on the in ferior land. An expansion of capitalist production and population, in his opinion, leads to the inclusion of land of inferior quality with the consequence of increasing differential rent and, therefore, lowering overall profit. The fall in profit reduces investment, and the process of accumulation finally leads to crisis and stagnation. Marx (and Friedrich Engels) continued with the tradition set up by Smith and Ricardo but also criticized it. Combining classical political economy approaches with different types of land rent for explaining different phenomena they insisted on the social dimension of rent and argued against the naturalization of it. They did this by analyzing the historical emergence of modern capit alist forms of land rent. Modern landed property was assumed to presuppose a disentanglement of direct producers from land. They were supposed to be con verted into workers who are free to sell their labor but also free from the ownership of land which would have allowed a subsistence economy. That means that land was defined as private property within specific societal structures.

Johann Heinrich von Thünen's Land Rent Theory

Heinrich von Thünen published his model of the isolated state in 1826. The second part of it came out in 1850. He was the first to explicitly bring land rent theory together with geography and in so doing explains the spatial structures of production within a coherent model. He took land rent in the form of differential rent as a key concept for explanation. This means that better land (in this case due to a more favorable location) allows the appropriation of higher rents because the conditions of production are relatively better. Von Thünen provided a model of marginal productivity for the simple case of an isolated state with one city center. He assumed that all the agricultural products had to be transported to the market place in the center of the city. A second simpli fying assumption was the existence of a homogenous hinterland. He intended to trace back the causes of rent to the given conditions of production and to explain how and why rents are determined by transport costs to the city center. Furthermore, his model demonstrated how rent determines the type of goods produced at different distances from the center. This is represented graphically by the so called Thünen rings.

Current Developments in Land Rent Theory

Modern approaches to land rent theory build either on the broader and more heterogeneous political economy tradition or on von Thünen's model. While modern theories based on the latter linked their modeling to neoclassical theory, the modern political economy tra dition in land rent theory resulted in a relatively higher degree of heterogeneity.

Land Rent in Neoclassical Economics and Urban Applications

In general, the principal insights of von Thünen's theo retical model are still at the core of modern approaches to the analysis of transport costs as determinants of land use and rent. His key argument that at each location the good providing the highest rent is produced and, hence, spatial structure is determined, is still a central starting point of theory. In 1909, Friedrich von Wieser developed a model conceptualizing rent within a neoclassical framework. This took the form of monopoly rent, as suming that households' demand, that is their capacity and willingness to pay for urban locations, differs. This was supposed to be a central mechanism to assign urban space to households. Nevertheless, this did not have a considerable impact on the debates that followed. The takeoff of land rent theory in the von Thünen tradition is to be dated back to 1964, when William Alonso published his book on location and land use. He was the first to fully integrate von Thünen's approach into a neoclassical model. In so doing, he provided a spatial model of eco nomic activity and founded a new field of research branded new urban economics. Rent is considered as a payment for land or more precisely for urban space. Type and intensity of land use and rent are described as a function of distance from the Central Business District (CBD) which determines a general economic and spatial equilibrium. While von Thünen's model is directed toward the explanation of location decisions for the case of production, Alonso integrated into his model repro ductive activities, such as the location decisions of households. At the core of this neoclassical model stands the idea of rational utility maximizing behavior not merely that of firms but also that of households. The model assumes that households choose locations at a specific distance to the CBD and thereby maximize their utility. Living closer to the city center implies, on the one hand, lower transport costs (in terms of money and time) but on the other hand, less space for living due to higher rents. At more distant locations rents are lower but transport costs are higher. Given a budgetary restriction households have to make their location decision in accordance with their preferences. This is represented by the bid price function which is a set of land prices a household is willing to pay at different distances while maintaining a constant level of satisfaction. As a consequence, households' preferences, budgetary restrictions, and transport costs determine rents and urban structure.

The models based on Alonso provide a deductive explanation of rental levels and regional structures. Moreover, as rents are higher in the CBD and lower in the periphery they are often assumed to be a good approximation to reality. Nevertheless, these models fail to explain the often very different spatial structures between different cities. This is due to a high degree of interdependence with a huge variety of relevant variables such as multiple forms of agglomeration benefits and costs which are difficult to be grasped in formal models. Another problem of such models is that they find it difficult deal with historical time, let alone the role of institutions, which are both crucial for the explanation of urban structures and their development. The models were used to argue that the maximization of utility or profits via the rent mechanism leads to an optimal structure of land use. The political conclusion, hence, usually was not to intervene in market forces. Nevertheless, the Alonso model has been refined since then. Relaxing the rigid assumptions has also led to more diverse conclusions for political intervention in urban and regional space.

Some of the rigid assumptions were abandoned and other factors such as transaction costs and externalities were included by authors such as Masahisa Fujita. These refined models focus not only on the spatial implications of rent but also on the question of whether political interventions such as rent controls have negative consequences on overall well being or not. While traditional Alonso type models usually conclude that any intervention in the private market is harmful, the modernized versions depict a more heterogeneous picture. Nevertheless, as demonstrated in the Cambridge controversy in capital the theoretical base of the core of neoclassical theory is rather unsound. The controversy highlighted the fact that the neoclassical concept of capital and the idea of marginal productivity and the production function describing the relation between inputs of capital, labor, (and land) presupposes a distribution of income among these factors. Therefore, it is theoretically inconsistent to explain the aggregate distribution of income between factors of production with the concept of marginal productivity. This means neoclassical theory is characterized by a substantial inconsistency. Notwithstanding, this theory continues at the core of new urban economics. Furthermore, it is frequently a criticism that the societal fundamentals of rent as well as the preferences of people are taken as given and, therefore, not subjected to reflection. While different empirical patterns of land use and rental structures are interpreted as resulting ultimately from a different set of individual preferences in market decisions, this is substantially different in modern approaches in the political economy tradition.

Modern Political Economy Approaches

An upswing in critical political economy theorizing has taken place since the end of the 1970s. This has provided an important stimulus for the further development of land rent theory. Compared to new urban economics, assumptions are less rigorous and theoretical approaches have a much higher degree of diversity. There are three main explanations for this: first, the political economy approach has a much broader scope in terms of the explaining phenomena related to land. This is due to the fact that political economy deals with economic, political, and social phenomena not in a separate manner but sees them as embedded into societal structures. Second, and related to the first point, the type of questions raised and the way in which they are answered is much more heterogeneous. Third, political economy itself is not one monolithic theory, such as neoclassical theory, but an approach with a broad diversity of theoretical paradigms.

Most of the modern approaches in a broader political economy tradition were oriented toward urban questions and dealt with different types of rent which were often defined in diverging ways. Nevertheless, on a general level, absolute rent, monopoly rent, extensive differential rent, and intensive differential rent are to be distinguished. The different types refer to different 'logics' and usually are not understood as accounting tools but rather as analytical approaches to arrive at a more sophisticated understanding of the role of land. Absolute rent was a concept put forward by Karl Marx and refers originally to a land tribute which is due to a lower organic composition of capital in agriculture. This means that production is relatively less capital intensive. Modern theorists considered this to be a key feature in the construction industry. This notion of absolute rent is based on the value theory of labor which was criticized for its theoretical inconsistencies. Besides this interpretation of absolute rent this type of rent was also considered as a class monopoly rent caused by the power of landowners to retain land from productive or reproductive use. As a theoretical concept absolute rent was mainly applied for the analysis of agriculture and resources. The concept of monopoly rent refers to the different capacities of different classes and social strata to pay for socially privileged space. Social groups which have more financial resources than others exclude the latter from their residential areas. Hence, monopoly rent is used to explain phenomena like urban segregation. Monopoly rent and its specific effects depend substantially on institutions such as rental contracts, housing policy, zoning, etc. Furthermore, monopoly rent is also used to explain spatial phenomena related to retail trade and service provision. Locations where higher income groups purchase or consume goods which have a monopoly character (e.g., exclusive wines and exclusive recreational areas) enable the seller to obtain surplus profit. This surplus profit due to monopolies is supposed to be converted into monopoly land rent. Extensive differential rent stems directly from the production sector and is a result of more favorable conditions of production at specific places. An example for this is retail space in prime sites which enables higher turnover and therefore reduces costs per unit. Surplus profits resulting from these privileged spaces are assumed to go into landowners' pockets in the form of extensive differential rent. Both, monopoly rent and extensive differential rent were used to analyze demand driven developments in the urban case. Intensive differential rent and, in part, absolute rent served to analyze the active role of the landed classes in general and the role of real estate developers and speculators in particular. Intensive differential rent is supposed to stem from a more intensive use of land compared to the prevailing norm. The introduction of multistory buildings is behavior by landowners or speculators which allows the attainment of intensive differential rent. Another way to obtain this form of rent is by changing the social value of certain spaces. This can take the form of gentrification or the production of gated communities in quarters which were formally occupied by social groups with a lower capacity to pay rent.

Neil Smith's rent gap theory represents a concrete approach to deal with a phenomenon caused by intensive differential rent. Instead of demand led explanations Smith proposed taking the speculators search for intensive rent as a key variable to explain gentrification. Gentrification as a phenomenon and Smith's explanation attracted the attention of numerous scholars and caused controversial discussions. He explained gentrification as a historical process which was linked via land rent to the development of capitalism. In so doing, he made a distinction between a site's capitalized land rent and its potential rent. Capitalized land rent, conceptualized as a deduction from surplus value, was assumed to represent the actual quantity of ground rent that was appropriated by the land owner, given the present land use. Potential land rent was defined as the amount of rent that could have been capitalized under the land's highest and best use. A reason for the difference between both values was seen in the disinvestment in certain places. This disinvestment led to a reduction of the overall surplus obtained at a plot of land and hence the actual rent. This mechanism was used to explain the deterioration of inner urban areas which was observable in many cities and went hand in hand with suburbanization. At the same time, the overall amount of capital applied to land increased and potential rent was supposed to experience a similar development. The growing difference between the actually capitalized land rent and the potential rent was supposed to be the driving force of gentrification because it created the possibility of surplus profits through reinvestment in the plot. This theoretical approach to land rent theory was frequently applied to address empirically the phenomenon of gentrification. In general, rent gap, and in particular its theoretical core, the difference between capitalized and potential land rent caused considerable controversies on how to measure these values. Blair Badcock and Steven C. Bourassa questioned whether the rent gap was an empirically useful concept at all. Smith responded that the rent gap was criticized as a neoclassical theory although it had never been intended to be one. The apparently opposing views were not only the result of terminological problems but of a real theoretical disagreement over land rent. One of the main questions was whether land use affected land rent on a particular plot or land rent affected the type of use of land – a crucial question in land rent theory. Daniel J. Hammel entered the debate and proposed a solution to that. He insisted that Smith's rent gap theory represented a theoretically consistent explanation of gentrification although the treatment of the relationship between land rent and land use was rather hidden in rent gap theory. Hammel pointed to the importance of scale in addressing this question. He argued that capitalized land rent had to be considered at least at two scales which he found in Smith's cycle of devalorization and his discussion of capitalized and potential rents. In Smith's approach, the neighborhood scale was crucial to explain potential land rent which was largely determined by location. Hammel argued that potential rent was determined at the metropolitan scale while capitalized land rent was determined at the neighborhood scale. Specific forms of land use in a neighborhood might have limited capitalized land rent. This conceptualization allows refuting much of the critique formulated against it and allows solving the problem of whether land use determines land rent and vice versa by insisting on different scales. Furthermore, Smith's rent gap theory was linked to his approach explaining uneven capitalist development in general. The geographical configuration of the landscape was assumed necessary for the survival of capitalism. Capitalism transformed space in its constant drive to accumulate larger quantities of social wealth. This was supposed to be associated with a dialectical interaction of a tendency toward differentiation and a tendency toward equalization of the levels and forms of production. In general, the rent gap theory is an important approach to land rent theory which helps to understand urban processes such as gentrification. Nevertheless, whether and how intensive differential rent is obtained depends heavily on the institutions regulating this type of rent seeking behavior.

The neo Ricardian approaches to land rent were based on Pierro Sraffa's systematization of David Ricardo's political economy approach, for which he provided a formal model. Land rent is conceptualized as external and internal differential rent as shown by Heinz D. Kurz and Neri Salvadori. Based on a neo Ricardian framework, Alan J. Scott developed a model for the urban context. He used Sraffa's theoretical approach assuming that land rent was caused by scarcity (external differential rent) and the application of additional capital to land (intensive differential rent). He showed how a high density of housing in inner city areas and a relatively lower density of housing in suburban areas are produced by the land rent mechanism. In contrast to Alonso, he explained urban patterns without taking preferences and utilities into account. Extensive and intensive differential rents in addition to transport costs made up the main factors which determined urban structures. Scott went well beyond traditional neo Ricardian interpretations of land rent and argued that rent had to be seen in the context of the capitalist mode of production. Building on the work by Doreen Massey and Alejandrina Catalano he insisted that specific types of landowners constituted specific sorts of social fractions. Scott also tried to integrate important urban phenomena such as urban planning into his framework. This made him exceptional in the neo Ricardian tradition, which, in general, was located at a relatively high level of abstraction and, therefore, abstracted from the analysis of institutional and political factors and their constitutive role for land rent. This was considerably different to the treatment of land in broader political economy traditions.

The distinction among the broader approaches in the political economy tradition along nomothetic and ideographic lines of theoretical enquiry has been investigated by Anne Haila. The ideographic approach was characterized by emphasizing the singularity of historical situations and social relations and by neglecting the existence of general laws. Michael Ball was one of the leading authors of this tradition. The merit of the ideographic tradition was that it directed the focus toward the analysis of concrete historical developments. Land rent was referred to as an act of payment of money and therefore interpreted as circulation of revenues, but not explained by relating rent to the process of production. The explanatory value of any general rent theory was largely denied. This was criticized and stood in sharp contrast to the prevailing nomothetic approach which was characterized by the search for laws in land rent theory. David Harvey is considered to be the central author of this nomothetic tradition. He built heavily on the different types of rents distinguished above. According to him and other authors, land rent was supposed to become increasingly a financial asset. Along with that a merger between capitalists and the landed class was observed. It was assumed that the landed classes had abandoned their rentist behavior and, therefore, opposing interests to capitalist development were neutralized. As such, land was assumed not to be an obstacle to the process of capitalist accumulation anymore. On the contrary, land rent was assumed to fulfill a coordinating function by assigning land to efficient uses and thereby boosting capitalist accumulation. Land rent was no longer supposed to be determined exclusively by production, but was assumed to be dependent on (international) investment flows in the built environment. This led to the neglect of the relation between economic accumulation, the rate of profit and land rent, and hence, the question of distribution. Moreover, within a nomothetic tradition specified in the above way, there was no systematic space for the analysis of contradictory interests and the treatment of politics left.

Both, the ideographic and nomothetic theoretical traditions – although offering important insights – have considerable shortcomings. Johannes Jäger proposed going beyond the nomothetic and ideographic traditions. He argued that the concrete context of rent should be analyzed and the fact that land rent represents a social relation should be taken seriously. Therefore, land rent was proposed to be analyzed embedded within the broader institutional context. In doing so, the specific historic interests and spatial dynamics related to land can be analyzed by building upon the different types of rent. Those different types of rent are assumed to be not invariant over time but to be changing according to their specific institutional embedding. In order to link land rent theory coherently to general developments it was proposed to integrate it into a medium range theory, in particular the regulationist approach. In regulation theory, capitalist accumulation is considered to be characterized by contradictions which temporally might be offset with an adequate regulation of its institutions or structural forms which land rent forms a part of. As land (or space) is required for any productive or reproductive economic activity it is necessary to have institutions such as land rent which regulate it. Patterns of urban development such as urban redevelopment and segregation can be analyzed linking them to distributional struggles between capitalists, landowners, and workers. In general, such an integrative conceptualization allows us to link the dynamics of economic accumulation and political regulation to urban geography and vice versa.

Scope of Explanation and Critique

Land rent is the particular institution which links the economy to geography and contributes to an understanding of urban and regional phenomena. Without doubt economic developments are crucial for the analysis of spatial structures and changes. Moreover, land rent is a category which helps to explain how geography, or the spatiality of economic activity itself, affects economic processes. There are two broad theoretical ways of dealing with land rent theory which differ substantially in scope of explanation, the underlying assumptions, and the conclusions drawn. On the one hand are the models based on von Thünen and Alonso and the work in the context of new urban economics which demonstrates a higher degree of rigidity but its scope of analysis is relatively narrow. Based on a neoclassical theoretical base, household's preferences, budgetary restrictions, and transport costs determine land rent and urban geography. While traditional approaches within this framework basically conclude that unregulated markets and land rent lead to an optimal outcome, more recent work, including externalities and market imperfections, arrives at more diverse conclusions. On the other hand, land rent theory in a political economy tradition shows a much higher degree of diversity at the cost of a higher heterogeneity between different approaches. It allows for a broader understanding of complex urban phenomena as it provides a comprehensive tool of analysis which links economic and social processes and (class) agent's strategies via different types of land rent to the production of geography. Taking the richness of land rent theory and its potential to explain urban and regional structures and development into account it is quite surprising that its use and further development in human geography is not very widespread today. At least in part this might be explained by the high degree of complexity of the different approaches.

See also: Capital and Space; Economic Geography, Quantitative; Gentrification; Location Theory; Natural Resources; Radical Political Economy; Regulation; Segregation.

Further Reading

Alonso, W. (1964). *Location and Land Use*. Cambridge, MA: Harvard University Press.

Badcock, B. (1990). On the nonexistence of the rent gap hypothesis. *Annals of the Association of American Geographers* 79, 125 145.

Ball, M. (1985). The urban rent question. *Environment and Planning A* 17, 503 525.

Bourassa, S. C. (1993). The rent gap debunked. *Urban Studies* 30, 1731 1744.

Fujita, M. (2003). *Urban Economic Theory*. Cambridge, MA: Cambridge University Press.

Haila, A. (1990). The theory of land rent at the crossroads. *Environment and Planning D: Society and Space* 8, 275 296.

Hammel, D. J. (1999). Re establishing the rent gap: An alternative view of capitalised land rent. *Urban Studies* 36, 1283 1293.

Harvey, D. (1982). *The Limits to Capital*. Oxford: Blackwell.

Jager, J. (2003). Urban land rent theory. A regulationist perspective. *International Journal of Urban and Regional Research* 27, 233 249.

Kurz, H. D. and Salvadori, N. (1998). The 'standard commodity' and Ricardo's search for an 'invariable measure of value'. In Kurz, H. D. & Salvadori, N. (eds.) *Understanding 'Classical' Economics. Studies in Long Period Theory*. London: Routledge.

Massey, D. and Catalano, A. (1978). *Capital and Land, Landownership by Capital in Great Britain*. London: Edward Arnold.

Scott, A. J. (1980). *The Urban Land Nexus and the State*. London: Pion.

Screpanti, E. and Zamagni, S. (1993). *An Outline of the History of Economic Thought*. Oxford: Clarendon Press.

Smith, N. (1979). Towards a theory of gentrification. *Journal of the American Planning Association* 45, 538 549.

Smith, N. (1984). *Uneven Development. Nature, Capital and the Production of Space*. Oxford: Basil Blackwell.

Sraffa, P. (1960). *The Production of Commodities by Means of Commodities*. Cambridge, MA: Cambridge University Press.

Von Thunen, H. (1966). *Von Thunen's Isolated State,* edited by Peter Hall. London: Pergamon Press.

Von Wieser, F. (1909). *Die Theorie der Städtischen Grundrente*. Wien: Deuticke.

Land Rights

R. Howitt, Macquarie University, Sydney, NSW, Australia

Glossary

Geographical Imagination Generally refers to the shared understanding of landscape, environment and social history and that provides a people or cultural group with a shared sense of territory and a common place in the world. In challenging the ways in which political values and environmental understanding affect geographical imagination, the discipline often seeks to nurture informed debate and misunderstanding.

Indigenous Peoples The first peoples to live in an area with customary ties and relationships to particular territories and customary laws and cultural forms that institutionalise and govern those relationships.

Indigenous Rights The passage of the United Nations Convention on the Rights of Indigenous Peoples provides a common standard for recognition of the rights of Indigenous Peoples. While many key struggles of Indigenous Peoples relate to protection of basic human rights, there is also a set of claims based on pre-existing and persistent links between people and territories and unresolved claims by colonial states to have usurped specific rights that underpin claims of Indigenous rights.

Postcolonialism Movements or studies that investigate, correct or mitigate the complex impacts and consequences of colonialism. It can also refer to independent governance arrangements in previously colonised states.

Treaty Rights Many Indigenous Peoples claim specific rights that were settled in treaties with occupying or colonising forces. These rights are often seen as anachronistic by modern states but have, for example in Aotearoa New Zealand, provided a powerful foundation for renegotiating the terms of coexistence between Indigenous and settler societies.

Introduction

The political movement advocating indigenous land rights has been an important element of postcolonial politics in many settler societies. The land rights movement argued that returning land to its traditional owners was fundamental to changing the place of indigenous peoples in settler nation states and addressing the historical legacies of colonial displacement of first nations. In many settler nations, indigenous peoples' land rights have been a difficult and divisive political issue, challenging widely held assumptions about national identity and future trajectories.

The idea of persistent indigenous rights to the land that forms the postcolonial nation state challenges ideas that lie at the heart of contemporary constructions of place, identity, and nation. In settler societies such as Australia, New Zealand, Canada, and the USA – the four nation states which voted against the Declaration of the Rights of Indigenous Peoples in the United Nations General Assembly in September 2007 – the land rights movement has been central to what Canadian scholar Michael Ignatieff characterized as the 'rights revolution' in 2000, which has redefined the idea of relationships between a nation state and its citizens since the 1960s.

In this context, the political movement for indigenous land rights has often framed a profound challenge to Eurocentric assumptions about the relationships between territory, identity, and governance that have been so fundamental to recent debates in political and cultural geography. The roots for the contemporary relevance of the land rights issue, however, are to be found much further back in colonial histories.

Land, Law, and Identity: Historical and Legal Constructions

Ideas about human territoriality have long been linked with the idea of possession of land. One of the challenging political questions arising from notions of territoriality is who, or what, belongs to whom? Does the land belong to the people – or do the people belong to the land? For many indigenous groups in settler nation states, the claims of a special relationship to land are central to their claims of identity – but the dominant legal systems of settler nation states have typically rejected such claims as poorly founded in law. For indigenous peoples, the social expression of the relationship is likely to be constructed as collective rather than individuated responsibilities and rights. The documentation is more likely to take the form of genealogy (in Maori, *whakapapa*), ritual, or story rather than a centrally recorded document. The claim is more likely to be founded upon a sense of 'time immemorial' or 'being born from the Earth' rather than derived from a claim of sovereignty. Does this reduce the validity or influence of the claim? To answer that question, it is necessary to consider the histories of possession and dispossession that arose from colonial occupation of the settler states.

Geographical expansion of European influence into the so called New World in the fifteenth and sixteenth centuries brought European property systems, which were undergoing rapid change as a result of the emergence of new systems of governance, the development of new wealth and the transformation of isolated kingdoms into externally oriented empires – into juxtaposition and often conflict with pre existing systems of property rights and access to land, resources, and wealth. The fate of precolonial property systems varied from place to place, but the treatment of property systems of indigenous and tribal groups, particularly nomadic peoples (whose failure to cultivate and subdue nature placed them on the wrong side of the biblical ejection of humanity from the Garden of Eden), is generally characterized by acts of denial, erasure, and removal. The universalizing claims of linear human progress that Christian Europe proclaimed as the basis of its own superiority to others, mixed with emergent notions in the nineteenth century with naïve Darwinian notions of fitness and privilege to produce extraordinarily barbaric and genocidal treatment of indigenous groups, with frontier wars that aimed at indigenous peoples as peoples. Customary legal systems and ancient jurisdictions of the New World were often dismissed as primitive and inferior. Even in the heartlands of Europe, the treatment of rights derived from customary and common law systems in the transition from feudalism to capitalism produced violent clearances and enclosure of commons, with powerful and propertied lords usurping the rights of commoners and marginalized kinfolk. Thus, in both the heartlands and the frontiers, the history of colonization was often written, as Karl Marx colorfully proclaimed, in letters of blood and fire. Customary law property systems not based on the allocation of separate and exclusive rights to identified and bounded spaces were often deemed by colonial powers too primitive to be considered equivalent to refined European notions of property and law. Consequently, these communal and collective rights were often simply denied, ignored, and erased as colonial systems developed.

Indigenous Property in the British Empire

Highly racialized hierarchies of humanity allowed emissaries of church and state to proclaim particular peoples as so primitive that law and property were unknown in their societies. This erasure by representation reached its zenith, perhaps, in the principle of *terra nullius* – land that belonged to no one. Australia, for example, was claimed by the British as property acquired not by conquest or cession but by settlement – because it was declared to be *terra nullius* by the occupying state power, in denial of the ancient jurisdiction already in place.

Under the legal principle of *terra nullius*, entrenched in Australian colonial law by the mid 1830s – around the same time as the US Supreme Court was shaping the doctrine that defined the First Nations of the USA as 'domestic dependent nations' – the British Crown claimed to acquire absolute sovereignty over the entire continent of Australia. Australia's colonial systems of property and resource rights were constructed as if Aboriginal peoples' interests were simply absent from the landscape.

In British North America, where claims of *terra nullius* would strain the truth too far, treaties were designed to achieve the same form of absolute possession. The American Revolution produced a series of new treaties, where First Nations provided treaties of peace and recognition that extended international recognition to the new republic. Western expansion of the United States produced more treaties – and more erasure of indigenous rights. Songwriter Buffy Sainte Marie captures the ethos in her powerful anthem *My Country 'Tis of Thy People Your Are Dying*:

> Hear how the bargain was made for the West,
> With her shivering babies in zero degrees.
> 'Blankets for your land' so the treaties attest.
> Oh well, blankets for land, that's a bargain indeed.
> (Buffy Sainte Marie, 1967 album)

In Aotearoa/New Zealand, the Maori land wars, which threatened the expulsion of the British from the islands, saw the signing of the 1840 Treaty of Waitangi, in which the British Crown claimed that the Maori chiefs had ceded sovereignty (and retained *kavanatanga* (governorship) to themselves). In Canada, massive British land grants in Quebec to the Hudson Bay Company were never secured by treaty nor confirmed by settlement – and yet the land involved was ceded to the Province of Quebec in 1912, conditional upon the making of treaties with the region's traditional owners. Failure to settle treaties at that time produced problems later for development of the area's hydroelectric potential and triggered Canada's modern treaty process with the negotiation of the James Bay and Northern Quebec Agreement in 1975. Similar land grants, for example, in Sabah in East Malaysia, provided the foundations for much more recent processes of dispossession through the operation of postcolonial government processes of planning, resource development, and land management. In the USA, despite recognition of treaty rights as the basis for a special relationship with the federal government, systemic corruption of that relationship through the Bureau of Indian Affairs has produced a legal maze of treaty breaches, illegal relocations, deceitful contractual arrangements, and imposed underdevelopment.

Indigenous Land Rights in the Twentieth Century

In the early twentieth century, indigenous interests in the USA, Canada, Australia, and New Zealand faced political and social marginalization. Entrenched anti indigenous racism ensured that indigenous people suffered unequal conditions in employment, housing, and education. For many indigenous groups, the best that could be expected was the cruel paternalism and grinding poverty of gov ernment reserves and reservations. Social policies were often more concerned with exercising control over in digenous people, separating children from families, and penalizing expressions of indigenous identity such as language, religious ceremony, or family relations. Re strictions on indigenous mobility, control over property, labor power, and wealth were all central to government programs aimed at removing indigenous peoples as separate and persistent entities in settler societies.

Through the early twentieth century, treaties with First Nations were generally simply ignored by govern ments, treated as anachronistic in the modern indus trializing state. Breaches of treaty obligations were often compelled by government policies, and state bureau cracies often completely failed in the roles with which they were charged. These roles typically involved pro tection, welfare, and advancement or with managing the property and interests of people whom the dominant legal system defined as wards of the state, or as too primitive to hold status as full citizens. In settled areas, the twentieth century typically saw a microgeography of dispossession, characterized by unsurveyed reservation boundaries being encroached by farmers and cities; lost documentation on land grants to individual indigenous people; government agencies disposing of indigenous land and resource assets in unfavorable arrangements (often involving criminal negligence and corruption); and poverty and trickery swindling people out of land through insufferable credit and contractual arrange ments. The opportunity to retain, or secure, return of land in such circumstances was difficult.

Decolonization after World War II left the question of indigenous rights in settler societies largely unaddressed. The fate of indigenous minorities in postcolonial states was considered an internal matter, largely immune to international comment or intervention. The enormous inequities between indigenous and settler populations in terms of health, education, employment, wealth, and well being reflected the failure of settler societies to place policies and practices that would distribute the benefits of national development to their indigenous citizens.

In areas that were more remote from the metropolitan heartlands of the settler societies, government reserves and reservations often covered vast areas of little interest to governments and commercial interests. It was often these areas of apparent wasteland that governments began to consider returning to some form of indigenous ownership when the liberalizing influence of the civil rights movement generally, and the land rights move ment in particular, began to shift the political landscape toward greater social acknowledgement of the persistent indigenous presence in the cultural landscapes of the settler societies. Ironically, however, by the early 1970s, these remote and isolated indigenous domains often became the target of resource explorers keen to acquire preemptive rights from governments, prior to any move to grant land rights to traditional owners of the lands. This period was widely seen as constituting a 'new re sources war' – a term coined by the US sociologist Al Gedicks in 1994. In more recent times, the coincidence of many indigenous domains with areas of high biodiversity and environmental value has heralded a new set of challenges with pressure to displace indigenous peoples from conservation lands, and limit their use of traditional resources.

Inequality was reflected in most official statistics – where states included indigenous people in their counts and samples. By the 1950s, indigenous populations in wealthy societies were poorer, less well, more un employed, less educated, more incarcerated, and lived shorter lives. Such inequality always has a specific geography and a history. It arose in particular circum stances in which the disadvantage resulting from loss of land was entrenched across generations into the patterns of the dominant settler society. In the Australian case, which in many ways reflects indigenous experience more broadly, the response of the dominant political discourses has involved narratives of absence, erasure, and denial. Both physical and cultural landscapes were emptied of the indigenous owners, the traces of the ancient juris dictions have been erased, and the persistent presence of indigenous peoples and their rights in contemporary landscape denied.

The lived cultural landscapes of settler societies are a complex reflection of the people to people and people to place relationships that structure everyday life. In multicultural, pluralist societies, which many of the set tler nation states have become, the values, perceptions, aspirations, and contributions of many groups shape those landscapes. Yet, the rights, interests, and values that predate the colonial state have proved difficult to ac knowledge and incorporate into these societies. The land rights movement has been one important element of the response.

By the late 1960s, under pressure from broad pro cesses of societal change reflected in the civil rights movement, increasingly liberal social legislation, and discomfort with overt racism, formal legislative responses to indigenous peoples' calls for land rights were considered. The land rights movement in Australia traces its origins deeply into the resistance and social justice movements supporting aboriginal and Torres Strait Islander people throughout the nation's colonial and postcolonial history. The post World War II movement traces origins to the strike of aboriginal pastoral workers in Western Australia's Pilbara region in 1950, and the Gurindji people's celebrated strike against appalling conditions and acquisition of their traditional land by the British company Vesteys in the Northern Territory in the early 1960s. In the Gove Land Rights case in the Northern Territory Supreme Court, Yolngu people from northeast Arnhem Land challenged the right of the Australian Government to grant land, which they felt was their own, as a mining lease to a Swiss aluminium company. While the 1971 judgment confirmed the principle of *terra nullius* as a foundation of Australian law, it also acknowledged the sophistication and organization of Yolngu customary law. It did, however, reject the Yolngu claim, arguing rather the Yolngu system of rights (in which people belong to the land rather than land belonging to people). In the wake of that decision, however, the incoming government in 1972 undertook to draft legislation to return land to Aboriginal ownership in the Northern Territory. The Aboriginal Land Rights (Northern Territory) Act of 1975 was finally introduced after much debate and protest. In the Australian case, Aboriginal people had established a tent embassy in the early 1970s to emphasize just how alienated they felt they were from the state. State and federal legislation was drafted in the early 1970s to return Aboriginal reserve lands in the Northern Territory and South Australia. In the case of the Northern Territory, the legislation also allowed for Aboriginal traditional owners to make claims to other lands, particularly Vacant Crown Lands – unused government lands – where they could demonstrate that they met the requirements of the legislation.

The provision of 'land rights' by legislation, however, has not produced equality – a matter that is seen as a failure of land rights by many critics. But in many cases, the postcolonial state's continued claim to have usurped all resource and development rights that might have been available to traditional owners as full beneficial owners of land meant that the legislation involved often tightly prescribed the sorts of rights granted by the state – and ensured that there would be no threat to postcolonial sovereignty. For example, the NSW Land Rights Act of 1983 allowed Aboriginal people in New South Wales (NSW) to claim land that was not required for any other purpose. The patchwork of land open to claim included decommissioned waste disposal sites, inaccessible land, land cut off by road developments and many areas of no value and little traditional significance beyond its status as part of a much larger historical territory. Inalienable freehold title (intended to reflect traditional values of land as mother) often meant that ownership of land could not be used to raise finance, and without resources, using the land for productive economic, social, or cultural purposes became difficult. In many places, exemptions from land taxes or local government rates, intended to reduce the financial burden on impoverished communities, was construed by local authorities to mean that local services were not provided. In remote areas, this sort of attitude embedded in entrenched patterns of structural racism and ingrained prejudice against indigenous peoples reinforces high costs, poor access to services, and poor economic and social outcomes.

In other jurisdictions, the return of traditional territories to indigenous ownership and control has relied on a range of judicial, parliamentary, and policy based initiatives. The principle of self determination, which underpinned the decolonization process in the former so called Third World, was rarely acknowledged in the settler societies as relevant to colonized people of the Fourth World of indigenous and minority groups in the main settler societies. While that principle, along with the principle of indigenous ownership of their traditional estates, is now acknowledged in the United Nations Declaration on the Rights of Indigenous Peoples, it remains unachievable in many jurisdictions.

Land Rights and Geographical Imagination

In postcolonial periods, indigenous groups have pursued land rights as a central element of their claims for recognition. Acknowledgement that customary rights to their traditional territories existed prior to colonial occupation has been difficult to secure, not least because such acknowledgement is often understood to undermine the postcolonial state's claims to sovereignty and may destabilize the legal foundations of government administration of property, resources, and environment.

In other words, the question of land rights underpins the social and geographical imaginary of the settler nation state. The persistent presence of indigenous interests belies the heroic, self congratulatory narratives of nation building that underpin identity and politics in the modern pluralist democracies. In national economies where land and natural resources provide both historical and contemporary foundations for national wealth, the land claims of displaced, but surviving indigenous groups unsettles the assumption that wealth is a product of hard work and good luck. The broken treaties, the racist violence, the genocidal evil of eugenic social policies, and the current failures to resolve these issues affirm that theft and violence have also been the foundations for settler nation prosperity.

In more recent times, following legal successes in establishing the persistence of Native title in Australia, resource rights in North America, and treaty rights in New Zealand, there has been a significant shift toward agreement making with indigenous groups, and the for mal protocols of the modern treaty process. More widely, the work of international representative and support groups in the nongovernment sector (e.g., IWGIA, Cul tural Survival, World Council of Indigenous Peoples, Inuit Circumpolar Council, etc.), the emergence of indigeneity as a significant identifier for marginalized peoples in many areas of Asia, South America, and Africa, establishment of a United Nations Permanent Forum on Indigenous Peoples and, after decades of deliberation and debate, passage of the Declaration on the Rights of In digenous Peoples, has focused attention on indigenous rights, including land rights. In many jurisdictions, the land rights struggle has emerged as a new focal point in the struggle for justice and identity. In the settler nation (nation states) states, it has been recognized that in both political and legal terms, indigenous rights may further unsettle the taken for granted narratives of settlement, property, and prosperity. New Zealand's concessions around resource and environmental management, off shore marine resources, and geothermal energy resources and Canada's new rounds of modern treaty negotiations, all suggest that the land rights issue continues to have wide currency in unequal societies. The emergence of new discourses around indigenous claims to saltwater country (i.e., claims to territorial waters of nation states), water, and resources that indicate the extent to which the issues that generated the land rights movements of the settler states remain unresolved.

In the twenty first century, the issues of indigenous rights cannot be limited to questions of land rights, nor can governments expect the international community to accept that their treatment of indigenous peoples is a matter of internal affairs. No longer focused solely on the return of traditional lands, indigenous peoples' movements challenge assumptions about human–nature relations; they highlight the importance of cultural diversity on a global scale, challenge the narratives of globalization, and their continued presence in their tra ditional territories and the cultural landscapes of the contemporary world challenge the geographical imagin ation to encompass much more than Eurocentric visions of development as the basis for understanding people's places in shaping the planet.

Conclusion

Indigenous peoples have often been overlooked in geo graphy's consideration of place. Key texts by Soja have narrated a 'cacophonous silence' around First Nations' place in the history of the United States. Yet, political movements advocating protection and return of in digenous land rights have provided and punctuated the development narratives of many nation states. Given the importance of the link between human territoriality, state sovereignty, and national identity, the lack of critical engagement with indigenous rights in the discipline of geography is perhaps surprising, but the establishment of an International Geographical Union Commission on Indigenous Peoples' Knowledges and Rights in 2006 in dicates that this is emerging as a significant element in the discipline in the twenty first century. The orientation of that commission toward both research and advocacy further suggests the relevance of the discipline to in digenous futures. The interplay of legal, anthropological, and political discourses in delivering and/or denying indigenous peoples rights to their traditional territories has certainly been an issue that has affected the historical geographies of many settler states. Particularly in the former British colonies, land rights as legal argument, cultural relationship, social movement, political prop osition, and policy program has significantly influenced national identity and politics. The history of land rights opens up valuable lines of thought on key geographical concepts such as human territoriality, boundaries, cul tural landscapes, and human–nature relations.

See also: Activism; Identity Politics; Indigeneity; Indigenous Geographies; National Spatialities; Place Names; Racism and Antiracism.

Further Reading

Asch, M. (1997). *Aboriginal and Treaty Rights in Canada: Essays on Law, Equality, and Respect for Difference*. Vancouver: University of British Columbia Press.

Blomely, N. (1994). *Law, space, and the geographies of power.* New Work: Guildford Press.

Brody, H. (2000). *The Other Side of Eden: Hunters, Farmers and the Shaping of the World*. Vancouver: Douglas & McIntyre.

Cant, G., Goodall, A. and Inns, J. (eds.) (2005). *Discourses and Silences: Indigenous Peoples, Risks and Resistance*. Christchurch: Department of Geography, University of Canterbury.

Gedicks, A. (1994). *The New Resource Wars: Native and Environmental Struggles Against Multinational Corporations*. Montreal: Black Rose Books.

Goodall, H. (1996). *Invasion to Embassy: Land in Aboriginal Politics in New South Wales, 1770 1972*. Sydney: Allen & Unwin.

Howitt, R. (2001). *Rethinking Resource Management: Justice, Sustainability and Indigenous Peoples*. London: Routledge.

Ignatieff, M. (2000). *The Rights Revolution*. Toronto: Anansi.

Ivison, D., Patton, P. and Sanders, W. (eds.) (2000). *Political Theory and the Rights of Indigenous Peoples*. Cambridge: Cambridge University Press.

Jaimes, M. A. (1992). *The State of Native America: Genocide, Colonization and Resistance*. Boston, MA: South End Press.

Johnson, J. T., Cant, G., Howitt, R. and Peters, E. (2007). Creating anti colonial geographies: Embracing indigenous peoples' knowledges and rights. *Geographical Research* 45(2), 117 120.

Sack, R. D. (1986). *Human Territoriality Its Theory and History*. Cambridge: Cambridge University Press.

Shaw, W. S. R., Herman, D. K. and Dobbs, R. (2006). Encountering indigeneity: Re imagining and decolonizing geography. *Geografiska Annaler: Series B, Human Geography* 88(3), 267 276.
Soja, E. W. (1989). *Postmodern Geographies: The Reassertion of Space in Critical Social Theory*. London: Verso.
Stewart Harawira, M. (2005). *The New Imperial Order: Indigenous Responses to Globalization*. London: Zed Books and Huia Publishers.
Wisner, B. and Stea, D. (eds.) (1984). *Special Issue on Indigenous Issues. Antipode* 16(2).
Young, E. (1995). *Third World in the First: Development and Indigenous Peoples*. London: Routledge.

Relevant Websites

http://www.pacificworlds.com
Association of American Geographers Indigenous Specialty Group.
http://www.yorku.ca
Canadian Association of Geographers Native Canadians Specialty Group.
http://www.cwis.org
Centre for World Indigenous Studies.
http://www.unl.edu
International Geographical Union Indigenous Peoples' Knowledges and Rights Commission.
http://www.iwgia.org
International Working Group on Indigenous Affairs (IWGIA, Copenhagen).
http://www.inuitcircumpolar.com
Inuit Circumpolar Council.
http://www.ohchr.org
United Nations High Commissioner for Human Rights (Declaration of the Rights of Indigenous Peoples).
http://www.un.org
United Nations Permanent Forum on Indigenous Peoples.

Landscape

J. Dubow, University of Sheffield, Sheffield, UK

Glossary

Cartesian The term refers to that dominant strand in Western thought which presupposes the essential independence and self-sufficiency of the rational human subject and argues for a theory of consciousness in terms of a detached, disembodied, contemplative knower. In this, it is a form of critical idealism which, based on the separation of mind and body, presents us with a model of reality which is rationally systematic but which fails to acknowledge the discontinuities between consciousness and the world.

Cultural Materialism A critical approach which developed in Britain in the late 1970s and 1980s, it is a term allied to Marxism, but which, in the productive tension of its two constituent terms, is resistant to the rigidities of economic determinism. While it maintains that there is no crude material reality beyond culture, it simultaneously asserts that cultural forms are embedded within the historical and material processes which formed them and in which they play a central role.

Esthetic First coined by the eighteenth-century German philosopher Alexander Baumgarten and assigned high priority in European thought since the Enlightenment, 'esthetics' most commonly refers to the tastes, categories, and criteria used in both the evaluation of art and various kinds of perceptual experience and effect. While it often assumes a universal and timeless validity, the realm of the esthetic has a clear, albeit contradictory, relation to ethics, politics, and cultural specificity.

Phenomenology A twentieth-century philosophical movement distinguished by a concentration on lived or embodied experience which reveals the sensory and perceptual 'meanings' of things that precede their abstraction in thought, knowledge, or theoretical interpretation. Exerting an immense influence on cultural and literary criticism, it has recently gained currency in cultural geography, especially in the field of landscape studies. Anti-Cartesian and anti-essentialist in impulse, it does not seek to understand the connection between an autonomous self (body) and an objective world, but rather the primordial encounter between them.

Representation A complex term, representation is conventionally defined as a symbol or image, or as the process of rendering something (an object, event, idea, or perception) intelligible and identifiable. With application to the worlds of both arts and science, representation is now understood less as a process of reproducing or describing a preexistent reality but as a complex set of meanings which often constitutes what counts and is valued as reality.

Ways of Seeing Coined in the 1970s by John Berger, a Marxist art theorist, the phrase refers not only to the perceptual relation between a viewing subject and viewed object but also to the historicity of this relation: the extent to which all visual experiences and representations are culturally and socially produced. The Renaissance 'invention' of geometric perspective, Cartesian notions of rational sight, and various forms of artistic and scientific naturalism, can thus all be analyzed in terms of their historical and ideological implications.

Introduction

The complexity of the term 'landscape' is matched only by the clamor of competing theoretical and methodological positions which have sought to understand it. Deployed variously in cultural geography, art history, landscape architecture, environmental studies, and even the archaeology, sociology, and anthropology of regions, territories, and locations, the term seems to suffer as much from a definitional overload as it does from a nagging ambiguity. However, with its etymological origins in feudal polities of customary law, land control, and communal occupation (*landschaft*), its later adaptation by the seventeenth century Dutch genre painters (*landschap*), and its role within the continental European school of morphological analyses (*landschaftsgeographie*), landscape's defining tension is summarily established: Is landscape the space in which we live, move, and have our being? Or is it the scenic framing of that space for our observation, contemplation, and appreciation? Alternatively, does landscape belong to the domain of utility or esthetics, and of objective spatial formation or pictorial representation? Most importantly, what does such a categorical distinction consist in? What is the historical genealogy of the divide and the weight of its social, cultural, and political significance?

From Cultural Materialism to the Textual Turn

From both a phenomenological and historical materialist perspective, Raymond Williams' seminal definition still resonates: "a working country is hardly ever a landscape."

Indeed, armed with the implications of a post Cartesian philosophy and with a Marxian understanding of mod ernity as coextensive with the development of industrial capitalism, Williams sees the philosophical separation of nature and art as the analog of the distinction between the spheres of production and consumption. Landscape, in this sense, is not only the process and product of certain property and social relations predicated on a commodity economy and its hidden practices of labor exploitation. Landscape is also a particular 'way of seeing' an estheticizing operation which solicits us to withdraw from a site, to 'view' it from afar: a distanciating move inseparable from a retreat to higher and safer ground, a rationalizing perception and a kind of refusal of whatever practical, political, or proximate involvements the scene might make on us. The implication, in short, is to understand landscape as a 'visual ideology'. While as a physical form it is a concrete materialization of social contradictions (a land owned by a privileged economic class but worked and maintained by another), it is also a system of esthetic and conventional ordering useful in the management of such contradictions.

Indeed, if landscape, as the term suggests, involves the cultural shaping or 'scaping' of the natural world and its sensed perception in order to render it politically functional, it has also relied on a host of related generic conventions. Chief among these is the tendency toward mimetic realism which, as Cosgrove notes, in composing space "according to the certainties of geometry" and the optical technologies of perspectival control, makes landscape the site of ideological mediation *par excellence*. So understood, landscape representation (as image, picture, or poetic and rhetorical trope) not only obscures the partiality of subjective vision in order to exhibit the world as an objective property of space, but it also effectively suspends the perceived scene from the flux of time and historical specificity to invest it with the (depoliticized) values of durability and duration. In short, what is at issue in such an account is how landscape as ideology functions to naturalize a scene, how it presents as given and inevitable all those events, actions, and accidents that constitute the 'real' structural dynamics of social relations and historical process. Accordingly, as Matless notes, for cultural materialists any critique of the idea of landscape involves lifting the 'veil' of idealized distortion, as John Barrell and Anne Bermingham do when they locate the 'invention' of the English rustic tradition in the context of the Enclosure Acts and the accompanying dispossession of the English peasantry which facilitated the rise and development of industrial capitalism. Alternatively, to critique the esthetic idealization of landscape, means being alive to the Marxian insight that ideology gains force to the degree to which its referents take material form, or to take seriously the injunction that culture is never more in evidence than whenever nature is announced.

If landscape, as the antithesis of 'land' or its idealized imagining, is crucial to the mystifying operations of capitalism, then, so too is the way in which it is in separable from the production of specific social and subjective identities. Thus, for instance, Barrell draws on a cultural materialist tradition in literary theory to argue that the taste for the panoramic or 'prospect' position so favored in the eighteenth century landscape esthetics is productive of that subject of civic authority who is not only capable to observe or 'frame' a view but who, from his visionary prospect, is able to abstract "substance from accident" and so grasp objects in the "form of their re lations to each other." As such, the landscape panorama does not merely denote a particular viewing position. It is, for Barrell, the very image of political liberalism. As a key move in the exercise of symbolic power, it attests to nothing less than that ideal figure who, possessed of a rationalizing distance, is thus capable of the unfettered mobility and disinterested generosity of true governance. Similarly, from an art historical perspective, Elizabeth Helsinger has argued the ways that the nineteenth cen tury landscape art of J. M. W. Turner does not just image a certain pictorial topography, rather, it represents an interlocking system of circulatory spaces associated with the dissonant class interests of burgeoning middle class tourism, the internal migration of a mobile labor force and a newly embattled hereditary aristocracy. In this, she argues, Turner's English landscapes and the figures that populate them not only negotiate conflicting class interests but also reflect back, in turn, on the very claims to political representativity and national identity in the public sphere of the nineteenth century Britain. Over the past three decades, cultural geographers have sig nificantly elaborated the connection between class an alysis and its imaginative cultural expression. For Denis Cosgrove, tracing the genealogy of the "landscape idea" from its original emergence in a Renaissance imaginary to its later Enlightened, Industrial, and Modern articu lations, landscape "represents a way in which classes of people have signified themselves and their world through their imagined relationships with nature, and through which they have underlined and communicated their own social role and that of others with respect to external nature." Similarly, for Stephen Daniels examining land scape designs and debates in Georgian and Regency England, the 'duplicity' of the genre resides in the extent to which it enfolds an individualized esthetic sensibility, on the one hand, and a carefully historicized class dis course, on the other. As Daniels sees it, the extent to which landscape, as a cultural image, compels the fiction of a sensual, organic unity with nature, is precisely the extent to which it conceals both the underlying material conditions of existence and adjudicates the separation that capitalist appropriation has brought between the human and natural worlds.

For both Cosgrove and Daniels, however, an analysis aimed at reading what Anne Bermingham calls the "class view of landscape" is ineffective, and reductionist, unless linked to the self fashioning of the subject as the projector and receptor of that view. Indeed, the implicit understanding of cultural materialist analysis is that the ideological force of landscape, as well as of the various generic and narrative typologies in which it may be represented (e.g., panoramic, pastoral, georgic, the sublime, and the picturesque), is not to be had simply by reading it as the product or record of the social relations that underwrite it. Instead, our attention is called to the kinds of spectatorial work, or the performative self positioning, evoked by a particular landscape site/sight at a particular historical moment. Landscape, in this sense, should not merely be placed in its correct periodized context in order to simply appreciate how a body of determinate signs may signify or reflect certain structures of cultural power, rather, landscape should be seen as itself a 'cultural practice', one that "makes history in both the real and represented environment," playing a formative role in the active production of subjects in both their material and textual engagement with the social world.

By the late 1980s and early 1990s, the growing appreciation of landscape as a cultural and textual practice – or the "[changing] of 'landscape' from a noun to a verb" in W. J. T. Mitchell's nifty phrase – began to take important lessons from the interpretative techniques of literary and cultural theory. In this, it further refined an earlier materialist critique by asking questions about readership, authorship, and the process of signification where such things refer to uncovering the hidden codes and unquestioned assumptions which structure how a landscape is communicated, experienced, and explored. While the move from a dominantly cultural materialist to textual 'turn' in geography is impossible to clearly discern – the recursive flow of intellectual movements being, of course, resistant to strict chronology or unprecedented insights – this interpretation of landscape was to encourage an even greater awareness of its discursive effects. That is, the extent to which all landscapes – whether material or imagined, physical or thought, the 'world itself' or as its mimetic duplication – are constructed within a realm of signs and their unstable meanings. Rehearsed most fully in the anthologies of Barnes and Duncan's book, *Writing Worlds,* and Duncan and Ley's book, *Place/Culture/Representation*, the key epistemological shift in these accounts was to insist that all landscapes as texts be understood as a system of culturally encoded signs. Following the post structuralist analyses of Roland Barthes, landscapes could now be seen as 'intertextual'; that is, woven through and cross hatched by a limitless force field of "citations, references, echoes, cultural languages, antecedent or contemporary," which together construct a ceaseless ensemble and thus a necessary deferment of meaning. What intertextuality pronounces, in short, is not simply that landscapes are representations – the graphic, pictorial, or rhetorical reproduction of a real world 'out there' to which they refer. At its furthest reaches, a landscape is a "representation of something that is already a representation" (Mitchell, 2002: 14). Long before it becomes the subject of a picture, it participates in a world in which cultural meanings and values are already inscribed and which we behold with an eye that is already implicated in, and mediated by, an endless field of discursive practice. We might interpret René Magritte's *La Conditione Humaine* in just this way (**Figure 1**). "This is how we see the world," said the Surrealist artist Magritte of his work in which an easel painting of a landscape has been grafted onto the outside view it depicts so that the two are doubled and made indistinguishable. Indeed, in a painting in which representation as artifice is set against and folded into a representation as natural fact, Magritte seems to speak directly to the intertextual idea that what the optical eye sees in a physical landscape – its empirical, phenomenal data – is already presupposed by an entire network of discursive determinations which enable its visual recognition. In short, here epistemology (what we know of the world) and ontology (the world itself) are neatly collapsed and conflated: the view outside the window being

Figure 1 René Magritte's *La Conditione Humaine*.

constituted through the painted image of it and by an eye that recursively reads, and rereads, one entwined in the other.

Critical Landscapes: Political Economy and Gender

If landscape names one of the ways in which spatial practices connected to deeper philosophical and historical movements, it also carries a particular political charge. The cultural materialism of Cosgrove and Daniels represents an expansive version of this critical acknowledgment, namely, the centrality of landscape esthetics not only to existing social processes but also to the complex operations of the economy. A more directly politicized version, offered by recalling an older foundational Marxism, however, has sought to commit landscape to a model of social existence proceeding upward from material life and thus to a theory of culture based on the determining force of the economy centered on relations of production and their associated class interests. The cultural geographer Don Mitchell, elaborating on the work of other radical North American geographers such as Smith and Zukin, offers what is arguably the most systematically materialist interpretation of landscape processes, especially in context of contemporary American land and labor systems. Arguing against the analyses of landscape as pictorial or textual medium – indeed, seeing such a culturalist interpretation as its own form of ideological reification, abstracting the codes of landscape's consumption from the actualities of its productive processes – Mitchell's perspective thus returns landscape to classical Marxism's first order priority; insisting that it is the underlying economic structure which – ontologically, epistemologically, and analytically – must form the basis for any landscape study. It is 'here' that the systemic conflict and violence of capitalist relations are centered – and, more importantly, 'here' that they may be contested. Mitchell's work on California's agricultural landscape thus focuses particularly on exploitative migrant labor practices, the breakdown of familial and community networks caused by a seasonally and permanently dislocated workforce, the obstacles created to disallow any effective labor organization, and the extent to which the enormous profitability of Californian agriculture and the affluence of its suburban expression directly depends on the manifest alienation of the laboring body. While Mitchell's concern is with the empirical specificities of the North American context, the objectivist import of his Marxian perspective, as well as his attempt to locate landscape study within a general societal analysis, means that the scope of his work relays to wider comparative networks and to questions raised by a longer genealogy of politically progressive scholarship. Thus his work on landscape has much to say, for example, to David Harvey's materialist geographies of justice and difference, political ecology, theories of uneven development, and not least, to certain analyses of newly globalized social order. More than this, Mitchell's understanding of landscape "constituted through a complex geography of violence" means that it reads, first and foremost, as call for a return to politics, conceived by him as form of intellectual labor embedded within the bigger picture of a liberatory materialism.

Writing the relationship of landscape to politics is not, however, the preserve of Marxist cultural geography. Scholarly debates of geographers are inseparable from politics in the widest sense of the term. Thus, all the partially visible philosophical, historical, and strictly political baggage geographers bring with them to an academic arena; the wider contentiousness implied by their position taking; the broader experiential dynamics of place, location, and the body that inevitably shadow their concerns, have together helped frame the study of landscape over recent decades. Feminist geography has certainly supplied one of the most far reaching accounts of landscape, in this regard. At one level, feminist geography concerned with the idea of landscape registers as one instance in the project of working through, and against, the binarized structure of a dominant cultural and geographic imagination – namely, the division between spaces of family, domesticity, and intimate life, on the one hand, and the performativities of a public and professional world, on the other. Related to this, feminist geography have been particularly concerned to reveal the material gendering of the industrialized urban landscape itself, which neither the available scripts of Marxism nor other materialist sociologies have adequately managed to address. Here, concerns ranging from the gendering of historical memory instantiated by the masculinist codes of public memorials and the differential spatiality of the workplace, to patterns of consumption and practices of mobility have all recognized the gendered nature of subjectivity and space while working hard to avoid the appropriation of an essentialist identity that the 'denaturing' charge of feminism is, after all, designed to attack.

It is in the engagement with visual observation and representation, however, that feminist geography has offered a meditation on 'landscape' in precisely the disciplinary sense of the term; that is, as a 'way of seeing', a process of cognition, a genre of picturing making, and, not least, as a cultural practice constituted around the epistemic equivalence of the 'male' and the 'gaze'. Informed by a generalized critique of the discursive transcoding of the male viewer as the bearer of reason/culture and woman as pacified body/nature, and further refined by psychoanalytic analyses of visual pleasure and the repression of desire, feminist geographers have worked to untie the Gordian knot between landscape, vision, and the body, while, at the same time, have exposed the gendered construction of the 'landscape idea' within the predominantly male legacy and logic of academic geography. Most productively, writers like Nash have sought to retain the "erotics, pleasure and power" of landscape esthetics, while reclaiming it for a feminist cultural geography that opens up a range of viewing positions and subjective positionalities that play on, but refuse the fixity of, normative gender identifications. From a feminist perspective, in short, landscape not only names the ability of a given social and cultural environment to consign some types of subjectivity to the margin, but also authorizes the alternative possibility of revisioning nature in terms of more complex, varied, and less determined lines of sight.

Other Landscapes: From the Nation-State to the Colonial Periphery

Landscape, however, is not just the site where we recognize our historical, cultural, and perceptual investment in space, but also where we redefine it. A powerful perspective developed through the 1990s was to deploy landscape not only as a critical tool useful in the analysis of culture and society but also as an active site of identity formation within both the 'invented tradition' of the nation state as well the expanded field of its colonial and imperial articulations.

Energized by the ferment of critical and philosophical ideas produced from the late 1960s onward by the post structuralist movement represented by Michel Foucault and Jacques Derrida, and further invigorated by the proliferation of postcolonial discourse analyses in the 1980s, the concepts of identity, language, space, and power breathed new life into understandings of how subjectivity might be thought and performed via specific landscape practices. Here, a concern with individual places, histories, texts, and esthetic genres comes to be elaborated alongside wider discursive and epistemological questions, not least the interrogation of landscape itself as an exemplary site where systems of European knowledge, politics, ethics, and esthetics are formed.

Informed, albeit implicitly, by key Foucauldian ideas of disciplinary and regulatory discourses embedded in particular spatial actions and activities, David Matless, for example, explores the connections between landscape, identity, and citizenship in the context of mid twentieth century England, focusing particularly on those embodied practices (e.g., naturism, organicism, planning, the Boy Scouts, hygiene, and health and fitness movements) through which the entwined languages of landscape and the self may be seen to work. Importantly, to see 'landscape and Englishness' in this way – that is, how the conduct of a particular historicized citizen subject is produced via the performance of particular landscape practices – means that neither term has recourse to any pregiven or original quality outside of its changing discursive articulation. Locating much of his work in the inter and postwar period, Matless is concerned to show how the bodily display of certain 'authentic' landscape practices – instantiated, for example, by the rustic 'organic' movement of the 1930s, or alternatively, by the figure of the planner preservationist "walking, cycling, camping, map reading" – is itself the means "whereby individual and nation might give form to itself environmentally," generating particular codes of moral and esthetic conduct necessary to a particular version of a proper, participatory, or 'landscaped' citizen.

To the extent that an account, like Matless', displays a dedication to detailed archival work and understands the English landscape, like the English subject, as the effect of historically specific spatial practices, the influence of Foucault is evident. However, it is perhaps only when one leaps the European fence to consider how landscape plays itself out beyond the metropolitan center, that a Foucauldian inflected account becomes most explicit. This is due in no small measure to the landmark work of Edward Said, which sought to recast a Foucauldian account of knowledge production into an analysis of European colonialism and imperialism. Thus, if at one end of the spectrum, the modern West has used landscapes to weave together regimes of value, morality, perception, and social power, at the other end, the history of military, administrative, territorial, and cultural expansion has placed landscape at the center of a postcolonial critique.

Indeed, we do not have to agree with the starkness of the assertion that "landscape is the [groundwork] and dreamwork of imperialism" (Mitchell, 1994: 8) to argue that looking at representation in the colonies is, perhaps, to have "privileged insight into what is most resilient, most dominant, and at the same time most politically constraining in the European landscape tradition" (Bunn, 1994: 128). Otherwise put, if we accept that landscape, as a way of seeing, depends on the controlling relation of sight to space, it is a relation that is necessarily complicated, under imperial conditions, because of the association between sight and surveillance. Allied to the critique of the older Enlightenment project of linking reason to illumination and sight to an ideal universal objectivity, much postcolonial writing has thus been concerned with how a certain surveillant gaze is implicit in the ways that European travelers, settlers, administrators, mapmakers, and artists attempted to comprehend a foreign landscape: evaluating its productive potential, classifying and categorizing its alien flora and fauna, inhabiting its spaces, projecting a future – and appropriating it for the European self.

Numerous geographers have taken up Said's Foucaldian inspired injunction to understand the colonial landscape as an 'imaginative geography', in particular through interrogating the discourses of exploration and scientific fieldwork, travel, cartography and navigation, pictorial convention, and, not least, geography itself as an institutional and intellectual project historically coincident with the imperial enterprise. At the same time, however, geographers have been careful to understand the colonial landscape not merely as an empty repository for the admission of European visions, or as a *tabula rasa* that appears willing to be named and claimed by the presence of Western knowledge systems. Colonial space, is after all the site of radical difference, a 'contact zone' filled with competing indigenous meanings and fractured by local contests over material and symbolic power. It is, in this sense, a "site of regular ontological shock" (Bunn, 1994: 129). It is a space whose topographical strangeness not only has to be rendered familiar or coded according to the tastes and conventions of a European landscape idiom (e.g., the pictorial conventions of the picturesque, the sublime, and the exotic). It is also a space in which the figure of the European confronts and negotiates its encounter with non European others. Here, Derrida's 'deconstruction' of the binary schema that underpins Western analytic distinctions and the metaphysical certainties and closures, preferences and repressions it enables, has been particularly useful to geographers concerned with destabilizing the hierarchical oppositions of West/East, self/other, center/periphery.

In short, while some cultural geographers have focused on the extent to which colonial landscapes are made legible and legitimate by the staging of various scientific, esthetic, and visual impositions, others have pushed the presumptions of Empire to its limits, pointing to the incoherences, absences, and instabilities immanent in the ideological and institutional system of imperialism itself. Thus, W. J. T. Mitchell, for example, has called for analyses of colonial landscapes that take account of "fractured images of unresolved ambivalence and unsuppressed resistance", and Dubow, among others, has used a phenomenological account in order to recover the somatic, psychic, and existential aspects of a colonial view prior to the ascendant disinterest of an imperial gaze. In such cases, the point is to understand an alien landscape not only as a site of cultural exchange or esthetic transformation but also one in which the bodily and perceptual particulars of the viewing body may be scrutinized for its moments of desire, disarray, and breakdown – and so to trouble the very epistemology of the sovereign subject written into European knowledge, theory, and history.

From a 'Way of Seeing' to a Bodily Encounter

If analyses of imperial and colonial landscapes have emphasized the effects of power, subjectivity, representation, and visuality, while also pointing to the impossibilities and failures of such abstractions, then a more recent development in cultural geography seeks to reinsert landscape into the more empirical, embodied, and everyday aspects of how it is practiced and experienced, moved through, and lived in. Although not as far removed from a post structuralist perspective as is sometimes supposed, nor indeed unrelated to a longer tradition of landscape writing itself (i.e., the vernacular landscapes of J. B. Jackson in the 1950s and 1960s), a return to the earlier twentieth century philosophical tradition of phenomenology has nonetheless opened the study of landscape to a range of reinvigorated agendas. Formally set in motion by a critique of that Cartesian ontology dividing the world between spirit and matter (or mind/body, subject/object, observer/observed) phenomenology is distinguished by a concentration on the primacy of lived experience – and the description of that experience – which reveal the connections through which we are related to the world prior to the work of theoretical analysis or tutored interpretation.

While, like all philosophical tendencies, phenomenology involves a variety of contradictory strains, a key source of inspiration for recent landscape studies has been Martin Heidegger's shift from 'pure' to 'existential' phenomenology with its insistence that human existence and understanding is more than rational cognition in its usual spectatorial or Cartesian sense. It is rather the process of a thoroughly 'worldly' agent in its living and affective engagement with the world – the living, affective engagement with landscape, for example. Indeed, if the persistent priority of Western knowledge and culture maintains the gazing and thinking subject as the source of agency and objective authority, the significance of phenomenology for landscape studies resides in its understanding of the inescapably corporeal nature of human being and of 'being in the world'. Here, the call is to reassert the subject's direct or pre reflective contact with the object world; to assert that it is the living and perceiving body itself that forms the basis or precondition of spatial experience and perception and not the work of a 'mindful' consciousness set over and against it. Accordingly, landscape is no longer rendered as static or mute matter capable of being observed by virtue of a detached distance, nor is its representation a matter of inert pictorialism capable of being 'read' via the codes of an abstract or derivative sign language. In this accounting, landscape is no longer a perspectival projection or a 'way of seeing' at all, nor is it a scene or setting within which a figure or form emerges and becomes culturally meaningful. Before or beyond any ordering in represen tation and epistemology, landscape is a space given along with what Maurice Merleau Ponty calls the primary 'intertwining' or 'enlacing' of body and vision. It is an intimate 'lifeworld' in which view and viewer, observer and observed, are indistinguishable from each other; an ongoing process in which self and space mutually con dition each other such that he/she 'who sees' and that 'which is seen' are immersed in intimate dialog. In this, it is not only the concept of landscape and its relation to vision which gets rethought but, also indissolubly, the 'coming into being' and 'coming into seeing' of the human subject itself.

If, as previously discussed, both a cultural materialist and textual approach to landscape were concerned with unveiling its apparently 'invisible' social and symbolic orderings, then a phenomenological approach is in some senses a return to the foundational question of nature/culture which underpins all landscape study. However, rather than posing the question as a problem of duality – of the separation of the material and the metaphoric, the real and its representation, interiority and exteriority, and the recursive tensions between them – phenomenology seeks to recover an understanding of landscape that precedes the work of such category distinctions. In this, however, it risks certain dangers. Vulnerable to a certain mythopoetic tenor which can invite a Romantic privil eging of the humanist subject, on the one hand, and the illusion of precultural perception, on the other, phe nomenological accounts of landscape have unsurprisingly drawn criticism from materialist, deconstructive, Fou cauldian, Deleuzian, and feminist geographers alike. Here, it is argued that a phenomenalist disavowal of social relations and its arrangements of power, or a reduction of this to the sensuously empirical, is a de politicized move in the characteristic manner of ideo logical thought. As ever, the issue is one of sensitivity and sophistication, of seeing the successive waves of land scape studies and the responses they engender as critical rather than (self) confirmatory interventions. Indeed, it is arguable that, in the most careful of hands, a phenom enological approach does not so much collapse the dis tinction between 'land' and 'scape' but, like the best of sort of radical or rooted critiques, provides a "means of communication between the space of the body and the realm of culture, between the determinations of nature and the weight of history."

Future Views and Other Avenues

With this possibility in mind, several recent trends in landscape study are finding ways of relating the material event to discursive signification and of finding new research agendas for their interplay. The theme of landscape and the politics of memory/memorialization is one such instance in which concepts of nationhood and experiences in identity are pressed close to the qualita tive densities of their historical ground, especially as these relate to the possibility of postcolonial, diasporic, and exilic self fashioning. Similarly, themes of bodily mobility, temporality, and their relation to biographic narrativity have creatively addressed the way that land scape animates the interconnectivity of environment, movement, history, and personal storytelling. In a related vein, the twinned forces of multisensory experience and its expression have come together in accounts of land scape which speak to the co constitution of self and space. Deeply localized, always contingent, empirically rich, and often archivally based, these are treatments in which landscape emerges as the perpetual doing and undoing of operations and objects, itineraries and maps, immediacies and mediations, sensations and ideas, things and thoughts. In this, it declares the affective enactments of both landscapes and their representations. In so doing, it brings into play landscape's less determined particulars – its textures, moods, sounds, pacings, dimensions – and calls for a response in the geographer's own performance – his/her grammars, poetics styles of thoughts, and forms of expression. If nothing else, landscape is precisely the space of this co constitution – and this tension.

See also: Culture; Nature.

Further Reading

Barnes, T. J. and Duncan, J. S. (eds.) (1992). *Writing Worlds*. London: Routledge.

Barnett, C. (1998). Impure and wordly geography: The Africanist discourse of the Royal Geographical Society, 1831 1873. *Transactions of the Institute of British Geographers NS* 23(2), 239 251.

Barrell, J. (1980). *The Dark Side of the Landscape*. Cambridge: Cambridge University Press.

Barrell, J. (1990). The public prospect and the private view: The politics of taste in eighteenth century Britain. In Pugh, S. (ed.) *Reading Landscape: Country City Capital*, pp 19 40. Manchester: Manchester University Press.

Bermingham, A. (1986). *Landscape and Ideology: The English Rustic Tradition, 1940 1860*. Berkeley, CA: University of California Press.

Blunt, A. (1994). *Travel, Gender and Imperialism: Mary Kingsley and West Africa*. New York: Guilford Press.

Bondi, L. (1992). Gender symbols and urban Landscapes. *Progress in Human Geography* 16(2), 157 170.

Bondi, L. (1998). Gender, class and urban space: Private and public space in contemporary urban landscape. *Urban Geography* 19(2), 160 185.

Bryant, R. (2001). Political ecology: A critical agenda for change? In Castree, N. & Braun, B. (eds.) *Social Nature: Theory, Practice and Politics*, pp 151 169. Oxford: Blackwell.

Bunn, D. (1994). Our wattled cot: Mercantile and domestic space in Thomas Pringle's African landscapes. In Mitchell, W. J. T. (ed.) *Landscape and Power* (2002, 2nd edn.), pp 127 73. London: Routledge.

Clayton, D. (2000). *Islands of Truth*. Vancouver, BC: University of British Columbia Press.

Cosgrove, D. (1984). *Social Formation and Symbolic Landscape*. Madison, WI: University of Wisconsin Press.

Cosgrove, D. (1985). Prospect, perspective and the evolution of the landscape idea. *Transactions of the Institute of British Geographers* 10(1), 45 62.

Cosgrove, D. and Daniels, S. (eds.) (1988). *The Iconography of Landscape*. Cambridge: Cambridge University Press.

Daniels, S. (1989). Marxism, culture and the duplicity of landscape. In Peet, R. & Thrift, N. (eds.) *New Models in Geography*, vol. 2, pp 196 220. London: Unwin Hyman.

Daniels, S. (1993). *Fields of Vision: Landscape Imagery and National Identity*. Cambridge: Polity Press.

Derrida, J. (1967). *Writing and Difference*, (transl. Bass A., 1978). London: Routledge & Kegan Paul.

Dewsbury, J. D., Wylie, J., Harrison, P. and Rose, M. (2002). Enacting geographies. *Geoforum* 32, 437 441.

Domosh, M. (1996). *Invented Cities: The Creation of Landscape in Nineteenth Century New York and Boston*. New Haven, CT: Yale University Press.

Driver, F. (2001). *Geography Militant*. Oxford: Blackwell.

Driver, F. and Martins, L. (eds.) (2005). *Tropical Visions in an Age of Empire*. Chicago, IL: University of Chicago Press.

Dubow, J. (2000). From a view on the world to a point of view in it: Rethinking sight, space and the colonial subject. *Interventions: International Journal of Postcolonial Studies* 2(1), 87 102.

Dubow, J. (2001). Rites of passage: Travel and the materiality of vision at the Cape of Good Hope. In Bender, B. & Winer, M. (eds.) *Contested Landscapes: Movement, Exile and Place*, pp 241 255. Oxford: Berg.

Duncan, J. and Ley, D. (1993). *Place/Representation/Culture*. London: Routledge.

Foucault, M. (1973). *The Order of Things: An Archaeology of the Human Sciences*. New York: Vintage.

Gregory, D. (1994). *Geographical Imaginations*. Oxford: Basil Blackwell.

Gregory, D. (1995). Between the book and the lamp: Imaginative geographies of Egypt 1849 1850. *Transactions of the Institute of British Geographers* 20, 29 57.

Harvey, D. (1996). *Justice, Nature and the Geography of Difference*. Oxford: Blackwell.

Helsinger, E. (1994). Turner and the representation of England. In Mitchell, W. J. T. (ed.) *Landscape and Power*, (2nd edn., 2002), pp 103 125. London: Routledge.

Kearns, G. (1997). The imperial subject: Geography and travel in the work of Mary Kingsley and Halford Mackinder. *Transactions of the Institute of British Geographers* 22, 450 472.

Lorimer, H. (2003). Telling small stories: Spaces of knowledge and the practice of geography. *Transactions of the Institute of British Geographers* 28, 197 218.

Lorimer, H. (2006). Herding memories of humans and animals. *Environment and Planning D: Society and Space* 24, 297 518.

Matless, D. (1989). *Landscape and Englishness*. London: Reaktion.

Matless, D. (1992). An occasion for geography: Landscape, representation and Foucault's corpus. *Environment and Planning D: Society and Space* 10, 41 56.

Mitchell, D. (1994). Landscape and surplus value: The making of the ordinary in Brentwood, C. A. *Environment and Planning D: Society and Space* 12, 7 30.

Mitchell, D. (1996). *The Lie of the Land: Migrant Workers and the Californian Landscape*. Minneapolis, MN: University of Minnesota Press.

Mitchell, W. J. T. (ed.) (1994). *Landscape and Power* (2nd edn.). London: Routledge.

Nash, C. (1996). Reclaiming vision: Looking and landscape and the body. *Gender, Place and Culture* 19, 227 247.

Pratt, M. L. (1992). *Imperial Eyes: Travel Literature and Transculturation*. London: Routledge.

Pugh, S. (ed.) (1990). *Reading Landscape: City Country Capital*. Chicago, IL: University of Chicago Press.

Rose, G. (1993). *Feminism and Geography*. Cambridge: Polity Press.

Rose, M. and Wylie, J. (2006). Animating landscape. *Environment and Planning D: Society and Space* 24, 475 479.

Said, E. (1978). *Orientalism: Western Conceptions of the Orient*. London: Penguin.

Said, E. (1993). *Culture and Imperialism*. London: Penguin.

Smith, N. (1990). *Uneven Development: Nature, Capital and the Production of Space*. Oxford: Basil Blackwell.

Williams, R. (1985). *The Country and the City*. London: Chatto and Windus.

Zukin, S. (1991). *Landscapes of Power: From Detroit to Disney World*. Berkeley, CA: University of California Press.

Landscape Iconography

S. Hoelscher, The University of Texas at Austin, Austin, TX, USA

Glossary

Iconography The description and interpretation of images or symbols in an attempt to explore social relations, cultural meaning, and political–economic power.

Image A two- or three-dimensional artifact that depicts or refers to something by having a similar appearance.

Intertextuality A concept derived originally from poststructural literary criticism that refers to the ways in which any given text or visual representation derives its meanings by other texts.

Representation The practices – both textual and visual – that create, communicate, and circulate meanings.

Visual Culture An interdisciplinary field of study that examines the features of culture that rely on visual images.

A key methodology within contemporary cultural and historical geography, landscape iconography seeks to describe and interpret visual images of the world. It treats such representations as social constructions and situates them in their cultural, historical, and political contexts with the intention of revealing what are often hidden and symbolic meanings. The method, first developed in the field of art history, initially focused on painting, but has since broadened dramatically to investigate expressive media such as photography, maps, and film, as well as the built landscape itself. In every case, iconographic interpretation demands a close formal reading of such cultural forms, a thorough understanding of the contexts surrounding their creation and subsequent circulation, and a comprehensive knowledge of their impact, which often embodies relations of power.

Introduction

The landscape iconography method in human geography traces its origins to the 1980s, when cultural and historical geographers began searching for new ways to investigate the humanly constructed world. Informed by Marxist cultural theory and the humanistic geography of the 1970s, these scholars had grown increasingly weary of what they saw to be the discipline's over reliance on a morphological approach to landscape studies. Such an approach – first developed by Carl Sauer and the Berkeley School – understood landscape purely as a physical entity, as the sum of its morphological components: all of an area's buildings, population, and infrastructure. This 'cultural landscape' could be mapped to illustrate the distribution of certain landscape features such as types of houses, and to demonstrate the directional flow of migrating cultural groups. The end result, for this kind of landscape study, was to show the 'impress of the works of man upon an area'.

By contrast, for several influential geographers writing in the 1980s, landscape was explored for its symbolic and ideological qualities. The concern for geographers like Stephen Daniels and Denis Cosgrove was to show how landscape is a "cultural image, a pictorial way of representing, structuring, or symbolizing surroundings". Landscape study, from this perspective, is less about the measurable features on the surface of the earth and more about landscape's symbolic character, especially as they appear in mediated representations. Furthermore, this perspective highlights a landscape's ideological associations, taking careful note of embedded power relations. Cultural landscapes, as explored from an iconographic methodology, are representations that inevitably serve someone's interest and, in this way, are never neutral or value free.

The result is a radically new and innovative way of investigating that most traditional of geographical concepts – the surface of the earth, or cultural landscape. Cosgrove and Daniels, in their 1988 influential edited volume that formalized the iconographic approach to landscape, urged geographers to consider all cultural landscapes as representations, whether that representation consisted of paint on a canvas, of ink on a map, or of the earth, stone, water, and vegetation on the ground. A monument to fallen soldiers in a city park might be more tangible than a landscape painting or photograph, but it is no more real or less imaginary. By thus broadening the definitions of both landscape and representation, the iconographic approach demands that its practitioners search beneath the surface of a painting, photograph, or urban park to lay bare the conditions and ideologies that went into the landscape's making. Such an approach does not deny the vast differences between types of representations – the social conditions that go into making a poem and a hydroelectric dam, of course, are quite different – but that, in every case, symbolic landscapes are understood to be storehouses of cultural meaning.

With cultural meaning comes politics. An iconographic approach to landscape recognizes that there is a politics to representation and that unearthing the symbolic qualities to landscape often involves exposing the power relations that helped create it. While the earliest studies of landscape iconography focused on the influence of class relations in creating landscape repre sentations, more recent work has examined an ever expanding set of interlocking identities, including: nationalism, colonialism, race, gender, and sexuality. The fruits of this work are geographical investigations that are thoroughly interdisciplinary in scope, transnational in scale, and evocative of their subject's complexity.

Borrowing from Art History: Toward an Iconographic Method

With few available models to show how a geographer might investigate the symbolic imagery of a landscape, early practitioners turned to the well developed field of art history. The work of German art historians Aby Warburg, Ernst Gombrich, and Erwin Panofsky ap pealed to geographers like Denis Cosgrove, who strove to interpret visual images emerging from Italian Re naissance. Erwin Panofsky, in particular, proved espe cially influential as he laid out a clear articulation of an iconographic method concerned with the "subject matter or meaning of works of art, as opposed to their form." Such a formulation proved useful for geographers seeking to interpret a landscape's multidimensional and contextualized meanings, instead of its morphological patterns.

For Panofsky, meaning or subject matter could only be understood by illuminating the belief systems and social milieu that first gave rise to the artwork. By emphasizing the contextual interpretation of pictorial symbols, Panofsky's was a project of historical recovery. In both his theoretical statements and in his own empirical work, he argued for assembling evidence about the symbols and signs that contemporary audiences would have known and used when viewing a painting. Such a prac tice, though painstaking and inevitably partial, proved necessary if a scholar was to lessen the distance between herself and her subject in the past.

In a series of oft cited essays that date to the middle decades of the twentieth century, Panofsky spelled out his method of interpreting visual images. Like an archaeologist digging through the earth's layers, he urged art historians to uncover the deep structures of cultural products. The first or primary level of visual interpret ation, though important, serves mainly to identity what is being depicted; clear observation and description are highly valued in this so called pre iconographic level. The secondary, or iconographic, level of interpretation recognizes that what is being depicted stands for some thing else, that it symbolizes something; hence, the Statue of Liberty is said to signify the American nation, or convention has it that a lamb in a Renaissance painting indicates Christ. However, it is only in the third or ico nological level of interpretation that scholars make use of interpretation to explore a visual image's more general cultural significance. Here, at the deepest level of 'in trinsic meaning', the artwork is shown to be governed by "underlying principles which reveal the basic attitude" of a wide range of social forces, from the nation and social class to religion and philosophical perspective.

Most geographers, in practice, conflate Panofsky's three levels in their writing, referring to their method of landscape interpretation simply as 'iconographic'. Nonetheless, the best of this scholarship implicitly rec ognizes the importance of thick, detailed description of what a scene depicts; it acknowledges the necessity of a systematic grounding in the historical context that gave rise of the image's symbolic codes; and it insists that this historical grounding extends to all matters of society and culture. Such a historically grounded method is inter textual, in that it depends on understanding the diverse types of historical sources that inform this artwork's creation. It is also interdisciplinary, in that it relies on the insights from complementary fields of knowledge.

Panofskian Iconography: An Example from American Landscape Painting

For an application of this method, we might look at a series of well known paintings by the American land scape artist, Thomas Cole (1801–48). Widely regarded as the founding father of a national school of American art, Cole was preoccupied with the question of how best to represent the values and ideals of the new republic, a country seemingly shackled to the Old World and bereft of its own culture and history. For Cole, as for con temporary writers like Washington Irving and James Fenimore Cooper, landscape served not just as a setting for historical events, but as the principal means to express social and moral issues. An iconographic approach to Thomas Cole's paintings would strive to unearth pre cisely those social and moral issues that found expression not only in canvas and paint, but also in the minds and attitudes of those who viewed his paintings in their ori ginal context. Stephen Daniel's discussion of landscape imagery and national identity in England and the United States offers one such example.

In preparing for his discussion of Thomas Cole's paintings, Daniels mines a wealth of primary historical sources that demonstrate both their contemporary re ception as well as the various influences on the painter, from political philosophy and economic pressures to patronage and other artists. Chief among these sources are Cole's own statements on national landscape and history in the form of letters, diaries, and public lectures. Daniels finds that in paintings like Cole's, trees, water, and mountains acquire powerful symbolic qualities and each is made to stand in for a range of meanings or 'associations' that viewing audiences would have immediately understood. Daniels then brings this rich historical context to bear on Cole's epic, five part painting cycle of *The Course of Empire*, completed between 1833 and 1836.

A firm advocate of a cyclical theory of civilization in which wealth and political liberty inevitably give rise to avarice and corruption, Cole devotes each painting to a stage in that cycle, all set in the same valley: *the Savage State*; *the Arcadian or Pastoral State*; *the Consummation of Empire*; *Destruction*; *Desolation*. That valley appears to be a scene from classical antiquity, especially in *the Pastoral State* (**Figure 1a**). In this scene, as in all five panels, a serrated escarpment crowned with a boulder rises over the sea and serves as a symbol of natural permanence amid relentless cultural change. The importance of the escarpment becomes apparent only in the third and central canvas, *Consummation* (**Figure 1b**), where it stands as virtually the only reminder of civilization's natural origins. Gone is both the close connection to nature and benign social relations, replaced now by abstract reason and power. The natural water inlet of the earlier panel has been diverted into ornate, artificial water fountains. Elaborate classical architecture hems in the busy harbor, in which Cole wanted his audiences to see an "a la mode New York." Cole foreshadows this pompous and corrupt scene with the patently American symbol of expansion, the right foregrounded tree stump of the *Pastoral State*. The following two panels, *Destruction* and *Desolation*, vividly point to the decline of empire with violence and devastation.

The series, Daniels notes, was an enormously ambitious undertaking for any American artist in the 1830s and resulted in a huge success for both Cole and American landscape painting more generally. Ironically, the public success of Cole's historical series was premised on misunderstanding – both unintentional and willful. As Daniels makes clear, Cole's underlying message was one of caution. Throughout his career, Cole remained ambivalent about the true nature of empire and the prospect of unfettered development. At the level of Panofsky's intrinsic meaning, these paintings represented a cautionary tale and one of inevitable imperial decline. Nevertheless, the governing elites who banked their fortunes on unlimited, upward progress saw the United States only in the first two panels, but resisted any suggestion that *Consumption*, much less *Destruction* or *Desolation*, represented speculation about America's future.

In this application of Panofsky's iconographic method, we find an approach that bears striking similarities to the discourse analysis and 'archeology' as sketched by Michel Foucault. Although there are vast differences between them – most notably in Panofsky's view that his method would reveal "essential tendencies of the human mind," a position that Foucault would find untenable – Panofsky's iconography approaches Foucault's intent in at least two ways. Both pinned considerable attention on the intertextual relationship between variously circulating texts and images. More importantly, both scholars demonstrated the need to dig below surface appearances to find deeper truths about social conditions. Panofsky's pursuit of 'intrinsic meaning', like Foucault's use of the term 'archeology', demands that students of visual culture go beyond superficial readings of the scene before them to lay bare the relationships and knowledge that construct social and geographic realities. From such a perspective, it becomes clear that Thomas Cole's real subject of *The Course of Empire* was American economic and political anxiety during the crucial nation forming decade of the 1830s – a point frequently denied by viewing audiences but one that is abundantly clear from a close iconographic analysis.

Landscape as a Way of Seeing

As decidedly American as Cole's painting cycle surely is, it also derives much of its inspiration from previously established European techniques, including those developed during the Italian Renaissance. For Erwin Panofsky, this was a time of intense productivity in the visual arts, when elaborate compendia of symbols were written for artists and patrons and which could then be decoded by trained art historians. For geographers like Denis Cosgrove, the period meant something else entirely. This was the time and place, along with northern Europe's Low Countries, where an emergent capitalism put its stamp on how the world was constructed, both materially and visually. Here, in the fifteenth and sixteenth century Venetian republic, a nascent urban merchant class began deploying new theories of space and geometry – most notably, linear or three dimensional perspective – and improved surveying techniques to more accurately and convincingly map their land holdings. New styles of painting also emerged from this rational and mathematical ordering of perspective, with the result being an entirely new way of depicting the world.

This historically specific 'landscape way of seeing' became an important tool for the appropriation of space, and thus a key concern for geographers deploying a landscape iconography methodology. In his now classic account of Italian landscapes, Cosgrove shows how this visual ideology was available solely to one group: the urban merchant class that commissioned the paintings, surveys, and mappings. As depicted through realistically appearing paintings, a landscape way of seeing became a powerful method to take control of potentially disputed land.

(a)

(b)

Figure 1 (a, b) Thomas Cole, *The Course of Empire*: *The Arcadian or Pastoral State* 39 × 63 inches, 1834; and *The Course of Empire: The Consummation of Empire*, 51 × 76 inches, 1835 36, oil on canvas. By the New York Historical Society. Source: www.wikipedia.org

Understood in this way, the iconography of a painted landscape can be closely entwined with what it represents on the ground. Indeed, it is the very relationship between expressive media like paintings and the material world that demands close scrutiny and gives this research its political urgency. Representations of landscapes through Renaissance painting and mapping justified to governing elites the seemingly inherent correctness – the natural ness and taken for grantedness – of their claims to property ownership. Furthermore, such a conception of land implies detachment and an 'outsider's' way of knowing. It follows that peasants and other workers who shaped the physical land are typically rendered invisible and erased from the view, lending the countryside an appearance of being unworked, at precisely the time of massive economic change. As Cosgrove's work clearly demonstrates, the resulting consolidation of power and property at this and other times of dramatic social transformation relied heavily on the visual appearance of landscapes.

Critical Themes in Contemporary Iconographic Research

Contemporary scholars who deploy the landscape iconography method have extended these insights into a wide range of geographical studies. The intensive nature of research necessary for successful iconographic interpretation lends itself to a case study approach; depth of analysis is generally valued over breadth of historical and geographic coverage. Moreover, general recognition that landscape meanings are inherently unstable across time and between groups, and that they are always open to negotiation and contestation, means that an exceptionally broad focus risks generating unsupportable generalizations. In this way, the landscape iconography method bears important similarities to an empirically grounded ethnography.

Although geographers continue to examine paintings as objects of iconographic analysis, the forms of expressive media available for iconographic interpretation have expanded well beyond this traditional medium. Virtually any component of visual culture may be probed for its competing symbolic meanings and, with this in mind, geographers have fruitfully explored film, television, sculpture, advertising, and comic strips. In every case, it is recognized that symbolic images do more than simply represent an external reality; rather, they are potent instruments in shaping that very reality. Among these, three stand out for their relevance for geographical research: maps, photographs, and the built landscape.

Maps

Maps have long been central to the practice of geography, of course, but only in recent years have they been subjected to the similar sort of iconographic analysis that we have seen in landscape painting. The traditional view of maps holds that they are transparent windows on the world, essentially objective visual depictions that stand above the messy realm of politics. Whatever inaccuracies that may exist, this view holds, are ones that can be explained simply by an earlier period's technological backwardness. Potent as this traditional view of cartography may be, it has been challenged by historical geographers, most notably J.B. Harley, who have shown how maps are better understood as social constructions that should be interrogated for both their intrinsic meanings and their ability to create knowledges designed to serve particular interests. Cartography, he argues, belongs to the terrain of the social world in which it is produced and should be approached iconographically.

Because maps operate behind a mask of seemingly neutral science – what Harley calls the map's sly rhetoric of neutrality – they are especially power visual media that hide and deny their social dimensions at the same time that they legitimate those very social relations. This is an especially relevant insight when viewing maps that have historical associations with state formation, capitalism, and imperialism, such as **Figure 2**. This map, compiled by John Bartholomew sometime in the 1850s, depicts the *British Empire Throughout the World, Exhibited in One View*, and is indicative of the way that "maps were used to legitimate the reality of conquest and empire." In order to understand a map like this, Harley suggests, it is imperative to attend closely its various iconographic elements including: stereotypical images of the empire's inhabitants, with *Natives of British Empire in Europe* at its center and increasingly 'exotic' peoples at its periphery; lettering and coloring that present the British possessions engraved "in a bolder character and colored Red;" and quantitative population data in the margins that would seem to demonstrate the immensity and legality of its holdings.

Photographs

Photographs operate in much the same way and recognizing the unique properties of this – and all – expressive media is fundamental to the iconographic method. Take, for example, this 1938 image of an isolated farm in the Texas panhandle by the American photographer, Dorothea Lange (**Figure 3**). At one level, it offers evidence of the sort of vernacular structures and agricultural patterns that have long interested cultural geographers. Pushing a bit further, a geographer might take notice of the abandonment of the dwelling and the particularly neat rows of contour plowed land surrounding it. Such descriptions, as important as they are, frequently neglect the fact that Lange's photograph of Texas is … a photograph, a representation of a three dimensional reality.

An iconographic approach to this image would thus take its objectness as a starting point. It would acknowledge that Lange created this image with a specific agenda in mind, that it served her and other's interests in competing ways, and that, over the years, it has been viewed in multiple contexts by different viewerships. As a key member of the President Franklin Roosevelt's Farm Security Administration (FSA), Lange was commissioned to photographically document the social and economic relationships of American agricultural labor. She became a severe critic of that system, using her photographs to expose its structural inequalities. This becomes evident when the photograph's full caption, as Lange intended it, it matched to the image itself: "Tractored out: Power farming displaces tenants from the land in the western

Figure 2 John Bartholomew *Map of the British Empire Throughout the World, Exhibited in One View*, 1:60 000 000, original size 26 × 49 cm, *c.* 1850s. The Library of Congress, Geography and Map Division, Call Number: G5730 185- .B3 TIL.

dry cotton area, Childress County, Texas Panhandle. June, 1938." Far from a value neutral picture of a Texas landscape, Lange's image simmers with indignation and moral conviction.

Built Environment

Just as a painting, a map, or a photograph may serve as a symbolic representation, so too can the built environment itself be examined for its iconographic properties. Landscapes throughout the world and through time have been constructed with specific ideals in mind, often by ideologically driven governing elites. By examining the individuals directly responsible for real estate development, by delving into the social world of those individuals, and by closely attending to the broad cultural and economic conditions, geographers like Mona Domosh have showed how, in the nineteenth century, cities like New York and Boston achieved their very distinctive look. Representing very different visions of urban life and culture, New York's dramatic skyscrapers and Boston's elite Back Bay became fundamental components of the social, economic, and spatial structuring of each city. Even more dramatic is Vienna's nineteenth century Ringstraße, a linear boulevard that encircles the Austrian capital's medieval core. Constructed largely by a newly empowered liberal government eager to display both its power and bourgeois values of enlightened reason and high culture, the Ringstraße encodes each of its major public structures with liberalism's value structure. Each building was constructed in the historical style considered to be most appropriate to those particular values: thus, the Parliament building in the classical Greek, the Rathaus in massive Gothic, the Burgtheater in Baroque style, and the University as a Renaissance structure (**Figure 4**).

Not all built environments are designed and constructed with such a clearly articulated representational system, of course. Nevertheless, even the most ordinary, everyday, and taken for granted landscapes carry

Figure 3 Dorothea Lange, "*Tractored Out*": *Power Farming Displaces Tenants from the Land in the Western Dry Cotton Area*, *Childress County*, *Texas Panhandle*. Silver nitrate negative, 4 × 5 inches. The Library of Congress, Farm Security Administration, Office of War Information Photograph Collection, LC-DIG-ppmsc-00232 DLC.

Figure 4 Unknown Photographer, *Vienna's Ringstraße*, *c*. 1890. The Perry-Castaneda Library, University of Texas at Austin. Source: "Wien: ein Spaziergang durch die Stadt und ihre Umgebung in 159 Bildern," plate 54. (Wien: Verlag von R. Lechner, 1910). Copy held in the Perry-Castañeda Library, The University of Texas at Austin.

symbolic meanings that can be interpreted for their iconographic intent and effect. American suburbs, for example, have long carried notions of the 'American dream' for both residents and developers alike, who have idealized the single family house and yard as embodiments of gender stereotyped 'family values'. But those same suburban landscapes may also be envisioned as prisons for women who feel confined by their relative isolation and as markers of exclusion that carry codes of social distinction and hierarchy. Not one meaning emerges from the complex array of suburban landscapes, but rather a multiplicity of understandings that entwine identity, race, gender, class, and politics.

Concluding Comments

The landscape iconography method in human geography has evolved from the preoccupation of a handful of historical geographers in the 1980s into one of the discipline's signature methodologies. Social, political, historical, and especially, cultural geographers have expanded the scope of landscape inquiry dramatically with the recognition that representations of the world can have significant impacts on how that world works. In every case, an iconographic approach to landscape shows how people create their own worlds and, along the way, construct visual representations of their individual and group beliefs, values, tensions, and fears.

See also: Archives; Critical Cartography; Ethnography; Movies and Films, Analysis of; Nationalism; Photographs; Representation, Politics of; Time and Historical Geography; Visualization, Feminist.

Further Reading

Cosgrove, D. (1998). *Social Formation and Symbolic Landscape*. Madison, WI: University of Wisconsin Press.
Cosgrove, D. and Daniels, S. (eds.) (1988). *The Iconography of Landscape: Essays on the Symbolic Representation, Design, and Use of Past Landscapes*. Cambridge: Cambridge University Press.
Daniels, S. (1993). *Fields of Vision: Landscape, Imagery, and National Identity in England and the United States*. Princeton, NJ: Princeton University Press.
Domosh, M. (1996). *Invented Cities: The Creation of Landscape in Nineteenth Century New York and Boston*. New Haven, CT: Yale University Press.
Duncan, J. and Duncan, N. (2004). *Landscapes of Privilege: The Politics of the Aesthetic in an American Suburb*. New York: Routledge.
Harley, J. B. (2001). *The New Nature of Maps: Essays in the History of Cartography*. Baltimore, MD: Johns Hopkins University Press.
Harvey, D. (2006). *Paris, Capital of Modernity*. New York: Routledge.
Hayden, D. (2003). *Building Suburbia: Green Fields and Urban Growth, 1820 2000*. New York: Vintage Books.
Matless, D. (1998). *Landscape and Englishness*. London: Reaktion.
Mitchell, W. J. T. (ed.) (1994). *Landscape and Power*. Chicago, IL: University of Chicago Press.
Panofsky, E. (1955). *Meaning in the Visual Arts*. Garden City, NY: Doubleday Anchor Books.
Rose, G. (2007). *Visual Methodologies: An Introduction to the Interpretation of Visual Materials*. London: Sage.
Schein, R. (ed.) (2006). *Landscape and Race in the United States*. New York: Routledge.
Schorske, C. (1981). *Fin De Siécle Vienna: Politics and Culture*. New York: Vintage Books.
Schwartz, J. and Ryan, J. (eds.) (2003). *Picturing Place: Photography and the Geographical Imagination*. London: I.B. Tauris.

Relevant Websites

http://www.artcyclopedia.com/
Artcyclopedia.
http://witcombe.sbc.edu/ARTHLinks.html
Art History Resources on the Web.
http://pro.corbis.com/default.aspx
Corbis.
http://www.maphistory.info/
Map History/History of Cartography.
http://www.mediahistory.umn.edu/
Media History Project.
http://www.nationalmediamuseum.org.uk/
National Media Museum.
http://oddens.geog.uu.nl/index.php
Oddens' Maps.
http://memory.loc.gov/ammem/index.html
The Library of Congress American Memory.
http://digitalgallery.nypl.org/nypldigital/index.cfm
The New York Public Library Digital Gallery.
http://www.loc.gov/rr/print/
The Prints and Photographs Reading Room of the Library of Congress.
http://www.rps.org/
The Royal Photographic Society.
http://americanart.si.edu/collections/exhibits/helios/index.html
The Smithsonian American Art Museum, Helios Photography Online.
http://visualstudies.duke.edu/
Visual Studies Initiative.

Landscape Perception

K. M. Morin, Bucknell University, Lewisburg, PA, USA

Glossary

Behavioral Geography A psychological turn in 1970s Anglophone human geography which emphasized the role of culture, cognitive processes, and decision-making variables as mediating the relationship between environment and spatial behavior.

Landscape After James Duncan, an ordered assemblage of objects visible from a particular location, which acts as a signifying system through which a social system is communicated, reproduced, experienced, or explored.

Perception Studies Focus on that portion of the total environment that is cognitively sensed and experienced, and that serves as a basis for decision making.

Politics of Landscape Landscapes carry symbolic or ideological meanings that reflect back and help produce social practices, lived relationships, and social identities, and also become sites of claiming or contesting authority over an area.

Spatial Science A quantitative approach to geography concerned with systematic observation and measurement of people, things, and events intended to enable the formulation of laws governing their spatial distribution.

Subjectivity One's sense of self. Humans perceive and ascribe meaning to places, develop emotional attachments to them, and interpret landscapes in highly subjective ways, that are, nonetheless, embedded within larger social structures.

Introduction: Terminology

Landscape

'Landscape' is a basic organizing concept in Anglophone cultural geography, but is equally foundational in fields as diverse as art, architecture, environmentalism, planning, and in the Earth sciences. It is important to keep in mind that landscape traditions differ across disciplines and places, and even within Anglophone cultural geography the concept of landscape carries much ambiguity and complexity, with probably hundreds of nuances to the term. The term 'landscape' has been (confusingly) conflated with numerous other geographical categories such as region, area, nature, place, scenery (particularly the rural countryside), topography or landform, and environment.

However, geographers have tended to emphasize the visual aspects of the physical world when they use the term landscape. Therefore, a useful way to begin thinking about landscape as distinct from these other geographical categories is as a portion of the Earth visible by an observer from a particular position or location – both physical location and social location. This prevailing connotation of landscape as a particularly visual form of spatial knowledge that can be taken 'in a single view' derives from sixteenth century Dutch landscape painting, with its emphasis on scenery. Today landscape continues to connote this visuality, although it is no longer confined to the single framed view or esthetic pleasure, and also invokes a greater concern for viscerality and experience.

Thus landscape may be thought of in the first instance as a 'thing' – an area or the appearance of an area, and the particular ways component parts of that area have been arranged to produce that appearance. From this vantage point we can talk about 'agricultural landscapes', 'urban landscapes', 'modern' and 'postmodern' landscapes, 'symbolic landscapes', 'ordinary landscapes', 'heritage landscapes', and so forth. But as will be discussed below, landscapes have both material and ideological aspects, and, in that sense, help produce and reproduce social norms and relations. Landscapes are represented in various media (film, paintings, advertising), and in turn are representations of lived relationships. Especially over the past 20 years, Anglophone cultural geographers have come to recognize how important representational practices are to the production of landscapes and hence, to social relations and social structure.

Perception

'Perception' is an equally amorphous term as it is used in Anglophone human geography. The concept primarily has been used within behavioral geography to refer to views of one's environment developed through cognitive processes. Study of perception in this sense has strong ties to psychology. People construct views of their worlds and move through space based on cognitive processes that capture and organize direct and indirect knowledge, experiences, and information acquired through the senses. The outcome is generally referred to 'environmental perception' (rather than landscape perception *per se*).

Despite the fact that most perception studies in geography have focused on systematic observation and measurement of sensory activity, cognitive categories, and cognitive processes, this article takes a much broader view of landscape perception. 'Perception' of landscapes can be studied in much broader social terms than merely or only as a product of cognition. We perceive (view, observe, know, understand) landscapes based on our individual subjectivity and identity, as well as through social systems within which we are embedded, such as socioeconomic class and religious belief systems. Landscape perception study in this latter sense is more sociological in orientation than psychological.

This article outlines two basic approaches to the study of landscape (not necessarily in the order they developed temporally): those whose methodology is quantitative and those more qualitative. 'Quantitative approaches' flourished and declined along with the spatial science revolution of the mid twentieth century, and are addressed first. 'Qualitative approaches' have in one form or another persisted throughout the twentieth century and into the twenty first, and are discussed second, in more depth.

Quantitative Methods

Landscape Perception: The Behavioral Approach

Mid twentieth century Anglophone human geography witnessed two important trends. One was a humanistic trend that stressed individuals' subjective experience, values, attitudes, and meanings involved in constructing views of the world. The second was spatial science, a quantitative or statistical turn in geography that stressed the importance of discovering process laws (such as 'central place theory') to explain the distance, distribution, and location of people, events, and things on the surface of the Earth. Behavioral geography can be understood as a subset or intersection of both of these trends: sharing spatial science's goal for developing predictive process laws – for example, to explain perceptions of what constituted hazardous landscapes – yet laying stress on the individual's subjective experience and worldview.

Behavioral geography represented a psychological turn in human geography that emphasized the role of culture, cognitive processes, and decision making variables as mediating the relationship between environment and spatial behavior. Within human geography *c.* 1960–70, 'perception' referred to an individual or aggregated group understanding, interpretation, or view of the surrounding environment based on information acquired through the senses and subsequently filtered through a number of cognitive processes. Theories were developed by aggregating individual surveys, interviews, questionnaires, and observations of behavior to the group level.

One of the most important advances made in this research was a critique of the so called 'rational economic man' upon which spatial science modeled human behavior. That is, that people's perception of and behavior in space could not be accounted for by the dominant model of profit maximizing 'rational' thinking: why do people take risks, such as living on flood plains, that seem to go against their own best interests? Perception studies emphasized personal preferences, superstitions, and other 'nonrational' mental constructs that are important variables in constructing world views.

Kevin Lynch, in his 1960 landmark study, *Image of the City*, studied what people knew of the physical/spatial environment in three American cities: Boston, Jersey City, and Los Angeles. Using sketch maps and interviews, Lynch identified five categories of cognitive information that people hold in their minds of the physical environment: paths, edges, districts, nodes, and landmarks, which are in turn used by people to orient themselves and to navigate within a place or region. *Image of the City* was important not only because it focused new attention on the perception of cities as a part of cognition, but also because it gave geographers a conceptual framework to study perception of the built environment.

Other studies, such as Downs and Stea's 1973 *Image and Environment*, argued that cognitive images are compiled in part through behavioral patterns, such that one's spatial orbit or activity space within the city helps form mental constructs of it – that is, one's perception. In addition, the more firsthand information people have of their environment, the more detailed and comprehensive their images will be. The study of images of Los Angeles held by three different groups (residents of Westwood, an affluent neighborhood; Avalon, a poor, inner city neighborhood; and Boyle Heights, a poor immigrant neighborhood) found that those whose activity space was constrained within the city held considerably smaller and less detailed cognitive 'maps' of it than those with more extensive spatial orbits. Thus, socioeconomic status, culture, as well as personal characteristics influenced spatial patterns and these in turn the cognitive perception.

This work effectively argued that perceptions have important consequences. People draw on their cognitive perceptions, for example, in making decisions about migrating from one area to another. Many studies of the settlement of North America found that the changing images of the Great Plains and West – as alternately arid and unattractive, or fertile and hospitable – greatly influenced the settlement patterns in the nineteenth century. Other studies similarly demonstrated the relative desirability of various locations in the US as places to live and work. Perception studies of this sort relied on a 'superorganic' or monolithic view of cultural influences (discussed momentarily). They were also based on the premise that it is impossible for an individual to see the environment as it 'actually' exists, since perception is filtered through cultural differences, the media, and the individuals' own subjective experiences, values, and goals. That perceptions of landscapes are somehow 'distorted' representations of reality was a contentious assertion likewise debated within qualitative studies of landscape perception.

Qualitative Methods

The Landscape School

Qualitative methods for studying landscape perception were introduced into American geography in the 1920s by Carl Sauer, especially with his *The Morphology of Landscape*. Sauer reacted against the environmental determinism of his day by arguing that it was collective human transformation of natural landscapes that produced what he called 'cultural landscapes'. While Sauer himself was more concerned with ecological processes set in motion by humans which produced, for example, agricultural practices and patterns, his more enduring influence was on a whole generation of cultural geographers associated with the Landscape School (aka Berkeley School) who used his empirical observation method to study the morphological features of landscapes as evidence of cultural difference. For decades his followers tended to study, for example, cultural artifacts such as house types and barn types to trace cultural hearth areas and diffusion of culture groups.

Mid twentieth century landscape studies in geography were also greatly influenced by the English historian W.G. Hoskins, who argued for detailed studies of landscape history, and the American geographer J.B. Jackson, who studied popular culture through vernacular landscapes such as trailer parks in the American Southwest. Jackson was founder of the popular *Landscape* magazine, published for 17 years beginning in 1951. In 1979 Donald Meinig edited a collection of works, *The Interpretation of Ordinary Landscapes*, written by some of the most notable landscape geographers working at the time – himself and J.B. Jackson, Peirce Lewis, David Lowenthal, Marwyn Samuels, David Sopher, and Yi Fu Tuan. This collection demonstrated both the continued interest in 'ordinary', everyday landscapes in Anglophone cultural geography, such as churches and houses, as well as how landscapes revealed social and personal tastes, aspirations, and ideologies. To Meinig, landscapes themselves could be read as collective social ideologies and processes: "symbols of the values, governing ideas, and underlying philosophies of a culture."

The Interpretation of Ordinary Landscapes also demonstrated one of landscape studies' most enduring metaphors, that of reading and interpreting landscapes as 'texts'. As texts, landscapes have elements arranged in a particular order that we can translate into language, grasp meaning, and 'read'. This approach to understanding landscape became particularly useful for reading landscapes as 'symbolic' representations of relationships among people and between people and the physical environment. Symbolic landscapes can be understood as representing particular values and aspirations of the people who build them, whose meanings are conveyed to the public and read (interpreted) by them.

For example, one might read the architectural features of Manhattan's skyline, particularly the height of skyscrapers that comprise it, as symbolic of economic power, technological ingenuity, and modernity and progress. After the September 2001 attacks on the World Trade Center, however, many commentators added to that list another set of more decentered understandings of what those skyscrapers symbolized – such as American greed, unjust foreign policy, and American vulnerability. The fact that these various meanings and interpretations coexisted simultaneously forced a recognition that not only could the same landscape carry vastly different meanings to different observers, but that the landscape itself was also a reference to a much larger set of social relationships, domestic and global, that required finetuned contextualization. Detailed study of such signs and symbols – the 'semiotics' of landscape – would in many ways structure landscape perception studies in the 1980s and 1990s.

Critical Social Theory

One of the most significant developments in Anglophone landscape studies from the 1980s toward was that advocated by the 'new' cultural geographers, a shift that began first in Europe within the context of an emerging cultural studies paradigm. The shape this discussion took was not so much a conversation between different approaches as much as it was a one sided rejection of the (old) Landscape School by adherents of the new.

By the last two decades of the twentieth century, the textual metaphor helped usher in a number of new questions related to how perceptions of landscape mediate social relationships. Informed by critical social theory, some geographers began challenging the assumption of their predecessors that cultural groups 'collectively' produced and read landscapes in the same way, as the preceding example of the lower Manhattan landscape illustrates. Perhaps the most far ranging critique of the Landscape School addressed its reliance on the 'superorganic' view of cultural influences on perception, popular in American Anthropology when Carl Sauer came of age. This view holds that culture is an external force operating in a deterministic way on a plane 'above' the human, imposing values on individuals such that all members of a culture came to share a customary way of life (and viewpoint). This notion of a collective culture both shaping and being shaped by landscapes came under criticism in the 1980s, to be replaced by the notion that landscapes are neither collectively produced, nor interpreted/perceived in the same way by members of a culture group. Rather, landscapes came to be thought of as sites through which competing authorities, per ceptions, interpretations, discourses, and knowledges intersect.

Thus, understanding the power of landscape to re inforce, challenge, or subvert the existing social order – the politics of landscape – has been of primary concern to many critical social theorists. They insist on acknowl edging the patterns and processes of hierarchical social organization responsible for the morphological features observed in landscapes. Denis Cosgrove, for example, in his *Social Formation and Symbolic Landscape*, defined land scape as a 'way of seeing' associated with the rise of capitalist property relations. He argued that the land scape concept enabled an erasure of class difference via media such as landscape paintings of landowners and their country property.

The works of James Duncan have been instrumental in clarifying the extent to which landscapes contain dif ferent meanings to different viewers, and how they act as 'intertextual' media through which often competing interpretations, discourses, and knowledges intersect. Critical social theorists such as Duncan have tended to highlight the extent to which multiple interpretations, or layers of meaning, are embedded within landscapes and their representations. This is important because land scape as a site of struggle for challenging the dominant social order often rests on effective interpretations. Therefore, it seems essential to recognize that meanings are not inherent in concrete objects or the physical world, but that they are socially ascribed to objects and that they change over time, and with the particular perspectives and social positioning of the viewer.

While power and authority seems to lie with those who can 'control' how landscapes are to be read, there is always room for contestation of that authority. In this sense, more recent landscape studies include a decidedly political component as they highlight the social and cultural conflicts and relationships, especially unequal power relations based on race/ethnicity, class, gender, and sexuality, that are involved in the creation, repre sentation, and interpretation of landscapes. Landscapes of graffiti offer a useful example. Dominant or hegemonic readings of graffiti, such as those articulated by the mayor's office or transportation authorities, might assert that graffiti is simply destruction of property, a crime against the city. But graffiti has also been variously understood as a means by those with no others to mark and stake out territory, or a means to challenge the existing social order by drawing visual attention to the situation of those economically and socially marginalized in the city. Alternative readings of landscapes always exist, and landscapes can always be read in ways not intended. Many graffiti makers think of themselves as creating art – an intention behind graffiti landscapes rarely acknowledged by more powerful voices.

As the foregoing discussion implies, one fundamental aspect of the critical social theory approach is that landscapes always carry a set of 'representational prac tices'. These refer to how people see, interpret, and represent the world around them as landscape, and how that represented landscape reflects and actually helps produce a set of lived relationships taking place on the ground.

Representation of landscapes can take many forms – narrative descriptions, drawings, paintings, maps, plan ning documents, engravings, photographs, and films, among others. Trevor Barnes and James Duncan's 1992 edited collection, *Writing Worlds: Discourse, Text & Metaphor in the Representation of Landscape*, examined nu merous such forms of landscape representation. These authors asserted that landscape representation and in terpretation required contextualization of author and audience, an outline of the rhetorics and tropes (figures of speech) employed to convey meanings, and an analysis of the processes by which readers become convinced that meanings conveyed are the 'natural' order of things in the world.

The largest monumental landscapes in the US, for example, carry laudatory messages about war heroes and military conquest. We can identify a set of social actors who produce such landscapes –historical societies, town planners, veterans groups, and so on. The messages de ployed by the monuments – a particular version of the past that celebrates masculinist cultural values – actively reproduce those values in the present and thus can shape social practice, such as by reproducing a culture of war. Of course, to reiterate, such values never go un challenged and can be undermined in numerous ways. The Vietnam War Memorial in Washington, DC became a site for contesting the celebration of war, for example, as it highlights the suffering and loss of war rather than triumphal conquest.

Much landscape work in recent years originating from a neo Marxian standpoint has highlighted questions about what exactly is the relationship between the con crete, physical, material world – the morphological as pect of landscape – and its representation. Don Mitchell has argued that landscape studies have relied too much on visuality, and he advocates, among other things, a focus on that which has been hidden from view, such as the histories of laborers whose work literally produces landscape. He worries that landscape studies that are concerned only with representations seem to leave the 'real world' of landscape modes of production and reproduction behind as objects of study. In his way of thinking, meanings produced in and through language, texts, discourses, iconography, and symbolism neglect the 'brute reality' of landscapes and thus represent a 'dangerous politics'. To illustrate, Mitchell's 1996 *The Lie of the Land* explores the role of labor within the expanding capitalist economy in southern California's San Joaquin Valley. He shows how landscapes literally 'get made', as well as how they are then re presented as landscape: in this case, as an esthetic, pastoral depiction of thriving agriculture (but otherwise invisible, exploited labor).

Mitchell's argument notwithstanding, it must be kept in mind that we only perceive landscapes through our readings of them; that is, that there is always a cultural filtering process that brings the 'reality' of landscapes to us through language and discourse. As Denis Cosgrove explained in 2006, "landscapes have an unquestionably material presence, yet they come into being only at the moment of their apprehension by an external observer, and thus have a complex poetics and politics." Thus, rather than set up a false dualism between the materialism of the likes of Mitchell and the postructuralism of the likes of Duncan and Gregory, it seems more productive to focus on the necessarily discursive constitution of the material world.

Representations, then, are not reflective or distortive images of some real, pre interpreted reality, but they themselves materially constitute reality. Thus, 'reality' is indistinguishable from its representation, and in this sense the much more important question is how representations of landscapes are produced and contested. In any case, this debate has subtly shifted in recent years to focus more on landscapes as fields of action, questioning how people are able to shape landscapes and be shaped by them, for instance, through control of legal systems. Epistemological differences remain between materialists and poststructuralists; though both 'sides' have moved beyond considering the question of the binary itself a departure point for analysis, there is little agreement on how to conceptually dismiss it.

Among those who have been able in recent years to bridge the divide between the materialists and those more attuned to discursive analysis of landscapes are those working within feminist geography. Much of the feminist landscape scholarship focuses on how women's relationships to landscapes (as experience, representation, or interpretation) differ from men's; or how landscapes construct, legitimate, reproduce, and contest gendered and sexualized identities. Gillian Rose's 1993 study of geography's traditions in fieldwork and landscape analysis is one feminist critique that has found its way into more mainstream human geography. Informed by a larger feminist corpus that highlights the positionality of the researcher (or observer) and situated and partial knowledges, Rose argues that geographical traditions in landscape study are deeply masculinist.

To Rose, geography's traditions involve a masculinist way of seeing landscape that is not just one of a relation of mastery or domination, but one of (white, bourgeois, heterosexual) pleasure in looking at landscape that has been constructed as feminine. Part of her commentary revolves around the same painting that Cosgrove deconstructed in his 1984 study, Thomas Gainsborough's 'Mr. and Mrs. Robert Andrews', *c.* 1748, which codified a particular way of seeing the land that helped naturalize and celebrate capitalist property and the rights of owners. Rose, however, rightly claims that Cosgrove's interpretation misses the different relationships that men and women had to the surrounding landscape; the painting reminds us that only men were landowners, and women's role was principally reproductive. In this and other landscape paintings, women appear passive or prostrate, as commodities of the male gaze. Not only do such landscape images themselves associate women with a feminized landscape, but as Rose points out, geographers reinforce sexism and masculinism by their inattentiveness to the impact that an analysis of gender roles and relations plays in landscape representation. These are not innocent, detached representations but they refract and reinforce lived gender roles and relations.

While much feminist work has demonstrated the mutual constitution of gendered landscapes and women's gendered identities, a recent turn in landscape studies has directed attention to relationships between men, masculinity, and landscape. Rachel Woodward, for example, examines the processes by which military masculinity and the landscape of Britain's rural countryside are mutually constituted. Woodward examines five sources of information – Army recruitment materials, general publicity, basic training information and videos, mass market paperbacks about military adventures, and television documentaries on military life. She shows how essential a particular construction of rurality itself is to the construction of 'warrior hero' – it is dangerous, rough, and hazardous. The rural countryside in the Army documentation is not that of idyllic community and nature in harmony, but is rather a harsh, threatening landscape against which the new recruit is pitted, and out of which his requisite physical and mental attributes will arise through its 'conquest'. Thus, this representation of the rural serves the duo purpose of articulating and legitimating one hegemonic type of military masculinity, as well as constructing the rural itself as a legitimate place to bear arms.

This kind of attention to how gender roles and relations are produced in and by landscape provides a useful model for further work on ways in which landscape helps produce and mediate national, ethnic, and sexual differences.

See also: Berkeley School; Landscape; Landscape Iconography; Representation and Re-presentation; Sauer, C.

Further Reading

Barnes, T. and Duncan, J. (eds.) (1992). *Writing Worlds: Discourse, Text and Metaphor in the Representation of Landscape*. London: Routledge.

Cosgrove, D. (1984). *Social Formation and Symbolic Landscapes*. London: Croom Helm.

Cosgrove, D. (2006). Modernity, community and the landscape idea. *Journal of Material Culture* 11, 49 66.

Downs, R. and Stea, D. (eds.) (1973). *Image and Environment: Cogni tive Mapping and Spatial Behavior*, pp 120 122. Chicago, IL: Aldine.

Duncan, J. (1990). *The City as Text: The Politics of Landscape Interpretation in the Kandyan Kingdom*. Cambridge: Cambridge University Press.

Jackson, J. B. (1990). The house in the vernacular landscape. In Conzen, M. (ed.) *The Making of the American Landscape*, pp 355 369. Boston, MA: Unwin Hyman.

Leighly, J. (ed.) (1963). *Land and Life: A Selection of the Writings of Carl Ortwin Sauer*, pp 315 350. Berkeley, CA: University of California Press.

Meinig, D. W. (ed.) (1979). *The Interpretation of Ordinary Landscapes: Geographical Essays*. New York: Oxford University Press.

Mitchell, D. (1996). *The Lie of the Land: Migrant Workers and the California Landscape*. Minneapolis, MN: University of Minnesota Press.

Monk, J. and Norwood, V. (eds.) (1987). *The Desert Is No Lady: Southwestern Landscapes in Women's Writing and Art*. New Haven, CT: Yale University Press.

Morin, K. M. (1998). Peak practices: Englishwomen's 'heroic' adventures in the 19th century American West. *Annals of the Association of American Geographers* 89, 489 514.

Morin, K. M. (2003). Landscape and environment: Representing and interpreting the world. In Holloway, S. L., Rice, S. P. & Valentine, G. (eds.) *Key Concepts in Geography*, pp 319 334. London: Sage.

Nash, C. (1996). Reclaiming vision: Looking at landscape and the body. *Gender, Place and Culture* 3, 149 169.

Olwig, K. (2002). *Landscape, Nature and the Body Politic: From Britain's Renaissance to America's New World*. Madison, WI: University Wisconsin Press.

Rose, G. (1993). *Feminism and Geography: The Limits of Geographical Knowledge*. Minneapolis, MN: University of Minnesota Press.

Sauer, C. (1925). *The Morphology of Landscape*. Berkeley, CA: University of California Press.

Woodward, R. (2000). Warrior heroes and little green MEN: Soldiers, military training, and the construction of rural masculinities. *Rural Sociology* 65, 640 657.

Language

R. Jones, Aberystwyth University, Aberystwyth, UK

Glossary

Bilingualism The ability to speak two languages.
Geography of Languages An attempt to show how language use varies spatially.
Language A system, which enables humans and other higher animals, as well as computers, to communicate with one another.
Languages of Geography The way in which the practice of the discipline of geography – either in terms of teaching or research – is influenced by language.
Monolingualism The ability to speak one language.
Multilingualism The ability to speak three or more languages.
Place Names The names that are used to identify particular places, features, or landmarks and which may incorporate and reflect identities, histories, and politics.

Introduction

The study of language, languages, and dialects has been an important feature of geography since its emergence as a modern discipline in the nineteenth century and continues to be a significant object of enquiry, particularly in the subject areas of cultural and political geography. For much of the nineteenth and twentieth centuries, the main academic focus has revolved around understanding what we may term 'the geography of language and/or languages'. In its most straightforward guise, this work has attempted to map the spatial distribution of language use at a variety of different geographic scales, ranging from the local to the national and the global. Such research has also paid especial heed to the geographic shifts in language use that have taken place over time. At the same time, some work on the geography of languages has sought to position language use within a broader set of questions relating to culture and individual and group identity. In this way, language use has been mapped as a way of enabling geographers to make broader claims concerning the (changing) geographic distribution and character of cultural or political groups. Fascinating work on the geography of languages continues to this day but it has been supplemented by two further, and related, areas of research. The first area of research is an academic interest in the language(s) of geography or, in other words, the way in which the use of a particular language influences the kind of teaching and/or research that is conducted by geographers. An emphasis is placed upon understanding how language use, and attendant cultures and identities, influence teaching and research cultures within the discipline of geography. Such themes have received the most sustained attention in recent years in terms of the ongoing debate concerning the dominant Anglocentricism of contemporary 'human geography', particularly in the context of allegedly 'international' research cultures. The second area of research, which has become important in recent years is a post structural academic interest in the connections between geography and language and which has focused in particular on issues relating to representation. Influenced by a so called 'linguistic turn' within philosophy and literary theory, a number of human geographers have begun to question the ontological certainties portrayed within the modern intellectual project. Derridean ideas concerning the inherently subjective signification of meaning through language have been used by a number of post structural geographers to illustrate the uncertainties and politics associated with all forms of (geographic) representation. Of course, there are important crossovers between these two more recent concerns. Both are grounded in post structural epistemological debates concerning the power laden qualities of language.

I use this threefold division as a way of structuring the remaining sections of this article. I begin by discussing the work that has been conducted on the geography of language(s), focusing in particular on the research by geographers that has sought to map language use and to relate this to issues of identity and politics. Following on from this, I then elaborate on the research that has been conducted on the language(s) of geography. In particular, I discuss the way in which languages and their related identities are implicated in geographic research and teaching. The final substantive section discusses the link between geography and language and shows how post structural concerns about representation have drawn much of their academic justification from philosophical debates relating to language. As a way of grounding these debates, I focus briefly on the use of language within recent geographic research on race and geopolitics.

Geography of Language(s)

Academic attempts to understand the distribution and relative characteristics of languages can be traced to the eighteenth and ninteenth centuries. During this period, comparative linguists sought to explain the linguistic characteristics of particular languages, their relationship to other cognate or different languages, and the geography of language use. William Jones and Edward Lluyd, for instance, attempted to explain the linguistic character of the various Celtic languages – Scots Gaelic, Irish, Manx, Welsh, Cornish, and Breton – during this period. Efforts were made to delineate the inherent linguistic qualities of these various languages, as well as explaining their relationship with one another. At the same time, such studies were also implicitly attempts to understand the relative geographic distribution of Celtic languages. Efforts to map languages – in terms of their linguistic characteristics and their spatial distribution – also evolved into academic studies of the association between language and identity. To return to the example of Celticism once more, many commentators during the nineteenth century used the existence of the various Celtic languages as a means of justifying racist beliefs in the necessary cultural differences that existed between Celts and Anglo Saxons. Celtic speakers, in this sense, were deemed to display particular racial characteristics – being excitable, wild, and prone to unprovoked acts of violence – which marked them out as distinct from the allegedly rational and scientific Anglo Saxons. Linguistic differences were thus translated into much broader commentaries on the alleged intrinsic identities associated with different kinds of people.

It is possible to witness these twin efforts – to map language use and to examine the link between language use and identity – in more recent geographic research on language. Work on the connections between language and identity politics, for instance, has attempted to demonstrate the way in which language use can be enrolled as a powerful marker of national identity. A number of contemporary states, in this respect, have promoted the use of their national language in order to differentiate themselves from other nations and states. A good example of the strong connections forged between language and national identity can be found in Ireland, where there has been a concerted effort to protect the Irish Gaelic language in the west of the country from further linguistic erosion through the designation of a Gaeltacht, in which support for the Irish language is supposed to be concentrated. The language's significance, of course, is the way in which it is incorporated into Irish nationalism as a means of distinguishing the Irish nation from Britain, Ireland's estwhile colonizer.

The use of language as a marker of national identity is never straightforward, nonetheless, as a number of authors have shown. In one context, we need to consider the way in which the use of different languages within particular states can lead to cultural and political conflict. The case of Canada is an especially significant one in this respect. Many authors have noted the importance of linguistic divisions within Canada, which reflects and has the potential to exacerbate broader cultural and political differences within the country. The relationship between a French speaking Quebec and the remaining provinces of Canada, which are predominantly English speaking, in particular, has been uneasy. The group identity that has existed and been promoted in Quebec has been strongly associated with the French language (and Catholicism). Here, we witness the way in which languages are implicated in broader cultural and political conflicts between different groups of people.

In another context, we need to think about how various nationalist organizations confer different kinds of significance on language within the discourse of nationalism. In more geographic contexts, nationalist organizations may designate certain places, landscapes, or territories as ones which fulfill the linguistic requirements associated with national identites, while other places, landscapes, and territories may be deemed to be marginal or irrelevant to the nation's linguistic project. The role of the Gaeltacht within Irish nationalism, discussed above, alludes to the great significance that is placed upon this region for the way in which Ireland is imagined geographically; other areas, because of their linguistic geographies, become marginal within Irish nationalist discourses. Similarly, work on Wales has shown how particular areas, most notably the Welsh speaking 'heartland', assume a position of great significance within the Welsh nationalist geographic imagination, while others are relegated to a secondary status (see **Figure 1**).

But if language has been enrolled within nationalist discourse as a key marker of national distinctiveness, it is clear that it has also featured as an important badge of identity politics at other scales. Much work in cultural geography in recent years has examined how cultural and linguistic diasporas are leading to the emergence of new forms of cultural and political identity. Language is connected to the emergence of diasporic identities in complex ways. At one extreme, language may be viewed as something that both shapes and is shaped by the fluid and antiessentialist forms of identity that can be associated with diasporic communities. Bielenburg has argued that "diasporic identities, transnational and subversive in character, challenge the security of identities defined, but also limited, by national boundaries." The multiple languages that are often spoken by diasporic communities, therefore, may both reflect and help to constitute these fluid and multiple identities.

Conversely, some have argued that diasporic communities are not necessarily based upon mobile and 'multiple attachments' of identity but, rather, may seek to emphasize the "essentialized notions of place and identity that they are supposed to transgress" (2005: 54). Carter's research in the United States, for instance, has shown how essentialized and highly defined notions of Croatian identity reemerged within the Croatian

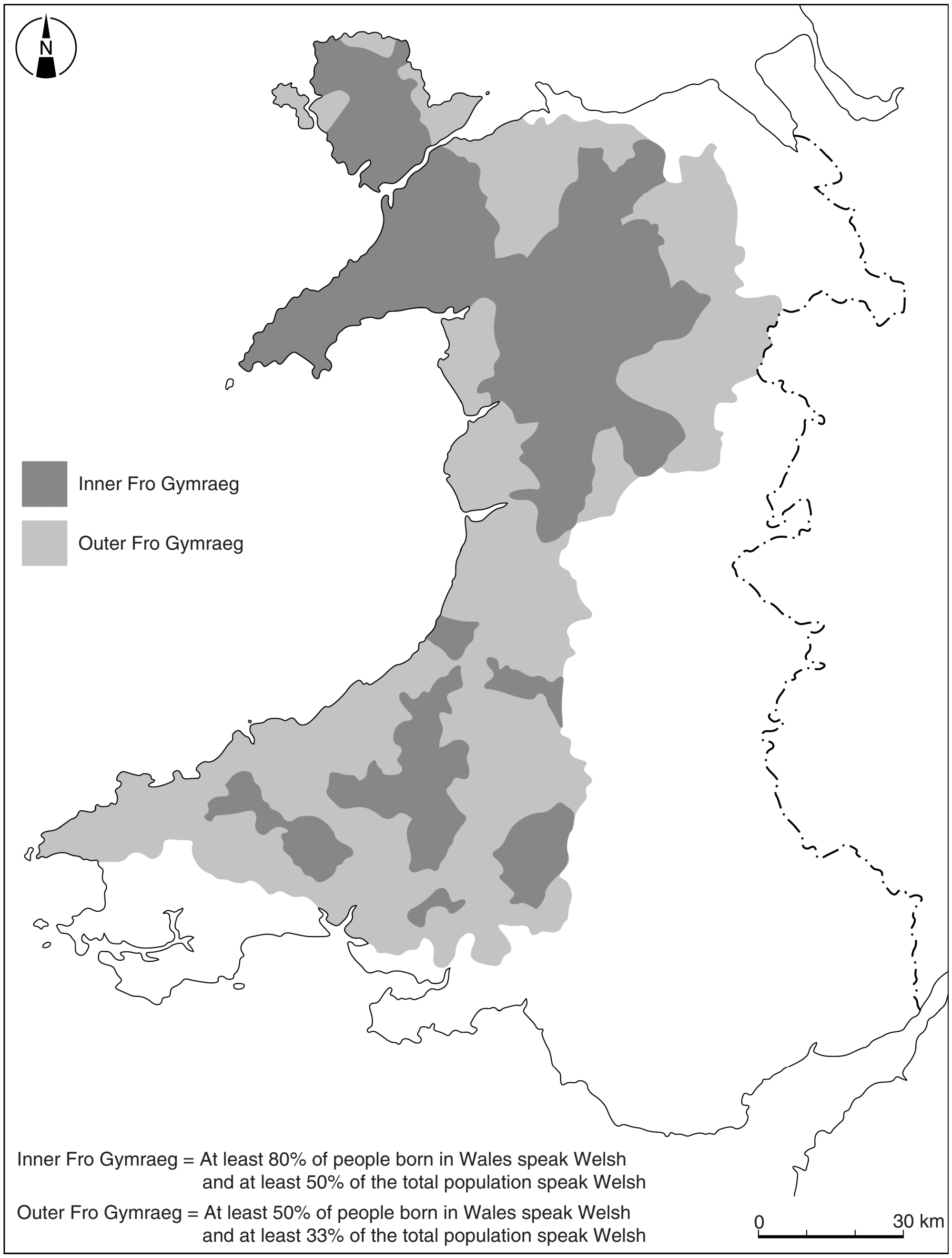

Figure 1 The differing interpretations of a Welsh-speaking 'heartland'. After Jones, R. and Fowler, C. (2007). Where is Wales? Narrating the territories and borders of the Welsh linguistic nation. *Institute of Geography and Earth Sciences*, Aberystwyth University.

diaspora at the time of the civil war in Yugoslavia. Sig nificantly, the leaders of the Croatian diaspora in the United States sought to garner popular support for their promotion of an increasingly exlcusive version of Cro atian diasporic identity through reference to the need to protect Croatian heritage, culture 'and language' from external pernicious influences. In this way, language has been enrolled as a marker of an exclusive group identity within the Croatian diaspora. At the same time, other work has considered how language is implicated in the reproduction of more local or regional identities. Some of the more interesting research in this area has explored the significance of local variations in language or dialect as indicators of regional cultures. Working in a tradition of cultural geography that would have been familiar to Carl Sauer, geographers, such as Zelinsky, have attempted to illustrate the existence of different cultures or vernacular regions within the United States. Although the distinctions between the various regions are subtle, they do exist, and are sustained by a variety of

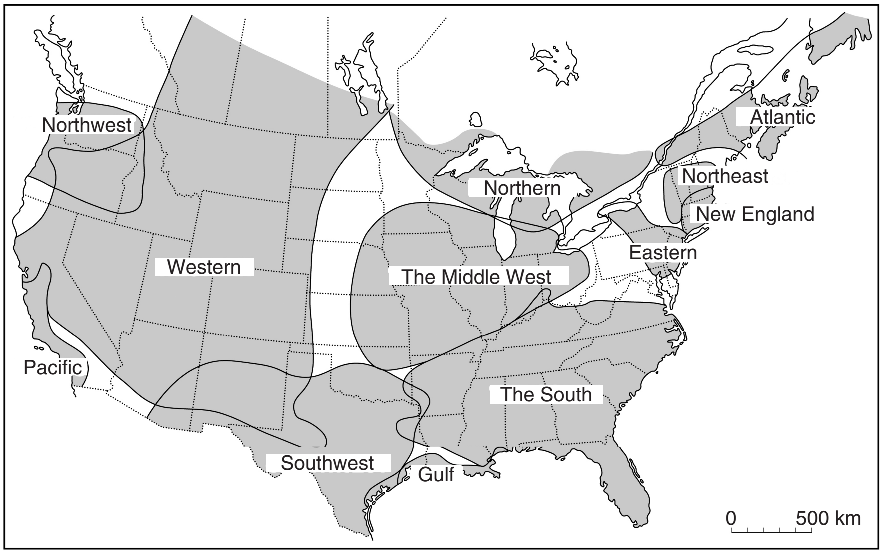

Figure 2 The vernacular regions of the United States. After Zelinsky, W. (1994). *Exploring the Beloved Country: Geographic Forays into American Society and Culture*. Iowa City, IO: University of Iowa Press.

different factors, most notably language and dialect, along with religion, foods, folklore, architecture, politics, and other cultural categories (see **Figure 2**).

One material marker of the connection between lan guage and identity is in the context of place names. Place names are important since they demonstrate how lan guage, as well as being a social and cultural practice, may also be inscribed onto the landscape. While the location of particular place names may be useful as a way of charting the existence of cultural and vernacular regions and linguistic shifts, it is equally pertinent as a means of illustrating the identity politics associated with language. Nash's work is important in this respect. She has exam ined the way in which identity politics in Northern Ireland has been connected to debates about place names. Naming and renaming places have been long standing colonial practices, most notably in the context of the replacement of indigenous place names by ones that derive from the metropolitan culture. In postcolonial contexts, therefore, debates about place names become a key site of political struggle. These debates are especially contentious in places like Northern Ireland, in which Protestants (identifying largely with a Unionist politics) and Catholics (supporting an Irish Nationalist politics) coexist. Nash's work is unusual, additionally, since she provides a more upbeat interpretation of the possibility for pride in local place names to act as linguistic 'bridges' between Protestants and Catholics within Northern Ireland.

Work on the politics of place names has also examined the connection between the act of naming places and other kinds of group identity. Alderman's research on the renaming of streets in commemoration of Martin Luther King Jr., for instance, has illustrated a tangled politics of naming within the African American community. Al derman notes how "debates over the scaling of King's memory revolve around the size of the named street, the street's level of prominence within a hierarchy of roads, and the degree to which the street transcends the spatial confines of the black community." What is significant about Alderman's research is the way in which it dem onstrates how even seemingly uncontentious acts of commemoration through naming become hotly debated and politicised acts. His work, indeed, shows how nam ing, as an act of memory, "can become a point of division and context…as activists seek to fulfill different political goals" (2003: 163).

Languages of Geography

If politics and power have become important threads within the study of the geography of languages, it is clear that similar concerns have been significant considerations in the work that has sought to understand the con nections between languages and the practice of geog raphy as a discipline. In general terms, such work derives much of its impetus from Olson's focus on a 'linguistic geometry', which structures the disciplinary engagement of human geographers with the world that they seek to study. Language and words, in this respect, possess a considerable cultural and political power to influence the kind of work conducted by geographers. Two interrelated areas of contemporary research within geography – at least in implicit terms – seek to take forward these ideas. The first area focuses on the influence of language(s) on the character of geographic research while the second examines the way in which languages are implicated in the learning environments created and reproduced within geography.

There has been a growing interest in recent years in examining the way in which language affects the type of research that is conducted by geographers. In a broad context, we need to appreciate how the meaning of geography – not only as a discipline but also as a term – varies from one linguistic context to another. Brunn, for instance, has charted how the meaning of the term 'geography' is dependent, to some degree, on language. Admittedly, while many European terms for 'geography' are rooted in the Greek words '*geo*' (earth) and '*graphos*' (writing) and thus ascribe a common meaning to 'geography', other languages' terms for 'geography' betray a different understanding of the term. For instance, no terms exist to depict the concept of 'geography' in many native American languages. In other languages, such as the Yoruba, there are separate terms for the study of physical and human relationships and processes. The different meanings ascribed to the term 'geography' within different linguistic groups, of necessity, impacts on practitioners' understanding of what the subject encompasses and affects the kinds of objects of enquiry that can legitimately be studied. In more specific contexts, the languages spoken by different groups can also affect the research process. Townsend, for instance, used a biographical approach as a means of understanding the geographies of women in the Mexican rain forest. Interestingly, she discusses the difficulties that she faced when trying to translate the women's life histories from a range of Mexican languages into an English research context. A loss of meaning was apparent in the translation process, which meant that it was difficult to represent the research subjects' actual experiences of living in the rain forest.

Townsend's study points to a broader issue concerning the cultural and political power associated with the English language in a contemporary human geography that is allegedly 'international' in character. A number of authors have commented on the increasing dominance of the English language and an Anglo American research agenda on the practice of human geography. In one regard, this dominance comes about as a result of Anglo American geographers' linguistic and geographic parochialness. The following quote, although directed toward the field of cultural studies, is an equally valid commentary on the research priorities of Anglophonic human geographers:

> As a field, cultural studies remains too centred in Anglophone perspectives, traditions and disciplinary histories. These problems are exacerbated by the apparent reluctance of many English speaking cultural studies scholars to grapple with empirical social and cultural contexts with which they are largely unfamiliar. (Grossberg, 2002: 1)

In another context, we need to think about the difficulties faced by non Anglophonic geographers in engaging in an academy that is dominated by the English language. Gutiérrez and López Nieva, for instance, have questioned the degree to which purportedly international journals in human geography are actually international in outlook. In addition, the mechanisms through which conferences are organized and papers are published, along with the more fundamental issues concerning the widespread lack of multilingualism among Anglophonic human geographers, mean that it can be difficult for non Anglophonic human geographers to make their voices heard on an international stage. The hegemonic dominance of an Anglo American research agenda within human geography, and the connotations of intellectual imperialism that come in its wake, seem set to remain a key area of debate within the discipline.

The dominance of an Anglophonic research culture also spills over into debates about the role played by languages within learning and teaching environments. Language, in this sense, influences how students engage with and acquire a range of geographical knowledges. Research on Wales, for instance, has shown how bilingual students, who study geography through the medium of English and Welsh, create a linguistic culture of geography because of their shared experience of specific practices and politics in bilingual education. The existence of this linguistic culture of geography influences the kinds of geography consumed by students in various classes and the range of geographical knowledges that they produce, whether in terms of presentations, essays, or extended research projects. The fact that bilingual students speak a different language – Welsh in this case – marks them out as separate from their monolingual English counterparts within the degree scheme. Furthermore, their bilingual status also enables these students to challenge some of the norms associated with an Anglophonic human geography. Such research further reinforces the need to comprehend the impact of languages on the teaching and learning environments that are created within human geography. The need to understand the influence of language on the reproduction of geographic knowledge in students, indeed, is likely to increase given the current emphasis on the international mobility of students within higher education.

Language and Geography

In this final section I consider the way in which philosophical and literary debates concerning language have been incorporated into human geography as part of a post structural academic project. Post structuralist epistemologies have largely been developed through recourse to philosophical debates concerning the character of language. Specifically, the main springboard for post structuralist thinkers was structuralist accounts of language. Saussure, a Swiss linguist and structuralist thinker, sought to understand the structures that informed all forms of communication, including language. For Saussure, "all languages [are] founded upon abstract regulations that shape the ways in which they are played, or manifested in practice" (Dixon and Jones, 2004: 82). The underlying tenor of Saussure's work is that the study of language provided a means of illustrating the sense of order and regularity that structured all forms of communication and, by implication, society and space in general. Post structuralists – most notably Derrida – have taken issue with such ideas and have sought to draw attention to the power relations that underpin the sense of order that is seemingly apparent in language and societies more generally. Derrida, for instance, argues that the structural qualities of language depend upon a power laden definition of a center and margin. Post structuralists, following Derrida, have questioned the way in which centers and margins – not only in language but also more broadly – are defined. Who has the social power to define the center and the margin? Who is marginalized through this process of definition? The center/margin binary, so popular with structuralists, also has the potential to limit what can be conceived of the world; things are either part of the center or the margin and there is little opportunity to occopy the space in between. Finally, structuralist modes of thinking about language have been questioned by post structuralists for their faith in the ability of a signifier (a word) to represent the signified (a thing). Post structuralists maintain that the language used to signify a particular thing is ultimately arbitrary and indeterminate, thus undermining any structuralist claim to a stability of meaning.

The impact of post structuralist epistemologies has been significant within human geography. Human geographers have contributed much to post structuralist concerns about representation, namely, how the real world is socially mediated "through ever present processes of signification." Language and discourse more generally are said to represent the world in socially mediated ways so that it is impossible to know the world without recourse to these representations. Some have questioned the value of such ideas, since they can lead to a 'crisis of representation', in which any notion of 'truth', 'objectivity', or authorial identity and integrity is seemingly lost. At face value, post structuralism would also seem to preclude strong evaluations or political intervention since such acts would necessarily invoke notions of 'truth' and 'objectivity'. Others have countered by arguing that the post structural academic project is inherently political since it seeks to illustrate the constructed character of all authority claims, which may be implicated in discourses and practices of domination.

Two brief examples will show the saliency of such claims. The first relates to the work conducted by post structuralist geographers on the idea of race. Racism is dependent upon the discursive construction of a center (usually white) and a margin or an other (usually non white). The aim of a post structural account of racism revolves around demonstrating how racial categories, far from being natural and given, are in fact socially constructed through the language used in political, academic, and popular discourse. Moreover, post structuralists also seek to illustrate the way in which notions of 'whiteness' and 'non whiteness' are relational and interdependent. In identifying 'whiteness' as a center, post structuralists are able to demonstrate and undermine the discourses and practices that help to promote and sustain white privilege.

The second brief example refers to the varied body of work that has been conducted in the subfield of critical geopolitics. Much of this work, significantly, has been informed by post structuralist concerns about language and discourse. O Tuathail and Agnew (1992: 198), for instance, have argued that the foundational premise of a critical geopolitics is that "geography is a social and historical discourse which is always intimately bound up with questions of politics and ideology" or, in other words, that 'geography is a form of power knowledge itself'. One recent manifestation of the use of language to construct 'power knowledge' has been in the context of the 'war on terror'. In the wake of the terrorist attacks on the United States that took place in September 2001, President George Bush was keen to use a language and discourse appropriated from the culture of the American West, referring to the need to 'smoke out' terrorists 'holed up' in the caves of Afghanistan. Another key set of discourses revolved around notions of civilization and barbarity. The United States was described as a civilized country of freedom and democracy, while the states or peoples supporting terrorism were described in a variety of media as an uncivilized 'other'. These brief examples demonstrate the key significance of language and discourse in 'scripting' geopolitical visions of the world.

Conclusions

The study of language and languages permeates contemporary human geography in a variety of different ways. It has been and still is an important object of enquiry. Various efforts have been made to map language use and to connect language to various forms of identity: national, transnational, and local. At the same time, a focus on linguistic issues has also impacted on more epistemological discussions within human geography. At one level, the fact that human geography is taught and researched through the medium of a variety of different languages raises important questions about the variation in the discipline from one linguistic context to another. Relatedly, post structuralist's attempts to work through the significance of language and discourse has been used as a means of challenging structuralist epistemologies and methodologies within the discipline. Taken together, the discussion shows the significance of linguistic con cerns for the practice of human geography in a number of different contexts.

See also: Cultural Politics; Discourse; Identity Politics.

Further Reading

Alderman, D. (2003). Street names and the scaling of memory: The politics of commemorating Martin Luther King, Jr within the African American community. *Area* 35, 163 173.

Bielenburg, A. (2000). Introduction. In Bielenburg, A. (ed.) *The Irish Diaspora*, pp 1 16. Harlow: Pearson Education.

Carter, S. (2005). The geopolitics of diaspora. *Area* 37, 54 63.

Clifford, J. (1997). *Routes: Travel and Translation in the Late Twentieth Century*. Cambridge, MA: Harvard University Press.

Desforges, L. and Jones, R. (2001). Bilingualism and geographical knowledge: A case study of students at the University of Wales, Aberystwyth. *Social and Cultural Geography* 2, 333 346.

Dixon, D. and Jones, J. P. (2004). Poststructuralism. In Duncan, J., Johnson, N. & Schein, R. (eds.) *The Companion to Cultural Geography*, pp 79 107. Oxford: Blackwell.

Grossberg, L. (2002). A special editorial for 16(1). *Cultural Studies* 16, 1 2.

Gutiérrez, J. and López Nieva, P. (2001). Are international journals of human geography really international? *Progress in Human Geography* 25, 53 69.

Johnson, N. (1997). Making space: Gaeltacht policy and the politics of identity. In Graham, B. (ed.) *In Search of Ireland: A Cultural Geography*, pp 151 173. London: Routledge.

Jones, R. and Fowler, C. (2007). Where is Wales? Narrating the territories and borders of the Welsh linguistic nation. *Regional Studies* 41, 89 101.

Kaplan, D. (1994). Population and politics in a plural society: The changing geography of Canada's linguistic groups. *Annals of the Association of American Geographers* 84, 46 67.

Kobayashi, A. (2004). Critical 'race' approaches to cultural geography. In Duncan, J., Johnson, N. & Schein, R. (eds.) *The Companion to Cultural Geography*, pp 238 249. Oxford: Blackwell.

Nash, C. (1999). Irish placenames: Post colonial locations. *Transactions of the Institute of British Geographers* 24, 457 480.

O Tuathail, G. and Agnew, J. (1992). Geopolitics and foreign policy: Practical geopolitical reasoning in American foreign policy. *Political Geography* 11, 190 204.

Townsend, J. (1995). *Women's Voices from the Rainforest*. London: Routledge. 1995.

Withers, C. W. J. (1984). *Gaelic in Scotland 1698 1981: The Geographical History of a Language*. Edinburgh: John Donald.

Zelinsky, W. (1994). *Exploring the Beloved Country: Geographic Forays into American Society and Culture*. Iowa City, IO: University of Iowa Press.

Relevant Websites

http://www.pobail.ie/en/
Department of Community, Rural and Gaeltacht Affairs, Ireland.

http://www.ocol clo.gc.ca/html/index e.php
Office of the Commissioner of Official Languages, Canada.

http://cymuned.net/blogsaesneg/?page id=21%22
Official website of Cymuned, the group campaigning to protect a Welsh speaking heartland.

http://www.celticleague.net/
The Celtic League.

Language and Research

A. Aylett and T. J. Barnes, University of British Columbia, Vancouver, BC, Canada

Glossary

Deconstruction It is associated with the post-structuralist, Jacque Derrida, and represents a technique to reveal the inherent instabilities and uncertainties of language.

Discourse A network of concepts, statements, and practices that collectively produce and authenticate particular knowledges and truths.

Hermeneutics The study and clarification of meaning wherever it is located.

Metaphor The transference of characteristics of one object to another, for example, waves in an ocean are like pulses of light.

Naïve Realism The idea that words are perfect representations of the world that they describe.

Narrative A particular kind of story in which its different elements are integrated, and through their integration become understandable one to another as well as to the whole.

Postmodernism A critical movement concerned to unseat the certainties of modernism by celebrating heterogeneity, openness, and novel forms of representation, including those found in language.

Post-Structuralism A post-World War II intellectual movement associated especially with a group of French intellectuals, one of whose concerns is with the unstable nature of language and its role in shaping subjectivity and identity.

Rhetoric The study of persuasive discourse: how words and expressions "influence men's souls" as Plato put it.

Text A cultural meaningful practice that can be interpreted and understood: from words to landscapes.

Introduction

Language is inescapable. To think about the meaning of language is already to use it. Furthermore, like Moliere's Msr. Jourdain who spoke prose all his life without knowing it, we use language all of our life, yet for the most part are unconscious of its operation. If we think about language at all, it is likely as a neutral and trans parent medium, a convenient means to connect the outside world of reality to the inside world of our mind. This larger view is known as naïve realism: the belief that pieces of language adhere to thoughts and objects in unassailable and essentially correct ways. Under this view, using language becomes simply the mechanical one of lining up words in the right order. All we need to know about a language is its vocabulary and rules of grammar. After that, it is like laying pipe, or electrical wire, "manual labor of the mind," as the novelist John Gregory Dunne once expressed it.

For much of the history of human geography, the disciplinary inclination has been either like Msr. Jourdain to avoid thinking about language, or to assert some form of naïve realism. Admittedly, there exists a long standing cottage industry, especially within American cultural geography, concerned with the categorization and map ping of languages and dialects, but even here naïve realism holds. Practitioners believe that language is like a material object, such as a barn or fence post, and its importance lies only in its physical geographical distribution.

Such a view has begun to change over the last 30 years, however. Starting in the late 1970s with political economy and humanistic geography, and from the 1990s with postmodernism and post structuralism, there has been increasing recognition of the utter centrality and complexity of language. Rather than simply reflecting the world, or our thoughts, language is riddled with social power, cultural norms, political imperatives, duplicity, and *non sequiturs*. As a topic, language is potentially as large as the discipline itself. Here, however, we confine ourselves to only four issues: a brief intellectual history of geographers' engagement with language; a discussion of various critical strategies used to reveal how language operates in geographical works; a review of substantive geographical studies of place, landscape, and language, frequently linked to issues of social power and identity; and finally, and perhaps the heir to those earlier cultural geography studies of the spatial distribution of languages, an assessment of the increasing hegemony of the English language as the *de facto lingua franca* of the discipline.

Geographers and Language: A Brief History

In many ways, naïve realism was made for geography. From its institutionalization in the late nineteenth cen tury, human geography was defined as descriptive, not explanatory, concerned with the concrete, and the par ticular, not the abstract and the general. Consequently, the naïve realist view accorded perfectly with the prac tices of geographers as they named, sorted, classified, and shuffled around bits of the world on pieces of paper. The culmination was American regional geography, and given justification by Richard Hartshorne in his tome *The Nature of Geography*. The discipline was defined there as a descriptive science of unique regions, and facilitated by a transparent language that could reflect without bias or distortion the exceptional character of the world.

Even though regional geography gave way sometime in the late 1950s to a scientific approach, 'spatial science', stressing explanation, abstraction, and generality, some version of naïve realism remained. Useful here is the distinction between natural and formal languages. Natural languages are organic and of the everyday, whereas formal languages are purposively designed for specialized ends. The quantitative revolution represented the replacement of an 'imprecise' natural language (hitherto used for describing unique geographical facts) by a more rigorous and exact formal language (used for general spatial explanation). Despite this, geographers still adhered to naïve realism. Language, albeit now as equations, numbers, and statistical formulae, mirrored reality. Indeed, the degree of correspondence could be scientifically proven to one significant decimal place or another.

The first signs of real change that geography might be ripe for a 'linguistic turn', which took language seriously, began in the 1970s, and associated with the introduction to the discipline of both political economy and a humanistic approach. Political economy emphasized the political nature of language – its power sodden character. Drawing upon Marx, David Harvey in *Social Justice and the City* argued that language, including the language of geography's quantitative revolution, was ideological – a smoke screen of words and symbols that obfuscated truth rather than revealing it. Harvey labeled such language 'counter revolutionary', contending that its purpose was only to forward the interests of a small elite at the expense of a large, powerless proletariat. For Harvey, language was not transparent but opaque, not innocent but culpable, skewed to meet the ends of only those with power. Only after the Revolution could language be redeemed.

In contrast, humanistic geography stressed less the issues of power and concealment than the fraught nature of establishing meaning within language. Drawing in particular on the German hermeneutic tradition, and rooting their substantive studies in ordinary places and landscapes, humanistic geographers in the important book *Humanistic Geography* edited by David Ley and Marwyn Samuels argued that interpretation was never straightforward; it was never merely reading labels stuck on the back of things. It was hard work, and, in the end, there was no end, and no final knock down interpretation. Words could always be otherwise, but there was no alternative. Language was the only game in town, and its perversity was what we were stuck with and stuck to. Moreover, unlike in Harvey's account, for humanistic geographers, there would never be redemption. Language was always wedged in an awkward corner, and we with it.

Gunnar Olsson drew brilliantly upon both political economic and humanistic positions to write geography's first book length treatment about the problems of language, *Birds in Egg*. Initially, an acolyte of the quantitative revolution, within a decade, however, Olsson was arguing that spatial science said less about the world it wrote about than the language it was written in. Like the humanistic geographers, Olsson thought we had no choice but "to bump our heads against the ceiling of language," but like the Marxists he thought that the resulting pain was distributed unevenly, favoring the powerful. His larger point: the language of geographers is neither naïve nor realist. Language is used strategically less to represent worlds than to make them, and in some cases to make them unjust.

Olsson's work was an important segue to post modernism and post structuralism, which made language central during the 'linguistic turn' that followed during the 1990s. As well as Ludwig Wittgenstein and Martin Heidegger, on whom Olsson had drawn, other key writers drawn into this later discussion by geographers included Michel Foucault, Jacque Derrida, and Jean Beaudrillard. Also on the margins of this work was the acknowledgment of a reinvigorated American pragmatism championed by Richard Rorty. While there were enormous differences among the positions of these various figures, there was agreement that language went all the way down. It was all that we have. For geographers, this resulted in an acute interest in how language worked, and its effects. Furthermore, to understand better the role of language, geographers began deploying techniques and methods from literary theory, such as discourse analysis and deconstruction. Such an interest was often combined with substantive studies concerned with showing how language operated in constructing especially place and landscape. In the process, geographers also pushed the limits of what counted as language, and thus what could be analyzed. There were written words, of course, but increasingly included were quite different texts such as equations and numbers, maps and atlases, and even architecture and building types. All were grist for the postmodern and post structural mill of analysis, interpretation, and scrutiny.

Methods and Techniques of Language Analysis

To understand how language operates, geographers have drawn upon a variety of literary and philosophical analyses. Four have been especially pertinent: rhetoric, metaphor, discourse, and deconstruction.

Rhetoric

Rhetoric is about understanding how words can be used persuasively. For the 'truth' of an argument does not necessarily win the day. It is the persuasive power of the language in which that truth is expressed that is critical. Specifically, under rhetoric, tropes such as metaphors, ironic asides, equations, jokes, citations, and anecdotes are the bases of authority and argument. This is in marked contrast to the belief of epistemology that avers, truth is established on the basis of a set of *a priori*, abstract criteria. For rhetoricians, however, truth emerges within only specific practices of persuasion on the ground, in discourse, and in conversation.

References to rhetoric by geographers have often been implicit (although explicit analyses do exist). Rhetoric was originally deployed within an offshoot of humanistic geography, the geography of literature, and concerned with how creative writers manipulate language to invoke the *genius loci* of a particular place or region. With much more of an edge has been work on how rhetoric has been used by geographers in their own writing to bolster arguments for maintaining the *status quo* or mounting intellectual change. Peter Taylor, for example, provided a compelling argument that the quantitative revolution in geography was successful because of its proponents' superior rhetoric and couched as a formal language (a theme taken up in a debate in 1994 between Trevor Barnes and Keith Bassett in *Environment and Planning A*).

Metaphor

Metaphor is one of the key elements of rhetoric. Arguments are persuasive when they are couched like an argument that has already been made in a different context. This is what metaphors do. They draw similarities between different contexts. The classic example is Isaac Newton's invocation of waves – hitherto reserved for describing and explaining liquids – to describe and explain quite a different context: light. Newton's larger strategy of likening one explanatory context to another is called 'metaphorical redescription'. Through metaphorical redescription, scientists and social scientists establish large (meta) frameworks of explanation that understands one class of phenomena in terms of the understanding of quite a different class of phenomena.

As an intellectual practice, metaphorical redescription is pervasive both in the sciences and social sciences. In the late nineteenth century, economists, for example, borrowed theories of physics to describe and explain the behavior of economic agents in rational choice theory. In geography, metaphorical redescription is found in theories of spatial interaction that likens it to gravity; in urban geography that, following the Chicago School, likens cities to plant biomes; and in industrial geography that likens economic regions to geological strata.

A problem with metaphorical redescription, however, are hidden assumptions frequently conveyed within the metaphor: postulates that might make sense in the original context, but are inappropriate in its new use. Consequently, some geographers have begun to 'exhume' geography's 'dead metaphors', inspecting them critically for their coherence, consistency, and compatibility in their new setting. Doing so often means scrutinizing the historical and material origins of the original metaphor that shaped its meaning. In this sense, taking metaphor seriously means "worlding them" as Neil Smith and Cindi Katz term it.

Discourse

The concept of 'discourse' further develops the analysis of language, especially its ability to create new truths and new worlds. Discourses are frameworks for understanding and directing different domains of social action. They are composed not only of specific vocabularies and rhetorical techniques, but also modes of thinking, grammars, rationalities, and even specific material practices that represent, interpret, and create new bits of reality. Although found in the structuralist works of Claude Levi Strauss, much contemporary discourse theory is based on the work of Michel Foucault that explicitly links power and language.

Foucault argues that by establishing criteria by which truth is judged, discourse is the primary channel through which power flows within societies, social groups, and academic disciplines. Power, communicated through discourse, facilitates certain modes of acting and thinking, while limiting others. Moreover, discourse functions at the level of taken for granted presuppositions about reality. As a result, discourses produce a reality by establishing what an individual can think, say, or do. In his work on criminality, insanity, and sexuality, for example, Foucault argued that discourses are not simply reflections, representations, of social phenomena, but define them, thus giving them a reality. Further, those discourses are held in place and given force by specific social institutions and practices of everyday life, as well as the physical sites (prisons, hospitals, and asylums) that embody them. As a vehicle for power, discourse is thus internalized by social actors, and supporting Foucault's contention that power is diffused across a multitude of self policing social actors, rather than exercised by a sovereign center.

Discourse analysis has been taken up across human geography, from political and economic geography to social, cultural, and environmental geography. It is possibly the most popular method of social analysis found within human geography. In every case, whatever the specific context, geographers are concerned to identify the constituents of the discourse, their origin, and most importantly their effects. The popularity of discourse as a critical concept, and the 'linguistic turn' more generally, has been criticized for sidelining the study of concrete problems on the one hand, and the continued role of capital and class in structuring power relationships on the other.

Deconstruction

Deconstruction is an approach to critical analysis developed from the writings of Jacques Derrida. Based on the Foucauldian argument that discourses are everywhere, deconstruction aims to destabilize them by a form of literary close reading that highlights their contradictions, exceptions, and *aporias*. Always attempting to find within a given discourse the means for its own critique, deconstruction pays close attention to the tensions between a discourse's stated meaning and the rhetorical techniques (underlying metaphorical assumptions, footnotes, and forms of argumentation) used to convey it. Deconstruction thus aims to reveal a text's multiple and contradictory meanings rather than searching for a stable and unitary meaning that corresponds to a truth or reality outside of language. Denying the possibility of this type of transcendental signification, deconstruction is based on a Saussurian conception of language as a 'system of differences' where, without a fixed center or truth to tether it, meaning can be made and remade through the infinite substitutions of 'freeplay'. In this way, deconstruction also challenges and destabilizes the relationship between power and knowledge established by Foucault.

There have been few sustained, focused studies making use of deconstruction in human geography in part because of the difficulty of such a task. Much more common is the invocation of deconstruction to identify contradictions within a given analysis, or a historical body of work and theory as in Derek Gregory's *Geographical Imaginations*. Sometimes, the word is used even more loosely by geographers simply to mean critique. Criticisms of deconstruction have focused on its perceived nihilistic rejection of rationalism and truth, and in geography at least often made by those within political economy.

Place, Landscape, and Language

In addition to these methodological reflections on language by geographers, there is also keen interest in the role language plays in constructing real places and landscapes, and, in turn, about how those processes are bound up with social power and identity.

Even early in the discipline's history, issues of language were important in describing place and landscape. For example, the British geographer A. J. Herbertson (1865–1915) and the French geographer Vidal de la Blache (1845–1918) both believed that there was an art to describing both. Words needed to be crafted, not mechanically bolted. Representing a place's *genius loci* in Herbertson's case, or a region's 'personality' in Vidal's, required delicacy, a deft writing hand.

Later work carried out from the early 1980s from both political economic and humanistic geographical traditions, however, was less concerned with the problems of using language to represent place and landscape than drawing upon literary theorists to understand how place and landscape were like a language – a text – that could be critically analyzed. Denis Cosgrove's *Social Formation and Symbolic Landscape* drew especially on Raymond Williams' Marxist literary theory to argue that social formations produced particular kinds of landscape. Williams had used the idea of a social formation to understand in his case the changing symbolic character of literature. Cosgrove did the same, except he was interested in the changing symbolic (later dubbed 'iconographic') form of the landscape. For example, the Italian Palladian landscape, or the landscape of English country houses, were symbols of an underlying social order based on private property and an associated inequitable distribution of power and resources. It was therefore necessary 'to read' the landscape, to decrypt its symbolic form, and grasp the social relations of power that lay beneath.

Treating the landscape as a text to be read also emerged from the humanistic perspective and which later melded into post structuralism and postmodernism (Roland Barthes was especially influential). In this work, though, there was no deep social formation that the text reflected. Instead, the text of landscape was a cultural and political product that emerged recursively from the specific context in which it was embedded, and which varied geographically and historically. That context might be late twentieth century inner city landscapes of Vancouver and two competing political visions, liberalism and conservatism, or the nineteenth century city of Kandy, Sri Lanka, and the dominance of a non negotiable religious order. As before, though, the point is to treat the landscape as a form of language. Consequently, its symbolism is readable through the prism of its larger context.

In addition to these studies of landscape, there have also been attempts to understand place in terms of meaning and language. Yi Fu Tuan's writing on *Topophilia*, and Edward Relph's book on *Place and Placelessness*, are both about how places achieve, or not, significance and import. In addition, Nicholas Entrikin in *The Betweenness of Place* explicitly draws upon the idea of literary narrative to make sense of place. For him, places are forms of 'emplotment', that is, narratives or stories that possess coherence such that one is able to make sense of how their individual elements are connected and joined. Place is not a haphazard collection of items, but holds together like an integrated story. Places are nar ratives to be told, and require to be analyzed as such.

Apart from attempts to analyze place and landscape as if they were a language, how people within a place and landscape actually use language has also attracted critical attention. Here language is treated as one of the inter esting constituent components forming place and land scape. Linguistic practices here act as a map of social divisions, structures of power, and forms of resistance situated within particular places. In his sociolinguistic geographies, Allan Pred, for example, shows how words both construct and reflect identities, occupations, and struggles that characterized the lives of particular people in particular places. Tracing changes in the lexicon of Stockholm's working class as they dealt with economic modernization, or the way in which language is impli cated in the racist practices of modern day Sweden, he argues that linguistic practice is central to the processes, institutions, and forms of consciousness that socialize place. Words and worlds are intimately linked.

Other works in this vein have shown how linguistic practices act within particular places and landscapes to create exclusion and intolerance. They have focused es pecially on gender, urban ethnolinguistic divisions, and racialized aspects of the cultural separation between rural and urban space. Politically, language is also deeply im plicated in an individual's or community's ability to claim citizenship, exercise rights, and participate in civil insti tutions. Reactions to these experiences of exclusion have led geographers to consider the formation of alternate, nonstate based, forms of political commitment and be longing, and operating across a variety of scales. Finally, there has been work on the way in which toponyms, that is, place names, reflect the interaction between culture, power, and place. Much attention has been given to changes in place names – whether in the nineteenth century Scotland, the West Bank, or in Latin America – and how it is bound up with power struggles over the control of place, meaning, and identity. Efforts to address these legacies by recovering and reinstating past place names have led geographers to consider the difficult questions around writing plural identities onto landscape.

The Language of Geography: English as *lingua franca*

A final issue is the power laden and non neutral nature of language as it functions within the discipline. Specif ically, there is a debate around the use of English as the discipline's *lingua franca*. At issue is the argument that the use of English within geography is not simply a neutral, convenient medium of communication. Rather, it is bound up with a specific worldview establishing flue ncy based geometries of power. That is, English as geography's *lingua franca* produces particular perceptions and depictions of the world, and our relationship to others, which would be different if another language was used.

The use of English as a spatializing practice has drawn the most criticism. While creating dominant spaces of communication centered in the English speaking world, the prevailing use of English also regulates academic practice beyond it through a deterritorialized and quasi universal system of norms and conventions. Specifically, within the English speaking community, a conflation is made between 'English' and 'international'. Such a con flation, however, obscures the fact that access to the 'international' is in fact controlled by a small number of journals dominated by and published in the English speaking world. Such an iniquity is further reinforced by key pieces in the structures of professional academia such as the Social Sciences Citation Index (SSCI), where non English speaking countries supply only between 5.1% and 7.4% of journal contributions.

The linguistic homogenization of the Anglo centric sphere of international academia also encourages an unwarranted conflation between English fluency and academic quality and legitimacy. In turn, this marginal izes the work of non English academics, and reduces their ability to shape agendas of publications and con ferences. Yet, another consequence of human geography's ironic blindness to its own place based character is the promotion of a specific vision of academia, and a cul turally specific set of epistemological concepts and areas of research. Consequently, the language and supporting academic institutions couched within the vocabularies of Anglo American social science are then presumed to be universally applicable. Terms like 'multicultural ism', 'immigrant', or 'postmodernism' circulate with and deploy specifically situated visions of what they describe, but take on the appearance of universality, thanks to the wide use afforded to them by the broad circulation of the English language.

Linguistic hegemony, and its associated disciplining practices of the kind we described, imagines and attempts to inscribe center–margin relationships between actors and locations, in this case, between an English Anglo America and a non English rest of the world. Such a model, however, has recently been called into question by issues of reception, resistance, and cultural significance, and which have shaped debates on cultural globalization more generally. Additionally, the core–periphery model has been troubled by the conscious adoption of English as a *lingua franca* by non native speakers, and used by them 'subversively' to bridge multiple linguistic groups and to facilitate participation in broader economic, political, cultural, and academic spheres.

Within this context, Claudio Minca argues that whether English operates as a tool of power or a bridge between diverse speakers is a question of volition. It is not the language itself but the contexts of power within which it is spoken that marginalize and homogenize. What is needed, he suggests, is the willingness to decontextualize and recontextualize knowledge and openness to sources of legitimacy outside the Anglo American tradition.

Conclusion

Much has changed since the naïve realist views of language that characterized geography's early engagements with language. As a consequence of the utilization of a variety of critical perspectives, human geographers no longer view language as a neutral vehicle, or simply verbal labels that adhere to the objects they designate. There is nothing simple about language or meaning. It is a vehicle for social power, a tool of persuasion, a means for representing, and producing particular realities, and central to a panoply of social institutions and practices that create space, place, and landscape. From these critical perspectives, and familiarity with the instruments of linguistic and textual analysis, geographers have understood the symbolic content of space, place, and landscape, and the social and political process that have shaped them. Language has been brought into geography, and, in turn, it has helped to illuminate the linguistic character of the geographical.

See also: Deconstruction; Discourse; Humanism/Humanistic Geography; Ideology; Landscape; Postmodernism/Postmodern Geography; Poststructuralism/Poststructuralist Geographies.

Further Reading

Barnes, T. J. and Duncan, J. S. (eds.) (1992). *Writing Worlds: Texts, Discourses and Metaphors in the Interpretation of Landscape*. London: Routledge.

Barnett, C. (2005). Life after Derrida. *Antipode* 37, 239 241.

Cosgrove, D. (1984). *Social Formation and Symbolic Landscape*. London: Croom Helm.

Curry, M. R. (1991). Postmodernism, language and the strains of modernism. *Annals of the Association of American Geographers* 81, 210 228.

Doel, M. (1999). *Poststructural Geographies: The Diabolical Art of Spatial Science*. Edinburgh: Edinburgh University Press.

Duncan, J. S. (1990). *The City as Text : The Politics of Landscape Interpretation in the Kandyan Kingdom*. Cambridge: Cambridge University Press.

Gregory, D. (1994). *Geographical Imaginations*. Oxford: Blackwell.

Minca, C. (2003). Critical peripheries. *Environment and Planning D: Society and Space* 21, 160 168.

Olsson, G. (1978). *Birds in Egg*. Ann Arbor, MI: University of Michigan.

Paasi, A. (2005). Globalisation, academic capitalism, and the uneven geographies of international journal publishing spaces. *Environment and Planning A* 37, 769 789.

Pred, A. (1990). *Lost Words and Lost Worlds: Modernity and the Language of Everyday Life in Late Nineteenth Century Stockholm*. New York: Cambridge University Press.

Smith, J. M. (1996). Geographical rhetoric: Modes and tropes of appeal. *Annals of the Association of American Geographers* 86, 1 20.

Smith, N. and Katz, C. (1993). Grounding metaphor: Towards a spatialized politics. In Keith, M. & Pile, S. (eds.) *Place and the Politics of Identity*, pp 67 83. London: Routledge.

Valentine, G. and Skelton, T. (2007). The right to be heard: Citizenship and language. *Political Geography* 26, 121 140.

Latin American Structuralist School

Cristóbal Kay, Institute of Social Studies, The Hague, The Netherlands

Glossary

Keynesian Economics It refers to the body of ideas inspired by the British economist John Maynard Keynes (1883–1946) particularly as set out in the *General Theory of Employment, Interest and Money* in which he argued that it was the duty of governments to achieve and maintain full employment. In the economic development literature, his ideas are used to justify state intervention in the economy due to the existence of market failures and the limitations of markets for achieving certain developmentalist goals.

Neoclassical Economics It originates largely from the writings of Alfred Marshall (1842–1924) and was subsequently developed in the twentieth century by John Hicks and Paul Samuelson, among many others. It believes in the economic rationality of individuals and in the efficiency of free markets. It argues against state intervention in markets (with some exceptions) as this leads to inefficiency and lower rates of economic growth. Today the term neoliberalism is more commonly used to refer to the promotion of free market economic policies.

Terms of Trade An index calculated by dividing an index of export prices by an index of import prices. It measures the purchasing power of exports in terms of imports.

The Latin American Structuralist School is one of the first major attempts by scholars and development practitioners from the developing world to systematically analyze the development processes and problems of their region from their own perspective instead of relying on those from the developed world. The foundational stone was laid by the Argentinean economist Raúl Prebisch with a study on Latin America, which he wrote in 1949 for the then recently established United Nation Economic Commission for Latin America (ECLA), later expanded to ECLAC to include the Caribbean. This document which created quite a stir at the time, due to the originality of its ideas and forcefulness of its policy recommendations, has had an enduring impact. Hirschman refers to it as the 'ECLA Manifesto'. In 1950, Prebisch was appointed as Executive Secretary of ECLA, which was located in Santiago de Chile. He gathered around him a group of young, bright social scientists (largely economists) from different countries of the region. The resulting studies on various development issues gradually gave shape to a new framework for analyzing development problems, a new vision of development and a new set of policy recommendations for the region.

Influences on, and of, Structuralism

The key ideas of what became known as the Latin American structuralist development school, usually shortened to 'structuralism', were formulated from the late 1940s to the 1960s, largely by ECLA staff. Latin American structuralism is also referred to as *desarrollismo* ('developmentalism' or 'growthmongering'), the 'center–periphery' paradigm or *cepalismo*, that is, CEPAL (Comisión Económica para América Latina) being the Spanish acronym of ECLA. During this period ECLA became increasingly influential not only in academic circles but particularly in government circles throughout the region. It provided a series of advisory services to government policy makers and trained hundreds of civil servants in development theory and policy, planning techniques, and economic and social statistics.

Besides Prebisch, the main contributors to the Latin American structuralist school are the Brazilian Celso Furtado, the Chileans Aníbal Pinto and Osvaldo Sunkel, and the Mexicans Juan Noyola and Victor Urquidi, all of whom are economists or political economists. Prebisch remained as head of ECLA until 1962. He then assumed the directorship of ECLA's sister organization, the UN's Latin American Institute for Economic and Social Planning (which goes by its Spanish acronym ILPES), which is located in the same compound as ECLA, until 1964. Subsequently, he was called on by the General Secretary of the UN to lead the United Nations Conference on Trade and Development (UNCTAD) in Geneva, which he did until 1969. Upon his retirement, he continued to be associated with ECLA by becoming 'in 1976' the founder and first director of its journal *CEPAL Review*, an association which continued until his death in 1986 at the age of 85. He was undoubtedly the pillar of structuralism. Furtado, who worked for 8 years in ECLA from 1949 to 1957, was the second most distinguished contributor to structuralism. His prolific writings are distinguished by their historical dimension. Although most of his structuralist writings were produced after he left ECLA, his passage through it shaped much of his thinking and he continued to exchange ideas with many of the ECLA staff throughout his lifetime.

Structuralist ideas are not confined to Latin America. Some elements of structuralism – such as the specific nature of underdeveloped regions and the problems they faced, the shortcomings of neoclassical theory, the prevalence of market failure, the need for state intervention to overcome the limitations of the market and to stimulate a process of industrialization – can be traced back to thinkers active before or during World War II, as imaginatively done by Joseph Love. Love also painstakingly explored the possible origins of Prebisch's ideas but much of this remains speculative, as Prebisch revealed little about his sources either in his writings or in the few interviews he gave. Be that as it may, the following Northern development economists can be mentioned as influencing Latin American structuralism: Albert Hirschman, Michal Kalecki, Kurt Mandelbaum (later Martin), Ragnar Nurkse, Paul Rosenstein Rodan, and Tibor Scitovsky. Some had studied the development problems of pre war Eastern Europe, which shared some characteristics of underdevelopment with Latin America. Some of these economists might not describe themselves as structuralists, as they were largely a heterodox group of economists influenced by Keynesianism, institutionalism, or Marxism. Furthermore, in the post war period the works of development economists such as Nobel Prize winners Arthur Lewis, Gunnar Myrdal, and Jan Tinbergen as well as the writings of Hollis Chenery and Paul Streeten share some affinities with structuralism. Although mainly influenced by Keynes, Hans Singer became associated with structuralism through the Prebisch–Singer thesis on the terms of trade.

Among the Anglo Saxon development economists, Dudley Seers is probably the one who most closely identified with, and contributed to, structuralism. His experience in ECLA from 1957 to 1961 transformed him from an economist specializing in economic statistics into a development economist who embraced structuralism and diffused it in the Anglo Saxon world and beyond. He was one of the first economists from the North to appreciate the originality and importance of structuralist ideas as can be gauged from his assessment "that the social sciences of Europe, especially in the periphery and semi periphery, would benefit from drawing on those of the 'Third World', especially Latin America (which experiences somewhat similar problems): these would be more useful than the social sciences of the 'core' countries."

The Role of ECLA and Leadership of Prebisch

Structuralism grew out of the desire to overcome Latin America's underdevelopment. The mainstream economic theories of the time, principally neoclassical and Keynesian, were found to be inadequate for this purpose as they did not reflect the reality of underdeveloped countries. In the 'ECLA Manifesto', Prebisch had already taken to task the classical theory of international trade based on comparative advantages for being flawed in its analysis on the gains from international trade. Following such a diagnosis, it was deemed imperative to gain a better understanding of the structures and processes shaping the Latin American economies and societies so as to be able to design more appropriate development policies. Such an enterprise required empirical investigations as well as new analytical categories and theories. While undoubtedly other critical social scientists in Latin America and beyond had reached similar conclusions, to ECLA's credit it took the initiative to overcome this deficiency by systematically building a database for each individual Latin American economy as well as for the region as a whole, producing annual reports on the state of the economy and later on the social conditions of the region. ECLA's analyses are firmly grounded in this empirical base and have fundamentally a practical purpose, as its mission as one of the regional organizations of the United Nations was to advise governments on economic and social policy. Thus, its mission differs from that of universities which have (or should have) greater freedom for critical thinking and whose role is more centrally concerned with the production of knowledge and theory. Being a United Nations organization imposes various limitations. For example, the studies of ECLA had to tread very carefully in their critiques of certain government policies such as the US. But Prebisch had the vision, enterprise, diplomatic skill, and intellectual clout to overcome as far as possible those limitations and to exploit the advantages afforded by an international organization belonging to the UN. He thus managed to build ECLA into the first major regional think tank as well as creating an epistemic community which gave birth to the Latin American structuralist school of development.

The special working conditions at ECLA facilitated the emergence of structuralism with its broad regional and international vision. ECLA enjoyed UN backing with its reasonably generous funding relative to other specialized development research institutions in the region, good employment conditions, a location in Chile which was one of the most stable democracies in Latin America and which had experienced a variety of original development models, and a multinational staff who shared both a common Spanish language (most Brazilian professionals understand and speak Spanish) and a common Iberian cultural heritage. Under the strong leadership of Prebisch, the talents of the ECLA staff were able to flourish and coalesce around a common mission.

The Structuralist Method of Analysis

To gain a better understanding of the Latin American development problem, the structuralists employed a method of analysis which was (1) historical – in the sense that to fully understand contemporary problems it was deemed necessary to explore their historical roots; (2) totalizing, as the analysis went beyond the local and national dimension to embrace the multiple links with the global and international system; and (3) structural, by arguing that the particular economic, social, and political structures of Latin America had been transformed by the colonial encounter and its characteristics and processes thereby differed from those of the developed countries. In this method, the notions of structure, process, and system are key elements in the analysis of the dynamics of underdevelopment and development. Structuralism, like development studies, also aimed to adopt an inter disciplinary or multidisciplinary approach. Although ECLA incorporated social issues in their studies, its failure to explore fully the class and political dimensions shaping the development process were due in part to the constraints of being a UN agency, which was supposed to remain neutral and confine its analysis and policy rec ommendations to the technocratic aspects.

In the structuralist view, the underdeveloped and developed countries are partial, but interdependent, structures which form one world system. The principal difference between these partial structures is that the developed countries have a greater degree of autonomy in shaping their development process, while the under developed countries are far more constrained in their development options as their development process is partly shaped by the dominant developed countries. Thus, the interdependence is between unequal partners as their interlinkages mainly benefit the dominant partner.

Thus, the structuralist approach differs not only from neoclassical growth theories but also from stage theories of economic growth as postulated by Walt W. Rostow and by sociology of modernization theorists for whom the underdeveloped countries are just at an inferior or earlier stage in their evolution than the developed countries. They thus postulate a similar path of development for the less developed countries to that taken by the more de veloped countries such as Great Britain or the USA. Meanwhile for the structuralists, the state of under development is part of the same global historical process which led to the development of the now developed countries. Therefore, underdevelopment is not part of a stage or of an evolutionary modernization process through which developed countries once also traversed, as both are part of the same process of global transformation.

Underdevelopment and development are thus two faces of the same world historical process, as both pro cesses are historically simultaneous and interlinked. In this account, the development strategy of the under developed countries should aim to seek a more equal relationship so as to enhance their capability for transforming their economy and society in the pursuit of higher standards of living for the majority of the popu lation. The gist of these ideas was first formulated by Prebisch in his seminal analysis of the unequal ex change between the 'center' and the 'periphery' as he referred to the developed and underdeveloped countries, respectively.

Unequal Exchange between Center and Periphery

The cornerstone of structuralism is Prebisch's center–periphery paradigm. Prebisch argued in his seminal text, published originally in Spanish in 1949, that the world economy is composed of a small number of core indus trialized countries, which he labeled as the 'center', and a large number of weaker nonindustrialized countries, which he labeled as the 'periphery'. While the center exported industrial commodities to the periphery, the latter exported primary commodities such as tropical agricultural products and mineral products to the former. According to the classical theory of international trade, such a specialization in production and international trade was to the benefit of both countries concerned as it exploited their respective comparative advantages. Fur thermore, according to this theory, the benefits of inter national trade would go mainly to the lower income countries and thus international trade would gradually close the income gap between the developed and the developing countries. However, Prebisch discovered that this was far from being the case. On the contrary, the income disparities between the center and the periphery were not only perpetuated but even widened through international trade. This situation arose largely because of the chronic long term tendency for the terms of trade of the primary producing countries of the periphery to deteriorate. This meant, for example, that over the years the periphery had to export an increasing quantity of primary products in exchange for the same amount of imported industrial commodities. There was thus a transfer of an economic surplus through the international market from the periphery to the center, the very op posite to what was argued by the theory of comparative advantages.

According to Prebisch, this long term deterioration in the terms of trade is due to the different structural characteristics of the product, labor, and capital markets in the center and periphery. Furthermore, technical progress is more intense in the center as its occurrence is more common and easily achieved in the industrial sector as compared to the primary producing sector. However, the greater organization of labor as well as the higher degree of monopolistic or oligopolistic power of indus trial firms in the center means that both capital and labor are able to retain most of the benefits derived from the increases in productivity through higher profits and wages. By contrast, in the periphery the existence of surplus labor and its weaker bargaining power combined with the higher degree of competition among primary producing enterprises means that, even though technical progress is more limited, most of its benefits are transferred to the center countries which import those primary commodities. As a consequence, the price of the primary commodities exported by the periphery to the center falls in relative terms compared to the price of the industrial commodities exported by the center to the periphery. As Singer (1950) aptly put it: "while the developed countries have the best of both worlds, as they are able to retain the benefits of their own technological progress as well as capture the benefits of the technological progress of the developing countries, the opposite is the case for the developing which have the worst of both worlds as they do not benefit from the technological progress generated in the developed countries nor are they able to retain the benefits of their own technological progress."

Singer reached conclusions similar to Prebisch, but he mainly explained the deterioration as arising from the lower income elasticity of demand for primary products as compared to that for industrial products. Thus, as incomes rose in the center they spent a proportionally declining share of their increased income on importing primary commodities from the periphery, while in the periphery an increase in income resulted in an increasing share of this increase being channeled toward importing industrial commodities. Thus, the demand for industrial products rose faster than the demand for primary products as incomes increased. This, among other factors, led to an increase in the price of industrial commodities as compared to primary commodities.

As both authors argued that the terms of trade deteriorated, a highly radical statement to make at the time as it was contrary to established international trade theory, their analysis became known as the Prebisch–Singer thesis. They were fiercely criticized by orthodox economists. While the Prebisch–Singer thesis has its flaws, it has stood the test of time remarkably well, although subsequent research has shown that the deterioration of the terms of trade has been less than originally estimated and for reasons which differ in part from those originally given. It should be borne in mind that neither Prebisch nor Singer argued that the periphery derived no benefits whatsoever from international trade but that the benefits of international trade were unequally distributed and seriously hampered the development efforts of the periphery.

When Prebisch became the secretary general of UNCTAD in 1964, which was his brainchild, he projected the structuralist ideas to the global stage. At UNCTAD, he hoped to be able to negotiate a more equal distribution of the benefits of international trade through a new trade policy for development. The aim was to usher a 'New International Economic Order' that would stabilize and improve the prices of the primary commodity exports from developing countries, open the doors in the North to industrial exports from the South, dismantle the agricultural subsidies to farmers in the North as well as the protectionist measures for agriculture in the North, and regulate the behavior of multinational corporations so as to ensure greater benefits for the host in the South. However, in 1969 Prebisch unexpectedly left UNCTAD as he felt increasingly frustrated with the inability of UNCTAD to make much headway toward creating a more equitable world due to differences between the countries of the South, but mainly due to the opposition from countries of the North, especially the USA, to measures which they perceived as being against their economic and political interests.

Some of the structuralist thinkers developed during the second half of the 1960s the structuralist stream within dependency theory (the other being the better known Marxist variant); structuralism, as well as dependency theory, had a major influence on world system theory which made use of the structuralist center and periphery terminology.

From an 'Outward-Oriented' to an 'Inward-Oriented' Development Process

To create higher rates of technological progress as well as retain a greater proportion of the benefits derived from technological progress, Prebisch became an ardent advocate of the industrialization of the periphery. This was to be achieved through protectionism such as tariff barriers which discouraged the import of certain industrial commodities so as to encourage their production locally. Such an industrialization process was referred to as 'import substitution industrialization' (ISI) as previously imported industrial commodities, principally consumer goods, were to be produced within the country. Other complementary measures, which largely involved state investment, were the development of the country's infrastructure such as roads, ports, electricity, housing and sanitation, as well as the establishment of some industries, like steel plants, which supplied key inputs to the new domestic industries. Usually, a state development bank was established to supply cheap credit to the emerging industrial entrepreneurs, as well as a state development corporation and/or a ministry for industry with the function of supporting the industrialization process with planning, investment, and technical advice.

In short, structuralists proposed a radical redirection in the region’s development process and strategy, shifting away from the ‘outward oriented’ development pattern which had existed since the colonial period toward an ‘inward oriented’ development pattern. Thereby the dynamics of the economy were to be redirected away from the production of primary commodities for the external market to the production of industrial commodities for the domestic market. Prebisch drew a lot of fire from the economic orthodoxy for daring to question their authority, as well as from the establishment in the US and other developed countries who saw him as a dangerous man, as they believed that his ideas would jeopardize their industrial export markets and bring higher prices for their imports of primary commodities. However, Prebisch was no Marxist or communist and was far from advocating revolutionary or socialist changes. Rather he was a passionate defender of the interests of developing countries and saw the need to reform the capitalist system to make it more egalitarian.

Latin America’s development pattern did shift to an inward directed ISI process. Without in any way denying the influence of the structuralist school on government policy, which certainly embraced an ISI strategy, many Latin American economies had already begun such a process as a consequence of the Great Depression of the early 1930s. But the structuralists did provide a useful rationale for governments to accelerate this process. When during the 1960s the ‘easy phase’ of ISI had been achieved and the process was giving signs of ‘exhaustion’, the structuralists recommended that governments should implement land reforms to improve agricultural efficiency. The resulting increase in food supplies would keep wages low as food prices would not rise, and increase profits and thus investment. It would also improve income distribution in the countryside and raise rural living standards thereby widening the internal market for industrial consumer goods giving a new boost to industry. However, agrarian reforms were only implemented in a limited way in some countries and failed to live up to expectations.

Structuralists also proposed to improve the efficiency of industry by the formation of regional common markets which would take advantage of economies of scale. Although some regional integration schemes were established, they generally failed to prosper. Structuralists were also early proponents of industrial exports so as to generate the necessary foreign exchange to import the required inputs for industry as well as for making it more competitive. But industrial exports failed to take off owing to insufficient government incentives, protectionist tariff barriers in the developed countries, and the lack of competitiveness. Governments failed to act on the warnings from structuralists about excessive protectionism and the lack of competitiveness of industry which they had already aired in the early 1960s. But by then the newly established industrial bourgeoisie had become too powerful and governments acquiesced to their continual demands for higher tariff barriers to protect their increasingly inefficient domestic industry.

The Rise of Neoliberalism and the Renewal of Structuralism

The mounting difficulties faced by the region’s industrialization process together with the debt crisis of the 1980s and the failure of governments to control inflation facilitated the rise of neoliberalism. Neoliberal economic ‘structural adjustment’ policies were implemented in most Latin American countries during the 1980s, except for Chile, where they had already been introduced in the mid 1970s. Neoliberal policies focused on achieving macroeconomic stability by controlling inflation and reducing government expenditure, by privatizing state enterprises which were regarded as particularly inefficient, and by liberalizing the economy by drastically reducing tariff barriers and opening the economy to the world market. Under neoliberal policies, the Latin American economies shifted back to an outward oriented development strategy in which the world market would determine the region’s comparative advantages.

The rise of neoliberalism, which also finds expression in the ‘Washington Consensus’, has led to the decline of structuralism. However, structuralist thinkers, largely from current or past ECLAC stock, began to respond to the neoliberal challenge and by the 1990s a series of publications appeared outlining a neostructuralist position. Neostructuralists take account of the new reality brought about by neoliberal globalization. They believe that through a state guided restructuring of the country’s links with the world market it is possible to achieve benefits for the majority of the population instead of ‘as at present’ for a privileged minority. This will require an internal restructuring of the economy and a major effort at technological innovation so as to achieve genuine and systemic competitiveness which is also sustainable, as it would not just rely on cheap labor and the export of unprocessed raw materials, as is currently largely the case, but on higher value added products. Thus, the state would have to undertake major investments in education and health so as to improve the quality and skills of the labor force as well as reducing poverty and promoting greater equity.

Neostructuralist ideas are beginning to have some influence on several of the more progressive governments in the region. To what extent neostructuralism will emerge as a genuine alternative to neoliberalism, instead of being co opted as the human face of a reformed neoliberalism, remains to be seen. Many of the key ideas of Prebisch, who became a fierce critic of neoliberalism, and those of the structuralists continue to illuminate the present, especially as the agenda for a more equitable and just international economic order has become more ur gent than ever.

In conclusion, the Latin American structuralist school has had important repercussions in theory and practice. It had the merit of opening the door to a critique of neo classical theory and made a major contribution to the emerging field of development studies as well as to de pendency and world systems theory. Structuralism or its current neostructuralist version might even offer the seeds for an alternative to the still hegemonic, although much questioned, neoliberal paradigm.

See also: Dependency; Development I; Development II; Modernization Theory; Neoliberalism and Development; Structural Adjustment; World-System.

Further Reading

Clarke, C. (2002). The Latin American structuralists. In Desai, V. & Potter, R. B. (eds.) *The companion to development studies*, pp 92 96. London: Arnold.

Colclough, C. (1991). Structuralism versus neo liberalism: An introduction. In Colclough, C. & Manor, J. (eds.) *States or markets?: Neo liberalism and the development policy debate*. Oxford: Clarendon Press.

Furtado, C. (1964). *Development and Underdevelopment: A Structural View of the Problems of Developed and Underdeveloped Countries*. University of California Press: Berkeley, CA.

Gwynne, R. N. and Kay, C. (2004). The alternatives to neoliberalism. In Gwynne, R. N. & Kay, C. (eds.) *Latin America transformed: Globalization and modernity*, pp 253 267. London: Hodder Arnold and New York: Oxford University Press.

Hirschman, A. (1961). Ideologies of economic development in Latin America. In Hirschman, A. (ed.) *Latin American issues: Essays and comments*, pp 3 42. New York: Twentieth Century Fund.

Kay, C. (1989). *Latin American theories of development and underdevelopment*. London: Routledge and New York.

Love, J. L. (1994). Economic ideas and ideologies in Latin America since 1930. In Bethell, L. (ed.) *The Cambridge history of Latin America, Vol. VI, Part 1, 1930 to the present*, pp 393 460. Cambridge: Cambridge University Press.

Love, J. L. (2005). The rise and decline of economic structuralism in Latin America: New dimensions. *Latin American Research Review* 40(3), 100 125.

Prebisch, R. (1950). *The economic development of Latin America and its principal problems*. Lake Success, New York: United Nations.

Prebisch, R. (1984). Five stages in my thinking on development. In Meier, G. M. & Seers, D. (eds.) *Pioneers in development*, pp 175 190. New York and Oxford: Oxford University Press.

Rodríguez, O. (2001). Prebisch: The continuing validity of his basic ideas. *CEPAL Review* 75, 39 50.

Seers, D. (1983). *The political economy of nationalism*. Oxford: Oxford University Press.

Singer, H. W. (1950). The distribution of gains between investing and borrowing countries. *American Economic Review* 40(2), 473 485.

Sunkel, O. (ed.) (1993). *Development from within: Toward a neostructuralist approach for Latin America*. Boulder, CO: Lynne Rienner Publishers.

Toye, J. and Toye, R. (2004). *The UN and global political economy: Trade, finance, and development*. Bloomington, IN: Indiana University Press.

Relevant Websites

http://www.eclac.cl
Economic Commission for Latin America and the Caribbean.

Law and Law Enforcement

D. Delaney, Amherst College, Amherst, MA, USA

Introduction

The relationship between law and human geography is both complex and, until recently, relatively unexamined. But the traces of legality are ubiquitous in landscapes. One need only consider the pervasiveness of official graffiti (the 'no trespassing' signs, 'no parking' signs, markers of juris dictional borders) and the density of legal rules and regulations that structure the spatialities of everyday life to recognize that law is everywhere. But, as we know, law is not everywhere the same. Elements of the legal, most obviously rules and rights, are among the primary social devices for giving spatial expression to significant di mensions of social difference. Further reflection on the legal constitution of property, home, public space, the city, the state, and 'the international' as entities that are in trinsically both spatial and legal is sufficient to suggest the thickness and complexity of legal geographies. The aca demic study of law and geography is primarily concerned with investigating the relationships of geographical or socio spatial phenomena to forms of social power and modes of meaning production associated with distinctively legal practices, institutions, ideologies, and forms of con sciousness. Insofar as geographers are interested in how human worlds are created, maintained, and transformed, the domain of the legal is of fundamental significance to the hows and whys of world making.

The Convergence of Law and Geography

Defining the boundaries of the field of inquiry is some what arbitrary. In the absence of institutional structures, or specialized journals, 'legal geography' is a relatively *ad hoc*, loose collection of scholars who work in a number of disciplines. For present purposes this field will be defined as consisting of self identified scholars who are engaged in sustained research on the general topic. It includes both geographers and scholars in other fields who have made a significant contribution and who regularly cite the work of professional geographers.

Within this amorphous cluster two principal orien tations can be identified. They are broadly distinguished by differing assumptions about the legal, the geographical, or spatial; implicit epistemological and methodological ap proaches; and normative or political commitments. These orientations will be identified as 'conventional' and 'crit ical' with no implication as to the relative merits of either.

Conventional approaches are so called simply be cause they operate within conventional understandings of law, geography, and scholarship. The objective of work in this vein is not to 'problematize' or 'theorize' so much as it is to describe. The methods are those of con ventional social science. Among its common tasks are identifying the geographical effects of changing rules or bringing to light differences among legal regimes through the methods of comparative law. This work may be reformist insofar as it seeks, through the methods of social science, to identify problems and propose more or less technical solutions. Such an approach is prominent in studies of land use, zoning law, and planning. Critical legal geography is characterized by a more self conscious engagement with social and legal theories and philosophies. It is oriented toward problematizing the (socio)legal and the (socio)spatial in terms of each other. It is often explicitly critical of existing social and political arrangements. This work is done by self identified progressive or radical scholars more interested in demystifying legal ideologies than in proposing marginal technical adjustments.

Until the 1980s the study of the connections between the legal and the geographic or spatial was undeveloped. That aspect of this relationship might be problematic or worthy of academic investigation or that intellectual engagement with this relationship might yield significant insights about human geography or law were simply not recognized. This was less the result of how the world was put together than of how knowledge about the world was organized. Legal scholars and human geographers had little in the way of common interpretive frameworks or vocabularies. This was despite the fact that legal schol arship and human geography each had strong inter disciplinary traditions upon which to build. The subfield organization of Anglo American human geography and its early articulated mission of synthesizing knowledge of history, politics, economics, and culture with space or place did not include a place for law. Likewise, at least in the United States, even though legal scholarship has had a robust tradition of 'law ands' (law and economics, law and sociology, law and politics, legal history, legal an thropology) going back to at least the 1920s, it did not recognize the value of geographical scholarship for ex ploring legal questions.

In the 1960s the world changed, and the perceived tasks of scholarship began to change as a result. In the 1970s the disciplines changed. Beginning in the 1980s, and gathering steam since, these changes were sufficient to bring the concerns, resources, and debates of law and human geography into each other's orbits. In the early 1970s the scope of geographical inquiry and the concerns of researchers expanded to include sustained engagement with social theoretical and philosophical resources that had previously been neglected. For present purposes the most significant development was the emergence of Marxist geography and other radical and critical streams of scholarship such as feminist geography that followed it. These streams brought issues of power, inequality, and social justice to the forefront of research agendas. They also built on well developed traditions of state theorizing. Important at this time as well were newer humanistic and phenomenological approaches to geographical inquiry. This strand eventually blossomed to include post structuralist and postmodernist modes of theorizing space. Significant here were increasingly sophisticated concerns with issues of meaning, interpretation, discourse, and problems of representation.

As a result of these two new streams of thought, geographers began to engage with novel questions and made use of ever expanding theoretical and methodological resources to bear on what, by the late 1970s, we recognized as the problems with space; that social space was not simply found but made, constructed, or produced; and that the conditions and effects of how it was produced had everything to do with the problems of power. It is important to recognize that these new perspectives and practices entailed more than simply 'importing' the vocabularies and debates of other disciplines into human geography. By re invigorating geographical inquiry, scholars came to a heightened awareness of the difference that space makes to a wide variety of social phenomena and problems. This, in turn, contributed to what has been identified as 'a spatial turn' in other disciplines which had the further effect of forging new interdisciplinary links among scholars across the academy.

These developments in human geography were paralleled by significant transformations in legal scholarship during the same period. While human geography was opening up to newer outside influences, so was Anglophone legal theory. Of great significance was the emergence of critical legal studies (CLS). Like radical and critical geographers, critical legal scholars took sustenance from the widening pool of critical social theorization and philosophy, including Marxism and post structuralism, and brought these to bear on legal analysis. As in human geography, issues of ideology, power and disempowerment, legitimation, and injustice became prominent in ways that departed significantly from conventional practices. Most importantly, critical legal scholars refused to accept the self legitimizing story about 'The Law' that orthodox scholars commonly told. Instead, they engaged in operations that they described as 'trashing', or revealing legal practices and artifacts to be quite the opposite of what they are conventionally described as being. Far from the general, neutral rules promulgated for the public good by public servants in efforts to create order, serve justice, and promote liberty ('the majesty of the law' story), the products of legal machinery were more clearly identified as deeply complicitous in structures of domination and subordination. Because the ordinary workings of liberal legal ideology frequently obscure this complicity, the intended result of trashing was de mystification. To the extent that liberal legal ideologies colonized everyday forms of social consciousness, critique was oriented toward de colonization. Relying on a more sophisticated theory of ideology and legitimation than can be presented here, most critical legal scholars also rejected a crudely 'instrumental' view of law and its relation to oppression and exploitation that reduces law to a blunt object with which the ruling class subordinates others. Rather, as legal scholars, they recognized that the workings of power through law do not always map one to one with the immediate interests of the powerful against those of the powerless, and that the relatively disempowered may also successfully deploy elements of the law (albeit less frequently) against the relatively empowered.

Reflecting the same broad social changes that resulted in the radicalization of some human geographers, other critical strands of legal scholarship, such as legal feminism and critical race theory, broadened the field of critique and spawned important internal debates. CLS also grew out of and engaged with a re invigorated project of socio legal studies – or the sociological study of law from outside the legal academy. As a result, many sociologists, anthropologists, political scientists, and historians began situating research projects within the new critical paradigms. Some were also influenced by the simultaneous spatial turn in the human sciences. By the 1990s legal geography emerged as a more self conscious endeavor. The work of Gordon Clark, especially his 1985 book, *Judges and the Cities*, was extremely influential in bringing these streams of scholarship together. Nicholas Blomley's *Law, Space and Geographies of Power*, published in 1994, was another seminal contribution. At present legal geography is less a subfield of either law or human geography than a trans disciplinary project as likely to be pursued by scholars outside either discipline as inside. It is also genuinely international. By early twenty first century there were a number of books, special issues of journals, symposia, an interdisciplinary reader, and countless articles and conference papers devoted to illuminating the complex interconnections and constitutive dynamics of the (socio)legal and the (socio) spatial. This diffuse pluralism has been very productive. However, it does give this endeavor a strongly *ad hoc* flavor.

Seeing the Legal in the Spatial

It is perhaps easiest to begin to appreciate the entanglements of the legal and the geographical through the simple observation of the density and complexity of formal legal rules that give form to the social spaces of everyday life. Consider 'home'. Even when one brackets issues of property and ownership, there is a multitude of legal directives governing, for example, who can enter a home under what conditions, what is permitted and forbidden on the inside, what may happen to a person who violates these strictures, and much more. That is, 'home' is a complexly constructed legal space and legal place. In the United States some of these rules derive from interpretations of the fourth amendment of the US Constitution. Other jurisdictions may derive similar rules from common law, case law, or statutes.

In Canada this is structured with reference to Section 8 of the Charter of Rights and Freedoms. The fourth amendment, and the thicket of legal commentary and interpretations of it, proclaims rules about when the police may enter a home, when a warrant from a judge is required and not required, what constitutes a 'search' or 'seizure', what may happen to evidence obtained in violation of these more primary rules, and so on. These rules, of course, must be interpreted and, thus, are revisable through interpretation. As rules and exceptions are re interpreted, rights are redistributed and the legal meanings of social space vis a vis power are transformed.

Other legal rules that constitute 'the home' as a legal space concern the actions of 'strangers' who are not state actors, for example, rules about trespass and burglary. Still other rules govern the intimate relations within home space, such as those relating to domestic violence (e.g., when the police may or must intervene, when a protective order is granted, and what the consequences of violating an order are) or child welfare. When one adds to these, the innumerable property regulations, rules governing aspects of the landlord/tenant relation, building codes, lending and insurance laws, civil rights legislation and more, the apparently 'simple' legal space of the home can be seen to be quite complex indeed. To some extent, these rules, and the spatial distribution of rights and powers that they effect, are obvious and well known. They are simply part of 'legal common sense'. But many rules are not well known. They may be obscure, highly technical, and known by only a few specialists. They may be esoteric (as is 'the castle doctrine', or 'the fruit of the poison tree doctrine'). They may be ambiguous or in tension with each other. This resolution, of course, requires interpretation, and relevant meanings may be amenable to a number of plausible interpretations. This conditions the politics of legal and spatial interpretation. Lawyers may be involved in crafting interpretations, judges involved in validating or invalidating them, prosecutors and the police in enforcing (or evading) them. Insofar as these practices are implicated in the interpretive restructuring of social space, they are of interest to legal geographers.

Generic and specific legal spaces (encompassing a wide range of instances, including property lots, municipalities, states, wilderness areas, drug free zones, indigenous reserves, and free trade areas) are configured in complex ensembles and constellations, and these constellations – and the interrelations among spaces – are the products of legal spatializations. They are the outcomes of specifically legal processes, practices, and performances through which distinctively legal meanings are embedded in landscapes. Recognition of these processes raises questions concerning how, why, and to what effect such spatializations and re spatializations are accomplished. An orthodox approach to these issues might refer to the intentions of those who were empowered (or who empowered themselves) to lay down the rules and to inscribe legal meanings onto social space. For example, it is common to refer to 'the intentions of the framers' to protect privacy and the security of the home in explicating fourth amendment restrictions on police searches as well as justifing exceptions to these same rules. Likewise, legislators are said to intend the regulation of protests in public space so as to preserve order. Obviously, any effort to stabilize the legal meanings of social space by reference to the intentions of actors who may be far removed in time and place is not without difficulties.

A subtler legal geographical analysis might inquire into the 'deeper' meanings of socio legal space. The aim here is to explicate legal spaces and configurations less as the results of intentional practices of individuals than as ideologically informed political or cultural artifacts. In this regard, it might be noted that many of the rules that are encoded in legal texts (statutes, constitutions, judicial opinions, and so on) are themselves the products of social and political contests. They are often crafted by agents who have access to legal institutions to be used against adversaries who do not have access. Certainly, white people have used legal instruments to produce the spaces of racial subordination (segregated spaces, native reserves, and colonies), and capital has produced legal spaces against the interests of labor (in the workplace, as well as in public space). These legalized and spatialized power relations can be understood simply in terms of material interests and domination. But they can also be examined in terms of ideologies and broader, more diffuse cultural meaning systems.

Insofar as elements of the legal are deeply implicated in the arrangement and assembledges of heterogenous spaces across extensive segments of social landscapes, legal geographical investigations can reveal important processes and practices through which such spatial assembledges are configured and reconfigured. This applies not only 'horizontally' but also 'vertically' through the legal configuration of spaces and landscapes associated with differing scales of reference (e.g., the personal, local, national, international, and global). Just as the legal constitution of discrete spaces conditions the operation and circulation of power with respect to actors and events in or out of these spaces, so legal phenomena are implicated in the social production of these more extensive, spatialized fields of power. These processes commonly operate through the adversarial construction of spatialized fields of legal meaning.

Social power, of course, takes many forms and has many social sources. Analytically economic, political, cultural, and other sources and expressions of social power are commonly distinguished from one another. The legal is, in important ways, distinctive from these. It is characterized by distinctive discourses, institutions, and practices. But it is not radically other than or entirely autonomous from these other structures of social power. Rather, distinctively legal forms of social power give expression to these other forms. One way this happens is through the redescription of power in terms of the discourses and categories of 'rights'. For example, forms of economic power are commonly translated into claims about property rights (i.e., the rights of investors, the right to strike, and consumer rights). But this is not simply a straightforward translation. As in many translations, some things are lost and some things are added. Frequently, what is added in advancing claims in terms of legal rights is a heightened presumption of legitimacy. Actions which have been deemed 'legal' through the ascription of rights are presumptively acceptable. Another characteristic that comes along with rights is enforceability: that is, the rights holder has presumptive access to the means of state coercion to enforce or protect these rights. Legal geography examines translation projects in connection with the geographical expression of social power.

The distinctively legal can also be understood as 'mediating' competing forms of power; for example, forms of economic power (ownership) and political power (democracy). Understanding this relation in terms of either translating or mediating opens up possible lines of inquiry into the workings of the legal as conditioning the social constitution of geographies of power. For example, if one is interested in the operations of geographies of capitalism (vis a vis patterns of investment and disinvestment, labor relations, class relations more generally, or in relation to the environment), it may be worthwhile to investigate the details of their construction and reconstruction through law at whatever scale of reference is relevant. Clearly one important point of ingress is through analyses of property: what 'rights' come with ownership? how are these rights created, regulated, and enforced? how might property rights in a given case be understood as limited by other kinds of rights and duties, such as the rights to organize and strike, the rights associated with privacy, or the requirements of environmental protection? Geographical reconfigurations take place through rights based practices, and contests among rights claimants shape the details of many political–economic spatial transformations. For example, much of the urban redevelopment in the United States takes place in the contexts of contests about the meaning of fifth amendment rules, rights, and procedures that govern the 'taking' of property by the state, and whether such 'taken' land can then give it to private corporate interests.

To take another example, consider the legal construction of geographies of race and racism. 'Race' is not a natural way of sorting different kinds of human beings but a dynamic, historical system of power that governs relations of domination, privilege, and subordination through processes of racialization. As examples as varied as apartheid, Jim Crow, White Australia, and indigenous reserves in different parts of the world demonstrate, racialization is commonly effected through processes of spatialization: separation, confinement, exclusion, expulsion, and forced removal. 'Races' (and the relationship among them) are also constituted by law, through the categories of legal discourse and the vectors of legal violence. One may investigate the politics of the legal constitution and reconstitution of these geographies of race and racism through analysis of the interpretive contests that aim to revise or reinforce them.

In a South African context such legal instruments as the Group Areas Act of 1950, the various versions of 'influx control' legislation which regulated the rural to urban migration of Africans, the complex legal texts that constituted the 'Bantustans' (the fictive 'homelands' of Africans) were all subject to interpretive maneuvers. The results of these maneuvers had the material effect of opening up or closing off spaces and places to African people. In a North American context similar analyses could look at the Alien Property Acts that prohibited 'Orientals' from owning agricultural land, or the Major Crimes Act of 1885 which extended federal police authority into 'Indian Country' (itself a legal construction), or *Plessey v. Ferguson* and other cases which validated the so called 'separate but equal' policies that underwrote the spatiality of anti black racism. These texts do not simply reference rules but invoke racist ideologies that are projected onto lived in landscapes and are legitimized through specifically legal ideologies. Such analyses can be enriched by detailed empirical or archival research into how actual persons navigated, accommodated, or resisted such spatio legal projects.

Another important feature of legal geographies of power (as such, or legal geographies of race, class, gender, etc.) is the 'sedimentation effect'. In common law jurisdictions the practical work of lawyers and judges frequently involves the interpretation of historical documents (constitutions, treaties, statutes, prior rulings, doctrines, contracts, deeds, and so on) that created or revised the legal meanings inscribed onto segments of social space. Contemporary disputes often turn on con tentious readings of these as they are imagined to be embedded in actual landscapes. For example, twenty first century legal arguments about the location, boundaries, and meanings of 'Indian Country' in the United States can turn on contested interpretations of the statutes that established a particular reservation in the nineteenth century, the executive orders that modified them, the treaties and contracts that were executed subsequent to these, prior cases that interpreted these and doctrines such as 'the domestic dependent nation', and the 'plenary power' doctrines that are key constituents of Federal Indian Law. Such traces of legality are imagined as 'frozen' in the past and, in a sense, embedded in the landscapes to which they confer legal meaning.

Contemporary conventional interpretations are therefore imagined as constrained or compelled by claims about what various actors (colonial administrators, local settlers, 'authorized' indigenous people) intended or understood in the remote past. These posited 'under standings' of legal space are privileged over con temporary desires, needs, and values. Insofar as these historical understandings (whether accurate or not) arose in explicitly colonialist or racist socio spatial projects, such projects can be given continued effect even though the ideologies that inform them would not be acceptable now. These are all outcomes of distinctively legal ideologies. However, it is also the case that because of the openness or indeterminacy of legal discourse, such sedimented meanings may be repudiated (in whole or part, tacitly or explicitly) and the relevant spatial fields of power thereby reconfigured. Legal geographical analysis can help to exhume such meanings or reveal contingent aspects of their excavation by legal actors.

Doing Legal Geography

Legal geography has three principal lines of inquiry: (1) How do distinctively legal practices shape or contribute to transformations of geographical phenomena? (2) How does geographical or socio spatial phenomena inform or condition the workings of the distinctively legal? (3) How can analysis of these reciprocal processes illuminate other questions of social research more broadly?

With respect to the first line of inquiry, much depends on the particular understanding of the legal that we bring to the task and how deeply into the legal we go in search of explanation. One might limit analyses of 'the legal' to the rules and regulations that are promulgated by formal legal institutions: for example, the US Fair Housing Act of 1968; sections of the Canadian Charter of Rights and Freedoms; and the Australian *Mabo v. Queensland* decision of 1988. But studies may be given greater sociological thickness through a more explicit examination of dis tinctively legal institutions and roles. Analysis of an issue or event can benefit from asking questions like: how do courts really work? What do judges and lawyers do when they argue or rule? What is policing or border patrolling really like? These kinds of questions can be addressed directly through empirical studies, including participant observation, or through engagement with relevant lit eratures. These thicker sociological treatments can be further enriched by more explicit investigation of dis tinctively legal discourses and ideologies that inform what these actors do. This kind of analysis need not be confined to formal, state legal actors. It can be extended to 'popular legal consciousness', that is, to how ordinary people think about, or through, the law. Important work along these lines has looked at property, at public space, and at the legal geographic construction of Israel/ Palestine. These distinctively legal images and dis positions can be extended or connected to more perva sive social ideologies, such as liberalism and its associated notions of self and community, or to prominent ideolo gies of statism and authority. These ideologies, in turn, may be connected to analyses of capitalism, patriarchy, racism, and other relevant structures of power.

Taking 'the legal' in varying degrees of 'thickness' helps to explicate the innumerable material practices and interventions that contribute to the formation and transformation of geographical phenomena. Such prac tices and ideologies participate in geographical trans formations in countless ways. Spaces are created, maintained, and revised through the often contested 'inscriptions' of abstract legal meanings (and associated images and ideologies) onto lived in places and land scapes. Borders are drawn and redrawn, people and places are put off limits, spatial identities are imposed or contested. Scholars in legal geography have pushed these lines of inquiry to explicate contexts and contests as di verse as those concerning political asylum, policing, gardening, homelessness, public housing, dissent, sexu ality, racial subordination, and land expropriation.

Just as there are methodological issues regarding how deeply to go into the legal, there are questions about how deeply to go into the geographical. Some of the geog raphy of law and geography sticks fairly close to the surface of mappable or locatable phenomena. But the interpretive resources of human geography have ex panded enormously over the last generation and what is taken to be geography has grown accordingly. Studies of the legal in the spatial may engage questions of scale and scaling, dimensions of geographical imaginations, or the workings and effects of spatial representations and discourses. This brings us to the second principal line of inquiry: investigating the spatiality in and of the legal. There are some obvious connections between legal phenomena and conventional conceptions of space. The very idea of law is often structured or framed by spatial or geographical images. Most obvious is the image of territorial sovereignty and its inscription of an inside/outside orientation to law. Sovereign space is imagined to contain, constrain, and limit the scope and reach of the law and, usually, to constitute a homogenous space over which the law of the land is operative. Basic order spatial images also condition such 'internal' dimensions of the legal as federalism, jurisdiction, and local government law. These are frequently described in terms of higher and lower levels of government and implicate conventional notions of scale.

Likewise, the international system of states constitutes a supra national space of reference for international law, and affects a global distribution of legal regimes. International law traditionally pertains to treaties and conventions with which member states of 'the international community' govern relations among themselves. But human rights activists and others have begun to make claims that international law implicates (and obligates) nonstate and so called 'private' international actors such as multinational corporations and nongovernmental organizations as well. Insofar as these claims are accepted and acted upon, international legality (structuring and structured by the spatiality of sovereignty) may be transforming into a global legality. That is, the spatiality of law on a planetary scale may be shifting toward a configuration in which the territorial state is merely one component among others. Both state and nonstate actors are resisting this posited de territorialization of law.

These common sense spatial structures of the legal are frequently simply assumed by scholars and others. However, they are increasingly being challenged and problematized from a number of perspectives. With respect to the fundamental image of territorial sovereignty, it is now commonly argued that while this conceptual device for imagining the space and geography of law may once have been accurate or serviceable, this is no longer the case. Whatever sovereignty once meant it means something quite different under conditions of globalization. That is, the foundational spaces of law (territorial states) are no longer the simple containers they were once presumed to be. To the extent that borders are increasingly understood as more 'permeable', elements of legality from 'outside' (such as regulatory authority or policing) may percolate into the insides.

Whether 'within' or 'without' (or, indeed, the very framing of inside and outside), the spatiality of law in this sense has always been a matter of interpretation, and interpretations are always multiple and variable. For example, some event that may appear to be a matter of 'national' (domestic) legality from one perspective can often be reinterpreted or reframed as being a matter of international legality (e.g., an event seen to implicate human rights). Likewise, what appears 'domestic' from one view can seem 'foreign' from another (e.g., an event concerning transnational flows). The same holds for taken for granted notions of federalism or the scaling of domestic law. Much recent work in human geography has critiqued inherited notions of scale and emphasized the social or political construction of scale. These ideas problematize conventional scalar conceptions of the spatiality of law. The political construction of scale often takes place through distinctively legal practices and institutions, most clearly in disputes about federalism or local government law itself. Legal geography tends to see space not as a self evident, objective matrix which contains law or which unproblematically distinguishes among legal regimes (either as among sovereign states or as between the domestic and the international). Rather, spatialities of law are themselves understood as the effects of spatio legal interpretive frameworks and practices.

As human geographers have increasingly attended to the significance of spatial and geographical discourses, representations and narratives, so legal geographers have relied on some of these more sophisticated interpretive resources in their analyses of the legal constitution of geographies of power. Social power and variable expressions of spatio legal consciousness and ideologies (e.g., property consciousness, forms of self and social consciousness associated with citizenship, privacy, publicness, criminality, and punishment) are always bound up with each other. Spatio legal discourse analysis is useful in understanding the unfolding of many kinds of disputes and events that are shaped by the confrontation between contending ways of imaging power in these ways. Of special interest to legal geographers are the ways in which these ideologically loaded socio spatial images in legal thought itself are implicated in contests about material spaces and lived in landscapes.

The third line of inquiry seeks to extend the relevance or utility of legal geography to broader social and social theoretic concerns. Beyond exploring these lines of reciprocity or mutual constitutivity of 'the legal' and 'the geographical' efforts have been made to examine how understanding the terms of mutual constitutivity may illuminate the workings of power generally.

There is very little that human geographers are interested in that does not implicate 'the legal' in one way or another. This does not mean that one must follow these lines of inquiry, just as one needn't explicitly foreground 'the economic' or 'the cultural' in every research project. But doing so may yield significant insights. Moreover, whether critical or conventional approaches are most appropriate depends on the task at hand. The worlds that human geographers seek to describe, explain, and assess are (at least as much) 'made' as 'found'. The legal is deeply enmeshed in the processes, practices, and pragmatics of world making. Much about the world is contingent and much of this contingency is illuminated through analysis of distinctively legal practices and events. The legal itself is contingent: on different, con tending forms of social power, on conceptual or ideo logical constraints, on inherited spatialities, some of which are materialized in the built environment. Ex ploring the dynamics of law and geography can reveal degrees of openness that may exist as well as elements of closure and foreclosure that condition what it is like to be in the world.

See also: Ideology; Radical Geography; Sovereignty; Policing.

Further Reading

Blomley, N. (1989). Text and context: Re thinking the law space nexus. *Progress in Human Geography* 13, 512 534.
Blomley, N. (1994). *Space, Law and the Geographies of Power*. New York: Guilford.
Blomley, N. (2004). *Unsettling the City: Urban Land and the Politics of Property*. New York: Routledge.
Blomley, N., Delaney, D. and Ford, R. (eds.) (2001). *The Law and Geographies Reader*. Oxford: Blackwell.
Chouinard, V. (1994). Geography, law and legal struggles: Which ways ahead? *Progress in Human Geography* 18, 415 440.
Clark, G. (1985). *Judges and the City: Interpreting Local Autonomy*. Chicago, IL: University of Chicago Press.
Clark, G. (1989). The geography of law. In Peet, R. & Thrift, N. (eds.) *New models in geography*, pp 310 337. London: Unwin Hyman.
Cooper, D. (1998). *Governing Out of Order: Space, Law and the Politics of Belonging*. London: Rivers Oram Press.
Delaney, D. (1998). *Race, Place and the Law: 1836 1948*. Austin, TX: University of Texas Press.
Ford, R. (1994). The boundaries of race: Political geography in legal analysis. *Harvard Law Review* 107, 1841 1921.
Ford, R. (1997). Law's territory: (A history of jurisdiction). *Michigan Law Review* 97, 843 930.
Herbert, S. (1997). *Policing Space: Territoriality and the Los Angeles Police Department*. Minneapolis, MN: University of Minnesota Press.
Holder, J. and Harrison, C. (2002). Connecting law and geography. In Holder, J. & Harrison, C. (eds.) *Current Legal Issues, Vol. 5: Law and Geography* pp 3 16. Oxford: Oxford University Press.
Kedar, A. (2001). The legal transformation of ethnic geography: Israeli law and the Palestinian landholder 1948 1967. *New York University Journal of International Law and Policy* 34, 923 1000.
Mitchell, D. (2003). *The Right to the City: Social Justice and the Fight for Public Space*. New York: Guilford Press.
Mitchell, D. (2005). Permitting protest: Parsing the fine geography of dissent in America. *International Journal of Urban and Regional Research* 29, 796 813.
Zick, T. (2006). Speech and spatial tactics. *Texas Law Review* 84, 581 651.

Learning Regions

B. T. Asheim, Lund University, Lund, Sweden

Glossary

Agglomeration Territorially concentrated location of economic activity exploiting external economies of scale (localization or urbanization economies).

Industrial District Local production systems characterized by agglomerated and closely networked small and medium enterprises (SMEs) exploiting localization economies.

Innovation Commercialization of inventions. Can be both new products and processes, as well as organizational innovation. Can be radical (new to the market) or incremental (new to the firm).

Learning Economies New phase in the development of capitalism where continuous learning is of strategic importance to remain competitive due to the rapid changes in the globalizing knowledge economy.

Localization Economies External economies of scale derived from co-location of industries belonging to the same or closely related sectors. Urbanization economies derive from external economies of scale from different sectors, typically found in urban areas. Regional clusters and industrial districts are examples of agglomerations exploiting localization economies.

Post-Fordism Characterized by customized, batch production (flexible specialization) in contrast to the standardized, mass production of Fordism.

R&D Research and development activities in order to develop innovations. Measured as R&D intensity by percent of turnover used for R&D. Carried out at universities, R&D institutes, and R&D department in firms. Innovation can however be based on other types of knowledge than R&D, for example, design. The knowledge base behind innovations is thus larger than R&D activities.

Regional Innovation System Systemic and long-term relations between organizations and institutions supporting and promoting knowledge exploration, learning, and competence building and organizations (e.g., firms) exploiting and using knowledge, learning, and competence building.

Technoeconomic Paradigm A technoeconomic paradigm describes a state when a new basic technology (e.g., information and communication technology (ICT)) has diffused throughout the whole of society, and has, thus, become the key technology of a society.

Third Italy Concept developed by the Italian sociologist, Bagnasco, in 1977 to describe the Central and Northeastern parts of Italy, which in contrast to the First Italy's (Lombardy, Piedmonte, and Liguria) Fordism had a regional economy of agglomerated SMEs in industrial districts characterized by post-Fordist production.

Introduction

For more than 25 years a strong case has been made that territorial agglomerations are growing in importance as a competitive location of economic activities in post Fordist learning economies. The main argument for this is that territorial agglomeration (or clustering) provides the best context for the promotion of innovative firms based on sticky knowledge and localized learning. This view on the importance of the regional level can find support from modern innovation theory, originating from evolutionary and institutional economics, which points to the increased importance of regional innovation systems, industrial districts, and technological districts. In the perspective of innovation as culturally and institutionally contextualized, strategic parts of learning processes emerge as highly lo calized, as opposed to placeless. Thus, local contexts can represent important parts of the knowledge infrastructure of firms and regions, underscoring the role of historical trajectories. However, research on clusters and regional innovation systems has also revealed that the regional level is neither always nor even normally sufficient for firms to stay innovative and competitive. This is a con sequence of firms' learning processes becoming increas ingly inserted into various forms of networks and innovation systems (at regional, national, and international levels), which increases the importance of firms' absorptive capacity for accessing nonlocal knowledge.

After years of influential research on the importance of territorial agglomerations for economic growth there is, thus, a need for an 'unpacking' strategy to disclose and reveal the contingencies, particularities, and specificities of the various contexts and environments where know ledge creation and innovation take place in order to obtain a better understanding of factors enabling or im peding these processes. So far, all the way from Marshall's writing on industrial districts, it has been assumed that business interactions (from exploiting localization economies) and knowledge flows were co occurring (and co located) phenomena. Furthermore, it has been maintained that local interactions and collective learning processes, or what is sometimes called 'local buzz', largely take care of themselves by just 'being there', while building 'global pipelines' to knowledge providers located outside the local milieu requires institutional and infrastructure support, as one cannot expect that it occur spontaneously.

It is this idea of an almost automatic shaping of endogenous learning and innovation capacity by just being co located in an agglomerated environment, which also lies behind Porter's understanding of how competitive advantage is 'created'. Recently, observers have questioned if cluster learning is a pervasive and 'collective' process only conditioned by territorial agglomeration as such. New research has shown empirically that there exists an uneven distribution of knowledge and selective interfirm learning due to the heterogeneity of firms' competence bases, which cannot be fully compensated by regional universities or other parts of a region's 'collective absorptive capacity'.

Thus, it is an important question if more planned and systemic approaches are needed in a globalizing knowledge economy in order for regional advantages to be deliberately 'constructed'. This argument is grounded in the fact that the contemporary globalizing knowledge economy – characterized by outsourcing/offshoring of both production and R&D, open innovation, dominating transnational corporations (TNCs), and intensified competition from developing economies of which China and India are the 'star' examples – is becoming more knowledge intensive, and explicit, codified knowledge is consequently increasing in importance. Simply leaving the question of how constructed advantage is attained just to the market or the 'territory' in the Marshallian way, when tacit knowledge was most important, is probably not enough.

This points to a new and more dynamic role for the public sector (including universities) generally and government and governance specifically in interaction with the private sector to promoting innovativeness and competitiveness. Governments and agencies at all spatial levels have increasingly become involved in seeking to stimulate innovation, and, consequently, innovation policy is put at the center of policies for promoting regional and national economic development. At the regional level supporting the formation of regional innovation systems, clusters, and learning regions have been looked upon as policy frameworks or models for implementation of long term, development strategies initiating learning based processes of innovation, change, and improvement.

What Is a 'Learning Region'?

In a newly published book *The Learning Region: Foundations, State of the Art, Future*, four articles are reprinted as representing the foundations of the concept of learning regions: Storper's 'Regional 'worlds' of production' from 1993; Florida's 'Toward the learning region' from 1995; Asheim's 'Industrial districts as learning regions' from 1996; and, finally, Morgan's 'The learning region: institutions, innovation and regional renewal' from 1997. This demonstrates that the concept of 'learning regions' is a product of the 1990s and reflects empirically the rapid economic development in places such as the 'Third Italy', which drew attention toward the importance of cooperation between small and medium enterprises (SMEs) in industrial districts and between firms and local authorities at the regional level in achieving international competitiveness. Theoretically, it reflects the definition of post Fordist societies as learning economies, where innovation is seen as basically a socially and territorially embedded, interactive learning process, which cannot be understood independent of its institutional and cultural contexts.

Even if these contributions in many ways have a common view on how to understand 'learning regions', underlining the important role of innovation, understood as contextualized social processes of interactive learning, they also disclose interesting differences. One such difference can be identified between the American and European approaches. While learning regions in a North American context are associated with the importance of the quality of the knowledge infrastructure of leading universities and research institutions in a knowledge based economy, producing, attracting, and retaining highly skilled workers (e.g., Silicon Valley), in a European context of learning economies the focus is much more on the role social capital and trust plays in promoting formal and informal interfirm networks and the process of interactive learning (e.g., industrial districts in the Third Italy). The broadness in the different interpretations of learning regions clearly also demonstrates that the concept can be fuzzy and the use of the concept both theoretically and practically rather flexible. In order to provide an analytically more precise understanding of the concept an attempt of identifying three important building blocks of the concept is presented in the following section.

The Building Blocks of the Concept of Learning Regions

The concept of learning regions has thus been used in at least three different contexts. The concept was first introduced by economic geographers in the mid 1990s, when they used it to emphasize the role played by cooperation and collective learning in regional clusters and networks in order to promote the innovativeness and competitiveness of firms and regions in the globalizing learning economy.

The second approach expressing (more indirectly) the idea of learning regions originates from the writings of evolutionary and institutional economics on the learning economy, where knowledge is considered the most fundamental resource and learning the most important process, thus making the learning capacity of an economy of strategic importance to its innovativeness and competitiveness. The concept of 'the learning economy' refers to the contemporary post Fordist economy dominated by the information and communication technology (ICT) related (information, computer, and telecommunication) technoeconomic paradigm in combination with flexible production methods and reflexive work organizations (i.e., learning work organizations and functional flexible workers). In addition the learning economy is firmly based on innovation as a crucial means of competition, where innovation is understood as interactive learning in contrast to the previous hegemonic linear model of innovation.

The third approach, which conceptualizes learning regions as regionally based development coalitions, has been developed by representatives of action oriented organizational research taking their knowledge of how to form intra and interfirm learning organizations based on broad participation out of the firm context and applying it at the regional level. By the concept 'development coalition' is meant a bottom up, horizontally based cooperation between different actors in a local or regional setting, based on a socially broad mobilization and participation of human agency.

When learning regions are defined in this way they resemble a regional innovation system broadly defined, incorporating the elements of a bottom up, interactive innovation model. The 'innovation system' concept can be understood in both a narrow as well as a broad way. A narrow definition of the innovation system primarily incorporates the R&D functions of universities, public and private research institutes and corporations, reflecting a top down model of innovation as exemplified by the triple helix approach. A broader conception of the innovation systems includes all parts and aspects of the economic structure and the organizational and institutional setup affecting learning and innovation in a region.

This type of regional innovation system may be denoted as 'territorially embedded regional innovation systems', where firms base their innovation activity mainly on localized learning processes stimulated by geographical, social, and cultural proximity, without much direct interaction with knowledge creating organizations (i.e., R&D institutes and universities).

The best examples of 'territorially embedded regional innovation systems' are networks of SMEs in industrial districts. Thus in Italy's Emilia Romagna, for example, the innovation system can be described as territorially embedded within that particular region. These territorially embedded systems provide bottom up, network based support through, for example, technology centers, innovation networks, or centers for real service providing market research and intelligence services.

The definition of learning regions as regional development coalitions is inspired by Scandinavian experiences, which have shown that flat and egalitarian organizations have the best prerequisites of being flexible and learning organizations, and that industrial relations characterized by strong involvement of functional flexible, central workers is important in order to have a working learning organization. Such organizations will also result in well functioning industrial relations, where all the employees (i.e., the (skilled) workers, as well as the managers) will have a certain degree of loyalty toward the firm. The way firms are organized also influences their capability to learn, and the change from hierarchies with vertical flows of information toward more flat organizations with horizontal flows of information is an important aspect of the learning economy.

The organized form of bringing the society inside the firm through learning organizations based on broad participation, and supported by labor market legislation, as well as a strong tradition of cooperation between the labor market organizations, is in many aspects the opposite way of achieving a fusion of the economy with the rest of society than the industrial district model, in which the firm is contextualized through its embeddedness in spatial structures of social relations. These contrasting models of contextualizing the firm also reflect the alternative interpretations of social capital, that is, as rooted in the 'civicness' of communities and informal social networks (i.e., bonding) in the industrial districts of the Third Italy or as embedded in formal labor market's organizations and institutions on the system level of society (i.e., bridging) in the Nordic countries.

Critique of Learning Regions

Taking into consideration the wide diffusion the learning region concept has attained, it is both quite natural and obvious that it has received quite a lot of criticism. The main criticism can be related to four topics: (1) regions cannot learn, only firms and organisations; (2) structural limits to learning in a capitalist global economy; (3) learning is only reproductive; and (4) social capital as an independent variable. The criticism that only firms and organizations can learn and not territorial units is shared by people coming from a variety of disciplines ranging from economics to regional studies. They argue that learning is an activity that requires agency and that such agency is represented by people working in various organizations. The counter argument is represented by an innovation system perspective partly arguing that interactive learning is socially and institutionally, and, thus, territorially contextualized, and partly that the formation of, for example, regional innovation systems according to a triple helix model place firms in a broader alliance of actors, which are territorially embedded. The danger of this is of course that the instrumental rationality of firms is taken over by the regions resulting in a one dimensional focus on development 'in' a region, and ignoring the development 'of' a region. This has clearly been the case in much of the research on endogenous, territorial development, but especially the learning region approach has in many cases more explicitly also taken the social cohesion and welfare dimensions into consideration.

The argument about structural limits to learning in a capitalist economy often has a reference to the neoliberal tendencies in the development of global capitalism, the last 25 years. This position, however, ignores the important research carried out by institutional economists and political scientists on the 'varieties of capitalism'. Their work has shown that it is not a question of only one type of capitalism, but of several (i.e., 'varieties'), and even if they share the basic, structural characteristics of the capitalist mode of production, their historical, concrete appearance and development can be quite different. They make the simple dichotomy between 'coordinated market economies', which can be found in the Nordic welfare states, as well as in Continental Europe, and 'liberal market economies', which is the Anglo American model. The variety of capitalism approach argues that different national institutional framework support different forms of economic activities with different social outcomes due to different types of institutional complementarities. This perspective supports the argument that coordinated market economies can demonstrate synergy between equity and efficiency, while liberal market economies disclose a contradiction between equity and efficiency. The dominance of learning work organizations in the Nordic countries, which is referred to in the next section, is also an example of the possibilities of learning taking place in the context of global capitalism.

The argument that learning can only be reproductive, that is, it only can imitate already existing phenomenon, ignores that learning also can be developmental and creative. The above mentioned learning work organizations show that autonomy of work of highly skilled workers provides a dynamic learning environment that also can be supportive of knowledge creation and innovation (see the next section). Finally, there is the criticism that the strong focus on social capital as an independent variable discloses important structural characteristics of the global capitalism. This position furthermore criticizes the view of social capital as a property of groups and regions, and prefers to look at it as a resource of individuals in struggle with others. Of course, social capital can be both, but if looking to the Nordic countries once again, European Union (EU)'s studies of social capital demonstrates a strong correlation between a high presence of social capital and trust of government, political authorities, and important societal institutions, which produces a strong social cohesion. This among other things shows up in more optimistic attitudes in the younger generation, for example, toward globalization, than in other European countries with a lower presence of social capital. In addition, the presence of social capital both as bonding and bridging represents a strong tradition of cooperation, which also is important for innovation as interactive learning, and adds to the high level of human capital in the workforce in a synergistic way.

Learning and Innovation in Regional Economies

According to the third perspective on learning regions as regionally based development coalitions, learning regions should be looked upon as a policy framework or strategy for formulations of long term partnership based development strategies initiating learning based processes of innovation, change, and improvement. In the promotion of such learning regions the interlinking of learning organizations ranging from work organizations inside firms via interfirm networks to different actors of the community, understood as 'regional development coalitions', will be of strategic importance. The attractiveness of the concept of learning regions to planners and politicians is to be found in the fact that it at one and the same time promises economic growth and job generation, as well as social cohesion. As such, learning regions may be analyzed as an answer and challenge at the regional level especially for regions with weak territorial competence bases to contemporary changes in the global economy, underlining the strategic role played by 'social capital's' emphasis on the social and cultural aspects of regions facilitating collective action for mutual benefit. Thus, it is not accidental that this approach to learning regions was used by the Regional Innovation Strategies pilot actions of the EU Commission as part of new policy developments in Europe since the 1990s to promote less developed regional economies within EU through innovation.

The concept of learning regions emphasizes the role of regional based learning organizations. The strategic role played by cooperation in a learning economy is underlined by the understanding of interactive learning as a fundamental aspect of the process of innovation. This broader understanding of innovation as a social, nonlinear, and interactive learning process (both reproductive and developmental) puts new emphasis on the role played by sociocultural and institutional structures in regional development, highlighting the significance of building social capital in order to foster cooperation, as well as promoting the principle of broad participation in intra and interfirm networks. They are no longer vestigial remnants of precapitalist civil societies. They are necessary prerequisites for firms and regions to be in novative and competitive in a post Fordist learning economy. Of strategic importance in this context is the capacity of people, organizations, networks, and regions to learn. The concept of a learning region can, thus, be used to describe a region characterized by innovative activity based on localized, interactive learning, and co operation promoted by organizational innovations in order to exploit learning based competitiveness. As al ready mentioned this implies a broad understanding of a regional innovation system as representing the territorial and institutional embeddednesss of learning organiza tions and interactive learning with a strong focus on competence building in addition to learning and innovation.

Thus, knowledge and innovation should not simply be equated with R&D. Innovative activities have much broader knowledge bases than just science based R&D, and there are many examples of nations and regions demonstrating a rapid economic growth and a high level of living standard with an industry competing on the bases of non R&D based, incremental innovations (e.g., Denmark and regions in The Third Italy). Thus, a re gion's knowledge base is larger than its science base, implying that arguing for an increasingly more know ledge intensive globalizing economy does not necessarily mean that innovation and competitiveness only become more dependent on R&D.

The distinction between non R&D and R&D based (regional) economies, furthermore, implies the use of different modes of innovation. One the one hand, we can talk about a broad definition of the mode of innovation as D(oing), U(sing), and I(nteracting) relying on informal processes of learning and experience based know how in most cases resembling a synthetic (or engineering based) knowledge base and/or a symbolic knowledge base in coordinated market economies. Such a mode of innov ation is typically found in non R&D based learning economies (e.g., Denmark), mainly producing user driven, incremental innovations. On the other hand, in a R&D, knowledge based economy one finds a more nar row definition of the mode of innovation as S(cience), T(echnology), and I(nnovation) based on the use of co dified scientific knowledge, which could broadly be as sociated with an analytical (or science based) knowledge base. This mode of innovation more often generates radical (product) innovations than a broadly defined mode of innovation and is typically found in liberal market economies. These two modes of innovation will also be differently manifested with regard to regional specialization and clustering.

However, on the firm levels these two modes of innov ation coexist, but are applied to different extent depending on the dominating knowledge base(s). Furthermore, re search shows that combining the two modes of innovation seems to be most efficient, that is, firms that have used the STI mode intensively may benefit from paying more at tention to the DUI mode and vice versa. Thus, the dis tinction between these two modes of innovation helps on the one hand to avoid a too one sided focus on promoting science based innovation of high technology firms at the expense of the role of learning and experience based, user driven innovation. On the other hand, it also indicates limits of such innovation strategies and, thus, emphasizes the need for firms in traditional manufacturing sectors and services more generally to link up with sources of codified know ledge in distributed knowledge networks.

Moreover, new research on the relationship between different types of work organization in EU and the im pact on job stress, worker satisfaction, labor market flexibility, learning, innovation, and patenting identifies four main forms of work organization: 'learning', 'lean', 'Taylorist', and 'simple structure'. This study shows a clear north–south divide with regard to the dominating forms of work organization with Northern Europe dominated by learning forms of work organization, while Southern Europe has work organizations characterized by either Taylorist or simple forms. The learning form of work organization, found in Scandinavia and the Netherlands, does not only result in less job stress and greater worker satisfaction, it also implies more labor market flexibility, superior conditions for learning and innovation, and even a larger propensity for patenting. The autonomy of work and learning dynamics found in this type of work organization not only supports repro ductive or adaptive learning, but also underpins devel opmental or creative learning. The positive impact of this form of work organization on innovation is confirmed by another study reporting that 'low road' practices using short term and temporary contracts, having a lack of employer commitment to job security, low levels of training, and so on are negatively correlated with in novation. In contrast, it is found that 'high road' work practices characterized by 'high commitment' organiza tions or 'transformed' workplaces are positively correl ated with innovation.

Conclusions: Learning Regions in a Globalizing Knowledge Economy

The continuous importance of the regional level for creating and retaining innovative firms in a globalizing knowledge economy is confirmed by results from a European comparative cluster survey, which shows that regional resources and collaboration are of major im portance in stimulating economic activity. In this study it was found that in many clusters, firms increasingly find relevant research activities and other supporting services inside the regional cluster boundaries.

Furthermore, we have argued that, under the right conditions, interactive learning has the capacity to promote and sustain regional innovativeness and com petitiveness. At the core of these conditions are cooper ation and nonmarket and noneconomic factors, including social capital, trust, and a tissue of supportive institutions. Therefore, it can be maintained that there is a potential for establishing learning regions within regions with a weak knowledge infrastructure through the use of ap propriate policy instruments, promoting learning, in novation, and endogenous development 'in' as well as 'of' regions.

However, as a result of increased global competition from newly industrialized countries such as Taiwan and South Korea, and rapidly catching up countries such as China and India, as well as increased codification of knowledge due to the growing knowledge intensity of the globalizing economy, the need for firms in Europe to enhance their competitiveness by combining a focus on product differentiation/innovation toward high quality/ value added products with cost efficiency has become evident. Subsequently firms apply an open innovation model and thus rely on sourcing – sometimes globally – for the best talents and knowledge while at the same time outsourcing or offshoring standardized labor intensive activities, but also increasingly R&D and innovation, to low cost countries. These development tendencies rep resent a serious threat and challenge to the endogenous learning region approach, which is clearly illustrated by the development in, for example, the Third Italy. The policy implications of this is simply that while the for mation of regional innovation systems, clusters, and learning regions still represent very important policy measures for the promotion of innovative and com petitive firms and regions, such measures do not consti tute a policy panacea, but have to be applied within a context of a multilevel approach to innovation policy.

See also: Industrial Districts; Locality Debates; Regional Innovation Systems.

Further Reading

Amin, A. and Thrift, N. (1995). Territoriality in the global political economy. *Nordisk Samhällsgeografisk Tidskrift* 20, 3 16.

Asheim, B. T. (1996). Industrial districts as 'learning regions': A condition for prosperity? *European Planning Studies* 4(4), 379 400.

Asheim, B. T. (2000). Industrial districts: The contributions of Marshall and beyond. In Clark, G., Feldman, M. & Gertler, M. (eds.) *The Oxford Handbook of Economic Geography*, pp 413 431. Oxford: Oxford University Press.

Asheim, B. T. (2001). Learning regions as development coalitions: Partnership as governance in European workfare states? *Concepts and Transformation. International Journal of Action Research and Organizational Renewal* 6(1), 73 101.

Asheim, B. T. and Coenen, L. (2005). Knowledge bases and regional innovation systems: Comparing nordic clusters. *Research Policy* 34(8), 1173 1190.

Asheim, B. T. and Gertler, M. S. (2005). The geography of innovation: Regional innovation systems. In Fagerberg, J., Mowery, D. & Nelson, R. (eds.) *The Oxford Handbook of Innovation*, pp 291 317. Oxford: Oxford University Press.

Asheim, B. T. and Isaksen, A. (2002). Regional innovation systems: The integration of local 'sticky' and global 'ubiquitous' knowledge. *Journal of Technology Transfer* 27, 77 86.

Asheim, B. T., Coenen, L., Moodysson, J. and Vang, J. (2007). Constructing knowledge based regional advantage: Implications for regional innovation policy. *International Journal of Entrepreneurship and Innovation Management* 7(2/3/4/5), 140 155.

Asheim, B. T., Isaksen, A., Nauwelaers, C. and Todtling, F. (eds.) (2003). *Regional Innovation Policy for Small Medium Enterprises*. Cheltenham: Edward Elgar.

Bathelt, H., Malmberg, A. and Maskell, P. (2004). Clusters and knowledge: Local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography* 28(1), 31 56.

Bellini, N. and Landabaso, M. (2007). Learning about innovation in Europe's regional policy. In Rutten, R. & Boekema, F. (eds.) *The Learning Region. Foundations, State of the Art, Future*, pp 231 251. Cheltenham: Edward Elgar.

Cooke, P. (2001). Regional innovation systems, clusters, and the knowledge economy. *Industrial and Corporate Change* 10(4), 945 974.

Cooke, P. (2007). Regional innovation systems, asymmetric knowledge and the legacies of learning. In Rutten, R. & Boekema, F. (eds.) *The Learning Region. Foundations, State of the Art, Future*, pp 184 205. Cheltenham: Edward Elgar.

Cooke, P., Boekholt, P. and Todtling, F. (2000). *The Governance of Innovation in Europe. Regional Perspectives on Global Competitiveness*. London: Pinter.

Ennals, R. and Gustavsen, B. (1999). *Work Organisation and Europe as a Development Coalition*. Amsterdam: John Benjamin's Publishing.

Etzkowitz, H. and Leydesdorff, L. (2000). The dynamics of innovation: From national systems and 'mode 2' to a triple helix of university industry government relations. *Research Policy* 29, 109 123.

Fagerberg, J., Mowery, D. and Nelson, R. (eds.) (2005). *The Oxford Handbook of Innovation*. Oxford: Oxford University Press.

Florida, R. (1995). Towards the learning region. *Futures* 27(5), 527 536.

Gertler, M. (2004). *Manufacturing Culture: The Institutional Geography of Industrial Practice*. Oxford: Oxford University Press.

Giuliani, E. and Bell, M. (2005). The micro determinants of meso level learning and innovation: Evidence from a Chilean wine cluster. *Research Policy* 34(1), 47 68.

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology* 91, 481 510.

Hadjimichalis, C. (2006). Non Economic factors in economic geography and in 'new regionalism': A sympathetic critique. *International Journal of Urban and Regional Research* 30(3), 690 704.

Hassink, R. (2007). The learning region: A constructive critique. In Rutten, R. & Boekema, F. (eds.) *The Learning Region. Foundations, State of the Art, Future*, pp 252 271. Cheltenham: Edward Elgar.

Hudson, R. (1999). The learning economy, the learning firm and the learning region: A sympathetic critique of the limits to learning. *European Urban and Regional Studies* 6(1), 59 72.

Isaksen, A. (2005). Regional clusters between local and non local relations: A comparative European study. In Lagendijk, A. & Oinas, P. (eds.) *Proximity, Distance and Diversity: Issues on Economic Interaction and Local Development*, pp 129 152. Aldershot: Ashgate.

Jensen, M. B., Johnson, B., Lorenz, E. and Lundvall, B. A. (2007). Forms of knowledge and modes of innovation. *Research Policy* 36, 680 693.

Lorenz, E. and Lundvall, B. A. (eds.) (2006). *How Europe's Economies Learn: Coordinating Competing Models*. Oxford: Oxford University Press.

Lorenz, E. and Valeyre, A. (2006). Organisational forms and innovative performance. In Lorenz, E. & Lundvall, B. A. (eds.) *How Europe's Economies Learn: Coordinating Competing Models*, pp 140 161. Oxford: Oxford University Press.

Lovering, J. (1999). Theory led by policy: The inadequacies of the new regionalism. *International Journal of Urban and Regional Research* 23(2), 379 395.

Lundvall, B. Å. (ed.) (1992). *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*. London: Pinter.

Lundvall, B. Å. (2008). National innovation systems   Analytical concept and development tool. *Industry and Innovation* 14(1), 95 119.

Lundvall, B. Å. and Johnson, B. (1994). The learning economy. *Journal of Industry Studies* 1, 23 42.

Mitchie, J. and Sheehan, M. (2003). Labour market deregulation, 'flexibility' and innovation. *Cambridge Journal of Economics* 27, 123 143.

Morgan, K. (1997). The learning region: Institutions, innovation and regional renewal. *Regional Studies* 31, 491 504.

Piore, M. and Sabel, C. (1984). *The Second Industrial Divide: Possibilities for Prosperity*. New York: Basic Books.

Porter, M. (1990). *The Competitive Advantage of Nations*. London: Macmillan.

Putnam, R. (1993). *Making Democracy Work*. Princeton, NJ: Princeton University Press.

Rutten, R. and Boekema, F. (eds.) (2007). *The Learning Region. Foundations, State of the Art, Future*. Cheltenham: Edward Elgar.

Storper, M. (1993). Regional 'worlds' of production. *Regional Studies* 27, 433 455.

Woolcock, M. (1998). Social capital and economic development: Toward a theoretical synthesis and policy framework. *Theory and Society* 27, 151 208.

Leisure

J. Mansvelt, Massey University, Palmerston North, New Zealand

Glossary

Carrying Capacity This embodies the notion of limits to resource use based on the premise that resource and land use can only sustain a certain number of actions/uses by individuals, beyond which that use will have detrimental effects on the resource, the environment, and the user's experience. Practically difficult to quantify and to conceptualize, it is now largely considered a chaotic concept.

Commodification The process whereby more and more aspects of social life previously organized outside of the market assumes the commodity form.

Recreations Viewed as a subset and expression of leisure, recreations are understood as activities which are undertaken in one's leisure time. These involve the engagement of one's body and mind in forms of leisure in recognized activities which may have both active and passive attributes. Examples of recreations include sports, shopping, reading, exercise, hobbies, and crafts. Early geographical research on leisure primarily focused on outdoor recreational pursuits.

Residual Time Time available for discretionary use, after the time taken for subsistence and work-related activities is subtracted.

Time-Budget Survey A survey of individual's activities over a period of time which can be used to reflect on the balance between leisure and work in participants' daily, weekly, or monthly routines.

Leisure encompasses a wide range of social and spatial practices, experiences, and meanings which are socially and materially constructed in specific time and place contexts. While many definitions of leisure have been formulated and debated, some common attributes exist. Leisure is frequently presented as a state of being or mind, an intrinsic condition in which a sense of agency is exhibited. The capacity to choose to be involved in particular forms of bodily and mental activity is also associated with leisure, as is some choice over the form and nature of one's involvement. A pleasurable and/or playful quality of experience is also common to definitions (**Figure 1**) though this may be expressed as a desire for, or be an outcome of, engagement – as in the sense of pleasure obtained after undertaking strenuous physical or mental activity, or the anticipation of an event. Leisure also involves voluntary or discretionary use of time. Though leisure is usually associated with free choice, freedoms to engage in leisure may be relative – with preference, entitlement, access, availability, opportunity, and practice varying greatly between people and place.

Leisure and recreation are often used as interchangeable concepts but traditionally both recreation and tourism have been understood as subsets of leisure processes, that is, as activities which embody characteristics of leisure but which are separated on the basis of the form and nature of activity. Recreations tend to be categorized as leisure pastimes (which may include formally organized sports), while tourism has emphasized pleasure oriented behaviors and recreations away from one's home region. However, there is considerable overlap between these definitions. Any differences between tourism and leisure in local or distant spaces may be less significant than how such spaces are produced, negotiated, known, imagined, and experienced. The connectedness of tourism and leisure is also reflected within the discipline of geography where a number of professional study groups, recently established journals, and publications combine studies of leisure and recreation with those of tourism.

Dimensions of Leisure

Geographers have drawn on various dimensions of leisure in their research and writing. These perspectives have not only provided different insights into the production, expression, practice, and embodiment of leisure but have consequences for what processes, spaces, and peoples are included or excluded.

Leisure as Residual Time

Here leisure is understood as time spent in leisure after work and subsistence needs have been accounted for, with leisure measured in quantity terms through instruments such as time budget surveys and work life balances. Leisure is placed in binary opposition to productive work, both different from, and a reward for work. Yet, time available for leisure does not equate to time spent in leisure. Restrictions on access to leisure forms and practices, fragmented leisure time, and a lack of individuals' entitlement suggest the leisure–work binary is simultaneously problematic and revealing.

A number of chronologies suggest that leisure became separated from work temporally and spatially with the Industrial Revolution. This separation was associated with the imperative for individuals to 'sell' their labor to those who owned the 'means of production' in locations away from sites of residence. Thorstein Veblen, writing

Figure 1 Leisure as play. Leisure is seen as involving elements of freedom, pleasure, and play.

about the rise of an elite leisure class at the turn of the nineteenth century, linked leisure to the class based division of society necessary for the material and social reproduction of capitalism. He noted the emergence of a wealthy class able to spend time in engaging in plea surable consumerist pursuits suggesting that leisure should be seen as nonproductive consumption of time. A view of leisure as residual time is underpinned by a Eurocentric and modernist trajectory of historical change, locating leisure's meaning and spatiality in the specific changes which preceded and followed the In dustrial Revolution, and understanding them in terms of Marxian analysis.

Conceptualizations of leisure as residual time have also been associated with leisure's substantiation as an un earned privilege for those not engaged in the formal labor market including the unemployed, retired, and unpaid family or care workers. A lack of entitlement has spatial expressions, with research in Anglo American contexts revealing the home for women in particular is a complex site of leisure and work, and one in which leisure practices were not easily separated or experienced from daily work tasks occurring within and outside home space. It has also been suggested that the negative construction of the leis ure concept through its association with play and con sumption has been a factor in the relative neglect of leisure as a significant topic of geographical inquiry.

Leisure as Activities

When leisure is viewed as activities as in tourism, shopping, sport, recreations, hobbies, it is the characteristics and form of activity that are emphasized. This may include cat egorization of provision (public, private, voluntary); the nature of activity (engaging the body actively or passively); its spatiality (locations of engagement); and form of par ticipation (individual, collective, involving production and/ or consumption of artifacts, experiences, etc.). Emphasis is on the product rather than the process of leisure. Reading or watching television, for example, may be classified as passive leisure forms, yet they may actively engage both mind and body, and may produce both pleasurable and anxious, sensations, feelings, and emotions.

Despite difficulties in categorization, much leisure research in geography has drawn on a view of leisure as activities. Places have tended to be seen as the location in which they occur, though recently more attention has been given to how spaces have a role in shaping leisure practices. A view of leisure as activities has underpinned surveys on leisure participation rates and land use. Leisure activities have also formed the basis of govern ment policy interventions based on equitable distri butions of and access to leisure, sustainable management of land, and healthy living campaigns.

Understanding leisure in terms of activities has been instrumental in highlighting how leisure may assume particular functions or roles in society. Leisure activities may have a role in legitimating and sustaining state he gemony, in promoting continued social and material production, and in the surveillance of its citizens. States for example, may regulate, provide, or facilitate various leisure activities, legislating against unacceptable activ ities and promoting desirable recreations. Leisure activ ities can have a role in disciplining bodies – producing healthy citizens, occupying the 'idle', socializing children, and providing compensation for the unemployed. They have also been used to stimulate social change and eco nomic growth. In Britain in the 1980s, for example, leisure and tourism activities were promoted by central and local governments as instruments of urban regener ation, a means of improving quality of life for residents, and of stimulating employment through private sector business growth, inward investment, tourism.

Initially geographers have tended to focus on com modified and/or public leisure pursuits and active rec reations. Focusing on leisure activities tends to privilege those more easily categorized which can have the effect of obscuring the significance of passive, private, mundane, relational, and even deviant leisure experiences in everyday life.

Leisure as Experience

Understanding the form and nature of leisure activities reveals nothing of the quality of experience and intrinsic value for the individual. Drawing initially on humanistic research into outdoor, urban, and home based recre ations, this dimension of leisure highlights how space is made meaningful and knowable through experiences of leisure, emphasizing the creative and active potential of leisure regardless of its characteristics as activities. Leisure is a subjective and reflexive part of being, be coming, knowing, and making meaning in the world. Geographers have implicitly drawn on this notion in researching how leisure activities (such as surfing the net, going to rave dance parties, or gardening) operate as mediums for the production, consumption, and negoti ation of identities.

Reflecting on leisure as a subjective quality of ex perience enables one to consider how leisure activities and places have come in turn, to be objectively and ex ternally constructed, a means of considering how tem poral, social, and spatial categorizations of leisure are formed, and how these may operate to define acceptable and deviant, inclusive and exclusionary practices, sub jects, and places (**Figure 2**). In colonial nineteenth century New Zealand, for example, the concept of leisure emerged primarily from a British/European co lonial heritage, resulting in some early settlers assuming indigenous activities such as 'weaving' were a form of leisure activity rather than a significant form of Maori material culture linked to communal livelihood and identity formation.

Leisure as Practice

A final dimension of leisure is its constitution as a practice and encounter between people as embodied subjects and place. This perspective draws on nonrepresentational and post structural theories to argue that leisure experiences and encounters are not simply understood via the mind (as a particular gaze or sub jective experience) but through the body. This helps overcome the tendency to conceptualize individuals as constrained 'users' of space, instead viewing them as active and sentient subjects engaged in bodily, repre sentational, cultural, and economic practices. Hegemonic constructions of power are both affirmed and resisted in the practicing of leisure, and leisure practices/encounters become mediums for shaping experience, ritual, emotion, memory, knowledge, and action in the making of places and identities. This approach has been significant in so cial and cultural geographies of leisure and tourism in recent years by providing a way of connecting changing socialities, spatialities, and subjectivities. A view of leisure as practice has challenged the primacy of the visual in creating tourist landscapes and experiences, for example, arguing that tourist subjects, places, and place meanings are not simply inscribed through a particular 'gaze' but are negotiated and shaped powerfully in the doing of tourism/leisure.

Recent Social and Spatial Transformations in Leisure

Changes in society and space continue to inform understandings of what leisure is and how it is manifest. Since the end of World War II a number of social and economic changes have been suggested as contributing factors in the expansion of leisure industries and spaces. These factors include: continued urbanization; increasing mobility of capital; labor and populations; technological change; car and home ownership; increasing standards of living and discretionary spend; shorter and more flexible working hours. In the West, statistics indicate that over the course of the nineteenth and twentieth centuries hours of weekly work have been reduced and become more fragmented and flexible. An increase in paid holi days and shifts toward service based economies led commentators in the 1970s and 1980s to predict the ar rival of a postindustrial society, one in which leisure would be prioritized over work. The continued social, political, and economic dominance of paid work, the existence of the income rich–time poor, and unevenness and disparity in the geographical extent of these trans formations has meant the predictions of the arrival of a leisure society have not been realized.

However, processes of globalization have brought other transformations, resulting in more complex webs of economic, cultural, and political leisure activity con necting people and places across nation states and pro ducing new forms of material and discursive connection. The growth of international sex tourism, environmental

Figure 2 The social construction of leisure: Maasai women in Tanzania doing bead work. Maasai women wear elaborate beaded necklaces, armlets, and earrings, etc.; easy to assume such practices are leisure, but those being made here are for sale to tourists. Credit: Regina Scheyvens.

impacts of resource and land use, and the formation of electronic ghettos are just some of the issues which reflect how people and places tied into complex and uneven relationships across a variety of scales. Though spaces, activities, and forms of leisure have altered over time, leisure remains a site for the production and expression of power, and a key marker of identity and difference.

The widespread availability of a range of mass produced and niche commodities; the development of new consumer products, services, and sites linked to particular lifestyles and identities; rituals associated with possession, repair, display, and use of commodities have meant leisure activities are increasingly subject to commodification. Global media conglomerates dominate coverage of major sporting events and the numbers of individuals employed in leisure, tourism, sport entertainment industries globally are growing. The growth of media, communication and information technologies has also meant greater exposure – though not necessarily personal access – to a range of new leisure practices, spaces, lifestyles, and identities. In some countries neoliberal forms of govern ance have seen the marketization of government provided leisure services with subsequent changes in delivery, cost, and access to leisure facilities and spaces.

Though predictions of a leisure society have not been fulfilled, leisure spaces are a visible part of landscapes in many contemporary societies. Their presence is often connected to commodified leisure practices with many of these spaces operating as sites of public though often privately owned leisure in the form of shopping malls, exhibitions, tourist spaces, theme parks, and multiplex cinemas (**Figure 3**). However, an emphasis on spectacular spaces and narcissistic, commodified, ostentatious leisure practices obscures the significant ways in which leisure also occurs in private and home spaces, and how mundane leisure practices such as eating, listening to music, or sharing time with friends are also productive of meaning, identities, social relations, and geographies which may extend beyond the locality in which they occur. Relations of gender, sexuality, class, disability, ethnicity, and age are also performed and expressed powerfully through leisure and social and spatial heterogeneity continues to be a feature of leisure practices and sites. Thus it is important to see many of the trends outlined here as generalizations, some with a much longer historical trajectory than their visibility as twentieth or twenty first century phenomenon might suggest.

Leisure Geographies

The development of leisure geographies – of scholarship within the geography discipline – can be linked to changing theoretical perspective and conceptualizations of leisure, as well as the social and spatial transformations outlined previously. While a broad chronology can be identified it should not be seen as rigid, with the different strands of research extending beyond the period in which they are identified and to different extents in different places. Positivist research, for example, was much stronger in the US, than in Europe. In Europe and the United Kingdom prior to 1980s tourism studies dominated leisure research, while in North America the emphasis was on recreation. In the German speaking world scholars have been as prolific as those in Anglo American countries with both the former West and East Germany following different research paths. The work of geographers outside European and Anglo American contexts has been less visible, with the consequence that leisure geographies themselves appear to have been dominated and framed through a small core of geographers working in First World contexts. There has also been a substantive volume of geographic work in tourism and recreation occurring outside formal academic departments which has contributed significantly to public policy, private sector, and government expertise.

1930s–1950s

Though leisure research has been undertaken by geographers from the 1930s, it has not had a strong presence within the discipline. Much of the early work constructed leisure as activities, primarily recreational pursuits based in the 'countryside' with a concern for issues of scenic amenity, supply, and demand. It is perhaps not surprising that landscapes and sites of leisure have long been a focus of studies in leisure. Early research was largely descriptive and focused on inventories of recreational use with emphasis on land management, planning, carrying capacities, resource use, and the mapping of aggregate spatial patterns.

Figure 3 Outdoor pedestrian shopping mall, Brisbane. Commodified shopping spaces as sites of leisure.

However, it was not only rural recreation activities which attracted geographers' attention. The development of tourist and coastal resorts and the economic contribution of recreation and tourism activities to regional development were also explored as were geographies associated with these such as traffic patterns and the purchase of second homes. Models adapted from economic and industrial geographies were applied to recreation and tourism using quantitative methodologies and techniques to explain patterns and behaviors. Examining leisure as activities occurring in places enabled mapping of land use and behavioral patterns which were often framed in relation to public policies concerning regulation and management of land, recreation, tourism, and health.

In the decades following World War II, increasing urbanization, mobility of populations, discretionary income, greater personal access to the motorcar, and availability of time and money for travel meant rural recreations and the effect of leisure participation on the landscape continued to be a focus of recreation research.

1960s–1970s

A bibliographic survey of literature on outdoor recreation in 1964 noted that research in the field still sat on the margins of geographical enterprise. The perceived frivolity of leisure research was a recurring theme in disciplinary reflections during the next two decades. Nevertheless leisure research began to have a more visible and vocal presence within the discipline during this period, influenced by increasing tension and awareness of the politics of resource, land, and consumer demands. More attention was given to recreational activity in urban areas and to the theorization of economic, behavioral relationships and perceptions, along with further work on resource and management implications in context of growing affluence in the Western world. During the 1960s and 1970s greater numbers of geographers identified themselves as leisure specialists and a number of textbooks on leisure and recreation were produced. Though post positivist theoretical approaches and qualitative research methods began to impact geography, research underpinned by positivist philosophy dominated leisure geographies until the 1970s.

Speculation about increasing significance of leisure in individuals' lives, increases in discretionary income, and decades of investment by governments in leisure services as part of welfare state policies meant there was optimism that leisure research might make a major contribution to society and geography. However, as many economies entered recession in the late 1970s and 1980s the future of leisure as a predominant sphere of social and academic life appeared less certain. Calls for the emergence of a distinct geography of leisure remained and by the late 1970s attention was shifting toward understanding leisure as experience as part of human need, desire, self actualization. Research informed by humanistic perspectives emphasized leisure as a qualitative and subjective experience important to the construction of place, informing explorations of leisure's relationship to lifestyle and social identity, and studies of 'micro' spaces such as home and garden.

1980s–1990s

In the early 1980s research on 'leisure lifestyles' coincided with acknowledgment that leisure was important in contributing to quality of life and to a quality of place. Though the effects of global recession impacted countries unevenly, they did stimulate research on the economic impacts of investment in leisure, recreation and tourism, sports geographies, and in particular their role in ameliorating some of the undesirable consequences of economic change. The early 1980s also brought explicit recognition of inequalities and deprivation in leisure provision. Access to and involvement in leisure was linked to functional rationales, with the social, spatial, political, and economic significance of state provision, voluntary leisure involvement, private sector leisure industries, and the professionalization of sport highlighted. Differential access to and uptake of leisure activities led to research on leisure constraints imposed externally to or internally by the individual. Much attention was given to the gendering of leisure, particularly in relation to Western women's experiences of raising nuclear families in the home, though it was scholars outside the geography discipline that contributed most to analyses of gender and leisure during this time. These early feminist studies also provided a critique of work–leisure binary recognizing its location in structures of production and paid work.

Attention to leisure constraints and relative freedoms in leisure were also informed by Marxian theorizing in which leisure could be conceptualized as part of the sphere of social reproduction, a means of creatively spending time to address alienation experienced as a consequence of the separation of one's labor from its products. Growing commodification also signaled the importance of the leisure sphere in stimulating purchase of commodities – a necessary part of the continuation of the capitalist mode of production. Leisure geographies broadened to include a whole range of spaces and places and much more attention was given to urban and home based spaces of recreation. The role of leisure as part of consumer and tourist landscapes, its significance in the development of cultural economies and in economic development was also explored. A small but significant body of work on sport emerged, research which continues to reflect on the economic and cultural significance of sporting practices and spaces.

1990s–2000s

The last decade of the twentieth century saw a more sustained engagement by geographers with leisure, especially with regard to its relationship to consumption (particularly through shopping, eating, and music). However these studies simultaneously pointed to the dangers of considering the connection between consumption and leisure as a necessary one, highlighting the ways in which leisure meaning, experiences, and practices are shaped across numerous spaces, practices, discourses in commodified and noncommodified moments. Shopping, for example, may be an act of mundane necessity and of commensality not necessarily the pleasurable, private, identity ridden, and superficial practice linked to a postmodern condition. The emergence of the 'new cultural geographies' has to some extent reinvigorated the theorization of leisure and its place within both society and the academy. While studies continue of leisure economics, patterns, and recreational land use, attention has also shifted to debates about 'environments' and landscapes as sites of social exclusion and sustainability. Renewed attention has been given to the theorization of leisure as a site of practice and encounter. This new emphasis has seen geographers engage with a range of theoretical approaches including post structural, postcolonial perspectives and actor network approaches. This has included examinations of the ways in which people and places are connected materially, discursively, and differentially in the production, consumption, and representation of leisure spaces and subjects, and in processes of identity formation, subjectivity, embodiment, and performativity.

One of the impacts of the new cultural geographies has been to highlight how leisure practices are not functions of singular and undifferentiated categories such as gender. Rather, leisure is shaped through practices which may involve complex intersections, formations, and narratives of gender, sexuality, disability, ethnicity, age, and socio economic status. These practices and encounters in turn create a range of landscapes in which people are both subject to and subjects of discourses. Though representation of leisure activities, choices, spaces, and times may be a significant aspect of contemporary leisure practice, attention to the role of leisure as practice in geography also acknowledges the importance of the body and emotion in construction and expression of leisure processes and places. In the twenty first century there is continued emphasis on leisure in disciplining and inscription of the body through activities as diverse as eating (**Figure 4**), body piercing, exercise, outdoor recreations, and holidays in a wide variety of contexts: the Internet, gym, street, shopping mall, gardens, beaches, resorts, theme parks, mountain and national parks, sports venues, homes, spaces of travel, and in rural areas. Space is no longer viewed as a location for leisure but is part of the way in which both place and leisure practice are shaped materially and metaphorically. This more fluid construction of space has enabled researchers to make linkages between scales previously seen as separate domains. The role of leisure in shaping responsible nation state citizens, for example, can be linked to particular practices at other scales – home provisioning, gardening, or listening to music. Scholars have also recognized the need to broaden examination of

Figure 4 Restaurant diners. In recent years geographers have focused on more mundane practices of leisure and the social, cultural, and material contexts in which they are shaped.

leisure to other contexts – to increase visibility of research undertaken of leisure in spaces outside Anglo American and World War I contexts and to continue to focus on a range of commodified and noncommodified spaces and practices of encounter.

Through research on activities, encounters, land scapes, and experiences of leisure, geographers have highlighted the role of leisure as a meaningful place making activity. Because leisure is expressed through many spaces, practices, imaginings, and experiences it provides a means of reflecting on and contesting par ticular binaries: between cultural and state politics, so cial and individual, public and private, material and representational, economic and cultural, production and consumption. While leisure geographers initially made a small contribution to human geography, the resurgence of both academic and empirical studies in the twenty first century mean leisure should be considered a sig nificant theoretical and political sphere, providing valuable insights into relationships between society and space.

See also: Consumption; Feminism/Feminist Geography; Humanism/Humanistic Geography; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Non-Representational Theory/Non-Representational Geographies; Postcolonialism/Postcolonial Geographies; Poststructuralism/Poststructuralist Geographies; Surveillance; Tourism.

Further Reading

Aitchison, C. (2006). The critical and the cultural: Explaining the divergent paths of leisure studies and tourism studies. *Leisure Studies*. 25(4), 417 422.

Aitchison, C. (2003). *Gender and leisure. Social and cultural perspectives*. London: Routledge.

Aitchison, C., MacLeod, N. E. and Shaw, S. J. (2000). *Leisure and tourism landscapes. Social and cultural geographies*. London: Routledge.

Brown, B. (2004). Hotdeskers and tourists: Geography as an everyday practical concern in work and leisure. *Geography* 89(1), 71 77.

Butler, R. (2004). Geographical research on tourism, recreation and leisure: Origins, eras and directions. *Tourism Geographies* 6(2), 143 162.

Crouch, D. (2000). Places around us: Embodied lay geographies in leisure and tourism. *Leisure Studies* 19, 63 76.

Crouch, D. (ed.) (1999). *Leisure/tourism geographies. Practices and geographical knowledge*. London: Routledge.

Hall, C. M. and Page, S. J. (2002). *The geography of tourism and recreation. Environment, place and space*. London: Routledge.

Haywood, L., Kew, F. and Bramham, P. (eds.) (1995). *Understanding leisure*, 2nd ed. Cheltenham: Stanley Thornes.

Kelly, J. R. and Freysinger, V. J. (2000). *21st Century leisure. Current issues*. Boston: Allyn and Bacon.

Kreisel, W. (2004). Geography of leisure and tourism research in the German speaking world: Three pillars to progress. *Tourism Geographies* 6(2), 163 185.

Terreell, S. (2004). *Access to geography series: The geography of sport and leisure*. London: Hodder & Stoughton.

Veblen, T. (1899). *The theory of the leisure class*. New York: Augustus.

Wolfe, R. I. (1964). Perspective on outdoor recreation: A bibliographic survey. *Geographical Review* 54(2), 203 208.

Relevant Websites

http://www.academyofleisuresciences.org
Academy of Leisure Sciences.

http://www.anzals.org.au
ANZALS, Australian and New Zealand Association for Leisure Studies.

http://www.geog.nau.edu
Commission on the Geography of Tourism, Leisure and Global Change, International Geographical Union.

http://www.nrpa.org
Journal of Leisure Research, National Recreation and Park Association.

http://www.tandf.co.uk
Leisure Sciences, Taylor & Francis Group.

http://www.tandf.co.uk
Leisure Studies, Taylor & Francis Group.

http://www.tandf.co.uk
Tourism Geographies, Taylor & Francis Group.

http://www.worldleisure.org
World Leisure Organization.

Lesbian Geographies

K. Browne and C. J. Nash, University of Brighton, Brighton, UK

Glossary

Lesbian This label is a contemporary or modern term often taken to refer to women who identify with or are identified as 'lesbians' within particular political, social, and/or cultural contexts. This usually pertains to their sexual and emotional practices, desires, relations, and relationships with women. However the category may be deployed even though the individual may not now be, or ever have been, engaging in same sex relationships or sexual practices. The concept of 'lesbian' and its relevance as a coherent or even useful category has come under close scrutiny by post-structural feminist and queer theorists.

Queer Critically examines the creation of normativities and looks to deconstruct these usually, but not solely, by investigating the normative links between gender, sex, sexuality, desire, primary relationships, and sexual practices. Queer theory has critically addressed homonormativity and heteronormativity using thinking that is nonessential, relational, and contingent in nature.

Strategic Essentialism Is the use of identities such as woman, man, lesbian, gay, or queer to represent what are heterogeneous groupings of individuals in order to effect social change for these groupings. Strategic essentialist actions usually involve individuals and groups uniting for a common cause even though their experiences of being a woman, man, lesbian, gay, etc. may be different, hierarchical, and contradictory.

Introduction

Lesbian geographies refer to scholarly research on the ways in which lesbians organize themselves in both rural and urban landscapes. These geographies are not coherent or fixed and have changed through time and are different across different spaces. Arguably, lesbian geographies initially emerged in response to the male dominance of early work on the geographies of sexualities. Work on lesbian geographies spans the subdisciplines of feminist geographies and geographies of sexualities. It often queries the heterosexuality of some feminist geographies and the male dominance of geographies of sexualities and more recently queer geographies. In contesting the normative positioning of white, middle class gay men in geographies of sexualities research, recognition of multiple social differences, including gender, that constitute the hierarchical sexualization of space has begun to surface. Recently, the very subject of enquiry – the lesbian – has come under scrutiny as scholars have engaged with feminist, post-structuralist and postcolonial analytic frameworks. The querying of lesbian geographies and the unfixing of gendered binaries within discussions of sexualities has problematized the gendered analysis of the sexualization of spaces. In this article we will outline the urban beginnings of lesbian geographies within the context of lesbian social movement politics within and beyond geographies before moving on to consider ruralities. The article will then examine the destabilization of 'lesbian' as an identity and the challenges posed by queer analyses to lesbian geographies and to gendered discussions of sexualities more broadly.

Lesbian Politics and Key Debates

Lesbian political activism and grassroots organizing has taken diverse and varied forms and has played a key historical and cultural role in the emergence of lesbian geographies and in research expressly directed at understanding lesbian spatial organizing. Lesbian political activism includes lesbian interventions in the broadest political spheres as well as the micro politics of lesbian spaces and engagements. Prior to examining various lesbian geographies, it is important to place these political activities within broader feminist and lesbian contexts. The relationship between this political activism and the constitution of lesbian geographies has been addressed by geographers, such as Valentine's examination of lesbian separatism in rural areas (see below) and Nash's discussions of queer women's bathhouses. However, there are many debates that occurred in other disciplines regarding lesbian politics that have yet to be addressed in geographical research. This is in part, because, work on lesbian geographies emerged after these debates had occurred in wider feminist political and academic circles.

One of the key feminist debates, occurring in the late twentieth century, concerned the need to recognize the diversity of the social category of 'women'. Feminist politics often used a homogenous understanding of 'women' and 'women's needs' that was actually based on knowledges grounded in white, heterosexual, middle-class, able bodied women's understandings. The homogeneity of women was contested by women of color, working class women, disabled women, and lesbian women who argued that the experience of being a woman was not the same for all women, rather gender was mediated by other social differences that could not simply be 'added on' to the category 'woman'. Instead these social differences, they argued, fundamentally recreated gendered experiences.

These internal divisions were often seen as a threat to the calls for equality that required a shared female identity under the sign 'woman'. This shared identity could be seen as fragmented and existing only for the advancement of specific causes, for example, equality legislation. Uniting for the purpose of advancing a particular agenda, whilst still recognizing the problems of essentializing and homogenizing the category 'woman', is called strategic essentialism. This can be seen in the geographical writings of early feminist geographers who argued for the inclusion of women in geography without addressing diversity between women. Yet it let to an invisibilization of lesbians in early feminist geographies.

In feminism more broadly, some scholars argued that heterosexuality itself was an inherent aspect of patriarchal social relations. Scholars and others engaged in feminist politics chose to engage in lesbian relationships in order to contest the heteropatriarchy of institutions such as marriage as well as address inequities between men and women in interpersonal relationships. This is termed as lesbian feminism and is associated with lesbian separatism where women chose to live in all female communities and communes, a form of spatial organizing constituting distinctive lesbian geographies. These separatist communities sought to establish different ways of living that were not hierarchized or based on individual family units. Relationships of power were critically addressed as women moved outside what was/is understood as the patriarchal dominance of men and the family unit as a site for the (literal) reproduction of this dominance.

The fragmentation of identities and categories such as 'women' and 'lesbian' associated with queer and other post structural conceptualizations of identities (see below) has had a dramatic effect on lesbian politics and wider engagements with gendered structures of power. Although there is some evidence of lesbian strategic essentialism, queer activism and political action has challenged the essential identities of 'woman' and lesbian that lesbian separatists and lesbian feminists have deployed. Nevertheless, there continues to be lesbian direct action groups, such as the lesbian avengers, and lesbian separatist communities and festivals continue to run, particularly in the USA. Lesbian geographies are only beginning to engage with the complexity of the historical and contemporary politicization and contestation of lesbian identities.

Urban Beginnings

Research on lesbian spatial organization surfaced in the scholarly literature considering the evolution of gay male residential and commercial districts in the post World War II period. Preliminary research focused on the spatial organization and location of gay bars and commercial establishments in several cities in the United States. The first major consideration in geographical work of gay male organization in urban landscapes was undertaken by Manuel Castells and Karen Murphy in their study of the gay male population in San Francisco. Although Castells and Murphy acknowledged that lesbians were present in the gay male neighborhood in San Francisco, Castells noted that lesbians did not organize spatially in the same way as gay men. He speculated that the 'inherent differences' between men and women themselves constituted the foundation for their seemingly distinctive relationships to urban spaces. Castells and Murphy claimed that unlike men, women rarely have these territorial aspirations: their world attaches more importance to relationships and their networks are one of solidarity and affection. In this, gay men acted 'first and foremost as men and lesbians as women' (1983:139). Subsequent geographical research quickly challenged this 'innate territoriality' theory with some scholars arguing that the lack of lesbian urban spaces was because society reacted differently to gay men. Gay men, they claimed, experienced different forms of oppression, making territorial strength and control an imperative for gay men.

For others, the lack of lesbian defined territories actually reflected men's and women's different economic and social circumstances. Lesbians, operating outside the accepted gender roles of wife and mother were arguably less able to compete economically and socially than men. Gay men, as men, however, were more financially and socially capable of organizing spatially and upgrading neighborhoods. The lack of visible, spatially based lesbian community had less to do with lesbians being 'less territorial' than gays and more to do with the gendered economic realities of the marketplace, including limited access to capital and other resources.

Research on lesbian spatial organization in urban areas began to emerge in the late 1980s. Some of the earliest work was on social organization of working class lesbian communities in the 1950s and the 1960s and documented highly organized butch/femme cultures in cities such as Buffalo, New York and Toronto, Canada. This working class lesbian visibility seemed to reflect the fact that it was more socially acceptable for working class women frequent downtown taverns and restaurants than middle or upper class women. At the same time, women's employment opportunities were also more restricted, limiting access to the economic resources needed for home ownership or business start up. Other research also demonstrated that women may not have been visible in the gay male ghetto or village due to overt hostility to the presence of women in gay male space. In Toronto, in the 1960s and early 1970s, for example, gay male restaurants bars and taverns often charged women a higher cover charge for admission, insisted on women having a male escort, and requiring women to wear skirts or dresses. Dress codes were designed to exclude 'butch lesbians' from gay male spaces as they were thought to be troublemakers and prone to starting fights.

Geographical research focusing on newly emergent lesbian residential organization began in the early 1990s. Initial research sought to examine lesbian residential concentrations and to compare these sociospatial patterns to those of gay men. This research was particularly interested in considering whether the absence of a publicly identifiable lesbian neighborhood reflected inherent gender differences in interests, needs, and values or reflected differences in the resources available to gay men and lesbians. Identifiable lesbian residential neighborhoods did exist although they did not have the associated lesbian controlled, commercial focal point. Research demonstrated lesbian neighborhoods tended to be found in downtown cores in countercultural or alternative neighborhood and by comparison, were much less visible than similar gay male areas and did not have the degree of territorial control and affiliated political and social activism. Researchers concluded that far from being less territorial than gay men, lesbians, as women and as primary caregivers, had different concerns about the kinds of areas that they occupied. City centers were not, in general, attractive locations for women because of safety concerns for both themselves and their children. Downtown areas were more economically attractive, containing lower cost housing, good public transit, and access to social services. Arguably, lesbians also did not have the financial capability to be involved in the gentrification process to the same extent as gay men – a specifically gendered and class related reasoning that draws on and reconfigures questions of sexual preference or orientation.

In addition to the obvious social and economic constraints on lesbians as women in terms of their locational choices, scholars also argued that lesbian politics in the 1970s and 1980s was closely allied with the women's movement and their ideological leanings were more global and less restricted to representation in the existing political system. Scholars argued that lesbians were less focused on controlling space for political power and more interested in addressing broader theoretical issues outside of the existing social structure. Nevertheless, research on urban spaces did demonstrate that lesbians used a number of different strategies for negotiating urban spaces and appropriating opportunities for socializing and political activism.

Academics began to rethink their approach to understanding the evolution of distinctly lesbian spaces starting with the basic premise that gay and lesbian spaces have evolved through completely distinctive processes and function in distinctive ways. Lesbians can appropriate and use public sites in visible ways, yet not have a territorial presence. Julie Podmore, for example, argues against the tendency of lesbian geographies to focus on the absence of women in heterosexual space or their seeming lack of expressly territorial clusterings. Arguably, how lesbians use public sites for interactions and communality may not fit within the territorial paradigms of urban geographies. Podmore demonstrates how women who exist outside heterosexuality, clearly use and experience 'heterosexual spaces' in ways that may not 'fit' traditional geographical investigations. From this perspective, research on lesbian geographies demonstrates a plethora of ways by which space is appropriated, used, negotiated, resisted, and recreated. Spatial appropriation and recoding through music, clothes, and language, as well as more political activities (such as 'dyke' marches) and desiring enactments (expressions of same sex desire) mark the resigning of space outside of heterosexual norms.

Lesbian Ruralities

Although initial work on lesbian geographies focused on urban areas, other work has considered the spatial organization of lesbians in rural communities. These studies have explored migration from rural areas to urban areas as well as lesbian lives in rural arenas. To date this work is less developed then the urban literature and this remains the case despite calls within rural geographies to explore those who are excluded when considering 'rural idylls' within dominant framings of sexuality, gender, class, and disability.

Although there remains a lacunae in understanding lesbian spaces beyond urban territories, some work has been undertaken. Valentine showed how some lesbian separatists in the 1970 saw the rural as an escape from 'man made' cities and an opportunity to develop communities away from patriarchy and exist closer to 'mother nature'. These separatist lesbian feminists appropriated the 'rural' as an ideal which allowed them to live separate from men in a self sustaining way. This also had symbolic and spiritual resonances with 'mother nature' and engaged with what can be understood as 'essential' women's qualities (caring, nurturing, and mothering) and other forms of ecofeminism. Valentine, however, problematizes these generic models of 'woman' and considers the divisions within these lesbian separatist communities. Despite the ethos of inclusion and equity that contested man made cities, there were and are issues of exclusion within these rural lesbian communes. For example, some communes had environments and workloads which were solely formulated for able bodied women.

Other studies considering rural lesbian lives examined gentrification by lesbians who were engaging in processes similar to nonlesbian migrants in terms of migration and rural gentrification. These point to the problematic assumptions of difference associated with 'otherness' in ruralities, where lesbians may engage in similar patterns of accumulation and capitalist consumptions practices.

These studies emphasize clustering and lesbian communities. They thereby replicate earlier urban work that assumed particular forms of territorialization as the sole focus of lesbian urban geographies. There has been a dearth of enquiries that have focused on rural lesbian lives and their spatial practices outside of the urban arena as well as a failure to explore the mutual constitution of urban and rural spaces beyond migrations to urban arenas.

The underdevelopment of rural lesbian geographies can perhaps be explained with reference to the critiques of urban literatures. Alongside, the focus on visible territories in cities, in particular the examination of 'gay ghettos', there is often an underlying assumption that cities are places of tolerance and acceptance. This exists in contrast to what are perceived as hostile rural spaces. The migration literature, in discussions of lesbian and gay sexualities, often contends that lesbians and gay men leave rural spaces for the safety of cities. Their participance in these studies tend to be men and women who have left rural areas in order to live in cities (often with identifiable 'gay' areas). Although these stories are pertinent and clearly have importance in the discussions of lesbian and gay geographies, there is a need to examine sexualities in the rural arena and perhaps beyond 'citified' identities such as lesbian and gay.

Where lesbians are often rendered invisible in discussions of urban landscapes, the absence of discussions of female appropriation and use of rural spaces is unsurprising. The dearth of investigations that address rural lesbian lives may also be due to the emphasis on the ownership and use of public spaces more generally. Interestingly, examinations of rural homosexualities have looked at public sex spaces and men who have sex with men. These studies have pointed to how identifications such as 'gay' can be read as urban and unrelated to sexual practices. This troubles the ubiquitous use of these categories as common identifiers of sexual identities and recognizes their spatial contingency, which in this case can differ between the urban and the rural. In recognizing the spatial associations of identity categories these writers have engaged with emerging work that is deconstructing and spatializing sexual identities and identifications.

Questioning 'Lesbian': Theorizing Sexuality and Identity in Space

Early work took a fairly untroubled view of the subject position 'lesbian' and of the spaces and places associated with those experiencing same sex desire. By the early 1990s, the geographical literature began developing a more nuanced and critical view of the complex interrelationships between identities, subjectivities and the constitution, maintenance and transgression of spaces. Much of this work drew on the work of Michel Foucault and postmodernist and post structuralist theorizing on the nature of identity, subjectivity, and the spatial. Feminist based research drawing on this work conceptualized a more nuanced and historically and culturally specific perspective on gendered and sexualized identities in circulation in specific places.

Scholars have demonstrated how the subjectivities and identities of women identifying as 'lesbians' have varied considerably in the period after World War II and that these various and shifting subject positions closely paralleled changing political and social activism. In this context, the previously taken for granted category of 'lesbian' could no longer be taken as the unproblematic starting point for geographical enquiry. Applications of postmodernist and post structuralist analytics also complicated geographical thinking on the nature and constitution of space itself. Scholars began theorizing the heterosexual nature of both public and private spaces and the ways in which power relations work to inscribe spaces with disciplinary discourses normalizing heterosexual gendered and sexual behaviors and practices. Normative expectations about gendered and sexual behavior in public spaces ensured that alternative sexual identities were rendered invisible or overtly eradicated from the landscape. Appropriation and occupation of urban spaces by non normative sexual and gendered groups such as gays and lesbians was reconceptualized as forms of transgression and resistance within relations of power. A focus on the operation of various modalities of power highlighted the uneven social relations within gay and lesbian spaces not only along gender but along class, race, age, and ethnic lines as well. This prompted a greater appreciation of the ways in which lesbian geographies offered not only the possibility of resistance from normative heterosexual inscriptions of place but contained their own internal limitations, exclusions, and repressions. Recent work has illustrated how a dominant gay and lesbian political agenda has resulted in the constitution of homonormative spaces in which particular political, social, and economic ways of 'being' gay or lesbian are privileged over others.

In questioning identity categories, those from the majority world have queried the ubiquitous use of the term 'lesbian' (as well as gay and queer) in understanding same sex relationships and non normative sexual lives. Geographical variations in how sexuality and gender are understood and lived have led scholars to argue that 'lesbian' (along with 'gay' and 'queer') as a minority world concept(s). We use the term majority/minority worlds here as what is termed the 'third/developing' world is actually the majority world in terms of population, pro duction etc., and to question the geographical accuracy of the term 'Western world'. This term is used also to challenge the hierarchization of terms such as 'first world', 'third world', and so on (see First World, Third World). Thus, 'lesbians' can be a concept that is located in the minority world (as well as a historical specific one) and one that may, or may not be acknowledged or understood in different historical periods and across the world. In some countries, such as Indonesia for example, although same sex relationships and sexual practices occur among women, lesbian (and gay) identities are seen as dangerous and 'not Indonesian'. These relationships can conform to certain gender roles regarding male and female responsibilities but as Wieringa has shown, these gender roles are not fixed within the boundaries of nor mative sexed bodies in Indonesian gender relations. 'Butch' women can engage in gendered relations such that relationships are not read as 'deviant' or 'lesbian'. From a post stonewall 'out' generation of minority world lesbian and gay activists, this is often read as 'repression', 'denial', and 'closeting', however, these terms need to be critically assessed in relation to geographical specificity and globalizing power relations. They rely on specific conceptualizations of sexualities built on particular forms of minority world identity politics.

The identity politics and identifications within the category 'lesbian' are also historically, as well as cultur ally, specific. Faderman documents how the identity category of 'lesbian' has only been recognized recently as an identity category. Tracing the existence of same sex desire and love relationships from the sixteenth century to the twentieth century, Faderman argues for the need to contextualize lesbian identities within particular his torical periods and cautions against reducing love rela tionships and same sex enactments to twentieth century conceptualizations of identities. More recently, trans women and transmen have been reclaiming identifi cations that were once historicized within a 'lesbian' paradigm, in particular cross dressing and women passing as men have come under scrutiny. These gender variant lives are the site of some contestation and are arguably not properly understood within contemporary con ceptualizations of gender variance and sexual identities. Yet, there continues to be a need to address the invisi blization of gender variant and women's same sex histories.

Queering Identities

Lesbian geographies, whether conceptualized as inherent or socially constructed, tended to presuppose a co herence and unity within the category 'lesbian'. In the last decade or so, scholars drawing in newly emergent analytical frameworks captured in the term 'queer the ory' further complicate understandings of both lesbian identities and geographies. Here, the term 'queer' in questioning hetero/homonormativities can harbor a variety of sexual and political articulations including lesbians, gays, bisexuals, bois, grrls, tranny boi's, trans women, transgenderists, transsexuals, transvestites, cross dressers, and drag queens. The 'queering' of lesbian geographies has involved the questioning of the category of 'woman' as embodied in a seemingly fixed biological body and therefore a collective sign under which female same sex desires, practices, and relationships could be placed. In fact, the body itself can become the literal site of physical reworking in ways that eradicate any clear correspondence between bodies/subjectivities and iden tities. With these rearticulations of bodies and gender and sexual performances of identities and subjectivities, les bian geographies can no longer be understood as places occupied by unproblematic female bodied 'women' act ing on same sex desire. Queer spaces, unlike their pre viously understood 'lesbian' spaces, contain and maintain a variety of differently embodied persons whose sexual and gendered as well as race, class, and ethic articulations are understood as transit and unstable.

Queer theorizing, in deconstructing the normative boundaries of gender and sexuality, questions the very possibilities of coherent female identifications under the sign 'woman'. This has been read by some as contesting the very possibilities of 'lesbian' just as this (and other) identity categories were beginning to gain recognition and possibilities of engagements with hegemonic power structures. Whilst playful forms of gender transgression are celebrated in queer theories and writings, there continues to be a tension between lesbian feminist sep aratists and transwomen in particular. These tensions are often reduced to discussions of 'essentialist' lesbian feminists and the celebration of gender variant identities. Essential ideals about sex and gender that link biological sex with womanhood have been questioned, yet there are still those who argue that social constructs have dis empowered women and these draw on biological norms that are different from trans experiences of womanhood. In these discussions the boundaries of sex, and con sequently, lesbian bodies and identifications are both contested and reestablished.

This clearly has implications for how 'lesbian' is de ployed, contested, recuperated, and reinscribed. Although queer is sometimes used to describe the breadth of les bian, gay, bisexual, trans, questioning, unsure, and gender variant, this is contested where queer is seen as anti normative. In this context, replacing the term 'lesbian' with 'queer woman' in order to recognize the fluidity of female embodiments, sexualities, and acknowledge the importance of gendered power relations in the con struction of lives, identities, and spaces is problematic.

This potential impasse, it could be argued, has led to a lesbian lacuna within queer geographies literature. Recently, research has begun to focus on intersection of gender and sexuality in public space. Some scholars argue that, in fact, it is gender transgression, as well as sexual orientation that is the marker for regulation and exclusion in pubic spaces.

Research examining public washrooms and transgressive female embodiments has argued that contesting gender identities also problematizes categories of sexualities. These categories can be recuperated at multiple sites from the repositioning of bodies that are read as male into categories of 'abnormal female' to public washrooms where the boundaries of gender are policed through the 'checking' of gender identities. What is interesting in this work is that it extends discussions of performative geographies to include gender and sexuality. Specifically, this work argues that gender and sex (along with sexualities) are constituted through different spaces. In taking on DeBeauvoir's call that one becomes a woman, this work contends that women (and men) are remade in part through the spaces they inhabit and the policing of these spaces. These spaces include lesbian spaces where women described being asked to leave because they were read as 'straight men'.

Conclusion

Engagements with lesbians in geography came slightly after the advent of gay geographies and examinations of gay (male) ghettos. There still continues to be a dearth of lesbian rural geographies but perhaps the biggest challenge to this area of enquiry is the birth and growth of the subject position 'queer' as well as the evolution of queer theories and practices. This is augmented by the use of deconstructive theoretical moves that contest both the sign 'woman' and the possibilities of 'lesbian'. Although gender continues to be a salient category of analysis, those who are lesbian identified can feel 'under siege' in academic and queer circles. The privileging of 'queer' gender and sexuality orientations and identifications particularly in youth cultures can render those who paved the way for such cultures, and particularly lesbian feminists as 'out of date' and 'passé'. This can be read as a dangerous move where queer engagements increasingly question gender boundaries and yet the gendered power relations that construct geographies and lived spatialities are underexplored.

See also: Gay Geographies; Heteronormativity; Queer Theory/Queer Geographies.

Further Reading

Adler, S. and Brenner, J. (1992). Gender and space: Lesbians and gay men in the city. *International Journal of Urban and Regional Research* 16(1), 24 34.

Bell, D. and Binnie, J. (2004). Authenticating queer space: Citizenship, urbanism and governance. *Urban Studies* 41(9), 1807 1820.

Bell, D., Binnie, J., Cream, J. and Valentine, G. (eds.) (1994). All hyped up and no place to go. *Gender, Place and Culture* 1(1), 31 47.

Browne, K., Lim, J. and Brown, G. (2007). *Geographies of Sexualities*. London: Ashgate.

Castells, M. and Murphy, K. (1983). *City and Culture: The San Francisco Experience. The City and the Grass Roots: A Cross Cultural Theory of Urban Social Movements*. London: Edward Arnold.

Davis, M. D. and Kennedy, E. L. (1993). *Boots of Leather, Slippers of Gold: The History of a Lesbian Community*. New York: Penguin Books.

De Beauvoir, S. (1949). *The Second Sex*. Harmondsworth: Penguin.

Faderman, L. (1981). *Surpassing the Love of Men: Romantic Friendships between Women from the Renaissance to the Present*. New York: Harper Collins Publishers.

Holcomb, B. (1986). Geography and urban women. *Urban Geography* 7, 448.

Knopp, L. (1990). Some theoretical implications of gay involvement in an urban land market. *Political Geography Quarterly* 9, 333 380.

Knopp, L. and Lauria, M. (1985). Towards an analysis of the role of gay communities in the urban renaissance. *Urban Geography* 6, 152 169.

Morgan, R. and Wieranga, S. (2005). *Tommy Boys, Lesbian Men and Ancestral Wives: Female Same Sex Practices in Africa*. Johannesburg: Jacana Media.

Nash, C. J. (2001). Siting lesbians: Sexuality, space and social organization. In Goldie, T. (ed.) *In a Queer Country: Gay and Lesbian Studies in the Canadian Context*. Arsenal Press: Vancouver.

Nash, C. J. and Bain, A. (2007). Reclaiming raunch: Spatializing queer identities at Toronto women's bathhouse events. *Social and Cultural Geography* 8(1), 16 42.

Podmore, J. (2001). Lesbians in the crowd: Gender, sexuality, and visibility along Montreal's Boulevard St. Laurent. *Gender, Place, and Culture* 24, 191 217.

Rothenburg, T. (1995). And she told two friends: Lesbians creating urban social space. In Bell, D. & Valentine, G. (eds.) *Mapping Desire: Geographies of Sexuality*, pp 165 181. London: Routledge.

Tattelman, I. (2000). Presenting a queer (bath)house. In Boone, J. A. (ed.) *Queer Frontiers: Millenial Geographies, Genders, and Generations*, pp 222 257. Madison, WI: University of Wisconsin Press.

Valentine, G. (1993). (Hetero)sexing space: Lesbians perceptions and experiences of everyday places. *Environment and Planning, D: Space and Society* 11, 395 413.

Valentine, G. (1993). Negotiating and managing multiple sexual identities: Lesbian time space. *Transactions* 18, 237 248.

Ley, D.

P. Jackson, University of Sheffield, Sheffield, UK

Ley, David (1947–)

David Ley was born in Swansea and educated at Windsor Grammar School before going to Oxford University to study geography at Jesus College, graduating in 1968 (**Figure 1**). He did his PhD with Peter Gould at Penn State University in the US and now teaches geography at the University of British Columbia in Vancouver. He has reflected on the way these people and places have shaped his career in a fascinating autobiographical essay in 2006.

As an urban geographer who focuses on the social and cultural dimensions of city life, Ley's work has been widely acknowledged for its theoretical and empirical contribution. Empirically, his work has dealt with issues of inner city housing and gentrification, immigration and ethnicity, with a particular interest in Canadian cities. More widely, he has contributed to geography's growing interest in social theory and to the discipline's recent 'cultural turn' with a particular interest in cultural representation of landscape and place. In many ways, his work has mirrored – and in some cases prefigured – changes within the recent intellectual history of human geography, from behavioral to humanistic approaches through Marxism and radical geography to questions of cultural representation and postmodernism.

Ley's work demonstrates a strong commitment to social justice, a blend of quantitative and qualitative research, and a belief that theory should be empirically accountable. All of these characteristics were evident in his PhD research on the social geography of inner city Philadelphia. The study was informed by current behavioral approaches to environmental cognition and mental mapping – including some fascinating maps of gang territoriality, urban stress surfaces, and spatial interactions within and between neighborhoods. It included survey data and analyses of secondary sources. However, the most original and compelling aspects of his research were rooted in ethnographic traditions of first hand fieldwork where his work was truly pioneering. His analysis of violent incidents between rival gangs demonstrated the close relationship between proximity and aggression but also gave a human face to these abstract spatial relationships. As part of this project, Ley mapped the distribution of urban graffiti as a marker of gangland territoriality. He recorded the exploits of rival gang members, where the more brazen the behavior the greater the return in status, and he studied the inner workings of a local community association. Published as *The Black Inner City as Frontier Outpost: Images and Behavior of a Philadelphia Neighborhood* (1974), it was one of the first studies in human geography to employ this kind of ethnographic approach, addressing serious social issues such as racism, gang violence, and urban regeneration through intensive fieldwork and a commitment to understanding people's subjective experience of place. The study came to be regarded as a 'classic' in human geography, celebrated as a turning point in the development of urban social geography, written with the authority of direct experience.

David Ley was appointed to the University of British Columbia (UBC) in 1972 where he began a collaboration with Marwyn Samuels which led to the publication of an influential collection of essays on humanistic geography (1976). The book offered an alternative to the then current emphasis on spatial analysis which, Ley argued, threatened to impoverish the subject, reducing it to 'a geography without man'. Drawing on a range of theoretical and philosophical sources, Ley proposed an approach to human geography that was rooted in three principles: anthropocentrism (where every object is an object 'for a subject'), intersubjectivity (the sharing of meaning contexts within

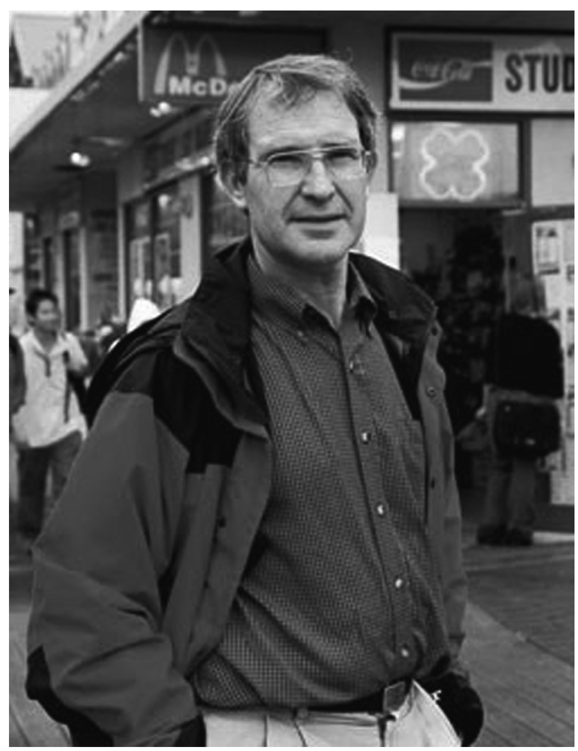

Figure 1 David Ley.

social groups), and the idea that social constraints impinge on individual human agency. It was at this juncture that Ley began to express his opposition to structural Marxism, bringing him into conflict with leading figures in radical geography such as David Harvey – see, in particular, his controversial paper with James Duncan.

Ley's commitment to a theoretically informed and ethnographically grounded human geography resulted in the publication of a ground breaking textbook on urban geography (*A Social Geography of the City*, 1983). This commitment was also apparent in Ley's evolving research agenda which moved on from issues of racialized urban inequality to new interests in inner city gentrification and the emergence of the 'new middle class'. This work brought him into heated debate with Neil Smith and others over whether the determinants of gentrification were primarily supply or demand led. Ley argued that Smith's supply led explanation, based on the identification of a 'rent gap' in urban housing markets, was insufficient to explain the processes observed in a range of cities across North America. Ley's work sought to balance supply side explanations with an understanding of housing demand from those with a postmodern taste for 'sensitive urban placemaking'. Ley's work in this field has been characterized by detailed empirical analysis of census data, supplemented by more qualitative studies of urban lifestyle. Indeed, unlike many of his contemporaries, Ley writes that he has never been persuaded that 'interpretation' and 'measurement' are in some manner incompatible.

Having played a pioneering role in geography's break with positivism and behavioral geography, Ley was also at the forefront of geography's so called 'cultural turn', exploring questions of discourse, representation, and cultural politics. He wrote a powerful editorial on culture's geographies with Derek Gregory and demonstrated the potential of this approach to the study of landscape spectacle and urban consumption through an analysis of world's fairs with Kris Olds.

Ley has subsequently played a prominent role in the 'Metropolis' project, serving as the director of UBC's Centre of Excellence for Research on Immigration and Integration in the Metropolis (1996–2003). The project demonstrated the value of academic research on a highly charged and politically sensitive topic, focusing in particular on Chinese immigration to Vancouver and on the economic, political, and cultural conflicts associated with this process. This work has also led to more general commentaries on the nature of transnational space, criticizing the 'global city' thesis for its economistic tendencies, the suppression of political and cultural domains, and the underdevelopment of human agency and everyday life.

David Ley's work is characterized by a strong sense of morality. It is theoretically informed and based on exacting standards of empirical scholarship. In recognition of his contributions to geography and social science, he has received numerous awards and distinctions. He has been a member of the Royal Society of Canada since 1998 and currently holds a Canada Research Chair in Geography, focusing on issues of urbanization and cultural diversity.

See also: Cultural Geography; Gentrification; Humanism/Humanistic Geography; Segregation, Urban; Territory and Territoriality.

Further Reading

Duncan, J. S. and Ley, D. (eds.) (1993). *Place/Culture/Representation*. London: Routledge.
Gregory, D. and Ley, D. (1988). Culture's geographies. *Environment and Planning D: Society and Space* 6, 115 116.
Jackson, P., Palm, R. and Ley, D. (1998). Classics in human geography re visited. *Progress in Human Geography* 22, 75 80.
Ley, D. (1974). *Association of American Geographers, Monograph No.7: The Black Inner City as Frontier Outpost: Images and Behavior of a Philadelphia Neighborhood*. Washington, DC: Association of American Geographers.
Ley, D. (1980). *School of Geography, Research Paper No. 24: Geography Without Man: A Humanistic Critique*. Oxford: School of Geography.
Ley, D. (1983). *A Social Geography of the City*. New York: Harper and Row.
Ley, D. (1987). Styles of the times: Liberal and neo conservative landscapes of inner Vancouver 1968 1996. *Journal of Historical Geography* 13, 40 56.
Ley, D. (1988). Interpretive social research in the inner city. In Eyles, J. (ed.) *Research in Human Geography*, pp 121 138. Oxford: Blackwell.
Ley, D. (1995). Between Europe and Asia: The case of the missing sequoias. *Ecumene* 2, 187 210.
Ley, D. (1996). *The New Middle Class and the Remaking of the Central City*. Oxford: Oxford University Press.
Ley, D. (2004). Transnational spaces and everyday lives. *Transactions, Institute of British Geographers* 29, 151 164.
Ley, D. (2006). Places and context. In Aitken, S. & Valentine, G. (eds.) *Approaches to Human Geography*, pp 178 183. London: Sage.
Ley, D. and Duncan, J. S. (1982). Structural Marxism and human geography: A critical assessment. *Annals, Association of American Geographers* 72, 30 59.
Ley, D. and Olds, K. (1988). Landscape as spectacle: Worlds fairs and the culture of heroic consumption. *Environment and Planning D: Society and Space* 6, 191 212.
Ley, D. and Samuels, M. S. (eds.) (1976). *Humanistic Geography: Problems and Prospects*. London: Croom Helm.
Ley, D. and Smith, H. (2000). Relations between deprivation and immigrant groups in large Canadian cities. *Urban Studies* 37, 37 62.
Ley, D. and Tuchener, J. (2001). Immigration, globalisation and house prices in Canada's gateway cities. *Housing Studies* 16, 199 223.

Relevant Websites

http://www.geog.ubc.ca/~dley/
(David Ley's Homepage)
http://www.riim.metropolis.net/
Metropolis Enhancing Policy Through Research.

Liberalism

E. Gilbert, University of Toronto, Toronto, ON, Canada

Glossary

Biopower A form of regulation exerted by the state or by institutions which attends to individual bodies and lives in ways that are designed to enable the protection of the population and the maximization of their quality of life.

Biopolitics Governance that is exerted through biopower, that is, through techniques of control targeted to individual lives and the population.

Citizenship The formal relationship between citizen and the state that in its ideal form denotes political membership comprising rights, responsibilities, and a sense of community belonging.

Democracy A form of government that is presumed to reflect the will of the people by way of popular representation, in which all citizens are purported to have equal rights and equal access to participation.

Freedom In liberalism, an ideal state of being whereby the individual, state, or market is able to act unhindered by constraints of any other individual, state, or market except perhaps when the safety of another is threatened.

Governmentality The practices through which subjects are governed, by themselves and by others, and which include techniques of power and expertise, such as accounting and mentalities or rationalities of rule.

Liberalization The loosening of regulations, often those enacted by the state, with respect to the governance of the market and social welfare.

Neoliberalism A form of governance that resurrects aspects of classical liberalism to affirm that individual liberty is achieved through expanded market competitiveness, a weakened state, and the 'responsibilization' of individuals and families for social welfare.

Post-Structuralism A loose term that captures a wide range of critical and theoretical approaches that have sought to destabilize totalizing narratives of truth, power, reason, and structures of meaning.

Introduction

The impact of liberalism on geography has been pervasive and profound. It is hardly an exaggeration that no area of Western, Anglo human geography has been left untouched, although from the late twentieth century much critical work has sought to undermine liberalism's key tenets. In part, the hegemony can be attributed to the scope and breadth of the concept, which refers at once to a philosophy, a political doctrine, and a political tradition, and includes theories of human nature, governance, and economy. To simplify greatly, liberalism champions individual 'liberty', for an individual conceived of as rational, egalitarian, and perfectible. How to limit the imposition of authority, particularly in the guise of state rule, is a key concern. So too is the role of the market, so much so that the legacy of liberalism today rests on the twin principles of individual and market freedoms, and how best to reconcile them when they diverge. Alongside these fundamental principles are powerful and programmatic ideals such as equality, freedom of expression, tolerance, universality, human rights, self government, liberal democracy, citizenship, public and private spheres, a free press, competition, private property, private enterprise, free trade, open markets, and the rule of law.

There is no definitive coherence to liberalism, however. It is complicated, confusing, and often contradictory. Liberalism has varied greatly over time and place, in ways that are often theoretically or practically consistent or reconcilable. It is also continually being reinvented and reimagined, often in light of that which has come before, lurching between elitism and radicalism. Thus, it is useful to follow Michel Foucault's approach which, rather than trying to define a singular concept, asks how liberalism aspires to govern. To what kind of 'liberty' do liberals aspire? For whom? From whom? Even with these questions in mind, liberalism's full complexities cannot be captured here. Three key concepts are thus considered – the individual, governance, and the market – and the ways that they have been used to secure and mobilize the governing of the population, as well as how they have been resisted. These concepts overlap, and are not easily separable, but doing so here provides an overview that is not simply historically linear or descriptive, but which gestures toward the variety of liberalisms across time and place even as its origins in the Anglo West are stressed. The importance of liberalism to geography and geographers is emphasized.

The Liberal Individual

If individual liberty is the foundational concept to liberalism, this begs the question: which individual? what liberty? These are the issues that this section will address; first, by examining four foundational principles of the liberal self – individualism, rationalism, egalitarianism, and ameliorism – before turning to the critiques that geographers and others have raised, drawing upon theories such as post structuralism, psychoanalysis, Marxism, feminism, and postcolonialism. Liberals have tended to presume that individuals are born free, and possess natural, inalienable, and usually God given rights. Humans thus have a prior ontology and epistemology: people are fully formed, self reliant, and capable of knowing and of governing themselves. John Locke, perhaps the most influential early contributor to liberal philosophy, helped shape this modern view. His *Treatises on Government*, written at the end of the seventeenth century, presents the individual as free and reflexive, in a state of nature much more harmonious than Hobbes's characterization of life as "nasty, brutish and short." Jean Jacques Rousseau also emphasizes human goodness and freedom; he worries, however, in *The Social Contract* (1762) that although "men are born free [...] everywhere they are in chains." Rousseau and Locke diverge with respect to how those chains could be loosened, but they both place the sovereign individual at center stage of their worldviews. John Stuart Mill also insists on an *a priori* principle of free individuals in *On Liberty*, which shaped his antipathy to authority and government.

Secondly, for liberals, reason provides both the basis for which the claim to individual liberty can be made, and the mechanism through which liberty can be realized and governance of the self justified. Even into the late twentieth century, reason prevails. In John Rawls's *A Theory of Justice* he posits an original position – a reworking of the mythical state of nature wherein humans are free, rational, and good – to provide a normative theory of the just society and its political institutions. In this ideal original position, individuals assume a rational impartiality because their self interests are surmounted, which leads to a consensus around social governance. Similarly, Jürgen Habermas' model of communicative action suggests that a just society will emerge out of the dialog of rational individuals in the public sphere, where individuals transcend all traces of their individuality. In the eighteenth century, Immanuel Kant had argued that reason was the best means to transcend one's passions. For Rawls and Habermas, however, it is not just individual liberty at stake, but that of the body politic.

Thirdly, the model of the rational individual infers a disembodied universalism that as we shall see has been extensively critiqued. It also, however, has provided the basis for an ethics (if not practice) of egalitarianism, used to expand the concept of natural rights across the population. Questions persist as to how to deal with pluralism. Wendy Brown outlines the ways that the concept of 'toleration' has been mobilized to this end, from the Reformation onward, and in figures as diverse as Spinoza and Milton. Locke's *Letter Concerning Toleration* appeals for the acceptance of religious difference, albeit not including Roman Catholics. His argument rests on a crucial differentiation between civil and political society: religion was deemed an individual and private matter, and hence tolerable because separate and distinct from the public sphere of common truth. Thomas Paine's agitation against slavery rests on similar principles of equal rights and tolerance. In *A Vindication of the Rights of Woman*, Mary Wollstonecraft argues that woman's inability to reason or to achieve liberty was because they were not given the opportunity to do so, not because of an innate incapacity. The distinction between public and private was retained, however, as she affirmed that woman's natural role was as wife and mother.

Underlying the arguments for women's rights and the abolition of slavery is a fourth principle, that of human progress and individual ameliorism. Nineteenth century utilitarians such as Mill and Jeremy Bentham enthused over the role of science and education in the perfectibility of the self and of society. Rousseau's semifictional *Émile* describes the best education for young men and women to fulfill their social roles, albeit again resting on a gendered differentiation of public and private spheres. Science and education were thus held out as progressive, even as they were used to reinforce discriminatory practices. There is also an implicit sense that the individual is self knowing and self knowable, and open to transformation. Into the twentieth century, the ethos of perfectibility has morphed into an emphasis on self improvement, one that entails a responsibilization of the individual (and the family) with a limited role for society or the state. The model of the autonomous, rational individual – ostensibly equal but needing improvement – continues to predominate, although as we will see below, in a more nefarious incarnation in the neoliberal *homo economicus*.

The liberal self has thus come to be associated with individualism, rationalism, egalitarianism, and ameliorism. Anglo geography has largely cohered with this model, for example, the universal, 'rational man' of positivism, and even the individualism celebrated in humanism. In the last several decades, however, attempts have been made to dislodge this construction of the self, and to problematize the appeals to liberty upon which liberalism is founded. The role of post structural theory, which gained ground in the discipline in the 1980s, has been central. The techniques of deconstruction associated with Jacques Derrida were wielded to reveal a narrative's inherent contradictions by uncovering how meanings are produced through binary oppositions, and the gaps and supplements to the text. This deferral of meaning and play of signifiers unsettles liberal presumptions of logic, truth, and certainty and the possibility of a rational, self knowing self. Individual autonomy is further challenged in that identity is conceived as relational, constituted in reference to an 'other'. The self is hence neither natural nor free; nor is there a single truth or reason through which liberty can be realized. Rather, post structuralists and postmodernists such as Jean Baudrillard and Jean François Lyotard celebrate disorder and disruption as they herald the death of the Enlightenment metannaratives of order, truth, reason, and logic that underlie liberal subjectivity.

The challenge to singular truth resonates with the much earlier writings of Friedrich Nietzsche who describes truth as "a mobile army of metaphors, metonyms and anthropomorphisms." In works such as *Beyond Good and Evil*, his attack extends still further to include self consciousness and free will; he remarks that the subject is but a 'grammatical fiction'. Drawing upon Nietzsche, Foucault's archeological studies displace the centrality of the human subject by turning to structural histories of medicine and the social sciences. In *The Order of Things*, he reflects upon the very recent history of the autonomous subject, and draws attention to the discursive contingency of the subject as an effect of power and discipline. These tactics are perhaps most clearly evident in institutions, whether for the 'insane', as Hester Parr and Chris Philo have examined, or the poor house described in Felix Driver's work. Spaces such as Chinatown work in still different ways, as Kay Anderson has revealed, to constitute marginal and dominant subject positions through racialized and gendered discourses.

Psychoanalytic theory has also unraveled the unity of the rational individual. Sigmund Freud's theory of the tripartite structure of the psyche into ego, superego, and id, undermines the certainty of a transparent self. Desires are repressed, denial and displacement are endemic, and dreams are laden with symbolism. Freud and his former colleague Jacques Lacan both suggest that identity formation is relational and conflictual, whether produced through the Oedipus complex, or during the 'mirror stage'. But Lacan takes issue with Freud's talking cure in that it extends the possibility of the restoration of a coherent self. He argues instead that the 'unconscious is structured like a language', an assertion that speaks to the complexity of the subject who cannot possibly be transparent to himself, but troubled with symbolism. Only a few geographers such as Liz Bondi, Heidi Nast, Steve Pile, and Gillian Rose have explicitly engaged with psychoanalytic theory to problematize the spatial constitution of bodies, sexuality, desire, and trauma.

If post structural and psychoanalytic theories decenter the autonomous and transparent liberal self, Marxist analysis redirects attention to the social constitution of the individual, enmeshed in the class based structures of capitalism. For Marx, liberal appeals to egalitarianism are impossible to realize in a private property system where ownership conveys power and influence. This invective informs David Harvey's seminal *Social Justice and the City* which argues for moving away from 'Liberal formulations' – the title of the first section of his book – to a revolutionary Marxist critique that will address social injustice by abolishing private property. The book thus positioned a new critical human geography in opposition to liberalism. Marxism, however, paid little explicit attention to the constitution of the subject and subjectivity. It could be inferred from the discussions of class, which shaped the intimate relations between the bourgeoisie and working class, that individuals are always social and political beings, situated at a particular historical moment – and not *a priori*, autonomous beings. But only with the writings of Marxist structuralist Louis Althusser does a more compelling critique of the self emerge. Drawing upon Freud and Lacan, he argues that the self is an ideological construct. His article 'Ideology and ideological state apparatuses', asserts that the self and citizen are made meaningful through their interpolation in social structures. The individual is thus produced through, constrained by, and even managed through such institutions.

Understanding the social constitution of the self was also the object of Anglo feminist geography of the 1970s, although much of the early work was resolutely liberal. The emphasis inside and outside the academy was on the extension of rights to women and the affirmation of women's place in the academy both as researcher and as research subject. Betty Friedan's *The Feminine Mystique* (1963) galvanized popular interest in the suburban anomie experienced by full time homemakers in the post World War II suburbs. Feminist geographers began to provide a more politicized critique of the public–private divide, often by drawing upon Marxism, although its relationship with feminism has often been uneasy. Suzanne McKenzie and Damaris Rose, among others, identified the centrality of the home and unpaid domestic labor to the material and ideological reproduction of capitalism. Addressing the public private divide that is inherent to liberalism – and the power politics of this differentiation – has been a key facet of feminism, as captured in the popular slogan 'the personal is political'. A more embodied feminist research has also emerged, attuned to subjectivity and positionality. Bodies themselves have become a focus of study as geographers work to unsettle conventional dualities between male/female and mind/body, as with Rachel Pain's exploration of women's fear of violence in public spaces. Isabel Dyck and Pamela Moss have also challenged the ableist presumptions of much social science research that disregards disability and illness. Research on the body has also informed studies of sexuality. David Bell, Jon Binnie, Glen Elder, Larry Knopp, and Gill Valentine have all examined gay and lesbian spatialities, and the impact of the visibility of bodies in the public sphere in an era when gay and lesbian activists were taking to the streets.

Research on the body has thus drawn attention to a wide range of subjectivities that are obscured by a liberal universalism that presumes a rational, disembodied self. Studies of race have made similar interventions, while activist scholars such as bell hooks and Audre Lorde have assailed feminist approaches that idealize the concept of 'woman' and ignore class and racial privilege. In contrast to Friedan's 'feminine mystique', hooks' moving piece on home as a site of resistance unpacked the suburbs as both the site of heterosexual, normative family and the dependence of this kind of family structure on black labor. Geographers Alastair Bonnett, Ruth Gilmore, Peter Jackson, and Laura Paulido have examined the ways that race is produced and reproduced in social and legal spaces such as schools and prisons. They challenge the liberal narratives of progress and ameliorism that govern these institutions and their imposition of dominant forms of subjectivity on the marginalized. The spatiality of racism needs to be interrogated, Linda Peake and Audrey Kobayashi argue, not only to reveal the hypocrisy of liberal universalism and its underlying racist practices, but also to illustrate that race is a human construction, with no biological justification, produced and reproduced through space.

The collection, *This Bridge Called My Back*, edited by Cherrie Moraga and Gloria Anzaldúa, records the tensions between liberal aspirations to pluralism and egalitarianism and the structural inequities that persist in the United States arising from its history of slave labor and dispossession from the land. The contributors demand a radical restructuring that can address this racism, not an integrationist liberalism that subsumes cultural differences, depoliticizes private cultural identities, obviates politics, and tries to remake the 'other' in one's own image. No less has been demanded by anticolonial and postcolonial writers. In the 1950s, legendary revolutionary and registered psychiatrist Franz Fanon railed against the internalization of the colonial subjugation by the colonized. From another perspective, Edward Said examines how Eurocentric imaginative geographies of the East as 'weak' and 'effeminate' were used to rationalize colonial and imperial expansion. His book *Orientalism* examines how discourses of the 'other' constitute the colonizing subject, whereas Fanon had suggested that colonial subjectivity was negated. More recently, postcolonial theorists such as Homi Bhabha, Robert Young, Ann Laura Stoler, and Anne McClintock have drawn upon post structuralism and psychoanalytic theory to frame the relations between colonizer and colonized in a more nuanced way that eschews binary oppositions, and explores the complexities of desire, eroticization, transgression, and hybrid subjectivity. But the question that Gayatri Chakravorty Spivak poignantly raises, as to whether the subaltern can find a voice given her interpolation in colonial and patriarchal discourse, is left hanging. The egalitarianism to which liberalism aspires never seems to be realizable for the marginalized.

Identity politics associated with studies of race and gender has raised important issues regarding the exclusionary practices of liberalism. At times, however, there has been a tendency to reinstate a liberal individualism, and the values of the perfectible and transparent self. Queer theory, by contrast, has sought to unsettle these narratives, while also raising opportunities for resistance. The attention to the social construction of the categories of gender, sex, and sexuality have challenged essentialized notions of identity and the identity politics that have emerged around fixed concepts of class, gender, and race. Judith Butler posits that the self is in a process of becoming, whose undoing is always possible. Butler's analysis of performativity reveals how gender is assembled through routine: subject positions are created, not just imposed, albeit enacted in a context of external constraints. She identifies the possibility of transgression, though humor and irony, for example, or resistant performances such as drag. Identities are thus fluid and in flux, not fixed, autonomous and self knowing, as conceived by liberalism. For some, the possibility that identities are complex, dynamic, and hybrid is cathartic because of the potential for reconstitution and reconfiguration that this holds out. For others, however, this characterization is too playful, too transient, and too partial to tackle material injustices. Along different lines, Katharyne Mitchell cautions against the 'hype of hybridity' because the aspiration to transformation and subversion can and has been co opted in the interests of capitalist accumulation.

Nonetheless, the trope of hybridity persists. In another strand of research, geographers have engaged with the concept to raise moral and ethical questions about the relation between the human and the nonhuman and to encourage a dissolution of those binaries. Donna Haraway's model of the cyborg, a technological hybrid of human and machine, works to complicate the distinctions between human and nonhuman, animate and inanimate, and nature and culture. Actor network theory (ANT) and scholars such as Michael Callon, Bruno Latour, and John Law draw from the philosophy of science, economy, and sociology to argue that agency (and power) are constituted in and through networks held together by immutable mobiles that help to make the networks durable. In this formulation, agency, albeit muted, is a capacity of both the animate and the inanimate. Sarah Whatmore has drawn upon ANT, Haraway, and the writings of Gilles Deleuze and Felix Guattari to push toward a reconceptualization of the 'autonomous self' that acknowledges a world woven together by human and nonhuman elements. Her work on 'becoming elephant', for example, addresses the performativity of the elephant subject in the recursive processes negotiated between actor and network. This attention to the 'more than human world' speaks to a broader contemporary anxiety around the constitution of the human subject.

The new work on hybrid natures unravels the binary opposition implicit to the concept of a liberal self which presumes that agency, rationality, egalitarianism, and ameliorism are uniquely human capacities. Identity politics, by contrast, has criticized the liberal individual because it is modeled on a white, masculine, and disembodied self who is idealized as rational, self interested, self maximizing, and self knowing. In some instances, critique has been used to expand liberal categories to make them more inclusive of marginalized populations, in effect, using liberalism's aspirations to universalism and egalitarianism against itself. Others, however, argue that paradox and prejudice are intrinsic to liberalism, so that it must be overthrown. Geographers have engaged with these debates from all sides, and they continue to challenge foundational liberal constructions such as the public–private divide, and the naturalization of gender and race. More cross cultural research could identify alternative epistemologies, for example, an aboriginal worldview or Chinese Buddhism, which each in quite different ways offer a more holistic and less individualized epistemology. Such comparisons might bring even closer into focus the liberal paradox whereby freedom from domination is emphasized at the very same time that material and discursive stratification is being reinforced. Liberalism's tenacity, however, speaks to the popular appeal of its claim to individualism, and the concepts of liberty that it promotes.

Governing through Liberty

If individuals are presumed to be free, rational, and perfectible, then how can society be ordered to maximize these qualities? Liberalism tends to argue that self rule is the best mechanism for ensuring the autonomy and freedom of the individual, and is a due right of rational beings. At the same time, classical liberalism suggests that maximizing individual freedom is also in the best interests of all, which requires a social ordering that is both individualizing and totalizing. A persistent question has been how to balance individual rights and social relations.

Political liberalism is usually traced back to seventeenth century England when social hierarchies were disrupted by the Glorious Revolution of 1640–88 which loosened the grip of religious authority, and bolstered popular sovereignty and parliamentary democracy. The aristocracy of the Whig party, heirs to liberal values after the English revolution, supported government through choice and consent – alongside personal and economic freedoms, and even religious toleration – although only for the propertied classes. The resonance with Locke's social contract is clear. Individuals agree to their political rule because although there is equality in the state of nature, social conflicts over natural rights, for example, property, need to be mediated. A state is thus a social necessity, although it needs to be limited, as outlined in Wilhelm von Humboldt's programmatic volume, *On the Limits to State Action*. John Stuart Mill suggested that the only legitimate justification for state constraint was to prevent harm, to the individual or to others.

Liberal thought thus arose alongside a breakdown in traditional authority and struggles to place power in the hands of the (propertied) people. It bore both progressive and conservative proclivities. What this has meant with respect to the role of the state has been of key interest to political philosophers. Writing in 1958, Isaiah Berlin differentiated between 'negative' freedom (freedom from coercion) and 'positive' freedom (freedom to self realize). Although it has since been argued that these freedoms coexist, liberals tend to emphasize one or the other. As noted above, classical thinkers such as Locke and Mill argue from the vantage of negative freedom: individual liberty is construed as freedom from state intervention, with the state playing only a minor role in the economy, only a little more of a role in matters of security and civil liberties, and almost no role in the private sphere. As we will see in more detail below, these ideas have been revived by neoclassical liberals such as Friedrich Hayek and Milton Friedman who argue that less state and more individual freedom is in the best interests of all. By contrast, the predominant liberalism of the early twentieth century, with a little nod to Rousseau, has emphasized positive freedoms. T. H. Green, L. T. Hobhouse, and John Rawls promote the state provision of social rights and services as a means to ensure that each member of the population has the same opportunities. Under welfare liberalism, resources are redistributed and services such as education and healthcare are provided to encourage a minimum standard of living. The social benefits and social security of this model are touted as egalitarian, even as redistribution helps to make capitalist inequities more palatable and hence more durable, to the advantage of the propertied class.

In both positive and negative models, liberalism is territorialized onto the nation state. Locke's writings, for example, and those of other contemporary liberal thinkers such as Thomas Paine, were formative to the push for national independence in the eighteenth century. The American Declaration of Independence (1776) and the French Declaration of the Rights of Man and of the Citizen (1789) were grounded in the natural rights of the individual and variants on Locke's life, liberty, and property, even as they took on different dimensions. US liberals trumpeted the importance of individual rights over class, whereas in Britain economic liberties were promoted. In France, the 1789 revolution's attack on the monarchy, aristocracy, and the Roman Catholic Church led to an emphasis on secularism and democracy. Liberalism also infused other national movements in Europe, enshrined, for example, in Spain's first constitution of 1812. In the Americas, Haiti declared its independence from France in 1791, drawing upon the aspirations to liberty of the French revolutions, just as the Spanish example would influence the South American wars of independence led by figures such as Símon Bolívar and José de san Martín.

Liberalism has thus been pivotal to movements for national independence. In turn, the nation state has become 'the' political container within which liberalism operates and within which individual rights and freedoms have been made achievable and understandable, even as it exhibits an antipathy to state rule. This is one of the key paradoxes between liberal ideals and forms of government. By and large, geographers have tended to reinforce the assumption that nation states are, as Neil Brenner has put it, the locus of power and economic management, although this has been challenged by globalization and neoliberalism. Some recent works, however, interrogate this assumption. Matt Sparke's *In the Space of Theory*, for example, draws upon antifoundational critiques of the production of truth and knowledge to challenge the mapping of social theory onto the nation state. As Clive Barnett and Murray Low remark, however, only in limited ways have geographers examined the connections between liberalism and democracy or democratization. Their collection, *Spaces of Democracy*, begins to attend to this lacunae, at the same time that it examines other scales of governance.

Foucault's work has been crucial to rethinking the state, governance, and power. He has instigated a strain of research on 'governmentality', that is 'the conduct of conduct', or how and why people govern themselves and others. Foucault identified in the liberal state a shift away from an emphasis on territorial sovereignty, to an interest in biopower, in the population and its governance. How to regulate life, rather than how to threaten death, is the object of the modern liberal state. Matt Hannah examines how the population is made knowable and calculable through discourses and techniques, for example, census and statistics, in late nineteenth century United States. As his work demonstrates, it is not that concerns for the population completely supplant sovereignty, but that governmental rule is reoriented toward ensuring the health, wealth, and happiness of the population. The state strives for social security, but it is also through this security that the state governs. Foucault also argues, however, that this rationality is not only disseminated by the state, but becomes internalized within the population. Individual and social institutions – the family, the workplace, nongovernmental organizations, etc. – all strive to maximize their quality of life, and assume responsibility for optimizing their own potential. Freedom is both a condition of and the outcome of this responsibilization.

Foucault identifies a two pronged strategy through which liberal technologies of power are deployed. The first, as noted above, entails the protection of the population from risks associated with the industrial economy (e.g., health programs, education reform, etc.), so that they can optimize their quality of life. The second is a more nefarious wielding of 'racial hygiene' whereby the population is assessed and ruled according to its perceived capacity to achieve responsible autonomy. Disciplinary power is applied both to reform those who do not meet the norms, and to excise those who are considered to be irredeemable, for example, in institutions such as asylums or prisons. The propagation of those elements of the population deemed most at risk, and hence risky to others, is hence limited. As the term 'racial hygiene' denotes, there is an inherently racial dimension to this social ordering, as racial differences are identified and mapped onto the population through techniques such as racial profiling and ghettoization.

Foucault's analytic framework identifies the illiberalism upon which the liberal state's rule of its population is founded. Other scholars have turned to examine the transnational dimensions of 'racial hygiene'. With echoes of Said's 'Orientalism', David Theo Goldberg illustrates that liberal governance manages its own population by identifying 'uncivilizable' populations beyond the state, largely codified through race. Geographers have been complicit in this process; at the beginning of the twentieth century, for example, environmental determinist Ellen Churchill Semple classified populations with respect to their land use and propensity to violence, naturalizing social relations in terms of geographical distribution. Liberals have also been, as Uday Singh Mehta documents, among the most ardent advocates of empire. British colonial India was a favorite laboratory. John Stuart Mill, who worked for the British East India Company for 35 years, approved of colonial rule because, he argued, Indian 'barbarians', were incapable of governing themselves. Thus, as Mariana Valverde has remarked, while Mill was progressive with regards to the expansion of rights at home, for example, to women, he had no compunction in rationalizing despotic rule for populations elsewhere. In settler societies, the consequences were often violent. Indigenous peoples were deemed 'uncivilizable' and were corralled onto reserves, or removed from their families and placed in residential schools as in Canada and Australia. These strategies were part of the bifurcated attempts to both discipline and train other populations through spatial segregation.

Geographers have been sensitive not only to this wider colonial history, but also to the complicity of the discipline. Felix Driver and David Livingstone have provided broad overviews of this participation, while others have addressed geographic tactics of rule and dispossession through maps, spatial segregation, and urban planning. The Royal Geographical Society (RGS) in Britain sponsored a number of expeditions by figures such as Charles Darwin, David Livingstone, Richard Francis Burton, Ernest Shackleton, and Sir Edmund Hillary. The explorers often had administrative impact. Livingstone, for example, opened up central Africa to missionaries who brought with them a form of biopoli tical care and control with their plans for health, edu cation, and work. New forms of geoeconomic empire reared their head when the RGS merged, after some controversy, with the academic based Institute of British Geographers; Shell's sponsorship of the new institution was contentious, largely because of its oil capitalism in Africa.

In the early twentieth century, geographers played a significant role in international affairs. The heartland theory of Halford Mackinder, which he submitted to the RGS in 1904, proposed a reworked geopolitics that placed Europe at the center of a nascent global worldview. Control of Eastern Europe, he argued, was necessary to world domination, which informed international and US containment policies. Almost his contemporary, geog rapher Isaiah Bowman, who directed the American Geo graphical Society for 20 years, played a parallel role in the US where he acted as presidential advisor first to Wood row Wilson and then to Franklin D. Roosevelt. In *American Empire*, Neil Smith uses the figure of Bowman to critique US aspirations to hegemonic globalism and the victory of liberalism over geography. This was the dawn of a new empire, being built on a global assertion of economic and moral liberalism. Although the new, internationalist drive appeared to triumph over space, it required the careful management of global geography. Area studies programs of the post World War II era, with their direct links to the military and foreign analysis, thrived in the impetus for knowledge of the world.

These academic policies and programs reinscribed the polarity between the domestic and the foreign, all the while enclosing the world within a liberal integrationist order. Domestically, Western states introduced a varying degree of redistributive welfare programs. Social security enabled the positive freedoms celebrated in the writings of T. H. Green and L. T. Hobhouse, as liberalism re invented itself as a superior moral force within the new geopolitical order. The international infrastructure that was created, including the General Agreement of Tariffs and Trade (GATT), the United Nations (UN), the International Monetary Fund (IMF), and the World Bank, was used to promote Western liberal economic and moral values around the world. The increased mobility of the population in this same era, however, revealed underlying social and cultural tensions. Hannah Arendt drew attention to the wave of displaced and stateless peoples in the aftermath of the two world wars, and the impact on people bereft of political status as a result of ethnic cleansing. When citizenship rights are grounded in the nation state where does that leave those whom the state has refused to recognize? This is one of the para doxes of liberal citizenship that Hannah Arendt exposed. Universal values appear to transcend the nation state, yet citizenship depends upon exclusionary practices that separate those with 'legitimate' claims to the state from the rest, internally and externally.

To put it somewhat differently, as Barbara Cruickshank has done, citizenship is a technology of state through which the population is governed and deemed governable (or not). Feminist scholars such as Ruth Lister and Iris Marion Young have examined the persistent gendering of citizenship that harkens back to the public–private divide of the social contract. Despite the expansion of programs post World War II, rights were dependent on workforce participation, a proviso that meant women were largely excluded except in their status as dependents of male breadwinners, while children still continue to be con sidered only citizens in the making. Citizenship, Daiva Stasiulis and Abigail Bakan forcefully insist, needs to be understood in terms of capitalism's uneven power relations whereby the rights and responsibilities associated with Western citizenship are predicated on the labor and sub jugation of the Global South. This kind of analysis is especially important given the ongoing political contro versy in the US around undocumented workers. Geog raphers have also examined these citizenship paradoxes, with attention to mobility and borders, as with Jennifer Hyndman's analysis of refugee camps in post war Africa, and Alison Mountz's work on the smuggling of Chinese migrants into Canada.

Ongoing research has also affirmed that formal citi zenship does not necessarily entail substantive rights. Drawing upon critical race theory, Yasmeen Abu Laban insists that bias is inherent to political participation, representation, and the allocation of resources, which perpetuate marginalization and discrimination. Multi cultural 'tolerance' also reproduces and naturalizes dif ferences: cultural traditions – for example, food and dance – are celebrated in the public sphere, but political structures remain intact, regulated by a universalized liberal reason. At the same time, however, the adherence to community, the rhetoric of tolerance, and the logic of egalitarianism have been used to expand citizenship claims for minority populations. Charles Taylor and Will Kylimcka have even proposed that group rights, for ex ample, aboriginal citizenship, are possible within the liberal nation state, as a means for reconciling demands for subnational self governance. At the transnational scale, the UN Declaration of Human Rights of 1948 has projected tolerance into the international arena, even if it remains very much state based in its application.

Immanuel Kant suggested that a cosmopolitan world order, with internationalized human rights, would ensure 'perpetual peace', an ethos that infused the early post World War II period. Yet, the reality has been quite the opposite. Violence and insecurity have long been constitutive elements of liberalism, rather than what happens when liberalism fails, as we have seen above with respect to the history of colonialism. If there is a coherence across liberalism, it is its tendency toward illiberalism – and the violent imposition of liberalism – in the name of liberty. The sovereign's power to render subjects as *homo sacer*, that is to remove their political rights so that their death can be ordered with impunity, brings violence to the fore. In particular, and compellingly for geographers, Giorgio Agamben's work in this area draws attention to those spaces of 'exception' where subjects are rendered into bare life: the quintessential example is that of the Nazi concentration camp. But, he argues, it is not simply that exception is located or locatable in particular spaces, but through the exception that the law is constituted. This reinforces the idea that politics, law, and violence are not separate but continuous realms, a perspective that further undermines the liberal premise of the neutral state. Derek Gregory develops Agamben's remarks to consider Guantánamo Bay as a state of exception. Gregory's analysis, however, situates Guantánamo within an international (not just national) politics and thereby attends to the colonial histories and international jurisprudence in which it is embroiled. He argues that it is not the law that is absent, but rather the application of human rights, a point which drives home the long legacy of practices of 'racial hygiene'. In the current 'war on terror' questions around human security are pressing, and have materialized themselves particularly at borders where the power of the nation state is highly visible. But even when formal war is not decreed, rights are suspended. Gerry Pratt documents the legal abandonment of women sex workers and domestic laborers, who because of their gender are outside the rational public sphere and consequentially more easily, and more surreptitiously, rendered into bare life.

With the closing of the Cold War and the fall of the Berlin Wall in 1989, Francis Fukuyama teleologically declared *The End of History and the Last Man.* Liberal capitalism, he argued, was now ascendant as the last, best model of social ordering and human political achievement. By contrast, Michael Hart and Antonio Negri argue in *Empire* that the moment of liberal democracy has passed, that it has undergone a crisis of legitimacy because of the military overreach of the United States. Writing before 9/11, their comments on new imperial forms seem especially prescient given the imposition of parliamentary democracy in Iraq in 2005, following the American led invasion, and the continuing violence there. The specter of older forms of colonialism and imperialism looms large as Western states impose their liberal norms and ideals in a reinvented world order. The Fukuyama style rhetoric that casts liberalism as the one rational choice, as morally superior, and as the quintessence of liberty is precisely that which enables violence to be perpetuated in liberalism's name.

Liberal Economies

If a key question of liberalism is how the individual is to be ruled (or not) within a just society, so too is the question of how the economy can best be governed (or not) to maximize individual liberty. Two key liberal economic principles are private property and free trade. For Locke, property is one of the three foundational natural rights alongside life and liberty. Property is deemed an inalienable, natural right because it arises out of labor over nature. Since people are autonomous, owners of themselves, they are also owners of their own labor and its products. It is to guarantee this right that people assent to being ruled, with the polity's chief role the preservation of one's property against the claims of others. Some notion of property rights have been enshrined in state laws since the American and French constitutions, and are now more universally recognized with the UN Declaration of Human Rights.

Property holders, it is argued, possess special social status, and an implicit moral and political authority, which is used to rationalize their dispossession of the land. This logic informed the push for enclosure that accompanied the changing farming economy of Britain and later, industrialization and urbanization. Common land was appropriated for private use, which reinforced the discrepancies between the landowning and the landless, who often rioted. In the colonial context, the result was an unreserved land grab. In settler colonies, where indigenous peoples did not share this concept of property rights and tilled the soil more lightly, land was routinely dispossessed. Capitalist claims to property rights thus not only externalized nature as that to be owned, but allocated resources in the interest of the bourgeoisie, an argument developed by Neil Smith in *Uneven Development.* Nicholas Blomley has examined more explicitly the continuing violence of liberal law which, constituted as it is through boundary making (e.g., frontier, survey, and grid), reinscribes the colonial geographies out of which it originates, and the differences between colonizer and colonized, and the owners and the landless. In the contemporary context, petro capitalism plays out a new kind of imperialism that is laced with violence, as Michael Watt describes, as the oil industries and local communities jostle over the land and its resources.

Free trade is a second foundational principle in liberal economics. Eric Sheppard notes that, despite historical–geographic differences in implementation, the arguments for (and against) free trade have remained largely continuous for over 200 years. Scottish Enlightenment thinkers, David Hume and Adam Smith, resisted the mercantilism of their day and its protectionism, and argued that free trade would stimulate economic growth. The pursuit of individual self interest in a competitive market, Smith asserted, would produce efficiencies that would benefit all society. The market would naturally regulate itself by an 'invisible hand'. The only role for the state was to provide some security, legal infrastructure, and limited public works to ensure that individuals could freely pursue their own economic interests. David Ricardo also was opposed to mercantilism. He argued vehemently against the British Corn Laws, in place since 1815, which protected the price of domestic grains against foreign imports. This subsidization, he argued, prevented the country from benefiting from open competition with other markets, and maximizing its production in terms of comparative advantage.

The shift away from mercantilism and toward liberal, *laissez faire* economics underpinned the development of free market capitalism in Britain. Interest in free trade also emerged elsewhere in Europe, for example, with the writings of Jean Baptiste Say in France and Anders Chydenius in Finland, and in the United States and Canada where a reciprocity treaty was in place between 1854 and 1866. By the late nineteenth century, however, a neoclassical economics had emerged. Drawing upon the economic liberalism of Smith and Ricardo – and retaining the principles of individualism, private property, and self adjusting markets – economists sought a scientific model for supply and demand, and production and consumption that would be in the best interests of all. Theories of consumer behavior were expounded based around a rational individual whose choices, it was assumed, would maximize their utility. The paradigm was thus rooted in liberal rationalism and perfectionism, with a nod to utilitarianism.

Economic liberalism, however, was contested. Socialist thought was percolating across Europe, particularly in Britain and in France. In Glasgow, Robert Owen promoted forms of collective living, and argued that the individual was a product of society, not an autonomous figure. The Chartist movement galvanized the working class to expand the franchise and their claims on the state. In France, socialism was promoted by Comte Saint Simon, who criticized social inequities and argued against both individualism and *laissez faire* markets. Pierre Joseph Proudhon promoted anarchy, and called the ownership of property theft. Most famously, Marx criticized capitalism's internal contradictions, and the (creative) destruction wrought by capital investment and disinvestment. He argued that the state was not neutral, but rather protected bourgeois interests such as private property. His alternative vision involved shared ownership of the means of production, which ran counter to the individualist, *laissez faire*, private property thrust of the capitalist economy. Marx was also wary of free trade: he feared that the abolishment of the Corn Laws was a ruse by manufacturers to drive down food prices and hence wages.

Marxist and socialist thought informed the radical geography of the 1970s, when there was also a brief revival of the anarchism of geographers Peter Kropotkin and Élisée Réclus. New research emerged on the spatial dimensions of capitalism in the urban landscape, drawing upon the methodologies of historical materialism. This rich body of work sought to depose economic geography's cohesion with liberalism, from its idealized model of liberal 'man' as rational, productive, and self maximizing, to its affirmation of private property. Bill Bunge, David Harvey, and Neil Smith examined the urban inequities of capital accumulation, from the de investment in inner cities, to the politics of reinvestment as part of gentrification. Feminist scholars, such as Doreen Massey and Linda McDowell, addressed the social structures underpinning the gendered dynamics of urban space. Recent work also attends to the contemporary reconfiguration of the nation state under globalization (and especially free trade) and neoliberalism.

The nation state, contra classical liberal theory, played a significant role in the economy at the beginning of the twentieth century, partly as a response to the radical critiques, but also because of all the accommodations demanded by the propertied classes. The sweeping changes of World War I and the Depression, when the failures of *laissez faire* liberalism were stark, gave further impetus to nationalist protectionism. Alternative ways of conceiving individual and economic liberty around 'positive' freedoms, which had previously been relegated to the background, were now being championed. In Britain, T. H. Green and L. T. Hobhouse advocated a qualified role for the state to maximize individual potential. The provision of basic necessities, argued John Dewey in the United States, would in turn contribute to the 'public good'. John Maynard Keynes provided the economic rationale for an expanded state role, by reorienting analysis to questions of employment. He argued that state intervention at times of high unemployment could help to stimulate consumption or investment, and thus help maximize individual and social freedoms. In the post World War II era, with varying degree, social welfare was implemented in Western liberal states. In Europe, a model of state planning, social programs, and public ownership was predominant, while in the US, Roosevelt's 'new deal' ushered in a series of federal programs for 'relief, recovery, and reform'. 'New deal' liberalism was configured around social liberalism, whereas in England liberalism was distilled through the Liberal Party and Prime Minister William Gladstone, a conservatism echoed in much of Europe.

In the post World War II era, nation states thus loomed large as they carved out a certain degree of economic autonomy. But they also participated in the creation of an international political order through multilateral institutions such as the UN, the Bretton Woods agreement, GATT, the IMF, and the World Bank. Political scientist John G. Ruggie described this as 'embedded liberalism' in that a 'grand bargain' was struck between state autonomy and external regulation, with an eye to social and political security. Although World War I had challenged the liberal premise that international interdependence would preempt war, the idea that a more open trading system would bring peaceful relations was revived. GATT was part of the program for post war economic recovery, and ushered in a nascent free trade, later expanded under its transformation into the World Trade Organization (WTO). These international and multilateral post war agreements consolidated a new globalized economic order, which by the 1980s would be underpinned by a revival of classical liberalism that has taken shape as 'neo'liberalism.

By the 1970s Keynesianism was being challenged by the dissolution of Bretton Woods; the oil crisis; and the shift to more open, internationalized markets. The Chicago School economists, drawing up on the *Ordoliberals* of 1930s Germany and the works of Friedrich Hayek in particular, refuted Keynesian interventionism. They attributed the crisis to fiscal policies, national debts, and government failures. Milton Friedman's advocacy of 'negative' freedoms – a minimal role for government in the economy, a prioritization of monetary policy over fiscal policy, and a general emphasis on individual liberty – began to gain acceptance. With the elections of Margaret Thatcher in 1979 and Ronald Reagan in 1980, *laissez faire* capitalism was revived. Unions were attacked, individualist economics were promoted through tax breaks and the reduction of social programs, and in the UK, industry was denationalized. Programs such as workfare, as Jamie Peck has detailed, began to replace rights based access to income support with a stress on individual self sufficiency.

With neoliberalism, the primacy of the individual was reinstated. But, as Graham Burchell argues, neoliberalism differs from classical liberalism in that it is not about determining the kind of government that is best able to optimize the 'natural' freedom of the individual. Rather, the object of neoliberal governance is to 'produce' the free individual – an individual who not only embodies all the liberal qualities of autonomy, rationalism, ameliorism, and egalitarianism, but also personifies market principles such as self maximization, efficiency, accountability, profitability, productivity, competitiveness, and entrepreneurialism. The revamped individual at the center of neoliberalism is *homo economicus*. Moreover, market values and expectations are also projected onto social and political institutions, such as governments and schools, which themselves become quasi enterprises whose 'products' (e.g., knowledge) can be consumed by their 'consumers' (e.g., students). The cost–benefit calculations demanded of these enterprises require additional forms of regulation. As Jamie Peck and Adam Tickell have pointed out, neoliberalism entails not just a succession of rollbacks (withdrawal or disinvestment), but also a rollout of regulatory forms that need more management and supervision. The workfare policies that Peck has examined, for example, require a de regulation of labor, but in a highly regulated way that is enforced by the disciplinary and penalizing measures of the state.

The use of disciplinary and punitive measures reveals yet another duplicity. Liberals have consistently claimed that liberalized economies engender peaceful relations, certainly at the state level. John Stuart Mill argued that commerce would bring about peaceful international relations, while anti Corn law proponent Richard Cobden proposed that free trade would help ensure peace (and hence prevent war). Increasing security cooperation between states after World War I and World War II was also infused with commonplace rhetoric that liberal states were less likely to wage war on one another. In a twist to this logic, Milton Friedman argued that free markets and individual freedoms would precipitate the fall of totalitarianism. But, as Thomas Friedman has recently suggested, the 'hidden hand' of globalization and corporate capitalism is only made possible by a 'hidden fist', in this case that of the US military. Liberalism often requires the use of force and generates insecurities, even within the domestic population. This is clearly in evidence in Chile where General Augusto Pinochet's Chicago inspired market reforms of the 1970s were implemented with the firm hand of the military state and thousands were killed and tortured. More recently the reconstruction of Iraq, which has been promoted as an opportunity to rebuild infrastructure and facilitate democratization, has been shot through with violence. The Coalition Provisional Authority, working with the US Department of Defense and UK authorities, has introduced new orders of council that liberalize the Iraqi market, including privatization, tax cuts, and openness to foreign investment (e.g., by US company Halliburton), in ways that were inimical to the previous Iraqi constitution. As the increasing violence and dissent there suggest, the liberalization of the economy has been far from successful, and the lives of the population are far from secure.

Privatization has also generated other kinds of insecurities. Natural resources have been especially impacted, although as James McCarthy and Scott Prudham reflect, little attention has been paid to the impact of neoliberalism on the environment, even as attacks on Keynesian spending and national investment have undermined the environmental legislation implemented in that era to limit unheeded development. They point out that forms of life are increasingly becoming commoditized through patents, genetic engineering, biopiracy and bioprospecting, private property rights around pollution, and user fees, all of which rely upon and expand a Lockean concept of individualized private property rights upheld by the state. This commoditization is transforming human–nature relations, but also concepts of the human. A new kind of biological citizenship is emerging, Nik Rose and Carlos Novas suggest, to reconcile the deterritorialization of the state, recalibrations of human life and value, new forms of biopolitical governance, and struggles and claims to biological rights premised on the very insecurities that these transformations entail.

Free trade has similarly generated insecurities, as the expansion of export processing zones (EPZs) suggest. EPZs allow free trade under preferential terms to encourage the relocation of manufacturing: limited bureaucracy, reduction or elimination of taxes, and the removal of import or export tariffs. As Melissa Wright has documented with respect to Mexican maquiladoras, however, the impact of these highly liberalized economic zones can be devastating on the employees. To attract investment, wages are reduced, unions are prohibited, and the rights of workers denied. Women have been made especially vulnerable, their labor cheapened, their lives devalued, and their agency denied. National borders have also become sites of human insecurity, even as free trade facilitates the movement of goods and capital. In tandem with the introduction of the North American Free Trade Agreement (NAFTA), for example, was Operation Gatekeeper, a border security initiative implemented in several US states. Joseph Nevins describes the regulations, border control agents, and surveillance techniques that were all deployed against 'illegal' Mexican immigrants. At the same time, NAFTA opened borders for professionals who best met the criteria of the marketized *homo economicus*: accountability, profitability, self scrutiny, and entrepreneurialism. These hierarchical policies are being reinscribed into the North American Security and Prosperity Partnership of 2005. As I have argued elsewhere, this NAFTA plus program exacerbates social insecurities in that it encourages greater private sector participation in the deepening integration and harmonization of economic and security sectors, but allows only limited mobility rights, either through expanded temporary work programs or in preregistered security programs for professionals.

The entrepreneurial rationality of neoliberalism, and its extension into social and political spheres, may generate a more secure field for trade and privatization, but creates more insecurities and anxieties for the population. Although Wendy Larner's caveat against universalizing and totalizing neoliberalisms must be heeded, some generalizations around private property and free trade orthodoxies can be made if only because of the ways that they have been imposed internationally by the post World War II infrastructure. Richard Peet has identified the ways that, since the Latin American debt crisis of the early 1980s, the institutions of the Washington Consensus have become much more interventionist with respect to industrializing economies. The 'unholy trinity' of IMF, World Bank, and WTO has institutionalized 'poverty' and 'underdevelopment' with its tonic of 'stabilization' and 'structural adjustment' programs that have imposed currency devaluation, privatization, de regulation, and free trade. For much of the Third World this has entailed a hollowing out of the nation state as credit and investment are made conditional on privatization, de regulation, tax reform, trade liberalization, and fiscal discipline. In Africa, James Ferguson describes, the rollback of the state has exacerbated political crisis and has led to more unrest, insecurity, and violence.

Opposition to neoliberalism and globalization is gaining ground. The post Washington Consensus, spearheaded by former World Bank economist Joseph Stiglitz, has sought to 'bring the state back in' through the backdoor of good governance, transparency, and accountability. Socialist governments in Latin America have opposed the Western imperialism of the Washington consensus and US foreign policy, although governments such as President Lula's in Brazil have found it difficult to implement anti neoliberal policies. The expansion of free trade has prompted extensive protests by civil society. The Zapatista uprising of 1994, in the wake of NAFTA, has been followed by the Battle in Seattle in 1999, in opposition to WTO negotiations, and at the Quebec City Summit in 2001, where discussions around the Free Trade Area of the Americas were taking place. In each of these instances, the protests have been met with intensive securitization and violent suppression, a reminder of the ongoing twinning of economic security with human insecurity. The opposition has had some success, even if incremental. Protests against this encyclopedia's publisher Elsevier, for example, prompted the company's disinvestment from arms fairs. Elsevier had argued that the defense industry is "necessary to the preservation of freedom and national security" and that "its exhibitions assisted in ensuring there is a licensed, regulated and open market." Academics, however, raised concerns about this deepening interpenetration of the university with securitization and militarization, and the compromise of academic and ethical integrity. The protests provided an opportunity to debate the deepening integration of universities into the military–industrial complex, an ongoing issue that needs to be addressed as liberal economies expand, security and military industries privatize, personal insecurities increase, and liberal expectations of the public sphere – of freedom of speech and political participation – are denied.

Conclusions

Geography has and continues to be complicit with liberalism in many ways: as an academic discipline, and with respect to teaching and learning. The discipline is entrenched in a history of colonialism, and has provided the science, techniques, and experts required to make empire possible and the dissemination of liberal values. Education is saturated with a principle of individual ameliorism. Universities have internalized neoliberal entrepreneurialism, embracing the categorization of students as consumers and teachers as learning providers. As an economy, the university is embedded in the military–industrial complex, from its research programs through to its pension plan schemes. The academic enterprise itself embodies liberal ideals, with individual scholarship and invention valorized alongside reason, universalism, and egalitarianism (albeit often in terms of a meritocracy). Academics, however, have come to question this complicity, and have challenged the presumptions invested in the liberal self, liberal rule, and liberal economics. Geographers have engaged in many of these debates, with a particular emphasis on how situated knowledge and localized research reveal the contradictions and complexities at the heart of the liberal project. This situated research helps to undermine the universalizing narratives of liberal hegemony, of the fault lines that have emerged in the wake of Fukayama's totalizing claims to liberal ascendancy, Samuel P. Huntington's narrative of a 'clash of civilizations', or Benjamin Barber's *Jihad vs. McWorld*. With the war on terror launched after September 11, 2001, the political imperatives of undermining these reconfigured 'Orientialist' discourses are all the more pressing. Continued attention to the challenges to equality, universality, politics, citizenship, economies, etc. as globalization and neoliberalism reterritorialize the domestic and international landscape, will illuminate the ongoing paradoxes of a shape shifting liberalism that constantly reconfigures itself in light of what has come before, and imagines itself anew.

See also: Gender, Historical Geographies of; Governmentality; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Neoliberalism; Patriarchy; Queer Theory/Queer Geographies; Sexuality.

Further Reading

Abu Laban, Y. (2002). Liberalism, multiculturalism, and the problem of essentialism. *Citizenship Studies* 6(4), 459 481.
Barry, A., Osborne, T. and Rose, N. (eds.) (1996). *Foucault and Political Reason: Liberalism, Neo Liberalism and Rationalities of Government*. Chicago, IL: University of Chicago Press.
Baumeister, A. T. (2000). *Liberalism and the 'Politics of Difference'*. Edinburgh: Edinburgh University Press.
Brown, B. (2006). *Regulating Aversion: Tolerance in the Age of Identity and Empire*. Princeton, NJ: Princeton University Press.
Cruikshank, B. (1999). *The Will to Empower: Democratic Citizens and Other Subjects*. Cornell: Cornell University Press.
Dean, D. (1999). *Governmentality: Power and Rule in Modern Society*. London: Sage.
Foucault, M. (1991). Governmentality. In Burchell, G., Gordon, C. & Miller, P. (eds.) *The Foucault Effect: Studies in Governmentality*, pp 87 104. Chicago, IL: The University of Chicago Press.
Hall, S. (1986). Variants of liberalism. In Donald, J. & Hall, S. (eds.) *Politics and Ideology*, pp 34 69. Milton Keynes: Open University Press.
Hindess, H. (2004). Liberalism what's in a name? In Larner, W. & Walters, W. (eds.) *Global Governmentality: Governing International Spaces*, pp 23 39. London: Routledge.
Locke, J. (2006). *John Locke: An Essay Concerning Toleration and Other Writings on Law and Politics, 1667 1683*. Oxford: Clarendon Press, first published in 1689.
Mehta, U. S. (1999). *Liberalism and Empire: A Study in Nineteenth Century British Liberal Thought*. Chicago, IL: Chicago University Press.
Mill, J. S. (2006). *On Liberty; and The Subjection of Women*. New York: Penguin, first published in 1859.
Ramsay, M. (1997). *What's Wrong with Liberalism? A Radical Critique of Liberal Political Philosophy*. London: Leicester University Press.
Richardson, R. L. (2001). *Contending Liberalisms in World Politics: Ideology and Power*. Boulder, CO: Lynne Reiner Publishers.
Smith, S. (2005). *The Endgame of Globalization*. New York: Routledge.

Relevant Websites

http://www.fordham.edu/halsall/mod/modsbook18.html
Internet Modern History Sourcebook.
http://www.britannica.com/EBchecked/topic/339173/liberalism#tab=active~checked%2Citems~checked&title=liberalism%20 %20Britannica%20Online%20Encyclopedia
Liberalism in the Online Brittanica Encyclopedia.
http://plato.stanford.edu/entries/liberalism/
'Liberalism' in the online Stanford Encyclopedia of Philosophy.

Life Course Approaches

R. Fincher, University of Melbourne, VIC, Australia

Glossary

Geography of the Lifeworld The interconnected lived experiences of individuals, over long time periods in particular locales. The everyday activities of the resting or moving body are a particular focus, along with the notion of 'dwelling' – the idea that one makes one's life in a place with reference to the history of that place and the rhythms of its environment.

Intergenerational Geography The transitions people make between their own life course phases, and/or the interactions between people in different life course phases or generations. The geographic element is in the manner in which these transitions and interactions are shaped by experiences of place or space.

Life Course The concept of a life as an enactment of certain phases one after the other but always with blurred temporal boundaries. The phases or stages through which people pass are defined and given meaning by social events in the time and place of relevance. Geographies of the life course always emphasize the spatial or place-based situatedness of the life course stages.

Life Cycle The concept of a life as a series of stages defined socially, but with an emphasis on the biological or age-related nature of the stages. Family-forming stages of the life cycle might be studied, or stages marked by reaching retirement age. Because birth and death mark the beginning and end of the life cycle, stages in between may be seem as developments toward an endpoint in a way that is more marked than in other concepts of lives.

The New Social Studies of Childhood The study of how children actively negotiate their lives in different contexts, thus creating their childhoods rather than these being imposed upon them.

Life course approaches in human geography take the view that lives are experienced by people in stages, and in particular that lived experiences are influenced strongly by people's presence in certain locations, environments, places, or spaces. Research in this tradition presents the realities of people's lives in line with these expectations of life stages and environmental situatedness, whether the people whose lives are described make such claims, or not. Life course approaches demonstrate as well how stage of life is a social and spatial construct, something created in social environments by institutions, traditions, and the general expectations of families and communities, and which we as people living a life course can come to take for granted.

In its emphasis on how spatial and environmental contexts influence people's lives directly and precisely, and the attention paid to describing this empirically, geographic work on the life course differs from research on life stage processes and contexts in other disciplines, which do not have this same contextual emphasis. Perhaps like other disciplines, though, in contemporary human geography there is not a large focus on studying lives in their entirety. Rather, in geographic approaches to the study of lives situated in times and places, attention is paid by researchers primarily to particular life stages (like youth, childhood, and older adulthood). In addition, in many areas of human geographic work there are assumptions made about life stages rather than those life stages themselves being the principal focus of the analysis.

What can we think of as the origin of life course studies in human geography? Who began analytically to situate lives, whole lives made up of stages and recognizable twists and turns, in times and places? The beginnings of this conscious focus in human geography may be found in the 1970s when phenomenological geography of the lifeworld began to explore in detail the lived experiences of individuals over time in particular locales. In the work of David Seamon and Graham Rowles, though there was a strong focus on describing the everyday activities of the resting or moving body grounded in particular spaces and routines, two other matters of importance were introduced. One was the notion of 'dwelling' – the idea that one makes one's life in a place with reference to the history of that place and the rhythms of its environment. The other was the finding that memory, particularly in older people, links lives over lifetimes to places. Both these matters refer to the life courses of people, individually or grouped, and introduce the notion of people's whole lives as being somehow coherent when one conceptualizes their placement in time and space.

Geographical research on the life course also had its beginnings in the development of time geography in the 1970s, to which the work of the Swedish geographer Torsten Hägerstrand made considerable contributions. Interested in the life biographies possible for individuals in certain local environments, in which facilities were of course timetabled to be accessible only at certain times (e.g., during business hours), research in this tradition analyzed the time budgets of different individuals and groups of people, showing in this way their life paths in time and space. Carlstein, Parkes, and Thrift edited a three volume collection of essays about time geography, in which volume 2 contains useful accounts of the Swedish work. Like phenomenological geography, how ever, time geography looked less at whole lives, and more at daily lives or shorter sections of lives Both phenom enological geography of the lifeworld and time–space geography were developed as a critical response to geographic research of the time on environmental per ception and cognition. In turn, upon (or in reaction to) these twin pillars of intellectual innovation in human geography in the 1970s, feminist geographers and those applying new forms of social theory in the discipline have built their analyses of lives and life courses.

Conceptual Distinctions

Two conceptual distinctions allow us to single out the geographic pursuit of life courses. First, how does life course differ from the other terms by which the periods of people's lives have been described in social science, such as life cycle, life span, and life history? Being clear about what these various terms mean, and how they have been used, helps define the approach generally being followed to produce geographies of the life course today. Second, if life course approaches have been used more frequently in other disciplines than geography, what has geography's precise contribution to understanding the life course been?

The different terms used in social science (including human geography) about lives offer a variety of ways of seeing the relationship of the individuals living their lives, to their contexts. Life course, our main concern, sees a life as an enactment of certain phases one after the other but always with blurred boundaries, in which the phases are defined and given meaning by social events in the time and place of relevance. There is no assumption here that every individual or particular group, even in one time and place, will experience the important events of that time and place in exactly the same way, and shape or time their actions identically. Nevertheless, general claims can be made from observations of many lives in certain contexts, so that overall a particular interpretation of life stages is given indicating that most people ex perience things that way. So – lives, in the life course approach, are produced in a particular time and space, and guided by particular events. They are socially (and, especially in geography, locationally) defined and rooted. Often, research has taken a major event as its starting point – like a World War – and examined the way that event and its social aftermath have appeared in the lives of people involved in that event.

Life cycle, a different term, has more explicit bio logical roots. (The term life span is the same.) Though still posing lives as a series of stages defined socially, life cycles are more explicitly age related, and there is a beginning with birth and an end with death to mark the cycle. Much emphasis is placed on the age related question of biological reproduction, with men's and women's life cycles differentially staged with reference to this. (Of course reproduction and the family are important matters for life course analyses as well, but are not defined so precisely with respect to age and the biology of aging.) In addition, policies like the age of retirement, or the age at which voting or driving a car are permitted to occur, are predicated on a view of the life cycle that sees maturity as occurring at a certain age, or (in the case of retirement) the end of one's useful working life as occurring at a certain age. Of course these age related judgments are assumptions based on the average situation across a large group; any one individual may differ markedly from this average.

Life history is yet another term from which to sep arate life course and life cycle. It is the biography of an individual. A life history is the property of an individual, and cannot be generalized to a population. It is evidence collected or reported about a life that might make ref erence to biological, historical, and place based events and stages. It could, however, take either a life cycle or a life course approach.

The concept of the life course is distinct from those of life cycle (or life span) and life history. But it is fair to say that echoes of all three concepts appear in most uses of any one of them. Different disciplines and theoretical perspectives have been associated with varied uses of these three terms about the study of lives. Developments in historical demography have advanced through work on the life histories of ordinary people in specific times and places. Developmental psychology, also, has used life histories to associate maturation stages more sensitively with aging. Research in the sociology of aging has dem onstrated the various ways that advancing through life stages, thus forming a life course, occurs in the context of different opportunities being presented and taken up or foregone. In important reviews, such as that by the sociologists Giele and Elder, of the social scientific use of the concept of life course, which has grown since the 1980s, the discipline of geography is rarely mentioned as a major contributor. Though the importance of lives being situated in various 'times and places' is often stressed, environments, places, and spaces are often subsumed in other disciplinary accounts as part of a person's or group's history. Other disciplines do not have the commitment to detailed investigation of environ ments or places as contributors to life courses, in the way that human geography does.

Contributions of Contemporary Geography on Life Courses

The contributions of contemporary geography to research on the life course come from its attention to the embeddedness of lives in sociospatial as well as temporal settings. How this embeddedness actually plays out, for people in places, is its focus. As mentioned already, rarely do we see in contemporary geographic work an account of the entirety of lives. Rather, certain segments of lives – such as childhood, adolescence, retirement – are singled out for analysis of their particular situatedness in a place–time. Advocates of longitudinal studies of whole lives would probably term contemporary studies of parts of lives cross sectional, seeing these studies as taking a temporal snapshot rather than considering the whole phenomenon of a life. But life histories, recounted as stories in an analysis to ground it in time as well as space, do also often give a longer perspective. This occurs particularly in contemporary geographic research on the contexts of women's adulthoods and their caring responsibilities, in geographies of children and childhood, and in the recognition that our public institutions make assumptions about life course stages that may not be accurate, socially or spatially. These are three areas of human geography in which a life course approach is being pursued successfully, as the following comments will attest.

In their pioneering collection from the 1990s *Full Circle*: *Geographies of Women over the Life Course*, Cindi Katz and Janice Monk record their conscious attempt to consider women's whole lives, not only across the full time span of those lives but also in terms of the range of experiences and activities that interact to form those lives. Their book seeks to include links between women's work of production and reproduction in a range of countries from different parts of the world of varying degrees of wealth. "This attempt to craft the fullness of lives in their precise locations," say Katz and Monk, "is opening up new ground for human geography." For their review of human geography to that time had found it failing to consider more than specific periods in women's lives (for example, ignoring the links between women's adolescence, middle years, and older age but rather focusing on the years of early adulthood). They had also found human geographies of women's lives concentrating disproportionately on matters like employment possibilities for women in Western countries (often cities).

Some of that lack of attention to the life course, as a set of linked and situated stages in life made up of interactions between a range of activities, is now being remedied, as we will see. But it is important to note that the body of work on women's adulthoods of caring along with paid work among myriad other responsibilities, even if it only implicitly references the life course as a conceptual basis, has been profoundly important in shedding light on the spatialities of this life course stage, and sometimes the transitions to and from that stage. It has revealed as well the variations in the allocations of caring tasks between families in different times and places. Nowhere is this better illustrated than in Katz and Monk's book, in which the tasks of women working for an economic livelihood ensuring social reproduction are situated both in numerous locations (in frontier farming areas of Colombia, in the rural Caribbean, as well as in small towns and major cities of the industrialized West) and in relation to transitions to and from other life stages.

It is interesting to ponder why it is that feminist geography, or geographies concentrating on the lives of women, are the places in our discipline in which matters of life course have appeared most frequently. Perhaps this is because of the manner in which the life cycle echoes through all life course research – with the stress of the life cycle concept on age related stages in life and their particular biological implications for childbearing and the production of families. Women's lives in many settings remain affiliated in policy and public expectations with childbearing and rearing, even if this is not the reality of every young adult woman's life. Men's lives do not seem to be studied from this point of view, in human geography.

Another area of recent geographic writing on the life course has been within what has been termed, by human geographers like Sarah Holloway and Gill Valentine, the new social studies of childhood. Geographers participating in this intellectual endeavor to articulate how children actively negotiate their lives in different contexts, thus creating their childhoods rather than these being imposed upon them, are having their concepts of spatiality taken up in the work of a range of other social scientists. Drawing on the accounts of Holloway and Valentine, it is evident how geographic understandings of the importance of place in the creation of children's lives may be illustrated, for example, by Katz's comparison of childhoods in the environments of rural Sudan with those of New York City, in which in some ways the former is freer for play than the latter. At a different scale, the situatedness of children's activities may be examined through studying children's use and production of everyday spaces, for example, in streets, homes, and schools, where limits and possibilities are negotiated daily by children. In addition, prominent spatial discourses in a locality, for example, statements made by a city council to encourage young people to use certain facilities or spaces, may create opportunities which previously had been unimaginable. In her important book about homeless youth in West Hollywood, Los Angeles, Susan Ruddick describes just such a situation, in which service providers managed to make homeless youth an accepted presence in that municipality. As do Holloway and Valentine, it is important to acknowledge the work by geographers over several decades on children's spatial cognition and perception, on which the new (more sociological than psychological) social studies of child hood build.

Moving to the societal rather than the individual scale, changing norms about youth and aging, even childhood, are basic to an evolving set of capitalist re lations. Ruddick has shown how the mantle of desirable youthfulness has been given, by institutions of marketing and public policy, to certain workforce situations and practices such as secondary labor market status and the need for lifelong learning and reskilling. This introduces the third contribution of human geography to life course studies – the identification of institutional assumptions about life courses, that are then able to be challenged if this is necessary.

Often implicitly or without fanfare, assumptions about the appropriate life course that individuals should follow guide the policies and institutional practices of govern ments, firms, and other organizations. This affects the opportunities and lives of individuals, groups, and places. Continuing to think of childhoods, the absence of chil dren's voices from urban planning has been remarked upon, in research emerging around the world (following a recent United Nations report) on child friendly cities. As already noted, Susan Ruddick has pointed out the con temporary pervasiveness of views of the desirability of youthfulness in labor forces, and the economic rationale of this in an expanding capitalism.

In another vein, looking at the institutions and in stitutional actors shaping the physical form of major cities, reference should be made to a pathbreaking 1970s study of great relevance to urban geographers: *Everything in Its Place: Social Order and Land Use in America* by the urban anthropologist Constance Perin. Perin found in the 1970s that urban planners, real estate agents, and developers viewed the suburbs and the inner cities of America as appropriate residential sites for people of particular life cycle stages. These city building pro fessionals created housing and community services in urban locations according to such views. One sees similar assumptions at work in the marketing and building of luxury, inner city, and high rise apartments in major Western cities to this day. Fincher has written on this topic recently. In the 1990s, urban geographers analyzed the creation of residential settings exclusively for older residents in a number of American muni cipalities, and the ethics and consequences of this. A particular contribution has been made, then, by geog raphers examining the institutions that help shape the physical and social features of particular places, exposing the assumptions about appropriate life courses that these institutions put into practice.

Issues for Continued Attention

What issues can be signaled as requiring further atten tion, in extending geographies of the life course? On the one hand, life course stage and life course transitions have generally been neglected in the studies of the social construction of difference pursued in human geography. Life course characteristics have often been assumed as present, but left relatively unexamined, in studies made of the spatialities of gender, ethnicity, and class. The call made by Katz and Monk in the early 1990s to focus on life courses, in their case the life courses of women, has not really been heeded, other than implicitly, in exam inations of difference. But on the other hand, there are two particular issues, within work that does focus on the conceptualization of age and life related stages and timing, that deserve to be mentioned here.

One has been eloquently argued by Vanderbeck, in a call for what he terms intergenerational geographies. Quite validly, Vanderbeck claims that age related re search in geography has tended to focus on particular age groups, usually the young or the old (though research on the group of adult women combining paid work and caring, on whom considerable focus by feminist geog raphers has been made, should be added). The transitions between life stages, much studied in the life course ap proaches used in other disciplines, have been less a concern to geographers. What this means is two things, says Vanderbeck: first, that intergenerational activity is not studied, and second, that societies and places are emerging in which generations, and age groups, are separated in much of their activity or even physically segregated by the location of their housing and the community facilities they use. Already mentioned has been the occurrence of the type of thinking that separates life course groups, in those institutions that see (and therefore construct) certain urban environments as ap propriate for people of certain life stages. Examples of this tendency, and its consequences, need to be docu mented as we create new geographies of the life course.

A second matter for attention is this. In studies of lives, human geographers tend to view the stages iden tified as relatively homogeneous. Childhood, adulthood, old age – these are generalized as quite commonly ex perienced within certain places and times. And yet we know that these are blunt categories – that there is 'de velopment' within adulthood, and there is, as Ruddick has said in her study of youth in Hollywood, room for a category of adolescence that is neither childlike nor adultlike. Recognizing that geographers do tend to study life course stage groups rather than whole lived lives, in the spatial contexts in which they are situated, how can these categories of childhood, adulthood, and old age be productively complicated? More precisely, is the com plication to be desired in a more nuanced human geography of the life course one that will come primarily from closer examination of the time and place in ques tion, or from closer examination of the activities and attitudes of the individuals within the group being studied? Efforts to embed this matter into contemporary theorizing about the differences produced over lifetimes would be welcomed.

See also: Children/Childhood.

Further Reading

Carlstein, T., Parkes, D. and Thrift, N. (eds.) (1978). *Timing Space and Spacing Time*, 3 vols. New York: John Wiley.

Fincher, R. (2007). Is high rise housing innovative? Developers' contradictory narratives of high rise housing in Melbourne. *Urban Studies* 44, 631 649.

Giele, J. Z. and Elder, G. H. (eds.) (1998). *Methods of Life Course Research: Qualitative and Quantitative Approaches*. Thousand Oaks, CA: Sage Publications.

Holloway, S. L. and Valentine, G. (eds.) (2000). *Children's Geographies: Playing, Living, Learning*. London: Routledge.

Katz, C. and Monk, J. (eds.) (1993). *Full Circle: Geographies of Women over the Life Course*. London: Routledge.

Perin, C. (1977). *Everything in Its Place: Social Order and Land Use in America*. Princeton, NJ: Princeton University Press.

Rowles, G. (1978). *Prisoners of Space? Exploring the Geographical Experience of Older People*. Boulder, CO: Westview Press.

Ruddick, S. M. (1996). *Young and Homeless in Hollywood: Mapping Social Identities*. New York: Routledge.

Ruddick, S. M. (2003). The politics of ageing: Globalization and the restructuring of youth and childhood. *Antipode* 35, 334 362.

Seamon, D. (1979). *A Geography of the Lifeworld*. New York: St Martin's Press.

Vanderbeck, R. M. (2007). Intergenerational geographies: Age relations, segregation and re engagements. *Geography Compass* 1, 200 221.

Literature

M. Brosseau, University of Ottawa, Ottawa, ON, Canada

Why should geographers be concerned with fictional literature? Shouldn't they be content with the study of the 'real' places that the 'real' world has to offer? After all, isn't geography a serious, if not scientific, enterprise dealing with empirical facts? If so, why should geographers be attracted to the fictional worlds, or fictive geographies, found in literary works? Generations of geographers have had different answers to these questions and intellectual motivations for resorting to literature have varied significantly. Although there is an emerging subfield in literary studies concerned with the geography of literary texts, this article deals specifically with the contributions of geography to the study of literature and the corollary contribution of literary analysis to the study of geography. Parallel in many ways to the discipline's recent interest in travel writing, the geographical analysis of fictive literature – or literary geography – constitutes a more or less autonomous constellation of research that has been the subject of successive critical reviews since the late 1970s.

Development of Literary Geography in the 1970s and 1980s

While geography did not begin a sustained engagement with literature until the 1970s, a number of earlier works point to the discipline's nascent interest in literature. Reviewers generally trace the origin of this interest to the publication of *Mill's guide to geographical books* in 1910. However, one could also refer to Vidal de La Blache's short paper on the geography of the Odyssey, which appeared in 1904, or even earlier, to Von Humboldt's *Cosmos* of 1847, which contains interesting insights into literature and painting. Yet, none of these authors clearly advocated the study of literary sources as a field of investigation for geographers. It was Wright who eventually sought to establish the relevance of such sources in his 1926 'plea for the history of geography'. For some, these early engagements with literature were above all an invitation to develop a more literary approach to geographical description.

Geographical Facts into Fiction: Regional Geography and Literature

Are novelists good geographers? Do they stick to the kinds of facts geographers would include in a textbook or a regional monograph? These are some of the first questions regional geographers began asking of literature in the early 1970s. The aim was to evaluate literature's documentary value, in particular that of nineteenth century realist novels that offered ample and seemingly authoritative descriptions of regional distinctiveness. In their engagement with literature, regional geographers sought to separate facts from fiction, an exercise referred to by Salter and Lloyd as the analysis of the "literal meaning of landscape." Although it was acknowledged that there can be no strict correspondence between the written and the actual landscape – a perfect mimesis so to speak – their analyses tried to establish whether or not the depiction of human–nature relationships, regional landscapes, and rural economies could be considered reliable and therefore worth quoting. It would seem that fiction or 'geographical freedom' was tolerated only in as far as it applied to the text's specific geographical referents (topographic features, place names, etc.), whereas the relationships that constitute the overall *genre de vie* depicted were considered accurate, at least in literary sources that stood up to the scrutiny of geographers. One could say that the focus of such an approach to literature was ultimately to harness the power of literary language to evoke landscapes and the 'personality' of regions, a quality that geography could benefit from once factual and fictional elements has been clearly distinguished. The attachment to geographical accuracy and verisimilitude is also understandable if one considers that the work in question was often motivated by a pedagogical spirit that sought to evaluate the potential of literary texts for teaching geography. Ultimately, the real value of literary sources to regional geography lay in what Chevalier refers to as the irreplaceable *supplément d'âme* that literature offers traditional geographical research. It is precisely from this perspective that humanistic geography's approach to literature evolved.

Imagination in Place: Literature and Humanistic Geography

In a bid to break the monotony of hard facts which had become the norm with the establishment of quantitative geography in the 1960s, humanistic geography sought to bring human experience, subjectivity, and agency, as well as notions of meaning and value back to the core of geographic inquiry. Consequently, experience of place and sense of place quickly became central concerns as humanistic geography turned to literature in an effort to establish a finer balance between geographical information and the human imagination. For many humanistic geographers literature represented, and in some case continues to represent, a somewhat mystical realm where the most concrete aspects of the outside world and the subjective experience thereof are blended in perfect harmony.

In the humanistic geography tradition of literary analysis, fiction is not considered a central feature of literary discourse. Due to its central concern with experience and sense of place, the thirst for an actual spatial referent, a real presence in a real place, remains crucial. However, the factual exactness with which this place is depicted is no longer important. The emphasis is on the experience, its intensity, and immediacy, and not so much on the place itself. The values, meanings, and feelings that the literary imagination associates with place are the focus. Similarly, for French speaking geographers with humanist leanings, literature was seen as a source to outline the contours of the *espace vécu*. Again, nineteenth century realist literature was the privileged source, for it was viewed as a form of mimetic transcription of the experience of places where the human–nature relationships reached their most harmonious expression. Having said this, as points of entry into the places and landscapes of the mind, other literatures were also used and more nuanced readings were sometimes achieved. For example, Porteous considers the imaginary landscapes of Malcolm Lowry's urban novels using the 'symbol' as a way to decipher the values attached to place, thus avoiding some of the most obvious traps of an unexamined realist reading, be it objective or subjective.

The role of humanistic geographers has been central in demonstrating the validity of literature for geographical inquiry. Criticized by radical geographers for its elitist and individualistic view of literature, humanistic geography nonetheless showed how the geographical imagination is not the sole property of academic geographers. In doing so, it placed the study of literature firmly on the agenda of human and cultural geography.

Social Justice in Literary Space: Radical Geography and Literature

As part of the disciplinary backlash against the quantitative orthodoxy associated with the spatial science approach that dominated geography during the 1960s, a radical current rooted in Marxism arose during the 1970s with the explicit objective of politicizing human geography. Although less abundant than work from regional and humanistic perspectives, radical geography also produced research on literature. The latter is characterized by an admittedly sociological perspective when compared to the resolutely individualistic perspective of the humanists. Still very much relying on a mimetic conception of language, radical geographers saw literature as a reflection of the social and spatial conditions of existence. Adopting a materialistic view of culture and literature, the fictive dimension of literary discourse was conceptualized as a form of departure from established reality, a departure to be criticized or celebrated, depending on the ideological viewpoint.

Cook, for example, puts the novel and its author in their original context of production in order to understand how the novel 'distorts' social and geographical reality. The novel's distortions are interpreted as the expression of ideological false consciousness, which results from the deforming filter of the author's class identity. Others, such as Olwig, try to understand "not so much the individual's apprehension of geographic reality as it actually is, but literature's social function in envisioning reality as it is not but ought to be, and with its potential, thereby, for stimulating change." From this perspective, the fictive dimension of literary sources is seen to have a liberating function because its 'difference' contradicts the monopoly of 'established reality'. As a departure from reality, literary fiction (1) shows the world from the taken for granted standpoint of the ruling class (bourgeois discourse), and (2) shows the world as it really is (critical discourse) or (3) shows what reality ought to be (revolutionary discourse). Overall, the significance of radical geography's engagement with literature lies in its insistence on the social relevance of literary production: literature is not the creation of a solitary genius, but a social production embedded in the materiality of social life.

Critiques of Regional, Humanistic, and Radical Geography Approaches to Literature

Brosseau's critical review of geography's engagement with literature during the 1970s and 1980s argues that regardless of whether literature was used to document regional characteristics, gain insight into human experience, or to critique hegemonic social structures and relationships, in all cases literature was approached 'instrumentally' for the purposes of confirming or illustrating established geographical hypotheses and epistemological standpoints. In other words, literature was co opted as a proving ground for geographical questions that had largely predetermined conclusions. Moreover, since the fictive dimension of literature was rarely problematized or considered significant, only a portion of literary texts was drawn upon – that which could be transformed, as Robinson (1987: 192) adequately puts it, into "a comforting readymade source for the social sciences." For Brosseau, the overall relationship to literary texts common to regional, humanistic, and radical geography not only denies the density, complexity, and ambiguity inherent to literary texts, but also explains why geographers were not usually attuned to the ways in which literature can disrupt traditional disciplinary perspectives or generate new geographical questions and 'novel' forms of answering them.

Furthermore, the discursive dimensions of literature – language (poetics, rhetoric, style, etc.) and form (literary convention, genre, composition, narrative structure, etc.) – were more or less ignored by these early approaches. After all, this is the stuff of literary critics and theorists, not geographers. For most geographers, the literary text was seen to yield its message unproblematically, and as such did not pose any challenge to a geographical reading. It is thought of in the realm of mimetic and transitive communication, not representation and all that this implies. The rejuvenation of cultural geography, with its more intense focus on issues of representation and textuality, has transformed the way many geographers approach literature (**Table 1**).

Literary Geography and the Cultural Turn

The Normalization, Sophistication, and Diversification of Literary Geography

At the beginning of the 1990s, geography's engagement with literature went through a triple transformation process: normalization (literature has become accepted as an almost matter of fact object of study for the discipline, so much so that literary geography has become textbook material), sophistication (geographical studies of literature are now informed by both literary theory and an array of critical theories, including feminism, post structuralism, postcolonialism, etc.), and diversification (geographers are analyzing a wider range of literary writings, including contemporary urban novels, children's fiction, crime fiction, science fiction, fantasy, poetry, comic books, etc.). This process of transformation, associated with a greater concern for discourse and textuality in human and cultural geography, has led to a 'reproblematization' of the less than clear boundaries between fiction and reality as well as between academic and literary discourse. This makes it all the more difficult to paint a coherent picture of geography's contemporary engagement with literature, for contemporary geography is more plural than ever.

Lines of Continuity and New Directions

Cultural and social geography's more intense engagement with literary theory, post structuralism, postcolonialism, feminist theory, and, more generally, with cultural studies as a whole has contributed to redefining geography's relationship with literature. However, the approaches that were developed in the 1970s and 1980s have not completely disappeared. The fascination with the referential quality of realist writing still prevails in some contemporary analyses, either for pedagogical reasons or for the benefit of historical geography. Humanistic studies of literature with a relatively unchanged focus on spatial experience and sense of place are still regularly contributed, albeit many incorporate contemporary themes, as in the case of a collection edited by King, Connell, and White that considers the intersection of identity, gender, travel, exile, and migration. They are also more inclined to engage with contemporary urban literatures than nineteenth century realist novels. In French speaking geography, the geographical imagination (not to be confused with the imaginative geographies described by Edward Said) remains an important area of research for humanistically minded geographers. Radical or materialist approaches, which never produced many case studies in the first place, have been recast into analyses that take into account not only class or race, but also gender, ethnicity and sexuality, identity politics, and resistance.

New directions in the geographical engagement with literature are difficult to organize into coherent streams of research given the plurality of contemporary studies in terms of literary genres studied, themes raised, critical stances adopted, and political prerogatives foregrounded. Notwithstanding, it is possible to categorize new directions along the following lines, which undercut the apparent impression of fragmentation associated with scholarship in a postmodern environment.

Alternative Geographic Epistemologies: Taking Literary Language and Fiction Seriously

The growing attention paid to discourse and textuality in geography and other social sciences has changed the way geographers approach literature. As mentioned earlier, despite their fundamental differences all of the various approaches to literary geography discussed above paid little attention to literary language and discourse *per se*. In the early 1990s, a growing number of geographers started to spend more time appreciating the relationship between textuality and the specific geographies that novels can generate. In simple terms, they started to take literary language seriously in their effort to understand how spaces, places, and peoples are represented in literature. As a result, they have become engaged in a more open ended, dialogical relationship with literature that is sensitive to the insights offered by literary difference: literary analysis is no longer seen as a monologic endeavour aimed at illustrating or demonstrating a preformulated hypothesis. Correspondingly, geographers no longer tend to see the fictive dimension of literature as a problematic barrier to overcome or neutralize, but rather as an important source with which to engage for epistemological insight.

Table 1 Geograph ca approaches to terature

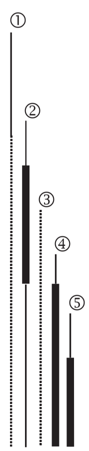

	Approach	Value	Dominant genre	Representation	Themes
1960	1- Regional	Documentary, pedagogical	Nineteenth-century realist and regional literature	Mimesis (reflection of geographical reality)	Fact into fiction, regional identity, landscape signatures
1970	2- Humanistic	Experiential, existential, symbolic	Nineteenth-century realist and regional literature	Indirect mimesis: reflection of the experience of geographical reality	Sense of place, experience of place, *espace vécu*
1980	3- Radical-Materialistic	Ideological, critical	Nineteenth-century realist, regional, and popular literatures	Mimesis (reflection of social conditions of production)	Social (in)justice in space (class, race, etc.)
1990	4- New culturalgeography A	Epistemological, cognitive, transformative	Modernist, crime/noir, contemporary, science fiction, etc.	Constructivist, post-structuralist	Writing the city, urban futures, geographical knowledge, and discourse
2000	5- New cultural geography B	Counter-hegemonic, critical	Colonial and postcolonial, non-Western, children's fiction, fantasy, etc.	Constructivist, post-structuralist	Politics of identity and difference (gender, class, ethnicity, sexuality, etc.)

Dominant; emerging/continuing; sporadic contributions.

It is only since the 'cultural turn' that geographers have started to draw attention to the radically different types of geographies literature can produce. In other words, studies are now increasingly attentive to how literature, and more specifically novels, not only reflect geographies already familiar to the researcher, but also generate unique geographies of their own, the analysis of which is capable of questioning the discursive rationality of academic geography. Brosseau, for example, examines the benefit of a dialogical approach in his analysis of the representation of urban life in the modernist novel *Manhattan Transfer* by John Dos Passos. In demonstrating that the fragmentation and rhythm of the urban experience is embedded in the very form of the text itself (composition, contrasted montage, etc.), he draws attention to the profound differences between scholarly and literary modes of textual representation. Similarly, Howell proposes that the alternative epistemologies of urban space and experience offered by crime fiction can be used as a basis for critiquing the epistemological assumptions and discursive strategies that characterize radical geography's scholarship on the city. On his part, Stainer argues that Ciaran Carson's novel *The Star Factory* subverts constructions of place, identity, and difference associated with sectarian nationalism in Belfast through an imaginative reinvention of urban geographical experience that emphasizes fluid and hybrid forms of identification, multiple viewpoints, and the transgression of territorial boundaries. Consequently, Stainer proposes that Carson's novel offers a progressive geographical imagination that acts as an alternative to the antagonism of sectarian politics. Finally, one can also point to a study by Kitchin and Kneale, which argues that science fiction novels constitute cognitive spaces in which possible sociospatial and sociotechnical futures are articulated free from the constraints of academic prediction. According to the authors, these future imaginings shape present day individual and institutional thought and practice, guiding the social and technological development of emerging spaces such as the Internet. Ultimately, the type of engagement with literature exemplified here questions or even destabilizes both the way we think about discourse, language, and representation and its impact on geographical knowledge. In doing so, it shows how the fictive geographies generated by literature cannot be separated from the textual form in which they are embedded.

Interventions in the Cultural Politics of Representation

Dissatisfied with 'traditional' American cultural geography's focus on material culture and landscape, and the epistemological tug of war between humanistic and radical approaches, many cultural geographers have worked to reconceptualize culture and explore ways to bridge the gap between what were once thought to be incompatible epistemological positions. This has resulted in a pluralization of approaches to literature as an object of geographical analysis as well as a diversification of the types of literary works studied. Nevertheless, these approaches do share a key characteristic: literature is not considered the creation of individual minds that are capable of articulating universal truths about the human condition, nor is it considered the simple product of authors conditioned by their position in an overarching social structure.

In a context in which culture is conceived of in terms of a 'web of meanings' subject to various lines of power, geographers acknowledge the partial and situated character of cultural representation, whether academic or literary. They have undertaken the task of teasing out the ways in which not only class and gender, but also race, ethnicity, national identity, and sexuality inform literary representation of peoples and places. For many contemporary geographers, literature has become one of the many signifying practices that not only reflect cultural phenomena (in a never perfect mimesis), but also contribute to the construction and circulation of situated cultural meanings. This implicates literature and its study in the politics of cultural representation, difference, and place. Two substreams of research can be identified within this particular way of conceptualizing literature. The first one is concerned specifically with the cultural politics of postcolonialism and the second with the cultural politics of identity and difference.

Cultural geography's engagement with postcolonial studies, and especially its emphasis on the production and diffusion of 'imaginative geographies', has transformed the way we look at literatures of the past. These have become conceptualized as culturally embedded discursive artifacts that can be deconstructed to expose the various ways through which images of 'otherness' and 'other' spaces were produced, diffused, and ultimately naturalized. Parallel in many ways to the renewed interest in 'travel writing', geographers studying literature have highlighted asymmetrical power relations that inform the representation of people and place, gender and ethnicity, and empire and periphery. Phillips, for example, discusses how the imaginative geographies generated in nineteenth century European adventure fiction are intimately connected to dominant ideologies of empire and masculinity prevalent at the time. In a similar vein, Hugill argues that American and British boys' novels of the Edwardian period played a significant socializing role by promoting manly, Christian values and in doing so, participated in the process of imperialism. Postcolonial theory has also proven a useful basis for teasing out the politics of representation in contemporary literature. For example, Balfe identifies the Orientalist tropes through which popular fantasy novels construct 'otherness' in relation to the Western 'Self' of heroic protagonists.

In recent contributions, literature is also considered as one of the many social discourses implicated in the politics of identity and difference. As such, literature and its interpretations can be conceptualized as a form of intervention into these politics. In this case, the term 'intervention' stems from the social and political agency of the authors (and sometimes that of the interpreter), who may promote, question, or resist culturally dominant meanings, and generally involves highlighting the situatedness or positionality (in terms of class, race, ethnicity, gender, sexuality, as well as any other layers defining identity) of literary discourse.

Feminist interpretations of literature were probably the first to open this type of inquiry. Earlier feminist studies, for example, demonstrated the relevance of a spatial reading of gendered spaces in literature and the power inherent to transgressing the private and public spheres. More recently, McKittrick has studied how gender and race in Toni Morrison's novel *The Bluest Eye* are constituted, naturalized, and eventually contested in relation to various spatial scales including the nation, the community, the home, the body, and the mind. Each of these scales is shown to have discursively regulated ex pectations of racialized and gendered behavior that are interpreted in different ways by Morrison's characters. Likewise, Hughes conceptualizes the fiction of Jamaica Kincaid both as a register of daily life in domestic space within the patriarchal context of British rule, and as an intentional political intervention that seeks to critique the power relations characteristic of colonial rule in the Caribbean.

In tune with contemporary debates about hybridity and identity, some have found Salman Rushdie's *The Satanic Verses*, to be a rich source in understanding the intricate, ambivalent, and hybrid national identities. In a similar vein, Jazeel examines the spatialized politics of sexuality and race in Shyam Selvadurai's *Funny Boy*. Jazeel proposes that Selvadurai's novel, "tackles Sri Lanka's entangled intersections of sexuality, race, and the politics of difference not just to show the spatialization of discourses that regulate sexuality or the inherent spati ality of tactics used to unsettle sexual discrimination, but more importantly to also unsettle the dialectics of div ision in post independent and post war Sri Lanka...." Although dominated by issues revolving around identity and difference, such studies of literature can also focus on other types of cultural contestation, such as Creswell's analysis of mobility as a form of resistance in Kerouac's of famous novel *On the road*.

Conclusion

It would be fair to say that, with its present range of themes and conceptual approaches, literary geography has become an intellectually mature nexus of study that now constitutes an autonomous subfield in human geography, its development having paralleled that of human and cultural geography in many ways. Con ceptualized variously as a documentary source, a tran scription of the experience of place, a reflection of the sociospatial conditions of its own production, a source of alternative geographic epistemology, or as a situated cultural discourse intervening within the politics of representation, literature has come to inform the research pursuits of many social and cultural geographers on a diverse range of questions. As such, one could just as well argue that literary geography does not constitute a sub field in its own right, but rather that the intersection between geography and literature has become one of many ways in which geographers can expand their understanding of the spatiality of social life. It is a question of perspective as to whether it is literature or geography itself that constitutes the ultimate object of inquiry when geographers turn their attention to literary works.

See also: Cultural Geography; Cultural Turn; Determinism/Environmental Determinism; Difference/ Politics of Difference; Discourse; Identity Politics; Postcolonialism/Postcolonial Geographies; Poststructuralism/Poststructuralist Geographies; Representation, Politics of; Text and Textuality; Text, Textual Analysis.

Further Reading

Balfe, M. (2004). Incredible geographies? Orientalism and genre fantasy. *Social and Cultural Geography* 5(1), 75 90.

Brosseau, M. (1994). Geography's literature. *Progress in Human Geography* 18(3), 333 353.

Brosseau, M. (1995). The city in textual form: Manhattan Transfer's New York. *Ecumene* 2(1), 89 114.

Brosseau, M. (1996). *Des romans géographes*. Paris: L'Harmattan.

Bureau, L. (1997). *Géographie de la nuit*. Montréal: L'Hexagone.

Chevalier, M. (1993). Géographie et littérature. In Chevalier, M. (ed.) *La littérature dans tous ses espaces*, pp 1 83. Paris: Editions du CNRS.

Cook, I. G. (1981). Consciousness and the novel: Fact or fiction in the works of D.H. Lawrence. In Pocock, D. C. D. (ed.) *Humanistic geography and literature*, pp 66 84. London: Croom Helm.

Cresswell, T. (1993). Mobility as resistance: A geographical reading of Kerouac's on the road. *Transactions of the Institute of British Geographers* 18, 249 262.

Frémont, A. (1981). Flaubert géographe. A propos d'Un coeur simple. *Etudes Normandes* 1, 49 64.

Géographie et Cultures, (2002). *Special Issue: Territoires littéraires. Géographie et Cultures* 44.

Géographie et Cultures, (2007). *Special Issue: Le Roman policier: Lieux et itinéraires. Géographie et Cultures* 61.

Hancock, C. (2002). Salman Rushdie, l'écriture contre le territoire. *Géographie et Cultures* 44, 43 62.

Howell, P. (1998). Crime and the city solution: Crime fiction, urban knowledge, and racial geography. *Antipode* 30(3), 357 378.

Hughes, R. B. (1999). Empire and domestic space in the fiction of Jamaica Kincaid. *Australian Geographical Studies* 37(1), 11 23.

Hugill, P. J. (1999). Imperialism and manliness in Edwardian boys novels. *Ecumene* 6(3), 318 340.

Jazeel, T. (2005). Because pigs can fly: Sexuality, race and the geographies of difference in Shyam Selvadurai's Funny Boy. *Gender, Place and Culture* 12(2), 231 249.

King, R., Connell, J. and White, P. (eds.) (1995). *Writing across worlds: Literature and migration*. London: Routledge.

Kitchin, R. and Kneale, J. (2001). Science fiction or future fact: Exploring imaginative geographies of the new millennium. *Progress in Human Geography* 25(1), 19 35.

McKittrick, K. (2000). Black, and 'cause I'm black, I'm blue: Transverse racial geographies in Toni Morrison's the bluest eye. *Gender, Place and Culture* 7(2), 125 142.

Monk, J. and Norwood, V. (1990). (Re)membering the Australian city: Urban landscapes in women's fiction. In Zonn, L. (ed.) *Place images in media*, pp 105 132. Maryland: Rowman and Littlefield.

Olwig, K. R. (1981). Literature and 'reality': The transformation of the Jutland heath. In Pocock, D. C. D. (ed.) *Humanistic geography and literature. Essays on the experience of place*, pp 47 65. London: Croom Helm.

Phillips, R. (1997). *Mapping men and empire: A geography of adventure*. London: Routledge.

Porteous, D. (1990). *Landscapes of the mind: Worlds of sense and metaphor*. Toronto: University of Toronto Press.

Preston, P. and Simpson Housley, P. (eds.) (1994). *Writing the city: Eden, Babylon and the New Jerusalem*. London: Routledge.

Robinson, B. S. (1987). The geography of a crossroads: Modernism, surrealism, and geography. In Mallory, W. E. & Simpson Housley, P. (eds.) *Geography and literature: A meeting of the disciplines*, pp 185 199. Syracuse: Syracuse University Press.

Salter, C. L. and Lloyd, W. J. (eds.) (1977). *Landscape in literature*. Resource paper for college geography. Washington, DC: Association of American geographers.

Sharp, J. (2000). Towards a critical analysis of fictive geographies. *Area* 32(3), 327 334.

Stainer, J. (2005). The possibility of non sectarian futures: Emerging disruptive identities of place in the Belfast of Ciaran Carson's the star factory. *Environment and Planning D Society and Space* 23(3), 373 394.

Relevant Websites

http://www.imaginingtoronto.com/
Intersections of Literature and Place in the Toronto Region.

http://placestudies.com
Place Studies.

Livelihoods

F. Owusu, Iowa State University, Ames, IA, USA

Glossary

Livelihood Capabilities, assets (including both material and social resources), and activities required for a means of living.

Livelihood Diversification The process by which families construct a diverse portfolio of activities and social support capabilities in order to survive and improve their standard of living.

Multilocal Livelihood Spatial dispersion of household members to enable the household to take advantage of economic opportunities in multiple rural and urban places or multiple countries.

Sustainable Livelihood A livelihood is sustainable when it can cope with and recover from stresses and shocks and maintain or enhance its capabilities and assets both now and in the future, while not undermining the natural resource base.

Introduction

Since the early 1990s, a stream of research in development geography that falls under the broad banner of livelihood studies has been used in planning, implementing, and evaluating research and development programs by international development agencies, nongovernmental organizations, and government agencies for poverty reduction, environmental management, rural development, and sustainability objectives. Contemporary livelihood studies have their roots in the household survival strategies studies of the 1980s, which were part of the shift away from the structural perspectives of the 1970s (the dependency theorists and neo Marxists). Survival strategies studies presented the poor not as passive victims of structural forces; rather, proponents argued that even though the poor make decisions that affect their survival within the confines of limiting structural conditions, they often have some room to maneuver. Consequently, the actions of individuals and households in livelihood studies were often considered strategic and well informed. Contemporary livelihood research broadens the discussion by exploring people's active involvement in responding to and enforcing change. Livelihood studies also focus on the relationship between macro level processes and domestic units, which makes them particularly useful for understanding the interplay of global and local factors (i.e., the global–local nexus).

The findings from livelihood studies across the developing world show that increasing numbers of people have opted for a development path characterized by income diversification. In fact, diversification of income sources is not limited only to poor people; rather, individuals and households across different classes in the developing world increasingly rely on multiple sources of income. Diversification of assets, income, and activities is also driven by multiple motives.

Livelihood research has also helped in highlighting intrahousehold differences, the problems involved in seeing the household as a homogenous unit of corresponding interests, and the fact that the interests of household members are not always consistent with broader family goals. In addition, the studies have helped in drawing attention to the increased blurring of traditional dichotomies by showing that many people simultaneously engage in urban and rural life, work in both the formal and the informal sectors, or commute from the countryside to urban centers (or vice versa) on a daily basis. Livelihood studies have also raised the issue of spatial dispersion of household members that often accompanies livelihood diversification by highlighting the proliferation of multilocal livelihoods, including the rise of nontraditional household living arrangements, and transnational networks.

The Livelihood Approach: History and Definition

The modern conception of livelihood in development geography can be traced to classical French geography, especially its focus on human–land relations. According to de Haan, the livelihood system, called *genre de vie* (a system of livelihood strategies of a human group in a specific region) in early twentieth century French geography, had a spatial identity. Yet, a livelihood system was seen to result more from social reality and the force of habit than determined by the natural environment. The concept however completely disappeared from development geography after World War II as most scholars turned their attention to the dominant structural perspectives of the time – dependency and neo Marxist approaches. The reemergence of livelihood studies began with the introduction of the concept of survival strategies in the 1970s, which highlighted the active, productive role of poor people and recognized their behavior and actions as both logical and well informed. 'Survival strategies' were used in reference to poor people's coping strategies in the face of economic crisis and to stress the rationality of their risk minimization strategies in unpredictable economies. The concept of survival strategies and its application to the household as an entity attracted a lot of criticisms, including the charge that it downplays the structured constraints on individuals and households, and privileges the idea that the actions and decisions of the poor reflect strategic rationality. Despite the criticisms, the concept continued to be used because it provided a means of bridging the gap between social and individual levels of analysis. While many early survival strategies studies ended in rather pessimistic conclusions, showing how poor households were increasingly excluded and marginalized from the benefits of economic growth, a new generation of more optimistic studies that approached households from a livelihoods perspective began to emerge in the early 1990s and focused on how people are able to survive. While most of these livelihood studies continue to draw inspiration from its classical man–land relation roots, many focus more on sustainable livelihoods, while others attempt to broaden the definition of the concept.

The renewed interest in livelihoods is led by many international agencies and is driven by the need to develop more effective policies to alleviate poverty and address environmental problems. The advisory panel of the World Commission on Environment and Development (WCED) first proposed the concept of 'sustainable livelihood security' in its 1987 discussions on resource ownership, basic needs, and rural livelihood security. In the discussions, 'livelihood' was defined as adequate stocks and flows of food and cash to meet basic needs. 'Security' was used to refer to secure ownership of or access to resources and income earning activities, including reserves and assets to offset risk, ease shocks, and meet contingencies. 'Sustainable' was used in reference to the maintenance or enhancement of resource productivity on long term bases. Thus, sustainable livelihoods were seen as entailing both equity and sustainability. The 1992 United Nations Conference on Environment and Development (Rio Summit) also used sustainable livelihoods to link socioeconomic and environmental concerns. Both the WCED and UN forums were important for moving international concern with environmental issues to a focus on people and their livelihood activities, and were central in placing these concerns within a sustainable development policy framework. Based on their research on food security and agroecological sustainability, Robert Chambers (at the Institute of Development Studies, Sussex, UK) and Gordon Conway (at the International Institute for Environment and Development, London) also helped popularize the concept of 'sustainable livelihoods'.

By the late 1990s, many in the donor community and many international organizations had adopted the sustainable livelihood approach as an operational tool to aid their work on poverty reduction. The approach was seen as representing a shift in development practice from needs based, resource centered solutions to people centered, holistic, and dynamic approaches that focus on people and their capacity to initiate and sustain positive change. The wide embrace of the approach by development theorists and practitioners, governments, nongovernmental agencies, and international organizations helped to broaden its definition and focus. For instance, the United Nations Development Program (UNDP) places greater priority on the role of technology; Oxfam emphasizes the right to a sustainable livelihood; CARE focuses on livelihoods at the household level and pays more attention to security of livelihoods rather than sustainability; the Society for International Development (SID) in Rome's Sustainable Livelihoods Project focuses on how to increase the effectiveness of grassroots and other kinds of civil society organizations; 'sustainable livelihoods' work at the Institute of Development Studies in the UK focuses on the sustainable rural livelihoods frameworks; while the Overseas Development Group at the University of East Anglia concentrates on diversification of livelihood activities.

The UK Department of International Development's (DFID) efforts in making sustainable livelihood a major focus of the United Kingdom's international development policy deserve some attention here. DFID's definition of sustainable livelihood, adapted from Chambers and Conway's 1992 work, states that "A livelihood comprises the capabilities, assets (including both material and social resources) and activities required for a means of living. A livelihood is sustainable when it can cope with and recover from stresses and shocks and maintain or enhance its capabilities and assets both now and in the future, while not undermining the natural resource base." DFID's sustainable livelihood framework, which has been widely used, has four main components (**Figure 1**):

1. Vulnerability context: people are conceived of living within a vulnerability context in which they are exposed to risks, through sudden shocks and temporal changes.
2. Livelihood assets: people have a number of capital assets, which they draw upon to make their livelihoods. These include natural capital – nature's goods and services; social capital – social networks and relationships of trust that influence the cohesiveness of people and societies; human capital – the status of individuals, including skills, knowledge, and labor; physical capital – such as transport, shelter, water, energy, and communications; and financial capital – including savings, income, and credit. These five capital assets together form the 'asset pentagon' which is used to assess people's overall asset base.
3. Livelihood strategies: the range and combination of activities and choices that people make in order to achieve their livelihood goals.

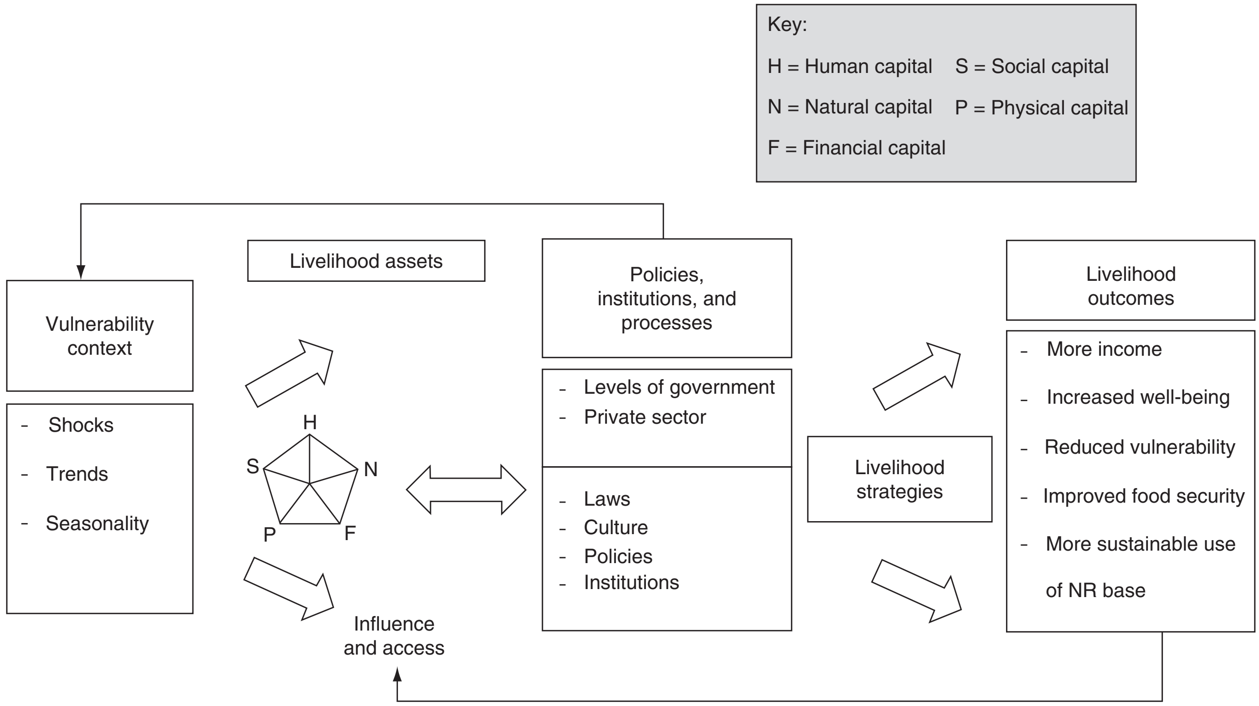

Figure 1 Sustainable livelihoods framework. Source: DFID Sustainable Livelihoods Guidance Sheets, Section 2.1 (available at Livelihoods Connect website: http://www.livelihoods.org/info/info guidancesheets.html).

4. Policies, institutions, and processes: these shape peo ple's access to assets and livelihood activities, as well as the vulnerability context in which they live. (Note the two way relationship between policies, insti tutions, and processes, on the one hand, and access to assets and the vulnerability context, on the other.)

The Unit of Analysis in Livelihood Studies

Development geographers have traditionally given much attention to the household. The household, defined as a social group which resides in the same place, shares the same meals, and makes joint or coordinated decisions over resource allocation and income pooling, is often con sidered the most important unit of analysis. In livelihood studies, increased attention was paid to household strat egies as a means of capturing the behavior of low income people. The household was considered as a single de cision making unit maximizing its welfare subject to a range of income earning opportunities and a set of re source constraints. The use of the household as the unit of study is considered useful for its potential to bridge the gap between microeconomics and its focus on the atom istic behavior of individuals and historical structuralism and its focus on the political economy of development. Critics, however, argue that rather than being a har monious entity that is pursuing an optimal balance, in dividual household members pursue individual ways to improve their situation. Therefore, it is important to recognize individual or intrahousehold level inequalities in which the well being and resource access of some household members especially women and children may be inferior to men; and also the inequalities of the broader levels of the extended family, the social group, and the community. Feminist scholars have criticized mainstream researchers for their failure to distinguish between indi vidual activities and household activities and for seeing everything as a livelihood strategy for all household members. They stress the importance of deconstructing the household decision making process, rather than as suming that such decisions reflect consensus. They therefore call for the abandonment of such a 'black box' view of household survival strategies, stressing that the individual and the household cannot be treated as iden tical and interchangeable, with individuals being per ceived as miniature households. In response to these criticisms, many livelihood studies focus on both the household and its individual members as units of study.

Livelihood Diversification

Livelihood diversification refers to the process by which households construct a diverse portfolio of activities and social support capabilities in order to survive and im prove their standard of living. This is based on the realization that a livelihood is more than just income; it encompasses income, both cash and in kind, as well as the social institutions (kin, family, compound, village, and so on), gender relations, and property rights required to support and sustain a given standard of living. Livelihood diversification also includes access to, and benefits derived from, social and public services provided by the state (education, health services, roads, water supplies, etc.). Livelihood diversification is therefore not synonymous with income diversification – the latter refers to the composition of household incomes at a given time (e.g., the income of a household at a given time may come from a combination of wages, profits from trading and sales of farm products) while the former refers to an active social process that allows households to engage in increasingly intricate portfolios of income generating activities over time (e.g., a member of the household may be employed in the formal sector, another may be self employed in the informal sector, and the entire household may also be involved in farming).

Livelihood diversification is not just a rural phenomenon, nor is it limited to developing countries. In fact, the practice has been found among urban dwellers in developing countries and in urban and farm families in the (post)industrial world. As agricultural price support systems and other subsidies in developed economies are removed, farmers are being compelled to diversify their sources of income. Similarly, the proliferation of part time and home based working patterns in the industrialized economies is evidence of livelihood diversification in urban areas. Although diversification in developing countries occurs for many different reasons, the underlying trends and processes include rural population growth that increases competition and diminishes access to productive assets such as land and hence declining returns to farming compared to other activities. The rationales for engaging in livelihood diversification include risk reduction; overcoming income instability caused by seasonality; improving food security; taking advantage of opportunities provided by nearby or more distant labor markets; generating cash in order to meet family objectives such as the education of children and, sometimes, the sheer necessity of survival following personal misfortune or natural and human disasters. Diversification is not just for the poor; rather, people of different socioeconomic backgrounds also engage in it but for different reasons. Thus, diversification may be for survival (as a safety valve for the poor), a means of ensuring security for those who feel vulnerable in the face of economic changes, or a means of accumulation for the not so poor. It may occur both as a deliberate household strategy and/or as an involuntary response to economic crisis.

Rural Livelihood Diversification

Contemporary rural development research and policy have shifted their focus from agrarian change to livelihood diversification. This shift is based on the realization that farming on its own rarely provides a sufficient means of survival in rural areas of low income countries. For this reason, most rural households depend on a diverse portfolio of activities and income sources in addition to crop and livestock production. Thus, rural livelihoods comprise one or several activities that variously provide food, cash, and other goods to satisfy a wide variety of human needs. Such activities may include cultivation, herding, hunting, gathering, and reciprocal or wage labor, trading and hawking, artisanal work (such as weaving and carving), processing, begging, and theft. A study of rural livelihoods shows that residents derive their incomes from multiple sources including 'farm income' (i.e., income from livestock and crops and comprises both consumption in kind of own farm output and cash income from the output sold); 'off farm income' (i.e., wage or exchange labor on other farms – within agriculture); and 'nonfarm income' (i.e., nonagricultural income sources, including nonfarm rural wage or salary employment, nonfarm rural self employment, rental income obtained from leasing land property, and remittances).

Urban Livelihood Diversification

Livelihood diversification has also been observed in urban areas. Although urban residents, especially in developing countries, have always diversified their sources of income, the economic crises that hit many developing countries in the 1970s and early 1980s and the subsequent adoption of neoliberal economic policies that resulted in the 'urbanization of poverty', have led to the proliferation of livelihood diversification among the urban poor as well as the not so poor segments of the population and among the informal as well as formal sector employees.

Life for many urban poor people in the developing world is a constant struggle to make a living. Whereas the rural poor can often produce most of their own food, urban residents generally depend on earned income to obtain food. Lack of access to agricultural land and other agricultural inputs often make it impossible for the urban poor to rely on urban agriculture. Thus, the urban poor make a living by severely reducing their consumption levels, or diversifying their livelihoods by finding ways to supplement their incomes or relying on social networks.

Similarly, public or private sector employees as well as private entrepreneurs have all found ways to diversify their sources of income, although the motivations for doing so vary. For instance, the increasing vulnerability of many public sector employees in developing countries due to a combination of factors, such as inadequate wages, threats of retrenchment, and the disappearance of guaranteed lifetime employment, have compelled them to diversify their sources of income. Private entrepreneurs also seek to diversify their sources of income in order to manage risk and to counteract economic shocks due to market fluctuations. However, lack of capital often limits the range of possibilities open to them as entrepreneurs.

The practice of multiple livelihood strategies among formal sector employees in many developing countries also deserves some attention. In many developing countries, the combined effect of the economic reforms of the 1980s and 1990s and the attendant increases in urban poverty have compelled many formal sector employees, either by necessity or choice, to join the informal sector in an effort to increase their income earning opportunities and diversify their sources of income. Studies have shown that people from the middle and professional classes that were previously assumed to be immune to the pressures of economic change and therefore lack the need to diversify their livelihood options now also rely on multiple sources of income. For instance, many formal sector workers supplement their incomes with part time informal sector jobs, such as taxi driving or petty trading. Other members of the formal sector employee's households may also supplement the family income by keeping gardens, selling prepared food, or trading. Such multiple livelihood strategies have become 'the way of doing things' in many cities, turning the informal sector into an avenue for 'part time' employment for formal sector employees and a source of additional income for many with full time employment in the formal sector.

Livelihood Diversification and Spatial Dispersion

Another trend linked to livelihood diversification, is the proliferation of multilocal livelihoods, including non traditional household living arrangements and transnational networks. Livelihood studies have helped highlight the limitations of the traditional view of the household by documenting the proliferation and the complexities of household arrangements. For instance, studies have shown that livelihood diversification often requires a more spatially extended understanding of the household. Spatial dispersion of household members allows them to take advantage of economic opportunities in multiple rural and urban places or multiple countries. For instance, 'split families' or 'divided households' are used to refer to rural households whose members work in urban centers or abroad, and the members are said to employ livelihood strategies that 'straddle' the rural and urban centers. Similarly, some urban households have 'dual or multiple residency' arrangements, in the sense that the household may have an urban home and a rural home close to the urban area to allow the members to spend most of their time in the urban home, but temporarily move to the rural residence when the need arises, for example, during harvest. Equally important to the understanding of the complexity of livelihood diversification and household arrangements is the role of non resident family household members to contributie to well being. These include different forms of migrants – urban migrants, circular migrants, seasonal migrants, etc. Migration seems to be a successful strategy, especially for those families with relatives working abroad because they are able to remit more money than internal migrants. Internal migrants often face a difficult time, because their incomes are unstable. As a result they struggle to survive and leave their families insecure. Unlike the international migrant group, this group is seldom able to escape from poverty and is unable to durably improve livelihoods by investing money in the local economy. These trends support emerging views in development geography that challenge the idea that regions are clearly demarcated and people are rooted in or tied to localities.

See also: Civil Society; Dependency; Informal Sector; Migration; Neoliberalism; Neoliberalism and Development; Nongovernmental Organizations; Poverty; Poverty, Rural; Remittances; Rio Summit; Sustainable Development; Vulnerability.

Further Reading

Carney, D. (ed.) (1988). *Sustainable Rural Livelihoods*: *What Contributions Can We Make?* Nottingham: Russell Press Ltd., for Department for International Development.

Carney, D. (1999). *Sustainable Livelihood Approaches Compared*. London: Department for International Development.

Chambers, R. (1995). *Poverty and Livelihoods: Whose Reality Counts?* Brighton: IDS.

Chambers, R. and Gordon, R. C. (1992). *Sustainable Rural Livelihoods*: *Practical Concepts for the 21st Century* University of Sussex, Institute of Development Studies, Working Paper No. 296, Brighton: IDS.

de Haan, L. and Zoomers, A. (2003). Development geography at the crossroads of livelihood and globalisation. *Tijdschrift voor Economische en Sociale Geografie* 94(3), 350 362.

Ellis, F. (1998). Survey article: Household strategies and rural livelihood diversification. *The Journal of Development Studies* 35(1), 1 38.

Ellis, F. (2000). *Rural Livelihoods and Diversity in Developing Countries*. Oxford: Oxford University Press.

Farrington, J., Carney, D., Ashley, C. and Turton, C. (1999). *Sustainable Livelihoods in Practice: Early Applications of Concepts in Rural Areas*. Natural Resources Perspectives 42. London: Overseas Development Institute.

Owusu, F. (2007). Conceptualizing livelihood strategies in African cities: Planning and development implications of multiple livelihood strategies. *Journal of Planning Education and Research* 26(4), 450 465.

Scoones, I. (1998). *Sustainable Rural Livelihoods*: *A Framework for Analysis*. University of Sussex, Institute of Development Studies, Working Paper No. 72, Brighton: IDS.

Relevant Websites

http://www.care international.org
CARE nonpolitical and nonsectarian humanitarian organization

dedicated to fighting global poverty. It operates in many African, Asian, Latin American, Middle Eastern, and Eastern European countries.

http://www.dfid.gov.uk
Department of International Development (DFID) part of the United Kingdom government that manages Britain's aid to poor countries.

http://www.livelihoods.org
Institute of Development Studies (IDS), University of Sussex, devoted to sustainable development issues, including documents, information on current events, links to organizations, and feedback on practical actions.

http://www.iisd.org
International Institute for Sustainable Development. IISD champions sustainable development around the world through innovation, partnerships, research, and communications.

http://www1.uea.ac.uk
Overseas Development Group a charitable company of the University of East Anglia that manages the research, training, and consultancy activities undertaken by the faculty members of the University's School of Development Studies.

http://www.oxfam.org
Oxfam International an international confederation of 13 independent nongovernment organizations dedicated to fighting poverty and related injustice around the world.

http://www.sidint.org
Society for International Development (SID) a global network of individuals and institutions concerned with development.

http://www.undp.org
United Nations Development Program (UNDP) a United Nations organization that focuses on finding solutions to global and national development challenges especially in poor countries.

Local Development

G. Garofoli, Insubria University, Varese, Italy

Glossary

Endogenous Development Development model based on the social capability of local (and national) actors which allows both valorization of local resources, development of productive interdependences, and control of transformation of social and economic system through accumulation and innovation processes.

External Economies The agglomeration of firms and industries reduces transaction costs and the presence of education and research facilities, of financial institutions, and efficient labor markets allow the access to strategic and specific assets; the interactions among different actors and institutions allow the production of new knowledge and professional competences which are available for local firms.

Growth Pole Model Model of industrialization (introduced by François Perroux) based on the crucial role of large innovative firms in propulsive industries.

Industrial District Model of industrialization based on networks of small- and medium-sized firms, with division of labor among firms and productive linkages. The balance of competition and cooperation among firms allows efficiency of small firms (through specialization) and their innovation orientation.

Linkages Backward and forward linkages have been introduced by Albert Hirschman to stress the opportunity for two sequences of productive interdependences (toward production of material inputs and equipment, on one side, and the transformation of production toward final markets, on the other) of the industrialization process, generating special pressure toward investments.

Local Productive System Productive organization model based on the use and accumulation of 'specific' resources which cannot be transferred in other areas; model based on the strict interaction between economy, society, and territory.

Milieu Innovateur Term and concept introduced by the French-speaking 'Groupe de Recherche Européen sur les Milieux Innovateurs' (Gremi); the *milieu innovateur* suggests that innovation is the product of spatial proximity and thick networks among innovative firms and social and institutional appendages.

The Starting Point

The rise of local development analyses and policies is mostly based upon two main phenomena: (1) the failure of traditional schemes of regional development (generally based on the attraction of large external firms and 'top down' strategies managed by national institutions), and (2) the starting up of new 'bottom up' development experiences and new patterns of local and regional development. Moreover, it is necessary to remember the role of European programs (cf. especially, the Leader program in rural areas) which has supported local development initiatives since the end of the 1980s.

Regional development schemes based on Perroux's growth pole model, characterized by the use of huge financial and fiscal incentives to external large firms and the idea of imitation and reproduction of the development conditions and organizations already experienced in developed regions, collapsed during the 1970s. This was due both to the failure of development goals (for lack of local economic integration and multiplier effects which determined the construction of the so called 'cathedrals in the desert') and to the fiscal crisis of the central state, which was no longer able to offer incentives to large firms in order to de localize their plants in depressed regions.

Moreover, the development of new industrial regions in peripheral areas, with different production organizations compared to earlier industrialized regions and countries, caused trouble to the center–periphery schemes and called for new interpretations.

Novelties

The emergence of new industrial regions and their better economic performances modified the previous hierarchy among regions, both at the international and national levels. The economic crisis and the de industrialization of some of the first comers – industrial and well developed regions – were coupled with the rise of new regions within the international division of labor and trade relationships.

This novelty was neither the consequence of capital mobility from developed regions nor the replication of previous models of production organizations. In other words, this process was not the consequence of simplistic regional reorganization due to market forces which reacted to different relative prices. On the contrary, it implied both the role of social and cultural variables and new forms of social regulation. The new industrial regions are usually the consequence of a dialectic process between market opportunities, mobilization of existing (often unused or badly used) resources, knowledge and learning, and new forms of production organization.

The local development approach recognizes, in fact, both the opportunity for different paths of development (e.g., based on small firms and systems of small firms, like industrial districts) and the role of local actors in managing the use of their resources, through the introduction of specific strategies.

This leads to a plurality of local models of development (i.e., opportunity for different strategies of local and regional development and for different economic policies using a wide set of tools) and allows to take into account the subjectivity of the local actors (i.e., individuals, firms, collective private organizations, intermediate institutions, public institutions, and the local and regional state) in their behavior and decision making process in relation to investments and long term interventions.

Local development cases were usually induced through spontaneous organization (this specifically implies 'bottom up' processes) but, at the same time, often benefited from the capability of introducing specific forms of social regulation and models of economic coordination.

The scale of local development – both for the analysis and for the construction of local strategies – could vary in consideration of different economic and social features (including density of population) and historical background of different localities. Practically, the definition of scale depends on the internal homogeneity and identity – from an economic, social, and cultural perspective – and external differentiation, in comparison with external areas. It depends, then, not only on structural characteristics (which could be measured with statistical indicators) but also on subjective perceptions of local actors. The definitions of local strategies are, in fact, very often decided on a voluntary basis which includes even the geographical definition of locality boundaries. Empirically, local development analyses and strategies refer to an intermediate scale between the municipal level and the administrative regional level. When the great urban areas are excluded, local development strategies are introduced in areas with a population range between 30 000 and 200 000 people.

A New Paradigm

Local development introduces a new paradigm in economic and social regulation as territory becomes a new regulatory bridge between the state and the market. Space has ceased to be a source of costs for firms, and instead, is assuming the role of favorable (or nonfavorable) environment, the creator of external economies (or diseconomies). In fact, territory plays an active role transforming a geographical space into a network of social relations and interactive actions producing new knowledge and competences. Territory works, then, as a meeting point of economic organizations (cf. the division of labor among firms and innovation trajectories), social regulation, and institutional governance. This explains why territory produces and takes into account external economies for individual firms, creating public goods through (both implicit and explicit) collective actions.

This means resources can be different and not standardized, strategies and actions can differ, pathways to development are different and open: territory matters. Regions and localities are not only portions of national systems but are also effective media of social and economic interactions among actors, interests, opportunities to singling out mobilization of resources, solving problems, producing decision making, creating firms, and organizing economic activities. The national system (its gross domestic product (GDP) level, its economic structure, and competitiveness) is just the outcome of the local and regional economic organization.

There are subjective and deliberated actions of local actors, very often collective actions accompany firms' strategies. This means, there exists a system logic and opportunity for intervention of the local state. Voluntarism and awareness about goals, opportunities, and threats are crucial elements in the governance of the transformation process.

There is space for economic development everywhere, even if past stories are important, producing some effects of 'path dependency'; in any case, the coherence of the local strategies and the capability of local actors, and their responsibility in finding solutions, are giving opportunities for socioeconomic voluntarism which is in opposition to geographical determinism.

The Theoretical Background

Local development analysis has been strengthened after the success stories and the literature on industrial districts (starting from the writings of Giacomo Becattini), which underlined the character of the model of productive organization that is opposite to that of the big company and 'development from above'. This would be a model based on dense clusters of small and medium sized enterprises (SMEs), with strong productive specialization (both at the local and firm's level), intense productive linkages which produce high division of labor among local firms, great capability in problem solving and especially, progressive production of new knowledge and professional competencies which determine a strong orientation to innovation and high quality products. Industrial districts, in their dynamic functioning mechanism, can produce dynamic competitive advantages which are embedded on the territory.

Piore and Sabel (in their book *The Second Industrial Divide,* published in 1984) draw strongly on the Italian example of industrial districts to oppose the model of flexible specialization to the Fordist model, by stressing the role of the historical alternative to mass production. Piore and Sabel have the great merit of breaking with the orthodox theoretical schemes. Scott and Storper also pay particular attention to the experience of industrial dis tricts in analyzing the changed relations between econ omy and territory.

The literature on industrial districts showed that small firms (when organized in networks and territorial or ganizations) could be effective (reaching high labor productivity, comparable to the values obtained in large firms) and innovative, breaking with traditional inter pretative schemes based on a simplistic (and determin istic) view of the principle of scale economies.

The model of industrial district denies the validity not only of a series of theoretical assumptions generally (and often uncritically) accepted, such as those concerning the lack of economic efficiency and innovation in small firms, but also of the rationale of regional development policies (and of development *tout court*) followed in both de veloped countries and developing countries.

Industrial districts disclosed (both theoretically and empirically) opportunities for alternative paths of de velopment (even driven by different sectors, not neces sarily industrial sectors) which could be strongly based on specific assets of the territory. Territory matters and local development are important.

The analysis of the productive organization of the industrial district and its social factors allows us to shed light on important new variables in the decisions (on lo calization, investment, and strategies) of economic actors that condition the processes of transformation of the local economy (and society) and consequently of the regional and national economy. Relations of cooperation among firms, relations between the system of production and the socioinstitutional system, the competencies and involve ment of workers in the productive organization of the firm and in the wider social model of the area, the role of specific local institutions that intervene to overcome de ficiencies (or 'failures') of the market (such as technolo gical centers, services centers, training schools, and local development agencies) become crucial. In other words, a social system of interrelations, of circulation of infor mation, and of production and reproduction of values organizes itself to permeate and characterize the mode of production. This means that many crucial factors are historically embedded in the local society and are there fore not easily transferable to other areas: the process of development acquires its definitive character as a 'social process' by refusing to appear only as a technical one.

In this way, territory plays an active role in the process of development as it includes all those historical, cultural, and social factors that are the basis of specific models of productive organization, of the continuous interaction among economic and social actors and, therefore, of the actual processes of economic and social transformation. Therefore, the emphasis on the territorially anchored organizational model of production hinges upon at least two dimensions:

1. the system of production, the relations among firms, and, therefore, the social division of labor (aspects such as ways of exchange of goods and services, transaction costs, production costs, increasing returns, dependence, or relative autonomy of supplier firms); and
2. the social and institutional base which allows that specific organizational form (such as worker consensus and participation; coherence of the strategies of spe cific local institutions; local policies of support and stimulus; reproduction of knowledge and values; and social investments necessary to reproduce external economies).

Local Productive System and Endogenous Development

The analysis of industrial districts opens the debate on crucial items which characterize both local economic and social structure and competitive factors within national and international division of labor. Industrial districts open, then, the general question of local development trajectories and the opportunity for alternative paths of development, including models without (or with marginal presence of) industrial production. Nowadays, a large literature exists on local development models based on agriculture (even with the so called 'agro industry districts'), tourism (with several different models of production organization and with different degrees of intersectoral integration), cultural activities (*cf.* the so called 'cultural districts'), environ mental management, and so on.

Moving ahead toward the functioning mechanism of these alternative paths and introducing the discussion on the autonomy's degree of local actors and institutions in the process of economic and social transformation, we can understand the role played by the concepts of local productive systems (LPSs) and endogenous development. These concepts are rather crucial both from an analytical perspective and from the point of view of development policies.

The goals for local development strategies and policies are mainly the use, valorization, and implementation of local resources (often specific resources, different from standard ones which are predominant in mainstream ap proaches) and the construction of dynamic competitive advantages (to maintain the sustainability and durability of economic and social organization) through the control of the accumulation process and innovation.

The approach on endogenous development stems from some seminal works, during the 1970s and beginning of

1980s, that allow a paradigm of the development 'from below' to emerge and, additionally, the approach connects itself to the literature on territorial and 'agropolitan' development and on eco development.

According to the recent literature on this theme, a model of endogenous development guarantees autonomy to the process of transformation of the local economic system, generally assuming the traits of a self sustaining process of development. The process of transformation is based, therefore, on local specificities and on the capacity to govern fundamental variables.

Economic development is always simultaneously a process of territorial restructuring, that is, it assumes, in many different shapes and forms, a geographic complexion. One of the most important expressions of this link between territory and development consists of dense regional agglomerations of firms and workers.

Three major conditions underlie this link, namely: (1) the existence of specific resources (and assets) that cannot be easily transferred elsewhere; (2) the emergence of a system logic that binds firms, collective actors, institutions, and workers together into a functioning economic order; and (3) the formation of project capacity, that is, an ability to deal with internal crises and react to external challenges, which requires, in turn, the establishment of coherent mechanisms of local economic governance.

We can describe this kind of territorial development in terms of the theory of LPSs. The literature on LPS initially stems from the analysis of successful industrial districts in economically advanced countries in the 1970s and 1980s, but it has continued to evolve in various directions, including an important step forward into issues of endogenous development.

The concept of LPSs can be generalized to any organizational model of economic activity rooted in geographic space and is based on the presence of external economies, specific resources, tacit knowledge, and mechanisms of social regulation. In this kind of system, the local milieu has a crucial relation to a number of interrelated variables, namely:

1. Production organization, that is, the operational structure of the local economic system as reflected in rules and modes of management that allow for 'division of labor' among firms. This, in turn, fosters the formation of productive linkages, social relationships, and cooperation rules. Where trust and loyalty prevail, costs of interfirm transaction are significantly reduced.
2. Professional skills and competencies are reproduced in the local area, both formally and informally. Social interaction and learning foster the acquisition of skills, and progressive learning stimulates new interests and abilities.
3. Diffusion of knowledge and information (about productive and managerial techniques, market outlets, local resources, competencies, and so on) wherein rules of competition and cooperation shape the diffusion of knowledge, on an involuntary basis in the first instance, on a voluntary basis in the second. Knowledge about local business conditions becomes a common heritage, a true public good.
4. Structures of social regulation help to solve common problems in the LPS and to negotiate points of political tension. These structures assume many different guises: governmental agencies, intermediate institutions, civil associations, private–public partnerships, and so on.

These points lead to the related issue of endogenous development, that is, the notion that the LPS possesses a degree of autonomy in regard to its internal structure and evolutionary course. This degree of autonomy derives from the decision making capacities of individual and institutional actors in the local area, and from their ability to control and internalize flows of knowledge and information.

A model of endogenous development is, in fact, based on the formation of 'social capability' at the level of the community of firms and institutions operating in the local sphere, through the progressive construction of a series of critical regional assets. Four such sets of assets (introduced by Garofoli in the book *Endogenous Development and Southern Europe,* published in 1992) are of special importance in the present discussion:

1. the first revolves around the valorization (i.e., use and implementation) of local resources, such as human resources, financial resources historically accumulated at the local level, material endowments, entrepreneurship, professional competencies, technological know how, and so on;
2. the second is concerned with the existence of (and ability to develop) interfirm productive interdependences (i.e., production of linkages effects) in the local area;
3. the third involves control of dynamic learning and innovation processes (i.e., reduction of X inefficiency and capability to introduce new products); and
4. the fourth is focused on the capacity to guide the process of accumulation at the local level, transforming financial resources into productive investments.

To be sure, endogenous regional development does not imply regional closure or imperviousness to external influences. On the contrary, it goes hand in hand with the insertion of the LPS in a wider economic environment (which implies progressive relations with the external) with multiple local impacts (new technologies, market shifts, national legislation, and so on). Endogenous development, in fact, means capacity to transform the social and economic system; capacity to react to external challenges; and capacity to introduce specific forms of social regulation, at the local level, that support these two processes. Endogenous development, in synthesis, is the capability to valorize local specific resources and knowledge and to internalize external knowledge; it is the capacity to innovate and produce 'collective intelligence' at the local level.

Endogenous development, in other words, refers simply to those elements of the productive system that, by reason of their collective order and mutual synergies, possess certain powers of local social choice and self determination.

Local Development and Innovation

This argument becomes clearer when we turn our attention to the role of innovation in development. Innovation is, at least in part, territorial in nature by reason of the interactive mechanisms and learning processes that drive it forward, and that themselves have a distinctive spatial expression. This general point can be found in a variety of theoretical contributions from the growth pole concept of Perroux (with its emphasis on lead firms as sources of new product innovation) to the more recent idea of regional innovation systems as proposed by analysts like Asheim and Gaffard. In the latter perspective, technological innovation is based on in territorially specific processes through the interaction of individuals and institutions in the course of everyday work and life. Territorial economic innovation and development are also, of necessity, path dependent as clarified by the evolutionary theory of innovation.

Two further distinctive lines of analysis have developed around the issue of the relations between innovation and territorial development. These can be identified in terms of (1) the technological district (introduced by Antonelli in 1986), and (2) the *milieu innovateur* as formulated by the *Groupe de Recherche Européen sur les Milieux Innovateurs* (GREMI). In the technological district approach, the process of technological change is underpinned by the existence of dense interactions between firms located in close proximity to one another. The concept of the *milieu innovateur* suggests that innovation is the product of an environment made up of the LPS and its various social and institutional appendages. The notions of the technological district and *milieu innovateur* are, of course, very close to one another. They share an evolutionary and nonlinear conception of technology and development; and they explicitly recognize the effects of spatial proximity in processes of technological innovation and the importance of externalities.

These remarks now bring us full circle back to the question of clustering, bearing in mind that "in a world of learning economies the specialization of firms and countries becomes increasingly important for economic performance" (*cf.* Lundvall and Johnson's book, *The Learning Economy* in 1994). In short, the clustering and specialization of firms are critical foundations of competitiveness and innovation, and the region is a true nexus of economic development. We must not forget, however, that LPSs are also sites of widespread, sometimes massive, market failure, so that collective action is called for as means of dealing with the consequent problems of inefficiency and misallocation. Much market failure in these systems consists of suboptimal information flows, inadequate supply of skills and competencies, and underprovision of certain critical services in circumstances where requisite scale thresholds are not satisfied by local conditions. Examples of this latter condition can be found in deficiencies regarding knowledge about distant or emerging markets, failures in the transfer of technological know how, or low levels of access to financial resources for SMEs.

Problems like these call for different forms of governmental action and institution building in civil society. These instances of collective action are in turn subject to processes of 'institutional and regional' learning, including the so called experimental regionalism proposed by Sabel in 1995. In the latter case, regions function like laboratories in which social experiments based on 'learning by doing' and 'learning by monitoring' are carried forward to various stages of completion by different kinds of corporate bodies.

The False Dichotomies in Local Development

Some dichotomies are often used in debate and discussions, just to put in opposition different perspectives, strategies, and policy orientation; it is possible to think about the following ones:

1. local and global;
2. competition and cooperation;
3. state and market; and
4. identity and (external) openness.

The literature and case studies on industrial districts showed that the opposition between local and global does not exist at all. The districts' firms are used to valorize specific resources and external economies of the territory combining them with international market opportunities: without external outlet there could not exist any opportunity for division of labor among firms and productive specialization.

At the same time, the balance between competition (which favors economic efficiency) and cooperation among firms, especially among complementary firms (which allows division of labor, problems solving, and continuous accumulation of knowledge and competencies) represents a structural characteristic of the industrial districts.

The capability to introduce specific forms of regu lation and collective actions, mainly as a reaction to market failures, represents a step forward to the eco nomic and social transformation of a local economy, giving new opportunities for local competitiveness and innovation. Often the introduction of intermediate in stitutions, introducing rules (*cf.* consortia among firms) and new competencies (*cf.* services and techno logical centers), is able to create conditions for market organization. This means collective action, often with the support of local and regional state, can foster market mechanisms. Collective actions and intermediate institutions can then create conditions for market effi ciency, clarifying that the market is often a social con struction. In conclusion, state and market are not in opposition.

Similarly, without identity of the local economic and social system (*cf.* common problems for local firms and local community) it is not possible to perceive oppor tunities in solving problems and to guarantee conditions for local policy intervention. But without the inter national comparison, the access to external information and knowledge, and the knowledge of external conditions of production and external markets, no room exists for effective local policies.

In conclusion, the dynamic dialectics and the interaction among the different couples of apparent dichotomies are able to determine transformation and development. Local development realization does not exist without this kind of interaction: all success stories of local development dem onstrate the existence of this dialectics.

The Typology of Local Productive Systems

At the end of the 1980s and beginning of 1990s, the literature faced the question of typologies of local models of development, and scholars sought to identify diverse typologies of LPSs. The objective of the discussion on typologies of development is not the exhaustive elabor ation of all possible models of local development. Rather, it is to allow a plurality of development models to emerge and reveal the potential bifurcations of processes of transformation and the roles (and responsibilities) of local social actors in controlling these transformations of local social and economic systems. This underlines the key role of forms of social regulation that are introduced (and that can be introduced) at the local level, and of local development policies.

Storper and Harrison (in *Research Policy*, in 1991) worked around the concepts of 'core' and 'ring' with four main typologies: all ring, no core; core–ring, with coordinating firm; core–ring, with lead firm; and all core, no ring, whereas Ann Markusen (in *Economic Geography*, in 1996) draws attention to 'sticky places' with four typologies: 'in dustrial districts', 'hub and spoke district', 'satellite plat form district', and 'state anchored district'. Garofoli (in *Industrializzazione diffusa in Lombardia*, in 1983) and Leborgne and Lipietz (in *Les régions qui gagnent*, in 1992) worked around three typologies of agglomerations of firms which, after the recent rediscovery of the question in recent years, especially in the discussion on clusters of small firms in less developed countries (LDCs), could be summarized with the following proposal:

1. clusters of small firms, specialized areas with static competitive advantages (agglomeration of specialized suppliers in areas with low wages) and weak ties among firms and lack of control over final markets;
2. LPSs, with division of labor among firms and con tinuous production of professional competencies; interaction and learning become strategic factors and allow a progressive shift toward the introduction of dynamic competitive advantages; and
3. system areas, with effective territorial integration network and presence of complementary firms which allow the introduction of new knowledge and innovation; voluntary decision making and partner ships for transformation and innovation; introduction of collective actions; and effective local state inter vention are able to produce dynamic competitive advantages.

In the last typology it is possible to include a lot of different areas which follow the 'high road to develop ment', like innovative industrial districts, local innovative systems, and technological poles.

The discussion on typologies is very important for the identification of strategies for local development, especially for the recent debate on LDCs. This kind of approach, in fact, could foster all the reflections on au tonomy of local economy, the upgrading paths and op portunities, the employability of labor force, innovation introduction, and self sustaining development processes. This could, then, help to single out goals and tools to launch effective strategies of local development.

The Role of the Territory: A Regulatory Tool

The impact of local development analysis and policies is crucial even in the light of more general issues, particu larly as they show that territory and the mesoeconomic level become strategic assets for development strategies at the national scale. Coordination and capability to mobilize knowledge, professional and technological competencies, decision making processes, and actors' behavior for eco nomic development in LDCs strictly depend on this kind of approach, which gives the opportunity to transform general economic development objectives into strategic actions, involving different actors and institutions (both private and public). The lack of capability in economic development strategies in LDCs is generally due to the incoherence between national goals and tools and media introduced to mobilize resources and competencies into an effective economic process, with induced and multi plier effects as a consequence of a plurality of connected decision making and investments.

All this means both a coherence at the horizontal level (among different local actors) and a coherence at the ver tical level (among different levels of government; i.e., along the institutional *filière*). This, perhaps, explains the success of the local (and endogenous) development approach in sev eral LDCs.

The success stories of local development are numer ous and show very different economic structures and different strategic items and tools in solving problems, launching local initiatives, and transforming economic and social processes. Different cases of industrial districts can be found, especially in Southern European countries, and some interesting cases of technological poles and of transformation and improvement of the quality of life in several rural areas can be found, even in LDCs. The huge number of analyses on the role of clusters and networks in emerging countries and the increasing introduction of development strategies in these countries (especially in Latin American countries and in the other continents) explain clearly the relevance of local development even for achieving effective results on more general objectives of economic development.

In all success case studies, it is often possible to identify a logic of planning of territorial integration and a very intense orientation to the use and valorization of local (mainly human) resources. In light of the above, the idea of a diffusion and multiplication of 'good practices' to solve economic and social problems in other areas is gathering strength. At the same time, it is important to have a correct perception of the different tools and schemes of development strategies and policies, looking at some successful European schemes (such as Leader, Urban, and Interreg programs) till new ones are intro duced in some countries (like territorial pacts and ter ritorial integrated programs) that rediscover some seminal concepts of the literature on local development.

Local Development Policies and Strategies: Some Guidelines

The main objectives for local development (and also for local development agencies) are the improvement of the existing firms (i.e., the increase in labor productivity, the improvement of the used human resources, the im provement in the strategic capability of firms, and a better position within national and international markets) and the formation of new enterprises (to improve the firms' density, increasing the diffusion of knowledge and the imitation of success stories, the complementary linkages among firms and the total employment, i.e., better use of existing resources).

This could clarify why, in spite of the increasing role of local development policies introduced at the national and supranational (*cf.* the European schemes, IADB, UNDP, and other international organizations schemes) level, the results in promoting effective development have often been weak, sometimes inducing failures of local develop ment initiatives. Very often, in fact, the mobilization of local actors and institutions was led looking for financial support rather than looking for the solution of crucial local economic and social problems, through their involvement and the construction of capabilities' empowerment.

This suggests the importance, first of all, of the cap ability to single out goals and opportunities for the local economy and society, that means capability to organize national and international comparisons to understand the relative position of the localities within the arena of the international division of labor.

The second necessary condition to launch effective local development strategies is the capability to animate local actors and mobilize resources, mainly the unused or badly used resources, as already underlined by Hirsch man in his seminal work in 1958.

The concept of 'LPS' is crucial here, as it introduces a system logic within the interaction of a large quantity of private economic actors and collective and public parties, thus helping to design development trajectories, alter native strategies, and transformation paths. Very often, the introduction of procedures of strategic planning, which can even assume the forms of decentralized social negotiation, could help the effectiveness of local devel opment strategies.

A few final comments are presented as a sort of methodological guideline. Earlier in the article, we have indirectly used some crucial concepts of development theory and strategies, which are useful to clarify that an effective local development process is based upon five different conditions:

1. the existence (and continuous production) of Hirschman's backward and forward linkages effects (complementary relationships and increasing division of labor are, then, crucial variables);
2. the organization of territorial integration (this will pro duce cumulative effects through multiplier mechanisms);
3. the achieving of a critical mass of economic activities (the local version of the Rosenstein–Rodan's 'big push;), which emphasizes the role of agglomerations;
4. the coordination and capability to introduce problem solving interventions, through collective actions; and

5. the mobilization of sequential and coordinated investments which are able to modify the profitability and feasibility of different investment decisions.

In conclusion, local development strategies and policies are a crucial tool for valorization of existing resources and reaching collective goals even at the national scale. Without the capability to mobilize resources and capabilities and without the coordination of different actors, at the mesoeconomic level, it could not be possible to obtain high social and economic performances at the national level. This is to say that national economic competitiveness and quality of the life at the national level depend on the capability to realize effective local development strategies in the best part of national territories.

See also: Agglomeration; Industrial Districts; Local Economic Development; Learning Regions; Networks; Regional Development, Endogenous; Regional Innovation Systems.

Further Reading

Becattini, G. (1998). *Distretti industriali e made in Italy*. Torino: Bollati Boringhieri.
Cooke, P. and Morgan, K. (1998). *The Associational Economy. Firms, Regions, and Innovation*. Oxford: Oxford University Press.
Courlet, C. (2001). *Territoires et Régions*. Paris: L'Harmattan.
Garofoli, G. (2002). Local development in Europe: Theoretical models and international comparisons. *European Urban and Regional Studies* 9(3), 225 239.
Schmitz, H. (ed.) (2004). *Local Enterprises in the Global Economy: Issues of Governance and Upgrading*. Cheltenham: Edward Elgar.
Scott, A.J. (1998). Regions and the World Economy. Oxford: Oxford University Press.
Scott, A. J. and Garofoli, G. (2007). The regional question in economic development. In Scott, A. J. & Garofoli, G. (eds.) *Development on the Ground: Clusters, Networks and Regions in Emerging Countries*. New York: Routledge.
Vazquez Barquero, A. (2002). *Endogenous Development*. New York: Routledge.

Local Economic Development

S. Conti P. Giaccaria, Università e Politecnico di Torino, Turin, Italy

Glossary

Competitiveness The capability of build and maintain an advantage versus other actors who share, at least partly, some of the aims and objectives of a given actor.

Human Development Index A synthetic measure of development, defined by UNDP, in order to take into account not only economic growth, but also the satisfaction of basic human needs (health, life expectancy, education, human right, etc.).

Local System A consistent set of actors (firms, local authorities, association, universities, etc.) and relationships, which produce some emergent properties (i.e., properties that are not belonging to the individual components), such as collective agency.

Methodological Individualism The belief that also complex organizations, such as the economy or the society, can be understood by analyzing a single representative agent, which is assumed to possess all the characters needed for an exhaustive comprehension of the organization itself.

Reification of the Territory The tendency to treat a given territory as a given, homogeneous, cohese system rather than a socially and constructed entity.

Territorial Competitiveness The capability of a territory to compete against each other. It can be either direct (territories competing to attract flows of events, residents, FDI, etc.) or indirect (territories sustaining the firms' competitiveness).

Transaction Costs According to Williamson's account, when a firm buy products and services on the market rather than making them internally, it faces a broad range of costs, which can vary from research costs to losses originated by supplier's mistrust and free-riding behaviors.

Introduction

It is no chance, in the context of contemporary economic and industrial geography, that one of the fundamental terms, both in research and praxis, is local economic development (LED), a synthetic concept that underlies a multitude of other terms around which there has been heated debate, such as industrial districts, industrial clusters, learning regions, innovative milieu, local production systems, regional motors and so on. In very general terms, we might define LED as a broad and heteroclite set of concepts and theories that share the common fundamental assumption that economic development processes do not depend on a unique and universal model of growth but rely on development paths which are historically and, more importantly, geographically bounded.

Against the background of growing awareness of the incapacity of the traditional models of analysis of regional development, LED aims to give meaning to the central role in contemporary development processes of an intermediate entity between the actor (the company, in particular) and the economic system as a whole. This intermediate entity has been labeled in several ways, sometimes with reference to the organization of production – cluster, district, innovation system – some other times with reference to its geographical scale – localities, local system. Despite the fact that such a great number of streams, schools and interpretations makes difficult to highlight a consistent and complete theoretical framework, most of the approaches to LED share at least two fundamental features: the role of relationality in supporting economic competitiveness and the emphasis on the qualitative rather than quantitative dimensions of development.

LED as Relational Process

As far as the issue of relationality is concerned, we can observe how LED theories entail skepticism about the possibility of grounding geographical and economic analysis on the behavior of the single, representative, rational, individual firm (i.e., a refusal of methodological individualism). The idea that the competition process is not under the complete control of the individual firm leads to highlight the role of inter firms relationships in supporting firm's competitiveness.

In the earliest accounts of LED, such emphasis on relationality concerns mainly economic relationships, such as input–output links and spatial division on labor to the local scale. On the theoretical ground, the reflection of the Californian school of Economic Geography (with Storper and Scott), grounded on Williamson's transaction costs and Piore and Sabel's flexible specialization, played a fundamental function in demonstrating how small and medium enterprises might achieve productivity and overcome their dimensional limits by specializing in few activities and by trading with other SMEs. On the more empirical side, Becattini's (re)discovery of Italian industrial districts and subsequent Porter's generalization through the concept of cluster drove to a renewed reflection on externalities and agglomeration economies.

Step by step, the interest in the relational dimension of the economy became wider and wider. More precisely, the focus shifted from mere economic relationality to a more social and cultural understanding of relationality, taking into account the issues of noneconomic linkages between firms (such as trust, institutions, and shared values) and of ties between firms and noneconomic actors (such as public administrations, universities, and associations). The idea that noneconomic relationships play a pivotal role in addressing LED processes was already explicit in the industrial district literature – for example, consider the Marshallian concept of 'industrial atmosphere' or the presence of noneconomic regulation in setting the price of intermediate goods within the district boundaries. The important feature, here, is that, since the 1990s, the emphasis on the noneconomic relationality became predominant. To establish collaboration between a company and a research center, or to set an exchange of information between the customer's and the supplier's technicians, relationships need to have characteristics once neglected by economic analysis: personal acquaintanceship, trust, continuity, sharing values, skills and languages, etc.

Such characteristics clearly depend on the location of the actors in the same territory. This is not a question of the importance of mere physical proximity which allows transaction costs to be reduced, creating pecuniary externalities – that is, advantages that can be immediately turned into some form of cost saving. Geographical proximity counts as a vital condition for the creation of organizational proximity, that is a community of economic and social actors, based on local customs and values, on the continuity of personal relations, on acquaintanceship and trust. The importance of proximity defines the centrality of the territory as the place where the economic community consolidates its own competitive advantage in everyday practice. In this sense, competitiveness cannot be separated from embeddedness in the territory.

Influenced by the knowledge based economy approach – that is the idea that in contemporary economy the main production factor is knowledge – some approaches have moved behind social embeddedness, developing the intuition that relationality might seize the form of specific cognitive processes that take place to the local and regional scale. In such perspective, through the recursivity of relationships, local actors become involved in some form of collective learning that allows them to produce the competencies they need in order to compete successfully. This cognitive account of relationality has been of course influenced by the emergence of high tech cluster, such as the Silicon Valley, where innovation occurs through interaction and spillovers among a number of specialized SMEs and top ranked universities. Nevertheless, it soon spread over the realm of hi tech industries to embrace as well traditional production, such as engineering or textile. Here, the concept of tacit knowledge – the kind of knowledge that cannot be easily codified and that consequently is embodied in practices and routines – is central. Following Michael Polanyi conceptualization, some geographers, mainly Scandinavian, have developed an original reflection on how tacit knowledge is proximity sensitive, that is, it cannot be transferred over distance and it needs social embeddedness in order to be exploited and reproduced. When the knowledge displayed in the innovation and production process relies on competences that are embodied in groups of workers or of firms – for instance know how – the only way to get access to such knowledge is localizing in proximity of such competence pools.

LED as a Qualitative Process

To summarize what some authors call the relational turn in economic geography, we might say that over the last 20 years the attention shifted from the intensity of input–output linkages to the qualities of immaterial, untraded relationships. This shift is mirrored by what we called the emphasis on qualitative rather than quantitative dimension of development. The consciousness that development processes are geographically uneven date back well before the spread of LED theorization. At least since the 1950s, economists and geographers have been dealing with the need of reducing the development gap between rich regions and less developed areas, both between countries and within the same country. Nevertheless, these approaches have been mainly focusing on the quantitative dimension of development, that is the growth of the regional economy in terms of macroeconomic indexes – such as employment, investments, income, consumption, and export. Of course, also LED is about the creation of shared wealth, but we cannot ignore the influence played by the debate about the extent and the meaning of development, showing up the limits of the growth based approaches.

Once again, the 1980s represent the relevant threshold to understand the qualitative turn that took place in economic geography. The identification between development and change was deeply criticized already in the previous decade, but such judgments were bounded in radical social and intellectual movements, such as Marxism, feminism, and ecologism. Since the 1980s, the skepticism toward growth based theories, instead, has been shared by a larger array of influential scholars and organizations, traditionally linked to mainstream growth based approaches. Just to consider the major changes in the attitude toward growth rhetoric, we can mention the work of

UN Brundtland that in 1983 coined the term 'sustainable development' to stress the need of conjugating economic growth with the preservation of natural resources and environment. Analogously, since the beginning of the 1990s, UNDP – the UN agency for development – has been reflecting on more complex definition of development overcoming the mere assessment of economic performances, outlining the human development index (HDI).

The idea is that the label 'development' has lost part of its attractiveness and it needs to be connoted by making explicit certain qualities we assume development should possess. We can talk about 'sustainable development' to mark that development should have certain environmental qualities, and 'human development' to denote that growth should take into account basic human needs – health, education, rights. In a similar way, scholars have been talking about LED to stress how development process should be sensitive to local and regional qualities.

In the case of LED, the emphasis on quality is related to the assessment of the role that difference and variety play in addressing the process of development. The refusal of standardized, universal, and hegemonic development models and the related belief that multiple development paths are possible are grounded on the awareness that territories are different. This means that the relational, material, and immaterial resources necessary to support firms' competitiveness vary from place to place. More deeply, it implies that different territories, even in case of a similar endowment in resources, have different institutional and cognitive competencies to combine and develop such resources.

The sensitivity of LED theories to the quality of development is strictly related to the failure of mainstream policies in reducing regional unevenness. Maybe with an excessive simplification, we can claim that regional policies in the 1950s – like Perroux's growth poles or Myrdal's cumulative model – were based on quantity rather than quality: building new infrastructures, attracting large corporations, lowering spatial and labour costs, etc. During the 1980s, not only regional inequalities were still at work, but something new became evident also in prosperous regions: different localities sharing a similar industrial tradition, dating back to the early stages of European industrialization, and with a similar quantitative endowment – infrastructure, large firms, resources – started to follow different development paths. While some of them – such as the Ruhr area – were still on the edge of the technological and competitive frontier, other – such as the British Midlands – were experiencing a process of deindustrialization and periferalization. In the same period, in Italy, the dualistic narratives opposing the wealthy North West to the deprived Mezzogiorno were challenged by the emergence of the Third Italy – the regions in Central and North East Italy, characterized by a more jeopardized economic geography of SMEs and industrial district. Even the South Italy was not any more described as a homogeneous flat land of poverty, needing massive public expenditure and integration with the industrialized North, as some seminal forms of industrial district were well recognisable, even in the census data.

As a consequence, the perspective of LED is more evolutionary and dynamic than the previous approaches to regional development: the focus in not any more on the stock of resources (infrastructures, firms, universities, labor pool) which are located in a given territory, but on the dynamics through which such resources are combined and mobilized in order to achieve the development goals. In other terms, the stocks – the hardware to use an ICT metaphor – are necessary but not sufficient without the software – the relations and the competencies – indispensable to make them work.

Another issue that is strictly associated with quality and variety is the consciousness that LED entails a higher degree of choice than previous regional development theories. If LED has to be understood with reference with some qualities that it should possess rather than mere growth, and if it depends on a broader range of factors and dynamics rather than the mere stock of resources, then LED is a process that cannot be taken for granted, but it depends on a multilevel series of choices, designing the development path of a given territory at a given time. Basically, with some simplification, we can highlight two realms where choice is urgent and unavoidable when dealing with LED.

The first issue is about the purposes and aims of development. Since its very origin, LED is committed with a broader notion of local prosperity that transcends the mere creation of economic wealth: communitarian participation, preservation of local values and identities, safeguarding of a certain degree of social cohesion and social mobility, and enhancement of the quality of life are all aspects that concur to broaden the spectrum of purposes that LED aims to achieve. On the one side, in its more critical and alternative narratives, LED can diverge so far from mainstream accounts for development, to foresee and accept the reduction of economic growth in order to achieve other goals, such as community's heritage and/or environment conservation, autonomy with respect to globalization, etc. In this perspective, some stream of local development reflection sounds somehow sympathetic with postdevelopment literature. On the other side, other actors involved in LED, such as local administrations, trade unions, entrepreneurs associations, are more concerned with the growth side of LED, improving the attractiveness of their territories and enhancing competition with other localities and regions.

The second realm where choice matters is related to the tools that are suitable to foster economic development to the local and regional scale. As we highlighted in the previous paragraphs, the factors that play a pivotal role cannot be interpreted any more in terms of standard equipment in material resources that alone secure eco nomic growth. Literature stresses that there are at least four areas to be covered in LED policies (infrastructure, inward investments, labor skills, and local firms). The novelty, as we stressed before, is that the success of the regional economy does not rely on the achievement of a critical mass in each of these areas, but on the fact that the four elements match with each other and they match altogether with the specific cultural, social, and political assets that make each place a unique geographical and historical product. As a consequence, policy design is becoming more and more a matter of choice in a context – the local and regional development potential – that is dynamic and evolutive: in a word, complex.

In LED theories and practices, the issue of choice is often treated in terms of governance. While the concept of government refers to a form of management of the public sector entirely entrusted to local and national political administrations, the idea underlying governance is based on a radically different perspective. When we talk of governance, attention is focused on a form of local government and management that is based on the interaction of many actors on the local scale: local and transnational companies, associations, labor unions, uni versities and research centers, in addition, obviously, to local and national institutions. In other terms, the choice of LED objectives, strategies, and actions is not – or it should not be – any more a matter for professional politicians and economic elites, but it entails a broader range of interests, representations, and projects. Hence, over the last 20 years we have witnessed the emergence of new decisional tools apt to secure transparency to the governance processes. For instance, new organizations, such as development agencies, have been created all over Europe in order to ease the partnership between public and private actors. Analogously, EU funding to regions has to be allocated more and more via participatory practices, cherishing local endogenous project making rather than top down development recipes.

The Holistic Dimension of LED and Its Risks: From Territorial Competitiveness to Territorial Reification

In this concluding section, we would like to highlight some of the extreme, and more ambiguous consequences, of LED theorization. In the previous sections we have argued that both a relational and a qualitative turn char acterized LED literature over the last 20 years. These two features turned out to deeply denote LED in holistic terms. If it is not possible to cut clear boundaries between the economic and the socio cultural relationships or be tween the governmental and the nongovernmental, we must adopt a holistic gaze which allows us to keep toge ther the different issues entailed in a LED process.

In general terms, holism in LED means that we cannot establish a clear hierarchy between actors, projects, strategies, and actions, but we have to pay attention to the way different actors, projects, strategies, and actions are put together in order to achieve some LED goals. To a deeper level, in many approaches, holism leads to the idea that local development entails some degree of col lective agency. This process is quite evidence if we consider the widespread use of the concept of 'system': local labor system, regional innovation system, and pro duction system are just some examples, but the systemic influence is evident also in other LED stream, such as localities studies or the milieu conceptualization. The system is first and foremost the level in which a relational dynamic occurs between different actors. It confers on both the actors and the system itself the capacity to as similate the disturbances and changes that arrive un ceasingly in the technological and competitive world, as well as collectively producing innovation and knowledge. The set of interactions between actors and territorial conditions is the origin of a system effect that appears as a particular technical, political and social atmosphere, climate and culture. It is thus difficult to define if not through concepts – such as milieu, industrial atmosphere or territorialised learning – that by expressing a process of perception, learning and continuous action, assumes multiple qualitative dimensions. This systemic nuance can be more or less explicit, but it usually brings the idea that a territory can be described as an agent with some emergent properties which do not belong to its individual components (individual citizens, firms, universities, pub lic administration, nongovernmental organizations, etc.).

Such an holistic account of the territory is quite evi dent in the emphasis that LED literature (but also a large part of contemporary social studies) put on the concept of community, assumed as something different from so ciety. In particular, communitarian institutions – such as personal acquaintances, mutual commitment and trust – are seen as a facilitator of those networking processes that reduce transaction cost and enhance technological spill overs. Analogously, we can mention, as an evidence of the holistic attitude of LED theories, the conceptualization of collective agency which is often behind the reflection on governance. Through governance and the definition of participatory process, LED is often interpreted as a process driven by some sort of collective agency, as if the territory itself – that is, the community of the stake holders – expresses its own projects and development paths.

What we would like to call to the reader's mind is that such a holistic conceptualization is, to some extent, am biguous as it can lead to quote opposite conclusions with respect to the original purposes of LED advocates. We can, for instance, consider the issue of territorial com petition, that is the idea that territories compete against each other in order to attract footloose globalized flows – of investments, dwellers, and events. Of course, territorial competition is, to a large extent, an output of global ization, as the pressure of global competition applies not only to the individual firms, but more and more to the whole territories.

At a deeper insight, talking about territorial com petitiveness is not an innocent metaphor: in its extreme consequences, it implies that the whole territory is (or should be) organized in order to be competitive. There is also another implication: the possibility that the territory is something with clear boundaries and a monolithic identity within them, just like other economic actors, like a consumer or a firm. A successful economic actor must be consistent in its purposes and consistently organized in order to achieve them. This brings us back to the holistic approach to LED, to what we might call the 'juridical personality' paradox, that is, the extreme pos sibility of attributing to a place some form of 'collective agency'. More explicitly, talking about the competitive ness of places also implies attributing a clear economic purpose to a social entity, a territory. This is undoubtedly a dangerous metaphor, as the importance of the external relations of companies brings into play the entire fabric of social and cultural relations of a place. Talking about the competitiveness of places implies assuming that these relations have not only economic consequences but also an economic purpose, that is, that the purpose of social and cultural relations is economic.

These features can explain why there is a large convergence about the importance of territorial competitiveness between heterogeneous groups, such as neo liberal policy makers and local development advo cates. The former find the competitiveness issue coherent with their overwhelming interpretation of economic interest as the engine moving and structuring human relationships, while the latter seem to be more attracted by the emphasis that territorial competitiveness puts on community, homogeneity, and communality of purposes within the local scale. In this perspective, we can fully grasp how pushing to the extreme consequence one of the main ideas proper of LED – attributing a collective agency to the territory – can bring us to conclusions which are somehow the opposite of what we were ex pecting. In fact, while the reflection on LED moves from the recognition of the relational, social, and cultural di mensions of the economic processes, the outcome of territorial competition is quite the opposite, making the territory an economic entity. Also it is quite striking to observe that under the label of LED we can find both ardent postdevelopmentalist, challenging infinitive growth ideology, and neo liberal institutions such as the World Bank and the OECD – both of them, it has to be noticed, played a fundamental role in inscribing LED in the mainstream developmentalist agenda.

More radically, even if we avoid the trap of attributing an economic finality to the territory, the holistic inter pretation of LED would bring us to another dead end, the reification of space, and territory (i.e., considering them as mere object of our experience, given once and forever, rather than socially and culturally constructed over time). In fact, assuming the territory as a system (i.e., an entity with its own boundaries able to perform in some degree as a collective agency) has some important consequences.

First of all, it follows that the territory expresses a certain level of homogeneity and cohesion. Better put, we might say that the LED debate tends to assume that territories possess an identity, in the twofold meaning of both a unique set of characters which make it different form the external world (the nonlocal or supra local) and a continuity over time, a sort of path dependence (not by chance one of the recurring terms in LED literature). Once again, LED risks leading to conclusions that are, to a certain extent, the opposite of its inspiration and starting point. In other terms, although LED theories insist on the variety of the development paths and the inexistence of a unique model, in many cases they emphasize the char acters of homogeneity, a consensus that a local system should possess in order to compete against other terri tories. Notwithstanding the attention paid to evolutionary economics and to the role played by ceaseless transfor mation and adaptation, path dependence risks locking the fate of a territory to the continuity with its past – either in terms of specialization or, in more sophisticated cases, to learning processes which can become obsolete and slow down rather than increase firms' competitiveness.

Second, another possible unintended consequence of LED scholars' attitude toward the reification of the ter ritory is evident in the fact that conflict is often pushed outside the boundaries of the local community, in the competition against other territories, while the local system is praised as the locus of consensus and cohesion. In other terms, overvaluation of cooperation and shared institutions within the territory hide the conflictual di mension of actor–network interaction. This emphasis on territorial collective agency and cohesion, and the con sequent underestimation of conflictual dynamics, is also evident in one of the cornerstones of LED, that is, the emphasis it puts on the concept of governance. This shift mirrors to some extent the crisis of traditional forms of political representation, not considering that the com munity and the so called civil society often hidden strong disequilibrium in terms of power. Also in this case, the outcome is opposite the original purposes of LED enthusiast: rather than producing democracy and participation, governance simply hides the interest of well known shareholders, reproducing well established hierarchies.

See also: Agglomeration; Competitiveness; Governance; Industrial Districts; Learning Regions.

Further Reading

Camagni, R. (ed.) (1991). *Innovation Networks: Spatial Perspectives*. London: Belhaven Pinter.
Cheshire, P. C. and Gordon, I. R. (eds.) (1995). *Territorial Competition in an Integrating Europe*. Avebury: Aldershot.
Conti, S. and Giaccaria, P. (2001). *Local Development and Competitiveness*. Dordrecht: Kluwer Academic Publishers.
Cooke, P. (ed.) (1989). *Localities: The Changing Face of Urban Britain*. London: Unwin Hyman.
Cooke, P. and Morgan, M. (1998). *The Associational Economy: Firms, Regions and Innovation*. Oxford: Oxford University Press.
Fujita, M., Krugman, P. and Venables, A. J. (2001). *The Spatial Economy: Cities, Regions and International Trade*. Cambridge, MA: MIT Press.
Lee, R. and Wills, J. (eds.) (1997). *Geographies of Economies*. London: Arnold.
Maillat, D., Quévit, M. and Senn, L. (eds.) (1993). *Réseaux d'innovation et milieux innovateurs. Un pari pour le développement régional*. Neuchatel: EDES.
Maskell, P., Eskelinen, H. and Hannibalsson, I. (1998). *Competitiveness, Localised Learning and Regional Development: Specialisation and Prosperity in Small Open Economies*. London: Routledge.
Pecqueur, B. (2000). *Le développement local*. Paris: La Découverte & Syrios.
Pike, A., Rodriguez Pose, A. and Tomaney, J. (2006). *Local and Regional Development*. London: Routledge.
Porter, M. E. (1990). *The Competitive Advantage of Nations*. New York: Free Press.
Pyke, F., Becattini, G. and Sengenberger, W. (eds.) (1990). *Industrial Districts and Local Economic Regeneration*. Geneva: International Institute for Labour Studies.
Storper, M. (1997). *The Regional World. Territorial Development in a Global Economy*. New York: Guilford Press.

Relevant Websites

http://www.ilsleda.org/home/
International Links and Services for Local Economic Development Agencies.
http://www.worldbank.org/urban/led
World Bank Local Economic Development.

Local Economic Development, Politics of

K. R. Cox, The Ohio State University, Columbus, OH, USA

Glossary

Growth Coalition A coalition of agents, public and private, with an interest in the economic growth of a particular geographic area. The idea is applied particularly to smaller-scale areas; this explains the common reference to local growth coalitions, and, less frequently, to regional growth coalitions.

Local Dependence Dependence of revenues on continuing location in a particular place in virtue of the expense and/or impossibility of relocation elsewhere. Often this is a result of the significance to business costs of investments of long life, the amortization of which depends on a continuing flow of value through the local economy. In other cases it can be a result of dependence on social relations that are built up slowly over time, like networks of relations with other firms or particular labor market relations, and which are quasi-fixed. Relocation would negatively affect both revenues and costs until the necessary connections had been reestablished. The idea can also be applied to local government, dependent on a local tax base and to workers; the latter again depend on localized social relations not just of the labor market sort but also with kin and friends.

State Structure A conception of the state as a structure of social relations of a legal and financial sort from which one can derive the various powers and susceptibilities of the different agents – political parties, civil servants, committees, government agencies, representatives – through which state decisions are made and implemented. These relations are more accurately sociospatial. State structure varies in ways which shed light on the different forms assumed by the politics of local economic development around the world.

Territorial Competition Competition, usually for inward investment, between entities that constitute themselves ideologically as representative of a broad cross section of local or regional interests, running the gamut from business to state agencies, to labor and to residents.

Urban Governance The practices and structures through which agents with varying interests in a particular urban area attempt to achieve some policy coherence with respect to realizing those interests.

In human geography, the interest in the politics of local economic development (LED) and the emergence of a literature on the topic date from approximately the early to mid 1980s. This was initially an American interest but it was then taken up in Britain. There have been studies in countries other than the United States and the United Kingdom and there is a continuing tension between particularity and claims of a more universal character. Work in this area has certainly contributed to widening horizons in economic and political geography but has yet to register adequately some of the more recent and influential trends in human geography as a whole. These would include both political ecology and what has come to be known as 'the cultural turn'. Most of the research into the politics of LED has been largely from a standpoint in political economy, and that understood in terms of its more left leaning variants.

Historical Background

The origins of the interest in the politics of LED in North America are diverse. As an academic interest, the early contributions were clearly from sociology and political science and to a lesser degree, economics. The early work of Harvey Molotch was particularly important. Other significant contributors include Susan Fainstein, John Mollenkopf, and Clarence Stone.

A landmark paper was Harvey Molotch's 'City as a growth machine'. In that paper he detailed the interest of professional property owners in the growth of cities, the subsequent formation of growth coalitions, and their involvement in the politics of urban growth as a means to enhancing property values. In human geography a decade later the work of David Harvey on urban politics established similar benchmarks for work in that field. These echoed Molotch's earlier insistence on the relation between the immobility of propertied interests and wider flows, particularly of people, but in a much more theorized and far reaching fashion. Not least Harvey showed how immobility was more than a matter of real property, that similar politics could be observed at other geographic scales, and how the politics of LED could be assimilated to the politics of capitalist development as a whole.

Regardless of particular disciplinary developments, though, the emergence of this interest corresponded with real world changes in the geography of uneven development. During the 1950s, but particularly during the 1960s, the urban reappears as an arena of serious tensions around the geography of development. Ongoing suburbanization in the context of the intense jurisdictional fragmentation of the American metropolitan area led to something known as the 'central city suburb fiscal disparities problem'. This in turn focused attention on redeveloping the central city, particularly its downtown areas, though subsequent urban renewal and highway projects were the object of popular opposition. With the growth of the Civil Rights Movement and concerns about black unemployment, these interests converged with ones in transforming Afro American geography: variously bringing employment into Afro American areas of the city or bringing blacks closer to where the job growth was occurring through a highly controversial process of 'opening up the suburbs'. This was a politics whose context was uneven development within metropolitan areas. Early work on these topics by political scientists and sociologists found echoes in human geography, particularly in the work of Julian Wolpert on location conflict, but for the time being the dominant voices were from outside geography.

During the 1970s geographically uneven development and studies of its politics underwent a scale shift. Attention now focused on what were seen as quite dramatic changes at interregional and international levels. Interregionally, the new spatial imaginary was dominated by the differentials in growth between the so called Coldbelt and Sunbelt. Instead of the distribution of employment and investment between central cities and suburbs, attention now turned to the competition between cities and states in different regions and, as the growth of Japan and the Asian Tigers came into view, between cities in different countries.

It is during this period that the competition for inward investment first captures the public and academic imaginations. An important expression of this, and resonating widely, at least in the relationships it pointed to, was the idea of 'capital versus communities'. This was articulated most clearly and insistently in the work of Bluestone and Harrison. The idea was a simple one. Capital had acquired a new mobility, interregional, international, and this ability to disinvest from one locality and invest in another had given it an enhanced capacity to drive hard bargains with labor and local government; the latter were the 'community' bit. The results were worker givebacks, financial incentives from local government to firms to locate in particular places which strained its ability to provide other services, a shortchanging of environmental regulations, and general stress in both workplace and living place. As the rhetoric of globalization took off, this was to receive a new emphasis in the form of what was predicted to be 'a race to the bottom'.

In the United States, these various concerns rapidly became public ones. The reasons for this have partly to do with the structure of the American state and partly to do with that of the media. When compared with Western Europe, both are quite remarkably decentralized. Local and state governments have to find major proportions of the money that they spend. LED has therefore loomed large on both local and state agendas. Furthermore, the structure of the federal state, as well as that of the individual constituent states, gives local representatives extraordinary powers to deliver in the form of (e.g.,) particular items of physical infrastructure, as well as obligations to do exactly that. The localness of the American media, particularly the newspapers, lends an additional urgency to specifically local issues. These comprise much of their bread and butter. In consequence, development politics assumes an unusual saliency in American cities and the reader finds it hard to avoid it.

From this standpoint, the relatively late arrival of an interest in development politics in Britain, both lay and academic, should not surprise. What made the difference was the re emergence of a regional problem subsequent to the dramatic changes in economic policy introduced by the various Thatcher governments in the 1980s. Local government now started to pay attention, though less for fiscal reasons and more for the electoral ones of being seen to deliver jobs. These changes in focus were recognized by David Harvey when he referred to the shift from urban managerialism – a policy emphasis which always had more resonance in the United Kingdom than in the United States – to urban entrepreneurialism. In the United States cities had been 'entrepreneurial' for a long time, but in the United Kingdom it was something of a novelty. In human geography this interest was registered by the so called 'localities project' examining what difference, if any, localities could make to respective economies through restructuring, attracting new firms, and so on. Later this interest was to receive reinforcement through restructurings of the British state and the devolution of some responsibility for LED.

The politics of LED has therefore 'traveled': from the other social sciences to human geography and from the United States to elsewhere, notably Britain. As a result it has changed, though not without difficulty. As we will see, the American context is quite *sui generis* and the sorts of conceptions that developed there are limited in their portability. In the hands of human geographers understandings underwent further transformations, though not all of these had to do with the characteristic operations of an enhanced spatialization. In particular, human geography was to rework the politics of economic development through a greater openness to more diverse streams of social theory.

Characteristic Framings

There is a highly influential conception of the politics of LED which goes back to the early work of Molotch, then his work with Logan, and that of Bluestone and Harrison and the 'capital versus communities' school. This is a conception which had clear appeal to human geographers because of its emphasis on local interests and wider fields of influences, including movements of capital. Central to it was indeed a relation between fixity and mobility. On the one hand, there were those with fixed assets of various sorts. For Molotch it was landowners who would constitute growth coalitions; for Bluestone and Harrison they included local governments with their stakes in local tax bases and the 'people' with their stakes in 'community'. On the other hand, there were various flows, a more global set of forces, population movements for Molotch, capital flows for Bluestone and Harrison which, if captured, would allow the enhancement of property values or the reproduction of communities. But in order to capture them, localities had to enter into competition with one other, offering various incentives of the sort indicated earlier. These in turn might generate local conflicts: struggles with residents and workers around issues of environmental disruption, plant closures, deterioration of amenities, and tax issues; and the disadvantaging of popular interests to the benefit of mobile capital and those landed interests that stood to gain from the exercise. This in turn could give rise to an ideological contestation as the local interests constituting the growth coalition responded with a vision of territorial struggle in which distributional issues within localities had to be set aside in favor of the more basic struggle with localities elsewhere.

There are a number of quite serious problems with this conception. It was, for a start, focused on a rather narrow understanding of 'development'. The idea that it could occur through the growth of existing 'export' activities was a rare intrusion into work on the politics of LED. Even the 'export' activities considered germane were quite limited in character. Yet shopping centers and residential developments could just as easily irrigate the rents of landowners and the local government's tax base as a major office or industrial employer. There was also some exaggeration of the degree to which capital was mobile, something that was taken advantage of, ironically enough, by more immobile branches of the export economy in 'their' quest for the sorts of 'business climate' that those pushing for inward investment were advocating.

Even granting the assumption of capital mobility, the central argument tended to overlook sharp variations in the vulnerability of local governments and workers to the demands of growth coalitions and their claims regarding that mobility. The fact was, and remains, that localities differ very considerably in these regards. So what they are willing to concede to a firm considering a location there, what they *have* to offer, is a variable. Growth interests in cities expanding on the backs of sunrise industries are unlikely to fight hard to keep a pipe casting plant open. However in a Youngstown, Ohio or Hammond, Indiana, that has little apart from that, and where the loss of employment would be unlikely to be made up for by the expansion of other firms, the story is likely to be very different.

In addition, and these concerns aside, this was not a conception of the politics of LED which could be transferred easily into a British context. For a start, and most significantly, the sorts of local interests and subsequent growth coalitions characteristic of the American case were not that easy to find. In the American instance there were indeed strong local interests, interests in the growth of particular local economies. In addition to Molotch's landed interests, these included gas and electric utilities and many banks. These all exhibited what is called a 'local dependence' – a dependence on enduring location in a particular place – which made them highly vulnerable to shifts in the wider space economy and important participants in local growth coalitions. However, in the United Kingdom, banks had national branching structures and historically, electricity and gas had been nationalized industries, so dependence on any particular place for the generation of revenues had been very weak. In an important sense, they had emancipated themselves from dependence on the fortunes of any one place by spreading their (geographic) risks. In addition, the sort of local 'fiscal' dependence characteristic of local government in the US, a dependence on local tax base, was much less in evidence. Rather a system of central government grants with a strong equalizing impulse worked against those sorts of sensitivities and weakened any local fiscal interest in development.

Furthermore, it is now clear that there are even wider differences having to do with the structure of respective states. The highly decentralized American state is in sharp contrast to its British counterpart. American federalism is of a radical sort, allocating very considerable powers and responsibilities to the states and the states in their turn have a similar relationship to local government. The separation of executive and legislative branches produces parties that are relatively weak compared to individual representatives. Along with the committees which any pending legislation must negotiate, there is very considerable latitude for them to work the system to local advantage. Localities, therefore, are nexuses of formal political power and this may be the crucial point. For while there are indeed strong local interests in the US in the utilities, property developers, local media empires, and sometimes the banks, without this decentralization of formal power they would have to look elsewhere, perhaps through national lobbying organizations, in order to achieve their goals. Conversely in Britain there are no such local articulations of formal power constituted by local government officials and MPs. The overlap of executive and legislative branches has meant strong political parties with only very limited scope for the expression of local interests, and a strong unitary state form has reinforced the centralization of state power. So again, the absence of strong local economic interests may be beside the point, though one might argue that if they had been more evident, then there would have been stronger pressures for a decentralization of the British state.

Even so, LED did become a public issue in Britain, as indeed elsewhere in Western Europe, and there have been moves – rather weak moves in practice – to decentralize responsibility for LED. Perhaps more significantly, interregional tensions of a developmental sort have given rise to a plethora of regional parties, spanning the left–right spectrum and including, *inter alia*, a revived Communist Party in East Germany going under the new name of the Left Party, the Northern League in Italy, the Scottish and Welsh Nationalist parties, respectively, the Vlamsblok in Belgium and the Christian Social Union in Bavaria. Again, state form seems significant. In the US, regional tensions can be accommodated by log rolling, and bipartisan coalitions; something that is a good deal less feasible in the more centralized states of Western Europe with their much stronger party systems.

These contrasts in institutional context have in turn provided a basis for some development of understanding of a more universalistic sort. The confrontation with the American literature and the experience on which it is based has also had some beneficial results; not least a realization that the politics of LED cannot be reduced to competition for inward investment. The 'Cambridge phenomenon', the growth of high tech industry in the Cambridge area, has continued to attract an attention not dissimilar to that which Saxenian earlier devoted to Silicon Valley. Industries in the biotechnology and software sectors have become a very significant element in the economic base of the Cambridge area. As in the Silicon Valley case, housing shortages, transportation problems, and resistance to further development have become an issue for the high tech sector and a challenge to its continuing expansion, generating a distinctive politics of local development. Still other work has focused on the politics of creating clusters to begin with; Helen Lawton Smith's work is notable and connects with another recent area of trans Atlantic interest, which is that of governance.

One of the distinctive features of work on the politics of local development in the 1990s was a growing appreciation of the way in which policy is underpinned by processes of local governance; processes that allow the negotiation of differences of interest between agents who might otherwise engage in mutually destructive contestation, and at the same time provide a source of ideological blandishment through which more popular opposition can be countered. Some of the original intellectual impetus for this work came from Clarence Stone and his idea of urban regimes. An urban regime would be formed by a particular local government and various corporate actors, though sometimes including representatives of organized labor. Their material basis was structural: the fact of a division of labor around development between the municipality with its tax base interests as well as its powers over physical infrastructure, taxes, and regulation on the one hand, and corporate interests – developers, would be 'inward' investors – with their capital as well as their needs for state cooperation and inputs on the other. To facilitate the functioning of the division of labor, however, perhaps to deepen it further, 'understandings' were needed: a shared vision of the development of the city, the mutual trust and to and fro through which that vision could be implemented. Stone's work, hugely influential as it has been, particularly among political scientists and among the urban studies fraternity, has not gone without criticism. Some of this has to do with the excessively dualistic nature of the social theory underpinning it; a rather uncritical lining up of various polarities underneath one another in which urban regimes are equated with agency and the political, and the wider flux of capital which they hope to capture, with structure and the economic. There have also been some significant shortcomings in terms of its understanding of space and this has provided human geographers with the opportunity to make a very distinctive contribution to this literature, though not always with explicit reference to urban regime theory.

Part of the problem has been the failure to conceive of governance beyond the level of a particular urban government. In the context of the highly fragmented metropolitan areas of the US where cooperation across jurisdictional lines is a *sine qua non* of LED, perhaps through the county, possibly through contractual relations between individual local governments, or through metropolitan planning bodies, this is a very odd oversight indeed. Harvey's earlier work on urban politics and his ideas about territorial structured coherence provided a framework for going beyond this and taking in the whole metropolitan area. He did not make reference to urban regime theory but he certainly pointed in the direction in which human geographers were later to take the idea of governance for LED.

Part of the context for this in the British instance has been the multiplication of state authorities with some sort of responsibility for local development, and typically in some sort of hierarchical fashion. Even before the new regional development agencies there were public–private partnerships which took in a number of local authorities. The division of authority has required some sort of coordination. The regional development agencies have been responsible for financial assistance but local authorities have retained control over land use regulation and the counties and metropolitan councils have been the ones with the site information submitted by private owners.

Beyond that, various agencies of the EU and the grants that 'they' can provide have provided an additional impetus for some sort of local, interjurisdictional organization.

In Britain, these sorts of local or even regional gov ernance relations have been quite dominant among various state agencies with overlapping or contiguous territorial responsibilities. In the United States, Stone's original idea of some sort of public–private organization retains some validity. The mediation of inward invest ment is carried out on a regional basis and involves co operative relations among local governments, local Chambers of Commerce and, providing the regional element through the scope of their respective service areas, the gas and electric utilities. The latter are key, since they control the information on possible sites. Very similar sorts of relations exist in the British case but, as is implied above, with counties and metropolitan councils performing the function that is carried out in the American instance by the utilities.

This particular turn in the understanding of gov ernance relations – cooperation among various agencies, public and private, with areas of interest at different geographical scales – has converged with and reinforced an interest in the politics of scale. This is a further direction in which theorizing the politics of LED has gone in the last decade. Indeed, scale has become a key word in this research area. The projects of growth coalitions, including those of local governments which are so often their political front, now have to be situated with respect to a state that has its own scalar division of labor – its own distribution of powers and responsibilities, fiscal, land use and regulatory in all manner of ways. This is not to say that this distribution is the key to understanding strategy. Powers and re sponsibilities can be bent to particular purposes, or even transformed, and the local balance of political forces may be such that mobilizing powers at higher levels of the state may be the only way in which the growth coalition can get its way. In this way, moreover, the significance of the state to understanding the politics of LED is reinforced.

The Current State of Play

From the standpoint of human geography, there remain some interesting omissions or weaknesses in the litera ture. It is surprising that the wider interest in political ecology has not had more impact on studies in the pol itics of LED, for there is obviously a relation to nature, and environmental groups have often been quick to point out a particular understanding of that relation. The same goes for questions of culture and identity: that what are called LED professionals have their own culture, and a quite masculinist one at that, is clear.

The other weakness that should be referred to is the Anglo American bias of the literature. Most of the work that has been done on the politics of local and regional development has taken as its empirical frame either American or British cases. This focus has, to be sure, encouraged useful comparative work but the relation between universality and particularity could obviously be enhanced through studies of this politics in other coun tries. France is more like the United Kingdom than it is the United States, but as Delphine Ancien has shown, it is by no means exactly the same and there are some very telling differences; the centralization of the Napoleonic state, the different balance of public and private, has had distinct effects and it would be hard to find parallels for them in the United Kingdom. We should also note Bae Gyoon Park's work on South Korea. Political parties there are implicitly regional in their support bases and this is strongly associated with the salience of regional development as a political issue: something that is cer tainly hard to imagine in either the British or the French cases. Clearly, and judging from just these two cases, exploration of other instances promises major rewards in pushing forward the boundaries of research in this area.

Despite these shortcomings and weaknesses, however, and returning to the origins of the literature in fields outside human geography, the contributions of human geographers have been of undeniable significance. Part of this is the enhanced sensitivity to the spatial and to the various ways it manifests itself in the politics of LED: beyond the simple relation of mobility and immobility, significant as that is, to understandings of the role of scale, of the differentiation of localities and the socio spatial, not just the social, construction of governance relations. There is also, however, the way in which human geographers have connected the literature with wider currents in social theory and in some contrast to the mainstream character of work in sociology, political sci ence, and economics. So while they may not have been first in the field in studying what is clearly an inter disciplinary focus of attention, they have invigorated it in distinctive and far reaching ways.

See also: Capitalism; Entrepreneurship; Governance, Urban; Local Development; Locality Debates; New Regionalism; Regional Development Models; Regional Development Theory; Regional Inequalities; Regional Innovation Systems; Uneven Regional Development; Urban Regimes.

Further Reading

Ancien, D. (2005). Local and regional development policy in France: Of changing conditions and forms, and enduring state centrality. *Space and Polity* 9, 217 236.

Bluestone, B. and Harrison, B. (1982). *The Deindustrialization of America: Plant Closings, Community Abandonment and the Dismantling of Basic Industries*. New York: Basic Books.
Cox, K. R. and Mair, A. J. (1988). Locality and community in the politics of local economic development. *Annals, Association of American Geographers* 78, 307 325.
Cox, K. R. and Townsend, A. R. (2005). The English politics of local economic development and the American model. *Regional Studies* 39, 541 553.
Cox, K. R. and Wood, A. M. (1997). Competition and cooperation in mediating the global: The case of local economic development. *Competition and Change* 2, 65 94.
Harvey, D. (1985). *The Urbanization of Capital*. Oxford: Blackwell.
Harvey, D. (1989). From managerialism to entrepreneurialism: The transformation of urban governance in late capitalism. *Geografiska Annaler* 71B, 3 17.
Lawton Smith, H. (1996). National laboratories and regional development: Case studies from the UK, France and Belgium. *Entrepreneurship and Regional Development* 8, 1 17.
Logan, J. and Molotch, H. (1987). *Urban Fortunes: The Political Economy of Place*. Los Angeles: University of California Press.
Molotch, H. (1976). City as a growth machine. *American Journal of Sociology* 82, 226 238.
Park, B. G. (2005). The territorial politics of regulation under state capitalism: Regional parties and the politics of local economic development in South Korea. *Space and Polity* 9, 237 259.
Pickvance, C. and Urry, J. (eds.) (1990). *Place, Policy and Politics: Do Localities Matter?* London: Unwin Hyman.
Stone, C. N. (1993). Urban regimes and the capacity to govern: A political economy approach. *Journal of Urban Affairs* 15, 1 28.
While, A., Jonas, A. E. G. and Gibbs, D. C. (2004). Unblocking the city? Growth pressures, collective provision, and the search for new spaces of governance in Greater Cambridge, England. *Environment and Planning A* 36, 279 304.
Wood, A. (2007). The politics of local and regional development. In Cox, K. R., Low, M. M. & Robinson, J. D. (eds.) *The Handbook of Political Geography*, ch. 32. London: Sage.

Local–Global

M. Haldrup, Roskilde University, Roskilde, Denmark

Glossary

Constructionist Perspectives Constructionism is a social theory of knowledge emphasizing that all knowledge is socially constructed, and is often contrasted to 'essentialism'. According to a constructionist position, there are no essential 'true' or 'real' facts, only constructions of knowledge that are negotiated and institutionalized in particular social contexts.

Disembedding It is a key concept in the work of sociologist Anthony Giddens. According to Giddens modernization and globalization means that social practices are 'lifted out' ('disembedded') from their local, social contexts and instead of being negotiated in face-to-face interaction is increasingly coordinated through and by abstract expert systems..

Institutional Economics School in Economic thought focusing on the role of manmade institutions (e.g., governments, organizations, culture, and so on) in the shaping of economic development or decline. Some of the key figures in the founding of institutional economics are T. Veblen, J. K. Galbriath, G. Myrdal, and more recently Douglass C. North.

Institutional Learning A key idea in the institutional economics of Douglass C. North is that economic trajectories are determined by institutions' ability to learn, adapt, and innovate. This idea has been highly influential in economic geography from the 1990s onward.

Locality Studies Analytical approach in geography and the social sciences more broadly that focuses on the dynamics and transformations of particular localities. Locality studies have been criticized for prioritizing empiricism over theoretical rigor thus being unable to generalize findings.

Marshallian Nodes/Clusters The observation originally by Alfred Marshall, that related industries seems to co-locate ('cluster') within the same regions. For example, the software industry in Silicon Valley or the several regions specialized in textile and footwear fashion in Northern Italy.

Neoliberalism Neoliberalism refers to the reemergence of economic liberalism in the 1980s and 1980s associated with 'Reaganomics' in the US and Thatcherism. The term is often related to a common strand of elements, such as the dismantling of welfare, privatization of public enterprises, tax cuts, and deregulation of labor markets. Since the end of the cold war, neoliberalist principles have been applied globally as part of post-socialist governance in East Central Europe, Russia, and other Soviet successor states as well as by the IMF policies toward Third World countries.

Relationality 'Relationality' is the basic assumption that social practices must be understood in terms of relations and contexts. In geography, the term refers to a general turn away from notions of absolutes space to an analytical approach in which practices, places, and processes are seen as relational outcomes of 'networked' connections and interactions between people and between people, institutions, and technologies, and so on.

Scale In geography, scale traditionally refers to the scale of maps. More recently, geographers have interrogated how scale is (politically and ideologically) produced. That is, for example, how global processes are produced as 'above' local, urban settings.

Time–Space Distanciation According to sociologist A. Giddens the disembedding of social practices from local, face-to-face interaction results in Time–space Distanciation. What popularly is referred to as globalization actually means that our daily practices are increasingly 'streched' in time and space as expert systems take over (part of) the coordination of our work and everyday life (see also 'disembedding).

Translocal Neologism aimed at capturing the fact that a growing number of socio-spatial processes cannot be conceived of within a single hierarchy of local–regional–global, but rather crisscross different geographically dispersed localities. Especially in migration studies (but also in tourist and leisure studies, urban studies, and so on) the term translocal has been used to capture such crisscrossing networks and processes.

Introduction

The concept of 'local–global' appears in three rather different disciplinary contexts. First, it is traditionally used within the philosophy of sciences to describe the generalizability of an axiom or a theory, that is, to determine whether a theory, sentence, or axiom has local or global validity. Optionally it may be used to signify whether an event has local or global impacts. This use is predominantly limited to analytical philosophy, computer science, and mathematics and will not be addressed further in this article. Second, the distinction between local–global knowledge has been used within the social sciences and human geography in a way which is slightly related to the discussion mentioned above to address debates about the universality or contextuality of scientific accounts. In particular, the notions of 'local knowledge' or the analogous term 'situated knowledge' have gained a widespread usage in some parts of (especially critical and post structural) human geography. The third and most important usage of the local–global dichotomy within the discipline of human geography is in relation to the study of social–spatial processes in geography and more broadly within the social sciences. Here the local–global has been conceived of as a dialectical relation tying together political–economic forces and social–cultural responses. This article concentrates on the two latter usages of the local–global dichotomy.

In geography, the conception of knowledge as profoundly situated, partial, and 'local' (as opposed to the ideal of positive, universal knowledge) has been an important inspiration in cultural, primarily feminist, geography, as well as in some contributions to economic geography. Clifford Geertz' notion of 'local knowledge' was decisive for the development of the contemporary sociology of knowledge, anthropology, ethnography, and cultural geography; it also inspired important discussions on methodology and the character of knowledge widely within the social sciences. This may be seen in relation to the rise of constructionist perspectives on geographical knowledge production in the 1990s. The most important usage of the local–global dichotomy in contemporary human geography is, however, closely related to the discussion of the changing conditions for local livelihood, everyday life, household economies, and so on popularly associated with 'globalization'. Hence, tensions within the global–local continuum have attracted much attention as an important feature of current political, social, cultural, economic, and environmental changes. As such, the local–global continuum is no longer a specialized field of interest for geographers but has moved to the central arena in popular debate as well as social theory.

This article begins by considering the (critical) discussion of the production of geographical knowledge which first and foremost has been raised by feminist geographers, but which also has contributed to critical accounts of geographical knowledge production in other fields of human geography. It then moves on to consider how the 'local–global' continuum became central as an (academic) response to transformations of the international division of labor which have taken place in the US and Western Europe from the 1970s onward, especially with the rise of a 'locality' approach to 'global' restructuring. After this, the article discusses more generally how human geography has conceived of 'the local' as a nexus of multiple forces, relations, and mechanisms, and traces how this understanding of the 'local' has found expression in recent economic geography. In doing this, it focuses on what has been coined the 'relational turn' in economic geography. After this is a brief outline of current ways of theorizing about the local–global continuum in terms of 'scale' and/or as dialectic processes such as what has been coined 'glocalization' in contemporary human geography. Then, the article broadens its perspective by engaging with some of the ways in which the local–global continuum has moved to the forefront of contemporary debate in politics and social theory by first discussing how the dichotomy has been perceived by what has commonly been known as 'antiglobalist' movements, and second, by showing how some of the most significant contemporary social theorists have placed the investigation of local–global relations central to the understanding of the dynamics of modernity. Finally, it addresses some of the central pitfalls and exaggerations of the 'local–global' relation regarding the character and impact of globalization in contemporary debates within human geography and the social sciences more broadly.

'Local Knowledge' and the (Im)possibility of a View from Nowhere

We begin by briefly considering the inspiration that the sociology of knowledge has had on how geographers conceive of 'local' and 'global'. This inspiration may not be the dominant usage of the local–global binary pair of terms in geography; however, the problems indicated by the local–global dichotomy within the sociology of knowledge have had important impacts throughout human geography.

In the sociology of knowledge, the local–global binary is a fundamental issue in relation to the production of knowledge: Is global, universal knowledge an ideal to strive for in social and scientific analysis or is it basically an illusion, a fallacious mirage that we should stop worrying about and instead concentrate on how to account for the different 'local' productions of knowledge that our everyday worlds as well as our disciplinary worlds are made of?

An important meaning of the local–global dichotomy here is in the general discussion of the character of knowledge in social science. As argued by anthropologist Clifford Geertz, the social world is inherently contextual and our knowledge of it is therefore as per definition partial and contextual too. This means that the idea of global, universal knowledge within the social sciences is an illusion. The skepticism toward the ideal of universal knowledge found in Geertz' interpretive anthropology is close to feminist scholar Donna Haraway's critique of 'the God trick' of positivist science. Indeed, Haraway's term 'situated knowledge', a term which forms the basis for her critique of the ideal of 'universal knowledge' in science, may be seen as analogous to the notion of 'local knowledge'. 'Universal knowledge' is seen by her as the dominance of one partial ('objective', masculine) perspective on the world over other possible perspectives. Her work more than Geertz, has been influential on the development of a critical perspective on knowledge production in geography. Like Geertz, she argues against the possibility of an objective 'view from nowhere' and emphasizes the contextual local and situated character of all knowledge production. Aaraway's and Geertz' use of the local–global dichotomy has inspired critical accounts of the production of geographical knowledge from the perspective of feminist geography in particular. Thus, feminist geographers have argued that key concepts and methods in geography such as space, place, mobility, nature, as well as the methodological dominance of fieldwork in geographical education are gendered ideological constructs that point to the profoundly situated character of knowledge production within the tradition of the geographical discipline. The critical insistence on the 'situated' and 'local' character of all geographical knowledge production must be understood in relation to the general increase in influence of post structuralism and constructionism in human geography from the 1990s onward.

The impact of the idea of knowledge as local and situated has however not been limited to feminist geography, but has also formed part of a broader critique of knowledge production and geography especially in some of the epistemological and ontological problems presented by the increasing use of geographical information systems (GIS) in geographical knowledge production and education. In addition to this, the local–global dichotomy as used within the philosophy of science/knowledge has also informed work within economic geography on embeddedness, networks, and innovation processes of firms and entrepreneurs. Here, the question of local–global knowledge has been framed as a matter of the character of knowledge and its ability to be codified (global) or transmitted in informal, face to face interactions as tacit (local) knowledge (see section titled 'The relational turn in economic geography'). A central argument in relation to this has been that as all knowledge production is basically local, partial, and situated, this fact must also be integrated within the scope of economic geography. This argument has been made regarding the importance of dealing with the local, situated knowledge of agents when analyzing their tactics and strategies in relation to innovation and entrepreneurship. This usage of the local–global dichotomy inspired by the sociology of knowledge may not be the dominant use of the terms in geography; however, this usage is a significant part of the theoretical debate within a range of subdisciplines.

The 'Locality Approach'

The use of the notion of 'local' related to critical discussions of the production of geographical knowledge outlined above is however not the dominant usage in contemporary human geography. Thus, the local–global continuum has most frequently been related to geographical discussions of the transformations of politics, livelihoods, and social and cultural relations on different scales. In this debate, it is possible to distinguish between two different ways of invoking the local–global terms: first, as an approach for analyzing international and global changes; second, as a field of study. Here we start by concentrating on the former and then address the latter in the following section.

In geography, the concept of the local–global relationship has first and foremost been used as an approach to analyze the impacts of internationalization and globalization on cities, regions, localities, and communities. Following from this, it is no surprise that the increased attention to the relationship between 'local' and 'global' dynamics was closely connected to the challenges of the New International Division of Labor which emerged in the late 1970s. The thesis of the new international division of work briefly assumed that an overarching shift in the spatial organization of global industry was taking place following the general economic downturn of the 1970s. In several countries, this shift showed up as a general decline in jobs in the manufacturing sector and a parallel growth in jobs in the service sector', a shift away from metropolitan centers and a relocation of (manufacturing) jobs away from the 'industrialized' countries of Western Europe and North America to Third World countries, typically located in Far East Asia. Throughout Western Europe, these dramatic changes caused worries regarding the economic prospects for welfare and economic growth and gave rise to a number of research initiatives such as the CURS (Changing Urban and Regional System in the UK) in Britain in the 1980s which aimed at identifying the local consequences of this relocation for localities, cities, and regions. Similar initiatives also developed, for example, in Scandinavia, to capture the local and regional consequences of the transformation of the welfare state induced by industrial decline and changes in livelihoods especially in peripheral regions. Of special significance here is the sustained, interdisciplinary Norwegian research focused on the shifting roles and functions of places and localities especially in the Arctic regions.

These research programs, commonly referred to as 'locality studies', also implied a partial departure from some of the most generalizing ambitions of structuralist Marxist geography. While this departure was seen by some as a worrying turn toward empiricism at the expense of theoretical rigor, the approach fruitfully gave rise to theoretical debate on the role of empirical research in geographical theory. Hence, the ‘locality approach’ has been decisive for developing and maturing the philosophical and methodological position of realism in geography.

Another important and lasting contribution of the ‘locality approach’ to current human geography has been the (re)introduction of ideographic approaches to geographical methodology following earlier influences from nomothetic positivism and Marxism in geography. Thus the locality approach has propelled a host of case studies in several countries introducing a multiplicity of ethnographic and anthropological inspired methods in geography often stressing the importance of acknowledging the particularity and contextuality of sociospatial phenomena as well as the significance of local knowledge and institutions in constituting ‘locality’ and spatial scale. In doing this, the ‘locality approach’ of the 1980s anticipated both later discussions of scale in geography (see below) as well as elements of the ‘cultural turn’ of the 1990s.

The Local–Global Nexus

While related to the emergence of ‘locality studies’, ‘the local’ and its embeddedness in global transformations has attracted increasing interest as a field of study in its own right. The dominant response from the discipline of geography and especially economic geography to the transformations of the spatial organization of labor tended to focus on issues such as the operation of Transnational Corporations (TNCs), capital flows, international commodity chains, the formation of foreign plants, and cross border corporate linkages. Researchers inspired by institutional economics and regulation however increasingly focused on local and regional scales. In doing this, the local (or in some writings ‘the region’) was seen as a ‘nexus’ for various social and economic processes coming together. This research especially pointed to the inter and intrafirm relationships within industrial regions and localities. The conceptual background for this was a resurgence of Alfred Marshall’s concept, the ‘industrial district’, as a description of territories consisting of a specialized working population engaged in a largely homogenic industrial structure with related industries such as the Lancashire cotton mills of the nineteenth century and characterized by common or complementary skills, normative orientations, and so on. Translated into the post War context, such industrial districts were ‘discovered’ especially around the fashion industry of Northern Italy and the semiconductor industry located in the San Francisco Bay Area. These ‘neo Marshallian nodes’ were largely understood as local and regional precursors of a new dynamic mode of capitalist regulation especially. The key idea in this literature was that the economic downtown experienced in the 1970s represented the end of the Fordist regulation and the start of a new post Fordist mode. Although this conception was also fiercely questioned, the view of the local as a nexus of social dynamics of various social, economic, and political (and cultural) character has been highly influential and especially important in bringing about a more contextual approach to the geographical study of local and regional transformations. This has been especially important in relation to understanding the relationship between regional development and noneconomic factors. Furthermore, the increased recognition of the local as an important context for ‘global’ networks and links has also pointed to the acknowledgment of how cultural elements such as family networks, work cultures, industrial heritage, institutions, and social ties shape in important ways the conditions for economic processes. Hence, the discussion of the local–global continuum in economic geography has, in recent years, formed an important stepping stone both for the development of a general critique of the ‘culture economy divide’ in geography as well as a more specific focus on the dynamic qualities of localities in regional economic change. This is particularly clear with regard to some of the developments in recent economic geography, especially those inspired by institutional economics and economic sociology (see below). However, the understanding of ‘the local’ as an important context for the unfolding of everyday lives, community relations, social and cultural identities such as gender, class, and race, as well as understanding the importance of the role of place in relation to these issues have, more generally, been themes at the forefront of geographical thought and analysis in recent years.

The Relational Turn in Economic Geography

As a more or less direct response to the changes in the spatial organization of labor described above, ‘mainstream’ economic geography also has gone through radical changes in recent years. One major influence has been the ‘culturalization’ of the very idea of economic organization and agency. Here, we illustrate these changes with two important developments. First, we briefly examine the empirical ‘discovery’ of the local and regional basis of economic change and the subsequent emergence of an inductive approach to the local ‘embeddedness’ of regional economies following the decline of structuralist Marxist approaches to economic geography. Second, we describe how this enhanced acknowledgment of the spatiality and the significance of social and cultural context for economic processes has matured into what can be described as a ‘relational turn’ in economic geography's understanding of the local–global continuum.

Taking our starting point in socioeconomic paradigms such as institutional economics, regulation theory, and economic sociology as well as in the empirical 'rediscovery' of the territorial basis for economics propagated by authors like P. Krugman (economics) and M. Storper (geography) in the 1990s, it is possible to identify a distinct research program for a new industrial geography emerging from the mid 1990s onward. This research program is most clearly pronounced in the so called Uppsala or localized learning school in economic geography. Inspired in particular by the economic sociologist Mark Granovetter's work on social networks and seeing 'local' networks and dependencies as the decisive element in the creation (and lack of creation) of prosperous regional and national economies, this approach identifies three research areas:

1. how economic agents (firms, companies, enterprises) relate to place and are spatially organized,
2. why (some) companies and branches tend to cluster in particular regions, and
3. why some high cost industries seem to maintain competitiveness within particular regional settings.

The key for answering these questions has been to focus on the local 'underpinnings' of learning processes and innovation, the local institutional and social framework for socioeconomic change. One major source of inspiration has been the concepts of 'tacit' and 'codified' (or 'explicit') knowledge adapted from theorist of science, M. Polanyi. Viewing innovation processes as driven by agents drawing on a conglomerate of codified and tacit knowledge, the central questions for economic geographers are what learning processes can be facilitated by codified knowledge only (which is seen as becoming increasingly ubiquitous and free floating as a consequence of globalization and enhanced technological capabilities of communication) and what processes rely on the use of tacit, informal, embodied, and embedded knowledge. The key assumption here is that the key 'engine' in innovation and economic change is to be found in 'tacit', non explicated forms of knowledge communicated and reproduced through informal networks with a high degree of spatial 'proximity'. Although there are significant differences in the importance delegated to noneconomic factors, for example, with regard to the significance of social capital for regional transformation in different versions of the approach, it is a basic assumption across the different contributions that international economic dynamics are inherently spatial and territorial and based on 'local' capabilities for learning, innovation, and adaptation. Thus, the localized learning approach is from the outset formulated in relation to issues of globalization as an approach aimed at analyzing the relationship between global flows (of capital, information, technologies, people, and so on) and the localized capabilities and processes facilitated by informal knowledge, institutions, and social capital. By drawing particular attention to the role of spatiality and networks in economic processes at local and regional levels, the 'localized learning' school of thought can be seen as part of a broader turn toward 'relationality' in contemporary geography. Thus, the 'discovery' of the territorial basis for local and regional development mentioned above not only meant the engagement with numerous empirical examples of specific variations of economic transformations but also paved the way for rethinking local/regional economies as the spatial outcomes of particular relationalities between technology, organizations, and territories. Indeed, this 'relational turn' can be seen as part of a general move toward relational understandings of geographic concepts in recent debate within the discipline.

With regard to economic geography, we see such developments in three interdependent theoretical movements. First, geographers have focused on the creation of industrial spaces such as industrial districts, clusters, and world cities (see above). Second, an increased awareness of the role played by social networks has uncovered the multiple, local–global tensions embedded in the differentiated production of space and also opened up for an understanding of networks as consisting of hybrid actor networks. Third, this engagement with the networked geographies and productions of differentiated spaces in economic development has also pointed to the need for engaging more directly with the production of geographical scale. In doing this, 'the relational turn' also means a reintroduction of the politics in economic geography as the politics of globalization and governance has become central for understanding the dynamics of local, urban, and regional transformations. At the same time, the increased attention given toward social networks and how actors are embedded in multiple discourses and hybrid networks, the relational turn in economic geography also signals a break with the culture–economy dichotomy in much of the geographical literature and calls for a more social and cultural approach to economic phenomena. Finally, the attention toward the production of geographical scale also points to broader discussions and developments in recent geographical thinking, developments that conceive of the local–global continuum as particular and produced scalar geographies rather than fixed hierarchies.

Hierarchies, Scales, or Networks

In geography, the discussion about the local–global continuum has often been addressed as a question of 'scale'. In everyday speech we often talk about local as being 'below' national or global structures and processes. However, geographical theory challenges us to rethink the local–global continuum in more relational terms and understand that geographical scale is a produced and contested phenomenon.

Contemporary discussions of 'scale' in geography took off with radical or Marxist geography contending that 'scale' basically was a product of the structures, mechanisms, and the struggles of capitalism. Being an effect of the contradictions within capitalist society, the hierarchy of local/regional, national, and global levels should be approached by paying attention to the 'political economy of scale', that is, the social and political struggle involved in constituting spatial hierarchies. Contemporary geographical thinking however argues that even if we acknowledge scale as an effect of the contradictions of our social and economic system, geographers, planners, politicians, and so on too often tend to assume an underlying hierarchy of spatial scale as something natural, given, and fixed. Thus, core concepts in geography such as regions/networks, urban communities, places, nation states, local/global, and so on seem to be connected to more or less predefined scales of operation. Against this assumption, recent theories in human geography have suggested that geographers ought to acknowledge that scale is socially constructed and therefore geography needs to pay particular attention to the socially and politically laden character of scalar expressions and the processes constituting these. Thus, it has been argued that 'scale' is an epistemological issue rather than an ontological one, that is, scale is primarily a product of people's social construction of reality rather than part of that 'reality'. Other responses to the role of 'scale' in geographical thought have stressed the uneven distribution of capabilities to operate at various scales. Particularly important has been the discussion of how mechanisms of globalization create new spaces for the politics of scale.

While the discussion of scale often seems to imply an idea of nested hierarchies of spatialities, another important inspiration for the theorization of local–global in contemporary human geography comes from analysis of migration, hybridity, and translocal 'networks'. Here it has been argued that the social networks of, for example, migrant groups across different cities produce new spaces of action that not only constitute the cultural context of their experiences and hybrid identity formations, but also create new 'translocal' spaces for political action and economic entrepreneurship. While this emphasis on the networked character of social networks, experiences, and activities across scales and localities suggests a network approach to the question of local–global relations rather than a scalar, it more importantly points to the necessity of understanding local–global relations as a politics of scale from 'below'. In recent debate, proponents of a scalar approach in geography have also emphasized the need to rethink scale in relational terms rather than as absolute, nested hierarchies in space.

An influential concept with regard to this has been the concept of 'glocalization'. This neologism coined simultaneously by sociologist R. Robertson and geographer E. Swyngedouw intends to capture the interdependent relation between local–global by indicating that all sociospatial processes may be viewed as simultaneously global and local. Like this, the concept forces us to think of the local–global in relational and dialectical terms and to avoid reducing complex sociospatial processes to simple dichotomies such as globalization and local adaptation. Others have similarly called attention to the multiple ways scales can be mobilized for different strategic purposes, for example, by 'jumping scale' by translating 'local' issues and concerns to other scales as well as the uneven distribution of how such abilities to 'jump scale' are distributed among different social actors such as versus neighborhood groups. In this way, it has been emphasized that the politics of scale also implies a social stratification between social actors able to operate across scales and actors more or less confined to 'local' spaces' (see also discussion below).

Global Challenges, Local Responses

In popular debate, the notions of local and global are often mobilized in order to produce specific countertactics against global transformations. Mobilizing the local scale against transformations perceived to come from 'above' or 'outside' is not a new phenomenon and social and urban geographies can provide a multiplicity of examples of how, for example, local communities and neighborhoods have been mobilized against/for urban renewal projects and so on. However, what is interesting in relation to the discussion of local–global dialectics is that this continuum has provided a key way of reasoning for contemporary grassroot movements related to both ecological issues and activities for a more just relationship between 'the West and the Third World'. Thus, the local–global continuum also appears in the form of political slogans like 'think globally, act locally'.

The increased awareness of the political, ideological, and social aspects of the local–global dialectics has, since the late 1990s, been an important part of the mobilization of nongovernmental organizations (NGOs) and grassroots movements against neoliberalism. The activities of antiglobalist movements have been a recurrent element of meetings such as the World Bank, World Trade Organization, (WTO) or the G8 countries (a group of the eight most influential nation states in the global economy). One of the founding moments in the antiglobalist movement was when activist Jose Bové in 1999 attacked a partially built McDonalds in France to raise awareness about the fact that fast food industry is an outlet for a global productivist agricultural system that both erodes the quality of its products (e.g., through introducing hormone treated beef) as well as eroding the livelihoods of farmers around the world. McDonalds and other primarily American based TNC's have become the prime targets having a strong symbolic value for what is commonly known as the antiglobalist movement.

Despite the common designation of such movements as 'antiglobalist', most supporters of the diverse and heterogeneous have been eager to emphasize that the target is not to defend 'the local' against global transformations or globalization as such – but rather, to attract attention to one particular form of globalization, that is, neoliberalism. Several activists (including Bové) have therefore preferred to describe themselves and their political actions as a 'global justice movement' or 'alter globalization' emphasizing the need for more socially just and fair modes of globalization. With major activities in relation to the WTO meeting in Seattle (1999) and G8 (meeting in Genoa (2001)), the antiglobalist movement has, despite criticism toward some of its supporters for resorting to violent and destructive forms of protest, been a driving force behind World Social Forum (an annual event since 2001) and the European Social Forum (an annual event since 2002) which has engaged hundreds of thousands of activists, NGOs, and politicians in discussing issues relating to alternatives to neoliberal globalization. In this, these movements have been important for driving through an increased public awareness of, for example, the mechanisms of global commodity chains and demands for more socially sustainable global trade relations between the West and the Third World. Examples of such activities span from the spread of global 'fair trade' products to the mobilization of popular support for 'global taxation' of capital flows propagated by the French based Association pour La Taxation des Transactions pour 1′ Aide aux Citoyens (ATTAC) movement which also has a strong foothold in several, primarily, European countries.

In social theory, the rise of 'antiglobalism' has also been understood more generally as part of a broader response to globalization and modernity. Thus, the rise of neonationalism, political Islam, and antiglobalism have been described as part of 'localist' tactics toward global capitalism and this approach to the local–global dialectics as new sociospatial axes of stratification and conflict have been a significant theme among contemporary social theorists (see below).

Global–Local in Wider Social Theory and Science

The concepts of local and global are one of the few examples in recent years of geographical discussions and concepts that have migrated from human geography into the social sciences and social theory more broadly. Hence from the early 1990s onward, in social theory and research, the local–global continuum has been increasingly used as an important tool to grasp contemporary processes of modernization and changing social and spatial patterns of inequality between territories, cities, and regions. Especially interesting has been the work of D. Harvey, A. Giddens, M. Castells, and Z. Bauman all of whom have used the local–global continuum as an entry into more broad social diagnoses of contemporary transformations. In doing this, discussions of changes in the global–local relationship have also moved away from specialist, geographical debate to the core of contemporary social theory.

Particularly important in this has been human geographer David Harvey's concept of time–space compression (1989) that represents globalization as a local–global dialectic. According to Harvey, the increasing capabilities for mobility which are made possible by innovations in infrastructure, transport, and communication technology and appears to be an inherent feature of modernity supports a process in which the world is experienced as a shrinking space. This process in which 'space is annihilated through time' is, according to Harvey, basically an effect of the dynamics of capitalism as described by K. Marx and F. Engels in the *Communist Manifesto's* famous observation (1842) of the modernizing forces of capitalism as a process in which 'all that is solid melts into air'. However, Harvey makes some important elaborations on this general diagnosis of capitalism. First, he argues that the apparent 'shrinking space' of the modern world is part of dialectic that is inherently unequal. The experience of the world as a space that is shrinking is at the same time the experience of particular social and economic classes. The main sources of social power in the modern (capitalist) world arise precisely from the control of time and space and the capability to manipulate these. In this way, time–space compression is not an equal process as the control of, for example, flows of capital, trade, and so on rests firmly with the economic upper classes of global capitalist society. Second (and related to this), time–space compression is also a dialectic inducing spatial inequalities between territories, cities, and regions by constantly creating new centers and peripheries within the international division of labor.

While being a geographer, David Harvey has had an enormous impact on the wider field of social theory. Especially his concept of time–space compression has, together with Giddens similar concept of time–space distanciation (1990), influenced social and cultural theories of globalization. According to Giddens, the social history of modernity is a process in which social action is disembedded and 'lifted out' of local, situated interaction. For Giddens, like Harvey, innovations in communication and transport technologies and growth in what he calls ‘expert systems’, by which he refers to systems that are able to coordinate actions across different spatiotemporal contexts (e.g., money), means that we increasingly have to rely on systems and processes not present in space and time and therefore beyond our control. This increasingly ‘stretches’ the spatiotemporal relations of which our actions are a part. At the same time, it creates increased ambivalence and reflexivity as a fundamental precondition for modern lifestyles and strategies. Although Giddens does not explicitly refer to the Harvey, the concepts may be conceived of as complementary and analogical concepts.

While both Harvey and Giddens emphasize the dialectical character of the local–global continuum, other significant contributions in recent social theory have paid more attention to the increasing polar character of ‘the global’ and ‘the local’ in the contemporary world. This way of conceptualizing the local–global relationship is most clear in the work of Manuel Castells. In his three volume work on *The Rise of the Network Society* (1996–98), Castells describes how technological innovations in transport and means of communication are the central vehicles in bringing about a new era in which space is annihilated through time. However, Castells goes one step further than Harvey and Giddens by contrasting current sociospatial transformations as caught within the dichotomy of two opposing logics or ‘spaces’ the ‘space of flows’ (i.e., the space of communication technologies, capital, and the social elite) and the ‘space of places’ (the space of bounded physical localities, situated everyday lives, and the social underclass). According to Castells, the rise of the network society means the emergence of a new ‘global’ dominant set of social, political, and economic relations in which the ability to flee spatial constraints is the characterizing feature. This in turn produces overwhelming and uncontrollable problems for political identities and entities that for a large part still draw their power from bounded ‘local’ territorial forms (such as nation states, regions, or urban neighborhoods). It is this dichotomy between an emergent non controllable ‘space of flows’ and an increasingly subordinate ‘space of places’ that explains the dynamics of contemporary political and economic changes. Thus, the fall of the Berlin Wall and the subsequent dissolution of the Soviet Union can, according to Castells, be explained by the inability to exercise control based on territorial bounded space ‘space of places’ over the flows of capital, technology, and goods ‘space of flows’ in the contemporary information based capitalism. This dynamic is furthermore seen as a constituting feature of the emergent ‘information age’ which creates new forms of inequality. Hence, a new global inequality is created around the pipelines and flows of the information age, decoupling large regions of the world such as the major part of the African continent from global growth and development.

The fourth major contribution of the local–global continuum to social theory has come from cultural critic and social theorist, Zygmunt Bauman, who in a number of publications on globalization, consumption, and identity (1998–2000) has emphasized how the changing balances of the local–global continuum drastically create new forms of sociospatial inequality. For Bauman, the game of globalization basically signifies a process in which not all can be winners. The transformations addressed by Harvey, Giddens, and Castells imply a logic in which some are losers and some are winners, ‘What appears as globalization for some means localization for others’. While the globe takes the shape of ‘the tourist’s oyster’ for the global elite, the underclass seems confined in their local life worlds. This view led Bauman to suggest social stratification today is increasingly constituted by the organization of space and the stratified abilities to transcend the constraints of the spatial boundaries imposed by locality. This means that while the global today appears as one frictionless space, for example, footloose capital or roaming tourists, for numerous others the ‘local’ is a fate they are imprisoned by. Examples of the latter could be workers of de industrialized cities in post Soviet Russia or inhabitants of squatter towns in Third World metropolises. In proposing this diagnosis, Bauman particularly emphasizes globalization as a class based ideological process closely associated with neoliberalism. In that sense, Bauman’s contribution can be seen in continuation of earlier Marxist critiques of capitalist modernity and it may not be a big surprise that he has had an almost iconographic status in (part of) the political anti globalization movement.

As the above contributions show, the changing relationships between local and global have transcended the disciplinary boundaries of human geography and become a major theme for social theories as well as contemporary political debate. Like this, important discussions regarding both critiques of capitalist (neoliberal) globalization and how nation states should respond to this has taken their starting point in a discussion of the ‘local global’ continuum as well as important attempts to theorize the dynamics of contemporary spatial political and economical change.

Geographies of Flows and Places

As is clear from the discussion outlined above, the debate of ‘local–global’ in human geography as well as broadly within the social sciences has been closely tied to the increased acknowledgement of the impacts of ‘globalization’. Therefore it is important to carefully avoid some of the central pitfalls in the way the idea of globalization is presented in both academic and popular discourse. Here, it may be useful to distinguish between the three different notions of globalization that seem to dominate the debate and consider how we can understand the changing relations of the local–global without exaggerated and oversimplified claims. Thus we may distinguish between three different tropes in debates of local–global relations, that is, as a question of:

- Scale
- Process
- Relations

A dominant understanding of local–global relations has been based on the notion of 'scale'. However, very often this idea of scale is presented like a fixed, absolute hierarchy of nested 'spaces' in which transformation and change seem to go in a unidirectional way from the top of the hierarchy toward the local. We see this especially in discussions of economic transformations and local development where the top down way of conceiving local–global relations has often been combined with a view leaving economic processes (at a global level) as the only dynamic forces changing conditions for social, political, and cultural relations at a 'local level'. This conception is widespread in public debate and has been the main target for recent critical discussions in human geography of the social construction of scale (see earlier section titled 'Hierarchies, scales, or networks').

The scalar understanding of global–local relations is sometimes combined with the second major way of conceiving local–global transformations, namely, as a question of 'process'. Thus, it has been argued that what has been coined 'globalization,' marks a whole new epoch for societies around the globe. While this has been seriously questioned by empirical research (e.g., by problematizing the extent to which contemporary patterns of trade is really 'global' or by pointing to the fact that with regard to transnational relations, we would have to acknowledge globalization as a pre modern phenomenon), the idea of globalization as a whole new epoch that modern nation states are about to enter into at the dawn of the twenty first century is often pronounced in social theoretical accounts as the one discussed above. This is evident, for example, in the idea that current societies are in a transitional phase between a space of places and an emerging and already existing space of places.

While both the scalar and the process oriented approach to understanding the changing local–global relations in different ways seem to exaggerate and oversimplify contemporary transformations, they do point to the fact that there are import transformations taking place at all scales in the local–global continuum. Thus they point to the need for a more thorough understanding of the 'relations' that change conditions for cultural, economic, social, and political relations at all levels. In particular, the increased connections across space facilitated by migration, business networks, transformations in information and communication technologies (ICTs), as well as infrastructure and transport technologies has attracted attention to the multiple 'translocal' relations that have altered the conditions at all levels in the local–global continuum.

Some authors, in particular the sociologists Ulrich Beck and John Urry have suggested that globalization should be understood as the emergence of a new independent 'condition' emerging in addition to place and locality bound everyday life; a condition that already penetrates all aspects of localized lives. Here it may be useful to dwell a bit on the arguments put forward in relation to this. In recent years, scholars from a wide range of disciplines such as political science, sociology, and anthropology have embraced the concept of 'cosmopolitanism' a conceptual framework for capturing current transformations of social relations, subjectivities, and political identities. While much of this debate has focused on cosmopolitanism as a new set of principles for political governance emphasizing commitment to global citizenship and cosmopolitan democracy, Beck and Urry have argued for the relevance of a 'cosmopolitan perspective' studies of social and cultural change too. The argument is that a new kind of nonelitist cosmopolitanism is on the rise – a cosmopolitanism building on the vast presence of translocal connections which all people are a part of. Thus, it is argued that the 'normative' (or 'philosophical') cosmopolitanism found in political theory and science may constructively be complemented with an engagement in tracing out how processes of 'cosmopolitanization', that is, the unintended side effects of increasingly transnational networks and flows are subjectively changing the spheres of habitual experience and everyday life. The day to day experience of the global through consumption, media, and so on transforms the experiential spheres of life worlds enforcing a globalization of emotions and empathy. This creates a transformation from within particular localities and nation states that transform everyday consciousness and identities significantly. Issues of global concern are becoming part of the everyday local experiences and the 'moral life worlds' of the people'. The increased mobility of people, images, and objects are not only brute observable movements between territorial organized regions of the globe but an ontological condition transforming social relations from within and seeming to produce a more 'networked' pattern of social life. Such mobilities transform the human condition on all scales, from the human body and its increasingly prosthetic connectedness to other bodies through technological gadgets such as hand free mobiles etc., to global processes of climate change. In emphasizing how everyday life and social networks are changed from within by processes of 'cosmopolitanization', this approach has been criticized for conflating cosmopolitanism and globalization and to avoid confusion, a better term for these processes might be 'banal globalization' as suggested by Urry. In this way we may acknowledge that 'banal globalization' of course not only produces cosmopolitanism but also enables 'banal orientalism' to be performed as the globe is increasingly turned into what has been termed 'a tourist oyster'. Apart from pointing out the obvious fact that 'banal globalization' is an unequal process in which not all people entertain the same mobility opportunities, cosmopolitanism also needs to be freed from the political, utopian connotations as well as from elitist experiences of frictionless travel and communication captured in Bauman's metaphor of the world as 'the tourists oyster' (see above). While the latter approach to the changing local–global relations clearly also possesses the pitfalls and risks of exaggeration and oversimplification, it does however clearly point out that these relations and their transformations should be approached as a relational aspect of the networks and mobilities that connect particular scales, localities, and places and as an empirical phenomenon rather than a theoretical dictum.

Conclusion

As we have seen in this entry, the local–global has an important theme in recent human geography. In critical discussions of the value laden and biased character of apparently 'neutral' and/or 'objective' geographical accounts, concepts, and representations, the local–global dualism has been important for bringing out a recognition of the fact that all knowledge and also geographical knowledge is inherently partial, biased, local, and situated and that the ideal of 'a view from nowhere' as the favorable position for science is nothing but an illusion. More importantly, the local–global is a central theme in debates about the changing conditions for industry, livelihoods, social and cultural identity formation, and political governance. Here the shifting balances of local factors and global forces were a central concern in human geography since the mid 1970s through the 1980s especially in relation to the changing global geographies of manufacturing that took place in those years. As outlined in this article, the local–global continuum has moved into the center of geographical, theoretical, and popular discourse. In recent social theory, the local–global continuum has been argued to be a central axis of social stratification as well as a fundamental element of the processes and dynamics of contemporary modernity. However, social theory seems often to assume a rather simplistic approach to the local–global dualism. As for theoretical debate within human geography, the dualistic concepts of local–global have been problematized. Thus it is a significant feature of recent disciplinary and theoretical debate that a relational turn has provided new ways of conceiving of local–global relations. Thus, more network oriented approaches have contested scalar and absolute approaches emphasizing the need to think of the local–global continuum in terms of translocal networks, flows, and mobilities. By offering a relational framework, such approaches have produced important new understandings of contemporary, political and economic transformations as well as of the spatiality of social networks and cultural identities. In doing this, the relational approach has helped to bridge a variety of established dichotomies in human geographic analysis and theorization. Hence, relational approaches have transcended hierarchical thinking as well as culture–economy and nature–society divides by incorporating elements of post stucturalist understandings of discourse and human–nonhuman relations into their conceptual framework. Despite these achievements, the network based approaches to the local–global continuum still seem to lack thorough understandings of power relations and inequalities. With regard to this, contributions from neo Marxist and political geography in particular have provided an interesting understanding of the local–global continuum as a dialectic process in which power relations, political struggle, and inequalities are inescapable elements.

See also: Actor-Network Theory/Network Geographies; Capitalism; Cosmopolitanism; Critical Realism/Critical Realist Geographies; Economic Geography; Feminism/Feminist Geography; Fordism, Post-Fordism and Flexible Specialization; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Mobility; Modernity; Neoliberalism; Nongovernmental Organizations; Positivism/Positivist Geography; Poststructuralism/Poststructuralist Geographies; Pragmatism/Pragmatist Geographies; Radical Geography; Regional Development and Noneconomic Factors; Scale; Social Capital; Social Capital, Place and Health.

Further Reading

Bauman, Z. (1998). *Globalization. The Human Consequences*. Cambridge: Polity Press.
Castells, M. (1996). *The Rise of the Network Society. The Information Age*, Vol. 1. Oxford: Blackwell.
Giddens, A. (1990). *The Consequences of Modernity*. Stanford, CA: Stanford University Press.
Harvey, D. (1989). *The Condition of Postmodernity*. Oxford: Blackwell.
Maskell, P. and Malmberg, A. (1999). Localised learning and industrial competitiveness. *Cambridge Journal of Economics* 23, 167 185.

Rose, G. (1993). *Feminism and Geography: The Limits of Geographical Knowledge*. Minneapolis, MN: University of Minnesota Press.

Smith, N. (1987). Dangers of the empirical turn: The CURS initiative. *Antipode* 19, 59 68.

Storper, M. (1995). The resurgence of regional economies, ten years later: The region as a nexus of untraded interdependencies. *European Urban and Regional Studies* 2(3), 191 221.

Swyngedouw, E. (1997). Neither global nor local: 'Glocalisation' and the politics of scale. In Cox, K. (ed.) *Spaces of Globalization*, pp 137 166. New York: Guilford Press.

Urry, J. (2004). *Global Complexities*. Cambridge: Polity Press.

Yeung, H. W. C. (2004). Rethinking relational economic geography. *Transactions of the Institute of British Geographers* 30(1), 37 51.

Relevant Websites

http://www.attac.org/
ATTAC.

http://www.globalpolicy.org/
Global Policy Forum.

Locality Debates

P. Cooke, Cardiff University, Cardiff, UK

Glossary

A 'Spatial Trap' May occur when a population finds itself caught in a region suffering industrial restructuring and job loss with no skills or experience of, for example, entrepreneurship that might help restructure the region endogenously. As early movers migrate to other regions, housing rents fall, particularly for owner-occupiers, and they have insufficient assets to purchase or afford housing rents elsewhere. Hence they are in a spatial trap, especially if little new industry moves in to take possible advantage of declining labor costs. A corollary of 'spatial trap' is 'spatial fix', where government incentives or subsidies repopulate such a newly deprived area with, for example, foreign direct investment in branch plant employment. This 'fix' enables the affected population to remain fixed in a now possibly improving regional or local economy.

Industrial Restructuring Occurs when change, usually negative, occurs in the geography of production, as when firms in advanced economies seek cheap labor zones en masse so that the industry must restructure by concentrating on higher value-added activities like research, design, or greater innovation. Industrial restructuring often has serious effects, both positive and negative, on regional development.

Regionalism Has three meanings: first it has political connotations, whereby political action in representation of specific regional (subnational) interests is mobilized to achieve relative or complete autonomy for a region; second, it is a school of geographic thought that sees regions either as 'culture areas' with key internal coherence and external differentiation from other 'culture areas' or as institutional constructs with economic, political, or administrative coherence similarly; third, it is misused journalistically to refer to collections of neighboring countries in geopolitical discourse. 'Georegions' is correct in this context.

Spatial Determinism Is the (misplaced) belief that spatial forces exert causal effects on human practices, for example, temperate climates favor economic growth, which is easily disproved nowadays.

Spatial Divisons of Labor Occur as industrial restructuring evolves. Thus low-wage, routine activities tend to concentrate in less-developed regions of a country or the world. More skilled, design or technical activities within firms or whole industries may concentrate in yet another region, while head office and R&D functions may concentrate in primate or regional cities of relevance to the industry in question.

Spatial Science Is the theorization and scientific analysis of distributions of phenomena such as agglomerations, including cities, industrial districts, agricultural zones, deserts, and natural systems over geographical space.

Introduction

The locality debates of the 1980s and 1990s were the most heated yet illuminating wrangles in human geography since those over 'environmental determinism' in the 1950s and the 'quantitative revolution' in the 1960s. The soul of the discipline seemed to be at stake, and the victory of the localities perspective over a desiccated and nihilistic structural Marxism that had threatened to remove spatiality from other than an epiphenomenal 'fix' for capitalist accumulation proved final. This observation may seem odd to those interested in and knowledgeable of the *dramatis personae* of the main vehicle fuelling the locality debates. This was the UK Economic and Social Research Council (ESRC) research program, the Changing Urban & Regional System (CURS), which the present author coordinated. The very titles in the preceding sentence express the politics of the day. Formerly the Social Science Research Council (SSRC), the name had been deliberately changed by Sir Keith Joseph, the Thatcherite minister of education and science on the spurious grounds that social research was nonscientific. The echoes of notions like 'scientific socialism' probably also clanged in the nightmares of the Mad Monk, whose recanting of his interventionist past had so buttressed Thatcherism. As well as establishing the right wing Centre for Policy Studies think tank, Joseph was the man who insisted his chief civil servants read *The wealth of nations* before offering him policy advice in his first Thatcher government stint as secretary of state for industry 1979–81.

The use of 'urban and regional system' as the object of inquiry of the ESRC research program was also not entirely value free. While city systems had been identified as early as Pred and such concepts influenced the quantitative revolution in geography and planning, the associated mathematization and behaviorism had produced little real 'theoretical' or 'policy' progress and were roundly criticized for this in a pathbreaking article by Doreen Massey. Thus it was not enough to consider relations among cities in the absence of theory addressing their 'regional' context and interactions. Moreover, the insights of Manuel Castells into a differentiation between the 'organization' of industrial production being analytically 'regional', while that of consumption constituted the 'urban' question had helpfully cleared much conceptual undergrowth. The coincidence of Massey's publication and the coming to power of the Thatcher and shortly after Reagan neoconservative governments was not circumstantial. Massey correctly foresaw how an absence of 'theory' in the work of spatial scientists up to then had left industrial regions in particular, vulnerable to the application of neoconservative political ideology in which, fundamentally, money is treated as more important than people – excepting only those who make it. Behaviorist models and their designers melted away in observing the disastrous changes caused by Thatcherism in the UK, including inner city riots across the country, the destruction of industrial relations law, and the decimation of regional employment in industry.

In discussions at the ESRC geography committee about the launching of a first research program into the effects of the damage wrought, the chair, Michael Wise of London School of Economics (LSE) is reported to have observed that as the geographical right had failed, now was the time to hand the intellectual torch to the left. This was to be under the intellectual stewardship of Wise's academic colleague, Doreen Massey, previously funded by an SSRC research fellowship to investigate the changing industrial geography of the UK. The original CURS program committee, chaired by Brian Robson and including Massey, Noel Boaden, Derek Lyddon, Duncan Gallie, and the author, as well as such future academic administration luminaries as Howard Newby, was sufficiently left leaning in 1984 to appoint the present author, then an avowed Gramscian as program coordinator. Immediately charged with writing four theoretical and methodological program framing briefings, the author rediscovered the postgraduate joys of burning the midnight oil. These were broadly set within a Massey flavored 'restructuring thesis' theoretical framework that arose from her fellowship findings.

Not surprisingly, many of the seven proposals finally selected for funding involved principal investigators familiar with a regional political economy approach, even though by no means all were economic geographers actively studying restructuring. Sociological teams such as that from Lancaster led by John Urry and Aston by Dennis Smith joined sociologically inflected ones from Durham, Kent, and Essex involving the likes of Huw Beynon in the Durham team (researching Middlesbrough) and urban sociologists Chris Pickvance, Nick Buck, and Michael Harloe in the teams studying Thanet (including welfare sociologist Peter Taylor Gooby) and Swindon, respectively. Indeed, of the 28 team members contributing to the core text only seven (25%) were practicing geographers. The other professional groupings were 12 sociologists (43%), seven planners (25%), and two economists (7%). The rather threadbare nature of theoretical economic geography at the time is thus revealed quite tellingly in these statistics. In what follows, the following lineaments and resonances of the locality debates will be elaborated upon. First, a fuller account of the origins of the localities analysis from the early 1980s will be attempted. Then the rise of locality studies and critical responses to the approach will be reviewed. Finally, attention will be devoted to the long legacy of locality studies, which includes new regionalist research, urban cultural and consumption research, and at a lesser level the 'scalar envelope' critique of spatiality in general.

Origins of Localities Analysis

In this section we will be concerned with three main issues. First, the critique of regional industrial geography that heralded the turn toward theorized spatial analysis of the locality kind. Second, a review of the importance of restructuring theory and recognition of ways in which 'local uniqueness' could be theorized in a general way by the development of local narratives within an encompassing theoretical discourse. Finally, reference will be made to this conspectus in relation to then ongoing debates in regional political economy.

For a considerable time there had been a rising sense of dissatisfaction with the products of – broadly neoclassical – theorems in economic geography and its subfield of industrial geography. It echoed a more widespread sense of the failure of geography to offer a more theoretically engaging canvas for the study of spatiality. Among the earliest such critics were the likes of Gunnar Olsson who produced a first explicit critique of geography's new obsession with formalistic modeling, quantitative techniques, and the so called 'spatial science'. A little later, Andrew Sayer produced a monograph entitled 'a critique of urban and regional modeling' which critiqued industrial geography for its reductionism, spatial determinism, and, accordingly, 'chaotic concepts'. About this time, radical urban and regional political economy associations had formed in the US and UK and these became the forum in both countries for the most advanced critique and retheorization of spatial issues, notably inequality and its causal mechanisms. The Union for Radical Political Economics (URPE) group and its journal *Review of Radical Political Economics* was founded in 1968 in the US and attracted involvement by the likes of the late Bennett Harrison, a leading critical thinker of his day, Ann Markusen, and Dick Walker. In the UK, the Conference of Socialist Economists (CSE) was founded in 1970, and from 1971 published *Bulletin of the CSE* which was in 1977 transformed into a refereed journal *Capital &*

Class that, like its US counterpart, continues to thrive. Articles by economic geographers published in *Capital & Class* include those of Doreen Massey and Andrew Sayer.

A key section of CSE was its 'Regionalism' group, which met regularly, usually at the Architectural Asso ciation in London, as well as in UK regional locations. Among the membership of the CSE Regionalism group were the present author, Doreen Massey, Richard Meegan, Ash Amin, Andrew Sayer, Mick Dunford, Kevin Morgan, Ray Hudson, Jim Lewis, Robin Murray, and John Urry, the last named also running the Lancaster Regionalism research group. At these meetings papers were presented and discussed in a seminar format, among these being early critiques of orthodox economic geog raphy and reformulations later influential in the rise of so called 'new regionalism' of which locality debates were a progenitor. Interestingly, many such CSE Re gionalism debates and critiques were on 'scale' in in dustry and on the nature of the state. The latter was criticized for its overweening and often misguided interventions in regions, industry, and public affairs more generally. Ironically, Thatcherism agreed, but with markedly different motivations for change. Trying to forge theory and practice that would facilitate action tailored to distinctive regional and local identities were thus CSE Regionalism group priorities. It will be re membered that this was the era of the failed UK de volution policy affecting Scotland and Wales, and of the rise of various left wing administrations in London and other large UK cities pursuing local policies in defense of their citizens against the depredations of Thatcherism. Many CSE Regionalism group members were actively advising such administrations, notably Doreen Massey and the Greater London Council (GLC). Group member Robin Murray was the architect of the GLC's alternative economic strategy at the time, aspects of which are covered in the work of Best.

At the heart of these debates was a project of region alism and localism informed by the industrial re structuring perspective. Economic geography at any given point in time was seen as a temporary resultant of pro cesses by which firms and industries found ways to cheapen labor costs and hence raise profitability. This produced 'sedimented' 'spatial divisions of labor' with each 'round of investment'. Hence, spatially differentiated labor markets existed and they were further differentiated by sociopolitical characteristics deriving from local work and cultural specificities. These were the base for differ ent kinds of localized response to crises of many kinds deriving from variable intensities and activation of local social capital. This in a nutshell was the theory of regional and local development which informed the subsequent locality debates. Of great importance to this theory, as discussed at CSE, was the accumulating evidence of successful, radical, regional, and local economic development activity in the Italian industrial districts as described by Brusco. A powerful analysis and presentation emphasizing the shift from Fordism to network forms of collective entrepreneurship in Italy was given to the Re gionalism group by CSE member Fergus Murray, in touch with Charles Sabel, subsequently published in *Capital & Class*. Later on Piore and Sabel underlined the new 'flexible specialization' associated with the same phe nomenon. Based on research in Italy, Portugal, and UK and with Portuguese research associate Artur da Rosa Pires, an early comparative 'clusters' study was published inspired by Murray's and Brusco's results. Thus the first locality debates were about the imputed demise of Fordist corporate control of the economy in the face of an ap parently rising radical and solidaristic entrepreneurship rooted in localities thriving under left wing regional polities – indeed in Italy, varieties of communist regional administration.

These ideas were being implemented in the GLC under the policy influence of Robin Murray and dis cussion in the Regionalism group was of the superiority, in terms of working conditions and profitability of the industrial district model over the corporate model practiced – in London, for example, in the furniture in dustry. However, the left intellectual imperative to sup port trade unions – until Thatcher at the height of their power in large corporations and the state – meant there was frequently a blind spot regarding the progressive potential of collective entrepreneurship Italian style. This dilemma is well captured in Dick Walker's com radely obituary for Bennett Harrison:

> Even Bennett was drawn away from the labour question to a diversionary fight over large and small firms……. however, he was derailed by a futile attempt to show that small firm clusters were not the wave of the future … the mistake of everyone, Right and Left, was to think that small firms are bound for the dustbin of history which is like saying that small creatures are less important than big ones in an ecosystem, just because the latter hog so much visible space (Walker, 2001).

Research informed by the restructuring and localities perspective began to be published in the geographic lit erature by others than the main progenitor, for example, Cooke with a paper on the way 'tertiarisation' or the rise of services industry was expressing new spatial divisions of labor to "produce new combinations of spatial and social differentiation at the sub regional level," and Morgan with his study of the fracturing of labor and lo cality contingent upon the privatization of the UK steel industry. Subsequently, there was a great burgeoning of such literature from all corners of the de industrializing world as the imposing 'restructuring thesis' was successfully tested out. This is explored in the section which follows.

The Rise of Locality Studies

In this section the following three aspects of locality studies and debates will be examined. The first of these concerns actual changes in regional policy under the neoconservative reforms set in train by the Thatcher governments. This is done not with forensic attention but more as a broad narrative of the pressures and shifts in trajectory typical of the time. Second, attention is paid to the more salient locality studies performed from a 'broadly restructuring and spatial divisions of labor' perspective. Finally, the key findings of the 'localities' research program and the controversies it provoked will be reviewed. In the space available, the two key debates – adding to the one concerning clusters and corporations already dealt with – will be reprised: the so called empiricist and postmodernist consumption geography turns.

So widespread was the concern among academics and policy makers at the wholesale changes visited upon government programs in support of less favored areas of the UK that even CSE Regionalism group sceptics bemoaned the rapidity of policy and funding change. Indeed, so draconian were the threats to anything other than the market mechanism as a means of fuelling economic development that the main professional body with a watch over this field – the Regional Studies Association – launched an inquiry, among the panel for which was the present author. Leading 'social democratic' and other panelists, for recall the Social Democrat Party had split from an increasingly left leaning Labour Party only two years earlier, included Derek Diamond, David Eversley, Peter Hall, John Goddard, John Rhodes, Michael Keating, and, as chair, Gerald Manners. The published report sought to reexamine the nature of current regional development problems. It was followed up by a book enlarging on the analysis. The report was one of the earliest to draw attention to the 'North–South divide' in UK prosperity. Much was made of the 'success' up to the 1973 oil price hike in transferring manufacturing jobs to development regions but 'failure' of policy to narrow regional gaps significantly thereafter. Manufacturing employment decline had accelerated bewilderingly, competitiveness was weak, and productivity faltering. The UK also joined the European Common Market in 1973 and manufacturing employment decline was set to continue for the next 30 years and more as the comparative advantage of the continent's industrial regions worked itself through the economy. This was usually to the UK's detriment given the more than 15 year advantage its competitor economies in the Common Market had already enjoyed regarding investment, modernization, and competition in an enlarged market.

The Thatcher reforms, which included swingeing cutbacks in regional grants (–40%, and covering 25% not 40% as previously of the working population), privatization of the older industries that had earlier been nationalized, and the dismantling of industrial relations legislation that had functioned as something of a cushion against precipitate employment decline – especially in negotiations with Labour governments – ushered in an intensification of manufacturing job loss. The report noted a retreat from regional policy, abolition of the Economic Planning Councils, and refusal to countenance regional strategic plans. In their place were offered new, more market friendly spatial policy initiatives such as the Urban Development Corporations (UDCs), Enterprise Zones (EZs), and localized enterprise agencies, subsequently rolled into Training & Enterprise Councils and other business support services (e.g., Business Links). The irony of these new 'local discourses' was not lost on the likes of CSE, but of course the 'Italianate' alternative economic strategies being practiced by the GLC were of such anathema to the Thatcher administration that in 1986 she personally arranged for the abolition of the GLC *tout court*. Fundamentally, the market mechanism became the key guide to economic development trajectories with a light hand on the interventionist tiller (e.g., UDCs and EZs) to smooth away impediments to market operations.

Numerous studies of restructuring effects upon localities were published in the UK and US, subsequently joined by work from other countries suffering a North–South divide such as Germany, albeit of lesser intensity than those appearing in the liberal market economies and much of the work of the 'California school' from Scott, Storper and Walker, and other Berkeley acolytes such as Meric Gertler, Annalee Saxenian, Amy Glasmeier, and Erica Schoenberger. Much of the aforementioned work reported postgraduate research that was being conducted contemporaneously with the UK work in the early 1980s and influenced somewhat by the presence as a visiting fellow of Doreen Massey hosted by Ann Markusen and others at the Berkeley Planning School. In the UK, collections such as those of Carney, Hudson and Lewis, Gregory and Urry, Thrift and Williams, and Allen and Massey opened up for discussion and exploration the nature of contemporary sociospatial differentiation and uneven development. The focus – often uniting geographers and sociologists – was upon theorized local distinctiveness under conditions of state invoked restructuring. These proved invaluable ingredients to the evolving 'localities' paradigm. By the 1980s, economic geographers worldwide were marveling at the innovation and creativity wrought upon the underwhelmingly descriptive modeling geography that had so successfully been critiqued and superseded. This was, of course, principally due to the development of interestingly sophisticated 'theory' which both engaged with and explained different spatial realities coherently, attracting economic geography stars of the future as it did so.

So what of the key findings of the actual CURS research program that stimulated so much debate, only a fraction of which can be touched upon here? The best place to look is the final chapter of Cooke which briefly summarizes the findings of the studies conducted. The first key finding was that the theoretical perspective adopted, taking an 'industrial restructuring through local labor market transformation' approach worked remarkably well, with rich quantitative and qualitative data that delineated internal and external articulations linking local and global spatial processes. In particular, second, the results confirmed something which a tradition of top down thinking in academe and government could never grasp, namely that localities were a little like laboratories in which unexpected compounds sometimes jelled with influential effect. It is noteworthy that at the end of the research local 'social capital' had been discovered, though we called it 'local proactivity'. This facilitated responses and initiatives by some, not all, localities to self improve their conditions of living and working within the constraint and opportunity sets locally available. Swindon and Liverpool were deemed managerially proactive policy wise to an increasing extent and at the time, late 1980s, many that had not been proactive much in the past, such as Cheltenham, Lancaster, and Thanet were becoming noticeably more politically so in the face of collective problems encountered. Industrial cities like Middlesbrough and Birmingham seemed the hardest hit in this respect, registering lower 'social capital' type effects than when they had been more prosperous. Third, there was much evidence of a rise in the 'networking' propensity of the localities, even the less proactive ones. Municipal enterprise, partnerships, and localized development agencies were among these more 'associative' forms then emerging. However, there was little sign then of 'clusters' though these seem to have burgeoned in many places in the subsequent decade with the decline of older corporate models, outsourcing of production, and the rise of collective entrepreneurship among innovative and creative small and medium enterprises (SMEs). Finally, these shifts seemed to presage something of a 'postmodern' localized developmental discourse as localities came to terms with the demise of the grand narratives of socialism, the welfare state, and the corporate 'job for life' assumption, not to mention the fragmentation typical of some of the worst de industrialisation instances.

Of the two critical positions in relation to this virtual hegemony in explaining change processes in economic geography that may be mentioned here, the first referred, bizarrely, to this collective, international, theoretically informed effort as a 'retreat' from theory and a new 'empirical' turn. The main burden of this critique was based on a fear that the "research would be unable to emerge from the morass of information" (Smith, 1987). However, as indicated, this fear was unfounded, not least because of the careful theoretical and methodological design of the research. Indeed what the critique signified was a fear on the part of the authors that data might 'disprove' theory, which given the ultraleftist positions of Neil Smith and his mentor David Harvey would have been disastrous for them. The 'scale' question reared up as it continues to do amongst this fraction, implying small is local and thus uninteresting while only big is important. We have noted above how one of the brethren, Dick Walker, did not fall into this 'spatial trap' and neither did the locality researchers. Of greater veracity was the bemoaning of a 'retreat from Marxist theory' (Smith, 1987) which, if true was prescient on our part, and if not true meant we were not allied with the somewhat over theorized version practiced by the critics in question. Of lesser direct import was the wide ranging critique, more focused on Cooke at that book's alignment of post Fordism with cultural postmodernism, the rise of consumption culture, and the rejection of grand narratives and modernist 'foundationalism'. Yet all these things have come to pass, whether always for the better is debatable. But we live nowadays in a media saturated world of 'floating signifiers' such as the celebrity obsession, reality TV, and rootless cultural production ('Pop Idol'). Bruce Springsteen's (2006) return to folk music may have marked the beginning of a movement to repair part of that 'postmodern' syndrome.

Conclusions: The Long Legacy of Locality Studies

Probably the longest lasting legacy of locality studies has been the rise of so called 'new regionalism'. Already spotted around the time of his return from Australia by Nigel Thrift this theorized regional political economy analysis was gaining ground rapidly as we have seen, in the new times of 'global localization'. The locality studies themselves and the comparative methodology that allowed spatial variety to be explained within a coherent and satisfying theoretical framework furthered this impulse. The next research project engaged in by the present author saw a return to the 'regional innovation systems' work first embarked upon with an SSRC fellowship in the early 1980s on this occasion comparing regional innovation in the UK and France. It concluded that indeed a process of global localization could be observed in both countries, despite their profoundly different 'varieties of capitalism'. This was followed by a further, definitively regional innovation systems comparison of four countries, which also identified and explicated the intraregional variety of industrial 'clusters' within regional innovation systems. Simultaneously, the first global comparison of regional innovation from a systems perspective derived from the 'localities' methodology was published, with a second edition arriving later to meet demand. Nowadays, of all the fields of innovation system studies, that focused upon regional innovation systems has the largest number of refereed publications. Of course, the range of studies of industrial clusters, following the first by regional scientist Stan Czamanski is now virtually innumerable, and by no means confined to economic geography.

Another field where there were resonances from the locality debates, and a further example of debate arising from the 'localities' methodology, concerned urban cultural and consumption research; what later became the so called 'cultural turn' in human geography. This arose largely from the opening perceived more strongly by the present author than most colleagues, with the possible exception of John Urry, into 'postmodern geographies'. How did this evolutionary element of 'related variety' emerge? First, recall the local discourses discoveries in the localities research findings and the 'antifoundationalism' associated with the demise of grand narratives associated with the modernity project in its various guises. That social model was undermined by verbal attacks from a more Gramscian left perspective represented to some extent in CSE positions, and it imploded under the neoconservative onslaughts of the Thatcher–Reagan era. The likes of David Harvey railed against all this, preferring, to his credit, not a Stalinist hierarchy as a preferred social model, but a nevertheless 'totalizing' discourse effacing social identity and variety, designed like Haussmann's Paris for maximum social control. The author debated this in person and in the special issue of *Society & Space* devoted to these and related issues then being uncovered by the locality debates and on the occasion of Harvey's somewhat underwhelming return to Britain to take a geography chair at Oxford. Both Nigel Thrift and the author noted his predilection to 'overtotalize' society, which he saw as utterly suffused by abstract capital and with no room for individuals. Even Marx did not go so far, it was pointed out. Later on, Simon Duncan asked the author to write a piece for *Geoforum* predicting the likely megatrends in geography for the 1990s. The author said it would be postmodernism and it would take the form of consumption studies, not least because the 'production' and labor market focus of the 'local restructuring' thesis had left especially urban geographers dissatisfied, not least given Castells' privileging of that sphere for urban analysis. In addition, it had to be admitted that the CURS localities were urban in the main, but although CURS contained in Chris Pickvance and Michael Harloe the first British academics to pick up on 'the French School', even they found Castells' structuralist Marxism hard to operationalize methodologically and the rest found it almost worse than Harvey.

By now work on 'creativity' has transmuted into studies of the 'creative class' that is as provocative a notion as 'localities' were a generation earlier. Lastly, debate still continues on the 'scale' question with a new generation of 'scalar envelopers' continuing to wish to efface spatial variety from a lofty and depressingly linear 'top down' perspective that has, naturally from an evolutionary economic geography perspective into which much of the 'localities' legacy has now comfortably folded, to be resisted.

See also: Network Regions.

Further Reading

Allen, J. and Massey, D. (1988). *The Economy in Question.* London, Sage.
Bathelt, H. (2003). Growth regimes in spatial perspective 1: Innovation, institutions and social systems. *Progress in Human Geography* 27, 789 804.
Best, M. (1990). *The New Competition*. Cambridge: Polity.
Braczyk, H., Cooke, P. and Heidenreich, M. (eds.) (1998). *Regional Innovation Systems*. London: UCL Press.
Brenner, N. (2001). The limits to scale? Methodological reflections on scalar structuration. *Progress in Human Geography* 25, 591 614.
Carlsson, B. (2006). Innovation systems: A survey of the literature from a Schumpeterian perspective. In Hanusch, H. & Pyka, A. (eds.) *The Elgar Companion to Neo Schumpeterian Economics*, pp 117 132. Cheltenham: Edward Elgar.
Carney, J., Hudson, R. and Lewis, J. (eds.) (1980). *Regions in Crisis*. London: Croom Helm.
Castells, M. (1977). *The Urban Question*. London: Edward Arnold.
Cooke, P. (1981). Tertiarisation and socio spatial differentiation in Wales. *Geoforum* 12, 319 330.
Cooke, P. (1983). *Theories of Planning & Spatial Development*. London: Hutchinson.
Cooke, P. (1985). Regional innovation policy: Problems and strategies in Britain and France. *Environment & Planning C: Government & Policy* 3, 253 267.
Cooke, P. (1987). Individuals, localities and postmodernism. *Environment & Planning D: Society & Space* 5, 408 412.
Cooke, P. (1989). Critical cosmopolitanism: Urban and regional studies into the 1990's. *Geoforum* 20, 5 17.
Cooke, P. (ed.) (1989). *Localities*. London: Unwin Hyman.
Cooke, P. (1990). *Back to the Future*. London: Unwin Hyman.
Cooke, P. (1992). Regional innovation systems: Competitive regulation in the new Europe. *Geoforum* 23, 365 382.
Cooke, P. and Morgan, K. (1998). *The Associational Economy*. Oxford: Oxford University Press.
Cooke, P. and da Rosa Pires, A. (1985). Productive decentralisation in three European regions. *Environment & Planning A* 17, 527 554.
Cooke, P., Heidenreich, M. and Braczyk, H. (eds.) (2004). *Regional Innovation Systems* (2nd edn.). London: Routledge.
Cooke, P., Moulaert, F., Swyngedouw, E., Weinstein, O. and Wells, P. (1992). *Towards Global Localisation*. London: UCL Press.
Czamanski, S. (1974). *Study of Clusters of Industries*. Halifax: Institute of Public Affairs, Dalhousie University.
Florida, R. (2002). *The Rise of the Creative Class*. New York: Basic Books.
Gertler, M. (1988). The limits to flexibility: Comments on the post Fordist vision of production and its geography. *Transactions of the Institute of British Geographers* 13, 419 432.
Glasmeier, A. (1988). Factors governing the development of high tech industry agglomerations: A tale of three cities. *Regional Studies* 22, 287 301.
Gregory, D. and Urry, J. (eds.) (1985). *Social Relations & Spatial Structures*. London: Macmillan.
Haggett, P. (1965). *Locational Analysis in Human Geography*. London: Arnold.

Hall, P. and Soskice, D. (eds.) (2001). *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage*. Oxford: Oxford University Press.
Harrison, B. (1974). *Urban Economic Development*. Washington: The Urban Institute.
Harvey, D. (1981). The spatial fix: Hegel, von Thunen and Marx. *Antipode* 13, 1 12.
Harvey, D. (1987). Three myths in search of a reality in urban studies. *Environment & Planning D: Society & Space* 5, 367 376.
Markusen, A. (1978). Class, rent and sectoral conflict: Uneven development in Western US boom towns. *Review of Radical Political Economy* 10, 117 129.
Massey, D. (1978). Regionalism: Some current issues. *Capital & Class* 6, 106 125.
Massey, D. (1979). In what sense a regional problem? *Regional Studies* 13, 233 243.
Massey, D. (1984). *Spatial Divisions of Labour*. London: Macmillan.
McLoughlin, B. (1969). *Urban & Regional Planning: A Systems Approach*. London: Faber.
Morgan, K. (1983). Restructuring steel: The crises of labour and locality. *International Journal of Urban & Regional Research* 7, 175 201.
Murray, F. (1983). The decentralisation of production: The decline of the mass collective worker. *Capital & Class* 19, 74 99.
Olsson, G. (1975). *Birds in egg*. Ann Arbor: University of Michigan.
Perry, D. and Watkins, A. (1979). *The Rise of the Sunbelt Cities*. London: Sage.
Piore, M. and Sabel, C. (1984). *The Second Industrial Divide*. New York: Basic Books.
Porter, M. (1998). *On competition*. Boston: Harvard Business School Press.
Pred, A. (1966). *The Spatial Dynamics of US Urban Industrial Growth,* 1800 1914. Cambridge: MIT Press.
Regional Studies Association (1983). *Report of an Inquiry into Regional Problems in the United Kingdom*. Norwich: Geobooks.
Saxenian, A. (1994). *Regional Advantage*. Cambridge: Harvard University Press.
Sayer, A. (1976). A critique of urban & regional modelling. *Progress in Planning* 6, 191 254.
Sayer, A. (1986). New developments in manufacturing: The just in time system. *Capital & Class* 30, 43 72.
Schoenberger, E. (1988). From Fordism to flexible accumulation: Technology, competitive strategies and international location. *Environment & Planning D: Society & Space* 6, 245 262.
Scott, A. (1988). *New Industrial Spaces*. London: Pion.
Smith, N. (1987). Dangers of the empirical turn: The CURS initiative. *Antipode* 19(1), 59 68.
Storper, M. and Walker, R. (1989). *The Capitalist Imperative: Territory, Technology & Industrial Growth*. Oxford: Blackwell.
Tabb, W. and Sawers, L. (1984). *Sunbelt/Snowbelt*. Oxford: Oxford University Press.
Thrift, N. (1983). On the determination of social action in space and time. *Environment & Planning D: Society & Space* 1, 23 57.
Thrift, N. (1987). No perfect symmetry. *Environment & Planning D: Society & Space* 5, 400 407.
Thrift, N. and Williams, P. (eds.) (1987). *Class & Space: The Making of Urban Society*. London: Routledge.
Walker, R. (1978). Two sources of uneven development under advanced capitalism: Spatial differentiation and capital mobility. *Review of Radical Political Economy* 10, 28 39.
Walker, R. (2001). Bennett Harrison: A life worth living. *Antipode* 33, 34 38.

Location Analysis

M. W. Horner, The Florida State University, Tallahassee, FL, USA

Glossary

Applied Geography The branch of geography stressing the practical application of geographic methods and concepts to real-world problem solving.

Geodemographics Field dealing with the development of methodologies for estimating the spatial variation in the socioeconomic and demographic characteristics of consumers.

Geographic Information Systems for Transportation (GIS-T) Specialized GIS software for dealing with transportation problems.

Gravity Model A model adapted from the natural sciences to help explain how people interact in space.

Operations Research Field concerned with the development of procedures for optimizing the function of various human activities.

Introduction

Location analysis addresses the spatial aspects of siting, managing, and evaluating private and public facilities, as well as investigates why urban economic systems are structured the way they are. Issues considered by location analysts in the private sector often include where to open a new grocery store, place a package distribution center, and locate various other retail activities. In the public domain, location analysts look at problems of siting new libraries, health centers, recycling depots, parks, and other community goods, based on some objective, quantifiable criteria. Implicit in many location analyses is the need to estimate the geographic market areas associated with certain planned or existing facilities.

Location analysis occupies a unique position within geography, the social sciences, and wider academic disciplines. Most geographers that study location analysis tend to have training in economic geography, urban geography, or transport geography. Additionally, geographic information systems (GISs) have provided analysts increased flexibility to build more sophisticated and in depth studies of location issues. Moreover, it is also not uncommon for location analysts to undertake training in advanced statistical and mathematical techniques. Many location problems are approached by determining an optimal, or best location (or set of locations), for facilities. Thus, location analysts routinely employ methods found in operations research and the engineering disciplines. Along the same lines, many scholars outside of geography have interests in location analysis. These people can be found in academic departments including, but not limited to, economics, ecology, civil engineering, industrial engineering, statistics, and others.

Location analysis as a professional field is one in which many geographers find employment outside of academia. Today, virtually all major retailers in the US, for example, engage in some form of location analysis to help improve their profitability. Analysts are hired to perform such tasks as determining possible locations for new stores, identifying potential customers to receive advertisements via mail, and recommending under performing stores for closing.

Historical Developments

The present field of location analysis has been shaped by a substantial body of work. Collectively, these contributions form the basis of what is known as 'location theory'. Location theory provides the conceptual and intellectual foundation of location analysis. Some of the earliest work in this vein was conducted in Germany by Johann Heinrich von Thunen, who in the early nineteenth century postulated agricultural rent theories relating market accessibility to transportation costs and land prices. Research in the first half of the twentieth century progressed forward, establishing several important principles for the delineation, analysis, and expectations about the geographies of markets. Important contributors to this growing field were Walter Isard, Augustus Losch, Alfred Weber, and Walter Christaller.

For example, during the early 1900s Alfred Weber laid a theoretical framework to help understand how industries choose to locate their facilities. Weber's research discussed the role of transportation and production costs in determining facility locations. One observation that arose from his work was that if production costs are assumed to be homogenous across the choice of sites, then the decision of where to locate the firm will be solely driven by transportation costs, such as the costs of moving raw materials to the firm for processing. This insight from his work influenced the structure of many location models developed heretofore, especially in the sense that transportation costs were the primary criteria for choosing locations. As a second example, Walter Christaller's work during the 1930s on establishing a 'central place theory' sought to explain the geographical patterns of settlements, while recognizing the hierarchical nature of places and their respective markets.

Christaller suggested that settlements would cast influence over nearby lands in relation to their relative sizes and scope of goods they sold.

Trade Area Delineation

Ideas from central place theory are routinely extended to determine trade areas for retail stores. A straightforward method of estimating the trade areas for very similar retail stores is to assume that people will travel to the nearest facility. Clothing retailers might make this assumption when deciding how to directly mail sale advertisements to prospective customers. Such trade area analyses are routinely conducted using GIS. As an example, **Figure 1** shows the locations of several retail stores along with an estimated trade area boundary around them. The area enclosing each store represents the area closest to that store. These special polygons are known as Theissen polygons.

Analytical techniques and theories inspired by the gravity model from physics and the natural sciences made their way into human geography early in the 1900s and have continued to draw attention in the literature. The basic gravity model of planetary physics stipulates that the interaction between two masses is proportional to the product of their masses and inversely related to the square of the distance between them. By substituting population size or some other activity attribute for the 'masses', one may estimate the interaction between two locations. Formally, the equation for the basic gravity model is

$$I_{ij} = \frac{P_i P_j}{d_{ij}^2} \qquad [1]$$

where I_{ij} is the interaction between locations i and j, P_i is the population at location i, P_j is the population at location j, and d_{ij} is the distance between locations i and j. One of the most significant aspects of gravity models (or as they are now known in geography – spatial interaction models) is that they formally specify the relationship between distance and the likelihood of interaction. Holding all other effects constant in the model [1], as distance increases, interaction decreases. This phenomenon is known as distance decay and is inherent to virtually all gravity models.

In empirical terms, interaction (I) in the gravity model refers to migration flows of people, commute to work trips, or some other form of human movement. The size of a population center (P) might be measured as the number of people living there, although when gravity-type models were first used in location analysis they were modified to contain terms that accounted for the

Figure 1 Example of market areas as based on Theissen polygons (Downtown Austin, TX).

characteristics of public and private facilities. For instance, characteristics of stores such as their square footage could be used to proxy their size or attractiveness.

Many extensions to the gravity model have been devised. For example, Reilly modified the basic gravity model such that it produces an estimate of the geographical point at which one location's dominance would give way to another. Given two centers of unequal size or status (e.g., cities), Reilly's 'law of retail gravitation' finds the exact geographical extent of the respective market areas they dominate. In notation this is

$$MB = \frac{d_{ij}}{1 + \sqrt{\frac{P_j}{P_i}}} \qquad [2]$$

where MB is the market breakpoint, and all other notation is as defined previously. This extension of the gravity model was important because it took into account the size of the centers (e.g., their population) and used that information to shape their market areas, whereas earlier approaches to this problem treated the centers as having equal weight and suggested the change in market area dominance would occur halfway between the locations. Other extensions to the basic gravity model looked to identify the 'potential' of areas as being markets for goods and services. An early contribution by C.D. Harris in the 1950s adapted the basic gravity model to quantify the market potential of areas. Formally, this is

$$A_i = \sum_{j\ 1}^{m} P_j d_{ij}^{-1} \qquad [3]$$

where A_i is the market potential of location i, m is the number of observational units, and all other notation is defined as previously. The measure in [3] is best thought of as an index that captures how accessible a location is to the population, P, at other locations. This has been a widely applied model because a variety of attributes can be used in place of P, such as retail square footage, total disposable income, total employment, etc. An example of a potential index appears in **Figure 2**.

The figure depicts a potential index as estimated for all counties in the continental United States, using the 1990 Census county population for P. The index is interpreted such that higher values indicate areas of greater market potential.

The pioneering efforts of Alan Wilson during the 1960s and 1970s formalized and synthesized some of the earlier work on gravity models and placed them on firmer theoretical footing by drawing parallels and connections between geography and statistical mechanics. Among Wilson's contributions, he specified a 'family' of

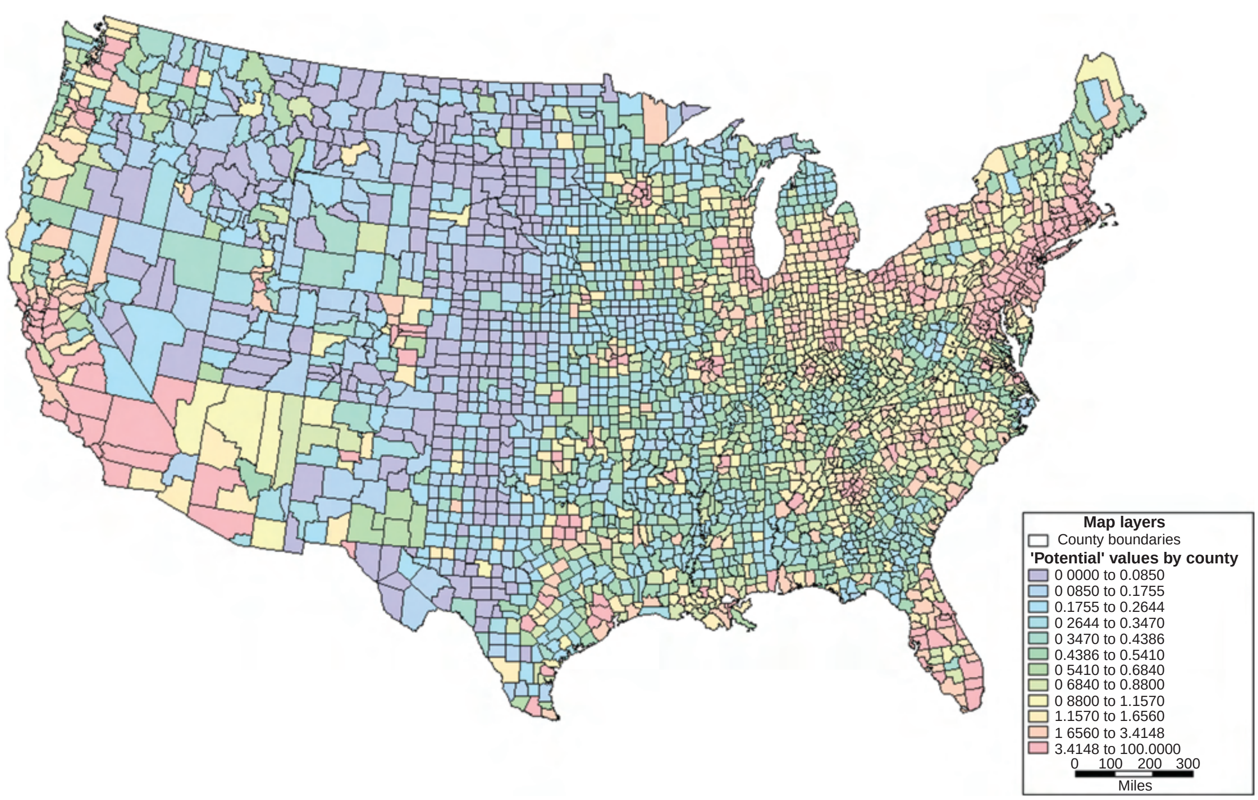

Figure 2 An example of a county level 'market potential' index.

spatial interaction models. This framework of spatial models was capable of modeling any one of four basic scenarios presented to the analyst. Two of these situ ations were particularly of interest to retail location analysts. Given a known population scattered across neighborhoods in an urban area, the 'origin constrained' spatial interaction model allows the analyst to predict which facility consumers will visit. Predictions by this type of model are constrained such that the number of people leaving each neighborhood is the same in the observed and predicted case. The 'destination constrained' model is quite similar, except that the des tination information is constrained such that the observed and predicted totals match. An example of this situation would be when a grocery retailer had data on the number of customers visiting its stores and wanted to ensure that any predictions about where its customers were coming from used these known counts.

Research on spatial interaction modeling has pro gressed since Wilson, with many of its developments impacting location analysis. For example, Alonso's theory of movement published in the 1970s further generalized the spatial interaction model, thereby allowing more modeling possibilities. As a second example, Fothering ham's competing destinations model formulated in the early 1980s recognized the interdependence of the at tractiveness of destinations and altered the spatial inter action model accordingly.

Normative Modeling

The field of location analysis also benefited from devel opments in normative modeling. Normative modeling is prescriptive in nature, looking to suggest a best case or optimal solution for a given decision situation. This is accomplished in location analysis by utilizing formal mathematical models for finding the optimal solution for a particular objective. For instance, Hitchcock in the 1940s formulated procedures for solving the transporta tion problem, a classic model used to minimize the total costs of moving goods between supply and demand lo cations. Often, an empirical shipment pattern or allo cation would be compared to the optimal allocation determined by the model in order to point out any in efficiency. In this sense, the optimal solution from the transportation problem is used as a diagnostic tool to guide future decision making.

One of the challenges faced by location analysts is the relative difficulty of solving some problems in an 'optimal' fashion. Essentially, as the number of observations in some of these problems grows even moderately large, a point is reached where the number of possible answers explodes. A standard facility location problem where the analyst must choose ten sites from a set of 150 candidates has more than 1.7×10^{15} possible combinations, with one or very few of them the 'best'. Fortunately, there have been improve ments in computing power coupled with the advent of specialized computer programs capable of intelligently solving such problems. These have given location analysts the means to approach such difficult problems.

Many of the spatial models that are developed in research and applied in practice today can be linked to concepts and ideas from classic location theory. The *p* median problem and the multifacility Weber problem are used to find optimal locations for new facilities such that transportation costs are minimized. These models can be formulated in either discrete (*p* median) or continuous (Weber) space. Continuous space models assume that facilities can be located anywhere in the study area. Discrete space models restrict facility place ment to a limited number of candidate sites, such as nodes on a network. Because most real world facility location problems more closely mimic the discrete case (i.e., facilities cannot be placed anywhere), continuous approaches are rarely used in practice. The *p* median model is among the most frequently applied discrete facility location problems. It optimally sites *p* facilities (where *p* is a user specified number) such that expected transportation costs between customers and facilities are minimized. It has been used in studies ranging from opening new grocery stores in US cities to setting up community health centers in developing countries.

Another important spatial model emanates from the need for one type of facility to provide 'coverage' to some other entity. Coverage in these models is usually deter mined based on whether an entity is within some pre determined distance or drive time of the facility providing the coverage. Although the literature on cov erage models is vast, there are two basic approaches found in practice. One is the 'location set covering model', where the objective is to provide complete cov erage to populations using a minimum number of facil ities. For example, this situation arises in planning for tornado warning siren systems. There, a minimum number of sirens must be sited such that all people in a study area are close enough to a siren to be able to hear a warning. The 'maximum coverage location model' also uses the concept of distance coverage, but it seeks to locate a fixed number of facilities such that the popu lation covered by those facilities within a given distance is maximized. In this sense, the maximum coverage model is very useful because many applications are constrained by the number of facilities that can actually be placed. For example, variants of the maximum cov erage problem can be used to select stop locations for new commuter rail facilities. In such a scenario, the number of stops to be sited is limited (having too many stops slows down the train), and it is assumed that sited stops 'cover' all of the demand for rail transit that is within some prespecified distance of them.

For many in geography and the location analysis community, these types of modeling approaches are col lectively referred to as location–allocation models. This is because models such as the *p* median involve a location component (where to place facilities) and an allocation component (how demand is allocated among the facilities). Many geographers and mathematical modelers have been active in location analysis and location–allocation model ing. Charles ReVelle was among the most prolific con tributors to the field. Beginning in the late 1960s ReVelle was involved with the creation of many location models for addressing problems in the private and public sectors. ReVelle became well known for his work with *p* median and coverage models. ReVelle's work was notably broad, as he created new approaches to solving location problems with both human and environmental implications.

Current Trends in Location Analysis

Location analysis research and applications of today differ from that of 30, 20, or even 10 years ago. One of the major developments is the recent widespread availability of highly functional, low cost GIS and related computing technology. Many location problems can be approached in a sophisti cated fashion from the confines of a desktop computer. Related to both the growth of GIS and spatial modeling, there has been continued development of the tools and techniques used to conduct market area analysis. Some of these tools take advantage of the increased power afforded by GIS to accurately represent real world environments including population demographics, transportation systems, and facility locations. Lastly, the growth and deployment of the commercial Internet has profoundly influenced the way location analyses are currently practiced.

Trade Area Analysis, Geodemographics, and GIS-T

Location analysts frequently need to determine the ex tent of trade areas, identify the characteristics of people living within those trade areas, and relate this infor mation to key variables about the facilities, such as their sales, production, etc. Trade area analyses have become more sophisticated as the range of computer based tools available to analysts has improved in quantity and quality. The ability of GIS to represent detailed trade area geographies and measure the characteristics of people within them to the household level makes it an important tool for businesses wanting analyze their markets.

Two principal developments have helped facilitate more robust trade area analysis. First, there has been an increase in the amount of detailed spatial data available about consumers, their characteristics, and their product preferences. Some of these data are obtainable from public sources, but many private data vendors sell geo demographic data. Much of these data are used for market segmentation purposes where the analyst seeks to determine if there is a pool of demand for a particular product based on the area's demographics. Geodemo graphic vendors build profiles of people in areas that account for their income characteristics, cultural afflu ence, political leanings, and many other dimensions.

Second, specialized goegraphic information systems for transportation, referred to as GIS T, have contributed to more insightful market area analysis. GIS Ts are par ticularly adept at managing and analyzing geographic data that involve transportation features. Because road net works are important for gaining realistic estimates of drive times or distances between stores and customers, as well as good distance measures between locations, GIS T has played a growing role in location analysis. **Figure 3** shows an example of a travel distance contour as constructed in a GIS T. It shows travel distance contours along a net work from a selected intersection (1, 2, 3, 4, and 5 miles).

The Commercial Internet

The rise of the commercial Internet over the last decade has had several implications for the field of location an alysis. Among the most prominent impacts, there has been a marked change in the importance of and restrictions imposed by distance in consumer behavior. Lower cost retail items such as books and music are routinely bought online, and increasingly, high cost items, such as auto mobiles, are also being purchased online. Distance is less of the deterrent it used to be. This means that trade area estimation methods that emphasize distance are less representative of consumer behavior today. This has prompted analysts to study the effects of the Internet on consumer behavior and develop new trade area methods.

Research on the structure and function of the net works comprising the Internet itself has attracted the attention of many location analysts. In the post 9/11 US, concerns about security enter into every aspect of society, including ensuring that adequate data transfer capabil ities are maintained in the event of a crisis. Thus, many location analysts have spent time developing method ologies that measure the Internet's capability to with stand catastrophic damages, as well as techniques for placing facilities to protect network infrastructure.

Contributions to Human Geography

The field of location analysis has made theoretical and applied contributions to human geography. On the

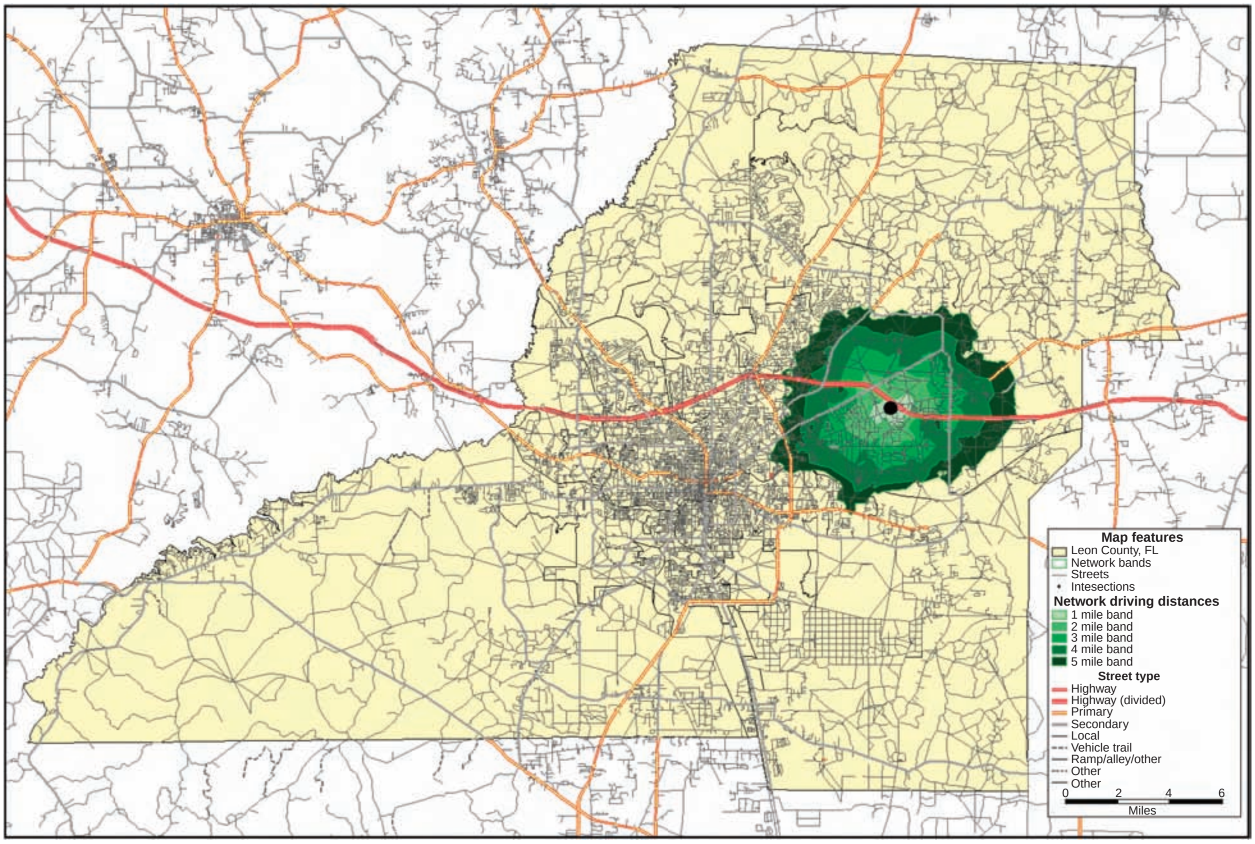

Figure 3 Travel distance bands of 1 5 miles from a selected intersection in Leon County, FL.

theoretical side, the basic insights into markets, places, and hierarchies that were established many decades ago persist today as constructs used to shape work in location analysis. Moreover, these ideas continue to influence students in introductory human geography courses, as well as more specialized economic and urban geography classes. In short, early research in location analysis con tributed substantially to the development of the view of geography as a spatial science that is concerned with the spatial organization and interaction of human activities.

Unquestionably, the techniques developed, adopted, and used by location analysts are largely quantitative in nature and inextricably linked to the quantitative revo lution in geography. Many of the statistical techniques popularized during the quantitative revolution, such as correlation analysis, multiple linear regression, and factor analysis, were advanced in location analyses through in vestigations into the relationships between cities, mar kets, and the hinterlands they dominate. Location analysts also gave human geography, via the quantitative revolution, insights into how spatial interaction could be measured and theorized. In that regard, geographers had a major impact on the development and refinement of the gravity model (spatial interaction model), which at tracted broad interdisciplinary interest. Location analysts played an important role in initiating a dialog with economists, engineers, and planners who were also working to adapt spatial interaction models to real world problem settings, such as highway traffic prediction and market area studies.

Location analysis has also been of interest to human geographers who consider themselves a part of the ap plied geography tradition. Applied geography stresses the application of geographical techniques and method ologies for practical problem solving. Through the ap plied geography interface, there has been exchange and collaboration between businesses and academics. The Applied Geography Conference (AGC) is one example of an annual forum where academics and business leaders meet to discuss research using location analysis techni ques. Because of the strong applied connection that lo cation analysis has enjoyed, students trained in location analysis have had success gaining employment in the private sector, which in turn has brought recognition to academic geography departments.

Retrospect and Prospect: Location Analysis Moving Forward

No field is without its limitations, and location analysis has been critically questioned on a number of grounds.

Many of the early models and theories developed for location analysis were crafted assuming a featureless, isotropic plane. In this sense, the geographic factors that might have affected location decisions, such as topo graphy, were not accounted for in the models. On the one hand, the rationale for making such simplifying as sumptions was to facilitate a formalized, rigorous treat ment of these location scenarios. This approach is similar to that in the natural sciences where, for example, the interaction of two atomic particles might be explored in a controlled experiment (e.g., in a vacuum chamber). However, many geographers saw the stylized scenarios from which these theories were derived as detached from real world contexts. Critics question the validity of these theories for explaining actual location decisions.

Location analysts have benefited from such criticisms. Analysts are developing more realistic location models that incorporate detailed population characteristics, transportation routes, and so on. GIS has been instru mental in this regard, because it has allowed location analysts to collect more detailed geographic data on which to base their models. There is a trend in the GIS and location analysis communities toward using more representative, elemental spatial data, when possible, thereby avoiding the difficulties and errors induced by using aggregate, low detail spatial datasets.

Location analysis was also criticized in human geog raphy for its strong adherence to positivistic method ologies, as post positivistic approaches became more widely accepted from the 1970s onward. Some very good criticisms were leveraged of location analysts' use of statistics, especially the use of techniques such as 'step wise' multiple regression, which seemed to be motivated by pragmatism rather than an underlying concern for the factors that might explain actual location behaviors. However, users of multivariate statistics today, including location analysts, have a more nuanced grasp of the ap propriate steps and considerations for model develop ment. This is evidenced in the voluminous body of work that seeks to account for the confounding and compli cating role of space in regression analysis.

In a related set of criticisms coming out of the quantitative revolution, location analysis in general was derided by post positivists for its close contacts and working relationships with governmental and business interests. Such alliances were out of sync with a sense of social relevance sweeping human geographic research in the late 1960s and 1970s. Location analysts were seen as perpetuators of the status quo and its capitalistic en deavors rather than taking up higher societal callings such as addressing social and environmental inequities. However, today location analysts investigate a much more diverse set of issues. Work for business and industry persists, but many location analysts explore socially relevant topics such as access to healthcare, equitable public service provision, and the protection of environ mentally sensitive lands.

Prospects for sustained interest in location analysis are good. The proliferation of GIS technologies has brought location analysis techniques to the masses. Lo cation analysts study a diverse set of problems and their insights continue to be in demand from government and industry.

See also: Hub Network Location; Location Theory; Network Analysis; Spatial Interaction Models.

Further Reading

Berry, B. J. L. and Parr, J. B. (1988). *Market centers and retail location: Theory and applications*. Riverside, NJ: Simon and Schuster.

Daskin, M. (1995). *Network and discrete location: Models, algorithms, and applications*. New York: Wiley.

Dawson, J. A. (ed.) (1980). *Retail geography*. New York: John Wiley and Sons.

Fotheringham, A. S. and O'Kelly, M. E. (1989). *Spatial interaction models: Formulations and applications*. Boston: Kluwer.

Greene, R. P. and Pick, J. P. (2006). *Exploring the urban community: A GIS approach*. Upper Saddle River, NJ: Prentice Hall.

Haggett, P. (1965). *Locational analysis in human geography*. London: Edward Arnold.

Horner, M. W. and O'Kelly, M. E. (2005). A combined cluster and interaction model: The hierarchical assignment problem. *Geographical Analysis* 37(3), 315 335.

Miller, H. J. and Shaw, S. (2001). *Geographic information systems for transportation*. New York: Oxford University Publishers.

Rodrigue, J. P., Comtois, C. and Slack, B. (2006). *The geography of transport systems*. New York: Routeledge.

Smith, D. M. (1981). *Industrial location*. New York: John Wiley and Sons.

Wilson, A. G. (1970). *Entropy in urban and regional modeling*. London: Pion publishers.

Location Theory

A. T. Murray, Arizona State University, Tempe, AZ, USA

Glossary

Agglomeration Economies Reductions in costs are possible by locating near-similar or related industries.

Co-location Two or more businesses (or industries) positioned next to or near each other.

Hierarchy Spatial arrangement of towns (or goods and services providers) where high-order items are found in select, larger places and low-order items in all places.

Market Area Geographic footprint representing the subregion where a good or service is consumed.

Optimum Location Geographic site(s) minimizing (or maximizing) a stipulated utility function or criteria.

Range Maximum distance that people will travel to consume a good or service.

Spatial Equilibrium Arrangement of businesses where each maintains a necessary threshold of customers.

Threshold Minimum market (or consumer base) needed to support a business.

Introduction

Location theory has long been a central component of economic geography, involving part explanation and part prediction regarding the location and interaction of people, goods, and services in both the private and public sectors. The explanation component arises when questions of 'why' and 'how' spatial patterns of activity have evolved over time. The prediction aspect comes about when certain knowledge is used to place, site, relocate, or otherwise plan for future service provision.

Location theory is the basis and/or context for studying how and why location decisions are made, either by companies, government agencies, or people, as well as provides the rationale for siting decision making and service allocation. This can involve the examination of access to goods and services, economic efficiencies, transportation considerations, service provision, etc. Classic work, as will be detailed below, has focused explicitly on land use and transportation costs, primarily in product manufacturing and delivery. More recent work in location theory builds upon this, but also has been broadened to address a range of issues, including assumptions made, model abstraction, computational complexity, as well as others. Given insights and strategic goals, location theory serves as the foundation for siting decision making and/or resource allocation, where modeling approaches are operationalized to support facility placement and service planning efforts.

Location theory is operationalized, or used in practice, through location analysis. While not the focus of this article, there are many methods that support location analysis, and briefly we note the three following types: (1) exploration of spatial information and spatial interaction, (2) descriptive assessment, and (3) normative (or prescriptive) decision making. At the most basic level, geographic information systems (GISs) and/or other cartographic mapping approaches can be used to explore spatial information associated with an urban or regional system of interest. This might involve qualitative assessment through mapping coordinates of entities (spatial objects), as well as the overlay of related spatial information, and could also entail a spatial examination of interacting flows (e.g., journey to work travel, market areas, migration, etc.). Moving toward greater quantification is the formal description of spatial arrangements. Examples of such descriptive methods include fitting distance decay functions, deriving location quotients, assessing excess commuting, among many others. Finally, there is the process of identifying the best locations and/or service patterns, which is known as the normative (or prescriptive) component of location analysis. Normative approaches have been used to plan for and site new fire stations, distribute emergency warning sirens, locate new retail outlets, align oil/gas transmission corridors, consolidate schools, and many others. Both the descriptive and normative elements of location analysis have traditionally been supported by quantitative methods and models, an overview of which is discussed elsewhere in this encyclopedia.

Explanation

In many ways location theory is the outgrowth of various aspects of economic theory, where encountered nuances like spatial heterogeneity, neighborhood effects, and human behavior and decision making were sought to be more fully understood. The foundations of location theory are generally regarded as including a focus on land use, industrial production, central places, and spatial competition. In fact, typically one person is attributed the progenitor of the major ideas and concepts for each. Specifically, the following are those individuals and their corresponding contribution area: land use – Johann von Thünen (1783–1850); industrial production – Alfred

Weber (1868–1958); central places – Walter Christaller (1893–1969) and August Lösch (1906–45); and spatial competition – Harold Hotelling (1895–1973). In what follows, an overview of each area is presented, high lighting some basic principles and assumptions.

Land Use

Observations regarding land use were fundamental to the location theory work of Thünen. Of particular interest was that agricultural land uses changed in systematic ways relative to the central market serving an area. The explanation for this systematic organization of space was economic in the sense of revenue potential. Given the price of land, costs for labor/inputs, and the associated distance to transport the harvested (or manufactured) product to the central market, it simply was not feasible for certain land uses to be too close to the central market. Whether a particular good warranted closer proximity to the central market depended on the price that could be obtained, less the costs involved to produce and transport it (land, labor, transportation, etc.).

The basic organization of a land use system observed by Thünen is summarized in **Figure 1**. These so called concentric circles reflected the general pattern of land uses at the time. What was most significant about this spatial organization is that there was an economic in terpretation to the pattern. In particular, the price of land, or the land rent, was higher closer to the city center (the urban industrial area). Taking into account that production/operating costs (including labor), transpor tation costs, and land rent all dictate what one would have to sell the product for to either break even or make a profit. Given high land rents in urban areas, as an ex ample, it likely would not make economic sense to grow rye in the city center as the product income would not be sufficient to make a living. The relationship between land rent and distance from the city center for a product (rye in this case) is simplified in **Figure 2**, reflecting the higher land rent closer to the city center. The sales price of rye would need to be higher if it were grown closer to the urban center, but could be lower if the rye is grown further away because of lower land rents (in addition to increased transportation costs). Different products have different land rent curves, thereby serving as an organ izing principle for various (competing) land uses.

The observations of Thünen are clearly based on a number of simplifying assumptions. One is that an isol ated state is assumed, where there is one market having no outside interactions (trade). Another assumption is that there are homogeneous land characteristics, where the land surrounding the market is entirely flat with

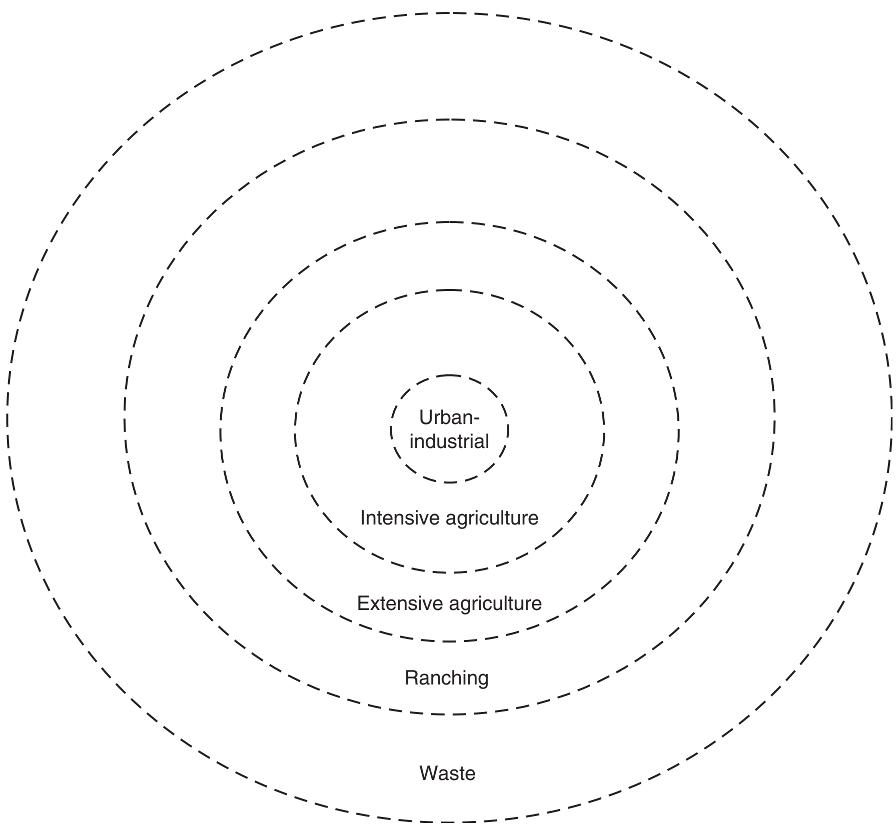

Figure 1 Land use system.

uniform fertility and absent of transportation infra structure, or rather that transportation is possible from anywhere to anywhere in any direction. With respect to human behavior, it is assumed that a farmer is interested in profit maximization.

The significance of the work of Thünen is that there is some rationale to spatial organization, and economics play an important role. While one can take issue with the simplifying assumptions made (and many have), the observed principles remain. Much current research continues on land use, examining organizing features, extending economic insights and relationships, and ex ploring formal approaches to model land use change.

Industrial Production

The major contribution of Weber to location theory was his work on industrial production. Of particular interest was where an industry (or firm) would locate in a region given material (and product) transportation costs, labor costs, and agglomeration economies. An important facet of Weber's work was formalizing how the optimum lo cation could be identified when cost of transportation was minimized. Weber simplified the problem as follows: given two locations from which raw materials (inputs) are obtained and one location (market) where the fin ished product is to be taken, find the optimal site for locating the industry (or factory). This basic problem is illustrated in **Figure 3**. For each given location (inputs and market), we know the quantity of materials to be moved and costs to transport are a function of the dis tance to be traveled.

While others are also credited with their work on this problem (Battista Calieri, 1598–1647; Pierre de Fermat, 1601–65; Evarista Torricelli, 1608–47; in particular), it was the formal and rigorous treatment by Weber in 1909 with respect to industrial location that is noteworthy. As illustrated in **Figure 3**, the transportation cost in this case is a function of the straight line (or Euclidean) distance and the amount of material transported, or rather a weighted travel distance. The optimum location relative to material transport is the point that minimizes this total cost. Such a point represents a spatial equilibrium, and is the most efficient placement of an industry with respect to transportation costs. To solve such a problem, one could consider each known location (inputs and market) as representing a force vector. A known location exerts a pull on the optimal location relative to the quantity of materials to be transported over some distance. With the three locations each exerting such a force, the point where the forces equal represents the optimal location. That is, an equilibrium state is reached such that no force vector is greater than the others. This location, producing an equilibrium state, is the location where total trans portation costs are minimized, thereby making it the optimal location. This point could be found using a mechanical model called a Varignon frame. It turns out that a mathematical approach exists as well for solving this problem, basically that of Weiszfeld in 1937, but will not be further detailed here.

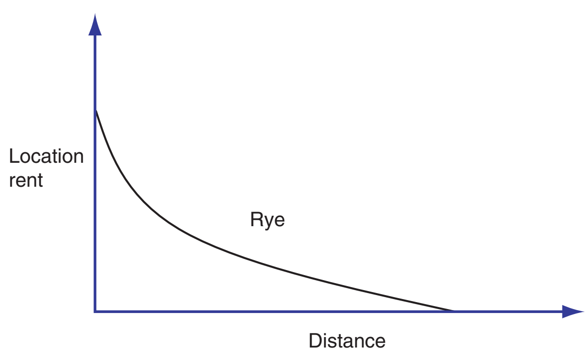

Figure 2 Land rent curve.

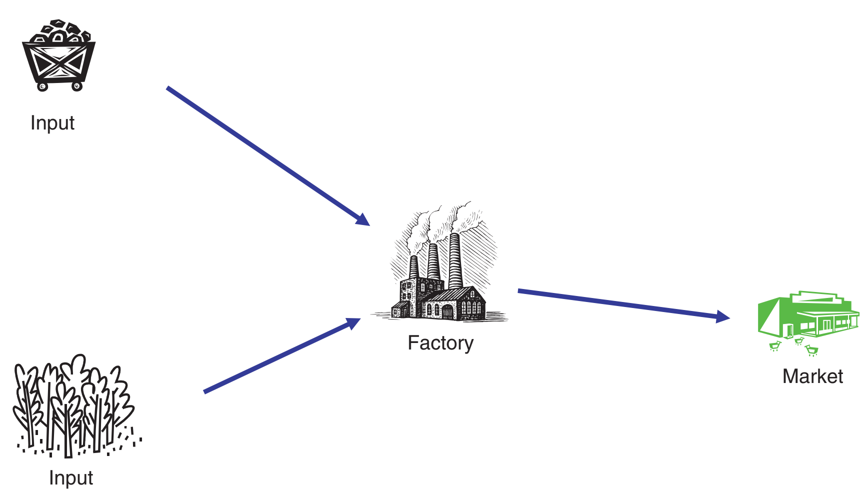

Figure 3 Transportation cost factors in industrial location.

As noted previously, labor costs and agglomeration economies were also of interest to Weber, and were seen as important facets of industrial production. Both issues were observed to be factors that could influence the optimal location of an industry, but were given less significance than the transportation cost factor. Of course if labor costs are relatively uniform, then there would be no influence. If they are not uniform, then a trade off between transportation costs and labor costs must be addressed. This is much the case with economies of agglomeration, where some reductions in costs are possible by locating near similar or related industries. However, this too is a situation where there would be trade offs with transportation and labor costs.

The work of Weber relied on many assumptions, either explicit or implicit. The more obvious were that raw material locations are given (known in advance) and the location and size of markets is given (known in advance). Relative specifically to transport cost, there is a reliance on transportation being a function of Euclidean distance. Beyond this, there was no competition and uniformity in culture, economic system, and political regime is assumed. There are additional assumptions regarding labor and agglomeration as well.

Of significance in the work of Weber on industrial production is formalizing the concept of transportation cost minimization, and establishing principles for identifying an optimum location. Also, recognizing potential trade offs between transport costs, labor costs, and economies of agglomeration were fundamental to location theory. Much research has been devoted to these themes, exploring their significance and expanding upon them. Ultimately one is left with direct connections to spatial efficiency, and equilibrium, in the location of industries, a notion that is similar to derived aspects of economic theory.

Central Places

The location theory contributions of Christaller in 1933 with respect to central places are well recognized, as is the work of Lösch in 1940. This work was interested in describing the relationship between central places (cities and towns) and the hinterlands they served. Of course such relationships are the result of urban evolution. In this regard, ordering principles of consumed goods and services were observed and described as hexagonal patterns, with cities and/or towns being in the center of associated hexagons. While Thünen was interested in the spatial organization of land uses, the focus of central places was on the distribution and/or use of goods and services. A good, say clothing, is produced and made available in a city, as an example. The demand for this good will be a function of its price and the travel cost for a consumer to purchase it. For a company to survive, and earn a living, it must sell its product. This means that some sustainable level of consumption of the good must be maintained. Given this, two important concepts emanate from the theoretical work of Christaller (and Lösch): threshold and range. The threshold of a good or service is the minimum market (total consumption) that is needed to bring a new firm (or producer or service provider or city) into existence and keep it operating. The range of a good is the maximum service distance that people will travel to buy that good or service. Both threshold and range reflect an explicit focus on demand for a good or service. These concepts can be further understood through the depiction of range in **Figure 4**, where the location of the business is shown (center) and the maximum travel range is indicated. This area, a circle in this case, reflects a potential trade area of businesses based upon the range. For the business to survive it must have sufficient demand for its product within this area. If it does not, the business will cease to exist. From a spatial perspective, the threshold should be achieved before the range is reached, if a business is to be viable.

At a broader spatial scale one is likely to see the emergence of patterns relative to range and threshold, and such patterns for a single good or service across a region are expected to be in the form of hexagons, as shown in **Figure 5**. Thus, a particular good or service would see a pattern similar to that shown in **Figure 5** under certain conditions, and underlying this are the associated ranges and thresholds because the independent businesses would not continue to exist unless each maintained the minimum customer base needed to make a living. The hexagonal pattern emerges because a viable business will maintain its customer base (threshold) within the product range and over the long term such

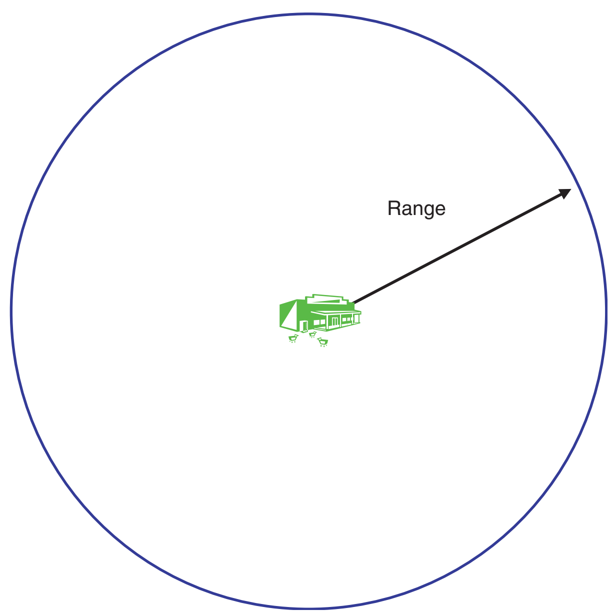

Figure 4 Product range.

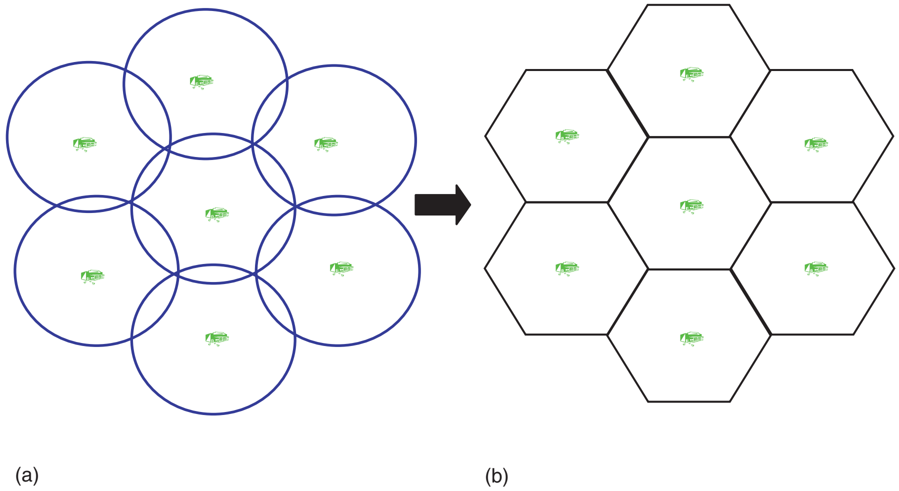

Figure 5 Hexagonal market area pattern.

patterns represent equilibrium states, reflecting a sus tainable number of businesses in a region.

Another facet of central places is the notion of hier archies. If this is viewed in terms of towns/cities, what we would see across a landscape are different size cities, not similarly sized as shown in **Figure 5**. Thus, there are small, medium, and large cities, and those in between as well, but in addition to the patterning in **Figure 5** there are hierarchical relationships that exist between these cities. In general, with respect to cities, larger centers offer more goods and services, so the range for these items may be larger. This means that smaller cities are offering less variety, and focus on more regularly used (or low order) goods and services. These more regularly used items are found in nearby larger cities as well, but the transportation costs to acquire them are prohibitive. Beyond this, sufficient demand exists for such goods and services in local areas (suburban or rural), meaning a necessary minimum threshold level of demand exists and it is priced reasonably compared to the price in a larger city that is further away (including transportation costs). Examples of high order goods and services found in larger cities might include government functions, farm equipment, higher education, specialized health care, etc.

The work on central places is not devoid of any underlying assumptions, and some of the previously noted issues are relied upon here as well. An obvious assumption in the hexagonal pattern is that consumers are minimizing distance with respect to transportation costs, and perceive the offered goods/services as equivalent. That is, a consumer will go to the closest center (or town), so there behavioral preference is ex ercised. On the business side, it is assumed that they seek the most advantageous locations. There are assumed economies of scale in operations related to system hier archies as well. As with previous work in location theory, researchers have taken issue with these and other relied upon assumptions. Thus, work has continued to extend the theory in a number of ways.

The significance of the work on central places relative to location theory is the prevailing factors of market areas, namely range and threshold. These characteristics, along with consumer and business behavior, ultimately result in emerging spatial patterns. Over time a spatial equilibrium emerges, and under certain conditions such patterns will be in the form of hexagons.

Spatial Competition

A final area of location theory is a focus on spatial competition, for which the contributions of Hotelling in 1929 are generally noted. Much of the previous work on land use and industrial production assumes some aspect of locational independence. However, spatial competition recognizes that an advantage can be gained by where one locates (or does not locate), creating a state of dis equilibrium. After all, businesses are in competition for customers. Of note in the work of Hotelling was an ex ample along the lines of that shown in **Figure 6**. Here, two hot dog vendors are operating along the shown street. The question is where is the best place to locate given a competitor on this street (knowing that the other vendor can also locate anywhere on the street). However, the issue of pricing is also a factor. Assume two given locations, market share can be manipulated by reducing the cost of the product, hot dogs in this case. By lowering the cost, more market share is possible. However, for the business to be viable for both vendors, a price equi librium will be achieved and it will be above the cost to run the operation. With respect to location, it was shown that the best place to locate was in the middle. In doing so, under idealized conditions it would ensure the vendor, say cart 1, of at least half of the total market. Half of the market is secured if both vendors are located halfway

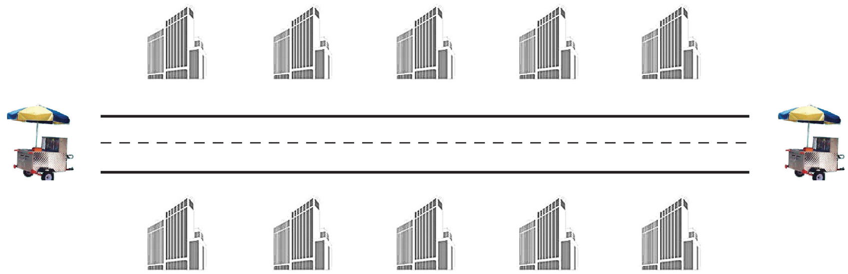

Figure 6 Two-vendor location along a linear market.

down the street next to each other (back to back in a linear market with uniformly distributed demand where people patronize the closest vendor). If the other vendor (cart 2) was located away from the middle, then the other vendor (cart 1) could gain additional demand purely on the basis of locational advantage. That is, the vendor in the middle (cart 1) would maintain half of the total de mand on the street plus half of the demand between the other vendor. Of course both vendors would eventually recognize this. Thus, at equilibrium both vendors would be next to each other, or co located, as this would ensure each an equal market share.

The work on spatial competition is an obvious attempt to deal with the complexities of businesses. A central issue in spatial competition is that the optimal location is not solely a function of cost minimization, but also takes into account demand (or market area) factors. Further, businesses must contend with game theoretic behavior, and such behavior (and associated costs) has implications for product pricing and market share. As with previous work, there are assumptions being relied upon. The simplified example highlights this point, but important insights are gained, forming at least part of the broader underpinnings of location theory.

Extensions

As noted previously, the above work has relied on as sumptions and simplifications to derive a theoretical basis for understanding land use, industrial production, central places, and spatial competition. Many researchers have devoted considerable effort to extending efforts in these areas, relaxing assumptions and putting into practice these theoretical underpinnings. Well recognized names contributing to location theory include Melvin Green hut, Walter Isard, Brian Berry, Martin Beckmann, but no doubt many others could be mentioned as well.

Prediction

Much of the noted location theory thus far relates to fairly explicit connections to economic theory. No overview of location theory could be devoid of recognizing other related advances associated with the formalization of basic ideas and/or mathematical ab stractions, both of which have contributed to the ability to predict patterns and/or locate future services.

The work by Alan Wilson formalizing spatial inter action approaches has been key for identifying market areas. Leon Cooper is noteworthy for his work on mod eling multiple facility siting, extending the basic problem posed by Weber. This is known as the location allocation problem. Louis Hakimi raised the issue of a more structured spatial environment, recognizing the limits imposed by actual transportation networks. These mod eling variants are known as the p center and p median problems. Further, Hakimi established optimality con ditions associated with location on a network. Charles ReVelle is recognized for his work on coverage modeling, where concepts of range are explicitly addressed as a mathematical abstraction of location. The two most well known of these modeling approaches are the location set covering problem and the maximal covering location problem.

Issues

Location theory contributes significantly to understand ing important current issues and helping develop plans to serve social needs and/or promote economic growth. It underlies effective emergency service plans, like locating hospitals, fire stations, warning sirens, etc. Location the ory also enables competitive services to be evaluated and explored, prior to siting the business or outlet. As an example, how will a market for a particular good respond if a new business is located in a particular place, and what will be the response of competitors. A final current context for which location theory is invaluable is exam ining environmental sustainability issues, like minimizing local impacts and determining how many activities (or species) are viable in an area.

More contemporary issues and contributions to lo cation theory involve addressing context, complexity, and representation of geographic space. The context issue is varied, including concerns regarding the phenomenon of interest varying across space, such as public equity, private industry objective, utility specification, con straining conditions, etc. By complexity we include issues of parameter fitting and interpretation in the context of spatial interaction, as well as model solvability. Finally, with respect to representation there are issues of how to represent space, either as objects (and which ones are important), whether data aggregation should be ap plied, how to represent spatial relationships like con nectivity and proximity, and appropriateness of scale of analysis.

Perhaps the greatest implications for location theory and its further evolution are GIS, and geographic infor mation science (GIScience) more generally. Effectively, more detailed information is available and accessible due to GIS maturation as a product and an industry. This has invariably raised questions about underlying assumptions related to location theory, but also has led to questions regarding what can be done and how this should be approached from a mathematical abstraction point of view.

See also: Spatial Analysis, Critical; Spatial Interaction Models.

Further Reading

Beckmann, M. (1968). *Location Theory*. New York: Random House.
Church, R. L. (1999). Location modelling and GIS. In Longley, P., Goodchild, M., Maguire, D. & Rhind, D. (eds.) *Geographical Information Systems* (2nd edn.), pp 293 303. New York: Wiley and Sons.
Dicken, P. and Lloyd, P. E. (1990). *Location in Space* (3rd edn.). New York: Harper and Row.
Isard, W. (1956). *Location and Space Economy*. Cambridge, MA: MIT Press.
Miller, H. J. (1996). GIS and geometric representation in facility location problems. *International Journal of Geographical Information Systems* 10, 791 816.
Murray, A. T. (2005). Geography in coverage modeling: Exploiting spatial structure to address complementary partial service of areas. *Annals of the Association of American Geographers* 95, 761 772.
Puu, T. (2003). *Mathematical Location and Land Use Theory* (2nd edn.). Berlin: Springer.
Smith, D. M. (1981). *Industrial Location* (2nd edn.). New York: Wiley.

Logistics

M. Hesse, University of Luxembourg, Luxembourg
J.-P. Rodrigue, Hofstra University, Hempstead, NY, USA

Glossary

Commodity Chain (Supply Chain) A functionally integrated network of production, trade, and service activities that covers all the stages in a supply chain, from the transformation of raw materials, through intermediate manufacturing stages, to the market. The chain is conceptualized as a series of nodes, linked by various types of transactions, such as sales and intrafirm transfers.

Distribution Centers Facility or a group of facilities that perform consolidation, warehousing, packaging, decomposition, and other functions linked with handling freight. Their main purpose is to provide value-added services to freight and they are a fundamental component of freight distribution.

Electronic Data Interchange (EDI) Communication mode for inter- and intrafirm data exchange in the freight forwarding and logistics business.

Gateways A location offering accessibility to a large system of circulation of freight, passengers, and/or information. Gateways reap advantage of a favorable physical location such as highway junctions, confluence of rivers, seaboards, and have been the object of a significant accumulation of transport infrastructures such as terminals and their links.

Hubs Central point for the collection, sorting, transshipment, and distribution of goods and passengers for a particular area. This concept comes from a term used in air transport for passengers, as well as freight. It describes collection and distribution through a single point ('hub and spoke' concept).

Intermodal Transportation The movement of goods in one and the same loading unit or road vehicle, which uses successively two or more modes of transport without handling the goods themselves in changing modes. Enables cargo to be consolidated into large units (e.g., containers) in order to have efficient cargo transfer between ships, barges, railcars, or truck chassis.

Platform/Modular Manufacturing Strategy in which a multinational corporation retains its core competencies, namely its research and development, retailing, marketing, and distribution, while subcontracting much of the manufacturing to the lowest bidders.

Radio Frequency Identification Device (RFID) Technology that uses small devices attached to objects that transmit data to a receiver. An alternative to bar coding used for identification and tracking purposes, notably for items shipped in units (boxes, containers, etc.), but can also be attached to an individual item. Main technical advantages include data storage capacity, read/write capability, and no line-of-sight requirements during scanning.

Supply Chain Management (SCM) The management of the whole commodity/supply chain, from suppliers, manufacturers, retailers, to the final customers. To achieve higher productivity and better returns, SCM tries to reduce inventory, increase transaction speeds, and satisfy the needs of the customers in terms of cost, quantity, quality, and delivery as much as possible.

Logistics involves a wide set of activities dedicated to the transformation and circulation of goods, from raw materials, manufacturing, to final market distribution, as well as the related information flows. Such activities are divided into two major functions which are materials management, with an emphasis on organization, and physical distribution, including transportation. By definition, materials management considers all the activities related in the manufacturing of commodities in all their stages of production along a supply or commodity chain. It includes production and marketing activities such as production planning, demand forecasting, purchasing, and inventory management. Physical distribution includes the broad range of activities involved in the movement of goods from points of production to final points of sale and consumption. Logistics has to provide for seamless flows between the different components of an increasingly fragmented creation of added value.

In the context of globalization, logistics includes the set of operations required for goods to be made available on markets or to specific destinations, which mainly include transportation, stock management, and order processing (**Figure 1**). Logistics is a multidimensional activity where added value is provided through the use of production, location, time, and control of elements of the supply chain. It reconciles production and consumption and its efficient use insures a close match between the needs of the customers and the capacity of the suppliers to provide them. Other terms have been used in close association with logistics, including supply chains and commodity chains, aiming at analyzing the set of interrelated activities needed for the manufacturing and marketing of certain products. Regarding logistical ex penditures, which include transportation, warehousing, inventory carrying, order processing, and administration, they represent about 10–15% of the total world gross domestic product (GDP). The most significant cost concerns transportation (39%), followed by warehousing (27%), and then inventory carrying (24%); they jointly account for 90% of all logistics costs. It is quite clear that considering those costs there are substantial interests for transport companies and manufacturers to improve their freight distribution.

On the basis of new information and communication technologies, distribution systems have become increas ingly driven by demand instead of by supply, implying a shift in the relative importance of specific logistical functions, mainly inventory, transport, and information systems (**Figure 2**). In a conventional situation, a supply driven distribution system is mainly based on the func tion of inventory. Production, often taking place in large batches, is simply 'pushed' down the supply chain with the hope that what is being produced will be consumed. Since elements of the supply chain are loosely integrated, parts and/or products must be stored to accommodate the chronology of the demand. In contrast, contemporary distribution systems are becoming increasingly demand driven (upstream). Under such circumstances, minimal inventories are maintained and most of it is circulating, thus the increasing importance of the transport com ponent in distribution. The operational management of such a system relies heavily on information systems to

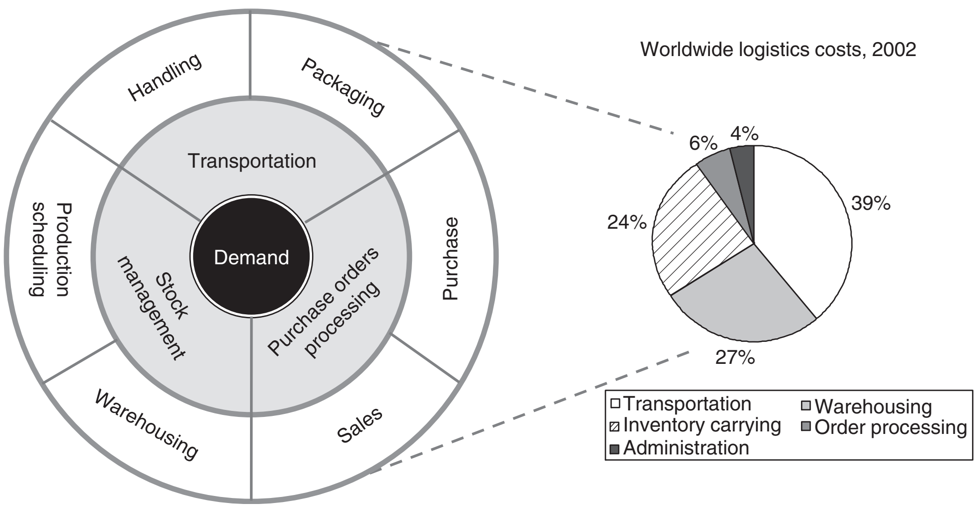

Figure 1 Logistics operations and global logistics costs.

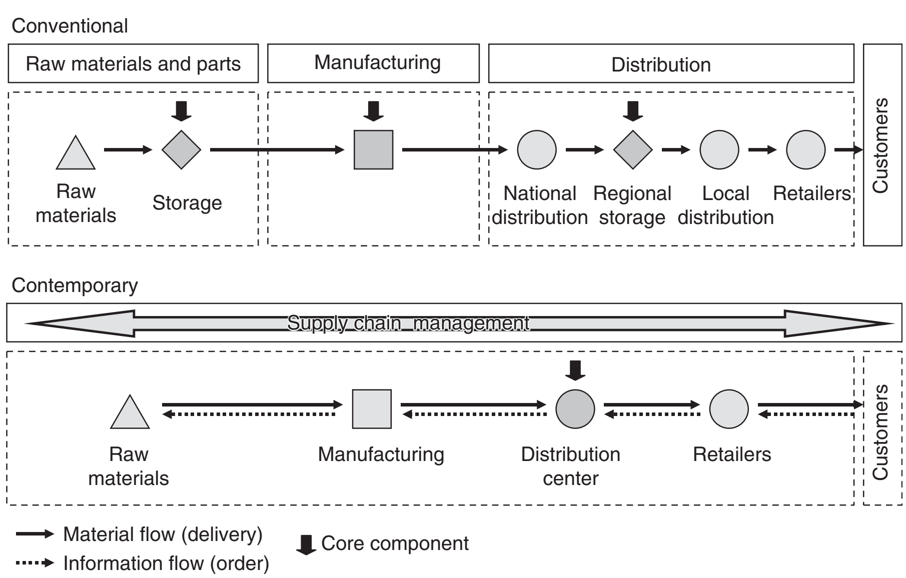

Figure 2 Traditional and modern supply chain arrangements.

insure that parts and/or products are delivered where and when they are required (on demand).

The Dimensions of Logistics

Organizational Perspective

Integrated supply chain management has developed in response to new modes of production, in the context of globalization, and with respect to a highly competitive market environment. The focus of supply chain management is shifting from maintaining inventories aimed at approximately satisfying a demand toward a comprehensive data collection system insuring that supply matches more closely with demand. This is mainly to be achieved through on demand distribution. Thus, physical flows also involve a significant amount of information flows. This trend was accelerated by the use of logistics, namely a better integration between transport modes and inventory control.

One of the main rationales for eliminating inventories is the need for cost reduction at all levels of corporate activity. Also essential is achieving a high market presence in terms of timely delivery, particularly since inventories no longer buffer against supply problems, disruptions in the transport chain, etc. In order to fulfill these requirements, specialized coordinators and integrators (third and fourth party logistics providers) have emerged, focusing on improving parts of the supply chain or providing their coordination and control. While push logistics involves a limited level of integration between suppliers, manufacturers, and distributors, a pull logistics system tries to achieve a higher level of efficiency through integration. Freight flows between components of the supply chain tend to be more frequent and in smaller batches. In addition, the sharing of demand dependant data helps better synchronize supply with demand. Reverse logistics also tends to be better integrated in the system to achieve a higher level of customer service, as well as to promote environmental strategies such as recycling.

The conventional forward channel in freight distribution is well understood with raw materials, parts and finished goods flowing from suppliers to producers, distributors and, finally, to consumers. There is also a reverse channel where waste, packages, and defective/obsolete products are 'climbing back' the supply chain (**Figure 3**). In some cases (such as a defective product), distributors will take back the merchandises, but in many others, a specialized segment of the distribution industry aims at collecting and then recycling goods and parts. Thus, reverse logistics (or reverse distribution) is concerned about the movements of previously shipped goods from customers back to manufacturers or distribution centers due to repairs, recycling, or returns.

Major changes in the manufacturing systems have led to the emergence of 'platform corporations' that bundle a core of research and development, finance, marketing, and distribution activities while removing (outsourcing) the manufacturing component, or having never had manufacturing in the first place. They have done so by focusing on the activities that provide the most added value and subcontracted the manufacturing of the products they design. Their core activities include research and development, finance, marketing, retail, and distribution. Many of them own globally recognized brand names and are actively involved in the development of new products. Their net worth is thus more a function of their brand names and capacity at innovation than from some tangible production assets. They outsource as much of the low margin work as possible and are flexible in their choice of suppliers. Thus, the term platform is used to characterize a mobile core establishing temporary relationships with manufacturers.

Platform corporations particularly thrive in the context of free trade; the mobility of factors of production is facilitated so that products can be produced wherever costs are the lowest and exported back to major markets.

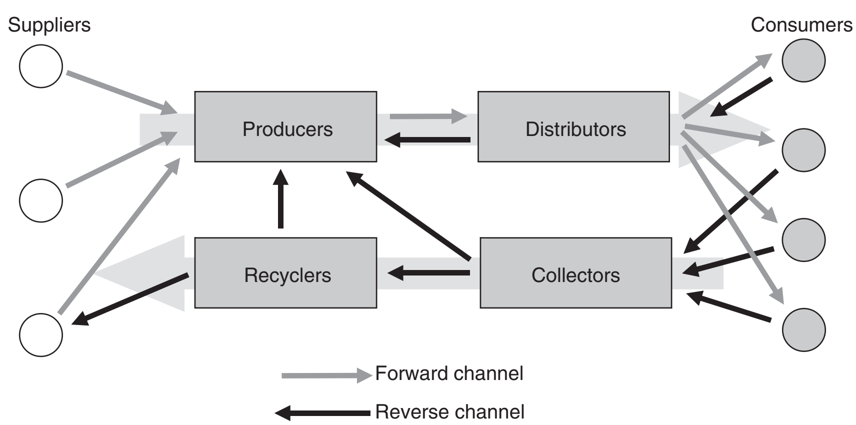

Figure 3 Forward and reverse distribution.

Telecommunications have also allowed corporations to decentralize their process while maintaining a level of control over their supply chain and informing their suppliers about changes in the demand. Since platform corporations provide the specification of the products/parts they require, there are a potentially high number of manufacturers that can bid to become suppliers. This works to the advantage of platform corporations by keeping costs low and it even creates a situation of recurrent overcapacity. Dell, IKEA, and Wal Mart are examples of such corporations that outsource the low margin manufacturing activities to low cost locations and establish effective networks of supply and production. Corporations are thus becoming 'logistical entities' since they manage a complex structure of production, distri bution, and retailing.

Technology

Modern distribution systems require a high level of con trol of their flows. Although this control is at start an organizational and managerial issue, its application re quires a set of technical tools and expertise. If technology can be defined by the level of control over matter, tech nology applied to logistics can be defined as the level of control of its flows, let them be physical and information related. An important technological change relates to intermodal transportation, particularly containerization, which has been shaping the logistics system in a funda mental way. Logistics and integrated transport systems are reciprocal endeavors. More recently, the application of new information and communication technologies (ICT) for improving the overall management of flows, particu larly their load units, has received attention. Thus, the physical, as well as the ICT parts of technological change are being underlined. The ICT component is particularly relevant as it helps strengthen the level of control dis tributors have over the supply chain. The technological dimension of logistics can thus be considered from five perspectives:

1. *Transportation modes*. Have been the object of very limited technological changes in recent decades. In some cases, modes have adapted to handle contain erized operations such as road and rail (e.g., double stacking). It is maritime shipping that has experienced the most significant technological change, which re quired the construction of an entirely new class of ships and the application of economies of scale to maritime container shipping. In the context of these changes, a global network of maritime shipping ser vicing large gateways has emerged.
2. *Transportation terminals*. The technological changes have become very significant with the construction of new terminal facilities operating on a high turnover basis. Better handling equipment lead to improvements in the velocity of freight at the terminals, which are among the most significant technological changes brought by logistics in materials movements.
3. *Distribution centers* (*DCs*). Technological changes im pacted over the location, design, and operation of distribution centers; the facilities handling the re quirements of modern distribution. From a locational standpoint, DCs mainly rely on trucking, implying a preference for suburban locations with road accessi bility. They have become one storey facilities de signed for higher throughput and less warehousing with specialized loading and unloading bays and sorting equipment.
4. *Load units*. Since logistics involves improving the ef ficiency of flows, load units have become particularly important. They are the basic physical management unit in freight distribution and take the form of pallets, swap bodies, semitrailers, and containers. Containers are the privileged load unit for long distance trade, but the growing complexity of logistics required a more specific level of load management. The use of bar codes and increasingly of Radio Frequency Iden tification Device (RFID) enables a high level of con trol of the load units in circulation.
5. *E commerce*. Consider the vast array of information processing changes brought by logistics. The com modity chain is linked with physical flows, as well as with information flows, notably through Electronic Data Interchange (EDI). Producers, distributors, and consumers are embedded in a web of reciprocal transactions. These transactions mostly take place virtually and their outcomes are physical flows. E commerce offers numerous advantages for the whole commodity chain, from consumers being exposed to better product information to manufacturers and distributors being able to adapt quickly to changes in the demand.

For logistics, ICT is particularly a time and embedded ness issue. Because of ICT, freight distribution is within a paradigm shift between inventory based logistics to re plenishment based logistics (**Figure 4**). The reliance is shifting from maintaining inventories aimed at approxi mately satisfying the demand to a comprehensive data collection system insuring, mainly through on demand transport, that supply matches with demand. This trend is accelerated by logistics, namely a better integration between the movement of goods on one hand and in ventory control on the other hand. While a push logistics system involves a limited level of integration between suppliers, manufacturers, and distributors, a pull logistics system tries to achieve a higher level of efficiency through integration of flows (i.e., information). Freight flows between components of the supply chain tend to be more frequent and in smaller batches. In addition, the sharing of demand dependant data (such as sales) helps better synchronize supply with demand. Reverse logistics also tends to be better integrated in the system to achieve a higher level of customer service, as well as to promote environmental strategies such as recycling. In several cases, third party logistics (3PL) providers are con tracting the supply chain management of some segments of the commodity chains.

Institutions and Regulation

In conventional logistics, materials management and physical distribution have been operated by manufacturers or retailers. The transport industries were constrained to a certain portion of the market, due to strong regulations pertaining to market access, multimodal ownership, and freight rates. As a consequence of de regulation, the market share of specialized service providers has increased tremendously through modal and intermodal penetration. According to recent organizational change in supply chain management, specialized subsidiaries like third and fourth party logistics (3PL and 4PL) providers have been taking over large portions of materials management from shippers in industry and retail (**Figure 5**). A 3PL is an asset based company that offers logistics and supply chain management services to its customers. It commonly owns and manages distribution centers and transport modes. A 4PL integrates the resources of producers, retailers, and 3PL providers in view to build a system wide improve ment in supply chain management. They are non asset based meaning that they mainly provide organizational expertise.

Major changes have occurred in the global trans portation regulatory framework, particularly in North America and in Europe through de regulation and pri vatization. Whereas transport services have traditionally been considered part of the national transport infra structure and thus been subject to strong regulation, most transport and logistics submarkets have been opened to competition since the 1980s. Formerly public or state owned companies like the Deutsche Post World Net now belong to the largest logistics corporations worldwide. As a consequence of de regulation, freight rates have de creased, productivity has improved, and markets have become integrated at a transnational level (e.g., North American Free Trade Agreement (NAFTA), European Community). In most cases, de regulation also permitted cabotage, foreign market entry, or multimodal ownership and thus improved the degree of integration of supply chains. Also, freight transport increased in response to lower freight rates and transnational integration based on diminishing barriers (customs, concessions). More

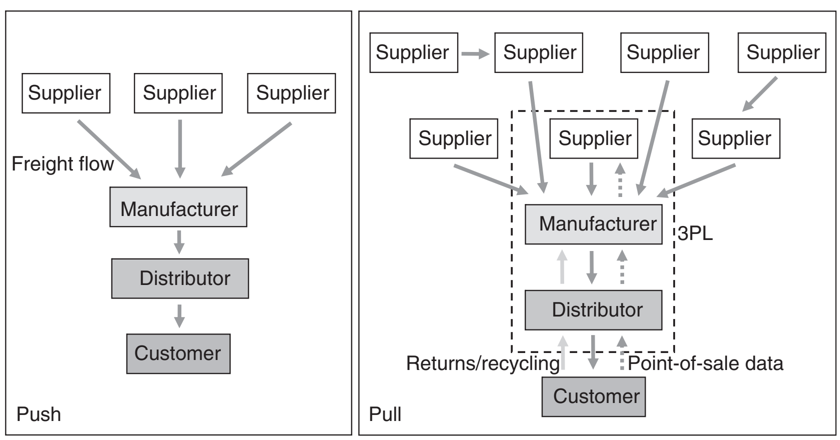

Figure 4 From push to pull logistics.

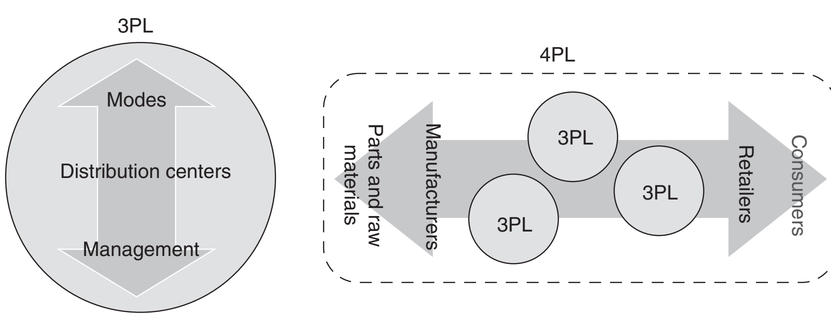

Figure 5 Third- and fourth-party logistics providers.

recently, the globalization of supply chains is further increasing, with multinational corporations offering worldwide services in logistics, terminal operations, etc. In particular, the specialization and globalization of port terminal operators is becoming highly important to understand spatial dynamics in the networks of global flows. Thus, both fragmentation and integration of supply chains can be observed. Also, as a general trend, the formerly strong link between corporate activity (e.g., terminal operations) and locations (e.g., main ports) is unraveling.

Regardless of the important role market de regulation has played in transport over the last decades, public policy remains a major factor of influence for the framework of distribution and thus on the freight traffic performance. This is due to the still vital role of nation states in providing and maintaining transport infrastructure, also to jurisdiction issues, particularly the enforcement of load and vehicle inspections, labor regulations, or vehicle noise and air emission standards. The main rationale used by public agencies to justify their involvement is increasingly related to the en vironmental sustainability of transport. Improved logis tics is commonly perceived as a strategy to achieve this goal. At the local and regional levels, zoning policies and building permits regulate the locational constraints and opportunities of the logistical firms. Yet the modern corporation appears to be quite powerful in selecting or even creating the locational environment that is needed. This usually exerts high pressure on local communities to offer appropriate conditions and to compete for corpo rate investments.

Territory and Infrastructures

The territorial development of the conventional logistics system was mainly determined by transport infra structure and the physical conditions under which it could be established and operated. In contrast, the spatial performance of modern logistics is an outcome of net work building and network design. Logistical networks are the major precondition for market coverage, spatial reach, and scale of distribution. They are increasingly hierarchical and configured at large scale, shifting from the local or regional level to national or even transna tional reach. The flow of goods depends on this particular characteristic, as does the locational system, according to the demand for cost reduction and economies of scale. Thus, the extent of goods flow and the number of dis tribution levels and places are inversely correlated: the more centralized the logistics network, the higher the amount of vehicle miles required to operate the system.

As a consequence of this trend toward hierarchy, distribution functions tend to be centralized, with dis tribution areas being extended. Thus, new logistics networks support the shift toward larger distribution centers, often serving significant transnational catch ments. However, this does not mean the demise of na tional or regional distribution centers, with some goods still requiring a three tier distribution system, with re gional, national, and international distribution centers. Besides the shift from multilevel structure toward a more large scale network structure, logistics networks have also been changing from direct to hub and spoke re lations, with central hubs being interconnected and serving subordinated nodes (**Figure 6**).

Concomitantly, facilities for logistics operations are much larger in size than before, the locations being se lected with particular respect to long distance accessi bility. Traditionally, freight distribution settled close to the places of production, such as in the manufacturing belt at the North American East Coast and in the Mid west, or in the old industrialized regions of England and continental Europe. Today, particularly large scale goods flows are directed through major gateways and hubs, mainly large ports and major airports, and at highway intersections with access to an extended market area. The changing geography of manufacturing and industrial production has been accompanied by a changing geog raphy of freight distribution. This is notably the case in Pacific Asia, where industrialization and integration to global trade have triggered the emergence of large scale high throughput nodes such as Hong Kong, Singapore, Shanghai, Busan, and Kaohsiung, now representing the largest container ports in the world. In the case of North America, those hubs or gateways are strategically located at the East and West Coasts, such as the Ports of the San Pedro Bay in Los Angeles, the Port of Seattle/Tacoma, and the Port of New York/New Jersey.

The expansion of such places is primarily due to the growth of trade and transport in general, supported by economic growth and the expansion of market areas due to globalization. Yet, the strategy of concentrating freight at hub locations is also becoming restricted, due to density, land constraints, and congested traffic arterials. Such limits to expansion and the scarce hinterland con nections of major hubs are considered the most important obstacle for further developing major hub locations. Consequently, inland hubs are becoming more and more important, where primarily road and air freight is con solidated. These new DC areas are mainly affiliated with the interstate network and air cargo facilities. Ware housing, trucking, freight forwarding, and air cargo activities are major indicators and drivers of this new distribution economy. One of such new inland hubs is emerging in North America along the Ohio River Valley, stretching across a corridor from Ohio and Indiana to Tennessee. Respective European develop ments favor strategic places in Southeast England, the Benelux countries, Germany, and eastern France.

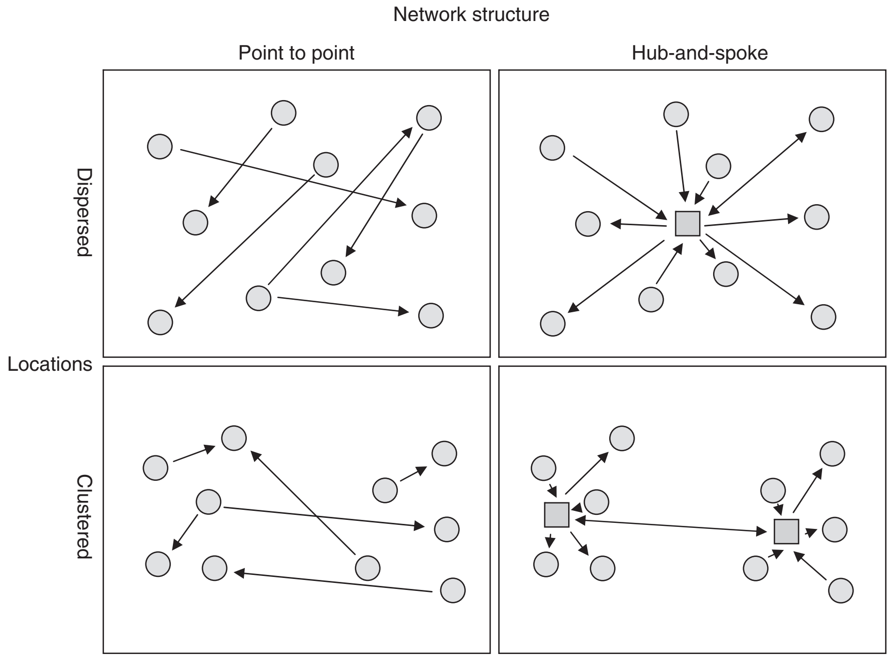

Figure 6 Freight distribution and network strategies.

The Netherlands has emerged as the most favored lo cation for European logistics, due to excellent accessi bility, advanced terminal and transport infrastructure, critical mass of logistics functions, and attractive operating conditions (vis à vis its neighbors). Schiphol Airport and the Port of Rotterdam are among the most important hubs for international freight flows in Europe.

Since many parts of the supply chain are now globally integrated, distribution centers tend to be the link be tween global sourcing and regional distribution. Thus the DC has become an interface between the geographies of manufacturing and retailing, consequently handling the distribution scale and scope. Innovations such as con tainerization and particularly developments in infor mation technologies have integrated all components of the chain. In response, major players in the distribution business (e.g., container shipping lines, freight forwarders, warehousing firms, terminal operators) are trying to control as many parts of the logistics chain as possible. Challenged by vertical and horizontal linkages, by mer gers, takeovers and strategic alliances, they need to stay competitive by increasing throughput and providing services at the lowest rates. As a result, the activity space of mainports is becoming relocated to low cost locations reaching far beyond traditional terminal sites and con necting more distant places of their hinterlands.

Outlook

The enduring growth in the movements of goods and in novations in the associated networks of logistics and freight distribution represent a dynamic part of economic glob alization. These issues have however been under repre sented in regional science and geographical studies. Freight transportation appears to have been dominantly the focus of geographers dealing with maritime, port, and rail transport issues. More recent developments in intermodal transportation, logistics networks, international trade, and the emergence of e commerce have substantially trans formed the freight transport sector. The same applies for changes in the institutional regimes and in corporate competition. These changes triggered a renewed interest from economists, social scientists, and geographers, placing the issue of freight distribution as a central concern of human geography, much beyond the traditional focus of transport geography. As technological change and global ization are moving forward, flows, circuits, and the associ ated spaces will receive further attention, since they are essential for the global network economy.

See also: Economies, Branch Plant; Firms; Hub Network Location; Transport and Globalization; Transport and Social Exclusion.

Further Reading

Brewer, A., Hensher, D. and Button, K. (2001). *Handbook of Logistics and Supply Chain Management*. London: Pergamon.
Dicken, P., Kelly, P., Olds, K. and Yeung, H. (2001). Chains and networks, territories and scale: Towards a relational framework of analyzing the global economy. *Global Networks* 1(2), 89 112.
Hesse, M. and Rodrigue, J. P. (2004). The transport geography of logistics and freight distribution. *Journal of Transport Geography* 12(3), 171 184.

Leinbach, T. and Capineri, C. (2006). Freight transport, seamlessness, and competitive advantage in the global economy. *European Journal of Transport and Infrastructure Research* 6(1), 23 38.

Notteboom, T. and Rodrigue, J. P. (2005). Port regionalization: Towards a new phase in port development. *Maritime Policy and Management* 32(3), 297 313.

Slack, B. (1998). Intermodal transportation. In Hoyle, B. & Knowles, R. (eds.) *Modern Transport Geography* (2nd edn.), pp 263 289 London: Wiley.

Slack, B. and Frémont, A. (2005). Transformation of port terminal operations: From the local to the global. *Transport Reviews* 25(1), 117 130.

Sturgeon, T. (2002). Modular production networks: A new American model of industrial organization. *Industrial and Corporate Change* 11(3), 451 496.

Visser, E. J. and Lambooy, J. G. (2004). A dynamic transaction cost perspective on fourth party logistic service development. *Geographische Zeitschrift* 92(1 2), 5 20.

Relevant Websites

http://www.cscmp.org
Council of Supply Chain Management Professionals.

http://www.logisticsmgmt.com
Logistics Management.

http://people.hofstra.edu/geotrans/
The Geography of Transport Systems, Hofstra University.

Longitudinal Methods (Cohort Analysis, Life Tables)

S. D. Withers, University of Washington, Seattle, WA, USA

Glossary

Cohort Any set of individuals sharing a commonality as a group in some way, usually by age.

Cohort Analysis A method for investigating the changes in patterns of behavior and attitudes of a cohort.

Event-History Analysis Various statistical methods for examining changes between successive states within an interval of time on the basis of a complete temporal record for a sample.

Hazard Rate The number of subjects experiencing a transition (change) divided by the number at risk of experiencing the transition (change).

Life Course Research The study of social processes extending over the individual life span.

Life Tables A table generated from age-specific mortality rates used to measure mortality, survivorship, and life expectancy.

Introduction

Human geographers are interested in understanding change over time and across space. As a discipline we have a rich tradition of documenting spatial and temporal variation, but more recent geographic inquiry has been concerned with the spatial and temporal processes and their interrelation in producing spatial and temporal variation. The development and application of longitudinal research methods in the social sciences, and in geography, have significantly advanced our understanding of temporal and spatial processes. Longitudinal research methods have been applied to a variety of substantive fields in geography. They are commonly applied in studies of the dynamics of residential duration and mobility, migration, immigrant settlement, life course research, intergenerational spatial arrangements, housing and home ownership, settlement patterns, neighborhood changes, residential segregation, regional economic health, labor market transitions, and consumer retail choices, to name a few. They have also been used in innovative ways to understand spatial and social diffusion. To date, longitudinal methods have led to important theoretical advances in our understanding of many socio temporal processes, and significant progress is being made in our understanding of dynamic spatial processes.

Although longitudinal methods have made a significant contribution to geographic inquiry, they were developed initially outside the discipline. This article discusses the development of longitudinal methods and the adoption of these methods within the field of geography. There is no single method longitudinal analysis. Longitudinal models frequently are termed hazard models, duration models, transition models or event history analysis. Two early precursors to event history analysis are survival studies using life tables and cohort analysis in life course research. Longitudinal methods are applicable for the study of temporal as well as spatial processes, although the former are developed more extensively. This article explains important terminology used in longitudinal modeling and numerous theoretical time dependencies. Longitudinal methods lend themselves to life course research due to the ability to empirically examine multiple time dependencies. These are frequently referred to as parallel events across various careers and trajectories. Next, a variety of longitudinal models are described representing parametric, semiparametric, and nonparametric approaches applied in continuous time and discrete time dimensions. Lastly, the extension of longitudinal models to spatial processes is reviewed.

The Origins and Foundations of Longitudinal Methods

The early developments of longitudinal methods were frequently called survival analysis since the substantive question was survival, literally. In the field of engineering, survival analysis (often called failure time models) was used to determine when a product would fail, or conversely how long it might last. In fields such as medicine and epidemiology, survival analysis was used to determine the potentially improved survival probabilities from specific treatments or medicines, compared to those not receiving the treatments. These earlier models assessed the probability of survival, also known as life expectancy.

Survival and the Life Table

The cornerstone of the demographic tradition of studying life expectancy is the life table. The life table describes the long run effects of a given set of age specific death rates on a cohort of people from birth until the death of the last survivor. A life table (**Table 1**) divides continuous time into discrete groupings of age intervals. For each age interval (column 1) the number in the

Table 1 L fe tab e for US ma es (1995)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Age interval	*Number of males in the population*	*Number of deaths in the population*	*Age-specific death rates in the interval*	*Probabilities of death (proportion of persons alive at beginning who die during interval)*	*Of 100 000 hypothetical people born alive*		*Number of years lived*		*Expectation of life*
					Number alive at beginning of interval	*Number dying during age interval*	*In the age interval*	*In this and all subsequent age intervals*	*Average number of years of life remaining at beginning of age interval*
x to $x+n$	$_nP_x$	$_nD_x$	$_nM_x$	$_nq_x$	l_x	$_nd_x$	$_nL_x$	T_x	e_x
0–1	1 969 872	16 622	0.00844	0.00831	100 000	831	99 283	7 254 216	72.5
1–5	8 055 333	3609	0.00045	0.00177	99 169	176	396 268	7 154 933	72.1
5–10	9 843 300	2212	0.00022	0.00113	98 993	112	494 659	6 758 665	68.3
10–15	9 685 241	3007	0.00031	0.00153	98 881	151	494 130	6 264 006	63.3
15–20	9 265 025	11 068	0.00119	0.00594	98 730	586	492 339	5 769 876	58.4
20–25	9 087 045	14 709	0.00162	0.00806	98 144	791	488 785	5 277 537	53.8
25–30	9 529 765	16 572	0.00174	0.00865	94 353	842	484 637	4 788 752	49.2
30–35	10 902 150	25 254	0.00232	0.01154	96 511	1114	479 803	4 304 115	44.6
35–40	11 071 207	32 339	0.00292	0.01462	95 397	1395	473 674	3 824 312	40.1
40–45	9 990 476	37 792	0.00378	0.01882	94 002	1769	465 893	3 350 638	35.6
45–50	8 559 836	42 609	0.00498	0.02458	92 233	2267	455 919	2 884 745	31.3
50–55	6 621 815	48 313	0.00730	0.03580	89 966	3221	442 290	2 428 826	27.0
55–60	5 317 251	58 848	0.01107	0.05398	86 745	4682	422 678	1 986 536	22.9
60–65	4 726 807	834 42	0.01766	0.08489	82 063	6966	393 799	1 563 858	19.1
65–70	4 505 822	119 415	0.02651	0.12480	75 097	9372	352 934	1 170 059	15.6
70–75	3 836 272	154 586	0.04031	0.18399	65 725	12 093	299 083	817 125	12.4
75–80	2 720 385	164 372	0.06044	0.26347	53 632	14 130	232 960	518 042	9.7
80–85	1 609 321	155 036	0.09636	0.38762	39 502	15 312	158 348	285 082	7.2
85+	1 016 875	182 823	0.17984	1.00000	24 190	24 190	126 734	126 734	5.2

Source National Center for Health Statistics unpublished tables Abridged Life Tables United States 1995 (http //www cdc gov/nchswww/datawh/statab/unpubd/mortabs htm accessed 1998) population is recorded (column 2), and the number of deaths is recorded (column 3) from which the age spe cific death rate in the interval can be calculated (column 4). The age specific probabilities of death are then cal culated as the proportion of people alive at the beginning of the interval that during the interval die (column 5). From these calculations, the probability of surviving to a specific age is calculated for a hypothetical population of 100 000 (called the Radix) (column 6). By multiplying the probability of death at each age interval with the number of people alive at the beginning of the interval, the number dying within this hypothetical population is calculated (column 7). The number of person years lived is then calculated for each of the age intervals (column 8) and then the cumulative number of years lived in each and all subsequent age intervals is calculated (column 9). The final calculation is the average number of years of life remaining at the beginning of the age interval, which is also known as life expectancy.

The goal of a life table is to measure the long run implications of the levels and age patterns of mortality at a moment in time. The table focuses on the age de pendence of rates and the expectations of the duration of life. In order to study variations within the population, separate tables are constructed for different groups. For example, males would be analyzed separately from fe males. Foreign born would be analyzed separately from native born, etc. Multistate and increment–decrement life table models are extensions of the life table method that handle multiple kinds of events. However, the gen eral approach becomes untenable if one wants to assess more than a few sources of population variation (het erogeneity). It is, however, invaluable in determining the survival rate over time.

Cohort Analysis and Life Course Research

Cohort analysis is another precursor to contemporary longitudinal analysis. It is a method of analyzing changes in behavior over time. A cohort is any set of individuals sharing a commonality as a group, usually grouped by age (birth cohort) but other important events can lead to cohort groupings, such as timing of marriage, graduation, and labor market entry. For a population at any given moment there are three explanatory temporal dimen sions: age, period, and cohort effects. Age effects refer to influences derived from the aging process; period effects are influences derived from the period of time such as a particular calendar year or specific moment in history; and cohort effects are influences derived from group membership. Being a particular age during a specific period constitutes the cohort effect. Together, age, per iod, and cohort effects fully specify the temporal di mensions of a population at any given moment.

Cohort analysis is longitudinal in the sense that a comparison is made of a cohort at two or more distinct periods of time to measure change. However, there are typically two limitations to cohort analysis. First, a difficulty with cohort analysis is that operationally age, period, and cohort effects are dependent measures since age can be derived from knowing the period and the cohort. The independent effect of any one of these elements is intractable. Second, the cohorts that are compared in a cohort analysis are usually separate sam ples. Typically, they are not the same individuals at two moments in time, but rather a representative sample of the cohort group at two time points. This is both an advantage and a disadvantage of cohort analysis. The advantage is that the data requirements of cohort analysis are relatively easily met with repeated cross sectional surveys. The disadvantage is that cohort changes can be readily influenced by unobserved variation between the two cohort samples. One way to overcome this limitation is to apply cohort analysis to panel data, which collects information from the same group of individuals over time.

The most prolific applications of longitudinal mod eling in geography fall within the broad scope of life course research, which is of little surprise given the demographic foundations of the development of these models. Life course research differs from life cycle studies of earlier decades. Life cycle studies tended to construct preconceived age specific behaviors across time, often with an emphasis on the normative se quencing of events across life spans. For example, a typical demographic life cycle model would depict a normative age specific progression from education, to leaving the parental home, to marriage, the child rearing, to empty nesting with corresponding achievements of homeownership, employment stability and promotion, and so forth. In contrast, life course research is concerned with variation across the life span in various parallel careers.

The life course is not the same as aging. It is a product of individuals' actions, institutions, and social and his torical forces. The value of a life course perspective is the appreciation that events in one aspect of life cannot be explained without reference to other aspects. Event his tory analysis has advanced life course research by em pirically examining the interdependencies of time and mutually constituted events in parallel domains. It pro vides a common conceptual framework to study multiple time dependencies in social life. The life course is a longitudinal paradigm that also applies to many social and spatial processes within geographic inquiry that are not of a demographic nature. The longitudinal approach lends itself to the study of change for any unit of analysis that exists over time. Applications within geographic research have studied subjects as diverse as individuals, households, firms, counties, business areas, market areas, business cycles, nations, elected representatives, settle ments, immigrants, and so forth.

Hazard Models for the Social Sciences

Geographers and other social scientists have a long tra dition of being interested in change. The purpose of longitudinal methods is to understand change. Another aspect of change that is vital, besides temporal variation, is change between states, which are often called events or transitions. A traditional life table portrays an un ambiguous single event (death). Yet many events of interest to geographers are much more complicated than a single nonrepeatable event. There are two important aspects of a longitudinal model: duration (time) and states (attributes). While survival analysis is suitable for determining the hazard over time for a single non repeatable event, it does not lend itself to the more complex substantive questions in social science research.

Throughout the 1970s and early 1980s, several ad vances were made in the statistical analysis of duration data making it more applicable to social science questions. Cox developed a regression model for life tables that in cludes a dependence of the hazard on time and a set of covariates. This model was a significant advancement in the field because it allowed the hazard to depend on time without having to determine in advance the structure of that dependence. In addition, statistical sociologists and economists developed standard hazard function methods that differentiated influences of unobserved heterogeneity from duration dependence. Initially, specialized software programs were written to estimate these newly developed hazard models. But, once these models were included in standard statistical software packages, the application of longitudinal methods in social science research exploded, and geography was no exception.

Considering Time and State Dependence

How does time influence the event in question? What is the dependence of the hazard on time? Longitudinal methods are particularly useful in determining how time (or duration) operates with respect to the question of interest. There are a number of intriguing questions that longitudinal methods enable researchers to explore. Time dependency can be measured in a variety of ways, including age, duration in a given state, duration in a previous state, exposure to a certain influence or insti tution, historical epochs, and the relative timing of historically meaningful moments such as 9/11, the as sassination of a public figure, or natural disasters and devastation.

In event history analysis, theoretical conceptions of transition probabilities focus on various types of state dependence. Markovian state dependence implies that the likelihood of a change in state is dependent on the current state occupied. For example, a renter is more likely to move than an owner. If the likelihood of a change in state is dependent on a previously occupied state, it is called a higher order Markovian state dependence. Occurrence dependence implies that the likelihood of a change in state is dependent on the number of times different states have been occupied. Continuing the example, a frequent mover who rents is more likely to move than an infrequent mover who rents.

In addition, time and state dependencies work in combination. For example, models of duration state de pendence imply the likelihood of a change in state is dependent on the length of time the current state has been occupied, commonly known as the principle of cumulative inertia. For example, a long time resident is less likely to move than someone who has recently moved. Similarly, a lagged duration state dependence implies a change in state is dependent on the length of time previous states have been occupied. In addition, the occurrence of one kind of event can increase or decrease the hazard of an other kind of event. Longitudinal models are sufficiently versatile to enable creative and complex theoretical spe cifications of the influence of time on events of interest.

The Language of Longitudinal Methods

Although there are many methods of longitudinal an alysis, there is a standard terminology used in these ap proaches. Commonly used are the terms event, duration, risk set, risk period, censoring, multistate transitions, competing events, episode, spell, trajectories, careers, hazard rate, and survival rate.

The event of interest must be well defined. By def inition, an event is a distinct, unambiguous change from one state to another. Longitudinal methods have broad applications because most phenomena can be categorized meaningfully into distinct events. So an event also as sumes a preceding time interval which represents its nonoccurrence. The interval of time during which the event does not occur is called duration. The set of in dividuals who are 'at risk' of the event occurring at each moment in time are called the risk set. The risk period is the complete interval of time during which a subject is at risk of having the event. It is the amount of exposure to risk. For example, by definition, upon getting married one enters the risk set of getting divorced (**Figure 1**, scenarios a, e). The risk period begins from the date of marriage until the end of the observation period, unless something occurs that removes the subject from risk without having the event.

Such occurrences are called censoring, specifically right censoring. If a subject were to die while still married

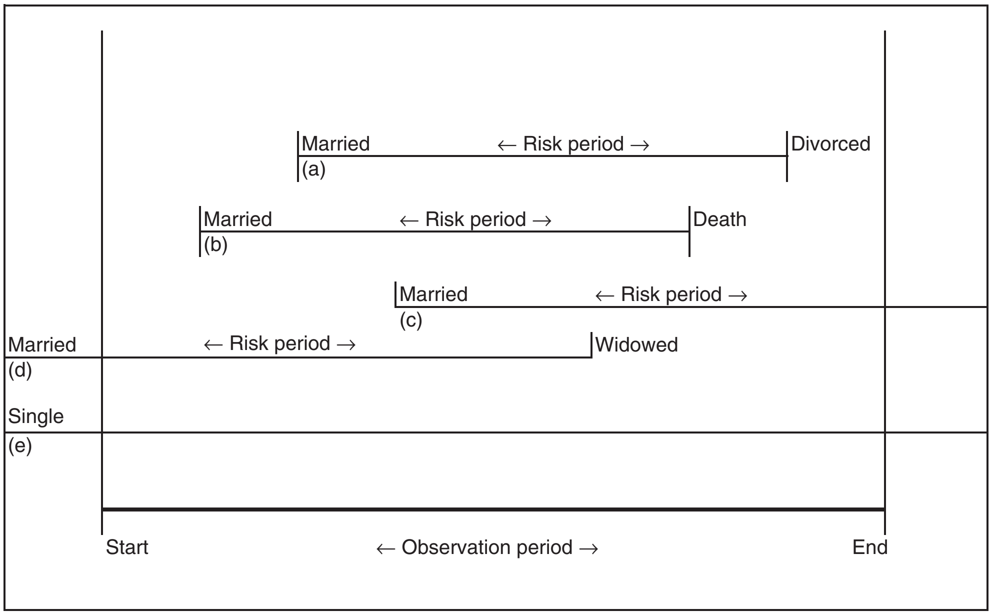

Figure 1 Hypothetical scenarios of the risk period for the risk of divorce across an observation period.

they would be censored since they are removed from the risk of the event of interest (**Figure 1**, scenario b). Standard regression methods cannot handle censored data, but longitudinal regression models are designed specifically to incorporate censored data. Anyone who remains at the end of the observation period without experiencing the event of interest is also right censored. The risk period is truncated at the end of the study (**Figure 1**, scenario c). Ideally, we observe people as they enter the risk set. However, it is not uncommon for people to enter the risk set before we observe them. In other words, they were already married when we started observing them (**Figure 1**, scenario d). These cases are censored (left censored) since there is information missing about the entire duration of their marriage.

For some risk periods their demarcation is not as obvious. For example, when does one become an adult? In a study of adult behaviors does one define an adult at age 18 (the legal context to vote and live independently), or age 21 (additional rights within the legal context), or upon moving out of the parental home (a behavioral context), or upon financial independence? The appro priate answer very much depends on the topic under study. Risk periods must also be given distinct un ambiguous measures and these measures need to reflect the theoretical context of the research and not arbitrary ages and dates.

Some event types are single state transitions but others are multistate transitions. For example, upon graduating from high school one could make the transi tion to work, to professional training, to college edu cation, or to travel, to name but a few options. Multiple event types must be defined unambiguously and are treated as competing events. Experiencing one of the competing events removes the subject from being at risk of the other event types. For example, if upon graduation you travel abroad for a year, you are no longer at risk of entering college directly after graduation. If you become widowed, you are no longer at risk of becoming divorced (**Figure 1**, scenario d).

Some events are repeatable. For example, most people change residence a number of times during their lifetimes. Each residence is referred to as an episode or a spell. The sequence of spells is called a career or a trajectory. People have numerous simultaneous trajectories (careers) throughout their lifetimes, such as educational trajectories (the sequence of schooling), housing trajectories (the se quence of homes and locations), household trajectories (the sequence of households one lives in over time), employment trajectories (the sequence of jobs held over time), and so on. Event history analysis determines if and how events and durations in one career influence events and durations in other careers across the life course.

The hazard rate $h(t)$ is an unobservable yet funda mental variable in longitudinal modeling. It is the rate at which events occur across time, given they did not al ready occur prior to time t. The usual expectation is that the hazard is not constant across time. The survivor function $s(t)$ is the probability of not having the event prior to time t. The unconditional instantaneous prob ability of having the event at time t is referred to as the probability density function. The hazard rate is the ratio of the probability density function to the survivor func tion. This is conceptually true of all longitudinal meth ods, but the specific functional form varies between the model types.

Types of Longitudinal Methods

The various types of longitudinal models can be dis tinguished by a few key elements. As discussed, there are differences between repeatable and nonrepeatable events, as well as single and multiple kinds of events. Other important distinctions are whether models are distri butional or regression methods; parametric, nonpara metric, or semiparametric methods; or whether they use continuous or discrete time measures. There is a wide variety of models called longitudinal methods, but here the focus is on five common methods: continuous time models, the Cox proportional hazards model, dis crete time models, log rate models, and spatial duration models.

Parametric Continuous-Time Models

Parametric models have distributional assumptions. In the case of continuous time longitudinal models, the parametric assumptions refer to the distribution of the underlying hazard rate. Recall the hazard rate is un observable yet controls the likelihood of events over time. These methods use a precise continuous measure of the duration of the risk period. Studies frequently use as small a unit of measure as is feasible. In life course re search it is common to have the calendar day, month and year of the start of the spell, and the day, month, and year of the end of the spell from which to calculate continu ous time duration in months. The standard hazard model that uses one parameter to characterize time dependence is called a Gompertz model, such that

$$\ln[h(t)] = a + \sum_i b_i X_i + ct \qquad [1]$$

where $h(t)$ is the hazard rate, a a constant, b the parameter estimates for a vector of explanatory variables X, and c the parameter for time dependence. Other time distri butions frequently tested are exponential, log logistic, uniform, normal distributions, and the Weibull model:

$$\ln[h(t)] = a + \sum_i b_i X_i + c[\ln(t)] \qquad [2]$$

In both of Gompertz and the Weibull model, the hazard changes continuously over time. It does increase and then decrease for example. The parameters are estimated by maximum likelihood estimation techniques. The data structure for these models is typically one record for each observation that includes the covariate measures and the duration measure.

These models are more popular in physical and bio medical sciences where frequently the main emphasis of inquiry is the functional form of the time dependence. These models are not as common in geography and the social sciences where theoretical time dependence is less of a concern. However, some geographers argue that these common parametric approaches should be explored more often in geographic inquiry to enhance our understanding of temporal processes.

Cox's Proportional Hazards Model

Influenced by the properties of the life table, in 1972 Cox developed a regression model that could handle the dual challenges of censored observations and time varying covariates. It is arguably the most popular methods of longitudinal analysis across the sciences. The model uses continuous time data. It is described as a semiparametric method. Cox's main innovation was a statistical method that partitions the likelihood function into two elements: one factor contains information about the coefficients, and the other factor contains information about time. Formally, the model can be written as

$$h_i(t) = h_0(t)\exp\left[\sum_k b_k X_{ik}(t)\right] \qquad [3]$$

where $h_0(t)$ is the baseline hazard function and $X_{ik}(t)$ is the vector of covariates which may change over time. These models are called proportional hazards because the model determines the underlying hazard for the baseline group (when values of the explanatory variables are null). This hazard remains the same for all groups. Hazards are increased or decreased on the basis of cov ariate values. This is only appropriate if the underlying time dependence is considered to be the same across covariate values. If this is not tenable, stratified models should be fit. For example, it is reasonable to believe that duration of residence influences renters differently than owners. One should stratify an analysis by tenure type (renters or owners) and run a separate Cox's model on each group.

Discrete-Time Event-History Analysis

Discrete time longitudinal methods do not have the precision regarding duration that continuous models exhibit. Instead, duration is measured discretely. Dis crete time measures commonly are derived from panel data which provide repeat measures on the same indi viduals at successive time periods. Discrete time models measure whether the event occurred within discrete time measures. For example, have you moved since last year? Did you have a change in family circumstances since last year? Because duration is measured in discrete units, it can be thought to approximate a continuous time dis tribution as a step function.

In a discrete time model, the hazard rate is the ratio of the number of events in the time period to the number of individuals at risk. The emphasis of these models is the specification of how the hazard rate depends on ex planatory variables. As for the time dependence, we can allow for any variation in the hazard by letting the intercept be different at each point in time, such that

$$\ln\left[\frac{h(t_i;X)}{1 - h(t_i;X)}\right] = a_i + \sum b_k X_k \qquad [4]$$

These models can be estimated with standard logistic regression procedures provided the data input structure is a person–period file, which lists a separate record of duration, time constant, and time varying covariates for each individual for each discrete time period they are in the risk set. There also needs to be a measure indicating whether the final record is an event or censored.

The time dependence is derived from the odds of the event occurring for the baseline group, which is the situation when all covariates equal zero. The parameters are estimated by maximum likelihood estimation. Par ameter estimates for covariates are then interpreted as increasing or decreasing the odds of the event relative to the baseline group. These models are called proportional hazard models because the dependence of the hazard on time is the same for all groups. Nonproportional hazard models can be specified by including an interaction be tween time and covariates. Including time varying cov ariates is particularly straightforward with these models.

Log-Rate Models

Hierarchical loglinear models are used in geographic research to analyze the association among categorical variables in complex contingency tables. A longitudinal variation of the loglinear model can be readily achieved by making two important changes to the saturated log linear model such that

$$\ln\left(\frac{F_{tij}^{TAB}}{W_{tij}^{TAB}}\right) = \lambda + \lambda_t^T + \lambda_i^A + \lambda_j^B + \lambda_{ti}^{TA} + \lambda_{tj}^{TB} + \lambda_{ij}^{AB} + \lambda_{tij}^{TAB} \qquad [5]$$

This is called a log rate model. As with a standard log linear model, F is the frequency of events, and T, A, and B are covariates. The presence of W in the denominator, which is the amount of exposure to risk, makes this a log rate model. In addition, this approach enables many theoretical hazard models to be empirically tested be cause one of the variables is time (T). Transformations of the time variable enable various parametric hazard models to be estimated. Proportional and nonpropor tional hazard models can also be estimates, since inter actions between covariates and time can be readily included in the hierarchical models. Log rate models are not as common as the previously described longitudinal methods but they provide a great deal of theoretical exploration without great demands for data, and show considerable promise in geographic research.

Spatial Duration Models

Spatial duration models follow the logic and mathematics of hazard models, but as applied to distance dependence rather than time dependence. Most of the conceptual foundations of temporal hazard models have spatial analogs. The spatial analog for the risk period (duration) is distance which is equal to zero at the start of the risk period. The end of the risk period occurs when a change in state occurs for a spatial process of interest some distance away from a substantively meaningful origin. Spatial duration (hazard) models can be constructed for continuous or discrete state space, and are particularly well suited to the study of diffusion processes, both spatial and social. However, very little research has examined spatial diffusion processes using an event his tory structure. Instead, these models have been applied mostly to spatial point patterns, and to spatial linkages, such as the spatial linkages between adult children and their elderly parents, settlement spacing across a region, and linkages in producer services, to name a few appli cations. Sparse is the research that has used spatial dur ation models to understand spatial processes, yet there is considerable potential for modeling diffusion processes. The parametric models of time dependence discussed previously lend themselves in a spatial context to esti mating different conceptions of how distance matters in diffusion processes. The dependence of the hazard across space can be conceptualized as constant, gradual, peaked, expansive, and so on. Creative applications could include studies of land use change along the urban/rural con tinuum or understanding spatial process of the spread of environmental hazards. Spatial duration models are a promising and underutilized approach for the discovery of spatial processes.

Conclusion

Longitudinal methods have been applied to a broad spectrum of geographic research, and have contributed to significant advancement in our theoretical understanding of spatial and temporal processes. Nonetheless, as a dis cipline we have barely begun to embrace the rich po tential of longitudinal models. We tend to study a single spell, rather than a trajectory. We tend to privilege one event type rather than mutually constituted events. We tend to focus on the explanatory role of covariates, and seldom interrogate the temporal process explicitly.

Promising, current research is embracing sequence analysis to study trajectories, simultaneous hazard equations to study the interdependence among events in parallel trajectories, and spatial hazard models to study spatial and social diffusion. Longitudinal methods will continue to advance our understanding of temporal and spatial processes in human geography.

See also: Demography; Regression, Linear and Nonlinear; Space-Time Modeling.

Further Reading

Allison, P. D. (1984). *Event History Analysis: Regression for Longitudinal Event Data*. Newbury Park: Sage Publication.

Blossfeld, H. P., Hammerle, A. and Mayer, K. U. (1989). *Event History Analysis: Statistical Theory and Application in the Social Sciences*. Princeton, NJ: Lawrence Erlbaum Associates.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society Series B* 34, 187 202.

Davies Withers, S. (1997). Methodological considerations in the analysis of residential mobility: A test of duration, state dependence, and associated events. *Geographical Analysis* 29, 352 372.

Glenn, N. D. (1977). *Cohort Analysis*. Beverly Hills: Sage Publication.

Holford, T. R. (1980). The analysis of rates and survivorship using log linear models. *Biometrics* 65, 159 165.

Mayer, K. U. and Tuma, N. B. (1990). *Event History Analysis in Life Course Research*. Madison, WI: University of Wisconsin Press.

Petersen, T. (1993). Recent advances in longitudinal methodology. *Annual Review of Sociology* 19, 425 454.

Plewis, I. (1985). *Analysing Change: Measurement and Explanation Using Longitudinal Data*. Chichester: John Wiley and Sons.

Siegel, J. S. and Swanson, D. (2004). *The Methods and Materials of Demography*. Amsterdam: Elsevier.

Singer, J. D. and Willett, J. B. (2003). *Applied Longitudinal Data Analysis*. New York: Oxford University Press.

Strang, D. and Tuma, N. B. (1993). Spatial and temporal heterogeneity in diffusion. *American Journal of Sociology* 99, 614 639.

Taris, T. W. (2000). *A Primer in Longitudinal Analysis*. London: Sage Publication.

Waldorf, B. (2003). Spatial patterns and processes in a longitudinal framework. *International Regional Science Review* 26, 269 288.

Wrigley, N. (1986). Quantitative methods: The era of longitudinal data analysis. *Progress in Human Geography* 10, 85 101.

Yamaguchi, K. (1991). *Event History Analysis*. Beverly Hills: Sage Publication.

Los Angeles School of Post-Modern Urbanism

B. Hooper, Radboud University Nijmegen, Nijmegen, The Netherlands

Introduction

In certain moments, in certain milieus, particular cities generate a number of issues that incite attention. The work produced by the Los Angeles School of postmodern ur banism (hereafter the LA School or LA scholars) is a diary of one such occasion. In the 1960s and 1970s, Los Angeles (LA), like other cities across the globe, began to experience the transformative effects of a nexus of historical–geo graphical developments which can be generalized as the following: decolonization, globalization, and capitalist re structuring; accelerated urbanization and immigration; the emergence of an increasingly heterogeneous population and an increase in tensions concerning 'the right to the city'; and the onset of what has been called the 'postmodern condition' – that is, a changing perception and experience of space and time and an intensified, sometimes cele bratory, consciousness of the 'new'.

It is this foment out of which a particular set of issues emerged and in response to which in the 1980s the LA School was brought into being. There are two ways to understand the LA School: as a broad based effort to retheorize the contemporary city in which LA itself is the 'school' – a pedagogical feast in which LA is allowed to teach – and as a narrower effort involving particular scholars. While the tendency has been to emphasize the latter, particularly as the school has aged, and clashing claims of ownership and authority have become evident, any accurate understanding of the LA School and its contributions must emphasize the critical massing of people and events which characterize the former. This means, first and foremost, recognizing the originary contributions of the millions of unnamed Southern California residents, documented and undocumented, whose everyday acts have constituted a profound re thinking of LA and the five county metropolitan region in which it is located (LA, Orange, Riverside, San Ber nardino, and Ventura). For the scholars and activists en gaged in this rethinking, of particular importance has been the presence of a supportive institutional framework of universities and research centers which have funded intellectual collaboration across a variety of disciplines and sites (these include but are not limited to the Uni versity of California at LA, Irvine, Riverside, and Santa Barbara; the University of Southern California; the Southern California Institute of Architecture; California Institute of the Arts; the LA Forum for Architecture and Urban Design; the Southern California Library and Institute for Social Research; and the Getty Center for the History of Art and Humanities). It is the dynamic gathering together of people, institutions, and events that yielded contemporary LA as an energized site – both laboratory and Rorschach blot – for investigating the new urban conditions.

Investigative Themes and Sites

In reviewing the accumulated results of the work of the LA School, two things become evident: no governing paradigm or methodology has exerted a hegemonic regulatory force; and despite the work's great diversity, several shared themes are identifiable at the level of phenomena investigated. It is the combination of these elements that is responsible for the extraordinary breadth of its contributions and its being identifiable as a par ticular body of work. With these clarifications in mind, ten thematic areas can be identified.

Postmodernity Is an Urban Condition

The nineteenth century thinkers bequeathed to the LA School the theoretical and esthetic practice of soldering investigations of the metropolis to those of modernity: not only siting modernity in the city, but also conceiving the city as modernity's symbol and most visible socio spatio temporal form. In the late twentieth century, LA, dubbed 'the capital of the twentieth century', was in ducted into this tradition as an exemplary site for investigating both the utopic and dystopic – or more accurately, the heterotopic – conditions of postmodernity. For LA scholars, as for their precursors, the identifying characteristics of modern and now postmodern con ditions are the combined effects of a fragmented, dis continuous experience of time, space, and causality, and an intensely mediated reality. These conditions, con ceptualized as most acute in the immediacy of metro politan relations, disturb notions of a linear temporality and a homogeneous, Cartesian spatiality, and produce the real as a 'precession of simulacra'. LA has been con structed as the incarnation and iconic representation of these dual conditions. In addition, two of the city's forms now circulate globally as icons of the icon. The first is the Bonaventure Hotel in which the visitor, like the new urban subject, is said to be lost in 'postmodern hyper space', acutely disoriented and unable to map her/his surroundings; the second is (of course) Disneyland as hyperreality; for Baudrillard and others, it is 'the perfect model of all the entangled orders of simulacra'.

The Postmodern Condition of Urban Knowledges

The LA School developed within a period marked by a deconstruction of modern Western knowledges and metanarratives and the correlated development of new, 'postmodern' epistemologies, ontologies, and modes of analysis. For LA scholars these developments contributed both to a critical consciousness about knowledge pro duction in general as well as to efforts to restructure urban theory in particular. For some, this restructuring has been driven by post Cartesian epistemological/ ontological considerations (no essentialisms, no founda tionalisms, no universal laws and principles, no master narratives, and no disembodied universal subject, a greater concern with mobility and flows and with the structurality of structures than with structures *per se*), while for others the mobilizing impulse has been more explicitly political – that is, a concern with difference and situatedness in theory production and an inclusion of voices repressed by hegemonic knowledges. However, while the questions raised by the general crisis of Western knowledges have been an influence upon LA scholars, in most cases the primary motive force for re thinking existing urban knowledges has been the misfit between dominant urban paradigms – both those de veloped in the early twentieth century in Europe and Chicago and those developed in the late twentieth cen tury as concerned with postindustrial cities or global cities such as New York, London, Tokyo, and Chicago – and LA's emergent reality.

Thus, if there is something that can be called 'post modern' vis à vis the LA School – and it is important to emphasize that the term 'postmodern' is highly contested – it is the attempt to analyze these new urban conditions, that is, those of contemporary LA, combined with the importation into urban analysis of differing kinds of knowledges (Marxism, semiotics, post structuralism, psychoanalysis, discourse theory, spatial theory, femi nism, postcolonialism, antiracist and gay and lesbian theory, rhizomatics, and so forth) and the inclusion of nontraditional sites of analysis (the lived city, urban bodies and embodiment, the domestic environment, the problematics of the everyday, subject and identity pro duction, the influences of mass media and the culture industry upon urban processes and events, etc.) that warrants the use of the term. This means that in joining 'postmodern' to 'the LA School', care must be taken to specify what is meant – a period or epoch, a specific kind of knowledge, a specific analytical site, particular urban processes, a particular experience of time and space – rather than generalizing across widely disparate analyses. Despite the frequent appendage of the adjective 'post modern' to the LA School, not all LA scholars would adopt this term to describe their works nor are they engaged in developing conceptual frameworks or re search methodologies grounded explicitly in what gen erally would be recognized as postmodern or post structural epistemologies and ontologies.

Space and the Spatial Turn

In the mid nineteenth century, Karl Marx described the transformation of everyday life in industrial modernity as the 'annihilation of space by time'. In postmodernity, 'annihilated' space returned. Michel Foucault stated that "the present epoch will perhaps above all be the epoch of space;" Fredric Jameson described a 'spatial turn' (1991: 154); and Edward Soja wrote that the 'present period' is 'complexly intertwined' with the 'reassertion of space' (1989: 5). Many LA scholars have been at the forefront of this spatial turn and have placed space and spatiality at the center of their analyses. Their primary endeavors have been directed toward a documentation of the spatial specificity of urbanism and urbanization, empirical explorations of the co constitutive (social) production of urban sociality, spatiality, and temporality, and a retheorization of urban spatial processes as gen erative of, rather than derivative from, larger social processes. For Soja it is the spatial that defines the LA School: "If you ain't spatial in an explicit way, you ain't a core member of whatever the LA group of scholars is called" (personal correspondence, September 2006).

Capitalist Restructuring and Globalization

For most, if not all, members of LA School, globalization and capitalist restructuring are the 'base', that is, the dominant forces driving the new urbanization processes. While there are differing degrees of causality assigned to these twinned restructuring processes, many LA scholars argue that the new modes of organizing and experiencing urban life evident in LA have been set in motion by the post 1970 upheavals of global capitalism. Much of this work has been concerned with the investigation of the restructuring of LA's geographies of production and con sumption: a transformation generally characterized as from Fordist to post Fordist and marked by the development of the region as one of the largest industrial agglomerations in the world. In line with the trends predominating in the region, LA scholars have focused on the development of LA's internationalized and not post industrial 'glocal' economy: specifically, the surge in information, service, and finance economies; the replacement of heavy manu facturing with high technology and craft industries, such as aerospace, garment, furniture, jewelry, and movie and television production; and the splitting of the labor market into a high paid/high skill corporate and professional elite and a larger mass of low skilled/low paid workers, predominantly immigrants.

Demographic Restructuring

A major analytical preoccupation of LA scholars has been the changing demography of the Southern California re gion. The population of LA has been transformed entirely by immigration and out migration in the last decades. The region's increasing heterogeneity, the changing com position, and locational patterns of its population, its in creasing density, its increasing homelessness and social and economic polarization – the abyss between the gated en claves of the wealthy and the favela like districts of the poor – have become emblematic of not only contemporary LA but also the global metropolitan future; characterized alternatively as the 'capital of the Third World', 'hetero topolis', 'prismatic metropolis', 'the ethnic quilt', and 'bi polar disorder'.

LA is now a city of minorities, changing from one of the whitest cities in the United States in 1970 (70% Anglo) to one of the world's most ethnically diverse in 1990 (LA County in 1990: Anglo, 40.8% (in 1980, 52.9%); Black, 11.2% (in 1980, 12.6%); Asian and Pacific Islander, 10.8% (in 1980, 6.1%); American Indian, 0.5% (in 1980, 0.6%); and Hispanic, 37.8% (in 1980, 27.6%). For Orange County in 1990: non Hispanic white, 64.5% (in 1980, 78.2%); Black, 1.8% (in 1980, 1.3%); American Indian, 0.5% (in 1980, 0.7%); Asian, 10.3% (1980, 4.8%); His panic, 23.4% (in 1980, 14.8%)). Despite its new 'diversity', however, LA retains its historically present and ethnically correlated social, spatial, and economic inequalities.

Posturban Morphologies

While 'the city' has remained the analytical dominant for LA scholars, they have de and reformed its meanings. LA's posturban form is a complex, networked urban ization in which older terms such as city, center, suburb, periphery, and hinterlands no longer apply. What has evolved over the last decades is a polycentric metro politan region characterized by dispersed patterns of growth which no single center dominates and in which edge cities proliferate. In fact, postsuburban Orange County, as the largest metropolitan area in the country with no central metropolis, may have exceeded LA as the new urban prototype. This polycentric, postmetropolitan, postsuburban conurbation is at the same time massive: a mega urban agglomeration in which borders between cities and counties blur. In 2000, the LA metropolitan region, one of only two 'megacities' in the imaginative geography typically named the 'North', had a population of over 15 million.

Hetero-Architecture

In contrast with the US East Coast's neomodernism, an innovative LA architectural community – including, for example, Frank Gehry, Frank Israel, Charles Moore, Morphosis, Eric Owen Moss, and Greg Lynn – has made important contributions to restructuring the spaces of LA. What Charles Jencks terms LA's new hetero archi tecture is characterized by a (post Cartesian) *ad hocism*, radical eclecticism, postmodernism, inclusionism, and technomorphism. Deploying a more politically charged redefinition of architecture, Teddy Cruz has taken the cross border metropolis of San Diego Tijuana as the focus of his research and experimentation. Cruz's con cepts of 'learning from Tijuana' and 'urban acupuncture' have produced particularly innovative designs in housing.

Utopic/Dystopic LA

The LA School has evidenced strong interest in an in herited bipolar view of LA as both utopic and dystopic. In contemporary LA, these designations are no longer tropes but have been built into the city as two highly visible technologies of social control, the Foucauldian carceral city and the Baudrillardian simulated city. Carceral LA is an urban panopticon of repression and surveillance which disciplines through the dystopic uto pia of privatized resources, securitized buildings, gated communities, electronic policing technologies, and a military like occupation of ghetto areas. The Bau drillardian panopticon of urban simulacra disciplines through the production of an urban imaginary of fantasy environments, such as Disneyland or Citywalk, which overcode the more dystopic realities of social and eco nomic polarization evident on LA's largely unwalked streets. While disciplinary technologies have always existed in cities, what is new, LA scholars suggest, is that the disciplinary effects of theme parks and prisons have not only merged but also escaped their former confines to become omnipresent and inescapable. The documented political effects of Foucauldian and Baudrillardian dys topic utopias are substantial: both have tended to operate to reproduce LA's ethnically correlated 'dual' political–cultural economy and, in the case of simulated reality, operate to overcode lived realities, thus valorizing certain realities and voiding others.

The Politics of Nature

While not a major area of interest for the LA scholars, the politics of nature has become an increasingly important one. At the forefront of this research are the works of FitzSimmons and Gottlieb on Los Angeles' institutional strategies of environmental control; Davis on the political history of the Los Angeles landscape; and Pulido on environmental racism and environmental justice. In light of both the massive environmental degradation, a city like Los Angeles produces, and Jameson's contentious statement: "Postmodernism is what you have when the modernization process is complete and nature is gone for good" (1991: ix), this area is ripe for further research.

Utopia and Dystopia Redux: The Right to the City

For some LA residents, the new demographic diversity has been interpreted as the loss of social cohesion and a virulent postmodern anomie, while for others it means new cultural hybrids, new cosmopolitical opportunities, and new possibilities for renegotiating citizenship rights and the urban social contract. Either way, dystopic or utopic, there can be no doubt that there has been an increase in an ethnically and class charged LA politics focused on 'the right to city', in which for many the right to survive (to be housed, fed, educated, and provided with adequate health care) are the stakes. The 1992 up rising following the not guilty verdict in the trial of police officers accused in the beating of Rodney King is often seen as emblematic, and symptomatic, of LA's long existing racial and ethnic inequalities, which have been further exacerbated by the competing claims of new immigrants for voice and space. For some, Los Angeles has become a test case for the negotiation of the urban future and social and spatial justice.

Los Angeles: The Exception or the Rule?

LA scholars inherited a tradition of LA exceptionalism: how have they reinterpreted it? Is LA the exception or the rule, like or unlike other cities? Is it prototype, archetype, or stereotype – or is it an 'ordinary city' that the LA School has marketed in hyperbolic terms? Is it the new urban paradigm, or is it the paradigm of the ex ception? While it is clear that all of the trends presented above are evident in numerous other cities, LA scholars have oscillated between positioning LA as a highly ex ceptional city, a harbinger city, and an exemplary city. Time will tell. Perhaps the greatest contribution LA scholars can make to urban studies is not in replacing one paradigm or school with another but in bequeathing LA as a set of lived problems and issues, it is worthwhile to explore.

See also: Dear, M. J.; Postmodernism/Postmodern Geography; Scott, A.; Soja, E.; Urbanism.

Further Reading

Baudrillard, J. (1997). *Simulacra and Simulation,* p 12. Ann Arbor, MI: University of Michigan Press.

Beauregard, R. (2003). City of superlatives. *City & Community* 2(3), 183 199.

Bobo, L. D., Oliver, M. L., Johnson, J. H. Jr. and Valenzuela, A. (2000). *Prismatic Metropolis: Inequality in Los Angeles*. New York: Russell Sage Foundation.

Brenner, N. (2003). Stereotypes, archetypes, and prototypes: Three uses of superlatives in contemporary urban studies. *City & Community* 2(3), 205 216.

Cenzatti, M. (1993). *Los Angeles and the L A. School.* Los Angeles Forum for Architecture and Urban Design.

Chase, J., Crawford, M. and Kaliski, J. (1999). *Everyday Urbanism*. New York: Monicelli Press.

Curry, J. and Kenney, M. (1999). The paradigmatic city: Postindustrial illusion and the Los Angeles School. *Antipode* 31(1), 1 28.

City 1/2: *It All Comes Together in Los Angeles Or*?

Davis, M. (1990). *City of Quartz: Excavating the Future in Los Angeles*. London: Verso.

Davis, M. (1996). How Eden lost its garden: A political history of the Los Angeles landscape. In Scott, A. J. & Soja, E. W. (eds.) *The City: Los Angeles and Urban Theory at the End of the Twentieth Century*, pp 160 185. Berkeley, CA: University of California Press.

Dear, M. (2001). *The Postmodern Urban Condition*. Oxford: Blackwell.

Dear, M. (2002). *From Chicago to L.A.: Re visioning Urban Theory.* Thousand Oaks, CA: Sage Publications.

Dear, M. and Flusty, S. (1997). The iron lotus: Los Angeles and postmodern urbanism. *Annals of the American Academy of Political and Social Science* 551(1), 151 163.

Dear, M. and Flusty, S. (1998). Postmodern urbanism. *Annals of the Association of American Geographers* 88(1), 50 72.

Dear, M., Schockman, H. E. and Hise, G. (eds.) (1996). *Rethinking Los Angeles*. Thousand Oaks, CA: Sage Publications.

Dunitz, R. J. (1993). *Street Gallery: Guide to 1000 Los Angeles Murals*. Los Angeles, CA: RID Enterprises.

Fitzsimmons, M. and Gottlieb, R. (1996). Bounding and binding metropolitan space: The ambiguous politics of nature in Los Angeles. In Scott, A. J. & Soja, E. W. (eds.) *The City: Los Angeles and Urban Theory at the End of the Twentieth Century*, pp 186 224. Berkeley, CA: University of California Press.

Foucault, M. (1986). Of other spaces. *Diacritics: A Review of Contemporary Criticism* 16(1), 22 27.

Fulton, W. (1997). *The Reluctant Metropolis: The Politics of Urban Growth in Los Angeles*. Point Arena, CA: Solano Press.

Garreau, J. (1991). *Edge City: Life on the New Frontier*. New York: Doubleday.

George, L. (1992). *No Crystal Stair: African Americans in the City of Angels*. New York: Doubleday.

Gooding Williams, R. (ed.) (1993). *Reading Rodney King, Reading Urban Uprising*. New York: Routledge.

Gottdeiner, M. (1995). *Postmodern Semiotics: Material Culture and the Forms of Postmodern Life*. Oxford: Blackwell.

Gottdiener, M. (2002). Urban analysis as merchandising: The 'L.A. School' and the understanding of metropolitan development. In Eade, J. & Mele, C. (eds.) *Understanding the City: Contemporary and Future Perspectives*. Oxford: Blackwell.

Gottdiener, M. and Kephart, G. (1991). The multinucleated metropolitan region. In Kling, R., Olin, M. & Poster, M. (eds.) *Postsuburban California: The Transformation of Orange County Since World War II*, pp 31 54. Berkeley, CA: University of California Press.

Hise, G. (1997). *Magnetic Los Angeles: Planning the Twentieth Century Metropolis*. Baltimore, MD: Johns Hopkins University Press.

Hooper, B. (2000). Bodies, cities, texts: The case of citizen Rodney King. In Soja, E. W. (ed.) *Postmetropolis: Critical Studies of Cities and Regions*, pp 359 371. London: Blackwell.

Jameson, F. (1991). *Postmodernism, or the Cultural Logic of Late Capitalism,* p 39. Durham: Duke University Press.

Jencks, C. (1993). *Heterotopolis: Los Angeles, the Riots, and the Strange Beauty of Hetero Architecture*. London: Academy Editions/ Ernst and Sohn.

Keil, R. (1998). *Los Angeles: Globalization, Urbanization and Social Struggles*. New York: Wiley.

Kenney, M. (2001). *Mapping Gay L.A.: The Intersection of Place and Politics*. Philadelphia, PA: Temple University Press.

Klein, N. (1997). *The History of Forgetting: Los Angeles and the Erasure of Memory*. New York: Haymarket Series/Verso.

Kling, R., Olin, M. and Poster, M. (eds.) (1991). *Postsuburban California: The Transformation of Orange County Since World War II*. Berkeley, CA: University of California Press.

Kling, R., Olin, S. and Poster, M. (eds.) (1995). *Postsuburban California: The Transformation of Orange County Since World War II*. Berkeley, CA: University of California Press.

Lavin, S. (2005). *Form Follows Libido: Architecture and Richard Neutra in a Psychoanalytic Culture*. Cambridge, MA: MIT Press.

Lefebvre, H. (1996). The right to the city. In Kofman, E. & Lebas, E. (trans.). *Writings on Cities.* Oxford: Blackwell.

Los Angeles Times, Staff (1992). *Understanding the Riots; Los Angeles Before and After the Rodney King Case*. Los Angeles: The Los Angeles Times.

Lynn, G. (1998). *Folds, Bodies and Blobs; Collected Essays*. Brussels: La Lettre vole.

Madhubuti, H. R. (1993). *Why L.A. Happened: Implications of the '92 Los Angeles Rebellion*. Chicago, IL: Third World Press.

MegaCity Task Force of the International Geographical Union, Department of Geography, University of Koln, Germany. http://www.megacities.uni koeln.de/documentation/start.htm Accessed in Aug. 2008.

Miller, D. W. (2000). The new urban studies. *Chronicle of Higher Education* 46(50), A15 A16.

Morrison, T. and Lacour, C. B. (eds.) (1997). *Birth of a Nation'hood: Gaze, Script, and Spectacle in the O.J. Simpson Case*. New York: Pantheon Books.

Ong, P. and Blumberg, E. (1996). Income and racial inequality in Los Angeles. In Scott, A. J. & Soja, E. W. (eds.) *The City: Los Angeles and Urban Theory at the End of the Twentieth Century*, pp 311 335. Berkeley, CA: University of California Press.

Pulido, L. (1996). Multiracial organizing among environmental justice activists in Los Angeles. In Dear, M., Schockman, H. E. & Hise, G. (eds.) *Rethinking Los Angeles*. Thousand Oaks, CA: Sage Publications.

Pulido, L. (2006). *Black, Brown, Yellow, and Left: Radical Activism in Los Angeles*. Berkeley, CA: University of California Press.

Reid, D. (ed.) (1994). *Sex, Death, and God in L.A*. Berkeley, CA: University of California Press.

Rieff, D. (1991). *Los Angeles: Capital of the Third World*. New York: Touchstone.

Ruddick, S. M. (1996). *Young and Restless in Hollywood: Mapping Social Identities*. London: Routledge.

Sandercock, L. (1998). *Towards Cosmopolis: Planning for Multicultural Cities*. Chichester: Wiley.

Sassen, S. (1991). *The Global City: New York, London, Tokyo*. Princeton, NJ: Princeton University Press.

Scott, A. J. (1993). *Technopolis: High Technology Industry and Regional Development in Southern California*. Berkeley, CA: University of California Press.

Scott, A. J. (1999). Los Angeles and the LA School: A response to curry and Kenney. *Antipode* 31(1), 29 36.

Scott, A. J. and Soja, E. W. (eds.) (1996). *The City: Los Angeles and Urban Theory at the End of the Twentieth Century*. Berkeley, CA: University of California Press.

Smith, A. D. (1994). *Twilight Los Angeles, 1992*. New York: Anchor Books.

Soja, E. (1989). *Postmodern Geographies: The Reassertion of Space in Critical Social Theory,* p 5. London: Verso.

Soja, E. W. (2000). *Postmetropolis: Critical Studies of Cities and Regions*. London: Blackwell.

Soja, E. W. and Scott, A. J. (1986). Los Angeles: Capital of the nineteenth century. *Society and Space* 4, 249 254.

Soja, E. and Scott, A. J. (1986). Los Angeles: Capital of the twentieth century. *Society and Space* 4, 249.

Song, M. H. (2005). *Strange Future: Pessimism and the 1992 Los Angeles Riots*. Durham: Duke University Press.

Sonnenshein, R. (1993). *Politics in Black and White: Race and Power in Los Angeles*. Princeton, NJ: Princeton University Press.

Storper, M. and Walker, R. (1989). *The Capitalist Imperative: Territory, Technology, and Industrial Growth*. Oxford: Basil Blackwell.

Turner, E. and Allen, P. (1991). *An Atlas of Population Patterns in Metropolitan Los Angeles and Orange Counties*. Northridge, CA: Center for Geographical Studies, California State University.

Valle, V. M. and Torres, R. D. (2000). *Latino Metropolis: Globalization and Community.* Minneapolis, MN: University of Minnesota Press.

Waldinger, R. and Bozorgmehr, M. (1996). *Ethnic Los Angeles*. New York: Russell Sage Foundation.

Wolch, J. and Dear, M. (1993). *Malign Neglect: Homelessness in an American City*. San Francisco, CA: Jossey Bass.

Lowenthal, D.

J. R. Gold, Oxford Brookes University, Oxford, UK

Glossary

Geosophy The study of geographical knowledge from any or all points of view.

Lowenthal, David (1923–)

If, to quote L. P. Hartley, "the past is a foreign country," then David Lowenthal is one of its most fluent and regular geographical visitors. From the outset, his work blended geographical and historical interests and meth odologies. Born 26 April 1923, Lowenthal was educated at Harvard University (BS History, 1943) and, after war service, at the University of California, Berkeley (MA Geography, 1950) and the University of Wisconsin – Madison (PhD History, 1953, although minoring in geography). His doctorate, completed in the History Department at Wisconsin under the supervision of the historian of ideas, Merle Curti, developed a topic first suggested by the cultural geographer Carl Sauer, with whom Lowenthal had studied at Berkeley. His first major appointment (Vassar College, 1952–56) combined the role of Assistant Professor in History with the headship of the Geography Department. From 1956 onward, he worked as a professional geographer – first as Research Associate at the American Geographical Society (AGS) in New York (1956–72), then as Professor of Geography at University College London (UCL, 1972–85, with subsequent Emeritus status) – but his concern with the past continued unabated. Indeed, the linking thread underpinning Lowenthal's wide ranging research always lay in the application of meticulous historical scholarship to subjects that included culture and landscape, per ception and environment, heritage and conservation, coupled with abiding interests in West Indian and American societies.

A key early theme in Lowenthal's research lay in identifying the origins and anatomy of thinking about the relationship between people and their environment. His doctorate on the nineteenth century Vermont geog rapher George Perkins Marsh highlighted Marsh's view of nature as a cultural construction and the extent of society's impact on the natural environment; ideas that have seen Marsh recast as a proto conservationist. Similarly, when reviewing the contribution made by John Kirtland Wright, an erstwhile colleague at the AGS, Lowenthal recognized that within Wright's all encompassing intellectual framework of 'geosophy' lay a style of inquiry that highlighted the impact of en vironmental ideas on thought and action and, ultimately, on the environment itself. A frequently cited paper on 'Geography, experience and imagination', published in 1961, saw Lowenthal draw on Wright's agenda to rec ognize that environmental knowledge reflected per ception rather than simply mirroring 'objective reality'. From this standpoint, people live in parochial worlds, shaped by selective perceptual filtering of environmen tal information but bound into the common realm of experience through culture.

The thinking behind his 1961 paper proved influen tial. Lowenthal developed the underlying ideas himself in a series of essays on the interpretation of landscape as a reflection of cultural norms (some written with Hugh Prince). These helped to reestablish 'landscape', often dismissed as a vague notion with little analytic value, as an important concept in geographical research – a view that influenced a subsequent generation of researchers that included Denis Cosgrove, Stephen Daniels, and Kenneth Olwig. Lowenthal's 1961 paper also served as one of the founding texts of the environmental per ception movement that emerged in geography and cog nate disciplines in the late 1960s, although its author was always ambivalent about approaches that went beyond studying perception of environment to assert linkage between perceived worlds and human behaviour (as with 'behavioural geography'). Less directly but still per ceptibly, echoes of the ideas of living in individually perceived worlds with shared cultural constructions res onated with the work of humanistic geographers and cultural geographers in the late 1970s and 1980s.

During the 1970s, Lowenthal became increasingly concerned with the features that made places and land scapes valued, characteristically assembling widely dis parate historical and contemporary examples to explore the connections between past societies and landscape. A 1975 paper on 'Past time, present place', adapted from his inaugural professorial lecture at UCL, emphasized the uses of the past and the way that people shape landscapes to conform with shared narratives that express tastes and values. This focus led partly to research on 'valued landscapes', but rather more directly suggested the ap proach found in his books *The past is a foreign country* and *The heritage crusade and the spoils of history*. The former explored the significance of the past for modern society, how we come to know it, and the ways in which we re construct it in the light of current needs. The latter pro vided a challenging account of the cultural politics of conservation, noting how history, appropriated and packaged as ‘heritage’, could be used for good or ill to support and legitimize narratives of, *inter alia*, national, ethnic, and religious identity. Addressing topics of increasing public concern, both books appealed to a wider audience than the normal readership for academic geography texts.

His writings are not without their critics. Mark Billinge, for example, argued that Lowenthal’s writing style represented a mild precursor of a tendency, especially amongst humanistic geographers, toward conflating overlong sentence construction and extravagant language with expression of deep meaning. Reviewers of his heritage books readily admired Lowenthal’s sophisticated and investigative recounting of detail, but sometimes felt that within his kaleidoscopic surveys lay an overemphasis on empiricism at the expense of broader theorizing. This orientation perhaps explained why these books appealed less to the ‘new cultural geography’, with its strong emphasis on critical theory, than to the growing field of heritage studies. Against these points, Lowenthal’s consistency as a scholar over more than half a century has contributed greatly to the content and reputation of historical geographical study. His success in forging links between geography and the humanities, in making a significant contribution toward the evolving multidisciplinary fields of environmental history and heritage studies, and furthering the understanding of place and landscape all point to the lasting value of his work.

See also: Behavioral Geography; Cultural Geography; Heritage; Humanism/Humanistic Geography; Landscape; Sense of Place.

Further Reading

Billinge, M. (1983). The Mandarin dialect: An essay on style in contemporary geographical writing. *Transactions of the Institute of British Geographers* 8, 400 420.

Lowenthal, D. (1958). *George Perkins Marsh: Versatile Vermonter*. New York: Columbia University Press.

Lowenthal, D. (1961). Geography, experience, and imagination: Towards a geographical epistemology. *Annals of the Association of American Geographers* 51, 241 260.

Lowenthal, D. (1972). *West Indian societies*. London: Oxford University Press.

Lowenthal, D. (1975). Past time, present place: Landscape and memory. *Geographical Review* 65, 1 36.

Lowenthal, D. (1982). Finding valued landscapes. *Progress in Human Geography* 2, 373 418.

Lowenthal, D. (1985). *The past is a foreign country*. Cambridge: Cambridge University Press.

Lowenthal, D. (1996). *The heritage crusade and the spoils of history*. London: Viking.

Lowenthal, D. and Bowden, M. (eds.) (1976). *Geographies of the mind: Essays in historical geosophy in honour of John Kirtland Wright*. New York: Oxford University Press.

Lowenthal, D. and Prince, H. C. (1964). The English landscape. *Geographical Review* 54, 309 346.

Olwig, K. R. (2003). Forum: The Lowenthal papers. *Annals of the Association of American Geographers* 93, 851.

Lusophone Geography

J. Gaspar, CEG, University of Lisbon, Lisbon, Portugal

Introduction

While it may be easy to define *lusofonia* as a set of Lusophone communities, it is far more difficult to define the territorial limits to its expression or its space.

The term *lusofonia* is derived from 'Lusophone', a Portuguese speaking person or a country for which Portuguese is an official language. It entered into com mon usage and quickly spread in the second half of the 1970s, following the independence of the Portuguese colonies in Africa. It could therefore be considered a postcolonial concept. *Lusofonia* also appears to be asso ciated with the notion of cultural convergence: "a set of existing cultural identities among the eight countries for which Portuguese is an official language."

This concept, however, is lacking insofar as it does not provide for any space for overlap between the political reality of the eight countries on the one hand, and the cultural and identity based dimension among Portuguese speakers on the other. Therefore, it is useful to distinguish between political *lusofonia*, as embodied by Lusophone countries, and Lusophone communities spread across the globe as a result of particular historical processes, with the largest communities comprising emigrants hailing from Portugal, followed by emigrants from Brazil and Cape Verde.

Though a concept that has acquired currency only in the last quarter of the twentieth century, it is surpris ing that it has been adopted not only among Lusophone communities but also outside of them, beyond the eight countries and the communities of the Portuguese speaking diaspora.

Identification with the concept of *lusofonia* is such that some universities in non Lusophone speaking coun tries have begun degree programs in *lusofonia*. Some authors include the concept with *Francophonie* and the Commonwealth.

Furthermore, there is evidence not only of an 'inte gration' of the diasporas under the umbrella of *lusofonia* but also that the concept's expansion correlates with migratory growth, particularly from Brazil and the young African countries for which Portuguese is an official language.

In regard to geographical studies where close links have long existed between the Lusophone communities of geographers, the emergence of this concept did not provoke new research paths or degree programs in 'Lusophone geography' or the 'geography of *Lusofonia*'.

One might suggest that geographers from Lusophone countries, particularly from Brazil and Portugal with their large number of established institutions, are late to adopt the innovative *lusofonia* concept. Furthermore, linguistic identity facilitates the standardization of geo graphical practice in Lusophone space, particularly as regards teaching.

Three Spaces and Three Times

The development of geography as a discipline in the many spaces encompassed by *lusofonia* is temporally and spatially contingent.

Portugal

Before its modern, academic institutionalization in 1911, the roots of Portuguese geography can be traced back to the fouteenth century with the definition of the new country's territorial limits and to the fifteenth century preparation and undertaking of maritime travel and ex ploration, which led to advances in cartography and de scriptive geography of lands that came to be known.

Another dynamic period in applied geography and cartography can be identified in the eighteenth century in the Portuguese *métropole*, with the modernization of the country and the reconstruction of Lisbon and the Algarve region following the 1755 earthquake, but above all with the occupation and colonization of Brazil, leading to the confluence of geographical and cartographic knowledge in order to support military action and economic ex ploitation that was reliant upon a planned network of urban centers and other bases.

The second half of the nineteenth century saw other 'opportunities' arise for the affirmation of geography, with the advances of Carl Ritter and Alexander von Humboldt. The Lisbon Geographical Society, established in 1875, was important for legitimating the occupation of African lands, particularly through its contribution to the 1885 Berlin Conference.

Brazil

Brazilian geography is a discipline that was constructed along with the country from the moment of independ ence in the last quarter of the nineteenth century. As it severed its colonial ties, geography – though implied in other discourses – was instrumental for the construction of a national identity and to affirm the legitimacy of the country's borders with the surrounding countries marked by Spanish decolonization.

Brazil presents a particularly interesting case of implantation, discussion, and development of modern geography in the mid nineteenth century, as it was not concurrent with its academic institutionalization nor with its introduction as an autonomous discipline at any educational level.

The founding concepts of modern geography, namely, those put forth by Ritter and Humboldt, assumed an important role in supporting the creation of a national identity, one based on the internal recognition of the territory and its occupation, and on the definition and affirmation of its political borders.

Meanwhile, new developments in geography in Eur ope and North America were integrated into the fringes of a formal academic system of geography teaching and research. However, this does not hinder the appearance of scientific studies that clearly accompany the devel opments stemming from the main European and North American schools. This is evidenced in the field of human geography with figures of recent international acclaim such as Josué de Castro and Milton Santos (**Figure 1**) that hail from scientific areas outside of geography.

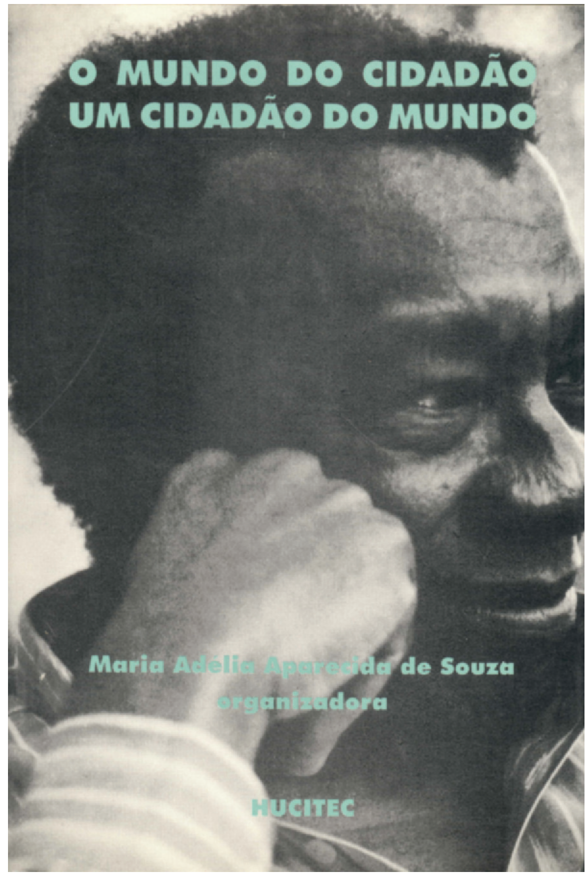

Figure 1 Milton Santos (1926 2001).

Geographical Studies in the New Postcolonial, Lusophone States

There is a significant 'geographical heritage' from the colonial period that has proven relevant for a series of important geographical studies in geography, from physical geography to urban geography, as well as a wealth of up to date cartographic and statistical re sources fundamental to the research in geography.

Autonomous university level teaching of geography was introduced with mixed success in the late 1960s, just prior to decolonization, at the universities in Mozambique and Angola. While the teaching body remained at Ernesto Mondlane University in Maputo, the department at Lu anda University's Lubango (Sá da Bandeira) campus fell completely apart due to the civil war following Angola's independence and could not easily reconstitute itself.

The continuity of the 'human geography school' in Mozambique and its continued participation in inter national geography networks is largely due to Manuel Araújo. His research has focused on three areas: internal migration, rural settlement systems, and urban geography.

Geography in higher education came into the scene for the first time in 1962 as part of a teacher training program. It was only in 1968 that an abbreviated undergraduate degree program in geography, with physical and human geography components, was offered. Following independence, in 1975, there was a time in which autonomy was lost over the teaching of geography. It could only be recovered in 1990 with the creation of a 5 year undergraduate degree program meant not only for training teachers but also regional and urban planners.

The 2001 curriculum reform reduced the under graduate degree program to 4 years and introduced a stronger planning orientation, renaming it a degree in geography, regional development, and the environment.

At Eduardo Mondlane University, there is also an interesting geography teaching component in the archi tecture and physical planning degree program.

In Angola, while teaching in geography was the same as that in Mozambique during the colonial period, the Instituto Superior de Ciências da Educação (Higher In stitute for Educational Sciences) exists today as part of Agostinho Neto University located on three campuses: Luanda, Huambo, and Lubango. The 4 year under graduate degree in geography includes a strong didactic and pedagogic component, though the university is considering a series of courses in regional and local planning as evidenced by its interest in geographic in formation system (GIS) and land use planning and en vironmental impact assessment.

A unique case can be found in Cape Verde, where a university structure has emerged along with the construction of the country following independence in relation to the needs of the country. Therefore, it was in the Instituto Superior de Educação (ISE; Higher Institute for Education), initially limited to the training of primary and secondary school teachers, that teaching geography at the university level began in the 1980s in order to prepare some students for university educations in Por tugal. Two 4 year undergraduate degree programs were created later on, one oriented toward teaching and the other toward planning. The ISE joined the newly estab lished Cape Verde University in 2000.

The Institutionalization of Human Geography in Portugal and Brazil

University level teaching of geography in Portugal goes back a long way in Coimbra University's College of Humanities. In Lisbon, it began in 1901 in the Higher Humanities Degree Program, in which human geography was favored, thanks to the influence of Prof. Silva Telles, who in 1904 defended his doctoral dissertation entitled *The Concept of Geographical Unites: An Introduction to Anthropogeography*.

But it was with university reform in 1911 that actual academic institutionalization of geography began, with its teaching being integrated into a 4 year undergraduate degree program in history and geography. The programs overseen by the universities in Lisbon and Coimbra were based upon the French university system, and their his torical and geographical content mainly inspired by French schools of thought, particularly in respect to human geography. However, German and Anglophone authors were not overlooked – rather, they offered guidance thanks in part to the influence of Silva Telles.

It was, however, in Coimbra that modern Portuguese human geography first came into being, due to the re search and teaching of Amorim Girão (1895–1960), who, starting from a regional approach, expanded the range of research topics in human geography to include popu lation, urban, and rural studies as well as making some pioneering steps toward lesser known areas, such as the study of religious spaces.

In 1930, in the context of reforms hastily advanced by the dictatorial regime installed in 1926, geography was separated from history and a 4 year degree program in geographical sciences was set up to include subjects from both the college of humanities and the college of sci ences. It sought to 'liberate' geography from the more ideological predilection of the historical component in favor of rendering it a more 'scientific' discipline like physics, mathematics, geology, etc.

While Jean Brunhes had a strong influence on the geographical thought of Amorim Girão, the affirmation of French geography came with the work of Vidal de la Blanche, whose *Principes de Geographie Humaine* in 1922 was translated and annotated by Alfredo Fernandes Martins, protégé and successor of Amorim Girão in Coimbra. Vidal de la Blanche also had a noteworthy influence on Orlando Ribeiro and the theoretical and epistemological orientations of the so called Lisbon School.

Interestingly, it was from this context that the best known name in modern Portuguese, Orlando Ribeiro (**Figure 2**), that geography emerged. Ribeiro had a solid grounding in the humanities and natural sciences thanks to his contacts with the French school, German geographers, and, though to a lesser extent, British and American geographers. In 1943 at Lisbon University, Ribeiro estab lished what would become Portugal's principal research body in geography, the Centro de Estudos Geográficos (CEG, Center for Geographical Studies). At CEG, he later organized the first International Geography Congress (IGU) following the end of the Second World War.

CEG came to play an important role in the production of geographical knowledge of the Portuguese colonies, from Cape Verde to Timor, as it received support from the Ministries of Overseas Possessions and Education. In this sense, we can speak of a certain continuity in geography's position in the affirmation of Portugal as a colonial power, which stems back to the Sociedade de Geografia (Portu guese Geographical Society) that was established in 1875

Figure 2 Orlando Ribeiro (1911 1997).

and remained relevant through academic geographical research until 1974.

The academic institutionalization of geography in Brazil occurred with the establishment of higher education programs at the universities in Rio de Janeiro and São Paulo in 1934 during a very specific political period characterized by the authoritarian '*Estado Novo*' (new state) regime led by Getúlio Vargas. The regime sought to modernize the country, emphasizing the importance of education, economic development, and nationalism.

Geography proved clearly highly instrumental in this context at three different levels: ideology via education; technology via contributions of geographical engineering; and physical geography via the greater acquisition of knowledge about resources and its ability to lend legitimacy to political measures geared toward territorial programs and interventions.

As some authors have noted, it was in this period that the modern device of state became duly territorialized. R. Morais observes, "It was in this way that a new material geography of the country was constructed, which was accompanied by a new symbolic construction of national identity: the national now clearly expressed as being in the realm of the state and officialdom."

At the same time, the state and the geographical community 'converged' their will as the geographical community came to see the opportunity and the necessity to organize themselves. The Associação dos Geógrafos Brasileiros (Association of Brazilian Geographers) was founded in 1935 on the initiative of civil society. The state was to later establish the Conselho Nacional de Geografia (National Geography Council) in 1937 and the Instituto Brasileiro de Geografia e Estatística (Brazilian Institute for Geography and Statistics) in 1939. These three institutions still exist today, having survived various political, social, and cultural transformations.

At the beginning, they received decisive support from European professors, among them were the French professors Pierre Deffontaines and Pierre Monbeig, who advanced the foundations of modern human geography in Brazil following the French masters Vidal de la Blanche and Jean Brunhes. The former began with São Paulo University (USP) and then moved on to Federal District University (UDF) in 1936, followed by Brazil University (UB) and the present day Federal University of Rio de Janeiro (UFRJ), while the latter had a longer stint at USP.

Lusophone Geography in the Second Half of the Twentieth Century

Main Areas of Research, Paradigms, and Praxis

Portugal

Until the late 1960s, human geography in Portugal followed the French 'possibilist' tradition, influenced by authors such as Pierre Gourou in the 'Geography of tropical regions', J. Beaujeu Garnier and Philippe Pinchemel in *Urban Geography*, and Jean Labasse and Michel Philiponeau in *Applied Geography*.

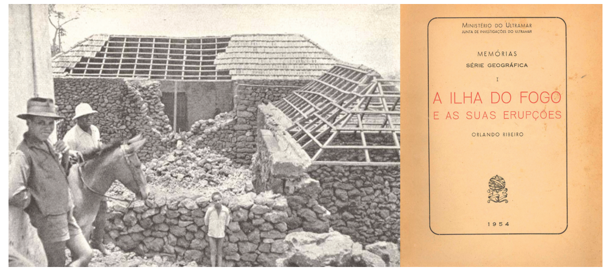

Figure 3 *A Ilha do Fogo* by O. Ribeiro.

Some of the most noteworthy monographs of the period were the extensive and deep reaching works dedicated to Portugal's colonial holdings, such as *A Ilha do Fogo* by O. Ribeiro (**Figure 3**), *Ilha de S. Tomé* (*The Island of São Tomé*) by F. Tenreiro, and *Santiago de Cabo Verde* (*Santiago in Cape Verde*) by J. do Amaral. To these we would add a work from the 1970s that comprises the most complete account of the origins of the Portuguese settlement in Southern Angola, *A Colonização das Terras Altas da Huila* by C. A. Medeiros.

The neopositivist new geography paradigm emerged in the 1960s and came to dominate by the mid 1970s, influenced in particular by T. Hägerstrand as well as by Anglophone authors like P. Haggett, B. Berry, and P. Gould. The first work undertaken in Portugal along these new lines was *Área de Influência de Évora* by J. Gaspar.

At the same time, a significant number of geographers of various origins chose Portugal as the site for human geography research, and in some cases lectured in Portuguese universities, mainly in Lisbon and Coimbra. This was the case with the French, such as Y. Leloup, M. Poinard, M. Drain, F. Guichard, Antoine Bailly, and S. Daveau, who settled in Portugal permanently, producing highly competent work; the Germans, such as H. Lautensach, P. Weber, and B. Freund; the Americans, such as D. Stanislavski and P. Gould; the Italian, Gaetano Ferro; and with the British, A. Williams and C. Jensen Butler, quite possibly the foreign geographer that most influenced the course of Portuguese human geography in the last 20 years.

Several young Portuguese geographers received scholarships to attend universities abroad, at first mainly in France, but later on and in greater numbers to the United Kingdom. Smaller numbers of students studied in Germany, Switzerland, Sweden, the United States, and the Netherlands. From the 1990s onward, a growing number of students sought out Spanish universities.

Human geography developed out of context of regional geography. Rural issues dominated at first, though later economic and urban geographies attracted more attention. An exception to this was the concise works by O. Ribeiro, *Portugal*: *O Mediterrâneo e o Atlântico* and *Mediterrâneo*.

Urban geography began with monograph accounts of the Portuguese and African cities of Oporto, Braga, Évora, Luanda, and Maputo. Yet, the 1970s marked the diversification of the subdiscipline, with emphasis placed on studies of urban networks and systems as well as on real estate, housing, commerce, culture, and urban renewal and rehabilitation.

In the 1980s and 1990s, economic geography (addressing topics such as industry, services, and tourism) with its different theoretical approaches – from new geography to neo Marxism, held significant weight in the production of Portuguese human geography.

Studies on modes of transport and their impact on the use and organization of space constitute a recurring theme in Portuguese human geography. Portugal's integration into the European Union (EU) and the reconfiguration of European space breathed new life into geographical studies on road, rail, air, and sea transport.

Population geography and, more recently, studies on internal and international migration first gained salience in the 1950s in Portugal but has evolved over time both theoretically and methodologically.

Social geography, which is a strand of human geography that concentrates particularly on social issues, has tackled topics like social segregation of space, family, and gender in relation to the labor market, and different social groups' access to goods and services.

From the 1980s onward, we see the emergence of health geography, sport geography, and other domains of spatial inquiry such as the spatial diffusion of innovation and future studies.

Though a recurring theme since the start of the twentieth century that has inspired a number of works in the Portuguese human geography, the study of landscapes never led to the development of a body of consolidated and coherent research.

At the theoretical level, it is difficult to discern dominant paradigms, as a range of theoretical and methodological approaches coexist – leading at times to confusion or contradictions – in what we might refer to as pluralist approaches. Thus, humanist, critical, and post structural approaches coexist with one another, while the emergence of an ecologist paradigm has become clear along with the renewal of neopositivist approaches made 'credible' by the potential of GIS.

Brazil

Brazil's socioeconomic development, the size of the country, and its demographics as well as its stabilized democratic regime has positioned Brazil among the world's greatest powers. Geography has both contributed to and is influenced by this process. Institutions like the Instituto Brasileiro de Geografia e Estatística (Brazilian Institute for Geography and Statistics) and the main university research centers have endowed the country with researchers, teachers, and planners spread throughout Brazil that are significantly engaged at the regional and local levels. Notwithstanding the importance of the large traditional centers of São Paulo and Rio de Janeiro, we can speak of a geographical bipolarization in Brazilian human geography, with very active groups from the north to the south of the country that are part of a variety of networks that regularly gather at scientific, technical, and professional meetings.

The north is anchored by the Federal University of Pernambuco (UFP) in Recife and the Federal University of Baía (UFB) in Salvador. From the UFP comes Manuel Correia de Andrade, an economist by training whose major contributions to Brazilian human geography have been recognized at the regional, national, and international scales. This university, home to a number of Brazilian geographers, has attracted geographers from abroad, such the French, Jan Bitoun, who chose to settle in Pernambuco. The UFB was host to the highly influential Milton Santos in the period before the military dictatorship, and, despite his exile from Brazil, his work was adopted by several disciples of different generations, such as Maria Auxiliadora da Silva, Rubens Toledo Júnior, and Pedro de Almeida Vasconcelos.

The Brasilia University is noteworthy for work in urban geography and urban and regional planning, as are the Federal University of Sergipe (UFS) for its tradition and progress in agrarian and rural geography and the Federal University of Rio Grande do Norte (UFRGN) for its recent advances in urban geography.

In the south, of the many universities in addition to those mentioned earlier, the Federal University of Santa Catarina (UFSC), Federal University of Paraná, and the Federal University of Rio Grande do Sul (UFRGS) are working toward strengthening their departments' work in human geography.

Urban studies, particularly the study of cities, has been one of the most privileged areas of Brazilian geography and is bound to the importance that cities, and the urban phenomenon have assumed during the occupation and shaping of the country throughout history. In fact, though it receives less attention than the Spanish colonization of America, the establishment of cities by the Portuguese over the sixteenth to eighteenth centuries proved fundamental to the economic, military, and political success of the appropriation of such a large territory. In postcolonial times, this instrument has continued to play a central role, turning Brazil from its earliest years into a highly urbanized country with a network of thriving, functional cities that lent a sense of cohesion to the land.

In 1994, Maurício de Almeida Abreu published an important essay on geographical study on the Brazilian city, in which he produces an exhaustive account of the works produced by Brazilian geographers in this area and correlates them with the trends and paradigms in urban geography over the last 50 years. According to Abreu, research in urban geography began in Brazil with the methodological essay by Pierre Monbeig that was pub lished in Portuguese on 'the geographical study of cities'. The essay 'would guide the thinking of countless Bra zilian geographers for more than a quarter of a century'.

This was followed by a large number of essays and monographs in the 1940s and 1950s on cities and other urban forms throughout Brazil into which the innovations promoted by the French school in Europe came to be introduced and in various ways influence Brazilian thought. This period is characterized by the number, range of focus, and open mindedness, and symbolized by the large body of work produced by Aroldo de Azevedo, from his first pioneering study in 1941 entitled 'Goiânia, uma cidade "criada"' (Goiânia, A 'created' city) to the concise four volume study on Brazilian cities (particu larly the study on São Paulo) he coordinated in 1958.

The French paradigm began to erode still in the 1950s and, for some authors, the Eighteenth IGU Congress held in Rio de Janeiro came to represent the start of this turn. However, this turn was still very influenced by the evolution French urban geography itself was undergoing. Indeed, geographers like J. Tricart, M. Rochefort, G. Chabot and, later, P. Claval, who were engaged in the renewal of French geography, had significant influence in Brazil. The renovation of urban studies undertaken by Brazilian geographers, such as P. P. Geiger and L. Bernardes, in the 1960s and 1970s represents one of French geography's greatest triumphs.

Influence from the Anglophone world started with visits from J. Friedman, B. Berry, and J. P. Cole and continued throughout the 1970s with visits from W. Bunge, M. Dacey, P. Haggett, R. Chorley, and D. Harvey.

Meanwhile, human geography accompanied the work being undertaken in other branches of geography on regional and urban planning processes, particularly dur ing the time of the military dictatorship which opened up a range of opportunities for geographers to involve themselves in land use planning.

The continuity of urban studies at the São Paulo University (USP) in the last half of the twentieth century are noteworthy, from the studies on regional urban sys tems (N. L. Müller) to the phenomenon of metropoli zation (M. A. Souza and A. F. Carlos).

In urban geography, among the countless high quality studies produced, it is important to mention the co herence and strength of A. Paviani's work, particularly that which was dedicated to the processes of growth and metropolization in the country's new capital, Brasília.

While there is continuity in the evolution of studies in urban geography, in line with the constant changes in the state of the art at the international level such as what led to the turn to strands of critical geography, something took place in the late 1970s that left a marked impact on Brazilian geography's orientations, particularly in the case of urban geography: the political openness of the regime in power, a prelude to the end of the dicta torship, allows Milton Santos to return from his long, yet scientifically rich and useful exile. Santos 'proposed' a 'new geography' that was tailored to the work that had meanwhile been undertaken by other Brazilian geographers.

The new geography and its quantitativist and theo retical approach were adopted, practiced, expanded, and diffused by the Instituto Brasileiro de Geografia e Esta tística (Brazilian Institute for Geography and Statistics) and São Paulo University's Geography Department at the Rio Claro Campus. While the Instituto Brasileiro de Geografia e Estatística tended toward practical appli cation (and, as some critics have emphasized, its success was mostly due to the 'services' that it provided the dictatorial regime in power from 1964 to 1978), São Paulo University's geography department at the Rio Claro Campus took a more academic direction with its pedagogical concerns and theoretical development, as can be seen in the academic journal founded in 1971 by A. Christofoletti and O. Ceron, *Boletim de Geografia Teor ética*, still in circulation today under the present, simpler title *Geografia* that was adopted in 1975.

It is the journal of the Associação de Geografia Teorética (Theoretical Geography Association) and is published thrice in a year in Portuguese with abstracts in English and French.

Its Role in the International Arena

Lusophone geography today has less impact on the international stage than might be expected, in spite of the relevance geography studies and teaching have in the Lusophone countries, particularly as regards Brazil and Portugal. This is due, to some extent, to linguistic issues. Unlike other countries, Portuguese and Brazilian geog raphers overwhelmingly chose to publish in Portuguese, which is particularly understandable in the Brazilian case since most of the geographical work produced has been geared toward the Brazilian geography community.

Yet, there is another factor that may have contributed to this predicament: the overwhelming role the French School played at the start and in the evolution of modern human geography in Portugal and Brazil which until very recently made French the lingua franca used by these countries in the production of geographical knowledge for international consumption. Even today, there remain very strong connections between the geography communities of Lusophone and Francophone countries.

However, if we consider the role that Lusophone geography has played in the scope of the International Geographical Union (IGU), in geography overall and in human geography in particular, it becomes clear that the two countries maintained a very active presence throughout the last half of the twentieth century.

Three vice presidents of the IGU have been Brazilian (H. O' Reilly Sternberg, S. Faissol, and B. K. Becker), and one has been Portuguese (O. Ribeiro). Ribeiro organized the first IGU Congress following World War II in 1949, which marked the IGU's growing international support and recognition. Furthermore, there has been an increased presence of Brazilian and Portuguese geographers in the activities of working groups and specialized commissions that comprise the IGU's scientific research networks. In the last two 4 year periods Portuguese and Brazilians have been on the steering committees of a large number of commissions.

Contemporary Geography in Lusophone Spaces

The Research State-of-the-Art in Human Geography

Research in human geography in Lusophone spaces at the start of the twenty first century is principally characterized by great topical diversity and a multiplicity of theoretical approaches.

While international recognition has increased, particularly in relation to the work produced by Portuguese and Brazilian geographers, the number of publications in international academic journals remains relatively limited. Therefore, it is from the contents of the main academic journals in Brazil and Portugal that we are best able to assess the state of the art Lusophone human geography. The other Lusophone countries that are engaged to some extent in human geography, Cape Verde and Mozambique, do not publish geography periodicals. However, Portugal and Brazil in particular publish a large number of academic journals in geography and are generally associated with university research centers. Despite the clearly uneven quality of the content, there is a large body of work covering a range of topics and theoretical and methodological approaches.

Economic geography and urban geography, with their multiple strands and scales of analysis, continue to be the privileged domains of human geography research. However, social and cultural issues have aroused growing interest, encouraging work on migratory movements, and their social, economic, and cultural consequences, the social production of space and landscapes. The interest in landscape stems from development in Anglophone humanist geography, though it sometimes makes interesting intersections with the work being developed in the French school, leading to new, particularly enriching perspectives when crossed with other disciplines, such as landscape architecture, sociology, tourism studies, urban planning, and local development. Agrarian and rural studies, which were important in Portugal and even more so in Brazil, have become particularly relevant in relation to work on the landscape.

Globalization and its spatial reflections at various scales are recurring themes in all Lusophone countries that do research in geography, particularly in Brazil and Portugal, and also in Cape Verde and Mozambique.

Gender and multiculturalism have also acquired greater attention, with gender being explored particularly in Brazil and multiculturalism being a hot topic in Portugal.

Tourism geography has acquired prominence in Portugal and is growing increasingly, more importantly in Brazil, and, to a lesser extent, in Cape Verde.

Sustainable development, with its various approaches concerning multiple geographical scales and risk assessment (both natural and technological) are also attracting attention.

One should note that Lusophone geographies, particularly the Brazilian and Portuguese strands, possess different matrices of influence (above all, Francophone and Anglophone) and broad international exposure and openness that facilitate a wide coverage of topics and theoretical orientations as well as being able to update itself continually, though at times its depth leaves something to be desired.

Furthermore, many departments and institutes engaged in teaching and research have kept the main branches of geography – physical and human – together, which today in some cases has led to their further integration in the scope of environmental studies and various aspects of applied geography, concerned as they are with the interplay between physical and human factors.

The 14th National Meeting of Brazilian geographers provided a fine context for observing the current concerns of a body of professionals no longer limited to academics and this group's international engagement, which privileges a pragmatic approach. This pragmatism is evident in the four transversal axes: (1) training the geographer, (2) teaching and research, (3) theory and practice, and (4) policy and development; as well as in the four topical axes: (1) frontiers, centralities, and peripheries; (2) schools of life versus schools of production; (3) technology and methods in contemporary geographical thought; and (4) regional integration, national projects, and resistance.

Examining the last two issues of two academic journals in geography representing different ages, traditions, and institutional frameworks: *Revista Brasileira de Geografia* (*Brazilian Journal of Geography*), which is a classic Brazilian geography journal that was 'renovated' in 2005, and *Boletim Goiano de Geografia* (*Goian Geography Bulletin*), which was founded in 1981, there are some similarities in terms of content – great diversity and the link maintained be tween human and physical geographies. The scope of the *Revista Brasileira de Geografia*, which is the mouthpiece of the Instituto Brasileiro de Geografia e Estatística (Bra zilian Institute for Geography and Statistics), generally surpasses geography, particularly human geography, yet it has remained the most continuous and solid reference in Brazilian geography. One issue contains articles on a wide variety of issues, from 'prostitution territories in public spaces located in the center of Rio de Janeiro' to mineral exploration, pollution, regionalism, cartography, and agriculture.

Portuguese academic journals are also geared toward integrating physical and human geographies, though generally more attention is given to issues related to human geography or the applied integration of physical and human geographical knowledges. Some of these journals often publish thematic issues which lends greater coherence to the overall content.

Geography and Power

Due to its intrinsic nature, geography constitutes a set of knowledges and practices important to the exercise of power, which is necessarily oriented toward space and territory. This was the case both in the colonial past and postcolonial present for Portugal, Brazil, and the other Lusophone countries.

Another issue is how professionals put this knowledge into practice, making it politically relevant. The role of geographers (professionals with university training ac quired in geographical institutes), particularly in Portu gal and Brazil, has gained importance through the practice of regional and urban planning and strategic planning, the processes of assessment of public policy, participation in land use management, and in the im plementation of territorial policy.

These activities are relevant at different scales and in concordance with their countries' own political and geo graphical particularities. For all of them, however, the local level (development, planning, and management) has been a privileged scale. In Portugal, due to the municipalist manner in which the state has been organized, there is an intermediate level of land use management that creates excellent opportunities to apply geographical knowledges in general and human geographical knowledge in par ticular. Brazil, on the other hand, is among the clearest cases throughout the world in which geography and pol itical power have worked together, a relationship that was cultivated for several years by the Instituto Brasileiro de Geografia e Estatística (Brazilian Institute for Geography and Statistics), a body that survived, more or less un affectedly, the periods of dictatorship and democracy. Perhaps the Instituto Brasileiro de Geografia e Estatística (IBGE), by contributing toward the production and dis semination of information on Brazil, was instrumental for both those in power and those of the opposition – as the latter never refrained from participating within the IBGE itself. Furthermore, the IBGE never lost its important role, due to its linkages to universities, its support of research and its work toward creating professional opportunities for geographers.

Portugal's integration in the EU in 1986 marked a moment in which it gained access to opportunities for geographers and geography in general to make an impact. Thus, though the end of the Portuguese dictator ship in 1974 should have led to a decline in the dis cipline's relevance and opportunities available to it due in part to the independence of its colonies, the very opposite occurred. The country was rebuilt and greater power was given to the municipalities starting in 1974, which led to high demand for the work undertaken by geographers.

Today, they comprise a significant portion of muni cipal bodies engaged in land use management and intermunicipal associations. Furthermore, membership in the EU meant that EU policies on regional and urban issues had to be implemented, thus opening up a fertile field of action for geography in the scope of EU pro grams, such as the URBAN Community Initiatives, the European Spatial Planning Observation Network (ESPON), the European Steel Design Education Pro gram (ESDP), the Study Program on European Spatial Planning (SPESP), the Interregional Cooperation Pro gram (INTERREG), and a host of other research and training programs.

See also: Africa; Postcolonialism/Postcolonial Geographies; Tropical Geography.

Further Reading

Abreu, M. A. (1994). O Estudo Geográfico Da Cidade No Brasil: Evoluçao e Avaliaçao. *Revista Brasileira de Geografia* 56(1/4), 21 122.

Almeida, R. S. (2004). O Pensamento Geográfico do IBGE no Contexto do Planejamento Estatal Brasileiro. In Martins, R. A., Martins, L. A. C., Silva, C. C. & Ferreira, J. M. H. (eds.) *Filosofia e História da Ciência no Cone Sul: 3 Encontro*, pp 410 415. Campinas: AFIC.

Amaral, I. do (1968). *Luanda Estudo de Geografia Urbana*. Lisboa: Junta. de. Investigaçoes do Ultramar.

Andrade, M. C. (1973). *A terra e o Homem no Nordeste* (3rd edn.). Sao Paulo: Ed. Brasiliense.

Becker, B. K. (1986). Geography in Brazil in the 1980s: Background and recent advances. *Progress in Human Geography* 2, 157 183.

Becker, B. K. and Egler, C. A. G. (1992). *Geography of the World Economy: Brazil A New Regional Power in the World Economy.* New York: Cambridge University Press.

Berry, B. and Pyle, G. (1972). Grandes Regioes e Tipos de Agricultura no Brasil. *Revista Brasileira de Geografia* 32(4), 5 39.
Dawsey, C. B. (2002). Geographical studies of Brazil in the United States and Canada from 1945 until 2000. *Geografia* 27(2).
Ferreira, A. B., Gaspar, J. and Medeiros,C. A. (1986). Notas em torno do desenvolvimento da Geografia em Portugal. *Anales de Geografia de la Universidad Complutense*, no. (6th edn.). Universidad Complutense, pp 63 79.
Gaspar, J. (1985). Portuguese human geography: From origins to recent developments. *Progress in Human Geography* 9(3), 315 330.
Gaspar, J. (2004). Applied geography in western and southern Europe. In Bailly, A. & Gibson, L. J. (eds.) *Applied Geography*, pp 151 168. Dordrecht: Kluwer Academic Publishers.
Geiger, P. P. (1988). Industrializaçao e urbanizaçao no Brasil, conhecimento e actuaçao da Geografia. *Revista Brasileira de Geografia* 50, 59 84.
Géographie, Economie, Société, 5(2) April 2003, Elsevier.
Girao, A. (1941). *Geografia de Portugal*. Porto: Portucalense.
Medeiros, C. A. (Dir.) (2006). *Geografia de Portugal* (4 Vols.). Lisboa: Círculo de Leitores.
Monteiro, C. A. F. (1980). *A Geografia no Brasil (1934 1977): Avaliação e Perspectiva*. Sao Paulo: Instituto de Geografia/USP.
Moraes, A. C. R. (1991). Notas sobre identidade nacional e institucionalizaçao da geografia no Brasil. *Estudos Históricos*, Rio de Janeiro 4(8).
Ribeiro, O., Lautensach, H. and Daveau, S. (1991). *Geografia de Portugal*. Lisboa: J. Sá da Costa.
Salgueiro, T. B. (1992). *A cidade em Portugal: Uma Geografia Urbana*. Lisboa: Ediçoes Afrontamento.
Sidaway, J. D. and Power, M. (2005). 'The tears of Portugal': Empire, identity, 'race', and destiny in Portuguese geopolitical narratives. *Environment and Planning D: Society and Space* 23(4), 527 554.

Relevant Websites

http://www.ageteo.org.br
Associaçao de Geografia Teorética.
http://www.apgeo.pt
Associaçao Portuguesa de Geógrafos.
http://www.ceg.ul.pt
Centro de Estudos Geográficos da Universidade de Lisboa.
http://www.cibergeo.org
Cibergeografia: Ensino, Pesquisa e Tecnologia.
http://www.geografia.fflch.usp.br
Departamento de Geografia FFLCH USP.
http://www.geografia.fcsh.unl.pt
Dep. Geografia e Planeamento Regional da Univ. Nova de Lisboa.
http://www.ub.es
GeoCrítica, Universitat de Barcelon.
http://www.ibge.gov.br
Instituto Brasileiro de Geografia e Estatistica.
http://www.uc.pt
Instituto de Estudos Geográficos, Universidade de Coimbra.
http://www.igeo.pt
Instituto Geografico Portugues.
http://www.mundogeo.com.br
Mundogeo.
http://www.ppgg.igeo.ufrj.br
Programa de Pós Graduaçao em Geografia PPGGUFRJ Página Principal.
http://www.uan angola.org
Universidade Agostinho Neto.
http://www.cniunicv.cv
Universidade de Cabo Verde.
http://www.uem.mz
Universidade Eduardo Mondlane.

M

Mackinder, H. J.

G. Kearns, Virginia Polytechnic and State University, Blacksburg, VA, USA

Mackinder, Halford (1861–1947)

Halford Mackinder (1861–1947) was born into a lower middle class family as the son of a medical doctor in the English market town of Gainsborough, Lincolnshire. After a decent scientific education at a school for the sons of doctors, he went to Oxford University where he studied physical science and took courses in law that prepared him for subsequent studies to qualify as a barrister in London. Neither his heart nor his head were in law and he had barely begun to practice when he was seduced by, first, education and, then, politics. Oxford University made a modest investment in the education of the working classes through sending peripatetic exten sion lecturers around the country to give night classes. From 1885, Mackinder was one of these teaching, ini tially, natural science and economic history, as local tastes dictated. Yet, this range of interests could, be believed, be synthesized as the discipline of geography, a subject taught badly at school as the memorizing of lists of capes, bays, capitals, rivers, and mountains, and taught not at all in English higher education, although well established in Germany as a field of academic research. Mackinder began to give classes on what he called a 'new geography', which would be adequate to the needs of Imperial Brit ain. In 1887, he was drawn into the national campaign to establish geography as a university subject, when he was invited to speak to the Royal Geographical Society (RGS) on 'The scope and methods of geography'. In the paper, published in the *Proceedings of the Royal Geographical Society* in 1887, he described a discipline that would trace the effects of environmental variation upon patterns of human activity. With funding from the RGS, Mackinder was set up as Reader in geography at Oxford and he began to try to find an audience for geography among students reading for other degrees. In this he was most successful with historians. Not until 1899 did Oxford create a School of Geography, where students could graduate as geographers, before proselytizing within British education the gospel of Professor Mackinder. The educational reformers had lost control of the RGS, which was only persuaded to stump up funds for this initiative after Mackinder had gained his bona fides as an explorer by completing the first European ascent of Mount Kenya in the summer of 1899.

From his involvement with adult education, Mackinder became from 1892 the principal of a new University Ex tension College at Reading, which grew to become a university. From his involvement with politics, came duties as part time lecturer at the London School of Economics (LSE) (1894) and then as full time lecturer (1902) and finally as director (1903). From 1887 to 1903, he had three simultaneous educational careers, at Oxford, Reading, and the LSE. In 1903 he resigned from Reading, in 1905 from Oxford, and in 1908 from the directorship of the LSE. At the age of 47, he turned to politics, almost full time.

Mackinder's interest in politics was evident from his days at Oxford where he followed men much wealthier and of higher social status than himself as president of the Oxford Union, a debating society where the politicians of tomorrow cut their teeth in, sometimes whimsical but always fierce, debate. His election was an extraordinary testimony to the hopes that many held out for his political future. Unlike, earlier presidents, such as his predecessor George Curzon, Mackinder did not have the private income needed to pursue a career in politics. Education was a passion but it did not pay well and it left little time for campaigning, after lectures given in Oxford, Reading, and London, or summers spent writing the new textbooks, such as *Britain and the British Seas*, in 1902, for the young subject. Mackinder was an imperialist and he supported, first, a group known as the Liberal Imperialists before, in 1903, hitching his star to the strong Britain policies of Joseph Chamberlain's tariff reform campaign. Britain, he argued, should use its empire to sustain its role in the world and this required keeping competing nations out of British and imperial markets. Britain no longer had the competitive advantage that could bear free trade competition and, in any case, many of its rivals were industrializing fast behind their own protectionist tariffs. In 1908, he received a retainer to work politically for 4 years for the cause of the economic integration of the Empire. In 1910 he was elected as a Liberal Unionist (Conservative) Member of Parliament for a Glasgow constituency. The imperial policy of building up the navy was very popular in this shipbuilding district.

Mackinder was now at the heart of imperial decision making. He did not cut the dash there that he had at the Oxford Union. His shyness seemed to have become worse after the failure of his marriage at the start of the century. His time lecturing seemed to have left him with little rhetoric and a passion for facts. His speeches were long, detailed, and boring. He found matters of imperial principle in subjects both great and, unfortunately, trivial. His final cause in Parliament concerned soda fountains and he argued that it was unfair that domestic producers of bottled soda paid duty whereas soda fountains of foreign origin dispensed limitless amounts of sweetened water without duty. This was hardly the sparkling career so many had anticipated. In the election of 1922, he was rejected by the voters in Glasgow, and his parliamentary career was over.

The Glasgow dockers who had been so keen on his shipbuilding rhetoric before World War I, were less keen on his anti Bolshevism after it. Yet, Mackinder's legacy rests in large part upon just this crusade and it was the only matter where he was given responsibility to act imperially in ways consonant with this writings. In 1917, Russia left World War I in the wake of its revolution. To the British, this was a betrayal of Russia's obligations to the anti German alliance and posed a dangerous example of revolution to British workers. The allies decided to reopen an eastern Front against Germany even though this meant fighting from within the territory of the nominally sovereign Russian state, now under Bolshevik rule. After the war, Curzon and Winston Churchill were among those who wanted to turn the allied troops directly upon the Bolshevik regime itself and establish in Russia a regime more friendly to the capitalist powers. There was opposition to this from soldiers waiting to be demobilized, from workers sympathetic to communism, and from members of the British cabinet worried about the cost of prolonging a wartime economy. At the moment when the anti Bolshevik forces appeared closest to an advance upon Moscow, Mackinder was selected by Curzon to go to South Russia as British High Commissioner in an attempt to knit together the various oppositional forces into a coherent group. He was asked to persuade them to attenuate their reactionary politics and accept some of the social reforms promised by socialists in order that they might be acceptable to workers and peasants in Russia, and their friends in Britain. Mackinder took 6 months to research this fully and went to Russia with plans for reorganizing the war torn economy of South Russia. By the time he arrived, he could only organize the evacuation of the families of counterrevolutionary officers for the game was up. Trotsky's Red Army was sweeping the White Russians out of the country. When he reported back to London, he could only resign and the great anti Bolshevik crusade was over, at least for the time being.

Mackinder continued to serve on many parliamentary commissions inquiring into economic issues, particularly relating to the Empire but his public role was minor. He continued to preside at geographical congresses and the award of the Patron's Medal of the RGS (1945) came near the end of a life devoted to the cause of the Empire in both education and politics. There is a fine biography of Mackinder written by Brian Blouet and published by Texas A&M University Press in 1987.

When Mackinder studied physical sciences at Oxford the Darwinian revolution was recent and exciting. Darwin's research methodology had been profoundly geographical. He examined the effect of the environment on natural selection and he examined species distributions for evidence of species shift. These ideas were applied to human society in ways that took natural selection away from its place in species selection and recast it as promoting class , race , and national selection. Social Darwinism justified (as competition) the role of force in human affairs, and justified (as natural selection) the primacy of the powerful in politics. Mackinder's view of the unity of Geography drew much on the environmentalism of Darwinism but his views on global political geography drew on the ideology of Social Darwinism. For Mackinder, geography comprehended a continuous chain of reasoning from the dynamics of Earth surface processes through to the unstable geographies of settlement, trade, and war. He emphasized the importance of a geographical vision – an environmentalist and strategic reading of human geographies.

His global perspective took up recent views of British authority as based on its sea power and questioned the permanence of this predominance. Mackinder argued that the interconnected world of the early twentieth century lacked the 'empty' spaces of the period of European colonization. Competition between European nations would no longer be fought out as a race to acquire ever more of the territory outside Europe as colonies. Instead, European nations would be brought face to face at every point on the globe. In this new era, the resources of the homeland would be the critical determinant of international competition. The balance was shifting from sea power to land power. If Britain, and what he called the Anglo Saxon race, were to prevail it would be in alliance with the land power of North America. Facing Britain would be either Germany or Russia, depending upon which of them managed to consolidate the resources of Eastern Europe and Western Russia. The land power of Europe derived, he argued, from an area, rich in minerals and soil fertility that was located beyond the reach of sea power. This was the inaccessible vast fastness of what he first termed the 'geographical pivot of history' and later in *Democratic Ideals and Reality* in 1919 called the 'heartland'. Here was the danger of Bolshevik Russia and Mackinder was one of the earliest to argue for the containment of the Soviet Union by denying it access to any warm water ports.

Unlike Isaiah Bowman for the United States or Paul Vidal de la Blache for France, Mackinder was not retained as a geographical expert to develop his country's geographical vision for the Europe that would be planned at the Paris Peace Conference. *Democratic Ideals and Reality* in 1919 is the advice that he would have given. He was suspicious of the 'idealism' of Woodrow Wilson's agenda for peace. The geographical realities, he believed, would produce new empires and a new global struggle. The geopolitical template could at best delay this by creating buffer states between Germany and Russia. His geopolitical strategy explained ways of containing the land power of Russia so that it could not serve as the basis for a sea power that might take communism globally. Mackinder's account of the land power basis for a global empire was taken up in Nazi Germany as a promise of success from the drive to occupy lands in Eastern Europe and western Russia. This avowal of his ideas by Karl Haushofer embarrassed Mackinder. It also embarrassed political geographers in the United States who tried unsuccessfully to establish differences in principle between Nazi geopolitics and Anglo American political geography. Yet Mackinder's ideas were soon dusted off and pressed into service as justification for the Cold War policy of Soviet containment. W. H. Parker published a good survey of Mackinder's geopolitical ideas as *Geography as an Aid to Statecraft* (Oxford: Clarendon Press, 1982). More recently, his ideas have served again to figure the global threat of Islamism and the need for global politics to be based on the realities of force and not the ideals of constitutions and treaties. Mackinder's geographical ideas sit ready to be used to justify a zero sum vision of world politics in which democracy can survive only by force of arms. This is global Social Darwinism.

See also: Darwinism (and Social Darwinism); Determinism/Environmental Determinism; Eurocentrism; Geopolitics; Kropotkin, P.; Lamarck(ian)ism; Nature, Historical Geographies of; Political Geography.

Further Reading

Blouet, B. W. (1976). Sir Halford Mackinder as British High Commissioner to south Russia, 1919 1920. *Geographical Journal* 142, 228 236.

Blouet, B. W. (1987). *Halford Mackinder: A Biography*. College Station, TX: Texas A&M University Press.

Gilbert, E. W. (1951). Seven lamps of geography: An appreciation of the teaching of Sir Halford J. Mackinder. *Geography* 36, 21 43.

Kearns, G. (1993). Fin de siècle geopolitics: Mackinder, Hobson, and theories of global closure. In Taylor, P. (ed.) *The Political Geography of the Twentieth Century*, pp 9 30. London: Belhaven Press.

Kearns, G. (1997). The imperial subject: Geography and travel in the work of Mary Kingsley and Halford Mackinder. *Transactions of the Institute of British Geographers* 22, 450 472.

Mackinder, H. J. (1887). The scope and methods of geography. *Proceedings of the Royal Geographical Society* 9, 141 160.

Mackinder, H. J. (1900). The great trade routes. *Journal of the Institute of Bankers* 21, 1 6. 137 155, 266 273.

Mackinder, H. J. (1902). *Britain and the British Seas*. London: Heinemann.

Mackinder, H. J. (1904). The geographical pivot of history. *Geographical Journal* 33, 421 437.

Mackinder, H. J. (1911). The teaching of geography from an imperial point of view, and the use which should be made of visual instruction. *Geographical Teacher* 6, 79 86.

Mackinder, H. J. (1919). *Democratic Ideals and Reality: A Study in the Politics of Reconstruction*. London: Constable.

Parker, W. H. (1982). *Mackinder: Geography as an Aid to Statecraft*. Oxford: Clarendon Press.

Ryan, J. R. (1994). Visualizing imperial geography: Halford Mackinder and the Colonial Office Visual Instruction Committee, 1902 1911. *Ecumene* 1, 157 176.

Toal, G. (1992). Putting Mackinder in his place: Material transformations and myth. *Political Geography* 11, 100 118.

Relevant Websites

http://onlinebooks.library.upenn.edu/webbin/book/lookupid?key=olbp22747
Democratic Ideals and Reality.

Malls/Retail Parks

G. A. Rice, University of Helsinki, Finland

Glossary

Anchor Store It is also known as a draw tenant that emerged in post-war America before subsequently being transplanted into Europe and the Far East. Initially, anchor stores formed one end of the 'dumbbell' shopping center design pioneered by Victor Gruen in the 1950s with the bar representing the smaller shops along the middle of the mall. In the UK, anchor tenants tend to be large food stores, such as Woolworths and Marks & Spencer, brought on board for financial stability and to attract smaller retailers into the remaining shop units. Due to the sheer amount of floor space, anchor stores pay the highest amount of rent while also setting the overall tone of the mall. Anchor stores in recently built malls are now more likely to include nonfood tenants. This is exemplified by the John Lewis Department Store in Glasgow's Buchanan Galleries and House of Fraser in Bluewater Kent, Southeast England. In North America, due to the sheer scale of malls, some developments are supported by a number of anchor stores; the West Edmonton Mall in Canada has eight anchor stores while the Mall of America in Minnesota has four.

Malls Often referred to as covered shopping centers in a complex or new buildings housing a variety of retail units with a number of anchor stores and entertainment facilities. This type of development is largely associated with the first prototype built in 1956 at Southdale outside Minneapolis by mall guru and architect Victor Gruen. Malls can be at in-town or out-of-town locations where they have become the cornerstones of suburban life. Malls at both locations can be further subdivided into a number of different types depending upon the size, scale, and extent of service function. Based on the experience of the US, 'regional malls' are defined as having at two department stores and a 100 shops and attracting customers as far as 20 miles away (smaller in the UK). 'Super regional malls' have at least five department stores and up to 300 shops attracting customers from up to 100 miles away. 'Mega malls' are even bigger in scale again in the West Edmonton Mall (WEM) and the Mall of America. In 2006, the 'mega mall' was dwarfed by the opening of a monolithic development, the Mall of Asia outside Manila in the Philippines.

Retail Parks These are also referred to as retail warehouse parks and are slightly different from malls in that they are broadly defined as a retail area (normally over 5000 m^2) containing a number of warehouse-style units selling brown and white goods. Brown goods include household entertainment appliances such as TVs or camcorders whereas white goods are more functional in washing machines and fridge freezers. Retail parks first began to appear in the early 1970s to take advantage of the fewer occupier restrictions operating at out-of-town locations. Retail parks are quite different from 'industrial estates' and 'business parks' which tend to rent out office suites or warehouse space as a node in a company's larger distribution network. Examples include Mallusk Industrial Estate outside Belfast and Hillington beside Glasgow airport.

Introduction

This entry gives a falvour of the work that human geographers and other scholars have conducted on malls and retail parks. In the former case the West Edmonton Mall (WEM) in Canada, the Mall of America in Minnesota, and the MetroCentre in the UK have been targeted to make theoretical links between urban geography, spectacle, and cultural form. This work has been supplemented by a parallel research agenda in the 'geographies of consumption' that tends to focus more on the 'everyday' banal material culture of specific places and their proximate neighborhoods. Through making links between consumption and identity, cultural and social geographers have studied malls in North London to suggest that through shopping, consumers are involved in a creative reworking of gender, ethnicity, class, and place. Despite being quite different beasts, retail parks have not been subject to the same level of microscopic scrutiny as malls.

The entry goes on to show that as available city centre space continues to deplete, human geographers have turned their attention to retail parks and out of town (regional) malls. The main thrust of this work is encapsulated by the new retail geography (NRG) and its emphasis on both the supply side and demand side of consumer landscapes. These recent consumer landscapes tend to be more polycentric and include malls such as Braehead outside Glasgow and retail parks such as Wolf Park at Temecula in Southwest California. If these off center sites have become destinations in their own right then the more general concern for critical urban geographers is the cultivation of 'downtown suburbia' whereby, out of town malls and retail parks mimetize and thus threaten the traditional city center retail offer. This enhances the competitve space economy fuelled by improved public fuelled by the rise of private car ownership that compounds the trappings of suburban living especially in areas where land is affordable.

Paradoxically, just as malls have fired the silver bullet into the heart of city centre retailing, they have also been erected as the golden ticket in the down town's fight back to regain its position at the top of the regional retail hierarchy. This so called consumption led urban re generation has become the default response to wide spread deindustrialisation, whereby former landscapes of production are mopped up and replaced by a new swathe of downtown malls and salubrious shopping districts. In the US, regeneration success has been attributed to specialty malls and their innovative designs. Sensitivity to user experience by architects like Jon Jerde is winning the 'sales per square foot' war between suburban and down town malls. However, this article concludes by suggesting that, despite the magnetism of recent city centre flagship regeneration projects, municipal authorities continue to struggle with *intra*regional competition between in town malls and out of malls and retail parks.

Origins and Evolution

Malls

When compared with retail parks, malls can claim a much longer history dating back to the shopping arcades of the early nineteenth century including Burlington in London, the Saint Hubert in Brussels, The Passage in St. Petersburg, and Galleria Vittorio Emanuele II in Milan. These arcades offered the first permanently cov ered shopping experience that would gratify the public who would be further captivated by the arrival of spec tacular world fairs and dazzling department stores in the middle of the nineteenth century. In both the US and Europe these 'cathedrals of consumption' presented the shopper's world as an escapist cocoon of salubrious luxury. Department stores were a consumer's paradise (especially aimed at middle class women) when compared with previous itinerant retailing and open markets that sold essential items such as vegetables, eggs, and poultry. The department store that is heralded as galvanizing more luxurious forms of consumption is the *Bon Marche* in Paris. This started life as a small store back in 1838 under the aegis of fabric salesman Aristide Boucicaut. By 1852, Boucicaut had established the *Bon Marche* as the pinnacle of Europe's consumer revolution. Here the idea of the department store was born and its success was replicated by the likes of Alexander Turney Stewart's Marble Palace in New York and the House of Fraser in Glasgow.

Although introduced some 100 years later, Victor Gruen's pioneering design for the world's first covered mall was inspired by the earlier department stores mainly because all commodities were sold from under one roof. Gruen's basic design had a large department store at one end with two parallel rows of shops and a pedestrian area in the middle. This development opended up in Southdale outside Minneapolis in 1956 and became the archetypal format for the postwar 'malling' of the US and Europe. This was also a period alsmost continuous de centralisation whereby large retail parks were developed often at out of town locations at the expense of proximate downtown areas. Since the 1990s, city centers have fought back in an effort to redirect retail capital and investment back into declining downtown areas. As an chor tenants in city center malls, department stores have been at the vanguard of various local strategies to bolster a more general 'back to the city movement'. The John Lewis Department Store in Glasgow's Buchanan Galleries' mall is an example of such a regeneration initiative that also points to the ways in which retail formats do not simply rise and fall over time but rather come to exist *in situ*.

Retail Parks

Unlike malls, retail parks tend to be predominantly lo cated on the outskirts of towns and city centers due to an abundance of (usually free) car parking. Although this car parking is shared as a communal facility, retail parks tend to be managed as an entire site or economic unit. They ar usually devoted to brown and white goods and Do It Yourself (D.I.Y) culture. D.I.Y retailers, such as B&Q and MFI, encourage people to embrace technoligies to fix problems in and around their houses without the help of professionals. Over and above this, retail parks typi cally accommodate electrical chain stores (Comet, PC World, and Electro World) and supermarkets (Tesco and Sainsbury) as major space units (MSUs). Examples of these retail parks include Broughton Park (Chester), Westcroft (Milton Keynes), and Strathkelvin (Glasgow) in the UK, Soroksár and Budakalász (Budapest) in Hungary, and Wolf Store (California) in the USA.

In the 1970s, retail parks really began to take advan tage of the opportunities associated with post war de centralization. Due to the sheer bulk of brown and white goods more floor space was required for merchandising, selling, and storage of overstock. Due to an ability to release large amounts of floor space, out of town lo cations became the key incubators for the growth of retail parks. While most total 'letting area' is devoted to malls, retail parks are predicted to increase over the next few years. In the UK, forecasts predict that retail parks will make up nearly one third of the total letting area com pared with almost two thirds for malls. More recently, retail parks and malls have teamed up to form super composite consumption sites on the edge of city centers. In conjunction with the rise of interstates in the US and motorway culture in the UK, the increasing popularity of the private car has made both out of town malls and retail parks more accessible. This continues to pose a threat to the traditional city center retail offer.

Malls as Spectacles and Cultural Forms

The West Edmonton Mall (WEM)

At the start of the 1990s a number of mall studies made links between urban geography and the incubation of spectacle as a means for profit. This incubation of spectacle was initially inspired by early Disney style theme parks but more recent research suggests that contemporary shopping malls are designed to reproduce similar experiences through cultivating a 'landscape of myth and elsewhereness'. The classic example is the WEM in Alberta, Canada. Using a number of specifically connected 'metonymical' and placial icons the desired effect is to create the impression that consumers of the mall are in a completely different destination. The idea is to manipulate real time and allow patrons to (re)create history in their own terms while glimpsing the continuity, harmony, and stability of humanity. 'Metonymical' icons include Versailles fountains (eighteenth century France) and a statue of Caesar (ancient Rome), with less emphasis on local Albertan heritage.

On a more sophisticated level, 'placial' icons attempt to simulate 'real' locations through near duplication that also support actual social engagement. In the WEM, the physical similarities between the fabricated World Water Park and 'real' Miami beach are such that mall patrons can swim, surf, and suntan as if they were on holiday. Other features, namely, the dolphin show and aquariums only have a slight resemblance to California Sea World but nonetheless they still emphasize the WEM's sense of elsewhereness.

The Mall of America

Another example that has been singled out to make links between urban geography and spectacle is the Mall of America in Bloomington, Minnesota. However, unlike its Canadian cousin, the focus here is on the 'dream house of the collectivity' whereby fantasies of authentic life are displaced onto commodities that are fetishized in the spatial, anthropological, and psychological senses. In the context of the Mall of America the assumption is that while shoppers get some pleasure from consumption, they often alternate between assertion and the denial of identity. The response from developers and designers has thus been to break the purported link between shopping and the mall itself. More specifically, the idea is to manufacture the illusion that more than shopping is going on. Over and above its 14 movie screens and key attractions that include the LEGO Imagination Centre and the Dinosaur Walk Museum, developers and designers have also attempted to re create a 'second nature' through cultivating the Mall of America as an acceptable 'third space' between work and the home.

Using any shopping mall to re create a 'second nature' or an acceptable 'third space' may seem like trying to catch lightening in a bottle. However, research shows that the Mall of America has had some relative success through exaggerating the differences between the 'harsh' world outside and enchanting interiors. This is directly linked to the design of the mall itself that masquerades as interconnected idealistic versions of different spaces and *civitas*. As a 'civic space' the Mall of America appears open to 'anyone' through the transplantation of an urban street. This tends to increase mall patronage as people tend to associate urban streets with the public realm, accessibility, and acceptability. The idealized public street in the Mall of America portrays a similar association as citizens can enjoy access, participatory entertainment, and opportunities for social mixing. However, research has shown that this idealized urban street is an example of a strongly bounded and purified space through managing out the conflicts associated with social and cultural differences. In reality, the idea is to support the spending habits of consumers while advocating policies that operate against street people, rowdy teenagers, and other undesirables.

The MetroCentre

In the UK, links have also been made between urban geography and spectacle but in different ways from the North American malls discussed above. Like the WEM in Alberta the MetroCentre opened up in Gateshead and expanded through a number of phases starting back in 1981 when a former coal 'staithe' (pier for loading coal) was designated an 'enterprise zone' (one of the first in the UK). At the end of the 1980s and early 1990s, a number of themed spaces were added, most notably the Roman Forum (now The Forum) and the Mediterranean Village (now The Studio). The most recent addition of the Metroland Theme Park connects the MetroCentre to its US and Canadian counterparts albeit on a much smaller scale. However, despite these similarities the MetroCentre case study differs from North American examples through placing more emphasis on malls as cultural forms.

An influential study of the MetroCentre suggests that the mall can be interpreted as encapsulating three levels of cultural form. First, the MetroCentre is marketed as a competitive tourist attraction so that individuals are carried beyond the immediate banalities associated with shopping. This does not just involve transforming the site itself but managing and directing the tourist quest through a multiplicity of meanings and possibilities of choice to reflect wider consumer culture. Second, the

MetroCentre deliberately uses highly idealized 'markers of neighborhood' whereby activities of community life can be acted out in the mall. These 'markers of neigh bourhood', including manicured spaces of living greenery peppered with benches, are an attempt to cultivate a 'second nature' not dissimilar to the WEM (see above). The third level of cultural form suggests that the MetroCentre is an entertainment environment that en courages consumers to become 'performers in their own drama'. This again parodies Disneyland and its rhetorical trick that excludes the purported terrors of urban life.

Malls, Place, and Identity: Brent Cross and Wood Green

The three mall case studies discussed above all make links between urban geography, spectacle, and cultural form but they have come under wider criticism for de voting insufficient attention to the actual 'practices of shopping'. A groundbreaking interdisciplinary study of two North London malls at Brent Cross and Wood Green addresses the tenor of this criticism through focusing on what consumers actually do when they visit the mall and what they understand as shopping. The study articulates the ways in which theories of identity are increasingly articulated in relation to particular spaces and places. This lends support to previous claims for a more pro gressive or 'global sense of place' by excavating the way in which contemporary urban geographies are shaped through porous networks of social relations, constructed through the specificity of their interaction with other places. Thus, the Brent Cross and Wood Green malls are used to making more general claims about the ways in which the 'practices of shopping' shape the identities of consumers that in turn take on particular urban geog raphies based upon the material culture of place.

Downtown Suburbia and Polycentric Urban Geography

Despite the differences in scale, size, and service function among the WEM, the Mall of America, the MetroCentre, and the two North London based malls, all share one common feature – an out of town location. This de centralization results in a more polycentric urban geog raphy in which malls reinforce the domestic value and physical order of suburban lifestyle. In the 1950s the ideal suburb consisted of a commuting father, housewife and two children who were stabilised through the purchase of houses, cars, furniture and electrical appliances. The origins of suburbia dates back to the eighteenth century so the addition of malls in the post 1945 era only in tensified the shift away from monocentric urban hier archies associated with central places and their associated shopping patterns. Before and during the two World Wars it was the city center that provided the most ex pansive range of (highest order) goods that the retail industry had to offer through department stores, and chain stores and the cooperative societies.

However, with the post war baby boom city centers became more crowded than ever which allegedly created racial tensions and put pressure on frontline public ser vices such as schools and the local welfare system. This intensifying pressure created the need for more space outside the city centers that eventually materialized as the 'new towns' up and down the UK and suburbia across the US. The resulting economic decline and subsequent 'human flight' helped create the demand for off center services in the form of malls and retail parks. The sub urban developments, such as Reston in Virginia, The Woodlands outside Houston, and Valencia north of Los Angeles, promised a better life to incoming residents. A consequence of this shift to suburban lifestyles was that long commutes for city center shopping purposes have been replaced by shorter shopping trips to proximate (out of town) malls and local retail parks. These shorter journeys are also due to the spatial switching of em ployment to new suburban offices, technopoles, and in dustrial parks. From Orange County in California to Grand Couronne around Paris new technologies in cluding microelectronics and light industries have gal vanized polycentric urban geographies that place jobs, town centers, and recreation closer to where suburbanites live.

Malls and retail parks have been crucial to the culti vation of suburban living. By repackaging and mimicking the city center services, malls and retail parks success fully provide a satisfactory alternative. This centrifugal process is buttressed by freeways and interstates (US) and motorways (UK and Europe) that serve exclusively as automobile connections between the suburban home and the mall (in some cases cutting the city center out of everyday life). Retail parks also benefited from this transport expansion through providing a wider choice of brown and white goods to the growing numbers of auto based consumers. Throughout the 1980s and 1990s, the mall's 'capturing of the city' inside its walls provided a controlled environment that was cleaner and safer (es pecially for children) than the real city center itself. In the US, the continuing growth of suburbia and its malls eventually evolved into 'edge cities' that include Ar lington outside Dallas/Fortworth and Broomfield to the northwest of Denver city center.

The Third California

The so called 'Third California' encapsulates the sub urban and polycentric trends noted above. As an example of explosive urban sprawl the area extends from the outer suburbs of Los Angeles to the foothills of the Sierra foothills in Northern California. Malls and retail parks are scattered throughout the region and help to support the growing population living and working in outlying metropolitan areas such as Bakersfield, Fresno, Sacra mento, and the so called Inland Empire. The East Hills Malls in Bakersfield, the Fashion Fair Mall in Fresno, and the Inland Empire's Riverside Plaza exemplify the type of comparative shopping experiences that pose a threat to the main urban center in Los Angeles. These malls and retail parks have had some success in attracting and re taining the skilled labour force integral to the Third California's evolution from agricultural past to an econ omy based upon construction, retailing and light industries.

New Retail Geography (NRG): Broadening the Research Agenda

As the proliferation of malls and retail parks in both North America and Europe evolved into more complex space economies a different approach was required to understand the wider context of retail change. One of the main responses was the emergence of subdiscipline in the mid 1990s that came to be known as the 'new retail geography'. Essentially, this is an attempt to fuse a political economy perspective with the cultural logic of retailing. This broad agenda has galvanized theoretical and empirical work ranging from the restructuring of retail capital, state regulation, national planning and its conflicting relationship with the cultures of consumption. The innovation associated with the NRG also lies in its ability to excavate the impacts of global supply side changes on urban and regional retail economies. Most of this work to date has largely been informed by UK based case studies associated with the so called 'store wars' and the 'food desert' debates of the 1990s. Despite this UK empirical focus, the NRG has successfully enhanced our more general understanding of how malls and retail parks relate to the planning literature and the various waves of decentralization that have been flanked by the rise of suburban services outside traditional city centers.

More specifically when compared with the intensive single mall case studies discussed above, new retail geographers tend to take a wider perspective to en capsulate the ways in which changes in the global retail operating environment work in and through patterns of consumption on the ground. This agenda defines out of town regional malls and retail parks as 'new urban forms' mainly because they have the potential to become 'boom burbs' and eventually evolve into US style 'edge cities'. More recently, however, new retail geographers have come more into line with the 'geographies of con sumption' research and its calls for 'on the ground' analysis. In exploring questions of identity a swathe of NRG literature has moved to challenge the simplicity of homogenizing global retail industry through demon strating the ways in which malls and retail parks are in tegral to constructing multiple identities.

Malls and Consumption-Led Urban Regeneration

As geographies of consumption research and new retail geographers continue to focus upon malls and retail parks, recent work has been informed by changes col lectively occurring as the so called 'back to the city' movement. Alongside the rise of suburban malls and retail parks many city centers began to fight back against the loss of trade to out of town developments. Throughout the 1990s, declining city centers have been aggressively promoting various regeneration programs in the form of city marketing, 'boosterism', and tax incentives to attract retail capital (and people) back into the city. Specialty malls, such as the Horton Plaza Center in San Diego, the Wedding Mall in Gurgaon, Princes Square in Glasgow, and Cobb Galleria in Atlanta, exemplify the nature of consumption led regeneration that now marks the Western city. Not only do these speciality malls shatter the traditional 'dumbbell' suburban mall design, they have also been used as leverage to attract the 'new middle classes' and other social groups that demanding social distinction through consumption.

As part of the efforts of city councils and municipal authorities to reinvent their downtown areas mass market malls have also been constructed to cater to nonspecialty shoppers. In Glasgow, this started in the 1980s with the St. Enoch Centre that eventually lost its domination to the Buchanan Galleries Shopping Centre. Covering 650 000 square feet with over 80 shops, the opening of second mall marked a significant addition to city center's retail offer. It is mass market malls such as Buchanan Galleries and the Cross Roads Plaza in Salt Lake City that are leading the crusade to put the downtown retail economy back on the regional map.

Outlook

The post war development of malls and retail parks are integral features of contemporary cities and their re gional hinterlands. While geographers have conducted more work on malls, there is now an increasing recog nition that other out of town consumption sites can also pose a threat to the city center economy. In conjunction with various types of regional malls retail parks are now just as likely to be accused of swinging the trade pen dulum away from traditional downtown areas. However, it is the captivating powers of mall design and the seductiveness of developments, including the WEM, the Mall of America, and the MetroCentre that have been most associated with sealing the fate of traditional downtown economies. Malls like these are characteristic of post war decentralization flanked by suburban life styles, fuelled by the convenience of private auto based consumption. However, in what may turn out to be a reversal of fortune, malls are now at the forefront of downtown renaissance. As city centers and municipal authorities fight back against 'retail flight', it is now near customary to see malls being used (to varying degrees of success) as part of wider urban regeneration agendas. Therefore, while out of town retail parks may remain a chimera for city centers, malls have the potential to sustain a more traditional urban hierarchy underpinned by a critical mass of retail capital in the city center.

See also: Consumption; Edge Cities; Retail Geographies; Suburbanization.

Further Reading

Chaney, D. (1990). Subtopia in Gateshead: The MetroCentre as a cultural form. *Theory, Culture and Society* 7, 49 68.

Crawford, M. (1992). The world in a shopping mall. In Sorkin, M. (ed.) *Variations on a Theme Park*. New York: Noonday Press.

Fishman, R. (2002). Bourgeois Utopias: Visions of suburbia. In Fainstein, S. & Campbell, S. (eds.) *Readings in Urban Theory* (2nd edn.). Boston, MA: Blackwell Publishing.

Frieden, B. J. and Saglyn, L. B. (1990). Downtown malls and the city agenda. *Society* 27, 42 49.

Goss, J. (1993). The 'magic of the mall': An analysis of form, function, and meaning in the contemporary retail built environment. *Annals of the Association of American Geographers* 83, 18 47.

Goss, J. (1999). Once upon a time in the commodity world: An unofficial guide to Mall of America. *Annals of the Association of American Geographers* 89, 45 75.

Goss, J. (2004). Geography of consumption I. *Progress in Human Geography* 28(3), 369 380.

Gottdiener, M. (1998). Consumption of space and spaces of consumption. *Architectural Design* 68, 12 15.

Gottdiener, M. (2003). Recapturing the centre: A semiotic analysis of shopping malls. In Cuthbert, A. R. (ed.) *Designing Cities: Critical Readings in Urban Design*. Boston, MA: Blackwell Publishing.

Guy, C. M. (1998). Off centre retailing in the UK: Prospects for the future and the implications for town centres. *Built Environment* 24(1), 16 31.

Hayden, D. (2003). *Building Suburbia: Green Fields and Urban Growth, 1820 2000*. New York: Pantheon Books.

Hopkins, J. S. P. (1990). West Edmonton Mall: Landscape of myths and elsewhereness. *Canadian Geographer* 34, 268 279.

Jackson, P. A. (1995). Guest editorial: Changing geographies of consumption. *Environment and Planning A* 27, 1875 1876.

Jackson, P. A. and Thrift, N. (1995). Geographies of consumption. In Miller, D. (ed.) *Acknowledging Consumption*. London: Routledge.

Key Note (February 2003). *Shopping Centres*. Key Note Market Assessments, UK.

Key Note (June 2006). *Shopping Centres*. Key Note Market Assessments, UK.

Kotkin, J. (2005). *The New Suburbanism: A Realists Guide to the American Future*. Costa Mesa, CA: The Planning Centre.

Kotkin, J. (2006). Hail to the suburban oasis. http://www.joelkotkin.com (accessed on 4 May 2008).

Kotkin, J. (2006). Presenting the new economic map of America. *The American* November/December (pp 31 36).

Kotkin, J. (2007). Suburban idyll. *Wall Street Journal* (April 19th).

Kotkin, J. (2007). The Myth of superstar cities. *Wall Street Journal* (February 13th).

Kotkin, J. and Frey, W. H. (2007) The Third California: The Golden State's New Frontier. The Brookings Institution Research Brief. The Brookings Institute, Washington DC.

Lowe, M. S. (2000). Britain's regional shopping centres: New urban forms? *Urban Studies* 37, 261 274.

Lowry, J. R. (1997). The life cycle of shopping centres. *Business Horizons* 40(1), 77 86.

Mansvelt, J. (2005). *Geographies of Consumption*. London: Sage.

Miles, S. and Miles, M. (2004). *Consuming Cities*. Basingstoke: Palgrave MacMillan.

Miller, D., Jackson, P. A., Thrift, N., Holbrook, B. and Rowlands, M. (1998). *Shopping Place and Identity*. London: Routledge.

Wrigley, N. and Lowe, M. (2002). *Reading Retail: A Geographical Perspective on Consumption Spaces*. London: Arnold.

Zukin, S. (2005). *Point of Purchase: How Shopping Changed American Culture*. London: Routledge.

Relevant Websites

http://www.brentcross.co.uk
Brent Cross Shopping Centre.

http://www.easternct.edu
Shopping Mall Studies at the Eastern Connecticut State University.

http://www.mallofamerica.com
The Mall of America.

http://www.metrocentre gateshead.co.uk
The MetroCentre Shopping Centre.

http://www.westedmall.com
The West Edmonton Mall.

http://www.themall.co.uk
Wood Green Shopping Centre.

Map Hacking

E. Mac Gillavry, Webmapper, Utrecht, The Netherlands

Glossary

Application Programming Interface (API) A source code interface that a computer system or program library provides, in order to support requests for services to be made of it, by a computer program.

Mashup Website or application that combines content from more than one source into an integrated experience.

Introduction

Over the last few decades, cartographers have been witness to the rapid democratization of mapping and the production of geographic information. For centuries, this has been the prerogative of official cartographic institutions with budgets large enough to fund the necessary mapping technologies and labor costs of highly trained specialists. At home, maps proved an instrument for nation building as nation states emerged. Maps were not only used in education to teach pupils about their nation, but even administrations used maps as a tool to govern their territory. Away from home, maps were an instrument for discovery, exploration, and colonization.

With the computer revolution of the 1970s, geographic information systems (GIS) moved from mainframe computers to desktop personal computers (PCs). From government and research, GIS spread into the business sector as a tool for operational planning and geographic analysis. Along with the uptake of the World Wide Web in the late 1990s, the Palo Alto Research Center (PARC) Xerox Map Server and Virtual Tourist websites served the first online maps to everyone's PC, followed by commercial mapping websites, such as MapQuest and Multimap.

These days, maps are more widely used than ever, thanks to the easy to use commercial mapping websites. Nevertheless, until recently individual citizens only rarely created maps, but rather used maps created by cartographers employed by national mapping agencies, academic institutions, and map publishers. GIS software remained expensive programs, restricted to the use of highly trained specialists.

Opening Up the Map

Recently, artists and programmers have taken advantage of the World Wide Web as a tool of mass distribution and pursue to bring mapping technologies, once used by highly trained specialists, to the people more directly. They make people aware of the power of creating maps by developing open source mapping tools and Web mapping applications to solve cartographic problems or to create alternative, or even subversive, cartographic representations. This practice is sometimes called 'map hacking'. Although these representations are based on mapping technologies, they did not spring from the official cartographic institutions or academic disciplines. They have been developed by artists and programmers, intrigued by the power of mapping technologies, to deliver meaningful information.

'Map hacking' is a term borrowed from online computer role playing games in which a large number of players interact with one another in a virtual world. The computer of the player always knows where the other players in the game are and what they are doing even though it does not normally display that information to the player. The practice of using an illegal program, that enables a player to modify the original game so that it displays the position and action of every unit on the entire map of the virtual world, is referred to as 'map hacking'. As other players are usually unaware of this, map hacking is regarded as a form of cheating. In this context, hacking is used in its original meaning: writing a software program to modify another program, providing access to functionality or features not intended, by the creators of that program, for the use of end users.

Among programmers, the term 'hacking' means to come up with a quick and dirty solution to a problem or a clever way to get something done or to develop a surprising and nontraditional use of technology. The term 'hacker' is taken very much as a compliment – referring to someone as being creative, having the technical ability to get things done. Although map hacking would suggest malicious tinkering with geographic databases, the current usage of map hacking derives from the latter meaning.

Drivers of Map Hacking

One of the main drivers of map hacking, as a practice to make people aware of the power of creating maps, was the lack of affordable and copyright free geographic data. Official cartographic institutions spend billions to create large scale base maps to fulfill their civil and military responsibilities. Nevertheless, geographic data created by US federal agencies is made available free of copyright and at the nominal cost of copying it.

In most other countries, however, geographic data has to be paid for and is subject to strict copyright laws. Since the market position of national mapping agencies is changing rapidly – their budgets being cut and status changed – their geographic databases have become their main capital assets. To survive in their new role, they have to protect their assets with copyright restrictions and inhibitory pricing structures. Map hacking has, therefore, become a practice to create copyright free geographic data and to bring together existing, freely available, sources of geographic data.

A successful example of this practice is Open StreetMap, a collaborative project to create a map that is free of charge and that anyone can manipulate and reuse. The project started in Great Britain, where geographic data from the national mapping agency Ordnance Survey is protected under Crown Copyright. Since anyone can contribute to OpenStreetMap, the coverage of the geo graphic data is growing and improving all over the world.

For map hackers to create geographic data, it is ne cessary to have the ability to measure the position of lo cations on the Earth's surface. The Global Positioning System (GPS) receivers used to be the domain of sur veyors and geodetic engineers, but nowadays, there are many cheap receivers, available for outdoor enthusiasts, that become smaller and smaller. As of 1 May 2000, Se lective Availability has been turned off so public users can locate themselves using GPS receivers with 'military pre cision'. Recently, there have been palmtops and even mo bile phones that have come with a built in GPS antenna. The ability to establish one's location is unprecedented.

Another driver of map hacking is the adoption of open source in the field of GIS. The IT industry has a long history of open source projects, from operating systems, web servers, and programming languages to applications. Only recently, has the Geographic Information (GI) in dustry started catching up. Open source databases, such as MySQL and PostgreSQL, now support spatial features. There are several libraries, for example, to convert be tween common spatial data formats, or to change pro jections that are included in various open source projects. Geographic Resources Analysis Support System (GRASS) has become a well known open source desktop GIS. For creating Web mapping applications, there are various open source projects, such as Deegree, GeoServer, and the UMN MapServer. Web clients such as Chameleon, ka Map, and OpenLayers allow developers to build compel ling interfaces to work with maps in one's web browser.

Most of the architectures for Web mapping appli cations are built around Open Geospatial Consortium (OGC) standards such as the Web Feature Server and Geography Markup Language specifications. With geo graphic data being transferred based upon open standards and clear interfaces, it is easy for map hackers to combine various geographic data sources and make interesting Web mapping applications, with no resources spent on buying software.

The power of creating maps is most apparent when mapping information that is not necessarily geographic in nature, but still relates to a position on the Earth's surface. The practice of adding a standard code to a piece of information in order to note its appropriate geographic location is called 'geo tagging'. First, Web log authors or 'bloggers' writing about similar topics started to form communities, but also listings of blogs, that were from the same country or even located in the same city, for example in Brighton, United Kingdom (http://www.brightonbloggers.com/), started to appear. Mapping the virtual space of the 'Blogosphere' onto the physical one, these listings put a sense of place back into cyber space. In a final step, the positional accuracy increased from country and town to the level of geographic co ordinates, while the geographic extent increased to cover the whole world with GeoURL. Bloggers are encouraged to manually add 'geotags' to their web pages to inform visitors about their location.

Some bloggers travel regularly, and blogs can be up dated using mobile devices – so called 'moblogging'. Location thus becomes another means to structure blog entries. Also, bloggers frequently write about locations they are visiting. Authors then geotagged not only their blogs, but also individual postings and pictures about the locations they visited. Nowadays, cameras can be enabled with GPS, so that digital photographs can automatically be geotagged. Geotagging has provided map hackers with a wealth of nongeographic information to plot on their base maps.

Google Maps Hacking

In February 2005, the Google Maps website was publicly launched. Like other commercial mapping websites, such as MapQuest, Multimap, and ViaMichelin, Google Maps offered mapping, directions, and local search. However, the user experience was unlike anything that came before it. Furthermore, it was bound to draw the attention of map hackers.

Three things set Google Maps apart from other commercial mapping websites. First, the mapping web site has a clean and responsive user interface, just like Google's search engine. Users enter keywords into a single input field and, within seconds, a set of matching results is presented on a map. Second, Google Maps does away with the refresh between every map view using fast loading, pre generated map images, so called 'tiles', that appear in view by simply dragging the map or zooming in. Since the page does not have to be reloaded, the mapping website allows for fast, nearly seamless, user interaction with the map. Finally, but most significant, is the programming underneath Google's mapping website. Since Google Maps makes use of JavaScipt that is in terpreted by the user's web browser, instead of running on a web server, map hackers gained easy access to the code and were able to figure out how to manipulate the program for their own uses.

Within days of the launch of Google Maps, developers had published detailed descriptions online of how to reverse engineer Google's mapping technology and combine it with your own geographic data or with geo graphic data from a third party on your own website, seemingly independent of either Google or the third party. These home grown, Web mapping applications were referred to as 'hacks', since Google had not sanc tioned the use of its code on external websites.

Early examples of Google Maps hacks include Paul Rademacher's Housingmaps.com and Adrian Holovaty's ChicagoCrime.org. In search of an apartment using the popular website Craigslist.org, a well known online classifieds directory, software engineer Paul Rademacher wanted to browse the listings geographically. He wrote a hack to create a new website that combined the classi fieds from Craigslist with the base maps from Google Maps for users to automatically visualize the results of their search on a map. Whereas Rademacher created a new Web mapping application, Andrian Holovaty simply improved on an existing Web mapping application built using traditional Internet GIS technology. To create the website ChicagoCrime.org, he hacked the online data base of crime incidents that is maintained by the Chicago Police Department and plots them on the base maps from Google Maps, to benefit from its almost seamless user interaction with the map.

Soon after these mapping hacks appeared all over the Web, Google sent out cease and desist notices because these hacks violated its terms of service. In their view, these hacks encouraged mass downloading of copy righted images and data, something the licenses with their map data providers did not cover. Consequently, many exciting and creative Web mapping applications were taken off line.

From Hacking to Mainstream: Mashups

The need for copyright free geographic data, the pro liferation of GPS enabled devices, the growing number of open source GIS projects, and the increasing amount of geo tagged information available online have facili tated many mapping hacks. With the publication of the book *Mapping Hacks*, computer books publisher O'Reilly aspired to bring these practices together. The book de scribes many open source mapping tools and Web mapping applications to solve cartographic problems and invites user to explore, create, to unearth shortcuts, and come up with fun and useful things. Most, but not all, involve the Web, and some require no programming by the user.

Map hacking has had a tremendous impact on the established GI industry. Particularly, ever since the first Google Maps hacks appeared, each layer of the multi tiered geospatial value chain tried to grasp these devel opments and had to react and readjust its business strategy. Traditionally, street level mapping, aerial pho tography, and satellite imagery are provided by official cartographic institutions, such as national mapping agencies, but also a few commercial suppliers like Automotive Navigation Data (AND), NavTeq, and Tele Atlas now play an important role in the geospatial value chain. The base data are processed and rendered using software from typical GIS vendors, such as Environ mental Systems Research Institute (ESRI) or MapInfo, or from companies like deCarta or MapSolute. On top of the mapping and imagery come the commercial mapping websites that have the applications and servers to deliver these to end users or business customers.

These tiers of the geospatial value chain are not fixed and there are interesting transitions between the base data suppliers, software vendors, mapping websites, and business customers. For example, AND provide hosted solutions for delivering mapping and travel directions straight to business customers. Michelin used to maintain a road network database mainly for its own paper road atlases, but realized it could benefit from this asset even more through its commercial mapping website Via Michelin, and furthermore, provide hosted solutions for its business customers.

Prior to 2005, visitors of commercial mapping web sites, such as MapQuest and Multimap, were able to look up travel directions and search for locations and busi nesses locations. While these searches were, typically, limited usually to the companies the websites had part nered with, the Google Maps hacks showed the potential of being able to combine these maps with any other online geographic data source instead of a Google business partner. Not surprisingly, then, Google Maps was the first mapping website to react. Sending out cease and desist letters first, Google soon not only recognized that they could gain exposure from these hacks, but also realized the potential of all these map hackers. For both Google and these developers to benefit, the Google Maps JavaScript based API was launched at the Where 2.0 conference in San Francisco in June 2005. Since the Google Maps API is free and comes with very lenient terms of service, it became even easier for map hackers to plot geographic data onto Google Maps.

Rather than mapping 'hacks', these new hybrid Web mapping applications that were created by combining third party geographic data with the base maps from Google Maps are referred to as mapping 'mashups'. In music, when you create a new song by taking the melody from one song and the lyrics from another, it is called a mashup. The term crossed over and is now used in the realm of the Web.

By opening its mapping website and facilitating map hackers to create their own mashups, Google has changed the rules of the online mapping industry. From then on, the rest of the commercial mapping websites were put on the defensive. They have been struggling to catch up with Google Maps and have only recently started to open up their maps to their business clients through JavaScript based mapping APIs. Nowadays, commercial mapping websites like MapQuest and Multimap all have set up developer networks to attract map hackers and let them create brand exposure from implementing their proprietary mapping APIs. Multimap has even gone so far as to integrate map hacking initiatives, such as Geonames and OpenStreetMap, into its Open API. Geonames is an online gazetteer, containing over 8 million geographical names, that is accessible free of charge through a number of Web services.

Not only commercial mapping websites have opened up to let developers create their own mapping mashups, but also GIS software vendors have embraced map hacking. For example, the ESRI ArcWeb Public Services program and deCarta's Developer Zone service allow developers to freely create Web mapping applications. The APIs that come with these services expose only a subset of their commercial APIs and come with more restrictive terms of service. Autodesk has taken a different approach to benefit from map hacking. By releasing the code for MapGuide, the company's online mapping technology, as open source, Autodesk aimed to benefit from the map hacking community to have faster innovation, more frequent releases, and lower developing costs.

Finally, national mapping agencies and commercial map data suppliers are opening up their base data to map hackers. Ordnance Survey, the British national mapping agency well known for its strict Crown Copyright on maps, has launched an alpha version of Open Space, another JavaScript based mapping API that allows developers to plot markers, lines, and polygons on top of its raster maps down to 1:25 000 and to search the gazetteer derived from its 1:50 000 map series. In 2007, the commercial map data supplier AND donated its base data of the Netherlands, China, and India to the OpenStreetMap Foundation. Developers can not only contribute to update the base map data by reporting map errors, but also are now free to use these rich base maps in their own mapping mashups.

Critiques

Map hacking is only effective in raising the awareness of the power of creating maps and facilitating the creation of alternative cartographic representations, when people have access to the technology (a powerful enough PC and Internet connection) and the knowledge to use it. There is a digital divide, because it takes awhile before new technology spreads from one social group (the map hackers) and is adopted by the wider public. Although the availability of mapping APIs has made it easier for people to create Web mapping applications, it still requires a significant amount of programming to actually create a mapping mashup that goes beyond simply plotting red pins on a base map. Nevertheless, there are more and more websites that provide utilities and small programs to help you plot points or GPS tracks or geocode addresses, further bringing down the barriers to starting creation of your own mapping mashups.

Not only does the digital divide put up impediments, but also the very mapping APIs themselves limit people in creating mapping mashups. These mapping APIs are not only based on proprietary and closed code, they also still depend upon the commercial provision of base map data and imagery. Despite the lenient terms of service of the various APIs, commercial map data suppliers constrain the usage and applications. For example, developers are still precluded from using these mapping APIs together with GPS receivers. Also, these commercial mapping APIs impose a uniform view of the world, since the creativity of developers is restricted by the functionality exposed in the closed code APIs. This restriction is referred to as 'red dot fever', because many mapping mashups typically allow developers to plot red dots on top of a limited range of pre rendered map styles. Finally, many mapping mashups are held at ransom because some of the commercial mapping websites reserve the right to include commercial advertising in, or associated with, the base maps and imagery provided through their mapping APIs at any future time.

Although mapping APIs facilitate the creation of a mapping mashups, it does also raise issues about the privacy, integrity, and accuracy of the information. Mapping mashups typically combine any third party data source with base mapping through a mapping API. Neither do the map hackers own the data or the map data, nor do the owners know in what way their data or base maps are being mashed up. Privacy is not only at stake when you need to supply credentials of your account with a photo hosting website to plot your holiday pictures on a map, but also when your house appears on the highly detailed aerial imagery of the mapping website. Furthermore, map hackers can easily spoof the data that comes from one source, thus jeopardizing the integrity of the data. Finally, whereas official cartographic institutions have built up a reputation for quality and accuracy over many centuries, companies that now underpin many mapping APIs lack this reputation. In particular, the aerial imagery supplied through their APIs is often misaligned, resulting in poor positional accuracy. Nevertheless, end users do not challenge the authority of these companies, perhaps because of their success in other areas, for example, their search engine.

Future

Map hacking as a practice to make people aware of the power of creating maps by developing open source mapping tools and Web mapping applications that solve cartographic problems or that facilitate people to create alternative, or even subversive, cartographic representations is certainly here to stay. Search engines, commercial mapping websites, GIS vendors, and base data providers have all contributed to pull map hacking away from the negative connotations and have facilitated mapping mashups by opening up their mapping APIs under very lenient terms of service. This has further raised the general interest in creating your own maps and putting a sense of place back into cyberspace.

Nevertheless, there remain issues around proprietary data, advertising, and closed code in these mapping APIs that limit the application of mapping mashups in some realms. Furthermore, privacy issues, data integrity, and accuracy are not always taken into account either. Map hacking will, therefore, still have a role to play, not only in the online mapping arena, but also in mobile mapping when telecom operators and mobile phone manufactures come to realize the potential map hackers will bring to mobile mapping, pedestrian navigation, and mobile local search.

See also: Critical Cartography; Internet/Web Mapping; Mapping, Commercial.

Further Reading

Erle, S., Gibson, R. and Walsh, J. (2005). *Mapping Hacks: Tips and Tools for Electronic Cartography*. Sebastopol, CA: O'Reilly Media.

Gibson, R. and Erle, S. (2006). *Google Maps Hacks. Tips and Tools for Geographic Searching and Remixing*. Sebastopol, CA: O'Reilly Media Inc.

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal* 69, 211 221.

Plewe, B. (1997). *GIS Online: Information Retrieval, Mapping, and the Internet*. Santa Fe, NM: Onword Press.

Relevant Websites

http://www.freeourdata.org.uk
Free Our Data: Make taxpayers' data available to them.

http://lists.burri.to/mailman/listinfo/geowanking
Geowankers mailing list.

http://code.google.com
Google Maps API.

http://www.mappinghacks.com
Mapping Hacks.

http://www.opengeodata.org
OpenGeoData.

http://www.osgeo.org
Open Source Geospatial Foundation.

http://conferences.oreillynet.com
O'Reilly Where 2.0 Conference.

http://www.planetgs.com
Planet Geospatial.

Map Interactivity

R. Edsall, University of Minnesota, Minneapolis, MN, USA

Glossary

Cartographic Model The general strategy for visual representation of a variable or dataset. Types of thematic maps (choropleth, isarithmic, etc.) are based on cartographic models.

Direct Manipulation From human–computer interaction, interface design enabling interaction with visual representations without use of menus, buttons, sliders, or other interface tools.

Drill Down From information visualization, move from summarized, aggregated, or classed information to more specific information.

Exploratory Data Analysis (EDA) From statistics, strategies for comprehending and analyzing large multivariate datasets without *a priori* assumptions or hypotheses. EDA research has resulted in innovative graphics and innovative ways of interacting with the graphics, such as brushing, linking, focusing, conditioning, and lensing (described herein).

Oblique View Also called 'bird's-eye view', a distorted but familiar perspective used in virtual environments and in-car GPS displays that simulates a forward-looking viewing angle from a short distance above the surface. Contrasts with plan-view, a downward-looking viewing angle from directly overhead (and adjusted to correct distortion) common on traditional maps.

Particularization Increasing the level of detail in cartographic representation via interaction. The opposite of generalization, particularization enables use of detailed information on demand.

Introduction

The addition of dynamic content in maps on computer screens or other types of technologies represented a major change in the design, use, and potential of geographic representations. Dynamic maps can be either animated or interactive (or both). Interactivity in maps is the more recent development; the advent of interactive maps resulted in a broadening of the capabilities of geographic representation, and placed many of the choices in symbolization, classification, generalization, and abstraction in the hands of the map user rather than the cartographer. In many ways, interactivity in maps allows the individual user to take on the role of the cartographer, though it cannot be said that user interaction has made the cartographer dispensable. Quite the contrary, crucial cartographic choices still need to be made and guidelines need to be followed in designing not only the graphic elements but also the interface tools for the interactive maps. This article examines the ways in which geographic representation is advanced through interactivity, and catalog the myriad interactivity forms in maps and how they support geographic work and communication. Also, applications with important or innovative interactive capabilities are briefly described. These applications, of course, are superseded rapidly with new versions or new concepts, however, the general inventory of reasons, functions, and modes of interactivity will continue to be applicable in the coming decades.

Interactive Map Functionality

Of course, humans have always 'interacted' with the maps they use, regardless of the map's medium. A child spins a globe, a military strategist draws lines, and a pedestrian rotates a map to align it with the direction of movement, a traveler annotates a map with personal notes about a visited place, and a meteorologist flips through several forecast charts to observe air movements. Interaction of any kind is the basis for any sort of discourse. Interactivity in computerized allows for a sort of dialog between a human and a machine, or, in a less direct but possibly more important way, between a researcher and data. In addition, a nontrivial advantage of interactivity on maps is the affective influence on the user: as maps respond to the actions of the user, they become more like the objects that are encountered in everyday life. This is (at least at present) still thrilling and novel. Interactive applications such as those described below are appealing partially because of the empowerment that the interactive techniques afford the user.

Maps incorporating interactive elements allow for the circumventing of many important limitations of static maps, and also allow for important connections to be made between different variables and between different representation forms.

It can be said that cartographers may not consider limitations of static maps as 'shortcomings', as map designers often exploit the fact that maps need to be abstracted to some extent in order to advance an argument, maintain readability, or preserve esthetic consistency. Besides, even with the addition of interactivity to begin to overcome these limitations, the affordances provided by interactivity remain choices of the designer of the representation (the map and the interface).

The following section briefly outlines some ways that, by adding interactive elements, a map's power and usefulness is increased.

Contextualize

A static map is necessarily limited by its geographic scope and scale. The extent of the Earth that is displayed on a static map is a choice of the cartographer and not the user, but it may be necessary for a user to understand the geographic context of the region. Interactive maps allow a user to browse around a region and to examine a phenomenon at various scales. Some interactive methods enabling contextualization include: (1) panning, (2) zooming out, and (3) flying through.

Panning by inspecting around a region at a constant scale. This is analogous to an astronomer who scans the skies by moving a telescope, or a biologist who scans a specimen by moving a microscope slide, or a geographer who scans a scene by moving his or her eyes on a page or a landscape.

Zooming out from a region. This allows the user to understand the context of the scene and reveal features and trends invisible at larger scales ('see the forest for the trees').

Flying through a virtual scene. Users of immersive environments can browse around a region, orienting to their surroundings in a fashion that may be more intuitive than that provided by traditional plan view maps. Though oblique (or ground level) views typical of virtual reality are not useful for contextualizing thematic information because of the necessary severe perspective distortion, their relative familiarity may enable more effective contextualization of location and reference information for some map users.

Panning and zooming are perhaps the most basic of interaction methods for maps and replicate the browsing and inspecting process with which most map users are familiar. They also reinforce the notion that the map is a representation of physical space onto a smaller map space. Though most typical, this space to space transformation is not the only option: some maps and graphics represent time (as an animation) or attributes (as a scatter plot). One can also pan and zoom based on these mappings, showing context of data before and after a time step, showing trends at coarser temporal resolutions, or showing bivariate or multivariate relationships across a broader range of attribute values.

Particularize

One of the necessary and useful steps in the cartographic process is generalization, where, for example, a city is represented as a dot, or a sinuous border is represented as a gentle curve. Interactivity allows for the level of generalization to be adjusted to suit the needs of the user. Common methods enabling 'particularization' (the opposite of generalization) include: (1) zooming in, (2) toggling on, and (3) adjusting intricacy.

Zooming in to a region with an automatic (and cartographer defined) change in the level of detail. A scale increase not accompanied by a particularization would be worthless – the black dot representing a city would simply become a bigger black dot. The level of detail change, made popular by the film *Powers of Ten*, is one of the biggest draws of virtual globes such as Google Earth, which allow for seamless contextualization and particularization, as users zoom from the global scale ('that's North America') to the human scale ('that's my backyard') and back.

Toggling on a detailed layer (e.g., secondary roads in addition to primary roads), which holds the scale constant but increases the level of detail. This method may seem desirable but it carries a price of cognitive and display overload – adding every street name of a major city on a small scale map of a country results in a useless tangle of words and lines. Including information on a map simply because it is available is one of the most common errors inexperienced GIS practitioners make.

Adjusting intricacy of marks on the map. The user may require more detail of a specific theme (e.g., rivers and streams) and it is conceivable that generalization techniques that would typically be implemented at the scale of the map would be reversed, revealing detail in shapes, linework, or aggregation on the fly.

Revealing detail from summarized or generalized representations is discussed in the context of data exploration below, but the concept is useful for a broader range of mapping and visualization tasks.

Resymbolize

Symbolization in cartography is a very broad ranging subject, and in many ways every mark on a map is, to some extent, a symbol (the map page itself 'symbolizes' reality). Thus, any manipulation to the look or content on the map could be considered a resymbolization, but for the purposes of this article, 'symbols' refer to the abstract representations of the thematic content of the map, or, more simply, those marks on the page that might be explained in a legend. While cartographic guidelines for static maps have been formulated over the last century to maximize understanding for communication across the broadest possible range of users, guidelines for interactive maps are formulated to maximize possible symbolic choices for individual users, who might be using the map to uncover unknown patterns. Some common ways to resymbolize interactively are: (1) reclassify, (2) recolor, (3) reproject, (4) remodel, and (5) reorder.

Reclassify by changing category breaks or number of categories in a classed map. Or by changing category membership of individual elements (e.g., changing a 'church' from a 'public building' to a 'house of worship'). This can reveal information lost using other categorization methods.

Recolor by swapping color schemes, color ramps, or other color choices. Swapping in a nontraditional scheme, for example, may highlight outliers or emphasize subtle differences.

Reproject by altering the transformation of the three dimensional Earth onto a computer screen, or by rotating a three dimensional scene on a virtual map. This will, of course, alter the distortion properties of the graphic, and may reveal unforeseen patterns and generate new insight. Modern in car navigation systems often allow the reprojection of the map display to show an oblique bird's eye view or a traditional plan view according to the preference of the user.

Remodel by varying the type of thematic map to represent data (e.g., from a choropleth to proportional symbol map). These thematic map types (also including dot density, dasy metric, flow, isarithmic, and others) are referred to as cartographic models, and each is designed to represent various data models optimally. However, in some cases, a change in cartographic model may prompt different ways of thinking about the information and facilitate insight.

Reorder by switching frames of an animation or moving the stacking order of data layers. One subtle form of symbol ization is the order in which elements of the map are presented visually. In an animation, frames are typically ordered chronologically, but this need not be so. In a GIS display, there is no conventional ordering of data layers (other than perhaps lowest to highest on the Earth – geology overlaid with land use overlaid with streets and buildings). In both cases, changing the order of the elements can result in new insights that would have re mained invisible without interactivity.

These are only a few of the many ways that symbol ization on a map can be altered using interactive tech niques. In many cases, as described below, established rules of static cartography may no longer be appropriate, particularly as the use of maps has expanded from means of public communication to tools of private data analysis.

Explore

While static maps have served the needs of geographers who wish to communicate an idea to a mass audience, maps designed for individuals to visualize large volumes of information with the goal of developing hypotheses – that is, maps for geographic visualization – are almost always interactive. The dialog between analyst and data is provided by the interactive elements whose purpose is to discover trends, isolate (or remove) outliers, visualize relationships, and access detailed information. Though it may be possible to accomplish some of these tasks through the use of static maps, the addition of inter activity has truly realized the use of maps for data ex ploration. Data exploration may be supported by interaction through: (1) drilling down, (2) manipulating visibility, (3) query, (4) focusing, and (5) lensing.

Drilling down, which is a move from summary, or aggregated, information to detailed data in order to examine something more thoroughly. Interaction with a classed polygon on a map, for example, might result in the display of detailed tab ular information (such as the actual data value) corres ponding to that place.

Manipulating visibility by either adjusting the transparency of a map element or toggling on and off a data layer. This is common in GIS and recalls John Snow's exploration of various data layers to solve the cholera epidemic in London in the 1850s.

Querying, which is, broadly speaking, a request for infor mation, can be done via command line querying languages, such as SQL, but it can also be accomplished through visual inter action. Examples of visual querying include 'brushing' and 'linking', two basic functions of multiple representation environments. Such environments (GIS is one) are characterized by several representations of the same or related data, providing context and perspective is im possible with a single representation. In most GIS, a subset of records of a large dataset is selectable through a spatial query or through an attribute query. This brush ing operation is often used in combination with linking, through which selected records in one view (e.g., the map) are also selected (and manipulated) in other rep resentations (e.g., tables, charts, and other maps).

Focusing on a subset of data values or spatial regions through eliminating (or generalizing) regions of less interest is another related visual querying operation. A specific form of focusing is called 'conditioning', through which attribute queries are performed on individual variables, displaying, for example, only those points between two values of one important variable. Conditioning reduces the amount of data displayed by displaying only those records that meet certain criteria.

Lensing simulates a fish eye lens or other physical magnifying tools to highlight a region while de emphasizing unlensed re gions. This adds detail to a region of interest, distorting its surroundings and reducing their visual importance but allowing them to remain visible. The lens may be as intuitive as a box attached to the cursor that exposes a satellite image or detailed map within the box in the corresponding location on the less detailed map. This is a form of particularization that is borrowed from the ex ploratory data analysis (EDA) literature in statistics.

Enabling the exploration of complex data is the goal of much current research, both within and outside geog raphy. New modes of interacting with these datasets are being created constantly. While the addition of inter activity on maps has profoundly improved the functionality and potential of all maps, it is most vital and most relevant for those maps designed for visual analysis and data exploration.

Means of Interacting with Maps

While the forms of interaction described above refer to computer map applications, humans have 'interacted' with the maps they use, regardless of the map's medium. A pedestrian rotates a map to align it with the direction of movement, a traveler annotates a map with personal notes about a visited place, and a meteorologist flips through several forecast charts to observe air movements. Now with computer enabled mapping, the ways we interact with maps are governed by the hardware and software interface tools of the computer.

Hardware Interaction

Most hardware interaction with present day computer maps (and with most other graphics applications) takes place with the mouse and the keyboard. Maps can respond to many different mouse actions, and in many cases the mouse movements require practice and significant dexterity. Of course, the basic mouse movement is the point and click, which (typically) selects the object below the mouse. Depending on the application and, sometimes, on the interface tool selected, doing this on a map can recenter, zoom in, highlight, select, drill down, link, or almost any other function. Using a mouse is significantly more complex, however, as a simple mouse over (pulling the mouse across an object on the screen without a click) might highlight an object or corresponding objects in different displays. Click dragging the mouse on the screen often results in a pan operation, moving the map underneath the cursor. Click dragging can also be used to select a group of close objects in a variety of shapes (a rectangular box or an irregular 'lasso'). Even within click dragging, there exist subtle variations in some applications: in Google Earth, a click drag with a stop will 'spin' the virtual globe and stop, but a click drag with a flick of the wrist at the end will spin the globe without stopping. Clicking the mouse button while holding a key of the keyboard down (e.g., the 'Alt' or 'Ctrl' key) might result in an alternate use of the tool being used (e.g., constraining the selection box to a square, or zooming out rather than in). Keys on keyboards can of course also be used to enter text annotations on computer maps or type in commands or data input.

Mouse and keyboard interactions, of course, require the computer on which the map is displayed to actually have a mouse and keyboard, but many interactive maps (and many more in the future) will exist on different computing media that will require new interaction tools and techniques. Trends in computing are toward smaller and flexible displays for use on cell phones, GPS navigation units, and hand held computers, and in high mobility systems, such as fighter pilot visors and scroll like rollable displays. These applications typically do not have familiar interface tools but rather rely on alternate interaction modes such as speech (using headset microphones and/or speech recognition software) and gesture (using data gloves and/or motion recognition software), or new interaction metaphors that replicate popular tools like iPods and game controllers (not unlike the video cassette recorder (VCR) and car–radio interface metaphors of today's popular applications). These forms of interaction require entirely new models of human–computer interaction, involving a formalization of speech, gesture, and motion so that a computer may understand enough to carry out the complex tasks required of visualization and (geo)graphic communication.

Software Interaction

Present map interaction occurs through interface widgets on the computer screen. These widgets include slider bars, buttons, menus, and text input lines. Also important are means of direct manipulation, where the movements of the mouse or other input device impact the display (and not via another interface widget). The transparency of the (primarily direct manipulation) Google Earth interface contrasts with the complexity of the (primarily tool and menu driven) interface of GIS (though GIS is, of course, far more extensive in its uses). Selecting, brushing, panning, and drilling down of objects on the map using simple mouse interactions are examples of direct manipulation that have made map interactivity intuitive and popular.

Mapping content delivered over the Internet has transformed dramatically in only a few years. One of the most noticeable changes is the seamless browsing that is now standard on commercial interactive maps, such as Yahoo and Google Maps. Only 5 years ago, zooming and panning required discrete steps: upon a user request (for a zoom in, for example), the map would disappear and then be replaced by an entirely new map. Presently, these counterintuitive discrete changes have been superseded by continuous scale and position changes, as adjacent regions and resolutions are 'buffered' or 'streamed' into the ever expanding memory of personal computers, and the browsing of maps has ironically become much more similar to our interaction with 'real world' paper maps (albeit incredibly rich, powerful, and detailed paper maps).

Notable Interactive Mapping Applications

Many of the capabilities and innovations described above exist in a variety of different applications with a variety of purposes and goals. In the previous section, several commercial mapping applications were mentioned. Google Earth (along with its web browser companion Google Maps, which is conceptually similar to, but more visually stunning and detailed than the original, widely used mapping application, MapQuest) has transformed the way the general public looks at spatial information though an easy to use interface. Google Earth is limited by its distorted nonglobal scale projection ('ortho graphic') and its uneven updating of imagery, but few would deny that the interaction provided by Google Earth makes it fun, educational, and rich with possi bilities. Microsoft has released Live Search, a mapping interface with an innovative remodeling/reprojection functionality that allows users to see detailed high resolution aerial photography (within a presently very limited group of cities) from an oblique bird's eye view. British tourists are familiar with the MultiMap interface, which uses a clever form of the lensing interaction described above, and US users can explore extensively with the Java based Acme Mapper application, which allows the overlay of US Geological Survey topographic maps and digital orthophotoquads onto current satellite imagery and vector maps. The examples above are primarily developed as Internet based maps; the work of Michael Peterson, who leads the International Carto graphic Association's Commission on Maps and the Internet, provides contextual frameworks for these applications.

Of course, geographers have been using interaction on maps for years through GIS interfaces. Though not many would argue that user interfaces of present GIS are in tuitive, they do provide a remarkable range of interaction possibilities that position GIS well as an environment for a wide variety of cartographic tasks. An alternate web based GIS called CommonGIS has been designed by a team led by Gennady and Natalia Andrienko for visual analytical tasks, with tremendous flexibility for resymbolization and exploration. An intriguing application that allows the interaction of generalization (see 'adjusting intricacy' above) is MapShaper, developed under the supervision of Mark Harrower at University of Wisconsin–Madison, which provides a simple slider bar that adjusts polygon generalization while preserving topology. 'Conditioned choropleth maps' such as those created by Daniel Carr and others exemplify the concept and advantages of focusing and conditioning in determining bivariate and multivariate relationships. More advanced spatial statistics are provided in GeoDa, a software environment developed by a team led by Luc Anselin that implements many of the ex ploratory brushing and linking techniques to coordinate multiple views. Multimodal speech and gesture interface research has resulted in prototypes such as the Penn State GeoVISTA Center's DAVE G, developed by Sven Fuhrmann and Alan MacEachren.

Research Challenges of Map Interactivity

Clearly, interactivity with maps is highly dependent on software and hardware advancements, about which many research questions will need to be posed. As computing becomes more and more ubiquitous, humans are inter acting with computer interfaces in ways that could barely be imagined only a decade ago. The rapid development of nanotechnology has led to mobile display devices that are controlled by touch sensitive screens (e.g., Apple's iPhone) or voice commands (e.g., in car navigation sys tems). The point and click mouse may soon be sup planted, even on desktop systems, with novel interaction methods such as eye tracking (the mouse follows the user's gaze), gesture tracking, or motion sensing (familiar to Wii video game users). These new interaction tech nologies will require analysis of their effective use for geographic representation and understanding.

The way humans use geographic representations is changing in important ways as we become better able to use computers to communicate instantaneously, globally, and with great detail. Successful collaborative work, for example, through web conferences, social networking, or virtual environments will require cartographers to understand important cognitive and usability issues. As examples, what happens when two individuals interact with a map being viewed by many people in contra dictory ways? What is the need for internationalization of mapping user interfaces to conform to cultural conven tions? Can sound or other senses other than vision be exploited to communicate or manipulate a map's con tent? These are but a small subset of research questions that will challenge cartographers and GIS scientists concerned with map interactivity in the coming years.

Summary

Interactivity in maps serves a wide variety of purposes and makes present maps more powerful than ever before. Many interaction forms used on popular interactive mapping applications exemplify the limitations of static paper maps. Interaction, for example, provides spatial and temporal context, allows for particularization and in creased detail, affords alternate symbolization strategies customized for individual users, and facilitates explo ratory spatial data analysis tasks. These general tasks, which are not mutually exclusive, form an organizational framework for categorizing and inventorying interaction methods on maps in the early twenty first century. In many ways, the provision of interactive tools and tech niques by a cartographer designing and interactive mapping system is itself a form of symbolization. Choosing which modes and functions of interaction to include for users can either facilitate or hinder understanding of information represented on dynamic mapping systems, much like the influence of traditional symbolization on static maps. Making proper choices when designing and utilizing such systems starts with an understanding of the present capabilities and future potential of interactive maps.

See also: Cartographic Animation; Digital Earth; Geovisualization; Internet/Web Mapping; Map Perception and Cognition; Mapping, Cyberspace; Maps.

Further Reading

Andrienko, G. and Andrienko, N. (1999). Interactive maps for visual data exploration. *International Journal of Geographical Information Science* 13(4), 355 374.

Carr, D. B., MacPherson, D., White, D. and MacEachren, A. M. (2005). Conditioned choropleth maps and hypothesis generation. *Annals of the Association of American Geographers* 95(1), 32 53.

Crampton, J. W. (2002). Interactivity types on geographic visualization. *Cartography and Geographic Information Science* 29(2), 85 98.

Edsall, R., Andrienko, G., Andrienko, N. and Buttenfield, B. (2008). Interactive maps for exploring spatial data. In Madden, M. (ed.) *Manual of Geographic Information Systems*. Bethesda, MD: ASPRS.

Harrower, M. and Bloch, M. (2006). MapShaper.org: A map generalization Web service. *IEEE Computer Graphics and Applications* 26(4), 22 27.

MacDougall, E. B. (1992). Exploratory analysis, dynamic statistical visualization, and geographic information systems. *Cartography and Geographic Information Systems* 19(4), 237 246.

Peterson, M. P. (1995). *Interactive and Animated Cartography*. Englewood Cliffs, NJ: Prentice Hall.

Rao, R. and Card, S. K. (1995) Exploring large tables with the table lens. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI'95)*, pp 403 404. Denver, CO: ACM.

Shneiderman, B. (1997). Direct manipulation for comprehensible, predictable, and controllable user interfaces. In *Proceedings of the ACM International Workshop on Intelligent User Interfaces* '97, pp 33 39. Orlando, FL: ACM.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison Wesley.

Unwin, A. R. and Hofmann, H. (1998). New interactive graphics tools for exploratory analysis of spatial data. In Carver, S. (ed.) *Innovations in GIS*, vol. 5, pp 46 55. London: Taylor & Francis.

Relevant Websites

http://mapper.acme.com
Acme Mapper.

http://www.commongis.com
CommonGIS.

http://www.geovista.psu.edu
Conditioned Choropleth Maps, Digital Government Quality Graphics, GoeVISTA.

https://www.geoda.uiuc.edu
GeoDa.

http://earth.google.com
Google Earth.

http://maps.live.com
Live Search.

http://www.mapshaper.org
MapShaper.

http://www.multimap.com
Multimap.

http://www.powersof10.com
Powers of Ten film and website.

Map Libraries and Archives

R. B. Parry, The University of Reading, Reading, UK

Glossary

Archive A store of documents or other material, usually of historical and/or rare items. In computer terms, archives also refer to copies of data held on electronic media for their long-term storage and future access.

Geodata/Geospatial Data Any data which have a spatial component.

Geolibrary A library of georeferenced objects (not exclusively maps) which can be searched by geographical location or area. A distributed geolibrary is one in which data and metadata may be retrieved from many distinct locations.

Metadata Data that describe the content, data definition and structural representation, extent (both geographic and temporal), spatial reference, quality, availability, status, and administration of a geographic data set.

Online Catalog Computer database that lists and describes the holdings of a library. Sometimes referred to as an 'online public access catalog' or ' OPAC'.

Open Source Mapping Mapping, generally computer mapping, which may be used free of any legal or technical restrictions.

Spatial Data Clearinghouse An online facility for searching numerous data collections using metadata tools.

Introduction

Map libraries are of many kinds. At the top of the hierarchy are major national libraries or archives re sponsible for the collection and conservation of maps of the home territory and beyond. Examples include the Geography and Map Division of the Library of Congress in the United States, the British Library Map Library in the UK, Library and Archives Canada, and departments within the National Library of Australia and the Bib liothèque Nationale de France. Many of these libraries are legal deposit libraries, giving them the legal means to acquire and build comprehensive collections through time of the nation's published mapping. In some coun tries legal deposit also extends to other, regional libraries. At the national level, maps related to the country's his tory and diplomatic activities are often held in a national archive rather than national library. Maps of local historic interest are held in local archives such as, in the UK, county record offices.

Map collections are also often found in university and similar educational institutions, either as part of the general library system or attached to a specific discipline such as geography, planning studies, or earth sciences. Such collections are usually concerned with the more specific research and teaching requirements of the insti tution. Important public access map collections are also associated with major geographical societies such as the American Geographical Society or the Royal Geo graphical Society, London. Manuscript maps and valu able collections of historical maps are often separated from a working map collection and held as 'special col lections' or within an archive department devoted to a broader range of historical material.

The rapid growth of map libraries in universities and the maturation of the profession of map librarianship in the Western world were a feature of the 1960s and early 1970s. This partly resulted from the widespread distri bution of surplus mapping following World War II and the escalation of mapping activities globally in the post war period. Many of the collections which developed in universities at this time depended heavily on the free distribution of such maps. In contrast, restrictions on official mapping in the Soviet bloc countries meant that detailed modern topographic maps of these countries remained unavailable publicly both internally and ex ternally until their release following the dissolution of the Soviet Union in 1989.

Listings and descriptions of a broad spectrum of map collections can be found in published directories such as the *World directory of map collections*, the *Guide to US map resources*, and the online *Directory of UK map collections*. Most major collections and many minor ones also have their own websites.

Map Collections in the Digital Era

In the predigital era geographical information was available in printed form, and map collections were val ued storehouses of geospatial data in this format. A major function of most national libraries and archives has been the collection of material of historical importance and its conservation for the future, as well as making it accessible to the public. University map libraries on the other hand have been less concerned with archiving and conser vation, focusing more on the needs of users requiring contemporary maps and geodata.

The last two decades of the twentieth century saw contemporary data resources increasingly move to a digital format, initially held on computer readable tapes and disks and ultimately on servers linked to and accessible via the web. Simultaneously many more sets of geographical data became available in geocoded form. In the UK it has been suggested that between 60% and 80% of all government data can be classed as 'geospatial'. In fact any spatially referenced data can be mapped, and this is not confined to government data – house prices, crime rates, traffic flows, noise pollution, health statistics, poverty and deprivation, lifestyle data, valued landscapes, and heritage sites are just a few examples of the range of mappable themes. Just as the arrival of digital data and the development of geographic information systems (GISs) heralded a radical change in the way maps are constructed and understood, so the nature of map libraries and the function of their curators were bound to change.

The major shift from paper maps to the acquisition and supply of digital maps and spatial data began substantially to affect libraries during the decade of the 1990s. Map libraries adapted most quickly to the new media in the US, where the numerous federal depository libraries began to receive large volumes of government data distributed digitally on CD ROMs. Here the use of GIS in libraries was stimulated by the Association of Research Libraries (ARL) GIS Literacy Project, started in 1992. GIS specialists were recruited and existing map librarians were reskilled to allow them to help their user community access and generate maps from digital data. This project was subsequently extended to Canada, where many university libraries have also begun to offer GIS services. In Britain and elsewhere, where government data were not always so freely available, map libraries were slower to respond. With the coming of the web however, it became impossible to ignore the increasing availability of maps online and the demands for digital data.

The digital era has affected not only the way data are stored and distributed, but also the way they are cataloged and made accessible. These, and associated issues of freedom of information, copyright protection, intellectual property rights (IPR), and data security, have all been concerns of increasing importance for map librarians.

Making Geodata Accessible to the User

During the 1990s and early 2000s, many major map libraries, which were already cataloging their new accessions on computer, undertook retrospective conversion of their earlier card catalogs to a digital format compliant with established machine readable cataloging (MARC) standards (developed to facilitate the exchange of bibliographic data in machine readable form) and Anglo American Cataloguing Rules (AARC2). Many of these catalogs became available as online public access catalogs (OPACs), or were added to cooperative catalogs such as COPAC, a union catalog giving access to the merged catalogs of a Consortium of Research Libraries in the UK.

During the same period, organizations outside the immediate sphere of libraries were developing metadata standards for the description of digital geodata sets. Thus the Dublin Core Metadata Initiative was set up to develop international and interoperable standards of data description and exchange in general, while in the United States the Federal Geographic Data Committee developed content standards for geospatial data as part of a larger remit to implement the National Spatial Data Infrastructure. Data searching facilities were also developed, such as GIgateway funded by the British Government to provide access to UK geospatial datasets, and the European Geoportal for European spatial data.

Essentially three kinds of digital data are of concern to map libraries and archives. First there is the stream of data emanating from national mapping organizations, and other government and nongovernment agencies in place of maps no longer distributed in printed paper format. Second there are statistical data such as the national censuses, and other thematic data which, through being both digital and spatially referenced, can be mapped or interrogated using GIS software. The third kind, much of which is being produced by the map libraries themselves, comprises maps, especially early maps, which have been scanned from paper originals. These latter afford collections an opportunity to showcase some of their prized possessions on the web. More importantly it makes early maps accessible in a digital environment and takes pressure off the use of originals. Given the epidemic of thefts suffered by major map libraries in the 1990s, this practice may also help to improve the security of the originals.

Many substantial collections of raster digitized maps are available on library websites: the Library of Congress, for example, has more than 10 000 maps online. The David Rumsey Collection and the extensive resources of the Perry Castañeda Library Map Collection at the University of Texas are other notable examples in the United States. In Britain, the British Library, The National Library of Scotland, and The Royal Geographical Society have been particularly active, using maps as a showcase through virtual exhibitions and themed tours, and providing online access to sometimes unique historical material. Historical statistical data, such as British decennial population censuses from the nineteenth century held in the National Archives, have also become available on the web.

In order to disseminate information about what has been digitized, and to avoid unnecessary duplication of effort, the Government Printing Office in the US set up a *Registry of U.S. Government Publication Digitization Projects* on the web to serve as a clearinghouse for digitized government data. Similar initiatives have been taken by map library organizations, such as the Map and Geography Round Table of the American Library Association (MAGERT) and the Western Association of Map Libraries (WAML) in the US and the British and Irish Committee for Map Information and Cataloguing Systems (BRICMICS) in the UK and Ireland.

To a large extent however, developments in access to digital geodata have proceeded outside the map library arena. Spatial data clearinghouses have been established by national mapping agencies, census bureaus distribute data from their own websites, and other nonlibrary organizations have developed the standards for metadata, data exchange formats, and facilities for online searching.

The Distributed Geolibrary

One direction in which map libraries have been moving is to operate as nodes in a distributed geolibrary system. The term geolibrary was coined by Michael Goodchild to denote a library of georeferenced objects or datasets which could be searched for and accessed through a catalog of metadata. Beginning 1994, he and his team developed the first extensive geolibrary, the Alexandria Digital Library (ADL) based at the University of California, Santa Barbara.

The idea of a distributed geolibrary is that libraries hold and share data appropriate to their geographic footprint, but make it searchable and accessible through the web. The term has subsequently been much used, often rather loosely, to describe initiatives, both national and local, to build online libraries of geodata. In Great Britain such a project, though centrally distributed and available only to subscribing institutions of higher education, emerged as the Digimap service run by the Data Library at the University of Edinburgh. Launched in 2000, and initially providing teachers and researchers with access to Ordnance Survey digital data, it has grown to include historical, land use, and geological mapping, and may expand further. Another example from Great Britain, in this case developed independently of a library but with the cooperation of map librarians, is the Great Britain Historical GIS with its web presence as *A vision of Britain through time.*

Conserving Geodata

As more and more maps and geodata began to be stored on magnetic media – and often only on magnetic media – concerns developed about the conservation and future proofing of data. Many initiatives have been taken to develop policies for the long term archiving and preservation of geospatial data. In the United States, a National Digital Information Infrastructure and Preservation Program was established under the Library of Congress, with several university map collections researching digital preservation issues. In the UK, a Digital Curation Centre was established in 2004 to support and advise higher and further education on matters of digital preservation. Issues concern the longevity of digital media, and the transferability and future proofing of data formats.

In the case of 'born digital' data, decisions have to be made and policies formulated as to what data sets should be preserved. Numerous initiatives have been set up and reports (e.g., the British Library's handbook on *Preservation management of digital materials*, or the Royal Statistical Society and UK Data Archive's report on *Preserving and sharing statistical material*) published to consider and advise on these questions. It is often difficult to recognize the possible future value of data collected for a specific application, and such data are easily lost. It is crucial, for example, to preserve boundaries of the spatial units to which statistical data can be mapped. At the same time, given the exponential growth in digital data, librarians and archivists have, as always, to make decisions as to what should be preserved. Geodata are just one part of a nation's digital memory, and the implications of capturing the latter are an ongoing issue for all archival collections.

As mentioned, digital preservation also extends to conserving the contents of paper maps through making digital surrogates. Major libraries have invested in high end flatbed scanners to copy what are often the most prized (but not necessarily the physically most vulnerable) early maps in their collections. Many maps in archives are included in bound volumes, making them difficult to scan effectively without putting the integrity of the volume at risk. To do justice to the originals, relatively high scanning resolutions must be adopted giving rise to very large data files and the need for high storage capacity, as well as problems of file transfer via the Internet. The latter can be solved through use of compression programs such as MrSID or Zoomify. Once captured such maps can also take on a new life within a GIS.

Threats and Responsibilities for Map Libraries in the Digital Era

Throughout the 1990s and early 2000s, meetings and conferences of map librarians invariably reflected the angst felt by the need for their role to be reassessed. Meanwhile some outside the profession began to question the need for map libraries, paper maps, and hence their keepers altogether. It became clear that some smaller university based map libraries which had not moved with the times were at risk of closure. This began to happen in the UK in the 1990s and with accelerating pace in the new millennium with closure or downsizing of many such collections. In the United States too, a reduction in the number of collections was noted in the latest edition of the *Guide to U.S. Map Resources*. Moreover, organizations using geodata have often overlooked the possible archival importance of their paper maps when moving to a digital system, and the maps have not always been saved from the skip.

Map libraries face many new challenges in the digital era, but they undoubtedly have an important role to play in safeguarding the paper past and preserving the digital future. One aspect of this is the need to conserve time slices of official mapping which in digital form is under continuous revision. A secondary function is to be pro active in the acquisition, archiving, and cataloging of data sets which might otherwise be lost to posterity altogether, products of the diffuse new mapping industry which has grown up with the developments in GIS and new map ping technologies.

In some countries, the commercial pricing of digital spatial data has impeded its acquisition by libraries. In the UK, the legal deposit libraries were successful in negotiating with Ordnance Survey a supply of annual digital 'snapshots' of its large scale digital database, be ginning 1999. Similar negotiations in Canada between the Canadian Association of Research Libraries and Natural Resources Canada were successful in providing access via the depository program to the National Topographic Database. New laws are being introduced to cover the legal deposit of digital data, and with the growth of open source mapping on the web, government organizations are under increasing pressure to make their data more freely available.

Map libraries which have embraced the digital revo lution have had to decide what level of service to offer users of digital geospatial data. No standard policy has emerged: models range from simply helping users locate and download data to provision of a full digital mapping service. To some extent, the level of service depends on the library's institutional framework, the requirements of the users, and available staffing.

With geodata deliverable directly to the desktop PC, and with the facility of web enabled GIS, it might still be argued that map libraries are redundant. In fact their role is crucial in safeguarding digital data for future gener ations and in preventing the loss of printed maps, dis carded on the assumption that 'someone else will have them'. In 2006, Canadian map libraries were successful in lobbying to reverse a decision by the Canadian Gov ernment to discontinue printing paper topographic maps. The role of map librarians themselves in interpreting the public's need for maps and promoting cartographic understanding has been enhanced by the need for them to serve as mediators between digital data and their users. While many collections will continue to be abandoned, and their resources dispersed or even destroyed, those that remain may become leaner but also hopefully fitter to meet the new demands of the information age.

See also: GIS and Cartography; GIScience and Systems; Integrated Spatial Data Infrastructure; Mapping, Distributed; Spatial Databases.

Further Reading

Boxall, J. (2004). Advances and trends in geospatial information accessibility Part 1: Geolibraries. *Journal of Map and Geography Libraries* 1(1), 7 39.

Boxall, J. (2006). Advances and trends in geospatial information accessibility Part 2: Policy dimensions. *Journal of Map and Geography Libraries* 3(1), 55 76.

Goodchild, M. F. (2004). The Alexandria digital library project: Review, assessment and prospects. *D Lib Magazine* 10(5): May 2007. http://www.dlib.org/dlib/may04/goodchild/05goodchild.html (accessed on 18 July 2007).

Jones, M. and Beagrie, N. (2001). *Preservation management of digital materials: A handbook*. London: British Library.

Keller, C. P. (2001). The map library's future. *Cartographic Perspectives* 38, 66 70.

Loiseaux, O. (ed.) (2000). *World directory of map collections* (4th edn.). Munich: Saur.

Marvin, J. and Fortin, M. (2006). Putting literacy into geographic information. *Bulletin of the Association of Canadian Map Libraries and Archives* 125, 3 9.

Panel on Distributed Libraries, National Research Council (1999). *Geolibraries*: *Spatial information resources*, *summary of workshop*. National Academies Press. http://books.nap.edu/html/geolibraries/ (accessed on 18 July 2007).

Parry, R. B. (2007). Map libraries. In Bowman, J. H. (ed.) *British librarianship and information work 2001 2005*, pp 214 230. Aldershot: Ashgate Publishing.

Parry, R. B. and Perkins, C. R (eds.) (2001). *The map library in the new millennium*. London: Library Association Publishing/Chicago American Library Association.

Royal Statistical Society and UK Data Archive (2002). *Preserving and sharing statistical material.* Colchester: UK Data Archive. http://www.data archive.ac.uk/news/publications/PreservingSharing.pdf (accessed on 18 July 2007).

Thiry, C. J. J. (ed.) (2006). *Guide to U.S. map resources* (3rd edn.). Lanham, MD: Scarecrow Press.

Zellmer, L. (2003). Chasing mayflies: Archiving spatial data. *Western Association of Map Libraries Bulletin* 34(2), 71 75.

Relevant Websites

http://www.cartography.org.uk
A Directory of UK Map Collections 4th Edition: Main Index. The British Cartographic Society, The Map Curators' Toolbox.

http://mapregistry.library.arizona.edu
ALA MAGERT Map Scanning Registry.

http://www.alexandria.sdc.ucsb.edu
Alexandria Digital Library (ADL).

http://www.davidrumsey.com
David Rumsey Map Collection.

http://www.dcc.ac.uk
DCC, Digital Curation Centre.

http://www.gigateway.org.uk/
Gigateway.
http://eu geoportal.jrc.it
INSPIRE, European Geo Portal.
http://www.lib.utexas.edu
Perry Castaneda Library Map Collection.
http://www.gpoaccess.gov/legacy/registry
Registry of U.S. Government Publication Digitization Projects.
http://www.geo.uu.nl
Universiteit Utrecht, GEO Geowetenschappen.
http://www.visionofbritain.org.uk/
Vision of Britain through Time.

Map Perception and Cognition

S. M. Freundschuh, National Science Foundation, Arlington, VA, USA

Glossary

Cognition It is the mental faculty or process of acquiring knowledge by the use of reasoning, intuition, or perception.

Euclidean Relating to Euclid's system of geometry, Euclidean refers to an *x*–*y* coordinate system where lines are parallel and perpendicular to each other.

Geographic Information System (GIS) It is a system for capturing, storing, managing, analyzing, and representing spatial information that is spatially referenced to the Earth.

Map Animation It is the rapid display of a sequence of maps in order to create an illusion of dynamic movement.

Orthophoto It is an aerial photograph (a picture of the Earth taken from directly overhead) that has been geometrically corrected ('orthorectified') such that the scale of the photograph is uniform, meaning that the photo can be considered equivalent to a map.

Psychophysical It is the branch of psychology that measures the effects of physical stimuli on people's sensory perceptions and mental states.

Selective Attention It is a type of attention which involves focusing on a specific aspect of an experience while ignoring other aspects.

Spatial It relates to people or objects that occupy and/or move about in space, or events that occur in environments, such as houses, neighborhoods, and cities.

What People See: Psychophysical Studies

Graphic Elements

Map perception research explores the perception and understanding of map symbols. This work, at its most basic, consists of presenting to people various map symbols (i.e., graphic marks) in an experimental setting, and measuring responses to these various symbols. The purpose of this type of research is to understand more fully what people see and what people understand from maps, and to use these results to inform the design of more effective maps. A useful framework to use for a discussion of psychophysical studies is the standard graphic elements that are used to design the graphic marks that represent features on maps. These elements are classified as either qualitative or quantitative. All graphic marks show the location and extent of a geographic feature or a spatial phenomenon. The qualitative graphic elements of 'shape', 'pattern arrangement', 'pattern orientation', and 'color hue' further distinguish between kinds of geographic features. The shape of a symbol refers to point and linear features, for example, a square representing a house or a dashed line representing a political boundary. Pattern arrangement involves the arrangement of graphic marks to show different areal features, for example, a stipple pattern on a geologic map to show sandstone, or a brick pattern to show limestone. Pattern orientation refers to the orientation of the graphic marks in areal features, for example, horizontal lines to show a cornfield and vertical lines to show soybeans. Color hue refers to the basic set of primary colors of red, orange, yellow, green, blue, and purple. Color hue is effective for showing different point, line, and areal features, for example, a black dot to show the location of a city, or blue to show water.

Quantitative graphic elements add either hierarchy and/or quantity information to a graphic mark, and include 'size', 'pattern texture or gray shades', 'color value', and 'color intensity'. Size of a symbol is a straightforward way to illustrate varying amounts of a phenomenon, for example, a narrow line to show low traffic flow and a wider line to show higher traffic flow. Pattern texture creates the visual impression of light textures that represent low quantities and dark textures that show higher quantities, for example, a country with a light gray to show low population density and a dark gray to show high density. Color value refers to the amount of black in a color, a low amount of black producing light colors that show low quantities and a higher amount of black producing darker colors that show higher quantities (e.g., using a light green to represent a young forest and a dark green to show older forests). Color intensity like value also creates the visual impression of light to dark, but instead refers to the amount of a color. For example, a light red to illustrate a low temperature range, with increasingly brilliant reds to show higher temperature ranges. Psychophysical studies of map perception have typically explored how people perceive these various graphic elements.

Early Psychophysical Studies

There are several well known psychophysical studies that have served as the foundation for much of the current work in map perception studies. One such study is the perception of proportional circles. Circles on a proportional symbol map are scaled (based on the area of the circle) to the value they represent. For example, a circle representing the value of 1000 would have twice the area of a circle representing the value of 500. A study by James Flannery in 1956 (published in 1971) showed various sized proportional circles to over 1000 people. Flannery found that subjects in his study increasingly underestimated associated circle values as circle size increases. To correct this underestimation, so that a map reader would more accurately perceive circle values, a formula was developed that increasingly inflated circle sizes as circle size increased. This formula is based on a calculation using a constant value known today as Flannery's constant. During the same time period, Robert Williams (1956) was also working on people's perception of map symbols, exploring a number of parameters that included symbol shape and size. Williams is most noted for his work on the perception of graytones. Graytone refers to the percentage of black (inked area) covering an area on a map. A 10% inked area would result in a very light graytone, whereas a 70% inked area would be a dark graytone. Graytones are often used on choropleth maps to show individual categories. Williams conducted an experiment in which subjects were shown various graytones, and from those he estimated the percentage of black of each graytone. The results demonstrated that people overestimate the percentage of black for graytones less than about 60% and underestimate the percentage of black for graytones over 60%. From this, Williams created a curve that matches the perceived graytone to the actual graytone, or the percentage of area inked. Work related to Williams' research on graytones is on the use of continuous tone maps to show choroplethic type data.

Color Perception

Cartographers have borrowed extensively from psychology and from graphic design with regard to color perception. Research on color perception on maps has included studies on the perceptual psychological system (eye–brain), as well as on people's subjective reactions to color. Color perception on a map is influenced by colors in the surrounding map environment. The result is color interactions, most notably simultaneous contrast and spatial effects, which can cause colors to appear different under differing map environments. Simultaneous color contrast occurs when a color, surrounded by another color, appears tinged by the complementary color of the surrounding color. For example, a gray symbol surrounded by blue will appear tinged with yellow, the complementary color of blue. The cartographer must be thoughtful when choosing and placing colors on a map to thereby minimize this effect. Spatial effects of color happen with high contrast between two colors that can subsequently affect the perceived depth of objects. For example, if yellow and purple symbols are placed on a black background, the yellow symbols will appear closer to the viewer than the purple symbols. This illusion is useful for establishing strong figure development on a map.

Certain colors can incite feelings and emotions, resulting in color preferences and implied associated meanings. Red–orange–yellow are considered warm colors, and green–blue–violet are considered cool colors. Feelings of aggression and excitement can be elicited by red–orange–yellow. Green–blue are often associated with feelings of peace and calm, whereas gray–brown–black are oftentimes associated with depression and sadness. Purple is often used to elicit emotions of dignity and royalty, and yellow with fun and cheerfulness.

Cynthia Brewer (2005) has conducted the most extensive work on color on maps. The result of this work was the development of a color selection tool called ColorBrewer. ColorBrewer is an online color selection tool that is designed to help mapmakers select effective color schemes for area–class maps and choropleth maps. The color schemes illustrated in this tool have been optimized for both sequential and qualitative data sets, and have been proven effective for class differentiation and feature identification.

Maps and Selective Attention

Research on 'selective attention' and maps, or in other words 'at what, and where do people look on maps?', has been important for understanding figure/ground issues and map design in general. Much of the experimental work on selective attention focused on the effect of Gestalt principles on 'figure' and 'ground' formation on the map scene. The figure should be what the map reader sees most prominent when looking at a map. The figure, in this sense, should visually stand out as a 'coherent whole' from the ground. This is critical for visual clarity of the mapped data. Formation of a figure is facilitated by a number of principles, including heterogeneity, contour, surroundedness, orientation, relative size, and convexity. A heterogeneous field is one that stands out as a figure against a background in a map. Heterogeneity can be achieved on maps using the perceptual grouping principles of 'proximity', where objects that are close together are seen as a group, and 'similarity' where objects of the same shape, color, or size are seen as a group. Objects on a map that have crisp or highly defined edges (contour) stand out against nonobjects, and subsequently are seen as figures. Places on a map that have a higher density of objects or features (surroundedness) stand out as a figure over areas that have a low density of objects. Related to this, because visual ground typically surrounds a figure, the smaller of two areas will be seen as the figure (relative size). Objects sharing the same direction (orientation) will create a visual form, and therefore be perceived as a figure. Finally, closed objects (closure) are seen as visual wholes, and therefore are perceived as a figure.

Eye movement studies, a method used to explore where people look on maps, have been useful in validating principles of figure/ground development. The basic design of an eye movement experiment is to place a map in front of someone, and photographically record on which part(s) of the map the person's eyes focus. This research has demonstrated that people focus on edges of features that have highly contrasting edges, or the edges of well formed regions. In other words, people tend to look more at places that have higher information content for them (the figure), and less at places that have lower information content (the ground).

Maps as Communication Devices

Early communication models for maps asserted that maps were more than a graphic to be viewed, and for the first time maps were linked to the field of communication theory. Rather than as a simple graphic, maps were seen as a means to pass on the cartographer's view of the world, via some symbol grammar, to the map reader. All map communication models possess the same basic components of 'source' (the geographic world), 'interpreter' (the cartographer), 'map' (symbol referent), and recipient (the map reader). The mapmaker, then, is seen as trying to communicate a purposeful message. Successful map communication would result if the map reader, after using the map, had the same 'worldview' of the mapped data as the mapmaker.

Communication models that were refined during the 1970s provided a theoretical framework that became a justification for empirical work in map perception and cognition. Cartographers and geographers focused on the cognitive role that maps played in spatial cognition. This focus became the foundation of 'scientific cartography', which is the science of conveying spatial (i.e., geographic) information from one person to another via maps. Research in scientific cartography ushered in research on 'semiotics' and 'linguistics', which are the study of the meanings of symbols and signs. It was at this time that French cartographer Jacques Bertin's (1967) work on semiology was discovered and embraced by the international cartographic community. Bertin's work introduced the then novel idea that certain map symbols have inherent characteristics that match characteristics of various geographic features and phenomena, and therefore can represent them more authentically. Bertin's work was not based on any research paradigm, but instead was based on his own intuitive commonsense about maps and graphic design. Subsequent research in cartography studied and verified Bertin's work on semiology of maps.

Since the publication of the first communication model for maps, much richer and more complex models have been developed. These refined models raised important questions about the thought processes used by the cartographer during the map design and creation phase of the models. These same models also began to raise critical questions about what the map user sees, and more importantly questions about what the map user understands and how the map user's spatial knowledge is augmented and/or modified. The encoding processes used by the cartographer to create the map, and the decoding processes used by the map reader to understand the map were considered to be two complementary cognitive systems. Research in this area began to consider that both the mapmaker and map reader possess prior knowledge – the cartographer prior to making the map, and the reader prior to using the map. There are differences in this prior knowledge, differences that are based on the amount and level of sophistication, and on error and bias. The unintended consequence of prior knowledge is miscommunication. The map user might see or realize a message that the cartographer did not intend, or, conversely, miss the message altogether.

Map Learning

Cognitive Maps

As early as 1913, the idea of a 'cognitive map' intrigued scientists, with spotty research taking place in the early 1900s. Research on cognitive maps became more of a central focus in the mid 1900s as psychologists and geographers alike began to see the relevance of cognitive maps to research in spatial cognition. The term cognitive map has been used in various ways, first by Edward Tolman in 1948 who proposed that humans construct map like representations in the brain, representations that provide a basis for navigation and wayfinding. Other meanings have equated cognitive maps to real world maps, or have instead suggested that they are functionally equivalent to maps, but may not take the form what we 'see' to be a map. For this article, a cognitive map is a mental representation that facilitates spatial learning, spatial thinking, and spatial behavior – or in other words, spatial cognition.

Spatial Knowledge

Maps have long been used for navigation and wayfinding. Interest in how to better facilitate these tasks began to develop in the 1970s, initially spawning much work on 'what information do people need to navigate?,' and 'what is the best way to present this information?' These two questions led to research on models of spatial knowledge, and on how people acquire new spatial information.

Early models for spatial knowledge and its acquisition defined landmark, procedural, and configurational information as the basic building blocks for spatial knowledge. Landmark knowledge provides the nodes (places in the environment) that are subsequently linked via procedural knowledge, or the routes taken between landmarks. Configurational knowledge provides the metric information, such as distances, direction, and orientation between landmarks and routes. Current theory suggests that these three kinds of knowledge are really interdependent, do not exist separately from each other, and are learned simultaneously.

Spatial Knowledge from Maps

Maps fit into this model of spatial knowledge in that maps provide distance, direction, and orientation (configurational) information most effectively. Spatial knowledge acquisition from maps facilitates more accurate Euclidean distance knowledge of an environment, whereas knowledge acquisition from navigating in an environment results in more accurate route distance knowledge. Knowledge acquisition from maps also facilitates more accurate object location knowledge, whereas acquisition from navigation results in more accurate orientation (direction) knowledge. Research has shown these same results when comparing knowledge acquisition from narratives that provide either procedural or survey knowledge.

Geographic Visualization

Geographic visualization (GVis) is still in its infancy, introduced widely with the publication in 1995 of Alan MacEachren's book *How Maps Work*. GVis offers a new paradigm within which to consider map perception and cognition. Conventional thinking in cartography was that the cartographer would interpret the world, choose appropriate graphic elements based on best practices, create a map with a specific message, and distribute to the map reader who was considered nothing more than a passive map consumer. This is the basis for initial map communication models. New map communication models add, at the very least, a loop wherein the map user becomes an active participant in the mapmaking process. GVis fosters greater interaction between map and map user, the result being an active and engaged map user. The map user is allowed and encouraged to explore the spatial data, to create multiple visualizations of the dataset, and to search for unknown spatial patterns and/or relationships in the map data. In essence, the map user becomes the cartographer, having control over many aspects of map creation.

New technologies such as WEB design and Internet software, graphic and animation software, and user interface design research have enabled this shift to the active map consumer. Map users can create, in real time, maps on the Internet that show locations of places and route directions between places, can overlay map symbols on orthophotos, can zoom in on any point on the Earth's surface, or can perform simple spatial analysis with basic geographic information systems functions. These same technologies have expanded our traditional graphic map elements to now allow the creation of maps that include the depiction of dynamic phenomena (animated maps), the use of sound to represent spatial data (sound maps), the use of touch to facilitate communication of spatial information (touch maps), and ultimately virtual reality that immerses the map user into a map space. These new ways to present map data provide information for multiple modalities, and potentially create new methods for acquiring spatial information with the possible result of creating new multiple cognitive representations.

Dynamic Design Elements

Until now, cartographers have relied on a handful of graphic elements to represent real world features and phenomenon on maps. Maps have essentially been a 'snapshot' in time, illustrating a single state of a particular geographic phenomenon. The map reader must infer dynamic processes from these static maps. Cartographers are now able to make animated maps, maps that can show both static states and dynamic processes. These dynamic maps are called animated maps. There are essentially six dynamic design elements that can be used to facilitate the design of animated maps. 'Display date' is the day and possibly time that the information being displayed in an animation frame was collected. 'Duration' refers to the length of time frames that are displayed in a map animation. 'Rate of display' illustrates changes in geographic processes that vary in intensity or amounts over time. 'Frequency' illustrates the number of geographic events, for example, snowfall, that occur over a period of time. 'Order' is the sequence of animation frames. Frames can be presented in chronological order, or the order can be modified to highlight a particular characteristic of a spatial process. Last is 'synchronization', which is the simultaneous display of two related spatial processes, therefore illustrating how the two processes are correlated. Spatial cognition is enhanced by the ability of the map user to control various aspects of the animation, aspects such as display speed, temporal range of the data, stopping and starting the animation, and temporal order of the display. The theory is that people will see spatial and temporal patterns that are not apparent or visible from static map images. This is still a growing area that needs much research to verify this theory.

Touch and Sound in Maps

Touch and sound design elements have also been developed – elements such as loudness, pitch, register, and duration for sound, and size, texture, orientation, and form for touch. Development in this area is relatively recent, and touch and sound variables have been used mostly in specialized mapping applications such as maps for visually impaired map users. Research is growing in this area with the development of 'natural screen interfaces' as one example. These interfaces are large sized digital displays that are touch sensitive. The map user can select a feature, or outline an area simply by touching the screen. These types of interfaces will continue to be developed, especially for planning environments where a group of decision or policymakers can interact with the map as a group to highlight areas, propose changes, or present various scenarios at with the touch of a finger.

Summary

This article provides a brief glimpse into the complex realm of map perception and cognition. It presents four periods of focus in map perception and cognition: psychophysical studies in perception, maps as communication devices, then maps and spatial cognition, and finally the development of GVis. The Further Reading section helps to explain the concepts here, as well as introduce new developments in map perception and cognition and point the reader to other readings in this fascinating field.

See also: Behavioral Geography; Cognitive Geography; Landscape Perception; Maps; Mental Maps.

Further Reading

Bertin, J. (1973). *Semiology of Graphics*. Berg, W. J. (trans.). Madison, WI: University of Wisconsin Press. (English translation of *Sémiologie Graphique* (1967), Paris: Gauthier Villars.)

Brewer, C. A. (2005). *Designing Better Maps: A Guide for GIS Users*. Redlands, CA: ESRI Press.

Flannery, J. J. (1956). The relative effectiveness of some common graduated point symbols in the presentation of quantitative data. *Canadian Cartographer* 8, 96 109.

Freundschuh, S. M. and Egenhofer, M. (1997). Human conceptions of spaces: Implications for GIS. *Transactions in GIS* 2, 361 375.

Kitchin, R. and Freundschuh, S. M. (eds.) (2000). *Cognitive Mapping: Past, Present and Future*. New York: Routledge.

MacEachren, A. M. (1995). *How Maps Work: Representation, Visualization and Design*. New York: Guilford Press.

Pomerantz, J. R. and Kubovy, M. (1986). Theoretical approaches to perceptual organization. In Boff, K. R., Kaufman, L. & Thomas, J. P. (eds.) *Handbook of Perception and Human Performance*, vol. 2, pp 36.1 36.46. New York: Wiley.

Robinson, A. A. (1952). *The Look of Maps*. Madison, WI: University of Wisconsin Press.

Steinke, T. R. (1987). Eye movement studies in cartography and related fields. *Cartographica* 24(2), 40 73.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review* 55(4), 189 208.

Williams, R. L. (1956). Statistical symbols for maps: Their design and relative values. ONR report. *Report of the Map Laboratory, Yale University*. Newhaven, CT: Defense Technical Information Center.

Map Types

D. Wood, Raleigh, NC, USA
J. B. Krygier, Ohio Wesleyan University, Delaware, OH, USA

Glossary

Applied Maps An early term, used by Eckert in 1925, for the category that would later become thematic maps.

Classification Systematic grouping based on assumed similarity of items.

Particular Maps Maps of some area less than the entire earth, as compared to Universal or General maps.

Reference Maps One of two grand classes of maps (the other being 'thematic maps') distinguished by a 'general-purpose' or focus on a collection of general environmental data (e.g., roads, boundaries, cities, states, etc.).

Thematic Maps One of two grand classes of maps (the other being 'reference maps') distinguished by a 'special-purpose' or focus on one or a few distributions of specific data (usually statistical). These maps are also known as applied maps, special-purpose maps, statistical maps.

Universal or General Maps Scale-based classification, including all maps of the world as a whole, as compared to particular maps.

That the classification of maps is not for the faint of heart was made clear as long ago as 1692 when, in a supplement to *Norte de la Navegación Hallado por el Quadrante de Reducción que Ofrece*, the Basque geographer Antonio Gaztañeta y de Iturribálzaga sorted maps into the following categories: (1) those hidden as state secrets; (2) rolled, not folded; (3) folded; (4) those which flake when handled; (5) those whose colors suppose precepts; (6) imaginary, or genuinely imaginary; (7) blank; (8) those that are oriented; (9) unoriented; (10) those that can be read like a book; (11) those of fusty smell; (12) shared among the faithful; (13) those capable of alarming the meek; and (14) additional copies. Nearly three hundred years later Helen Wallis and Arthur Robinson ventured a more capacious categorization in their *Cartographical Innovations* of 1987. The 191 map types they identified included the following: (1) chorographic; (2) fan; (3) globe, pocket; (4) imaginary; (5) plan, plat; (6) tactile; (7) enclosure; (8) moral statistics; (9) sanitary; (10) tithe; (11) eclipse; (12) magnetic north; (13) wave directions; (14) flow line; (15) point symbol, divided circle; (16) spot height; (17) hand colored; (18) lettering, type inserted; (19) map surface, ceramic; (20) map surface, metal; (21) anaglyptographically reproduced; (22) inked; and (23) atlas, island. One of these two map type classifications is fictional and one is real, and it is difficult to ascertain which is which.

To classify, of course, is human and doubtless there are as many classifications of maps as there are types of maps in the classifications. But human action is always motivated, and during the twentieth century what motivated the dominant classification of maps – the division into two overarching types, the 'general purpose' (or 'reference') map and the 'special purpose' (or 'thematic') map – was the need to isolate and make visible a practice of making small scale, typically statistical maps that could be justified as a subject in a university curriculum. The classification emphasized a division of labor between technicians, who were responsible for the reference (topographic, base) maps, and scholars, who created the thematic (special purpose, applied) maps. Excluded thereby were those who were neither technically trained nor academically educated to make maps, that is, everyone else. This division of maps and their making arose in the nineteenth century as part of an effort to rationalize a scientific cartography in the service of a geography struggling to validate its status as a science. Today this 'great types' classification is withering along with the academic enterprise that called it into being. The terms of the classification are largely meaningless to the vast majority of technically and intellectually competent people making and working with maps outside the cartographic establishment as, indeed, to most of those within it.

The division of maps into types began early in the history of mapmaking, but far more often than not, it was a division based on differences in scale. On the one hand were maps of the world as a whole, that is, 'universal' or 'general' maps. On the other hand were 'particular' maps, that is, maps of continents, regions, countries, or even smaller parts. Here, for instance, from his *Dictionarium Britannicum* of 1730, is Nathan Bailey's definition of a map: "A plain figure, representing the several parts of the surface of the earth, according to the laws of perspective, or it is a projection of the surface of the globe, or part thereof in *plano*, describing the several countries, islands, seas, rivers, with the situation of cities, woods, hills, etc. Universal maps, are such as exhibit the whole surface of the earth, or the two hemispheres. Particular maps, are such as exhibit some particular part or region thereof."

This scale based typology served from the sixteenth century into the nineteenth when mapmaking began to be called upon to support the ambitions of geography to situate itself in the scientific academy. Inspired by the examples, as well as the labors, of Carl Ritter and Alexander von Humboldt, Hermann Berghaus began producing high quality maps of climate, hydrography, vegetation, anthropology, ethnography, and other topics collected in his *Physikalischer Atlas* of 1838, which was later distilled, revised, and redrafted as *The Physical Atlas* by A. K. Johnston in 1848. During the later part of the nineteenth century such 'applied maps' as these were increasingly distinguished from less narrowly focused 'geographic maps' at every scale. In his *On the Nature of Maps and Map Logic* of 1908, Max Eckert refined these map categories, distinguishing geographically 'concrete' maps that "reproduce facts as they exist in nature, such as the distribution of land and water and of heights and depressions" from geographically 'abstract' maps that "present, in cartographic form, the results of scientific induction and deduction and in most cases, can be traced back to the study of the scientist."

In 1925, Eckert expanded on his effort to establish the abstract maps as scientific in a discussion of 'applied maps' in the second volume of his *Die Kartenwissenschaft*: "The applied map design is done at the desk of a scholar, because the practical cartographer has done enough in drawing a perfect base map," Eckert wrote. "Only seldom does the real cartographer proceed to the field of applied cartography. It is generally known that he has other work to do. Moreover he has no time to care about scientific problems and their translation into cartographic form which is a full time occupation, because he is already totally occupied with his manual, but nevertheless scientifically guided work." As Eckert concluded, "The matter of applied map design is the very task of a geographer." Erwin Raisz, in the first textbook of cartography to appear in English, his *General Cartography* of 1938, distinguished between a similar pair of categories that he called 'general' and 'special'. He refined this distinction through a second edition (in 1948), and in a 1962 revision that he called *Principles of Cartography*.

By then, however, Nikolaus Creutzberg had rechristened the 'special' category as 'thematic' (in a paper of 1953), and Raisz now incorporated the new term in his revision: "Maps," Raisz wrote in 1962, "are of many kinds. Perhaps the most important difference is between serial and individual maps. Large scale topographic maps and charts come in sets and are usually made in government offices with highly specialized equipment and broken down to jobs with rather rigid standards. In the second class we have maps often on smaller scale which the individual can design and draw. In the first, the technical training is the more important; in the second, the knowledge of geography and certain ability in graphic expression." This second category Raisz now broke down into (1) charts; (2) thematic or single factor maps; (3) land use maps; (4) city maps; (5) transportation; (6) political and historical maps; (7) maps of the various sciences; (8) maps for illustrations and advertising; and (9) cadastral maps. Marking the growing importance of the thematic category was Eduard Imhof's simultaneous publication of *Thenatische Kartographie*; and soon the 'single factor map' was the subject of Erik Arnberger's *Handbuch der Thematischen Kartographie* in 1966, Werner Witt's *Thematische Kartographie* in 1967 (with its second edition in 1970), and Sylvie Rimbert's *Leçons de Cartographie Thématique* of 1968.

Implicit in this classification was a narrative about the genesis of maps. Initially this had three phases. Raisz, for example, wrote in his 1938 text that: "The process of revealing the Earth's pattern has three phases: The surveyor measures the land, the cartographer collects the measurements and renders them on a map, and the geographer interprets the facts thus displayed." The problem with this version was that it minimized the role of the cartographer, and Arthur Robinson soon collapsed the three phases into two. In *Elements of Cartography* (1953) – which through its six editions would become the defining textbook for Anglo American cartography in the second half of the twentieth century – Robinson reconceived the process as follows: "The entire field of mapmaking is usually thought of as consisting of two distinct phases," he wrote. "The first is concerned with the detailed large scale topographic mapping of the land or charting of the sea. The remaining large proportion of cartographic activity is less clearly defined, being usually thought of merely as smaller scale, special cartography, or simply as 'not' the first mentioned."

That is, Raisz's 'surveying' was aggregated to topographic mapping, and his 'interpretation' of what was soon to be rechristened thematic mapping. "Topographic mappers," Robinson went on, "make maps from field or air survey and are concerned with such things as the shape of the earth, height of sea level, land elevations, and exact and detailed locational information. Generally speaking, this group, which includes the great national survey organizations, national land offices, and most military mapping organizations, makes the basic maps from which the other group starts." This other group does not make maps from surveys but "using the detailed maps, compiles from them the basic data required and then proceeds to add relationships, generalizations, and a host of other kinds of material. To this group belong the geographers, historians, economists, and many others of the social and physical sciences who are seeking to understand and interpret the social and physical complex on the earth's surface."

This version of cartographic genesis actually creates three groups. In the first, of course, are those responsible for topographic mapping. Typically government employees, these work with highly specialized equipment at carefully defined tasks, including surveying, drafting, engraving, and printing. That is, these mapmakers are technicians and laborers. Their adherence to strict standards, however, results in precision and accuracy. The second group uses the first's data to interpret social and physical patterns. These mapmakers are scientists, uni versity people, professionals, and thinkers. However, because their intellectual work is based on the careful labor of the topographers, it inherits the accuracy and precision of these technicians. Everyone else falls into a third group that is neither technically proficient nor educationally equipped to make maps. This renders any map they might make of doubtful value.

Valorized above all by this typology were university cartographers and what was soon universally known as the thematic map. As it brought the thematic map to prominence, the typology also created a novel map type rarely cataloged, yet highly prominent in the literature. This was the base map. The base map was what university cartographers compiled from the technical work of the topographers: "All special purpose maps are made on the foundation of a base map," Robinson wrote in his first edition, where the base map was the subject of an entire chapter. "This base map is compiled first, and the ac curacy with which it is made determines in large part the accuracy of the final map." The base map fails to appear in cartographic typologies, however, because once the university cartographer has performed his interpretation, the base map disappears.

Classifications are systematic segmentations of the world. Ideally, they are consistent, clearly demarcated, and complete; in other words, they obey unique classifi catory principles, consist of mutually exclusive cat egories, and have a slot for everything in their purview. It may be the case that no classification in existence fully satisfies these requirements, but the schemes of Eckert, Raisz, Robinson, Imhof, Arnberger, and the others fall wildly short of the mark. The supporting story of how maps are produced is broadly untrue and historically it is emphatically false. For example, the corpus of maps made prior to the inauguration of large scale topographic mapping was obviously not based on it, nor were the vast majority of later maps based on precedent mapping tra ditions, including urban cadasters, railway maps pro duced by houses like Rand McNally, early highway maps, small scale thematic maps in atlases of the nine teenth and early twentieth centuries, maps of diseases at large and small scales, Sanborn insurance maps, most planning maps, illustrative and advertising maps of all kinds, and so on. Indeed, it is hard to say to what extent even today the origin myth has much validity.

None of these problems, however, mattered at mid century when university cartographers were attempting to justify their positions on university faculties. As stu dents of classification have long observed, among other things, classifications are about struggles for professional authority. Foreclosing one labeling option as they preset others, categories valorize this point of view and silence those. Valorized by the map types constructed by Eckert, Raisz, Robinson, Imhof and the rest, were academic mapmakers like themselves and the maps that they alone made: the thematic maps that were shifted by this aca demic classification from a marginal position in the world of maps to center stage.

Ensuing developments were dramatic. Robinson had not used the word 'thematic' in the first (1953) edition of his text, but Imhof, Arnberger, Witt, and Rimbert had all published their thematic cartography texts by the time Robinson published his third edition (with Randall Sale in 1969). Dispensing with efforts to classify map types ("To attempt to catalog with precision the infinite number of kinds and uses of map is an impossible task"), Robinson immediately launched into a history of car tography. Where in the earlier editions this history had moved from 'The Beginnings of Cartography' through 'The Early Modern Period' to 'Twentieth Century Car tography', in the third edition it moved from 'The Be ginnings of Cartography' through 'The Dark Ages', 'The Renaissance', and 'The Early Modern Period' to 'The Rise of Thematic Cartography'. "In addition to the nautical chart and the topographic map," Robinson now wrote, "a third great class, the thematic map, was added to the repertoire of cartography by the early nineteenth century." Noting that in the past the thematic map had been called the 'special purpose map', Robinson claimed that, "Its main objective is specifically to com municate geographic concepts such as the distribution of densities, relative magnitudes, gradients, spatial rela tionships, movements, and all the myriad interrelation ships and aspects among the distributional characteristics of the earth's phenomena." At this point in the text Robinson recapitulated the substance of his earlier 'two phase' description of the field, but when he reached the second, dependent phase, he added, "The other category, which includes thematic cartography … ."

By the time of his text's fifth edition, with Sale, Joel Morrison, and Phillip Muehrcke (in 1984), the moves Robinson had made in the third edition had been so lidified. Among other things, Imhof's textbook had gone into a second edition (in 1972), Arnberger had sup plemented his *Handbuch* with his *Thematische Kartographie* (in 1977), Barbara Petchenik had provided psychological justification for the claims of thematic mappers in her *From Place to Space: The Psychological Achievement of Thematic Mapping* (in 1979), and Robinson himself had published *Early Thematic Mapping in the History of Cartography* (in 1982). This last meant that a map type that had existed for scarcely more than a generation now had a history which, in a burst of retrospective reclassification, rele gated most of the history of mapmaking to 'The Devel opment of the Base Map', while it hitched the history of thematic mapping to the prestigious history of science. Finally, the first edition of Borden Dent's *Principles of Thematic Map Design* was to appear the following year (1985) with its definitive opening: "Maps are graphic representations of the cultural and physical environ ment," Dent intoned. "Two subclasses of maps exist: general purpose (reference) maps and thematic maps. This text concerns the design of the thematic map."

Can it be surprising that Robinson now felt em powered to risk a classification of maps? While con tinuing to acknowledge that the variety of maps was unlimited, there were, he now ventured, "recognizable groupings of objectives and uses for maps, which permit us to catalogue them to some degree." He discussed these under three headings: scale, function, and subject. Scale varied, Robinson noted; and there was no limit to the possible subjects of maps; but when it came to function, there were three classes: general maps, thematic maps, and charts. Since Robinson had now isolated scale as an independent factor, he could be more subtle than he had been even as recently as 1969. Large scale general maps are still usually topographic, but Robinson acknowledged that much larger scale maps are often required by en gineers, and that small scale general maps of states, countries, and continents also exist. General maps are typified by the portrayal of "things such as roads, settlements, boundaries, watercourses, elevations, coast lines, and bodies of water." Thematic maps, which now may be large, as well as small scale, "concentrate on the spatial variations of the form of a single attribute, or the relationship among several." Charts remained segregated in a separate class to serve the needs of nautical and aeronautical navigation.

The triumphant progress of the thematic map con tinued. Arnberger's *Thematische Kartographie* went into a second edition (in 1987), and Dent's *Principles of Thematic Map Design*, now called *Cartography: Thematic Map Design*, went into a second (1990), third (1993), fourth (1996), and fifth edition (1999). In 1992, Judith Tyner published *Introduction to Thematic Map Design*. Tyner's slant on map classification was individual. While acknowledging that there were three classes, Tyner made these out to be general purpose, special purpose, and thematic. General purpose maps, she proposed, "do not emphasize one type of feature over another," while special purpose maps "are created for a very specific type of user. Geologic, soil, and cadastral maps are included here," along with all navi gational maps. Thematic maps, Tyner allowed, "have been called a variety of names (special subject, statistical, distribution, and data maps) but the term 'thematic' is now generally accepted." She stressed a point subdued since Raisz: "Although general purpose and special purpose maps are produced by cartographic agencies, institutions, and firms (frequently by teams of specialists such as surveyors, photogrammatrists, designers, and cartographers), a thematic map, even if produced by a similar agency, is probably the work of only one or two people."

There were, of course, dissenting voices. In his *Cartographic Design and Production* (of 1973), J. S. Keates noted that the "expression 'thematic' does suggest that the subject matter deals with a particular theme or subject, but as this is true of all maps it is not particularly helpful in determining a category." In his later *Under standing Maps* (1982, with a second edition in 1996), Keates also argued that cartography had arbitrarily lim ited its scope with its emphasis on the thematic map. John Campbell acknowledged the reference/thematic dis tinction in his *Introductory Cartography* (of 1984) but he also observed that the "problem with dividing maps into reference and thematic types is that there is no clear cut dividing line between the two." Phillip Gersmehl echoed this sentiment in his *The Language of Maps* (1991) when he noted that, "The distinction between reference and the matic is thus more than a little blurry."

Despite such blurring and polite internal discussions about things like Tyner's special purpose maps, the orthodoxy of the reference/thematic distinction, and the history and the production hierarchy it entailed (in cluding cartography positions on university faculties), seemed secure as the 1990s opened; but in fact, heterodox arguments had been gaining adherents throughout the 1980s, and these came from a confounding number of directions. Critics such as Denis Wood (especially in his *The Power of Maps* of 1992) and Brian Harley (in a number of essays collected in 2001 as *The New Nature of Maps*) called into question the relevance of the classification to anything but the authority it granted academic cartog raphers, and Wood in particular stressed its inadequacy as a formal classification. Others, including Robert Rund strom (in articles he began publishing in 1990) and Doug Aberley (in his *Boundaries of Home: Mapping for Local Em powerment* of 1993), attacked the marginalization of non professional mappers. They insisted that everyone was capable of making maps, and they theorized a growing body of mapping by First Peoples, bioregional activists, English parishes, and others, that has matured into a fully fledged, counter mapping movement. Artists, too, who had worked with maps for decades, began to produce art maps in astonishing numbers. Many of these were ex plicitly constructed to call into question the authority of the received mapping tradition, and to deny, confound, ignore, or contest received map types.

At the same time the academic discipline built around the handcrafting of thematic maps was being ever more rapidly consumed by what gradually became known as Geographic Information Science. GIS software, particu larly once it spread to personal computers and the internet, made it possible for anyone to make maps (or at least the map types the software allowed you to make).

Since the software contained the basic intellectual knowledge of the academic cartographer (e.g., the func tions and defaults involved in thematic mapping) and eased the technical skills required to make maps, it fur ther empowered mapmakers outside the profession who, without a need to justify a position in the academy, found little utility in the reference/thematic distinction.

But the distinction is losing its force even within what remains of cartography. In 1999, Terry Slocum published *Thematic Cartography and Visualization*, sure to be the last in the lineage of comprehensive thematic cartography texts that was inaugurated with Imhof's *Thenatische Kar tographie* in 1962. In Slocum's second edition, retitled *Thematic Cartography and Geographic Visualization* (with Robert McMaster, Fritz Kessler, and Hugh Howard, 2005), Slocum pretty much dissolves the distinction when he admits that, "Although cartographers commonly dis tinguish between general reference and thematic maps, they do so largely for the convenience of categorizing maps. The general reference map also can be viewed as a thematic map in which multiple attributes are displayed simultaneously; thus, the general reference map can be termed a multivariate thematic map. Furthermore, al though the major emphasis of general reference maps is on 'location' of spatial phenomena, they can also portray the 'spatial pattern' of a particular attribute (e.g., the pattern of drainage on a USGS topographic sheet)." Cynthia Brewer's *Designing Better Maps* (2005) rarely uses the term 'thematic' and never defines it, while John Krygier and Denis Wood intentionally excluded the terms 'thematic', 'reference', and 'base map' from their *Making Maps* (2005).

To classify remains human, however, and the classi fication of maps endures unabatingly. The Cambridge Historical Commission files maps under the following categories: (1) Raisz Maps, (2) First Period Reconstructed Maps, (3) Harvard Maps, (4) Insurance Maps, (5) Library of Congress Maps, (6) Sanborn Maps, (7) USGS Quad Maps, (8) Ward Boundary Maps, (9) Transit Maps, and (10) Miscellaneous and Correspondence Maps. In *Map ping Hacks: Tips and Tools for Electronic Cartography* (2005), which details an array of techniques, tricks, and hacks that anyone can try, authors Schuyler Erle, Rich Gibson, and Jo Walsh sort mapping into nine categories: (1) mapping your life, (2) your neighborhood, (3) your world, (4) mapping on the web, (5) mapping with gadgets, (6) mapping on your desktop, (7) names and places, (8) building the geospatial web, and (9) mapping with other people, a classification that reflects the burgeoning interest in mapping outside academic cartography and its growing relevance to everyone's life. In their *Else/Where: Mapping* (2006), Janet Abrams and Peter Hall attend to the diverse creative community working with maps, which they divide into four classes: (1) mapping net works, (2) mapping conversations, (3) mapping terri tories, and (4) mapping mapping. Their typology embraces the most traditional of maps as well as many far beyond anything academic cartographers would agree were maps. Not only are such distinctions as that be tween thematic and reference maps missing from these books, so is any reference to the academic cartography literature. Except as it is embedded in GIS software, it is almost as though the world of mapmaking dreamed by Eckert and built by Raisz, Robinson, Imhof, Arnberger, and others, had never even existed.

The future proposes a further relaxation of typolo gical rigor. If anything defines early twenty first century mapping it is surely the Google Map mashup. A mashup combines content from multiple sources into a single hybrid. While maps have always been mashups (tradi tionally known as compilations), technology such as Google's application programming interface allows just about anyone to create maps from diverse data sources created by themselves and others. As such mapping technology continues to evolve and grows in accessibility, and as more people make increasingly sophisticated and diverse maps, new map types and new classifications of map types are sure to follow.

Mapping Agencies

C. Perkins, University of Manchester, Manchester, UK

Glossary

Cartography Practices, processes, and discipline concerned with transforming and representing a real world in a mapped form.

Framework Data Geographical template upon which thematic data may be draped.

National Mapping Agency Organization with responsibility for collecting and maintaining framework spatial data and mapping relating to a nation-state.

National Spatial Data Infrastructure A system for acquiring, processing, storing, distributing, and improving the use of geospatial data encompassing people, technology, and policy standards.

Introduction

The modern history of cartography is strongly associated with the rise of specialist mapping agencies concerned with the creation and dissemination of different kinds of maps and spatial data. In the early modern period in Western Europe mapping emerged as an endeavor at once individual and corporate – individual in the sense that named cartographers and surveyors were able to survey, draft, and publish their own maps – but corporate because of the increasing complexity of mapping processes and pressures toward the development of specialist mapping skills, that were beyond the capability of any single individual. Social, economic, and political changes encouraged a trend away from individual mapping activity, and toward a set of mapping practices increasingly carried out by specialist agencies. By the nineteenth century mapmaking had become something that organizations did. Individuals might deploy maps, using them as practical spatial devices, but it was agencies that made these products. Recent technological and social changes are beginning to alter the relative significance of individual and agency: people and organizations are now increasingly making their own maps once again, instead of relying upon a generic product made by an external agency.

This article explores this changing institutional context, describing the dimensions across which mapping agencies vary. It highlights a case study of contemporary institutional forms in which mapping activity is being carried out, in order to illustrate the significant impact of local and time specific factors. By stressing practices and local processes, competing roles of mapping can be explored, and the significance of an institutional approach to cartography can be assessed.

Charles Handy has argued that all organizations work as micro societies, with varying cultures, motivations, and decision making structures, in which individuals exercise different roles, subject to different power relations, and in which leadership propels change. The direction and trajectory of change is almost impossible to predict since an agency collaborates or competes with many other organizations, operating in a wider social context. Actors in organizations operate in a complex network. All organizations operate by deploying formal rules that regulate individual behavior, but it is often more symbolic rituals that govern everyday practice. Technologies involving machines, methods, and knowledge are deployed to advance agency goals. Economic relations govern successful operation: labor is organized into a particular mode of production. Van den Hoonaard and Nichols have suggested that a good way of understanding some of these interactions and processes of change as they relate to mapping agencies is to think in terms of different 'map worlds' – social contexts, involving collaborative behavior, a division of labor, particular technologies, conventional practices, and different actors and agents, which together predominate in different places and at different times throughout the history of cartography.

So, to understand the significance of a mapping agency requires a multidisciplinary approach drawing upon economics, sociology, political science, management theory, psychology, and sociology: the notion of map worlds draws on all of these. There are, however, often rather specific characteristics shared by mapping agencies that stem from the nature of maps and spatial data as media. Maps and spatial data are a form of information and as Ian Masser has shown this has interesting policy implications. Mapping agencies may sell mapping as a commodity, or license access to the products they make, but the map base also serves as a resource like capital, land, or labor, which has an ongoing value for a customer, and may be deployed by an agency to add value to its own processes. The mapping itself is the most significant asset for many mapping agencies. Mapping also provides an infrastructure, so that unlike other commodities it necessarily supports essential or progressive social activities. So other agencies, like the nation state or individuals, rely upon mapping even if they do not directly acquire it. Given these characteristics and the need for a multidisciplinary approach we can now turn to things that influence the nature of mapping organizations, and how these together make different map worlds.

Dimensions of Organizational Difference

In an extended overview published at the start in 2000, Parry and Perkins identified the 2000 most important agencies in the world that were then actively involved in the production of mapping. In addition there are of course many other organizations that support the creation of coded spatial information, without being directly involved in its production, as well as agencies that store mapping, encourage the use of mapping, or represent the interests of those who map. A process of technological change continues to democratize mapping activity and widen the range of agencies that make mapping. There are significant regional disparities in this process. Most mapping takes place in the developed world and most agencies here play increasingly specialist roles: in 1996 Perkins and Parry identified 246 mapping agencies active in the UK alone.

Some kinds of mapping organizations are easy to classify, but many do not conform to obvious stereotypes. The nature of the agencies changes over time reflecting local culture, politics, and technology. **Table 1** highlights some of the many dimensions across which these agencies vary.

'The legal context' in which an organization operates has significant impacts. Workers' rights, minimum wages, health and safety, patents, and trademarks, all limit the degrees of freedom an agency is able to exercise. Land use planning legislation may restrict expansion opportunities. Competition legislation may impact on pricing policies. Legislation about individual rights to information may limit access to information by protecting copyright, or it may protect individuals from organizational intrusion, through data protection legislation. Copying of derived information from earlier surveys was the mainstay of the era of copper engraved printed mapping in the seventeenth and eighteenth centuries – a more sophisticated protection now hinders reuse, and mapping agencies extract significant royalty income from those who use their data. Data protection legislation may also limit mapping activity: data vendors may wish to combine data sets, but if this reveals individual identities then combination may be illegal.

On the other hand, legislation may place certain government information in the public domain, for example, in the case of American citizens' rights of access to federal data, and in many Western liberal democracies where 'freedom of information' legislation increasingly guarantees citizens rights to information that might otherwise be deemed secret. Many governments, however, enact legislation to protect official secrets: mapping is often regarded as a powerful strategic asset to be guarded with care. The significant differences in charging regimes between American and European official mapping agencies are a direct result of different legislative contexts. In the UK copyright legislation protects government data, and the Official Secrets Act limits public access rights: spatial data are expensive, and it has been argued this discourages commercial applications. In contrast in the USA a major secondary sector of specialist mapping firms adds value to heavily subsidized federal data sets released into the public domain.

Other legislation regulates particular kinds of mapping. Official organizations are established by statute. Legislation on property transfer and its registration strongly influences the form of cadastral mapping activity. Liability legislation is a powerful influence on the working practices of charting agencies, which need to depict navigational hazards, as accurately and frequently as is possible.

'The economic context' in which an agency operates is fundamental. At a national scale the relative prosperity of a nation is a good indicator of the health of its mapping agencies, and most of Africa even today remains very badly served by indigenous mapping capacity, as a consequence of debt burdens, an inadequately trained local workforce, and inadequate infrastructure. Sporadic aid projects and uncertain funding lead to institutional uncertainty, and fewer major mapping programs. Many mapping agencies in the developed world are also not in direct control of their budgets, and must ask a parent body for resources to carry out mapping. Agencies across Western Europe, North America, and Australasia

Table 1 Mapping agencies: attributes

Cultural context
Legal context
Copyright
Freedom of information
Secrecy
Privacy
Economic context
Political context and governance
Military/civilian
Subordinate/independent
Technological context
Product focus
Area
Scale
Theme
Specialist/generalist
Nature of agency
Size
Longevity
Size and adaptability
Organizational framework

benefited from Keynesian economic policies in the aftermath of World War II, in which their expenditure was underwritten by the state. Official mapping was seen as underpinning national administration and progress. These policies, however, were soon abandoned under neoliberal economic orthodoxies: economic pressure on the public sector and efficiency gains are now the norm. There has been a clear trend in the period since the mid 1970s for official agencies to gain more financial freedom, with a consequent increase in the amount of contracting out of activities such as digitizing, to areas with lower labor costs, as well as an increased amount of product franchising and joint ventures to minimize risk.

Smaller independent mapping consultancies have benefited from these trends, but competitive tendering for project work characterizes many of their operations. Decisions about pricing policies, and whether maps or mapping data should be sold, or licensed for use, are not always in the hands of any mapping agency. Nor is the level of profit.

'The technological context' in which a mapping agency operates is also fundamental. Major shifts from one mode of production to another characterize the history of cartography – technologies of data collection, mass production, data manipulation, and dissemination alter the nature of the mapping agency, its products, working practices, and users. Theocratic worldviews reliant on manuscript mapping fostered an elite image in Medieval times, with limited mass impact. The inside world of map ateliers copying earlier published maps allowed an esthetic image to be widely diffused throughout Europe in the seventeenth and eighteenth centuries. Scientific mapping practices grew in significance in the seventeenth century and encouraged the development of nineteenth century field practices of triangulation for military survey. Photogrammetric sourcing of data from aerial coverage for mass topographic publication in the mid twentieth century, shifted the focus of mapmaking back inside, and facilitated the operation of mass produced official mapping, in factory like production lines. Widespread adoption of digital production, telecommunications, global positioning, and the web instigated radical changes and a much more flexible organizational context, in the period from the last two decades of the twentieth century. These technologies have been adopted in changing economic, political, and social contexts: innovation is uneven across time and space.

The 'governance and political context' of these organizations often plays a crucial role in their operations. Few mapping agencies have been autonomous: parent bodies dictate policy. Official mapping of different nations is almost always the responsibility of a national mapping agency charged with responsibility for maintaining geodetic control, and compiling mapping that reflects officially specified needs. These agencies emerged at different dates in different nations – often tracing their histories back to a military past, but usually anchored in a requirement to carry out topographic surveys. National mapping agencies reflect local strategic concerns. At first, official mapping served the requirements of military forces to prosecute warfare or defend the nation, and of colonial administrators to manage land and plan resource exploitation. In the period after World War II there was a general retreat from military control of national mapping agencies, with a growing civilian, economic, and administrative rationale. More recent trends in the developed world have seen official mapping agencies gain an increased freedom from state control, reduced support from public sources, and increasingly commercial rationales. Indeed, commercial governance has characterized much of the history of mapping agencies. In the period before the nation state became involved the links between commerce and mapping were particularly important – money was made by mapping, and a map trade was developed that represented the interests of the wealthy.

In addition to these wider social concerns the nature of the mapping agency itself is important. The most obvious dimension concerns what they map. 'Different kinds of maps' often characterize a particular institution: geodetic control, topographic mapping, or framework data is maintained by national mapping agencies; cadastral mapping by cadastral surveys; hydrographic and bathymetric charts by hydrographic agencies; census mapping by statistical offices; geological mapping by Earth science agencies; environmental mapping by environmental agencies, etc. Commercial map publishers release tourist, town, and motoring mapping. Lists of these organizations are available in *World Mapping Today* (2000) and on many websites.

In the face of technological and social change these formerly discrete categories are, however, no longer either fixed or separate. Precise responsibilities vary. Some national mapping agencies also carry out hydrographic charting; others produce land use data sets in addition to topographic data; some specialize in maintaining only essential national data, ignoring more commercial activities; some carry out thematic mapping and publish or maintain a national atlas; others have a narrowly defined topographic role.

The 'scale and resolution' at which mapping is published also varies, and reflects local historical precedent. An agency's spatial reach and external collaboration is also important, in particular because of impact on 'standardization and interoperability'. Some organizations map the world, others their nation, others a region in a federal nation, others a local authority. Earlier attempts at systematic global mapping that depended upon collaboration between agencies foundered on inadequate communication and competing ideologies, but more recent technological progress, and increasing global concerns, has seen the creation of many global databases. In the wake of these globalizing tendencies has come increasing pressure toward standardization, and inter operability. For example, the European Union (EU) Infrastructure for Spatial Information in Europe (INSPIRE) directive, enacted in May 2007, seeks to trigger the creation of a European spatial information infrastructure to allow users to identify and access spatial information from a wide range of sources, in an inter operable way, for a variety of uses. INSPIRE places the onus on states to deliver National Spatial Data Infra structures (NSDIs) that are aligned across state bound aries, and in so doing inevitably reduces local autonomy.

The ‘frequency of product update’ is a further im portant consideration and also depends upon resources, precedent, and corporate strategy. Real time update in relation to the mapping of traffic flows, contrasts strongly with attributes of databases that change much less fre quently, such as relief. Some organizations focus on ‘one part of the production process’, concentrating upon, for example, land surveying, or the wholesaling of finished products, or the provision of a piece of visualization software. Others provide a comprehensive service from survey through to publication for a particular spatial remit.

Almost all mapping agencies invest significant re sources in their specialist data, and given this investment, products are unlikely to be completely abandoned unless there is a very good reason. It can be hard to escape from the tram lines of past practice: development of the large scale digital Master Map in the UK, for example, reflects the late nineteenth and twentieth century history of large scale national mapping, as well as innovative ap plications of digital mapping. Agencies have a varying ‘longevity’. Some national mapping agencies date back to the eighteenth century: Ordnance Survey can trace its history to the establishment of a map office by the Board of Ordnance in the Tower of London in 1791. Others are considerably more ephemeral, with short corporate lives. Organizational longevity can be an indicator of carto graphic quality: official British mapping nowadays reflects a sustained investment by the nation in its official map ping over the previous century and a half. Even if a product line is consistently maintained the agency re sponsible for its creation may well change its name many times and gain or lose mapping responsibilities. Com mercial mapmaking illustrates a history characterized by a complex genealogy of takeovers and market failures, in which institutional flux predominates over longevity.

The ‘size of the agency’ can make a significant dif ference to the nature of its operations. After the Russian revolution (1917) giant monolithic Russian mapping agencies standardized practice across the huge geo graphical extent of the former Soviet Union, and sought to establish global mapping standards, that a smaller agency would have found difficult to achieve. In contrast, smaller agencies are more able to respond to change, but are more vulnerable to external pressures.

‘Institutional organization’ is another significant factor. Map ateliers producing printed mapping in Western Europe in the seventeenth and eighteenth centuries were often organized as family dynasties, in which printing plates passed from one generation to another, and in which craft skills were important. On the other hand, official mapping in the mid twentieth century relied strongly on a functional division of labor, in which a large and relatively unskilled in house workforce was mar shaled by managers to deliver standardized products; surveyors in the field collected information; draftsmen at the center created a map; printers set up the plates to publish the output; customer relations were often out sourced to agents, etc. Agencies adopted Taylorist ap proaches to labor – ‘rational’ mass production maximized efficiency in the ‘map factories’ of national mapping agencies. These management structures distanced work ers from completed products, and map users from spe cifications. More recent shifts have seen reductions in the numbers of unskilled people employed in mapping agencies; a trend toward flatter management structures with greater accountability; and market led organizations that are much more responsive to customer demand.

The Flexible Spatial Information Economy

To illustrate how these different factors work, this section provides a case study of the contemporary operation of mapping agencies in the developed world. Over a decade has passed since the last detailed overview of the chan ging role of mapping agencies, described in the first chapter of *Mapping the UK: Maps and Spatial Data for the 21st Century*. In this short period the pace of change has increased, and there has been a considerable turnover of organizations. **Table 2** shows some of the kinds of mapping agencies, highlighting newly emerging roles and organizations. The old certainties of surveying, com piling, producing, remaking, selling, storing, buying, and using the paper map, in separate contexts, and often in separate agencies, are disappearing. Old organizations have had to change or go out of business; new agencies with new roles have emerged. Key factors in this dyna mism have been the blurring of boundaries between agencies, more flexible working practices, and a growing internationalization of service provision.

The nature of mapping has changed radically over the last 30 years, in ways that have altered how mapping agencies operate. Revolutions in telecommunications and personal computing have complemented new technolo gies for data acquisition, at a time of economic liberal ization, globalization, and increasing social disparities.

Table 2 Change in mapping organizations

Typical organizations in 1975	*Typical organizations in 2007*
Official mapping agency	*Official mapping agency*
Production-driven, functionally organized, public-funded, hard-copy maps, e.g., Ordnance Survey	Market-oriented, flexible, public private partnerships, digital framework data, e.g., Ordnance Survey
Commercial map publisher	*Portals*
Specialist market-oriented compilation and publication of derived hard-copy maps and atlases, e.g., Phillips	Free web-served access to third-party mapping, with capability for user design, e.g., Google Earth
Map shop	*Value-added retailers*
Hard-copy map sales, e.g., Stanfords	Multiple data sets customized for targeted market, e.g., Landmark Information Group
Map library	*Spatial data clearinghouse*
Classified hard-copy collections of published mapping, e.g., British Library	Online access to spatial resources with search engine access, e.g., EDINA Digimap
Cartographic service company	*Cartographic service companies*
Survey companies, drafting companies, freelance cartographers, e.g., Fairy Surveys	Mapping consultancies offering turn-key services, e.g., Lovell Johns
Cartographic society	*Social networking sites*
Regulated membership, meetings and conferences, training, e.g., British Cartographic Society	Blogs, e-mail listings, virtual communities, e.g., The Map Room

There has been an unprecedented profusion of data sources: civilian high resolution digital image sources are now available which compete with mapping data sets formerly tightly controlled by national governments. A decade ago this was an unrealized possibility. Global positioning facilitates unprecedented geodetic accuracy, and also encourages the creation of new spatial information sources, created as a by product of other systems, instead of being designed as mapping *per se*. Geographic information systems (GISs) have matured, offering carefully targeted services that increasingly supplant formerly generic mapping. Information is now shared over the World Wide Web, often cutting out the need for acquiring data from a supplier, with mapping increasingly targeted toward the specific requirements of identified user groups. Location based mapping served *in situ* on mobile devices threatens to further revolutionize the ways we deploy mapping.

Consumers in the developed world now frequently make their own maps, deploying data served over the web, and merged with their own information, to create targeted and closely focused products. Complex interaction, visualization, animated mapping, data sharing, and multiple representations across media are replacing formerly fixed specifications and hard copy mapping. New products with flexible specifications are being served to meet market need, with digital products sourced from multiple origins. These trends were all possible in 1996, but have all gathered pace since then.

It is unsurprising that mapping agencies have responded to these changes. A formerly labor intensive industry in which little changed and where timescale was largely irrelevant, has changed into one relying on capital and technology intensive systems, where time and money dictate strategy. This technology is increasingly expensive, targets are now monitored, and performance constantly audited against quantifiable outcomes. The market rather than government now dictates priorities for many mapping agencies, and cost recovery is replacing state financial support. The former monopolistic, producer led mapping sector is increasingly facing competition from new suppliers in the private sector, and often from new global players. Official agencies no longer control quality – they regulate and seek to manage risk. In the developed world, older organizations still survive, but often with radically altered roles. Economic pressures increasingly encourage agencies to outsource tasks that would have been carried out in house in the past – from 2005, for example, the United States Geological Survey and the National Atlas program in the USA have increasingly been forced to consider which aspects of their mapping might be better achieved in the private sector, leading to concerns over whether national goals might get sacrificed in a drive for greater economy. Collaboration and partnership dictate where authority and control ruled in the past.

Military uses of spatial data continue to drive developments: new Western mapping of Afghanistan and Iraq reflects the continuing power of military demand. Strategic roles for mapping have risen in significance in the aftermath of 9/11, with increased calls for real time mapping surveillance of critical infrastructure, and for emergency planning. National mapping agencies, whose missions underpin successful operation of the nation state, continue to be significant, despite increasing commercialization. They still provide framework data, but in most developed countries now maintain digital spatial databases, instead of fixed format maps. Mapping data are licensed for outside use and served over the web from digital clearinghouses. NSDIs, facilitating access to spatial information, are growing in importance. Over 80 NSDIs existed by 2004, often bringing together data sets that were formerly separate and disseminating these through web served data clearinghouses. Statistical agencies still disseminate socioeconomic data, albeit often in digital form over the web, while Earth science agencies increasingly serve a research role, in which systematic survey has been replaced by targeted support for minerals interests or environmental planning.

An increasing specialization has taken place in the world of commercial mapmaking, with ownership be coming more concentrated, at the same time as newly specialist services focused on niche markets deliver new products. Official agencies collaborate as partners with value added retailers. For almost every sector of the economy there are now competing and carefully targeted specialist mapping products – a recent article about golf mapping, for example, identified 85 different companies active in a highly specialized field. Specialist agencies bring together multiple data sets from different orga nizations to target these markets.

Agencies selling maps and spatial data have also changed. The map shop still survives but is now com plemented by specialist web based retailers, who have been able to develop global rather than local markets. The web allows publishers to reduce dissemination costs and as a consequence more producers are selling digital or hard copy products themselves than was the case in the mid 1990s. A new kind of web based distribution service has grown apace in this period, allowing users to search for and download targeted data from specialist spatial data warehouses.

Desktop mapping software and web served mapping make it much easier to cut out the services of a cartog rapher and membership of cartographic societies has fallen. Instead, new virtual communities on the web share concerns about mapping in blogs, web pages, podcasts, and e mail groups.

The most significant new players in the global mapping market, however, are Internet portals offering democratized, high resolution mapping and image data from user friendly interfaces. The funding model for services such as Google Maps and Google Earth, Windows Live, Yahoo Maps, MapQuest, or MultiMap lies in advertising revenue rather than in direct sales of spatial products or in the licensing of data. They target everyday mapping needs for finding business or postal locations, coupled with simple route planning and overlay capability in systems that offer a cut down GIS capability to people in the street for the first time. Mapping backgrounds are sourced from many different national and global players, such as National Aero nautics and Space Administration (NASA), Digital Globe, and TeleAtlas. Google, Microsoft, and Yahoo have each released the application programming inter face (API) code for these services, encouraging a pro fusion of locally created, but web shared 'mash ups', that combine other data with the high resolution image or map backdrop served from the portal. National mapping agencies have also started to release the API code, for example, from Ordnance Survey in its OpenSpace product. The paradox of Internet mapping portals however, is that despite local empowerment, the image and map backdrop itself is increasingly a global product, under the control of significant multinational agencies.

Conclusions

The future of mapping agencies is uncertain. One future vision might see a largely decentralized set of agencies, serving multiple and customized layers of spatial data over the web, through a series of competing but targeted web portals, accessed by context aware mobile devices, offering the potential of ubiquitous mapping, in which a user's position is always on the map. Data in this scenario may in some respects be shared but remain in the control of distant and powerful agencies with agendas of their own. Alternative open source projects offer a rather different organizational vision of the future. Loose web based collaboration and individual sharing of community owned data are already growing apace in projects such as OpenStreetMap. Collaborative mapping following a Wikipedia like model deploys technologies to develop and share web based mapping, instead of making maps for private profit. Whatever configuration of agen cies emerges to meet this need, we can be certain that local circumstances at particular times will be critical in determining trajectories of change.

See also: Mapping, Commercial; Mapping, Distributed; Maps.

Further Reading

Handy, C. (1993). *Understanding Organizations* (4th edn.). London: Penguin Books.
Masser, I. (1998). *Governments and Geographic Information*. London: Taylor and Francis.
McHaffie, P. H. (1995). Manufacturing metaphors: Public cartography, the market and democracy. In Pickles, J. (ed.) *Ground Truth*, pp 113 129. New York: Guilford Press.
Parry, R. B. and Perkins, C. R. (2000). *World Mapping Today*. London: Bowker Saur.
Pedley, M. S. (2005). *The Commerce of Cartography*. Chicago, IL: University of Chicago Press.
Perkins, C. R. and Parry, R. B. (1996). *Mapping the UK: Maps and Spatial Data for the 21st Century*. London: Bowker Saur.
Pickles, J. (2004). *A History of Spaces: Mapping Cartographic Reason and the Over Coded World*. London: Routledge.
Rhind, D. (ed.) (1997). *Framework for the World*. Cambridge: GeoInformation International.
Van Hoonaard, W. C. and Nichols, S. (1999). Map worlds: A conceptual framework for the study of gender and cartography. *Proceedings of the International Cartographic Association Conference*, *Ottawa*.

Relevant Websites

http://www.cambridgeconference.com
Conference Proceedings Thursday 19 July, 2007 Conference Information, Cambridge Conference (Provides the Most Useful

Current Overview of Papers about the Changing Role of National Mapping Agencies).

http://www.eurogeographics.org
EuroGeographics (Provides Useful Comparative Data about Every National Mapping Agency in Europe).

http://www.usgs.gov
Maps, Imagery, and Publications, US Geological Survey (Gives Access to American Official Mapping Activities).

http://www.openstreetmap.org
OpenStreetMap is a Free Editable Map of the Whole World.

http://www.ordnancesurvey.co.uk/oswebsite/
Ordnance Survey Great Britain's National Mapping Agency (Is Probably the Richest Website for a Contemporary National Mapping Agency).

http://whc.unesco.org/en/mapagencies
The Most Complete Online Listing with Links to the Contact Details of National Mapping Agencies.

Mapping, Commercial

B. Hunt, Santa Barbara, CA, USA

Glossary

Custom Cartography It is the mapping produced by a commercial enterprise for a client and not for production and distribution into the marketplace.

In-Car Navigation System It is an electronic device offering driving directions using a GPS system. It can be installed by the auto-manufacturer or sold as an aftermarket product.

Large-Scale Mapping It is a mapping which shows great detail in a small area, for example, street maps.

Professional Quality Street Atlas These are some published atlases that are created for professional users, such as emergency services personnel, real estate agents, and delivery people. Professional quality products are distinguished from mass-market products, which do not have the same level of attention to data quality and are designed for sale to a less discerning public.

Small-Scale Mapping It is the mapping which covers a wide area but offers less detail, for example, a map of a continent.

Commercial Mapping

Hundreds of companies worldwide produce and publish maps and atlases. Some are large companies with hun dreds of employees, but most are very small businesses or just individuals. Together, they comprise an industry selling between $3 and $5 billion maps. This article re views commercial map publishing in the world today.

Maps for Every Purpose

Often, people assume that maps are published to go somewhere, to travel to a distant destination, or to find an address. While most commercial maps are used for navigation, they are also used in teaching, for decoration, and for planning purposes. There are many professional uses too, such as to survey rights of way, record land ownership, establish sales and marketing territories, and strategize military maneuvers. The wide range of uses results in a variety of map products.

Road and street maps and atlases are the most com mon and basic commercial mapping sold worldwide. Large scale plans for urban areas and smaller scale maps of countries and regions often focus on the network of highways and streets used by vehicles, the dominant mode of transportation worldwide. City street maps allow users to find an address and navigate the city, while smaller scale road maps help users plan routes through the countryside or from one city to another. About half of the production from commercial map publishers is for use by drivers.

Studies of map users show that the fear of getting lost worries many drivers. It is second only to their fear of accidents, according to one survey. Many people feel that the presence of a map in their vehicle will help prevent them from getting lost or guide them out of danger should they find themselves in unfamiliar surroundings. Sadly, many people cannot read the maps they have, for they lack the skills and understanding to use the tool properly. Nonetheless, maps are often purchased for use as a talisman to protect the driver from navigation errors.

Aeronautical and nautical charts, the primary navi gation mapping for the nonautomotive markets, are also an important component of the mapping industry. Jep pesen, a subsidiary of Boeing, is the world's leading commercial publisher of aeronautical charts. The com pany is based near Denver, Colorado. Nautical charts, however, are primarily produced by government agen cies. The National Oceanic and Atmospheric Adminis tration (NOAA) is the primary producer of nautical charts in the US. The UK Hydrographic Office serves this function in Britain. Smaller commercial publishers, such as Imray, Laurie, Norie and Wilson, publish nautical charts and books where opportunities remain in the market.

The remainder of the commercial map industry is divided into many smaller markets, publishing products for a wide array of purposes. For example, maps are produced to interpret historical events, to aid in planning a university education, to project an image of world citizenship, or to teach children about their world. Folded paper maps of battlefields help readers of history understand the decisions taken by soldiers and their leaders, providing an insight into the cause of the result. A wall map showing the thousands of colleges and uni versities in the United States hangs on the wall of school guidance offices, helping counselors explain the array of choices for graduating high school students. Wall maps of the world are used for several reasons, allowing a busi ness or individual to decorate using an object that con veys their interest in the wider world, and conveying their citizenship or interest in environmental concerns. Parents also use them to encourage their children to learn about the world.

Types of Commercial Mapping Companies: Size and Function

Commercial map publishers vary in size from just one person to companies with hundreds of people. However, worldwide, all are privately held companies. None are publicly held, which means that none have shares traded on a stock exchange. Therefore, there is no information about total revenues or profits. Based on private con versations with departing employees during the past 20 years, though, some reasonable guesses can be made. The largest commercial map publisher is probably Rand McNally in the USA with total sales around $250 million, mostly from the sale of printed paper maps and atlases. Worldwide, Langenscheidt Verlag of Germany is ap proximately the same size, counting sales from map and atlas publishing in Germany, Britain, and the USA together with publishing of travel guidebooks and lan guage dictionaries.

The larger map publishers rely on the sales of street and road maps. Most of the commercial map publishers focus on the travel market, providing maps and atlases for those on vacation or holiday, and business travelers. Specialty map publishers providing maps for hiking, biking, and other outdoor activities; thematic products for historians and railroad buffs; and wall map publishers of wall art are usually much smaller companies. A third class of commercial cartography provides custom carto graphic services for businesses. Custom cartographers also tend to be smaller companies, although several have between 10 and 50 employees.

The Size of the Commercial Map Publishing Market and Current Sales Trends

The size of the map publishing industry is difficult to measure. During the 1990s, the International Map Trade Association conducted market research in North America. The surveys were taken annually for several years. Researchers asked informants about their purchases of maps and atlases in shopping malls across the US and Canada. The results indicated that producers sold $500 million of maps, atlases, and globes to retail shops in the two countries. Assuming retailers double the price before selling to the public, the map market in North America was worth about $1 billion in the 1990s.

The market has grown by a quarter or a third since then, but has stagnated most recently. Publishers of professional quality street atlases have seen flat or de clining sales in the last few years, especially in Europe, where in car navigation systems are more accurate and trustworthy. Widespread use of the Internet for travel research and quick answers to geographical questions has stopped the growth of paper maps, although the wide dissemination of map images on websites has also promoted increasing use of cartographic products. These opposite forces have combined to maintain steady sales of paper maps through the first decade of the twenty first century, but the future direction of the market is less clear. Paper remains a useful medium for cartography, providing a wider viewing area than a computer screen and instant availability of the information without need for electronics.

Commercial map publishing in Europe is more widespread and successful than in North America for many reasons. The size of the market is probably more than $2 billion, although no data exist to prove it. This guess is based on the wider acceptance of maps, perhaps due to higher quality products (more accurate and cartographically more advanced), better geographical education and knowledge, more time for holiday travel, and roads following random directions. In contrast, much of North America uses a simple grid system for roads that makes maps less necessary for navigation.

Commercial map publishing in the rest of the world probably does not have more than $1 billion in sales, although Japan especially has a strong market for street and road atlases and wide distribution in bookstores and shops throughout the country. Perhaps the system of naming the buildings and addressing them around a block of structures rather than giving them street names and numbers results in a reliance on high quality mapping.

Taken together, the world map and atlas publishing market, including globes, is probably between $3 and $5 billion in sales per year.

A Survey of Map and Atlas Publishing in North America

As noted above, Rand McNally is the largest commercial map publisher in North America, with sales of around a quarter of a million dollars. The company is based in Skokie, Illinois, a suburb of Chicago. Founded in 1845 to produce railroad timetables, it expanded in the early twentieth century to produce road maps. This was pri marily a premium business from the 1920s until the 1970s, when petroleum companies no longer purchased maps to give away at service stations. Rand McNally's business relies heavily on two markets. First, Rand pub lishes the leading commercial road atlas of the United States, selling over 6 million annually. Second, his com pany is the leading distributor of folded city, state, and regional maps into mass market retail chains across the nation. Rand relies on a fleet of vehicles and independent sales agents to inventory and re stock customer stock. Rand's mass retailers, such as Wal Mart, Target, Office Mart, and Borders, allow the publisher to manage the map section independently.

The second largest map publisher and distributor is the portion of the Langenscheidt mapping business in the US. The American portion has sales of around $100 million. Andreas Langenscheidt, owner of the company and resident of Germany, purchased several independent map publishers during the past decades. Today, the Langenscheidt family includes American Map Corp (AMC), Hagstrom, and Hammond, all based in metropolitan New York City; ADC, based in the Wash ington, DC area; Arrow, located in Boston; Trakker, op erating near Miami; Creative Sales, based in the Chicago area; and Patton, based in suburban Philadelphia. Each subsidiary has maintained a separate identity until re cently, when some of the brands have been replaced with ADC. Trakker, Creative Sales, and Patton have been moving in this direction.

Another half dozen North American commercial map publishers and distributors have sales between $10 and $30 million. DeLorme, headquartered in Freeport, Maine, publishes a popular series of state road atlases. Universal Map, with offices near Lansing, Michigan, and Orlando, Florida, publishes road and street maps. In September 2007, Universal filed for Chapter 11 bankruptcy pro tection. National Geographic publishes wall maps and has a strong series of outdoor recreation maps from offices near Denver, Colorado. Mapsco, a publisher of pro fessional quality street atlases based near Dallas, Texas, has significant map and atlas sales in Texas and surrounding states. Map Link, North America's wholesaler of maps, distributes products from most commercial map pub lishers to bookstores and other retails. Replogle, the largest globe manufacturer and based near Chicago, is probably the largest of this next group of commercial publishers. Globes are a useful cartographic and geographic product, but their market is quite different from that of maps and atlases. Globes are bought to decorate and furnish homes and offices instead of being used in navigation.

A Survey of Map and Atlas Publishing in Europe

Unlike North America's single market, Europe's map and atlas market is fragmented. Publishers primarily create products for their home country. Because of national borders and language barriers, each major country in Europe has strong commercial map and atlas producers. Also, the government mapping agencies for a couple of major European nations have transformed themselves into very strong commercial publishers.

Germany is the largest market within Europe. Until July 2007, three major commercial publishers competed in the market. MairDumont, based near Stuttgart, pub lishes guidebooks as well as maps and atlases. They publish German city plans, regional mapping at inter mediate scale, and an international series of country road maps. MairDumont acquired the primary competing map programs for Germany from CartoTravel and ADAC in 2007. CartoTravel published Falk plans for German cities and regions. ADAC is the automobile club of Germany. Falk maps and the ADAC map program is now published and distributed by MairDumont. Other cartographic products in Germany are available from Internationales Landkartenhaus (ILH), also known as Geo Center, a distributor based in Stuttgart. ILH maintains a complete cartographic warehouse, managing a large inventory of titles for academic and worldwide distribution as well as travel titles for the German retailer.

The British Ordnance Survey (OS) dominates map ping in that nation, integrating large scale surveying with small scale publishing. Based in Southampton, OS em ployed nearly a thousand people as recently as 20 years ago. The OS today has a turnover of around 100 million pounds, making it still one of the largest map publishers in the world, despite major layoffs. Britain has several important map publishing companies, including the cartographic division of Harper Collins (formerly Bar tholomew), Phillips, and the Automobile Association (AA). Maps are also widely sold by wholesalers and dis tributors. GeoCenter UK, a Langenscheidt company, is the largest of the specialty map and travel distributors.

The French market has two giants competing for the commercial map and atlas market. The French govern ment's privatized mapping agency, the Institut Geo graphique National (IGN), located in Paris, produces street maps and atlases, providing complete coverage for the nation at 1:100 000 and also 1:250 000. IGN also offers road atlases of France. Similarly, Michelin, the multi national tire manufacturer, also produces street and road mapping for all of France. The two publishers appear to target products to match their competitor's offerings. Both are widespread in shops throughout France. Other commercial publishers in France are smaller companies with specialized maps. A wholesaler based in Tours, Cartotheque, distributes many of the remaining products.

Three commercial Italian map publishers provide the market with most of the city plans, regional maps, and atlases of Italy. Touring Club Italiano, based near Milan, distributes through regional warehouses. DeAgostini, a reference publisher, also produces regional maps and atlases for the country. Litografia Artistica Cartografia (LAC) of Florence produces province maps and city plans for Italy. The market for Italian maps is dominated by the competition between these three companies, al though Belletti and Studio FMB also publish city and regional maps for the country.

Other Western European nations typically have one strong mapping company or agency. Hallwag and Kuemmerly, owned by MairDumont, dominates the Swiss market for maps, while Freytag and Berndt has been Austria's leading map publisher since the nineteenth century. The market leader in some nations is the government mapping agency, such as the Irish OS, or the auto club, such as ANWB in the Netherlands, or a private, commercial company, such as Turinta in Portugal.

Eastern Europe has many emerging commercial publishers. During the Cold War, Hungary's government had a very extensive mapping agency, Cartographia, which has been privatized. Cartographia now operates as a fully commercial enterprise, but is much smaller than before. Trained cartographers abound in Hungary; so, many small map publishing companies have emerged, led by Gizella Bassa and Andras Szarvas, among others. Similar stories exist in the Czech Republic, where Shocart has established itself as a local leader. Jana Seta publishes maps in Latvia, Topkart is a wholesaler and distributor of maps in Poland, and smaller publishers exist in the remaining countries. Many of the older commercial companies based in Germany and Austria have also opened cartographic centers in the Czech Republic, Slovakia, and Hungary in order to utilize the trained cartographers in the region at lower wage rates.

A Survey of Map and Atlas Publishing in Asia, Africa, Latin America, and Oceania

A few other significant commercial map and atlas publishers have emerged in other parts of the world. Japan is well mapped with street and regional atlas products produced in Japanese for the local market. Two publishers, Shobunsha and Kodansha, produce maps sometimes in English for visitors, but the vast majority of maps produced are in Japanese only.

Commercial mapping exists and flourishes in China. Although produced by state sponsored commercial firms, it is produced and sold in the same way commercial mapping is in the US and Europe. The market varies from small scale national atlases and country maps to professional grade street atlases and city plans. There are also surveying, GIS, and mapping activities at the province and city level. Their products are widely distributed and used by a growing middle class.

Beetle Map, a Korean map publisher, has produced illustrated drawings and folded maps supported by advertising and premium sales. Periplus, a Singapore publisher of guides and maps, has emerged as a leader in map and atlas production in East Asia. Smaller map publishers in Thailand, Malaysia, India, Iran, Lebanon, and Israel produce titles for their local markets. Some are distributed to the world community through Map Link in North America and ILH Geo Center in Europe.

Mapping is difficult to obtain in much of Africa, although Struik has emerged as the major commercial mapping company in South Africa. It is generally known by its brands, Map Studio and New Holland.

Small commercial map publishers exist throughout Latin America. Guia Roji has published Mexican state and city maps for decades. Guia Quatros Rodas produces an annual guide and map for Brazil. Publishers in Chile, Peru, and Argentina provide major city street mapping and country road maps for their respective nations.

Four Australian map publishers dominate the market for maps and travel guides. The world's most widely distributed English language guidebook publisher, Lonely Planet, is based in Melbourne. Now owned by the BBC, Lonely Planet publishes a series of destination travel maps for major world cities; maps are a small part of the company's business. Universal Map Enterprises, based in Sydney and now owned by Telstra, the Australian telephone company, is the leading publisher of professional quality street atlases. Hema Maps of Brisbane is the Australian leader in travel maps, producing a national road atlas, state and regional maps, and city plans for the entire region, including New Zealand and nearby island nations. Explore Australia publishes a series of state and city plans, which accompanies their market leading guidebooks.

A Survey of Custom Cartographers

Many commercial mapping companies are not publishers, or publishing is only a sideline. Instead, most of these companies produce cartography primarily under contract for other businesses. The companies each specialize in different areas. The range of services provided can be illustrated by a brief survey of the North American custom cartographers.

GeoNova is the new name for one of the largest custom cartographic companies in the US. Formerly part of MapQuest, this company has about 50 employees and maintains a detailed database of the US. Michelin, Langenscheidt, and Barnes and Noble have all contracted for products from MapQuest during the past decade. In addition to creating products for other map publishers, GeoNova also creates custom graphics to illustrate geography textbooks and other reference publications.

There are two other contract cartography businesses with over 20 employees. Mapping Specialists of Madison, Wisconsin, provide thematic and street mapping as well as special graphics for textbooks and atlases. Maps.com of Santa Barbara focuses on education and media, providing special cartographic images for current events, curriculum, and textbooks. These larger contract cartography businesses often provide mapping to telephone book publishers, publish a few titles, and create custom wall maps for other businesses for use in sales and marketing.

Many smaller contract cartographers exist throughout North America. Most are just one or two people, preparing maps and graphics for book publishers, the media, and local government. Some have become specialized, such as Stuart Allan's firm, Allan Cartography of Medford, Oregon. A true cartographic craftsman, Stuart has patiently improved the cartographic presentation of physical relief over decades of tinkering. Similarly, Nat Case of Hedberg Maps and Alex Tait of International Mapping Associates have achieved a reputation for graphic innovation and data precision. The craftsmanship and care of these custom cartographers contrast with the lesser attention to cartographic excellence from the majority of small cartographers. Most premium and custom cartography is printed inside folded maps, brochures, or pamphlets that are supported by advertising. The quality of the cartography contained in a map distributed at no cost is usually less than that contained in a product sold in a shop or purchased online.

The market estimates are 'not' reliable and cannot be tied to a date. They are just the author's best guess from talking to everyone in the map industry.

See also: Mapping, Distributed; Maps.

Relevant Websites

www.hemamaps.com.au
Hema.
www.ign.fr
Institut Geographique National.
www.maptrade.org
International Map Trade Association. The monthly map reports give a lot of information about the map industry.
www.langenscheidt.com
Langenscheidt.
www.mairdumont.com
Mairs Dumont.
www.nationalgeographic.com
National Geographic.
www.ordnancesurvey.co.uk
Ordnance Survey.
https://peripluspublishinggroup.com
Periplus.
www.randmcnally.com
Rand McNally.
www.shobunsha.co.jp
Shobunsha.

Mapping, Cyberspace

M. Zook, University of Kentucky, Lexington, KY, USA
M. Dodge, University of Manchester, Manchester, UK

Glossary

Cyberspace A term coined by science fiction author William Gibson, now used to refer to the space of interaction and information available through the computer networks, such as the Internet.

Information Space A generic description for all types of electronic communications networks such as peer to peer file exchange, instant messaging, and mobile devices.

Map Mash-Up The practice of dynamically combining two or more Internet-based services into a new map.

Spatialization Map-like interfaces that employ spatial metaphors to represent data that are not necessarily spatial.

Introduction

Over the past 30 years many different mapmakers, from a diverse range of academic, technical, and institutional backgrounds, have mapped various aspects of cyberspace. Cyberspace cartographies are one of the significant areas of creativity in contemporary mapmaking, with a considerable amount of experimentation with visual metaphors, survey methods, data sources, and, above all, novel forms of user interaction with map representations. Indeed, because cyberspace lacks established conventions of representations, it has proven to be a ripe domain for innovation in cartographic theory and practice such as facilitating new types of participatory, bottom up, mapping activity. At the same time, it offers powerful new means for surveilling and mapping social activity and can increase the capability of government and corporate interests to control what is represented.

Mapping cyberspace can be usefully categorized into three distinct cartographic modes:

- maps in cyberspace;
- maps of cyberspace; and
- maps for cyberspace.

The first mode, maps in cyberspace, involves putting existing forms of geographic cartography online to widen access and add user interactivity. Maps in this mode utilize conventional spatial conceptualizations and are drawn in much the same way as long established paper based maps. They often, however, depend upon the availability and organization of geospatial data in cyberspace, thereby affecting how the content of these maps is prioritized. The second mode, maps of cyberspace, documents the infrastructures and operations of cyberspace itself, and is, therefore, primarily defined by the subject of the maps rather than the spatial conceptionalization behind the map or the tasks undertaken with them. The resulting maps encompass a multitude of graphic forms, some of which appear quite uncartographic in a normative sense, such as topological network graphs. The last mode, maps for cyberspace, produces maps designed to navigate through the virtual spaces of cyberspace. These maps are mostly created through the spatialization of nongeographic information structures to produce map like interfaces to support interactive browsing and searching. As such this mode is primarily defined by the task to which the maps are put rather than their subject or spatial conception.

As with all categorization, this threefold classification is a simplification. The boundaries between idealized map modes are not always clear cut and, as noted below, there are cases in which examples can be placed in different classes depending on the context in which they are used. Thus, cyberspace mapping is best viewed as a wide continuum of different representations running from static online cartographic maps of geographic space at one end to immersive spatial interfaces for the navigation of purely virtual space at the other.

Maps in Cyberspace

Mapping in this mode has already produced demonstrable utility and commercial viability in putting 'real world' cartography online through developments in web mapping portals and Internet based geographic information system (GIS) services. For example, many popular mapping services allow users to locate addresses or obtain driving directions for trips including options for avoiding road construction or other delays. While most of these are commercially driven, they are generally free to end users and rely on advertising and sponsorship for revenue generation (**Figure 1**). Because of the continuity with previous forms of mapping, this mode has received much attention from cartography practitioners and academic researchers. This work is mostly of a technical nature, that is, adapting existing practices to the new media and developing appropriate forms of user interactivity.

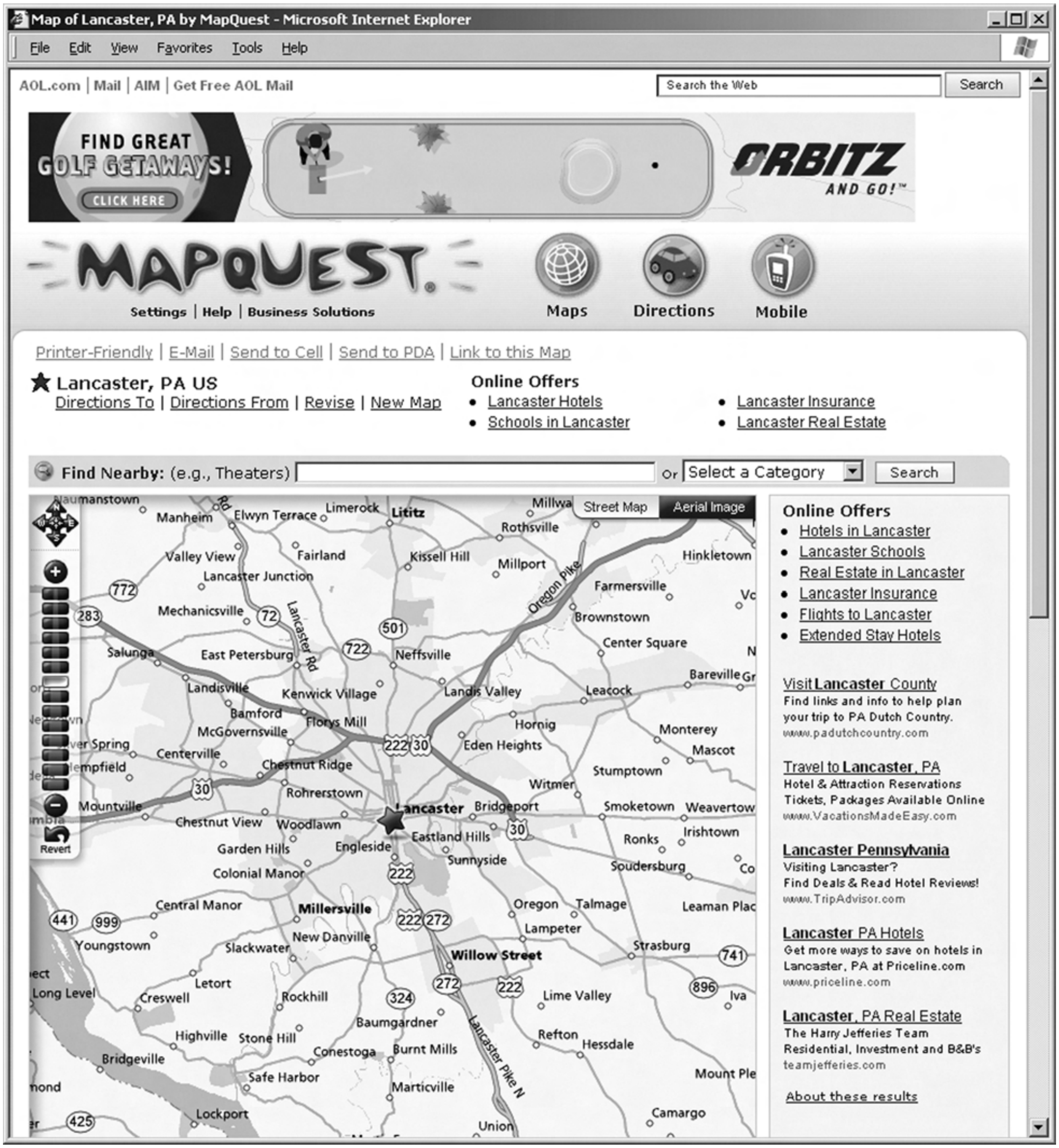

Figure 1 A typical 'maps in cyberspace' example showing a road network representation, generated by the MapQuest service. Source: author screenshot.

In addition to providing driving directions and the ability to find and map nearby facilities, the emergence of virtual globes with aerial and satellite imagery, such as Google Earth and NASA's World Wind, represents a seductive 3 D interface for exploring geographic space ranging from simple search to identifying (and com menting on) points of interest (**Figure 2**). Thus, maps in cyberspace are creating the means for a more dynamic and interactive mapping that, arguably, transforms people from passive consumers of cartography into much more active map users. Indeed, online mapping services such as Google's 'My Maps' allow Internet users to create their own maps and share them with the entire online com munity. While potentially empowering for those with access, persistent gap between the 'connected' and those who are offline or unskilled in the use of these

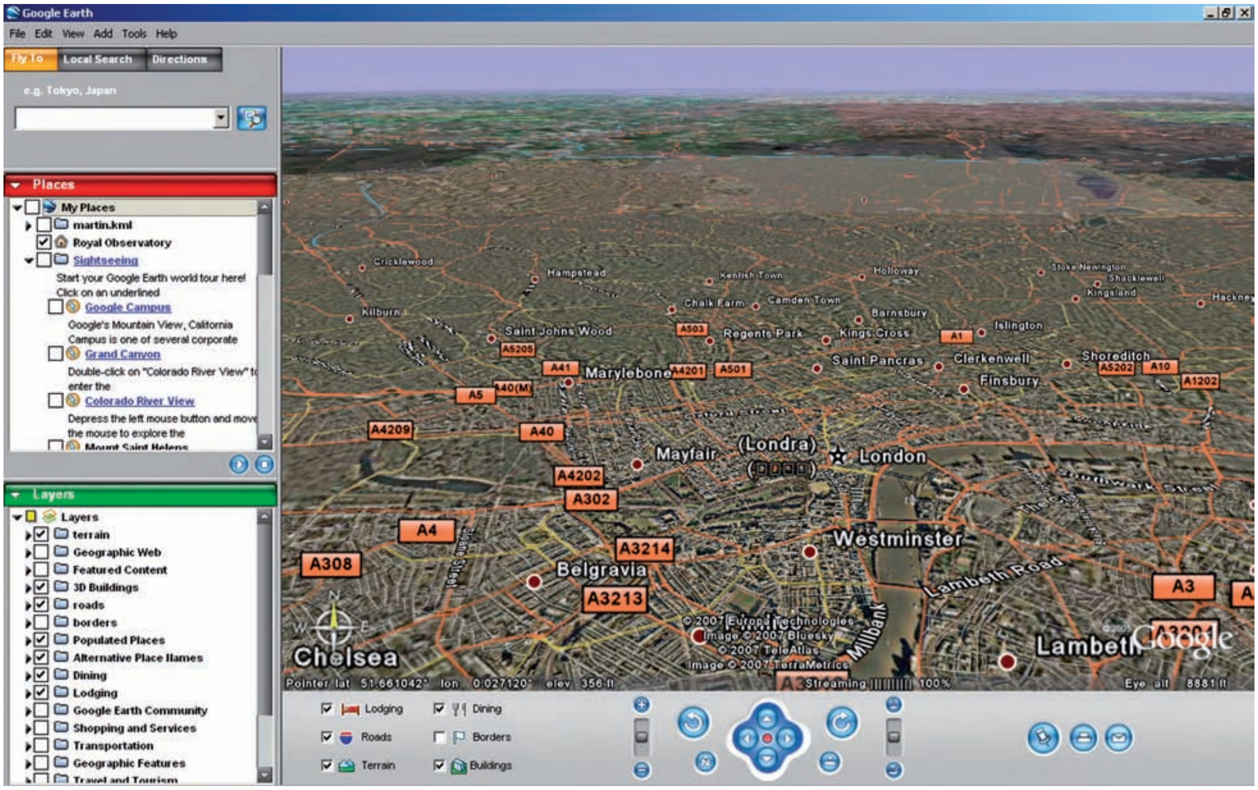

Figure 2 An example of virtual globe display, showing central London seen through Google Earth. Source: author screenshot.

technologies means that active participation in online mapping remains highly differentiated. While this gap is not novel, the replication of the digital divide in the use and annotation of online maps raises questions of whose representations and understandings of place will be mapped and prioritized.

For example, a number of hobbyists and hackers have begun to generate 'map mash ups', the combination of multiple online data sources into a new integrated service that maps a specific phenomenon. Examples include locating apartment listings from craigslist.org and mapping user reported prices at gasoline stations. Mash ups illustrate how the combination of digital data available on the Internet with basic mapping software can create new and useful interfaces to the world, albeit the nature of these interfaces is closely tied to who is online and who is not. Moreover, this expanded mapmaking capability is accompanied by a number of related concerns about control over the organization and representation of many maps in cyberspace. For example, databases of businesses and other activities utilized by these maps are often organized at the postal code level to allow for faster processing and map generation. Thus, despite the serious shortcomings of postal codes as areal units, they are becoming a 'naturalized' division used for these maps.

A more serious matter is the way in which the presence and popularity of a business or organization in cyberspace determines the order in which it appears in a map. Google Maps, for instance, uses a hybrid measure of distance that combines physical distance and ranking in cyberspace to determine the order in which businesses appear on maps. Entities with large and highly ranked websites can be prioritized over those with little to no presence in cyberspace even if the latter entities are physically closer to the searched location. Additionally, since maps in cyberspace are largely commercial products, this ranking process (including the possibility of paid placement) is opaque to users. In short, the control over these maps and the algorithms used to generate them is vested in private companies without accountability to the public who uses them.

This privatization of mapping marks a shift in cartographic power away from governments who have long naturalized their agendas via maps designed as instruments of state control. The shift to corporate control of maps, albeit troubling in its own right, challenges the states' ability to shape what appears on maps of their territory. For example, some states such as Greece, Turkey, Pakistan, and South Korea have complained when imagery of sensitive military and government sites has appeared in online mapping services such as Google Earth. While this loosening of state control over mapping is emancipatory, the power to control maps nevertheless remains concentrated. A potential means to place control of maps within the citizenry is web enabled 'democratic cartography' such as the 'OpenStreetMap' project, that is, a grassroots effort to create an open source topographic database in the public realm. While much more participatory and less passive than other mapping methodologies it nevertheless presupposes that contributors are on the 'connected' side of the digital divide. Although issues of quality, reliability, and scalability remain, the ability to leverage cyberspace to organize group efforts to create open source spatial data highlights the potentially revolutionary nature of maps in cyberspace.

Maps of Cyberspace

The maps of cyberspace mode encompass graphic representations that facilitate the spatial understanding of the materiality of cyberspace itself, that is, showing the geographic patterns of network infrastructure, content production, and/or the distribution of users. This narrow focus of subject matter is a key characteristic of maps of cyberspace.

Many of the maps produced in this mode use the semiotics of mainstream cartography, for example, cable routes as colored lines on a geographic base map or thematic mapping that spatially represents statistical data on usage (**Figure 3**). However, other products of the maps of cyberspace mode go beyond conventional cartographic conceptions and use nongeographic forms of representations, for example, non Euclidean visualizations of the topological structure of network infrastructures (**Figure 4**). Such abstract graphs illustrate the relational connectivity between Internet nodes rather than their position in geographic space. In addition, a number of maps of cyberspace also expand beyond the two dimensions of mainstream cartography and utilize 3 D graphics and virtual reality interfaces, for example, visualizing web server traffic as 'skyscrapers' on a virtual globe (**Figure 5**). While an innovation in cartographic practice these maps often perform poorly in actually conveying useful information.

Few of the makers of maps of cyberspace call themselves cartographers; rather they are a diverse group ranging from individual programmers to governmental agencies. Unsurprisingly, the most prolific group are operators of data networks who make maps of their infrastructure to accomplish particular and immediate goals such as analyzing traffic or projecting growth. Much of this work is never made public. This lack of publicly available maps of cyberspace, tied to the 'invisibility' of telecommunications infrastructure more generally, compels many to become cyberspace mapmakers in order to do their jobs. This scarcity of maps has made individual authorship of maps commonplace, particularly because the cyberspace infrastructures can be used to map themselves in quite innovative ways and at very low costs.

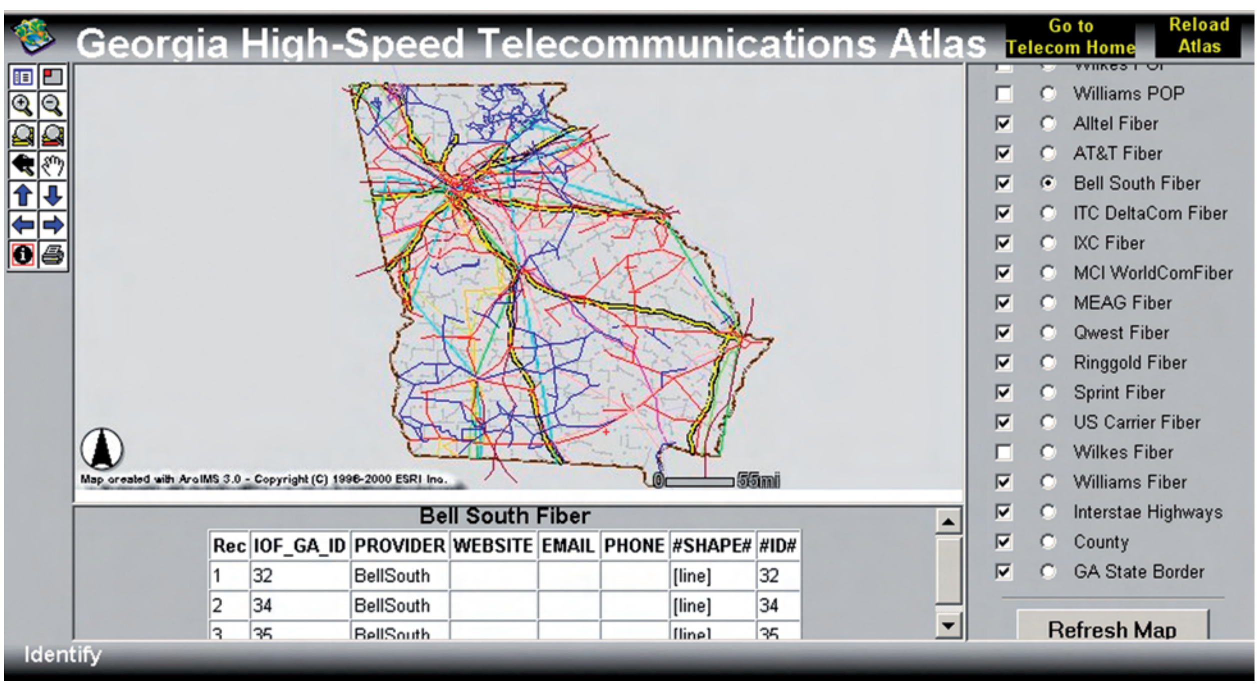

Figure 3 A telecommunications atlas of network infrastructure in Georgia, USA, is a typical example of 'maps of cyberspace' mode. However, it was disseminated using web mapping technology that is characteristic of the 'maps in cyberspace' mode. The atlas was produced by university researchers at Georgia Tech as an information resource for regional economic development. Source: author screenshot; service no longer available.

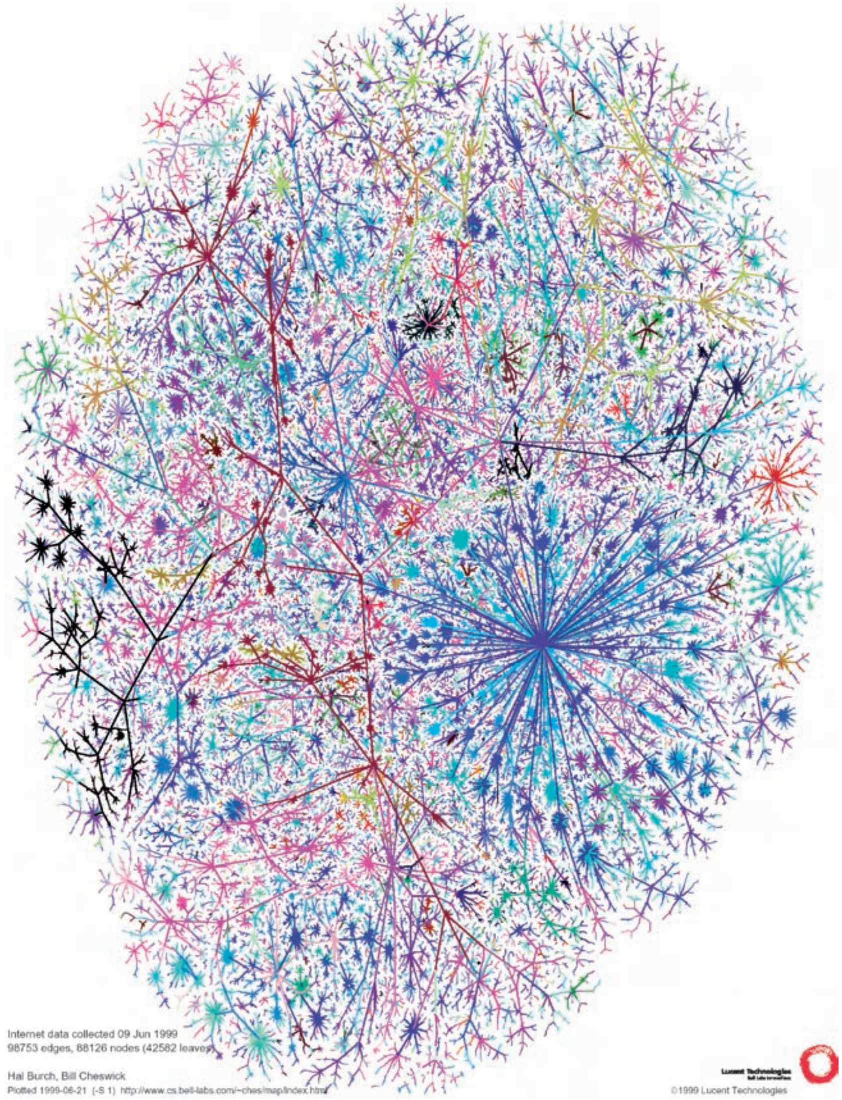

Figure 4 A ‘map of cyberspace’ example which graphs the core of the Internet ‘cloud’ in topological terms. The color coding of nodes seeks to highlight zones that share common network addresses and likely corporate ownership. Source: Bill Cheswick, Lucent Technologies, 1999.

This allows novel opportunities for, so called, super empowered individuals to chart vast swathes of cyber space with minimal resources, utilizing software to automate the surveying process and reducing the burden of visualizing large volumes of data.

Given the diversity of producers of maps of cyberspace it is not surprising that they serve multiple normative purposes. These range from documenting cyberspace in frastructure and flow, to cartographic propaganda by companies with vested financial interests in the expansion of cyberspace, to maps used in policy analysis. Infra structure based maps of cyberspace can show complex computer networks over several spatial scales, from the street up to the global scale and are used to maintain physical hardware and manage network flows more ef fectively. The propaganda maps are tied to the market goals and are deployed as persuasive devices to support the rhetoric of expansionism and as a means to exert sovereignty of private capital over public cyberspaces. A cursory examination of most Internet service provider (ISP) websites reveal bright, colorful, and visually ar resting maps used to highlight the advantages of the latest communications technology to prospective investors and potential customers (**Figure 6**). An additional important purpose of maps of cyberspace is academic and policy analysis of the expansion of information flows and cyberspace. Pioneered by geographer Jean Gottmann’s work on intercity telephone call patterns in the 1960s, much of this work focuses on explaining the exponential growth in Internet infrastructures, connectivity, and usage. Visual summary presentation using statistical charts and geographic maps is common.

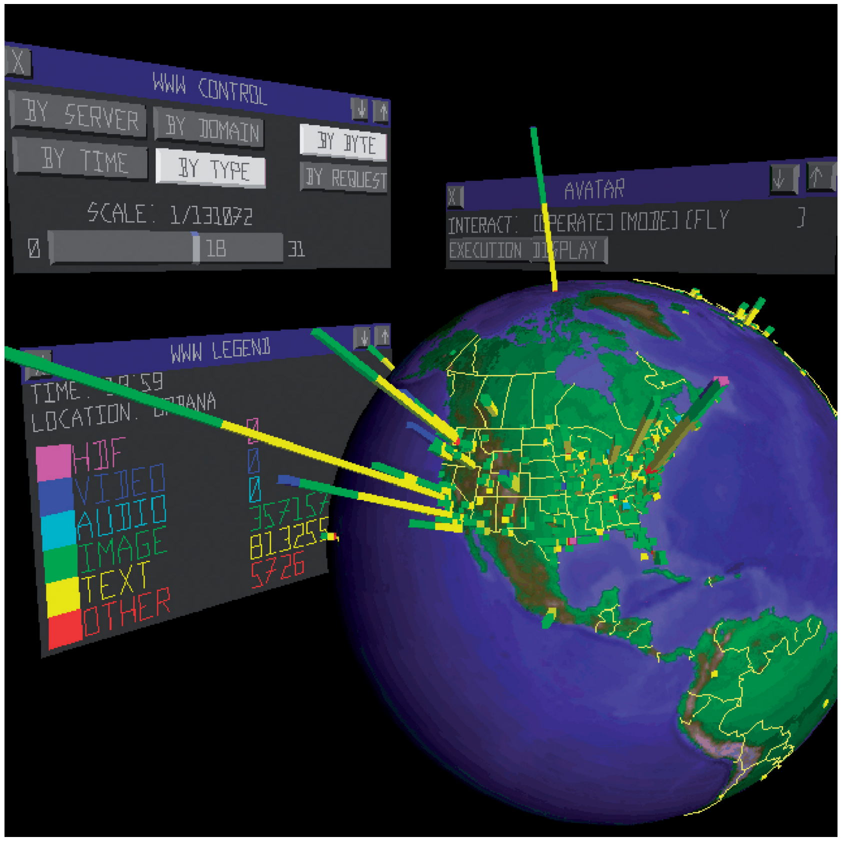

Figure 5 A 'map of cyberspace' showing aggregate traffic on a web server as geocoded bar drawn onto a globe. It could be interrogated within a virtual reality interface. From Lamm, S. E., Reed, D. A. and Scullin, W. H. (1996). Real-time geographic visualization of World Wide Web traffic. *Proceedings of Fifth International World Wide Web Conference*, Paris, France, 6 10 May.

In addition to these three specific purposes, maps of cyberspace taken as a whole have a significant pedagogic utility in challenging the naïve notions that virtual interactions spell the death of distance or render geographic location meaningless. Maps of cyberspace reveal the complex intersections between virtual space and geographic space at various scales. A number of these maps can provide insights into who owns and controls the supporting infrastructure, how and from where content is being produced, and how the uneven diffusion of infrastructure affects cost, speed, reliability, and ability to connect. These maps illustrate how infrastructure is concentrated at the global scale in certain countries (such as the US, UK, and Scandinavia), at the national scale in certain regions (e.g., Silicon Valley, the west London M4 corridor, the Helsinki metropolitan area), and even at localized neighborhood clusters within 'high tech' cities like San Francisco or New York, for example, Matthew Zook's mapping of Internet domain name ownership (**Figure** 7).

Despite the utility of these efforts, the available maps of cyberspace provide, at best, only a partial view of cyberspace given that data are limited in many areas, for example, information on Internet traffic flows between and within cities is unavailable to researchers. Moreover, mappable information of cyberspace is actually diminishing as the growing diversity, size, and privatization of cyberspace, as well as post 9/11 security concerns are making it harder to survey and represent it legibly.

Maps for Cyberspace

The extent and usage of cyberspace have grown very rapidly in the last decade. With so many distinct virtual spaces and users online, cyberspace has become an enormous and often confusing entity that can be difficult to cognize and navigate. The maps for cyberspace mode focus on helping people understand the structures of online spaces of information and social interaction, rendering them in visual form and enabling people to navigate through them. In other words, maps become interfaces of exploration 'inside the wires', rather than representations of how the 'wires' themselves are arranged and produced. They map immaterial information rather than material infrastructures.

Cyberspace has meaningful informational structures to be surveyed, calculated, and mapped. For example, the

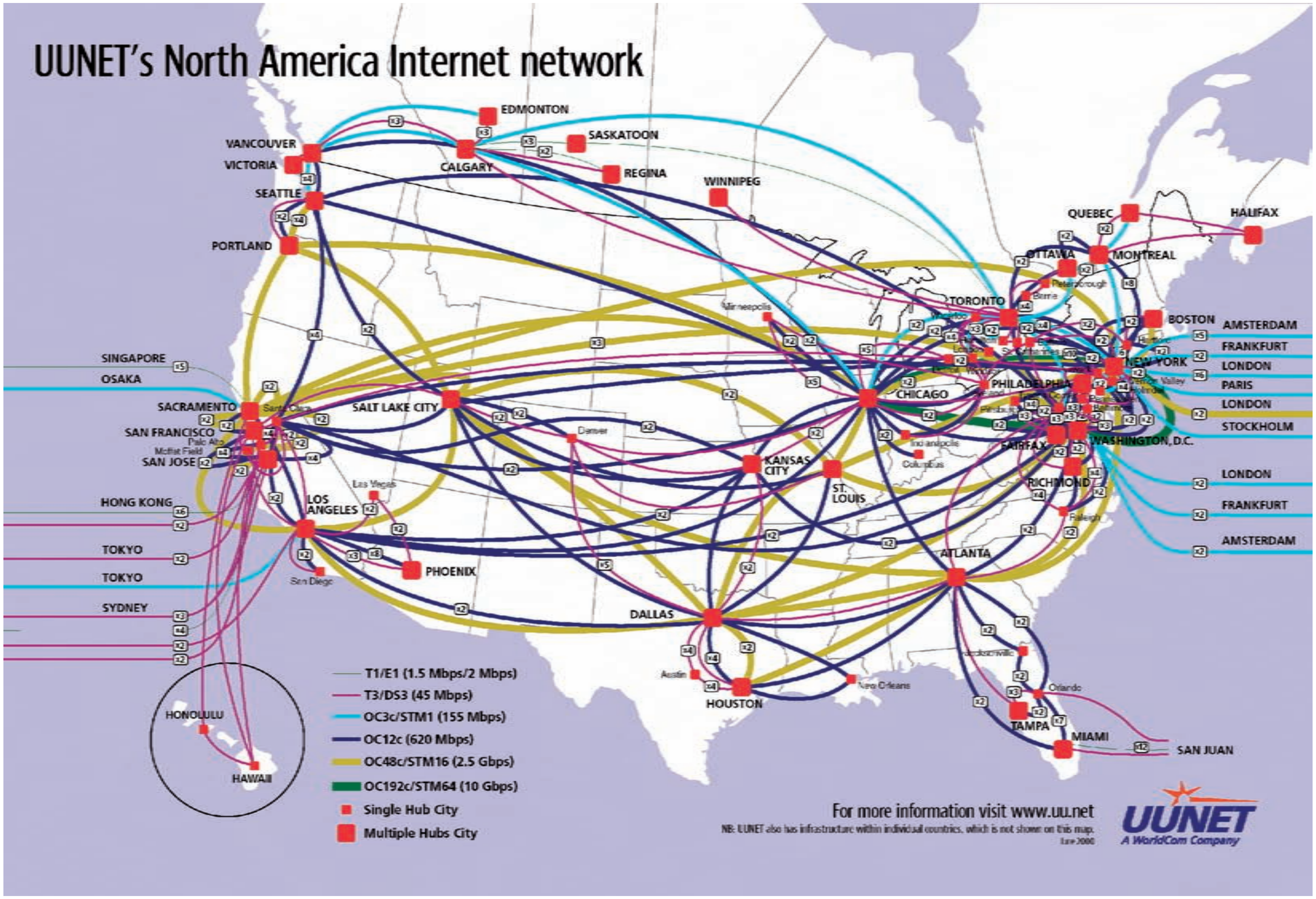

Figure 6 A map of cyberspace' produced for market ng Wor dCom's UUNET network nfrastructure. Source: or g na y pub shed on corporate webs te, June 2000. No onger ava ab e.

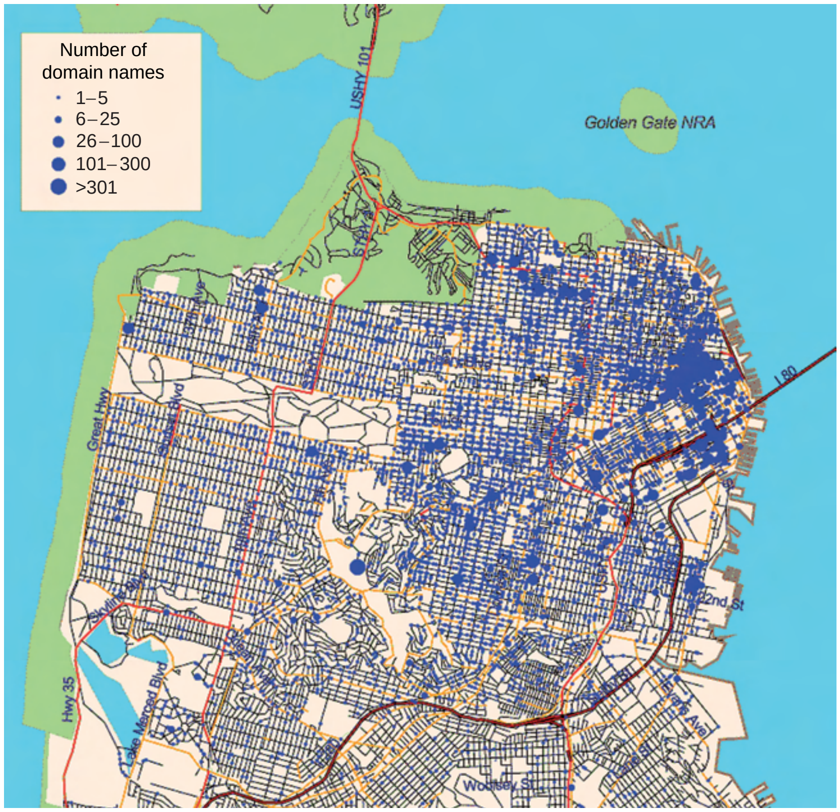

Figure 7 A 'map of cyberspace' representing the number of Internet domain names within San Francisco, based on street address of the registered owners. Source: Matthew Zook.

semantic similarity between content, affinity ties of differing strengths in online social networks, and turn taking in mediated conversations have all been successfully transformed into map like representations. Contributions by cartographers or geographers have been minimal, instead a diverse group of mapmakers, including graphic designers, sociologists, information scientists, librarians, and interface engineers (mostly located within academia) have led this work. These efforts, however, have largely failed to produce workable maps suitable for widespread public usage.

Many of these projects, particularly within computer science, emerge from information navigation studies which focus on creating more efficient means of human–computer interaction. Online spaces, such as the web, simply provide a conveniently accessible, large scale test bed for this work. Another important focus is the information design community that structures the architecture of the online content such as site maps on websites (**Figure 8**). Valuable and eclectic contributions have also come from new media artists developing interactive maps as works of art and as virtualized architectural spaces.

Maps for cyberspace have utility since they render intangible virtual media, composed of immaterial software algorithms and database records into visually tangible spaces. Even though one cannot 'touch' hypertext, for example, it is possible to plot its structures on screen to aid user navigation. Depending on their scale and design, information maps can give people a unique cognition of a space otherwise difficult to understand. As such, mapping of information space offers three distinct and interlinked advantages:

- creating a sense of the whole information space;
- supporting ad hoc interactive user exploration; and
- revealing hidden connections between data objects.

In a metaphorical sense, these maps enable users to get a 'bird's eye view' of an information space. Such overview visualization, displayed on a single screen for cognition at a glance, is particularly important given that most online information seeking is via unstructured and poorly formulated browsing and foraging techniques. These maps try to provide an intuitive and meaningful interface to the structures of information space not only in terms of direct relationships between documents (via citations or hyperlinks, for example) but also in terms of shared themes, semantic connections, and common usage within the document's content. These structures and relationships are usually completely hidden in the presentation of conventional interfaces, like web browsers. Yet this is often where one finds insight and answers, in the

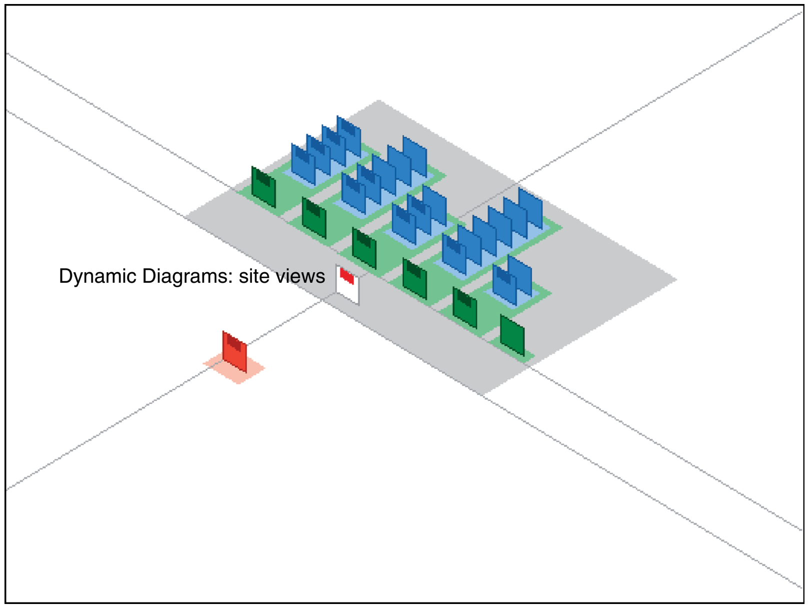

Figure 8 A 'map for cyberspace' system developed by Dynamic Diagrams provided as an interactive spatialization to aid website navigation. Source: author screenshot.

visual–cognitive assimilation of how the mosaic of available information fits together. The effective power of maps for cyberspace comes from showing these relationships to users to enable them to make better navigational decisions.

Developments in the field of information visualization in the last decade have proved particularly fertile through processes of spatialization. These are map like interfaces that employ explicit spatial metaphors to represent data that are not spatial. Spatialization renders large amounts of abstract data (usually textual corpus) into a more comprehensible, compact visual form by generating meaningful synthetic spatial structure (such as distance on the map display scaled according to a metric of lexical similarity between data items) and applying cartographic design concepts from topographic mapping and thematic cartography. Some of the most cartographic looking examples have used the conventions of hill shading and terrain contouring to create browseable virtual landscapes (**Figure 9**). There is scope for greater involvement of cartographers in information visualization to develop improved spatializations. In turn, developments in spatialization driven by computing and information science specialities are feeding back into mainstream mapping practice, pushing the boundaries of cartographic theory.

Despite these potential advantages, creating workable spatializations faces significant epistemological challenges. This is particularly the case because cyberspace's many information spaces overlap, but often in *ad hoc* and unplanned ways, giving rise to complex, nonhierarchical and dynamic structures that are not easily surveyed or visualized. Moreover, cyberspace is not a homogenous or static phenomenon, but a myriad of rapidly evolving digital databases, communications channels, and media platforms, with distinct forms of virtual interaction. These are inherently transient landscapes where changes are 'hidden' until one encounters them. Change can happen instantaneously, for example, deleting a web page leaves behind no trace (unless archived elsewhere). The lack of reciprocity in relations means an information node can vanish without notice or notification to any other party.

These issues of information mutability and transience are likely to grow, and become obfuscated by the increasing use of encryption and *ad hoc* distributed architectures (e.g., peer to peer file sharing, mobile devices, and Wi Fi networks) making information mapping even harder. The task of generating even a basic index of parts of cyberspace continues to tax the largest corporations and government agencies. Web search engines, for example, struggle to keep pace with the growth and mutability of just this one part of cyberspace. In contrast to maps of geographic space which may remain valid for decades (e.g., topographic maps), the shelf life for many cyberspace maps is short. To combat this, maps for cyberspace could be structured to dynamically represent virtual space in real time, much like a radar map for tracking weather patterns.

A final challenge for spatializing cyberspace relates to the nature of the space where space time laws of physics have little meaning. This is because information spaces are purely relational and are solely the productions of their designers and users and only have attributes of geographic (Euclidean) space if explicitly programmed. Thus, many information spaces violate two principal assumptions of

Figure 9 A ‘map for cyberspace’ called NewsMaps that produced an information landscape to visualize and navigate new stories. Source: author screenshot; service no longer available.

modern (Western) cartography. The first of these are the Cartesian properties of space as continuous, ordered, and reciprocal; there are no sudden gaps in the landscape, and the Euclidean notion of distance holds true, that is, the distance from A to B will be the same as from B to A. Yet parts of cyberspace are discontinuous, lacking linear or ganization and in some cases elements can have multiple locations. The second assumption is that the map is not the territory but a representation of it, that is, the territory has a separate, ongoing existence and meaning beyond the map. Yet there are virtual spaces, such as created in hypertext, where in a literal and functional sense the map ‘is’ the territory. Cartesian logic collapses and there is no reality independent of the representation. This can be experienced in the experimental 3 D fly through spatia lizations of hypertext, such as Apple’s HotSauce navi gation map interface (**Figure 10**).

A last important point to consider is the ethics of maps for cyberspace when they are used to visualize the pat terns of online social interaction between people (such as conversations and activities in virtual worlds) in order to gain insight into user behavior. These visualizations are double edged in that they may help inform the social life of an online community, but simultaneously they repre sent a type of cartographic surveillance which highlights interactions previously hidden in unused log files and databases. In ethical terms the act of mapping itself may constitute an invasion of privacy and infringement of personal rights, particularly if the appeal of some online social spaces is their sense of anonymity. For example, the public release by America Online (AOL) in August 2006 of user searches showed how revealing seemingly inno cent online interactions could be and how easily indi viduals could be identified. Thus, in some senses, these maps can shift the information spaces they chart from what their users consider semiprivate spaces to public spaces, thus changing the nature of the space itself.

Conclusion

Cyberspace mapping continues to rapidly develop, with visualization encompassing all types of information space. As wireless access diffuses and mobile devices become more sophisticated, the line dividing Internet space versus other information space is growing more indistinct. Moreover, as these electronic spaces are in corporated into our daily lived routines, we are in creasingly navigating through hybrid spaces in which the physical and digital meld together.

All three modes of cyberspace mapping discussed create innovative forms of representation and have

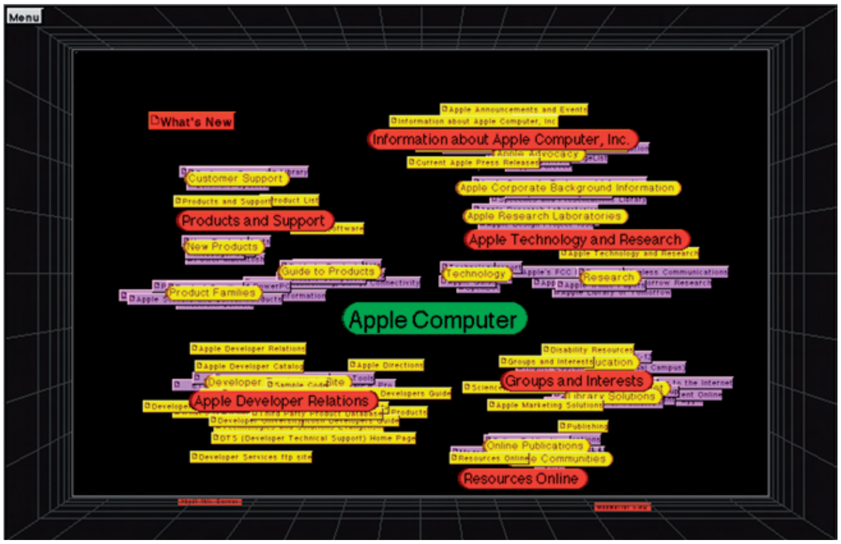

Figure 10 A 'map for cyberspace' called HotSauce that presented websites to users via a fly-through interface. Source: author screenshot.

expanded the frontiers of cartography in terms of what is mappable and how users interact with maps. Alongside these achievements there remain age old questions of the power and the politics of maps. While cyberspace is in some ways democratizing mapmaking, it has simultaneously provided governments and corporations with new opportunities for mapping and control. Maps continue to be effective governmentality tools when the cartographic gaze is harnessed by the state to discipline people and by corporations to market to them. For example, web browsing leaves detailed digital data trails that reveal everywhere a user has 'been' and what they read online. The resulting data, which generally reside with private companies, can be mined and visualized. The ways such trails and traces are used to build data profiles are particularly threatening because they open many new axes of discrimination for powerful public and private interests. Thus, while mapping cyberspace has opened new opportunities for spatial understanding, it also brings up larger questions of power and control. Who is making and using the maps? How is use of these maps monitored? How does the use of the map change the maps and mapmaking?

See also: Critical Cartography; Cyberspace/Cyberculture; GIS, Public Participation; Internet-Based Measurement; Internet/Web Mapping; Map Hacking; Map Types; Mapping, Topographic.

Further Reading

Card, S. K., Mackinlay, J. D. and Shneiderman, B. (1999). *Readings in information visualization: Using vision to think*. San Francisco: Morgan Kaufmann Publishers.

Castells, M. (2001). *The internet galaxy*. Oxford: Oxford University Press.

Crampton, J. W. (2003). *The political mapping of cyberspace*. Edinburgh: Edinburgh University Press.

Dodge, M. and Kitchin, R. (2000). *Mapping cyberspace*. London: Routledge.

Dodge, M. and Kitchin, R. (2001). *Atlas of cyberspace*. London: Addison Wesley.

Fabrikant, S. I. and Buttenfield, B. P. (2001). Formalizing semantic spaces for information access. *Annals of the Association of American Geographers* 91(2), 263 280.

Geowanking, Discussion of innovation geospatial technologies and open source mapmaking. http://lists.burri.to/mailman/listinfo/geowanking (accessed on 17 July 2007).

Gibson, R. and Erle, S. (2006). *Google map hacks*. Sebastopol, CA: O'Reilly.

Gottmann, J. (1961). *Megalopolis: The urbanized northeastern seaboard of the United States*. New York: Twentieth Century Fund.

Harpold, T. (1999). Dark continents: Critique of internet metageographies. *Postmodern Culture* 9(2), January 1999.

Kahn, P. (2000). *Mapping websites: Designing digital media*. London: Rotovision.

Kraak, M. J. and Brown, A. (2001). *Web cartography*. London: Taylor and Francis.

Lamm, S. E., Reed, D. A. and Scullin, W. H. (1996). Real time geographic visualization of World Wide Web traffic. *Proceedings of Fifth International World Wide Web Conference*, Paris, France, 6 10 May.

Peterson, M. P. (2003). *Maps and the internet*. Amsterdam: Elsevier Science.

Pickles, J. (2004). *A history of spaces*. London: Routledge.

Wise, J. A. (1999). The ecological approach to text visualization. *Journal of the American Society for Information Science* 50(13), 1224 1233.

Zook, M. A. (2005). *The geography of the internet industry*. Oxford: Blackwell.

Relevant Websites

http://www.cybergeography.org/atlas
Atlas of Cyberspaces (maintained by Martin Dodge).

http://www.caida.org
Cooperative Association for Internet Data Analysis.

http://www.housingmaps.com
Example of Map Mash up (Apartment listings).
http://www.mywikimap.com
Example of Map Mash up (Gasoline prices).
http://bbs.keyhole.com
Google Earth Community Bulletin Board.
http://bbs.keyhole.com
MapQuest, an AOL company.
http://worldwind.arc.nasa.gov
National Aeronatics and Space Administration, Learning Technologies, World Wind 1.4.
http://www.openstreetmap.org
OpenStreetMap Project.
http://www.telegeography.com
TeleGeography Research, A Research Division of PriMetrica, Inc, Products and Services, Maps.
http://www.visualcomplexity.com
Visual Complexity Gallery of Internet Network and WWW Spatialization.

Mapping, Distributed

W. Cartwright, RMIT University, Melbourne, VIC, Australia

Glossary

Distributed Mapping The use of intranets and the Internet for providing mapping products from image repositories or provided on demand, generated from geographical databases.

Web 2.0 "The transformation of the original Web of static documents into a collection of pages that still look like documents, but are actually interfaces to full-fledged computing platforms" (O'Reilly, 2004).

Web Mapping The use of the World Wide Web as a publishing medium for maps and map-related products.

The revolution to information provision that the Internet, and more particularly the World Wide Web (Web), has spawned has changed forever how information products are viewed. They are now wanted, not demanded, almost immediately. In newspaper terms, this would be described as wanting information 'before the ink has dried', but for digital information this is probably best described as wanting information 'before the data collection sensor has cooled'! Advances in data collection and telecommunications ensure that collected data are quickly and faithfully transmitted. Processing procedures and equipment, map 'construction', and 'rendering' software and geographical information delivery systems now provide the ability to deliver on demand geo information products in almost real time.

With the arrival of the Web, graphical access method to information was made available. The first browser was not all that dissimilar to today's Internet Explorer or Firefox counterpart, and a current day user of the Web could easily adapt to this original manifestation. Some of the early Web mapping packages used text heavy interfaces to list the available mapping inventory. Once the hypertext markup language (HTML) file was 'clicked', the usual means of viewing geographical information was via a collection of scanned maps. Many simple map access sites were developed and, while powerful media access tools were provided, the reliance of just scanned maps somewhat limited their effectiveness. Problems with scanned maps have been identified to be image quality degradation, warping from improper scanning, coarse scanning resolutions, and over reduction that renders many maps unreadable. However, users have accepted these products due to two factors: lower cost (or free) and time (almost immediate delivery of products).

Also, quite early, in Web mapping terms, another new genre of 'published' map was made available, like products from MapQuest. MapQuest has probably produced the most impressive product for finding streets and business locations, especially in the USA. Using this resource, users can pick a country, zoom in to part of it, and then down to street level. If two addresses in the USA are known, then both maps and route instructions can be generated and viewed. While a complete coverage of the USA is available, a sparser street database is available for other countries. Nevertheless, the MapQuest product is a perfect example of Web delivered information.

In general terms, mapping services available on the Web include:

1. map and image collections;
2. downloadable data storages and online map generation services;
3. hybrid products;
4. virtual globes;
5. ubiquitous mapping; and
6. collaborative publishing using Web 2.0.

The subsequent sections of this article briefly describe these products.

Map and Image Collections

The extent of map libraries on the Web can be illustrated by the sheer number listed in the University of Minnesota's web page. It includes details of sites (numerous) in the USA as well as global libraries that provide Web access. Early implementations were scanned collections of images and maps, and the Central Intelligence Agency (CIA) World Fact Book, for example, made available maps of almost any part of the world. The Perry Castaneda Library (PCL) map collection (University of Texas at Austin) provided excellent collections of scanned maps, along with other pertinent information. Another large site to access geospatial information was the Alexandria Digital Library. It focused on the provision of spatially indexed information via the Web. It contains a collection of geographically referenced materials and services for accessing those collections. The project was further developed via the Alexandria Digital Earth Prototype (ADEPT), funded for 1999–2004 by the US National Science.

This type of Web resource is extremely helpful where access to rare or unique maps would otherwise be difficult or impossible. For example, Oxford University's Bodleian Library, a repository to numerous historical artifacts related mainly to Oxford and Oxfordshire, makes available via the Web a number of rare map fac similes. These high resolution scanned images may be used by scholars in papers without the need to formally request copyright clearance.

Downloadable Data Storages and Online Map-Generation Services

Publishers that have traditionally published their infor mation as paper maps and books now use the Web to provide extra information to support their paper publi cations. Initially, digital geospatial information files were made available to access and download online. Web re positories were established by both governmental and private mapping organizations to streamline how these products were marketed, sold, and delivered. These sites were developed with the express purpose of making maps more readily available to the general public and pro fessional map users. They allowed the information to be made available with little cost to the providing organ ization due to the 'hands free' nature of Web delivery.

MapQuest produced a product for finding streets and business locations. This Web resource allowed users to pick a country, zoom in to part of the country, and then keep moving down to street level. MapQuest product is a perfect example of early Web delivered information using the Internet: maps were quickly delivered, but in elegant and bland. They were about information pro vision in a conventional, standard manner.

Later, these services were supplemented, and in some cases replaced, by online map generation services com monly known as Internet map servers. These Web delivered software services are utilized to publish maps, data, and metadata and to provide basic geographic in formation system (GIS) functionality. These systems are referred to as 'dynamic map generation systems', as a map is generated in response to each client request. Map servers deliver map images (generated after a user spe cifies a particular area to map), extracted features from a database, routes and route maps, and complete atlases. Online map generation services are 'built' from a number of essential components:

1. data and databases;
2. servers;
3. web server;
4. map server;
5. client applications;
6. browser;
7. stand alone viewers; and
8. GIS applications.

The system works whereby a user selects desired map features. This request is then sent via the Web to the Web server, which passes this request to the map server. Once the map server receives this request, a map is 'generated' and an image is produced. The image location and file name is returned to the Web server, which the client computer uploads and displays. The mapping system relies on HTML functions, enhanced with JavaScript, as well as advanced HTML clients like ASP, .NET, Cold Fusion, etc. 'Stand alone' systems are also available like ARC Explorer, from Environmental Systems Research Institute (ESRI).

Generally, one of two routes is used to develop Web server services: develop packages using open source software (like the Minnesota Map Server); or employ proprietary software like ESRI's ArcIMS, Autodesk's MapGuide, MapInfo's MapXTreme, or Intergraph's Geomedia Web Map. Using software the Minnesota Map Server requires development using packages like php, whereas commercial software packages are generally 'plug and play'.

Hybrid Products

Combined discrete/distributed products, which publish on the Web on discrete media, such as CD ROM and DVD, as well as paper, have also been developed. In many cases these were developed to ensure that the products could be delivered irrespective of whether an Internet service was available. Also, some product design strat egies were to place the archival product elements on discrete media and delivered 'updates' and current in formation via the Internet. These products include the US Geological Survey's (USGS) National Atlas of the United States, published on both CD ROM and the Web, the Atlas of Switzerland published in five languages using CD ROM, online via the Web and elegantly published in a bound paper atlas and Understanding USA, a paper atlas that had a complementary website that incorporated all of the map images from the paper atlas, but as interactive maps, plus Virtual Reality Modeling Lan guage (VRML) three dimensional (3 D) diagrams that could be viewed as 3 D objects and rotated, zoomed, etc.

Digital Globes

Digital globes are a fairly recent phenomenon. They provide access to information in a manner that simulates how users would use a physical globe, but the image is virtual. They are delivered as stand alone globes or as Web delivered products. According to Riedl, there con sists of three types of digital globes:

1. virtual hyperglobes – visualization of the digital image on a virtual globe body in virtual space;

2. tactile hyperglobes – visualization of the digital image on a physical (touch sensitive) globe body in real space; and
3. hologlobes – visualization of the digital image on a virtual globe body in real space.

The globes can be delivered on desktop systems as nonimmersive systems or as fully immersive systems using high end 3 D VR systems.

An early product was described as 'the ultimate world globe': T Vision, created by ART+COM, Berlin. A large projection screen displayed an image of the Earth, compiled from satellite data. The interface was a 1 m diameter sphere, used like a large trackball, which the user can spin while the large screen keeps pace with appropriate displays. A 3 D mouse gave access to finer resolution. Weather patterns could be viewed in real time and the user can 'fly' to any spot on Earth for a more detailed view. More recently, commercial software pro viders like ESRI make available products like *ArcGlobe* (in 2004 as part of *ArcGIS 9*) and National Aeronautics and Space Administration (NASA) developed World Wind as an open source product. However, the biggest impact in this area of geographical information delivery has been *Google Earth*, a free package from the Web search engine company Google. Google purchased the technology from the little known company Keyhole and then made this product the phenomenon it became in the mid 2000s. It provides detailed maps and images of most of the world, and these images can be viewed as both 2 D and 3 D images. Following the success of Google Earth, now unashamedly an advertising medium, the Microsoft Corporation launched its rival Digital Globe product, Virtual Earth.

Ubiquitous Mapping

Ubiquitous mapping has been built on the concepts of ubiquitous computing, the 'third wave' in computing, where computers are omnipresent and interconnected in a ubiquitous network. We now see this type of computing in the form of handheld personal computers (PCs), mo bile phones, wireless sensors, radio tags, and Wi Fi. Ubiquitous mapping is being delivered via cellular tele phone systems, through the use of wireless Internet zones and sensors that upload current data like train timetables, etc., to users who have subscribed to a service. Such services are deemed to be ubiquitous when a user does not need to 'log in' or actively connect to a service. Ac cording to Morita, ubiquitous mapping provides the ability to create and use maps any place at any time to resolve spatial problems.

New methods for depicting the Earth and its cultural and natural attributes have developed and many other graphic and nongraphic formats have become available to complement maps for presentation. This includes simple in car alphanumeric readouts, audio guided city guides, and short message service/multimedia messaging service (SMS/MMS) perhaps with Global Positioning Systems (GPSs) enabled cellular telephones that deliver dir ectional and locational information to users.

Collaborative Publishing Using Web 2.0

Web 2.0 is the use of the Web by individuals and groups of individuals to provide and share information by util izing the Web in a different manner. The way that users of Web 2.0 communicate differently is that they do not require materials packaged by publishers – they do this themselves, and, they are computer literate, equipped with versatile computers and appropriate software and, perhaps most importantly, ready to use the Web in dif ferent ways.

It has been made possible by three broad technology trends, according to Roush:

1. inexpensive internet access;
2. inexpensive wireless computing devices; and
3. the Web as a platform for personal publishing and social software.

Mapping packages delivered using social software and Web 2.0 include free maps and images for reuse from OpenStreetMap, GPS traces – a public collection of road centerlines – also from OpenStreetMap, a library pro viding a common application programming interface (API) for Google, Yahoo! and Microsoft's javascript mapping APIs from MapStraction and Tile Engine v0.5, a Web mapping engine for insertion into individual sites that present maps with clickable thumbtacks from CivicMaps. The availability of this free software allows individual users to create 'mashups' – collections of maps using the resources provided via the Web.

Web hybrid publishing using API, Web feeds, or JavaScript has added to the genre of Web delivered maps. Mashups allow maps to be produced by mixing services delivered through a third party using a publicly access ible and usable interface or an API. Perhaps the most widely used mapping application is that provided by Google Maps. It provides base maps of almost anywhere that can be used as an 'underlay' for individual anno tation with default symbols or specially created sym bology. Map views are available as topographic or street maps (perhaps the most widely used maps are street maps), imagery (satellite or aerial), hybrid (maps plus imagery), and street (360° views of some locations). Google based applications are also available, for ex ample, Quikmaps and PlaceOpedia. They allow user producers to generate information overlays and map annotations that could be described as 'geo notes'. As well as user producer maps, 'mainstream' publishers are also using mashups for Web publishing. For example, the *New York Times* uses mashups to illustrate its travel sections.

See also: Cyberspace/Cyberculture; GIS, Public Participation; Internet/Web Mapping.

Further Reading

Baard, M. (2003). A connection in every spot. Wired news, http://www.wired.com/news/print/0,1294,60831,00.html (accessed in April 2008).

Bestor, T. (1996). 'Real time globe', p 44. *Wired*.

Morita, T. (2004). Ubiquitous mapping in Tokyo. In *Proceedings of the First International Joint Workshop on Ubiquitous, Pervasive and Internet Mapping*, pp 7 16. Japan.

Riedl, A. (2006). Digital globes. In Cartwright, W. E., Peterson, M. P. & Gartner, G. (eds.) *Multimedia Cartography* (2nd edn.), pp 255 266. Heidelberg: Springer Verlag.

Roush, W. (2005). Social machines. *MIT Technology Review* 108(8), 45 53.

Weiser, M. (1996). Ubiquitous computing, http://sandbox.xerox.com/hypertext/weiser/UbiHome.html (accessed in April 2008).

Relevant Websites

http://maps.civicactions.net
CivicMaps.

http://www.mapstraction.com
MapStraction, Javascript mapping abstraction library.

http://www.openstreetmap.org
OpenStreetMap.

Mapping, Non-Western

M. W. Pearce, Ohio University, Athens, OH, USA

Glossary

Cosmographical Map A map of sacred space.
Geomancy The practice of divination or siting based on a reading of the location of spiritual energy or 'qi' in the landscape.
Graticule The spherical network of latitude and longitude lines on the earth.
Performative Cartography Spoken, chanted, gestural, danced, or any other form of ephemeral and embodied mapping.
Processual Cartography A mode of mapping which prioritizes process over product.

Introduction

Non Western mapping is a collective term used in map history to describe the cartographic traditions, indigenous or nonindigenous, of Africa, Asia, Oceania, and the Americas. It replaces 'primitive mapping', the term for any form of mapping perceived to be either foundational to or less developed than European cartography. Instead, the term non Western mapping emphasizes that these practices are separate map traditions in their own right, although they may have either influenced or been influenced by Western mapping. The term tends to refer to mapping from the past rather than the present, and as a result includes a wide variety of cartographic material through history and from every continent.

To categorize this disparate material as 'non Western' is problematic on several levels. By placing the map traditions in a single category, they are stripped of the cultural context from which they derive meaning and usefulness and which identifies them as mapping. Without context, the category is dominated by forms recognizable to Western eyes as mapping, based not only on the visual resemblance of those maps to European maps, but on their political and economic resemblance as well, rather than on their own culturally significant terms. That this category includes at least 2000 years of mapmaking is problematic because the definition of which cultures are 'Western' or 'non Western' is one which quickly erodes over time. Finally, the term casts those separate traditions in the negative, the 'not' Western, perpetuating a Eurocentric perspective of the field.

Approaches to Non-Western Mapping in Geography

With a few exceptions, such as Bruno Adler's 1910 study Karty pervobytnykh narodov ('Maps of primitive peoples'), the history of cartography as written in the West initially included little regarding Euro American mapping. Evidence of 'primitive' mapping was anecdotal and partial, and because it was incompletely perceived or understood, non Western mapping was categorized with the other partially documented history – that of early European maps assumed to be the gestational predecessors of a later, matured European cartographic tradition. As such, non Western map traditions were evaluated in terms of their 'contribution' to a modern, monolithic science of cartography centered in Europe.

This situation shifted in the 1970s and 1980s, influenced by the theoretical contributions of ethnohistorical scholarship in history and anthropology, the postmodern critique of the text and discourse, and postcolonial theory. Brian Harley, influenced by the work of Michel Foucault and Jacques Derrida, began publishing a series of articles on deconstruction and the colonial cartographic enterprise in Western map history. But it was the History of Cartography Project, based at the University of Wisconsin Madison, which built on these theoretical and historical foundations to take the greatest steps toward reversing the marginalization of non Western cartographies from map history.

As originally conceived by Harley and David Woodward, the project was to write a complete history of cartography in four volumes, with *Volume 2* devoted to non Western mapping. In their words, Harley and Woodward's mission included the goal "to redefine and expand the canon of early maps" (Woodward and Harley, 1994) by taking a global view of the mapping impulse in non Western societies. They sought to improve on the limitations of histories of cartography available at that time, written by Leo Bagrow, Lloyd Brown, George Crone, and R. V. Tooley, which either curtailed or ignored the map traditions of Africa, Asia, the Middle East, and the indigenous cultures of the Americas and the Pacific. To achieve this expansion of the canon required the reconceptualization of the very definition of map, mapping, and mapmaking. As they wrote, "the study of early maps in non Western societies cannot be confined to examples mirroring the familiar characteristics of European cartography" (Woodward and Harley, 1994). Rather than compiling research on non Western material culture resembling European cartographic 'forms', Harley and Woodward focused on examples of non Western cultural practices having a mapping 'function' in a particular culture or time period, both to shift the emphasis away from mapping as an exclusively mathematical activity to mapping as both mathematical and cultural, as well as include those maps with little formal resemblance to Western maps.

Between the project's inception in 1977 and the publication of the final book of *Volume 2* in 1998, the editors' theoretical conception of mapping and cartography widened as new theoretical and historical scholarships emerged. One significant change in approach was from a view of non Western map traditions as foundational to European map history, to a view of those traditions as separate and distinct ways of knowing. A second significant change was a shift away from an emphasis on the accuracy of maps, that is, how well do they depict measured angles and distances of Euclidean space, to an emphasis on the usability of maps, or, how well do they work for the function for which they were intended? When the editorship shifted to David Woodward and G. Malcolm Lewis for the volume on indigenous cartography, the definition of cartography was opened wider still to recognize the significance of processual and performative cartographies as modes of mapping central to indigenous societies. As these conceptual widenings fostered the expansion of the global canon of the history of cartography, their single volume on non Western mapping expanded to three books, and by the publication of the third book in 1998, the project had decentered European maps and mapping as the defining forms for map history.

Several other influential studies of non Western mapping have expanded the theoretical foundations of map history as well as non Western historiography. In 1989, David Turnbull's exhibition and book *Maps Are Territories: Science Is an Atlas* explored the similarities between maps and scientific theories as a basis for a comparative discussion of Western and aboriginal mapping. In 1991, Robert Rundstrom clearly articulated the differences between indigenous and nonindigenous mapping as processual versus archival practices in his article 'Mapping, postmodernism, indigenous people and the changing direction of North American cartography' for *Cartographica*. In 1993, Matthew Edney's 'Cartography without 'progress': reinterpreting the nature and historical development of mapmaking', also for *Cartographica*, provided a necessary framework for the conceptualization of Western map history as a blending of different modes of mapping, rather than a singular and monolithic map history based on the progress of mathematical accuracy. And Thongchai Winichakul's 1994 study *Siam Mapped* demonstrated that the creation of national identity could be read in the conflicting views of space between traditional Thai cartographic discourse and Western cartographic discourse.

Despite the definitive contributions of the History of Cartography Project, and the wealth of subsequent scholarship on non Western cartographies, the questions of where to draw the boundaries around the definition of map, which activities constitute mapping, and whether mapping practices across cultures exist at equal or hierarchical levels, all continue to be the subject of debate. Geographers have demonstrated that spatiality is fundamental to culture, and that all cultures have sign systems for communication and representation of spatiality. But are these sign systems all mapping? The boundary delineating which materials or processes may be defined as mapping, and which are more appropriately defined as way finding or spatial awareness, dance or storytelling, continues to be contested. Some geographers, such as Thomas Bassett in his writing on African cartographies for the History of Cartography Project, have demonstrated that as long as a Eurocentric definition of map exists, those traditions which least resemble that definition will remain entirely omitted from the canon of map history. Others, such as Denis Wood in his writing for the *Encyclopaedia of the History of Science, Technology, and Medicine in Non Western Cultures*, maintain that the idea of the map is inherently Eurocentric because the map is by definition an artifact of European scientific ontology and epistemology. David Turnbull, in his work *Masons, Tricksters and Cartographers*, has sought to mediate the question of boundaries by firmly locating comparative cartographic histories in the sociology of scientific knowledge.

As scholars continue to connect previously marginalized or forgotten realms of map history to the global canon, and as indigenous peoples worldwide continue to map their territories using traditional methods, the scope of non Western mapping continues to expand and deepen. For some regions of the world, such as China, there is a wealth of published cartographic history; for other regions, including Mongolia, Sri Lanka, and most of the continent of Africa, little to nothing is known in the West. This imbalance may reflect the depth and extent of cartographic tradition in a particular place, but it also may reflect the extent to which Western scholarship has focused on this place, or the types of questions which have been asked, or the particular field of study through which the mapping has been re represented. The field of non Western mapping is riddled with gaps. As the map traditions of the world are written into the canon and these gaps are filled, and as the very concept of Western mapping and Western science is dismantled, the category of non Western mapping erodes and eventually will cease to exist.

Characteristics of Western Mapping

As John Pickles and David Turnbull have written, that which is referred to as Western mapping (also, modern cartography, or Ptolemaic cartography) is an assemblage of ideas about natural and correct cartographic practices. Based in the larger assemblage of Western science, Western mapping assumes an ontology of space as observable, rational, and measurable, and its correct and natural depiction as a uniform, homogenous abstraction. This abstraction is achieved through the mathematical projection of the map from spherical to planar surface and the control of the distortions of that projection through the use of the graticule as a spherical coordinate system and grid. The grid is also used as a scaling device for measuring and fixing horizontal distances in the map. Western cartography also emphasizes a uniformity of orientation to the cardinal directions within the space of the map, as well as a uniform perspective, either from the decentered or orthogonal view of the grid, or from the converging, linear perspective view of the aerial oblique map.

Projection, grid, and fixity of orientation, scale, and perspective combine to effect a particular visual aesthetic of the cartographic gaze, that is, the quality of omniscience, authorlessness, and modernity which enables and empowers not only sovereignty but also domination and imperialism. Brian Harley, Denis Cosgrove, Matthew Edney, John Pickles, Gillian Rose, and Denis Wood have each explored the agency of this visual aesthetic, the way the map is read, perceived, and used by the map reader and in turn shapes society, as characteristic of Western cartography. Although, as a result of this scholarship, it may now be self evident that Western cartographic characteristics are culturally constructed and epistemologically driven, they remain so pervasive and ingrained in geography as it is both taught and practised that it is perhaps difficult to imagine the alternatives.

Characteristics of Non-Western Mapping

It is in non Western mapping that these alternatives may be found. As James Blaut established, mapping is a universal cultural activity. Yet, each society develops a cartography unique in function, mode, economy, materials, and symbolization. One map may depict the stars, another, the earth, and another, the stars and earth fused together inscribed on a monastery wall, in a palm leaf, scroll, basket, tattoo, or arrangement of sand. A non Western map is an archived record on cloth of taxes owed to a central government, elsewhere it is material left over from a single ceremony, or the gestures of the ceremony itself.

Non Western mapping may share similarities with Western mapping as well. For example, non Western mapping has been incorporated into and become part of Western mapping, as when the mathematical precision of Islamic geodesy provided the foundation for precise geodetic measure in the West. In addition, land survey, cadastral mapping, and resource mapping to support urban planning, engineering projects, and the administration of governments and empires can be found throughout the history of non Western cartographies. Cadastral mapping and land survey played a central role in the administration of the Ottoman Empire from the fifteenth to eighteenth centuries, for example. In China, hydrological resource mapping formed a significant cartographic tradition during the Qing Dynasty. And in 1916, King Njoya of the kingdom of Bamum (present day Cameroon) oversaw a 60 person survey team for a large scale topographical survey of his kingdom.

When Western and non Western mapping are compared side by side, two differentiating factors come forward. The first is the concept of map as a projection from sphere to plane, and the use of the graticule as a coordinate system. Although Islamic cartographers explored and developed Ptolemy's concepts of projection and graticule in their world maps, no other non Western mapping tradition uses these concepts. Though many Asian societies utilized precise mathematical techniques for measuring distance and angle, none of these cultures perceived mathematical cartography as the ideal outcome for cartography. European use of both projection and graticule was known in China at least from the arrival of the Jesuits onward; yet as Cordell Yee has established, Chinese cartographers did not consider the spherical shape of the earth and the use of a graticule to transfer points from that sphere as relevant to their work. Though many non Western cultures made use of grids, the presence of a grid in the map does not necessarily indicate this particular function. In China, maps from the Song Dynasty onward made use of square grids for achieving both a measured and uniform scale across the map, and as a device for alignment of overlapping features between maps, but not as a coordinate system. In Southeast Asia, eighteenth century Burmese government cadastral, military, and administrative maps made extensive use of grids for geographic feature transposition, but not as a coordinate system.

The second general difference is the concept of connectivity between sacred and profane space. Non Western maps are often called cosmographical because they incorporate religious or spiritual geographies with physical, political, and cultural geographies. Although this practice also infused mapping in late medieval Europe, part of the gradual transition to what became known as Western cartography during the fifteenth and sixteenth centuries was the removal of sacred geographies from the space of the map. The connectivity of the sacred and profane can be found throughout the non Western world (although again the exception is Islam). As with the concepts of projection and graticule, mapmakers' exposure to Western maps in which such separation was depicted, did not necessarily result in the adoption of that practice. Throughout Asia, European maps depicting nonsacred, abstract space were considered to be in accurate. The concept of sacred and profane space as inextricable has shaped the great map traditions of the non Western world, including the mandala maps of Tibet, the aboriginal dreamings of Australia, the geo mantic maps of China, the Luba lukasa maps of the Democratic Republic of Congo, the shapes and forces maps of Korea, and the Jain tirtha maps, to name only a few. Islamic mapmakers incorporated the spiritual in another sense, using mathematical precision to locate the correct direction of Mecca, for example, or the study of geography as a spiritual practice as directed by the Quran.

Fostering this connectivity between the sacred and profane are the map techniques considered to be distinctly non Western in nature, that is, those practices not con ventionally used in the West. Some of these techniques are shared across knowledge systems; others are unique to one particular time and place. Similarities in mapping tech niques across non Western cultures can be seen in indi geneity as it exists on every continent, in the spread of cultural influences from region to region as in Asia and Islam, and in shared religious belief systems as in societies practising Buddhist and Islamic cartographies. Those mapping techniques uniquely conventional to non West ern cartographies include processual mapping through dance and chant, the incorporation of the landscape into the map, radiating coordinate system, scale and direction shift, multiple perspectives, mirroring, and incorporation of temporality.

Processual Cartographies

Many non Western map traditions come from indigen ous cultures. Indigenous mapping is performative,

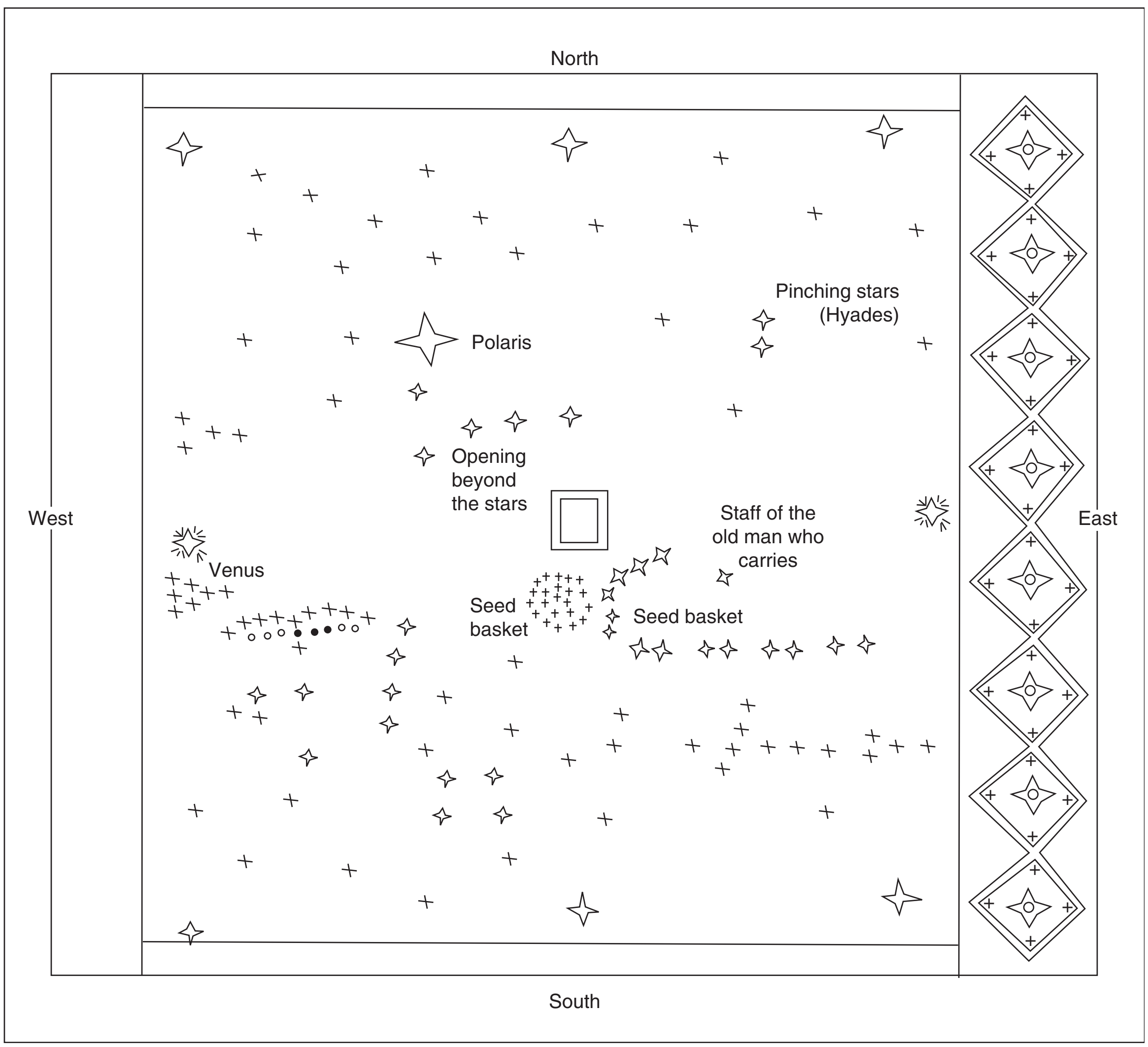

Figure 1 The Sky. Reproduced from Griffin-Pierce, T. (1995). *Earth Is My Mother, Sky Is My Father: Space, Time, and Astronomy in Navajo Sandpainting.* Albuquerque: University of New Mexico Press, p. 121.

emphasizing mapping as process rather than as the arti facts of process. Traditional aboriginal mapping, for ex ample, is primarily a performative cartographic process, and the maps on bark and sand or any other object which come from these performances are intended to be temporary. In West Africa, Bozo elders mapped the Niger River watershed on the ground during ceremonies marking the New Year. In North America, the Dine Navajo mapping in chantway ceremonies is created from sand and destroyed at the end of the ceremony. The drawing in **Figure 1** reconstructs the sandpainting 'The Sky' from Big Starway ceremony.

In performative mapping, the map itself may be en coded through dance. For the Barasana people in Colom bia, cosmographical spatial representation is incorporated into clothing and ornamentation and choreographed to represent celestial geography as it is danced in the long house. In the Arctic, Saami people inscribed their clothing and drums for dances with cosmographical maps of spir itual and migratory travels. Processual mapping may also be encoded through chant, as in the map traditions of Arawakan speaking peoples. For the Arawakan Wakuénai, mapping through the *málikai* chants create the community through naming places and the events of the ancestors Kuwái and Amáru. In Melanesia, toponymy is mapped through chant in specific sequences and directions, as well as through the incorporation of body paint and shell or nament arrangements and the choreography of the dances.

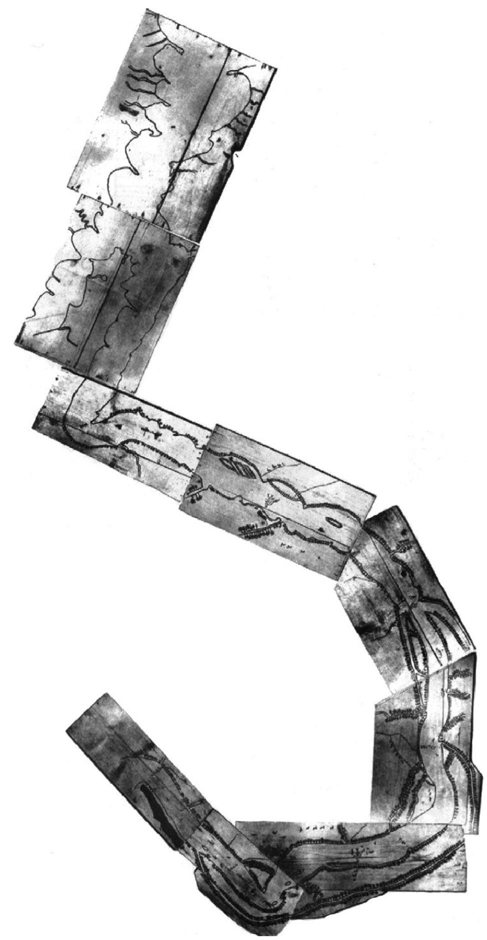

Figure 2 Chukchi map boards joined together. From Woodward, D. and Lewis, G. M. (eds) (1998). *The History of Cartography. Vol. 2, Book 3: Cartography in the Traditional African, American, Arctic, Australian, and Pacific Societies*. Chicago: University of Chicago Press, p. 343.

Landscape as Map

Indigenous processual cartography also often draws the landscape into the map, or incorporates the landscape as a cartographic archive. Indigenous place naming, and the recollection of those place names by visiting and using mnemonic devices and stories to reconstruct the se quence and connections between place names is a mode of mapping using landscape as the vehicle for the map. Siberian peoples inscribed maps on birch bark and placed them in the landscape as directional guides like map kiosks, or painted the maps directly on trees. These message maps depicted hydrology in aerial perspective, with hunting or paddling scenes depicted in profile to show the shapes of animals and vegetation. A more permanent version of these maps can be seen in Chukchi and Mansi peoples' wooden engraved maps (**Figure 2**). Among Algonkian speaking peoples of the North and northeastern regions is the similar tradition of map making on birch bark. These maps conventionally depict the outlines of physical geographic features in aerial perspective and cultural and built features in profile. These maps were used in the landscape for showing the way, like map kiosks, and in the Great Lakes region they were used for the Mide practice of mapping the locations of spirits and path to the spirit world. In aboriginal tradition, maps were painted or engraved on rocks across the landscape by the ancestors. The maps left by these ancestral mapmakers are permanent, not ephemeral, and depict the features of both physical topographies and Dreaming tracks.

Radiating Coordinate System

Incan society used the landscape as a coordinate system composed not of perpendicular meridians and parallels, as in the Western tradition of the graticule, but of radi ating lines. These radiating lines or ze'qes were punc tuated by holy sites or wak'as. A wak'a could be a topographical feature, such as a cave, spring, or field, or a built cultural feature such as a temple, tomb, or carved stone. The ze'qes, mostly straight but occasionally bending to take in different wak'as, were associated with different social groups or provinces and served both re ligious and astronomical functions as well as defined water rights based on kinship. R. Tom Zuidema has

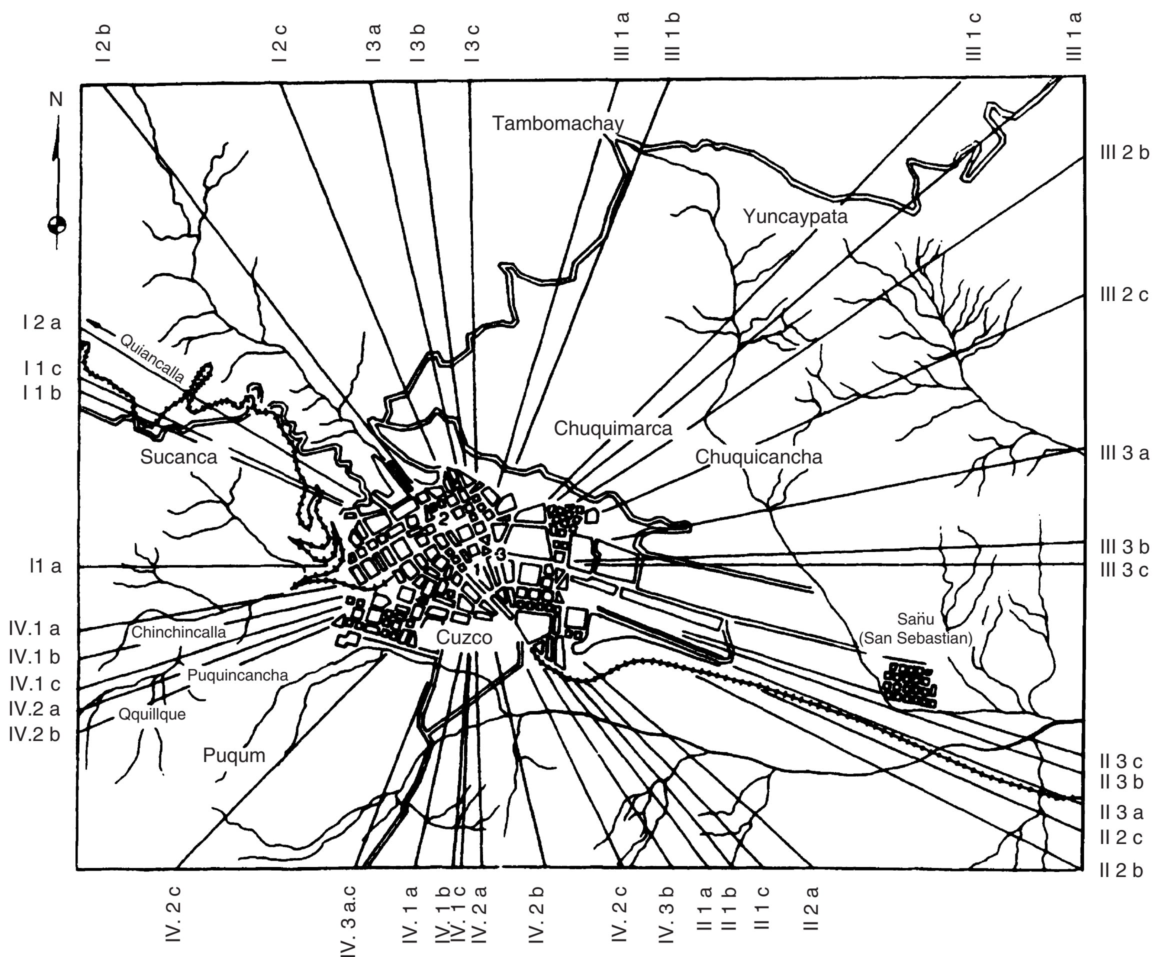

Figure 3 Map of the city of Cuzco with ze'qes superposed. From Aveni, A. F. (1977). *Native American Astronomy*. Austin: University of Texas Press, p. 252.

documented the locational function of the ze'qe system, in particular its use to define the relative directions of those geographical features located beyond the horizon. Though many Andean towns had such a ze'qe system, the largest radiated from the center of the Incan empire at Cuzco (**Figure 3**). In the sixteenth century, Bernabé Cobo recorded 332 wak'as on the Cuzco ze'qes. Together the ze'qes and wak'as formed a coordinate system for a map of the Incan empire to be read from Cuzco. The pattern of locational, radiating lines is a system repeated in Incan notation through the portable knotted cords of the khipu. Both the ze'qe system and khipu technologies were existing Andean practices adapted by the Inca.

Scale Shift

Another technique found in non Western cartographies is the use of scale shifts as a means of placing emphasis on important places in the map. This practice can be found throughout Asian cartography, where scale was used as a stylistic device to prioritize certain places in the map. In South Asian, Southeast Asian, and Tibetan cartographies, the shift of scale for either the clearer representation of particular kinds of geographical features, or for emphasis on important places, is a standard attribute of the map. In Tibet, regional maps dating from the eighteenth century depicted a shift to sacred from profane space through selective increases and decreases in scale across the map, as in an eighteenth century map of sacred places in Central Tibet (**Figure 4**). Shifts in scale do not neces sarily indicate a disregard for measured distances. For example, several extant regional maps were completed in Delhi and Agra during the rule of the Mughal emperors, including one from the late seventeenth or early eight eenth century depicting about 500 named features for what is now Pakistan, with parts of India and Afghanistan (**Figure 5**). In this map, all distances between settlements are given, for military logistics, yet both scale and dir ection vary within the image.

Direction Shift

Direction shift is another technique for depiction of the shift from the sacred to the profane. Direction shift for

Figure 4 The Potala and other principal holy places of Central Tibet. Musèe National des Arts Asiatiques Guimet, Paris. Reproduced from Woodward, D. and Harley, J. B. (eds.) (1994). *The History of Cartography. Vol. 2, Book 2: Cartography in the Traditional East and Southeast Asian Societies*. Chicago: University of Chicago Press, plate 34.

sacred space is a technique of eighteenth century Tibetan regional maps, and in South Asia, Kashmiri regional maps depicting the Vale of Kashmir use the convention of shifting orientation and perspectives. This practice is also found in East Africa, where Tigrean maps portray the physical and cosmographical landscapes within the same map using concentric circles like other African carto graphies, but incorporating within those concentric cir cles a combination of multiple direction orientations. In the Tigrean map in **Figure 6**, the outer circle is oriented to the north, the inner circle is oriented to the west, and the wind rose is oriented to the east.

Multiple Perspectives

The use of perspective in ways other than the Western usage of both the orthogonal as well as linear converging perspective also characterizes non Western mapping. A common technique is the depiction of linear features from an aerial perspective, and point features from a profile perspective. This differentiation of feature sym bolization using perspective is common in traditional cartographies throughout the world. But other instances of perspective shift in non Western mapping include the use of multiple perspectives, the use of linear perspective

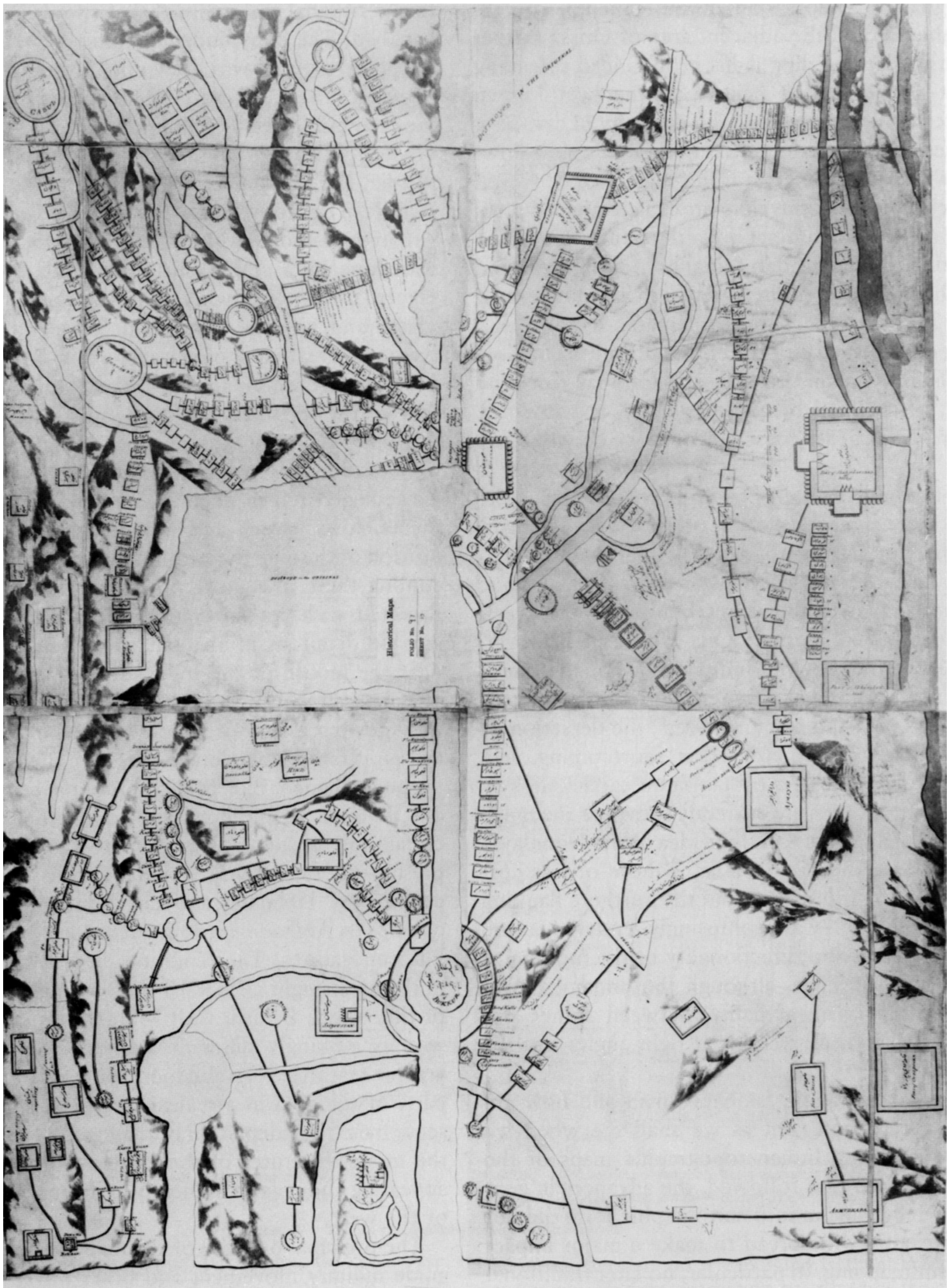

Figure 5 Copy of a Mughal map in two sheets, *c.* 1792 95, from an original cc 1650 1730. National Archives of India, New Delhi. (F. 97/10,11.). Reproduced from Woodward, D. and Harley, J. B. (eds.) (1987). *The History of Cartography. Vol. 2, Book 1: Cartography in the Traditional Islamic and South Asian Societies*. Chicago: University of Chicago Press, p. 406.

diverging rather than converging, and the use of sub terranean perspective.

Across the traditional Asian cartographies, the use of a single perspective was considered poor technique; in stead, Asian cartographic language emphasized the use of multiple perspectives within the same map as natural and correct. Chinese mapping incorporated the perspective techniques of landscape painting, including the concept of variable viewpoint as in many maps of the nineteenth century, including the *Tai Shan Quan Tu* (Complete map

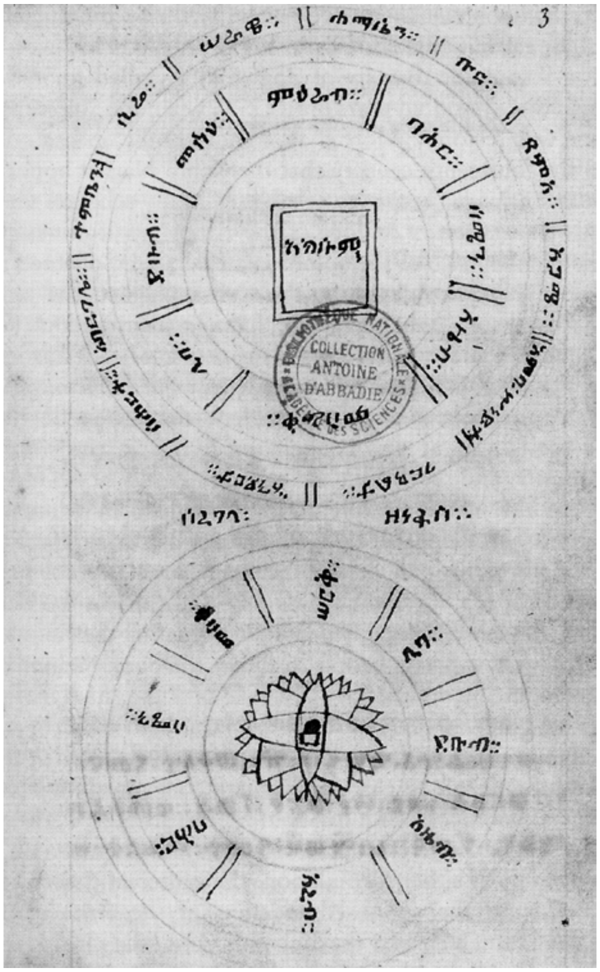

Figure 6 Tigrean Circle map and wind rose. Collected by Antoine Thomas d'Abbadie. Bibliothèque Nationale du France, Paris, Collection Antoine d'Abbadie, 1859 (no. 225, fol. 3). Reproduced from Woodward, D. and Lewis, G. M., (eds.) (1998). *The History of Cartography. Vol. 2, Book 3: Cartography in the Traditional African, American, Arctic, Australian, and Pacific Societies*. Chicago: University of Chicago Press, p. 29.

Figure 7 Rubbing of a stone map, Tai Shan Quan Tu (complete Map of Tai Shan [Mount Tai]). Field Museum of Natural History, Chicago (235581). Reproduced in Woodward, D. and Harley, J. B. (eds.) (1994). *The History of Cartography. Vol. 2, Book 2: Cartography in the Traditional East and Southeast Asian Societies*. Chicago: University of Chicago Press, p. 146.

of Tai Shan) (**Figure 7**). In this map, features are represented in oblique perspective yet not from any single, fixed viewpoint as in the technique of oblique perspective in Western mapping. The eighteenth century regional geographies of Tibet also shifted perspective depending on what was represented and made use of divergent, as opposed to convergent, perspective in oblique views. In the map of the Potala at Lhasa in **Figure 8**, divergent perspective aids the pilgrim to Lhasa by mapping the pilgrim's perspective at any given point around the Potala.

In South Asia, many cosmographic maps combining aerial and profile perspectives center on Mount Meru and map features in vertical space, rather than the Western convention of depicting features in horizontal space. Buddhist mapping in Southeast Asia also uses the technique of combining profile and aerial perspectives in vertical space for maps centered on Mount Meru, such as in the example from the *Trai Phum* (Story of Three Worlds), a fourteenth century accordion style Thai manuscript which includes a map of Asia (**Figure 9**). Kashmiri regional cartography also uses shifting perspectives to depict the Vale of Kashmir.

The aboriginal maps known as the *dhulaŋ*, or bark maps, which combine detailed physical landscapes with *djalkiri*, the ancestral Dreaming tracks, also make use of a particular technique of perspective. The mapped Dreaming tracks depict combinations of the political, cadastral, and spiritual landscapes; in the *dhulaŋ*, these tracks may be symbolized from either the aerial perspective or from the subterranean perspective, that is, the perspective of the spirits inside the earth looking up to

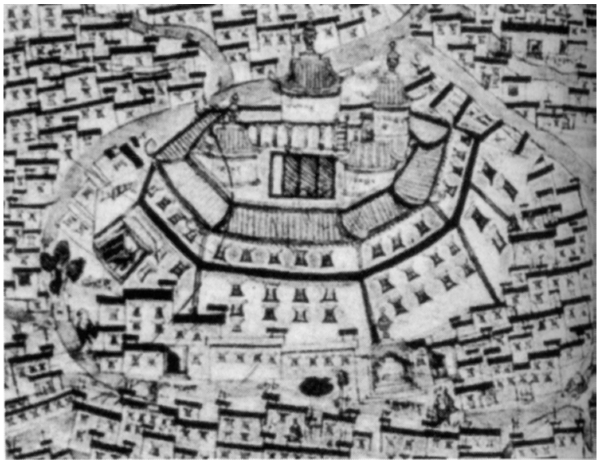

Figure 8 The Potala of Lhasa. From Waddell, L. A. *The Buddhism of Tibet, or Lamaism, with Its Mystic Cults, Symbolism, and Mythology, and Its Relation to Indian Buddhism* (2nd edn.). Cambridge: W. Heffer, 1934 rep. 1958, facing 287.

the surface. A shift to subterranean perspective is again a shift to sacred space: the map portrays the world as seen through the eyes of the ancestors.

Mirroring

In the Plains region of North America, the Pawnee, Lakota, and Cheyenne peoples are well known for their mapping tradition which emphasizes the connectivity between the celestial and terrestrial. Celestial geography is a map of terrestrial geography, and vice versa, and each is the home of the spirits. This mode of cosmographical mapping, called mirroring, can be seen in a Pawnee star chart or a Cheyenne hide. The structure of the tipi and the layout of a Lakota community also mirror celestial geography. In South America, Desana people use mir roring in their mapping of celestial features on the physical, social, and built features of the landscape. Mirroring is also used in West African cartography, where the Dogon cosmographical maps on cave walls can also be found in other Dogon design forms such as weaving, and in the layout of villages.

Incorporating Temporality in Mapped Space

The incorporation of temporal space into the geo graphical space of the map is another non Western technique. In southwest North America, the incorpor ation of calendrical time with spatial relationships is seen in Kiowa cartography. The Kiowa tradition of painting on hide depicts the passage of time through inscriptions for each annual cycle; these inscriptions often include mapping. Indigenous peoples of Mesoamerica produced what were later called 'lienzos', maps which combined the cultural and historical landscapes of a town, including the migration and political histories, on painted cloth.

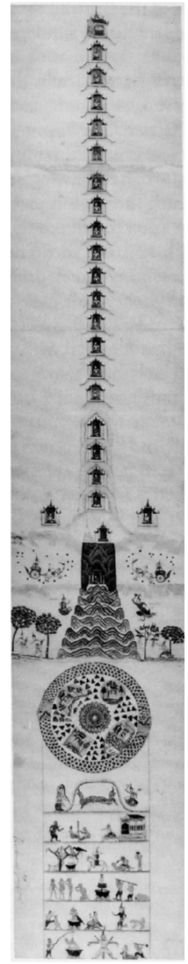

Figure 9 The vertically arrayed universe. Oriental and India Office Collections, British Library, London (Add. MS. 27370, fol. 5). Reproduced from Woodward, D. and Harley, J. B. (eds.) (1994). *The History of Cartography. Vol. 2, Book 2: Cartography in the Traditional East and Southeast Asian Societies*. Chicago: University of Chicago Press, p. 719.

Lienzos depicted the relative locations of geographic features and their place names, and de emphasized measured distances. The *Codex Xolotl* (*c.* 1540 C.E.) and the *Mapas de Cuauhtinchan* are well known lienzos. Mix tec community maps included detailed genealogical in formation, as in the lienzo of Zacatepec (*c.* 1540–60) (**Figure 10**). In this map, place name symbols anchor the perimeter of the central rectangle, while three gener ations of royal genealogy move through the entire map, fusing history and geography in the same space. In digenous peoples of Mesoamerica continue to map with lienzos today.

Regional Scholarship and Collections

For the Islamic cartographies of North Africa and the Middle East, the foremost scholar is Fuat Sezgin, and his

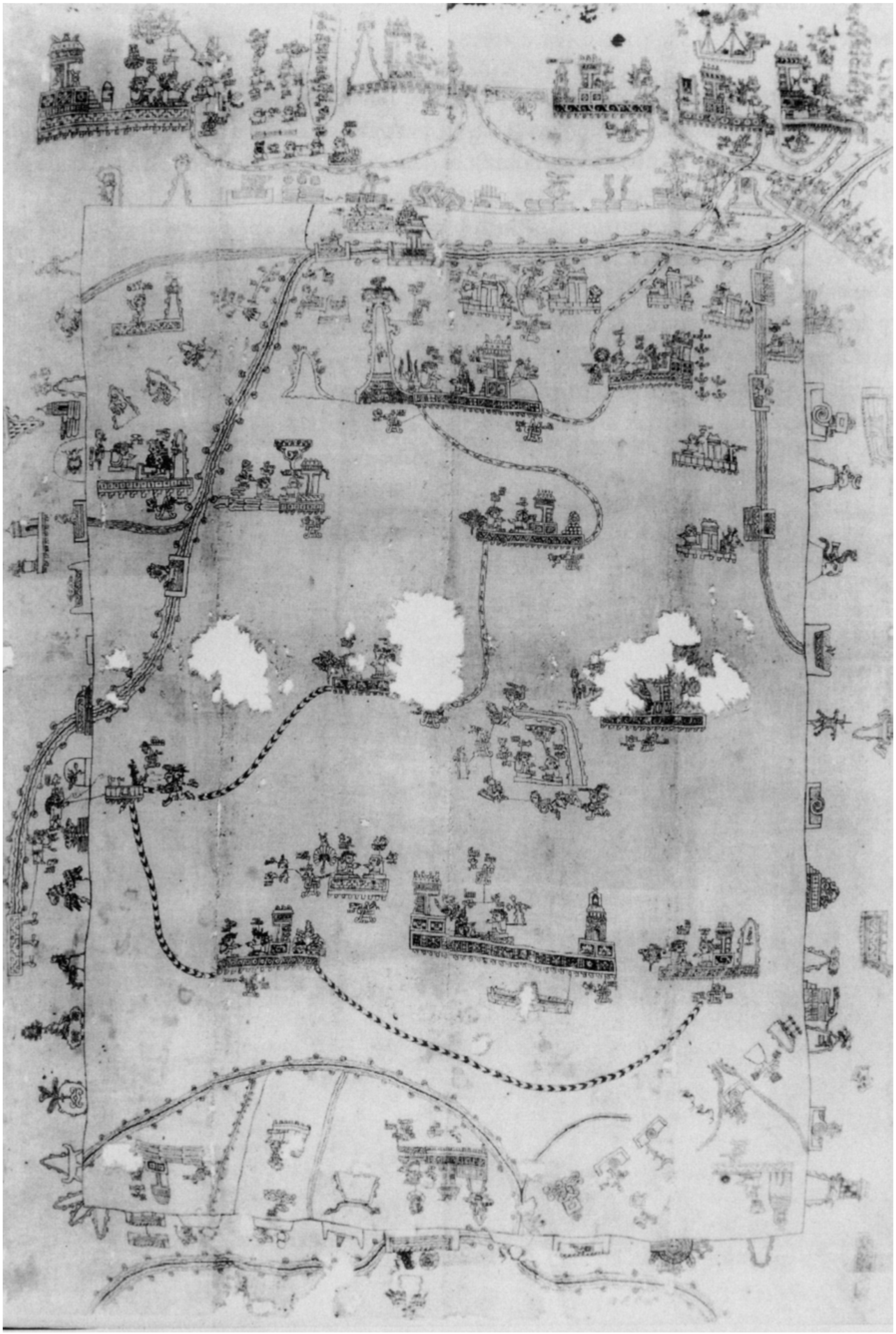

Figure 10 The Lienzo de Zacatepec, *c.* 1540 60. Museo Nacional de Antropología, Mexico (No. 35 63). Reproduced from Mundy, B. E. (1996). *The Mapping of New Spain: Indigenous Cartography and the Maps of the Relaciones Geográficas*. Chicago: University of Chicago Press, p. 109.

Vol. 11 of the *Geschichte des Arabischen Schrifttums*. Other scholars of Middle Eastern and North African map his tory include S. Maqbul Ahmad, Youssouf Kamal, Ahmet T. Karamustafa, Konrad Miller, J. M. Rogers, Emilie Savage Smith, Svat Soucek, and Gerald R. Tibbetts. Collections of Ottoman maps are in the Topkapi Sarayi Müzesi Kütüphanesi, Istanbul, and the British Library, London.

Major scholars of East Asian map history include Funakoshi Akio, Yi Ch'an, Edouard Chavannes, Chen Cheng siang, Mei ling Hsu, Miyajima Kazuhiko, Gari Ledyard, Joseph Needham, Jeon Sang woon, Joseph E. Schwartzberg, Nathan Sivin, F. Richard Stephenson, Kazutaka Unno, Cao Wanru, Cordell D. K. Yee, Marcia Yonemoto, Wang Yong, and H. Yoon. Major collections of Tibetan mapping are the Wise Collection of the Oriental and India Office Collections, British Library, and the Harrer Collection, Völkerkundemuseum der Universität Zürich; for the cartography of Japan, the Mitsui Col lection, East Asian Library, and University of California at Berkeley.

Known major collections of South Asian cartography include the Maharaja Sawai Man Singh II Museum in Jaipur, and in London the British Library and the Hodgson Collection of the India Office Library and Records. Major scholars include B. Arunachalam, Susan Gole, Harka Gurung, Irfan Habib, Joseph E. Schwartz berg, and Maya Prasad Tripathi. For Southeast Asia, scholars include Truong Buu Lam, Joseph E. Schwartz berg, and John K. Whitmore. Collections include the Royal Commonwealth Society Library, The British Li brary, National Archives of India, and the collection of U Maung Maung Tin, Mandalay.

Major scholars of aboriginal, Maori, and Oceanian map history include Phillip Lionel Barton, Giselle Byrnes, William Davenport, Ben Finney, David Lewis, Nancy Munn, Eric Kline Silverman, Peter Sutton, and David Turnbull. Major collections of Marshall Island stick charts include the Museum für Völkerkunde in Berlin, and the Godeffroy Collection at the Museum für Völkerkunde in Hamburg. Major collections of historic aboriginal maps include a collection solicited by Norman B. Tindale between 1930–1953 and housed at the South Australian Museum; the Mountford Sheard Collection at the State Library of South Australia, collected from Central Australia by Charles Pearcy Mountford in 1940; and drawings collected by Ronald Murray Berndt between 1939 and 1980s housed at the Berndt Museum of Anthropology at the University of Western Australia.

Most of the academic record of indigenous mapping of North America has been shaped and described by G. Malcolm Lewis; other scholars include Barbara Belyea, Elizabeth Hill Boone, Tom Cummins, William Gustav Gartner, Walter D. Mignolo, Barbara E. Mundy, Marion Oettinger, Jr., Benjamin S. Orlove, Richard Ruggles, Robert Rundstrom, Rainer Vollmar, Gregory Waselkov, and Neil L. Whitehead. Scholars of indigenous mapping of the Arctic include Elena Okladnikova, Alexi V. Post nikov, Robert Rundstrom, and G. M. Vasilevich. The first exhibition of native maps of North America, organized by G. Malcolm Lewis and Mark Warhus, was the Amer indian and Inuit Maps Exhibition Project of 1996.

For regional overviews of non Western map tradi tions, a concise alternative to the three books of *Vol. 2* in the *History of Cartography* is offered by the *Encyclopaedia of the History of Science, Technology, and Medicine in Non Western Cultures*, edited by Helaine Selin.

See also: Cartography in Islamic Societies; Cartography, History of; Indigenous Mapping; Maps; Science and Scientism, Cartography.

Further Reading

Blaut, J. M., Stea, D., Spencer, C. and Blades, M. (2003). Mapping as a cultural and cognitive universal. *Annals of the Association of American Geographers* 93(1), 165 185.

Edney, M. H. (1993). Cartography without 'progress': Reinventing the nature and historical development of mapmaking. *Cartographica* 30(2/3), 54 68.

Mundy, B. E. (1996). *The Mapping of New Spain: Indigenous Cartography and the Maps of the Relaciones Geográficas*. Chicago, IL: University of Chicago Press.

Needham, J. (1954 85). *Science and Civilization in China*, 6 vols. Cambridge: Cambridge University Press.

Rundstrom, R. A. (1991). Mapping, postmodernism, indigenous people and the changing direction of North American cartography. *Cartographica* 28(2), 1 12.

Selin, H. (ed.) (1997). *Encyclopaedia of the History of Science, Technology, and Medicine in Non Western Cultures*. Boston, MA: Kluwer.

Sezgin, F. (2000). *Geschichte des Arabischen Schrifttums*, vol. XI. Frankfurt: BRILL.

Thongchai, W. (1994). *Siam Mapped: A History of the Geo Body of a Nation*. Honolulu, HI: University of Hawai'i Press.

Turnbull, D. (1989). *Maps Are Territories: Science Is an Atlas*. Chicago, IL: University of Chicago Press.

Turnbull, D. (2000). *Masons, Tricksters, and Cartographers. Comparative Studies in the Sociology of Scientific and Indigenous Knowledge*. London: Routledge.

Woodward, D. (1995). Mapping map history: The History of Cartography Project at the University of Wisconsin Madison. *Wisconsin Academy Review* 42(1), 4 10.

Woodward, D. and Harley, J. B. (eds.) (1987). *The History of Cartography, Vol. 2, Book 1: Cartography in the Traditional Islamic and South Asian Societies*. Chicago, IL: University of Chicago Press.

Woodward, D. and Harley, J. B. (eds.) (1994). *The History of Cartography, Vol. 2, Book 2: Cartography in the Traditional East and Southeast Asian Societies.* Chicago, IL: University of Chicago Press.

Woodward, D. and Lewis, G. M. (eds.) (1998). *The History of Cartography, Vol. 2, Book 3: Cartography in the Traditional African, American, Arctic, Australian, and Pacific Societies.* Chicago, IL: University of Chicago Press.

Relevant Websites

http://classes.bnf.fr/idrisi/index.htm
al Idrîsî la Mèditerranèe au XII^{e} siècle.

http://www.fas.harvard.edu
China Historical GIS, Harvard College Faculty of Arts and Sciences.
http://www.davidrumsey.com
David Rumsey Map Collection.
http://wason.library.cornell.edu
Historical City Maps of Tianjin, Charles W. Wason Collection on East Asia, Cornell University.
http://www.davidrumsey.com/japan/
Japanese Historical Maps, East Asian Library, University of california, Berkeley.
http://angel.library.ubc.ca
Japanese Maps of the Tokugawa Era, University of British Columbia Library.
http://benares.uni hd.de
Kailasanatha Sukula's Mirror of Kasl.
http://www.lib.kobe u.ac.jp/dlib/e index.html
Maps from the Sumida Maritime Materials Collection.
http://jpimg.digital.archives.go.jp/kouseisai/index e.html
National Archives of Japan, Digital Archives, Maps.
http://www.soas.ac.uk
Needham Research Institute.
http://islamic cartography.blogspot.com
Photos of Islamic Cartography.
http://www.prep.mcneese.edu/engr/engr321/preis/piri r∼1.htm
The Piri Reis Map Project.
http://virtualshanghai.ish lyon.cnrs.fr
Virtual Shanghai: Shanghai Urban Space in Time.
http://www.smith.edu
Vistas: Visual Culture in Spanish America 1520 1820, by Dana Leibsohn and Barbara Mundy, Smith College.

Mapping, Philosophy

C. Perkins, University of Manchester, Manchester, UK

Glossary

Cartography Practices, processes, and discipline concerned with transforming and representing a real world in a mapped form.

Epistemology The study of knowledge and its construction.

Ontology The study of being and existence.

Semiology The study of signs and sign systems and their cultural and social connotations.

Traditional Representation This is something that stands for something else and is a medium through which meaning is communicated. In post-structuralist/postmodernist philosophy, however, representation involves the deployment of power, rather than any necessary semiotic link to something else.

Introduction

Mapping is a way of knowing the world and of bringing the unknown into the realms of the known. Jim Blaut's work shows how, throughout history, mapping has played an important role in human culture, serving as one of the defining characteristics of being human, a shared way of thinking, a way of relating to the world outside, and a universal language. Histories of cartography have sought to understand these changing ways of mapping and have also been imbued with different philosophies. For example, Denis Cosgrove's work has explored a genealogy of how people have imagined and represented the globe over the ages. Meanwhile, cartography emerged in the late twentieth century as a specialist discipline, deploying different ways of understanding these processes. Ideas and philosophies change; the ways we imagine the world and mapping change with them.

We all map things out, and it matters whether you are on the map or not: mapping in language is about claiming, making a statement, controlling, representing, and purporting to know the world. Mapping stands between us and the world out there, but it may also be understood as reflecting and embodying different approaches to that world: there are analogies between knowledge and mapping. Both are argued over and communicated; both are employed to carry out work; both are creative; both are enacted; both rely upon a belief system; both are shared and social, etc. These analogies are complex and contested. They rest upon the relations of mapping and philosophy.

Philosophy matters because of the nature of mapping, its assumptions, social and technological contexts, esthetics, politics, and practices. Mapping is an everyday activity and tool, and it has also been employed and theorized in different ways in different times and places. There are different ways of understanding the relations between mapping practice, the world being mapped, the map as an entity, and the human capacity to engage in mapping, and it is these differences that form the focus of this article. The many different audiences and voices speaking about the map's significance partly explain the rich variation in the ways in which it has been understood. The ideas underpinning mapping have themselves shaped different peoples' knowledge of the world, altered how they have employed the map as an artifact, and engaged in mapping behavior in different contexts.

Philosophical difference underpins the practical application of mapping and how we understand mapping as an activity. For example, adopting a scientific approach to mapping allows different insights to be reached from those that might flow from a more critical form of knowledge in which the social function of mapping is emphasized. Within scientific approaches, different assumptions about the ontological status of the map can lead to profoundly different approaches to mapping: the behavioral scientist will emphasize human factors, the software engineer might stress data organization; the information visualization specialist might emphasize the ways in which symbology operates. Difference also resides within social theoretical critiques of mapping: an emphasis on deconstruction is likely to lead to different maps and understanding, from those that emerge from an anthropological concern with practice.

This article discusses the implications of different philosophical approaches to mapping, exploring the ways in which people have sought to understand the medium and its practices, and relate these to the world of ideas. It explores the ends (and beginnings) of mapping, and its relations to cartography as a discourse, by simplifying and generalizing complex and difficult arguments into a map of the field. It starts with a key to the concepts and explores the epistemological and ontological status of mapping, and some of the different strands of thought that might be relevant to this process. The rest of the article details how these principles have directly and indirectly contributed to the multifaceted practices and products associated with mapping in different times and places.

The author's argument is inevitably partial. It offers a guide, rather than a definitive picture. Illustrations and examples are employed to make issues easier to follow. Just as the best maps are personal statements, so the intention here is to tell the author's story of the relations between philosophy and mapping, exploring the genealogy of the ideas, assessing the contributions of key thinkers, classifying and simplifying, but also stressing how human, complex, and relational the map has been as part of our material culture. By revealing the assumptions underpinning different ways of mapping and different ways of understanding mapping in different times and places, we can begin to understand the significance of wider intellectual traditions and explore direct links between the world of ideas and the world of the map.

Epistemologies and Ontologies

Mapping is not only epistemological but also deeply ontological. It is both a way of thinking about the world – offering a framework for knowledge – and a set of assertions about the world itself – resulting in things with an ontic, or ontogenic status. Philosophers and theorists have written about the significance of mapping, but mapping itself produces knowledge and has an ontological status. This philosophical distinction between the nature of the knowledge claims that mapping is able to make, and the status of the practice and artifact itself, is fundamental to any consideration of the relations between mapping and philosophy. But what creates difference? How and why is mapping problematic for those who think about its status?

To answer these questions is to frame subsequent arguments about mapping and philosophy. Here the author would be introducing three ways of approaching these questions, by considering some of the dimensions across which philosophical differences are constituted; by situating Habermas's notions of knowledge domains vis à vis mapping; and by reflecting upon more recent ideas of the day to day operation of mapping as an assemblage of knowledge communities. It should then be possible to explore how these ideas have been worked through in practice, by considering the context in which philosophies have been used. Following Gillian Rose, this means focusing upon sites through which meaning might be constructed: in mapping itself, and in consumption and production practices; and also through the modalities of technology, composition, and social context which strongly influence what mapping comes to mean.

Negotiating the Binaries

A useful way of understanding these questions is to explore some of the dimensions across which philosophical debate is made. **Table 1** illustrates some important binary distinctions that strongly influence views on the epistemological and ontological status of mapping: judging a philosophy against these distinctions provides an often unspoken set of rules for knowing the world, or in the present case, for arguing about the status of mapping. These distinctions are clearly implicated with each other; an emphasis upon the map as representation, for example, is also often strongly associated with the quest for general explanation, with a progressive search for order, with Cartesian distinctions between the map and the territory it claims to represent, with rationality, and indeed with the very act of setting up dualistic categories. By exploring how these dimensions work, we can begin to explain how and why mapping and philosophy are related in such complex ways.

The mind body distinction is often a fundamental influence on how people think about the world. If the mind is separate from the body, then instrumental reason becomes possible: the map can be separated from the messy and subjective contingencies that flow from an embodied view of mapping. Science and reason become possible and a godlike view from nowhere can represent the world in an objective fashion, like a uniform topographic survey. On the other hand, assuming a unity of mind and body focuses attention on different more hybrid and subjective qualities of mapping, rendering problematic distinctions between the observer and observed.

Table 1 Rules for knowing the world: Binary opposites around which ideas coalesce

Mind	Body
Rational	Irrational
Empirical	Theoretical
Absolute	Relative
Nomothetic	Ideographic
Ideological	Material
Subjective	Objective
Knower	World
Random	Ordered
Individual	Society
Regulation	Freedom
Text	Context
Map	Territory
Structure	Agency
Process	Form
Production	Consumption
Representation	Practice
Real	Hyper-real
Holistic	Atomistic
Functional	Symbolic
Vague	Precise
Immutable	Fluid
Essence	Immanence
Static	Becoming
Network	Node
Nature	Culture

The question of whether knowledge is unique or whether the world might be subject to more general theorizing also has fundamental implications for mapping. An ideographic emphasis on uniqueness has frequently pervaded theorizing about mapping in the history of cartography: if each map was different, and described a unique place, searching for general principles that might govern design, or explain use, or direct production would be doomed to fail. Instead, mapping became the ultimate expression of descriptive endeavor, an empirical technique for documenting difference. Artistic approaches to mapping that privilege the subjective may be strongly compatible with this kind of interpretation. On the other hand, a more nomothetic approach which emphasizes laws, denies difference, and risks reifying artificially theorized models or generalizations, while at the same time offering the possibility of progress. Debate continues around the nature of generalizations about mapping, whether they are concerned with holistic approaches or break up systems into parts, whether they are stochastic or regular, invariant or contingent, natural or cultural, objective or subjective, functional or symbolic, etc. It is clear, however, that since World War II, a number of different scientific orthodoxies have pervaded the world of mapping that almost all trade on this notion of searching for common explanations. The formal discipline of cartography has emerged from this process. Yet the paradoxically everyday ideas of geography and mapping as factual knowledge survive, and reflect the continuing power of ideographic and empirical thought.

An emphasis upon mapping as a text or discourse emerged in the early 1990s, in stark opposition to the more practical and technologically driven search for generalization. Social theoretical ways of understanding mapping often emphasize the discursive power of the medium, stressing deconstruction, and the social and cultural work that mapping achieves. The power of mapping becomes more important than the empirical search for verifiable generalization. A textual emphasis almost inevitably also draws attention to the need for social and historical contextual study.

Structural explanations of the significance of mapping have also strongly influenced understandings of mapping. Insight might stem from class relations, from cultural practice, from psychoanalysis, or linguistics: for example, semiotic approaches to mapping have been a powerful and influential way of approaching the medium and its messages. There is an ongoing debate in relation to mapping over how the agency of an individual might be reconciled with this kind of approach, given that structural approaches often posit fundamental and inevitable forces underpinning all mapping. There is also a continuing debate over the philosophical basis of the structural critique. For example, is it grounded in a materialist view of the world, or in a more ideological reading of the human condition? A clear tension also exists between explanations of mapping that value the local and the subjective, as against those that value objective and universal explanations. There may also be conflicts between philosophies that argue for mapping knowledge as a social force, as against those that depend upon more humanistic and individual ways of knowing the world. Understanding mapping as a form of regulation and surveillant force may not be compatible with envisaging mapping processes as emancipatory and empowering for an individual.

The distinction between forces making the world, or forces consuming it, also has a strong resonance in philosophical debates around mapping. The cultural turn described elsewhere in this encyclopaedia has encouraged a growing emphasis upon the contexts in which maps operate, encouraging a shift away from theorizing about production and toward philosophies of mapping grounded in consumption. The map reader becomes more important than the map publisher. Technological change and the democratization of cartographic practice have also encouraged this changed emphasis. Associated with this shift has been the increasingly nuanced drift toward poststructuralist ways of knowing the world, which distrusts all encompassing knowledge claims. Instead of a belief in absolute space, or a socially constructed world, an alternative way of understanding mapping has suggested the relativity of everything, in a universe where notions of reality come to be replaced by simulation and in which the play of images replaces visual work, or in which speed of change itself gains an agency.

The status of the map as an entity also causes debate. On the one hand, many theorists have argued that maps exist and have an essence from which we can derive understanding. Recently, Kitchin and Dodge, on the other hand, have rejected this ontological position in favor of an ontogenic status for the object in which its properties are called into being when it is employed, instead of residing in its materiality. Mapping, in this view, is always becoming, and is always a process. Instead of relying upon a precisely specified notion of space, its inter subjective, relational, and multivalent qualities are important in mapping practice, and these qualities may be vague, rather than precise.

Recent emphasis on nonrepresentational ways of understanding the world has begun to offer another changing way of rethinking mapping. Also drawing upon actions and ideas of immanence, instead of essence, these approaches focus on the practices and processes of mapping in which meaning is constituted, rather than on representations or discourse. Forms are called into being, with mapping being fluid, and meaning being mutable. Mapping is enacted, action and emotion become important, and the affectual qualities of mapping are matter. The narrative of mapping moves its audience and is itself moved by the context. The emphasis of this work shifts away from the nodes or vectors, toward the flows that allow a mapping network to operate. Relations between actors or actants matter more than the phenomena themselves.

The Domains of Knowledge

In practice, some of these dimensions are more important than others in the creation of knowledge claims about the world. Habermas claims it is possible to distinguish be tween three major types of knowledge, each of which makes claims in particular domains, and each of which has very different aims. **Table 2** maps out these dis tinctions and makes links to the kinds of insight about the world that flow from scientific–technical explanation, historical–hermeneutic thought, and critical knowledge.

Scientific knowledge is inevitably instrumental–techni cal, implicitly seeking explanation, and control over the world. A view of mapping as scientific relies upon reason and generates particular kinds of knowledge, while denying the validity of other truth claims. Western technical practice is valorized, and protected from critique, by an ideology that reinforces the teleological power of technical applied practice. Well designed maps have to follow 'good practice'. Good maps 'tell the truth'. Other mapping is marginalized; expert knowledge alone is deemed valid. The neutrality, distance, and objectivity implied in this kind of knowledge are well suited to many of the contexts in which Western mapping and mapping technologies have been employed. It can be useful to deny human agency and reduce complexity to rules that produce results. Much Western mapping adopted these approaches after World War II, and instrumental–technical approaches to mapping remain preeminent among practitioners.

In contrast, historical–hermeneutic approaches to mapping offer interpretation instead of explanation and evoke feelings emotion and insight. This kind of know ledge has never been frozen out of approaches to map ping. It lurks in notions of design quality, in perceptions of mapping, and in the inspirational qualities of the medium, with its power to fire the imagination. Despite the dominance of science, the rise of geographical in formation system (GIS), and the power of the Web, al ternative human centered approaches to mapping are of increasing importance. Artists are reimagining the map, non Western mapping traditions are being rediscovered and reenacted, and the geographical imagination being remapped in diverse fashions.

Table 2 Domains of knowledge

Knowledge type	*Domain*	*Aim*
Instrumental technical	Reason	Control/ prediction
Historical hermeneutic	Understanding, feeling, emotion, etc.	Interpretation
Critical	Disclosure of false beliefs	Emancipation

Critical knowledge also never disappeared during the hegemony of scientific mapping. Here, a normative aim is central to the project, emancipating people, freeing them from false consciousness, and advancing the human spirit. These critical and radical approaches to mapping became reinvigorated in the 1990s when they came to be reflected in a diverse array of theorizing and in a widening raft of counter mapping enacted by communities employing mapping as a tool to reassert their rights.

So philosophies generate different kinds of mapping and knowledge about mapping. It is to recent theorists of the sociology of science that we must turn in order to begin to explore how these differences become estab lished as social practices, and to show how mapping can work in knowledge assemblages.

Knowledge Communities and Mapping

Theorists in science and technology studies have re cently argued that knowledge is constructed and dis seminated through a complex and shared array of social processes that might be termed an assemblage. Actors in this array follow carefully orchestrated ways of doing, whether it is work or play, which are at once taken for granted, but also intensely politically charged. In these contexts, a technology becomes a flow that allows dif ferences to be reconciled. David Turnbull has per suasively argued that mapping works like this: in different periods and cultures the map has served as an artifact around which practices coalesce and in which meaning becomes fixed. So the map has immutable qualities that it can carry into new contexts, signifying, for example, Western colonial values in the process of administering and controlling other cultures and spaces.

In this view of knowledge communities, mapping practices are protected from critique by a complex array of actions, flows, artifacts, actors, and processes, but also constantly renegotiated. The practice of mapping, whether it is scientific–technical, historical–hermeneutic, or critical depends upon the context, but also influences that context. Different assemblages interact: the Western mapping tradition adapts to others it encounters, and also adopts elements of these traditions, approaches, and philosophies. Within that tradition, networks or webs of practice enact different mapping philosophies. In the world of media, mapping time is paramount: maps are ephemeral and constantly change; the Google map hack celebrates local power in the face of the global corporation; the military Department of Terrestrial Magnetism (DTM) enacts the strategic targeting of a technophilic dream; the planner's land use GIS celebrates bureaucratized administrative control; the artist's mapping performance subverts and celebrates humanity and difference. Each artifact is deployed in a network of practice: philosophies are made and remade, rather than existing as de corporalized, *a priori*, ideologies.

So philosophies underpinning mapping practice become a matter of everyday events and transactions. A rule book, an apparently inconsequential ritual, a chance discovery, juxtaposition, and feeling, all reinforce and change ways of imagining mapping. Mapping itself, and philosophies of mapping, have an agency of their own, capable of enacting change. The assemblage, however, is itself not immune from change. It is constantly morphing into something new, changing into new forms, haunted by traces of the past, but always becoming something new.

So the story of changing mapping philosophies always needs to be considered as historically contingent as timed, placed, and cultured. A web of interacting possibilities in which everything is complex, nothing is inevitable, and everything is possible. The rest of this article illustrates this contingency at work, by considering how different mapping ideas have been worked through and enacted in different contexts.

The Practice of Science

Everyday Empirical Mapping Practice

For much of its history, maps have been a touchstone for geography, employed as an objective tool to describe regional distributions, and as a source for geographical ideas. Practitioners of the geographic art needed to learn how maps were produced and designed, in order to be able to employ them in their exploration, trade, administration, or research. From the mapping of the Roman Empire onward, this practical approach to mapping as a tool has been hugely influential.

Following its establishment as an academic discipline in the late nineteenth century, geography adopted this practical approach. Geographers studied technologies of production and design, debated the construction of different map projections, devised new kinds of thematic maps, and interpreted the landscape using topographic surveys. The kind of science underpinning this technology was rarely considered: mapping was simply assumed to be mimetic and factual. Philosophizing about mapping was something best left to philosophers. Mapping was simply a method, a key tool in the geographers' kit, with no politics, and the vast majority of people who make and use maps still mostly assume mapping is 'neutral' and apolitical. This aphilosophical view rendered mapping as a technology, designed through the appliance of science to work for its users.

However, a rather different scientific approach to mapping can also be traced back through the history of cartography. Ancient Greek cartography created practical maps of the Western world, but it also defined some of the scientific principles on which mapping could progress. For example, the spherical nature of the globe was proposed and scientific experiment based upon astronomical observation verified the validity of the assertion. Pythagorean trigonometry was applied to the development of accurate surveying technology. Mathematical descriptions of map projections were developed for the first time. Grids were invented and applied. Claudius Ptolomy's *Geographia* systematized scientific principles into the first published guide to map making. The world was cast as knowable, observable, and subject to representation. Science allowed progress.

This more abstract and theoretical approach to mapping has ebbed and flowed ever since. Technological 'advances' in map making have depended upon scientific knowledge, but scientific knowledge is itself a social creation, with changing values and practices. It is in the period after World War II that these differences in the philosophy of science first became significant for mapping, with the emergence of 'mapping sciences' in a separate disciplinary configuration.

In the 1950s the first sustained attempts began to emerge in the USA, to justify why and how mapping might be scientific. This scientific reorientation was a more political project than an explicit application of any single philosophy of science. Space, following Kant, became a container; geometry and mathematical language were privileged over social or humanistic ways of understanding the world. In this view the history of cartography was a story of progress. Over time, maps had become more and more accurate, knowledge improved, and implicitly it was assumed that everything could be known. The artifact and individual innovation were what mattered. Cartographic research practice never explicitly adopted positivist orthodoxies, and while it may have implicitly employed critical rationalist methods, testing and falsifying hypotheses, few significant working theories emerged. In this taken for granted framework, mapping science was practical and applied. Numerous small advances built a discipline.

Robinsonian Functionalism

The Anglo American discipline of cartography emerged in the decades after the War, employing the rhetoric of 'good' cartographic science, to advance its status. Institutions grew with vested interests in these practices. Separate graduate programs were developed, PhDs proliferated, drawing offices were established in universities, cartographic researchers were appointed, national societies were formed, and a new discipline developed. The

International Cartographic Association was founded in 1959 to help overcome barriers in the profession by encouraging the free exchange of theoretical and technical information, and charged with aims that were both scholarly and scientific in remit.

Probably the most influential scientific approach to the new discipline was advanced by Arthur Robinson, whose *The Look of Maps*, was published in 1952. Robinsonian orthodoxies came to be represented in the much reprinted and classic textbook *Elements of Cartography*. Robinson suggested that an instrumental approach to mapping grounded in experimental psychology might be the best way for cartography to gain intellectual respectability and develop a rigorously derived and empirically tested body of generalizations appropriate for growing the new subject, and for the application of this knowledge in the new world of the Cold War. Robinson adopted a view of the mind as an information processing device. Drawing upon Claude Shannon's work in information theory, complexity of meaning was simplified into an approach focusing on input, transfer, and output of information about the world. Social context was deemed to be irrelevant, the world existed independent of the observer and researcher. The cartographer was separate from the user and a single optimal map could be produced to meet every need. It was assumed that uses of the map could be identified and that the sole purpose of mapping was to reduce error. Framed by an empiricist ideology, this inward facing research agenda aimed to reduce signal distortion. Art and beauty had no place in this functional cartographic universe. Research aims were normative and focused upon improving map designs by carefully controlled scientific experimentation. Issues, such as how to represent location, direction, and distance; how to select information; how best to symbolize these data; how to combine these symbols together; and what kind of map to publish, were worthy of attention.

Cartographic Communication

Out of this context, in the late 1960s and 1970s, emerged an increasingly sophisticated and often over complex series of attempts to develop cartographic communication as a theoretical framework to direct research. **Figure 1** illustrates the emphasis of this approach. The cartographer chooses what to include in a map and how to show it, and the reader interprets this map by relating it to their own prior knowledge. Communication models encouraged researchers to look beyond a functional analysis of map design, exploring filters that might hinder communication. For researchers such as Head, or Schlichtmann, the map artifact itself became the focus of study, with a semiotic focus to the research, instead of a functional emphasis. For Chris Board the map was a conceptual, as well as a functional, model of the world. As models of cartographic communication multiplied, so also attention increasingly focussed on the map reader, with cognitive research seeking to understand how maps worked, in the sense of how readers behaved, and how they interpreted different aspects of mapping. Researchers proposed that mental maps existed, which might be investigated in order to understand environmental cognition. It was assumed that map reading depended, in large part, upon these cognitive structures and processes. In practice, most research continued to investigate filters in the centre of this system, concerned with the cartographers' design practice, and with the initial stages of readers extracting information from the map. Little work addressed either what should be mapped, or how mapping was employed socially. Most research adopted a naively behavioralist approach.

Other strands of scientific research into mapping emphasized the technologies that might be employed. Waldo Tobler's analytical cartography emerged in the early 1970s, offering a mathematical way of knowing the world, and laying the foundations for the emergence of geographic information science. This analytical approach sought progress through the application of mathematical models and the subsequent application of technology so as to create new conceptual bases for mapping the world. The ethos of most Anglo American research, however, still positioned the map user as an individual, apolitical recipient, and the cartographer as a technician delivering spatially accurate value free products, based upon narrowly identified and carefully controlled laboratory based experiments, that gradually and incrementally improved knowledge of how the world operated.

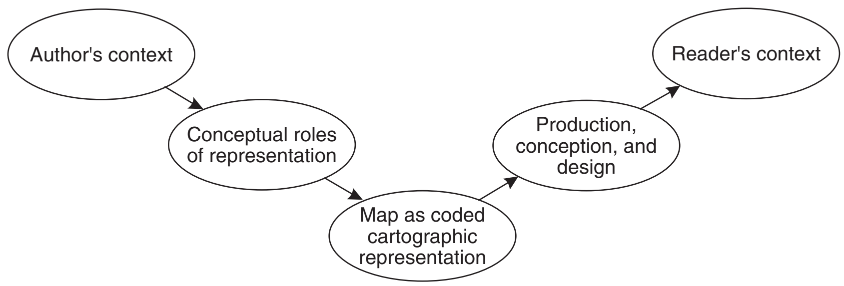

Figure 1 Cartographic communication.

In other contexts, different theoretical positions were adopted. The French disciplinary tradition, for example, was much less influenced by Robinsonian functionalism and empirical research. Semiotic approaches were much more influential in this context, and may be traced back to the influential theories of Jacques Bertin. In 1969, Bertin derived from first principles, rather than from empirical research, a set of visual variables, which might be manipulated by designers concerned with the effective design of mapping and other visualizations.

While Anglo American orthodoxies separated the academic pursuit of cartographic knowledge, from its commercial or strategic practice by publishers or national mapping agencies, cartographic science in the Eastern bloc explicitly sought to apply technical knowledge about mapping to state interests. Here mapping sciences were advanced as part of the socialist triumph over nature. The emphasis was much more explicitly upon producing im proved practical tools, and in particular upon thematic map production, rather than discovering how mapping functioned. The normative goal was to advance revo lutionary progress, through carefully controlled and pre scribed research agendas.

Toward Representation

By the mid 1980s the power of cartographic communi cation as an organizing framework was beginning to wane. Technological changes rendered problematic a single authoritative view of the world at a time when data were becoming much more readily available, and when technologies for the manipulation and dissemination of mapping were also being revolutionized. Users could become mappers: many possible mappings could be made and the map was no longer the end form. Mapping needed to be understood as much more of a process than was possible in communication models.

Digital mapping technologies separated display from printing and removed the constraint of fixed specifi cations. GIS has increasingly supplanted technical as pects of cartography. Digital position, elevation, and attribute data could be captured from remotely sensed sources, and easily stored and manipulated in a digital form. Precision and accuracy increased apace. Imagery could be generated to provide real time updates of changing contexts. The displays of these data are no longer just static maps, and animated displays incorpor ating time into the mapped world are now commonplace. Tobler's analytical cartography has led to a profound technical revolution that also changed ways of thinking about mapping.

We can all choose which patterns to display in these new maps. We can include what we want, and show it how we want. The Web allows these images to be widely shared and disseminated at low cost. Barriers between viewer and viewed are blurring: immersive virtual reality systems allow the user to become a participant in the mapped landscape, instead of an onlooker. These tech nologies are also increasingly mobile: maps on phones or personal digital assistants (PDAs), and incorporating in formation derived from global positioning system (GPS), now plot 'your' changing real location as you move through the world. A whole range of possible ways of seeing the world may now be created by the users of the image, instead of a single carefully crafted view from an authoritative producer. We can all explore 'our own' visualizations, instead of having to accept a fixed image created by someone else.

In the face of these profound challenges, a second hegemonic approach to mapping had replaced cartog raphic communication by the mid 1990s as the scientific orthodoxy. The linear inevitability of communication has been supplanted by a multifaceted and multi layered merging of cognitive and semiotic approaches, centered around representation and strongly influenced by the work of Alan MacEachren. Articulating ideas grounded in Peircean semiotics, this approach recognized the need for a much less literal and functional positioning of mapping. The iconic representation of this approach is the drawing of cartography cubed that appears in **Figure 2**. The dimensions of interactivity, the kind of knowledge, and the social nature of the process show the three key ways in which scientific understanding has been repositioned. Mapping can now be investigated as collaborative, the social context beyond map reading *per se* can be charted, and the process of knowing explored. And mapping is one of many kinds of visualization.

Representation focuses much less on artifacts, and much more upon process. It is grounded in a more realist science, recognizing contingency and context, compatible with a universe where science has to come to terms with uncertainty and where outcome may depend on complex local circumstance. It is, of course, still a scientific ap proach, reliant upon Western ways of seeing and upon technologies of vision. It still depends upon the neutrality of scientific experimentation and a representational view of the world. Other ways of mapping exist beyond this still narrowly defined frame and it is to these that we now turn.

Humanizing Mapping

As a sign system, mapping may be read in different ways by different people and there has been an increasing appreciation that science has to take this social and in dividual context into account, instead of offering uni versal explanations. Meaning matters and depends on culture as well as form, content, technology, or politics. Mapping speaks differently to different people. The

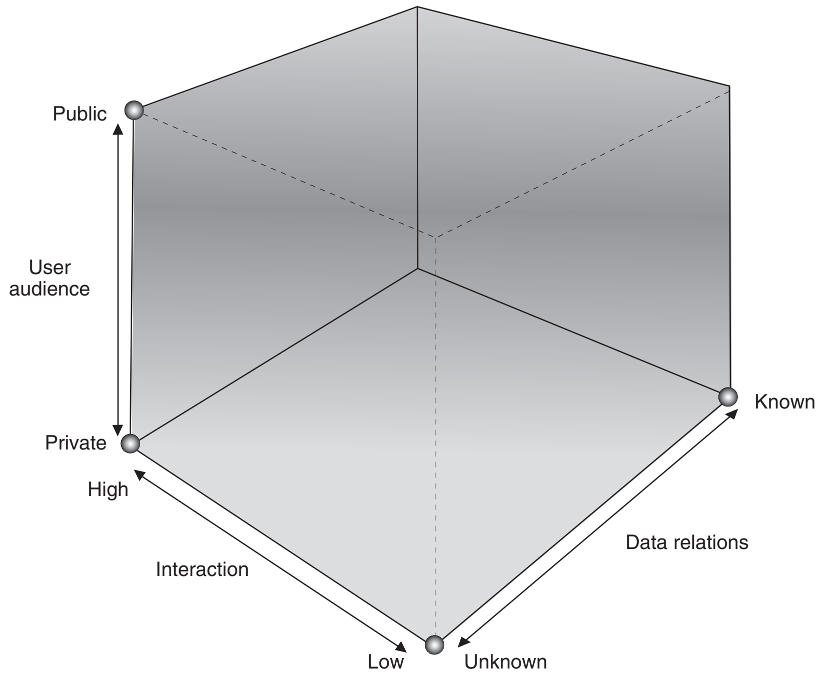

Figure 2 The mapping space of representation: MacEachren's cube.

meanings of the image are constituted in a context that is complex and personal, so the individual can then be cast as an active force in mapping.

Subjective Practices

Cross cultural studies of map skill acquisition suggest that individual mapping skills may be a shared cultural universal, learnt as a practical device in many different cultures, but practiced by individuals according to their own needs. In a long series of empirical studies, Jim Blaut has shown how children acquire these mapping skills in different cultures. A subjective interpretation of mapping can go beyond empirical studies, and be considered a matter of practice and knowledge.

In the practical world of design, it has long been recognized that esthetic factors are important for the communication of information in mapped form. Col lectable antiquarian maps are valued as objects of beauty, acquired for their rarity and esthetic qualities. The visual display of a map is important: maps decorate as well as inform. In a reaction against the functional Robinsonian orthodoxy that dominated much of the practice during the early years of the American birth of cartography as a separate discipline, attempts were made in the 1970s and 1980s to incorporate art into scientific mapping practice. Following the lead of researchers, such as John Keates, scientific practices sought to systematize artistic qualities such as layout, visual hierarchy, name placement, bal ance, and appropriate levels of generalization. It was recognized that subjective combinations of different visual variables resulted in designs that articulated par ticular points of view.

It is also widely recognized that mass production and the corporatization of cartographic practice led to a de cline in design quality and that the rise of GIS and web mapping has exacerbated this retreat from the esthetic. Online maps are mostly functional but ugly. But map users are still able to recognize artistic quality in all mapping: individual, subjective, holistic, and personal qualities of designs are often contrasted favorably with more utilitarian, depersonalized, analytical, and func tional qualities associated with mapmaking as a scientific technology. And aspects of mystery, emotion, and taste are beyond science: map users are able to recognize good designs even if they do not know why these designs are good.

The person reading the map interprets this esthetic according to the context in which the map is discursively positioned, and according to who they are. A vibrant agit prop image on an alternative web blog may well be read differently from a pared down black and white infor mation graphic in *The Times*. Esthetics depend upon the historical and cultural context of the image. At different times and in different places, the worlds of the artist and those of the cartographer have overlapped. In the heyday of decorative cartography in the seventeenth century artists and mapmakers shared similar goals, spaces, and visual devices. Following the rise of mass produced technical mapping, and the twentieth century invention of cartography as a science, a much more separate ex istence became the accepted orthodoxy. In the 1990s, however, artists increasingly began to use mapping as a device to draw attention to the way in which individuals relate to the world. Surrealists, pop artists, situationists, land artists, conceptual artists, community artists, digital media artists, and live artists have all employed mapping in their work. And non Western alternative mapping traditions grounded in practice and subjectivity also challenge the appliance of science. Aboriginal art speaks with a different esthetic to a Western map of the same area. It is not only what is mapped and how features are mapped that dictate how the image is employed, reading the map goes beyond the esthetic. Elements in designs may allow a map to be dated or even placed, but the individual can play a much more active role in this process, and the look of a map immediately encourages individual reflection and feeling.

Humanistic Approaches to Mapping

If mapping practice is imbued with human conceptions beyond science, then the same is true for mapping thought. Phenomenological approaches to mapping reject the narrow certainties of functional Cartesian thinking, and challenge empiricism as a way of explaining the world. Interpretation becomes more important than ex planation; mapping narrates a personal story, one of many possible ways of knowing the world. Interpretive approaches to mapping are often imbued with an esthetic concern. Some theorists have argued that the forms of the visual medium itself are sufficient to interpret. James Elkins, for example, suggests that the relations of writing, notation, and pictorial form within any image may be analyzed, and regards maps as what he terms 'schemata'. Mapping as a medium clearly carries connotations that are separate from other forms, such as pictures, for ex ample. An iconographic approach, also derived from art history, is another useful way of analyzing the subjective connotations of mapping. Three levels of analysis here are possible: the forms of symbols may be described; their meaning read in a literal sense; or alternatively inter preted and related to social function in a tertiary iconographic analysis.

Other humanistic interpretive devices stem from lit erary approaches and a textual view of mapping. Der ridean deconstruction offers one approach to mapping, comprising the identification, and understanding of underlying (but often unarticulated but implicit) as sumptions or ideas, upon which the text rests, and which strongly influence what we think or believe about the mapping. It encourages a consideration of absences as well as presences, focusing upon textual contingencies, paradoxes, and contradictions and argues for rejection of dualistic thinking. John Pickles has argued that the her meneutic circle may be employed to analyze mapping. Mapping may be read in a critical process by exploring the relations between mapping as a text, its language, the cultural context, the authors' world, and the world being represented. Of course, a subjective, polyphonous, and multi vocal map can emerge from this process and tex tual analysis is inevitably interpretive, with meaning constantly being constructed.

These interpretive approaches to mapping share a focus upon the cultural context in which mapping is employed. They increasingly stress the ambiguity of the map reading process, and the need to consider contexts beyond the map. They also tend to reject narrowly structural readings of mapping, valuing instead phe nomenological investigation, and increasingly a nuanced and post structural distrust of universal explanations.

Critical and Emancipatory Cartographies

There has been an increasing recognition that the image of the world can never fully represent the world, because the process of representing is itself part of the world being represented. The mapmaker and the map would them selves have to be part of the map of the world, and can never, despite the illusion conveyed by the godlike per spective of the image, ever be truly separated from the image. Science becomes a set of cultural practices, with a subjective viewpoint, and a politics. It becomes a view from somewhere, rather than a mirror like representation.

In the late 1980s, attention gradually focused on the positioned nature of cartography as a discipline. The work of Brian Harley began to question how mapping operated as a powerful discourse. In a series of papers published between 1980 and his death in 1991, Harley began to challenge scientific orthodoxy and focus atten tion on the social role of mapping technologies. He proposed a new research agenda concerned with the roles maps play in different societies, arguing that maps often reinforce the *status quo* or the interests of the powerful, and that we should investigate the historical and social context in which mapping has been employed. In this view, cartography was not necessarily what cartographers said it was. Instead, Harley argued that we can only understand the history of cartography if we interrogate the forces at play around the mapping.

This social constructivist critique sometimes articu lated structural explanations for mapping, which sought understanding beneath the apparent surface of observ able evidence. For example, David Harvey's Marxist analysis of the role of mapping in time space compression examined the role of global images in the expansion of European colonial powers, and situated these as re flections of a changing mode of production. Drawing on linguistic structural thought, Denis Wood employed Barthean semiotics to persuasively argue that the power of maps lay in the interests they represented. Mapping in this view always has a political purpose, and this 'interest' often leads to people being 'off the map'. Wood argued mapping works through a shared cultural reading of a number of different codes in every map, which may be analyzed in a semiotic process to reveal the power behind the map. These interests all too often led to subjugation, oppression, control, and commerce. Through economic relations, legal evidence, governance, or social practice, the power of maps continues to be used to control. It has been argued that many of the social roles played by mapping stem from the modernist project, and that mapping partly constitutes modernist enterprise itself. By examining different categories across which power might be articulated, contextual studies can reveal how maps reflect but also constitute different kinds of political relation. Colonialism, ownership, national identity, race, military power, administration, and gender have all been theorized as playing key roles in mapping relations.

The Power of Contextual Approaches

Historically mapping has guided the explorer, fostering a sense of inevitable progress, and encouraging belief in the march of civilization. Local knowledge has been translated into tools to serve the needs of the colonizer, with new territories scripted as blank spaces, empty and available for the civilizing Western explorer to claim, name, subjugate, and colonize. Projection and design have been used to naturalize the political process of imperial control and sell imperial values to citizens at home. The continuing progress of colonial adventures is mapped out nowadays in our news broadcasts, and on the Internet, but the imperial rhetoric of control, governance, management of territory, and creation of new imperial landscapes remains the same. The colonial project relies on the map, and in turn the map relies on colonial aspirations.

This imperial context also shows how maps are part of the geographical imagination of every nation, contributing along with censuses, museums, and the printed word to what Benedict Anderson has termed a sense of imagined community that was enormously powerful in fostering the modern development of the nation state. Despite globalization, national mapping agencies still publish mapping designed to different specifications, in a cultural cartography that parallels national landscapes. National atlases serve as cultural ambassadors for their nations and ethnic stereotypes are perpetuated through the fixing of place names on officially published mapping.

The property relationship is also closely supported by mapping. Throughout history, the claims, rights, and desires of landowners are reflected in maps and in technological changes in mapping. The estate map served as a practical tool for the administration of a landed estate, as a means of social control, and as an esthetic statement for its owner. Digital land information systems allow land to be reapportioned but also reflect the power of the regime. Mapping itself can be a property with a value, often accumulated over many years and assiduously defended by its owners. Publication or dissemination is not always in the interests of the powerful: leaving out information from a map may also serve economic interests and reify social relations.

Military interests have exerted substantial influences on mapping throughout the history of cartography. Most of the significant technological innovations now accepted as routine in the mapping sciences, such as remote sensing, GIS, terrain analysis, and GPS based technologies, have trickled down from the military. The Web has also been strongly influenced by military research. Location based services stem from military investment and innovation. Multimedia mapping research is fired by military needs for battleground simulation. In high tech warfare, military personnel are more distanced from the act of killing if trained with simulators that depersonalize conflict, turning the reality in it into a computer game, where the real landscape morphs increasingly into the virtual.

It has also been argued that visual technologies are employed to facilitate social governance and that mapping plays an important role in this process by offering a technology that helps to 'discipline' model citizens by marking acceptable and unacceptable forms of behavior and spaces. National states are much more interested in tracking their citizens after the 9/11 attacks and a renewed emphasis upon mapping and GIS as surveillant technologies can be seen as part of a strategic and rational governmental response to risks, which parallels the invention of thematic mapping as a technology to reinforce a moral order in early nineteenth century Europe. People become cast as a problem, or as subjects demanding attention and control, in the interests of the state or corporation. Foucauldian interpretations of mapping as surveillant power knowledge are strongly compatible with a society in which closed circuit television (CCTV) increasingly captures and places representations of individual behavior and in which real time mapping places people and things. Mapping technologies have the potential to appear to reveal everything, but at the same time there may be moral, political, or social reasons for not revealing spatial information. There are still huge variations between different nations over what is released into the public domain and many nations still wish to keep secret the existence of strategic facilities. Rights to privacy may conflict with the all seeing power of spy satellite, or the real time Web based tracking mapping of children or criminals.

The authoritative scientific, 'all knowing' gaze of the map has been cast as masculine. In this view, mapping becomes analogous to a landscape or nature, created as an entity to be seen, controlled, and subject to male power, de corporalized and de personalized, doing male work. Mapping practices appear to reinforce this interpretation. In the history of cartography, most maps have been made by men. Where women were involved in the past, they were often in positions where they could exercise very little power over the medium. When women were depicted on maps, they were often exoticized or eroticized. The world represented on maps has also tended to be a world of stereotypical male interests: the built environment, emotionless and factual, revealed to be managed. And there are gender differences in the way maps are read. However, recent interpretations suggest these generalizations may oversimplify – a feminist GIS might empower very different gazes, new themes can be mapped, in new ways. A more hybrid vision becomes possible in which power may be subverted.

Critical Cartography

At first, these challenges to science operated as an empirical and social critique of mapping practice, rather than as a theoretical challenge to mapping as scientific or technical knowledge *per se*. A more critical alternative began to emerge from this process, however, in which the focus shifted toward different philosophical ways of understanding powerful mappings. For example, Harley later moved on to employ Derridian deconstruction and Foucauldian notions of power knowledge in the discursive construction of the mapping process. In a Foucauldian sense, mapping offers a very appropriate metaphor to explain the heterotopic associations of contemporary spaces. Power becomes something that may be used to control, but which may also be employed to resist and create alternatives. Post structuralist approaches show us that governance and oppression include the seeds of resistant possibilities. And in the last decade there has been a widespread recognition of the power of post structuralist critical approaches. Now there are also postcolonial mappings encompassing new ways of understanding a globalized world of flows and linkages, orientalist mappings informed by Edward Said's work, as well as feminist mapping theory. A critical cartography has begun to emerge.

This critical project does not only encompass writing about mapping. New mapping has been published to disseminate alternative views as part of a 'counter mapping' process. For the last 15 years there has been a profusion of these 'alternative' maps: reaffirming the rights of indigenous peoples; arguing local cases in resource struggles; encouraging community involvement in sustainable lifestyles; reasserting the role of the past in contemporary contexts; or celebrating the esthetic and the local in an age apparently dominated by uniform and mechanized production and global style. The World Wide Web has offered a new collaborative medium and forum through which alternatives may be disseminated and which cuts across the limiting forces of traditional institutions. These projects challenge orthodoxies by often suggesting that theory and practice might be combined in a form of praxis, which seeks to change the world while rewriting accepted ways of theorizing about the world. And in turn, practice has become increasingly important in thinking about mapping.

Relational Thinking: Toward Practice

The artistic and critical turns toward mapping described above are taking place in a time of rapid technological and cultural change, reflecting territorial uncertainty in the aftermath of the collapse of the Soviet Union; the global war on terror; as well as challenges to American cultural hegemony and neoliberal orthodoxies. Flux in the information economy; the profusion and dynamism of products and media; increasingly privatized multi channel and multi mediated cultures, and the rise of virtual places and spaces have all encouraged a focus on, and questioning of, the apparently fixed certainties of the map. The blurring of contemporary identities and of divides between genders, ages, social roles, and between nature and culture are all central for new mapping philosophies. These shifting ways of thinking about mapping also reflect a more general rise in relativist approaches to knowledge, and in particular an emphasis upon performance, relations, and affect, instead of essence.

Hyper-Reality and Relational Thinking

Mapping allows an exploration of the ambivalence of contemporary sign systems, in which reality and the map are increasingly blurred and where the real and the hyper real coexist. The TV news graphic of conflict in Afghanistan incorporates mapping, pictures, animated symbols, satellite imagery, text, and is interpreted in a voice over delivered by the news anchor, where the events that it narrates are inseparable from the 'infotainment' on the screen. The play of images becomes a central concern in this Baudrillardian economy of signs, where imitation replaces representation and where new media remediates older geographical imaginations. The inter textual field of the map may be read in many different ways. Mapping may, in some contexts, be seen as constantly echoing and remaking past tropes. In others, such as the worlds of gaming, it is at the forefront of creating new imaginaries, in which the active subject is immersed as an actor in the virtual landscape or map. The map becomes the territory and carries us with it on a journey of change. We can zoom in onto the landscape and almost become part of the action.

Maps are also mobile and networked, an outcome of a technology that allows cultural messages to be transformed and shared. The development of Actor Network Theory in science studies has done much to encourage a consideration of the transformative potential of technologies. In his early work, Bruno Latour reflected on the immutable and mobile nature of mapping, the ways in which the medium has been able to carry fixed messages into new contexts, and in so doing support the expansion of modernism. David Turnbull stressed the indexical qualities of mapping, the ways inscriptions brought the world home to different people and he also charted encounters between different assemblages of mapping knowledge. In this approach a thick description of the context is needed to unpack the relations and flows between different actors and actants. Researchers have to trace the inscriptions left behind by different actors, but also need to carry out ethnographic work in the field. More recent actor network approaches stress the ambiguities of mapping work, and the complex negotiations in different knowledge communities, through which mapping works.

The Affective Performance Of Mapping

Mapping also reminds us about the pervasive and problematic relations between subjects and objects: mapping can be embodied and its performance reveals the (e)motive and constitutive nature of action. Employing Judith Butler's notions of performance, Del Casino and Hanna use the construct to argue against binary thinking, and in favor of spaces and mapping as being co constituted. A hybrid mapping emerges from this process, at once produced and consumed, cultured but also natured.

Other theorists, such as Kitchin and Dodge, argue that mapping calls spaces into being and is, itself, always called into being to do particular sets of work. Mapping in this view is always emergent, and as practice is concerned with processes not representations.

For some this relational turn has been theorized as nonrepresentational, drawing on Deleuzian ideas. Mapping can then be imagined as a suite of cultural practices involving action and affects. This kind of approach reflects a philosophical shift toward performance and mobility and away from essence. This rethinking of cartography is supported by historical and contemporary work. Researchers concerned with historical contexts increasingly stress the interplay between place, times, actions, and theory. Mapping in different cultures reflects three traditions: an internal or cognitive set of behaviors involving thinking about space; a material culture in which mapping is recorded as an artifact or object; and a performative tradition where space may be enacted through gesture, ritual, song, speech, dance, or poetry. In any cultural context there will be a different blend of these elements. So interpreting mapping means considering the context in which mapping takes place, and the ways in which it is performed. Instead of focusing on artifacts, or esthetics, or human agency, or the politics of mapping, a nonrepresentational performative approach questions who employs mapping, to what purposes, and focuses upon mapping actions. In early modernity, the culture of printing allowed mass production and dissemination of mapping, and encouraged the map to play its often oppressive role, but detailed examination is needed to chart the actions that this culture brought about. In contexts where encounters occur between different mapping traditions, for example, the inter textual field of mapping might best be interpreted through a postcolonial approach: Matt Sparke's reading of the interplay between North American Indian mapping, colonial alternatives, and nation building in Canada is informed be a consideration of Homi Bhaba's work. Performance is seldom simple; actions are complex and negotiated and mapping interrelates with other cultural and geographical texts.

Contemporary research is also starting to employ ethnographic approaches to carry out empirical investigations to investigate how everyday mapping is practiced and its relations to identity, the construction of different spaces, and the task oriented behavior it supports. Researchers have started to critique representational approaches to mapping and way finding. Instead of positing the existence of a mental map, way finding may be possible because of cultural practice and might be best understood as a narrative reenactment of past mobility. Places are called into being as we move through them, spaces co constructed through mapping practices. In this kind of approach, mapping becomes a set of social actions.

Praxis

In this performative and relational cartography, technologies need to be observed in specific local contexts, and seen in action. Resistant mappings may be understood in a similar fashion to more oppressive mappings that have received the bulk of the attention of historians of cartography, and the worlds of mapping knowledge and mapping practice increasingly blur. Theory and practice merge in a form of critical praxis. For example, the Web brings together mappers as users in collaborative projects grounded in space, in mapping hacks and 'mash ups', which subvert existing power structures, and in collaborative urban exploration. It also encourages a sharing of different ways of thinking about the map. New technologies create new stories and performances and evoke many different emotions. Instead of the rational behavior assumed by the scientific model, or the esthetic experience evoked by artists, or the oppressive behavior implied in Harleyian takes on the medium, a performa tive cartography imagines many different possibilities. Mapping may be angry, boring, celebratory, exciting, happy, painful, persuasive, playful, proud, reassuring, revealing, restless, sad, or wistful. As human practice we can make it what we want. And the limits to what may be known about mapping are similarly boundless.

Summary

This article shows how philosophy matters for mapping and demonstrates how changing approaches to the me dium have been mapped out in practices, artifacts, and in writing. The apparent neutrality of the mapped image continues to exert a strange power through the esthetics of design, but despite the hegemonic technological power of scientific and technological progress, mapping is still imagined, made, and remade in shifting and ambiguous ways that can evoke powerful emotions and are enacted in every part of the world.

Philosophy and mapping morph into one another in ways that are beyond a brief encyclopaedic summary. Their relations are complex. The neat subject headings in the argument above are, in practice, fluid and often ambiguous: critical approaches have historically often also engaged with the human subject; scientific ways of knowing the world have always had to come to terms with esthetics and with imaginative power; post struc turalist thought emerged as a response to structuralist critiques of science; recent performative approaches to mapping depend on relations between human and nonhuman actors and technologies, etc. Different philosophies coexist and cross fertilize one another. The world of ideas constantly interplays with the material world and with the world of practice and new maps emerge from this contested process. Human beings will continue to try to make sense of this complexity by advancing new ways of thinking about the world, and creating new mapping.

See also: Cartography, History of; Maps; Poststructuralism/Poststructuralist Geographies; Representation-Mapping; Semiotics.

Further Reading

Blaut, J., Stea, D., Spencer, C. and Blades, M. (2003). Mapping as a cultural and cognitive universal. *Annals of the Association of American Geographers* 93(1), 165 185.

Cosgrove, D. (ed.) (1998). *Mappings*. London: Reaktion Books.

Crampton, J. (2001). Maps as social constructions: Power, communication and visualization. *Progress in Human Geography* 25, 235 252.

Del Casino, V. and Hanna, S. P. (2005). Beyond the binaries: A methodological intervention for interrogating maps as representational practices. *ACME: An International E Journal for Critical Geographies* 4(1), 34 56.

Habermas, J. (1978). *Knowledge and Human Interest*. Cambridge: Polity Press.

Harley, J. B. (2001). *The New Nature of Maps*. Baltimore, MD: Johns Hopkins University Press.

Harley, J. B. and Woodward, D. (1987). *The History of Cartography.* Chicago, IL: University of Chicago Press.

Law, J. (2004). *After Method: Mess in Social Science Research*. London: Routledge.

MacEachren, A. M. (1995). *How Maps Work*. New York: Guildford Press.

Perkins, C. (2006). Mapping. In Douglas, I., Huggett, R. & Perkins, C. (eds.) *Companion Encyclopaedia of Geography* (2nd edn.) pp 555 571. London: Routledge.

Pickles, J. (2004). *A History of Spaces: Mapping Cartographic Reason and the Over Coded World*. London: Routledge.

Robinson, A., Morrison, J. L., Muehrcke, P. C. A., Kimerling, J. and Guptill, S. C. (1995). *Elements of Cartography* (6th edn.). London: Wiley.

Rose, G. (2007). *Visual Methodologies* (2nd edn.). London: Sage.

Turnbull, D. (1993). *Maps are Territories*. Chicago, IL: University of Chicago Press.

Wood, D. (1992). *The Power of Maps*. London: Routledge.

Mapping, Race and Ethnicity

H. Winlow, Bath Spa University, Bath, UK

Glossary

Anthropogeographer An individual who works at the intersection of the disciplines of anthropology and geography, using physical racial categorizations to explore the relationships between race and space.

Anthropometric Mapping Mapping of individual racial traits, such as cephalic index, pigmentation, or stature.

Ethnic Mapping Mapping of aspects of ethnicity, including religion, language, nationality, and culture.

Otherness Term used to describe the Western tradition of categorizing the Western self against the exotic other through binary categorization, such as civilized/savage, advanced/primitive.

Racial Mapping Mapping of supposedly discrete racial groups, which may be based on a combination of physical traits.

Sign System A set of signs or symbols used within map texts and contributing to the internal power of maps. Elements of sign systems include data division, projection, orientation, shading, color, symbols for specific features, use of arrows and lines, legends, and marginalia.

Social Construction Idea that categories such as race, ethnicity, and gender are socially produced and are not fixed.

Introduction

The cartographic representation of race and ethnicity encompasses a range of categorization systems and mapping techniques which have been used to convey notions of difference. This discussion focuses on selected maps from the mid nineteenth century onwards, a period which coincided with a rise in thematic cartography, an emphasis on scientific representation, and the development of increasingly sophisticated cartographic forms. Critiques of colonial mapping have demonstrated the ways in which cartographies have reinforced notions of overseas others and how Western processes of mapmaking have created cartographic silences in relation to native peoples. Alongside this erasure of ethnic groups from the landscape, there has been a separate history of racial cartography that has involved the deliberate portrayal of difference within the map text, a form of mapping that has also served to reinforce notions of 'otherness'.

The establishment of evolutionary theories in the nineteenth century led to an obsession with delineating human racial types, which were often portrayed in cartographic form. Individual anthropologists and geographers developed their own sign systems to produce racial and anthropometric maps at global, continental, and regional scales. The increasing use of statistical and cartographic techniques for state control of the social body in the late nineteenth and early twentieth centuries led to the use of ethnic cartography for delineating political boundaries, for example following World War I, a trend which remains evident in recent political cartographies. Individual cartographers have been involved in the social construction of race, which has been reinforced through cartographic power.

It is informative to provide a brief historical overview of the cartographic portrayal of race preceding the nineteenth century, before analyzing more recent mappings of race and ethnicity.

Mapping Difference

Notions of racial difference can be traced to travel accounts of the Greek and Roman periods. Descriptions of fantastic and monstrous races date back at least to Herodotus in the fifth century BC, and are also found in the work of Megasthenes, Ctesias of Cnidos, and Alexander the Great, in the fourth century BC, all of whom traveled to India. Pliny the Elder, whose work influenced Medieval thought, produced the 36 volume *Natural History* (first century AD) which consolidated many earlier works of geographical lore and included detailed descriptions of the monstrous races purported to exist at this time. Types described here included the blemmyae (with faces on their chests, lacking head and neck), the sciopods (shadow foot, with one leg and a single great foot), and the cynocephali (dog head). Pliny's bizarre descriptions of human varieties were transmitted to the Medieval period, partly because of their exotic appeal and partly because the existence of some groups could be explained anthropologically – the idea of the blemmyae appears to have resulted from the observation of a tribe of this name, who carried large ornamental shields, giving the impression of headless men.

Myths of monstrous races were reinforced in Medieval *mappaemundi*. Both the tripartite T O maps (which included the three known continents and were of eastern orientation) and the zonal world maps (based on a spherical world and divided into climatic zones) contained depictions of monstrous races. In the thirteenth century

Psalter map 14 monstrous races appear below the Nile, while in the Ebstorf map (*c.* 1240) 24 monstrous races appear, and in the Hereford map (*c.* 1290) 20 races appear, each located on the periphery of the known world. In Macrobian zonal maps, dating from the late Greco Roman period, the sphericity of the Earth was central to the concept of the antipodes, enabling cartographers to place the monstrous races opposite to and below Europe. An example of a twelfth century Macrobian world map (**Figure 1**) confined the sciopod to the antipodal region. Through cartographic representation, Australia was portrayed as a continent of perversity and otherness, in advance of its discovery by Western Europeans.

As new geographical discoveries were made by Europeans during the Renaissance, the myth of the monstrous races gradually eroded, but a fascination with racial difference remained. Alongside the mapping of new territories, Renaissance maps typically contained representations of savage races in the marginalia – a tradition that continued into the nineteenth century. By the mid nineteenth century the whole globe, excluding the polar regions, had been explored, ordered, and cartographically reproduced. Myths of fantastic peoples had been dispelled but ideas of moral and physical difference remained, and ethnocentric assumptions about moral and physical differences were applied to contemporary populations. Following the acceptance of evolutionary theories, concepts of race hardened and the Victorian fascination with classification and measurement was applied to humanity. Some groups, such as Australian aboriginals, became regarded as the missing link of evolutionary history. Concepts of racial hierarchy were supported by an emphasis on linear progressionism and were reinforced through Victorian visual culture, including art, photography, and cartography.

Racial categorization started to become a feature of supposedly scientific maps in the early nineteenth century, with the introduction of thematic cartography. By the 1900s a rich set of techniques existed for mapping human population distributions, including population density, migration, longevity, language, and religion. In 1852 Heinrich Berghaus published his *Physikalischer Atlas*,

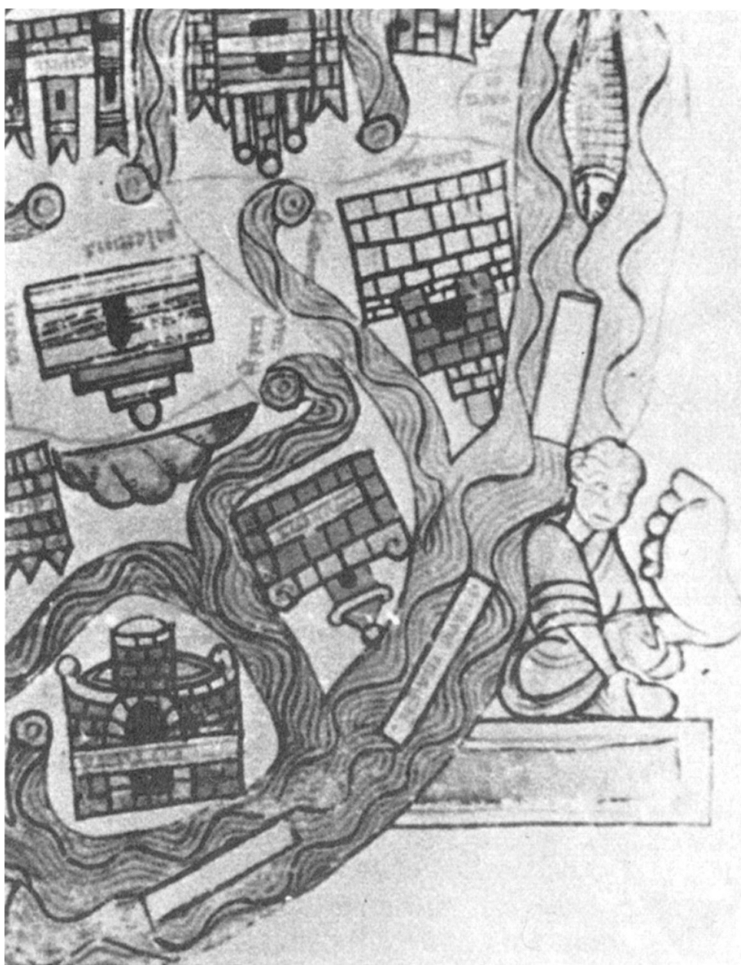

Figure 1 Sciopod on a world map, Beatus of Liebana. From Friedman, J. B. (2000). *The Monstrous Races in Medieval Art and Thought*. 2nd edn. Syracuse, NY: Syracuse University Press.

an early example of thematic cartography, which included distributional maps of physical features, meteorology, and human populations. Nineteen ethnographic plates portrayed ethnic and linguistic geographies – for example, a map of North America showed the distribution of 34 Indian tribes, using color codes. **Figure 2** shows Berghaus' map of the human races and is an example of how racial others were portrayed in the marginalia of the map text, images which are culturally embedded in the discourse of the map.

Definitions

This discussion centers on maps produced from the late nineteenth century onwards, an era coinciding with the formation and development of modern academic disciplines, including geography, anthropology, and cartography. Within this context, it is useful to define the terminology related to mapping race and ethnicity.

Racial Mapping

The concept of racial mapping is applied to those representations where the author has attempted to prove the existence of several discrete races, each combining a number of physical traits, such as stature or pigmentation. Examples of individuals who used this approach include American anthropologist W. Z. Ripley who argued for a tripartite division of Europe, and Australian geographer Griffith Taylor who developed a zones and strata approach to understand race.

Anthropometric Mapping

The label anthropometric applies to the mapping of individual traits (without necessarily being assigned to a

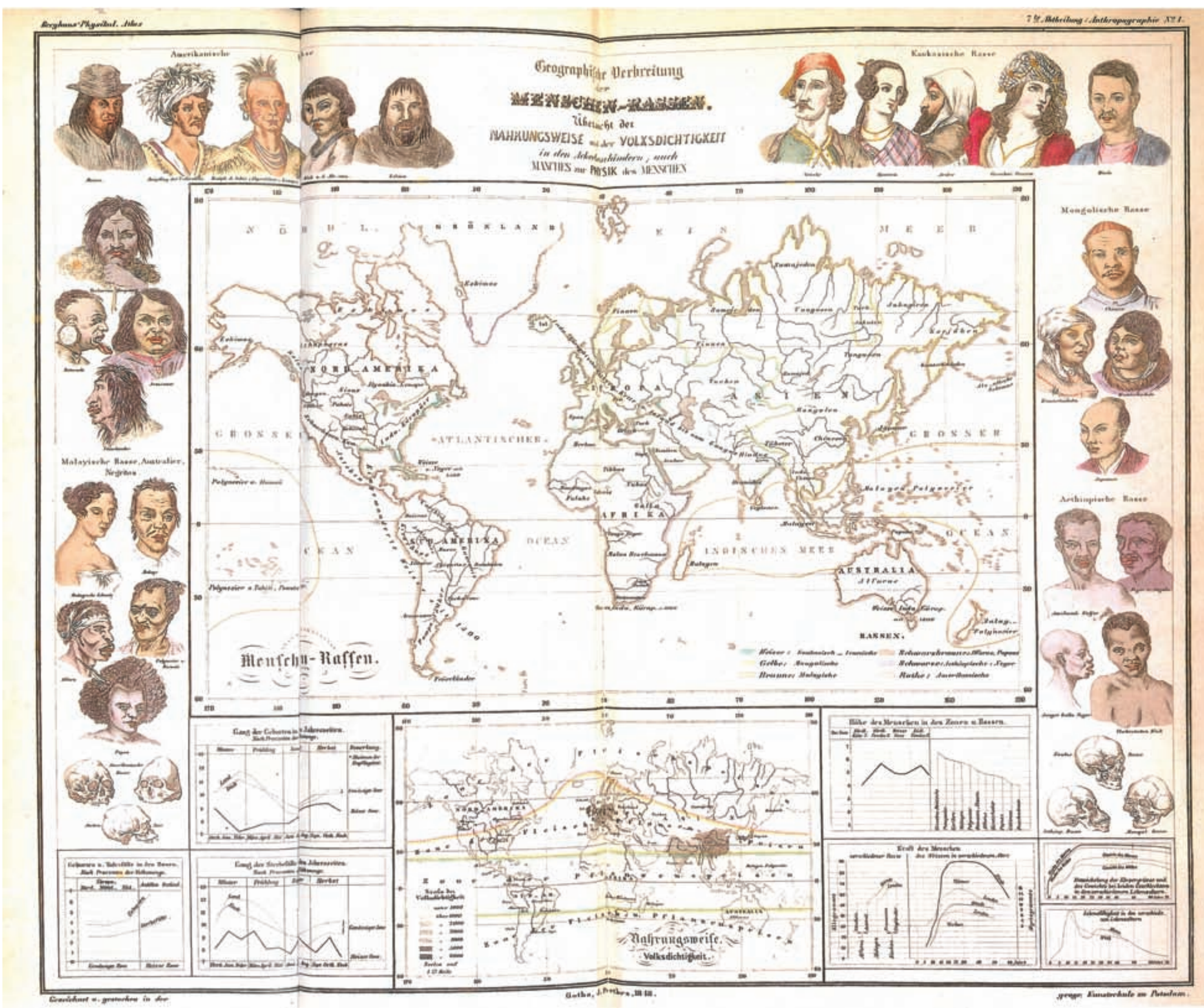

Figure 2 Map of ethnic divisions of the human race, Berghaus 1852. From Whitfield, P. (1994). *The Image of the World: 20 centuries of World Maps*. London: British Library. © British Library Board. All rights reserved (Maps 49.e.32).

racial group). This involves the production of maps that focused on single racial characteristics, including ceph alic index, stature, and pigmentation. The mapping of single traits was again evidenced in the work of Ripley, and in Britain in the work of physical anthropologists, including John Beddoe, John Gray, and James Tocher, and geographers, such as Herbert John Fleure.

Ethnic Mapping

Ethnic mapping involves the mapping of aspects of eth nicity (which may overlap with biological differences). This type of mapping includes a focus on cultural differences, including religious background, perceived nationality, and language. A trend of ethnographic mapping developed in nineteenth century atlases (as illustrated by Berghaus) while later forms of ethnic cartography, based on statistical data, are represented in the US mapping of boundaries for the 1919 Paris Peace Conference. Data on ethnic group, nationality, language, and religion are collected in many modern censuses in relation to small geographic units, allowing these distributions to be mapped.

Social Construction of Race

Kobayashi has argued that the idea of race as we understand it can be traced to Enlightenment develop ments in scientific thinking. Immanuel Kant's stress on the links between skin color and distance from the equator fed into new ways of thinking about race, and his assumptions about links between darker skin color and inferior intellectual qualities reflected a hierarchy of people and places. Similarly, Montesquieu's 1748 *The Spirit of the Laws* emphasized the influence of environ mental factors, such as climate and soil, on cultural characteristics and reintroduced the idea of the molding power of the environment which dated to the writings of Hippocrates and Aristotle. The Enlightenment concept of The Great Chain of Being reinforced ideas of racial hierarchy and progress. Victorian racial classificatory systems also had their roots in the Enlightenment period and stemmed in part from the acceptance of the Linnaean classificatory system. Carolus Linnaeus' *System in Nature*, first published in 1735, designated animals into mammals and reptiles, identified subgroups of mammals, and Linnaeus was the first to suggest an ape ancestry for the human species. In his second edition, of 1740, Lin naeus delineated 4 categories below the species level, defined as white Europeans, red Americans, yellow Asians, and black Africans (which remained throughout later editions). A fifth subspecies, labeled *Homo sapiens monstrosus*, included groups which he was unable to fit into his other classifications, including a selection of imaginary peoples.

In the nineteenth century the idea of discrete race types, linked to geographical region, became more strictly delineated, assuming an additional sense that initially appeared more scientific. The theory of natural selection, with its emphasis on environmental adaptation, was widely accepted but the branching tree metaphor in Darwin's work was ignored in favor of linear progress and Victorian scientists assumed an innate link between biological, socio cultural, and intellectual development. In the early twentieth century, a few anthropologists, such as Franz Boas, questioned these assumptions, noting that cultural change seemed to develop independent of biology. However, the wider scientific paradigm con tinued to focus on attempts to classify both biological and cultural development. It was only after World War II and the Holocaust (which incorporated eugenic ideas advo cated by racial scientists) that biologists and anthro pologists fully rejected racial classificatory systems based on physical traits. Kobayashi has noted that, despite the defeat of Nazism and rejection of racism following the war, there was a systematic denial of racial issues within geographical circles. Racial issues began to be addressed, mainly in relation to spatial inequalities in the late 1960s and 1970s. Following the critical turn of the 1980s and 1990s, human geographers have come to understand the idea of race as socially constituted over time – races are social constructions and not discrete biological cat egories. The acceptance of this idea among social sci entists has been reinforced by recent work in the field of genetics, which has been unable to find any significant and consistent divisions between biological groups on the basis of race or ethnicity.

Racial Measurement

Racial hierarchies were constructed and legitimated through Victorian racial science: physical anthropometry allowed race to be measured and classified. A number of methods were devised for calculating racial type, in cluding phrenology, facial angle, and cephalic index (breadth of skull expressed as a percentage of length). The idea that races could be ranked on a continuous scale of civilization was applied to biological difference and extrapolated to sociocultural and intellectual devel opment. The socially (and arbitrarily) constructed boundaries, between different racial groups and traits, were reinforced through cartography. Many ways of in terpreting, dividing, and mapping the same data were possible, since physical changes occur on a continuous scale. Stocking has observed that the terms dolicho cephalic (long headed), mesocephalic, and brachy cephalic (broad headed) to denote skull shape were arbitrary categories established by anthropologists. Equally, other anthropometric characteristics occurred on a continuous scale, with the result that different classificatory systems were produced from similar data.

Maps and Power

Harley's work on the deconstruction of the map, which incorporates Foucault's emphasis on power/knowledge, provides some means of understanding how the power of the map was harnessed to reinforce racial taxonomies. Maps are often regarded as neutral, scientific, and accurate reflections of the world. The fact that maps became an acceptable part of academic discourse meant that their legitimacy went unquestioned for many years. It is now widely acknowledged that maps are cultural texts re flecting wider social, cultural, and political contexts.

Methodologically, Harley's concepts of internal and external power can be utilized to critique a wide range of maps. Internal power is contained within the map text, the key to which is "cartographic progress," which includes the use of sign systems, which highlight and suppress elements through the use of shading, lines, and projection, resulting in the hierarchicalization and standardization of the landscape (Harley 1989, 13). Wood's explanations of the emergence of cartographic sign systems and of the links between signifier and signified provide a theoretical and practical framework for understanding maps as nar ratives, as well as understanding how maps become nat uralized. The signifier (element in the map text) becomes confused with the signified (element being represented). However, it is not only the internal linguistic elements of the map that need to be open to critical analysis. Behind most maps is a patron; historically, the monarchy, the government, and the church have all initiated mapmaking schemes. External "power is exerted on cartography" (Harley 1989, 12), is often centralized and bureaucratic, and is imposed from above.

Critique of Racial, Anthropometric, and Ethnic Cartography

The mapping of race and ethnicity by geographers has involved the division of statistical data into categories and the mapping of these categories using a range of cartographic devices, including projection, orientation, shading, coloring, lines and arrows, and labeling. The cartographic representation of race has taken place across all geographical scales and cannot be divorced from wider political contexts, or from the situatedness of individual geographers. This is demonstrated below through a focus on two mapping trends: anthropometric and racial maps produced by anthropogeographers in the late nineteenth and early twentieth centuries; and the use of ethnicity data in twentieth century boundary de lineation. In the first area, Griffith Taylor's world scale zones and strata theory; W. Z. Ripley's tripartite division of Europe; and regional scale anthropometric mapping in Scotland. In the second area I will examine the mapping of the Balkan states following World War I and consider the recent post Yugoslav mapping of the area.

Taylor's Zones and Strata Theory of World Evolution

Griffith Taylor's (1880–1963) racial geography stemmed from a combination of geological theory, nineteenth cen tury racial science, and environmental determinism. *En vironment and Race*, published in 1927, represented many of his principal geographical conclusions and examined the origins and distribution of racial groups. Here, Taylor argued that the ice ages and interglacial periods had triggered human evolution, and that human groups had been pushed out down three continental migration cor ridors (Eur Africa, Australasia, and America) – ideas posited in W. D. Matthew's 1915 *Climate and Evolution*. Basing his categories upon cephalic index, Taylor argued for six divisions: negrito, negro, Australoid, Mediterranean, early Alpine, and late Alpine. He argued that the primitive long headed negrito group had been pushed from a center of evolution in Asia after the first ice age, and that the Mediterranean and the broad headed Alpine groups had evolved after the fourth glaciation – remaining closest to the center of dispersal, where temperate climates had promoted racial advancement.

Taylor applied the geological metaphor of zones and strata to the migration and development of human groups. The layers of strata were defined as artifacts from previous cultures and it was argued that each continental peninsula would contain zones of increasingly primitive peoples. The order of evolution was assumed to be the same, whether moving out from the center through the zones, or moving down through the strata. In the zones where racial evolution was considered most advanced, the buried strata of more primitive tribes would be more numerous. Taylor's cartographic elaboration of this the ory is illustrated in **Figure 3**. A triangular feature in the bottom left corner of the map indicates Taylor's racial taxonomy, shown as strata ranging from negrito to late Alpine. The use of the north polar projection emphasized the migration of the races outward from a central core, while arrows indicate the three migratory channels (de vices previously utilized by Matthew).

Throughout his career, Taylor manipulated carto graphic images and developed his own methods of graphicacy. A distinctive feature was his use of isopleths, a meteorological technique which he used to depict racial and linguistic dispersion. In his migration zone map (**Figure 3**), isopleths are used to distinguish between different strata emphasizing radial dispersion. The lines

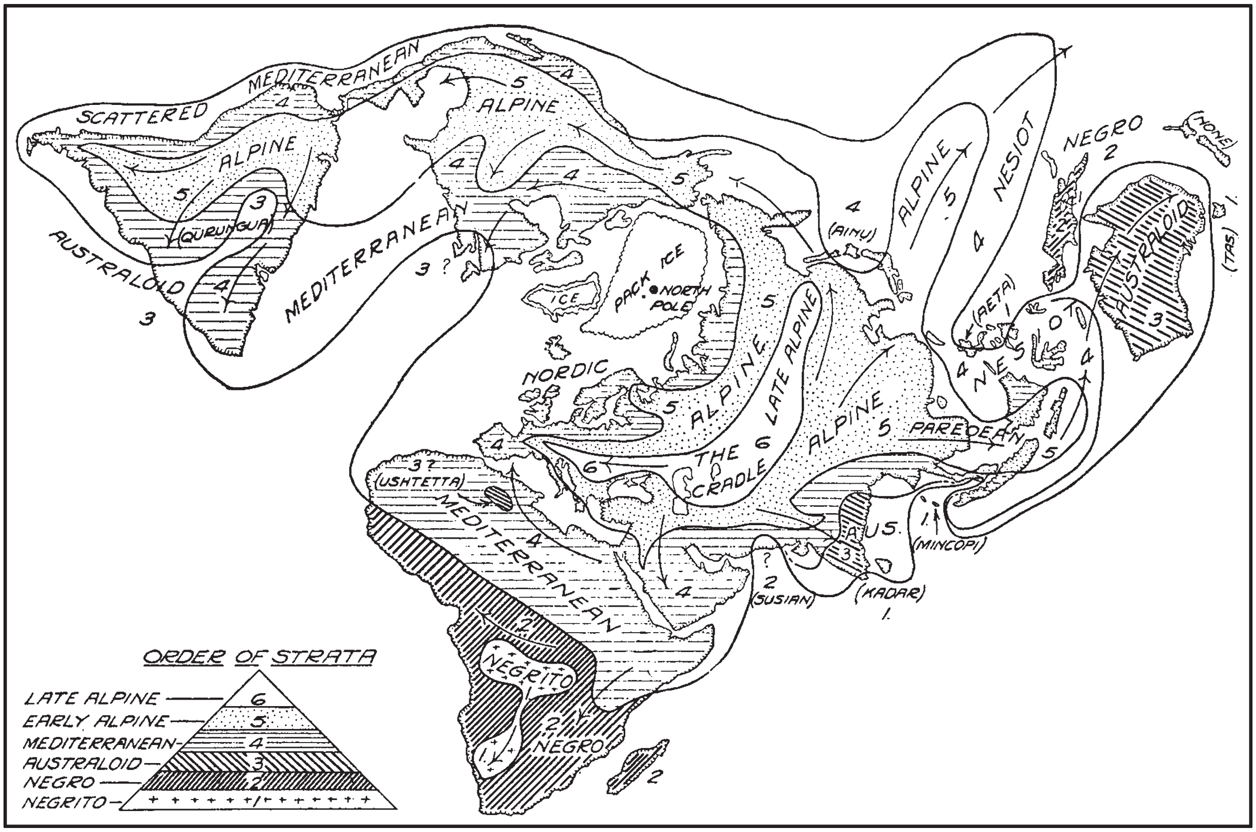

Figure 3 Migration zone classification of the races of man. From Taylor, T. G. (1919). Climatic cycles and evolution. *Geographical Review* 8(6), 289 328.

on this image are isocephalic lines, indicating areas of equal cephalic index.

Taylor's work borrowed from the field of geological mapping, as part of a wider system of scientific metaphor. Besides his zones and strata model, Taylor mapped racial and linguistic distribution using the concept of shatter belts, the lava flow analogy, and geological sections. He used block diagrams to make comparisons between the evolution and distribution of transport, mammals, and human races. Block diagrams of linguistic evolution reinforced Taylor's assumptions about primitive language groups, and he applied the geological construct of the inlier to the Basque, the Galcha, the Welsh, and the Breton groups, again implying the existence of a physical and mental race hierarchy.

Taylor's racial cartography was sustained by the view that cartography was a neutral science, a tool which could be used to display racial truths based on anthropometric data. The incorporation of geological metaphor reinforced these ideas.

Mapping the Races of Europe

The racial theories of W. Z. Ripley (1867–1941) incorporated anthropological and geographical elements, evolutionary theories, and a reliance on statistics. In *The Races of Europe* Ripley promoted a new tripartite taxonomy for Europe, which challenged the traditional idea of one European race and suggestions of multiple European races. Statistics – which had been collected by European anthropologists throughout the nineteenth century – were used to draw conclusions about biological race types and to link social and moral issues to these traits. Central to Ripley's investigations was the notion of pure race, or type, and Ripley argued that the most reliable methods for delineating these types were cephalic index, pigmentation, and stature. His three European races included the Teutonic, found in northeastern Europe, characterized by tall stature, fair pigmentation, and a long head; the Alpine, found in central Europe, characterized by medium stature and pigmentation and a broad head; and the Mediterranean, found in southern Europe, characterized by dark pigmentation and a long head. It was assumed that every individual belonged to one of these categories or resulted from an intermixture of them.

Ripley's cartography legitimated his racial categorizations and emphasized the idea that particular racial groups inhabited particular spaces. In producing his maps Ripley consciously adopted a method that, he argued, rendered all data comparable (although his statistics had been gathered from a wide range of sources, using a variety of collection methods). Views expressed in an 1899 article on mapmaking, and the collation of the maps in *The Races of Europe*, illustrate a desire to represent the racial truth, which involved his objectification of human individuals.

Figure 4 shows Ripley's map of brunette types for Germany. The portrayal of increasing pigmentation from north to south (representing Teutons in the north and moving to an intermediate type in the south) is particularly dominant in this map, and could serve to reinforce assumptions of a biological racial hierarchy, where advanced races are associated with temperate northern latitudes and reflecting traditional assumptions, including Kant's theories that pigmentation is related to distance from the equator. The internal power of the map text lies in the imposition of geometrical divisions on the population and the landscape, as well as the use of shading that reinforces color divisions. Ripley's maps were supported by external power manifested in the cultural acceptance that races could be identified scien tifically, and that they could be hierarchically classified.

In *The Races of Europe*, Ripley's maps effectively demonstrated the existence of a tripartite scheme in Europe. His later work drew on these assumptions and involved a metaphorical mapping of social and moral geographies onto the US landscape. In the late nine teenth century US increased immigration, from eastern and southern Europe, was regarded as a problem by the elite. There were fears that increasing rates of divorce and suicide and low birth rates among Anglo Saxons could lead to the race suicide of this group. Hints of these fears were apparent in *The Races of Europe* where Ripley discussed the persistence of the darker (Mediterranean) types in city environments and mapped the occurrence of suicide and divorce, in relation to cephalic index. While initially refraining from stressing a hierarchical element

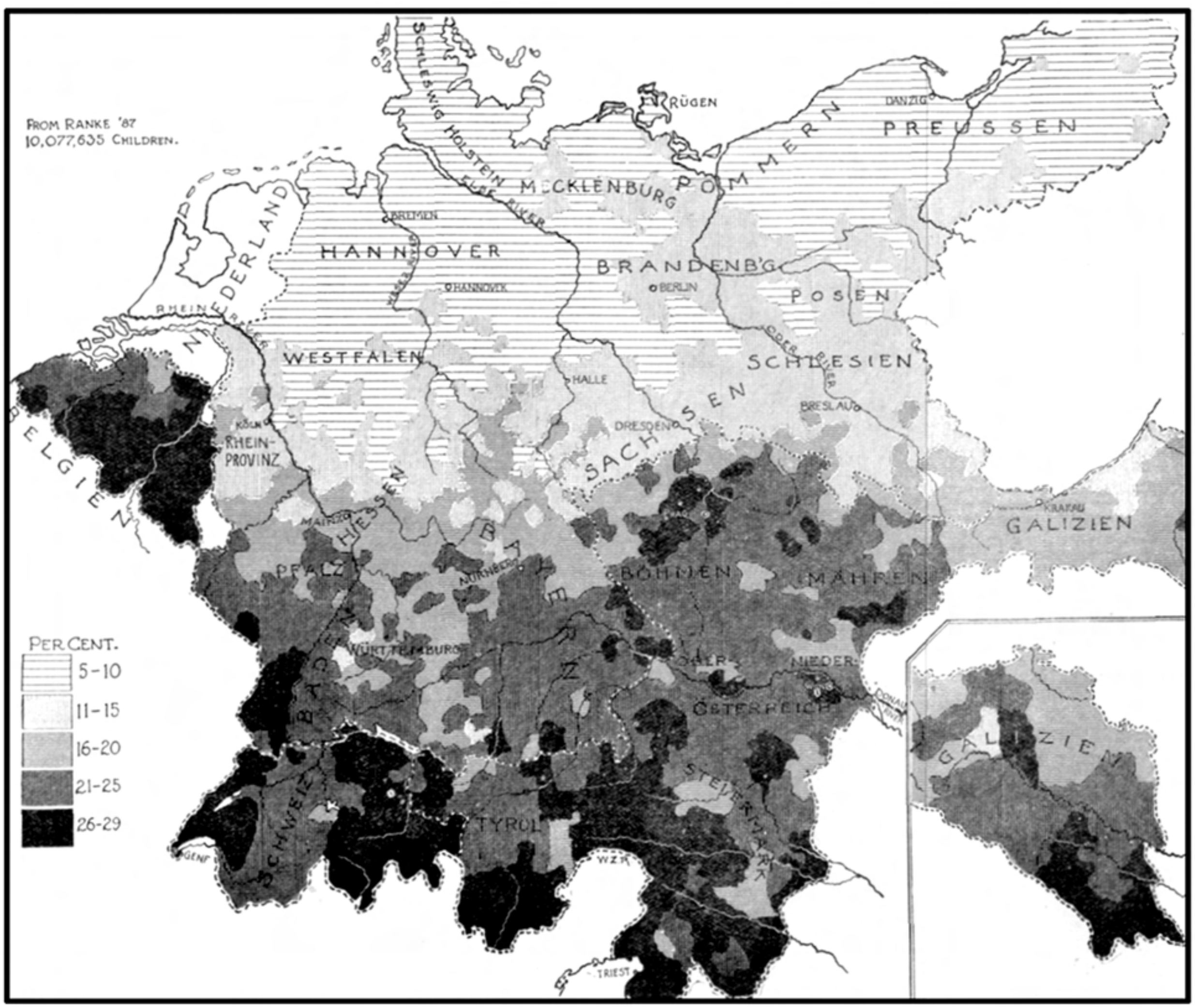

Figure 4 Relative frequency of brunette types for Germany. From Ripley, W. Z. (1899). *The Races of Europe: A Sociological Study*. London: Kegan Paul, Trench, Trubner.

in his scheme, this became implicit in Ripley's later work, allowing his divisions to be used in support of immi gration restriction. A 1913 map produced by Ripley for the *New York Times* highlighted the immigrant problem. This image, based on 1910 census data, showed the percentage of foreign born whites in the total population and accompanied an article titled 'Census figures disclose grave racial problems'. Here, darker shading denoted the greatest dominance of the Mediterranean immigrants, and was shaded according to state (and not city) boundaries, visually emphasizing the concentration of immigrant groups.

Mapping Regional Identities

The regional mapping of racial traits is evidenced by the anthropometric mapping of the Scottish and Welsh populations in the UK. In the nineteenth century, the industrious characteristics of the Teutons were con trasted with the image of the Celts as backwards. Sep arately, the development of Gaelic pride romanticized Celtic life and culture, which was seen as isolated from modernity. These developments resulted in a confused notion of Scottish ethnic identity, where lowlanders were associated with English hegemony and highlanders were constructed as culturally and racially other.

A series of articles mapping anthropometric traits in Scotland were published in the *Journal of the Royal An thropological Institute* (*of Great Britain and Ireland*) in the early twentieth century. This included work by Scottish an thropologists James Tocher and James Gray and British anthropologist John Beddoe. Far from denigrating the Celtic roots of the population Tocher regarded Scotland as having many areas of good stock, particularly in rural areas. Tocher and Gray were motivated to collect stat istical population data across Scotland. Similar motivations led H. J. Fleure, who felt that Welsh rural and cultural traditions were central to Welsh identity, to map racial traits in Wales. In a 1900 paper, Tocher and Gray mapped the pigmentation characteristics (of hair and eyes) of the East Aberdeenshire population based on statistics gathered at the 1895 Mintlaw gathering and at schools across the region. Pigmentation characteristics were mapped by plotting percentages of the type of pigmentation for each school. Numbers were then treated as if they were heights on a relief map (in a way similar to Taylor's construction of isocephalic lines) and then shaded to impose visual clarity on the data. Several areas on the maps displayed highly regular boundary lines reflecting the use of math ematical formula employed rather than any natural div isions between racial types.

In a 1907 article, Gray produced 21 anthropometric maps based on pigmentation data from approximately 500 000 school children across Scotland. He observed that the intervals between the contour lines were shaded in order to make clear the various densities of pigmen tation. Again, a geometrically imposed classification was clear in the maps produced. Here, Gray also introduced his idea of divergence maps which were produced for boys' hair and eye pigmentation (**Figure 5**). These were constructed using logarithms based on pigmentation frequencies on an area basis and the final values were again used to define the contour intervals. The maps were intended to reveal how the population was distributed within the country and to identify which areas were made up of native Scots (darker pigmentation) and which of Norse invaders (lighter pigmentation). Gray's statistics later formed the basis of John Beddoe's 1908 map based on the compound index of nigrescence, where a high nigrescence index (a quasi algebraic equation invented by Beddoe) was equated with the Celtic population.

The division of the population on the basis of hair and eye color was an artificial construct. Other formulas could be, and were, used to produce widely varying re sults, depending on the values given to each character istic. Anthropometry reflected a search for characters that would display the correct ranking of race, which reflected the assumptions made by individual racial scientists.

Ethnicity and Boundary Making in the Balkans

Crampton has investigated the relationship between cartographic knowledge, biopolitics, and the spatialization of race, through a study of maps produced by the US for the 1919 Paris Peace Conference. The Inquiry, established in 1917 and comprised mostly of academics (including geographers Isaiah Bowman and Ellen Churchill Semple), was charged with determining American policy. The aims of the Inquiry were to collect statistical information re lating to territorial claims and to produce maps of possible future political boundaries. Racial groups across Europe were identified and maps were produced for each area showing the boundaries for mixed and doubtful zones. Racial boundary lines were added where local authorities agreed and a system for delineating the boundaries of disputed territories was devised. The final system relied solely on the supposedly neutral American and British maps, rejecting other maps (Serbian, Bulgarian, Greek, and Albanian) as propaganda. While the Inquiry focused on mapping race, aspects of perceived nationality, lan guage, and religion (terms that were used in an ambiguous and sometimes interchangeable fashion) influenced de cisions. Many boundary decisions were made on linguistic grounds, an ethnic, rather than a biological category. A black book, produced for use by US conference repre sentatives, recommended boundaries across Europe and included 22 maps, many of which focused on eastern and

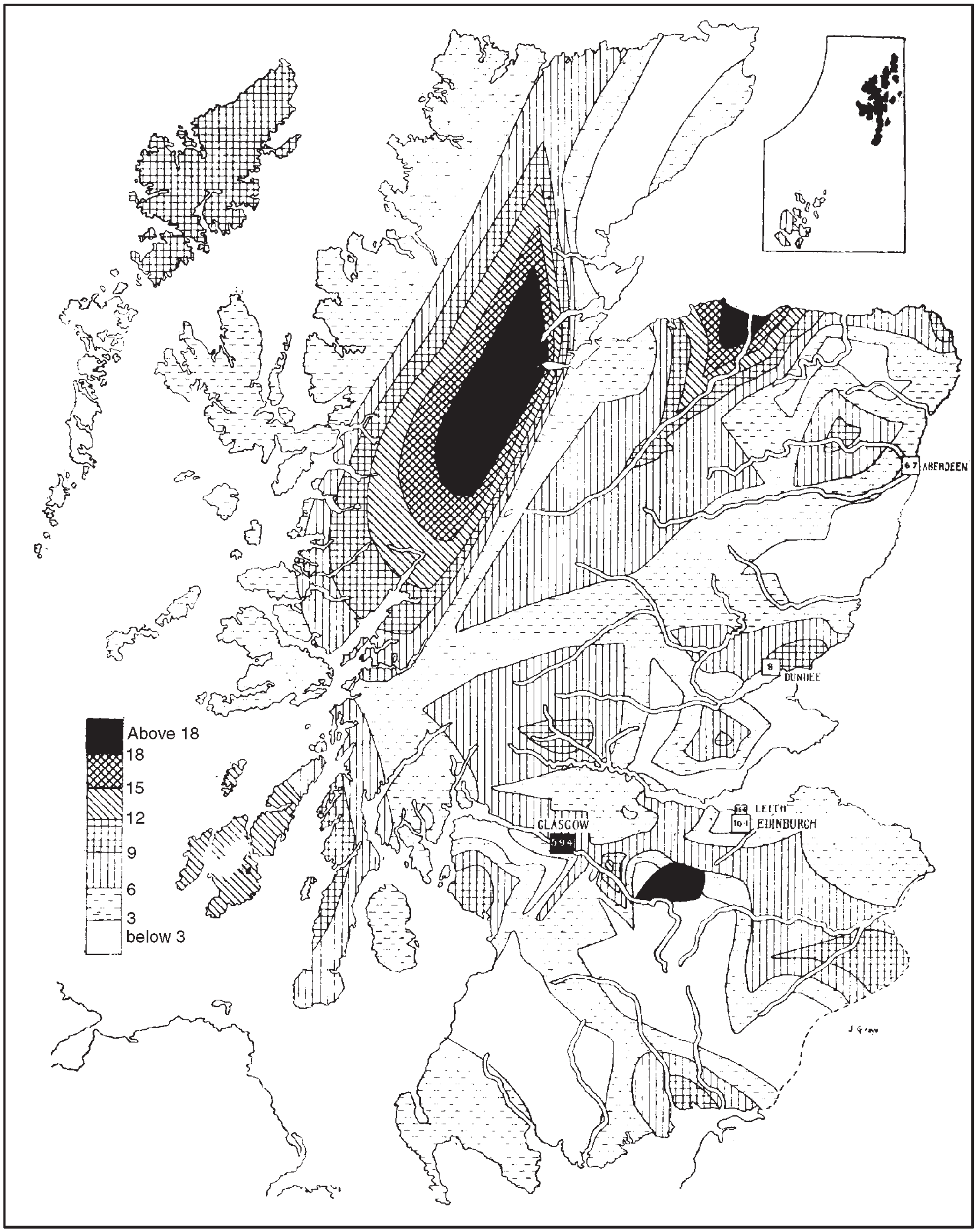

Figure 5 Divergence map for boys' eyes, Scotland. From Gray, J. (1907). Memoir on the pigmentation survey of Scotland. *Journal of the Royal Anthropological Institute* 30, 104 124.

southeastern Europe – the main area in dispute being the Balkans region. For Yugoslavia the black book recom mended that boundaries follow linguistic divisions, apart from in two areas. The Inquiry adopted Jovan Cvijic's (1865–1927) suggestion that the northern Albanians be united with western Serbia and southeastern Montenegro, under Yugoslavia and ignored local claims. In Istria, a line was placed further to the east than the linguistic boundary placing 370 000 Yugoslavs in Italy. The boundary line was disputed after World War II, resulting in the killing and deportation of Italians. The Inquiry delineated boundaries across Europe, which was assumed to reflect an underlying racial and ethnic partition – categories defined by out siders, and not always relevant at the local level.

Jansen has analyzed the role of nationality statistics and maps in representations of the post Yugoslav wars, noting that most academic work in the field has included pre war 1991 census data on nationality per republic. Maps of ethnic distribution, based on this data, served the purposes of the engineers of ethnic cleansing, but critical observers have used these maps to contrast the complexity of the pre war situation with the neat post war maps – testifying to the processes by which territories were homogenized. Yansen critiques the uncontextualized use of the pre war maps, which entails dangers of misrepresentation. The power embedded in the map texts, and their utility for political purposes, was served through cartographic sign systems. The use of color to differentiate ethnic groups represented majority territories and did not denote a re gion of only one national group, representing anything between 50% and 100% of the population (based on 1991 data). In some instances, maps of Bosnia showed a group as belonging to an area, which included less than 50% of that group. Some of the map areas with no absolute majority were represented as blank indicating a misleading sense that homogeneity is the norm, as apparently represented in the colored areas. Jansen (2005, 48) concludes that: "Cartography cannot be seen in isolation from struggles for power/knowledge: mapmakers make choices about what to show or not and they select a certain overlap between statistics and territory."

Race, Ethnicity, and Genetics

While the acceptance of race as a social construction is widespread among social scientists, the idea of race and ethnicity as definable categories remain widely accepted in society. Categories are reinforced through patterns of group identification; the media; and the use of ethnic categories in national censuses – which reflect govern mental concerns of population control, as well as the immigration history of the country concerned (for ex ample, the 2001 UK census include categories for Indian, Pakistani, and Bangladeshi). A recent UK study by Smart *et al.* reveals that the census categories (as defined by the Office of National Statistics) are widely applied in gen etic medical research. While some biologists have re jected the concept of race as meaningless, others rely on it extensively. Some scientists argue that current race categories should continue to be used at present (partly due to the portability of these categories into other fields and the current lack of genetically precise boundaries). Others recognize that the messy categories of race and ethnicity make it illogical to continue to rely on these divisions and argue for the need to use a different basis of group identification, based on current genetic knowledge (which would translate less well into the public consciousness).

Marks, Tishkoff and Kidd, and Jorde and Wooding have argued that broad geographical groupings can be defined, such as European and African, but note that these categories are overlapping and continuous, with genetically intermixed groups existing at the boundaries of geographical areas. Populations are thus related to geographical clines and are understood in relation both to local adaptation to climate and to interbreeding. However, discrete boundaries, as defined by national censuses, continue to be used. Marks reports that at tempts to map racial populations using genetic criteria have resulted in different classifications and mapped representations (reflecting failed past attempts to find a single classificatory system).

Reflections on past scientific practices reinforce the idea that present science is also underpinned by societal concerns, which means that ethical considerations should be at the forefront of research based on racial and ethnic categorizations. Often a focus on disease in ethnic groups (e.g., sickle cell anemia in the west African population) can result in groups which have been historically discrimin ated against, being selected for research. Cultural values influence scientific analysis, including genetic analyses, and scientists continue to bear a social responsibility for conclusions reached in analyses of the human species.

Conclusions

Through a selection of cartographic texts this discussion has illustrated that there has been a tradition of mapping racial difference, dating at least to the Medieval period. Through external power, based on the assumption that categories can be easily delineated and defined spatially, and through internal power, inherent in the sign systems of cartographic texts, cartography has been used to re inforce ideas of racial hierarchy, to support imperialism, and to legitimate nationality claims. While anthropo metric measurement became outdated in the mid twen tieth century, modern governments continue to monitor the social body through the collection of data based on ethnic group, nationality, and religion, and genetic re search currently relies on national classifications of eth nicity. These systems are based on the assumption that every individual can be neatly classified and mapped, leaving little space for ideas of contested identities, which have now become accepted in wider academic and social discourse. Current understandings of nationality re inforce the idea that identity is bounded by territory when in fact identity is often hybrid and contested, and transcends neatly imposed geographical boundaries.

See also: Cartography, History of; Darwinism (and Social Darwinism); Ethnicity; Lamarck(ian)ism; Maps and the State; Other/Otherness; Race.

Further Reading

Crampton, J. W. (2006). The cartographic calculation of space: Race mapping and the Balkans at the Paris Peace Conference of 1919. *Social and Cultural Geography* 7(5), 731 752.

Crampton, J. W. (2007). Maps, race and Foucault: Eugenics and territorialization following World War One. In Crampton, J. W. & Elden, S. (eds.) *Space, Knowledge and Power: Foucault and Geography*, pp 223 244. Aldershot: Ashgate.

Friedman, J. B. (2000). *The Monstrous Races in Medieval Art and Thought* (2nd edn.). Syracuse, NY: Syracuse University Press.

Gray, J. (1907). Memoir on the pigmentation survey of Scotland. *Journal of the Royal Anthropological Institute* 30, 104 124.

Harley, J. B. (1988). Maps, knowledge and power. In Cosgrove, D. & Daniels, S. S. (eds.) *The Iconography of Landscape: Essays on the Symbolic Representation, Design, and Use of Past Environments*, pp 277 312. Cambridge: Cambridge University Press.

Harley, J. B. (1989). Deconstructing the map. *Cartographica* 26(2), 1 20. http://utpjournals.metapress.com/content/e635782717579t53 (accessed in Apr. 2008)

Harley, J. B. (2001). *The New Nature of Maps: Essays in the History of Cartography*. Baltimore, MD: Johns Hopkins University Press.

Jansen, S. (2005). National numbers in context: Maps and stats in representations of the post Yugoslav wars. *Identities: Global Studies in Culture and Power* 12, 45 68.

Jorde, L. B. and Wooding, S. P. (2004). Genetic variation, classification and 'race'. *Nature Genetics* 36, S13 S15. http://www.nature.com/ng/journal/v36/n11s/index.html (accessed in Apr. 2008).

Kobayashi, A. (2003). The construction of geographical knowledge: Racialization, spatialization. In Anderson, K., Domosh, M., Pile, S. & Thrift, T. (eds.) *Handbook of Cultural Geography*, pp 544 556. London: Sage.

Marks, J. (1995). *Human Diversity: Genes, Race and History*. New York: Aldine De Gruyter.

Martin, P., Ashcroft, R., Ellison, G. T. H., Smart, A. and Tutton, R. (2007). *Reviving 'Racial Medicine'? The Use of Race/Ethnicity in Genetics and Biomedical Research, and the Implications for Science and Healthcare*. London: St George's, University of London.

Oldroyd, D. R. (1994). Griffith Taylor and his views on race, environment, and settlement, and the peopling of Australia. In Branagan, D. F. & McNally, G. H. (eds.) *Useful and Curious Geological Enquiries Beyond the World. Pacific Asia Historical Themes: The 19th INHIGEO Symposium*, pp 251 274. Sydney, NSW, Australia, 4 8 July. Springwood, NSW: International Commission on the History of Geological Sciences, Conference Publications.

Ripley, W. Z. (1899). *The Races of Europe: A Sociological Study*. London: Kegan Paul, Trench, Trubner.

Ryan, S. (1996). *The Cartographic Eye: How Explorers Saw Australia*. Cambridge: Cambridge University Press.

Smart, A., Tutton, R., Martin, P., Ellison, T. H. and Ashcroft, R. (2008). The standardisation of race and ethnicity in biomedical science editorials and UK biobanks. *Social Studies of Science* 38(3), 407 423.

Stocking, G. W. Jr. (1968). *Race, Culture and Evolution*. Chicago, IL: Chicago University Press.

Taylor, T. G. (1919). Climatic cycles and evolution. *Geographical Review* 8(6), 289 328.

Tishkoff, S. A. and Kidd, K. K. (2004). Implications of biogeography of human populations for 'race' and medicine. *Nature Genetics* 36, S21 S27. http://www.nature.com/ng/journal/v36/n11s/index.html (accessed in Apr. 2008).

Whitfield, P. (1994). *The Image of the World: 20 Centuries of World Maps*. London: British Library.

Winlow, H. (2001). Anthropometric cartography: Constructing Scottish racial identity in the early twentieth century. *Journal of Historical Geography* 27(4), 507 528.

Winlow, H. (2006). Mapping moral geographies: W. Z. Ripley's races of Europe and the U.S. *Annals of the Association of American Geographers* 96(1), 119 141.

Winlow, H. (in press). Griffith Taylor's cartographic experiment: Racial strata and geological anthropometry. *Geographical Research*.

Wood, D. (1993). *The Power of Maps*. London: Routledge.

Woodward, D. (1987). Medieval *Mappaemundi*. In Harley, J. B. & Woodward, D. (eds.) *The History of Cartography, Volume One: Cartography in Prehistoric, Ancient and Medieval Europe and the Mediterranean* pp 286 370. Chicago, IL: University of Chicago Press.

Relevant Websites

www.censusscope.org
Census Scope Website, US. Link to maps of race and ethnicity based on data from the 2000 census.

http://www.genome.gov
National Human Genome Research Institute, US.

http://www.lib.utexas.edu
Perry Castaneda Library Map Collection, University of Texas, Austin, US. Collection of maps of Bosnia and the former Yugoslavia, including maps of ethnicity.

Mapping, Topographic

P. Collier, University of Portsmouth, Portsmouth, UK

Glossary

Geodesy The science concerned with the shape and size of the Earth.

Hachures Lines on a map showing relief, normally running down the slope with the length of the hachure showing the length of slope, the thickness indicating the steepness and the density indicating the degree of slope.

Lithography Literally printing from stone, but usually printing from a greasy ink image on a printing plate, where noninked areas are kept wet to stop printing ink adhering.

Map Projection A mathematical transformation between a spheroid, representing the surface of the Earth, and the plane of the map.

Meridian Arc A part of a meridian line joining the two Poles, usually used to determine the shape and size of the local spheroid.

Spheroid A mathematic shape approximating to the shape of the Earth. Different spheroids are used in different parts of the Earth due to the irregular shape of the Earth.

Introduction

Topographic maps have become so widely available and used by geographers that their existence is taken for granted. However, even 100 years ago such mapping existed for only a few countries in Europe, India, and Java. Scattered coverage existed for parts of North Africa and the Middle East and even the United States had very little coverage. In the rest of the world topographic mapping was nonexistent. Before 1900 it was not even clear what constituted a topographic map. Wheeler, writing in 1885, used the term, but did not attempt to define it. The distinction must have been unclear in the minds of many at that time since a number of writers were at pains to define exactly what they meant by topographic maps. One of the clearest was that given by the British Colonial Survey Committee, the body charged with the oversight of mapping by colonial survey departments. In 1906 they defined a topographic(al) map as "an accurate representation of a portion of the earth's surface showing all the accidents of the surface, both natural and artificial, which can be depicted on the scale selected." Both Hinks, writing in 1913, and Beaman, in 1928, noted the importance of scale, with Beaman also noting that the map is designed to portray "certain selected features." That is, since a topographic map cannot show everything, what it does show is a function of the scale of the map and a deliberate process of selection.

Topographic maps are normally produced by national mapping agencies on behalf of national governments, although some countries have relied upon outside bodies, such as private survey companies, to meet their needs. Topographic maps are designed to meet a country's basic general purpose mapping needs, and to act as a base for showing thematic information, for example, on soils or geological maps. The normal scale range for topographic mapping is from 1:5000 to 1:250 000, although some larger scale maps may be referred to as topographic, as may some at scales of 1:500 000 or even 1:1 000 000. However, in general, maps at scales larger than 1:5000 are normally referred to as plans, and those at scales smaller than 1:250 000 are referred to as geographical maps.

The topographic maps of the present day, whether on paper, or in digital form, are the products of a long evolutionary process. This process has left an indelible mark on both the look and the contents of topographic map. Without an understanding of the process by which the maps evolved it is not possible to understand why they look the way they do.

The need and ability to create graphic representations of the surroundings seems to be inherent in human societies. It is, however, a long way from that drive to the decision to create accurately scaled representations of the world. The first step in the process of making accurately scaled representations probably resulted from the need to measure tracts of land to delineate property boundaries and for taxation, so called cadastral surveys. Surviving texts from classical antiquity, such as the *Corpus Agrimensores*, describe this process, but Roman maps of wider areas show little regard to geometric fidelity.

The earliest attempts to create maps that are accurately scaled and oriented, and which point the way to topographic maps, were in the Renaissance. To create an accurate map it is first necessary to create a robust framework of control points to which all detail can be related. The first practical control frameworks were created using the technique of triangulation, first described by Gemma Frisius in 1533. Willebrord Snell (Snellius) first applied the technique successfully over a large area in his triangulation of Holland in the early seventeenth century. In 1571, Leonard Digges provided the first description of the plane table, a simple and cost effective tool for surveying the detail that is shown on the map (**Figures 1** and **2**).

To create accurate maps of extensive areas it is not enough to have a technique for control surveys and equipment for the detail surveys. It is also necessary to have some way of determining the precise position on the surface of the Earth of the area being surveyed, and to be able to transform measurements made on the curved surface of the Earth into the positions of points on a flat sheet of paper. Astronomy provided the only solution to the problem of providing the precise position of a point on the surface of the Earth. However, to carry out astronomical observations of sufficient precision required a good telescope and an accurate timepiece. Neither of these was available before the middle of the seventeenth century. The need for a way to transform measurements from the curved Earth onto a flat sheet of paper had been recognized by the ancients. Ptolemy, working in Alexandria, had provided instructions for the construction of projections to carry out this process. However, these early projections were unsuited to accurate mapping at medium to large scale. The first usable projections were to come from the work of the Cassinis, four generations of a single family that was to transform mapmaking and create the first real topographic survey.

The Origins of Modern Topographic Maps

France was the first country to attempt to create an accurate map of its territory using uniform standard and symbols. To do this, the first necessity was to determine the precise length of a degree of latitude. Abbé Picard had attempted this by measuring the meridian through Paris. However, when his work was extended to the Channel and the Pyrenees it cast doubt on the precise shape of the Earth. Jacques Cassini supported Newton's theory that the Earth was a flattened sphere, and the extension of the meridian arc seemed to support this view. Further observations of meridian arcs were made in Lapland and South America, confirming Newton's theory. This meant that a simple projection from a spheroid to a flat plane was not practical for the whole of the Earth. What was needed were a series of spheroids, each of which provided the best fit in different parts of the world. As the true shape of the Earth was more accurately defined it became necessary to refine the existing spheroids, or to develop new ones.

César François Cassini de Thury was responsible for filling in the topographic detail on the framework largely surveyed by his father, Jacques Cassini de Thury. All but a few sheets had been completed by the time of his death in 1784. The work was finished by Jean, the son of César François, with the 182 sheets being completed in 1793, and published at a scale of 1:86 400 (*une ligne pour cent toise*). France was thus the first country to have a true topographic survey covering its entire territory.

Even before the French topographic map had been completed other European countries were following their example. Systematic survey work was initiated in Austria in 1762, but this was originally conceived as a cadastral survey. In Prussia a start was made on triangulation in 1781, followed by detail survey in 1785, but little was achieved before the disruption of the Napoleonic Wars. Norway's first military topographic survey was ordered in 1773, but it only covered the southern part of the country. In Sweden a start was made in 1758, but it remained limited in scope until well into the twentieth century. In Britain the first systematic triangulation was carried out in 1784, as part of a scheme to link the London and Paris observatories, but the national survey was not established until 1791 (see **Figure 3**). In most other European countries a start on systematic mapping did not start until after the Napoleonic Wars.

Outside of Europe, only in India was any attempt being made to create systematic mapping cover. The initial drive was financial as the East India Company realized that income derived from rents and land taxes could be as profitable as trade. In each of the presidencies, Madras, Bombay, and Bengal, survey departments were established to carry out cadastral mapping. Quite quickly it was realized that there was also a need for a general purpose topographic map and a start was made on a suitable triangulation scheme covering the whole subcontinent. Under directors such as George Everest, the Great Trigonometrical Survey became one of the most important technical and scientific endeavors of the nineteenth century and laid the foundations for much of modern geodesy.

Topographic Maps in the Twentieth Century

By 1900 most states in Europe had programs for topographic mapping, and a start had been made in North America, East Asia, and parts of Africa. At the rate of progress seen in the nineteenth century it would have taken several centuries to complete world coverage. However, two factors interacted to speed up the rate of mapping that meant that topographic mapping covered most of the world by the end of the twentieth century. The first factor was that innovations in the process of data collection considerably enhanced the speed at which mapping could be produced. The most important of these was the development of air survey. Although there had been attempts to take aerial photography in the nineteenth century, and to use terrestrial photography for mapmaking, it was the combination of the aeroplane and the camera in the early years of the twentieth century

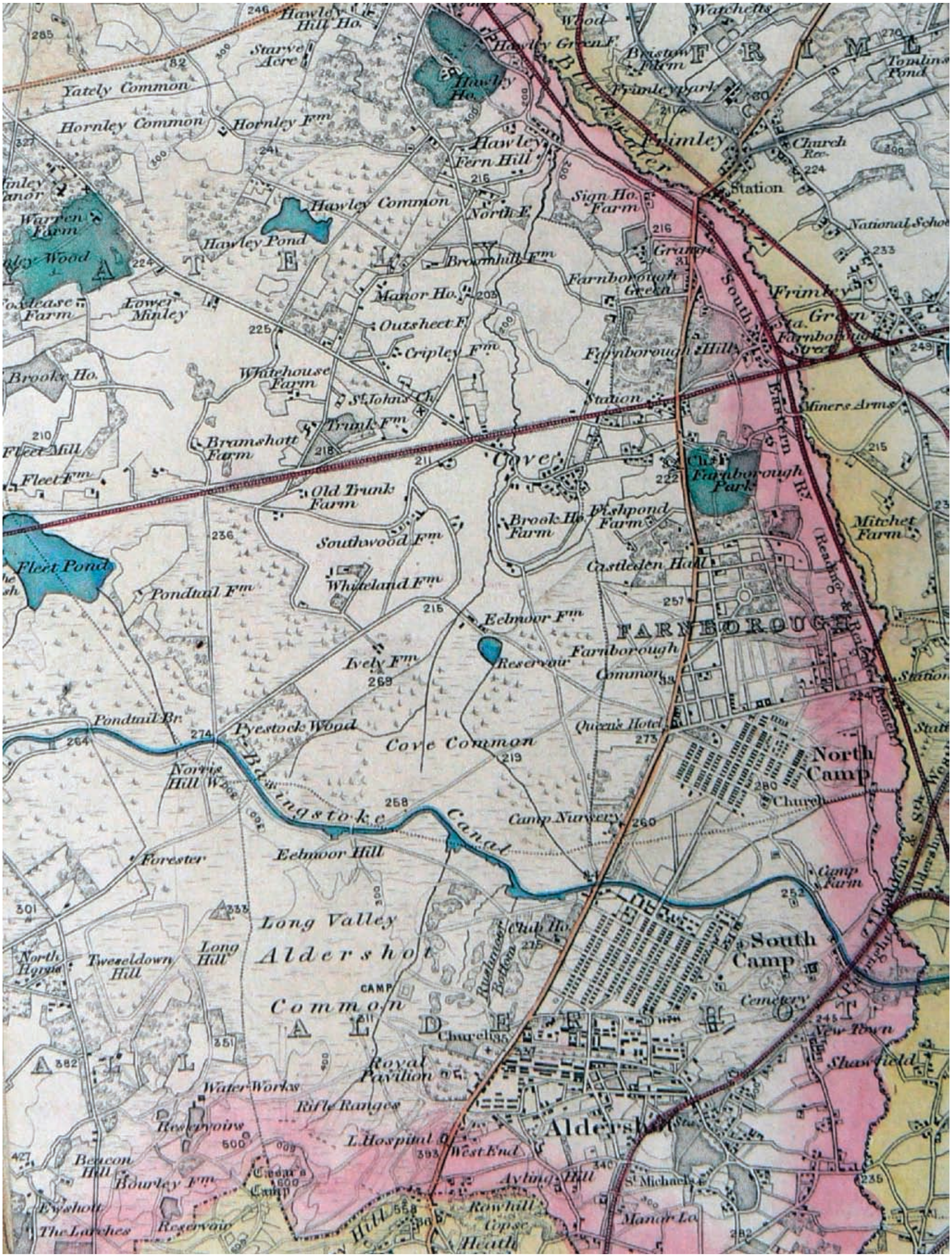

Figure 1 Extract of a 1st Series Ordnance Survey 1-inch map of the Aldershot area, hand colored.

Figure 2 Extract of a French 1:80 000, showing the heavy hachuring which tended to obscure other detail.

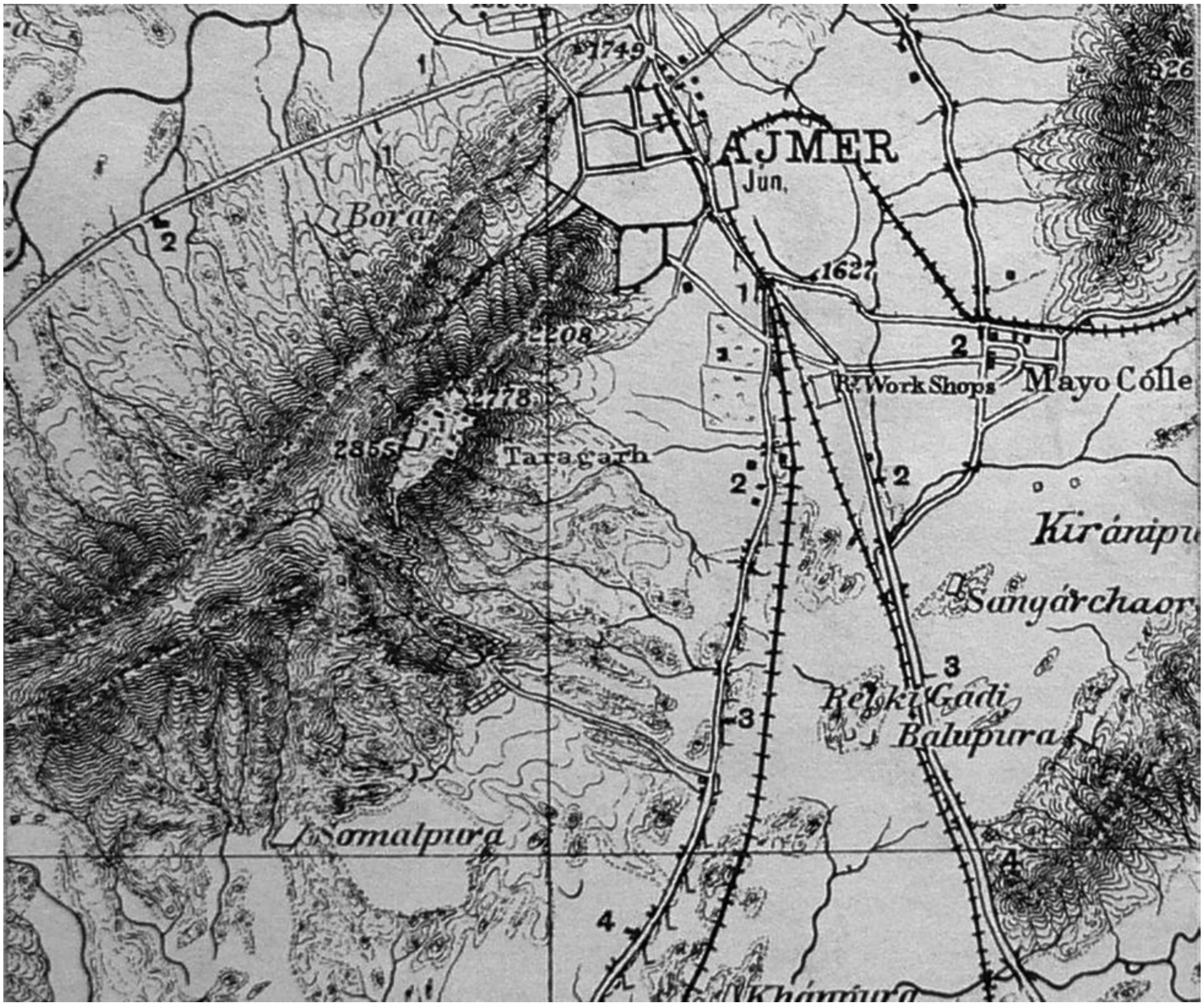

Figure 3 Extract of an Indian 1-inch map from about 1880, showing the typical method of relief depiction.

that made air survey into a practical proposition. The second factor was the global conflicts of the twentieth century that created an unprecedented requirement for topographic mapping for military purposes. The combination of air survey and military requirement first occurred during World War I, when combatant nations needed to create maps of territory occupied by their enemies. Aerial photography offered the only solution to this problem. In Europe aerial photography was mostly used to revise the topographic maps that already existed, but in the Middle East, in Palestine and Mesopotamia, it was used to create mapping of previously unmapped areas.

Progress in the development of air survey was slow in the inter war period, not helped by the financial stringency of the time, but organizations such as the US Geological Survey and the Tennessee Valley Authority started regular topographic mapping programs based on air survey in the mid 1930s. Following the outbreak of World War II, and the domination of Western and Central Europe by Nazi Germany, British and American forces became almost wholly reliant on air survey to produce the maps required for the liberation of the occupied countries.

Following the end of World War II and the outbreak of the Cold War, the Western allies embarked upon ambitious schemes to create mapping of any areas in which they might need to intervene. In part, these programs also fulfilled the needs of the countries mapped for maps as basic tools of development. At the same time as the Western allies were busy with their mapping programs, the countries of the Soviet Bloc were busy creating up to date maps of their own territories, and then going on to map their potential allies outside of Eurasia. Perhaps the most ambitious program of all was in the Soviet Union, which set out to map all of its territory at a scale of 1:25 000. This ambitious program was not long completed when the Soviet Union was itself dissolved.

However, there is now a program in place to map the whole of the Russian Federation at 1:10 000. None of this would have been possible without the use of air survey.

An important factor in the development of most na tional mapping agencies, and the kinds of maps they produced, was the fact that many of the agencies had their origins in military survey. Before 1900 civilians were in control of mapping in only two countries, the United States and Canada. This military control of mapping had two important consequences, the topographic maps were produced primarily for military users, and in some countries topographic maps were classified and not available to the public. Although the Cassinis had funded the first national topographic map by public subscription, during the Napoleonic period the national survey be came a military responsibility and the maps were classified. Military control of mapping did not cease in France until 1940. British colonial survey departments were, in theory, civilian organizations, but the directors were frequently serving army officers, and they took their orders from the Colonial Survey Committee, which was dominated by the War Office. In the French colonies, mapping was the responsibility of the Service Geo graphique de l'Armee. As survey departments were es tablished in South America, it was almost always as a branch of the General Staff, under full military control. The only exceptions to this pattern were the white settler colonies in Australia, South Africa, and New Zealand, where, from the first, the survey departments were under civilian control.

The impact of military requirements on topographic maps was not limited to those produced by military controlled survey departments. Even the United States Geological Survey was influenced by military needs. Their definition of woodland is divided into five types of which the first, 'woods brushwood', is "Areas containing tree cover or brush that is potential tree cover. The growth must be at least 6 feet tall and dense enough to afford cover for troops."

The Content and Appearance of Topographic Maps

The contents and appearance of topographic maps has always been a function of the technical limitations of the time and some innate conservatism within the organ izations producing them. When the Cassinis introduced the first national topographic maps, there was only one way of reproducing the maps, by printing from engraved copper plates. The engraving of the copper plates was an expensive and time consuming task, performed by skilled engravers. The map was then printed as a black image on paper using a flatbed printing press, one copy at a time. The technology of the time did not permit printing in more than one color. If a colored map were required, it would have to be hand colored using watercolors or ink.

In the early nineteenth century a new printing pro cess, lithography, was gradually being adopted. In litho graphic printing an image was drawn on the flat surface of a block of fine grained limestone using greasy ink. If the noninked surface of the stone were made wet, ink for printing would only adhere to the inked surface. A sheet of paper laid on the stone would then take up the image. The preparation of the stone for printing was a much simpler process than the engraving of copper plates. However, it was not until the twentieth century that lithographic printing finally replaced the use of engraved plates.

This is clearly demonstrated in the treatment of relief. The Cassinis used hachures to depict relief. This was normal practice in the eighteenth and early nineteenth centuries and all the national surveys that were initiated before the middle of the nineteenth century used hach ures (**Figures 4** and **5**), although there were differences in their mode of use. In general, the length of the hachure indicated the length of the slope, the thickness indicated the steepness, and the density could be used to determine the degree of slope. Sometimes the impression of relief could be enhanced by the use of 'shadow hachures' where a light source is assumed to be illuminating the surface from the left. Left facing slopes had the hachures drawn more lightly giving the impression that right facing slopes were in shadow. The best examples of this use of hachuring were on the '*Topographische Karte de Schweiz*' the so called '*Dufourkarte*', produced between 1842 and 1864 in Switzerland.

By the middle of the nineteenth century a number of national mapping agencies had started to use contours to depict relief on their maps. This change was largely driven by a demand for more precise height information of topographic maps. However, a combination of financial constraint and the inherent conservatism alluded to above, led to many countries continuing to use hachuring for relief portrayal until the second half of the twentieth century. In Britain, contours were first introduced in the mid 1850s, but exactly how and when they should be used was still being debated in the early 1890s. In France, hachures continued to be used until 1:50 000 scale sheets finally replaced those at a scale of 1:80 000, a process that took several decades. In Germany, the 1:100 000 *Kartes des Deutschen Reiches*, continued to be published with hach ures into the 1950s, as did some of the 1:75 000 scale sheets (see **Figure 6**), enlarged to 1:50 000, published in Austria. Hachures continue to be used, even on modern mapping, to show slopes, such as on embankments or in cuttings, which are important, but would not necessarily be shown by the contours.

When hachuring was first introduced, the hachures would have been engraved, along with all of the rest of

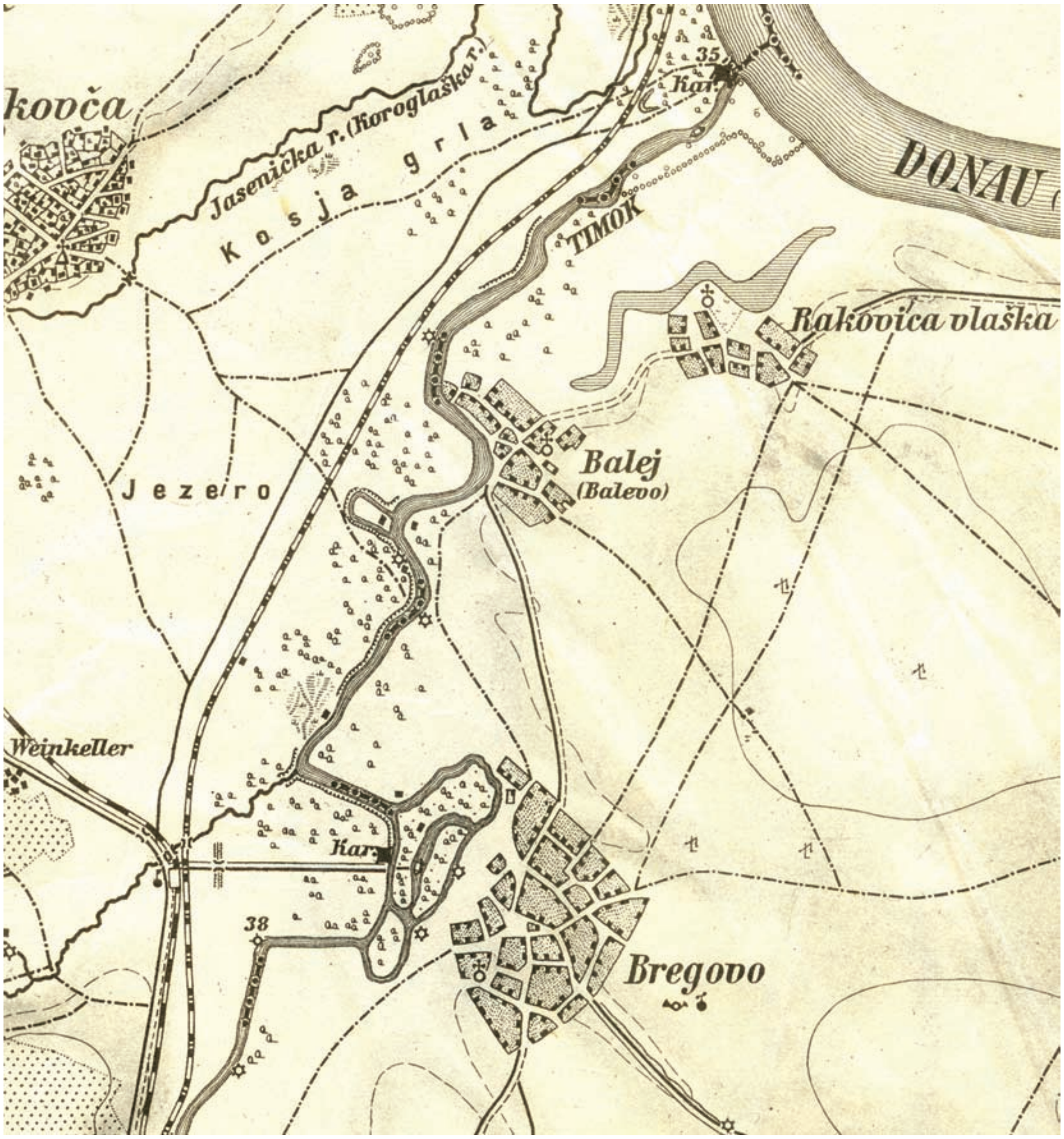

Figure 4 Extract of an Austrian 1:75 000 map from the start of World War I.

the map contents, on the copper printing plates used at the time. This was an extremely time consuming oper ation and requiring very skilled labor. On the other hand, the field survey work, which produced the drawings from which the engraved plates were made, was carried out relatively quickly. The introduction of contouring re quired much more fieldwork on the part of the land surveyors, with the physical measurement of huge numbers of points on the ground and the pegging out of contours. The drawing of the contour lines for repro duction was, in contrast, a relatively quick and easy process (**Figures 7** and **8**).

The widespread adoption of air survey techniques in the middle of the twentieth century meant that meas uring contour lines became simple and cost effective. In addition, the contours lines could all be plotted directly, without the need for interpolation, as had been the case with land survey methods. This led to a significant in crease in the accuracy of the contours shown on maps (see **Figure 9**).

The planimetric detail shown on maps, such as coastlines, drainage, communications, and settlements have hardly changed much since the surveys of the late eighteenth century, although the ways in which they are shown have changed. Unlike relief, which was shown as contours even on monochrome maps, the introduction of color printing had far reaching effects on the portrayal of planimetric information. As the earliest maps were monochrome, it was only possible to discriminate be tween different types of linear features by the kinds of lines used. For example, the coastline was commonly shown as a series of parallel lines. The high tide line was

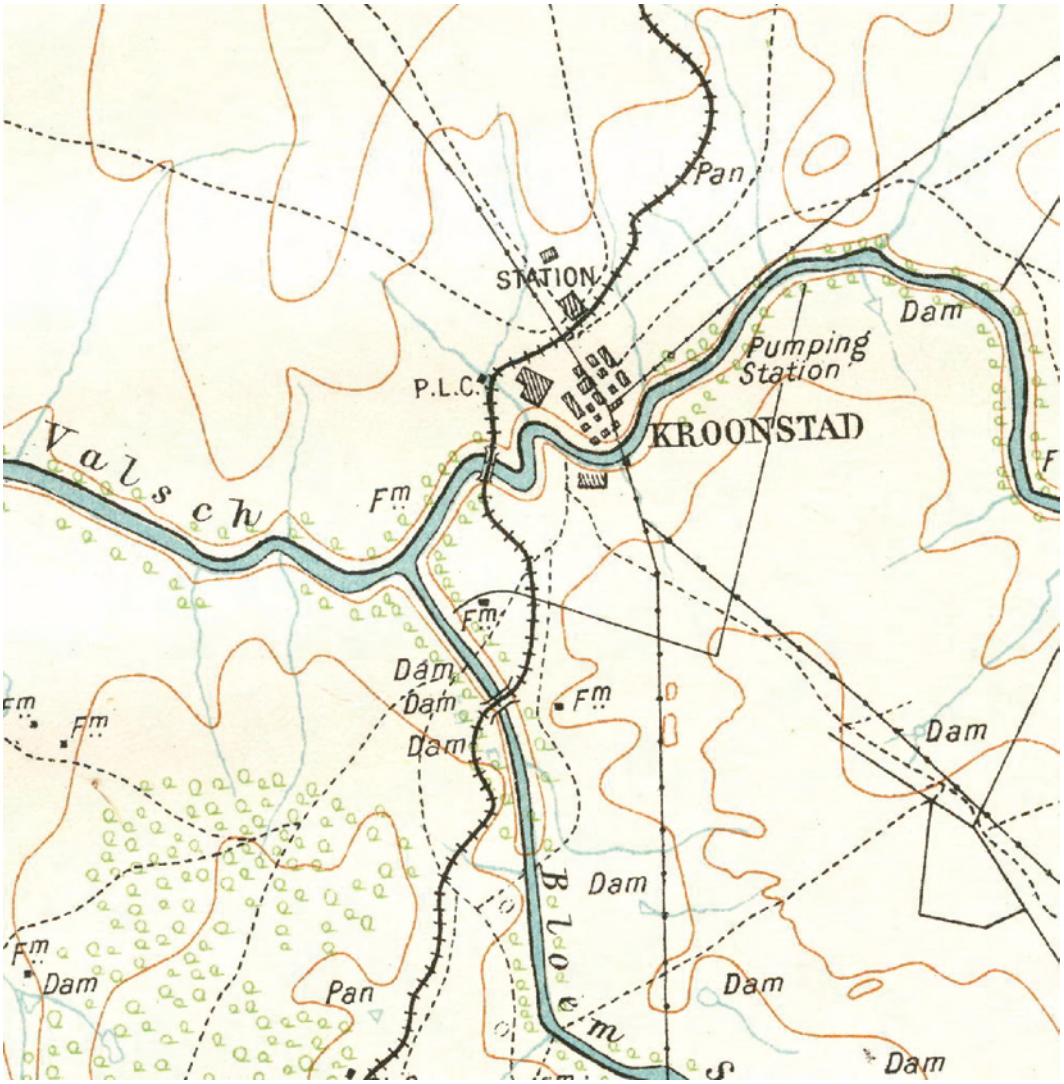

Figure 5 Extract of an Orange River Colony 1 inch to 2 miles sheet from about 1904.

a thick line, with a series of gradually thinner parallel lines representing the beach. The low tide line was then shown to be another thick line, followed by another series of gradually thinner parallel lines in the sea. The intro duction of color allowed the sea to be printed in blue, either as a solid area of color or with a vignette, gradually going from solid blue at the coastline to the white of the paper for the sea areas away from the coast. This meant that the high and low water lines could be shown as simple solid blue lines.

Rivers were shown as a solid line, usually becoming two parallel lines if the river was above a certain width. Ephemeral streams, such as in *wadis*, were usually shown as a pecked line. Again, the introduction of color meant that rivers could be shown in blue, permitting solid black lines to be used for other features without risk of mis interpretation by the map user. This was particularly im portant with early topographic maps, as it was not normal practice to include a key to the line types and symbols used. Most early maps showed roads as a pair of parallel solid casing lines, while tracks were shown as parallel pecked casing lines. It was not usually possible to show footpaths on monochrome maps. Whereas roads and tracks cut through the hachures, leaving bare paper between the casing lines, where shown, the pecks of footpaths were simply superimposed on the hachures, and could be dif ficult to identify in areas of strong relief. However, it was recognized that footpaths were an important feature, es pecially as topographic maps got increasingly used for recreational purposes. The introduction of color printing, and the abandonment of hachuring meant that a wider range of line types could be used on the maps, without the risk of confusion, or them being difficult to see against a heavy background of hachures.

With the introduction of a wider range of line types, each with its own particular meaning, and with the wider

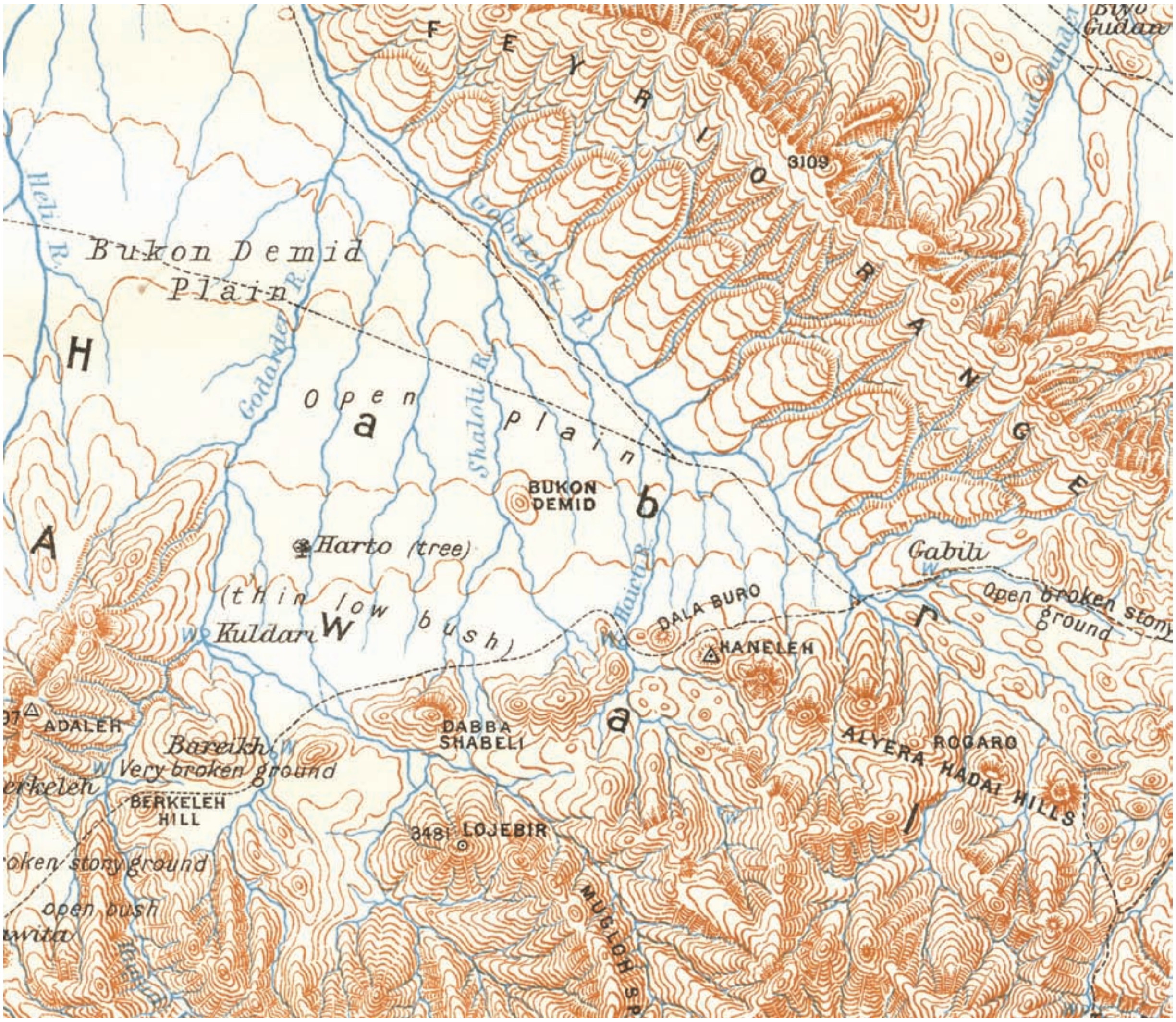

Figure 6 Extract of a 1:125 000 sheet of Somaliland, early twentieth century.

use of topographic maps by the general public, it became essential to introduce a key to the symbolization. Another factor driving the content of topographic maps, and the increasing use of symbols, was the demand by the armies for more information to be show. This was demonstrated very clearly by ‘Report of Committee on a Military Map of the United Kingdom’ produced by the British War Office in 1892. A series of recommendations were made in this report relating to the kinds of features to be shown on a proposed military map and the types of symbols to be used. The map that emerged from the committee’s report was to become the 3rd Series Ordnance Survey map of the early twentieth century. The use of hachures was dropped for the 4th Series, and the resultant map is clearly the direct ancestor of today’s Ordnance Survey 1:50 000 topographic maps. While the link between military mapping and civil topographic mapping has always been close, that link is rarely demonstrated quite so clearly as it was with the 3rd Series Ordnance Survey map (**Figure 10**).

Different countries adopted different approaches to the depiction of line features on maps once color had been adopted. Some, like the Ordnance Survey, used a color fill to show different categories of roads. Others, such as the Swiss, used difference in the thickness of the road casings to differentiate between different categories. The use of color makes the differences much clearer, and by the late twentieth century, most countries had adopted this approach.

Some point and area symbols had always been shown on topographic maps, although rarely explained in a key. The introduction of keys allowed the introduction of a wider range of symbols. The use of color has also helped in allowing a wider range of features to be shown. Point symbols were introduced to show the presence of important features and landmarks, which would help in

Figure 7 Extract of an Ordnance Survey 3rd Series sheet, showing the characteristic depiction of relief.

Figure 8 Plane table and telescopic alidade, the standard equipment of the topographic surveyor until the mid-twentieth century.

Figure 9 A modern total station, still used for topographic survey work in small areas.

navigation. The choice varied from country to country, but usually included major public buildings, places of worship, bus and railway stations. Area symbols also vary from country to country but usually include built up areas, woodland, and a variety of other land cover types. Many of the symbols used are not self explanatory, meaning their use would not be possible without a key.

In addition to graphic symbols, the body of the map will also contain lettering. As with other features, the use of lettering has evolved over time. On early topographic maps, lettering was almost entirely restricted to place names. Most mapping agencies used a combination of lettering styles and sizes to indicate a hierarchy of settlements. This was frequently related to both population size and to the administrative importance of a settlement. Most topographic maps today, also include the names of administrative areas, sometimes also including various kinds of administrative boundaries. Prominent features, such as hills, mountains, valleys, headland, and bays will also be named, within the limits dictated by the scale of the map. Drainage features, such as streams, rivers, and lakes, will also be named if the scale of the map permits.

The grid is another feature of modern topographic mapping which has its origins in military uses. Prior to World War I no topographic map was gridded. The usual practice was to show the geographical coordinates of the sheet corners, and some indication of geographical coordinates along the edge of the sheet. This practice meant that it was difficult to give a precise position of an object on the map to another user. During World War I, when artillery was forced to use maps to fire at unseen targets, systems of grids were introduced to allow artillery observers to direct the fire onto the required targets. These were found so useful as a way of referencing information, that the use of grids quickly became normal practice even for civilian mapping.

Whether topographic maps will continue to be produced in their current form is open to question. The capability to print 'on demand' from digital databases, and only showing the features required by the customer, and in the style chosen by the customer, may in the long term lead to the end of mass publication of topographic maps. The incorporation of digital topographic maps in Global Positioning System (GPS) based navigation systems may also lead to a radical rethinking of the contents. Topographic maps need to be simplified, by the removal of unnecessary information, if they are to be easily read on the small screen of a navigation system. This will also undermine the general purpose nature of the traditional topographic map.

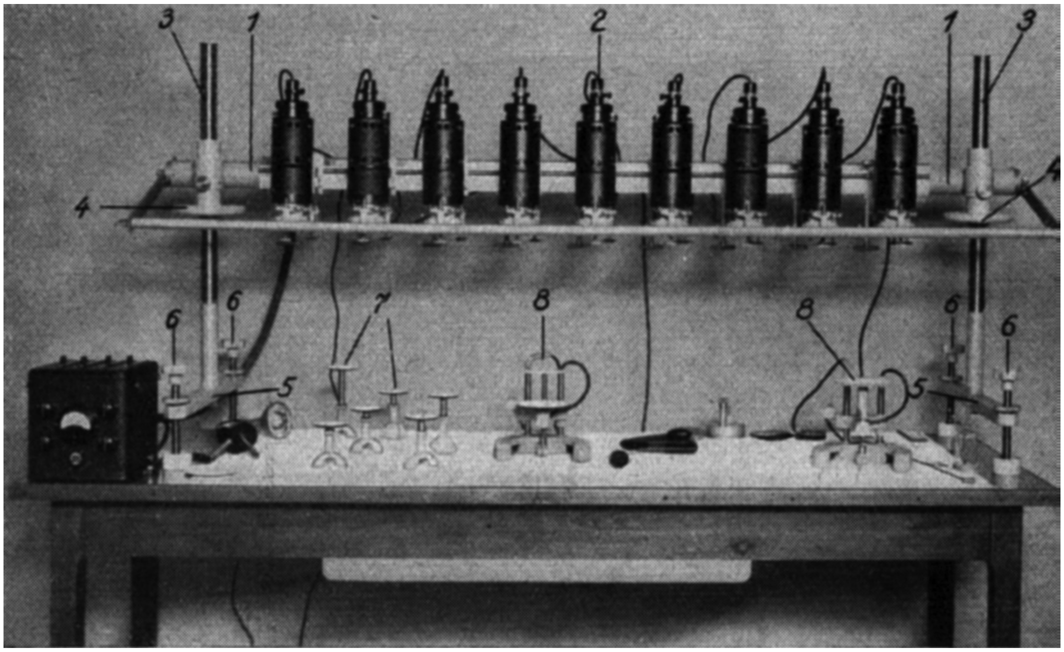

Figure 10 Multiplex, which made air survey cost-effective for topographic mapping.

See also: Cartography, History of; GIS and Cartography; Map Types; Maps and the State.

Further Reading

Collier, P. (2002). The impact of technological change in data capture on topographic mapping between 1900 and 1939. *Cartography and Geographic Information Science* 29, 155 174.

Collier, P., Forrest, D. and Pearson, A. (2003). The representation of topographic information on maps Relief. *Cartographic Journal* 40, 17 26.

Collier, P., Pearson, A. and Forrest, D. (1998). The representation of topographic information on maps Vegetation and rural land use. *Cartographic Journal* 35, 191 197.

Edney, M. (1997). *Mapping an Empire*. Chicago, IL: University of Chicago Press.

Forrest, D., Pearson, A. and Collier, P. (1997). The representation of topographic information on maps The coastal environment. *Cartographic Journal* 34, 77 85.

Harley, J. B. (1975). *Ordnance Survey Maps: A Descriptive Manual*. London: HMSO.

Harvey, P. D. A. (1980). *The History of Topographical Maps: Symbols, Pictures and Surveys*. London: Thames and Hudson.

Imhof, E. (1982). *Cartographic Relief Representation*. Berlin: Walter de Gruter.

Jagomagi, J. and Mardiste, H. (1994). Maps and mapping in Estonia. *GeoJournal* 33(1), 81 90.

Keates, J. S. (1989). *Cartographic Design and Production*. Harlow: Longman.

Kretschmer, I. (1991). The mapping of Austria in the twentieth century. *Imago Mundi* 43, 9 20.

McGrath, G. (1966). The representation of vegetation on topographic maps. *Cartographic Journal* 2, 87 89.

Parry, R. B. and Perkins, C. R. (1987). *World Mapping Today*. London: Butterworth.

Parry, R. B. and Perkins, C. R. (2000). *World Mapping Today* (2nd edn.). London: Bowker Saur.

Seymour, W. A. (1981). *A History of the Ordnance Survey*. Folkestone: Dawson.

Thompson, M. M. (1979). *Maps for America*. Washington, DC: United States Department of the Interior.

Thrower, N. J. W. (1996). *Maps and Civilisation: Cartography in Culture and Society*. Chicago, IL: University of Chicago Press.

Wilford, J. N. (1981). *The Mapmakers*. London: Junction Books.

Yashchenko, V. R. (1990). Topographic survey of the USSR completed at a scale of 1:25 000. *Mapping Sciences and Remote Sensing* 27(1), 3 8.

Relevant Websites

http://www.ngi.be
Belgian Institut Geographique National.

http://www.ordnancesurvey.co.uk
British Ordnance Survey.

http://www.gsc.nrcan.gc.ca
Geological Survey of Canada.

http://www.ga.gov.au
Geoscience Australia.

http://www.ign.fr
Institut Géographique National (IGN).

http://www.usgs.gov
United States Geological Survey.

Maps

D. Wood, Raleigh, NC, USA
J. Krygier, Ohio Wesleyan University, Delaware, OH, USA

Glossary

History of Cartography Study of the history of maps, map theory, and human context.
Mapmaking Practice of making maps, typically seen as divorced from theory and philosophical issues.
Maps Tangible and intangible objects that locate environmental and human features.

Introduction

Maps, as human cultural artifacts like cars, tables, belt buckles, and spoons, are more readily exemplified than defined. This has not deterred people from trying to define them, however, at least not since the seventeenth century when simultaneously, in places as far flung as New Spain, Japan, Russia, England, and Massachusetts, maps and mapmaking first became common. A recent collection of hundreds of definitions of the English word 'map', from 1649 to 1996, makes it plain that during this period 'map' has been more or less constantly defined as "a representation of a part of the earth's surface." (Similar collections could be made in Spanish, Japanese, Chinese, Russian, and other languages.) This nearly unanimous definition, however, has to be understood not as an outsider's impartial description of the nature and role of the map, but as a projection, as it were, of the map itself – the map as it would like to be seen – just as people project their own cultural norms into definitions of 'civilized', and those of the 'other' into definitions of 'barbarian'.

Defining the map as a representation of a part of the Earth's surface naturalizes the map, and this has the effect of universalizing it; it also obscures its origins in the rise of the state; and it ignores its role in the establishment and maintenance of social relations in those societies where it exists. Naturalized in this way, maps seem ordinary and unremarkable, indeed necessary; and for this reason contemporary scholars, immersed as they are in their own world of maps, widely conflate maps and mapmaking with fundamental human abilities like orientation, wayfinding, and other features of spatial intelligence (what a recent commentator has suggested calling "our awareness of our familiar environment"). Because of this conflation these scholars take it to be an attack on a population, as a denigration or denial of its cognitive or cultural abilities, to deny that it makes or uses maps. But just as people long lived, and as many continue to live, without writing, nonetheless carrying on a rich human life, so people have long lived and many continue to live without maps. People create maps only when the nature of their social relations calls for them. The social relations that most insistently call for maps are those of modern states everywhere in the world.

The Map Discourse Function

People make maps to discover their minds and to connect themselves. These are also the reasons people talk. But when talk becomes inadequate, either because the discourse gets too complicated, or there are too many people, or they are separated by too great distances, people develop alternative forms of communication. For the past 30 000 years people have been making artifacts that anticipate the sorts of things that we call badges, genealogies, inventories, almanacs, histories, itineraries, and maps today – 'anticipate' because the distinctions we now draw so automatically among these very different discourse functions took long to evolve, and have often only recently reached their current forms. Paleolithic peoples bundled these discourse functions together on incised bones. Since then, elaborating on Paleolithic achievements, people have constructed an ever widening repertoire of cultural forms – clothing, rituals, pottery, painting, sculpture, architecture, drawing, writing, books, prints, film, etc. – within which to encode evermore elaborate communications. Paralleling the growth in these forms have been comparable expansions in the powers of sign systems – gestural, sculptural, pictorial, pictographic, symbolic, numeric, syllabic, consonantal, alphabetic, and others – often mixing and overlapping these in rich syntheses of functions, forms, and meanings.

Among these, the map is a comparatively novel synthesis. Most English speakers use 'map' in a straightforward way to describe an artifact, still most commonly printed on paper though increasingly taking electronic form, that selectively links places in the world ('theres') to other kinds of things ('thises') – to taxes and voting rights, to species abundance or incidence of rainfall, to the presence or absence of roadways, etc. – for the purpose of underwriting the reproduction (or contestation) of the social relations of power. That is, maps are more or less permanent, more or less graphic artifacts supporting the descriptive function in human discourse linking territory to other things, advancing in the process the interests of those making (or controlling the making) of the maps. Relatively recent phenomena, these map artifacts have comparatively shallow roots in human history. Almost all maps ever made have been made during the past hundred years, the vast majority in the past few decades. So many maps are made today, and in such numbers, that no one any longer has any idea how many. The numbers printed daily just by the world's newspapers are effectively un countable, but easily number in the billions. In contrast, maps surviving from everywhere in the world for all of human history prior to the rise of the modern state number, in a very inclusive definition of the map, in the very low thousands, as if all the humans on the planet had made a single map each year – here one, there another – across the preceding couple of millennia.

Paralleling the explosion in map numbers that took place with the rise of the modern state was a corres ponding penetration of the map into ever further reaches of daily life. If there *is* a sense in which maps may be said to have existed in the medieval and ancient worlds, they were largely confined to sporadic large scale property control, and rare small scale cosmological speculation, functions (with the exceptional appearance in China of a military mapping function as early as the third century AD, and the late medieval appearance in the Medi terranean of a coastal sailing function). But in sharp contrast, beginning in the sixteenth century, vast swaths of territory were increasingly subjected to systematic surveys by newly self conscious states. In 1559, for ex ample, the Hapsburg Emperor, Philip II of Spain, com missioned a detailed survey of his possessions in The Netherlands, in 1566 of those in Spain, in 1575 of those in southern Italy, and in 1577 of those in New Spain; in 1591, the Japanese hegemon, Toyotomi Hideyoshi, ordered all *daimyo* to submit summary cadastral records and maps for the construction of a countrywide *cadastre*, and in 1604, the shogun Tokugawa Ieyasu ordered the submission of a second set of cadastral and cartographic documents; in 1663, Louis XIV's minister for home af fairs, Jean Baptiste Colbert, commissioned the collection of surveys and maps to cover all of France; and in 1666, the governor of Siberia commissioned the mapping of Siberia. Most early modern states initiated similar pro jects. If all the commissions were not completed as ini tially intended – for example, Philip's, of New Spain, was not, returns from Hideyoshi's request were spotty – these efforts did lay the ground for increasingly comprehensive and intrusive surveys, including the nineteenth century inauguration of national topographic mapping programs – programs widely completed during the twentieth cen tury – and the production of fire and insurance atlases that detailed plans of individual houses, and even con struction details of heating systems.

Today, we map the weather in something approaching real time, the locations of sex offenders, the residences of donors to political parties and the size of their donations, school attendance zones, atmospheric ozone, the con version of rain forest to farmland, the route to any cin ema from your home address, regularly updated locations of roadblocks in the West Bank, reported instances of the West Nile virus, yesterday's crimes sorted by type of crime, the locations of tomorrow's highway construction delays, deaths in Iraq, cell phone towers, the tax value of homes, bus routes, bike paths, election returns by pre cincts, counties, and states, consumer preferences by ZIP code, etc. – Is there something we *do not* map? In fact so pervasive and taken for granted are maps that it will be hard for most readers of this volume to accept the *recency* (and continued relative isolation) of their general use, or to appreciate the *explosion* in their numbers that we continue to experience today.

The History of Mapmaking

Even more recent than maps has been an interest in their history, datable in its current form only to the 1980s. Earlier histories wed the interests of twentieth century academic cartographers to a preexisting European anti quarianism dominated by a nationalist passion for dec orative printed maps of the fifteenth to eighteenth centuries. This spawned a hero saga (Demosthenes, Ptolemy, Mercator, the Casinis, Minard, Edes Harrison, etc.) that plotted cartographic progress from humble origins in Mesopotamia to the putative accomplishments of the Greeks and Romans, the rediscovery of which during the European Renaissance led directly to the development of the triumphant scientific cartography that swept the world in the wake of Western colonialism (Lewis and Clark, Livingstone, GIS, etc.).

As we now acknowledge, this story is false in almost every particular. Although the oldest surviving un contested map is Babylonian, this map is in no way the 'origin' of mapmaking, which was originated as called for again and again around the world. Maps such as the Babylonians and Egyptians did make were not 'built on' by Greek, Roman, or subsequent 'European' mapmaking, most of which was independently invented and re invented. Indeed, Greco Roman contributions to the history of mapmaking have been unconscionably ex aggerated: if ancient Greeks actually made any maps at all, there are no surviving ancient Greek maps, but a few from the Romans. In any case, most subsequent 'Euro pean' mapmaking was in no way indebted to either of these, nor was 'European' mapmaking ever the 'scientific' enterprise it has been claimed to be. It was first and foremost a profoundly ideological activity serving na tional identity building, colonialist, and other interests; and was paralleled by similarly motivated mapmaking elsewhere in the world.

In general, looking at the historical record through the lens of contemporary mapmaking radically distorts the importance of maps for the administration of the great 'historical civilizations' by assuming they must have done things the way we do them. This has not only led his torians to assume that when people wrote about mapping they must have made maps, and that where one map survives, a hundred must have been made, but to pos tulate 'mapmaking' traditions where instead there were probably traditions of 'cosmological speculation', tradi tions of 'property control', traditions of 'centralized management', traditions of 'military mapping', and perhaps others, including, for instance, the discourse function fulfilled by 'geomantic site location'; but none precipitating 'the idea of the map', that is, for most readers of this book, so 'self evidently' common to them all. Other 'maps' appearing in the historical record likely played no part in any of these traditions, but instead arose from isolated efforts to address unique problems (laying new drains, defending property in a law suit, etc.). That is, they were based on no prior model, left no progeny, and so are akin to what geneticists call a 'sport'; which explains why they are so hard to pigeonhole as, *precisely*, map, plan, or drawing (a good example would be the plan and diagram of Canterbury Cathedral, *c.* 1153–61). As their existence and the rest of the record attests, map making was a marginal activity for all these peoples, among whom the functions served by mapmaking today, '*to the extent that they existed at all*, were served by other, typically scripted and/or numeric forms of inventory and control. This is to say that the historical record is spotty not because survival rates were low – which is in any case difficult to entertain given the higher survival rates for so many other, far less consequential artifacts – but because maps were actually infrequently made.

This of course is why uncontested maps more than 500 years old are rare at any scale from anywhere in the world. Cosmographical diagrams are more common (they are nonetheless extremely rare), and large scale plans more common still (though again the numbers are ab solutely tiny), but prior to the fifteenth century small scale geographic maps are rare almost to the point of nonexistence in any cultural tradition except that of China where they begin to appear in the twelfth century and rapidly become common. But then no unquestioned map of any kind predates the second millennium BC. Whether prehistoric humans made maps is uncertain, because the interpretation of their artifacts is mired in controversy; though if they did not it was not because they were unable, but because the discourse function served by maps either was not called for, or was fused with other discourse functions in a synthesis not recog nized as maplike today. Reputable scholars used to assert the recently discredited maplike qualities of the wall painting at Çatalhöyuk (6200 BC), and a similar case has been made for the petroglyphs at Valcamonica (2500 BC) and elsewhere, but if prehistoric humans did make maps – which is doubtful – they were neither made often nor in very many places; they likely served broadly pictorial, religious, ritual, symbolic, and/or magical functions; and their production was discontinuous with the practice of mapmaking encountered in historic populations.

The oldest extant maps about which there is scholarly consensus are, as noted, Babylonian. Dozens of large scale, Babylonian, cuneiform maps and plans survive from the second and third millennium BC, but only a couple of small scale maps survive, and these from the first millennium BC. The existence of the so called Turin gold mining map from about 1150 BC is the sole survivor of a putative Egyptian mapmaking tradition of roughly similar age that otherwise is represented only by cosmographical diagrams and pictures of gardens, canals, and other features. Recent scholarship posits an Indic tradition of mapmaking stretching back to the first mil lennium, but the earliest extant artifacts are an allegorical wall sculpture from about AD 400, and a Jain cosmo graphical diagram of the thirteenth century AD. There is textual evidence of a Hindu tradition of cosmographical globe construction dating from the first millennium BC, but again no actual globes predate the fifteenth century AD. In China, three maps survive from the second cen tury BC, but few others until the twelfth century AD when, again, maps first become common. Evidence also suggests a Tibetan mapmaking tradition rooted in the first millennium BC, though again, with the exception of a *mandala* transmitted to Japan in the ninth century AD, no survivors predate the eighteenth century. Textual evidence also supports a Hellenistic mapmaking tradi tion, but no maps survive of any character. Except for medieval European copies of Roman itineraries, no small scale Roman maps survive, despite the elaborate instructions for producing them in Ptolemy's *Geography*, and even large scale survey and property maps do not exist in abundance. That is, with respect to the ancient world there are many more textual suggestions that something 'like' mapmaking was carried out than there are surviving artifacts, the numbers of which, with the exception of Babylonian and Roman plats and surveys, *may be counted on the fingers of two hands*. That is, map making was comparatively widespread and uncommon everywhere.

The record is not much different for the medieval world. Islamic scholars elaborated sophisticated theore tical schemes for the construction of maps from the seventh century on, but if any were made, none survive from periods prior to the tenth century, and maps remain rare until the fifteenth and sixteenth centuries. In medieval Europe, handfuls of cosmographical diagrams and large scale plans are extant from the seventh century, but with the exception of the late medieval portolan charts, maps were otherwise unknown. There is textual evidence of relatively small scale mapmaking in Japan as long back as the seventh century AD, but again, nothing survives; maps of state allocations of arable property are extant from the eighth century, but no maps are common until the sixteenth. Textual evidence supports a map making tradition in Vietnam as early as the eleventh century, but again no artifactual maps predate the fif teenth century. The oldest surviving Malay maps are from the sixteenth century. No Mesoamerican maps predate the Conquest, though again there is ample reason to assume a preexistent tradition of cosmographical diagrams and some evidence of limited property (or 'community') mapping among the Nahua, Mixtec, Otomi, Zapotec, Totonac, Huastec, Chinantec, Cuicatec, and Mazatec. No indubitable maps made prior to the fifteenth century survive from sub Saharan Africa, South America, Australia, Oceania, or North America, though in many places the record was systematically destroyed, and historical research may yet uncover evidence of mapmaking traditions unknown today (sporadic map making at large and very small scales is predicted for all civilizations with large bureaucracies and extensively mediated relationships). Despite these lacunae, the re cord suggests that large scale property maps and small scale cosmographical diagrams were made rarely, but with increasing frequency, everywhere in the world since the third millennium BC. Other mini traditions seem to develop often, only to die out again, except in China. The limited number of extant map artifacts clearly sketches the tenuous hold any of these discourse functions had in the notational repertoire of these societies.

The significance of these data is obvious. Human so cieties did not need maps and got on handily without them for hundreds of thousands of years. But during the last two or three millennia BC, larger, more complicated societies including Babylonia, Egypt, perhaps the Indic societies centered on Mohenjo Daro and Harappa, and China began to articulate, sporadically and apparently independently, but among and continuous with other indigenous textual productions – memorial inscriptions, memory aids, almanacs, genealogies, inventories, his tories, and descriptions of routes and territory (in mix tures of sculptural, pictorial, pictographic, syllabic, consonantal, and/or alphabetic forms) – graphic notation systems linking 'location' with 'rights and obligations' (as in the large scale property maps), and with 'speculative attributes' of the larger environment (as in cosmo graphical diagrams). Similar graphic notation systems filling broadly similar social functions emerged fitfully in other ancient civilizations around the world, again ap parently independently, although extensive trade and other connections among these groups are acknowledged and cross fertilization undoubtedly took place. The ar ticulation of such similar notation systems in so many of these societies strongly supports the notion that map discourse functions of this character inevitably emerge in societies whose increasing size and complication call for them (the specialization required for making maps de mands at least a population of the size maps permit to function), of which, again, the best example is China. But the sporadic nature of this articulation no less strongly suggests that at the size and degree of complication reached by most ancient civilizations, the map discourse function as it has come to evolve could be satisfied by other, better established discourse functions (generally scripted and/or numeric); and so the map discourse function failed to establish itself no matter how many times it was seeded. The map discourse function is no where well rooted until the rise of the early modern state (which in China may mean the Song) with which it co evolves as an instrument of polity, to assess taxes, wage war, facilitate communications, and exploit strategic resources.

Limitations of the Ascription 'Maps'

While it is not 'wrong' to refer to these early graphic notation systems as maps, it is anachronistic. It is critical to accept, as already intimated, that these graphics were not emitted 'as maps' by those who made them. To ima gine this would be to see them through the conceptual filter created by modern mapmaking. For example, early 'map' artifacts were generally free of the 'heightened spatiality' so characteristic of what most readers of this book think of as maps, and there is no evidence that they were discriminated from other graphic textual pro ductions on this ground. Until modern times no society distinguished – or made – such maps as distinct from religious icons, *mandalas*, landscape painting, construction drawings, itineraries, and so on; and current scholarship stresses the continuity between religious iconography and that which materialized on the earliest maps. For example, the Chinese word *tu*, frequently translated into 'map', can also be translated 'picture', 'diagram', or 'chart'. *Tu* of 'geographical' subjects may have had poems pain ted on them as was common on 'paintings' of other subjects. This not only reflects the conceptual continuity that tied together the Chinese practices of what today even the Chinese think of as discrete genres ('painting', 'mapping', 'drawing', etc.), but that synthesis of painting, calligraphy, and poetry that so effectively distinguishes, say, a Ming painting from that of the European Renais sance (that, say, of Wen Cheng ming from that of Mi chelangelo). This synthesis lent Chinese *tu* an explicitly expressive character inconceivable in twentieth century conceptualizations of mapmaking, even in China.

Such inclusiveness characterizes other words fre quently translated into 'map', including the Arabic *naq shah* (painting, any kind of visual representation), its

Indian derivation *naksha* (picture, plan, general de scription, official report, etc.), the Sanskrit *chitra* or *ale khya* (painting, picture, delineation, etc.), the Latin *mappa* (cloth) and *carta* (formal document), the Mexican *lienzo* (linen, cloth, canvas, etc.), and the Aboriginal Australian *dhulan* (painting, map, diagram, graphic representation, etc.). Not only do these broadly inclusive terms not draw the distinctions among types of graphic production made by contemporary map using populations, but they refer at the same time to graphic systems that mingled what most readers of this volume keep apart. Mesoamerican *lienzos*, for example, did not privilege space as the maps of modern states do, but rather in their 'community maps' drew history and territory together (or, perhaps, from their perspective did not rip history and territory apart). Where the Mixtec made do with one, modern states insisted on using 'three' or 'four' discrete discourse functions (thus: plat, deed, title search, genealogy, etc.). Were the Mixtec discourse function to exist today, it would more likely be termed 'map history' or 'pictorial genealogy' than 'map'. Another example: Jain cosmo graphical diagrams mingled places most readers of this volume would locate on topographic survey sheets with places where 'release' is possible (places contemporary cartographers would not even locate in 'space'), thereby constructing, in the words of one expert, "a gigantic theater where transmigrations and reincarnations take place." Unlike the artifacts most readers of this volume think about as map – artifacts that discourse about the socio spatial territory we mutually inhabit – Jain cosmographical diagrams discourse about destiny. A third example: in their cosmographical diagrams, medieval Christian Europeans fused the historical commitment of the *lienzos* with the teleological orientation of the Jain cosmograms to create, in the words of another expert, "a visual narrative of Christian history cast in a geographical framework." Again, most of those reading this book would break this out into separate discourse functions (say, painting, history, maps, etc.).

None of these ways is better or worse, or more or less 'advanced', but they *are* differentially capacitated to fa cilitate life in populations of different sizes, with different rates of social and geographic mobility, and different degrees of labor specialization and hierarchic integration. (Breaking up discourse functions enables specialization, which in turn supports hierarchic integration. This in turn permits higher rates of growth and mobility without loss of social integrity.) There is, here, no question of quality (or even utility); there is no 'contest'. The *lienzos* served the Mixtec, as their cosmographical diagrams did the Jains and the medieval Christians, every bit as ef fectively as topographic surveys serve the interests of the modern nation state. The discourse functions a society evolves (chooses, or has forced on it) depend on what kind of society it is. What is at stake, ultimately, are the differences in social structure that in the cases of the Mixtec, Jain, and medieval Christian called for pictorial genealogies and cosmographical diagrams, but in the case of modern states call for topographical surveys and the construction of the 'idea' of cartography that such sur veys seemingly entail.

The Rise of Mapmaking in the Early Modern State

But indeed, few of the graphic notations produced in ancient or medieval civilizations would be considered maps today, whether we spell that 'map' (as in the United States), *mapa* (as in Mexico), *carte* (as in France), *kharitah* (as in Turkish or Arabic), *mana chitra* (as in Eastern India), or *chizu* (as in Japan). Maps construed as, or theorized in the light of, topographic surveys gained currency only in the last 300 years or so, and within this period, only in relatively stable states with entrenched, centralized bureaucracies and well established acad emies. Though few people used maps in AD 1400, by AD 1600, people around the world found them indispensable. There is a divide here that is impossible to evade. Recall the dates at which maps *really* begin to appear in the historical record: Islamic artifacts may date to the tenth century, but maps do not become common until the *fif teenth* and *sixteenth* centuries; the oldest surviving map of China may be from the second century BC, but maps are not common until the twelfth and only become abundant in the *seventeenth* century; large scale Japanese maps may survive from the eighth century, but national and pro vincial maps only begin appearing in the late *sixteenth* century and are not common until the *seventeenth*; the oldest surviving Hindu globe is from the *fifteenth* century; Vietnamese and European maps become plentiful only in the *fifteenth* and *sixteenth* centuries; Mesoamerican maps survive largely from the *sixteenth* century; and Malay maps from the *sixteenth* century. Again and again we find large, centralized societies, from everywhere in the world, inaugurating *mapmaking* traditions during their transition to the early modern state (a transition China may have begun in the Song).

For mapmaking, this transition has had the recent attention of scholars working on Japan, China, Thailand, Russia, Europe, the North American colonies, New Spain, and elsewhere; and there is reason to believe that not dissimilar processes were at work in all societies struggling with a more or less common socioeconomic transformation, no matter how diverse the individual histories. Undoubtedly, the nascent European mapmak ing tradition was transported around the globe; but the ability it demonstrated to 'import' material from other traditions (well documented, e.g., in the cases of Islamic, Burmese, Chinese, and Japanese mapmaking) and the 'ease of its apparent adoption', seems actually to describe a 'merging' of mapmaking traditions that we are prepared today to acknowledge as at equivalent levels of devel opment, a merging into a kind of 'transnational' or 'worldwide' tradition that differentiated not West from East, but modern nation states from the smaller face to face societies out of which they evolved and which they would soon enough gobble up.

Intriguingly, the functions the new maps initially served were not those that might strike us as obvious (e.g., wayfinding); nor were the state functions they did ini tially serve newly created in the fifteenth century (they were functions that had been previously served by scripted forms, even by talk). For example, in 1602 the *duc de Lesdiguières* commented to Henry IV of France that, "Your majesty will understand much better than I can set it out in writing, if [you] will look at the map of Dau phiné with the Piedmont border," while Michelangelo complained that if only the Hapsburg Emperor, Charles V (r. 1519–58), "… had ordered a drawing to be made of the course of the river Rhône, he would not have met with loses so severe, nor retired with his army so dis arrayed." Charles, in fact, did use maps, extensively. About the very battle to which Michelangelo referred, Martin de Bellay wrote of seeing Charles, "Studying the maps of the Alps and the lower region of Provence so enthusiastically that the emperor had convinced himself that he already possessed the land in the same way he owned the map." More generally, Marshall Vieilleville observed in the 1560s, apropos the campaigns of Henry II of France that, "A military commander must no more move without a map than a pilot or galley captain, unless he wants to court disaster," though the most general admonition seems to have been Castiglione's of the 1520s to the effect that there were, "matters, the which though a manne were liable to keep in mynde (and that is a harde matter to doe) yet can he not shew them to others" without a map or painting. These anecdotes illustrate the growing currency of maps in the early modern period, but they also make very clear their novelty. Indeed the anecdotes seem actually to catch a more comprehensive discourse function – could we call it 'describing'? – in the very act of differentiating. These anecdotes presage a steep increase in the use of maps for military, adminis trative, and speculative humanistic purposes.

Why? What happened after AD 1400 that called people to start making maps? The canonical answers, with their focus on so called 'scientific' mapmaking and dependence on the presumption of a European excep tionalism, can be dismissed as largely irrelevant even for Europe, where the substantive explosion in mapmaking took place outside their purview. Canonical accounts have always focused on small scale mapping of the world – and on the heroic growth of European knowledge about the world – but such accounts have no bearing on the explosion in large scale mapmaking of local areas that accounted for the overwhelming bulk of new maps. Consider the northern Italian plains. Extant maps pre dating the fifteenth century can be counted on the fingers of one hand, but in the sixteenth century, mapmaking explodes. For example, less than a 1% of the 10 000 maps archived by the Venetian state predates AD 1565, and almost half – that is, thousands of maps – were com missioned solely by the Office of Rural Lands, an office founded in 1566. In another example, only a dozen maps among the 10 000 archived by the Florentine state pre dates AD 1565, where the bulk of archival maps, devoted to property control, dates from the seventeenth century. The offices commissioning most of the vast Milanese archive of over 76 000 maps were also founded in the sixteenth century: Acque (sixteenth century–1801), Acque e strade (1574–1801), and Confini (1518–1802). Identical accounts can be given for the Papal States and Naples.

But they can also be given for the rest of Europe. In England where the history of medieval mapmaking is particularly well known, scholars have identified no more than 35 domestic maps produced before 1500. Yet by no later than 1540, England's Henry VIII had available to him maps for a wide variety of purposes; in 1574 Christopher Saxton began the publication of his atlas of English counties; and in 1593 John Norden began the publication of his series of county topographies. Indeed such a mass of maps – and other papers – had been generated during the sixteenth century that in 1610 a State Paper Office had to be established to marshal them. In France, where only ten domestic maps have been found that predate 1500, maps began to be used for military purposes early in the sixteenth century, and their use gradually expanded until, under Henry IV (r. 1589–1610), the country was more or less systematically mapped. Indeed during the sixteenth century, mapmak ing took such root in France that, in 1663, Louis XIV and his chief minister, Colbert, could envision using maps for military and naval purposes, for making political and judicial decisions (especially about jurisdictions, an ob vious use today), for economic and financial planning (mines, canals, fiscal divisions, etc.), and for establishing the boundaries of ecclesiastical dioceses. By then, there were also plenty of presses capable of printing and dis tributing maps of every size and character.

But similar accounts – which in Europe could be re peated for the Spain and Austria of the Hapsburgs, the United Provinces of Maurice of Nassau, the Sweden of Gustav Adolph, and the Poland and Lithuania of Mikołaj Krzysztof Radziwiłł – can also be given for emergent states elsewhere in the world. The case of Japan is ex emplary. Extant maps from classic and medieval Japan are sufficiently numerous (well over 200, according to the latest scholarship) to suggest that Japan had one of the most robust mapmaking traditions in the premodern world. Certainly, no European polity has anything like its record to display. Though most of the extant maps are large scale maps of local property holdings (the earliest from the eighth century), a map of Japan attributed to the Buddhist priest Gyōki is believed to have been made during the early classical period. Although Gyōki type maps were occasionally reproduced in the medieval era, there is no evidence after the ninth century of either resumed national surveying or efforts to revise the classical prototype. The possibility does exist that a second national mapmaking effort took place in the late twelfth century but, again, if any such maps were actually drafted, none survives. As we have seen, however, in 1591, Toyotomi Hideyoshi ordered all *daimyo* to submit summary cadastral records (*gozen chō*) and maps for the construction of a countrywide *cadastre*; his successor, Tokugawa Ieyasu, repeated the order only 13 years later; and other surveys followed. In the late 1630s, a national map assembled from provincial surveys was released to commercial printers who issued it as a woodblock atlas. By 1700, literally ‘thousands’ of Japanese maps covering, in the words of a recent scholar, “virtually every domestic subject and in virtually every format,” had issued from government offices and commercial printers.

A different story, but to a similar end, can be told about Russia beginning with Ivan IV (r. 1533–84), both the large scale mapmaking involved in Muscovite property litigation and the small scale mapping of Siberia; about the mapping of New Spain; and about the late seventeenth and early eighteenth century mapping of the British colonies in North America. Indeed variants of the story can be told about every corner of the globe. In the words of another contemporary expert, “Medieval societies rarely produced maps. This generalization holds historically throughout Eurasia, from England to Japan. Mapping was not a routine part of any official transactions or procedures in medieval times,” and this can be expanded to the rest of the world. Yet shortly thereafter, maps were not only routine parts of numerous procedures, but they were being made in mind boggling numbers.

The explanations for this explosion in mapmaking vary, of course, from place to place. In the case of Italy, for example, a quantitative analysis argues that, “three discontinuities – times of increased mapping production – stand out: the late fifteenth century, the mid sixteenth century, and the late seventeenth century,” each of them marked by both increased rationalization of bureaucracies and pronounced upturns in the economy, most notably the ‘Italian Indian summer’ of the fifteenth century, and the late seventeenth century’s recovery from its long economic crisis. In the Japanese case, an expert draws attention to the simultaneous disorientation and reorientation that characterized sixteenth century Japan: “On the one hand, warfare wiped out not only the geography of the medieval polity but many of the petty lordships formed in its wake. Sweeping campaigns and mass transfers made governors into strangers in their own lands. On the other hand, administrative change advanced a model of integration,” and she also draws attention to the importance of a spike in urbanization. Russian map making, a scholar has recently argued, “allows us to invert the way we have come to imagine the relationship between central state mapping projects and local interests,” adding that, “In an immense, unmanageable land where centralization could never have set roots without the participation and support of local communities, maps brought local knowledge to the service of the central state,” and this was as true for the large scale Muscovite property mapping as it was for the mapping of Siberia too. The general implication that mapmaking emerges as a rationalizing tool of control during periods of relative or increasing prosperity in early state economies is broadly supported by the evidence from the Hapsburg, Bourbon, and Tudor realms, from Southeast Asia, and the North American English colonies as well.

This is all doubtless true, yet all the bureaucratic functions fulfilled by the maps during this period ‘could’ have been carried out in other ways, as they largely had been during the later middle ages. As experts on cadastral mapping have reminded us, maps are not indispensable even for *cadastres*, pointing out that even today there is no comprehensive map based *cadastre* in states like Norway and the United Kingdom where it could certainly be expected. Attempting to explain what prompted the adoption of cadastral mapping during the early modern period by so many states, these experts point out that, “Conviction of the merits of mapping was a precondition for mapping itself.” This is actually a theme – variously put – in much contemporary scholarship, where a particularly significant merit was the ability of the map to figure the new state itself.

It is important to accept that if the map was essentially a novel function during the fifteenth, sixteenth, and seventeenth centuries, so was the state itself. Although we take the state for granted today, as we do maps, nothing like it existed in earlier periods. Doubtless there were earlier polities that resemble the modern state in many ways – the Greek *polis* does, the Roman Empire does, China does under the Tang – but they differ from the modern state in essential ways as well, and in any case, the modern state did not derive directly from any of them. Although, again like maps, the state too is more readily exemplified than defined, experts on the state can point to a number of characteristics that states invariably possess, among which the development of more or less permanent, more or less impersonal political institutions is paramount. Evolving from a period in which loyalty had been directed to one’s lord, to one’s immediate community, and to one's family; and that was typified by a powerful sense of mutual obligations among face to face acquaintances, this new political structure with its im personal institutions and ultimately abstract character required new forms for its embodiment.

Contemporary scholarship is unanimous that the map possessed an all but unique power to give the abstract idea of the state some kind of form, initially to those living within the embryonic state, later to those outside it. In fact, one recent commentator has urged that, "Ever since Abraham Ortelius and Gerard Mercator published their world maps and atlases in the sixteenth century, single sheet maps had presented the sovereign states as visually and territorially unified constructs." More par ticularly, it has been argued about Russia that, "From the point of view of the state, and as experienced by its subjects, mapping the heartlands and the frontier con stituted two pieces of a single project: the creation and imaginative consolidation of a territorial tsarist empire;" about Japan that, the "nascent state struggling for survival used two general programs of registration – the cadastral survey and the cartographic survey – to put on paper, and in the minds of participants, the tropes of union;" and about the young United States that, "the image of the national map was one of the few visual artifacts dem onstrating what many perceived to be either an abstract or even untenable fiction, namely that there could be a national union between disjointed regions and politically disparate people." Such maps also spoke to outsiders, as in the case of Qing China where it has been argued that, "Using scaled maps, easily interpretable by anyone trained in the same map idiom, was an effective way to stake out claims of empire to an encroaching Europe; the Kangxi atlas defined what China was territorially to the rest of the early modern world;" and in the case of Britain, whose imperial maps sought "to persuade the maps' readers on either side of the Atlantic of British ownership rights regarding the North American contin ent." Similar conclusions have been reached with regard to early modern – and even much later – mapping pro grams in France, Thailand, and elsewhere.

The most striking feature about these assertions is their insistence on the map as an artifact that 'con structed' the state, that helped to bring the state into being. It is almost as if it were the map that conjured the state 'as such' into existence, out of the disjointed rabble of the American colonies, out of the far flung possessions of Chinese emperors, out of the territories of the recently warring *daimyo* of Japan, out of the disparate peoples of tsarist Russia. One researcher has termed this map made construct the 'geo body' and has characterized the emergence of Thailand's geo body as "a victory of mapping." The geo body is produced by mapping in three related but distinct ways. In the first place, the very act of mapping requires that the state be something mappable, that is, a geo body, a 'thing', which is to say, with edges, with borders. State borders are brought into being through mapping, both by the imperative 'to be mapped' and 'through' the medium of mapping. Second, these borders establish a shape, the shape of the nation, the nation's visual form; and this mapped shape rapidly becomes iconic, totemic, "the map as logo" as one stu dent of nationalism has put it. For example, in the young United States "the national map permeated American material culture": maps of the new nation were prom inently featured in portraits (where the maps stressed the sitters' identities as Americans); they decorated the walls of American homes and schools; they were integrated into textbooks and didactic puzzles; and they were dis played in public offices, coffee houses, and taverns. Third, the map through its presentation of the state as 'an ex istent thing' obscures the origins of the state in history, in effect assuming, and so projecting, the prior existence of the geo body. This was especially useful for colonial regimes which claimed to 'inherit' ancient geo bodies which, as one theorist of nationalism has put it, the co lonial regimes then constructed by drawing "historical maps designed to demonstrate, in the new cartographic discourse, the antiquity of specific, tightly bounded ter ritorial units" that had, in fact, not previously existed. This, in turn, promotes rhetoric about the inviolability, and so the necessity of defending, borders, which returns us to our first consideration. It was these interlocking benefits – that created the geo body and gave form to the state – that convinced leaders of early modern states of the merits of mapping, and that constituted its necessary precondition.

Large scale, local mapping may seem far removed from these considerations, but, in fact, the mapping of the state, the mapping of the larger world within which it situates itself, and the mapping of the local worlds it nurtures, are reciprocally linked. For example, in Japan Hideyoshi conceived of mapmaking as a localized and incremental program which, while doubtless an expression of control, was also and more importantly, through the collaborative, ongoing labor itself, an in strument of conversion: "Precisely because union was fractious and unfamiliar, cartography served the con querors by instilling a fugitive idea of cohesion, not by reflecting any palpable reality … In this way Hideyoshi and his successors not only normalized a nascent polity but invented, and instructed countless participants in the very imagining of 'our country'." Similarly, in Russia, the unabashedly local maps made during litigation over property "represent the authority of the central state in the provinces. They exhibit the skill of the central state apparatus at extending its influence and bringing its routinized practices and language to the local arena. The interests of center and periphery intersect in the use of the maps." At the same time, the state emerges stronger against the image of other states in a world context. So Jesuit maps, for instance, through their depictions of alien worlds, provoked a heightened consciousness of 'our country' in the case of Japan: "A 'Japan' assumed its strong cartographic profile as attention to the globe and lands that were 'not Japan' reoriented the geographical imagination." Recent scholarship in Russian and Chinese mapmaking has stressed the importance to each of its awareness of the other. Russian envoys to the Qing court were making maps of China as early as 1682. Later the Kangxi Emperor made a gift of the atlas he had commissioned of China to Peter the Great, to impress the Tsar both with the state of the art science the Qing Emperor patronized and of Qing claims to territory. Examples such as these can be multiplied almost endlessly as the number of states proliferated during the nineteenth century.

What cannot be overlooked in all of this is what it was about the map that endowed it with its ability to embody this novel entity, the state. It certainly could not be, as the canonical definition would have us believe, the map's ability to "represent a part of the earth's surface," since scholars are unanimous about the fact that maps 'constructed' the state, brought it into being, forcing into existence previously nonexistent borders (especially well documented for the United States, Russia, and Thailand), giving shape to the shapeless (as in the case of China), unifying the disparate (as we have seen for Japan, Russia, and the United States), furnishing form for what was in essence no more than a dream (the dream of every early modern state). But thinking about the map as a representation was from the beginning no more than a mask, a cloak, a way of making this 'creative' aspect of the map … disappear. From the beginning, it was essential that the state appear to be a fact of nature, something real, enduring – to, at all costs, obscure its recent origins in violence, its tenuous hold on tomorrow – and the map granted this: "We no more than show what exists," said the map (even today it says this about situations in dispute, as between Pakistan and India, Israel and Palestine, India and China). What the map thereby avoided saying was, "*Exists*, yes, but only on this map which, in fact, *affirms* its existence."

This is what the map actually does, affirms the existence of everything on it. "This is here," the map says, "and that is there," affirming as it does the precedent existence of the thing in question 'together' with its location; and so at the same time saying, "Here is this" and "Over there is that," in the process making 'powerful' existence claims. In effect, the map is actually a system of propositions (a proposition is a statement affirming or denying the existence of something), an argument about existence; and if it started with paddy fields and long fields and manor lands and with the states these made up and the world these states composed, or wanted to imagine, wanted 'everyone' to imagine they composed, the map has gone on to a long career rich in the affirmation of the existence of a bewildering variety of things, some whose existence we continue to affirm (e.g., all the nation states we have mentioned), some we have come to deny (the island continent of California, the Northwest Passage, the open polar sea, etc.), but, in any case, things very hard to imagine without the creative intercession of the map (geologic strata, frontal weather systems, the hole in the ozone, etc.).

When we said earlier that most English speakers use 'map' in a straightforward way to describe an artifact that selectively links places in the world ('theres') to other kinds of things ('thises'), we deliberately failed to draw attention to the propositional character of these links; but it was precisely this propositional character in a vehicle that likewise failed to draw attention to it (that cloaked the propositions in facticity) that, as we have seen, made maps useful to the early modern state which, for precisely this reason, heavily promoted their use. Propositions supported by evidence and argument, or even merely sufficiently often repeated, soon enough solidify into facts. That the world is a sphere seems to lack the provisional character we expect in our propositions, and though it remains eminently a proposition, it feels like something else. So today do continents (but is it Europe and Asia or … Eurasia?), nation states (despite the fact that they get hammered together and fall apart with unnerving frequency: think Czechoslovakia, 1918–92; the United Arab Republic 1958–61; or the Soviet Union, 1922–91), and coastlines (though they all move around). Some maps are a little more open about their propositional character: the maps of proposed legislative districts, of proposed subdivisions, of next year's school attendance zones, of which states are going red and which blue, even that map you just downloaded of the route to that convention you are going to … Is it accurate? Do all those roads exist? Can you really make a right turn there?

Like all maps, it assures you that you can, but what does it say down there in the small print in the corner? "You may find that construction projects, traffic, or other events may cause road conditions to differ from the map results"? Hmmmm … When we said earlier that the "maps selectively link places in the world ('theres') to other kinds of things ('thises')," we added "for the purpose of underwriting the reproduction (or contestation) of the social relations of power," and this capacity of maps for ignoring construals of reality alternative to those they propose – and the facticity they thereby manage to project – is what most substantively underwrites the reproduction of the social relations of power. Effective maps can absorb change without changing. For the continuously evolving early modern state this was critical. New data could be absorbed into apparently timeless frames, and stripped of disturbing novelty. A Japanese scholar has noted that the issue "is succinctly conveyed in the phrase 'newly revised,' which became a commonplace in the titles of the information library [of seventeenth century Japan]. The words insisted that something new in a text was new enough to merit special attention, though not quite new enough to merit a fresh beginning. Something fundamental survived – something suscep tible to revision rather than reimagination," and she has gone on to generalize about the way "expectation re mains the most powerful preservative of models. Map makers and map users learn to expect the kind of maps they are accustomed to seeing. In the end, then, the strength of models is the facility to frustrate, as either unthinkable or perverse, the revision of their underlying conceptions. An alternative representation of Edo [Tokyo] would have required not so much new evidence as a new vision. Had commercial mapmakers accorded privilege to commercial wards rather than martial man sions, they would have projected a rival plot: this is a financial and mercantile capital (say), administered through the neighborhood associations of townspeople, where entertainment is a major enterprise. For that leap, they needed no fresh data. They needed a radical phil osophy." We still do. All of us.

See also: Cartography, History of; Map Types; Mapping, Philosophy; Maps and the State.

Further Reading

Anderson, B. (1991). *Imagined Communities: Reflections on the Origin and Spread of Nationalism* (revised edn.). London: Verso.

Berry, M. E. (2006). *Japan in Print: Information and Nation in the Early Modern Period*. Berkeley, CA: University of California Press.

Bruckner, M. (2006). *The Geographic Revolution in Early America: Maps, Literacy, and National Identity.* Chapel Hill, NC: University of North Carolina Press.

Buisseret, D. (1992). *Monarchs, Ministers and Maps: The Emergence of Cartography as a Tool of Government in Early Modern Europe*. Chicago, IL: University of Chicago Press.

Edney, M. (1993). Cartography without progress: Reinterpreting the nature and historical development of mapmaking. *Cartographica* 30(2/3), 54 68.

Harley, B. and Woodward, D. (eds.) (1987 present). *The History of Cartography*, 4 vols. Chicago, IN: University of Chicago Press.

Hostetler, L. (2001). *Qing Colonial Enterprise: Ethnography and Cartography in Early Modern China*. Chicago, IL: University of Chicago Press.

Kivelson, V. (2006). *Cartographies of Tsardom: The Land and Its Meaning in Seventeenth Century Russia*. Ithaca, NY: Cornell University Press.

Mundy, B. (1996). *The Mapping of New Spain: Indigenous Cartography and the Maps of the Relaciones Geográficas*. Chicago, IL: University of Chicago Press.

Seed, P. (1995). *Ceremonies of Possession in Europe's Conquest of the New World, 1492 1640*. Cambridge: Cambridge University Press.

Turnbull, D. and Watson, H. (1989). *Maps Are Territories: Science Is an Atlas : A Portfolio of Exhibits*. Geelong, Victoria: Deakin University.

Winichakul, T. (1994). *Siam Mapped: A History of the Geo Body of a Nation*. Honolulu: University of Hawaii Press.

Wood, D. and Fels, J. (2007). *The Natures of Maps: Cartographic Constructions of the Natural World*. Chicago, IL: University of Chicago Press.

Maps and Governance

F. Harvey, University of Minnesota, Minneapolis, MN, USA

Traditional View: Using Maps to Govern

In the traditional view, maps are documents used to govern (e.g., the book *Maps for America*). Maps are factual and their accuracy is increasing with technological improvements. Governments have always made use of maps, even if some doubts were associated with a map because of utility in governing. Without traveling, a map makes it possible for a trained reader to make sense of an area and use that in formation to help make decisions. This capability is in trinsic to governance and the visual acumen of the map is crucial in complementing or even replacing voluminous paper documents. Millenia ago the Chinese government relied on maps; now, it is fair to say all governments use maps in almost every facet of their activities. A local government uses maps, the CIA uses maps, the UN uses maps. Maps are ubiquitous in government. As J. B. Harley wrote, a map is "…graphic representation that facilitates a spatial understanding of things, concepts, conditions, pro cesses or events in the human world." Taken from a functional viewpoint, maps are crucial documents for collecting, analyzing, and communicating information.

Maps are also central in creating iconic images that governments rely on to communicate. This is a key reason why maps are often one of the main graphical ornaments on a local council's chambers, in the mission control center for a satellite mission, and the logos of international organizations. The Western understanding of the role of maps in governance focuses on their use for communicating information. Of course, a map reflects the considerations and knowledge of its creators and points to social understanding of the time. As noted by Norman Thrower, "viewed in its development through time, the map is a sensitive indicator of the changing thought of man, and few of his works seem to be such an excellent mirror of culture and civilization" (Thrower, 1996).

Spatial data infrastructures (SDIs) are now under development around the world (INSPIRE and the NSDI in the US are often cited examples). Maps play a central role in motivating governments to support and fund these costly infrastructures as they enable the production of maps for myriad uses in government and attempt to maximize efficiencies and hold down costs through pol icies that emphasize reuse of existing data.

Critical Perspectives: Maps Act in Use

In most recent analysis of governance and mapping, the power of mapping is the central concern. Maps are not merely documents, neutral in nature, but subject to dif ferent interpretations. Maps act in various ways. Denis Wood's writing on the power of maps offers rich insights into how the merging of the rich and endless complexity of the world into a single piece of paper or display creates a powerful image or icon. Still, a map remains selective and greatly limited contrasted to our experiences and engagements with the world, but it remains a key means for understanding and analyzing the world. This power is very attractive and lucrative; its misuse and abuse are issues that continue to engage Wood and others.

Michel Foucault was interested in governance and the spatiality of power and offers some good starting points for reconsidering the map. The analysis of governance and mapping relies specifically on bio power, power/knowledge, and governmentality. Bio power compro mises various techniques into technology that enables the State to control entire populations and undergirds the nation state and capitalism. Instead of earlier techniques of control that emphasizes physical punishment in re sponse to transgressions, bio power focuses on techniques that emphasize the protection of life, the regulation of the body, and the production of other technologies of power. Foucault's famous research was of sexuality, yet arguably mapping becomes a key technology of bio power for aligning actions with the State's interests. Power/knowledge is Foucault's concept to integrate power and knowledge. Power, for Foucault, is based on knowledge and shapes the reproduction of knowledge along the lines of its intentions. Power/knowledge, in the scheme of the modern society that emphasizes repression over oppression, is never a centralized good, but is always decentralized, relativistic, and unstable in situated rela tionships. Governmentality is related to both the con cepts of bio power and power/knowledge and takes various forms according to historical periods and specific power configurations. In neoliberal States, govern mentality production of knowledge assists in the creation of self regulating selves whose actions are aligned with dominant forms of bio power. Governmentality com prises institutions, programs, and productions of the population and the exercise of power upon the human body(ies). Indeed, as Foucault discusses, power is a re lationship of force. Each map serves as a nexus for aligning and configuring various bio power and power/knowledge among de centered and conflict laden mem bers of societies.

Two caveats are important to record to this brief at tempt at a synopsis of some of Foucault's thinking. First, bio power, power/knowledge, and governmentality are never singular and monolithic hegemonies. Each faces contradictions among members of societies and is dynamic and ever changing. Second, technology should be understood broadly as an extension of human capabilities by other means and is by no means restricted to material artifacts that affect material change, but also includes organizational technologies.

Following Foucault and later Anglo American scholarship to engagements with the question of how power is made, the opening question of this article must be inverted and pursued in terms of the relationship of mapping to governance practices: if without a map there is no modern government, the analysis of governance of mapping is called on to examine how mapping prefigures and maintains the authority and power of governance in the face of social, economic, and political change and conflict. This article takes the argument that a key element of successful strategies that maintain sovereign power involves the delegation of authority to nonhumans, mainly information technologies, as well as technological and regulative organizations of coordinated administrative activity, or bureaucracy. Humans are key to maintaining flexible arrangements and responding to unexpected problems and changes as well as assuring a 'human face'.

Foucault in the early 1970s was primarily interested in the study of how power manifests itself and produces discourses of truth. His studies of the 'how of power' involved, on the one hand, the creation of rules of right that delineate power and, on the other hand, the powerful effects that discourses of truth have. To engage power as something that circulates and functions only when it is part of relationships, Foucault stresses the importance of ascending analyses of power in its individual trajectories with a focus on agents of power and transformations of power. Studying the 'how of power' involves thorough examinations of the mechanisms that circulate, form, and organize knowledge. The analysis leads to insights into local forms of subjugation and apparatuses of knowledge.

Matthew Hannah's study of the development of governmentality in the nineteenth century exemplifies this type of analysis in geography. This work studies modern state formation on a national scale occurring at both the localized and state levels with two goals: first, to show that Foucault's ideas about governmentality help to understand the historical geography of modern America; second, to illustrate how spatial the activities of governmentality at a national scale are. Hannah's study centers on a key figure of the late nineteenth century, Francis Amasa Walker, who, among other pursuits, was superintendent of the 1870 and 1880 censuses and president of MIT. Analytically, Hannah is thoughtfully aware that any focus on Walker is only to engage deeper structural problematics for engaging the larger network of relationships that involve mapping, surveying, and censuses in the establishment of territorial mastery. Engaging the questions how society pursues the establishment of control over space and the importance of statistics demonstrates mechanisms of power, he offers considerable insight into the roles of mapping for governance.

Drawing as well on Foucault's concepts, similar issues are developed by Crampton in his study on the mapping of cyberspace. Crampton shows in this book how the use of thematic mapping has played a central role in enabling the state to shift the focus of governance from people to populations, by tracing their development and refinement since the late eighteenth century for political means. This book specifically offers an articulation of Foucault's governmentality concept as the intersection between disciplinary technologies that regulate the masses, and the micro technologies that the individual may use (or be used upon the individual) to shape and reshape the self. Consulting a map to establish the reference for an activity, in this sense, amounts to the tangible exercise of Foucault's governmentality. This is true for addressing the minutiae of daily life in bureaucracies and esoteric questions that seem to idly spring out of unfettered curiosity. In both cases, maps become key agents for configuring individual knowledge and action.

The mundane day to day drudgery of mapping to maintain governmentality's mastery of territory is most evident in the cadastre, which records land ownership, land rights and obligations, as well as their boundaries. A cadastral map, following Kain and Baigent, is an instrument of control that consolidates state power and obliterates earlier relationships to land ownership and access. Echoing Foucault's thinking about the perpetual nature of conflict in society, they also highlight that the use of cadastral mapping is characterized by struggles as turbulent as struggles for political power. Katherine Verdery's study of land restitution in Romania following the demise of Ceausescu's communist regime in 1991 demonstrates that the conflicts surrounding the cadastre lack a clear binary, but involve the maintenance of flexible arrangements to respond to conflicting claims to land ownership.

Is to Govern to Map?

By reversing the question 'Is to govern to map?,' which suggests maps are an essential part, yet still subordinate to the activities of government, into 'Is to map is to govern' the problem becomes a matter of understanding the various roles maps take in the configurations of government power and authority in practice. From this vantage point, government mapping produces the geographic relationships and configures a network of humans and nonhumans. As David Turnbull has written, the map precedes the territory and that before the map the territory was not governed. In other words, while the map appears to be the graphical representation of a political sovereign's claim to that territory and the governments organization of the territory, in the manifold relationships of power it serves as the means of distributing administrative capacities and political influences. Each map becomes the nexus of a particular configuration of power, a material trace of relationships.

We now return to the concept of centers of calculation developed by Bruno Latour in *Science in Action*, published in 1987, a Foucault influenced approach to conceiving the flows of power and knowledge in networks. Accumulation of knowledge is a key part of a center of calculation. Scientists, who are exploring, discovering, and creating, prepare representations that show the new knowledge in relationship to existing knowledge. Other scientists add to the accumulation. One of Latour's favorite examples is the scientist explorer who heads out to a distant land, makes notes and maps, possibly even engaging with natives, and returns to the empire with a map in hand that consolidates different forms of knowledge acquired in the remote territory. Other scientist explorers will bring back additional knowledge thereby enlarging the accumulation. Latour describes accumulation as a cycle involving the drawing of maps. They figure significantly as technical inscription devices in the practices of representation and production. Control over the inscription devices enhances the power of individuals and institutions in the center of calculation. It allows for decisions that determine the validity of new and old knowledge and the promotion or dismissal of individuals. With control centers of calculation are increasingly capable of facilitating the network of scientists and bureaucrats who support the apparatus of power. Latour is interested here primarily in schematized cycles of accumulation as the action of science, but his examination of Laperouse's cartography of Asia in the hands of French government officials offers insight into the role of mapping for government. The maps Laperouse returns to Paris make it possible for the government to act at a distance.

The ability to act at a distance is enhanced with maps and is a key reason for them to be significant for government. This provides the motive for governments to develop cartographic resources in the service of the state. Development and control over accumulation is a key aspect of power and for enhancing governmental significance of maps. Foucault, in his work, is concerned with an analysis that considers how power develops in the transition from feudal to modern forms of state and society. Whereas in a feudal state the sovereign relied on repression (Hegel, Freud, Reich) to retain and maintain power, in the modern state it relies on oppression. Oppression means the control of institutions that produce individuals who pursue activities that retain and maintain the state's power. Sheer and total reliance on maps to learn and know areas we are unable to experience, or areas that may be even impossible to experience, underscores the centrality of maps in providing knowledge that is central to the maintenance of government control and power.

For this reason, mapping figures centrally in attempts, by individuals with governmental power, to control a country from the center and manage activities in the country and engagements with other countries. For Foucault, governmental power takes on the specific form of governmentality, the collection of institutions, analyses, and studies that result in formation of specific governmental apparatuses that undergirds the development and maintenance of governmental power. In Foucault's historical analysis, governmentality has led to the development of the population as a marker in reference to government, the object of government intervention, and the object of governmental techniques.

Alan Scott and Timothy Mitchell have both examined the role of mapping technologies in the practices of governmentality. Mapping in government allows actors the possibilities for pursuing their decisions while constraining these decisions to only the rational choices represented by maps and other codified inscription devices. For Scott, even the mundane cadastral map is more than a set of maps for calculating taxation, "… when [maps are] allied with state power, [they] would enable much of the reality they depicted to be remade … a state cadastral map … does not merely describe a system of land tenure; it creates such a system through its ability to give its categories the force of law" (Scott 1998). Governance and mapping are always a hand and its glove acting together to further the state's interests and power.

Mapping involves both people and things. Latour's Center's of Calculation concept puts a spotlight on the key role that nonhumans have for governance. As Latour and others have argued, nonhumans are considerably more resilient and adhere to orders, regulations, instructions, and programs with far less variability and arbitrariness. They are far less flexible than humans and also lack abilities to respond to key characteristics of human social and political interactions. As for industry, where robotic devices have replaced or supplemented humans in most forms of production, the use of nonhuman technologies and bureaucratic organization has become ubiquitous. Particularly, information technologies, GIS for mapping, have become commonplace fixtures in administrations. Nigel Thrift has pointed to some of these issues.

For Latour, 'centers of calculation' are sites that bring together multiple relationships of an actor network. All sites are local and connected. Actors bring ingredients together and networks remain from traces of actors' ac tivities. Any network involves relationships among humans and relationships among nonhumans. Nonhu mans, such as maps, or immutable mobiles in Latour's terminology, are especially important in aligning inter ests, stabilizing existing networks, and extending net works to include new actors and sites. Sites and networks are understood as elements of the connections of power.

Summary

This article has presented an argument resting on the theoretical work of Foucault and others that mapping lies at the core of governance. Maps are central to governance and can be combined, collated, and compared in myriad ways that alter the distribution of power in a government and state. Recently, John Pickles has also offered an en gagement with this most basic of geographical adminis trations of governance with this considerations of the governmentality of the cadastre. His analysis points in conjunction with Turnbull's work that mapping is an activity of creating geographic knowledge, abilities for action, possibilities for coordination, and the means of iconic identity production.

It lies in the interest of the sovereign to maintain and create further nonhumans and minimize the roles of humans. Ethically and morally an engagement with the important role of maps for government must consider the use of cartographic surveillance to monitor individuals and areas. Government relies on mapping – without a critical engagement with these government practices we may find rights and abilities gravely constrained by mapping.

The roles of mapping in the practices of governance should be considered as configurations of power, authority that involves nonhumans and humans. Starting from these elements, Foucault and other theorists of geography and politics challenge geographers and others to think not of a single map as an artifact of governance, but instead think of each and every map as a nexus that helps configure relationships of power and authority. This is a viewpoint that thinkers from diverse disciplines have advanced in the last 30 years. For example, Bruno Latour, in a number of his works, presents the concept of centers of calculation, the means of accelerating the accumulation of knowledge, which make it possible for a sovereign or government in Paris to jointly bring together various information sources, review situations and make decisions about the planning and allocation of land in parts of Tahiti. This is certainly one of many theoretical plinths influenced by Foucault's work for engaging the polymorphous roles of mapping in the practices of governance.

In spite of all attempts to establish constantly up to date mapping in administration with the goal of stabilizing sovereign power, exigencies, problems, and errors lead to discrepancies in mapping activities. As Foucault writes power is always accompanied by resist ance. The places of resistance to mapping are manifold, both outside of government, and perhaps even more strongly within government. Resistance outside of gov ernment is perhaps most often of a passive nature: the never ending changes to the natural and built up en vironments mean that even the most constant gaze of mapping produces knowledge that no longer corres ponds to the environment when it called on in a center of calculation. Rivers meander and people tear down buildings and rebuild. Change is constant; the only question is whether the change is significant for the questions being raised. Resistance in government comes from a variety of sources, both active and passive. Most active resistance comes from bureaucratic in fighting and struggles over access to resources. Maps are a source of power and therefore will be a part of many govern mental conflicts over the delegation of responsibilities for mapping. Passive resistance comes more often from the trials necessary to negotiate the bureaucratic ap paratuses put into place to help assure the distribution of power. Discrepancies in mapping and the persistent question of accuracy come in two major forms. Errors in administrative processing and manipulations of infor mation in the system to produce a paper world that is distinct from the lived world of the people involved. The political dimensions of human social activity and the limits of governmentality constitute an important aspect of all types of resistance. No matter what type of pol itical organization a society follows, the distribution of resources, the means of controlling centers of calcu lation, and the ability to train future specialists involve political questions.

Whatever forms the resistance takes, humans are key to their resolution. Disciplines and governmentality are key to the continued vitality of the uses of mapping in governments. The training of specialists to work in gov ernment and the education of the population to assure its activities support the state help the government deal with conflicts and steadily work on the improvement of mapping technologies.

See also: Colonialism I; Colonialism II; Governance; Integrated Spatial Data Infrastructure.

Further Reading

Crampton, J. (2003). *The Political of Mapping of Cyberspace*. Chicago, IL: The University of Chicago Press.

Crampton, J. W. and Elden, S. (2007). *Space, Knowledge and Power: Foucault and Geography*. Aldershot, England: Hampshire; Burlington, VT: Ashgate.

Foucault, M. (1991). Governmentality. In Burchell, G., Gordon, C. & Miller, P. (eds.) *The Foucault Effect: Studies in Governmentality*, pp 76 86. Chicago, IL: University of Chicago Press.

Foucault, M. (1997). *Society Must Be Defended* (Macey, D., trans.). New York: Picador.
Hannah, M. (2000). *Governmentality and the Mastery of Territory in Nineteenth Century America*. Cambridge: Cambridge University Press.
Harley, J. B. (1989). Deconstructing the map. *Cartographica* 26, 1 29.
Harley, J. B. (1991). The New History of Cartography. UNESCO. http://findarticles.com/p/articles/mi m1310/is 1991 June/ai 10940835 (accessed April 2008).
Kain, R. J. P. and Baigent, E. (1992). *The Cadastral Map in the Service of the State: A History of Property Mapping*. Chicago: University of Chicago Press.
Mitchell, T. (2002). *Rule of Experts. Egypt, Techno Politics, Modernity*. Berkeley, CA: University of California Press.
Pickles, J. (2004). *A History of Spaces. Cartographic Reason, Mapping, and the Geo Coded World*. New York: Routledge.
Rose, N. (1999). *Powers of Freedom*. Cambridge: Cambridge University Press.
Scott, J. C. (1998). *Seeing Like a State*. New Haven, CT: Yale University Press.
Thrower, N. J. W. (1996). *Maps and Civilization: Cartography in Culture and Society*. Chicago, IL: University of Chicago Press.
Turnbull, D. (1989). *Maps Are Territories. Science Is an Atlas*. Chicago, IL: University of Chicago Press.
Turnbull, D. (1998). Mapping encounters and (en)countering maps: A critical examination of cartographic resistance. *Knowledge and Society*, vol. 11, pp 15 43. London: JAI Press.
Verdery, K. (2003). *The Vanishing Hectare: Property and Value in Postsocialist Transylvania*. Ithaca, NY: Cornell University Press.

Relevant Websites

http://monarch.gsu.edu
The Politics of Mapping Jeremy W. Crampton, Georgia State University.

Maps and Protest

D. Wood, Raleigh, NC, USA
J. Krygier, Ohio Wesleyan University, Delaware, OH, USA

Glossary

Activism Aggressive action taken for political or social purposes.
Dissent Disagreement with prevailing ideas or forces; lack of conformity.
Hegemony Domination of one group or person or idea over others.
Protest Similar to activism; specific actions taken to counterprevailing political or social forces.
Satire Sarcasm, irony, or ridicule used to undermine prevailing political or social forces.

Protest maps come in three registers, that of the office, that of the streets, and that of the press.

In the official register are maps made to establish, advance, or illustrate official protests. These protests are often of other maps. In Hillsborough County, Florida, for example, a protest map is attached to a protest, which is a form on which objections can be raised to flood risk designations proposed by recently resurveyed flood maps. Residents of Austin, Texas, are being encouraged to file similar protest maps, as indeed is anyone who feels ill served by the National Flood Insurance Program maps being updated as part of its Map Modernization Program by the Federal Insurance and Mitigation Administration (FIMA). FIMA has even prepared *A Guide for Community Officials: Appeals and Protests to National Flood Insurance Program Maps* that describes the form that protest maps must take. In Stillwater, Oklahoma, on the other hand, protest maps are prepared by city staff to help planning commissioners prepare for public hearings on proposed street improvement districts. These protest maps are of properties whose owners are protesting inclusion in proposed improvement districts. Such usages of 'protest map' are widespread.

In the register of the streets are maps 'of' or 'to' protests. That is, these are maps that let you know how to get to protests. As a genre they seem to have come into their own during the Republican National Convention of 2004 when seemingly every newspaper, as well as every blog, carried maps not only of the convention sites, but of the protest sites as well. These rapidly became known as protest maps, so that invitations such as this have become common: "If your group wants to be represented at the event (table, leaflets, protest maps, etc.), please get in touch with us." The etymology is apparent in: "I would like to invite you to attend our protest. Maps of campus are available and protesters are asked to arrive at the mall by 12:45 because, while Bush is inside, the event media will have nothing to cover except the protesters outside," as also in, "Hi everybody. There is a protest on the Balnagown estate of Mohamed Al Fayed on Sunday 27th April at 12 noon. All are invited for a peaceful protest. Maps provided. We will cause no damage and leave only footprints."

A protest in essence is a solemn declaration of opinion, usually of dissent, and this sense is nicely focused by the official protest maps with their need to be "certified by a registered professional engineer or a licensed land sur veyor" (unless derived from 'authoritative sources' such as the Bureau of Land Reclamation or a State department of highways and transportation). The process is formal, carefully framed, and the maps that resolve the protests have the force of law. These maps are members of a whole class of documents in an enormous system of dispute resolution that runs from the complaint counter of your neighborhood big box retailer up through the appellate courts. But protests may also be registered in such dra matic, typically collective forms as strikes, boycotts, rallies, and marches, may even involve violence, and these are the sorts of protests the maps on the table with the leaflets etc., are directing people to. What we are referring to as the register of the press includes maps that, like official protest maps, are actual protests (not merely of or to them), but that at the same time are distinctly unofficial (often anti official) and partake of the noisy, public, self consciously rhetorical character of street protests, more oriented toward ferment than resolution.

Doubtless there are earlier protest maps in this register, but perhaps the most famous is "The Gerry mander: a new species of monster, which appeared in Essex South District in Jan. 1812" (**Figure 1**). With a few strokes of the pen the map transformed a recently con figured Massachusetts electoral district into a kind of winged salamander, with a name that combined that of the lizard with that of Massachusetts Governor Elbridge Gerry. Engraved by Elkanah Tisdale for the 26 March 1812 issue of the *Boston Gazette*, the map was widely re printed by Federalist sympathizers (broadsides appeared immediately) protesting the redistricting scheme that gave Gerry's Republicans, if not Gerry himself, a decisive advantage in the upcoming state elections.

Tisdale's map is sometimes thought about as a metaphorical or satirical map, but then satire – trenchant wit, irony, or sarcasm used to expose and discredit vice or

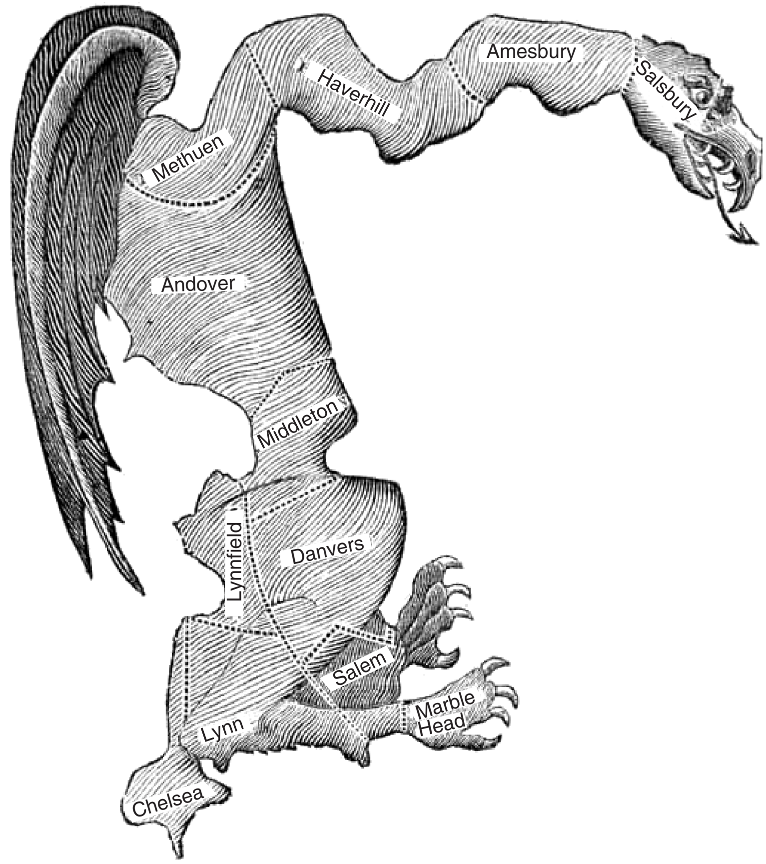

Figure 1 Elkanah Tisdale's original Gerry-mander, as it appeared in a Broadside published in Salem Massachusetts in 1812. From Parton, J. (1877). *Caricature and Other Comic Art*, p 316. New York: Harper and Brothers.

folly – is a frequent companion of protest. Heavier on the sarcasm but playing in a related key is McArthur's 'Universal corrective map of the world'. Stuart McArthur, an Australian, was 12 years old when a teacher told him it was wrong to orient a world map he had drawn south up. He was 15 when, an exchange student in Japan, he was ridiculed by his fellow American exchange students "for coming from the bottom of the world." He was 21 when he published the map that fulfilled the vow he had taken in Japan to set things straight, protesting, with his map, not only the usual north up orientation but people's prejudice against the south: "Never again," a text on the map declares, "to suffer the perpetual onslaught of 'downunder' jokes – implications from Northern nations that the height of a country's prestige is determined by its equivalent spatial location on a conventional map of the world."

In a world dominated by maps oriented north up, it may happen that 'any' map with south on top comes to be taken as a kind of protest against the hegemonic point of view, but protest usually involves an awareness of not only what it is in favor of (south up), but of what it is opposed to (north up). In fact, protest is 'often' clearest about what it is against (when 'protest' is used affirmatively, as in a phrase like 'protested his innocence', it is always in the face of denial or doubt). What makes McArthur's 'Universal corrective map of the world' a protest map is the 'corrective' in the title which in escapably brings to mind the view being corrected, just as the 'Gerry mander' brings to mind the shape of electoral districts less easily transformed into lizards. It is rarely in protest maps in this register that facts are being con tested, as they always are in the official register. Instead it is the way the facts are framed. 'Gerry mander' does not question the shape of the new electoral district. It insists that we look at it askance. The corrective map does not propose a new world. It asks that we look at the world anew.

Here is another example. In a discussion paper of 1971, the Detroit Geographical Expedition published a map it had compiled of the 'Citywide pattern of chil dren's pedestrian deaths and injuries by automobiles'. The map displayed the deaths and injuries as dots on a background of Detroit streets. It also indicated the lo cation of the city's black population with a meandering dashed line. It is not hard to see that most of the kids killed by cars lived in black neighborhoods, but this is a conclusion someone looking at the map has to draw. A

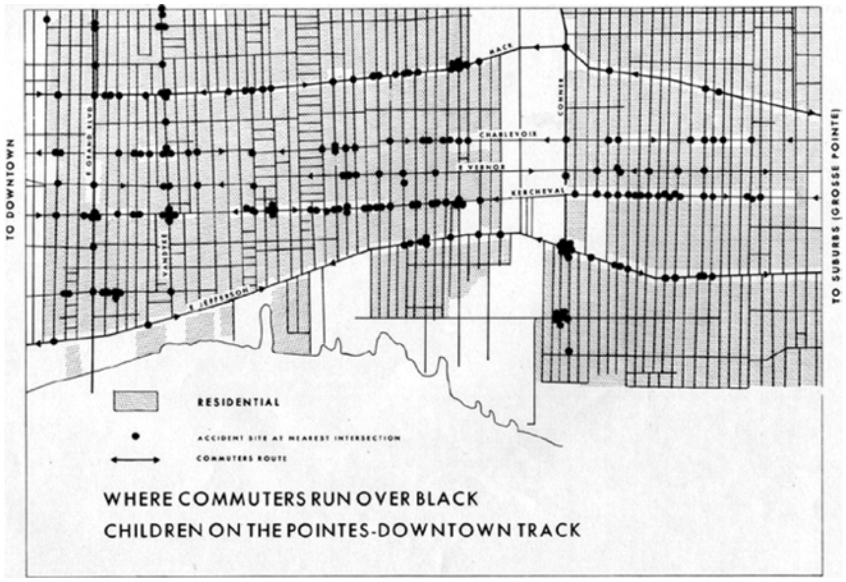

Figure 2 Where commuters run over black children on the Pointes-Downtown track. From The Detroit Geographical Expedition and Institute, Field Notes: Discussion Paper No. 3: The Geography of the Children of Detroit, Detroit, Michigan, 1971.

couple of pages farther on the expedition zoomed in to give us a map that drew the conclusion for us: "Where commuters run over black children on the Pointes Downtown Track" (**Figure 2**). There is no mistaking the protest here. The map no longer displays a 'pattern' but locates crime scenes, and the deaths are no longer caused by automobiles. Indeed, the automobiles have vanished to be replaced by their drivers who are specifically characterized as commuters. Any Detroiter would have known that these commuters were white and on their way between work downtown and home in the exclusive Pointes communities to the east. That is, this is a map of where white people, as they rush to and from work, run over black children. That is, it is a map of where white adults kill black kids. It is a map of racist infanticide, a racial child murder map.

Again, 'Where commuters run over black children on the Pointes Downtown track' proposed no data that had not been on the less inflammatory 'Citywide pattern of children's pedestrian deaths and injuries by automobiles'. It did, however, ask that we think about the data differently. Gwendolyn Warren wrote about this difference in an article that accompanied the maps: "The way the city is situated, there is the central place downtown and then there are rings which go outside of that and the big ring right outside downtown Detroit is the Black community. All the area about a mile going out from downtown Detroit is one way traffic and runs right through the heart of the Black community. And on one specific corner in six months there were six children killed by commuter traffic. But, naturally, these deaths of the children or the injuries or whatever it happened to be were disguised as something else. They never said that a certain business man who was working for Burroughs downtown who was on his way to Southfield went through the Black community by way of this commuter traffic and killed my people – Black children. Even in the information which the police keep, we couldn't get that information. We had to use political people in order to use them as a means of getting information from the police department in order to find out exactly what time, where, and how, and who killed that child. The fact that it actually establishes a pattern proves it is not 'accidental'."

The protest, the anger, is on the surface here. It radiates from the map. A similar anger radiates from the maps in *The Nuclear War Atlas* published in 1982 by the Society for Human Exploration as a two sided poster, text on one side, 28 maps on the other. It folded, William Bunge, its principal author recalls, "into a 5 in. × 8 in. size designed for peace demonstrations where it was abundantly sold." Black, white, and red, the very design and layout of the maps were inflammatory, but the reframing of data that had been culled from a variety of impeccably reserved sources (e.g., *Progress in Nuclear Energy*, *Health Physics*, *Child Psychology*) was largely carried out by the titling: 'The march of doom', 'Patriotic poisoning', 'The sea of cancer'. 'The sea of cancer' was a map of the US. largely covered with red stippling that indicated areas that would be exposed to 100 or more rem of radiation in a full nuclear war. "Not only will most of the United States be washed in immediate radiation," read the caption, "but even the white areas on the map are safe only in the sense that people in the open escape short term damage but not long term. The cancer is everywhere." A map of the world makes the point that the boundaries of a missile armed United States and Soviet Union are global: "To state the new geographic reality using the militaristic language of the 1980s, 'The Russians are not coming. They are already here.' At least they are straight up in the sky above us and thus are bounded by the earth's surface, not 'contained' by boundary lines. They can kill anything on the earth's surface and for a con siderable depth below it; the Americans likewise. 'Con tainment' has been a mathematically proven bankruptcy for almost twenty years." By dramatically reframing simple truisms like these, *The Nuclear War Atlas* gave people whole new grounds for protesting nuclear weapons.

If the poster edition of *The Nuclear War Atlas* had been self published and largely distributed by its author, Michael Kidron and Ronald Segal's *The State of the World Atlas* of 1981 was a Pluto Press project distributed in the United States by Simon and Schuster that has gone through eight subsequent editions and spawned an entire family of protest atlases (*The State of War and Peace Atlas*, 1983; *The New State of the World Atlas*, 1984; *The State of Women in the World Atlas*, 1986; *The State of the Earth Atlas*, 1990; *The State of Religion Atlas*, 1993, and so on, each with multiple editions). The initial object of Kidron and Segal's protest was the state: "It is our contention that that the destructive aspects of the state have come cru cially to exceed the constructive ones," they wrote in the atlas' introduction. Their maps documented the way states have proliferated and expanded into the remaining nonstate areas of the planet. Their maps documented the military preoccupation of states and their squandering of resources on war. Their maps documented the unequal distribution of state resources, and the impact of the state on labor, society in general, and the environment.

The maps made no pretense about being neutral, and if the first edition's maps framed their subjects with titles like 'The state invades the sky', 'Arms for the sake of power', 'Bullets and blackboards' (mapping the ratio be tween soldiers and teachers), 'Slumland', 'Fouling the nest', and 'The dying earth', the second edition added incendiary subtitles for any who might have missed the point. For example, below the title, 'Scourges of the state' – a map of prisoners, capital punishment, state assassin ations, and torture – the second edition now appended: "All states are armed against their citizens. Many states use exceptional methods to terrorize them." A quarter of the volume was taken up by notes about Kidron and Segal's data sources. One of these commenced, "The morally repulsive priorities of the state can be illustrated in many ways; but perhaps nowhere more eloquently than in the comparison between expenditure on prep arations to promote injury or death and expenditures to heal and sustain life." It is always possible to disagree with Kidron and Segal, but it is not possible to mistake their point which, after all, was the point of their reframing.

Maps in this register do not have to be literally run through a press, nor of course do they have to be pro testing the state of things in the world. Maps of this type have become ubiquitous on the web where often they are protesting other maps. An already notorious example involves the mapping of the results of the US presidential election of 2004. The night of the election, even before the Kerry/Edwards campaign had conceded, maps were published displaying the states in which a majority had voted for the Republicans in red, and those in which a majority had voted for the Democrats in blue. These maps showed a fringe of blue states in the far west, along the Great Lakes, and in the northeast, and the rest of the country, almost all of it from the Coastal Ranges in the west to the Atlantic in the east, in red. America, these maps proclaimed, was Republican. But as Michael Gastner, Cosma Shalizi, and Mark Newman soon pointed out, the maps were misleading. They failed to recognize that most of the red states had small populations and the blue ones large. Blue states might take up less space, but they were filled with many more people which is, after all, what matters in an election. So Gastner, Shalizi, and Newman made an election results cartogram, that is, a map that displays places sized according to some attri bute other than their spatial extent. Their cartogram made it clearer that, though majorities in Montana, Wyoming, Utah, Colorado, and 27 other states might have voted for Bush, almost as many people in the more heavily populated states of California, New York, Illinois, Pennsylvania, and 15 others had voted for Kerry; and that instead of most of the country voting for Bush, nearly half had voted for Kerry.

USA Today published a map of the election's results by counties instead of states. While its map was less strik ingly divided, it again was overwhelmingly red. This too was then remade as a cartogram that magnified blue counties to reflect their greater populations, and it too made the election seem a far more equal contest. But even after allowing for the population differences, there was still more red than blue on the map which, since the national vote had been almost equally split, should not have been the case. The problem turned out to be that the amount of red was skewed by counties in which only a 'slim' majority had voted Republican. Robert Vanderbei suggested that one way to deal with this would be to use shades of purple to indicate the actual percentage of voters in each county. On Vanderbei's map, there were only a few red and blue counties. Most of the country was one shade or another of purple. The maps of Gastner, Shalizi, Newman, and Vanderbei protested not only the crude 'maps' of simple majorities, but also the 'impres sions' people took away from them, that the country, split between a vast Republic heartland and a Democratic fringe, had handed Bush some sort of mandate. There was in fact no clear regional split at all, and instead of handing Bush a mandate, Americans had reelected him by the slimmest of margins.

Protest maps can concern themselves with lesser issues as well. In 1931, Harry Beck made a sketch in an exercise book of a map of the London Underground. Beck's ambition was to make the system intelligible by reducing its routes to vertical, horizontal, and diagonal lines, by increasing the scale of its route dense center, and by eliminating surface detail except for an equally stylized Thames. Beck's map, many times revised and issued in uncountable forms and numbers, in time be came an icon, not only of the London Underground, but of modern design itself. In 1992 the conceptual artist, Simon Patterson, produced a lithograph called *The Great Bear*. Except for its title, which is a common name for the constellation Ursa Major, the print reproduced the contemporary version of Beck's map, except when you looked closely you realized Patterson had replaced the station names with those of philosophers, actors, polit icians, and others that we sometimes think about as 'stars'. For example, Patterson renamed the stops on the Bakerloo (brown) line after engineers, those on one branch of the Northern (black) line after musicians, and those on the other branch after movie stars. The Fra Angelico station stands where the Saints line crosses the Italian painters line; and Geoff Hurst, on the Footballers line, is only a stop away from William Randolph Hearst on the Louis line. Patterson has worked in an equivalent fashion with paint chips, the periodic table, electric circuit diagrams, slide rules, air traffic route maps, con stellations, and The Last Supper (e.g., *The Last Supper Arranged According to the Flat Back Four Formation* (*Jesus Christ in Goal*)). "I like," Patterson says, "disrupting something people take as read."

The Great Bear soon became an icon in its own right (a copy hangs in the Tate) and it attracted its own imitators: there was the *London Undergrub* (all the stations named after food), the *Undergroans* map (an impolite version), the *Untergrund* map (in German), an 'Anagrams' map (in which anagrams had been made of all the station names), *The Company Sponsored Map* (with the names changed to match, or nearly match, the name of well known com panies), an upside down version, and one on which the underground lines had been flipped over the Thames so that south London now had most of the lines. In 2005 Thomas David Baker produced the *Moviemaker Tube Map*: "I liked *The Great Bear*," Baker writes, "but I didn't like the way when a station was both on the Artist and the Footballer line that the replacement person was just an Artist or a Footballer, but not both. Doing it for movies – using Director, Actor, Cinematographer, etc., for the lines – meant I could make sure that each individual repre senting an interchange had done the job represented by each line that goes through that station." In early 2006 the *Musical Map* appeared in *The Guardian* with each line named for a type of music (soul, reggae, pop, etc.) and each station after an artist of that type (The Four Tops, Peter Tosh, U2, and so on). Artists at intersecting stations had to fall into the mixed genre of the intersecting lines (e.g., Prince is found where the Funk line intersects the Pop line). This map prompted the creation of still other versions.

Geoff Marshall, a tube fanatic (and holder of the world record for going round the entire system in the least amount of time), decided to gather these maps together into a folder, 'Silly tube maps', on his website. In addition to maps on which the names had been changed, Marshall posted a score of others (a map of stations with toilets, a map showing travel times between stations, a map on which dotted walk lines connected stations less than 500 m apart), and links to still others, including the London Tube Map Archive with its three dozen versions. Marshall made variations of his own, including *The Real Underground* that showed which portions of the Under ground 'were' underground, and a map on which the station names had been omitted, which became the basis for further variations on the part of others (the 'London undergrub' had in fact been inspired by Marshall's site). In March 2006, a lawyer representing Transport for London, which owns the tube map, threatened to shut Marshall's site down unless Marshall removed "ALL images which infringe my client's intellectual property [by] midnight on Monday 13 March." Marshall imme diately posted the threat which stirred a storm of con troversy and finally … a protest map.

Needless to say, it is in the style of Beck's tube map. However here the stations have been renamed, 'in', 'March', '2006', 'Transport', and so on, to create the sentence, "in March 2006 Transport for London's lawyers suddenly took offence to tube maps designed in the style of the Great Bear by Turner Prize nominated artist Simon Patterson being hosted by world record holding tube enthusiast Geoff Marshall and used legal bullying to force their removal. We think the people responsible for this decision are," and here you have to pull back to see the larger pattern made by the lines and stations which reads, in Harry Beckese, 'Wankers' (**Figure 3**).

The 'Wankers' map embodies everything we have come to expect of a protest map. Its creator, who wishes to remain anonymous, spelled out his motivations for us: "I read Geoff's blog on the subject, and I thought, 'what a colossal waste of public money' or words to that effect. Nothing he was doing was harming TfL, if anything, it was the reverse, and here they were, setting the packs of highly paid lawyers on to him, with my [expletive deleted] money. Particularly as it was just after the *Guardian* had published a map with musical artists, and made a big fuss of it. I think that's what inspired some people to create other maps – I saw some of the less functional maps as artistic endeavors. TfL knew he

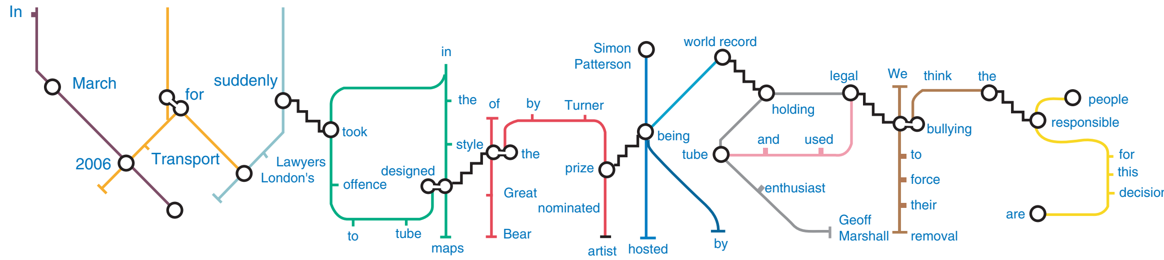

Figure 3 The Wankers map as posted on Geoff Marshall's website: http://www.geofftech.co.uk/tube/sillymaps/. The map was made by an anonymous contributor.

couldn't afford to defend himself; it seemed like corporate bullying, and it just stuck in my craw; even if you put the most benign view on their actions, they show a great misunderstanding of the internet and the difference between commercial websites, and personal sites; Geoff doesn't even carry any advertising, even though he gets a whole lot of traffic. At the time, I think it was Saturday 11 March, I had a fair amount of free time, and not much else to do, so I set about with a graphic package designing it. I thought it might cheer Geoff up a bit."

The map also makes it clear how hopeless the categorization of maps is. More than enough ink has been spilt already over whether Beck's original *Underground* map is a map or not, but only pedants refer to it as 'the Diagram'. What type of map it is, however, is another question. Helen Wallis and Arthur Robinson may have regarded it as a 'Route map', whereas Erwin Raisz would have been more likely to call it a 'Transportation map'. It is possible that none of these would have thought Patterson's *The Great Bear* was a map at all. Certainly, none of them had a category for art maps, much less art map parodies (or parodies of a parody, for whatever else it is *The Great Bear* is certainly a parody). Thinking about the 'Wankers' map as a protest, however, gets at its motivation (as well as its content and form), and motivation in the end is what really matters about all the maps we have discussed here. Like the anonymous creator of 'Wankers', they have all been moved by the perception of injustice, to the voters of Massachusetts, to Australians, to the black children of Detroit, to the inhabitants of an earth threatened by nuclear holocaust, to victims of the state system, to Geoff Marshall. Thinking about these maps from the perspective of motivation gets at aspects of them other ways of thinking about them cannot. But then perhaps all maps could be profitably approached from this angle. We would recommend it.

See also: Activism; Cartography, History of; Heritage; Inequality; Mapping, Philosophy; Maps.

Further Reading

Bunge, W. (1988). *The Nuclear War Atlas*. New York: Blackwell.

Gordon, A., Paqlen, T., Rogers, H. *et al*. (2008). *An Atlas of Radical Cartography*. Los Angeles, CA: Journal of Aesthetics and Protest Press.

Kidron, M. and Segal, R. (1981). *The State of the World Atlas*. London: Penguin.

Parton, J. (1877). *Comic Art*, p 316. New York: Harper and Brother.

Maps and the State

M. Farish, University of Toronto, Toronto, ON, Canada

Glossary

Apartheid Cartography A form of spatial reasoning associated with the apartheid (meaning 'separation' in Afrikaans) state of South Africa (1948–93), which featured the division of the country into racial groups and a hierarchy of separate residential areas for each group.

Banal Nationalism A term coined by the author, Michael Billig referring to the everyday, habitual, and unnoticed reproduction of nationalist sentiment, whether in the use of money, the reading of newspapers, the presence of flags in public spaces, and so on.

Geo-Body A term coined by Winichakul (1994) to encapsulate the territoriality of national culture, since the most crucial and constitutive element of a nation-state is its territorial home.

Governmentality An intriguing and evasive concept developed by the French philosopher, Michel Foucault to account for the expansion of the acts and arts of government associated with modern bureaucratic states, which treated their citizens as members of a legible population subject to arrangement and scrutiny.

Mappae Mundi A general term for Medieval European maps of the world, schematic rather than scientific representations which were often arranged to show the known world according to religious principles of order, with Jerusalem at the center and the heavens at the top of the frame.

Print-Capitalism A term coined by Benedict Anderson in his classic study *Imagined Communities* to capture the relationship between printing technology, or mechanical reproduction, and the commodification of printed texts, a relationship which facilitated the growth of nationalism.

Sovereignty The exclusive right to exercise political authority over a geographic region.

Territorial Trap A phrase coined by the political geographer, John Agnew to challenge the overwhelming attention paid to the state as an organizational political and geographical principle, since the state is a historical construction and certainly not the only scale at which politics operates.

Introduction

Maps cleaving the world into discrete state territories are ubiquitous (**Figure 1**). They can be found in school classrooms, newspapers and electronic media, and popular atlases. These proliferating and banal objects reinforce an entrenched assumption that sovereign, homogeneous states are the world's definitive political units – an assumption furthered by the common habit of granting each state a distinct color. However, this division is a relatively recent phenomenon in human history. Medieval makers and readers of the few maps in circulation, for instance, would have been baffled by this configuration. Thus the history of the modern state, with its bounded expanse of territory, is inseparable from maps and mapping. For the state to become a common, unremarkable geographic feature, particular forms of cartography were required, cartography that was both public and presumed to be scientific. Conversely, placing the state at the heart of a history of cartography is also revealing, as it prevents maps from being treated as inert manifestations of scientific principles, and instead demands a critical approach concerned with the power of maps to shape life.

Although its association with the consolidation of power means that 'the state' has ancient roots, the contemporary understanding of a state is usually historically coupled with the monarchical dynasties of Early Modern Europe, dynasties which gave way over varying intervals to more representative forms of government. The identification and policing of political boundaries was crucial to the solidification of what is now routinely called the state system. Within these limits sat an expanse of land, likely far larger than earlier feudal parcels, over which state institutions claimed authority, or, in Max Weber's famous phrase, a monopoly on legitimate violence. This sovereignty, however tenuous at times, has always been certified in various ways by cartography: a map that establishes a relationship between a state and its territory is an extremely potent device.

Maps can strengthen or undermine this relationship. They can be used for defensive or expansive purposes. The establishment of modern empires, by states in Europe and elsewhere, is a clear example of the latter. Since at least the late fifteenth century, colonial campaigns have been prompted, aided, and justified by cartography. However, addressing the link between maps and states requires genealogical care: neither are universal or innocent entities. A now well established scholarly literature has demonstrated the power of maps to reinforce or challenge dominant perceptions of geography. Maps, in this revisionist history of cartography, do not stand as records of reason but are instead treated as instruments active in the production of truths. Concurrent work in political geography, meanwhile, has used the concept of

Figure 1 The familiar world of states (2006).

the territorial trap to show how the idea of a state – just one form of political organization among many – has been naturalized to the detriment of other actors and scales. Politics is certainly not reducible to the state, nor is the state a unitary creature. Therefore, while the map and the state are closely aligned, both historically and philosophically, they are not irrevocably so.

Identifying and untangling connections between maps and states has been one of the tasks of the massive and ongoing History of Cartography project. Although the initiative has varied since its beginnings in the late 1970s, its founding purpose, to replace the Whiggish idea of maps as simple indicators of cumulative progress with an approach rooted in social theory, remains crucial to understanding the role of maps as tools of governance. Every cartographer would acknowledge that maps are selective representations. However, it is quite another thing to question the politics of mapmaking, or the uses to which maps can be put. If cartography is understood only as a precise science, made increasingly accurate by new technologies, then the political dimensions of maps are not always foregrounded. The representational distortions and limitations of maps have quite material effects. If mapping is understood as an activity which encourages certain types of geographic awareness and discourages others, then it seems quite relevant to examine the ways in which the state, as a particularly naturalized political entity, has been cartographically fashioned.

Spaces of the State System

Mappae mundi such as the famed Hereford map (dating to about 1300) are quite different from maps published during the Renaissance. The medieval kingdoms portrayed in more recent historical atlases are certainly cartographic impositions, in that their rulers and subjects would not have understood space in such a way. Rather than conceiving the world through the sort of visual lens offered by a map, those who claimed possession of a domain likely knew it as a series of places, routes, and distances. This was quite different than the perspective later offered by Renaissance maps, which turned geography into a surface through a cartographic gaze that could, in theory, remain standardized while scales or locations were shifted. The version of space produced by modern cartography is thus equivalent to the time measured by the modern clock.

Of the many elements contributing to the development of cartography in Renaissance Europe, one of the most crucial was what the sociologist, Benedict Anderson calls print capitalism, which permitted maps and other texts to be reproduced at a much greater pace, and for much larger audiences. Still, those who were most interested in maps were those seeking to gain and maintain power. In return, innovative or persuasive maps earned their makers support and patronage. Given the nature of this early market for maps, it is not surprising that they quickly acquired a close association with colonialism and war.

Even three decades ago histories of cartography would have likely been Eurocentric and progress driven, nodding to Greek and Roman antecedents before gathering steam in the Medieval Period, which could be usefully contrasted with the more scientific cartography that followed. Introducing the concept of the state renders these histories even more partial, but also more contextual and contingent, since mapping must then be treated as a political practice. And yet just as there is no singular definition of the state, there should not be a dichotomy between maps associated with states and those that are not. As European travelers first arrived in what was called the New World, or the Americas, they were certainly aware of and aided by indigenous forms of cartography. In 1540, the French explorer, Jacques Cartier was famously shown a stick diagram of part of the St. Lawrence River system. Such maps were no doubt practically useful, but less valuable when Cartier and others wished, as they often did, to lay claim to specific spaces, particularly a form of claiming that would resonate in the capitals of Europe where the Americas (named for the Italian explorer and merchant, Amerigo Vespucci) were being ruthlessly divided, renamed, and contested.

Granting New World possessions names such as Hispaniola (the island now containing Haiti and the Dominican Republic), as Christopher Columbus did in 1492, suggests the specific lines of authority that were being drawn across the Atlantic, precisely because Columbus had also presented plans for his ambitious journey to potential sponsors in Portugal, Italy, and England. The willingness of the Spanish monarchy to support his controversial proposals gave the country an early advantage in the European competition for overseas sources of trade and plunder. Less than a century later, King Philip II of Spain instigated a substantial review of Spanish possessions in the Americas and beyond. Called the *Relaciones Geográficas*, this survey depended heavily on indigenous artists for the maps of what was dubbed New Spain. However, this reliance did not last long, as Spanish colonial administrators sought standardized documentation that would enforce their claims to territory. Although colonial encounters certainly produced intertwined forms of geographic knowledge, native mapping methods were increasingly pronounced irrelevant within the sphere of imperial rivalry. European techniques, and European terms, were believed to be both superior and necessary.

The excitement and fear brought about by exploratory voyages and interactions with diverse forms of nature and culture was, not surprisingly, met with attempts to control the resulting maps. A hydrographic office was established in Lisbon at the end of the fifteenth century to manage and shape the information returned by, and provided to, the ambitious Portuguese navigators venturing into the Indian Ocean and beyond. Despite the close relationship between religion and empire, mapmaking was, by the sixteenth century, an increasingly secular exercise, the province of nonchurch bureaucracies. The ever more common and formal practice of diplomacy required new and secular cartographies to settle border disputes. In 1607, France's King Henry IV demanded a map of his coasts and frontiers, leading to the appointment of royal surveyors and ultimately small scale maps of the entire kingdom. The increasingly perfectionist geometry of military fortifications also heightened the need for mathematically accurate scales. As the historian, Jeremy Black concludes, in Early Modern Europe maps were overwhelmingly devoted to government and the maintenance of order.

When, in 1569, the Flemish cartographer, Gerardus Mercator (born Gheert Cremer) created a projection that rendered the globe as a cylinder, treating the meridians as parallel (rather than convergent), he ushered in a type of map, still common today, that drastically enlarged the temperate regions of the world – and reduced the size of the area known as the tropics. Although his projection was extremely useful for navigation, because constant straight lines could be drawn across a chart, its distortions magnified the size (and thus the prominence) of Europe, particularly Spain precisely as it was establishing the first global empire. Europe was at the top and in the center of Mercator's world map – a position of choice. All projections are distortions, but this orientation reflects additional, more political concerns. And while Mercator's era, one of early circumnavigation attempts, was also a period of increased interest in globes – he was himself a skilled globe maker – flat maps were obviously better at displaying geographic information at various scales, including that of an isolated political unit such as the state.

What is defined as 'geography' is always the result of struggle, including efforts to organize and administer territory. As the monarchies of sixteenth century Europe began to directly pursue policies of royal absolutism, their attempts to consolidate unitary authority over space met, and in many cases triumphed over, competing forms of geographical knowledge. The vertical, hierarchical order of *mappae mundi* (with Jerusalem at the center of the map, and the celestial realm at the top) was superseded by a horizontal cartographic plane, associated not just with states and sovereignty but with the shifts wrought by intellectual titans such as Galileo, Descartes, and Newton. If it was possible to understand space as horizontal, then it was also possible to divide or compartmentalize that space into units such as states. As Gearóid Ó'Tuathail has shown, the expansion of Elizabethan England into the island of Ireland in the sixteenth century required maps for strategy, subjugation, and subsequent governance. With maps, Ireland was turned, for the English, from an indecipherable and disorienting realm into a legible imperial possession. In so doing, the geographical entity of Ireland was fashioned anew, but so too was England, since the inevitable accompanying cultural distinctions – wherein the Irish were classified as irrevocably inferior – reflected the English desire for progress and advancement. Ireland, in other words, required a cartographic rewriting that was part of a much broader campaign to bring it under the sign of government.

It is easy to envision the symbolic authority of a map clutched in the hands of a monarch, politician, or navigator. However, using cartographic knowledge was impossible without institutions that worked toward its production and preservation. In 1573 Christopher Saxton was directed by Queen Elizabeth I to survey England and Wales. His efforts produced, in the form of both an atlas (printed in 1579) and a wall map (1583), an image of the territory that proved remarkably durable (**Figure 2**). However, the creation of formal cartographic agencies and archives that could sponsor field surveys lasting for many years, even decades, and then house the results, was not a simple process, and unfolded over the next several centuries. In 1747, King Louis XV commissioned a map of France at the scale of 1 to 86 400, a prodigious task that stretched on past the beginnings of the French Revolution in 1789. Supervised by two generations of the Cassini family, this mapping effort served as the basis for the 1791 *Atlas National*, and influenced other state driven surveying projects of the period. The English Board of Ordnance, for instance, initiated a one inch to the mile survey in 1790, a project which lasted over eight decades.

Such monumental initiatives were state driven and more importantly public. They stoked the burgeoning civic nationalism of the eighteenth and nineteenth centuries. However, these national map surveys were also proclamations of cartographic accuracy. The 180 Cassini map sheets covered an impressive 1300 square feet, providing exacting views of the state. These sheets were instrumental in the postrevolutionary effort to divide France into equal area *départements* that were based, in a profoundly modern gesture, on physical geographic features rather than cultural or historical associations. With the cartographic information provided by surveys, additional acts of government were feasible, from taxation to engineering projects. Equally, such maps were essential to the definition of a state's territory as a homogenous space. Whereas monarchs had literally embodied their realms, modern cartography introduced an alternate, less individual figure into the capitalist public sphere: what Thongchai Winichakul has neatly dubbed the 'geo body' of the nation. State cartography thus democratized access to geographic information, but only by prioritizing military, government, and property owning interests in map content and symbolism.

Extending the State

The global maps of the last half millennium narrate a contest between states over the territory of the world

Figure 2 'Anglia' from Christopher Saxton's *Atlas of the Counties of England and Wales* (*c.* 1579) (Special Collections Department, Glasgow University Library).

deemed strategically or economically valuable. However, it should not be forgotten amidst the romance of exploration and imperial competition that even the most hegemonic states possess borders which are constantly in need of reassertion. Indeed, the power of maps is perhaps most clearly found in their ability to create a state and its boundaries before such lines are well established in a formal sense (through the establishment of border patrols or posts, for instance). As the historian of cartography, J. B. Harley explained in an important essay, names such as New England, New France, and New Spain were on maps before these places were actively settled by Europeans. And once in place – though always subject to change – such names, and their reification on maps, encouraged and justified the appropriation of territory from its indigenous residents. The contrast between the frequently nomadic lives of these peoples – whether in Australia, Africa, North America, or elsewhere – and the fixity of colonial cartography could not have been clearer, with the result that nomadic behavior was, to put it mildly, discouraged.

Maps featuring names, enclosures, and hierarchies of space were fundamental to the exercise of colonial power. Imperialists divided the world with little regard to the nuances of local human geography, let alone environmental fragility. In hindsight, maps provide fascinating and disturbing portraits of the search for and occupation of territory in the Age of Empire. Mapping legitimized imperialism by rendering parts of the globe as empty or uncivilized. However, it was the combination of this superiority with the increasingly scientific status of cartography (not to mention the development of a formal academic discipline of geography) that made maps such useful tools of empire. A map can be a particularly clear way to practice what is sometimes called 'othering': distinguishing 'us' and our territory from 'them' and their territory. If this cartographic distinction is combined with a cultural equivalent, the recipe for colonization and its justification is secure. Maps have also always been essential to the perpetuation of colonialism, allowing imperial states to comprehend distant regions in familiar cartographic language.

The 1648 Peace of Westphalia, ending the Thirty Years War, is of such importance to international relations historians and theorists because it is said to have formally inaugurated the 'state system', along with the idea of noninterference – that a state's internal affairs are its sole, sovereign responsibility. Even less orthodox scholars have acknowledged the momentous character of Westphalia. The French philosopher, Michel Foucault observed that its balance of power approach resulted in new frontiers, new divisions, and new regulations for the relations between states, at least those European states considered on equal terms. War, it was believed, would henceforth occur selectively, supplemented and avoided by state diplomacy, and states would balance one another through such instruments as permanent armies. Within each state, meanwhile, communities would be regulated by policing – in the broad sense of the term – as the public face of authority and administration. According to Foucault, the objective of policing was to assure the life of the state and its subjects; to watch over, govern, and ultimately strengthen the state. This was a personal and statistical imperative, to be sure, but it was also most certainly a cartographic one. The emergence of the state system solidified and depended on the idea of 'population', as both discrete individuals and an aggregated national unit, but it also defined the category of 'territory' in the same terms: both could be given qualities, or properties, to be enclosed, controlled, divided, and of course mapped.

The distortions of maps are less interesting than their role in broader systems of spatial organization and understanding – what Foucault called regimes of truth. When Napoleon invaded Egypt in 1798, the French army produced a 47 sheet map of the region for the *Description de l'Egypte*, an ambitious attempt to gather "all the information necessary for making known the modern condition of Egypt." The imperative behind the colonial mapping of Egypt was, paradoxically, driven by the Enlightenment desire for universality; conquest, in other words, could be scientific. Numerical and visual, concerned with both measurement and representation, maps reinforced the belief that the world could be rationally ordered according to a single set of immutable principles. This belief was derived from a selective appropriation of classical traditions, a move that is clear in the *Description*, in which the glory of imperial France is linked to that of ancient Egypt. Except when rendered in negative terms, contemporary Egyptians are overwhelmingly absent from the text. The conquest of Egypt was, by this logic, returning civilization to its origins, and the *Description*, with its systematic and detailed maps, was the symbol of this gift. As a sign of progress, scientific cartography rationalized dominance. Nowhere in the *Description de l'Egypte* is there a record of the force used to subjugate Egyptian people and territory; this violence lies below the surface of the maps.

That mapping played such an important role in French occupied Egypt was no surprise; military geographers also followed Napoleon's armies as they battled European powers from Portugal to Russia. Such constant campaigning required the repeated redrawing of frontiers as states were created or altered. Rivals such as Britain hastened to draft maps that would aid in defending a potential French invasion. However, Britain was also engaged in an unprecedented set of imperial campaigns, which resulted, by the nineteenth century, in the famed type of world maps with British possessions colored red or pink. In other, more precarious European regions lacking independence, such as what is now Finland or Bulgaria, the publication of nineteenth century atlases stood as gestures of defiance (against Russian and Otto man rule, respectively). In nineteenth century Siam (now Thailand), despite the absence of colonial rule the presence of British troops in nearby Burma encouraged the adoption of a rapid 'Westernization' policy under King Mongkut, who was fascinated by astronomy and cartography. Self defined nations that lacked their own states were actively seeking territorial privileges. De pending on what it displays, a map can establish claims to a homeland or rebut them, using historical evidence to create a genealogy of a nation rooted in a particular space. And while the dangerous insinuation that the na tion, as a community, is inextricable from the state is often associated with political declarations (Benito Mussolini's "everything in the state," for instance), such pronouncements do not encompass the banal or everyday aspects of nationalism, including the appearance of national maps on postage stamps or classroom walls. However, as the examples in this paragraph show, writing historical geography without reference to states is a challenging exercise.

Nineteenth century imperial competition and na tionalist differentiation also resulted in more explicitly propagandistic maps, representations that discarded sci entific or objective pretenses in favor of overt persuasion or more subtle manipulation. The most striking of these turned political regions into caricatures of singular human or animal figures. Ascribing a particular person ality or identity to a state was not novel, but colonialism, alongside the increased attention paid to anthropology, race, and evolution, made this cartoonish cartography more prevalent and powerful. The desire for cartographic precision was thus countered by the need to use straightforward methods that would educate, persuade, and amuse the public. Lurid propaganda maps, un surprisingly, have been durable. *Prelude to War* (1942), the first of Frank Capra's War Department sponsored *Why We Fight* documentary series, indicted Americans for as suming that isolation was still possible in an era of global hazards. In the film, maps of Germany, Italy, and Japan were transformed into threatening symbols. Several years later, American periodicals frequently featured maps depicting the Soviet Union as a bear, its paw extended menacingly into Asia and the Middle East.

As a geographic technology, then, the map brings the cultural entity of the nation into contact with the polit ical vehicle of the state. Some of the most important agents for this conjunction have been national geographic societies. In the United States, the organization of the same name was founded in 1888, and some 9 months later the first issue of the *National Geographic Magazine* (later shortened to *National Geographic*) was released. The first issue with an added map was, notably, published in 1899. The map featured the Philippines, which had been ceded to the United States from Spain a year earlier. The fledging American Empire clearly required cartographic explanation and justification.

Surveying the American West

Several decades before the establishment of the National Geographic Society, Jedediah Morse had called for, and then produced, new textbooks oriented toward a par ticularly American understanding of geography. For this, his position in the history of American geography is a familiar one. But an equally important figure in the consolidation of the United States as a national space was the educator, Emma Hart Willard (1787–1870). Deter mined to create a richer concept of history for the new nation, Willard was also influenced by Morse's geography, and was fascinated by the ability of maps to 'place his tory'. She was writing and teaching at precisely the time that historical atlases, with their unique ability to deepen both the historical and geographic character of a nation, first proliferated. Willard's hugely popular textbook, *History of the United States, or the Republic of America*, re printed almost annually from 1828 to 1860, was accom panied by a series of maps, arranged in chronological order and featuring the same general area. As the his torian, Susan Schulten notes, this was the territory "that would *become* the United States." Projecting sovereignty back into the past, this cartography was explicitly na tional, lending legitimacy and clarity to a state whose boundaries were still very much in flux. Willard was clearly aware of the role maps play in creating a national geo body that is politically solidified around the struc ture of the state.

The nineteenth century was also the period of sub stantial exploration and occupation of the American West an activity which relied upon and produced extensive cartographic information. Institutional support for map ping expeditions was provided by the Army Corps of Topographical Engineers, yet another example of the close affiliation between the state and military geography. Mapping activity was often conducted alongside cam paigns to subdue native groups such as the Navajo, and was more generally essential to establishing control over the territory of the West.

In his 1868 book, *The Parks and Mountain of Colorado*, Samuel Bowles described "maps from Washington, that put down only what is absolutely, scientifically known," leaving "a great blank space" in the area of the Colorado Plateau. This was, of course, a cartographic fabrication, since such blankness, as in the case of colonial Australia, depended on both erasures and the containment of a blank region by its opposite: known space. However, the deliberate presentation of regions as empty invited survey teams to fill them. In the case of Colorado the task was conducted by the explorer, scientist, and bureaucrat, John Wesley Powell. Filling in the last gap on the con tinental map secured Powell's reputation as an iconic figure in the history of the American West. Despite minimal initial support from Washington, it was still clear to him and his crew that they could only complete their task by effectively taking the Plateau to the capital in the form of data and narrative.

Recounted by Powell in adventurous language, his expedition quickly garnered significant publicity, leading Congress to belatedly finance the venture and in corporate his work into the sphere of government survey science. The resulting 'Geographical and Geological Survey of the Rocky Mountain Region', or Powell Survey, was an impressive and interdisciplinary effort, yielding, it was believed, information of great practical value to various business interests and arms of the government. The ultimate aim was the production of more suitable maps which would hasten settlement and resource de velopment. The complications of Powell's project, and of data circulation between the West and Washington, were overwritten by claims of professional certainty. By the mid 1870s, all Western surveys were to be conducted in a systemic, uniform fashion, funneling into a new *Atlas of the Territories of the United States* produced by the De partment of the Interior. While this atlas was never fin ished, the intentions behind it certainly lingered on, and various surveys were eventually incorporated into the United States Geological Survey (USGS), established in 1879 and directed by Powell from 1881 to 1894.

To move from a complex local scale to a clear state wide view necessarily involves simplification, such that the state supported map, that exemplary project of legi bility, will show only that which interests the state – details crucial to the governance of populations and territory. This is itself a simplistic presentation of states, which are constellations of diverse institutions, indi viduals, and practices, but the aim of legibility and the inevitable failure or limits of this aim are crucial to understanding the relationship between maps and the state.

While the major Western surveys were underway, the economist, Francis Amasa Walker was finishing his own influential *Statistical Atlas* (1874) for the US Census Bureau. Walker directed both the 9th and 10th American Censuses (1870 and 1880), and was also a commissioner for Indian Affairs in the early 1870s. His thematic maps were some of the first cartographic constructions to fully rationalize the West as American. Once 'unknown' re gions were perceived and represented as closed, bounded by known space, and subject to a property regime, this regime was verifiable. Mapping the Colorado Plateau, then, gave order to an uncertain and shifting landscape precisely when it was deemed most necessary.

The *Statistical Atlas*, according to geographers such as Matthew Hannah and Jeremy Crampton, was a particu larly significant representation of relationships between society and space. Thematic maps – including the choropleths first produced in the early nineteenth cen tury – were not strictly the tools of states, but they were nonetheless governmental in the way they displayed calculable populations within a territorial field. Walker's *Atlas*, therefore, stands as a superb example of a more general nineteenth century phenomenon: the con junction of a state's desire for information about its constituents (manifested in the census) with the calcu lation and measurement made possible by the nascent discipline of statistics, and also the drive to display the results of this juxtaposition in cartographic form.

A survey which produces and justifies a particular property regime is almost certainly rewriting an earlier geography. In settler colonies from Australia to the Americas, to do this with conviction required a double move. First, older regimes, often indigenous, were ig nored in favor of the seemingly rational and sensible system introduced by a comprehensive survey. Second, these indigenous populations were incorporated into the survey, not as equivalent to settlers, but as historical ar tifacts. Powell's survey included considerations of native languages, culture, technology, and archaeology. He is now praised as a progressive promoter of ecological values, given his opposition to those he did not trust as stewards of the West. However, his mapping of this region to make it legible and valuable from afar also had more violent effects.

The State Goes Global

With the end of the nineteenth century, imperialism seemed to reach a sort of cartographic terminus. In his legendary 1904 address to Britain's Royal Geographical Society, the British geographer and Conservative polit ician, Sir Halford Mackinder described an increasingly closed world, and used a map to illustrate what he called "the geographical pivot of history" rooted in the Eurasian heartland. This cartographic 'reality' was, for a loyal supporter of the British Empire such as Mackinder, troubling, as it seemed to signal the increasing margin alization of Britain next to the might of Germany and Russia.

As armies expanded in advance of World War I, so did the collection of intelligence, which depended directly on the production of maps. Much of this cartographic ma terial went unviewed by the public in belligerent states. However, it was clear to mapmakers that their influence on political life was growing, and that maps could be useful tools in shaping geopolitical beliefs. Whereas the readers of newspapers at the beginning of the nineteenth century would not have expected to encounter any maps, a century later these features were expected. And one of the crucial means of unifying the nation has historically been the newspaper.

The timing of Mackinder's address was both aus picious and unfortunate. He certainly anticipated the emergence of the Soviet Union as a superpower, but he had little to say about the United States. Nor could he predict the impact of the airplane on geopolitical theorizing. Indeed, by the time of World War I, aerial photography had dramatically altered cartographic con ventions. Maps aiding artillery and bombing attacks were improved, largely owing to the fact that a region such as the Western Front could be rapidly remapped from the air, so as to account for shifts in position. Photo inter preters could review dozens of images of the same target, taken from different angles.

Not surprisingly, the World Wars also featured, and prompted, an explosion of propaganda mapping. By the 1930s, German atlases were already supporting the ex pansionistic racism of the Nazi Party, whose officials also encouraged the public display of maps depicting German populations living in foreign states, on the one hand, and maps suggesting Germany's vulnerability to a host of threats, from France to the Soviet Union, on the other. But to draw a line between a category of propaganda maps and a more scientific variety is a mistake; it is more appro priate to treat them on a spectrum of approaches and intentions. A map provides a particular set of ways – crude or clever, orthodox or radical – to address a geopolitical situation. According to the historical geographer, Michael Heffernan, the mapping efforts of various national geo graphic societies – in London, Paris, New York, and elsewhere – during World War I not only indicate how central maps were to wartime strategy, but also confirm that cartography played a fundamental role in political decision making before, during, and after the war. For the 'Inquiry' accompanying the post war Paris Peace Con ference, an American cartographic group led by the geographer, Mark Jefferson contributed hundreds of maps.

During World War II and the early Cold War, the Earth itself became an object of geopolitical scrutiny. However, such universal views were always directed from somewhere, and more than often these global views were American in origin. The tendency to visualize the Earth as a natural or cultural whole was of course much older, but the world maps of mid twentieth century geopolitics were more fully realized and systematic versions of such perspectives. ('Maps', in this case as always, were not just visual artifacts; the speeches of Cold War political leaders, for instance, were rife with spatial imagery that led to such familiar concepts as containment.) The American popular cartography of the period presented a particularly impressive vision of a national globe. The appearance of such imagery in periodicals, to take an obvious example, was not directly demanded by the American government, but this is less important than the way such maps naturalized the American state as a (be nevolent) global hegemon.

Some aspects of mid twentieth century globalism were cosmopolitan, appealing to geographic scales be yond the state. Many of the contributors to the 1944 collection *Compass of the World* were determined, as the humanist writer and Librarian of Congress, Archibald MacLeish wrote, to make the globe "truly round: a globe in practice, not in theory." This was an optimistic goal, given the growing predominance of a realist international relations discourse that invoked Westphalia and other earlier episodes in support of a state centric balance of power. And, indeed, it was in the arena of cartography that *Compass of the World* fell short as an idealistic exercise. The book's maps were created by the prolific Richard Edes Harrison, familiar to readers of *Fortune* and *Life*. His illustrations presented views from a fixed point beyond the Earth rendered at unusual angles, an illusory approach achieved by first photographing a large globe. The uncluttered appearance of Harrison's iconic maps departed from familiar conventions so much that they were more akin to advertising than anything previously classified as cartography. Just as importantly, these images turned viewers into pilots, a perspective fitting exactly with the emergence of the 'air age' and the terrible tri umph of air power in World War II. Harrison's com positions were in fact reprinted in American military periodicals and used to train actual airmen, but they outlasted the war itself, remaining popular well into the next decade (**Figure 3**).

Harrison's maps were certainly flexible enough that they could reinforce state boundaries or reveal the limits of these divisions. But as World War bled into Cold War, the first of these options began to dominate the second. And Harrison was hardly alone in this bias. In *The Geography of the Peace*, his short 1944 manifesto on the shape of the post war spatial order, the prominent Yale University geographer, Nicholas Spykman expressed af fection for a particular global projection, produced by the American Geographical Society, which could be easily used to portray the United States as both central and besieged by other landmasses. For his part, Harrison produced his own companion volume to *Compass of the World* in 1944. *Look at the World*: *The Fortune Atlas for World Strategy* was most certainly the most popular American atlas produced for and during World War II. Harrison's ability to translate the novelty of conflicts into graphic images was radical, but these visual sensibilities did not necessitate a cartographic de centering of the state – or the United States. Rather, his projections revealed, as he put it, that the "entire conflict pivots around the U.S.," and that the world had divided into two camps, "those who are for us…and those who are against us."

Figure 3 'Outward from the U.S.' from Harrison, R. E. (1944). *Look at the World: The Fortune Atlas for World Strategy*. New York: Knopf, 13.

Not surprisingly, Harrison's methods prompted criticism from other cartographers and students of geography. Predictably, some confronted him for failing to conform to standardized cartographic conventions, or for the sacrifice of mathematical precision. Harrison responded, appropriately, that all mapmaking was subjective. But in so doing, he was also implicitly acknowledging the veracity of another accusation: that his work was propagandistic. By drawing the United States closer to enemy spaces and forces converging on North America, Harrison's maps rendered the state automatically internationalist, forced to contend with new forms of combat and aviation technology. He and others considering the cartography of global war were not afraid to admit that the map could be a weapon. Secure in the knowledge that their representations were helping to shape the contours of a world at war, they chose what they thought to be the appropriate perspective, the necessary truth required for successful strategic planning – by the strategists of their own states.

World War II was a boon not only for Harrison, but for cartography more generally. War, quite simply, encourages public interest in maps – and specific types of maps. In a February 1942 radio address, President Franklin Roosevelt – himself lavishly provisioned by the National Geographic Society – famously instructed Americans to "take out and spread before you a map of the whole earth." Geographical awareness, at least that which could be derived from either a globe or one of the many war atlases sold to the public, became a valuable and much discussed commodity. Following Roosevelt's address, *Publisher's Weekly* printed a story titled "Presidents' Speech Causes Great Demand for Maps," suggesting that "the demand was almost as great as it was just after Pearl Harbor." Rand McNally's map store in New York hastily installed two window displays on 'Geography for Defense', one focused on national defense, the other internationalist.

Also in 1942, Roosevelt (and British Prime Minister, Winston Churchill) received a Christmas gift from Army Chief of Staff, George C. Marshall: a 50 inch, 750 pound globe, the brainchild of William Donovan, Director of the Office of Strategic Services (OSS). Looking back on the task, the cartographer Arthur Robinson mused that sitting next to the giant globe would have given Roosevelt and Churchill "a view rather like that of an astronaut" (**Figure 4**). This view is what the cultural geographer, Denis Cosgrove has called the Apollonian eye, a disinterested consideration of the Earth from above. Rising beyond the parochial planet might render the world's political divisions irrelevant or at least suspect. But it is important to recall that even the famed Whole Earth photograph, which has become the symbol of both environmentalism and corporate globalization alike, was

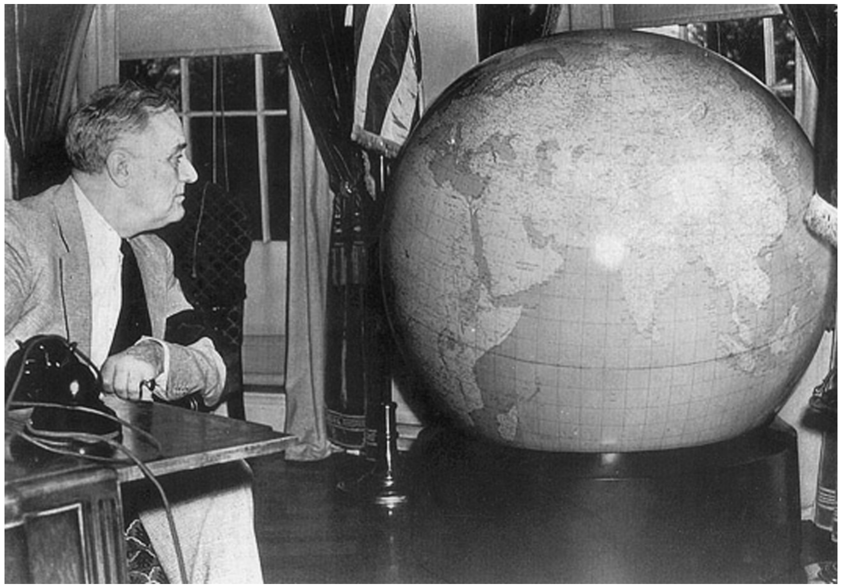

Figure 4 American President, Franklin Roosevelt next to his 50-inch globe, December 1942 (General Collections, US Library of Congress).

taken by an Apollo 17 astronaut in 1972. As Senator (and future President) Lyndon Johnson put it in 1958, the year NASA was established, it was imperative that America secure "the ultimate position – the position of total control over the earth that lies somewhere out in space."

When not designing presidential globes, Arthur Robinson headed the substantial OSS Map Division, which compiled and produced various forms of strategic cartographic information. Initially supplied by the Library of Congress, the Department of State, and the Army Map Service, the Division supplemented its collection after a national radio appeal by OSS Director, Donovan. The millions of items eventually accumulated by the busy Map Division necessitated the development of expanded classification and cataloging standards which lasted long after the war. The OSS eventually transformed into the Central Intelligence Agency (CIA), a prominent hub for Cold War maps and geographic information.

The July 1956 issue of *Scientific American* reported that the US Army Map Service had just calculated the "longest line ever surveyed," from Finland to southern Africa. The most important result of the exercise was a revision of the Earth's estimated radius. Not only would maps become more accurate, the article implied, but the new measurements would be used to plot the trajectories of satellites, notably for the 1957–58 International Geo physical Year (IGY), and also for more explicitly (if un mentioned) military campaigns. These satellites were, in 1956, still the stuff of science fiction scenarios and clas sified speculative reports, and when a space vehicle was actually launched a year later it was, alarmingly, not an American product. However, these sorts of media stories reveal the degree to which global mapping, by the time of the Cold War, had overwhelmingly become the province of 'one' state.

According to the cartographic historian, John Cloud, the mapping technologies of today, from geodesy and remote sensing to geographic information systems (GISs), can be comprehensively traced to the Cold War United States. All of these were tied to American in telligence and military objectives: studying the Soviet Union with an eye to potential conflict, but also other strategically significant regions where the superpower dispute at the heart of the Cold War was frequently contested by proxy. If maps are shot through with power, then the 'silences' of maps, often enforced by an infra structure of secrecy, are worth attending to. The Cold War is particularly challenging for students of secrecy, not just because it is recent history, but because it has led to a situation, arguably more prevalent today than ever, where, as Cloud (2002: 261–282) writes, "the products of highly classified technologies," restricted rigorously by various arms of the American state, "are displayed can didly as completely unclassified maps and data." Thus, even the ostensible democratization and proliferation of maps (and the ability to create them) can conceal the ways in which these maps, and the devices and agencies responsible for their creation, are anything but public.

The Present State of Cartography

Maps can now be generated faster and in larger numbers than at any time in history. Projections and perspectives can be switched easily, and ever more data can be fed into mapping software. While innovative programs such as

Google Earth give individuals – at least those with access to a computer – the opportunity to create participatory geographies, this relative transparency is still limited compared to the cartographic power of state intelligence agencies and militaries, particularly those of the United States. Visit the website of the National Geospatial Intelligence Agency (NGA) and an official history cites the Lewis and Clark expedition of the early nineteenth century as the first example of American 'geospatial intelligence'. The timeline then glides through World War I, the establishment of the CIA in 1947, and the U 2 spy plane and CORONA satellite programs before arriving at the formation of the unified Defense Mapping Agency (DMA) in 1972.

The DMA produced over 66 million maps in support of the First Gulf War (1990–91), a staggering figure. However, further integration of mapping and imagery analysis was required, leading to the 1996 creation of the National Imagery and Mapping Agency (NIMA), a joint CIA Department of Defense agency mandated to produce ever more geospatial intelligence. Its name was changed to the NGA in 2004, but not before, as the website states, NIMA participated in the resolution of border disputes in South America, the Middle East, and the Balkans. These seemingly innocuous interventions were thoroughly geopolitical and state centric. Unmentioned on the website is the controversial role of American satellite technology in locating supposed weapons of mass destruction in Iraq – also a state, but one that, because of its 'failed' or 'rogue' status, was deemed less sovereign and invaded in 2003. The NGA is today the most obvious and substantial example of the close, and violent, connection between mapping and the state. But additionally, in another reaffirmation of the longstanding connections between geoscience and the state, several sizable commercial GIS firms have also become fixtures in the burgeoning 'homeland security' market.

If, as political geographers have suggested, we understand geography as geo graphy, an active form of Earth writing, then the act of mapmaking is effectively intelligence gathering. What, after all, is a census or a similar exercise of administration without a territorial frame? A map does not just show the surface of a territory; it arranges the features of that space in a specific order that has been established to benefit a map's creators. And the correlation of cartography with intelligence is an old one; the late fifteenth and early sixteenth century voyages of Columbus led to a collection of maps which his Spanish sponsors attempted to keep secret.

The more 'advanced' democracies of today are hardly more open with their cartographic information, despite the pervasive preconception that 'our' maps stand for accuracy and neutrality. Activists and curiosity seekers worldwide delight in uncovering facilities – often associated with militaries – that do not exist on government produced maps. Such erasures are justified by the invocation of national security, and indeed it is a mistake to simply counter these claims with a demand for scientific accuracy. Here again, cartographic scholars have suggested an alternate approach: that we consider how efficiently maps, especially those produced by dominant political actors and agencies, convey what they are supposed to show. The failure to show certain facilities may well be a mark of efficiency. So would the refusal to acknowledge a map's author, or its stated purpose. Unmasking the erasures, silences, or biases of a purportedly scientific or authoritative map is ultimately to challenge the state and its geographic foundations.

Cartographic Violence

The preceding historical sketches point to several broader conclusions about the role of cartography and its relation to violence. As the locus of sovereignty moved from the individual feudal ruler to state institutions, the body politic (or, later, the nation) was given a geographic dimension in addition to its necessary history. Geography thus lies at the very heart of the idea of the state. And the state's claim to sovereignty over its territory – in essence the ultimate form of property right – depends on 'seeing' that territory clearly, in cartographic and other forms, just as a single property owner does with a local map.

Without this correlation of geography with political power, sovereignty would become a fiction. Thus, to speak of politics without reference to geography – or geopolitics – is to perpetuate an abstraction. Equally, however, to claim territory – and the word territory is itself suggestive – is to rule through force, as the history of unsuccessful internal rebellions has shown. To fully rule a territory requires the authority to violently control it if necessary. This is the source of Weber's well known maxim: a state holds a monopoly over the legitimate use of violence within its territory, meaning that violence perpetrated by other actors is illegitimate.

Historians use the example of piracy to illustrate this argument: once treated as legitimate or semilegitimate alongside other forms of violence, and frequently supported by monarchs or cities, piracy was at some juncture cast outside of the sphere of legitimate violence, likely when, in the eighteenth century, the British East India Company asked the Royal Navy to protect company ships against British pirates! More recently, discussions of legitimate violence have figured prominently in the moral and legal prosecution of the American 'War on Terror', with certain 'combatants' restricted from holding human rights because they are apparently affiliated with no state. Like pirates, many defined as terrorists are classified as stateless, and to be stateless in any sense today is understood as a damaging trait.

Those individuals and groups who are mapped outside of a state – and this does not necessarily have to be a physical exclusion – are also abandoned by its legal order, an abandonment which can be violent indeed. The special, accepted status of state violence results in dis proportionate attention paid to wars between states, whe ther in historical scholarship or contemporary media accounts of conflict, despite the evidence that most forms of armed struggle do not fit this model. The bias favoring state versus state wars results in the contrasting ignorance or perceived illegitimacy of intrastate disputes which often feature groups that do not fit cartographic conventions; they are literally off the map. It has become all too easy recently to give these nonstate actors a generic name, in distinction to specific states: terrorists. This definition of terrorism is in fact the product of the state system: a wide variety of actions grouped under this category have been consistently set up in distinction to, and outside of, the system.

Examples of the affiliation between cartography and state violence are legion. One of the most chilling must certainly be the case of apartheid South Africa, featuring the Bantustans, or homelands, established for the majority African population. Promoted in a language of ethnic preservation, these homelands embodied the territorialization of racism, combining appeals to un changing purity and geographical isolation. Significantly, the post apartheid state was remapped into a structure which unsettled the previous linkages between identity and geography. 'Apartheid cartography' has not been limited to South Africa. The political geographer, David Campbell has deconstructed similar forms of mapping applied to postconflict Bosnia during the 1990s, and the broader relationship between maps and identity based exclusion is a deep one. In a striking intervention the architect Eyal Weizman (2007) has recently demonstrated how Israel's occupation of Palestine rests on a "politics of verticality": the separation of a patchwork landscape of settlements in the West Bank has been achieved through the securing of hilltop sites, airspace, subterranean re sources, and a series of bridges and tunnels, all of which are separate from – and more specifically over and under – Palestinian sovereignty. While extraordinary, this case is not unique, since sovereignty has frequently been treated in three dimensions, particularly in colonial contexts. Still, Weizman's own dissident cartography shows just how limited, and thus insufficiently critical, the persistent horizontal approach to space prevalent in modern car tography can be.

De-colonizing Maps and States

Considering the historical associations between maps and states is necessarily a study of cartographic authority; state produced and state oriented maps are among the most powerful forms of geographical representation. However maps, or at least cartographic tropes, also ap pear prominently in more subversive realms, such as that of postcolonial literature. That they frequently appear there in an ironic or humorous light is all the more interesting. These maps – literal or figurative – not only serve to interrogate official histories of colonialism, but they also challenge the unquestioned nationalism lingering within hybrid postcolonial states. This carto graphic revisionism has been a source of some import ance for scholars with the conventional history of cartography in their sights. The cartography that became a colonial reality in locations such as India was itself a heterogeneous product, crafted and brought together by various individuals. There may have been an overarching goal – to distribute or organize Indians in positions that benefited the colonial authority – but this was always, in its fullest extent, an unattainable objective. In this sense the affiliation of maps and the state, and the supremacy of this affiliation, have never been complete.

Independence from colonialism, meanwhile, has fre quently been heralded by the publication of national atlases. These texts, and more generally the use of maps by residents of the world's less developed regions, have largely escaped the attention of cartographic scholars. Postcolonial atlases still carry lingering traces of coloni alism: many of them have been produced by multi national publishing firms based in the world's richest countries, and more importantly, their maps frequently reflect standards established by colonial powers. The publication of a national atlas as a monument to in dependence may thus be understood in part as a further reification of the state as a geopolitical unit. However, despite the global reach of certain cartographic conven tions – certain ways of seeing – maps still carry different meanings in different locations, and there are of course varying forms of cartographic literacy and exposure to maps even within states.

See also: Cartography, History of; Colonialism I; Colonialism II; Critical Geopolitics; GIScience and Systems; Governmentality; Imperialistic Geographies; Military Geographies; Nationalism; Scale; State; Surveying.

Further Reading

Anderson, B. (2006). *Imagined Communities: Reflections on the Origin and Spread of Nationalism* (rev. edn.). London: Verso.
Biggs, M. (1999). Putting the state on the map: Cartography, territory, and European state formation. *Comparative Studies in Society and History* 41, 374 405.
Black, J. (2003). *Visions of the World: A History of Maps*. London: Mitchell Beazley.

Campbell, D. (1999). Apartheid cartography: The political anthropology and spatial effects of international diplomacy in Bosnia. *Political Geography* 18, 395 435.

Cloud, J. (2002). American cartographic transformation during the Cold War. *Cartography and Geographic Information Science* 29, 261 282.

Collier, P. and Inkpen, R. (2003). The Royal Geographical Society and the development of surveying, 1870 1914. *Journal of Historical Geography* 29, 93 108.

Cosgrove, D. (2001). *Apollo's Eye: A Cartographic Genealogy of the Earth in the Western Imagination*. Baltimore, MD: John Hopkins University Press.

Crampton, J. W. (2004). GIS and geographic governance: Reconstructing the choropleth map. *Cartographica* 39, 41 53.

Crampton, J. W. (2004). Exploring the history of cartography in the twentieth century. *Imago Mundi* 56, 200 206.

Edney, M. H. (1997). *Mapping an Empire: The Geographical Construction of British India, 1765 1843*. Chicago, IL: The University of Chicago Press.

Elden, S. (2007). Governmentality, calculation, territory. *Environment and Planning D: Society and Space* 25, 562 580.

Godlewska, A. (1995). Map, text and image The mentality of enlightened conquerors: A new look at the Description de L'Egypte. *Transactions of the Institute of British Geographers* 20, 5 28.

Hannah, M. (2000). *Governmentality and the Mastery of Territory in Nineteenth Century America*. Cambridge: Cambridge University Press.

Harley, J. B. (2001). *The New Nature of Maps: Essays in the History of Cartography*. Baltimore, MD: The Johns Hopkins University Press.

Harrison, R. E. (1943). The war of the maps: A famous cartographer surveys the field. *The Saturday Review of Literature,* (August), 24 27.

Harrison, R. E. (1944). *Look at the World: The Fortune Atlas for World Strategy*. New York: Knopf.

Heffernan, M. (2002). The politics of the map in the early twentieth century. *Cartography and Geographic Information Science* 29, 207 226.

Kirsch, S. (2002). John Wesley Powell and the mapping of the Colorado Plateau, 1869 1879: Survey science, geographical solutions, and the economy of environmental values. *Annals of the Association of American Geographers* 92, 548 572.

Mundy, B. E. (2000). *The Mapping of New Spain: Indigenous Cartography and the Maps of the Relaciones Geográficas*. Chicago, IL: University of Chicago Press.

Neocleous, M. (2003). Off the map: On violence and cartography. *European Journal of Social Theory* 6, 409 425.

Ó Tuathail, G. (1996). *Critical Geopolitics: The Politics of Writing Global Space*. Minneapolis, MN: University of Minnesota Press.

Pickles, J. (2004). *A History of Spaces: Cartographic Reason, Mapping and the Geo coded World*. London and New York: Routledge.

Robinson, A. H. (1997). The President's globe. *Imago Mundi* 49, 143 152.

Schulten, S. (2007). Emma Willard and the graphic foundations of American history. *Journal of Historical Geography* 33, 542 564.

Schulten, S. (2001). *The Geographical Imagination in America, 1880 1950*. Chicago, IL: University of Chicago Press.

Scott, J. C. (1998). *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven, CT: Yale University Press.

Sparke, M. (1998). A map that roared and an original atlas: Canada, cartography, and the narration of nation. *Annals of the Association of American Geographers* 88, 463 495.

Weigert, H. and Stefansson, V. (1944). *Compass of the World: A Symposium on Political Geography*. New York: Macmillan.

Weizman, E. (2007). *Hollow Land: Israel's Architecture of Occupation*. London: Verso.

Winichakul, T. (1994). *Siam Mapped: A History of the Geo body of a Nation*. Honolulu, HI: University of Hawaii Press.

Wood, D. (1992). *The Power of Maps*. New York: Guilford Press.

Relevant Websites

http://www.special.lib.gla.ac.uk
Christopher Saxton's Atlas of the Counties of England and Wales (*c*. 1579), Special Collections Department, Glasgow University Library.

http://www.newberry.org
Cold War Popular Magazine Cartography: A Virtual Slide Set, Newberry Library: Smith Center Publication.

http://www.dur.ac.uk
International Boundaries Research Unit, Durham University.

http://www.lib.utexas.edu
Perry Castaneda Library Map Collection, University of Texas Austin.

http://www.sunysb.edu/libmap/coordinates/seriesb/no6/b6.htm
Recent Trends in the History of Cartography: An Annotated Bibliography by Matthew Edney, Stony Brook University.

http://www.lib.utexas.edu
Relaciones Geográficas Collection, University of Texas Austin.

http://www.geography.wisc.edu/histcart
The History of Cartography Project, Department of Geography, University of Wisconsin Madison.

http://www.nima.mil
US National Geospatial Intelligence Agency.

Markov Chain Analysis

W. Li and C. Zhang, University of Connecticut, Storrs, CT, USA

Glossary

Algorithm A step-by-step problem-solving procedure, especially, an established, recursive, computational procedure for solving a problem.

Correlation Statistical techniques or measures which show whether, and how strongly, pairs of variables are related.

Geostatistics A collection of statistical methods which were traditionally used in geosciences, describing spatial correlation among sample data and using it in various types of spatial models.

Markov Chain Random Field A single Markov chain that moves or jumps in a space and at any location, interacts with its nearest known neighbors in different directions, and decides its state by the interactions.

Nearest Known Neighbor A location with known value which is the nearest along one direction to the location to be estimated and which may be a sampled location or a previously estimated location.

Nonlinear Property of a kind of system whose behavior is not expressible as a sum of the behaviors of its descriptors.

Range The distance in which the difference of the variogram or transiogram from the sill gets neglectable.

Simulation An imitation of some real thing or process, which generally entails representing certain key characteristics or behaviors of a selected physical or abstract system.

Sill The limit of the variogram or transiogram tending to infinite lag distances.

Transiogram It refers to a transition probability diagram. Theoretically, it is defined as a two-point conditional probability function $p_{ij}(h)$ over the distance lag h.

Introduction

Probabilistic models are appropriate in human geography for obvious reasons: geographical concepts and mea surements tend to be inexact, geographical relationships are often complicated and poorly understood, and geo graphical manifestations of human behavior remain un predictable despite much research. It is convenient to treat the processes of changes as random in aggregate and to describe them in a form of some stochastic process describing the changes in the locational pattern. The Markov chain theory is the simplest process of this kind.

Since Brown's analysis of innovation diffusion in 1963, Markov chains entered the geographical literature. Then one dimensional Markov chains became popular tools in the late 1960s and early 1970s in human geography for describing and modeling social mobility, such as migration, city growth, and changes in population distribution, residential structure, transport networks, industrial structure and pattern, and land use. Collins *et al.* provided a review on applications of Markov models in geography in 1974. However, the applications of Markov chains in human geography faded relatively since the late 1970s. The main reason may be related with the limitations of one dimensional first order stationary Markov chains and the difficulties in constructing high order or multidimensional Markov chain models.

Recently, a Markov chain random field (MCRF) theory was suggested. This new Markov chain theory extended a single Markov chain for multidimensional modeling. The measure for MCRFs is called 'transiogram' (i.e., *transi*tion probability dia*gram*), rather than the conventionally used 'transition probability matrix' (TPM). Transiogram solves the major technical problem in estimation of transition probabilities in many geographical applications of Markov models. It overcomes the limitation of TPM that it can only be used to provide estimates for time periods equal to the time interval of the input time series data or lag dis tances equal to the sampling interval of the input spatial data. For example, a 1981–85 land use change matrix can only provide quinquennial estimates. However, transio gram provides a way for estimating continuous transition probabilities (with any time or space separation) from samples. Transiogram also provides a visual spatial rela tionship measure for categorical data.

One-Dimensional Markov Chains

A Markov chain represents a sequence of random vari ables X_1, X_2, X_3, … with the Markov property. The Markov property means that given the present state, the future state is independent of past states. Here a 'state' refers to a category defined by users, such as a type of land use or a status (e.g., move, stay) of a resident. For a first order Markov chain, we have

$$\begin{aligned}&\Pr(X_{n+1} = x | X_n = x_n, \ldots, X_1 = x_1)\\&\quad = \Pr(X_{n+1} = x | X_n = x_n)\end{aligned} \qquad [1]$$

The possible values of X_i form a finite set of states S, called the state space of the chain. The description tool of a Markov chain is its TPM. For a stationary Markov chain, its TPM does not change with time (or space). Below is a TPM of a discrete time (or space), two state stationary first order Markov chain defined on the state space $S=(1, 2)$:

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \qquad [2]$$

This means that if the present state of the Markov chain is 1, it will have a probability of 0.7 moving to the same state (i.e., state 1) and a probability of 0.3 moving to state 2 at the next time (or space) step. If the present state of the Markov chain is 2, it will have a probability of 0.6 moving to the same state and a probabilty of 0.4 moving to state 1 at the next time (or space) step.

Stationary Markov chains have a property – after sufficient transition steps, their transition probabilities reach an unchangeable state and the different rows be come the same. Such a row of transition probabilities is called an equilibrium vector, where each element w_i represents the 'proportion' of the corresponding state i. The equilibrium vector can be obtained by multiplying the TPM with itself sufficient times until the resultant TPM has no changes in its element values.

The above TPM in eqn. [2] has diagonal entries (i.e., p_{11} and p_{22}), which means the Markov chain may move from a state to the same state (i.e., self transition). If the self transition probability of a Markov chain on state i is p_{ii}, the probability that the number of consecutive self transitions of the Markov chain on state i is k, which obeys the geometric distribution

$$P(T_i = k) = \theta_i(1 \quad \theta_i)^{k-1}, k = 1, 2, \ldots \qquad [3]$$

where $\theta_i = (1 - p_{ii})$. The geometric distribution is a dis crete analog of an exponential distribution. When using such a Markov chain model to simulate the spatial se quence of a phenomenon, the sequences of states, such as *AAABBBBAABBBBB*…, are generated, and the numbers of consecutive self transitions of a state such as *B* obey the geometric distribution.

The number of self transitions on a state represents the 'wait' time (or distance) of the Markov chain on the state. For studying land use change, this 'wait' time refers to the time (e.g., years) of the land to be continuously used in one way (e.g., cropping). In many cases, 'wait' time may not follow a geometric distribution. Thus, the above Markov chain model may not be suitable for all cases. An appropriate Markov chain model for cases where the geometric distribution cannot be applied may be an embedded Markov chain, which considers only transitions between different states in its TPM and deals with 'wait' time in a user defined way.

For a three state, stationary, embedded Markov chain, its TPM can be expressed as

$$Q = \begin{bmatrix} & q_{12} & q_{13} \\ q_{21} & & q_{23} \\ q_{31} & q_{32} & \end{bmatrix} \qquad [4]$$

In the above TPM, only transitions between different states are counted and self transitions are ignored. When using such a Markov chain model to simulate the spatial sequence of a phenomenon, only the sequences of dif ferent states, such as *ABCBACBACA*…, are generated. To describe the 'wait' time of each state, an appropriate probability distribution chosen by users, such as a log normal distribution, may be used. We found no literature in human geography that used the embedded Markov chain approach. But such a method may find its useful ness in the future because 'wait' time in some phenomena in human geography may not fit a geometric or ex ponential distribution.

Transiograms

A transiogram refers to a transition probability diagram (a curve or a series of points, of transition probabilities with increasing separate distance) for characterizing the autocorrelation of a class (or state or category) or the cross correlation from a class to another class. Limi tations in data collection make it impossible to estimate a continuous measure directly from sparse sample data. Transiogram overcomes the data limitation and can be used to detect and model spatial autocorrelation or cross correlation at different scales from sample data through inferring transition probability models. A transiogram can be represented as a transition (or conditional) probability function on a continuous lag $\mathbf{h}$:

$$p_{ij}(\mathbf{h}) = \Pr(z(\mathbf{x}+\mathbf{h}) = j | z(\mathbf{x}) = i) \qquad [5]$$

where z is a reliazation of the random z at a specific location $\mathbf{x}$. Here Z may be one dimensional or multi dimensional. The second order stationarity assumption is applicable here so that $p_{ij}(\mathbf{h})$ is dependent only on the lag $\mathbf{h}$, not on the location $\mathbf{x}$. An auto transiogram $p_{ij}(\mathbf{h})$ represents the self dependence (i.e., auto correlation) of a single class i and a cross transiogram $p_{ij}(\mathbf{h})$ $(i \neq j)$ rep resents the cross dependence of class j on class i. Here class i is called a head class and class j is called a tail classes. i and j are not interchangeable here because cross transiogram is normally asymmetric.

Transiograms directly estimated from sample data are called 'experimental transiograms'. Continuous transio gram models can be acquired by using mathematical models to fit experimental transiograms. Basic math ematical models for modeling experimental transiograms include the exponential model and the spherical model. Expert knowledge may be incorporated in the estimation of transiogram models. Here expert knowledge refers to the knowledge of experts in parameter estimation of transiogram models, which typically include sills, ranges, and model types (e.g., exponential, spherical). **Figure 1** shows that an experimental autotransiogram and an experimental cross transiogram are approximately fitted by basic mathematical models. Here, the sill of a transiogram refers to the stable height that the transiogram gradually approaches. Theoretically, the sill of a transiogram should be equal to the proportion of the corresponding tail class. The range of a transiogram refers to the lag distance where the transiogram approaches the sill. When the separate distance between two data is less than the range, they are considered spatially dependent (or correlated); otherwise, they may be considered independent.

Transiograms have several uses. One use is to estimate transition probabilities, particularly multistep transition probabilities, from insufficient data. When data are insufficient, transition probabilities estimated from data are inaccurate. Under this situation, reasonable transition probabilities can be obtained by using mathematical models to fit experimental transiograms; thus characteristics of a large set of geographic data (e.g., population) can be inferred using a small sample set. For example, the racial migration probabilities in a large city may be inferred from a small sample set by using transiograms.

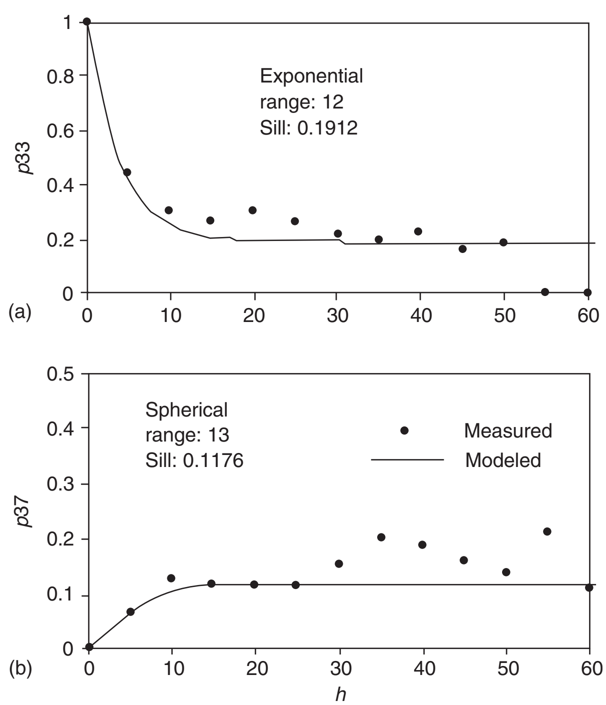

Figure 1 Transiogram modeling by mathematical models: (a) autotransiogram, (b) cross-transiogram. The scales along h axis are number of pixels.

The second use is to get continuous lag transition probabilities from sample data through modeling. Usually sample data can only provide us transition probabilities with one or several specific lags (e.g., one step or n step transition probabilities), which is a severe limitation when transition probabilities at other lags are needed. The third use of transiograms is to characterize spatial (or temporal) heterogeneity of discrete variables. Correlation ranges and sills of transiograms and transiogram shapes are all reflections of spatial heterogeneity. In addition, class polygon size (mean length) may also be inferred from autotransiograms. Thus, transiogram may be used to infer polygon sizes of different land use classes (e.g., residential, commercial, industry, park, and agriculture) or different urban planning district sizes with certain residential characteristics (low, medium, and high density). The fourth use of transiograms is to provide input transition probabilities to Markov chain simulation. Transiogram models can provide transition probabilities with any lag distance. The fifth use of transiograms is for data mining in large datasets. Through estimating transiograms from a large dataset of a categorical variable, one may find autocorrelation properties of single classes and complex relationships between different classes hid in the dataset. Thus, it may be used to determine if the magnitude or frequency level of some phenomena differ from one location to another. Therefore, it is expected that transiograms will be a useful spatial statistical tool in human geography to make generalizations concerning complex spatial patterns for studies of diseases, urban growth, land use change, residential mobility, socio-spatial segregation, and vehicle traffic, etc.

Markov Chain Random Fields

The MCRF theory extends a single Markov chain for any dimensional modeling. The general solution of the conditional probability distribution of a MCRF Z at an unsampled location x was derived as

$$\Pr(z(\mathbf{x}) = k | z(x_1) = l_1, \ldots, z(\mathbf{x_m}) = l_m) = \frac{\prod_{i=2}^{m} p_{kl_i}(\mathbf{h_i}) \cdot p_{l_1 k}(\mathbf{h_1})}{\sum_{f=1}^{n} \left[\prod_{i=2}^{m} p_{fl_i}(\mathbf{h_i}) \cdot p_{l_1 f}(\mathbf{h_1})\right]} \quad [6]$$

where $p_{kl_i}(\mathbf{h}_i)$ represents a transition probability in the ith direction from state k to state l_i with a lag $\mathbf{h}_i$; $\mathbf{x}_1$ represents the neighbor from which the Markov chain moves to the current location x; m represents the number of nearest known neighbors (or locations); k, l_i, and f represent states in the state space $S=(1, \ldots, n)$; $\mathbf{h}_i$ is the distance from the current location to its nearest known neighbor $\mathbf{x}_i$. With increasing lag h, any $p_{kl}(\mathbf{h})$ forms a transiogram which represents spatial (auto or cross) correlation of classes. It can be seen that the conditional probability distribution equation of the MCRF is actually composed of transiograms. With changes of the number m and the directions of nearest known neighbors, the above general solution actually includes a set of different MCRF models for one dimensional and multidimen sional modeling. It is clear that MCRF models are nonlinear.

In practical use, however, the above general solution cannot be simply implemented directly, because it is necessary to consider a limited number of nearest known neighbors and the conditional independence of nearest known neighbors in cardinal directions for optimal simulations. Normally, considering four orthogonal car dinal directions is sufficient for two dimensional mod eling. So if only the nearest data locations in four cardinal directions are considered, the MCRF model in eqn. [6] is simplified as

$$\Pr(z(\mathbf{x}) = k | z(\mathbf{x}_1) = l, z(\mathbf{x}_2) = m, z(\mathbf{x}_3) = q, z(\mathbf{x}_4) = o) \\ = \frac{p_{ko}(\mathbf{h}_4) \cdot p_{kq}(\mathbf{h}_3) \cdot p_{km}(\mathbf{h}_2) \cdot p_{lk}(\mathbf{h}_1)}{\sum_{j=1}^{n} \left[p_{fo}(\mathbf{h}_4) \cdot p_{fq}(\mathbf{h}_3) \cdot p_{fm}(\mathbf{h}_2) \cdot p_{lf}(\mathbf{h}_1) \right]} \qquad [7]$$

where 1, 2, 3, and 4 represent the four cardinal directions considered. In directions 2, 3, and 4, transitions are from the current unknown location $\mathbf{x}$ to its nearest known neighbors, but in direction 1 (i.e., the coming direction of the Markov chain), the transition is from the nearest known neighbor $\mathbf{x}_1$ to the current location $\mathbf{x}$.

In simulation, a cardinal direction is replaced by a search sector to cover the whole search area (usually a circle) (**Figure 2**). Thus, four search sectors that equally split a search circle may be used to search for the nearest known neighbors, one from each search sector, for esti mating the conditional probability distribution of a random variable at an unsampled location. In addition, because the nearest known neighbors found within a search radius may not always reach four, the needed MCRF models may be further simplified from eqn. [7]. A Markov chain sequential simulation algorithm can be used to conduct simulation.

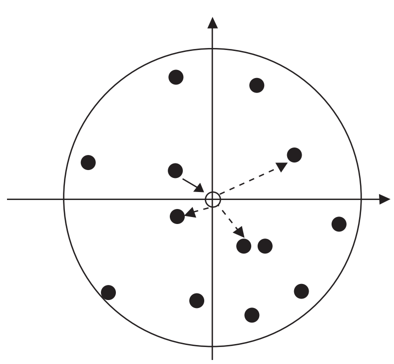

Figure 2 The searching sectors and neighborhood used in the random-path Markov chain sequential simulation algorithm.

Figure 3 shows simulated results of seven land cover classes using a random path sequential simulation algo rithm. The simulation was conditioned on a random sample set of 130 point data in a 35 km^2 area (a 295 by 295 lattice). Apparently, the patterns in simulated realizations are polygonal. Polygonal patterns are in accordance with the custom of area class mapping and are also convenient for human understanding and data processing using GIS tools. Different realizations are imitative of each other but with apparent differences in details, which demonstrates the different possible con figurations of the land cover classes in reality. The op timal map based on maximum occurrence probabilities represents the best estimate based on samples. When samples are very sparse or simulation is unconditional, sills of transiogram models will have strong influence on the proportions of classes in simulated results. But with the number of conditioning samples increasing, the proportions of classes in the conditioning sample set will play a major role in determining simulated class proportions.

Although single realizations can represent spatial uncertainty by their differences in spatial distribution of classes, a more accurate and vivid way to quantify and demonstrate spatial uncertainty is using occurrence probability maps, which can be estimated from a large number of realizations or directly calculated. As shown in **Figure 4**, the maximum occurrence probability map represents the purity (or quality) of the optimal map, and occurrence probability maps of single classes represent uncertainty of spatial distribution of each class. In the maximum occurrence probability map, the white–gray stripes actually indicate the approximate location of class boundaries, which are called 'transition zones'.

MCRF models have been used to model categorical geographical variables and may also be extended to model continuous geographical variables. It is expected that MCRF models will be useful in varied branches of human geography, particularly those involving the spread of disease (epidemiology), the practice of commerce plan ning, and exploration of traffic patterns in an urban core. However, how to use them and exactly where to use them are issues for human geographers to explore. For example, if one wants to map the spatial distribution of apartment rents or land prices in a city from a sample dataset, one may consider using two dimensional MCRF models.

Issues

One dimensional Markov chains have been used for decades in human geography and were proved useful in describing and modeling geographical phenomena and processes, such as patterns of human migration, repro ductive behavior, and intergenerational occupational

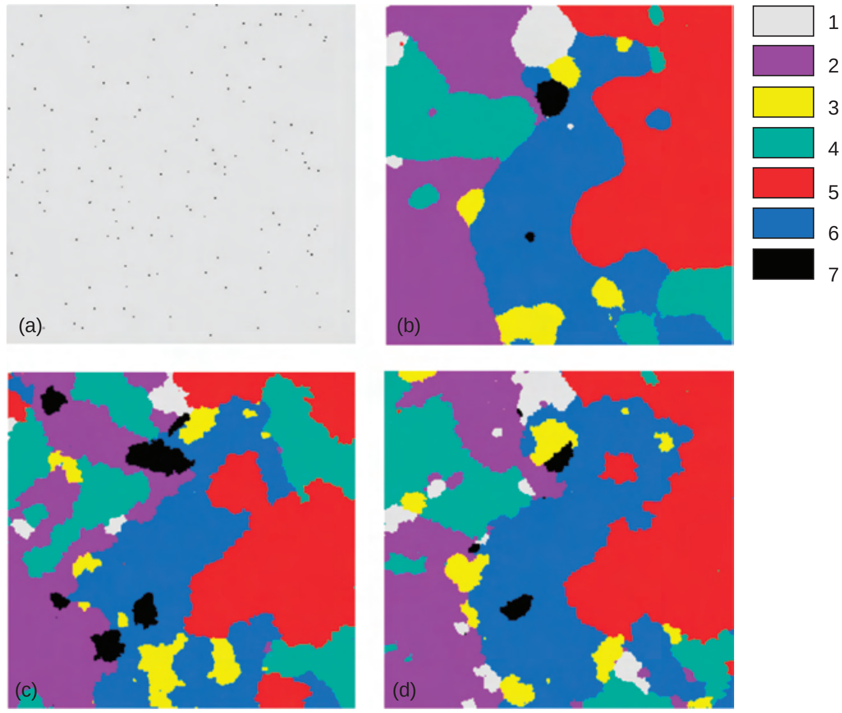

Figure 3 Simulated results using a random-path Markov chain geostatistical algorithm, conditioned on a random sample set of 130 points: (a) samples; (b) optimal map; (c) and (d) realizations.

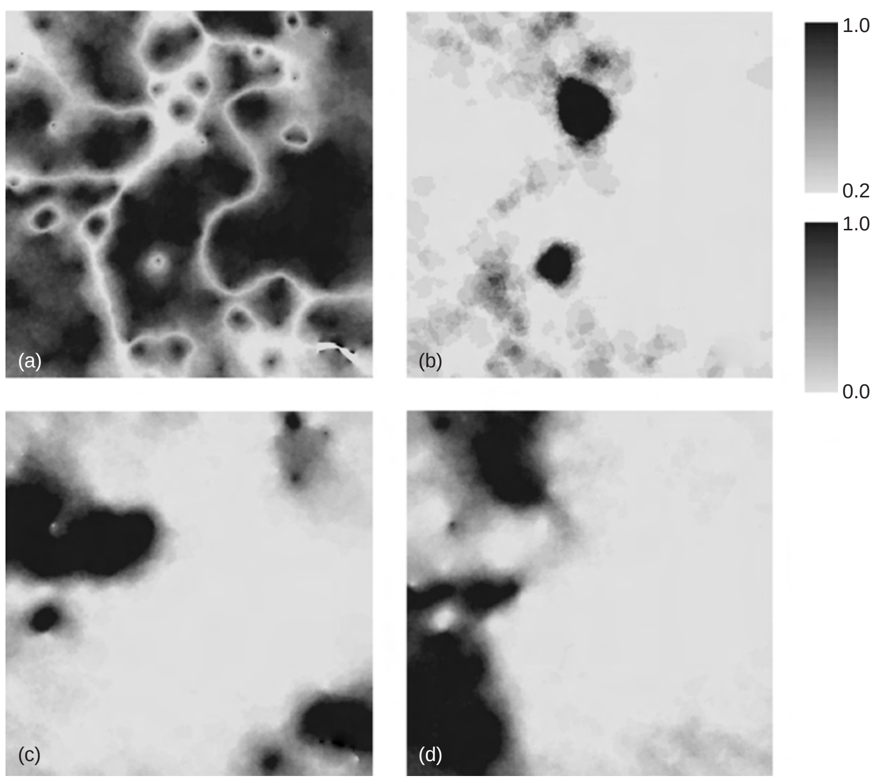

Figure 4 Occurrence probability maps generated using a random-path Markov chain geostatistical algorithm, conditioned on a random sample set of 130 points (correspond to **Figure 3**): (a) maximum occurrence probability map; (b), (c), and (d) occurrence probability maps of classes 7, 4, and 2, respectively.

mobility. However, it is still difficult to build high order Markov chain models to deal with the complex property of social phenomena or processes.

Transiogram provides a way for estimating continuous transition probabilities from sparse samples and expert knowledge and for spatial heterogeneity characterization of categories. Considering the previous applications of TPM in human geography and the advantages of tran siogram over TPM, it is expected that transiogram will also be a useful tool in human geography for measuring spatiotemporal changes. However, transiogram is just a two point spatial measure. For measuring complex geo graphical phenomena and processes, it is obvious that multipoint statistics are needed.

There is a need for methods of analyzing large scale spatiotemporal dependence in geographical studies be cause strong spatiotemporal dependence or autocorrela tion exists in areally distributed variables. Because MCRF theory extended a single Markov chain for multidimen sional simulation, it should be useful in dealing with multidimensional human geographical phenomena and processes, such as describing and predicting changes in land use, population distribution, residential structure, transport networks, and industrial structure and pattern. While many geographical phenomena and processes in volve both spatial and temporal dependences, integrating the time dimension and space dimensions into the same MCRF model will be desirable.

See also: First Law of Geography; Kriging and Variogram Models; Monte Carlo Simulation; Simulation; Spatial Interpolation; Uncertainty.

Further Reading

Brown, L. A. (1970). On the use of Markov chains in movement research. *Economic Geography* 46, 393 403.

Clark, W. A. V. (1965). Markov chain analysis in geography: An application to the movement of rental housing areas. *Annals of the Association of American Geographers* 55, 351 359.

Collins, L., Drewett, R. and Ferguson, R. (1974). Markov models in geography. *The Statistician* 23, 179 209.

Li, W. (2007). Transiograms for characterizing spatial variability of soil classes. *Soil Science Society of America Journal* 71(3), 881 893.

Li, W. and Zhang, C. (2007). A random path Markov chain algorithm for simulating categorical soil variables from random point samples. *Soil Science Society of America Journal* 71, 656 668.

Richardson, H. W. (1973). A Markov chain model of interregional savings and capital growth. *Journal of Regional Science* 13, 17 27.

Tang, J., Wang, L. and Yao, Z. (2007). Spatio temporal urban landscape change analysis using the Markov chain model and a modified genetic algorithm. *International Journal of Remote Sensing* 28, 3255 3271.

Relevant Websites

http://en.wikipedia.org/wiki/Markov chain
Classical Markov chains.

Marxism/Marxist Geography I

A. Cumbers, University of Glasgow, Glasgow, UK

Glossary

Capitalism A particular mode of production, dominant since the eighteenth century, based around the private ownership of the means of production and the related need for people to sell their own labor power to make a living.

Commodity Fetishism The term used critically by Marxists to describe mainstream economists' failure to acknowledge the social relations (and exploitation) that underlie the production of commodities.

Dialectics and Class Struggle The form that social relations take under a Marxist perspective, whereby society is composed of two fundamentally opposing classes (exploiter and exploited). Through class struggle, society is transformed through time from one mode of production to another (e.g., feudalism to capitalism, capitalism to communism, etc.) as the oppressed class (e.g., workers) eventually overthrows and liberates itself from the oppressor (e.g., the capitalist class).

Spatial Fix The geographical stabilization of capitalist production through particular organizational forms and arrangements for periods of time necessary to secure accumulation and surplus value.

Surplus Value The difference between what labor is paid in production and the total return extracted by the capitalist, which, according to Marxists, represents the real exploitation of labor.

Uneven Development This is the tendency under capitalism for some places to grow very rapidly while others experience decline. The two countervailing tendencies – of growth and decline – are fundamentally related and intrinsic to capitalist forms of production.

Introduction

The adoption of Marxism as a theoretical perspective and a political project in the 1970s was the single most important development in the evolution of a critical human geography. Its introduction into the discipline was closely associated with David Harvey, who had hitherto been a leading figure in the development of a more scientific approach to geography in the 1960s. Marxism (and a radical political economy perspective) dominated critical geographic thoughts during the 1970s and 1980s because of the insights it provided into the way social and spatial inequalities are underpinned by class struggle and conflict. Following the writings of Karl Marx (1818–83), a Marxist perspective perceives capitalist society to be divided into two opposing classes: the capitalists who own the means of production (e.g., land, factories, equipment, etc); and the mass of workers who own nothing beyond the ability to sell their own labor. The category of 'free labor' therefore distinguishes capitalism from earlier class based societies (e.g., feudalism, slavery) where the mass of workers was completely in thrall to an elite. While capitalism is more progressive than earlier societies, it nevertheless leads to an exploitative set of relations between capital and labor. Capitalists, in competition with each other, are forced to increase their profits through the production process. Workers, having no independent means to earn a living, feed, and clothe themselves and their families, have to sell their 'labor power' to capitalists to receive an income.

Subsequently, Marxism been criticized for its economic reductionism, determinism, and neglect of other dimensions of social relations (e.g., gender, ethnicity, cultural, and national identity). Its influence on geographers has also declined with the growing distrust of 'grand narratives' and the emergence of poststructural and postmodern discourses that seem more in keeping with geography's traditional interest in spatial diversity and difference. Nevertheless, while the mainstream of human geography may have shifted away from Marxism, it remains pivotal to a critical human geography. For those committed to alternative perspectives that challenge uneven power relations and dominant narratives, Marxism remains an important point of departure.

Marxism has been particularly influential in geography because it offers a detailed theoretical perspective on the tendency for the economy to produce patterns of geographically uneven development. This is in direct opposition to neoclassical economics (NCE), which became a key influence in human geography, particularly in the development of economic geography and regional science, in the 1950s and 1960s. NCE assumes that spatial inequalities are eradicated by the operation of market forces, if left to operate freely without regulation or interference, ensuring the optimal distribution of economic wealth. Adam Smith, the father of modern economics, referred to these forces as the "invisible hand" and subsequently, mainstream economics has held the view that the economy is governed by self regulating markets where the forces of demand and supply work to eradicate spatial inequalities over time. Differences are eradicated and the economy will return to a spatial equilibrium. In opposition to this, Marx and other nonmainstream economists (such as Joseph Schumpeter and John Maynard Keynes) developed a very different conception of the economy as being fundamentally unstable and prone to periodic crises due to its own contradictions. Market forces do not naturally adjust to imbalances between demand and supply, but often allow these to persist and worsen in time without some form of intervention or regulation. Spatially, these same tendencies will produce patterns of inequality between places which may change over time, but, critically, the spatial logic is one of unfolding processes of uneven development rather than the equalization of difference in the long run.

Marx's approach is first and foremost a historical one, resulting in the use of the term historical geographical materialism, as argued, alongside many others, by Erik Swyngedouw in 2000. Marxist geographers view economic development in terms of changing forms of social and political organization in time and space, characterized by different forms of class struggle. Capitalism as a particular 'mode of production' emerged out of a feudal society in Western Europe from the fifteenth century onward, and over the past 200 years has spread inexorably to the point that one can now talk of a global capitalist economy.

Fundamentals of a Marxist Approach

For Marx, capitalist accumulation is underpinned by the tension between capital (employers) and labor (employees). Marx was the first to recognize the centrality of labor, and the importance of the 'labor process', to the operation of capitalism. Marx recognized that human labor is fundamentally different to other commodities as the source of surplus value or profit, which in turn drives the process of economic development (referred to as capital accumulation) under capitalism. A key observation was that the surplus value accruing to the holders of capital from the manufacture and sale of commodities was based upon the exploitation of labor. Surplus value is created by labor which Marx refers to as 'variable capital'. Other factors of production (e.g., land, machines and raw materials) are all fixed, in the sense that the capitalist pays a price that is then a fixed cost in the production process. Labor, however, is not fixed. The capitalist pays a price for labor (wages) but can then extract more effort from labor in the course of producing commodities than what has been paid. This gives rise to surplus value, which operates either through the capitalist making labor work for longer than the real exchange rate of labor (absolute surplus value) or by making labor more productive, usually through reorganization of production through management or technical innovation (relative surplus value). Hence, surplus value is not the 'just' return to entrepreneurs for risk taking, as classical economists prior to Marx believed, but is instead the result of a fundamentally unequal exchange in the labor market. How does this exploitation come about? In a brilliant uncovering of the inner workings of the capitalist economy, Marx elaborates his theory in the pages of his major work *Capital*. Here, we will sum up the main points.

Marxism and Dialectical Reasoning

An important starting point in Marxist analysis is the dialectical method. At the most basic level, a dialectical perspective emphasizes the importance of processes and relationships for interpreting the world, rather than fixed things or categories. Historical change is driven by the tensions between opposing forces, usually represented in the form thesis–antithesis–synthesis, where thesis represents the original position, the antithesis its negative relation, and the synthesis the resolution of the contradiction. Mapping this onto Marx's understanding of social change, this is most obviously expressed in the *Communist Manifesto*, co written with his long time comrade and collaborator Friedrich Engels:

> The history of all hitherto existing society is the history of class struggles. Freeman and slave, patrician and plebeian, lord and serf, guild master and journeyman, in a word, oppressor and oppressed, stood in constant opposition to one another … The modern bourgeois society that has sprouted from the ruins of feudal society has not done away with class antagonisms. It has simplified the class antagonisms. It has but established new classes, new conditions of oppressions, new forms of struggle in place of the old ones. (Marx and Engels, 1967: 79, 80)

Throughout history, social change has been driven by the battle between opposing classes, the dominant and the dominated, with the subservient class eventually overpowering the ruling class to create new conditions of society. Many critics have accused Marx of teleological thinking in this respect, suggesting that his key dialectical insight is the inevitability of the working class or proletariat overpowering the bourgeoisie or capitalist class to create a utopian communist society. There are, however, many different interpretations of Marx; commonly, some, such as Richard Peet in his 1998 textbook for geographers, make a distinction between the early philosophical Marx and the later political economy work. It is in the latter that arguably his most important contribution lies, as an analyst of capitalism and its contradictions, rather than as an architect of communism.

Using dialectics, Marx provides new insights into the way opposing class relations shape the economy and the process of commodity production. In his most sophisticated and powerful analysis of capitalism, *Capital*

(Volume 1), Marx starts off with a long discussion of the commodity form and what it represents. In capitalist societies, commodities are the useful products of labor that are bought and sold. These include objects as diverse as an automobile, banana, or haircut. Like the classical political economists, such as Adam Smith and David Ricardo, before him, but in contrast to mainstream economists from the mid nineteenth century onward, Marx based his analysis of commodities on the 'Labour Theory of Value'. The value of a commodity from this perspective is related to the labor needed to produce it. However, where Smith and Ricardo related the value of labor to the work of individuals, Marx started from the perspective of society as a whole, introducing the concept of 'abstract labor'. For the classical economists, the wages of labor reflect the true value of the labor supplied. For Marx, the wages of labor do not reflect the real value of labor, but instead reflect the value of the total amount of labor needed to produce a commodity ('abstract labor') minus the degree of exploitation of labor by capital. The key point here is that the production of commodities is underpinned by social relations, representing the class struggle between capital and labor. It is not the free and equal exchange associated with a mainstream economic perspective on markets.

Use Value and Exchange Value in Commodity Production

Out of this basic premise, Marx develops the concepts of 'use value' and 'exchange value' symbolizing the twofold character of all commodities. Use values signal the practical value of commodities – a shoe can be worn, bread eaten, etc. The exchange value, in contrast, reflects the value of a commodity in exchange for another commodity, that is, its market price. This twofold character of commodities becomes evident in what is known as the 'circuit of capital', where the exchange value of a commodity (C) is first realized in exchange for money (M). However, when that money is exchanged for another commodity (C), exchange value turns back into use value (**Figure 1**). Where Marx differs, in relation to Smith and Ricardo, is in seeing this circuit as underpinned by social relations (and power). He is famously scathing of the classical political economists (CPE) for viewing exchange value as the value between things rather than about relations between people, something that has become known as 'commodity fetishism'. He notes:

> Man's reflections on the forms of social life, and consequently, also, his scientific analysis of those forms, take a course directly opposite to that of their actual historical development. He begins, *post festum*, with the results of the process of development ready to hand before him. The characters that stamp products as commodities, and whose establishment is a necessary preliminary to the circulation of commodities, have already acquired the stability of natural, self understood forms of social life, before man seeks to decipher, not their historical character, for in his eyes, they are immutable, but their meaning. Consequently, it was the analysis of the price of commodities that alone led to the determination of the magnitude of value, and it was the common expression of all commodities in money that alone led to the establishment of their characters as values. It is, however, just this ultimate money form of the world of commodities that actually conceals, instead of disclosing, the social character of private labour, and the social relations between the individual producers. (Marx, 1906: 87)

C – M – C

Figure 1 The circuit of capital.

In its simplest sense, we can think of exchange value as the perspective of capital and use value as the perspective of labor. From labor's point of view, commodities represent the need to consume in order to achieve basic social reproduction (e.g., food, clothing, housing, energy, etc.). For capitalists, their value is in terms of what they can be sold for to create profit, which can, in turn, be reinvested. When thought of in these terms, there is clearly an underlying tension between use value and exchange value. A concrete example of expressing this would be in the tensions between economic growth and environmental sustainability. Capital's logic is to encourage greater and greater consumption, production, and exploitation of resources in the pursuit of expanded surplus value, whereas the logic from the point of view of labor (and the planet) is to continue to reproduce the conditions for human survival. Production for value is fundamentally at odds with creating the sustainable conditions for human development.

In deepening his analysis, Marx thinks of labor in two opposing senses: as socially useful labor and as 'abstract labor'. Socially useful labor or work exists in all human societies throughout history, whatever the forms of social relations (e.g., slavery, feudalism, etc.). All societies, as a necessary condition of human existence, have to work: basic tasks, such as obtaining food, making clothes, constructing shelter, and producing forms of usable energy, have to be performed to ensure the reproduction of the species. In even the most primitive human societies, simple efficiency savings give rise to a division of labor between different people, often on a gender basis, which becomes more complex and hierarchical as societies become more advanced. Specific to the mode of production that is capitalism, however, human labor becomes geared toward commodity production for exchange value or 'abstract labor'. Under capitalism, all work becomes

$$M \ldots C \left\{ \begin{array}{l} MP \\ \\ \qquad P \ldots C^{*} \ldots M + \Delta \\ \\ LP \end{array} \right.$$

Figure 2 The process of production within the circuit of capital. From Castree, N., Coe, N., Ward, K. and Samers, M. (2003). *Spaces of Work: Global Capitalism and Geographies of Labour*, p 28. London: Sage.

subservient to producing surplus value and "only has meaning and only appears as a social relation when it is embodied in a product that is exchanged" (Cleaver, 2000: 111). In the process, the individual worker becomes alienated from the product of his/her labor.

A fundamental difference between Marx and the CPE tradition is the way Marx reveals the process of commodity production as being both riddled with contradictions and underpinned by antagonistic class relations. It is through the exploitation of labor in the production process that capital deepens and sustains surplus value in the circuit of capital (**Figure 2**). Capital in its money form (M) is transformed into commodity form (C) by purchasing the means of production (MP) – factories, machines, materials, etc. – and labor power (LP). The means of production (MP) and labor power (LP) are then combined in the production process, under the supervision of the owners of capital or their managers and representatives, to produce a commodity for sale (C′) – for example, a car, house or, even, a haircut. This commodity is sold for the initial money outlay plus a profit, or surplus value.

The Labor Process

Following on from this, capital needs not just to purchase labor but also to control it, achieved through the labor process, if it is to increase surplus value. Hence, the division of labor under capitalism is not geared primarily toward the efficient organization of work to improve the general wealth of society (although this may occur at certain points in time and space) but is intrinsically a function of the imperative to create the conditions to realize surplus value. Opportunities to expand surplus value also exist within the production process itself and its associated labor process. To increase surplus value, capital needs either to reduce the costs of production or increase the price of the final product – the commodity for sale. In highly competitive markets, it is clearly more difficult to do the latter, so attention then falls upon reducing costs and raising the productivity of the means of production. Whereas mainstream economics largely ignores what happens in production – when labor is paid a wage, this is assumed to reflect its market value and this is the end of the matter – for Marxists what happens in production and in the labor process is critical to understanding the basic conflict between capital and labor. Because of the importance of the labor process in extracting surplus value, capital will always be driven to increase labor productivity and extract more value from labor, which is in a fundamental conflict with labor's need to earn enough to make a decent living. There is, therefore, an imperative for the capitalist to take control over the labor process from the worker, introducing forms of management which allow work to be organized on more capitalistic lines.

The Historical Development of Capitalism and Long-Turn Tendencies toward Crisis

Struggle for control over the labor process becomes a critical dimension of power relations and class struggle in the capitalist economy. For Marx and Engels, the emergence of the factory system in the late eighteenth and early nineteenth centuries and the accumulation of workers under one roof and under managerial control – and away from the household where early industrial craft work was organized according to noncapitalist principles – was a critical moment in the transformation of the capitalist labor process. A second critical transformation occurred later in the nineteenth century when employers began to pursue a more scientific approach toward the organization of work, observing and closely supervising job tasks with the view to developing more productive systems. This led to an increased 'technical division of labor' with the realization that productivity could be increased if work became increasingly fragmented into its component tasks allowing workers to become more specialized and proficient. Scientific management or Taylorism – after its most famous advocate, the engineer, F. W. Taylor – also encouraged greater experimentation with new techniques to increase productivity, the most famous being the introduction of the moving assembly line by Henry Ford at his automobile factory in Dearborn, Michigan in 1911.

Alongside the exploitative nature of capital–labor relationships, the other key dynamic process driving the capitalist economy is competition between capitals (firms). In Marx's view, the competition between firms is fierce and over time leads to a Darwinian survival of the fittest. Weaker and less competitive firms are driven out of the market or acquired by stronger rivals leading to an increasing concentration of economic power and the growth of larger and larger firms. Later Marxist theorists, such as the US economists Paul Baran and Paul Sweezy, writing in the mid twentieth century and documenting the rise of the large modern corporation, coined the term 'monopoly capitalism', for an advanced stage of capitalism in which the economy becomes increasingly dominated by a small number of large firms that exercise enormous market power.

For the system as a whole however, intensified competition over the longer term can lead to overproduction and a fall in the rate of profit/value as markets becomes saturated. Overproduction can lead to economic crises as overcapacity results in firms cutting back on employment, which in turn generates unemployment and a collapse in demand for products for the simple reason that workers are also the main consumers of capitalism. This situation can very quickly turn into a vicious downward spiral as firms' attempts to cut costs and reduce capacity further in response to a reduction in the demand for their products. This combination of circumstances was exactly what happened in the 1930s in the Great Depression when a crisis of overproduction led to the Wall Street Crash in the US. Because the US had experienced a major boom in the 1920s, pulling in investment from Europe in particular, the effects of the crash and the subsequent bankruptcy of thousands of firms led to a worldwide economic downturn and mass unemployment in many countries. Indeed, the Depression has been blamed as a critical factor in the rise of Nazism in Germany as the business classes sought refuge from the threat of Communism in Fascism. Only state intervention, taking different forms in different countries, and the increased spending on military production in the lead up to World War II injected some much needed demand back into the major economies to stimulate growth.

Key Schisms in Marxist thought and Practice

Revolution versus Reformism

In the century and a half since Marx was writing there has been considerable debate within and outside Marxism about the interpretation of his ideas and their continuing relevance. There is not the space here to do justice to all the various controversies, so we will highlight two of the most important dividing lines. In doing so, it is important to emphasize the evolution of Marxist thought as political praxis, that is, in being reformulated in relation to broader social and political events. One of the earliest schisms within Marxism, which continues to have important political and ideological ramifications to this day, is between a broadly defined social democratic or reformist position and a more revolutionary perspective. The schism emerged in what was known as the 'Second International' (of the international working class movement) at the turn of the nineteenth and in the early part of the twentieth centuries and was related to the thorny issue of how socialism might come about as industrial capitalism gathered apace.

These tensions were most evident within the German Social Democratic Party (SDP), which was the largest and most powerful Marxist inspired political party in the world in the period immediately prior to, and just after, World War I. The party was split between revolutionary theorists such as Rosa Luxemburg and Karl Liebnecht who argued for overt and violent class struggle to overthrow the capitalist system, and a range of more reformist positions, taken by intellectuals such as Karl Kautsky and Eduard Bernstein, who argued for struggle and social change within capitalism. Behind such political differences are different theoretical understandings of the relations between classes, the role of the state under capitalism, imperialism, and the relationship between political practice and broader economic forces. They were to lead to a split within the SDP and the foundation of the Communist Party in 1918. Outside Germany, the revolutionary position gained ascendancy in Russia, leading to the Bolshevik Revolution in 1917. Theoretically, however, both positions have been accused of economic determinism in their reading of Marx, in the sense that economic crisis within capitalism is seen as inevitable, because of its own contradictions, and the political question is about the tactics needed to prepare the working class for the coming struggle.

Structure versus Agency

The second key divide in Marxist thinking relates to this relationship between structure and agency, between those who emphasize the primacy of underlying structures and those who emphasize the autonomy of human action. Again, these are complex debates and, in the interests of brevity, we will try to summarize the key differences. At one end of the spectrum are those such as the French Community Party theorist, Louis Althusser, who, writing in the 1960s in reaction to the more humanist Marxism of Sartre on the one hand, and to the totalitarian nature of the Soviet Union and its crude version of Marxism on the other hand, tried to reformulate Marxism to deal with issues of ideology and culture. He rejected the economic determinism of earlier Marxists and recognized that key institutions in society (e.g., the state, church, and media) have some autonomy from capitalist processes. The role of ideology is critical for Althusser in shaping human behavior so that, while individuals may feel that they are acting of their own free will, the institutions of capitalism ultimately constrain individual agency to the system's requirements. Althusser's base superstructure distinction is important here where the base relates to the forces of production (underlying class relations) and the superstructure or relations of production, reproduced through institutions such as the state. Ultimately, in the last instance, the base has causal power.

In opposition to this structuralist perspective, critics such as E.P. Thompson make the point that the Althus serians ultimately strip the working class of any agency to resist and overthrow capitalism because everyone is sub jugated by powerful underlying forces. Moreover, Althus ser's attempt to create theoretically sophisticated, yet highly abstract, Marxist concepts served the purpose of reifying philosophy over politics while also separating out theory from concrete empirical events. Thompson himself castigated the Althusserians in his famous 1978 polemic *The Poverty of Theory*, arguing forcefully for a Marxism that remains rooted in the everyday consciousness of existing social relations. In a key passage, he notes that:

> Experience arises spontaneously within social being, but it does not arise without thought; it arises because men and women are rational, and they think about what is happening to them and their world … What we mean is that changes taking place within social being, would give rise to changed experience; and this experience is de termining, in the sense that it exerts pressure upon social consciousness, proposes new questions, and affords much of the material which the more elaborate intellectual exercises are about. (Thompson, 1978: 200)

In other words, working class people become active subjects in their own right. Moreover, class based agency springs directly from material and cultural experiences and is always contingent upon the particular circum stances and processes that it is subjected to. This means that class itself (and by implication, class struggle) does not take on a fixed, ahistorical form but should be understood in more dynamic and fluid terms, reflecting the historical (and spatial) context of capitalism.

The point to be made here is that many of the sub sequent critiques of Marxist work in geography and elsewhere address a particular structuralist or Althus serian reading of Marx, notably James Duncan and David Ley writing in 1982, whereas Marx can be interpreted in very different terms as a more agency centered per spective with implications for how we think about space and place. In this respect, it has to be borne in mind that Marx's purpose in his more important work, *Capital*, is not primarily about developing a theoretical critique of capitalism in terms of iron laws and reified concepts, but to develop an alternative, political perspective – a weapon to put into the hands of the working class.

Culture versus Economy

A related debate has long raged between more orthodox Marxists (such as the Althusserians) and newer schools of thought over the relationships between the role of dis course or culture and the material or economic in the shaping of human and class consciousness. A dissatisfaction with the tradition handed down from Engels onward – of reading off class consciousness and identity simplistically from economic subject positions – led to a new wave of Marxist work, from the 1920s onward, that was concerned with the interrelationships between culture, class con sciousness, and political strategy. A range of diverse the orists, from Lukacs, Gramsci, the Frankfurt School of Adorno, Horkheimer, Marcuse, and latterly Habermas, grappled with the importance of culture, propaganda, and a range of other nonmaterial social strategies and practices to the operation of power and control in capitalist societies. Such accounts led to more sophisticated perspectives on the role of the state, the media and civil society institutions (such as the church, trade unions, etc.) in the maintenance of social order.

Perhaps the most important of these influences in geography is the Italian theorist, Antonio Gramsci, who, working from a Marxist perspective in the 1920s and 1930s, sought to understand how the mass revolutionary movements in Western Europe failed to follow the Soviet Union down the path of Communism, and in the case of Germany, Italy, and later, Spain and France, resulted in the emergence of Fascist dictatorships. Gramsci's key contribution was the recognition that ruling classes ex ercise control over the mass of the population through subtle forms of consent, through which culture and ideology are used to persuade, rather than coerce, people into believing that the state, or elite groups, is acting in their best interests. Successful capitalist states are, therefore, able to achieve hegemony through co opting different classes into an alliance for long periods of time. Popular culture, newspapers, television, etc., are im portant weapons for elite actors to propagate certain common sense views that are important to maintain power, such as market values, the dubious link between democracy, freedom and capitalism, etc. For revolution aries such as Gramsci, counter hegemony needs to be constructed in the discursive and cultural realms, re ferred to as 'war of maneuver', as well as the material and the economic, if the working class is to overthrow their subjugation by the bourgeoisie.

Fault-Lines within Geography

These debates have attracted variable attention from geographers, reflecting the different concerns that have exercised the discipline. There has been little written explicitly by geographers about the schism between revolution and reform, although there have been lively debates, from time to time, between those who have ar gued for policy intervention within the capitalist system and those, such as David Harvey, who have argued for a more revolutionary geography. Debates concerning structure–agency relations have been more extensive and profound for the development of the discipline. A recurring set of themes surrounds the implications of a Marxist view of society, predicated on underlying social structures in shaping human agency, for geographers' traditional concerns with areal differentiation and regional cultures. While the latter would imply a sensitivity to the uniqueness of place and a more open ended, multifaceted sense of agency, a more structuralist account would resonate more with the spatial science tradition of developing a more generalist set of laws and processes that constrain human behavior.

The 1980s was a period of particularly intensive theorizing and discussion over the relations between society and space, culture and economy, class and social identity, signified in the setting up of the journal *Environment and Planning D: Society and Space* in 1983. The edited collection *Social Relations* and *Spatial Structures* (Gregory and Urry 1985), which included many protagonists from geography (Harvey, Massey, Allan Pred, Richard Walker, etc.) and influential sociologists such as Anthony Giddens, Ray Pahl, John Urry, and Peter Saunders who were beginning to write about the role of space in shaping social and economic life, was a landmark text of the time. Structure–agency tensions within geography led many researchers to depart from more orthodox Marxist perspectives toward more middle ground positions through the embrace of critical realism (the work of Andrew Sayer in particular); structuration theory (e.g., the work of Allan Pred); and the French anthropologist, Pierre Bourdieu's, concept of habitus.

The Construction of a Marxist Geography

The Relevance of Marxism in Human Geography

Marxist ideas had little influence in human geography until the 1960s when a younger generation of geographers became increasingly frustrated at the inability of contemporary geographical theory to contribute to the pressing social problems of the day. Key issues, such as racial tensions in US cities, the Vietnam War (symbolizing the imperialism of US foreign policy), gender inequalities, and the rediscovery of poverty in inner city ghettos, were largely absent from geographical research, leading David Harvey to call for a revolution in geographic thought:

> The quantitative revolution has run its course, and diminishing marginal returns are apparently setting in … There is an ecological problem, an urban problem, an international trade problem, and yet we seem incapable of saying anything of depth or profundity about any of them. (Harvey, 1973: 128 29)

In response to these pressing social issues, a group of geographers in the US particularly sought to fashion a new radical geography. This movement began at Clark University in Massachusetts, where a group of postgraduate students, led by Richard Peet, launched *Antipode: A Radical Journal of Geography* in 1969. This radical new geography turned increasingly to political economy and Marxism for its intellectual foundations because it gave a coherent and appealingly holistic analysis about social relations and class that could be applied to a range of different settings.

While the writings of Marx provided only a few scattered comments and insights into the geography of capitalism, Marxist geographers, notably Harvey, sought to build on these by developing a distinctively Marxist analysis of geographical change. From this perspective, the human landscape is shaped by the conflict laden relationship between capital and labor, mediated by the state, providing a deeper and more political explanation for spatial variation than spatial science.

Uneven Geographical Development and the Spatial Fix

A key geographical insight from Marxism is the way that capitalism produces uneven development across space. Marx recognized uneven development to be a basic feature of capitalism in the way that certain places prospered and attracted investment at the expense of others. Uneven development can be seen at a variety of different scales: from the global, in the relations between North and South and the tendency for capital to flow to particular hotspots at the expense of other places (e.g., Southeast Asia during the 1980s); to the very local, in the way that cities become divided between rich suburbs and poor ghettoes or *favellas*. At the same time, however, capital accumulation processes are extremely dynamic so that a core region during one phase of growth may fall into decline in the next phase. Theory is unable to predict which regions will suffer this fate as outcomes are not predetermined but are ultimately the result of the interaction between technological changes and changing class relations, both within individual places and across space.

Harvey has made an important contribution to understanding the geographical dimensions of the capitalist economy in his book, *The Limits to Capital*, published in 1982. This is by far the most serious and sophisticated attempt to develop a geographical perspective on Marxist political economy. As Trevor Barnes has put it: "Harvey's contribution is in showing that one of the limits to capital, one of the obstacles to the generation of surplus value, is geography. Space and place are part of the furniture of the crises of capitalism" (Barnes, 2004: 409). Harvey showed that capital, in its abstract form as a process for creating surplus value, is characterized by two opposing tendencies: the need for mobility, on the one hand, to seek out more profitable locations than existing centers of production, and the need to fix investment in one place, on the other hand, for a sustained period – or what he refers to as the 'spatial fix' – to allow the realization of surplus value in the first place. There is, in this sense, a geographical dimension to the dialectical tensions running through capital. Thus, capital is never completely mobile, but has to put down roots in particular places to realize the conditions for continued accumulation. At the same time, some types of capital (e.g., multinational corporations, currency speculation, etc.) are clearly more mobile than others, which are more tied by social and cultural relations to particular places (e.g., family owned small firms com pared to the branch plants of multinational corporations). Despite this, capital (in the abstract sense) retains a spatial advantage over labor, which is more place bound by virtue of its need to secure the reproduction of itself through family and community ties.

Another major contribution has been the geo graphically nuanced theory of uneven development put forward by Neil Smith, in 1990. Processes of uneven development, according to Smith, are the result of a dialectic of spatial differentiation and equalization that is central to the logic of capitalism, transforming the complex mosaic of landscapes inherited from pre capit alist systems. Capital moves to areas that offer higher profits for investors, resulting in the economic develop ment of these areas. The geographical concentration of production in such locations results in differentiation as they experience rapid development while other regions are left behind. At the same time, the tendency toward equalization reflects the importance of expanding the market for commodities, implying a need to develop newly incorporated colonies and territories so as to generate the income to underpin consumption.

The process of economic development in a particular region tends to undermine its own foundations, leading to higher wages, rising land prices, lower unemployment, and the development of trade unions, reducing profit rates. In other regions, underdevelopment leads to low wages, high unemployment, and the absence of trade unions, creating a basis for profit that attracts capital investment. Over time, capital will 'see saw' from de veloped to underdeveloped areas, jumping between lo cations in its efforts to maintain profit levels. It is this movement of capital that creates patterns of uneven de velopment. Smith develops a theory of the production of space under capitalism which leads to the emergence of three primary geographical scales of economic and pol itical organization: the urban, the national, and the global. The dynamic nature of the uneven development process is most pronounced at the urban scale at which capital is most mobile, resulting in, for example, the rapid gentri fication (upgrading through the attraction of investment and new middle class residents) of previously declining inner city areas. Conversely, patterns of uneven devel opment exhibit most stability at the global scale where the divide between developed and developing counties remains as wide as ever, although East Asia has risen to the core of the world economy through sustained eco nomic growth since the 1960s.

The Spatial Divisions of Labor Approach

Another geographer who has been at the forefront of developing Marxist ideas in human geography is Doreen Massey. Massey's landmark text *Spatial Divisions of Labour* was one of the first attempts by an economic geographer to interpret the changing location of industry in a par ticular empirical context (Great Britain from the indus trial revolution through to the 1970s) in terms of a Marxist political economy approach. Her main contri bution was to link insights about geographically uneven development to the changing organizational structure of capitalism in the twentieth century, grounded in an understanding of the Marxist labor process. She argued that with the decline of traditional heavy industry and the increasing dominance of large corporations in most sectors of the economy, the economic geography of Britain in the twentieth century was becoming less characterized by sectoral differences between places (e.g., shipbuilding in Clydeside and Tyneside, coal mining in South Wales, and financial services in London) but in creasingly by occupational differences reflecting how different places were linked together in a spatial division of labor.

Though there is no predetermined outcome, and corporate strategies will vary according to aspects such as the technical requirements of different industries and the ownership structure of firms, Massey noted a growing tendency for corporations to separate out, geographically, the higher level jobs in areas like senior management and research and development from the lower level and more routine jobs such as the processing or final assembly of products. Through the emergence of multi plant cor porations, this division of labor takes on an explicitly spatial form, with companies locating the higher order functions in cities and regions where there are large pools of highly educated and well qualified workers (London and the Southeast), with lower order functions locating increasingly in those regions and places where costs (especially wage rates) are lowest (e.g., Northern England and Scotland). For Massey, this uneven development is an expression of the underlying social relations of capital ism: class relations within capitalism take on an inter regional form with some regions.

However, Massey's approach is not one sided. Places and regions are not mere passive recipients of corporate location decisions but over time and, reflecting the outcome of past waves or 'layers' of economic restructuring processes, develop particular regional cultures that will reflect the outcome of political and economic struggles between different groups. Thus a region such as South Wales developed a strong and militant culture in the early part of the twentieth century as a result of the development of a strong labor movement associated with the mining industry. This in turn can have broader geographical consequences. The emergence of strong labor movements and trade unions in Britain's industrial regions (such as South Wales, the North East of England, and the Central Belt of Scotland) were key to the rise of the Labour Party at a national level during the twentieth century and the social reforms (notably the setting up of a National Health Service) passed by the Labour Government at the end of World War II. Massey's reminder that places are the active site of class struggle also resonates with the work of social geographers who have been influenced by the insights of Marxism on class formation for understanding how social relations are shaped outside the workplace in the broader community. Thus, there is a body of work which explores how class relations interact with other forms of social relations (e.g., gender, race, and ethnicity) to produce and reproduce social divisions and inequalities (e.g., over housing, public services, and education) in particular places (e.g., the Nigel Thrift and Peter Williams collection in 1987).

The emergence of regions with strong trade unions, that are able to bargain for higher wages and improved working conditions, can also have repercussions for the future location decisions of corporations, encouraging them to seek out new locations where labor is less well organized. In the United States, this trend has become particularly evident since the late 1960s with manufacturing companies closing down plants in the more unionized older industrial heartlands of the north and east (increasingly referred to as the 'rustbelt') and setting up new plants in the south and west of the country (the 'sunbelt'), where the absence of an organized labor movement has allowed them to pay lower wage rates and reorganize working practices to the advantage of the management.

An associated trend has been an increase in the relocation of production overseas to lower wage locations in Third World countries. This development, ongoing since the 1960s in lower technology sectors such as clothing and textiles, but now spreading to more advanced manufacturing and even some service sectors, is sometimes termed a new international division of labor. While the recent geographical expansion of capitalism is inescapable, the mobility of individual firms is often overplayed. In some cases, firms use the threat of relocation to extract concessions over labor at existing locations, while in many other situations, cost is not the most important location factor and other factors such as the availability of skilled labor, sunk costs through past investment, locally specific industrial knowledge, are as likely to encourage firms to remain *in situ*.

Debates About the Relations between Marxism, Social Relations, and Space

Massey's emphasis upon the uniqueness and politics of place alongside broader spatial processes helped stimulate what became known as the 'localities' debate in the 1980s. Up to this point, the work of Harvey, Massey, Smith, and others were important in developing sophisticated accounts of how capitalism produces a space economy, but geographers were becoming increasingly concerned with the way that space itself is an active dimension in the production and reproduction of capitalist and class forms, rather than viewing social relations as primary. A central theme in the localities debate was the degree to which particular places are the site of complex social relations that are not easily read off from broader spatial and social processes, as Phil Cooke argued in 1986 – a direct challenge to the earlier Marxist accounts – but accused by some, such as Neil Smith in 1987, of leading away from a critical spatial theory and back to an earlier generation of empiricism.

Another approach that sought to give equal weighting to social and spatial relations was the body of work concerned with what came to be known as 'spatial dialectics', concerned with how social relations interact with space to create particular concrete and territorialized forms of capitalism. The work of Ed Soja and his 'sociospatial dialectic', initially outlined in 1980, was particularly influential in this regard in calling for "an extensive and flexible rethinking of both theory and practice, a reconstruction which will continue to draw upon the achievements of Marx but which must also be more directly attuned to the specificity of contemporary capitalist (and socialist) spatiality and temporality" (Soja, 1985: 91).

Soja insists on the need for Marxists and critical social theorists to more fully integrate space into their analysis of capitalism from the outset, rather than as an additional component that could be added when confronted with the more concrete outcomes of capitalist restructuring. He is particularly critical of Marxist theorists, such as Lefebvre and Mandel, who recognize the importance of space, but only in terms of a strategy or tactic through which capital deals with its own contradictions. Space in Soja's dialectical schema has three dimensions which overlap and interrelate: physical, mental, and social. Most obviously, the space of nature (physical) is made social through accumulation processes. Similarly, interpretation, cognition, and understanding are inherently spatial and are themselves produced out of spatialized social relations. Thus, it follows that "spatiality and temporality, human geography and human history, intersect in a complex social process which creates a constantly evolving historical sequence of spatialities, a spatio temporal structuration of social life which gives form not only to the grand moments of societal development but also to the recursive practices of day to day activity" (Soja, 1985: 94).

The Backlash against Marxist Ideas in Critical Human Geography

By the late 1980s, Marxist geography was coming under attack from new forms of social theorizing, primarily poststructuralism, postmodernism, and postcolonialism. The main focus of these critiques was Marxism's claims for a set of universal truths and concepts (e.g., capital, class) with which to understand economy and society. Grand theories or 'meta narratives' (big stories) claiming to uncover the changing organization of society are rejected as a product of the privileged position and authority of the observer rather than being accepted as objective representations of the realities that they purport to explain. Instead of functioning as a set of universal truths, knowledge should be regarded as partial and situated in particular places and times. Postmodernists in particular reject conventional notions of scientific rationality and progress, favoring an open interplay of multiple local knowledges; a perspective that has obvious appeal to geographers. Marxists were also criticized for being out of touch with the 'new times' of the 1980s, marked by the dominance of neoliberal ideas, particularly in the UK and US, and the collapse of the Soviet Union, which, however far it might be from Marx's ideal, was regarded as the standard bearer for communism in practice.

Three main criticisms were leveled at Marxist geography. In the first place, it was accused of neglecting agency in favor of being overly deterministic in its explanation. Secondly, those Marxist geographers who did take agency seriously were accused of reducing agency to class position rather than recognizing its multi dimensional nature. In this regard, some of the most trenchant criticisms came from those dissatisfied with Marxist geographers' neglect of other social categories, like gender and race, in shaping human agency. Harvey, in particular, was the subject of ferocious attacks by feminist geographers in the early 1990s. Massey, as we have seen a Marxist herself in her earlier incarnation, described Harvey's approach as a "steamroller of an analysis which insists that capital and labour … are all there is to it … It is absolutely stated that everything must be subsumed to a question of class … one's identity and the struggles we are engaged in are more multi faceted than Harvey's position is capable of conceiving" (Massey, 1991: 54–55). A third criticism related to the Marxist emphasis upon the determining role of economic forces and relations over culture and discourse, which was criticized for offering an – at best – partial understanding of social relations.

Although these are important and relevant criticisms, it is debatable whether Marxist geography was really as economically determinist as its critics allege. To take the most celebrated theorist, Harvey's work has mirrored that of Marx in some respects in moving between theoretical abstraction (in the *Limits*) and more empirical work where time, space, and place complicate a Marxist analysis, as underlying structural tendencies interact with grounded and contingent realities. Harvey has himself responded in a number of works that seek to engage with his critics, notably with contributions in 1989, 1996, and 2000, although he continues to be attacked for his materialist reductionism and adherence to a dialectical method. There is a danger that such debates become a 'dialog of the deaf rather' than a more productive engagement that recognizes the potential for a more nuanced, actor centered and politicized form of Marxism. Richard Walker makes a good point in this respect with his sympathetic critique of Harvey for trying to "'oversell' the logic of capital [as] the driving force behind almost everything" (Walker, 2004: 439).

Marxist Geography in the 1990s

In response to some of the criticisms outlined above, Marxist geography itself has developed considerably from the early 1990s onward and, in some senses, has actually retreated from the more abstract works of Harvey and Smith to more grounded geographical research. Researchers have also sought to respond to some of the charges of economic reductionism by filling in the gaps left by earlier analysis, particularly through work on the state, labor, and the environment.

An important body of work on the role of the state has been inspired by the French regulation school typified by the work of Aglietta, and developed in the more recent work of the political sociologist Bob Jessop. The Regulation Approach grew out of the 1970s and the productivity crisis that emerged in Western Europe and North America after two decades of economic growth centered upon the spread of mass production techniques (Fordism). Faced with a growing economic downturn, unemployment, and inflation, a group of economists at the Centre for Mathematical Economic Forecasting Studies Applied to Planning (CEPREMAP) became frustrated at the paucity of both mainstream and Marxist approaches in explaining the crisis, and instead, began to explore the way the economy is 'socially regulated'. Focus turned to the way particular social customs, norms, and practices play an important part in shaping economic life to create stable forms over the longer term. Wider processes of regulation find expression in specific institutional arrangements which mediate and manage the underlying contradictions of the capitalist system, expressed in the form of periodic crises, enabling renewed growth to occur. This occurs through the coming together and consolidation of specific modes of regulation, referring to the institutions and conventions which shape the process of capitalist development. Regulation is focused on five key aspects of capitalism in particular: labor and the wage relation, forms of competition and business organization, the monetary system, the state, and the international regime. When these act in concert, a period of stable growth, known as a 'regime of accumulation', ensues.

Although Marxist geographers began to draw upon regulationist thinking in the 1980s – particularly in studies of industrial restructuring (e.g., the Mick Dunford and Dianne Perrons text of 1982 or the Allan Scott and Michael Storper collection of 1986) – it was not until the 1990s that more spatially sophisticated regulationist accounts began to appear (e.g., the contributions by Mick Dunford in 1990 and Ash Amin in 1994), initially concerned with critically assessing the uneven geography of state regulation in response to processes of economic and political restructuring. The work of Jamie Peck and Adam Tickell, notably in 1994, has been particularly notable here in identifying the gaps in regulationist thought, especially the tendencies to underplay sub national variations in modes of regulation, both historically and through contemporary changes brought by neoliberal restructuring and the shift from Fordism to more flexible forms of employment. More recently, there has been a focus upon the geographical unevenness of neoliberalism itself as a mode of social regulation, as in Peck and Tickell's 2002 paper.

Another body of Marxist work, a new subfield of labor geography cultivated in varying ways by Andrew Herod, Don Mitchell, and Jane Wills, has attempted to correct for the capital centric Marxism of the 1980s by exploring how workers also shape capitalist landscapes. As one of the most prominent of these new labor geographers puts it:

> … the production of the geography of capitalism is not the sole prerogative of capital. Understanding only how capital is structured and operates is not sufficient to understand the making of the geography of capitalism. For sure, this does not mean that labor is free to construct landscapes as it pleases, for its agency is restricted just as is capital's by history, geography, by structures that it cannot control, and by the actions of its opponents. But it does mean that a more active conception of workers' geographical agency must be incorporated into explanations of how economic landscapes come to look and function the way they do. (Herod, 2001: 34)

While there is a long established tradition within Marxist geography of studies that explore how workers seek to 'defend place' against the threat of plant closure and industrial restructuring, Herod and others have been pursuing research that highlights the more proactive role of trade unions in creating their own 'spatial fixes'. Three themes have been particularly evident in labor geography: attempt by trade unions to internationalize their operations to come to grips with global capital; the way workers and unions develop their own strategies to respond to economic restructuring; and finally, new attempts to organize workers at the local scale in service sector activities.

Finally, there has been some productive work on Marxist understandings of nature and environment by the likes of Noel Castree, writing in 1995 and 2002, attempting to go beyond earlier conceptions in which nature was perceived as passively transformed by capitalist processes, without any sense of its own agency to a reconciliation of political economy with the very real ontological existence of other actors or 'actants', both human and nonhuman (as insisted upon by actor network theory). One of the most important contributions has been Richard Peet and Michael Watts's 1996 edited collection *Liberation Ecologies*, which seeks to integrate certain poststructuralist concerns – such as the importance of discourse, alternative understandings and knowledges, and the specificities of place – within a continuing political economy perspective on the uneven and environmentally destructive tendencies of capitalism. Critically, by locating their collection largely within environmental and indigenous peoples' struggles in the global South, they are able to show the continuing importance, both theoretically and politically, of a Marxist critique of capitalism, yet, at the same time, one that embraces the plurality of actually existing class struggles.

The Continuing Relevance of a Marxist Perspective to the Geographical Imagination

Since the late 1990s there has been something of a return to Marx and political economy, not least because of the growing evidence of the destructive and uneven consequences of unregulated global capitalism. With the 150th anniversary of the publication of *The Communist Manifesto* in 1998 and the financial crises then engulfing East Asia and Russia, even normally sober and conservative publications, such as the *Financial Times* and

New Yorker, published articles heralding Marx's thought. Although Wall Street's discovery of Marx was predictably short lived, listeners voted Marx the greatest philosopher of all time in a BBC Radio 4 poll in 2005. Ultimately, Marxist political economy is still relevant because of its value as a framework for understanding the evolution of the global capitalist system, a point appreciated by many in the recent global justice movement, notably John Holloway writing in 2005. Marx's primary contribution to knowledge was as an analyst of capitalism not as an architect of communism. While he offered only a few scattered comments about geography, this has been rec tified by Marxist geographers like Harvey and Smith with their own theories of spatially uneven development and historical geographical materialism. In the absence of other approaches which can match its historical geographical reach and analytical purchase, Marxism provides the most suitable framework for analyzing the 'big questions' concerning the economic geography of global capitalism. Marxism also retains a strong sense of social and political commitment, emphasizing issues of inequality and injustice and the need to change the world as well as interpret it.

At the same time, Marxist geography, as currently practised, continues to have its limitations and weak nesses. There remains a need to provide a stronger sense of agency, and integrate cultural, discursive, and insti tutional insights within a broader materialist under standing of contemporary capitalism and its historical development. In short, what is required is a Marxism that does not claim to have a monopoly on the truth, but is receptive to insights from other perspectives and which evolves in line with capitalism as its object of analysis. Ironically, given the rather hackneyed critiques of Marxist geography, one area of productive engagement might be over the current meaning of resistance and class struggle, perhaps as signposted by Cindi Katz in her 2004 monograph. Jessop's 2004 criticism of Harvey's *Limits to Capital* is apposite here, in that it is overly concerned with understanding the geography of capitalism from the point of view of capital (and thus an emphasis upon circulation, finance, etc.) rather than as class struggle and the perspective of labor.

Understanding capitalism from the perspective of labor and the different forms of resistance that emerge to global capital in particular places, how the local is woven into broader global networks in myriad, complex, and contested ways, suggests a fertile agenda for Marxist and other critical geographers. Such a perspective would depart from the still strong influences of Althusserian Marxism, to recapture more of the everyday experiences of, and resistances to, capitalism, and would understand concepts such as capital and class, not in reified terms, but as more fluid categories that change and evolve in time and space. However, it would still be a struggle conceived in dialectical terms, but a struggle which is itself framed by discursive as well as material resources. Critically, following the tradition of 'Open Marxism', the key to understanding capitalism is to see it, not just in terms of the exploitative relationship between labor and capital through which surplus value is created, but also to view it in terms of the alienation and the dehuman ization that accompanies commodification: the increasing subjugation of everything (people, animals, plants, the entire Earth) to commodity production. Resistance to capital is therefore a resistance to this subjugation and always an attempt to recover humanity and dignity. This is an important point because it means that we could take a broader view of labor and work, which are not just confined to an idealized industrialized proletariat and reified class positions, but extended to all forms of human activity (integrating production, social repro duction and consumption) and incorporating all the di verse ways in which economic and social life are increasingly subservient to commodity production. As Holloway puts it: "the theory of value proclaims the subjugation of doing, the fact that human, creative doing is reduced in capitalism to the dehumanising process of abstract labour, of value production" (Holloway, 2005: 186).

See also: Actor-Network Theory/Network Geographies; Cultural Materialism; Dialectical Reasoning and Dialectical Materialism; Historical-Geographical Materialism; Marxism/Marxist Geography II; Radical Geography; Structural Marxism.

Further Reading

Aglietta, M. (1979). *A Theory of Capitalist Regulation: The US Experience*. London: New Left Books.
Althusser, L. and Balibar, E. (1968). *Reading Capital*. London: New Left Books.
Boyer, R. (1990). *The Regulation School: A Critical Introduction*. New York: Columbia University Press.
Castree, N. (1995). The nature of produced nature: Materiality and knowledge in Marxism. *Antipode* 27, 342 362.
Castree, N. (2002). False antitheses? Marxism, nature and actor networks. *Antipode* 34, 11 46.
Castree, N., Coe, N., Ward, K. and Samers, M. (2003). *Spaces of Work: Global Capitalism and Geographies of Labour*. London: Sage.
Cleaver, H. (2000). *Reading Capital Politically* (2nd edn.). Edinburgh: AK Press.
Cooke, P. (1986). The changing urban and regional system in the UK. *Regional Studies* 20, 243 251.
Duncan, J. and Ley, D. (1982). Structural Marxism and human geography: A critical assessment. *Annals of the Association of American Geographers* 72, 30 59.
Dunford, M. (1990). Theories of regulation. *Society and Space* 8, 297 321.
Dunford, M. and Perrons, D. (1982). *The Arena of Capital*. Basingstoke: Macmillan.
Gramsci, A. (1971). *Selections from the Prison Notebooks*. London: Lawrence and Wishart.
Gregory, D. and Urry, J. (1985). *Social Relations and Spatial Structures*. Basingstoke: Macmillan.

Harvey, D. (1973). *Social Justice and the City*. Baltimore, MD: Johns Hopkins University Press.
Harvey, D. (1982). *The Limits to Capital*. Oxford: Basil Blackwell.
Harvey, D. (1989). *The Condition of Postmodernity*. Oxford: Blackwell.
Harvey, D. (1996). *Justice, Nature and the Geography of Difference*. Oxford: Blackwell.
Herod, A. (2001). *Labour Geographies: Workers and the Landscape of Capitalism*. New York: Guilford Press.
Holloway, J. (2005). *Change the World without Taking Power: The Meaning of Revolution Today* (2nd edn.). London: Pluto.
Jessop, B. (1994). Post Fordism and the state. In Amin, A. (ed.) *Post Fordism: A Reader*, pp 251 279. Oxford: Blackwell.
Jessop, B. (2004). On the limits to the limits to capital. *Antipode* 34, 481 495.
Katz, C. (2004). *Growing up Global: Economic Restructuring and Children's Everyday Lives*. Minneapolis, MN: University of Minnesota Press.
Latour, B. (1993). *We Have Never Been Modern*. London: Harvester Wheatsheaf.
Marx, K. (1906). *Capital: A Critique of Political Economy* (1st American edn.). New York: Random House.
Marx, K. and Engels, F. (1967). *The Communist Manifesto*. Harmondsworth: Penguin.
Mitchell, D. (1996). *The Lie of the Land: Migrant Workers and the California Landscape*. Minneapolis, MN: University of Minnesota Press.
Peck, J. and Tickell, A. (1994). Searching for a new institutional fix: The after Fordist crisis and the global local disorder. In Amin, A. (ed.) *Post Fordism: A Reader*, pp 280 315. Oxford: Blackwell.
Peck, J. and Tickell, A. (2002). Neoliberalising space. *Antipode* 34, 380 404.
Peet, R. (1981). Spatial dialectics and Marxist geography. *Progress in Human Geography* 5, 105 110.
Peet, R. (1998). *Modern Geographical Thought,* chs. 3 5. Oxford: Blackwell.
Peet, R. and Watts, M. (1996). *Liberation Ecologies: Environment, Development. Social Movements*. London: Routledge.
Polanyi, K. (1944). *The Great Transformation: The Political and Economic Origins of Our Time*. Boston, MA: Beacon Press.
Scott, A. and Storper, M. (1986). *Production, Work and Territory: The Geographical Anatomy of Industrial Capitalism*. London: Allen and Unwin.
Scott, A. J. and Storper, M. (eds.) (1986). *Production, Work and Territory*. London: Allen and Unwin.
Smith, N. (1987). Dangers of the empirical turn: Some comments on the CURS initiative. *Antipode* 19, 59 68.
Smith, N. (1990). *Uneven Development: Nature, Capitalism and the Production of Space* (2nd edn.). Oxford: Basil Blackwell.
Soja, E. (1980). The socio spatial dialectic. *Annals of the Association of American Geographers* 70, 207 225.
Thompson, E. (1978). *The Poverty of Theory and Other Essays*. London: Merlin.
Thrift, N. and Williams, P. (eds.) (1987). *Class and Space: The Making of Urban Society*. London: Routledge Kegan Paul.
Walker, R. (2004). The spectre of Marxism: The return of the limits to capital. *Antipode* 34, 434 443.

Marxism/Marxist Geography II

A. Jones, University of London, London, UK

Glossary

Classical (Western) Marxism The form of Marxism that emerged in the nineteenth and early twentieth century derived form Marx's writings.

Dialectical Thinking A method of argument or thought that systematically counterposes contradictory ideas with a view to combining and resolving them into a coherent synthesis.

Epistemology A theory of what constitutes valid knowledge.

Marxian Analysis The implementation of the concepts and theoretical arguments grounded in and/or derived from Marx and Marxist schools of thought.

Marxisant Analysis Theories or concepts that loosely draw on Marxist ideas but do not necessarily adhere to Marxist epistemology.

Modernist Epistemology A theory of knowledge based around the principles of rationalism, structures, truth, and universalism.

Neo-Marxism A loose term covering a range of twentieth-century revisions of classical Marxian analysis but still adhering to the major epistemological tenets of classical Marxism.

New Social Movements (NSMs) A term referring to a range of social movements that emerged in (Western) societies since the 1960s and which depart from conventional politics and political organization, for example, the civil rights movement.

Ontology A theory of the nature of what exists that permits knowledge to be possible.

Postmodern Thought A broad school of thought within philosophy, the arts, and social sciences characterized by skepticism toward the claims of grand theory of the modern era along with its underpinning epistemology.

Post-Marxism An assemblage of relatively diverse attempts to overcome the limitations of classical Western and neo-Marxisms.

Poststructuralism A broad school of philosophical thought stemming back to the 1960s which in rejecting the epistemology of structural linguistics led to critiques of rational subjectivity, truth and knowledge.

Revised Marxist Geography A broad range of attempts within human geography to overcome the critiques of classical Marxist approaches to the discipline.

Structuralism A strand of twentieth-century philosophical thought seeking to explain human language and behavior through the existence of underlying structures of which they are part.

Teleology The belief in or the perception of purposeful development toward an end.

Introduction

The development of Marxist thinking within human geography has been the subject of radical and contentious debates over the last 30 years. The reintroduction of the classic concerns of Marxist analysis is discussed in the first article (*Marxism/Marxist Geography I*), and undoubtedly represents a major shift toward the radicalization of human geography during the 1970s and 1980s. Through this period, questions of class, labor, and the transformation of material and social spaces became embedded in the mainstream debates of much human geography. However, what is also clear is that the subject was never really dominated by 'classical Marxism' as other strands of social and political thought had been. Rather, the evolution of Marxist geographical thought has been both diverse and influenced by quite different strands of Marxian thought – for example, both 'New Left' and Gramscian Marxist thinking. Different strands of Marxist thought produced distinct approaches within human geography and in that sense it is dangerous to simplistically assume that one, singular Marxist geography is distinguishable. To do so would run the risk of producing a caricatured Marxist geography that does not do justice to the complexity and nuances of much Marxian thought in human geography.

Nevertheless, by the late 1980s, the position of various Marxist geographies as the basis for the hegemonic 'radical' school of thought within the subject was certainly being challenged. This occurred principally for two interrelated reasons. The first of these concerned the status of Marxism as a philosophical and theoretical school of thought. During the 1980s, classical approaches to (Western) Marxism came under increasing attack and criticism at the epistemological level. Much of this criticism originated from philosophy and other social science disciplines, yet the critique of Marxist epistemology gradually permeated human geography and is captured in the idea of the 'postmodern turn'. Second, and not unrelated to these philosophical critiques, contemporary world events during the 1980s provided support and weight to the idea that Marxism was problematic and unable to offer empirically accurate or theoretically powerful insight into the trajectory of human development. Through the decade, events from the triumph of New Right politics in Europe and North America to the divergent pathways of developing countries suggested Marxist analysis was disconnected from actual events in the 'real' world. Perhaps most importantly, the events of 1989 with the collapse of the Soviet Union, the fall of the Berlin wall, and the demise of most of the so called communist 'Second World' seemed to herald the triumph of liberal capitalism and its values. This was taken by many as the death knell of traditional Marxist arguments concerning the tendency of capitalism toward crisis and the transformation of political systems to socialist values. As an intellectual and theoretical tradition, (Western) Marxism entered a period of crisis and decline. In human geography, as in other social science disciplines, this precipitated a huge crisis of confidence in Marxism.

This article addresses this next stage in the evolution of Marxist thinking within human geography, and the implications that it has and continues to have for contemporary Marxist geographies. Clearly an element of linear chronology in relation to the evolution of human geography is present insofar as much of the work covered by this article came after the more classical approaches of the 1970s and 1980s, and before the post structuralist and arguably full blown post Marxist geographies of the last decade. However, such a linear chronology of thought within the subject is not as neat as it may first appear, and significant overlaps, inconsistencies, and discontinuities exist. These difficulties in producing a neat linear chronology of Marxist human geography are perhaps most typified by the issues of what constitutes a revised Marxist as opposed to a truly post Marxist human geography. The 1990s saw several Marxist geographers' attempt to adapt their concepts and theoretical arguments in response to the burgeoning 'postmodern critique'. Such adaptations can be seen, depending upon your perspective, as little more than a continuation and reconfiguration of classical approaches to Marxist geography, revised Marxist geographies, or shifts amounting to post Marxist geographies. Different thinkers regard themselves as occupying different positions in this respect, and so the precise nature of post Marxist human geography remains the subject of debate. To understand all of these developments, it is necessary to consider the postmodern critique of Marxist thinking in greater depth.

Critiques of Marxist Thought

The philosophical critiques of Marxism that permeated human geography in the 1980s in fact stemmed from a long standing set of debates within continental philosophy that stem back to at least the 1960s. In particular, a string of French philosophers (Derrida, Foucault, Deleuze, Guattari, Lacan, Lyotard, and Baudrillard) and other social scientific thinkers form a broad body of thought leveling a variety of critiques at classical Marxism. In essence, these critiques are ontological and epistemological in nature insofar as they call into question the underlying assumptions made by Marxist analysis. Taken as a whole, they represent what can be broadly labeled the 'postmodern turn' in social scientific thinking, although much of the critique focuses around a more narrowly defined set of so called post structural philosophers. It is important to realize, however, that many of the key thinkers within post structuralism doubted or denied this label, and that it represents a diverse and sometimes contradictory body of thought. Certainly not all post structuralist thinkers rejected Marxian analysis outright. For example, while many geographical thinkers take Gilles Deleuze and Felix Guattari's major contributions *Anti Oedipus: Capitalism and Schizophrenia* and *A Thousand Plateaus* – in 1973 and 1982 respectively – as seminal post structural texts, they are also grounded in (revised) Marxian concepts. Similarly, even key post structuralist critics of aspects of Marxist analysis (such as Jacques Derrida) have continued to argue for revision rather than rejection of Marxist epistemology.

Perhaps the most important issue in relation to the critique of Marxism is, however, the fact that there was no one singular 'pure' form of Marxism which wholly represented the object of the critique. Marxist thinking was already heterodox and riven by (often) vigorous disagreement about key concepts and assumptions. These debates and divisions had already echoed through human geography before the crisis of Marxism arose. An example would be E P Thompson's famous critique during the 1970s of structuralist Marxism that echoed through much historical geography. In this respect, therefore, there never was one singular Marxism or Marxist geography to which all the emerging critiques applied to equally. It is nearer the truth to suggest that different critiques were relevant to 'some aspects' of different strands of Marxist (geographical) thought but that there was no one Marxism that was equally vulnerable to all of them.

Furthermore, while it is easy to impose a divide between 'philosophical' and 'world historical' strands to the crisis of Marxist thought, such a divide risks significant misrepresentation. In reality, the development of ideas within post structuralist and other strands of philosophy and the social sciences were themselves inextricably bound to events in the 'real' world. The events of the late 1960s in the US and Europe which witness the birth of new social movements (NSMs) (environmentalism, civil rights) were heavily influential on the philosophical debates that formed the center of the emergence of post structuralism, and the wider postmodernist movement. It is difficult to separate the inspirations behind Michel Foucault's thinking on sexuality and power from his political activity, and likewise Deleuze and Guattari's arguments about capitalism from their anti Fascist political agenda. Such elisions between thought, practice, and personal politics perhaps go some way to explaining why their relationship with Marxism is a complex mixture of revision and critique. This situation continued through the 1970s and 1980s, and so it is simplistic to reduce the emergence of either post structuralism or post Marxisms to a simple linear chronological story. Likewise, it is also important to remember that several key thinkers sought to distance themselves from Marxist analysis in the 1970s and 1980s only to return to its epistemological tenets in the 1990s.

The following discussion therefore identifies at least eight major themes of critique that have questioned Marxist thought in recent decades. These are not exhaustive, nor necessarily discrete from each other, but represent some of the major (ongoing) challenges for Marxism and Marxian analysis within human geography (as elsewhere in the social sciences).

Economic Determinism

The first criticism concerns what is argued to be the crude economic determinism inherent in Marxian analysis. During the 1980s, Ernest Laclau and Chantal Mouffe argued that Marxist analysis reduces its explanation of the nature of society and societal change unequivocally to the economy. This has been termed 'the economic monist causality of Marxism', where Marxism is accused of rooting all explanation for societal change in economic factors. Such a critique echoes Barry Hindess and Paul Hirst's argument that there is no necessary correspondence between the working class and socialism. They argue that the relative autonomy between the political and the ideological produces a multiplicity of popular struggles around universal human goals and a concern for discourse and nonclass practices. Such a critique draws on world historical developments in terms of attempts to reformulate Marxian analysis to accommodate the emergence of the NSMs since the late 1960s. Certainly, by the 1980s, traditional Marxism's strong reliance on economic centered explanation had been widely attacked within human geography. For example, Stuart Corbridge extended this line of critique to the neo Marxisms of the 1970s and 1980s. He argues that even these reformulated forms of Marxism exhibit similar problematic tendencies in that they understand societal development purely as a consequence of the logic of capital. Furthermore, in line with critiques of rationalism, Corbridge argues that neo Marxist theory (in the context of development studies) perpetuates a reliance on an epistemologically weak understanding of causation resulting from the influence of underlying economic structures.

By the late 1980s, the argument that Marxist approaches were couched in economic determinism had also become more widely associated with the so called 'cultural turn' within human geography. The Marxist geographies of the 1970s and 1980s were seen as far too narrow in terms of the breadth of factors through which they theorized society and economy. Social and cultural issues were being ignored with the result that Marxist geography could not offer adequate explanation for the development of later twentieth century world society.

Marxism as a Grand Theory

A second related criticism comes at the philosophical level. A range of post structuralist and postmodern thinkers – for example, Lyotard, Foucault, Laclau, Derrida, Deleuze – criticize the 'grand vision' of Marxism and argue that it presents a totalizing discourse that crushes the multiple differences of human experience. This critique has a number of diverse strands that also span linguistics, history, literary theory, cultural theory, and feminist thought. Jean Francois Lyotard's 'incredulity at the metanarrative' thus represents the emblematic critique of this epistemological tendency within Marxism. Lyotard argues that the privileged position of the theorist or observer (or any subject) as being able to regard the universal truth is untenable. The North American philosopher Richard Rorty terms this a rejection of the 'God's eye view'. Contrary to this bedrock of modernist epistemology, the postmodernists argue that interpretations and the authority of the theorist or observer are socially constituted, contingent, and partial. This 'crisis of representation' undermines the privileged epistemological position and totalizing claims of Marxism about being able to represent the truth of the world. In contrast, the postmodern pluralist perspective requires an openness to multiple positions and a range of voices, and rejects the idea that changeless, universal, foundational relationships exist. Such a position entails a vigorous rejection of much of the universalizing claims of Western Marxism around the formation of class, class politics, and the nature of capitalism. It thus represents a wholesale attack at the epistemological foundations of much Marxist analysis which it sees as trying to conceptualize and understand diverse societal conditions through an inflexible, rigid, and inappropriate theoretical metaframework. In Derrida's terms, traditional Marxian analysis 'does violence' to our understanding of the diversity of human experience under contingent conditions in time and space. Like all such 'grand theory', it thus smothers understanding of the complexity of societies, economies, and their evolution.

Structuralism and System

A further related dimension to the critique addresses the structuralist epistemological foundations of Marxism. As a grand theoretical framework based around economic relations, Marxism is also criticized for its epistemological structuralism. Marx's theorization in *Capital* identifies a series of underlying collective relations in society that emerge from the development of capitalism (primarily as a conflict between exploitation and resistance in society). In particular, he identifies structural connections which form the link between the social relations of production and the form of the state. These structural factors limit and direct the kind of state which comes into existence. Later forms of Marxism certainly deployed these concepts of structural forces in a more rigid and deterministic manner than Marx himself, and certain strands of classical Marxism appear to lose the subtlety and sophistication of Marx's original exposition of these structural relations.

Nevertheless, the post structuralist turn in philosophy and the social sciences represents a strong critique of this idea of underlying structures, certainly in any rigid form. Classical Marxism's concepts of base infrastructure, superstructure, and the mode of production are all questioned as reductionist generalizations that fail to adequately capture the complexity of contemporary world society. Postmodern thinking is skeptical of the very coherence of such structures as singular entities that have agency. Thus, the greatest problem of all is the tendency within various (classical) Western Marxist schools of thought to find or imply causality within these structural relations. Various postmodern thinkers argue that concepts of structure are crude reductions of causality (i.e., how and why social transformations occur) and overdetermine the multiple actors and influences that shape diverse outcomes in societal development.

Related to this argument is the issue of the utility of the concept of the 'system' as applied to societal development. In a similar manner to the skepticism surrounding the existence and coherence of underlying structures, Marxism has also been criticized for assuming that social and economic relations can be meaningfully understood as systemic in form. Work within the history and philosophy of scientific thought suggests that the metaphor of the system was 'borrowed' from the natural and biological sciences in the nineteenth century and has subsequently been inappropriately applied to socioeconomic relations in Marxian thought.

Teleology

Marxism is mechanistic insofar as it construes history as a series of stages and modes of production (traditional societies, feudalism, capitalism, socialism). This 'teleology' has been heavily criticized. In essence, a range of critical strands around this issue again point to the historical rootedness of Marxism in nineteenth century concepts. Marx's original theoretical argument is that socialism, achieved through revolution, would be the inevitable outcome of the chronological development of capitalism. He contends that the inherent crises of capitalism in all societies produce the conditions that ferment class formation, class consciousness, and eventually a revolutionary transformation to a socialist society. Just as the conditions for the emergence of capitalism were created by the ongoing development of feudal societies, so the conditions for socialism are produced by capitalism. Long a point of debate in classical Marxism, later twentieth century critiques argued that Marxism inappropriately lifted concepts and metaphors from the natural sciences in the nineteenth century (notably ideas of evolution). However, undoubtedly, the root teleological tendencies within Marx stem from the philosophy of Hegel whose conception of dialectical development clearly informs Marx's arguments that modes of production develop through a process of dialectical materialism.

Eurocentricism

By the 1980s, both Marxist geography and Marxist social science more generally were under attack for representing a Eurocentric theoretical explanation for world historical development. In particular, the developing postcolonial critique from thinkers in the less developed world suggested that classical (Western) Marxism is historically and geographicaly specific in its origins, concepts, and analytical framework. Marxism is thus deemed an inappropriate theoretical tool for understanding the development of non Western societies as development thinkers argue that developing countries have not and will not necessarily follow a pattern of development similar to the Western countries of the global North. Marxism has therefore become cast as a Eurocentric interpretation of world historical development, and one that is not sensitive to the different societal and cultural context of countries in the global South.

Class Reductionism

A range of critical voices from the 1960s onward argued that classical and neo Marxisms are problematic because of their tendency to reduce societal development to the relations between different social classes. Given the primacy of the dialectical conflict between the proletariat (working class) and capitalist class in classical Marxism as a mechanism producing change, by the later twentieth century a range of thinkers argued that the evolution of contemporary societies could not be adequately explained through class relations alone. Again, this line of critique was bound to world historical developments and in particular the emergence of NSMs based around identity politics or issues from the 1960s. Such collective political movements transcend class boundaries and by implication call into question the capacity of Marxian analysis to explain the development of new social formations. Furthermore, by the 1980s, the notion of capitalism as a singular economic system was also regarded as outmoded with an emerging literature across the social sciences arguing that multiple forms of capitalism exist and that there exist considerable differences between different varieties of capitalism. For example, development theorists, drawing on a postcolonial perspective, also question whether labor in less developed countries can be adequately conceptualized in Marxism's epistemological framework. The implication of these arguments is a growing skepticism as to whether it is possible and/or desirable for the principal object of Marxian analysis – capitalism – to be successfully demarcated as a singular concept with similar characteristics.

Finally, and related, feminists also argued that traditional Marxian conceptions of class remained heavily gendered and ignored the complex class position of women in the later twentieth century. By the 1980s, this line of critique has permeated human geography with feminist geographers like Linda McDowell arguing that the new class formations of Western societies did not fit classical concepts as women and men's class identities have been transformed by de industrialization, identity politics, and cultural change.

Gender

Seventh, within human geography and across the social sciences, feminist thinkers argue that classical Marxism ignores important issues of gender. Feminists questioned both the notion of what constitutes 'work' in the economy and also how those who undertake it can be defined. They argue that Marxian concepts of labor as paid work represent a narrow and partial conception of work as a practice undertaken in contemporary societies. In Western societies, feminists point to the way in which Marxian analysis (and other classic frameworks) ignores the economic value of unpaid work (much undertaken by women).

However, feminist criticisms of Marxism have extended to all aspects of Marxist epistemology. More recent feminist thought, drawing on postmodern and post structuralist thinking, has developed a wide ranging and multipronged assault on both the validity of the concepts deployed by Marxism to accurately represent the late twentieth century world, and also as a consequence wider propositions Marxism made about the systemic and structural evolution of societies and economies. The feminist critique destabilizes both the epistemological boundaries that defined concepts like class and also the theoretical claims about the outcomes of processes of class action, consciousness, or struggle within societies. What began therefore as a criticism that women were marginalized or ignored in Marxian analysis in the 1970s, developed, by the 1990s, into a wholesale skepticism about the capacity of Marxian analysis to theorize any wider truths about socioeconomic development. For example, within human geography, feminist thinkers such as J. K. Gibson Graham argue for a radical break from traditional frameworks of Marxian analysis. Gibson Graham's 1996 book *The End of Capitalism* (*As We Knew It*), which also draws heavily on the post structuralist critiques of Marxism, thus presents a significant epistemological departure from Marxist geographies of the 1970s and 1980s.

Environmental Issues

Finally, although by no means less significant, is the critique of Marxism that emerged from environmental thought. The environmental movement and its politics represented one of the NSMs from the 1960s that challenged the conventional view of societal development espoused by classical Western Marxism. Marxism was criticized for its exclusive theoretical focus on the economic and material conditions in society, and for ignoring issues of environmental degradation and the relationship of environmental issues to social justice. Again, as with other new social movements, the new issue based political struggles in the developing global environmentalism appeared to represent a challenge to classical Marxian conceptions of the mechanisms behind societal change and the way in which that would be achieved. From the 1960s, the environmental movement transcended class boundaries and traditional Marxian analysis struggled to theorize the causes and solutions to environmental problems. Within human geography, this critique grew louder during the later 1970s and 1980s as Marxist geography continued to focus on traditional Marxian problems – social justice, urban change, and the nature of capitalism in the world economy. However, the absence of environmental analysis in Western Marxism also provoked one of the earlier revisions to Marxism in the form of political ecology which emerged from neo Marxian development theory during the 1970s.

The Question of Post-Marxism(s)

As the above discussion indicates, the crisis of Marxism produced an enormous and wide range of responses through the 1970s, 1980s, and 1990s across the social sciences as well as in human geography. What it also produced, by the 1990s, was a vigorous debate within and beyond human geography as to what post Marxism might represent. Various literatures began to describe themselves, or be described as, 'post Marxist' but disagreement as to what this actually meant was inherent in the debate. Ernest Laclau and Chantal Mouffe offer a good typology of what must be considered in this term when they argue that their position is both 'post' Marxist and post 'Marxist'. The former ('post' Marxism) refers to the obvious 'after Marxism' of a perspective that has abandoned the main principles of Marxism. This, they argue, corresponds to the position of postmodernist thinkers including Lyotard and Baudrillard. However, the latter category (post 'Marxism') refers to attempts to graft recent theoretical developments in post structuralism, deconstruction, postmodernism, and feminism onto Marxism. This approach, which Laclau and Mouffe state as their goal, aims to make a revised Marxism relevant to a new cultural climate that is no longer responding to classical Western Marxist doctrine. In this respect, post Marxism in any form can be differentiated from neo Marxism(s) in that it is much more critical of classical Marxism and takes into account changes in world politics and international relations since World War II. However, as Stuart Sim points out, this kind of post Marxism can look like 'a somewhat unstable coalition of interests' which are held together (or not) by the 'angle of intervention'. The next section will develop this issue in relation to (post)Marxist geographies.

Ron Johnston identifies three schools of post Marxist thought from outside the discipline that have permeated human geography. These are not distinct schools of post Marxism *per se*, but represent major approaches to development. First, during the 1980s there were attempts to develop a more rigorous form of (post)Marxian analysis which was termed 'analytical Marxism'. This school of thought emerged within political economy and derived from the work of the Canadian philosopher Gerald Cohen. The major goal of this project is to combine Marxian concepts and theories with rational choice and methodological individualism, and analytical Marxism rejects any form of dogmatic classical Western Marxism. Second, a string of political science theorists argue for a post Marxist approach which recognizes the autonomy of politics and economy, and thus is founded on a different notion of causality and a more sophisticated conception of capitalism as a system dependent upon and shaped by many conditions, not just the relations of production. The work of Laclau and Mouffe as well as Hirst and Hindess falls within this category. Third, and influential within human geography, the (largely) French school of regulation theorists developed an approach that sought to avoid the teleology and essentialism of Marxism by focusing on 'mid level' concepts. In particular, the regulation theorists seek to theorize the historically specific institutional configurations of state form, capital, and labor which give rise to equally historically specific (and unstable) regimes of accumulation. Human geography has drawn heavily on this approach in debates on industrial and economic organization, and its relationship to societal development. A key example is the debate over the transition in Western economies from a Fordist to a post Fordist regime of accumulation during the 1980s and 1990s.

However, these three schools of thought in fact represent only a partial array of what can be seen to be the 'purer' post Marxist approaches within the discipline. These schools of thought remain close in terms of their lineage to Western Marxism in both epistemological terms. Others have argued for a much wider definition of what represents post Marxism to include all the different schools of thought that draw more weakly on Marxian analysis but depart more radically form the epistemological approach of Western Marxism. This includes many strands of recent human geographical thought that clearly draw on a Marxian tradition but also engage heavily with post structural and postmodern thinking. The next section will thus consider a number of the most influential strands of 'post Marxist' geography and assess whether they warrant this label, or whether they might be better categorized as 'revised Marxist' geographies or even 'in between' Marxist and post structuralist geography.

Revised or Post-Marxist Geographies?

Since the late 1980s, a series of attempts have been made to salvage Marxian analysis within human geography from the crisis of traditional Marxism while retaining many of the major epistemological and conceptual strengths of the framework. In the terminology of Laclau and Mouffe, they represent post 'Marxism' insofar as they do not represent a total abandonment of Marxist principles, concepts, and epistemology but seek rather to graft Marxist concepts and elements of a Marxist epistemology onto other epistemological approaches. For the most part, these 'post ' or 'revised' Marxist geographies thus represent attempts to produce new hybrid epistemologies that combine elements of Marxian analysis with postmodern and post structural thinking. In that sense they are certainly 'revised' forms of Marxist geography more than they are 'post' Marxist. The following discussion considers four of the most significant strands to these revised Marxist geographies.

Historical-Geographical Materialism

Probably, the most notable and most clearly demarcated strand of revised Marxist geography centers on the conceptual framework of historical geographical materialism. This has been largely developed by and is associated with the work of David Harvey. His arguments for a historical geographical materialist approach run back to the 1970s and 1980s. Harvey played an important part in bringing Marxian analysis to the forefront of thought in (radical) human geography. As the first article discusses, his books *Social Justice and the City* in 1973, *The Limits to Capital* in 1982, and *The Urban Question* in 1985, apply what is a geographical version of Western Marxism to the issues of social justice, the world capitalist economy, and urban development. By the later 1980s, Harvey also turned attention to the cultural development of late capitalism in his *Condition of Post Modernity*. It is at this point that he first begins to develop detailed arguments for a revised Marxian spatial epistemology that seeks to reconfigure Marxist concepts and principles – such as the processes of capital accumulation – with a sensitivity to geographical differences. He argues in effect that the crisis of Marxism rests around the failure of classical Marxist historical materialism to grasp the significance of shifts in political economy, state function, cultural practices, and in the temporal configuration of social relations. The crisis of Marxism can thus be overcome if historical materialism takes geography seriously.

During the 1990s, Harvey developed this historical geographical materialism as the central plank of his response to the critiques of Marxism associated with the postmodern turn. Two distinct strands to this are identifiable. First, in his *Justice, Nature and the Geography of Difference*, Harvey takes up the criticism of environmental or green theorists, and seeks to apply his Marxist approach to contemporary debates about nature, society, and the global environment. This is discussed further below under eco Marxist geographies. Second, Harvey also responds at length to the critiques of Marxism proposed by post structuralist and postmodern thinkers. This response strikes back at the core of the 'post ' critique insofar as it challenges the claims concerning of the crisis of representation, the nature of theorizing, and the cultural turn. Harvey argues that postmodern human geography runs the risk of becoming de radicalized and incapable of producing politically engaged theory. In this respect he positions himself 'against' the post structuralist shift in human geography and beyond 'whose sole purpose seems to be to take irreconcilable difference, incommunicability, particularity, and irreducible individualism'. However, he also argues that some of the criticisms forwarded by postmodern thought are pertinent and need to be heeded. Importantly, for Harvey, the forms of knowledge produced by postmodernism and post structuralism are not incompatible with Marxian analysis. He contends that 'historical geographical and materialist theory, because it deals in totalities, particularize motion and fixity in a certain way', offers the prospect of embracing these approaches 'within its frame'. Central to this is the power of dialectical thinking, which Harvey argues is much closer to deconstruction than many postmodern thinkers acknowledge. Contrary to the postmodern critique, therefore, his stated goal is to rebuild Marxian metatheory in such a way as to incorporate an understanding of spatiality and temporality (and socioecological issues).

In sum, Harvey proposes his historical geographical materialism as an epistemological framework that retains the powerful insights emergent from Marxian analysis while absorbing and adapting to the post structuralist and postmodern critique. His contention is that this represents a reinvigorated and revised Marxist geography approach that will enable human geographers to conceptualize contemporary global society in a manner compatible with a progressive radical and empowering politics that promotes greater social and environmental justice. This framework therefore aims to retain a vision of socialism understood as an alternative vision of 'how society will work', 'how socioecological relations will unfold', and 'how human potentialities can be realized'. To achieve this, the 'extreme post structuralist fantasy' of thinkers like Derrida needs to be dismissed, and replaced with a new revised Marxist geographical approach that provides a basis to theorize and politically engage with contemporary society's circumstances. This position has been labeled a 'post Marxist' one, although Harvey would not see it as such. Thus, whether Harvey's historical geographical materialism is therefore a revised Marxist or post Marxist geography remains a moot point, and one which will be further considered in the next section.

Revised Marxist Political and Economic Geographies

Broadening out the scope of Marxist geographies from the more narrowly defined historical geographical materialist approach, a second strand of revised Marxist geographies can be grouped together around the wider approaches to political and political–economic theory within human geography. These areas of human geography all seek in various ways to continue to draw upon a Marxian framework to an extent, while also incorporating revised epistemological frameworks and reconfigured concepts to adapt to the earlier critiques of Marxism. At least four schools of thought warrant attention.

First, within development geography, there have been continued attempts to reformulate Marxian theoretical frameworks since the crisis of Marxism in the late 1980s. The 1980s was described by several commentators as the 'lost decade of development', and neo Marxian schools of thought about development theory – such as Andre

Gunter Frank's dependency theory – are seen to be seriously undermined by the critique of Marxism. This produced an 'impasse' in development theory which geographers responded to, but were certainly not necessarily in agreement over. On the one hand, during the 1990s, development geography was pushed toward post structuralist thought that questions the very notion of 'development' as a grand narrative (hence the term 'post development') and concerned itself with issues of power, discourse, multiplicity, and difference. On the other, thinkers such as Richard Peet are critical of these post Marxist approaches to development geography. Instead, in his 1990 book *Global Capitalism*, he outlines a vigorous defence of the utility of a (modestly) revised structural Marxism for theorizing development. He argues that much of the postmodern critique of Marxism's structuralism and teleology are invalid. For Peet, the critique of Marxism is based on a simplistic reading of classical Marxism that fails to appreciate its sensitivity to how systemic phenomena develop, their dynamic nature, and the processes by which they are reproduced. In essence, as with Harvey's revised historical geographical materialism, Peet sees much of the postmodern critique of Marxism (and hence Marxist geographies) as being misguided in its view that a Marxian approach is dogmatically teleological or simplistically structuralist. He therefore argues that the death of structural Marxism 'has been exaggerated' and that there is a need to continue to theorize societal structures rather than adopt a postmodern 'antistructuralist position'.

Second, and in part linked to the revised Marxian approaches to development theory, another school of geographical thought continued to work within the epistemological framework of world systems analysis. Proposed by Immanuel Wallerstein in the late 1970s, world systems analysis in its original form represents a reconfiguration and spatialization of classical Western Marxism. In particular, it articulates the historical processes by which capitalism extended to the planetary scale in geographical terms, as well as seeks to theorize the geographical manifestations of exploitation between core and peripheral regions of the world economy. Through the 1980s and 1990s, human geographers continued to implement this approach in theories of societal change in the less developed world. Furthermore, world systems approaches sought to absorb a number of the critiques of Marxist epistemology through a more dynamic and relational view of world capitalist system. In this respect, in the late 1990s, theorists such as Peter Taylor were advocating a political geography grounded in a world systems approach that remained Marxian but which has revised the nature of the concepts used from those proposed by Wallerstein in the 1970s.

A third school of thought within political and economic geography that continues to draw on the Marxian tradition to some degree is concerned with the ideological and conceptual basis to contemporary global capitalism and its politics. This school of thought gravitates broadly around the concept of neoliberalism where a range of geographical thinkers continue to couch their analysis in stronger or weaker Marxian frameworks to varying degrees. On the one hand, Harvey's perspective on neoliberalism remains very much grounded in his historical geographical materialism. In contrast, others such as Jamie Peck and Adam Tickell – in seeking to theorize the spatial and institutional manifestations of neoliberalism – are clearly informed by Marxian concepts but also integrate other epistemological perspectives from human geography and the social sciences.

Fourth, the arguments of Andrew Sayer in his 1995 book *Radical Political Economy* also inform a strand of human geography concerned with radical theory. Sayer's previous work had for some time sought to develop a more sophisticated path through the debates between modern and postmodern approaches to theory and epistemology – notably in his arguments for the utility of a critical realist methodological approach. His contention with respect to political economic analysis is that the crisis of Marxism has provoked a dangerous reaction in the social sciences. For Sayer, while a reduction in the dogmatism of totalizing grand theory is welcome, it is dangerous to drop 'any theory which claims to explain extensive and interlinked systems'. Postmodern critiques of structural theories are in his view 'likely to divide and fragment understanding' rather than illuminate. He argues that the postmodern critique thus runs the risk of dumping Marxist analysis on dogmatic *a priori* grounds that it is a totalizing discourse without critically engaging with its substantive content.

Sayer's diagnosis has common ground with Peet, but his proposed solution differs. Rather than a revitalized structuralist Marxist geography, he calls for a 'loose post Marxist' consensus around a revised Marxian epistemology that he suggests many empirical researchers have in fact already moved toward in light of the critique of Marxism in any case. This involves 'the usage of markedly diluted and de totalized versions of Marxist concepts and an increasing eclecticism'. It therefore represents an abandonment of, among other things, 'the privileged status of class, the labor theory of value, and the old time religion of alienation' as well as' de prioritizing the existential qualities of labor, adopting a more positive attitude to consumption', and admitting 'the importance of nonclass oppression and difference'. He thus also argues that Marxism is not incompatible with discourses of postmodern social science, contending that 'the discourses of political philosophy are not separate, but deeply intertwined'. His 'reformulated post Marxist political economy' thus seeks to 'go beyond' elements of several of these discourses.

Marxian Urban Geography

A third strand to revised Marxist human geography centers around urban theorization. Urban geography as a subdiscipline arguably suffered a crisis of identity in the early 1990s as the postmodern and cultural turn under mined perspectives seeking to map generalized patterns of urban form and the processes shaping them. However, a loose school of urban geography has sought sub sequently to continue to use a revised Marxian informed epistemology to theorize urban development and process. At least two significant threads are worth identifying.

First, urban geography continues to utilize a Marxian perspective in theorizing urban form, planning, and de velopment. Important to many of these contributions is the influential work of French Marxist philosopher Henri Lefebvre and, in particular, his books *The Production of Space* and *Writing on Cities* which were published in Eng lish in 1991 and 1996 respectively. Lefebvre's work offers a Marxian epistemological framework for conceptualizing the nature of space as a socially produced phenomenon. Furthermore, Lefebvre was long critical of a rigid, structuralist Marxist approach and thus his socially and culturally sensitive Marxian epistemology is easier to align with revised Marxist perspective in light of the postmodern critique. Lefebvre's approach has permeated many forms of revised Marxist human geography, but it is most evident in urban geographical theorization as Lefebvre himself applied his arguments most directly to issues of urban planning and form. For example, the work of Neil Smith on public space in contemporary cities draws heavily on Lefebvrian notions. Smith analyses how public spaces are disappearing and being redefined as centers of private commerce and consumption, and even surveillance and police control. He also expands this definition of 'public' to the transnational, to democratic spaces of dissent. Such arguments draw upon Lefebvre but also are clearly influenced by a range of postmodern thought including that of Foucault on power, politics, and surveillance. Smith's work on gentrification also integrates Marxian concepts with insights from the cultural turn in human geography. In the *The New Urban Frontier* (1996), he moves beyond his earlier Marxian theoretical frame work to argue that gentrification should not be under stood as the simple outcome of the new middle class tastes and a demand for urban living, but rather as part of a much larger shift in the political economy and culture of the late twentieth century. Smith's urban geographic perspective is thus both within a Marxian lineage while also drawing on postmodern and post structuralist ideas.

Second, although overlapping, another thread of urban geography has continued to use a revised Marxian per spective to theorize urban politics and social change. Once more, Harvey's work continues to seek to deploy his his torical geographical materialist perspective to 'the urban question'. In considering contemporary globalization and uneven geographical development in the global economy, Harvey considers their impact on urban form and land scapes in a framework that remains clearly informed by his revised geographical historical materialism. His arguments, for example, in *Spaces of Hope* seeks to understand how the imaginaries of contemporary urban development of the built environment are shaped by neoliberalism and how Marxian analysis can provide a critical framework through which to reflect on and engage with those imaginaries. Human geographers have likewise also continued to draw on the work of Manuel Castells whose analysis of the network society continues to draw on his earlier Marxian urban political economy, but whose work since the 1990s has used Marxian concepts in new ways that are informed by Lefebvre and social and cultural theory.

Eco-Marxist Geographies

The fourth and final strand of revised Marxian human geography that has generated an identifiable literature can be loosely described as a revised eco Marxist human geography. The growing response by radical human geographers to the postmodern charge that Marxism has paid little attention to environmental politics and problems represents the common theme through this work. Much of this thought has drawn on a burgeoning eco Marxist lit erature beyond human geography, although it has also developed the Marxian perspectives on the production of nature developed by Neil Smith and others in the 1980s. More recent contributions represent a form of revised Marxist geography insofar as they have increasingly sought to engage with postmodern epistemological arguments concerning society/nature and new forms of politics. Such eco Marxist human geography has thus been concerned, for example, with debates about environmental political movements, the rights of indigenous groups in the de veloping world, and environmental values.

In conceptual terms, Harvey's contribution has again been one of the most important in this strand of thinking too. In *Justice, Nature and the Geography of Difference*, he argues that there is a need for 'a much more unified language' for 'exercising the joint responsibility toward nature within the social and biological/physical sciences'. Harvey's proposition is to develop a geographical eco socialist politics that seeks to conceptualize 'what hap pens in the environment today' as 'highly dependent upon capitalist behaviors, institutions, activities and power structures'. The task of this environmental version of historical geographical materialism is to theorize how socialism can conceptualize 'the duality of ecological–social transformations as part of a far more coherent project' than previous environmental thought. Harvey thus offers a blueprint for a revised eco Marxist geography whose conceptual tools are those of Marxian analysis, but which also absorbs many of the critical arguments of postmodern thought and ecopolitical thought. That is not to suggest, Harvey argues, that wholesale agreement can be reached between a Marxian perspective and ecological thinking, but that a dialectical engagement in the debate with Marxist thought will produce a new (revised) framework for addressing the challenges that exist in the global environment and its relationship with global society.

Critiques of Revisionist Marxist Geography

The revised forms of Marxist geography that emerged in human geography during the 1990s continue to be widely cited and influential within the discipline today. However, they also received criticism from a range of theorists broadly aligned with the post structuralist based schools of thought.

First, feminist geographers have continued to criticize Marxian approaches to human geography for a lack of sensitivity to diversity and difference in contemporary social and economic life. In particular, there is continued dissatisfaction with the political possibilities that even revised Marxian human geographies present in the twenty first century. For example, J K Gibson Graham's arguments in *The End of Capitalism* (*as we knew it*) continue to question whether Marxian political economy can successfully capture the multiple and fractured forms of identity behind collective organization and action in today's world. One of the most enduring arguments developed is that Marxian analysis continues to represent capitalism as a unified, singular totality, rather than a multiplicity and plural, partial, feature of society. Gibson Graham argues that this prevents successful radical thinking and precludes adequate understanding of what alternatives to capitalism might be possible.

The implied consequence is a rejection of the revisionist perspectives that seek to combine Marxian and postmodern conceptions of class and other Marxian concepts. Instead, while retaining a weak Marxian echo in drawing on the political–economic tradition, Gibson Graham's new epistemological framework draws heavily on poststructuralism and is 'post' Marxist insofar as in its significant departure from Marxian structuralist principles and radical revision of Marxist concepts. Gibson Graham's approach reconceives identity and political action beyond a class based politics and is opposed to conceptions of societal development that are shaped by rigid structural imperatives.

Second, and similarly, other contributions drawing on post structuralism have challenged the view that revised Marxist geographies can be successful in reconciling any kind of Marxian epistemology with the post structural and postmodern critique. In this line of argument, revised approaches such as Harvey's historical geographical materialist perspective have only engaged with the 'post critique' at 'a relatively superficial level'. It is contended that many revised Marxist geographies thus only pay 'lip service' to the language of post structuralist philosophy rather than absorbing its epistemological contentions about how the social world can and should be conceptualized. Two fundamental criticisms are most significant. On the one hand, Harvey's concepts have been criticized for remaining essentially unchanged from the rigid and inflexible conceptual lexicon of classical Western Marxism. On the other, his argument that dialectical thought is little different to the deconstructive and fluid conceptions of multiplicity in post structuralist approaches is also doubted.

More recently, Doreen Massey has continued to make a case against human geography seeking to base itself around 'grand narratives' such as Marxism in arguing that the future of time space is 'open' and not closed or scripted according to a grand narrative. For Massey, human geography should thus embrace 'the very possibility of the intersection' and 'juxtaposition' of 'different conditions, influences, and stories'. Such a position appears to be at odds with a revised Marxist geography, at least in any of its 'stronger' epistemological versions.

Third, more recently, debates on the nature of scale within human geography have also developed a critical stance toward much longstanding and revised Marxist geography. The ontological status of the concept of scale is the key issue here. From a post structuralist stance, geographer's adoption of actor network approaches has provoked skepticism as to whether, in Bruno Latour's terms, the world's complexity can be captured by "notions of levels, layers, territories [and] spheres." Latour argues that the world should not be thought of as made up of discrete levels of bounded spaces which fit together, as if scales could be stacked together.

Such arguments represent a challenge to much existing and revised Marxist human geography which has made extensive use of scale as an organizing concept for understanding class, political action, social justice, and globalization. Many thinkers who have reformulated their Marxian perspective and sought to overcome many of the critiques associated with the crisis of Marxian still continue to rely heavily on scalar epistemologies. In different ways, the work of Smith, Harvey, Eric Swyngedouw, and Kevin Cox all deploy Marxian concepts through theorizations of scale that post structuralist geographical thought of Gibson Graham and Massey challenges.

Fourth, there is continued debate about the nature of eco Marxian approaches to human geography. The most significant challenge to eco Marxist geography has come from a strand of post structuralist thought using an actor network approach. Actor network theory (ANT) has developed from the sociology of scientific knowledge concerned with the way in which knowledge of the natural world is created and its relationship with the research process. Key thinkers include Bruno Latour, Michel Serre, John Law, and Michel Callon. ANT's key criticism of eco Marxist geographies (as with social scientific thought more generally) centers on conceptions of nature–society relations. ANT argues that conventional social scientific theory imposes a false binary divide between nature on the one hand and society on the other. Latour, Law and others argue that this false divide is disempowering, and produces partial and problematic conceptions of power, agency, and practice. Latour, in particular, has developed a strong critical stance with implications for Marxian (and other modernist) conceptions of scale. Latour argues that the existing (modernist) conceptions of 'stacked' or 'nested scales' already discussed – where local places are intersected by 'big scalar' processes at the regional, national, or global level – are highly problematic. Contray to this epistemology, Latour argues instead that socioeconomic life is better understood as constituted through complex sets of relations between humans and nonhumans, and that 'actors', 'power', and 'agency' all emerge from multiple and diverse sets of associations which form networks of varying strength and size. These actor networks permeate every aspect of contemporary social and economic life and transect every scale from the local to the global, as well as all the spaces in between.

Within human geography, thinkers such as Sarah Whatmore have argued that despite seeking to overcome the lack of environmental concepts in classical Marxism, eco Marxist geographies fail to adequately conceptualize the nature of issues such as environmental politics because it continues to rely on rigid, structuralist, and reductive conceptual binaries. In essence, the ANT critique once again returns to the wider challenge that revised Marxist geographies fail to overcome the full force of the post structuralist critique. However, there is no consensus in this debate, and Noel Castree has argued that the ANT critique does not represent an insurmountable challenge for eco Marxist geography. Castree argues that 'strong' versions of ANT go too far in undermining the powerful insights of Marxian political economy. He contends that by revising Marxist epistemology along 'relational' lines, Harvey, Swyngedouw, Smith, and others, in fact, go a long way toward successfully collapsing both the nature–society and global–local binaries. This final point leads, then, to the concluding section.

The Future of (Post-) Marxist Geographies

The extent to which Marxist geographies continue to present a radical, coherent, and effective framework for theorizing remains an open and unresolved debate. Clearly, a significant number of key thinkers continue to make detailed arguments for retaining a revised Marxist approach to human geography. Furthermore, it is also clear that many more geographical thinkers continue to rely heavily on a Marxian epistemology of some form, if not wedded to the concepts of any distinct school of thought within what can be labeled classical, neo or post Marxism. Sayer's argument that much empirical, social, scientific research is informed by a loose post Marxism remains true for human geography. In that sense, to reiterate Laclau and Mouffe's terminology, much Marxist geography is more post 'Marxist'. The discipline therefore remains to a large extent sympathetic to Derrida's (1994) proposition that there is a continued need 'to take inspiration from a certain spirit of Marxism…which must not be renounced'. This is the spirit of Marxian analysis not "riven to the body of Marxist doctrine [or] to its supposed systematic, metaphysical, or ontological totality," but to 'its fundamental concepts of labor, mode of production, and social class'. However, while this spirit may live on in human geography, this has not prevented often vigorous debate and disagreement with what forms it should take.

In conclusion, the future of Marxist geography might thus be argued to be both in one sense assured, and in another sense, precarious. The likelihood that a weak Marxian epistemological framework will inform much future human geographical thinking seems highly likely. Marxist concepts still represent a common and widely used conceptual lexicon for most human geographers. Furthermore, given their diversity, it has been possible for different strands of Marxist thinking to evade (or even answer) the critiques discussed earlier. This clearly presents the opportunity for more plausible post Marxisms to arise and explain why Marxist geographical thought is far from dead. However, whether the stronger revised Marxist geographies of David Harvey or Neil Smith will continue to weather the storm of post structuralist critique seems more uncertain. The attempts to reconcile these approaches by thinkers such as Noel Castree have undoubtedly further complicated this debate, but have not ended the disagreement. Castree may be correct to argue that the differences between Marxism and post structuralist approaches such as ANT have been overstated but others remain skeptical. As the entry on *poststructuralism/post structuralist geographies* discusses, much post structuralist thought in human geography remains at odds with revised Marxist epistemologies. Furthermore, a significant strand of recent influential thought in the subject – such as that of Massey – appears to be reaffirming the argument against the kind of metatheory that Marxian analysis represents. Yet what kind of radical human geography, that can fully address Marxist geographies' politically engaged theoretical and progressive goals, can emerge from post structuralist geographical thought remains as yet also unclear.

See also: Actor-Network Theory/Network Geographies; Dialectical Reasoning and Dialectical Materialism; Historical-Geographical Materialism; Marxism/Marxist Geography I; Poststructuralism/Poststructuralist Geographies; Radical Geography; Structural Marxism.

Further Reading

Amin, A. (1994). *Post Fordism: A Reader*. Oxford: Blackwell.

Baudrillard, J. (1990). *Cool Memories* (trans. Turner, C.). London: Verso.

Castells, M. (2001). *The Rise of the Network Society* (2nd edn.). Oxford: Blackwell.

Castree, N. (2002). False antithesis? Marxism, nature and actor networks. *Antipode* 111 146.

Cohen, G. (2000). *If You're an Egalitarian, How Come You're So Rich?* Cambridge, MA: Harvard University Press.

Corbridge, S. (1989). Marxism, Post Marxism and the Geography of Development. In Peet, R. & Thrift, N. (eds.) *New Models in Geography*, pp 224 256. London: Unwin Hyman.

Cox, K. (2005). From Marxist Geography to Critical Geography and Back Again. http://geog www.sbs.ohio state.edu (accessed July 2008).

Deleuze, G. and Guattari, F. (1973). *Anti Oedipus: Capitalism and Schizophrenia*. London: Athlone.

Deleuze, G. and Guattari, F. (1982). *A Thousand Plateaus*. London: Athlone.

Derrida, J. (1997). *Specters of Marx*. London: Routledge.

Escobar, A. (1994). *Encountering Development: The Making and Unmaking of the Third World*. Princeton, NJ: Princeton University Press.

Foucault, M. (1980). In Gordon, C. (ed.) *Power/Knowledge: Selected Interviews and Other Writings*. Brighton: Harvester.

Gibson Graham, J. K. (1996). *The End of Capitalism (as we knew it)*. Oxford: Blackwell.

Gibson Graham, J. K. (2006). *A Post Capitalist Politics*. Minneapolis, MN: University of Minnesota Press.

Harvey, D. (1989). *The Condition of Postmodernity*. Oxford: Blackwell.

Harvey, D. (1996). *Justice, Nature and the Geography of Difference*. Oxford: Blackwell.

Harvey, D. (1999). On fatal flaws and fatal distractions. *Progress in Human Geography* 23(4), 557 562.

Harvey, D. (2000). *Spaces of Hope*. Edinburgh: Edinburgh University Press.

Hindess, B. and Hirst, P. (1977). *Mode of Production and Social Formations*. Basingstoke: Macmillan.

Johnston, R., Gregory, D., Pratt, G. and Watts, M. (2000). 'Post marxism'. In *The Dictionary of Human Geography* (4th edn.), pp 619 620. Oxford: Blackwell.

Jones, A. (1999). Dialectics and difference: Against Harvey's dialectical post Marxism. *Progress in Human Geography* 23(4), 529 556.

Jones, A. (1999). Notes from the deck of the 'post modern Titanic': A response to David Harvey. *Progress in Human Geography* 24(4), 562 563.

Lacan, J. (2001). *Ecrits: A Selection*. London: Routledge.

Laclau, E. and Mouffe, C. (1985). *Hegemony and Socialist Strategy: Towards a Radical Democratic Politics*. London: Verso.

Latour, B. (1993). *We Have Never Been Modern*. Brighton: Harvester Wheatsheaf.

Lefebvre, H. (1991). *The Production of Space*. Oxford: Blackwell.

Lefebvre, H., Elden, S., Lebas, E. and Kofman, E. (2006). *Henri Lefebvre: Key Writings*. London: Athlone.

Lyotard, F. (1993). *Libidinal Economy* (trans. Hamilton Grant, I.). London: Athlone.

Massey, D. (2005). *For Space*. London: Sage.

McDowell, L. (1991). Life without Father and Ford: The new gender order of post Fordism. *Transactions of the Institute of British Geographers* 16, 400 419.

Peck, J. and Tickell, A. (2001). Neoliberalizing space. *Antipode* 34, 380 404.

Peet, R. (1998). *Modern Geographical Thought*. Oxford: Blackwell.

Rorty, R. (1991). *Essays on Heidegger and Others: Philosophical Papers*, vol. 2. Chicago, IL: Chicago University Press.

Sayer, A. (1995). *Radical Political Economy: A Critique*. Oxford: Blackwell.

Sim, S. (1998). *Post Marxism: A Reader*. Edinburgh: Edinburgh University Press.

Smith, N. (1992). Geography, difference and the politics of scale. In Doherty, J., Graham, E. & Malek, M. (eds.) *Postmodernism and the Social Sciences*, pp 57 79. New York: St. Martin's Press.

Smith, N. (1996). *The New Urban Frontier: Gentrification and the Revanchist City*. London: Routledge.

Swyngedouw, E. (1997). Neither global nor local: Globalization and the politics of scale. In Cox, K. (ed.) *Spaces of Globalization*, pp 137 166. New York: Guilford Press.

Taylor, P. (1992). Understanding global inequalities: A world systems' approach. *Geography* 77, 10 21.

Thompson, E. P. (1976). *The Poverty of Theory*. Penguin: Harmondsworth.

Masculinism

A. Bain, York University, Toronto, ON, Canada

Glossary

Duppy Feminism A version of male feminism as practiced by male academic geographers who are theoretically, philosophically, and practically committed to feminism, but who may fail to fulfill these tenets in all of their work and actions.

Female Masculinity A term used to refer a range of subject positions (e.g., drag king; butch; tomboy; female-to-male) that perform masculinity in different ways and, in so doing, challenge the notion that masculinity can only be attributed to a supposedly self-evident male body.

Hegemonic Masculinity The version of masculinity that is most highly valued, legitimated, and respected in society; it works to disempower women and to subordinate other men.

Masculinism Pervasive patriarchal ideologies of masculine cultural dominance.

Masculinist Rationality A form of knowledge that emerged during the scientific revolution that assumes a knower who is white, bourgeois, heterosexual, and male; and a knower who can separate mind from body to produce supposedly unbiased and objective knowledge.

Men's Studies An interdisciplinary academic field of study that focuses on men's experiences, interests, and issues.

Patriarchal Dividend The general advantage that men gain from the overall subordination of women.

Reflexivity A self-critical, introspective consideration of how the positionality of the researcher and his/her subject(s) and the relations of power between them may influence the research process.

The Gaze A form of knowledge identified by feminists as masculinist for the unidirectional way in which people and things are transformed into objects of consumption by the vision of a viewer in a privileged position of power.

Transgender A term used to describe people whose mental gender does not necessarily conform to their physical anatomy, including transvestites; transsexuals; males-to-females; females-to-males; and intersexuals.

Introduction

Research in geography on gender was initiated by feminist geographers in the mid 1970s. Much of this early work drew on feminist politics and theories to examine how gender relations are structured and transformed across space. Particular attention was directed toward developing a critical discourse about women's oppression in society and an understanding of the ways in which women's oppression has been perpetuated in the development of geographical theory and knowledge. In subsequent decades, feminist geographers challenged the notion of universal womanhood and sought to theorize the ways in which multiple and intersectional axes of difference – class, racialization, citizenship, sexuality, age, religion, and ability – fractured identities and produced complex geographies. Out of the ensuing celebration of diversity and difference within feminist social theory came a recognition that gender is not merely an attribute of femininity and therefore only of interest to women academics and women students; men and masculinities are just as gendered as women and femininities. Consequently, the feminist debate was broadened to not just indirectly imply, but rather to directly engage ideologies of masculinity. Some feminist geographers in conjunction with feminist male scholars within other disciplines began to explicitly focus their research attention on men, masculinity, and men's powers, practices, and identities. This effort to open up lines of scholarly debate to include analyses of the meanings and enactments of dominant identities that have typically been taken for granted (e.g., whiteness and heterosexuality) could also be seen in other academic fields. Through the 1990s, a geographical, empirically based literature emerged and coalesced around theorizations of the constructions, performances, and practices of men and masculinities.

Theorizations of Men and Masculinities

Outside of the discipline of geography, theorizations of masculinity have been associated with the establishment in the late 1970s of men's studies as an interdisciplinary academic field of study variously regarded as a complement or a challenge to women's studies. Men's studies were established in conjunction with the men's movement to respond to the perceived advantages that women had received in their struggle for equality through feminist political action and the accompanying loss of male privilege and clear gender roles. The men's movement (variously composed of: the antisexist men's movement; the men's rights movement; and the mythopoetic men's movement) is a consciousness raising effort to encourage men to name their experience as 'masculine' and to affirm their male gender and sexual identities. A network of men's groups and related journals and magazines has gradually developed around this movement.

Critical research in men's studies has grown dramatically in the last two decades, dominated, in particular, by research produced in the disciplines of sociology and psychology. It has focused on men's experiences, interests, and issues and has opened a range of topics to political, media, academic, and policy debate that include men's relation to economic restructuring; new nationalisms; health, life course, and household changes; fatherhood; consumption patterns; educational underachievement; sexualities; and violences. These debates have been studied in different disciplinary contexts (e.g., sociology, psychology, geography, social policy, political science, cultural studies, literature) and have raised fundamental questions about men and masculinities. Questions such as: what is a man? How do men obtain and maintain power? How do men embody and perform masculinity? Is there a crisis of masculinity? These sorts of fundamental questions have helped to render men more gendered in theoretical and practical terms, which in turn has created opportunities to challenge previously taken for granted assumptions about men's power, authority, and social practices. Men's studies has made an important contribution to gender research in geography in that it has helped to foster the interrogation of masculinity from different perspectives.

A touchstone for much of the work on masculinities, both within the discipline of geography and without, is research by the Australian sociologist Robert Connell who has written widely on the multiple ways of being masculine. As an important male masculinity theorist, Connell has played a significant role in advancing understandings of the ways in which masculinity is (re)constructed within the context of historically and geographically differentiated social and gender relations. Connell helped to develop the now widely used term 'hegemonic masculinity' – a term which itself has become hegemonic. Drawing on Antonio Gramsci's concept of 'hegemony', which refers to the cultural dynamics whereby a group claims and maintains a dominant social position, Connell defines hegemonic masculinity as the version of masculinity that is most highly valued, legitimated, and respected in society. Connell argues that in contemporary Western societies, it is rational, middle class, and heterosexual masculinity that is the normative standard or the dominant form of masculinity against which other forms of masculinity are subordinated. Hegemonic cultural norms of what it is to be a man tell men what is appropriate clothing, conduct, body shape, values, and aspirations if they wish to be recognized in society as suitably masculine. Culturally sanctioned ways of being male are generally associated with evidence of power, physical strength, authoritativeness, confidence, and (hetero)sexual prowess. While few men may actually meet the normative standard, he suggests that the majority of men actually benefit from its hegemony because they profit from the 'patriarchal dividend', the general advantage that men gain from the overall subordination of women. Connell goes on to explain that hegemonic masculinity is a constructed cultural and economic force that is both a personal lived experience and a collective project supported by different institutions (e.g., the state, the workplace, and the school). However, hegemonic versions of appropriate ways of being masculine can be destabilized, Connell suggests, by changes in social norms and mores. Dissenting masculinities, such as those performed by men who incarnate traits such as compassion, affection, kindness, cooperation, and homoeroticism in their everyday lives, can challenge hegemonic masculinity. The idea of hegemonic masculinity is important, particularly when coupled with the notion of the pluralization of masculinities, because it emphasizes that despite the variety of different ways of doing masculinity, essentialized ideas about what it means to be a man remain deeply embedded and profoundly influential in society.

Geography of Masculinities

Although the discipline of geography has historically been dominated by men, critical analysis of masculinity, particularly that which is produced by men, has been slow to develop within the scholarly geographical literature. It was not until the early 1990s, two decades after the rise of feminist geographies and its accompanying interest in gender and sexualities, that a geography of masculinities began to emerge. Influenced by the work of Connell, numerous geographers, particularly those working in the feminist cultural geographical tradition, began to explore variations in masculinity across time and space. After more than a decade of research, it could be argued that work on the geography of masculinities has reached a critical mass.

Peter Jackson is widely recognized for the central role he played in laying a solid foundation for a geography of masculinities. He was the first geographer to explicitly advocate and detail the intellectual necessity and political utility of men undertaking a sustained critique of men. He encouraged geographers to explore the contradictions inherent in masculinity in order to reveal the spatial structures that support dominant forms of masculinity. To this end, he outlined an agenda for 'mapping masculinities' that would aim to document and to challenge the social construction, and the temporal and geographical contingency of masculinity through an examination of the relationship between masculinity, economic activity, and identity formation; the role of the public and private sphere in varying expressions of masculinities; gender construction at different spatial scales; homophobia; child abuse; and male organizational fraternalization. Certainly, the relationship between masculinity and economic activity has become a significant area of empirical investigation within the discipline of geography. Inspired by the work of Jackson and Connell, Linda McDowell has played a crucial lead role in the development of geographical research examining the intersections of masculinity, work, and class. Additional geographic research by Doreen Massey as well as Glendon Smith and Hilary Winchester has further confirmed the significance of work as a key place for the reproduction of men's power and masculinities.

More recent geographical research, such as that by Alison Bain, has focused on men in families, and in particular, on fathers and fatherhood. It draws on a broader interdisciplinary body of literature written from a legal, a family studies, and a health science perspective that approaches fatherhood as a social construction and explores the ways in which men's roles and authority within the family have shifted through Western history. There remains substantial scope to further explore the geographies of men in families – the gender division of domestic labor, parenting, and emotional work – particularly through a more diverse lens that examines differences across cultures, classes, and sexualities. Within the geographies of sexualities literature, the significant emergent topics of transgender and female masculinity have much to contribute to any discussion of masculinities, a discussion that also needs to address the Anglo American bias in much of the work on the geography of masculinities.

The names of female geographers mentioned in previous paragraphs suggest that a noteworthy amount of geographical research on masculinities has been undertaken by women. However, Peter Jackson's original request was for men to take advantage of their insider perspective to undertake work on men. Male geographers, as Jeff Hopkins has documented, have been slow to respond to Jackson's request. Nevertheless, important exceptions include the work of Lawrence Berg, Alistair Bonnett, Peter Hopkins, and Richard Phillips as well as significant work on homosexual masculinities by David Bell, Jon Binnie, Gavin Brown, Michael Brown, and Larry Knopp. When speculating on why critical studies by men on masculinism have been rather limited relative to the number of males in the discipline, Jeff Hopkins suggests part of the problem could be generational lag (i.e., older, established male geographers pursuing research interests that they may have established when sexism went unchallenged); credibility concerns for heterosexual males; disinterest; and reluctance to challenge a privileged position in society in a way that would erode access to power. Clearly, there is scope within geography for men and women of different identities and philosophical perspectives to produce work that is written and read from a perspective that is critical of masculinism.

Masculinism in Geography

Simply stated, masculinism has been conceptualized by geographers as pervasive patriarchal ideologies of masculine cultural dominance. It should be noted that it can be difficult for men and women to identify masculinism and to challenge it in practice. One of the first usages of the term masculinism in geography was by Gillian Rose in her book *Feminism and Geography*. In this groundbreaking book about the gender of geography, Rose powerfully argues that the geographic tradition has been socially and ideologically structured by male dominance – a set of processes that she unites under the label 'masculinism'. She persuasively illustrates how masculinism reveals itself not only in the topics that geographers choose to study, the conceptual frameworks that geographers employ in their research, and the epistemological claims to exhaustive knowledge that geographers make, but also in the career structure and teaching strategies of the discipline. Rose boldly states that the discipline of geography is masculinist. By this she means that geographic knowledge has been distorted by false objectivism, authoritativeness, and exhaustiveness as well as patronizing views of women and non Western people. As Lawrence Berg has pointed out, the attachment of geography to objectivity can still be seen in the blind peer review process for publication in scholarly journals – an anonymous review process is often required in which author and referee remain unknown to one another. In this process it is believed that objectivity equates with impartiality, detachment, disembodiedness, and anonymity.

Masculinist work, feminists argue, is not pluralist, sensitive, or dynamic; instead, it excludes, marginalizes, and silences other subjectivities and interpretations in the process of knowledge production about the world. It does so, feminists maintain, through a reliance on 'masculinist rationality' and the power of 'the gaze'. Masculinist rationality is a form of knowledge that emerged during the scientific revolution and the Enlightenment. Masculinist rationality assumes a knower who is white, bourgeois, heterosexual, and male; a knower who can separate his mind from his body; a knower who is not influenced by his experiences, values, passions, or emotions; and a knower who considers his thoughts to be unbiased and objective. This masculinist knower never problematizes his own positionality nor considers the potential partiality of his perspective. Consequently, he can comfortably assume a master subject position that allows him to confidently generalize about the world. Such an epistemological position is termed masculinist, not because it is inherent or essential to men, but because of the historical link between value neutral objectivist social science and its embodiment in white, bourgeois, hetero sexual males.

The gaze has also been identified as a masculinist form of knowledge and power by feminists. The gaze refers to the unidirectional process by which people and things are transformed into objects of consumption by the vision of a viewer in a privileged position of power. In relations of looking, the viewer has the power to control what is seen and how it is seen. Feminist geographers have argued that landscapes, women, and non Western people have been represented as feminine and passive by the desire, pleasure, and power of the active male gaze. The power of the masculine gaze to seek to render space and people transparent is central to Rose's exploration of the mas culinism of contemporary geographical discourse.

Rose uses a critique of time geography and humanistic geography in order to demonstrate how diverse masculine subjectivities have influenced the production of academic geographical knowledge. Time geography, as developed by the Swedish geographer Törsten Hägerstrand, uses representational diagrams of the everyday paths taken by individuals through time and space to reveal the tem porospatial structuring of social life. Rose swiftly reveals the hegemonic masculinity inherent within time geo graphy's minimalist rendering of the body as a neutral vessel without skin color, sexual desire, or emotion and its minimalist rendering of space as an infinitely knowable, transparent, and freeing medium of social life. The re search methods employed by time geographers are characterized by Rose as masculinist for the way in which researchers take spaces and bodies for granted as universal and, in so doing, exclude different subjectivities and socialities from knowledge. However, this need not ne cessarily be the case. Inspired by Rose's critique, Mei Po Kwan has productively demonstrated through her 'body maps' how geographic information systems (GISs) methods can be used to visually reveal aspects of women's everyday lives. She proposes reimagining lines repre senting women's life paths in space–time not as abstract lines in transparent Cartesian space, but as body in scriptions of oppressive power relations that express women's corporeality and embodied subjectivities. Kwan persuasively argues that the representational possibilities of geospatial technologies (GT) (e.g., GIS, global pos itioning systems, and remote sensing), when combined with geospatial practices that consider emotion, feelings, values, and ethnics, can offer creative possibilities for challenging the masculinism of time geography.

Rose goes on to explain that the masculinism inherent in humanistic geography is more paradoxical. She char acterizes the research methods employed by humanistic geographers as masculinist for the way in which researchers reject scientific rationality and dualisms and superficially embrace personal emotion and self reflexivity as a means to bolster their own power and knowledge claims about place. The masculinism of humanistic geography, Rose maintains, is apparent in its authoritative and exhaustive knowledge claims and disinterest in the broader social power relations of exploitation and oppression that struc ture experiences of places. The fundamental geographical concept of place, Rose argues, is theorized in humanistic geography in terms of an implicit masculine norm that ignores the possibility of other experiences of place.

For Rose, feminist geography provides a direct chal lenge to masculinism and a vision of a different kind of geography – a geography that is not based on exhaustive claims to knowledge nor on the exclusion of other mo dalities of knowing nor dependent on a relationship of dominance and subordination, but rather a geography based on a celebration of differences and possibilities. Rose advocates that men listen to, read, discuss, and learn from feminism. Some male geographers have done just that. David Butz and Lawrence Berg developed the term 'duppy feminism' to describe male academic geographers who are theoretically, philosophically, and practically committed to feminism, but who may fail to fulfill some aspect of these tenets in all of their work and actions.

Jeff Hopkins also engages with Rose's proposition by advocating for a critical cultural geography of men and masculinities that is informed by feminist thought and methodological practice. A critical cultural geography of masculinities, he proposes, would take as its ideological cornerstone the notion that gender inequality exists and that gender imbalance in social relations should be eradicated. Its ultimate goal, however, would be to challenge multiple forms of masculinism. As such, it would confront patriarchal and chauvinistic forms of masculinity that women as well as men (particularly those men who do not conform to unwritten codes of manliness) find oppressive and exploitative. It would take an anti homophobic and an anti mysogynistic stance that recognizes that multiple male sexualities and identities exist and must be respected. And, lastly, it would inter pret knowledge as a social and embodied process that is partial, and constituted by and inscribed with power re lations and emotions.

Challenging Masculinist Research Practices with Feminist Methodologies

Feminist geographers have used a range of different in clusive research methods to investigate a variety of re search questions. Consequently, there is general agreement that there is no singular feminist method; rather, there are key characteristics that form the foun dational building blocks of feminist methodologies. Feminist research methodologies have been set up in direct opposition to masculinist research practices.

First, feminist research has emancipatory goals. Where masculinist research may work 'for' particular social groups, feminist research aims to work 'with' women and other marginalized groups to validate and to give voice to their experiences in ways that challenge social inequalities and promote social transformation of gendered power relations. Thus, feminist research seeks to be relevant, accessible, and participatory.

Second, feminist research recognizes that all knowledge is socially constructed from a particular subject position – it cannot always be fixed and known through detached observation and testing. Feminist researchers seek to dismantle the hierarchies of power relations between researcher and researched inherent in masculinist research; they do so by fostering open and emotional relationships with research subjects that challenge the top down establishment of research priorities, questions, and knowledges associated with masculinist positivist science. Feminist researchers often employ collaborative and nonexploitive methods (e.g., intensive interviewing, participant observation, recording of personal narratives, and life histories) that provide ready access to individual's voices and more intimate micro spatial scales. Feminist researchers also engage in a process of self reflexivity in order to acknowledge some of the ways in which social location influences power relations, perceptions, and interpretations. In this way, the situatedness and partiality of knowledge is celebrated in place of objectivity.

Third, feminist research redefines 'the field' within fieldwork to include the politics of everyday worlds and intimate, familiar settings (e.g., body, home, workplace) that researchers regularly inhabit. Vision driven, experiential fieldwork as a research method and a pedagogical tool is central to the practice of academic geography. Undergraduate students are taken on field trips as a means of introducing them to geography, a discipline once characterized by Carl Sauer as a science of observation. Experience 'in the field', the place where geographers 'go' to 'do' research, constitutes an important part of the professional identities of geographers – the experience can confer a sense of authority, legitimacy, and belonging to the discipline. Traditional fieldwork, however, has been characterized by feminists as a masculine enterprise within geography for several reasons. First, fieldwork is often described in heroic terms as a 'character building rite of passage' into the academic world that involves great stamina and intellectual discipline. Second, the field has frequently been treated as a wild place to be secured and kept separate and contained by the individual fieldworker. Third, fieldwork is rooted in masculinist notions of objectivity – an unbiased observer ensuring analytical distance from the object of study in order to collect and interpret reliable data that in turn produces reliable knowledge. Feminist researchers maintain that they are always 'in the field' and by rendering the personal political they aim to participate in a continual process of working toward social change.

Methods of Researching Men and Masculinities

In light of the substantial body of writing by feminist geographers on how to undertake research on gender, it is worthwhile considering how geographers have sought to understand different cultures of masculinity. It would be fair to argue that while men and masculinities have become research objects in geography, the ways in which research on men and masculinities has been undertaken by geographers have remained largely unexamined. Questions that could use consideration include: Are there particular methodologies that are more appropriate than others for studying masculinities? Are there particular writing strategies that are more appropriate than others for representing the voices of men so as to neither malign nor romanticize them?

A survey of the geographical literature published in the last decade on masculinities reveals that geographers have relied quite heavily on representations of men and masculinities in a variety of media (e.g., film, television, magazines, newspapers, song lyrics, art, and advertising) as a source of data. A second significant source of data is semistructured interviews. Whether studying merchant bankers, engineers, gay men, young male school leavers, college students, househusbands, or fathers, interviews remain one of the most common research methods used by geographers to elicit the views and opinions of men of different ages, ethnicities, sexualities, religions, and occupations. When young male academics are studying young men, single sex focus groups and participant observation have also been used. Within this diverse empirical literature on the geographies of masculinities there remain only a handful of instances in which geographers have reflected in any substantial, meaningful, and critical way on the process of conducting research on and with men – particularly noteworthy is work by Linda McDowell and Robert Vanderbeck.

One of the important concerns to be raised by geographers of masculinities relates to interpreting men's silences on particular issues. Masculinity is a topic that people may know about intuitively, but might be unable to comfortably articulate or define clearly. Masculinity and manhood are multidimensional, fluctuating constructs understood in relation to femininity, womanhood, and a system of gender practices; they are not something to be measured and/or cataloged. When asked to discuss particular personal and emotional issues relating to understandings of masculinities, some men may not say very much or they may have learned disdain; but that does not mean that more is not going on below the conversational surface. Within a dominant male culture, it may be difficult for men to acknowledge and relate to questions about emotions such as sadness, fear, vulnerability, or powerlessness. It is, therefore, imperative that researchers develop novel approaches for better exploring personal and emotional issues that men may learn to keep private. In instances where men may not be particularly verbally adept, it is useful to reflect on what strategies researchers can use to penetrate men's silences and to encourage them to talk without putting words in their mouths. Repeat interviews are one strategy that some geographers have found helpful because they provide the time necessary for individuals to relax and to build up trust and to reflect on and to revisit topics in conversation. In interviews, it is important that researchers not succumb to the temptation to take at face value men's accounts and testimonies as revealing particular masculinities without getting at how and why men come to feel and think about themselves in particular ways. It is also important to provide opportunities to explore the tensions that men may feel in living out different masculinities, tensions that might not be readily apparent in an interview setting. Interviews rely on the spoken word as the primary means of expression and communication where other more visual, written, or ethnographic methodologies might be more effective at revealing situated gender practices, performances, and identities.

Another significant concern raised by geographers of masculinities relates to how researchers (re)write themselves into their fieldwork narrative. In a thought provoking article, Robert Vanderbeck reflects on how male researchers have negotiated masculinities during fieldwork. He draws on his own experience conducting ethnographic research as a volunteer. He discusses the experience of not fitting in, of not being able to easily establish a sense of male camaraderie, and of not conforming to hegemonic gender ideals. His masculinity, for example, was critiqued by informants who frequently scrutinized elements of his bodily performance – the timber of his voice, his diction, his mannerisms, and his attire – and dismissively ascribed a homosexual label. This experience, Vanderbeck documents, undermined his confidence and credibility as a fieldworker. Vanderbeck insightfully concludes from a review of the social science fieldwork literature that the range of masculinities expressed therein is limited and that male social scientists tend to write about themselves in ways that reinforce their own hegemonic masculine position.

To conclude, the degree of self reflexivity apparent in writing by male geographers who research masculinities is limited. It is too easy, as Matthew Sparke explains, to assume the role of 'an academic male tourist' who makes bold 'as a ism' announcements of the 'straight white man variety' rather than interrogating the complex contradictions of one's own positionality. Admittedly, there is the danger that too much reflexivity can be self indulgent, and little more than public privileging of the researcher's masculine prowess. Notwithstanding, there is certainly scope for researchers of masculinities to explore in a more sustained and critical manner the masculinism that is deeply embedded in the research process, with an eye to responsibly transforming the way in which research on men is undertaken in the discipline of geography.

See also: Feminist Methodologies; Masculinities.

Further Reading

Bain, A. L. (2007). Claiming and controlling space: Combining heterosexual fatherhood and artistic practice. *Gender, Place and Culture* 14, 249 266.

Berg, L. (2001). Masculinism, emplacement, and positionality in peer review. *Professional Geographer* 53, 511 521.

Berg, L. and Longhurst, R. (2003). Placing masculinities and geography. *Gender, Place and Culture* 10, 351 360.

Butz, D. and Berg, L. (2002). Paradoxical space: Geography, men and duppy feminism. In Moss, P. (ed.) *Feminist Geography in Practice: Research and Methods*, pp 87 102. Oxford: Blackwell Publishers.

Connell, R. W. (1995). *Masculinities*. Berkeley, CA: University of California Press.

Hopkins, J. (2000). Signs of masculinism in an 'uneasy' place: Advertising for 'big brothers'. *Gender, Place and Culture* 7, 31 55.

Jackson, P. (1991). The cultural politics of masculinity: Towards a social geography. *Transactions of the Institute of British Geographers* 16, 199 213.

Kwan, M. P. (2002). Feminist visualization: Re envisioning GIS as a method in feminist geographic research. *Annals of the Association of American Geographers* 92, 645 661.

Kwan, M. P. (2007). Affecting geospatial technologies: Toward a feminist politics of emotion. *The Professional Geographer* 59, 22 34.

McDowell, L. (2001). Working with young men. *Geographical Review* 91, 201 214.

Rose, G. (1993). *Feminism and Geography: The Limits of Geographical Knowledge*. Minneapolis, MN: University of Minnesota Press.

Sparke, M. (1996). Displacing the field in fieldwork: Masculinity, metaphor and space. In Duncan, N. (ed.) *BodySpace: Destabilizing Geographies of Gender and Sexuality*, pp 212 233. London: Routledge.

Sundberg, J. (2003). Masculinist epistemologies and the politics of fieldwork in Latin Americanist geography. *Professional Geographer* 55, 180 190.

Vanderbeck, R. (2005). Masculinities and fieldwork: Widening the discussion. *Gender, Place and Culture* 12, 387 402.

Masculinities

B. van Hoven, University of Groningen, Groningen, The Netherlands
P. Hopkins, Newcastle University, Newcastle, UK

Glossary

Hegemony The process whereby particular behaviors and characterstics become dominant through subtle everyday practices and processes which encourage subordinate groups to accept and take on certain dominant qualities and attitudes leading to the marginalization and exclusion of particular practices, attitudes, and values.

Identity Markers of social group membership – for example, class, race, age, and ability – that are written on people's bodies, enacted through everyday performances and contested through political action.

Ideology A set of values or beliefs which (re)produce behaviors, attitudes, and ideas about how people should (or should not) live their lives.

Imaginative Geographies Representations of places that are a result of images and imaginings. Edward Said discusses the uneven power relations evoked and emphasized by such representations in his critique of representations of the Orient.

Masculinism Theories and practices that prioritize the experiences of men and reproduce power structures advantageous to men as a group are called masculinist. Masculinist (political) movements, however, focus on disadvantages men experience in society, for example, regarding custody or criminal sentencing.

Sex Role Theory Sex roles are seen as the prototypes into which men's and women's identities are forged through socialization. Men are measured by their success or failure to comply with social norms of masculinity, such as to be strong, successful, capable, reliable, and in control.

Introduction

Masculinity has been defined both as a form of identity (i.e., a process of identification whereby particular traits, personal qualities, and attributes are seen to be masculine (or not)) and as a form of ideology (i.e., a doctrine or set of beliefs dominant in institutions, society, and culture at large). As such, masculinity is a social, historical, and political construct. Since the performance of masculinity interrelates with space, place, and time, that is, the same person can reveal a different kind of masculinity and/or femininity at different times and in different contexts, it is more appropriate to speak of masculinities.

Just as work about whiteness and heterosexuality was brought to the attention of scholars interested in race and ethnicity and sexuality respectively in the 1980s, interest in men and masculinities became a topic of interest among academics whose work focused upon gender studies. Research on masculinities often departs from Robert Connell's writing on masculinities in general and hegemonic masculinity in particular. Connell's own research in high schools in 1982 laid the foundation for his thinking about the production and experiences of masculinities and hierarchical structures within these. Following an 'inventory' of writings on masculinity, Connell published a concept for 'hegemonic masculinity' which he defined as actions and practices that permitted and continued men's domination over women. Hegemonic masculinity was achieved through culture, institutions, and persuasion and was changeable over time (and space).

Connell and Messerschmidt acknowledge work on masculinities that preceded their own, in particular work within psychoanalysis (including Freud's work on the Wolf Man) and work drawing on sex role theory which sought to appreciate the social nature of masculinity. Elsewhere, the period after World War II has often been regarded as a time of change for men which caused a conflict in role demands and expectations. This was largely due to women's entry into the labor market, and expectations for men to be able to economically support their family, be (physically) strong figures of authority, and adopt forms of behavior in the service sector traditionally associated with women. The problem for men, then, was that they had no alternative within the dualistic structure of masculine–feminine, the feminine being inferior. Key critiques of the sex role theory (by feminist scholars) included the lack of attention for structural analyses of power relations, the dominance of middle class, white, heterosexual males as norm and its ignorance to issues of resistance and history. Connell's work on hegemonic and multiple masculinities provided a new framework for the study of masculinities within the context of gender studies.

Work on (hegemonic) masculinities has received increasing critique in recent years, contending that the concept of masculinity essentializes male–female difference; it naturalizes the body rather than seeing it as discursively produced. The concept essentially makes the subject invisible. More recently, then, social scientists, including human geographers, have spent much time exploring and emphasizing the complex ways in which masculinities are socially constructed, produced, con sumed, and performed.

The Origin of the Study of Masculinities in Geography

Retracing the emergence of the first published work on masculinities in geography, Peter Jackson's work is often regarded as setting the groundwork for later research. The motivation by geographers to engage with mascu linities has been varied and includes both personal and political reasons with some seeing the work as part of the shifting focus of feminism and others regarding a focus upon men and masculinities as a response to feminism. The fact that most articles on masculinities within geography – in particular the earlier work – were pub lished in the feminist journal *Gender, Place and Culture* also supports this. But perhaps one (a shifting focus in femi nism) does not necessarily exclude the other (a response to feminism) for many geographers studying mascu linities draw on feminist scholarship. It is worth drawing attention to some of the purposes male geographers – who are still in a relative minority in the emerging field of geographies of masculinities – have stated for studying masculinities. Peter Jackson has noted that the feminist and gay movement agenda of the late 1980s aimed to redress the balance of power between men and women, but that those men who view themselves as more emancipated and who wish to challenge male supremacy find little guidance and support in literature. Lawrence Berg describes his starting point for discussing mascu linities within the binary discourse of theory and empirical investigation in New Zealand as follows: "As geo autobiography, [this article] represents my views of human geography in Aotearoa (New Zealand), based on my experiences as a Canadian postgraduate student undertaking doctoral studies at a New Zealand university [...] One of my purposes is to resist this hegemonic construction of masculinity and thereby reduce my participation in the (re)constitution of masculinist geographies" (Breg, 1994: 245–261).

More recently, van Hoven and Hörschelmann high lighted the contested source of interest in men and masculinities in their introduction to *Spaces of Mascu linities* by acknowledging that geography had been a masculinist discipline but one with little attention to men's experiences as men. Although some current work on masculinities is also political, most literature deals with men's experiences as men addressing a wide range of concerns about men's everyday lives, masculine iden tities, and gendered performances.

Over 15 years now since Jackson's landmark inter ventions, work about the geographies of men and mas culinities is reaching a critical mass in human geography. Jackson's interest in masculinities emerged in the context of calls for new emphases within cultural geography. Endorsing earlier work by humanist geographers, Denis Cosgrove and Peter Jackson sketched 'new directions in cultural geography' (with the 'old' largely being repre sented by Carl Sauer's work). A key critique was the lack of plurality in approaches to culture and a lack of at tention to ideas of hegemony and resistance. Cosgrove and Jackson's own work adopted interpretive approaches drawing on social anthropology and cultural studies. Bringing together his attention about plurality and power relations, and the significance of representational spaces, Jackson produced the first studies on masculinities in geography. Through exploring representations of men in advertisements, Jackson points to differences in social constructions of (ideal) masculinities over space and time. Using an example from a study in a textile mill in West Yorkshire, Jackson also illustrates the effect of representations, or imaginative geographies, on the ma terial spaces of both white female and black male mill workers. This led Jackson to the conclusion that in de veloping a social geography of gender, the construction of historically and geographically specific forms of mascu linity requires mapping in order to identify potential lines for political struggles. The spatial structures that underpin and reinforce contemporary patriarchal gender relations should also be a part of this mapping.

Developments within Geographies of Masculinities

Given the extent of current work in geography on mas culinities, we use this contribution to offer a flavor of some of the important contributions to understandings in this field, rather than offering a comprehensive overview.

Representation and Empire

Many of the earlier contributions to the study of mas culinities in geography also address representational spaces, specifically addressing artistic and cinematic representations and advertising, for example, advertise ments in magazines, or advertisements in the context of campaigns (such as the *Big Brother* campaign or army recruitment). Here, we draw attention to the work on representations in the context of the British Empire. *Soldier Heroes* by Dawson and *Mapping Men and Empire* by Phillips are books that explore narratives of adventure, Empire, and masculinity. Based on a range of novels depicting adventures of the imperial frontier, Phillips provides a description of patriarchal social relations embedded within these. Dawson approach is more nar row in that it focuses on two men (Havelock and

Lawrence) and deeper because it focuses on the ways in which these men are represented in several cultural forms, for example, journalism, biography, fiction, and film ranging from Victorian times until the present. More recently, though, others have drawn attention to the impact of the adventure story, as ways to promote man liness, on nonfiction, that is, the writing of the plantation in mid nineteenth century Ceylon. This work highlights that where the realization of the imperialist adventure was constrained by the material conditions of place, it was the struggle with rather than the conquering of the elements that counted in aiming to achieve manliness. Yet, for certain narratives to make sense, and for the struggle to be successful, the separation of male and fe male spheres, mentally and/or physically, was required.

Masculinities in Transition?

That this kind of artificial separation of spheres is not a thing from the past transpires from work which is set in more recent times. In studies on places in transition, insights are offered into how a separation of spheres is used to make sense of and retain (patriarchal and mas culinist) narratives. Examples discussed in *Spaces of Masculinities*, for instance, are from Lebanon and India. The way in which men in Lebanon, whose position in the family had been challenged due to urbanization pro cesses, aimed to resolve contradictions in the family, at least temporarily, was to draw on cultural values of honor and shame during the Lebanese civil wars. A discourse was created that blamed a hated Other for the misery of urban life and redefined men as protectors of 'their' women, thus justifying the repression of women. In India, (middle class) men have begun to aspire to Western masculinities as portrayed through the global media. Although this contrasted with more traditional norms and expectations of their families, they managed to re solve anxieties by emphasizing that male identities are rooted in public cosmopolitan spaces. The role of women in maintaining this role for men, then, was to perform home based religious and familial duties and becoming bearers of tradition. The participation of women in eth nic violence in Lebanon or globalization in India, or a focus on perhaps men's role in the home sphere, would have destabilized the narrative of the male 'hero'. Again, there are parallels with work on men and Empire. The power of the state and state institutions to structure gender relations and to (attempt to) fix gender identities emerges in the above examples and others such as the military, the prison, or the police. Even at smaller scales, studies have addressed ways in which institutions struc ture gender relations and construct gendered identities as required opposites. This poses all kinds of problems for those, for example, entering the labor market.

Linda McDowell's work on working class men has illustrated some of the everyday experiences of young men in times of socio economic transition and change, in particular in reference to what has been called second modernity. Importantly, she illustrated the impact of structural factors on young people's personal biographies. McDowell's work broadens and deepens literatures on issues facing working class men which include studies about gangs bikers, hooligans, yobs, and lads. McDowell's research involved a series of interviews with working class young men, school leavers from Cambridge and Sheffield, who were seeking employment under the newly emerging conditions of second modernity. Rather than developing personal biographies marked by oppor tunities and choices, though, McDowell found their ac cess labor effected by social and spatial differences. One of the main findings of the research was the desire of the young men to achieve 'domestic conformity', a version of 'working class respectability'. The young men consulted were continually involved in a process of constructing themselves as masculine, a construction that places a great deal of emphasis on waged labor.

McDowell's work has introduced feminist approaches and masculinities to economic geography and thus il lustrated the value of such approaches to other sub disciplines within geography. Little and Leyshon have made a similar argument for rural geography by noting that research has addressed the ways in which mascu linity and femininity are constructed and performed in the context of rural society and culture but that rural embodiment has remained underexplored. The role of nature in the construction and representation of both dominant and marginal rural masculinities has been the focus of a variety of work. The masculine identities of young men living in rural Australia ('beyond the me tropolis') have recently been researched by Jane Kenway through ethnographic work in rural areas. This research identified three main groups of young men according to the leisure pursuits and practices they engaged with on a daily basis: sacrosanct, subversive, and scorned mascu linities. Sacrosanct masculinities were performed by men who participated in a range of 'embedded and endorsed leisure/pleasures' such as activities with motor vehicles and playing football. Subversive masculinities were those associated with 'unendorsed leisure/pleasures' such as those associated with participation in specific forms of sport such as skateboarding. Finally, scorned mascu linities were inhabited by men who participated in 'dis embedded and unendorsed' activities such as those connected with music and computers. In exploring these issues, this work provides a strong challenge to metro centric work about masculinities by highlighting how young men in rural areas negotiate global discourses, and local habits and histories in the management and ar ticulation of their masculine identities.

Masculinities and Intersecting Social Identities

Besides seeking to understand how masculinities influence and are influenced by social and economic change, much work in human geography has also looked at the ways in which masculinities intersect with other social identities and the geographical scales that shape and are shaped by these interactions. As noted above, the importance of the body as a site of masculine identity has emerged as focus of much attention. Although Peter Jackson began to discuss race and masculinities in his work on advertising (as has work on postcolonial geographies), and McDowell has highlighted the importance of the well conditioned male body as important factors in the world of finance, it was Robyn Longhurst's projects on 'shitting, shaving, and shampooing' that put men's bodies and processes of embodiment on the map. Some work on disabilities, for example, has drawn attention to how disability can threaten a man's sense of masculinity as they are often reliant on other people and so lack independence.

Work on the geographies of age, too, promises interesting insights into and contestations of masculinities by drawing attention to the ways in which gender shapes the everyday experiences of young people and age shapes the everyday gendered experiences. Research in educational contexts has provided insights into how and why boys 'do' masculinity, and feminist geographers have demonstrated differences in children's geographies (i.e., their radius of activity) depending on their urban or local context, or depending on their location in developing or developed countries. Rachel Pain's recent research about rematerializing risk and fear for children has built on this. She has shown that boys report higher levels of victimization, and whereas girls fears tended to be more localized, boys often feared places more distant such as other neighborhoods (as a result of the greater spatial freedom often afforded to boys) and were more secure in a group environment.

Last but not least, Kathe Browne's research (in *Spaces of Masculinities*, for example) on drag kings who 'do' masculine on stage and women who 'are' masculine, that is, who are mistaken for men, in their everyday lives offers not only an addition to work on embodiment but provides a starting point for research into female masculinities and thus a further destablization of dominant narratives.

Conclusion: Futures of Masculinities in Geography

Earlier, we defined masculinity as a form of identity and as a form of ideology. We discussed ways in which geographers explored masculine identities in different social and cultural context and also drew attention to the ways in which (state) institutions structure gender relations and identities. Much research has drawn on cultural studies and feminist scholarship as a framework of reference. Such research has done much to destabilize dominant narratives of (hegemonic) masculinity and instead argued that there is a mutually constitutive relationship between masculinities and class, 'race', sexuality, age, and ability.

However, although research emphasizes that masculinities are continuously negotiated, and continuously shifting and changing depending on class, 'race', sexuality, age, ability, and location, it seems perhaps less has changed than one might expect. Looking at research over time, for example, it is notable that as early as 1957, in a discussion of the 'new burdens of masculinity', Mayer Hacker emphasized the significance of masculinity to men and the inextricable link of male sexual performance to self worth. In comparison, she noted femininity was less important to women. It emerged from research by Levant *et al.* that although there are some differences in attitudes to masculinity ideology based on geographic location and race, the greatest difference is based on gender. Women then are far less traditional in their views of the male role than men are. Last but not least, Courtenay more recently demonstrated that the resources available for constructing masculinities – and the signifiers of 'true' masculinity – have a negative impact on men's health. Demonstrating masculinity in different ways implies crossing over socially constructed boundaries and risking critique or even physical violence. In addition, in the context of a patriarchal society that awards certain behavioral traits, demonstrating idealized forms of gender helps men assume more powerful positions. Hartley (in Messner, 1998: 155–276) observed "What boy in his right senses would not give his all to escape this alternative to the male role? For many, unfortunately, the scramble to escape takes on all the aspects of panic, and the outward semblance of non femininity is achieved at a tremendous cost of anxiety and self alienation." The original quote stems from the late 1950s but its significance remains unchanged.

Reverting to Jackson's call for mapping masculinities, we might note that a considerable amount of 'mapping' has been achieved and that interesting new lines of research have emerged. But although more now is known about the lived experiences of men of a variety of backgrounds – at least in academic circles – the question arises about how helpful this research has been in changing gender relations in society and the mechanisms of patriarchal, masculinist institutions. The clarification of Jackson's 'lines of political struggle' then, has received somewhat less attention. One line of struggle appears to be our very own work environment. Some discussion – motivated largely by feminist geographers and the exclusion of women from and within the academy – has addressed the conditions of work for geographers already.

Such arguments have addressed the performance of masculinities in academia and the ways in which different forms of gender discrimination as well as the gendered conduct of academic staff work to support men's con tribution and marginalize women. Yet, the organization of panel discussions on the institutionalization of gender and the associated normalization of masculinities in the workplace at international geographers conferences demonstrate that policy appears to be lacking behind.

We would not advocate that our work, mapping mas culinities, has been of no consequence. Instead, we would encourage researchers from all areas of geography to consider more explicitly gendered identities and masculine ideologies and the policy relevance such work might have. Perhaps more work needs to include collaborations with organizations and institutions in order to raise awareness and impact strategy and policy. Work on representational spaces and everyday lived experiences remains vital in critically accessing structural conditions for producing specific identity positions as well as the role of agency.

See also: Capitalism; Cultural Politics; Cultural Turn; Feminism and Work; Feminism/Feminist Geography; Gender and Rurality; Gender, Historical Geographies of; Gender in the City; Heritage and Identity; Identity Politics; Jackson, P.; Masculinism; McDowell, L.; Natures, Gendered; Parenting/Motherhood/Fatherhood; Patriarchy; Social Geography.

Further Reading

Berg, L. (1994). Masculinity, place and a binary discourse of 'theory' and 'empirical investigation' in the human geography of Aotearoa/New Zealand. *Gender, Place and Culture: A Journal of Feminist Geography* 1(2), 245 261.

Berg, L. and Longhurst, R. (2003). Placing masculinities in geography. *Gender, Place and Culture* 10, 351 360.

Bondi, L. (1990). Feminism, postmodernism and geography: Space for women. *Antipode* 22(2), 156 167.

Connell, R. (2005). *Masculinities* (2nd edn). Cambridge: Polity Press.

Connell, R. and Messerschmidt, J. W. (2005). Hegemonic masculinity: Rethinking the concept. *Gender and Society* 19(6), 829 859.

Cosgrove, D. and Jackson, P. (1987). New directions in cultural geography. *Area* 19, 95 101.

Courtnenay, W. H. (2000). Constructions of masculinity and their influence on men's well being: A theory of gender and health. *Social Science and Medicine* 50, 1385 1401.

Dawson, G. (1994). *Soldier Heroes: British Adventure, Empire and the Imagining of Masculinities*. London: Routledge.

Flood, M. (2008). XY: Men, Masculinities and Gender Policy. http://www.xyonline.net/links.shtml (accessed in Sep. 2008).

Flood, M. (2008). The Men's Bibliography, 18th edn. http://mensbiblio.xyonline.net (accessed in Sep. 2008).

Hacker, H. M. (1957). The new burden of masculinity. *Marriage and Family Living* 19(3), 227 233.

Jackson, P. (1991). The cultural politics of masculinity: Towards a social geography. *Transactions of the Institute of British Geographers* 16, 199 221.

Jackson, P. (1994). Black male: Advertising and the cultural politics of masculinity. *Gender, Place and Culture* 1, 49 59.

Kenway, J., Kraack, A. and Hickey Moody, A. (2006). *Masculinity Beyond the Metropolis*. New York: Palgrave MacMillan.

Little, J. and Leyshon, M. (2003). Embodied rural geographies: Developing research agendas. *Progress in Human Geography* 27(3), 257 272.

Levant, R. F., Majors, R. G. and Kelley, M. L. (1998). Masculinity ideology among young African American and European American women and men in different regions of the United States. *Cultural Diversity and Mental Health* 4(3), 227 236.

Longhurst, R. (2000). Geography and gender, masculinities, male identity and men. *Progress in Human Geography* 24(3), 439 444.

McDowell, L. (2002). Transitions to work: Masculine identities, youth inequality and labour market change. *Gender, Place and Culture: A Journal of Feminist Geography* 9(1), 39 59.

McDowell, L. (2003). *Redundant Masculinities: Employment Change and White Working Class Youth*. Oxford: Blackwell.

Messner, M. A. (1998). The limits of the male sex role: An analysis of the men's liberation and the men's rights movements' discourse. *Gender and Society* 12(3), 155 276.

MRC (2008). Men's Resource Centre for Change. http://www.mrcforchange.org/voicemale.html (accessed in Sep. 2008).

Oxfam (2008). Resources. Gender Equality and Men. http://www.oxfam.org.uk/resources/learning/gender/gem.html (accessed in Sep. 2008).

Pain, R. (2006). Paranoid parenting? Rematerializing risk and fear for children. *Social and Cultural Geography* 7(2), 221 243.

Phillips, R. (1997). *Mapping Men and Empire: A Geography of Adventure*. London: Routledge.

UNDP (2000). Gender and Development. Men and Gender Equality. http://sdnhq.undp.org/gender/programmes/men/men statement.html

van Hoven, B. and Horschelmann, K. (2005). *Spaces of Masculinities*. London: Routledge.

Relevant Websites

http://www.southasianmasculinities.org
Exploring Masculinities. A South Asian Travelling Seminar.

http://www.genderjustice.org.za
Sonke Gender Justice Network. HIV/AIDS, Gender Equality, Human Rights.

Massey, D.

T. Skelton, National University of Singapore, Singapore

Massey, Doreen (1944–)

Doreen Massey became professor of geography in the Faculty of Social Sciences at the Open University (UK) in 1982 from where she consolidated and developed her exceptionally productive and international career as a leading theoretician in geography and the social sciences (**Figure 1**). Alongside this post she holds 22 international positions as a member of editorial, advisory, and management boards. Professor Massey has been awarded several honors in recognition of her outstanding contribution to geography and the social sciences. She was awarded the Prix Vautrin Lud (Nobel de Géographie) in 1998 and the Anders Retzius Medal in Gold by the Swedish Society of Anthropologists and Geographers in 2003. These combine with many accolades awarded by British universities, academies, and societies.

Massey was born to working class parents in Wythenshawe in the northwest of Britain. She has written about this homeplace of growing up, as well as her other significant 'home', Kilburn in London. Being 'northern' has been and remains an important part of Massey's personal and spatial identity. In her early career it would be difficult to describe Massey as a 'geographer', but there were always tangential relations with the subject. From her undergraduate degree at Oxford she worked in London as a principal scientific officer at the Centre for Environmental Studies (CES). A sabbatical allowed her to complete a Masters in Regional Science at the University of Pennsylvania. She returned to her job at the CES and held a variety of responsibilities alongside this post. She was a member of the editorial advisory board for *Environment and Planning A* (1968–86) and a part time tutor for the Architecture Association, the Polytechnic of Central London, and the London School of Economics (LSE).

Figure 1 Doreen Massey.

Her reconnection with geography as a disciplinary space appeared to take place in 1980 when she was awarded a senior research fellowship in the department of geography at the LSE (1980–82). She then held a 3 month visiting professorship in the department of geography at the University of California, Berkeley, before taking up her Open University professorship.

Doreen Massey has played a profoundly influential role within geography and, as we might expect from the nonboundedness of her theorizations, throughout the social sciences. Her intellectual influence spans beyond the actual geography of Britain. She has held, or holds, a range of membership positions in Brazil, Finland, Ireland, Mexico, Nicaragua, Spain, and the US. She has held memberships at editorial and advisory board status for seven geography journals and three in the wider social sciences. She co founded (and edited) *Soundings: A Journal of Politics and Culture* in 1998. She has strong connections with the Royal Geographical Society, the British Academy, and is a founding academician of the Academy of Learned Societies in the Social Sciences. Massey has evaluated and challenged the meaning and production of geography through her membership of two Higher Education Funding Council for England (HEFCE) Research Assessment Panels (1995–97, 1999–2001) and through numerous 'conversations' across the human–physical geography divide.

Massey's phenomenal CV testifies to her ability to communicate and profess her geographical and social imagination to diverse and differentiated audiences. Her career is still in a process of 'becoming' and charts a tremendous 'activity space' with international interactions throughout Africa, Australia, Europe, New

Zealand, Scandinavia, and the Americas (North and South). To date Massey has published 26 authored, co authored, and edited books (several have been translated into Greek, Italian, Japanese, Polish, and Portuguese). She has 78 chapters in books and collections and 130 journal articles (ranging from the top academic journals in the social sciences to more populist and pedagogic journals such as *Primary Geography*). In her career she has made innumerable presentations to an incredibly varied array of audiences ranging from academic conferences, local geographical societies, action groups, art colleges, trade unions, politicians, and bishops. She has reached even more diverse audiences through her public intervention work on television programs, British Broadcasting Corporation (BBC) radio (local, national, and the world service), and the Australian Broadcasting Corporation.

Continually interwoven within her academic publications and presentations are political interventions which take the form of political lobbying, activism, and a range of nonacademic writing. Indeed, Massey's work places great importance on the need to create new spaces and places for political change; articulates political discourses that provide a genuine openness for the future; and demands a politics which challenges 'inevitable trajectories'. For Massey, writing is very firmly connected with doing.

In view of the expansive collection of Massey's publications and her concomitant impact on geography, it is unsurprising that Massey's scholarship is iterative and weaves and threads its way through and into itself. She returns to theorizations, themes, and concepts throughout the chronological timeline of her work. In later work she expands and develops ideas and arguments put forward in earlier pieces (e.g., the 'regional problem' discussion which she has written on from the 1970s to the present). Given the recursive nature of her work we need to loop forward and backward in order to capture the ebb and flow of Massey's thinking as it relates to urban/industrial/economic geography; her conceptualizations of space and of place; and her work on social relations and spatiality.

Massey's urban, industrial/economic geography work began in the 1970s. Her approach within these strands of geography has always offered critical conceptual engagements and interrogations of the taken for granted and she has consequently done a great deal to develop and stretch these subdisciplines. Her critical urban and economic geography continues to be present in much of her subsequent work, culminating in her 2007 book, *World City* which questions the power and reach (or lack of reach) of London's wealth. Initial inspection of her early (1970s–1980s) economic geography work makes it appear very UK centric, with a focus on debates and representations about regional inequalities and a critical engagement with regional policy. However, Massey's international perspective, which is evident throughout all her work, is established even within this early writing. In 1979 Massey asked the question 'in what sense a regional problem?' and introduced the term 'spatial division of labor'. This conceptualization was later to be famously extended in her book, *Spatial Divisions of Labour* (1984), described as one of the most cited books in human geography. She coined the term 'spatial division of labor' to critique the normative assumption that economic activity responds to geographical inequality in the conditions of production to maximize profits. Massey argued that what this interpretation ignores is the ways in which different forms of economic activity use existing geographical inequalities so that profits can be maximized. Hence geographical unevenness will be used in different ways by different sectors and may even play a role in structuring capital. As such, there will always be different so called regional problems. Spatial divisions of labor are not static but evolving as capital and production shift and change. Additionally Massey stressed that spatial divisions of labor are formed through national and international/global processes. This conceptualization of labor provided a means by which the dramatic economic and industrial transformations taking place in industrialist capitalist countries, from the 1960s onward, could be analyzed through a focus on spatial organization, not just on patterns of employment. It also provided a framework for investigating the relation between the spatial and the social; for example, gender relations and inequalities, employment and the exploitative tactics of capital, and how these are spatialized. Massey established the argument that changes in industrial and economic structures (including employment) have to be examined through class and economic relations (the social relations of capitalist production) at both national and international scales.

Massey's theorizations of space and place are probably what she is best known for. This part of her work has been translated into a range of languages and has a global reach. While space and place as concepts are present in her early work, the 1990s marked a very productive period in which Massey critically engaged with the meaning of both. Her interrogations have consistently argued against the taken for grantedness of what space and place mean, the neglect of space compared to time, the false separations between the local and the global, and notions of spaces/places being fixed, bounded, and closed. In two extensively cited essays, Massey expounded the possibility of a 'global' and 'progressive sense of place' at the same time as demanding recognition of the 'power geometries', produced by different actors and structures, that are played out in space and place. She teased apart what is meant by place in an era of time–space compression and asked how place can remain important, particularly to people, and provide a positive sense of rootedness without being reactionary. Places can be conceptualized as moments of articulation in networks of social relations and hence have internal tensions, contradictions, and harmonies. Places are theorized as processes and as meeting points. However, place is not just produced by what goes on within it but it is about its relational position with what goes on beyond; the local and the global constantly connect and collide.

In *For Space* (and elsewhere) Massey pulled together many of the threads of her quest for meaningful theorizations of space (and also place) and called for careful analysis of the way in which space/place are produced by the materialities of power. She articulated three key propositions for space. First is to understand space as a product of interrelations where relations are embedded practices that are constructed through politics, identity, and spatialities. Second, to imagine space as a sphere of possibility for the existence of multiplicity and heterogeneity which means recognizing the non inevitability of certain trajectories and universalizations of how the world is and will be. Multiplicity requires the spatial and without multiplicity there is no space, they are co constitutive; space requires the existence of plurality. Likewise place is permeable and a constellation of trajectories. The third key proposition is that space is always in process, always becoming, it is an open system. Such openness is a requirement for possible futures in which politics can make a difference. Space could be defined/imagined as a 'simultaneity of stories so far'; if space is recognized as a sphere of identities and entities and of the relations between them, then it is always being made and consequently creates the possibility of a politics around the negotiation of these relations. For Massey, one possible politics to strive toward is the 'responsibilities of place'. Globalization has local origins, it is produced in local places, and consequently it is possible to ask what kind of local political strategies can lead to the constitution of 'global responsibilities'.

Massey's work on social identities effectively laces together her analysis of the spatialization of capital (her urban and economic geography work) with her theorizations of space and place because it looks at inequalities, connections, relations, and lived materialities of people's lives and also of the discipline of geography itself. In *Flexible Sexism* Massey takes geographers to task for their writing on postmodernism and geography and their elision of feminism. Dualisms constructed within the high technology sectors are examined for the ways in which they construct a particular form of masculinity, one associated with temporal and spatial flexibility. In this type of employment masculinity is associated with transcendence and femininity with immanence, but the public and private dualism is not quite as evident as expected. This demonstrates porosity and potentially a more progressive type of time–space. An examination of the spatial construction of youth cultures in the Yucatán, Mexico, provides an insight into debates about hybrid cultures; the interaction of the local and the global and the openness of cultural formations. Massey argues that this is evidence of the ways in which the social relations that constitute space are organized into constellations of temporary coherence that are always and everywhere imbued with power.

See also: Economic Geography; Industrial Location; Place; Spatial Division of Labor.

Further Reading

Massey, D. (1979). In what sense a regional problem? *Regional Studies* and re printed in Agnew, J., Livingstone, D. N. & Rogers, A. (eds.) (1996) *Human Geography: An Essential Anthology* 13, 233 243. Oxford: Blackwell.

Massey, D. (1984). *Spatial Divisions of Labour: Social Structures and the Geography of Production*. Basingstoke: Macmillan.

Massey, D. (1991). A global sense of place. *Marxism Today,* June, pp 24 29.

Massey, D. (1993). Power Geometry and a progressive sense of place. In Bird, J., Curtis, B., Putnam, T., Robertson, G. & Tickner, L. (eds.) *Mapping the Futures: Local Cultures, Global Change*, pp 57 69. London: Routledge.

Massey, D. (1994). *Space, Place and Gender.* Cambridge: Polity Press.

Massey, D. (1995a). The conceptualisation of place. In Massey, D. & Jess, P. (eds.) *A Place in the World?* pp 46 85. Oxford: Oxford University Press and The Open University.

Massey, D. (1995b). Masculinity, dualisms and high technology. *Transactions of the Institute of British Geographers* 20, 487 499.

Massey, D. (1998). Spatial construction of youth cultures. In Skelton, T. & Valentine, G. (eds.) *Cool Places: Geographies of Youth Cultures*, pp 121 129. London: Routledge.

Massey, D. (1999a). Space time, 'science' and the relationship between physical geography and human geography. *Transactions of the Institute of British Geographers* 24, 261 276.

Massey, D. (1999b). Negotiating disciplinary boundaries. *Current Sociology* 47, 5 12.

Massey, D. (2001). Geography on the agenda. *Progress in Human Geography* 25, 5 17.

Massey, D. (2004). The responsibilities of place. *Local Economy* 19, 97 101.

Massey, D. (2005). *For Space*. London: Sage.

Massey, D. (2006). Landscape as a provocation: Reflections on moving mountains. *Journal of Material Culture* 11, 33 48.

Massey, D. (2007). *World City.* London: Polity Press.

Relevant Websites

http://www.bbc.co.uk
A Festival of Ideas for the Future Open University Radio Lecture (Thursday 9 November 2006 21:30 22:30 (Radio 3)).

http://www.ica.org.uk
Doreen Massey: London Inside Out. Pioneering radical geographer Massey takes on the view of London as a happy global microcosm (9 September 2005).

Material Culture

D. P. Tolia-Kelly, Durham University, Durham, UK

Glossary

Marxism A political ideology based on the writings of Karl Marx, arguing that material relations in society can be defined through understanding the ownership of means of production in any society. Capitalism itself was critiqued for providing economic conditions for the overthrow of social rule toward a democratic state where workers owned the means of production and thus were freed from chains of exploitation, alienation, and oppression.

Post-Structuralism A philosophical account of society that challenges Marxism as being dependent on defined, narrow structures that do not reflect social reality. A philosophy espoused by those considered 'radical' but with a conservative approach to social change and reform.

Material Geographies

The study of material culture within the discipline of geography has been rooted in ideological concerns about the writing of 'real' histories of society and engaging with the 'real' in its spaces. A materialist approach has considered 'matter' at the heart of its research. 'Matter' is positioned in these accounts as solid evidence, artifact, 'mirror', and ultimately 'super object', where the place and nature of 'materials' in society are used to figure out what is happening in society and its social relations, or what has occurred in the past. 'Material culture' has operated as a 'litmus test' to what is going on right now and what has gone before, spatially and temporally. A materialist orientation to the study of society and space has had several phases in the history of geography. At the beginning of the twenty first century, we saw a 'return to matter' as a reaction to the 'cultural turn' of the 1990s. This call was essentially a call for a return to 'things' as opposed to discourse, narrative, semiotics, and the visual, and the notion of language as being the engine of social science. Instead, 'materiality' has become the vanguard of a political consciousness that seeks to be politically inclusive, a counter to the dominant, masculinist way of doing geography, and a mode which can operate as a tool to investigate geographies of the everyday, the past, and of the geographies in the process 'of becoming'. The materialists in the social sciences have a varying array of doing geography where the 'object' of research can be varied: 'nature'; 'the urban'; the museum; the body; visual culture; all that which is consumed, produced, and decayed; and that which is conceptualized as 'landscape', 'the domestic scene', and 'waste'. In terms of scale, the orientation toward any 'matter', such as the figure of 'nature', can incorporate a study in various scales – from the scale of global ecology to seed, to plant cell, and to atom. 'Matter' is at once tangible and intangible, visible and absent, decayed and in the process of becoming, evoking sentimentality and mundaneness.

This article traces the ways in which 'material culture' has been important in the study of geography as a discipline and how it is engaged within current literatures in geography today. To understand this involves knowing the value of 'material culture' in research across the social sciences including archaeology, anthropology, and cultural studies. In contemporary research, material culture is positioned as a 'gateway of knowing', through its texture, its sensory presence, and as a textual record of what has gone before. For example, a church, a wall, or gravestone in the landscape are artifacts of their times and their presence is a tangible record of what has gone before; this is a very different way of thinking history to the scholarship of a historian's narrative. Using a Marxist lens in any society, wealth is linked to the ownership of the 'means of producing' wealth, such as factories, land, and mining rights. This is true in the twenty first century; those that own land, industrial production, and/or access to oil are those that lead our global economy. Considering material production gave rise to a dual notion of society – 'producers of wealth' (laborers), and those that controlled the 'means of wealth production'. Those that produced the goods were the 'poor' and those that did not produce were 'the rich'.

The collective presence of the materials produced were considered as evidence of the wealth of that society, produced through labor and the distribution of that wealth based on who owned what materials, what of quality, and where. For example, in the twenty first century, if one were to map, on a global scale the number of people who owned a car, house, credit card, or bank account, one would end up with a map of global 'development'. Materiality in its simplistic mode becomes an artifact of measuring wealth, health, and distribution of economic 'goods' and 'bads'.

Material geographies is a reaction within the discipline against the 'dematerializing' trends of the 'cultural turn'. However, as Matthew Kearnes reminds us, it is important to note that the deployment of 'matter' is not always more attentive than the thing itself. Material cultures represent a focus on the 'thingyness' of the 'bump into able' world. But it does reflect a belief in the presence of material cultures as signifiers of the 'real', that is, power relations, social relations, and political geographies where the study of objects secures a research process situated in the world of humanity. The accounts below trace where material geography is situated cur rently. It begins with an account of historical materialism and an engagement with various framings of 'material culture' research within the discipline.

Thinking about material culture as the evidence of social relations is rooted in Marxist ideology – namely, historical materialism. Matter in this political ideology becomes an external social referent; to point out that material cultures are the evidence of social relations in cluding wealth, poverty, and oppression is located in Marxist thinking. The production and consumption of materials in any society become the evidence of how well that society is doing. Collective social wealth is reflected in the development in production of social goods that are beyond basic subsistence needs of a society; art, architecture, technology, and academic development thus represent an advancement of society. Marx argued that "relics of bygone instruments of labor possess the same importance for the investigation of extinct economic forms of society, as do fossil bones for the determination of extinct species of animals." Thus, material cultures could be artifacts through which social organization and social progress could be determined. Humanity in this account has a unitary history, where struggles between producers of wealth and goods are in struggle with owners of wealth and the means of production. A unitary approach to history thus means that there is a pattern to history, a mode of thinking about history where the past can tell us about the present, where the material cultures of the present can tell us about the present and the past – this is known as dialectical materialism.

History and Material Cultures/Historical Materialism

Marxism has influenced the ways in which archaeology, anthropology, and geography have been studied over the last few 100 years. These disciplines in the twentieth century reorientated themselves to engage with 'material culture' as a form of text through which 'other worlds, peoples, and places' could be narrated. This form of archaeology, anthropology, and history shifted the lens of these disciplines to engage them with 'whole' societies; from being groups of privileged antiquarians, they be came disciplines for scholars of material culture. The politics of this shift reflected a move away from elite collections and Whiggish accounts of the past, beyond the historical narration of history through 'king's lists'.

Elite relations and narrations with objects are reflected in the nineteenth century practice of collecting and their 'cabinets of curiosities'. These collections of artifacts were ways of recording the world and its cultures, each cabinet became a microcosm of the world and its curious peoples and natures. These collections on the domestic scale and the national scale of museum spaces were objectifications of authoritative knowledge, and reflected an imperial taxonomy of the world and the cultural 'development' of its people. This conceptualization of the world ordered civilizations on a linear spectrum – at one end were nativized 'others' through to the English, Vic torian society, which held the most advanced position; this is termed as 'unilineasm'. These artifacts or 'super objects' were reified and figured through a dominant and particular way of defining societies and their 'progress', culminating in collections such as that held at the British museum. The social, contextual, and economic placing and values of these curiosities were secondary to their positioning as 'artifacts of the world' – all encompassing, intellectualized, reified accounts, removed from the social meanings and situatedness of these material cultures. This is what was embraced in the twentieth century in the form of cultural materialism.

Cultural Materialism

The study of material cultures has not, however, been singular in its nature. Although some contemporary scholars are influenced by Marx, many scholars who have different political orientations occupy the realms of scholarship that have been rooted in Marx and shaped by 'materialism'. To understand these shifts, the following sections summarize varied approaches to material cul tures that shape the modern discipline.

Raymond Williams coined the term 'cultural materi alist'. Cultural materialism is a theoretical movement which emerged in the early 1980s along with new his toricism. Williams viewed culture as a productive pro cess, for him the study of the literatures of any age would reflect the dominant values of that society. Material cultures, such as the novel, the painting, the architecture of a city, were thus situated as cultural materials that could narrate the political struggles of a society, and its economic, social, and cultural systems of governance, in Foucauldian terms. The importance of cultural materi alism is that it seeks to unearth alternative narrations of the dominant histories told, genealogically. Materials of culture were defined as a tool of hegemonic political structures in a society that utilized and appropriated canonical and historically important texts in an attempt to validate or inscribe certain values on the cultural imaginary. Postcolonial writers such as Leela Gandhi and Edward Said have argued that the English novel is an example of the ways in which the values of 'orientalism' were promoted and secured as part of the economic and cultural consolidation of the Empire. The novel thus is not a mere reflection of those values, but becomes a diagnostic and political tool.

Anthropology and the Super-Object

Williams has influenced geographers, particularly in the analysis of landscape in relation to specific social and economic regimes. In work on materialist accounts of 'landscape', scholars have considered representations of national landscapes and other material cultures that circulate to consolidate 'landscape values' in the public sphere. Recognition that landscape has been cultivated as an esthetic object within which material relations of economy and power are obscured is key to a cultural materialist approach in geography. Studies of the icon ography of landscape painting, for example, have laid bare the ideological basis of dominant visual cultures within society. In particular, 'landscape' as a 'way of seeing' the world has the power to consolidate national values and social understandings of national citizenship within cultural discourse. As the work on art history in geography has shown, landscape allows national class identities and interests to be muffled – hidden within symbolic depictions of nation. From a cultural materialist perspective, it is evident that landscape representations in art, literature, and other forms of cultural production can support a process of 'ideological mystification'. The heterogeneity of national cultures and their territories of enfranchisement is veiled. Landscape is thus an instru ment of cultural power. It has a mask of innocent ideal ism, occluding moral, ideological, and political questions. The dismantling of the ideology behind landscape rep resentation is central to understanding its power in the colonial and imperialist projects which have their legacy in the nationalisms of postcolonial nations and their contemporary iconographies. A cultural materialist per spective has been crucial in consolidating an interest in cultures of resistance, 'difference', and the possibilities of alternative cultures of enfranchisement.

Anthropological writings in the field of material cul tures have been influential in geographical research. The work of Danny Miller from the 1980s onward, (although rooted in Marxism) has highlighted the shift from a Marxist ontology that insisted that being human was constituted in the 'act of production' toward a concern with the role of consumption in society in the con struction of themselves and others through relations with cultural forms. Consumption of goods and cultural materials is a phenomenon of developed capitalist economies – those evolving societies with regard to fully operational consumer centered infrastructures in China and India are regarded as 'in process' where consumer objects will become fully sovereign over and beyond philosophies that have seemingly shaped both of these societies in the past. Anthropological concerns over the social are reorientated through the conceptualization of material cultures as mattering to kinship relations; ob jectification is thus enlivened to give 'things' a social life and biography. 'Why things matter?' is a key question of this critique of Marxist reductionism.

Post-Materialist Materialism

The terms 'matter' and 'materiality' are often conflated in the political imperatives of social science's approach – 'because something is material, it must be material' is a view challenged by post structuralist thinkers, who have engaged with materiality from this perspective. For post structuralist researchers, 'things' cannot be read as text or be relied upon to be understood as artifact, or as repre sentational (of social trends, of a memory, of an event, etc.). Instead 'matter' and 'things' are always in process; we cannot 'know' a thing as 'knowable' and treat this as evidential or artifactual because these are simply ac cepted values denoted through language and narrative. It is argued that these mythologies about things are those that become solidified, and are as material as things themselves. The 'social life of things' from this per spective is thus reduced to newly 'textualized' accounts; before we approach an object, we bring a story to it that has been formed through other textual encounters; thus the textual encounters shape our engagement with this object. Material cultures thus are only meaningful at the point of each encounter – at the point 'of becoming' not fixed, prismatic, or defined through their role in eco nomic relations. This evaluation is inevitably inspired by Arjun Appadurai, but rejects the material object as bounded and singular.

Appadurai has called for (in the footsteps of Mausse and Marx) an enlivened approach to material cultures across the social sciences through his view that material cultures have a 'social life', and are not benign, deadened tools that are functional in a taxonomic record, non agentic, and in our control, but regarded as things which shape our identities and thus, in turn, are shaped and given identity beyond their function. However, the re lations between 'things' and the body remain problem atic. Reducing things to text delimits their presencing effect in social space, yet delimiting them to post struc tural or phenomenological accounts reduces their agency or role in the fabric of societal systems of governance, social control, and alienation. Politics and material cul tures meet again at this juncture; thus the value of ma terial cultures is signified by the political values that one holds. The nineteenth century cabinets of curiosities reflect a need for a universalist perspective on defining humanity, modernity, and approach to philosophies of existence. A postmodern and/or post structuralist view of history and culture does not always require the figure of materiality; materiality does not automatically matter or have meaning. In these accounts, meanings are in process and thus are not so readily tangible. This represents another intellectual and, namely, ontological turn, whereby thinking, feeling, and conceptualizing 'material cultures' supercede their eventfulness in everyday life. This could be thought of as a philosophy where the immaterial is privileged as a condition of material understandings, where the 'matter' or presence of the thing is uncertain, denied, or continually contingent; the contingency of material cultures' eventfulness/meaning/liveliness as being relational to the absent. In this account, the objects that are 'nonagentic' or 'immemorialized' are both productive and reductive. This is an antithesis to the politics of an integrative approach to material cultures embodied in the approaches used in the work of anthropology, archaeology, and geography, that continue today when dealing with, for example, excavations, monuments, health, landslides, transport networks, kinship networks, and race geographies.

Material Cultures, Memory, Loss, and Nostalgia

The value in approaching material cultures, such as ruins, decaying texts, paintings, and waste matter itself, focuses attention on to the discarded and divested. These are sites and objects usually occluded. This offers new life to the obsolete through acknowledging their liveliness and power in the cultural sphere; new esthetics of these material cultures are discovered and old ones evoked. For some researchers on ruins and waste, the disordering effects of ruination expose the ways in which the material world is defined, valued, and ordered. For writers such as Tim Edensor, "waste is part of an excess, impossible to totally erase, which contains rich potential for reinterpretation and reuse because it is under determined. Through processes of decay and non human intervention, objects in ruins gradually transform their character and lose their discreteness." Again, there are temporal slippages between past and present, and spatial slippages between object and bodies.

With these forms of material cultures, memory, loss, and nostalgia feed into the ways in which 'waste' matters. Things, people, and events consigned to the past continue to haunt the present. Also, 'loss' has come to be seen as a core reason for this return to material cultures as an intellectual force. In the nineteenth century, archaeologists and anthropologists were keen to collect the cultural materials of societies at risk from colonial expansion and modernity. The collection of material cultures was seen to be an act of preservation of a 'lost world' of curiosities. In contemporary geographical research on memory, material cultures are keenly positioned as the harbingers of postcolonial cultures that may be lost or histories that may not be articulated. The shifts toward the absent within the present has allowed the articulation of languages, voices, cultural practices, worlds individuals, and subjective experiences of those already outside of the academic realm. In the case of material artifacts held in museums, calls for their repatriation and renarration 'against' those narrations of imperial science are ironically the means through which these 'other' material tales are articulated.

For some writers, memory, nostalgia, and loss are at the heart of much material cultural studies. There is a view (arguable) that embedded herein is a yearning for a knowable modern society, where things were not contingent, but concrete, ordered, and robust. The compounding effect of speed and time–space compression has caused a cultural loss of individual and social infrastructure that was eulogized in the Victorian era. Material cultures today represent multifarious events, spaces, times, and subjectivities – unknowable and intangible.

Visual Culture

Visual culture has been the site of research within the realms of 'material cultures'. The aim here is to consider visual objects beyond their meanings as 'texts'. For many researchers, the visual material cultures analyzed have been dealt with in the tradition of cultural materialism. More recently, theories of affect and emotional geographies have influenced the ways in which interactions with art and objects are meaningful beyond the text. Biographical research on these material cultures trace the paint, paper, printing presses, etc., that were employed in the processes of production and display. In these research processes, there is a collapsing of the encounter between the objects, their identity, and that of the subject encountering it. In Appadurai's view, India as a developing consumer society is a good site for the excavation of meanings between objects and bodies. He argues that things in India never lose some of the magic of their human makers, owners, or handlers; and following Marx, both things and humans share the mystery of the commodity and the underlying evidence of labor. For him the sheer profusion of things in India makes it impossible to set art apart from its wider context and makes it difficult to distinguish art clearly from the objects of everyday life. If this is true, then the global art market is purely and arbitrarily valuing commodities that have no relation to 'art'. Visual culture thus collapses into a material world of cultural meanings, not globally recognized and universal, but locally developed, subtle, and nuanced. Situated values of material cultures are thus the only way to understand their effectiveness, their affective value, and ethnographic role in the creation of dialog about the world we live in. These investigations between whether 'art' matters and its 'effects' result in a very simple taxonomy of visual culture based on 'Does it make you feel?'

Materializing the Body in Human Geography

Peter Jackson in the 1990s argued that the emphasis within our discipline should be on when and where the materiality of material culture makes a difference; this is in opposition to Marxist accounts where material cultures were assumed important in an *a priori* manner. As elucidated in this article, the very nature of 'matter' is questioned within some corners of the discipline; in others with the aid of 'material chemists', the nature of things considered 'toxic waste' are being explored at an atomic scale. The very matter of material cultures are being attended to in a very different way from that of the Sauerian attendance to landscape or of Cosgrove's original attention to landscape as a 'way of seeing'. However, without political reflexivity in the research process, the very figure of matter dissolves into rhetoric and mythology. The very fact is that we are material flesh, with material processes of life and thought, from memory synapses to reproductive potential to the potential to act against the very existence of the world we live in. Degeneration in the realms of human rights, ecological integrity, democratic process, and the material needs of survival are not being met in some corners of the world. These cannot be mythologized, or be deemed 'contingent'. Therefore, the argument here is that the realm of material cultures should retain an attunement to why material cultures make a difference. Under the regime of Nazism, Agamben has demonstrated that social governance allowed the material conditions for the denial of life; the Holocaust was material – in the material loss of life, and the political means through which societal governance created states of exception for particular bodies and citizens.

Thus the project of writing geography requires a commitment to things, presence and text; otherwise there is always the risk of the past collapsing into a set of narratives and mythologies that are claimed by Holocaust deniers equally to those that recover the truths of genocide. Matter can act as evidence, albeit enriched by contingencies, absences, imaginings, and reawakenings of geographies of the past, present, and future.

See also: Cosmopolitanism; Cultural Capital; Material, The.

Further Reading

Appadurai, A. (2006). The thing itself. *Public Culture* 18(1), 15 21.
Bennett, J. (2004). The force of things: Steps towards an ecology of matter. *Political Theory* 32(3), 347 372.
Bingham, N. (2006). Bees, butterflies, and bacteria: Biotechnology and the politics of nonhuman friendship. *Environment and Planning A* 38(3), 483 498.
Cook, I. and Harrison, M. (2007). Follow the thing: West Indian hot pepper sauce. *Space and Culture* 10(1), 40 63.
DeSilvey, C. (2006). Observed decay: Telling stories with mutable things. *Journal of Material Culture* 11(3), 318 338.
Edensor, T. (2005). The ghosts of industrial ruins: Ordering and disordering memory in excessive space. *Environment and Planning D: Society and Space* 26(6), 829 849.
Edwards, E., Gosden, C. and Philips, R. (2006). *Sensible Objects: Colonialism, Museums and Material Culture*. Oxford: Berg.
Hirsch, E. and O'Hanlon, M. (1995). *The Anthropology of Landscape: Perspectives on Place and Space*. Oxford: Oxford University Press.
Holloway, J. and Hones, S. (2007). Muji, materiality and mundane geographies. *Environment and Planning A* 39(3), 555 569.
Jackson, P. (2000). Rematerializing social and cultural geography. *Social and Cultural Geography* 1(1), 9 14.
Kearnes, M. B. (2003). Geographies that matter The rhetorical deployment of physicality? *Social and Cultural Geography* 4, 139 152.
Malafouris, L. (2004). The cognitive basis of material engagement: Where brain, body and culture conflate. In DeMarrais, E., Gosden, C. & Renfrew, C. (eds.) *Rethinking Materiality: The Engagement of Mind with the Material World*, ch. 5, pp 53 61. Cambridge: McDonald Institute Monographs.
Miller, D. (1998). *Material Cultures*. London: UCL Press.
Rose, G. (2004). Everyone's cuddled up and it just looks really nice: The emotional geography of some mums and their family photos. *Social and Cultural Geography* 5, 549 564.
Victor, B. (2002). *The Material Culture Reader*. London: Berg.
Whatmore, S. (2006). Materialist returns: Practising cultural geography in and for a more than human world. *Cultural Geographies* 13(4), 600 609.
Wright, C. (2004). Material and memory: Photography in the Western Solomon Islands. *Journal of Material Culture* 9(1), 73 85.

Relevant Websites

http://www.ashgate.com
Ashgate Publishing.
http://www.siaamc.org.uk
Islamic Material Cultures.
http://mcu.sagepub.com
Journal of Material Culture, Sage.
http://larsdatter.com
Medieval Material Culture Blog, Mesterinde Karen Larsdatter.
http://www.materialculture.wisc.edu
The material culture program, Material Culture Institute, Wiconsin, US.
http://www.ucl.ac.uk
UCL Centre for Museums, Hertiage and Material Culture Studies Lanched, University College London, Institute of Material Cultures.

Material, The

A. Tay, University of Bristol, Bristol, UK

Glossary

Assemblage The coming together of artifacts, bodies, and things in a complex, nonlinear, and heterogeneous fashion. It signals an ontological refocus on the structure of matter as an emerging collectivity in which the processes of its formation and expression do not threaten the identity of the individual components.

Material Tangible, sensate matter with distinct physical form, shape, and size at any one temporal moment.

Material Landscape Objects depicted in space and/ or place based on the understanding that their presence was made possible through enactment by related things and networks.

Quasi-Object The recognition of things having distributed agencies and potentialities for enrolment into networks such that they may be considered as social ties. A quasi-object, in other words, does not have an existential ideal. Quasi-object is often used in conjunction with 'quasi-subject', and more recently, the singular reference term hybrid.

Second-Hand Culture The realm of reciprocal exchange or purchasing of goods and possessions that have passed through the production/consumption cycle as least once before. Circulating in and around antique store as it does on ebay.com, it typifies an alternate and ingenuous space to mainstream capitalism.

Stuff Geography A term coined by geographer Ian Cook *et al.* to refer to an eclectic collection of mundane, taken-for-granted things that, put under the spotlight, reveal their spatially contingent biographies and interactions with human liveliness.

Objects and Landscapes

Many objects in human landscapes literally form the basis of our historical scholarship. The accepted genesis of the subdiscipline of cultural geography – The Berkeley School under the direction of Carl Sauer (1889–1975) – cuts an investigative figure to survey distant, and not so distant lands, recording in careful detail a cultural topography of landscapes filled with things familiar and peculiar. In the 1920s, maps were used extensively by the Chicago School of Urban Sociology to define, for example, housing settlement patterns. Such annotated and demarcated maps aided the field of urban ecology in pioneering the zoning and planning practices of many American cities and in decisions regarding the physical placements of various commercial, industrial, and residential infrastructures. A bounded city was thus viewed as an organism: it needed to be ordered and cleaned up just as it needed systemic respite in the form of greenbelts, parks, and trees. Allusions to such objects of nature were to figure substantially in humanistic studies such as that of Yi Fu Tuan in examining humans' interactions with their environment.

Claims to such objects, however, did not go uncontested. Urban geographers speak of the 'power of maps', pointing to those which purport to show clustering according to social indicators, such as age, ethnicity, and income levels, as in fact material constructions of race and moral ideologies. Just as cities can be nursed to health, they can fall or remain 'sick'. Sociologist Mike Davis, for example, shows how in a quest to maintain safety, sanitation, and esthetics in Los Angeles, bum proof benches, stoop seats, and untimely water sprinklers in the park serve as objects that make the city inhabitable for those on society's margins. Similarly, Jane M. Jacobs looks at high rise residential buildings as what she terms 'artifactual modernities'– emblems of progress and development of urbanized areas. When they become the reason for increased incidence of suicides known as 'high rise leaps', however, they pathologically show how 'things fall apart'. This is also linked more broadly to the work of postcolonial geographers like Brenda Yeoh, who refer to colonial landscapes as terrains of conflict and resistance. Here she considers disputed objects such as headstones and street signs as symptomatic of the nation as "the contended object in everyday discourse involving both colonists and colonized groups." Indeed, as another study by Divya Tolia Kelly's shows, even idealized 'colonial' landscape photographs and paintings can refer to a multifaceted desire to belong, remember, and perform new citizenship identities.

In a way, this goes to demonstrate that it is not just things but that places themselves may be considered worthy objects. The physical constructs of their architectural form may manifest a need for spatially specific negotiations as they simultaneously draw attention to wider political, legal, and theoretical claims as metonymic place objects. Some examples here include the work by geographers Don Mitchell on the basketball courts in People's Park, Berkeley; the California prison by Ruth Gilmore; Guantanamo Bay's detention centre in Claudio Minca's study; and Vancouver's public library in Loretta Lees' work.

Object as Commodity

Since the 1970s, the term 'object' in geography has been very closely associated with the idea of the commodity, a terminology which recalls most strongly Marxist studies. Here, the object is, unlike Hegel's spirit or what Marx himself calls the specter, decidedly tangible, grounded, and realized. In it rawest form, the object is the human laborer. Divorced from their own sensories, 'dead labor' is transformed into commodities which are then sold for a profit. Marxists are thus interested in the dynamics of this relationship. In particular, they wish to uncover and demonstrate how exactly the terms of exchange are unequal: 'labor power', it seems, has a peculiarly unique 'use value' of creating more than is necessary to reproduce itself; and it is not recompensed as such. In this way, the commodity, deconstructed, becomes the very transformative material necessary to govern radical politics and inform social structures. As geographer David Harvey says "the grapes that sit upon a supermarket shelves are mute; we cannot see the fingerprints of exploitation upon them or tell immediately what parts of the world they are from."

Critical geographers, especially those of a Marxist leaning, have taken up Harvey's implicit challenge to expose the 'exploitation' that have been produced and sustained in, and because of, space. It helps, as Michael Watts (2001: 182–212) says, that "commodities are preeminently geographical objects." By looking through the prism of oil in Nigeria, his research is shot through with accounts of pipes and poverty, discrepant local politics and a complacent global myopia, and violence fueled by oil's unceasing desire and demand as it is by its consummate commodity form – money. It highlights a paradox of the plenty – that those in closest physical proximity are also those for whom, economically, socially, and politically, the oil is unreachable – and speaks of this 'petrocapitalism' as a fairy tale turned nightmare: oil is a commodity that for many literally dictates life and death. In less (perhaps) severe circumstances, similar studies have been undertaken on the globally traded commodities such as tea and coffee, as well as that of the myriad studies on water, which is often associated with the work of geographer Eric Swyngedouw.

Geographers also point to the wider engagement of commodities, not just with labor but also with the environment, or more pertinently 'nature'. Here nature is not cast as a serene respite but a 'factor of production', valued not for what it is but what it can provide: nature is commodified. The work of Margaret Fitzsimmons and Noel Castree, among others, demonstrates how nature undergoes this process of commodification. Similar to what environmental activist Vandana Shiva terms 'biopiracy' in regard to the patenting of the basmati rice in India, researchers have pointed to the troubling ways in which indigenous knowledge and resources are appropriated from nature for profit purposes. Usually carried out without due acknowledgement or even at the expense of local communities to whom their ownership belongs, they question if 'selling nature to save it' through practices such as bioprospecting is a pragmatic step toward environmental conservation or rather a further step toward plundering nature. The contestations that are being fought by various organizations and communities over animal and plant species recall philosopher Michel Foucault, for when nature, in this case, is made a commodity: 'knowledge is power'.

Questions, furthermore, surround any unproblematic conception of what, indeed, constitutes 'nature'. For one, there is a risk of hypostatizing it as a universalized material object in a way that returns labor to its abstraction. It also appears to define, by contrast, a readily accepted concept of culture. It was perhaps this fear that geography's engagement with the material was solely around the idea of what James Duncan calls 'object fetishism' that ignited what later came to be known as the 'cultural turn' in geography. This was a reaction, sometimes unfairly, to versions of two subdisciplines that were mentioned before: cultural landscape studies and its uncritical description of objects, and Marxist geography that was so consumed by ideology that an object could not be anything but a festishized commodity. Together with such key influences like Denis Cosgrove and Stephen Daniels, they combined the visibility of landscape objects with the critical lenses of left geography to demonstrate how objects should not be reified as simply cultural or political. Rather it was necessary that these depicted objects underwent intense scrutiny in aid of uncovering the 'what, who, and how' of representation through discursive practices. As a study of the 'iconography of landscapes', texts were favored as the medium in which one situated and positioned an object and reflexively hinged meaning to depending on the prevailing social and cultural contexts.

The reading these objects met with prolific efficiency of a built form that was gaining popular and academic interest at that time: the mega shopping mall. Here, a singular internalized space could shift the focus from the production of objects to their seamless presentation. Seamless did not necessarily mean coherent, for here was a chance, as the work of sociologist Rob Shields and geographer Jon Goss show in their discussion of the West Edmonton Mall in Canada and the Mall of America in Minnesota, to create a symbolically idealized landscape far away from the harsh realities of everyday life. Thematized and estheticized, material allusions of landscapes, places, and things were brought together to imply fun and pastiche, fantasy over reality, and a compressed space/time of everywhere and nowhere. According to sociologists Jean Baudrillard and Mark Gottdiener, text was overtaken by semiotics and the sign value of objects eclipsed the commodity's use value. When the sign appeared more real than the actual object it represented, it could finally escape its signifier and float detached from its object.

This dismemberment of the sign from the object took its toll. In 2000, a seminal paper by geographer Peter Jackson called for a 'rematerializing (of) social and cultural geography'. This paper, often cited as a reference point to the 'material turn' in geography, was not a call to raze the landscape of its erroneous theoretical or objectified leanings but, quite simply, to build a bridge. Mirroring a sturdy bridge, objects would constantly demonstrate a physical resilience in geography; as its metaphoric structure suggests, it was to allow access to a multitude of things and communication between theories; as passage, it spoke of an always unfinished business of things and the necessity of crossing divides to make things matter more. This programmatic essay was loosely based on two important approaches rising from contemporary engagements with the commodity.

First, extending the work of modern thinkers such as Georg Simmel and Walter Benjamin, there has been an impetus to bring together 'use' and 'symbolic' values such that the focus is on the social and cultural context of an object's 'exchange' value, which itself was already a marriage of concerns regarding the reproduction of society (classic sociology) and the meaning of its objects (classic anthropology). Attention is paid not just to the political and economic reality of an object but also to the cultural understanding and practice that produced and informed it. To recognize it as such, it is necessary to 'follow the social life of things'. This idea, drawing on the work of anthropologist Arjun Appadurai, traces the history and biography of things as it simultaneously details the reciprocity between the economic and the cultural 'regimes of value'. In this way, it contributes to at least three preoccupations regarding things that geographers have picked up on: (1) that things are fluid, often in motion, and are multiply positioned; (2) that sometimes there is a need for both surface and depth accounting of an object; and (3) that as shown in the different approaches of scholars such as George Henderson and Neil Maycroft, the semantics of the value categories themselves (what does 'use' mean?) can be opened up and questioned.

Second, while the commodity implies consumption, not all consumption involves the act of purchase. Indeed, the strong alliance of material culture and consumption studies, with geography, turns this issue on its head such that the importance of the material object is not as a commodity but rather the commodity as one form of many others to be consumed. The work of anthropologist turned material culturalist' Daniel Miller is very influential in this alliance. By relating the basic interaction between people and things as having undergone a process of externalization (alienation of goods from institutions of origin) followed by sublation (society's reappropriation of externalized form in forming the social subject), a Pandora's box of materials, contexts, means, meanings, and practices is proliferated. As he shows with the iconic Coca Cola and its varying proper, mythic, and popular identity, the figure of 'value' can be broadly understood and widely embraced. In this way, subsequent studies have focused on the interactions between 'real' and 'virtual' (Internet, GPS, computer games, and artificial intelligence) spaces, varying situations (epochal and everyday) and the corresponding political, ethical, and esthetical concerns that inform and constitute these spaces. Geography then is coming to terms with its material landscape – in a way that things are no more fixed in time and space than they are in everyday life; no more made than they are creative and imaginatively used; no more objectified than the lenses through which they are discursively and practically conceived; and no more a reductionist approach of the material to the social as the social is embedded with material things.

Networking Materials/Material Networks

Objects that constitute the material landscape brought into being by networks. The first of two sets of networks, aligned closely with consumption studies, places strong emphasis on the types of connections that elucidate these objects. Here, the collaborative project of food scholarship and geography has produced some of the most lucid accounts of the discipline's 'material turn', utilizing concepts such as 'commodity chains', 'circuits', and 'food networks' to interrogate the nature of these connections. Some of these are portrayed as journeys, literally, through the eyes of chickens, pigs, and cows through filières from farm to the dining table. The 'invisible mouth' of consumers, to use Stewart Lockie's term, chomps its way through these journeys too, becoming aware of chains lengthened by distribution and retail networks, as well as the global movements of fish and imitation crab, French beans, and sushi, to name just some examples, as illustrated in the work of Becky Mansfield, Susan Freidberg, and Ted Bestor, respectively.

What these linkages often provide is an insight into the slice of life of the food object at hand. They may detail, for example, how the various material risks that are posed – antibiotics, bovine spongiform encephalopathy (BSE), *Escherichia coli*, dioxin, and hormones – refract along the way not just the constitutive change to the (becoming) food but also its physical, psychological, and emotional impact on those along the chain. When these dividual components or attributes of food are put under the microscope, the choices that can be, or are made, are accentuated. The work of David and Michael Goodman, among others, point to how the inadequacies of conventional food systems have led to an array of alternative food networks. Avoidance of the risks mentioned above, for instance, has led to the proliferation of organic networks such as that evidenced in the work by geographer Julie Guthman. Farmer's markets and local food cooperatives, to name but two places, have at heart an interest in shortening commodity chains, but strengthening their networks. The traveling food object, in these cases and more generally, meets with issues of knowledge, in both expert and profane forms; trust brought about by network 'transparency'; and imagination, when attempts are made to spur vicarious imaginaries of transnational food items that congregate on a supermarket food shelf. The term 'circuit' thus is preferred to emphasize the food item's lack of unidirectional movement – interactions with people and other entities to and fro inform the network as much as the object passing through it.

These circuits have presented geographers with points of entry into explicating networked objects more generally. Ian Cook *et al.* consider what they call the 'stuff geography' of mundane, everyday things, to profit by delving into such openings to manifest their myriad connections: these have resulted in fresh understandings of human liveliness in networks with ballet shoes, mobile phones, and spectacles, to give a few examples; or, as their work on bananas and papaya show more explicitly, it may present them with a political moment to 'get with the fetish'. Geographer Elaine Hartwick's call to do just that, for example, makes possible not just the carrying out of 'engagement tactics' within networks, but shows how it is precisely these networks that allow for such political action. Like Alex Hughes' analysis of auditing the Kenyan cut flower industry shows, the issue of responsibility, discussed by Clive Barnett and Paul Cloke among others, is one of the strongest axes on which many a material object pivots its political and ethical strength. The numerous accounts of fair trade products as a show of ethics are a testament to the strength of an object's advocacy.

The second set of networks follows the theoretical preponderance of actor–network theory (ANT). Associated most prominently within geography with the work of Bruno Latour, ANT emerged from the discipline of the social studies of scientific knowledge in which the practice of science in places like laboratories was found to betray a whole range of set categories. Pertinently, this act of boundary keeping was considered to have overestimated the ontologies of both the 'subject' and 'object'. To take the case of the latter, objects are not considered to have existential ideals. They are constantly in a triage of stability, materiality, and fluidity, and are labeled as objects mainly because they are, in the words of ANT proponent John Law, 'an effect of (relatively) stable arrays or networks of relations'. In this sense, while they resemble the first set of networks as utterly connected objects, ANT emphasizes, in this second set, their utterly entangled natures such that they should be more accurately termed 'quasi object' and 'quasi subject'. Both are also said to be 'actants'; that is, they both have 'agency' to effect, and enact a network. By extension, they subscribe to what has been termed a materialist semiotics. As described by Ann Game and Nick Bingham among others, it means that all entities including texts may, without any predetermined privilege, be seen to impact on their corresponding networks.

Geographers such as Jonathan Murdoch have probed further to investigate how the degree of embeddedness of objects, or as is often the case in food networks, the quality of objects may be portrayed in this seemingly 'flat' understanding of a network that bears no power relation. Geoffrey Bowker and Susan Star, in their book *Sorting Things Out*, have attempted to answer this question. Networks have different degrees of ordering and stability which relate to the ease of performing a network into being. Concordantly, they have always exhibited 'geographies of enablement and constraint'. Indeed, the authors use a hospital setting to show how list making, classification, and labeling (e.g., the Nursing Intervention Classification) can produce a 'Jacob's ladder' phenomenon that normatively privileges or excludes a range of treatments which them affects a patient's material reality through variations in physical care all the way to her/his success in medical insurance claims.

The etymology of 'thing', as philosopher Martin Heidegger proclaimed, has always meant gathering around and deliberating on a contested or pertinent matter. A thing therefore cannot simply be a fact or an artifact. In this way, an assembly of things is made present through its configurations of connection, frequencies of circulation, and positioning within the network. It is also through such modes of ordering that the 'other'– quasi objects and quasi subjects – is configured: through disposal, encumbrance, expulsion, or by sheer absence. In either case, the presence or absence of a 'thing' is always already as a hybrid, as Sarah Whatmore shows in her book *Hybrid Geographies*. Here the ideas of vitalist philosophers such as Gilles Deleuze and Felix Guattari are employed to show how through the metabolic exchanges between human and nonhuman matters, an assembly of things gather and emerge as an assemblage. This assemblage is no more simply the material objects that have enjoined them together but a complex engagement which cannot do without these material objects. What such an approach also does, therefore, is interrogate material entities at the boundaries of their prescribed categories. The work of 'new animal geographies' is particularly productive in this matter. In Hayden Lorimer's close analysis of the reindeer, Jamie Lorimer's discussion of concrakes, Sarah Whatmore on the elephant, and both Kay Anderson and Lewis Holloway on the human/animal collective in zoos and agricultural shows, the previously undifferentiated matter of an 'animal' is opened up to reveal the intermingling of species idiosyncrasy, cultures of knowledge, and contextual practices.

Embodiment

The embodiment of material objects, in this section on the 'material turn' in geography, brings together new and established scholarship within geography, which speaks of the direction that objects, things, matter, and artifacts may take as an ongoing engagement with the discipline. The first of this may be considered a reply to a broadly feminist critique that the conceptual notion of 'hybrid' in network geographies has led to a loss of subjectivity and a place for the corporeal being. Here, following the work of geographers such as Alison Blunt and Robyn Dowling, the setting of the home and its 'homely' objects can provide an insight into how a seemingly resilient ordering of domesticity is filled with places and things that reciprocally inform the entwinement of fear and safety, violence and intimacy, alienation and belonging, and nostalgia and progress. These places, which could include the exterior space of the garden as detailed by Russell Hitchings or the interior space of kitchen as Janet Floyd's research shows, are filled with various sociotechnical devices – not only home furnishings but also refrigerators and canned food – which proliferate assemblages that pay greater credence to the performativity of domestic objects and bodies.

The second considers the influential work of geographers Louise Crewe and Nicky Gregson on what they call the 'second hand culture' of things. Here, the machined working of capitalism is overlain with ingenious and alternate spaces such as car boots and garages, charity and retrothemed shops which seek to, as it were, extend the social life of things. Including objects as diverse as clothing, furniture, toys and do it yourself (DIY) products, the embodied qualities of both the human and nonhuman materials necessitate work and effort on these very objects in view of attaining certain forms of disembodiment and erasure. This, they have asserted, is carried out through rituals of divestment, such as cleaning and purification, to ensure that the object is 'reenchanted' and ready to enter, once again, new cycles of consumption. However, for better or worse, traces remain as 'new' and 'second hand' cultures do not make up the totality of the world's networks. Even discarded, unwanted, and expulsed hybrids create their own orderings, as explicated by geographer Tim Edensor in his discussion of waste matter. Here, he looks at how the 'congealed life' of these various things embodies within them potentialities which transform them from being artifacts toward nondiscrete entities exhibiting sensual and esthetic properties.

What this leads to, in the third instance, is an understanding of an object beyond its use, sign, and exchange value. As Karl Marx himself often affirmed, there would be no social transformation without an attendance to the sensory qualities of labor. Geographers have, by extension, concerned themselves with the similar sensory qualities of material objects. They have employed the ethologist James Gibson's concept of 'affordance' to refer to the way a material object's properties is disposed, or 'afforded' to those which it comes into contact with. In a way that remembers that the embodied object is always discursive, creative in a variety of ways and able to assume multiple identities, the 'affordance' of an object is a sum total of its practical (social arrangement) and expressive (affective) order. It is thus not just that human geographers have started paying attention to sensory objects but coincidently that these objects have in their material nature provided for these sensory experiences. Indeed, those attending to what is known as geographies of affect have explored these material forces and metabolic exchanges between human/nonhuman bodies and the corresponding feelings and emotions that are expressed in such encounters.

The fourth aspect, which is connected to the scholarship of affective geographies, is one that takes increasing care in examining the scale at which the various constituents of matter take place. Geographer Matthew Kearnes (2003: 139–152), for example, cautions against posing matter as a "universally undifferentiated conditionality." Rather, attention should be paid to the scale at which the seemingly imperceptible has an 'expressive physicality', which is in fact affectively perceptible. This, it has been suggested, is the nanoscale at which atoms hold within them their affective capacity. Research on the resultant nanotechnologies that are produced joins up neatly with some earlier scholarship in which Nikolas Rose, among others, have referred to as the (bio)politics of life. In this way, the hybrid forms of human/nonhuman are not just added on to the bodies in the guise of pacemakers or artificial limbs but also partake of an assemblage to become the very materials of living, and indeed life itself. Research in geography like that carried out by Bruce Braun on the avian flu and Ebola virus has begun to make headway into the political and ethical business of life's objects.

Finally, new material objects that are continually surfacing and networked into the discipline often attempt to stretch the limits of the object's ephemeral yet efficacious possibility. A final example here is Ben Anderson's work in music as a figure of matter which attenuates a feeling of boredom. Music notes and instruments that bear out the beat of the actual material object – in this case the radio player – transmit waves that move sound to the ears and its subsequent bodily experience and modification (of feeling less bored). As unruly and irreducible as the link to the effect that is produced might be, there is no doubt that the material assemblage that brought about this change of mood is integral, and that the persistence of sheer material objects contributed to this effect.

New Material Entanglements

An exclamation by Noel Castree regarding the notion of nature may be aptly subjected to a word substitution to show the vagaries and social life of material objects within the discipline of human geography. The word and material trope here is car: 'The car is dead. Long live the car!' From its genesis, as detailed in the work by Gerry Kearnes, the car has been an object that pervasively carries out time–space compression like no other. It has aided patterns of urban development especially, and unfortunately contributes to the exploitation of oil as a hotly contested commodity. Metaphorically, it has inspired Peter Jackson's reference to the 'traffic in things' as both a movement between use and exchange value as well as an indication that this movement is not unidirectional. As a symbol, it represents in various cultural context freedom and wealth and is discursively referenced as a sign of modernity. Its sign, however, cannot be divorced from its use value for a car does not amount to much if it does not move. Because the car journeys, it allows for the presence of mega shopping malls in the outskirts of the city as it allows for a trip to the zoo to see 'nature'. It is also popularly enacted as part of an array of virtual reality games. Often, it is the movement and the presencing of the driver as it is of the car. It is thus a sociotechnical device that is a set of relations between a number of things: car batteries, electrons, engineers, laboratories, and so on. In this way, it is an assemblage of things, parts of which may be called into action, such as car boots for the sale of second hand things. Finally, the car is accommodating to new ways of travel and engagements. Methodologically, as shown especially in the work of Eric Laurier, it opens up new textual and conversational analysis of and within the car with the aid of contemporary research tools such as the video camera which enhances the sensory qualities of this thing held together by a relatively stable configuration called 'the car'.

Conclusion

Notably, the usage of the terms objects, matter, material, and thing has endured some interchangeability and slippage here. The literature discussed has mostly borne out preferences for certain terminology either through distinction from earlier sets of ideas or through a purposeful ambiguity to question or disrupt fixed meanings and usages of what are necessarily complex terms. Like in the discussion of the car in the previous section, the word object and its closely related word 'objectify' has fallen out of favor for it implies a bounded entity or essentialized nature or identity (as in 'object of desire'). It remains, however, the most shorthand way of initiating an assemblage or network. Matter as a physicality, meanwhile, attests to an urgent importance in the discipline to recognize the varying scales at which things matter, an expressiveness that is suitable regarding global consumption as it does molecular viruses. The term material then deftly assembles all of geography's things: it forges an unruly alliance with the discipline's politics, ethics, and esthetics; it splices the workings of identity, knowledge, and power through these engagements; it enacts networks in motion and creates sociotechnical devices that challenge the categories of object and subject as it does the boundary between humans and nonhumans. There is a liveliness in material objects: things matter to geography, and they are certainly here to stay.

See also: Actor-Network Theory/Network Geographies; Affect; Animal Geographies; Body, The; Commodity Chains; Consumption; Cultural Turn; Fair Trade; Food Networks; Food Networks, Alternative; Human-Nonhuman; Jackson, P.; Landscape; Malls/Retail Parks; Material Culture; Networks; Sensorium; Social Studies of Scientific Knowledge; Subjectivity.

Further Reading

Anderson, B. (2004). Time stilled space slowed: How boredom matter. *Geoforum* 35, 739 754.

Appadurai, A. (ed.) (1986). *The Social Life of Things: Commodities in Cultural Perspective*. Cambridge: Cambridge University Press.

Blunt, A. and Dowling, R. (2006). *Home*. Oxon: Routledge.

Bowker, G. C. and Star, S. L. (1999). *Sorting Things Out: Classification and Its Consequence*. Cambridge, MA: MIT Press.

Braun, B. (2007). Biopolitics and the molecularization of life. *Cultural Geographies* 14(1), 6 28.

Clarke, D. B., Doel, M. A. and Housiaux, K. M. (eds.) (2003). *The Consumption Reader*. London: Routledge.

Cook, I. *et al*. (2004). Follow the thing: Papaya. *Antipode* 36(4), 642 664.

Cook, I., Evans, J., Griffiths, H., Morris, R. and Wrathmell, S. (2007). 'It's more than just what it is': Defetishizing commodities, expanding fields, mobilising change…. *Geoforum* 38, 1113 1126.

Edensor, T. (2005). Waste matter the debris of industrial ruins and the disordering of the material world. *Journal of Material Culture* 10(3), 311 332.

Gregson, N. and Crewe, L. (2003). *Second Hand Culture*. Oxford: Berg.

Jackson, P. (2000). Rematerializing social and cultural geography. *Social and Cultural Geography* 1(1), 9 14.

Kearnes, M. (2003). Geographies that matter the rhetorical deployment of physicality? *Social and Cultural Geography* 4(2), 139 152.

Law, J. (2002). Objects and spaces. *Theory, Culture and Society* 19(5/6), 91 105.

Miller, D. (1989). *Material Culture and Mass Consumption*. Oxford: Blackwell.

Watts, M. (2001). Petro violence: Nation, identity and extraction in Nigeria and Ecuador. In Peluso, N. & Watts, M. (eds.) *Violent Environments*, pp 182 212. New York: Cornell University Press.

Relevant Websites

http://www.consume.bbk.ac.uk
Cultures of Consumption.

http://mcu.sagepub.com
Journal of Material Culture.

McDowell, L.

A. Jones, Birkbeck, University of London, London, UK

Linda McDowell (**Figure 1**) has made a sustained and important contribution to feminist, social, and economic geography over the course of her career. She is best known as a key proponent of feminist geography, and her empirical work has ranged across issues such as the gendered nature of work and labor markets, identity, and transnational migration. More widely, she has made important contributions to theoretical debates in human geography on gender and place, as well as having made significant interventions in debates surrounding qualitative methods and their value for human geographic research.

Linda McDowell was born in Stockport in Greater Manchester in 1949, and then educated at Cambridge University, graduating in 1971. She studied for masters in planning (1971–73) at the University of London and moved to work with the University of Kent at Canterbury for her first job. She then moved to the Institute of Community Studies in the East End of London, and

Figure 1 Linda McDowell.

during her time there she decided to register for a PhD at the Bartlett School of Planning, University College London, where she had done her masters degree. She did this as a part time student, supervised by Peter Cowan and working on housing change in two London boroughs. Through this period she joined the Open University as a lecturer and also had two young children. The experience of balancing raising a young family and studying for a PhD has been undoubtedly influential in her subsequent academic research interests.

Even before finishing her PhD, she took a lectureship at the Open University where her research interests continued to focus on housing and social transformation during the early 1980s. However, through the 1980s an interest in gender relations and feminist theory came to the fore, notably beginning with a paper on feminism in the first issue of the journal *Society and Space* which was joint authored with Manuel Castells. By the time she moved to the University of Cambridge in 1991, she was arguing the case for much greater attention to be paid to gender within human geographical research and the need for a more explicit feminist human geography. This perspective formed the basis for her groundbreaking study of the gendered nature of working practices in the City of London during the 1990s. In *Capital Culture*, McDowell argued that the workplaces of investment banks in London remain highly gendered environments. Theoretically, she explored how the few successful women at the top of the banking sector are those who have successfully defined themselves as 'honorary men' in a masculine world. She thus reveals the complexity of the gendered performances that constitute the City workplace, and in doing so challenges much of the preceding theoretical debate about women's disadvantage in the labor market.

Following on from this study, *Gender, Identity and Place* examined the wider relationships between gender and place at a range of scales from the body to the nation. McDowell narrated the history, geography, and politics of feminist geography since its inception, and outlined the important theoretical shifts that began to appear in the 1990s. She argued, as has been the case in other disciplines, that these shifts pivoted around the recognition of the "theoretical and political importance of…multiplicity and difference," and in particular, the need to challenge "the phallocentrism and universalism of Western discourse" (McDowell, 1999: 215). For Linda McDowell this shift was occurring most significantly in two areas: the general "turn" to questions of meaning and representation in addition to "the more usual interest of (feminist) geographers in material social relations" (McDowell, 1999: 245) and, second, the movement away from the idea of "feminist geography" as it was conceived in the 1970s, to pluralist "feminist geographies."

After a period as vice principal of Newnham College in Cambridge between 1997 and 1999, Linda McDowell moved to a chair, first at the London School of Economics, and subsequently, at University College London. During this period, her empirical attention moved from business elites in key positions of power and influence to those experiencing labor market marginalization. In *Redundant Masculinities*, she examines gender 'as a social, cultural and economic force' in the lives of white working class young men in two contrasting UK cities. Based on a series of extensive interviews with young men, she shows how UK media images of working class young men as 'idlers, layabouts, or yobs' profoundly misrepresents the complex intersection of economic and sociocultural change. She shows how these young men "on the verge of adulthood" can "no longer rely on the moral certainties that con structed men as breadwinners, whose identities were constructed in the main in the workplace, and women as primary carers of dependants in the home" (McDowell, 2003: 24). The argument she constructs through exam ining the various dimensions of these young men's lives is that they are "neither rebels nor yobs, neither failures nor successes, but instead they lead careful, constrained local lives" which are based in attempts "to construct a version of acceptable working class masculinity in the face of economic circumstances that militate against their efforts" (McDowell, 2003: 237).

In many ways, Linda McDowell's distinctly geo graphical take on socioeconomic change and gender has constantly returned to the theme of how material and contextual circumstances at the wider scale intersects with microscale issues around gender, bodies, social practices, and specific circumstances. Her historically grounded life narrative research, presented in the book *Hard Labour*, takes a similar but different cut at theorizing the experiences of female Latvian migrant voluntary workers in the UK. McDowell's research is based on lengthy interviews with some of the 3000 or so Latvian women who came to work in the UK in the immediate aftermath of the chaos and political upheaval of World War II. Here, she analyzes the hybrid identities of these migrants, and the relationship their sense of identity had with the 'women's work' they came to the UK to undertake. She analyzes the strategies these women de veloped for coping with hard manual work, and the ne gotiations between their "residence in British working class neighborhoods and their determination to reconstruct an imaginary version of a pre war bucolic Latvian communal identity" (McDowell, 2005: 8). Her recurrent theoretical interest in the relationship between gendered identities, economic circumstance, and cultural change are brought into particularly stark relief as she considers the 'betweenness' of these migrant women in post war Britain as they negotiated their dual roles as wives and as participants in the labor market.

Throughout her career, Linda McDowell has also made a number of important contributions to methodo logical debates within human geography. Her inter disciplinary approach has led her to argue for a more effective combination of quantitative and qualitative techniques, and her research has made wide use of re search techniques such as ethnography and life histories, which are not widely adopted in contemporary human geography. As such, she has made important contributions concerning the positionality of the researcher in the re search process, elite research methods, and the difficulties associated with different types of qualitative method.

Overall, Linda McDowell is most widely known for her major contribution to the development of feminist (economic) geographies, but the wide ranging scope of her work means that it is often cited across the full breadth of subdisciplines within and beyond human geography. Her work has also been translated into French, Spanish, Chinese, and German. Taking a chair at the University of Oxford in 2004, she has continued to act as a prolific contributor to debates on gender, labor markets, identity, and the interaction between economic and societal transformations. She is thus notable as a figure within the discipline who, in the view of David Harvey, is one of the relatively few geographers to con sistently produce major theoretically infused books grounded in empirical work.

See also: Economic Geography; Feminism/Feminist Geography; Harvey, D.; Masculinities; Migration; Oxbridge Geographies; World/Global Cities.

Further Reading

Harvey, D. (2006). The geographies of critical geography. *Transactions of the Institute of British Geographers* 31(4), 409 412.

McDowell, L. (1989). Gender divisions. In Hamnett, C., McDowell, L. & Sarre, P. (eds.) *The Changing Social Structure*, pp 158 198. London: Hodder and Stoughton.

McDowell, L. (1991). Life without father and Ford: The new gender order of post Fordism. *Transactions of the Institute of British Geographers* 16, 400 419.

McDowell, L. (1991). Multiple voices: Speaking from inside and outside the project. *Antipode* 24, 56 72.

McDowell, L. (1992). Doing gender: Feminism, feminists and research methods in human geography. *Transactions of the Institute of British Geographers* 17, 399 416.

McDowell, L. (1997). *Capital Culture: Gender at Work in the City of London*. Oxford: Blackwell.

McDowell, L. (1998). Elites in the City of London: Some methodological considerations. *Environment and Planning A* 30, 2133 2146.

McDowell, L. (1999). *Gender, Identity and Place*. Cambridge: Polity.

McDowell, L. and Sharp, J. (eds.) (1999). *A Feminist Glossary of Human Geography*. London: Arnold.
McDowell, L. (2001). Father and Ford revisited: Gender, class and employment change in the new millennium. *Transactions of the Institute of British Geographers* 24, 448 464.
McDowell, L. (2001). It's that Linda again: Ethical and practical issues in research with young men. *Ethics, Place and Environment* 4, 88 100.
McDowell, L. (2003). *Redundant Masculinities*. Oxford: Blackwell.
McDowell, L., Perrons, D., Fagan, C., Ray, K. and Ward, K. (eds.) (2005). *Gender Divisions and Working Time in the New Economy*. London: Edward Elgar.
McDowell, L. (2005). Love, money and gender divisions of labour. *Journal of Economic Geography* 5(3), 365 379.
McDowell, L. (2005). *Hard Labour: The Forgotten Voices of Latvian Migrant Volunteer Workers*. London: UCL Press.
McDowell, L. and Massey, D. (1984). A woman's place. In Massey D. & Allen, J. (eds.) *Geography Matters!*, pp 195 215. Cambridge: CUP.

Relevant Websites

http://www.geog.ox.ac.uk/staff/lmcdowell.html
Lind McDowell's homepage.

Media

D. B. Clarke, Swansea University, Swansea, UK

Glossary

Media Means of communication, particularly mass communication (the press, radio, television, etc.).

Mediascape The 'media landscape', in any one of three senses: (1) the makeup of the media as a whole; (2) the world portrayed by or in the media, either implying disjuncture between it and the world it purportedly represents or in a more phenomenological sense; and (3) a mediatized landscape within which individuals may trigger digital media (for purposes of surveillance, consumer targeting, or audio-visual tours).

Mediasphere The totality of media forms, their relational configuration, and dynamic evolution; an 'ecosystem' for media.

Mediatized Affected or altered by the presence of the media, particularly in the sense of the reconfiguration of social relations, institutions, and, especially, events.

Multimedia Associated with information-technology applications, the combined use of more than one media form, such as text, graphics, audio, video, animation, and interactivity.

New Media New technological means of communication, especially those associated with electronic, digital, and information technologies (e.g., the Internet).

Introduction

The term 'media' (the plural of 'medium', a Latinate term meaning 'middle', now commonly used in the singular as a mass noun) refers to the heterogeneous array of means of communication characterizing human societies, especially various means of mass communication (print, radio, television, etc.) characterizing modern societies. The term 'media' is, in fact, a truncated form of 'media of communication', which effectively highlights the 'channels' through which communication takes place. While 'communication' once signaled physical transit as readily as the transmission of information ('communications' may still refer to road and rail networks, air transport links, etc.), the term 'media' signifies those technologies that somehow substitute for physical movement in the conveyance of information of one kind or another; technologies that accomplish the disassociation of transportation and transmission (e.g., telegraph lines, television broadcasts, the Internet, etc.). Definitions such as this are, however, somewhat misleading, for as Marshall McLuhan famously proclaimed, "the medium *is* the message." Only the most naïve of understandings would conceive of a medium as a strictly neutral channel for conveying information. Consider the differences between the news reported in print, on the radio, and on television: the properties of various media (e.g., their capacity to show images, still or moving) fundamentally affect the reporting of news stories. Form affects content in the most profound of ways; while content cannot exist independently of form. Thus, language itself may be mediated in two broadly independent ways, corresponding to its spoken and written forms. The former typically makes use of sound waves; the latter relies upon marks on the page. Jacques Derrida famously argued that spoken language is in no sense an authentic, originary form, verbalizing pure conscious intent, with written language a mere supplement and subsequent development. All language is capable of being written (it possesses a 'graphematic' quality) even before writing *sensu stricto* appears on the scene. Writing may, however, be regarded as something like a prosthesis, extending the human voice – and human memory – beyond its normal capacities (the reason one writes letters, makes a will, takes lecture notes, and makes shopping lists). Harold Innis categorized writing as a 'time biased' medium, dedicated to transcending the present. Nonetheless, speech relies precisely on the same mediation through signifiers as writing does; albeit signifiers of a different kind, utilizing different media.

Most media may be thought of as having their basis in technologies that serve in some way as prostheses, extending particular human capacities beyond their previous bounds (this, at least, is the basis of McLuhan's understanding). For example, the invention of the telephone bypassed the need for face to face interaction, allowing speech to be transmitted (initially down phone lines; latterly by wireless means) beyond the previous constraints of space (i.e., those set by the distance a voice would carry). In Innis' terms, the telephone exemplifies 'space biased' media, dedicated to 'shrinking' space; while recording technologies, starting with the gramophone, extend the temporal bounds of speech – in much the same manner as writing (as exemplified by the grooves etched into wax cylinders on the earliest gramophones). Modern telephone systems (those with 'voicemail') incorporate both these extensions. The interdependence of space and time does, however, render Innis' conceptualization somewhat shaky. Insofar as writing transcends the present, it permits communication without co presence in space – thereby extending the spatial range of communication (as exemplified by the mail system). Just as the mail system and the telephone created possibilities for the coordination of far more complex and distanciated sets of social relations than was hitherto possible, the media in general serve to 'stretch' social systems across space and time (as Anthony Giddens would put it) or to 'compress' time–space (as David Harvey might prefer).

It is not difficult to see why McLuhan should have enthusiastically coined the term 'global village' to encapsulate the possibilities he envisaged in the wake of the exponential development of media technologies, particularly with the advent of electronic means of communication. Here, however, the dangers of technological determinism, famously signaled by Raymond Williams, should not go unremarked. The capacities that particular media technologies have the potential to actualize may or may not be realized. Yet while different societies, at different times and in different places, may take up and respond to the media in different ways, justifiable concerns over technological determinism should not detract from the ways in which particular media technologies do indeed serve to 'reconfigure' space and time, as contributors to the volume edited by Brunn and Leinbach sought to demonstrate. As Paul Virilio notes, the theoretical limit of the transmission technologies that succeeded transport technologies is the speed of light – canceling out distance; permitting near instantaneous contact; creating a kind of 'telepresence'. With the appropriate technology, a heart surgeon in Berlin might operate on a patient in Johannesburg without leaving home. Yet while revealing that physical distances have always been a function of the speed (and cost) at which they could be overcome, Virilio remains more a technological indeterminist than a technological determinist, attuned to the fact that the media never simply mediate; they form part of an increasingly complex, 'mediatized' world. "Why is there somewhere rather than nowhere?" is the question that Virilio forces us to ask.

Mediaspheres: Power and Influence

One significant quality of most modern media technologies has been their tendency toward 'massification'. While writing using a quill or biro permits the transmission of information without the need for co presence in time or space, the invention of the printing press (*c.* 1450) ramified this capacity to the *n*th degree – with truly revolutionary implications. Little wonder that McLuhan should have referred to the 'Gutenberg galaxy' to convey the astronomical impact of a medium that permitted the transmission of information and ideas to far greater numbers of people than ever before. Yet despite McLuhan's maxim, the relationship between print media and knowledge is not straightforward. It is telling that the most famous book produced by the inventor of the printing press was the Gutenberg Bible. As a technology originating in deeply religious times, the printing press did not immediately or automatically overthrow prevailing ideas and beliefs. Nonetheless, McLuhan's point still carries weight. **Figures 1** and **2** plot subscriptions to Diderot's and d'Alembert's *Encyclopédie*, a product of Enlightenment Europe, published in France and intended to overthrow traditional knowledge and to undermine its religious basis. The distance decay effect and concentration of uptake around urban areas notwithstanding, these maps illustrate the extensive reach made possible by relatively cheap reprographic technology (the maps record subscriptions to the mass produced quarto and octavo editions of 1777–82), as well as the permeability of national boundaries.

Later audiovisual media, like cinema and television, significantly amplify these same properties, reaching mass audiences with increasing rapidity. For McLuhan, the electronic age was marked by a speed that imploded space. The point is well illustrated by the news media. The instantaneous 24 hour global television news coverage now offered by many of the world's largest media corporations may not be evenly accessible over the surface of the Earth (**Figures 3** and **4**), and there are significant concerns about who controls the news, but this is profoundly different from a situation where the dissemination of news relied upon word of mouth, which meant that news tended to be concentrated within major settlements, diffusing elsewhere slowly, irregularly, and unevenly. Ironically, the pace of the media today has accelerated to the point where news coverage increasingly 'anticipates' events: reporting on what is set to take place, filling in time while waiting for something to happen (dubbed 'newzak', after Muzak).

Between wall to wall television news coverage and the public proclamations of town criers, newspapers, radio, and cinema newsreels demonstrate the contingent nature of technological change: the venerable institution of the press survives, even if that relative upstart, the newsreel, proved to be a vanishing mediator *en route* to television news. The forerunners of today's newspapers have existed since the early sixteenth century, though low literacy levels meant that town criers were not immediately redundant. The survival of the daily newspaper, an institution dating back to the early eighteenth century, relates to consumer habit, convenience, marketing, commercial interests, and myriad other factors that have created a highly complex media environment. This complexity is likely to persist, even in the wake of 'technological convergence' (the ability of digital devices to approximate to other forms of media). Online newspapers have not stopped people from

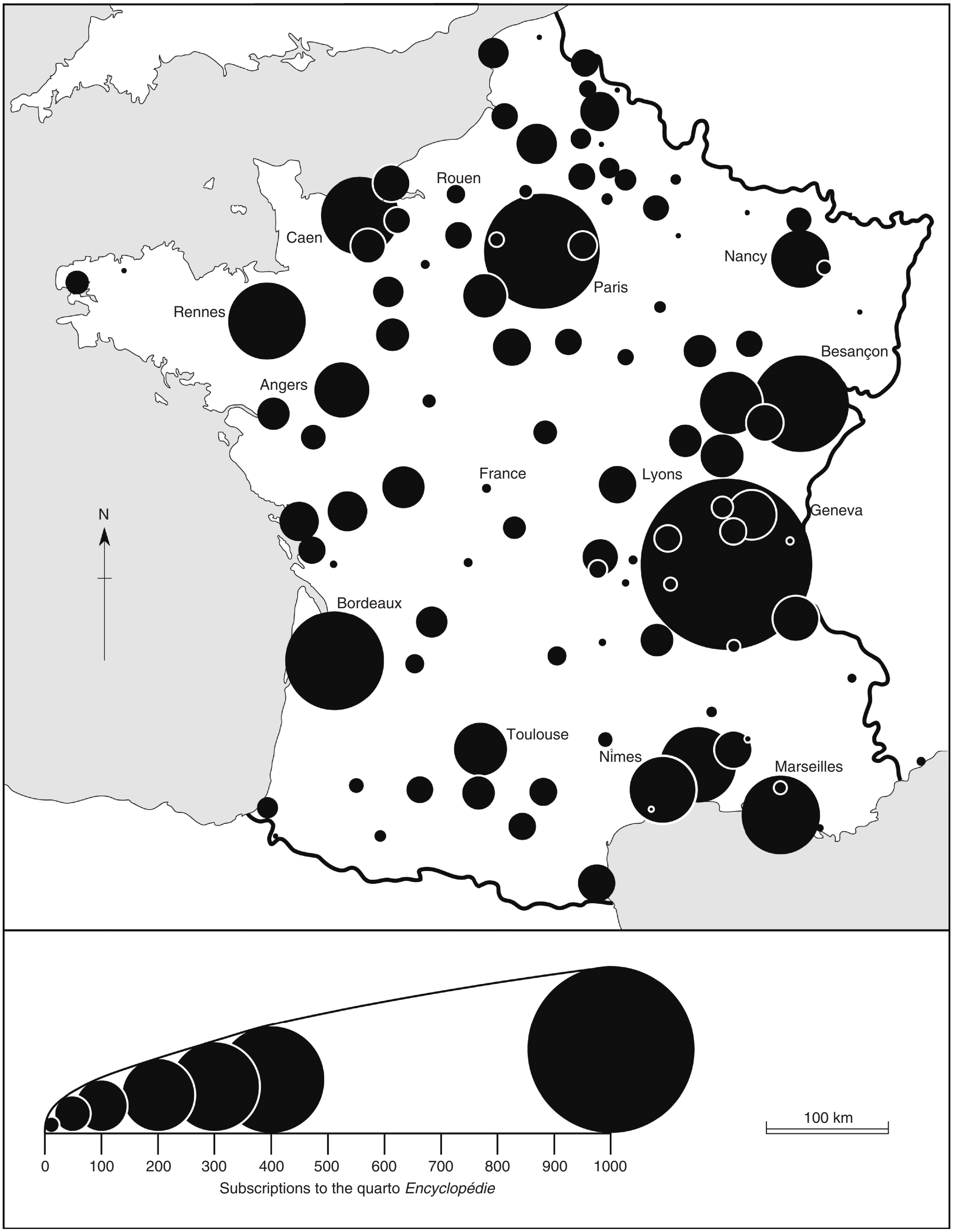

Figure 1 The distribution of the quarto edition of the *Encyclopédie* in prerevolutionary France. Adapted from Darnton, R. (1973). The *Encyclopédie* wars of prerevolutionary France. *The American Historical Review* 78(5), 1331 1352.

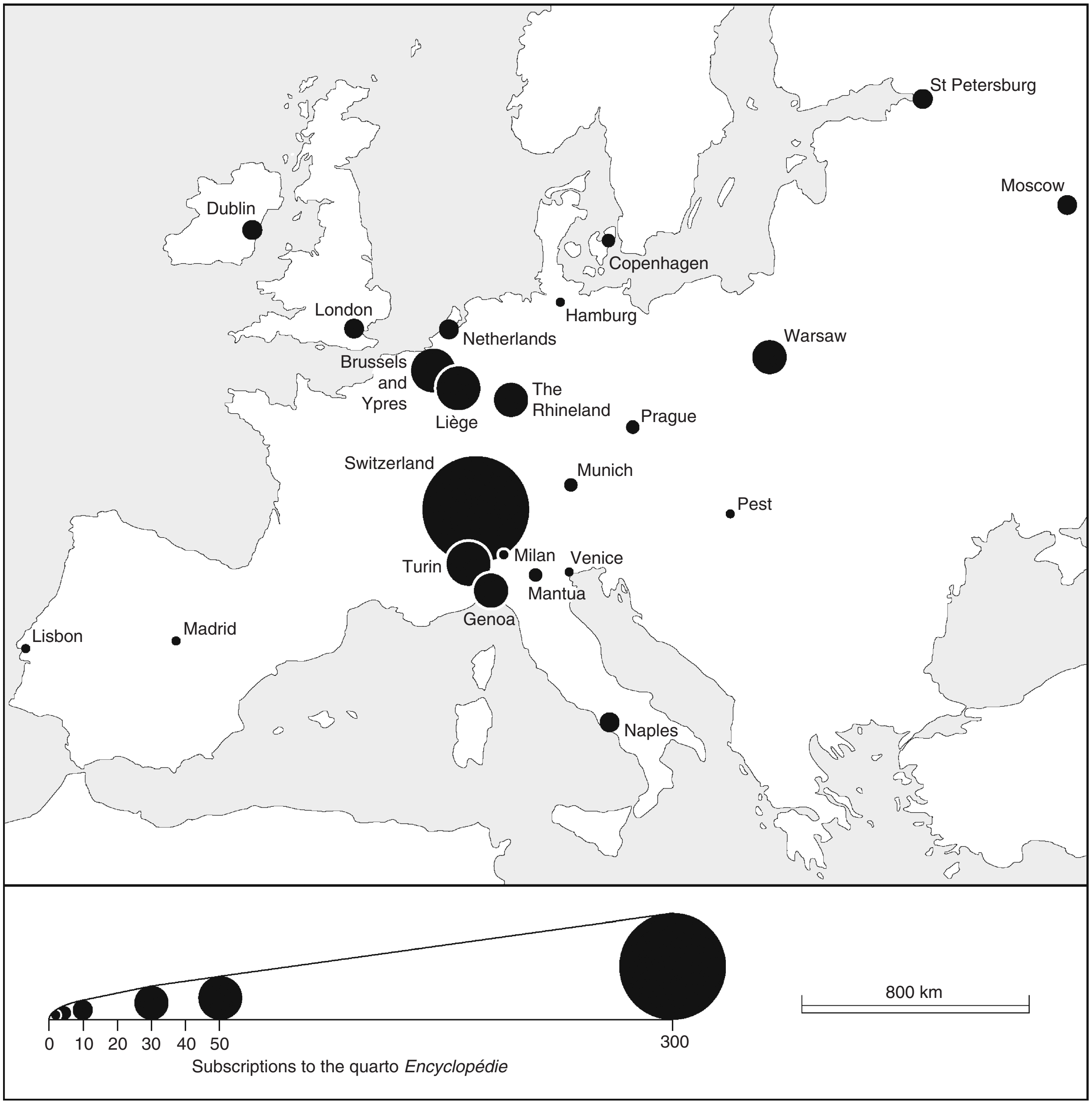

Figure 2 The diffusion of the quarto edition of *Encyclopédie* beyond France. From Darnton, R. (1979). *The Business of Enlightenment: Publishing History of the Encyclopédie, 1775 1800*, p 301. Cambridge, MA: Belknap Press.

reading the paper form, for example. Likewise, despite many dire predictions to the contrary, television did not kill off the cinema (the fate of the newsreel notwith standing). Hollywood's enormous economic and cultural influence, as documented by Allan Scott, is in fact in creasingly tied to other media, as a result of the dra matic concentration of capital in some of the world's largest media conglomerates (a process known as 'media convergence'), though many national film industries – 'Bollywood', Korean cinema, etc. – are also enjoying sig nificant growth and increasingly transnational distribution.

The media environment in its totality has been dubbed the 'mediasphere' by Régis Debray. Debray's coinage is intended to operate in much the same way that the term 'biosphere' works with respect to life, implying an ecosystem for media – subject to dynamic evolution and generally organized around the latest media technology to rise to dominance. There have been numerous attempts to trace out media chron ologies. Debray's conception entertains three key pos sibilities to date: the 'logosphere' (dominated by orality, with writing permitting the diffusion of information);

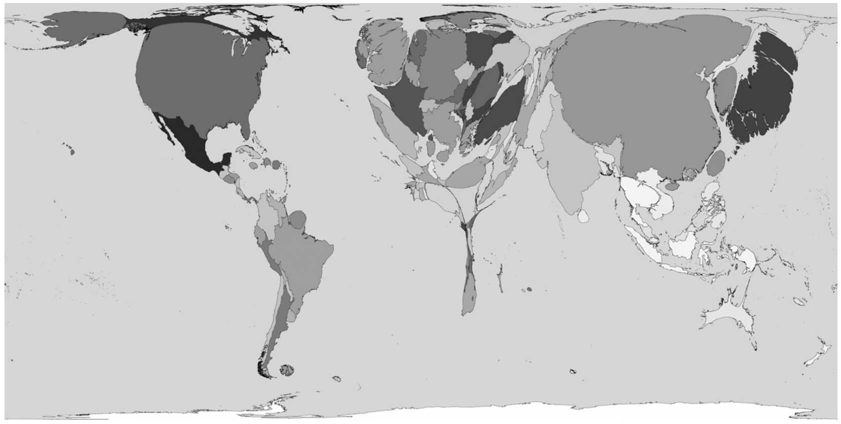

Figure 3 The world mapped according to the number of television sets in use (2005). Note: territory size proportionate to share the sum total of television sets in use. Copyright 2006 SASI Group (University of Sheffield) and Mark Newman (University of Michigan) (http://www.worldmapper.org).

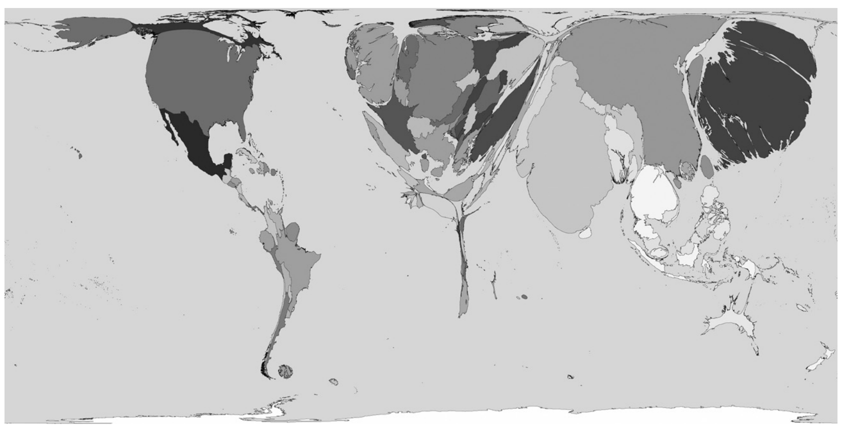

Figure 4 The world mapped according to the circulation of daily newspapers (2005). Note: territory size proportionate to the sum total of daily newspapers in circulation. Copyright 2006 SASI Group (University of Sheffield) and Mark Newman (University of Michigan) (http://www.worldmapper.org).

the 'graphosphere' (dominated by print technology); and the 'videosphere' (associated with the rise to dominance of various audiovisual media). One might reasonably extend this schema by adding a fourth, the 'cybersphere', to designate a mediasphere dominated by digital information processing and new media technologies. However, to the extent that Debray's work remains more or less faithful to the media chronologies established by McLuhan and Innis, a more radical understanding might be gleaned from Friedrich Kittler's dismissal of McLuhan's starting point on the media as bodily prostheses. Rather than seeing media technologies as arising from human needs, Kittler insists on the 'medial *a priori*': the network of media technologies that underpins the capacities of a given culture, shaping subjectivities and corporeal capacities in the first place. Kittler's conception highlights our increasing dependence on the mediation of social life through media technologies. It also delivers a different media chronology, identifying a discontinuity brought about by the advent of analog media technologies (e.g., the cinema, the gramophone), which culminate in digital information technologies. Whereas the preceding culture, underpinned by writing and the printing press, relied on filtering reality through a sign system (such that recording anything in the world involved its transcription into preexisting linguistic terms), a culture possessed of analog or digital technologies has no need of such reductive transcription. It is capable of 'sampling' the world and 'remixing' it. The upshot is an increasingly complex situation, marked by the proliferation, mutation, and acceleration of media technologies, while direct access to the underlying programming languages becomes increasingly unattainable. To the extent that it seriously reworks the terrain charted by the likes of McLuhan, Kittler's work has significant implications for considering the media's reconfiguration of space and time.

Media Representations: Texts and Audiences

Medium theory, following the spirit of McLuhan's dictum that "the medium is the message," and epitomized in the work of Joshua Meyrowitz, sees the media as largely destructive of a sense of place. In distancing themselves from the idea that media messages are over determined by the medium, most media studies attentive to geography adopt a more empirically directed stance. On the one hand, such work emphasizes the spatial variability of media 'provision' and 'usage' (including phenomena such as the overleaping of intermediate technological forms, exemplified by the uptake of the mobile phone in Africa). On the other hand, such work has highlighted place based differences in audience 'responses' to particular media texts, especially the local 'reception' of global media products (US soap operas in Trinidad; James Bond films in China; etc.). A further tradition of work concerns itself with a more esthetic evaluation of media texts, focusing on the geography 'in' rather than the geography 'of' the media (why Marlboro cigarette adverts feature cowboys in wide open spaces; why sci fi films adopt particular conventions to make the diegetic future cities they portray recognizable as futuristic and as cities). Empirically oriented work on media provision, usage, effects, and texts derives from a variety of traditions of study, ensuring a shifting terrain that is difficult to negotiate. Changes in media technologies, moreover, constantly alter the meaningfulness of particular terms and debates. For example, what does it mean that television, once a mass medium dedicated to broadcasting entertainment and 'infotainment' – more recently giving way to 'narrowcasting' to 'target audiences' – is now most prolific as a means of surveillance, in the form of closed circuit television (CCTV)? Likewise, what are the implications of the fact that, despite the predominant focus on the audience in media studies, new media technologies have 'users' rather than 'audiences', forged in the image of Alvin Toffler's 'prosumers' (who *pro*duce and con*sume* at one and the same time)? Or again, how far does the experience of cyberspace, as an experience of inhabiting a series of virtual worlds, question the validity of thinking in terms of representations of space? A better sense of these issues will emerge from following the contours of existing work and drawing attention to some of its limitations.

Regrettably, much existing work on the media relating to space and place does merely adopt predefined areas as convenient settings for the study of media provision and reception, empirically documenting media usage within particular national spaces, for instance. Recognition of the permeability of national boundaries to increasingly global media flows, however, often underlies such studies – a theme that is brought into sharp focus by the work of authors as varied as David Morley, Kevin Robins, Nick Couldry, and Mackenzie Wark. While the local reception of media products produced elsewhere is sometimes framed in terms of cultural imperialism, given the global dominance of the US media corporations (**Table 1**), a considerable body of work has aimed to demonstrate that matters are far more complex 'on the ground', showing how local understandings may differ significantly from those intended by those responsible for producing media texts. This focus on audience 'readings' takes its cue from the British cultural studies tradition, specifically from the insistence on the 'relative autonomy' of culture instilled by Stuart Hall and the Birmingham Centre for Contemporary Cultural Studies. This amounted to an attempt to reconcile the 'dominant ideology' encoded into media messages (and 'decoded' by media theorists) with the possibility of other decodings (e.g., oppositional readings), largely by focusing on 'actual' media subjects rather than relying on media theorists. This tradition was given a geographical airing in the volume edited by Jacqueline Burgess and John Gold. More recent work in this vein has jettisoned the dominant ideology thesis in favor of a 'postmodern' emphasis on the pleasurable, empowering potential of popular culture. There are, however, dangers to this approach, not least because, in attempting to resist characterizing media subjects as dupes, there is a tendency to assume that virtually any meaning can be read into any text. Undoubtedly, people respond differently to media texts, and the determinants of their decoding practices are affected by a variety of 'situational', 'geographical' factors. Yet work emphasizing this has an unfortunate tendency to impose a thoroughly Western, humanist model of the subject as freely constitutive onto other cultures – neglecting the fact that meanings produce subjects as much as subjects produce meanings. Despite seeming to ignore the readings of 'actual' subjects in favor of 'abstract' readings proffered by theorists, a focus on how media texts 'position' subjects may offer a more convincing account of the consequences of media globalization than accounts purporting to give voice to 'local' appropriations of meaning.

In the face of widespread media globalization, a significant amount of work concerned with space and place

Table 1 The world's largest media corporations

	Annual revenue (US$ billion)	
	2001	*2006*
Time Warner	27.3	43.7
The Walt Disney Corporation	23.4	31.4
News Corporation	13.6	23.9
Bertelsmann AG	13.3	22.9
CBS (Viacom)	20.2	14.4

Source: Company reports.

in the media, such as that collected in the volume edited by Leo Zonn, has been given over to considering the accuracy or otherwise of the media's portrayal of place, whether these representations appear in the news; in documentaries; or even in avowedly fictional works, which make no claim to re present reality. A small but growing body of work, such as that of David Crouch and others, has developed this idea in an alternative direction, suggesting that reality is being transformed by its attempts to live up to its media image, particularly in the cases of tourist places (e.g., 'setjetters' make movie locations their holiday destinations, demanding not to be disappointed with the reality they find there). There is an underlying problem with all such work, however, insofar as it assumes that reality and its representation are somehow separate; that re presentations, as the term itself would have us believe, passively reflect a preexisting reality – either accurately or with added distortion.

It is closer to the mark, however, to recognize that so called 'representations' partake of reality, actively shaping the world. Given that we have no direct access to reality; that the world is never immediately present but always mediated by human cognition; the very idea of representation, as a derivative image or copy of reality, is a misnomer. There is a kind of ontological 'flatness' between trees and photographs (of trees), cities and films (of cities), terrorist attacks and news reports (of terrorist attacks). Indeed, this is where the power of the media becomes most evident as a force in the contemporary world. Consider, for example, the 2001 al Qaeda attacks on the Pentagon in Washington and New York's World Trade Center. There can be little doubt that these attacks were deliberately planned as a 'media event'; that the perpetrators were fully aware of the global news coverage their actions would attract: the planes struck the screen before they struck the Twin Towers. This kind of situation writ large might be associated with the work of Jean Baudrillard. Often mistaken as an advocate of the idea that reality is increasingly entrained to its media image, hence becoming increasingly insubstantial, Baudrillard in fact argues that the distinction between image and reality has imploded in the opposite direction: that there is too much reality – or hyperreality. An example of this excess of reality is provided by that media staple, the opinion poll. Opinions are definitively not facts: they are meant to be argued with; to provoke a response. Yet opinion polls treat opinions as facts, to be counted up and analyzed, creating a 'truth' against which one can no longer legitimately argue. The weight of opinion becomes a dead weight. Far from enhancing communication, 'means of communication' inhibit the possibility of any response. The media no longer communicate, they excommunicate – as attempts to forge a communion of the ex communicated epitomized by canned laughter, TV and radio phone ins, and 'interactive voting' neatly illustrate.

Baudrillard's arguments notwithstanding, increasing appeal to 'representation' in the sense of democratic representation has been made in work dealing with the relations between the media, space, and place. For many, the issue centers on the way in which the media has colonized the public sphere. This relates in no small part to the fact that most media institutions are commercial organizations, ultimately dependent on advertising money, as Armand Mattelart and Michael Chanan make clear. With media capital being progressively concentrated into a handful of media giants, the media is increasingly characterized by power without responsibility – a formulation classically alluding to media barons, like Rupert Murdoch of the Australian based News Corporation. More generally, the globalization of the media has seen a purely commercial model rapaciously crowding out erstwhile alternatives, notably the public service broadcasting that once dominated the national spaces of many 'mixed economies'. The state ownership and control of the media that characterized the former Eastern bloc countries has not quite been consigned to history, but the increasing permeability of national boundaries accompanying changes in media technologies render such state control more difficult than was previously the case (the rise in state surveillance of Internet use notwithstanding). Where the commercial media, ostensibly free from overt political interference, has become dominant, however, serious concerns arise over the kind of freedom this affords. The debate owes much to Jürgen Habermas's notion of the public sphere, while a good deal of it focuses specifically on the possibilities of community media, typically, though not necessarily, operative at a local level. Such issues have been debated most compellingly in particular political contexts, such as post apartheid South Africa, as exemplified in the work of Clive Barnett. Yet the extent to which the media as a whole might be opened up to more democratic control remains contentious. New media technologies create possibilities that hold out the hope of more democratic forms, such as the 'blog' (or weblog). Yet while there are examples of the power of 'blogs' to bypass mainstream media and cross international boundaries (such as the case of Salam Pax, the 'Baghdad Blogger'), this potential is often short lived. On the one hand, the state rapidly catches up with technological change, as witnessed in the acquiescence of Google to state Internet surveillance in China. On the other hand, such developments are quickly appropriated and assimilated by the mainstream media: 'bloggers' are co opted as journalists, while grainy videophone images are adopted as signs of authenticity, eyewitness accounts adding to the authority of global media corporations (the irony being that telepresence apparently demands evidence of embodied presence).

Being and Being Connected

A media saturated world entails that, alongside our im mediate environment, we increasingly inhabit a media tized environment, created by television, the telephone, the Internet, etc. The interface or interaction between these environments is, however, complex. Work on 'media rituals' suggests that the media operates in the secular sphere in much the same way as rituals that function to mark the boundary between the sacred and the profane: media events like state funerals, royal weddings, world cups, or terrorist attacks work, at the level of collective consciousness, to forge some degree of social integration. If such work is dogged by functionalism, it is nonetheless true that media usage is far from an isolated individual act. The sense in which the media increasingly mediate any form of sociality is, perhaps, the most significant as pect of any geography of the media, for all media trans form social relations based on proximity in space. One daring prediction, made shortly after the invention of the telephone (and greeted with incredulity at the time), speculated that "one day, every small town in America will have one." The telephone was initially seen as a broadcast medium, like radio (or 'wireless telegraphy'). Today, as Manuel Castells and colleagues have noted, access to the media at fixed points in space is increasingly being surpassed by mobile communications that use multimedia devices to connect people wirelessly into a network that is largely indifferent to proximity in space. Yet the increasing importance of connection to people who are not co present in space, to information stored elsewhere, and to entertainment that does not rely on the proximity of others is always only ever enabled, rather than determined, by media technologies. To that extent, the form of the media and the spatiality it promotes is symptomatic of the shape of the social world as a whole.

See also: Cyberspace/Cyberculture; Discourse; Discourse Analysis; Movies and Films, Analysis of; Telecommunications.

Further Reading

Balnaves, M., Donald, J. and Donald, S. H. (2001). *The Penguin Atlas of Media and Information*. New York: Penguin.
Barnett, C. (2003). *Culture and Democracy: Media, Space and Representation*. Edinburgh: Edinburgh University Press.
Brunn, S. D. and Leinbach, T. R. (eds.) (1991). *Collapsing Space and Time: Geographic Aspects of Communications and Information*. London: Harper Collins.
Burgess, J. and Gold, J. R. (eds.) (1985). *Geography, the Media and Popular Culture*. London: Croom Helm.
Castells, M., Qiu, J. L., Fernandez Ardevol, M. and Sey, A. (2006). *Mobile Communication and Society: A Global Perspective*. Cambridge, MA: MIT Press.
Cavell, R. (2002). *McLuhan in Space: A Cultural Geography*. Toronto: University of Toronto Press.
Couldry, N. and McCarthy, A. (eds.) (2003). *Media Space: Place, Scale and Culture in a Media Age*. London: Routledge.
Crouch, D., Thompson, F. and Jackson, R. (eds.) (2005). *The Media and the Tourist Imagination: Converging Cultures*. London: Routledge.
Darnton, R. (1973). The *Encyclopédie* wars of prerevolutionary France. *The American Historical Review* 78(5), 1331 1352.
Darnton, R. (1979). *The Business of Enlightenment: Publishing History of the Encyclopédie, 1775 1800*, p 301. Cambridge, MA: Belknap Press.
Howley, K. (2006). *Community Media: People, Places, and Communication Technologies*. Cambridge: Cambridge University Press.
Mattelart, A. and Chanan, M. (1991). *Advertising International: The Privatisation of Public Space*. London: Routledge.
Meyrowitz, J. (1987). *No Sense of Place: The Impact of the Electronic Media on Social Behavior*. Oxford: Oxford University Press.
Morley, D. (2006). *Media, Modernity and Technology: The Geography of the New*. London: Routledge.
Morley, D. and Robins, K. (2004). *Spaces of Identity: Global Media, Electronic Landscapes and Cultural Boundaries*. London: Routledge.
Scott, A. J. (2005). *On Hollywood: The Place, the Industry*. Princeton, NJ: Princeton University Press.
Wark, M. (1994). *Virtual Geography: Living with Global Media Events*. Bloomington: Indiana University Press.
Zonn, L. (ed.) (1990). *Place Images in Media: Portrayal, Experience and Meaning*. Savage, MD: Rowman and Littlefield.

Medical Geography

R. Earickson, University of Maryland, Baltimore, MD, USA

Glossary

Carrying Capacity The maximum number of people that can be supported on Earth, or in any of its regions.

Endemic Disease A disease that is constantly present to a greater or lesser degree in people of a certain class or in people and/or animals living in a particular location.

Environmental Physiology The study of the adaptation of individuals to environmental change and stress.

Ethnomedicine The study of the beliefs and practices concerning illness stemming from indigenous cultural development; it observes and describes hygienic, preventive, and healing practices, also taking into account temporal, and spatial references.

Holistic Medicine A system of healthcare based on a concept of the 'whole' person as one whose body, mind, spirit, and emotions are in balance with the environment. As such, treatment may include conventional medicine and various nontraditional methods of diagnosis and therapy. Personal responsibility for health is stressed, and surgery and prescription drugs are avoided whenever possible.

Iatrogenesis Refers to the causation of a state of ill health or adverse effect or complication caused by or resulting from medical treatment, unintended effects of health and medical behaviors or policies, or conditions in medical facilities that themselves make people sick.

Political Ecology The study of political struggles for control over natural resources, or of political encounters whose outcome is determined by differential access to natural resources.

Prevalence In epidemiology, it is the total number of cases of a given disease or disorder in a specified population at a specified time regardless of when the illness began. Thus, the prevalence of tuberculosis in New York City since 2000 includes the cumulative number of New Yorkers suffering from the disease since that year. This is often confused with incidence, which refers to only new cases diagnosed during a particular year.

Social Capital It involves feelings of trust and safety, the quality of connections among people living in neighborhoods, networks of family and friends, and the extent to which people are engaged in community groups, clubs, and other organizations.

Synoptic Climatology A holistic approach to the study of applied climatological issues that affect humans and other organisms around the world. Synoptic climatologists attempt to characterize an entire weather situation that exists in a given area at a given time, to gain a better understanding of both the atmospheric environment and its effects on the organisms that experience it.

Introduction

Medical geography uses the concepts and paradigms of the discipline of geography to investigate human–environmental relationships of disease, nutrition, and medical care systems, of which maps and geographic information science (GIScience) are arguably most prominent today. Therefore, site, situation, place, location, region, and all spatial concepts may be compared to the distribution and dispersion of most diseases, nutrition as a factor in sickness and health, and the supply and demand for healthcare resources. One of the predominant concepts of geography as a science is that it examines relationships between peoples and their environments in holistic terms. In the particular subfield of medical geography, the focus is upon those interactions that bear mainly, but not exclusively, upon human health within a variety of cultural systems and a diverse biosphere.

In the geographic tradition, medical geography is highly interdisciplinary and integrative. Its practitioners emanate from, or integrate with a range of disciplines, including anthropology, cartography, economics, epidemiology, history, human ecology, medicine, political science, psychology, public health, and sociology. Medical geographers are also, foremost, problem solvers who deal with both environmentally based problems and those of anthropogenic origin. Their workplaces include academia, private, and public organizations.

Origins of the Discipline

Most writings about medical geography attribute the origin of the academic discipline to the 24 chapters of *Airs*, *Waters*, *Places* from the collection titled the *Hippocratic Corpus*, or *Hippocratic Collection*. Many mistakenly attribute these writings to Hippocrates himself, when in fact he wrote none of them. While the vast majority of scholars believe that Hippocrates himself wrote *Airs*, *Waters*, *Places*, a lack of evidence to support this means that the works in this collection must be presumed anonymous.

Geographic historian Frank Barrett has written exten sively and eloquently about the origins and development of medical geography, covering the period from the mid fifth century BC to the mid twentieth century AD. It is clear from this record that there was some interest in the geography of disease earlier. The diffusion of information about diseases, nutrition, diet, and geographical variations in medical systems began to take off only in the eighteenth century, because of explorations, scientific experimen tation, and advances in printing. Thanks to a few physicians who wrote volumes on the subject, a tremendous amount of information about variations in disease and nutrition in the known world, and possibly the first disease map of the world, was added to the literature during this century.

The emergence of germ theory and the works of Pasteur, Darwin, Koch, and others produced writings in the nineteenth century that brought to the forefront the idea that infectious diseases are caused by microorgan isms, and added significant knowledge about variations in disease distribution and causation throughout the world. Up to this time, physicians had believed that the geo graphical variation of disease was due in large part to separation of the races, but Darwin's work made such a view untenable. Geographers were beginning to observe associations between geographical distribution and vari ations in diseases, medical treatments, and diet. In the latter half of this century, empirical research into disease, using numerical data, reinforced the new scientific para digms that spun off Darwin's theories.

The late nineteenth and early twentieth centuries saw the ascendance of the well known concept of environ mental determinism. Writers like Ellsworth Huntington claimed that hot, humid climates encouraged lassitude and elevated the risk of tropical infectious diseases. Cli mate 'explained' why indigenous peoples in the tropics were 'dull in thought and slow in action'. On the other hand, new discoveries in bacteriology and public health challenged these views with alternative, optimistic ex planations. This period also ushered in improvements in public health. Increased life expectancy in the Western world during this time was due more to preventive than to curative medicine.

The mid twentieth century brought the French sur geon Jacques May to America with a mandate from a geographical society to explore geographical influences on diseases. He acknowledged that epidemiology and parasitology predated medical geography, but explained that he was hoping to establish correlations between geographical aspects of physical, social, and biological factors and causative agents, vectors, intermediate hosts, reservoirs, and humans. Today, such associations are often referred to as 'ecological correlation'. May encouraged such a statistical approach, but acknowledged that reliable statistical data at a suitable scale was difficult to acquire – a lesson that would be visited on health and social sciences, and public health to the present day. He also recognized the complexity of these associations and that it was easier to establish statistical probability than to show biological links between diseases and suspected causative factors. May considered mapping the distri bution of diseases worldwide to be an important task for geographers at that time. His *Atlas of Diseases* certainly was a significant contribution to the field. He also pub lished *The Ecology of Human Disease* and edited a second volume titled *Studies in Disease Ecology*. May became a geographer through his mandate, but his true allegiance was to ecology, and a number of medical geographers who followed him would adopt and refine the mantle of disease ecology. Within a few decades, geographers like Andrew Learmonth would effectively define medical geography as the ecology of infectious disease. The work of those whose career defining years occurred since 1970 is described later.

Human Ecology and Disease

Geographers who have worked in the paradigm of human ecology consider their research as a study of cultural, environmental, and biological variables as a synergistic ecosystem affecting health. Ecosystems occur in space and exist in time; they are dynamic, interrelated, and multi dimensional. The fact that life is a series of spatial and temporal changes makes human ecology, like some vir uses, a moving target and thus a continual challenge. In her textbook, *Medical Geography*, Melinda Meade sets out a framework for the study of human health. Basically, there are three broadly conceived components in the human health ecosystem – habitat, population, and behavior.

Humans live and play out their roles in places; that place is 'habitat'. Microhabitats, such as houses and workplaces, schools, recreation spaces, transportation ve hicles, and macrohabitats such as deserts and rainforests are instrumental to illness and health. Some examples of illness in habitat are *Escherichia coli* outbreaks in restaurants or through inadequate food processing, SARS or influenza transmission among passengers in transport vehicles, and cholera epidemics following natural disasters.

'Population' includes the characteristics, status, and conditions of people as organisms. One aspect of Mal thusian dictum states that populations tend to increase geometrically. For example, in sub Saharan Africa, about 41 new births per 1000 population occur annually, while only 17 per 1000 die during that time. This is a natural increase rate of 2.4%. At this rate, the estimated popu lation of this region in 2005, 752 million, will increase to over 1.1 billion in 2025 and over 1.7 billion by 2050. Sub Saharan Africa's fertility rate is over 5.5 (the average number of children born to all women in the region over their reproductive lifetime). Twenty years earlier, the fertility rate there was over 6.5, and in some nations in the region, this rate exceeded 8! Thus, the reproductive potential for humans is considerable.

Opposing this biotic potential is the totality of environmental forces, abiotic and biotic, that prevent that potential from being realized. Abiotic factors are essentially what nature provides, including the physical geography of a place – climate, soils, and water quantity and quality determine the availability of nutrients, and the potential for food production. Biotic factors include humans, animals, insects, parasites, etc. – their competition for resources and their abilities to carve out niches in the ecosystem. Medical geographers take into account the variation in physical and cultural factors that affect human health worldwide. Included among these factors are the earth's 'carrying capacity', prospects for food production, soil destruction and desertification, genetically based differences, human 'environmental physiology' variations in the aging process, and cultural behavior.

It is difficult to exaggerate the effect of culture in human populations. For instance, culture appears to have a much greater effect on population growth than biological factors. Consider the impacts of abortion, infanticide, contraception, notions about coital frequency or extramarital sex, age of first marriage, and commonality of divorce on natality and related health services. Japan is an interesting case study in demonstrating the interplay of cultural factors in regulating natality. A complex web of legal and political history has led to a ban in Japan on chemical contraceptives commonly accepted in the Western world. The result is almost a total dependence on condoms, the rhythm method, and abortion to control birth. Japan, along with many nations in Europe, has one of the lowest fertility rates on earth. At the same time, there is a dramatic increase in the life expectancies in these (and other) countries. There is concern in these nations about the dearth of children and the future of labor availability, not to mention the exploding demand for elderly health services.

Humans face numerous stressors in their health ecosystem simultaneously. According to Meade, cultural behavior interacts with habitat and population in several ways – habitat conditions, wealth and technological attainment, the risk for health hazards, and the myriad differences in customs, nutritional, and immunologic status. All of these vary geographically and temporally. Add to these the propensity for inequalities in society, and one can appreciate the complexity of processes that affect disease and health in populations.

Malnutrition and Communicable Disease

From the point of view of a medical geographer, a 'landscape' is a place with a complex expression of abiotic and biotic processes. When one knows how to analyze its components, geographic and otherwise, one can usually determine the potential for any health (or other) problem there. This section explores medical geography's approaches to the problems of malnutrition and communicable disease.

Malnutrition may be defined as any type of poor nutrition, including having too much or too little food, or having an improper balance of nutrients. Whereas there are arguably enough food resources on earth to feed all people a diet with necessary nutrients, malnutrition is all too common. One way of viewing it is to characterize malnutrition as a problem of distribution, not production. Variability in ability to produce food at the local level is an important factor in creating pockets of malnutrition. In a sense, malnutrition is a measure of ecological failure; it is system feedback indicating that a population has not successfully adapted to its environment on either a temporary or permanent basis. Persons suffering from malnutrition (deficiencies or excesses) are more likely to get sick. Jacques May devoted some 20 years to the study of the ecology of malnutrition and left a voluminous literature on the subject.

Medical geographers have associated nutritional hazards with cultivation pattern, seasonal, or cultural conflicts. Topography, weather, soil, water supply and quality impose limits on what can be cultivated in any given place. The agricultural patterns of the world are determined by these factors. Culture has created distinct geographic dietary regions. Food production evolved from specific cultural hearths and then spread over paths of cultural contact. Over time, it became apparent that diversity and variety in diet were necessary to provide a full range of nutrients. A diet limited to staple foods is often associated with poverty, and so regional diets are often associated with specific nutritional deficiencies. Nutritional deficiencies have resulted in, to name but a few, 'kwashiorkor' (insufficient protein), 'marasmus' (insufficient calories), 'rickets' (vitamin D deficiency), and 'goitre' (iodine deficiency).

The power of prescription, prohibition, and cultural preference influences food consumption patterns in most parts of the world. Consider religious injunctions against eating pork in the Middle East, and cattle in India. The recent history of the native Inuit of Canada is instructional. Over the past century, the Inuit have gone through successive stages of culture contact with white Canadians. Once nomadic hunters, Canada's native peoples live today in modern towns and hamlets, working as wage earners. There is a running debate whether these people were propelled by provincial politics to settle in towns or whether they moved voluntarily to seek more secure access to food, supplies, and medical care. The Inuit perceive some of these settlements to be crowded and violent compared with their previous lifestyle, with overall lower quality of life, including housing shortages, pollution, noise, high prices, dangerous roads, and general pressure to drop certain traditions of the past and adopt trends in the White culture, such as eating 'commercial food'. Modern television advertising drives a thriving market for 'fast food', that is high in fat, salt, and calories, which arguably leads to excess body weight and obesity.

It has become apparent that many people would rather find a dietary supplement for this than to simply alter their diet, a condition that has profited drug manufacturers. Geographers research things like super market accessibility, since supermarkets that offer a wide range of fresh produce, whole grain products, and un processed foods at less expensive prices than conveni ence stores, are an important source of affordable and nutritious foods – one key to improved health.

The broad term, communicable disease, is used by the US Centers for Disease Control and Prevention to in clude all diseases attributable to a specific infectious agent (e.g., viral, bacterial, or parasitic) that is transmitted from an infected host to a susceptible host, whether directly or indirectly through an intermediary vector. Some diseases are spread by direct contact, where coughs, sneezes, or simple close contact with infected individuals can serve to pass the disease causing agents.

Though certainly not exclusively, medical geog raphers have paid substantial attention to vector borne disease, since such diseases have been recorded world wide and are persistent. Vectors are the vehicles by which parasites are transferred from an infected to a susceptible host. The most common vectors are arthropods – insects, spiders, and crustaceans among others. The geographic interest is understandable, since arthropods have rather specific niches in earth's ecosystem – mosquitoes and flies are at home in warm regions, while ticks, fleas, and lice are important vectors at high altitudes and latitudes. A good example is Malaria, which is caused by plasmodial parasites. Malaria may be transmitted to humans only if the vector, the Anopheles mosquito, is present to intro duce the parasite into the human bloodstream. Malaria, with over one million deaths annually worldwide, kills more people than any other vectored disease. Other common vectored tropical diseases (Chagas disease, dengue fever, lymphatic filariasis, leishmaniasis, oncho cerciasis, schistosomiasis, and trypanosomiasis) together are responsible for over 100 000 deaths annually. Several human activities may promote the emergence and spread of communicable diseases:

- encroachment on vector habitats, such as construction of settlements in brush and/or forested areas;
- changes in agriculture, such as introducing a large scale single product operation in an area formerly not in agricultural use, and/or sparsely occupied by in dividual households;
- destruction of forests through deforestation, and in fusions of settlement and transportation;
- uncontrolled urbanization, particularly in regions of developing countries where communicable diseases are 'endemic'. Typically, rapid growth concentrates large numbers of people in crowded areas with poor sanitation, which accelerates the transmission of communicable disease; and
- modern transportation, such as ships and other cargo carriers often harbor unintended 'passengers', such as arthropods and rats that carry infected microbes to destinations that formerly had no cases of certain communicable diseases. With international jet aircraft, the potential exists for a person or persons infected with a disease like SARS or influenza to carry the disease halfway round the globe before their first symptoms appear, by which time they may have infected others through exposure through coughing or casual contact.

Since spatial 'diffusion' is a topic of geographic import ance, it follows that studies of the diffusion of disease are common in the medical geography literature. Under standing the mechanisms that influence the spread of disease and its spatial pattern is at the core of the geog raphic study of diffusion. Most studies of this form of diffusion have emphasized an explanation of events in a space–time context. Variables such as distance, the lo cation of habitats, and the distribution of jobs and facil ities all influence the level and direction of interaction. Diffusion patterns are determined by the configuration of the networks that encourage movement and barriers that discourage it. Whether a malady spreads only locally, such as *E. coli*, or diffuses as great world pandemics, such as HIV/AIDS, disease threats challenge the development of predictive models that can benefit human populations. Such models have been developed both by medical geographers and epidemiologists. Some of the work of Rodrick Wallace and Peter Gould is instructive in this regard. **Figure 1** is a classic simple ecological correlation between disease diffusion and distance.

Human Health and Economic Development

Some resources essential to human health are basically nonrenewable (e.g., energy fuels, certain plant and animal species) and others are only slowly renewed (e.g., soil, water, clean air, and certain vaccines). Humans have created ecological problems by overuse and misuse of resources in the long term. Due to the continued growth of population in some regions and the desirability of some environments over others for economic develop ment, people have largely tended to maximize their short term gains in capital of all kinds rather than

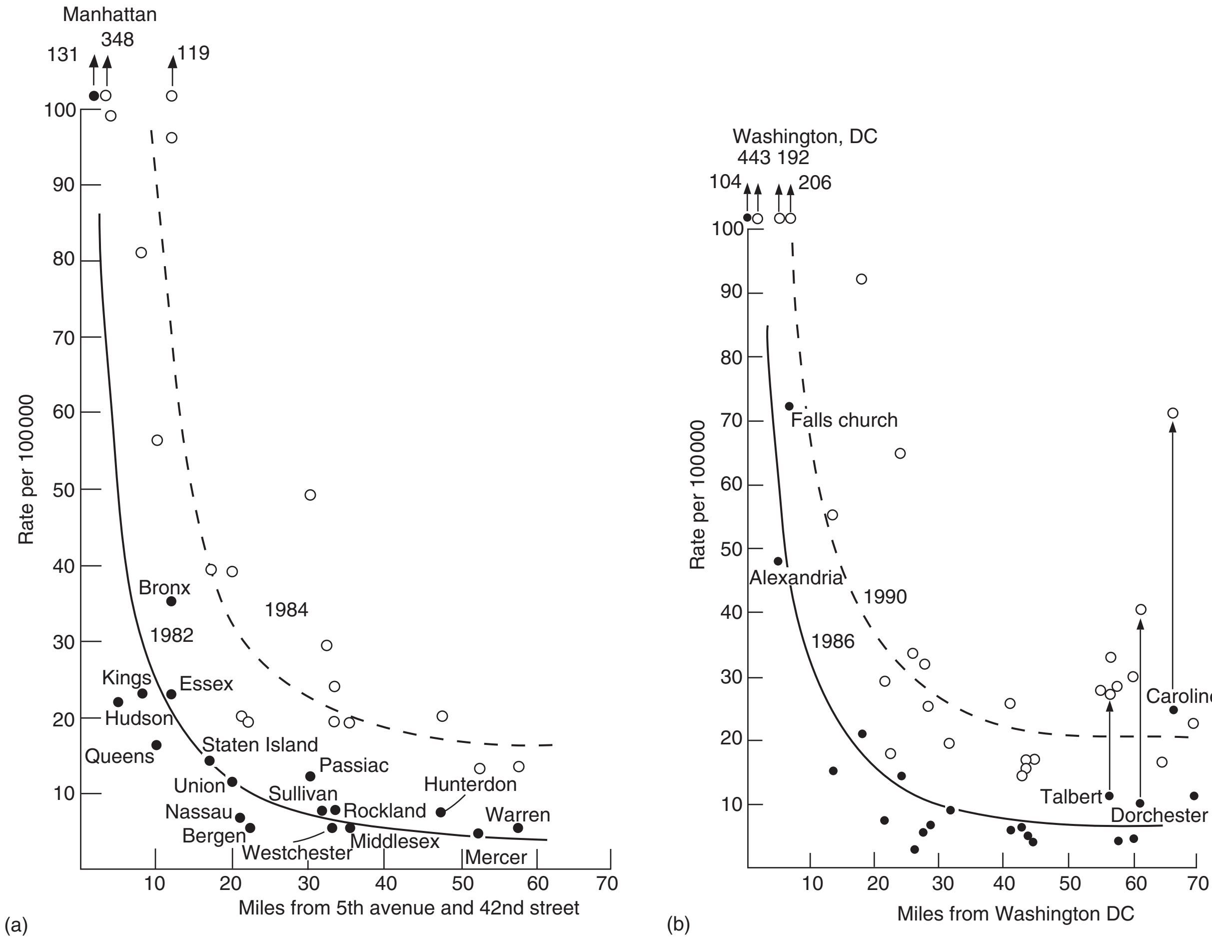

Figure 1 (a) The rate of AIDS cases per 100 000 population as a function of county distance from Fifth Avenue and 42nd Street in New York for 1982 and 1984. (b) Same as (a) for the Washington, DC area, 1986 and 1990. Reprinted from Wallace, R., Fullilove, M. T., Fullilove, R., Gould, P., and Wallace, D. (1994). Will AIDS be contained within U.S. minority urban populations? *Social Science and Medicine* 39, 1051 1062.

optimize long term conservation of resources. The result is inequities in morbidity, mortality, and the distribution of resources. High income countries, with less than 20% of the world's population, have a per capita gross national product well over 60 times that of low income less developed nations, which have over half the world's population but a life expectancy some 15 years less, and the gap continues to widen. Inequities can and do lead to unequal suffering from lack of necessities, morbidity, and mortality. Medical geography has long had an interest in the spatial distribution of health inequities and more recently to explanations that give some prominence to place. Some of these are the direct or indirect results of uncontrolled or unregulated economic development both at global and local levels.

Economic development has frequently been associated with the 'epidemiologic transition'. Disease patterns are related to many things, but especially socioeconomic status. It is well known that, with the advent of the Industrial Revolution, a gradual shift in mortality occurred, from communicable diseases to lifestyle diseases. Over the centuries, poor, rural populations have been drawn to industrial centers to take advantage of job opportunities in mining and manufacturing. As a result, they have left places where communicable diseases were responsible for most deaths and migrated to places where chronic and degenerative diseases became the major causes of mortality. As disease patterns change, the average age of death shifts from childhood and young adulthood to older ages, with corresponding increases in life expectancy. Chronic diseases have been linked to lifestyle changes (dietary changes, additional stress, sedentary employment, etc.) that are associated with economic development. A most recent consequence of epidemiologic transition is the reemergence of infectious diseases previously thought to be under control, many of which are antibiotic resistant, in addition to the rapid emergence of a number of 'new' diseases.

In underdeveloped tropical and subtropical regions, the effects of development frequently involve human conflict with the natural environment. Without sufficient resources to stave off illness from infectious disease, or lacking political will and/or economic means to prevent contamination, people succumb to any number of mal adies. Countries that have promoted large population resettlement schemes as a way of developing their na tional resources have subjected large numbers of people to lifestyle changes that were detrimental to their health. Settlement of the agricultural frontiers of the Amazon basin in South America is instructive. Agribusiness practices there were responsible for the spread of on chocerciasis and shistosomiasis into areas where they had previously been insignificant. Hughes and Hunter, and Stock's writings on development and health in Africa were an early introduction to the 'political ecology' of disease. They described some of the political, environ mental, and biological interactions that are necessary to understand patterns and changes in disease distribution in the tropics. Mayer has written further about the gen eral application of the concept to medical geography.

Lest we think that developmental health hazards are predominantly rural, consider that urbanization involves considerable demolition and construction, which means open drainage and septic ditches, garbage dumps, and water storage. Everywhere, we are creating new habitats for insects and they are becoming highly insecticide resistant. Dengue fever, for example, is spreading in the Caribbean and Central America. Chagas disease has followed people to cities where vectors thrive in poor urban housing and domiciliated animals.

Globalization, with its economic, social, environ mental, and policy dimensions, is also of interest to medical geography. In various places, we find that glob alization has exacerbated inequality in income distri bution and worsening of health status for low income populations. Health policies, both national and local, have been affected by development. The economic 'miracles' of China and India, for example, are obscured by some of the world's most severe environmental problems, which are taking a large toll on these nations' human and ecological health. Furthermore, modern technology has provided a transportation system that globalizes the disease infection process. There are no places on the habitable earth that are more than 30 h apart by air transportation. This phenomenon has been described as a virtual viral superhighway. On the positive side, globalization has promoted greater diffusion of communications on, among other things, women and children's rights and the HIV/AIDS pandemic.

Vital research into the fundamental changes in earth's ecosystems and how they impact health are ongoing in medical geography. A fertile area of research involves the health consequences of global warming. Combined land and sea temperatures have been going up worldwide since the mid nineteenth century. Drastic alterations in the distribution of vegetation are expected to occur more quickly than in the past. Experts anticipate changes in climate that are likely to lead to health problems related to heat stress, natural weather disasters and flooding, food shortages, changes in vector distribution, and, con sequently, infectious disease patterns. Tick borne dis eases and maladies caused by tropical and subtropical waterborne parasites are likely to expand their range into currently temperate zones.

Many health problems are likely to be indirect. Epi sodes of food shortage or prolonged heat stress may result from violent behavior associated with social cata clysm. Heat related mortality has been widely docu mented, especially among the poor and elderly. Global warming may disrupt ocean currents or alter trade wind or jet stream patterns, leading to an increase in the fre quency and severity of violent weather. Much medical geographic research may be found in the field of *Synoptic Climatology.* At the University of Delaware, USA, Center for Climatic Research Synoptic Climatology Lab, lab research projects include, among other things, the im pacts of weather and pollution, development of a heat stress index for public use, and a synoptic approach to monitoring surface ozone.

Before the nineteenth century, there was no sustain ability science, and there were few regulatory bodies with any authority. In the Western world, the direction and pace of improvement of environmental health was a matter for visionaries, pioneers, and philanthropic busi ness leaders. Today, public health has developed to the point where there is no excuse for rampant development without regard to its effects on health. Unsustainable economic development is well documented and un mistakably visible in real pain and suffering among the poor, working class, and beyond. The social and eco nomic costs of the damage being wrought are also much more measurable, as indeed are the likely additional costs if nothing is done. Medical geography continues to monitor global surveillance, and to promote sound health and development planning.

Chronic/Lifestyle Disease and Environmental Health

Economically developed countries have reached the end of the classic mortality transition, when death is caused mainly by degenerative and lifestyle diseases such as heart disease, cancer, stroke, diabetes, accident, homicide, etc. Risk of infant and child mortality is low and life expectancy is typically well into the 70s. Much of the research literature in medical geography, and some of its most highly developed methodology, are focused on patterns of health conditions and disease etiology in North America, Europe, and Japan. These patterns are increasingly shared with recently industrialized coun tries, especially in Asia. Modern, affluent nations also have the trained personnel and the financial capacity to engage the distribution, causation, treatment, and pre vention of the pathogenic complexes of the twenty first century.

There is a substantial literature on health inequities. Much of this focuses on inequities that are expressed in terms of social class, gender, and ethnicity. Medical geographers have paid particular attention to the spatial patterns of health inequities and to explanations that give some prominence to place. Many maladies appear to occur consistently more often among the poor than among the affluent. There seems to be a socio pathological complex made up mainly in stress, lifestyle, diet, housing, pollution, and aging infrastructure. Some diseases, such as tuberculosis and alcoholism, are tradi tionally associated with low income populations. There are numerous correlations between health problems, measures of poverty, and place in the literature. A wide range of diseases and disorders appear to be associated with unemployment, low income, limited education, and inadequate diet. Also found in the mix are variables such as sanitation, overcrowding, trash accumulation, traffic, accidents, industrial hazards, and pollution. Common symptoms and behaviors include stress, alienation, in security, depression, drug and alcohol abuse, smoking, and spousal and child abuse.

There is a growing body of research that appears to confirm that a lack of 'social capital', broadly conceived, contributes to poor health and even early death. Recent quantitative and qualitative studies in several Western nations demonstrate that observed and self reported ill health is associated not only with the traditional stressors of poor diet, smoking, and low income, but also with broken ties to family, friends, and familiar social organ izations. In New York City and elsewhere, when low income, powerless citizens are displaced from familiar communities, their social networks are shattered, and rudimentary public services are undependable or non existent. Crime, substance abuse, violent death, and the incidence of diseases like HIV and tuberculosis tend to rise precipitously.

Many researchers argue that people experience poor health because they are more likely to participate in risky lifestyles or behaviors. Since these behaviors are socio geographically patterned – lower socioeconomic status groups are more likely to engage in unhealthy behaviors – such habits are likely to be key determinants of spatial health inequities. A major health risk associated with modernized environments is a tendency for increased adiposity (amount of body fat). Excess consumption of fats, salt, and sugar are markers for heart disease, and low income populations in the United States and other Western nations are prone to consume less healthy diets. Smoking is well established as a risk factor for much chronic diseases, especially lung cancer and heart disease. Poor dental hygiene is a marker for oral diseases and related disorders. Obesity, ill health behaviors, and dia betes are associated with people with common cultural backgrounds and spatial distributions.

Disability research has emerged in medical geography as a socially or culturally constructed form of exclusion – for instance, activist movements demanding signs in Braille for the blind or access ramps for people confined to wheelchairs. However, disabilities and impairments are more nuanced, as explained by Wilton and Evans. They may be examined from the viewpoints of several different conceptual models: medical, social, human rights, and cultural feminist. Geographers describe the active role of the sociospatial environment in determining what con ditions or impairments are experienced as 'disabling'. Disability may be defined in terms of physical impair ments, chronic illness, sensory impairment, and intel lectual/learning conditions.

Geographers became interested in mental health late in the 1970s, mainly as a result of the worldwide move ment to close asylums and create community mental health centers and 'halfway houses' for former hospital patients. A few early influential studies came out of the United Kingdom (an analysis of the distribution of schizophrenia) and the United States (two classic eco logical studies of distribution and correlates of mental ill health and facility location). Christopher Smith found a spatial dimension and landscape expression to com munity mental health. Landmarks and the layout or spatial design of a community provided orientation and facilitated a sense of identity. Facilities such as com munity centers, places for parties, churches, and libraries promoted rootedness (social capital) and provided means to transcend the daily conditions of life. More recently, studies have turned to investigations of community atti tudes toward mental health facilities and patients, and the conflict between centers and neighborhood residents – the now well known NIMBY (not in my back yard) syndrome. Also appearing in the literature are studies about social support and contacts for the homeless, im migrants, and individuals of different classes, ethnicity, and gender. All have expression in space and place.

Closely associated with mental health studies are in vestigations of the spatial distribution of substance abuse (drug and alcohol addition). In early 2006, and again in 2009, the Association of American Geographers col laborated with the US National Institute on Drug Abuse to sponsor a multidisciplinary symposium. Themes ad dressed included:

- spatial patterns of drug use and addiction,
- linking spatial models with neuroscience and genetics in drug abuse research,

- locational analyses of drug addiction treatment and service delivery facilities, and
- use of geographic information systems (GIS) to better understand and respond to drug addiction.

A publication of these symposium papers appears in the 'Further reading' section.

HIV/AIDS, despite its communicable quality, is in cluded in this section because its spread is a function of human behavior and lifestyle – not transmitted by vec tors, and not confined to any particular ecological biome. No other pandemic in recent times has had the profound impact on people and institutions as has human im munodeficiency virus (HIV) and the acquired immuno deficiency syndrome (AIDS). The attitude among many people in the developed world that infectious diseases were to a great extent a problem of the past was forcibly changed in the 1980s as AIDS began to diffuse world wide. The earliest cases of AIDS in the Western world were found among gay men, particularly in major urban centers in the United States and Europe. The disease eventually began to be transmitted to women and their children. The disease is spread by contact with con taminated blood products and unsterilized needles used by drug abusers. Geographers have published a number of maps illustrating the correlates to HIV/AIDS (see **Figure 2**). Much work remains for all disciplines and organizations to provide sufficient context and resulting insight into this most recent plague.

A substantial geographic literature exists on the sub ject of environment and health. The 1962 publication *Silent Spring* rippled throughout the public and scientific communities and initiated a growing wave of research into the linkages between environment and human health. Despite this mass of research, there is still limited hard scientific proof that some illnesses are caused by the contamination that human activities add to the environ ment. Most scientists are still concerned about emerging lifestyle diseases and are committed to genetic research as the next magic bullet, while the professed associations between human health and environmental exposures remain highly contentious.

Despite the scientific uncertainties in the linkages between environmental exposures and health outcomes, some public health studies suggest that such exposures are important mediating factors explaining the relation ship between low socioeconomic status and ill health and premature mortality. Millions die each year from environment related conditions. The most prominent threats lurk in places that should be safe – home, school, and neighborhood. The evidence is persuasive that modernization leads to increases in serious chronic dis ease rates, with these increases related to the increased pollution, general stress, and adiposity that comes with modernized life.

Healthcare Systems, Distribution, and Access to Health Resources

When we speak of healthcare 'systems', it is important to differentiate between the way healthcare is defined and the way in which that care is organized and financed. Basically, healthcare may be differentiated on the basis of where it fits on a model scale ranging from 'biomedical' to 'holistic'. At the biomedical end, we find practitioners applying the principles of the natural sciences, especially biology and physiology, to clinical medicine. For the most part, this involves diagnosis and curative treatment of the patient's illness or disorder. In the holistic range, we find practitioners and patients in a cooperative relationship, leading to some optimal attainment of physical as well as mental health and well being. It may or may not involve drug and/or surgical treatment, but in large part, it places prevention on an equal footing with curative treatment. It also focuses on education and responsibility among patients to achieve health balance and well being. Care and caregiving occur in a variety of settings, including hospitals, physician offices, clinics, workplaces, schools, homes, and village settings. Some researchers observe the disciplinary and health policy changes that underlie care and caregiving. Geography tends to focus on the places and spaces of care and how these are used and understood. An examination of the experiences of patients and prac titioners alike are part of the study of medical care.

Professional health systems originated in the ancient Middle Eastern world. Ayurvedic medicine, for instance, can be traced back to the migration of Aryans into the Indus Valley around 2000 BC. Today, we refer to Ayur vedic and comparable medical practices as 'traditional', and biomedicine as 'modern'. Biomedicine itself arose out of ancient Greek medical practices and advanced rapidly with the development of the scientific method. Nine teenth century advances in bacteriology, along with the advent of germ theory, elevated the status of biomedicine. Several so called traditional systems still exist in various forms and under a range of names in both the developing and the developed world. We refer to the study of these systems as 'ethnomedicine', but the more inclusive term 'alternative medicine' is widely used today, and usually refers to therapeutic practices that are not part of the conventional biomedical school curriculum. Alternative therapies include, but are not limited to, disciplines such as folk medicine, herbal medicine, homeopathy, faith healing, new age healing, chiropractic, acupuncture, and naturopathy.

Healthcare delivery involves or intersects with a large part of human culture and the environment. Healthcare geography is concerned with: (1) how places differ in terms of the needs and demands for healthcare, (2) cul tural, political, and economic circumstances that affect how much care is actually needed and used, (3) inequities

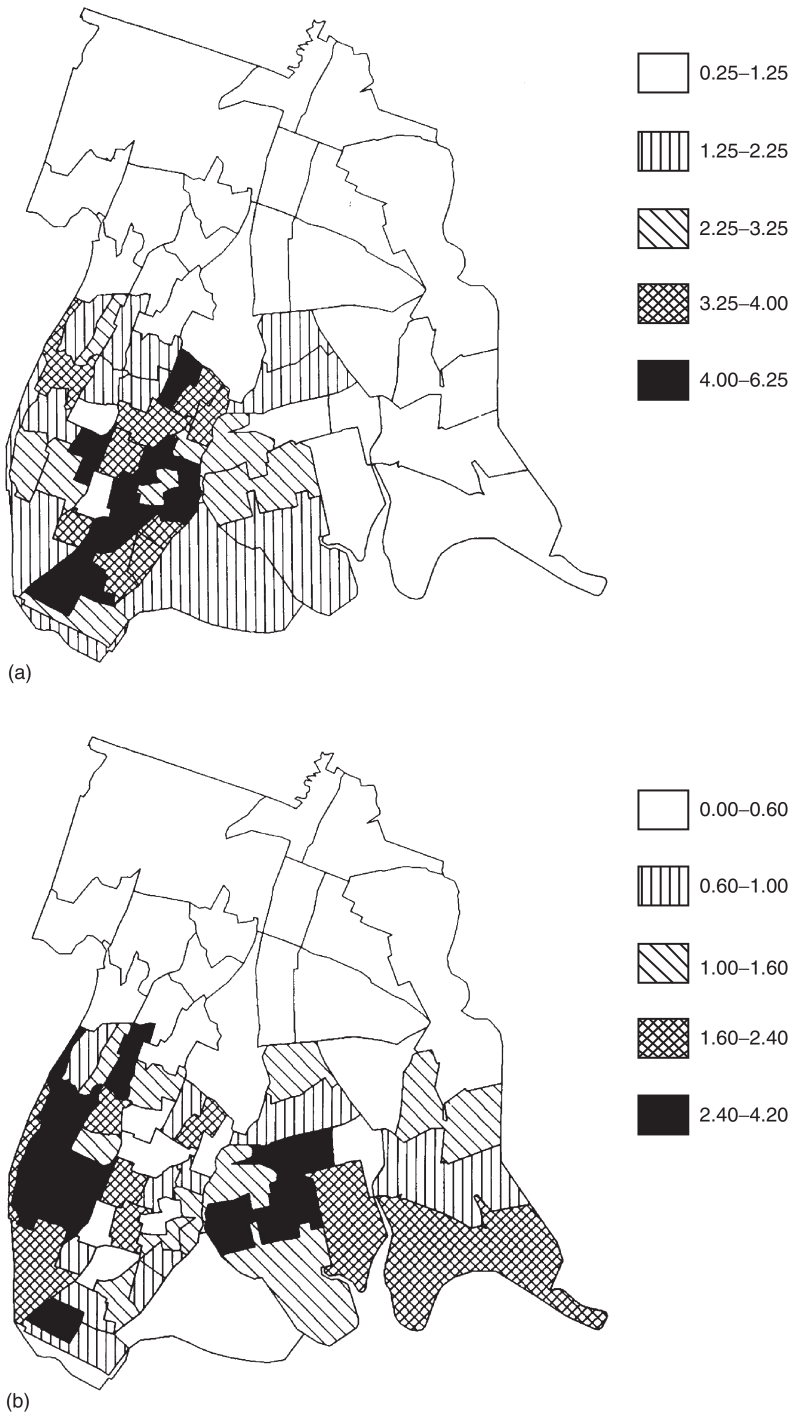

Figure 2 The Bronx, New York: (a) annual number of drug-related deaths, by quintiles; (b) HIV deaths per 100 000 population, cumulative to 1988. Adapted from Wallace, R. and Fullilove, M. T. (1991). AIDS deaths in the Bronx 1983 88: Spatiotemporal analysis from a sociogeographic perspective. *Environment and Planning A* 23, 1701 1723.

in the provision of care, and (4) methods and models that address spatial inequities and offer spatial solutions. A more recent interest in place involves the features of place that make them beneficial to health. These features can refer to environmental quality, such as clean air, mineral water, or magnificent topography. Or, they may rest on the qualities of structures, such as cathedrals and temples. Such places gain wide reputations for providing a perceived improvement in one's well being – a socially constructed characteristic.

Medical geography has paid attention to the socio political aspects of healthcare. The historical accounts of national medical systems are punctuated with their legal and policy characteristics. Gatrell, for example, contrasts the collectivist ideology of Britain's National Health Service to the anti collectivist ideology of the US medical system, and Christopher Smith writes comprehensively about the privatization of the Chinese medical system that replaced a collectivist medical organization when that nation modernized its political and economic systems. We are reminded almost daily that medical care practitioners and facilities are not always successful in healing patients – antibiotics, which are prescribed to cure, can be dangerous in certain circumstances; multinational pharmaceutical conglomerates appear to foster public dependence on their products, which are sometimes made available without adequate testing; and hospitals have been known to produce virulent infections in otherwise healthy patients. These are examples of 'iatrogenesis'.

Medical geography's interest in health services restructuring illustrates the multidisciplinary approach of health/medical care delivery. The various branches of public health, especially health administration and education, are involved in the study of medical care, and the contributions from the social and behavioral sciences (anthropology, sociology, economics, history, political science, and psychology) are obvious and well known. Restructuring is usually a response to some perceived flaw in the present delivery of care. Market forces that affect the delivery of health/medical care are significantly changing the financial base and functional role of public health departments, hospitals, and other medical service organizations.

The concept of inequality reverberates through the health services literature. Meade and Earickson, Curtis, and others discuss at great length the equality of distribution within predefined geographic spaces and inequity. Healthcare resource distribution has been studied at scales ranging from the nation to the neighborhood. Tools such as the location quotient, Gini index, and physician population ratio enable us to measure crude inequities at most scales. A recent statistical approach called multilevel modeling now enables simultaneous statistical comparison of variables between different geographic scales. The question of what is equitable in the availability of healthcare cannot be answered in simplistic terms because, among other things, there is considerable variation in demand for service and need for service, and variation in access to service ranges from mass exclusion to universal provision, depending on what kind of care is at issue, and the quality of care being provided.

Medical geographic research includes an impressive array of studies that address patient access to health/medical care. Accessibility refers to the extent to which people can easily obtain health services when needed. It may be examined from the economic or political viewpoint, which usually highlights healthcare equity, or from the geographic viewpoint, which deals with the location of service facilities. Facilities are conventionally separated into different levels of healthcare provision, usually into three categories: (1) primary care, provided in the home, a physician office, clinic, or health center; (2) secondary care, provided in a hospital setting, where patients are admitted for treatment that goes beyond what is available in a primary care setting; and (3) tertiary care, in which specialized care is dispensed that may not be available in either categories 1, or 2. Examples include facilities that specialize in diseases like cancer or Alzheimer's. Medical/healthcare accessibility varies enormously between developed and developing countries.

Underlying the problems of inequitable or inadequate health service provision in modern society means going beyond the economics and politics of medical resources, need, and demand. Health and society are interconnected, and the basic determinants of health have far more to do with social class, family and peer values, education, risk taking, and environmental hazards than with access to practitioners and their technology. Health is governed by ubiquitous forces we may not be prepared to change: food production and eating patterns; social uses of alcohol, tobacco, and other substances; dependence on automobiles; sedentary lifestyle; and so on. Thus, modern biomedical care may be virtually irrelevant to how we should deliver adequate and equitable care at a cost that society can reasonably bear.

Future Directions and Issues in Medical/Health Geography

Medical geography is a dynamic field of study. From the early narrow positivist human ecological construction of illness and treatment to the latest alternative theoretical and methodological approaches, it is clear that geographers in this specialty must remain involved in such up to date issues as how medical systems and practice change over time and space and how the integration of traditional and modern medicine might be achieved to improve health for all. Medical geography will increasingly continue to overlap with epidemiology and other social and behavioral science disciplines' interests in health. GIS, for instance, has already become a valuable tool in health and healthcare research, regardless of its disciplinary home. In another example, a book of readings on globalization and health, edited by ecological anthropologist Greg Guest, offers important insights to medical and health geographers – insights that were virtually unheard of a half century ago.

In its infancy, medical geography tended to focus heavily on infectious disease. With developments in medicine and biology, that focus diffused into lifestyle diseases and disorders. More recently, studies of degenerative diseases and unhealthy behavior have surfaced. In the future, we will see increased attention being paid to the health effects of globalization. As Guest, and others point out, natural resources are being exploited at an alarming and unsustainable rate. At the same time, chemical waste products from energy production and manufacturing are altering biological ecosystems worldwide. Societies worldwide are feeling the effects of globalization, where events and actions in one world region can no longer be considered in isolation. People in North America share air and water through global circulation patterns with people in Europe and Asia. Changes in atmospheric chemistry and planetary physics are affecting biological and social systems, with the combined changes having direct and indirect impacts on human health and well being. Since the latter half of the twentieth century, there is an inexorable aging of the world's population, particularly in developed nations. Echoing the writings of postmodern health geographers, which reflects the interests of cultural geography and draws on social theory, we note an emerging literature that articulates the dynamics between older people and the places in which they live and are cared for. The health geography of the future will increasingly investigate these systems, and the reactive behaviors of people and societies worldwide.

Until the 1980s, quantitative studies of disease patterns tended to focus almost exclusively on aggregate spatial patterns. Studies of infectious disease, exemplified by the work of May and Hunter, come readily to mind. More recently, attention has switched to quantitative analyses that are based on individual events, and the spatial arrangement of point patterns that represent collections of individuals with disease. This methodological approach, however, pays little attention to what the points on the map really represent. Map symbols are not inanimate objects; they represent real people. Spatial arrangements of people and their ecosystems have shed considerable light on disease causation, but positivist approaches that made up so much of medical geography in the past give scant attention to the feelings, experiences, beliefs, and attitudes that individual humans have. The medical/health geography of the future will feature human expressions of their health status and needs (through qualitative methodologies) rather than mere collections of map symbols and statistical analyses that accompany them. The same may be said about research involving healthcare and health facilities.

Finally, health geographers should become more involved with political activism on behalf of planet Earth and the health of its citizenry. For example, the dominance of the petroleum industry worldwide has created a crisis within the US economy and adversely affected the financial sector and industrial and agricultural sectors in most nations. Recreating a political economic system that promotes development that sustains human health will require creating new rules, alternative sources of wealth, and transformed institutions. The current international economic system arguably does not support the sustainable use of global common resources, and health for all. The current political framework favors transnational corporations and current trade agreements, at the expense of environmental stability and global public health. Taking a stand for the latter may entail career risks, but the weight of health geographers' considerable collective education can benefit citizens worldwide by helping to achieve a goal of better health for all in a sustainable world.

See also: Ageing and Health; Care/Caregiving; Chronic Disease; Communicable Diseases, Globalization of; Complementary and Alternative Medicine; Disability and Chronic Illness; Disease Diffusion; Environmental Health; Epidemiological Transition; Gender and Health; Health and Development; Health Geography; Health Inequalities; Health Services Restructuring; Health Systems and Health Services; Healthcare Accessibility; HIV/AIDS in Developed Countries; Housing, Neighborhoods and Health; Indigenous Health and Medicine; Mental Health; Social Capital, Place and Health; Therapeutic Landscapes; Voluntary Sector; Wellbeing.

Further Reading

Barrett, F. A. (2000). *Geographical Monographs, No. 23: Disease and Geography: The History of an Idea*. Toronto, ON: York University.

Cliff, A. and Haggett, P. (1984). Island epidemics. *Scientific American* 250(5), 138 147.

Curtis, S. (2004). *Health and Inequality: Geographical Perspectives*. London: Sage Publications.

Dyck, I. (1998). Women with disabilities and everyday geographies: Home space and the contested body. In Kearns, R. A. & Gesler, W. M. (eds.) *Putting Health into Place: Landscape, Identity and Well Being*, pp 102 109. Syracuse, NY: University of Syracuse Press.

Eyles, J. and Consitt, N. (2004). What's at risk? Environmental influences on human health. *Environment* 46(8), 24 39.

Garrett, L. (2000). *Betrayal of Trust: The Collapse of Global Public Health*. New York: Hyperion.

Gatrell, A. C. (2002). *Geographies of Health: An Introduction*. Oxford: Blackwell Publishing.

Gould, P. and Wallace, R. (1994). Spatial structures and scientific paradoxes in the AIDS pandemic. *Geografiska Annaler* 76B, 105 116.

Guest, G. (ed.) (2005). *Globalization, Health and the Environment: An Integrated Perspective*. New York: Rowman and Littlefield.

Hughes, C. C. and Hunter, J. M. (1970). Disease and development in Africa. *Social Science and Medicine* 3, 443 493.

May, J. M. and McLellan, D. L. (1974). *The Ecology of Malnutrition in Western South America*. New York: Hafner.

Mayer, J. (1996). The political ecology of disease as a new focus for medical geography. *Progress in Human Geography* 20, 441 456.

McLafferty, S. L. (2003). GIS and health care. *Annual Review of Public Health* 24, 25 42.

Meade, M. S. and Earickson, R. (2000). *Medical Geography* (2nd edn.). New York: Guilford Press.
Moon, G., Subramanian, S. V., Jones, K., Duncan, C. and Twigg, L. (2005). Area based studies and the evaluation of multilevel influences on health outcomes. In Bowling, A. & Ebrahim, S. (eds.) *Handbook of Health Research Methods*, pp 266 292. Berkshire: Open University Press.
National Research Council (2001). *Under the Weather: Climate, Ecosystems, and Infectious Disease*. Washington, DC: National Academy Press.
Smith, C. (1998). Modernization and health care in contemporary China. *Health and Place* 4, 125 139.
Stock, R. (1986). 'Disease and development' or 'the underdevelopment of health': A critical review of geographical perspectives on African health problems. *Social Science and Medicine* 23, 689 700.
Wallace, R. and Fullilove, M. T. (1991). AIDS deaths in the Bronx 1983 88: Spatiotemporal analysis from a sociogeographic perspective. *Environment and Planning A* 23, 1701 1723.
Wallace, R., Fullilove, M. T., Fullilove, R., Gould, P. and Wallace, D. (1994). Will AIDS be contained within U.S. minority urban populations? *Social Science and Medicine* 39, 1051 1062.
Whiteis, D. G. (1998). Thirdworld medicine in first world cities: Capital accumulation, uneven development and public health. *Social Science and Medicine* 47, 795 808.

Relevant Websites

http://www.userpages.umbc.edu/~earickso/Index.html
Association of American Geographers, Health Geography Specialty Group.
http://www.associations.eas.ualberta.ca/caghealth/index.html
Canadian Association of Geographers, Geography of Health and Health Care Specialty Group.
http://www.isc.temple.edu/neighbor/service/
Disability Specialty Group of the Association of American Geographers.
http://www.esri.com/index.html
ESRI GIS and Mapping Software.
http://www.ca.dir.yahoo.com/Science/Geography/Human Geo graphy/Medical Geography/Health Atlases/
Medical Geography Health Atlases.
http://www.en.wikipedia.org/wiki/Medical Geography
Health Geography from Wikipedia.
http://www.rgs.org/OurWork/Research+and+Higher+Education/ResearchGroups/Research+Groups+A+ +M/Geo graphy+of+Health.htm
Royal Geographical Society Geography of Health Research Group.
http://www.eoearth.org/by/topic/environmental health
The encyclopedia of earth: Environmental health.

Medieval Geography

K. D. Lilley, Queen's University Belfast, Belfast, UK

Geographers have increasingly neglected their subject's medieval history, and so much of the material used in this article derives from beyond geography, from literary historians, historians of science and art, and philologists. Indeed, in these areas, 'medieval geography' has seen something of a renaissance, as evidenced both by recent editions of medieval geographical texts as well as critical studies of medieval geography and mapping. What follows draws upon what both geographers and non geographers have written on geography's medieval history. The first half of the discussion is a historio graphical exercise which is long overdue and serves to reveal how important medieval geography was to those geographers shaping the discipline's identity during the formative stages of its development as an academic sub ject in the early twentieth century. Following this, the second area of discussion is on geographical thought 'in' the Middle Ages, particularly: (1) the place of geography in medieval learning, (2) the sources and influences of geographical ideas in the Latin West, and (3) geo graphical works of the Middle Ages and their visual and textual representation of terrestrial space. While the temporal and geographical focus is on Europe from the fifth through to the fifteenth centuries CE, clearly within this context geography and geographical knowledge of the Latin West also touches upon the Islamic, Byzantine, and Judaic worlds of this period.

A Modern Historiography of Medieval Geography

In the UK, historical geographers who have worked on the medieval period have by and large taken an approach which looks at geographies of settlements, landscapes, and peoples with the aim of using historical evidence to reconstruct these past geographies. On the whole then, the main characteristic of more recent work by human geographers has been on geographies 'of' the Middle Ages rather than geographies 'in' the Middle Ages, whereas before World War II, Anglophone geographers whose interest was in the medieval past were concerned more with geography 'in' the Middle Ages. This section focuses on the latter.

The early twentieth century saw the publication of three lengthy studies on geography in the Middle Ages. The first of the three studies appeared at the turn of the twentieth century in three separate volumes by Raymond Beazley, then lecturer in the history of geography and historical geography at the University of Oxford. Beaz ley's volumes went under the title *The Dawn of Modern Geography*, itself telling, for clearly he saw geography's medieval past as an integral part of 'modern geography', marking its 'dawn', its beginnings ('modern history' at Oxford also takes in the Middle Ages). The three vol umes that Beazley wrote cover a wide period, from the early to the late Middle Ages, and his approach is to combine what would now be described as geographical knowledges of the period, as well as geography *per se*, encapsulated in the subtitle he chose, a 'History of medieval exploration and geographical science'. He saw them as "an attempt to open up comparatively fresh fields of historical and geographical inquiry" (1906: iii). At the time they were relatively well received. In the *American Historical Review*, Stevenson considered Beazley's third volume, which dealt with the period 1260–1420, to be a "truly excellent work." Beazley himself saw "the geo graphical progress of the Middle Ages" as "essentially connected with the extension of Europe and Christen dom into its present dominion over the best and largest part of the earth," (1897: 1) and hence a justification of European and Christian dominance in his own imperial age. There were Victorian resonances in Beazley's ac count of "the medieval expansion of Europe," in "the wonderful diffusion of Christianity through missionary travel," and the "part taken by pilgrimage in explor ation," for example, though he seems unaware of this as he remarks "we shall have especially to notice many ideas and circumstances somewhat strange to us of the present day."

The three volumes of the *Dawn of Modern Geography* are not easily summarized in a few lines, but in essence they deal with 'practical exploration' and 'geographical study' within the period of the Middle Ages, each volume dedicated to a particular era (CE 300–900, 900–1260, and 1260–1420). On the whole, his concern is with the former, with descriptions of travel and pilgrimage, exploration and Crusading, and commerce and missionaries, while on latter he notes that "there is in general during this time a lack of geographical enterprise or study for the sake of knowledge," (1897: 2), and tends to summarize it under what he calls 'geographical theory' for each of the three main periods he deals with, with some additional chap ters on 'non Christian geography', principally that of Islamic scholars. In this 'geographical theory', Beazley describes the geographical content of works by the likes of neoplatonists such as Macrobius and Martianus

Capella, whom he calls 'geographers', noting their debts to 'classical knowledge' as well as Christian thinking, two hallmarks of early medieval science in general. This geographical theory covers geographical knowledge of the time, as understood by Beazley, gleaned by him from the early medieval accounts of what he terms 'patristic geography' and those later 'scholastic' and 'schoolmen' geographers, which included descriptions of the Earth, such as its shape and size, its content and character, both in textual accounts of 'Earth knowledge', in cosmography and natural philosophy, and in their associated imagery, in the form of maps. Under the heading of 'Geographical theory' Beazley thus combined both medieval geography *per se*, as it was represented in contemporary texts, as well as geographical knowledge more broadly defined. Beazley's efforts were judged 'most timely' in Stevenson's review of the work, and "without doubt the best that has yet appeared on the subject." Beazley himself recognized the novelty of what he was doing when he admitted "that no attempt has yet been made to deal with this subject as a whole" (Beazley, 1897: ix). Certainly his efforts provided a basis for the two subsequent studies of medieval geography that appeared a few years later.

In his introduction to a later edition of J. K. Wright's *Geographical Lore at the Time of the Crusades*, Clarence Glacken notes that "the emphasis of Beazley's *Dawn of Medieval Geography* was 'not on theory or concepts but on the history of medieval travel and exploration'" (Glacken, 1965: viii). He makes a case that Wright's book "has a secure place in the history of geography and in the history of ideas" because "it is distinct from a history of exploration and a history of science, even though it possesses some characteristics of both" (Glacken, 1965: ix). By 1965, Wright was a well established figure of American geography, having been elected president of the Association of American Geographers in 1946 and whose ideas on "the study of the world as people conceive of and imagine it" – what he called 'geosophy' – were influential in what would become 'humanistic geography'. Wright saw Beazley's three volume study as "the most extensive and satisfactory work on medieval geography as a whole," (Wright, 1925: 507) and yet in his preface notes "to the more theoretical aspects of geographical knowledge in this age he gives but meagre space." When Wright wrote *Geographical Lore* he was doing so under the close influence of his dissertation supervisor, Charles Homer Haskins, the notable historian of medieval science, and it is, therefore, perhaps of no surprise that Wright's focus is on the 'time of the Crusades', the twelfth and thirteenth centuries in particular, which also formed Haskins' chief interest, as exemplified in his *Renaissance of the Twelfth Century* and *Studies in the History of Mediaeval Science*. Wright's focus was different to Beazley's not only in terms of his more restricted temporal coverage but also in the subject matter he aimed to cover, "to illustrate and trace the origins of the most characteristic geographical ideas current in Western Europe at the height of the Middle Ages" (Wright, 1965: xvii). His approach in so doing is again broad, for "the history of geography" he notes "leads its students into many fields."

The title Wright chose for his book was *Geographical Lore*, not knowledge, or theory, or science, which he recognized required some explanation. He saw it offering wider scope 'than most definitions of geography', covering "what was known, believed and felt about the origins, present condition, and distribution of the geographical elements of the earth," such as "theories of the creation of the earth, of its size, shape and movements," as well as "the various physiographic features," and "theories of the regions of the earth's surface" (Wright, 1925: 1–2). For this reason, he explains, "the geographical lore of the Middle Ages involves a wider range of subjects in space, as well as in time, than is now included in geography," a mixture of "cosmology, astronomy, astrology [and] theology" (Wright, 1925: 2). Again the names and works he cites are familiar ones, those discussed earlier by Beazley, those from the 'ancient world', from Western Christendom, and 'the Moslems' in the early Middle Ages, who are seen to have contributed 'geographical ideas' circulating later in the twelfth and thirteenth centuries. In the later parts of the book, Wright then takes a novel approach. Rather than "to try to arrange the material chronologically," he adopts "a topical and regional arrangement," and covers cosmogony and cosmology, the atmosphere, the waters, the lands, 'astronomical geography', cartography, and regional geography, as if trying to replicate how such subjects were treated at the time in works such as those he cites: borrowing therefore a "medieval scheme of learning" (Wright, 1925: 4). His emphasis though is on knitting together various sources to 'illustrate' geographical lore, and so create his own medieval geography.

Wright's medieval geography, then, is a description of the Earth as seen through the eyes of medieval authors, rather than an analysis of the 'idea' of geography that was present in works of the time. Hence, *Geographical Lore* is more concerned with what the state of geographical knowledge was in the Latin West at the time of the Crusades, and the sources of that knowledge and what it encompassed, and less about how those who wrote their accounts perceived geography to be. His justification for this, he says, is that "geography in the Middle Ages did not form a distinct and separate science," and that "the student who learned anything of geography learned it incidentally to the study of other subjects," tangentially as it were, for "even the word 'geography' was scarcely ever used" (Wright, 1925: 127). This incidental geography, gained by medieval students in their learning of various aspects of philosophy and theology, is how Wright presents his own medieval geography to modern students. This 'classification of knowledge' approach would have been familiar to medieval students, and his use of it is a sign perhaps of his own sensitivity to the subject matter and a reflection of his particular geosophy of the Middle Ages.

The last of the three principal books of the early twentieth century on medieval geography is Kimble's *Geography of the Middle Ages*, a work that tries to compress into one single volume the wide temporal range of Beazley's three volumes, and Wright's detailed subject coverage – an ambitious project. At the time reviewers saw it in this light, as a "little volume" that "does not break much new ground" but which "should be of par ticular value to those who wish to obtain a comprehen sive yet concise and well balanced concept of medieval geography as a whole." What this 'medieval geography' comprised for Kimble is not explicitly discussed, except in his preface where he states that his book: "essays to follow the development of earth knowledge as it was affected, on the one hand, by the prevailing ideologies, and on the other by the results of human enterprise," by which he means the influences of classical, Christian, and Islamic sources of 'geography', and the 'expanding hori zons' of the medieval West through travel to and within Africa and Asia.

Kimble's aim was to contribute to "the history of geographical thought," (Kimble, 1938: v) as "the contri bution of the Middle Ages to the advancement of geo graphical studies has still to be adequately appreciated." What his treatment amounts to is a description of 'world knowledge' in the Middle Ages, as told through con temporary sources, including again the familiar neo platonic canons of the early Middle Ages cited by both Wright and Beazley before him, together with extracts from later works, what he calls 'geographical studies', meaning works such as John of Sacrobosco's *Treatise on the Sphere* and the *Speculum Naturale* of Vincent of Beauvais, in which accounts of the Earth and its size and origins are given. Kimble's 'topical' rather than chronological struc ture troubled Lynn Thorndike (an eminent historian of science and magic) in his review of the work, but not as much as Kimble's use of sources and "general back ground of intellectual history against which he sets his account of geographical thought and knowledge." Thorndike seems to be saying Kimble is out of his depth, for his book "has nothing to offer the reader which may not be found elsewhere" and had "not done full justice to the intellectual life of the Middle Ages." In his view, Kimble's *Geography of the Middle Ages* was found wanting. Perhaps this explains his subsequent move away from medieval geography, toward a newer emergent 'regional geography', perhaps fostered also by his physical move from geography at the University of Reading, where he had prepared his book, to a new academic position at McGill University in Canada where he was appointed their first professor of geography, in 1944, having re ceived favorable support from Carl Sauer.

Kimble's move away from the medieval period was matched also by both Wright and Beazley, neither of whom pursued the subject of medieval geographies further into their careers. Beazley became professor of history at the University of Birmingham in 1909 and subsequently worked on the nineteenth century Euro pean history, while Wright's interests moved more toward his pursuit of 'geosophy'. Curiously, this shift by medieval historical geographers, from an interest early in their careers in the Middle Ages to more recent periods, and to more conceptual matters, is a recurring feature of the post war period too. The pre war studies of medieval geography by Beazley, Wright, and Kimble have not carried through, however, into their successors' work on the Middle Ages. Instead, medieval geography in the pre war sense faded from the discipline as those who had championed its cause moved on to other things in the pursuit of their academic careers. Looking across their work, there are characteristics in this pre war medieval geography that are broadly similar, and this perhaps in part accounts for its passing; a repetition of a story already told, perhaps told better by those doing more detailed work elsewhere in the history of science and religion (including Haskins and Thorndike), and seemingly made superfluous by the ostensibly all en compassing treatments of Wright and Beazley. Durand's review of Kimble's *Geography in the Middle Ages* concludes that "it is apparent that the specialist is occasionally in danger of finding himself [*sic*] in the position of a mere amateur," and "pending a reliable synthesis it is perhaps wise to give up the attempt to establish generalisations in these matters": enough of a disincentive to put off any young budding historian of geography who might be tempted to venture in the footsteps of Beazley, Wright, and Kimble into the Middle Ages.

Geography's medieval history was being written about in the early twentieth century, because it was seen to be very much an important part of the 'geographical tradi tion'. It was helping to form the discipline's own identity as geography was becoming a defined academic field, and in this sense it deserves to be more thoroughly studied in its own right by those with an interest in geography's recent history. The 'medieval geography' that was being presented is a mixture of what today might be called 'geographical knowledge' of the Middle Ages, whether from travel accounts or cosmographical works, as well as more narrowly defined 'geography *per se*', at least in dis cussions by Beazley on geographical theory and Wright on geographical lore, but always the two strands are mud died. Kimble's work was described by Durand as "an essay on the *Kulturgeschichte* of medieval geography," but the question of what geography meant to writers in the Middle Ages is not fully addressed. Wright's observation that the word 'geography' was scarcely ever used in the Middle Ages and is mentioned only in passing without consideration why this might be the case, or what its significance might be in trying to write of geography's medieval history and medieval geography.

It seems then that rather than presenting geography's medieval history, the three key studies on medieval geography of the pre war era are essentially histories of what geographical knowledge existed in the Middle Ages. While this interest has continued among historical geographers in the field of cartographic history, for example, through the work of David Woodward on medieval maps and mapping, it has fallen increasingly to nongeographers to write geography's medieval history, as geographers prefer to turn their attention to later periods. Indeed, this eliding of the Middle Ages is occurring in historical geography more generally. Philologists and historians, rather than geographers, are now picking up where Kimble, Wright, and Beazley left off in the pre World War II era. Perhaps one of the reasons why they have not been taken up and built upon since then is because they are seen to have already covered the ground sufficiently, though the fact that these works are rarely cited today, and their contents generally not mentioned in recent historiographies of geography, would suggest instead that they are simply forgotten contributions.

Geographical Thought in the Middle Ages

The modern historiography of medieval geography reveals the problems facing those writing about geographical thought in the Middle Ages. It is always going to be easier to begin geography's intellectual history in the modern era when the word 'geography' gained wider currency in written texts and circulation within and beyond Europe. What follows is an attempt to summarize the nature of geography in medieval learning, rather than range widely across what 'geographical knowledge' existed in the Middle Ages. The latter would result in the kind of exercise that Wright engaged in, with consideration given to what people in medieval Europe thought of mountain formation or how earthquakes occurred, for example, regardless whether such phenomena were then understood or recognized as belonging to 'geography'. Similarly with the medieval traveler and pilgrim accounts that Beazley wrote of. The alternative is to adopt a narrower perspective, to take a bookish approach that focuses more on the sources used in the Middle Ages to learn about the physical world, particularly its form and content. This is the aim here, to review briefly, through recent critical studies, those contemporary works that provided the Latin West with geographical thought. The connection that existed between scholarly accounts of the world and the geographical imaginations of the inhabitants of medieval Europe is an exercise that would reveal much about both, but it is beyond the scope of this article. Instead, what follows gives a flavor of geographical thought in the Middle Ages, highlighting first the issues of placing geography as an intellectual activity, second, identifying those sources that were especially influential during the Middle Ages as geographical accounts, and third, the stylistic characteristics of geographical works written and consumed by medieval Europe's intellectuals.

The Place of Geography in Medieval Learning

The word 'geography' is of Greek derivation, not Latin. Yet the language of learning in medieval Europe was predominantly Latin, so as a word it rarely appeared on the written page. To take a strictly etymological and textual approach to locate the subject of 'geography' in the Middle Ages could prove a long and disheartening search therefore, though not entirely fruitless. Lozovsky has identified cases in some Carolingian manuscripts and glosses where a Latinized term for 'geography' does appear. More commonly instead, the Latin phrase, *orbis descriptio*, the description of the Earth, is how contemporaries in the Latin West translated 'geography'. With this in mind and established, Lozovsky trawled the canonical works of encyclopedists and cosmographers of the early Middle Ages to tease out their conceptions of what *orbis descriptio* encompassed. She thus notes that sometimes, "in encyclopaedias and tracts titled *On the Nature of Things*, it constituted part of the knowledge about the created world, belonging to *physica*," while in other cases it was placed "in the quadrivium as part of geometry," and so "tracts specially devoted to geography, often without mentioning this directly, were perceived as treating part of the knowledge about the created world" (Lozovsky, 2000: 33).

This ambivalent placing of geography within medieval learning did not mean that there was a lack of coherence and continuity in *orbis descriptio* accounts. Far from it, for as Lozovsky notes, "though geographical knowledge lacked a definition or a disciplinary status and functioned in various contexts, it still possessed a certain self sufficiency as a subject; special 'geographical' texts testify to this" (Lozovsky, 2000: 3); and hence, she argues, "without trying to give a definition of geography in the early Middle Ages, we should follow the ideas implicit in geographical descriptions of the time, particularly in self sufficient ones, such as geographical tracts or chapters in encyclopaedias," in those texts that "include the description of the earth as a whole and of its various regions." The subject matter of this 'geography' was clear enough to those who wrote about it at the time, whether it featured in accounts of creation, in exegesis, in geometry, or in history. There was such a thing then as 'geography *per se*' in medieval thought, "sufficient enough to become a subject of special treatises" (Lozovsky, 2000: 28–34). If there was a common theme to this *orbis descriptio* of the early Middle Ages it was 'on the shape and size of the Earth', and 'place descriptions' (topography and chorography), recurring issues that required exposition and illustration often with the use of diagrams and maps.

Influences and Sources of Geographical Thought in the Latin West

Broadly, there were three main sources of influence on geographical thought in the Latin West, from the classical and late Antique world, from Biblical and Judaic traditions, and from Islamic scholarship. Throughout the Middle Ages these molded what was known and understood of the wider world and its geographical content and makeup, and also had a bearing on how geography was written about and represented visually in medieval sources. Of the classical sources, Ptolemy's *Geography* is usually singled out by geography's historians as one 'lost' to medieval scholarship, only to be 'rediscovered' later in the fifteenth century, gaining further in popularity by being printed in the sixteenth century. But thanks to copies made by Arabic scholars, aspects of Ptolemy's work were known in the medieval West, such as his system of the universe and his tables of coordinates, while in Constantinople, from 1300, his *Geography* was being studied in its own right. The Ptolemaic legacies evident in the Middle Ages exemplify the ongoing contact between cultures and traditions that placed the Latin West within wider circuits of geographical knowledge, fostered through certain centers of exchange and learning such as Toledo, Cordoba, Sicily, and Ravenna where Islamic, Christian, and Jewish texts were translated and reworked, providing a link between 'East' and 'West'. They provided a route for classical Greek texts on cosmography, geometry, and astronomy to enter the realm of medieval learning, including Platonic and Aristotelian works. Of course, Latin texts inherited from Roman sources were more readily available and thus more widely consumed around Christian Europe. Key influences, for example, were the Latin encyclopedic texts, such as Seneca's *Natural Questions* and Pliny's *Natural History*, the latter in particular including books on 'regional geography' and which (via a later derivative work by Solinus) were especially influential in the Latin West.

The Latin encyclopedic tradition formed the foundation of geographical thought through the early Middle Ages. Here neoplatonist accounts of the world were based upon Plato's *Timaeus* which was translated into Latin by Calcidius in the fourth century, and whose commentary and translation provided later scholars with an important source of Greek cosmology. As well as Calcidius, a Platonic conception of the world was also conveyed in other key Latin texts of the late Antique period, notably Macrobius' *Commentary on Cicero's Dream of Scipio* and Martianus Capella's *Marriage of Philology and Mercury*, both of whom were writing in the fifth century and describing the Earth and the heavens in ways that did not refute Christian doctrine. This same Latin encyclopedic tradition continued on into the early Middle Ages through the work of Boethius, Cassiodorus, and Isidore of Seville, each becoming key sources of geographical information and ideas, and providing a basis for later descriptions of the Earth (see below). In their treatment of geography there were continuities between the Latin encyclopedists because of their use of common earlier classical sources. For example, the influential section in Macrobius' *Commentary*, 'on the inhabited portions on Earth', derived from earlier neoplatonic works (particularly Posidonius and Porphyry), while Martianus Capella's discussion of 'geometry' was largely a treatment of geographical material borrowed from Solinus and Pliny, describing the Earth and the features of its habitable parts. Their work was in turn used as a basis for later, medieval geographical accounts. Thus, with its books on the seven liberal arts, the *Marriage of Philology and Mercury* "became the foundation of medieval education," (Stahl, 1962: 172) while an allegorical style made it appealing to later medieval neoplatonists. Additionally, "as an authority on geography," Stahl notes, "Macrobius was extremely influential in the Middle Ages" (1952, 20). For example, "the maps that accompanied manuscripts of his *Commentary* became models for one of the commonest types of medieval *mappae mundi*, the so called 'zone maps', " (Stahl, 1952: 200–216) while the views on the extent of the Earth's habitations that Boethius later set out in his *Consolation of Philosophy* repeated those of Macrobius (though he attributes them to Ptolemy). As their view of the world was repeated and embedded in geographical accounts, both in textual and visual forms, such geographical information derived from Martianus and Macrobius was continually reinforced. Thus, the form and content of the world as shown in later medieval 'world maps', such as the Hereford *mappamundi* with its depiction of the 'monstrous traits of more distant folk', was attributed to the influence of these encyclopedists and their borrowings from classical sources.

Latin Christian works of the late Antique period were also important and influential sources of geographical material in the Middle Ages, and given particular weight because of their patristic authority. Augustine, Orosius, and Isidore of Seville, who all were highly influential in early Christian theology, included in their work descriptions of the Earth, and, importantly, related these to their faith, so providing a doctrinal basis for studying the Earth and its content. Geography had particular theological resonance in Christian belief, not just because of the lands, places, and peoples mentioned in scripture and their relation to the life of Jesus, but because knowing about the world was also to know God. The Earth was His creation after all. This created physical world was understood to be like a book, and it was through the work of the natural philosophers who studied it that then allowed medieval Christians to see and understand these divine realities. Lozovsky, too, stresses the important exegetical function of geography for Christians in the early Middle Ages, that images of the Earth were an aid to contemplation, particularly on Biblical events associated with particular regions or localities, and also on the "the question whether Paradise should be understood in the literal sense (as a place on earth) or in the allegorical sense." Such matters concerned Augustine, for example, who wrote of it in his study of Genesis, and occupied Isidore in his *Etymologies*. Their discussions sought to establish the location of Paradise geographically, and thus geography had significance in dealing with fundamental facets of Christian doctrine. It also featured in historical accounts of the late Antique and early medieval period.

These same Christian authorities wrote accounts of the perceived turbulent times they were living in during the fifth, sixth, and seventh centuries, particularly the 'Sack of Rome' to the Goths in 410 and the threat that this was seen to pose to Christianity and Christians. It was the stimulus for Augustine's *City of God* and his arguments 'against the Pagans', as well as other works on the Goths by Cassiodorus, Jordanes, Isidore of Seville, and Orosius.

The 'histories' written by these Christian authorities were prefaced by geographical descriptions that drew upon classical sources of geography, such as Pliny, Ptolemy, Strabo, and Solinus. Of them, Orosius' *Seven Books of History against the Pagans* came first, written as a response to Augustine's *City of God*, and as Lozovsky remarks, provided "a long and fairly detailed geographical survey of the known world" which itself subsequently "became a school text and an authority for later geographical writings" by being copied separately (Lozovsky, 2000: 71, 78). Jordanes' *Gothic History* was written in the 550s with an account of Scandza, an island that was 'central to his story as the place of the Goth's origin', about which he drew upon Ptolemy and provided a geographical description that also used Orosius, and similarly Jordanes' *History* provided an important source of classical material for later authors. One of the more influential of these was Isidore of Seville's *Etymologies* (see below). Following in the encyclopedic tradition as already noted but at the same time deriving "his geographical information from Orosius, Pliny, and Solinus," Isidore's *Etymologies* was a "synthesis of Christian and classical knowledge" that "proved to be very influential in the Middle Ages" as Lozovsky notes, by being "widely copied and studied at schools" throughout Europe (Lozovsky, 2000: 107). These Christian historiographies of late Antiquity and the early Middle Ages were thus important not only for the way they positioned geography as key part of historical writing, but also for what knowledge they passed on 'as conduits' to later Christian authorities of the Middle Ages, such as Bede, from earlier and often lost geographies.

In the so called Dark Ages of European history, in the post Roman world, geography formed an important aspect of Christian thinking as authorities looked to earlier classical sources for information about the world around them. Lozovsky points out that this "formal geographical knowledge in the early Middle Ages followed its own principles, very different from those that people used in their everyday life" (Lozovsky, 2000: 153). It seems the reason why later Christian authors of the Middle Ages continued to draw upon classical sources, rather than looking to contemporary accounts of the world, was because of their perceived authority. The names of Ptolemy, Pliny, Solinus, Macrobius, and Martianus were as much a foundation of Christian learning and knowledge as Augustine, Orosius, and Jordanes (though concerning the formal hierarchy of learned authority in the Middle Ages, Augustine and Orosius would be regarded much higher that the first four, who were pagan writers). Through them a consensus of geographical information existed about the created world which led medieval Christians toward a better understanding of the Creator. This is the conclusion also reached by Merrills on the survival of earlier geographies through the historical writing of patristic authorities, though he also recognizes that "each of these geographical compositions was shaped indelibly by the historical works in which they appeared" (Merrills, 2005: 311), so that rather than simply ossifying earlier geographical thought these later sources reworked it to suit the circumstances of their authors.

The geographical sources that had particular influence throughout the Middle Ages were attractive to later scholars because of their Roman and Christian authority, but this does not mean that other ideas and sources were ignored. While Islamic texts on geography such as those by Al Mas'ūdī and Abū Zayd al Balkhī had no circulation in Latin Christian Europe, it was through Arabic contacts with Christian scholars that texts based on Ptolemy's *Geography* were made available later in the Middle Ages, especially from the twelfth century onward, as well as related studies of astronomy and geometry, again through Arabic versions of Greek texts. Though geography's historians have long recognized 'the contribution of the Moslems', particularly in making available Aristotle's works, the wider influence that Arabic sources had on geographical thought in Europe during the later Middle Ages is only now getting the serious attention it deserves, notably in the context of cartography. Sources of medieval geographical thought are due a careful reappraisal which has begun already for the period of the early Middle Ages, thanks to the efforts of Lozovsky and Merrills. While reverence was certainly given to the earlier authorities of Antiquity and the high Middle Ages, the period from the ninth to the fourteenth centuries was also one of great intellectual and scholarly enterprise which produced its own works on geography, but which only now are becoming the focus for critical analysis. With further such studies, it may soon be possible to examine more thoroughly the sources and influences of geographical thought in the Latin West, but until then the tendency will be to place emphasis on those canon ical geographies that were considered so authoritative in the Middle Ages.

Geographical Works and Their Medieval Circulation

In writing of geography's medieval history, there is a strong tendency among geographers and historians to separate text and maps. Yet, as Edson has pointed out, a more holistic approach is preferable, where medieval mapping and writing are seen as one, as indeed they were for those originally engaged in *orbis descriptio*. Geo graphical works in the Middle Ages were written with accompanying visual depictions so it seems artificial therefore to divorce geographical images from their texts. Sometimes these depictions appeared as marginalia, to illustrate a particular point, as Lozovsky shows in the case of one ninth century commentary on Martianus Capella's *Marriage of Philology and Mercury* where the manuscript has two figures drawn in the margins to in dicate the shape of the Earth. In other cases a depiction forms part of the textual description, as in another Martianus commentary of tenth or eleventh century date, again identified by Lozovsky, which has a 'map' of the world with the three continents of Asia, Europe, and Africa drawn to a T O configuration placed at the point where these are discussed in the text. Of course, some medieval world maps existed independently from (known) accompanying texts – the Hereford *mappamundi* being an example – though even they were understood as 'texts', albeit in visual graphic form. Whereas maps have been pulled out from texts (sometimes literally) and studied in their own right for the 'geography' they con tain, the meaningful place of these visual geographies is alongside the texts they were produced for, in context, as both were a means of describing terrestrial space in geographical works. The connections between visual and textual geographies then begin to reveal some interesting traits, as Lavezzo has recently demonstrated in her study of English imaginings of Britain's place at the edge of the wider world.

What follows here then is a brief consideration of the visual and textual content of geographical works, first looking at academic descriptions of the Earth that primarily belonged to scholarly study, and second ac counts of topography and chorography that drew upon such academic texts. Both reveal the integral nature of visual and textual geographies in the Middle Ages, a defining feature perhaps, while at the same time illus trating the symbiosis between geographical works which arose through the circulation of their normalized ac counts of the world.

Visualizing and Textualizing Terrestrial Space 1: Academic Descriptions of the Earth

Of the foundational works that provided scholars and intellectuals of the Middle Ages with descriptions of the Earth (*orbis descriptio*), it is those of the encyclopedic tradition that arguably saw the most use, influence, and circulation in the Latin West. Of them, on the basis of surviving manuscripts and copies, Macrobius and Mar tianus were undoubtedly widely consulted in centers of learning, in the cathedral schools and universities around Europe. In the same genre and also important was Isidore of Seville's *Etymologies*.

Isidore wrote the *Etymologies* between *c.* CE 615 and 630s, toward the end of his life while he was Bishop of Seville, and as its title suggests it is a compilation of many different subjects, an etymological explanation of them and particular terms brought together in 20 books cov ering in encyclopedic form "the accumulated learning of the classical world," including geographical material. With nearly a thousand manuscript copies of it surviving, the *Etymologies* was very widely circulated across medieval Europe and was soon known beyond Isidore's own Visi gothic Spain, in Gaul, and Ireland, for example. Lozovsky notes the influence of the *Etymologies* had not just as knowledge of the created Earth and its 'combination of Christian and classical pictures of the world', but in its treatment of this material. It therefore provides a useful indication of the nature and content of geography subject matter that was obtainable and being consumed during the Middle Ages.

Although the term 'geography' does not appear any where in the *Etymologies*, three of its books cover geo graphical topics, Books XIII on 'The cosmos and its parts', Book XIV on 'The earth and its parts', and Book XV on 'Buildings and fields'. These derive mainly from Isidore's use of Pliny, Martianus, Solinus, and Orosius, and form part of a larger group within the work that, based on their content, might be called *On the Nature of Things*. This is a common theme in the Middle Ages, in those encyclopedic works that likewise dealt with the physical world such as Bartholomaeus Anglicus' *The Properties of Things* (see below).

What did Isidore's geographical material include? 'On the cosmos and its parts' begins with a description of the world as a creation by God and proceeds to describe the sky, including meteorological phenomenon such as lightening, thunder, rainbows, and clouds, and the 'different kinds of water', such as the sea, rivers, lakes, and pools. Within his treatment and categorizing of these subjects there is a certain spatial logic. He starts with the character of the world as a whole, and its composition from atoms and elements, and then proceeds with what is above the Earth, in the sky and air, and then what is on the Earth, its waters. The terrestrial is further continued in the following book on 'The earth and its parts'. This begins by describing the Earth ('*de terra*'), "placed in the central region of the world, standing fast in the centre equidistant from all other parts of the sky," Isidore explaining that it is a 'globe' (*orbis*) because of the roundness of its shape, and that it 'is divided into three parts' made up of the three continents of Asia, Europe, and Africa, which he proceeds to examine consecutively. For each continent he gives an account of their lands, their rivers, and their cities, as well as the derivation of their names. Thus for Gaul, he notes it "is so called from the whiteness of its people," and that "the mountains and the chilliness of the sky keep the heat of the sun from this region, so that the whiteness of bodies does not darken in colour," hence giving readers not only a description of land and people but also an explanation. Having dealt with the three continents, Isidore turns to 'islands', where he begins with a description of Britain, "an island in the Ocean, cut off from the whole globe by the intervening sea," before considering other islands elsewhere, such as Cadiz and those of the Mediterranean, which he was surely more familiar with. After the islands come descriptions of features of the Earth, 'promontories', 'mountains and other terms for landforms', including volcanoes and mountain ranges, as well as 'lower regions', such as caves and 'fumeroles'.

Through his account of the Earth and its parts, Isidore works from the large scale down to the small. He starts with a description of the Earth as a whole, then its constituent parts, the three continents, followed by lesser geographical units, the islands. In Book XIII, having described the three continents, Isidore adds a short explanation of this hierarchy when he differentiates between the scale of the Earth, its lands, and the provinces as "parts of those, as Phygria is part of Asia, Raetia part of Gaul, and Baetica part of Spain." There is, then, a hierarchical treatment in his spatial order of things. This is also evident in Book XIV, 'On buildings and fields', which he starts with cities, then turns to 'famous towns', then types of buildings and then 'parts of buildings', moving also from the urban to the rural. Isidore's image of the world is thus an ordered and orderly one. Besides in the respective contents of Books XII, XIII, and XIV, this is evident through the sequencing of the three books themselves, moving from the cosmos, to the Earth, to its buildings and fields. The spatial order of the created world is translated by Isidore into a textual order on the written page. His books, like the Earth and its parts, thus follow a hierarchical placing of subjects. From the structure and content of the geographical material in the *Etymologies* we gain not only insight into what Isidore considered important and relevant information about the physical and human world around him, but also gain a sense of his own spatial thinking, evident in how he seeks to order subjects according to differing spatial scales and hierarchies, from large to small, from high to low.

Isidore's *imago mundi* was projected throughout the Latin West by being used in later textual geographies as well as visualized through *mappaemundi*. Isidore had himself intended to make use of a figure in Book III to illustrate 'the shape of the world', pointing out "the shape of the world is shown in this way." This he dealt with as part of astronomy, preempting his book on the cosmos and its parts that came later. Similarly, his brief treatment of geometry comes earlier, included under mathematics (also in Book III) where he relates it to 'measuring the Earth', from whence 'geometry takes its name'. Lozovsky notes this division of Isidore's geographical material and suggests, "as a result, there is no unified image of geographical space in Isidore's *Etymologies*." Yet it may be suggested while his treatment of Earth description forms a series of discrete parts, they add up to an overall unifying image of the physical and human world, one that was repeated in subsequent textual and visual descriptions of terrestrial space. It was copied 'word for word' by Hrabanus Maurus in the ninth century in his *De naturis rerum*, for example, and was generally popular among Irish and English monastic encyclopaedists of the eighth and ninth centuries. Later, too, Isidorian geography was drawn upon by the author of *De imagine mundi*, of around 1100, attributed to Honorius Augustodunensis, one later manuscript of which includes a 'world map with more than 200 place names and legends'. It also was the basis for the early thirteenth century *De proprietatibus rerum* by Bartholomaeus Anglicus, itself a popular source, drawn upon, for example, by Ranulf Higden in his *Polychronicon* of the mid fourteenth century. These textual works themselves often included in them *mappaemundi*, further repeating and reinforcing the textual geographies of Isidore, while medieval manuscript copies of Isidore's own work showed in map form the parts of the Earth he described in the *Etymologies*. As these interconnected textual and visual accounts of terrestrial space were circulated and recirculated across the Latin West, so too were the academic geographies compiled by Isidore and others like him, such as Martianus and Macrobius and their precursors.

Visualizing and Textualizing Terrestrial Space 2: Topography and Chorography

Ptolemy's 'spatial order' was differentiated into cosmography and geography, and 'chorography', which "pictured the form and character of localized spaces and places.” In the Middle Ages, it was ‘topography’ (*topographia*), as distinct from *orbis descriptio*, that referred to writing on specific localities and regions, as in the case of Gerald of Wales’ *History and Topography of Ireland*. Writing during the later twelfth century, Gerald had studied in Paris and traveled in England and Ireland on royal business for Henry II, but spent much of his time in Wales where he was born and where he served as a churchman, bringing him into regular contact with visitors and residents. Gerald also wrote a *Description of Wales*, similar in scope and style to his earlier *Topography of Ireland*. Both works drew on personal observations and local testimonies gathered through Gerald’s own travels and duties, but more broadly they relate to a long tradition in the Middle Ages of combining histories and geographies that Merrills traces through from the ‘canonical geographical writing’ of ‘Christian geographers’ of late Antiquity, such as the work of Orosius, Jordanes, and Isidore (see above). In Britain, this tradition was perpetuated especially through Bede’s *Ecclesiastical History of the English People*, as well as by Gerald of Wales’ topographical works on Wales and Ireland, and latterly Ranulf Higden’s *Polychronicon*, or ‘universal history’. All sought to place Britain in historical and geographical context, drawing upon earlier geographical sources, and in the process describing the character of the island and its relationship with the wider world. It is a tradition, Lavezzo argues, which is evident not only in textual accounts of Britain but also in contemporary visual representations and was instrumental in constructing and defining English self identity throughout the later Middle Ages.

Gerald begins the *History and Topography of Ireland* with an account of the island’s position and its overall dimensions in relation to Britain, and then proceeds to describe its characteristics, quoting both Bede and Solinus and contradicting them openly. He then turns to name Ireland’s “nine principal and magnificent rivers,” noting the division of the island “into five almost equal parts,” and the abundance of its lakes and islands, but leaving it to the third and final part of the work to give an account of Ireland’s history and people. In his *Description of Wales* Gerald takes a similar approach, dealing again first with “the length and breadth of Wales, its natural features and its rugged surface,” “before turning to consider the” “three more or less equal parts” of Wales, the genealogy of the Welsh princes of the three parts together with their administrative organization, and then in some detail names its rivers and mountain ranges, touching also on “the fertility of Wales and its attractiveness,” as he had previously in his *Topography of Ireland*. In defining the characteristics of Wales and Ireland, Gerald follows a common model, therefore, of addressing matters of geography before history. Geography it seems is given due prominence.

Gerald’s academic debt to Bede’s *Ecclesiastical History* is not only as a source of material but also as a model for the structure he follows in his own work. Bede’s *Ecclesiastical History* had great status and popularity among later medieval historiographers, being copied from the eighth century onward not just within Britain but across Europe too, and without too much adaptation. The ‘geographical introduction’ to the *Ecclesiastical History* is addressed thoroughly by Merrills, who notes how “few modern scholars have expended much energy on Bede the geographer” (Merrills, 2005: 249–73), probably because his sources, “a patchwork compilation of Pliny, Orosius, and Gildas,” appear to be so unremarkable and orthodox. Bede’s introduction starts with an account of Britain’s position “opposite Germany, Gaul, and Spain,” as well giving the island’s overall dimensions, thus placing it in relation to continental Europe, and the wider world. He continues by describing the fertility of the land, its rivers, emphasizing their productivity, noting “the country was once famous for its twenty eight noble cities as well as innumerable fortified places equally well guarded by the strongest of walls and towers, gates and locks” (Merrills, 2005: 329). It is thus a laudatory description, reflecting Bede’s interest in his own country, as a land as well as a people, for he also pays particular attention not just to Britain’s relations to the outside world but also to the relationships of those within it, as well as their neighbors in Ireland. Recently, the way Bede made use of British and Irish geographical marginality, as a metaphor for their spiritual purity, has attracted particular comment. Indeed, Bede’s geographical introduction as a whole is seen by Merrills as “an almost eschatological geography” that owed much to the Roman and Christian authorities he relied upon.

In their descriptions of Britain and Ireland, Gerald of Wales and Bede share in a ‘geographic exceptionalism’ that celebrated the islands’ spatial distance and physical isolation from the rest of the Christian world, a point Lavezzo develops. She suggests that Gerald’s as well as Bede’s description and perception of Ireland as a favorable place was closely tied to their ideas about what it meant to be geographically peripheral, to be on the edge. Their views of Ireland as a land of “milk, honey, and vines” emphasize its fertility and abundance, ideas taken from Solinus and Isidore, and stood as geographical metaphors for the island’s perceived sanctity and ‘divine favor’. As with Isidore’s *Etymologies*, it was not just geographical and topographical details that got copied into later works such as Gerald’s *Topography*, but spatial imaginaries too. For Lavezzo this is made all the more intriguing as she identifies a complex paradox between Gerald’s favorable description of Ireland in his *Topography* and his otherwise wholly negative representation of the Irish people, who he sees as dirty and barbaric, a contradiction whereby “Gerald transforms the natural purity that, in Bede, serves as a tool by which the writer elevates Ireland, into a means of demonizing the Irish." Gerald's descriptions of Ireland and the Irish are by no means innocent or passive illustrations but are themselves highly motivated accounts and emotively charged. He was writing on the 'character of the country' not for amusement but for political purpose, as was the case with his *Description of Wales*, where he concludes by giving explicit advice to the English on 'how the Welsh can be conquered', and 'how Wales should be governed once it is conquered'. He wrote the *History and Topography of Ireland* soon after participating in the English conquest of Ire land in the reign of Henry II, having accompanied the king's son, John, in his attack on Waterford in 1185.

While Gerald's 'geographical introductions' can be seen in the context of perpetuating an established medieval historiographical tradition of linking geography and history, as with Bede's description of Britain and its 'eschatological geography', and like those earlier histories of Isidore and Orosius, there are within them inherent spatial imaginaries that reveal something of their author's thinking. With Gerald and Bede, their apparent geo graphical remoteness, writing as they were at the per ceived edge of the world, perhaps comes through in the accounts of the lands they seek to describe. Yet, at the same time, despite their own marginal locations in Brit ain they were politically and intellectually connected, both at the time they were writing with contemporary ecclesiastical and aristocratic networks, and also across both time and space through the past sources they drew upon and in the spatial imaginaries they shared and passed on. These networks stretched the agency of par ticular geographical ideas across the realm. Higden, for example, was called by King Edward III to attend court and to bring with him his chronicles, though for what purpose is not recorded, while an 'estate book' of the lands that were the property of a Master Henry de Bray, farming in the east midlands of England around 1300, included at its start a 'description of the world' (*descriptio mundi*) and the three continents, a list of English counties, bishoprics, and kings, as well as an account of Henry's own land.

In addition to textual descriptions of Britain and Ireland's peripheral place in relation to the wider world, Lavezzo argues that the same English spatial imaginary was manifest visually in maps accompanying historio graphical works of the later Middle Ages, including Gerald's *Topography*, as well as Ranulf Higden's *Poly chronicon*. These images likewise placed Britain at the geographical margins of the world, and in her view reveal how Gerald and others were thinking geographically about the place of England (specifically), and the English, in relation to Ireland and continental Europe, and con structing cultural identities through the relative placing of others. With Gerald and Higden, maps and text are thus working together to mutual effect. They are also relating *topography* – as place description – and *geography* – as Earth description, and hence connecting the local to regional and to global scales, reinforcing a tiered spatial hierarchy as Isidore made evident in his *Etymologies*. Rather than being two separate genres then, 'academic geographies' and 'topographies' were as interwoven as history and geography were. The nature of these inter woven medieval geographies offers the opportunity to enhance our understanding of the ways in which con ceptions of space and place were socially constructed and mediated in textual and visual geographies of the Middle Ages.

Conclusion

Geographical thought and ideas of the Middle Ages should no longer be seen as a hiatus in geography's his tory. Rather, 'medieval geography' has connections, both empirical and conceptual, with those geographies of the ancient world that it borrowed from, and those of the modern age that it continued to influence. Indeed, through the work of geographers such as Beazley, Wright, and Kimble in the early twentieth century, it could be argued that medieval geography helped shape the modern discipline, at least in its Anglo American manifestation.

From studying medieval geography, what becomes clear are the continuities running throughout geog raphy's history, from Antiquity to the present day, in the circulation of geographical texts, as well as in the spatial construction of knowledge. They are evident, for ex ample, in charting a 'cartographic genealogy of the Earth' that spans the ancient, medieval, and modern worlds. Indeed, such temporal–geographical connections and continuities point to the need to avoid artificially di viding between ancient, medieval, and modern geog raphies. A more holistic and integrated approach to writing geography's history is required. The debt owed to earlier 'encyclopedic geographies' in the classification of geographical knowledge and 'encyclopedism' of the early modern period is a case in point. These continuities deserve much greater scrutiny and appreciation among human geographers, both to challenge enduring as sumptions about geography in the Middle Ages and to reverse the growing neglect shown by human geog raphers toward studying the medieval period.

As literary and intellectual historians are demon strating, geography's medieval history is due a careful critical reappraisal, and here human geographers have a continuing contribution to make, for example, in bringing to bear their ongoing debates about the 'geographical tradition' and of 'putting science in its place'. Such thinking will not only help to deepen understanding of the nature of medieval geography but it will also help to ensure that geographers do not lose this part of their discipline's past. There are contributions to be made by human geographers, therefore, in the telling of their own subject's medieval history. From reviewing the recent work of nongeographers one area where such a contri bution might be particularly insightful is in exploring the spatialities of 'medieval geography', not simply what geography was in the Middle Ages, and what was covered by 'geographical' texts and images, but more the signifi cance of what these can tell us about geographical cul tures of contact, learning, and dissemination of ideas and knowledge in the Latin West and beyond. This requires a study of 'geography *per se*' as well as 'geographical knowledge', and an approach that crosses disciplinary boundaries and differences. The medieval geographies outlined above hint at this potential.

See also: Cartography, History of; Cartography in Islamic Societies; Cosgrove, D.; Enlightenment Geography; Feudalism and Feudal Society; Geography, History of; Historical Geography; Knowledge and Education, Historical Geographies of; Medieval Historical Geographies; Wright, J. K.

Further Reading

Barney, S. A. Lewis, W. J., Beach, J. A. and Berghof, O. (eds. and trans.) (2006). *The Etymologies of Isidore of Seville*. Cambridge: Cambridge University Press.
Beazley, C. R. (1897). *The Dawn of Modern Geography*. London: John Murray.
Beazley, C. R. (1901). *The Dawn of Modern Geography*, vol. II. London: Henry Frowde.
Beazley, C. R. (1906). *The Dawn of Modern Geography*, vol. III. Oxford: Clarendon.
Cosgrove, D. E. (2001). *Apollo's Eye. A Cartographic Genealogy of the Earth in the Western Imagination*. London: John Hopkins University Press.
Edson, E. (1997). *Mapping Time and Space. How Medieval Mapmakers Viewed Their World*. London: The British Library.
Edson, E. and Savage Smith, E. (2004). *Medieval Views of the Cosmos. Picturing the Universe in the Christian and Islamic Middle Ages*. Oxford: The Bodleian Library.
Howe, N. (2008). *Writing the Map of Anglo Saxon England: Essays in Cultural Geography*. New Haven, CT: Yale University Press.
Kimble, G. H. T. (1938). *Geography in the Middle Ages*. London: Methuen.
Kline, N. R. (2001). *Maps of Medieval Thought: The Hereford Paradigm*. Rochester, NY: Boydell.
Lavezzo, K. (2006). *Angels on the Edge of the World. Geography, Literature, and English Community, 1000 1534*. London: Cornell University Press.
Livingstone, D. (1992). *The Geographical Tradition: Episodes in the History of a Contested Enterprise*. Oxford: Blackwell.
Livingstone, D. (2003). *Putting Science in Its Place: Geographies of Scientific Knowledge*. Chicago, IL: University of Chicago Press.
Lozovsky, N. (2000). *The Earth Is Our Book. Geographical Knowledge in the Latin West ca. 400 1000*. Ann Arbor, MI: University of Michigan Press.
Merrills, A. H. (2005). *History and Geography in Late Antiquity*. Cambridge: Cambridge University Press.
O'Meara, J. J. (ed. and trans.) (1982). *The History and Topography of Ireland: Gerald of Wales*. London: Penguin.
Stahl, W. H. (ed. and trans.) (1952). *Macrobius Commentary on the Dream of Scipio*. New York: Columbia University Press.
Stahl, W. H. (1962). *Roman Science. Origins, Development and Influence to the Later Middle Ages*. Madison, WI: University of Wisconsin Press.
Stahl, W. H. and Johnson, R. (1977). *Martianus Capella and the Seven Liberal Arts, vol. II, The Marriage of Philology and Mercury*. New York: Columbia University Press.
Taylor, J. (1966). *The Universal Chronicle of Ranulf Higden*. Oxford: Clarendon Press.
Thorpe, L. (ed. and trans.) (1978). *Gerald of Wales: The Journey Through Wales and the Description of Wales*. London: Penguin.
Willis, D. (ed.) (1916). *The Estate Book of Henry de Bray of Harleston, Co. Northants (c.1289 1340)*. London: Royal Historical Society.
Wright, J. K. (1925). *The Geographical Lore of the Time of the Crusades. A Study in the History of Medieval Science and Tradition in Western Europe* (1965 edn.). New York: Dover.

Relevant Websites

http://home.newadvent.org/cathen/
Catholic Encyclopaedia.
http://www.collectbritain.co.uk/
Collect Britain (The British Library, London).
http://www.davidrumsey.com/
David Rumsey Map Collection.
http://www.maphistory.info/
History of Cartography.
http://www.fordham.edu/halsall/sbook.html
Internet Medieval Sourcebook' (Fordham University, USA).
http://cosmos.bodley.ox.ac.uk/hms/home.php
Medieval Islamic Views of the Cosmos (Bodleian Library, Oxford).

Medieval Historical Geographies

R. Jones, Aberystwyth University, Aberystwyth, UK

Glossary

Demesne The land held directly under the lord's control in the feudal system.

Feudalism A social and spatial order in which power was devolved from a lord to a vassal and which was underpinned by a peasant economy.

Kinship The real and imagined extended familial relationships between groups of people.

Peasant Lowly individuals, who provided goods and services for the land that they received from their lord.

Vassal An individual who governed a tract of land on behalf of his or her overlord.

Introduction: What Is the Medieval Period?

The medieval period or Middle Ages, extending from approximately AD 500 to approximately AD 1500, was a crucial phase in the development of political, cultural, economic, and ecological aspects of human societies. In political terms, the medieval period witnessed the emergence of early forms of territorial control. With regard to culture, more highly defined ethnic communities of people emerged during the Middle Ages. Fundamental developments also occurred in an economic context with the emergence of early forms of capitalist activities based on towns and cities. Finally, significant ecological changes took place during the medieval period, most clearly in terms of an increased human exploitation of natural resources. Importantly, geography was implicated in each of these developments. Various kinds of geography reflected these crucial developments for humankind but, at the same time, they also helped to shape the character of these developments.

Before proceeding to outline the key historical geographies of the medieval period, it is necessary to think through our use of the term 'medieval' or 'Middle Ages'. In the first place, we need to appreciate the length of the period encapsulated by these terms.

A short article that deals with medieval historical geographies can be criticized for skating over an extremely long period of history in too few words. Clearly, political, cultural, economic, and ecological forms and practices would have varied considerably between the collapse of the Roman Empire in approximately AD 500 and the beginnings of the modern period in approximately AD 1500. In this respect, it is difficult to encompass all of the social and spatial forms that existed within this period in one short article. The aim in this article, therefore, is to draw attention to some of the key changes that took place during the medieval period in a variety of different contexts and to show how each of them possessed important geographies.

In the second place, we also need to focus on another important idea implied within the use of the term '*Middle* Ages' or '*medi*eval'; namely that the medieval period is a distinct epoch between two other periods. We need to think, therefore, about character of the relationship between the medieval period and both the earlier classical period and later early modern period. One can elucidate evidence for a temporal disjuncture between the classical and medieval periods, for instance, in the fact that towns and cities within Europe, which had been key nodes within a Roman imperial network, fell into disrepair and were gradually abandoned during the early medieval period. And yet, Roman towns and cities, in spite of a certain change in function, were still used throughout Europe in the centuries following the breakup of the Roman Empire. In this respect, it is difficult to sustain the claim that the beginnings of the medieval period represented a radical shift in the character of human societies. There were important overlaps between the classical and early medieval period. At the same time, some have maintained that the shift to the modern period (experienced around AD 1500) also signaled fundamental changes to the various sociospatial organizations and processes operating within society. Various authors have contended, for instance, that forms of political and cultural identity were radically reconfigured during the modern period. Equally, there is a strong sense that the modern period, especially in Europe, witnessed the formation of distinctive and new forms of polities. Others have shown that there was much continuity between the medieval and modern periods and, indeed, it is possible to show that each of the allegedly 'modern' forms of politics, culture, and economy described above possessed roots that were laid down during the medieval period. Taken together, such arguments mean that we should avoid thinking of the medieval period as a totally distinct and separate period of human history. Rather, it is a period of human history that sustained many of the political, economic, cultural, and ecological forms and practices that were prevalent during the classical period. Similarly, the medieval period also laid the foundations for many of the sociospatial forms and practices that were to assume greater significance during the modern period.

The article is structured as follows. The first substantive section examines the way in which various writers sought to think geographically during the Middle Ages and focuses in particular on the various efforts made to describe or write about the Earth, its peoples, and places. The following section then proceeds to discuss some of the ways in which contemporary geographers have attempted to study the various historical geographies of the medieval period and examines in detail issues relating to political transformation, the processes of urbanization, and the environmental histories of the Middle Ages. Brief conclusions then follow in which some broader implications of medieval historical geographies for Geography as a discipline are discussed.

Thinking Geographically during the Middle Ages

The attempts that were made to think and write geographically during the Middle Ages were associated, in large measure, with the practices of geographical mobility that occurred during the period. New experiences, brought about by travel and mobility, could inspire individuals to write detailed accounts of alternative polities, cultures, economies, and ecologies that they had encountered. At the same time, geographical narratives about particular peoples and places could inspire further acts of mobility. In this section, three different contexts within which geographical writing and mobility have been implicated are highlighted. The first relates to a largely internal European dynamic of mobility that encouraged people to write about people and places. The second focuses on the process whereby European scholars began to write geographically about lands that they had experienced outside Europe. This is an important example, since it presages the geographical writing that underpinned much of the European imperial acts committed during the modern period. Finally, an example that challenges an Eurocentrism that is implied in the first two examples is discussed. The third example focuses on a set of mobile practices and a geographical narrative that took place wholly outside Europe.

Many historians have noted that a key dynamic of the European Middle Ages was the way in which Frankish and Norman lords and noblemen extended the territorial reach outward from their 'heartlands' in northern France to more outlying areas of Europe. These acts of conquest and colonization, which took place between the eleventh and fourteenth centuries, were intimately tied up with acts of geographical writing. On the one hand, the encounters between Frankish and Normand lords and knights and other people, living in the more 'peripheral' areas of Europe, encouraged the former to write damning accounts of the latter. In turn, these geographical writings could justify further acts of expansion and colonization by these Frankish and Norman lords. Echoing the use of forms of othering within a modern imperial project, the key tropes that were used to structure geographical writings concerning a European 'periphery' were those relating to barbarity. Postcolonial critics of medieval Europe, for instance, have drawn particular attention to constructions of religious difference, which marked out internal and external others, notably through binaries of Christian/Jew and Christianity/Islam, respectively. Some of the most startling illustrations of this process are the various condemnations of the ways in which these 'peripheral' peoples lived. For instance, Gunther of Pairis referred to the early medieval Poles as:

> a people whose culture is crude; terrible of face, with a frightening brutality in their customs. They rage with a horrible sound, they are ferocious, threatening. Quick of hand, lacking reason, used to plundering. They scarcely behave like men, their cruelty is worse than the horror of wild beasts.

If these 'peripheral' were barbarous, as many writers asserted, then they needed to be 'civilized' by the representatives of an European 'core'. Of course, the most striking example of this connection between geographical writing, mobility, and a 'civilizing' project was that associated with the Crusades (AD 1096–AD 1291). Over nearly 200 years, hundreds of thousands of representatives of an European conquering and civilizing project traveled to the eastern Mediterranean to appropriate lands that were deemed to be central to the political and cultural identity of a Christian Europe.

The second example provides a different take on the connection between mobility and geographical writing during the Middle Ages by discussing one of the first attempts made by an European to write a geographical account of lands lying outside Europe. Written at the very end of the thirteen century, Marco Polo's *Travels* provide a 'description of the world', which extended from the area of Venetian influence in the Mediterranean to the Far East and from the Persian Gulf to the far north of Russia (see **Figure 1**). Marco Polo's ability to write such a far reaching geographical account derived in large measure from the military domination and political stability brought about by the creation of the Mongol Empire from the beginning of the thirteenth century onward. The Mongol Empire's control of most of the Eurasian continent led to an increased access for Europeans to trade routes, which led to the Far East. Venetians, such as Marco Polo's father and uncle, made the most of these new trading opportunities by traveling into, and trading with, the various lands controlled by the

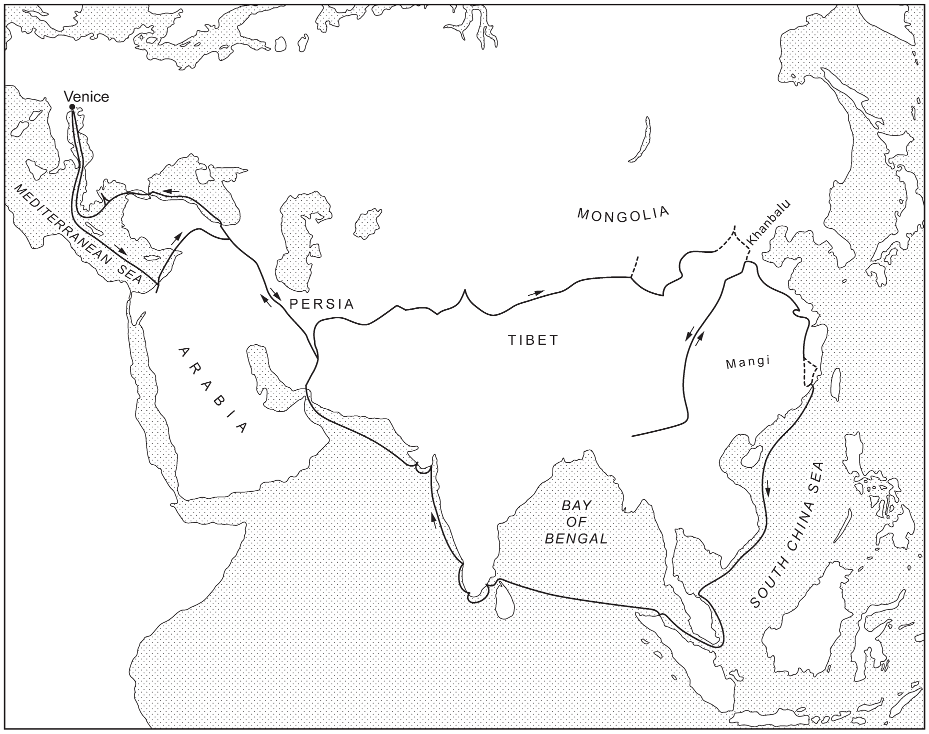

Figure 1 The Travels of Marco Polo.

Mongol Empire. Indeed, Marco Polo himself gained much support from the last of the great rulers of the Mongol Empire, Kublai Khan, and was sponsored in his travels throughout the empire. The *Travels* itself is a combination of Marco Polo's experiences of traveling throughout the empire over a number of years and a series of knowledges and narratives that he collected on his travels. What is striking about Marco Polo's account of his movements is its ethnographic qualities. His de tailed narrative makes reference to the religious beliefs, the character, the trading activities, the styles of agri culture, and so on practiced by people living within particular cities. His narrative is equally concerned with explaining the history of particular peoples or places. He provides a detailed description, for instance, of how the Province of Manji was conquered by the Grand Khan and incorporated into the Mongol Empire. Following on from this, Marco Polo also tries to explicate the political economy that underpinned the Empire. For instance, he consistently notes the military and political structures put in place by the Khans to enable them to rule distant localities. Similarly, he discusses the paper money economy and the taxation regimes that helped to sustain the Empire's war machine. Finally, his narrative is also concerned with outling the way in which what we may term the physical geography of the Empire contributed to its sustenance. Marco Polo, for instance, describes in detail the various rivers that he encountered on his tra vels, as well as the different kinds of crops and foodstuffs grown in different regions. The latter discussion even includes an account of the high quality of the rhubarb grown in one upland region of the Province of Manji! In total, Marco Polo's account of his *Travels* in the Far East are based on a remarkably sensitive and wide ranging appreciation of the distinctive character of particular places and, as such, can be viewed as a forerunner of the kind of regional geography that became prevalent in human geography during the first half of the twentieth century. What is equally significant about his account is the way in which it is underpinned by an obvious ad miration of the social and spatial order put in place by the rulers of the Mongol Empire.

The final example of geographical writing that is discussed is slightly different in that it is based on the extensive travels of a Muslim judge, Ibn Battutah, throughout much of North Africa and Asia during the fourteenth century. The significance of this example is that it helps to challenge some of the Eurocentric as sumptions concerning the Middle Ages. It demonstrates the way in which an alternative Muslim world and worldview existed – and, indeed, was highly important – during the medieval period. Ibn Battutah's main motiv ation for commencing his travels was the requirement for all able bodied Muslims to complete a pilgrimage to Mecca at least once during their lifetime. His ability to travel was conditioned, in large measure, by the same factors that affected Marco Polo; namely the existence of an extensive, relatively stable and, by the fourteenth century, largely Muslim land based Mongol Empire. The latter factor was highly significant given Ibn Battutah's profession as a Muslim judge. His highly portable skills and knowledges gave him the ability to gain employment in a large proportion of the towns and cities that he visited on his travels. Ibn Battutah's descriptions of the places that he visited mark him out not only as a geog rapher but also an ethnographer, a historian, and botanist. His account of southern India may be taken as an illus trative example of geographic writing. At one level, he is concerned with outling some of the broader 'political geographies' that characterize the region, whether in terms of the ethnic and religious makeup of the region or in terms of forms of political control and subordination. At the other extreme, he also provides detailed accounts of some of the more mundane practices of the inhabit ants. One fascinating section, for instance, outlines the manner and order in which different foods were served within a particular banquet. In a similar fashion to Marco Polo, Ibn Battutah was interested in both the human and the physical geographies of the places that he visited. He drew attention, for example, to the types of vegetation that grew within particular places but was also keen on showing the way in which these were then farmed productively and traded. Contrary to Marco Polo, Ibn Battutah's accounts of the places that he visited are underpinned by his own personal identity as a Muslim. He notes the location of mosques, the numbers of Muslim people living within particular places, as well as the way in which the Muslim faith influenced the various religious and secular practices that he witnessed. Even though Ibn Battutah had been diverted many times from his original pilgrimage to Mecca, his extensive travels over a total of 29 years were still predicated by his worldview as a Muslim.

What similarities can we note in these various medi eval geographical writings? First, each of these different types of geographical accounts are linked in some way to mobility. Geographical writing could either affect prac tices of mobility or it could be produced as a result of an individual's or a group's mobility. Second, one can also cast some aspersions on the veracity of some of these accounts. Most clearly, the geographical writings about 'peripheral' European peoples and places were instances of propaganda and did not truly reflect their actual character. The accounts by Marco Polo and Ibn Battutah have also been criticized by medieval and more con temporary commentators for their lack of accuracy. Third, there is very little sense of the collection of what we may term geographic data as a basis for these nar ratives. They are, at best, based upon the personal ob servations of the traveler. Fourth, and following on from the previous point, the various geographical writings of the medieval period – particularly those of Marco Polo and Ibn Battutah – are highly descriptive in nature. There is little effort here to explain the people, places, and processes witnessed in the accounts, apart from through reference to religious worldviews. Not all of these issues are shortcomings as such; they merely demonstrate that the type of geographic writing com mitted during the Middle Ages was of a different char acter to that conducted on the medieval period today. It is these more contemporary studies of the medieval period that are discussed in the following section.

Thinking Geographically about the Middle Ages

The study of the historical geographies of the medieval period has been prevalent within the discipline of geography for much of the past 100 years. The greatest medieval geographer, arguably, was Sir Henry Clifford Darby, the first professor of geography at Liverpool (1945–49), and who was later professor at University College London (1949–66), and at Cambridge (1966–76). Darby conceived that historical geography should be concerned with understanding the historic changes that had affected landscapes and regions. Moreover, it was an approach that was based upon using contemporary methods and approaches as a way of understanding his toric data. This approach was utilized in his early re search on the medieval Fenland of the East Anglia of England but came to the fore more fully and more famously in the context of his editorship of *The Domesday geography of England*, in which he sought to use an un paralleled data source (the Domesday Survey of England, conducted in 1086) as a way of illustrating the way in which various places, located throughout England and Wales, had differed from each other during the medieval period. Of course, the promotion of the study of medieval historical geography did not come about single handedly. Other notable academics were involved in this project, most notably Emrys Bowen at Aberystwyth, Bruce Campbell at Belfast, Della Hooke and Terry Slater at Birmingham, Glanville Jones at Leeds, and Brian Roberts at Durham.

The study of medieval historical geographies, there fore, represents a varied, if rather limited, aspect of contemporary human geography. At present, medieval historical geographers examine various themes relating to medieval political geography, cultural geography, eco nomic geography, population and land use, and ecology. It is impossible to review the variety of work being conducted on these themes in the remaining paragraphs of this section. Instead, the focus is on three illustrative aspects of the research currently being conducted on medieval historical geographies in order to provide a sense of the range of different ways in which human geographers are grappling with the social and spatial forms that characterized the Middle Ages.

Towns and Urbanization during the Middle Ages

The study of towns and processes of urbanization has been an important aspect of research on the Middle Ages in geography and other related disciplines. Research on this theme has been varied. In a quite straightforward context, geographers have been able to show how different medieval towns varied considerably in terms of their characteristics. The character of towns and cities, for instance, varied considerably during the early Middle Ages. The main reason for such a variation was the distance of the town or city from the geographical core of the Roman Empire. The decline in the functional vitality of urban centers occurred to a far greater extent in the early Middle Ages in parts of Northern and Western Europe, for instance, than it did in southern France and Italy. At the same time, we need to appreciate how the character of towns and urban areas changed in a temporal sense. It has been argued, for instance, that urban areas shifted from being merely centers for trade during the eighth and ninth centuries (described as *emporia* in Latin and often possessing the Germanic suffix *wik/wiic/wich*) to being centers for trade and production for much of the remaining medieval period. In this sense, we need to acquire a temporal and geographic sensitivity if we are fully to understand the variety of urban forms that existed during the Middle Ages.

Equally, geographers have sought to demonstrate the important geographies that were associated with towns during the Middle Ages. First of all, it is evident that medieval towns were centers for trade over different geographical scales. Italian cities, such as Genoa and Venice, for instance, lay at the heart of long distance networks of trade that extended out to much of the known world and which imported a range of expensive goods – most notably spices and silk – from Asia into Europe. Indeed, the existence of such 'international' trading routes acted as the motivation for Marco Polo's extensive travels throughout much of Asia during the thirteenth century, discussed above. Similarly, high quality textiles and metal goods were exported from a variety of towns and ports located in northwestern Europe to Italy and a variety of different trading towns located along the Baltic. Such 'international trade' helped to foster long distance connections between various towns and cities but it was also dependent upon the development of stronger associations between towns and their immediate hinterland. The growing economic and trading significance of such towns, especially during the early Middle Ages, was predicated in large measure on the high levels of craft production that occurred within their immediate hinterlands. Work that has charted the character of such local and long distance trade illustrates the significant topological connections that characterized the medieval period and acts as a corrective to popular assumptions concerning the localized nature of medieval life.

A further key aspect of the process of urbanization that took place during the Middle Ages was its role as a facilitator of certain social order. The creation of guilds as means of regulating craft production, for instance, brought about a certain degree of socioeconomic order within towns and cities. We can also think about the role played by towns and cities in promoting order in a more sociocultural context. Towns and cities have always been connected with ideas of civilization and, in this respect, the formation of towns or boroughs in medieval Europe was viewed as a mechanism through which 'peripheral' and allegedly 'barbaric' people could be 'civilized'. A number of historical geographers, for instance, have shown how the process of creating boroughs in Wales and Ireland was used by Anglo Norman kings and lords in order to conquer and subdue the Welsh and Irish people. The creation of new boroughs, possessing liberal social and economic rights, could be used to attract incomers from the core areas of the Anglo Norman Empire. The arrival of these individuals brought a greater stability to the Anglo Norman presence within these 'peripheral' areas. At the same time, this process of urbanization possessed important geographies. We witness this most clearly in the way in which these newly created boroughs were designated as 'pure' spaces within which Irish or Welsh people were not allowed to venture. In the settlement of New Ross in Ireland, for instance, the town's original Irish inhabitants were expelled and were forcibly relocated to a suburb which lay outside the town walls (see **Figure 2**).

Medieval Territorial Power

Equally important has been the research conducted by geographers and others on the changing political and territorial forms that characterized the Middle Ages. Two key changes affected European political space during the medieval period. The first was the reemergence of larger and relatively more stable political formations. The small scale, haphazard, and temporary political formations of the early Middle Ages gradually gave way to a series of larger and more stable political forms; so much so, indeed, that the European political map was far simpler by 1500. It contained far fewer but larger political units. Secondly, there was a shift in the way in which political space was conceived, becoming far more territorial in its makeup. The political formations of the early Middle Ages were conceived primarily as groups of people and only secondarily as territorial entities. During the early Middle Ages, ties of kinship helped to bind

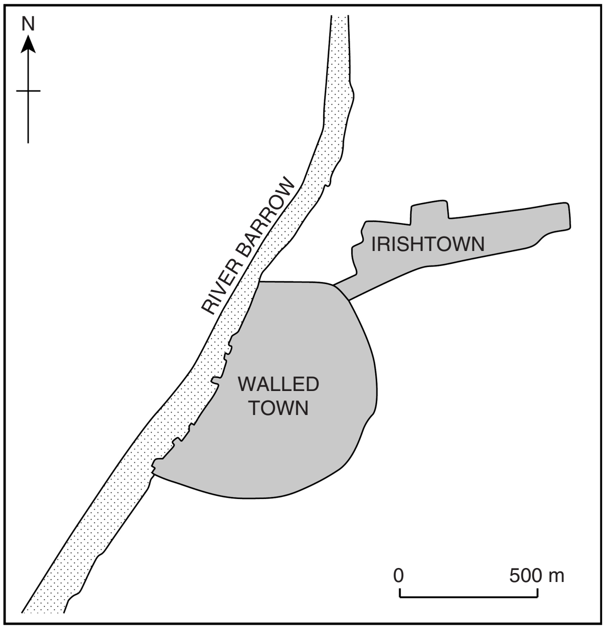

Figure 2 Social and spatial segregation in New Ross, Ireland.

groups of people together. The feudal system of order, which dominated much of Europe during the medieval period was also, in essence, a mechanism of governing space that was dependent on social relations between various groups of people; most notably between lord and vassal and between lord and peasant. The political formations of the later medieval period, however, began to be viewed, first and foremost, as territorial states. Both these changes are, of course, interrelated. The development of larger polities throughout the Middle Ages justified the creation of more bureaucratic and territorialized forms of state governance while the development of these new means of governance enabled the existence and survival of larger and more stable polities.

Given the significance of the control of territory for these emerging polities, geographers have been at the forefront of debates concerning these important political changes. Attempts have been made, for instance, to map out the growing territorialization of power experienced in Wales between the eleventh and the thirteenth centuries. The precise definition of lordships and territory became increasingly apparent in thirteenth century Wales, especially in the period subsequent to the Anglo Norman conquest of 1277–83. Following the conquest, it increasingly became the norm for Anglo Normans to lobby their king so that he could receive specific territories from him. For instance, Roger Lestrange in 1283 pleaded with the king, asking whether he could ‘obtain for him Penlyn or Edeyrnyon, or one of them; or he would prefer Maillor Saysnek: he would very much prefer to come to Maillor Saysnek’. Of course, what enabled this carve up of the Welsh landscape was an exact definition of the economic and political rights and dues associated with precisely defined territories. It is in this context that we witness the increasing territorialization of political power during the Middle Ages. Interestingly, these developments were mirrored in many ways by equally fundamental changes to the governance of Church lands. It has been shown how the economic and political power of the Church in medieval Cornwall, for instance, was originally promoted and latterly underpinned by the emergence of a territorialized network of parishes. Here, once again, we witness the way in which the development of new forms of political and economic authority within the Middle Ages were intimately tied in with a growing territorialization of power. Geographers, in this respect, have led the efforts to understand the discourses and practices that facilitated this process.

The Environmental Histories of the Middle Ages

Finally, let us discuss briefly the way in which human geographers and others are beginning to focus on a rather neglected aspect of medieval history, namely the environmental history of the Middle Ages. Work in this area is beginning to examine: the opportunities and constraints provided by the environment during the Middle Ages; the environmental impact of medieval societies, focusing particularly on the way in which the environment was transformed during the Middle Ages; and the way in which various environmental hazards impacted on the development of medieval societies.

For the sake of brevity, let us examine two instances, which illustrate the connections between medieval societies and environmental history. The first example shows the impact of medieval forms of social and economic governance on the environment, as well as some of the environmental opportunities and constraints that helped to transform medieval society in Europe during the later Middle Ages. Work by historical geographers and others has begun to examine the significance of the environment as a constraining factor in the reproduction of feudalism as a means of socioeconomic order. Feudalism was predicated on a decentralized form of political control as various lords acted as their overlord’s local representatives or vassals. In more economic contexts, feudalism was based upon a system in which peasants and slaves were required to fulfill certain obligations for their lord; by paying money rents, providing food renders, or working the land held directly in the lord’s hands (*demesne*). In return, the lord offered a certain amount of land to his peasants so that they could subsist. As population numbers increased between the tenth century and the end of the thirteenth century, and as a ‘parasitic class’ of lords sought to increase the amount of money rents, renders, and services that they received from their peasants, so the environmental pressures placed upon the land increased. Peasant production decreased and, as a result, famines and widespread human disease became more prevalent, especially during the fourteenth century. The argument, in essence, is that the environmental constraints of the feudal system ultimately brought about its collapse as a means of organizing medieval society.

The second example, once again, focuses on the interrelationship between medieval political forms and the environment but does so in a far more targeted manner. Recent research has examined the way in which the political systems that have been present in western highlands of Guatemala over time have influenced the types and the range of varieties of maize contained within the region. For much of the postclassic Mayan Period (AD 900 to AD 1524), Mayan society in the western highlands of Guatemala was organized in the form of segmentary lineages. This form of social organization is based upon notions of kinship. The main social and political unit was that of a relatively small lineage group but the social system also allowed for temporary alliances to be formed between different lineage groups, especially during the times of crisis. Significantly, the existence of localized social political units such as these impacted on the crop varieties – most notably maize – that were produced within the western highlands. Although a certain amount of trade in portable and high value goods took place within this society, trade in low value and high volume items, such as maize seeds, was extremely limited. The lack of a meaningful overarching political structure in the western highlands of Guatemala during the postclassical period, in effect, maintained a high degree of crop variety, which can still be detected today. Significantly, the variety of maize decreased in the period subsequent to the Spanish conquest of Guatemala (AD 1525–AD1821). This shift was not due to the imposition of a unified colonial political and social system – and a related increase in trade and exchange of maize varieties – as one might expect but rather because of the catastrophic decline in the native population, which took place between 1520 and 1770; it is estimated that the population declined by about 90% during this period. Fewer inhabitants and fewer farmers, as a result of this decline, meant that fewer varieties of maize were being produced and sustained over time. In this context, we witness how the environmental and social histories of the medieval period could combine in a kaleidescope of influences.

Conclusions

In this article, the key features of a medieval historical geography have been outlined. The article began by discussing the significance of geographic writing during the medieval period. Contributions by people like Marco Polo and Ibn Battutah were brought about by political and social changes that took place during the thirteenth and fourteenth centuries, and their accounts have much to say about the character of people, places, and processes during this period. The study of medieval historical geographies in the contemporary period was also discussed. Three illustrative areas of contemporary research were highlighted; relating to medieval urbanization, medieval territorialities, and medieval environmental histories. Interesting and innovative research by geographers is being conducted in each of these contexts.

Let us conclude the article by raising two more problematic issues concerning the study of medieval historical geographies. The first issue relates to the tendency for studies of medieval historical geographies to display an Eurocentrism. Let us be conscious, in this regard, to the fact that the current article is largely Eurocentric in its outlook, dealing as it does with case studies linked to an European feudal society. To a certain extent, such an outlook is understandable since the use of the terms 'medieval' and 'Middle Ages' is based upon a temporal framework that is rooted within the historic trajectories of Europe. At the same time, there is scope to challenge the Eurocentrism of studies of medieval historical geographies by showing how the sociospatial formations that were prevalent in medieval Europe also existed – albeit perhaps in different formats – in other parts of the world. Indeed, it is likely that these extra European medieval historical geographies have already been examined by geographers, historians, and others working within alternative national and linguistic traditions. In this regard, we may undermine the Eurocentrism of Anglo American and European understandings of the meaning of medieval historical geography merely by increasing our academic dialog with colleagues working within other countries. The second issue that needs to be raised is the relative decline in the number of studies being conducted by historical geographers on the medieval period. The reasons for this decline are not known but may be due to combination of factors. First, the lack and intractability of empirical sources for the Middle Ages may make it difficult for some to conceive of a research career that focuses on medieval historical geographies. This is especially pertinent since some of the linguistic skills – in Latin and Greek, for instance – necessary for medieval research is less prominent within school education than in the past. Linked to this is a second possible explanation, namely that the processes and sociospatial formations of the medieval period are so different from those of the more recent past and the present that to study them would require the development of a totally new set of geographical knowledges. Third, and building on the previous point, researchers increasingly may believe that medieval historical geographies are irrelevant as a means of understanding some of the key political, economic, social, cultural, and ecological challenges that face us in the contemporary world. And yet, one could argue that there is much to be gained by maintaining or, perhaps, re extending our temporal horizons so that they take ac count of the medieval period. Many of the sociospatial formations with which we are familiar with today were pre saged during the Middle Ages. One important way of understanding some of the processes affecting human societies and environments in the contemporary world is by gaining a fuller comprehension of how they have been altered in the past, including the medieval period.

See also: Historical Geography; Medieval Geography.

Further Reading

Bartlett, R. (1993). *The Making of Europe: Conquest, Colonization and Cultural Change 950 1350*. London: Penguin.

Darby, H. C. (1952 1977). *The Domesday Geography of England*, vols. 1 7. Cambridge: Cambridge University Press.

Dodgshon, R. A. (1998). The socio political map of Europe, 400 1500. In Dodgshon, R. A. & Butlin, R. A. (eds.) *An Historical Geography of Europe*, pp 54 72. Oxford: Oxford University Press.

Harvey, D. (2000). Landscape organization, identity and change: Territoriality and hagiography in medieval west Cornwall. *Landscape Research* 25, 201 212.

Jones, R. (2007). *People/States/Territories: The Political Geographies of British State Transformation*, pp. 49 79. Oxford: Blackwell.

Lilley, K. (2000). Non urbe, non vico, non castris: Territorial control and the colonization and urbanization of Wales and Ireland under Anglo Norman lordship. *Journal of Historical Geography* 26, 517 531.

Mackintosh Smith, T. (ed.) (2003). *The Travels of Ibn Battutah*. London: Picador.

Moore, J. W. (2002). The crisis of feudalism: An environmental history. *Organization and Environment* 15, 296 317.

van Etten, J. (2006). Molding maize: The shaping of a crop diversity landscape in the western highlands of Guatemala. *Journal of Historical Geography* 32, 689 711.

Verhulst, A. (1998). Towns and trade, 400 1500. In Dodgshon, R. A. & Butlin, R. A. (eds.) *An Historical Geography of Europe*, pp 100 114. Oxford: Oxford University Press.

Relevant Websites

http://www.the orb.net
The ORB, the Online Reference Book for the Middle Ages.

Mega-Cities

Y.-M. Yeung, The Chinese University of Hong Kong, Hong Kong, People's Republic of China

Glossary

Extended Metropolitan Region A term coined by McGee to describe the mega-urban region found in many countries in the Asia Pacific region. 'Desakota' is the term used to highlight the stage of rural–urban transition.

Global City A class of particularized cities that derive and thrive on the process of globalization. Saskia Sassen's seminal study on the global cities of London, New York, and Tokyo in 1991 has since been much extended.

Growth Triangle A form of subregional economic cooperative entity often involving three or more countries, capitalizing on their shared complementarities and geographical strengths, applied especially to the Asia Pacific region since the 1980s.

Mega-City A very large city consisting of at least 10 million inhabitants.

Mega-Urban Region A region built around one or more mega-cities, stretching as far as 100 km from the city. The transition from urban to rural landscapes is highly visible.

Megalopolis A term used by Jean Gottmann in his 1957 study of the northeastern seaboard of North America from Boston to Philadelphia, characterized by a more or less continuous stretch of dense urban settlements.

Meta-City A class of exceptionally large cities with a population of at least 20 million.

Third World City A city located in the developing world.

Urban Corridor A spatial form highlighting the spread of a number of cities over an area and showing a directional orientation.

Urban Planning A discipline or profession that concerns itself with several dimensions of the built and social environments in an urban setting.

World City A city that gains celebrity as it performs certain critical economic or other functions having a direct or tangible effect on global affairs.

Introduction

The twentieth century was unprecedented in the rapid growth and urbanization of the planet's human popu lation, particularly during the latter half of the century and in developing countries. The global urban popu lation, which stood at 0.73 billion in 1950, almost quad rupled to 2.84 billion in 2000, and to 3.15 billion in 2005, for an annual growth rate of 2.65% in the period 1950–2005. During the same period, the world's population grew at a rate of 1.71% per year; thus, its urban popu lation grew more than twice as quickly as its rural population. The urban population in more developed regions was 0.42 billion in 1950 versus 0.31 billion in less developed regions. By 2000, the relative weights had reversed, and the respective figures were 0.87 billion and 1.97 billion, further increasing to 0.90 billion and 2.25 billion, respectively, in 2005. In the period 1950–2005, the urban population in less developed regions grew at a rate of 3.61% per year, nearly three times as rapidly as in more developed regions, where the figure was 1.37%.

As the world experienced rapid urbanization, cities themselves continue to grow rapidly, as measured com monly by the size of the population but lately by other criteria as well. Giant cities have appeared, but 'mega cities' is the preferred term to describe them because the United Nations has applied a quantitative description of these cities. Mega cities have been shown to have de veloped in tandem with the region within which they are situated, hence the emergence of mega urban regions. Mega cities are also world cities or global cities if they perform vital functions for the global economy or in other domains. With globalization and urbanization expected to accelerate further, mega cities, especially in developing countries, are confronted with many challenges.

The Rise of Mega-Cities

As urbanization gathered speed during the twentieth century, academics attempted to track the growth of large cities. At first, the aim was to measure a city of notable demographic weight when it reached the size of 1 million or more inhabitants. In 1950, there were only 78 such cities, but in 2000 the figure had grown to 400, a more than fivefold increase within half a century. By 2010, there are expected to be 511 cities with a population of at least 1 million.

Since the 1980s, some cities have grown to become so large that a different concept had to be devised to track their growth. The United Nations uses the figure of 10 million, upgraded from the initial threshold of 8 million, to designate a new class of very large cities – mega cities. The rapid rise of mega cities is clear from the fact that in

1950, there were only two such cities (New York and Tokyo), but in 2005 the number had mushroomed to 20, accounting for 9% of the world's urban population. Since 2005, two more are on their way to that size category. By 2015, it is projected that there will be 22 mega cities in the world, of which 17 will be in developing countries (**Table 1**).

In fact, mega cities have been growing more slowly than the world's urban population as a whole, which averaged an annual growth rate of 2.4% in the period 1975–2005. Only Dhaka, Lagos, Delhi, Karachi, Jakarta, Mumbai (Bombay), and Manila grew more quickly. Between 2005 and 2015, only six of the expected 22 mega cities are projected to grow more than the estimated average annual growth rate of the world's urban population of 1.9%. These cities are Lagos, Dhaka, Karachi, Jakarta, Guangzhou, and Delhi. It is clear from these lists that Asia and, to a lesser extent, Africa are the two continents that will witness a continued rapid growth of mega cities. Yet in 2005, these continents were the world's least urbanized, at 40% for Asia and 38% for Africa. They are expected to reach a level of urbanization of 54% and 51%, respectively, in 2030. From 1990 to 2005, African mega cities topped the world in terms of rate of growth (**Table 1**).

Table 1 also shows another spatial shift of mega cities over time. Since 1950, the proportion of mega cities in developed countries relative to those in developing countries has declined relentlessly. By 2005, the dominance of developing countries in the realm of mega cities had become indisputable. In the developing world, the rise of mega cities in Asia is especially striking: 16 mega cities

Table 1 The world's mega-cities, 1950 2015 (the 30 largest urban agglomerations)

Mega-city	*Population (in millions)*				*Average annual rate of growth 1990 2005 (%)*		*Percentage of the country's total urban population 2005 (%)*
	1950	*1990*	*2005*	*2015*			
Africa						3.63	
Cairo	2.49	9.06	11.13	13.14	1.38		35.15
Lagos	0.29	4.76	10.89	16.14	5.67		17.19
Latin America						2.17	
Mexico City	2.88	15.31	19.41	21.57	1.59		23.87
Sao Paulo	2.33	14.78	18.33	20.54	1.45		11.67
Buenos Aires	5.10	10.51	12.55	13.40	1.19		35.95
Rio de Janeiro	2.95	9.60	11.47	12.77	1.19		7.31
Santa Fe de Bogota	0.68	4.91	7.75	8.93	3.09		23.38
Lima	0.97	5.83	7.19	8.03	1.41		35.40
Asia						2.90	
Mumbai (Bombay)	2.86	12.31	18.20	21.87	2.64		5.74
Delhi	1.37	8.21	15.05	18.60	4.12		4.75
Shanghai	6.07	8.21	14.50	17.23	3.86		2.73
Kolkata (Calcutta)	4.51	10.89	14.28	16.98	1.82		4.51
Jakarta	1.45	7.65	13.22	16.82	3.71		12.33
Dhaka	0.42	6.53	12.43	16.84	4.38		34.96
Karachi	1.05	7.15	11.61	15.16	3.28		21.09
Beijing	4.33	7.36	10.72	12.85	2.54		2.02
Manila	1.54	7.97	10.69	12.92	1.98		20.53
Istanbul	0.97	6.55	9.71	11.21	2.66		19.72
Seoul	1.02	10.54	9.65	9.55	– 0.59		24.98
Guangzhou	1.49	3.92	8.43	10.42	5.24		1.59
Tehran	1.04	6.37	7.31	8.43	0.92		15.71
Shenzhen	0.17	0.88	7.23	8.96	15.07		1.36
More-developed regions						0.63	
Tokyo	11.28	32.53	35.20	35.49	0.53		41.76
New York-Newark	12.34	16.09	18.72	19.88	1.01		7.77
Los Angeles	4.05	10.88	12.30	13.10	0.82		5.11
Osaka-Kobe	4.15	11.04	11.27	11.31	0.14		13.37
Moscow	5.36	9.05	10.65	11.02	1.09		10.19
Paris	5.42	9.33	9.82	9.86	0.34		21.16
Chicago	5.00	7.37	8.81	9.47	1.20		3.66
London	8.36	7.65	8.51	8.62	0.71		15.90

Source: Data from United Nations, Population Division (2006). World urbanization prospects: The 2005 revisions, pp 46 55 and 126 129. *Report Of United Nations*, *Population Division*. New York: United Nations.

appeared in Asia in 2005. The growth of Shenzhen from a tiny border town of 0.33 million inhabitants in 1978 to a major city of 7.23 million in 2005 has been phenomenal, with an average annual growth rate of 15.07% between 1990 and 2005. **Figure 1** shows the spread of mega cities across the world, with a very clear concentration of them in the 'south'.

As **Table 1** also reveals, in 2005, some cities had a population of close to 20 million. By 2015, there are expected to be more such cities. A new term – meta city – has been coined to denote this class of 'hyper city'. It is a new type of human settlement above and beyond the scale of the mega city, spurred by economic develop ment and increases in population. A meta city gradually swallows up rural areas, cities, and towns, to become a multinucleated urban form with one identity. The rise of meta cities will pose new challenges for humankind. Such cities are so gigantic in size that managing them and serving their inhabitants is a daunting task in terms of the resources, skills, and technology required. Meta cities are larger in population than entire countries. To wit, the population of Greater Mumbai exceeds that of Norway and Sweden combined. Yet, the world's largest cities account for only 4% of the world's total population. In 2005, Tokyo remained the world's largest meta city, with a population of 35.20 million. By 2015, Mexico City, Sao Paulo, and Mumbai are anticipated to join the league of meta cities. They will be further joined in 2020 by Delhi, New York, Dhaka, Jakarta, and Lagos.

Mega-Urban Regions

Mega cities have grown so rapidly that their economic and other activities are prone to spill over their tradi tional boundaries. Given the cross border nature of many global economic activities, the concept of the traditional city is no longer a realistic one for measuring the population of a very large city and their activities. New concepts and statistics gathering frameworks have to be designed to capture the reality of such cities.

Foremost among such attempts are studies that have focused on mega cities in the Asia Pacific region, which are acting as the nucleus of a large range of economic, planning, and social activities that can extend to as far as 100 km from the city's center. Between the city and the surrounding countryside, there is a great deal of inter action, which is constantly changing in mode and in tensity. The peri urban zone immediately surrounding the city is an area of maximum rural–urban interaction. McGee pioneered the concept of the extended metro politan region (EMR) and, along with his associates, has applied it, with satisfactory results, to many countries in the Asia Pacific region. Jakarta, Bangkok, Manila, Ho Chi Minh City, Hong Kong–Guangzhou, Shanghai, and Beijing have been found to support this formulation in the empirical studies that have been undertaken.

It has been observed that the expansion and merging of highly urbanized zones remains an important trend in many parts of the world. The continued coalescence of Boston–New York–Philadelphia–Baltimore–Washington (Jean Gottmann's megalopolis), Los Angeles–San Diego, Chicago–Milwaukee, Tokyo–Nagoya–Osaka–Kobe–Kyoto, Amsterdam–the Hague–Rotterdam (Rand stad, Holland), Hong Kong–Shenzhen–Guangzhou, and Shanghai–Hangzhou–Suzhou are examples. By extending their reach and resources, they will increase their influ ence and imprint on the global economy. In the 1990s, Canada saw the forced amalgamation of several municipal entities in the provinces of Ontario and Quebec to form larger and new municipalities. Toronto, Ottawa, and Montreal have reorganized and expanded into new en tities, but they are still a long way from becoming true mega cities, notwithstanding the mislabeling of the urban expansion process by the local media.

Looked at from a broader spatial scale, mega cities often form units in a wider and larger regional grouping. Within the Asia Pacific region and, to a certain extent, in Europe, subregional cooperative economic zones in volving several countries have been formed. Known in Asia as growth triangles, the best and most successful examples are the Southern China Growth Triangle, with the participation of Hong Kong, Taiwan, and Guang dong/Fujian; and the SIJORI Growth Triangle, involving Singapore, Johore (southern Malaysia), and Riau Islands (Indonesia). The likening of the urban system in Europe to grapes or bananas similarly emphasizes the inter relationships of mega cities in different regional settings. In the same vein, the term, 'urban corridor', is used to refer to the interconnection between mega cities that has been observed to occur in many regions of the world. Within the Asian context, the formation of the BESETO (Beijing–Seoul–Tokyo) urban corridor is most striking and advanced, but a continuous corridor along the Western Pacific Rim from Tokyo along the coastal area southward to Java in Indonesia is emerging. Given the recent strong economic growth of China and the ASEAN (Association of Southeast Asian Nations) countries, this continental urban belt may well become real in a few decades.

From Mega-City to World City

Mega cities and world cities began to gain prominence in the latter part of the twentieth century, but for different reasons. Mega cities are important by virtue of the size of their population, and because of their concentration of economic, political, and technological power. World cities, on the other hand, are famous for certain critical

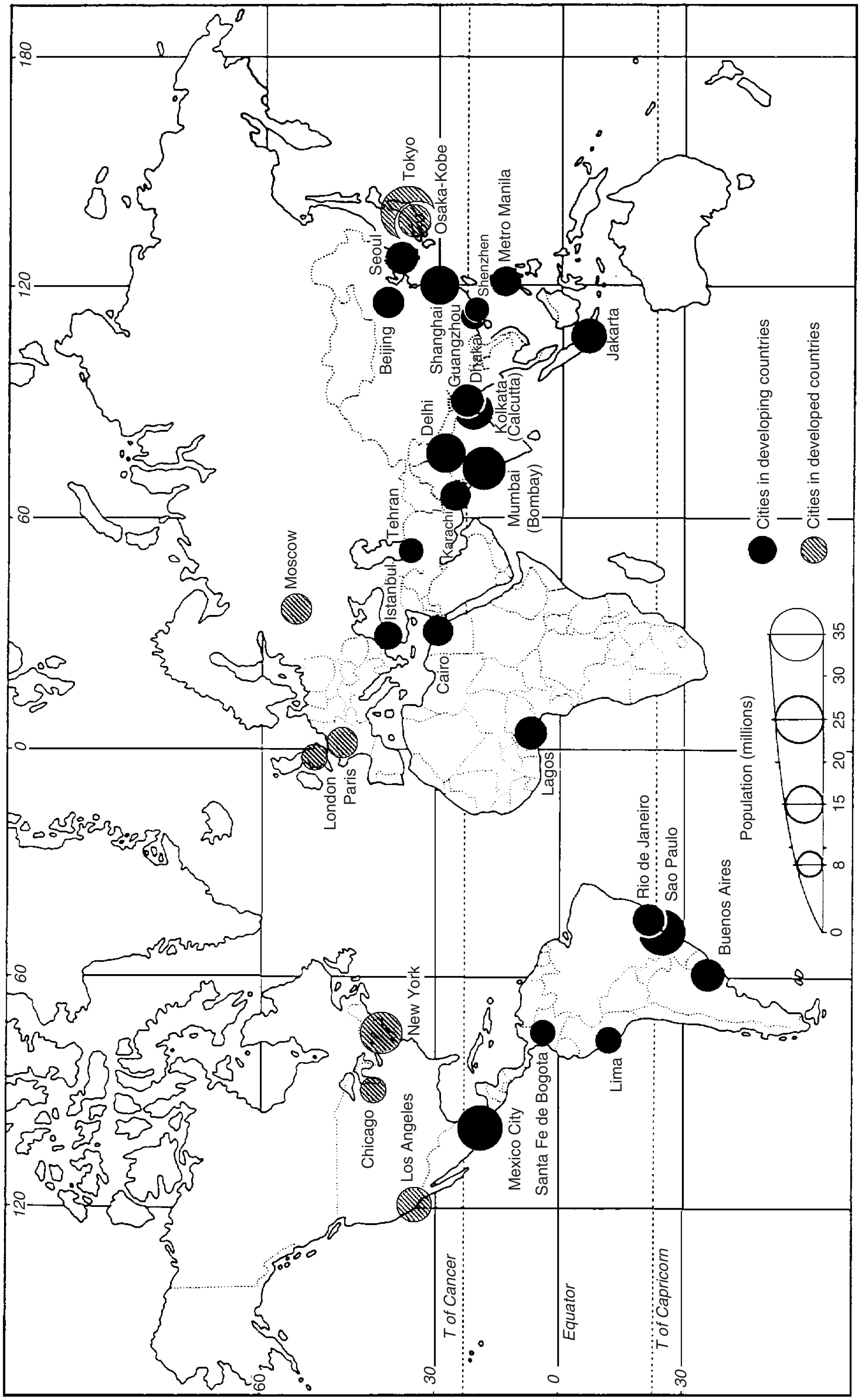

Figure 1 Mega-c t es of the wor d, *c.* 2005 (the 30 argest urban agg omerat on).

functions that they perform in the global economy, or for the worldwide impact that they have in a noneconomic sense. Whenever the lists of these two types of cities are compared, some overlap is observed but they tend to encompass different cities.

World cities, or global cities, are a class of particu larized cities that drive and thrive on processes of glob alization. They are command posts in the organization of the global economy, key locations for finance and spe cialized service firms that have replaced manufacturing as the leading economic sector in that city, sites of pro duction and innovation in leading industries, and markets for products and innovations that have been produced. These multiple functions of world cities connect them with other world cities in the region and in other parts of the world, in a functional and effective network. World cities are part of the global economy, defined as one where capital flows, labor markets, commodity markets, information, raw materials, management, and organizat ions are internationally and fully interdependent through out the world, even on a daily basis.

The new global economy has impacted cities in de veloped and developing countries in different ways. In industrially advanced economies, the loss of manu facturing jobs means that cities must restructure their economies in order to survive, as shown by the experi ence of Detroit, Chicago, New York, Manchester, and Osaka. Equally, other cities such as Los Angeles, Miami, Berlin, and Vienna have discovered new roles and are thriving in the new global environment. More specific ally, Berlin has assumed the new role, and Vienna has recaptured its old one, of international business center for the central European region. New York, London, and Tokyo remain the world's leading financial centers, but the recent rise of Hong Kong, Mumbai, and Shanghai is noteworthy. The Hong Kong stock market has climbed to new heights on the back of the endeavor of Chinese firms to 'go out' through a listing in Hong Kong. The volume of initial public offerings (IPOs) in Hong Kong in 2006 exceeded that of New York.

The rise of world cities in the developing world is, in part, a result of the fact that the new global economy is being erected on a new international division of labor (NIDL). Cities in developing countries are becoming centers of productive activities and, consequently, thea ters of capital accumulation. NIDL has been premised on four factors: a strong desire among firms to utilize less expensive sources of labor and more profitable locations for production; the response of firms to the growing bargaining power of certain developing countries in hosting industrialization; the expansion abroad of firms from developed economies in order to reduce costs and meet growing international competition; and the re location overseas of firms or segments of their operations to sidestep well organized labor and government regulations. As a result of the above four developments, the relocation of firms from developed to developing countries over the past few decades has been sustained and massive. One indication of this global shift is the growing and large scale phenomenon of foreign direct investment (FDI) in developing countries over the past few decades, focused sharply on their cities. Over the past decade, China has emerged as a top destination for FDI, part of the remarkable economic transition that has been taking place in that country since the 1980s.

Under the new global economy, competition among cities has intensified. Many research institutes have been tracking urban competitiveness, regionally or globally, on an annual basis. Indeed, many cities have been engaged in improving themselves, striving to achieve the status of a world city if they are not yet there, in a process that may be called world city formation. Their efforts are going into building a strong physical and social infrastructure. For example, erecting architecturally distinctive office buildings conveys an image of power and prestige; and constructing major international airports, superfast trains, and information highways will sustain the global reach of a world city. Many Asian cities such as Osaka, Hong Kong, Seoul, Kuala Lumpur, Shanghai, and Bangkok, have recently built mammoth, state of the art airports. Similarly, new container ports have been built in Shanghai, Shenzhen, and Tanjung Pelepas (in southern Malaysia). Fast trains now connect Taipei and Kaoshiung, Seoul and Pusan, and Shanghai and Beijing (under con struction). Many of these Asian cities are preparing themselves to become world cities, if they are not yet in that class.

Cities also compete in their entrepreneurial skills through 'place making'. They are engaged in a com petition to become internationally famous in order to improve their chances of attracting investment, tourists, and other resources. The right to host the Olympic Games, Asian Games, World Cup, World Expo, and other high profile events is fiercely contested. The positive effects of staging these world events have indeed pro pelled some cities to new heights of economic and social development, as hosting the Olympics has done for Tokyo, Seoul, Los Angeles, Sydney, and Athens. At an other level, Disneyland in Paris, Tokyo, and Hong Kong, and Universal Studios in Osaka, are examples of successful investments in mass entertainment, capital izing on the global brand names of the institutions in question.

Planning and Managing Mega-Cities

With the increasing concentration of mega cities in de veloping countries, the challenges of planning and managing them are becoming ever more daunting because these countries often lack the necessary resources, expertise, and experience to organize and run such cities. The magnitude and scope of the pertinent problems are unprecedented in the history of humankind.

In the 1990s, the Population Policy Section of the Population Division of the United Nations supported a mega city research project involving 20 such cities in developing countries. Several of its important findings may be summarized here. First, it was found that although big cities have some advantages that could make them attractive to certain types of industries, the empirical evidence for their overall economic attractiveness was inconclusive. However, most mega cities perform national and/or international economic functions, such as government, banking, insurance, and finance, depending on their relative connectivity to the global economy. Manufacturing employment figured prominently in Asian mega cities, ranging from 23% in Manila and Dhaka, to 35–40% in Mumbai, Kolkata (Calcutta), and Seoul. The informal sector is a key contributor to urban employment, providing 7 out of 10 jobs in Latin America and the Caribbean.

Second, mega cities recognize the need to adopt spatial strategies that would allow them to adjust to their growing size and functions. Many are striving to achieve a polycentric structure by promoting subcenters. Nevertheless, Cairo, Jakarta, Dhaka, and Chennai (Madras) have not succeeded thus far. Only Seoul appears to have achieved a measure of control over growth and some directional strategy by adopting physical measures, such as new towns, industrial estates, and green belts, complemented by social and tax policies.

Third, these and other studies have reaffirmed a widespread and severe lack of basic urban services in many mega cities, including a chronic shortage of housing, inadequacies in both the quantity and quality of the water supply, antiquated sewerage facilities, poor solid waste disposal, and polluted air and water. Squatter settlements and slums have proliferated. Some 40% of the inhabitants of Mexico City live in such conditions, while in Bogota the figure is 50% and in Manila 32%.

Other recent inquiries have pointed to an increasing economic and social polarization within mega cities, in part the result of accelerating processes of globalization. Urban poverty is grave and entrenched, a condition often viewed as an affront to human progress in an age when technology and globalization have brought unprecedented wealth and achievements. Crime and a deteriorating social order have emerged as societal problems, especially in mega cities in Africa, which also have to combat the endemic health problem of HIV/AIDS. Terrorism has become a new urban pathology that is plaguing mega cities, notably in the developed world, as a consequence of the terrorist attacks on New York and Washington, DC in September 2001.

In view of the plight of urban areas in the developing world, the United Nations and its relevant agencies have lent resources, commitment, and energy to understanding and are trying to ameliorate urban problems. These agencies, such as the World Bank, United Nations Development Program, UN HABITAT, and Urban Management Program; and other agencies such as Canada's International Development Research Centre and the Ford Foundation have been particularly active since the 1980s, when the problems of the mega cities in developing countries came to the fore. Much of the work that has been accomplished thus far has proceeded in the spirit of 'learning by doing', as there were no precedents from which to draw lessons. A guiding philosophy for much of the effort has been the firm belief that mega cities in developing countries have much to learn from one another. Thus, the approach of sharing experiences by building a network was followed, with the construction of a new platform on which all actors could play out their parts meaningfully and purposefully. The UN Conference on Human Settlements held in Vancouver in 1976 was the first such global event that brought politicians, planners, practitioners, academics, activists, media representatives, and others to the rare opportunity of comparing experiences and charting the way forward. This was followed by the City Summit held in Istanbul in 1996 and by Habitat III (World Urban Forum) in 2006, again in Vancouver. All of this underlined the commonality of the urban experience across the globe and the participants found much to learn from one another.

With the collective efforts that have been devoted to combating urban problems in developing countries, mega cities have now become cities of hope. Recently, both international organizations and scholars conducting research on mega cities, have been emphasizing issues of governance. Many governments across the developing world have collapsed because of fast paced political changes. These changes in government are a sign of the increasing aspirations of people to participate in governance, and an expression of greater demands for democracy and for transparency in public policy. In this respect, the role of nongovernmental organizations (NGOs) has been critical and civil society is emerging as a force recognized by the authorities and international agencies. Yet, cross border crime, massive speculative financial flows, and unwelcome cultural influences are some examples of what urban governments of today have to deal with on a regular basis. In this sense, urban governance is occasionally in a state of siege.

The mega cities of today are witnessing fundamental changes in the way they are governed, in how new technologies are changing urban life, and in how tolerable they are to live in.

To begin with, one must note the growing application of new technologies in the planning of urban areas.

Public–private partnerships and competition are often involved in the process. Malaysia's Multimedia Super Corridor, which stretches from Kuala Lumpur's Petronas Twin Towers to the International Airport and includes Putrajaya, the administrative capital, and Cyberjaya, the hi tech park, is a classic example of combining planning principles with new technologies.

Information technology (IT) is increasingly a part of urban life, as the Internet has become an indispensable tool in business transactions, the dissemination of infor mation, the gathering of data, the provision of enter tainment, and so on. Urban societies are fast becoming networked societies, with flows of ideas, money, and in formation traveling almost instantaneously within a city and across the world. The rise of the networked trans national corporation is an especially significant phe nomenon, not only in developed economies but recently also in newly developed economies such as South Korea, Singapore, Hong Kong, Taiwan, Brazil, and India.

Future Challenges

The twenty first century is an urban century in that, for the first time, any time after 2007 there will be more people living in urban than in rural areas. In other words, the human population has crossed the rural–urban divide and the urban population has exceeded the rural one. Mega cities in developing countries will be a major force in directing global processes and shaping the global economy.

Yet, the challenges confronting mega cities are im mense. Mega cities are operating in a fast paced atmos phere of global change, in which people, capital, goods, and ideas are traveling at speeds never seen before. Massive and speedy monetary flows, triggered by com puter registered transactions and speculative invest ments, contributed to the global stock market collapse in 1987 and to the Asian financial crisis in 1997. This is one downside of globalization. Globalization is also exerting a marginalizing influence between cities and nations, as well as within cities. In many cities and nations, urban poverty remains a daunting problem amid growing prosperity. The divide between the haves and the have nots has grown within cities, between nations, and across the globe.

There will be more freedoms and more constraints in the mega cities of the future. Greater freedoms will be extended to individuals and institutions because they will be networked electronically. Wired interactions will supplement face to face contacts. This will also affect urban lifestyles, as people can work at home, shop by computer, and travel using smart cards. The well heeled are likely to have property and business interests in several cities or countries, and families to be similarly dispersed, out of preference. Citizens of these cities will be increasingly prone to 'think globally and act locally'. Indeed, social organizations and political institutions face the challenge of having to capture the loyalties of people who have highly diverse orientations. In such a fluid and shifting milieu, urban governance will become ever more challenging, as apart from having to serve local con stituents, urban administrations must deal with, and share power with, multinational corporations that are becoming larger and more powerful. Urban governments need to devolve power to local units and adopt a flexible approach to involve talented people in a civil society.

Above all, the mega cities of the future will have to improve their livability, in the areas of health, education, the environment, and, more generally, the quality of life. All survey data to date show that human development or similar indicators are positively skewed toward cities in developed countries. Mega cities in developing countries tend to score fairly or poorly in these indicators. The question of the livability of mega cities should be raised alongside that of their economic, environmental, and cultural sustainability. It will be incumbent upon these urban governments to rise to these challenges.

See also: Planning, Urban; Third World Cities; World/ Global Cities.

Further Reading

Brennan, E. and Richardon, H. W. (1989). Asian megacity characteristics, problems, and policies. *International Regional Science Review* 12(2), 117 129.

Doggan, M. and Kasarda, J. D. (eds.) (1988). *The Metropolis Era: A World of Giant Cities*. Newbury Park, CA: Sage Publications.

Fuchs, R. J., Brennan, E., Chamie, J., Lo, F. C. and Uitto, J. I. (eds.) (1994). *Mega Cities Growth and the Future*. Tokyo: United Nations University Press.

Laquian, A. A. (2005). *Beyond Metropolis: The Planning and Governance of Asia's Mega Urban Regions*. Washington, DC: Woodrow Wilson Center Press.

Lo, F. C. and Yeung, Y. M. (eds.) (1996). *Emerging World Cities in Pacific Asia*. Tokyo: United Nations University Press.

Lo, F. C. and Yeung, Y. M. (1998). *Globalization and the World of Large Cities*. Tokyo: United Nations University Press.

McGee, T. G. (1991). The emergence of Desakota regions in Asia: Expanding a hypothesis. In Ginsburg, N., Koppel, B. & McGee, T. G. (eds.) *The Extended Metropolis: Settlement Transition in Asia*, pp 2 25. Honolulu, HI: University of Hawaii Press.

Montgomery, M. R., Stren, R., Cohen, B. and Reed, H. E. (eds.) (2003). *Cities Transformed: Demographic Change and Its Implications in the Developing World*. Washington, DC: National Academies Press.

United Nations Centre for Human Settlement (2001). *Cities in a Globalizing World*. London: Earthscan Publications.

UN HABITAT (2006). The state of the world's cities report 2006/2007. *Report of UN HABITAT*. Nairobi: UN HABITAT.

United Nations, Population Division (2006). World urbanization prospects: The 2005 revisions, pp 46 55 and 126 129. *Report Of United Nations*, *Population Division*. New York: United Nations.

Yeung, Y. M. (1997). Geography in the age of mega cities. *International Social Science Journal* 159(151), 91 104.

Yeung, Y. M. (2002). Asia Pacific urbanism under globalization. In Yeung, Y. M. (ed.) *New Challenges for Development and*

Modernization: Hong Kong and the Asia Pacific Region in the New Millennium, pp 275 293. Hong Kong: Chinese University Press.
Yeung, Y. M. (2004). *Enhancing the Livability of Asian Mega Cities*, 30 pp Occasional Paper No. 148, Hong Kong Institute of Asia Pacific Studies, The Chinese University of Hong Kong.

Relevant Websites

http://www.megacities.nl
Megacities.
http://magma.nationalgeographic.com
Megacities @ National Geographic Magazine.
http://www.megacities.uni koeln.de
Mega City Taskforce of the International Geographical Union (IGU).
http://www.radicalcartography.net
Radical Cartography: Maps of US Megacities.
http://www.megacitiesproject.org
The Mega Cities Project.
http://www.un.org
UN Department of Economic and Social Affairs, Population Division.

Meinig, D.

K. M. Dunn, School of Social Sciences, Penrith, NSW, Australia

Meinig, Donald (1924–)

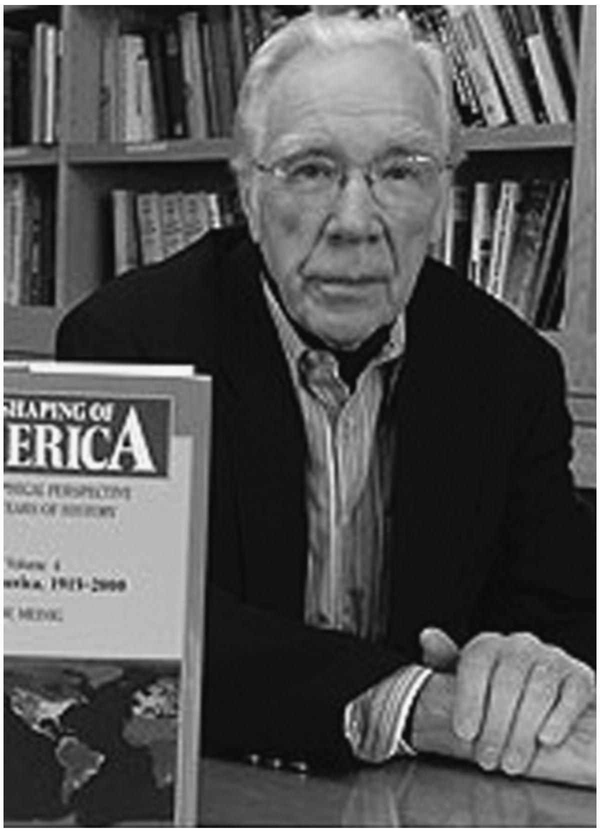

Donald William Meinig's texts demonstrate his com prehensive understanding of social processes, economic development, and biophysical conditions. They also betray his family background, born on a farmstead on the eastern edge of Palouse county, a famous grain area in Washington county where Meinig's geographical imagination was ignited by his youthful explorations of its human and physical topographies. Plans to go into the US foreign service were abandoned, but his geo graphical studies were interrupted by wartime and a period in the US Engineering Core. His gradual ex posure to UK, and later Australian, historical geography (through a Fulbright fellowship) was to extend his geographical horizons.

Meinig (who acquired his PhD at Washington, 1953) worked for 10 years at the University of Utah (US). Since 1959, he served at Syracuse University, where the Department of Geography evolved to become a prominent global node of geographical scholarship. During 1958, he spent a year in the Department of Geography at The University of Adelaide, Australia. His research on the South Australian wheat frontier became the basis of his 1962 book *On the Margins of the Good Earth*. Meinig traversed several thousand miles, visiting every township in the region, as well as exhaustively using archives (State Parliamentary records and muni cipal newspapers for the entire era). Field observations and archive mining were the career mainstays of Meinig's research method. Short hand descriptions of his basic approach include immersion and study from the bottom up. The general intent of his work was to encourage stronger appreciations of place and a clearer understanding of its dynamic nature.

Meinig's analysis of the colonization of the South Australian wheat belt, north west from Adelaide in the 1870s and 1880s, revealed his subtle ability to describe the links between agricultural development, transport, settlement, and biophysical conditions. The South Aus tralian Government passed laws preventing farming beyond the Goyder's rainfall line, where rainfall was unreliable for farming, and the vegetation only appro priate for pastoral leaseholds. Meinig's analysis of local newspaper editorials and parliamentary debates traced how good rains and bumper crops had generated pressure on the government to expand the farming frontier. Goyder's observations were publicly ridiculed, and the theory that 'rains will always follow the plough' gained a popular following. Political pressure saw Goyder's line erased, sheep pastoralists had their lands resumed, and the wheat farming frontier expanded in a land scramble. For 2 years there were great successes among the wheat farmers. But 3 years of drought produced wide scale abandonment in the 'marginal lands' north of Goyder's line. Bio physical limits were a continued theme in Meinig's 1971 text *Southwest,* where he noted the tem pering effects of natural hazards like drought on over development within fragile zones. He remarked on how a sudden influx of European agriculturalists would be followed by a straggling exodus, a quick testing of the land's productivity by ordinary people, often at great personal cost.

Meinig's work never shied from the political tensions associated with colonization. This was most engagingly demonstrated in his book *Southwest,* in which he described the fortunes of three different peoples in the southwest of the United States. The three were: Indigenous Americans (with their stubborn occupation and cultural mainten ance); Hispanic Americans (their largely unrecorded cultural imprint and unrecognised 'traditional' occu pancy), and the 500 year European presence (with the dramatic changes in resource use and the environment). Meinig's historical geography has a rigor and honesty that proscribes a whitewash. He commented on how expansions of Hispanic and Anglo Americans were often at the cost of Indigenous people, confining the latter to reservations and inhospitable lands. In 1968, he described the Anglo settlement in the Great Columbian Plain as a white invasion. He correctly characterized the cross cultural encounters in America as "harsh, and often bloody". Surprisingly, his Australian work is silent on the presence and history of Indigenous people. Later, he would preface the first of his four volume description of the *Shaping of America* with a justification of his unveiling of the ugly sides of colonization.

Meinig lamented the lack of a comprehensive de scription and analysis of American national history, and his work addressed precisely that gap. The catalyst for Meinig's book *The Great Columbia Plain* was his own life history in the Palouse region, followed by his graduate research on the Walla area. The *Columbia Plain* and *Southwest*, along with his book on Texas, and his esteemed paper on the Mormon history and presence (the most cited of his publications), comprised the springboard for his ambitious four volume historical geography of the United States. These volumes covered first Atlantic America (1986), then Continental America (1992), Transcontinental America (1995), and finally Global America (2004). One measure of his contribution in this field is the multiple citations of his work in *A Scholars Guide to Geographical Writing on the American and Canadian Past*. In the Preface to *Southwest* (1971, page vi) Andrew Clark described Donald Meinig as "one of the most evocative interpreters of the evolution of the American West."

In a recent book titled *Landscapes* there is a comment on the timeliness of Meinig's recommendations for studying landscapes. Meinig advocated the study of ordinary landscapes and he outlined how readings of landscapes are multiplicitous. The former recom mendation was to eschew a pre occupation with exotica, and to repatriate our research gaze to the landscapes and people around us. The second was a harbinger of a fundamental principle of post structuralism – that there are many truths, and these truth claims are in a contest. Meinig said "It will soon be apparent that even though we gather together and look in the same direction at the same instant, we will not – we cannot – see the same landscape". Despite this multiplicity, Meinig did not abandon students to chaos and cacophony. He saw landscapes as texts to be read in a rigorous way. Meinig articulated seven principles for historical geography (context, holism, localism, systems, tensions, dynamism, and growth). And he edited the text in which Lewis' seven axioms for landscape reading were presented. In *The Great Columbia Plain*, Meinig exhaustively detailed the sequences of human occupancy in that remarkable region of Washington State. He described the economies and cultures of Indigenous people (such as the Spokane, Nez Perce, and Yakima peoples) as well as the diverse re ligious and cultural histories of white fur traders and missionaries from Canada and the eastern US. The varied and competing political economies of hunter gatherers, fur traders, military administrations, gold miners, pas toralists, and wheat farmers were subtly described. It is an excellent guide to all researchers on how to carry out a study of sequent occupation.

Donald Meinig has an exceptional talent for the powerful telling of stories about landscapes and peoples. He has been a supporter of more personalized, honest, and creative renderings of geography, outlined most fully in his address to the Institute of British Geog raphers in 1983, which is widely cited by geographers undertaking studies of fiction, reflexive analyses, and using situated knowledge. Depth, honesty, openness, and rigor were markers of the way a geographer ought to read landscapes. This is indeed how Donald Meinig interpreted the landscapes of the United States and of Australia.

See also: Cultural Geography; Historical Geography; Humanism/Humanistic Geography; Indigenous Geographies; Landscape.

Further Reading

Conzen, M. P., Rumney, T. A. and Wynn, G. (1993). *A Scholar's Guide to Geographical Writing on the American Past*. Chicago, IL: University of Chicago Press.

Meinig, D. W. (1962). *On the Margins of the Good Earth: The South Australian Wheat Frontier 1869 1884. Monograph Series, No. 2*. Chicago, IL: Association of American Geographers.

Meinig, D. W. (1965). The Mormon culture region: Strategies and patterns in the geography of the American west, 1847 1964. *Annals of the Association of American Geographers* 55(2), 191 220.

Meinig, D. W. (1968). *The Great Columbia Plain: A historical geography, 1805 1910*. Seattle, WA: University of Washington Press.

Meinig, D. W. (1969). *Imperial Texas: An interpretative Essay in Cultural Geography*. Austin, TX: University of Texas Press.

Meinig, D. W. (1971). *Southwest: Three peoples in Geographical Change 1600 1970*. New York: Oxford University Press.

Meinig, D. W. (ed.) (1979). *The Interpretation of Ordinary Landscapes: Geographical Essays*. New York: Oxford University Press.

Meinig, D. W. (1983). Geography as an art. *Transactions of the Institute of British Geographers* 8(3), 314 328.

Meinig D. W. (1986). *The Shaping of America: A Geographical Perspective on 500 Years of History, Volume 1, Atlantic America, 1492 1800*. New Haven: Yale University Press.

Meinig, D. W. (1992). *The Shaping of America: A Geographical Perspective on 500 Years of History, Volume 2, Continental America, 1800 1867*. New Haven: Yale University Press.

Meinig, D. W. (1995). *The Shaping of America: A Geographical Perspective on 500 Years of History, Volume 3: Transcontinental America, 1850 1915*. New Haven: Yale University Press.
Meinig, D. W. (2004). *The Shaping of America: A Geographical Perspective on 500 Years of History, Volume 4: Global America, 1915 2000*. New Haven: Yale University Press.
Winchester, H. P. W., Kong, L. and Dunn, K. M. (2003). *Landscapes: Ways of Imagining the World*, Ch 2. London: Pearson Education.

Relevant Websites

http://www.maxwell.syr.edu/geo/faculty current/meinig.htm
Donald Meinig's Homepage.

Memorials and Monuments

D. H. Alderman, East Carolina University, Greenville, NC, USA
O. J. Dwyer, Indiana University, Indianapolis, IN, USA

Glossary

Counter hegemonic monument A monument created to challenge a dominant interpretation of the past, demonstrating how memorials can be used as tools of social resistance.
Mythologize To create a myth, not necessarily in the sense of fabricating history but in converting memory into a widely shared and socially important identification with the past.
Sanctification The creation of a sacred place for remembering an event, person, or group, accompanied by the building of memorials and monuments.
Symbolic Accretion The appending of supportive or antagonistic commemorative elements onto existing memorials and monuments.
Symbolic Annihilation The manner in which memorials and monuments fail to represent (or misrepresent) the identities and histories of marginalized social groups.
Symbolic Capital The accumulation of prestige, honor, and influence through the manipulation and control of symbols and cultural representations, including memorials and monuments.

Introduction

Memorials and monuments are of increasing interest to geographers, growing out of recognition of the social nature of commemoration and the important role that space and place play in the process and politics of remembering. Geographers envision these public symbols as part of larger cultural landscapes that not only reflect certain perspectives on the past but also work to legitimate them as part of the normative social order. Geographers analyze memorials and monuments through three conceptual lenses. The 'text' metaphor emphasizes a critical reading of the histories and ideologies given voice (and silenced) in the content and form of memorials and monuments, as well as the dynamic nature of (re)inscribing memory into space. The 'arena' metaphor focuses on the capacity of memorials and monuments to serve as sites for social groups to debate actively the meaning of history and compete for control over the commemorative process as part of larger struggles over identity. The 'performance' metaphor recognizes the important role that bodily enactments, commemorative rituals, and cultural displays occupy in constituting and bringing meaning to memorials and monuments, suggesting that the body itself constitutes a place of memory.

Social and Geographic Nature of Commemoration

The study of memorials and monuments is of increasing interest to geographers as well as scholars from a wide range of fields in the humanities and social sciences. Informing much of this interest is an understanding that memory and its manifestations cannot be fully explained in terms of cognitive psychology. While the mechanics of remembering and forgetting is a matter of brain chemistry and individual predilection, what comes to count as being worthy of commemoration is mediated within society. Many of our recollections of the past are collective memories that are constructed and transmitted to us through group interactions and a variety of cultural practices. What is commemorated is not synonymous with all that has gone on in the past. Rather, what is defined as memorable or historically significant is open to social control, negotiation, and contestation.

The social or collective interpretation of the past is constituted, in part, through the construction of material sites of memory, which include memorials and monuments. The conventional distinction between monuments and memorials – the former supposedly characterized by triumph and the latter embodying loss – is somewhat unhelpful. Monuments are one kind of memorial text, taking their place alongside a wide range of media designed to facilitate remembering and forgetting of the past. Typically situated in public space, monuments include a host of material culture elements associated with collective memory, for example, street signs, historical markers, landmarks, statuary, preserved sites, and parks. Together, they constitute what may be termed the memorial or heritage landscape.

Memorials and monuments are important symbolic conduits for not just expressing certain versions of history but casting legitimacy upon them. They give the past a tangibility and familiarity – making the history they commemorate appear to be part of the natural and taken for granted order of things. Memorials and monuments wield further social influence by serving as tourist destinations, civic gathering places, and settings for everyday activity. The power of commemorative place naming, for example, comes from the manner in which history is inscribed into our daily vocabulary, both verbal and visual. Appearing on road signs, addresses, advertising billboards, and maps, the past is constantly made part of one's spatial and historical frame of reference. Like all cultural landscapes, memorials and monuments have a normative power, at once reflecting and reproducing social ideas about the past.

Memorials and monuments influence how people remember and value the past, in part, because of their apparent permanence and the common impression that they are impartial recorders of history. In reality, memorial landscapes narrate history in selective and controlled ways – hiding as much as they reveal. Con sequently, the social process of remembering is accom panied, simultaneously, by a process of forgetting – an excluding of other historical narratives from public consideration and recognition. Because memorials and monuments often reflect the values and worldviews of the dominant social class, they tend to exclude the histories of minority or subaltern groups. Although not usually detectable from the surface, memorials and monuments hold deeper stories about how they were created, by whom, and for what ideological purpose.

Memorials and monuments obviously represent his tory, but it is wrong to see them as completely couched in the past. They are also mirrors of more contemporary events, issues, and tensions. Rather than having a fixed, static meaning, memorials and monuments are in a constant process of becoming as present social needs and ideological interests change. They fall in and out of public favor as opinions about the past shift. The re writing of history is often carried out through a re scripting of the identity and meaning of places of memory. While it is true that much of this rewriting is done by those in power, it is also done by everyday people as they seek to turn the memorial landscape into a site for struggle and resistance. There is increasing evidence of traditionally marginalized groups building counter hegemonic monuments that challenge the dominant historical narratives that frequently exclude them. Despite the impression that memorial landscapes are somehow frozen in time, they are perhaps better seen as open ended, conditionally malleable symbolic systems.

Geographers have made significant contributions to the study of memorials and monuments by recognizing how collective memory is shaped by the politically contestable nature of space and place. Where the past is remembered is not incidental but actively shapes the memorialization process. Every memorial and monument has a placement and relative location that may confirm, erode, contradict, or render mute the intended meanings of the memorial's producers. A memorial's placement refers to the specific condition of its site, for example, its visibility, accessibility, symbolic elements, and its ad jacency to other parts of the landscape. In contrast, a memorial's relative location or situation is examined more broadly in relation to the rest of the city or countryside. Relevant issues here include its location vis à vis the area's mosaic of race , gender , and class based patterns; its proximity to power filled sites such as the central business district or other memorials; the flow pattern of tourists and visitors who visit – and just as importantly, those who do not.

Public commemoration is not simply about deter mining the appropriateness of remembering the past in a certain way; it also involves deciding where best to emplace that memory within the broader cultural land scape. Sometimes memorials and monuments become embroiled in controversy when antagonistic memories are perceived as infringing upon one another's symbolic space. The importance that location plays in shaping the meaning and politics of commemoration is well illus trated in the context of ongoing efforts to celebrate the Civil Rights Movement in the United States. Because of white opposition, African Americans have struggled and often been unsuccessful in placing monuments to the Movement within the traditional core of memorial space in cities. For the most part, Main Street, the county courthouse, and city hall remain devoted to remembering white dominated historical narratives. Many of America's roadways named for Martin Luther King, Jr., are side streets or portions of roads located within poor, black areas of cities. The placement of these monuments re inforce the importance of traditional racial boundaries and has, in some instances, changed a street's symbolic meaning from a point of African American pride to yet another reminder of continued inequality and injustice (**Figure 1**).

Approaches to Studying Memorials and Monuments

Memorial as Text

Geographic interest in memorials and monuments has grown out of a concern for how the past is remembered, as well as a more general concern with methods for in terpreting cultural landscapes. In particular, there has been a movement toward analyzing the landscape and its symbolic elements as documents that are written and read by social actors and groups within their own specific historical and ideological contexts. The 'text' metaphor has been one of the most dominant models for inter preting cultural landscapes, in general and those specific to memorials.

In a manner analogous to inscribing words on a page, the textual approach to studying memorials understands commemoration as the process of manifesting stories on

Figure 1 Martin Luther King Jr. Drive in Greenville, North Carolina (USA) marks an area that is predominantly African-American with limited commercial development, whereas East Fifth is mostly white and more upscale. Originally black leaders wanted all of Fifth Street renamed not just part of it and have expressed deep frustration over the city's segregation of King's memory. Where the past is remembered is not incidental but actively shapes the meaning and politics of the memorialization process. Photograph by Derek Alderman.

and through the landscape. This approach mobilizes a series of characteristic questions that seek to understand what historical discourses are represented and given authority via the memorial. These questions include: What is said and not said about the past? Whose history is remembered or forgotten through the spatial inscription of memory? To what extent does a monument silence certain accounts of the past while it gives voice to others? Does the differential treatment of histories tell us something about power relations and patterns of inequality within historical and contemporary societies? And to what extent do commemorative silences perpetuate these unequal power relations into the future?

In the case of Australia and the United States, for example, memorials and monuments have traditionally been built or 'written' in ways that ignore or misrepresent the history of indigenous peoples. In these narratives, white explorers and colonizers are cast as civilized heroes while natives are depicted as primitive and inferior, if they are shown to exist at all. The monuments found in Adelaide's Prince Henry Gardens – one of South Australia's most important symbolic sites – show clear evidence of this type of symbolic marginalization. All of the statues, busts, and memorials that geographers have examined honor people of European, immigrant origin. Commemorative texts can also take on important gender patterns. Only a few of the monuments located at Prince Henry Gardens memorialize a female. In addition to being historically incomplete, these exclusions assist in reproducing the social and cultural order, contributing to the further subordination and invisibility of under represented groups. How history is written and made socially important through memorials and monuments is not simply a matter of semantics but vital to achieving fairness and preventing the symbolic annihilation of marginalized social groups and their historical identities.

When interpreting memorials as texts it is important to note the dynamic nature of the 'writing' process. As was the case when paper was scarce and commonly reused time after time, memorial landscapes can take on the appearance of a palimpsest in which elements of past and present writing can be seen. This susceptibility to overwriting, embellishment, and erasure can also be thought of in terms of what has been called symbolic accretion. In a manner analogous to the geologic process of sedimentation, memorials and monuments undergo symbolic accretion as different historical meanings are layered onto them over time, thus challenging the notion that these symbols have a final, established meaning. Symbolic accretion often involves grafting a memorial with commemorative themes that ally with or confirm its original meaning. Yet, accretion can also be used to append alternative, potentially antagonistic narratives to memorials in ways that disrupt or contradict dominant interpretations of the past. Symbolic accretion is not only a way of analyzing the stratigraphy of ideologies written onto memorial landscapes. It also represents a political

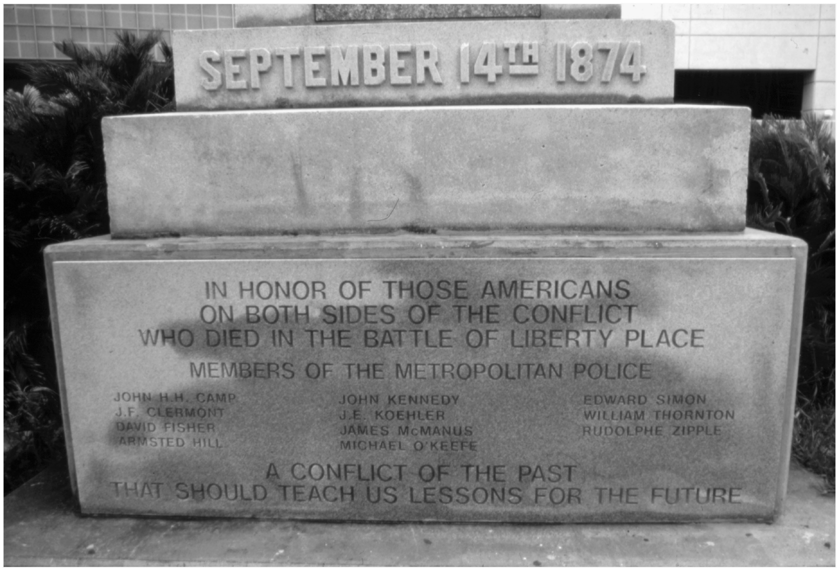

Figure 2 Long interpreted as offensive to New Orleans' black majority, the Liberty Monument was moved to a less prominent location and affixed with a new inscription. The inscription, pictured here, recognizes not only the White League but also those who fought against the racist organization. Monuments and memorial undergo symbolic accretion as different and sometimes contradictory historical meanings are layered onto them over time. Photograph by Owen Dwyer.

strategy used by groups to raise the visibility of their commemorative vision via an established memorial presence.

The once celebrated Liberty Monument in New Orleans, Louisiana, USA, offers a case in point. Erected in 1891 to commemorate the White League, a paramilitary organization that spearheaded the drive to disenfranchise African Americans in Reconstruction era Louisiana, the Liberty Monument was removed from the city's main thoroughfare in 1989 under the pretense of street repairs. Civil rights activists had long argued that the monument celebrated an illegal action and was an offense to the city's black majority. Following the protests by an odd coalition of white supremacists and historic preservationists, the monument was re erected but this time at a new, less prominent location. Importantly, the re placed monument included a new plaque, one that calls into question the memorial's original inscription by praising those who died fighting the White League.

The White League presumably did not envision that its monument would one day become a rallying point for civil rights activists, an antimonument to white supremacy. Yet no memorial or monument speaks for itself; each one is dependent upon its audience to support – or betray – its vision of the past. Similar to the interpretation of a book, 'readers' of monuments will react differently to what is ostensibly the same story. The fracas over the Liberty Monument demonstrated that insofar as a memorial claims to present the authoritative version of the past, it furnishes different audiences with the raw material for a new round of criticism regarding what is 'true' or 'real'. Beneath the appearance of historical consensus and stability, monuments – and by implication the meaning and significance of the events they represent – are the product of and conduit for ongoing debate (**Figure 2**).

Memorial as Arena

Recognition of the multiple ways in which memorials and monuments can be authored, read, and acted upon leads to an examination of the politicized nature of public commemoration. This approach utilizes the metaphor of 'arena' to direct attention to the political struggles and debates that frequently revolve around the representation of the past through the landscape. While collective memories are under constant reconstruction and reinterpretation, the process is not done freely or without constraints. People's ability to commemorate the past in certain ways is limited by competition and conflict with other people wishing to narrate the past in a different way. The potential struggle and contest over whose conception of the past will prevail constitutes the politics of memory. In this respect, memorials and monuments are places for social actors and groups to debate and negotiate the right to decide what is remembered and what version of the past will be made visible to the public.

In emphasizing how memorials and monuments are arenas for the politics of memory, one must recognize that historical representation is not only a product of social power but also a tool or resource for achieving it. The transmission of a commonly accepted conception of history is vital to legitimizing established political orders.

By the same token, challenging the dominance of existing ideologies or ways of seeing the world requires the creation and diffusion of alternative interpretations of the past. It is little wonder, then, that wholesale changes in the memorial landscape often accompany major shifts in political ideology and power within countries. Governments often use public symbols to legitimate certain nationalistic interpretations of history. The memorials and monuments of an earlier regime are replaced by commemorative icons that mythologize a new set of heroes, military campaigns, and ideological causes. Accompanying these changes are decisions about what to do with now illegitimate memorials. Should they be destroyed, stored out of sight, or relocated to places that encourage public contemplation rather than emulation? Nowhere has this process been made more apparent than in Central and Eastern Europe after fall of communism. While these commemorative alterations are often represented in the media as dramatic and victorious in tone, the image belies the political struggles inherent in reinventing national history and tradition. In post Soviet Russia, political elites competed with themselves and the larger public over the control and transformation of memorials and monuments from the previous communist era. These places of memory became important pieces of symbolic capital within the new political order as state leaders vied to maintain or gain greater prestige, legitimacy, and influence (**Figure 3**).

Commemoration is especially controversial when it involves remembering trauma, atrocity, and violence. While some places of tragedy are obliterated, others go through a process of sanctification in which people make them into sacred places, not in a strict religious sense but in terms of creating highly meaningful and respected memorials or monuments. Public remembrance of atrocity is necessary as a tool for facilitating social compensation to victimized groups, moral reflection among the larger society, and public education. Memorials and monuments related to Holocaust are instructive here, particularly in light of the competing historical meanings and interpretations that have been projected onto them. Despite Auschwitz's significance to Jews, the Polish Catholic Church has waged a controversial appropriation of the camp's symbolic space since 1970s. The martyrdom of Polish Catholics displaced representations of Jewish victimization. Atrocity is a difficult commemorative topic because it requires identifying victims, perpetrators, and heroes. While these roles would appear to be clearly set, they can be defined in quite fluid and sometimes opposing ways. Buchenwald became the site of such debates following revelations that the former Nazi concentration camp had also been used as a Soviet detention camp for Germans after World War II and that Nazi perpetrators were among the victims of the Soviet camp.

The past becomes contentious because of the close connections between memory and identity. How we imagine ourselves in the present is intimately linked to how we remember and represent ourselves in the past, providing justification for why we should be recognized and respected publicly. Heritage is inherently dissonant or open to discord or disagreement. Expressing one's

Figure 3 As with many political revolutions, a commemorative reorganization accompanied the 1989 fall of the Communist government in Hungary. New memorials were created and previously neglected ones were restored. Many Communist-era monuments were moved from public places to Statue Park, an outdoor museum near Budapest. The statue of Lenin in the foreground, previously located at the main entrance of the Cspel Iron Works, was actually rescued and hidden by workers before it could be officially toppled. Reproduced with permission from Hungary-Statue Park.

heritage invariably means that another, different identification with the past is disinherited, excluded, or degraded.

Memorials and monuments play an important role in constructing a sense of belonging to a national heritage, although these commemorations can be a site of tension between nationalism and other competing lines of identity. The memorial landscape can also express notions of regional identity, which is also open for debate and reinterpretation among social groups. Recent public controversy in the southeastern United States over the public display and meaning of the Confederate battle flag provides a good example.

The Confederate flag, which was used by slaveholding southern states during the United States Civil War, has become a lightning rod for competing interpretations about the symbol's meaning and, ultimately, the historical reasons for why the South seceded from the Union. On the one hand, many white southerners see the flag as a symbol of their ancestors' heroic fight for independence from an intrusive federal government, hence suggesting that slavery was not the 'real' reason for the Civil War. On the other hand, many African Americans view the flag as a symbol of racism and a commemoration of past efforts to preserve slavery and white supremacy. They argue that the mere presence of these memorials to the Confederacy not only marginalizes blacks from any meaningful discussion of the South's history but also openly conflicts with their ongoing efforts to construct a more empowering image of black history built upon civil rights and social justice. In this respect, the Confederate flag also functions as an arena for struggles over racial identity as black and white southerners compete to decide whose interpretation of the past will prevail. The arena metaphor, with its emphasis on contest and spectacle, prompts us to consider more closely how people present or display their connections with the past to the larger public.

Memorial as Performance

While geographers have most often focused on the textual and contested nature of memorials and monuments, some have emphasized the importance of 'performance' to commemoration. The metaphor of performance directs our attention to the ways in which memorial landscapes serve as a stage, literally and figuratively, for a wide range of performances such as public dramas, rituals, historical re enactments, marches and protests, pageants, civic ceremonies, and festivals. It is not just that these performances happen in or at places of memory. Rather, the memorial landscape is constituted, shaped, and made important through the bodily performance and display of collective memories. For instance, since the late 1970s, the activist organization Madres de Plaza de Mayo (Mothers of the Plaza) has carried out collective rituals such as occupying and marching in city plazas. They do this in the memory of children abducted and killed by the formerly oppressive Argentine government, forcing the country to come to terms with its troubled past. Through the bodily performances of the 'mothers', such as wearing white scarves in remembrance of the diapers of their lost children, ordinary plazas are converted into emotionally and politically charged memorials. The body itself can be viewed as a place for commemoration, as well as political expression (**Figure 4**).

Performance is particularly important within the heritage tourism industry. Host communities are often concerned with projecting a positive image to tourists, one that actively displays a distinctive and engaging local history. In many cases, existing memorials, monuments, and historic sites become stages for historical tours and re enactments that carry out this commodification or selling of the past. The packaging of the memorial landscape through performance is not confined to the drama created by actors and historical guides but includes the everyday, seemingly mundane practices of tourism workers as they simply talk about the past with visitors. Festivals and pageants are community performances increasingly used to highlight the historical identity of places. These landscape spectacles appear to be innocent if not often overly commercialized representations of the past. Yet, they – like any memorial or monument – highlight selective visions of the past. In 1996, Bristol, England, hosted the International Festival of the Sea in an attempt to market its maritime heritage. Festival organizers excluded certain histories from promotional performances, namely the port city's role in slavery and imperialism. Even more striking was the city's forced removal of traveler populations – itinerant groups who roam the countryside – from the festival site for fear of offending tourists. In this case, Bristol's festival organizers presented a sanitized image of the city, prompting us to consider how commemoration involves not only a selective performance of certain historical narratives (over others) but also decisions about who should (or should not) be part of these performances.

The sum of these different approaches is that memorials and monuments do more than simply reflect ideas about the past and secure them to the ground. They extend these ideas into the world and actively participate in how people write, read, debate, and perform history in selective and varying ways. Memorials and monuments are inherently political. They can be manipulated by those social groups in control while also providing a platform of resistance and struggle for those traditionally powerless. The past is a potentially contested terrain and where memorials and monuments are located – in our sense of time and our sense of place – plays a critical

Figure 4 Members of Madres De Plaza de Mayo hold each other and walk together at the central plaza in La Plata, Argentina. These female activists demonstrate the important role that rituals, bodily performances, and displays of collective memory play in converting ordinary public spaces into emotionally and politically charged memorials. Reproduced from Bosco, F. J. (2006). The Madres de Plaza de Mayo and three decades of human rights activism: Embeddedness, emotions, and social movements. *Annals of the Association of American Geographers* 96(2), 357, with permission from Blackwell Publishing.

role in shaping what (and who) is ultimately remembered and forgotten.

See also: Cultural Politics; Festival and Spectacle; Heritage; Landscape Iconography; Material Culture; Memory; National Spatialities; Street Names and Iconography.

Further Reading

Alderman, D. H. (2002). Street names as memorial arenas: The reputational politics of commemorating Martin Luther King Jr. in a Georgia county. *Historical Geography* 30, 99 120.

Atkinson, D. and Laurier, E. (1998). A sanitized city? Social exclusion at Bristol's 1996 international festival of the sea. *Geoforum* 29, 199 206.

Azaryahu, M. (2003). RePlacing memory: The reorientation of Buchenwald. *Cultural Geographies* 10, 1 20.

Bosco, F. J. (2004). Human rights and scaled performances of memory: Conflicts among the Madres de Plaza de Mayo in Argentina. *Social and Cultural Geography* 5, 381 402.

Bosco, F. J. (2006). The Madres de Plaza de Mayo and three decades of human rights' activism: Embeddedness, emotions, and social movements. *Annals of the Association of American Geographers* 96(2), 357.

Charlesworth, A. (1994). Contesting places of memory: The case of Auschwitz. *Environment and Planning D: Society and Space* 12, 579 593.

Dwyer, O. J. (2004). Symbolic accretion and commemoration. *Social and Cultural Geography* 5, 419 435.

Florida Center for Instructional Technology (2005). Teachers Guide to the Holocaust. http://fcit.usf.edu/holocaust/resource/resource.htm (accessed on 20 August 2007).

Foote, K. E. (2003). *Shadowed ground: America's landscapes of violence and tragedy* (2nd edn.). Austin: University of Texas Press.

Forest, B. and Johnson, J. (2002). Unraveling the threads of history: Soviet era monuments and post Soviet identity in Moscow. *Annals of the Association of American Geographers* 92, 524 547.

Graham, B., Ashworth, G. J. and Tunbridge, J. E. (2000). *A geography of heritage: Power, culture and economy*. London: Arnold.

Hay, I., Hughes, A. and Tutton, M. (2004). Monuments, memory, and marginalization in Adelaide's Prince Henry Gardens. *Geografiska Annaler B* 86, 201 216.

Hoelscher, S. D. (1998). *Heritage on stage: The invention of ethnic place in America's little Switzerland*. Madison: University of Wisconsin Press.

Johnson, N. (1995). Cast in stone: Monuments, geography, and nationalism. *Environment and Planning D: Society and Space* 13, 51 65.

Leib, J. I., Webster, G. R. and Webster, R. H. (2000). Rebel with a cause? Iconography and public memory in the southern United States. *Geojournal* 52, 303 310.

Till, K. E. (2003). Places of memory. In Agnew, J., Mitchell, K. & Toal, G. (eds.) *A companion to political geography*, pp 289 301. Malden: Blackwell.

Relevant Websites

http://www.auschwitz.org.pl
Auschwitz Birkenau Memorial and Museum.
http://www.battleflag.org
Battleflag! Preserving Confederate Memory.
http://www.highwaymarker.org
Driving History Home: Historical Highway Markers.
http://www.english heritage.org.uk
English Heritage.
http://www.explore.canada.travel
Heritage Festivals of Canada, Canadian Tourism Commission.

http://www.h net.org
H Memory Discussion Network, H Net: Humanities and Social Sciences Online.
http://www.scrapbookpages.com
Holocaust Sites.
http://www.nationaltrust.org
National Trust for Historic Preservation.
http://www.ncmarkers.com
North Carolina Highway Historical Marker Program.
http://www.oklahomacitynationalmemorial.org
Oklahoma City National Memorial and Museum.
http://sites of memory.de
Sites of Memory by Mark Hatlie.
http://www.szoborpark.hu
Statue Park, Budapest, Hungary.
http://www.uvm.edu
Ten Questions to Ask at a Historic Site, University of Vermont.
http://www.skp.com.au/memorials2/
War Memorials in Australia.
http://www.cr.nps.gov
We Shall Overcome: Civil Rights Places of United States.
http://www.buildthememorial.org
World Trade Center Memorial Foundation.
http://www.worldwartwohrs.org
World War II Historical Re enactment Society.

Memory

S. McDowell, University of Ulster, Coleraine, UK

Glossary

Contestation It is to refute, dispute, or disagree.

Cultural Landscape The environments which reflect the interactions between populations and their surroundings.

Heritage The selective use of the past as a resource for the present and future; an assemblage of myths, histories, values, and behaviors.

Identity It is the collective aspect of a set of characteristics by which a person or thing is definitively known.

Power It is the production of effects over something or someone.

Representation It means to communicate, depict, symbolize, or portray.

Territory It is a geographic space containing a homogeneous or collective group.

The Nuances of Memory

Because of memory's integral importance to our sense of self, scholars of many hues have sought to understand it using a plurality of theoretical and methodological ap proaches. The study of memory has proliferated among a number of disciplines to such an extent that it is no longer sufficient perhaps to discuss it as some kind of unitary or abstract entity. There are, for example, mul tiple subgenres to memory: official, unofficial, public, private, collective, communal, local, national, societal, historical, emotional, urban, and postmemory – to name but a few, all of which have attracted the attention not only of psychologists but also of historians, anthropolo gists, sociologists, and increasingly, geographers.

Primarily and more traditionally, academics have thought about memory in terms of scale, from the indi vidual or private which focuses on personal experiences or recollections of the past; the local or communal, which draws on key events or symbolic moments that have occurred within close knit groups; to societal memory which describes narratives of the past which are sympa thetic to a broader, loosely interconnected population. On that same scale is public and national memory. The former emanates from the fusion of official and localized cultural expressions, which is often a manifestation of political and social relationships. It is, moreover, a process that is not only negotiated by official or national groups but also by the media, academics, heritage institutions, and local community organizations. The latter – national memory – commonly derives from the state and its in stitutions representing the perceived hegemonic needs and values of the general public. Frequently thought of in conjunction with official memory, national memory is fundamentally linked to nation building and concentrates on specific aspects of the past which helped form the nation. Consequently, the state is usually the official mediator of national memory and, as such, it assumes responsibility on instructing the nation what to remem ber. The African National Congress (ANC) government, for example, after taking power in South Africa after apartheid, set about establishing 6 days of national remembrance. Unofficial memory typically used to describe memories that are not state structured and thought of as a binary opposite to official or national memory is arguably a somewhat ambiguous and dan gerous term, in that it infers a certain lack of legitimacy or authority. This is, of course, refuted by many groups and individuals who regard their own individual, local, or communal interpretations of the past to be just as valid and 'official' as that of the state or other officially sanc tioned forms of remembering.

More recent subgenres of memory, triggered, perhaps, by new cross disciplinary theoretical approaches and in sights into the multifaceted and contested nature of re membering, include literal, exemplary, emotional, and postmemory. Literal memories are those recollections of the past attached solely to violent or disturbing events which do not extend to other similar events. Exemplary memories, conversely, extend their reference to other similar occurrences, thus making it easier to detract les sons from the past to inform the future. The study of 'postmemory' also looks at the effects of violence on memory and is particularly pertinent in analyses of Holocaust memory or the remembering of war. It de scribes a form of memory that has been circulated through other people's experiences of a violent event. In post memory, memories are passed down through generations to be represented by people who have no personal at tachment to the memory. Subsequently, they seek to reuse, reenact, and re represent those memories in order to feel closer to their ancestors. The term emotional memory, meanwhile, has been increasingly employed by scholars to describe the transgenerational remembering of traumatic events. At the heart of such varied trajectories has been an attempt to dissect the multiple meanings and functions of a complex and challenging concept.

Memory and Geography

Phenomenological approaches to memory have pondered over whether or not remembering is something that is done, as opposed to something almost innate that people have, and scholars have questioned the extent to which the latter are different from our manipulation of them. Indeed, a growing body of work supports the argument that memories are not necessarily embedded cognitively; rather they are produced through social and spatial processes. The influence of place on the process of remembering has long been the subject of geographic interest, and geographers are specifically concerned with the representational dimensions of memory and how it is produced through social activities of remembrance, such as commemoration. It has been convincingly argued by many that all memories have a geography as they are recalled or evoked, for the most part, through specific sites and attach our recollections and experiences to particular places. When most of us think of the 11 September attacks in 2001, for example, we are instantly transported to the devastated site of Ground Zero – our minds replay the shocking scenes of the Twin Towers engulfed in flames. Equally recalling Martin Luther King's "I have a dream speech" is now virtually synonymous with the National Mall in Washington and the backdrop of the Lincoln Memorial. Geographers, conceptualizing memory as spatially constituted, seek it in a place and a location and argue that there are strong connections between physical locations and remembering and commemorating the past. Concentration camps and battlefields, sites of bombings and assassinations, and the locations of speeches and celebrations are places that incite our memories. Places are best understood as environments that have been invested with meaning. They are, therefore, the locations with which people connect, either physically or emotionally, and are bound up in notions of belonging (or not belonging), ownership, and consequently, identity. Geographical studies of places of memory have centered on four specific areas: landscapes and commemoration, identity and heritage, the politics of remembrance, and increasingly the relationship between memory and territory.

Dissecting Memory in the Cultural Landscape

The making and marking of place underscores the potential of cultural landscapes (which in geography can be understood to constitute those environments which reflect the interactions between populations and their surroundings) to act as tangible places to dissect the meanings and functions of memory. The desire to mark 'place' is often strong after a tragedy. People often flock to the sites of deaths or killings, bringing with them ephemeral commemorative paraphernalia – flowers, prayers, pictures, and poems. The instantaneous commemoration of the 1995 Oklahoma bombing is illustrative of this point. Only hours after the bombing, the site's peripheries became almost sacred space. Onlookers brought wreaths; toys, and messages of condolences. As the days progressed, ideas for a permanent memorial flooded into the governor's offices from all over America. The site not only united the country in grief but also communicated messages of resilience and defiance. The wake of the 2006 London bombings evoked a similar response. Again the need to sanctify each site became immediate in the days and weeks following the attacks. Landscapes then, as representative entities, play critically important roles in shaping how the public interprets, defines, and debates its past. Simon Schama's work on the relationship between landscape and memory has been particularly successful in propagating this approach.

A long established geographic concept is, of course, the idea of landscape as a 'text' or a communicational resource, as something to be read and interpreted. Any reading of it must then focus its features and layers, primarily on the visual images and symbols embodied within it. It is universally agreed that visual matter in particular is fundamentally important to the cultural construction of identity, and scholars are largely agreed that what is visible is ubiquitous, reinforcing the notion that the cultural landscape is a dynamic resource in the interpretation and articulation of identities and memories. Geographers have led the field in dissecting the visual or iconographic features of the cultural landscape, such as public buildings, monuments, plaques, plinths, graffiti, and street names. As inscriptions of memory which find tangible representation in the landscapes around us, they commemorate and champion partisan interpretations of the past and present onto public places. As such, they represent what is valued (and what is not) within a society, and can be read as icons of identity and spatializations of history; they are, what Pierre Nora famously termed *Lieux de mémoire* (sites of memory), which possess the ability to embody the past and stimulate remembering in fixed locales. Groups and individuals within societies reconstruct signs and symbols and read them in different contexts, thus transforming their reference and meaning. The cultural landscape then, and the memories and meanings it embodies like society, is in a constant mode of flux, as it consistently develops and mutates. Rarely passive, it is subject to reappropriation, interpretation, contestation, and destruction. Thus, cultural landscapes as sites of memories are representative, emblematic, evocative, and expressive, and reveal the contestable and diverse nature of acknowledging the past.

Memory, Identity, and Heritage

The themes of identity and heritage have formed integral components of geographical approaches to understanding the meanings and functions of memory. Remembering the past is, of course, an essential part of the present, and is important for a number of reasons, one of the most significant being the articulation of identity. It is universally agreed that memory is something that is essential to our understanding as human beings, and therefore is inexorably tied to our sense of identity and to our ability to construct and deconstruct identities in the present. In cognitive terms, scholars believe that a person's identity is shaped only by that which is accessible to remembrance. For a group or nation to have a collective identity, it has to have a shared narrative of the past, an agreed acceptance of all those events which have formed the group over time, including knowledge of the group's origins. Particular emphasis may be placed upon specific moments or events which resonate strongly with the group's identity. These collective beliefs are crucially important in fostering a sense of togetherness and cultural solidarity. National cohesion, in particular, requires a shared awareness and identity that is endorsed through common historical experience. This gives weight to the idea that memory is socially constructed and negotiated in the present through a variety of different media and processes. Never static, it is malleable to the influences of the present and the requirements of the future.

Remembering the past is invariably selective as it yields to the demands of specific interests and ideologies at particular times and in particular spaces. As such, the past is continuously reshaped for contemporary identity requirements. This manipulation of memory is commonplace within every social order as it continuously retells and reinvents its past, often at the expense of historical accuracy. America's commemorative landscape had until the 1960s and 1970s completely omitted the narratives of Native Americans, African Americans, and Hispanics, to name but a few. Memories are typically modified to suit particular situations and times, and do not then always correlate with historical truths. Throughout the selection process, memories are reconfigured, becoming infused with myths. Such distortions are rarely accidental as the dimensions of 'forgetting' in the construction of identity are just as crucial as remembering. In many societies, the exorcism of dark, potentially damaging, or shameful memories is an integral part of maintaining a collective identity and ensuring the continuing unification of the group. So too are memories which threaten to disrupt a shared interpretation of the past. Silences within narratives of memory have become fertile ground for geographic analysis. Kenneth Foote's work on landscapes of violence and tragedy in the US explores text on the theme of forgetting. One of Foote's case studies examines the deliberate erasure of the seventeenth century witch trials in Salem from public memory, illustrative of how shameful episodes in a nation's history can become erased or obliterated from public consciousness.

However, the past is, as geographers have noted, not simply subject to processes of remembering and forgetting, but its participants are also profusely sensitive to its potential in the present. If we understand memory as something that is alive, partisan, and negotiated in the present, then we must too accept that it is inexorably linked to the heritage process. Heritage is a term that is employed to describe the selective use of the past in the present. Conceived both spatially and socially it is an assemblage of myths, values, nostalgia, and traditions, and is carefully tailored to meet the needs of contemporary societies, be they cultural, social, political, or economic. David Lowenthal's works *The Past Is a Foreign Country* and the *Heritage Crusade and the Spoils of History* have, arguably, been two of the most important texts written on the complex relationship between memory, identity, and heritage. Heritage, Lowenthal believes, conjures feelings of inheritance, familiarity, commonality, and continuity and speaks to our sense of understanding. It can too be concretized in specific sites through the insertion of commemorative symbols, situated performances, or rituals. It is (like memory) constructed, a cultural product, or tool that is alive, negotiated, and contested. Just like amnesiacs who lose their sense of self because they cannot remember who they were, without our ability to remember, identity, culture, and heritage is lost. Through acts of remembrance, such as commemorations, cultures and traditions are formed, maintained, and destroyed, as memories are manipulated and exaggerated.

The Politics of Memory

The study of places of memory can reveal how different political perspectives are expressed through spatial practices and how meaningful place is produced and reproduced through those practices. Geographers have written extensively on the specific capacity of places of memory to represent power, two in particular: Andrew Charlesworth, who has explored the meaning of Holocaust memory, and Karen E. Till, who has published widely on the politics of memory, have made substantial contributions to this field. Collectively, they argue that *lieux de mémoire* such as monuments, memorials, and buildings, which have been invested with meaning, carry conscious and subconscious messages, and are subject to competing interests. Their place in public life translates powerful ideological messages that are never apolitical and ensures the messages they communicate are open to contested interpretations. Charlesworth's research on

Auschwitz (as one of the most potent symbols of the Holocaust) shows how malleable sites of memory are to manipulation and how susceptible they are to power struggles. His examination of the contestation over symbolic space at Auschwitz highlights the individual attempts by both the Catholic Church and the former Polish communist state to claim ownership over the site for their own needs.

Dominant ideologies often evoke the past through spatial and temporal representations to reinforce their influence over society, dictating and defining what is remembered (and consequently forgotten). Those with the most at stake politically, or those with the greatest ability to exercise power, have a vested interest in the production of partisan sites of memory and bring the past into focus to legitimize a present social order. This can be seen in the postsocialist reconstruction of identity in Central and Eastern Europe, where after the fall of communism there was a return to presocialist values and national identities. Through evoking the presocialist period, the renaming of streets in places like Romania bolstered the postsocialist regime's claims to power. Instrumental to communal or group solidarity is a sense of unity only achieved through a shared memory or common experience. At the core of all social orders is a sense of continuity over time and space, sustained by nostalgia for the past and the careful tailoring of memory by powerful agents to fulfill the identity requirements of the present. Sites of memory, such as monuments, plaques, museums, and symbolic architectural spaces, as static and permanent reminders of the past concretized in the present are largely constructed by national governments to represent hegemonic values that cultivate notions of national identity and frame ideas and histories of the nation. As sites of civic construction, they instruct citizens what to value concerning their national heritage and public responsibilities. Such sites represent and embody power, greatness, resistance, memory, and loss. Citizens participate in the production and reproduction of the past through a variety of spatial practices, including wreath laying, attending memorial services, marching, and a number of other performative rituals. Reenacting and repeating the past at these sites not only consolidates the legitimacy of the memory being evoked but also reifies ownership of the space and reinforces the power of the group.

While the past has the power to bond communities, societies, and nations in the present, it can also operate as sites of resistance or division, particularly if the memory evoked has no resonance with groups or individuals within a particular social order. Remembering the past is not always consensual, and the creation of meaningful place is habitually fraught with contestation. Till's work on the struggle to find an appropriate past for the 'new' Berlin's present is particularly good at illustrating the difficulties of finding common ground during the commemorative process. It is, for example, unlikely that all of a nation's population will offer equal support to the emphasis placed upon the remembering of a particular aspect of the past, and competing interests frequently collide over physical representations of power in the cultural landscape. In some instances, groups and individuals who contest the memory being evoked (or forgotten) will work to undermine or manipulate the memorial site or create their own separate important place which is indicative of their own interpretation of the past and understandings of the present. Take, for example, the efforts to challenge official interpretations of the past in Guatemala, which has been plagued by violence resulting from conflict between the state and guerrilla forces since the 1970s. In February 1982, the Guatemalan army killed 177 women and children and buried them in a mass grave (the killings later became colloquially known as the Rio Negro Massacre). Twelve years later, the tiny community of Rabinal Baja Verapuz constructed a memorial entitled the 'Monument of Truth' which, for the first time, told the 'truth' about the massacre and the government's involvement in rape, maiming, and murder. After only 2 weeks, the memorial was smashed by state forces in an effort to silence conflictual representations of what had taken place. Undeterred, the community erected another memorial the following year on the thirteenth anniversary of the killings. The second memorial was constructed from steel and concrete, testament to the community's wishes for its story to 'be told'. Names of the 70 children who were killed were etched onto the steel front of the memorial while the inscription on an accompanying plaque recalled the brutality of the event: "Children who were smashed against the rock." The eventual siting of the memorial to the victims of the Rio Negro Massacre signaled the creation of new spatial arenas through which to challenge interpretations about violence.

Memory and Territory

Inextricably linked to the representation of memory are questions of territory and territoriality. If, as some geographers have suggested, all memories are metaphors of physical locations, it is important to consider the territorial implications of remembering the past. A territory is a demarcated geographic space that typically contains some kind of homogeneous group, sharing a collective identity or heritage. Territories control and mobilize groups and their resources inside defined boundaries where access is often strictly controlled (both physically and imaginatively). Identity, for example, is largely expressed through territoriality as groups, communities, and populations identify or connect with particular places, landscapes, or sites. Memory therefore is fundamentally important in legitimizing or justifying territorial claims, as it infers a sense of natural or historical belonging (or not belonging), and generates a feeling of rootedness, familiarity, and continuity. Within societies then, various social orders insert symbols or perform activities which resonate with their interpretations of the past and their narratives of identity, which simultaneously incite remembering and bound or affirm territory. The conflation of the past and present through poignant sites of memory or other spatial practices reveals the potency of memory to concretize territorial demands over spaces of production. This can be clearly seen in segregated places, such as Northern Ireland, where ethnonationalist groups insert commemorative monuments along interfaces to demarcate their respective territories. Effectively, such spatializations of memory constitute territorial markers which are strategically placed or performed within and outside a geographic space. Physical boundaries, reenactments, or patterned movements and the gradual manipulation and construction of meaningful place confirm ownership over specific spaces, demarcating dichotomous boundaries for those who live within territories, as well as for those who live outside the territory (and whose narratives of the past do not resonate with those of the homogeneous group).

Summary

In sum, memory is a process which is inexorably bound to our present day sense of self. We evoke it for validation, legitimization, and unity, and we draw on it in order to challenge, refute, and undermine. It is political and often territorial, serving certain agencies and groups through communicating narratives of inclusion and exclusion, and continuity and instability.

See also: Genealogy and Family History; Heritage and Culture; Heritage and Identity; Nationalism, Historical Geography of; Street Names and Iconography; Time and Historical Geography; War, Historical Geography and.

Further Reading

Charlesworth, A. (1994). Contesting places of memory: The case of Auschwitz. *Environment and Planning D: Society and Space* 12(5), 579 593.

Dywer, O. J. (2000). Interpreting the civil rights movement: Place, memory and conflict. *Professional Geographer* 52, 660 671.

Foote, K. (2003). *Shadowed Ground: America's Landscapes of Violence and Tragedy*. Austin, TX: University of Texas Press.

Harvey, D. (2000). Landscape organization, identity and change: Territoriality and hagiography in medieval West Cornwall. *Landscape Research* 25(2), 201 212.

Hoelscher, S. and Alderman, D. H. (2004). Memory and place: Geographies of a critical relationship. *Social and Cultural Geography* 5, 347 355.

Lowenthal, D. (1985). *The Past Is a Foreign Country*. Cambridge: Cambridge University Press.

Lowenthal, D. (1996). *The Heritage Crusade and the Spoils of History*. Cambridge: Cambridge University Press.

Nora, P. (1989). Between memory and history: *Les Lieux de Mémoire*. *Representations* 26, 7 25.

Riceour, P. (2004). *Memory, History and Forgetting* (transl. Ballemy, K. and Pellauer, D.). Chicago, IL: University of Chicago Press.

Samuel, R. (1996). *Theatres of Memory: Past and Present in Contemporary Culture*. London: Verso.

Schama, S. (1995). *Landscape and Memory*. New York: Vintage.

Till, K. (2005). *The New Berlin: Memory, Politics and Place*. Minneapolis, MN: University of Minnesota Press.

Mental Health

C. J. Smith, University at Albany, SUNY, Albany, NY, USA

Glossary

Deinstitutionalization A long-term trend which has seen fewer people living and being treated as 'patients' in mental hospitals. This trend is associated with the closing of many public hospitals and the subsequent transfer of patients to community-based mental health services.

'Deviant' Behavior Behavior that is significantly different from the traditional or generally accepted standard.

Mental Health A state of emotional and psychological well-being in which an individual is able to function effectively in society, and meet the ordinary demands of everyday life.

Mental Illness General label used to describe any of a number of psychiatric disorders that are related to untypical/deviant behavior.

Post-Asylum Geographies The geographic study of mental illness and mental health in community settings.

Psychiatric Ghettoes Geographically restricted areas, usually in urban core locations, that are characterized by an excess concentration of people experiencing mental health problems and of services and facilities catering to their needs.

Sink Estates The use of 'sink' in this context (in the UK) is intended to describe public ('council') housing estates that become places of last resort for individuals and families that have nowhere else to go and can not afford to buy properties on the private market. A 'sink' estate is normally associated with a high degree of socio-economic deprivation, which often goes hand in hand with criminal activity and a range of other social problems

Spatial Epidemiology Study of the origins and spread of disease within a population, emphasizing geographical variables.

Stress/Psychological Stress A mentally or emotionally disruptive or upsetting condition occurring in response to adverse external influences and capable of affecting physical and psychological health.

The 'Drift' Hypothesis Assumption or observation that a localized concentration of a phenomenon (e.g., mental illness) is caused by human movement into the locality.

Introduction

In the early 1970s, some human geographers began to lobby for the serious consideration and study of social problems that had previously been ignored by their colleagues. As part of what might be referred to as the emergence of a social conscience within the discipline, the literature began to feature some reports of geographers' efforts to explore the dimensions of human suffering and what sociologists refer to as 'deviant' behavior. One component of this exploration was an approach to the constellation of spatial issues related to mental health and the onset of mental illness. Since that time, a significant body of research has emerged that can be referred to collectively as 'mental health geographies'. The early work did not register as a significant component in the larger realm of human geography output, and researchers approached the issues from a variety of conceptual and methodological stances, which made it difficult work to categorize accurately. With the passage of time, however, it has been possible to identify two overlapping categories in the work: one focusing on the distributional geographies (the 'spatial epidemiology') of mental illness, the other involved with the changing locational characteristics of the provision of care for the mentally ill, including interpretations of the effectiveness of such care, and community reactions to the siting of mental health services in community (nonhospital) locations.

In definitional terms, the majority of geographers working in this field appears to have taken as given the labels and definitions used in defining different categories of diagnosed mental illness. Mental health, on the other hand, is generally considered to be the absence of mental illness, either in a given individual or in a community. The research conducted during the last three decades shows that geographers have been eclectic in defining who it is they are studying and what their problems have been. For obvious reasons, mostly associated with the difficulties of data collection, many researchers have chosen either to make use of official diagnoses or to accept the implications of such definitions by focusing their work on a population housed in a facility whose usage assumes such definitions. In later studies, some researchers have gone beyond or outside these traditional psychiatric labels and have tried to describe and define the populations they are interested in by using terms such as 'stress' and 'psychological distress'. It is implicit that such researchers prefer to focus their work on the problems associated with everyday life, and on situations in which individuals experience a series of adjustment difficulties that are not incorporated into or measured by the formal diagnostic categories of mental illness.

Three 'Waves' of Mental Health Research?

The majority of the early research on mental health and mental illness was conducted in cities, and much of it was designed to test the hypothesis that specific urban conditions are associated with the onset of mental illness. It was the urban sociologists of the Chicago School who initiated such research, which involved mapping the distribution of specific mental illness categories in the city. In many of the published studies, the data pointed to a marked distance decay in the prevalence of mental disorders with increasing distance from either the city center or from a specific facility such as a mental hospital. The assumption was that many persons with a potential or actual mental illness condition would, in the absence of viable support systems, be housed in residential neighborhoods where living conditions were suboptimal. It was assumed that social isolation and other non normative living conditions in such neighborhoods were causal factors in the onset of their illness, and that if they returned to such neighborhoods after hospitalization, they would undoubtedly experience additional difficulties. The general view, in other words, was of individuals who are 'alone in the crowd', spatially proximate to others yet socially distant from them for a variety of behavioral and attitudinal reasons, and shunned by the wider society.

Geographers who began to map psychiatric case registers, beginning in the early 1970s, also found higher rates of diagnosed mental illness in downtown areas, which supported some sort of 'breeder' hypothesis, suggesting that life in such neighborhoods was dangerously stressful. Such studies became more sophisticated over time, but the conclusions were essentially the same: it seemed that people experiencing mental health problems were more likely to live in inner city neighborhoods, creating the rudiments what some geographers would later come to refer to as 'pyschiatric ghettoes' that would be home to a combination of facilities providing services for the mentally ill.

Most of the early work in this subfield was highly positivist in its approach, and generally undertheorized: researchers were looking for spatial patterns in mental illness rates and correlations with some obvious socioeconomic characteristics in an effort to test hypotheses about causation. In general, the researchers were able to map such patterns, at least crudely, and with some degree of statistical significance they could describe and interpret the harmful nature of 'stressful' localities, which in many instances were to be found in urban downtown areas and along 'skid row'. Although few of the findings were remarkable or unexpected, the results of this early work made the case that patterns of mental illness, for so long defined in purely individual terms, could realistically be explored at the larger geographical or 'contextual' level, suggesting that place and location might be integral to the onset of mental health problems.

At intervals, an interesting spatial twist was added to or offered in place of this generalized conclusion, suggesting that at least part of the observed geographic pattern of mental illness resulted from human population movement, or 'drift', with the relocation of some people closer to mental health facilities being interpreted as an indicator of their deteriorating mental health. Some studies pointed to yet another potential explanation, namely, that for simple reasons of financial expediency and a lack of alternatives, psychiatric patients were being released (or 'dumped') into urban neighborhoods where opposition to their presence was minimal. To some extent this conclusion was interpreted as having less inherent interest for geographers, because the observed locational pattern was a contrived one, rather than the result of human choices. But it was also evident that living in the neighborhoods identified in this way represented an additional set of 'stressors' for people experiencing difficulties coping with everyday life, so it was not surprising to find high rates of recidivism. There is also a more positive interpretation to be made from the findings of such studies, based on the assumption that such neighborhoods provide a more normative or even positive living environment for the mentally ill. As the reasoning goes: it is precisely in neighborhoods like this (as opposed to more suburban and/or more upmarket locations) that supportive social networks of individuals experiencing mental health problems may emerge, alongside a clustering of mental health services and facilities which have been located there as a last resort. Such places may be seen by outsiders as the worst possible environment for the mentally ill; but from another perspective they may be thought of in more positive terms, as perhaps the only and maybe even the best place for such people to live during difficult times.

The other branch of the mental health geographies' subfield has been associated with studies of the spatial characteristics and implications of the facilities, both institutional and community based, where treatment for the mentally ill has been provided. The earliest among these studies can be grouped broadly under the heading of 'asylum geographies', which focused on 'lunatic asylums', later relabeled as 'mental hospitals'. Such institutions first appeared in Western societies in the eighteenth and nineteenth centuries, and were intended to separate the mentally ill from the mainstream of society, while at the same time providing some form of 'treatment' for their symptoms. In the 1970s and 1980s, this work would dovetail with what might be called 'post asylum geographies', made up of studies that had a number of overlapping research themes, including: (1) efforts to interpret the causes and consequences of the closure of the old asylums and the subsequent 'deinstitutionalization' of the mentally ill; (2) investigations into the provision and effectiveness of alternative care facilities provided in a variety of community based treatment programs; (3) descriptions and assessments of the emergence of spatial concentrations of former mental patients and service programs in specific urban locations, which became known simply as 'service dependent ghettoes'; (4) studies investigating the impact of such community facilities on their host neighborhoods, which were accompanied by some of the earliest examples of what would become widely known as the not in my backyard (NIMBY) phenomenon; and (5) analyses of the successes and failures in terms of community adaptation and 'coping' among the already and the never hospitalized mentally ill. This last approach saw geographers exploring the dimensions and operations of community based social networks and the provision of local social supports, which can be interpreted, with hindsight, as prescient attempts to evaluate the effectiveness of locality based social capital, as well as being among the earliest examples of the emerging 'therapeutic landscape' studies that would later come to the fore in the broader field of health geography.

All of this work would later be described as part of the 'first wave' of mental health geographies, which were involved primarily with locational analysis studies of mental illness and the mapping of the geographic distributions of mental disorders. The work represented a range of different theoretical and empirical approaches to the study of social problems, and it should be interpreted as a welcome, if rare, foray into important areas of social policy by human geographers. Much research is still conducted in the tradition of this 'first wave', although the results are anything but consistent. A recent study in Canada, for example, reinforced the much earlier findings about the propensity of people experiencing mental health problems to be more mobile than a matched cohort, and more likely to move ('drift') from the suburbs to the inner city. The contemporary literature also continues to feature studies that are able to identify clear cut and significant evidence that certain urban neighborhoods act as 'stressors' and are likely to influence different measure of mental illness. Interestingly, other recent studies, in particular, one conducted by an interdisciplinary team of researchers at Bristol University which included some well known geographers, have disputed such findings. In this study the authors found that neighborhood level factors accounted for less than 1% of the between individual variations in the data, leading them to conclude that place *per se* does not appear to have an important effect on the geography of mental illness.

Although such debates linger, and it seems likely that geographers will continue to offer their quantitative and mapping expertise to such studies, it was already clear – by the start of the 1990s – that another branch or 'wave' of mental health geographies was underway. In this new work, the research emphasis was shifting away from large scale, data based studies that either explored the effects of place/locality on measures of mental health, or involved spatial/structural critiques of mental health policies. The evidence suggests that the choice of topics and the methods in use in the newer work was influenced by work being conducted by geographers and nongeographers in the fields of cultural and disability studies. Some of the work falling into the category of mental health geography was beginning to focus on issues of identity, with the central questions being asked – shifting from a preoccupation with officially diagnosed categories of 'deviance' to a concern with 'difference'. Where the first 'wave' of mental health geographies had been dominated by empirical and quantitative examples of spatial science, the second 'wave' seemed to owe more to new developments within social theory: researchers were using mostly qualitative methods in small scale studies, and recognizing that mental health is structured by broader sociospatial forces.

This effectively shifted the mode of investigation from an analysis of space and geographic patterns to what has been referred to as "more interpretative engagements with the fusions of disability, identity and place." This reorientation, which accompanied rather than replaced research in the first 'wave' tradition, was probably being influenced in part by the 'cultural turn' within human geography, and in part by broader shifts in research foci suggested by contemporary work in postcolonial and feminist studies, postmodernism, and post structuralism. Studies within the second 'wave' can be characterized rather generally as falling into two very broad groups: one focusing on human differences and the way such differences are 'spatialized', in the sense of being played out and reacted to in space; the other group including studies that have been concerned with the ways people with mental health problems encounter space and place. A similar tendency can be identified at roughly the same time within the broader realm of health geographies.

Recognizing the value and appreciating the theoretical significance of such contributions should not imply that there was no value to the more positivist approaches associated with the first 'wave' of geographies of mental health, and as noted earlier, such studies continue to appear in the literature. It can be argued, however, that the earlier studies represented only a scratching at the surface of what could be achieved in this subfield. It also is evident that such studies provided the context for and the background to the newer and 'softer' approaches that were more attuned to qualitative studies of human suffering. The content matter of studies conducted in the second 'wave' was not always and never entirely 'new' to geographers: in fact, some of them can be linked quite clearly to earlier studies that focused on identifying everyday life stresses and the issues associated with 'coping in the community'. The newer research projects, however, often had an entirely new theoretical base; they were usually carried out at a much smaller scale, and the methodologies were invariably quite different.

Directions for the Future

A third 'wave' of mental health geographies that was thought to be on the horizon at the turn of the millennium does not yet appear to have materialized fully, although recent debates about the issue of policy 'relevance' within the broader realm of human geography indicate that some reorientation is being called for. What was significantly missing from the subfield of mental health geographies at that time (the early 1990s), and is perhaps still missing today, is the serious engagement of researchers with the important policy issues of the day. There are a number of reasons why geographers have steered clear of such work, including the traditional cynicism academics have for the 'dirty work' of politics and the more mundane work associated with 'applied' geography. Geographers with a spatial science and/or positivist leaning might have been easily put off, either in the realm of policy by a refusal to accept their relatively simplistic findings, or in the realm of theory by complaints about technological fixes, spatial fetishism, and ecological fallacies. At the same time, it seems likely that scholars engaged in the more reflexive work of the later era (the second 'wave') might have deliberately steered clear of large scale public policy debates, preferring to stay within the comfort zone of small scale qualitative studies.

As noted earlier, in the face of this reluctance, there have been calls from scholars working in several different subfields within human geography, for a 'policy turn' to match – and some would argue to compensate for the inadequacies of – the earlier 'cultural turn'. One pair of reviewers, for example, complained that although many human geographers are busily surveying and mapping geographies of 'exclusion', very few of them are actively involved in attempts to bring about significant change. This claim has been made quite specifically in the realm of urban geography in the recent past; and in an earlier review of the work being done in the health field, geographers have called for greater emphasis on 'critical geographies' of health in an attempt to suggest future research activities. What was involved here was a recommendation for new ideas and practices within the subfield, linked – by a commitment to emancipatory politics – to a broader call for progressive social change, and the application of critical theories to the pursuit of applied geographical research.

If the agenda recommended here is to be extended to the realm of mental health research, geographers will need to become more ambitious in the types of research they embark on, more willing to explore a variety of theories, and perhaps of most importance, flexible enough to move across scales when circumstances call for it. Geographers working toward a critical geography of mental health might consider operating at two fundamentally different but clearly interrelated scales: one looking at the larger theoretical picture, and considering the spatial/structural influences on mental health and ill health; the other operating at the individual level, and calling for geographers to be actively involved in changing the existing power relationships that constrain human agency.

The first suggestion here would involve interpreting the occurrence of mental health (and other social) problems within the larger context of economic globalization and neoliberalism. This is by no means a new suggestion for human geographers, who have in fact been writing about and warning of the pervasive spread of neoliberalism for well over a decade, but apart from a paragraph here and there little emphasis has been placed on the human consequences of neoliberal policies, or on their impacts on the access to and delivery of human services. One exception to this is the work of a group exploring the role of neoliberal policies in the creation and maintenance of 'social exclusion', while combining the spatial perspectives of urban planning with the service delivery perspectives of healthcare providers and practitioners. This work is focused in Britain, where the downside of the current capitalist triumphalism has seen an increasing number of people facing greater economic risk and instability. As the authors recognize, spatial issues have been at the heart of this process, as a result of ever widening flows of commodities, finance, and production, and the continual creation of new patterns of spatially uneven development and instability.

The argument here is consistent with David Harvey's view that neoliberalism should be thought of primarily as a political project which results in the elevation of elite status for a privileged minority, matched by the rapidly declining status of the working classes and the unemployed. Neoliberal theorists, in recommending and pushing their cause, have painted a dark picture of the impacts of earlier welfare oriented regimes, as expensive and wasteful, sapping individual enterprise and independence, limiting choices, and offering few in centives for self provision and personal responsibility. As neoliberal ideas gradually took hold, the role of the be nevolent state in providing income support and public services was eroded, and constraints on markets, such as trade unions, taxes, and subsidies, were eliminated. Social regulation was shifted effectively from the national level to the individual level. The assumption was that poverty would eventually be lessened by continuing economic growth, made possible by a freer and more cost effective allocation of resources, and guided by market principles and newly emerging entrepreneurialism. It was hoped that the freeing up of labor markets would provide new opportunities for the poor and unemployed, and that the overall benefits of faster economic growth, 'trickling down' in all directions, would ultimately reduce the level of poverty. As has been extensively documented, however, in many places this did not happen: in Britain particu larly, poverty increased massively, especially among people on fixed incomes, along with rising rates of unemployment and the emergence of an increasingly 'casualized' workforce. A large number of the un employed were effectively cut off from the labor market by social and spatial barriers, especially damaging in the case of deteriorating housing areas (like the 'sink' estates in British cities). The urban poor, rather than dis appearing, were increasingly on/in view, and often were perceived as an unruly and potentially dangerous 'underclass' group, of little use to the future economy, and acting as a permanent drain on it. Under such cir cumstances it would not be surprising to find rising rates of mental health related problems, made worse by the denuding of the landscape of service delivery that has accompanied neoliberal rule.

It is evident from a review of the research conducted within the field of health and mental health geography, that many (possibly a majority of) recent studies have begun from the assumption that a combination of poverty and inequality in some form or other is at the root of higher rates of health related problems in certain areas, which we can also assume are exacerbated by declining access to treatment and care options at the local level. It is clear, however, that to place the emphasis on the ex istence of inequality within a specific location (the local) ignores the reality of the much broader set of de terminants of health and welfare problems operating at the larger scale (the global). In a recent report and ana lysis of this issue at a global level, for example, one re searcher argues that the emergence of neoliberalism and its related social policies have had dire consequences for health and human welfare in a range of localities around the world. As the argument suggests: neoliberal phil osophies and policies to some extent encourage in equalities, which are believed to have a positive motivational force for labor. Such policies tend to be 'individualist' in the sense that they have often been as sociated with attempts to stifle and eradicate a range of collective or state actions in favor of pushing for indi vidual or family 'responsibility' and self provision, while at the same time they work to undermine the institutions that could reduce inequalities, and inhibit the forces that might be able to ameliorate or 'buffer' the worst effects of such inequalities through social provision. An empirical analysis suggests that in parts of the world where neo liberal policies have been resisted or applied with less fervor, people have better health, measured in a number of different ways, as well greater access to and a wider variety of health care services.

A theoretical stance on the corrosive role of in equality, however, is only one of the initial steps toward a critical geography of mental health. A review of the work conducted in contemporary social geography has re cently suggested that if geographers are to pursue policy relevant research, they need to begin by showing more interest in issues of power and politics, both in theoretical terms and through their own action oriented involve ments. The implications here for mental health geog raphies are quite promising. Individuals who have recurring mental health problems are hampered by a number of empirical realities, some of which have been the bread and butter concerns of geographers working within the framework of the earlier 'waves' of the sub discipline. For example, mental health services are almost universally poorly funded; mental illness remains mis understood and the mentally ill are still significantly stigmatized; and individuals with recurring mental illness are more often than not consigned to live with spatial isolation, underemployment, and a denial of their rights. In political terms the power of the mentally ill, and their freedom to act, is inhibited by the dispersion of political power among the many interest groups concerned with issues of mental illness, and also by the widespread exclusion of the mentally ill from full participation in economic, social, and political life. The net result is that mental health concerns rarely make it onto political agendas, and proactive policy options are not often considered. Also, as several mental health geographers have been finding, especially those working within the second 'wave' of studies described above, many people experiencing mental health problems have difficulty ar ticulating (and sometimes even recognizing) their own needs, and they tend to have very rudimentary knowl edge about what is available to/for them in the community.

One possible major step in the right direction – and this could serve as a clarion call of sorts for mental health geographers – would be to focus on enhancing the in dividual agency needed to challenge this fundamental lack of power and knowledge. There are a number of options that fall directly in the path of mental health geography research, including: efforts to focus on rights issues among the mentally ill; attempts to enhance national mental health legislation; improvements in the realm of advocacy, empowerment, and guardianship for the mentally ill; and better 'governance' in terms of higher accountability and quality control in the provision of mental health services. Consistent with such research activity would be more direct action, such as efforts to mobilize the mentally ill politically through voter registration programs and the creation of more active and powerful interest groups.

Conclusion

Research conducted by geographers during the past three decades has brought some crucial and cutting edge attention to the study of mental health issues, including the focus on the spatiality of difference, and the role of place in identity formation. This work has significantly enhanced our understandings of the importance of space in the realm of mental health and the provision of mental health care. In spite of the new and richly theorized approaches, mental health geographers, like the majority of human geographers, appear reluctant to become active in grass roots politics or in issues of mental health policy. In the meantime individuals with mental health problems seem destined to continue to experience socio/spatial 'exclusion', they will continue to have their human rights denied, and they will continue to suffer from their lack of political capital and economic freedom.

See also: Health Geography; Medical Geography; Welfare Reform.

Further Reading

Coburn, D. (2004). Beyond the income inequality hypothesis: Class, neo liberalism, and health inequalities. *Social Science and Medicine* 58, 41 56.

Dear, M. (1977). Psychiatric patients and the inner city. *Annals of the Association of American Geographers* 67, 588 594.

Dear, M. (1992). Understanding and overcoming the NIMBY syndrome. *Journal of the American Planning Association* 58, 288 299.

Dear, M., Clark, G. and Clark, S. (1979). Economic cycles and mental health care policy: An examination of the macro context for social service planning. *Social Science and Medicine C* 13, 43 53.

Faris, R. E. and Dunham, H. W. (1939). *Mental Disorders in Urban Areas*. Chicago, IL: University of Chicago Press.

Giggs, J. A. (1973). The distribution of schizophrenics in Nottingham. *Transactions of the Institute of British Geographers* 59, 55 76.

Gough, J., Eisenschitz, A. and McCullouch, A. (2006). *Spaces of Social Exclusion*. London: Routledge.

Harvey, D. (2005). *A Brief History of Neoliberalism*. Oxford: Oxford University Press.

Imrie, R. (2004). Urban geography, relevance and resistance to the 'policy turn'. *Urban Geography* 25, 697 708.

Jones, P. A. (2007). Researching the geography of mental wellbeing. *Journal of Mental Health* 16, 439 443.

Kearns, R. A. and Gesler, W. M. (1998). *Putting Health into Place: Landscape, Identity and Well Being*. Syracuse, NY: Syracuse University Press.

Kearns, R. and Moon, G. (2002). From medical to health geography: Novelty, place and theory after a decade of change. *Progress in Human Geography* 26, 605 625.

Kelly, B. D. (2006). The power gap: Freedom, power and mental illness. *Social Science and Medicine* 63, 2118 2128.

Kitchin, R. M. and Hubbard, P. J. (1999). Research, action and 'critical' geographies. *Area* 31, 195 198.

Pain, R. (2006). Social geography: Seven deadly myths in policy research. *Progress in Human Geography* 30, 250 259.

Parr, H. (2004). Medical geography: Critical medical and health geography? *Progress in Human Geography* 28, 246 257.

Parr, H., Philo, C. and Burns, N. (2004). Social geographies of rural mental health: Experiencing inclusions and exclusions. *Transactions of the Institute of British Geographers NS* 29, 401 419.

Philo, C. (1987). 'Fit localities for an asylum': The historical geography of the nineteenth century 'mad business' as viewed through the pages of the asylum journal. *Journal of Historical Geography* 13, 389 415.

Philo, C. (2004). *A Geographical History of Institutional Provision for the Insane from Medieval Times to the 1860s in England and Wales: The Space Reserved for Insanity*. Lampeter: Edwin Mellen Press.

Propper, C., Jones, K., Bolster, A., Burgess, S., Johnston, R. and Sarker, R. (2005). Local neighbourhood and mental health: Evidence from the UK. *Social Science and Medicine* 61, 2065 2083.

Smith, C. J. (1975). Being mentally ill in the asylum or the ghetto. *Antipode* 7, 53 55.

Smith, C. J. (1981). Residential proximity and community acceptance of the mentally ill. *Journal of Operational Psychiatry* 12, 2 12.

Smyth, F. (2005). Medical geography: Therapeutic places, spaces and networks. *Progress in Human Geography* 29, 488 495.

Wilton, R. (2004). Putting policy into practice? Poverty and people with serious mental illness. *Social Science and Medicine* 58, 25 39.

Wolch, J. and Philo, C. (2000). From distributions of deviance to definitions of difference: Past and future mental health geographies. *Health and Place* 6, 137 157.

Wolpert, E. and Wolpert, J. (1974). From asylum to ghetto. *Antipode* 6, 63 76.

Relevant Websites

http://www.ges.gla.ac.uk:443/research/groups/humangeography
For an introduction to the work of one of the leading researchers in this sub field, Dr Chris Philo, see the website describing the research group he is involved with at the University of Glasgow: the Human Geography Research Group (HGRG).

http://www.dundee.ac.uk
Social Inclusion and Mental Health, which describes a three year research program funded by the UK Economic and Social Research Council (from 2004 2007; grant number RES 000 27 0043 "Embodied Geographies of Inclusion: Placing Difference"). The chief researcher on this project was Dr Hester Parr (Department of Geography, University of Dundee).

Mental Maps

S. Bell, University of Saskatchewan, Saskatoon, SK, Canada

Glossary

Behavioral Geography The branch of geography concerned with human spatial behavior and the processes that underlie spatial decision making and spatial actions.

Cartography The field of geography devoted to the creation of maps.

Cognitive Map The internal spatial representation of the world as we know it, and the accompanying affective responses that this knowledge evokes.

Direct Mapping Process of having a person draw a map or map-like figure, closely related to sketch mapping.

Georeferencing The act of assigning geographic coordinates to spatial information. In mental and sketch mapping this process can allow the integration of hand-drawn information with other existing geospatial information.

Indirect Mapping Process of using responses to nonmap drawing tasks to create a map-like representation of a geographic space.

Mental Map Preference surfaces generated by asking people about their attitudes and perceptions of different places.

Sketch Maps Maps drawn or created by hand or other free-drawing techniques.

Introduction

Mental maps are external map like products used to represent what we know about the world. Geographers have long been interested in how we understand the world around us. Knowledge of the spaces and places we experience directly and indirectly on a daily basis is the foundation for how the world is known and how spatial decisions are made. Much of what we know about the world and the feelings this knowledge elicits (both from memory and as a result of experiences as they happen) has at least some basis in our 'cognitive map', the internal spatial representation of the world as we know it, and the accompanying affective responses that this knowledge evokes. The world in which we live can produce visceral responses related to places we have visited or about which we have learned. Furthermore, spatial information as coded and recalled in memory can be quite different from reality in terms of accuracy and completeness. That both spatial and place based representations stored in memory and acquired through experience can be so divergent from reality makes studying this knowledge compelling. One approach that geographers and others have used extensively is the process of mapping what is known and felt about the world. This mapping can be done in variety of ways including both direct and indirect map creation. Direct map creation involves drawing or building a map or map like representation, whereas indirect one involves the creation of a map based on responses to a series of questions or tasks.

Maps are simplified representations or models of referent objects. In geography the referent object is al most always the Earth or some subset of the Earth's surface. In addition to being simplifications of the phys ical and human nature of the Earth, maps also include unique cartographic elements that are added to increase the utility of the map itself. These include but are not limited to labels, symbols, inset maps, legends, scale bars, north arrows, etc. Most people think of maps as paper renditions of a geographic area. These remain popular and are probably still the most common type of map but new types of maps have been emerging that are dynamic and take advantage of new computer, communication, and storage technology. Including these new map types, there are three broad traditions within cartography: the storage tradition, the communication tradition, and the analytical tradition. These traditions are mentioned here because they are also represented in our 'cognitive map'; we use our cognitive map to store information and knowledge about the world we know; we use our cog nitive map as the basis for communicating information about the world as we know it; and we solve problems everyday by analyzing the spatial information we have stored in our cognitive map. Understanding the nature of our cognitive map is one of the central concerns of be havioral geography, that branch of geography dedicated to understand human spatial behavior and human spatial decision making.

The process of using and creating maps that represent our knowledge of the world is known by many terms. Mental maps, sketch maps, diagrams, multidimensional scaling, verbal descriptions, and navigational tracking are all techniques that have been used to develop external representations (outside the cognitive map) of the world as it is known by individuals and groups. Each technique provides unique information about how we perceive, code, integrate, and recall information and knowledge about the world around us. None should be considered exclusively preferred over the others and each should be considered in light of the geographic question of interest. Some techniques, such as sketch mapping, drawing, and base map labeling, record responses in a map like form ("draw a map of the Brown University Campus" or "label these points on the map"), whereas others rely on de cidedly non map like responses ("which of these two locations is closer to X" or "list these places in order of preference") in order to derive a final product that looks like a traditional map.

Mental Maps and Sketch Maps

Peter Gould and Robert White were two of the first geographers to use the term 'mental map'. While 1974 marks the publication date of Gould and White's seminal book *Mental Maps*, it should not be considered the date by which we judge the term's age. There is no doubt mental mapping was entering the discipline's lexicon well before 1974 and had been used to refer to things other than physical maps. In fact, there is some confusion among the terms cognitive map, mental map, mental image, internal cognitive representation, and others. Gould used the term mental map to refer to preference surfaces generated by asking people from one place to rate their preference for selected other places. The preference ratings were then used to pro duce preference surfaces that were represented on ac curate cartographic base maps. These products are well known and have logically resulted in the assignment of the term mental map to this type of external represen tation of our cognitive map (**Figure 1**). The responses themselves were not map like but were associated with individual places which supported the creation of representations of aggregate responses that were very map like. Thus, the starting point for this discussion of mental maps will be a map like externalization (pre sented in true cartographic form) based on nonmap responses.

Many techniques share Gould and White's approach to producing a representation of what is known about the world. Some approaches, such as multidimensional scal ing, produce map like representations based on nonmap input. By having respondents indicate paired proximities an algorithmic technique can be used to create a metric spatial array (actual locations on a two dimensional (2 D) representation) of points that corresponds to a nonmetric assessment (judgments of the proximity pairs of places on a subjective scale) of the spatial relationships among a set of points. For geographers, this was useful because the output was a metric representation of real world points that was different from the actual distribution of those points, suggesting that the differences represented the deviation of the cognitive map from the real world dis tribution of the places in question.

A technique widely used both within and beyond geography has been that of sketch mapping. Sketch mapping consists of drawing map like representations under constrained and unconstrained conditions. Sketch

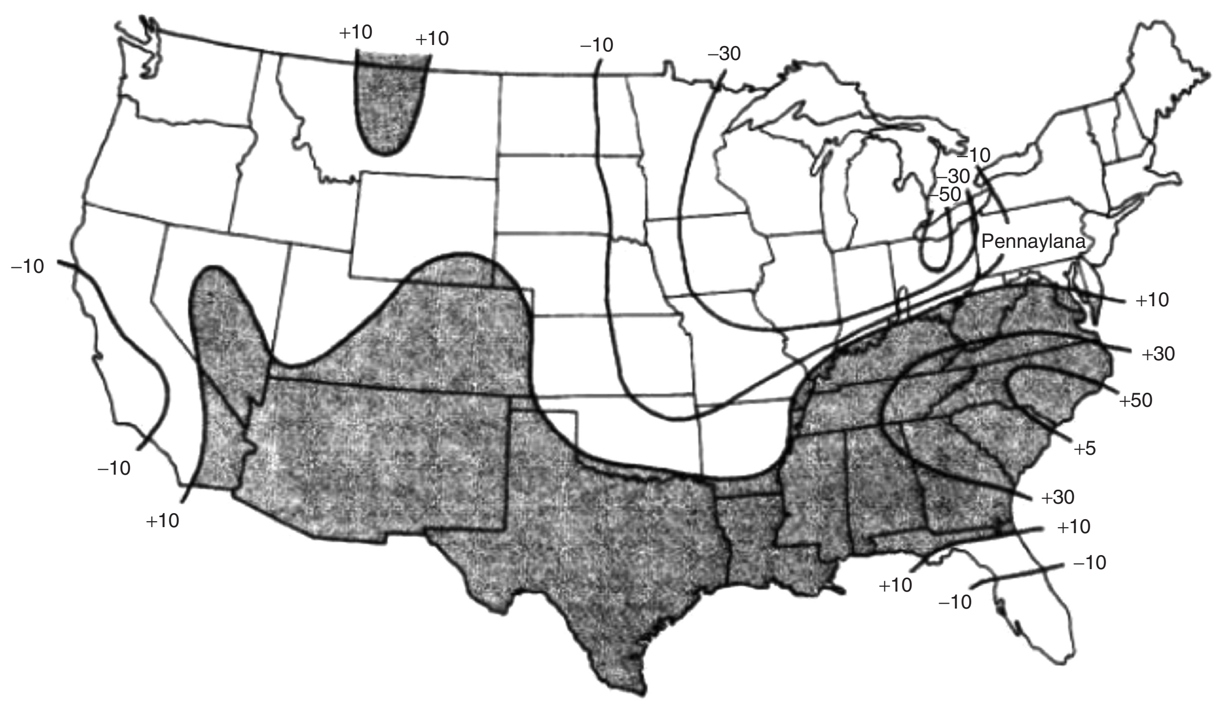

Figure 1 Mental map of North Carolina geographic preference.

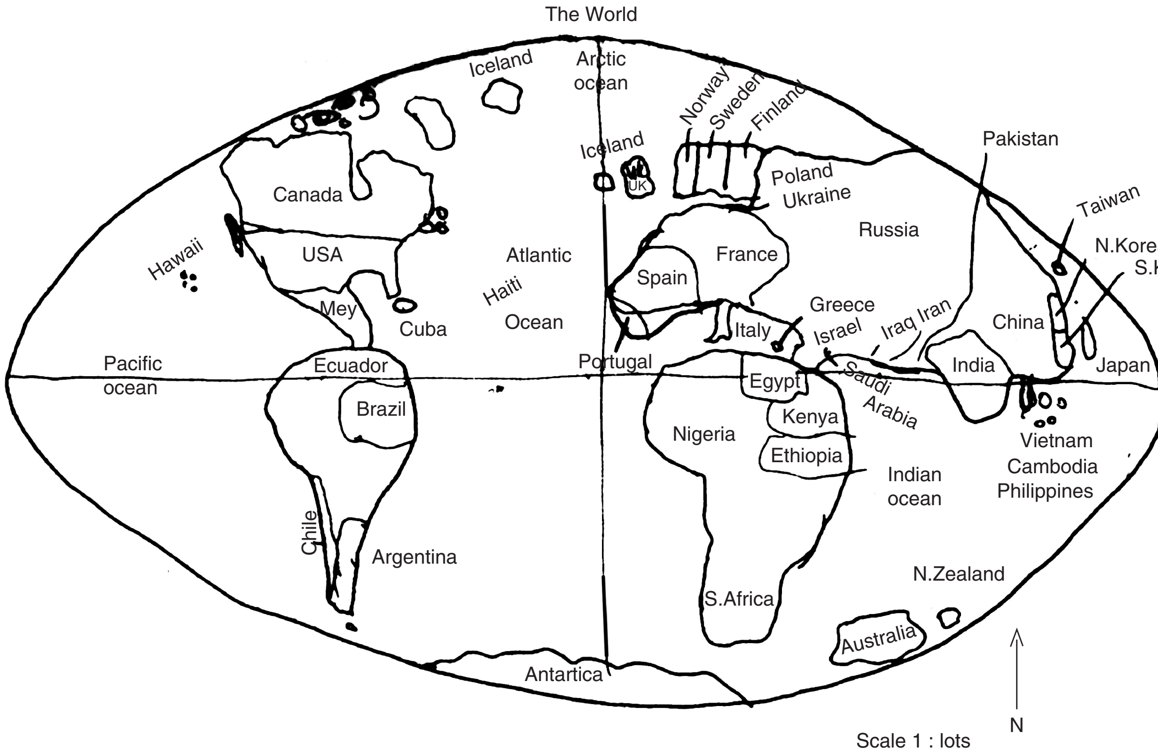

Figure 2 A free sketch of the world resulting from the instruction "draw a map of the world labeling as many countries as you can." This is an example of a direct mapping activity in which the map 'product' is a direct result of the task the person was asked to complete.

mapping can range from free sketching tasks during which maps are drawn on blank paper with limited instructions (such as "draw a map of your hometown" or "draw a map of the world labeling as many countries as you can") to the placement of a symbol or symbols on a cartographic map (**Figure 2**). While the latter is the product of a person putting pen to paper, it involves very little 'sketching'; it does not provide a map based representation of the individual's spatial or geographic knowledge. It does, however, indicate where in the world that individual thinks something is located, which is potentially more useful (and interesting) than a sketch map that could be impossible to interpret. While sketch maps have a history almost beyond knowing (the first sketch maps could have been drawn in the sand or scratched onto rock walls), Kevin Lynch's *The Image of the City* was the first thorough examination of drawings of urban spaces that helped delineate what people know about their home cities and more importantly provide a classification scheme for this knowledge.

Does an Ideal Term Exist?

The term mental map has been introduced as the representation of a person's, or group's, preference for different parts of the world and is an example of one approach that uses the mapping paradigm for the examination of our internal understanding of the world in which we live (our cognitive map). That it refers to an explicit, and elegant, method for representing people's preferences for different places and how these preferences vary depending on where one lives should not be muddied by referring to all map based approaches as mental mapping. However, the more universal term, sketch mapping, through the inclusion of the term sketch in the name, might imply that drawing is a necessary component. Most of these techniques have in common the objective of elucidating the nature of spatial knowledge held by the person performing the task. Researchers in geography and other disciplines use these techniques for a variety of reasons. Initially the techniques are used to uncover the spatial knowledge held by the individual and the structure and arrangement of that information. In turn, this can be used to understand the processes that are the foundation for how the human brain encodes, stores, accesses, and uses spatial knowledge as well as the state of that knowledge (accuracy, abundance, bias, etc.). The researcher may or may not be interested in a specific place on the Earth's surface.

Mapping for Knowledge Assessment

As indicated above, mapping techniques can add to our understanding of how information is coded, stored, accessed, and used by the human brain. Mapping tech niques can also help us evaluate what is known about the world; the 2006 National Geographic–Roper survey on geographic literacy is an example of how sketching and mapping can be used to elucidate what is known about the world by individuals and groups. In the actual Na tional Geographic–Roper survey, no maps were created but questions were asked regarding information on several maps which could be used to produce ‘maps of understanding’. Unfortunately no such maps have been produced based on the National Geographic–Roper survey results. The results in the final report of the 2006 Geographic Literacy Study are typical of the type of information that can be gained from mental and sketch mapping exercises targeted at distilling what is known about world, regional, or local geography. The benefit of this ‘no drawing’ approach to mapping is that the experimenter can control potentially confounding vari ables. Such variables include drawing ability, relative scale, and locational. By having participants label maps or identify country and/or geographic outlines, a valid measure of geographic knowledge can be established. Such techniques have been used to determine the relative lack of geographic knowledge held by North American school children in comparison to school children around the world and even differences between age groups and genders.

Sketch mapping can also be used as a training aid in geographic education. By increasing student skill and experience with maps and mapping, their ability to use such tools for developing geographic knowledge also increases. Geography for Life, the national standards for geographic education, outlines several ways that map ping can be incorporated into the essential spatial and geographic themes for acquiring and improving geo graphic knowledge. Building activities that incorporate mapping is an intuitive approach to increasing geo graphic literacy and can be done in conjunction with subject matter that is cognate to geography, art, math, civics, writing, science – just a few examples of academic areas that have some relationship to geography and mapping.

Beyond the use of mapping activities in education, the study of how spatial information is acquired and used in everyday life is a central concern of geographers inter ested in spatial cognition, spatial behavior, and cognitive mapping. A core goal of each is an improved under standing of how spatial knowledge is structured in the brain and what processes operate on this knowledge in order to solve spatial and geographic problems. These problems generally involve the mundane day to day activities we all perform, such as traveling from one place to another (either local or distant), giving directions to others, or deciding what shops, services, and other lo cations to visit and in what order.

Other Applications of Mental Maps and Sketch Mapping

Geographers use several general types of mapping tasks that fall within the scope of mental mapping and sketch mapping. As described above, mental mapping is the process of collecting, in a variety of ways, responses re lated to geographic places with the goal of generating a map like representation, while sketch mapping is the process of producing spatial products (drawings, models, etc.) that represent geographic places. For the most part mental maps look more like traditional cartographic maps and communicate something about what is known about the places on those maps; sketch maps usually do not look like formal maps, a characteristic that is the result of relying on the artistic ability of the person creating the map as well as their (limited) knowledge of the world. This necessarily constrains what a sketch map can be used for; it must involve the presentation of spatial information in a map like form. Mental mapping on the other hand can focus more on the nonspatial nature of places.

A further distinction between sketch and mental mapping would be the extent to which the participant is directly/explicitly involved in the creation of the map; direct model (map) building requires the participants to create their own map, whereas indirect model (map) building is the result of responses to questions or through the creation of lists of the relationships among places. Several examples of each have been indicated already: direct model (map) building can be accomplished through map drawing, 3 D model building, and computer drawings, as well as labeling and augmenting outline maps. Indirect model (map) building can be accom plished through means such as simple lists, preference lists and ratings (as by Gould), paired comparisons, Likert scale ratings, verbal descriptions, and picture taking.

Implication for the Use of Sketch Maps

Sketch mapping has been studied in many ways and used to study many topics. The reliability of the general technique has been confirmed, suggesting it is a viable means of testing and retesting geographic understanding. In other applications of the technique both drawing and 3 D model building has been used to better understand how both blind and sighted people acquire knowledge about their surroundings. Furthermore, the nature of geographic knowledge held by people around the world and of different cultures has been explored in an im portant series of studies by Aaarinen which explores not only the ability of sketch maps to expose what is known about the world but also represents an initial examination of the role that drawing ability and shape recognition play in the use and generation of sketch maps. Whether a sketch mapping or mental mapping activity is a valid method for a given research question is less well known. Very little research has been reported on the validity of sketch mapping for examining or evaluating geographic knowledge. A critical component of any sketch mapping task, and one not shared with less artistic mapping ap proaches, is the role that drawing ability or artistic ability plays in one's ability to communicate or represent what they know about the space being drawn.

Geographic researchers interested in using a sketch mapping approach should carefully consider several factors related to their topic of study as well as what their expectations are in regards to the mapping output. While researcher must remain aware of the costs associated with confounding variables and the validity of their instru ment for a given study, there is no doubt that sketch mapping offers the enticing promise of a rich array of data as well as the visual and map like nature of the output. For these reasons it is important that the style of mapping exercise be carefully considered with respect to spatial limits and type of task. Limits can include boundaries (neat lines) and preliminary or seed infor mation (country outlines, landmarks, street networks, etc.) that will guide the participant and provide add itional tools during analysis. For instance, if a researcher is interested in how much knowledge a group has of world geography he or she might consider a sketch mapping task that presents only country outlines. This will provide an outline for the drawing of mountains, rivers, lakes, or other features, as well as the labeling of countries without the potentially confounding effect of participant drawing ability. However, such a task will not allow for the examination of participants' knowledge of country and continent shape and necessarily provides information that might give some participants clues to features they otherwise would not have recalled.

When researchers have specific and narrow objectives they should use discretion in the development of the mapping task. Limiting freedom becomes a valuable tool so that confounding variables can be eliminated. This might involve asking participants to include only specific information, such as roads or navigational landmarks in a task associated with spatial knowledge acquisition in a recently experienced environment, or to ask them only about specific parts of the world in a world map drawing or labeling exercise. It is in these situations that indirect modeling tasks might prove more efficient.

Using Mental Maps

Researchers often use nonmapping (indirect model building) techniques to collect data that has a spatial dimension and that will lend itself to being mapped. This might include country names recalled or labeled on an outline map that can be enumerated and used to generate a choropleth map of knowledge of world countries. An other example might be lists of addresses visited during a day out completing errands that can be geocoded and mapped. Geo coding is the process of assigning co ordinates from a continuous global coordinate system, like latitude and longitude to street addresses. If the latter data are presented in order, the output map can also communicate the pattern of travel through the environ ment. Interestingly, neither of these examples tells us much about the nature of our cognitive map, although they do tell us a little bit about how it is used and give us a small sample of the knowledge stored there.

Sketch maps can be scored in many ways. Nominal data can be recorded simply by noting the presence or absence of certain features on the map, particularly those features that are important for solving spatial or geo graphic tasks or features related to the expressed purpose of the map. Enumerating the types of features or the number of different types of features will provide an additional level of data detail. Direct measurement of the map is also possible. If a path or other real world object has a certain shape or has edges that meet at specific angles, the measured angle on a drawn map can be compared to the actual angle and error between the two can be recorded. A more basic technique would be to compare with real world relationships the spatial rela tionships between and among features drawn on a map. In Canada, the province of Saskatchewan is to the east (right) of Alberta and to the west (left) of Manitoba; the presence or absence of each of these relationships can be recorded to confirm that Saskatchewan was placed in the correct 'topological' location in Canada. Furthermore, Saskatoon is a city in Saskatchewan; confirming that the point feature Saskatoon is contained by the polygon Saskatchewan will produce a viable score for that piece of information. If an overall 'map quality' score is desired, independent map raters can be used to assess each map's 'quality'. These assessments should be used carefully as there are opportunities for bias in terms of what consti tutes good and bad sketch maps. In general, the use of mapping techniques in the assessment of geographic knowledge should be undertaken with great care and thoughtful consideration regarding the nature of the task and the goal of the research.

The Future of Mapping Our Geographic Knowledge

Maps and mapping exercises have been used to better understand what we know about the world and have been used experimentally by geographers and nongeographers in a variety of setting to learn more about our cognitive map. Much of this potential will be realized as new ap proaches to mapping our knowledge emerge. Ongoing developments in geographic information systems (GISs), computations cartography, and computer technology represent distinct opportunities for researchers interested in mapping as a research tool. Drawing maps on tablet computers will soon support the recording of not only the spatial dimension but the temporal as well. Such advances will reveal important information about the sequence of events that produce a finished sketch map. Existing GIS software tools can already be used to georeference elements of drawn maps and if done carefully can be used to integrate maps from multiple individuals to compare differences and elucidate com monalities. Some of these techniques are already in use and have produced a better understanding of the level of uncertainty in our communal cognitive representations of everyday geographic terms like downtown or neighborhood.

The state of the art of mental mapping and sketch mapping today will serve as a fertile ground for future mapping applications and research. The historic interest in sketch mapping and our mental representations of the world along with the long standing importance of maps to geography serves as the best evidence for the con tinued use of new and innovative techniques to bridge what is known (cognitive map) and what can be com municated (sketch and mental maps).

See also: Behavioral Geography; Cognitive Geography.

Further Reading

Downs, R. M. and Stea, D. (1973). *Image and Environment: Cognitive Mapping and Spatial Behavior*. Chicago, IL: Aldine.

Geography Education Standards Project (US), American Geographical Society of New York, Association of American Geographers, National Council for Geographic Education, and National Geographic Society (US). (1994). *Geography for Life: National Geography Standards 1994*. Washington DC: National Geographic Research and Exploration.

Golledge, R. G. and Stimson, R. J. (1997). *Spatial Behavior: A Geographic Perspective*. New York: Guilford Press.

Gould, P. and White, R. (1974). *Mental Maps*. Harmondsworth: Penguin.

Jacobson, R. D. (1998). Cognitive mapping without sight: Four preliminary studies of spatial learning. *Journal of Environmental Psychology* 18(3), 289 305.

Kitchin, R. M. (1994). Cognitive maps: What are they and why study them? *Journal of Environmental Psychology* 14(1), 1 19.

Kitchin, R. M. and Freundschuh, S. M. (eds.) (2000). *Cognitive Mapping: Past, Present, and Future*. London: Routledge.

Lynch, K. (1960). *The Image of the City*. Cambridge, MA: MIT Press.

Montello, D. R., Goodchild, M., Gottsegen, J. and Fohl, P. (2003). Where's downtown? Behavioral methods for determining referents of vague spatial queries. *Spatial Cognition and Computation* 3(2 3), 185 204.

Moore, G. T. and Golledge, R. G. (eds.) (1976). *Environmental Knowing: Theories, Research and Methods*. Stroudsburg: Dowdenm Hutchinson and Ross.

NGS, R. P. A. (2006). *Final Report: National Geographic Roper Public Affairs 2006 Geographic Literacy Study*. Washington, DC: National Geographic Education Foundation.

Saarinen, T. (1973). Student views of the world. In Downs, R. & Stea, D. (eds.) *Image and Environment: Cognitive Mapping and Spatial Behavior*. Chicago, IL: Aldine.

Taketa, R. (1996). Using field sketch mapping to teach basic mapping concepts in elementary school geography. *Journal of Geography* 95(3), 126 130.

Tuan, Y. F. (1975). Images and Mental Maps. *Annals of the Association of American Geographers* 65(2), 205 213.

Métropole d'équilibre

G. Burgel, University of Paris 10, Nanterre, France

Glossary

Centrality Major concept of urban geography concerning both the material densification of space at various scales (town center, monocentrality or multicentrality of urban regions, capital, or national metropolis) and the polarization of economical, social, and political powers.

DATAR – *Délégation à l'Aménagement du Territoire et à l'Action Régionale* (in France) State institution founded in 1963. For over 40 years, it has been the inspiration and instrument behind all French public policies regarding the voluntaristic organization of the national territory. A major milestone in the terminology, it was replaced in 2006 by the DIACT (*Délégation Interministérielle à l'Aménagement et à la Compétitivité des Territoires*).

Decentralization Generic term designating both the deliberate dispersion of economic activities from the initial concentration hubs (industrial and tertiary decentralization) and the transfer of skills from the central state institutions to subordinate territorial levels: regions, departments, and municipalities. In the latter case, the term deconcentration is often preferred.

Metropolitan Area Extension of the metropolis which tends to combine the problems of internal spatial organization and the functional role of the large town.

Metropolization Complex urban concentration process involving the size of a large town, its economic and demographic growth rate and, more and more frequently if not always, its ability to control and manage companies, especially in terms of international and financial attractiveness.

Urban Framework System connecting the various towns of a given territory, based on their relations in terms of size, quantity and quality, services, and dimension of attractiveness space.

Over the last half century, the metropolitan growth centers have proved extremely successful in terms of fashion, if not in terms of efficiency, especially in Europe with the regional planning and development policies whenever the aim was to deliberately stop the hyper trophia of a national capital considered to be too invasive. One after the other, Hungary with Budapest, Rumania with Bucharest, and Greece with Athens have tried to provoke urban oppositions and territorial rebalancing programs faced with the sharp growth of their main growth centers, both demographic and economic. The reference model remains France; however, where from 1963, the state tried to impose metropolitan growth centers to counterbalance the assumed over importance of Paris. Experience is an interesting laboratory for emergence of a planning ideology, creative implemen tation of spatial intervention tools and shifts in action, or even complete about turns, caused by changing eco nomic mechanisms, especially the globalization of ex changes and the construction of Europe.

Exceptional Emergence Conditions

Structurally after World War II, France represented an ideal setting for the invention of metropolitan growth centers. The urban system was durably marked by structural characteristics and anomalies in the normal distribution of town sizes, which had drawn the attention of observers for a number of years. 'Reilly's law' draws up a regular relation between the rank and the size of towns in a country, which is generally a straight line on x/y coordinate axes. In the case of France, the weight of the capital region is so large that we observe a sharp drop of the curve on reaching the second town in the classifi cation. This regularity of 'Reilly's Law', which struck all observers, has never really been explained, but its ex ceptions have led to questioning and even concern. First, the weight of the capital region – in 1954, in a France that, under the effect of the baby boom and the start of the 'thirty glorious years' (strong economic growth from 1950–73), had just crossed over the historical threshold of 40 million inhabitants, the Paris agglomeration (Paris and its suburbs) already exceeded 7 million residents. Immediately afterwards, however, none of the major provincial towns reached 1 million inhabitants (Lyon, Marseilles, Lille). Conversely, there were nearly a hun dred small and middle sized towns, which had from 50 000 to some 200 000 inhabitants at the time, fairly evenly distributed over the national territory, and which formed the true fabric of the French spatial framework. The architecture of this original device has not signifi cantly changed today, if we consider a relatively uniform translation movement due to the increased urban growth from 1950 to 1980: the population of the capital had flattened out at 10.6 million inhabitants, all the largest regional metropolises had now reached 1.3 million in habitants; and numerous medium agglomerations had grown from 100 000 to 500 000 residents mainly due to the fusion of the expansion areas of the towns forming them. These remarkable steady situations were the result of very long term historical processes: the millenary creation of a nation state around its political and eco nomic capital ensured its hypertrophic growth while sterilizing the ascension of competing metropolises; the departmentalization sought by the 1789 revolutionaries, to guarantee the equality of territories and above all the relay of the central influx, created and strengthened a dense network of prefectures. The book by Jean François Gravier *Paris et le Désert Français* systemized and popu larized these dissymmetries for decades.

This original situation was theorized in the 1950s through the publication, delayed due to the unfavorable international climate, of analyses on the hierarchized urban networks. The American W. J. Reilly modeled the attractiveness of the commercial framework in American towns. A. Lösch made a general hypothesis on the or ganization of economic spaces. It was the German W. Christaller, however, who found the greatest success in geographical studies in France. His empirism, with his analyses located in Bavaria, seduced the partisans of a French School of Geography, fervent believers in the principles of regional monographs. While his system of territorial operation – seven level hierarchy of urban centers, gravitational attractiveness of services, and hex agonal nesting of town polarization areas – attracted all those who, since the publication of *Les Villes* in 1947 by Georges Chabot, had endeavored to set up a system of general urban geography through the functions of the agglomerations. In France, these geographic approaches were corroborated in particular by studies in the field of spatial economy (François Perroux, Raymond Boude ville), which emphasized the regional polarization gen erated by innovation and investment, often of public origin.

This gave rise in the early 1960s to an intense movement of intellectual and political activity, com bining indiscriminately trends of thought often inspired by Marxism and a regal voluntarism initiated by the power of General de Gaulle, then at the peak of his power (creation of the 5th Plan). Some, in particular supporters of the geographer Pierre George, strived to prove the historical and social origin of hierarchized urban networks. The most famous were Michel Roche fort and Raymond Dugrand. At the same time, engineers and senior civil servants – for example, Paul Delouvrier, appointed 'proconsul' in Algeria at the end of the French colonization before becoming the leading figure behind the Paris new towns – were the builders of a voluntary policy aimed at correcting the imbalances generated by unconstrained capitalistic development. They all finally came to the conclusion that the idea of a certain degree of controlled spatial egalitarianism would guarantee a more stable social equity and a safer redistribution of the growth product. This consensus was reflected in the Hautreux, Rochefort, Lecourt report, which represented the real birth certificate of the metropolitan growth centers, the initial charter of the DATAR (*Délégation à l'Aménagement du Territoire et à l'Action Régionale*), which was in the process of being set up and which used it as a basic theoretical reference for decades.

Major Accomplishments of the DATAR

Statistical analyses and comparisons with the assumed model for the creation of urban networks led to the identification of the major exception in the French sys tem: the lack of demographic size and structuring func tions in the large regional metropolises (**Figure 1**). The solution was to strengthen and stimulate them through infrastructures and large public investments to 'coun terbalance' the Parisian capital. Eight urban areas were therefore selected, partly because they were the next largest after the Paris agglomeration and partly because they were relatively distant from the capital and well distributed across the national territory, aspects con sidered to be an additional guarantee of spatial efficiency. From north to south and east to west, Lille, with its ties in the Roubaix and Tourcoing conurbation, the Nancy Metz bipole in Lorraine, Strasbourg in Alsace, Lyon and its two extensions Saint Etienne and Grenoble, Marseilles on the Mediterranean Arc, Toulouse and Bordeaux in the southwest, and Nantes Saint Nazaire on the Atlantic coast.

Although ambitious, the project was unambiguous. The primary objective was to open up the provinces, still considered to be off centered and isolated by the star pattern formed by the royal roads and then the railways and further reinforced by the creation of the motorway network, radiating out from Paris. The decision was made to connect Paris to the major provincial metropolises by fast rail transport lines. This was the start of the French TGV (high speed train) adventure, admired by the entire world. The first TGV line, which would link Paris and Lyon in just 2 h, was launched in 1974 and inaugurated by President François Mitterrand in 1981. Thirty years later, all the metropolitan growth centers, except Tou louse, are in the process of being connected to the TGV network: in June 2007 the latest town, Strasbourg, one of the two European capitals, is 2 h 20 min from Paris. The TGV represented an outstanding technological achieve ment considerably reducing time and space, placing France at the center of the European means of interurban transport to the United Kingdom (Eurostar), Benelux and Germany (Thalys), Switzerland and Italy, and the Spanish peninsula. Paradoxically, however, the original intention of the advocates of the system has been di verted: the fast links obviously run in both directions and extending the TGV network increases the Paris

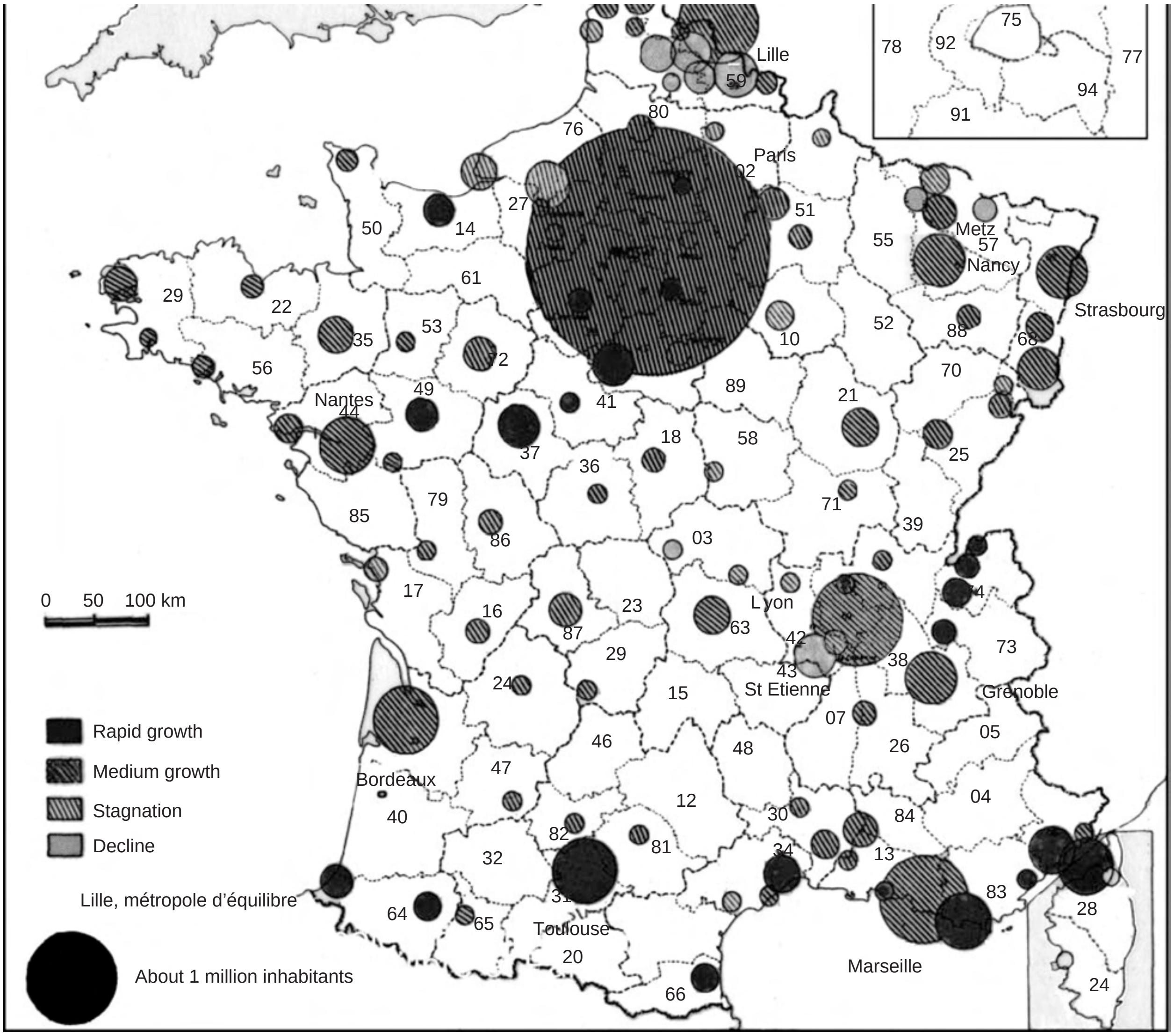

The French urban network (2000)

Figure 1 The French urban network (2000).

centrality, making the French capital the hub of European exchanges.

The same uncertainty can be observed in the rationales applied to the metropolitan growth centers. The fathers of the theory of hierarchized urban networks considered services to be the driving force making the towns more attractive. Typically, in a way that reflects the financial crazes of the time, often forgotten today, the DATAR chose to place emphasis on the leading role played by industry. In a rather late analysis of its action, experts pointed out that "urban growth, the requirements for services and administration and the determination to reverse worrying sociopolitical imbalances, in addition to the criteria of a strictly industrial economic rationality, justified a policy promoting the major towns." It is clear, however, that the period of this policy focusing on metropolitan growth centers corresponded to a second industrial age and especially to a need to increase the number of sites to match the extension of heavy industry, surrounded by a rich and complex sociospatial fabric – research centers, universities, and urban equipment. This example demonstrates a clear lack of foresight in the economic dynamisms already implemented and in the ability of the state mechanisms to promote or adapt them. From the very start of the 1970s, the number of industrial crises (coal mining, iron and steel, textile, automobile) increased, tertiary activities took over from the manufacturing industries, and in particular, company competitiveness escaped the yoke of the political authorities.

Still, in the same way as the large infrastructural equipment built the framework of the national territory, decentralization of the public services formed the keystone of the state's attempt to promote the attractiveness of the metropolitan growth centers. Often foreshadowed many years earlier, the achievements continued over more than three decades, regulated by budget restrictions or resistance from the professional spheres. The central administrations, health, and culture therefore benefited from provincial sites obtained increasingly easily, the thinner the symbol or the faster the pace of requirements and creations: civil status of French people born abroad at Nantes, *Hôpital des Grands Brûlés* (serious burns hospital) and *Centre International de Recherche sur le Cancer* (International Agency for Research on Cancer) at Lyon or the *Villeurbanne TNP* (Popular National Theater) near Lyon. However, the tensions were somewhat more tangible for the Paris *Grandes Ecoles* and research. While Lille readily accepted the *Institut Supérieur d'Électronique* and Toulouse, obviously the *École Nationale de l'Aviation Civile*, the transfer to Lyon of some of the *Écoles Normales Supérieures* was made much more reluctantly. In the 1990s, the transfer of the *École Nationale d'Administration* to Strasbourg resulted for many years in split site education, costly for teachers and students alike, between Paris and the Alsace metropolis, in the same way as the European parliamentary sessions between Brussels! It is true that this represented an attack – justifiably(?) – on the actual bases of the central State.

In addition, the anti Paris policy, led by the DATAR, was to be a catalyst as active in the direction taken as direct promotion of the metropolitan growth centers. The implementation in 1955 of an 'approval' procedure authorized the administration (decentralization committee) to create or extend activity premises in the capital region, according to a standard of complex and permanently raised thresholds (e.g., 500 m^2 initially for industry, and then 1500 m^2 in 1972 and 3000 m^2 in 1985). In fact, these inflexible directives, much criticized by the suburban elected representatives who once again perceived them as the source of their current economic and social drawbacks, all the more since they had been relaxed for the Paris new towns, were of more benefit to the small and medium towns on the outskirts of Paris (Dreux, Évreux, Beauvais, Orléans) than to the much more remote provincial metropolises.

Lastly, the metropolitan growth centers were to serve as test benches for the new urban governance in the areas of regional integration and intercommunality. The need to consider territorial planning on an extended spatial scale, so that the regions surrounding the agglomerations promoted could benefit from the fallout from the public investments allocated, led to a broad definition of the intervention perimeters. From 1966, therefore, OREAMs (Organizations for the study of metropolitan area development) were set up in several areas around metropolitan growth centers (Marseilles, Nord, Nantes Saint Nazaire, Nancy Metz, Lyon Saint Etienne). Organizations focusing on analysis and proposal more than action, under the authority of the regional prefects, they nonetheless foreshadowed an integrated vision of regional development. Similarly, the law of 31 December 1966, which set up 'urban communities' to group together under a single authority communes within a large agglomeration, was applied by governmental decision in four out of the eight metropolitan growth centers (Lyon, Lille, Bordeaux, and Strasbourg). This led, at a very early stage, to the emergence of the management instruments that were to spread 30 years later in a radically different decentralization political context: planning schemes under the control of the regions, communal groups on the initiative of the territorial collectivities.

Shifts of a Regional Development Policy

Despite the massive resources implemented and the undeniable achievements, the policy of the metropolitan growth centers remained above all marked by the scale of disillusions and shifts with respect to the declared intentions. The heavy urbanization trends, upheaval of the centrality's political philosophies and logics, explain these shifts. The first error was probably due to a failure to place the theory of hierarchized urban networks into a historical perspective. Advocates of the fascinating generalization of uniform space geometry, the authors did not appreciate the fact that the spatial models described corresponded more to the maturity of the industrial societies than to the emergence of a civilization based on information and communication. Discontinuity of the territories, increased material and immaterial mobilities, and changing mentalities obliterated the traditional notions of hierarchy and centrality. In France, from the 1950s to the 1990s, the demographic growth of the Paris region no longer stifled the French dynamism and the metropolitan growth centers were far from reaping the reward of the hopes placed in them (**Table 1**). The middle sized towns, assumed to offer a happy medium between living environment and economic efficiency, were triumphant. To a large extent, the DATAR, without giving up the objective of fighting against the capital's hypertrophia, changed its aim and followed the trend: in 1972, state contracts were drawn up with the middle sized towns, then with the small towns, and lastly with the districts, an old notion of rural geography, resurrected for the requirements of technocratic development. To a large extent, the geographic fragmentation of regional policies weakened the legitimacy of the metropolitan growth centers.

Another reason for the changing logics was institutional, related to application of the decentralization laws in 1982–83. Quite paradoxically, transferring the skill of some of the central state authorities to the regions

Table 1 Triumph of the French middle-sized towns over the metropolitan growth centers in the second half of the twentieth century (in million inhabitants)

	Total population	*Demographic gains*			
	(1990)	*1911 54*	*1954 82*	*1982 90*	*1990 99*
Paris agglomeration	9.3	1.7	2.3	0.4	0.3
8 metropolitan growth centers	6	0.9	2.1	0.2	0.5
Other towns >50 000 inhabitants	13.5	1.7	6.3	0.3	0.8
Total of towns >5000 inhabitants	38.5	5.4	13.5	1.3	1.6
Metropolitan France	56.6	3.2	12.6	2.3	1.9

and departments, in particular in terms of economic developments and territorial planning, affected a policy mainly aimed at decentralizing the bases of economic growth. The context was quite different, however. The metropolitan growth centers of 1963 corresponded to a deliberate attempt by the Jacobin system to correct its own exaggerations. From the 1980s, the state's financial and political withdrawal allowed a spirit of competition to grow between the regions and the towns. The metropolitan growth centers were inspired by a unique model. Metropolization now existed in various forms, with unequal resources and varying degrees of success.

These transformations were all marked with the seal of a radical upheaval of urban centrality, due to globalization of the economy and a weakening of the national states, especially through the construction of Europe. Towns were considered less in terms of their immediate and internal relations with the nearby spaces and their relations of subordination or command with the other agglomerations in a pyramidal hierarchy, than in terms of a system of more or less remote relations with centers of the same size. Paris was therefore not so much the hypertrophic capital of France to be challenged, but more a global city competing with London, New York, or Tokyo, to be supported as much as possible. Similarly, the regional metropolises were much more concerned with developing their own future individually rather than with playing a role in a national system, which must be entirely rebalanced. The current French metropolization process no longer consists exclusively of the higher levels of the demographic hierarchy. It is based on the activities, professions, and social layers that require the large town to develop and prosper: art, banking and insurance, company management, research, and telecommunications. At the same time, the principle of innovation centrality is maintained (very large towns) and a modernity distribution cycle is speeded up.

This set of complex processes finally explains why the current fate of the eight French metropolitan growth centers of 1963 seems less related to the univocal results of a planning policy than to the heritages and projects managed by each one. Three of them appear to have a privileged situation, although for different reasons. Lyon plays on its historical destiny as an international metropolis at the traditional crossroads of the Mediterranean, Northern, and Alpine influences. Lille benefits from its architectural heritage, the futuristic Euralille, and its new position as 'European hub' between Paris, London, and the Randstadt. And Toulouse sees its global destiny tied to that of Airbus, the major competitor of the American Boeing. Two other more regional successes may appear in a favorable position, due to their peripheral situation, which was for many years a handicap in an era of construction of national economies before becoming an asset in a world open to the outside. Strasbourg is now promoting its status as European capital in a climate of political and economic construction drawn toward the East by institutional expansion. A sea breeze is once again blowing over Nantes from the Atlantic coast, at a time when Saint Nazaire has revived its industrial tradition in the commissioning of large cruise ships. Finally, the last three are struggling to recover from crises, after seeing them through more or less successfully. Bordeaux seems to be coping the best, following a long period of lethargy after losing the colonial traffic and an eclipse of its local political aedileship, a fatal blow in the new urban landscape. Nancy associated with Metz is struggling to overcome the environmental and human handicaps of historical continental industrialization (iron and steel). And notwithstanding the state's repeated efforts (Euromed for high tech reconversion of the docks), Marseilles is finding it difficult not to be overshadowed by its regional rivals, despite the fact that they are less privileged in terms of institutional policies: to the West, the 'gifted' Montpellier is top of the class in technopolitan economy (information technology, health), while to the east, Nice combines its touristic charms with the virtues of Sophia Antipolis, the French Silicon Valley. Proof, if need be, that the policy of the metropolitan growth centers has not been easy.

Overall, the metropolitan growth centers were subjected to a tremendous life size trial of a theoretical model designed to reconstruct the space organization and attractiveness of towns and a policy of interventionism applied from above. For a while, at the height of the industrial systems, they appeared to satisfy a universal

Table 2 Urban performance increases with town size: greater metropolitan employment in French towns (% of the active population)

Total number of jobs in the urban area (in millions)	*Number of urban areas*	*1990*	*1999*	*Trends 1990 99*
5.1	1 (Paris)	14.3	16.0	+11.7
0.2 0.7	11	8.4	9.5	+20.4
0.1 0.2	21	5.7	6.4	+20.0
All urban areas	354	8.2	9.0	+14.8

and steadfast geometry of the urban networks and the efficiency of state voluntarism. They already neglected both the historical and social processes responsible for the production of regional territories and urban materialities, as well as the innovation power of the local environments. The sudden acceleration of technological, economic, cultural, and societal changes shook these theoretical approaches and disoriented their applications. This voluntarist policy of regional planning will have left behind it the testimonial of ambition, always necessary, irrespective of the direction chosen, the risk assumed, the material equipment structuring the territories and, probably even more, a spirit of revival, which played a role in the current success of the metropolises.

Conclusion

Despite this relative theoretical failure of the policy of metropolitan growth centers, the French example has constantly fascinated numerous countries, which experienced the same symptoms of territorial imbalance in relation to a hypertrophic political and economic capital (**Table 2**). The DATAR and its experts did a great deal to export and adapt the measures taken, while the French state was considered a model of efficiency and successful compromise between respect for free market economy and public interventionism. Adopted often long after the initial idea and often with fewer means, these principles brought the same, often disappointing, results everywhere. This demonstrates that in modern times, the popularity of a policy depends less on the accuracy of its theoretical bases than on the communication that accompanies it.

See also: Networks, Urban.

Further Reading

Burgel, G. (1999). *Paris, Avenir de la France*. Paris: l'Aube.

Burgel, G. (2003). La ville contemporaine de la seconde guerre mondiale à nos jours. In Pinol, J. L. (ed.) *Histoire de l'Europe Urbaine* (livre 6). Paris: Seuil.

Burgel, G. (2006). *La Revanche des Villes*. Paris: Hachette.

Burgel, G. (2008). *Paris meurt il?* Paris: Perrin.

Chabot, G. (1947). *Les Villes*. Paris: A. Colin.

Christaller (1933). *Die Zentralen Orte in Süddeutschland*.

Dugrand, R. (1963). *Villes et Campagnes en Bas Languedoc. Le Réseau Urbain du Bas Languedoc Méditerranéen*. Paris: University Presses of France.

Gravier, J. F. (1947). *Paris et le Désert Français*. Paris: Flammarion.

Hautreux, J., Rochefort, M. and Lecourt, R. (1963). *Le Niveau Supérieur de l'Armature Urbaine Française*.

Losch (1931). *Die Räumliche Ordnung der Wirtschaft*.

Merlin, P. and Choay, F. (eds.) (1988). *Dictionnaire de l'Urbanisme et de l'aménagement*. Paris: University Presses of France.

Reilly, W. J. (1931). *The Law of Retail Gravitation*. New York: Knickerbocker Press.

Rochefort, M. (1960). *L'organisation Urbaine de l'Alsace*.

Rochefort, M. (1993). *Métropoles en Déséquilibre?* Paris: Economica.

Rochefort, M. (2002). *Special Issue: Des Métropoles d'Equilibre aux Métropoles d'Aujourd'hui, Strates. Parcours dans la Recherche Urbaine. un Géographe Engagé*.

Roncayolo, M. (ed.) (2001). La ville aujourd'hui, mutations urbaines, décentralisation et crise du citadin. In *Histoire de la France Urbaine*. Paris: Seuil.

Sassen, S. (1991). *The Global City*. Princeton, NJ: Princeton University Press.

Middle East and North Africa

M. E. Bonine, University of Arizona, Tucson, AZ, USA

Glossary

Desertification The process of the degradation of the land due to desiccation, caused by both natural climatic factors and anthropogenic factors.

Identicide The policy to erase the cultural identity of a group or people.

Islamic City The view that Islam has been influential in the creation, morphology, and character of cities during historical times and even today, a model first proposed by Orientalists.

Medina The traditional or 'old city' in the Middle East and North Africa, specifically referring to the cities of the Arab world.

Narrative of Loss The typical architecture/urban narrative of the Middle Eastern city, that due to the loss of a past great civilization and plundering by colonization that the cities are in a perpetual state of underdevelopment.

Orientalism The concept as developed by Edward Said that postulates that the image of the Orient and its peoples were invented and formulated by writers, scholars, and artists of the West and which created a racist, colonist characterization of the Other.

Petroleum Urbanization The extremely rapid urbanization of the Arab states of the Gulf Cooperation Council (GCC) for which massive numbers of foreign workers are essential.

Purdah The seclusion, separation, and sheltering of women from nonrelative men.

Qanat A water supply and irrigation system in which a near-horizontal underground channel taps an aquifer at a higher elevation, and the water flows by gravity to the village and fields; especially still prominent in Iran.

***Qanat* Irrigation Cultures** When a rural, village society and many of the economic and social relationships are influenced by the cycle and flow of the *qanat* and the irrigation practices.

Urbicide The specific policy and strategy for the destruction of urban fabric and cities in order to control, displace, or eliminate the population.

Introduction

The Middle East and North Africa, as one of the world cultural regions or realms, has been defined in various ways that have resulted in rather different combinations of nation states. This region sometimes is designated as only Southwest Asia, which would be those Arab states plus Turkey and Iran – and sometimes Afghanistan. Egypt might be part of such a region, which then might also be called the Middle East or the Near East, the latter term still preferred by archaeologists focused on studying the ancient cultures of this area. Rest of the states of North Africa, from Morocco to Libya, are also principally Muslim, Arab, and with similar arid environments, and hence the region becomes the Middle East and North Africa. It encompasses most of what is called the Arab world, except that Turkey and Iran are not Arab, and there are other states on the margins, such as Mauritania or Somalia, that are Muslim and arid but which usually are not considered part of this world cultural region. The complexities of just trying to define the Middle East have led some to say that it is somewhat an indefinable region.

Desertification and Degradation Narratives

The Middle East and North Africa is comprised principally of arid and semiarid environments. Human occupation, including numerous empires and 'civilizations', has for millennia impacted these fragile landscapes, and the typical view until recently has been that humans have overused this environment, and that anthropogenic (human) factors have caused the desertification of many parts of the region. Nomads are particularly accused of overgrazing the land with their herds of goats, sheep, and camels, which leads to the loss of the scant vegetation, causing erosion and denuding of the landscape. These degradation narratives were particularly prevalent in the 1960s and 1970s (and later) for explaining the extreme famines that impacted the peoples of the Sahel in West Africa. The United Nations Conference on Desertification (UNCOD) held in Nairobi in 1977 represented this view, which led to national policies and programs to 'combat' desertification, as well as the 1994 United Nations Convention to Combat Desertification (UNCCD). Although numerous countries of the Middle East and North Africa (and elsewhere) have embraced this narrative, the debate about the 'encroaching desert' remains quite controversial. By the later 1990s, especially due to the 'greening of the Sahel' caused by the return of rains to much of West Africa, the debate about desertification has begun to incorporate the evolving understanding of climate variability and that natural drought and climate changes may have as much as if not more impact than anthropogenic influences.

More recently, Diana Davis and others have begun to examine some of the historical roots of the degradation narrative, particularly how it developed during the French colonial period in North Africa in the nineteenth and early twentieth centuries. After the French invaded Algeria in 1830, they soon began to compare the rather dry and bleak environment around them with the accepted myth derived from the ancient Greek and Roman writers: that North Africa was once one of the most fertile regions of the world, including one of the principal granaries of Rome. Why would the environment have changed? What caused the desertification and loss of forests that were believed to have occurred since the Roman period? Arab nomads were the answer. Also using the writings of Ibn Khaldoun and other Arab historians, the French soon asserted that the fertile land of North Africa, including the existence of extensive forests, became desertified by overgrazing and the cutting down and burning of forests by hordes of Arab nomads.

The degradation narrative was then used by the French to implement their colonial policies in the name of environmental protection. Forests were taken over by the colonial French government and the indigenous use of forest criminalized. Large areas of agricultural and pastoral lands were appropriated. The control of pastoral nomads became one of the priorities of French policies and enforcement, all in the name of not only preserving the environment but also to bring back what was considered North Africa's fertile land of the past. The colonial policies also gave much of the appropriated land to French farmers – who sometimes did attempt farming in some rather marginal environments where previously only nomads had grazed their herds, and whose farming practices then in actuality often did induce desertification! The colonial environmental narrative not only lasted throughout the French colonial period but also formed the basis of the environmental policies of the independent states of North Africa after World War II, as well as continuing to be influential for present day programs, policies, aid, and practices.

The Scarce Resource: Water Wars and Conflicts

The Middle East and North Africa has a considerable deficit of moisture throughout most of the year in the region, and the scarcity of water has led to many conflicts and even wars over this precious but absolutely necessary resource. Populations continue to increase in every nation of the region, requiring each year even more water for domestic use, industry, and irrigated agriculture. Conflict over water can be at the village level where the individual allocations of irrigation water might be in dispute, or the quarrel might be between states over the division of the waters of major rivers such as the Nile or the Tigris and Euphrates.

Although some areas of the Middle East, such as central Turkey, have considerable agriculture from dry land (rainfed) farming, much of the region must be irrigated to produce crops. Traditional irrigation of local agricultural fields in villages usually revolves around a cycle of water in which the users have a specific number of time shares of water for irrigating their crops. It is not a volume of water but a specific length of time of flow, which might vary in the total amount of water available depending on the season or the climatic conditions influencing the specific aquifer or source of water. During drought seasons and years, there is less flow and hence less land can be irrigated. Water is used by flood irrigation, sometimes called border or strip irrigation, where water is conveyed in open channels and crops are irrigated by inundation of fields having low ridges to control the water.

Within Iran the *qanat* was (and is) used throughout much of the central plateau. This irrigation system obtained water by a near horizontal underground channels which were dug into aquifers at a higher elevation than where the water exited at the village and the cultivated fields. Water flows by gravity and usually the flow is determined by local climatic conditions. Historically, conflicts might occur where a new *qanat* channel was dug beneath an existing underground *qanat*, which would be able to drain some of the water from that *qanat*. Since about the 1960s, much of the conflict has been between existing *qanat* systems and new tube wells with modern pumps. The more efficient pumps would often cause the demise of *qanats* downstream by lowering the water table below the intake aquifer of the *qanat*. In some valleys where dozens of villages once depended on *qanats*, all of these systems went dry. Also, whereas the *qanat* water was shared water by villagers based upon rights to irrigate, the water from wells was owned by individuals and sold by volume. Unfortunately, many wells also went dry after a few years, creating a drastic decline in the availability of water – and in agriculture production. The village social structure traditionally had been based upon the flow and cycle of water from the *qanat*, and so these irrigation cultures and many of the social relations were also impacted by the demise of the *qanat* systems.

All of the major international rivers of the Middle East and North Africa are highly contested. The Nile for thousands of years has been used principally by Egypt, and it is only in the nineteenth and mainly twentieth centuries that conflict has emerged over its waters. Egypt has claimed the majority of the water of the Nile based upon its traditional use of the Nile, this usufruct creating the rights to the water. Yet, as irrigation and agriculture schemes have developed in Sudan, this nation wants to use more of the water, from both the White Nile and the Blue Nile, for its own purposes. Then, as Ethiopia continues to develop and its population increases, it also will want to use the waters of the Blue Nile, which originates from Lake Tana in Ethiopia and is, in fact, the principal source of the main flood of the Nile in Egypt that occurs at the end of every summer. Development schemes and the desire to control more of the water led Egypt to build the Aswan Dam at the beginning of the twentieth century, and finally the Aswan High Dam completed in 1971. The sharing of waters of the Nile between Egypt and Sudan has been quite contentious, and this is without Ethiopia yet developing any major use of the Blue Nile.

Similarly, the waters of the Tigris and especially the Euphrates rivers are highly contested. Both rivers originate in Turkey, and as Turkey develops major dams and irrigation projects (the GAP Project), there is less water for downstream Syria and Iraq. The Euphrates is particularly affected by the dams in Turkey and Syria, leaving Iraq as the last recipient of the water, despite its longer historic right to use the river.

Although not a major river in terms of flow, the Jordan River and the waters underneath the Palestinian Authority (West Bank) are perhaps the most contested waters of the Middle East. The source of the upper Jordan River are three major springs, the Banaias in Syria, the Hasbani in Lebanon, and the Dan in Israel. A major tributary of the Jordan, the Yarmuk River, originates in Jordan and Syria, as well as forming part of the border between Israeli occupied Golan Heights and the countries of Syria and Jordan. It is the groundwater of the West Bank that is even more contentious, for much of it flows from there westward into Israel, which depends upon this water as part of its scarce water supply (up to 40% of the groundwater that Israel uses by some estimates). The limitations on the drilling of any new wells by Palestinians after the Israelis occupied the West Bank in 1967 caused considerable hardship for the Palestinians, which contrast considerably with the extensive wells (and swimming pools) allowed for the Jewish settlements in the West Bank. Even after the establishment of the Palestinian Authority, Israel still has the final control of the allocation of water in the West Bank.

Edward Said, Orientalism and the Islamic City

The work of Edward Said (1935–2003) has had a profound impact not only in human geography but also for most other disciplines of the humanities and social sciences. Said provided a powerful intellectual framework showing how power and politics shaped the representation and interpretation of culture. His best known work, *Orientalism* in 1978, along with such other works as *Covering Islam* in 1981 and *Culture and Imperialism* in 1993, depicts how the West represented – and really discovered and invented – the East (the Orient). Orientalism was created by Western archaeologists, artists, and writers in the eighteenth and nineteenth centuries, and it became part of the discourse of domination of the East by the West. The Middle East constituted the principal region for Said's examination and critique of the West's Orient, although India and other Asian 'Oriental' cultures are included as well. His concentration on culture and power influenced contemporary human geography by creating a greater focus on race, power, and representation, including imperial landscapes. This legacy also has had a major impact in furthering the colonial discourse and contributing to the development of postcolonial theory – even if much of that theory has been to counter and object to Said's Orientalism arguments.

Edward Said also had another, but related, passion. As a Palestinian his writings and energies were also directed toward the plight of the Palestinian Arabs, focusing on the injustices that have been inflicted on these people throughout the twentieth century and continuing today. He linked his critique of Orientalism with how the Palestinians were represented in the West. The Palestinian Arabs, as one of the uncivilized and degenerate 'Orientals', were seen as backward, while the Zionists were depicted as capable and civilized, besides the acceptance in the West of the prophesied claims to the biblical 'promised land'. Palestinians were associated with terrorism, ignorance, and despotism, while the Zionists were associated with democracy and liberal ideas. Hence, related to his Orientalism discourse, Said emphasized how the nature of power and representation meant that the Palestinians were being represented by others – and as the Other. His works, such as *The Question of Palestine* in 1980, in fact not only brought worldwide attention to the plight of the Palestinians but also provided a framework for countering identicide, the concerted attempt to erase historical Palestine and the Palestinian cultural identity – and, hence, giving the Palestinians a voice to the international community.

Orientalists, who also can be defined as those earlier scholars with knowledge of the languages and cultures of the Middle East and North Africa, were responsible for creating the concept of the Islamic city. This model was based upon the belief that Islam as a religion and way of life had a specific influence of the urban built environment as well as the inhabitants themselves. The French Orientalists, Georges Marçais, William Marçais, and Jean Sauvaget, for instance, developed these ideas in the 1920s and 1930s, which culminated in Gustave von Grunebaum's once influential 1955 article, 'The structure of the Muslim town'. More recently, the use of primary documents, such as *waqf* (charitable endowment) deeds or *sijills* (*sharia* court records), has led to much more sophisticated analyses of traditional cities in the Middle East. *Sharia* law has, in fact, been significant for the morphology of Muslim cities, as Basim Hakim has shown for Tunis, where the heights of buildings, widths of streets, and ownership of shared walls were regulated by *sharia* and decisions by a *qadi*, the Islamic judge. In some instances, the *qibla* or the direction of prayer to Mecca influenced not only the layout of mosques but also the orientation of the traditional Medina itself.

The Islamic city or the Middle Eastern city also has been interpreted by the narrative of loss, that these cities have declined from a previous glorious past. Such a narrative and the models of the Islamic city or the Arab city or the Iranian city in the contemporary world eschews an examination of these cities as global cities which are impacted significantly by forces of globalization and internationalization.

Rapid Urbanization, Oil, and the Impact of Globalization

Similar to rest of the developing world, rapid urbanization has been significant in the Middle East. Major cities, particularly the capitals and primate cities, have had tremendous growth as fertility rates remain some of the highest in the world (but declining) and rural and foreign migrants flock to the urban areas mainly for work. Several megacities have emerged in the region, with Cairo's population from 15 to 18 million, Tehran at least 12 million, and Istanbul over 10 million. Shantytowns and illegal settlements have accompanied much of the rapid urbanization, while refugee camps are part of the urban landscape of cities such as Amman, Beirut, and Khartoum.

The Arab states of the Arabian Peninsula and the Gulf Corporation Council (GCC) have undergone some of the most rapid urbanization anywhere in the world because of their oil wealth and small indigenous populations. Particularly as the price of oil began to climb in the early 1970s, this petroleum urbanization created a transformation of these cities in Saudi Arabia, Kuwait, Bahrain, Qatar, United Arab Emirates, and Oman. Development was extremely rapid and the native populations were inadequate and not sufficiently skilled, so that millions of foreign laborers have come to the GCC states to work. The forces of globalization have created a pool of foreign labor, which in fact constitutes the majority of the labor force in all six of these Arab states (and where foreigners are the majority of the total population in Kuwait, Qatar, and the United Arab Emirates).

In the first decades after World War II as these oil economies began to develop, most of the foreign migrant workers were Arabs – such as from Egypt, Yemen, Lebanon, and Jordan, including many Palestinian refugees resulting from the Arab–Israeli conflict. In more recent decades most foreign workers are 'Asians', coming from India, Sri Lanka, Nepal, Indonesia, the Philippines, and other countries. Many Arab and Asian foreigners have been working in one of the GCC states for decades, but it is virtually impossible for them to obtain citizenship in these 'oil welfare' states. The Gulf also has turned into cities of gleaming skyscrapers. Abu Dhabi and Dubai are representatives of the forces of global capital – and oil wealth – massive five star hotels, indoor skiing facilities, and the construction of the world's tallest building attest to (what some call) excesses of wealth and development.

The Arab–Israeli Conflict: Urbicide and Identicide and the Palestinians

Ever since the establishment of Israel in 1948 the Israelis and Palestinians (and other Arabs) have been in conflict, and this has remained one of the most difficult and intractable disputes in modern times. Water rights, refugee status, and 'right of return' of the displaced Palestinians, the status of Jerusalem, Jewish settlements in the West Bank, and many other issues divide these two peoples. Although hundreds of scholarly – and not so scholarly – books have been written by individuals from all disciplines, geographers also have contributed their perspectives. In particular, articles by Ghazi Falah and others in the journal he founded, *The Arab World Geographer*, have provided valuable information and insights into the conflict.

An example is the focus on urbicide, the specific policy and strategy for the destruction of urban fabric and cities in order to control, displace, or eliminate the population. Originally identified as a term for the planned and implemented destruction of (parts of) specific cities in Bosnia, such as Mostar, more recently, urbicide has been seen as a relevant concept related to the Palestinian cities of the West Bank and Gaza. Specifically, the policies of the Israeli government toward the Palestinian Arab population in the Palestinian Authority have been an attempt to demonize the Palestinian built environment as premodern, formless, and dangerous, hostile places, where terrorists and suicide bombers abound. Hence, it becomes acceptable (as Israeli policy) to destroy such built structures. Besides destruction, planning and implementation of new (Israeli) building structures and neighborhoods and new settlements can also be an integral part of urbicide.

The conflict between Israel and the Palestinians, in fact, has long provided ample evidence of violence and the destruction of built environments. Originally, with the establishment of the State of Israel and the resultant war in 1948–49, the exodus of much of the Palestinian population from the new boundaries of the state resulted in hundreds of villages being either razed or reoccupied with new Jewish inhabitants. Many Palestinian settlements became ruins and even those that remained occupied by Arabs sometimes disappeared from Israeli maps. The demolishing of individual houses of Palestinians also became a policy of the Israeli military after the 1987 and 2000 'Intifadas', where residences which housed the family of a suicide bomber (or other individuals deemed a threat) were often bulldozed as 'punishment'. Since the 1967 war, at least 12 000 Palestinian houses in the 'occupied territories' (including East Jerusalem) have been destroyed in either individual or massive housing demolitions.

More recently, urbicide has become Israeli policy in the Palestinian Territories, particularly since the 2000 al Aqsa Intifada. In 2002, the Israeli Defence Forces occupied much of the Palestinian Territories, allegedly to counter Palestinian suicide bombers and Islamic terrorism (including leading members of Hamas). Besides destruction of urban fabric by missiles, tanks, and bulldozers, urbicide is also brought about by other measures which strangle and isolate the population of this perceived Other. A matrix of control has been created by the Israeli policies and military, consisting of Israeli settlements, bypass roads, and tunnels, and more recently by the construction of the Wall of Separation which sometimes even separates Palestinians from their own agricultural land. Major highways in Israel often will not have offramps to Arab settlements and usually there are no signposts signifying that Palestinian settlements even exist. Such highways may also be built principally on confiscated Palestinian agricultural land and these thoroughfares for the Israelis often can be barriers for the Palestinians to move between their residences and their cultivated lands.

Many of these policies become part of the strategy to marginalize the Palestinians – a policy of identicide. A concerted effort has been made to separate the Palestinians from their cultural and material background – from their identity. The destruction in 2002 of much of Nablus is an example of the policies and implantation which have created urbicide and attempted identicide in this largest Palestinian city in the West Bank.

Gendered Space: Private and Public Space

Societies in the Middle East and North Africa tend to be the most segregated societies by gender in the world, although customs certainly differ by subregion, religion, class, lifestyles, and education. Predominately Muslim groups and peoples occupy this cultural region, and so the influence of Islam becomes central to understanding many of the social practices and patterns. Pastoral nomads have been the focus of many studies in the past, where it was emphasized how nomadic women tend to be 'more liberated' than their counterparts in the cities – including not being veiled as well as often working alongside men (which is true for many rural women as well). These past studies also sometimes simplified the roles of women and men, with the males often considered as the principal herders and managers of all productive resources, while the women were seen as primarily the purveyors of domestic space. However, perspectives by human geography and feminist geography have contributed newer understandings of the role of gender and space among nomads, for instance. Nomad women, in fact, have considerable ethnoveterinary knowledge, often greater than the men. They play a major role in the daily care of livestock, and because of their extensive contact with their animals they have considerable knowledge of livestock diseases and indigenous range ecology. Yet, women are often ignored in government (and outside aid) funded programs and projects, which presuppose that the livestock production is entirely in the hands of the men, and hence, any information and funds for improvements are given only to the males of a specific nomadic group. Many nomads also have undergone sedentarization where they engage in more intensive agropastoral production, and these changes transform many of the household production strategies as well as the relationships between male and female.

Discussions of the veil and *purdah*, as well as women's rights, have in particular been a focus of much scholarship in the Middle East and North Africa. The *hijab* or veil worn by women designates protection and a boundary with the outside world, signifying the private world that encloses – and protects – a woman. Yet, it has been one of the principal symbols of oppression of Muslim women as interpreted by many writers and feminist scholars in the West. Women's rights in the Middle East, in fact, have been a constant struggle throughout the twentieth century and until today. Providing education for women and reforming divorce and polygamy laws show progress, but only partially and with differing results from country to country. One of the most symbolic reforms, the abolition of the veil, was carried out in such secular states as Ataturk's Turkey or Pahlavi Iran, but the reemergence of Islam and Islamists the last several decades has resulted in reveiling and the donning of Islamic dress. Many more women in Cairo are wearing the veil today compared to two decades ago, while since the end of the 1970s in the Islamic Republic of Iran, women must wear a *chador* over their head and body when venturing out into public space. Whereas the adoption of the veil and Islamic dress was once mainly voluntary and a personal decision, the rise of Islamic fundamentalists and their conception of an Islamic society and state have limited the choices available to many women today.

Whereas the duality of public and private space was once used to describe men's space versus women's space in the Middle East and North Africa, and Muslim societies in general, within the last several decades the perspectives coming from feminist scholars and particularly ideas from feminist political ecology and postmodern deconstruction have shown a much more complicated reality. Both men and women negotiate private space and women can control at times public space. At the household level both sexes have particular vested interests in the resources that they manage. The fact that many men of rural communities, including settled nomads, work for months or even years outside of the home, results in significant alterations in the role of the women in terms of the control of resources and production at the domestic level. As nomadic groups settle and develop new livelihood strategies, they retain some herding but engage in farming and labor migration. The nomadic identity begins to transform as the domestic and public spaces change.

The New Media and the Public Sphere

Within just the last several decades the advancements in media and communication, similar to rest of the world, not only have profoundly affected the lives of individuals but also have transformed the public sphere. Terrorist organizations such as al Qa'ida now make and distribute sophisticated training camp tapes, as well as producing tapes for the world media. That media also includes new Arabic language satellite television stations, such as Al Jazeera, being broadcast from Qatar in the Persian Gulf. Such media provides different voices and perspectives than those which come from the BBC, CNN, or CBS, and Al Jazeera in particular has become not only a voice in and for the Arab world, but also it has emerged as a major global provider of international news and critical perspectives. Arab societies across all of North Africa and the Middle East now watch these new stations, engendering substantive discussions and debates. On the lighter side, Egyptian soap operas have become extremely popular, as episode after episode mesmerizes much of the Arab world.

The new media might provide a threat to the state in some instances, for governments can no longer control what news and programs are being watched by their citizens, even though satellite dishes have at times been outlawed, as in Saudi Arabia and Iran. In other instances, the emergence of the new media has enabled Islamic customs, doctrines, and practices to be discussed and questioned. What is considered proper Islam and how to live an Islamic life are promulgated on the television, while counter arguments are made by other religious authorities. Audiocassette tapes and videos of antiregime clergy circulate in some countries. Islamic educational videos and CD ROMs are available throughout the region. What is most significant, however, is that the masses in the Middle East and North Africa now have ready access to news and information; it is no longer the purview of just the government or the educated elites.

The Internet also has transformed the public sphere. Internet cafes have become crowded with youths and the cafes have become new, important public spaces. A multitude of websites are now available to this new generation. Islam is also part of the Internet and is another way that various viewpoints of doctrine or interpretation are spread. Diversity of opinions and discussions of numerous topics has created new social networks and contacts. Various groups, including Islamist and even terrorist organizations also utilize the Internet, many having their own websites. In about 2005, Arab blogging began in earnest and many new websites advocated greater political transparency for specific states and the Arab world in general. Governments and censors have countered and the battle for controlling cyberspace – and for reforming authoritarian governments – has begun. In some states, such as Saudi Arabia, Syria, Yemen, and Libya, the filtering and blocking of websites for political or moral reasons is done by the government, often by using only one provider controlled by the state.

Conclusion

Stereotypes of the Middle East revolve around Islamic terrorists, rich oil sheikhs (with large harems), stoic nomads riding camels across burning sand dunes, and oppressed veiled and sheltered women shunning public spaces. The realities, however, are much more complex and subtle. These societies not only have embraced modern media and technologies, but also in many instances new lifestyles are being combined with a new commitment to the religion of Islam. The peoples of the Middle East and North Africa are in large part still adapting to a postcolonial world, and yet are still often being defined by the West in an Orientalist mode. It is true that many inequalities exist within these societies: between the state and the ruled, between majorities and ethnic and religious minorities, between man and woman. The intrusion of the West also has created contested political space, epitomized at the state level by the enduring Arab–Israeli conflict where the territorial and political struggles have polarized much of the region – and even the world.

See also: Desertification; Gender, Historical Geographies of; Islamic Urbanism; Poststructuralism/Poststructuralist Geographies.

Further Reading

Abujidi, N. and Verschure, H. (2006). Military occupation as urbicide by 'construction and destruction': The case of Nablus, Palestine. *The Arab World Geographer/Le Géographe du Monde Arabe* 9(2), 126 154.

Bonine, M. E. (1990). The sacred direction and city structure: A preliminary analysis of the Islamic cities of Morocco. *Muqarnas: An Annual on Islamic Art and Architecture* 7, 50 72.

Bonine, M. E. (1996). Qanats and rural societies: Sustainable agriculture and irrigation cultures in contemporary Iran. In Mabry, J. B. (ed.) *Canals and Communities: Small Scale Irrigation Systems*, pp 183 209. Tucson, AZ: University of Arizona Press.

Bonine, M. E. (2005). Islamic urbanism, urbanites, and the Middle Eastern city. In Choueiri, Y. M. (ed.) *A Companion to the History of the Middle East*, pp 393 406. Oxford: Blackwell Publishing.

Davis, D. K. (2007). *Resurrecting the Granary of Rome: Environmental History and French Colonial Expansion in North Africa*. Athens, OH: Ohio University Press.

Eickelman, D. F. and Anderson, J. W. (eds.) (2003). *New Media in the Muslim World: The Emerging Public Sphere* (2nd edn.). Bloomington, IN: Indiana University Press.

Elsheshtawy, Y. (ed.) (2004). *Planning Middle Eastern Cities: An Urban Kaleidoscope in a Globalizing World*. London: Routledge.

Falah, G. W. (editor in chief) (2004). *Special Issue on the Work and Legacy of Edward Said: The Arab World Geographer/Le Géographe du Monde Arabe* 7(1 2).

Falah, G. W. and Nagel, C. (eds.) (2005). *Geographies of Muslim Women: Gender, Religion, and Space*. New York: The Guilford Press.

Hakim, B. S. (1986). *Arabic Islamic Cities: Building and Planning Principles*. London: Routledge and Kegan Paul.

Kapiszewski, A. (2001). *Nationals and Expatriates: Population and Labour Dilemmas of the Gulf Cooperation Council States*. Reading: Ithaca Press.

MacLeod, A. E. (1997). The new veiling and urban crisis: Symbolic politics in Cairo. In Bonine, M. E. (ed.) *Population, Poverty, and Politics in Middle East Cities*, pp 304 325. Gaianesvill, FL: University Press of Florida.

Meharg, S. (2001). Identicide and cultural cannibalism: Warfare's appetite for symbolic peace. *Peace Research Journal* 33, 89 98.

Sakr, N. (ed.) (2007). *Arab Media and Political Renewal: Community, Legitimacy and Public Life*. London: I.B. Tauris.

Relevant Websites

http://ddc.aub.edu.lb
Digital Document Collection at the American University of Beirut.

http://www.irinnews.org
IRIN: Humanitarian news and analysis from the UN Office for the Coordination of Humanitarian Affairs.

http://www.lib.utexas.edu
Maps from the University of Texas at Austin, Perry Castaneda Library.

http://www.middle east online.com
Middle East Online the link page is relevant newspapers, Middle East and North Africa search engines, etc.

Migrant Workers

C. Cindy Fan, University of California, Los Angeles, CA, USA

Glossary

Circular Migration Moves by a migrant between the sending and receiving areas many times over a period of time.

Dependency Ratio The ratio of the young and the old to the working-age population.

Ethnic Enclave A place within a city or a larger community within which members of a particular ethnic group tend to concentrate.

Guest Worker Program A program to recruit workers from another country to ease labor shortage in the host area, with the expectation that the migrant workers will eventually return home.

Human Smuggling The procurement of illegal entry of a person into a country, usually with the objective of making a profit.

Human Trafficking The recruitment, transportation, or receipt of persons through deception or coercion, usually for the purpose of exploitation or forced labor.

Remittances The portion of migrant incomes that is sent home.

Replacement Migration International migration that is needed to offset population aging, population decline, and shrinkage in the working-age population.

Sojourner A migrant seeking temporary paid employment in another country who intends to return home eventually.

Xenophobia Fear of strangers and persons from a different country, ethnicity, or culture.

Introduction

Migration refers to the movement of people across a specified boundary for the purpose of establishing a new or semipermanent residence. Migrants are commonly understood as people moving over relatively long distances, as opposed to movers who move over short distances such as changing residences within the same city. Migrant workers are migrants who move for work and are also referred to as labor migrants.

As early as the late nineteenth century, E.G. Ravenstein had observed that most migrants left their homes in order to search for work of a more remunerative or attractive kind. Study after study since then have confirmed that the search for work is the most important motive of migration. Although people do migrate for other reasons – family reunification, displacement by war, political persecution, etc. – labor migrants consistently constitute the preponderance of international and internal migrants all over the world.

The Economic Geography of Labor Migration

Push and Pull

From a theoretical perspective, the flow of labor is a result of uneven regional development. Neoclassical economic theory, which assumes free flows of factors of production, predicts that workers will migrate from low wage to high wage areas. Over the past several decades, the gap in per capita income between high income countries and low income countries has continued to widen. Not surprisingly, the level of international migration has also increased, so that today, more people are residing outside their country of birth than ever before. The International Organization for Migration (IOM) estimated that by 2005 there were at least 190 million international migrants. The more developed regions in the world – in particular, North America and Europe – consistently have the largest net gains in population from immigration and the less developed regions – Africa, Asia, and Latin America – have the largest losses from emigration.

Flows from less developed areas to more developed areas also occur between neighboring countries and within countries. Argentina, for example, has long been a magnet of migrant workers from poorer Latin American countries, including Chile, Bolivia, Paraguay, Brazil, and Uruguay, although Argentines are themselves migrating to Spain and Italy. In the US, since the 1960s, economic growth in the Pacific Coast states and the South Atlantic coastal states and job loss in the old manufacturing states in the Northeast and Midwest have sustained a westward and southward movement of the population, often referred to as the Rustbelt (or Snowbelt)–Sunbelt shift. The longstanding eastward migration in Russia, and the inland coastal migration in China, both represent movements from poorer areas toward destinations that offer better economic and job opportunities. The push from the origin and pull by the destination are, therefore, primarily defined in terms of where the jobs are.

What neoclassical theory fails to predict is the impact of labor flows on uneven development. According to the theory, migrant flows are expected to change the labor supply and demand in the sending and receiving areas such that wages, as well as levels of economic develop ment, will eventually equalize over space. Despite the trend toward integration of markets on a global scale and increased international migration, the gap between rich and poor countries continues to widen and will in turn reinforce the push and pull of international migration.

Labor Shortage

Migrant workers move predominately from areas lacking jobs or desirable jobs to where labor shortage exists and where the attractive jobs are. Both economic and demo graphic reasons account for labor shortage. Places that experience rapid economic growth and job expansion often have labor demand beyond the capacity of the local workforce. Recruitment of labor may be from nearby or distant areas. Since the 1960s, Mexicans have migrated northward to work in the maquiladoras – foreign owned assembly plants near the US border. During the 1950s and 1960s, Germany and other Western European countries that experienced economic boom implemented guest worker programs to recruit unskilled assembly line workers from Yugoslavia, Turkey, and other Southern and Eastern European countries. Historical ties and geo graphic proximity explain the flows of low skilled labor from Tongo and Cook Islands to work in agriculture and manufacturing in New Zealand. Oil rich countries such as Saudi Arabia have long recruited skilled and unskilled workers from adjacent and distant countries alike. Labor recruitment has also involved human smuggling and trafficking. Between the end of the fifteenth century and the middle of the nineteenth century, millions of Africans were taken against their will to work as slaves in the Americas. Even today, children from Burkina Faso and Mali are sold to plantation owners in Cote d'Ivorie. Many Eastern European and African women have been lured to Western Europe by brokers who subsequently forced them into prostitution.

A mismatch between the labor demand in specific sectors and the local workforce may generate flows of workers with special skills. For example, foreign born workers constitute a significant proportion of the skilled workforce in the US information technology industry. Similarly, the longstanding and high demand for nurses in the US has exerted a migratory pull for foreign nurses, especially from the Philippines but also increasingly from Canada, the United Kingdom, India, Korea, and Nigeria. An aging population in the US will exacerbate this de mand into the next several decades.

Demographically, changes in the age structure may result in labor shortage. Primarily because of aging, the dependency ratio in Japan and many European countries, such as Germany and Italy, is rapidly increasing. To make things worse, these countries have had low levels of fer tility and are experiencing or will experience population decline. Projections of severe labor shortage have set off a heated debate about immigration. While some point out that replacement migration – in the form of massive immigration to the above countries – can be a solution, others argue for increasing labor productivity as a pre ferred means to offset workforce shrinkage.

Household Economics and Strategies

The neoclassical approach focuses on the individual as the unit of analysis, but perspectives that emphasize household economics and strategies show that migration decisions are often made in order to maximize benefits for the entire family or household. Labor migration is an important means to increase household income and re duce risk. China's floating population – people who are not living in their places of registration – is estimated to be between 100 and 150 million and consists primarily of rural migrants working in urban areas. They are pushed from the farmland because of large labor surplus in agriculture, and they are attracted to job opportunities in manufacturing and services in the booming towns and cities. Leaving some family members behind in agri culture and engaging in circular migration, these migrant workers are diversifying the household's income sources. Most important, their remittances – mostly used for living expenses and agricultural input and to finance major household projects such as building or renovating a house – are crucial for the household's well being.

Many, if not most, international migrant workers send home remittances. IOM estimated that total international remittances to developing countries have increased considerably over the past two or three decades and are amounting to about US$100 billion today, a figure much higher than official aid flows to these countries. In some labor exporting countries (such as Mexico, El Salvador, Indonesia, the Philippines, Turkey, Morocco, Jordan, and Yemen), remittances account for a significant proportion of the national economy. A possible cost to households that receive remittances is they may become dependent on this source of income and hence vulnerable to eco nomic changes not only in their own country but also in the country where the migrant works.

The household approach is also important for understanding the role of gender in labor migration and how migration in turn affects gender roles and relations. The conventional model, which assumes that the hus band's career is the determining factor of family migra tion and the wife is the tied mover, is quickly outdated, as women enter the labor force in increasing and large numbers. In some cases, women migrants are the majority in the migration stream because they are of demand for specific types of work. For example, large numbers of women from the Philippines, Indonesia, and Sri Lanka are employed in Hong Kong, Taiwan, and Singapore, where wage increases over the past decades have resulted in shortage of low skilled labor in domestic work and low end services. These migrant women send home large portions of their income and have become the main breadwinners of their families. It is not clear, however, if this has had real impact on women's status in societies where women are traditionally considered inferior to men. In other contexts, women's career has become an important factor in family migration. The gradual decline in mobility (across counties) in the US, for example, is in part explained by increased participation of women in full time jobs and thus families' reduced readiness to make moves in order to advance the husbands' careers.

The Politics of Labor Migration

When labor migration takes place, it may initially be considered a temporary solution, by the migrant workers, as well as the receiving areas that recruit them. However, as migrants' duration of stay increases, as family members join them, and as migrants raise families and have children, they may become permanent *de facto*, if not *de jure*, members of the host society. Issues about migrants' rights, citizenship, and social acceptance are often sources of tension that are especially challenging to policy makers.

Many labor importing countries have had discriminatory immigration policies. Between the 1850s and 1870s, poor rural migrants from Southern China were contracted as coolies to work as miners, railroad builders, farmers, and laborers in the US. The vast majority of these sojourners were men, who intended to return home eventually. By the 1880s, prejudice and economic slowdown culminated into the 1882 Chinese Exclusion Act, which was the first nationality based immigration law in the US and which suspended immigration of Chinese laborers and barred those already in the US from becoming citizens. The 1924 Immigration Act, furthermore, established a national origins quota system, which effectively banned all immigration from Asia, except for a small number of diplomats and students. It was not until the 1965 Immigration and Nationality Act Amendments that national origin, race, and ancestry were replaced by family reunification and occupational skills as bases for immigration to the US. Australia and New Zealand, likewise, in the 1970s and 1980s, replaced admission policies that favored certain national origins, especially the United Kingdom, by policies that employed skill and family reunification criteria. Debates on labor immigration almost always invoke one or more of the following three concerns.

Labor Demand and Supply

Changes in labor demand and supply often result in changes in labor immigration policy. For example, the 1970s Oil Crisis and subsequent economic recession prompted the end of guest worker programs in Western Europe, where instead measures were taken to foster return migration of unskilled labor. Many migrant workers and their family members were left jobless. In response, France and Italy have periodically legalized undocumented immigrants, while Germany and the United Kingdom have launched programs that attracted high skilled immigrant workers. Similarly, in the 1970s, because of the departure of manufacturing jobs, entry of the baby boom cohorts into the workforce, and high levels of unemployment, Australia's immigration policies shifted to emphasizing skills in demand. Indeed, a human capital model of immigration, one that attracts migrants that are well educated and bring valuable skills, is increasingly popular.

In the US and other labor importing countries, a controversial and highly political issue is whether immigrants have adverse effects on the wages and employment of native born workers. The anti immigration view sees immigrants as displacing American workers, especially low skilled minorities who are already disadvantaged to begin with. Those holding an opposing view argue that immigrants are taking jobs that American workers will not or cannot do and that labor market discrimination and segmentation are more important explanations for the unemployment of native workers. Immigrants from and via Mexico to the US, for example, are primarily engaged in agriculture and other low wage, manual jobs that most Americans are reluctant to take. Findings from research, including econometric studies, are not conclusive, although there is some evidence that the labor market effects of unskilled immigrants in the US are small but they may be significant in particular industries, such as garment and meatpacking. A similar debate exists in countries experiencing large scale rural–urban migration. Rural migrants in Chinese cities are blamed for taking away jobs from the urban unemployed, although the former primarily work in the so called 3 D (dirty, dangerous, difficult) jobs that are shunned by urbanites.

Societal Security

Societal security refers to the sense of security derived from cultural homogeneity, a strong identity, and confidence in the continuation of traditional patterns of language, custom, value, lifestyle, and standard of living. Immigrants may be perceived as intruders and outsiders who can threaten the age old process of creation and evolution of nation states. The Chinese Exclusion Act, for example, was based on the notion that individuals from certain national origins and of certain races were unassimilable and should therefore be prohibited from becoming part of the host society. Ethnic tensions in

Europe, likewise, have xenophobic roots. The Paris riots in 2005 are a vivid reminder of the persistent dis crimination toward immigrant workers and their des cendants. Although Hong Kong had become part of China in 1997 and is no longer a British colony, children born on the Chinese mainland to Hong Kong parents still must wait years before having the right of abode in Hong Kong.

Circumstances under which immigrant workers came, and their roles in the host country's economy, are central to the assimilation discourse. Japan has long maintained itself as a highly homogeneous country; yet, many Kor eans born and raised in Japan and three or four gener ations removed from their Korean descent do not have Japanese citizenship. Most are descendants of Korean migrant workers forced to come to work in mines or factories in the first half of the twentieth century, when Korea was a colony of Japan and during which Japan suppressed Korean culture and language. Japan's natur alization practices have historically been strict; persons of non Japanese descent born in Japan are not automatically given citizenship. At the same time, many Koreans living in Japan choose not to naturalize for fear that their ethnic and cultural identity will disappear and they will lose political power. For centuries, and especially during the European colonial period in the nineteenth century, large numbers of migrants from heavily populated provinces of Southern China were recruited to work in mines and plantations in Southeast Asia. They became traders, entrepreneurs, and middlemen between the colonists and the natives, and thus were perceived as part of imperialist exploitation. Since then, Chinese migrants and their descendants have played important roles in the host countries' economies and in some cases have dominated business and trade. Yet, they are subject to discriminatory policies such as restrictions on college admission, pro fessional occupations, and political participation. At the same time, the Chinese are criticized for their reluctance to integrate. Mixed marriages are not common. Ethnic tension, when fueled by economic problems such as unemployment, can escalate into violence, as seen in the 1998 anti Chinese riots in Indonesia.

National Security

The third concern is national security, which involves both issues of illegal immigration and border control, as well as more specifically responses to the September 11 (2001) events. Illegal immigrants primarily are attracted to destinations of economic opportunities and intend to search for work there. They are, however, seen as law violators since they arrive undocumented or by illegal means or they overstay without official status. As the volume, scale, and sophistication of human smuggling increase, it is also increasingly connected to transnational organized crime. Throughout the 1980s and 1990s, ships holding dozens, even hundreds, of illegal immigrants were found off the coasts of or arriving Europe and North America. Some of these ships sank or ran aground; some made it to the shore; and some arrived with mi grants who did not survive the journey. The Fujian province of Southern China, which has had a long history of labor out migration, has been a major source of smuggled workers to the US and Europe. Typically, the Fujianese are from poor, rural areas and do not fit im migration criteria that emphasize skills and family re unification. The fees that the smugglers (snakeheads) charge are in the order of tens of thousands of US dollars; thus, most migrants who successfully arrive are required to work for months or even years to pay back the debt. As these migrants find work and settle down, they send back good, if not completely accurate, news of their lives, which reduce the perceived risk of smuggling and en courage more to come. Migrant networks not only con vey information but also help potential migrants to fund the trip. Most Fujianese illegal immigrants living in New York have relatives who settled in the US earlier, who could provide loans and other types of assistance. Net work factors become more important as the migration stream matures. Indeed, globalization has greatly ex panded the means by which migrants maintain ties with their origins and reinforced the role of migrant networks in sustaining specific migration streams for a long period of time.

Illegal immigration from or via Mexico to the US, which has increased sharply since the 1980s, has been of much concern to the American policy makers and public. Illegal border crossing is seen as an act undermining the rules of entry and the national sovereignty of the re ceiving state, thus also reinforcing the negative stereo types associated with immigrants as a whole. Despite the US government's efforts in deterring their entry, many migrants have successfully crossed the border, by turning to smugglers to help them enter through deserts and mountains, via the so called banzai runs with dozens of other migrants through highways, or through other means. In March 2006, the US House passed new legis lation that would make illegal immigration a felony, impose new penalties on employers who hire illegal immigrants, and erect fences along one third of the US–Mexican border. A massive rally in downtown Los Angeles, reportedly involving more than half a million people, and demonstrations in Phoenix, Milwaukee, and other large cities, sent a strong message in protest of the legislation. In response to the political pressure, the Bush administration proposed a guest worker program but it has hitherto not received strong support.

International terrorism, and especially the 9/11 events, has driven international migration right to the top of the national security agenda, worldwide and especially in the

US. How did the 9/11 terrorists enter and live in the US is clearly a migration question and has implications not only for border control but also for immigrants as a whole. Responses by the US government included developing better technologies of border control, holding hundreds of predominantly Middle Eastern men without charges for long periods, and implementation of an electronic foreign student tracking system in educational institutions. While no concrete data exists, anecdotal evidence suggests that the US government has tightened approval of visas to foreign students, who may otherwise add to the high skilled workforce in the labor market. All of the above, together with profiling practices adopted by airport and other authorities, reinforce the perceived connection between security threats and immigrants and may further heighten racism on the grounds of national security. Despite the fact that the 9/11 hijackers were terrorists and not labor migrants, the increased fear of terrorism at the beginning of the twenty first century is likely to result in more restrictive immigration policies not only in the US but also in Europe.

The Impacts of Labor Migration

Demographic Impacts

Migration is selective, and labor migration is highly selective. Most migrant workers are young and in the working age, and thus their departure may result in a bipolar age structure in the sending area. In many Chinese villages that have experienced out migration of rural labor to cities, for example, the elderly and young children predominate. On the other hand, young migrant workers will make the population at the destination more youthful. Such changes in the age structure may have effects on birth and death rates. In places experiencing massive out migration, birth rate will decrease and death rate will increase; and in places receiving large numbers of young migrant workers, birth rate will increase and death rate will decrease. Also, migrant workers' fertility level may be higher than that of the population at the destination, as in the case of immigrants in the US. However, in most countries, the number of immigrants is not large enough to have significant impact on the overall fertility rate.

In general, migration is more selective of men than women. This is especially the case for labor migration over long distance and to new areas of development. Sex ratio in California during the nineteenth century, for example, was extremely high because the vast majority of the early settlers in the Western, frontier, areas of the US were men. However, sex selectivity of labor migration is heavily dependent on the jobs available to migrants. The maquiladoras near the US–Mexican border, for example, hire mostly women as factory workers.

Economic Impacts

The economic costs and benefits of labor migration are often at the center of debates on migration and immigration but are difficult to calculate. Standard economic theory considers migration from low wage to high wage areas a means to allocate scarce labor resources to their highest value use. In this view, the receiving economy gains and expands because of labor in migration, and migrant workers also gain in the form of higher wages. Local workers most similar to the migrant workers may experience wage decline and job loss, but these negative effects tend to be small. It is extremely difficult to answer the question whether the costs of migrant workers to the host economy, including their demand for public services such as schools and health care, are greater or smaller than the economic value they generate. The answer depends heavily on migrants' age of arrival and their years of education. Because most migrant workers are young and in their working and taxpaying years, it is generally believed that their benefits are greater than their costs to the receiving economy. In addition, because migrant workers tend to concentrate at the extremes of the education ladder, involving those with high levels of education and skills, as well as those with low levels of education and skills, they may exacerbate the socioeconomic inequality in the receiving society.

The economic impacts of migrant workers on the sending areas are usually mixed. The net effect on human capital depends on the migrants' level of skill and whether they are needed at home. Most developing countries welcome the opportunity to export their excess unskilled workers in order to relieve unemployment pressures and generate remittances. However, when skilled workers leave in massive numbers the risk of a brain drain is of concern. Some African countries, such as Malawi, train nurses and health care workers to world class standards only to lose them to countries that offer higher wages. The impacts of return migration are also difficult to discern. If migrants are returning to rest and retire, then their economic impacts on the sending areas would be small. However, migrants may return with new technologies, skills, and capital and ideas of use to them and to the sending areas, in which case they are well positioned to create jobs and promote economic growth in the home community. There is evidence that more of the highly skilled international migrants today are returning to their home countries, especially when the latter are experiencing rapid economic growth and are offering attractive financial and career incentives. A significant number of students originating from China and finishing higher education in Western countries, for example, have chosen to return. In addition, globalization has fostered a new type of transnational migrants, mostly highly skilled and some employed by multinational or transnational corporations, who maintain dual or multiple lives and identities by living and working in the host society while maintaining close ties with the donor society.

Social Impacts

Compared to demographic and economic impacts, the social impacts of labor migration are even harder to as sess, in part because social changes are usually gradual and are results of multiple factors. Nonetheless, some common observations are notable. Social relationships and social organizations in the sending areas may change. Inasmuch as labor migration is more selective of men than women, the women that are left behind in the home community may take on new roles and have greater autonomy in making household decisions. On the other hand, women may also become increasingly dependent on their husbands as the latter become the sole or most important wage earners. Migrant remittances may be spent on education for children and improving the health and livelihood of the entire family. Migration networks, as well as the circular migration and return of rural–urban migrants, may introduce ideas and practices from urban areas and eventually bring about lifestyle changes in rural sending communities.

Popular destinations of labor migration tend to attract migrants from different origins; and as migrant workers settle down and bring their families and raise children, the population in the host society becomes more diverse. The social consequences of increased diversity, especially in relation to immigration, are often controversial and hotly debated. Conventional thought emphasizes an evolution ary process of integration, including first of all adaptation and adjustment, followed by acculturation, and finally full assimilation of immigrants into the host society. Inter marriage with a member of the host society is often used as an indicator of assimilation. Critics of the above model question the inevitability of assimilation. An alternative model is multiculturalism or pluralism, in which immi grants retain their cultural and ethnic identities while living and working as full members of the host community. These different views have been driving the debates about public policy, such as those on bilingual education in US schools.

Geographically, immigrant communities have created distinct spaces and places defined by their identities and cultures. The term ethnic enclave is used to describe places such as Chinatown, Little Italy, and Little Havana in US cities where unskilled immigrant workers and their families first settled and were later joined by more im migrants from the same origin countries. These ethnic enclaves have become permanent fixtures of labor im porting countries and bases for an ethnic economy, which facilitates immigrants' access to capital, protected mar kets, and pools of cheap labor. Critics of ethnic enclaves, however, argue that these places hamper immigrants' ability to adjust to and assimilate into the host society. Nevertheless, more recent immigrants with higher edu cation and skills appear to be less dependent on tradi tional ethnic enclaves while at the same time boosting new communities that appeal to a middle class and transnational lifestyle. The term ethnoburb, for example, has been used to describe suburban areas in the US where many professionals recently immigrating from Mainland China, Taiwan, and Hong Kong live, work, and shop.

Conclusion

So long as uneven regional development exists, labor flows within and between countries will continue. Migrant work and remittances can be an important source of in come increase and diversification. Once a migration stream occurs, it can be sustained for a long time by migrant networks. Globalization has further boosted mobility worldwide and the transnational flows of mi grants with multiple ties and identities. Thus, labor mi gration cannot be switched on and off like tap water. Furthermore, historical experience shows that many mi grant workers recruited to temporarily relieve labor shortage eventually become permanent members of the host society. The impacts of labor migration on sending areas are mixed. The long term economic impacts of labor migration on both the receiving areas and the mi grant workers are positive, although there may be small, negative effects on local workers with similar skills. Demographically, young, immigrant workers can help to offset the workforce shrinkage in aging societies. Yet, the prevalence of social exclusion of immigrants, com pounded by security concerns over illegal immigration, border control, and international terrorism, has continued to define labor migration as a controversial, political, and highly contested issue.

See also: Emigration; Ethnic Economies; Ethnicity; Immigration I; Migration; Mobility; Multiculturalism; Racism and Antiracism, Transnationalism and Labor Geography.

Further Reading

Aronowitz, A. A. (2001). Smuggling and trafficking in human beings: The phenomenon, the markets that drive it and the organisations that promote it. *European Journal on Criminal Policy and Research* 9(2), 163 195.

Beaverstock, J. V. (2005). Transnational elites in the city: British highly skilled inter company transferees in New York City's financial district. *Journal of Ethnic and Migration Studies* 31(2), 245 268.

Castles, S. and Miller, M. J. (1998). *The age of migration*. New York, NY: The Guilford Press.

Clark, W. A. V. (1986). *Human migration*. Beverly Hills, CA: Sage Publications.

Fan, C. C. (2002). The elite, the natives, and the outsiders: Migration and labor market segmentation in urban China. *Annals of the Association of American Geographers* 92(1), 103 124.

International Organization for Migration (2005). *World migration 2005: Costs and benefits of international migration*. Geneva: International Organization for Migration.

Li, W. (1998). Los Angeles's Chinese ethnobourb: From ethnic service center to global economy outpost. *Urban Geography* 19(6), 502 517.

Lohrmann, R. (2000). Migrants, refugees and insecurity: Current threats to peace? *International Migration* 38(4), 3 22.

Martin, P., Abella, M. and Kuptsch, C. (2005). *Managing labor migration in the twenty first century*. New Haven, CT: Yale University Press.

Martin, P. and Midgley, E. (2003). Immigration: Shaping and reshaping America. *Population Bulletin* 58(2), 3 37.

Martin, P. and Widgren, J. (2002). International migration: Facing the challenge. *Population Bulletin* 57(1), 1 40.

Parfit, M. (1998). Human migration. *National Geographic* 194(4), 6 35.

Ravenstein, E. G. (1889). The laws of migration. *Journal of the Royal Statistical Society* 52, 241 301.

Schoenholtz, A. I. (2003). Transatlantic dialogue on terrorism and international migration. *International Migration* 41(4), 173 192.

Silvey, R. and Lawson, V. (1999). Placing the migrant. *Annals of the Association of American Geographers* 89(1), 121 132.

United Nations (2004). *World population prospects: The 2004 revision, Vol. III: Analytical report*. New York, NY: United Nations.

Weeks, J. R. (2005). *Population: An introduction to concepts and issues* (9th edn.). Belmont, CA: Wadsworth. (Chapter 7: The Migration Transition)

Yeoh, B. S. A., Huang, S. and Willis, K. (2000). Global cities, transnational flows and gender dimensions, the view from Singapore. *Tijdschrift Voor Economische En Sociale Geografie* 91(2), 147 158.

Relevant Websites

http://www.ilo.org
International Labour Organization.

http://www.iom.int
International Organization for Migration.

http://www.prb.org
Population Reference Bureau.

http://www.un.org
United Nations, Department of Economic and Social Affairs, Population Division.

Migration

P. Boyle, University of St Andrews, St Andrews, UK

Introduction

Migration is a topic which has attracted the attention of geographers, and other social scientists, for centuries. It is inherently spatial, involving the redistribution of individuals, families, and groups within and between countries. Through migration different communities are linked together in more or less formal networks of flows and the implications of these streams may impact on both the migrants and the places through which they move. Migrants are often among the most entrepreneurial and dynamic members of society, while others may be exploited and abused during forced relocation.

Consequently, migration research is incredibly diverse, including work that focuses on: the migrant and the factors that influence their decision making; the impact of migration upon migrants; the spatial patterns of migration and how they might be explained; the impacts on the origins and destination that migrants move between; and the various economic, social, and political implications of such movements. Migration research is therefore theoretically and empirically diverse and, with certain recent advances in social theory emphasizing the importance of mobility as a central construct in understanding contemporary lifestyles, it seems likely that migration will continue to be a central topic in the geographical imagination.

Virtually, everyone moves home one or more times in their lives and many will migrate over longer distances several times. For some migration is an exhilarating, liberating, and profitable experience. For others, moving may be forced upon them, embarked upon as a last resort or represent failure. Migration therefore plays a critical part in people's life courses, usually being associated with major life changes such as the move from the parental home into further education, the formation of a cohabiting or married union, a change in job, or the choice of a place to spend retirement. For those moving abroad, migration can represent even more drastic changes including separation from friends and family, the stress of the move itself, and the potential difficulties associated with coming to terms with a different cultural setting. Home and citizenship can carry contradictory and ambiguous meanings for immigrants, particularly for those who lead 'transnational' lives, which transcend localities and nations in different ways and at different times. The rules about access to citizenship vary between countries with France requiring 'cultural assimilation' and others such as Canada, the UK, and the US encouraging a more 'multicultural' approach which recognizes the advantages of ethnic groups maintaining their cultural, language, and social traditions. Thus, international migrants may challenge the dominant ideas of citizenship, particularly when they are illegal, temporary, or spend time divided between different contexts.

Of course, migration also impacts on those caught up in the moves of related individuals. Family members may be 'forced' to move home when a key figure in the household decides to take up an opportunity elsewhere, while others may suffer from being left behind while parents circulate between work at home and abroad. Migrations have significant implications for society, both at the origin and destination. The increasing politicization of immigration means that it makes the headlines in many countries on a daily basis, with some right wing political parties focusing their agendas on the tensions that arise when immigrant groups arrive in what appear from the press headlines to be large numbers. Some maintain that large scale international flows of migrants may have a direct and significant impact upon the internal migration of members of the host society and that the ripples of migration that result may be long term. Elsewhere divergent migrant identities can result in contradictory responses to incomers and outgoers. Take some of the more remote areas of Britain in the Scottish Highlands and Islands, where the out migration of young adults and the attendant implications for the labor force is lamented, while the in migration of 'outsiders' and the potential implications for the cultural identity of places may also be a matter of concern.

Migration is therefore a diverse and challenging topic. For individuals and families, it concerns everyday aspects of their lives, while migrant groups can challenge and influence the cultural and political underpinnings of communities and nations. Further, there is an increasing awareness that migrations should be classified less as events occurring at critical points in a person's life course but, instead, should be regarded as processes which are interrelated and which themselves help to shape the cultural and political identities of both people and places.

Defining Migration and Migrants

How do we decide who is a migrant? Despite the breadth of research on the topic, defining migration is not as simple as it might at first appear. We need to distinguish migration, which has a sense of permanency and involves moves to 'distant' places, from 'residential mobility' which is usually assumed to involve shorter distance moves perhaps for housing reasons, or from temporary mobility such as holidays and short term relocations. Of course, what we mean by 'long' or 'short', 'temporary' or 'permanent' vary between studies and are often influenced as much by the data resource at hand, as by theoretically guided principles. By definition censuses which usually include questions about a person's location 1 or 5 years prior to enumeration treat any move as a 'migration' although some researchers choose to separate flows that cross the boundary of some administrative area (migrations) from those that occur within these zones (residential mobility). Even then the problem remains that some moves within an administrative zone may be over a longer distance and represent more of a cultural wrench than moves between neighboring zones.

These rather technical deliberations are important for more quantitative studies of migration, but more qualitative approaches have often focused on migration as a cultural rather than a discrete event. For some, the attempt to distinguish migration from mobility or travel is an unnecessary and problematic attempt to separate blurred concepts as the points of departure and arrival are increasingly difficult to define. Thus, while those moving between nations have traditionally been referred to as 'international migrants', there is increasing recognition that this term is imbued with a rather positivist sense of relatively permanent moves from one fixed point to another. It fails to capture the fluidity of movement which many cross national movers experience and 'transnational migration' has gradually become recognized as a term which better reflects the changing experiences of many migrants during the twentieth century. For these people international moves are only part of a biography of movement between places, with some moves being more permanent than others, and many being part of a regular circulation between different places. Such transnational migrants are often bilingual and move easily between these different cultural settings, pursuing economic and political interests in both places. Whether this really is a new social phenomenon, or simply a new insight into our understanding of a process where migrants actively maintain simultaneous, multi stranded social relations linking their place of origin and destination, encouraging us to rethink the notion that people and their identities are firmly linked to specific places and questioning the idea that these identities are stable, continues to be debated. Critics suggest that transnational practices have existed for some time and that eventual integration should not be precluded from studies of international flows. The 'myth of return' is an old fashioned term which still has some purchase for those migrant groups who gradually come to terms with the inevitability of their new homes, despite a yearning for their place of upbringing which may never recede.

Another key distinction in the migration literature is whether the move is voluntary or forced. This over arching distinction helps to distinguish between certain groups, such as 'refugees', who may be forced to leave home because of the political situation, and those who make well planned decisions to move elsewhere voluntarily. Again, even this simple distinction is problematic. According to the 1951 United Nations (UN) Convention definition, refugee status is only conferred upon those who have a well founded fear of 'being persecuted for reasons of race, religion, nationality or political opinion' and who move between nations. Movers within countries are not strictly included, and there are far more 'internally displaced persons' than there are international 'asylum seekers' who have moved and are awaiting judgment on whether they have achieved refugee status. Nor are people forced to move for nonpolitical forms of persecution, such as those who face extreme hardship because of their ethnicity (the 'ethnic cleansing' in the Balkans during the 1990s is an extreme example of this), or for other reasons outside their control such as environmental disasters. Partly because of some of these problems, the definitions of refugee and asylum seekers that are adopted vary by country, often influenced by the political context at the time. Certainly, the press struggles to distinguish correctly between 'asylum seekers', 'refugees', 'economic migrants', and 'illegal' or 'irregular' migrants, as these terms seem to be used interchangeably, often in rather dismissive and contradictory ways. More recently, though, while the United Nations High Commissioner for Refugees (UNHCR) has not managed to revise the 1951 Convention, which has the major advantage that such a large number of countries are signatories, it does take a more pragmatic view of who should be treated as deserving of refugee status.

Even the use of the term 'voluntary' is not without its difficulties. Taking the example of internal, within nation, migration, the neoclassical assumption that people make relatively 'rational', voluntary decisions to move, perhaps in search of employment, fails to capture the subtleties of many migrants' experiences. The stereotypical 'speculative' economic migrant who moves in search of better prospects is a rare beast as most employment related moves are 'contracted' with the migrant moving once an employment opportunity has been secured. Even if the move can be regarded as 'voluntary' for the person stimulating the decision, for the other family members who are bound up in this move the decision may well be involuntary. More significantly, the term voluntary also implies that migrants act out these decisions, devoid of the various family, employment, and political structures that shape everyday decision making. Many moves are in fact influenced by the decisions of others, even if the final decision to move may apparently rest with the migrant.

Thus, population migration is a diverse and complex subject which is difficult to define. It captures a myriad of different moves which are stimulated by various factors, some of which are more observable than others. It involves everyday experiences, such as single people moving to taking up new jobs in a different location, as well as some of the most profound changes to global society, such as the Jewish diaspora which has involved the relocation of thousands over centuries, whose ancestors may retain a strong yearning to return 'home'. It is no wonder, then, that the research conducted on the geography of migration is also extremely varied.

A Brief History of Thinking on Migration

The philosophical approaches to geographical migration research have matched most of the major evolutions in ideas that have swept through the discipline of geography as a whole over the past few decades. Indeed, migration research has often been at the forefront of these philosophical developments. Much of the early research on migration focused on the identification of spatial patterns of migration and regularities in the characteristics of migrants which helped in our understanding of the causes and consequences of migration flows. In the nineteenth century, Ravenstein developed his 11 major 'laws of migration' which summarized the regularities that could be identified from data gathered from the 1871 and 1881 censuses. Even today some of these ring true as, for example, the majority of migrants continue to move over short distances and large towns still grow more from migration than from natural increase. Other general statements about migration have also had a long lasting influence on the subject. Lee's work in the 1960s developed Ravenstein's laws into a series of hypotheses concerning the volume of migration and its streams and counter streams, with a focus on the 'push' factors at the origin which stimulated moves, the 'pull' factors at the destination which attract movers, and the obstacles that 'intervening opportunities' between any pair of potential origins and destinations might present.

The development of such empirical laws that apparently governed the observed patterns of migration continued into the 1970s and Zelinsky's 'mobility transition' was an ambitious attempt to generalize the transitions that take place in the rates and scale of migration as society modernizes. This is perhaps the best example of a 'grand theory' of migration. The first, preindustrial, phase was characterized as having relatively limited and generally local levels of migration. Migration rates were expected to increase considerably in the modern period, but then to wane again in the fifth, postindustrial, phase when circulation becomes more common than migration *per se*. While such attempts to define an overarching model of global migration patterns would be regarded as naïve nowadays, this approach certainly shaped much of the research that followed during the next couple of decades. The focus on circulation in the last phase of modernization chimes with the recent focus on transnational migration.

During the 1970s and 1980s, insights from neoclassical economics began to dominate the more quantitative approaches to geographical migration research. More emphasis focused on employment related moves with the assumption that workers respond to wage variations between places in rational ways – people were expected to move from low to high wage and from high to low unemployment places. This type of approach was popularized with the research by Todaro and others working in the global south where rural to urban migration was burgeoning. The 'human capital' model was a key extension of this work which focused on individual decision making and posited that rational potential migrants would assess the anticipated future stream of monetary and psychic consequences of migration. Although the majority of the subsequent research focused on anticipated differences in wage rates, the human capital model was a much broader approach which also recognized that a broader temporal perspective was required to understand migration decisions. Apparently irrational moves which were financially detrimental to the migrant household in the short term may well have made sense in the longer term, such that migration could be regarded as a form of investment based on anticipated gains.

Despite the conceptual advance of the human capital approach, its 'gender neutrality' has been argued to be unrealistic. In fact, there is growing evidence from the 'family migration' literature, which focuses on internal migration within countries, that moves tend to benefit the man more than the woman, at least in terms of labor market outcomes. The human capital model also posits a rather unrealistic assumption that people acquire and act upon perfect information. The approach removes the hurly burly of daily life from migration decision making and treats the migrant as a passive receptor of information. More behavioral accounts therefore developed during the 1960s and 1970s which stressed the mechanisms behind individual acts of migration and put the individual agent at the heart of the process. The role of perceptions and the fallibility of the rational decision making model were emphasized in these approaches as it was recognized that people do not always take 'perfect' decisions. Naturally, these philosophical leanings encouraged more qualitative approaches to be adopted, particularly as more strictly 'humanist' accounts developed which focused attention squarely on the beliefs, aspirations, and obligations held by migrants themselves, rather than the more objective characteristics, such as income and education, which can be measured for these individuals. Thus, while it remains important to under stand the economic incentives of migration, given that so many flows are driven by employment related factors, more realistic models which incorporate gender roles and individual attitudes and expectations have become more common.

None of the approaches described above took much account of the institutional and political structures which influenced migration flows. More structural approaches emphasize the conceptually deeper processes which underpin society, and which require theoretical insights, rather than the collection of observable characteristics. Examples of this work included studies of the role of colonialism and capitalism and the global implications for international migration between countries which have experienced uneven development. For example, the slave trade and its implications for millions of African slaves were organized through colonial structures which linked Africa and the New World, while the rise of world cities and skilled expatriate migration is a clear example where migration provides a linking mechanism for global cap ital. The rather extreme economic differences that exist between countries in the global north and south have not diminished in relative terms, even though economic conditions in many poorer countries have improved in the last few decades. Employment rates and wage gaps are often cited as the major reasons for stimulating mobility in such unequal contexts, with international migrants often being attracted to work in lower status, more poorly paid work in countries where the wages for these jobs are considerably more than could be earned at 'home'. In a growing number of cases, national governments in poorer countries play an important role in encouraging young adults to seek work abroad, in the hope that remittances will be returned. At the same time governments in over 30 receiving countries actively promote the immigration of skilled workers through various programs such as the 'Business Immigration Programme' in Canada or the 'fresh talent' initiative in Scotland. Through these bur eaucratic practices, immigrants are categorized along a spectrum of desirability which reflect the imagined needs of nation states. Thus, nation states play a powerful role in mediating migration, by scripting the identities of transnational subjects, while the very existence of the nation state rests on the states' ability to police borders and to define and restrict membership through citizen ship and other legal processes.

Alternative structures which have influenced thinking in migration research include 'patriarchy' which repre sents the subordination of women through a set of male dominated system of social structures. Unequal power relations within households mean that migration is often undertaken on behalf of the man, while the women (and children) may be obliged to support such a move. It seems that 'gender roles', rather than 'relative resources' of partners, drive such decision making, as described below. Certainly, more feminist approaches have gained considerable purchase in a post structural literature which has recognized the growing importance of women in global migration streams. It has been argued that until the 1980s, migration was conceptualized largely as a male phenomenon, with discussions of migrant workers con figured through masculinized definitions of mobility and economic activity. Some of the original theories of mi gration, such as the 'push/pull' approach, can be criti cized for their emphasis on the autonomy and agency of male migrant workers and their focus on employment related mobility. Various approaches, including the 'household strategies' perspective, have since emphasized the role of women and the household context within which migration decisions are made. Thus, increasing attention has been paid to the feminization of migration, particularly in the transnational migration literature where women are acknowledged to play an increasingly important role as primary migrants and this has led to a reappraisal of theoretical explanations of migration.

Overall, migration research has a long and impressive history which has been guided from most of the major philosophical advances that have occurred within the discipline of geography. There has been a growing rec ognition of the need to move beyond an objective in terpretation of migration as a reaction to a series of push and pull factors in relation to economic forces, with migrants' identities gradually being acknowledged as important factors shaping mobility and the interpretation of economic variables. It has also been recognized that simple models of discrete migration events may not capture the diversity of migration experiences which are interlinked throughout a person's life course. Not surprisingly, given this breadth of theoretical work, the methods used to explore migration also vary considerably.

Methods of Migration Analysis

The methods used to analyze migration can be broadly divided into quantitative and qualitative approaches, al though there is a growing body of more mixed method research. Quantitative studies include descriptive ap proaches, where flows of migrants may be summarized using different types of rate such as in and out migra tion rates which compare the size of flow to the popu lation at the destination or origin, respectively. Other descriptive measures used to summarize migration pat terns include those such as 'demographic effectiveness' which compare the size of net flows to places with the total in and out flows, or measures of segregation such as the 'dissimilarity index' which describe the geographical mixing of different population subgroups such as ethnic or migrant groups.

However, quantitative migration research has also been at the heart of some of the more advanced modeling research that has been conducted by geographers. These include aggregate models of migrants flows between places, many of which are underpinned by the application of Newton's 'gravity model' which assumes that flows between different areas are proportional to the size (or populations) of these areas and inversely proportional to the distance between them. These 'spatial interaction models' are statistical approaches which can be fitted using different modeling assumptions and are now commonly based on the Poisson or negative binomial distribution. They can be used to identify the factors that influence migration flows, as well as to identify residual flows which are much larger or smaller than might be expected.

Micro level quantitative models of individual decision making and migrant behavior have also been applied, which deal with the individual level factors that stimulate behavior. Such 'discrete choice' models simply distinguish between those who do and do not migrate using a binary outcome variable. Rather more complex, and potentially much more enlightening, are quantitative 'longitudinal' models where information about individual's behavior is captured at more than one point in their lifetimes. Panel data sets, where a sample of respondents is repeatedly interviewed on a frequent (often annual) basis, are one rich source of longitudinal data. Other sources include birth cohort data sets, which have been successfully established in the UK and have involved capturing everyone born in a particular week and then re interviewing them at different critical points in their lives, and linked administrative data which make use of routine census and other data sets to build up comprehensive pictures of people's lives through time. A major advantage of longitudinal data is that migration events can be regarded as a series of transitions which occur during a person's life course. Rather than focusing on cross sectional groups of migrants and nonmigrants (as is necessary in many analyses), migration can be linked to previous characteristics and to future outcomes in 'event history analyses'. Longitudinal methods have therefore been applied to numerous migration topics including, for example, housing careers, employment careers, and the links between migration and subsequent demographic events such as fertility.

Of course, qualitative approaches have also been widely implemented in migration research. Interviewing is a standard approach which offers the value of being able to collect the primary data that are required. Most secondary data, even those collected in the richest of panel data sets, usually only include a small number of questions on migration behavior and the underlying causes. Thus, quantitative studies that rely on secondary data often include relatively limited information about the underlying causes of migration decision making. Interviews and more ethnographic techniques have been usefully applied to various migration questions and these allow a greater depth of understanding about the complex range of factors that influence migration processes.

Biographical approaches have also been advocated in recent years. More than simply ethnographic studies, this approach is philosophically different, drawing upon structurationist theory and attempting to collect information about the wider contextual processes that influence migration decision making. Explicitly recognizing the structural constraints within which migrations occur, it focuses on three major issues. First, migration is seen as an action in time, rather than as a discrete event. This encourages more longitudinal conceptualizations of migration processes, be they quantitative or qualitative. Second, each migration may be influenced by a range of overlapping causes, rather than being explained by a single (usually economic) factor. Some secondary data sets may include a question about why people moved to their current location, but the assumption that such moves can be attributed to a single factor may well be naïve. A move to a particular city may have been stimulated by a job opportunity, but the choice of the precise residential location will have been influenced by a range of other factors such as the cost of housing, the quality of local schooling, and more environmental considerations related to the residential neighborhood. Third, migration is seen as a cultural event, where migrant actors are embedded within, and themselves influence, a cultural and political context. This approach therefore moves beyond the more neoclassical assumptions which treat migrants as relatively isolated decision makers. Thus, this approach explores the narratives of everyday life within which migration is situated.

The breadth of migration research, both philosophically and empirically, has resulted in a wide body of interdisciplinary work that geographers have made a major contribution to. We understand a great deal about the dominant patterns of flows, the characteristics of migrants, and the causes and consequences of migration. One rather obvious, but crucial, observation is that migrants are a selective group – not everyone migrates and some people migrate considerably more than others.

Migration and Selection

Migration is a selective process which involves some population subgroups more than others. Ravenstein's laws were an early attempt to summarize some of the distinctive characteristics of migrants, focusing mainly on their demographic circumstances. Today, migration is still more common among certain groups than others. For example, in the global south, migration tended to be dominated by men who moved in search of employment opportunities, often spending time divided between rural and urban areas where employment opportunities arose at different times in the year. However, we have already seen that a growing number of migrants in poorer nations are now women for whom more varied employment opportunities are available and, in relation to gender at least, it is clear that the selective nature of migration has changed. Indeed, in some cases it is more that the perception of migration needs reconsideration. The example of Irish migrants to Britain is a good example where the statistical evidence shows that women have dominated these flows for many decades, but the dominant popular stereotype has been of the 'mick' or 'paddy' image and, until recently, academic research has also focused primarily on the male migrants.

Migration is also selective of age. Migrant 'age schedules' are often plotted to demonstrate how the rates of migration change through the life course. Focusing on internal, within nation, movers we know that migration is relatively common for young children, as families often move home to accommodate the growing size of the household and the different requirements that children have. Family migration then declines as children enter school and become embedded in social networks which are difficult to break. It then rises to its peaks in early adulthood, with those leaving home for work and education, gradually declining then with age until, for some, there is a small peak around or just prior to the age of retirement.

However, migration is not only selective of certain demographic characteristics. The highly educated, those with higher incomes and people in certain types of occupation, such as the armed forces, are also more likely to move around. Indeed, some researchers divide migrants into 'movers' and 'stayers', the former moving quite frequently during their lifetimes and the latter being much less likely to move at all. For some stayers 'cumulative inertia' may set in as the longer a person or family resides at a particular place, the less likely it is that they will move on.

One growing area of migration research concerns health selective mobility. On the whole, migrants are usually healthier than nonmigrants, which is often referred to as the 'healthy migration effect'. For example, international migrants are usually found to have better health than those they leave behind. For some groups, such as those moving from poorer to richer countries, their health may not be as good as that observed among those they join. For other migrant groups, their health is better than both those they leave behind and those they join. A number of studies have therefore considered migrant health as an opportunity to tease out the influence of 'nature or nurture' on different diseases. If genetic effects are most important, we would expect that the rates of disease in a particular migrant group would remain similar to the rates observed in their home nation. On the other hand, if the rates of disease among the migrant group change over time to be closer to the rates in the destination society, then environmental factors are likely to play some role. This effect has been observed for numerous conditions including breast cancer among Japanese women migrants into the US, for example. Rates of this disease are low in Japan and high in the US and Japanese women who relocate into the US experience higher rates than those they leave behind suggesting that some aspect of the environment is having an influence.

Similar selection effects have also been observed for migrants within nations, with young adult migrants tending to be healthier than nonmigrants. However, this reverses in old age, when migrants tend to be less healthy than nonmigrants as many may be prompted to move into care or to be close to relatives when health begins to fail.

Such selection effects are important as they have implications for more general studies in which migration may not be the central focus. Numerous geographical and epidemiological studies ignore this potential biasing effect of health selective mobility. A simple geographical example would be a study which compares health outcomes by the deprivation category of the residential area, with the obvious expectation that health will be worse in more deprived places. However, while it is quite plausible, and indeed likely, that living in deprived areas will lead to poorer health, it is also possible that those in poor health are more likely to end up living in more deprived areas, because of health selective migration. Those in good health are more likely to leave deprived areas, and those in poor health may be more likely to end up or remain in these poorer places. Thus, at least part of the relationship between health and deprivation may be caused by these mobility patterns and, if ignored, would result in inaccurate conclusions about the causal influence of deprivation on poor health. This issue is a useful example of how valuable a longitudinal approach can be, as this allows the direction of effect to be considered in ways that are impossible using cross sectional data.

Of course, the selectivity of migration varies between different contexts and by the type of migration that is involved. Thus, we would expect different types of movers to leave different origins and choose different destinations. Young adults may be attracted to the bright lights of urban life, while families with, or expecting to have, children may be more likely to move away from urban centers into suburbs or surrounding rural areas. Thus, migration selectivity underpins the patterns and scale of migration that are observed.

The Scale and Patterns of Migration

The broad patterns of migration are well known to us, although new and diverse sets of migration streams are forever appearing. Mobility has become such a wide spread practice that some talk of the 'age of migration', as few people worldwide have no experience at all of mi gration and its effects. Despite the apparent scale of these flows, though, estimates suggest that only around 3% of the global population is an international migrant, a sur prisingly low figure given the regular headlines in many countries which highlight the ever increasing numbers of immigrants. There is also growing evidence that migra tion rates have not increased as dramatically over time as was once thought, at least within nations.

Providing entirely reliable estimates of the global scope and scale of migration is virtually impossible. Even estimates of local flows within many developed countries are difficult to produce accurately, particularly in coun tries where no formal registration system exists. The population estimates derived from national censuses, which provide the only national count of the population in some countries, are known to be least accurate for the most mobile groups such as young, single adults. Recently, it has become clear that the official statistics on flows into and out of the UK, for example, are particularly prob lematic as no formal system for recording such migrants is in place and small scale surveys of people passing through (air)ports are relied upon in the absence of better data. The estimates of flows between countries are even more difficult to derive if irregular migration is common. The number of Mexicans who have passed the border into the US may be in the millions, but it is extremely difficult to provide an accurate estimate of the true scope of this economically and politically important migrant stream.

Of course, although we have an impression that mi gration has reached unprecedented levels, which may be true of some origin–destination flows, it is certainly the case that major movements of people have occurred for centuries. Within eighteenth century industrial Europe and the US industrialization and the lure of the city resulted in massive flows of people from rural areas into urban centers. There was a limit to the number of people that could be accommodated in these cities and by the end of the nineteenth and beginning of the twentieth century, 'suburbanization' became the dominant pattern. Originally home for the working classes, the city suburbs became increasingly regarded as attractive, family friendly residential environments so that by the 1950s the suburbs became a culturally distinctive and popular site for a growing migrant middle class. Between the 1960s and 1990s, 'counterurbanization', or the net decline of the cities and net growth of the more accessible rural areas beyond the city limits through migration had become the dominant spatial pattern in much of Europe and the US. This represented a radical reversal of the population concentration that preceded this period and most suggest that this was caused by a mixture of factors including quality of life decisions, economic restructuring, and industrial (re)location.

However, from the 1980s onwards this counterurban pattern has slowed, and in the US at least some metro politan areas began to grow once more. Thus, some cities have now reversed the pattern of net out migration, as inner city areas 'gentrify' and attract a new group of mainly young inner city inhabitants, attracted by the various leisure and social opportunities afforded there. Some suggest that this reversal in fortunes for certain urban centers is strongly related to economic cycles and that job opportunities are underpinning this trend, while others point to the growing importance of immigration and the higher fertility rates displayed among these groups in certain cities. For example, the population growth of some global cities, such as London, has been driven more by international than internal migrants. Ei ther way, it is clear that the dominant counterurbanization process has slowed and that for some groups at least, urban areas offer attractive, if only temporary, locations.

The lure of the city in countries in the global south has long been the dominant migration flow and in many countries it continues to be biased toward primate cities, although this focus seems to be diminishing to some degree. One of the most significant flows of migrant workers occurs within China, where there are estimated to be around 140 million internal migrants which, when compared to the 200 million international migrants worldwide, demonstrates the scale of this particular mi grant flow. This is despite the fact that the 'floating population' of movers from rural to urban areas are not formally acknowledged in the Chinese *hukuo* household registration system where they are treated as temporary migrants and as a result are channeled into the poorest paying jobs and are not entitled to the subsidized services available to urban residents.

The dominant patterns of international migration are a mixture of historically consistent flows which involve linked networks of particular origins and destinations, and new patterns which often emerge as a result of political restructuring or unrest. According to the UN, international migrants are those who have resided in a different country for at least one year. Major flows through history included the estimated 12 million Afri can slaves who moved to the New World during the eighteenth and nineteenth centuries; another 12 million economic migrants who were processed through Ellis Island between the 1850s and 1930s on their way to a new life in the US; and 1 million migrants from Britain who made their way to Australia as 'Ten Pound Poms' whose passage and a £10 grant was paid by the Australian government.

One estimate from the UN suggests that in 2005 there were around 200 million international migrants, although this is likely to be a considerable underestimate as 'illegal' or 'irregular' migrants are extremely difficult to collect reliable data on. In certain contexts, the scale of this migration is remarkable. Something like 35 million Chinese currently live outside China, while in 2000 it was estimated that around 20% of all migrants lived in the US. In addition to these large flows of reasonably well documented migrants, the number of irregular migrants seems to be growing. For ex ample, there are purported to be around 10 million irregular migrants in the US, most of whom originated in Mexico.

As described above, many migrants are forced to move because of the political circumstances. The definition of refugees which was achieved in the early 1950s was driven by the need to resettle many who had been forced to migrate because of the Nazi threat. By the 1960s, major refugee flows were growing in Africa and by the 1970s Southeast Asia began to experience large flows, particu larly associated with Vietnam and the 'boat people' who fled following the communist victory in 1975 and the creation of the Socialist Republic of Vietnam. In the 1990s, a significant change was the growth of refugee flows in the global north, including Bosnia and Kosovo. Refu gees continue to make up a major part of recent inter national migration flows – in 2005 there were estimated to be around 9 million refugees worldwide. Particularly large flows include those from Afghanistan, Sudan, the Democratic Republic of Congo, and Somalia. However, while the scale of these moves remains large, there are considerably fewer refugees today than in the last couple of decades – in 1990 there were reported to be around 17 million worldwide. While the public impression in many richer countries is that they are increasingly expected to bear the burden of refugee flows, the vast majority of such migrants originate in and remain in the poorest parts of the world. Flows within Africa, some of which occur over short distances between neighboring countries, still make up the majority of refugee movements worldwide.

It is indeed a mobile age but, as argued above, the proportion of the world's population who are inter national migrants remains relatively small. This is caused by a number of factors, including: the inability of many of the world's poorest to move elsewhere; simple inertia as most people are content with the circumstances in which they live and never contemplate long distance inter national moves; and the role of governments who often influence migration across their borders with some, such as that in the Philippines, actively encouraging inter national labor mobility and others, such as those in a number of communist countries, strongly resisting it.

While the geography of major migration flows has changed over time, some of the underlying trends asso ciated with these moves have also changed. A far higher proportion of international migrants than ever before are women and this trend is likely to continue in the future. In the past, most international women migrants moved to join their partners as 'family reunification' was a major cause of much mobility. Nowadays women are far more likely to move independently and many are the primary breadwinners for their families. Many of these transna tional migrants are engaged in the reproductive, rather than productive, sphere, working as maids, or in enter tainment, healthcare, or the service industries; for some parts of the world, such as the Philippines, these migrants make a major contribution to local and national econ omies. Indeed, in many parts of the world the scale of domestic worker migration (the 'maid trade') far out weighs the movement of professional workers.

There has also been a change in the origins and destinations of migration. Some traditional centers of out migration have more recently become attractive lo cations for in migrants – North African is one example which lost, and continues to lose, many migrants to Southern Europe, but which is now becoming a popular destination for migrants from sub Saharan Africa. While most international migrations were (or were assumed to be in the literature of the time) permanent, far more people move temporarily. Today, there is a growing emphasis on the role of 'circulation' with large numbers of migrants leading dual lives divided between countries where they may spend months or years at a time.

These large scale movements of economic and polit ical migrants have implications not only for the migrant groups themselves, but also for the origin and destination areas which lose and receive them and in many cases the impacts of migration can be quite severe.

The Impacts of Migration

As discussed above, for many people migration is asso ciated with new opportunities and can be regarded as a positive experience which can be exciting and chal lenging. For others, though, migration may be a negative experience, perhaps forced upon them, by circumstances outside their control. For both host societies, who have to cope with what the right wing press often describe as 'migrant floods', and origin societies, who may suffer the implications of a 'brain drain' that accompanies skilled out migration, the impacts may be hugely significant at a societal level. Consequently, migration has impacts on the migrant themselves, but it also impacts upon the places they move between. It is important to realize that these impacts will not be static, but will vary over time.

The Migrants and Their Families

Moving is a potentially stressful event, occurring rela tively rarely for most and involving considerable planning and expense as well as causing the breakup of social networks and the formation of new ones. Much migration occurs because the migrant(s) involved aim to improve their quality of life, perhaps through moving into a more suitable property or by choosing a more desirable neighborhood. These types of moves may well be over relatively short distances and are usually described as 'residential mobility' rather than migration. Longer distance moves are more likely to lead to greater disruption for the migrants involved, particularly when moving to a country which uses a different language, or to a place where there are few social contacts or support networks. Moving into new locations may even bring with it the risk of being introduced to new health risks, although in many cases the migrants themselves are often feared as the harbingers of disease.

These more negative consequences of migration are much more likely when the moves are forced or impelled, particularly if those involved have no choice but to leave possessions and, often, other family members behind. 'Human trafficking' is perhaps the most extreme example, and according to United Nations Children's Fund (UNICEF), over 1 million children per year are trafficked into the modern day equivalent of slavery. These abducted children are secretly transported across borders and sold like commodities or trafficked within countries for the sole purpose of exploitation. For these vulnerable individuals, migration represents the removal of basic human rights and an unpredictable and dangerous future. It can involve being forced into prostitution, illicit international adoption, or recruitment into armed forces. Among trafficked adults most are women, mainly forced into the sex industry. Trafficking is different to 'smuggling' which involves migrants voluntarily embarking on a move organized by smugglers. However, while these types of moves may not result in long term exploitation, on some occasions they can if the payment to the smuggler is not completed before the move.

Of course, the impact of migration may not be so severe, and in many cases moving home may involve a mixture of benefits and costs. Families who move because of a change of job of one adult (usually the man) may well have weighed up the potential pros and cons, although it seems unlikely that they will do so in an entirely rational way as posited by some neoclassical theories. Even so, it is probable that long distance moves will only be made if the benefits of the new opportunity appear to outweigh the attractions of staying put. One outcome of such 'family migration' is that women usually suffer from these moves, at least in terms of labor market outcomes. Thus, the consensus in a literature which spans over four decades is that women are more likely to be out of work, have smaller incomes, and work fewer hours than their nonmigrant women counterparts. Thus, the terms 'tied migrant' or 'trailing spouse' have been used to describe those who may lose out from migration individually, even if the family as a whole benefits, and in most cases they are women. 'Tied stayers', on the other hand, are those (mainly men) who are discouraged from moving because of the implications it would have for others in the family, although defining and providing estimates of this group is rather more difficult to achieve.

Coming to conclusions about the benefits or costs to certain individuals is not always a simple task. Take, for example, the many thousands of transnational migrant women who leave their families to work elsewhere, often abroad, in order to send remittances home to their (transnational) families. Caught between two worlds, these women may inhabit rather contradictory spaces, often regarded as both insiders and outsiders at the same time. They may experience a sense of 'placelessness', 'floating' between locations and finding it difficult to identify with either place as 'home'. The marginalized position of migrant domestic women at the edge of society raises the intersecting relations of gender, ethnicity, class, and citizenship and there is clear evidence that these women suffer discrimination and varying legal rights across destination countries. Taking the Asian example, female migrant workers in Singapore are constrained by tight labor contract systems and little tolerance for civil activism, while in Hong Kong quite frequent demonstrations by domestic workers are tolerated and collective action therefore provides an opportunity to challenge some of the worst discriminatory practices.

To certain researchers, these women are to some degree 'forced' to move by feelings of family obligation, and living away from family and friends in circumstances that are often quite difficult may have serious emotional implications for the migrants and their close relatives 'left behind'. The impacts on their children and the elderly may be particularly severe as the family is reconfigured and transformed by these migratory processes. On the other hand, many of these women enjoy the economic independence associated with such moves, as well as the escape from traditional attitudes concerning women's roles at home. For these transnational women, life can be empowered by migration and the economic and social independence that it brings.

Thus, while certain types of migration can be summarized as being broadly positive or negative, many moves bring with them a mixture of costs and benefits, which are likely to vary for different household members. Similarly, it is not simple to determine whether migration has a positive or negative influence on the societies between which migrants move.

The Origin and Destination Societies

The various impacts of international migration on the host and origin societies have been well documented and they involve geographical, economic, and more socio-political implications. For origin countries a major impact of migration relates to 'remittances'. The World Bank suggests that migrant labor earns around US$20 trillion and in 2005 that about US$200 billion was sent home by migrants, much of it being returned to poorer countries where most migrants originate. Including informal remittances might lead to a doubling of that figure. The channels through which remittances are returned are diverse and a migrant industry has developed which facilitates the movement of people, their possessions, and income between nations. Some transfers are passed through formal channels and can be recorded, but most are informal and estimating their size is very difficult. Countries such as Mexico and the Philippines are major recipients of remittances, while the US, Saudi Arabia, and some Western European countries are major sources. How these are utilized in the origin country varies considerably between communities and families. In some cases remittances are invested entrepreneurially in businesses, or community based enterprises. In other cases, their uses may be less obviously productive, being spent on consumer goods or, in certain scenarios, being used to support civil campaigns or conflicts.

Overall, the redistribution of wealth and the consequent financial benefits that such remittances provide may seem obvious, but certain commentators regret the dependency that some place on these income streams. However, such migrations also supply a variety of 'social remittances' as new ideas may pass between origins and destinations, and migrants can continue to play a major role in political outcomes, as many retain their rights to vote while abroad. Much less research has considered the social and emotional impact on children and older people as members of their families leave for potentially long periods of time. The implications for caring are potentially serious, particularly as such a high proportion of transnational migrants are now women.

One drawback of these flows of migrants for origin societies is the fact that migration is selective (see above). The 'brain drain' refers to the loss of skilled migrants that some countries suffer. Although these migrants may make an important contribution to their home societies through the remittances that they provide, those that leave are often among the most skilled and educated. In some cases, there are fears that this will have deleterious implications for the labor market at home and the actions of richer countries who actively attract certain types of key workers when they are suffering shortages (the UK National Health Service has sought nurses from abroad for example) can make this problem particularly vexing, leading to calls that some form of compensation should be paid to the, usually poorer, countries who have invested considerable resources training these individuals. Other countries, such as Singapore, have launched impressive public education campaigns, designed to encourage workers to take advantage of the skills training that can be acquired abroad, but emphasizing that their heart should remain in Singapore. This 'global city state' aims to embed itself in an extraterritorial terrain of network relations, from which profit can be extracted and one way of extending its influence is through the encouragement of inward and outward flows of capital, goods, services, information, and people.

In relation to the destinations, many argue that migrants are among the most entrepreneurial of groups and that most societies benefit from their presence. Most migrants are economically active and in many countries within Europe and elsewhere in the developed world where the population is aging and birth rates are well below replacement level, there is a growing recognition that migrant labor may be the only solution to the growing 'demographic deficit'. It is no surprise that so many countries actively encourage skilled immigrants to enter, but at the same time they actively discourage the entry of other groups with different backgrounds.

Thus, despite the benefits that most migrants bring they are also threatening for some in the destination societies. Migrant bodies are often depicted as diseased, racialized, criminalized, and undeserving. They may compete for jobs, one consequence of which is the depression of wage rates. Many complain that they make undue use of public services, especially healthcare. The impression of immigrants is therefore contradictory. While celebrating the diversity of ethnic difference in some aspects of daily life, such as the increasing breadth of cuisine available in most Western countries, fears of being under siege from groups who will resist being 'assimilated' are commonly portrayed in the media. The irony of this perspective in New World countries where the white majority are themselves descendents of immigrant groups is rarely acknowledged.

Research on the geography of immigrant settlement contributes to the tendency among some to maintain a 'them' and 'us' perspective with its focus on segregation, ghettoization, and 'proximity to whites'. Whether ghettos ever existed remains a debated point and, in an increasingly mobile world, the ability of ethnic groups to retain a cohesive community without clustering geographically is becoming more recognized. More recent research, particularly in the US, which apparently links the flows of incoming international migrants into key 'port of entry' states and cities with the selective out-migration of lower class whites who find the competition for jobs and housing too difficult, helps to fuel such concerns. Referred to by some as a 'demographic balkanization', critics have argued that the link between these two broad migration processes is at best weak (the overlap between immigrant and native employment markets is not great in most places) and that the use of such terminology is particularly problematic and in flammatory suggesting, as it does, an immigrant induced breakup of a unified nation.

In some countries migrant groups are regarded with suspicion and even feared as the links between migration and security are emphasized. Recent terrorist atrocities in many countries have fueled a right wing reaction against 'others' who appear to threaten national sovereignty. It is certainly true that the political significance of migration, and particularly irregular migration, has been and con tinues to be a major concern for many societies. Per ceptions of 'floods' of migrants are fueled by certain political parties and press organizations, and the sur prising success of certain right wing politicians, such as Jean Marie Le Pen in France, Pauline Hanson in Queensland, Australia, and the late Pim Fortuyn in The Netherlands reminds us of the level of support that these groups can muster.

The impacts of migration on both migrants and the places through which they move are often significant but rarely simple to elucidate. In many cases migration is as sociated with a series of costs and benefits, which are often very difficult to quantify or judge against one another.

Conclusion

The topic of migration has attracted the attention of geographers for many years as spatial mobility and movement are inherently geographical. The approaches taken to understand population migration are theoreti cally and empirically diverse. The contribution of geog raphers to our understanding of the patterns and impacts of such flows is significant and, as advances in social theory have been adopted in the discipline, the study of migration has developed in new and exciting ways.

Emerging migration trends which are likely to figure in both academic and popular discourses in years to come include the growing importance of women in inter national flows, the increasing importance of skilled mi grants in the political agendas of origin and destination societies, the growing number of students who move easily across borders to improve their skills, the in creasing reliance on remittance income in some of the poorer countries, the role of global cities as stopping points for migrant elites and the ever more mobile 'creative class', the growing importance of Asia in terms of global migrant flows, the religious, cultural, and ethnic tensions that arise from immigration, and the perceived threat to security that immigrants bring. These subjects raise significant questions of political, economic, and social change.

Further, theoretical thinking in the social sciences more generally is likely to become increasingly influenced by the appreciation that mobility in general is becoming a defining element of global society. Although migration rates may not have changed as much over recent times as people imagine, the significance of migration in helping to shape and, for some, undermine the nation state is be coming increasingly recognized. Post structural approaches to the geographies of the nation state stress the importance of the categorization of immigrants, particularly in relation to the way that international borders are policed. As 'here' and 'there' breakdown for many transnational migrant groups, concepts of hybridity, borderland, and flow become increasingly pertinent in recent geographic research agendas, such that mobility, or 'migrancy', has become a central theme for many commentators in the social sci ences. Indeed, some argue that the sanctity of the nation state is eroded by the international 'space of flows' of capital, commodities, information and labor, as the circu larity of transnationalism challenges the concept of a state with guarded borders which has a goal of reproducing national citizens.

There is an increasing recognition that migration is a complex and multifaceted process, rather than as a dis crete event; the concept of migration is more fluid than many commentators have assumed in the past. The temptation to treat migration events as discrete and in dependent of one another is therefore waning as migra tion is being regarded more a process, both for the migrants and the places through which they move. Indi vidual biographies can be mapped as a series of discrete events, but they can also be regarded as journeys which consist of a series of interrelated start and stopping points that overlap and intersect with the journeys of others. The irony, of course, is that while transnational migrants are apparently less likely to forge meaningful bonds with destination societies, which for some might appear to reduce the 'threat' of immigration at least in terms of absolute numbers, it also raises the prospect of a growing body of incomers who are less inclined to adopt national norms and values than were previous waves of immi grants. The tensions associated with immigration for destination country governments therefore become in creasingly difficult to manage, as economic migrants may be welcomed while other, potentially more vulnerable, groups are resisted. The development of innovative means of measuring, modeling, and understanding these migration patterns and processes represents a challenge for future geographical understanding of this important topic.

See also: Demography; Emigration; Immigration II.

Further Reading

Boyle, P. J., Halfacree, K. H. and Robinson, V. (1998). *Exploring Contemporary Migration*. London: Longman.

Castles, S. and Miller, M. J. (2003). *The Age of Migration: International Population Movements in the Modern World*. New York: Guilford Press.
Clarke, W. A. V. (1998). *The California Cauldron: Immigration and the Fortunes of Local Communities*. New York: Guilford Press.
Harris, J. R. and Todaro, M. P. (1970). Migration, unemployment and development: A two sector analysis. *The American Economic Review* 60, 126–142.
Khoser, K. (2007). *International Migration: A Very Short Introduction*. Oxford: Oxford University Press.
Lee, E. S. (1966). A theory of migration. *Demography* 3, 47–57.
Massey, D. S. and Kouaouci, A. (1999). *Worlds in Motion: Understanding International Migration at the End of the Millennium*. Oxford: Oxford University Press.
Portes, A. and Rumbaut, R. G. (2006). *Immigrant America*. Berkley, CA: University of California Press.
Ravenstein, E. G. (1885). The laws of migration. *Journal of the Statistical Society of London* 48, 167–235.
Sjaastad, L. A. (1962). The costs and returns of human migration. *The Journal of Political Economy* 70, 80–93.
Vertovec, S. and Cohen, R. (1999). *Migration, Diasporas and Transnationalism*. Aldershot: Edward Elgar.
Zelinsky, W. (1971). The hypothesis of the mobility transition. *Geographical Review* 61, 219–249.

Relevant Websites

http://www.forcedmigration.org
Forced Migration Online.
http://www.migrationwatchuk.org
Migration Watch.
https://genographic.nationalgeographic.com/genographic/index.html
National Geographic Genographic Project.
http://www.refugeecouncil.org.uk
Refugee Council.
http://www.unhcr.org
UNHCR.

Migration, Historical Geographies of

W. Jenkins, York University, Toronto, ON, Canada

Glossary

Assimilation A process whereby a minority group gradually adopts the attitudes, customs, and identity of the prevailing majority culture, thereby shedding its own distinctive cultural practices and identity.

Cultural Landscape The visual result of human interaction with the natural world as represented, for example, in the farms, fields, houses, villages, and cities that dot the Earth.

Diaspora The dispersion of a people from their original homeland and the common ties and identities that such a people form across their various sites of settlement.

Indentured Labor Where 'unfree' laborers are placed under contract to work for another person for a specified period of time, often without pay, and in exchange for accommodation and/or free passage to a new country.

Moral Panic A sudden increase in public perception of the possible threat to societal interests and values due to the activities of certain sections of society, such as youth and/or immigrants, as reported through the media (e.g., newspapers and television).

Pogrom An organized attack, often a massacre, against a minority group and often carried out with the tacit approval of official authorities.

Racialization The process whereby racial categories are created and assigned to a previously unclassified relationship, social practice, or group; its effect is to provide 'race' with a clear social meaning in a way that divides societies.

Remittance Money that is sent from one place or person to another, often from a family member abroad to family or relatives back home.

Introduction

Migration, as studied within geography and other social sciences, generally refers to the permanent and semi permanent movements of individuals and groups across space. Such moves may range from the short distances involved with changing addresses within a city, town, street, or even building, to more long distance relocations across states and continents. Volume wise, movements also occur at both individual and group based levels. The reasons for such migrations are clearly many and do not always have a planned or predictable quality. Because of its often profound impacts on the welfare of peoples and the reshaping of places around the globe in the past as well as the present, migration continues to be a key topic of interest not just for geographers but also for anthro pologists, economists, historians, psychologists, and soci ologists. For geographers at least, and those who pursue historical geography in particular, long standing ap proaches to the study of migration have ranged from quantifying the numbers of migrants moving within and between various nations and regions to evaluating the social and cultural impacts of those migrants on ordinary landscapes, rural and urban, in more local contexts. In recent years, however, topics such as the representation of migrants and their placement within wider discourses of, for example, race and nationalism have surfaced within the geographical literature.

Conceiving Migration

The forces and circumstances that have motivated people to move around the globe throughout history are many and varied. Analysts of migration, such as economic historians and geographers, traditionally adopted models distinguishing between the origin and destination regions of migrants. These models typically involved to varying degrees of sophistication the quantitative comparison of an array of economic, social, and political variables within each region (e.g., wage rates, underemployment levels, land prices, and population densities) that were held to inform peoples' migration decisions. Within this framework, a series of 'push' factors in places of origin were contrasted with 'pull' forces attracting migrants to destination societies. The roots of such approaches lie in nineteenth century neoclassical economics as well as the work of scholars such as E.G. Ravenstein, who sought to formulate a series of 'migration laws'. Among the most noteworthy of Ravenstein's 'laws' were his claims that the major cause of migration was economic, that most migration occurred over short distances, and that mi gration proceeded mostly from rural to urban areas. While it is true that many migrants, with an eye to bettering their economic situations, compared the op portunities existing at home with those offered else where, the factors affecting migrations were frequently more complex. At a broad level, the contexts behind movements can be differentiated between 'voluntary' and 'involuntary' where the enterprising moves of 'adventurer traders', for example, contrast clearly with the forced evacuations brought on by events such as environmental disasters, civil strife, and the raiding and trading of slaves.

Although the push/pull framework draws useful conceptual attention to the conditions behind peoples' decisions to move, it oversimplifies migration as a straight forward movement from A to B while neglecting its human and experiential dimensions. Migrants have proved to be more restless and transient beings than previously assumed. Micro scale research has revealed migration to be more of a stage process with stepwise, circular, and return patterns of movement. Fixed descriptions of mi grant groups as 'sojourners' or 'settlers' are therefore of limited analytical value; migrants can oscillate between such categories at several points in their lives. Traditional approaches to migration have also neglected the specifi cities of political–geographical contexts, such as the his torical role of states and empires in channeling and regulating the flows of subjects, citizens, exiles, and guests. As geopolitical entities, they could not only expel sections of their population but also control the mobility of others resident within their borders or between their colonies.

Broad conceptions of the historical migration process can thus be moved forward in two ways. First, the context that leads to migration can be considered as a continuum whereby those possessed with the freedom to make their own independent decisions to move are located at one extreme, and those without such freedoms at the other. Second, the complex geographies of migration, occurring at varying intensities and over different lengths of time, can be viewed within the framework of a 'migration system' that describes a cluster of moves from a well defined region of origin to one or more receiving regions over a given period. As political, economic, or environ mental conditions change in any of these societies, the dynamics of movement within the migration system be come transformed and redirected, or the system ceases altogether.

Migration systems were sustained in various ways. For one thing, opportunities needed to exist in destinations, whether to spur the movement of forced labor or to entice 'free settlers'. Among the latter especially, personal correspondence relayed migrants' informal commentaries about living conditions and the likelihood of success to those they left behind. The connections these engendered between kin, relations, friends, and communities in origin and destination localities that spurred further movements to the latter have been termed 'chain migration'. Mailed tickets for passage and financial remittances were crucial to this process. At another level, policy and propaganda by private shipping and railroad companies and govern ment ministries played crucial roles in maintaining mi gration flows as long as opportunities remained in destination societies (**Figure 1**). The images presented within such documents of rich and cultivated landscapes bristling with promise and profits were often exagger ations of reality with descriptions of haphazard weather patterns kept to a minimum. Whether it was a personal letter or a government issued pamphlet, potential mi grants were learning about, and forming imaginations of, places and landscapes elsewhere. For those escaping op pressive political regimes, slippages of detail were not as important as being free itself.

Historians and historical geographers have typically viewed the Atlantic Ocean as a key arena in which several migration systems of varying magnitude became elaborated between Europe, Africa, and the Americas from the seventeenth century onward. The most in famous of these is undoubtedly the forced movement of slaves from West Africa from the seventeenth to the early nineteenth centuries. Over the same time period and beyond it, Europeans moved in dramatic numbers to the Americas as well. In the case of the United States, migrants responded to the powerful place centered

Figure 1 Poster: Canada 160 Acres fritt land till hvarje Nybyggare, *ca.* 1905 06 (Norwegian version of Canadian immigration poster offering 160 acres of free land on the prairies). Source: Library and Archives Canada/Department of Immigration fonds/C-132141.

myth that was (and is) ‘the American Dream’ with its ‘promise of freedom’ from the economic, political, and religious oppressions they experienced in their home lands. In the nineteenth century North Atlantic, for example, the flight of Irish peasants from the mid century potato famine was followed in later decades by the labor migrations of Italians and the movement of Jewish families from their ‘pale of settlement’ within the Russian Empire as a result of continued persecution and pogroms.

This scholarly focus on the Atlantic is gradually becoming balanced by studies dealing with other parts of the world. The Pacific Rim, for example, also witnessed similar displacements of people during the last two centuries. Under constant threat of drought and civil strife in the southern provinces of China in the second half of the nineteenth century, for example, those who could flee spread themselves out to other parts of Asia as well as to the west coasts of the Americas from Vancouver to Santiago. Other studies have been insightful regarding the gender composition of migrant flows. While many of the aforementioned Chinese were male, Filipino females settled themselves in North America, Western Europe, and the oil rich Gulf States in the second half of the twentieth century. In many ways encouraged by their own government, these labor migrants have become concentrated in the areas of domestic service and healthcare in their destination societies.

At a global scale, the mechanics of empire building fostered intercontinental migration systems, principally to satisfy settlement and labor demands across various sites of colonization. With Britain’s abolition of the slave trade in 1807 and the ending of practiced slavery in its colonial possessions in 1833, the initiation of new migration systems of indentured labor from southern Asia were crucial to maintaining labor forces on plantations of sugar and other commodities in the Caribbean, southern Africa, and Fiji, intensifying the multicultural composition of these societies. It is not surprising that some historians have referred to this movement as a ‘second slavery’. Other South Asian migrants who were not indentured were also recruited to build railroads and develop mercantile activity in eastern Africa during the nineteenth and early twentieth centuries (**Figure 2**). These examples illustrate well the ‘political economy’ interpretation of migration that views flows of migrant labor as integral to global capitalism’s unceasing search for new markets while minimizing labor costs. Migrants were thus located at the center of a network of power that drew the interests of national and imperial governments together with those of private capital, the latter comprising variously shipping and railroad interests, agricultural producers, resource extractors, or urban captains of industry.

Mapping, Measuring, and Representing Migration

Historical geographers and other social scientists have used a variety of sources to understand not only the large scale effects of migration in different parts of the world through time, but also the lived experiences of migrants. National census returns, cartographic surveys, and the papers and reports of state parliamentary committees, in both manuscript and published form, have facilitated studies of internal migration, emigration, and immigration in urban and rural contexts over the past two centuries and more. Surviving passenger lists from steamship companies, as well as a variety of ecclesiastical records, have often shed important light upon the regional origins of migrants and their context of departure.

At a more qualitative level, newspaper accounts offer national and local perspectives on, and reactions to, migration in the context of receiving, or ‘host’ societies. They contain evidence on how migrants were represented and perceived by host societies not only in the text of editorials but also in images such as political cartoons. Within the evolving ‘media culture’ of the nineteenth century, such cartoons often depicted migrants in an unflattering light which served to underline their essential differences from their hosts and cast doubt on their abilities to adapt to their new cultural contexts. In some instances, newspaper reports of migrants as disease carriers sparked moral panics and street demonstrations against ‘new foreigners’, with the widespread use of terms such as the latter serving to underline who belonged and who did not.

Letters, diaries, autobiographies, and historical novels offer other important qualitative insights on the private lives of migrants and their journeys. For studies of recent migrants, oral history has been invaluable in mapping and contextualizing the places they experienced and the challenges they faced within each. Scholars have also used the cultural landscape itself as a field based archive to study the ‘imprints’ of migrants in various places, with house architecture, graveyards, and place names providing evidence of sociodemographic patterns as well as tastes and traditions among migrants. Such exercises in ‘landscape reading’, while not without their shortcomings, can nevertheless serve as an important supplement to what is gained through indoor research in archives and libraries.

The Consequences of Migration

While migration has often been discussed in relation to the creation of new states and societies, this frequently occurred at the expense of preexisting ones. Scholars have usually illustrated this point by considering the

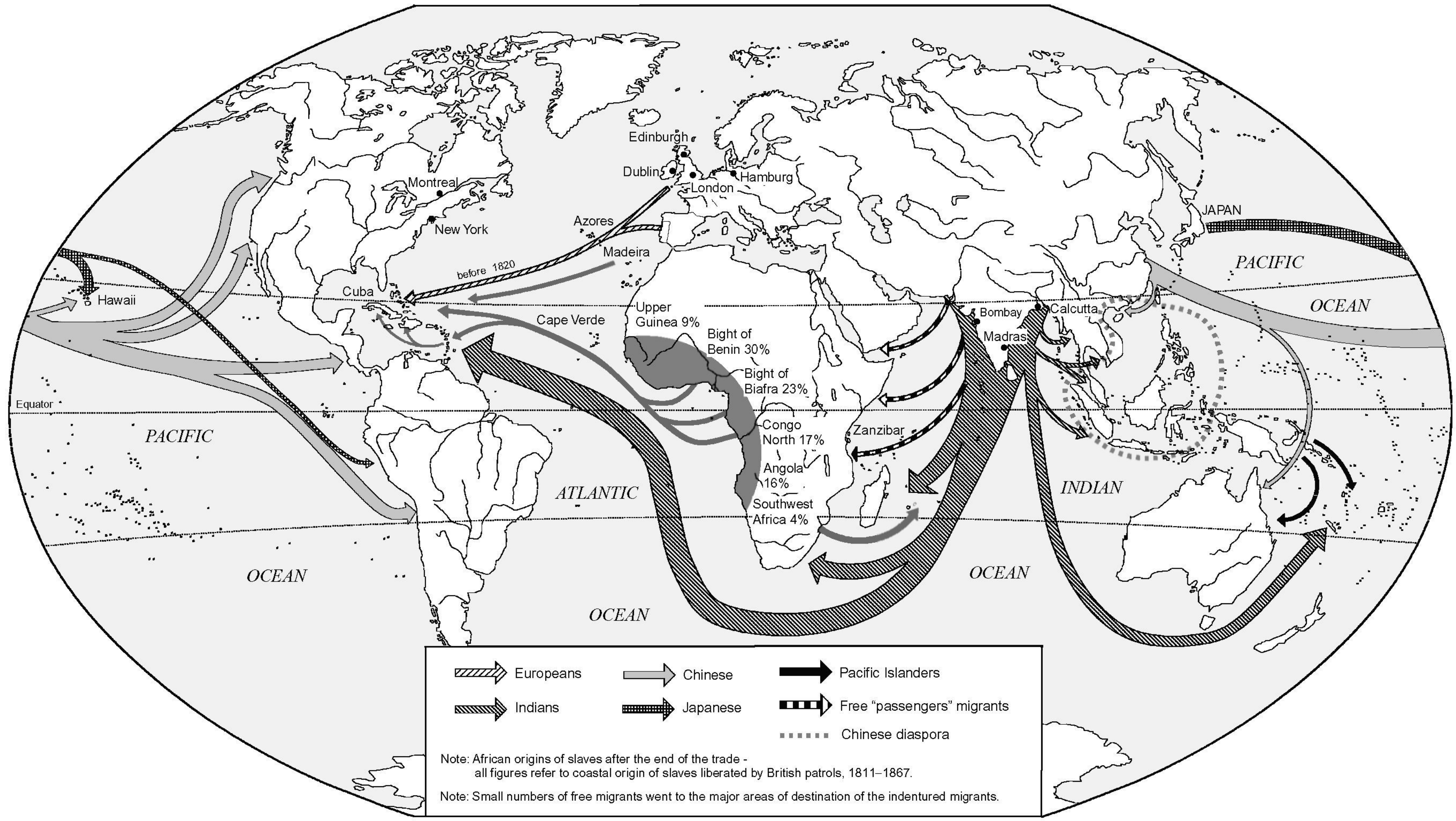

Figure 2 Principal overseas migrations, 1830s–1919, of Asian indentured and free migrants. From Hoerder, D. (2002). *Cultures in Contact: World Migrations in the Second Millennium.* Durham: Duke University Press. Hoerder's map is adapted from Northrup, D. (1995). *Indentured Labor in the Age of Imperialism, 1834–1922.* Cambridge: Cambridge University Press. Maps 1 and 5 from de Gèrard, C., Jan, M. and Rageau, J.-P. (1994). *Atlas Historique des Migrations.* Paris: Le Seuil.

encounter between European migrants and indigenous peoples in various parts of the world where land was not only expropriated by the migrants on their own terms but the pathogens they carried had deadly effects on local populations. This 'disease impact' of conquerors and settlers was especially apparent in the Spanish and Portuguese colonies of Latin America, where pandemics of smallpox, measles, and typhus devastated preexisting populations in the sixteenth and seventeenth centuries. Many indigenous survivors would be on the move over succeeding centuries, marshaled into labor forces for the mining of mercury and silver in the highlands of Bolivia and Peru, for example. At later times, the successful introduction of 'new settlers' prompted the development of techniques designed to gain knowledge about, and subsequently regulate the movements of those dispossessed from their lands. Here, the role of the American government in the aftermath of the nineteenth century displacement of Native American tribes from the Great Plains is instructive. The system of territorial reservations that resulted enabled the monitoring of movements of what was regarded as a potentially troublesome population while aiming to assimilate them to the ways of 'civilization'. The ability to move could be viewed as an act of subversion as much as an act of freedom.

The capacity of migrants to dominate the political contexts of their new settings was not always so assured, however. In these instances, migrants were recognized as minorities and their patterns of social incorporation and acceptance took many forms. The retention of separate cultural identities by migrants was often as much an imposed creation of host societies as it was of the migrants themselves, since the former were usually quick to form judgments about the capacity of the latter to fit in or 'assimilate'. Cities, especially port cities, were places where migrant visibility was heightened and issues of 'migrant difference' were played out in the search for work and shelter. Prevailing ideas about racial difference also informed the reaction to different groups of migrants. While the 'Chinatowns' of San Francisco and Vancouver emerged as minority migrant enclaves in atmospheres of hostility and prejudice in the nineteenth and early twentieth centuries, that of Singapore emerged in a situation where the Chinese became a demographic majority within a British colony. The creation of such urban diversities from migration was nothing new, however. Amsterdam and Vienna were profoundly affected by surges of immigration in earlier centuries, while the Venetian 'ghetto' of the sixteenth century represented an early attempt to deliberately segregate an outcast cultural group (Jews) within the boundaries of the city.

In other places and at other times, the contribution of migrants to the shaping of landscapes was far less recognized, especially if those migrants were not permanent settlers. Depending on the context, their participation in labor markets brought them into contact with urban employers, rural landowners, native born workers, and other migrants of different origins. Migrants worked rural lands not just as pioneer farmers but also as low paid laborers, whether on the tea plantations of Darjeeling or the citrus groves of central California. The creation of popular idyllic images of landscapes such as California often blinded viewing audiences to the often bitter struggles between capital and labor that underlay their production and reproduction through time. In California, the East Asian and Mexican laboring men who had to tolerate the deplorable squalor of housing camps were (and remain) hidden in precisely this way. Viewed and treated as 'expendable foreigners', credit for the landscape's beauty was accorded to the landowners and not the workers. The latter situation was similar for migrants from southern Africa converging on the diamond and gold mines of the South African Rand from the late nineteenth century onward. Again, these migrant activities illustrate their role in underpinning capitalist societies whether through the cultivation of land, the exploitation of natural resources, or the development of urban industries. Despite the intermittent hostility of regional and state level legislation, employers could not afford to do without such cheap sources of labor though it is important to remember that these 'foreigners', far from always being docile servants of capitalism, were often also important transmitters of radical ideas who engaged in strikes and other struggles to improve their conditions of work as much as they were important agents in the transformation of material landscapes.

Migration has therefore, brought about a plethora of cultural interactions and intermixing in many places, with obvious consequences for territorial identities, and the idea of the 'foreigner' connects to the place of migrants within the development of national political systems and identities over the past two centuries. With more pronounced and popular ideas of national identity, citizenship, culture, and sovereignty circulating through the period, issues related to the comings and goings of people within the 'national territory' received increased attention from state governments and institutions. Immigration entered the policy arena and, as already noted, became directly linked to economic development and nation building, especially in parts of the so called 'New World' (e.g., Argentina, Australia, Brazil, Canada, New Zealand, and the United States). What nation builders saw as the 'blank spaces' of cultivable land in these countries was aggressively promoted in order to stimulate migrant imaginations, particularly in Europe. In the United States, immigration and the promised freedom of internal movement served an ideological purpose in glorifying the libertarian and democratic values at the heart of the nation's constitution. But the emphasis on

Figure 3 Train carrying refugees in the aftermath of the partition of India and Pakistan in August 1947.

attracting Europeans to America betrayed the essential racialization behind this ideology, illustrated with the banning of Chinese immigrants in 1882 and the afore mentioned confinement of Native peoples to remote reservations. There and in Canada a hierarchy of immi grant preference, based on prevailing notions of race, prevailed until the 1960s in order to preserve fixed no tions of a white, English speaking, and Christian North American continent.

In other parts of the globe, emerging ideas about nationhood also became formulated with the ideal of national 'cultural congruence' in mind. An imperative existed, in other words, to align the geography of the 'national people' with the boundaries of the state in such a way that would produce such congruence. Through the narration and popularization of ideas about who counted among the national 'imagined community', displace ments and expulsions were often resorted to in order to ensure the removal of the excluded 'them'; 'purified geographies' of the nation could thus become realized. Such quests for national congruence have led to the movements of populations and selective immigration and citizenship policies in many countries over the past century and more. As is well known, the aftermath of World Wars I and II witnessed the redrawing of the political geography of Eastern and Central Europe along these idealized 'nation state' lines and, in the case of the Balkan states, led to later ethnoreligious conflict. World War II resulted in massive population transfers within the same region while the decimated Jewish population looked to the new state of Israel as the place to forge new lives. Post war decolonization within the British Empire also resulted in the redrawing of borders and movement of peoples. The mass transfers of some 5 million Hindus, Muslims, and Sikhs across the newly created border between India and Pakistan between August and No vember 1947 exemplifies how the imperative to equate nation and residence displaced many (**Figure 3**). This episode also exposed the gruesome effects of inter communal violence when people traveling by train and road were exposed to attack by the opposite community; an estimated 200 000 adult and child migrants were killed during these dramatic months. The now familiar idea of the migrant as a refugee took shape as a result of episodes like this.

Migrants, Identity and the 'Homeland'

Migration was not just about arrival, it was also about departure. Migrants' affections for, and relationships with, the lands they left were varied. Some looked back more than others, forever feeling a melancholic sense of exile in their destination societies; others, thankful for getting out when they did, were far less sentimental. Not all intended to leave their 'old countries' behind forever either and by the twentieth century, many could afford to travel the long distances home to visit and, finances permitting, resettle permanently. One American gov ernment report from the early twentieth century cited southern Italians as the group with the highest rate of return. The loyalties of these return migrants remained evidently weighted in favor of their region or nation of origin. Such patterns of migration were not always wel comed by their American hosts who expected migrants to settle within and embrace the values of their new nation, the 'cradle of liberty'. The image of the opportunistic transnational labor migrant was born out of such debates, where his or her 'flexibility' was not necessarily seen as a good thing.

For others, temporary migration was part of a more clear cut career trajectory. The 'homeland' was rarely forgotten by those who forged their careers on its behalf in an imperial context as governors, civil servants, sol diers, and others. Empire building depended on the migration of these sorts of personnel from the metro politan center to the colonial periphery for the purposes of government, administration, law and economic de velopment, a process that became more formalized in the case of the British and French Empires in the late nineteenth and early twentieth centuries. Some formed part of ruling elites in more than one colony, for example, Thomas Pope Hennessy, who served as British governor in both Barbados and Hong Kong. In colonies with for midable summer climates such as India, British families were provided with periodic visits to the more familiar and 'suitable' climatic environment of their place of birth in the early twentieth century, although such visits often made clear that imagining and experiencing Britain as 'home' were often two different things. After years abroad, it was not uncommon for many of these Britons to feel a sense of displacement once the everyday reality of 'home' had to be negotiated. Only then did the emo tional relationships they had forged in India come into focus, leading to a somewhat hybridized sense of be longing. Spending extended periods in the colonies did not necessarily engender a prolonged sense of 'exile' for many of these movers, and while some 'went native', the majority still anticipated Britain as their final burial place.

The previous section discussed the challenges mi grants faced when encountering the assertion of nation alist identities in their places of settlement, especially when such identities revolved around fixed notions of culture and race. It is worth remembering, however, that such migrants did not remain passive to, and in some cases initiated, similar nationalistic impulses in their homelands of origin. Scholars in both the humanities and social sciences now use diaspora as a summary term to describe the often wide 'scattering' of people to places separate from what they recognize as their ancestral 'homeland'. Although this remains a rather under re searched area among geographers, they are nevertheless developing understandings of the multiple ways in which migrants and their descendants continued to imagine these homelands and the significance of these imagin ations. For some, such actions hardly extended beyond the symbolic cultural nostalgia felt during annual festi vals, while others assumed more active stances in the politics of their homelands. This was particularly the case where projects of nationalist 'liberation' remained un resolved. Among the key actors in late nineteenth cen tury Irish nationalist agitation, for example, were the urbanized Irish migrants in North America who provided money to support not only political campaigns at home but also more radical initiatives such as the purchase and use of weaponry. By the early twentieth century, the Irish and Poles in the United States had become well known as proactive campaigners for the national independence of their homelands, taking their passions into streets and meeting halls as well as in print. Migrant or 'ethnic' leaders such as newspaper editors, religious figureheads, intellectuals, and other members of the middle class were important in such instances. Interpretations of migration as enforced exile caused by imperial injustice and mis management were particularly effective in mobilizing support among rank and file migrants for diasporic na tionalist projects such as these, while the culture and 'imagined community' of the homeland was relentlessly translated through journalism, art, vernacular theatre, literature, poetry, music, personal letters, and often stirring occasions such as the street parades and lectures of visiting agitators.

Conclusion

While the above discussion has concentrated explicitly on the migration of people, it is clear that the movements of these same people facilitated the mobility and ex changes of other entities across geographical space such as texts, animals, foodstuffs, and plants which contributed to the circulation of different forms of knowledge as well as the transformation of landscapes and societies in dif ferent parts of the world. While there has not been suf ficient space here to cover these dimensions of migration, it would not be inaccurate to say that migration now forms one aspect of a wider interest among geographers on 'mobilities'. While migration continues to be an area of concern for historically minded geographers, it is be coming increasingly bound within wider inquiries about the ideologies and practices of capitalism, imperialism, and nationalism as well as the development and de ployment of socio political categorizations such as race, ethnicity, and citizenship. As far as understanding mi gration as a process in itself is concerned, traditional push/pull perspectives have given way to more sophis ticated understandings of migration systems and the consequences of movements on intercultural relations in different places. The coming decades should see such historical geographical studies of migration move into the post World War II era while continuing to embrace parts of the world beyond the Euro American Atlantic axis; they should also seek to deepen their understand ings of migrant diasporas in a global context character ized by intensified interconnection.

See also: Archives; Ghettos; Imperialistic Geographies; Migrant Workers; Multicultural City; Nationalism, Historical Geography of; Remittances.

Further Reading

Anderson, K. J. (1991). *Vancouver's Chinatown: Racial Discourse in Canada 1875 1980*. Montreal: McGill Queen's University Press.

Baily, S. L. (1999). *Immigrants in the Lands of Promise: Italians in Buenos Aires and New York City, 1870 1914*. Ithaca, NY: Cornell University Press.

Bodnar, J. (1985). *The Transplanted: A History of Immigrants in Urban America*. Bloomington, IN: Indiana University Press.

de Gèrard, C., Jan, M. and Rageau, J. P. (1994). *Atlas Historique des Migrations*. Paris: Le Seuil.

Gowans, G. (2003). Imperial geographies of home: Memsahibs and Miss Sahibs in India and Britain, 1915 1947. *Cultural Geographies* 10, 424 441.

Grigg, D. B. (1978). E.G. Ravenstein and the 'the laws of migration'. *Journal of Historical Geography* 3, 41 54.

Hannah, M. G. (2000). *Governmentality and the Mastery of Territory in Nineteenth Century America*. Cambridge: Cambridge University Press.

Hoerder, D. (2002). *Cultures in Contact: World Migration in the Second Millennium*. Durham: Duke University Press.

Houston, C. and Smyth, W. J. (1990). *Irish Emigration and Canadian Settlement: Patterns, Links and Letters*. Toronto: University of Toronto Press.

Jacobson, M. F. (1995). *Special Sorrows: The Diasporic Imagination of Irish, Polish, and Jewish Immigrants in the United States*. Cambridge, MA: Harvard University Press.

Knowles, V. (1992). *Strangers at Our Gates: Canadian Immigration and Immigration Policy, 1540 1990*. Oxford: Dundurn Press.

Lambert, D. and Lester, A. (eds.) (2006). *Colonial Lives across the British Empire: Imperial Careering in the Long Nineteenth Century*. Cambridge: Cambridge University Press.

Mitchell, D. (1996). *The Lie of the Land: Migrant Workers and the California Landscape*. Minneapolis, MN: University of Minnesota Press.

Northrup, D. (1995). *Indentured Labor in the Age of Imperialism, 1834 1922*. Cambridge: Cambridge University Press.

Nugent, W. (1992). *Crossings: The Great Transatlantic Migrations, 1870 1914*. Bloomington, IN: Indiana University Press.

Ostergren, R. C. (1988). *A Community Transplanted: The Trans Atlantic Experience of a Swedish Immigrant Settlement in the Upper Middle West 1835 1915*. Madison, WI: University of Wisconsin Press.

Robinson, D. (ed.) (1990). *Migration in Colonial Spanish America*. Cambridge: Cambridge University Press.

Robinson, V. (ed.) (1996). *Geography and Migration*. Cheltenham: Elgar.

Western, J. C. (1992). *A Passage to England: Barbadian Londoners Speak of Home*. Minneapolis, MN: University of Minnesota Press.

Withers, C. W. J. (1998). *Urban Highlanders: Highland Lowland Migration and Urban Gaelic Culture 1700 1900*. East Lothian: Tuckwell.

Relevant Websites

http://www.history.ac.uk
General resource website for scholarly work on world migration history through the Institute for Historical Research in London, Institute of Historical Research.

http://www.ucalgary.ca
Peopling North America: population movements and migrations, The Applied History Research Group, University of Calgary.

http://popdis.org
Population displacement in Eastern Europe and the USSR in the 20th century.

http://www.loc.gov
The African American mosaic: A research guide for the study of black history and culture, including migration during the era of slavery, Library of Congress.

http://www.let.leidenuniv.nl
The history of international migrations, especially those in Europe and the Western world, Universiteit Leiden.

Military and Geography

M. Farish, University of Toronto, Toronto, ON, Canada

Glossary

Imperial Archive A term used by Thomas Richards in his 1993 book of the same title to refer to "a fantasy of knowledge collected and united in the service of state and Empire," a fantasy seen as necessary in order to grapple with the predicament of controlling distant territories.

Militarism The extension of military influence into civilian spheres of life, most apparent when nation-states express expansionist or imperialist tendencies.

Military–Industrial–Media–Entertainment Network A term coined by James Der Derian, in 2001, to describe the expanding alliances between the now-traditional military–industrial complex and the media and entertainment sectors, particularly in the United States.

Modernity The condition of being modern, but also a loosely defined historical epoch characterized by a particular approach to history and geography, an approach that emerged in Europe in the sixteenth and seventeenth centuries and was globally dominant by the middle of the twentieth century.

Introduction

The French geographer Yves Lacoste famously wrote that "geography serves, first and foremost, to wage war." (Lacoste, 1976). He was referring to the fact that geography has always been essential to acts of organized violence. Military history is littered with examples of armies and planners mixing terrain and tactics to great or disastrous effect. But Lacoste was also suggesting something else: that despite its emergence as a valid intellectual field, geography was, then and now, still primarily a military idea and tool.

This claim might seem surprising to a student pursuing a degree in geography today. The fourth edition of the influential *Dictionary of Human Geography* (2000) includes no entry on military themes, and university departments, where the discipline has been given institutional credibility, rarely feature specialists who call themselves military geographers. Such individuals are much more likely to be found elsewhere, in armed forces academies, research institutes, and working for the US Department of Defense or its equivalent in other countries. Their sessions at major conferences, such as the Annual Meeting of the Association of American Geographers, are equally distinct, additionally attended by only a handful of curious onlookers or determined critics. Military geographers are rarely listed in the pantheon of personalities trotted out during courses on the intellectual trajectory of geography, although many distinguished geographers have had military ties.

But as a number of geographers, inspired by Lacoste and other perceptive predecessors, have begun to recently document, the history of geographical ideas is inextricably tied to military imperatives, and geography is much more than that practised in academic settings. The American National Geospatial Intelligence Agency (NGA), formerly the National Imagery and Mapping Agency, which has extensive ties with both the defense and intelligence communities, is reputedly the country's largest employer of cartographers. In the age of virtual reality recruitment tools, fluid news network simulations of combat zones, and the electronic proliferation of satellite photography through programs such as Google Earth, the world's conflicts (at least those given prominence in the media), from Afghanistan to Iraq and beyond, have produced a dizzying catalog of geographical images and ideas associated with the practice of war. Militaries, in other words, continue to be central to the production and popularization of geographical knowledge – not only to formal cartographic representations, but to more elusive spatial principles, such as the perception that the world is a composite of hostile environments.

Toward a History of Modern Military Geography

Despite the substantial, ongoing production of military geographies, and despite the recent re examination of geography's history from a more critical perspective, it is fair to suggest that geographers have yet to adequately account for the abiding associations between their field and the contemplation and conduct of war. While self described military geographers are unlikely to publish in many of the discipline's leading journals, this has not always held true. Perhaps, this shift is due to the changed nature of conflict, which at least as practised by the world's more advanced armed forces does not seem to reach as deeply into popular consciousness as it once did – a distance that may well be deliberate. Or perhaps it is due to a simultaneous expansion and fragmentation of geographical inquiry, heralding exciting and welcome debates while paradoxically also encouraging compartmentalization into like minded collectives.

Whatever the case, it is certainly instructive to consider the historical trajectory of military geography, a term that does not encompass all that is at stake here, but certainly provides a welcome focus.

The Geographic Arts of War

Military geographers writing on the history of their field typically gather steam with the emergence of modern warfare, presumably traceable to the formation of a modern epoch in the fifteenth and sixteenth centuries. But then many military geographers are less concerned with the details of historical cases than they are with timeless principles of war, both philosophical and environmental. If there is one constant in the history of military geography, it is a forceful separation of culture from nature, resulting in a regional division of the world into generalized natural categories. Far more malleable than terrain and climate, in this model, is technological development, or the use of technology to minimize or even overcome environmental constraints. (Culture, however, is never entirely absent from such views, particularly but not exclusively when military geographers turn to urban landscapes.) The globe, in this worldview, becomes a finite space, and while knowledge does not necessarily guarantee victories, it is certainly the case that military geography takes knowledge as a key to power, and that geography is first and foremost a descriptive science, whose analytical properties are understood only in the most pragmatic terms.

‘Geography Militant’

A full survey of military geography would reach far back into the histories of armed conflict and empires, and to geographers such as the Greek scholar Strabo. But military geography as it is studied and taught today is a product of the eighteenth and nineteenth centuries, the period Joseph Conrad famously referred to with the descriptor ‘Geography Militant’.

In Britain, for example, the first intelligence agency, the straightforwardly named Depot of Military Knowledge, was created during the Napoleonic Wars to accumulate geographic information, including maps, on the military power and terrain of foreign countries. Although this activity did not continue consistently throughout the nineteenth century, it certainly prefaced more ambitious twentieth century attempts at gathering geographic intelligence, and also reflected the creation of an imperial archive necessary for the management of Britain’s global possessions. By the late nineteenth century, the Intelligence Branch, a new government department, was collecting thousands of books and maps each year, and participating in the famous Anglo Russian ‘Great Game’ in Central Asia. But in a matter of decades Britain’s empire was in decline, and the geographical excitement that gripped the country at the apex of its power – excitement fueled by exploratory initiatives and romanticized campaigns – was waning. While this spelled the decline of several smaller geographical societies, the geopolitical rivalries between several European powers ultimately resulted in a much more direct role for military geography.

The organization of the Royal Geography Society (RGS) in 1830 and its nineteenth century activities were inextricable from the colonial operations of the British state, such that the Society became a useful guide for imperial expansion. And the ‘New Geography’ identified with the establishment of geographical education at Oxford and Cambridge Universities in the late nineteenth century was also found to be practically useful to a ‘New Army’ in the years leading up to World War I. But a formal field of military geography was much more prominent in the US, where geographical study could be found at the West Point Military Academy shortly after its establishment in 1802.

The World Wars

Innumerable examples could be used to demonstrate the close relationship between the history of military geography and the history of cartography, but one of the most suggestive must surely be the role of maps in World War I, a subject that has been ably documented by the historical geographer Michael Heffernan. By the beginning of the twentieth century, one of the largest map collections in the world was housed at the London headquarters of the RGS. On the eve of the German invasion of Belgium and France in 1914, the Society opened this collection to the British War Office, particularly the Geographical Section of the Office’s General Staff, which essentially merged with the Ordnance Survey during the war years. This combined operation supervised the production of some 32 million map sheets (the equivalent of 21 000 per day), mostly large scale trench maps. Meanwhile, the RGS became the site of frequent lectures on the war, some critical of British leaders for insufficient use of the country’s geographic experts. Many of these lectures employed startling maps that documented the potential spread of German power over the globe – maps that invariably found their way to the popular press.

In another wartime center of calculation, the *Société de Géographie de Paris*, while much smaller than the RGS, was also providing its map collection and related resources to the French Army. Many leading French geographers were recruited to serve on a new commission that would produce maps and reports on various European regions for the French General Staff. Monthly public meetings of the *Société* were devoted to the war, and German geographers were harshly criticized for their apparent devotion to imperialist causes. French geographers went on to produce a highly influential series of reports on French colonial policy and the future political geography of Europe, focused on the cartographic expression of national sovereignty. Paris was also the scene of the 1919 Peace Conference, an event featuring the work of The Inquiry – a vast effort to create a rational and scientific archive of geographic information at the American Geographical Society in New York in the late stages of the war.

The global scope of World War II prompted tremendous interest in an exhausting roster of places, a quest for geographic knowledge that frequently familiarized unfamiliar landscapes in the language of military interests. This meant that the intelligence gathering efforts of World War I were expanded in scope and direction, and that for the first time, the systematic study of regions with an eye to their strategic value emerged as a subject of scholarly significance. In the US, reports were published suggesting that social scientists with knowledge of the globe's areas were almost as significant, and as unusual, as officers familiar with combat in the same locations.

This apparent lack of cosmopolitan experience, a condition invoked repeatedly by worried pundits, drove attempts to accumulate geographic information and teams that could usefully analyze this data, components that were assembled in important wartime clearing houses, such as the Office of Strategic Services (OSS), the forerunner of the Central Intelligence Agency, and the Smithsonian Institution's Ethnogeographic Board. While neither was an explicitly military agency, each provided significant assistance to arms of the American military. The Ethnogeographic Board, for instance, produced a booklet titled *Survival on Land and Sea*, in 1943, which was distributed in the hundreds of thousands to the armed forces. Condensed enough to be carried easily by a soldier on foreign duty, the book contained "the main things that a man should know about living in wild countries." This invocation of untamed places suggests a preoccupation with hostile environments, both natural and cultural, and their control and defeat through the application of knowledge compiled by experts and then passed on to troops in the field. This preoccupation characterized American military activity throughout the twentieth century, particularly from World War II onward.

Although it clearly echoed earlier forms of travel literature, including works composed in the service or the shadow of empire, *Survival on Land and Sea* was just one example of a burgeoning military literature on endurance issued during a war of unprecedented destruction, mobility, and scope. These novelties were captured in a *National Geographic* article, titled 'Fit to fight anywhere', also published in 1943. It documented the mental and physiological challenges of American soldiers warring "from hot tropic swamps to cold Aleutian fogs," but also various attempts to recreate such "strange conditions" in climate and fatigue laboratories. And while nontemperate climates were being simulated on domestic soil, so called 'combat scientists' were descending on the spheres of war, particularly the Pacific, where members of the Office of Field Service tackled problems of transportation, communications, climate, and insects. Troops with time to spare could aid this military science, particularly if they possessed a copy of the Smithsonian's *Field Collector's Manual in Natural History* (1944), by gathering specimens and shipping them back to the US for accumulation and study.

Military geography, therefore, had become much more than the narrow consideration of tactics and terrain. Dozens of geographers with little direct interest in military topics were called to Washington and other outposts of the American war effort. They worked, in particular, at the OSS, where the stirrings of geography's 'quantitative revolution' – still over a decade away from sweeping through universities in the US and elsewhere – were evident in both the interdisciplinary connections needed to address wartime intelligence and the occasional marginality of geographic perspectives in the face of challenging social scientific problems. In Britain, substantial contributions were made by numerous geographers to the colossal *Naval Intelligence Handbooks*, published in 58 volumes between 1941 and 1946, and regarded as the single, most substantial geographical endeavor ever attempted. In these and other cases, wartime service acted as a springboard into university careers for young scholars, and solidified the reputation of others. Richard Hartshorne, who held a senior position with the OSS, lectured at the National War College in the 1950s, and many other prominent American geographers received small slices of the enormous military funding available to scientific and social scientific researchers during the early Cold War. In sum, World War II, wrote Kirk Stone, a veteran of the OSS and Naval Reserve, "was the best thing that has happened to geography since the birth of Strabo."

The Cold War and Beyond

In his Military Geography contribution to the influential volume *American Geography: Inventory and Prospect*, in 1954, Joseph Russell repeated the familiar argument that at the dawn of World War II, American knowledge of the globe's strategic regions was inadequate. After documenting the extensive role of geographers in the war effort, Russell made the case for military geography as not only useful to the conduct of the Cold War, but also as a field that could, through its attention to particularly 'extreme' environments, provide the combination of practicality and integration seemingly required of the social sciences.

This was an ambitious claim, and in fact the trajectory of military geography diverged from that of the discipline in subsequent decades. Academic geographers continued to advise and conduct research for national militaries, and geography's quantitative revolution of the 1950s and 1960s owed much, particularly in the US, to military funding and the work of scientists and social scientists with ties to the seemingly omnipresent military–industrial complex. Advances in the technical practice of cartography, remote sensing, and related geographic sciences were directly linked to military initiatives in geodesy, space science, and other fields. But such deep connections were rarely acknowledged, and, at least until the 1970s, challenged even less frequently. Thus, as military geography was gaining legitimacy as an applied subject that could contribute variously to Cold War tactical doctrine, scientific intelligence gathering, or, more tangentially, the burgeoning field of area studies, the depoliticization of the discipline was creating a 'spatial science' that seemed, on the surface, distant from explicitly military concerns. (As historians of the quantitative revolution have documented, the geography of this period was distant from virtually all worldly concerns.)

The inevitable critique of abstraction and positivism emerging from social and cultural upheavals of the 1960s and 1970s was directed less at geography's Cold War affiliations than at its general lack of political purchase, although certain individuals, such as Yves Lacoste, specifically identified the dangers of military geography. For all the philosophical engagement with theories of social justice or humanism, very little of the emergent critical geography of the 1970s was concerned – as scholars in many other fields were – with the Vietnam War. While this is a surprising absence, it is crucial to recall that the ties between geography and militarism were hardly limited to self described professional geographers. As the American armed forces and those organizations resisting them offered competing versions of 'militant tropicality' in Southeast Asia, scholars and activists were challenging the legitimacy of classic Cold War ideas, such as modernization theory, which at its worst offered social scientific justifications for military interventions in the name of 'stability' and the endless fight against a voracious international communism.

In a more direct sense, military support and sponsorship was responsible for a succession of new geographic ways of seeing that continue to be closely affiliated with defense departments worldwide. The suite of systems and technologies comprising what is now called GIScience – including geodesy, photogrammetry and remote sensing, and cartography – emerged out of a series of overlapping classified initiatives in the Cold War US. With the mobilization of cartography and geography during World War II, spatial knowledge and the means of its production were brought much more directly under state control.

Contemporary Military Geographies

Despite acknowledging drastic differences in technology, war planning, and the range of duties given soldiers, studies in military geography today remain concerned, as one classic 1966 textbook put it, with "how geography may be used in the conduct of military affairs," and with "how geographic conditions may influence or constrain the exercise of military power." By these standards, military geography is a mere tool, able only to question the proper exercise of power – still understood overwhelmingly as national power – but not the assumptions built into the idea of power. The assertions that military geography is a type of applied study or that it is the common sense consideration of facts do not excuse its narrowness. Even while acknowledging the changed circumstances and scales of war, it has clearly failed to keep pace with theoretical and methodological changes within the wider world of geographical scholarship, perhaps with the exception of most GIScience.

Some leading American military geographers are aware of – and even embrace – this stasis, attributing the isolation of military geography to the radicalization of academic inquiry during the 1960s and 1970s. Regardless of this claim's veracity, it is still intriguing that the impact of intellectual transformations on military geography was so minimal. This would be unimportant were it not for military geography's ongoing, intimate affiliation with organized violence. Its proponents and productions, moreover, perpetuate the proximity of defense industries to universities. For this precise reason, it is a mistake to homogenize military geography, particularly as it is taught; there are regional differences, to be sure, but more importantly there is always some room, no matter what pedagogical approach is adopted, for debate and dissent within a teaching environment. This is worth recalling in any consideration of the perceived gap between military geographers and their critics.

Some of these critics have recently begun to examine the geographies of armed conflict – the spatial and territorial dimensions of violence – from quite a different direction. One important strand of this literature addresses the contemporary strategic preoccupation with cities, including the recycling of long standing Orientalist tropes to frame the urban spaces of the Arab world. Complex social landscapes, such as Baghdad, have been turned into empty military targets through media representations of precision bombing campaigns, leaving cities as little more than nests of terrorists or destinations for destructive weaponry. The devastating impact of these weapons on the everyday geography of Baghdad and other flashpoints of the War on Terror has gone underreported in many major media outlets, and a prominent television station that did feature street level images of carnage, the Qatar based Al Jazeera, had its Baghdad offices bombed. Meanwhile, graphically sophisticated video games produced by the American military allow users to immerse themselves in replicas of these same environments, where they are charged with occupying and pacifying such stereotypical Arab cities. All of this suggests that to speak of military geography in the twenty first century is really to grapple with the seemingly infinite varieties of what James Der Derian has called the Military–Industrial–Media–Entertainment Network.

While hardly novel, the urbanization of war has also recently prompted broader investigations into the idea of security and the consequences of conflict for civilians, including particularly threatened or powerless groups. This suggests, again, that what might fall under the category of military geography is much more than terrain and tactics – and even the consequences of conflict. Themes such as military land use, the political economy of militarism, military citizenship, the many troubling associations between militarization and gender, and cultural representations of militarism are all crucial to understanding the tremendous imprint of military geographies beyond armed conflict. Military agencies continue to draw a substantial number of trained geographers, and to produce some of the most powerful geographies in circulation, but the extent of militarism is such that it influences places and lives seemingly distant from battlefields or bases.

Although military geography is undoubtedly practised, and military geographers employed, by countless armed forces – whether national or not – this ubiquity must be measured against profound imbalances in historical and contemporary militarization. In other words, a country such as the US, which accounts for an astonishing half of the world's total military spending, is undoubtedly able to employ more military geographers (especially when understood broadly), and to practice more technologically sophisticated, though not necessarily successful, forms of geographical analysis. More research is certainly needed on the role of military geography in China, home to the world's largest standing army, or in North Korea, where one quarter of the population (an estimated 6 000 000 people) are active or reserve troops. These are impressive and disturbing figures, but so too is the estimate that the total defense spending of the US exceeded one trillion dollars in fiscal year 2007, and that outlays for the Pentagon alone (about half of that spending) were some 7 times that of China's entire defense budget. Such figures are invariably part of speculations by a number of prominent commentators on the costs and consequences, domestic and international, of American imperialism.

Militarism's Geographies

Unsurprisingly, the shocking acts of terrorism on 11 September 2001 and the subsequent launch of a global and seemingly perpetual War on Terror jolted geographers into additional scrutiny of the field's military affiliations. The comforting myth that countries such as the United States or United Kingdom were 'at peace' was definitively shattered. Responses have ranged from wide ranging assessments of geographical knowledge mobilized in the service of controversial and devastating military campaigns to debates over the ethics of scholarly publishing – debates which focused on this very *Encyclopedia.* Recent international conferences have featured numerous sessions devoted to the critical analysis of militarism, security, geopolitics, and related topics. Interdisciplinary connections are being forged with anthropologists, historians, sociologists, and others with similar interests and aims.

In the spirit of critical inquiry and intellectual engagement, these developments in geography and beyond are welcome. Yet it is unclear whether they signal any reduction in the gap between academic geography and those types of geography sponsored by Departments of Defense. More importantly, much work remains to be done on both the geography of military activities and the even more pervasive (and often more elusive) geographies of militarism, defined basically as the extension of military influence into civilian spheres of life – to the point that in many societies the distinction is impossible to discern. Further research on the specific spaces of military power, the banal geographies of militarism, and the moral geographies of violence is thus urgently necessary.

See also: Critical Geography; Critical Geopolitics; Geography, History of; Geopolitics; Maps and the State; Orientalism; Political Geography.

Further Reading

Barnes, T. J. and Farish, M. (2006). Between regions: Science, militarism, and American geography from World War to Cold War. *Annals of the Association of American Geographers* 96, 807 826.

Chatterton, P. and Featherstone, D. (2007). Intervention: Elsevier, critical geography and the arms trade. *Political Geography* 26, 3 7.

Cloud, J. (2002). American cartographic transformations during the Cold War. *Cartography and Geographic Information Science* 29, 261 282.

Clout, H. and Gosme, C. (2003). The naval intelligence handbooks: A monument in geographical writing. *Progress in Human Geography* 27, 153 173.

Collins, J. M. (1998). *Military Geography for Professionals and the Public*. Washington, DC: National Defense University Press.

Cowen, D. and Gilbert, E. (eds.) (2007). *War, Citizenship, Territory*. New York: Routledge.

Der Derian, J. (2001). *Virtuous War: Mapping the Military Industrial Media Entertainment Network*. Boulder, CO: Westview.

Driver, F. (2001). *Geography Militant: Cultures of Exploration and Empire*. Oxford, UK: Blackwell.

Enloe, C. (2000). *Maneuvers: The International Politics of Militarizing Women's Lives*. Berkeley, CA: University of California Press.

Farish, M. (2005). Archiving areas: The ethnogeographic board and the Second World War. *Annals of the Association of American Geographers* 95, 663 679.

Flint, C. (ed.) (2005). *The Geography of War and Peace: From Death Camps to Diplomats*. Oxford, UK: Oxford University Press.

Graham, S. (2008). *Cities under Siege: The New Military Urbanism*. New York: Routledge.

Graham, S. (ed.) (2004). *Cities, War and Terrorism: Towards an Urban Geopolitics*. Malden, MA: Blackwell.

Gregory, D. (2004). *The Colonial Present: Afghanistan, Palestine, Iraq*. Malden, MA: Blackwell.

Heffernan, M. (1996). Geography, cartography and military intelligence: The Royal Geographical Society and the First World War. *Transactions of the Institute of British Geographers* 21, 504 533.

Hewitt, K. (1994). 'When the great planes came and made ashes of our city…': Towards an oral geography of the disasters of war. *Antipode* 26, 1 34.

Lacoste, Y. (1976). *La Géographie, ça sert, d'abord, à faire la guerre*. Paris: Maspéro.

Palka, E. J. and Galgano, F. A. (2005). *Military Geography: From Peace to War*. New York: McGraw Hill.

Peltier, L. C. and Pearcy, G. E. (1966). *Military Geography*. Princeton, NJ: D. Van Nostrand.

Russell, J. A. (1954). Military geography. In James, P. E. & Jones, C. F. (eds.) *American Geography: Inventory and Prospect*, pp 484 495. Syracuse, NY: Syracuse University Press.

Smith, N. (1992). History and philosophy of geography: Real wars, theory wars. *Progress in Human Geography* 16, 257 271.

Stoddart, D. R. (1992). Geography and war: The 'new geography' and the 'new army' in England, 1899 1914. *Political Geography* 11, 87 99.

Stone, K. H. (1979). Geography's wartime service. *Annals of the Association of American Geographers* 69, 89 96.

Wallace, M. (2006). *Preliminary Bibliography of Recent Research on Militarization and Demilitarization*. Center for the Study of Militarization, Brown University. http://www.watsoninstitute.org/pub/Militarization Demilitarization Bibliography.pdf (accessed Aug. 2008).

Woodward, R. (2004). *Military Geographies*. Malden, MA: Blackwell.

Woodward, R. (2005). From military geography to militarism's geographies: Disciplinary engagements with the geographies of militarism and military activities. *Progress in Human Geography* 29, 718 740.

Relevant Websites

http://www.antiwar.com
Antiwar.

http://www.cdi.org
Center for Defense Information.

http://www.globalsecurity.org
Globalsecurity.org page on military topics.

http://www.nga.mil
US National Geospatial Intelligence Agency.

Military Geographies

R. Woodward, Newcastle University, Newcastle upon Tyne, UK

Glossary

Militarism The extension of military influence into civilian social, economic, political, and cultural life.
Military Activities The activities undertaken by military forces at the behest of the nation-state for defense and security purposes.

Introduction

Military geography is defined in two distinct ways. A long standing definition takes military geography to be the application of geographic information, tools, and techniques to the solution of military problems. A more contemporary definition takes military geographies to mean the study of the plurality of impacts of military activities and militarism on space, place, environment, and landscape. This second definition sees the relationship between militarism and geography as dynamic and interactive; militarism and military activities are both geographically expressed, in that they shape the world, and geographically constituted, in that they come into being in the form that they do as a consequence of the material and discursive conditions in which they evolve.

Both approaches – named here as military geography and military geographies for the sake of easy distinction – share a core concern with military activities and militarism as inherently spatial, and an understanding that these spatialities are multiple in cause and consequence. Both approaches to military geography are defined around military institutions and their practices as their primary object of inquiry, in contrast to, for example, political geography's focus on the state as the location of and actor in the exercise of military power. Both approaches recognize that the geographies of militarism and military activities are features of both the active and actual pursuit of armed force, and of the processes and practices surrounding the development and maintenance of military power beyond the immediate exercise of violence. Both approaches problematize the binary of war and peace and focus instead on the consistency of military institutions and the variety of ways in which military power is embedded and expressed.

These two approaches, however, speak to rather different positions on the role of academic inquiry. Traditional military geography is most usually a scholarship in the service of the military objectives of the state, and sees its engagement with military institutions as politically and morally unproblematic or unremarkable. The approach of military geographies, in contrast, is associated with a more critical position on the application of geographical knowledge, with an explicit focus on how military power is developed and deployed at a range of scales, and a self conscious moral positionality on the military production of specific configurations of social relations in space. Much of this work takes a highly critical view of the deployment of material resources and discursive strategies in the militarized pursuit of territorial control. It should be noted, though, that the extent to which antimilitarist and/or pacifist politics are promoted varies greatly between writers. It should also be noted that many writers on the spatialities of military activities and militarism see themselves neither as military geographers nor even as geographers; that their research and writing is included for discussion here is a testament to the interdisciplinary nature of military and spatial science, rather than to the coherence of intradisciplinary labeling.

The terms 'militarism', 'military activities', and 'militarization' have not been overused in Anglophone human geography, which arguably has avoided explicit engagement with military institutions and their effects. Militarism can be defined as the extension of military influence into civilian social, political, and economic spheres, with the associated prioritizing, promotion, and preservation of a nation's armed forces. Militarization can be defined as the processes through which military influence and priorities are extended to civilian life. Military activities are defined as something rather different and more practical, namely the activities that (usually) the nation state requires for its defense and security, or for military offensives and interventions beyond its borders.

It follows from these definitions that we can distinguish between the geographies of military activities, which are the patterning of both material entities and social relations across space as a consequence of the production and reproduction of military capabilities; and the geographies of militarism, defined as the shaping of civilian space and social relations by military objectives, rationales, and structures, either as the intended consequence of the extension of military influence into nominally civilian spheres of social life, or as a byproduct (perhaps unintended, sometimes not) of that process.

A Brief History of Geography's Engagement with Military Activities and Militarism

Military endeavor requires geographical knowledge; the conquest of territory and the assertion of control over space through violent means makes necessary an understanding of that territory's physical features and social structures. Organized warfare has therefore always conscripted geographical knowledge and the tools, techniques, and labor of geographers. As modes of warfare have evolved and changed, in terms of both the objectives and targets of military violence, and in terms of the development of technologies of organized violence, so also have the modes of geographical knowledge changed. Furthermore, ways of conceptualizing space according to the requirements of military strategy have also evolved, in conjunction with technological and objective development of war. This connection between geographers and military endeavor is identifiable across a range of historical contexts, from the ancient Chinese world to the conquests of medieval Arab powers.

In nineteenth century Western Europe the relationship between the imperial state and the emergent discipline of geography was close and mutually constitutive; in order to facilitate imperial expansion, the territorial objects of desire needed to be mapped, described, and known. For example, the British and German imperial adventures of the late nineteenth century could not have proceeded without the input of Halford MacKinder and Friedrich Ratzel. During the global conflicts of 1914–18 and 1939–45, in both Britain and the United States of America, geographers contributed to the war effort through the compilation of knowledge about overseas territories for their respective governments and military strategists.

In the post World War II era in Anglophone and Francophone human geography, specifically from the 1960s onwards, we can identify a rupture in the trajectory of military geography, the origins of the two rather different interpretations of the subdiscipline identified above. A more traditional and conservative military geography has proceeded virtually unchanged from its nineteenth century roots. So, for example, the application of geographical concepts to military endeavors continues to the present with the engagement of strategists within the Israeli Defense Forces with spatial theory to facilitate military tactics and strategy in the Occupied Territories, with the use of geographic information system (GIS) and remote sensing and other visual imaging techniques in Afghanistan by the armed forces of the United States and allies in their search for Taliban operatives, and with the descriptive physical and cultural geographies of Iraq produced by American military geographers in assistance to military occupation of the country.

A more critical military geography has its genesis in the emergence of radical and dissident politics emergent in the 1960s, inspired by social change, civil rights movements, and the rediscovery within the social sciences and human geography of Marxist theory as a conceptual underpinning for a geography alert to social justice issues. Critiques, for example, of United States Armed Forces' military strategies during the Vietnam War, particularly those targeted against civilians, emphasized an oppositional and critical antimilitarism politics in their descriptions of the effects of military activities on agricultural areas. However, with the exception of a self identified applied military geography, Anglophone human geography to an extent disengaged with questions of militarism and its geographies *per se*, attending instead to a focus on the outcomes of military activities and the role of the state.

Themes in Contemporary Military Geography

The range of themes that have emerged as either part of an explicit military geography/geographies project, or that could be incorporated within these distinct bodies of work, are diverse. Eight stand out and are discussed here.

A first theme is that of the historical geography of past military campaigns, a 'terrain and tactics' approach which takes as its area of study the influence of the impacts of geology, geomorphology, topography, biogeography, weather events, and climate on military campaigns and on specific battles. This approach is identified strongly with both traditional military geography, and also with geographical military history. Allied to this theme are studies which seek to assess the resulting impacts on landscapes and landforms of military activities, including the construction of fortifications and the effects of explosive munitions, and attempts to bring contemporary understanding to bear on past military events through archaeological and geoarchaeological tools and techniques. A key feature of much of this body of work is the use of topographic analysis tools and methodologies to establish how the natural environment affects the conduct of military campaigns. Evaluations of contemporary and historical geographies of military events have also examined the impacts on human populations of armed conflict and organized violence. Work in this area has focused, for example, on the disruptions to economic development and social cohesion, on health, welfare and education systems, on disease epidemics, and on industrial and agricultural productivity disruptions caused by warfare.

A second theme follows from this, and concerns the development of GIS tools and methodologies explicitly for military use. Particularly significant here are mapping systems – many of which have developed through col laborative military–academic endeavor – techniques for terrain analysis for combat purposes, and strategies for visualizing terrain and territory. Changing modes of warfare have been instrumental here in pushing the de velopment of techniques for visualizing and mapping urban warfare environments, as a prerequisite for military operations in urban terrain.

A third theme – and significant for revealing the na ture of human geography's engagement with military issues – is that on the spatialities of the pursuit of armed conflict, and the consequences of that. Although usually identified in disciplinary terms as a feature of political geography and critical geopolitics (few writing on these issues would self identify as military geographers), it merits mention because it constitutes a coherent body of work around military issues. Topics within this theme would include the causes of armed dispute between states, the wider regional and global power balances predicated on military interventions, and the con sequences for territorial integrity, borders, and bound aries of the pursuit of the state's objectives through armed violence by nation states. A critical geopolitics literature has emerged which assesses the militarized activities of nation state and subnational groupings in terms of their performative and representational strategies.

A fourth theme engages with the consequences for the spatiality of military activities of changing modes of warfare and motivations for the deployment of armed force. There is much debate within military science and related fields as to whether observable differences in the means of imposing armed force since the end of the Cold War (1989 onwards) have constituted some sort of revolution in military affairs. This debate centers around whether changes in military practice and capabilities have been solely or almost exclusively technologically driven, including developments in weapons, communi cations, intelligence and surveillance systems, in trans portation and mobilities, and in strategies for force deployment, or whether this revolution is more funda mental to the causes of armed conflict. This alternative argument sees post Cold War conflicts around the world as remarkable for the rupture they represent with the past in terms of their origins, rationales, civilian in volvement, and symmetry relative to opposing forces. Shifts include (but are not limited to) the rise of ethno nationalist conflicts, of religiously identified conflicts, and resource wars (particularly over oil), and at a more ab stract level, changing notions of security including en vironmental and resource security. A good example of how the spatialities of changing modes and motivations for armed conflict have been addressed is research and writing on the urbanization of military activities, in cluding weapons systems and force deployment tactics developed specifically to counter an unequal opposing force in urban terrain, strategies of spatial planning at the urban and regional levels explicitly to counter and de fend against new forms of armed violence, and strategies for the targeting of urban infrastructure (urbicide) as an explicit tactic for armed forces.

A fifth theme concerns the economic geographies of military activity and military resource use. All military forces are major users of land and material resources, and the utilization of both shapes both the form and function of militaries themselves, and the form and function of affected localities. In terms of land use, the military ownership and control of large areas of land is arguably necessary for the maintenance and reproduction of military power. Such land uses may be very extensive indeed – commentators estimate that an average of 1% of all land is used for military purposes in many advanced capitalist nation states. Such land uses may be contested of course, with competing claims for land use drawing heavily on alternative visions of both the material utility and sociocultural significance of such spaces. In addition, military land uses extend to overseas bases; the United States Armed Forces, for example, maintain a ring of military bases spanning around the globe. Again, inter pretations of these land uses are contested; are these a legitimate strategy for securing the interests of the United States, or a strategy for global hegemony forced by overwhelming military economic power interests? The consequences of both domestic and foreign military bases, in terms of their impacts on the economic geog raphies of localities and regions, are marked by their unpredictability. The type of facility, the nature of a region in which it is located, and the strength and density of the economic, political, and sociocultural connections which embed a facility in its locality will profoundly affect the military economic geographies imprinted upon space and place.

The economic geographies of militarism extend also to the defense industry, that sector engaged with the manufacture of weapons systems and other military matérial, and the networks and supply chains through which components are brought together, and manu factured objects distributed nationally and globally. Re search work on the changing economic geographies of the defense industrial sector has emphasized linkages between an increasingly globalized industrial sector and national state security objectives, highlighting the ten sions between the two. Of particular significance here is the capacity of the defense industrial sector to dictate military force deployment capabilities at a quite funda mental level, through supply led economies within the sector. The dependency of manufacturing regions on defense industrial activity has also been analyzed, with the technological developments in weapons systems driving a changing geography of production and distribution alongside trends in the distribution of manufacturing and R&D nationally and globally. Also significant here is research which suggests the limited capacity most advanced capitalist economies have for defense industrial conversion and diversification. Most recently, the privatization and outsourcing of state security functions under neoliberal economic regimes, particularly in logistics and services for military personnel on base and in theater, raises profound questions about the movement of the locus of control over security provision away from the nation state, that institution which traditionally has had sole legitimate authority over the deployment of violence. This issue is particularly marked with the increased reliance of private security personnel to perform security related functions in lieu of enlisted troops.

A sixth theme concerns the social geographies of militarism and military activities, an issue revolving around civilian responses to and engagements with military presence, practices, and personnel, and around the social and cultural geographies produced, expressed, and embodied by military personnel themselves. The social and cultural consequences of military bases, for example, impact on local populations in ways dependent not only on the nature of the economic linkages between a facility and its host population, but also on social relations beyond the base wire. Degrees of harmony and integration at one extreme, and over social disruption and abuse at another mark the variability of impacts. Military bases, for example, are commonly associated with high levels of dependency among local populations on illegal and unregulated economic activities, such as trade in drugs, alcohol and armaments, and prostitution. The abuse and exploitation of certain local social groups, particularly women, by soldiers is a feature of many overseas bases. Domestic military bases generate their own social relations of security and danger, trust and mistrust.

Geographies of resistance and opposition also follow the military control of space. Such activities, ranging from localized disputes over specific military uses for particular sites to wider campaigns against a military presence in itself, are significant for the alternative conceptualizations of the use of space which they invoke. Protests against specific uses include resistance to activities on the grounds of their impacts on, for example, natural environments, valued landscapes, and social and economic well being. Antimilitarist protests against a military presence (such as the US military presence in Okinawa) or against specific military installations (such as nuclear weapons systems, or military intelligence sites) invoke moral objections to a more abstract notion of the militarization of space and place.

A further set of cultural and social geographies of militarism and military activities relate to the spatialities of gender relations and gender identities within military institutions. Military institutions, particularly residential military bases, have social dynamics that may be starkly gendered in their spatial expression, and these in turn may be overlain by military–civilian differences. An example would be the differential access to and use of space in those military installations which function simultaneously as military garrison, military installation, and domestic residence for the families of military personnel. The gender dynamics within armed forces are themselves spatialized through the designation of gendered spaces, and the construction and reproduction of gendered military identities through the use of space. An example would be the embodied practices of military training as a means of constituting particular military masculine identities, and the fact of the practice of this across challenging spaces and landscapes.

A seventh theme is that of the environmental impacts of military activities, and the politics of military environmentalism. Armed forces are key users of global natural resources, including energy resources. Furthermore, military activities have pronounced environmental effects in both conflict scenarios, and in nonconflict situations through the environmental effects of the ongoing maintenance of the capacity of military institutions to exercise lethal force. There is little consensus as to whether a value judgment can be made about conflict versus nonconflict environmental damage; although the environmental impacts of armed conflict can be profound – witness the torching of Kuwaiti oil wells during the 1991 Gulf War – the ongoing and normalized use of domestic and overseas facilities by military facilities carries with it the risk of long standing and deleterious environmental impact.

The environmental impacts of military activities take many forms. Pollution is a key issue and includes both the visible rubbish and detritus of military activities such as the remains of ordnance (and unexploded ordnance) on training areas, and the hazardous littering of unexploded ordnance, particularly landmines, in former conflict sites. Nonvisible pollution hazards follow military forces around the contemporary world, and is the biggest single environmental issue associated with military bases. Chemical pollution includes toxicity to natural environments from the fuels, lubricants, solvents, and explosives used in the maintenance and testing of military vehicles, munitions, and weapons systems. Chemical and radiological hazards have been noted from the use of specific weapons, such as the use of the Agent Orange herbicide by the US Armed Forces during the Vietnam War, and the use of depleted uranium tipped projectiles during the 1991 Gulf War. The testing of nuclear weapons has in effect rendered some former testing sites uninhabitable, places such as Bikini Atoll in the South Pacific, and the weapons testing grounds of the Southwestern United States. A further environmental impact is that produced through environmental modification, caused by explosives impacts in testing and in conflict, the destruction of specific habitats and the impact of noise, footfall, and tracked vehicular movements.

The politics of military environmentalism have provoked critical commentary from geographers interested in the military representation of armed conflict and related activities. At what point do the environmental impacts of military activities come to be described as destructive? Are there distinctions to be drawn on moral grounds between the environmental impacts of military activities versus other industrial, commercial, or leisure related land uses? Arguably, armed forces are just one of many institutions, particularly in advanced capitalist nation states, which cause both environmental modification and damage, and which seek to justify or rationalize such damage with reference to the necessity of such activities. What is significant here is not so much that environmental damage occurs but the reference to security and defense objectives as an overriding justification for such impacts.

The eighth and final theme to be considered here is the cultural geographies of militarism, particularly the construction and representation of military landscapes. Anglophone human geography's focus on landscape as a form that is both constructed and legible has provided a powerful impetus to studies of the ways in which military power is physically inscribed onto landscapes. The militarized construction and representation of landscapes is a strategic act. Work on the impacts of military nuclear testing in the Southwest USA, for example, emphasizes the representational practices of the US Armed Forces and Department of Defense in describing these desert areas as uninhabitable and unproductive spaces, a strategy of legitimation for the testing of nuclear weapons and a counter to local indigenous representations of this space as a living, working, productive environment. Another example would be the significant body of work that has built up around sites of memory, mourning, and memorialization, which fix into (usually urban) landscapes specific interpretations of past conflicts (see, for example, the strategies of memorialization that followed World War I, in the construction of war graves in former battlefields, and the placing of war memorials in cities, towns, and villages back home). Notable too is the fluidity of meaning around such places, visible in the changing national interpretations of former Nazi concentration camps in Central Europe, and the incorporation within national monuments of mourning and memorialization of multiple categories of participants of former wars. No longer are soldiers the only ones so honored.

Current Issues of Theory and Practice in the Study of Military Geographies

Four issues of theory and practice in the study of military geographies dominate current work in the field. The first is recognition of the need for a theorized approach to the pursuit of military power across space, one which recognizes the conceptualization of space and spatiality deployed by military forces (whether in operations or in basing) and is alert to the power geometries that such conceptualizations entail. The second issue relates to access to information and data on military activities, across the spectrum of such activities. A paucity of information on activities and impacts across the social, environmental, and economic spheres has long limited the nature of investigations into military geographies and the geographies of militarism. The third issue is the politics of the engagement of the academy with military power and military institutions. For a more traditional military geography, that engagement is relatively unproblematic. Within the critical political tradition in human geography, along a spectrum from pacifism to critical observation, issues of engagement with military institutions raise moral questions about the role of academic geographers. Sometimes those issues are starkly clear – whether or not researchers undertake research funded by military institutions, for example. Sometimes those issues are more complex; this encyclopedia, for example, is published by Reed Elsevier, a multinational corporation which through a subsidiary company ran a major biennial arms trade exhibition in London, the Defence Systems and Equipment International (DSEi). This connection led to debate within and beyond geography on whether academics should support through their publishing in Reed Elsevier books and journals, a key player in the international arms trade. The fourth issue – and this is entirely speculative – concerns the influence of wider global political developments on philosophical and conceptual trajectories within human geography. As the consequences become more apparent of the enormous human costs of the unrestrained use of military force by the US state and allies in Afghanistan and the Middle East, we should consider carefully how the discipline of geography can respond adequately to what many name the 'Long War', what it might mean to place war, and the outcomes of armed violence at the center of our geographical inquiries.

See also: Critical Geopolitics; Geography, History of; Geopolitics.

Further Reading

Caldwell, D. R., Ehlen, J. and Harmon, R. S. (eds.) (2005). *Studies in Military Geography and Geology*. London: Kluwer.

Clout, H. and Gosme, C. (2003). The naval intelligence handbooks: A monument in geographical writing. *Progress in Human Geography* 27, 153 173.

Flint, C. (ed.) (2005). *The Geography of War and Peace: From Death Camps to Diplomats*. Oxford: Oxford University Press.

Gilbert, E. and Cowan, D. (eds.) (2007). *War, Citizenship, Territory*. London: Routledge.
Graham, S. (ed.) (2004). *Cities, War and Terrorism: Towards an Urban Geopolitics*. Oxford: Blackwell.
Kaldor, M. (2000). *Global Insecurity: Restructuring the Global Military Sector*. London: Pinter.
Kuletz, V. (1998). *The Tainted Desert: Environmental and Social Ruin in the American West*. London: Routledge.
Lacoste, Y. (1973). *La Géographie, ça sert, d'abord, à Faire la Guerre*. Paris: Maspéro.
Mamadouh, V. (2005). Geography and war, geographers and peace. In Flint, C. (ed.) *The Geography of War and Peace: From Death Camps to Diplomats*, pp 26 60. Oxford: Oxford University Press.
Palka, E. J. and Galgano, F. A. (eds.) (2000). *The Scope of Military Geography: Across the Spectrum from Peacetime to War*. New York: McGraw Hill, Primis Custom Publishing.
Shapiro, M. J. (1997). *Violent Cartographies: Mapping Cultures of War*. Minneapolis, MN: University of Minnesota Press.
Woodward, R. (2004). *Military Geographies*. Oxford: Blackwell.
Woodward, R. (2005). From military geography to militarism's geographies: Disciplinary engagements with the geographies of militarism and military activities. *Progress in Human Geography* 29(6), 718 740.

Mixed and Multiple Methods

J. H. McKendrick, Glasgow Caledonian University, Glasgow, UK

Glossary

Analytical Levels (Analytical Scales) They are realms of enquiry, each of which is associated with a different geography, for example, personal, familial, neighborhood, regional, national, and global. The objectives of enquiry and the manner in which structural forces are relevant to enquiry vary across these levels.

Ecological Fallacy The error of ascribing characteristics to individuals on the basis of where they live.

Epistemology This refers to systems of thought, which describes different ways of viewing, understanding, and appraising the social world.

Multilevel Modeling An approach to data analysis which simultaneously allows the relative significance of different analytical levels to be ascertained, for example, by incorporating personal, familial, neighborhood, and structural variables in the analysis.

Introduction

Multimethod research describes the application of more than one method to engage a research question. However, as demonstrated in this article, this deceptively simple description belies a wide range of approaches to research.

Toward the end of the 1990s, there was a growing interest in multimethod research in human geography, with seminars, conference sessions, and special editions of journals debating this approach to research. The catalyst for this was dissatisfaction with the entrenchment of researchers into either side of a qualitative/quantitative divide, which was emerging as the manner in which qualitative research designs were mainstreamed in human geography was implicitly – and often explicitly – positioned in opposition to quantitative research designs.

There has been little direct discussion of multimethod research in geography since the 1990s. Textbooks of contemporary research methods in human geography pay little attention to multimethod research. Although textbooks tend to introduce students to quantitative research and qualitative research, there is virtually no discussion of bringing these together in multimethod research. Thus, the multimethod debates of the 1990s appear to have had limited impact on how students are introduced to research. One possible reason for the limited impact of the multimethod debates of the 1990s was that they primarily focused on using multimethod research to bridge the qualitative/quantitative divide, which is only one of several reasons for pursuing multimethod research.

Although fading from the fore in methodological debate, multimethod research continues to be widely practised by human geographers. Indeed, multimethod research is not common in geography and should be considered as one of the defining features of the regional geography paradigm that prevailed in the second half of the nineteenth and the first half of the twentieth century.

An Example of Multimethod Research in Human Geography

Bradford and McKendrick's exploratory study of the commodification of children's play space in the UK demonstrates the utility of multimethod research in human geography. Fieldwork commenced with the aim of reviewing developments in the commercial provision of playgrounds for young children in North West England. Key questions were addressed with respect to social exclusion, children's role as clients and consumers of services, family leisure, and children's 'place' in the built environment. A carefully balanced and interlinked research design encompassing questionnaire surveys, interviews, field observations, and video recordings was administered to children, parents, service providers, and play center staff. In summary, six commercial providers of children's playgrounds were profiled and research was conducted in ten field sites in the Greater Manchester area; 872 families were surveyed, of which 30 were interviewed (adults and children separately), and, in turn, from which seven participated in a second interview. This second interview was based on the family's video diary of a day visit to a commercial playground. Themes which cut across each of the methods emerged to shed insight into disabled children's experience of play space, children's birthday parties, family decision making, and the role of playgrounds in urban landscapes.

Definition: The Parameters of Multimethod Research

Mixed, Multi, Multiple, or Integrated?

Leaving aside the grammatical questions of whether the concept should be hyphenated (e.g., multi method), separated (multi method), or joined (multimethod), and whether or not the concept should be abbreviated (i.e., multi method) or (not multiple method), there are two substantive issues pertaining to adjectival part of the concept that must be appreciated.

First, a distinction is sometimes drawn between multimethod research and mixed method research. Mixed method research is sometimes specifically used to describe hybrid applications in which two or more approaches are used concurrently (at the same stage in the research, at the same time, in the same place, and with the same group being researched). Hybridity can refer to a single research instrument collecting different types of data, for example, a survey comprising questions with fixed responses (amenable to quantitative data analysis) and questions with open responses (amenable to qualitative data analysis). Hybridity might also refer to using a single research instrument to collect data, which are themselves amenable to both quantitative and qualitative data analysis, for example, subjecting responses to open ended questions to quantitative data analysis (frequency of response) and qualitative data analysis (interpretative understanding of relationships and structure). When defined in opposition to mixed method research, multimethod research describes the independent application of more than one method in a research design. However, multimethod research is not always defined in contrast to mixed method research and is also understood to be a generic description for hybrid applications (described here as mixed method research) and the independent application of more than one method.

Second, integrated research is sometimes used by those who wish to emphasize the intentional deployment of more than one method in a research design. Although most proponents of multimethod research would agree that research should not be described as such if it merely comprises simple collections of qualitative or quantitative data without planned use of diverse methods with a serious attempt to integrate the conclusions, other commentators are less insistent on this point.

Stage and Level (Scale)

Multimethod research is often staged and is described more comprehensively as multimethod, multistage research. The archetypical multimethod project is multistage in design, that is, the questionnaire survey (stage 1), which identifies key informants who are interviewed (stage 2) to explore details of key findings emerging from the survey.

More controversial is whether multimethod research should be used to describe research that spans more than one analytical level (scale). Some geographers have argued that research should not be described as multimethod if it spans more than one analytical level. Most concern here is directed at researchers transposing findings from areal levels (neighborhood, city, etc.) to social levels (individuals and households) and committing an ecological fallacy. There are two problems with not considering multilevel research to be multimethod research on these grounds. First, it is reasoned that such research cannot be described as multimethod as the different components of this research (at different analytical levels) do not address the same issues. However, this critique rests an overly narrow definition of a research question in which research is not considered to be whole if it involves an analysis over more than one analytical level. This is particularly limiting for the geographer who may be more inclined than other social scientists to directly explore relationships across analytical levels, for example, making sense of household behavior in the context of neighborhood norms. Similarly, and equally problematically, this critique also argues against the premises of multilevel modeling. Second, this critique tends to assume that researchers working across analytical levels will be prone to transposing results across levels. This wrongly makes the assumption that the researcher will transpose findings, as opposed to using them to contextualize or inform understanding at other analytical levels. As for multistage research, a more comprehensive description for research involving more than one method across analytical levels is multimethod, multilevel research.

Multimethod Research, Multimethod Research Design, and Multimodal Research

The sophistication of the multimethod approach, or the ends to which it can put, can be described as progressing through three tiers, with each subsequent tier incorporating and extending the scope of the former.

First, at the most simplistic level, multimethod research is when more than one research method is used to address a research problem. Second, multimethod research designs not only comprise more than one method, but they also deliberately incorporate more than one method in research that consciously plans to capitalize upon the strengths afforded from using complementary approaches. Finally, multimodal research designs refer to when research deliberately embraces qualitative approaches and quantitative approaches to research.

The higher orders of this hierarchy are open to critique. Some might question the practical limits to which knowledge can be advanced by bringing together findings from diverse methods (multimethod research design), while others might question the epistemological validity of embracing qualitative and quantitative approaches (multimodal research). These 'key debates' are discussed below.

Uses of Multimethod Research

There are four reasons why human geographers practised multimethod research, most of which comprise several dimensions.

Data

Unsurprisingly, most researchers refer to data considerations when accounting for their use of multimethod research.

First and foremost, multimethod research affords the possibility for achieving congruence through data triangulation, that is, to enhance the credibility of research by using different methods to generate similar, consistent, or convergent findings, thereby reaching the same conclusions.

Second, multimethod research can be used to extend the comprehensiveness of research by combining methods, each of which generates particular data, which when brought together are complementary and can broaden understanding of the issue at hand by enriching, expanding, clarifying, or illustrating.

Third, multimethod research can be used to highlight meaningful inconsistencies between findings produced through different approaches to the same research question. Here, the objective is to reflect upon, understand, and explore why different realities are constructed (and not to source 'the truth').

Fourth, explanation can be pursued through multimethod research. Here, the objective is not merely to extend the knowledge base. Rather, the objective is to design a subsequent stage in the research to specifically seek explanation for emergent findings (which may be unclear or controversial) from an earlier phase of the research.

Finally, and in a similar vein, multimethod research can be used to test hypotheses, which emerged from an earlier phase of research.

Philosophy of Scholarship

Multimethod research is consistent with traditional and some contemporary ideals of scholarship.

First, the breadth of understanding that is provided by multimethod research is consistent with approaches to geographical scholarship that preceded the systematic tradition. More specifically, multimethod research was routinely used to construct geographical understanding in regional geography. Second, multimethod research is also consistent with more recent concerns to 'give voice' to marginalized groups and to promote 'polyvocal' research. Similarly, a commitment to ethnography on the grounds that it offers a depth to understanding through immersion in the field is inherently a commitment to multimethod research, as participant observation is invariably buttressed by nonparticipant observation and document analysis.

However, although a multiplicity of methods are implicated in these endeavors, research would only be understood as multimethod if it involved the researcher actively using more than one method to explore research questions.

Epistemological Preferences

Multimethod research has been criticized by some on epistemological grounds – these debates are considered below. Here, it is important to note that there are three epistemological positions, which range from being tolerant of multimethod research to advocating this approach.

First, from an aparadigmatic stance, practical decisions about research design and method should be influenced by the demands of the research and not by philosophical assumptions. Here, epistemology should not determine practice. Second, a dialectical stance acknowledges that philosophical assumptions inform research design, and proposes that these assumptions should be intentionally mixed to overcome limitations and partiality to produce more comprehensive and insightful research. Finally, a pragmatic stance advocates an inclusive framework which embraces multiple assumptions (and diverse methods).

Political

Finally, it should be recognized that the reasons for pursuing multimethod research may owe less to research design *per se*, than to ulterior motives. First, as outlined above, some of those whose primary concerns are epistemological or with particular ideals of geographical scholarship may be more inclined to pursue multimethod research to realize their goals. Second, multimethod research may be used as a learning tool for new researchers, for example, using different methods to engage the same question on geographical field courses. Finally, multimethod research may be used to gain the confidence of an audience. It has been argued that policymakers have been more wary of conclusions drawn from small sample, in depth, qualitative investigations; careful integration of complementary large scale quantitative survey results may be helpful in gaining the trust of an otherwise skeptical audience.

Styles of Multimethod Research

Attempts to devise a typology of have been troubled by the number of factors against which multimethod research can be categorized.

First, multimethod research can be classified according to the number of component studies/methods that are used.

Second, method type is a second basis for classification. The archetype multimethod research is the survey questionnaire combined with the in depth (key informant) interview. However, there is a wide range of multimethod research possibilities beyond this, ranging from combining very different types of method (e.g., controlled field experiment and ethnographic participant observation) to combining methods that are merely slight variations of a methodological genre, for example, 'focus group interview' data with data from in depth 'personal interviews'.

Third, research can be distinguished according to whether methods are applied concurrently or sequentially. In concurrent designs, data are collected to test for congruence, to collect complementary data, or to highlight meaningful inconsistencies. In sequential designs, later strands can fulfill each of these goals or they can be used to explain or test the findings of the previous strand.

There may also be grounds for further subclassification by juxtaposing timing and method type, for example, distinguishing the large scale survey which precedes the key informant interview (perhaps to probe for explanation for why there is an association between survey variables) from the key informant interview which precedes the large scale survey (perhaps when the interview is used to generate response options for survey questions).

Fourth, the relative importance of the respective methods in multimethod research can be used as a basis for classification. Although all methods make an important contribution to the research, some components of the multimethod research may be more dominant or may be deemed to be more important to the overall project. On the other hand, the component parts may be considered to be of equal worth.

Fifth, multimethod research can be distinguished according to the point at which findings are integrated, that is, throughout the study or kept separate until the end.

Sixth, multimethod research can vary according to whether it has independence or interdependence between methods. That is, whether subsequent methods are applied without reference to preceding methods, or whether subsequent methods follow on from and develop preceding methods. On the one hand, it might be argued that independence is needed to ensure that results are not influenced by initial research. In contrast, it could be argued that interdependency is needed to ensure that the full complexity of a research problem is addressed.

Seventh, whether or not the research is undertaken by a team (perhaps with their own methodological specialisms) or by a sole researcher (by necessity, with competence across methods) is also used as a basis to classify multimethod research.

Eighth, a distinction can be drawn between a 'multimethod project' and a 'project or program of research that utilizes multiple methods'. Although both may pursue the same methods to the same ends, in the former, the multimethod work is designed as a single project, whereas, in the latter, the second stage research builds upon the work of an earlier research project.

Ninth, many classifications of multimethod research make reference to whether the design works within or across methods. As discussed above for hybridity, 'within method' multimethod research is when a single research instrument collects different types of data or collects data which is amenable to both quantitative and qualitative data analysis. 'Across method' multimethod research is that in which different types of data are collected through contrasting research methods.

Tenth, the nature of the multiple in multimethod research can vary. There is a tendency to literally interpret 'method' as 'technique' when discussing multimethod research. However, it should be acknowledged that the multiple in question could pertain to how the method is applied, that is, the research could use multiple sampling frames, data sources, or analytical techniques.

Finally, and as was also discussed previously, multimethod research can be applied within or across analytical levels.

Key Debates in Multimethod Research

Criticism of multimethod research has been directed at practicalities and its philosophical underpinnings.

Practical Challenges

The first weakness with multimethod research is to acknowledge that it is more demanding in terms of time, financial cost, and researcher skills than single method applications. Not all researchers are equally comfortable or confident in applying a range of methods. In particular, this makes multimethod research a less realistic proposition for small scale research, such as that required of undergraduate dissertations.

Second, the pursuit of triangulation for congruence makes it difficult to deal with disparity between datasets. In particular, there is a need to avoid the tendency to try and identify which set of data is the most valid.

Third, proponents of multimethod research should avoid the tendency to present 'breadth' of study as an inherent virtue. While the search for more comprehensive understanding is a laudable goal that could be achieved through multimethod research, multimethod research is only a virtue if the design affords the subsequent/concurrent method to add substantially to what is known from other components of the research. Similarly, it should not be assumed that measurement error will always be reduced through multimethod research.

Fourth, there is a danger that, in seeking to capitalize upon the complementary strengths of different methods, the pursuit of multimethod research encourages a more limited application of method than might otherwise prevail.

Deterministic Conjunctions of Epistemology and Method

The most fundamental criticism of multimethod research is that the application of more than one method in

Table 1 How ep stemo ogy nforms, rather than prec udes, methodo og ca strategy: Examp es from m grat on stud es

Research tradition	*Objectives*	*Migration example*	*Methods and application*			
			Survey/questionnaire	*Interview*	*Fieldwork*	*Nonreactive*
Pos t v st	Estab sh emp r ca regu ar t es wh ch are assumed to be of genera (un versa) s gn ficance	(Patterns reasons for and outcomes ar s ng from) Counter urban zat on from SMSAs	**Mathematical modeling of census migration data to provide a national overview and trend projection**	Content ana ys s of responses to a structured nterv ew to estab sh why househo ds dec ded to m grate	F e dwork survey n urban fr nge to dent fy su tab e s tes for 'm grant' res dent a deve opments (based on nterv ew?)	Comparat ve systemat c ana ys s of m grant and nonm grant ch dren's wr t ng about what they ke most about the r (new) home
Human st c	Va or ze human exper ence and seek to understand the mean ng va ue and human s gn ficance of events	'Return' m grat on of 'Amer can' Jew sh fam es to srae	Deta ed soc a att tudes quest onna re to estab sh and understand the persona context of (nonfunct ona reasons for) th s m grat on	n-depth nterv ew to exp ore the mean ng of th s m grat on exper ence to each nd v dua m grant	**Ethnographic, participant observation to share the emotions, experiences, and significance of this particular migration**	Ana ys s of reported accounts (autob ography speeches magaz ne art c es) of return m grat on to s tuate the persona s gn ficance of the act n ts w der soc o-po t ca context
Rea st	dent fy the structures wh ch generate outcomes v a mechan sms (necessary causa powers) under 'spec fic cont ngent cond t ons'	M grat on assoc ated w th decentra zat on of government departments from London to other parts of Br ta n	Expans on method data mode ng wh ch be ng attent ve to context enab es the spec fic t es of each department's m grat on to be so ated from the genera exper ence	**In-depth interview with senior civil servants to uncover the underlying processes behind these departmental migrations**	Part c pant observat on w th n the c rc e of dec s on makers to estab sh the processes wh ch contr ve to produce the resu tant m grat on	nterpretat on of a ser es of nterna departmenta br efing papers to chart the tempora deve opment (fe course) of the processes at work
(Fem n st) Standpo nt Theory	Know edge s soc a y constructed Estab sh a successor sc ence n wh ch unpr v eged know edge (e women's) s recogn zed and va or zed	M grat on exper ences of women partners (of) ma e company execut ves	mp ementat on of a survey des gned by women us ng the r (fam ar) anguage to generate an overv ew of m grat on exper ences that they co ect ve y deem to be mportant	**Focus group interview to share and rationalize migration experiences among the group**	Unobtrus ve observat on of fam y processes nvo ved n these m grat ons undertaken wh e fu fi ng a serv ce-sector funct on (remova s estate agency house search mortgage ender etc)	nterpretat on of exper ences conta ned w th etters comm ss oned for a women's journa on the subject of 'Me and my partner on the move'
Postmodern st	Estab sh that the 'mu t p e' pos t on ng of the author (or reader) has nfluenced the product on (or nterpretat on) of the narrat ve	M grat on of e te 'Th rd Wor d' women	Log- near mode ng of m grat on data to est mate the s gn ficance of d fferent 'pos t ons' on the propens ty to m grate	n-depth nterv ew w th the women to 'unpack' the r rat ona zat ons of the r m grat ons	Unobtrus ve access to m grant's exper ences v a a serv ce funct on (e g abor market consu tant nt remova contractor)	**Unpack autobiographies, personal journals, and/or letters of migrants which discuss the migration experience**

McKendrick J H (1999) Multi-method research An introduction to its application in population geography *Professional Geographer* 51(1) 40–49 research can undermine epistemological validity. Proponents of multimethod research have observed that deterministic (or at least conservative) conjunctions of epistemology and method hold sway. In contrast, they argue that the range of methods that can be used within any epistemological tradition is greater than is generally assumed. This expansive view of methodological possibilities should not be interpreted as a claim that 'all methods are equal'. On the contrary, **Table 1** uses bold typeface to suggest that, for each research tradition, some methods are more useful than others, and that the preferred method varies between research traditions. Thus, the survey questionnaire is particularly useful for the positivist, the interview is likewise useful for the realist and the standpoint theorist, those working within a humanistic perspective would favor experiential fieldwork, while the postmodernist would be inclined to utilize 'documented' sources. These conjunctions are, and should be, the foundation of most research endeavor. This should not preclude the application of other methods, most obviously not only for circumstances where data availability is a problem, but also for circumstances whereby 'alternative' methods provide complementary insights.

The Future of Multimethod Research in Human Geography

Multimethod research is an attempt to combine research methods to address a particular research problem. It is a generic term that encompasses a wide range of research strategies: it may be deployed strategically, it may be used over the course of a research project, and it may breach the qualitative/quantitative divide or it may be practised within each camp. This article has clarified the essential character of multimethod research.

Multimethod research has become something of an absent presence within human geography. That is, it is widely practiced, but is not openly discussed in methodological debates. It presents a range of opportunities, both substantive and strategic, and methodologically, is an approach that is both tenable and viable. Multimethod research fulfills a number of research objectives ('tactical' and compensatory), is consistent with traditional models of scholarship, offers a model of research which destabilizes methodological dualisms, and, perhaps most importantly of all, provides the means to improve the effectiveness with which these issues are addressed.

See also: Interviews: In-Depth, Semi-Structured; Quantitative Methodologies; Transcripts (Coding and Analysis).

Further Reading

Graham, E. (1999). Breaking out: The opportunities and challenges of multi method research in population geography. *Professional Geographer* 51(1), 76 89.

Hoggart, K., Lees, L. and Davies, A. (2002). Multi method and multi level analysis. *Researching Human Geography*, pp 65 73. London: Arnold.

Lawson, V. (1995). The politics of difference: Examining the qualitative/quantitative dualism in poststructuralist feminist research. *Professional Geographer* 47(4), 449 457.

McKendrick, J. H. (1999). Multi method research: An introduction to its application in population geography. *Professional Geographer* 51(1), 40 49.

McKendrick, J. H., Bradford, M. G. and Fielder, A. V. (1999). Privatization of collective play spaces in the UK. *Built Environment* 25(1), 44 57.

McKendrick, J. H., Bradford, M. G. and Fielder, A. V. (2000). Enabling play or sustaining exclusion: Commercial playgrounds and disabled children. *The North West Geographer* 3(2), 32 49.

McKendrick, J. H., Bradford, M. G. and Fielder, A. V. (2000). Time for a party! Making sense of the commercialisation of leisure space for children. In Holloway, S. & Valentine, G. (eds.) *Children's Geographies: Living, Playing, Learning*, pp 100 119. London: Routledge.

McKendrick, J. H., Bradford, M. G. and Fielder, A. V. (2000). Kid customer? Commercialisation of playspace and commodification of childhood. *Childhood: A Global Journal of Child Research* 7(3), 295 315.

Philip, L. J. (1998). Combining quantitative and qualitiative approaches to social research in human geography An impossible mixture? *Environment and Planning A* 30, 261 276.

Rocheleau, D. (1995). Maps, numbers, text and context: Mixing methods in feminist political ecology. *The Professional Geographer* 47(4), 458 466.

Mobility

P. Merriman, Aberystwyth University, Aberystwyth, UK

Introduction

Mobility and movement are of concern to academics working across the humanities and the social, physical, natural, and medical sciences. Physical geographers have concerned themselves with the movements of water, ice sheets, soil, plant species, and radiation, to name just a few examples, while mobility has been a long standing concern of human geographers working on a broad range of issues, from migration, transport, and trade, to imperial expansion, travel writing, tourism, globalization, and cultural diffusion. The next section examines how geographers have approached the issues of mobility and fixity, before showing how the emergence of anti essentialist post structuralist thinking, participative methods, and an interdisciplinary concern with themes such as embodiment, globalization, transnational migration, tourism, mobile communications, the Internet, performance, and the spaces of the car and airport, has led to a resurgence of writings focusing on mobility across the social sciences. The subsequent section examines the power relationships surrounding particular kinds of mobilities, and the discourses of exclusion which have gathered around the mobilities of such groups as gypsies, travelers, tramps, and immigrants. Then we explore how particular mobilities and senses of movement are accounted for in geographical writings on place and landscape. Mobilities have commonly been held to erode distinctive senses of place, generating feelings of placelessness, but in recent years geographers have come to theorize places and landscapes as dynamic, practiced and performed through the movements of all manner of things. The final section discusses recent work on cultures of transport, focusing on the geographies of the car and road.

Geography, Fixity, and Mobility

Geography – a term whose etymology can be traced through Greek, Latin, and subsequently French – is often defined as Earth writing. When I talk to friends who have a limited knowledge of academic geography, they frequently suggest that geography is about physical landscapes, maps, or places. It is said that geography is, or should be, concerned with locating things and events, particularly physical landscape features. Fixity and locatedness emerge at the heart of such imaginaries of what I and other the geographers do, and in these definitions, fixity is frequently constructed in tension or opposition to mobility. Things move or stay still. They are stationary or mobile. As Tim Cresswell has shown, throughout the history of geography, academics have developed a range of theoretical approaches and methodological tools for accounting for and exploring the relationships between mobility and fixity.

Mobility has been framed as unfortunate, abnormal, and problematic, and as inevitable, normal, and productive. In the 1960s, positivist constructions of human geography as a spatial science frequently constructed fixity or locatedness as the norm or ideal. Where movements occurred, they were constructed as rational, functional, and purposeful movements between fixed points. In the 1970s, time geography examined the mobilities of people engaged in everyday activities, while Marxist thinkers have long emphasized patterns of mobility, flow, and circulation in society and the economy. In focusing on place, many humanistic geographers constructed mobility as an eroding force or threat to the integrity of places. Human geographers working in the areas of transport geography, migration studies, and tourism have also done much to advance geographical research on particular kinds of mobilities. Early studies tended to focus on the causes and effects of particular mobilities, or on the quantitative dimensions of mobilities. However, since the late 1980s and 1990s there have emerged a large number of qualitative studies of mobility marked by a new wave of research on tourism, transport and migration, and the emergence of an interdisciplinary interest in themes such as globalization, transnational migration, diasporic cultures, mobile communication technologies, the Internet, performance, and postcolonialism.

With these developments, and the increasing purchase of anti essentialist post structuralist thinking, and participative and performative methods, it has been suggested by Mimi Sheller and John Urry that the social sciences are seeing a 'mobilities turn' and the emergence of a 'new mobilities paradigm'. Mobilities research does appear to be flourishing, but it is important to be cautious about such proclamations. Mobilities research is not that new. Much contemporary research builds upon long established and broader reaching developments in the social sciences. New and emerging research questions do not necessarily lead to a 'turning' of disciplines and research agendas, or the establishment of new paradigms. It is important, therefore, not to overstate the influence of these new agendas, or to overlook more firmly established lines of research. Indeed, while 2006 saw the launch of the interdisciplinary social science journal *Mobilities*, and the publication of a number of important books on mobilities, we should not forget that the *Journal of Transport History* was first published in 1953, *Transportation Research* in 1967, and the *Journal of Transport Geography* in 1993.

Mobilities research does not simply provide a new unproblematic nonessentialist way of thinking, tracing narratives built around metaphors of mobility and fluidity, and allowing academics to move away from sedentary theories rooted in ideas of fixed, solid essences. This was the mistake made by a generation of post modern, post structuralist, and feminist theorists who embraced a 'nomadic metaphysics' that celebrated (and frequently romanticized) the transgressive mobilities of the migrant, traveler, nomad, or theorist in an attempt to depart from a 'sedentarist metaphysics'. There is a politics and geography of power bound up with practices and discourses of both mobility and fixity, and as Tim Cresswell has asserted, it is important for human geographers and other social scientists to expose these power relations and examine the social, spatial, and material processes by which particular mobilities are produced. Modern Western societies may appear to emerge and gain life through the mobilities of all manner of things, but our worlds are punctuated by a broad range of material things and sedentary beliefs. In Western nations it is commonly assumed that their inhabitants will have nationalities, fixed dwellings, postal addresses, and either lease or own property. The geographies of mobilities are inseparable from particular materialities, whether of transport infrastructures, passports, human bodies, border fences, or walking boots. What is more, when one travels, one does not simply travel across the landscape. Mobilities rework, shape, animate, and perform places and landscapes.

Mobility, Politics, and Exclusion

Advertisers, manufacturers, and academics delight in informing us that our communications and movements are speeding up, the world is shrinking, time–space compression is occurring, and that there is an 'end to geography'/'geography is history'. We are said to be increasingly mobile, living in a 'hyper mobile' world, and yet through such assertions we are provided with little sense of the differential experiences and impacts of mobilities. The geographies and histories of mobilities are often missing, and popular cultural commentators seem to overlook the long and tortured history and complex geography of globalization. As Doreen Massey points out, we do not all have access to e mail, television, public transport, private vehicles, or immigration visas. New connectivities and mobilities produce geographies of exclusion, disconnection, inequality, and immobility. Pacific islands which were once key stopping points for passing ships are now by passed by intercontinental air liners. In rural Britain, cuts in public transport provision and post office closures have had a detrimental impact on the communications and mobility of vulnerable groups without access to cars or the Internet. Indeed, while the movements of the business traveler, commuter, Western tourist, quarantined animal, and airmail letter may be facilitated by businesses and governments, the mobilities of less empowered minorities are frequently overlooked, while the movements of groups such as refugees, gypsies, and migrant workers are closely regulated and commonly criticized. People and things, then, are not always free to move. Human and nonhuman bodies have different capacities to move, or purposes for moving. The mobilities of human bodies, water, oil, and many other things may appear to be governed by 'natural laws', but their movements are enmeshed in complex political, economic, and cultural geographic processes. Governments, large companies, academics, and a broad range of individuals have long been concerned with how mobilities are performed in different ways and at different scales, from the microspaces of the human and animal body, to the scale of the nation state or globe.

Governing Mobilities

Mobile entities are incessantly identified, classified, and monitored. The mobilities of commodities are tracked by logistics experts and distribution companies using computers, Universal Product Code (UPC) bar codes, radio frequency identification (RFID) tags, and written inventories. Customs and excise agencies attempt to stem the flow of illegal drugs, untaxed cigarettes and alcohol, or unauthorized agricultural imports, tracking their movements using international policing networks, informers, and more or less sophisticated searching technologies. Animals may be quarantined. In the UK certain species of domestic pet (including dogs, cats, and ferrets), once microchipped and vaccinated, can be issued passports to travel within the European Union (EU). Authorities are continually trying to categorize mobile objects and individualize mobile subjects, and judgments are made about the movements of particular things.

When I fly to the USA for an academic conference, as a British citizen I am required to stand in an immigration queue designated for non US citizens. I am required to have a machine readable passport which can be swiped, and might soon be required to have one of the new generations of biometric passports featuring a computer chip which can store facial biometric data (including key facial measurements and characteristics), a digitally encoded photograph, and other personal information. And yet, I am afforded certain privileges not possessed by all non US citizens. As a citizen of one of 27 ‘developed’ (the majority Western European) countries, I do not need to apply for a visa. I simply queue up with my completed I 94W visa waiver form, answer a few questions, submit my passport and fingerprints for scanning and, hopefully, I am granted permission to enter for up to 90 days. Non Western travelers, and those deemed to be a moral, physical, or biological threat to the USA and its citizens, receive a quite different treatment, while throughout the world thousands of migrants scale fences, cross deserts, rivers, and seas, or travel in the back of lorries in an attempt to traverse national borders. People smuggling has emerged as a high profit activity for criminal gangs.

My experiences at US immigration checkpoints are predictable, occasional events, but my mobilities and mobile communications are governed and monitored in many other seemingly mundane but significant ways. My movements through the airport, as with the street, railway network, or shopping center, can be monitored or registered with closed circuit television (CCTV) cameras, pedestrian, or vehicle flow counters, at check in desks, ticket barriers or toll booths, and predicted and guided by architects with computer simulation models, computer aided design packages, and a variety of architectural and design features. Software and computing are increasingly being embedded into the built environment of Western countries. Credit card purchases and supermarket loyalty cards collect data on mobility and purchasing patterns, generating information which is highly valued by marketers and consumer profilers. Our mobilities are registered by the movements of our mobile phone, in addition to the calls we make. The movements and locations of our bodies, identification documents, financial transactions, and private technologies, then, are tracked by an array of government and corporate agencies. We ‘move’ and are ‘traced’ through transportation networks, border control points, electric cables, radio waves, and computer databases. But while the scale and capabilities of tracking technologies may be changing, authorities have long been concerned with tracking the movements of individuals and groups. The pervasiveness of computer, digital, and optical technologies which are used for tracking mobilities should not lead us to overlook a much longer history of ‘technologies’ for governing mobile citizens and maintaining extensive political and economic networks: from the history of customs agencies and imperial trading companies, to the history of passports, visas, and radar.

The mobilities and spaces of car driving in Britain are particularly instructive. Long before government agencies, transport engineers and the police utilized automatic speed cameras, CCTV cameras, number plate recognition technologies, and computer databases to monitor and regulate vehicle drivers, their movements have been identified, individualized, and regulated through the use of vehicle number plates, driving tests and licenses, codes of conduct, police speed traps, ‘sleeping policemen’ (speed control ramps), traffic lights, roundabouts, and a whole host of other technologies. Governments and the police have long recognized that to govern mobile vehicle drivers across a fairly extensive network of roads, they need to combine juridical and disciplinary techniques of government with liberal principles which encourage drivers to practice techniques of self government, controlling the movements of themselves and their vehicles. Experienced drivers appear to embody the skills and practice the etiquettes (both formal and informal) of driving in particular environments, performing the act of driving and governing their own movements and conduct in automatized, precognitive, and unreflexive ways. They may subtly resist or reinterpret the formal rules of the road, but the potentially fatal consequences of inattentiveness and rule breaking lead the majority of motorists to drive in more or less socially acceptable ways.

Mobility and Exclusion

By engaging in particular strategies or government programs, and deploying a broad range of technologies, individuals and authorities aim to exclude particular mobilities, activities, entities, and individuals from the public/private sphere, and minimize the risk of certain events occurring. The mobilities of illegal immigrants, drug smugglers, travelers infected with diseases such as SARS (severe acute respiratory syndrome), and suspected terrorists are approached with varying degrees of urgency and intensity by authorities, who ‘profile’ and ‘screen’ travelers and categorize those deemed to be ‘high risk’: whether this is due to their name, ethnicity, religion, nationality, travel itinerary, or their body temperature. This section focuses on the mobilities of two groups who are frequently excluded or criticized by political and cultural commentators: first, asylum seekers and economic migrants; and second, tramps, gypsies, and travelers.

In the UK, asylum seekers have been consistently constructed by the right wing and tabloid press as out of place: as unwelcome immigrants whose claims to refugee status are untested or unproven, and who are ‘scrounging from the state’. Asylum seekers are not yet proven to be ‘legitimate refugees’ (according to the 1951 UN Convention) entitled to protection from political persecution, and reactionary social commentators frequently label them as ‘bogus asylum seekers’. The stability of British society and Britain’s economy is presented as being under threat from such mobilities, and the dramatic and emotive language which journalists and others use draws upon metaphors of uncontrolled flows of fluids or the spread of infectious diseases. Headlines shout about 'floods', 'tides', and 'waves' of immigrants, and this language is not limited to the mobility of asylum seekers. With the accession of ten countries (including eight East European countries) to the EU in 2004, commentators focused their attention on the 600 000 or so individuals who came to the UK in the first two years (May 2004–June 2006). Tabloid newspapers suggested that 'the tide' or 'surge' of EU migrants needed halting, as the British economy, health service, and Briton's jobs were said to be under threat. There is, of course, a long history of hostility to immigration and immigrant mobilities, in the UK and elsewhere. Prejudices and political and juridical discourses have gathered around the movements of people for hundreds of years, and entire academic subdisciplines and theoretical debates – including studies of migration, colonialism and postcolonialism, diasporic cultures, and Orientalism – focus their sights on the social, economic, cultural, and political imagination and effects of migration and travel. National histories are crosscut with narratives and memories of immigrant and emigrant mobilities, and the diasporic and hybrid cultures which emerge are frequently defined by, or seen to be practiced through, those very mobilities and their spaces. We are urged to trace the 'routes' shaping cultural formations, rather than searching for the 'roots' of cultures and identities. Attention is focusing on the 'Atlantic geographies' and 'Black Atlantic' cultures weaving North America, Europe, Africa, and beyond, demonstrating how mobilities are practiced and experienced in strikingly different ways. The histories and geographies of mobility and migration reflect complex power relations, and geographers have done a great deal to help expose such power relations.

Mobilities are frequently constructed as central to, and essential to, capitalist development and circulation, but it is also common for mobilities which do not appear to have a (legitimate) purpose or destination to be romanticized as a way to escape, or criticized as a threat to, the routines of sedentary societies. In the USA, mobility has long been central to both mainstream and countercultural imaginaries of the nation. Mobility was central to narratives of pioneering Westward expansion, and both road transport and railway travel (as well as shipping) became embroiled in accounts of nation formation. The road, in particular, has been portrayed as a space of freedom, escape, and resistance for a broad range of individuals: families fleeing the depression and poverty of the dustbowl states in the 1930s (as depicted in Dorethea Lange's photographs, or John Steinbeck's 1939 book *The Grapes of Wrath*); men escaping the responsibilities and mundane realities of domestic life in Jack Kerouac's 1957 book *On the Road*; or the two women escaping abusive relationships in the 1993 film *Thelma and Louise*. Such narratives are often highly gendered, and it has been argued that academics must not simply affirm or romanticize popular narratives which equate mobility with freedom and resistance. People stay put and resist. Others may have very little freedom or choice in their movements. What's more, there are a whole series of mobilities – for example, those of refugees and asylum seekers – which are more commonly criticized than romanticized.

As Tim Cresswell shows in *The Tramp in America* (2001), during the 1870s a moral panic gathered around the mobilities of the tramp. The completion of an extensive network of railway lines in the American West, coupled with a fluctuating economy and increasingly mobile workforce, acted to facilitate the movement of homeless and jobless men and women across America. In 1875, Josiah Flynt estimated that as many as 10 000 tramps rode the boxcars every night, and the mobilities of these tramps were presented as unpredictable, threatening, immoral, and pathological. A broad range of discourses gathered around the movements and actions of the tramp. Sociologists sought to define and understand the life world of the tramp. Laws were enacted to control the mobilities of tramps and vagrants, although such categorizations were conveniently applied to a broad range of individuals deemed to be 'out of place', enabling the police to arrest the unemployed and working class activists. Despite the presence of a significant number of female tramps and hobos, the tramp was frequently constructed as a male figure who posed a threat to women and the moral sanctity of the spaces of the home. The figure of the tramp was constituted as degenerate and pathological by eugenicists and medical experts. In film and cartoons, most famously *The Tramp* (1915), starring Charlie Chaplin, widespread stereotypes of tramps and hobos were used to illicit laughter, but unlike more official narratives Chaplin's character emerged as a hero representing resistance to authority. Cresswell's writings reveal how the tramp was 'made up' in a variety of ways, but while tramping was constructed as a modern problem in the USA in the 1870s, England and other countries had seen panics over vagrancy and homelessness as early as the late fifteenth century.

In the UK the mobilities of gypsies and travelers have long been constructed as a threat to society and capitalist spatiality. In contrast to romanticized stereotypes of 'traditional' Romany gypsies traversing the countryside in horse drawn wooden caravans, contemporary gypsies and travelers equipped with motor vehicles and modern caravans are frequently constructed as a polluting presence who threaten the planned order of a 'Middle English' countryside. Gypsies and travelers are presented as being out of place in the contemporary English countryside, and in 1994 Britain's Conservative government enacted a Criminal Justice and Public Order Bill which included clauses designed to control the mobilities and actions of gypsies – along with New Age Travelers, ravers, hunt saboteurs, and environmental campaigners. During discussions of the Bill, British Members of Parliament suggested that the legislation was not aimed at 'real gypsies', but at New Age Travelers who, in their view, had chosen to drop out of society, leave urban areas and paid employment, and adopt a mobile lifestyle. New Age Travelers were labeled as dirty scroungers who did not belong in the countryside, and the Act entitled police forces to set up road blocks and control the movements of groups of travelers, ravers, road protestors, and hunt saboteurs. As geographer David Sibley pointed out, the Criminal Justice and Public Order Bill was 'emphatically concerned with controlling movement' (1994, 301).

Mobility, Place, and Landscape

Mobility and Placelessness

Geography is frequently presented as a discipline concerned with places or landscapes, and it has been suggested that increasing levels of mobility might herald the 'end of geography' and the erosion of place. In his 1976 book *Place and Placelessness* the humanistic geographer Edward Relph suggested that increasing mobility had brought about the extension of placelessness in the twentieth century. The modern infrastructures of the airport, railway, and roads were seen to cut across the landscape, generating feelings of placelessness. In the mid 1990s, French anthropologist Marc Augé developed a similar argument in his book *Non Places* (1995). Augé argued that 'nonplaces' were characteristic of an era of 'supermodernity', with an excess of information, increasingly mediated social relations, a shrinkage of space, and people becoming increasingly closed off to the world. Nonplaces were seen to be fairly generic ubiquitous spaces of detachment and solitariness, and common examples include spaces of mobility such as motorways, airports, and hotel lobbies, as well as other spaces of modern consumption and circulation, like shopping centers, theme parks, and industrial estates. But the questions to ask here are: is it useful to approach such environments as nonplaces or as placeless? And do we need to delineate a new species of place – especially a negative term, that is, nonplace – to account for people's experiences or engagements when moving through such environments? Social scientists, including geographers, have mixed opinions on the usefulness of these terms. On the one hand, Augé and others do appear to articulate or refract the somewhat unreflexive and distracted ways in which some travelers move through such spaces, paying little attention to environments they inhabit and traverse, and forgetting their actions. Modern spaces of mobility are frequently designed to facilitate such journeys and transactions, and our familiarity with such environments and modes of travel brings about shifts in our senses of being and dwelling in travel, and shifts in the spatialities and phenomenologies of inhabiting and moving. On the other hand, a broad range of other academics have asserted that people inhabit such spaces in a diverse array of ways, and airports, motorways, and shopping centers are highly variegated, politicized, and policed spaces. Airports may be occupied by regular business travelers who reflect very little on their surroundings, but they are also inhabited by thousands of workers, first time travelers, and a not insignificant number of homeless people who authorities frequently attempt to exclude from such spaces.

Experiences of distraction, boredom, solitariness, and detachment are characteristics of all manner of environments and the production of many different places, despite the globalization and standardization of architectural styles and the fact that a large proportion of travelers may not actually reflect on or remember their surroundings. People may feel isolated or bored at home, at work, or in picturesque villages (which are frequently upheld as the archetypal, meaningful social place). What's more, while spaces of mobility may appear to be asocial environments where people rarely converse with one another, and where there is little sense of community, academics such as Marc Augé operate with rather narrow and conservative definitions of both 'the social' and 'place', overlooking the ways in which different mobilities may be conceptualized and enacted as practices of connection, and the importance of all manner of material technologies in social relations. Mobile telephones, letters, e mails, light houses, or indeed car horns, wing mirrors, and indicator lights, are integral to the social relations and communications of mobile actors, and mobilities are produced and performed through an array of more or less transitory, individualized and embodied social relations and associations with these and other things.

As an array of geographers such as Tim Cresswell, Kevin Hetherington, Doreen Massey, and Nigel Thrift have suggested, we need to rework our definitions of place, taking account of complex spatialities and temporalities, and the ongoing mobilities of people, nonhuman things, and information into and out of places. Rather than thinking of places as fixed, rooted, frozen in time and space, and closed off to the world, we need to think of places as relational, open to the world, and 'in process'. Places are constantly performed through the 'gathering' of materials and movements. It is useful, here, not just to think of 'places' as concretized 'nouns' or 'things', but to think about the ongoing, incessant 'placing' of the world, and the processes which lead to the emergence of both highly reflexive and mundane/backgrounded attachments, associations, and significations. An understanding of different mobilities and mobile processes can therefore highlight the importance of conceptualizing place as a noun and a verb. Places are continually undergoing processes of 'placing' and the same could be said of landscapes.

Landscape and Mobility

Landscape, mobility, and movement are closely, if subtly, entwined. Landscapes change, viewpoints shift, and people inhabit and move through their surroundings in different ways. Throughout the long history of esthetic engagements with – and commentaries on and in – the landscape, writers, landscape gardeners, painters, film makers, and academics have, to varying extents, taken seriously the ways in which people move through particular landscapes. Landscape designers have accounted for the viewpoints and esthetic experiences of those who walk or drive through landscapes. Artists, too, have engaged with the embodied movements and esthetic sensibilities of the mobile viewer, developing particular techniques for reflecting or expressing the dynamism and movement associated with particular landscape perspectives. J. M. W. Turner's atmospheric landscapes, Giacomo Balla's abstract futurist paintings of passing motor cars, Laszlo Moholy Nagy's experimental color traffic photographs, and Julian Opie's video installations of British motorways, to mention four quite different examples, use a range of artistic mediums and strategies to refract different senses of movement in the landscape. J. B. Jackson's writings on America's vernacular landscapes reflect a similar sensibility, as he attempted to articulate the viewpoint of the mobile observer who traversed the landscape by car, truck, or motorcycle. This genealogy of sensibilities to mobility, and the difference that mobilities make to understanding and encountering landscapes, is complex and varied, and in the past decade or so academics have advanced similar arguments in suggesting how geographers, anthropologists, and others should approach and theorize landscape. During the cultural turn of the 1980s, geographers writing about landscape drew increasingly upon techniques such as iconography and textual analysis, engaging with debates in the disciplines of art history, anthropology, literary studies, history, and cultural studies. Landscape was approached by geographers such as Denis Cosgrove and Stephen Daniels as being represented or textualized in different media and materials, and bound up with particular visual cultures and ways of seeing. They rarely interpreted landscape as being fixed or static (although they were often accused of this), and their work frequently articulated the ways in which people moved through the landscape, and how landscape representations were produced, circulated, encountered, recognized, and consumed in particular settings. Nevertheless, since the early 1990s anthropologists and archaeologists such as Barbara Bender, Tim Ingold, and Christopher Tilly have drawn upon phenomenological writings in an attempt to move away from a focus on representations of landscape. It has been stressed that landscapes are not 'static representations' but are in process, in movement, and in a state of becoming. In their attempts to comprehend particular fieldwork encounters they have focused on the diverse ways people dwell in, inhabit, and move through (especially walk through) the landscape. More recently, geographers such as John Wylie have examined the performative and affective dimensions of walking through the landscape, and throughout these academic engagements it is suggested that reflective accounts of our embodied mobilities provide more satisfactory non representational or processual ways of apprehending the ongoing constitution of both self and landscape. As with writings on place, then, academics are increasingly suggesting that we should approach landscapes as open and produced through particular mobilities and movements, paying attention to the embodied mobilities by which we inhabit, traverse, and view the landscape. As art historian W. J. T. Mitchell has suggested, we should consider landscape as a verb rather than (or as well as) a noun.

Travel and Transport

Geographers have held a long standing fascination with travel, and the history of the discipline is closely entwined with histories of travel and exploration. The majority of British undergraduate geography students will go on a field class and undertake fieldwork during their degrees, and physical or imaginative travel to other places appears to be a central concern of many geographers. As geographers have developed an increasing interest in the representational practices and embodied mobilities of their own discipline, so they also have been paying increasing attention to the social and cultural geographies of travel, travel writing, tourism, and transportation. Geographers have examined the imaginative geographies which are refracted and produced through travel writing, the geographies associated with different cultures of exploration, and they have developed a sustained interest in the environmental, social, economic, and cultural dimensions of tourism. This section provides a slightly different focus, examining the social and cultural geographies of transportation, and particularly the increasing attention paid by geographers and social scientists to 'automobilities', and the geographies of the car and road.

Transport, Society, and Space

Transportation forms part of many people's daily lives, whether we travel by motor car, bus, train, airplane, bicycle, foot, or consume commodities which have been moved by such methods. The past few centuries have seen massive changes in the capacity of many people to travel and communicate across great distances. In addition to the expansion of modern communications technologies such as the postal system, telegraph, telephone, and satellite technologies, the expansion of railway networks, the rise of the private motor car, and the growth of airline travel has helped to transform perceptions of space, time, speed, and distance. In Britain, improvements in coaching technologies and roads helped to increase stage coach speeds in the late eighteenth and early nineteenth centuries, while the railway construction boom of the 1830s and 1840s led to a massive reduction in journey times, and the emergence of distinctly new experiences and geographies. Britain's railway companies were among the nation's largest employers in the nineteenth century, and they facilitated the mobility of passengers and important commodities. The railways brought a new technology and vast engineering and architectural structures into Britain's rural and urban landscapes. Some found these spaces and technologies exciting, others found them frightening, while many first class passengers found the public nature of this mass mode of transport depersonalizing and at times embarrassing – as they were unaccustomed to sharing compartments with strangers. In an era characterized by locally measured clock time, the demand for accurate railway timetabling led to the adoption of a standard time across each company's network and eventually the entire nation. The railways became associated with a broader 'industrialization' of space and time, while the speed of railway travel led to the emergence of new ways of looking at the passing landscape which Wolfgang Schivelbusch has termed 'panoramic perception': where landscapes lose their depth, and passengers are forced to look or glance differently at the evanescent scenes scrolling by.

Transport technologies have had a significant impact on the geographies of work, leisure, and homelife. As urban areas have expanded so new forms of public transport such as the horse drawn tram and bus, electric tram, electric railways, underground railways, and motor bus have helped to speed up urban travel times, making it more practical and less onerous to commute to work. As Greater London expanded in the early twentieth century so the new deep level electric underground railway companies attempted to generate new business by promoting suburban and country living and the ease and economy of commuting. This was a time when an increasing number of people from urban areas were discovering rural areas, and this discovery was facilitated through increases in leisure time, and increasing numbers of tour guides. It also reflected the increasing popularity of walking, bicycling, train excursions, and (later) motoring. As urban areas expanded so railway and bus companies extended their networks out toward villages surrounding large cities such as London, facilitating travel across greater distances for leisure and business. Of course, it would be too simplistic to suggest that new transport technologies caused urban expansion or vice versa, but the close alignment of urban and transport expansion, and the new forms of urban and suburban development which arose in the twentieth century, fueled accounts which suggested that mobility and movement were threats to the particularities of place. The movements of cars, people, and commodities, to name three examples, have variously been seen to erode local distinctiveness, community cohesion, senses of place, and to generate feelings of placelessness. As suggested in the previous section on mobility, place, and landscape, such accounts are overly simplistic.

It is not only urban geographies which have been closely aligned with changes in transport technologies. The growth of railway travel, bicycling, and car travel have had far reaching effects on rural areas, positive and negative. At a very different scale, the geographies of empire have long been entwined with the geographies of transportation, from shipping and railways, to the motor car and air travel. New transport technologies facilitated the weaving together of empires, and such things as railway networks and modern roads have variously been seen to be essential to the governance, policing, and development of colonies. Shipping technologies were central to the geographies of colonial conquest, expansion, and trade, and they facilitated the forging of the 'Atlantic geographies' mentioned earlier, and the transportation of both voluntary and involuntary imperial travelers. Enslaved Africans faced the possibility of death and serious illness in the horrific conditions onboard ships traversing the 'middle passage' between Africa and the Americas. In the later nineteenth and early twentieth centuries shipping companies such as British India and P&O (Peninsular and Oriental Steam Navigation Company) transported mail, goods, and more fortunate imperial travelers between Britain and its empire. In the 1890s and early 1900s leisure cruises were becoming increasingly popular, while after World War I, P&O and other companies actively promoted their services to the thousands of hopeful emigrants seeking new lives in countries such as Australia.

In the 1920s and 1930s, commercial airlines like Britain's Imperial Airways and Dutch company KLM developed air routes linking their respective home nations with their empires, and with the growth in mass long distance air travel from the 1960s we see the rise of new Atlantic, Pacific, and global geographies. The growth in air travel has given rise to new geographies and geographical imaginations, altering people's perceptions of distance, and more recently becoming enmeshed in debates about emissions/global warming and terrorism. Western academics, students, or members of the public tend to take their mobilities for granted. We inhabit a variety of spaces of travel, transport, and mobility on a more or less regular basis, and we may not inhabit such spaces with any great excitement. We may see particular modes of transport as merely functional, as delivering us from A to B, and we may not reflect on the geographies or spatialities of our movements. Nevertheless, an understanding of the mobilities and transportation of a range of people, ideas, and things is integral to our understanding of the geographies of many more seemingly stable processes and fixed locations. Mobility and stasis, fixity and fluidity, are relative, relational, and variegated conceptions.

New transport and communication technologies are aiding the development of new practices and forms of mobility, but it is important not only to remember that many people do not have access to new technologies and forms of mobility, but that many older forms of mobility and communication are still widely practiced (including the postal service, train travel, walking, and cycling). Walking is still a vital form of mobility for the vast majority of people, not only on a micro scale 'in all societies' (moving around the house, workplace, the local neighborhood), but for those without access to private cars or public transport in Western and non Western societies. In rural Africa the vast majority of trips are undertaken locally on foot, while many of those completed for domestic purposes – such as collecting water, fuel, and food – are carried out by women. Feminist scholars have undertaken extensive research on the gendering of such mobilities, exploring the embodied practices of women, the gendering of metaphors of mobility, gender differences in experiences of mobility, differentiated patterns of access, and much more.

Automobilities, or the Geographies of the Car

Transport is big business, and particular modes have attracted detailed attention from economists and environmentalists. Transport economists, engineers, psychologists, architects, and transport geographers have attempted to understand, model, and facilitate the movement of people and goods through transport networks, but it is only more recently that sociologists have begun to pay attention to the more qualitative social dimensions of transport and travel. One key literature has been the critical writings which have emerged on 'automobilities', and specifically the sociologies of the car and driving.

Economic geographers, transport geographers, and environmental geographers have a fairly long standing interest in the motor car. They have traced the shift from Fordist to post Fordist modes of production. They have modeled the flows of motor cars along roads, and they have traced the growth in motor car ownership and use. Geographers have long recognized that motor vehicles matter to environmental geographies, and yet social and cultural geographers, along with sociologists, have arrived somewhat late in examining the social relations and practices surrounding one of the most significant material objects in the lives of several billion people worldwide. Cars and other motor vehicles matter, not just because of their resource use, emissions, production innovations, or significance as globalized commodities. Cars also matter in our social lives and cultural relations, and for over a century they have been enmeshed in complex globalized, nationalized, localized, gendered, sexualized, and racialized discourses of identity formation, desire, exclusion, and stereotyping. Indeed, one would argue that if environmentalists, politicians, and economists want to understand why many people appear to have an unrelenting addiction to driving and attachment to their cars, it is vital that they take seriously the embodied practices, social relations, ontological formulations, spatialities, and discourses of rights which are enacted when driving or owning cars.

Social scientists have been paying increasing attention to the complex and variegated social relations associated with driving and the spaces of the road. Sociologists have stressed that although motor vehicles appear to provide an unrivaled capacity to facilitate autonomous mobility (or automobility), a driver's movements are contingent upon an extensive network of actors responsible for manufacturing, servicing, fueling, and regulating vehicles, as well as planning, constructing, and maintaining roads. Motor vehicles appear to be popular and unrivaled because they combine a sense of autonomous and independent mobility, with feelings of privacy, power, and speed. When stepping into their vehicles humans become something else, and a number of sociologists have examined how car drivers are somewhat hybrid or cyborg figures, whose movements, spatialities, and very being emerges through the joint action of people in cars. What's more, when they are fully trained, gain confidence, and have become familiar with their vehicles, drivers perform the act of driving in relatively automatic, unconscious, unreflexive ways. As with actions such as typing, drivers appear to embody their driving knowledge and skills and perform these actions in relatively unthinking and effortless ways, while the very process of driving gives rise to detached engagements with one's surroundings and a perpetual forgetting of one's actions and movements through the landscape. Driving entails rather disengaged, distracted, or unreflexive modes of attention. And yet despite this somewhat detached involvement, a broad range of academics have shown how driving entails significant affective engagements, producing distinctive emotional geographies associated with intense feelings of anger, excitement, and fear. The design of motor vehicles creates limits to how drivers can express their intentions and emotions, and drivers often get frustrated or angry at the apparent obliviousness, deafness, or blindness of other motorists to their actions and feelings.

The emergence of hybrid beings and ontological formations such as the car driver is intimately associated with the distinctive spatialities of the car, road, and driving. Drivers must develop an awareness of what is going on inside, alongside, in front of, and behind their vehicle, judging the width of their vehicle, its capacity to break and accelerate, glancing in mirrors, and expressing their intentions using indicator lights. Drivers simul taneously inhabit the public and private realm, occupying a personalized and privatized space (often infused with the personal soundscape of one's music collection), while at the same time traversing the public highways and engaging in actions which have major consequences to oneself and others. Motor vehicles have had a major impact on the geographies and sociologies of the road, for with the domination of many urban roads by petrol and diesel driven vehicles, other road users such as pedes trians, cyclists, and indeed animals have gradually been forced on to the verge or specially constructed pave ments/sidewalks. Organizations such as the British anti car group Reclaim the Streets have argued that this re sults in the privatization of an important public space. Indeed, urban geographers and sociologists are beginning to recognize that mobilities more generally, and cars in particular, are a fundamental part of the life (and, per haps, death, in the case of cars) of cities, with the flow of all manner of things across their boundaries. We often encounter cities when moving through them – at, below, or above ground level. We engage with buildings kinesthetically and visually through our everyday mobi lities: walking through buildings and driving past land marks. The proximity of homes or office sites to underground stations or major roads can impact upon their value or perceived development potential. Think about the strategic location of billboards at major road junctions, or the impact of the diagrammatic plan of London's tube network upon people's imaginations of that city. Different mobilities are fundamental to how cities function, and how they are encountered as vibrant, energetic, exciting, congested, polluted, or scary places.

In the past, academics have attempted to discern globalized patterns in the consumption of cars, but there are distinctively localized and nationalized geographies to the consumption of vehicles and the spaces of the road. Anthropologists, sociologists, and geographers have shown how people drive, inhabit, and consume their vehicles differently in different societies and cultures. For example, the Pitjantjatjara aboriginal peoples of South Australia have incorporated cars into their material cul ture and cultural practices in specific ways, using, valuing, exchanging, abandoning, recycling, and navigating them in ways that may appear unfamiliar to Western observers. Cars, as material objects and spaces, become embroiled in highly distinctive social relations, rituals, and practices of mobility and consumption.

Social scientists may have been rather slow in exam ining the multifarious ways in which cars have become embroiled in our lives, but commercial organizations and cultural commentators have long recognized the signifi cance of the motor car in our cultural lives. Novelists, playwrights, photographers, artists, pop musicians, film makers, and toy manufacturers have engaged with humanity's fascination with driving and cars. Hundreds of films (some of which may be classified as road movies) have explored the activities of driving and the con sequences of automobilities. In literature, writers have explored the automotive passions that are expressed by a wide range of fictional characters, from the reckless ar istocratic Toad in Kenneth Grahame's *The Wind in the Willows* (1908), to James Ballard and Vaughan searching for their next car crash in J. G. Ballard's *Crash* (1973). As these different examples highlight, motor vehicles (es pecially cars) are enmeshed in our worlds in multiple ways, giving rise to economic, social, cultural, political, and environmental impacts and effects.

Conclusions

Mobility and movement are a fundamental part of our ongoing engagements with the world, and these mobilities are being explored by academics working across the hu manities and social sciences, as well as the physical, nat ural, and medical sciences. Human geographers have held a long standing interest in different mobilities, focusing on geographical patterns of migration, transportation, and tourism, but in the last decade or so there has been a resurgence of writings on mobility in human geography and across the social sciences. With the increasing pur chase of anti essentialist, post structuralist thinking, par ticipative and performative methods, and the emergence of an interdisciplinary interest in themes such as global ization, transnational migration, diasporic cultures, com munication technologies, and the geographies of the car and airport, so the topic of mobility has assumed a more prominent position in a wide range of international, interdisciplinary debates and agendas. Mobilities have been celebrated for their transgressive potential, but social scientists have argued that we must not see a language of mobility, flux, and movement as a simple solution to overcome sedentarist thinking. Instead, we should attempt to reveal the complex politics of mobility, examining how different mobilities are produced, practiced, and regulated. Mobilities and materialities are closely entwined, and mobilities rework places and landscapes on an ongoing and incessant basis. Mobilities are important to human geography and our human geographies, and the flows, circulation, and mobilities of all manner of things will no doubt appear and reappear throughout this encyclopedia.

See also: Distance; Fluidity-Fixity; Landscape; Migration; Migration, Historical Geographies of; Non-Representational Theory/Non-Representational Geographies; Place; Tourism; Trade, Transport and Communications, Historical Geographies of; Transport and Social Exclusion; Travel and Travel-Writing.

Further Reading

Augé, M. (1995). *Non Places: Introduction to an Anthropology of Supermodernity*. London: Verso.
Bohm, S., Jones, C., Land, C. and Paterson, M. (eds.) (2006). *Against Automobility*. Oxford: Blackwell Publishing/Sociological Review.
Cresswell, T. (1993). Mobility as resistance: A geographical reading of Kerouac's "On the Road". *Transactions of the Institute of British Geographers* 18, 249 262.
Cresswell, T. (2001). The production of mobilities. *New Formations* 43, 11 25.
Cresswell, T. (2001). *The Tramp in America*. London: Reaktion.
Cresswell, T. (2006). *On the Move: Mobility in the Modern Western World*. London: Routledge.
Cresswell, T. and Merriman, P. (eds.) (2009). *Mobilities: Practices, Spaces, Subjects*. Aldershot: Ashgate.
Featherstone, M., Thrift, N. and Urry, J. (eds.) (2005). *Automobilities*. London: Sage.
Kaplan, C. (1996). *Questions of Travel*. Durham, NC: Duke University Press.
Law, R. (1999). Beyond 'women and transport': Towards new geographies of gender and daily mobility. *Progress in Human Geography* 23, 567 588.
Massey, D. (1991). A global sense of place. *Marxism Today* June, 24 29.
Merriman, P. (2004). Driving places: Marc Augé, non places and the geographies of England's M1 motorway. *Theory, Culture, and Society* 21(4 5), 145 167.
Merriman, P. (2007). *Driving Spaces: A Cultural Historical Geography of England's M1 Motorway*. Oxford: Blackwell Publishing.
Pooley, C., Turnbull, J. and Adams, M. (2005). *A Mobile Century? Changes in Everyday Mobility in Britain in the Twentieth Century*. Aldershot: Ashgate.
Relph, E. (1976). *Place and Placelessness*. London: Pion.
Schivelbusch, W. (1986). *The Railway Journey: The Industrialisation of Time and Space in the 19th Century*. Oxford: Berg.
Sheller, M. and Urry, J. (2000). The city and the car. *International Journal of Urban and Regional Research* 24, 727 757.
Sheller, M. and Urry, J. (2006). The new mobilities paradigm. *Environment and Planning A* 38, 207 226.
Sibley, D. (1994). The sin of transgression. *Area* 26, 300 303.
Thrift, N. (1996). *Spatial Formations*. London: Sage.
Urry, J. (2000). *Sociology beyond Societies: Mobilities for the Twenty First Century*. London: Routledge.
Urry, J. (2002). Mobility and proximity. *Sociology* 36, 255 274.
Urry, J. (2007). *Mobilities*. Cambridge: Polity.
Wylie, J. (2005). A single day's walking: Narrating self and landscape on the South West Coast Path. *Transactions of the Institute of British Geographers* 30, 234 247.
Young, D. (2001). The life and death of cars: Private vehicles on the Pitjantjatjara lands, South Australia. In Miller, D. (ed.) *Car Cultures*, pp 35 57. Oxford: Berg.

Relevant Websites

http://www.tandf.co.uk
Mobilities (three times a year), Routledge: ISSN 1745 0101.

Mobility, History of Everyday

C. G. Pooley, Lancaster University, Lancaster, UK

Glossary

Automobility The dominance of the motor car in twenty-first-century society, not only for transport, but also as a social, economic, and cultural icon.

Everyday Mobility The routine journeys that most people undertake on a daily basis, including travel to school, to work, to shop to visit friends, and for leisure.

Mobility Transition The notion that all countries will move through a transition from low to high levels of daily mobility and residential migration, in parallel with changes in the demographic transition.

New Mobility Paradigm The concept that in twenty-first-century Britain mobility has itself become a significant activity rather than simply a means to travel from place to place.

Path Dependency The concept that certain key technological, institutional, or other developments have long-lasting effects shaping later developments and hindering further innovation.

Space–Time Pathways The ways in which individuals tend to travel over specific routes, regulated by external constraints, and often mutually interdependent.

Virtual Mobility The concept that development of the Internet, mobile phones, and related mobile communications allow people to interact and be 'mobile' without physical movement or face-to-face contact.

Definitions

Mobility is central to almost all activities and few people do not engage in some aspects of everyday mobility. In all parts of the world, children travel to school; adults commute to work; most people travel to visit friends and relatives, to shop, to participate in sports or entertainments, or to go on a holiday. Most of the time such movements are a taken for granted part of everyday life but, possibly for this reason, the history of everyday mobility has received much less attention than other aspects of population movement such as long distance migration. However, it can be argued that everyday mobility should not be viewed as a distinct activity, separate from residential migration, but rather as part of a mobility continuum that ranges from short distance everyday moves to international migration and virtual mobilities (**Figure 1**). Moreover, it can also be suggested that there is increasing interaction, and blurring of distinctions, between these different types of mobility. Thus, for instance, greater mobility opportunities in the twentieth century allowed at least some people to exchange commuting for residential migration, as they preferred to travel longer on a daily basis rather than move to a house closer to work. In the later twentieth century residential migration to live in a desired school catchment, thus influencing the journey to school, has also been common in some locations. In both the past and the present, some people effectively live in two (or more) places and thus the concept of 'home', and the journey from home to work, becomes indistinct. For instance, seasonal migrants may live and work in a variety of places; students live in the parental home and at college; and children whose parents live apart may have different homes, and hence different journeys to school, depending on which parent

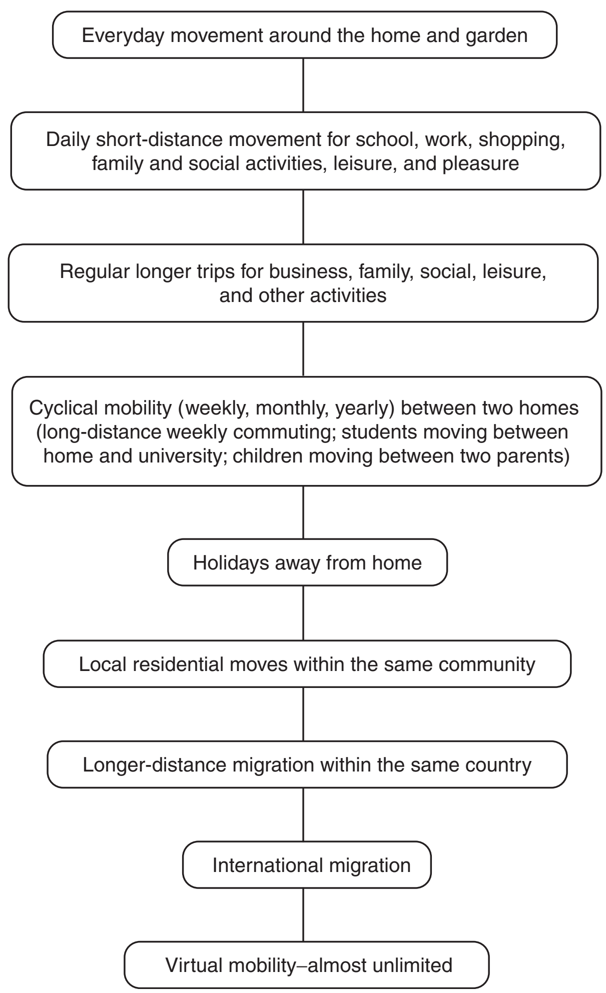

Figure 1 The mobility continuum.

they are living with at the time. In addition, in the twenty first century many activities that previously re quired mobility can now be undertaken without leaving the computer on your desk. Thus, virtual mobility allows contact with relatives on the other side of the world, the purchase of goods on line, and the acquisition of infor mation that previously would have required a visit to a library. In these senses concepts of mobility are mutable and everyday mobility should be seen as part of a much larger set of interacting processes.

Why Is Mobility Important?

At its most basic everyday mobility is the process that allows society, economy, and family life to function. It is the glue that holds everything together. In this sense nei ther its significance, nor its relevance to human geography, need further justification. However, it can also be suggested that mobility has a number of costs which further increase the significance of everyday travel beyond its role as a process that allows society to function. At least five sets of mobility costs can be identified (**Figure 2**). Everyday mobility imposes costs on individuals in terms of the time spent traveling, the implications of this for personal health and well being (through, for instance, stress and exposure to pollution), and the consequences of disrupted family routines with often complicated arrangements for child care or meals. High levels of mobility also impose costs on society, culture, and economy. The economic costs may be seen in loss of effective working time due to long com muting journeys, or in impacts on healthcare provision through work related stress. Lack of leisure time may impact on both individual quality of life and societal co hesion as people feel unable to participate in neighbor hood activities. The environmental costs of mobility are obvious, especially in car dominated cultures with high

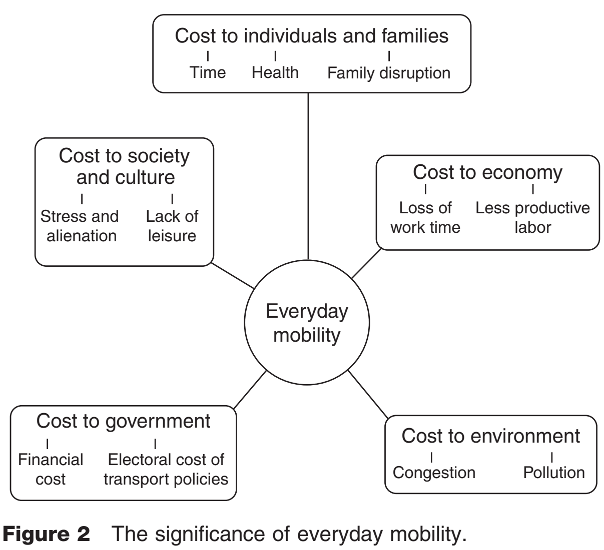

Figure 2 The significance of everyday mobility.

levels of congestion and pollution, while governments have to deal with the political costs through impacts on the exchequer and the electoral costs of transport planning decisions that may be unpopular with some. For instance, restrictions on motor vehicles may please environmental ists but anger many individual motorists. Thus, it can be suggested that everyday mobility, however and wherever it is undertaken, is likely to not only have significant impacts on and relevance to individuals and their families, but also wider implications for governments and other institutions that seek to regulate, in various ways, aspects of the economy, society, and environment.

Theorizing Mobility

Traditionally, studies of mobility have operated within quite confined disciplinary spaces. For instance, research on the engineering and technological aspects of travel and transport has rarely interacted directly with cultural and sociological studies of travel behaviors. Geographers have contributed to this debate only at the margins, fo cusing principally on mobility as a spatial process, and with more interest in the implications of movement ra ther than in travel itself. In recent years, however, these barriers have begun to disappear and there has been increasingly fruitful interaction between mobility re searchers from a variety of disciplines. This section highlights just a selection of the approaches that have been used. Much research on mobility emphasizes change: that levels of mobility – not only of people but also of goods, services, ideas, and information – have increased dramatically over the past few decades and that there is an almost inevitable link between increased mobility and high levels of economic development. As long ago as 1971, the American geographer Wilbur Zelinsky encapsulated some of these ideas in his ‘Hy pothesis of the mobility transition’ which argued that as societies ‘modernized’, they experienced higher levels of both residential migration and population circulation (everyday mobility). In common with other stage models of ‘development’, Zelinsky’s ideas have attracted much criticism, but this assumption of almost inevitable change toward ever higher levels of mobility (sometimes termed hypermobility) remains both pervasive and persuasive. Geographers have also contributed significantly to mo bility studies through their work on space and time. In the 1970s, the Swedish geographer Torsten Hägerstrand developed his influential ideas about space–time path ways, and such ideas have remained especially influential in studies of space, place, and mobility. More recent re search on mobility in geography has drawn heavily on theories developed in sociology and cultural studies. For instance, the works of Bourdieu and de Certeau have been used to explore the ways in which everyday mobility links to performativity. Every time we step outside our front door, we consider how we look, and how we might be viewed, and thus commonplace jour neys can become a performance that is embedded in everyday life.

While much research on transport has utilized an en gineering approach, focusing on technological solutions and modal splits, mobility researchers have focused mainly on the ways in which particular technologies, es pecially the motor car, have achieved dominance leading to cultures of 'automobility'. Concepts of path dependency have been used to explain how and why the automobile became such a dominant feature of twentieth century travel, with increasing emphasis not only on the en vironmental consequences of car dependence, but also on the cultural implications as the automobile itself becomes a symbol of social life. For instance, it has been argued that the car has become much more than a means of transport, and is now a possession and space that can be personalized and utilized to make statements about the self. Mobility researchers have also engaged directly with discourses on risk. It can be argued that high levels of mobility con tribute to perceptions of risk at a variety of levels. These range from the environmental damage done by emissions from cars, planes, and other modes of transport to research on individual and societal perceptions of risk that have increasingly restricted mobility for some groups that are perceived as vulnerable. Thus, despite increased mobility opportunities, it is argued that young children find their freedom of movement more restricted than in the past. It has been suggested by John Urry that, in the twenty first century, mobility is at the core of society and he posits a 'new mobility paradigm' which suggests that rather than travel being a means to an end, mobility has itself become the significant activity. Travel itself (for instance going for a car ride) can be a leisure pursuit; many people use travel time to work productively; and new mobile communi cations enable people to continue most everyday activities while on the move. Moreover, virtual space allows people to 'travel' without actually moving. Thus, functional and symbolic aspects of corporeal travel and mobility activities become increasingly intertwined, with mobility (either physical or virtual) central to this process. Although none of these trends are entirely new, it is usually argued that all have increased markedly over the last century and with a notable acceleration from the late twentieth century.

Global Geographies of Mobility

It is usually argued that globalization has made distant places more similar and has led to the dominance of Western cultures and technologies around the globe. What, then, are the implications of globalization for mobility experiences in different parts of the world? Certainly, there are clear global influences. The automobile is increasingly dominant and desired in all parts of the world, mobility potential has undoubtedly increased almost everywhere, the environmental implications of, especially, increased air travel certainly have global implications, and new mobile communications are available in all parts of the world. However, it can be suggested that there are two factors that may mediate these trends. First, in the context of everyday mobility – rather than, for instance, long distance migra tion – it can be suggested that processes of globalization and technological change have rather less significance. For most people, in most societies and at most times in at least the recent past, the majority of everyday activities are undertaken close to home and do not require sophisticated or new forms of mobility. For instance, in twenty first century Europe and Africa, most young children attend a school relatively close to their home and, even in a car dependent country such as Britain, walking remains the single most important means of traveling to school for primary school children. Longer distance trips remain relatively rare and, even in rich European countries, most people will only use air travel once or twice a year (and some never fly). Thus, it can be suggested that for everyday mobility there are fewer inherent differences over time and space than (for example) with regard to residential migration, and thus globalization processes have had less opportunity to generate change. Second, it can be argued that despite globalization trends there remain massive global inequalities in access to everyday transport. While it is assumed that in the West most people can gain access to a car if they wish (though some may choose not to for a variety of reasons), in many poorer parts of the world the automobile, although often privileged by planners, remains accessible to only a tiny minority of the population. Throughout much of Africa and parts of Asia and South America, the usual means of travel for most people are on foot, by bicycle or on public transport. The degree to which globalized and Westernized forms of travel have become truly universal should not be overstated, and there is little evidence that these inequalities in the global geography of everyday travel are going to change rapidly.

Change Over Time

Generalization about long term changes in everyday travel is difficult due to a lack of detailed data. Because everyday mobility is usually viewed as a mundane activity, and because it has relatively few long term impacts, it has rarely been systematically recorded even in rich countries of the world. Apart from in the recent past, statistics for all countries are hard to find and rarely reliable. More use fully, it is suggested that the history of everyday travel can be conceived as passing through four phases that represent different approaches to mobility. These are not meant to be 'developmental' in the conventional sense, and there is certainly no suggestion that all countries will encounter all stages, or that they will necessarily follow a particular order. They are arranged in a broad chronology that fits with the history of everyday travel for most people in many developed countries of the world, but in most times and places more than one phase could occur concurrently as different people had varied mobility experiences.

For much of human history, everyday transport has been organic, based on human and animal power. In many poorer parts of the world walking, horses or oxen still provide much everyday transport. During the second half of the nineteenth century in Europe, relatively low technology forms of transport (trains, trams, buses, and bicycles) became progressively more widespread, but due to cost constraints were only rarely utilized by the poor. Faster and more private forms of transport (especially cars) remained the preserve of the elite. However, from the 1920s widespread use of both cycles (especially by men) and public transport was common in European cities. During the twentieth century, and especially after about 1950, a combination of wider access to existing forms of transport (especially the motor car) and increased availability of newer technologies (especially planes) increased mobility potential enormously in most rich countries of the world. However, in poorer countries, although such transport is available it remains restricted to a minority. It can be argued that we may now be entering a fourth phase of green or virtual mobility, as concerns about the environmental and resource impacts of high levels of mobility may lead both to the implementation of new technologies and/or the replacement of physical movement by virtual movement. However, although both are currently possible, as yet there is little real evidence of significant change in mobility behavior. One enduring feature of everyday travel is the relatively unchanging nature of mobility inequalities. In most time periods and most parts of the world, some can move more freely than others. In general, mobility is easiest for those on high incomes, and those living in richer countries and/or in urban areas. In most countries and time periods women have experienced more mobility disadvantage than men, while both the young and the elderly are likely to have restricted mobility. It is argued that, although there have been massive changes in transport technology, and in access to mobility opportunities, over the past century, many of the core characteristics of everyday mobility have changed relatively little.

Modal Change

In the rich countries of the world there has been an undeniable decline in the use of walking as a means of transport. Indeed, it can be argued that one of the reasons why pedestrians are rarely catered for in transport planning is because walking is now viewed principally as a leisure activity rather than a means of transport, and thus it is seen to be of secondary importance. Thus, in Britain personal travel on foot fell from 34.8% of all trips in 1975/76 to 23.5% in 2005, and travel to school on foot by children aged 5–10 fell from 73.5% to 49% over the same period. However, the same statistics can be viewed from another perspective. In Britain in 2005, 70% of all journeys of less than 1 mile were on foot, and one third of all travel by children was on foot. A recent survey of everyday mobility in Manchester confirmed the continued importance of walking for children's travel, with 50% of travel to school by children aged 10/11 undertaken on foot in 2002/3 compared to 76% in the 1940s. It can thus be argued that, despite a decline in the absolute amount of travel undertaken on foot, walking continues to be an important mode of everyday travel. In most of the poorest countries of the world, walking remains the only available means of transport for many people in both urban and rural areas. With very low rates of car ownership, most long trips are undertaken by public transport and thus walking is usually also a secondary component of these journeys. Thus, in cities such as Dar es Salaam and Nairobi walking accounted for around 46% of all trips, and in rural areas dependence on walking even for quite long journeys is even greater. However, despite the numerical importance of travel on foot, transport policies have often mirrored those of the Western world, giving prominence to road and other transport schemes that favor the urban elite, and perpetuate mobility inequalities.

Cycling and walking are often linked together as accessible, cheap, and environmentally friendly travel modes. They also have a shared experience of marginalization by motorized transport. However, in reality cycling and walking are very different. Most importantly, whereas most people without mobility impairment can walk, a bicycle is a moderately sophisticated item of engineering that by no means all can or want to use. Indeed, it can be suggested that historically use of the bicycle has more in common with the evolution of car use than it does with walking. In Europe and America bicycles developed as leisure vehicles for the rich in the late nineteenth century, but by the early twentieth century were cheap enough to become widely used for much everyday mobility, especially by men. Thus, in Germany bicycle use peaked in the late 1920s, with the cycle accounting for 15–40% of the modal split in most German cities in the 1930s, rising to 60% in Berlin. Similarly, in Britain in the 1930s and 1940s, some 20% of journeys to work were undertaken by bicycle and the cycle was the single most important form of transport for men traveling to work. In the second half of the twentieth century, bicycle use fell dramatically in most European cities with bikes accounting for only 5% of travel in Germany by the late 1960s, and for just 1.3% of all trips in Britain in 2005. However, in the Netherlands cycling still accounts for some 6.6% of traveler km compared to just 0.6% in the UK. In poorer countries of the world cycling has played, and to a large extent still does play, a much more significant role in urban transport. Shanghai has been cited as the cycling capital of the world, and bicycles form an important mode of transport for short to medium journeys throughout China. Indian cities also record very high rates of cycling, and the bicycle accounted for over 38% of trips in Delhi in 1999. In many ways the situation in Chinese and Indian cities is not unlike that in much of Europe in the 1920s and 1930s. Bicycles are a preferred means of transport because they are relatively cheap and allow rapid movement around a congested city. However, as in Europe, the position of cyclists is fragile. Despite low rates of car ownership, the streets of Asian cities are congested and cycling is dangerous. Moreover, cycling is often seen as an outdated mode of transport, and city planners and many cyclists themselves equate wealth and development with the acquisition of a motor cycle or car.

National statistics clearly demonstrate a decline in the use of public transport in most rich countries in the world with, over the past half century, private car traffic growing almost equally at the expense of both walking and public transport. Thus, whereas in 1966, 40.3% of all journeys in Britain were undertaken by bus or train, by 2005 this had fallen to just 8.2%. The decline in public transport use has been especially marked in the USA: although 38.8% of work trips are made on public trans port systems in Europe, in the USA the figure is just 9%. In most countries the decline in public transport use has been especially marked in rural areas. Public transport plays a much larger role in everyday mobility in most poor countries of the world. This is clearly demonstrated in India. In Mumbai motorized public transport modes account for 40.9% of all travel, compared to just 9.3% by motorized private travel, and in Chennai (Tamil Nadu) 44.5% of trips are by the transit system compared to 8.8% in private motor vehicles. In contrast, in Melbourne (Australia) motorized private travel modes accounted for 73.8% of travel compared to just 7.1% by motorized public modes. However, despite the obvious importance of public transport in the cities of developing countries, it can be argued that this transport mode is facing similar pressures to those seen in the richer west. As people become more affluent then buses are deserted in favor of more comfortable taxis or, in some cases, private cars, and long established (if sometimes poorly managed) public transport systems are increasingly being threatened by Western style transport planning schemes that see fixed route systems in particular as obsolete, and which pri oritize private motorized transport for a minority.

The extent to which private motor transport has come to dominate everyday mobility in almost all richer countries of the world, and in many urban areas in poorer countries, can hardly be overstated. In the USA motor ized private transport accounts for 88.5% of all trips; in Australia and New Zealand it is 79.1%; in the Middle East 55.9%; Western Europe 49.7%; High Income Asia 41.6%; in Low Income Asia and Latin America around 35%; in Africa 32.3%; and in Eastern Europe 26.8%. In all areas car dependence has increased over the past few decades. Many studies have demonstrated the degree to which once a car is owned people are very reluctant not to use it. There is also a strong opposition to schemes which impose financial penalties on car use. In Britain such attitudes can be identified from about the 1950s when car use began to change from being mainly a leisure activity to an everyday means of travel to work and for other personal business. Today such attitudes are even more firmly ingrained, and are expressed in op position to policies such as congestion charging, road pricing, and workplace parking charges. However, there is also extensive evidence that the car does not necessarily bring improved access and mobility to all, that where policies to restrict car use and to provide alternative modes of transport are put into place, the dominance of the car can be challenged without restricting personal mobility; and that increased incomes do not necessarily lead to greater vehicle use.

Conclusions

It can be argued that there is a strong sense of path dependency in the history of everyday mobility. Once certain key modes of transport become established they acquire such economic, social, and cultural significance that it is very difficult to change travel behavior without dramatic intervention. This has certainly been the case with the motor vehicle where, in all parts of the world, but to varying degrees, automobility has become the dominant aspiration of most people, and provision for motor vehicles has been a major plank of transport policies. However, a historical perspective on everyday mobility also suggests some other scenarios. First, examination of changes in everyday mobility over time shows that although mobility potential has increased dramatically, and most people do undertake more long journeys than used to be the case, for many people, for most of the time, everyday travel (for instance to school or work, to shop or visit friends) is relatively short distance and repetitive. In this sense the structure of everyday travel has changed much less than might be expected. Second, where changes in travel behavior have been forced on people (for instance the withdrawal of tram systems in British cities in the 1940s), most travelers very readily shifted to other forms of transport. Thus, it can be argued that most everyday travel remains short distance, predictable, and adaptable; and that it may be possible to reduce car dependency for short trips in urban areas without significantly altering travel behavior.

See also: Historical Geographies, Urban; Industry, Historical Geographies of; Mobility; Transport, Public.

Further Reading

Adams, J. (2005). The limits to integration: Hypermobility, a challenge to governance. In Lyall, C. & Tait, J. (eds.) *New Modes of Governance: Developing an Integrated Policy Approach to Science, Technology, Risk and the Environment*, pp 123 138. Aldershot: Ashgate.

de Certeau, M. (1984). *The Practice of Everyday Life*. Berkley, CA: University of California Press.

Cresswell, T. (2001). The production of mobilities. *New Formations* 43, 4 28.

Featherstone, M., Thrift, N. and Urry, J. (eds.) (2004). *Special Issue: Automobilities. Theory, Culture and Society* 21(4 5), 1 284.

Hagerstrand, T. (1975). Space, time and human conditions. In Karlqvist, A., Lundqvist, L. & Snickars, F. (eds.) *Dynamic Allocation of Urban Space*, pp 3 14. Farnborough: Saxon House.

Hillman, M., Adams, J. and Whitelegg, J. (1990). *One False Move …: A Study of Children's Independent Mobility*. London: Policy Studies Institute.

Kenworthy, J. and Laube, F. (2002). Urban transport patterns in a global sample of cities and their linkages to transport infrastructure, land use, economics and environment. *World Transport Policy and Practice* 8, 5 19.

Kenyon, S., Lyons, G. and Rafferty, J. (2002). Transport and social exclusion: Investigating the possibility of promoting inclusion through virtual mobility. *Journal of Transport Geography* 10, 207 219.

Marshall, S. and Banister, D. (2000). Travel reduction strategies: Intentions and outcomes. *Transportation Research Part A* 34, 321 338.

Pooley, C., Turnbull, J. and Adams, M. (2005). *A Mobile Century?: Changes in Everyday Mobility in Britain in the Twentieth Century*. Aldershot: Ashgate.

Schafer, A. and Victor, D. (1997). The past and future of global mobility. *Scientific American.* October, 36 39.

Thrift, N. (1995). A hyperactive world. In Johnston, R., Taylor, P. & Watts, M. (eds.) *Geographies of Global Change: Remapping the World in the Late Twentieth Century*, pp 18 35. Oxford: Blackwell.

Urry, J. (2000). *Sociology Beyond Societies: Mobilities for the Twenty First Century*. London: Routledge.

Whitelegg, J. (2003). *Earthscan Reader on World Transport Policy and Practice*. London: Earthscan.

Zelinsky, W. (1971). The hypothesis of the mobility transition. *Geographical Review* 61, 219 249.

Relevant Websites

http://www.lancs.ac.uk
Centre for Mobilities Research.

http://www.t2m.org/en/
International Association for the History of Transport, Traffic & Mobility.

http://www.cosmobilities.net/
The Cosmobilities Network.

http://www.dft.gov.uk/
UK Department for Transport.

Modern City

J. R. Gold, Oxford Brookes University, Oxford, UK

Glossary

Human Ecology The area of ecological inquiry that deals with the interactions between human beings and their environments.

Modernism A shared international tendency within the arts that emerged in the last decade of the nineteenth century and the first three decades of the twentieth century, which sought to express the spirit of modernity through radically different forms of representation that often challenged established practices.

Modernity Quality or condition of being modern, the sense of living in modern times.

Modernization The process of social transformation induced by the switch from agrarian to industrial modes of production.

Urbanization The relative growth of the population living in towns as opposed to rural areas.

Introduction

Any use of the term 'modern city' needs careful definition. In one sense, it need not imply anything more than that which is current but the term often gains added potency from the epithet 'modern' signifying a sense of opposition. The modern city is not just that which is here and now but also something that provides a sharp contrast with what has gone before. In this sense, labeling a city as 'modern' reflects the fact that cities are normally the places where economic, social, and cultural innovations first appear. Most of these innovations are passing fashions that quickly fade, but others can constitute pervasive new trends that bring far reaching changes to both the city and the wider society.

Judgments about the nature and consequences of these changes reflect the attitudes that observers adopt toward 'modernity' – the sense of living in modern times. Some have viewed modernity in a positive light, stressing the benefits of abandoning the constraints of the past, liberating progressive thought, and fostering creativity. Others interpret modernity pessimistically, identifying emerging trends as likely to alienate the city's residents and undermine the quality of life. Judgments about modernity, of course, can vary over time. The typical changes brought by city center reconstruction in the 1960s, for example, were overwhelmingly identified at the time as the march of progress and as bringing much needed order to urban chaos. Two decades later, the same developments were roundly condemned for misunderstanding how cities function, as abandoning fundamental values, and as being motivated by commercial greed and bureaucratic arrogance.

This sense of ambiguity underpins both principal uses of the term modern city as it is employed here. The first sees the modern city as synonymous with the 'industrial city', relating the term to a period lasting from roughly the 1820s to the 1970s, in which the city was reshaped by the modernizing influences that occurred in the wake of demographic and technological change. The second sees the modern city as a late stage in the development of the industrial city, coterminous with the rise of modernism in architecture and planning and, therefore, restricted to the three decades after World War II.

The Industrial City

Formulating an understanding of the industrial city necessarily invites comparison with its preindustrial counterpart. The latter was an outpost in a sea of rurality. Typically walled, with narrow streets and tightly packed buildings, it maintained close links with the countryside, notably through ties of landownership and by serving as a marketplace. The preindustrial city was socially stratified, with power concentrated in the hands of ruling families or occupational groups (especially guilds), who were themselves associated with long established areas ('quarters') of the city. Its craft economy saw trades clustered in specific quarters but there was no functional differentiation in land use, with an intermingling of houses, shops, workshops, and markets.

The advent of industrialization, rapid demographic increase, and urbanization (the relative growth of the population living in towns as opposed to rural areas) brought irrevocable change. The city became the prime forum for modernization – the process of social transformation induced by the switch from agrarian to industrial, especially mechanized, modes of production. The loosening of the class boundaries that typified the preindustrial city allowed new networks of social relationships, derived from occupational specialization of labor, to flourish. In these circumstances, the way of life for the working classes became dictated by the factory clock, with people taking their identity from their place in the workforce.

The interplay between quantitative and qualitative aspects of modern urban life fascinated social theorists.

City dwellers would be in contact, directly and indirectly, with thousands of other people, rather than just a few hundred as in rural society. It was believed that such increases in the scale of urban life had a profound, albeit mostly deleterious, impact on the way that people lived. The German social theorist Fernand Tönnies, for example, contrasted rural and urban lifestyles, arguing that the former was characterized by the warm primary social relationships of community (*Gemeinschaft*) and the latter by the cold, impersonal, and contractual relationships of society (*Gesellschaft*). His compatriot Georg Simmel accepted that there were positive elements to city life, particularly through removing the constraints of small town life and allowing individuals to assert their uniqueness, but expressed concern about the price of adjusting to the new environment. He argued that the slower tempo of rural society permitted unconscious and habitual behavior patterns, whereas the modern city forced people to devise conscious strategies if they were to cope with the sensory overload that resulted from the quickened pace of life. Simmel argued that this condition caused a distinct 'metropolitan personality', manifested in the detached and blasé approach that people adopted when dealing with their fellow citizens.

His ideas, in turn, influenced American social theorists, particularly Robert Park who had attended lectures that Simmel gave in Berlin in 1899. Park took the idea that individuals and groups had to adjust to come to terms with modernity although, unlike Simmel, believed that the city was also a product of nature, especially human nature. From that standpoint, Park proposed that human society might be studied using concepts analogous to those employed by ecologists when studying plant communities. Park moved to the University of Chicago in 1914, where his proposal for the study of 'human ecology' strongly influenced the approach to sociological inquiry adopted by the University of Chicago's Department of Sociology and Anthropology. The so called Chicago School of social researchers used their home city as laboratory, applying ecological concepts to fieldwork based studies of, *inter alia*, community formation, gangs, hoboes, social movements, and race relations. They recognized that distinct communities, identified by their shared social characteristics, were clustered in geographically delimited portions of the city. Following ecological theory, these were termed 'natural areas' and constituted the 'niches' which different communities occupied within the city.

The idea of natural areas served as the cornerstone for generalizations about the spatial organization that lay beneath the superficially chaotic appearance of the modern city. The most famous exposition of these principles was the concentric zone theory, developed by Edward Burgess in 1924 (**Figure 1**). Chicago, located on the shoreline of Lake Michigan, had steadily expanded outward for over a century. Burgess interpreted that experience to propose a model of five ever expanding concentric rings. At its heart was the central business district (CBD) known in Chicago as 'the Loop', although the model was generalized to address the tendencies of any town or city to expand radially from its center. Burgess noted that Chicago's growth was fueled by the arrival of poor immigrants, who had little option but to accept cheaper housing in the run down districts close to the urban core. This area, known as the 'zone in transition', not only housed the city's transient elements and underclass, but also saw the greatest competition for land and resources due to the pressure on land caused by the growth of business and light manufacturing. As time went by, the poor could take advantage of their rising standards of living to move out to locations where housing was better and crimeless, with their place taken by newer immigrants who, like their predecessors, had to take what was on offer. Poorer areas, therefore, were located near the center and more prosperous districts near the city's periphery. Processes of change occurred through sequences of 'invasion' (the spread of a community into an area previously occupied by another) and 'succession' (the resulting changes in the composition or structure of an ecological community).

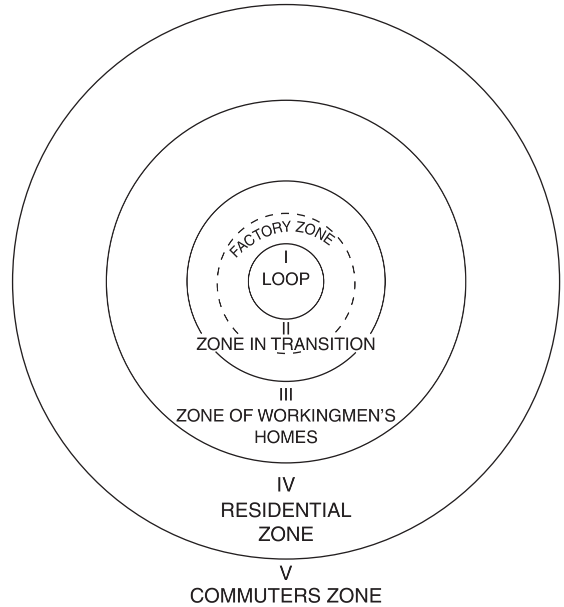

Figure 1 The growth of the city. Source: Park, R. E., Burgess, E. W. and McKenzie, R. D. (1925). *The City*. 51p. Chicago: University of Chicago.

The concentric zone model showed that the modern city had an underlying sociospatial structure vital to its functioning rather than being an inchoate mass of loosely related parts. Later theorists attempted to revise and extend Burgess' formulation. In 1939, Homer Hoyt added sectoral principles (development in wedges out from the city center), although he was more interested in the lo cation of higher priced housing than developing general models. In 1945, Chauncy Harris and Edward Ullman suggested that cities might have multiple nuclei rather than a single center. Collectively, their ideas influenced thinking about the structure of the modern city and, indeed, excited another generation of urban theorists in the 1960s and 1970s with the rise of spatial analysis.

Nevertheless, the models had long drawn adverse comment for their Darwinistic and ecological basis, with critics arguing that they oversimplified urban land use patterns and failed to explain how natural areas formed in the first place. In addition, although belatedly realized, ecological models effectively described the past. For most US cities, the CBD had reached its peak by the late 1920s and no longer supplied the engine by which the dy namics of the commercial economy generated waves of invasion and succession. Later, the onset of deindus trialization further eroded the value of ecological models by removing other key dynamics of growth. Yet long before that time, new developments in planning and architectural practices, themselves rooted in the logic of industrialism, had arisen that appropriated the term modern city for the period of postwar development characterized by radical urban reconstruction. Those practices were rooted in the rise of modernism.

Modernism

While there is no single consensual definition for 'mod ernism', it was essentially a shared international tendency that emerged within the arts in the last decade of the nineteenth century and the first three decades of the twentieth century. Manifested in the work of disparate and largely independent movements that challenged the established order, modernists sought to express the spirit of modernity, in both its optimistic and despairing modes. In terms of representation, modernism prompted a search for forms that expressed the needs and chal lenges of modern times against those of the past. Seen in this way, modernism was defined primarily by forms and esthetics. These included atonality in music, abstraction in art, free verse in poetry, streams of consciousness in novels, and geometric and unadorned shapes in archi tecture. At another level, modernism was identifiable by the attitudes that underpinned the new representations, which included attitudes toward the past, the arts and their relationship with industry, and the conduct of everyday life. Modernism, thus, was not just a matter of painting, sculpting, or designing buildings in radical ways; it also signified that the individual artist or designer had adopted moral positions guided by the 'right spirit'.

At the outset, then, the development of modernism in general and architectural modernism in particular was founded in critique. Modern architects rejected the conventional role of the architect as restricted to being the designer of special buildings. Rather they believed that for architecture to work, it was necessary to take a holistic view and become involved in the design of the everyday environment at all scales from interior design to the city and beyond. Modernists castigated traditional building practices for placing pursuit of style over function and for failing to take advantage of the potential of new constructional materials, such as steel, glass, and concrete. Above all, they criticized what they saw as the muddle and confusion of the industrial city, which had grown without formal plan and with minimal concern for the welfare of the working population.

Instead, architectural modernism embraced rationalist approaches toward the ordering of urban space, with de velopment guided by comprehensive ('master') planning. Modernists celebrated the potential of industrialized building methods as offering the opportunity to remake the modern city and offer better conditions for its in habitants. The dwelling, from this point of view, could be as precisely, quickly, and efficiently made as any manu factured product. The 'modern movement' saw archi tecture as responding primarily to what the building was supposed to do. In emphasizing function rather than being deflected by considerations of culture or style, modernism therefore claimed to offer principles suitable for universal application. These principles in due course would lead to visions of the modern city that combined buildings, traffic movement systems, and land use in ways quite different from anything previously attempted (**Figure 2**).

The Making of the Modern

There were many hearths of modernism. North Ameri can cities, particularly the late nineteenth century Chi cago, made their contribution. In 1885, for instance, the 'Loop' gained the first recognizable skyscraper with the completion of the ten story 'Home Insurance Building'. Designed by a local architect William LeBaron Jenney, it showed the development potential of 'iron and steel' framed structures that freed buildings from the height restrictions imposed by the load bearing strengths of conventional materials. Chicago's CBD quickly took on a pronounced vertical dimension (**Figure 3**), with the city acting as the proving ground for building forms that would later change the appearance of the downtowns of large cities throughout the world.

Another important source of influence came from the public housing estates of West European cities, particu larly Rotterdam and Amsterdam in the Netherlands and Munich, Hanover, and Frankfurt in Germany. Housing was always of particular concern for modernists, with social housing programs offering the opportunity to raise

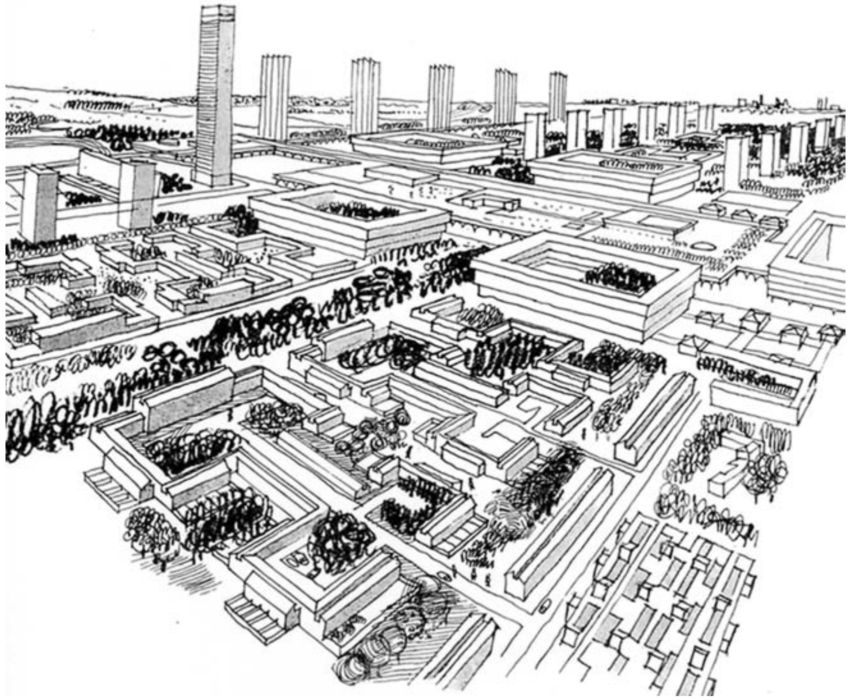

Figure 2 Scheme for the regeneration of Boston Manor (1958). General view from the southeast. Source: *Architecture and Building*, 1958, vol. 33, p 337 (journal ceased publication in 1960, authors deceased, copyright untraceable).

the living standards of the working classes. The city of Frankfurt, for instance, created eight major housing developments between 1925 and 1929, providing 15 174 living units – roughly 11% of Frankfurt's total housing stock – primarily in peripheral estates of row housing. Designed by the Architect's Department under Ernst May, the estates were notable both for innovations in prefabricated concrete slab construction and for use of 'minimum habitation' (*Existenzminimum*) space standards to maximize the number of dwelling units created. Achieving these standards was dependent on extensive use of ingenious built in storage, foldaway beds, and kitchen design. Around 10 000 of the dwellings, for instance, were equipped with the industrialized and ergonomic 'Frankfurt kitchen' which was intended to modernize and simplify household chores.

These and similar developments provided ingredients for the visionary city schemes produced by European modernists immediately before and after World War I. The German architects Bruno Taut (1880–1938) and Ludwig Hilberseimer (1885–1967), the Swiss architect Le Corbusier (1887–1965), Russian urban decentralists such as Nikolai A. Miliutin (1889–1942) and Mikhail Okhitovitch (1886–1937), and the Italian Futurists Antonio Sant'Elia (1888–1916) and Mario Chiattone (1891–1957) all contributed images of urban futures that responded to the opportunities made available by science and advanced technology. Their chosen designs varied profoundly in spatial organization and esthetics, but broadly

Figure 3 Steel-framed buildings in 'the Loop', Chicago.

contributed two prototypic visions of what the future city might look like. One type exploited the potential of mass transportation, with the components of the city arranged into corridors along high speed transport lines, simultaneously offering enhanced mobility while retaining proximity to open countryside (due to the narrowness of the strips of development). The other type comprised the 'city of towers', in which the modern city appeared as a geometrically arranged and high density agglomeration with a pronounced vertical dimension.

Further development of these ideas came from the activities of the *Congrés Internationaux d'Architecture Moderne* (CIAM). Founded in 1928 as an international body for promoting and disseminating modern architecture, CIAM was the most important forum for discussion of ideas about modern city design in the interwar period. Its meetings and manifestoes consistently asserted the architects' right to be involved in city planning. Although far from being harmonious or single minded, the organization promoted an analysis that saw architecture as appropriate if fit for purpose. CIAM's approach took four elements – work, housing, transportation, and recreation – as the key functions of the city and held that the city worked best if these were allocated to separate land use zones or channels. Although this left aside the difficult question of how to deal with town centers, CIAM's functionalist recipe appealed as a way to end the conflicts produced by mixed land use and looked to segregation of pedestrians and vehicles as a rational way to design cities for the motor age.

Making the Modern City

Economic and social conditions after the war were not immediately favorable to the rapid reconstruction of the modern city. Nations that had experienced wartime destruction and the upheavals caused by mass flows of migrants necessarily prioritized resources for urgent action to renovate the urban fabric, deal with housing shortages, and repair infrastructure. In most cases it took until the late 1950s, when the immediate problems of economic austerity had receded, for policymakers to tackle the larger question of pervasive urban problems. When doing so, they could take advantage of a social consensus that favored planned and comprehensive approaches rather than piecemeal action and which saw the industrial city, once the symbol of modernity, as now associated with the failure of the old order. The ethos of the times, combined with rising standards of living and growing awareness of the challenge posed by rapidly increasing car ownership, favored urban reconstruction and modernization. The moment was therefore propitious for the spread of modernist thinking about the broad framework of the city and its various components – workplaces, leisure facilities, city centers, hospitals and health centers, schools, universities, housing, and roads.

The first step in comprehensive redevelopment was wholesale clearance of dilapidated districts, although official definitions of 'dilapidation' frequently bore the hallmarks of convenience if an area of existing development stood in the way of rebuilding a town center or constructing new housing estates. Planners routinely sought to recreate urban structure by adopting principles of single land use zoning. Housing areas were conceived as 'neighborhoods' in which good design supposedly led to the flourishing of communal life and, particularly when dealing with social housing, gave previously deprived tenants an opportunity to experience a markedly improved standard of life. New housing estates were emblematic of the modern city, at their best providing an exciting imagery of tall buildings set in green open space – the embodiment of a 'city in a park' (**Figure 4**). Shortage of resources, however, frequently led to approaches that neglected the necessary provision of social and communal facilities and concentrated on producing the maximum number of dwellings at lowest cost. This was particularly the case in cities that experienced shortage of suitable land for building.

Traffic arrangements saw the introduction of traffic free precincts in association with high capacity roadways and multistory car parks. There was extensive use of vertical segregation of pedestrians and vehicles by means of flyovers, underpasses, and subways. Sensitivity to the subtleties of place was secondary to the pursuit of spatial order, often leaving major historic landmarks and buildings as isolated and forlorn islands in a sea of change. Important changes also occurred in the appearance of the built environment. Skylines became punctuated by tall buildings, with the use of modern materials and the adoption of novel building technologies contributing to a new visual vocabulary of nontraditional built forms. This affected the construction and appearance of houses, offices, factories, transport facilities, bridges, and street furniture. Concrete became a favored building material. Official handbooks provided guidelines for landscaping, signage, and even ways of trimming trees fringing the highways.

Architects and planners, of course, rarely had a free hand to impose blueprints derived from visionary principles. In Britain, for example, town center renewal scarcely occurred before the 1960s, with the passage of time since the end of the war serving to foreclose options. To elaborate, the property boom had seen piecemeal accumulation of office blocks and partial redevelopment of high streets of many towns and cities in a manner that paid little attention to the wider needs of the town. Highway authorities, for their part, had also created roads without close coordination with plans for the built environment. In Birmingham, for instance, the city's

Figure 4 Flats at Alton West, Roehampton, London (LCC Architects Department, 1954 1958).

engineers pushed forward the long awaited 'inner ring road' in 1956–57, deliberately choosing not to adopt a plan for the central areas in case its stipulations deterred investment by the commercial sector. Instead, they sought to capitalize by offering potential developers sites along the line of the 'ring road'. This led, for example, to the Smallbrook Ringway, one of Britain's first urban motorways, becoming lined with continuous ribbons of offices and shops – an extraordinary negation of the principles of pedestrian–vehicle segregation so cherished within modernist circles.

This saga of insensitive road planning was not atypical, either in Britain or elsewhere. The private car was then closely associated with personal mobility and freedom of choice. A powerful coalition of politicians, planning and design professionals, and commercial interests connected with the road industry shared an ideological commitment to supply led solutions for road problems. Few challenged the prevailing view which held that, with suitable surgery, the older urban cores could accommodate modern traffic and communications flows on an indefinite basis. After all, city planners and local politicians associated such surgery with the pursuit of progress, with cities needing to change to meet the demands of the internal combustion engine. Carving urban motorways through the knotted entrails of the inner city and building high capacity ring roads ranked along with remodeling city centers and employing high rise flats as precisely the actions expected of progressive municipalities.

The Demise of the Modern City

The confidence with which these goals were pursued would soon be severely dented by wholesale criticism. To some extent, the critique had already emerged *before* the main phase of reconstruction occurred. In the late 1950s, Jane Jacobs, an associate editor of the *Architectural Record*, took issue with prevailing approaches to urban renewal. As outlined in her 1961 book *The Death and Life of Great American Cities*, Jacobs saw the city's vitality threatened by a utopian and predominantly visual artistic approach to redevelopment that failed to take account of urban complexity and intensity. During the early 1960s, studies on both sides of the Atlantic made it clear that comprehensive redevelopment policies failed to take social factors into account that, in some instances, compensated for the physically blighted nature of areas earmarked for clearance. These studies made little impact at the time, but their importance was reassessed, when a raft of other problems became manifest in the late 1960s and early 1970s. City center redevelopment schemes repeatedly failed to resolve how to devise attractive town centers that gave expression to civic identity and placed narrowly defined functionality before liveability. Road developments, especially urban motorways, added to urban pollution and created impenetrable barriers within the districts they bisected. Pedestrianized precincts suffered from visual sterility and placelessness. The reputation of social housing, especially estates of tall blocks, collapsed in the face of high profile reportage of constructional, social, financial, and management problems.

Taken together, these criticisms led to the paradox of the modern city becoming a historic term. Just as the modern city *sensu lato* became superseded when processes of deindustrialization finally removed the intimate relationship between modernity and industrialism, so too did the strength of criticism directed toward modernist intervention render the modern city *sensu stricto* a thing of the past. Henceforth, commentators would have to resort to the prefix 'post' – postmodern and postindustrial – to meet the linguistic problem of how to describe what came after the modern city.

See also: Urban Architecture; Urban Representation/ Imagination; Urbanization.

Further Reading

Brown, N. (2001). 'Robert Park and Ernest Burgess: Urban Ecology Studies, 1925. http://www.csiss.org/classics/content (accessed June 2008).

Cherry, G. E. (1988). *Cities and Plans: The Shaping of Urban Britain in the Nineteenth and Twentieth Centuries*. London: Edward Arnold.

Deckker, T. (ed.) (2000). *The Modern City Revisited*. London: Routledge.

Donald, J. (1999). *Imagining the Modern City*. London: Athlone Press.

Gold, J. R. (1997). *The Experience of Modernism: Modern Architects and the Future City, 1928 1953*. London: E. and F.N. Spon.

Gold, J. R. (2007). *The Practice of Modernism: Modern Architects and Urban Transformation, 1954 1972*. London: Routledge.

Hall, P. (2002). *Cities of Tomorrow: An Intellectual History of Urban Planning and Design in the Twentieth Century* (3rd edn.). Oxford: Blackwell.

Hvattum, M. and Hermansen, C. (eds.) (2004). *Tracing Modernity: Manifestations of the Modern in Architecture and the City*. London: Routledge.

Jacobs, J. (1961). *The Death and Life of Great American Cities*. New York: Random House.

Le Corbusier. (1967). *The Radiant City*. London: Faber and Faber.

Lindner, R. (1990). *The Reportage of Urban Culture: Robert Park and the Chicago School*. Cambridge: Cambridge University Press.

Mumford, E. (2000). *The CIAM Discourse on Urbanism, 1928 1960*. Cambridge, MA: MIT Press.

Punter, J. and Carmona, M. (1997). *The Design Dimension of Planning: Theory, Content and Best Practice for Design Policies*. London: E. and F.N. Spon.

Ravetz, A. (1980). *Remaking Cities: Contradictions of the Recent Urban Environment*. London: Croom Helm.

Robson, B. T. (1969). *Urban Analysis: A Study of City Structure with Special Reference to Sunderland*. Cambridge: Cambridge University Press.

Ward, S. V. (2004). *Planning and Urban Change* (2nd edn.). London: Sage.

Relevant Websites

http://www.open2.net
From Here to Modernity, BBC/OU Open2.net.

http://www.lib.uchicago.edu
Robert Park: Sociology, University of Chicago Centennial Catalogues.

Modernity

D. Linehan, University College Cork, Cork, Ireland

Introduction

It is simply not possible to say when precisely modernity begins, but it is widely accepted that a highly articulated version of modernity emerged in Europe from the fifteenth century onward when three key developments combined to usher in a new social order. These three developments were: the expansion of scientific and humanist knowledge, the rise of capitalism, and the growth of strong centralized states. The creation of the modern world as a consequence of these developments has been intensely geographic in nature. Together with urbanization, colonialism, and globalization, modernity involved extensive contact and mixing between world cultures, the creation of new geographic divisions of labor, the constitution of new spatial relationships, and profound transformations of nature and the environments. The incremental and often contested consolidation of modernity has steadily transformed commerce, communication, and culture and altered public and private transactions in social life. Bonded together by a strong belief in progress, the relationships were mutually re enforcing and represented a powerful discourse about a new way of organizing political, economic, and social life. Although these relationships have gone through many different changes, they remain at the heart of the project of modernity. Dispersed between the foci of knowledge, capitalism, and the state, geographical thought and practices can be regarded as key agents through which modernity has been formed, coordinated, and produced. Simultaneously, the conditions created within modernity have reshaped how people perceived global relationships, imagined the environment, lived in places, and gone about organizing the economy. An appreciation of modernity from this perspective challenges the very basis upon which space and place are conceived and encourages more nuanced and 'relational' approaches to thinking about human geography.

Modernity, Knowledge, and Geography

For many observers, the intellectual foundations of modernity lie in the transformations in the production of knowledge between the Renaissance and the Enlightenment. This historical period was characterized by the rising dominance of individualism and, in contrast to the fatalism of the medieval period, tended to promote an optimistic outlook on the potential of humankind. Based broadly upon the rediscovery of knowledge from classical antiquity and the Arab world made during the Renaissance, the cultural and technological innovations of the Enlightenment enabled the creation of a scientific worldview that ushered in new regimes of order, progress, and rationality. The Enlightenment is sometimes portrayed as a rupture in the history of human development, following which people in the West increasingly imagined themselves in novel ways. This development established a new and powerful social imaginary whereby it was assumed that modern society was on a progressive trajectory, and through rational, scientific endeavor, was destined to transcend its limitations and rise to a higher stage of human development. In contrast to the premodern individual who may have viewed herself and the world through the lens of myth or religious belief, the modern person increasingly drew upon the tenets of humanism and science, and in time, would be asked to become more reflexive and self conscious in outlook. This new sensibility was well reflected in the literature and art from this period onward. The rise of portraiture, the proliferation of personal diaries, or the self conscious hero in William Shakespeare's *Hamlet* (written in 1600) who asks himself 'to be or not to be' can all be regarded as cultural artifacts generated among some of the new spaces of modernity which proliferated from this time onward – spaces such as the art gallery, the coffee house, and the theater.

Geographical ideas also changed radically during the Enlightenment, and concepts of nature, time, and space were fundamentally re ordered and reflected strongly the impulse of modernity. The Enlightenment worldview helped naturalize the essentially representational category of geometric space as 'fact'. This awarded new powers to the abstractions of the mapmaker, the surveyor, and the architect, whose work could be used to powerfully envision the state of the world (**Figure 1**). Developments in geographic science, notably in new techniques and theory in navigation, surveying and mapping, became a key way to both perform and represent a modern worldview. Through, for example, the consumption of globes, atlases, or travelogs, geographical knowledge enabled literate Western individuals and societies to see themselves as modern and distinguish themselves from the so called primitive 'others' (**Figure 2**). Geographical science also provided a set of techniques to plan and implement the conquest and reorganization of terrestrial resources. For instance, by calculating and measuring landholdings, topographical surveying underwrote the legal appropriation of the commons and the transformation of communal and ancestral

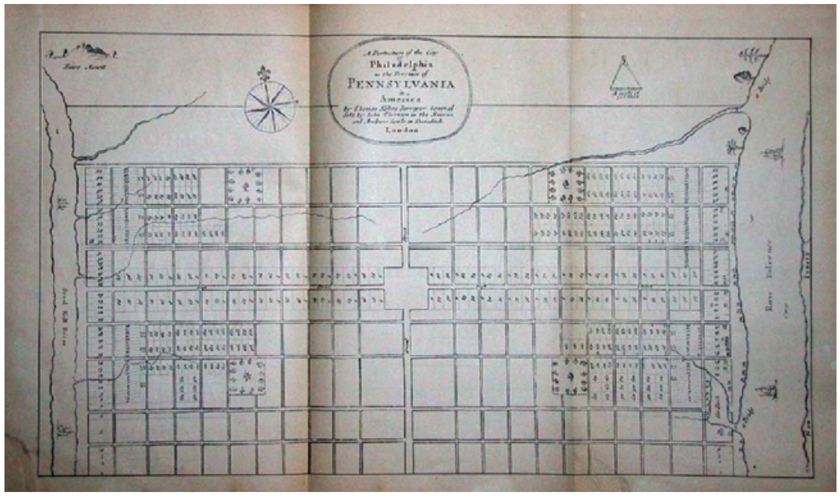

Figure 1 Plan of the city of Philadelphia (1683). Devised by William Penn, this conception for Philadelphia is among the earliest examples of utopian city planning. As illustrated here, one of the most explicit expressions of Enlightenment thinking was in the use of grids to organize urban development and human settlement. Grids appeared frequently in colonial contexts, underwriting the distribution of land and urbanization in Australia, North America, and Africa.

Figure 2 One of the several large globes made for Louis XIV of France by the Italian geographer and cartographer, Vincenzo Coronelli, in 1683.

land into private property. Geographical science also enabled European exploration, moderated the risks in global sea travel, and in innumerable ways supported the appropriation of resources and peoples exploited by colonialism. In this regard, it can be said that modernity became closely associated with a spatial logic that informed the organization of resources and investment and played a key role in the distribution of land.

More broadly, within modernity, the creation and application of knowledge became instrumental to the rational organization of society. Consequently, ideas of progress and the use of scientific knowledge were foundational attributes of most forms of planned urban transformation in the nineteenth and the twentieth centuries. From Thomas More's *Utopia* to urban plans devised by Ebenezer Howard or Le Corbusier, Enlightenment values have had an enduring effect on shaping space and society (**Figure 3**). These ideals became impressed into the planning of cities like Vienna, Paris, and Barcelona, which in the eighteenth and nineteenth centuries laid out new systems of streets,

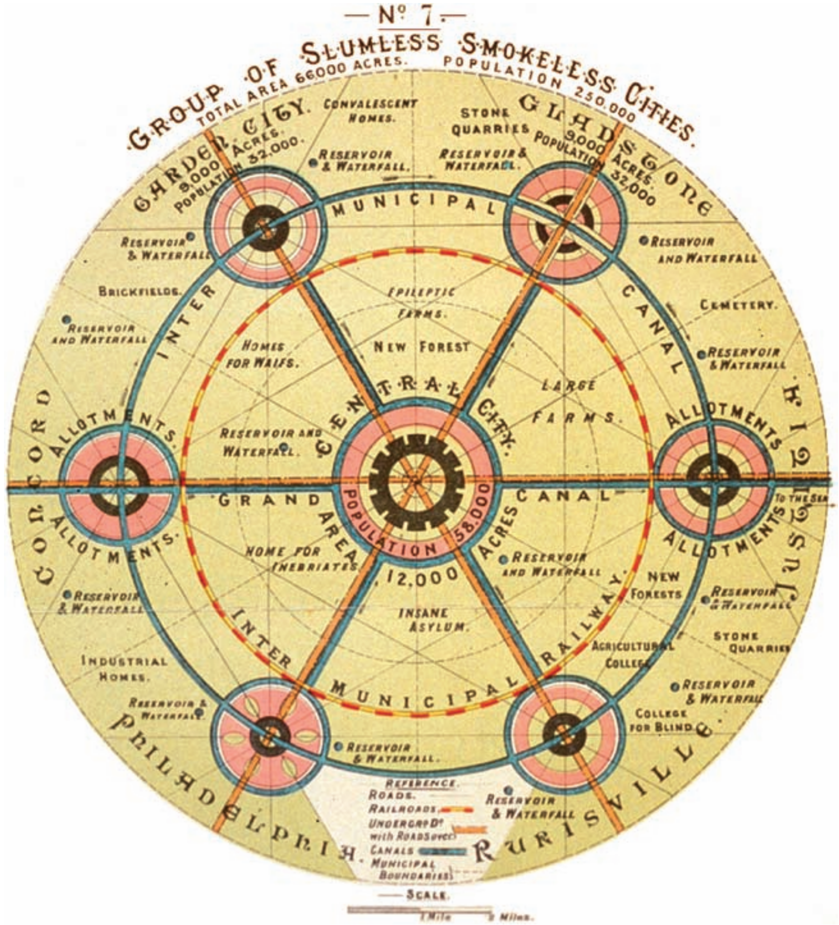

Figure 3 Illustration from *Garden Cities of To-Morrow* (London, 1902). Ebenezer Howard's diagrammatic illustrations of how a model garden city should be planned were always understood to be idealizations, but nevertheless, were enthralled by Enlightenment and Utopian thinking. These geometric designs embraced the problems and potential of modernity, and presented new visions of order. Howard offered a workable plan for a progressive urban society, which through spatial planning would address the alienation of human society from nature by integrating cities into the surrounding environment.

promenades, boulevards, and park networks. These new modern spaces were intended to reflect the civilized and modernized outlook of their populations and more particularly, the modernizing impulse of their political elites. The spatial signature of this progressive culture can still be seen in the symbolic cores of many European cities, such as Paris, Madrid, and St Petersburg, crowded with monumental architecture, spectacular landscapes, and progressive urban institutions such as galleries, exhibition halls, and museums, all committed to the heroic celebration of human endeavor and creativity, core humanist attributes.

However in line with the ambiguous and contradictory qualities of modernity, it has been observed that the pursuit of order has often involved profound social disruption. Modernity was often achieved with severe costs to liberty, life, and ecological sustainability. In spite of its claim to progress, modernity held within it, the seeds of barbarism. In his lectures on 'physical geography', the Enlightenment philosopher and geographer, Immanuel Kant concurred with another pillar of Enlightenment thinking, David Hume, that white Europeans were superior in racial characteristics to Asians and Africans. Both these key thinkers declared that Africans were incapable of rational thought, cultural expression, and were to all extents 'outside humanity'. These kinds of positions embedded inside the Enlightenment project helped underwrite European arguments for the acquisition of colonial lands and ensured slavery was maintained until the mid nineteenth century as the basis for the commercial exploitation of resources in the so called 'New Worlds'. As such, the utopian promises of the Enlightenment were rarely attained. Indeed, they were often distorted, as, for example, in the 'civilizing' but more often than not, racist mission of European colonialists in Asia and Africa in the nineteenth century, or the rational but brutal organization of concentration camps during the twentieth century. Consequently, modernity – in particular, how it represented by Enlightenment thinking and its inheritors – may be understood as a duplicitous event, promising stability and progress, but often presenting a stressful, destabilizing, and precarious experience. This was particularly the case for the peasantry, the poor, and indigenous peoples, exploited by colonialism, harmed by racism, and excluded from the comfortable orbit of bourgeois protection and social life.

Modernity, Capitalism, and Space

Underwriting modern forms of knowledge was the dynamic of capitalism, which as outlined above, often instrumentalized the potential of Enlightenment thinking in the pursuit of profit and higher levels of productivity. The emergence and consolidation of capitalism led to sweeping changes in the organization of social and political life. It specifically led to the creation and destruction of markets and the social and cultural institutions that supported and thrived within these new arrangements. Many geographers have drawn upon Karl Marx's observations on the role played by new forms of production and consumption in upheaval of spatial, temporal, and social relationships (**Figure 4**). With capitalism as its engine, these conditions gradually transformed the human experience through the instrumental re casting of nature, economy, and personal life. Capitalism thrived in the new technological and political conditions of modernity and created new ways of organizing money and credit, agriculture, work, and trade. These transformations all found profound expressions in the landscape. The transformative power of capitalist modernity is well revealed in the international development of sugar and cotton cultivation between 1600 and 1800. These developments created a vast and highly interconnected Atlantic trade, involving the systematic enslavement of millions of African peoples, who were modernized against their will by coercive discipline and forced transportation to work as slaves in vast plantations, whose construction involved the obliteration of indigenous people and the destruction of natural habitats. The wholesale disruption of social life, the oppression of the individual, and the creation of intricate geographies of dependency, inequality, and exploitation involved within such systems was regularly expressed in the histories of capitalism. Developments such as manufacturing, mining, and plantation based agriculture frequently undermined the rights of workers and the poor while simultaneously threatening or destroying the ecological sustainability of the environment to extract profit.

In addition to extraordinary impacts on localities and their communities, for Marx, the power of industrial capitalism resulted in the 'annihilation of space'. This process generally described the ways in which the spatial boundaries and obstacles to trade and travel were broken down by investment in new transportation infrastructure and communication technology. Following the development of reliable trading routes within Europe,

Figure 4 Johannes Stradanus, detail from *Nova Reperta*, *c*. 1600. One of the most significant developments within modernity was the acceptance by the populace of 'standard time', whereby modernity was made tangible through the adoption of a highly regulated way of conceiving time. In England, for example, from the early 1500s, public timekeeping was widespread, with 80% of market towns serviced with a public clock. These developments, together with the eventual widespread adoption of wrist watches in the early twentieth century, literally connected new technologies of time-keeping to the body, and restructured people-conscious organizations of reality.

technological innovations with global impacts, such as the development of canals, steamships, railways, and telecommunications, made distance culturally and economically less significant. Places became increasingly interconnected, their populations more mobile, and the goods and ideas they produced and consumed increasingly integrated into extended circuits of trade, communication, and knowledge. The sociologist Anthony Giddens has noted that advances provided by such technologies redefined the meaning of geography and radically shifted the foundational basis of place. It followed that in modernity, the very meaning of 'presence' shifted and relations between those activities embedded in the locality and those of 'absent others' elsewhere increasingly shaped the nature of place and of human consciousness itself. This process ensured that social and economic relationships were lifted out of local contexts and shaped in a wider field of social relations. Thus, when Europe and America were first connected by telegraph cable in the 1860s, the Earl of Carlisle could stand on the windswept shore of the West of Ireland where the telegraph came ashore and declare: "Hitherto the inhabitants of the two worlds have associated perhaps in the chilling atmosphere of distance with each other at a sort of bowing distance; but now we can be hand to hand, grasp to grasp, pulse to pulse."

In time, cities such as Buenos Aires, Berlin, Shanghai, or Mumbai were dramatically transformed by the extraordinary power of industrial capitalism and the impacts it had on world trade. These developments were often rapid and disruptive and in themselves came to signify the experience of modernity. In Paris, reconstruction undertaken in the mid nineteenth century effectively separated the financial and commercial districts from residential and recreational areas in the city. This transformation destroyed many areas of the medieval townscape and involved a radical displacement of many elements of the city's population. Under capitalist modernity, urban dwellers often faced the contradiction of being thrilled and emancipated by new urban spaces, but also felt disempowered and excluded from the new metropolis that emerged. There has been no shortage of artists and social commentators who conceived a link between the new social and environmental conditions of the city and psychological state of the urban dweller. Whether concerned to interpret the effects on the individual to exposure to city crowds, the acceleration of personal life or people's separation from nature, a range of classic interventions by Simmel (1901), Mumford (1938), and Wirth (1964) postulated new inter relationships between self and society within modernity. George Simmel's seminal essay, 'The metropolis and mental life', based primarily on the transformations he witnessed in his own hometown Berlin during the late nineteenth century, identified the emergence of an essentially modern attitude of the urban dweller, based on indifference and rationality. Hence, as much as the metropolis symbolized progress and enlightenment, it has also been recognized that the impacts of modernity in the city also resulted in the loss of community, the erosion of tradition, intense individualization, and in its worst forms, social atomization. These characteristics were captured by Marshall Berman's (1982) classic text *All That Is Solid Melts into Air*, where he famously refers to modernity as storm of perpetual disintegration and renewal, propelling people into a life of struggle, anxiety, and ambivalence, in which they are faced with the task of constantly revising the basis of their identity. At the same time, however, other observers have emphasized that the modern city was also a space of liberation, of freedom, and security. This has been particularly noted in the case of twentieth century Western women, who were offered new roles and ways of life in the new economies of the modern city. The city then became a space of liberation, particularly from the restrictions of women's lives in traditional, and often, patriarchal rural societies. Such apparent contradictions in the experience of modern city life reflect squarely the problematic meanings of modernity itself.

The State and Modernity

As a social and ideological force, modernity has always found fertile grounds for development within the opportunities created by the modern state. In contrast to stateless societies characterized by informal modes of governance, incoherent political boundaries, and a dependency on custom to resolve disputes, the modern states consolidated their rule over a strongly defined territory, and were able to expand their bureaucracies, standardize the application of the law and taxation, and particularly through the control of the military and the police enhance their capacity for surveillance and control. This process led to the domination of physical space, the control of land, and the constitution of the legal and illegal subjects of the state – 'the citizen' and 'the alien'. Importantly, this compulsion to create order did not stay fused to state institutions but percolated throughout society where obligations to 'progress' and indeed to 'civilization' shaped, either through compulsion or persuasion, the daily practices of individuals and organizations. The social theorist Jürgen Habermas argued that many of the fundamentals of the modern civil society were ensured through the creation of opportunities for debate and discussion, in what he termed the 'public sphere'. His work demonstrated that the 'public sphere' emerged in a number of institutions during the eighteenth century. From the salons and drawing rooms of Paris frequented by artists, actors, and politicians, to London coffee houses filled with merchants, sailors, and bankers, new styles and practices of communicating emerged and were protected within the embryonic contexts of liberal democracy. Together with the growth of literacy and the print media, these spaces functioned as points for debate and information exchange. They also became sites where new tastes, values, and norms of conduct were diffused through society, ensuring the populace was always aware of what was fashionable and 'new'.

In a comparable context, the work of the philosopher Michel Foucault on the history of subjectivity, governmentality, and space has also been highly influential in shaping the geographers' understanding of the relationships between identity and spatiality within modernity. An important pillar in Foucault's observations on modernity is the issue of surveillance, whose constancy in modern society, he suggested, profoundly altered people's behavior and mentalities to such a degree that individuals internalized the rules and aspirations of the modern state. As part of the 'civilizing process' of modernity, key institutions such as the civil service, the police, and schools emerged as agents of modernity, scrutinized the population, and regulated in turn new forms of conduct in the populace. These behaviors included observance to social and personal hygiene, 'proper' public behavior, and control on personal morality. Within the modern state, it has been acknowledged that these civilizing values and practices shaped the disposition of people in a broad range of urban institutions, such as the factory,

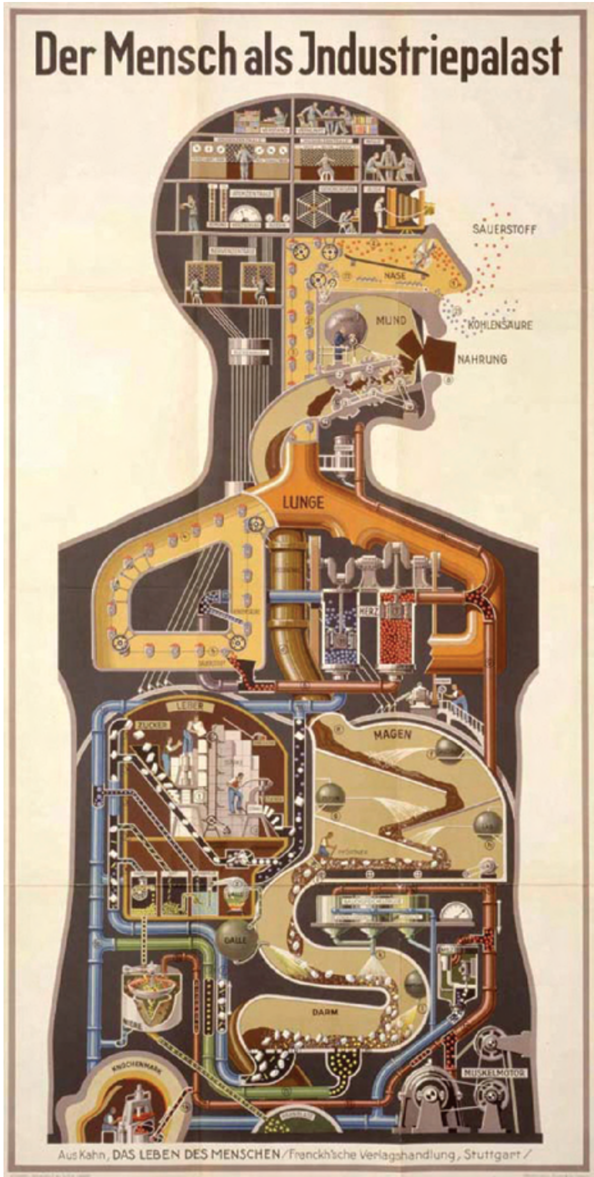

Figure 5 *Der Mensch als Industriepalast.* (Man as an industrial palace) (1926). Devised by the illustrator Fritz Kahn. Here, he located industrial modernity within the body the digestive system is imagined as a chemical factory, nerves were picturized as telephone wires, and eyes as cameras. Symbolically merging the body with industrial technology epitomized the qualities of 'high modernity', reflecting full confidence in modernist, technologically based solutions to social issues.

museums, exhibitions, art galleries, department stores, and dancehalls, where the modern body and the modern subject were embedded into the project of modernity (**Figure 5**).

Modernity Futures

Located around the hub of knowledge, capitalism, and the state, geographical practices and thought have been consistently bound up with the project of modernity. Although these relationships continue to go through many changes, the importance of modernity for geographical research remains core to understanding the qualities and impacts of contemporary life. Undoubtedly, today modernity is again at the crossroads. As a consequence of ongoing social, political, and economic transformation, some commentators have argued that there is a crisis over whether society as it is envisaged within modernity can now be identified, notably as categories such as 'locality' and 'community' seem undone by rampant globalization. At the heart of these concerns are critical questions about the nature of social bonds, the institutional order, and collective life itself, which may or may not have fundamentally changed. It is clear however that in spite of the postmodernist critique, which sought to challenge the basis of modernism, modernity itself remains a set of processes which inevitably transforms everyday consciousness and identities. It continues to impact profoundly on global development and it continues to mutate and adapt to local contexts and cultures. As such, the geographical engagement with modernity means embracing a far more subtle notion of place than presented in traditional ideas of the discipline, with their particularly bounded notions of place, and geographical scale. Engaging the consequences of modernity ensures that geographers will continue to invest in place but not as traditional mapmakers whose cartography sealed the difference between people and their culture with thick inky lines. Modernity's geographies show that even as things fall apart, people and place are forever healing, creating new landscapes, new challenges, and new conditions for the human experience.

See also: Capitalism; Enlightenment Geography; Modern City; Public Space.

Further Reading

Berman, M. (1993). *All That Is Solid Melts into Air. The Experience of Modernity*. London: Verso.

Bunnell, T. (2004). *Malaysia, Modernity and the Multimedia Super Corridor: A Critical Geography of Intelligent Landscapes*. London: Routledge.

Entriken, J. N. (1991). *The Betweenness of Place: Towards a Geography of Modernity*. Baltimore: John Hopkins Press.

Giddens, A. (1990). *The Consequences of Modernity*. Cambridge: Polity.

Habermas, J. (1989). *The Structural Transformation of the Public Sphere*. Cambridge: Polity Press.

Harris, C. (1991). Power, modernity, and historical geography. *Annals of the Association of American Geographers* 81(4), 671 683.

Harvey, D. (2003). *Paris: Capital of Modernity*. New York: Routledge.

Kern, S. (1983). *The Culture of Time and Space, 1880 1918*. Boston: Harvard University Press.

Nye, D. (1997). *Narratives and Spaces: Technology and the Construction of American Culture*. Exeter: University of Exeter Press.

Ogborn, M. (1998). *Spaces of Modernity: London's Geographies, 1680 1780*. London & New York: Guilford Press.

Pred, A. (1990). *Lost Words and Lost Worlds: Modernity and the Language of Everyday Life in Late Nineteenth Century Stockholm*. Cambridge: Cambridge University Press.
Strohmayer, U. (1997). Technology, modernity and the re structuring of everyday geographies. *Geografiska Annaler* (Series B) 79(3), 155 169.
Wilson, E. (1985). *Adorned in Dreams Fashion and Modernity*. New Brunswick: Rudgers University Press.
Wirth, L. (1964). *Urbanism as a Way of Life*. Chicago: Chicago University Press.

Relevant Websites

http://www.ggy.bris.ac.uk
Clocks and Clock Time in England.

Modernization Theory

R. N. Gwynne, University of Birmingham, Birmingham, UK

Modernization theory refers to a body of theory that became prominent in the 1950s and 1960s in relation to understanding issues of economic and social develop ment and in creating policies that would assist economic and social transitions in poorer countries. This body of theory had significant influence in policy making from the 1950s to the 1970s. The various components of modernization theory received critiques from the outset, but their influence within policy making endured for a significant period of time.

However, the set of theories that tried to interpret the process of modernization in the 1950s and 1960s was by no means a coherent set of ideas but a wide ranging set of work from theorists located in a range of disciplines – as well as countries and universities. Many subsequent cri tiques of modernization theory have not fully examined the wide and varied range of contributions, preferring to focus on one or two prime candidates for their analysis. This brief article will attempt to examine a greater range of modernization theorists than normally come under scrutiny.

In retrospect, it could be argued that the most sig nificant contributions in modernization theory came from economic theorists, on the one hand, and sociologists, on the other. This article will seek to reevaluate the full range of modernization theory and will reexamine these literatures in terms of two categories – an analysis of the set of economic theories that attempted to explore the key features of modernization followed by the socio logical version.

It is first useful to put modernization theory into a geopolitical context, and more specifically that of the Cold War. After the end of World War II, the United States became the hegemonic power in the capitalist world but its former ally in that war, the Soviet Union, rapidly emerged as the only rival – and with a markedly different ideological agenda; the Second World became the shorthand for all those countries within this Soviet bloc. Preston argues that the bipolarity of the Cold War era set the context for policy makers and academics in what became known as the First World (USA, Western Europe, and their allies). Ideas and policies for devel opment in the so called Third World (the world's poorer countries in Latin America, Africa, and Asia) had to concern a distinctive 'free world' approach to what be came known as the modernization of economies and societies; the containment of communism and aid donor competition (between the USA and the USSR) were additional factors. Part of modernization theory specif ically set up agendas and policies for a noncommunist pathway to development at the height of the Cold War – as most famously in Walt Rostow's influential book of 1960, *The stages of economic growth: A non communist manifesto*.

However, the roots of modernization theory go further back than the aftermath of World War II. All modern ization theories start with an implicit or explicit refer ence to a dichotomy between two ideal types: the traditional or backward society and the modern society. This was a concern of theorists from at least the mid nineteenth century. Most modernization theorists then argued that these two ideal types of social structure were historically connected through an evolutionary process which followed certain general laws.

The Economic Theories on Modernization

In the 1950s and 1960s the most influential economic theorist on modernization was undoubtedly Walt Rostow. He was one of the few people who combined being an academic with a significant political role – as key advisor to the Kennedy and Johnson administrations of the 1960s (between 1966 and 1969 Rostow held the post of National Security Adviser to the White House). In his biographical chapter, Menzel saw Rostow as a convinced liberal with a missionary like zeal which was expressed equally in his commitment to development and his anticommunism.

Rostow argued that, based on his analysis of the prior experience of the UK, USA, and other industrialized economies, all societies need to pass through five stages of economic growth. The process started with a tradi tional society dominated economically by agriculture and low productivity. The second stage involved achieving the preconditions for takeoff in which Rostow highlighted both political and economic themes – the political con stitution of a national state, the expansion of trade, and an increase in the rate of investment. Rostow saw the evo lution from stage two to stage three (he used an appro priate metaphor of the epoch, takeoff, to describe this stage) as the most problematic; with takeoff economic growth becomes a permanent feature of society, the rate of investment goes up to 10% or more and technological change becomes embedded within society. The next stage, the road to maturity, was conceptualized as a long historical period as every aspect of the economy is modernized and makes use of the new technologies.

A fifth stage of high mass consumption sees the economy more oriented to consumer goods and services. Whether Rostow's contribution constitutes a theory at all rather than merely a taxonomy of economic stages is one of the fundamental criticisms and the empirical validity did not often tally – some countries had elements of takeoff but did not achieve sustained economic growth, for example.

In the 1950s and 1960s there were few 'modernization' theorists who chose to counter Rostow's arguments on economic grounds. One of the few was Alexander Gerschenkron, a Russian Jew who had arrived in the US via Austria in 1938; his Russian and East European heritage was key to understanding his arguments. Gerschenkron argued that economic modernization could be an important force if linked to the concept of backwardness. As other countries modernize and backwardness in a country deepens, the underprivileged society will become increasingly sensitive to the contrast between itself and the successfully modernizing state. He argued that the role of the state, modernizing elites, and new ideologies could galvanize a society toward a 'spurt' of industrialism and accelerated growth.

Gootenberg has subsequently argued that there were two core concepts underlying Gerschenkron's approach. One was the paradoxical advantages of backwardness; countries that tardily industrialized could import advanced scientific and industrial techniques that put them on the cutting edge of world product cycles and competition. Second, there was the concept of substitutes. Looking back on East European industrialization (as opposed to that of the Industrial Revolution in England) Gerschenkron argued that if an element or prerequisite for growth were lacking (such as formal capital markets), these societies were adept at developing new institutions, such as state industrial banks, as a substitute.

Two clear distinctions can be drawn between Rostow's and Gerschenkron's versions of economic modernization. First, Gerschenkron privileged historical discontinuity over linearly conceived continuities. Theories of backwardness could lead to different pathways; relative backwardness could be overcome by novel strategies for accelerated growth. In other words, history could ride roughshod over any neat recipes of prerequisites and takeoffs. Second, Gerschenkron was explicitly relational. Later industrializers were different precisely because they were affected by the first industrializers.

Gerschenkron's relational approach was much more widely accepted by Latin American economic modernizers than the arguments of Rostow – even though (unlike Rostow) Gerschenkron made little attempt to directly transfer his arguments into the development policy arena. This was left to Albert O. Hirschman, an economist of a similar Middle European origin. Hirschman is best known for his theory of unbalanced growth, whereby a pattern of accelerated growth may be stimulated by concentrating development efforts on key industries and locations in lagging regions. Hirschman's books and his conceptualization of economic linkages can be understood as putting Gerschenkron's unorthodox substitutes and stage skipping into the policy framework of development. Furthermore, Hirschman can be seen to make the link with geographies of modernization as he argued that growth must also be geographically unbalanced, taking place through 'master industries' which have the potential to generate linkages in and around the location of a key industrial plant.

Growth thus becomes concentrated in certain regions of a country. Hirschman was optimistic about growth effects trickling down to other regions where industrial investment was not being made; the regional development problem becomes a spatial case of unbalanced growth. It was also presumed that the state would intervene to influence the geographies of unbalanced growth whenever they became serious national concerns.

Whereas Hirschman was distinctly optimistic about the nature of unbalanced growth, Myrdal, another economist with an interest in the spatial implications of economic modernization, was much more pessimistic. One of Myrdal's key concepts for the spatial incidence of activity in free market economies was the 'process of cumulative causation': economic activities become concentrated in certain centers, to the detriment of other spaces within the country and are reinforced by movements of capital, goods, and people. Myrdal called these movements spread (from centers to periphery) and backwash (from periphery to centers) and saw negative backwash effects (such as outflow of surpluses from periphery to centers) as the dominant force creating increasing inequality with the modernization of economies.

The first geographer (or regional development planner) to become involved in these debates was John Friedmann who very much systematized the work of Hirschman and Myrdal and developed a clear exposition of the core–periphery model within economically modernizing countries. His work on regional planning in 1960s Venezuela still provides the classic empirical work on regional development planning within the modernization framework (Venezuela was seen as experiencing Rostovian takeoff at the time). Venezuela's massive oil resources were being funneled into the growth pole of Ciudad Guayana, a new city being built on the basis of steel and aluminum in the country's 'undeveloped' east. This regional investment was seen as a way of reducing the excessive centralization of the economy in Caracas.

A number of quantitative geographers then appeared on the scene to explore the themes of modernization and space. The work of Gould, Riddell, and Soja introduced (respectively) modernization surfaces, trend surface analysis, and the geography of modernization. The work was not only noted for a heavy emphasis on quantitative method but also for a circularity of argument. Modernization was linked to central place theory and there was an inherent tautology behind the quantification. The measurable modernization variables were essentially urban in nature and hence it was no explanatory surprise to find that 'modernization' came through the urban central places of the developing countries in question. The resounding critiques that such prominent though narrow treatments of modernization received have meant that geographers rarely ventured back to discuss the concept of modernization as developed by other social science disciplines, and particularly the versions developed from within sociology.

The Sociological Version

It could be argued that economic versions of modernization and their spatial ramifications give only a partial view of the development process. There are some economic criteria that can have very contradictory effects. The resource conundrum is one example. Significant investment in a nation's natural resources such as coal or oil has been important in guiding the economic pathway of some countries (USA, UK, Germany) but virtually insignificant as a factor in many others (Japan, Taiwan, South Korea). Meanwhile, there are many countries with large scale investments in and ownership of natural resources that have found it difficult to achieve a sustained form of economic development – Venezuela and Nigeria would be two examples. For example, Venezuela, after being seen as a strong candidate for takeoff during the 1960s and 1970s, subsequently became a country fluctuating between decline and stagnation.

In contrast to the economic focus, another strand of modernization theory argued that cultural and social variables are more significant than economic ones. The idea of a transition between two polar types of society (traditional and modern) has a long historical tradition within the social sciences. The work of Max Weber in particular was very influential for the postwar modernizers. By using Weber's classifications one can construct an ideal type of a traditional society, with the predominance of a traditional type of action (action determined by well rooted custom) and a traditional type of authority (whose domination is based on the belief in the everyday routine as an inviolable norm of conduct). Weber made the point of comparing this traditional society with what he called the new 'rational' society.

Parsons extended this Weberian framework by arguing that social relations and roles can be determined in terms of five dimensions which present polar alternatives. These are the 'pattern variables' which help to define the differences between traditional and modern societies and four of them were particularly significant for his analysis:

1. *Affectivity versus affective neutrality in terms of roles*. Essentially this means the difference between a role having immediate gratification in its performance or being affectively neutral.
2. *Ascription versus achievement*. Some roles provide status according to their physical and nonachievable social attributes (class, sex, age, family) while in contrast roles can provide status according to and depending on performance and merit.
3. *Diffusion versus specificity*. Some relations are functionally diffuse in that they cover a series of unspecified dimensions (friendship, family roles) in contrast to others which are functionally specific in that their content is clearly defined (within a firm, for example).
4. *Particularism versus universalism*. Role expectations can be defined for individuals in terms of their 'particular' situation which cannot be transferred (family relations) while other role expectations can be determined as being more 'universal' in character (purchaser and seller within a market transaction).

In this way Parsons argued that every society possesses a set of 'role expectations' or the roles which the individuals within the society expect one another to perform. Parsons noted that in traditional societies roles tended to be affective, ascriptive, diffuse, and particularistic. Meanwhile in modern societies roles can become affectively neutral, performance oriented, specific, and more universal. Each society protects this series of role expectations by its system of rewards and retributions.

Together, the role expectations and the system of rewards and retributions constitute a 'social value system' which has a substantial ability to condition the activities of society's individual members. One interpretation was that mobile societies in which the talented individual could rise by achievement irrespective of background would provide a more suitable social value system for entrepreneurial recruitment than hierarchic, traditional societies. For economic growth to occur within a free market society, it was important that the behavior patterns of capitalism (risk taking, profit motivation, the impersonal nonfamilial organization) should blend with the society's prevailing value system and receive wide social approval. Considerable empirical evidence to support Parsonian logic was provided from such diverse cases as the United States and Japan since the 1870s.

In retrospect, one key problem of Parsonian theory was that the concept of social approval exaggerated the cohesion displayed by societies as they experienced modernization and economic change. This problem was taken up in the Parsonian tradition by an academic who wrote most of his significant sociological work in Spanish and hence has been largely ignored by most critiques of modernization theory. Gino Germani was born in Italy but migrated to Argentina to escape fascism in 1934 (he had been imprisoned for over a year for distributing antifascist propaganda) and wrote his seminal work in Spanish in 1966.

In contrast to Parsons, Gino Germani saw social change within a context of economic growth as asynchronous (or not synchronized). In other words, social institutions, social groups, and social values change at different velocities with economic growth. Thus, social forms coexist within society that belong to different epochs and stages of modernization. Some parts of society remain fairly traditional and backward but exist alongside others which have already become modern. Germani thus argued that the process of modernization is conflictive and is lived as a crisis which divides groups and institutions.

Germani's thesis was threefold and was strongly influenced by Parsons. First, in preindustrial societies most actions are prescriptive whereas in modern societies there is a predominance of elective actions. Second, in traditional societies change tends to be a violation of traditional norms and therefore is abnormal and rare; in an industrial society, on the contrary, change becomes a normal phenomenon. Finally, traditional societies possess an undifferentiated structure with few institutions performing many functions. In modern societies, each function tends to be performed by a specialized institution which results in a highly differentiated structure.

Germani argued that these three main changes of the modernization process must also occur in the sphere of knowledge, science, and technology. Instrumental rationality, separation from theology and philosophy, and increasing specialization must guide the production and development of knowledge. In the sphere of the economy, new specific and autonomous institutions must appear, operating according to principles of rationality and efficiency. Social stratification must change from a system based on inherited attributes to one which works according to norms of achievement and performance. This entails a shift from low levels of social mobility to high levels of social mobility. In Latin America the lack of social mobility was a key theme of the 1960s, a time in which elites were seen to be very much in control in most Latin American societies. The contrasts and comparisons between Italy under Mussolini and Argentina under Peron always fascinated Germani, particularly the conflictive nature of 'modernization' in both cases. The implication from Germani's analysis that there should be an increase in the political participation of the popular strata gained widespread support, particularly within Latin American universities and student political movements.

The key difference between Parsons and Germani was that the latter saw modernization as asynchronic. Germani distinguished several types of asynchrony: geographical (as with cores and peripheries), institutional, intergroup, and motivational. Furthermore, Germani's idea was that due to asynchronic change contemporary developing countries suffer from many cleavages and problems which did not exist in those countries which developed earlier. This gives an interesting link to the Gerschenkronian version of modernization and economic change in that the situation in developing countries is radically different from that which existed in advanced nations in the first stages of their development.

Concluding Remarks

Modernization theory was a body of theory that dates from the 1950s and 1960s.

It undoubtedly had strong ideological undertones and part of its thinking was framed by the exigencies of the Cold War and the perceived division of the world into two competing ideological blocks. However, a number of interesting debates were formed within this period as between Rostovian and Gerschenkronian versions of economic change or the differences between Parsons and Germani over the conflictive or consensual nature of societies experiencing modernization.

Subsequent critiques of modernization theory have tended to identify modernization theory with the unilineal approach of Rostow in the economic sphere or the consensual perspective of Parsons in the sociological. However, both these approaches were severely criticized from within the contemporary framework of modernization theorists, a theme that is 'conveniently' ignored by subsequent critics. One aim of this article was to show that the body of modernization theory was not a consistent set of ideas but one with significant variations of disciplinary emphasis and of policy recommendations.

Furthermore, the questions of modernity and modernization are still crucial areas of teaching and research in countries grappling with the process of modernization. In this context, the approaches set out by Germani and the linked theories and policies of Gerschenkron and Hirschman have had a more long lived influence within the Latin American region than those of Parsons and Rostow. Although the term 'modernity' has tended to replace that of 'modernization', the search for the theoretical and empirical underpinnings of the economic and social forces behind the modernizing process is still significant.

It may be true to say that the strength of modernization theory coincided with the establishment and growth of many key global and regional institutions of development, such as the World Bank in Washington or the United Nations Economic Commission for Latin

America in Santiago. These institutions became strongly framed by some of the key ideas of modernization theory. Examples include the perception that industrialization has a close relationship with development and the importance of establishing national and regional plans to orchestrate the process of industrialization and economic growth. Subsequently these issues have proved the ingredients for part of the postdevelopment critique.

However, by the early twenty first century, it has been the work of the less ideological theorists of the time, such as Gerschenkron and Germani, that have 'lasted' in that they can still inform contemporary theorists, particularly in Latin America. Unfortunately many of the trenchant critiques of modernization theory have conveniently airbrushed out (or ignored) these writers in terms of the canon of modernization theory. While Rostow and Parsons are seen as prime candidates for critique, the more nuanced interpretations of Gerschenkron, Hirschman, and Germani are often not tackled. This could be part of the crisis of 'area studies' in the universities of North America and the UK; the declining status of regional approaches could be seen to dim the chances of a 'Gerschenkronian inspired' political economy, which belongs with geographically and culturally embedded disciplinary traditions.

See also: First World; Postdevelopment; Second World; Third World.

Further Reading

Friedmann, J. (1966). *Regional development policy: A case study of Venezuela*. Cambridge, MA: MIT Press.

Gaile, G. (2006). John Friedmann. In Simon, D. (ed.) *Fifty key thinkers on development*, pp 101 106. London: Routledge.

Germani, G. (1966). *Política y sociedad en una época de transición*. Buenos Aires: Paidos.

Gerschenkron, A. (1962). *Economic backwardness in historical perspective*. Cambridge, MA: Harvard University Press.

Gootenberg, P. (2001). *Hijos* of Dr. Gerschenkron: 'Latecomer' conceptions in Latin American economic history. In Centeno, M. A. & López Alves, F. (eds.) *The other mirror: Grand theory through the lens of Latin America*, pp 55 80. Princeton, NJ: Princeton University Press.

Gould, P. R. (1970). Tanzania, 1920 63: The spatial impress of the modernization process. *World Politics* 22, 149 170.

Gwynne, R. N. (2006). Alexander Gerschenkron. In Simon, D. (ed.) *Fifty key thinkers on development*, pp 116 121. London: Routledge.

Hirschman, A. O. (1958). *The strategy of economic development*. New Haven, CT: Yale University Press.

Larrain, J. (1989). *Theories of development: Capitalism, colonialism and dependency*. Cambridge: Polity Press.

Menzel, U. (2006). Walt Whitman Rostow. In Simon, D. (ed.) *Fifty key thinkers on development*, pp 211 217. London: Routledge.

Parsons, T. (1951). *The social system*. Glencoe, IL: The Free Press.

Preston, P. W. (1996). *Development theory: An introduction*. Oxford: Blackwell.

Radcliffe, S. (2006). Gunnar Myrdal. In Simon, D. (ed.) *Fifty key thinkers on development*, pp 181 187. London: Routledge.

Riddell, J. B. (1970). *The spatial dynamics of modernization in Sierra Leone: Structure, diffusion and response*. Evanston, IL: Northwestern University Press.

Rostow, W. W. (1960). *The stages of economic growth: A non communist manifesto*. Cambridge: Cambridge University Press.

Soja, E. W. (1968). *The geography of modernization in Kenya: A spatial analysis of social, economic and political change*. Syracuse, NY: Syracuse University Press.

Modifiable Areal Unit Problem

D. W. Wong, George Mason University, Fairfax, VA, USA

Glossary

Ecological Fallacy Erroneously inferring individual situation or behavior based upon aggregated data.

Geographically Weighted Regression A form of local regression weighted by neighboring values and the weights are functions of distances to neighbors.

Geostatistics A branch of spatial statistics which focuses on statistical spatial interpolation methods such as kriging.

Raster Data Geographical surfaces and features are represented by layers of grid cells. Each grid layer represents a theme. It is also known as grid data or field data.

Spatial Autocorrelation Correlation of observation values over space. Positive spatial autocorrelation will exhibit a clustering pattern, whereas negative spatial autocorrelation will exhibit a disperse pattern.

Spatial Interpolation The procedure or technique to create a spatial surface based upon values from sampled point locations.

Spatial Resolution The size of spatial unit for which the data are captured or tabulated.

Vector Data Geographical features are represented by the three geometric primitives of point, line, and polygons.

Introduction

Geographical data possess unique characteristics that are not present in aspatial data. One such characteristic is that they are scale dependent. Spatial data can be gathered at, or tabulated for, different spatial resolutions, such as different pixel sizes in remote sensing images, or different census areal units in population census. Therefore, we can partition space into different scale levels. Even if we keep the spatial scale somewhat constant, there are many, or even infinite ways to form regions and data can be tabulated according to these various partitioning schemes. In other words, the boundaries of regions can be modified at different scale levels or according to different zonal patterns at about the same scale, and data can be tabulated according to these different zonal systems. But when data tabulated according to different zonal systems are analyzed, it is unlikely that they will provide consistent results even though the same variables are used and the same areas are analyzed. The variability or inconsistency of the analytical results is mainly due to the fact that we can modify areal unit boundaries and thus the problem is known as the modifiable areal unit problem (MAUP).

One may easily recognize that as long as we do not use individual level data, but use aggregated data, the MAUP will arise. In other social science disciplines, individual observations are often aggregated into groups according to various socioeconomic or demographic variables. The MAUP effect may not be apparent in those cases. But when the observations are grouped by location, the MAUP will emerge. In fact, as long as the data possess location information, the MAUP will be relevant. Therefore, the MAUP is a concern not just in human geography, but also in political science in the context of voting, in sociology when studying population characteristics, and in economics when studying regional economics issues. It is equally applicable to the analysis of physical environment.

Definitions and Effects of the MAUP

The term MAUP was first coined by Stan Openshaw and Peter Taylor when they experienced the range of *r* squared (measure of correlation) that one may obtain by aggregating data to different scale levels and according to different spatial partitioning systems. Specifically, the MAUP refers to two types of variability or inconsistency in data analysis. Smaller areal units can be merged to form larger but fewer areal units. This may occur in social science when a number of census blocks are grouped together to form a census block group, and a number of block groups are grouped to form a census tract in the US census geography. In remote sensing, the raw data captured by the censors may have 30 m resolution, but can be aggregated to form grid cells with 60 m or 90 m resolution. In this case, data are aggregated or tabulated into larger areal unit with lower spatial resolution. Therefore, multiple data sets can be tabulated for different geographical scale levels. The areal units at different scale levels form a geographical hierarchy. The same analysis can be performed using data at different scale levels, but will likely have different results. The variability or inconsistency in the results from data at different scales is known as the scale effect – a subproblem of the MAUP.

Another subproblem of the MAUP is the zoning effect. When the spatial scale is somewhat fixed or the number of areal units is kept constant, the region can be delineated into a number of areal units, but the bound aries of delineating units can be drawn in many ways. Or in more common situations, the boundaries have been drawn already, but over time, the boundaries need to be adjusted while maintaining the original number of units. This often happens in modifying census enumeration units for censuses. When populations of local areas change, the boundaries of census units including voting districts are redrawn to maintain roughly the equal number of people within each unit. The number of areal units stays the same, but the boundaries partitioning the region change. When data sets for the same area, but tabulated according to different partitioning systems, are analyzed, they will likely produce inconsistent results. This is known as the zoning effect, another subproblem of the MAUP.

There is a general impression that zoning effect is more relevant to vector format data than to raster data, but in fact, the zoning effect may affect raster format data in a subtle manner. It is true that in raster data, the boundaries of cells are fixed as long as the cell size is determined, unless irregular cell size such as quadtree or other variable grid formats are used. However, where the grid system is placed over the study region it is not an obvious issue. The grid system can shift slightly in dif ferent directions but still cover the study region ad equately. Multiple grid systems can be used with the same grid size (scale), and data from these systems may not yield the same analysis results.

Since the discovery of the scale effect in 1930s and the formalization of the problems under the MAUP umbrella in 1970s, most research efforts have been focusing on the scope and effects. That is what types of analysis will be subject to the MAUP and the extent of the impacts. With a few exceptions, the general consensus is that the MAUP is quite pervasive across various classical statistical techniques, spatial analytical procedures, spatial models, and spatial statistics. While the MAUP effects are easily recognized in most descriptive statistics, much focus was put on its impacts on correlation analysis. Openshaw and Taylor conducted the monumental experiment, which later resulted in the labeling of the MAUP, and demon strated that data created through various spatial aggre gation methods offered correlation coefficients over a wide range of value. In general, correlation coefficients are smaller (lower correlation) between variables for data tabulated for smaller spatial units (data with higher spatial resolution). Larger spatial units (lower spatial resolution) tend to raise the level of correlation among variables.

The results of the correlation experiment conducted by Openshaw and Taylor have significant implications for many types of quantitative analysis and modeling used in geographical research. Correlation is the basis of almost all classical statistical methods, including regression analysis. Thus, all classical statistics based upon correl ations of variables are subject to the MAUP. Subsequent studies have shown that relationships among variables can switch from one direction (for instance, slightly negative) to another (strong positive) when smaller areal units were aggregated to form larger units.

While the MAUP effects, especially the scale effect on classical statistical analysis, are quite obvious when their impacts on correlation analysis are identified, their im pacts on spatial analytical techniques and spatial models are less transparent. As a result, many geographers have investigated how MAUP can affect different types of spatial models, including location–allocation analysis, spatial interaction models, input–output models, and a few others. The general consensus is that the MAUP effects are pervasive among spatial models. In some cases, spatial versions of certain analyses are more sensitive to the MAUP effects than their nonspatial counterparts.

Sources of the MAUP

Enumerating the impacts of the MAUP is useful in order to recognize the potential pitfalls in spatial analysis. It is, however, necessary to understand what causes the MAUP. In general, one could claim that when individual obser vations (which can be humans or physical objects) are used in the analysis, then the MAUP effects will be un likely to appear. When observations are grouped together, and data are summarized or aggregated for the groups of individuals, the analysis results based upon the grouped data will be different from those obtained from analyzing individual level data.

Nevertheless, there are at least two general ap proaches to group or aggregate individual level data. On the one hand, individuals can be classified into groups according to certain groups of the individuals, such as socioeconomic status or racial–ethnic identity, and these characteristics do not utilize location information. On the other, individuals can be grouped together based upon where they are, or they are grouped by regions to pro duce data at the aggregated levels. But in geographical space, there are many, and theoretically infinite ways to delineate boundaries to form regions, and smaller regions (such as counties in the US) can be aggregated to form larger regions (states). MAUP refers to the latter type of aggregation based upon space and many data in geo graphical research are provided in spatially aggregated formats or sometimes also known as ecological data. Many social science disciplines, such as political science (e.g., in voting), economics (e.g., in regional develop ment), and sociology (e.g., population analysis) also utilize spatially aggregated data frequently.

In general, we can aggregate individual level data into smaller units (more in numbers) or larger units (less in numbers). Analysis results from all these levels will not be consistent, as the information content of the data has changed or been lost through the aggregation process. The ideal situation is that when we aggregate individual observations into a group, all individual observations are identical. Then the group values will represent individual values perfectly. In this case, there is no loss of information content between the individual and aggregated level data. Unfortunately, in reality, individual observations do vary even if we try to group highly similar observations or small regions together. Still, if individual observations are not identical, the aggregated data will lose their accuracy in describing individuals. In other words, how well the aggregated data can describe individual observations is inversely related to the MAUP effects. In social science, this is often known as the ecological fallacy problem, which refers to the inaccuracies of inferring individual situation or behavior based upon aggregated data.

When individual observations are grouped together, they are all represented by some summary measures. Often, measures of central tendency, such as mean or median, will be used. Average height or weight is often used to describe a population, and median household income is often used to reflect the socioeconomic condition of a neighborhood. When one single value is used to represent all individual observations, variation among observations is not captured by the data anymore. When more variation exists among individuals, the aggregated value will be a poorer describer of individual observations. The extreme highs and lows will be ignored at the aggregated level, and they are all replaced by a value somewhere in between. When one examines the aggregated values, they will be less varied than the original values, or the data are smoothed through the aggregation process.

In the spatial context, individuals within the neighborhood are grouped together and aggregated data are used to represent the characteristics of the neighborhood. Neighborhoods in the vicinity can be aggregated together to form regions, and the grouped data at the neighborhood level will be aggregated with other neighborhoods to form data at the region level. Through this spatial hierarchical aggregation process, data at multiple levels can be created. But as the data are aggregated to the higher levels (fewer in numbers of areal units), the data will have less variation as they go through the smoothing process successively. In the US Census, individual observations within a census block are aggregated to provide the census data at the block level. Several blocks can merge together to form a block group, and several block groups will form a census tract. Above the tract level, the census geography includes counties, states, divisions, and regions, following one hierarchical order. We can expect the scale effect when examining census data at different census levels.

The same smoothing process is applicable to explain the zoning effect when numbers of units are about the same, but boundaries are drawn in different ways. Assume that we have different ways to partition the region, and there are some underlying patterns of similarity among observations or subregions. Depending on how the boundaries are drawn, if dissimilar observations are grouped together, we expect a large degree of smoothing. But if the boundaries coincide with the underlying patterns of the observations to group similar observations together, then little smoothing may occur. Therefore, the magnitude of zoning effect depends upon the variation among observations, the spatial pattern of observations, and to what extent the boundaries match the pattern.

In **Figures 1a** and **1b**, observation values are distributed randomly, but two zonal systems are used to aggregate them. Given the randomness of the data, the two zonal systems will still produce different aggregate level data sets, but their difference will probably be small. Significant smoothing will occur within the zonal units. However in **Figures 1c** and **1d**, observation values are distributed in a clustering manner. While the two zonal systems imposed upon them are not dramatically different, the zoning system matches the clustering pattern of the data in **Figure 1c** perfectly, and the zoning system in **Figure 1d** slightly mismatches the pattern. In **Figure 1c**, no smoothing will occur within each zonal unit. However in **Figure 1d**, each zonal unit includes mostly identical observations with a few different observations. Smoothing or averaging will occur within each unit to eliminate the differences. The magnitude of smoothing also depends upon how different the four values are. If they are dramatically different, more smoothing will happen. But if they are just slightly different from each other, the magnitude of smoothing will be small.

To summarize, the magnitude of the MAUP effects is related to the internal homogeneity of areal units. Smaller areal units with identical or very similar values will generate less ‘scale effect’ or less smoothing when they are aggregated to form larger units. Zoning effect will also be minimal when only similar units are grouped to form different zonal patterns. It is to be noted that in spatial statistics, homogeneity or similarity across units is referred to as spatial autocorrelation. Because the degrees of similarity across units affect the MAUP effect, therefore, MAUP and spatial autocorrelation are related.

Potential Solutions

The literature on the MAUP is repleted by studies assessing its impacts. Literature in recent decades has switched the focus to identify solutions to the MAUP. In early years, two somewhat extreme approaches were suggested. The core issue of the MAUP is the

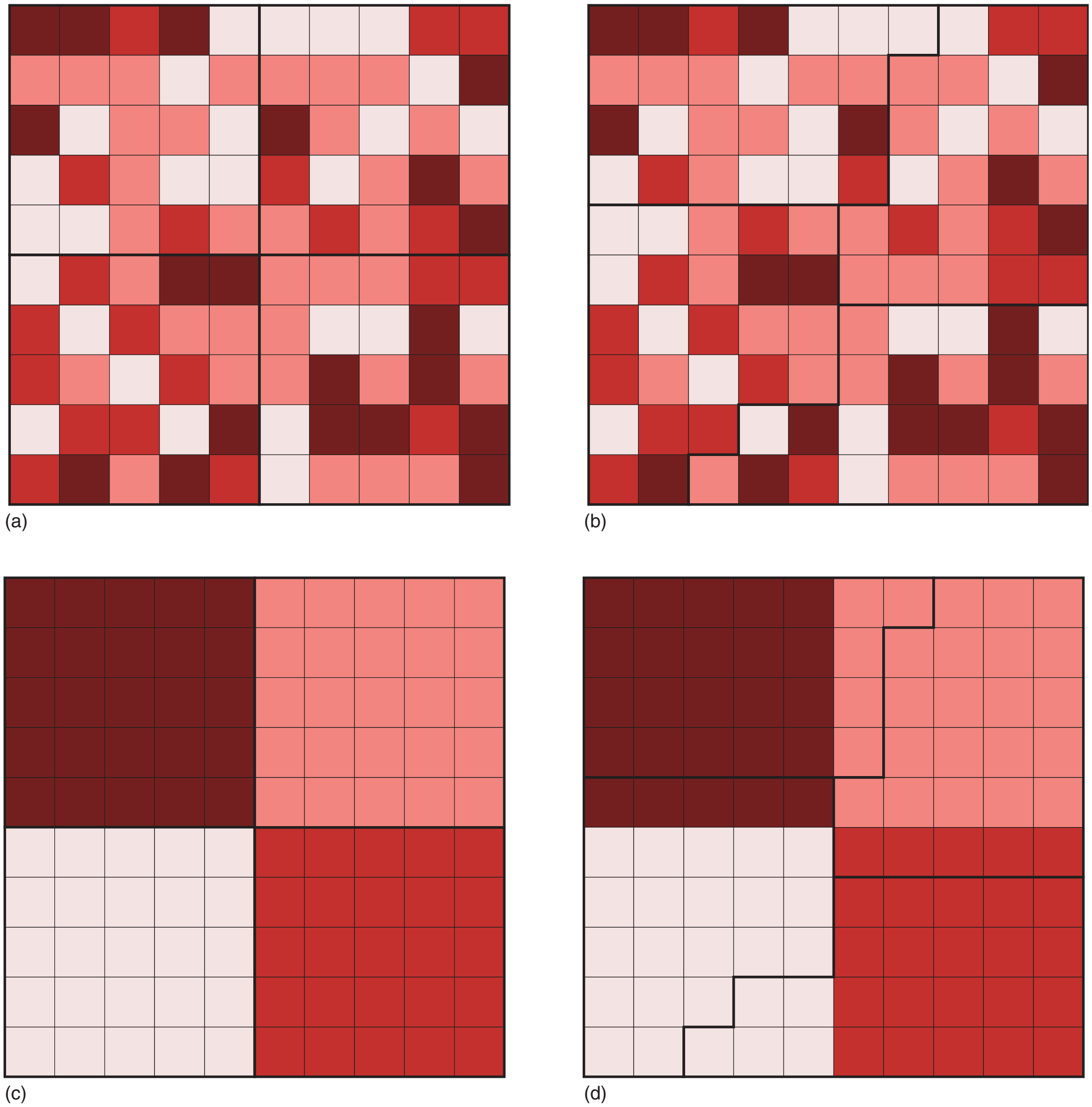

Figure 1 (a, b) Random distribution of observation values with two zoning systems; (c, d) clustering of observation values with two zoning systems.

inconsistency in analytical results when data are tabu lated at different scales or zonal systems. The ideal goal is to obtain consistent results regardless of the scale or zonal patterns. It is quite clear that data at different aggregation levels or for different zoning systems will be different due to spatial smoothing, and there is not much the analyst can do about it. Then, the burden lies on finding ana lytical techniques that are frame or scale independent. The other extreme is to admit that the MAUP does exist and currently we do not have methods to handle it successfully. At the minimum, we should acknowledge its existence by showing the variability of the results when data at different scales or from different zonal systems are used. The spirit is to show the uncertainty of results due to the spatial nature of the data.

Between these two extremes, many attempts and ap proaches have been investigated to address the MAUP. Some researchers argued that among the two sub problems of the MAUP, zoning effect is easier to handle as it can be treated as a spatial interpolation problem. Two different zonal systems for the same area will give different analysis results. In order to resolve the in consistency, we may use spatial interpolation technique to estimate the data on a new zonal pattern based upon the two existing sets of zonal data. On the other hand, scale effect has been the focus of many endeavors to handle the MAUP. Developing scale independent or scale insensitive spatial analytical techniques is a noble call, but so far has limited success. Only a few methods seem to demonstrate such property to some extent, but they were developed for specific applications and have limited usage. The migration model suggested by Waldo Tobler is one of the few successful attempts. It is very unlikely that we will be able to develop general tools or models that are scale independent.

Instead of developing techniques to tackle the MAUP, another approach is to deal with the data. The MAUP stems from the fact that data can be aggregated spatial in many ways. The question is which aggregation is the most preferable based upon certain criteria. The approach is characterized by developing optimal zoning systems according to different modeling objectives or criteria. For instance, if the objective is to identify a fixed number of depots to minimize the total delivery cost to a set of census tracts as represented by the center points of the tracts, the optimal zoning approach will identify other possible zonal patterns comparable to the census tract boundaries, but can offer the lowest total delivery cost. In general, techniques associated with this approach have been quite mature and are operational. However, when different criteria are used, different zoning systems have to be developed.

Some researchers adopted a statistical approach to model the scale effect by incorporating the correlations between units within the zone or region. The essence of this approach, as demonstrated by the correlation experiment by Openshaw and Taylor, and the fact that spatial smoothing as the source of the MAUP, is to capture the correlation within zones or regions at different scale levels. These multiple levels of correlation are taken into account in the statistical models when aggregated data are analyzed statistically. The concept is sound, but computationally very intensive and difficult to track. As a result, it does not offer a practical solution. In between searching for scale independent spatial analytical techniques and merely acknowledging the presence of the MAUP, a middle ground approach is to document the sources of the MAUP. This is very much a spatial accounting framework with the objective of identifying where in the region the discrepancies arise at different scales and how much these are attributable to various locations. In other words, this approach intends to identify where and how much spatial smoothing takes place in the hierarchy of zonal system from small units to larger units. This approach has been demonstrated in the study of measuring segregation at different spatial scales and has the potential to be expanded to other applications.

Because ecological data are commonly used, the MAUP is also known in other disciplines. In political science, especially in the voting literature, the MAUP is often discussed in the context of ecological fallacy and the error bound technique has been suggested to handle the ecological fallacy issue by providing a margin of error when inferring individual situations based upon aggregated data. Thus, the method provides information about the magnitude of possible error in using aggregated data while the results are intended to explain individual behavior. This method indirectly deals with the MAUP. However, most geographers who are knowledgeable in the MAUP are not optimistic about this method.

Recent studies also suggest identifying the geographical extent or spatial scale of homogeneous areas in order to model the scale effect. This direction very much applies geostatistical techniques, including variogram analysis and kriging. By examining the variogram, we may be able to identify the spatial extent (range in the variogram) of areas with similar values. Thus, the argument is that creating regions smaller than the spatial extent of homogenous areas may minimize the scale effect. Because of the raster or grid nature of the analysis, this approach has raised some interests among remote sensing scientists and geoscientists. On the other hand, to better understand the pattern of spatial heterogeneity (the dual of homogeneity), it was suggested that geographically weighted regression (GWR) may have the potential to help address the MAUP problem. GWR can derive a set of parameter estimates of a regression model for each areal unit. By examining the spatial variation of these parameters, we may obtain some insights on how spatially aggregating these units may create the MAUP.

Conclusions and Future Directions

In short, the MAUP effects are pervasive, and their magnitudes vary, depending upon different spatial variables. Many attempts have been made to handle the MAUP effects, but so far, we have limited success. It is the first step to acknowledge the fact that analysis result based upon one scale level or one zoning system is just one out of many possible outcomes. Searching for results that are stable over geographical scales or across different zoning systems is still an admirable goal. However, it is unlikely that the generic solutions will be available across different types of analysis or across disciplines. But because the nature of the problem is very much the same regardless of the analytical techniques adopted or the context of analysis, specific solutions or techniques for different contexts dealing with the MAUP have to be built upon the conceptual understanding of the MAUP as discussed here.

See also: Ecological Fallacy; Scale Analytical; Spatial Autocorrelation.

Further Reading

Arbia, G. (1989). *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems*. Dordrecht: Kluwer.

Fisher, P. F. and Langford, M. (1995). Modelling the errors in areal interpolation between zonal systems by Monte Carlo simulation. *Environment and Planning A* 27, 211 224.

Fotheringham, A. S. (1989). Scale independent spatial analysis. In Goodchild, M. F. & Gopal, S. (eds.) *Accuracy of Spatial Databases*, pp 221 228. London: Taylor and Francis.

Fotheringham, A. S., Brunsdon, C. and Charlton, M. E. (2000). *Quantitative Geography: Perspectives on Spatial Analysis*. London: Sage.

Fotheringham, A. S. and Wong, D. W. S. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A* 23, 1025 1044.

King, G. (1997). *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton, NJ: Princeton University Press.

Openshaw, S. (1984). *The Modifiable Areal Unit Problem*. *CATMOG*, 38. Norwich: Geobooks.

Openshaw, S. and Taylor, P. J. (1979). A million or so correlation coefficients: Three experiments on the modifiable areal unit problem. In Wrigley, N. (ed.) *Statistical Applications in the Spatial Sciences*, pp 127 144. Pion: London.

Quattrochi, D. A. and Goodchild, M. F. (eds.) (1997). *Scale in Remote Sensing and GIS*. CRC Press. Boca Raton, FL.

Tate, N. J. and Atkinson, P. M. (eds.) (2001). *Modelling Scale in Geographical Information Sciences*. Wiley and Sons: London.

Tobler, W. R. (1989). Frame independent spatial analysis. In Goodchild, M. F. & Gopal, S. (eds.) *Accuracy of Spatial Databases*, pp 115 122. Taylor and Francis: London.

Wong, D. W. (1995). Aggregation effects in geo referenced data. In Arlinghaus, S. L. & Griffith, D. A. (eds.) *Practical Handbook of Spatial Statistics*. CRC Press: Boca Raton, FL.

Wong, D. W. (2003). Spatial decomposition of segregation Indices: A framework toward measuring segregation at multiple levels. *Geographical Analysis* 35, 179 194.

Wong, D. W. (2006). Ecological fallacy. In Warf, B. (ed.) *Encyclopedia of Human Geography*, pp 117 118. Sage Publications: Thousand Oaks, CA.

Wong, D. W. and Amrhein, C. G. (1996). Research on the MAUP: Old wine in a new bottle or real breakthrough? *Geographical Systems* 3, 73 77.

Monte Carlo Simulation

S. Ratick, Clark University, Worcester, MA, USA
G. Schwarz, Universidad Nacional de Colombia, Bogota, DC, Colombia

Glossary

Central Limit Theorem It states that the sum of a large number of independent random variables has a Gaussian distribution. For this reason, many phenomena in the natural and social sciences can be modeled using a Gaussian distribution.
Conditional Joint Distribution The probability distribution of a collection of random variables conditional on the (i.e., given) values of one or more related random variables.
Correlation The covariance between two variables normalized to have values between –1 and 1 (similarly for spatial correlation).
Covariance It is a measure of the degree to which two random variables vary together: positive covariance implies that when one variable is above its mean the other variable tends to be above its mean (tends to be below its mean for negative covariance).
Cumulative Distribution Function (CDF) It completely describes the probability distribution of a random variable. For continuous random variables, the CDF at a particular value of the random variable is equal to the integral of the probability density function (PDF) from negative infinity to that value. The CDF is 0 at negative infinity and 1 at positive infinity.
Gaussian or Normal Distribution A type of PDF that is of great importance in the natural and social sciences because of the central limit theorem.
Inferential Statistics It uses a sample obtained from a population to make inferences about the true but unknown characteristics of that population.
Latin Hypercube Sampling A stratified sampling technique designed to ensure an even representation of the entire probability distribution in the sample for a Monte Carlo simulation.
Map Interpolation A method of calculating new map data points from a set of known map data points (usual methods involve weighted combinations of the known points).
Probability Density Function (PDF) It describes the relationship between the values of continuous random variables and corresponding probabilities, expressed as the area (integral) below the PDF. The integral over all values of the random variable is 1.
Probability Distribution Function It assigns each random event (value of a discrete random variable) a probability between 0 and 1. The sum of all probabilities is 1.
Pseudo-Random Numbers The random numbers generated by a software algorithm. In Monte Carlo simulations, pseudo-random numbers are used to draw samples from probability distributions.
Random Field Mapped values of random numbers; these can be spatially correlated suggesting that adjacent values may not differ as much as those further apart.
Random Sample A sampling method that ensures that each member of the population has an equal chance of becoming part of the sample.
Sampling Distribution The distribution of a statistic that occurs when the statistic is calculated a number of times for a given population.
Spatial Covariance The covariance of two random variables (or two values of the same variable) that are a specified distance apart. A related measure is the semi-variogram.
Variance It is a measure of the spread of a random variable about its mean.

History of the Monte Carlo Method

The Monte Carlo method can be used to address any mathematical problem or model that is too complex, time consuming, or resource intensive to solve analytically. Instead of tackling the numerical problem directly, the Monte Carlo method allows the researcher to obtain an approximation of the solution through setting up an experiment of statistical sampling. As the name indicates, the method borrows from games of chance such as those played at the famous casinos of Monte Carlo in Monaco.

The formal development of the Monte Carlo method dates to World War II, when physicists working on the atomic bomb in the Manhattan Project needed to analyze neutron transport in fissile material. While these multidimensional problems were too complex to solve analytically, John von Neumann and Stanislaw Ulam suggested that experiments using random samples and executed on the then newly available computers may provide sufficiently accurate approximations to the desired solutions. Of course, mathematicians in the seventeenth and eighteenth century had already developed similar ideas by observing games of chance and repeating the corresponding experiments. In this respect, Ulam and von Neumann were formalizing a research tool based on a long standing mathematical tradition that dealt with the analysis of random events.

The spread of the Monte Carlo method was aided by two developments since the 1970s. First, theoretical work on computational complexity showed that the Monte Carlo method is not just capable of providing an ap proximation that is within a specified statistical accuracy of the exact solution, but that for a large class of problems this approximate result can always be obtained in less time than the exact solution. Second, the rapid devel opment in computing power gave researchers the tools to implement more elaborate Monte Carlo simulations for models in a variety of areas that previously could not be tackled mathematically.

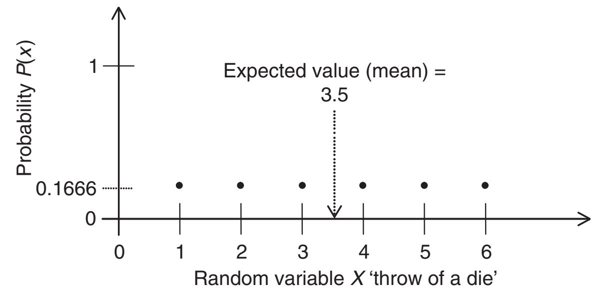

Figure 1 Probability distribution function for throw of a die.

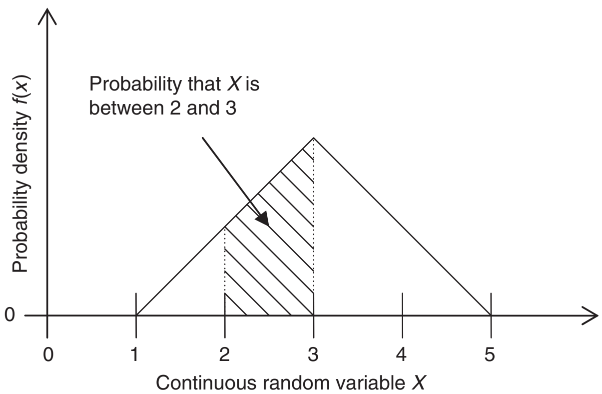

Figure 2 Probability density function for a triangular distribution.

Principles of the Monte Carlo Method

Uncertainty and Probabilities

Games of chance and the applications of the Monte Carlo method in geography have one central feature in common: the outcome of the system under study is not known in advance, that is, there is uncertainty involved. At the throw of a die, for instance, the result is not known before the die comes to rest on the table. What is known, however, are the possible outcomes of this experiment, namely, the num bers one through six. Theoretically, one could attempt to measure the precise direction, angle, and speed of the throw; the material, weight, and balance of the die; the friction of the surface and the air; and many more par ameters to calculate the result – but that attempt is bound to be too complicated and doomed for failure. Instead, statistical theory is employed to analyze the random ex periment 'throw of a die' and make statements about the probabilities of its outcomes. In a similar fashion, the Monte Carlo method is used to gain knowledge about the out comes of a wide variety of systems that involve uncertainty.

Unless the die is rigged, the probability of each of the six numbers occurring should intuitively be one sixths or 16.667%. The sum of the probability of all simple events is always 100%, since something must happen with the die. In other words, each possible event is assigned a probability value between 0 and 1, indicating the likeli hood of occurrence. The mathematical function that represents this assignment is called a probability distri bution function (**Figure 1**).

In general terms, a probability distribution function describes the relationship between the values of a random variable and the corresponding probabilities. The ex pected value of a random variable, also called the mean of the distribution, is the average result of the random variable after conducting the experiment an infinite number of times. While it is impossible to empirically conduct an experiment until infinity, throwing the die a sufficiently large number of times and calculating the average value of the results should come reasonably close to the true expected value: since all six outcomes are equally likely, the average should be 3.5.

While the experiment 'throw of a die' has discrete outcomes and the corresponding random variable X can only assume integer values from 1 to 6, statistical theory generally deals with continuous random variables that can assume any value. Here, the distribution function becomes a 'probability density function' (PDF), but the principle remains the same: a PDF maps events to probabilities, which are represented as the area below the density function (**Figure 2**). The sum of all probabilities over all possible simple events must be 1. Mathematically, this is equivalent to a PDF's integral being 1.

If we graph the probabilities of the random variable taking on a value less than or equal to a specific value x, $P(X \leq x)$, we obtain the 'cumulative distribution function' (CDF). This is equal to the integral of the PDF from negative infinity to x. The CDF completely describes the probability distribution of a random variable, and can be used to directly read off the probability of a random event (**Figure 3**).

A very common PDF is the 'Gaussian' or 'normal distribution' in the form of a bell curve (**Figure 4**). Its ubiquity in statistics stems from the 'central limit theo rem', according to which the sum of a large number of independent random variables, no matter what their in dividual PDFs, is distributed according to the Gaussian distribution. On this basis, many applications of the Monte Carlo method in geography assume that the input random variables have a Gaussian probability distribution.

Depending on the precise form of the PDF, the ex pected value or mean can be very hard to calculate in advance. Instead, it may be more practical to simply run the experiment a sufficiently large number of times, note the values (realizations) of the random variable, and calculate their average. This average serves as an estimate for the true expected value. Estimating true values from a random sample in this fashion is the basis of inferential statistics. Applying the Monte Carlo method thus in volves drawing a sample from a random variable (with a certain PDF) and calculating the desired output, in this case the average as an estimator for the expected value. Usually, the Monte Carlo method is used to calculate output values that are much more complex than simple averages. Similarly, the method is often applied to more than one random variable and PDF as inputs. The ac curacy of the estimator obtained through the Monte Carlo method depends only on the number of real izations, that is, the size of the sample. In the case of estimating the mean, for instance, the standard error of the estimate decreases with the square root of the sample size.

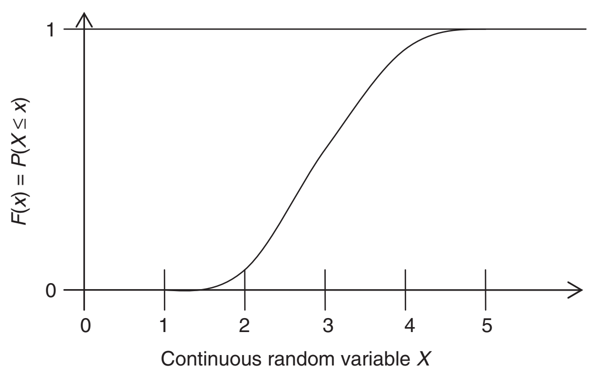

Figure 3 Cumulative distribution function for a triangular distribution.

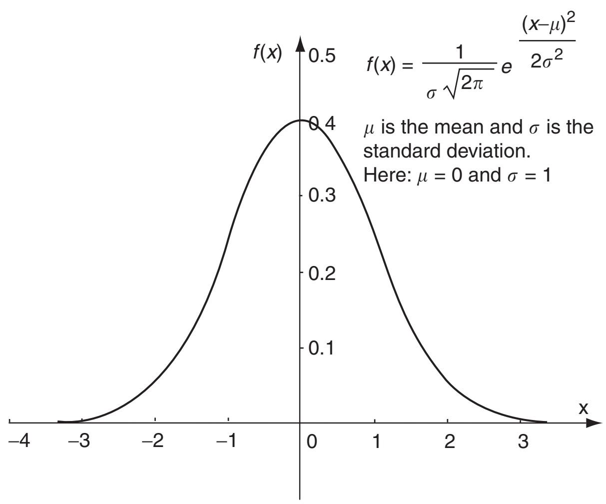

Figure 4 The Gaussian probability density function.

Statistical Sampling

The Monte Carlo method relies on realizations (draws) from a PDF. Ideally, to correctly apply the Monte Carlo method and obtain valid results, the sampling method employed should be completely random. In a random sample, each draw must be independent of every other draw, that is, there must be no correlation between samples. Previous observations of the random variable have no bearing on future draws. The number of real izations has to be sufficiently large to accurately repre sent the distribution of the input variables. If the experiment involves the throw of a die, this condition is easily fulfilled by throwing the same die a large number of times. If the Monte Carlo method is to be applied to a continuous PDF that only exists as a software model in a computer, as is usually the case, it is considerably harder to obtain truly independent observations of the random variable.

Computers are inherently deterministic machines, so to apply the Monte Carlo method in the framework of a computer based simulation model is not trivial with re gard to sampling. Indeed, implementing a good random number generator has been one of the main challenges of computer engineering. As true random numbers can only be implemented as physical processes, the term pseudo random numbers is used to designate random numbers that are generated by software functions. They are not completely random, since the sequence of numbers only depends on the algorithm and a seed value, and is otherwise deterministic. Several solutions exist to obtain a seed value; for instance, current processor load or computer temperature may be used. Modern computer algorithms are then able to provide pseudo random numbers that are statistically indistinguishable from a true random sample of the desired PDF. Their quality with respect to patterns and repetitions varies consider ably between programming languages, however.

Another sampling technique used in the context of the Monte Carlo method is Latin hypercube sampling. This stratified sampling method is useful to reduce the size of the necessary sample, but still accurately depicts the PDF of the input random variables. To obtain a Latin hypercube sample, the range of values for each random variable is divided into a number of sections equal to the size of the sample desired. Within these intervals, one value is randomly drawn in accordance with the segment of the PDF within that interval. The process is repeated in each interval and for each input random variable (**Figure 5**). As a final step, the samples of each variable are randomly combined to form the set of input values for a specific realization of the Monte Carlo simulation. For instance, the first set of values might be made up of the sample of interval 1 of the first random variable X_1 and of interval 8 of the second random variable X_2. In this way, a Latin hypercube sample helps to ensure that the Monte Carlo simulation takes account of the entire length of the distribution, including unlikely but extreme values in a smooth representation of the distribution.

Applying the Monte Carlo Method: An Example

The Monte Carlo method can be used to analyze a wide variety of phenomena in geography. Consider, for example, the gravity model, which aims to represent interactions between two cities. In analogy with Newton's law of gravitation, the model supposes that two objects attract each other in proportion to their masses and in inverse proportion to the distance separating them. For cities, the mass is usually represented by population, while the interactions under study can include traffic or trade. Mathematically, the model looks as follows:

$$F_{ij} = kP_iP_jd_{ij}^{-a}$$

where i and j are the two cities, F the traffic between them, P their population sizes, and d the distance between them. Constants k and a are used to adjust the model to a particular situation, where parameter a determines the strength of the distance decay in the interaction. In the original gravity model, k is 1 and a is 2. **Table 1** shows the parameter values used in this example, where two cities were 45 km apart and parameter k was kept constant at 1. The remaining parameters P (city size) and a were the subjects of the Monte Carlo simulation, and **Table 1** shows the mean values of their probability distribution.

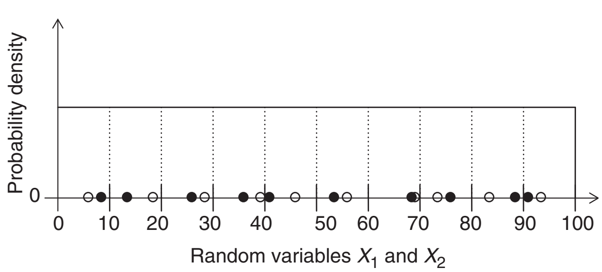

Figure 5 Latin hypercube sampling. ○, samples for X_1; ●, samples for X_2.

City sizes P_i and P_j had Gaussian PDFs, with standard deviations (a measure of the spread of the distribution around the mean) set to 10% of the mean, that is, 3000 for P_i and 4000 for P_j (**Figures 6** and **7**). The distance decay parameter (a) had a triangular distribution with average 2; minimum 1.8 and maximum 2.2 (**Figure 8**).

Table 1 Parameter values and means for the Monte Carlo simulation example

k	1
d	45
a	2
P_i	30 000
P_j	40 000

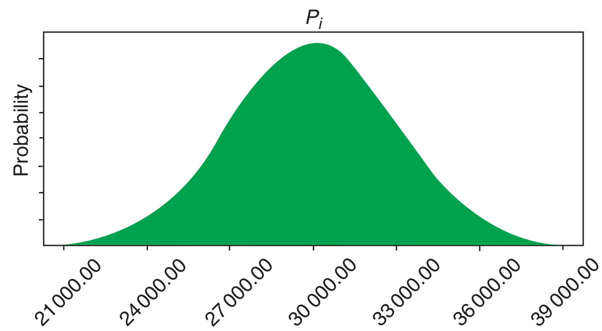

Figure 6 Probability density function of P_i.

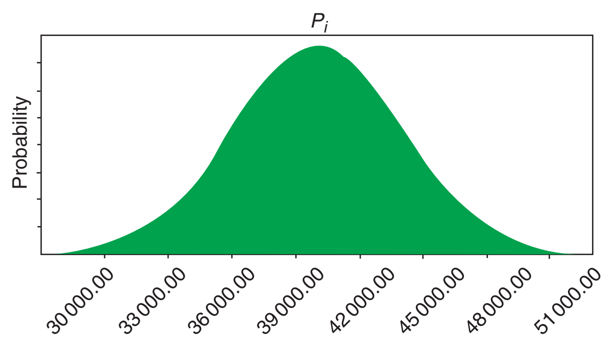

Figure 7 Probability density function of P_j.

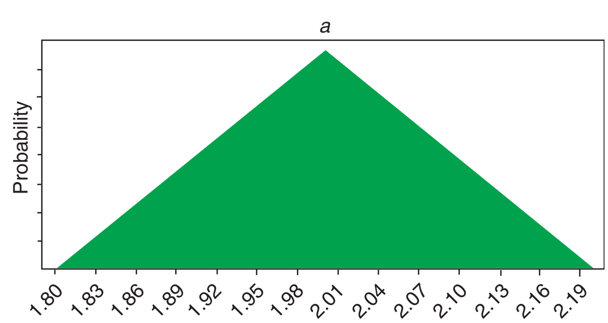

Figure 8 Probability density function of parameter a.

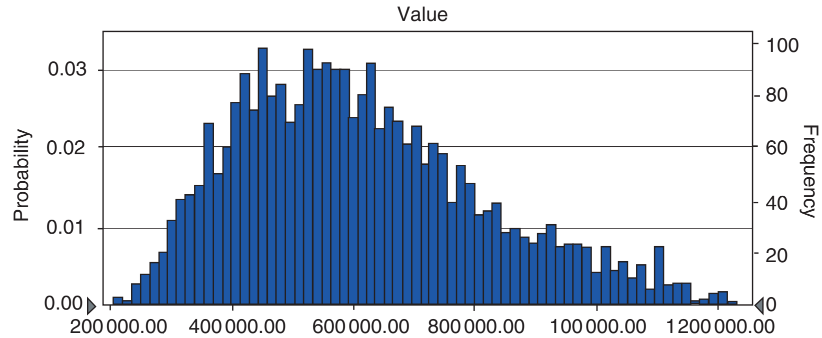

Figure 9 Distribution of interaction values F_{ij} (traffic) after 3000 Monte Carlo runs.

Having defined these three input PDFs, the Monte Carlo simulation consists in drawing a large number of random samples from each of these and calculating the output value for each sample: in this case the traffic F_{ij} between cities. **Figure 9** shows the distribution of output values that the Monte Carlo method produced after 3000 such draws, conducted on a computer using pseudo random numbers.

The mean of this distribution is 622 315.77, and its standard deviation is 216 933.48. As is apparent from **Figure 9**, the distribution is not symmetrical, so that the median value is 587 328.44, and thus differs from the mean. The minimum traffic value is 205 164.62 and the maximum 1 565 252.37, showing the considerable spread of this distribution. Two cities with populations of around 30 000 and 40 000 and a slightly varying distance decay parameter between 1.8 and 2.2 can thus show quite different levels of traffic or other interactions between them. Despite the simplicity of the gravitation model's assumptions, the Monte Carlo method shows the wide variety of interaction values that can result in practice.

The Monte Carlo method made it possible to obtain this distribution of output values with relative ease. It can be applied to models of any complexity and with any number of input random variables. In this example, the output distribution could be obtained through calculus, but with many more complex models, Monte Carlo simulation is often the only way to gain a full overview of the outcomes of a system.

Applications of the Monte Carlo Method

There have been numerous and diverse applications of Monte Carlo methods that include, for example, science (computing multidimensional integrals and model sensitivity analysis), education (teaching and research), business (portfolio management and product life cycle analysis), environment (probabilistic risk assessment and resource allocation), health (delivery of services and epidemiology), government (project choice and standard setting), engineering (design and project management), and energy (utility management and methods for hydrogen production). Application areas seem limited only by imagination and computing power. We focus, here, on a few specific applications to help illustrate the use of the technique in geography, and to highlight some of the issues in its implementation.

Environmental Risks and the Use of Probabilistic Risk Assessment

Monte Carlo methods have been widely used to perform environmental probabilistic risks assessments (PRAs). This is because estimating the environmental risks to which people are exposed, and their associated con sequences, often requires the mathematical combination of numerous factors and processes that may be beyond the capability to calculate in a closed form analytical solution. The complexities of these interactions, and the variability (e.g., variations in the exposure and sensitivity of populations to environmental pollutants) and un certainty (e.g., our lack of knowledge about important models, parameters, or human behavior) inherent in the process can be addressed within the Monte Carlo framework (**Figure 10**). Both variability and uncertainty can be expressed as PDFs.

The random variables (V_1, V_2, …, V_n) represent population exposure variables (e.g., body weight, the frequency, and type of exposure to environmental pol lutants) that are combined in a Monte Carlo analysis to provide a measure of the distribution of risk (a PDF showing the risks to different proportions of the popu lation). These types of risk distributions have been used to help set standards for pollution control (**Figure 11**) or to determine the need for remedial action. Decisions about response can be made with respect to the determination of what may be a reasonable exposure of the population to the risks of the site (if the 95th per centile is considered reasonable exposure, then action would be taken in case A where more people will be exposed to the risk, and not at B where fewer will).

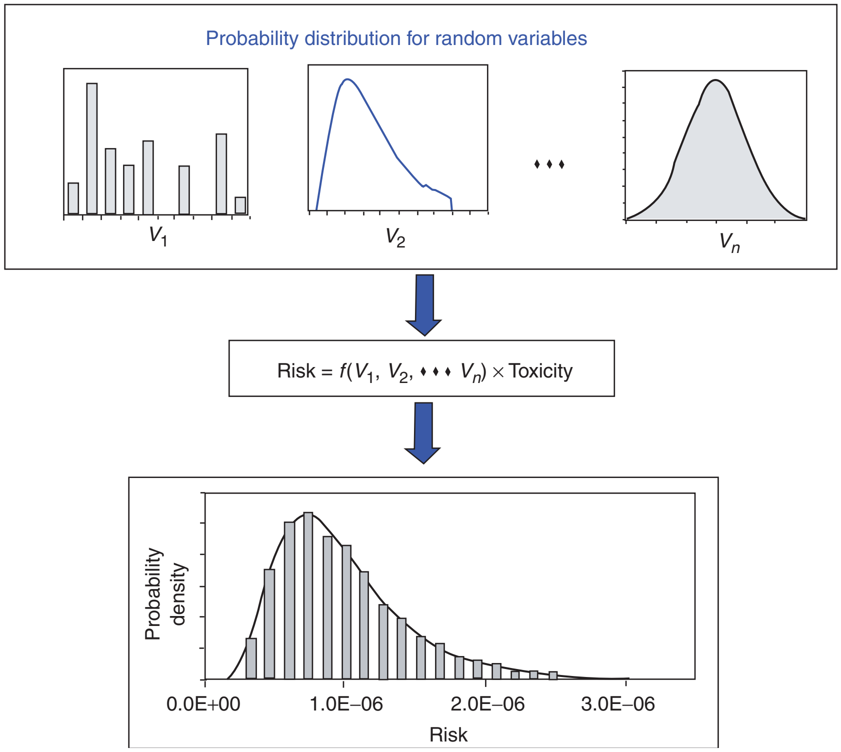

Figure 10 Schematic of a Monte Carlo analysis for probabilistic risks assessment.

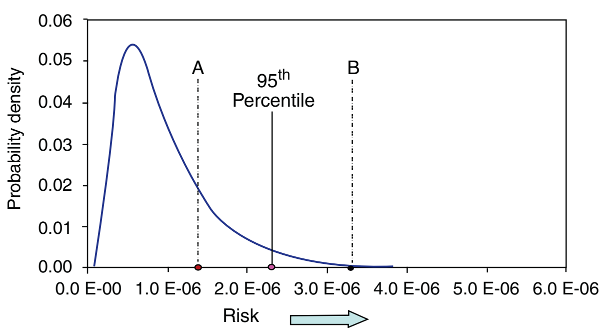

Figure 11 Using the results of a probabilistic risks assessment with Monte Carlo analysis to determine if response to the risk is warranted.

The generic application of the PRA can be used for other types of hazards: floods, fires, and acts of malice, for example. The benefits of conducting a PRA through Monte Carlo analysis have been expressed as providing more complete information about the characterization of the variability of exposure to risks, providing quantitative information about confidence levels for risk estimates (reflecting uncertainty), the ability to help determine important factors through sensitivity analysis, and the ability to assess the value of additional information to the decisions about standard setting and response to environmental risks. Concerns about the use of Monte Carlo methods in PRA include: the potential lack of transparency in the workings of the model, the consequences of hidden assumptions about the choices of the PDFs for specific parameters, and the decision makers' lack of familiarity with the technique.

Bootstrapping

Bootstrapping is a type of statistical resampling that employs Monte Carlo methods applied to observed data. This computer intensive method can be used to deter mine the sampling distribution of a summary statistic (mean, median, and standard deviation) or relationship (correlation and regression coefficient) when these sam pling distributions are extremely difficult to obtain ana lytically. In particular, bootstrapping can, in principle, generate an estimate of both the bias and standard error for any statistical parameter of interest. In geostatistics, bootstrapping is often used as a basis to calculate stat istical significance and confidence intervals for examining spatial autocorrelation and testing for spatial clustering.

There are two main ways in which the bootstrap technique can be applied. In the 'parametric' method, a knowledge of the data's distributional form (e.g., Gauss ian and exponential) is required. When it is available, such distributional information contributes to greater precision. In the more common 'nonparametric' version of bootstrapping, no distributional assumption is needed.

Bootstrapping involves drawing a series of random samples from the original sample – with replacement – a very large number of times. The statistic or relationship of interest is then calculated for each of the bootstrap samples. The resulting distribution of the calculated values then provides an estimate of the sampling distribution of the statistic or relationship of interest.

As an example, **Table 2** lists the original random sample of 20 individuals used to estimate the correlation between income levels and years of education, and the first two of 10 000 bootstrap samples drawn from the original sample. Notice that the bootstrap samples do not contain all the original observations, but do repeat some of the original data points. **Figure 12** shows a histogram of the sampling distribution for the correlation coefficient from the 10 000 bootstrap samples. From this distribution, it is possible to estimate the bias and standard error of the estimated correlation coefficient (0.627) calculated from the original sample. For example, to estimate bias we can calculate the mean value of the correlation based on the distribution of bootstrap samples (0.614), and subtract the original correlation; this results in an estimated bias of 0.013.

The development of powerful computing capabilities and specialized software for sampling has fostered more widespread use of bootstrapping. As a result, limitations on the number of bootstrap samples have all but disappeared, and no longer pose an obstacle to obtaining any desired level of precision. The only real limitation is the size of the original sample (e.g., 20 in our illustration). As the sample size increases, not only will the estimated parameter become more accurate, but the bootstrap empirical distribution will also better represent the true underlying distribution of the population being studied.

Sensitivity Analysis in Managing Model Uncertainty

Many applications of computer based models rely on sensitivity analysis to determine the effects of uncertain parameters or model structures on the output of those models. The results of sensitivity analysis may be used to help calibrate or validate models, to identify parameters for which additional research and measurement should be undertaken, and to help determine how a model may be restructured. Sensitivity analysis is performed through random sampling of distributions representing the uncertainty in the model parameters. Model 'uncertainty' is used to describe a wide variety of issues with model construction and use such as: measurement error, variability, scaling and aggregation, and model structure (**Table 3**). The probability distributions describing uncertainties are developed either objectively (e.g., through observation and measurement) or subjectively (e.g., through the conjoined opinions of experts). In Monte Carlo sensitivity analysis, parameters may be varied singly or in combination. A value is drawn from the distribution for each parameter under study and then input into the model. The results of model runs are compared (e.g., through statistical analysis such as correlation analysis) to produce the contribution of the parameters to the obtained distribution of the model outputs (similar to the distribution results for the gravity model example above in response to the probability distribution of the input parameters).

Table 2 Original and first two bootstrap samples

Correlation 0.627			*Correlation 0.619*			*Correlation 0.402*		
Original sample	*Years of education*	*Annual income ($)*	*First bootstrap sample*	*Years of education*	*Annual income ($)*	*Second bootstrap sample*	*Years of education*	*Annual income ($)*
1	16.3	48 170	5	21.8	69 087	13	17.3	61 110
2	12.7	50 480	4	11.6	53 489	6	16.1	50 896
3	11.0	35 672	2	12.7	50 480	20	20.1	56 182
4	11.6	53 489	9	17.5	57 071	2	12.7	50 480
5	21.8	69 087	11	16.9	45 516	5	21.8	69 087
6	16.1	50 896	8	20.4	90 559	14	16.7	55 763
7	14.5	34 034	18	21.1	86 769	9	17.5	57 071
8	20.4	90 559	11	16.9	45 516	6	16.1	50 896
9	17.5	57 071	11	16.9	45 516	13	17.3	61 110
10	20.6	63 761	16	12.6	34 020	21	20.1	56 182
11	16.9	45 516	4	11.6	53 489	9	17.5	57 071
12	14.2	50 451	7	14.5	34 034	11	16.9	45 516
13	17.3	61 110	1	16.3	48 170	5	21.8	69 087
14	16.7	55 763	17	13.6	54 394	13	17.3	61 110
15	12.8	78 025	7	14.5	34 034	5	21.8	69 087
16	12.6	34 020	16	12.6	34 020	6	16.1	50 896
17	13.6	54 394	2	12.7	50 480	12	14.2	50 451
18	21.1	86 769	10	20.6	63 761	15	12.8	78 025
19	21.3	85 408	4	11.6	53 489	19	21.3	85 408
20	20.1	56 182	17	13.6	54 394	20	20.1	56 182

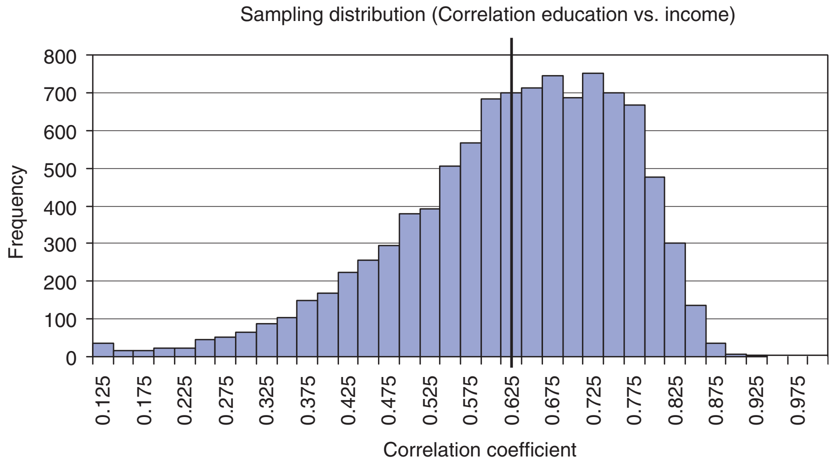

Figure 12 Sampling distribution for the correlation coefficient (10 000 Bootstrap samples) line represents the correlation coefficient for the original sample.

Table 3 Sources of uncertainty

Source	*Description*	*Examples*
Measurement error	Random error (precision), systematic error (bias)	Bias in historical measurements of speed of light, or mean density of Earth
Variability	Systematic differences (i.e., over space or over time), inherent randomness	Spatial variations in precipitation, unpredictability of weather
Model structure	Functional form of relationship (linear or nonlinear, specified predictor variables)	Crop climate model (linear or quadratic, missing climate variables)
Scaling/ aggregation	Spatial or temporal scale (or level of aggregation) of model or data	Tuning model parameters to spatial or temporal scales. Models developed at national scale are used for regional analysis

Reproduced from: Katz, R. W. (2002). Techniques for estimating uncertainty in climate change scenarios and impact studies. *Climate Research* 20, 169.

Conditional Simulation

Many map interpolation techniques (e.g., nearest neighbor polygonal, triangular, inverse distance, and kriging) that are used to provide estimates for unsampled locations, produce maps whose spatial description is often different from the real map, and from the sample points used to create the estimated map. One problem with using the map of estimated values as an estimate of the map of true values is that, because of the interpolation method, the former is often smoother than the latter. As a result of this smoothing (the degree of which depends on a number of factors, including the density of sampling), the map of estimated values will not accurately represent the magnitude or sharpness of peaks and troughs in the real map.

Conditional simulation is a method for generating spatial simulations that reproduce the spatial variability in the true map (as estimated from the observations), and that are also constrained to pass through the data (i.e., the values simulated at sample locations are equal to the observed sample values). Conditional simulation involves Monte Carlo simulation from the joint conditional distribution of values at unsampled locations (e.g., along a fine grid of locations) given the values at sample locations. The basic assumption is that the underlying random field, possibly after transformation, is Gaussian and that the mean and spatial covariance given to the corresponding estimates are based on the original data. **Figure 13** is a map of elevations kriged from 500 sample elevation data (kriging is a weighted sum interpolation technique that exploits the spatial covariance function $C(h)$ or the semivariogram $\gamma(h)$, and which also produces a measure of the variance for each estimate). **Figures 14** and **15** are two realizations of the true elevations produced by conditional simulation. Notice the increased 'roughness' of the simulations compared to the kriged map, and that, although they exhibit the same degree of roughness, the simulated maps are different. Each realization represents an estimate of the true map that can be used, for example, to determine the effects of spatial processes (e.g., flooding or pollution deposition) that would operate over an area. Conditional simulation has also been used to determine the value of obtaining additional samples.

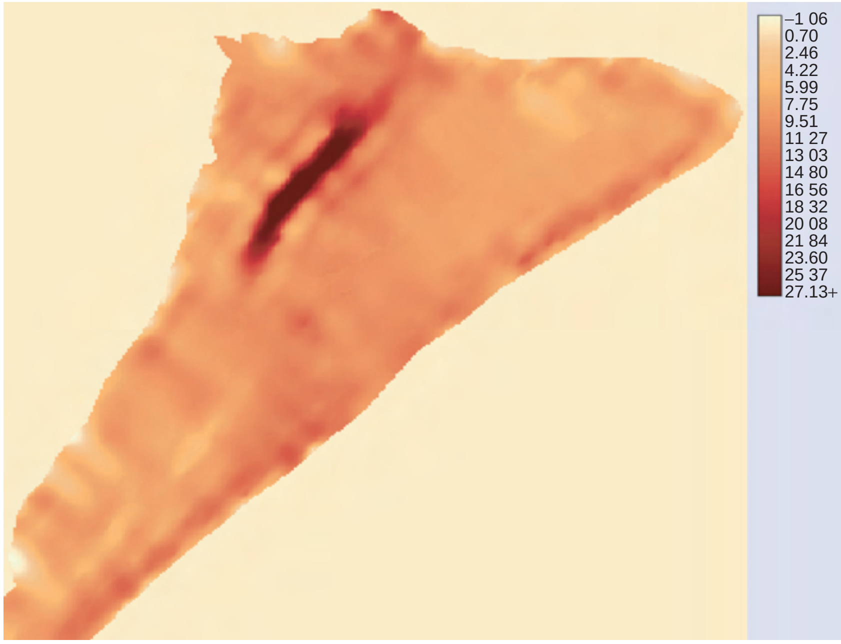

Figure 13 Map of elevations obtained by kriging 500 sample points.

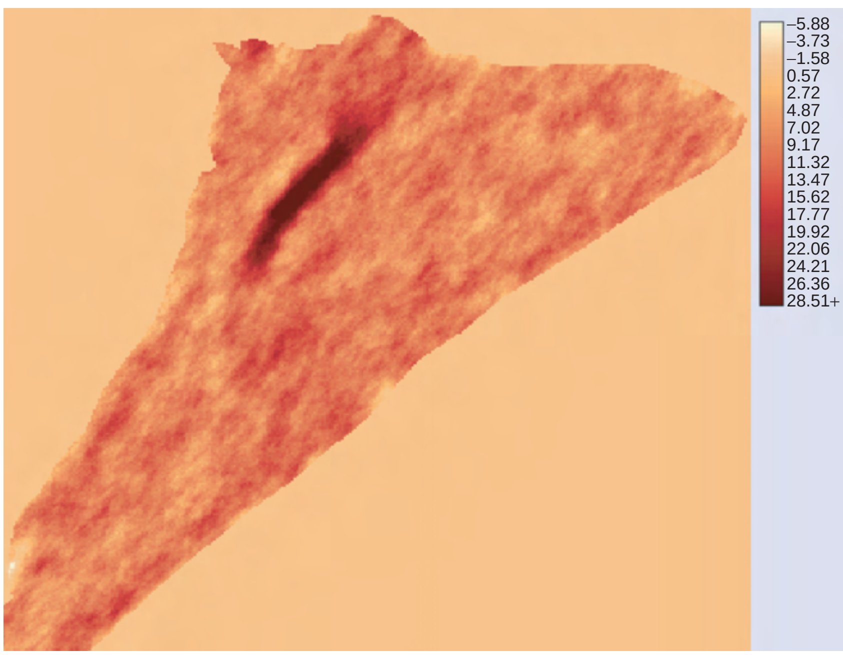

Figure 14 One realization of elevations using conditional simulation.

Conclusion

The Monte Carlo method permits the analysis of a wide variety of mathematical problems relatively simply. Its main advantage is that other methods are often more costly, impracticable, or impossible to implement. With the advent of more powerful computing capabilities, the Monte Carlo method provides a simple and effective approach to experimentally determine the outcomes of complex models and approximate solutions to intricate mathematical problems. Monte Carlo methods thus play a large role in the many subfields of geography that routinely employ mathematical models. They are especially useful in analyzing large spatial models and geographical data exhibiting high levels of complexity.

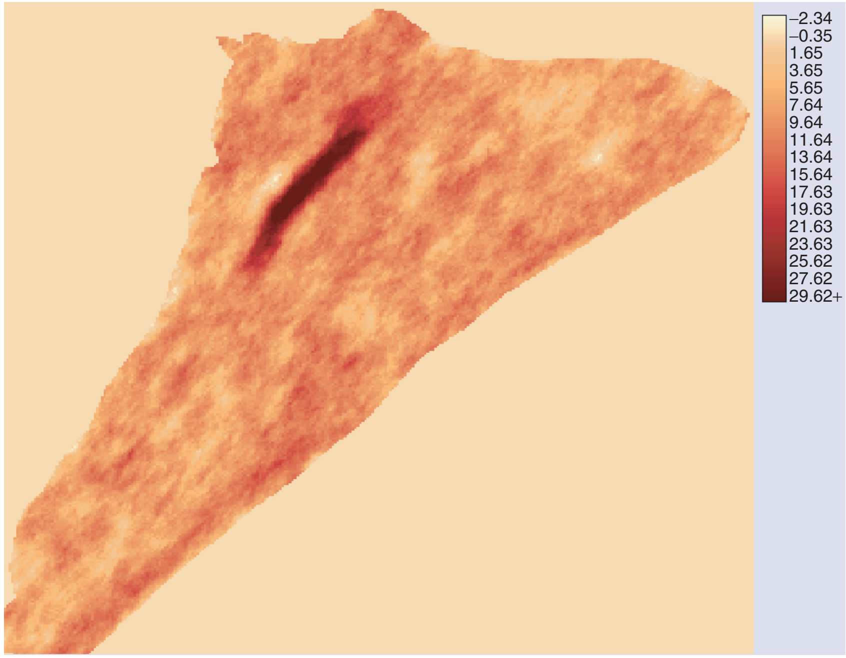

Figure 15 Another realization of elevations using conditional simulation.

The main disadvantage of the Monte Carlo method is that it provides approximate solutions that are only as good as the input assumptions and sampling technique used. Errors in the sampling method, or a sample size that is too small, may lead to unwarranted conclusions. Furthermore, as the complexity of the problem increases and a higher accuracy of the output is desired, the size of the required sample increases exponentially. In such cases, running a Monte Carlo simulation on a desktop computer may be resource intensive and time consuming.

See also: Kriging and Variogram Models; Location Analysis; Quantitative Methodologies; Sampling; Spatial Clustering, Detection and Analysis of; Spatial Interaction Models; Spatial Interpolation; Statistics, Inferential.

Further Reading

Aoyama, Y., Ratick, S. J. and Schwarz, G. (2005). Modeling the impact of business to business electronic commerce on the organization of the logistics industry. *Geographical Analysis* 37(1), 46 68.
Burmaster, D. E. and Anderson, P. D. (1994). Principles of good practice for the use of Monte Carlo techniques in human health and ecological risk assessments. *Risk Analysis* 14(4), 477 482.
Eckhardt, R. (1987). *Special Issue: Stan Ulam, John von Neumann, and the Monte Carlo Method. Los Alamos Science* 15, 131 137.
Efron, B. and Tibshirani, R. J. (1993). An Introduction to the Bootstrap. New York: Chapman & Hall.
Fishman, G. (1996). Monte Carlo: Concepts, Algorithms and Applications. New York: Springer Verlag.
Hattis, D. H. and Burmaster, D. E. (1994). Assessment of variability and uncertainty distributions for practical risk analysis. *Risk Analysis* 14(5), 713 730.
Kalos, M. H. and Whitlock, P. A. (1986). *Monte Carlo Methods*, vol. 1. New York: Wiley.
Manson, S., Ratick, S. J. and Solow, A. R. (2002). Decision making and uncertainty: Bayesian analysis of potential flood heights. *Geographical Analysis* 34(2), 112 129.
Oak Ridge National Labs (1995). Introduction to Monte Carlo methods, Computational Science Education Project. http://www.ipp.mpg.de/de/for/bereiche/stellarator/Comp sci/CompScience/csep/csep1.phy.ornl.gov/mc/mc.html (accessed Jul. 2008).
Office of Emergency and Remedial Response, US Environmental Protection Agency (2001). Overview of probabilistic approach to risk assessment. In Risk Assessment Guidance for Superfund, Vol. III, part A: Process for Conducting Probabilistic Risk Assessment. pp 1.1 1.34. US Environmental Protection Agency.
Rubinstein, R. Y. (1981). *Wiley Series in Probability and Mathematical Statistics: Simulation and the Monte Carlo Method*. New York: Wiley.
Simon, J.L. (1997). *Resampling: The New Statistics*, 2nd edn.). http://www.resample.com/content/text/index.shtml (accessed Jul. 2008) (An online text that can be read on its own or can accompany a computer program originated by Simon, Resampling Stats).
Sobol, I.M., Messer, R. (tr.), Stone, J. (tr.) and Fortini, P. (tr.). (1974). *The Monte Carlo method* (2nd Russian edn.). Chicago, IL: The University of Chicago Press.
Wyss, G. D. and Jorgensen, K. H. (1998). *A User's Guide to LHS: Sandia's Latin Hypercube Sampling Software*. Albuquerque, NM: Sandia National Laboratories.

Relevant Websites

http://random.mat.sbg.ac.at
WWW virtual library: Random numbers and Monte Carlo methods, Random number Generator.

Moral Economies

L. McDowell, University of Oxford, Oxford, UK

Glossary

Absolute and Relative Poverty Alternative ways of conceptualizing poverty. The former depends on an absolute definition such as the income needed to purchase a basket of goods or a particular income level below which people or households are poor. Common standards are, for example, falling below the average income or state benefit levels, plus 20%. Relative poverty as the term indicates is a more flexible concept. Poverty is defined in relation to what is an acceptable standard within a particular geographical area (be it nation, state, or city) at a particular moment in time. As societies become more affluent poverty lines also rise.

Caring Labor/Ethics of Care This term captures the distinction between work undertaken for financial remuneration and work that is about caring for others, without expectations of reciprocity or payment. It is typically employed in feminist analyses of domestic labor.

Ethics A system of moral values.

Gendered Moral Rationalities The notion that men and women operate on the basis of different moral systems; expanded in the work of geographer Simon Duncan to refer to differences 'between' women, especially mothers, in the ways in which they assess the decision to combine waged employment and unpaid caring labors.

Normative Standards An established standard that typically is universal, or is not dependent upon the context.

Introduction: Definitional Contradictions

At first sight the juxtaposition of the terms moral and economy(ies) seems paradoxical, even farcical – surely morality is about relations of obligations, about mutual respect and support, about knowing right from wrong, about caring for others whether members of an immediate family or community or the distant strangers whose needs are great. Economics, on the other hand, as everyone knows, is the dismal science, based on assumptions about rational economic man, who knows the price of everything (but the value of nothing) and who, in capitalist societies, makes cold and calculating decisions based on the maximization of profits, extracting as much financial advantage as possible from transactions that occur between unknown, faceless others. Even when defined in a less prejudicial way – making economies, being careful – there is a touch of the penny pinching, the mean and the dour, although perhaps the term moral itself embodies implications of a somewhat careful, rather joyless approach to life rather than one of casual abandon and hedonism. How appropriate then that both the originator of the science of political economy – Adam Smith – and the current prime minister in the UK – Gordon Brown – are both Scots, unable to escape the associations with a Calvinist culture of moral certitude and careful prudence in economic matters.

However, a concern with moral worth is evident more widely in Britain (and other capitalist societies) in the early twenty first century as policy debates about the appropriateness of a strong Protestant ethic centered in participation in waged work influence economic and social policy decisions. Similarly, the need to respect others is a central idea in the debate about (the problems with) British youth and 'esteem' – a very Victorian notion – is a key measure of the worth of a British academic in the latest round of a rational calculation of what each of us is worth. It is clear that strong notions of what counts as valuable, what sorts of behaviors are acceptable, how judgments are made about worth structure most arenas of life. In economic terms these judgments about value are perhaps particularly clear as, in the formal economy at least, monetary values are placed on economic activities.

Defining Morality

Defining morality, however, even in association with the second term economies, is a task that is not easy. The older certainty of moral philosophers has been challenged by postmodern theorists who have pointed out the relative nature of key terms. Normative notions such as morality are neither uncontested nor universal; instead what is moral is context and situated, dependent on the circumstances and the actors involved. Nevertheless the centrality of the notion of what is good and just lies behind the evaluation of whether an action is moral or not. Western economic notions about the social basis of lending and borrowing, based on the regulation of interest rates – a way of repaying risk – have no purchase in economies based on Islamic beliefs, for example, and while the analysis of this article shall be restricted to capitalist Western economies, it is clear that even here there is debate about what is included within the term the economy and so as the subject material of the discipline. As many have argued, feminist scholars have challenged the very definition of what is an economy, showing how the exclusion of unwaged work from its purview resulted in a neglect of the essential contributions of domestic and voluntary labor from the total social reproduction of economic relations in the widest sense.

Returning to definitions of morality, a distinction discussed (but actually rejected) by the geographer Andrew Sayer in his book about the moral significance of class is helpful. There he argues that in common sense terms morality is simply what kinds of behavior are regarded as good, how we should treat others, and be treated by them. In philosophy, there is a large and contested debate about normative conceptions of what is good but as Sayer suggests, in everyday life these norms are generally less coherent and explicit than philosophers typically assume. Sayer also notes that morality is sometimes distinguished from ethics, and that this distinction implies a difference between more formal norms or rules and informal everyday ideas of the good life based on experiences. Rather confusingly, though, philosophers tend to use both terms either interchangeably or even reversing the sense.

The original distinction between formal morality and everyday ethics comes from the work of Hegel who saw a difference between an abstract universal morality (*moralität*) and a more situated notion of good (*sittlichkeit*) related to individuals' position in the communities in which they live. In his work on class relations, Sayer uses the terms ethics and morality interchangeably but also notes that liberal philosophers tend to associate ethics with good and morality with right. It is preferable in some contexts to keep this Hegelian distinction; also, the term ethics is less judgmental than the term 'moral' or morality with its universalist or singular associations with a particular version of what is right. The term ethics perhaps better captures the complex, situated, and interdependent decisions that people currently make about the best way to live and to relate to others. Thus in feminist inspired work about relations of caring and how these activities are both essential to but undervalued by economic activities, the term 'ethics of care' seems preferable than the notion of 'gendered moral rationalities' that some geographers have adopted. However, Sayer and other commentators agree, that in most cases the terms are used interchangeably.

Defining the Economy

Defining the term economy raises equally difficult issues. It has been common practice in mainstream Anglo American economics to equate economic relations with relations of production, exchange, and distribution within capitalist societies in which a cash nexus is dominant. Thus, activities studied by neoclassical economists and economic geographers typically have included all those forms of work that have a monetary value. However, as students of peasant economies or of domestic labor undertaken in the home have long argued, all sorts of production and exchanges take place outside the cash economy, or what might more generally be defined as formal market relations. Exchanges in kind, informal employment, cooking and cleaning 'for love' within households are also part of an economy and in their own way are as structured by moral beliefs, rules, and norms as are formal market relations, regulated by institutional structures and legislative controls. Indeed, the moral basis of these forms of work and exchange is perhaps rather clearer than in the market, as they depend on affection and mutual care rather than on profit (although also on exploitation and forms of interpersonal control and domination).

All forms of economic behavior, however, have ethical implications as they are embedded within social interactions and so beliefs about the basis on which an interaction takes place, its fairness (or not), the relative position of the individuals involved and what is due to them, and a judgment of the potential outcomes. As an object of study then 'moral economy' (or economies) might 'either' be the study of the set of ethical beliefs and normative assumptions that structure any kind of economic action 'or' a subset of economics in which a particular (universal?) view of morality is the underlying principle of exchange. As an example, we might compare free market or neoliberal capitalism as practiced at present in the US and UK which is based on a set of beliefs that values free markets, competition, and free trade more highly than the achievement of a less inequitable distribution of incomes with, say, green economics which is based on a particular valuing of the long term sustainable use of resources and their preservation for future generations over immediate profits. While the latter arguments are clearly based on an identifiable version of a moral judgment, the former, despite its attempt to present itself as 'value free' is equally evidently located within a set of moral beliefs about how the world ought to work.

What is perhaps surprising is how long the notion of a value free, positivist version of economics was dominant in twentieth century Western social science, despite the self evident involvement of economics in the sense of the production, distribution, and consumption of goods and income necessary to live at a particular standard in questions of who should get what – an evidently moral question that lies at the center of political or moral philosophy. Nevertheless, one of the marked features of the sort of positivist social science that dominated mainstream geography and economics throughout all but the last quarter of the twentieth century was the apparent absence of ideas about value or morality. Positivist social science was represented as objective and factual, based on the measurement and prediction of aggregate social behaviors. It made no claims to be 'moral' but simply to measure 'what is'. But as the critiques that developed in geography from the late 1960s onward from, among others, Marxist and feminist geographers and economists, and later postcolonial theorists, have documented in detail, this apparent objectivity disguises relations of exploitation and oppression as mere facts. Ideas about market rationality, for example, or meritocratic societies where achievement is seen as a measure of worth rather than privilege, are just as value dependent as the notions about class privilege and structured patterns of masculine domination that inform Marxist and feminist social science, respectively. Normative discourses – as Marxist and feminist scholarship are imply a difference between what is and what ought to be: an argument that positivist social scientists tended to dismiss as 'political', rather than using the less judgmental term 'normative'.

The Origins of the Term 'Moral Economy'

Although the focus of this article is capitalist economies, and on the interplay between cultural beliefs, political judgments, and economic activities, the term moral economy initially was developed in the context of traditional, peasant societies, in which customs, culture, and traditions influenced the social relations of production, exchange, and waged labor. Indeed, it has been argued that economic relations based on the principles of social justice, fairness, even goodness, are only possible in such small, closely knit societies based on face to face interactions. However, work by E. P. Thompson on the development of British capitalism during the eighteenth century and by other historians and sociologists has documented the ways in which a concept of a 'moral economy' might be utilized to resist authority and to justify struggles against capitalist exploitation. In the context of contemporary capitalist economies, it is often suggested that a moral economy is one in which there is some form of balance between ethical norms and economic principles in order to achieve a degree of social justice. It is in this sense that the term is used here, documenting the ways in which moral questions about what 'ought to be' have been addressed in post war Britain. Also, a closer look has been taken at how moral arguments that lie behind demands for greater income equality and universal provision for risks have influenced economic policy. All states have to answer questions about the degree of income inequality and differences in overall standards of living that are acceptable and justifiable at different times. Who deserves what forms of support and on what moral basis?

In the development of Western industrial economies, for example, until perhaps the late 1950s, it was assumed that every able bodied man had the responsibility to engage in wage labor in order to provide for his own needs and those of his family, whereas women and children had a claim on his income through the rights and obligations of marriage and kinship. In the second half of the twentieth century and into the new millennium these rights and responsibilities have been transformed, altering notions about gender and household obligation, as well as maternal care. New notions about the extent to which caring responsibilities are economic issues are now important, as the concept of a work/life balance enters economic discourse, in part replacing or at least complementing discussion about who should enter the labor force and for what financial rewards.

Moral Philosophy and Political and Moral Economy

At the heart of economics lie questions not only about the production of goods and services but interrelated issues about who is entitled to or deserves what standard of living – that is, questions about distribution and consumption. Clearly in conditions of great scarcity and want, ethical notions about fairness may be overridden by pressures simply to survive, but in modern Western societies, more complex issues about what is a good society, what is fair and just lie at the heart of the modern capitalist economy. Indeed, these moral principles may be contested and overridden by other considerations – arguments about maximizing profit, for example, or about the need to provide incentives for the very rich to work harder (or at all), or for the unemployed to enter the labor market. What is clear, however, is that in complex societies, systems of market controls and regulations have been established in order to ensure that the moral principles that lie behind the distribution of incomes between the working and nonworking population are enacted and enforced. These sets of social regulations include the institutions and regulations both of the tax system and welfare provision (identified in the French regulations school as a key part of the control of capitalism and necessary for the maintenance of profit), as well as the banking system, private property rights, labor law and workers' organizations, and all those institutional regulations that define 'the rules of the game'. All these rules are necessarily based on moral economic norms about what is right, or perhaps rather what is acceptable. What is always contested, as the economy is based on social relations of unequal power, is the extent to which forms of intervention should 'interfere' (as free marketeers claim) with the untrammeled rules of the market, in order to alter the control of and distribution of profits, for 'the greater good' or to raise the standard of living of the least fortunate. It is in this sense that one has to examine the different moral principles that lie behind arguments about and the methods used to regulate the distribution of incomes from participation in waged work at a number of key moments in the development of a British welfare state. These discussions tended to be corralled to a specialist sub branch of economics – welfare economics – or to be seen as the domain for social policy rather than economic issues *per se* but this division is less and less tenable.

One of the founding fathers of the discipline of economics is Adam Smith, whose *Wealth of Nations* established the basic principles of the new discipline of political economy. As Smith argued and other later nineteenth century political economists extended, one of the purposes of economics as a theoretical discipline is to establish 'the rules of decency' – how the economic world should work, how owners and workers should be rewarded, and how its outputs should be distributed. The establishment of rules about what is fair and just, or acceptable is a historical and contested process, the outcome of long struggles about, for example, the length of the working day, about workplace health and safety regulations, and about fair and minimum wages. Although many of these issues are still the focus of struggle – legislation introducing a basic minimum wage was not passed until 2001 in the UK, for example, – political economy and economics as disciplinary frameworks influencing policy played an important part in establishing the normalcy of economic principles – about the nature and legitimacy of profit, for example. Now a set of neoclassical principles are taken for granted: what were once matters of contestation have become normalized, as if they were basic timeless facts of the world. Thus, as Sayer noted, questions of validity have become turned into questions of behavior, although there still remains key areas of disagreement about, for example, the right to sell body parts or the ownership of genetic information and material.

In the next two sections of the article, focus is laid on questions about the distribution of income, looking back half a century or so to compare the arguments made then by the new Labour Government elected after the end of World War II with those made by the New Labour Government, initially elected in 1997, about economic inequality and the normative principles that lie behind an acceptable degree of inequality between individuals. Here the work of the political sociologist W. G. Runciman is particularly drawn upon.

Moral Principles Post-World War II

Although many economic historians and social policy analysts regard the post war settlement introduced by Attlee's Government as establishing a new moral basis for the British economy and state, according to Runciman, the reforms rather completed the processes of the moral transformation that had begun under Lloyd George between 1915 and 1922. Whether new or a continuation of older normative beliefs and standards, however, four principles underlay the settlement which all challenged previous Conservative ideas about providing for risk and about acceptable degrees of income inequality and security. The principles transformed the moral basis of providing for the poor and less fortunate members of British society. First, the old insurance principle of contributions made by individuals against risk was to be replaced by universal, across the board provision financed by central government. Welfare provision was to be on the basis of need, regardless of ability to pay and free at the point of access. Second, progressive taxation was to be introduced, although the Labour Party stopped short of the abolition of private property. The third principle was to protect the legitimate interests of organized labor, especially the least well paid and most exploited workers, shifting the balance from the long domination of capital, and aiming to produce relatively narrow pay differentials after tax between different categories of workers and management and bosses. The final principle was one connected to wider macro level economic management and more centralized planning.

While these moral principles had huge general appeal after the deprivations of World War II, as the decades passed and the working population became more affluent, the general consensus began to evaporate, allowing not only a long period of Conservative domination during the middle decades of the twentieth century but ultimately the election of the radical right government headed by Margaret Thatcher that established a new set of moral foundations for economic and social policy. Runciman suggests that the seeds of the decline of the post war consensus lay in the changing economic circumstances of the second half of the twentieth century, illustrating the interconnections of political philosophies and economic conditions. First, all forms of social services – schools, housing, and especially the National Health Service (NHS) – were unable to meet the seemingly unlimited rising demands, the huge costs, and growing bureaucracies needed to provide and allocate the services. Second, progressive taxation of incomes became less popular as with growing affluence more manual workers found themselves paying tax and national insurance contributions. Third, trade unions lost a great deal of their moral legitimacy. Strike actions felt by the public became unpopular and more generally less affluent and less organized workers, especially women workers, felt let down by the masculinized union movement and its focus on the skilled male workers. Thus the widespread acceptance of the post war moral compact began to vanish.

In 1979 Margaret Thatcher became the prime minister with a completely different moral agenda for the economy – she had a vehement belief in the superiority of free, competitive markets, in individualized market provision of services, in 'rolling back' the state, and an implacable hostility to the Trade Union Movement. Declaring there is no such thing as society, she summarized her economic aims as turning Britain back from a 'give it to me' to a 'do it yourself' society. These radically different normative beliefs have prevailed under New Labour since 1979. Despite the rediscovery of poverty in the 1960s by social policy commentators on the left, the moral beliefs of 'old' Labour lost their legitimacy, in large part, according to Runciman, because of the relative affluence of the British working class as a whole and the new forms of consumption and consumerism made possible by rising incomes, notwithstanding the widening gap between the least and most affluent individuals and households. Redistributive arguments no longer had moral force.

The latter part of the twentieth century was marked in affluent nation states by the growing significance of consumption both as a motor of economic growth and as a marker of social position and self identity. As many commentators have noted, the UK and the US have become consumerist societies marked by desire for instant gratification. A range of goods is available – and lifestyles constructed through their purchase – that was unimaginable in the earlier post war decades when the political left was motivated by a more puritanical view of the world. The post war welfare state was based on a moralizing notion of the basic necessities of life. And although the plain living, chapel going, good doing, high thinking wing of the Labour Party was never completely dominant, it lost its legitimacy as the century progressed. As the working class began to travel, expand the range of goods in their homes, spend more money on leisure and the purchase of private pleasure, a new rhetoric based on both class condescension and a belief that the absolute difference between the highest and the lowest incomes was less important once absolute poverty had (virtually) disappeared became dominant. Indeed a moral discourse about 'spendthrifts' who fail to save for their old age, about 'benefit scroungers' who live off state benefits rather than entering the labor market, about 'yobs' of all ages who drunkenly threaten the decent middle class in the public spaces of British cities has become the hegemonic representation of the unacceptable fraction of the working class, who are too loud, too brash, and so undeserving. To this radical change must also be added a more complex, multicultural society in which both women and men have more opportunities, in which religion has become an important social divide, and in which at least until recently opportunities for social mobility seemed the main mechanism for self improvement rather than the state provision of a basic standard of living. Thus the moral certainties of post war economic policy have vanished in the face of a neoliberal rhetoric about individual rights to pursue profit and income rather than to support aversion of social moral obligations. In the early years of the new millennium, however, the British economist Richard Layard has suggested that the pursuit of happiness rather than profit should be the overriding moral principle behind economic policy in the UK. Greater wealth, he has argued, is not correlated with satisfaction and happiness.

Ethical Consumption

A strand of that ethical Puritanism that marked left wing movements in the post war period, however, has reemerged today, fused with a growing concern about distributional issues at a global scale rather than within nation states. Thus the key moral issues of the early twenty first century are as much about the global distribution of income and resources, the location of production, the exploitation of workers, and the effects of rapid industrialization on the climate, for example, as they are about national policies and in the main are based on a moral belief in the principles of environmental sustainability. Ideas about ethical banking and socially responsible investment (SRI) in which fund managers make decisions about stock market portfolios based on social and/or environmental criteria are spreading. Another of the recent movements in the West based on normative beliefs about entitlement and distributional questions is the ethical consumer movement which takes a range of forms from boycotts of certain products and/or retailers to moves to buy locally and eat only seasonal products. Thus there is a growing number of non meat eaters in Western societies, as well as boycotts of retail chains such as Wal Mart and the fast food outlets such as the MacDonald Corporation.

These movements raise interesting ethical issues, as indeed does the wider sustainability movement. Who sets the standards? Are they equally appropriate for rich and poor consumers, for rich and poor nations? Recent work on Wal Mart is an interesting example. This is a US based and originally family owned chain of stores (that now owns ASDA in the UK) that owns thousands of stores in low income communities selling goods at rock bottom prices. This is achieved both through extremely aggressive purchasing policies and through the payment of low wages to its employees. The moral costs of these policies seem clear: suppliers, especially in the developing world, receive inadequate recompense, US and UK workers have to claim benefits to survive on poverty wages and so the state is subsidizing the chain and competition from small retailers is vanquished, wiping out the traditional main street in small town America.

Wal Mart is also anti union. However, the low prices at Wal Mart clearly are important for the poor in the US. It has been estimated, for example, that food and dry goods at Wal Mart are on average 27% cheaper than at tradi tional grocery stores. The poor benefit directly and all US consumers indirectly from the anti inflationary ef fects of low prices. It is not so simple, however, when other factors enter the moral equation and when the timescale is longer as the farming and working conditions in the developing world that enable Wal Mart to operate its low cost policy are morally and environmentally un sustainable. Action is needed at a larger scale than in dividual boycotts, using international bodies to insist on adequate labor conditions in Third World factories, for example. Normative standards, levels of minimum pay, and codes of conduct are no longer (if they ever were) moral issues that can be addressed within the bounds of the nation state. The new international division of labor in a globalizing world raises increasingly complex ques tions for moral principles on which to base economic decisions.

See also: Consumption; Gender, Historical Geographies of; McDowell, L.; Neoliberalism.

Further Reading

de Goede, M. (2005). *Virtue, Fortune and Faith: A Genealogy of Finance*. London: University of Minnesota Press.
Fishman, C. (2006). *The Wal Mart Effect*. London: Allen Lane.
McDowell, L. (2004). Work, workfare, work/life balance and an ethics of care. *Progress in Human Geography* 28(2), 145 163.
Runciman, W. G. (2006). What happened to the Labour Party? *London Review of Books* 28(12), 17 21.
Sayer, A. (2005). *The Moral Significance of Class*. Cambridge: Cambridge University Press.
Thompson, E. P. (1993). *Customs in Common: Studies in Traditional Popular Culture*. London: New Press.

Relevant Websites

http://www.ifat.org
International Federation for Alternative Trade.
http://www.fairtrade.org.uk
The Fairtrade Foundation.

Moral Landscapes

G. Setten, Norwegian University of Science and Technology, Trondheim, Norway
K. M. Brown, Macaulay Land Use Research Institute, Craigiebuckler, Aberdeen, UK

Glossary

Landscape Landscape is both a material fact and a representational social construct. Landscape is where substantive material forms and the unfolding social practices – including the particular ideological representation of landscape as a perspectival 'view' – are mutually and recursively constituted. Consequently, landscape – as form, representation, and meaning – is a site of contention and struggle. Such struggles not only shape landscapes but also involve attempts to naturalize them, making them seem inevitable, ordinary, and even necessary. This allows landscape to function as a medium through which identities are negotiated and reproduced, as well as a space in which the legitimacy and moral authority of particular people and ways of life are claimed and contested.

Moral Geographies The term moral geographies has come to denote the often, but not always, empirical study of how aspects of morality relate to spatial and temporal patterns and social relations. Related work has examined how moral judgments or assumptions are made about particular people, things, and modes of bodily conduct as appropriate or inappropriate, or in or out of place, in relation to particular spaces, places, and landscapes, and underlines the spatial peculiarity and geographical constitution of morally underpinned values and behaviors. A key theme of moral geographies is how we deal with the variety of potentially conflicting notions of how we ought to be in, and engage with, the world.

Morality Morality is commonly conceptualized as the sphere of significant normative concerns that focuses on how the world 'ought' to be. Morality, thus, refers to the beliefs, actions, and behaviors that reflect and underpin people's conceptions of what is just and unjust, appropriate and inappropriate, right and good. Moral practices and beliefs, are commonly (albeit not universally) distinguished from ethics conceived as moral or normative theory. Although moral issues pervade all realms of human life and are deeply enmeshed with everyday social interactions, discourses, and practices, the extent to which we routinely engage in moral evaluation is rarely acknowledged. Indeed, many of these can be all the more powerful for any apparent lack of moral content; a veneer of objectivity and self-evidence tend to make the underlying moral judgments invisible.

The concept of moral landscapes addresses the inter relationship between landscapes and moral values and judgments; it concerns how particular symbolic and material landscapes both shape and reflect notions of 'right/wrong', 'good/bad', 'appropriate/inappropriate', and 'natural/unnatural' in relation to particular people, practices, and things. It also concerns the ways in which certain moral boundaries are naturalized in, and through, landscapes, in the interplay of their material and repre sentational forms and related significations.

Moral Geographies and Landscapes

Geographers' contribution to the explicit elucidation of moral issues peaked around the turn of the century. Con necting the moral and the spatial was at the heart of works being undertaken which can broadly be identified as 'moral geography'. A key advocate for this connection, David M. Smith, held that geographers had a key role to play in examining "the contextual thickening of moral concepts in the particular (local) circumstances of differentiated human being." Political philosopher, Michael Walzer's distinction between 'thick' and 'thin' morality thus offered geog raphers an opportunity to explore the co constitution of morality with particular spaces, places, and landscapes that philosophers have pointed to but little explored:

> This dualism is, I think, an internal feature of every mor ality. Philosophers most often describe it in terms of a (thin) set of universal principles adapted (thickly) to these or those historical circumstances. … Morality is [however] thick from the beginning, culturally integrated, fully resonant, and it reveals itself thinly only on special occasions, when moral language is turned to specific purposes (1994: 4).

Although the theme of moral geography is not new, the term moral geographies was first used in the late 1980s by Felix Driver to refer to the workings of moral assumptions regarding human conduct and its relation ship to the environment in nineteenth century social science. Since then, numerous works have made clear that the human–environment and society–nature rela tionships that lie at the heart of the discipline of geog raphy have fundamental moral underpinnings. For this reason, Philo, in *New Words, New Worlds: Reconceptualising Social and Cultural Geography*, urged geographers to:

> establish the geography of everyday moralities given by the different moral assumptions and supporting

> arguments that particular peoples in particular places make about 'good' and 'bad'/'right' and 'wrong'/'just' and 'unjust'/'worthy' and 'unworthy'. There can be little doubt that these assumptions and arguments do vary considerably from one nation to the next, from one community to the next, and one street to the next: and it is also evident that the lines of variation overlap in many ways with variations between people and places in terms of social class, ethnic status, religious belief and political affiliation (to name but a few possibilities) (1991: 16).

Indeed, it seems that moral issues are inescapably manifest in many key geographical concepts. For example, Philo notes that "moral assumptions and arguments often have built into their very heart thinking about space, place, environment, landscape," (Philo, 1991), and Matless points to the numerous moral assumptions that are often "taken for granted regarding space, time, place, nature, land, etc." The precise ways in which humans position themselves and others with respect to notions of land, environment, and nature are thus replete with normativity and moral assumptions. Coming from a different, but related angle, commentators such as Gregory have pointed to the morally differentiated character of geographical reality. As Smith indicates, moral values not only vary among individuals and groups, and hence from place to place, as part of our world of difference, but are not fixed or immutable, and so can also change through time.

Work on moral geographies in general, and landscape in particular draws attention to the way morality underpins the fundamental relationship people have with land; how they see it, how they engage with it, and how they allow others to engage with it, that is, how landscape is used to both prohibit and enable certain behavior. The morally loaded use of the term land use has also been highlighted, particularly in relation to planning which is founded on a normative morality concerning the 'right' or 'best' use of land, and also in relation to the practical use of land that necessarily reflects and produces moral judgments relating to the conduct of people in the environment. Notions concerning the 'use' and 'misuse' of land, and competing formulations of appropriate behavior in the landscape, reflect moral assumptions about environment and society, and are influential in shaping particular landscapes. When particular landscapes come to be associated with particular moral values, particular groups or activities are given legitimacy while others are undermined (i.e., labeled as immoral), which in turn affects the characteristics of those landscapes. Other authors have focused on moral landscapes as materializations of moral approval or condemnation of particular practices.

Not only are moral geographies expressed through different behaviors or conduct in the landscape, a moral vocabulary of landscape has emerged. Terms such as vulgar, disorder, and alien, to name just a few have become part of a geographic language that seeks to describe processes of transgression or the breaking of codes of conducts. Any work on moral geography and landscape thus concerns the delineation of limits against which events/behavior/conduct can be judged and which are always place and time specific. It is against this background that four discernable, yet interlinked, contemporary loci of geographical thought that deal with moral landscapes can be identified.

Moral Landscapes as a Way of Doing Human Geography

Conduct in Place

In this approach, landscape forms the stage upon which moral judgments of people and practices are made; the moral evaluation of acceptable modes of bodily conduct or of appropriate roles and identities always takes place explicitly or implicitly with respect to particular landscapes. Consequently, particular combinations of individuals, groups, and behaviors assumed to belong to, or be in place in, one landscape may be deemed quite out of place in another.

A key exponent of this approach is David Matless. He links the construction of the 'character' of particular places, or rather how their character ought to be, to the conduct of particular groups and contested notions of what combinations of roles, identities and behaviors, and even sounds, are in keeping with that character or not. He illustrates how particular landscapes are complicit in shaping national and class identities, as well as allied assumptions about acceptable modes of conduct, with implications for who can stake a cultural claim over particular spaces. There have also been attempts to link these ideas to animal geographies, such as those uncovering some of the moralities that make certain hunting practices acceptable in some landscapes and not others, and showing that landscapes are crucial in determining who can do what, in what style, and where. Matless' approach is also echoed in discussions of how music legitimates particular landscapes, for example, as they are underlain with moralities of nationhood and citizenship.

Such work links particularly closely with research on geographies of exclusion and transgression, which demonstrates how such boundaries of acceptability relating to what or who is 'in' or 'out of place' may only become visible when transgressed. Practices and identities transgressive of a moral order can occur both purposely and unintentionally, but have radical potential in that they can denaturalize and disrupt the existing order and its taken for granted assumptions underlying who and what is included and excluded in the use of everyday space.

Thus, the allied processes of subject formation are often highly contested as attempts to assign marginal identities to particular (often marginal) spaces, consequently communicating how places are made and resisted. A key device noted by several commentators is concealing and sanitizing issues of 'who belongs' or 'to whom do things belong' by focusing on issues of 'where things belong' or 'what behaviors belong'.

Moral Practice and Landscape

This second loci of moral landscapes and geography differs from the first in that it moves beyond a staged understanding of landscape. It is an approach based on the assumption that people's relationship to their physical surroundings is expressed through practice, and moral judgments about such practices in particular landscapes help shape landscapes themselves. Ideas of appropriate and inappropriate practice are thus moulded into the physical landscape, making landscapes and human practices a co constituted morally charged process. Emphasis is placed on how culture–nature relationships reflect and produce moral judgments, and landscapes being both a representation and a reflection of such relationships, can help us understand how the production and meaning of a lived and practiced landscape becomes a moral landscape.

Gunhild Setten has demonstrated for Jæren in Norway that what is considered appropriate or inappropriate, good or bad, practice in relation to a particular landscape differs between different groups, for example, between farmers and environmental bureaucrats. Moral values and arguments are relative not only to specific histories and geographies, but also to how one comes to share moral assumptions with others. The operation of implicitly or explicitly shared moral assumptions among a group of people concerning right and wrong in human action can be thought of as a moral order.

Ordering nature is a key component both in farming and within the public planning and regulatory bodies of the agricultural sector. Setten has thus argued that farming practices cannot be understood in isolation from the interweaving of such practices ordering nature, and from the concept of how the landscape becomes the subject of contested codes of conduct through these ordering practices. This suggests the importance of identifying how landscapes are the product of rules aimed at orienting and producing practices that ultimately are cast as natural or unnatural, moral and immoral.

Works on landscape, practice, and morality have demonstrated the need to raise questions related to whose morals, whose ideas of good or appropriate behavior, are allowed to dominate landscape discourse. These works have also demonstrated the role of landscape narratives in justifying different types of conduct.

Landscape as Polity

One of the most cherished landscape discourses within human geography evolves around landscape's scenic nature, that is, landscape as a visual idea. An often competing discourse defines landscape as a physical, territorial entity, as real and tangible land. Kenneth Olwig has shown that landscape is not simply territory or scenery, but also an expression of law, justice, and culture, that is, the amalgamation of community and place. Etymologically, the very term landscape was originally built around notions of morality, justice, and law, evidenced in Scandinavian and Germanic territorial history. With reference to medieval Scandinavia, Olwig demonstrates how traditions, customs, and institutions were woven into the character and conditions of a land, and that these 'lands', *Landschaften* (in German) or districts developed their own identity. Law, customs, and morality were central in the shaping of the land by its people; thus 'landscape' referred to the polity within which this took place. The early concept of *Landschaft* was consequently an expression of legal institutions and processes. Up until the nineteenth century, on the North Friesland coast of Schleswig Holstein, one can find examples of landscape territories that formed polities with their own representative assemblies and hence a large degree of self government with local rules derived from customary law. These polities were in German referred to as *Landschaften.*

At present, employing landscape as polity represents a potent force in taking seriously what Olwig termed "the substantive nature of landscape" (Olwig, 1996). According to Olwig, a substantive understanding of landscape recognizes the importance, both historically and contemporarily, of community, culture, law, morality, and custom in the shaping of human existence. A substantive landscape thus comes to be articulated through a polity's ideals and practices of law and justice. Consequently, a substantive landscape does not champion a scenic understanding over a physical, territorial understanding of landscape. It rather aims at bringing to the fore how landscape is always a site of contention in the intersection of physical and symbolic powers. There are strong links between what is here identified as the loci of 'moral practice and landscape' and 'landscape as social justice', respectively and a substantive understanding of landscape. Setten has used this understanding in developing ideas about landscapes of practice and landscapes as moral territories. Don Mitchell argues for the importance of substantiveness in relation to social justice and landscape.

Landscape and Social Justice

According to Don Mitchell, landscapes are made and remade through labor, exploitation, and struggle, and are naturalized by the work that ideological landscapes do in making invisible the associated material injustices.

Consequently, the way we conceive of landscapes and our relationship to them affects the ways in which we are able to make and defend claims to them. Although rarely making explicit reference to the term 'moral', the landscape theory of Don Mitchell is infused by dimensions of morality and normativity in three distinct, but intimately linked, ways. First, notions of morality are implicated in the mechanisms through which particular constructions of landscape assert and justify particular orderings of landscape and associated social relations. For example, Mitchell elaborates how landscape is always more complex than its morphology implies because it is produced and reproduced ideologically through normative judgments and assumptions; powerful actors make and represent the landscape as they think it ought to be – such as the means by which surplus value will be extracted – and then lend their 'vision' legitimacy by imbuing it with morally saturated definitions of what is right, proper, just, or natural in particular circumstances. He emphasizes the dual mechanism of how such moralities work to enable certain values – and therefore certain groups such as immigrant workers or homeless people – to be 'written out' or made invisible in the landscape, while simultaneously naturalizing their erasure.

Second, underpinning Mitchell's work is a profound concern for the outcomes of the processes of landscape production and reproduction, particularly with regard to geographies of justice and injustice, and whose interests are ultimately served by manipulating and controlling particular people, resources, and behaviors in such ways. The material manifestation of landscapes, and their role as a concretization of social relations, means that struggles over its various forms, meanings, and representations impinge on real people's bodies and lives and the very structures and conditions of existence. Landscape's ability to conceal the labor and social relations through which it is both made and made known, allows us to easily overlook how landscapes of beauty, bounty, and privilege are only made possible by the production of landscapes of destruction, exploitation, and alienation elsewhere. Questions of landscape and justice also thus involve questions of property, in terms of shaping who can – and cannot – make legitimate claims to occupy, appropriate, or alienate land.

Third, a strong normative element permeates Mitchell's assertion that we, as landscape researchers, must take a critical stance on what to investigate and how. Our role as scholars in representing and producing landscapes should not be overlooked. He condemns the attempt of some scholars to produce 'morally neutral' landscapes, and their associated failure to acknowledge the material implications of their approach because, as he asserts in *The Lie of the Land*,

> [t]he landscape will be defined by someone. The only questions are who claims a voice in doing so, and how that voice is amplified and transmitted in all manner of discourses, popular, academic, or on the ground (1996: 199).

The moral imperative is clear in his illustration of how we can actively contribute to the production of less degenerate landscapes, and steer landscape studies explicitly in the direction of social justice. It is also evident in showing how we are in a position both to give visibility (or not) to marginalized groups, and to draw attention to multiscale linkages and geographical interdependencies that often cause one landscape to thrive at the expense of another. Furthermore, Mitchell and others assert that such integration of landscape and social justice concerns must be made with greater explicit reference to normative theory anchored in moral and political philosophy. Yet others have extended the reassertion of materiality in Mitchell's work by drawing attention to the agency of the landscape itself.

From Moral Landscapes to Landscapes of Justice, Ethics, and Care

Works on moral landscapes demonstrate that in order to make the coupling of landscape and morality a relevant concern for human geography, it can be useful to address issues that go beyond the realm of the landscape itself. In fact, the attachment of moral reasoning to landscapes in debates within human geography has moved on to a more explicit and broader focus on ethics, care, and social justice. Approaches falling under the rubric of moral landscapes have consequently become but one way of doing human geographies that produce, and constitute the product of, moral concerns. In the process, the moral status of landscape seems to have been weakened, in that landscape becomes just a metaphor for other key concepts and concerns in geography, namely those of place, space, nature, and environment. Parallel to this development is that notions of morality have come to be central to a number of recent trends in human geography that are occupied with notions of justice, ethics, and care. Diverse works on space, place, nature, and environment, as seen from, for example, feminist , antiracist , and queer identified geographers, have made clear that it is no longer a question of whether geographers should concern themselves with moral and ethical issues, but in what ways must such issues be taken forward as part of human geography's agenda.

See also: Cultural Geography; Ethical Issues in Research; Landscape.

Further Reading

Brown, K. M. (2007). Understanding the materialities and moralities of property: Reworking collective claims to land. *Transactions of the Institute of British Geographers* 32, 507 522.

Cresswell, T. (1996). *In Place/out of Place: Geography, Ideology and Transgression*. Minneapolis: University of Minnesota Press.
Driver, F. (1988). Moral geographies: Social science and the urban environment in mid nineteenth century England. *Transactions of the Institute of British Geographers* 13, 275 287.
Jones, M. (2006). Landscape, law and justice concepts and issues. *Norsk Geografisk Tidsskrift Norwegian Journal of Geography* 60, 1 14.
Kobayashi, A. and Proctor, J. D. (2004). How far have we cared? Developments in the geography of values, justice, and ethics. In Gaile, G. & Willmott, C. (eds.) *Geography in America at the Dawn of the 21st Century*, pp 723 731. Oxford: Oxford University Press.
Matless, D. (1998). *Landscape and Englishness*. London: Reaktion Books.
Mitchell, D. (1996). *The Lie of the Land: Migrant Workers and the California Landscape*. Minneapolis: University of Minnesota Press.
Mitchell, D. (2003). Cultural landscapes: Just landscapes or landscapes of justice? *Progress in Human Geography* 27, 787 796.
Olwig, K. R. (1996). Recovering the substantive nature of landscape. *Annals of the Association of American Geographers* 86, 630 653.
Philo, C. (comp.) (1991). *New Words, New Worlds: Reconceptualising Social and Cultural Geography*. Lampeter: Department of Geography, St David's University College.
Sack, R. D. (1997). *Homo Geographicus: A Framework for Action, Awareness, and Moral Concern*. Baltimore: Johns Hopkins University Press.
Setten, G. (2004). The habitus, the rule and the moral landscape. *Cultural Geographies* 11, 389 415.
Smith, D. M. (1997). Geography and ethics: A moral turn? *Progress in Human Geography* 21, 583 590.
Smith, D. M. (2000). *Moral Geographies: Ethics in a World of Difference*. Edinburgh: Edinburgh University Press.
Tuan, Y. F. (1989). *Morality and Imagination: Paradoxes of Progress*. Madison, WI: The University of Wisconsin Press.
Walzer, M. (1994). *Thick and Thin. Moral Argument at Home and Abroad*. Notre Dame: University of Notre Dame Press.

Relevant Websites

http://www.sagepub.com
Cultural Geographies.
http://www.tandf.co.uk
Ethics, Place and Environment. A Journal of Philosophy and Geography.
http://www.personal.ecu.edu
The Ethics, Justice, and Human Rights Specialty Group of the Association of American Geographers.

Movies and Films, Analysis of

S. C. Aitken, San Diego State University, San Diego, CA, USA, and National Technical University of Norway, Trondheim, Norway

Glossary

Auteurism A theory that links films primarily to the creativity of their directors.

Deconstruction The unpacking/taking apart of movies to suggest the complex ways in which they relate to societal values and politics.

Discourse Analysis The study of institutional (e.g., Hollywood) and representational (e.g., movie images) narratives as they relate to the creation of power structures in society.

Experimental Methodologies Post-structural methods that reject grand theories and focus on nonrepresentational ways of knowing.

Filmic Pertaining to film, movies, and cinematic images.

Hermeneutics An interpretation of texts that attempt to maintain the integrity of the authors' intent.

Male Gaze The active visual attention of male characters (and audiences) on passive female representations.

Mirror Stage (Oedipal Stage) The stage at which individuals recognize themselves as monadic beings, separate from their environments (and mothers).

Nonrepresentational Theory Post-structural theories that challenge the power of representation (and cognition) and focus instead on nonrepresentational ways of knowing such as embodied emotions.

Political Unconscious Esthetic and psychic processes that underlie larger economic and social structures.

Introduction

The geographic study of film focuses on interpreting and explaining the ways people, places, spaces, and things are embodied and embedded in moving pictures and the ways cinematic visions influence and transform our geographic world. Geographers arrived relatively late to the study of films and movies: Study of the interrelations between filmic spaces and the politics of culture, and the use of film as a means for understanding our place in the world was largely unexplored in geography prior to the 1990s.

When systematic geographic study of film took off in the 1990s, it was colored by methods prevalent at that time. These included semiotics, hermeneutics, psychoanalysis, feminism (particularly Freudian analysis), and Marxism. With a grab bag of borrowed methods, geographers explored the ways places and spaces were represented in films, the ways characters interacted with each other and their cinematic environments, and the ways environments took on active roles rather than providing stages or muses for central characters. Early studies were largely empirical, using specific films to elaborate person–environment relations in landscapes from deserts to ice flows to cities. Movie narratives and the roles of characters were used to elaborate larger social and cultural issues of interest to geographers, such as gender and ethnic relations. Fiction films were studied for their cultural relevance and geographic realism. Documentary films were critiqued for their authenticity, their geopolitical bite, and their use of landscapes.

The last decade witnessed a diminution of the authority of formal and relatively structured methods in favor of post structural and experimental methods. A large motivation for this move was to engage culture worlds in all their complexity and to resist the temptation of reducing them to one or two explanatory axes (e.g., the power of the phallus and the inevitability of neoliberal agendas). A further motivation was to get beyond the power of images as representations and look more closely at the ways images affected people and cultures from nonrepresentational standpoints. The study of movies underwent further significant transformation and gained increasing acceptance as a part of geography when the introduction of technologies such as the digital versatile disc (DVD) format and the Internet facilitated relatively easy access to movies, scripts, storybooks and commentaries.

Gauged by number of geographic articles and books that incorporate analysis of films and movies, it seems reasonable to suggest that the study of film within the discipline of geography has now come of age. More important than the quantity of articles and special sessions at conferences is a quality of work that results in important debates on the character of filmic representation and meaning production. This resulted in the development of both spatial ontologies of film and filmic ontologies of space, with study going beyond the postmodern idyll that there is nothing but image/representation. This theoretical sophistication offsets the narrow methodological empiricism of earlier work, dispelling notions that geographers either naively embrace certain films as tools for representing geographic concepts (landscape, space, place, and so forth) or they unabashedly borrow from film criticism and cultural studies to help elaborate geographic questions with little methodological or theoretical concern.

Structural Methods

Structural methods are important for answering geographic questions that presuppose the world is ordered in a certain way. Many moviemakers indulge, for example, in systems of signs as a way of communicating with audiences. A full appreciation of Robert Rodriguez's movies such as *Sin City* (2005) or, to a lesser extent, his earlier *El Mariachi* trilogy, requires an understanding of the use of space in graphic comics, the timing of narrative in pulp fiction, and the ways 1930s film noire social realism translates into twenty first century adolescent fantasy. Alternatively, it may be suggested that Rodriguez's work defies sophisticated analysis and is structured simply as entertainment, but a suggestion of this kind requires an understanding of the ways contemporary audiences read and appreciate his work. What passes for entertainment is very different today from the 1930s and so, questions may be raised about the continued popularity of gothic and noire genres. In either case, structured ways of studying films and film contexts are elaborated well methodologically through textual methods, Marxism, feminism, and psychoanalysis.

Textual Methods: Semiotics and Hermeneutics

In the early 1970s, Paul Ricoeur suggested that social and political institutions inscribe values, meanings, and behaviors in much the same way as texts. This metaphor was quickly adopted by geographers interested in 'reading' landscapes. With the advent of film geography as a bonified part of the discipline, early researchers pointed out that this 'life as text' metaphor is applicable to social and cultural productions such as fiction and documentary movies. To aid in the reading of cinematic texts, geographers turned to linguistics and literary studies.

Semiology is intent upon uncovering the meaning of texts by dealing with what signs are and how they function. Swiss linguist Ferdinand de Saussure suggested that any sign could be divided into two components: the signifier and the signified. For example, the image of pages dropping off a calendar (the signifier) denotes the passage of time (the signified). The relationship between the signifier and the signified is arbitrary but may be fixed for a certain time by the nuances of political culture. The image of the moving calendar was widely accepted in the 1940s and 1950s, but is now seen as an over used relic, and today this semiotic technique is most likely to be used in satires or 'retro' movies. American philosopher, Charles Sanders Peirce, elaborated the triad icon–index–symbol to describe how signs are created, interact, and operate within films. For Peirce, any one film sign or frame may be constituted as an icon, an index or a symbol to varying degrees. The sign is an icon when there is resemblance (e.g., a portrait signifies a particular person); it is an index if there is a physical, causal, or symptomatic connection (e.g., smoke signifies fire); and it is a symbol if the connection must be learnt (e.g., the color green signifies go). If Peirce's work enables an expanded, variable application of semiotics, then Umberto Eco's development of the 'open' text in his *Opera aperta* ('Open Work') provides a methodology for understanding more complex fields of meaning, rather than simple strings of meaning. From this, movies are understood as open, internally dynamic, and culturally and psychologically engaged fields. Eco's novel *The Name of the Rose* (1980, later made into a movie starring Sean Connery), for example, is about a series of murders in a medieval monastery. It ends with semiotic beauty as the main protagonist attempts to reconstruct a library and create a space for cultural meaning based on scraps of evidence and the combination of random pieces of information.

Methodological concepts used by geographers interested in semiology include intertextuality, metaphor, and metonym. Intertextuality refers to the conscious and unconscious employment of previously created texts. Robert Rodriguez, for example, consciously uses the spectacularly violent comic book texts of Frank Miller in *Sin City*, whereas the text/story line of Sophocle's *Oedipus Rex*, written around 425 BC, may be so well rehearsed in Western culture that it goes virtually unnoticed when it shows up in a variety of movies (e.g., sons killing fathers in *Bladerunner* (1982), *Star Wars Episode III* (1983), and *My Own Private Idaho* (1991)). With metaphors, the relationships between two or more things are suggested through analogy (e.g., movies are texts), whereas with metonyms they are suggested through associations between names. In *Traffic* (2000), Director Steven Soderbergh uses orange/red and blue filters to signify respectively the danger and passion of drug trafficking across the Tijuana border crossing and its consumption by bored suburban middle class teenagers in Ohio. That danger and passion are signified by orange/red and suburban boredom and drudgery are signified by blue are understood only within a specific set of cultural codes.

As a textual method, hermeneutics is based on methodological principles from literary criticism that incorporate the context of the creator of the text as well as the reader. And so, movies are read in terms of the larger social and cultural contexts of their times as well as the proclivities of screen writers, directors, producers, and actors. For hermeneuticians, the integrity of the movie must be preserved so that meaning is derived from it and not projected onto it by the researcher. Auteur theory is a particular form of hermeneutics that prescribes the authorship (auteur) of films to the vision of directors who create seeming unique, distinctive, and recognizable movies. In theory, an auteur's films are identifiable regardless of genre. The term was first ap plied as a cinematic concept in François Truffaut's 1954 essay 'A certain tendency of the French cinema'. Some feminists criticize auteur theory with the suggestion that it elaborates patriarchal domination in moviemaking, because the majority of directors who are singled out for their vision are men.

Marxism

Whereas semiotics and hermeneutics are mostly used to analyze visual representations and spaces already pro duced, Marxist methodologies are concerned with the production as well as the consumption of images and narratives. Of particular interest are questions about who owns and controls the production of movies, and who stands to gain most from their distribution. How are the directors, writers, actors, artists, and other creators of movies, and how are cinemas and the consumers of movies affected by the patterns of ownership and control of the movie industry?

In classical Marxism, there are assumptions of underlying economic structures with their own inner logics, and an understanding of these logics enables re searchers to analyze everyday cultural elements. In the 1980s, concern was raised that the complexity of culture was lost to an endless recoding of property relations that are reduced to an explanation based upon class conflict. Concerned about classical Marxism's inability to inter pret small variations in cultural meaning, Fredric Jame son used literature and film studies to argue that there are important underlying esthetic and psychic processes – which he called the 'political unconscious' – that are co created along with larger economic forces. This consciousness is socially produced, and is filtered through the minds of men and women to the extent that dominant ideologies – and particularly capitalism – can be con tested. In *The Geopolitical Aesthetic* (1992), Jameson uses the works of French filmmaker Jean Luc Godard and Filipino filmmaker Kidlat Tahimik, among others, to argue that film analysis is one of the primary method ologies for exposing the importance of space in the construction of social and economic relations in late capitalism. Jameson's Marxist inspired approach to film reinvigorated studies of visual culture through the 1980s and 1990s, and paved the way for psychoanalytic and post structural frameworks.

Feminism

Feminist analysis of movies got its first major impetus from Laura Mulvey's (1975) essay 'Visual pleasure and narrative cinema'. She famously argued that voyeurism and what she called 'the male gaze' are foundational in Hollywood movies to the extent that the male character always actively looks and the female character is passive as the object of attention. The gaze, Mulvey argues, seeks to exercise power over its subject: as the 'maker of meaning', the male subject (either the leading man or the audience or both) gazes at the female image and marks 'her' as the 'bearer of guilt'. Fetishism, she goes on to note, involves the adoption of a female image as a spectacle that is so complete (e.g., Marylin Monroe) that it disavows Freud's theorizing of the male castration complex. Rather than wanting to punish the female other as the 'bearer of guilt', the fetishist raises it to a spec tacular object of desire. All of this positions the re searcher to understand filmmakers' preoccupation with various forms of male regression – physical, psycho logical, and historical – that connect nostalgia for the past and childhood with male fears for the body. Two ex amples point out the power of this argument: male fears related to the castration complex are represented by disembowelments and other forms of mutilation in slasher horror movies; and looking at the Medusa's head wriggling with phallic snakes turns men to stone – a point discussed by Freud in a 1955 essay – and is represented by Busby Berkeley's famous 1930s musicals featuring women dancers with exotic head dresses.

Post Mulveyian thinking focuses on the paradox that although masculinity derives considerable social and sexual power from representational castration and other male fears of the body, feminist film theory has problematically equated masculinity with activity, voyeurism, sadism, and story, and femininity with passivity, exhibitionism, maso chism, and spectacle. Carol Clover points to this paradox in an essay entitled 'Her body, himself: gender in the slasher film' by noting that many men accept female victim heroes – who she characterizes as 'final girl'. In examining this paradox, she argues for multiple spectator positions. Mulveyian theory constructs a monolithic male subject as universal and uncontestable and, at one level, this repre sentational convention provides a means through which the private structuring of desire can be represented in public form. But fears and desires embodied in images of disem bowelments are not just about castration fears, they also reflect and refract what Kaja Silverman notes as a series of male masochistic fantasies, and these fantasies are a pleas urable point of identification for viewers irrespective of gender. If Mulvey's preoccupation is with cinema's power to gender fix its spectators in a society that is already con structed under patriarchal dominance, then Clover suggests that perhaps the reverse is also possible: cinema offers its spectators shifting and multiple positions and resistance to larger hegemonic norms. Anne Frieberg, in *Window Shop ping,* moves this argument further by arguing that the dark spaces of cinema enabled women to go out in public to embrace or resist spectacle. This is about a space of dif ference that draws heavily on postcolonial theory because, for Clover and Friedberg, the spectator's identification is shifting, unshackled from an imperious Hollywood that contrives racial identities, ethnicities, genders, or sexual preferences. Clover's and Silverman's suggestions in com bination further frustrate the persistence of a comfortable methodological dichotomy aligning masculinity with domination and femininity with submission.

Psychoanalysis

Christian Metz was one of the first film theorists to apply psychoanalytic methodologies. He looked at the 'sus pension of disbelief' enabled by film, noting that what the film viewer witnesses is not real but an 'impression of reality' that is deeply wound into the collective psyche of viewers. For example, so called social pathologies such as agoraphobia, vertigo, hysteria, and schizophrenia are embedded in the landscapes of many movies to the extent that the space created marks collective responses to what is not normal/hegemonic as defined by society (and, problematically, the discipline of psychology). What the work of fantasies and phobias in film calls into question are the assumptions of (1) a discrete individual and an external threatening object and (2) some standard around which normality and humanness are judged.

Psychoanalysis and feminism are strongly linked in film analysis that counters essentialized standard of maleness. In her seminal work, Mulvey uses Jacque Lacan's linguistic reinterpretation of Freud to argue that our lack is made good by film, which restores us to an imaginary wholeness (at least for a short while) through a simulated and constructed reality. Lacan argues that this lack is precipitated by what he calls the 'mirror stage' (Freud's Oedipal stage) when language moves us from oneness with our environment (mother) to create a self/other dichotomy. Mulvey argues that in classic Holly wood films, women are never represented as 'self', but rather as 'other', the dark forbidden continent. An im portant and obvious direction here is suggested by speculating that men, also, are objects of desire and mystery. Within this context, the genre of the 'buddy movie' (e.g., *Gallipoli* 1981, *Tango and Cash*, 1989; the Leathal Weapon Series with Mel Gibson and Danny Glover) looms large as espousing an active homosexual eroticism. In this vein, it may be argued that male hys teria is a disassociation from hegemonic masculinity, and its allegories, the buddy movie/road movie, is a spatial performance of a redoubtably disenfranchised spirit.

There is some criticism of a cine psychoanalytic focus on interiority and pathology because it detracts from larger, exterior political and economic problems. Of late, psychoanalytic, Marxist and feminist perspectives have joined forces with post structural theory to provide a methodological critique of the pervasive and problematic assumptions that contemporary visual methodologies are powered solely by representations and how we react to them.

Beyond Representation

In the late 1990s, concern arose that there was too much emphasis on the power of representations and images, at the cost of understanding the complex material contexts of lived experience. Some postmodern writers, for ex ample, argued that public loss of faith in the so called authentic (the real) can be laid at the door of the visual media and, in particular, moviemakers: the copy was confused with reality. Alternatively, post structural wri ters emphasize that representations do not have a life of their own; instead, they are given meaning through the actions, thoughts, and material experiences of people.

Discourse analysis – based upon the writings of Michel Foucault – looks at the social consequences of represen tation by focusing on the ways power is constructed through representation rather than focusing on the rep resentations themselves while at the same time not losing site of the importance of the ways political identities are constructed. Foucault's work is often recognized with Jacques Derrida's methods of deconstruction as some of the first forays into post structural analysis. Discursive practices – the ways meanings are connected through representations and actions – are forms of disciplining and so discourses, at heart, are about power and knowledge production. Discourses produce subjects (directors, actors, and viewers), as well as images (special effects) and tech nologies (surround sound and DVDs).

Digitally Versatile

Geographical discourse analysis as applied to film is about the special language of movies and how that lan guage is practiced to elaborate spatial relations, emotions, and uneven development and the power of places. It is about how subjects and landscapes are produced through the media of film, and it is also about the production of technologies such as the Internet and DVDs.

Often, the Internet facilitates the illegal download of movies prior to their release dates. Web pages are sometimes created for particular movies that encourage exploration of wide ranging critiques and commentaries. DVDs, and particularly special collectors' editions, often include along with the main feature, special features such as director's commentaries or discussions between screen writers and actors as well as technical and special effects information. Of course, as informative as it is, this supplementary information is hugely contextualized and is often vetted by the film studio that produces and markets the DVD. Nonetheless, these added features offer researchers new insights into the larger discourses that play through a particular film.

Moving Images and Experimental Methods

In *Engaging Film*, geographers Tim Cresswell and Debbie Dixon argue that a post structural analysis of film is about engaging mobility, which can be thought of in a broad sense as a certain attitude, at times openly radical and at times quietly critical, toward fixed notions of people and places. That is, an emphasis on mobility suggests a certain skepticism in regard to stability, root edness, surety, and order.

Tom Conley's *Cartographic Cinema* (2007) engages movies in a provocative and quite straightforward way. By focusing specifically on the ways maps are represented in a series of movies, he goes on to suggest the need to go beyond representation by looking at what Giuliana Bruno (2001) calls tender mappings. These evoke carto graphies of affect that take viewers on journeys that go beyond cinematic technologies and dominant dis courses. For the geographer interested in analyzing film, these journeys are simultaneously internal and external, esthetic and political.

Geographers interested in analyzing film now use experimental methods to get at the nonrepresentational, affective, mobile components of film and cinema, with particular focus on the work of Giuliana Bruno and Gilles Deleuze. The corpus of Giuliana Bruno's work is about taking journeys and flights of fancy that elaborate the ways that certain films resist and rework patriarchal, urban structures and military fictions. To achieve these ends, her work is couched in material landscapes. While Deleuze is also characteristically post structural in his rejection of grand theories and structured methods, his response – unlike that of Derrida and Foucault – is not a painstaking, genealogical analysis and forensic criticism aimed at revealing life in all its complexity. He advocates experimental methodologies that do not try to 'represent' the world. For Deleuze, representation is a negative procedure because it distorts, judges, and ultimately destroys what it portends to reveal. Alternatively, his notion of affective disruptions and how those play out in filmic spaces highlight multiple subjectivities that resist, are one with, and are differentiated from seeming hege monic forms. As such, they add to the world in a positive way. In *Cinema 1*, Deleuze suggests that affective meth odologies present 'pre verbal intelligible content' that is not about story lines, linguistically based semiotics, a universal language of film, or some existential or Lacanian lack in the viewer. Experimental perspectives move re searchers beyond narrowly circumscribed structural ar guments about cinema genres, or mechanistic notions of how images work, to a more nuanced appraisal that draws heavily from circuitous, flexible, and transformative no tions of space and time. Proponents argue that it enables a liberatory connection with lived experience, both within the cinema theater and beyond it, to the images and material conditions that continuously produce and reproduce daily life. Experimental methodologies en courage tools such as emotions that enable anger against the structures that would impede freedom of expression, and joy for what comprises the best possible spatial and filmic worlds. This relates in important ways to a focus on cinema from a nonrepresentational perspective that evokes tender mappings.

See also: Discourse Analysis; Feminism/Feminist Geography; Film; Semiotics.

Further Reading

Aitken, S. C. and Dixon, D. (2006). Imagining geographies of film *Erdkunde*. *Band* 60, 326 336.
Aitken, S. C. and Zonn, L. E. (eds.) (1994). *Place, Power, Situation and Spectacle: A Geography of Film*. Lanham, MD: Rowman and Littlefield.
Bruno, G. (2001). *Atlas of Emotion: Journeys in Art, Architecture, and Film*. New York: Verso.
Clarke, D. B. (1997). *The Cinematic City*. London: Routledge.
Clover, C. (1989). Her body, himself: Gender in the slasher film. In Donald, J. (ed.) *Fanatasy and the Cinema*, pp 91 133. London: British Film Industry.
Conley, T. (2007). *Cartographic Cinema*. Minneapolis: University of Minnesota Press.
Creswell, T. and Dixon, D. (eds.) (2002). *Engaging Film: Geographies of Mobility and Identity*. Lanham, MD: Rowman and Littlefield.
Deleuze, G. (1986). *Cinema 1: The Movement Image* (trans. Tomlinson H. and Habberjam B.). London: The Athlone Press.
Friedberg, A. (1993). *Window Shopping: Cinema and the Postmodern*. Berkeley, CA: University of California Press.
Jameson, F. (1992). *The Geopolitical Aesthetic: Cinema and Space in the World System*. Bloomington, IN: Indiana University Press.
Metz, C. (1974). *Language and Cinema*. The Hague: Mouton.
Mulvey, L. (1975). Visual pleasure and narrative cinema. *Screen* 16(3), 6 18.
Silverman, K. (1992). *Male Subjectivity at the Margins*. London: Routledge.
Virilio, P. (1989). *War and Cinema: The Logistics of Perception* (trans. Camiller P.). London: Verso.

Multicultural City

H. Hoernig, Université de Québec à Montréal, Montréal, QC, Canada
M. Walton-Roberts, Wilfrid Laurier University, Waterloo, ON, Canada

Glossary

Assimilation Process of acquiring characteristics and qualities of the dominant group through behavior and/or residential settlement patterns.

Cosmopolis An urban region inhabited by people of different national and ethnic backgrounds; also used by Leonie Sandercock to refer to an ideal city where all people coexist respectfully and harmoniously, amidst social, cultural, and racial diversity.

Culture A term of many definitions, here defined as a shared and dynamic system of meaning and power through which people live out and recreate social life.

Difference In relation to identity, refers to a wide spectrum of key human qualities related to individual and social characteristics such as gender, race, ability, sexuality, etc.

Ethnic Enclave A model of ethnic settlement where dense clustering of residential settlement of a single ethnic group occurs, accompanied by economic and social activity.

Ethnoburb A model proposing a form of suburban multiethnic settlement formed by national and local contextual factors and global restructuring, with characteristics of both a traditional ethnic enclave and a global economic outpost, and socially stratified by inter- and intraethnic differentiation.

Heterolocalism A model proposing that through transportation systems and communications technology, ethnic groups maintain their identities and communities across regionally dispersed residential settlement patterns punctuated by concentrated nodes of social or commercial activity.

Hybridity New social and cultural products and processes which arise from the mixing, crossing, or displacing of boundaries between people, media, and materials.

Integration Process of incorporating newcomers into social, economic, political, and spatial spheres of society.

Multiculturalism Idea that cultural diversity is integral to society.

Segregation Exclusion of a social group from the rest of society, usually used in reference to spatial residential patterns, but can also relate to economic, social, and political spheres.

Introduction

As urbanization processes continue to dominate regional and national development, migration and immigration trends are transforming the sociocultural composition of urban regions and introducing a growing spectrum of cultural diversity. As a result the multicultural city, as an urban demographic reality, is an increasingly common phenomenon throughout the world. For example, the major immigrant receiving metropolitan areas of London, Toronto, New York, and Dubai are each home to more than over 100 ethnic and linguistic communities. Other cities combine both national and international sources of cultural diversity such as Johannesburg, Delhi, and the city state of Singapore.

In its most basic application, the multicultural city concept denotes a city of many cultural groups. In contrast to those urban societies with more homogeneous populations, this diversity has direct consequences for the social, cultural, economic, and political life of the city. However, in its ideological sense, the multicultural city also conveys multiculturalism, that is, the idea that cultural difference in the city can be socially sanctioned and celebrated. In other words in an ideal multicultural society different cultural groups coexist and simultaneously maintain their cultural integrity and readily participate in an equitable, urban society.

The Historical Multicultural City

Multicultural cities in a demographic sense are not new. Urban interethnic mixing has accompanied urban development throughout history. Scholars have employed various taxonomies to explore the historical evolution of cities and their associated ethnic heterogeneity, with the most effective approaches linking the political and economic context of the time to the nature of ethnic relations within the city. A general model of such cities would include four types. The first includes preindustrial or ancient cities such as Rome or Constantinople where ethnic and religious pluralism existed and the cities could be considered, by some measures, multicultural and cosmopolitan. The second general category includes colonial cities as illustrated in Furnivall's work on plural societies in the Dutch East Indies. Colonial processes initiated the movement of various groups primarily for economic purposes, and their settlement in the city was marked by a racialized hierarchy that often delimited

Figure 1 Entry to Vancouver's inner city Chinatown. Source: Zhi Xi Zhuang.

settlement and employment patterns. Furnivall argued that these societies were characterized by power hierarchies and ethnic plurality, where ethnic groups experienced only limited social interaction. A third type can be termed industrial cities, such as Chicago, that in the early twentieth century became the focus for seminal work on ethnicity and urban form. More recent scholarship has examined the postindustrial city, shaped by globalization and transnational flows of migrants, capital, and goods. Recent work on the cities of Vancouver and Los Angeles, for example, examines the socioeconomic and political transformations that accompany rapid increased global migration, and demonstrates the specificity of associated urban change.

Historic and geographic context are also critical in understanding the changing discourses attached to ethnic difference within the city. In late nineteenth century Vancouver, Canada, for example, official and quotidian forms of discrimination contributed to the formation of an inward looking ethnically and economically segregated ethnic Chinese community. Over the past century economic, social, and political shifts, including dramatic changes in immigration and equity policies, have resulted in tremendous changes in Chinese immigrant settlement patterns and economic activities. For example, Chinese immigration to Vancouver in the early 1990s was dominated by wealthy Hong Kong and Taiwanese migrants fleeing the uncertainty of Hong Kong's 1997 repatriation to China. Settlement during this period was directed to suburban, rather than traditional inner city neighborhoods, and was accompanied by a burst in transnational economic activity, including high technology, real estate, and professional services. These developments led to the emergence of new urban spatial forms such as the 'ethnoburb', the 'monster house', retail condominium style shopping malls, and significant inner city redevelopment characteristics of postindustrial landscapes. Over the last century there has been a transformation in the discourses associated with ethnic Chinese immigrant settlement, revealing complex intersections of ethnicity and class. Additionally, there has been a radical change in the built environment related to Chinese immigration, as **Figures 1** and **2** demonstrate.

The Contemporary Multicultural City

The multicultural city can be divided into three broad categories based upon sources of diversity. There are those long standing multicultural cities of multiethnic states of Europe, Africa, and Asia, which continue to diversify due to national migration and urbanization. Second, there are the more recently developed cities of the settler societies of North and South America, Australia, and New Zealand; where over the past century immigration has contributed extensively to growth and development of cities in these areas. Third, there are those cities where cultural diversity includes both national ethnic groups and those from immigrant sources.

Within these three groups, there are several sources of urban cultural difference. First, there are indigenous populations, those original inhabitants of countries affected by colonialization and settlement, resulting in a majority national settler population and minority

Figure 2 False Creek development by the Chinese diasporic entrepreneur Li Ka Shing in Vancouver, Canada. Source: Margaret Walton-Roberts.

indigenous population, for example, countries of North America and Australasia. Colonialism has also left a legacy of ethnic diversity in these countries due to forms of voluntary, enforced, and indentured labor mobility, such as Fiji, parts of the Caribbean, East Africa, and South America. Second are the national ethnic popu lations of those long standing multiethnic nations in cluding India, most African and Asian countries, and some European nations, for example, Belgium, and the United Kingdom. This category would also include mi grant ethnic minorities such as the Irish travelers, and the Roma (gypsies). The third category includes inter national ethnic populations derived from more recent waves of immigration which have taken place within a wider context of colonial and neocolonial connections.

Understanding Urban Multiculturalism

The term multicultural has been applied widely and therefore it is important to examine the various ways in which it has been used. The concept culture itself is a complex term of numerous definitions. In this context, it can be understood to refer to a learned set of interrelated beliefs, ideas, and behaviors shared by a group of people. Although culture is shared and therefore a social phe nomenon, scholars of recent decades emphasize that culture is dynamic and contested and insist that it must be contextualized and understood within broader social processes. In addition to culture, race and ethnicity are key concepts to this field of inquiry. Race (scientifically invalid but socially salient) refers to human phenotype or physical characteristics, often skin color, and ethnicity refers to the boundaries defined between social groups on the basis of certain shared features such as language and history. The three terms are used in overlapping ways by urban residents and scholars in reference to urban multicultural issues.

The concept of the multicultural city is significant because it signals a particular understanding of the city that includes cultural difference. Many conceptual izations of the city do not explicitly recognize cultural difference. In some urban discourses cultural diversity has been merely ignored or deemed irrelevant, while in others cultural difference is assumed to be a temporary condition while cultural minorities assimilate to the majority population. The use of the term multicultural city brings cultural diversity to the forefront as a per manent condition.

Many scholars describe several levels or uses of the concept multiculturalism. These uses can be directly applied to various themes of multicultural city scholar ship as illustrated in **Table 1**. First, the term multi culturalism can be used to describe a demographic condition. This description of the general characteristics of a society's population includes the statistical exercise of counting and describing 'cultural' bodies. The second use refers to the social reality of multiculturalism that is, the 'on the ground' description and analysis of the intra and intergroup relations. The third use refers to the ideological or normative sense of the concept. Here, multiculturalism is a way of thinking about how people of different cultures should be able to function in society, and how they should interact with one another. The fourth is a political use and includes the national, re gional, and local politics involved in such debates, as well

Table 1 Uses of the concept of multiculturalism as they apply to the multicultural cities

Uses of the concept multiculturalism	*Application to the multicultural city*
Demographic	Urban demographic; national, regional, and local residential settlement patterns
Social reality	Social inclusion or exclusion; social and economic integration or segregation (voluntary or involuntary); uraban inter- and intraethnic relations
Normative	Urban normative discourse of the multicultural city, e.g., Sandercock's cosmopolis
Political	Urban policy: e.g., land-use planning, urban infrastructure and services; ethnic minority participation in urban citizenship
Critical discourse	Critique of urban ethnic and racial relations; urban social movements which challenge unequitable urban policies or management practice, e.g., housing, policing, immigrant support services, etc.

as resulting policies and statutes designed to manage cultural diversity. Finally, the fifth use of multiculturalism can be as critical discourse. Here multiculturalism as critique or debate can be seen as the means by which marginalized peoples can criticize and resist oppressive forces or structures in society. In this way, a transformative multiculturalism is a movement which draws upon cultural difference to destabilize the status quo and work toward more equitable, intercultural relations.

Another important dimension to the concept of the multicultural city is the evolution of ideas related to multiculturalism. These ideas have been examined, negotiated, translated into policy and programs, and vociferously debated in various ways over time and within different national, regional, and urban contexts. For example, in Canada, one of the few countries that has established official federal policies, legislation, and programs related to multiculturalism, scholars have noted major shifts in the ways that multiculturalism has been interpreted and implemented over the past 40 years of its official existence. While there have been several renderings of the Canadian chronology, most versions recount the ways in which the policy shifted from its somewhat superficial focus on only cultural events and programs ('song and dance' multiculturalism), to a more critical stance in which policy and programs are designed to address issues related to social exclusion, discrimination, and equity. Debates in various other immigrant receiving countries such as Australia, the UK, and European nations have also been dynamic, characterized by varying waves of support and opposition including vigorous debates concerning national and regional identity, specifically related to issues such as the freedom of religion (e.g., 'the *foulard* or headscarf affair' in France), the freedom of speech, and the costs and benefits of immigration.

Finally, aspects of multiculturalism can be categorized by three issues arising from cultural diversity, each with a corresponding set of implications that are summarized in **Table 2**. One set of issues relates to the recognition of cultural difference; characterized by a demand for cultural recognition that requires some type of accommodation. A second set of issues relates to discrimination and racism under existing polity where claims are made for fair treatment and equal human rights. The third set of issues occurs at the intersection of ethnicity with one or more other dimensions of difference. In these cases, it is the combination of two or more characteristics that results in the marginalization of an individual or group.

The Multicultural City: Urban Structure and Function

The relationship between multiethnic populations and urban structure and function has long been a preoccupation of urban scholars. Various models of the North American city such as those developed by the Chicago school (e.g., Burgess' concentric zone model) included residential patterns along ethnic and racial lines into their conceptualization of urban structure and change. Early models were premised upon ecological processes of invasion and succession or filtering. Scholars have criticized the Chicago school for its ecological approach, and its assumption that broad social categories of race or ethnicity are stable coherent classifications that can be uncritically mapped onto the urban landscape.

More recent work on social difference in the city has highlighted the ways in which urban political economic and sociocultural processes have contributed to the social construction of ethnic and racial identity, as well as the complex and dynamic ways in which social identities become 'fixed' in the built form. More recent scholarly approaches to human geography have also recognized that culture forms are only one of the many axes of difference (e.g., gender, class, ability, race, religion, age, etc.) that intersect to determine the tremendous dynamic social complexity of urban life. Such issues become of the greatest concern when this intersection contributes significantly to social marginalization or exclusion. Most scholarship addresses this theme directly or indirectly by examining the ways in which groups face barriers, and experience or overcome their social and economic disadvantage. A second, related theme is urban citizenship, a term which encompasses both the rights and responsibilities of urban residents, including political and social

Table 2 Three readings of issues related to urban cultural diversity

	Issue	*Group claims*	*Urban issues*
Accomodation	Cultural difference	Claims to cultural recognition	Architecture and land-use planning issues (housing, places of worship, retail); heritage and memorial sites; recreation and leisure facilities; cultural celebrations (festivals, parades); language rights; freedom of religion (dress, religious practice etc.)
Equity	Discrimination/ marginalization	Claims to human rights/equity	Equitable access to urban infrastructure, services, and programs; access to affordable housing; employment equity; participation in urban governance
Intersectionality	Social marginalization or exclusion due to the combination of more than one dimension of identity	Claims to equity, in recognition of the multiple or interacting barriers faced by individuals or groups	Problems experienced by specific groups, e.g., immigrant, single mothers or disabled people of an ethnic minority, usually in relation to other issues such as housing, employment, or support services

participation in urban, civil life. We explore these in more detail below.

Spatial Patterns

One way to investigate cultural diversity has been to examine the ways in which different groups are distrib uted across the urban landscape, usually querying the correlations of their locations with other variables such as education or date of arrival. There are several scales at which this investigation takes place. One approach is to examine a country as a whole, and study settlement patterns across a national urban system. In most cases, settlement patterns are not uniform, but rather, reflect the influence of other economic, political, and social factors related to conditions of entry or of migration and settlement processes. Most immigrant receiving coun tries have one or more gateway cities, adjacent to a port of entry, hosting higher numbers of foreign born resi dents, for example, New York, Vancouver, and Sydney. Many multiethnic countries have greater levels of cul tural diversity in those cities supporting high rates of economic activity, for example, Accra, Johannesburg, or Mumbai.

A second approach is to examine intra urban patterns of settlement. Classic North American studies of recent immigrants have examined ethnic enclaves, or areas of concentrated co ethnic residential settlement and social and economic activity (see **Figures 3** and **4**). Many human geography studies of recent immigrants have concerned the relationships between particular groups (as defined by different variables, e.g., country of origin, race, etc.), economic success or deprivation (as mediated by factors such as education and income) and intra urban residential location. Residential segregation, both volun tary and involuntary, has received particular attention in relation to its implications for urban interethnic relations. Scholars have also examined ethnic settlement in relation to broader urban patterns of dispersal and recentraliza tion, such as Li's ethnoburb model or Zelinsky and Lee's concept of heterolocalism. Scholarship has also addressed issues related to informal settlements, often located on peripheral sites and populated primarily by migrants and the poor. While this has been primarily a developing world issue, it has also been relevant to developed world cities where problems of homelessness and affordable housing supply are acute.

A more finely grained analysis has been used in yet a third approach. Here scholars have examined individual trajectories and the micro sites of multicultural inter action and exchange, to investigate on the ground oc currences of interethnic negotiation in specific places such as workplaces, schools, parks, and streets or in ref erence to particular events like a conflict or a decision making situation. These studies reveal the active and dynamic construction of both boundaries and alliances between individuals and groups upon multiple grounds. Such studies also indicate the role played by political structures and institutions in mediating, managing, and facilitating such interactions.

Economic and Political Issues

Multicultural populations have implications for the urban economy in both its formal and informal sectors. Economic integration is often a critical concern for urban ethnic minorities and a central component to the settlement and integration process for newcomers. Many geography studies have explored various patterns of economic activity related to immigrant, migrant, or ethnic minority groups. Again these studies explore a host of factors related to economic integration. Work in the area of global cities has highlighted the social po larizing effects of globalization through the concentration

Figure 3 Scenes from a suburban ethnoreligious residential enclave: the Ahmadiyya Peace Village in Vaughan (Ontario, Canada), a neighborhood developed by an Ahmadiyya Muslim community. Source: Heidi Hoernig.

of employment, particularly for recent immigrants, in low paying service sector jobs. Other scholarship has documented the difficulties many minorities and recent immigrants encounter upon entering the formal labor market, for a variety of reasons (e.g., language and cultural barriers, systemic racism, lack of recognition of foreign credentials, insufficient education, etc.). As a result, many pursue self employment through small business development. The economies of multicultural cities also include a range of informal economic activities including street vending, sex trade work, and begging. Scholars have explored the ways in which race, ethnicity, and culture have been significant in the participation of and outcomes for certain groups involved in these activities. Cities also draw upon cultural diversity in various ways such as in the promotion of tourism, in the portrayal of positive images for consumption (e.g., a downtown cosmopolitan lifestyle), or in strategies designed to attract additional economic development and immigration.

Singapore is a good example of a city that has diligently 'imagineered' itself as a cultural city, and cultural diversity has been one of the central tropes the city authorities have utilized in order to compete for tourism and business investment in the city. In reshaping the city processes of heritage designation and gentrification have, however, ironically worked to displace the very segments of the population this process of urban boosterism seeks to revalorize. The example of Singapore demonstrates the tension between the cultural city as façade and the multicultural city as home.

The participation of different ethnoracial groups in urban political arenas has been another fertile area of scholarship. Scholars have also examined situations in which urban governments integrate cultural diversity into policy development, programs, and service delivery, including the use of multiple languages and the official recognition or celebration of significant cultural and religious events. Culture emerges as an important political issue in those multiethnic cities where indigenous or traditional political institutions, as well as representatives of the formal government, participate in local decision making. Urban politics are also woven into struggles for access to municipal services for immigrant, migrant, or cultural minority groups; in areas such as affordable

Figure 4 Proposed Punjabi retail and service commercial plaza adjacent to Sri Guru Singh Sabha gurdwara on Airport Road, Malton (Mississauga, Ontario, Canada). The sign reads 'Great Punjab Business Centre: Where Business and Culture Come Together'. Source: Heidi Hoernig.

housing, social and employment services, and urban infrastructure provision.

Social and Cultural Issues

Within the city certain groups may be disadvantaged by cultural or ethnic difference, but ethnocultural groups can also benefit from institutional completeness, where multiple co ethnic organizations create social and political capital for the group. Research has explored how intragroup differences (such as gender or age) affect resource distribution, and this line of enquiry represents an analysis that is more responsive to the intersection of identities.

Cultural issues can be viewed through specific features which transform how people experience the multicultural city. Language is a key issue expressed in various ways: through public sector channels when officially recognized or through group specific means such as through ethnic media, business (see **Figure 5**), social and cultural networks. The current tension in southern US border towns over the use of Spanish illustrates this dimension well. Religion is also often a central component to urban cultural difference. It plays out spatially through features such as residential patterns, places of worship (see **Figure 6**), religious schools and colleges, shrines, bookstores, and retirement homes. An overlooked but omnipresent aspect of the multicultural city is time. Cultures are frequently characterized by distinct temporalities, often in association with religion. Daily cycles and calendars, often following a lunar rather than the Gregorian calendar, structure the flow of individual and group activities. Multicultural cities are therefore also multitemporal spaces, although in most cases minority groups must adapt their temporal sequence to the dominant temporal culture.

Festivals, holidays, and other cultural and religious events are a familiar expression of urban multicultural temporalities. Scholarship on multicultural festivals can range from urban boosterism and place marketing to critical examinations of what Hage terms 'cosmo multiculturalism', where ethnic difference is contained and made politically palatable for the middle class through the spectacle of food and the festival.

Urban Planning, Development, and Management

One final important approach to the study of the multicultural city has been that of urban processes of planning, development, and management. This area includes

Figure 5 Multilingual signage (Korean, Chinese, English, and Persian) of a multiethnic retail and service commercial strip along Yonge Street, North York (Toronto, Ontario, Canada). Source: Heidi Hoernig.

Figure 6 Ethnoreligious pluralism along Highway No. 5 in Richmond (Canada), where 15 places of worship are located within two blocks between Blundell Road and Stevenston Highway. Five religions are represented in mandirs (Hindu), a temple (Buddhist), a gurdwara (Sikh), a mosque (Muslim), and churches (Christian). Shown above in this photo is the Gurdwara Nanak Niwas (Sikh), and the Az-Zahraa Islamic Centre (Shia). Source: Heidi Hoernig.

official public sector actors and organizations, as well as those in private and nongovernmental sectors. A body of scholarship has coalesced around issues that arise when multiethnic residents have differing beliefs and behaviors related to basic urban planning and development issues. This includes a group of works which examines conflict in multiethnic cities. A range of issues have been ex plored through conflicts that have erupted over devel opment of ethnic retail, housing, places of worship, funeral homes, and heritage sites. Conflicts have been attributed to a number of causes including cultural fac tors, racism, and fear. While land use conflict is common to urban development, this scholarship has demonstrated some of the practical challenges of planning and man aging a multiethnic community. Key issues in the area of urban governance amidst cultural diversity include the need to reexamine public interests in view of culturally diverse needs and the importance of civic structures and processes that allow urban residents access to and in volvement in public decision making. Scholars have documented those few city governments which have made significant progress in developing innovative methods for addressing some of these issues.

For other multicultural cities, mostly in developing world cities, the most acute urban planning and devel opment issues concern basic urban infrastructure and service delivery for water, waste disposal, electricity, and transportation. Here again, exclusion from these critical services has resulted from the intersection of cultural, ethnic, and racial factors with other political and eco nomic variables.

Key and Emerging Issues

The concept of the multicultural city is germane to urban scholarship in an increasingly globalizing and ur banizing world. Several issues identified will continue to influence the direction and scope of this work. One central question concerns the ways in which cultural, ethnic, and racial difference is conceptualized and ana lyzed. Social science scholars have drawn methodological and analytical attention to factors of both stasis and change, in their understanding of the multiple ways in which difference is constructed. Scholarship related to the multicultural city also faces the challenge of under standing the interplay between social and cultural pro cesses related to identity and representation and the political economic processes addressing the distribution of power and resources.

International comparative studies provide both inter esting and challenging questions about the ways in which issues related to cultural diversity emerge in different cities. Although scholarship in both the global North and South include many studies that simultaneously examine issues of cultural difference with other concerns (e.g., housing, governance), scholarship related to the multi cultural city in developed countries commands its own growing subfield. To date in many of the studies of the global South, multi and interethnic issues are largely overshadowed by the overwhelming need to address more basic issues concerning access to urban infra structure, services, and poverty. However, there are ex ceptions, for example, violent clashes in Baroda, India, between Muslims and Hindus have been interpreted through the lens of the multicultural city.

Several related concepts also highlight future areas of research related to the multicultural city. One concerns the general question of how cultural difference intersects with other dimensions of difference. As discussed above, many issues related to culture, ethnicity, and race be come most relevant when they coincide with other characteristics such as income or gender. This raises the question of the usefulness of the concept of multicultural city, where perhaps an open focus on difference is more conceptually and analytically effective. Another import ant concept is that of transnationalism, a social phe nomenon which draws further complexities into our understanding of how culture plays out in various re lations that take place across time and space. Third, the concept of cosmopolitanism, in its cultural variant, con cerns an important characteristic of some urban residents, as well as an emerging discourse put forth by urban residents and decision makers. Finally there are some interesting lines of enquiry to be drawn between modern day and historical multicultural cities, especially with reference to the economic, political, and cultural di mensions of interethnic relations and governance practices. Scholars have also pondered the utility of examining the historical antecedents of the multicultural city and contrasting the relatively positive social and political experiences of ethnic groups in ancient multi cultural cities, versus the exclusions and divisions evident in many modern metropolises today.

Globalization, immigration, and urbanization trends underscore the importance of the multicultural city in the twenty first century, not only as an ethical ideal, but increasingly as an economic imperative. Understanding and improving the human capacity to peacefully coexist in multicultural communities is crucial to future just and prosperous urban development. Scholarship that strives to understand how issues of identity and difference are socially constructed in the urban milieu are of increasing relevance both academically and politically. Past urban riots such as those throughout French cities (2005) and in Sydney (2005), Baroda (2002), Lagos (2002), Oldham (UK, 2001), Mumbai (1993), and Los Angeles (1992) are jarring reminders of the potential spiraling costs of unattended interethnic conflict and socioeconomic inequities for ethnic minorities.

See also: Cosmopolitanism; Ethnicity; Immigration I; Multiculturalism; Race; Religion/Spirituality/Faith; Social Geography; Transnationality.

Further Reading

Amin, A. (2002). Ethnicity and the multicultural city: Living with diversity. *Environment and Planning A* 34, 959 980.

Anderson, K. (1991). *Vancouver's Chinatown: Racial discourse in Canada*. Toronto: McGill Queens University Press.

Binnie, J., Holloway, J., Millington, S. and Young, C. (eds.) (2006). *Cosmopolitan urbanism*. London: Routledge.

Briggs, X. d. S. (2004). Civilization in color: The multicultural city in three millennia. *City and Community* 3, 311 342.

Burayidi, M. (ed.) (2000). *Urban planning in a multicultural society*. Westport, CT: Praeger.

Fincher, R. and Jacobs, J. (eds.) (1998). *Cities of difference*. New York: Guilford Press.

Furnivall, J. S. (1948). *Colonial policy and practice*. Cambridge: Cambridge University Press.

Grillo, R. D. (2000). Plural cities in comparative perspective. *Ethnic and Racial Studies* 23, 957 981.

Hage, G. (1998). *White nation: Fantasies of white supremacy in a multicultural society*. New York: Routledge.

Laguerre, M. (2003). *Urban multiculturalism and globalization in New York City: An analysis of diasporic temporalities*. New York: Palgrave MacMillan.

Li, W. (1998). Anatomy of a new ethnic settlement: The Chinese ethnoburb in Los Angeles. *Urban Studies* 35, 479 501.

Olds, K. (2001). *Globalization and urban change: Capital, culture, and Pacific Rim mega projects*. Oxford: Oxford University Press.

Patel, A. (2003). Cultural diversity and conflict in multicultural cities *Fondazione Eni Enrico Mattei Research Paper Series* 100. Available at http://www.feem.it/Feem/Pub/Publications/WPapers/default.htm

Qadeer, M. (1997). Pluralistic planning for multicultural cities: The Canadian practice. *Journal of the American Planning Association* 63, 481 494.

Rogers, A. and Tillie, J. (eds.) (2001). *Research in Migration and Ethnic Relations Series. Multicultural policies and modes of citizenship in European cities*. Aldershot, UK: Ashgate.

Roseman, C., Laux, H. D. and Thieme, G. (eds.) (1996). *EthniCity: Geographic perspectives on ethnic change in modern cities*. Lanham, MD: Rowman & Littlefield.

Sandercock, L. (2003). *Cosmopolis II: Mongrel cities in the 21st century*. London: Continuum.

Yeoh, B. (2004). Cosmopolitanism and its exclusions in Singapore. *Urban Studies* 41, 2431 2445.

Zelinksy, W. and Lee, B. (1998). Hererolocalism: An alternative model of the sociospatial behaviour of immigrant ethnic communities. *International Journal of Population Geography* 4, 281 298.

Relevant Websites

http://www.immi.gov.au
Australian Government Department of Immigration and Citizenship.

http://www.canadianheritage.gc.ca
Department of Canadian Heritage.

http://www.metropolis.net
Metropolis: An International Forum for Research and Policy on Migration, Diversity, and Changing Cities.

http://portal.unesco.org
UNESCO International Migration and Multicultural Policies.

http://www.unhabitat.org
UN Habitat: State of the World's Cities Report.

http://www.urban.org
Urban Institute, A Nonpartisan Economic and Social Policy Research Organization.

Multiculturalism

J. Clayton, University of Sunderland, Sunderland, UK

Glossary

Assimilation The incorporation of previously culturally distinct minority groups into the cultural majority of a nation.
Heterogeneity The existence of social and cultural differences within a given population.
Identity Politics Political action carried out on behalf of a specifically defined social or cultural group.
Integration The manner in which migrant and minority groups are included and treated within 'host' societies.
Pluralism A society which is defined in terms of an affirmation and acceptance of social and cultural diversity.

Introduction

The term 'multiculturalism' is now commonplace within contemporary academic as well as popular discourse, being as it is at the heart of a number of contentious and passionate debates around; nationhood, belonging, community, immigration, race, ethnicity, religion, equality, and human rights. It remains a controversial concept and practice supported by many, but increasingly dismissed by those from a range of political traditions and theoretical standpoints. Multiculturalism is far from a straightforward idea and has been interpreted by geographers, other social scientists, and policymakers in a variety of respects, often dependent upon the motivations of its use and the context of its employment. As a consequence, its contours often remain ill defined and unclear.

Broadly speaking, there are two main strands of thought in geography and related disciplines of sociology, politics, and cultural studies as to what multiculturalism entails. The first of these is that multiculturalism stands for the very condition of living with social and cultural diversity within specific nations and places, that is, the social fact of the proximate coexistence of those identifying as different in a variety of racial, ethnic, and religious respects. The second strand refers to a set of formal institutional practices which juggle notions of equality, fairness, and justice while at the same time recognizing the presence and rights of racially, ethnically, and religiously defined groups, in an attempt to provide for populations with varying needs. Both of these definitions recognize social and cultural differences as significant to an understanding of the term; however, the latter stresses the normative recognition of difference and the inclusion of personal and collective cultural markers within the public realm where they are protected, catered for, but also legitimized and reified.

Multiculturalism as Diversity

The premise upon which multiculturalism as a political mode of integration is based is that there exists in modern nation states heterogeneous populations who do not share racial or ethnic affiliations, cultural practices, or geographical attachments. Multiculturalism therefore speaks of a condition of demographic pluralism, something which clearly defines any society in some shape or form, but focuses here upon racial, ethnic, and religious differences.

Shifting patterns of social and cultural diversity have historically been driven by prevailing patterns of internal as well as international migration. Motivations for population movements vary considerably but in general terms have been driven by opportunity, force, economic inequality, and survival as a consequence of, among other processes, urbanization, industrialization, slavery, conflict, and natural disaster. Given histories of economic and political domination and the relative instability of today's less developed countries, it is no surprise that the majority of such movements has taken place either between countries in the global south or from these countries to the global north. The most common destinations for international migrants in recent years include the United States, Canada, Australia, and countries in Western Europe, including Germany, France, and the UK. However, in terms of the destination of refugees, who have a well founded fear of persecution in their home nation, 70% currently live in less developed countries outside of North America or Europe. The complexity of today's multicultural societies across the world can be understood as a legacy of historical processes which have resulted in population movements, but also as a continuing product of accelerated globalization.

Despite the history of multiculturalism as a social fact and its spatial pervasiveness, it has only really gained currency and entered into academic and public discourse in the Western world in recent decades. This can be seen as a direct consequence of and response to growing racial, ethnic, and religious diversity in countries, including the UK, the United States, Canada, and Australia, which have been unofficially and more officially marked out as 'white' national spaces. While what we know today as the British Isles has always been a place successively settled by groups of various origins, the arrival of migrants from the Caribbean in the 1950s and then from the Indian sub Continent in the decades that followed, signaled the beginning of the UK conceived of as a multicultural nation. In this sense multiculturalism does not just refer to the coexistence of socially and culturally distinct groups, but also the arrival and presence of those visually marked out as 'different', seen to disrupt some imagined monoculturalism. In the case of the UK, this mono culturalism has been equated with its 'whiteness'.

In the case of the UK, there is no denying that these migration patterns have had a significant impact on the face of the nation. Of a population now over 59 million, 7.9%, that is, 4 635 296, identify as belonging to 'non white' ethnic groups, mostly, but by no means entirely composed of individuals with heritages traced back to the former colonies in the Caribbean, India, Pakistan, and Bangladesh. In the context of Canada this figure is much higher at 18.9%, and in Australia nearly one quarter of the population (24.6%) now identifies as 'non white'. In recent years, the picture in each of these nations has become even more complex. For instance, restrictions placed on the movement of global economic migrants, refugees, and asylum seekers have become more strin gent, and countries such as the UK have also seen the arrival of new migrant groups from an expanded Euro pean Union. While political reaction to the growth of minority ethnic groups, particularly 'non white' groups, in these countries has been hostile, particularly where far right political parties have gained ground, there is increasing acceptance that multiculturalism is now a defining feature of these societies.

Multicultural Cities

Along with considerations of the impact of diversity upon nation states, the concept of multiculturalism has also been used to describe and analyze the experience of cities. As most migrant communities initially gravitate toward urban areas, attracted by the prospects of work, the availability of services, and the importance of securing support networks, cities have become the focal point of attention among geographers examining multiculturalism in action. For example, 45% of the UK's ethnic minority population now live in London. Indeed, it has been suggested that in the context of multiculturalism cities have become more im portant than nations in defining the identities of indi viduals, senses of belonging, and notions of citizenship. Cities with global connections are becoming the hubs of multicultural activity and the arenas where the challenges of living with diversity are enacted through the everyday lives of its citizens.

Geographers have focused their attention upon the relationship between multiculturalism and cities by em phasizing the spatial effects of racial, ethnic, and cultural 'difference' as played out upon the urban environment. Those working through positivist traditions have pro ceeded to map the residential settlement and location of minority ethnic communities through the analysis of census data and by measuring levels of ethnic isolation by using segregation indices. There has also been an em phasis on the processes through which patterns of race and residence emerge, focusing upon the political economy of residential location and the structural con ditions under which these patterns develop. This work has highlighted the relation between space and 'race', particularly through an emphasis upon the residentially divided geographies of multicultural cities as a result of positions of inequality, discrimination, and 'community' formation. However, such approaches have also attracted criticism. For instance, it has been argued that ethnic concentration is often treated as an inherently prob lematic phenomenon and that complex identities are often reduced to cartographic patterns, neglecting more detailed personal and collective histories, relations of power and the significance of global networks, and distant places beyond the boundaries of the urban neighborhood.

On the other hand, as part of an accepted 'cultural turn' within the discipline and with reference to emerg ing literatures on the construction of racialized identities through social practice and globalization, geographers have placed greater attention on the experiential aspects of multicultural cities. In doing so, geography has bor rowed from cultural studies, urban sociologists, and ethnographers by addressing the realities of multi culturalism as lived out. Geographers have also recog nized the need to take more seriously the productive capacities of urban multicultural spaces and their role in the reproduction of diverse identities as well as the local translation of global processes as everyday situated phenomena. Other emerging avenues of research include the effects of transnational commodity cultures in cre ating new forms of ethnicized identities, the socio spatial construction, and negotiation of minority and 'white' ethnicities, particularly in terms of youth subcultures and cross cultural exchange and racialized representations of place through which powerful meanings ascribed to places work to reinforce specific interpretations of 'race'. In particular, these studies have been successful in opening up the importance of place and urban sites as powerful elements within processes of racialization.

Multiculturalism as Public Policy/Mode of Integration

Multiculturalism can also be identified as one version of an ideological and institutional response to the per ceived challenges which diversity and reactions to diversity pose. Much of the current literature devoted to this definition of multiculturalism has emerged from political philosophy and considers in both abstract as well more empirical ways how modern nation states may and indeed do deal with the changing relationships between the individual, communities, and the nation state in the context of global and societal change. By and large, this is approached through a liberal political lens but also takes into account strands of communitarian thought and the significance of identity politics, by outlining the possibilities of accommodating the diverse needs and practices of minority groups while maintaining core democratic ideals and valuing the rights of the individual.

A significant shift from assimilationist policies in several Western nations was witnessed in the late 1960s and 1970s amidst rapid economic restructuring, a broad political and ideological struggle between right and left, and the growing influence of social movements which highlighted the marginal position of working class and minority communities. The dominance of assimilation as the prevailing mode of integration of minority ethnic groups in the period preceding this demanded that minority communities adjust their social and cultural practices, customs, and beliefs in order to fit in with those of the national majority. This was based upon a desire for some sort of level playing field through the secularization of the public realm (although the public realm itself is never entirely neutral) but was also driven by a longing to preserve an 'authentic' sense of national identity. However, given pervasive cultures of racism and the power relations involved in the integration of migrant communities, it was increasingly clear that migrant groups could not obtain fair treatment.

Not all nations in this period underwent this shift away from assimilation and for those that did their trajectories were often divergent. In France, for example, with its national culture historically rooted in republican ideals of the secularization of the public realm and the United States with its overriding notion of 'melting pot' assimilation, specific versions of integration which valued allegiance to the nation above all other forms of identification became entrenched. However, particularly in the UK, Canada, Australia, Sweden, and the Netherlands, multiculturalism as public policy began to represent the official government approach to the challenges of diversity as part of a concerted effort by those on the political left to undermine assimilationist, nationalistic, and reactionary sentiment. In effect this was a challenge both to the idea of homogenous nations and to the structural inequalities and discrimination experienced by those outside of the national majority, something which was not solely imposed, but also developed from below by those struggling to deal with the realities of unequal societies.

In the case of Australia, the adoption of multiculturalism as an official policy in the early 1970s came as a result of persistent lobbying activities on behalf of activists directly involved with and concerned for the welfare of marginalized minority ethnic groups. These activists, including welfare workers and academics, were able to impress upon the national government at the time the need to push a multicultural agenda to the forefront of public policy by attempting to involve various minority groups. It took several years for the government to become persuaded by this argument eventually bringing in a multicultural agenda firstly in their dealings with ethnic affairs policy and then into areas of education and social policy by the mid 1970s. By 1975 the federal departments of labor and immigration, social security, education, the media, and the Attorney Generals department were all under the influence of a policy of multiculturalism.

In Canada during the 1960s and 1970s, there was also a growing concern over issues of inequality which were increasingly stark, particularly highlighted by those politically aligned to the left. As a consequence, the country witnessed many changes in terms of its social policy as regards public healthcare, pension schemes, support of the unemployed, and expansion of the education sector. The introduction of multiculturalism as an official policy can be seen as part of these changes along with the increasing pressure exerted on the government by minority French speaking Canadians in their campaigns for a bilingual nation.

Within the context of the UK, academic and political debate in this period was largely concerned with the manner in which New Commonwealth immigrants could be incorporated into mainstream British society, recognizing the disadvantaged positions of minority communities and their racialized exclusion in terms of housing, employment, education, healthcare, and other social services, particularly in major urban centers. While antidiscrimination legislation in 1965, 1968, and 1976 marked the beginning of the adoption of multicultural policies by the British government, it was clear that the impetus for these changes grew out of antiracist protest and grass roots movements, a cause which was then taken up by the Labour party. It was out of these struggles that later multicultural policies emerged. The political will behind these ideas were at once emancipatory and universalist, appealing to ideas of inclusion, equality, and social justice but at the same time they endorsed a form of particularlism which desired that the needs of minority groups be met and provided for within the public domain.

Despite these developments, by the 1980s a Conservative controlled national government remained opposed to the concept of multiculturalism. In defiance of this, it was local Labour controlled councils, particularly in multiethnic urban areas, who took up multiculturalism as a guiding principle, developing in different ways municipal programs which attempted to meet the needs of minority communities. For instance, programs focused upon the incorporation of minority ethnic groups into local political processes, expansion of multicultural education, provision of language services targeted at specific minority groups, the opening of community centers, and the relaxation of planning regulations which allowed for the construction of greater numbers of places of worship.

The End of Multiculturalism?

There is growing recognition by academics and commentators that this phase of multiculturalism is now largely in retreat. In current debates around minority rights and national identity in several countries, multiculturalism has come under increasing attack from those speaking from a variety of political positions and academic viewpoints.

Those who saw multiculturalism as emerging out of an antiracist struggle; a fight for the recognition, human rights, and equal treatment of minority groups have grown disillusioned with a celebratory form of multiculturalism which seems to ignore its political roots. On the other hand and in line with a return to assimilation, multiculturalism has been accused of being responsible for the development of poor community relations. In the case of the UK, this has been seen in the aftermath of urban 'riots' in the north of England in 2001. Official responses to these 'riots' identified that policies of multiculturalism and self segregation led to the establishment of parallel lives and mistrust in these divided communities which culminated in this unrest. This has been identified as part of a shift in a government agenda and a growing discontent among other leading figures such as the head of the Commission for Racial Equality who have argued that a policy of multiculturalism leads to fragmentation rather than integration. Multiculturalism is now considered as the problem and not the solution. Such criticisms have also been countered by arguments that the sources of discontent in minority communities, have not emerged from official policies of accommodation, recognition, and provision but through the persistence of racism, inequality, and successive underinvestment in the UK's urban areas.

What is seen to guide this move away from multiculturalism and toward assimilation in the case of the UK is the idea that a stronger and more coherent national identity and sense of Britishness is needed to counter the differences seen to threaten the nation. This must be seen not only in the context of the disturbances mentioned above but also in the aftermath of 9/11 and 7/7 which entrenched this fear and suspicion of difference, particularly in religious terms, and has drastically altered the debate and the governments direction. The language of this cohesion discourse emphasizes the importance of common values, social order, solidarity, interaction, and a shared sense of belonging as key to the development of more harmonious inter ethnic relations. What alarms some scholars and activists is that the very basis of multiculturalism, as a platform for greater racial equality and social justice, is becoming undermined by increasing calls for restriction of migrants, refugees, and asylum seekers alongside increasing demands placed on minority ethnic groups to conform to some prescribed and narrow version of Britishness.

Multiculturalism and Identity

Challenges to the concept of multiculturalism have not only emerged from those who believe that multicultural societies have become too separated. Questions have also been raised by those who believe that institutional practices have actually reinforced the position of minority groups through the categorization of identities as 'other'. This can particularly be seen in the context of urban entrepreneurialism and the need for locales to attract capital through the promotion of attractive cosmopolitan assets. One example of this is the employment of multiculturalism as a place marketing tool and the commercialization of particular forms of difference. This is seen in the development of ethnic quarters such as Chinatowns in many Western cities, as well as the growth of multicultural festivals, officially endorsed practices which have imposed particular versions of ethnic identities that are attractive to the majority and which are economically viable. At the same time multiculturalism has come to stand for the exoticization of ethnic difference, while simultaneously neglecting the dominant position and formation of white identities as 'the norm'.

It is also contended that the definition of minority and migrant groups on the basis of strict racial, ethnic, and cultural terms within the discourse of multiculturalism has worked to draw boundaries around individuals, fixing identities as stable with essential and unchanging characteristics. Postmodern and anti essentialist theories stress that identities are mobile, multiple and contextual, and open to change, countering communitarian philosophies of multiculturalism in that they conceal the realities of hybrid identities and multiple cultural influences as well as the dynamic possibilities of an individual's identity. The dangers of fixing identities are that they become naturalized and associated with inherent biological and cultural traits, rather than seen as products of political struggles and contestations in specific geographic contexts. The meanings and significance of particular ethnic labels for instance vary throughout history, but also according to national context and from place to place. It is also argued that the multifaceted and multidimensional aspects of identity, cross cut by a multitude of social categorisations such as race, ethnicity, religion, social class, gender, age, and sexuality, mean that individuals cannot so easily be pigeon holed. Such cri tiques do not deny the importance of difference, but the idea that it can be so easily pinned down by highlighting how individuals are in states of liminality and ambiguity, on the threshold of various forms of identification which are constantly being redefined and renegotiated through their everyday lives.

While theoretical approaches, which aid an under standing of the constructed and nonessential character of racial and ethnic identities, are vital to an appreciation of the possibilities and dynamics of multiculturalism, there is a tendency to ignore the realities of lives which operate in relation to specific and identifiable racial and ethnic groups. For instance, the employment of universal and undifferentiated figures, such as the 'migrant' or 'nomad' in the celebratory language of postmodernism, neglects the structural inequalities and experiences of marginality which construct the lived realities of racialized indi viduals and which formed the basis for political multi culturalism in the first place.

What seems to be particularly valuable from such theorizations though is that the identities which consti tute multiculturalism as a social fact and as public policy should not be taken for granted, but recognized as at once imposed, performed, and actively negotiated. Indeed, what is clear is that definitions of multiculturalism and the way in which geographers and other social scientists have attempted to outline the boundaries of this slippery concept depend largely upon the manner in which the identities of individuals and racial, ethnic, and religious groups are conceptualized.

See also: Belonging; Citizenship; Difference/Politics of Difference; Ethnicity; Festival and Spectacle; Globalization, Economic; Hybridity; Identity Politics; Immigration I; Local-Global; Multicultural City; Nationalism; Race; Racism and Antiracism; Segregation.

Further Reading

Alexander, C. and Alleyne, B. (2002). Introduction: Framing difference: Racial and ethnic studies in twenty first century Britain. *Ethnic and Racial Studies* 25(4), 541 551.

Amin, A. (2002). Ethnicity and the multicultural city: Living with diversity. *Environment and Planning A* 34(6), 959 980.

Anderson, K. (1991). *Vancouver's Chinatown, Racial Discourse in Canada, 1875 1980*. Kingston: McGill Queen's University Press.

Back, L. (1996). *New Ethnicities and Urban Culture: Racisms and Multiculture in Young Lives*. London: UCL Press.

Back, L., Keith, M., Khan, A., Shukra, K. and Solomos, J. (2002). New labour's white heart: Politics, multiculturalism and the return of assimilation. *Political Quarterly* 73(4), 445 454.

Bennett, D. (ed.) (1998). *Multicultural States: Rethinking Difference and Identity*. London: Routledge.

Goldberg, D. (ed.) (1994). *Multiculturalism: A Critical Reader*. Oxford: Blackwell.

Jackson, P. and Penrose, J. (eds.) (1994). *Constructions of Race, Place, and Nation*. Minneapolis, MN: University of Minnesota Press.

Kymlicka, W. (1995). *Multicultural Citizenship*. Oxford: Oxford University Press.

MacLaughlin, J. (1998). Racism, ethnicity and multiculturalism in contemporary Europe: A review essay. *Political Geography* 17(8), 1013 1024.

Mitchell, K. (2004). Geographies of identity: Multiculturalism unplugged. *Progress in Human Geography* 28(5), 641 665.

Nash, C. (2003). Cultural geography: Anti racist geographies. *Progress in Human Geography* 27(5), 637 648.

Parekh, B. (2000). *The Future of Multi Ethnic Britain*. London: Profile.

Rogers, A. and Tillie, J. (eds.) (2001). *Multicultural Policies and Modes of Citizenship in European Cities*. Aldershot: Ashgate.

Sandercock, L. (1998). *Towards Cosmopolis*. London: Wiley.

Schuster, L. and Solomos, J. (2004). Race, immigration and asylum: New labour's agenda and its consequences. *Ethnicities* 4(2), 267 287.

Sundstrum, R. R. (2003). Race and place: Social space in the production of human kinds. *Philosophy and Geography* 6(1), 83 95.

Relevant Websites

http://www.irr.org.uk
Institute of Race Relations.

http://www.cre.gov.uk
The Commission for Racial Equality.

http://www.runnymedetrust.org
The Runnymede Trust, Intelligence for a Multi Ethnic Britian.

http://www.statistics.gov.uk
UK Statistics Authority.

http://www.unesco.org
United Nations Educational, Scientific and Cultural Organization: Social and Human Sciences, International Journal on Multiculture Societies.

Multidimensional Scaling

W. M. Bowen, Cleveland State University, Cleveland, OH, USA

Glossary

Attribute Space An objective space which contains a set of stimuli object and in which the dimensions are defined by the traits or characteristics of those stimuli (contrast with 'psychological space').

Binary Choice Judgment The product of a judgment task in which a respondent is asked to select and commit to one of two possible alternatives.

Dimension The degree of gradation or quantification along one of a set of mutually exclusive traits or characteristics.

Dissimilarity The psychological distance between two stimuli.

Ideal Point An ideal level of attributes for objects of choice around which a decision-maker's utilities decrease monotonically.

Individual Differences Models Psychologically oriented models that represent individuals in terms that explicitly recognize that each individual is in some respects like all other individuals, in some respects like some other individuals, and in some respects like no other individuals.

Judgment Data The products of a formally structured process in which respondents are asked to provide their judgments of a set of stimuli and in which the judgments are rendered formally so that they are in a form appropriate for communication, interpretation, and processing.

Liking Ratings The products of a judgment task in which respondents are asked to order a set of stimuli by providing answers to questions such as "On a scale of 1 to 5, how much do you like stimulus *x*?"

Multidimensional Scaling Any of a variety of methods for producing a geometrical representation of a data matrix.

Pairwise Comparison Judgments The products of a judgment task in which respondents are presented with all possible stimulus pairs and asked to indicate that member of the pair who in their judgment possesses more of the attribute being scaled.

Preference-Ratio Judgments The products of a judgment task in which respondents are presented with all possible stimulus pairs and asked to indicate the degree to which one of the stimuli is preferred to the other using ratio-scaled numerals.

Psychological Space A subjective space in which a set of points are identified that correspond to respondents' perceptions of or preferences for a set of stimuli. The dimensions of the psychological space may differ from those in the corresponding attribute space (see 'attribute space').

Unfolding Model A model in which a space is constructed to contain a set of points for a corresponding set of respondents and another set of points for the stimulus objects that they judge, all on the assumption that the respondent's preferences are 'folded' around the location of the respondent's own position (the ideal point) in the space.

Introduction

From walking across a room to driving a car to work to selecting a residential or industrial location, human spatial behavior is based upon hundreds of spatial choices. These choices invariably involve negotiating sets of stimuli that have both physical (objective) and perceived (subjective) attributes or dimensions. The stimuli are objective in the sense that they are presented to an individual's sensory organs as elements of his or her environment. The stimuli are subjective insofar as, in order to respond deliberately to their environments, individuals must perceive and in some degree discriminate between and interpret the stimuli they encounter in terms of their previously acquired spatial knowledge and understanding of the relevant environment. Successful spatial decision making does not necessarily require that the subjective dimensions completely and perfectly represent the objective ones. Nor does it require that individuals perceive and simultaneously consider all of the stimuli objectively present in their environments. The only necessity for individuals is to select and respond to a subset of stimuli that enables them to predict the outcomes of their choices, so that they can achieve their purposes. Geographical research that has as its aim the study of spatial behavior must thus take at least two sets of considerations into account. First, it must account for both the objective and subjective aspects of the sets of stimuli with respect to which individuals deliberate and respond in making their predictions and choices. Second, it must account for the possibility that individuals select and respond to subsets of stimuli from their environment, and that their views of their environment are not a complete and perfect representation of that environment. Multidimensional scaling aims to meet both of these requirements. In doing so, it enables powerful systematic analysis of the perceptions and preferences that underlie individuals' spatial choices and behaviors.

Multidimensional scaling (MDS) models comprise a subset of a general class of scaling models, all of which are oriented around the development of instruments for measuring or ordering stimulus objects with respect to their attributes or traits. The term 'multidimensional scaling' in particular refers to a family of data analytic methods used to geometrically represent dissimilarity or preference relations between elements of a set of objects or stimuli in low dimensional multidimensional space. The following sections describe the historical and conceptual foundations of MDS, give an overview of the MDS research process, and characterize four specific purposes for which MDS may be used in human geography. Examples of the wide range of application areas related to human geography include cognitive mapping, environmental planning, cartography, choice of transportation mode, migration, route learning and way finding, consumer behavior, site selection, and spatial decision making.

Historical and Conceptual Foundations of MDS

The historical and conceptual foundations of MDS are found in late nineteenth and early twentieth century empirical discrimination experiments in psychophysics and psychometrics. The researchers conducting those experiments were interested in characterizing empirical relationships between physical magnitudes such as the amount of light in a room or the weight of an object, and the corresponding values on related sensory scales. For instance, the researchers would ask the respondents in their experiments to judge the weight, pitch, color, or brightness of physical objects (generally referred to in the MDS literature as 'stimulus objects'). Then, they would make inferences about the relationship between the physical magnitudes and the corresponding judgments of the respondents.

The initial insight that led to the development of MDS models in particular occurred in 1938 when Richardson proposed an analogy between the psychological concept of dissimilarity and the geometric concept of distance. Richardson's idea was that the Euclidean distance axioms suggest intuitively plausible parallels between psychological dissimilarity and geometric distance. The parallels can be seen clearly by explicit consideration of these axioms. The axioms stipulate specifically that for a mathematical function of d defined over pairs of objects (i, j) to be a Euclidean distance function, it must satisfy the following four axioms:

$$d(i, i) = 0 \quad [1]$$

$$d(i, j) \geq 0 \quad [2]$$

$$d(i, j) = d(j, i) \quad [3]$$

$$d(i, j) + d(j, k) \geq d(i, k) \quad [4]$$

Accordingly, in terms of psychological dissimilarity, the first axiom implies that each stimulus object must be identical to itself. The second one implies that when two stimulus objects are not identical, their dissimilarity must be greater than zero. The third one implies that a stimulus object i must be exactly as dissimilar from object j as object j is from object i. All three of these are highly intuitive. While the fourth axiom is less intuitive than the others, this did not keep the overall analogy from catching on. The analogy between geometric space and psychological space became the impetus for all future MDS models.

The analogy between geometric space and psychological space suggested a distance model for dissimilarity in which a researcher might start with a respondent's numerical reports of his or her subjective judgments of dissimilarity between pairs in a set of stimulus objects. Then, by treating the judgments as distances and imposing a requirement of geometric coherency upon them, the researcher might work backward from the judgments to infer the geometric structure of the psychological space that must have been used by the respondent in generating them. According to this model, the full set of dissimilarity judgments between stimulus objects i and stimulus objects j (δ_{ij}) can be represented by the standard Euclidean distance function:

$$\delta_{ij} = d_{ij} = \left(\sum_{k=1}^{K} (x_{ik} - x_{jk})^2 \right)^{1/2} \quad [5]$$

Here, δ_{ij} is a datum (a reported numerical proximity judgment of the perceived dissimilarity between stimuli objects i and j), d_{ij} is the corresponding theoretical distance between x_{ik} and x_{jk} (x_{ik} and x_{jk} are the kth coordinates of the ith and jth objects or stimuli, respectively) in the psychological space used by the respondent in generating the judgments.

While the analogy between geometric space and psychological space established the concept of the distance model, it did not stipulate an algorithm with which to solve the implied operational problem of starting with a set of distances and deriving a set of coordinates from them. Torgerson, in 1952, first stipulated the solution to this problem. Torgerson recognized that if the distances are to meaningfully represent the psychological structure of the stimuli as they are configured in the mind(s) of the respondent(s), then the stimulus coordinates must be algebraically derived from the distances. Thus, the problem he faced was in a sense the opposite of the one commonly solved by grade school students. The challenge in the commonplace problem is to start with a given set of coordinates and, from them, to derive the distances between them. But in the distance model of dissimilarity, the distances are given, not the coordinates. The problem therefore is to start with these distances and, from nothing else but the distances and the rules of algebraic reasoning, to work backwards through the Euclidean distance formula to derive a corresponding set of stimulus coordinates that accounts for the distances. Torgerson was the first to show how to do this. Once he did so, it became possible to estimate the values of the theoretical quantities $x_{ik} - x_{jk}$ from the judgment data, and thus to meaningfully describe aspects of the psychological space from which the judgments must logically have originated.

Significant progress in MDS theory and practice followed over the next few decades. Advancements in mathematical psychology provided a compelling psychological and theoretical rationale for MDS methods. These include a theory of data in which data are viewed as relations between points in a space. They also include the 'unfolding' model of choice according to which there is an ideal level of attributes for objects of choice, and a decision maker's utilities decrease monotonically on both sides of this 'ideal point'. Researchers working directly on the analytical basis of MDS overcame several major restrictions on the mathematical assumptions about the judgment data implicit in Torgerson's algorithms. One such advancement, known as ordinal MDS, relaxed the assumption that the judgments are proportional to the distances. Another was a procedure which relaxed the assumption that the researcher has only one data matrix of judgments. This set up the possibility of doing MDS analyses of individual differences in cognition and perception. Ways to isolate and estimate individual difference weights were discovered, enabling operational characterization of variation in judged stimulus structure across individuals, occasions, settings, or treatment conditions. A probabilistic, multidimensional version of Coombs' unfolding model for the scaling of spatial data was stipulated and introduced into geography. Computer programs for implementing the newly developed algorithms were developed and disseminated.

Once MDS matured in the 1970s and 1980s, research aimed toward improving upon and enhancing MDS techniques has focused largely in two areas. First is the area of confirmatory modeling, in which the objective is to compare certain properties of an estimated configuration with those of an expected configuration. Second is probabilistic MDS in which the objective is to test hypotheses about how decision makers view stimuli and make decisions, and to realistically represent the richness of individuals' cognitive processes when presented with complex stimuli. For instance, probabilistic MDS can be used to model differential uncertainty about respondents' perceptions and preferences, thus enabling a researcher to avoid the error of confounding sensory and hedonic variability, on the one hand, with actual perceptions and preferences on the other.

Today, MDS is a sophisticated set of scaling methods, most of which use dissimilarity judgments to investigate respondent perceptions of stimuli. Alternatively, some MDS analyses are based upon liking ratings, preference ratios, and binary choice judgments and are oriented toward investigating respondent evaluations of stimuli. In either case, technically nuanced MDS models and techniques are available. Many current MDS analyses represent the stimulus structure of judgments obtained from several individuals in a stimulus space shared by the group. Alternatively, many estimate individual difference parameters to characterize variation in judged stimulus structure across individuals. Most MDS analyses are deterministic in the sense that every configuration of points in the solution space is uniquely determined by the parameters in the mathematical MDS model and the judgments. But a growing number are probabilistic in the sense that they assume that randomness is present in the judgments. Rather than stipulating a configuration of stimulus points described by unique values, probabilistic MDS analyses describe the stimulus locations using probability distributions. Numerous texts and book chapters describing and explaining the rationale and details of the wide array of technical varieties, applications, and software packages for MDS are readily available.

Overview of the MDS Research Process

The MDS research process can be divided into a number of phases, the first of which is the researcher's formulation of the research problem. At the outset of the process, the geographical concepts and problems to be investigated are specified. Next, the researcher collects judgment data from the set of respondents and uses a computer program to analyze the data. Finally, the results are interpreted and reported. **Figure 1** depicts an overview of the MDS research process.

The researcher's objective during the problem formulation phase is to identify the particular set of stimuli of interest. The idea is to identify those stimuli that comprise the relevant aspects of the perceptual/preference space in the mind of the respondent(s). This is

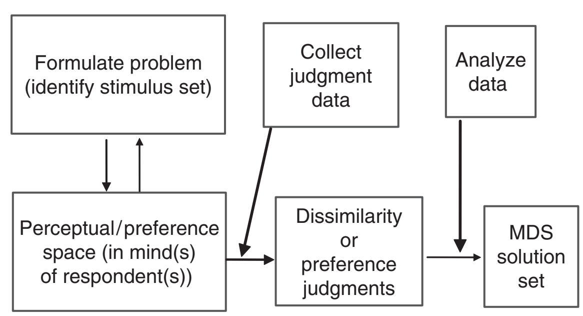

Figure 1 The MDS research process.

dictated by the purpose of the research. If the research problem is about, say, how people select a residential location, then the stimuli might be pictures and descriptions of a set of eight to ten residences. If the research problem has to do with peoples' perceptions of various neighborhoods, then the stimuli might reflect salient aspects of the relevant neighborhoods. In any case, the stimuli are selected by the researcher to represent those objective aspects of the relevant environment about which the subjective construction of the stimuli on behalf of the respondents is of interest in terms of the research problem. Most often, these are the stimuli that form the sensory basis for the particular behavioral responses of interest in the research problem. Ideally, the stimuli reflect all of the important aspects of the concept under study as accurately as possible.

Once the stimulus set has been identified, the researcher is ready to collect judgment data from respondents. During this phase, the respondents are asked a series of questions. Their answers comprise the data. Specifically, for each stimulus pair they are asked to judge dissimilarity or preference on a numerical scale (e.g., a seven point Likert scale from very similar to very dissimilar). When the stimuli set contains n stimuli, this produces a matrix that contains $(n^* (n-1))/2$ numerical pairwise comparison judgments.

In the next phase, the data are analyzed using an MDS computer program. The main objective of the computer program is to construct a model of the data designed to illustrate the relationships between judgments. The data are judgment data, specifically measures of dissimilarity or preference for the pairs (i, j) of n stimuli. The model illustrates the relationships between these judgments by geometrically representing them in geometric space. Because the judgments reflect the respondents' views of the stimuli, the relationships between judgments are implicitly also the relationships between stimuli.

During the analysis phase, the researcher faces a number of choices which, when taken together, collectively determine the particular approach to the analysis. Choices that the researcher faces include those between metric MDS, which deals with interval or ratio judgments, or nonmetric MDS, which deals with ordinal judgments. Choices must also be made between dissimilarity, liking rating, or preference analysis, group or individual analysis. Several different algorithms are feasible for several of the particular approaches. Dissimilarity judgments reflect the degree of dissimilarity between stimuli. Liking rating judgments are a species of dissimilarity judgment in which the distance in the MDS solution space is between the stimulus object and an ideal point. Preference ratio and binary choice judgments reflect the respondent's preference for one stimulus in relation to another.

Choices also must be made in the analysis phase between exploratory or confirmatory analysis, and deterministic or probabilistic analysis. Exploratory MDS investigates the psychological structures formed by individuals and groups in relation to the stimuli without any prior expectations about the properties of the MDS solution. Confirmatory MDS is used when more is known substantively about the stimuli or the relations between stimuli, so that the researcher can meaningfully impose additional restrictions or constraints upon the MDS model. With deterministic MDS, the analysis is restricted formally to a description of the geometric configuration of the stimuli. In probabilistic MDS the analysis becomes not only descriptive but also inferential in the sense that it allows hypothesis testing about the properties of the solution (e.g., the number of dimensions in the solution space). In each case, the appropriate alternative depends upon the purpose of the research study.

MDS programs create structural graphics of the stimuli. **Figure 2** illustrates one such graphic, produced by the ALSCAL algorithm in SPSS. The graphic plots each stimulus, in this case each neighborhood, in a two dimensional space. In this case, one dimension represents population, and the other percent in professional occupations, both of which can be measured objectively. The dimensions could just as easily have been perceptual, such as neighborhood status and ease of access. In any case, the proximity of stimuli to each other along any given dimension indicates either how similar they are or how preferred they are in relation to each other along that dimension. In this case, it can be seen that Newbury, Trumbull, Concord, and Clarks are all similar along dimension 1, with Geneva being the least similar to them. Along dimension 2, Geneva is quite similar to Newbury, Trumbull, and Concord, and least similar to Clarks and Perry. Accordingly, by examining the structural graphic, insight may be gained into the similarities and differences among stimuli. In many applications, the dimensions must be labeled by the researcher, a task which requires subjective judgment on behalf of the researcher and is often very challenging. The results must then be interpreted in terms of the research question.

Purposes for MDS in Human Geography

One purpose for MDS in human geography is as an exploratory method for visualization of geographic data. The idea is for the analyst to look at the configuration of stimulus points to find rules that enable description of the distributions of the stimuli in the attribute space. This is of especial interest in situations in which data are not linked to an explicit theory with which to predict the magnitudes of the stimuli or patterns within them. It may also be of value when the purpose of the investigation includes documenting knowledge organization, assessing the compatibility of knowledge organization, or providing metrics for decision making.

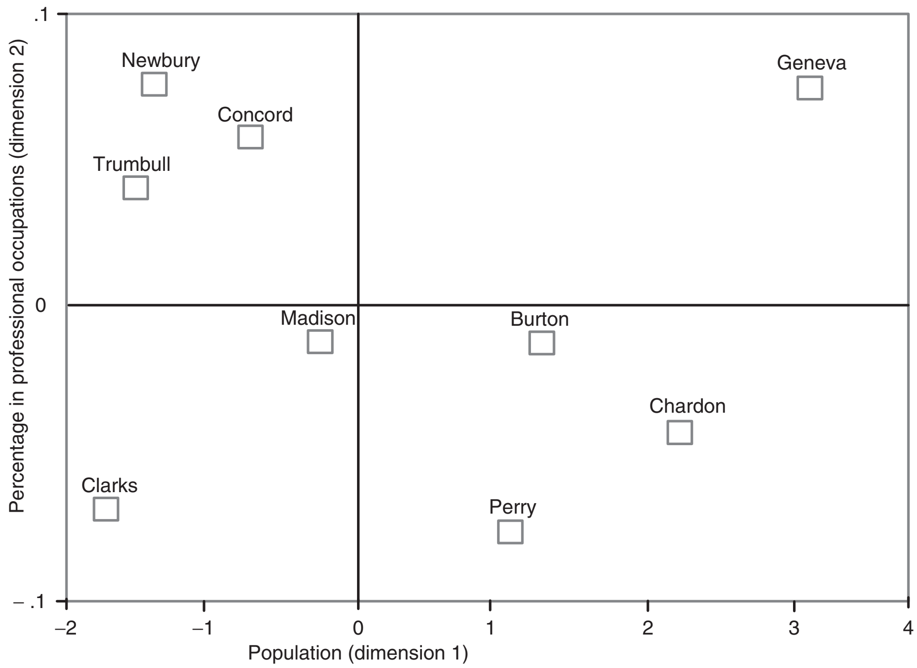

Figure 2 Two-dimensional MDS configuration for a set of imaginary neighborhood.

Another purpose for which MDS may be used is as an analytic technique for testing hypotheses about the structure of the relations between objects or stimuli. MDS is useful in this respect when theory predicts characteristics of the configuration of stimulus points in MDS space. For instance, a researcher investigating the acquisition of cognitive maps might hypothesize that geographical features which impede movement will lead to increased judged distance estimates between places on different sides of the barrier. Hills and trees might be expected to lead to overestimated distances, as might the numbers of turns or intersections along a route. Conversely, regularity in the layout of roads and places in flat areas devoid of features might lead to underestimated distances. In this situation, the effects of these barriers on the judged distance estimates could possibly be found in the configuration of stimulus points in the MDS solution. All else equal, those places on different sides of the barrier in geographical space would show up further apart in the MDS solution space, and those on the same side of the barrier would show up closer together. Thus, MDS can serve as an analytic technique with which to test hypotheses about the dimensions, factors, or attributes according to which the stimuli are conceptually distinguished in the theory.

A third purpose for MDS is exploring and documenting the psychological structures formed by individuals and groups in relation to various stimuli within their environments. Researchers in the fields of marketing and consumer psychology use MDS for this purpose when they investigate consumer perceptions of the similarity between various products or brands, as well as in much of their market positioning, new product planning, pricing, and advertising research. Researchers in human geography could use MDS for this purpose in a wide range of applications including, but by no means limited to, research into the geography of personality traits, learning and acquisition of cognitive maps, individual differences in residential area evaluation, the perceptions, preferences, and attitudes of residents in neighborhoods in which an urban freeway is proposed, spatial shopping choices and behaviors, perceptions of regional economic and environmental milieu, and the use of geographical metaphors in information archiving and access. Recently developed probabilistic MDS algorithms are especially promising in this regard. Their ability to factor differential uncertainty about respondents' perceptions and preferences into the MDS solution makes a high degree of realism possible. Also by enabling hypothesis tests of aspects of the relevant psychological structures, they allow researchers to challenge their own assumptions about such factors as the degree of psychological similarity between individuals, the degree of differential familiarity with stimuli among subsets of respondents, the number of dimensions used by respondents in the psychological processes of generating dissimilarity or preference judgments, or the presence of dominant attributes in forming judgments, among others.

A fourth purpose for MDS in human geography is in application as a multi attribute model of spatial decision making. Specifically, MDS can be used in conjunction with the unfolding model to operationalize a conception of spatial decision making in which alternatives that are closer to an individual or group's ideal point are preferred to those that are further away. Accordingly, researchers investigating the perceptions of preferences of respondents making spatial choices can use MDS in a relatively passive way as a measuring device and analytical tool. They might, for instance, use it to describe the psychological structure of the perceptions and preferences of an individual or group of people who made a choice to select, say, a particular site for a retail outlet or hazardous waste facility, or who selected a particular route for navigating through a neighborhood. Alternatively, decision makers in the process of making spatial choices can use MDS more actively as a part of a decision support system designed to help implement and manage the choice process in the face of multiple and/or complex alternatives, multiple criteria, uncertainty, and more information than can be systematically processed by the unaided human brain. In this regard, MDS serves not only as a convenient technical tool for augmenting the unaided human cognitive capacity for storing and processing information about the decision situation. It also offers a more general hypothesis about the rationale underlying spatial decision making processes. Examples of human geographical research areas in which MDS can be used as a model of spatial decision making include, but are not limited to site selection, route selection, spatial cognition, migration, and spatial learning.

Overall, there are far more potential applications of MDS in human geography than is commonly recognized. Partially, this is attributable to the fact that spatial behavior and spatial choice is only one out of a fairly large number of alternative research areas in human geography. Partially, it is attributable to the fact that a relatively small group of human geographers base much of their research upon recognition that the dimensions of human geographical systems themselves are often intellectual constructs that can take a multiplicity of forms. In this respect, the potential of MDS in human geography will be realized only once human geographers more widely recognize and take interest in the influences on spatial behavior attributable to the human mind's attribution of dimension to human geographical systems. Partially it is also attributable to the fact that MDS models and methods are not widely studied, understood, and utilized within human geography. In this respect, the potential for MDS to enable greater effectiveness at describing, understanding, and comprehending the processes of perception, thought, and choice that underlie spatial behavior will be realized only once the constituent models and methods are more widely studied and understood within the field.

See also: Maps; Quantitative Methodologies; Scale; Spatial Science.

Further Reading

Borg, I. and Groenen, P. J. F. (2005). *Modern Multidimensional Scaling: Theory and Applications* (2nd edn.). New York: Springer.
Coombs, C. H. (1964). *A Theory of Data*. New York: John Wiley and Sons.
Davison, M. (1983). *Multidimensional Scaling*. New York: Wiley.
Gatrell, A. C. (1983). *Distance and Space: A Geographical Perspective*. New York: Oxford University Press.
Golledge, R. G. and Stimson, R. J. (1997). *Spatial Behavior: A Geographic Perspective*. New York: Guilford Press.
Haynes, K. E., Bowen, W. M., Arieira, C. R., Burhans, S., Salem, P. L. and Shafie, H. (2000). Intelligent transportation systems benefit priorities: An application to the Woodrow Wilson Bridge. *Journal of Transport Geography* 8, 129 139.
Lloyd, R. (1976). Cognition, preference, and behavior in space. *Economic Geography* 52(3), 241 253.
MacKay, D. (1983). Alternative probabilistic scaling models for spatial data. *Geographical Analysis* 15, 173 186.
MacKay, D. and Zinnes, J. L. (1981). Probabilistic scaling of spatial distance judgments. *Geographical Analysis* 13, 21 37.
MacKay, D. and Zinnes, J. L. (1988). Probabilistic multidimensional scaling of spatial preferences. In Golledge, R. G. & Timmermans, H. (eds.) *Behavioural Modelling in Geography and Planning*. New York: Croom Helm.
Richardson, M. W. (1938). Multidimensional psychophysics. *Psychological Bulletin* 35, 659 660.
Rushton, G. (1969). Analysis of spatial behavior by revealed space preferences. *Annals of the Association of American Geographers* 59, 391 400.
Torgerson, W. S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika* 17, 401 419.
Zeleny, M. (1982). *Multiple Criteria Decision Making*. New York: McGraw Hill.
Zinnes, J. L. and Griggs, R. A. (1973). Probabilistic, multidimensional unfolding analysis. *Psychometrika* 39(3), 327 350.

Relevant Websites

http://www.analytictech.com/borgatti/mds.htm
Multidimensional Scaling.
http://www.proscal.com
Proscal tool for analysing consumer choices.

N

Nation

J. Penrose, University of Edinburgh, Edinburgh, UK

Glossary

Category A group of things, people, or actions that share one or more common attributes; a mechanism for ordering and representing the world.

Nation The largest self-defined collectivity of people whose members *believe* that they share some form of territorial association and that they are genealogically related.

Nation-State A geopolitical entity whose legitimacy – claims to sovereignty and political power – is based on its capacity to represent a distinctive nation.

Nationalism The ideology which argues that the boundaries of a state should be congruent with those of a nation.

Power The ability to control one's self, other people, nonhumans, space, and events.

Primordial Existing at the beginning of time and space.

Romanticist An intellectual movement that originated in late eighteenth-century Europe which was characterized by a rejection of Enlightenment ideas and norms, especially the primacy accorded to reason.

Sovereignty Supreme power which is usually considered to be absolute and unlimited, and usually exercised within a territory.

Territoriality The attempt to influence or control people, nonhumans, phenomena, and relationships by delimiting and asserting control over a geographic area called a territory.

Introduction

Nation is one of the most used and least understood words in the English language. Ironically, a good part of the power and resilience of what are called nations stems from the ambiguity associated with the concept of nation over time and in both academic and everyday realms. Today, nation is commonly used to mean things as diverse as a people, a state, citizens of a state, and a country – sometimes within a single newspaper or journal article. This diversity of meaning has itself become an ideological tool because it allows nation to be deployed to serve a wide range of ideological agendas and, perhaps paradoxically, to accommodate either shifts in government policy or the absence of policy altogether. For example, multiculturalism legislation frequently oscillates between viewing the nation as a homogeneous entity and a plural one. Similarly, the nation is portrayed as something that is defined by cultural group membership in some contexts and by citizenship in others. This lack of clarity, nestled comfortably in semantic confusion, can be expedient in the short term for politicians who want to avoid identifying (or alienating) their support base and making clear cut policy commitments. However, because nations are also used to justify the current distribution of global geopolitical power, and this in a world where their relevance is increasingly suspect, it is important to be clearer about what this term means (or at least, how it is used).

To advance this quest for greater clarity, this article outlines three key theoretical conceptions of nation and their relationship to geography and geographical thought. This is followed by a discussion of the political and personal significance of nations which demonstrates how nations have become linked with power on a variety of scales. These associations with power are then related to efforts to construct and maintain the nation to reinforce the world order based on nation states. In the final section, limitations of the nation and challenges to it are explored with an eye to assessing its future relevance.

Theoretical Conceptions of Nation

The term 'nation' has been in common use in English since the latter part of the thirteenth century when it was used primarily to connote broad divisions of humanity as defined by notions of race. By the late eighteenth century, the Enlightenment along with revolutions in America and especially France had combined to produce a new political context where popular sovereignty was introduced as a means of addressing problems ascribed to old, self serving, dynastic, and religious empires. The ideas that government should be by and for the people – instead of an elite few – represented a monumental shift in political thought and the application of these ideas demanded clarification of just what was meant by 'the people'. It is this process of clarification that vaulted the concept of nation into its current position of prodigious power; a position that is reinforced by its deeply embedded ambiguity.

In general, three prevailing usages of the term 'nation' have emerged from the Enlightenment context of the eighteenth century. In each case, the understanding of what a nation is, is often linked to ideas about when, and much less often where, nations first appeared. The first conception of nation is often referred to as primordialism and was originally associated with the Romanticists who viewed nations as very similar to volk groups; each of these relatively small groups constituted 'a people' and they were seen as the fundamental, natural, units of humanity. According to scholars like Vico (1668–1744), Fichte (1762–1814), and, especially, Herder (1744–1803), each nation would, if given the chance, devise its own means of governing itself in ways which would ensure the survival and development of the group as a distinctive entity.

Although social construction theory has discredited the idea that nations are primordial or natural, it has also revealed how and why so many people continue to see them as such (see below). Today, the understanding of nation as a fundamentally cultural entity is sometimes reflected in use of the terms 'ethnic nation' or 'cultural nation'. Such nations are usually conceptualized as ancient (Walker Connor is an important exception here), and they are commonly defined as involuntary associations that can be identified by objective attributes such as language, religion, descent, and sometimes 'race'. All of these attributes, but especially those related to kinship, are seen to produce – and be maintained by – deep emotional and psychological bonds. Their grounding in essentialized notions of culture means that these nations are also seen as respectful of cultural differences but this can be manifested in exclusivity and a regressive preoccupation with the past and artificially fixed notions of tradition. Here, it is very important to distinguish between those who believe that nations defined by these qualities actually exist as empirical realities (often proponents of particular nationalist movements), and those who see cultural nations as constructed entities that rely on individual identification with the nation, and shared imaginings of it, for its survival. Until quite recently, it was not uncommon to link cultural nations with the geographical regions of Eastern and Central Europe but this view has been challenged by current scholarship.

The second conception of nation is usually termed modernist and it uses the term to describe 'the people' (instead of 'a people'), usually as the inhabitants and/or citizens of a state. This perspective is commonly seen to originate with thinkers like Rousseau (1712–78), whose preoccupation with developing a new form of state still required that he define the people who would be the bearers of his newfound notion of popular sovereignty. For Rousseau, and for his intellectual heirs, the focus is on membership in a political unit rather than a cultural one and, gradually, this understanding has come to find expression in the term 'civic nation'.

One of the key distinctions between the modernist and the primordialist perspectives on the nation is the dating of their origins. Not surprisingly, all modernists are united in the belief that nations are exclusively modern phenomena. For scholars like Ernest Gellner and John Breuilly, nations represent a functional response to the new requirements of industrialization and capitalism – requirements that included a populace (and labor pool) that was relatively unified by a shared mass culture and a common language. These nations are conceptualized as voluntary ones that are based on civic ties and shared political principles. As this suggests, they are promoted as inclusive entities that seek to transcend cultural differences and thus, so the logic goes, they are rational and progressive groups that are modern and forward looking. The geographical association of these so called civic nations has normally been confined to Western Europe and, more recently some of its white settler colonies like Canada. Importantly, these geographical associations, like the distinction between ethnic nations and their civic counterparts, have been increasingly challenged by recent scholarship.

To some extent, the third and most recent theoretical perspective on the concept of nation is a product of unease with the growing polarization between primordialist and modernist approaches. The term 'ethno symbolist' has been coined to denote scholars who reject the primordialist view that nations are essential and continuous, as well as the modernist view that they are nothing more than modern inventions. Although the term ethno symbolist rests uneasily with its key proponents, people like John Hutchinson, John Armstrong, and, especially, Anthony D. Smith, it does reflect their preoccupation with the symbolic legacy that premodern ethnic identities – often of very long standing – hold for today's nations. In general, ethno symbolists view nations as social and historical phenomena (rather than natural or 'given'), which are defined, not so much by real or even imagined ties of kinship as by lines of cultural affinity that become embodied in symbols, myths, memories, and values that are distinctive to a particular cultural group and that unite them in a sense of shared destiny. They acknowledge the power of the modern era to transform premodern ethnics (ethnic communities) into nations, but insist that nations cannot be understood without reference to their living ethno symbolic legacy.

Each of these three approaches is quite clear about what a nation is and about when nations emerged but, from a geographical perspective, the absence of any sustained consideration of the where of nations, of their spatial and territorial qualities, is conspicuous. This lacuna is probably a product of the tendency for debates about the concept of nation to be dominated by sociologists, political scientists and, increasingly, anthropologists. However, a closer look at each of the three key perspectives reveals more that is of geographical relevance than the simple tendency to associate different kinds of nations with particular geographical regions. For primordialists, nations are products of the merging of a people and the territory that they occupy (or that they once occupied and continue to claim). Here it is the emotive power of a group's attachment to a particular piece of land that is seen to be the primary factor in the formation of a territory and in strategies to preserve it. In contrast, modernists believe that the geographical borders of a state delineate the territory of a nation. Here it is the material resources of a territory, including the symbolic significance of controlling it, that has primary influence in setting the boundaries of the state and, hence, the nation. For ethno symbolists, territory achieves its significance as a symbol of cultural continuity over time and space. Its value is primarily in what it signifies and communicates rather than in any objective reality.

The first two approaches perpetuate traditional practices of political geography by treating boundaries as naturalized lines or demarcations of the limits of sovereignty. Although they differ about how the boundaries of the nation are constructed and about which kind of boundaries (cultural or political ones), should define the borders of nation states, neither perspective actively questions the unit of the nation state *per se*. Thus, both conceptions of nation become complicit in reinforcing what John Agnew has called the territorial trap. In other words, they perpetuate the assumption that nation states are fixed units of sovereign space; that they serve as containers of society; and that there is a clear and unproblematic division between the domestic realm within the state and the foreign realm outside of it. The territorial trap stifles more flexible and liberating conceptions of the nation because it reinforces the assumption of an inalienable link between a single nation and a single territory. This view disregards the fact that the social and political processes which influence the formation and development of nations, operate on a variety of scales; it ignores the growing prevalence of multiple and transnational national identities; and it precludes the sharing of space as a solution to geopolitical conflict.

Ethno symbolism seems to offer potential for acknowledging the dynamic relations between the social world and space but this has yet to be realized fully. By viewing territory as a symbol of cultural continuity it is possible to imagine alternative relationships between people and places, ones that are not premised upon exclusive and inalienable rights to set pieces of land. For example, it becomes possible to imagine different nations sharing territory without either territory or nations losing their individual significance. Having said this, the ethno symbolist approach, like the other two, still has a largely unacknowledged tendency to reify the anthropological notion of culture as something that is ontologically real and, though complex, not problematic in its own right. As long as this is the case, notions of cultural distinctiveness will continue to inform constructions of both nations and the nation states which they legitimize. However, even if new conceptions of culture are devised, the geographical processes of defining boundaries and effecting exclusions will always remain central to the material practices, ideologies, and discourses through which social groups called nations (however defined), and the identities that they support, are constituted.

The Significance of Nations

Nations acquired their global political significance through the spread of the ideology of nationalism and the associated rise of nation states. Nationalism, as an ideology, argues that the boundaries of the nation should be coterminous with those of the state. The application of this ideology produces nation states.

Depending on one's perspective, congruence between the boundaries of the nation and the state is desirable either because it is the best way of ensuring the survival of nations or because it is the best way of improving the efficacy of states. Where the former view held sway, it was the (putative) boundaries of nations that set the boundaries of the nation state. Sometimes the application of this process led to the merging of territories that were inhabited by groups sharing a common language and/or culture to form a single nation state (e.g., Germany and Italy). In other cases, larger (often imperial) territories that were home to numerous cultural groups, were divided into a number of smaller polities, each representing or seeking to represent a single nation (e.g., Estonia, Bulgaria, and Slovenia). In contrast, where the primary drive for nation state formation was the creation of a political context in which new political principles like popular sovereignty could be applied, the (putative) boundaries of the state defined the borders of the nation state. In these cases, the nation tended to be defined as the citizenry of the state and its function was to ensure the continuity and legitimacy of that state.

Regardless of how the nation was conceptualized and how it was integrated into the process of nation state formation the general concept attained huge political significance because it legitimized nation states. The success of the earliest nation states, along with their growing links with democracy, inspired emulation on an ever wider scale. Today, all existing states call themselves nation states and are generally recognized as such by others. It is currently impossible to lay a claim to a state without grounding this claim in the demonstrable capacity to represent a nation. Although processes of globalization have challenged the sovereignty of nation states through the formation of suprastate organizations, membership in most of these entities remains confined to nation states. Similarly, pressures which threaten to fragment nation states from within usually mount their challenges to established nation states by applying the same nationalist ideology that legitimizes the existing state. Thus, nations retain profound political significance because they constitute the sole basis for claiming formal political power and for legitimizing the nation states that form the cornerstone of the current world order.

The power that nations acquire through their global geopolitical significance is paralleled by their very personal significance. Indeed, many scholars argue that without the constant identification of individuals with their own particular nation these entities would cease to exist. Variations on this idea can be found in Renan's famous proclamation that the nation is a daily plebiscite; in Benedict Anderson's description of nations as imagined communities; and in Walker Connor's conviction that nations exist when people believe that they do. The idea that the essence of a nation is the belief in it, rather than any objective characteristics of it, has been a very important one in understanding nations and their ability to inspire and sustain nationalist movements and/or nation states.

One of the ways in which people come to believe in nations is through processes that internalize them; as the individual becomes part of the nation, the nation becomes part of the individual. When this happens, the survival of the self is seen to be contingent upon the survival of the nation and thus people will go to incredible lengths to defend their nation. All processes that strengthen personal identification with a particular nation also contribute to the reification of the general concept of nation and its validity as a means of ordering the world. Thus, there is an important, though largely invisible, symbiosis between nation as a category of personal identification and as a category of global political power. Both work together to represent the world as coherently structured, and to reinforce the centrality of the concept of nation in maintaining this structure. Nations give people a means of locating themselves in the world and of distinguishing themselves from others. In the process, they fulfill the less salubrious function of effecting inclusions and exclusions, empowerment and disempowerment: they become a convenient and convincing way of deciding who gets what, where, and of legitimating any resulting inequalities.

Building Nations

The fact that nations are usually perceived as natural, or at least as profoundly judicious divisions of the earth's people and territory, has been instrumental in entrenching their positions of power in the modern world. In reality, however, this perception of nations, and the nations themselves, are both more accurately viewed as what Michael Billig describes as 'invented permanencies'. They are products of elaborate and constant processes of social construction. To some extent this is necessitated by the fact none of the world's roughly 200 bounded nation states can claim that they comprise a single culture. Thus, in order to fulfill their function of legitimizing a given state, national unity (if not uniformity) must be pursued. Importantly, the nations that have come to be associated with any given state have been imagined very seldom as a composite of the state's constituent cultures; instead the unifying nation is usually a replica of the most dominant group (or nation). Ironically, members of other nations within the state have to endure the inequalities associated with downplaying their alternative group membership and assimilating with the dominant one if they hope to achieve anything like equal access to resources. This is one more manifestation of the complex power relations that actually underlie the seemingly unproblematic category of nation.

In most cases, nation state formation began by creating a national scale; something that is now largely taken for granted but which, arguably, did not exist prior to the ascendance of nation states. Nation states transformed medieval Europe's imprecise political boundaries into very exact ones and they did much the same thing with systems of communication. It was only in the age of nationalism, with its emphasis on bounding lands and peoples, that there was any reason for standardizing spelling and grammar – for creating languages – or for extending these conventions across the entire territory of the state. The application of nationalist ideology naturalized the notion of discrete languages even as it invented them and these new national tongues were deployed as both evidence that a distinctive nation existed and as a means of constructing one. Actively encouraging all of the population of a nation state to speak in the same way reinforced the emerging sense of a national scale and was instrumental in all other nation building practices.

One other very early process in creating a national scale involved reinforcing claims of sovereignty by establishing a strong, centralized power structure, usually in the form of something called a national government. The creation of a political system that was dominated by national parties was normally paralleled by the introduction of an electoral system which also worked to reinforce the national government as the locus of political activity and power. Ultimately, the entrenchment of a centralized power structure also gave governments the institutional basis for producing legislation that could, in turn, protect the authority of the state and advance efforts to create a uniform nation within it.

Nation building practices involve both encouraging uniformity in the population of the state and inspiring personal identification with the new nation state and, consequently, loyalty to it. For example, a national army brings together conscripts from different regions, encourages them to bond with one another through shared training experiences, and indoctrinates them with an overriding allegiance to their common state which they share a duty to defend. Similarly, systems of compulsory education create a forum for standardizing language, for emphasizing national interpretations of the past, present, and future, and thereby socializing children across any given country into imagining themselves as members of the same nation. The development of national media is also a fundamental element of most nation building agendas. Through print, radio, television, and now the internet, media expose the population of the entire state to the same national events and they disseminate national perspectives on international developments. In Canada, the evening news program presented by the publicly owned Canadian Broadcasting Corporation is actually called 'The National'. Finally, the creation of national symbols is an effective mechanism for linking other nation building strategies together while also heightening their visibility and significance. For example, national anthems and flags tap into and mobilize emotive responses to the nation as a source of both community and personal identity and they figure prominently in the activities of the military and schools. Each time that a national anthem is sung or a national flag is flown, the existence of the corresponding nation is reasserted and personal connections and loyalties to it are both aroused and validated. Individuals perform the nation and it is reified.

As Billig has noted, flags, anthems, and other symbols of nations are frequently banal but, if anything, this increases their potency in effecting the ideological work of nationalism. In combination, practices of nation building reinforce belief in the existence of nations and of a national scale, and they also shore up the illusion of precise boundaries between us and them, domestic and foreign. The banality of much nationalism is paralleled by the constancy of nation building activities – they occur all the time and in an unimaginable array of forms and contexts. This is partly because the imagining and performance of nations is a requirement of their continued existence but also because persistent internal divisions must be smoothed back into a veneer of uniformity.

In addition, states must work hard to maintain imagined permanencies in ever changing contexts. For example, nation states frequently become entrapped in the contradictions between their economic and cultural goals. For many, the immigration of 'foreigners' is essential to continued economic expansion and, consequently, to the state's fulfillment of its obligation to pursue the best interests of the nation that it represents. Yet, in fulfilling this obligation, the concomitant need for cultural uniformity is undermined. Attempts to reconcile these contradictory demands frequently involve immigration policies that privilege those migrants who are deemed most likely to assimilate with the national culture. In contrast, those who are considered just too different are, at best, granted temporary visas so that their labor and creativity can be exploited even as their membership in the country that these contributions support is denied. Where these selections and exclusions were once framed in explicitly racist terms, they are now couched in terms of the need to protect the nation (usually undefined) from inappropriate cultural practices, and so called bogus asylum seekers. In some cases, citizenship tests are being introduced to exact at least the illusion of commitment to the dominant nation, and in others, different categories of citizenship have been devised in order to confer differential rights within a single nation state. Importantly, the granting of citizenship does not preclude ostracization from membership in the nation. In all of these ways, individual nation states work hard to maintain the illusion of a homogeneous nation that legitimizes both that particular state and the global system based around nation states.

The Future of the Nation

Despite their resilience to date, there are several reasons for some proactive reflection on the future of nation states. First, both the ideal concept of nation and its empirical manifestations are demonstrably unsound and this means that nations constitute a very dubious basis for distributing global geopolitical power. This assessment is based on the fact that neither human beings nor the territories that they occupy are easily divided into nations – either by nature or by human intervention. For a growing number of people, affinity with more than one nation is the norm and many areas of land are claimed with equal legitimacy by more than one group. Moreover, attempts to impose a coincidence between the borders of states and nations are the single most important source of conflict both within and between nation states.

Second, global acceptance of nationalist ideology, despite the manifest impossibility of effecting it in practice, means that states expend huge energies and resources on shoring up the nation in order to retain the legitimacy of the state. Thus, even as more and more states become characterized by increased pluralism their abilities to respond openly to this changing reality remain constrained by the logic of nationalism. These restrictions are reflected in things like the contortions associated with immigration policy, as described above, and also in attempts to simply redefine nation as a civic entity. This strategy is designed to circumvent the growing incongruity between nationalist ideology and the nation states it legitimizes – and this is more than the usual problem of the discrepancy between ideal types and their empirical expressions. Taking recourse in the notion that nations can be defined by political principles and shared political culture has three very important consequences. First, the privileged position of the dominant culture group is overlooked; second, the huge power that has long been associated with cultural conceptions of nation is denied; and third, the need to address directly the limitations of a world based on nationalism is avoided.

A third reason for thinking about the future of nations (both conceptually and empirically) is that, the assertion that nations exist in any concrete sense (including their imagining as such) contributes to the reification of the view that the world is ordered coherently by categories. This perpetuates processes of categorization – of drawing precise boundaries – that are deployed to effect the unequal distribution of material and symbolic resources on scales that range from the global to the personal. A focus on the national context also makes it possible to ignore those sociospatial relations that transcend geographical scales and that demand reconceptualization of the world in ways that are not contingent on ostensibly fixed categories. As a consequence, alternative and potentially liberating ways of thinking about the relationships between people, and between people and place, remain overlooked as well. Geography has an important role to play in moving the world beyond the limitations of this category based perspective and this can begin with a concerted critique of how the nation is used, both conceptually and empirically, to regulate power and identity.

See also: Citizenship; Colonialism, Internal; Cultural Geography; Cultural Politics; Deconstruction; Difference/Politics of Difference; Economies, Borderland; Ethnic Conflict; Ethnicity; Fatherland/Homeland; Geopolitics; Globalization and Transnational Corporations; Hegemony; Identity Politics; Ideology; Immigration I; Immigration II; Land Rights; Migration; Migrant Workers; Modernity; Multiculturalism; National Parks; National Schools of Geography; National Spatialities; Nationalism, Historical Geography of; Nationalism; Other/Otherness; Place; Place, Politics of; Political Boundaries; Race; Racism and Antiracism; Refugees and Displacement; Representation, Politics of; Self-Other; Social Movements; Sovereignty; Space I; State; Symbolism, Iconography; Systems Theory; Territory and Territoriality; Transnational Ethnic Networks; Transnationalism; Transnationality; Urban Planning and Human Geography; World-System.

Further Reading

Agnew, J. and Corbridge, S. (1995). *Mastering Space: Hegemony, Territory and International Political Economy*. London/New York: Routledge.

Anderson, B. (1991). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso.

Anderson, J. (1986). Nationalism and geography. In Anderson, J. (ed.) *The Rise of the Modern State*, pp 115 142. Brighton: Harvester Press.

Billig, M. (1995). *Banal Nationalism*. London: Sage.

Breuilly, J. (1993). *Nationalism and the State* (2nd edn.). Manchester: Manchester University Press.

Connor, W. (1994). *Ethnonationalism. The Quest for Understanding*. Princeton: Princeton University Press.

Gellner, E. (1983). *Nations and Nationalism*. Oxford: Basil Blackwell.

Johnson, N. C. (2002). The renaissance of nationalism. In Johnston, R. J., Taylor, P. J. & Watts, M. J. (eds.) *Geographies of Global Change* (2nd edn.), pp 130 142. Oxford: Blackwell.

Lowenthal, D. (1994). European and English landscapes as national symbols. In Hooson, D. (ed.) *Geography and National Identity*, pp 15 38. Oxford: Blackwell.

Marden, P. (1997). Geographies of dissent: Globalization, identity and the nation. *Political Geography* 16(1), 37 64.

Ozkirimli, U. (2005). *Contemporary Debates on Nationalism: A Critical Engagement*. London: Palgrave Macmillan.

Penrose, J. (1994). 'Mons pays ce n'est pas un pays' full stop. The concept of nation as a challenge to the nationalist aspirations of the Parti Québécois. *Political Geography* 13(2), 195 203.

Penrose, J. (2002). Nations, states and homelands: Territory and territoriality in nationalist thought. *Nations and Nationalism* 8(3), 277 297.

Smith, A. D. (1986). *The Ethnic Origins of Nations*. London: Penguin.

Smith, A. D. (1995). *Nations and Nationalism in a Global Era*. Cambridge: Polity Press.

Storey, D. (2001). *Territory: The Claiming of Space*. Harlow: Prentice Hall/Pearson Education.

Relevant Websites

http://www.lse.ac.uk
Association for the Study of Ethnicity and Nationalism.

http://www.nationalismproject.org
The Nationalism Project.

http://www.learner.org
The Power of Place: Geography for the 21st Century.

National Parks

N. Curry, Countryside and Community Research Institute, Cheltenham, UK

Definitions

National parks (subsequently, Parks) have been defined by the World Conservation Union (International Union for the Conservation of Nature and Natural Resources, IUCN) as one of seven types of protected area requiring a specific kind of management (see **Table 1**). This clas sification was introduced in 1992 (involving scientists from 181 countries) as a rationalization of some 140 terms found to be used globally for protected areas. Parks (IUCN Category II) are to be managed mainly for eco system protection and recreation. They are natural areas of land or sea (or both), designated to:

(1) protect ecological integrity of ecosystems for present and future generations;
(2) exclude exploitation or occupation inimical to the purposes of designation of the area; and
(3) provide a foundation for spiritual, scientific, edu cational, recreational, and visitor opportunities, all of which must be environmentally and culturally compatible.

Global definitions such as this are bound to lead to ex ceptions. Some of the more salient are considered here. Some Parks are 'not national'. The Gaza – Kruger – Gonarezhou area between Mozambique, South Africa, and Zimbabwe, for example, in crossing international borders, is termed by the IUCN a 'trans boundary' or 'trans frontier' Park. These are important for their inte grated management systems in relation to wildlife, but they require international protocols, can lead to national disagreement, and threaten international security. The IUCN also recognizes Parks for Peace, defined as: "trans boundary protected areas that are formally dedicated to the protection and maintenance of biological diversity and of natural and associated cultural resources, and to the promotion of peace and cooperation" (IUCN, 1994).

Some Parks fall into 'the wrong IUCN classification'. The English and Welsh Parks, for example, are classified as IUCN Category V protected areas (protected land scape/seascape) because they are managed mainly for conservation and recreation and do not have a dominant ecological or ecosystem component. To complicate matters further, some Parks 'fall into more than one protected area category'. The largest national park in Poland, Biebrza, is also designated as a wetland of global significance under the Ramsar Convention. The Tatra Park (one of nine in Slovakia) is also part of a UNESCO Biospheric Reserve. The Royal Chitwan National Park in Nepal too is a World Heritage Site.

Conversely, areas 'not designated as Parks are ac corded Parks status'. In the USA, National Historic Re serves (such as Ebley's Landing), National Recreation Areas (Santa Monica Mountains), and National Sea shores (Point Reyes and Canaveral) are accorded Park status. Whilst the US National Park Service (founded in 1916) manages 390 sites, only 58 of them actually carry the designation by name. In Turkey too, the 33 named Parks are supplemented by Natural Parks, Nature Re serves, and Mature Reserves that are considered in the same way in planning and management terms.

Some Parks 'prohibit or discourage recreation'. In the Sundarbans National Park, spanning both India and Bangladesh, the unique mangrove vegetation and the viable Bengal tiger population mean that visitors are not allowed in the Park. The Vanoise National Park in the Alps, the first French Park, was actually set up in 1963 to protect the area 'against' tourism development. It now has a central zone where recreation is discouraged and a peripheral zone twice the size, used to deflect visitor pressure from the former. Other Parks have been con sidered, effectively, to be 'nonexistent'. A proportion of the world's Parks exist in name only as there has been no active management of them at all in pursuit of Park ob jectives. Known as 'paper parks', these are under constant threat of damage and destruction.

The world over, therefore, what constitutes a Park is far from uniform (despite the IUCN definition): not all

Table 1 IUCN protected area categories

Category Ia	Strict Nature Reserve: protected area managed mainly for science
Category Ib	Wilderness Area protected area managed mainly for wilderness protection
Category II	National Park: protected area managed mainly for ecosystem protection and recreation
Category III	Natural Monument: protected area managed mainly for conservation of specific natural features
Category IV	Habitat/Species Management Area: protected area managed mainly for conservation through management intervention
Category V	Protected Landscape/Seascape: protected area managed mainly for landscape/seascape conservation and recreation
Category VI	Managed Resource Protected Area: protected area managed mainly for the sustainable use of natural ecosystems

Source: IUCN (1994). *Guidelines for Protected Areas Management Categories,* 261pp. Cambridge, UK and Gland, Switzerland: IUCN.

areas that possess Park characteristics are called Parks and some areas called Parks do not possess Park character istics. If there is any commonality across Parks in practice, it relates to conserving special places for human benefit.

Nature and Extent

The IUCN estimates that there are more than 44 000 of their protected areas, worldwide, covering some 10.1% of the world's terrestrial surface. In 2003, the best estimate of the number of Parks within this number, given the problems of definition discussed above, is 3881, covering nearly 4.5 million km^2, distributed regionally as in **Table 2** below. Over time, both the number and area of Parks is becoming a smaller proportion of all IUCN protected area classifications. Parks (IUCN protected area Category II) represented only 3.8% of the 'number' of protected areas in the IUCN seven point protected area classification in 2003, but 23.6% of their 'area', as they have been commonly established to protect larger areas at the ecosystem and landscape level.

Despite their generally large size, only two of the largest 20 IUCN protected areas are Parks: Northeast Greenland (established in 1974), which is the largest protected area of all at 972 000 km^2 and the Tassili N'Ajier Park in Algeria, the nineteenth largest protected area at 72 000 km^2. The most common on the 'big list' are IUCN Category VI protected areas (see **Table 1**), Managed Resource Protected Areas, where unmodified natural systems predominate (e.g., the Great Barrier Reef Marine Park in Australia and the Venezuelan Alto Ori noco Casiquiare Biosphere Reserve).

As an example of the way in which the geography of national parks is constantly changing, three new Parks were created in 2005 in the far east of Russia, to provide refuge for the Amur tiger. This was part of a program to introduce 30 000 km^2 of protected areas in the region of Krasnoyarsk, itself only part of a system of 21 new federal protected areas, pledged by the Russian government for the first decade of the millennium.

History

Early Thinkers

The notion of the 'national park' developed in a number of countries separately during the nineteenth century, but it was essentially a Western phenomenon with a diverse range of underlying values concerned with esthetics, culture, amenity, access, and science. Consistent with the views of Byron and Coleridge, Wordsworth, in his *Guide to the Lakes of 1810*, described the English Lake District as a "national property … for every man (sic) … who has an eye to perceive and a heart to enjoy." He was later to be concerned that the lakes would be overrun when the train came to Windermere in 1844. In a crowded country, the 'freedom to roam' value was the driver for the first legal proposal for Parks in England, with Bryce's 1884 Access to Mountains Bill. Although the Bill failed, it marked the start of an active campaign for Parks for recreation purposes. By 1936, the Standing Committee on National Parks had been formed as a voluntary body, to argue, in government, the case for Parks.

The painter George Catlin, in his travels though the American West, wrote in 1832 that the Native Americans

Table 2 National parks: their significance in different world regions

Region	*% of the land area of the region that is covered by an IUCN protected area*[a]	*National parks as a % of the number all IUCN protected areas in the region*	*% of all IUCN protected areas in the region that are national parks, by area*
Antarctic	0.5	2.4	0.2
Australia and New Zealand	14.8	7.3	24.8
Caribbean	29.6	11.5	39.0
Central America	27.9	11.5	19.7
East Asia	8.8	4.7	10.4
Eastern & Southern Africa	17.2	4.5	25.7
Europe	14.6	0.6	13.1
North Africa & Middle East	9.9	6.4	17.0
North America	20.8	10.2	36.7
North Eurasia	8.2	1.1	19.8
Pacific	3.7	6.2	27.0
South America & Brazil	22.2	14.3	17.5
South Asia	6.9	9.4	22.3
South east Asia	16.4	9.6	28.1
West and Central Africa	8.7	3.6	31.8

[a] These figures do not include marine protected areas because they are not part of the land surface.

Source: IUCN/UNEP (2003). *2003 United Nations List of Protected Areas*, pp 37 44. Cambridge, UK and Gland, Switzerland: IUCN; Cambridge, UK: UNEP World Conservation Monitoring Centre.

in the United States might be preserved "by some great protecting policy of government … in a magnificent park … a nation's park, containing man and beast, in all the wild and freshness of their nature's beauty!" Culture, amenity, and ecosystems were to be the originating val ues in this 'uncrowded' land. Similar ideas for preserving nature were expressed in Sweden by Baron Adolf Erik Nordenskiöld in 1880 and the Scottish American natur alist, John Muir, was also to prove inspirational in de veloping the Parks on both sides of the Atlantic, being an early advocate of conservation, environmentalism, and the animal rights movement.

New Parks for Diverse Reasons

Yosemite was designated in 1864 (advocated by Olmstead and signed through an Act of American Congress by President Lincoln), principally "for the benefit and en joyment of the people." Yellowstone followed in 1872. Because of its remoteness, its management had become the responsibility of the American Federal Government (rather than an individual state) by 1890. It had the support of conservationists and politicians, but also business. The Northern Pacific Railroad was keen to exploit the tourism potential of the Park. In the same year, Sequoia Park was established as a wilderness sanc tuary to protect groves of giant sequoia trees that were being destroyed by logging. The Everglades, in contrast, were designated for their abundance and variety of wildlife, rather than for scenic or historic values.

The 'Western' motivations of public enjoyment pre vailed in other early Parks. The Australian Royal Na tional Park was designated south of Sydney in 1879, laid to parkland, recreating the soft English countryside: the 'home' counties for many a new in migrant. Its purpose was recreation: the 'lungs' of Sydney. The Banff National Park, too, was introduced in Canada in 1885 built around the tourism potential of the hot springs rather than, at that time, its conservation value.

Public access motivations were combined with wil derness and cultural reasons for designation in New Zealand. Tongariro was designated in 1887, driven by its important Maori cultural and spiritual associations (and gifted by Maori to the nation) as well as its volcanic features. More recently, Whanganu River and Paparora National Parks, designated in the 1980s, and the transi tion of Northwest Nelson Forest Park into the Kahur angu National Park in 1996, brought the current land based Parks in New Zealand to 13. Ten are in moun tainous regions not considered suitable for settlement; this facilitates their maintenance and management as wilderness areas.

Other designations were made before World War I in both the Northern and Southern hemispheres, but it has been suggested that many of these were more to do with national prestige, or 'monumentalism' on the part of governments, than with any deep desire to conserve nature.

The 'wilderness' and 'enjoyment' motivations were echoed in Europe. The first nine of Europe's 359 Parks were set up as a system in Sweden in 1909 embracing mountain lands, virgin forest, deciduous forest, swamps, archipelago, and old agricultural landscapes. In England and Wales, the National Parks and Access to the Coun tryside Act of 1949 led the first Parks to be designated in 1951 (Peak District, Lake District, Snowdonia, and Dartmoor), very much in pursuit of access and amenity. Recent British additions in 2002 (Loch Lomond and the Trossachs), 2003 (Cairngorms), and 2005 (the New Forest) bring the total in Britain to 14.

Preserving game reserves was a driver in South Africa, with the introduction of the first Park in 1926 between the Sabie and Olifants rivers. Today, South African National Parks is responsible for nearly 400 000 km^2 of protected land in 20 Parks, with purposes now embracing tourism development consonant with conservation objectives. **Figure 1** below illustrates the growth in Park designations from these early developments. **Table 3** provides ex amples of the diversity of principal motivations for des ignating Parks at other times. These suggest that the origins of national parks do not reflect the IUCN's attempt to provide a global definition of national parks.

Organization and Planning

The World Conservation Union, more properly known as the IUCN, was set up in 1948 and imposes some global norms on the development of Parks through its 140 member countries. Membership is patchy, however, and is particularly low in the Pacific region. Its Policy,

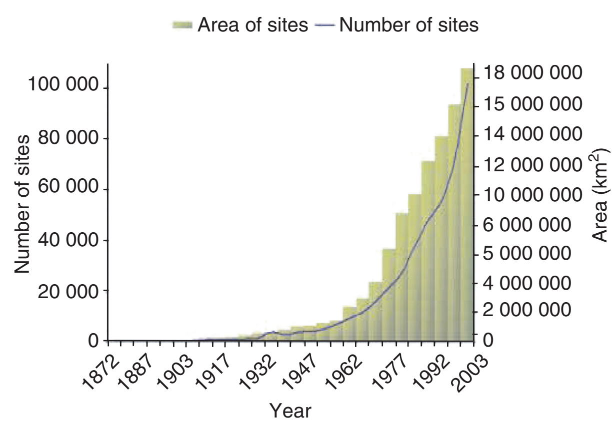

Figure 1 Cumulative growth in protected areas by 5-year increment, 1872 2003. Reproduced from PAGE 26, IUCN/UNEP (2003). *2003 United Nations List of Protected Areas*. Cambridge, UK and Gland, Switzerland; Cambridge, UK: IUCN; UNEP World Conservation Monitoring Centre, with permission from IUCN.

Table 3 Examples of principal motivations for designating Parks, by timeline

Country	*Timeline*	*Original motivation*
Canada	First Park in 1885, 42 today	To protect ecosystems and manage them for visitors to appreciate, understand and enjoy.
Uganda	First Park in 1938	To strengthen the conservation status of the ecosystem.
Finland	First 3 Parks in 1956, 35 today	To safeguard diversity of nature and sites of Finnish history.
Turkey	First Park in 1958, 35 today	To protect archaeological and historical features.
Costa Rica	First Parks Santa Rosa and the Poas Volcano, in 1971	To control deforestation.
Alaska	First 10 Parks in 1980, 17 today	To protect representative natural, cultural and historic features.
Scotland	First Park in 2003, 3 today	To protect culture, heritage, beauty and amenity, open air recreation and sustain local communities.

Source: Author, culled from national web sites.

Biodiversity and International Agreements Unit was es tablished in 2002 to orchestrate policy delivery for Parks. The United Nations Environment Programme World Conservation Monitoring Centre (UNEP WCMC) also advocates biodiversity through policy, data analysis, and support.

The IUCN holds a World Park Congress every 10 years, to bring together protected area specialists, man agers, and experts. It acts as the focus for determining new action and policy. The inaugural World Park Con gress, the "First World Conference on National Parks" was held in Seattle from 30 June to 7 July 1962. The purpose of this conference was to establish a more effective international understanding of Parks and to encourage further development of the Park movement on a worldwide scale. Issues discussed included the effects of humans on wildlife, species extinction, the religious significance and esthetic meaning of certain parks and wilderness, international supervision of boundary parks, the economic benefits of tourism, the role of national parks in scientific studies, and the practical problems of park management.

The fifth IUCN World Parks Congress (WPC) was held in Durban, South Africa in September 2003. With close to 3000 participants from 160 countries, it was the largest and most diverse gathering of protected area ex perts in history. The challenge before the 2003 Congress was to demonstrate how protected areas are relevant to the broader economic, social, and environmental agenda for humankind in the twenty first century. The congress called for stronger and more innovative management for biodiversity and sustainable development, with a renewed emphasis on locality and indigenous peoples. The involvement of young people in park protection was to be developed alongside a stronger financial basis for management.

Regional Action

Those signing up to the IUCN Parks processes are un evenly distributed throughout the world, seemingly based more on historical accident than the values of particular resources. But this imbalance is gradually being redressed. The Pacific Island states, for example, produced National Biodiversity Strategies and Action Plans (NBSAPs) as required under the Convention on Biological Diversity. These define the biodiversity needs of each country and identify a number of priority areas where IUCN expertise could assist. In particular, a transfer of expertise is encouraged from Australia and New Zealand, whom the IUCN considers to be world leaders in best practice for environmental management. Both countries also have active environmental non governmental organization (NGO) communities that provide strong advocacy and a generally high level of expertise.

In contrast to the Pacific, Europe in general has a well developed system of Parks, and a range of European legislation and multilateral agreements for their pro tection. The effectiveness of this protection is not uni form, however, and the main thrust of the IUCN European Programme (2005–08) is to move toward im proving this protection in the weakest areas.

Support Groups and Networks

In addition to global bodies and national agencies with a responsibility for the administration and management of Parks, there is a range of networks and foundations that provide support. Globally, for example, the networks of Marine Protected Areas and World Heritage Sites (sponsored by UNESCO) and the World Network of Biosphere Reserves both are active in orchestrating planning and management expertise and experience.

A number of countries have their own Parks 'foun dations', independent of government, to ensure the pur suit of Park objectives. Thus, the New Zealand National Parks and Conservation Foundation seeks to broaden the base of stakeholders in Park management and the American National Parks Foundation creates partner ships and has a fundraising role. The Australian Foun dation raises funds for specific Park projects and the African Parks Foundation (established in Holland) seeks to match effective management with sensitive economic development.

In the UK too, the Council for National Parks (and its Scottish equivalent) seeks to protect, enhance, and pro mote the understanding of the nation's Parks and acts as an umbrella for over 40 amenity groups. The Association of National Parks is a federation of the 14 UK Parks, working collectively. At the European level, the PAN Parks Foundation brings all the protected area stake holders together for information sharing and good prac tice. EUROPARC, the umbrella organization for all authorities who manage Europe's protected areas (na tional parks, regional parks, nature parks, and biosphere reserves) in 38 countries, has the common aim of pro tecting Europe's unique variety of wildlife, habitats, and landscapes.

Landownership

The IUCN suggests that the ownership and management of Parks should be with the "highest competent authority of the nation". Despite this implied 'public authority' preference for ownership, patterns are diverse. There is much private ownership. Small farmers own most of the Narew Park and over half of the Biefrza Park in Poland, and in England and Wales, private ownership dominates but is not exclusive, as **Figure 2** indicates. Leasing from the state is a dominant tenure in the Biscayne Park in Florida for privatively owned settlements used for weekend recreation. In the Richtersveld Park in South Africa, all of the land is owned by people living within the Park, and is leased 'to' South African National Parks with considerable economic benefit.

In the Asian regions, much Parkland is in communal ownership amongst kinship groups and clans, who have a historic right to the resources on that land. This is a common traditional form of ownership that still prevails in many Park areas. In some Parks the ownership is de termined according to land use. Thus in the Cat Tien National Park in Vietnam, agricultural land is under private ownership but forest land remains with the state, and ownership changes with change in use. In the Ben ta'al National Park, established in the 1940s as Lebanon's first Park, a cooperation was formed amongst landowners, villagers, and government over the mutual ownership of the Park. But in many areas, such as in the Pacific, it is not clear who actually owns Parkland, as much of it is not titled at all.

The dynamics of ownership are equally complicated. Many African Parks were established under colonial re gimes, where the forced removal of local communities was common practice. This has led to the more recent reassertion of land claims in Parks such as Augrabies Falls, Kruger, Vaalbos, and Dongolo, similar to the res titution of some Parks in New Zealand, to Maori, under the Treaty of Watangai. Alongside forced removal, the state has actively purchased land for Park purposes, as in the Pieniny National Park in Slovakia, and independent charities have also purchased land and then ceded it to the state as has occurred in Costa Rica.

In fact, the ownership of 'rights' in Parks is as critical as the ownership of the land itself. The state can curtail rights without actually owning the land. In African Parks, for example, state control over hunting and resource exploitation shapes the appearance and condition of Parks, often with problematic economic consequences as in the Kisite Marine National Park in Kenya. In England, state control over development is stringent. Dispossession and restitution have also occurred in relation to rights as well as to land.

In the Parks of Australia, for example, Aboriginal people have been dispossessed of customary rights (hunting, fishing, the taking of traditional foods, etc.) and in a number of cases in land restitution, many of the rights over land remain with the state as, for example, in the

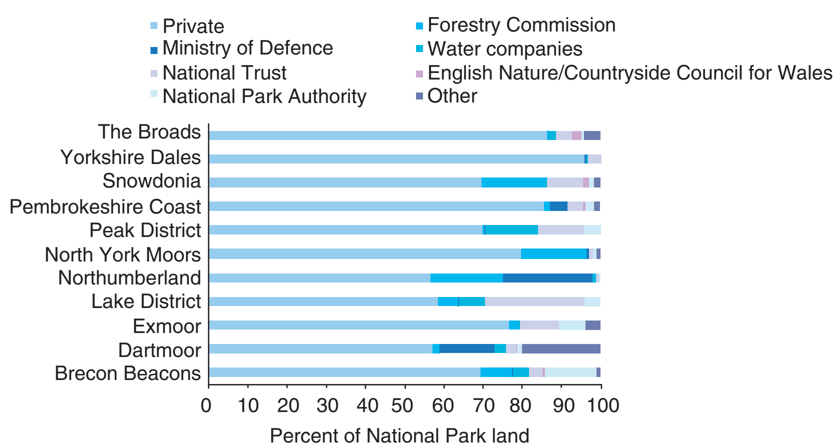

Figure 2 Landownership in the national parks of England and Wales.

2006 repatriation of Biamanga and Gulaga, two coastal Parks. In some of the trans boundary Parks in southern Africa, for example, repatriated land may not be used for farming or mining or contain permanent settlements. It may be used for religious and cultural purposes, however. Rights are also bought and sold. In Costa Rica and many European countries, state payments are made for en vironmental services from private landholders.

Management Issues

At the global level, the largest tension in the management of Parks is that between conservation and development. Because three quarters of the IUCN's protected areas are in the developing world, they represent a particular commitment for governments who face challenges re lating to developmental issues and priorities. In the re gion bounded by Sudan, Eritrea, Djibouti, Ethiopia, Kenya, Uganda, and Somalia, for example, an Inter governmental Authority for Development has constantly to balance the needs of a poor population with the needs of conservation.

In the Pacific priority region, the challenge is to get protected areas in general onto the political agenda. Political instability in the region has weakened the ability to balance development demands and the conservation of resources and ecosystems. Many Pacific countries do not have the resources to implement international conser vation agreements, and aid funding does not make this a priority over economic development and health. The Pacific Islands Round Table on nature conservation, formed in 1998, is the only body in the region with a specific conservation remit.

Sometimes, local opposition to protected areas can contribute to their lack of management. In other cases governments themselves can ignore protected area status when they contain valuable resources. In yet others, such as in the Zapata National Park in Cuba, economic live lihoods (hunting, tree felling, and agricultural intensifi cation) can damage the ecological value of Parks.

One of the more successful approaches to amelior ating this tension between conservation and development has been to place emphasis on 'community based ap proaches' in pursuit of the sustainable husbandry of natural resources. At Mt. Elgon, Uganda, for example, there has been a shift from traditional protectionist ap proaches (which had not been effective) to sustainable development ones through collaborative schemes, in volving local people in resource management and con servation. This has been successful at stimulating cultural and economic values in tandem.

Khunjerab National Park in Pakistan is one of the highest Parks in the world and has developed community based conservation to increase the population of snow leopards and Marco Polo sheep. In Kisite Marine National Park and Mpunguti Marine Nature Reserve in Kenya, the balance between resource exploitation (particularly fishing and tourism), social infrastructure (particularly health and education), and conservation is being constantly adjusted in response to community requirements.

Similar schemes can be found in the Langtang Na tional Park in Nepal where traditional land use practices are sustained within the local community as a means of preserving the natural environment. Community man agement is also active in the Bundala and Udawalawe National Parks in Lanka. The Darwin Initiative (funded by the British Government), too, has led to the estab lishment of botanical gardens in Tam Do National Park in Vietnam, through a community involvement scheme.

In a number of cases, community involvement has been combined with processes of visitor control. In some cases, this has involved using entry fee charging to keep visitor numbers down. In Kenya, Park entry fees and day trip charges have been used since 1975 both to control visitor use but also to raise revenues for a number of community projects. Entry fees are payable at all Parks in Nepal including the Mount Sagarmatha (Everest) Na tional Park to control visitor use and to generate local revenues. In further deference to community values, in some Nepalese parks, things such as smoking, chewing tobacco, and consuming alcohol are not allowed, for re ligious reasons. Visitor numbers are also rationed in other ways. Special permits are required, for example, in Yala (Ruhuna) National Park in Sri Lanka for this purpose.

In contrast, some parks have a tradition of more in tensive recreation activity of benefit to the local com munity. The Tatra National Park in Slovakia, for example, has a range of preexisting spa centers, including sanatoria, health centers, cure houses, and spa hotels. Here, a commercial return is a central feature of the Park, as a means of stimulating the vibrancy of the local community.

Other significant management issues include the de velopment of sustainable ecosystems that, in some way, can also manage other resources to good effect. In the Cat Tien National Park in Vietnam, for example, man agement regimes are designed both to improve eco systems and also to have an active role in the control of flooding and the protection of water resources.

Finally, bio security and disease control are critical issues in some Parks, particularly where, as in Africa, there are large migratory animals. Bio security can also be threatened by unplanned colonization and oil exploration as in the Yasuni National Park in Ecuador. There are also specific ethnobotany programs in place in the Ayubia National Park, for example, in the Himalayas to protect biodiversity in the Himalayan moist temperate forest.

See also: Ecology; Ecotourism; Nation.

Further Reading

Bishop, K., Green, M. and Phillips, A. (1998). *Models of National Parks*, Scottish Natural Heritage, Review Number 105. SNH, Inverness, Scotland.

Blunden, J. and Curry, N. R. (1990). *A People's Charter? 40 Years of the 1949 National Parks and Access to the Countryside Act*. London: Her Majesty's Stationery Office.

Bruner, A. G., Gullison, R. E., Rice, R. E. and da Fonseca, G. A. B. (2001). Effectiveness of parks in protecting tropical biodiversity. *Science* 291(5501), 125 128.

Eagles, P. J. F. and Mc Cool, S. F. (2002). *Tourism in National Parks and Protected Areas: Planning and Management*. Oxford: CABI Publishing.

Ghimire, K. B. and Pimbert, M. P. (eds.) (1997). *Social Change and Conservation: Environmental Politics and Impacts of National Parks and Protected Areas*. London: United Nations Research Institute for Social Development, Earthscan (ISBN:1853834106).

Hardin, G. (1968). The tragedy of the commons. *Science* 162(3859), 1243 1248.

IUCN (1994). *Guidelines for Protected Areas Management Categories*, 261pp. Cambridge, UK and Gland, Switzerland: IUCN.

IUCN/UNEP (2003). *2003 United Nations List of Protected Areas*. Cambridge, UK and Gland, Switzerland: IUCN. Cambridge, UK: UNEP World Conservation Monitoring Centre.

MacEwen, A. and MacEwen, M. (1982). *The Resource Management Series Number 5: National Parks: Conservation of Cosmetics?* London: Allen and Unwin (ISBN 0 04 719003 5).

Runte, A. E. (1997). *National Parks: The American Experience* (3rd edn.). Lincoln, NE: University of Nebraska Press (ISBN:0803289634.)

Wells, M., Brandon, K. and Hannah, L. J. (1992). *People and Parks: Linking Protected Area Management with Local Communities*. Washington, DC: World Bank (ISBN 08 213 2053X).

West, P. C. and Brechin, S. R. (1991). *Resident Peoples and National Parks: Social Dilemmas and Strategies in International Conservation*. Tucson, AZ: University of Arizona Press (ISBN 08 165 11284).

Relevant Websites

http://www.iucn.org
General information about the IUCN can be found here.

http://www.nationalparks.gov.uk
Parks in England and Wales, National Prks.

http://www.env.go.jp
Parks in Japan, Ministry of Environment, Government of Japan.

http://www.caingram.info
Parks in Turkey, Alan Ingram.

http://www.unep wcmc.org
The World Database on Protected Areas can be found here, WCPA.

http://www.nps.gov
U.S. National Park Service, International Programs Quarterly Bulletin, October December 2006, Office of International Affairs, NPS (An international bulletin for Parks containing current issues can be found here).

National Schools of Geography

P. Claval, Université de Paris-Sorbonne, Paris, France

A few years ago, Derek Gregory summed up the history of modern geography in the following way:

> In the eighteenth century, geography is supposedly a 'European science'; in the nineteenth century distinctive schools of geography are identified with particular national traditions ('the French school' and the 'German school'); and in the twentieth century particular places are often used as markers of the intellectual landscape ('the Chicago school', 'the Berkeley school', and 'the Los Angeles school'). But the global circulation of information [...] has made many of these parochialisms unusually problematic. No doubt we have retained some of them [...], but there is nevertheless an important sense in which our 'local knowledge' simply isn't local any more. (Gregory, 1994: 8 9).

Gregory's view of the evolution of our discipline was oversimplified, but he was right when opposing a time when geography was a European common endeavor, a time when national schools were active, and a time when parochialisms were progressively rubbed out. Such an evolution raised important problems: why did the geographic tradition become fragmented into national schools sometime in the nineteenth century? How did these national schools evolve in the twentieth century? Why did they recede during the last generation?

Other questions may be asked concerning the national schools: Did all the nations have a 'national school'? Were they all built along the same principles? What was the role of national intellectual traditions and institutional context in their genesis? What was the effect of the political history of nations and their periods of war, dictatures, and foreign occupation? In the way geography was conceived, what favored fragmentation?

Geography before Its Fragmentation into National Schools

Until the mid nineteenth century, geography developed in a context utterly different from the contemporary one. From the Renaissance, the main task of the discipline had been to build an increasingly accurate view of the world. As reminded by Anne Godlewska, geography was mainly molded by this cartographic imperative. During the sixteenth and seventeenth centuries, it was impossible to get direct and accurate measures of longitudes. As a result, geographers kept busy in analyzing travel books in order to evaluate distances and indirectly longitudes. They spent most of their time in libraries and ignored fieldwork.

The improvement of astronomical methods of longitude determination and John Harrison's invention of the marine chronometer caused a deep crisis in geography: new aims had to be developed. Because of the new facilities offered for mapping and the possibility to have natural scientists aboard, Western states developed maritime explorations: the travels of Cook in Britain, Bougainville in France, and Behring in Russia were the most famous. These national expeditions were conceived along parallel lines: there were close relations between the academies that were in charge of their scientific programs. Out of the description of the Earth's surface provided by the natural scientists involved in these explorations, a new form of geography developed, best exemplified at the beginning of the nineteenth century by Alexander von Humboldt. The geography he practiced knew no boundaries: a Prussian subject, trained in different German universities, he explored Spanish America and settled in Paris for 25 years for editing the material he and Aimé Bonpland had gathered. He was instrumental in the creation of a new form of scientific institution, which played an important role in the development of explorations and the expansion of European nations during the nineteenth century: the society of geography.

The popularity of geography during the eighteenth and early nineteenth centuries did not come from its content. It resulted from the comparisons between the landscapes, mores, institutions, or economic activities it allowed: hence, the strong interest in geography of many philosophers, Montesquieu, Rousseau, or Turgot in France, and Kant in Germany. Another role for geography was also explored at that time: for Herder, progress was a reality, but the paths followed by the different peoples were not similar; every one had to invent his or her own way through their ability in mobilizing the environment in which it lived. The national dimension of geography was in this way stressed, but Herder's ideas did not prevent the discipline to preserve its 'European' character until mid nineteenth century, as shown in the work of his followers, Carl Ritter, for instance.

The Emergence of National Schools: The Institutional Context

National schools of geography appeared in the second half of the nineteenth century – two of them, at least, the

German one from the 1860s or 1870s and the French one in the 1880s and 1890s. There were several reasons for such an evolution.

The early rise of a German school of geography was the result of the leading role of two geographers in the building of the modern form of the discipline: Alexander von Humboldt and Carl Ritter had created an intellectual tradition, which was missing elsewhere. The founders of the new discipline in Western countries were often former students of Ritter or Humboldt in the University of Berlin: Louis Agassiz and Arnold Guyot for the United States, or Elisée Reclus for France. Trained as a historian, Paul Vidal de la Blache became a geographer when discovering the volume of Ritter's *Allegemein vergleichende Geographie* on Asia Minor.

German geographers were well ahead of the others, but by 1850, there was still nothing as a German school. The creation of national schools resulted from the new forms of institutionalization undergone by the discipline at the time when nation states were becoming the main form of political organization.

Some of the components of the new institutional setting predated mid nineteenth century. Societies of geography had appeared in the 1820s, but their number mainly jumped after 1870. Their recruitment changed: instead of grouping mainly explorers, missionaries, topographical officers of armies and navies, and independent scholars, they attracted a growing number of businessmen, interested in the economic opening of the world and the expansion of European direct or indirect rule over 'backward' countries. The transformation of the economic conception of international life gave in this way a new significance to the modern nation states of the West – but the lines developed by the societies of geography were too similar to induce a fragmentation of geographical knowledge into national schools.

What was much more important was the transformation of geography into an academic discipline. Here again, Germany played a decisive role, thanks to the reform of Prussian Universities conceived by Wilhelm von Humboldt in 1808. For the first time, universities were mainly conceived as research institutions. Each science had to be given the means to organize its own development. The potential of such a reform on scientific achievements in geography became evident after 1871, when Prussia transformed Germany into an Empire: within a few years, departments of geography were founded all over the country. The professors who were recruited for the new chairs had different backgrounds, but they shared common goals: to give the new generations of German people a clear idea of what was Germany and to help to the creation of a colonial empire and the building of a strong German navy.

Since the unity of Germany was built on the defeat of France at the end of the Franco Prussian war of 1870–71, France experienced a similar rise of national militantism. The development of new institutions took a few years. By the 1880s and 1890s, French government was busy in modernizing French Universities – partly along the German line – and multiplying chairs of geography. Most of the new professors had been trained as historians, but they were motivated by the desire to build a stronger France through its colonial expansion and industrial development. They were as much involved in the reconstruction of French territorial integrity (which meant the restitution of Alsace and Lorraine, annexed by Germany in 1871) as their German colleagues in the fulfillment of German territorial unity.

In a way, these institutional transformations explained the early development of a strong nucleus of academic geographers in both Germany and France during the last decades of the nineteenth century. Did it mean that it was already possible to speak of a German and a French school of geography by 1880? No: French geographers relied mainly on German sources. They were working for the defense of French interests, but relied on the same intellectual tools as their German colleagues.

The dynamics of academic life was, however, conducive to the development of particularisms. Professors were free to choose their research field and use the sources and methods they preferred. Since travel was still expansive and time consuming, most scholars had few possibilities to visit faraway countries and to meet regularly foreign colleagues. International congresses of geography were still attracting more generals, admirals, or diplomats than university teachers. It was often easier to visit the colonial empire of one's own country than to travel really abroad.

At the national scale, there were forces that countered the fragmentation of the scientific enterprise. Since professors were generally selected by commissions, which included colleagues coming from other universities, academic careers were judged according to national rules. In France, the selection of secondary school teachers in geography (and history) was based on a competitive national examination: it forced all the university teachers in geography to work along the same lines and to cover the same basic program.

From 1905, a French national interuniversity field trip was organized every year for graduate students and university teachers coming from all over the country: it was conducive to a national codification of fieldwork.

Between the 1850s and 1880s, there was only one scientific journal in geography: the *Pettermanns' Mitteilungen*. It was published in Gotha, a small city in central Germany and written in German. The quality of the maps and data it published was such that it really enjoyed an international audience.

At the end of the nineteenth century, the situation became to change. Scientific national geographic journals were launched in most industrialized countries: *Annales de Géographie* in France in 1891, *Geographische Zeitschrift* in Germany in 1895, and the *Annals of the Association of American Geographers* and *The Geographical Review* in the United States in 1910. The main orientations of these journals were given by national authors.

Because of the way academic life was built, a mechanism of national differentiation began to operate from the 1870s and was reinforced during the last decades of the nineteenth century and early twentieth century in most Western countries. Academic geography was increasingly rehearsed on national scenes. The national intellectual life influenced it strongly. It was difficult for a Spanish geographer to escape the philosophy of nature developed in Germany by Krause: imported in Spain in the 1850s, it had become the official philosophy taught in Spanish Universities. This explained the weight given to the management of nature by the Spanish intelligentsia of the time.

The institutional dynamics was not, however, the only factor responsible for the development of national schools.

The Emergence of National Schools: The Role of Human Geography in Germany and France

Human geography appeared in the 1880s and 1890s. Ratzel coined the term 'anthropogeographie' in 1882. French geographers began to speak of *la géographie humaine* from 1895.

Human distributions were studied by geographers since a long time. Thanks to the progress of ethnography and folklore studies, the knowledge of the diversity of tools, know hows, techniques, and human settlements had much progressed from the beginning of the nineteenth century. Ritter's position analysis did not stress only the role of atmospheric and oceanic currents in the shaping of the diversity of the Earth, it focused also on the significance of 'circulation' and the role of steamships in contracting the Earth's surface.

What motivated the birth of human geography as a semiautonomous field of enquiry within geography was, however, Darwin's interpretation of evolution. Human geography was born as a kind of ecology of human groups in their relations to environments. The main aim of the new discipline was to answer a fundamental question: up to what point were human beings and human activities a reflection of environmental conditions? Human geography was born from a deterministic interpretation of man/milieu relationships.

It soon appeared that the influence of milieus on human groups could not be described as a one way process. Ratzel, who had created human geography in an environmentalist perspective, held in fact an ambiguous position. In many cases, he stressed the strong control the physical constraints exercised on peoples, especially nature peoples. He had learned from Ritter, however, the role of exchange in escaping local conditions: cultured peoples freed themselves from the tyranny of nature, thanks to their sense of space, their *Raumsinn*, and the institution which allowed for mastering distance: the state.

The evolution in France was different, but the result was similar: the environmentalist perspective which had given human geography its *raison d'être* was transformed. In order to analyze the relations, human beings developed with their environment. Vidal de la Blache had introduced in the 1880s the idea of *genre de vie*, that is, the way people exploited the local resources. It was a tool not only for exploring ecological constraints but also for appreciating their limits: ways of life traveled with the groups who had developed them, and applied to environments for which they had not been devised; technical innovation allowed the introduction of new technologies and new links with the local environment. Last but not least, and as for Ratzel, trade allowed local people to find elsewhere what they were unable to produce locally. Because of the Lamarckian interpretation of evolutionism, which was very popular in the French geography in the 1890s, the idea of adaptation to local conditions seemed more significant than that of natural selection.

In Germany as well as in France, human geography lost almost immediately the scientific ambition out of which it was born: when doing field work, scholars discovered the wide variety of solutions different groups gave to similar environmental conditions. The interest of geography stemmed from the detailed views it offered on man/milieu relationships, their variety, their stability for long periods, their crises, etc. Because of the introduction of such a possibilist perspective, human geography had ceased to be a nomothetic discipline. It was transformed into an idiographic one.

In such a context, the study of human geography could be built on three paradigms since the environmentalist one has ceased to be exclusive: (1) it could be conceived in Kantian terms as a science of the regional differentiation of the Earth's surface; (2) it could be presented in an evolutionist perspective as the analysis of man/milieu relationships; and (3) since many geographers disagreed with the breakdown of geography into human and physical parts, a new interest developed in landscapes.

These three conceptions could be combined. In Germany, emphasis was mainly put on the regional approach, according to Hettner's views, and landscape studies along the lines developed by Schlüter. The Ratzelian legacy was lighter, except in political geography. In France, geography mainly combined the regional approach with the man/milieu perspective: the starting point of research was not the landscape, but density maps, since they showed in a simplified way how the environment was used by human groups.

There was another possibility to combine the three main approaches in human geography, but it did not develop in the 1890s or early 1900s. Its success came later and elsewhere, in North America. For Carl Sauer, the aim of human geography was mainly to focus on man/milieu relationships and to analyze them through landscape expertise. It meant that geographers were as much interested in the transformation of the 'natural' com ponents of landscapes by men as in the influence of en vironment on human behavior.

The strength of the German and French schools of geography came from the fact that they expressed two ways of combining the paradigms of human geography when environmental determinism ceased to appear as the only entry to the discipline. The Berkeley school had a status similar to the German and the French ones, even if it was not a national school.

National Schools of Geography Elsewhere

The development of national schools resulted from the collapse of the environmentalist conceptions as de veloped in the 1880s and 1890s: human geography could no more be defined by its capacity to answer one central question – the influence of milieus on human beings and human behaviors. Three perspectives competed in the new discipline. They could be – and were – combined in three main ways; hence, the emphasis given to the French and German schools – which were genuine national schools – and the Berkeley one, which was not a national school.

Geographers who were responsible for the develop ment of the discipline in the last decade of the nineteenth century and the beginning of the twentieth chose as models either the German or the French school, or tried to combine them. Their conceptions were also influenced by the earlier geographical traditions existing in their countries and the *Weltanschauung* of their elites.

The influence of the French school was strong in Belgium, the Netherlands, Britain, Spain, Portugal, Latin America, and Rumania. The influence of the German school was prominent in Austria and the former countries of the Austro Hungarian Empire, Russia, the Baltic States, Scandinavia, and Japan. There were also countries that looked at both France and Germany. In Italy, intel lectual contacts were traditionally stronger with France, but the geographers of Trentino and Trieste (regions which were under Austrian rule until World War I) were trained in Vienna and introduced German views. In Scandinavia and the Baltic States, French geography played a complementary role to the German one.

Britain had played a pioneering role in the great ex plorations of the eighteenth and nineteenth centuries. The London Missionary Society had been partly re sponsible for the creation of the Royal Geographical Society. British geography was in this way closely asso ciated with imperial expansion. There was not a trauma similar to that of the Franco Prussian war of 1870 for accelerating the institutionalization of the discipline. It was only in the 1880s that a new interest developed in the teaching of geography as a tool for strengthening the British and imperial identities.

The nineteenth century tradition of exploration kept a more important role in Britain than elsewhere. Most of the ideas used by early British academic geographers were imported from France. Patrick Geddes served as middleman between the two countries: he was closer to Elisée Reclus, on one hand, and Le Playan sociologists on the other, than to the Vidalian school. Hence the em phasis given to field work – through the influence of Le Play house, in London. The French regional approach was also imported and at the same time, reinterpreted. Out of these foundations, British geography soon de veloped original orientations. In the inter war period, historical geography began to be one of its main origin alities, thanks to Darby. From the time of Patrick Geddes, the interest for planning was strong. It was strengthened in the 1930s by the preparation of a land use survey organized by Dudley Stamp. During World War II, it appeared as an invaluable tool for the British Air Force in charge of locating hundreds of new military airports, and for the government, which had to develop the production of cereals for feeding the British population.

It was certainly more difficult to define the British school of geography than the German or French ones, but its specificities kept consolidating during the first half of twentieth century.

Spain provides another example of the formation of national schools in the first half of the twentieth century. The enthusiasm for geography resulted, in the second half of the nineteenth century, from the success of the Krausian philosophy of nature. The main problem for Spanish people was to break with the spiral of decline, which had started in the seventeenth century and had been accelerated by the Napoleonic invasion. The problem was not, as in Germany or France, a problem of boundaries. It was a problem of development. In order to transform their countries into a modern one, Spanish elites were convinced that Spanish people had to become more aware of the severe constraints created by Spanish climate and the low rainfall of a large part of the country. Hence the emphasis on *excursionismo* – field trips organ ized for a wide middle class public – and geographical education.

Spanish intellectuals knew that in order to avoid too strong an influence of France, it was good for them to look toward Germany. Many German textbooks were in this way translated into Spanish in the 1920s and early 1930s. Geographical proximity and the widespread knowledge of the French language gave, however, to the French school its dominant influence.

The impact of the Civil War was important in this respect. Many Spanish intellectuals (and geographers) were killed or had to leave their countries – most of them settled in France, Mexico, and other Latin American countries. In the early 1940s, Franco's regime launched a policy of scientific research, where geography played an important role – was it not necessary to remind the country of its past, its former empire? Was it not useful to teach Spanish people the best way to develop their resources? As a result, the regional approach appeared as a conception which could be shared at the same time by Franquist and non Franquist geographers. Unfortunately, it was also a time when regional geography started to decline in France and elsewhere.

Japan provides another fascinating example of the shaping of a national school. As shown by Keiichi Takeuchi, there was an old geographic tradition in the country, which did not disappear completely with the Meiji revolution. As early as 1872, geography was introduced, however, in the Japanese school curricula. Japanese people were eager to mobilize all the Western forms of knowledge. They translated many Western books. Bunjiro Koto (1856–1935) had been sent to Germany to study geology in the early 1880s. Back to Japan, he was appointed as a professor of the new National University of Tokyo. He participated in the launching of the *Chigaku Zasshi* (*Journal of Geography*) in 1889 and in the foundation of the Geographical Society of Tokyo in 1891. He introduced Ratzel's anthropogeography in Japan. Two of his former students, Naomasa Yamasaki and Takuji Ozawa, headed the two main departments of geography in Japan before World War II.

The filiation between German and Japanese geographies was evident, but as shown by Takeuchi, other influences played. For instance, Japanese scholars discovered earlier than in the West the significance of von Thünen's analysis of agricultural location. Japanese geographers looked also toward America.

Japanese geography developed in an increasingly imperialist atmosphere: it hampered the development of the discipline in the 1930s.

In a way, geography was a successful discipline in the United States from the start: Thomas Jefferson's *Notes on the State of Virginia* and Samuel Morse's textbooks testified of the fact. Geography did not play, however, the same role as in European countries: American consciousness was not built on a shared feeling of territorial belonging, but on the democratic experience of founding a new nation: hence the reduced place given to geography both at the elementary and secondary schools level.

During the nineteenth century, geography played an important role in the East Coast universities, thanks to the immigration of a few scholars, Louis Agassiz and Arnold Guyot. Both were Swiss born, both had been trained in Berlin. They did not succeed in initiating a national school on the German model. At the end of the nineteenth century, American geography was mainly personified by William Morris Davies: it was centered on geomorphology. After World War I, interest shifted toward human and economic geography. It was really the time when something like a national school appeared in the United States. In the Middle West Universities, where geography flourished at that time, the main problem was a methodological one: how to codify field work, to use cartographic evidence, and to develop enquiries? Hence the shaping, from the early 1920, of a Middle West school of geography, which differed from the European ones by the emphasis it puts on methods rather than on content. Hartshorne tried to graft it on the Hettnerian tradition of regional studies, but American studies did not conform to this German model.

At the same time, Carl Sauer developed an interpretation of human geography, which combined two of the three main paradigms of the geography of that time: man/milieu relationships and landscape analysis. Since the Berkeley school was based on a deeper reflection on the nature of geography, it had more influence abroad than the Middle West school.

Changing Conditions and the Disparition of National Schools

After World War II, it became increasingly difficult to speak of geography in terms of national schools. There were several reasons for that.

The conditions of mobility within the academic sphere changed. Travel costs dropped and thanks to airplane, the time needed to visit a faraway country was much reduced. In their international relations, geographers relied increasingly on only one language, English, instead of several – German, French, and English. Since the countries of the former British Empire and the United States needed geographers, an international market of geographical expertise appeared: it concerned first North America, Britain, the Southern Hemisphere dominions, and countries of Northern Europe like the Netherlands and Sweden. It is now increasingly a world market.

The conceptions of geography evolved rapidly. The 'new geography' of the 1960s broke with the plurality of coexisting paradigms that characterized the first half of the twentieth century and proposed a unified theoretical spatial view of the discipline. Humanist and radical approaches criticized the idea of geography as a spatial science but did not restore the plurality of paradigms on which the national schools had prospered. For humanist as well as for radical geographers, the ambition is to open a new perspective on the discipline. The idea that it could be based on a mixture of different paradigms does not appear as a serious one.

Up to what point are the geographers born in Anglo phone countries provided with a serious advantage in a world where the English language is the only one to enjoy a global use? Since the academic capital they build is more widely recognized than that achieved by colleagues using other languages, is there not a threat of intellectual imperialism? The danger is a real one: new ideas for mulated in other languages are often ignored for some time in the English speaking world and their authorship is sometimes attributed to Anglophone scholars: con cerning the role of Marshallian districts and clusters in contemporary economic geography, Porter's study of 1990 is more widely quoted than the previous studies of Italian scholars like Bagnasco or Becattini. Thanks to Paasi's study, many Anglophone geographers are now conscious of the problem. It is particularly serious at a time when geography is experiencing a cultural turn. Even if scien tific research and economic international relations are increasingly expressed in English, and if modern media are responsible for the universal diffusions of novels, movies, or TV series using the English language, the cultural diversity of the world has not disappeared. Would it not be easier to take it into account if more consider ation was given to the languages ordinary people – and many scholars – use in their daily life?

It is increasingly difficult to defend the idea of a plurality of scientific approaches within geography. In order to become conscious of new problems and launch new orientations, some places offer more opportunities than others. It is the conviction shared by some of the geographers who work in Los Angeles: they believe that the center of world history has moved West and is now located in southern California; they enjoy the privilege to be in the best place to build an 'avant garde' view of the discipline. It is the reason for which some people speak of a Los Angeles school of geography. The reality behind such a school is much different from that existed in the first half of the twentieth century. There are no similar ities between the Berkeley school of the 1930s and the Los Angeles school of the 1990s.

See also: Geography, History of; Landscape; Nationalism.

Further Reading

Bagnasco, A. (1977). *Tre Italia. La Problematica Territoriale Dello Sviluppo Italiano*. Bologne: Il Mulino.

Becattini, G. (dir.) (1987). Mercato e Forze Locali: Il Distritto Industriale. Bologne: Il Mulino.

Berdoulay, V. and Soubeyran, O. (1991). Lamarck, Darwin et Vidal: Aux fondements naturalistes de la géographie. *Annales de Géographie* 100, (561 562), 617 634.

Buttman, G. (1977). *Friedrich Ratzel. Leben und Werke Eines Deutschen Geographer*. Stuttgart: Wissenschaftliche Verlagsgesellschaft.

Claval, P. (1984). *Géographie Humaine et Économique Contemporaine*. Paris: PUF.

Claval, P. (1998). *Histoire de la Géographie Française de 1870 à Nos Jours*. Paris: Nathan.

Eisel, U. (1980). Die Entwicklung der Anthropogeographie von einer 'Raumwissentschaft' zur 'Gesellschaftswissenschaft, Kassel, Kasseler Schriften zur Geographie und Planung, n° 17.

Freeman, T. W. (1980). *A History of Modern British Geography*. London: Longman.

Godlewska, A. (1999). *Geography Unbound. When Description Fell to Theory*. Chicago: Chicago University Press.

Gómez Mendoza, J. (2006). Géographie régionale et identité nationale: Penser la géographie espagnole selon un modèle étranger. In Claval, P. (ed.) *Géographies et Géographes*, pp 134 154. Paris: L'Harmattan.

Gregory, D. (1994). *Geographical Imaginations*. Oxford: Blackwell.

Hettner, A. (1905). Das wesen und methoden der geographie. *Geographische Zeitschrift* 11, 549 553.

Johnston, R. J. (1979). *Geography and Geographers. Anglo American Human Geography Since 1945*. London: Arnold.

Livingstone, D. (1992). *The Geographical Tradition*. Oxford: Blackwell.

López Morillas, J. (1956). *El Krausismo en España. Perfil de una Aventura Intelectual*. Madrid: Fondo de Cultural economica.

Martin, G. (2005). *All Possible Worlds. A History of Geographical Ideas*. London: Oxford University Press.

Paasi, A. (2005). Globalization, academic capitalism, and the uneven geographies of international journal publishing spaces. *Environment and Planning A* 37(5), 769 789.

Platt, R. S. (1959). *Field Studies in American Geography*. Chicago, Department of Geography Research Paper No. 61.

Porter, P. (1990). *The Competitive Advantage of Nations*. New York: Free Press.

Ratzel, F. (1882 91). *Anthropogeographie, Oder Anwendung von Erdkund auf die Geschichte*, vol. 2. Stuttgart: Engleborn.

Sauer, C. O. (1963). In Leighly, J. (ed.) *Land and Life. A Selection from the Writings of Carl Ortwin Sauer*. Berkeley: University of California Press.

Schluter, O. (1899). Bemerkungen zur Siedlungsgeographie. *Geographisch Zeitschrift* 5, 65 84.

Schultz, H. D. (1980). *Die Deutschsprachige Geographie von 1800 bis 1970. Ein Beitrag zur Geschichte ihrer Methodologie*. Berlin: Selbsvertrag der Geographischen Institut der Freien Universitat Berlin.

Stamp, L. D. (1931). The land utilization survey of Britain. *Geographical Journal* 78, 40 53.

Takeuchi, K. (2000). *Modern Japanese Geography*. Tokyo: Kokon Shoin.

Vidal de la Blache, P. (1886). Des rapports entre les populations et le climat sur les bords européens de la Méditerranée. *Revue de Géographie* 19, 409 419.

National Spatialities

T. Edensor, Manchester Metropolitan University, Manchester, UK

Introduction

To conceive the nation in spatial terms is a complex matter that brings together a number of processes and theoretical approaches. For the relationship between space and national identity is variegated and multiscaled, producing a complex geography that is constituted by borders, symbolic areas and sites, serialities, constellations, pathways, dwelling places, and everyday fixtures. The national is evident not only in widely recognized grand landscapes and famous sites, but also in the mundane spaces of everyday life. This article discusses a range of different national spatializing processes that interweave with each other to consolidate a strong cognitive, sensual, habitual, and affective sense of national identity, and provide a common sense spatial matrix which draws people and places together in spectacular and banal ways; and accordingly, discusses spatial boundaries, preferred ideological landscapes, iconic sites, serial spaces, everyday realms of national habit, the notion of 'home', and media representations of national space.

The Nation as Bounded Space

The nation continues to be the preeminent spatial construct in a world in which space is divided up into national portions. The nation is spatially distinguished as a bounded entity, possessing borders which mark it as separate from other nations. Borders ostensibly enclose a definable population subject to a hegemonic administration in the form of a discrete political system holding sway over the whole of this space but which, in a world of nations, is expected to respect the sovereignty of other nations. These borders are also imagined to enclose a particular and separate culture, a notion that is informed by hegemonic ways of differentiating and classifying cultural differences that are subordinate to the nation, and conceived as part of national cultural variety.

The nation masquerades as a historical entity, its borders providing a common sense existence that locates the imagined community in time and space. Yet despite the appearance of fixity, nations continually change: they fragment, emerge, conjoin with each other, lose parts and expand, as any glance at a historical sequence of political maps throughout the twentieth century demonstrates. For the nation state may be threatened by large, supranational federations which organize around trade, social legislation, and law (for instance, as with the European Community) promoting widespread fears about the demise of national political autonomy and culture; by autonomy seeking regions contained within its borders (such as Scotland, the Basque Country, and Kashmir); and by globalization, which produces global flows of goods, information, images, and people from elsewhere, a proliferation of extranational diasporic identities, virtual spaces such as 'cyberspace', and the development of planetary consciousness or global awareness, prompted by greater knowledge of challenges and threats to the world as a whole.

Although we continue to live in world of nations, defensive forms of national identity and rigidly guarded borders are increasingly threatened, since nations now need to have more permeable borders from the financial, informational, commodity, and cultural flows which circulate the globe. Despite these undoubtedly powerful processes, claims that the nation as space of primary belonging is in decline are exaggerated. For national identity can be reconstituted in diaspora, can forge new cultural constructions of difference out of the confrontation with otherness, and can morph into innumerable spatial forms. Accordingly, at a practical and imaginary level, national geographies continue to predominate over other political spatial entities.

Ideological National Landscapes

It is difficult to imagine a nation without conjuring up a particular rural landscape. Ireland has become synonymous with its 'West Coast', the USA with its 'Wild West' and Norway with its 'fiords'. Argentina is inevitably linked with images of the pampas, with: gauchos riding across the grasslands, but the deserts, swamps, and mountains of Argentina tend to be overlooked. These specific landscapes are loaded with selective symbolic values and mooted national virtues, and often stand as symbols of historic continuity. Stamped with the modern construction of separate urban and rural realms, wherein the *gesellschaft* of the city is contrasted with a highly romanticized rural *gemeinschaft*, such 'national landscape ideologies' symbolize the *genius loci* of the nation, and is understood as the realm of forebears who typically tamed, the land and were nurtured here.

These iconic, privileged landscapes continually circulate through popular culture: in tourist campaigns, films, advertisements, and painting. Yet they are highly selective. For instance, the rural scenes that are held to epitomize Englishness, containing a reliable mix of churches, thatched or half timbered cottages, rose laden gardens, village greens, games of cricket, country pubs, hedgerows, golden fields of grain, plough and horses, and hunting scenes, tend to be geographically confined to the South of the country. Lowenthal identifies four esteemed attributes which are encoded in the national landscape ideology encoded in these representations, namely, 'in sularity', 'domestication', 'stability', and a 'sense of order'. In suggesting a historical fixity, such ideologies purify space and act to stabilize a sense of national identity in times of rapid change and ignore rural developments in agribusiness, commuter communities, and the rise of tourism, which have decentered this notion of rural idyll.

Iconic Sites

As with the selectivity of ideologically loaded landscapes, nations also contain particularly iconic sites which sym bolize particular values. Such sites often commemorate historical events or processes (monuments such as the Statue of Liberty, Arc de Triomphe, and Nelson's Column), or provide evidence of a national 'golden age' (such as Stonehenge, the Great Pyramids, and the Taj Mahal), or conversely, celebrate the modernity and pro gress of the nation (such as the Empire State Building, Petronas Towers, and Sydney Opera House). In addition, there are a host of sites which symbolize official power: the royal palaces, halls of justice, military edifices, presi dents' houses, parliamentary buildings, and so on, which provide the materialized spaces of national rule.

These spatial attractors often occur in an ensemble of related sites, to constitute ceremonial points of reference or 'memoryscapes' around which circuits of memory are organized. For instance, a tourist trip to Paris involves gazing upon a number of emblematic places, including the Eiffel Tower, the Place de Concord, Notre Dame Cathedral, the Arc de Triomphe, the Champs Elysees, and the Louvre. These are noteworthy both as signifiers of France for outsiders and as ideological statements about Frenchness and the republic within France. How ever, the meaning and power of such sites may become obsolete, as with the statues in Eastern European nations which no longer reflect national sentiment following the fall of communism. Despite the desires of the powerful to imprint particular ideological constructions of national identity upon space, such aims may be thwarted by changing notions of belonging and style. This is par ticularly apparent in stone statues of military heroes, philanthropists, and statesmen which have been critiqued by feminists as a highly gendered form of spatial repre sentation. Alternatively, monuments and memorials may be interpreted from a number of different perspectives, and yet in these cases, such symbolic sites can be widely shared as a cultural resource although the values which they represent can be contested.

In addition to these iconic sites, other symbolic places of popular assembly such as squares, beaches, commons, parks, and thoroughfares serve as venue for communal endeavors such as festivals, demonstrations, tourism, and informal gathering. Such sites can be places of proximity, where people go to look at and meet other co nationals. Certain spaces of assembly associated with national identity, such as Times Square in New York City, Trafalgar Square in London, the Djma el Fna in Marrakesh, Gateway of India in Bombay, and the Zocalo in Mexico City, are venues for seething motion and a multiplicity of activities, identities and sights, and are places where collective rituals (such as New Year's Eve celebrations) are carried out. In contrast to the rather purified, single purpose spaces of state power, they are more inclusive realms which allow the play of cultural diversity, providing a space in which tourists and in habitants mingle, people picnic and protest, gaze, and perform music or magic, sell goods and services, and simply 'hang out'.

Other popular highly symbolic national sites include sports stadia (Wembley, the Maracana, Madison Square Gardens, and the Stadio Olimpico), theaters, museums, and historic districts. In the case of the former, it might be argued that since sport is increasingly powerful as a focus for national identity and prestige, they have become particularly significant in recent years. All such sites, however, tend to be nonexclusive in the social mixing of co nationals they allow, and the values they symbolize are rarely passed down by a cultural elite but emerge from familiar interaction with such spaces in conditions of co presence with fellow nationals.

Serial Landscapes

Most accounts of the relationship between space and national identity have focused on the kinds of symbolic landscapes and sites discussed above. However, few de pictions explore the more mundane spatial features of everyday experience which are equally important in constructing and sustaining national identity. For ir respective of the effects of globalization, most of us live in recognizable worlds, distinguished by distinct material structures, and the predictable distribution of objects and institutional arrangements. Within these inhabited realms, surrounded by familiar things, routes and fixtures, we make our home by the accretion of habitual enactions, by our familiar engagement with the space in which we live. While these more intimate spaces may be considered as the agglomeration of regional or ethnic synchronicity, shared understandings and collective rituals, familiarities also extend across a larger scale, stitching the local and the national together through their serial reproduction across national space.

The semiotic imprint of familiar features constitutes a sense of being in place in most locations within the nation, and this is facilitated by vastly enhanced mobility which has enabled people to travel across national space. However, these fixtures are not only read as signs, but are also felt and sensed in unreflexive fashion. There is a plethora of everyday, mundane signifiers which are noticeably not present when we travel outside the familiar national space and into other national spaces. These recurring institutions, vernacular features, and everyday fixtures are embedded in local contexts but persist throughout the nation as serial features. For instance, the British landscape contains a host of unremarkable but predictable elements. The institutional matrix of everyday life is signified by familiar commercial and bureaucratic notices which indicate where services can be procured or commodities purchased. Likewise, service provision is marked by (red) telephone boxes and post boxes, distinctly designed equipment for conveying flows of power and water – including grids, fire hydrants, street lighting, guttering, telegraph poles and pylons, identifiable road signs and artifacts, styles of domestic architecture, garden ornamentation, and home décor which tend to fall within a recognizable vernacular range. Such elements rarely confound expectations of how space will appear and feel, and when this occurs, unfamiliar features stand out in relief against this normative spatial context. Particular fixtures are occasionally more overtly celebrated as quotidian features of national identity and also may be sought out by foreign tourists in search of signs of otherness, of unfamiliar national identity. Here we can identify mundane features such as British pubs, French cafes, and Italian churches.

These mundane signifiers are also accompanied by the recognizable forms of flora and fauna which recur throughout most environments. The circling black kites throughout urban parts of India are unremarkable, and a host of other unspectacular animals and plants are also part of everyday landscapes in most national spaces, rarely commented upon except when they are no longer so common. In addition to these living elements, serial landscapes produce distinct soundscapes, sonic geographies that provide a backdrop to everyday life, so that the distinctive sounds of traffic, muezzins or church bells, birdsong, and music also install a national sense of being and belonging.

While such serial features provide anchors for spatial identity, they should not be imagined as testifying to a static landscape. As the production of space goes through dynamic cycles, inhabitants must also accommodate themselves to new developments. The awareness of change and growing familiarity with the new creates a complex spatial network. But as people adapt to transformation, they are apt to domesticate changes so that they are incorporated into familiar national spatial contexts.

National 'Taskscapes'

Above, the emphasis is on the routine experience of space, the ways in which particular actions recreate space, and the modes of habituation which render it familiar. After all, the most common spatial experience inheres in everyday, familiar space, that realm which forms an unquestioned backdrop to daily tasks, pleasures, and routine movement. The street, neighborhood, and local institutions and facilities form the terrain on which quotidian practices are unreflexively carried out. Such mundane space is structured and organized to enable continuity and stability, and equally, the meaning and functions of such stable environments are reproduced at a daily level by the repetitive, collective, sensual, practical, and unreflexive performances of its inhabitants. To emphasize, for much of the time we unreflexively perform embodied habits in accordance with the affordances of place, with the constraints and opportunities of space, and such enactions become embedded over time.

At one level, regular individual time–space paths are marked upon familiar space and these intersect with spatial and temporal patterns of collective routine practice at sites such as local shops, bars, cafes, and garages. Such routines become sedimented both in the landscape and in the habit body to provide geographies of continuity constitued by paths, constellations of co presence, fixtures, meeting points, and intersections. Linked by the roads down which people drive and the pavements that they walk along, venues for shopping, services, relaxation, and entertainment are spaces of circulation in which people coordinate and synchronize activities, stabilizing social relations in time–space. The concept of the 'taskscape' captures the space within which these unreflexive modes of being in the world take place, for it is a space in which a practical, multisensory knowing facilitates spatial practice and coordinates collective activity. The development of an everyday practical orientation means that mundane practices are not static but allow for the adaptation of habits from the past as part of the ongoing process through which space is (re)produced.

Crucially though, when these practices are contested by others who do not do things the way we do them, or space becomes defamiliarized by redevelopment or the intrusion of 'foreign' elements, there is the potential to become confused and threatened. When enduring fixtures, familiar pathways, characteristic features, and landscapes are removed or radically altered, a sense of disorientation can result. For instance, the removal of red telephone boxes along the streets of British towns and cities caused fear that the mundane environment was under threat. Likewise, the perceived differences in household décor, eating habits, musical preferences, religious practices, and clothing conventions may appear strikingly different when they appear in landscapes of habit in which esthetic, functional, and practical norms have become embedded and they may be experienced as practices that defamiliarize the local.

In addition to the perceived threats to stability by the intrusion of these unfamiliar elements, taskscapes provide stages for the enaction of functional know how. For instance, shared practical knowledge ensures that locally we know how to take a bus, join a library, buy a favorite brand of chocolate bar, go for an alcoholic drink, park a car and drive it at an appropriate speed. This form of spatial knowledge, largely learned in the familiar local world, extends to the national level, since certain templates of spatial organization and the serialities mentioned above extend across the nation (in retail outlets, public transport systems, entertainment venues, car driving, service infrastructures, and provision) so it is likely that most of the shared, learned, and rehearsed, practical competencies that are required for the performance of a host of mundane practices can be reproduced in other national locations with little need for reflection or concentration. However, when traveling to unfamiliar, foreign spaces for the first time, the practical knowledge that is so integral to reproducing an awareness of identity, a sense of place and its actual characteristics may well be irrelevant. It now becomes difficult to get things done because the familiar fixtures in the landscape are nowhere to be seen, the functions of shops and other services are baffling, the movement across space by public transport is confusing, social practices appear obscure, and one's own performances in space are neither practical nor understood. Even carrying out the most ordinarily automatic, mundane endeavors – such as buying a chocolate bar or driving – may present a considerable challenge, in which case habitual unreflexivity may be replaced by intense concentration or uncomfortable self consciousness. These underacknowledged spatial and performative dimensions are rarely the subject of academic analysis yet they highlight how national identity devolves to the level of the local, and is extended throughout national space in everyday settings.

Homely Space

The multiscalar dimension of national identity is well illustrated by the host of spatialities which surround the notion of 'home', which can variously refer to house, neighborhood, village, city region, or nation and is thus a term which links differently scaled affective spaces. For instance, 'good housekeeping' can be applied to both the family finances and the policies of national government, and the national football team plays 'at home' and 'away'. Yet constructions of home are integral to the boundaries of space making, notably in distinguishing the national from the space of the 'other'.

Perhaps most saliently, home is usually conceived of as a place of comfort where we are most relaxed and unselfconscious, the place where we 'make ourselves at home' in the world according to social and aesthetic conventions about conviviality, domesticity, and furnishing and decorating space, all practices which demarcate domestic territorial boundaries. Homemaking is able to domesticate elements from elsewhere, establishing conditional and contained connections with other places where the home remains the hub of such relations. Again, domestic practices produce taskscapes where we develop bodily habits and routines around everyday fixtures in an organized space, and a practical competence, for instance, in understanding how electric appliances work. Once more, these unreflexive routines only become apparent when we enter domestic spaces in which different codes inhere.

Besides practical knowledge though, the sense of comfort and affective belonging that emerges from the home is particularly founded upon the ways in which it is sensually experienced and embedded in memory. The 'felicitous' areas of childhood play within domestic space, such as bedrooms, attics and parlors, with textures and micro atmosphere absorbed in childhood are complemented by the those of the immediate locality, the bus shelters, dens, woods, back streets, parks and fields, and other sites of childish adventure and adult experience to provide a rich source of often involuntary memory. Thus, the texture of the world, the feel of familiar spaces underfoot, the barely perceptible subtleties of climate, vegetation, and built environment are part of a sense of place which extends from home and garden to the wider locality and beyond.

As with other spaces, the home must be continually reproduced by domestic work and by social and cultural activities, and it is always liable to be threatened by the presence of otherness. Defensive operations debar the unwelcome from entering the 'homeland' at state level, and they are also mobilized at a smaller scale where cultural, ethnic, and 'racial' difference is marked in discriminatory housing markets, employment, and social services. Those perceived as 'outsiders' are often situated in marginal places where their difference can be contained (and desired). These defensive responses indicate that although the home is often conceived as an idealized refuge, it can be a site of both oppression and liberty, and it can restore and stifle.

Increasingly, a sense of homeliness may not only be achieved by a situatedness in particular physical space but may also be reached via homely networks of people and information. Yet while it is progressively more common for people to identify several 'homes' as social and cultural networks become more complex, familiar spatial reference points are still sought. The importance of feeling comfortably 'at home' in space, and in understanding home as the site from which national tastes, rituals, good habits, and modes of comportment should be learned and reproduced is particularly evident in expatriate and migrant communities. Accordingly, such communities are apt to forge networks with the 'homeland', whether virtual or material, and reproduce cooking, decorating, convivial, and esthetic national conventions in the realm of the 'other'. Home and its national(ist) manifestation can, therefore, be simultaneously grounded and mobile, not merely present in national space, but exported and reestablished.

Media Representations of National Space

Having discussed the symbolic, social, and lived dimensions of inhabiting and reproducing national space, it is vital to consider a further dimension, namely, the representation of national space, for all the characteristics identified above also loom large in the representations articulated in popular and 'high' cultural forms. In a mediatized age, the proliferation of images and verbiage also include representations of national space, ranging from the stereotypical and archetypal, to the incidental and everyday. Certain iconic sites are represented both in paintings and photography, and in tourist marketing and popular films, continually underlining their symbolic value. More typically, national landscape ideologies feature in mythic cinematic portrayals, such as the grand American landscapes featured in Western films that are further loaded with cultural values concerning ruggedness, the wild and yet a land that can be conquered, and provide the setting for equally symbolic acts of heroic individualism. Less obviously, but equally importantly, the imagined community of the nation is reproduced in the plethora of everyday television programs, advertisements, and films which indiscriminately feature mundane scenes of daily life. For instance, popular British soap operas all feature a central pub, domestic interiors, corner shops, and cafes which can consolidate a sense of belonging to the normative spaces of British life. Such recurrent images are part of the production of the imaginary geographies of the nation, and produce intertextuality, or an interspatiality, in that they conjure up spaces that are familiar to the national viewer. A familiarity with the seriality of national space is thus further consolidated by its reproduction via representations across the media. Moreover, in this case, the television links home and domesticity in that such images are consumed in the domestic, mundane context of the home.

Conclusion

National spatialities are multiscalar and complex, folding into each other to reproduce the resonance of national identity. The nation is domesticated, replicated in local contexts, and practiced in everyday spaces. It is present in iconic sites, esteemed national landscapes, serial spaces, and in the unreflexive reproduction and sensual experience of familiar space, all of which are further entrenched as national by their portrayal in popular media representations. Modern nation building has entailed the incorporation of all internal differences, so that whatever regional and ethnic differences may predate the nation's formation, they all become subservient to, and part of, the greater national entity. Region, city, village all remain tied to the nation as a larger ontological and practical framework, and are absorbed into this code (and hierarchy) of larger significance. Enhanced mobility and mediatization means that such scenes are commonly witnessed by co nationals, and this extends to foster a global international recognition.

In recent years, there have been several accounts that point to the increasing placelessness wrought by globalization. Through global processes, it is asserted, the national, regional, and local specificities of place are erased in the production of homogeneity and 'serial monotony' to produce 'nonplaces'. However, while these assignations appear to characterize certain spaces such as international airports and shopping malls, it must be contended that global processes are typically contextualized within localities, where rather than overwhelming local space, they are domesticated, adopted via codes of spatial ordering. Other arguments point to the disembedding of local practices and the redistribution of identities away from the nation and into localisms, groups with a global membership, virtual communities and lifestyle gatherings. However, while the nation has undoubtedly been challenged as the primary space of loyalty and identity, it retains its power through the institutionalized repetition of everyday spatial practice and the seriality of landscape, as well as it continued hold on popular geographic imagination through flows of media. Moreover, in the face of volatile and proliferating global flows of all kinds, the nation remains a point of anchorage, a space which promises epistemological and ontological security amidst conditions of ceaseless flow.

See also: Globalization, Economic; Nationalism; Scale.

Further Reading

Anderson, B. (1983). *Imagined Communities*. London: Verso.
Cubitt, G. (1998). *Imagining Nations*. Manchester: Manchester University Press.

Daniels, S. (1993). *Fields of Vision: Landscape Imagery and National Identity in England and the United States*. Cambridge: Polity.
Edensor, T. (2002). *National Identity, Popular Culture and Everyday Life*. Oxford: Berg.
Edensor, T. (2006). Reconsidering national temporalities: Institutional times, everyday routines, serial spaces and synchronicities. *European Journal of Social Theory* 9(4), 525–545.
Frykman, J. and Lofgren, O. (eds.) (1996). *Forces of Habit: Exploring Everyday Culture*. Lund: Lund University Press.
Hage, G. (1996). The spatial imaginary of national practices. *Environment and Planning D: Society and Space* 14, 463–485.
Harrison, P. (2000). Making sense: Embodiment and the sensibilities of the everyday. *Environment and Planning D: Society and Space* 18, 497–517.
Herzfeld, M. (1997). *Cultural Intimacy: Social Poetics in the Nation State*. London: Routledge.
Hooson, D. (ed.) (1994). *Geography and National Identity*. Oxford: Blackwell.
Ingold, T. and Kurttila, T. (2000). Perceiving the environment in Finnish Lapland. *Body and Society* 6(3–4), 183–196.
Johnson, N. (1995). Cast in stone: Monuments, geography and nationalism. *Environment and Planning D: Society and Space* 13, 51–65.
Matless, D. (1999). *Landscape and Englishness*. London: Reaktion.
Ross, K. (1995). *Fast Cars, Clean Bodies: Decolonisation and the Reordering of French Culture*. London: MIT Press.
Taylor, T. (1994). *A Dream of England: Landscape, Photography and the Tourist's Imagination*. Manchester: Manchester University Press.

Nationalism

D. H. Kaplan, Kent State University, Kent, OH, USA

Glossary

Banal Nationalism The everyday manner in which nationalism and national identity are expressed and permeate the lives of the citizens of a nation.

Decolonization National independence of colonies from an empire.

Diaspora Originally applied to the Jewish population, it has been used to refer to all national groups that reside, often in a dispersed fashion, away from their original homeland.

Discursive Landscape A combination of textual and visual landscape elements including narratives, images of the landscape, and various symbols of national identity that are found on the land.

Irredentist Nationalism Effort to secede from an existing state in order to join with a neighboring state.

Nation In its modern usage, a territorially based community of people tied together by cultural similarity (whether ancestral or created), a shared historical legacy, and a sense of commonality. All nations either possess or seek to possess a political state.

State Territorially based political entity that contains a government, a permanent resident population, sovereignty, and clearly delimited boundaries.

Substate Nationalism Effort to create a sovereign state from a culturally distinct region.

Supranational Referring to an international organization that includes several political states. Member states retain their sovereignty, but must abide by some of the decisions of the supranational organization, often made by majority vote.

Is it possible to conceive of nationalism without geography? Nationalism is a political ideology that is based on trying to transform national identity into some form of a political state. It is a territorial strategy because any political state requires some type of sovereign space that it can control and within which it can exercise legitimate force. Without territory a political state would be boundless, spaceless, and placeless – it would lose its rationale.

Some common definitions of nationalism and national identity do not explicitly include the ideas of space, territory, or geography. One of the most succinct definitions of the nationalist doctrine holds that

> community is naturally divided into nations, that nations are known by certain characteristics which can be ascertained, and that the only legitimate type of government is natural self government. (Kedourie, 1960)

The role of territory in this definition is undetermined. Medieval references to 'nations' were more likely to evoke ethnic groups, culturally distinct but not territorially based. Yet other definitions have clearly included the importance of territory or space in their ideas of nationalism. Sir Henry Maine in the nineteenth century argued that territory, along with kinship, undergirded political unions from the ancient or modern era, a reference to what might today be considered a 'nation state'. A modern scholar of nationalism, Anthony Smith, has argued that nationalist desires are based on the possession of land and any national cohesion requires the possession of territory. Geographers have long maintained that the nationalist project is fundamentally an ideology that seeks to construct a national space out of social space.

The primary debates about nationalism have to do with the constitution of the nation itself. The key fissure is between those who see the nation as something of long duration and based on a longstanding ethnic identity – these are termed primordialists or perennialists. Those who see nations as socially constructed and of relatively recent duration are termed social constructivists. In fact, the distinction is one of emphasis since all but the most radical primordialists would agree that nations are not 'essential' human attributes and most social constructivists admit that national identity is a most durable identity that is normally based around some core elements. National construction can stem from things such as economic modernization, print capitalism, political instrumentalism, and other things. National identity is something that has to be continually maintained, and one of the goals of nationalism is to keep the flames of nationhood alive. Institutions facilitate this process. The national media can help to create a sense of unity and of difference with outsiders. Similarly literature can be a very important force for national unity as well. The educational system can help to inculcate young children in a common national history and national identity, and throughout this of course we can discuss how significant is language in creating a commonality as well.

Geographic discussions of nationalism generally assume that it is a spatially implicit ideology and that territory is a vehicle which supports nationalist conceptions and nationalist actions. When you consider the language that nationalists use to describe their nation, the vitality of space is evident. Terms such as homeland, territory, motherland, fatherland, even land itself – these are all geographically based. Even those nationalist groups who do not happen to occupy the same territory – for instance, diaspora groups dispersed far from their ancestral lands – are anchored to the prospect of some sort of territorial consolidation in the future. Perhaps the best known example of this would be the Zionist move ment formulated by Jews, most of who lived well beyond the territory that they sought to occupy and control. A strong territorial imperative imbues most and perhaps all nationalist movements. It is difficult to think of an ex ception. Nationalism can be examined based on a variety of geographic concepts. The purpose of this article is to show how certain geographic themes such as those of space, place, centrality and peripherality, landscape, and scale fit into geographic expressions of nationalism.

Nationalism and Spatial Identities

There exist a variety of different nationalisms. One common form of nationalism, termed substate national ism, relates to efforts to create a whole state out of a particular culturally distinct region. Another form of na tionalism, termed irredentist nationalism, seeks secession from an existing state in order to join with a neighboring state. The efforts of colonies to gain independence from a mother country or a larger empire is termed de colon ization. Yet another form may be to rework the national character of the state itself; in other words attempt to re create the national identity and constitution of the state. This nationalism *in situ* can be a very powerful force in creating new types of politiconational identities. An ex ample of this from recent times would be the nature of the Iranian revolution in which Iran went from a state with more modernist and pro Western inclinations to a state that was essentially theocratic.

One thing that binds these different nationalist movements together is the intrinsic importance of terri tory. In most cases, nationalist movements arise from groups that already occupy (but perhaps do not control) a particular chunk of territory. One example would be the Basque movement found within northeastern Spain and southwestern France. While Basques occupy this area, they do not enjoy as much control as they might wish. Yet the definition of a Basque homeland is up for debate. Their task lies in choosing which territorial conception forms the basis of a new Basque nation state and the focus of their political goals. There are no obvious physical demarcations such as on an island nation. Quite the op posite since the Basque nation straddles a significant physical border in the Pyrénées Mountains, as well as an existing international border between France and Spain. Here the Basque spatial imagination has a lot to do with various narratives, rhetoric, symbols, and practicality. So the conceived territorial extent of the Basque nation has shifted from the former Kingdom of Navarre – a large entity that includes the Basque heartland, as well as much territory where Basques are not in the majority – to Euskal Herria, a limited territory with a more substantive Basque population. Euskal Herria has been institutional ized as well, since it includes territory demarcated by past administrative boundaries, including six historical Spanish *fueros* and three historical Basque provinces in France. The international boundary further complicates how Basques conceive of their own nationalist goals. Spanish Basques have been radicalized by the excesses of the Franco regime and by the strength of the dominant Basque political party. French Basques view their situ ation differently as part of a centralized French state.

Different segments of a single nation carry strikingly different ideas of what a nationalist project should entail. Within Israel, so called neo Zionists see the West Bank territories now occupied by Israel as an essential part of the Israeli spatial identity. In fact, some of these settlers would consider any pullback from such territory as tan tamount to a national dismemberment. For other Israelis, the vast majority, this territory is liable to negotiation. These post Zionists are concerned with the integrity of the Israeli state and see the state and nation as co terminous, but they are willing to concede some territory in order to secure greater degrees of security and peace. Does this mean that they espouse a nationalism that is not territorial? Not really, but they do offer a view of national territorial identity that is more permeable, and more flexible. The core identities placed in Israel proper are more important here than those identities forged at the borderland.

Boundaries of National Identity

In geography, there is a long legacy of research that looks into borderlands and of how borderlands can be con structed. Geographers were involved in Southeastern Europe after World War I carving a network of bound aries that would best reflect the nature of the region's ethnic geography. More recently geographers have fo cused on how boundaries can operate at a number of different levels. For example, we might think about boundaries as something that clearly distinguishes be tween those who are part of the nation and all outsiders. In that sense, boundaries become something of a wall. However, we also need to think about how these walls are built and how they are perceived. Boundaries are things that become a big part of what has been described as 'a discursive landscape'.

Consider boundary crossings. In some situations a crossing can be a major event, involving the presentation of official documents and perhaps even a series of back ground checks. In other situations, a boundary crossing may go unnoticed. The Schengen Agreement replaced a series of European border crossings with nothing more than signs, making travel between signatory states com pletely effortless. At the same time, the elimination of physical boundary crossings between various European states augmented the significance of the boundary sur rounding the European Union itself. Geopolitical cir cumstances have quite clearly changed the meaning of boundaries in this respect. A good example of how this might work at a more meso scale would be to consider the boundary between Finland and what now is Russia. This boundary straddles an area known as Karelia, found on both sides of the present day boundary and which contains a distinct regional group. When Finland was a part of the Russian Empire during most of the nineteenth century this boundary was fairly open. In 1939, the boundary was moved sharply west, forcing the inhabit ants to evacuate behind the new Finnish–Soviet line. For the next several decades, it became highly militarized and difficult to cross. According to Paasi, Finnish Karelia was a boundary region loaded with nostalgia for past unity and the utopianism of a lost future. Now, changing geopolitical circumstances have enabled the boundary to open up again to cooperative endeavors and to increased border crossings, and have enabled former residents to visit their ancestral towns and villages on the Russian side. Yet the boundary here now carries another meaning as it also represents the increased divisions marking off the relatively wealthy European Union from the much poorer countries on the outside.

Many nationalist goals focus around the boundary because this is the line that demarcates one state from another. However, it is more than simply a separation of political entities. It reflects the level to which nationalists have aspired to possess a certain extent of land. It also shows how nationalists seek to heighten the differences between their nation and those of others who live outside the boundary. This can take several different effects. Nationalists may attempt to enforce cultural homo geneity within the boundaries of areas that they control. So, for example, Mussolini inherited a larger Italian state after World War I – including areas that were not cul turally Italian – but he worked aggressively to nationalize the entire territory by forcing a series of name changes, language changes, place name changes, or even outright evictions in order to promote national uniformity.

Minority groups fight back by offering rival versions of the boundary. Irredentism, covered in another article in this encyclopedia, seeks to move the international boundary to allow minorities in one state to join up with their brethren in the state next door. Of course bound aries can also be places of accommodation and even cooperation. These are areas which reflect a kind of common cause between different national communities as demonstrated in the promotion of cross border crossings, institutions, and symbols of mutual interest. Within Europe a series of Regios or Euroregions have been es tablished around state borders which legitimize this level of cooperation. Outside of Europe, we see more informal aspects of cross border community. The border between the United States and Mexico, for example, has long been an area where the two national cultures seem to blend together. Evidence of Mexican culture is particularly evident in those areas along the Rio Grande River, which play host to American born Latinos (some who have occupied the area for generations), new Mexican immi grants, and the Anglo population. This 'MexAmerican' region blends ethnicity, culture, and even national iden tity (see **Figure 1**).

Landscape and Nationalism

Within core regions, nationalism shines most vividly through the presence of distinctive landscape types. Such landscape elements reflect the changing nature of na tionalism over time and the contested nature of particular nationalist symbols. Political geographers are always interested in how various attitudes, ideologies, thought processes are manifested spatially and tangibly. For in stance, Edensor has produced an exhaustive list of dif ferent national elements as present in British everyday life. Along with such grand structures such as London's Big Ben and the Tower Bridge, there are also elements like canals, fish and chip shops, hedgerows, and other elements that communicate Britishness in different and often changing ways.

A fascinating study of nationalist monuments in Moscow asks how nationalist monuments constructed in one political era are dealt with in subsequent political eras. How does post Soviet Russia handle monuments that were constructed during the Soviet period, and which in many cases were intended to legitimize and glorify a different and even derogated society? There are clear distinctions made between those monuments. Some glorify an aspect of nationalism that can be shared by all regardless of ideology. This would include many monu ments to World War II heroes (see **Figure 2**) and to the 'Great Patriotic War' that transcends ideological dis sonance. There are other monuments which have been disavowed, particularly those that speak to the glory of the Soviet economic system. And then there are those monuments which are contested: some groups, for ex ample, still honor Nikolai Lenin whereas he symbolizes a discredited system to others. A similar level of ambiva lence surrounds Japanese wartime monuments, especially the Yasukuni Shrine in Tokyo. Some Japanese still seek to glorify Japan's imperial past and to honor the soldiers

Figure 1 Charro Days, held annually in Brownsville, Texas has been held since 1938 and is intended to highlight the Mexican culture of the region. From http://www.pbase.com/lahuasteca/image/75245072.

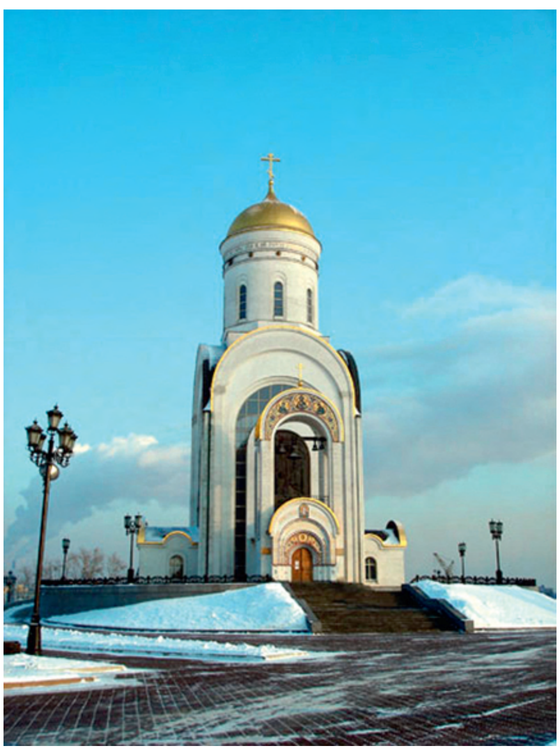

Figure 2 A photograph of the Victory Park monument in Moscow. Shows a monument to the Soviet effort in World War II that has been co-opted by the Russian and Moscow governments. From http://www.nterra.ru/index.php?page 1&lang eng&path mos id&id 46.

who fought during World War II. Others see the shrine as an example of rampant militarism that has no place in the ideology of modern Japan.

Landscape elements show up clearly between nation ally distinct portions of a particular state. Many countries in the world, while appearing in some respects as singular entities, are in fact divided between different national groups and the territories they possess. When this occurs, these divisions are manifested by differences in the land scape, and often these differences have a lot to do with the particular goals and aspirations of national groups.

The island country of Cyprus shows how this might work on the ground. While Cyprus has long been divided between Greek and Turkish speaking Cypriots, the sep aration before 1974 was more localized. In 1974, Cyprus was invaded by Turkey who carved out a distinct territory on the northern third of the island (see **Figure 3**). Greek Cypriots were forced to leave this territory, whereas Turkish Cypriots who lived south of the new partition line were enticed to move into much of the land vacated by the Greek Cypriot population in the northern third. Today the differences between these two areas are stark. For one thing, each section has been required to replicate insti tutions that at one point were shared by the entire country. Airports, roads, seaports, and other facilities are duplicated on both halves of the island. At the same time, landscape elements reflect separate nationalist objectives of Turkish and Greek Cypriots. For the Greek Cypriots their national goal is simply a reunification of the island dominated by the Greek speaking population. They have not altered many

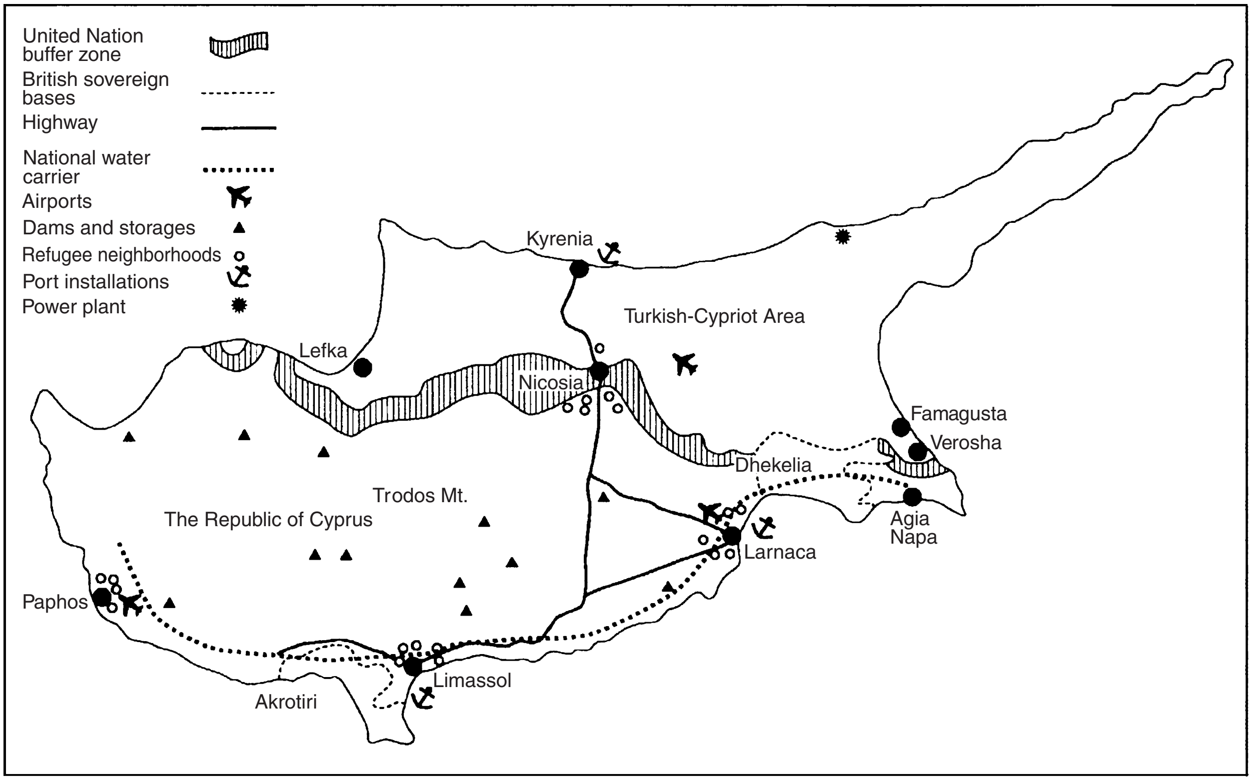

Figure 3 Map of the Cyprus partition from Kliot, N. and Mansfield, Y. (1997). The political landscape of partition: The case of Cyprus. *Political Geography* 16, 495 521.

aspects of the Turkish landscape, instead choosing to retain place names, monuments, and other types of elements that profess Cyprus's continued biculturalism. Within the Turkish half of the island however, all Greek evidence has been eradicated. The notion is that here we have a new society, a Turkish Cypriot society, that is ethnically homogeneous in every way. And the landscape reflects that.

Nationalism and Scale

Geographic scale is particularly relevant to the phe nomenon of national identity. National identities are nested to a considerable extent, such that a smaller identity may be contained within a larger identity and that identity may in turn be contained within a yet larger identity. This violates the political agenda of nationalism which demands absolute loyalty to a single national en tity at a single geographic scale. However, political geographers have come to recognize that identity and loyalty operate in more complicated ways.

Identities and loyalties at several spatial scales may operate harmoniously, but quite often they create tension and potential conflict. They may also be asymmetrical in the sense that the identity of one group may be at a different spatial scale than the identity of another group. The situation of Cyprus, mentioned above, is one example. Another example would be the 'double vision' in the spatial identities of East and West Germans at the height of the Cold War. Just as East Germany was seeking to reconstitute itself as a new, distinct nation, West Germany continued to adhere to an identity that in cluded the East German territory, as well as once German lands given over to Poland and the Soviet Union. This was not a case of aggression – eventual German unification was a peaceful process and all claims further east had ceased to be represented – but it does show the extent to which parts of a separated or partitioned country can evince starkly different spatial views.

We can also see this happening in the case of many substate nationalisms that emerge in the world. It is es timated that there are some 5000 putative nations. Yet, there are only 200 or so states. That means that there are many potential national movements that have not been successful. One result of this can be a national separatist movement, in which a substate national group pre dominates within a particular territory and seeks to form their own state. In the case of Québec, this distinction in identities and in nationalist agendas is clearly apparent. Within the last 40–50 years, the French Canadian population has become vested in the province of Québec itself. Before that, identity was much more fixed on re ligion and to some extent, language. Today, the province of Québec has assumed many of the functions of a sovereign state. However, outright separatism has been stymied by the presence of several groups who would rather see Québec continue as part of Canada. Along with the ever diminishing Anglophone population and the substantial immigrant population, this includes populations of indigenous peoples. Ironically, these indigenous peoples have argued that, should Québec secede from Canada, they would secede from Québec.

Such cases reflect challenges to nationalism emerging from a lower spatial scale. Then what of challenges that come from above? Globalization, for example, has led some scholars to describe the end of the nation state. In light of the recurring importance of nationalism in people's lives, this claim seems overblown. However, it is possible to conceive of how certain supranational entities that – while they may not supersede the significance of the nation state – have the capacity to fundamentally transform it. Certainly, empires played that role in past ages, but these prevailed mostly in a prenationalist era. The best example of this within the modern world would be the development of the European Union. The idea of Europe includes national and local loyalties, but also promotes an overarching sense of European identity. While the European Union continues to grow and to assume more authority, it is still true that a majority of people in most member states feel far more identity with their country than with Europe as a whole. Yet the idea of Europe will continue to present at least a friendly challenge to the national identities and nationalisms contained within it. It may even facilitate the expression of a variety of substate nationalisms, which can flourish in a time when national sovereignty is weakened and when the European Union as a whole provides a protective umbrella for all manner of small nation regions.

Currency is a good way to reflect national sentiments, since it often presents the types of icons and symbols that the entire nation can feel proud of. Money puts a country in its best light. It is also part of what has been described as 'banal nationalism'. So what of the euro, a symbol of nations agreeing to sublimate their monetary symbol to a higher organization? While the euro became the primary currency for most members of the European Union in 1999, it was only in 2002 that the euro entered everyday life in physical coins and bills. All members of the European Monetary Union exchange the same currency. However, the insignia on the individual coins and notes is customized for individual countries. While all euro coins and notes are able to circulate among all member states of the European Monetary Union, within each country most of the currency tends to bear the national insignia. In this way, the balance between supranational European identity and individual national identities are neatly manifested in a very common everyday object.

Conclusion

Discussions of nationalism often involve debates regarding the nature of nationalism. Questions relate to defining nations and nationalism and whether nationalism emerged long before the present age or is a relatively new phenomenon. Categorizations of nationalism, from questions of a civic nationalism, to an ethnic nationalism, to the more pernicious forms of nationalism representing some of the worst excesses of our era, are also explored. An additional issue involves whether nationalism is a peculiarly 'Western' phenomenon that arose in Europe and which was then imposed on the rest of the world, or whether indigenous forms of nationalism outside of the West may be identified.

These debates must be understood by geographers seeking to investigate nationalism and the websites listed below should help provide a good start to this broader investigation. This article focuses more on the particular contributions and concerns of political geographers – and these tend to revolve around how nationalism conceives of and manipulates space.

Space is a vital aspect of nationalism, but is also quite problematic. This contrasts with the spatial nature of a political state. Political states are fixed, at least at a particular moment in time. A boundary is neatly marked and crossing is made difficult or easy depending on the political relations between the adjoining countries. There is usually a designated capital where the state apparatus is contained.

In the case of nationalism there is no clear delimitation. National conceptions may be coterminous with the extent of the state – and nationalists may act as if the matter is settled – but nations are usually surrounded by much messier boundaries and contain land that is subject to a variety of different interpretations. The role of narratives, landscape elements, and such banal items as currency reflect these different conceptions. In fact, they are often introduced as a way to cement a preferred version of nationalism. Other symbols may be evoke a rival conception of nationalism, perhaps that of an unfulfilled national group. But because nationalism is something that stems ultimately inside peoples minds, its definition will always be fairly fluid. It has been the task of geography within the last 20 years or so to make sense of that fluidity and how it creates the national spaces we all must share.

See also: Borderlands; Critical Geopolitics; Diaspora; Fatherland/Homeland; Geopolitics; Governmentality; Irredentism; Landscape Iconography; Nation; National Spatialities; Political Boundaries; Scale; Sovereignty; Territory and Territoriality.

Further Reading

Beck, J. M. (2006). Geopolitical imaginations of the Basque Homeland. *Geopolitics* 11, 507 528.

Edensor, T. (2002). *National Identity, Popular Culture and Everyday Life*. Oxford: Berg.
Forest, B. and Johnson, J. (2002). Unraveling the threads of history: Soviet era monuments and post Soviet national identity and Moscow. *Annals of the Association of American Geographers* 92, 524 547.
Hakli, J. (2001). The politics of belonging: Complexities of identity in the Catalan borderland. *Geografiska Annaler B* 83B, 111 120.
Herb, G. (2004). Double vision: Territorial strategies in the construction of national identities in Germany, 1949 1979. *Annals of the Association of American Geographers* 94, 140 164.
Herb, G. and Kaplan, D. (eds.) (1999). *Nested Identities: Nationalism, Territory and Scale*. Lanham, MD: Rowman and Littlefield.
Kaiser, R. J. (2002). Homeland making and the territorialization of national identity. In Conversi, D. (ed.) *Ethnonationalism in the Contemporary World: Walker Connor and a Study of Nationalism*. London: Routledge.
Kaplan, D. (1994). Two nations in search of a state. *Annals of the Association of American Geographers* 84, 585 606.
Kedourie, E. (1960). *Nationalism*. London: Hutchinson.
Kepka, J. and Murphy, A. (2002). Euroregions in comparative perspective. In Kaplan, D. & Hakli, J. (eds.) *Boundaries and Place: European Borderlands in Geographical Context*, pp 50 67. Lanham, MD: Rowman and Littlefield.
Kliot, N. and Mansfield, Y. (1997). The political landscape of partition: The case of Cyprus. *Political Geography* 16, 495 521.
Newman, D. (2001). From national to post national territorial identities in Israel Palestine. *Geojournal* 53, 235 246.
Paasi, A. (1996). *Territories, Boundaries and Consciousness: The Changing Geographies of the Finnish Russian Border*. Chichester: Wiley.
Raento, P. (2004). Striking stories: A political geography of Euro coinage. *Political Geography* 23, 929 956.
Storey, D. (2001). *Territory: The Claiming of Space*. Harlow: Prentice Hall.
White, G. (2004). *Nation, State, and Territory: Origins, Evolutions, Relationships*, Vol. 1. Lanham, MD: Rowman and Littlefield.

Relevant Websites

http://www.fordham.edu
Internet Modern History Sourcebook: Nationalism, Fordham, University (some good links and references).
http://en.wikipedia.org
Nationalism, Wikipedia (a comprehensive definition).
http://www.tamilnation.org
Tamil Nation and Beyond (for an interesting justification of how nationalism applies to this one group).
http://www.nationalismproject.org
The Nationalism project (the primary website on the topic).
http://www.aag.org
The Online Center for Global Geography Education, Association of American Geographers (course module on nationalism).

Nationalism, Historical Geography of

M. A. Busteed, University of Manchester, Manchester, UK

Glossary

Cultural Hegemony Interrelated set of ideas and values generally accepted as being 'normal' and 'natural' in society.
Essentialism Belief that a group is defined by possession of exclusive immutable characteristics.
Hybridity Cultural milieu produced by interaction and reciprocal borrowing between contending groups.
Liminality Sense of being propelled into ethereal sensory group events above and beyond daily routine.
Othering Process of designating a group as characterized by totally negative characteristics.
Postcolonialism Impact of former colonial rule on culture and society of ex-colony.
Public Space Routes and places in urban areas open to use for both everyday purposes and ritual occasions.
Resistance Opposition to hegemonic discourse.
Social Construction Flexible self-definition devised by group at particular stage in development.
Subaltern Originally lower-ranking officer in armed services, appropriated for use in cultural studies to designate group rejecting hegemonic project.

Introduction

A nation may be defined as a group of people who believe they share a common past, certain cultural characteristics, a sense of differentiation from other groups, and a deep rooted attachment to a territory which they regard as exclusively theirs. Nationalism is the shared sense of belonging together. It is usually accompanied by a belief in the right of self determination, involving the control of the nation's territory and the apparatus of the state and the right to infuse state institutions with the values of the controlling national group, creating a nation state.

Approaches

Nationalism has probably been the most potent force in the making and remaking of the political map of the globe over the past 250 years. Some of the long established states of Western Europe were characterized by largely xenophobic forms of low level proto nationalism from late medieval times, but in its modern form nationalism originated in late eighteenth and early nineteenth century Western Europe as part of the Romantic Movement. The literature on nationalism is vast, derived from a great range of disciplines and shows no sign of diminishing. For a long time, discussion and analysis accepted the phenomenon in the terms adopted by nationalistic advocates. These argued their nation had possessed their distinctive identity and territory from earliest times, had survived through tribulation and triumph, sometimes experiencing a 'golden age' of cultural and political achievement, at others surviving great tribulation through implied providential intervention. Particular characteristics of the nation were highlighted as exclusive defining qualities. Sometimes these were aspects of 'national character', but quite often they were cultural features such as a distinctive language or religion, with dress, food, and folklore as supplementary defining features. This essentialist approach was notable in the political geography texts of the mid twentieth century, which frequently mapped language and religion and, comparing these spatial patterns with the territory occupied by the nation state, treated any mismatch as the root cause of territorial disputes, thereby implicitly accepting essentialist assumptions and arguments.

However, from the 'cultural turn' in so many of the humanities and social sciences in the 1980s and 1990s, there came a more flexible approach better able to accommodate the notably varied and dynamic nature of nations. Nations were now conceptualized as 'imagined communities' whose imagining is a social construction born out of the circumstances of time and place and capable of renewal and reinvention in response to societal change. This line of analysis meant that a historical approach was particularly appropriate and fruitful and it was used to deconstruct almost every aspect of nationalist projects. It is argued that a nation's history, so often presented as objective telling of factual events, was in fact a deliberately crafted nationalistic narrative transmitted to subsequent generations via the school system. These narratives were revealed as acts of creative self justification seeking to legitimize the nation by selectively constructing an historic, distinctive, and territorially rooted identity. Continuity with the past was used to imply legitimacy for the contemporary nation, its aspirations, and its current regime. To bolster the aura of historical respectability, authorities indulged in what is termed the 'invention of tradition' whereby public ceremonies were crafted to mark significant occasions. Events were devised utilizing allegedly ancient ritual, liturgy, motifs, and insignia, when in fact analysis revealed they were often based on relatively recent happenings. Thus, in the United Kingdom, the installation of the Prince of Wales in 1969 was supposedly based on an event of 1301, when in fact the model was the investiture of 1911.

Constructing the National Hegemony

This approach is well illustrated in the deconstruction of attitudes to those defined out of the nation. A xenophobic element is common to all nationalisms, possibly because national groups seem particularly sure of what characteristics they do not possess. This process of 'othering' helps with self identification as people differentiate themselves from their neighbors and are particularly vociferous in pointing up the contrasts with those they regard as historic enemies. Every possible vice of character, culture, and behavior is assigned to the great 'other', especially in wartime propaganda, and the consequent stereotyping enters deep and lingers long in popular culture. Nonetheless, it is argued that such contrasts are merely policy instruments appropriate to the current national interest. As interests alter, so also do the definitions of the 'other', and earlier enemies become friends and even allies. English and later British popular nationalism had long been defined around the alleged qualities flowing from Protestantism, particularly individual responsibility and freedom of judgment, speech and worship. This construct was presented as the antithesis of Roman Catholicism, which was designated as foreign, dictatorial, clerically dominated, and subversive of native British rights and freedoms. An accompanying narrative drew attention to long struggles against European Catholic powers, especially Spain, occasionally Austria, but particularly France and to the traitorous subversive intentions of native and nearby Catholics, as demonstrated by the Gunpowder Plot of 1605 and successive Irish Catholic rebellions, especially 1641. Commemorative annual celebrations, bonfires, and sermons and popular song and folklore reiterated the theme. But, by the late nineteenth century Catholics had been redefined into Britishness because the national interest had altered. Not only was France now evolving into an ally, but the argument goes on, there was now a project to present the growing British Empire as an inclusive institution under whose benign tutelage peoples from a vast range of ethnic, cultural, and religious backgrounds could find respectful protection and guidance and share the benefits of white, Christian British civilization and technology.

Challenging and Buttressing Hegemony

It is perhaps in analysis of the colonial and postcolonial situations that deconstruction of nationalist projects has proven most fruitful, particularly in the analysis of the role of cultural panic and the rise of reactive cultural nationalism. The impact of modernization following the imposition of colonial rule not only seriously disrupted indigenous patterns of economic, social and cultural geography, but also stimulated the growth of native middle class intelligentsia. Finding their upward mobility blocked by the prejudices and power structures of the occupying power, they frequently set about the creation of alternative structures based on indigenous features and values. In particular, they focused on those cultural features which they regarded as least contaminated by colonial rule and around them they constructed an alternative world outlook. The features in question were often characteristics of the oral popular culture of remote rural regions relatively untouched by urban commercial development but in danger of extinction. These were rescued, recorded in written form, often for the first time, popularized through voluntary grassroots organizations, and presented as the true cultural traditions of the people. It has been pointed out that in reality they had often been archaic features on the point of being abandoned and in many cases languages in particular such as Greek under the Ottoman Empire and Irish under the British administration in Ireland had to be modernized and standardized in grammar, vocabulary, and pronunciation in order to become usable. Many in such movements found cultural nationalism alone was not enough and made the transition into politics where they often provided the vanguard of political nationalist movements and the first generation of leaders in the newly independent nation state. They made their project of nationalistic cultural revival a defining feature of the new state, the policies of Irish governments after independence in 1922 in placing language revival and the preservation of the surviving Irish speaking districts being notable examples.

The establishment and defense of cultural hegemony are presented as a central project of all nation states, old and new. A key element is the creation and perpetuation of the myth of cultural homogeneity. In reality, all societies are characterized by internal cleavages based not only on cultural differences such as language and religion, but on age, gender, sexuality, region, and class. Anxiety about the potential for state disintegration, reinforced by administrative convenience, has led national governments not only to the definition and vigorous defense of 'national' characteristics, but constant efforts to imprint them on all the population, sometimes by explicitly banning alternative cultural motifs. In addition to a national education system, governments have also utilized national cartographic surveys to alter local place names, deliberately excising minority languages or sometimes accidentally bowdlerizing them as surveyors with little or no knowledge of the local language have misheard pronunciation of local nomenclature. Such a process occurred in the nineteenth century with the work of the British Ordnance Survey in the Scottish Highlands and the Gaelic speaking districts of Western Ireland. Study has also revealed the ways in which governments have also gone to great lengths to devise and display a range of national iconography, such as national flags, anthems, mottoes, slogans, and colors, as well as coins, currency notes, postage stamps, and official documentation bearing the images of rulers, heroes, and significant landscape scenes. Constant daily encounters with the display and circulation of such elements of what has been described as 'banal nationalism', nevertheless, renders their message 'normal', a triumph of mute but powerful symbolism.

Performance and Monument in Public Space

It is argued national governments do not rely on such subliminal messages alone. From the earliest days of nationalism the significance of public space and landscapes for transmission of the nation's message was realized. The central areas of large cities, particularly capitals, were laid out or refashioned in the grand style. Buildings housing the head of state or government, the legislature, judiciary, and significant ministries were laid out in grandiose and visually imposing fashion. The spaces and roadways thus created are utilized to present and perform selected aspects of national life. From quite early in human history, processions have been used to express group identity and sentiment. Originally probably religious in motivation, they were quickly adapted for secular nationalist purposes, with participants assembling at selected points, often dressed and adorned in distinctive fashion and carrying emblems of membership or office, moving along distinctive routes and converging on significant sites where carefully crafted ceremonies were performed. The constant use of the same points of assembly, routeways, and ceremony consecrated them into sacred sites and spaces, the repeated reenactment of the ceremony elevated it into sacred secular ritual and the constant repetition of wording transformed it into liturgy. The carefully structured sequence of movements and stillness, speech, music, and silences created a sense of liminality, but such events were held in the open air to engage public, if passive, participation, a process enormously boosted by the advent of national radio and television. A realization of the multisensory appeal of such occasions, it is claimed, explains why governments stage events deliberately designed to attract and hold attention by spectacle, the choreographed combination of color, movement, and sound, involving units of the armed forces and their equipment together with carefully selected 'representative' civilian elements, often acting out significant events in 'the national story' in tableau form, all calculated to stimulate anticipation, excitement, involvement, and national pride.

Studies have also demonstrated how governments adorn the landscape of public space with static monuments deliberately calculated to transmit nationalistic messages to generations of passersby, some of which also act as markers and rallying points for public ceremonial. Several dimensions of these artefacts have been analyzed. They can take a variety of forms based on mythical individuals, past rulers or heroes, anonymous representative individuals or groups, or they can be fashioned as plinths embellished with iconic symbols. They may be placed on or near significant routeways or in specially reserved spaces. They are constructed of enduring materials, often bronze or marble, designed not merely to withstand erosion by weather or traffic, but also to express the eternal character of the nation. Size and design are calculated to express the importance of the event or individual and to make them a visible center of attention not only to casual passersby but for the crowds during ceremonial occasions. The stance portrayed when sculpting individuals is literally larger than life to enhance visibility and the emotion conveyed varies from sorrow and suffering through dignified declamation or endurance to aggression or triumph. In colonial situations, architecture and statuary have emphasized the massive resources, military might, and technological superiority of the conqueror, and by implication the cultural, moral, and technological inferiority of the native population, points driven home by the publications of the burgeoning nineteenth century geographical societies in the metropolitan centers.

Subaltern Resistance

Study has emphasized how such celebratory artefacts and events rarely go unchallenged. Running parallel to almost every effort to establish and consolidate a cultural hegemony, there is a narrative of resistance mounted by subaltern groups presenting an alternative scenario which may be constructed around their particular nationalist outlook or issues of gender, class, sexuality, age, or the environment. Consequently, in many ways resistance has been presented as the obverse of the hegemonic social construction. Location is seen as a key dimension to resistance. In light of the powers of policing and surveillance exercised by the state, some subaltern groups choose to perform acts of resistance in peripheral spaces where state power is weakest. However, given the significance attached to public space by the hegemonic power, resistant groups frequently choose to express their alternative outlook in these same venues on the grounds that this will be a very public subversion of hegemonic pretensions, likely to attract maximum attention. These acts can take many forms. On occasion they can involve disruption of official ceremony as in the newly in dependent Irish Free State in the 1920s when the re publicans who had lost the civil war of 1921–22 over the treaty with Britain disrupted Armistice Day com memorations by the Irish who had served in the British forces in World War I. Such tactics invite forcible sup pression, though this in turn can generate useful propa ganda. The tactic of a mocking alternative aiming to subvert official events by ridicule and laughter has a long history. It has the advantage that the use of force in suppression can seem gross overreaction. Authorities traditionally responded by allowing such alternative events to proceed, treating them as useful and harmless ventilation of pressure, to the point where they became incorporated into the official events as carnival.

However, it has been argued, subaltern groups also respond by staging much more serious happenings which in organization, structure, and discipline and choice of date and venue are deliberately calculated as very public alternatives contesting the hegemonic calendar and nar rative and official use of public space. Thus in Dublin the main thoroughfare, Sackville Street (named after an eighteenth century Lord Lieutenant, the British royal representative in the country) saw the erection of a statue to the British naval hero Admiral Lord Nelson in 1809. Subsequently in 1864, the foundation stone for a monu ment to the Irish nationalist leader Daniel O'Connell was laid in the same street and eventually unveiled in 1882 after a vast procession which wound its way around various locations in the city associated with O'Connell's political career. By the early twentieth century the street was adorned with monuments to a range of nationalist and Catholic personalities. In other cases, the subaltern group may choose to trace out and adorn a totally con trasting alternative field of public spaces and sacred sites specific to their narrative. In South Africa the highly theatrical Afrikaaner reenactment of the Great Trek in the centenary year of 1937 culminated in the laying of the foundation stone of the massive Voortrekker Monu ment, finally unveiled in 1949, the year after the first Nationalist Party election victory.

Regime Change

It is perhaps in situations of regime change or shifts in ideology that radical changes in the national narrative are most notable. This may be signified by changes in the educational curricula, official iconography, the com memorative calendar, and public statuary. Monuments left over from the previous regime may be neglected, removed from prominent positions to obscure locations, or even destroyed, either by popular assault or official decision. These alterations to the monumental landscape are a notable feature of the postcolonial situation. In independent Ireland after 1922, statues to British mon archs and military men were vandalized and eventually removed quite quickly, though Nelson's pillar lingered in central Dublin until blown up by a reviving Irish Re publican Army (IRA) in 1966. The new regimes also proceed to make their narrative central to the curricu lum, install their distinctive commemorative calendar, and embellish the commemorative landscape with their monuments and statuary. Particular emphasis is given to the ideologists, events, and heroes of the struggle for independence. However, it has also been pointed out how even within the new dispensation there can be dissident groups who contest the discourse and mount their own events. Commemorations at the grave of the eighteenth century Irish republican hero Wolfe Tone are held each year though at different times, by groups from the Irish government, Sinn Fein and, latterly, dissident republicans disillusioned by the compromises flowing from the Northern Ireland peace process. In the late 1980s and early 1990s, there were notable examples of the re placement of communist inspired events and icons by the iconography of the newly installed national governments of Central and Eastern Europe. There was also a notable trend to rename public thoroughfares in honor of events and personalities significant in the national rather than the communist narrative.

Changing Worlds

The late twentieth and early twenty first centuries also brought challenges to some of the long established na tion states around the North Atlantic. Research has re vealed how they sought to respond to the challenges mounted by regional groups and sections of their citi zenry long unhappy with their political status and the ways in which their narrative had been excised from of ficial versions. In the United States moves toward the gradual incorporation of the Afro American population into full citizenship led to moves to create memorials to the process of emancipation in the form of museums, statues, and the renaming of streets. The process pro voked dissent and resistance within the black population itself, centered on the need to recognize the role of women in the civil rights movement and the extent to which the focus on race had obscured class issues.

On the other side of the North Atlantic several de velopments sparked rethinking on national identities. Former imperial powers had to cope with the fact that they had lost their imperial territories to nationalist in surgency and often did so by rewriting the historical narrative. The results could take diametrically opposite trajectories. In the case of Britain the transition from empire to commonwealth was presented as a painless process of calculated tutoring on democratic government, with the military campaigns against nationalist move ments excised. In Portugal by contrast the emphasis was on the gallantry displayed in the army's valiant struggle to retain the colonies, with little mention of the African role in either the Portuguese or the African nationalist forces. The indigenous peoples in North America and other territories long settled by Europeans also began to challenge hegemonic identity constructions.

In addition, the development of the European Union, the impact of economic and cultural globalization, and contrasts in regional material well being led to varying degrees of regional devolution in all the major powers of Western Europe. This in turn provoked questions about the nature of the regional identities in question and brought to the surface tensions within regional move ments which had previously been subsumed. In Northern Ireland after the partition of the island in 1920 the Unionist Party regime dominated public life until its collapse in 1972. For 50 years the official commemor ations and icons of the province were those of the almost exclusively Protestant and vehemently pro British Unionist majority, while those of the largely Catholic Irish nationalist minority were marginalized and in some cases declared illegal. As the province struggled toward a political settlement in the 1990s, increasing efforts were made to accommodate both cultural traditions. Among other things this involved finding more space for the Irish language and the nationalist narrative. This in turn in volved some restriction on the use of public space by Unionists, who found the adjustments hard to accept. Their sense of loss of authority, imminent threat from the Irish Republic and armed republican groups, and betrayal by British governments who did not appreciate their outlook led some of the more hardline loyalist elements to cast around for alternative constructions of identity involving a mélange of Ulster regional and local iden tities, class conflict, fragments of Celtic mythology, the Ulster role in the battles of World War I, and the lan guage of the early seventeenth century Protestant settlers.

Toward Hybridity

A recurring debate on the nature of nationalist constructs concerned the extent to which a truly subaltern voice could be heard. One school of thought argues the strength of the state power of cultural imposition and oppression and surveillance is so great that any acts of resistance are fleeting, minor, and of no long term con sequence. Alongside this it has been argued that the asymmetry in the power relationship between the hegemonic and the subaltern is so great that any act of resistance has to be expressed in ways that compromise the true message of the subaltern. Alternatively it can be argued that the asymmetry is more apparent than real, that panopticism is an illusion and that in every regime there are spaces for resistance which can be exploited by individuals and groups knowledgeable about local places and traditions. The result is interaction between the hegemonic and the subaltern which subtly alters both. The subaltern group is compelled to fashion its resistance in accordance with rules laid down by the hegemonic, and the hegemonic regime has been successfully chal lenged to accommodate resistance. It is asserted this process reveals that, like all constructs, both the hege monic regime and the alternatives are hybrid, constantly interacting and impacting on each other, renewing and redefining themselves as they adjust to political, social, and cultural changes. This argument, the obverse of es sentialism, states that hybridity is inherent in all social constructions, despite their perpetual efforts to present themselves as eternal and immutable. It proved particu larly fruitful in study of diasporic identities, crafted from a blend of elements from their country of origin and their new homeland.

Reflection

The shift of emphasis in studies of nationalism has cer tainly produced a much more flexible conceptual frame work, better suited to the great variety found among nations, the inherently dynamic nature of human society, their heterogeneity, and the perpetual interaction between human groups. It has clearly underlined the key role of change and thus the importance of the historical approach to the understanding of a sentiment which has moved large numbers to superhuman effort and ultimate sacrifice and repeatedly refashioned the political map of the globe.

See also: Ethnicity and Resistance, Historical Geographies of; Heritage and Culture; Heritage and Identity; National Spatialities; Nationalism.

Further Reading

Alderman, D. H. (2003). Street names and the scaling of memory: The politics of commemorating Martin Luther King jr. within the African American community. *Area* 35, 163 173.

Anderson, B. (1983). *Imagined communities: Reflections on the origin and spread of nationalism*. London: Verso. (Revied Edition 1991).

Busteed, M. (2005). Parading the green procession as subaltern resistance in Manchester in 1867. *Political Geography* 24, 903 933.

Colley, L. (1992). *Britons: Forging the nation 1707 1837*. New Haven, NJ: Yale University Press.

Crampton, A. (2001). The Voortrekker monument, the birth of apartheid and beyond. *Political Geography* 22, 221 246.

Gordon, D. A. and Osborne, B. S. (2004). Constructing national identity in Canada's capital, 1900 2000: Confederation Square and the

National War Memorial. *Journal of Historical Geography* 30, 618 642.
Graham, B. and Shirlow, P. (2002). The battle of the Somme and Ulster memory and identity. *Political Geography* 21, 881 904.
Hobsbawm, E. J. (1990). *Nations and nationalism since 1780: Programme, myth and reality*. Cambridge: Cambridge University Press.
Hobsbawm, E. J. and Ranger, T. (1983). *The invention of tradition*. Cambridge: Cambridge University Press.
Hutchinson, J. (1987). *The Gaelic revival and the creation of the Irish nation state*. London: Allen & Unwin.
Kong, L. and Yeoh, B. S. A. (1997). The construction of national identity through the production of ritual and spectacle: An analysis of National Day parades in Singapore. *Political Geography* 16, 213 239.
Jackson, P. and Penrose, J. (eds.) (1993). *Constructions of race, place and nation*. London: UCL Press.
Kearns, G. (2001). "Educate that holy hatred": Place, nationalism and identity in the Irish nationalism of John Mitchel. *Political Geography* 20, 885 911.
Light, D. (2004). Street names in Bucharest, 1990 1997: Exploring the modern historical geographies of post socialist change. *Journal of Historical Geography* 30, 154 172.
Pile, S. and Keith, M. (eds.) (1997). *Geographies of resistance*. London: Routledge.
Staun, M. S. (2000). The Paris geographical society constructs the other, 1821 1850. *Journal of Historical Geography* 26, 222 238.
Whelan, Y. (2003). *Reinventing modern Dublin: Streetscape, iconography and the politics of identity*. Dublin: University College Dublin Press.

Relevant Websites

http://www.nationalismproject.org
The Nationalism Project.

Natural Resources

G. Bridge, University of Manchester, Manchester, UK

Glossary

Adaptive Management A process of decision making in the face of uncertainty, characterized by the monitoring, evaluation, and adjustment of management actions to take account of feedback. Increasingly applied to flow resources such as fisheries and game species.

Ambient Resource A class of flow resource for which resource use degrades neither the amount nor quality of the resource (e.g., solar radiation or wave action). Nonexhaustible by definition.

Critical Zone Resource A class of flow resource for which use can exceed the regenerative capacity of the resource (e.g., groundwater or fish populations). The 'critical zone' refers to the threshold beyond which further use undermines the capacity of the resource to regenerate. Renewable but potentially exhaustible.

Flow Resource Also known as a renewable resource, refers to a component of the natural world that serves a socially valuable function and which has the capacity for regeneration. Current use need not diminish future availability. Subdivided into 'ambient' and 'critical zone' resources, based on the impact of use on regenerative capacity.

Resource Governance In general, the sociopolitical processes – administrative fiat, public participation, interest-group politics, and stakeholder consultation – through which decisions about natural resources are made. More specifically, the exercise of political–economic power in society through decisions about the design, manipulation, and control of natural resource systems.

Resource Imaginary A hegemonic discursive formation – combining social constructions of nature and various cultural norms around value – through which the natural world is viewed primarily in instrumental terms and appraised as a set of discrete, dissociable parts whose value is determined by society's wants. Frequently interpreted uncritically as a 'common-sense' view of nature and its relation to society, the particularity of this imaginary is seen by comparison with alternative ways of valuing the nonhuman world, such as those articulated by animals' rights associations, proponents of deep ecology, and some indigenous societies.

Stock Resource Also known as a nonrenewable resource, refers to a component of the natural world that serves a socially valuable function and which for all practical purposes has a finite supply. Current use diminishes future availability. Stock resources are exhaustible by definition.

Introduction

Natural resources is a deceptively peaceable term that finds widespread application within human geography. It describes products of biological, ecological, or geological processes that satisfy human wants. Examples include game species, soils, mineral ores, timber, and water. Classical political economy describes the raw materials and productive energies of the nonhuman world as gifts of nature, and contemporary accounts draw a strong distinction between resources and manufactured goods. Natural resources are not restricted to direct material inputs to economic life. They include a vast range of ecosystem services that are not directly consumed by use, but which are necessary for economic production and/or the maintenance of life (such as carbon sequestration, flood attenuation, and the maintenance of biodiversity). Many of these services provide important 'sink' functions by assimilating wastes produced during the use of environmental goods. In addition, there are other nonextractive ways in which physical environments and species can be considered to provide forms of utility for society (or parts thereof): recreational amenity, esthetic appreciation, and spiritual inspiration imply the attribution to the natural world of a complex range of value systems. Natural resources, then, is an inclusive category reflecting the different – and potentially conflicting – ways in which societies appraise the utility of the biophysical world.

Natural resources straddle two distinct epistemological traditions within geography. An extensive body of applied research on environmental evaluation and the management of water resources, rangelands, forests, and fisheries exemplifies geography's contributions to a tradition of producing instrumental forms of knowledge (such as those associated with regional planning). The primary purpose of knowledge in this managerial tradition is to understand environmental systems in ways that facilitate their management to deliver socially desirable goals. However, geography also has a robust tradition of critical inquiry which examines how ideas about nature are integral to the production of social differentiation and domination. And for this critical tradition, the notion of natural resources is problematic. The distinctions and differentiations enabled by the category 'resources' – between productive, valued assets, and unproductive 'wastes,' for example – play a key role in the organization of contemporary society and have their origins in the revolution of socionatural relations associated with the emergence of capitalism. From a critical perspective, the designation 'natural' obscures the significant ideological, political, and economic work that must be done to transform nature into a resource: as Hudson quipped "natural resources are not naturally resources." Geographical research, then, frequently illustrates how the capacity to claim nature as a resource is structured by access to technology, wealth, and social power.

Classification

A distinctive vocabulary is available for differentiating the qualities and properties of a vast range of natural resources. An elementary classification distinguishes between biotic (living) and abiotic (nonliving) resources: the former includes an enormous diversity of faunal and floral species capable of biological forms of reproduction; the latter includes an array of productive (water, minerals) and ambient resources (solar radiation), some of which – groundwater, beach gravels – have the capacity for regeneration by virtue of the physical systems of which they are part. Soil – the most extensive terrestrial resource – is a mixture of biotic and abiotic components. An equally basic classificatory scheme centers on the spatial extent of the resource: atmospheric nitrogen and oxygen, for example, occur everywhere on Earth and, accordingly, are considered ubiquities. Most other resources are constrained in their geographical range and frequency of occurrence and are localized to some degree. Localization is often a function of wide variations in resource quality: while aluminium is the second most abundant metallic element and occurs in clays the world over, it is found in commercial concentrations only in a small number of locations. Another popular axis of classification centers on the mobility of the resource and its ease of 'capture'. Land, water power sites, and coal deposits are examples of fixed resources. Groundwater, oil, game birds, and whales, by contrast, are fugitive resources whose mobilities resist conventional practices of enclosure and make it difficult to enforce exclusive ownership.

The distinction most commonly drawn, however, is between stock (nonrenewable) and flow (renewable) resources based on the rates at which biophysical materials regenerate (**Figure 1**). Stock resources exist in finite supply: since removing these resources diminishes the physical stock available for future use, all nonrenewable resources are exhaustible by definition. Stock resources are subdivided into those consumed by use (such as fuels) and those which, once extracted from the environment, may be recycled (most applications of metals). Resources that possess significant regenerative capacity are classified as flow resources or renewables. These may be subdivided according to whether or not the renewable resource is exhaustible. For so called ambient resources – such as solar radiation, wind, or waves – use of the resource degrades neither the amount of the resource nor its quality: this class of renewable resources is, therefore, nonexhaustible. Most flow resources, however, can be exhausted if consumption exceeds the rate of regeneration. This category of exhaustible renewables includes groundwater, fish, game species, and soil resources, all of which can be 'mined' to depletion/extinction if the rate of harvesting exceeds biophysical regeneration. Because a

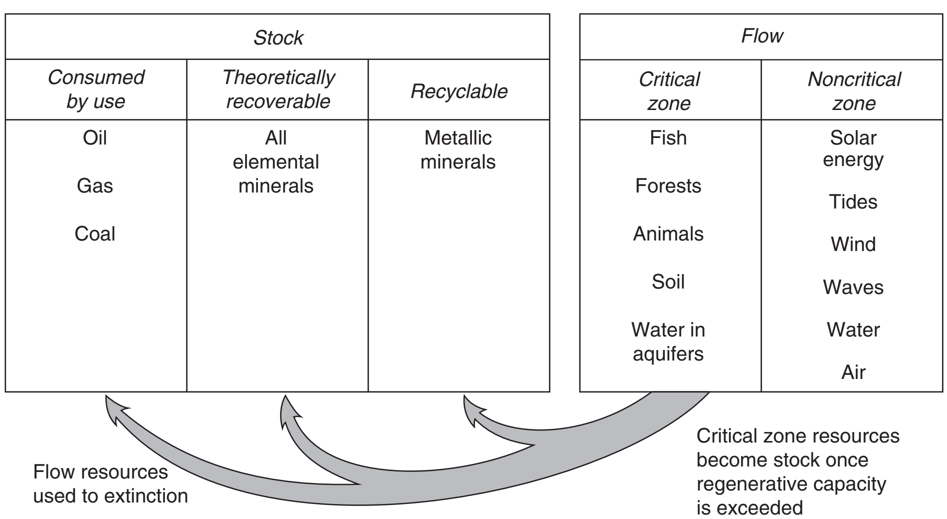

Figure 1 A basic classification of resources: Stock vs. flow. Source: Rees, J. (1991). Natural Resources: Allocation, Economics and Policy. London: Routledge.

threshold exists beyond which further consumption exceeds the capacity of the resource to regenerate, exhaustible renewables are also known as 'critical zone' resources. Many contemporary environmental challenges relate to the failure to manage critical zone resources effectively, and to maintain rates of extraction and use within the boundaries of the biophysical system.

In practice the distinction between renewable/nonrenewable is a matter of the timescale one uses to evaluate the regenerative capacities of biological and/or physical systems. Fossil fuel accumulations are currently being laid down in depositional settings around the world, yet oil and gas are considered stock resources because their generation occurs over extremely long timescales (millions of years) and is massively exceeded by the rate at which the resource is used. Timber, water, soils, and even some construction materials (sand and gravel) are capable of renewing themselves over more meaningful timescales, say 10–1000 years. For many biotic and abiotic resources, the question of their renewability is dependent not only on natural regeneration times but on levels of social investment. Studies of traditional agriculture, for example, show human labor playing a vital role in maintaining soil fertility, and contemporary agroindustrial practices invest large energy and financial resources (in the form of synthetic fertilizers, or pumping for irrigation or drainage) to maintain the utility of the soil resource and offset the effects of nutrient depletion, soil erosion, and soil degradation. Similarly, development of an extensive infrastructure for collecting and recycling aluminium would enable society to treat this metal more like a flow resource than a stock. For this reason, analysts prefer to think of a resource continuum which stretches from those nonrenewable stock resources which are consumed by use (such as fossil fuels) to the infinitely renewable ambient resources of solar energy, tidal power, and wind.

Cultural Appraisals

A geographical approach to resources recognizes resources as 'cultural appraisals' of the nonhuman world. This idea, which can be traced in the writings of nineteenth century political economists such as Ricardo and Mill, is articulated most clearly in American resource geographer Erich Zimmerman's phrase "resources are not; they become" (Zimmermann, 1933). Zimmerman argued against the vernacular view of resources as unchanging physical substances, and proposed instead that nature acquired the status of a resource only when society recognized it to have some functional value. Resources, then, are a dynamic social category. Components of the nonhuman world slip into (uranium, coltan, hydrogen) and out of (osiers, esparto grass, guano, sperm whales) this category over time and space. The remarkable history of natural gas (methane), for example, demonstrates how a single substance can be considered as variously hazardous waste, neutral stuff, or a valuable natural resource depending on knowledge, price, social norms and expectations (regarding pollution), and the availability of alternatives.

Resources occupy an uneasy – although analytically very productive – boundary zone between the conventional human and physical branches of geography. Like their close cousins natural hazards (negative resources), natural resources are hybrid forms which exist only at the interface between physical and social systems. Simultaneously social and physical, resources are socionatures that can just as easily be described by the terminologies and metrics of physical science as of social science: urban drinking water, for example, can be subject to chemical, physical, epidemiological, or sociological analysis. By understanding natural resources as cultural appraisals, and undertaking empirical research into the lived experiences of resource users, human geography has made significant contributions to three specific areas of analysis: resource scarcity, resource access, and knowledge.

Scarcity

Dependent on natural resources for food, materials, and fuel, all societies face the prospect of shortages or 'running out' from time to time. Regional, national, and planetary resource inventories have been the stock in trade of applied resource geography since the birth of the discipline. Despite the discipline's historic roots in environmental determinism, modern human geography has sought repeatedly to identify how demand for natural resources is related to the ways in which a society is organized. From Zimmerman to Harvey to Howitt, critical human geographers have adopted a relational approach to natural resources, illustrating how what comes to be considered as a resource is at least as much a matter of social, political, and economic conditions as it is any intrinsic or inherent properties. As a consequence, geographical research on the availability of food, minerals, and other resources has often taken issue with the neo Malthusian assumption that scarcity is an external physical condition setting the bounds of human possibility.

A relational approach to resource availability comes in 'weak' and 'strong' versions. The weak version highlights how resource availability rarely takes the form of an absolute physical limit, but is typically experienced as an economic constraint. The most meaningful measure of resource scarcity/availability, therefore, emerges as a function of knowledge and price. This relationship is captured for mineral resources by the McKelvy Box (developed by the United States Geological Survey).

The box's two axes frame a space – the resource base – characterized by an increasing uncertainty of geological knowledge in one direction and, in the other, a decreasing feasibility of economic recovery (see **Figure 2**). Mineral deposits for which there is a high degree of geological certainty and which are economically recoverable (those in the top left corner of the box) are distinguished from the remainder of the resource base by their designation as reserves. Technological or social changes that decrease the costs of recovery and/or that increase the certainty of geological knowledge can lead to resources being reappraised as reserves, effectively increasing availability. Crude oil reserves, for example, wax and wane over time depending on the price of oil and the prevalence of different production and consumption technologies. Sustained high oil prices create incentives for new exploration and production (which, all other things being equal, drives an increase in reserves) and also create incentives for demand reduction. Resource availability, then, has a dynamic quality that emerges from a society's capacity for technological innovation, behavioral change, and collective action. The weak version is common to the market response model which characterizes many so called Cornucopian refutations of physical resource constraints (such as those popularized by Lomborg). It explains the apparent paradox of 'increased' physical reserves for most mineral resources over the last century in the face of prodigious rises in mineral production and consumption.

The strong variant of a relational approach to resources also accepts that scarcity is an economic rather than a physical condition. Unlike the weak version, however, it examines the ways in which social inequality – that is, the internal differentiation of societies around, for example, household income or access to decision making – influences resource availability/scarcity. Time and again experience shows that rarely do the rich and powerful starve, go without fuel for cooking or heating, or have only limited access to clean water. Drawing on such observations and often working in the tradition of structural political economy, geographers have sought to denaturalize resource scarcity by examining how resource availability is mediated by wealth and power. The experience of drought in an industrialized, intensively water managed landscape, for example, is shown to be the result not of geographic or temporal variability in rainfall or even spectacular mismanagement, but as the outcome of (economic) priorities adopted by the companies and agencies charged with water supply. By locating the limits to resource availability internal to the economy (although not exclusively so), geographical research on resources is able to pry open a political space in which to discuss how apparent shortages might be resolved not by locating additional supply, but by changing the way resources are allocated within society, the uses to which they are put, and the social groups who influence critical resource decisions.

Access

Access to natural resources is a foundation for human welfare. Unprecedented economic growth and extensive social improvement experienced in many societies since the mid 1800s rest on a vast mobilization of materials and energy. For many people, lack of access to food, water, and basic materials remains a stubbornly brute fact, and one not restricted to the developing world (e.g., fuel poverty – where households spend over 10% of their income on heating – is a significant social issue in Europe). A focus on social access – as opposed to inventorying resource availability – highlights the social mechanisms that enable and constrain the ability of nations, groups, or individuals to derive benefit from resources. The significance of access is made particularly clear in settings like the Niger Delta where poverty coincides with spectacular resource abundance.

Cumulative production	Identified resources			Undiscovered resources	
	Demonstrated		Inferred	Probability range	
	Measured	Indicated		Hypothetical	Speculative
Economic	Reserves	Inferred reserves			
Marginally economic	Marginal reserves	Inferred marginal reserves			
Subeconomic	Demonstrated subeconomic reserves	Inferred subeconomic reserves			
Other resources	Includes nonconventional and low-grade materials				

Figure 2 The McKelvy box: Reserves vs. resources. Source: United States Geological Survey.

Resource access may be analyzed at several different spatial scales. Access and control can be considered primarily an issue of national resource security and geopolitical power. For example, from the end of the nineteenth century the growing commercial might of the US rested, in part, on its ability to negotiate favorable terms of access for US companies to the biological and mineralogical wealth of the Caribbean, Pacific, and South America (subsequently transformed into croplands, plantations, mines, and pastures). Interstate competition has ensured that national resource security remains a core component of international realpolitik, notably around access to oil and gas reserves, in negotiations between upstream and downstream riparian states, and in the challenges to 'environmental security' resulting from ecological degradation and social displacement.

Human geography, however, has made only modest contributions to research on interstate resource conflicts and the construction of international resource regimes. While acknowledging the importance of state policies to the administration of resources, geographical research focuses on the multiple spatialities and temporalities of resource conflicts that extend beyond those of the state, exploring a range of ways in which natural resources can become objects of political struggle. Contributions by Peluso and Watts to political ecology, for example, explicitly reject the reduction of resource and environmental security questions to issues of interstate competition. Through research on the lived experience of households and communities dependent on agricultural, forest, or water resources, geographical scholarship on resource access explores the social institutions through which people come to use resources and the implications of institutional change for livelihood security. Institutions of property feature strongly in this work because of the important role these formal and customary structures play in governing rights of ownership, access, and exclusion. Research by geographers, environmental anthropologists, and institutional economists has documented the diversity of property arrangements – private, state, open access and common property – through which renewable resources such as forests, fisheries, and grasslands are managed.

A common focal point for research on resource access institutions is Garrett Hardin's 'tragedy of the commons', a highly influential metaphor which has been used to assert that collective resource management will inevitably fail and that, as a consequence, private or state control of resources is preferable. Research by geographers and others challenges this assertion, and demonstrates that collective resource management regimes are widespread, are not restricted to developing countries, and can be resilient in the face of economic and environmental change. In practice the commons are not the open access, free for all regimes assumed by the tragedy of the commons: rather what distinguishes common property from open access regimes are the customs, rule systems, and enforcement mechanisms that condition individual access to and use of the resource. The past decade has seen renewed interest in the negotiated rule systems, indigenous environmental knowledges, and adaptive capacities that can characterize common property management regimes. Given the failure of many 'top down' development and conservation initiatives, the relative success of common property resource regimes has provided the intellectual justification for community based natural resource management, and for hybrid management arrangements involving state and local groups. Such co management regimes are now ubiquitous, their uptake driven partly by evidence of successful self management at small scales, the failures and exclusions of state management, and the problematic nature of local management when environmental systems and social drivers occupy translocal scales.

Knowledge

Although conflicts over natural resources frequently concern the politics of access and allocation, in many of the most intractable social struggles it is the practices through which lands and other socionatures come to be defined 'as resources' that are at stake. While chimpanzees (laboratory experimentation), exposed uplands (wind power sites), and tropical agro diversity (germplasm) are strikingly heterogeneous, what they share in common is that their status 'as resources' (noted in parentheses) is something achieved through cultural, economic, and political work. The apparent objectivity and universality of the 'resource imaginary' disguises its ethnocentric particularity, and its silencing of alternative conceptualizations of the nonhuman world. The definition of ancestral lands as a commercial mineral resource, for example, overlooks the preexisting land uses and land ownership that must be undone to establish mines and realize the exchange value of the ore. It is for this reason that a number of authors refer to the 'violence' of the resource imaginary, documenting the practices through which environments and resources are made intelligible to states and/or firms and ultimately are made subject to manipulation and control.

Located at the nexus of knowledge, practice, and authority, resources are a focal point for human geography's encounter with post colonialism. Research on knowledge, power, and resources has taken two slightly different paths in human geography. A significant body of work follows a Foucauldian route, tracing the moments through which knowledge of the biophysical world becomes the means for expanding administrative power. Another strand maps the stunning diversity of rationalities and cultural knowledge systems in which landscapes and species are enmeshed. This work indicates the range of attachments and meanings invested in natural re sources, the significant roles these attachments can play in the creation of cultural identities and subjectivities, and how the social value of socionatural systems can massively exceed that captured by utilitarian and in strumental modes of valuation. The implication of this work is that social struggles over resources are not always over a stable, shared category (e.g., in disputes over re source distribution or allocation) but often arise because components of the biophysical world are embedded in multiple knowledge systems.

Resource Governance

Social concerns regarding the use and management of natural resources have varied over time and space. Four major governance problematics – exploitation, conser vation, sustainability, and adaptive management – can be identified over the last 100 years. These are not mutu ally exclusive and contemporary resource management practices often seek to address several problematics simultaneously.

From Exploitation to Conservation

The acquisition and exploitation of natural resources is a problematic common to all societies. Historically re source exploitation is closely associated with practices of state formation (e.g., the centrality of oil exploitation to the growth of the Venezuelan state from the 1930s on ward) or the extension of state control (e.g., Khrushchev's 'Virgin Lands' campaign to plow up the northern steppes of Kazakhstan during the 1950s). Indeed, until the late nineteenth century the state's interest in resources lay almost exclusively in securing the supply of raw materials to the domestic economy through territorial means. State actions assisted primitive accumulation and extended monopoly control by defining and defending private and public property rights, subsidizing the costs of resource exploration by privateers, and deploying legal, political, and sometimes military means to deny access to com peting nations to critical or strategic resources. Terri torial development policies from the Russian Far East to the Indonesian archipelago continue to regard resource endowments as natural advantages to be exploited in the pursuit of growth and the reduction of regional dis parities. Policymakers in poor countries and economic ally deprived regions frequently describe investment in their agricultural, forestry, or mineral resource sectors as a key that will unlock national/regional potential and set in motion the virtuous cycle of socioeconomic change promised by post war modernization theory. Increased recognition of the challenges associated with resource based development – a complex set of arguments often reductively articulated as the 'resource curse thesis' – and vigorous critiques of modernization theory have not dimmed the enthusiasm for resource exploitation as a cornerstone of national development policy in much of the developing world.

In the early twentieth century the state's interest in natural resources expanded from exploitation and terri torial control to include a concern for orderly flow of resources into the economy over time. Marking this shift is the growth of scientific approaches toward the conser vation of resource stocks and the vesting of these new capacities for conservation in the administrative state. In much of Europe, the European colonial territories, and the United States this shift began to occur toward the end of the nineteenth century as a response to the massive appropriation of natural resources associated with indus trialization and colonial trade. The emergence of resource management as an administrative responsibility of the state marks an important historical point in the ecological modernization of the state. It acknowledges that resource scarcity or abundance are, at least in part, the outcome of social organization, and the beginning of efforts by the state to promote the production – via practices of hus bandry, stewardship, and conservation – of environmental goods and services that hitherto had been taken as free inputs (such as the fertility of soil, seed stocks, or game species). The elevation of resource management to a specialized branch of knowledge – what historian Samuel Hays terms an 'applied science of possibilities' – has multiple origins in the experiences of estate and colonial administrators struggling to work with the spatial and temporal variability of ecological systems. It is most clearly exemplified, however, by emergence of the Con servation Movement in the United States in the early twentieth century, a movement personified by Gifford Pinchot, the first professionally trained forester in the United States and first Chief of the US Forest Service. Conservation came to dominate thinking about resources during the twentieth century. Its emergence at the end of the nineteenth century coincided with the transition from an 'empty' world of expanding frontiers to a much fuller world in which there were limited possibilities for further territorial expansion via the acquisition of lands and re sources occupied by non Western peoples.

From Sustainability to Adaptive Management

Much academic and policy work on renewable and nonrenewable resources works within the conservation paradigm. However, the collective experience of com munities around the world with conservation has shown that the future oriented time preferences of conservation often conflict with more immediate time preferences of resource users. Prioritizing resource conservation can impact present patterns and rates of resource use to the point that, in some settings, livelihoods and life chances of groups dependent on access to the conserved resource are seriously compromised. Conflicts over implementing conservation raise questions of social justice (who) and equity (how much) that complicate conservation's pri mary concern with futurity and the time preferences of use (when). Such conflicts raise important democratic questions about for whom resources are being conserved and which social groups have a right to be consulted – or stronger yet, have rights which require their consent – on how resources should be used.

Conflicts over resource and environmental conser vation are experienced at a range of geographical scales, but it has been in the context of international develop ment that the call for an alternative agenda of conser vation with development – later to become sustainable development – has been most forcefully expressed. Facing calls from industrialized countries to reduce tro pical deforestation and modify resource and pollution intensive patterns of industrialization, developing coun tries sought to vigorously protect their access to national and international resources against charges of green imperialism. Popularized by the Brundtland Report of 1989 and the Earth Summit of 1992 (although formulated in a series of publications that followed the Brandt Report on global inequality in 1980), sustainable devel opment has provided a new set of concepts, metrics, and institutions through which resource governance is now imagined. Although the meaning of sustainable devel opment has been somewhat debased through overuse, in its original form it modifies the objectives of resource governance in three significant ways. First, it rejects postmaterial forms of environmentalism by insisting that for many of the world's poor the primary development challenge is not reducing resource use but expanding resource availability (clean water, food, cooking and heating fuel). Second, it imagines resource management responsibilities at geographical scales that range from the local community to the planetary (a geographical con sciousness institutionalized, for example, in the Agenda 21 planning framework). And third, because of its rec ognition of multiple scales of action, sustainable devel opment introduces a new cast of potential actors that extend both above and below those of the administrative state. Becasue of the way sustainable development di versifies actors and scales – encompassing community groups, multinational corporations, and supranational regulatory authorities alike – it has become a dominant problematic for thinking about resource governance in the contemporary period.

Developments in ecology over the past two decades have driven a reappraisal of many ecological models underpinning practices of conservation and sustainable development. Research on several commercial resources, most notably fish and game species, has demonstrated sharp fluctuations in population size that cannot be sat isfactorily explained by equilibrium based theories of conventional systems ecology. By showing the prevalence of disturbance events in ecological systems and the capacity of these systems for 'surprise', this work throws into question many of the assumptions of spatial and temporal regularity within conventional conversation and resource management models (such as those associ ated with management of predator–prey relationships, carrying capacity, or the assessment of biodiversity–stability relationships). New research emphasizing the importance of change, risk, and uncertainty in the man agement of resources has given rise to adaptive resource management as an alternative framework for the gov ernance of natural resources. Adaptive management ex plicitly recognizes uncertainty and seeks to optimize decision making around natural resources via an iterative process of intervention, monitoring, evaluation, and adaptation in the light of new information. Recent re search on the governance of a range of resources – from fisheries to global climate – emphasizes the possibilities for improving resilience and reducing vulnerability by designing responsive and flexible social mechanisms that allow adaptation as new information becomes available.

Critical Resource Geography: An Economic Geography Research Agenda

Anglophone economic geography has addressed a wide range of concerns in the post war period but has avoided a sustained engagement with natural resource questions. An earlier period of geographical inquiry had no such aversion: it inventoried, cataloged, and speculated about resources with an unswerving instrumentalism. The paradoxical legacies of this earlier period are a richly detailed set of resource descriptions, assessments, and maps that can provide a fascinating account of the evo lution of a global resource economy; and a heterodox tradition of critical inquiry – located on the margins of economic geography – which considers how 'resource geographies' are structured by economic, political, and social processes and, therefore, owe a debt to more than nature alone. These legacies provide the foundation for a 'critical resource geography' which engages questions of knowledge, scarcity, governance, and sustainability that are at the heart of modern human geography. Critical resource geography (1) accounts for the political, eco nomic, and cultural processes by which particular com ponents of the natural world become produced as resources, and subsequently proliferate through the economy in the form of commodities; and (2) examines how the biophysical and material properties of resources influence the social organization of their production.

Substantive changes are currently taking place in the scale, rate, and pattern of resource production and con sumption. From palm oil and tar sands to fisheries and carbon offsets, these changes raise important social, ecological, and political issues and call out for the at tention of a socially engaged economic geography. Understanding the geographies and histories of specific resources can be a productive place from which to start. Shifts in the scale and direction of resource flows, for example, can reveal the extent to which established re gional and global economic geographies are being re made. Similarly, analyses of energy, water, and food distribution at a range of scales can show how resource availability has less to do with the raw productivity of 'natural' systems and much more to do with the ways in which commodity systems are socially organized. Critical resource geography takes up the challenge of producing empirically grounded, theoretically informed accounts of emergent resource geographies and their social, eco nomic, and ecological implications. In doing so, it prob lematizes the treatment of 'nature' within modern economic geography, a field which has largely defined itself by bracketing out nature as an object of inquiry.

See also: Biodiversity; Conservation and Ecology; Environmental Hazards; Environmental Security; National Parks; Nature; Natures, Postcolonial; Political Ecology; Sustainable Development; Water Management.

Further Reading

Bakker, K. and Bridge, G. (2007). Resource regulation. In Cox, K., Low, M. & Robinson, J. (eds.) *Handbook of Political Geography*, pp 219 234. London: Sage.

Folke, C., Hahn, T., Olsson, P. and Norberg, J. (2005). Adaptive governance of social ecological systems. *Annual Review of Environment and Resources* 30, 441 473.

Harvey, D. (1974). Population, resources, and the ideology of science. *Economic Geography* 50(3), 256 277.

Hays, S. (1959). *Conservation and the Gospel of Efficiency The Progressive Conservation Movement 1898 1929*. Cambridge: Harvard University Press.

Howitt, R. (2001). *Rethinking Resource Management: Justice, Sustainability and Indigenous Peoples*. London: Routledge.

Hudson, R. (2001). *Producing Places*. New York: Guilford Press.

Lomborg, B. (2001). *The Skeptical Environmentalist*. Cambridge: Cambridge University Press.

Peet, R. and Watts, M. (eds.) (2004). *Liberation Ecologies: Environment, Development, Social Movements*. London and New York: Routledge.

Peluso, N. and Watts, M. (eds.) (2001). *Violent Environments*. Ithaca, NY: Cornell University Press.

Rees, J. (1991). *Natural Resources: Allocation, Economics and Policy*. London: Routledge.

Swyngedouw, E. (2004). *Social Power and the Urbanization of Water: Flows of Power*. Oxford: Oxford University Press.

Walters, C. (1986). *Adaptive Management of Renewable Resources*. New York: Macmillan.

White, R. (1996). *The Organic Machine: The Remaking of the Columbia River*. New York: Hill and Wang.

Zimmerer, K. (1994). Human geography and the 'new ecology': The prospect and promise of integration. *Annals of the Association of American Geographers* 84(1), 108 125.

Zimmermann, E. (1933). *World Resources and Industries*. New York: Harper Row.

Relevant Websites

http://www.atlas.aaas.org
American Association for the Advancement of Science, Atlas of Population and Environment.

http://www.millenniumassessment.org
Millennium Ecosystem Assessment.

http://www.resalliance.org
Resilience Alliance.

http://www.rff.org
Resources for the Future.

http://minerals.usgs.gov/minerals/pubs/myb.html
United States Geological Survey Minerals Yearbook, USGS.

http://www.wri.org
World Resources Institute.

http://www.worldwatch.org
World Watch Institute.

Naturalistic Testing

R. D. Jacobson, University of Calgary, Calgary, AB, Canada

Glossary

Case Study An approach where the researcher explores a single entity, 'the case', which is bounded in time and activity. Examples are an event, a process, or a particular social group. Detailed information is collected over a sustained period of time about the case using a variety of qualitative data collection procedures.

External Validity It refers to the generalizability of the research findings. To what degree are the conclusions from one study that are based on a specific population, context, settings, and variables extendable to different contexts, and to additional populations? To what extent can the results obtained from studies be expanded beyond the studies' boundaries?

Grounded Theory A theoretical and methodological research approach, which places a central emphasis upon the data; the construction of any theory is grounded upon the data from which it is derived. In this context any knowledge is generated through inductive reasoning.

Ideographic Research concerned with the unique and the particular, such as the study of a single individual case, or more broadly, a series of case studies.

Internal Validity It is the extent to which variations in an outcome (dependent) variable can be attributed to controlled variation in an independent variable. A causal connection between independent and dependent variables is usually assumed.

Methodological Pluralism The incorporation of multiple methodological approaches to provide a form of data triangulation from diverse methods. These methods may have different ontological implications and foundations; however, by their integrated use they offer a more legitimate, reliable, and valid approach to research.

Natural Experiment Research undertaken in as close to a natural or undisturbed setting, without intervention from the researcher in a scenario uncontrolled by the researcher, but believed to be sufficiently powerful as to reveal the causal relationships between the researched event and changes in outcomes associated with the event.

Phenomenological Studies A qualitative method of research involving the examination of human experience through the detailed description of the people being studied. The methods involve studying a small number of individuals through extensive human interaction to develop patterns and relationships of meaning. As the research seeks to understand human experiences, the approach produces a descriptive synthesis narrative of the phenomena under study.

Reliability It is the consistency of a set of measurements; reliability does not necessarily imply validity. In experimental settings, it is the extent to which the measurements of a test remain consistent over repeated tests of the same subject under identical conditions.

Thick Description A term referring to interpretation, that is based upon the researcher being immersed into the setting to uncover the different meanings, and then being able to ascribe the interpretations and interactions of these meanings into a rich, constructive, and layered manner.

Introduction

Naturalistic inquiry or testing aims to leverage the benefits of conducting research in a natural setting in order to provide a rigorous contextual evaluation of the problem or phenomena under research scrutiny. It therefore is a predominantly qualitative research methodology. This is in contrast with controlled experiment inquiry in which the researcher manipulates the independent variables with some explicit control over other factors in order to observe the effects on the dependent variables.

As the meaning derived from research is closely related to the research methods and context, naturalistic testing occurs in a 'natural setting', that demands that the research be mediated or undertaken by a researcher that is flexible and adaptable to the indeterminate nature of the research. In this manner the researcher is working 'on the fly' building upon their tacit knowledge through a wide range of qualitative techniques. These would include the wide range of qualitative analysis techniques available to a human geographer. A sample list in the broadest sense would include, but is not limited to, interviews, observations, fieldwork, ethnography, case studies, content and document analysis, and so forth. In the most classic and pure scenario of naturalistic inquiry the phenomena under investigation are sampled in a purposeful manner, in order to negotiate the problems of random sampling; the data collected are then analyzed in an inductive manner, leading to the development of a grounded theory. From this, a projection of the next steps in the research is proposed in an emergent design. The sampling, analysis, grounded theory, and emergent design process loop are then repeated through several iterations, until a theory becomes stable, and a degree of redundancy is achieved. Clearly this process is bounded by the practical resource constraints imposed upon the research, such as time. The outcomes of the research are negotiated through a series of checks with respondents who have acted as sources of information. Different patterns emerging from the data are reconciled through this process. In the classic naturalistic model the information is then gathered to provide a case study report. The results are therefore primarily an interpretation of the idiographic information. These results may be tentatively applied or extended beyond the specifics of the case study through a careful consideration of credibility, transferability, dependability, and confirmability. These terms are the naturalistic analogs to the more conventional criteria of internal and external validity, reliability, and objectivity.

Context

Naturalistic inquiry emerged as an alternative approach to the very controlled characteristics of empirical positivist based methodologies. Naturalistic testing could be considered at one end of a continuum, where the other end is a highly controlled clinical style experiment. The rise of naturalistic inquiry has been considered by some researchers as a significant paradigm shift from positivist based quantitative inquiry. In modern positivist, logical, and other similar systems of inquiry, experience is taken to be objective, testable, and independent of theoretical explanation. Serious artificial constructions lead to models that yield explanation in the sense of logic. In natural science this assumes that the objects and the investigator are external, and not related. In other words, the language that is used within these studies is exact, universal, and generalizable. Little meaning is ascribed to local anomalies as these may only generate problems, when the application of universal categories is applied to a specific or particular circumstance. By contrast, naturalistic inquiry seeks to focus on how and why, within the context where, the data are gathered from within the setting rather than as an external controlled set of research events.

Philosophy of Naturalistic Inquiry

The philosophy underpinning naturalistic inquiry is one of moving away from positivist natural science methodologies to axioms that better fit the phenomenon of sociobehavioral research. Although the 'naturalist paradigm' emerged from educational and sociological researchers, it has analogous properties to postmodern and post structuralist advances in geographic theory.

The nature of reality (ontology) is assumed that there are many multiple constructed realities, multiple truths, and perspectives that can only be studied holistically in an effort to seek out understanding or to make enquiries about this reality, or these multiple realities as outcomes will inevitably diverge. In other words, each inquiry potentially can raise more questions than it is able to answer, although a level of understanding is achieved. The predictability and control of these realities are difficult. This departs from a traditional positivist view that there is a tangible reality 'out there', and that this singular reality can be subdivided into independent variables and processes. These fragments of reality can be studied independently of each other, and that inquiry can converge toward this reality.

Within positivist epistemology, the relationship of the knower to the known, the researcher, and the object of inquiry are independent. However, the naturalistic framework has the implicit assumption that the enquirer and the object of the inquiry interact to influence one another. In this way there is a relationship, they are inseparable, and there is not the discreet dualism implicit in controlled studies.

The aim of inquiry is to develop a geographic body of knowledge. In the naturalistic mode this is done in order to develop an idiographic body of knowledge that form and inform a series of working hypotheses that describe a specific context. This is the key issue and problem with naturalistic research: while there is meaning and understanding gained at the individual level, the generalization, applicability, and transferability of this research to a more general population or wider context is problematic. By contrast in positivistic inquiry, the aim is to generate a nomothetic body of knowledge, that are highly generalizable, and that are independent of time and context. In a controlled experiment the generalizability is implied by the controls established. However, the data gathered are often artificial, and cannot be reinserted into the holistic environment from which it was isolated.

Within naturalistic research the cause and effect are not always discretely separatable as there are mutual interactions between the researcher and their study.

Within the inquiry, both the research and those being researched are in a state of mutual collaboration and simultaneous shaping and reshaping so it becomes impossible and complicated to distinguish and isolate causes from their effects.

Inquiry is not viewed as value free. Its axiology is considered to be value bound in at least four ways: (1) the nature of the inquiry is influenced by the researchers' values, such as through the choice of the subject matter; (2) research is clearly influenced by the choice of framework and paradigm that guides the investigation into the problem; (3) the research is influenced by the choice of any background or substantive theory utilized to collect the data and then subsequently analyze the data and later interpret the findings; and (4) the content, methods, and context of the inquiry are all value laden. Any research is influenced by the values and experiences of the researcher, and particularly how these are exam ined in the context of the research setting.

Components of Naturalistic Inquiry

Outlined below are the key components of naturalist inquiry as proposed by Lincoln and Guba. The justifi cation of these components is based upon both their coherence and interdependence and also their logical dependence upon the axioms outlined in the philosophy section above.

The natural setting is a prerequisite for the re searchers' experimental methodology. This is necessary because the naturalistic ontology suggests that wholes cannot be understood in isolation from their context and that they can't be separated to be studied in their con stituent parts, because the whole is more than the sum of the parts. There is also the belief that the act of observing and/or recording of data involve interacting with the subject under investigation and have an effect on the data being collected. Undertaking research in this context means that it is important for determining the fullest understanding of the entity. But this also has a limitation as to the transferability of this finding to other contexts.

The researcher is acknowledged as an active partici pant in gathering knowledge and experimental data. The use of the qualitative instrument is necessary because the human is capable of grasping and evaluating the mean ings of mutually shaping subtlety that can be only ap preciated by an individual. In other words, the individual interactions are value based, and only an individual can take account of these values and the resultant biases.

The use of tacit knowledge, that which is intuitive and felt, as well as propositional knowledge in the form of language is assumed. Because much of the interaction between the investigator and the respondent, participant, subject, or object of inquiry happens at this level, the nuances of multiple realities can only be appreciated through this method and tacit knowledge also acknow ledges the mirroring and the mimicking of the patterns of the investigator.

Qualitative methods are often used because of their ability to deal with multiple realities. However, multiple methodologies including triangulation are encouraged where approaches using combinations of qualitative and quantitative methodologies are a mechanism of re inforcing and shaping the validity of any methodology.

Purposeful sampling used as representational or ran dom sampling may exclude more deviant or outlying cases. By using purposeful sampling, a greater array of overall realities may be uncovered. In addition, grounded theory can be developed that takes adequate account of specific experiments on local conditions. The context and the values thus improve the options for transferability, as local conditions have been acknowledged.

By using inductive data analysis, the investigator re spondent or object relationship interaction becomes ex plicit, recognizable, and accountable, and the values of these interactions can become an explicit part of the analytical structure, as biases can be explicitly acknow ledged, and negotiated.

Grounded theory is the guiding theory because it is grounded in the data; it provides the best fit with the data under investigation. In other words, there is no primary theory that could possibly encompass all of the potential multiple realities. The experimenter enters the situation as neutrally as possible, and drives his/her theory from close analysis of his/her data and emergent patterns.

The design of the research is expected to emerge, in other words, to flow and cascade rather than to be arti ficially constructed ahead of time. It is assumed that it would be inconceivable for the experimenter to know ahead of time, all of the values and scenarios that would be encountered and to be able to foresee the mutual shaping, and to see from the subject to what may happen throughout the experiments and the subsequent outcomes.

The outcomes of the research are intended to be 'ne gotiated'. In other words, it is the human subject con structions of reality that the enquirer seeks to reconstruct. Thus, the validity and the outcome of the inquiry depend on the nature of the quality of the interaction between the known and unknown, in the meaning and construction of the data. The tenet here is that the working hypotheses that might apply in a specific given context, are best to be verified and confirmed by the people who exist and live in that context.

The case study reporting mode is often used using thick description, as it provides a basis for individual naturalistic generalizations. This allows for a holistic picture of the values and positions, and the relationships of the investigator. The substantive theory of the natur alistic methodological paradigm has to acknowledge the local contextual values.

Data are generally interpreted ideographically, that is, in terms of the particulars of the situation rather than nomothetically, in terms of law like generalizations, as the interpretations of the data depend upon the local context for their validity. Applications beyond the specific contexts of the study are likely to be only tentative, due to the mixture of mutually shaping influences and dif ferences from location to location, different value sys tems, and different site specific information. As a result, the boundaries of the research are usually set in some mediated way between the investigator and the focus of the research.

Alternatives to the traditional measures of validity are sought; these include credibility, transferability, dependability, and confirmability. Trustworthiness seeks to address issues of truth, applicability, consistency, and neutrality. Overall the research conducted within a naturalistic framework aims to provide an emergent paradigm whereby understanding, interaction, and analyses move from simple to complex, from a causality that is linear to mutual, and from objective to perspective.

Conducting Naturalistic Inquiry

The process of designing a naturalistic inquiry, as with any research design involves planning the structure and strategy of the investigation, structured in a certain way as to obtain answers to the relevant research questions and to minimize and control variables. This is a multistage process involving the conception and overall plan of the study, the hypotheses that will be investigated, the variables to be included, the relationships expected to be found among those variables – this is equivalent to formally setting hypotheses – the methods and instruments of data collection, and the modes and mechanisms for data analysis. The researchers operating under this framework often prepare contingency and alternative plans to try and anticipate any emergent problems within the research design.

An example of naturalistic inquiry is in a multimethod approach to investigating the understanding of geographic space by people without the use of vision. In this context, data were gathered using a mixture of qualitative and quantitative methodologies in both artificial and natural(istic) settings. As the overall aim of the research was to develop a broad understanding of how people without sight can navigate through space in their everyday environments, it made conceptual and methodological sense to investigate how individuals were able to negotiate traveling through their environment, by conducting research in the urban setting through which people routinely travel. The research would provide a more realistic and complete picture as to the salient problems, issues, and abilities of the individuals concerned. However, the transferability and replicability of these findings are limited by the context, and data collection methods. Questions would arise: How representative of the whole population was the sample used? Were the problems experienced by this group representative of problems in other areas, or were they specific to this particular locale? In terms of measuring people's ability to learn a route through using multiple methodologies involving several trials, working in natural and experimental settings, including, observing peoples ability to renavigate the route, to provide distance estimation and to point to locations along the route, to construct maps and models of the route, and to provide descriptions of the route of these results, a very comprehensive dataset was collected. However, the data collection could be critiqued by its very collection, that is, was compromised, or contaminated by a lack of control in the natural setting. For example, the participants were able to augment their navigation skills with additional information such as passing traffic, wind, and so on, that may vary from one user to another. The strength of using this method was not to isolate in an artificial environment, but to obtain a realistic and holistic picture of the abilities of people to undertake these navigation and related tasks in a natural environment. Clearly, blind and visually impaired people need to navigate to points of employment, education, and other locations for shopping and social activities, rather than between points in a highly controlled laboratory setting up. The information gained from the highly controlled laboratory setting, such as an open space like a gymnasium whereby individuals are studied in great detail, including reaction times and accuracy in navigating between points, provides a different level of information that is also valuable and this information is at the other end of the continuum from naturalistic testing.

One of the strengths of a naturalistic method is in the acknowledgement of these different positions and the interactions between the investigator and the participants. This is not viewed as an intrusion, or a weakness, but leading to more of an opportunity that can be explored and exploited to further explore the contextual reality of any given experimental situation.

Critiques

The most significant critique of a naturalistic testing appears through its use of inductive reasoning, relating to how realistic and reliable are the inferences made. Even if researchers can't validate knowledge in an *a priori* matter, it is however nontrivial and significant that it can be validated at all by any methodology.

There are a set of theoretical and practical problems that have been identified with induction and inductive reasoning that occur at a philosophical and at a methodological level.

Classic problems with induction arise with seeking data that support and will confirm a hypothesis. Universal statements, such as 'all crows are black', could never be conclusively verified. However, with the discovery of one nonblack crow it could be conclusively refuted. Thus, falsification has been introduced as a mechanism for avoiding the problems with induction. However, when dealing with research hypotheses they are not treated in isolation from a background theory, their social, political, and contextual settings.

Researchers employing naturalistic inquiry have to demonstrate the rigor of their analysis and their data collection methods. However, in response to these criticisms, they frequently contend that the widely used hypothesis testing in conventional scientific method is not itself free from criticism and has problems as no hypothesis is tested in isolation. There are other matters undergoing investigation besides the hypothesis itself, for example, the reliability of the testing, whether the hypothesis has been derived in a valid manner from the relevant theory, and whether the assumptions of the theory themselves were valid. Therefore, the falsification of a hypothesis may not be due to the weakness of the hypothesis that it uses, but due to other matters. In addition scholars seldom seek to test hypotheses with the sole intention of falsifying them. If that were so, if the failure to confirm a hypothesis were grounds for rejecting it, then nearly all theories in science would be rejected because at some point, they can be falsified by some observational data. Therefore, in practice, researchers tend to accept the theory that has the greatest explanatory potential, in combination with the fewest problems.

Scholars studying the history and philosophy of science have commented that scientists may not choose paradigms on the basis of data alone. There is a significant contribution from the underlying social and political forces, the context and historical analysis that lead to these changes in an ecological analysis of the relationship between theories and evidence.

Place of Naturalistic Testing in Human Geography

Frequently, human geographers have commented that falsification in this abstract hypotheses testing is more suited to the physical sciences, because of the nature of quantitative observations, measurements, data, and methods available to them. Naturalistic testing provides the human geographer with a paradigm and context within which to conduct research in natural settings involving complex interactions between humans, society, and their environment. It fits within the recent debates about how geography is practiced, the ethics of doing research, and a framework for establishing and clarifying who the research is on, by, for, or with. For example, it provides, some resolutions to the problems associated with ethnography and participant observation as they relate to a colonial anthropological legacy, and the issues around the power and gaze of the researcher, by providing a mechanism for the acknowledgement of this power, and methodology for giving voice to the focus of the research. As naturalistic testing relies heavily on the natural environment and the acknowledgement of the context research, there is a clear effort to understand the relationship between the observer and the observed on how these are overlapped and intertwined. This is highly relevant to debates in modern social theory and social philosophy in geographic methods. The issues revolve around power, space, and subjectivity, and the power–knowledge epistemology put forward by Focault. Naturalist methodologies contribute to the idea that all knowledge is to a degree situated. Further refinement and sensitivity to context suggests that all theories are to a degree limited and partial, and have emerged from their own context and circumstances from which they respond and are researched. Naturalistic methodologies have a way to deal with negotiating admissible knowledge, providing opportunities for improving the coherence of data, and integrating the political and methodological by avoiding the search for natural laws through the acknowledgement of the subjectivity of the research. Stringent efforts are made to complete the research in a rigorous manner as possible through multiple and redundant mechanisms for validating and triangulating, approaches, data, and analyses.

See also: Content Analysis; Cross-Cultural Research; Diaries (Video, Audio or Written); Discourse Analysis; Ethical Issues in Research; Ethnography; Experimental Design; Fieldwork; Focus Groups; Grounded Theory; Hypothesis Testing; Mixed and Multiple Methods; Participant Observation; Participatory Action Research; Questionnaire Survey; Situated Knowledge, Reflexivity; Subjectivity.

Further Reading

Alexander, P. A. and Winne, P. H. (2006). *Handbook of Educational Psychology*. Mahwah, NJ: Erlbaum.

Baxter, J. and Eyles, J. (1997). Evaluating qualitative research in social geography: Establishing 'rigour' in interview analysis. *Transactions of the Institute of British Geographers* 22(4), 505 525.

Ericsson, K. A. (2006). *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge: Cambridge University Press.

Erlandson, D. A. (1993). *Doing Naturalistic Inquiry: A Guide to Methods*. Newbury Park, CA: Sage.

Eyles, J. and Smith, D. M. (1988). *Qualitative Methods in Human Geography*. Cambridge: Polity.

Golledge, R. G., Jacobson, R. D., Kitchin, R. M. and Blades, M. (2000). Cognitive maps, spatial abilities, and human wayfinding. *Geographical Review of Japan, Series B* 73(2), 93 104.

Hay, I. (2005). *Qualitative Research Methods in Human Geography.* Oxford: Oxford University Press.

Hesse Biber, S. N. and Leavy, P. (2006). *Emergent Methods in Social Research*. Thousand Oaks, CA: Sage Publications.

Lincoln, Y. S. and Guba, E. G. (1985). *Naturalistic Inquiry*. Thousand Oaks, CA: Sage Publications.

Mantzavinos, C. (2005). *Naturalistic Hermeneutics*. Cambridge: Cambridge University Press.

Reis, H. T. (1983). *Naturalistic Approaches to Studying Social Interaction*. San Francisco, CA: Jossey Bass.

Shimony, A. and Nails, D. (1987). *Naturalistic Epistemology: A Symposium of Two Decades*. Dordrecht, the Netherlands: Reidel.

Williams, D. D. (1986). *Naturalistic Evaluation*. San Francisco, CA: Jossey Bass.

Relevant Websites

http://www.geog.buffalo.edu
Department of Geography, University at Buffalo: Association of American Geographers, Qualitative Research Specialty Group.

http://www.nova.edu
Nova Southeastern University: *The Qualitative Report*, journal dedicated to qualitative research along with comprehensive web resources.

http://www.uofaweb.ualberta.ca
University of Alberta: International Institute for Qualitative Methodology.

http://www.ualberta.ca
University of Alberta: *International Journal of Qualitative Methods*.

Nature

K. R. Olwig, Swedish University of Agricultural Sciences, Alnarp, Sweden

Glossary

Cosmography, Cosmology Historically, the geographer as cosmographer tried to synthesize all kinds of information about the Earth and its place in the universe in order to create a total picture of the cosmos as an orderly system. Cosmography seeks to describe the universe whereas cosmology sought to understand and explain the nature of the universe, which has been described. It might thus be said that cosmography creates a picture of the universe, whereas cosmology is concerned with our perspective upon that picture – our world picture (*Weltanschauung*) or 'world view'.

Enlightenment An eighteenth-century Western European intellectual movement emphasizing the 'light' of reason, rationality, and science in the understanding of human culture and the natural world and in encouraging its progress. Proponents of the Enlightenment were thus critical of Europe's *ancien régimes* as dominated by the aristocracy and justified by ideologies rooted in metaphysics and religion.

Renaissance It is usually seen to stretch from the time of its origination in Italy in the fourteen hundreds to the time when it gradually penetrated Northern Europe and England in the fifteen and early sixteen hundreds. Renaissance means rebirth and even though the term was not used at the time, it is appropriate because the people of the time saw themselves as giving a new birth to the classical antiquity of ancient Greece and Rome – the period which ended with the decline of the Roman empire a millennium before the Renaissance.

Sensorium At Newton's time sensorium meant, according to the dictionary, "the brain or a part of the brain regarded as the seat of the mind," which in modern terms refer to "the parts of the brain that are concerned with the reception and interpretation of sensory stimuli." For Newton, who was fascinated by optics, space was the 'sensorium' of God who in turn was the architect of nature.

The concept of nature is notorious for its complexity, and for this reason most attempts to explicate the meaning of the term end up with a confusing catalog of its many, seemingly disjointed, meanings. To avoid this problem this article focuses on the rationale by which meanings of nature have been created over time. Particular attention will be paid to the role played by geographers in gen erating these differing senses of nature, and the role they have, in turn, played in shaping geography. It is useful, in this context, to think of nature as a metaconcept and to think of geography in metageographical terms.

The prefix meta indicates that the concept is, as the dictionary puts it, of "a higher logical type," and when it is attached to the name of a discipline it designates an approach that "can deal critically with the nature, structure, or behavior" of the discipline to which the prefix is attached, as in 'metalanguage', 'metatheory' or, as in this case, metageography. Note the dictionary uses 'nature' virtually synonymously with 'meta ', suggesting that metageography is concerned with the 'nature' of geography. Geography itself, however, might be called a metascience because it creates an overarching world picture, or *weltanschauung*, as the Germans call it. This is why geography is not just a university discipline, but a school subject intended to mold the minds of school children. In times past, as shall be seen, geography was sometimes called 'cosmography', because it was, as the dictionary puts it, "a science that deals with the consti tution of the whole order of nature or the figure, dis position, and relation of all of its various parts." Cosmography, in turn, is foundational to the branch of philosophy called 'cosmology' that deals with the pro cesses of nature and the relation of its parts. The nature of geography is thus more than mountains, rivers, and plants; it can also be the metanature by which people understand the character of their universe.

This article seeks to make it easier to understand both the nature of geography and the nature in geography (which is to say the geomorphology, the meteorology, the landscape, etc.), by showing how geographers' use of the concept of nature continually flip flops between this metanature and more familiar physical meanings of nature.

The Metanature of Nature and Geography

One of the characteristics of the concept of nature as a metaconcept that makes it both so difficult to define, and so engaging, is its dichotomous cum dialectical character. A dichotomy is characterized, as the dictionary defines it, by "division into two parts, classes, or groups especially into two groups mutually exclusive or opposed by con tradiction," whereas a dialectic in its basic (original Greek) sense meant, according to the dictionary, "the theory and practice of weighing and reconciling juxta posed or contradictory arguments for the purpose of arriving at truth especially through discussion and de bate." Nature, thus, is commonly opposed to culture, so that to understand nature you need to understand how it might be understood as the converse of culture (a concept which rivals nature in its complexity). This dichotomy might be understood as a static dualism, so that culture is the negative of nature, but it can also be understood in a more relational and dialectical sense in which the notion of difference is fundamental. When nature is thought of in such dialectical terms the use of oppositions is primarily a device with which to think, as Plato demonstrated, but when it is dichotomized it is frozen (or reified) into fixed dualistic relations.

The most fundamental and volatile division within geography is arguably that between physical geography as a natural science concerned with natural phenomena, and human geography, which is a social and cultural science. The word 'physical' derives ultimately from the Greek word for nature *physis*, "the source of growth or change inherent in or construed as nature: something that grows or becomes," as the dictionary defines it. Physical is, in practice, usually synonymous with nature, and in some languages, such as those of Scandinavia, physical geography is thus called *naturgeografi*, 'nature geography'. Some geography departments are divided into physical and human geographical sections, and some geographical journals have separate series for physical and human geography. For some geographers this division is taken to mean a rigid dichotomy, to the point that they may even reject the relevance of the one form of geography to the other. Others, however, may see precisely the society/nature dialectic as key to the discipline. The issue, furthermore, is not simply philosophical. Dissension between physical and human geographers has ripped apart geography departments, and arguably caused some to close or become divided as separate departments, or to become divided as subdivisions of other departments, such as geology and sociology.

The Nature of Landscape as Nature

The apprehension of landscape as scenic space is arguably foundational to modern geography's dichotomous notion of nature in practice and in theory. The word apprehension is used here deliberately to mean "a perception that is comparatively simple, direct, and immediate and has as its object something considered to be directly and nondiscursively understandable," as defined by the dictionary. Practice has likewise been deliberately placed prior to theory. Many people in geography and related fields (such as landscape architecture and planning) are, in fact, practitioners who master, for example, forms of representation such as cartography, geographic information system (GIS), or draftsmanship, and then apply these tools to the research, planning, and management of the geographical environment. Other geographers are employed at universities to teach these skills. But even though there is a theoretical foundation for these endeavors, it is not a foundation that requires an elevated philosophical understanding of the idea of nature or space. For people basing their practice upon maps, GIS, or landscape diagrams, nature is made up of the objects apprehended via these representations, and space is the space of the map or the landscape diagram. This article will thus be concerned both with the practice of geographers, including representational practices, and the apprehension of nature to which they give rise, as well as with more theoretical and nonrepresentational understandings of the nature of, and in, geography.

The forms of representation that have subsequently guided the history of geography's engagement with the concept of nature derive largely from the Renaissance. If one wants to understand why contemporary geography, for example, is divided into physical and human geography, or questions of essentialism, sexism, or nationalist ideology, it is therefore necessary to know how these issues were built into geographical representations of nature that date way back in the Renaissance, when modern geography, and the idea of the modern itself, began to take shape. For this reason this article devotes important space to the Renaissance.

The term landscape in the Germanic languages, including English, originally referred to a region, or place, and its polity, and when the Renaissance Dutch, in particular, developed a genre of painting devoted to the representation of such places, this genre became known as 'landscape painting', much as the genre involving painting people became known as 'portrait painting'. In Samuel Johnson's (1709–84) classic English dictionary from 1755 the first definition of landscape was thus given as: "A region; the prospect of a country." The term, however, is also applied to a genre of painting that developed at about the same time (in the fifteenth and sixteenth centuries) in Italy and which used the newly developed technique of central point perspective to depict what were usually idealized scenes, rather than the landscapes of particular places or regions. Since these scenes could contain the same elements as the Dutch paintings, for example, hills, valleys, rivers, it is easy to confuse the two types of painting. This second type of landscape is covered by the second definition for landscape in Johnson's dictionary: "A picture, representing an extent of space, with the various objects in it." This second type of landscape thus begins its existence as a form of representation that generates a space within which objects can be located, whereas the first type of landscape began as a place called a landscape, which only thereafter became represented in prospects. Landscape in the second sense, that which grew out of Italian Renaissance art, created a pictorial space that was by no means limited to the frames of oil paintings, but which also played an important role in providing the back ground and surroundings for a new form of theater in which scenery, painted to illude the perspective of a spatial depth disappearing into infinity, frames the action. The scenes depicted in these representations were typi cally those of an idealized nature, as inspired by the literary classics and the Bible, but it is important to note that the form of spatial representation itself was also seen to be an expression of nature. This second, spatial, sense of landscape will first be focused upon here because it played a key role in the development of modern geog raphy. At the end of this article the sense of landscape as region, place and people will be returned to as a means to understanding the alternative, nonmodern ideas of nature that emerged at the end of the twentieth century.

The representation of the world in terms of landscape scenery creates an image that is characteristic of the modern perception of nature, complete with the di chotomies that have been seen to characterize modern geography. What we typically see in a representation of landscape as spatial scenery, as in the theater, is a layered structure in which the stage floor provides the scenic surface of the Earth with a characteristic geomorphology of hills, valleys, etc. Upon this surface is a layer of flora, and within and upon this flora live differing species of fauna. This is the layer of 'natural' landscape that one typically finds described as the bottom, foundational layer of the landscape in theoretical works on landscape geography, as well as in geography textbooks in the schools. Part and parcel of this foundational natural landscape is a superior layer of atmosphere with a climate that is characteristic of the scene's particular location on the globe (a heath may thus have dark clouds and a tro pical scene bright sunshine). Sandwiched between the terrestrial natural landscape and its atmospheric and ce lestial counterpart is the layer of the cultural landscape, and it is on this scenic stage that the geography of human existence is spatially located, or 'blocked' to use the theatrical term, and plays out its history.

In the theater the scenery often sets the stage for the action, signaling, for example, that the human relations depicted will be stormy, and there is likewise a long standing geographical tradition that sees the natural landscape as providing the foundation for, and thereby determining, the cultural landscape – so that stormy places might be seen to produce stormy peoples. This same scenic structure, however, also allows both drama tists and geographers alike to conceptualize the layer of culture as being unaffected by, or in opposition to, nature. And, finally, it allows for a more dialectical understanding of the relation between the layers of nature and culture where nature influences people, and people modify na ture, so that a stormy heath might be transformed into a peaceful farming landscape. The same layered structure can thus be interpreted in different ways. What is characteristic of the apprehension of such scenes is that they can seem obvious because they are based on a per ception that is comparatively simple, direct, and imme diate and has as its object something considered to be directly and nondiscursively understandable. This idea of landscape as scenery, and nature as landscape scenery, is so easy to take for granted that it seems 'natural'. This understanding of nature, however, is absolutely the most modern definition (number 15 in this case) to be found in a modern standard dictionary: "15 c: the aspect of out of doors (as a landscape): natural scenery." It was only in the late seventeenth century, in fact, that people really begin to use the word nature synonymously with landscape.

If we look to the oldest senses of nature to be found in a standard dictionary we find meanings such as:

> 2 a: the essential character or constitution of something,

or

> 4: a creative and controlling agent, force, or principle operating in something and determining wholly or chiefly its constitution, development, and well being: a : such a force or agency in the universe acting as a creative guiding intelligence : a set of principles held to be es tablished for the regulation of the universe or observed in its operation. (Merriam Webster, 1968)

Nature in these senses is not the things located in the space of the landscape, the flora, the fauna, the clouds, etc., but the abstract forces, principles, and laws which create the world and give its essential character. When people speak of the nature of something in the meta sense, they may be implying that it gets its character from the operation of these principles – for example, the nature of a river is to flow downhill – and the implication may be highly normative – for example, it is the nature of women to nurture children. Nature may be the most normative word in the language because what is natural is normal and what is unnatural is abnormal. The word nature is thus highly ideologically loaded, if often only implicitly.

The 'Birth' of Nature

The power of the abstract meta notions of nature listed above is hidden in the Latin root of nature, *nat* , which we also find in words like natal and nativity, but also in nation, the native born. The philosopher John Passmore reminds us that:

> The word 'nature' derives, it should be remembered, from the Latin *nascere*, with such meanings as 'to be born,' to 'come into being'. Its etymology suggests, that is, the embryonic, the potential rather than the actual. (John Passmore, 1974)

Today people think of nature in terms of actual things, like plants, so that the natural science of botany studies plants, but in its original sense the study of the nature of plants was the study of the natural principles that bring potential plants into being. Today people think of these principles as natural laws, so in this way the older sense of nature is still very much alive, except that people today are more likely to think of the plants, rather than the abstract principles that create material phenomena, such as the plants, as the primary nature. When nature is thought of in terms of principles we tend to be concerned with the laws governing, as the dictionary puts it, "the essential character or constitution of something," and hence with norms. The essential nature of something might thus be termed its inborn natal character that determines its growth and being.

In landscape apprehended as scenic space, the natural things in the picture – the geology, the plants, the animals, etc. – are those physical things which nature has generated. What is easy to overlook is that these things are represented within 'an extent of space' which also exemplifies a form of nature. As was the case with the laws and principles of nature, we do not tend to see the invisible space that has organized the natural elements in the picture. Thus, when we look at a typical landscape diagram in a physical geography or landscape ecology textbook we tend to focus on the objects depicted, such as the hills, the rivers, etc. The arrows in the picture, however, may help us to see, for example, the relationships between the geomorphologic processes governing the erosion of mountains into hills, the sedimentary soil in the valleys, the biogeography of the plants and animals, and the potentiality to grow differing types of crops on the river's flood plain. The spatial arrangement of the different elements in such landscape diagrams can imply, without our being necessarily conscious of this, that there is a natural causality between, in this case, the soil quality and the crops grown. This is because such diagrams are usually concerned with the vertical relationship between the natural foundation of the landscape in the soil and the other elements in the landscape, such as flora, fauna, climate, and crops. What can be easily overlooked, however, is the more horizontal spatial relationships between, for example, the farm economy and the demands of a distant market, which may, in fact, have an important determining effect on what is grown. The image thus appears to give a neutral objective picture of the relations between the physical and cultural phenomena depicted, but the spatial juxtaposition of the different things in the picture may, in fact, give a biased impression of what is determining the character of the landscape. The bias reflects the interests and intentions of the people behind the image, be they natural scientists, cultural historians, environmentalists, economists or, particularly in the nineteenth and early twentieth centuries, nationalist ideologists, intent on convincing people that their nation naturally sprang from the native soil. It is thus important to examine the way that space expresses nature, even when it is not visible nature.

Nature as Cosmological Intercourse

The fact that the root of nature means birth suggests the metaphor of mother nature (or the goddess *Natura*) as the female figure who gives birth to the world. But since it takes a male and a female to create offspring, there must also be a father nature. One of the oldest themes in cosmology is the idea of a celestial male figure that gives birth to the world through intercourse with an Earthly female figure. This idea was given artistic form by the Roman poet Virgil (70 BC–19 BC), who may well be the single most important influence upon Western artistic representations of nature, including those in the Renaissance Italian landscape paintings. One of the earliest recorded uses of the word nature synonymously with landscape was thus a late seventeenth century English translation by John Dryden of Virgil's *Georgics*, which is essentially a poetic handbook on the arts of agriculture. In the *Georgics* Virgil gives the following poetic form to the natural intercourse between earth and sky which gives birth to the world in the spring when:

> … the swelling earth
> Demands the seed of life. Then Father Air
> With fruitful rains omnipotent descends
> Upon the bosom of his smiling bride,
> And … greatly feeds
> Her teeming womb. …;
> The gentle moisture freely flows. …
> Thus dawned, I trow, the birthday of the world. (II, 380 420)

Ancient philosophers and poets drew upon myth to explain the cosmological principles behind what we would now call science. Virgil's 'birthday of the world' thus recalls, the cosmogony of Hesiod (*c.* 700 BC), where the world is born of the intercourse between a male sky god, Ouranos, and a female Earth goddess, Gaia, as well as the 'big bang' theories of modern cosmology, not to mention the 'Gaia hypothesis' of James Lovelock.

A poem such as Virgil's gives both flesh and blood and spatial perspective to the philosophical distinction between *natura naturans* and *natura naturata*. Northrop Frye, a literary scholar with an eye for cultural archetypes, has explicated the difference between the two Latin notions of nature in terms of looking down as opposed to looking up. Hunters and gatherers, or farmers, when looking downward would see how the richness of game or plant growth depended upon complex terrestrial factors such as moisture, soil, elevation, etc. Looking up, however, they would see that the Earth's fertility, and even that of female beings, correlated with the regular lawful movement of the heavenly bodies. This looking down and up eventu ally leads, according to Frye, to the distinction made in Latin between *natura naturans* and *natura naturata*. In the sexual creation myths, like that recounted by Virgil, the Earth mother might be regarded as a symbolic repre sentation of *natura naturans*, nature as a bursting forth of life and energy, its divine personalities the animating spirits of trees, mountains, rivers, and stones. Looking up, however, one sees a cycle of the Sun and the moon that implies a fixed lawful nature that never changes. This regularity, according to Frye, suggests the intelligent creation of an ideal nature in which becoming is sub ordinated to the being of a stabilizing power that controls and plans according to the laws of *natura naturata*. In Renaissance geography, following Frye, *natura naturata* would be represented by space, and *natura naturans* by the things located within that space.

Geography's Nature

The word geography combines the Greek geo , meaning Earth, with graphy meaning writing or representation. Geography traditionally has combined writing with other forms of representation, such as maps and pictures, in order to convey an image of the Earth. The Earth is a planet, and for this reason geographers have been con cerned with depicting not only the Earth, but also the relationship between the Earth and the heavenly bodies with which the Earth shares the space of the cosmos. The diurnal movement of the Sun, the change in the rela tionship to the Sun that takes place through the seasons, and the importance of the relationship between a par ticular location on the Earth and the Sun (e.g., a location on the equator) are of importance to the geography of a location. The relationship to the Moon is also important, particularly with regard to the tides. The relationship between the Earth an other heavenly bodies is less ob vious, but we must not forget that astrology, which is concerned with the interrelations between a multitude of heavenly bodies, was regarded as a high science for many centuries. Finally, it should be noted that one of geo graphy's most ubiquitous tools, the map, is usually based on the plotting in of coordinates on a grid according to celestial bearings. The nature of geography as a discipline is thus very much concerned with not just the Earth, but the Earth in relation to the cosmos.

A key event establishing the birth of the modern world picture in the Renaissance was the rediscovery in the early fourteenth century of an ancient classic text called *The Geography* by the second century Alexandrian Greek geographer, astronomer and astrologer, Claudius Ptol emy. According to Joan Gadol, a scholar of the Italian Renaissance:

> The systematic origins of Renaissance art and of the Copernican astronomy can be found in a movement of thought which may be properly called a "Ptolemaic re naissance" even though Ptolemy was to be deprived of his authority because of it. We have seen how Alberti established the rules of artistic representation by modi fying Ptolemy's principles of projection. When scientific "pictures" of the world came to be constructed according to these same principles, modern astronomy and geog raphy began their rise. (Joan Gadol, 1969)

With the rediscovery of Ptolemy's geography came the techniques from which modern mapping has evolved, which is basically the use of a graticule made of lines of latitude and longitude superimposed upon the surface of the globe, or a flattened section of the globe, to create a gridded surface upon which the location of points on the Earth can be plotted according to celestial coordinates. If these locations are along the coast of a body of land, and if they are connected by a line, what one then gets is a map with the outline of this coast. Ptolemaic cartography thus involved the making of maps of both the heavens and the Earth, and they were therefore very much rep resentations of the relationship between the *natura nat urata* of the heavens, with their unchanging lawful motion, and the *natura naturans* of the terrestrial world in which nature is undergoing a constant process of be coming. In the Renaissance Ptolemy's geography was termed a cosmography because it was considered to be a general description of the world or of the universe and an expression of what the dictionary calls "the science that deals with the constitution of the whole order of nature or the figure, disposition, and relation of all of its various parts."

For Ptolemy geography looks at the position rather than the quality, noting the relation of distances every where, and mathematics was thus an important part of geography, dealing with conditions on Earth that were correlated to those in the cosmos:

> For geography must first consider the form of the whole earth as well as its size and its position with reference to the heavens, so that it may be able to tell the size and nature of the known portion of the earth and under what parallel circles of the celestial sphere each place lies. From this it will be possible to learn the length of nights and days, the fixed stars that are overhead and those that at all times are above or below the horizons, as well as all other information that we include in an account of habitable regions. (Ptolemy, 1948)

The importance of Ptolemy's geography for the es tablishment of the modern scientific 'picture' of the world, referred to above, lies in the creation of perspective techniques through the modification, by

Renaissance polymaths like Leon Battista Alberti (1404–72), of the Ptolemaic map's top down projection to a more oblique angle. Ptolemy's map could be termed a (geo)graphic representation of an extent of space, with the objects within it, and the same applies to the use of the modified cartographic techniques used to represent landscape as scenery.

The Nature of and in Landscape

A Renaissance proponent of the new science of per spective painting was Leonardo da Vinci (1452–1519), an artist cum natural scientist. By using the underlying laws of geometry to impart a cosmic order to the organic terrestrial world, the science of perspective was, for da Vinci (as quoted by Richard Turner), both an expression and a representation of nature:

> If you despise painting, which is the sole imitator of all viable works of nature, you certainly will be despising a subtle invention which brings philosophy and subtle speculation to bear on the nature of all forms sea and land, plants and animals, grasses and flowers which are enveloped in light and shade. Truly painting is a science, the true born child of nature. For painting is born of nature; to be more correct we should call it the grand child of nature, since all visible things were brought forth by nature and these, her children, gave birth to painting. Therefore we may justly speak of it as the grandchild of nature and as related to God. (Turner, 1966: 25 26)

This view of painting explains why da Vinci regarded the science of perspective, which he helped develop, to "be preferred to all the discourses and systems of human learning." This can be seen in his painting of the popular artistic theme of the event known as the 'Annunciation', which took place when the archangel Gabriel made a birth announcement (on 25 March 1 BC, nine months before Christmas) to the Virgin Mary, informing her that she is to give birth to the child of God. Prior to the development of perspective, painters usually showed a scene reminiscent of Virgil's poetic description of cosmic coitus between celestial and terrestrial nature in the spring. What we see thus is a scene in which the be clouded figure of the heavenly father sends forth golden beams that descend upon the body of Mary. In the per spectival versions of this same scene, by contrast, God in the heavens above has disappeared, and instead the eye is transported out to the horizon into an infinite point in space by lines of perspective that frame and infuse the landscape scene upon which Gabriel and Mary act, as if in a theater. In the perspectival representation, the nat ural principle represented by the seminal light rays of God is replaced by the geometrical laws of an invisible spatial framework that prestructures the visible world. Perspectival representation thus made it possible to abandon the ancient conception of a dualistic nature, polarized between an eternal heavenly nature and a temporal Earthly nature, and to rethink these relations in terms of a corporal Earthly realm that is infused and penetrated by celestial geometrical principles (previously reserved for the heavens).

The significance of the transition in the representation of nature in painting exemplified by the transformation of the Annunciation was, as Godal stated, that "when sci entific 'pictures' of the world came to be constructed according to these same principles, modern astronomy and geography began their rise." Through to the Middle Ages people thought that the laws governing the 'perfect' movements of the heavenly bodies, circling around the Earth, were different from those that applied to the 'imperfect' realm of the Earth. The lines of perspective of da Vinci's perspectival space, however, connect Mary and her Earthly surroundings to the infinite space of the heavens, thus breaking the barrier between the Earthly and the heavenly. This same abstract space, working behind the scenes, to link heaven and Earth, can be seen during the Enlightenment in the work of Isaac Newton (1642–1727), a proponent of both modern astronomy and geography. For Newton, the same laws of motion through absolute geometric space applied both to the cosmic space of the heavens and to the space of terrestrial phenomena on Earth. Newton thus thought of space as the 'sensorium' of God because geometric space provided the lens like framework through which the underlying laws of nature, as established by God, were apprehended.

Newton was a strong supporter of modern geo graphical science – to the point that he supervised the translation and publication of the Ptolemy inspired geography of the Netherlands geographer, Bernhardus Varenius (1622–50). By following the precepts laid down by Ptolemy, Varenius came to provide an important model for the rise of modern scientific geography. Mapping and the optics of perspective, be it in geography or art, functioned as a kind of spatial sensorium through which a particular central, singular, and unified picture of the nature of the world could be formed. This framing of phenomena facilitates their treatment in terms of meas urable spatial relationships governed by natural laws, such as those of physics and geometry.

Nature, Landscape and the Birth of Modern Academic Geography

It was largely through the efforts of Alexander v. Hum boldt (1769–1859), one of the most famous scientists of the early nineteenth century, that geography was estab lished as a university discipline with strong roots in the natural sciences. The fact that he chose to call his piece de resistance *The Cosmos* harks back to the cosmographic character of Ptolemy's geography. Humboldt's geography represented a colossal synthesis of the interactions between the layered elements of the landscape from the soil to plants, to hydrography, and so on, and in producing this synthesis he made an important contribution to the subsciences encompassed by this cosmic world picture.

From the time of v. Humboldt to the mid twentieth century, geography was thought of as a discipline bridging the natural and human sciences. Landscape was a key term in this geography. It was particularly in Germany, however, that landscape geography was formalized as a science closely following the layered structure of landscape represented as an area of space with the objects in it. This tradition lives on particularly in Germany (and in nations within its cultural sphere), but also more generally in physical geography, as well as in ancillary subdisciplines such as landscape ecology. In these forms of geography nature is understood primarily in terms of physical things, the geomorphology of mountains, valleys, soil, flora and fauna, etc. It is characteristic of this tradition that its practitioners tend to focus upon the things represented within the landscape framework, but pay relatively little attention to the space that enframes their idea of landscape and acts as a sensorium through which the natural relations within it are apprehended. Humboldtian geography, however, embodied a more complex conceptualization of nature than is reflected in these relatively mechanical forms of landscape geography.

Nature and Difference in Geography

Humboldt's approach was clearly inspired by landscape art that literally formed, for him, a 'scientific picture' of the world. 'Science', as Humboldt (1849–58) put in the Cosmos, "is the labour of the mind applied to nature, but the external world has not real existence for us beyond the image reflected within ourselves through the medium of the senses," and by this he largely meant the pictorial image. The medium of the senses, however, is very much influenced by the imagination. He thus praises views hidden by clouds "knowing that his half transparent vapoury veil imparts to the scene a certain charm from the power exercised by the imagination over the domain of the senses." For Humboldt (1849–58), "the character of the landscape, and of every imposing scene in nature, depends … materially upon the mutual relation of the ideas and sentiments simultaneously excited in the mind of the observer."

The scenic landscape structure used by v. Humboldt worked to reframe ancient questions concerning the relationship between nature and culture, by focusing attention on the interaction between the physical and cultural elements of the landscape. This helped raise the question of whether society was shaped by God's and nature's teleological plan, as Carl Ritter (1779–1859) argued, or whether it was society that actively shaped its natural environment and its destiny. The American geographer George Perkins Marsh (1801–82) in his classic 1864 *Man and Nature: Or, Physical Geography as Modified by Human Action*, praised the 'more philosophical' and 'imaginative' character of Humboldt's and Ritter's, 'new school' of geography. Thus he noted that "the most interesting field of speculation, thrown open by the new school to the cultivators of this attractive study," was inquiry into "how far external physical conditions … have influenced the social life and progress of man." For Marsh (1864) the answer was that, whereas others "think that the earth made man, man in fact made the earth." By this he meant that man has "reacted upon organized and inorganic nature, and thereby modified, if not determined, the material structure of his earthly home" (George Perkins Marsh, 1864). For Marsh, as for Humboldt, this issue was not simply one of physical relations, but a question requiring an appreciation of 'the relations between mind and matter' as they changed through time – also into the future. Marsh is credited with having had an important influence on the subsequent development of the conservation movement, which, of course, required not just that people be made aware of physical facts, but also that their philosophical and esthetic understanding of the relationship between society and nature undergo a transformation. As Marsh put it, echoing Humboldt:

> To the natural philosopher, the descriptive poet, the painter, and the sculptor, as well as to the common observer, the power most important to cultivate, and, at the same time, hardest to acquire, is that of seeing what is before him. Sight is a faculty; seeing, an art. (George Perkins Marsh, 1864)

This might seem altmodisch a century and a half later, but the same basic techniques of cartographic and landscape representation as those developed by Humboldt, combined with rhetoric in the vein of Marsh, helped gain Al Gore the 2007 Nobel prize for his work on global warming.

The society/nature issue became a key 'problematic' in nineteenth and early twentieth century geography that encompassed geographers such as Friedrich Ratzel (1844–1904), Ellen Churchill Semple (1863–1932), and Ellsworth Huntington (1876–1947), who favored environmental determinism, as well as geographers such as Elisée Reclus (1830–1905), Joachim Frederik Schouw (1789–1852), and Vidal de la Blache (1845–1918), who, like Marsh, preferred to emphasize the active role of society in transforming nature, for better or worse. Environmental determinism played an important role in naturalizing nationalistic ideological claims of an essential blood and soil bond between the people of the nation and the environment encompassed by the state. After the nationalistic excesses of World War II this tended to give landscape geography of this kind a bad name. The critical tradition embodied in the work of Marsh, however, played an important role in the twentieth century revival of the concept of landscape by the so called 'Berkeley School' of cultural and historical geography founded by Carl Sauer (1889–1975). From the perspective of America the ques tion of perception became particularly vital because of the highly diverse ways the natural environment of the 'New World' was perceived through time by settlers from different cultural backgrounds, and the ways that these perceptions, in turn, differed from those of the native populations for whom this world was hardly 'new'. A pivotal moment for this tradition occurred with the publication of the massive 1956 conference compendium, *Man's Role in Changing the Face of the Earth* that provided an important point of departure for key contributions to the society/nature discourse ranging from Clarence Glacken's 1967 classic study, *Traces on the Rhodian Shore: Nature and Culture in Western Thought* – to David Low enthal's studies of the work of George Perkins Marsh and environmental perception, to the philosophical approach to nature in relation to place and landscape made by Yi Fu Tuan. By mid twentieth century, however, geography understood as the study of landscape was being eclipsed within mainstream geography by a redefinition of geog raphy as the science of space.

The Nature of Space

Space was key to landscape geography, but not its focus. The focus of geography changed dramatically in the mid twentieth century with geography's 'spatial revolution'. To understand this shift it helps to return to the time of Humboldt, and the work of a contemporary phil osopher who also taught geography, Immanuel Kant (1724–1804).

Kant became involved in geography's re visioning as a spatial discipline through the influence of Richard Hart shorne (1889–1992), the leading theoretician of mid twentieth century Anglo American geography. Hartshorne essentially proposed that geography abandon the concept of landscape and, with it, the concern with the historical process of society/nature interaction, and replace it with a notion of geography as 'spatial science'. Hartshorne found the precedence for his approach in the geography of Kant, who he felt had based his philosophy on an epistemology that separated space from time as the overarching cat egories through which people perceived and structured their world. Put simply, this implied that geography should be concerned with space, and time could then be left to the historians. In Hartshorne's view Kant's ideas influenced the ideas of Humboldt, as well as an extensive pantheon of largely German geographical authorities. Though Hartshorne's comprehension of Kant, and his influence on geography, can be questioned, it is nevertheless notable that Kant framework for understanding space resembles the classical tradition descended from Ptolemy.

For Kant the cartographic preconception of the world acted as a 'systematic' (or structural, to use the equivalent modern term) *Vorbegriff*. With Kant, as with his Renais sance predecessors, the map provided the basis for a conception of the world as landscape scene, as in a theater. According to Kant:

> The world is the substratum and the stage [Schauplatz] on which the play [Spiel] of our abilities takes place. It is the foundation upon which our modes of knowledge [Erkenntnisse] are acquired and used … . In addition we must get to know the objects of our experience as a whole, so that our modes of knowledge do not constitute an aggregate but a system [structure]. Because in the system [structure] the whole precedes the parts; in the aggregate, however, the parts precede [the whole]. … The whole is here the world, the scene upon which all our experiences are placed. (Immanuel Kant, 1802)

Through knowledge of the geography of this land scape scene, Kant (1802) argued, "we anticipate our fu ture experience, which we will later encounter in the world" and, in this way, gain "a pre conceptualization of everything …. [T]he perception of the world requires more than merely seeing the world." The idea of the whole as *Vorbegriff* is, for Kant, "architectural; it creates the sciences" (quoted by Paul Richards). As Kant phrased it:

> Space is not something objective and real, nor is it a substance or an accident, or a relation, but it is subjective and ideal and proceeds from the nature of the mind by an unchanging law, as a schema for co ordinating with each other absolutely all things externally sensed. (Immanuel Kant as cited in Paul Richards, 1974)

This statement brings to mind Newton's spatial sen sorium, except that the sensorium no longer belongs to God or to nature, but to the nature of the mind.

The Nature of Post-Hartshornian Space

By focusing on space Hartshorne appeared to liberate geography from landscape, and hence questions of na ture/society relations. Once thus liberated the way was open for a latter generation of geographers to drop Hartshorne's concern with the study of the areal distri bution of geographical phenomena (including natural phenomena), and concentrate on spatial analysis. Though the movement to redefine geography as a quantitative spatial science was cast as a 'revolution', it would be more correct to term it a focusing (as appropriate to an era that called itself 'the space age') on the spatial notion of na ture that was at the heart of Ptolemy's geography and landscape conceptualized as scenic space. Space became the 'sensorium' of late mid twentieth century geography and it was very much the absolute space of Newton, the map, and landscape scenery.

By the early 1970s theoretically oriented geographers began to question the idea of absolute space as inherited from Newton, but that did not mean geography had freed itself from Kant. Kant's concept of space combined elements of Isaac Newton's pregiven absolute space and the German notion of Raum (which is closer to the English notion of room), as formulated by the German mathematician and philosopher, and chief critic of Newton, Gottfried Wilhelm Leibniz (1646–1716). Kant phrased the relationship between the two conceptions of space this way:

> Most Germans, following Leibniz, hold that space is nothing but a relation between existing things, vanishing if the things vanish. Most geometricians, following the English, conceive it as a boundless receptacle of possible things. The receptacle theory is an empty figment. The first opinion comes into conflict with geometry. (Immanuel Kant, 1802)

Put simply, Kant solved the contradiction by treating space, in the 'English' sense, as a transcendent concept, rooted in intuition. It helps order and give meaning to knowledge gained through the senses, but it cannot predetermine the knowledge of concrete things gained through the senses. When Kant writes of 'differentiating' the regions of space he therefore notes that our general, rational, knowledge of this space is nevertheless of use only when we are able to comprehend it in terms of the relational space of the front back, left right, organization of our bodies.

The most influential critic of the spatial science of the quantitative geographers was David Harvey who argued for the substitution for Newton's absolute space of a Leibniz inspired conception of 'relational space'. The connection between space and the idea of nature and landscape was not lost, however, on geographers working within the Marxist tradition that Harvey helped generate. Neil Smith, a student of Harvey, played a particularly central role in elucidating the connection between uneven social and economic development, nature as a socially mediated phenomenon, and the creation of relational space. Im portant Marxist oriented geographers who are concerned with both space and nature are Don Mitchell, Margaret Fitzsimmons, and Noel Castree. Geography has sub sequently seen a proliferation of spaces, too many to number here, but a salient feature of most of them, that links them back to space as the sensorium of God and nature, is the tendency to write of space in the singular, just as God and nature are written of in the singular. The end result of the spatialization of geography, however, is not a geographic discipline in which one of its spaces reigned supreme, but rather a geographic discourse in which, for example, the relational space of capital accumulation is counterposed to the absolute space of the globe.

'Nonmodern' Natures

This article has hitherto focused upon notions of nature that are explicable in terms of a cosmology that can be traced back to Ptolemy (1948), and which became apprehensible through the representational medium of the map and the landscape as "a picture, representing an extent of space, with the various objects in it," to use Johnson's definition. The 'postmodern' era of the late twentieth and early twenty first century, however, has seen a critique of the foundations of this notion of landscape. This critique has forced geographers to re think their relationship to a modernity that arguably began in the Renaissance when, as Gadol (1969) put it: "scientific 'pictures' of the world came to be constructed according to these same [Ptolemaic] principles, [and] modern astronomy and geography began their rise." The critique of this 'modern' notion of nature came initially from a number of quarters, among them feminists, en vironmentalists, and science studies.

An early manifestation of the critique of geography's nature was the questioning by feminist geographers, and others, of the implied gendering of nature in phrases such as 'man and nature', 'man's role in changing the face of the Earth', etc. This critique is not entirely fair, since the word 'man' can be regarded simply as a shortening of 'human' that is not necessarily gendered male (the Swedish equivalent is gendered female). The implication that those using the term 'man', or found gazing at a landscape scene, are necessarily sexist and politically incorrect is unfortunate. The gendered cosmology of nature discussed above suggests, nevertheless, that there is an underlying mindset that is likely to interpret the expression 'man and nature' as a relationship in which man is viewed as active and cerebral, whereas nature is regarded as gendered passive, feminine, sensual, and corporal. These ideas, feminists have argued, continue to permeate geographical discourse, imbedded as they are in terms such as virgin nature, the penetrating gaze, etc. This gendered use of nature is not only dualistic, it is essentialistic and normative, and hence ideologically loaded. It implies, for example, that women, by nature, should be passive sensual creatures whose body naturally should be devoted to procreation.

Environmentalists, many of whom share the feminist critique, have been opposed to the polarization between society and nature that has attended the emphasis upon preserving virgin, untrammeled wilderness as the epitome of nature. Environmentalists have also been concerned with the polarized relationship between the natural and human sciences, and this has drawn some of them to the discipline known as 'science studies', in which the French anthropologist Bruno Latour is a leading figure. Latour has argued that modernism, as it emerged from the Renaissance and Enlightenment, made the promise that science, by isolating nature as an object of study, would be able to transform and control that nature to the benefit of society. The contemporary environmental crisis, however, suggests to Latour that this separation and objectification of nature, which thereby generates a nature/society dualism, may be a source of environmental problems, and this has led him to look for nonmodern alternatives to modernity. Within geography Sarah Whatmore's replacement of the society/nature dualism with a notion of a hybrid nature, incorporating society, is the most important example of this approach. The critique of modernity is in itself a critique of dualistic thinking because modernity posits itself as the progressive half of the dualism, modern/traditional. Just as 'man' is often posited as the active driving force that cultivates and makes fruitful a passive, feminine nature, so urban modernity has been posited as the driving force bringing progress to societies dominated by traditional rural, and hence natural, values.

Much of the critique of geography's concept of nature has focused on the role of representation. When landscape is conceptualized as a spatial representation of nature, the viewer apprehending the landscape is automatically placed outside of, and alienated from, this landscape and nature, which becomes objectified as scenery. Such a structure both obfuscates the role of the viewer within landscape and nature and implies the ability to manipulate such an objectified nature as so much spatial scenery. Part of the problem, when critiquing this model of nature, is its very success. This model of nature, as noted, essentially underlies Al Gore's slide show that won him the 2007 Nobel Peace Prize. The problem is not that this approach is not effective or scientifically useful, but that it is limited by the fact that the people who view such representations tend to be left, as it were, standing outside the picture as observers, when people are, in fact, an integral part of a nature that is not a picture at all, but a complex network of practices, including human practices. An approach that places people outside the picture will inevitably have difficulty explaining the myriad of ways that social practices are implicated in the processes, including the representative practices that produce both global warming and the science that studies it.

Nonrepresentational Nature

One solution to the way forms of representation, such as the landscape scene and the map, limit the understanding of nature, is to try think outside the box (or grid) of representation. If the ' graphy' in geography is defined as meaning "writing or representation," geography, by definition, cannot be nonrepresentational. This does not mean, however, that geography has not long grappled with the issue of phenomena that were not amenable to geographical representation. This is an issue that has been of particular concern with regard to the geographical concept of *choros*, which meant "land," among other things (such as place), in ancient Greek. This is the notion of land or region that was at the root of Johnson's first definition of landscape as region, and hence as something which can be represented, but which preexists representation. Landscape defined as a representation of space, on the other hand, is in origin a form of representation.

The heart of chorography, *choros*, appears to defy representation (though not artistic 'presentation'). One way of explaining the concept of *choros* (which is also often spelled *chora*) is to liken it to the cipher zero because it is a 'nothing that is'. Zero is a cipher that holds a place, even though it is nothing in and of itself, making it, from a representational point of view, an absent presence. A land, understood as *choros*/*chora*, thus holds a place where things can take place, grow, and become, much in the way that *physis* or *natura naturans* grows and becomes. This is very different from the way things are located in a frozen state of being as objects located in a spatial grid in the space of a map or a landscape scene. Geography has thus long contained discourses and concepts that have run against the grain of its use of graphic representation to comprehend the world, and thus, when present day geographers begin to take an interest in concepts such as *choros*/*chora* or *natura naturans*, they are engaging with issues that are not foreign to geography.

Inspired by the thinking of people such as Bruno Latour, Jacques Derrida, Baruch Spinoza (1632–37), and Giles Deluze, for whom concepts like *choros*/*chora*, and *natura naturans* are vital, there is today an increasing interest in nonmodern, nondualistic, and nonrepresentational conceptions of nature among geographers. The geography that most geographers practice, however, is one that is still rooted in forms of representation like the map and landscape scene. These are representations that have shaped the nature of, and in, modern geography. The question of what the nature of the geography of the future will be is now being debated, but unless the ideas of its critics manage to inform the practice of geographical practitioners, the nature of, and in, geography will remain much as it has since the Renaissance, and the dichotomies that divide geography and its apprehension of nature will persist.

See also: Ideology; Landscape; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Nationalism; Non-Representational Theory/Non-Representational Geographies; Quantitative Revolution.

Further Reading

Castree, N. (2005). *Nature*. New York: Routledge.
Castree, N. and Braun, B. (eds.) (2001). *Social Nature*. Oxford: Blackwell.
Cronon, W. (ed.) (1995). *Uncommon Ground: Toward Reinventing Nature*. New York: Norton.
Frye, N. (1980). *Creation and Recreation*. Toronto, ON: University of Toronto Press.
Gadol, J. (1969). *Leon Battista Alberti: Universal Man of the Early Renaissance*. Chicago, IL: University of Chicago Press. Quote p 157.
Johnson, S. (1755). *A Dictionary of the English Language*. [Reprinted 1968]. London: W. Strahan.
Kant, I. (1802). A translation of the introduction to Kant's *'Physische Geographie*'. In May, J. A. (ed.) *Kant's Concept of Geography and Its Relation to Recent Geographical Thought*, pp 255 264. [Reprinted 1970]. Toronto, ON: University of Toronto Press.
Marsh, G. P. (1864). *Man and Nature*: *Or*, *Physical Geography as Modified by Human Action*. [Reprinted 1965]. Cambridge, MA: Belknap Press.
Merriam Webster (1968). *Webster's Third New International Dictionary of the English Language, Unabridged*. Springfield, MA: G. & C. Merriam.
Olwig, K. R. (1993). Sexual cosmology: Nation and landscape at the conceptual interstices of nature and culture, or: What does landscape really mean? In Bender, B. (ed.) *Landscape: Politics and Perspectives*, pp 307 343. Oxford: Berg.
Olwig, K. R. (2002). *Landscape, Nature and the Body Politic: From Britain's Renaissance to America's New World*. Madison, WI: University of Wisconsin Press.
Olwig, K. R. (2006). Global ground zero. Place, landscape and nothingness. In Terkenli, T. S. & d'Hauteserre, A. M. (eds.) *Landscapes of the New Cultural Economy*, pp 171 192. Dordrecht: Kluwer.
Passmore, J. (1974). *Man's Responsibility for Nature*. New York: Charles Scribner's Sons.
Ptolemy, C. (1948). Geographike Uphegesis (exerpts). In Drabkin, I. E. & Cohen, M. R. (eds.) *A Sourcebook in Greek Science*, pp 162 165. Cambridge, MA: Harvard University Press.
Richards, P. (1974). Kant's geography and mental maps. *Transactions of the Institute of British Geographers* 61, 1 16.
Smith, N. (1984). *Uneven Development: Nature, Capital and the Production of Space*. Oxford: Basil Blackwell.
Thrift, N. (2007). *Non Representational Theory: Space, Politics, Affect*. London: Routledge.
Turner, R. A. (1966). *The Vision of Landscape in Renaissance Italy*. Princeton, NJ: Princeton University Press. Quote on pp 25 26.
von Humboldt, A. (1849 58). *Cosmos: A Sketch of a Physical Description of the Universe*. London: Henry G. Bohn.
Whatmore, S. (2002). *Hybrid Geographies: Natures, Cultures, Spaces*. London: Sage.

Nature, Historical Geographies of

P. Holland and A. Wearing, University of Otago, Dunedin, New Zealand

Glossary

Archipelago A cluster of oceanic islands.

Biodiversity It is commonly represented by the number of taxonomically distinct species in an area.

Biota The living things, plants as well as animals native to an area.

Carrying Capacity The number of individuals of a particular species that an area can support without additional resources.

Cataloger One who makes lists of the valuable, useful, and interesting things found in an area.

Chronicler One who lists key events in a society's history.

Ecological Footprint A measure of human dependency on environmental resources, often expressed as the area of productive land required to support a resident's lifestyle.

Ecological Services The economic value of the services that nature provides the residents of an area: for example, reduction in potential flood damage through natural control of runoff in a forested catchment after heavy rain.

Ecosystem The functional web of living things in an area, together with the physical resources and flows of energy that support them.

Eden The Garden of Eden as described in the First Book of Genesis of the Bible, and the home of Adam and Eve before their fall from grace.

Macro-Environment A large, environmentally distinctive area: for example, a belt of coastal sand dunes.

Systematics Scientific methods for naming and classifying living things.

Nature across the Centuries

The meanings that people have ascribed to 'nature' and the significance of nature to people have shifted back and forth over the centuries, but the belief that nature is a divine gift can be traced back to Mesopotamian and Egyptian creation mythologies and still remains central to many of the world's great religions. A complementary belief holds that nature is the stage given by the deity for people to transform and improve during their lifetime: by wholly or partially transforming native ecosystems, importing plants and animals for productive, esthetic, or recreational purposes, and recreating the Garden of Eden on Earth. This belief has often entailed making lists of the natural features of different places, usually before extracting, using, or relocating them: Strabo's account of the Roman world two millennia ago, the Doomsday Book for England late in the eleventh century, and the writings of Marco Polo about Central and Eastern Asia during the late thirteenth and early fourteenth centuries are early examples of this approach.

With the resurgence of mathematics and physics during the seventeenth century, particularly through the investigations of Robert Boyle, Galileo, Johannes Kepler, and Isaac Newton, European scholars like Georges Buffon began to feel that just as the orbit of a celestial body and the timing of an eclipse can be predicted with the aid of scientific axioms, analytical tools, and theories, so can nature be understood and explained without reference to the First Book of Genesis. In effect, they were drawing upon and extending the ideas of Greek scholars in the Mediterranean basin 1500 to 2000 years earlier. This revitalized current of thinking about nature ramified during the Enlightenment and afterwards, with Alexander von Humboldt a major influence. He pointed to fruitful new ways to investigate nature, including reciprocal relationships between nature and people. His books were widely read, and his scientific work was admired by Charles Darwin. When the latter's *The Origin of Species* was published in 1859 it catalyzed probing analyses of nature and fostered widespread debate about the Eden story, special creation, and catastrophic change. Across the whole of science, logical positivism – with its philosophical roots in deduction – came to be held in higher regard than the inductive or pattern seeking approaches that had dominated scholarship for centuries.

In 1911 the American geographer, Ellen Churchill Semple, wrote that "Man has been so noisy about the way he has 'conquered Nature', and Nature has been so silent in her persistent influence over man" (p 2). Like many authorities before her, but few since then, Semple saw nature as gendered. A decade later the social geographer Jean Brunhes described the European tradition of natural geography as an amalgam of the atmospheric sciences, biogeography, geomorphology, and hydrology, bringing it close to what is now termed physical geography. He saw wind and water as the prime natural forces acting on people, warned against expecting natural geography to explain much about individual actions, and argued strenuously against a determining role for nature on society. In 1930 the American geographers, C. A. Huntington and F. A. Carlson, declared that humans are

Table 1 Thinking about nature persistent currents in Western and other traditions

Wilderness, where nature is sublime
A place for spiritual growth
A mythical site of stories and traditions
A place of delight and surprise worthy of exploration and depiction in words and images
A dynamic and responsive 'land organism', sometimes termed Gaia, subject to ecological principles
A field of care requiring protection for and rehabilitation of native species and environments
An arena of diverse opportunities and risks for people
An esthetically pleasing and functional landscape where selected native and introduced species have been deployed in an environment modified by people
A wild domain for people to subdue

subject to the same forces as plants and animals, and that the human environment comprises five elements: the locational, climatic, physiographic, biological, and social.

The years since 1945 have been a time of extra ordinary pluralism in how environmentalists, including geographers, have viewed nature (see **Table 1**) but three influential publications stand out: *A Sand County Almanac* by Aldo Leopold, *Man's Role in Changing the Face of the Earth* edited by William Thomas, and *Silent Spring* by Rachael Carson. Earlier writers had shown how people have persistently. Misunderstood and mispresented na ture. Among the most influential of them were two Americans, John Muir and George Perkins Marsh. In *Man and Nature*, the latter declared ... that their activities can cause. Their impact was, however, greater during ... the revolution in environmental thinking of the 1960s and 1970s.

In recent times, what people understand about nature has been enriched by other's beliefs and traditions. To F. S. C. Northrop, who was impressed by Hindu culture and religious thought, nature "is a vast, spread out, going on ness, vague and indeterminate at its outer fringes, ablaze with diverse colors, and issuing forth manifest sounds, fragrances, and flavors" (p 1055). A half century later the American poet and environmentalist, Gary Snyder wrote that "nature means everything. The agricultural, the urban, the wild mountains and the forests and the many stars in the sky … Nature is our reality" (pp 15–16).

Methods and Materials

An historical account of nature can tell the modern reader as much about the writer's practical concerns and intellectual position as about nature itself. How we see the world is influenced by our knowledge and experience, the availability of terminology that captures the functions and origins of things observed in the landscape, access to recording and measuring devices, and the reasons for reporting what we sense. Over the centuries, as knowledge of the world grew more detailed so people asked progressively more probing questions about their surroundings. An early concern was to locate sources of useful items for exchange or trade. Later, scholars grew interested in differences between people and territories across the globe, and this inspired deeper questions about alliances, beliefs, and our place in the cosmos. The En lightenment crystalized approaches to surveys of nature, and the published accounts that followed were as much concerned with the origins, functions, and interrelations of the components of nature as with their appearances, names, and distributions. It is inevitable that there will be gaps in early accounts of nature but they are what we perceive, and not necessarily what the original observers regarded, as silences in the historical record.

A useful approach, and one that has long been em ployed by historical geographers, is for researchers to imagine that they are living in the area during the time of interest and to seek answers to the following questions. What did people know about nature and how did they obtain that information? What were their decision mak ing and governance structures? How did livelihood and nature intersect and interact? What were people's belief systems and what effect did they have on their activities? When supplemented with information from oral history and folk traditions, answers to those questions may en sure a deeper understanding of nature and our place in it.

Until Alexander von Humboldt's pioneering research in South America was published, most accounts of nature were partial in their thematic and areal coverage. It may be possible for contemporary historical geographers to recast, extend, and amplify such accounts with information from other sources: for example, travelers' tales, diaries and letters, newspapers and magazines, financial and other records, paintings and sketches, photographs and maps. What must, however, be kept in mind is that these sources tell us only what struck the observer as noteworthy at the time. For example, an old newspaper might refer to a major snowstorm on a certain day and at a particular place, but without formal weather observations from a network of stations we cannot know if it was confined to that place or unique to that year. Paradoxically, of all the information we need for a reasonably comprehensive historical geog raphy of nature what was not recorded may be of greatest potential value. The further back the period of interest, the greater the challenge to the historical geographer to fill any perceived gaps.

Nature, Islands, and Historical Geography

Jean Brunhes was interested in the 'natural entities' of human geography and regarded islands as instances of them. In his thinking, islands are similar to 'habitat islands' in the Equilibrium Island Biogeography model of Robert MacArthur and Edward Wilson: in other words, large or small areas surrounded by profoundly different environ ments, whether young or geologically old, and at various distances from potential sources of living things. Many oceanic islands have good documentary and fossil records that cover an often short period of human settlement. They also show the impact of human settlers who directly or indirectly altered environmental conditions to support their beliefs and ways of life, as Paul Bahn and John Flenley showed in *Easter Island*, *Earth Island*.

In the following sections, changing representations of nature are illustrated with particular reference to three clusters of low latitude, oceanic islands: the Maldives, the Society Islands, and the Seychelles. All have a relatively short history of human settlement: long enough for residents to leave their mark, but not long enough for people directly or indirectly to have erased most traces of the natural environment. An early account of the Mal dives, an archipelago in the Indian Ocean, was written by the North African geographer, Ibn Battuta who traversed the Islamic world during the fourteenth century. A few European mariners had previously visited the Society Islands in the central Pacific, but the journal of the British naturalist, Joseph Banks is a much fuller account of in digenous nature and society during the second half of the eighteenth century. The third example comes from the American biogeographer, Jonathan Sauer's study of peo ple and plants in the Seychelles, a cluster of high and low islands in the western Indian Ocean. All three accounts were written after comparably short periods of residence. They differ, however, in the ideas, experiences, and ap proaches of the three authors.

A Representation of Nature in the Islamic World

The index volume for Ibn Battuta's account of his travels from North Africa to China, and from Sri Lanka to Central Asia, contains 4450 entries of which only about 9% concern nature. The latter include topographic fea tures, macro environments, minerals, weather and hy drology, disease and health, pests, domesticated and wild animals, cultivated and native plants (see **Table 2**). The remaining entries are names of important people, king doms and administrative divisions, political alliances, and the like. Ibn Battuta was a careful observer and his de scription of the Islamic world during the fourteenth century hints at its vast longitudinal spread, chief macro environments, economic minerals, environmental, bio logical, and ethnic diversity, and large native vertebrates.

Ibn Battuta left a particularly detailed account of the Maldives and their profoundly transformed environments where coconut palms provided food, fiber, building materials, and shelter, where the only significant agri cultural crop was millet, and where survival depended upon access to the resources of land and sea. He described the archipelago of almost 2000 large and small islands as "one of the wonders of the world" (p 822), and was im pressed by the fertility of the soil; branches thrust into the ground quickly took root and sent out new shoots. Unlike Muluk, which he described as "one of the finest of the islands in greenness and fertility" (p 846), many parts of the archipelago were sparsely vegetated, including one small island with only a few coconut palms, some bananas, and two crows. Imported musk, ambergris, and rose water were used for personal adornment, and most of the fruit trees were native to other parts of the Asian Tropics. Of the lanes between houses, he wrote that "to walk in them is like walking in an orchard" (p 826). Exports of coconut fiber, cowry shells, and textiles covered the cost of im ported pottery and brass household goods, gold, and silver ornaments, and food items that could not be raised locally. He described one feast that included imported rice, clarified butter, and salt meat, as well as local fish, chickens, bananas, and coconut honey. For another feast, which Ibn Battuta gave important visitors from Sri Lanka, the Wazir provided five sheep that had been imported from India. After eating, his guests were served imported areca nut and betel, before a plate of Macassar sandalwood shavings was carried round for them to rub on their hands.

Table 2 Relative numbers of references to elements of nature as recorded by Ibn Battuta (the Islamic world and the Maldives) and Joseph Banks (the Society Islands)

Feature	*The Islamic world*	*The Maldives*	*The Society Islands*
Location and geology			
Geographical features	4%	3%	9%
Minerals	7	3	2
Physiography			
Macro-environments	3	-	8
Weather, climate, and hydrology			
Weather and hydrology	10	-	13
Biology			
Plants and plant products	46	69	49
Animals and animal products	26	22	16
Pests and diseases	4	3	3

Sources: Bivar, A. D. H. (2000). *The Travels of Ibn Battuta A.D. 1325 1354*, *Index to Volumes I IV.* London: The Hakluyt Society and Beaglehole, J. G. (ed.) (1963). *The Endeavour Journal of Joseph Banks*, (2nd edn.). Sydney, NSW: Public Library of New South Wales and Angus and Robertson.

To a twenty first century reader, Ibn Battuta's account of the Maldives is like looking at the world through a piece of coarsely woven cloth. Some important elements of nature are clear, others are blurred, and many are not evident, but as an account of the social and political structures of his times it is informative. Ibn Battuta, like Strabo 13 centuries earlier, was a servant of princes, a cataloger, and a chronicler. His geographical account of society and nature in the Islamic world during the fourteenth century was written in the spirit of how he viewed his responsibilities to his political masters.

A Glimpse of Eden

By August 1768, when Captain James Cook set out on from England on his first voyage of discovery into the Southern Hemisphere, science, technology, and phil osophy had greatly progressed since Ibn Battuta's time four centuries earlier. Jean Jacques Rousseau was in vestigating the effects of climate and soil on human so ciety, while Carl Linnaeus and his students were undertaking expeditions to distant places to collect plants for formal naming and scientific study. There was great interest among European gardeners in propagating the rare and beautiful plants collected by botanical explorers from distant places, and the foundations were being established for the great European botanical gardens and private estates. Nature was increasingly viewed by scholars as the proper realm for painstaking, scientific inquiry. There was, however, a growing tension between those who accepted the biblical representation of nature as the deity's gift to people and those to whom nature was a unified suite of diverse but interconnected phenomena subject to the laws of science. Banks was aware of those intellectual positions and was later to become a key figure in the Royal Society of London. His account of the So ciety Islands is remarkable for its detailed representation of a human society and an environment that had de veloped in virtual isolation from Asia and the Americas but which were on the threshold of a long period of catastrophic change. Banks wrote about human mores, governance, and culture, made frequent reference to plants and animals, and described the islands' macro environments. He also commented on the form, con struction, and spatial arrangements of houses, and noted that although many paths wound through the settled lowlands only a few penetrated the forested hill country. He even followed one river valley into the mountainous interior in the hope of finding inland settlements, but was not successful.

Banks has left us an account of Edenic nature in part of Polynesia, and he employed artists to depict the transformed and wild landscapes visited by the *Endeavour*. Several of their sketches and paintings of Tahitian landscapes have been reproduced in Bank's published journal but they do not convey the sense of exuberant nature that is so evident in his journal or in the paintings of William Hodges a decade later. On 13 April 1769, his first day in Tahiti, he encountered groves of breadfruit and coconut and wrote in his journal that "the scene we saw was the truest picture of an arcadia of which we were going to be kings that the imagination can form" (Beaglehole, 1963: 252), a place where "benevolent na ture has not only supplyd them [the Tahitians] with nescessaries but with abundance of superfluities" (Bea glehole, 1963: 342).

> No countrey can boast such delightfull walks as this, the whole plains where the people live are coverd with groves of Breadfruit and cocoa nut trees without underwood; these are intersected in all directions by the paths which go from one house to the other, so the whole countrey is a shade than which nothing can be more gratefull in a climate where the sun has so powerfull an influence. (Beaglehole, 1963: 340)

He even calculated that ten mature breadfuit trees would satisfy a Tahitian family's annual needs.

Although members of Cook's party were usually able to obtain sufficient fresh food, there were times when the islanders had none or only a little to spare, notably when breadfruit was out of season. Banks seems not to have appreciated that even in Arcadia there are seasonal limits to plant growth and a finite environmental carrying capacity. His account contains many references to reefs and the challenges they posed to people in large and small boats, and is a good source of information about environmental conditions and their spatial variability. He had brought vegetable seeds from Rio de Janeiro to Tahiti and, showing a scientist's experimental thinking, planted them "in as many varieties of soil as I could chuse" (p 308). He also found that residents had planted breadfruit trees on the less fertile lower hill slopes to ensure a supply of food when plantations on the coastal strip failed during drought or after a severe storm, but he estimated that these reserves would not be sufficient to cover a major emergency.

Banks, like Ibn Battuta before him, was writing about transplanted and transformed nature (see **Table 2**) and he, too, recognized the diverse uses to which people may put a small number of plants specially introduced for the pur pose: for example, ropes and threads, building materials, clothing, and mats were made from breadfruit, coconut, fig, hibiscus, nettle, and paper mulberry plants. Ocean and lagoon were the main indigenous sources of animal protein and they were supplemented by domestic chickens and dogs, both originally from Southeast Asia. Unlike Ibn Battuta's account, however, Banks' writing reveals a sci entist fully engaged with what he saw and collected. With the exception of coral, he observed that stones carried the "visible marks of fire" and concluded that Tahiti "owes its original to a volcano which now no longer burns" (p 308). He was struck by the economic importance of introduced food, fiber, and decorative plants, the many species of fish – more than in the seas around Great Britain – and the very few native land plant and animal species. Banks was one of the earliest European biologists to collect in a tropical oceanic archipelago, with its characteristically small and specialized native biota of a dozen or so cosmopolitan or wide ranging species, and a small number of endemic species, and so his observations do not come as a surprise to readers 250 years later. On Huahine, Banks and Daniel Solander, the latter a student of Linnaeus, found only "10 or 12 new plants, a few insects indeed and a species of scorpions which we did not see at Otahite" (p 317). On Ulhietea the two were "very agreeably entertaind by the reception we have met with from the people, tho we were not fortunate enough to meet with one new plant" (p 323). On another island their walk "did not turn out very profitable as we found only two plants that we had not seen before" (p 326), and a seine net cast into the lagoon yielded few fish for them to preserve and describe.

Plants, People, and Nature in the Seychelles

Our third example comes from the findings of Jonathan Sauer, who investigated relations between plants and people on the Seychelles' coast during the 1960s. The archipelago had been settled by early in the eighteenth century, and after 1770 was successively colonized by France and Great Britain. The islands owe their historically recent occupation to distance from the lands that flank the Indian Ocean and isolation from the main trade routes between India, the Middle East, and Africa. The two dozen larger islands have an often rugged terrain on a granite base, and some coral. The climate is maritime equatorial, dominated by latitudinal shifts in the trade winds, and subject to relatively few storms. Unlike the sea that surrounds the archipelago, land areas have only a modest biodiversity relative to their size and environmental heterogeneity: 80 native species of fern, 170 species of flowering plants, a dozen endemic land birds, a bat, several large reptiles, numerous insects, and other invertebrates.

> A unique aboriginal vegetation did develop in the mountains and interior lowlands of the larger islands: magnificent rain forests, palm groves, and screwpine thickets. These were composed mainly of endemic species, presumably derived from waif immigrants arriving by means of unlikely and seldom reported dispersals. Considering the mountainous, humid, equatorial environment, the inland communities were extraordinarily poor floristically. They proved highly vulnerable to human disturbance and few relics survive today. (Sauer, 1967: 93)

Sauer (1967) concluded that differences in geological substrate and varying exposure to the prevailing weather systems accounted for many observed differences in vegetation structure and composition. He did, however, comment on the transformational activities of successive groups of settlers: "The Seychelles coast thus had acquired a rich natural assemblage of over fifty cosmopolitan species, forming a resilient vegetation which was to be reshaped but by no means ruined by human intervention" (p 94). Over the period of human settlement, successful plant introductions had virtually doubled the size of the islands' flora. Sauer noted 62 introduced plant species, many of which had entered as stock raised at the Mauritius Botanical Garden. They included cinnamon and other spices, coffee, cotton, patchouli, sweet potato, tobacco, maize, and the vanilla orchid. As in other Indian Ocean and Pacific Basin islands, the small and versatile flora of native and introduced species was used for boat building, as well as for cordage, tanning, dye stuffs, fish poison, medicinal, and ceremonial purposes.

With the aid of words and statistics, maps and diagrams that depict the results of field studies on transects running inland from the coast, Sauer presents a richly detailed, culturally aware, and scientifically informed representation of a transformed and transplanted landscape, and a pleasingly full account of wild nature on the Seychelles.

Historical Geographies of Nature: Problems and Progression

For many centuries the metaphor of the garden has permeated thinking about nature and settled landscapes. In traditional Japanese culture, gardens are idealized and profoundly controlled representations of individuals living in harmonious relations with nature. In contrast, a formal European garden laid out in close conformity with Euclidean geometry epitomizes human dominion over nature. During the nineteenth century in early colonial New Zealand, wild nature lay beyond the fence that surrounded the orderly arrangements of homesteads and out buildings, lawns and vegetable gardens, orchards and shelter plantings. Over time, settlers' highly managed and economically functional environments spread across their properties and through the rural landscape. Recently, rural people have shown interest in a more diversified realm where closely managed agricultural systems and modified native ecosystems form an esthetically pleasing and environmentally sustainable mosaic (see **Figure 1**) such as can be found today in parts

Figure 1 Edging toward a biologically more diversified and structurally more complex landscape in the Catlins region of southern New Zealand. Native shrubs and bracken fern persist on the hill crests, intact or only lightly cut-over native forest predominates on upper slopes, successional communities replace the former forest cover on mid-slopes, and intensively farmed land occupies the lowest slopes and valley floors. Unlike the steeper hill country, the lowlands have been almost completely transformed by people, plants, and animals, and little remains there of the native ecosystems that prevailed in the late nineteenth century. Source: A. Wearing.

of the New England states of the USA, the British Isles, and Western Europe. The vector of change may be turning away from human dominion over nature toward coexistence with natural forces and elements.

A historical geography of nature is inevitably a close reflection of successive residents' belief systems and the writer's intellectual position. While people view nature as a stage given to them to transform they will show little abstract interest in its components, let alone functions and interactions. Developments in the scientific disciplines of systematics, ecology, and environmental studies, supported and extended by the publications of humanists like Simon Sharma and Gary Snyder, have helped change that stance and are fostering an expanded and more nuanced appreciation of nature. One recent current in thinking involves a return to the notion of value, typically expressed as the financial worth of ecological services provided by plants, animals, and ecosystems. Several national analyses have been completed and most show that the value of such services exceeds the gross economic product.

A historical geography of nature needs to be sensitive to the dynamic relations between living things and their environments. It should refer to the area's weather systems and climates, describe those terrain features and principal landforms that affect the distributions of plants and animals, show how the soil cover may vary in its capacity to support plant growth, account for the origins and vulnerability of the native biota and ecosystems, and evaluate the transformational capacity of introduced species and alien technologies. Comparisons should be drawn with situations elsewhere, and attempts made to explain the findings by reference to bodies of theory in the humanities as well as in the biological, Earth, environmental, and social sciences. The investigator should also take note of annual and less frequent environmental challenges – such as extremes in temperature, changing water availability, and strong winds – and identify the key features of wild and transformed systems. As most such observations were first made in the middle of the nineteenth century, it is inevitable that any account of nature written before then will be partial.

A recurrent theme in the historical geographic literature is that while nature may be silent it is also responsive. James Lovelock, who proposed the Gaia model, has argued that human societies will continue to face the prospect of a system inimical to their survival until they learn how to strike the balance between off take and what the ecosystem can supply. By knowing how successive societies have interacted with nature we could be better placed to respond to the societal and environmental challenges of the twenty first century.

See also: Environment, Historical Geography of; Exploration; Imperialistic Geographies; Nature, History of; von Humboldt, A.

Further Reading

Beaglehole, J. C. (ed.) (1963). *The Endeavour Journal of Joseph Banks* (2nd edn.). Sydney, NSW: Public Library of New South Wales and Angus and Robertson.

Bivar, A. D. H. (2000). *The Travels of Ibn Battuta A.D. 1325 1354, Index to Volumes I IV*. London: The Hakluyt Society.
Gibb, H. A. R. and Beckingham, C. F. (1994). *The Travels of Ibn Battuta A.D. 1325 1354, Volume IV*. London: The Hakluyt Society.
Glacken, C. J. (1967). *Traces on the Rhodian Shore: Nature and Culture in Western Thought from Ancient Times to the End of the Eighteenth Century*. Berkeley, CA: University of California Press.
Guthrie Smith, H. (1921). *Tutira, the Story of a New Zealand Sheep Station* (1st edn.). Edinburgh: Blackwood.
Hacking, I. (2007). Almost zero. *London Review of Books* 10 May, 29 30.
Hadot, P. (2006). *The Veil of Isis: An Essay on the History of the Idea of Nature*, Chase, M. (trans.). Cambridge, MA: Harvard University Press.
Mann, C. C. (2005). *1491: New Revelations of the Americas Before Columbus*. New York: Alfred A Knopf.
Naylor, S. (2006). Historical geography: Natures, landscapes, environments. *Progress in Human Geography* 30, 792 802.
Sauer, J. D. (1967). *Plants and Man on the Seychelles Coast: A Study in Historical Biogeography*. Madison, WI: University of Wisconsin Press.
Schama, S. (1995). *Landscape and Memory*. Toronto, ON: Random House of Canada.
Semple, E. C. (1911). *Influences of Geographic Environment on the Basis of Ratzel's System of Anthropo Geography*. New York: Henry Holt.
Snyder, G. (2006). Writers and the war against nature: Above all we need human beings who love the world. *Resurgence* 239, 12 17.
Thomas, W. L. (ed.) (1956). *Man's Role in Changing the Face of the Earth*. Chicago, IL: Chicago University Press. (The source of the F. S. C. Northrop quotation.)
White, G. (1789). *The Natural History and Antiquities of Selbourne* (by Scolar Press). London: B. White. (Reprinted by Scolar Press.)
White, L. (1967). The historical roots of our ecological crisis. *Science* 155, 1203 1207.

Relevant Websites

http://www.hakluyt.com
Hakluyt Society.

Nature, History of

M. Tanskanen, University of Joensuu, Joensuu, Finland

Glossary

Berkeley Geography Graduate School It refers to the loose association of like-minded geographers associated with Carl O. Sauer (1889–1975) and his viewpoints and predilections. Sauer had a long career in the Department of Geography at the University of California, Berkeley and he fostered a distinctive school of geography grounded in biophysical, cultural, and historical approaches.

Cultural Ecology It studies the adaptive processes by which the nature of society and an unpredictable number of features of culture are affected by the basic adjustment through which human beings utilize a given environment.

Environmental Determinism The doctrine that individual human actions, beliefs, and values are controlled or determined by the ambient environment. According to environmental determinism, societies, cultures, and civilizations are also products of their environments.

Environmental History It is the study of humans and nature and their past interrelationships. Environmental historians base their understanding of human and nature relations primarily on historical methodology, but often borrow from the work of scientists and scholars in fields outside of history.

Introduction

The most traditional way of approaching history of nature is to define it concretely as a continuance of natural phenomena from the perspective of the natural sciences, for example, periods of climate, hydrology, and other natural systems. Especially as the climate change is recognized as one of the greatest environmental, social, and economic threats facing the planet, an interest in history of nature has awoken in many sectors. The academic world is trying to discover what is happening in nature and how the climatic processes have changed.

It is obvious that human actions have caused the climate change. Therefore, one of the most important questions concerning the habitable Earth and human beings' relationship to it is in what manner humanity has changed the Earth from its hypothetically pristine condition. Instead of through nature's systems, history of nature can be approached also from the standpoint of humanity. Climate change is a common topic in everyday conversations. It is easy to find a person who has made a connection between exceptional weather phenomena and climate change and, at the same time, given thought to history of nature and to the timescales of nature. In academic interest, it is essential to ask how people perceive the living world and environment around them. An important additional question is which time tied attitudes to nature have influenced human action at different times.

Geography deals with all these questions and it is the strength of the discipline. In regard to climate change, a holistic geographical analysis of the questions insists on exact measurements, mathematical calculations, and different models and also studies of how human beings' attitudes to nature have transformed over time. Scales of nature and scales of human beings have to be observed.

The aim of the text is to focus on the ways in which human geographers have paid attention to history of nature, how it has been defined in human geography, and what its role is in the present human geographical research. First, the concept of nature is briefly introduced. It is done for the purpose of clarifying what geographers mean when talking about nature. The topic of the article is then approached from the angle of historical geography, a subfield of human geography. History of nature is a vast theme and there is no sense in trying to summarize the whole field in one article, therefore historical geography has been selected to exemplify the topic. The text focuses on the ways how nature and history of nature have become a part of research in historical geography.

Concept of Nature

Nature as a concept is elusive, complex, promiscuous, and also familiar. In the broadest sense, nature can be drawn parallel to the natural world or physical universe. Nature refers to the phenomena of the physical world and also to life in general. In many contexts, it precludes manufactured objects and human interaction but it has also been linked to human beings and attempts to describe the complexity of human life (e.g., human nature).

Nature has often been called one of the most complicated words in the English language. It is a strongly dualistic concept that is close to the concept of culture. On the one hand, nature is external, nonhuman reality, somehow pure and God given but, on the other hand, it is an abstraction that incorporates human, as well as nonhuman spheres of reality.

The complexity of nature has led to its division into first, second, and third natures. First, nature refers to a given, pristine nature of biophysical processes and forms, to genuine and original nature untouched by humans. Society has evolved from that nature and the impact of humanity has been so pervasive that first nature has been replaced with second nature which contains both culture and environment. Second, nature is partly biophysical and partly social. Third, nature comprises the information flows and representational signs, symbols, and discourses of nature and it also symbolizes human domination over nature. It materializes, for example, in the symbolic and tamed nature of urban parks and gardens. In addition to first, second, and third natures there have been definitions of fourth nature which refers to the situation where the passed over nature takes back its place and the boundaries between human space and wilderness are blurred.

One can argue that nature and culture are contiguous and the major subfields and the key concepts of human geography are focusing on the intersections between them. Traditionally, one of geographers' main questions is whether people are separate from nature or subject to the rules of nature. The relationship between humanity and nature (or humanity and environment) is an eternal question from which huge amounts of geographical researches have arisen. Predominantly, geographers have been more sensitive to the complexity of nature and less willing to see it as a dualistic concept. In a theoretical sense, the emphasis has mainly been on the interpretations of the relationship between the concepts of nature and culture.

Some human geographers have even suggested that nature is socially constructed and nature does not have an existence separate from culture. The basis of this idea is that all knowledge of nature is contextual and socially constituted and people can understand nature only through their own culturally constructed definitions of it. Especially geographers influenced by the so called new ecology are apt to talk about nature–society hybrids rather than about two interacting domains. At the same time, an idea of understanding nature and culture as gendered has emerged, with nature given a passive feminine character and culture a masculine empowered one.

History of Nature in Historical Geography

Historical geography seeks to understand geographies of the past and also how the past impinges on the present. Historical geography derives theoretical viewpoints, subject matters, and methodological tools from both geography and history and it works at the boundary of these academic disciplines. It is a hybrid approach that examines land use, landscapes, environments, human–environment relations, spaces and places historically, and the geographies that have changed over time. The aim here is to exemplify how history of nature has been present in historical geography research. Historical geography as a theme is approached through introducing some authoritative geographers and their classics from the viewpoint of how nature and history of nature have developed as a part of research. The topic is also examined through the discursive question of the invisibility (or even loss) of nature in human geographical studies. The end of this section illustrates the role that history of nature has in the culturally and socially oriented study of the Finnish mire landscape. Finnish mire landscape is a good example for underlining the differences between the timescales of nature and humanity and also for emphasizing the notion of the four natures.

Roots of Nature-Oriented Historical Geography

Historical geography has developed from several approaches in a succession of overlapping periods of innovation. The idea of nature oriented historical geography has its European roots, for example, in the literary productions of German geographers Alexander von Humboldt and Carl Ritter, and the American roots in the works of lawyer George Perkins Marsh and historian Frederick Jackson Turner.

In *American history and its geographic conditions* (1903) Ellen Churchill Semple described the ways in which the natural environment of the United States had conditioned the course of American history from the seventeenth to the nineteenth centuries. The book reviewed developments in American history, including the settlement of the Atlantic seaboard colonies, the development of cities, and the expansion of road and rail systems. Semple argued that each of these developments had been structured by the natural features of the land. The idea of environmental determinism, a view that the physical environment, rather than social conditions, determines culture, was strongly present, and the connection between geography and history was indisputable. Semple's research was subject to a great deal of critique but she has, nevertheless, been raised to a pioneer status in the study of human–environment interaction. She was among the first to detail the ways the natural environment has impacted the course of human history.

In historical landscape geography, Carl Ortwin Sauer in his well known publication *The morphology of landscape* (1925) was concerned with the way human culture has fashioned and modified natural landscapes to produce cultural landscapes. Sauer was critical of environmental determinism and proposed an approach called landscape morphology or cultural history. That approach involved inductive gathering of facts on human impact on landscape over time. Sauer had a firm idea of how a landscape could be presented not only as a region but also as a chronological space processed by both natural forces and cultural impacts. According to him, the aim of landscape research was to trace the process in which a pristine natural landscape changed into a human occupied cultural landscape.

Sauer believed history is as important to geography as physical terrain and that there is such a thing as a humane use of land. He contemplated the Earth in all its vast physical and cultural variety and its changes through both geological and historical time. Sauer specialized in studying causal relationships between the elements of the natural environment and the activities or creations of human beings, and after his retirement the so called Berkeley geography graduate school developed the research into cultural ecology. Sauer was one of the first landscape geographers who did not only want to observe and outline the physical landscape but also to understand the background connections of landscape elements and the variety of interrelations within a landscape. He was also ahead of his time when his thoughts included a burgeoning anxiety about the impoverishment of biodiversity caused by human action.

When examining the interplay of the human–environment relationship through the classics of empirical studies, one has to call to mind Henry Clifford Darby's cross section method as one of the best known methods in historical geography. It was introduced in *Domesday England* (1977) and is dualistic by nature. During the first stage, a researcher tries to reconstruct historical cross section points from an exact geographical area. Depending on the area, materials can consist of, for example, old maps, aerial photos, and modern geographic information system (GIS) generated maps. The second stage consists of analyzing processes of change by writing vertical themes, or narratives, in which the idea is to explain the change in landscape by paying attention to the context in which the change has happened.

During the second stage of the cross section method, one needs to observe both natural and socioeconomic processes. It is also essential to pay attention to timescales because the scales of nature are different from the scales of humanity. Furthermore, it is necessary to remember that it is impossible to trace the present exclusively from the past. In certain cases, the present includes several alternatives and only one is going to be realized. However, the past is contiguous with the present, defining it and also diminishing the number of existing possibilities. Both history of nature and history of humanity include information that supervises the future of the units to which it is related.

The idea of nature and knowledge of the place of humans in nature clarify in Clarence J. Glacken's monumental work *Traces on the Rhodian shore* (1967). He rightly insisted that it is impossible to understand nature without integration with an understanding of history. According to him, nature and history become empty ideas if isolated from each other. He related social and natural phenomena to the imagined dichotomy of humanity and nature. His main thesis on how nature influences culture and vice versa has later become groundwork for many geographical, philosophical, historical, and environmental studies.

Cryptic Role of Nature in Geography

The impact of the above mentioned authoritative geographers and classics of historical geography has been remarkable. For example, from their perceptions, the concept of nature and the history of the idea of nature have arisen as an inseparable part of studies of historical geography. It can also be argued that geographical research has often been strongly directed by history of nature.

Nature as a resource for human life has traditionally been one of the main themes in geography. Research themes concerning with mineral wealth, use of fossil fuels, loggings of rain forests, use of boreal forest belt, and exploitation of sea areas have been largely examined by several disciplines and especially in geography they have been subjects of numerous studies. In physical geography, viewpoints have varied from positivism oriented geological, climatologic, and ecological approaches to predominantly statistical and nowadays GIS based examinations of the exploitation of natural resources.

In human geographical research, viewpoints have varied from structuralism based local community studies to constructive value and attitude studies and further to methodologically humanistic examinations concerning individual experiences and feelings connected to the use of natural resources. In human geographical studies, the focus has been on humanity, on an individual, or on a certain group of people, but the knowledge of nature as an inherent element of research is fundamental. The problem is that the role of nature and history of nature as a supervisor of studies has not necessarily been spelled out in research reports, and it is easy to arrive at the impression that nature has been ignored.

The invisibility of nature in human geographical research has also resulted in critique. From the 1980s, especially environmental historians have criticized historical geographers for reducing the physical environment to nothing more than a passive stage on which socioeconomic forces battle through a series of conflicts.

Nature has been conceptualized to an obscure abstraction and that has led to a situation where the role of history of nature is almost impossible to outline. In many studies it is easy to detect the influence of history of nature as merely a supervisor without special emphasis. The focus of studies can be, for example, on the transformation processes of cultural landscape but often human action is outlined in reflection to nature and history of nature.

Geography, contrary to most disciplines, has not fully compartmentalized society and nature and maybe this causes an illusion that geographers have lost nature *per se* from their studies concerning the relationship between humanity and nature. Raymond William's (1980) dictum "the idea of nature contains, though often unnoticed, an extraordinary amount of human history" has been espoused by geographers and this is a part reason for emphasizing human history more than history of nature in human geographical research. Under pressure of criticism from environmental historians and sociologists, geographers can defend themselves by revealing the character of the concept of nature which, having been conceptualized into an abstraction, is not easy to understand without familiarity with it.

The vitality of nature expresses in many present studies and publications by human geographers. For example, Noel Castree's *Nature* (2005) examines critically the role that geography has played in developing societal understandings of the concept of nature and highlights the diverse aspects of the nature that geographers examine within a common frame of analysis. One of Castree's main arguments is that nature often appears in geographical discourse through collateral concepts where it figures as a real but ghostly presence.

Finnish Mire Landscape as Example

The exploitation of boreal forest lands has been a fascinating geographical question, especially from the Northern European perspective. After World War II, the intensive use of forests and mires has produced many local and spatial conflicts, and questions of politics of nature, regulations of using forests and mires, and issues of social and environmental justice have arisen into the core of human geographical research. When focusing on these questions, wide knowledge of both natural and social contexts is needed even if nature in itself (forests and mires) seems to be reduced to only a passive forum of conversation.

When studying the changes in Finnish mire landscapes drained for forestry, one needs sufficient knowledge of the natural history and morphological features of mires. When analyzing the cultural character of present mire landscape, it is not enough to gather statistics of its drainage for forestry and try to find the socioeconomic context from archival sources. It insists also the knowledge of mire formation and rate of peat accumulation and the understanding of the role of time in all those processes. The present mire landscape is both a reflection of human action during the last few centuries and also an expression of natural processes that have been taking place for thousands of years.

Finnish mires are postglacial formations and many natural processes have transformed them into the physical shape which has, in many ways, been usable for human beings. When people first came to the north, mire nature was in its pristine condition untouched by humans and the use of mires adapted to the circumstances of nature.

The impact of humanity increased from the eighteenth century and first nature replaced with second nature which contains both the Finnish society and the natural processes of boreal forest area. Especially technical development in the second part of the twentieth century has made extensive exploitation of mires possible. Mires have become an economically valuable resource that is utilized in many ways, for example, in forestry, agriculture, and peat harvesting. In the process of building an affluent Finnish society, the purpose to increase productive forest land led to a situation where almost 60% of over 10 million Finnish mire hectares were drained for forestry and the change in mire landscape was huge. In that extremely intensive project from the early 1950s to the mid 1990s, the natural characteristics of mires were ignored.

After the drainage period, the forest certification has not allowed new drainage at all and that has led to a situation where nature has started to change more natural like. The productivity of mire forests is almost like in forests with mineral soils and many old mire areas are looking like boreal forests. In many areas, very intensively used and the passed over mire nature is taking back its natural character even if it can never reach its pristine nature. At the same time, it is more difficult to recognize the boundaries between human and natural mire space.

Conclusion

Nature is one of the main concepts in geography and the question of the idea of nature – and the history of the idea of nature – exists in the core of human geography. History of nature is not an established course of study in human geographical studies and human geographers have hardly anything to contribute if the field is understood as a concrete continuance of natural phenomena. History of nature, as exemplified by historical geography, directs human geography which is mainly based on the dualism of humanity and nature and in which human beings' relationship to nature is one of the most important research themes.

With the present environmental problems, such as the extensive issue of climate change, the question of the relationship between humanity and nature has come to the fore and the need for holistic research of environmental subjects has grown. There is also a need for interdisciplinary research that is founded on specialization and on capability to cooperate. For example, human geography concentrates on conceptualizing the terms of nature and humanity instead of focusing on exact natural sciences, such as geology, climatology, and ecology, which are capable of creating new information concerning the systems and rules of existing nature. Through dialog, it is possible to reach a better understanding of problems that cannot be solved by one discipline alone. It is important to remember that the recognition and definition of seemingly nature based environmental problems always take place in social processes.

See also: Culture; Culture/Natures; Historical Geography; Landscape; Nature; Nature, Historical Geographies of.

Further Reading

Castree, N. (2005). *Nature*. London: Routledge.

Darby, H. C. (1977). *Domesday England*. Cambridge: Cambridge University Press.

Glacken, C. J. (1967). *Traces on the Rhodian shore. Nature and culture in Western thought from ancient times to the end of the eighteenth century*. Berkeley, CA: University of California Press.

Lehtinen, A. A. (2006). *Postcolonialism, multitude, and the politics of nature: On the changing geographies of the European North*. Lanham, MD: University Press of America.

Sauer, C. O. (1965). The morfology of landscape. In Leighley, J. (ed.) *Land and life: A selection from the writings of Carl Ortwin Sauer*, (2nd edn.), pp 315 350. Berkeley, CA: University of California Press.

Semple, E. C. (1903). *American history and its geographic conditions*. Boston, MA: Houghton Mifflinm.

Smith, N. and O'Keefe, P. (1980). Geography, Marx and the concept of nature. *Antipode* 12(2), 30 39.

Vasander, H. (2006). The use of mires for agriculture and forestry In Lindholm, T. & Heikkila, R. (eds.) *Finland Land of mires. The Finnish Environment 23* (1st edn). pp 173 178. Helsinki, Finland: Finnish Environment Institute.

Williams, M. (1994). The relations of environmental history and historical geography. *Journal of Historical Geography* 20(1), 3 21.

Williams, R. (1980). *Problems in materialism and culture: Selected essays*. London: Verso Editions.

Nature, Performing

S. Fullagar, Griffith University, Nathan, QLD, Australia

Glossary

Binary Opposition Refers to the dualistic structure of western thought that makes one term (nature) inferior to and dependent on the other (culture).

Discourse Circumscribes the performance of meaning through language within particular power–knowledge relations.

Embodiment Acknowledges the biological and cultural dimensions of human subjectivity and knowledge of the world.

Hybridity Is the mixing of differences, the impure identity and space in-between.

Ontology Refers to assumptions about the nature of being.

Introduction

Through the influence of postmodern and post structural thinking over the last 20 years human geography has contributed to key debates about nature–society relations. The concept of performance has been central to the line of questioning that has critiqued the foundational ideas about nature as the opposite of human identity and culture. By exposing the ontological assumptions that have informed ideas about nature, geography has opened up different ways of thinking 'through' nature–culture relations within the spaces of everyday life, global networks, and knowledge practices (mapping, categorizing, research). As a result nature can no longer be naively thought of as an essentially given organic or inorganic sphere that resides beyond the reach of cultural meaning, representations, and activities that individuals 'do' to create their identities as human beings of various kinds. Performance, or more specifically performativity, following the seminal influence of Judith Butler's 1990 book, *Gender Trouble*, describes how individual identities are enacted and thus constituted through the discourses, normative practices, and embodied relations of everyday life. We exist by 'doing,' and hence meaning about ourselves and others is performed within the historical context of the present that preexists us as individual agents. For example, our gender as women, men, transgender, or other variations of identity is not simply assigned by an inherent meaning in our biology, but rather is performed through social norms and scripts that we engage with and live through, but do not necessarily 'choose' freely. By extension, our notions of 'human nature' are performatively constituted in relation to the otherness and difference of nonhuman nature with whom we interact (forests, animals, climate). Performance signifies a break with the assumptions that have informed both positivist/scientistic and humanist traditions of thought about human subjects and the natural world.

The notion of performance has enabled geographers to theorize the tensions informing how human and nature/nonhuman identities and relations are 'made up' through cultural processes and power relations involving language (discourses, narratives, metaphors) and bodies (emotion/affect, unconscious, senses). In this way, arguments move beyond conventional positions that (1) assume nature is different and given by a presupposed natural order (the essentialist argument about nature's purity that humans have degraded), and (2) that nature and our relations to it are determined by the social order (a political economy argument that nature is primarily a resource for global capitalism that we have become alienated from). Both of these perspectives assume that the 'truth' of nature can be 'represented' as objective knowledge that embraces a version of either scientistic or constructionist thinking. In the search for geographical truth there is a desire for self certainty in knowledge of nature as a thing, an object to be measured and rationally assessed. The performative turn in geography counters this rationalist tradition by acknowledging the profound uncertainty of 'knowing' and the potential for developing new ways of understanding and engaging with nature, and human identity within it.

Performance also moves geographical thinking beyond a reliance upon the intentional subject or agent by understanding identity formation as produced through the use of language and historically specific discourses of nature–culture relations. In this way, to perform humanness through a relation with nature is not simply the expression of a voluntaristic conscious choice although it does clearly involve decisions of value, ethics, and social distinction (e.g., walking in the forest, eating meat, or creating pollution). We perform a range of identities that are marked by social spatial location (e.g., gender, class, sexuality, disability, ethnicity) within a myriad of images and experiences of different kinds of nature (the inorganic environment, plant and animal life, visceral bodies). In order to understand how our identities are constituted as individuals, geographers used the concept of performance in different ways to explore how identity formation occurs through embodied actions and cultural locations. For example, researchers have examined how the ‘green identities’ of environmentalists have been performed through particular scripts, codes of conduct, and spatial networks that characterize individual choices within this middle class social movement. Other researchers, use performance to signify how the experience of nature is negotiated through feeling and ideas, as well as a staged event aimed at revealing the performativity of nature in context. The English allotment garden has been explored in relation to how everyday gardeners from different social backgrounds engaged with the cultivation of nature and through this embodied practice their own, and each others, identities. Some analyses also include artistic performances of nature and culture to unsettle ideas about how nature is practiced and recreated through senses. In this way the allotment garden figures within geography a site where knowledge is performed through different registers (theory making, artistic creation, gardening practice) and brings a range of senses (taste, feel, hearing) into play that counter the dominance of vision that has historically framed nature as landscape. These examples demonstrate how the notion of performance itself is not fixed within one theoretical perspective, and that geographers have crossed disciplinary boundaries to make nature ‘thinkable’ in new ways. This shift has also elicited critique and protestations about the limitations of performativity for understanding how nature is exploited, degraded, and regulated within global capitalism. Questions have also been raised by those working with ideas of performance about the problems of ‘uncritically’ applying them to geography, performance is often used by geographers in a way that privileges a humanist subject because the tensions between the subjects’ ability to exercise agency and the discourses that regulate these very actions are not resolved in Butler’s own work. Feminist writing has significantly contributed to the exploration of these ideas within geography.

The Cultural Turn toward Performance

Historically nature has been both the object (physical) and subject (human) of geographical knowledge. Radical and post structural theories have informed ongoing debates over nature and environment as highly contested concepts. Performativity has emerged as part of the cultural turn in geography, which has been influenced by broader post structural and feminist theories. The work of Foucault and Nietzsche, along with linguistic theorists such as Austin, has contributed to investigations of identity as an effect of performance. In this sense identity is reiterated through performative speech acts that normalize how we act in relation to nature, and ourselves as we articulate our position within and beyond it. Thus ideas of performativity offer a radical critique of any originary notion of interiority (essential human nature) and by extension open up questions about the cultural production of nonhuman nature as we have come to know it. This is not to deny the materiality of nature (that plants and animals exist beyond language) but rather to point out the effects of dominant discourses on relations between nature and identity (e.g., romanticism, economic rationalism, conservation). In this sense discourses constitute what and how we ‘know’ nature, rather than language naively being seen in terms of the representation of unmediated reality or truth. However, the emphasis on language has been subject to much critique by nonrepresentational theorists who argue that the materiality of nature and embodiment needs to be understood as having an intelligibility that exceeds capture in language. Nonrepresentational theory is further discussed below in relation to embodying performance.

Performative Resistance

Ecologists and philosophers have argued that within western nations the performance of human identity has been structured by a binary opposition of culture/nature. Within this relation nature is positioned as an inferiorized, feminized otherness – a background to the unfolding drama of human mastery, as can be seen in Val Plumwood’s 1993 book *Feminism and the Mastery of Nature*. Anthropocentrism is an extension of this way of thinking about nature as a reflection of human identified qualities and use values. Geographers have explored of how culture/nature relations and processes of othering also map across marginalized human others (e.g., indigenous peoples) who are positioned as closer to nature, the body, and the feminine. Hence, human and nonhuman natures are performed through an intersecting web of discourses and sociospatial materialities that constitute identities through particular and universal relations. Nature may be performed through local practices (e.g., consumption and place making) but is also mediated through the networks of globalization, postcolonial histories, and the market logic of advanced capitalism.

Understanding human identity as performative also brings into question the ways in which nature is constituted through everyday interactions that can unsettle or reinforce such oppositions. Extending this line of thinking geographical knowledge has explored how human–nature relations and identities are performed through particular spatial practices, classification activities, and symbols (e.g., travel, conservation, mapping etc.). Understanding the performance of human and nonhuman natures involves recognizing the repetitive iterations of sameness, as well as the subversive possibilities of different engagements, affects, and transformations of the present. In this way emphasis is placed on the resistances, spaces, and practices that subvert human centered thinking and actions that imply denying exploitative or nostalgic relations with nature. Performing nature in this sense involves paying attention to the generative possibilities and relations with nature that move us beyond the logics of degeneration and destruction. For example, the idea of the 'performativity of life' has been used to shift thinking beyond either a total socialization or demise of nature thesis. This involves considering how thinking differently about invasive species that cross global borders (ferals) poses new questions about how to understand disruption and change within human and nonhuman ecosystems without recourse to an ontological distinction between nature and culture or fixed borders that define belonging or exclusion.

Along with the notion of generative possibilities is performative resistance as a means through which counter practices and identities can be produced via different political mobilizations organized around threats to nature. Environmental protests on local and global levels embody the performative impulse that disturbs normalized explanations of problems and solutions. Other forms of engagement work through the desire to perform alternative modes that recreate everyday life for ecocitizens, individually and collectively (e.g., community gardens, bushland preservation groups, alternative power generation, and cycleway lobby groups). Exploring the performative also enables a critical, and subversive, analysis of the various permutations of nature based discourses that inform ecotourism ventures, environmentally friendly consumption, nuclear energy as a 'clean' option, and the environmental performance bottom line. Rather than emulate a romanticized notion of resistance with its humanist subject at heart (often embraced in studies of marginalized groups), a performative approach considers how power is exercised in multiple ways to reinforce dominant norms as well as to disrupt and open up other possibilities for living differently. Performativity seeks out the multiple, explores diversity, and questions any singular representation of the claims of social groups or nonhuman nature.

Embodying Performance

By exploring how identity is performed in relation to nature, theorists also employ the concept of embodiment. The body figures as a key site through which nature and culture, biology and sociality contribute to the formation of identity. It is also through our bodies that we come to experience and know the natural world, both through the senses (vision, touch, hearing) and the mediated images of culture (nature documentaries, magazines, blogs, websites). Hence, the lived, feeling body is also a site through which ethical relations to nature have been explored through the notion of becoming. Identity is performed through the body, which can be thought of as composed of multiple experiences, relations, and thus modes of becoming human. Nonrepresentational theorists have shifted the focus from the body as object to the performance of body practices of becoming, as in Nigel Thrift's 1996 book, *Spatial Formations*. Everyday practices that connect self and nature, such as rambling, riding, dancing, rock climbing, eating, etc. are performative embodied knowledges that offer the opportunity to think about what embodied experience brings to understanding. They are practices that enable the human–nonhuman relation to be explored in its specificity. For example, the desire to rock climb evokes questions about whether the climber is becoming part of the rock face or is the rock face appropriated by the climber in the quest for a masterful identity. These different modalities involve different ethical relations, self, and other identifications, and practices of care or use in relation to nature (e.g., using natural or artificial holds, navigating pristine bushland). The act of rock climbing is also performed within a broader assemblage of networks, land ownership, flows of information, and commodity exchange. Nonrepresentational approaches draw upon the philosophical work of Deleuze and Guattari, relational insights from actor network theory, sociologies of emotion and affect, as well as a focus on the productive nature of power in the work of Foucault, Benjamin, and de Certeau. This is a move away from traditions that emphasize the representation and capture of meaning where geographical knowledge seeks to fix truth and reiterate its own authority. By attending to the performative presentations, manifestations, and interconnecting experiences, nature–culture relations can be understood as continually in process. Embodied knowledge of nature can be explored in ways that seek to make apparent the taken for granted ways in which we know the world.

In extending this line of thinking Sarah Whatmore in her 2002 book, *Hybrid Geographies*, also employs postcolonial theory to explore notions of identity that move beyond oppositional relations with nature and embrace hybridity. Performing hybrid identities involves an acknowledgement of the diffuse, multiple forces that compose the self as both natural and cultural. Deleuze and Guattari's notion of becoming nature or becoming animal has been used to explore how the other is conjured up through networks and spatial practices (e.g., the global trade in nonhuman animals). Emphasis is given to the assemblages of desires, multiple relations, and embodied intensities or flows that connect human and nonhuman nature. This approach contrasts with political economy perspectives that presume that power works in largely repressive rather than generative ways and human/nonhuman actors move with knowable intention. Such a shift away from understanding the becoming subject (human and nonhuman nature) as an intentional agent has been crucial to opening up ways of understanding the otherness of nonhuman nature. While discursive frames may limit how nature is imagined there are possibilities for engaging with, and being moved by, the specificity and difference of animals, plants, and ecosystems in ways that escape capture in systems of representation. These relations of multiplicity can give rise to new alliances that move human relations with nature into new kinds of ethical possibilities for recognizing difference in the other and within the self. These moments of becoming nature may not be intentional or desired, for example, experiences of being attacked by crocodiles, sharks, or bears when engaged in ecotourism activities. They may, however, work to disrupt the positioning of animals as objects of the tourist gaze and bring into play the shared materiality and embodied vulnerability evoked by the near death experience.

Performing Ethical Relations

The question of how to perform different ethical relations has become central to a geographical reimagining of nature–society, human–nonhuman identities. Such a shift toward ethics also involves a critique of universalized moral codes that emphasize individualized rights or normalized practice over the more fluid, relational, and sociospatial contexts. In a move away from simply adopting a position of moral authority, a focus on the possibilities of transformative understanding and practices relating to nature has been identified within cultural geographies of hope. For example, researchers have considered the embodied emotion or affect of particular tree places in performing a more mindful relation with nature in everyday life. In contrast to much work that has explored animal geographies, reflection on the affects of trees on human nature moves debates beyond the realm of sentient beings to consider the subtle interplay between living surfaces and textures. The ethical relation can connect through the desire to care for, with, and through the experience of nature as an otherness that is both of and distinct from human identity. Such geographical mediations move beyond the binary opposition between nature and culture but do not abandon notions of specificity and connection. Instead these approaches offer different ways to consider how the boundaries between the human and nonhuman are figured and refigured within the spatial practices and flows of everyday life.

Conclusion

Performance has been central to the critique of ideas about nature as the opposite of human identity and society. By exposing the ontological assumptions that have informed ideas about nature, geographers have opened up different ways of thinking 'through' nature–culture relations within the spaces of everyday life, global networks, and knowledge related to both research and practice. Critical questions have been posed by performative approaches that ask, 'how do we understand human identity formation in relation to existing and emerging knowledges about nature as distinct and yet also connected (genetically modified food, climate change)?' What might be the effects of thinking about nature–culture relations, beyond binary oppositions, for the creation of different kinds of resistances, interventions, and hybridities? How are certain kinds of becoming nature, becoming animal practices evoked by embodied affects and emotions that move the self beyond normalized responses? These questions identify a tension that exists for many theorists in this area, between acknowledging the discursive mediations of performance and exploring the generative potential of embodied experience as it exceeds closure within language. For critics of performative ideas there remain questions about how power and domination are accounted for within globalized networks of capital that treats nature as a resource or externality. There are also concerns relating to the capacities of individual and collective subjects to bring about change that has structural effects on inequalities and exploitation of the marginalized, both human and nonhuman. What these debates have offered human geography as a discipline area is the opportunity to refigure its own borders and engage with a range of ideas about performing nature. This new discursive space promises no neat solutions to increasingly complex environmental and social problems, but it does generate possibilities for other ways of thinking through geographies of hope and critical awareness.

See also: Hybridity.

Further Reading

Butler, J. (1990). *Gender Trouble*. London: Routledge.

Clark, N. (2003). Feral ecologies: Performing life on the colonial periphery. *Sociological Review* 51(2), 163 182.

Cloke, P. and Jones, O. (2003). Grounding ethical mindfulness for/in nature: Trees in their places. *Ethics, Place and Environment* 6(3), 195 214.

Crouch, D. (2003). Performances and constitutions of natures: A consideration of the performance of lay geographies. *Sociological Review* 51(2), 17 30.

Deleuze, G. and Guattari, F. (1987). *A Thousand Plateaus*: *Capitalism and Schizophrenia,* Massumi, B. (trans.). Minneapolis, MN: University of Minnesota.

Franklin, A. (2002). *Nature and Social Theory.* London: Sage.

Fullagar, S. (2000). Desiring nature: Identity and becoming in narratives of travel. *Cultural Values* 4(1), 58 76.

Horton, D. (2003). Green distinctions: The performance of identity among environmental activists. *Sociological Review* 51(2), 63 77.

Instone, L. (2004). Situating nature: On doing cultural geographies of Australian nature. *Australian Geographer* 35(2), 31 43.
Jarmal, T., Everett, J. and Dann, G. (2003). Ecological rationalization and performative resistance in natural area destinations. *Tourist Studies* 3(2), 143 169.
Nelson, L. (1999). Bodies (and spaces) do matter: The limits of performativity. *Gender, Place and Culture* 6(4), 331 353.
Philo, C. and Wilbert, C. (2000). *Animal Spaces: Beastly Places*. London: Routledge.
Plumwood, V. (1993). *Feminism and the Mastery of Nature*. London: Routledge.
Szerszynski, B., Heim, W. and Waterton, C. (eds.) (2003). *Nature Performed: Environment, Culture, and Performance*. Oxford: Blackwell.
Thrift, N. (1996). *Spatial Formations*. London: Sage.
Whatmore, S. (2002). *Hybrid Geographies: Natures, Cultures, Spaces*. London: Sage.

Nature, Social

N. Argent, University of New England, Armidale, NSW, Australia

Glossary

Deconstruction Method central to post-structuralist thought. Seeks to destabilize so-called objective claims to knowledge and truth by revealing how multiple meanings can be made of texts advancing such claims, and also by identifying hidden ideological assumptions by authors within texts.

Epistemology Body of thought concerned with the development of knowledge, including the means and methods for judging 'facts'.

Hybridity Body of thought that emphasizes the co-mixing and co-constitution of so-called natural and social phenomena in, for example, networks of food production and consumption and also the physical composition of human bodies after transplants of animal tissue or prostheses.

Ontology Theory of what exists and what meaningful statements we can make about existing entities.

Post-Structuralism It refers to the work of French thinkers such as Derrida and Foucault. Against structuralist theories of society in rejecting totalizing, all-encompassing explanations of social and economic behavior, instead asserting that meaning circulates in uncertain and unstable ways via texts and images.

Introduction

The body of thought labeled 'social nature' emerged in the 1990s, posing fundamental challenges to the philosophy and practice of human geography, and of geography more broadly. The social nature literature highlighted the ways in which a trenchant and pervasive philosophical dualism originating in Kantian and Baconian thought had maintained its grip of geography – a discipline whose *raison d'etré* supposedly lay in the integration of the natural and social sciences – through its various paradigm shifts throughout the twentieth century. Dualist conceptions of nature and society as materially and logically separate entities of the world could be seen in the environmental determinism of early twentieth century geography; the later cultural geography school led by Carl Sauer; and the spatial science of the post World War II quantitative revolution. These anthropocentric and technocentric worldviews were challenged by the counter revolution of Marxian influenced political economy during the 1970s. Here, nature was brought back into human geography as a product of waves of capitalist accumulation. Nonetheless, this political economy stance drew criticisms that it retained an implicit dualist conception of nature and society (i.e., nature as something produced, and impacted upon, by society and economy).

At the heart of the social nature, literature was an ethical project aimed at discrediting philosophical dualism. This dualism was seen to contain the origins of many of the major environmental dilemmas facing current and future generations. Drawing upon an earlier body of sociological thought on 'social construction', writers such as Noel Castree and David Demeritt insisted upon seeing nature as socially constructed and, more importantly the social and the natural as essentially interpenetrated by each other. That is, it is impossible to completely understand human society without an appreciation of the many ways in which it depends upon, interacts with, and defines the natural world. Similarly, social nature theorists insist that we cannot make sense of the biophysical world and its many component entities and processes without an understanding of how these have been defined by and incorporated into social, cultural, and economic systems. In order to develop a reasonable understanding of this deeply philosophical body of thought, it is vital to be clear about what we mean by the term 'social nature'. What does 'social construction' mean in this context, and what 'nature' – given that the word is a homonym with many subtly different meanings – is being socially constructed?

Social construction is a form of critique first developed by sociologists in the late 1960s and early 1970s. Challenging the dominant structuralist/functionalist approach to identifying and discussing social problems in contemporary sociology, social constructionists argued that perceived social pathologies such as domestic violence, graffiti, public loitering, and the like were widely considered as aberrant behaviors at least in part due to the claims made about these acts by third parties, for example, those affected by the vandalism or academics studying the causes and incidence of criminal behavior. For instance, the act of vandalism did not automatically register in the broader public's consciousness as a social problem; it became one via 'knowledge claims' made about vandalism by concerned groups within society (e.g., local government, business owners).

The social constructionist approach is today perhaps more popularly associated with feminist scholarship. Second wave feminists have focused on, *inter alia*, the social construction (by patriarchy) of sex and gender and have critiqued the persistent and taken for granted direct correspondence between sex (e.g., female) and gender (e.g., woman) that forms a key structuring principle of societies the world over. Post structuralist feminists, such as Judith Butler, draw attention to the many mundane, day to day acts performed by women – modes of dress, public participation and comportment, domestic roles and the like – that help create, conform to, and therefore reproduce an ideal of womanhood. Further, these forms of behavior and thinking cannot be traced back to dif ferences in chromosomal make up, hormones, or genital assignment. Rather, they must be traced back to the ex pression of unequal social powers.

A gendered body can be seen to be a complex amalgam of the social and physical (or natural) in space, rather than just a biological organism, or an individual woman or man. It is this determined antidualist perspective, with its em phasis on hybridity – a co mixing of the natural and the social – that has been adopted, adapted, and imported into the literature on 'social nature'.

In what ways was this critical perspective brought to bear on nature? 'Nature' has a number of alternative meanings and it is important to clarify which definition we are arguing is socially constructed. Three primary definitions can be recognized: (1) external, (2) intrinsic, and (3) universal (see **Table 1**). The external, material realm of nature is the version of nature that we are perhaps most familiar with. In Marxian thought, this is sometimes known as 'first nature', the primordial world prior to the emergence of humans and the material re source base of their sustenance. This is contrasted with 'second nature', which incorporates human beings and the products of their labor processes, including cities, farm, transport networks, and the like into 'first nature'.

The second definition comes to us from the process of abstraction where the natural (as in definition 1 above) and the social worlds are conceptualized as being con stituted by a small number of key structural factors or principles. These are considered indispensable elements of a phenomenon's existence and can be contrasted with a larger number of nonessential or contingent factors that may or not be found depending upon the circumstances. For example, if we were interested to find out what the nature of the capitalist rental market we would isolate in thought those factors and entities that are considered as having causal powers essential to the operation of the process under investigation (see **Figure 1**). Here, 'nature' refers to an essential quality of something.

The third and final definition of 'nature' refers to the largely inbuilt forces which drive the natural and social worlds. This definition is bound up in the oft heard statement, 'It's only human nature' or 'It is only natural (that something should happen in the way that it has)' or, conversely, that some behavior or outcome is 'unnatural'. Which definition of nature has the socially constructed nature literature been most concerned with? The short answer is 'all' but in the following we will consider ex amples of the approach applied to each of the three types of nature just outlined.

There are two main philosophical 'camps' in the social nature literature. Both approaches hone in on positivist, naïve, realist accounts of supposedly natural processes, laws, and relations, seeking to de naturalize taken for granted arguments and supposedly 'iron laws' regarding nature and its intersections with human society. However, each school of thought entails quite strongly contrasting ontological beliefs and epistemological commitments.

Table 1 Alternative definitions of 'nature'

i	*External nature:* the biophysical realm, constituted by humans and nonhuman flora and fauna and nonbiological matter.
ii	*Intrinsic nature:* the causal powers, or essences, that make 'things' what they are, and not something else.
iii	*Universal nature:* as in *ii* but refers more generally to the bundle of largely instinctive capacities or reflexes that act as driving forces, e.g., 'It's only human nature for someone to abhor torture', or 'I didn't really think about rescuing him from the fire it was just second nature'.

From Castree, N. and Braun, B. (eds.) (2001). *Social Nature: Theory, Practice, and Politics.* Malden, MA: Blackwell; Demeritt, D. (2002). What is the 'social construction of nature?' A typology and sympathetic critique. *Progress in Human Geography* 26(6), 767 790.

Social Construction-as-Refutation

The first school of thought has been labeled by David Demeritt as 'social construction as refutation'. This fo cuses on the deconstruction of taken for granted con cepts and bodies of thought in a fashion akin to the feminist example considered above. Perhaps the earliest sustained contributions in this field emanate from Marxian scholarship in which rural and urban environ ments were argued to be the effective production of capitalism.

An instructive piece of research that falls within this field is Goodman, Sorj, and Wilkinson's analysis of the evolving industrialization of Western developed nation agriculture. Goodman *et al.*'s historical analysis of the industrialization of developed world agriculture regarded the application of industrial production techniques to farming as following two major trajectories: 'appro priationism' and 'substitutionism'. Both strategies aimed at overcoming the barriers posed to capitalist accumu lation by nature: the spatial fixity of land and the seasonal nature of land based production.

Appropriationism refers to the various industrial tech niques applied to farming which have come to displace either human labor (e.g., mechanical harvesting displacing the labor intensive chain of reaping, threshing, bagging, and sowing) or the biophysical production process of the

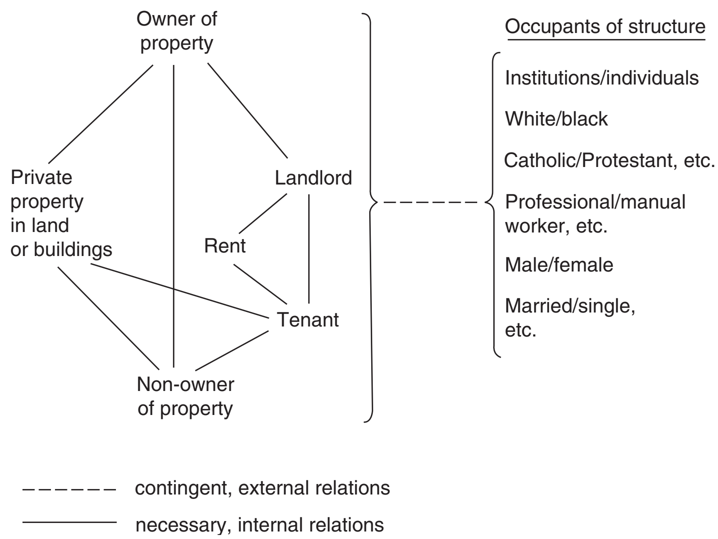

Figure 1 Nature ii: Causal factors in the structure of a capitalist rental market. From Sayer, A. (1992). *Method in social science: A realist approach*. London: Routledge.

agricultural commodity itself (e.g., the introduction of chemical fertilizers replacing farm animal manure).

By contrast, substitutionism involves the effective rendering down of agriculture to little more than a provider of inputs to the industrial production of food and fiber. Margarine production in the late 1800s was an early precursor of substitutionism, with plant oils and other compounds combined in various ways to produce a relatively cheap (for consumers) and cost efficient (for manufacturers) alternative to butter at a time when ani mal fats were expensive. In the contemporary era, more advanced scientific techniques have permitted the manufacture of surrogate foods (e.g., meat analogs) via the use of single cell proteins and soya isolates. However, it is the more recent revolution in biotechnology – par ticularly genetic engineering – that offers a further quantum leap in the full capital penetration of farm production and the undermining of land based agri cultural production. These methods open up promising opportunities for even more efficient production, and for custom grown foods that better meet human and animal nutritional needs.

Construction-as-Philosophical Critique

The second major approach – construction as philosophical critique – is associated with more postmodernist and post structuralist stances, and con centrates upon the role of language and discourse in framing the interpretations of the 'real world' by scien tists and how these interpretations become regarded as truths. Unlike more philosophically materialist work, though, scholarship in this category tends to emphasize the discursive creation of supposedly natural phenomena and causal processes. Also, this type of critique is also generally more forthright in its advocacy of nondualist modes of thought.

Four distinct versions of critique in this category can be distinguished, though the differences between each type are, in many cases, actually very subtle. For sim plicity's sake, this article only deals with three of these: the sociology of scientific knowledge, actor network theory, and post structuralism.

Sociology of Scientific Knowledge and Actor-Network Theory

The sociology of scientific knowledge (SSK) and actor network theory (ANT) literatures are closely related yet distinct, sharing an interest in revealing the human toil involved in the creation of scientific knowledge. Given that many aspects of the natural world, such as magnetic fields, changes to the ozone layer, and genetics are es sentially unobservable and hence unknowable except through their humanly perceived effects, SSK theorists are concerned with the actual labor involved in dis covering these truths. Researchers in this field also focus on revealing the power relations that become entangled in contested claims within cognate areas of science, and how these struggles influence the practice of scientific discovery. Borrowing from the pioneering work of Donna Haraway, writers in this field insist upon the historical specificity, social situatedness, and partial character of scientific knowledge.

A useful illustration of the research foci and approach used in SSK is Kay Anderson's studies of race, though this is not to say that Anderson would consider herself an adherent to this school. Much of Anderson's work since the 1990s has been concerned with exploring the role of science and scientists in drawing, and maintaining, the boundaries between the natural and the social (and cul tural). Arguing against a stable, universal conception of nature, she criticizes the supposedly scientific basis of the common sense notion that human races are substantially different from each other (e.g., as *Homo erectus* is deemed to be different from modern *Homo sapiens*). Further, though, in wanting to take scholarship in this field beyond all dualistic conceptions of, for example, human/animal, or society/nature, Anderson aims to denaturalize 'race' and shed light on the scientific notion that humans are different and also superior to animals.

The closely related field of study, ANT, shares the same ontological and epistemological beliefs as SSK but is perhaps better known for studies which 'flesh out' and map in concrete terms the complex co mingling of 'actants' (i.e., the many and various human and nonhuman constituents included in an actor network) in everyday life. One of the central philosophical, and ethical, stances of ANT is its agnosticism toward the possession and execution of power (and, hence, agency) in such networks: human agency is not privileged in this research. There are three distinctive elements of ANT. The first, 'hybridity', refers to the heterogeneous composition of networks (humans and nonhuman biological beings and inorganic artefacts) as well as the conception of power relations which does not, *a priori*, privilege human subjectivity. The second element, 'collectivity', relates to the essentially collective and relational nature of network activity: for the network to be sustained through time and space each actant to some degree depends upon all of the others. Third and finally, 'durability' covers the relative internal strength and resilience of network relationships, espe cially where these are stretched across vast distances and contrasting cultural contexts. In all of these senses, then, ANT is almost inherently a part of the socially con structing nature body of knowledge.

An instructive example of the ANT approach, and one in which these three elements are clearly evident, is a study of a river restoration project in southern England by Sally Eden and colleagues. This study focused upon attempts by a group of concerned academics (the River Restoration Project (RPP) team) to renovate the River Cole such that its hydrology and ecology more closely resembled its pre industrial era state. The authors pre sent their study as a staged production in which each of the actors (see **Figure 2**) – the RPP team; engineers; consultants; local 'stakeholders'; the National Trust; the National Rivers Authority; relevant maps, plans, and computer models; the riverine flora and fauna; and, of course, the River Cole itself. These entities are neces sarily drawn in, or enrolled into, the network by dint of their internal attributes and capacities (their 'natures' as in definition 1 above) and by the way in which the res toration project evolved. The network therefore became inherently 'hybrid'.

Although the authors are careful not to judge the success or otherwise of the restoration effort – to do so would be logically incompatible with ANT – they do nonetheless show how potential problems and conflicts between the many actants bound together in the network are negotiated through and mediated. Through an it erative process of stating, clarifying, and re negotiating its aims, the RPP – as the organizational center of the network – helped to ensure the 'durability' of the net work, at least until the new stream channels were built and scientifically and esthetically evaluated.

The nonhierarchical presentation of the network as a collectivity accentuates the ANT perspective on the

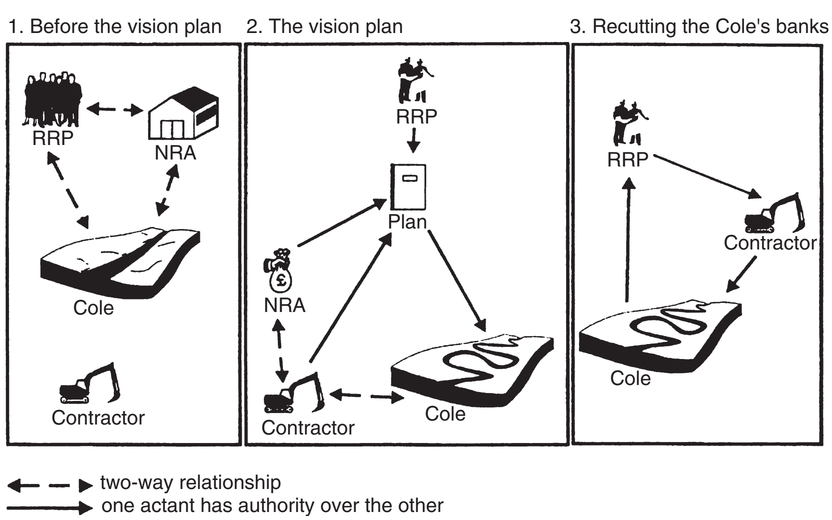

Figure 2 The actor-network of the River Cole's restoration. From Castree, N. and Braun, B. (eds.) (2001). *Social Nature: Theory, Practice, and Politics.* Malden, MA: Blackwell.

ontological equivalence of actants regardless of the hu manness or nonhumanness. But it also shows how some of the key human actants in the network almost sub consciously conceive of agency and power as an attribute that is definitely not restricted to human subjects. In a revealing passage, the RRP General Manager discusses the need for ongoing monitoring and possible engin eering works after the initial river channel works and ascribes agency and autonomy to the river:

> … we knew what we would have to do would be to react in the following spring after construction to what the river told us. We knew the river would tell us an awful lot in the first winters of flooding, we weren't [to] play God, we could go so far into giving the river the sort of shape and regime that we felt it would be happy with.

The final approach considered in this subsection is that of post structuralism. As already noted above, post structuralist 'takes' on nature (definitions 1, 2, and 3), in contrast to the Marxian standpoint, insist upon seeing society/nature relationships through the lens of a re lational and discursive ontology in which the existence of, and our ability to 'see', entities is dependent upon the existence of other entities, like that of the maps and computer models developed by the RRP and consultants in the discussion of ANT just considered. In other words, post structuralist 'social nature' writers insist that understandings of nature (and of society nature inter relationships) are the product of socially , culturally and historically situated discourses. Importantly, different social groups will produce distinctive and often com peting discourses because of their members' backgrounds and characteristics (e.g., ethnicity, birthplace, class affili ation). This entails that meanings of 'nature' are rarely stable and always relational. Again, perhaps the best way to gain an appreciation of the major knowledge claims of a post structuralist stance is to briefly review work done in this vein.

Bruce Braun and Joel Wainwright explore con structions of nature in the coastal forests of British Columbia (BC), Canada as the effects of social and political power. In particular, they reveal how a new ideology and practice of forestry management of 'sus tained yield' emerged after World War II and thus shaped a new vision of the forests. However, rather than critique the doctrine of 'sustained yield' contained within the landmark 1945 Sloan Report on the basis of its long term impact on the ecology of the forests, or on its position regarding the merits of private versus public ownership of forests, the authors focus on the key operating prin ciple contained within 'sustained yield' – the 'working circle' (see **Figure 3**).

Within the pages of the Sloan Report the expert knowledges of cartography, statistics, biogeography, and accounting were implicated in rendering visible the BC forests as little more than a resource. However, the ap parent widespread influence of the report was also dependent upon a receptive colonial capitalist mindset; itself preconditioned by generations of landscape and resource evaluation, together with the demarcation and forceful occupation of lands in which the intertwined physical and cultural geographies of First Nations peo ples were either obliterated or overlain by European settlement.

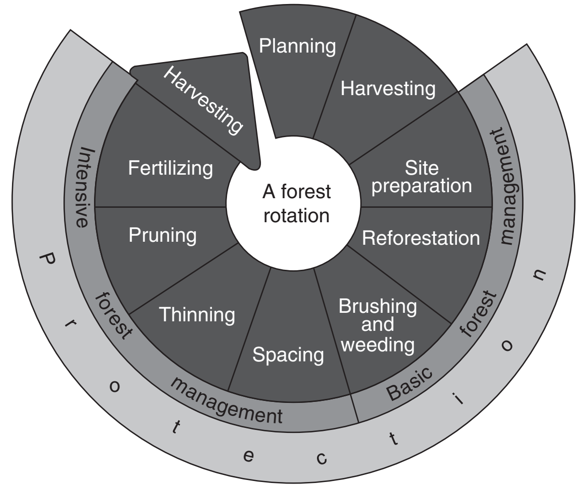

Figure 3 The 'working circles' model of 'sustained yield' forestry management in British Columbia forests.

Drawing upon Derrida's notion of (hau)ntologies, the cultural geographies of First Nations peoples could be seen to form a constitutive outside – an alternative ontology which is marginalized but nevertheless provides a definitional standpoint against which a dominant ontology (e.g., 'sustained yield') draws its power and meaning – to the instrumentalist rationality contained within Sloan's report. This alternative ontology came to broader public awareness from the 1970s as a loose grouping of environmentalists, local First Nations peo ple, and other local residents voiced their protests against the orthodoxy of 'sustained yield' across numerous sites in BC, culminating in the successful campaign of the Tla o qui aht during the mid 1980s.

Critiques of the Social Nature Literature

Two broad types of criticism of the social nature litera ture and approach can be identified: one of a political/ realist nature, the other more purely philosophical/ logical.

The first set of criticisms tackles 'head on' the anti essentialist and anti foundationalist claims of social na ture writers. Here, writers have attacked what they regard as the dangerous political and ethical relativism of such research. For them, the deconstruction of scientific research findings confirming, *inter alia*, ongoing biodiversity decline, or enhanced global warming simply risks the continued exploitation of the environment by industry, and actively undermines organized political efforts to advance new, less exploitative modes of development. In other words, according to this view, social nature theorists risk playing into the strategic hands of corporate capital, who advance a 'business as usual' approach to these potential global dilemmas. It is important to note, though, that the majority of social nature writers generally eschew a radically idealist ontological stance regarding nature and society. Although most would argue that what we regard as nature is not only heavily imprinted by human society (as in the Marxist 'production of nature' thesis), and that our ways of perceiving and speaking about society and nature are also social constructions in the sense of being concept dependent and culturally influenced, very few would hold that there is no independent material entity such as biophysical nature.

The second deeper philosophical criticism to be made of social nature goes to the core of its ontological and epistemological standpoints. The two are obviously tightly articulated in that epistemological arguments depend fundamentally upon what is agreed to exist. As already noted, social constructionists are generally forced to navigate a philosophical knife edge in arguing that supposedly objective truths about nature are partial, situated knowledge while simultaneously holding to the notion that a material nature cannot exist outside of our impressions of it.

An example of this delicate balancing act can be seen in 'natural' disasters research. Some have claimed that such disasters are really social constructs, for without the presence of people in the affected regions, no disaster would exist. For critical realists this is a logically flawed argument. To be sure, a powerful combination of, *inter alia*, poverty, international debt, inadequate housing and public infrastructure, and national and regional dependences upon land based production all help to explain why people come to be concentrated in environmental settings which are both predisposed to extreme environmental conditions (e.g., monsoons) and poorly equipped to deal with the human dimensions of a 'natural' hazard. Laudable as the political and ethical claims of social constructionism are, in this respect they commit an epistemic fallacy. This involves confusing ontological with epistemological issues. That is, what is thought to exist is an ontological matter but judgments about what are causal factors, and what are merely contingent factors, in existing things are concerns of epistemology. Critics argue that social constructionists who confuse these two dimensions therefore inadequately conceptualize the objects of their research and thereby are likely to make mistaken claims about cause, effect, and contingency.

Conclusion: Toward Hybridity?

By the early years of a new millennium, the social nature school appeared to have become bogged down by the, at times, torturous philosophical arguments just considered, together with the almost inescapable problem of how to write about society and nature in a nondualistic fashion without starting with the two as separate entities which then have to be rethought as interpenetrated by each other. Some of its chief protagonists acknowledged that while social nature studies had generated a vibrant and insightful set of research initiatives, the philosophical limits of its major tenets had been reached. With the array of problems attending binary conceptualizations and dualist ontologies now firmly laid out, for some the way forward and past the philosophical impasse that characterized debates between realists (of whatever stripe) and social constructionists lies in the not dissimilar study of hybridity. Heavily influenced by SSK and ANT, as well as the writing of Donna Haraway on the cyborg, hybridity presupposes the ever present mingling of the human and the natural, explicitly acknowledging the agency of humans and nonhuman flora and fauna in their interminglings within, for instance, food chains, colonization processes, transplant surgery, and zoo management.

See also: Actor-Network Theory/Network Geographies; Feminism/Feminist Geography; Nature-Culture; Natures, Postcolonial; Philosophy and Human Geography; Physical Geography and Human Geography.

Further Reading

Anderson, K. (2001). The nature of race. In Castree, N. & Braun, B. (eds.) *Social Nature: Theory, Practice, and Politics*, pp 64 83. Malden, MA: Blackwell.
Braun, B. and Castree, N. (1998). *Remaking Reality: Nature at the Millenium*. London: Routledge.
Braun, B. and Wainwright, J. (2001). Nature, poststructuralism, and politics. In Castree, N. & Braun, B. (eds.) *Social Nature: Theory, Practice, and Politics*, pp 41 63. Malden, MA: Blackwell.
Castree, N. and Braun, B. (eds.) (2001). *Social Nature: Theory, Practice, and Politics*. Malden, MA: Blackwell.
Demeritt, D. (2002). What is the 'social construction of nature?' A typology and sympathetic critique. *Progress in Human Geography* 26(6), 767 790.
Eden, S., Tunstall, S. and Tapsell, S. (2000). Translating nature: River restoration as nature culture. *Environment and Planning D: Society and Space* 18, 257 273.
Goodman, D., Sorj, B. and Wilkinson, J. (1987). *From Farming to Biotechnology: A Theory of Agro Industrial Development*. Oxford: Basil Blackwell.
Haraway, D. (1991). *Simians, Cyborgs, and Women: The Reinvention of Nature*. New York: Routledge.
Whatmore, S. (1999). Hybrid geographies: Rethinking the 'human' in human geography. In Massey, D., Allen, J. & Sarre, P. (eds.) *Human Geography Today*, pp 22 39. Cambridge: Polity Press.
Whatmore, S. (2002). *Hybrid Geographies: Natures, Cultures, Spaces*. London: Sage Publications.

Nature-Culture

O. Jones, Countryside and Community Research Institute CCRI, Cheltenham, UK

Glossary

Actants Entities that have effects upon other entities through combination effects. Thus a person and hammer combining in the act of the human hammering other objects comprise an actant. The point is that the entities that 'act' in the world are not solely restricted to the human, or indeed to the animate, living being, but are now enlarged to include many other nonhuman 'things' enrolled into actions.

Actor-Network Theory An approach often associated with the work of Bruno Latour and John Law which sees any and all worldly formations as produced by integrated networks of differing actors, or rather actants, which include humans and nonhumans. It makes no ontological distinction between nature, humans, and technology.

Affect Systems of the body such as emotion, balance, senses, which underpin life and allow interaction.

Agency The capacity to act creatively.

Dialectics An approach to philosophy and critical thought which takes contradicting ideas about a matter in hand and seeks to synthesize a new understanding from the examination of the opposing positions.

Dwelling A theory about life-in-environment and time which foregrounds a sense of being among the world. Drawing from phenomenology and especially Heidegger, it emphasizes the co-constitution of subject and object, self and environment. It suggests the condition of being human is one of being in and among the world – rather than one of a separate thinking self.

First and Second Nature The idea of first nature is that of pure nature before any human interference. Second nature is that which emerges as humans make changes to the environment.

New Dialectics A version of Marxist dialectics pioneered chiefly by the geographer David Harvey.

Social Constructivism, Social Constructionism Approaches that argue: first, that societies shape the world in which they act; and second, that we (humans) can have no direct access to reality independent of society, so our understanding of the world is in fact shaped and created by our social structures.

Theoretical Hybridity An attack on dualistic, pure categories, instead suggesting that most bodies or systems are derived from heterogeneous sources. It is a deliberate reappropriation of a biological term to invoke the fusing, or in more charged terms the miscegenation, of categories previously held apart.

Introduction

'Nature culture': we have 'nature', we have a hyphen, and then we have 'culture'. The job of the hyphen is to reunite these two realms. They have been crudely and violently divided by modern knowledge which might be written as 'nature/culture'; two realms separated by a slash which represents a whole range of ways in which they have been forced apart. Reuniting them is a very big task for such a small symbol, for the division has been, and largely remains, a ubiquitous foundation of modern knowledge. This dualism's brutalist architecture is clearly visible in a number of forms; for example, in the division of the social and natural sciences, the denial of agency in nature, and the exclusion of nature from dominant historical, political, and ethical formulations.

Geography can be regarded as an unusual (and promising) discipline because of the way that it bridges between these two realms, dealing with both 'the human' and 'the physical'. But this structure within geography is itself a symptom of the nature/culture worldview. Indeed, (sub)disciplinary, theoretical, and methodological specializations within 'human' and 'physical' geography often widen that divide rather than the reverse. However, geography remains very well placed to play its part in the process of ontological healing which is needed to reunite our divided and damaged world. Calls for this reunification are being made (in different ways) by some eminent figures within geography, for example, David Harvey, Doreen Massey, Nigel Thrift, Sarah Whatmore, and Margret Fitzsimmons.

The nature/culture division has profound implications for not only how we understand the world but also for how we act in and upon it. Some of these implications appear to be very troubling and problematic for the very well-being of life on Earth. A range of environmental thinkers argue that the perilous situations that many sectors of global society seem to face, in terms of ecological, economic, and cultural sustainability, are, in part, driven by difficulties generated by the nature/culture divide. The stakes are very high in this philosophical naming game. But can the difference between 'hyphen' and 'slash' really be that important? In answer, we might draw upon the environmentalist philosopher Mary Midgley, who compares philosophy to plumbing! If a bad smell develops in your house, or ominous damp patches appear in the ceiling, you call a plumber who might pull up the floor boards and investigate the hidden systems which service the obvious and visible facts of taps, toilets, sinks, and so on. The problem might well be some blockage in the hidden workings of the system. If worrying signs develop in the way society is working, we need to 'pull up the floorboards', look beneath everyday understandings and practices for old conceptual infrastructures which may have gone wrong. The nature/culture divide is one such – a blockage which needs to be cleared.

The nature/culture divide is then a pervasive prob lematic of modern knowledges. There are a whole host of vexed questions about the risks, and the rights and wrongs of how culture(s) engage(s) with nature in nu merous ways, such as resource extraction, food pro duction, genetic modification, power generation, nature conservation, understanding and practices of bodies, identities, landscapes, and cities. All of these engagements weave across the so called divide, operating through networks of organic beings, technology, science, industry, economic, politics, and culture. Important approaches in geography and related subjects are therefore now arguing that nature and culture are not, and have never been, divided in terms of everyday lived life on our beautiful blue green brown gray white planet. There is a move away from thinking about nature/culture as ontologically discrete categories (as still represented in much geo graphical writing). Instead, the world and the many for mations which compose it, be they continents, cities, industries, habitats, bodies (human, nonhuman organic, nonhuman technological), are all hybrid assemblages of, and in, heterogeneous entanglements or networks. These entanglements criss cross the so called divide with such intensity, and moment by moment and place by place frequency, that its only meaning stems from its falsity. Our knowledges, disciplines, methodologies, politics, and moralities need to be adjusted to respond to this chal lenge. Alongside the imperatives and concerns that derive from the problematic nature/culture/worldview, there are new forms of charm, beauty, and 'enchantment' to be found in nature culture understandings and practices, as Jane Bennett vividly sets out.

Below, we will consider how the division was not al ways thus; some of its characteristics and consequences; and how it is now being called into question for a number of reasons and in a number of ways. We consider two ways in which the dualism might be tackled – by making everything 'culture' or everything 'nature'. But these 'one way' moves are also called into question as they deny the complexity and heterogeneity of life. Then we gather together seven 'two way' approaches to treating nature culture; actor network theory (ANT), hybridity, new dialectics, new ecologies, dwelling, animal geog raphies, and new readings of place. We then inspect four key challenges flowing from these various approaches: agencies of nature culture, ethics of nature culture, politics of nature culture, and disciplines and method ologies of nature culture; and finally, we address the idea of differences in nature culture and some of the spaces in which they are enacted, including bodies, cities, coun tryside/wilderness, and spaces of biodiversity.

The Great Divide

The nature/culture divide is driven by a number of forces. First, modern knowledge systems like to divide, define, classify, and explain. This has been the very job of science, and was quite possibly a necessary step in making science and rational knowledge as we know them. The nature/culture divide is one of the fundamental divisions upon which modern knowledge has been based. It has been seen fit to understand the natural as one set of relatively discrete systems, for example, through biology, chemistry, physics, and geology. These have processes operating within them which can be under stood as functioning independently of social life, there fore they can be studied without considering the social. In an attempt to understand the complexities of the 'natural world', it seems to make sense to isolate and specialize.

Conversely, and somewhat understandably, humans tend to think of themselves and their societies as rather different from nature, and rather special. This specialness rests largely on our apparently unique capacities – our high levels of self consciousness, our use of language, rational thought, developed knowledges, tools and tech nologies, and the subsequent production of culture. The dazzling complexity of both individual and collective human life, it seems, also merits a set of specialized studies – the social sciences and humanities, from which natural processes are generally excluded. Politics and ethics have been thought of predominately as stuff of this realm – of people and for people. If nature is a set of mechanical, predetermined, unreflexive processes (albeit highly complex), then politics and ethics are not 'of' nature, or 'for' nature, but can only be 'about' nature; as resources, for example.

Cartesian Dualism

Descartes is often cited as the founder of modern phi losophy and science. His contributions to mathematics and logic were part of the foundations that he laid. So too was the famous mind/body dualism. This did not exactly split humans and nature apart, but it split mind, thought, and language apart from the nature of the human body and certainly from the rest of nature. Mind, thought, and language were the defining characteristics of self and humanity, and had to be understood as such. The rest of nature operated through mechanical, automated, material systems. Other, related, dualisms such as subject/object and agency/structure, where the cultural is the separate, primary actor and nature (the rest of the world) an outside, passive recipient of action, also became part of the modernist architecture of knowledge. Ecofeminist thinkers, for example Val Plumwood and Vandana Shiva, suggest that the Cartesian scientific revolution was a particularly masculinist view of knowledge, self, and world in how it sets up nature as a separate, external world to be objectively studied, controlled, and manipulated. This divided worldview, some argue, has also been buttressed and built upon by certain religions, particularly the Abrahamic, monotheistic religions (Judaism, Christianity, and Islam), as they too have tended to separate 'man' from nature.

Reaping the Whirlwind?

There is a philosophical imperative here which argues that this bifocal reading of the world is deeply flawed. And this imperative spills out of the realm of philosophy and into some of the most pressing practical questions global society faces today. The emergence of environmentalism and ideas of sustainable development in the latter decades of the twentieth century came about as a result of increasing evidence that global society has been on an unsustainable trajectory since (at least) the Cartesian revolution, and the subsequent industrial/capitalist 'revolutions' which are now increasingly globalized. It might thus seem perverse and dangerous to try to dissolve the idea of (separate) nature as the apparent global environmental crisis deepens, and global society is belatedly waking up to this predicament. The specter of anthropogenically induced climate change is just one of a number of profound concerns which are novel to the human story thus far. Deforestation, the overfishing of the oceans, the degrading of soil and water resources, pollution, and the overall decline in biodiversity are other very pressing matters. Nature is in trouble. Should we be attacking it conceptually as well?

For some, the answer is that we need to dissolve the nature/culture divide in order to avert disaster and to better understand our relationships with and within biophysical systems. Our conceptualization of nature as separate from, 'and subordinate to', culture has allowed corrosive entanglements, often emanating from industrial cultures, to ensnare and overwhelm the beings, spaces, and processes which comprise 'the natural realm'. Nature and culture become imagined and studied as separate, and so much of the traffic which has routinely crossed the 'divide' goes unnoticed and unpoliced. Fresh waves of intense and novel mixings of 'social' and 'natural' elements, which are poorly understood and which still go unpoliced, may throw up further major problems which are global in reach. And, there is a very real concern that 'nature' will not go quietly. Extreme anthropogenic disturbances to biosphere systems will possibly throw up new patterns and processes, such as the melting of the ice caps, sea level rises, and realignments of climate zones and ocean currents, which will not be at all sympathetic to many settled, cultural, and ecological arrangements. A number of environmental writers are pointing out that this has already started: for example, Elizabeth Kolbert and Mark Lynas.

Learning the Long Game!!

The increasingly detailed evidence of past geophysical processes now being gathered by environmental archaeologists, when put alongside histories of human societies, shows that what were once hard to explain rises and falls of power, success, and failure in social/political systems, can sometimes be attributed to long term variations and cycles in biogeological systems such as shifting climate zones, or the fall out from extreme events like super volcano eruptions. Nature and society have always been much more entangled than our histories have envisaged. Nature is not merely a passive stage for human history but rather one of the lead players in the show. Megasystems such as the Sun–Earth–Moon interactions, ocean currents, and exchanges between ocean and atmosphere shape the very conditions within which the human social exists. Time is a vital aspect of all this. The temporalities of cosmology and geomorphology are extremely hard to read in terms of social time, and it is only quite recent discoveries which allow us to read these huge rhythms and cycles in which 'social' life is enframed. To better grasp the entanglements of nature culture, we need to readjust the temporal horizons and sensitivities of our understandings.

It Was/Is Not Always Thus!!

Finally in this section, it is worth noting that the nature/culture divide has not always been in place, and in some nonmodern societies was never established. Excellent histories of environmental thought, for example, by David Pepper, take us back to the inception of the dualism and to what was in place beforehand where nature culture conceptualizations were very different. We are told of the 'great chain of being' where humans, animals, plants, and natural materials were in one overarching, hierarchical family with god at its head and angels taking their places too. In nonmodern societies, and in other theological constructions of life, again nature and culture are not divided but part of grand cycles or narratives of creation, life, and death.

Two simple points are to be made. The first is that under these differing conceptualizations, differing practices, politics, and ethics of nature society emerged showing how basic understandings of how the world 'is' then underpin actions, politics, and ethics. The second is that recognizing these alternative models shows that the nature/culture divide, and the view of nature that it creates, is not 'natural' or inevitable. Understandings of nature culture relationships have changed, can change, and will change again. The nature/culture divide is showing distinct signs of wear and tear as it comes increasingly besieged from a range of perspectives. Various branches of knowledge – philosophical, environmental, social theory/science, the natural sciences – need to play their part in questioning the 'divides' that seem to have served us so ill, and in trying to unravel the myriad nature culture entanglements which in fact make the world.

All Nature or All Culture?

One way to dissolve the divide is to simply say 'all is nature', or, 'all is culture'. The divide disappears. On the face of it, good cases can be made for both these positions. It is reasonable to think that there was a time before humans evolved that all was natural. Humans and their culture have evolved from nature and thus can be seen as extensions of nature. They are just very particular forms of nature. Even the most cultural, artificial things one can think of, say, the Manhattan skyline, is composed of natural elements which have been assembled by human animals. Perhaps they should be treated and studied as such – as natural objects, not that dissimilar from the towering termite mounds of Africa. Alternatively, all could be said to be culture. The very idea of nature, our knowledges of biology, chemistry, physics, and so on are all productions of human mind, thought, and language. In this sense, everything we know we have created. We can only think of nature once culture has invented it. Before Newton had his encounter with the apple, gravity simply did not exist as a force to be reckoned with. These 'one way' approaches to dissolving the divide are discussed below. They are important because they point to the huge constructive power of both nature and culture. After these are tackled, approaches which offer more 'symmetrical' views are considered.

All Is Nature: New Life Sciences

It could be seen as ironic that some of the more recent insights and discoveries of natural science have challenged the great divide and make a strong case for all being nature. They show the great extent to which human life and culture has emerged from, and exists as little more than faint flickerings within processes of biochemical existence. Foremost among these are developments from Darwin's theories of evolution which place the cultural firmly as an extension of the natural. This is not only in terms of how we evolved but in how many cultural forms stem from more basic animal functions. Donna Haraway pithily asserts that biology and evolutionary theory have reduced the line between humans and animals to a 'faint trace'. Theories of affect as far back as Spinoza, as well as more recent psychological and neurological insights, all show that rationality, language, and consciousness are not isolated from the body and bodily processes, such as the unconscious and the emotional, but rather are emergent from it. At this point, culture becomes little more than a thin film of extension on these natural processes. Supporting and extending these interconnections is the recent decoding of DNA, apparently showing the mechanisms by which the forms of all life unfold. The line between humans and not only animals, but all living things, and between determinism and free will becomes increasingly faint.

Yet, further layers of interrelation can be wrapped around this unity of living things. All life, and the very stuff of our planet, has been spun out of the processes of cosmological evolution. The popular music star Moby sings 'we are all made of stars'. All the complex atomic elements which go to make up the Earth and life on it were produced by the nuclear fusion of distant, ancient stars. These exploded, scattering the elements into vast stellar dust/gas clouds which, eons later, compress and form planets; then, perhaps, life on them, then, perhaps, culture from that life. Is the cultural hence merely vivid and fleeting blossomings emanating from vast and deeply embedded natural processes? This raises the question of 'determinism'. Is all present and future action shaped by a determinable progression from past situations – 'natural determinism'? Is human nature defined by our environment – 'environmental determinism'; or our biology – 'biological determinism'; or a combination thereof? Perhaps, all we can say here is that these are some of the most profound and challenging questions in science and philosophy, and that many cases are made for varying degrees of freedom of action at any given moment, albeit set within the shaping contexts of unfolding processes.

All Is Culture: Social Construction and the End of Nature

In human geography and the social sciences, rather predictably perhaps, it has been more fashionable to see culture framing nature rather than vice versa. This is the insight of social construction, and we think about it here because it forms another sustained attack on the nature/culture dualism. In this case, the divide is collapsed by overwhelming the apparently natural with flows of constructive power and practice from the realm of the social. Popular accounts such as Bill McKibben's *The End of*

Nature depict culture overwhelming nature so that no part of it remains untouched. The atmosphere has been changed; even the frozen wildernesses of the poles con tain traces of pollution. Nature has lost its purity and independence, and that is what made it 'nature' in the first place. The argument is that nature and culture once did stand apart, and that separation was key to their very being, but culture has now corrupted nature in qualitative and quantitative terms. McKibben subsequently qualified his argument, but it rests on restating the problematic dualized view of nature as different and separate from culture.

Social constructivist work, as geographers such as Noel Castree have pointed out, instead argues that 'nature' never was or is, simply natural. Rather, it is a collection of powerful cultural ideas which become understood as 'real' or 'natural'. If nature is created by culture, then the cre ation is commonly denied, or forgotten, and the product seen as outside culture. This process, for some, constitutes the very 'essence' of nature, which is really 'discursively constructed all the way down'. William Cronon shows that such processes even apply to wilderness, the space where a quintessentially pure nature is supposed to exist. He shows that the idea of wilderness was a construction of modern society and its troubled relation with its en vironment, as industrialism and urbanism developed, along with the colonial expansion of European culture into the North American west (and elsewhere). To see these places as wilderness was to deny the complex his tories of the landscapes which had for millennia included the presence and effects of aboriginal populations.

Noel Castree and Bruce Braun point out that demystifying nature in this way is intellectually and politically radical and liberating. They further add that "there is no generic social constructionist position, only specific modalities of social construction" emanating from differing cultures in differing places and times which generate a whole range of contested and contesting natures. Rather than one untouched, unchanging 'nature', this work points to the creation of a hybrid 'social nature' which is constructed and can come in a number of forms, ones that Castree lists as, 'knowing nature', 'engaging natures', and 'remade nature'. These are about the ways in which we (humans) understand what is natural, and how these understandings are played out as we encounter and remake that nature. This becomes an ever more pressing question as politics play out in bodies and identities, and as scientific and technological inter ventions remake nature right down to, and beyond, the bone. The phrase 'artifactual natures' has been coined to represent this idea of nature which is 'purposefully en gineered'. Geographers such as David Demeritt stress that we must ask what types of artifactual natures are now being developed, by whom, and with what consequences?

Even advocates of constructionism admit that we are now in a postconstructionist era in geography and be yond. Critiques of social nature can accuse it of what Castree calls, hyperconstructionism, the idea that there is no stable, object, material nature 'out there'. Geographers such as Chris Philo and Sarah Whatmore suggest that constructionism has resulted in geographies of nature which empty nature of its vitality and agency to an extent where (at worst) the world is rendered as an exclusively human achievement in which 'nature' is 'swallowed up in the hubris of social construction'. Demeritt argues that conflicts over ideas of social nature arise partly out of the great complexity and varieties of meanings attached to the components of the concept; and he seeks for a clearer position by pointing to ideas of 'constrained construction' where there is a material process, say, climate change, and then a whole range of scientific, political, and cultural forces and perspectives complexly and discordantly clustered around it.

Despite its apparent decline, Neil Smith's assertion that the insight of social construction cannot be bypassed or set aside seems hard to gainsay:

> The central and undeniable insight is that the authori tative appeal to reality as the ground of truth claims is always filtered through the social muslin of represen tations gathered into discourse (no matter how liberal and permissive the muslin may be) and that no kind of purely extra social authority is available for arbitrating the shape and dynamism of nature (Smith, 1998: 273).

This view of the power of thought, language, and culture at the heart of constructive discourse does need to be heeded, but in ways somehow still allowing nature, life, and power (and, of course, a pre/extra human his tory). The 'one way' approaches reviewed so far reveal the power of both nature 'and' culture, but only at the expense of the other. The seven alternative approaches to which we now turn, in differing ways, seek more 'sym metrical' ways of dealing with nature culture.

Reweaving the Torn Asunder World: Seven Approaches to Nature-Culture

What is in a hyphen? Or, in what ways are the relatively settled, separated ideas of nature and culture being linked or re entangled? Do we need to see nature and culture as somehow extant, but thoroughly inter penetrated? Or, do we need to go further still, to a point where we think about the world, and facets of it, in ways which do not even recognize the old division and old terms at all, but rather, in ways which only allow for many other, smaller differences? What kind of categories, languages, methodologies, politics, and ethics are needed to construct such a new worldview? This section outlines seven approaches that variously seek to tangle and to combine categories of nature and culture.

Actor-Network Theory: A World of Networks

Advocates of actor network theory (ANT) such as Bruno Latour have mounted a concerted attack on modernist modes of classifications. They insist that the persistent separation of the world into cultural (or social) and natural denies the entangled 'nature' of everyday material life and allows dangerous entanglements to flourish unnoticed and unpoliced. ANT has sought to develop a 'symmetrical' view across the previously inscribed nature/culture/technology divides. This symmetry dissolves not only the nature/culture dualism, but also that of subject/object and agency/structure. Indeed, Latour has recently written that we should now completely bypass the old dualisms. We should no longer treat them as starting points for discussions (as so many do) or even as grounds of debate and attack. We should get on instead with tackling the world of actualities – networks or assemblages which contain unique, complex, and changing populations of people, organisms, things, substances, and processes. ANT hence argues that all manner of things (as many as you can imagine) are variously entangled together in specific formations or networks in the making of the world. These networks produce any given achievement in the world, be it education, power generation, food production, politics, music, and so on. Four points can be made about these networks which underpin the relational world view more generally.

First, the networks that ANT envisages make up the entirety of the unfolding fabric of life. They come in many forms and scales. They are unstable and prone to breakdown, and lots of effort goes into stabilizing and repairing them. Many networks fail or just moulder away, others constantly emerge. Second, networks make space rather than trace across previously present 'empty space' which is waiting for life to fill it. Straightforwardly Euclidean notions of space are rendered topological – space is crumpled, lumpy, folded. For example, Latour says that a journey on a modern train across Europe is an entirely different time–space matter than a perilous hike through thick jungle. Old notions of space – place as bounded locality, distance and nearness, local and global – are all problematized. Offices and computers in distant world cities may be in effect closer to each other than they are to areas of poverty which might be physically just down the road.

Third, all elements of the network are actors, or actants, which have agency or actancy (both latter terms are intended to de center the human subject). Actants' identities and qualities are not innate but are relational – emergent from the network into which they are bound. Fourth, power needs to be read as located throughout the network rather that just at key centers. ANT is particularly interested in devices which connect and can effectively transmit agency/power from one part of the network to another. How are actants enrolled into a network? How are they held in place? What manner of translations and translating devices are needed to allow differing types of actants to communicate and thereby maintain network stability? The historical argument is that the nature/society divide was a creation of modernity which ignored the true conditions of connected, networked, relational life. If 'modern' means nature/society divided, then 'we have never been', and never will be 'modern', as Latour famously put it. This is not just some esoteric philosophical argument. Latour points out that our divided vision of the world has made us blind to the traffic which criss crosses between the apparently separate realms of nature/society, and also to the monstrous formations which can thus form.

ANT has been questioned for its lack of interest in uneven power relations and thus in victimization. Nigel Thrift, while acknowledging its insights, has suggested that it fails to deal with ideas of place. Nick Bingham with Thrift suggest that it misses 'the sizzle of the event', that is, the complexities of encounter between entities. ANT also has a strongly technological inflection which seems to under represent organic living things. Some have questioned whether it is reasonable to treat different types of actants in networks – for example, machines and animals – in the same (truly symmetrical) way. The life of animals here seems to be denied, almost as in social constructionist approaches. As in social construction, however, the power of the central insight of this approach needs to be carefully heeded as we contemplate nature culture assemblages that make up the world. It seems undeniable that everyday processes unfolding in the world do involve a whole host of actants from right across the spectrum of existence working together (ideas, texts, chemicals, machines, organisms, processes, finances, and so on) – all being assembled together in forms of 'heterogeneous engineering', a key ANT motif. The study of these networks requires new approaches which inevitably break out of settled disciplines and disciplinary 'regions', and requires new suites of methodologies (although specific methodologies will remain useful as forensic techniques of investigation).

Hybridity: Impure Life

Hybridity relates very closely to ANT, but it also brings certain nuances and emphases which deserve specific attention outside, or along with, the 'strictly network' approach. A dictionary will tell you that hybrid means something like, 'anything derived from heterogeneous sources'. Often, it refers to plants or animals which are interbred rather than pure in origin. The heterogeneity, the impurity, of 'theoretical hybridity' is an attack on dualized and purified identities and categories, notably that of nature/culture. ANT networks are clearly hybrid in their combining of a variety of actants into assemblages, but hybridity adds to this vision in two ways. First, it stresses that individual bodies are never pure. It becomes too easy to see networks as assembled from elements of nature, technology, and culture whose form and identity still retain the foundations of the modern settlement. Second, hybridity stresses more open, other, spontaneous, and unruly forms of becoming than ANT sometime does. This emerges in part from hybridity's intersecting of biophilosophy with feminist theories of fleshy living bodies, as in the work of Whatmore.

One famous example of a hybrid body is Haraway's 'cyborg', who is part machine, part woman, and part animal. She/it deals in and through instinct (nature), language, body, and technology. This might seem a bit fanciful to some. But at much more prosaic levels, we humans are all cyborgs: people with hearing aids, glasses, and a range of medical implants are to a small, but significant effect, hybrid entities. Their capacities are changed; nature (the body) and technology blur into each other. The same could be said for all the clothes and tools that we use every day: these cannot but shape our very capacities. Gilles Deleuze and Felix Guattari offer hybrid views of conjoined bodies formed of human, nonhuman, and technology making new entities with new capacities for action. For example, a rider, a horse, and riding technology (saddle, stirrup, and reins) together become a new entity which has new power and space/life making potentials (think of the Mongol empire).

The hybrid geographies of Sarah Whatmore focus not only on the ecologies of life, and the material relations through which they are articulated, but also on how these are contested and remade through knowledge, politics, governance, and ethics. Interdisciplinary experimentation between previously isolated approaches, and the scrambling of 'expert' and 'lay' knowledges, are thus to the fore in her recent research projects. Geographers such as John Murdoch and Paul Cloke use the term hybridity (and co constitution and co construction) as a broad approach of which ANT is one part. They argue that hybrid approaches focus our attention on the ways in which the previously held apart worlds of nature and society routinely and inevitably mix.

At the heart of the notion of hybridity, and ANT, are questions about the balance between relationally and individuality. Whatmore is keen to increase the pressure on the notion of the autonomous individual (human and nonhuman) and the mind body, subject object, self other divides, in order to build knowledges, politics, and ethics of affective intercorporeality. This is moving toward ideas of 'geophilosophy' which are best exemplified by the work of Deleuze and Guattari. These authors build a new set of conceptual tools and categories which can address the complexities, hybridities, and fluidities of the entangling world. New spatial languages such as smooth and striated space, lines of flight, flows and grids are developed; new terms for being – becoming animal, bodies without organs – are developed. In terms of logics of knowledge and logics of networks, rhizomic replaces arboreal hierarchy. Unlike trees, rhizomes are nonlinear and nonhierarchical; they can grow and branch at any point, in any direction. Some proponents of the hybridity approach feel that it becomes problematic in the way that everything becomes, somewhat perversely, the same; in that, everything is just made up of disparate elements which have come together in some way or other. The way around this problem is to take notice of the differences in the precise ways that specific hybrids are formed, noting crucial differences in their spacings, timings, and capacities. This noticing of difference begins to divide the world up again, but not along the old line(s) and not on the old scales, and we will return to this point later. We now consider five further approaches in which nature/culture is re rendered into nature culture.

New Dialectics; Certain Kinds of Networks (Capitalism)

Marxism has had a central interest in nature and nature culture relationships from its inception, not least in how capitalism alienates society from nature and renders nature as property. Latterly, various attempts have been made to project Marxism into ecologism (or vice versa), seeking to make 'green Marxism'. At the heart of the Marxist approach is dialectical materialism in which nature is embedded as a generative force. Particularly as developed by David Harvey, new dialectics attempts to extend these key trajectories and embrace a more fully relational, hybrid, view of the world. As such, it makes a sustained attack on both the nature/society dualism 'and' the dualism of space and time. Here, as in ANT and hybridity, there is a relational view of actants. They are not separate, or in possession of innate, stable identities; rather, their natures are relationally inscribed. But Harvey, from his Marxist base, suggests that capitalism infuses the majority of networks, or relations, which have constructed the modern world, arguing that this fundamental process needs to be confronted. Humans and nonhumans in 'socio ecological formations', as Castree sums up, "become the 'arteries' through which an invisible process of ceaseless capital value expansion operates." His materialist formation then sees not only society making nature in its image but transforming nature, and nature in turn dialectically reworking society. There is a resistance from natural elements as they are forcefully enrolled into capital accumulation networks, which sets up the dialectic dynamic. Castree offers fish farming as an example of this kind of dynamic between economy and nature. It seems to be a logical market response to the challenges and expenses of sea fishing and fish stock crisis and resulting quota systems, but the many well documented problems with fish farming re veal natural elements not simply conforming to capitalist logics but instead, subverting, resisting, and outflanking the impositions placed upon them.

Margaret Fitzsimmons supports this focus on indus trial capitalism which over the last 200 years has, as she puts it, ushered in "massive ecological change … writing over the landscapes and lifeways of other human cul tures." She points to the four basic interactions among living organisms: competition and struggle; adaptation into niches; collaboration and cooperation; and environ mental transformation and how, through the study of their entanglements, Harvey's purpose is to find

> … a way of depicting the fundamental physical and biological conditions and processes that work through all social, cultural and economic projects … in such a way as not to render those physical and biological elements as a banal and passive background … [so that] … somehow the artificial break between 'society' and 'nature' must be eroded, rendered porous and eventually dissolved (Harvey, 1996: 192).

From this base, Fitzsimmons, after Harvey, advocates academic interaction between the natural and social sciences, and solidarity with active socio ecological projects that address both justice and difference.

Castree feels that the impact of the new dialectics has been limited by the turn away from Marxism in geog raphy and the social sciences. That said, even the new dialectics is thought by some to slip back inadvertently toward, or never properly to shake off, a dualized view of nature/culture embedded in dialectical reasoning. An other criticism is that relations are defined almost in evitably as conflict, whereas other approaches are keen to seek out positive sum relationships between, say, econ omy and ecology. Nevertheless, this perspective offers a focus upon the key role of industrial capitalism in shaping the spaces and networks of the world in which human and nonhuman elements are relationally articu lated. Felix Guattari, in his book *The Three Ecologies*, makes it plain that it is capitalism which is denuding cultural, psychological, and ecological diversity to the extent that we are witnessing 'ecocide'.

New Ecologies: Mobile, Impure Natures

Since it was coined by the nineteenth century German biologist Ernst Haeckel, ecology has been regarded as one of the natural sciences and a branch of biology. It focuses on the study of organisms and their relationships with their environment (including other organisms). Thus mostly, but not exclusively, it has focused on the nature side of nature/culture. Given its focus on rela tionships, however, on occasion it has moved toward a thoroughly relational view of life and drawn in other disciplines in order to effectively map life, meaning that it has, in effect, bumped up against the limits of narrow disciplinary foci and the nature/culture worldview. Ideas of new ecology have emerged which question established ideas of separate nature 'and' various assumptions about that nature, particularly assumptions of equilibrium – that nature is in harmony and balance – and that nature is a pure system which in contact with society 'inevitably' corrupts, distorts, and lessens. Instead, there is an as sumption of nonequilibrium dynamics, spatial and tem poral variation, instability, complexity, uncertainty, and even chaotic fluctuations as being more the norm for biophysical systems. This is thought to be so for both apparently natural and apparently human impacted spaces.

Along with a greater emphasis on disequilibria in new ecology, there is a (related) assumption that systems are more open than closed, with exchange and flux between systems occurring constantly. There is concern for spatial and temporal dynamics developed in detailed and situ ated analyses of 'people in places', using, in particular, historical analysis as a way of explaining environmental change across time and space. There is a rejection of the view of nature as separate realm into which human life intrudes, and any correspondingly simple idea of 'first nature' being superseded by 'second nature'. New ecol ogy hence regards ecosystems and habitats as systems which may be more or less open or closed, stable or volatile, certain or uncertain, fleeting or enduring, spa tially focused or diffused, and in which a whole host of agents are interacting, including humans. The challenge from new ecology, according to (Castree, 2005: 235), is "to regard human actors as always already 'part of' complex and changeable biophysical systems" whose actions do not necessarily 'corrupt' or reduce nature. New ecology offers a more hopeful, but still challenging, view that nature–human interactions can be positive. It is not hard to find at least some obvious examples: Bill Adams, for instance, points out that some of the richest habitats in the UK in terms of biodiversity are adaptive, such as chalk downland grazed by domestic animals, and coppiced woodlands. Here, economic production has gone hand in hand with the production of rich (in bio diversity terms) natures. These hybrid ecologies where both culture and nature seem to flourish have often been relatively neglected in scientific and political agendas.

New ecology also moves the purview into terrains less often regarded as natural such as cities which, with their mosaics of spaces such as gardens, parks, allotments, derelict land, transport network ‘verges’, and landscaped spaces such as car parks, can offer a much richer ecology than intensively farmed, but apparently green, rural landscapes. Thus, even the quintessentially social spaces of the city can be ecologically rich, and places of a new cosmopolitan politics of human with nonhuman conviviality. Recent studies have shown that the biodiversity hotspots of Germany are not to be found in the countryside or nature reserves, but in cities such as Berlin. Given the rise of the urban world, this is a vital realization, and it also offers wider lessons in the possibilities of human and nonhuman flourishing. One of the challenges of new ecologies, however, is that a recognition of complexity and uncertainty shows how (simple) prediction, management, and control are unlikely, if not impossible.

Dwelling: Human and Nonhuman Life as Emergent, Relational, and Emplaced

If new ecology sees human action as part of the natural world, then perspectives on dwelling tend to see nature as part of humanity. Dwelling is about addressing life in terms of being in the world as an active, embodied, immediate, yet also temporal (enduring) relational process. It is about living body in environment (space and place) which is sensing, responding, engaging, exchanging, remembering, knowing, and doing. Dwelling offers a more organicist view of the world than that of ANT. Here, we have the world constructed of the many co minglings of nature culture actants within the everyday dwellings of particular worldly locations. Sarah Whatmore and Steve Hinchliffe have discussed how dwelling is about the ways in which “humans and other animals make themselves at home in the world through a bodily register of ecological conduct.” Springing from the later work of Martin Heidegger and related phenomenological approaches, dwelling offers a ground from which life (human and nonhuman) can be rethought away from Cartesian derived dualisms. Dwelling differs from social construction approaches through its stressing of the physical, the relational, the sensed, the orchestration of body and space/environment.

The work of anthropologist Tim Ingold is pivotal to the current upturn in interests in dwelling in geography, sociology, and beyond. For Ingold, dwelling is a perspective that treats the immersion of the organism person in an environment or life world as an inescapable condition of existence. The world “continually comes into being around the inhabitant, and its manifold constituents take on a significance through their incorporation into a regular pattern of life activity” (Ingold, 2000: 153). He suggests that nobody and nothing has ever crossed the ‘/’ of nature/culture because it simply does not exist. Dwelling is often explained as a switch from a ‘building perspective’ to the ‘dwelling perspective’. The former rests on the notion that human thought and action is somehow isolated from the world and thus in a position to impose on the world, literally, to build upon it ideally generated blueprints. The latter points out that any form of life emerges from the world, and that there is never a gap through which thought/practice can completely free itself. The two are always entangled; meanings do not overlay the world but are immanent in the contexts of engagements which perform it moment by moment. This turn has also been mapped as a switch from ‘thinking space’ to ‘thinking place’. Both human and nonhumans are given active roles in this interplay of body and environment.

Ideas of dwelling are potentially bound up with the notions of home, local, and (rural) rootedness. This sits uneasily with the apparently mobile, speeded up, stretched out nature of much contemporary (urban) life. But in such life, time–space deepened experience remains inevitable, and new forms of repeated encounters still abound, such as commuting to and from work. Examples of life without enduring relationships with places such as home (of some kind or other), workspaces, or cities, and with patterns of rhythm and repetition, are nonexistent. Dwelling then inevitably leads to an emphasis on temporality/process in the consideration of landscapes. To capture the relational, emergent, creation of space, Ingold coins the term ‘taskscape’, where the spatiotemporal patterning of the environment takes relatively settled physical form through repeated practice. But this is to make the point that all landscapes/places are taskscapes, in which case the distinctions between them are ultimately dissolved.

Sociologies of nature suggest that dwelling overcomes conflicts between ‘realist’ and ‘idealist’ approaches to nature and environment. Adrian Franklin places dwelling at the heart of a new anthropology of nature which pushes toward an animated, turbulent vision of the world unfolding in a burgeoning far reaching (in time and space) interfolding of processes. He suggests that the basic building blocks of this anthropology are “unmediated perceptual knowledge, practical experience and knowledge of the world, the technologies that link humans and nonhumans, the aesthetic and sensual composition of experience and the cultural choices that are made in reference to these” (Franklin, 2002: 71–72). Uses of dwelling in geography by Paul Cloke and Owain Jones problematize a risk that we might replace nature/culture dualisms with a dwelt/authentic–undwelt/inauthentic life dualism which is present in Heidegger, which can lead to a dualism between bounded space and network. John Wylie further adds that dwelling must confront the novel, the fleeting, the singular, and the moment in emplaced becoming, as well as the longer patterns of familiarity, practice, habit upon which dwelling has tended to focus.

Animal Geographies

One important development in notions of dwelling is the use of the concept to consider the lives of nonhuman animals. The dwelling of animals as well as of humans, and the continuities between them, are important threads in Ingold's work. Geographers such as Thrift use the related notion of Von Uexküll's life world to stress the particular intelligences and spatial practices of dwelt (animal) life. More generally, the somewhat oxymoronish focus on animals within 'human geography' emerges from a recognition of intense mixings and intimacies between humans and nature in the particular forms of animals, and the myriad spatialities of differing animals' presences in differing societies. The thrust of this can be traced in two edited collections from a decade ago, one by Jennifer Wolch and Jody Emel and the other by Chris Philo and Chris Wilbert, which study how animals are implicated in the social in a vastly diverse range of ways; how the acknowledgement and study of these implications has been marginal(ized); and how the questions and implications posed range from the ecological, through the political, to the ethical.

In animal geography, David Matless and various colleagues have developed ideas of nature culture by considering what they call 'animal landscapes'. Expanding examples of hunting and other forms of animal–human interactions historically embedded in place and landscape at a differing range of scales, they show how relations with animals can be a powerful part of human identity and place formation. These processes vary from example to example, and will almost certainly contain conflict as animals are constructed and acted upon differently. Matless uses the example of hunting and wild fowling in the Broadlands (UK) where various historical renditions of the landscape, from topographical writings to museum displays, can chart these 'animal landscapes' and their changing formations as new actors (e.g., new species) and new cultural ideas are folded into the mix of landscape. In another study which uses otter hunting as an example, hybrid and animal geographies are combined and developed in a number of ways. The geographical context of encounter, the nature of the animals involved, the class and culture of the human actors, all fold into highly dense, and highly specific, geographies of culture–nature–animality–place.

Places as Entanglements

Place is a complex and somewhat fuzzy term, but in various guises it has a track record of scrambling, to some extent at least, the nature/culture divide. Places, in terms of somehow distinguished local spaces, have obvious material, physical 'and' cultural dimensions. There has long been an interest in the interplay of these dimensions in the very formation of place. Humanist geographers such as J. Nicholas Entriken have sought to develop ideas of place which also scramble simply dualized notions of objectivity and subjectivity, but beneath all this scrambling the nature/culture divide often tacitly remains as place beds down into being 'humanized' space and nature. A concerted effort has hence been seen to rejuvenate approaches to place in human geography which are keen to jettison any notion of them as simply bounded, static, social spaces, all too easily demarcated on a map by a line. The aim instead is to propose places as temporal processes where all manners of trajectories – of people, nonhumans, economies, technologies, ideas, and more – contingently settle out into distinctive local patterns. These are always changing, yet enduring for now, until forced apart, perhaps through new flows of forces from elsewhere, as they always remain networked into the wider world. Thus these views try simultaneously to hang on to both a topographical, located, dwelt view of life and a topological, networked view of life.

As Ash Amin and Nigel Thrift summarize, "places are best thought of not so much as enduring sites but as moments of encounter, not so much as 'presents', fixed in space and time, but as variable events; twists and fluxes of interrelation." Both the 'social' and 'natural' elements are on the move in these views of place. Doreen Massey, in her major text written in 2005, is at pains to stress the dynamic nature of nature. It is not just a case of social flows whirling through, and tangling with, more fixed grounds of nature. Indeed, if a long enough view is taken, even land itself can be seen swirling across the surface of the globe through processes of plate tectonics. Many other natural processes operate in flows, rhythms, and velocities equally unamenable to everyday human apprehension, yet also which remain highly various, such as ice ages, sun spot cycles, and long term weather patterns. These we need science to read. Other rhythms such as planetary movements and the seasons, and corresponding movements/rhythms of oceans, animals, and plants are amenable to 'ordinary' sensing and memory. All these velocities and rhythms flow together into ongoing makings of place.

All manner of actants thus bring their agency to the formation of place, which is one key outcome emerging from eddies or entanglements in trillions of intermeshing flows patterning space–time. Cities, houses, offices, parks, cars, desks can all be seen in this way. Stephan Harrison *et al.* argue just this; "[a]ll kinds of things can come together in the world and, in that process of encounter and settling down into at least a short term equilibrium, they can creatively produce new kinds of organisations that are greater than the sum of their parts." Thrift incorporates a whole range of approaches, including dwelling and ANT into what he calls 'ecologies of place'. To this, he adds yet further entanglements of memories, longing, affect, and even hauntings. Place can be a receptacle which holds together all of these rich entanglements of the social, the natural, the material, the imaginary, the past, and the present.

Four Key Challenges of Nature-Culture

As we have already set out, the modern settlement of nature/culture sits at the very foundation of modern knowledge along with related dualisms such as subject/object and agency/structure. It is inevitable that, if these foundations are successfully undermined, the ramifications will be far reaching and a lot of rebuilding work will be required. Understandings of agencies, ethics, politics, disciplines, and methodologies need to be urgently reworked as we change nature/culture to nature culture.

Agencies of Nature-Culture

In the modern settlement, agency is seen as a distinctly human capacity. It is understood as the capacity of the human subject, using thought, language, and free will, to act on the world. Even higher order animals have been excluded from ideas of agency – they act by complex instinct alone in ways from which they cannot creatively depart. Below them, lower order animals, plants, and materials are deemed purely mechanical followers of set processes. This privileging of human agency has been a ubiquitous factor in modern knowledge and a key architecture of the nature/society divide. Margaret FitzSimmons and David Goodman claim that it has been commonplace in social theory to ignore the specific 'agency' and 'materiality' of nature. We have discussed agency already in the approaches to nature culture outlined above, but it is important to note that they all, albeit to different extent and in somewhat differing ways, challenge the privileging of human agency. As this is a key architecture of nature/culture, it has become a prime target for those seeking to move to nature culture.

Environmental philosophers such as Val Plumwood and Arne Naess have long argued that agency needs to be given back to nature. In Plumwood's terms, "once nature is reconceived as capable of agency and intentionality, and human identity is reconceived in less polarised and disembodied ways, the great gulf which Cartesian thought established between the conscious, mindful human sphere and the mindless, clockwork natural one disappears." Put at its simplest, 'we' humans would not be here without (among many other things) the sun (and other stars), the moon, the millions of microbes which occupy our bodies, the plants, and animals with which we share the world – without, in short, nonhumans of many a hue. Life itself is a creative force. Surely, this is a form of agency?

There are however challenges to thinking about the agencies of nonhumans alongside those of humans. The answer is not to deny humans agency, but rather to rethink what agency is and how it is enacted. There is now a concern to redefine agency in ways which allow it to arise through nature society assemblages, and which recognize the specific, embodied, agencies of beings and things other than human. This is not to deny the uniquely distinctive capacities of humans, but rather to expand the notion of agency. Whatmore (1999: 26) suggests that agency should be seen as "a relational achievement, involving the creative presence of organic beings, technological devices and discursive codes." Latour argues that ideas of agency need to be disaggregated in order to account for the differing ways in which the global population of things can creatively act. According to him, "there might exist many metaphysical shades between full causality and sheer non existence [in terms of agency]: Things might authorise, allow, afford, encourage, permit, suggest, influence, block, render possible, forbid and so on" (Latour, 2004: 226). He feels that "no science of the social can even begin if the question of who and what participates in action is not first opened up, even though it might mean letting elements enter that, for lack of a better term, we call non humans"(Latour, 2004: 226).

Two further brief points need to be made about agency in nature culture. The movement of attributing agency to nonhumans is in a way an extension of the idea of human agency. It is the capacity of individuals to act. Latour talks of 'things' acting. The more relational processes set out above seek to destabilize this view by saying it is the relational interaction which generates agency, not from individuals alone but between things. We also need to recognize the agency of processes, for example, of photosynthesis and of tidal fluctuations. In a world of process, these are the major creative forces which flow through things and in which things cause turbulence. The second point relates to the idea of place set out above and is also returned to below in the discussion of sites. If the world is made up of mixings, a vast interrelating set of mixings across space and time which are always in flux (in a whole range of velocities), then the contingency, the 'turned upness' of those mixings has an agency. What we are bumping up against in the course of our everyday life makes a difference. This is the agency of chance and of difference.

Ethics of Nature-Culture

The opening up of questions of agency goes hand in hand with the opening up of questions of ethics. The environmental philosopher Holmes Rolston notes how the common assumption persists that ethics are for people. People are both the subject and the object of ethics, in the sense that only humans are deliberative moral agents and also that humans have ethical obli gations only to other humans. Under modernism, ethical consideration has largely operated only up to the sup posed boarder of nature, or stopped very soon after it. Environmental political philosophies such as ecofemi nism and deep ecology, however, have long contended that the application of ethical concern solely or princi pally to the culture side of the nature/culture dualism alone is deeply unethical in itself, and a key driver of unsustainable practices. Shiva argues that the scientific revolution in Europe transformed views of nature and in doing so removed all ethical and cognitive constraints against its violation and exploitation. Questions of animal rights form a small but significant disturbance at this nature/society boundary.

Reconfiguring ethics to work throughout the entire body of nature culture is then a primary and again urgent task. Lawrence Buell argues that the environmental crisis involves a crisis of the modern imagination, the amelio ration of which depends on finding better ways of imagining nature and humanity's relation to it. This re imagining requires, as Jim Cheney puts it, an "act of considerable moral imagination for those raised in the heart of the monster, the Western dualism of moral in siders and outsiders." We need forms of ethics which can follow entanglements, relationality, and hybridity through nature culture, and which are thoroughly embedded in ongoing practice so as to acknowledge the vitality, agency, and value of human and nonhuman life. We need some how to heed the voices of 'nature' and let them speak in our (human) political and ethical deliberations. Michel Serres has eloquently stated that, through exclusively so cial contracts, we have abandoned the bond that connected us to the world. He asks "what language do the things of the world speak that we might come to an understanding of them contractually?" And he answers that, in fact, "the Earth speaks to us in terms of forces, bonds and inter actions, [] Each of the partners in symbiosis thus owes life to the other, on pain of death." This heeding of nature will include acknowledgement of difference and otherness not only between people but between all humans and non humans, who remain also interdependently bound toge ther in life on Earth. Geographers such as William Lynn have opened up the idea of the geography of ethics, while Owain Jones has pointed to the ethical challenges of life which come in very different scales and forms of em bodiment and very different spaces, and even elements, such as the oceans.

Politics of Nature-Culture

Many political implications and questions stem from the newly broken grounds of nature culture. Advocates of a nonmodern nature culture worldview have little faith in modern politics. In terms of thought and deed, modern politics are outmoded, narrow, hopelessly out of step with how the world becomes moment by moment, day by day, and place by place. That said, ANT and sister relational approaches have been critiqued in turn for providing an uncertain ground on which to build any form of politics. The first answer to this difficulty is that, to reveal the relational, hybrid nature of everyday formations, and to begin to break up the modernist settlement of nature/culture, is a political act in itself. Key nonmodern thin kers such as Latour suggest that not only can effective science (both natural and social) only begin once the divide is breached, the same goes for politics too. In his key early work, Latour set out the notion of 'the parlia ment of things', which has been taken up as a motif for a new kind of politics in which hybridity, complexity, and process are central.

In his more recent idea of *Dingpolitik*, Latour is seeking new practices of politics, which form through and around networks, collectives, and 'issues of concern', where the more than human is given voice, where pol itical assemblies are mobile (such as portable legislature buildings) and of multiple form. The grand succession narrative of adversarial, ideologically founded, con ventional politics is replaced or broken up into many streams flowing at once. No longer are we faced with a simple, entrenched frontline between the left (and en vironment) and capitalism. 'Capitalisms', and other in stitutional arrangements, now take many forms, large and small actions with novel, local alliances and objectives are occurring worldwide. There is, as Latour puts it, 'a pix elization of politics'. New forms of politics and gover nance are hence being proposed through which the active agencies of things/nature can represent themselves, or at least be better represented. Jonathan Murdoch sets out some principles of 'ecological planning' where the pro cesses and forces of nature are built into the very fabric of planning processes. Hinchliffe *et al.* in the context of researching nature in the city, call for a cosmopolitan politics of conviviality. They draw upon the cosmopo litics of Isabelle Stengers which is a politics in which the recognition of nonhumans into the body politic generates new formations of scientific and political practices, as well as more democratic distributions of expertise.

Disciplines and Methodologies of Nature-Culture

The nature culture worldview also sets profound chal lenges for settled patterns of academia, disciplines and, to an extent, the methodologies that they employ. This does not entail a complete abandonment of the many forms of expertise that have devolved to describe aspects of the world. It is rather a redefinition of how, where, and in what ways those expertise are deployed. If particular formations in the world, say networks of food pro duction/consumption, involve a whole set of interacting processes and elements, including the biophysical, social, economic, political, cultural, and technological (as they clearly do), then the forensic skills of the natural and social sciences are still needed; but they need to be employed first in the acknowledgement that they are studying a symptom, a pulse in a larger body (rather than an entire body), and second in such a way which can communicate with other investigators. Thus we come to multidisciplinary and interdisciplinary approaches. But Whatmore warns that this needs to be about more than just placing disciplines side by side, it is also about changing the ontological and political basis of knowledge. We also need to embrace the pragmatist and non representational realization that knowledge is not simply a representation of the world from the outside, but a creative practice in the world which changes it. If all is flows, processes, entanglements, and hybridity, in dif fering networks, assemblages, and places, then we need, as Marcel Henaff puts it, "procedural methodologies, taking seriously the particularities of the sites, the un predictability of circumstances, the uneven patterns of landscapes, and the hazardous nature of becoming."

Sites of Nature-Culture

With new approaches to nature culture set out above, and new views of agencies, ethics, politics, and methods, we can begin, at last, to confront the legacies of geog raphies of nature/culture and the false divisions that it engendered. Geography has a distinct advantage in this respect, since one highly productive way of thinking nature culture is through spaces, or sites, of one kind or another. Geographers do not see the world as generally divided into ontological sections, but a world made up of many different places 'and other' forms and types of sites, patterns, and processes. There are many types we could choose to consider, but here, as particularly telling ex amples, we briefly consider bodies, cities, countrysides, and spaces of biodiversity.

Bodies (human) are primary sites of nature culture. They are at once intensely natural and cultural. We (humans) are entirely dependent upon a whole host of continual exchanges between our physical bodies and the wider environment that occur though various systems. For example, consider the millions of microbes which live in and on our bodies. The desire for a sterile, 'germ free', environment, one which is reflected/fueled by the pro motion of many products impregnated with indis criminate antibacterial agents, is a reflection of the imagined separation of human life from nature. It is a dangerous illusion. Some have argued that the rise of a number of health problems in developed societies is partly put down to the oversterilization, purification, and isolation of human bodies, thus breaking the relational, trans body functions on which healthy life depends. Additionally, the 'cultural' manifestations of bodies (ra tional thought, language, self identity, free will, voluntary movement) are emergent from and dependent upon all manner of affective systems which include memory, emotion, and various physiological and biomechanical systems which are also common through the nonhuman world. The complexity of the nature culture within us all remains one of the great challenges to knowledge of ourselves and (our place in) the world. The reunification of nature/culture perhaps has to start with the body, just as the Cartesian dualism started at the bodily level.

Cities, like New York or Mumbai, can be, as Harvey points out, read as natural as well as cultural phenomena. They are raised out of wood, sand, iron, and the like. The point is not that these apotheoses of urbanity are espe cially 'natural'. The same could be said for any city, town, or village, any house, or any artifact which, in the end (or beginning), emerges from raw material and chemical compounds extracted from the Earth and/or from living organisms. The point is that these cities and the spectacles that they offer (such as the Manhattan skyline), perhaps the most extreme, intense forms of cultural artifice seemingly a million miles 'away' from nature, can in a very real sense be seen as extrusions of natural substances and processes through complex dies which have created 'culture'. Not only are they made from 'raw materials', but structural techniques such as columns, beams, and arches began their lives as mimics of natural arrangements and still adhere to the 'the laws of nature'. They are assembled by the work of social organisms as much as is animal architecture. And of course, cities still have physical geographies, for example, topographies of hills, valleys, and watersheds. Furthermore, they teem with nonhuman life ranging from viruses, bacteria, plants, and animals, whose members overwhelmingly outnumber the human population. As already mentioned, cities can be bio diversity hotspots and more hospitable to many creatures than the countryside that might surround them. If human geographies of the city have tended to downplay the natural, then the study of physical geography has tended equally to ignore the urban. Again, as in bodies, cities are sites where nature culture is very vividly and excitingly evident, once we choose to see them in these new ways.

Countrysides. In the instance of 'countryside', notably the British countryside, the idea of nature culture is perhaps less novel, for it has long been claimed that the British countryside is as much a thing of culture as it is nature. Most of the land (and the very form of the land) has been adapted over centuries of agriculture, forestry, and other land uses. But the specificities of nature still hold great sway in how things go; for example, the changing bands of rock types which help determine soil, vegetation, and landscape form. The patterns of the country are not simply inscribed by culture, but rather they are outcomes of differing trajectories of various beings and processes coming together, in various ways at various speeds and momentums, and more or less conflictually, and perhaps heading off in new relational directions.

(Spaces of) biodiversities, spaces of difference. In her 2002 book *Hybrid Geographies*, Whatmore identifies two readings of biodiversity, one which follows the nature/culture divide, another which does not. In the former, nature is seen as a pure, separate realm. Under this view, spaces of rich, pristine, biodiversity are identified, and conservation efforts are based on separating and protecting these important areas from human society. This policy is even to the point of evicting indigenous peoples from spaces such as national parks. The alternative view which Whatmore finds in United Nations literature for the 1993 World Food Day sees things differently. This claims that humanity's place in nature is "still not widely understood. Human influences on the environment are all pervasive; even those ecosystems that appear most 'natural' have been altered directly or indirectly during the course of time. Starting some 12 000 years ago, our forebear, as farmers, fisherman, hunters and foresters, have created a rich diversity of productive ecosystems." From this remarkable, hybrid nature culture view of biodiversity, Whatmore concludes that there is no 'state of nature', only richly inhabited ecologies in which the "precious metal of bio diversity is intimately bound up with the diversities of cultural practices." These alternative definitions of biodiversity–human relationships, Whatmore argues, are important in terms of proprietorship and governance. How nature is named and seen makes a difference to who can own, control, or shape it. Here, though, we are more interested in the idea of difference in terms of how the world is always being made and unmade through distinctive and particular ecologies of nature culture.

This hybrid view of biodiversity thus presents us with a range of different, rich, biocultural ecologies. They return us to ideas raised earlier of ecological and cultural diversity being conjoined. A number of, perhaps surprising, bedfellows are seeking to focus on these kinds of links. Common Ground, the UK based arts and environment organization, has devoted much of its effort in identifying and promoting what it calls 'local diversity', where biodiversity, such as local fruit varieties, plant species, breeds of livestock, are intimately bound into local economic, cultural, and social practices. Common Ground therefore fears the loss of this richness through homogenization, be it ecological, social, cultural, and/or economic. Guattari, in *The Three Ecologies*, similarly sees ecological 'ecocide' going hand in hand with the devastation of cultural and psychological diversity at the hand of what could be called modernist capitalism.

Conclusions

The last point raised above is very important. In relational, hybrid spaces and processes, cultural (economic) processes and ecological processes can both flourish. For example, in the developed world, modern nature conservation efforts often involve separating and protecting nature, perhaps in nature reserves, from socioeconomic activities (not least agriculture). But this denies the fact that much of the 'nature' to be protected was created by agriculture and other land uses in the first place. Now, in the UK, some producers and some conservation bodies are seeking to reorientate production to ways that echo the past (e.g., less intensive systems using traditional methods) but which also 'intentionally' seek to produce nature as well as commodities: in other words, endeavoring to promulgate processes of nature culture which are positive sum games for both sides of the supposed divide.

If nature and culture are held apart, and encounters between them are seen as zero sum games (initially for nature, but eventually for all), then planetary prospects looks gloomy. With the global population rising past 6.7 billion people and more of the planet following the consumption styles of the developed world, not least in dietary terms, culture will literally consume some elements of nature (such as old growth forests). But nature cannot really be displaced; rather, it will be rendered into different, less diverse forms. These forms, in turn, might consume culture. Rather than thinking nature/culture, then, we need to think and act upon nature culture formations which may come in a myriad array of combinations and spaces. We need to heed their composition and their processes. We need, as in new ecology, to assess if these assemblages are bringers of growth, well being, richness and diversity, or otherwise. We need sites, networks, spaces, and places where different nature cultures can flourish – bodies, parks, streets, ponds, houses, rivers, gardens, oceans, windowboxes, forests, towns, cities, farmland, countrysides, wilderness, markets ….

See also: Actor-Network Theory/Network Geographies; Animal Geographies; Nature, Historical Geographies of; Nature, History of; Nature, Performing; Nature, Social; Posthumanism/Posthumanistic Geographies; Time and Historical Geography; Urban Habitats/Nature; Wilderness.

Further Reading

Adams, W. (1996). *Future Nature*. London: Earthscan.
Bennett, J. (2001). *The Enchantment of Modern Life: Attachments, Crossings and Ethics*. Princeton, NJ: Princeton University Press.
Bingham, N. and Thrift, N. (2000). Some new instructions for travellers: The geography of Bruno Latour and Michel Serres. In Crang, M. & Thrift, N. (eds.) *Thinking Space*, pp 281 301. London: Routledge.

Braun, B. (2003). Nature and culture: On the career of a false problem. In Duncan, J. S., Johnson, N. C. & Schein, R. H. (eds.) *A Companion to Cultural Geography*, pp 149 179. Oxford: Blackwell.
Braun, B. and Castree, N. (eds.) (1998). *Remaking Reality: Nature at the Millenium*. London: Routledge.
Castree, N. (2005). *Nature*. London: Routledge.
Castree, N. and Braun, B. (eds.) (2001). *Social Nature: Theory, Practice and Politics*. Oxford: Blackwell.
Castree, N. and Braun, B. (2006). Constructing rural natures. In Cloke, P., Marsden, T. & Mooney, P. H. (eds.) *Handbook Rural Studies*, pp 161 170. London: Sage.
Cheney, J. (1999). The journey home. In Weston, A. (ed.) *An Invitation to Environmental Philosophy*, pp 141 168. New York: Oxford University Press.
Cloke, P. (2003). Knowing ruralities. In Cloke, P. (ed.) *Country Visions*, pp 195 217. Harlow: Pearson.
Cloke, P. and Jones, O. (2001). Dwelling, place, and landscape: An orchard in Somerset. *Environment and Planning A* 33, 649 666.
Cronon, W. (ed.) (1996). *Uncommon Ground: Rethinking the Human Place in Nature*. New York: W. W. Norton.
Deleuze, G. and Guattari, F. (1987). *A Thousand Plateaus. Capitalism and Schizophrenia*. London: The Athlone Press.
Demeritt, D. (2001). Being constructive about nature. In Castree, N. & Braun, B. (eds.) *Social Nature: Theory, Practice and Politics*, pp 22 40. Oxford: Blackwell.
Fitzsimmons, M. (2004). Engaging ecologies. In Cloke, P., Goodwin, M. & Crang, P. (eds.) *Envisioning Human Geography*, pp 30 47. London: Arnold.
Fitzsimmons, M. and Goodman, D. (1998). Incorporating nature: Environmental narratives and the reproduction of food. In Braun, B. & Castree, N. (eds.) *Remaking Reality: Nature at the Millennium*, pp 194 220. London: Routledge.
Franklin, A. (2002). *Social Nature*. London: Sage.
Guattari, F. (2000). *The Three Ecologies*. London: The Athlone Press.
Haraway, D. (1991). *Simians, Cyborgs, and Women: The Reinvention of Nature*. New York: Chapman & Hall.
Haraway, D. (1992). A manifesto for cyborgs: Science, technology, and social feminism in the 1980s. In Nicholson, L. J. (ed.) *Feminism/ Postmodernism*, pp 190 233. London: Routledge.
Harrison, S., Pile, S. and Thrift, N. (2004). *Patterned Ground: Entanglements of Nature and Culture*. London: Reaktion Books.
Harvey, D. (1996). *Justice, Nature, and the Geography of Difference*. Oxford: Blackwell.
Henaff, M. (1997). Of stones, angels and humans: Michel Serres and the global city. *SubStance* 83, 59 80.
Hinchliffe, S., Kearnes, M. B., Degen, M. and Whatmore, S. (2005). Urban wild things a cosmopolitical experiment. *Environment and Planning D: Society and Space* 23(5), 643 658.
Ingold, T. (1993). The temporality of landscape. *World Archaeology* 25, 152 174.
Ingold, T. (1995). Building, dwelling, living: How people and animals make themselves at home in the world. In Strathern, M. (ed.) *Shifting Contexts: Transformations in Anthropological Knowledge*, pp 57 80. London: Routledge.
Ingold, T. (2000). *The Perception of the Environment. Essays in Livelihood, Dwelling and Skill*. London: Routledge.
Kolbert, E. (2007). *Field Notes from a Catastrophe: A Frontline Report on Climate Change*. London: Bloomsbury.
Latour, B. (1993). *We Have Never Been Modern*. Hemel Hempstead: Harvester/Wheatsheaf.
Latour, B. (2004). Non humans. In Harrison, S., Pile, S. & Thrift, N. (eds.) *Patterned Ground: Entanglements of nature and culture*, pp 224 227. London: Reaktion Books.
Latour, B. (2007). From Realpolitik to Dingpolitik or How to Make Things Public. http://www.bruno latour.fr/articles/article/96 DINGPOLITIK2.html (accessed August 2008).
Lovelock, J. (2000). *Gaia: A New Look at Life on Earth*. Oxford: Oxford University Press.
Lynas, M. (2008). *Six Degrees: Our Future on a Hotter Planet*. London: HarperPerennial.
Massey, D. (2005). *For Space*. London: Sage.
Matless, D. (2000). Versions of animal human: Broadland, *c.* 1945 1970. In Philo, C. & Wilbert, C. (eds.) *Animal Geographies: New Geographies of Human Animal Relations*, pp 115 140. London: Routledge.
Matless, D., Merchant, P. and Watkins, C. (2005). Animal Landscapes: Otters and wildfowl in England 1945 1970. *Transactions of the Institute of British Geographers NS* 30, 191 205.
McKibben, B. (1990). *The End of Nature*. London: Penguin Books.
Midgley, M. (1996). *Utopias, Dolphins and Computers: Problems of Philosophical Plumbing*. London: Routledge.
Murdoch, J. (2003). Co constructing the countryside: Hybrid networks and the extensive self. In Cloke, P. (ed.) *Country Visions*, pp 263 282. Harlow: Pearson.
Naess, A. (1997). Sustainable development and the deep ecology movement. In Baker, S., Kousis, M., Richardson, D. & Young, S. (eds.) *The Politics of Sustainable Development: Theory, Policy and Practice within the European Union*, pp 61 71. London: Routledge.
Pepper, D. (1996). *Modern Environmentalism: An Introduction*. London: Routledge.
Philo, C. and Wilbert, C. (2000). Animal spaces, beastly places: An introduction. In Philo, C. & Wilbert, C. (eds.) *Animal Spaces, Beastly Spaces: New Geographies of Human Animal Relations*, pp 1 34. London: Routledge.
Plumwood, V. (1993). *Feminism and the Mastery of Nature*. London: Routledge.
Plumwood, V. (2002). *Environmental Culture: The Ecological Crisis of Reason*. London: Routledge.
Royston, H. (1999). Ethics on the home planet. In Weston, A. (ed.) *An Invitation to Environmental Philosophy*, pp 107 140. New York: Oxford University Press.
Serres, M. (1995). *The Natural Contract*. Ann Arbor, MI: Michigan University Press.
Shiva, V. (1988). *Staying Alive: Women, Ecology and Development*. London: Zed Books.
Smith, N. (1998). Nature at the millenium: Production and re enchantment. In Braun, B. & Castree, N. (eds.) *Remaking Reality: Nature at the Millenium*, pp 269 280. London: Routledge.
Thrift, N. (1999). Steps to an ecology of place. In Massey, D., Sarre, P. & Allen, J. (eds.) *Human Geography Today*, pp 295 352. Oxford: Polity.
Thrift, N. (2008). *Non Representational Theory: Space; Politics; Affect*. London: Routledge.
Whatmore, S. (1999). Rethinking the 'human' in human geography. In Massey, D., Sarre, P. & Allen, J. (eds.) *Human Geography Today*, pp 22 41. Oxford: Polity.
Whatmore, S. (2002). *Hybrid Geographies: Natures, Cultures, Spaces*. London: Sage.
Whatmore, S. and Hinchcliffe, S. (2003). Living cities: Making space for urban nature, Soundings. *Journal of Politics and Culture* http://www.open.ac.uk/socialsciences/habitable cities/ published papers and papers in progress (accessed on August 2008).
Wolch, J. and Emel, J. (eds.) (1998). *Animal Geographies: Place, Politics and Identity in the Nature Culture Borderlands*. London: Verso.
Zimmerer, K. (1994). Human geography and the 'new ecology'. *Annals of the Association of American Geographers* 1, 108 125.

Relevant Websites

http://parliamentofthings.info
A Parliament of Things.
http://www.bruno latour.fr
Bruno Latour.
http://web.archive.org
Internet Archive: Wayback Machine.
http://www.lancs.ac.uk
Lancestor University.
www.open.ac.uk
The Open University: Distance Learning Courses and Adult Education.

Natures, Charismatic

J. Lorimer, Kings College, London, UK

Glossary

Jizz A term used by naturalists that refers to the unique combination of properties of an organism that allows its ready identification and differentiation from others. The etymology of jizz is contested, but it is generally believed to be a corruption of the plane-spotters acronym GISS – general impression of size and shape.

Jouissance The pleasure experienced in the presence of meaning – the satisfaction of completing an intellectual project.

Nonhuman Charisma The distinctive properties of a nonhuman that distinguish it relative to human perception and affect individual and social attachments. It is relational and comprises ecological, practical, and esthetic components.

Introduction and Definitions

Popular bestiaries of the nonhuman world, both ancient and modern, are inhabited by a relatively small number of organisms – lions, elephants, whales, and snakes, to name a few of the most common. These animals receive disproportionate attention across animal spaces, ranging from the gladiatorial arenas of the Roman Empire to the databases of the International Union for the Conservation of Nature and Natural Resources (IUCN) Species Survival Commission. Similarly, different cultures invest enormous symbolic meaning in "awesome and awe full" landscapes – across space and time, a few nonhuman archetypes such as forests, swamps, mountains, and waterfalls have been variously eulogized and vilified.

In differing ways, these places and organisms can be understood as charismatic – they possess and perform nonhuman charisma. In its most general definition, nonhuman charisma refers to those properties of a place or organism that distinguish it relative to human perception. Although these properties result in myriad human attachments and affections – ranging from love to disgust and wonder to indifference – the charisma of a place or organism refers exclusively to their ability to get noticed. Nonhuman charisma therefore primarily relates to esthetics, but this encompasses the whole body and its full range of senses. Furthermore, nonhuman charisma is both too lively and material to be a social construction and too variable to be naturally determined – it relates to both the agency of the nonhuman being witnessed and the social structure in which the witness is enmeshed.

The adjective 'charismatic' is currently most commonly applied to nonhumans by nature conservationists, who strategically employ charismatic species as flagships to raise funds for and awareness of broader conservation initiatives. However, they have not sought to explore the parameters of this charisma. Cultural and historical geographers of nature and animal geographers have become interested in nonhuman charisma as part of the broader materialist turn with their subdisciplines. Geographers of nature have explored nonhuman charisma as a manifestation of nonhuman, or 'more than human' agency, and have traced a few of its various historical and contemporary manifestations. Geographers and sociologists of science have also traced the affective underpinnings and performative character of encounters between humans and nonhumans in the field. They have demonstrated that attention to the charisma of different organisms and places provides useful insights into popular and scientific ontologies, epistemologies, and ethics.

Theorizing Charisma

There is a long history of interest in human charisma in sociology and social theory, which stems largely from the influential work of Max Weber. Drawing on articulations in the Apostle Paul's letter to the Corinthians, Weber understood charisma as a 'gift of grace': a valuable, yet threatened property in the increasingly 'disenchanted' and bureaucratic modern world. Weber and his students carried out a series of studies that traced the social effects of human charisma, especially within large organizations. These studies did not seek to generalize about the characteristics of human charisma, nor did they attend to charisma in the nonhuman realm.

Elsewhere, the charisma of organisms and places has been a long standing concern of a small group of evolutionary biologists, led by the preeminent conservationist E. O. Wilson. In a number of controversial books, Wilson and his colleague Stephen Kellert have argued that there exist 'biophilic' and 'biophobic' tendencies among humans. These are commonly expressed across all cultures toward different nonhumans, and result from and are perpetuated by the genetic advantages they confer on those expressing them. Charisma is not god given, but it is innate. So for example, they argue that humans favor parkland landscapes because they would have afforded good visibility and hunting opportunities in the past. In contrast, humans experience biophobia toward snakes and spiders, because they are poisonous.

Wilson and Kellert's theories have been contested by social anthropologists who point to the important shaping effects of culture and representations on individual perceptions of nonhuman charisma, and the difficulties in finding genes for snake avoidance, for example. Instead, such social constructivist accounts replace the formers' environmental determinism, with webs of symbolism, which together construct perception. Charisma is here the result of religion, folklore, and other modes of social signification. However important as they are, such constructivist accounts go full circle and efface the fleshy and earthy materiality of the places and organisms being encountered; charisma is here reduced to narration and iconography.

More recently, cultural geographers have sought to find a path between these realist and constructivist approaches that reconciles the material and the immaterial facets of social behavior and environmental perception. Here they have drawn on a variety of approaches and philosophies, including ethology, phenomenology, post structuralism, and vitalism. These new, 'more than human', 'posthuman', or 'nonrepresentational geographies' emphasize the bodily skills, affects, and technologies that surround and mediate encounters between humans and nonhumans. This approach has shed interesting light on the character of charismatic natures. Nonhuman charisma is here understood to encompass both the ecological affordances of a place or organism and the affective forces that flow between humans and nonhumans in embodied encounters. In contrast to an understanding of charisma as innate or god given, here charisma is relational, and emerges within particular social and material contexts.

Typology of Charismatic Natures

Taking such a more than human approach, it is possible to disaggregate the characteristics of nonhuman charisma into a three part typology, which comprises its ecological, esthetic, and corporeal dimensions. This division is designed as a heuristic schematic in which the three dimensions interweave and overlap.

Ecological Charisma

Ecological charisma refers to the ease with which a place or organism is perceived by the witnessing human subject. To trace the ecological charisma of a nonhuman requires an ethological perspective. Ethology starts from an understanding of a being – human or otherwise – as an ecological entity, immersed in its environment, and performing a number of core behavioral characteristics. Ethologists have termed these the affordances of an organism, and taken together they establish its Umwelt. The intersections of the Umwelten of different organisms determine their possible interactions within an ecological complex.

An ethological perspective on human–environment interactions foregrounds the important common properties of human bodies. These are the specific human competencies that, in Katharine Hayles' terms, frame the 'cusp' through which we make sense of the world; they can be understood as the human body's primary affordances. All humans are warm blooded mammals. Most humans are bipedal, between 1.4 and 1.9 m tall, land dwelling, diurnal, and ocular centric (but in possession of five senses). Human sensory organs only make use of small portions of the electromagnetic, acoustic, and olfactory spectra. Although many, though not all, of these constraints have been stretched or overcome by the extension of the human body through technologies, these developments have only occurred relatively recently and are still expensive and unwieldy to use in the field.

Understood in this ethological fashion, we can see how the physiological and phenomenological configuration of the average human body puts in place a range of filtering mechanisms that disproportionately endow certain organisms and places with ecological charisma. While these can be overcome and are by no means deterministic, the ecological affordances of different organisms currently intersect more or less easily with those of humans. These intersections determine the detectability of an organism and the accessibility of a place and the ease with which an interested human is able to tune in to its behavior or ecology.

The important affordances which determine an organism's detectability include a range of parameters that influence its visibility, including size, color, shape, speed, and degree of movement. They also include aural characteristics such as the presence or absence of a noise, call, or song, and the frequency and magnitude of this sound. Taken together, these affordances constitute what natural historians call an organism's 'jizz'. Jizz refers to the unique combination of properties to an organism that allows its ready identification and differentiation from others.

Furthermore, the nature and frequency of any human–nonhuman encounter relates to the intersections between the space–time rhythms of the two organisms. These intersections are controlled by factors such as an organism's seasonality, migration patterns, day–night ecology, and distribution on land, in the air, or under the sea. Differences in the relative detectability and accessibility of nonhumans would account for the disproportionate human knowledge of birds and mammals, over nematodes and bacteria, for example. This pattern is illustrated in **Table 1**, which shows the degree to which

Table 1 Taxonomic partialities in conservation research

Group	*Prevalence in nature (%)*	*Prevalence in conservation literature (%)*	*Ratio of prevalence (r)*
Plants	18	20	1.1
Vertebrates	3	69	23
Birds	19	39	2.0
Mammals	9	40	4.4
Fish	48	8	0.2
Reptiles	15	8	0.5
Amphibians	9	6	0.7
Invertebrates	79	11	0.1
Mollusks	6	19	3.2
Crustaceans	3	8	2.7
Insects	80	68	0.8
Butterflies and moths	15	48	3.2
Beetles	39	26	0.7
Other insects	46	26	0.6
Other invertebrates	11	6	0.5

Source: Clark, J. and May, R. (2002). Taxonomic bias in conservation research. *Science* 297 (5579), 191 192.

different taxa are proportionately represented in the conservation literature.

The ecological charisma of a place or organism is also closely related to its practical and economic utility to humans. Those organisms and places that help sustain or threaten agricultural livelihoods often have long cultural histories. However, it is important to separate such utilitarian properties from those encompassed by the concept of charisma. Charisma relates more to the esthetic properties of a nonhuman, rather than its eco nomic value, though the two are closely interwoven.

Esthetic Charisma

The next two types of nonhuman charisma – esthetic and corporeal charisma – relate to the affective properties of places and organisms. They build on the intersections between the different affordances of nonhumans to trace the affective forces that flow between humans and non humans. Here encounters between humans and nonhu mans are understood as acts of becoming, where humans and nonhumans evolve and are involved through mul tiple processes of territorialization and deterritorializa tion over differing timescales. Such encounters are understood as both skilled and emotional, involving a variety of forms of knowledge outwith the confines of modern rationality.

Esthetic charisma refers to the distinguishing prop erties of a place or organism's behavior and appearance that trigger particular emotions in those humans who encounter it, either in the flesh or as a textual repre sentation. It catalyzes individuals' ethical sensibilities and is layered over and interacts with ecological charisma in different ways. It is difficult to generalize about the af fective properties of different places and organisms as there is an enormous variety in the responses of different individuals and cultural groups. However, there are two axes along which we can explore different manifestations, which relate to the forms and the temporalities of the nonhumans encountered.

In relation to the forms of nonhuman, especially when it comes to individual organisms, there is a powerful axis of anthropomorphism at work that orientates people's ethical sensibilities. For many people, the organisms they feel the greatest affection toward are those that most closely approximate an anatomical or social behavioral norm. Konrad Lorenz has argued for the existence of instinctive preferences to organisms that look like human babies – the cuddly species much favored by certain forms of animal rights advocacy. Other philosophers have drawn attention to different anatomical components as the touchstone of ethical concern. Levinas foregrounds the face as the site of reciprocity, while Heidegger emphasizes the hand. Monstrous and autonomous or ganisms that depart from such anatomical norms are often widely vilified – notice the popular mobilization of invertebrates in dystopic science fiction and horror films (**Figure 1**).

Furthermore, as James Hillman points out, organisms whose mode of social organization renders the individual superfluous to the multitude – like ants – are often widely distrusted. Here there is a strong, sometimes phobic fear of the unknown, which echoes the import ance Lorenz places on anthropomorphism. However, in the case of insects, it is their radical alterity to humans in terms of size, ecology, physiology, esthetics, and modes of social organization that engenders popular feelings of antipathy and distrust. In many ways insects have the characteristics of what Julia Kristeva has termed the 'abject', where the abject is understood as the breakdown of meaning that results from being confronted and

Figure 1 Insect horror poster.

overwhelmed by the other. This abjection is far from universal and there are some who are much more strongly drawn to nonhuman alterity and rejoice in the difference of the nonhuman realm. For them this difference suggests fascinating modes of becoming otherwise.

It is also possible to identify another important axis along which esthetic charisma is arrayed. This relates to the temporalities of the nonhuman realm – the rates at which a place, organism, or ecological complex changes. Despite the best efforts of Darwin and his successors in disequilibrium ecology, for many people in the West, nature is still understood as a timeless and pure place beyond and radically separated from human culture; it is in balance or equilibrium. Stable nonhuman rhythms and patterns are naturalized and thus moralized to provide the foundations for codes of social behavior. Certain people and practices are in or out of sync, depending on their harmony with perceived natural rhythms. Those adhering to this balance of nature temporality favor stable pastoral landscapes, or purified wildernesses, where nature runs its own course at incremental timescales.

In contrast, others look at the nonhuman realm and see only chaos, disorder, and hybridity. For them, the nonhumans to admire are those characterized by feral transgression or apocalyptic potential. They rejoice in the unpredictability and autonomy of nonhumans and resent efforts to efface nonhuman difference. For some, this sensibility results in what Stephen Baker has termed, 'anthropomorphobia', a dislike of anthropomorphism and a Nietzschean antipathy toward the domestic sphere. This awe in the face of the sublime and the celebration of nonhuman tricksters has a long cultural history, going back at least to the romantic period.

Without going as far as to prescribe universal modes of biophilia and biophobia, it is nonetheless possible to identify two important axes, which together help chart the parameters of esthetic charisma. Different individuals and cultural groups place different places and organisms at differing points within a matrix formed by these axes – however, it is clear that those nonhumans at the furthest edges will often be the most charismatic – the baby faced panda as opposed to the giant squid, or Constable's hay meadows against urban wastelands, to give two of the numerous possible contrasts.

Corporeal Charisma

Corporeal or practical charisma relates to the affections arising from practical, embodied, and sensory interactions between humans and nonhumans in different places. In contrast to esthetic charisma, which generally has its effects in an instant, corporeal charisma involves visceral becomings that take place over a longer period of time. There is a huge diversity of possible types of corporeal charisma, which relate to the enormous diversity of the modes that exist for encountering the nonhuman realm. These range from production to consumption, from detailed study to adrenaline fueled adventure, and are greatly affected by the technological infrastructures that surround any encounter.

There are two wide ranging examples of corporeal charisma that relate to the pursuit of knowledge about the nonhuman realm – these can be termed epiphanies and jouissance. Professional and amateur naturalists are often inspired in their work by early encounters with particular species or places. They have an epiphany with a nonhuman in which they are lifted up or deterritorialized. This event plants a seed and generates the enthusiasm for later encounters and research. Jouissance, on the other hand, is more protracted, and describes the pleasure experienced in the presence of meaning. It is the sense of achievement gained from unraveling a mystery, or generating a revelation, of completing a list or of finding a lost species. Many naturalists and ethologists working on higher order mammals experience jouissance through the recognition

Figure 2 Human elephant encounter.

and reciprocity they enjoy in their interactions with their research subjects (**Figure 2**).

Both jouissance and epiphanies are offered up un equally by different nonhumans as a result of differences in ecological charisma – for example, it is more common to have an epiphany with an eagle, than with a nematode. These are just two examples, and there are many more forms of corporeal charisma that have yet to be described in detail.

Nonhuman Charisma and Ontology

This understanding of nonhuman charisma – as a mode of nonhuman agency – implies a bounded relational ontology. Unlike the determinism of biophilia, the char acter of charisma is not fixed; however, neither is it di vorced from the material and ecological properties of the nonhuman and reduced to a social construction. The recognizable jizz and accessibility of some organisms and landscapes – like birds or mountain peaks – implies the existence of an intuitive ontology of forms and tempor alities that is closely tied to ecological charisma. This intuitive ontology affects the categories used to carve up the nonhuman realm – into units like species and habi tats. Many of the current working definitions of species are closely tailored to higher order organisms and fall down when applied to lesser groups; similarly, schemes for classifying vegetation essentialize stable climax communities over more dynamic ecological complexes. An intuitive ontology also strongly effects the scope of existing knowledge about nonhuman organisms – the accessibility of birds, over nematodes, for example, means that their ecology and behavior are much more thor oughly understood by natural scientists.

However, the ecological charisma of an organism is not solely deterministic of human familiarity and modes of categorization. With the development of increasingly sophisticated technological infrastructures for tuning in to the nonhuman world, previously inaccessible places and organisms are receiving more attention. Innovations in film making and the growing popularity of televised natures have also subtly changed the landscapes of non human popularity and familiarity. In some cases, this has reinforced previous dispositions – toward the cuddly and ferocious and against the creepy, for example. However, television has also given unprecedented attention to previously inaccessible spaces – like the deep sea or subterranean regions.

Charisma, Epistemology, and Conservation

Certain charismatic organisms are strategically employed in nature conservation as 'flagship species' to raise fund for and awareness of wider ecological problems. Perhaps the most famous of these is the Giant Panda, whose image has been adopted as the iconic brand of the World Wildlife Fund (WWF) (**Figure 3**). Under the umbrella of the panda brand, the WWF raises money and campaigns not only for the species, but also for all global bio diversity and, by extension, policies to ameliorate climate change, reduce deforestation, and prevent water pol lution. Resources raised from the charisma of the panda trickledown all over the world. This strategy is also used

Figure 3 The WWF 'panda brand'. Source: www.wwf.org, permission granted July 2006.

on a smaller scale, where charismatic species are employed to help protect the less charismatic organisms with which they share their habitat.

As flagship species, such charismatic nonhumans are employed as boundary objects that mediate between several different natural epistemologies. In the context of nature conservation, the esthetic charisma of the Panda harnesses the more the rational vernacular affections of the general public and helps to make them commensurable with the objective epistemology of conservation science. Such a strategy helps to overcome the long standing antipathy between science and emotion in nature conservation and allows groups with very different interests in the nonhumans in questions to work together toward a common end.

Nonhuman charisma also operates outwith the exotic realms of conservation in the more mundane spaces of pet keeping. Here animals serve as significant others in the emotional lives of their owners. For many people, such domesticated beings bridge the worlds of humans and animals, and establish epistemological and ethical linkages through the wonderful and affectionate encounters they afford. Geographers are beginning to explore these interactions, but more work needs to be done on the character of and relationships between everyday and exotic nonhuman charisma.

Charisma and Ethics

Esthetic and corporeal charisma are differentially offered up by various nonhumans and together sketch out an alternative taxonomy to the nonhuman realm. This taxonomy does not understand organisms as objects, ordered arboreally into cladistic hierarchies. Instead, it is open to the creative force of nonhumans in their encounters with humans. Such nonhuman charisma plays a vital role in compeling and orientating ethical sensibilities toward the nonhuman realm. Nonhuman charisma acts as the catalyst that compels people to care for, ignore, or abhor different components of the nonhuman realm.

The two axes of esthetic charisma outlined above suggest that such ethical sensibilities are strongly orientated toward and against various modes of nonhuman similarity and difference. For some animal rights theorists, like Peter Singer, it is the similarity of certain nonhumans to a human ideal that earns them ethical status. For other environmentalists, who emulate Henry Thoreau, it is the radical and incomprehensible alterity and chaotic dynamism of certain nonhumans that earns their wonder and respect. Either way, the character of nonhuman charisma strongly affects the scope and intensity of ethical concern.

Conclusions

Recent interest in charismatic natures should be understood as part of wider return to the material and materialist analysis in cultural geography, after the excesses of the 'cultural turn'. This work is highlighting the oft neglected agencies of nonhumans – including bodies, organisms, landscapes, and technologies – in the emergence of everyday life. Theories of nonhuman charisma provide one such materialist mode of engaging with the nonhuman realm, and this approach offers a number of novel insights into popular and scientific ontologies, epistemologies, and ethics. It makes clear the strong influence of charisma on what people know about the nonhuman world and how they act within it. There is still a great deal of research to be done on the nature of nonhuman charisma and its social and ecological effects across different geographical contexts, and this area offers several interesting avenues for future enquiry.

See also: Actor-Network Theory/Network Geographies; Affect; Animal Geographies; Becoming; Body, The; Conservation and Ecology; Dwelling; Emotional Geographies; Human-Nonhuman; Material, The;

Nature; Nature, Performing; Non-Representational Theory/Non-Representational Geographies; Posthumanism/Posthumanistic Geographies.

Further Reading

Baker, S. (2000). *The Postmodern Animal*. London: Reaktion Books.

Bennett, J. (2001). *The Enchantment of Modern Life: Attachments, Crossings, and Ethics*. Princeton, NJ: Princeton University Press.

Clark, J. and May, R. (2002). Taxonomic bias in conservation research. *Science* 297(5579), 191 192.

Hayles, N. K. (1995). Searching for common ground. In Soule, M. & Lease, G. (eds.) *Reinventing Nature? Responses to Postmodern Deconstruction*, pp 45 60. Washington, DC: Island Press.

Hillman, J. (1988). Going bugs. *Spring: A Journal of Archetype and Culture* 1988, 40 72.

Jones, O. (2000). (Un)ethical geographies of human non human relations: Encounters, collectives and spaces. In Philo, C. & Wilbert, C. (eds.) *Animal Spaces, Beastly Places*, pp 268 291. London: Routledge.

Kellert, S. and Wilson, E. O. (1993). *The Biophilia Hypothesis*. Washington, DC: Island Press.

Kristeva, J. (1982). *Powers of Horror: An Essay on Abjection*. New York: Columbia University Press.

Lorimer, J. (2007). Nonhuman charisma. *Environment and Planning D: Society and Space* 25(4), 911 932.

Macnaghten, P. and Urry, J. (1998). *Contested Natures*. London: Sage.

Thrift, N. (2005). From born to made: Technology, biology, space. *Transactions of the Institute of British Geographers* 30, 463 476.

Whatmore, S. (2002). *Hybrid Geographies: Natures, Cultures, Spaces*. London: Sage.

Natures, Gendered

C. Radel, Utah State University, Logan, UT, USA

Glossary

Discourse Social practices and communications which produce (and reproduce) knowledge about the world, making the world intelligible to oneself and to others.

Ecofeminism A range of theoretical and activist positions which connect the oppression of women with the destruction of nature.

Environmental Relations Another term for human–environment relations – interactions between people and their biophysical environments.

Essentialism The view that gender differences are the result of essential biologically based differences between women and men.

Gender A description of socially constructed differences between women and men. Gender was formerly used in contrast to sex (as biologically based differences), but recent feminist theory no longer draws this distinction, seeing sex as likewise not occurring outside of a social framework of understanding.

Patriarchy A social system in which men (and male characteristics) are deemed superior to women (and female characteristics) and exercise authority over them.

Social Construction The view that gender is not biologically determined; rather, gender is defined and achieved through social processes that articulate and enforce appropriate gender roles and behaviors.

Introduction

Geographers have long expressed an interest in understanding the relationship of people to their biotic and natural physical environments. In human geography, as a discipline, the focus on the human–environment relationship has a long history, with a scope of analysis ranging from societies to households to individuals. For much of geography's history, women were subsumed into a household unit or were an invisible appendage to the male environmental agent. The development of feminist theory beyond the borders of geography initially led to a broadening of focus within geography's examination of the human–environment relationship to include women's experiences. Ecofeminist theory played an influential role in geography's early treatment of women's relationships to nature and the environment, with much of the work celebrating women's connections to nature and to natural landscapes as different from men's.

The subsequent explosion of political ecology as a subfield of geographic inquiry paved the way for the development of a cultural–political ecology approach that opened up the household unit and its relationship to the environment to a consideration of the place of gender and gender relations. Political ecology explicitly incorporates analyses of power relations and the social distribution of natural resources into the study and understanding of land use decision making and behavior, and of human–environment relations more broadly. This opening of the household unit allowed inquiry into the both cooperative and conflicting relationships between men and women in the context of household and community environmental relations. Much of the early work in political ecology that incorporated gender relations and differences was linked to research on agricultural and environmental change and often explored women's access to and control over resources, particularly land and trees.

The Influence of Ecofeminism

Before the development of gender focused research coming out of cultural–political ecology, early work in geography that focused on women/gender and the environment had been strongly influenced by US and international ecofeminist scholarship. Women's relationship to the environment and to natural resources was held up as different from that of men's and was celebrated. Ecofeminist thought in the US developed during the 1970s along many different lines and in a variety of forms. It is possible, however, to identify two main strands of ecofeminist thought, both of which argue for women's gender based interest in environmental conservation or sustainability: cultural ecofeminism (sometimes labeled spiritual ecofeminism) and social ecofeminism.

The cultural ecofeminist strand, which dominated initially, based women's interest in environmental conservation or well being on women's inherent caring or nurturing nature and on a shared subjugation under patriarchal systems. This position is frequently labeled as essentialist, as it bases its understanding of the relationship between women and environment in large part on an essential (or presocial) characterization of women as closer to nature through, for example, the biological realities of human reproduction. The dualisms of nature/culture and woman/man are frequently viewed as parallel in Western society, with culture and man in positions of dominance following the scientific revolution. Under cultural ecofeminism, these dualisms are not rejected but instead are taken as a source of power, aiming to reverse the positions of dominance. The association of women and nature as equivalent sides of dualistic pairs is one sense in which we can consider nature as gendered, and one common image that embodies this association is Earth as Mother.

The predominantly essentialist strand of cultural ecofeminism was partially replaced during the 1990s by a social ecofeminist strand that was primarily constructivist in nature. This ecofeminist position firmly centered women's interests in conservation and the environment in their socially constructed gender roles of childcare and the gendered division of labor. Women were not considered inherently closer to nature; rather, the sociocultural assignment of specific reproductive and productive tasks to women placed them closer to nature through their labor. In the developed world, the focus has been primarily on child rearing or other caring roles, and these roles have been proffered as an explanation for women's grassroots activism in environmental health issues. In the Third World, women are generalized as having primary responsibility for subsistence production and natural resource tasks related directly to meeting immediate human needs, such as the provision of drinking water and the collection of firewood for meal preparations. These gender based labor responsibilities are thought to position women as the first to become aware of environmental degradation and to place their gender based interests in line with those of environmental conservation. A gender roles approach, as articulated in the 1990s, claims that women's roles and responsibilities make them resource users and managers, giving them both interests in natural resources and special knowledge of those natural resources.

Keeping this general outline of ecofeminism in mind, it is possible to identify at least three primary influences the different variants of ecofeminist thought have had on geography's treatment of the connection between gender and the human–environment relationship. These influences are summarized below:

1. *An examination of nature as feminine.* The dualistic association of man with culture and woman with nature was initially embraced as a source of power and women's relationship to the environment identified as superior to men's. This initial embracing of nature as female later developed into more critical assessments of the ideas of nature or of landscapes as female, with examinations of the sociocultural mechanisms by which nature becomes feminized and explorations of what impact this feminization has on human–environment relations more broadly.
2. *A focus on shared oppressions.* Ideas of positionality or the situatedness of knowledge coming out of feminist critiques of science contributed to a treatment of the relationship between women and nature as different from the relationship between men and nature. Since women are oppressed through gender structures, much as nature is oppressed by male dominated society, women have a different perspective on and knowledge of the environment. This perspective and knowledge was (and continues to be) viewed by some as superior to men's; however, more recently, this perspective and knowledge is viewed simply as different and as an outcome of all the various aspects of a person's identity, including, but not restricted to, gender.
3. *An assumption of women's privileged access to nature through gender roles and responsibilities.* Women's material connections to local environments through their labor can be considered the central core to understanding women's relationships to environments. The approach taken can vary from an essentialist position which highlights women's child rearing and subsistence production to a constructivist position which examines gendered divisions of labor specific to a time and place. A more essentialist position views women as natural environmental caretakers or stewards, whose own interests coincide with those of nature. The constructivist position, on the other hand, does not assume women's privileged access to nature.

Ecofeminism, especially as formulated in the 1970s and 1980s, received much criticism from feminist scholars for its essentialist leanings placing women as closer to nature. Scholars and activists critical of ecofeminism found the continued association of women with nature (as opposed to culture) as problematically reinforcing existing patriarchal structures that oppress women. Third World feminist scholars criticized the essentialist leanings of ecofeminism for creating a single unified image of woman, ignoring the differences among women and the intersections of gender with other social structures such as class, age, and ethnicity.

Feminist geographers and geographers interested in the political ecology of resource access and control likewise have found ecofeminism to be a problematic basis for understanding the gendered aspects of nature and environments. Vandana Shiva has been the foremost shaper of ecofeminist thought with respect to Third World women and environments, and for political ecologists working in the rural Third World her theoretical position has been particularly problematic. For example, Shiva's 1988 publication, *Staying Alive: Women, Ecology and Survival in India*, explored the negative effects of the Green Revolution in India on both rural women and the environment and identified a feminine principle which she argued is necessary for maintaining humanity's balance with nature. This feminine principle represents rural Third World women's distinct spirituality and relationship to the environment that is crucial for the realization of a just and sustainable development. Many geographers view Shiva's argument as a romanticization of both the feminine and the indigenous as antidotes to a destructive modernist and capitalist development.

More recently, there have been several scholarly attempts to reclaim the promise of ecofeminism for addressing gender based oppression and environmental destruction. In 1999, two ecofeminist theorists (Catriona Sandilands and Noel Sturgeon) issued important calls for salvaging ecofeminism, specifically as political action, distinguishable from a theoretical and analytical tool. One body of work (Sandilands) attempts to move ecofeminism forward as a project in radical democracy, with the potential to destabilize hegemonic categories, identities, and discourses. In this effort, it draws upon the constructivist strand of ecofeminism, recognizing ecofeminism's potential to destabilize, but also identifying significant risks in the strategic application of essentialism for political activist purposes. As argued by many feminist scholars, essentialism reifies existing identities and discourses, in effect legitimizing and bolstering them.

A second body of work (Sturgeon) takes a critical look back at the role played by ecofeminism in international debates during the 1990s around women and environments. It highlights the positive contributions of ecofeminism as an international political discourse, arguing that the intervention of scholars like Shiva into the discourse on women, environment, and development came at a crucial moment in time. Prior to Shiva's intervention, international discourse had positioned poor women primarily as destroyers of the environment, through their role in population growth and in the necessary provision of basic household needs. This work argues that Shiva's ecofeminist theory and feminine principle should be placed into a historical and political context that led to the transformation of women into agents for positive environmental change. In this perspective, ecofeminist intervention opened a political space for the participation of women in sustainable development and in environmental conservation as experts, instead of as villains or victims.

Alternative Formulations of Women–Environment Relationships

Scholars in disciplines outside geography have also formulated critiques of ecofeminism based on case study research in the rural Third World. These scholars argue that ecofeminism's gender and racial essentialisms legitimize discourses that position women as labor resources for sustainable development projects in the developing world. Some of the most important alternative conceptualizations of the relationship between women and environments have been formulated by scholars such as Bina Agarwal, Cecile Jackson, and Melissa Leach.

Agarwal, coming from a primarily economics background, termed her theoretical approach feminist environmentalism to set it apart from ecofeminism. In her seminal works in 1992 and 2000, she outlines the bases for women's gender based interests in environmental sustainability. Agarwal argues that women may act in the interests of conservation more than men, due to (1) a gendered distribution of natural resource rights, with women more likely to be dependent on common property resources, such as community forests; and (2) a gendered division of labor, with women more likely to be in charge of childcare and the provision of basic household needs. Thus, women and men differ in two critical respects: (1) the extent of and (2) the nature of their dependence on natural resources, and this translates into gender based differences in their relationships to environments and natural resources.

Although feminist environmentalism conceives of gendered human–environment relationships as culturally constructed, context or place specific, and intersecting with other social relations based on class, ethnicity, and age, Agarwal's work nonetheless can evoke an undifferentiated image of the rural peasant woman and her environmental concerns. As a result, feminist environmentalism exhibits clear commonalities with ecofeminism, and it is only in the case of extreme need that the poor peasant woman acts against the interests of conservation and environmental protection, for example, by cutting green trees for fuel in order to prepare food for her children.

The works of both Cecile Jackson and Melissa Leach (both anthropologists by training) are, on the other hand, explicit in conceiving gendered environmental relations that allow for some women's interests to oppose the interests of conservation. Jackson highlights the diversity of women's environmental relations, as an outcome of historical change and differences among women, arguing that women act as environmental agents in both positive and negative ways. Leach's case study on the Mende of Gola, in Sierra Leone, explores how human–environment relations interact with social relations, including gender. What these conceptualizations have in common with each other, and further extend from feminist environmentalism, is the intersection of gender with other social constructions of difference among groups of peoples, such that human–environment relationships can vary as much (or even more) among women as they do between men and women. Thus, perhaps as problematic as an essentialist approach to gender and environment is one that is based upon an uncritical, overly generalized, and fixed understanding of women's roles. These theorists also critique the gender roles approach for having focused exclusively on women, leaving invisible men and men's environmental relations, despite a reality in which women's changing interests and relationships to environments are in part an outcome of their changing relationships with men and with male controlled institutions. Further, a gender roles approach, through its generalizations, tends to produce two images of Third World rural women: (1) women as victims of negative environmental change and (2) women as fixers of nature, such that conservation projects justifiably can harness women's labor (thereby exploiting women and their cheap labor for environmental goals).

Feminist Political Ecology

Feminist political ecology developed in geography through the 1990s, drawing upon empirical field research in political ecology and a constructivist feminist theory of gender. Feminist political ecology presents a strong critique of the essentialist aspects of ecofeminism for a failure to adequately address both the material and the ideological bases for women's environmental actions, and draws upon the alternative conceptualizations (described above) developing outside the discipline of geography. The key element to the approach of feminist political ecology is a grounding of an understanding of women's relation to the environment in the material, historical, sociocultural, and political realities of specific places in the tradition of political ecology. Fundamentally, feminist political ecology is the analytical and theoretical perspective that women and men differ in their experience of and relationship to environments, and that this difference is important in part due to the social construction of gender difference. Thus, the second sense in which we can conceptualize nature as gendered is not that nature is female; but rather, that our relationship with nature is intertwined with our experience of gender.

The work of political ecologists led to the identification of an important lesson in the search to understand women's relation to the environment: the need to focus on gender relations as opposed to purely on women's roles and to see these relations as both affecting and being affected by rural environmental change. This understanding in turn has led to a number of different themes in the study of women's relation to the environment, three of which have been particularly important in geography over the last decade: (1) the variable impact of sustainable development and conservation projects on women and on gender relations; (2) the critical examination of the universalized claims of women's shared interest in environmental protection, based upon empirical research with women in specific places; and (3) the mismatch of women's resource rights with women's growing responsibilities in resource use and management, especially with men's migration.

In 1996, Dianne Rocheleau and colleagues articulated a framework for the emerging subfield of feminist political ecology within political ecology, defining three topics of inquiry within a focus on gendered differences in environmental issues:

1. *Gendered knowledge.* Individuals' knowledge of the environment is an outcome of their material interaction with that environment, which in turn is an outcome of who a person is and the role he/she plays in the household and in the larger community or society. Gender is one important element in a person's identity which structures his/her environmental interactions and thereby his/her environmental knowledge. In other words, women and men experience the environment differently through their labor and social positions and develop different environmental knowledge: knowing about different aspects or knowing about the same aspect in different ways.
2. *Gendered resource and environmental quality rights and responsibilities.* Access to resources, or to environmental quality, also varies according to a person's identity, and gender has played an important role in many places around the world in defining a person's access to and control over different natural resources. Although land tenure is the classic example, feminist political ecologists are also interested in finer nuances of resource tenure, examining, for example, different gender defined rights to different parts of a tree. An interest in how environmental quality rights may intersect with issues of gender inequality connects feminist political ecology to environmental justice literature in the developed world, which although initially focused on questions of race, class, and ethnicity has recently begun an examination of gender inequality in the experience of environmental degradation and risk. In addition to environmental rights, responsibility for natural resource based labor is frequently gendered, reflecting broader gendered divisions of labor. Feminist political ecology views both environmental rights and responsibilities as gendered, but acknowledges the intersection of gender with other social divisions such that different women may experience different rights and responsibilities, and may experience changes in those rights and responsibilities over the course of their lives.
3. *Gendered environmental politics and grassroots activism.* Gender differences in knowledge, rights, and responsibility translate into differences in environmental interests or stakes with regard to environmental change. Although women make up a minority in many of the traditional environmental conservation groups in the

> developed world, they are often in the majority in environmental grassroots movements. Environmental sociologists have found limited evidence of gender differences in environmental attitudes and behaviors overall, but there is evidence that women and men differ in the types of environmental issues with which they engage politically, as well as in the manner in which they carry out their political and activist interests. In many instances of women's engagement with grassroots environmental movements, essentialist ecofeminist positions (particularly those related to motherhood) have been strategically deployed. Feminist political ecologists, and feminist geographers more generally, often have ambiguous feelings about this strategic deployment of gender essentialism.

Since the late 1990s, scholarship in feminist political ecology, and in cultural and political ecology more broadly, has continued a quest to understand how human–environment relationships intersect with various social relationships, such as class, ethnicity, and gender. Geographers have articulated this intersection in a variety of forms and have continued to emphasize that gender relations are mutually constructed with environmental relations. Increasingly, geographers examining human–environment relations are stressing the role of ideology together with materiality, exploring, for example, how social identity (including gender identity) is created with and through environmental relations, with specific material outcomes. With respect to the gendered nature of environmental relations, there is a developing exploration of women's agency to (re)define their identities, including their identities in relation to environments, and the material consequences of these processes.

Feminist theory has played an important role in geography's broader examination of human–environment relations and of nature–society interactions, beyond the bounds of feminist political ecology. Donna Haraway has been particularly influential in this respect, with her destabilization of the nature/culture dualism through the metaphor of the cyborg, a creature both animal and machine. Geographers like Sarah Whatmore also destabilize the nature/culture dualism through the concept of hybrid geographies, in which the boundary between human and nonhuman is transgressed.

Some Related Fields of Inquiry in which Human Geographers Participate

Women and Sustainable Development

Women and sustainable development (or women, environment, and development) as a field of inquiry, but also especially as a practice, emerged out of the two international fields of women and environments, and women and development. Shiva's assertion of a feminine principle and her romanticization of the poor rural woman as living in harmony with nature were particularly influential in sustainable development discourse and practice as it developed through the late 1980s and the 1990s for nongovernmental, governmental, and multilateral development organizations. Shiva and others wrote widely of the Chipko movement in India and the Greenbelt movement in Kenya, both of which became symbols, first in ecofeminist writing and then in international policy discussions, of women's closeness to nature and their proclivity to protect the environment.

The critiques of ecofeminism's essentialism have since been incorporated into international policy oriented discussions on women, environment, and development, with constructivist arguments replacing essentialist ones. Yet, discussions continue either to universalize the experience of rural Third World women, with women's environmental interests based upon a generalization of women's roles, or to ignore gender and gender based power inequalities altogether. The growing body of work in geography (as well as in disciplines such as anthropology) that explores the complex intersections of environmental relations and gender relations in greater detail and diversity, to date, have had little influence on the discourse of sustainable development and the formulation and implementation of conservation and development projects. As women, environment, and development discussions positioned women in sustainable development discourse as universal caretakers and knowledge repositories, on the ground conservation projects fail to include even basic gender analysis, further supporting universalized notions of the relation of women to the environment.

Gender and Agriculture

Another related field has been the study of gender and agriculture, which began with the seminal work of Esther Boserup in the early 1970s. The field began with an exploration of the various labor roles of women in agricultural production, and has since also focused on women's access to and control over land, both in terms of formalized (legal) land tenure and in terms of practice. Most of this work has taken place in the context of the Third World. Recent debates concerning women's access to and control over land center on whether or not increased gender equity in land rights will necessarily lead to women's gender empowerment. Recent work in geography in the field of gender and agriculture in the developed world has focused on women's participation in organic or sustainable farming movements, bringing in questions of gender identity, differential capital access, and social constructions of farming and nature.

Gender and Sustainable Livelihoods

Robert Chambers and other colleagues developed the concept of sustainable livelihoods in the early 1990s as a means to understand poverty and inform poverty reduction policies. A typical livelihoods framework is based on identifying different types of capital (e.g., natural, social, physical, financial, and human) and how these are combined by individuals and households in their various efforts to make a living. A number of scholars and practitioners have noted that women and men have differential access to and control over these capital pools, and that these differences mean that men and women face different challenges with respect to creating and maintaining sustainable livelihoods. As a result, men and women may adopt different livelihood strategies. Recently in geography, Ann Oberhauser and colleagues have argued that these strategies are more than specific economic practices, and that women's livelihood strategies include struggles to transform social identities and relationships.

See also: Agriculture, Sustainable; Conservation and Ecology; Culture/Natures; Environment; Feminism/Feminist Geography; Feminist Geography, Prehistory of; Gender and Rurality; Hybridity; Livelihoods; Natural Resources; Nature; Nature, Social; Nature-Culture; Patriarchy; Peasant Agriculture; Political Ecology; Sustainable Development.

Further Reading

Agarwal, B. (1992). The gender and environment debate: Lessons from India. *Feminist Studies* 18, 119 158.

Buckingham, S. (2004). Ecofeminism in the twenty first century. *The Geographical Journal* 170, 146 154.

Domosh, M. and Seager, J. (2001). *Putting Women in Place: Feminist Geographers Make Sense of the World* (ch. 6: The environment). New York: The Guilford Press.

Jackson, C. (1993). Doing what comes naturally? Women and environment in development. *World Development* 21, 1947 1963.

Leach, M. (1994). *Rainforest Relations: Gender and Resource Use Among the Mende of Gola, Sierra Leone*. Washington, DC: Smithsonian Institution Press.

Leach, M. (2007). Earth mother myths and other ecofeminist fables: How a strategic notion rose and fell. *Development and Change* 38, 67 85.

Nightingale, A. (2006). The nature of gender: Work, gender, and environment. *Environment and Planning D: Society and Space* 24, 165 185.

Oberhauser, A. M., Mandel, J. L. and Hapke, H. M. (2004). Gendered livelihoods in diverse global contexts: An introduction. *Gender, Place, and Culture* 11, 205 208.

Rocheleau, D., Thomas Slayter, B. and Wangari, E. (eds.) (1996). *Feminist Political Ecology: Global Issues and Local Experiences*. New York: Routledge.

Sandilands, C. (1999). *The Good Natured Feminist: Ecofeminism and the Quest for Democracy*. Minneapolis, MN: University of Minnesota Press.

Shiva, V. (1989). *Staying Alive: Women, Ecology, and Development*. London: Zed Books.

Sturgeon, N. (1999). Ecofeminist appropriations and transnational environmentalisms. *Identities* 6, 255 279.

Trauger, A. (2004). 'Because they can do the work': Women farmers and sustainable agriculture in Pennsylvania, USA. *Gender, Place and Culture* 11, 289 307.

Relevant Websites

http://www.energia.org
Energia International Network on Gender and Sustainable Energy.

http://www.fao.org/gender/
Food and Agriculture Organization of the United Nations (FAO), Gender and Food Security.

http://www.genanet.de
GENANET Focal Point Gender, Environment, Sustainability.

http://www.genderandwater.org
Gender and Water Alliance.

http://www.unep.org/gender env/
United Nations Environment Programme (UNEP), Gender and the Environment.

http://www.gendercc.net
Women for Climate Justice.

http://www.wedo.org
Women's Environment and Development Organization (WEDO).

http://www.genderandenvironment.org
World Conservation Union (IUCN), Gender and Environment.

Natures, Postcolonial

M. Sioh, DePaul University, Chicago, IL, USA

Glossary

Essentialism An identity as a given and immutable.

Global South An alternative term to ‘developing countries’ referring to poorer countries that are mostly located in the Southern Hemisphere. Most of these countries are also former European colonies.

Objectivism A doctrine that holds that there is a reliable correspondence between reality and an external world that is observable.

Overdetermined A post-structuralist term signaling originary conditions that will lead to a certain outcome.

Paradigm A concept, less defined and structured than a model, used by Thomas Kuhn to refer to a worldview.

Relativism A theory that meaning is filtered through our cultural and historical background and context and often used in opposition to objectivism.

Social Construction A school of thought that claims that reality is an artifact of its social and cultural context and often used in opposition to essentialism.

Stereotype An oversimplified conception of members of a group.

Introduction

The study of postcolonial nature is a theoretical engagement with the impacts of colonialism on reconceptualizing nature, usually forests and wild species, not only in developing countries but also in parts of the developed world where racial minorities have to grapple with a legacy of colonialism. Nature, in its popular sense, is taken to mean essence or inherent properties, external to human society, or the opposite of culture. Postcolonialism is a school of cultural criticism concerned with the impact of colonialism on both colonized and colonizing peoples in the present, as opposed to postcolonialism which refers to the chronological break after political independence. Postcolonial nature studies follow as a critique and an alternative paradigm to sustainable development, and more generally, the environmental movement. Postcolonial nature studies is a hybrid of nature studies and postcolonialism, complex areas of inquiry derived from, and given impetus by, the intellectual currents of post structuralism, postmodernism, and social studies of science that became influential in the social sciences and humanities in the 1980s and 1990s. While definitions of both the terms ‘postcolonial’ and ‘nature’ are contested, multiple, and fractured, postcolonial nature as a problematic is characterized by its definition with respect to or in contrast with the stereotype of the native.

The political significance of the widespread intellectual acceptance of nature’s ontological character has been its use to justify morally reprehensible programs of conquest and material appropriation. This trend was most clearly seen in the late nineteenth to early twentieth century vogue for environmental determinism which coincided with the era of high imperialism. In its most basic articulation, environmental determinism justified colonialism, because if we are what nature makes us, then the populations of certain parts of the world are born to conquer and lead just as others are born to be conquered and led. The rise of natural history as a discipline was the result of European colonial activity by scientist explorers, such as James Cook, Alexander von Humboldt, Joseph Banks, and, most famously, Charles Darwin. Consequently, scholars have also traced the origins of the modern conservation movement to the European tropical island colonies inspired by the search for the Garden of Eden, itself derived from the Zoroastrian, and therefore non European, notions of *Pairidaeza* (paradise) and *garodaman* (garden).

Postcolonial nature studies is concerned with the epistemological and ecological consequences of this construction of ‘natural’ history. Despite the twinning of natural history with political ventures, even when environmental determinism was succeeded by cultural and landscape ecology in the second half of the twentieth century, nature conceptions continued to be viewed as politically neutral. Sustainable development is now the prevalent model for managing nature in developed countries that is being exported to the rest of the world, including societies where respecting nature was the custom before the advent of Western capitalism and modernity. Sustainable development is seen by aid and development agencies as especially urgent in developing countries because environmental degradation there is perceived, ahistorically and contextually in isolation, to be caused by a combination of high fertility rates, ignorance, and irresponsible governments.

Marxist influenced political ecologists have responded to this neutral conceptualization of nature by arguing that nature is a product of profit driven transformation. Some Marxists have gone beyond arguing that modern capitalism simply exploits and transforms nature to argue that modern capitalism produces nature, and that in turn, nature is used for political legitimation of social relations. Cultural geographers and historians too have studied the role of nature as a symbolic source in Western cultures for representing social ordering, or in contrast, the role of nature as a common heritage in masking social ordering. Particular aspects of nature can be used to symbolize ontological group identity, such as the giant redwood groves in the United States that came to symbolize the national magnitude as well as the ancient lineage of the new nation in the nineteenth century. Thus, conservation of the redwoods became synonymous with patriotism. Alternatively, nature conservation is seen as originating in the colonies where reformist scientists attempted to recreate Eden to compensate for the industrial destruction of nature in their home countries. However, the explicit theorization of cultural constructions of postcolonial nature emerged out of the fields of ecofeminism, post structuralism, postcolonialism, and social studies of science.

Intellectual Influences

Ecofeminism

Ecofeminism links the domination of nature to that of women. In analyzing descriptive and normative statements about nature and women, ecofeminists observe that changes in the latter are often preceded by changes in the former. Ecofeminists identify the turning point in the constructions of nature and women as the Enlightenment, before which nature was identified as a nurturing mother and the center of an organic cosmology. This view is still central to non European ecological models. From the sixteenth century onward and with the advent of the Scientific and Industrial Revolutions, ecofeminists argue that women and nature were viewed as wild resources to be tamed and exploited materially, psychologically, and recreationally by harried men/entrepreneurs. But ecofeminists in post colonial countries emphasize that their societies experience the triple domination that pose women, natives, and nature as inferior to men, Westerners, and technology. For example, the introduction of land enclosures and highly capitalized, monocultural cropping for profit during colonialism devalued native women more than men because women were engaged in subsistence agriculture and gathering wild forest products. The destruction of women's livelihoods and the unequal distribution of resource access were justified in the name of theories of progress, gender, race, and nature. These theories ascertained the natural abilities of certain groups to participate in the introduced agricultural practices and deemed a material world not incorporated into the capitalist system as wasteful. Ecofeminists from postcolonial societies argue that such theories thought to be universal and objective are actually masculinist, ethnocentric, and historically specific, and policies informed by them will only perpetuate the current social inequalities and loss of biodiversity resulting from the accelerating spatial expansion of monocultural cash cropping.

Post-Structuralism and Discourse

The challenge to the notion of an essential identity by ecofeminists, and feminists in general, signaled the opening of the broader debate about the theoretical basis of knowledge. This set of challenges is often somewhat simplistically referred to as post structuralist, a mainly French school of literary theory that critiques the rational, objective, autonomous, self aware subject of humanist thought. Theory has a privileged position because it governs the rules of judging knowledge: what counts as evidence and what constitutes an argument. But theory's privileged position depends upon the existence of an underlying, unchanging external truth associated with nature. The claim of a fixed, external nature has been challenged by Michel Foucault as an historically constituted idea that emerged during the Enlightenment. Hence, the social context of enquiry, rather than the world in which the investigation takes place, constructs what we find, as opposed to us discovering a reality fully formed. Where ideology argues for single imposed truth, post structuralists, after Foucault, prefer the term discourse. The term discourse refers to a system of signs that constructs meanings that are then represented as essential truths. Discourse works by selecting and organizing information, establishing relevances, defining unities, and describing hierarchical relations between them. Yet, for discourse to appear legitimate, meaning must appear ontologically derived, so the rules for organizing discourse must be obscured. Because perceptions of racial, class, and gender differences and hierarchies frequently encode each other, interlocking discourses strengthen each other to form, eventually, a regime of truth. The post structuralist turn in the social sciences and humanities has meant that theories are viewed more as stories with their own assumptions, contradictions, and silences that challenge the correspondence between reality and representation. Instead, knowledge is perceived as partial, contextual, situated, and reflecting power relations. Terms such as 'nature' and 'native' are constructed through language and are unstable in communication. Yet, even well meaning policies are enacted based on assuming the immutability of these terms, thus continuing the conditions that overdetermined the past environmental degradation.

Postcolonialism and the Natural 'Other'

Postcolonial studies emphasize that the social relations bequeathed by colonialism are more than just political or economic but also a cultural process imagined and energized through language. The interest in the legacy of colonial representations can be traced to post structural theories inflected through Edward Said's classic critique of Orientalism, which questions how ideas acquired the status of natural truths during the period of high imperialism in the nineteenth and twentieth centuries. Orientalism supplied colonialism with an enunciative capacity which was transmitted from one generation to the next through repetition and institutional embeddedness. Postcolonialists argue that political independence has been merely cosmetic and that colonially dictated representations have been redeployed and reaccentuated through administrative forms, institutions, and relationships. Said argued that recurrent, systematic tropes and asymmetries in representation – which he described as a discourse – ran through colonial texts. Colonial culture has to be understood as a material network of connections between power and knowledge. The Orient exists for the West, and is constructed by, and in relation to, the West. It is a mirror image of what is inferior and alien ('Other') to the West. Orientalism is a style of thought based upon an ontological and epistemological distinction made between 'the Orient' and 'the Occident' in which an essentialized image of a typical Oriental is represented as culturally and, ultimately, biologically inferior. In the same manner, postcolonial nature scholars have analyzed the way nature, even in its positive constructions, was conceived as a native dwelling 'other' space to the spaces of colonial towns and plantations, and as the ultimate visible symbol of the difference between European civilization and its wild 'Other'.

While postcolonial theorists agree that neocolonial power depends upon its authority to fracture the subjectivity of its native subordinates by encouraging them to identify with Western values and cultural norms, critics of Said such as Homi Bhabha argue that the critique of Orientalism asserts too much intentionality and direction in the flow of power. Instead, Bhabha emphasized the instability and hybridity of identity and ambivalence in neocolonial relations. For Bhabha, the point of stereotypes is precisely that the endless repetitions assuage colonial insecurity by fixing identity. But even stereotypes can have a temporary purchase on imagination that allows a discourse to judge and control. Stereotypes derive their power precisely because they are unstable and ambivalent and can be modified to change with times. The colonized and postcolonized themselves accept certain stereotypes of power and show their ambivalence in their desire for Western tools and symbols of power – science and technology. Postcolonial scholars instead urge us to be attentive to how signification is affected by particular sites and contexts of enunciation, or what Donna Haraway has called 'situated knowledge'. In Bhabha's 'mimicry' concept, native copying of white behavior is never quite the unadulterated homage to a superior culture, because the performance and enunciation are always different from the original and intended. In the course of translation from their sources to the rest of the world, scientific ideas, narratives, and theories are rearticulated into hybrids that dilute the intentions of power. For postcolonial nature scholarship, Western notions of nature conservation are examples of the slippage between, and hybridization of ideas as paradigms of resource management and sustainable development are absorbed, yet undermined, through being spliced onto notions of national sovereignty and local practices of territorialization. These notions of conservation may sometimes work against Western power, but they almost always work against indigenous tribal groups who experience a double colonization.

Gayatri Spivak takes up the issue of addressing multiple levels of colonization by using deconstruction techniques to examine contradictions in official texts and subplots and minor characters for the different stories they tell us. While postcolonial nature studies have relied on deconstruction of archival documents and unofficial texts such as letters and diaries to open up different narratives of colonial and postcolonial natures, arguably, it is Spivak's notion of 'strategic essentialism' that has been most usefully deployed in practical struggles over control of material resources. Strategic essentialism refers to a strategy that while strong differences may exist between members of ethnic minority groups, it is sometimes advantageous for them to temporarily 'essentialize' themselves and bring forward their group identity in a simplified way to achieve certain goals. Spivak has since disavowed the ways in which she claims the notion of strategic essentialism have been mis used. This has been used not only by certain native groups to emphasize their suitability to manage natural resources in Western societies but also by tribal groups in the Global South in alliance with international conservation groups. However, Spivak has cautioned postcolonial scholars on whether it is ever possible to represent anyone or anything else accurately or whether we should abandon attempts to represent an 'inaccessible blankness'.

Social Studies of Science and Construction of Nature and Natives

Deconstructing representations of race and gender have achieved significant purchase in the intellectual arena, but the application of the same methodology has engendered heavy resistance in the context of nature studies. While not claiming that the material world is entirely socially constructed, because nature cannot express itself, post structuralists, who would include social constructionists in their ranks, contend that there can be no objective view of nature since we are always left with human representations that construct nature through language, in particular those of scientists. Critics of the social constructionists contend that this is tantamount to stating that all knowledge is relative, and that there can be no standards for judging among knowledge claims. Because science is used to enforce moral codes and justify rewards, then controversies about scientific knowledge have repercussions for moral discourse. Nature in postcolonial countries is triply disadvantaged through the arbitration of mainly male, white scientists and indigenous scientists who disavow native knowledge as part of their claim to legitimacy through an alliance with Western scientific rationality. Within the field of social studies of science, proponents contend that scientific enquiry was more subjective, flexible, and historically contingent than was previously acknowledged, in terms of who belongs to a community, whose views are noted, and how consensus is reached. The newer generation of social scientists (1980s onward) studying science differs from the older generation who claimed that although knowledge is subjective, truth can still be known through experience, whereas the newer generation argues that the 'truth' of a statement is based on the degree of its links to, and embeddedness within, other statements. The focus on representation and relativism as opposed to subjective and experiential knowledge has implications for postcolonial nature scholars who have shifted methodological emphasis to archival sources in place of empirically measurable information.

Understanding why certain representations of nature prevail becomes a matter of determining the degree of its links to, and embeddedness within, other discourses such as gender, colonialism, nationalism, and global capitalism. One way of approaching the issues of linkages and embeddedness is as a network made up of various discourses that interact, sometimes in contradiction to each other, but with an overall discernable direction. The process of acquiring allies, both human and nonhuman (such as machines) can be examined using Bruno Latour's network theory to understand how discourses are translated into practices of power. Latour's network theory proposes that in order to acquire the status of truth, a paradigm must acquire allies. Like the post structuralists, Latour sees a scientific claim as a regime, which acquires the status of truth when it has collected enough allies, which can be in the form of interlocking discourses. Fact construction is a collective activity, and consensus lies in the creation of a network of commensurabilities, meaning that participants must agree on the general aim of the project at hand even if on a point by point basis they do not agree. Within contemporary nature discourses, sustainable development is the corporate identity that governs the way the material world can be shaped and deployed, because its ambiguous definition allows it to mean different things to different constituencies. Simultaneously, sustainable development incorporates the language and concepts of powerful existing discourses of colonialism, race, gender, class, science, and progress, which while not unchanging over time, have been reaccentuated to subvert non Western views of nature.

Interlocking Discourses: Postcolonial Natures

Modern dualistic and individualistic rationalities underpin environmental research and practice, which have been redeployed to the detriment of non Western practices of nature conservation. While discourse does not unilaterally determine what can be said about, and done to, nature, a whole network of interests is involved whenever 'nature' is invoked. In summarizing the theoretical treatments above, the various significations of nature stem from its place in social relations of power. Postcolonial nature is given meaning through specific material and discursive practices; its physical transformation embodies the (neo)colonial relations of power through which dominant images are produced. Colonial power relations were starkly symbolized in opposing landscape formations of forests and plantations. Nature, symbolized by forests, was cast as the negative 'other' of the plantation's modernity and role as an oasis of scientific order, just as natives were cast as the negative 'other' to Europeans. These images are projections of nonspatial signifiers onto the physical space of nature. As the physical environment changed, forms of consciousness – perceiving, symbolizing, and analyzing nature – were also reconstructed. In the process, the colonial and capitalist transformations legitimated a set of symbols that placed Europeans and technology above natives and nature. Imperialism was not just the result of the power to take over territory but also a justificatory regime for its actions.

Postcolonial nature is an idea that has a history and a tradition of thought, imagery, and vocabulary that have given it reality and presence in society in the same manner as contemporary discourses on the Global South that are beyond any correspondence with a real place. Nature has been made to stand for many things: as a resource in itself for export, as an impediment to capitalist agriculture to be replaced with plantations, farms, and ranches, and as a source of fear and a symbol of anxieties about national identity. Moreover, the dominance of the discourses that turned nature and the Global South into negative places lies in their 'knitted together strength', that is, the very close ties between socio economic and political institutions. Applying actor network theory to postcolonial nature, we see that the status of nature depended on its role within the colonial enterprise, as well as the larger project of progress upon which postcolonial nationalism would depend. The fol lowing are common tropes in narratives of the post colonial nature–native nexus.

Resource Management and Environmental Stewardship

The most dominant, and still the most dominant, trope of colonial and postcolonial nature is its economic potential. During the colonial era, European perceptions of land scapes were always inflected through the needs of empire. Natives were seen as too lazy and/or too ignorant to exploit nature to its fullest potential. Although the value of plantation (or ranch) products, and timber was eco nomic, the justification for replacing the forests (and grasslands) went beyond the purely economic to integrate modern concepts of rational order and scientific progress. While exploiting nature is now associated with con temporary environmental degradation rather than being praiseworthy, appropriation and exploitation of nature nonetheless goes on. The forestry industry and the en vironmental movement represent 'nature' or 'wilderness' as a realm separate from 'culture' while relocating it within the abstract space of the global commons through representations and rhetorical practices of responsible stewardship in which Western trained scientists, bur eaucrats, and capitalists are better managers of nature than native peoples.

Liberal Western attitudes toward the nature–native nexus take two forms: romanticization and alliance with tribal groups as the last innocent links to nature or an tagonism and hostility toward many governments in the Global South as irresponsible stewards of global com mons and oppressors of tribal groups. Both forms involve essentialization of natives either through a fetishization of the exotic in the case of tribal peoples and their 'oneness' with nature, or the subtle implications of ignorance, oppression, or corruption, or all three in the case of post colonial states that do not comply with Western en vironmental conservation models. Of course, the trade in stereotypes need not necessarily be unidirectional or disempowering. Employing strategic essentialism allows disempowered groups to represent themselves as the best equipped to halt environmental damage wrought by ex ternal sources and conserve nature because of their perceived special kinship with it. This is not to deny the validity of the experience and situated knowledge but to highlight that if disempowered groups are to achieve their aims of controlling their immediate environment, they need to incorporate the language of power.

Territorializing Nature

From the perspective of governments of developing countries, Western environmentalism is a neocolonial interference in its sovereignty and the championing of tribal peoples and knowledge a Trojan horse for under mining the attempts of developing societies to modernize and escape poverty. This view was most forcefully ar ticulated in the 1990s by Malaysia's then Prime Minister Dr. Mahathir, who harks back to the colonial en couragement of native innocence and unworldliness that had disastrous consequences – the theft of native lands and resources by Europeans. Historically, colonial con servation authorities saw indigenous peoples as the main cause of environmental destruction through native agri cultural practices and hunting (which Europeans referred to as 'poaching'). Hence, conservationist ideas in post colonial states have been associated with political conservatism. Indigenous or traditional ecological knowledge, then, can only be essentialized and praised when there is nothing to fear from it. Ironically, given the stress on neocolonialism, for tribal peoples, the loss of ancestral lands through appropriation or physical deg radation has resulted in a condition of double coloniza tion, first from the Europeans and then from the majority ethnic population in the post colonial nation states.

But in the latter half of the twentieth century onward, nature construction has proceeded in tandem with state building. Transforming the physical landscape has been, and is, closely related, to political landscape formation. According to Max Weber, an important function of the state then is to demarcate where the boundaries of its territory lie and to maintain order within those bound aries. In the context of forest based decolonization wars and the Cold War in the 1940s through to the 1970s, nature in the form of the tropical forest was socially constructed as undesirable. Since the state's boundary is the limit of the state's coercive supremacy, then the for est, as beyond the control of the state, was perceived as a frontier that had to be territorialized. Forests were the antonym to the controlled domain of the colony or the nation state and the synecdoche for insurgency. During the Cold War, decolonization military campaigns often segued into insurgencies in many newly independent countries. Once seen as securely territorialized by the colonial powers, post colonial states became new terri tories to be colonized as frontiers between the West and Communist Russia and China. These forest based in surgencies became the symbol of this unterritorialized world. The US Secretary of State, Dean Rusk, resisted troop pullout from Vietnam by claiming that the alter native was a world situation that disintegrated into the 'law of the jungle'.

Finally, post colonial states needed a positive pro jection of territorialization and some postcolonial nature scholarship has focused on large scale agricultural schemes as constructions of a new nature. Multiple repetitions of agricultural rituals across space act as positive productions of state power, while making it fruitful for capitalism. The scale, organization, and standardization of state supported agricultural schemes marked them as the 'new nature' for a modern nation produced through scientific techniques.

Criticisms

In general, criticisms on postcolonial nature studies have focused on the overemphasis on textuality and culture. Marxists, especially, criticize its methodological lack of empirical groundedness and paying insufficient attention to other more traditional configurations of power such as capital. The multidisciplinary nature of postcolonial nature studies also poses the challenge of how to avoid incoherence in such a wide open field. Moreover, much work remains to be done in the area of contesting con structions of nature in the Global South. Ironically, the criticisms themselves reflect the deeply rooted problem at the core of the debates on essentialized notions of authentic representation and epistemic legitimacy taking place across the humanities and social sciences whenever the terms 'native' and 'nature' come into play. However, the fact that postcolonial nature studies is a hybrid of different schools of research augers well for its devel opment as an alternative and more respectful paradigm for the relationships between society and nature, and within societies by heeding different voices.

See also: Land Change Science; Nature; Political Ecology; Postcolonialism/Postcolonial Geographies; Poststructuralism/Poststructuralist Geographies; Sustainable Development.

Further Reading

Demeritt, D. (1994). The nature of metaphors in cultural geography and environmental history. *Progress in Human Geography* 18, 163 185.
Escobar, A. (1995). *Encountering Development: The Making and Unmaking of the Third World*. Princeton, NJ: Princeton University Press.
Grove, R. H. (1995). *Green Imperialism: Colonial Expansion, Tropical Island Edens and the Origins of Environmentalism, 1600 1860*. Cambridge: Cambridge University Press.
Gupta, A. (1998). *Postcolonial Developments: Agriculture in the Making of Modern India*. Durham, NJ: Duke University Press.
Marglin, S. (1996). Farmers, seedsmen, and scientists: Systems of agriculture and systems of knowledge. In Marglin, F. E. & Marglin, S. A. (eds.) *Decolonizing Knowledge: From Development to Dialogue*, pp 185 248. Oxford: Clarendon Press.
Merchant, C. (1989). *Ecological Revolutions: Nature, Gender and Science in New England*. Chapel Hill, NC: The University of North Carolina Press.
Moore Gilbert, B. (1997). *Postcolonial Theory: Contexts, Practices, Politics*. New York: Verso.
Morgan, S. (1996). *Place Matters: Gendered Geography in Victorian Women's Travel Books about Southeast Asia*. New Brunswick, NJ: Rutgers University Press.
Prakash, G. (1999). *Another Reason: Science and the Imagination of Modern India*. Princeton, NJ: Princeton University Press.
Said, E. (1978). *Orientalism*. New York: Pantheon.
Schama, S. (1995). *Landscape and Memory*. Toronto, ON: Random House.
Shiva, V. (1988). *Staying Alive: Women, Ecology and Development*. London: Zed Books.
Sioh, M. (1998). Authorizing the Malaysian rainforest: Configuring space, contesting claims and conquering imaginaries. *Ecumene* 5, 144 166.
Sioh, M. (2004). An ecology of postcoloniality: Disciplining nature and society in Malaya, 1948 1957. *Journal of Historical Geography* 30, 729 746.
Willems Braun, B. (1997). Buried epistemologies: The politics of nature in (post) colonial British Columbia. *Annals of the Association of American Geographers* 87, 3 31.

Neighborhood Change

J. Kenny, University of Wisconsin–Milwaukee, Milwaukee, WI, USA

Glossary

Factorial Ecology Drawing upon the theoretical framework of social area analysis, the application of factor analysis and principal-components analysis to study urban residential patterns.

Family Life Course/Lifecycle Assuming a traditional nuclear family, households may be classified according to stages in the family life course or lifecycle (e.g., single, childbearing, and empty nesters) that carry with them expectations for housing needs and preferences.

Filter-Down Process As theorized by land economist, Homer Hoyt, a process whereby high-income households continually move out to more desirable housing, leaving their houses vacant and creating a chain of occupancy for middle-income households, who then leave their housing to be occupied by lower-income households.

Gentrification The conversion of a low-income district into one dominated by middle- to high-income residents.

Housing Submarkets Distinctive types of housing in localized areas of a city that tend to be occupied by households with similar socioeconomic characteristics.

Human Ecology Introduced by the Chicago School of urban sociology in the early twentieth century, this perspective examines the social and demographic compositions of neighborhoods assuming that the sociospatial patterns reflect the competition that exists among various groups for favorable locations.

Social Area Analysis Approach to modeling the internal structure of metropolitan areas that evaluates how family characteristics, economic rank, and ethnic status interrelate to produce sociospatial patterns.

Definitions

Despite the familiarity of the word, neighborhood as a focus of research involves a degree of ambiguity, since it contains both geographic (place oriented) and social (people oriented) aspects. Many use the word interchangeably with community, another term that is difficult to pin down. However, as generally defined, the neighborhood focuses on a residential area – houses and their immediate environment, including local institutions such as churches, schools, and local services – demarcated by subjective boundaries. Evaluated functionally, these boundaries are determined by the intersection of household types, ethnic, and lifestyle characteristics, and social class associated with area residents.

Although the dynamic nature of neighborhoods is inherent in the concept, attention to change further complicates this discussion, since an effort to evaluate neighborhood character (or neighborhood quality) might involve a focus on physical decline or improvement, a consideration of shifting demographic characteristics, or both. Traditionally, emphasis has been given to housing submarkets and the role of residential mobility in shaping and reshaping neighborhood character within a metropolitan area. More recently, researchers have also considered the ability of neighborhood residents to organize in the face of unwanted change. Various themes in urban geographic research, as demonstrated below, highlight the extent to which geographic studies have reassigned significance to discussions of neighborhood change as both disciplinary research agendas and the conditions in cities themselves are transformed.

Research Themes

Among the earliest analyses of urban space focus on issues of neighborhood changes. The credit for this generally goes to members of the early twentieth century Chicago School of Sociology and the study of human ecology. Borrowing concepts from animal and plant ecology, Chicago sociologists developed their analyses by collecting extensive ethnographies of various social and ethnic groups to document communities and territory, and by mapping out social difference to assess urban growth and neighborhood change. 'Natural areas', that is, territorial units roughly the equivalent of neighborhoods, served as the basic unit of analysis reflecting the relative attractiveness of districts within the city and the competitive power of a particular residential group. Human ecological analyses assume that urban areas constantly experience shocks to their social–spatial arrangements as various groups expand and adjust territorially through processes of 'invasion', 'dominance', and 'succession'.

Attempting to find evidence of the city's internal structure in the patchwork of Chicago's neighborhoods, sociologist Ernest Burgess contributed the widely recognized concentric zone model to the social sciences in the 1920s and with it a generalization of urban growth based on the dynamics of change at the neighborhood level. The concentric zone model, consisting of rings radiating outward from the central business district, represented greater economic and social status with each successive residential ring. Underpinning this pattern was a history of successive waves of immigrants drawn to inexpensive housing and jobs in the industrial city and the consequent expansion outward of the 'zone of transition' (defined as the central city's factory zone and adjacent areas of deterioration). While these zones imply competitive conditions, the 'natural areas' (or neighborhoods) within any particular zone were characterized as symbiotic, as expressed in the residents' shared sentiments, traditions, and local interests.

Numerous critics of Burgess' sociospatial model offered alternative perspectives on neighborhood change. In the late 1930s, land economist Homer Hoyt reformulated the residential model by changing the zones to sectors and arguing that the desire of the wealthy for new homes fueled urban growth and development. Sectors developed, he noted, as higher status uses developed adjacent to one another moving outward along transportation corridors or other favorable landscape features. His critique of the zonal model stemmed from his study of patterns of rental values in American cities. Sponsored by the Federal Housing Administration in the last years of the Depression, his study sought to predict mortgage lending risks in different types of neighborhoods by identifying the relationship between socioeconomic characteristics of residential populations and the stability of neighborhoods. Hoyt theorized that the central dynamic of the housing market depended on the transfer of obsolete housing from higher income households, who vacate in order to occupy more fashionable dwellings, to middle income households. Thus, he contributed the concept of the filter down process and the term 'filtering' to housing market vocabulary, arguing that status and the desire for novelty serves as the mechanism for change by driving urban growth and creating a chain of vacancies in the housing stock that moves previously owned units down the social scale.

Burgess' and Hoyt's models both drew from the Chicago School of Sociology's close examination of the city's changing neighborhood conditions. Yet when geographers Chauncy Harris and Edward Ullman introduced them to urban geography in their seminal article on the nature of cities in 1945, the models' iconic status as generalizations of the city's internal structure tended to dwarf the human ecology research agenda. Particularly when introduced into the study of mid twentieth century North American urban geography, static zonal patterns rather than dynamic processes typified the models' interpretation and application.

The influence of human ecology on British social geography persisted more effectively and attention to urban residential patterns was sustained as a dominant aspect of research through the 1960s and 1970s in Britain. British geographer Ceri Peach acknowledged the Chicago School's influence by applying the term spatial sociology to work that used spatial analysis to clarify social structures. Methodologically, it borrowed from forms of analysis developed by sociologists in the 1950s. On the one hand, a reinterpretation of social area analysis – factorial ecology – analyzed social differentiation through multivariate statistics. On the other hand, this spatial sociology addressed issues of segregation through measures of social separation such as the index of dissimilarity. Each approach served to vindicate the school's association of social difference with spatial distance and provided both a new vocabulary and new measures for academic and applied research related to socially fragmented cities. The futures of these methodologies varied.

Social area analysis was developed by sociologists Eshref Shevsky and Wendell Bell in the 1950s. Theorizing increasing social differentiation within urban areas, the researcher evaluates census tracts to assign them a position on a scale for each of three indexes: family characteristics, economic status (or social rank), and ethnic status (or segregation index). This method of classifying and describing different neighborhoods links the three constructs to the social space of the city and highlights residential patterns associated with changing urban lifestyles. The focus on residential areas underscored the researchers' interest in the neighborhood as an important context for community formation, group identity, and the shaping of individual attitudes.

Geography's adaptation of social area analysis, factorial ecology, gained ground at the advent of the quantitative revolution in the 1960s and 1970s as computers offered greater opportunity to manage a large number of variables related to the three indexes of social variation. The conceptualization of these indexes is taken up below in discussions of neighborhood lifecycles and housing markets. While factorial ecology analysis can handle a great deal of information, and continues to be used in marketing work to define a metropolitan area's residential structure, it falls short of explicitly accounting for neighborhood change since it tends to produce data snapshots. Although still in use, the peak influence of factorial ecology within urban geography ended in the 1970s as geographers sought more critical perspectives on neighborhood change, particularly as questions related to racial segregation and gentrification assumed greater significance.

Indexes of segregation, again pioneered by sociologist in the 1950s, continue to be an important measure of social integration of various racial and ethnic groups and the character of particular neighborhoods. The index of dissimilarity – the most commonly used of the indexes, which includes the index of segregation and the index of isolation (also called the index of dominance) – compares the spatial distribution of two groups to answer a question. What percentage of one group would have to move to other areas within the city to achieve a proportional distribution comparable to their proportional distribution within the city as a whole? This simple statistic provides a means of addressing changes in neighborhoods over time as well as suggesting the relative assimilation of particular ethnic and racial groups. For example, geographer Ceri Peach drew upon such figures in recent (and ongoing) debates as to whether segregation of South Asian Mus lims in British cities is growing or declining. Although undoubtedly powerful statistics for addressing particular conditions, interpretations related to individuals' as similation are frequently made from such figures – and this possibility suggests just one of the many critiques associated with the Chicago School and its positivist approach.

When inferences are made about individuals, based on aggregated spatial patterns, an error can be produced, known as an ecological fallacy. An individual's behavior cannot be assumed based on knowledge of the group's statistics. This error of interpretation is well known to introductory statistics students, although, unfortunately, knowledge of the term does not safeguard research. Perhaps a more critical issue related to such analysis is that it is difficult to infer process from the pattern asso ciated with group information. Behavioral geography's interest in residential choice modeling responded to this particular challenge as it attempted to model individual decision making.

Behavioral geography's influence on studies of neighborhood change reflects both the contemporary critiques of geography as a spatial science and the par ticular social conditions of the 1960s. By the second half of the twentieth century, increasing complexity and social polarization in Western cities underscored once again the patterns of investment and inequality in urban areas and made housing abandonment and racial segregation sig nificant themes in urban social geography. As numerous, large American cities experienced change in their racial composition, studies examined the 'tipping point' in housing markets. That is, they questioned at what point do residents respond to the in migration of a new mi nority population by suddenly and rapidly leaving the residential area themselves. The 'prisoners' dilemma', one of the simplest forms of gaming models, articulated the logic used by property owners with incomplete know ledge in the face of changing conditions. Assumptions regarding the relationship between a neighborhood's ra cial characteristics and property values influenced the homeowner's expectations of the future and, with im perfect knowledge of other's plans, resulted in a choice to sell. Such analyses contributed to an understanding of the self fulfilling prophecy involved with such panic over racial change. Over time, more serious consideration would be given to the social construction of race and the construction of racial identity with consequences affecting households' choices and constraints. Kay Anderson's seminal study of Vancouver's Chinatown's 100 year history, for example, explores the extent to which segregated urban districts must be evaluated by considering the impact of racial ideology and insti tutional practices.

Whereas neighborhood decline, segregation and housing abandonment proved important issues from the 1960s, rising property values in certain older, central city neighborhoods drew increasing research attention from the 1970s onward. Over the last several decades, the examination of the processes involved in gentrification further demonstrates the extent to which neighborhood change is influenced by a complex set of factors. Gen trification, a term coined as early as 1964 to address the class transformation of older central city neighborhoods, represents an important aspect of contemporary urban spatial restructuring. Neil Smith introduced the concept of a rent gap as a means of addressing why some neighborhoods might be profitable to redevelop while others are not. The rent gap refers to the difference between actual ground rent obtained from existing land uses and the potential rent that could be obtained if the centrally located property were redeveloped for higher income residential use. If the rent gap serves as the underlying economic driver for gentrification, others emphasize the emerging demands on the part of a new middle class and changing cultural values. David Ley, for example, examined the role of gentrifiers in the process, questioning the extent to which changing cultural values and attitudes might draw people to older, working class neighborhoods. He argued that gentrification was a spa tial expression of new cultural values and examined the particular sense of place associated with such urban neighborhoods. Contemporary research tends to benefit from these two aspects of the early gentrification litera ture. Thus, the supply side explanation – focusing on the roles of developers and mortgage lenders as they seek profitable investments and the support provided by government agencies seeking an 'urban renaissance' – combines with the demand side understanding of new lifestyle choices and cultures of consumption among members of a new middle class. An ongoing research agenda emphasizes the importance of a 'geography of gentrification' that maintains a sensitivity to the specific contexts and scales of gentrification rather than trying to reduce it to a single, generalizable process.

The contrast between the earliest work on neighbor hood change, as exemplified by Burgess and Hoyt, and current work on gentrification and urban form put to rest any notion of universal, persistent trends in the housing market. Instead, neighborhoods are influenced by un certain metropolitan wide factors related to the regional economy, technological innovation, population and im migration trends, state and federal government policy, and emerging social and cultural values.

Neighborhood Lifecycles

In the face of such complexity, research on neighborhood change might begin with attention to three components that influence the individual household's choices and, ultimately, through the aggregate effect of neighbors' choices and neighborhood lifecycles. These include the aging of the physical environment, the aging of the residential population, and the consumer values associated with housing submarkets. In monetary terms, the quality of the physical environment involves the initial investment made and its ongoing maintenance. Common sense suggests that the quality of the initial construction would influence the probability of its continued maintenance at least in the near term. Until relatively recently, assumptions about aging neighborhoods tended to emphasize physical decline accompanied by declining real estate values resulting in out migration from the central city. Smith's work on the rent gap as it relates to gentrification demonstrates, however, that investors and developers evaluate disinvestment in the built environment and opportunities for redevelopment of central city real estate based on changing trends in the central city housing market.

Related to such changing trends, the potential for obsolescence of any particular unit is also a matter of its functional and stylistic character. Housing units may be described as functionally obsolete based on the occupant's demand for new amenities or changing technologies. Changing family structure, that is the size of the family and age of its members, also influences the household's assessment of the structure as functionally viable. Stylistic obsolescence acknowledges the fads associated with desireable architectural styles and the potential that changing tastes may trigger the desire for new housing, as highlighted by Homer Hoyt's research on upper income households. Hoyt's term filtering introduces a means of linking the mobility of individual households to neighborhood change with its emphasis on the movement of wealthier households up the housing scale (into new construction) and the consequent filtering of (their) houses down the social scale. The movement of households creates a vacancy chain with each move associated with an opportunity for those with incomes a rung below the previous owner. The ultimate consequence for aging structures in the vacancy chain would be a lack of value to anyone and abandonment. Financial obsolescence is another form that influences individual household behavior. For instance, a rise in family income or the effect of inflation reduces the relative size of mortgage repayments, thus encouraging the desire for a move into a larger and/or more expensive unit. In the language of the real estate market, a 'starter house' suggests a combination of functional and financial obsolescence that would trigger such a move.

Just as the aging of the built environment suggests conditions for neighborhood change, the aging of a neighborhood's residents also introduces conditions for neighborhood change. This is based on assumptions about the behavior of residents, particularly homeowners, over time. Mobility is associated with certain periods of an individual's life. Traditional habits associated with the nuclear family would suggest that younger, single individuals and young families would move more often. Influencing the decision to move includes a household's (perceived) need for space, desire to own rather than rent, and the (perceived) need for different environmental amenities for families. On the "contrary", families with children will be more sedentary, and particularly homeowners will stay in their houses until age affects their ability to maintain their own household. Because neighborhoods tend to be characterized as appropriate to different stages of the family life course (or family lifecycle), there are examples of housing areas with units built about the same time where residents 'gray' together. As a consequence, the aging of area residents influences opportunities for neighborhood change when they depart.

Arguably, as demonstrated by researchers such as Patricia Gober, changes in household structure related to the family life cycle are less dominant today than they were in the post World War II city. For the baby boomers and suceeding generations, lifestyle concerns rather than family life cycle offered greater explanatory value. Producers and consumers respond to new opportunities and changing lifestyles. Three categories characterize differences in consumption that include family oriented people, careerists, and consumerists. Family oriented people conform to behavior associated with the previously dominant concept of the family life course. Careerists tend to be highly mobile and status conscious with a focus on prestige neighborhoods. Consumerists may not place the same priority on professional requirements but they do prioritize the material benefits of modern society in their housing choices and seek the amenities of more fashionable neighborhoods. These categories reflect a decidedly middle class orientation in a discussion of housing markets and neighborhood life cycle. Clearly, lower income households are primarily influenced by income rather than defined by lifestyle choice.

Urban Segregation

The differentiation of residential areas might be interpreted as one primarily based on income – a form of economic segregation so naturalized in capitalist economies that it is rarely discussed in terms of segregation. Instead, discussion of segregation emphasizes the categories of difference associated with race and ethnicity. Although the Chicago School's optimistic assumptions regarding assimilation of different groups and its consensualist view of society dominated social geography for a long period, greater attention was given to persistent segregation and racism over time. To describe instances of voluntary and enforced segregation based on assumptions of either choice or constraint, scholars refined their vocabulary to distinguish among: (1) colonies, a temporary stage in the assimilation of groups into the wider sociospatial fabric of a city; (2) enclaves, more permanent ethnic settlements based on the internal cohesion of their members; and (3) ghettos, settlements that form in response to the negative conditions and external forces facing a particular ethnic or racial group. The distinction between these categories – particularly between enclaves and colonies – as has been raised in the work of Ceri Peach regarding South Asian Muslims in British cities influences interpretations of social fragmentation. In the United States, attention to the expansion and contraction of poverty neighborhoods – called impacted ghettos – considers particularly the contributions of sociologists Manuel Castell, whose work evaluates the economic and social exclusion of marginal groups, and William J. Wilson, who has advanced the concept of an underclass. Contemporary research acknowledges the complex of factors influencing changing conditions within segregated neighborhoods, including attention to economic opportunity, social and cultural constraints, and institutional influences on government and business practices.

Housing Submarkets

Housing, as both shelter and investment, is a complex package. The discussion above focused on individual units and consumers. The stability of any particular neighborhood, however, depends upon the dynamics of the local real estate market, that is, the demand for housing of particular types and price, the institutions involved in property transactions, and the people served. The role of institutional barriers and social gatekeepers in shaping spatial submarkets requires attention to the role of developers, mortgage finance agencies, realtors, and government. In part, housing submarkets result because developers build homes with similar physical chacteristics adjacent to one another based on the construction benefits of economies of mass production. Local land use regulations tend to further reinforce the class homogeneity of submarkets as a means of stabilizing investments for financial institutions and property owners. Furthermore, spatially based attributes contribute to housing quality and, as a consequence, units in close proximity share certain qualities that define property value.

Class and race were key factors in defining housing markets in the twentieth century city, and thus served as important variables in discussion of neighborhood change. Today, the proliferation of lifestyle choices further contributes to the transformation of the residential landscape. Paul Knox provides one illustration of this in his case study of Washington, DC. Drawing together a sociological literature on the culture of consumption – particularly influenced by the sociologist Pierre Bourdieu and his concept of habitus, Knox provides an overview of the development industry's response to the demands of the new bourgeoisie. Both sociologists Sharon Zukin's and Rosalyn Deutsch's research also have proven influential in their respective work on postmodernism's 'society of the spectacle' and the associated impact on defining fashionable neighborhoods with consequential change. The niche marketing of urban landscapes reflects the creation of products in response to the fragmentation of classes and thus influences the fortunes of new and existing neighborhoods in a metropolitan housing market.

Neighborhood Organizing

Although the literature on neighborhood change focuses primarily on residential mobility and competing submarkets, certain neighborhoods maintain a relative stability despite pressures to do otherwise. The ability of neighborhoods to organize in the face of unwanted change has also received attention in urban geographic research. The conceptual framework for evaluating locational conflict and its influence at the individual and neighborhood level identifies three potential options: voice, loyalty, and exit. In those instances that residents actively respond to a particular change, they exercise voice whether that is through formal, legal channels or through covert, illegal resistance. When resignation to a particular change is involved, traditionally the misnomer loyalty has been applied. This resignation suggests a perceived inability to respond and, consequently, suggests negative consequences for the individual and potential change in the neighborhood. Those who are able to exercise other choices might select the exit option. The combined effect of many individuals pursuing this option also results in neighborhood change. Recent studies on the effectiveness of local organizing to direct local land use development, such as the work of Mark Purcell and Deborah Martin, examines neighborhood activism as an expression of a group's vision for their neighborhood. The activists invest their social values in a spatial vision for the neighborhood. Whether their voice meets with success depends on the group's relative power in a land use competition with an outcome that is likely influenced by various scales of negotiation and accountability.

See also: Chicago School; Gentrification; Social Justice, Urban; Urbanization.

Further Reading

Anderson, K. (1991). *Vancouver's Chinatown: Racial Discourse in Canada, 1875 1980*. Montreal, QC: McGill University Press.
Berry, B. and Rees, P. (1969). Factorial ecology of Calcutta. *American Journal of Sociology* 74, 445 491.
Burgess, E. (1925). The growth of the city: An introduction to a research project. In Park, R. E., Burgess, E. W. & Mackensie, R. D. (eds.) *The City*, pp 47 62. Chicago, IL: University of Chicago Press.
Gober, P., McHugh, K. and Reid, N. (1991). Phoenix in flux: Household instability, residential mobility, and neighborhood change. *Annals of the Association of American Geographers* 81, 80 88.
Harris, C. and Ullman, E. (1945). The nature of cities. *Annals of the American Academy of Political and Social Science* 242, 7 17.
Hoyt, H. (1939). *The Structure and Growth of Residential Neighborhoods in American Cities*. Washington, DC: United States Government Printing Office.
Knox, P. (1991). The restless urban landscape: Economic and sociocultural change and the transformation of metropolitan Washington, DC. *Annals of American Geographers* 81, 181 209.
Lees, L., Slater, T. and Wyly, E. (2007). *Gentrification*. New York: Routledge.
Ley, D. (1996). *The New Middle Class and the Remaking of the Central City*. Oxford: Oxford University Press.
Martin, D. (2004). Reconstructing urban politics: Neighborhood activism in land use change. *Urban Affairs Review* 39(5), 589 612.
Peach, C. (2006). Islam, ethnicity and South Asian religions in the London 2001 census. *Transactions of the Institute of British Geographers NS* 31(3), 353 370.
Purcell, M. (2001). Neighborhood activism among homeowners as a politics of space. *Professional Geographer* 53(2), 178 194.
Shevsky, E. and Bell, W. (1955). *Social Area Analysis: Theory, Illustrative Application, and Computational Procedures*. Stanford, CA: Stanford University Press.
Smith, N. (1996). *The New Urban Frontier: Gentrification and the Revanchist City*. London: Routledge.

Relevant Websites

http://www2.urban.org/nnip/ncua/index.htm
Neighborhood Change in Urban America.

Neighborhood Effects

Sang-Il Lee, Seoul National University, Seoul, South Korea

Glossary

Buffer Zones Areas of user-defined width surrounding geographical features.

Dummy Variable A numerical variable used in regression analysis to represent categories.

GeoCoding The process of positing a geographical feature on a map as a point by using a code representing its location.

Logistic Regression A regression analysis used for categorical dependent variables.

Measurement Scale The physical size of the spatial unit in a study, sometimes called spatial resolution.

Misspecification Error A modeling error in a regression framework committed mainly by omitting relevant variables.

Operational Scale The spatial extent to which a spatial process operates.

Spatial Dependence A particular relationship between geographical similarity of locations and numerical similarity of values observed on them.

Spatial Heterogeneity A lack of spatial uniformity in mathematical and statistical properties.

Introduction

Neighborhood effects refer to socio geographical milieu's influences on the ways in which people think and act. In other words, the neighborhood effect is a structural factor which may drive individuals in a place to possess a collective behavioral disposition leading to similar behavioral outcomes or choices. In social sciences, neighborhood effects have been regarded as a synonym or a particular type of contextual effects. The concept of neighborhood effects has provided a conceptual foundation for those who seek to emphasize the importance of socio geographical contexts in explaining various types of behavioral outcomes such as educational achievements, adolescent misconducts, crimes, and so on. The concept has also been utilized in housing studies, epidemiology, psychology, and so on.

It is, however, in the field of political geography that the most attention has been given to the concept. Political geographers have used the concept in order to explain why different people in the same place vote in similar ways while similar people in different places vote in different ways. It is generally accepted in particular by political scientists that individuals' voting decisions are largely determined by who they are. However, if the voting decisions are also associated with where they live or have been socialized even after the effects of who they are were controlled, one might argue that there exist geographical contexts' influences on the political behaviors and the influences might be conceptualized as neighborhood effects. On the contrary, if who they are explains the voting patterns or if the explanatory power of where they live is not sufficiently large in comparison with that of who they are, one might argue that neighborhood effects do not exist.

Subsequently, this article briefly discusses some theoretical underpinnings which the concept may rely on, and then moves to some methodological issues as to how to formulate and model neighborhood effects. All the discussions revolve around voting behaviors.

Theoretical Underpinnings

How could socio geographical contexts possibly drive individuals in a place to possess a collective behavioral disposition? It has been known that the conceptual origins of neighborhood effects lie in the work of a Swede and an American political scientist, H. Tingsten and V. O. Key, in 1930s to 1940s. It is by Kevin Cox, David Reynolds, and Butler and Stokes in the late 1960s that the concept emerged as a keystone for electoral studies.

Kevin Cox, in his seminal 1969 paper, identified four factors (or biases), on the basis of a theory of communication and information flow, which influence people within a socio spatially structured information network. The geographical distance bias refers to the fact that people living nearby are more likely to interact than those apart. The acquaintance circle bias facilitates more frequent and intensive interaction and thus tends to lead to more opinion agreement between members of the same acquaintance circle. This bias is closely related to the first one; people who live close together are more likely to be members in the same acquaintance circle than those who live further apart. In practice, the information flows within a neighborhood are often diffusive, influential, and persuasive such that the situations may heighten the possibility that some having minority opinion are converted to the dominant majority opinion. The forced field bias indicates some situations whereby an established political culture in a local area appears to be quite coercive enough to mobilize opinions in a particular way. This is most clearly observed where place based political parties are pervasive. The reciprocity bias refers to the fact that the chance of information being passed between two individuals is greater where there already is a tie between them. For example, the relationships between parents and children, siblings, husbands and wives, friends, and colleagues provide a greater channel for information exchange. All these factors can be articulated in spatial terms, and collectively constitute the fundamental rationale for the concept of neighborhood effects.

In sociology and other social sciences, there are several theories about how neighborhood effects occur, some of which are congruent with Cox's notions, but others not. Contagion theory underlines the power of peer group influences to spread particular types of behaviors. Collective socialization theory emphasizes how neighborhoods provide role models and sources of social control. Institutional theory stresses the roles of various institutions located in local areas such as schools, businesses, political organizations, social service agencies, and so on. Competition theory and relative deprivation theory are less relevant here. On the basis of all these theoretical contemplations, five processes by which neighborhood effects are produced can be listed as follows (after Johnson *et al.*, 2005):

1. Local social interaction: the classic neighborhood effect. This involves a process of 'conversion through conversation' or 'those who talk together vote together'; "I talk with them and vote as they do."
2. Environmental selection: according to this process people choose to live (to the extent that they can) among people they wish to associate with; "I want to be like them so I live with them."
3. Emulation: in this process, people choose to behave like their neighbors, even without interacting with them, on the basis of observed (or inferred) behavioral patterns; "I live among them and want to be like them."
4. Environmental observation: people see and hear about issues in their local area and vote with their neighbors accordingly, in order to promote local interest; "what I observe around me makes me vote with them."
5. Local pressure: political parties actively seek support through canvassing and local campaigns, and people may therefore be influenced by the intensity of that local pressure to vote in a particular way; "they want me to vote for them here."

Debates

There have been two large scale debates on whether neighborhood effects or contextual effects really matter, first in *Political Geography Quarterly* and second in *Political Geography*. Those debates were seen as conflicts between two academic fields – geography and political science. In the 1987 debate, political scientists reported that if individual factors were controlled, contextual effects would vanish. In contrast, political geographers demonstrated that substantial contextual effects remained when individual factors were held constant. The former argues that neighborhood effects are mere myth or embellishment, while the latter contends that the alleged pure individuals have already been saturated with contextual effects.

In the 1996 debate, Gary King, a political scientist, argued "The *geographical variation* is usually quite large to begin with, but after we control for what we have learned about voters, there isn't much left for *contextual effects*." (159–164) and "Geographical variation yes, contextual effects no." (159–164). John Agnew (1996: 129–146), a political geographer, contended "The concept of geographical context can be used to draw attention to the spatial situatedness of human action in contrast to the non spatial sorting of people out into categories based on census and other classification schemes that inspires most conventional social science." He further argued, "[T]he hierarchical geographical context or place channels the flow of interests, influence and identity out of which political activities emanate." and thus "political behavior is inevitably structured by a changing configuration of social geographical influences…" (Agnew, 1996: 129–146).

It is worthwhile to indicate here that the boundary between individual effects and contextual effects is often blurred; what are usually regarded as individual variables are indeed dependent upon contextual variables in many cases. For example, an individual's union membership may be significantly affected by local norms and traditions; individual income and education levels are often dependent upon the quality of schools and jobs in the local neighborhood; an individual's status in the occupational ladder is likely to be constrained and shaped by the local labor market economy.

In a methodological sense, the essence of the arguments raised by those who are against the geographical/contextual effects is that empirical models favoring the effects are committing a kind of misspecification error; that is, contextual variables turn out to be significant simply because important individual variables are omitted. Thus, what is at issue in the debates is methodological, rather than theoretical.

Methodologies

The methodologies for formulating and estimating neighborhood effects may be classified into two categories: area based and individual based. One could think of a typical research framework from each of the categories. A study with aggregated data may involve a regression analysis in order to explain spatial variation of voting results with various independent variables which are usually derived from census data. Although this kind of analysis could provide a formula to predict the voting outcomes at a locality on the basis of the locality's characteristics, it gives little insight into neighborhood effects. As a matter of fact, neighborhood effects reside intact in a set of residuals. On the other hand, a study with individual data may also involve a regression model, usually logistic, by positing an individual's choice in a binary format on the left hand side and their attributes on the right hand side which usually include:

1. socioeconomic status (occupational class, educational attainment, housing tenure, employment status, in come, etc.);
2. ascribed characteristics (sex, age, race, ethnicity, etc.); and
3. family background (parents' political propensity and socioeconomic status).

This approach, however, completely ignores the neigh borhood effects and has been overwhelmingly adopted in political scientists' literature.

Methods for Aggregate Data

A study based on aggregate data examines the neigh borhood effects in reliance on an area's collective char acteristics which are expected to explain collective outcomes under investigation. Prior to a discussion on methodological principles utilizing area data, a dis tinction between compositional variables and contextual variables needs to be made. In general, a compositional variable refers to an individual's socioeconomic charac teristic whereas a contextual variable relates to location which provides a milieu within which an individual lives and is socialized. When aggregate data are considered, this distinction needs to be furthered. In a conceptual sense, compositional variables are associated with who people in a place are and what they have, whereas con textual variables describe environmental properties, physical and/or human, that a place possesses as a macro scale entity. For example, geographical variations in death rates are not only due to different age structures of areas (the compositional effect), but also due to some detrimental attributes of the areas that may have a direct or indirect effect on people's health (the contextual ef fect). It should be noted, however, that the two types of variables are dependent upon each other such that it is often impossible to separate one from the other.

One way of tackling neighborhood effects with ag gregate data is to regress socioeconomic variables (usually compositional variables) of local areas plus 'locational variables' on voting outcomes. The locational variables are mainly associated with where an area stands as a spatial object in relation to other areas or specific features. For example, they could be each area's distance from a focal point, for example, central business district (CBD), or dummy codes indicating membership in buffer zones or to particular geographical categories based on either ad ministrative regionalization or functional classification, or an areal unit's identity itself. If the locational variables have a statistically significant explanatory power with other socioeconomic variables being held constant, the existence of neighborhood effects would be confirmed. For example, it has been acknowledged that suburban areas are more pro conservative than their inner city counterparts. One can follow this analytical line by in corporating a series of higher level regional dummies along with socioeconomic characteristics of basic spatial units. A significant level of heterogeneity in the regression coefficients of the dummy variables could evidence, at least partially, the point that geographical contexts count.

A more intelligent way of tackling neighborhood ef fects in the regression framework using aggregate data might be to attempt to define a set of contextual variables rather than the simple, locational variables, and then to put them on the right hand side along with a set of compositional variables. Contextual variables may in clude economic experiences, socio cultural practices, capital–labor relationships, and the level of social capital of a place. If the set of contextual variables has sufficient explanatory power in comparison with the set of com positional variables, one can say that there exists a neighborhood effect. One problem in implementing this framework is that it is often difficult to measure genuine contextual variables and to separate them from com positional variables. A carefully designed set of con textual variables could identify neighborhood effects which would otherwise be hidden in the residuals.

There are other types of area based analyses utilizing aggregate datasets, some variables of which are derived from individual level survey data. We know the pro portion of votes favoring a particular political party from aggregate data, but do not know what percentage of people belonging to a particular social class actually vote for the party. By possessing that kind of information, one may be able to devise a way of proving the existence of neighborhood effects. It has been found that the more dominant the class is in an area, the greater the pro portion of its members who vote along the expected class lines. This echoes the finding that the more residentially segregated cities would show greater polarization be tween social classes.

In comparison with individual based research, area based studies provide indirect or 'circumstantial' evi dences for neighborhood effects, basically because they are never completely free from the problem of ecological fallacy, though there have been some attempts to bridge the aggregate–individual gap. They also are subject to modifiable areal unit problem (MAUP); there is no such thing as a set of 'natural' spatial units; therefore all the statistical results are dependent upon the scale and configuration of spatial units engaged.

Methods for Individual Data

Purely individual based research can be seen as a simple extension of the primitive approach; the dataset is com posed of individuals' characteristics plus their locational properties. The locational properties simply refer to which regional unit the individuals' residential locations belong to; the regionalization scheme could vary ac cording to what spatial scale is considered. In most cases, a logistic regression model for a binary dependent vari able is utilized to see if the locational variables really count when the individual characteristic variables are held constant. If the individual variables as well as the locational variables are all significantly related to elect oral choice, then one could conclude that contextual effects on voting are important.

Individual based methods, in many cases, combine individual level survey data and aggregate census data. Rather focusing on where people live, the hybrid ap proach pays more attention to the characteristics that an area possesses. Those areal characteristics are captured by compositional variables (rarely by contextual ones) which usually include ones related to socioeconomic perform ance, class or occupational structure, and demographic structure. The variables are first obtained from census in an aggregated form at a particular spatial scale; and the variables are then coded onto the individual level data according to individuals' residential location, so that each individual is allocated a separate set of areal attributes. Once all the variables are prepared, a logistic model is fitted.

One of the most advanced methodologies to imple ment the hybrid model is multilevel modeling. It is contended that multilevel models operate at more than one scale, so that a single model can handle the micro scale of people and the macro scale of places simul taneously. Multilevel modeling is expected to resolve the analytical dilemma not only between individual scale and aggregate scale but among different aggregate scales. Here it may be sufficient to provide how a multilevel model is specified for an empirical study. One might take a three level model where level one is the individual level, level two the local level, and level three the re gional level. Independent variables at the level one may include some individual variables such as age, occu pational class, housing tenure, employment status, and so on, and compositional (and/or contextual) variables at the level two may include unemployment rate, un employment rate change, and proportion of workers in a particular industry. By conducting a multilevel modeling, one may find that the relationships of the dependent variable not only with each of the independent variables at the level one but with each of ones in the level two vary region to region (the level three). Empirical studies utilizing the multilevel modeling report that typically 10–20% of the variance in the response is attributed to contextual effects.

Some Lessons from Spatial Data Analysis

Recent advances in spatial data analysis and geographical information have increasingly provided Science (GISc) new insights into neighborhood effect research, though their applicability still remains limited. One of the most important sources of the limitation is that there are some conceptual and methodological difficulties in applying spatial data analysis techniques to neighborhood effect modeling. In spatial data analysis terms, the neighbor hood effect is associated not with first order effects but with second order effects. In general, first order effects relate to a global or deterministic or large scale trend, while second order effects refer to local or stochastic or small scale variations, resulting from the spatial de pendence in the process, that is, the tendency of neigh boring locations to be correlated with each other in terms of the deviations from the first order effects. In short, the first order effects is associated more with compositional effects while the second order with contextual effect.

In the individual based approach, some second order point pattern analysis tools could provide a promising room for application. When voters' locations are geo coded, Ripley's K function can gauge the degree of the spatial clustering of voting patterns. A bivariate K function, however, could provide more insights into the nature of neighborhood effects. For example, if the geo coded voters' locations are given a categorical value, one for supporters of a particular party, and two for supporters of the other party, then we have two sets of point patterns. By applying a bivariate K function, one could gain an understating of whether they are spatially independent or not. In a regression setting, aggregate or individual, spatial linear regression models such as various types of spatial autoregressive models, could edify the problem of spatial autocorrelation in residuals endemic to the ordinary least square (OLS) regression framework which has over whelmingly been adopted in the neighborhood effect studies.

In the cases of aggregate data, various techniques of exploratory spatial data analysis (ESDA), particularly utilizing local measures of spatial association such as local Moran's I_i and Getis Ord's G_i^*, could broaden the analytical spectrum allowing for exploring spatial dependence and spatial heterogeneity which voting patterns usually display. Such ESDA techniques, however, may be more useful when a particular assumption on the spatial scale at which neighborhood effects operate is satisfied: the operational scale should be larger than the measurement scale, that is, the (average) size of spatial units. Only when the assumption turns out to be con ceptually and practically sustainable, (positive) spatial dependence observed in voting patterns can be regarded as a circumstantial hint of the existence of 'inter local' neighborhood effects as opposed to 'intra local'. How ever, it should be noted that many of ESDA techniques do not allow for heterogeneity in the first order effect, that is, constant compositional effect over space, which wouldn't be tenable in most cases under investigation. Thus, genuine (inter local) contextual effects can be tackled only when the first order, compositional parts are removed from the original variance of voting patterns.

Geographically weighted regression (GWR) can be seen as a method to deal with intralocal and interlocal neighborhood effects simultaneously. One of the most important rationales of GWR is that local areas possess their own peculiarities which are irreducible to the global trend. Thus, GWR allows for spatial heterogeneity of causal relationships between variables, that is, spatially drifting regression parameters. Meanwhile, GWR fulfills the task by having values at original locations referencing ones at their adjacent locations in the estimation process. In the context of neighborhood effects, GWR can show how spatially different the relationship of the dependent variable with each of not only compositional variables but contextual variables is (intra local neighborhood effects) and how similar each of the relationships between ad jacent local areas (inter local neighborhood effects) is.

Some critical issues for future research need to be addressed. The first issue revolves around how to define a neighborhood: Is there any natural foundation on which neighborhoods are delineated? How can we be sure that our data units are congruent with the concept of neighborhood? What is the spatial extent of a typical neighborhood? Some GIS based approaches such as an automated neighborhood identification procedures open up a promising space for this issue. Second, there is an other scale issue, that is, the multi scale nature of neighborhood effects. To gain understanding of the whole picture of neighborhood effects, it may be desir able to investigate the hierarchical and/or interlocking nature of the effects occurring at different spatial scales. Recent advances in multilevel modeling could provide a viable starting point to address this issue. The third issue is about how to specify neighborhood characteristics, either compositional or contextual. In order to ad equately model neighborhood effects, it is crucial to possess explanatory variables which properly measure socioeconomic and environmental conditions of a place. Some efforts have recently been made to devise com posite measures of neighborhood characteristics such as measures of areal deprivation by exploiting a variety of data sources in a GIS environment.

See also: Cox, K.; Ecological Fallacy; Electoral Geography; Exploratory Spatial Data Analysis; First Law of Geography; Geographically Weighted Regression; Georeferencing, Geocoding; GIScience and Systems; Modifiable Areal Unit Problem; Point Pattern Analysis; Regionalisations, Everyday; Regression, Linear and Nonlinear; Spatial Autocorrelation; Spatial Clustering, Detection and Analysis of; Spatially Autoregressive Models; Statistics, Overview; Statistics, Spatial.

Further Reading

Agnew, J. A. (1996). Mapping politics: How context counts in electoral geography. *Political Geography* 15, 129 146.

Anselin, L. and Cho, W. K. T. (2002). Spatial effects and ecological inference. *Political Analysis* 10, 276 297.

Bailey, T. and Gatrell, A. C. (1995). *Interactive Spatial Data Analysis*. Harlow: Longman.

Cox, K. R. (1969). The voting decision in a spatial context. *Progress in Geography* 1, 81 117.

Dietz, R. D. (2002). The estimation of neighborhood effects in the social sciences: An interdisciplinary approach. *Social Science Research* 31, 539 575.

Eagles, M. (ed.) (1995). *Spatial and Contextual Models in Political Research*. London: Taylor and Francis.

Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University Press.

Johnston, R. J. and Pattie, C. J. (2006). *Putting Voters in Their Place: Geography and Elections in Great Britain*. New York: Oxford University Press.

Johnston, R. J., Propper, C., Sarker, R. *et al*. (2005). Neighbourhood social capital and neighbourhood effects. *Environment and Planning A* 37, 1443 1459.

Jones, K. and Duncan, C. (1996). People and places: The multilevel model as a general framework for the quantitative analysis of geographical data. In Longley, P. & Batty, M. (eds.) *Spatial Analysis: Modelling in a GIS Environment*, pp 79 104. Cambridge: GeoInformation International.

King, G. (1996). Why context should not count. *Political Geography* 15, 159 164.

McAllister, I. (1987). Social context, turnout, and the vote: Australian and British comparisons. *Political Geography Quarterly* 6, 17 30.

O'Loughlin, J. (2003). Spatial analysis in political geography. In Agnew, J., Mitchell, K. & Toal, G. (eds.) *A Companion to Political Geography*, pp 30 46. Oxford: Blackwell Publishing.

Taylor, P. J. and Johnston, R. J. (1979). *Geography of Elections*. London: Penguin.

Neighborhoods and Community

J. Flint, Sheffield Hallam University, Sheffield, UK

Glossary

Broken Windows A theory developed within criminology that environmental incivilities will result in further crimes and the decline of a neighborhood.

Collective Efficacy The combined capacity of groups of individuals to influence their local environment or social circumstances.

Communitarianism A form of social theory and political organization based on membership of, and engagement within, a community.

Ecological Fallacy The fallacy of making inferences about individuals from aggregated data.

Residualization The increasing concentration of deprived and vulnerable populations within a neighborhood.

Social Capital Forms of collective networks and reciprocal relationships that bring social benefits to a group of individuals.

Spatial Mismatch Spatial barriers of access, such as poor transport connections that prevent individuals from taking up certain employment opportunities.

Tipping Point A theory that once a particular population forms a certain proportion of residents in a neighborhood, rapid and significant change will then occur in the neighborhood.

Underclass Theory A theory that certain deprived populations are becoming increasingly socially and culturally isolated from 'mainstream' society.

Introduction

Neighborhood and community are two of the most frequently used terms in urban geography. They are multidimensional forms of human organization with no universally agreed definition to describe them. Neighborhoods and community are synonymous with urban settlements. Although neighborhood and community are not interchangeable, and indeed one may exist without the other, it is the sociospatial interaction of neighborhood and community that has been the focus of urban geography, dating back to Mayhew's study of London's rookeries in the nineteenth century and the Chicago School's exploration of neighborhood formation in the 1920s. It is primarily through neighborhoods that urban spaces become identified as urban places, with their own histories, trajectories, and meanings for residents.

If the city is conceptualized as a form of community or communities existing within urban space, then neighborhoods are the smallest territorial scale at which the communal processes of urban life are usually organized and studied in urban geography. In this conceptualization, neighborhoods may be regarded as the blocks upon which the economic, social, political, and cultural elements of the city are built. Neighborhoods and community have become increasingly prominent in both urban geography and urban policy, based on an understanding that neighborhoods serve crucial functions within urban systems and are in turn affected by wider changes in urban and global forces. Contemporary studies within urban geography suggest that neighborhoods and communities, and our understanding of what these mean, are being transformed at an unprecedented scale and pace.

Defining Neighborhoods and Community

Neighborhoods comprise the smallest unit of urban social territory and political organization. There is no universally accepted definition within urban geography of what constitutes an urban neighborhood. The smallest unit of a neighborhood is typically defined as an area of up to 10 min walk from one's home. A neighborhood may be enclosed by physical boundaries, including railway lines, major roads, and fences. It may also comprise residential properties of a similar built form or housing tenure. Residents of a neighborhood may share similar characteristics, including ethnicity, occupation, and income, which come to define the identity of the neighborhood. Neighborhoods are also demarcated by a range of agencies and organizations for the purposes of urban administration, service delivery, and governance. Within neighborhoods, there are often subterritories of particular streets, and neighborhoods are also situated within wider localities, often referred to as wards or districts, which in turn are located within larger settlements and regions. Neighborhoods are multidimensional forms of urban organization, comprising a series of spatially based attributes and processes relating to the built environment, the economic, social, and cultural characteristics of the population residing within them, the interactions between residents and the forms of employment, services, and facilities located within neighborhoods, and how these are utilized.

Community is an equally complex and contested term. Community relates to a form of social organization based upon some commonality between individuals, which re sults in them being defined as members of such a com munity and simultaneously demarcates others who are not members of the community. Membership of a community is differentiated from membership of a family grouping or legally defined citizenship of a nation state and is therefore regarded as a separate form of social identity with related social processes. Most individuals are members of diverse and overlapping communities. Communities need not be primarily spatially based. They can relate to religious af filiation (as in the global Umma of Islam), political beliefs, sexual orientation, occupation, and cultural and leisure interests. The growth of global communication technology has been particularly important in establishing 'online' communities which reduce the centrality of space to community identity. However, many communities are linked to defined spaces and places at a range of spatial scales, including the international (e.g., the European Union) and nation states. The interaction of place and community at the local level has received particular at tention among urban geographers and urban governments, leading often to a conflation of neighborhood and com munity and a focus upon the combination of social and spatial dynamics within neighborhoods. The extent to which individuals identify with communities and neigh borhoods varies, as does the importance of neighborhoods and communities for individuals' lives. It is therefore im portant to avoid the ecological fallacy of equating an in dividual's residence with his/her personal attributes or behaviors.

The Functions of Neighborhoods and Communities

Neighborhoods serve a number of functions. Their pre dominant function is as a site of residence and home lo cation for individuals and families. This is often, but not necessarily, linked to individual psychosocial processes of belonging, identity, security, and familiarity. The emer gence of neighborhoods in the historical growth of urban areas was often linked to their role as the sites of initial residence for new populations. The migration of rural populations to cities due to industrialization characterized the development of urban neighborhoods in the developed world in the nineteenth and twentieth century and char acterizes the rapid urban growth of developing nations into the twenty first century. The close proximity of home to work was a feature of the development of human settle ments which continued through the Industrial Revolution. However, in the post World War II era, advances in transport and communication technologies and changing economic activities reduced the link between work and residential neighborhood, epitomized in the growth of commuting. This was often combined by deliberate urban planning policies that sought to separate residential neighborhoods from industrial and commercial zones in cities. However, further evolution of the economy and in formation technology and concerns about environmental sustainability may result in a renewed proximity between neighborhood and place of employment.

Neighborhoods also serve a primary function in housing markets and housing policies. They serve as a sorting mechanism that links households to employment opportunities, educational facilities, available health ser vices, or ethnic and cultural support mechanisms. The residential choice of neighborhood is heavily mediated by house prices and income, and the composition of most neighborhoods therefore reflects differential economic and social status. Urban geography research suggests that the neighborhood is equally, or even more, important than the characteristics of individual properties in driving resi dential location decisions. The provision of subsidized, affordable housing by public or quasi public agencies also results in the development of new neighborhoods, for example, the slum clearances and construction of large scale periphery housing estates, which were a dominant feature of urban planning in many cities during the twentieth century.

Neighborhoods also serve as sites of consumption and civic engagement. They provide some local retail and commercial services, although these have often reduced due to the growth of edge of town retail parks. Neigh borhoods may also provide specific facilities linked to ethnic, cultural, or religious affiliations. They provide sites of leisure and recreation, including public spaces, and are the locations where public services such as po licing, environmental cleansing, and education are de livered. In addition to organizational sites for public service delivery and management, neighborhoods also serve a civic and political function as the arena for en gagement between urban residents and governance structures and the lowest tier of political representation. The recent trend in many cities for decentralized and devolved urban administration has resulted in both neighborhood management and neighborhood forums for decision making gaining increasing prominence. The functions that a neighborhood serves vary between indi viduals and also change over the life course, for example, the attributes required from a neighborhood by a young, single person is likely to be different to those required by a household with children. The extent to which an indi vidual is reliant on the neighborhood to provide his/her employment, friendship networks, and access to services and facilities is also a function of economic status, so that a neighborhood of residence and the attributes contained within it will matter more to some individuals.

Neighborhoods are also arenas for socialization, in cluding family and friendship networks, and also provide sites of social solidarity and protection. The concept of neighboring refers to social interaction between non family members sharing a spatial proximity, while the concept of community is differentiated from the relationship between citizen and state. As such, neighboring and community are regarded as organic processes developing and occurring within neighborhoods. The acts of neighboring range from the exchange of pleasantries in the street to reciprocal favors, such as guarding properties or childcare. Community, strongly influenced by the concept of communitarianism, denotes both a membership of a social body and the social processes that arise from this membership. In addition to community serving as a site of identity and belonging within the arena of the neighborhood, the processes of community are envisaged as fostering social solidarity and providing the motivations and means for collective endeavor, including political engagement. This is based on a common attachment to a neighborhood (and its future) and a shared sense of purpose among its residents that overrides other social divisions and allegiances to other forms of community. However, the work of urban geographers, from Harvey Warren Zorbaugh in the 1920s and David Ley in the 1970s through to contemporary studies, has identified the complexities and disunities in everyday behavior and social interactions within neighborhoods. Despite this, the promotion of community, enacted spatially through neighborhoods, has been a prominent pillar of contemporary urban policy in many nations.

Neighborhood and Community Careers

Urban geographers have identified that neighborhoods and communities, like individuals, have careers. That is, neighborhoods evolve over time, both internally and in relation to other neighborhoods in an urban area. These careers can result in changes in the function of neighborhoods, their populations, and their built environment. The study of neighborhood careers has a history dating back to the Chicago School's identification of zones of transition through which neighborhoods became the site of new populations and economic activities and subsequently evolved to serve new functions in the urban system, such as the 'frontier outpost' role of Black inner city neighborhoods in the United States. A contemporary manifestation of neighborhood careers, which has been subject to considerable attention in urban geography, is the processes of gentrification occurring in many cities of the world, whereby previously working class and/or ethnic minority inner city neighborhoods are being transformed by new populations, businesses, services, and facilities. Gentrification and regeneration may represent the 'rise' of a neighborhood, associated with improvements in the physical environment, falling crime rates, higher educational attainment, rising house prices, improved public services, and an enhanced reputation of the neighborhood, and its inhabitants, in the wider urban area. Conversely, urban geographers have also studied neighborhoods that decline, leading to social disorganization, the abandonment and demolition of properties, rising crime levels, the withdrawal of commercial and public services, and a stigmatization of the neighborhood and its inhabitants. A contemporary example of this process is what is usually termed the residualization of public housing estates in many countries, resulting in a growing concentration of vulnerable and deprived residents. Specific, and contested, concepts such as broken windows and population tipping point theories have been developed to describe the micro processes within neighborhoods that lead to a spiral of economic, social, and physical decline.

The role of community in neighborhood careers is uncertain. A strong community, stable population, and long term residency are regarded as signifiers of successful neighborhoods, while rapid population turnover is often a signifier of a distressed neighborhood. However, collective coping mechanisms among residents are also a feature of many declining neighborhoods. Changes in neighborhoods are linked to wider economic and societal shifts. For example, the long term evolution of some neighborhoods in the northern United States was linked to economic developments in both the north and south of the country, the resultant migration of the Black population to inner urban areas and the subsequent 'white flight' to the suburbs and then further economic and social restructuring which increasingly disadvantaged inner city neighborhoods. The link between neighborhood and employment opportunities is also emphasized in spatial mismatch economic theories that suggest residents in some deprived neighborhoods are unable to access new employment opportunities in urban areas due to their physical isolation and the poor or expensive transport infrastructure serving these neighborhoods.

One of the challenges facing urban planners is the extent to which policy should focus on changing the fortunes of neighborhoods as static places or the mobile populations who reside within them, linked to the functions that particular neighborhoods serve. For example, some neighborhoods have retained long established ethnic populations and identities, including many Jewish communities. Other neighborhoods experience shifts in their population and identity, for example, the historic Irish and Italian, and more recently Korean and Hispanic neighborhoods of New York. This also applies to the economic, housing tenure, and life course function of neighborhoods. Neighborhoods dominated by public or private rented housing may remain problematic as sites of low income and social problems, but may be regarded as necessary transitory locations for individuals prior to them moving to other neighborhoods and housing tenures. The emphasis on spatially targeted neighbor hood regeneration and area based initiatives in urban policy also faces this dilemma, with the evidence sug gesting that many of the improvements resulting from interventions benefit individuals who then leave a de prived neighborhood to be replaced by new vulnerable or deprived households, resulting in little overall change in the circumstances of the neighborhood. Some urban re newal programs, such as the Hope VI and Moving to Opportunity programs in the United States, are explicitly based on relocating households to more affluent neigh borhoods and communities in order to further their life opportunities.

Neighborhood and Community Effects

The concept of neighborhood effects is influential in contemporary urban geography and urban policy. Neigh borhood effects describe the advantages or disadvantages that a place of residence has on the life circumstances and opportunities of an individual. Neighborhood effects act as a related but independent variable to individual charac teristics such as age, gender, ethnicity, and social class. Neighborhood effects have therefore been identified as explaining the differential trajectories of neighborhoods with initially similar socioeconomic and demographic population profiles.

The prominence given to neighborhood effects re flects both the cultural and spatial turns in human geography. Neighborhood and community and the social and cultural processes occurring within them have be come more important as the concept of poverty has broadened from material deprivation to a focus on a wider social exclusion that includes both individual and collective spatially based dimensions. The 1990s saw a reemergence of the 1960s 'rediscovery' of poverty and a focus within urban geography upon cultures of neigh borhood poverty. The school of underclass theorists claimed the existence of specific forms of collective at titudes and behaviors that resulted in the sociospatial distancing of deprived communities from the rest of their societies. While urban geographers disputed the extent to which such cultures were caused by internal neighbor hood processes or were the result of wider economic and social restructuring, underclass related theories have been influential within urban policy programs aimed at reversing neighborhood decline.

The growing concentration of poverty within par ticular urban neighborhoods and the concentration ef fects that are conceptualized as arising from this process have led many urban policy regimes to attempt to in crease the levels of social mix within neighborhoods, including providing a range of housing tenures, income levels, household types, and ethnic diversity. However, the international research evidence suggests that the re sultant spatial proximity does not subsequently lead to improved social interaction between social classes or fa cilitate a growing sense of community based around a collective neighborhood identity.

Social Capital and Collective Efficacy

Two concepts linked to neighborhoods and community have recently been influential in urban geography and urban policy making: social capital and collective efficacy. Social capital refers to the forms of trust, reciprocity, and collective capacity that exist between networks of indi viduals and neighborhood based organizations. As with other forms of capital, social capital is a resource which communities may draw upon to further opportunities and services within neighborhoods. Social capital may operate informally between neighbors in order to strengthen social processes and a sense of belonging or be used by local communities to influence economic and political decision making to achieve improvements within neighborhoods. Social capital has been promoted as a neighborhood and community based mechanism for complementing economic and social regeneration in deprived neighborhoods and for revitalizing civic en gagement. However, social capital is also theorized as a mechanism for ensuring the status of more affluent and elite neighborhoods. There is an important distinction between social capital that enables residents to cope in deprived neighborhoods, for example, exchanging favors or lending each other small amounts of money, and social capital that enables residents and neighborhoods to ad vance their circumstances by enhancing employment and educational opportunities or improving the quality of public services within neighborhoods.

Collective efficacy develops social capital theory by focusing on the outcomes, rather than the presence, of collective processes or shared norms and values within neighborhoods. Primarily developed in criminology, it has been applied within urban geography to studies of how communities use mechanisms of formal and informal social control to regulate behavior and the use of public space within neighborhoods. These studies have demon strated that levels of collective efficacy vary significantly between neighborhoods, and, as with the neighborhood effects literature, suggest that collective efficacy is not related straightforwardly to the income, ethnicity, or age of residents. The importance of social capital and col lective efficacy theories lies in their identification of the significance of the sociospatial interaction of communal processes within urban neighborhoods. While social cap ital and collective efficacy remain influential in urban policy and planning, they have been criticized by some urban geographers for failing to capture the complexities and divisions that exist within communities and that play out at the scale of the neighborhood.

Neighborhoods and Urban Segregation

Neighborhoods are a physical and spatial manifestation of the wider economic and social forces acting upon them. Although there are contemporary concerns that urban neighborhoods are becoming increasingly socially segregated along community fault lines, it is important to note that cities, including classical period Rome, have always had neighborhoods that have been socially and spatially segregated on the lines of social status, income, and ethnicity. Indeed, urban geographers have long identified that the delimitation of boundaries and territoriality are inherent features of the social construction of neighborhood. However, international urban geography research evidence does suggest that many neighborhoods are becoming more internally homogeneous, while concentrations of wealth and poverty are becoming more pronounced, thereby increasing the neighborhood effects on life chances. Two other contemporary developments in neighborhood formation have also received much academic attention in relation to urban segregation.

First, neighborhoods are becoming less physically as well as socially permeable. This is symbolized in the growth of gated communities which prevent nonresidents entering neighborhoods and which challenges the traditional urban concepts of public space, fluidity, and universally accessible, if differentiated neighborhoods. Gated developments also reflect the growth of community governance through which local neighborhoods seek to govern themselves, insulate their residents from risk, and consume private rather than public services. In addition to changing the traditional built form and sociospatial processes or urban areas, these developments also challenge the role of neighborhoods and local communities as the first tier of political and civic engagement that links residents to the wider municipality.

Second, there are concerns, particularly in Western Europe, that neighborhoods are becoming increasingly segregated along ethno religious lines, with a resultant undermining of national belonging and social cohesion. In the UK, the concept of parallel lives denotes the spatial and social isolation between neighborhoods comprising different ethnic groups, in which strong internal community ties are premised on the exclusion of outsiders. The work of urban geographers such as Fred Boal illustrates that ethnic and religious segregation within neighborhoods is not a contemporary development and that differentiation of residence is influenced by population change and migration and linked to wider notions of community identity, belonging, and interaction.

Changing Forms of Neighborhoods and Communities

Urban geographers have identified the scale of changes occurring in urban localities across the globe and manifested most visibly at the neighborhood level. Neighborhoods have always formed and reformed, but the contemporary pace of change is greater than its historical precedents. Neighborhoods are more susceptible to rapid local economic restructuring and to increasing population flows within and between nation states. While neighborhoods and communities are posited as providing important anchors for social identity, familiarity, and belonging, the sociospatial processes occurring within them mean that traditional certainties about neighbors and access to housing and employment are diminishing. As individuals become more mobile and are required to become more flexible and responsive, it is likely that many will live in more neighborhoods during their life course than previous generations. However, mobility and choice are far from universal, and neighborhoods in the developed and developing world are sites of exclusion and disadvantage as well as places of opportunity. Just as cities increasingly compete in a global economy, neighborhoods, including those within the same city, are also in competition for forms of capital, making the trajectories of individual neighborhoods uncertain.

The reductions in space and time manifested in processes of globalization are also evident in how neighborhoods are identified and defined, for example, through online residential search sites, while information technology also creates new opportunities for the formation of communities. However, these technologies are still linked to the local sociospatial manifestation of forces that define what a neighborhood is. Economic development, population change, and technological advances dominated the historic formation and evolution of neighborhoods and local communities in the nineteenth and twentieth centuries. Environmental sustainability may well prove to be an equally important factor for neighborhoods and communities in the future.

See also: Chicago School; Ecological Fallacy; Gentrification; Governance, Urban; Planning, Urban; Public Spaces, Urban; Regeneration to Renaissance; Segregation, Urban; Urban Policy.

Further Reading

Boal, F. W., Poole, M. A. and Murray, R. C. (1978). *Religious Segregation and Residential Decisions in Belfast*. Belfast: Queens University.

Cohen, A. P. (1985). *The Symbolic Construction of Community.* London: Tavistock.

Crow, G. and Allan, G. (1994). *Community Life. An Introduction to Local Social Relations*. Hemel Hempstead: Harvester Wheatsheaf.

Etzioni, A. (1995). *The Spirit of Community: Rights, Responsibilities and the Communitarian Agenda*. London: Fontana.
Jacobs, J. (1961). *The Death and Life of Great American Cities*. Harmondsworth: Penguin.
Kearns, A. and Parkinson, M. (2001). The significance of neighbourhood. *Urban Studies* 38, 2103 2110.
Lees, L., Slater, T. and Wyly, E. (2008). *Gentrification*. New York: Routledge/Taylor and Francis Group.
Ley, D. (1974). *The Black Inner City as Frontier Outpost: Images and Behavior of a Philadelphia Neighborhood*. Washington, DC: Association of American Geographers.
Park, R., Burgess, E. W. and McKenzie, R. D. (1925). *The City*. Chicago, IL: University of Chicago Press.
Putnam, R. (2000). *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon and Schuster.
Sampson, R. J. (2003). Collective efficacy. In Christensen, K. & Levinson, D. (eds.) *Encyclopedia of Community*, vol. 1, pp 205. London: Sage.
Sampson, R. J., Morenoff, J. and Gannon Rowley, T. (2002). Assessing neighborhood effects: Social processes and new directions in research. *Annual Review of Sociology* 28, 443 478.
Suttles, G. (1972). *The Social Construction of Community*. Chicago, IL: University of Chicago Press.
Taylor, P. (1999). Places, spaces and Macys: Place space tensions in the political geography of modernities. *Progress in Human Geography* 23, 7 26.
Zorbaugh, H. W. (1929). *The Gold Coast and the Slum A Sociological Study of Chicago's Near North Side*. Chicago, IL: University of Chicago Press.

Relevant Websites

http://www.neighbourhoodcentre.org.uk
ESRC Centre for Neighbourhood Research.
http://www.icpsr.umich.edu
Project on Human Development in Chicago Neighborhoods, Inter University Consortium for Political and Social Research.

Neocolonialism

M. Watts, UC Berkeley, Berkeley, CA, USA

Introduction

Standing at the heart of the term neocolonialism is a sort of paradox. Kwame Nkrumah, the anticolonial leader and first president of Ghana and the first theorist of neocolonialism as such, defined the condition as:

> modern attempts to perpetuate colonialism while at the same time talking about 'freedom' (Kwame Nkrumah, 1964: 41).

In his formulation it was the 'last stage' of imperialism because it emerged in the context of the Cold War (in which the Third World became the site of proxy anti communist struggles) and the deepening militancy of ex colonial territories. This was no longer 'naked colonialism' but rather more invisible modalities – economic, ideological, political, and cultural – in which colonial exploitation was perpetuated. For Nkrumah, the control by imperial powers over nominally independent states was achieved through new forms of corporate and especially financial forms of capital, by a psychological dependency among Third World elites, by the effects of what he called 'limited wars' and by the capitulation of African, Latin American, and Asian leadership to the hegemonic forces of the former colonial states.

While Nkrumah first gave voice to the concept of neocolonialism as an analytical device, the substance of the term was an integral part of African anti imperialist theorizing. Leopold Senghor, Alioune Diop, and Albert Memmi all articulated similar sorts of ideas in the 1950s and 1960s. In 1967, it was Franz Fanon, in particular, in his searing account of Algeria and what he called "the pitfalls of nationalist consciousness," who laid out of profound map of the forms and norms of the neocolonial condition. The Algerian war of liberation played an absolutely indispensable role in the formation of neocolonialism as a category of thought, and French intellectuals – most obviously Jean Paul Sartre in 1964 (reprinted 2001) – also contributed to the theorization of, in his case, French neocolonial rule and the deployment of violence (the assassination of Patrice Lumumba in particular) as its particular instrumentality.

Neocolonialism as Exploitation

Central to the idea of neocolonialism is the persistence of 'colonial' (or more properly imperial) and therefore exploitative relations. There are three broad senses in which exploitation is usually deployed in this regard: first, as a modern form of industrial or commercial land use (or mineral extraction) derived probably from its Latin origins as an arrangement or explanation. The second meaning is a generic sense of unjustness or oppression, referring to a potentially wide ranging social relations across time and space (exploitation of women by men, workers by capitalists, slaves by slave owners, low by high castes, serfs by feudal overlords). The range of opinion here can be substantial ranging from the Kantian view that exploitation refers to the instrumental treatment of humans to coerced activity to psychological harm. The analytical content is of less concern than its moral standing and the moral force of the reasoning (whether and how the state or groups of states should prohibit exploitative transactions or refuse to enforce such agreements). The third meaning which is central to neocolonial theorizing is explicitly analytical in the sense that it purports to provide a theoretical and conceptual ground on which the claim – A exploits B when A takes unfair advantage of B – can be assessed. In philosophical terms, one can say that social science seeks to understand the truth conditions under which such a claim can be made of a particular social setting.

The theorization of exploitation in neocolonialism was originally (and to a large extent still is) associated with the work of Karl Marx and the notion of surplus appropriation and the labor theory of value, and subsequently theories of imperialism. Marx's account identifies a fundamental contradiction at the heart of capitalism – a contradiction between two great classes (workers and owners of capital) which is fundamentally an exploitative relation shaped by the appropriation of surplus. Unlike feudalism in which surplus appropriation is transparent (in the forms of taxes and levies made by landowners and lords backed by the power of the church and crown), surplus value is obscured in the capitalist labor process. Marx argues that labor is the only source of value, and value is the embodiment of a quantum of socially necessary labor. It is the difference between the sale of a worker's labor power and the amount of labor necessary to reproduce it that is the source of surplus value. The means by which capital extracts this surplus value under capitalism – through the working day, labor intensification, and enhancing labor productivity – coupled to the changing relations between variable and constant capital determine, in Marx's view, the extent, degree, and forms of exploitation. In the first volume of *Capital* Marx identifies the origins of surplus value in the organization of production (the social relations of pro duction so called). In volume II Marx explains how ex ploitation affects the circulation of capital, and in volume III he traces the division of the total product of ex ploitation among its beneficiaries and the contradiction so created. In Marxist theory, two kinds of material interest – interests securing material welfare and interests enhancing economic power – are linked through ex ploitation (exploiters simultaneously obtain greater eco nomic welfare and greater economic power by retaining control over the social allocation of surplus through in vestments). Members of a class, in short, hold a common set of interests, and therefore have common interests with respect to the process of exploitation.

A long line of Marx inspired theorizing has, of course, attempted to grasp exploitative relations between coun tries, and this contributed directly to Nkrumah's original formulation. This is the heart of theories of imperialism, whether as the coercive extraction of surplus through colonial states through unequal exchange or through the imperial operation of transnational banks and multilateral development institutions (the World Bank and the International Monetary Fund). The so called 'anti glob alization' movement (especially focusing on institutions like the World Trade Organization) and the 'sweatshop movements' (focusing on transnational firms such as Nike) are contemporary exemplars of a politics of exploitation linking advanced capitalist states and transnational com panies with the poverty and immiseration of the Global South against a backdrop of neoliberalism and free trade.

In the Marxian tradition, there has been in general an abandonment of the labor theory of value – away from Elster's view that "workers are exploited if they work longer hours than the number of hours employed in the goods they consume" – toward John Roemer's notion that a group (or country) is exploited if it has "some con ditionally feasible alternative under which its members would be better off". Perhaps the central figure in de veloping these arguments is Erik Olin Wright. Building on the work of Roemer, Wright distinguishes four types of assets, the unequal control or ownership of which constitute four distinct forms of exploitation: labor power assets (feudal exploitation), capital assets (capitalist ex ploitation), organization assets (statist exploitation), and skill assets (socialist exploitation). While Wright and others have used this approach to grasp contemporary exploitation within so called advanced capitalism, there are obvious resonances between the operation of neo colonialism and his quartet of forms of exploitation.

Neocolonialism and the Process of Decolonization

Decolonization refers to the process, often long, tortuous, and violent, by which colonies achieve their national aspirations for political independence from the colonial metropolitan power. Decolonization can be understood as the period of later colonialism but implicit in the notion of neocolonialism is the idea that decolonization was incomplete or perhaps aborted to perpetuate a form of metropolitan or imperial hegemony. Modern coloni alism in its various norms and forms extended over the period from the fifteenth to the twentieth centuries. In the New World, which had been subjected to Spanish, French, Portugese, and Dutch colonial rule in the 'first age of colonialism', the first wave of decolonization oc curred in the eighteenth century. In this regard, the so called 'classical age of imperialism' in the last quarter of the nineteenth century was short, the first decoloniza tions of the second wave being achieved after the end of World War II. The two cycles of imperialism both con cluded with a limited phase of decolonization followed by the rapid collapse of empires and an irresistible push to political independence. What neocolonialism offered was a critical sense of how this process occurred and whether it was in any sense complete.

The first challenge to the first wave of colonization came in 1776 as British North American colonies de clared independence. While Britain maintained its Caribbean and Canadian colonies, the Napoleonic up heavals in Europe so weakened Spain and Portugal that European settlers from Mexico to Chile expelled their imperial masters. By 1825, the Spanish and Portugese empires were dead. In the subsequent 115 years up to World War II, decolonization was limited to Cuba in 1898 and two groups of British colonies: the white settler colonies (Canada, Australia, New Zealand, and South Africa) granted internal autonomy and finally full sov ereignty in 1931, and Egypt and Iraq after World War I. World War II marked the death knell for European col onization: India's separation from the British, Indonesia from the Dutch, the remaining Arab mandated terri tories, and Indochina from the French. The independ ence of Ghana in 1957 marked an avalanche of liberations in Africa, though the process was not complete until 1990 (Nambia). Between 1945 and 1989 over 100 new in dependent states were created.

Decolonization is a process marked by the achievement of political independence but the duration, depth, and character of decolonization movements vary substantially. In some African colonies, colonization was barely accom plished and resistance movements of varying degrees of organization and institutionalization attended the entire colonial project. In other cases, an organized anticolonial and nationalist movement came late, accompanied by a rapid and hastily assembled set of political negotiations in which it is clear that the metropolitan power wished to hand over the reigns of power with utmost expedience (Nigeria). In others, it took a war of liberation, a bloody armed struggle by leftist guerillas or nationalist agitators pitted against white settlers or intransigent colonial states (as in Laos, Vietnam, and Zimbabwe).

One of the problems with analyzing decolonization, as Fred Cooper notes, is that the story 'lends itself to be read backward and to privilege the process of ending colonial rule over anything else that was happening in those years'. It should also be said that any account of decolonization – or for that matter neocolonialism – presumes an account, or a theory, of colonialism itself: top down interpretations take colonial projects at face value, whereas the nationalist account denies any reality to the goal of modernization which the colonial state is purported to bring. In general, decolonization is seen as either (1) 'self government as an outcome of negotiated preparation and vision from above' by a colonial state apparatus, or (2) as a 'nationalist triumph from below' in which power is wrested (violently or otherwise) from recalcitrant colonizers. In practice, decolonization was an enormously complex process involving something of each, and shaped both by the peculiarities of colonialism itself and the particular setting in world time in which the nationalist drive began.

There are two forms of decolonization which rest on what one might call nationalist triumph. The first is built upon social mobilization in which a patchwork of anticolonial resistances and movements (many of which are synonymous with colonial conquest itself) are sewn together into a unified nationalist movement by a western educated elite (Malaysia, Ghana, or Aden). Mobilization occurred across a wide and eclectic range of organizations – trade unions, professional groups, and ethnic associations – bringing them into political parties and propelled by a leadership focused on racism, on liberation, and the sense of national identity of the colony, given its own history and culture. The second is revolutionary – Franz Fanon is its most powerful and articulate spokesman – in which the vanguard is not western educated elites or indeed workers, but the peasants and *lumpenproletariat*. It rested upon violence and rejection of any semblance of neocoloniaism. Decolonization rejected bourgeois nationalism (of the first sort); rather, as Fanon put it, "the last shall be first and the first last. Decolonization is the putting into practice of this sentence."

Both views depict nationalism as subsuming all other struggles and hence obscures and misses much history; both posit a 'true cause', as Cooper puts it, in which there is little truck with opposition. Mamdani's enormously influential book on Africa makes the important point that decolonization posed the possibility of breaking with the tradition of European colonial indirect rule (what he called 'decentralized despotism') in which African custom granted enormous powers to local systems of traditional (and therefore cultural) authority, and developing instead a sort of civic nationalism in which cultural politics did not play a key role. Most African states continued the colonial model in which African colonial subjects were granted a racial equality and citizenship rights but in which 'indigenes' were simultaneously a sort of bonus. In the historiography of the period, the nationalist road to self government tends to take for granted the depth and appeal of a national identity. It is precisely the shallowness of these nationalisms in the postcolonial period which reveals how limited is the simple nationalist account of decolonization itself. In practice, decolonization occurred in the context of all manner of contradictions and tensions between the national question and other social questions.

There is also a narrative of decolonization which has a singular vision but from the side of the colonial state. It was the colonial bureaucracy, long before nationalist parties arose, which shaped self government on a calculus of interest and power derived from an older conception of colonial rule (New Zealand and Canada) as a stepping stone to Independence. In this view, Africa by 1947 had already been set on the road to decolonization – this is a classic instant of Whig history – in spite of the fact that the colonial offices typically saw early African leaders as schoolboys or demagogues. Another version of the *dirigiste* theory is rendered through the cold calculation of money and cost. It was the decision making rationale of accountants estimating costs and gains – and who in particular gained – against the backdrop of imperial power's economic performance after World War II which sealed the fate of the colonies.

In all of these accounts – for India as much as Indonesia or Iraq – colonialism is as monolithic as the explanations themselves. There is a reduction involved in seeing Indians or Kenyans as colonial subjects or as national or proto nationalist actors. An alternative approach pursued by the so called subaltern school sees colonialism as a contra metropolitan project, moving against trends to exercise power under universal social practices and norms. In other words, the hegemonic project of colonialism fragmented as colonial rule attached itself to local idioms of power. From this experience characterized by hybrid forms of identity, of blurred boundaries, and contradictory practices, the process of decolonization must necessarily look more complex than simply self rule managed from above by the colonial state or mobilized from below by nationalist forces. In the same way, the shift from decolonization to neocolonialism operated through similar sorts of complexity and intermixing in which hegemonic metropolitan power was exercised through a battery of institutions and practices ('development', 'capitalism', and 'socialism').

Neocolonialism as a Contemporary Category of Analysis

The term neocolonialism has fallen out of fashion since the 1970s (with perhaps the exception of a flood of Soviet

Marxist inspired, and largely ideological, studies pub lished up to the mid 1980s). It is quite true however that the broad thrust of critical development work from the late 1960s onward – dependency theory emerging from Latin America, the calls for a New International Economic Order, unequal exchange theory, and the French Marx isant modes of production debate, and Wallersteinian world systems theory to take handful of the most prom inent trends in development discourse – all spoke of re lations of exploitation between former colonial states and the advanced capitalist core. Indeed, any theory of im perialism almost by definition presumes the exercise of powerful forms of dependency among First and Third World states and to this extent the analytic core of neo colonialism identified by Nkrumah has been central to any critical and left wing account of global political economy.

More recently in the so called anti globalization protests – the movement of movements – the critique of corporate power, of the expanded role of finance capital in the impoverishment of the Global South, and the imperialist role of multilateral development institutions such as the International Monetary Fund (IMF) and World Bank is entirely consistent in substance with Nkrumah's account. The revivification of neocolonialism is especially clear in the World Social Forum (WSF). First convened in January 2001 in Porto Alegre, Brazil, the WSF in an annual meeting held by members of the so called anti globalization movements – sometimes dubbed the 'movement of movements' – to provide a setting in which global and national campaigns can be coordinated, shared, and refined. It is not an organization or a united front but "an open meeting place for re flective thinking, democratic debate…by groups and movements of civil society that are opposed to neolibera lism and to domination of the world by any capital or any form of imperialism." The WSF has grown substantially from its first meeting in Brazil. Subsequent meetings in 2002 and 2003 were also held in Porto Alegre, and thereafter in Mumbai (2004), Porto Alegre (2005), and Nairobi (2007). In 2006, a 'polycentric forum' was held in Bamako (Mali), Caracas (Venezuela), and Karachi (Pakistan). In 2001, 12 000 people attended the WSF; in 2007, the number had grown to 60 000 registered at tendees, and 1400 organizations representing 110 coun tries. The WSF has also prompted the establishment of a number of regional fora – the Asian Social Forum, the Mediterranean Social Forum, and in 2007, the first US Social Forum – though not all of them stand in a similar relation to the 'parent body'.

The genealogy of the WSF is complex. The fact that four of the 11 WSF meetings have been held in Porto Alegre – a city with strong connections to the Brazilian Left and the Workers Party and the home to an in novative model of local government and participatory democracy (so called participatory budgeting) – says much about the broad ideological thrust of the Forum. Ideologically, it clearly stands against neoliberalism and free market capitalism; it is of the Left but looks for new and different models of economic and political organi zation drawing from a vast array of experiments embracing the landless workers movements, anti dam struggles, indigenous peoples, and anti corporate and multilateral struggles. The idea of a global convention of anticapitalist movements was in part driven by the desire to provide a counterweight to the World Economic Forum held every year in Davos and by the difficulty of organizing mass protest in Switzerland capable of gen erating sufficient media coverage to challenge the pre vailing hegemony of free market discourse and practice. The protests against the World Bank and IMF annual meetings in 1999 and thereafter – most notably in Seattle, Genoa, and Washington, DC – were an important milestone in the move toward an alternative forum for civic movements opposing unfettered capitalism around the world.

It is impossible, however, to understand the WSF outside of the counter revolution in development thin king and relatedly the growing dominance of neolib eralism (free markets, free trade, privatization, and state cut backs). The abandonment of Keynesian models of capitalist development – marked by the ascendancy of Ronald Reagan, Margaret Thatcher, and Helmut Kohl – and the rapid adoption of the economic ideas associated with Frederick Hayek and Milton Friedmann and the Mont Pelerin Society had massive and direct implications for the Global South beginning in the 1980s with the massive onslaught of structural adjustment and stabili zation programs. It was out of this combination of 'eco nomic reform' (viz., the rapid liberalization of state led development), 'shock therapy', and in many places mas sive economic recession (e.g., the early 1980s and the late 1990s) that the plethora of movements, often arraigned against the privatizations of various commons, arose. In contradistinction to the triumphalism (and purported inevitability) of globalization that dominated the 1990s, the WSF stood for, in their own language, "another world is possible" rather than "there is no alternative". In some circles, the WSF is held up as a shining example of what Negri and Hardt call 'the multitude'. The term neo colonialism does not appear prominently in the WSF but the affinities are clear. The language now is of 'recolon ization' (as the process) and 'empowerment' and 'alter native development' (as the solution).

Conclusion

Neocolonialism more than anything was a key marker of a certain sort of 1960s Third World nationalism. Neo colonialism was a by product of its largely African and

Marxist origins, of the Bandung movement, and of the contradictions of decolonization as it unfolded in the wake of World War II and in the heart of the Cold War. For a while, it was central to thinking of the theories of imperialism within a Marxist frame but it fell out of intellectual fashion. The so called neoliberal counter revolution and the devastating consequences of structural adjustment and economic reforms on a large part of the Global South (and Africa, in particular) has given neocolonialism a shot in the arm (seen in the various forms of anti or alternative globalization movements). The anti IMF movements across Africa in the 1980s and 1990s had, in this respect, more than a tinge of Nkrumah's original intent. There are also close affinities between neocolonialism in the Nkrumah sense and postcolonial theory , but the latter always distanced itself from the sort of determinism and historical telos that accompanied so much of Leninist and orthodox Marxist accounts of empire.

See also: Colonialism I; Colonialism II; Dependency; Neoliberalism and Development; Structural Adjustment.

Further Reading

Arrighi, G. and Pearce, B. (1972). *Unequal Exchange*. London: Monthly Review Press.
Brewer, J. (1987). Exploitation in the new Marxism of collective action. *Sociological Review* 35, 84 96.
Buchanan, A. (1985). *Ethics, Efficiency and the Market*. Totowa, NJ: Rowman and Allanheld.
Chamberlain, M. (1985). *Decolonization*. Oxford: Blackwell.
Cooper, F. (1997). *Decolonization and African Society*. Cambridge: Cambridge University Press.
Elster, J. (ed.) (1986). *Karl Marx: A Reader*. London: Cambridge University Press.
Fanon, F. (1967). *The Wretched of the Earth*. London: Penguin.
Guha, S. and Spivak, G. (eds.) (1988). *Subaltern Studies*. Oxford: Oxford University Press.
Hardt, M. and Negri, A. (2003). *The Multitude*. Cambridge, MA: Harvard University Press.
Harvey, D. (2005). *A Brief History of Neoliberalism*. London: Clarendon.
Hill, J. (1994). Exploitation. *Cornell Law Review* 79, 631 699.
Leite, J. (2005). *The World Social Forum*. London: Haymarket Books.
Lenin, V. (1916). *Imperialism*. Moscow: Progress Publishers.
Mamdani, M. (1995). *Citizen and Subject*. Princeton, NJ: Princeton University Press.
Marx, K. (1992). *Capital*, Vols. I III. London: Penguin.
Mertes, T. (2005). *The Movement of Movements*. London: Verso.
Moore, B. (1973). *Reflections on the Causes of Human Misery*. Boston, MA: Beacon Press.
Nkrumah, K. (1965). *Neocolonialism: The Last Stage of Imperialism*. London: Thomas.
Rey, P. P. (1977). *Les Alliances de Classes*. Paris: Maspero.
Roemer, J. (1986). An historical materialist alternative to welfarism. In Elster, J. & Hylland, A. (eds.) *Foundations of Social Choice Theory*, pp 133 164. Cambridge: Cambridge University Press.
Sartre, J. P. (2001). *Colonialism and Neocolonialism* (first published in 1964). London: Routledge.
Starr, A. (2005). *Global Revolt*. London: Zed.
Wallerstein, I. (1977). *World Systems Theory*. London: Academic.
Wright, E. (1985). *Classes*. London: Verso.
Young, R. (2001). *Postcolonialism*. Oxford: Blackwell.

Relevant Websites

www.forumsocialmundial.org.br
World Social Forum.

Neoliberal Economic Strategies

R. Le Heron, The University of Auckland, Auckland, New Zealand

Glossary

Accumulation by Dispossession Arising from corporations and elites exercising class power, to set in motion and benefit from privatization and financialization processes in the economy.

At-A-Distance Governance The development of relationships of coordination and control based on a local acceptance of remote rules, conventions, calculative practices, procedures, and so on through institutionalized power geometries. The effects are often redistributive as well as generative of accumulation.

Commons It denotes dimensions of biophysical processes (e.g., land, water, and air) that are outside of private ownership but available to enter production processes with few constraints other than the productivity of the biophysical processes. Appropriation of commons takes many forms (e.g., of land, fisheries, airspace, intellectual property, and genes).

Contracts Legal arrangements expedite the number, frequency, repeatability, and formalization of transactions among economic actors. It is usually applied to external arrangements among discrete actors (e.g., firm to firm and contracted workers in rural and urban activities) but increasing to internal transactions within businesses, government, public–private organizations, and NGOs, to insert new layers of audit and performance in the conduct of activities.

Development Project Distinguishing mid-twentieth-century model of nationally organized economic growth, instituted on a world scale that provided relatively stable conditions for especially national accumulation, with negotiated social contracts between capital and labor in nation-states.

Enclosure Traditionally seen as the systematic institutional bounding of access to, conditions of use of and rights over the distribution of product from biophysical processes (e.g., land and soil, sea, and air) that along with marketization is integral to capitalist development.

Entrepreneurialism (Including Urban Entrepreneurialism) A vision of social activity and organization holding that individuals through their special creative and innovative qualities contribute to growth and profit of their spheres of activity. The vision downplays the importance of contextual constraints for individual initiative, the role of contingency in shaping entrepreneurial activity, and power relationships in constructing criteria and justifications for success.

Financialization The elaboration of the capitalist financial system following widespread deregulation, beyond that required for international trade and productive investment flows, leading to speculation, fraud, predation, and thievery on unprecedented scales, with major redistributive effects in favor of corporate and elite interests.

Globalization Project A political and economic construct of mainly elite and corporate interests, framed by a dominant idea, market rule on a global scale, to organize global economic growth, having context-specific interpretations, practices, and distributive and redistributive implications, favoring particular groups.

Governance It refers to governing practices, often at multiple geographical scales, that are distributed internally and externally (e.g., in and between entities) in economic and institutional/political processes, marking the blurring of boundaries between and within public and private sectors and masking the interplay of interests and the distribution and redistribution of gains and losses.

Hegemonic Project A lived system of meanings and values affirming geographically grounded practices that are taken for granted, naturalized, normalized, and dominant in a context (used initially with reference to the nation-state but now applied internationally to, e.g., neoliberalism).

Post-Structural Political Economy An emerging strategy in the social sciences that explicitly and deliberatively deploys post-structural approaches with structural approaches in research.

User Pays A mechanism for assigning the costs of access to and use of resources, goods, services, ideas, and so on to users, regardless of the user's ability to pay. One social effect is a redistribution of social surplus from poorer to wealthier sections of the population.

Introduction

This article begins with a note of caution on what can be realistically accomplished and an observation about where it is thought that geographers as knowledge workers exploring neoliberalizing practices are likely to head in the next few years. The immediate unease is over the phrasing neoliberal economic strategies, because this gives the impression of being an unproblematic classifi cation of a set of actions, maneuvers, or campaigns, as most dictionaries indicate, to introduce greater freedom in political institutions relating to the economy. Some of concern this article has stems from living in a country that has had recent, rapid, and deep economic re structuring, much described by human geographers in cluding the author of this article as neoliberal in character. We wish to disturb the commonsense position by arguing that each word bundled into neoliberal eco nomic strategies must be approached carefully. We do so by thinking geographically, especially with regard to the multiple and ever changing politics of the spatial scales, connections, and boundaries that attend claims, be haviors, and activities associated with neoliberal pro cesses. The accumulated insights of economic geography risk being disabled by exercises such as attempting to slavishly fill the neoliberal economic strategies 'box' unless such exercises are finely contextualized. We con tend that a process disposition toward knowledge pro duction leads us from neoliberalism to neoliberalization to neoliberalizing practices, which helps with how we put our growing geographic knowledge to work. This leads to the guarded optimism that a powerful new direction of geographic research on neoliberal processes is appearing regarding humans and nonhumans in a socionatural universe, drawing from and adding to the critical litera ture developed already by urban, economic, and devel opment geographers.

The geographers' penchant for empirical exploration in response to alarming reports on the condition of planet Earth has seen growing attention on resource issues. One exciting thread is the interest is Earth processes where economic geographers are using the think kit of a re lational ontology to explicate qualities of the Earth that enable where and how life in its diverse forms inhabits the planet. This may seem some distance from neoliberal economic strategies. The argument is that as human geographers have come to know neoliberal processes better they are revealing both a whole range of new understandings and possibilities and also how such knowledge confers different political and ethical hori zons. This is a very different intellectual frontier from the twentieth century. Intriguingly, the recent preoccupation with neoliberalism appears to have rather fortuitously coincided with the Earth processes turn. Importantly, human geography has on hand insights of post structural political economy cum post structural political ecology that allow very different questions to be asked, in and from differently understood contexts.

Distinctions and Dangers

The push in the early part of twentieth century to de velop the national economy, via the national develop ment project, as a set of activities to support populations (using politically agreed rules so as to bring forward claims for improvement and for arriving at social settle ments) had by the end of twentieth century been overwhelmed by a new organizing mythology, the globalization project, wherein national populations should be competing to participate in the global eco nomy, through competition. This point of entry into neoliberal processes paints a broad picture of a capital centric project holding sway. Three themes to keep in mind about this reality are the high levels of national and international concentration in most economic activities, the spread of business models into government and community activities, and the character of engagement among the domains of business, government, and com munity, especially the activities of corporate and elite interests. This economic landscape fundamentally alters the terms under which resource access, allocation, and distribution are handled. In short, investment decisions made in the name of accumulation and profits are very central. The overarching questions are still who is getting what, where, by what means, and with what effects upon human and nonhuman others. Major interests are in clined to neoliberal priorities because they appear to provide more favorable conditions for profits for local, national, and international interests, at the scales needed especially by large entities.

The term 'neoliberal' is recent. Conventionally, it embodies the overriding principles of marketization (making markets when none existed) and liberalization (freeing up markets from the restraints of governments), which together offer a framework for facilitating and managing globalizing economic processes (connecting markets to enable greater prospects of economic growth). The speed with which neoliberal ideas have proliferated has astounded commentators. On one level the emphasis on markets in a liberal order is a revisiting (one meaning of neo) of thinking deeply entrenched in Western polit ical economy. In orthodox terms, the neoliberal goal when counterpoised against the early to late twentieth century experience of state led national development is about replacing politicized hierarchies and bureaucracies with depoliticized markets.

On another level the pro market movement has pre cipitated an enormous struggle over knowledge pro duction. Why? Because the very idea of freeing up political institutions challenges the established conven tions, attitudes, and commitments toward particular styles of politics that contributed to the successes of the national development vision. Some pressing challenges include the politics of whether politics should be integral to theoretical and empirical analysis (the end of tradi tional political economy?); the politics over what is ac ceptable to be named, located, othered, or vilified when discussing markets (silencing opposition?); the politics around the acceptability of depoliticizing market insti tutions (legitimating less open practices?); the politics of how much economy should be included in politics as well as how much politics should intrude into the economy (the influence of the cultural turn?); the politics of the boundaries between what can be represented about the world and performed in the world using knowledge so produced (keeping the academy and policy and community separate?); and the ethics of how researchers might conduct themselves with respect to politics (how politics might be made visible?). Clearly, ethics and politics in producing knowledge deserve to be a paramount concern.

One (ethical and political) intervention aimed at shaping how knowledge is produced is to unpack the category of neoliberal economic strategies. **Figure 1** summarizes three stylized distinctions – neoliberalism, neoliberalization, and neoliberalizing practices. The figure achieves several things. First, it illustrates the general idea that the search for 'isms' is a *cul de sac*, by noting different forms of disconnection from actual experience. A consequence is that the notion of neoliberalism is too abstract for practical politics. Second, while 'izations' edge thinking toward encompassing wider processes, the characterization is still that logically, someday, the process will be completed. One implication is that knowledge production is still separated from specific circumstances and so all the 'ifs', 'buts', and 'maybes' of lived experiences are left out of discussion. This takes us to the third point. By highlighting practices, that is, procedures understood in everyday contexts, which can be pointed to, named, challenged, and looked at from all angles, be shown to contain inconsistencies and so on, new spaces are opened that encourage engagement on different terms.

Noel Castree (2006: 1–6) submits that in analysis of neoliberal processes the relevant question is "What comprises the neoliberal component of this complex situation?" Each situation will be a distinct mixture of wider as well as more local social relations, some neoliberal. We can go on to inquire, "Are the relationships and interactions of the neoliberal component an especially significant or defining feature in an overdetermined situation?" Castree (2006: 1–6) But there are more issues.

When we undertake research on neoliberal processes in our different contexts what and how are we naming? This is important because naming provides targets for political critique (as in naming the 'need' for a better framework for accumulation that is a strong message of neoliberalism). Castree suggests we are doing two things. First, accepting that we can pin down strategies, policies, programs, plans, and projects, in or between the private and public domains, then we are examining contingently occurring processes and outcomes. His point is that the processes and outcomes may well have been different had the neoliberal elements not been present. Second, the focus of research is not to find some temporary variant of a pure form. Rather, interest is in the articulation between certain neoliberal policies and many other processes.

Neoliberalism (*noun*)	*Neoliberalization* (*noun*)	*Neoliberalizing practices* (*verb*)
Thought abstraction Monolithic exogenous processes transforming places from outside Pure and formal construct discussed as a model/ paradigm Universal, applicable everywhere, yet no such thing geographically Encourages search for similarities in different settings Assumes particular behaviours of actors	Finite process Sense of predefined sequence Closure when new state arrived at different starting dates different contexts and conditions plurality of forms variants but still with similarities still hard to find Actors shown to have adopted a range of behaviors in different contexts	In-the-making overdetermined (more-than-neoliberal context), other processes implicated coordinated contingent open to intervention continually negotiated counter-tendencies integral no scales where geographic difference ends and spatial similarity begins Neoliberalism understood as impure at all geographical scales Actor behaviours continually being reconstituted Practices made possible by active input from sets of administrators, experts, strategic brokers, and other translators who show the practices as imaginable and workable

Figure 1 Contesting categories of neoliberal.

Globalizing and Neoliberalizing Economic Processes

There is little doubt that globalizing and neoliberalizing processes need to be assessed against the backdrop of macro and structural knowledge of long run geog raphical and historical processes. But first a reality check – the late twentieth century champagne glass image of the world's resources. At the top of the glass is the gen erous cup of plenty, 80% of the world's resources con trolled and consumed by 20% of the world's population, at the thin base, 1–2% of resources available for 80% of the world's population, the constricted stem highlighting the gap of grief of such unevenness in worldly wealth.

Whatever one's politics, this image should be salutary. More so because the overarching dilemma of the post 1945 era was the reluctance of the so called developed countries to significantly alter the highly unequal access to wealth creation that characterized the twentieth century world economy at the time. The Bretton Woods institutions of the long post war boom formed ostensibly to deliver global development were captured by de veloped country interests. Crucially the aspirations of shared development, held by developing countries, did not eventuate. With the failure of the Bretton Woods agreement came a period of instability and crisis; aban donment of the gold standard, the oil crisis and inflation, global recession, and financial turmoil. The General Agreement on Tariff and Trade (GATT) framework came under strain as all manner of groups (e.g., farmers, unions, small businesses, and citizens) saw they had much to lose from ever more open markets. Non Tariff Bar riers were adopted to manage international trade and the US was seen as wielding too much economic power in favor of its transnational corporations (TNCs). Closure in 1994 of the Uruguay Round of GATT seeded the World Trade Organization (WTO), which was immedi ately criticized by less developed countries as not elim inating unequal power in the global economy. One consequence of the neoliberalizing movement was that it opened up space for nongovernment organization (NGO) and civil society resistance.

What was achieved then? Viewed from within any country the answer has to be reasonably positive for some decades. The construction of the national development project gave substantial gains to people in the developed countries and to a lesser extent many developing coun tries through state intervention in societal investment processes and a refocusing to a degree of business be haviors to meet national development goals. Bob Jessop argues that this was a time when directive (e.g., nation alization of industries) and supportive (e.g., subsidies, assistance, incentives, and import barriers) actions by the nation states were a departure from earlier patterns of more *laissez faire* government involvement. This extended into regional and local government. This directional change was widely adopted, with the so called Western model preferred over alternatives, such as development based on empowering local cultures. Underlying this was a vision of an economy powered by industrialization and the promise of technological abundance. The overriding metric of this organizing myth was outcomes and gains (e.g., aggregate growth) but this masks what people are giving up or losing as they experience development. Moreover, it cloaks the international structural background. The cumulative effects worldwide of individual national developments have included further mal distribution of wealth, deterioration of the bio physical environment, the paradox of Southern countries aspiring to Western style growth and wishing not to be just sources of natural resources, and the contradiction of sustainable development where global managers are ar guing for access to and use of resources in ways that do not compromise their accumulation of economic wealth.

Thus, the second half of the twentieth century was for most countries not a great half century. Indeed, for many countries and a high proportion of the world's popu lation, the last 50 years have been ones of deception at the hands of developed countries and evolving TNCs. A development framework was found operationally to be a framework that retained structural unevenness. Why then, as so many have asked, did neoliberal ideology spread so dramatically? Who in different settings viewed the neoliberal turn as especially attractive? How was a relatively diffuse ideology translated into specific gov ernmental actions and diffused so rapidly? Who and what groups were able to mobilize against the shift to neo liberal practices? These questions set the scene for the next section.

Overviews on the Directionality and Content of Neoliberalizing Economic Strategies

The past decade has seen vigorous efforts by researchers from geography and sociology, particularly to establish the changing contours of neoliberalizing processes in ways that illuminate the underlying spatiality of such processes, regardless of how they might be conceptual ized. A number of concerns have framed the debate. Bob Jessop for example argues from a state theoretical reg ulationist approach that we need to go beyond neoli beralism to interpret recent changes. He begins by highlighting a gradation, from policy adjustments (im proving performance of an accumulation regime and mode of regulation), to regime shift (the introduction of new economic and political principles) and radical sys tem transformation (a move from state socialism to capitalist social formation). In ideal type terms he labels the post 1945 period as Keynesian Welfare National State (KWNS), and places recent developments under the rubric of the Schumpeterian Workfare Postnational Regime (SWPR). Using overlapping aspects of economic regulation he identifies four SWPRs: neoliberalism which promotes market led economic and social restructuring; neostatism involving a market conforming but state sponsored approach to restructuring; neocorporatism consisting of a negotiated approach to restructuring by public, private, and third sector actors with the aim of balancing cooperation and competition; and neocommunitarianism which focuses on links between the market and the state via the third sector in securing economic development and social cohesion.

This preliminary overview has been extended in several interesting ways. Following Jessop, Neil Brenner picks up on the distinction between state projects intended to alter the state's institutional arrangements and state strategies that are aimed at the circuit of capital and business and civil society or the community. Brenner's interest is European city system urban entrepreneurialism. For this context he concludes that in contrast to the KWNS which sought to equalize the distribution of population, industry, and infrastructure (spatial balance), the hallmark of the glocalizing state era is the reconcentration of capacities for economic development in strategic subnational sites such as cities, city regions, and industrial districts. This has promoted much activity around local economic development. Work by human geographers mostly on urban entrepreneurialism, however, has meant a neglect of rural–urban links, especially around property, lifestyle, food movements, and so on.

In a more encompassing review Adam Tickell and Jamie Peck sketch three phases to the historical geography of neoliberalism (**Figure 2**). The first phase, proto neoliberalism, relates to a fledgling ideology that was, to use their words, not destined to translate into the ascendancy of neoliberalism. They contend, at least for the UK, that the Labour Government of 1976 going 'cap in hand' to the International Monetary Fund (IMF) for emergency financial aid helped create conditions for a regime shift. By declaring that 'there is no alternative' (since abbreviated as TINA), Margaret Thatcher established the necessity for a shift. The IMF's conditions amounted to a First World form of structural adjustment subsequently imposed on many developing countries in the 1980s (e.g., constraining of inflation, reduced money supply, and government spending cuts introducing market favoring rules). Their second phase, roll back neoliberalism, approximates the New Zealand experience in the mid 1980s. Facing severe balance of payments difficulties and a currency crisis, heavy assistance and subsidies for a once highly competitive agricultural sector, unsustainable fiscal spending with resistance to further taxes building up, the failure of 'think big' national development projects, and so on, the incoming Labour government of 1984 in New Zealand (elected on a caring society platform) announced unprecedented state sector and economic policy reforms (e.g., a floating exchange rate, removal of agricultural subsidies, downsizing and reorganization of government departments, and opening of the economy through rapid reductions in tariffs and quotas that had shielded an import substitution manufacturing base). This intervention in an interventionist state was a time of dismantling an existing order, of widespread 'pain before gain' as equivalence among sectors in the economy was established. But it was also a celebratory moment. A book, *Farming without Subsidies*, published in 1991 by the recently privatized government printing office, showed the world evidence that New Zealand's farming industries with minimal producer subsidies were thriving under the new regime. The third phase, roll out neoliberalism, denotes a consolidation, through extending principles and developing new economic policy agenda. As Tickell and Peck remark one of the far reaching effects has been attempts to remove key economic policy issues from political processes and enshrine user pays. **Figure 3** illustrates how neoliberalization can be seen to have extended beyond islands of experimentation to embedding in wider policy networks, institutional circuits, and political processes. Privatization for instance is extending to a growing number of sectors; not just industrial utilities like energy, water, and transport, but health, education, media, communications, pensions, prisons, and defense. Human geographers have been active in examining economic strategies in a range of countries, sectors, and industries. A number of special issues of *Antipode*, *Environment and Planning A,* and *Geoforum* extend this knowledge base. A major empirical omission is the study of financialization, which with privatization is regarded by David Harvey as a hallmark of the neoliberal era. What the figure does not show is the growing tension around financialization that is expressed in new contradictions among owners, workers, and managers and the reluctance of the state to resolve any impasse that results.

Significantly, however, much of the literature on neoliberal processes has a curious state centric bias, though for all this, it skirts the question of why and how the restructured state which continues to have deep tentacles in every sphere of the economy is not mobilized to resolve issues. Developments in business are infrequently studied. That TNCs and small and medium enterprises have been actively developing their own trajectories, more in spite of than because of state projects and strategies during the neoliberal era, seems to have escaped many researchers. Given that even under strong state intervention, businesses could not be held to socially defined goals, the greater apparent freedoms

	Proto-neoliberalism	*Roll-back neoliberalism*	*Roll-out neoliberalism*
Periodization	Pre-1980	1980s/early 1990s	Since early 1990s–
Mode of political practice	Extra-state project	Statization	Hegemony/state-building
Dominant discourses	Anti-Keynesianism/state failure	Small government deregulation	Paternal state/free economy
Key institutions	Liberal think tanks	Governing parties	State cadres
Mode of political rationality	Ideological critique	Ideological project	Technocratic management
Sources of resistance	Keynesian orthodoxy	Organized labour	Cyber-activitists
Intellectual frontier	Monetarist economics	Supply-side economics	Bourgeois sociology
Totemic figures	Milton Friedman, Hayek, Pinochet	Thatcher and Reagan	Clinton, Blair, Schröder, Greenspan
Principal agents	Theorists, philosophers	Vanguard politicians, political appointees	Policy-functionaries, technopols
Intellectual elite relations	Confrontation	Conciliation	Cooptation
Service delivery	Spending cuts	Privatization	Marketization
Labour regulation	Crisis of full employment	Mass unemployment	Full employability
Fiscal posture	Stagflation	Tight money/liberal credit	Persistent deflation
State finances	Fiscal crisis	Systemic indebtedness	Debt repayment
Geographic heartlands	Chicago	London and Washington, DC	Brussels, London, and Washington, DC
Geographic frontiers	Santiago	Brussels	Paris, Berlin, Hong Kong, Singapore, and Johannesburg
Space of resistance		British cities and coalfields, North American rustbelt	Anti-globalization confrontations, France, Malaysia
Scalar constitution		National	Glocal
Financial discipline	Inflation	Structural adjustment	Standards and codes
Ethic	Individualism	Amoral marketization	Moral authoritarianism

Figure 2 Modes of neoliberalization. From Tickell, A. and Peck, J. (2003). Making global rules. Globalization or neoliberalization? In Peck, J. & Yeung, H. (eds.) *Remaking the Global Economy*, pp 163 181. London: Sage, Table 10.1 Modes of neoliberalization, p 169.

available in the neoliberalizing global environment to organizations (and their management) that exceed in size many nation states should mean shining the spotlight onto business governance strategies. Investigations of global production networks, global supply chains, and transnational business behavior from a variety of perspectives (e.g., fair trade, ethical trading, organic production, corporate social responsibility, and issues of public health) are gradually being linked to the literature on neoliberalism. These developments are not independent of the new spatialities that Brenner identifies. Thus clusters, industrial districts and innovation systems, local food provisioning, and short supply chains, may have strong local territorial traces but they are also linked into global production networks and global supply chains, whether largely internal or external to MNC and TNC organization. The spatiality and the redistributive aspects of these connections is a shadowy presence in writings on neoliberalizing processes.

Understanding and Coexisting with Neoliberalizing Processes

The general drift in the literature toward the explication of neoliberalizing processes took a sharp leap forward with the publication of *Neoliberalization: States, Networks and People*. This book targets three concerns: the orthodox terrain of mainstream economic development and alternatives (a state centric exposition); the opening of conceptual space between states and markets (achieved by examining intermediaries); and consideration of states and subjectivities (emphasizing the constituted nature of understandings and behaviors). The book's aspiration is gaining leverage for alternative political directions. To this end, it identifies four ways of understanding neoliberalization that to some degree signals a shift from representative knowledge about neoliberal processes to knowledge on how to perform differently in the context of neoliberalizing processes. The understandings in brief

		Spatial Relations	
		Shallow neoliberalization ...pursuit of locally specific strategies	*Deep neoliberalization ... interlocal logics and reflexive relations*
Institutional	Roll-back neoliberalism ...its destructive and deregulatory moment	*Attacks on inherited Keynesian-welfarist structures, coupled with primitive deregulation of markets,* including: • Monetarist macroeconomic management • Primitive marketization: dogmatic deregulation and privatization • Place-specific assults on institutional and spatial strongholds of welfare statism and social collectivism (e.g.,social service cuts, deunionization)	*Extension of neoliberal strategies to the international domain,* including: • Reductions in overseas aid • Imposition of structural adjustment programs • Initial liberazation of financial markets and trading relations • External imposition of neoliberal strategies • Intensification of coercive pressure emanating from international markets
forms	Roll-out neoliberalism ...its creative and proactive moment	*Proactive statecraft and institution building in service of neoliberal goals,* including: • Invasive moral reregulation of the urban poor • Expansion of penal state apparatus and social control policies • Continued crisis management of deregulated and privatized sectors • Extension of, and experimentation with, market-complementing forms of regulation • Technocratic economic regulation within neoliberal parameters	*Normalization of neoliberal logics and premises in interlocal relations,* including: • 'Fast policy' development through inter-local transfer, learning, and emulation • Appropriation of networking forms of governance and policy development • Institutionalization of competitive globalism (e.g., WTO dispute resolution architecture, GATS proposals regarding public ownership) • Posture of permanent adaptability and reflexivity in fields like urban governance and social/penal policy

Figure 3 Spaces of neoliberalization. From Tickell, A. and Peck, J. (2003). Making global rules. Globalization or neoliberalization? In Peck, J. & Yeung, H. (eds.) *Remaking the Global Economy*, pp 163 181. London: Sage, Table 10.2 Spaces of neoliberalization, p 170.

are: as an ideological hegemonic project (e.g., the dis cursive development of the idea, and how it was natur alized); as systematic shifts in class power (e.g., who is able to decide on investment decisions, who captures surplus and on whom as a result burdens fall, and who can stall the mobilization of state action); as changes in boundaries of the state (e.g., involving the appearance of NGOs, private–public partnerships, the concept of joined up government, and subsidiarity or at what level gov ernance is best undertaken); and as governmentality (e.g., especially the development of new governing practices, such as benchmarking, best practice, standards, certifi cation, audit, and the accompanying subjectivities that are constituted). In some respects, the NGO and civil society movements are losing momentum, revealing the new challenge of reactivating state engagement to deal with ethically inspired pressures.

The idea of governing mentalities is especially helpful as it provides a framework better suited to probing ethics and politics in contexts where choice is a guiding maximum. The governmentality framework is in a sense a way to deal with the multiplicity of networks that are activated and nurtured during episodes of heightened restructuring/transition and as a result of continuing re adjustment. But part of the difficulty of grappling with neoliberalism has been the fact that regulation theory, which spotlights the mediation of economic processes under the interventionist state tends to confine politics to the state and to ascribe prescriptive logics to politics. This is not to suggest abandoning macro and meso structural frameworks but to recognize the different tasks of political intervention. The ethics movement is challenging all spheres of economic life, from the personal (e.g., how does one behave in relation to declared values), valuing the nonhuman (e.g., animal welfare and ecosystems), calling for new evaluate metrics (e.g., ecological footprints based on land resources used to support territorial populations, food miles that calibrate carbon contributions rather than price, and General Progress Indicators instead of gross domestic product), to the corporate (e.g., insisting on accountability such as corporate social responsibility and rewriting the rules of performance).

The utility of the concept of the political project is being explored by economic geographers Wendy Larner, Nick Lewis, and Richard Le Heron who are seeking to understand the genesis and momentum of political initiatives in a neoliberalizing environment. A political project is a speculative imaginary defined by a number of dimensions that delineate how projects are simul taneously constituted and constitutive. Political projects (always of particular interests) spring from rationalities that justify and legitimate claims and actions, an in telligibility to others that attracts attention and com mitment, relate directly or indirectly to spaces that because they are identifiable become potentially gov ernable, providing governable subjects who can be en rolled, through translations of key ideas into immediate circumstances and practices, with the help of technolo gies that facilitate action and assist with knowing the world. In this characterization, political projects are re lational constructs or emerging networks giving new meaning to economic strategies, in the sense of seeking and gaining visibility and resources through action and the availability of a resource to be networked or aligned discursively into other political projects.

Progressive Spaces

It has been argued that economic geographers have come to engage with neoliberal economic strategies in in creasingly sophisticated ways. This is one key point of **Figure 1**. This growing process centered understanding is that what might start as a neoliberalizing economic strategy can be negotiated (prioritizing ethical inter ventions) and mediated (from political interventions) into something with progressive dimensions. This engage ment has consequences, however. One is that there has been a shift on the part of a growing number of re searchers from a politics of resistance mentality to a politics of emergence ethos. Some of the shift in theo retical thinking needed to more fully activate this movement has been highlighted by Castree. To this must be added the full weight of the most recent Gibson Graham post structural political economy intervention, *A Postcapitalist Politics*. They strongly favor focusing on what can be done through mapping community assets rather than retaining the needs centered approach of traditional political economy, the latter being complicit with forestalling instead of activating new political and economic directions. It is ironic that instituting more shared practices of understanding and making choices over valuing and using group assets could easily be classified as neoliberal like development of self man agement. The progressive potential of neoliberal at a distant technologies of governance should also not be discounted. By taking the ideas of subjectivity seriously (derived from the post structural literature), economic geographers are now equipped with capacities to reflex ively explore the connecting and making of place–place co constitutive processes (e.g., neoliberal and other pro cesses). At present, this is being propelled by the ethical movement, which is after all, people taking care over their choices and the ramifications and implications of choices made by them and others. In perhaps one of the more startling of contingencies in the history of geog raphy Bruno Latour, speaking at the British Sociological Association in 2007, called for making explicit the fra gility of the life support systems that make our spheres of existence possible. His call excites the notion that insights from the trajectory of economic geography's engagement with neoliberalizing economic strategies are part of what geography can contribute to this new ethical and political project around knowledge production.

See also: Capitalism; Competitiveness; Consumption; Firms; Global Commodity Chains; Livelihoods; Natural Resources; Privatization; Transnational Elites.

Further Readings

Brenner, N. and Theodore, N. (2002). *Spaces of Neoliberalism: Urban Restructuring in North America and Western Europe*. Oxford: Blackwell.

Castree, N. (2006). From neoliberalism to neoliberalisation: Consolations, confusions and necessary illusions. *Environment and Planning A* 38, 1 6.

England, K. and Ward, K. (eds.) (2007). *Neoliberalization. States, Networks, People*. Oxford: Blackwell.

Gibson Graham, J. K. (2006). *A Postcapitalist Politics*. Minneapolis, MN: University of Minnesota Press.

Harvey, D. (2006). Neo liberalism as creative destruction. *Geografiska Annaler* 88B(2), 145 158.

Heynen, N., McCarthy, J., Prudham, S. and Robbins, P. (eds.) (2007). *Neoliberal Environments. False Promises and Unnatural Consequences*. Oxford: Routledge.

Jessop, B. (2002). Liberalism, neoliberalism, and urban governance: A state theoretical perspective. *Antipode* 34(3), 453 472.

Larner, W. (2003). Neoliberalism. *Environment and Planning D: Society and Space* 21, 509 512.

Le Heron, R. B., Rees, E., Massey, E., Bruges, M. and Thrush, S. (2008). Improving fisheries management in New Zealand: Developing dialogue between fisheries science and management (FSM) and ecosystem science and management (ESM). *Geoforum* 39(1), 48 61

Leitner, H., Peck, J. and Sheppard, E. (2006). *Contesting Neoliberalism: Urban Frontiers*. London: Guilford.

Lewis, N. I., Larner, W. and Le Heron, R. (2008). The New Zealand designer fashion industry: Making industries and co constituting political projects. *Transactions Institute of British Geographers* 33, 42 59.

McCarthy, J. and Prudham, S. (2004). Neoliberal nature and the nature of neoliberalism. *Geoforum* 35(3), 275 283.

McMichael, P. (2000). *Development and Social Change. A Global Perspective*. Thousand Oaks: Pine Forge Press.

Mansfield, B. (2007). Property, markets, and dispossession: Development quota as neoliberalism, social justice, both and neither. *Antipode* 39, 479 499.

Peck, J. (2004). Geography and public policy: Construction of neoliberalism. *Progress in Human Geography* 28, 392 405.
Prudham, S. (2007). The fictions of autonomous invention: Accumulation by dispossession, commodification and life patents in Canada. *Antipode* 39, 406 428.
Radcliffe, S. (2005). Neoliberalism as we know it, but not in conditions of its own choosing: A commentary. *Environment and Planning A* 37, 323 329.
Rankin, K. (2004). *The Cultural Politics of Markets: Economic Liberalization and Social Change in Nepal*. Toronto: University of Toronto Press.
St Martin, K. (2007). The difference that class makes: Neoliberalization and non capitalism in the fishing industry of New England. *Antipode* 39, 527 548.
Tickell, A. and Peck, J. (2003). Making global rules. Globalization or neoliberalization? In Peck, J. & Yeung, H. (eds.) *Remaking the Global Economy*, pp 163 181. London: Sage.
Wilson, D. (2004). Toward a contingent urban neoliberalism. *Urban Geography* 25(8), 771 783.

Neoliberalism

W. Larner, University of Bristol, Bristol, UK

Glossary

Deregulation Reduction or elimination of state regulation over economic and social activities, and their replacement with market modes of governance.

Governmentality Analytical approach that emphasizes heterogeneous political problematizations and governmental technologies in constituting political formations.

Marketization Emphasis on quasi-entrepreneurial and market modes of action for the governing of economic and social life.

Neoliberalism Short-hand term referring to the new emphasis on markets, minimal states, and individual choice in contemporary political–economic formations.

Neoliberalization Analytical approach that emphasizes the contested political processes associated with neoliberalism.

Privatization Repositioning of public assets into the private sector, usually in an effort to increase efficiency and accountability.

Introduction

Neoliberalism is a short hand term referring to a new emphasis on markets, minimal states, and individual choice as a means of ensuring economic and social well being. Over the last decade geographers have paid con siderable attention to this change. They are concerned to identify how a political–economic transformation of this scope was able to occur, the consequences for economic, social, political, and environmental relations in a range of places and at a variety of scales, and the prospects that remain for social justice. More generally, neoliberalism has become a keyword across the social sciences, in cluding development studies, sociology, anthropology, and political studies. It is also widely used among policy makers and activists resisting the general move to mar ket based policies.

Despite its increasing popularity in geography and more generally there remains a great deal of confusion about neoliberalism. What exactly does this term cap ture? Is it an activist and academic caricature of a prag matic policy consensus that now has adherents from across the traditional left–right political spectrum? Does the term provide useful analytical purchase for those attempting to understand the ways in which economic and social relations are currently understood and acted upon? Is neoliberalism an economic and social successor to Keynesian welfarism, or a more pragmatic temporary political accommodation? Is it more usefully seen as in volving the qualitative transformation of states, spaces, and subjectivities? The answers to these questions depend on how neoliberalism is understood both theoretically and politically.

The Rise of a 'Neoliberal Era'

The term neoliberalism first appeared in social science literatures in the context of attempts to describe a dis tinctive set of policies deployed at the level of the nation state during the 1970s and 1980s. During this period, conservative and social democratic governments alike were involved in debates over the relative roles of states and markets. Whereas during the post war period, the state provision of goods and services to a national population was understood as the best means of ensuring social well being, by the 1980s markets were increasingly understood to be a better way of organizing economic activity because they were associated with competition, economic efficiency, and choice. In conjunction with this general shift toward the tenet of 'more market', deregu lation and privatization became central themes in polit ical programs.

Yet it is rare to find the actual term neoliberalism used in the political discussions of this period. Only in Latin America was the term widely used to describe market oriented reform programs, most notably in Pinochet's Chile. Elsewhere much more specific terms tended to be used, including political programs named after specific politicians – Thatcherism, Reaganomics, Rogernomics – or terms that focused on specific aspects of the economic processes in dispute – economic rationalism, monetarism, deregulation, privatization, and marketization. However, by the 1990s neoliberalism began to be more widely used as a *post facto* description of the general shift in policy paradigms that took place in the 1970s and 1980s. It re mains most common in those countries that have ex perienced dramatic state sector reform programs, notably those in Latin America (Chile, Mexico, Argentina) and the group of countries formerly known as the liberal welfare states (United Kingdom, Canada, New Zealand, Australia). Only recently has the term entered into popular political debates in the United States.

Whereas initially discussions focused on neoliberalism as a nation state project, more recently neoliberalism has begun to be used interchangeably with the term globalization to refer to the process of opening up national economies to global organizations such as the IMF and World Bank. Indeed, increasingly the two terms are now often linked through the phrase 'neoliberal globalization'. As with its developed country counterpart, this discussion was a response to political events in the 'real world'. By the late 1990s this language was being used to describe structural adjustment processes, in particular the imposition of market oriented reform programs associated with the so called Washington Consensus. The response to the so called Asian crisis generated further debate about the role of states and markets, as did efforts to understand the experiences of 'transitional economies' such as those in Eastern Europe and South Africa. In each case the term neoliberalism gave analysts a means of describing the new emphasis on the introduction of markets as a mode of governance in both economic and social domains.

In the last 5 years references to neoliberalism have proliferated. In both critical social science and activist literatures neoliberalism has replaced the earlier labels that referred to specific politicians and/or political projects. The domains in which analyses of neoliberalism are found have also expanded, now encompassing transnational, national, urban, and local scales. For academics, neoliberalism is now providing an analytical means to link macro level discussions of global economic change with micro level discussions of subjectivities. Like its important predecessor, post Fordism, which moved from an account of changes in economic production to that of 'New Times' which encompassed economic, political, and social processes, neoliberalism appears to have moved from being a descriptor of contemporary policy frameworks to that of the current period as a whole. References to the 'neoliberal era', 'neoliberal state', 'neoliberal city', and even 'neoliberal nature' now abound. For activists neoliberalism has become a means of labeling an increasingly ubiquitous set of market oriented policies that are understood to be responsible for growing numbers of economic, political, social, and ecological problems.

Theorizing Neoliberalism

As neoliberalism has become increasingly pervasive, so too have the intellectual tools used to analyze its origins and effects become increasingly disparate. Over the last decade, theorizations of neoliberalism have gradually broadened from that of a political philosophy based on the writings of Hayek and Friedman, to a policy program premised on marketization and a minimal state, to a new form of economic and social rule premised on 'government at a distance'. Analyses of neoliberalism have also begun to colonize debates, such as those around Third Wayism and good governance, which began as discrete endeavors. There is also an emergent discussion of new subjects and spaces, manifest in terms such as 'neoliberal citizenship' and 'neoliberalizing spaces'. Today, neoliberalism is variously understood as policy program, hegemonic ideology, and governmentality, and increasing numbers of analysts are working across these literatures in their accounts.

The analysis of policy frameworks remains the most common approach to neoliberalism in academic literatures. In this approach the focus is on the new emphasis on the market provisioning of formerly 'public' goods and services. Analysts attribute this shift in policy agendas to the capture of key institutions and political actors by a political ideology that rests on five values: the individual, freedom of choice, market security, *laissez faire*, and minimal government. These theoretical ideas are understood to underpin the new institutional economics (built on public choice theory, transactions cost theory, and principal agency theory) which, together with a new emphasis on managerialism, comprise the intellectual basis of the neoliberal challenge to Keynesian welfarism, and provided a diverse range of governments with the theoretical impetus for deregulation, marketization, and privatization. This is not to suggest that the rise of neoliberalism was inevitable or straightforward. Comparative institutionalists have revealed the variety of national paths through which these new ideas were able to work their way from relative obscurity to influence and political power. Diverse literatures on epistemic communities, transnational advocacy networks, global public policy networks, and global knowledge networks have all been deployed to explain how these ideas have become internationally influential.

Marxist and neo Marxist readings of neoliberalism link these policy changes to broader analyses of changes in nature of capitalism. For these commentators, it is the broader workings of capital that underpin the shift from Keynesian welfarism and developmentalism toward political programs favoring the relatively unfettered operation of markets. This has reestablished the conditions for sustained capitalist accumulation, privileging the dominant classes and restoring their income. The question of why the rest of us have acquiesced to this state of affairs has been the focus of sustained attention from neo Gramscians. These scholars argue that neoliberalism has been able to constitute subject positions from which its discourses about the world make sense to people in a range of different social positions. There are two foci in these discussions: first, the relationships between classes within the so called 'hegemonic bloc'; second, those between the state and civil society. The former focuses on the way in which particular constellations of political forces come together around a set of understandings. For example, Stephen Gill has drawn attention to the relationships between different fractions of capital, political actors, and key elements in the state apparatus in his efforts to explain the rise of neoliberalism at the global level. In the case of the latter, attention focuses on the way in which hegemonic political discourses articulate with broader social understandings. For example, Stuart Hall's classic analysis of Thatcherism argued that it represented an ideological hegemony based on the tenets of neoliberalism and neoconservatism.

Socialist–feminist analysts have built on these claims, emphasizing that new political and ideological forms emerge out of struggle and contestation rather than being simply imposed in a top–down manner. In doing so, they have drawn attention to the role of oppositional movements and broader social processes in determining the contingent outcomes of neoliberalism. While neoliberalism has indeed reinforced the power of finance capital and increased social polarization, equally importantly it has been shaped by the 'mainstreaming' – albeit in compromised fashions and with unexpected outcomes – of a wide range of social movement claims. Seen together, the contributions of neo Gramscian and feminists have encouraged closer inspection of particular neoliberal political projects that have revealed complex and hybrid political imaginaries, rather than the straightforward implementation of a unified and coherent philosophy.

Neo Foucauldian work on governmentality brings yet another perspective to accounts of neoliberalism. This literature argues that while neoliberalism may mean less government, it does not follow that there is less governance. While on one hand neoliberalism problematizes the state and is concerned to specify its limits through the invocation of individual choice, on the other hand it involves forms of governance that encourage both institutions and individuals to conform to the norms of the market. This literature has inspired innovative analyses of neoliberalism, which show how 'advanced liberal' strategies of rule, found in diverse realms including development programs, workplaces, educational institutions, and health and welfare agencies, encourage people to see themselves as individualized and active subjects responsible for enhancing their own well being. It has encouraged commentators to move beyond the terrain of the state in explaining the exercise of new forms of power and the implications for governmental, political, and ethical processes. The distinction between advanced liberalism and neoliberalism has also emphasized that strategies and techniques premised on 'governance at a distance' can be linked to a variety of political projects.

The varying contributions of these diverse theoretical literatures have broadened analyses of neoliberalism. Whereas early accounts were dominated by analyses in which the focus of attention was on the state either as specific set of political institutions or as acting in a relatively straightforward way on behalf of capital in imposing a coherent political project, more recently there has been a tendency to understand neoliberalism as involving historically and geographically contingent political processes, and to conceptualize it as a contested assemblage of diverse discourses and techniques out of which new political forms, spaces, and subjects are being constituted. Correspondingly, interest in neoliberalism among the social sciences has diversified from the structural accounts most often associated with political science, policy studies, and political economy to the ethnographic and embodied approaches more likely to be found in the domains of human geography, sociology, and cultural anthropology.

Geographies of Neoliberalism

If the Dictionary of Human Geography is an accurate reflection of changing disciplinary interests, then neoliberalism had come to the attention of human geographers by the late 1990s. Initial contributions focused on debates about the so called 'New Right' in the United Kingdom and United States, with a parallel but less high profile discussion taking place in development studies focused on the changing models of Third World development. For these early observers the extension of market relations was highly problematic. More specifically, deregulation and privatization were identified as transferring power away from democratically elected governments with a mandate to ensure universal service provision, toward private capital concerned primarily with furthering opportunities for accumulation. In turn, the shift from public to private sector and the new emphasis on individual choice was understood to erode the foundations of both national economies and traditional social solidarities. It was also widely accepted that neoliberalism was increasing spatial and social polarization, with women, visible minorities, and indigenous peoples being particularly disadvantaged.

In subsequent years considerable energy was spent criticizing the move away from Keynesian welfarist and developmentalist approaches in the developed and developing world, respectively and tracing the deleterious consequences of the new approach for politics, people, and places. Understandings of neoliberalism itself, however, were rarely questioned. Despite the debates and diversity within the broader theoretical literature on neoliberalism, early geographical contributions to this literature were dominated by Marxist and neo Marxist accounts in which it was usually assumed that neoliberalism was a relatively coherent reform program initiated and rationalized through a particular ideological framework, the widespread adoption of which could be attributed to the influence of key politicians and/or political organizations. Consequently in these accounts the state was seen as having a key role in guaranteeing and regulating the spread of neoliberal ideology and practice.

More recently, greater attention has been paid to the historically and geographically specific nature of 'actually existing' neoliberalism and the diverse forms this political project takes in different places. This has given rise to an important distinction between abstract accounts of neo liberalism premised on the model of unfettered market activity and freely choosing subjects, and neoliberaliza tion which involves multiply determined, geographically uneven, politically contested processes. Geographical scholarship has increasingly focused on the latter. This work has not only helped identify the diverse reasons for the emphatic turn to neoliberalism, it has also em phasized how neoliberalism is associated with the rescaling of political processes across supranational, na tional, urban, and local levels, and the way in which the new orientation toward markets and individualism has been incorporated into the common sense understand ings of the world.

This research has helped draw attention to mutations and contradictions within neoliberalism itself. For ex ample, as political ecologists began to examine the im plications of neoliberal processes for socionature, they emphasized the need to delineate carefully between processes such as deregulation, commodification, mar ketization, and privatization. Feminist geographers underlined the fact that women and ethnic minorities are not simply the victims of the political–economic pro cesses associated with neoliberalism, but rather they play an active role in resisting and thus shaping these new state forms and institutional configurations. Relatedly, there is now a growing theoretical and empirical dis cussion of what is being variously named as 'roll out', 'inclusive', 'social', and 'after' neoliberalism. Debates about these new developments range from claims that they are simply 'flanking mechanisms' to compensate for the inadequacies of the market, to arguments that stra tegically targeted resistance has resulted in the trans formation of the overall political project.

Seen collectively, these contributions suggest that human geographers have begun to take neoliberalism seriously. It is now widely accepted that there is no going back to the Keynesian welfarist and import substitution approaches that characterized post war development strategies. The research of the last decade has sensitized geographers to the need to identify and examine the different forms of neoliberalism over time and space, to trace the flows and networks through which neoliberal discourses and techniques have been disseminated, to analyze the ways in which neoliberalism is mutating, and to interrogate the various forms in which neoliberalism has been instantiated and embodied. Thus whereas early studies saw neoliberalism as a coherent policy agenda and focused on documenting the negative consequences of the new emphasis on the market, a more diverse range of theoretical and political approaches are now being de ployed in the effort of human geographers to research the new spaces and subjects emerging in the context of what is now understood to be a fundamentally new political formation.

There remain, however, two notable tendencies features to the geographic literature on neoliberalism relative to other disciplines. The first is a lingering commitment to methodological nationalism. Only re cently has the focus of geographical attention shifted from the domestic politics of particular country case studies to the globalizing processes driven by international insti tutions and actors. More research projects have begun to work across North/South divides, resulting in a more careful tracing of the intellectual, policy, and practitioner networks that underpin the global expansion of neoliberal ideas, and their subsequent manifestation in government policies and programs. The analyses that are emerging show that despite the assumptions in accepted accounts, the movements of neoliberal ideas are not one way and they do not necessarily flow in the directions expected. But much more work remains to be done in this area, and a deeper engagement with the existent literatures on policy transfer and global networks is needed.

There is also a tendency to see debates about neoli beralism as the continuing preserve of Marxist and neo Marxists. Certainly, the concept of neoliberalism is one overwhelmingly mobilized and deployed by left wing academics and political activists, and it is important to note that theoretical and political architects of neoli beralism don't name themselves in this way. However, analysts of neoliberalism are not the only well inten tioned contemporary political critics who inadvertently redraw the very configurations and effects of power they seek to vanquish. And certainly the ideological and ma terial processes associated with neoliberalism resist easy dismissal or glib characterizations. In this context, more geographical analyses that engage with the debates and themes found in wider theoretical and social science literatures on neoliberalism are needed.

The Politics of Neoliberalism

A broader engagement with the diverse bodies of scholarship on neoliberalism would also suggest that we should be wary of foreclosing political debates on this topic too early. This is particularly important given the growing prominence of discussions of neoliberalism in the United States. Understandably, for many critical commentators in this context neoliberalism remains in delibly linked to a right wing moral project. David Harvey, for example, recently argued that there is a risk that the moral discourses of cultural nationalism, Chris tianity, and family values are seen as providing a new basis for social solidarity in the United States in a context where neoliberalism has eroded more traditional social solidarities. Elsewhere, however, there are countervailing trends. In Latin America, for example, important variants of a developmental state based on a partnering ethos are emerging in the wake of ‘more market’ neoliberalism. Similarly, in liberal welfare states such as the United Kingdom, Canada, and New Zealand there is a new political impetus that emphasizes the building of human and social capital through strategic future focused in vestments. This suggests we should avoid setting up neoliberalism as a preconstituted explanation for con temporary political changes, and remember that neoli beralism is not the only political formation found in the current period.

If we accept arguments about multiple and contested processes associated with new political formations, we are also forced to consider carefully the assumption that the neoliberal state is an inherently unstable political form. If this new political formation is shaped by multiple pol itical projects and oppositional claims then it may be – as all state forms are – an uneasy political compromise. In our analyses of new modes of economic and social gov ernance, we need to know much more about the ways in which new political rationales and organizational forms are constituted and consolidated in particular places. How has neoliberalism articulated with other political projects? How do different versions of neoliberalism draw people into regulatory relationships with the state? Are different people being drawn in different ways? What of the traditional political alliances of both the left and right? How are these alliances being reconfigured by the foregrounding of new social actors? What forms of social subjectivity are being assumed and constituted? What are the gendered and racialized paradoxes that are emerging?

What are the implications of highlighting complexity and contradictions in analyses of neoliberalism? In con structing neoliberalism as a monolithic apparatus that is completely knowable and in full control of policy makers, many critical analyses inadvertently reconstruct its he gemony. The recognition that contemporary political formations are inherently contested, rather than as suming we live in a monolithic neoliberal world, is a more empowering way of assessing contemporary polit ical conditions and potentialities. Above all, it is the programmatic coherence of neoliberalism we should challenge. Indeed, it may be more useful to see the current moment as involving profound experimentation rather than the rolling out of a political agenda. Not only would it help us overcome the hopelessness generated by monolithic accounts of the ‘neoliberal project’, it would allow us to think carefully about the multiple forms that political strategies, techniques, and subjects are taking. In this context the need for more rigorous theorization and further careful research should be clear.

See also: Governmentality; Political Ecology; Social Geography.

Further Reading

Barry, A., Osborne, T. and Rose, N. (1996). *Foucault and political reason: Liberalism, neo liberalism, and rationalities of government*. London: University of Chicago Press.

Brenner, N. and Theodore, N. (eds.) (2002). *Spaces of neoliberalism; urban restructuring in North America and Western Europe*. London: Blackwell Publishing.

Castree, N. (2006). From neoliberalism to neoliberalization: Consolations, confusions and necessary illusions. *Environment and Planning A* 38, 1 6.

Dumenil, G. and Levy, D. (2004). *Capital resurgent: Roots of the neoliberal revolution*. Cambridge, MA: Harvard University Press.

Gill, S. (1995). Globalization, market civilisation and disciplinary neoliberalism. *Millennium* 24(3), 399 423.

Hall, S. (1988). The toad in the garden: Thatcherism among the theorists. In Nelson, C. & Grossberg, L. (eds.) *Marxism and the interpretation of culture*. London: Macmillan Education.

Harvey, D. (2005). *A brief history of neoliberalism*. Oxford: Oxford University Press.

Hindess, B. (2002). Neoliberal citizenship. *Citizenship Studies* 6, 127 143.

Larner, W. (2000). Neo liberalism: Policy, ideology, governmentality. *Studies in Political Economy* 63, 5 26.

Laurie, N. and Bondi, L. (eds.) (2005). *Working the spaces of neoliberalism*. London: Blackwell Publishing.

Mitchell, K. (2004). *Crossing the neoliberal line: Pacific rim migration and the metropolis*. Philadelphia, PA: Temple University Press.

Peck, J. and Tickell, A. (2002). Neoliberalising space. *Antipode* 34, 380 404.

Peet, R. (2003). *Unholy trinity: The IMF, World Bank and WTO*. London: Zed Books.

Sharma, A. and Gupta, A. (eds.) (2006). *The anthropology of the state*. London: Blackwell Publishing.

Relevant Websites

http://www.globalissues.org/TradeRelated/FreeTrade/Neoliberalism.asp
Social, Political, Economic and Environmental Issues.

http://www.alternet.org/tags/neoliberalism/
Online Journalism Community.

Neoliberalism and Development

W. E. Murray, Victoria University of Wellington, Wellington, New Zealand

The terms – neoliberalism and development – are contested: views on interaction between the two turn on one's political perspective. Neoliberal thinkers believe that the policies that have been diffused across the Global South based on their ideas have brought large benefits, especially in the economic sphere – although some would argue that there have been positive effects in the political and social fields too. There appears to be a mounting body of evidence, however, that suggests that neoliberalism has eroded development prospects, where development is defined as social change toward a situation where people can exercise their full human potential.

Neoliberalism, Globalization, and Development

Neoliberalism is an economic, political, and it can be argued, cultural paradigm, which is all but hegemonic at the governmental scale across the globe. Echoing back to early capitalist thinking, neoliberalism argues that governments fail inherently. Such state intervention in the economy causes crowding out and inefficiency leading the capitalist system to function less effectively, lowering competitiveness, and ultimately prospects for economic development. According to neoliberal thinking the most effective way to optimize welfare is through a privatized, deregulated economy that is open to trade and inward investment. This allows comparative advantage to work on a global scale, thus maximizing total global welfare. Neoliberal policy facilitates the linking of processes at different scales articulating the local into the global and vice versa, thereby stimulating greater economic openness and efficiency. In essence, neoliberalism promotes politicoeconomic Darwinism on a world scale, and the model thrives on the notion that only the fittest economic units survive and that the benefits of efficiency gains will eventually trickle down to all members of society.

Neoliberalism is a political ideology, promoting negative liberty (freedom to purse individual goals without intrusion from others), which seeks to optimize welfare through operation of what neoclassical economists called the 'invisible hand'. In the sense that neoliberalism is bound up as part of capitalistic individualism, it is a cultural paradigm, also. Early neoclassical political economy was clear about the central regulatory function of the state, while neoliberalism is far more mistrusting of its role. The culture of neoliberalism forms part of the wider regulation of the capitalist system and is tied to Western ideas of modernization, civilization, individualism, materialism, accumulation, and rationality. In terms of its relationship with theoretical perspectives on development it is borne of enlightenment 'doctrines of development', which have had a major impact on the Global South since the advent of colonialism and through the neocolonial/postcolonial modernization period since World War II.

Neoliberal ideas have fed into recent discourses of good governance, which ostensibly have to do with optimizing the rules and interventions that regulate and govern societies and polities, but arguably have had more to do with the advancement of free market economics, under the cloak of liberal democracy. There is no reason why good governance should have this particular connotation, some might argue that good governance is best exemplified in the 'developmental state'. For example, arguably the most impressive developmental progress historically has not come through free markets alone but through state intervention and strategic political economy reflecting popular aspirations. However, neoliberal thinkers link state developmentalism at best to inefficiency and at worst to corruption; thus, arguments for rolling back the state have come to dominate policy reform across the world.

Neoliberalism is intimately tied to 'globalization as currently practiced', that is to say, corporate globalization. It can be argued that the model that has taken hold since the 1980s is the most recent turn in the globalization agenda which has moved through two waves – colonial and postcolonial – which are further subdivided respectively into two phases, mercantilist and industrial and modernist and neoliberalism, respectively. Neoliberalism reduces barriers to economic flows which opens doors to investment and trade. This leads to the greater penetration of transnational corporations (TNCs) into the Global South and the articulation of peripheral economies into the global capitalist system, often under unfavorable conditions. Globalization has often been used as a synonym for neoliberalism by hyperglobalists who argue there is no alternative. These are really just arguments for a particular model, which is masqueraded as inevitable, that are designed to further the self interest of those that promote it. As geographers well know, there are many different possible globalizations and the neoliberal model can and is being resisted from the grassroots upward. The reign of neoliberalism and the spell it has cast on orthodox development thinking need not be everlasting.

Historical Roots of Neoliberalism – Laboratory Chile

The ideas of neoliberalism emanate, in large part, from the University of Chicago's school of economics and are associated with Milton Friedman and other monetarist economists – although the principles have been elaborated beyond supply side economics. The history of the diffusion of this specific type of economics is in many ways serendipitous, suggesting that the notion that 'there is no alternative' is fallacious. Following the Chilean coup of 1973 the military junta led by Augusto Pinochet had no idea of how to run an economy, but they knew that they wanted to reverse the Chilean 'road to socialism' of Salvador Allende's 'Frente Popular' which they had toppled. They called on the services of a number of right wing Chilean economists who had been trained at the University of Chicago during the late 1960s, some of whom had been sponsored by the anticommunist inspired, US backed funding program 'Alliance for Progress in the Americas'. They developed a radical restructuring model based on neoliberal principles and received advice directly from Friedman and economists close to him. The Chicago Boys, as they are now known in Chilean folklore (many of them survive till today and one, Joaquin Lavin, came close to being voted president in 1999 and 2005), recommended aggressive opening of the economy, devaluation, sharp reductions in public service expenditure, and privatization of most nationalized industries. In the event the dictatorship did not privatize copper, as this was important strategically and financially to them; Chile has the largest copper supplies in the world, and 10% of revenues were set aside for the military.

In a few short years Chile became the most open and 'free' economy in the world. This led to an inflow of foreign investment and, later, a surge in exports of primary products and nontraditional agricultural items destined for the Northern Hemisphere. These flows were largely controlled by foreign firms and often mediated through large economic groupings built in part from the elite families established in colonial times. Chile's macroeconomic 'miracle' came at great social, political, and environmental cost, and some have also questioned the economic performance of the model. Given the attendant cuts in social expenditure combined with the shock exposure to global forces, the socioeconomy was radically restructured. Small business, small farmers, the poor, and generally those most in need suffered grave consequences, furthermore, the natural environment was not accounted for in the new model and sustainability was threatened by the clamor to export vast quantities as quickly as possible. In terms of the model's stated objectives of lowering inflation, increasing exports and competitiveness the policy was a success and it was argued that eventually this success would trickle down to the masses if the economy were freed up sufficiently to allow this.

Based in part on the macroeconomic success of the Chilean model it was applied in the UK in 1979 under Thatcher, to the USA in 1980 under Reagan, and followed by New Zealand in 1984 as the post World War II welfare consensus was eroded by the recessionary impacts of the oil crisis. The diffusion of the policy across the Global South was made possible through structural adjustment policies (SAPs) applied to the periphery by financial institutions, primarily the International Monetary Fund (IMF) and the World Bank and later some aid donor agencies and the World Trade Organization (WTO). As explored below, the consequences were devastating to the lives of millions of people. It is instructive that in many cases in the 1980s, neoliberalism was brought in under military dictatorships – in democratic systems in poorer countries they may not have been politically sustainable.

Neoliberalism and the Political Economy of Global Development

The neoliberal agenda has supporters in many quarters, not least in big transnational business and rightist political parties and organizations that back it. In academia supporters tend to come from schools of business and management in which social issues are not always at the forefront. Proponents of neoliberalism often argue that the success of East Asian countries – breaking away from primary product production and into high technology manufacturing, services, and ultimately into the top ranks of the global economy – proves that the theory works. Others have countered that the East Asian model has involved market forces and processes, closely monitored, sequenced, and directed by developmental states that have intervened actively to create the conditions necessary for capitalist growth. In a way, there really is no such thing as pure neoliberalism – all economies are regulated to a degree and deregulation can be conceptualized a 'reregulation' toward lower levels of state involvement. There is a continuum between structuralist approaches to development and neoliberal approaches, becoming progressively less interventionist as we move toward the latter.

At the scale of the whole globe, the world economy remains mixed. Despite the fact that neoliberalism has been promoted as the solution for development ills in the South, Western countries themselves have continued to protect their agriculture and industry behind high tariffs and quota walls, especially in the European Union (EU), Japan, and the USA. In this sense, the world system exhibits neoliberal elements juxtaposed with highly regulated components designed to promote the interests of those that have most influence over the rules of the global economy. In this sense the world economy is in fact highly 'un free' for many, especially as it traps much of the Global South into a primary product producing role which offers very few prospects for long lasting social and economic progress, in spite of fleeting commodity booms such as that which has characterized the late 2000s.

The differences in levels of 'development' between nation states and individuals at the extremes is large and growing. Neoliberalism and the networks it creates in crease these disparities. The opportunities for those plugged into the net with the abilities to compete are substantial. However, the gaps are becoming wider and more are falling through them. This has the potential to lead to global conflict. There are thus geopolitical and economic, as well as moral arguments for reforming the system. Indeed, post September 11th 2001, a number of commentators were arguing that the root cause of the conflict was indeed development inequalities. While this vastly oversimplifies the factors behind the current global instability, it does point to the fragility of peaceful re lations under the neoliberal model which is so adept, and indeed is predicated on, concentrating wealth.

The Rolling Out of Structural Adjustment Policies and the Rolling Back of the State in the 1980s and 1990s

Neoliberalism has risen to global hegemony at a time when alternatives have been sidelined. The end of the Cold War helped solidify the place of the paradigm across the poor world and was seen as a triumph of rational and democratic values. From the mid 1980s onward neoliberalism eclipsed a range of alternative development approaches, including dependency and structuralism. The backbone of this diffusion has been the rolling out of SAPs and World Bank and IMF loans, which together with some aid flows were made con ditional on the implementation of neoliberal policy from the mid 1980s onward. Following the reform of the General Agreement on Tariffs and Trade (GATT) one of the principal backers of neoliberal reform has been the WTO, which now has the power to punish signatory countries that fail to deregulate and open their econ omies, even if they are very poor and industrially underdeveloped, lest social and economic intervention 'distorts' international trade.

The watershed in terms of the implementation of SAPs was the 1982 oil crisis, which led to the debt crisis (which it could be argued still persists). In the 1970s international lending to poorer countries, on the back of petrodollar deposits generated by the first oil hike in 1973, led to an escalation in debt. At the time, given the expansive money supply, interest rates were low. Following the second oil hike however, inflation and re cession bit into the Western economies and they used supply side policies, including interest rate rises to re store equilibrium. These rates filtered through the system and left the indebted nations with enormous debt bur dens – in Latin America in the 1980s debt outstripped annual GDP in many countries. In 1982 Mexico de faulted and in response a SAP was imposed on the country in return for a rescheduling and extension of debt. SAPs were heralded as the new path toward eco nomic self sufficiency and modernization, as opposed to the Keynesian influenced state developmentalism ap proach that had dominated since the end of World War II. In reality they were designed to restore stability to and expand the profits of the global financial system upon which the prosperity of the elite in the West and their associates in the South was, and remains, dependent.

The adoption was rapid across the Global South – after all, countries had little option but to heed the financial institutions that were in turn funded and driven by Western political and economic powers. By 1990, for example, every Latin American and most African coun tries had adopted neoliberal policies. In this way the World Bank, and its major backers gained neocolonial power over the economy policy of nation states. By the end of the 1980s a good deal of the Global South had signed up to SAPs, thus surrendering a substantial level of economic sovereignty to external powers with a very particular worldview.

The Nature of Neoliberal Development Models

The nature of SAPs has been relatively constant across time and space – leading to accusations of a 'one size fits all' approach to their design. Indeed, even former insiders at the World Bank such as former chief economist, Joseph Stiglitz, have argued this point. Essentially, SAPs con tained four cornerstones: (1) 'downsize' – reduce taxation and public expenditure to lower deficits and raise in centives for private enterprise; (2) 'privatize' – sell off state owned assets to introduce the profit motive into the economy; (3) 'deregulate' – reduce state involvement in order to maximize efficiency and allow the free market to find its natural equilibrium; and (4) 'globalize' – reduce protectionist measures, such as tariffs, to allow foreign investment, facilitate the transfer of technology, attain economies of scale in exports, and ultimately stimulate competitiveness on the global stage. In isolation, many of the objectives of SAPs are laudable. The unfortunate reality is that economies have often been launched into a global system in which the industrialized economies have had decades and centuries to evolve their economies; thus, it is enormously difficult for poorer countries to participate without becoming locked into low value added production as dictated by comparative advantage. In East Asia this comparative advantage was turned into competitive advantage through selective intervention. In most countries, however, adherence to the model has led to little structural economic progress.

If the impacts of neoliberalism in the West in the 1980s were harsh, they were devastating in the South. Poverty and inequality were, by definition already much higher in the periphery and the new policies exacerbated these divisions. It is recognized by both supporters and detractors that there are negative impacts when neoliberal policies are first applied. Austerity is required in order to put SAPs in place. Countries adopting SAPs will experience rising inequality, poverty, unemployment, and informal sector activity in the short term at the very least. Neoliberals have consistently argued that these are merely transitory problems of 'adjustment'. There is gathering consensus, even among some centrist neoliberals, that empirically this has not been observed. The toughest critics have argued that SAPs have often left countries worse off in terms of holistic development measures, including environmental impacts. Studies have shown that it is often women, children, and the poor that have suffered the most devastating consequences in terms of loss of livelihoods and increased vulnerability. Furthermore, these impacts have been shown to be long lasting. The continued marginalization of the developing world in an economic sense can be empirically illustrated as countries in Africa, and parts of Asia, Latin America, and Oceania continue to lag behind. Partly based on the empirically observed failure of SAPs there has been a shift away ostensibly at least from extreme neoliberalism toward what might be termed 'third way' policy in some financial institutions and some countries. The case of Latin America, explored below, illustrates the trajectory from extreme SAPs to 'neoliberalism with a human face'.

Neoliberalism and beyond in Latin America

Arguably, in Latin America, the impacts of neoliberalism have been felt more acutely than anywhere else. Given the depth of the debt crisis in the early 1980s, SAPs were almost universally adopted by the middle of the decade. The combination of the austerity measures accompanying the SAPs and the debt burden itself led to a 'lost decade of development' whereby previous gains in living standards were halted or reversed. Across the continent income growth fell to 1.4% on average, compared to 5% for the previous decade, becoming negative in places such as Argentina and Bolivia. Unemployment and poverty rose and inflation skyrocketed in many countries. Due in part to this economic instability dictatorships maintained or grabbed the reigns of power in many places.

Based on SAPs, Latin American countries were encouraged to exploit their comparative advantage which was and is almost exclusively found in primary and agricultural production for export. Despite industrialization gains in the largest markets of Brazil and Mexico based largely on the ISI policies of the 1960s, this policy compounded the region's historical role as a resource periphery for the global economy. This had regressive developmental impacts including declining long run terms of trade, Dutch disease (overvalued exchange rates), the evolution of enclave economies, and vulnerability to commodity boom and bust cycles among other things.

Neoliberal restructuring had a particularly profound effect on agriculture in Latin America. Most of the Latin American economies have relative cost advantages in agriculture and neoliberalism exploits these. There was a major shift away from domestic, food and traditional production in the 1980s and 1990s, toward export oriented, exotic, and nontraditional production. Forestry, fisheries, and agriculture formed the backbone of this change. This has had profound impacts in rural localities across the continent as the signals of the global market are transmitted into local spaces. Some such trends provided export opportunities for farmers. For many rural dwellers, however, globalization in the countryside has had negative impacts including loss of land, concentration of ownership, increased vulnerability to external trends, and environmental nonsustainability.

Following the end of the Cold War and the implicit, and sometimes explicit, support of the US and its allies of the neoliberal, dictatorships of the 1980s declined. This has led to the watering down of the neoliberal agenda in the continent and the revival – to an extent – of regional cooperation efforts. Throughout the 1990s, given that electorates now had the vote, governments had to make concessions to social policy in ways which some have argued hark back to the structuralist period of the 1960s and 1970s. Designed in order to ensure the political sustainability of the model, this has been termed 'neoliberalism with a human face' or 'neostructuralism'. After the Asian Crash of 1997, an economic downturn occurred which led eventually to particularly harsh outcomes in Argentina, Brazil, and Uruguay. The left and center left blamed this on IMF policies and, more broadly, the neoliberal Washington Consensus. Neoliberals blamed internal factors, especially corruption. This crisis solidified the critique of neoliberalism at all levels of society from the grassroots upward and resistance movements became increasingly heard across the continent.

The coming together of popular and political concern about the impacts of neoliberalism has led to a shift to 'center left' politics in many countries from the early 2000s, and the victory of presidential candidates with centrist and left leaning tendencies including among others, Lagos (2000–06) and Bachelet (2007–) in Chile, Toledo (2001–06) in Peru, Morales (2006–) in Bolivia, Lula de Silva in Brazil (2003–), Chavez in Venezuela (1999–), Vazquez (2005–) in Uruguay, and Kirchner (2003–08) and Fernandez de Kirchner in Argentina (2008–) among others. It is certain that in order to win elections in Latin America, it is necessary to adopt a post neoliberal stance, and this is bound up in part with a desire to distance the continent from the currently highly unpopular United States. Ironically, the USA has meddled politically in Latin America less than ever over the last decade and thus the space to distance countries from it has been created. In the Argentinean election of 2007, the eventual winner, the center left Peronist influenced Justicialist candidate, Cristina Fernandez de Kirchner, campaigned using anti IMF television advertisements which played on nationalistic feelings.

Whether the shifts in policy in Latin America can be termed 'leftist' or even 'center leftist' has been called into question. There can be little doubt that the policies of Chavez's Bolivarian revolution, involving nationalization and poverty reduction, fall further to the left than many of the others, but this has been made possible by the country's ample supply of oil and thus influence over the global economy. Other ostensibly leftist governments such as that of Morales in Bolivia or Ortega (2007–) in Nicaragua could be viewed as populist. In other cases governments are decidedly centrist by historical and global measures, and despite anti IMF words of leaders such as Kirchner, they are still very accommodating of the global capitalist system. The free market remains a central pillar, and as a consequence the general model in Chile, Argentina, Uruguay, and Brazil is, in fact, further right than the structuralist inspired policies of the 1960s. Renovated left leaders (*los renovados*) have argued that global conditions have changed, ISI policies failed, globalization is here to stay, and that the market can be tempered and optimized. Critics might charge that governments across Latin America talk anti neoliberalism while still pursing many policies that have been inspired by it.

Although in a small handful of economies there has been a shift away from primary product export reliance, in most there has not. The resource curse thesis argues that resource endowment can, under certain conditions, actually be an impediment to development progress – in contrast to more traditional neoclassical and neoliberal ideas. This is the most damaging legacy of the years of pure neoliberalism in the continent and one which the current 'neostructural' model – whether we define it as a continuance of neoliberalism or something else – is very unlikely to solve. The spectacular yet ultimately non sustainable short term windfalls that are accompanying the commodity boom of the mid 2000s may actually delay the necessary strategic diversification of the Latin American economies unless the lessons of economic history in this respect are heeded.

The Future of Neoliberalism – Resistance and Reinvention

The Global South has been the site of intense struggles over neoliberalism, and resistance movements abound. There is also a rich vein of development theory that has emanated from the periphery – including dependency and structuralism that still has much to say about the current situation despite the disdain that has been piled on it by neoliberal orthodox economists. There have been international institutional attempts at the government scale to resist and reform including, it could be argued, the formation of trade blocs in the South such as Mercado Comun del Sur (MERCOSUR) and South African Development Community (SADC) – although it could be posited that these merely perpetuate and diffuse neoliberalism further. Other examples include the continued policies of Organization of Petroleum Exporting Countries (OPEC), collective action by developing countries against the hypocrisy of Western agricultural policy, and more recently, efforts led by Venezuela's Chavez to build anti neoliberal alliance across the Southern Hemisphere incorporating countries that are out of favor in the West such as Iran and Cuba. Whatever one thinks of Chavez's attempts in this regard they signal a mounting consensus in the South against neoliberal inspired ideas and the dominance of the Organization for Economic Cooperation and Development (OECD) countries.

Over the recent past struggles which go back many decades, have coalesced at the grassroots – at least in the eyes of the popular media – to form the antiglobalization movement. To the extent that such a movement can be discerned it is motivated by a common abhorrence to neoliberalism – or 'globalization as currently practiced'. There has been growing popular dissent and the formation of transnational nongovernmental organizations (NGOs) and new social movements intended to resist neoliberalism. This has been particularly prevalent in Latin America, due to the profound impacts of neoliberalism there, its association with rightist dictatorships, and the history of popular struggle. This is embodied most obviously in the Zapatista movement, which some credit as the foundation of the antiglobalization movement. The lines between antiglobalization, anti neoliberal, anti Western, and anti US 'movements' have become increasingly hard to discern. Popular struggles have become far broader than opposing neoliberalism, focusing most recently on the Iraq War, as well as more general antipoverty and pro environment causes. At the root of many such movements there is a disdain for free market economics and US led economic and political dominance which, in the opinion of the protestors, has brought regressive impacts across most of the world and underlies specific geopolitical moments such as the War on Terror and the so called fight for democracy in the Middle East.

In the case of lending institutions there has been a marginal tinkering with the policies of the Washington Consensus partly in response to the concerns of new social movements and political change. The World Bank's approach to development is ostensibly far more rounded than it once was, focusing on broader sustainability in a social and environmental sense, drawing attention to poverty and inequality as embodied in the Poverty Reduction Strategy Papers, and promoting the notion of good governance. Notwithstanding, the major institutions such as the WTO and the IMF still pursue an essentially neoliberal doctrine of development. It could also be argued that underlying the World Bank's ostensibly holistic policy is the core idea that it is free markets, free trade, and competition that are deserving of promotion; the road to this might be slightly altered but the eventual destination is the same. Such shifts then have been largely rhetorical, barely hiding the fact that at the core, development finance is about expanding the influence of capitalism across the globe, opening the 'emerging' economies as places for business. This might well bring benefits for some sections of society, but ultimately it is tied up with discourses concerning the inevitability of globalization and the need for poorer countries to embrace it or to sink.

Do organizations such as 'Jubilee 2000', events such as 'Live 8', and endeavors by the UK government to write off the debt of the poorest nations point toward a new popular and parliamentary politics when it comes to neoliberalism and development? Will neoliberalism continue to dominate and thus foster corporate globalization or is it being meaningfully reformed? Can we talk of a shift toward a post neoliberal development policy? The shifts in Latin America and beyond indicate that extreme neoliberalism is no longer popular in the Global South itself, not that it ever was beyond the elites. Indications are that despite promising superficial signs the underlying approach remains much the same. Neoliberalism is both facilitated by and facilitates globalization. It has allowed the penetration of multinational companies who locate parts of the commodity chain in peripheral areas. Neoliberalism has had the effect of widening income gaps at virtually all scales of analysis and the world economy thrives on this unevenness. It may be that the powerful interests that have colluded to make neoliberalism dominant – big business, conservative politics, and the institutions that represent their aspiration such as the WTO – are difficult to defeat. This raises questions over what we mean by development, and who should be responsible for defining it and designing policy. Development geographers have a duty to continue to map the impacts of neoliberalism, but more crucially they need to engage directly in the formation of viable alternative models if the inequity and nonsustainability of neoliberal development practice is to be challenged.

See also: Anti-Geopolitics; Civil Society; Colonialism, Internal; Debt; Dependency; Desertification; Development I; Development II; Famine; HIV/AIDS in Developing Countries; Intermediate Technology; Neoliberalism; Orientalism; Postcolonialism/Postcolonial Geographies.

Further Reading

Auty, R. M. (1995). *Patterns of Development: Resources, Policy and Economic growth*. London: Edward Arnold.
Gwynne, R. N. and Kay, C. (2000). Views from the periphery: Future of neoliberalism in Latin America. *Third World Quarterly* 21(1), 121 156.
Gwynne, R. N. and Kay, C. (2004). *Latin America Transformed: Globalization and Modernity,* (2nd edn.). London: Arnold.
Harvey, D. (2007). *A Brief History of Neoliberalism*. Oxford: Oxford University Press.
Kay, C. (1989). *Latin American Theories of Development and Underdevelopment*. London: Routledge.
Kelsey, J. (1995). *The New Zealand Experiment: A World Model for Structural Adjustment?* Auckland: Auckland University Press.
Murray, W. E. (2004). Neocolonialism. In Forsyth, T. J. (ed.) *Encyclopedia of International Development*. London: Routledge.
Murray, W. E. (2006). *Geographies of Globalization*. London: Routledge.
Peck, J. and Tickell, A. (2002). Neoliberalizing space. *Antipode* 34, 380 404.
Petras, J. F. (1999). *The Left Strikes Back: Class Conflict in Latin America in the Age of Neoliberalism*. Boulder, CO: Westview Press.
Potter, R. B., Binns, T., Elliot, J. A. and Smith, D. (2004). *Geographies of Development* (2nd edn.). Harlow: Prentice Hall.
Power, M. (2003). *Rethinking Development Geographies*. London: Routledge.
Preston, P. W. (1997). *Development Theory*. Oxford: Blackwell.
Smith, N. (1984). *Uneven Development*. Oxford: Blackwell.
Stiglitz, J. E. (2002). *Globalization and Its Discontents*. London: Penguin Press.
Storey, D., Bulloch, H. and Overton, J. D. (2005). The poverty consensus: Some limitations of the popular agenda. *Progress in Development Studies* 5(1), 30 44.

Relevant Websites

http://www.hdr.undp.org
Human Development Report, UNDP.
http://www.econ.worldbank.org
World Development Reports, World Bank.

Neoliberalism, Urban

W. Larner, University of Bristol, Bristol, UK

Glossary

Creative City Approach to urban policy that privileges the role of culture and creativity in new forms of economic development.

Gentrification Process of urban neighborhood change in which relatively affluent residents displace lower income groups.

Neoliberalization Analytical approach that emphasizes the contested political processes associated with neoliberalism.

Neoliberalism Shorthand term referring to the new emphasis on markets, minimal states, and individual choice in contemporary political-economic formations.

Revanchist City Urban politics that combines punitive approaches with moral discussions about public order.

Shadow State New relationship between state and voluntary sector in which the latter is increasingly involved in developing policy and service delivery.

Urban Entrepreneurialism Urban economic development strategy that involves political coalitions between local states and private sector actors.

Introduction

Over the last decade, increasing numbers of geographers have engaged in debates about neoliberalism. Like their colleagues elsewhere in the social sciences, they have been interested in examining the implications of new political emphasis on markets, minimal states, and individual choice as a means of ensuring economic and social well being. Of particular interest to geographers have been the political economic and sociospatial consequences of this new political formation. It is argued that whereas Keynesian welfarism saw the nation state as the primary basis for economic, social, and political processes, neoliberalism has been associated with the contradictory rescaling of these processes to both supra and subnational scales. It is in this context that urban geography has become the disciplinary locus for many geographical accounts of neoliberalism, and changes in the nature of contemporary cities the focus of many analyses. In the most general terms, the argument is that the new market oriented formulations associated with neoliberalism have given rise to a new role for cities in managing the relationships between global flows and networks and local economies and societies. In turn, this has had major implications for urban governance, the urban landscapes, the built environment, and urban subjectivities.

Accounts of what are now commonly called 'neoliberal cities' build on early work by David Harvey and Manuel Castells during the late 1970s in which they identified the strategic roles that cities play in capitalist societies, rather than seeing contemporary urban spaces in simply empirical terms. This work stimulated the rise of a new research agenda in which Marxist and neo Marxist perspectives began to be deployed to theorize the city. Correspondingly, in the early 1980s, there was a considerable resurgence in urban geography, as geographers began to explore the shift in urban governance from the redistributive policies associated with Keynesian welfarism to a new emphasis on generating economic growth by encouraging competitive markets and entrepreneurial activities. These new theoretical perspectives reshaped analyses of urban development, drawing attention to both the broader structural processes that were altering relationships between cities and nation states, and the way in which the class based interests of hegemonic urban actors influenced these processes. Today, many geographical analyses continue to interrogate the relationships between political economic restructuring and new forms of urban governance, including the transformation of policies and programs, the constitution of new spaces and subjects, and the rise of city based resistance struggles.

For many urban geographers, neoliberalism has emerged as a useful umbrella term to describe these changes in the urban environment, in particular, the new emphasis on international competitiveness, marketization, and economic growth, and the effects this has had on different groups within the city. The consequence is that discussions about topics as varied as de industrialization, urban entrepreneurialism, the informational city, global cities, gentrification, and sociospatial polarization are all now likely to be framed by macro accounts of neoliberalism. However, like the use of the term neoliberalism in geography more generally, this is a *post facto* argument. It wasn't until the late 1990s that neoliberalism began to emerge as a broad term to describe new forms of urban governance involving market based political formations. Indeed, the landmark text is Neil Brenner and Nik Theodore's *Spaces of Neoliberalism*, which was not published until 2002. Moreover, this text explicitly focuses on urban restructuring in North America and Western Europe. Today, however, increasing numbers of urban geographers are drawing on the broader literatures of neoliberalism in their efforts to analyze widespread shifts in urban governance from welfarism to entrepreneurialism, from government to governance, and from redistribution to recognition, in cities across the world.

Neoliberalism or Neoliberalization?

In geographical accounts, neoliberalism is a critical term that is most often understood to be about a political ideology that privileges markets and the role of indi viduals in ensuring economic and social well being. The origins of this formation are usually traced to the post war period and the writings of philosophers and econo mists such as Hayek and Freidman. It is then traced through the national experiments of the 1980s and 1990s, particularly the political programs known as Thatcher ism and Reaganomics, which involved a critique of interventionist economic policies and the welfare state. The new approach was made manifest in regulatory frameworks that privileged the role of markets, and fiscal austerity and privatization programs designed to shrink the public sector. Neoliberalism is now used as a general descriptor for these market based institutional shifts and policy realignments. However, whereas classic liberalism saw markets as naturally occurring phenomenon, neoli beralism is premised on the idea that the conditions can be created that will encourage people to act in a more marketlike fashion. Neoliberalism is thus both an analysis of the world and a view of how it should work. It em bodies assumptions about the appropriate role of the state, markets, and individuals. It is a political discourse that both represents and constitutes economic and social relations in particular forms, and has transformed dom inant political imaginaries over the last two decades.

Geographers have made important contributions to the debates about neoliberalism by distinguishing be tween neoliberalism as a universalizing ideology and the specificities of 'actually existing neoliberalism' as a lo cated and contingent phenomenon. More recently, these discussions have been complemented by the term neo liberalization which emphasizes the contested nature of political economic processes in explaining the turn toward 'more market'. These conceptual distinctions have allowed geographers to do what they do best – examine historically and geographically specific processes in par ticular places to explain why, from Johannesburg to Moscow, neoliberalism now appears to be the new common sense in urban politics. These analyses show that the state is now only one of many institutions in volved in urban governance, and that a raft of new entities has emerged to take on powers and responsi bilities once provided more centrally. These entities range from the corporate business partners involved in public–private partnerships that provide infrastructural services, through to the nongovernmental organizations and social movements who have become increasingly involved in social and environmental policies and service provision. Moreover, this new approach to urban gov ernance is not simply premised on encouraging more markets in order to facilitate the processes of economic exchange; rather, it involves a wide range of economic, social, political, and environmental processes that are premised on negotiated agreements among diverse urban organizations and actors, making manifest a new em phasis on self managing, autonomous, urban subjects.

Theoretical Approaches to Neoliberalism

Not surprisingly, given the intellectual and disciplinary legacies identified above, the literature on urban neoli beralism continues to be dominated by Marxist and neo Marxist political economy approaches, particularly those associated with regulationist tradition. In these accounts, neoliberalism is understood to be a set of political shifts associated with the changing nature of the capitalist economy. It is argued that the increasing globalization of economic processes that followed the end of the long boom in the 1970s reflected the efforts of capitalists to find new ways to restore profitability. In turn, the growing tendency to organize production across national bound aries and the rise of globalized financial systems forced governments to abandon their commitment to full em ployment and an inclusive social welfare system in order to focus on enhancing economic efficiency and inter national competitiveness. Urban geographers have linked these changes to shifts in the governance of cities. They argue that in the context of economic globalization, cities have become engines of economic growth as well as centers of innovation, giving rise to a new political emphasis on market activities and individual enterprise.

More conceptually, these scholars identify a shift from the nation state formations associated with Keynesian welfarism to the heterogeneous scalar formations asso ciated with neoliberalism. In these theoretical debates the work of Bob Jessop, supplemented more recently by that of Neil Brenner, has been influential, particularly as the 'urban question' and the 'scale question' have become increasingly intertwined. It is argued that the move be yond methodological nationalism is associated with a new role for cities as sites of economy, politics, and so ciety. This new scale is socially constituted, produced by the spatial imaginaries and political strategies of a wide range of actors. As the politics of scale became the focus of attention in these discussions, so too did questions of scale jumping, used to refer to the ways in which political actors appeal both upward and downward in their efforts to broaden the scope of political struggle, become more central. Finally, as the debate between those who understand scale as the new spatial ontology structuring urban politics and those who would privilege the role of networks has come to dominate debates about the nature of the neoliberal city, so too has social movement theory entered this terrain and greater attention has been paid to questions of the discursive frames and political coalitions that underpin efforts to resist neoliberalism. That said, in the vast majority of these accounts, the state, aligned with corporate power, is understood to be the primary archi tect of neoliberalism.

However, this political economy scholarship is not the only approach to urban neoliberalism. Also influential has been the neo Foucauldian governmentality literature. Developed out of the late writings of French philosopher Michel Foucault, this work is concerned with theorizing political power as an assemblage of discourses and techniques. There are three aspects to these discussions. First, this literature focuses on problematizations – an examination of the historically and geographically spe cific ways in which urban problems are defined and so lutions imagined. These discussions have highlighted a new role for the state as a partner in government whose role is to facilitate and support the self governing of both entities and individuals. Second, this literature has drawn attention to the specificity of the objects and subjects of governance identified and constituted through urban policies and programs. Of particular interest here has been the shift away from conceptions of urban govern ance premised on universalizing conceptions of society, toward multiple and heterogeneous conceptions of community, and the new emphasis on urban citizens as active, entrepreneurial subjects. Third, increasing atten tion has been paid to the strategies and techniques used in urban governance. Of particular interest are the rela tively mundane techniques – such as land use plans, zoning, and community consultations – through which particular objects and subjects of governance are consti tuted. These analyses have also identified the new em phasis on 'governing at a distance' through budgets, targets, standards, and objectives, all overseen by moni toring and audit. Correspondingly, it is argued that there is a new relationship between expertise and politics as accountancy, audit, enumeration, calculating, and evaluation take on a new and heightened emphasis.

Interestingly, given the important role feminist geographers played in debates about gentrification during the 1980s and 1990s, they have been less visible in recent debates about urban neoliberalism. In part, this may re flect the gender politics of the so called 'neoliberal era' in that, both men and women are now involved as paid workers in market based politics of gentrification and consumption led urban redevelopments. Gay consumer spaces are also part of the new urban renaissance and have been rescripted as sites of spectacle and con sumption in, for example, Paddington in Sydney and Soho in London. Recently however, a new generation of urban geographers concerned with the gendering of neoliberalism have begun to interrogate how the 'work, life, play' rhetoric of city officials is premised on de gendered understandings of the urban subject and over looks the ongoing demands of social reproduction, including housework, caring, and emotional labor. At tention has also been paid to the gendered injustices that have arisen as a consequence of neoliberal policies such as the deliberate targeting of sex workers as an 'un desirable' group in newly gentrified areas of the inner city. More conceptually, feminist theory – together with post structuralist and postcolonial readings of citizenship – has informed important analyses of the 'making up' of urban citizen subjects, for example, in work associated with the journal *Citizenship Studies*. In these accounts, attention has focused on the way in which market based strategies for managing different popu lations are linked to changes in political subjectivities among those affected by such policy shifts.

Geographical Studies of Neoliberalism

Just as theoretical influences have become more diverse in the debates about urban neoliberalism, so too has the focus of empirical studies. Early accounts focused on the corporate led revival of central city areas based on megaprojects in the areas of entertainment, tourism, and retail. It was shown how run down inner city areas were being revitalized through private sector investment, particularly through activities that involved conspicuous consumption in the form of shopping centers, cultural venues, and tourist attractions. Examples of this 'urban entrepreneuralism' were widespread, including the re vitalization of London's docklands and Auckland's waterfront. The aim of these developments, it was argued, was to recreate the city as both a focus of market oriented economic growth and a space of elite consumption. One result was increased competition between cities and a new emphasis on what has come to be called 'place marketing' as cities attempted to reposition themselves in the globalizing markets for both capital investment and people. Today, the legacy of this new emphasis can be clearly seen in the proliferation of international league tables that compare cities in the search of the 'best place to do business' or the 'most livable city'. Significantly, these comparisons are no longer to other cities within the same nation state; rather, they are to other cities in the global 'space of flows' (another term for which we should thank Manuel Castells).

One intellectual site in which the politics of the neoliberal transformation has received particular attention is in accounts of urban regimes. The focus of these analyses is on a ‘new urban politics’ in which political elites engage in the courting of private investors. Efforts have been made to identify the class coalitions that come together around efforts to develop enterprise zones and public–private partnerships, and the relationships between city officials and private sector actors through which financing, construction, and management of these new spaces takes place. Initially these accounts understood the new involvement of the private sector in urban governance as representing the ‘rolling back’ of the state, but now it is increasingly recognized that these changes are more usefully understood as involving the political, institutional, and geographical reorganization of the local state. This encouraged a rethinking of changes in urban governance during the 1980s and 1990s. Whereas the 1980s was a period of lean government, in which cost cutting, cutbacks in public services, and the privatization of urban infrastructure was the primary focus of what was then called the entrepreneurial city, more recently, greater attention has focused on the creation of new ‘flanking mechanisms’ in which efforts are made to displace the worst effects of ‘tooth and claw’ neoliberalism. Jamie Peck and Adam Tickell distinguished between ‘roll back’ and ‘roll out’ neoliberalism in their efforts to understand this shift – a classification that is now entrenched in the literature on urban governance.

These accounts of shifts in the nature of urban governance have segued into a focus on what is sometimes described as the shift from government to governance. In these debates, urban geographers have engaged more actively with their colleagues in political science and policy studies. There are two aspects to this literature. First, there are important discussions that show how the business led networks that were established to recreate the conditions for economic growth have given rise to new clusters of economic activity, involving not only relationships between local states and firms, but also between educational institutions and infrastructural developments. New economic spaces and activities are a corollary of these relationships; for example, the recent proliferation of business incubators as a means of growing small entrepreneurial firms makes manifest these new relationships between state, business, and educational institutions. The conclusion of these studies is that we are seeing networked forms of entrepreneurial modes of governing premised on the need to insert urban activities into global flows and networks.

Second, a parallel and related discussion has developed around the role of social movements and not for profits organizations in urban governance. As budgetary pressures have tightened and local states increasingly contracted out services, a new role has emerged for the community sector in social policy and provisioning. In the early 1990s Jennifer Wolch called this phenomenon the ‘shadow state’. In the years since she made this observation, new relationships between state and so called ‘third sector’ organizations have proliferated. Associations, advisory groups, forums, regeneration agencies, social enterprises, and partnerships have emerged as forums in which urban governance decisions are made. The extensive literature on the relationships between neoliberalism and these ‘bottom up’ organizations has revealed not only the ways in which these new relationships alter all those involved, but also how they are giving rise to new urban spaces and subjects premised on heterogeneous, overlapping, and multiple conceptions of community. Debates over whether or not social movements and third sector organizations have simply been co opted to neoliberal agendas, or whether new spaces for politics emerge through such initiatives, continue apace.

These changes in the nature of urban governance link into a second set of discussions that now occur under the umbrella of neoliberalism. Over the last two decades there has been considerable attention paid to the relationships between neoliberalism and increasing socioeconomic and sociospatial disparities, and their manifestations in the urban setting. There are a number of features to these debates. First, debates about neoliberalism have reframed discussions of gentrification. Initially, gentrification was assumed to be confined to so called global cities such as New York and London in which increasing mobility had given rise to a new constituency of financiers, technocrats, and other transnational subjects who were able to purchase high price residential property in the inner city. However, over the last two decades, it has become clear that gentrification is a much more generalized process in which middle class inhabitants and the accoutrements they demand have underpinned the redevelopment of massive swaths of Western cities through new forms of housing, retail precincts, and consumption spaces. It is widely accepted that this process is associated with high levels of exploitative real estate development, and has had diverse geographical implications.

Second, attention has turned to the urban inhabitants who have been displaced by gentrification, increasing property prices, and new expectations of neighborhoods. Neil Smith’s description of the ‘revanchist city’ has been particularly influential in this respect. He explicitly associates the rise of a divided city with the neoliberal repudiation of welfarism and concerted efforts to expand programs premised on individual responsibility and self reliance. While his account was developed to explain events in New York City during the 1990s, analyses developed from this work have explored the impacts of neoliberalism in Western cities more generally. They show that urban pro growth agendas inevitably intensify social and territorial inequalities. The growth of these inequalities explains phenomenon such as structural unemployment, the rise of workfare, and the use of day laborers in major US cities. The geography of inequality has also been addressed; various studies have shown how groups such as squatters, 'squegee kids', street people, and sex trade workers have been targeted as 'disorderly populations' and forced to more marginal spaces in the city. Efforts to 'clean up' the city are also associated with new strategies of policing and surveillance involving CCTV cameras and the privatization of formerly public spaces. The consequence, it is argued, is that contemporary Western cities are becoming spaces of consumption premised on the growth of 'spatial apart heid' as poor, unemployed, racialized, and homeless people are banished to peripheral housing estates, far flung suburbs, or ghettos. The result is a new geog raphy of exclusion and fear based on race, gender, class, and sexuality.

Finally, cities feature in the literature on neoliberal ism in accounts of contemporary subject formation. The argument is that cities are replacing nation states in the construction of social identities. This is not just an analysis of the decline of the social subjectivities associated with welfarism and social citizenship; it is also an argument that new forms of urban governance are premised on an image of the urban citizen as calculating, self responsible, autonomous, and un encumbered. In the discussion above, we have already seen that new forms of urban governance are premised on a specification of the urban subjects as the discerning consumers of ever increasing numbers of goods and services, responsible for their own destiny. So too are the less well off seen as responsible for their own misfortune; the unemployed, the homeless, and the poor are increasingly seen as responsible for their own fates. Correspondingly, the consequence of neoliberal efforts to 'govern through freedom' is not only a new politics of inclusion and exclusion, but also new forms of citi zenship and new ways of being political are also emerging.

It is this last point that opens up the analysis of neoliberalism and counters the pessimistic vision that emerges in so many of the discussions above. Whereas it is agreed that cities play a key role in neoliberal re structuring, it can also be argued that they have been central to both resistance against neoliberalism and sites of alternative practice. For example, recent years have seen a concerted effort to involve community members into decisions about social service provision and neigh borhood revitalization. Whereas the literature on neolib eralism has focused on the 'rolling back of the state', a longer historical view might see the shift to governance as a move away from the top down planning agendas associated with urban renewal in the 1950s and 1960s toward a 'bottom up' strategy in which urban citizens are now able to make meaningful decisions about different aspects of their lives. This is not a straightforward pro cess; these new efforts to govern mobilize people as entrepreneurs and consumers as well as neighborhood members and residents. There is also the risk that this new approach reasserts the neotraditional sense of place and community that is often associated with the rise of gated communities variously based on lifestyle, prestige, and security. What is clear, however, is that the use of community to redraw subjectivities and create new entrepreneurial modes of citizenship is creating multiple and contradictory effects. How these articulate with the constitution of global labor forces drawn into the city through both paid work and social reproduction, giving rise to what Aihwa Ong has recently called 'states of exception', remains to be explored in the literature on urban neoliberalism.

Conclusion

Cities are understood to be the crucibles in which political and ideological strategies associated with neoliberalism are invented and experimented with. Consequently urban neoliberalism has recently been the focus of considerable attention among geographers. From early discussions of grand scale projects premised on marketization to the more recent developments associated with new under standings of culture and the 'creative city', neoliberalism has proved a useful label to capture key features of con temporary urban processes. There are, however, growing difficulties and significant silences in this literature.

First, neoliberalism is too often treated as a unified totalizing set of discourses and practices. One con sequence is that too much is now being included under the label of neoliberalism. We hear surprisingly little about the other modalities of power in contemporary cities – disciplinary, pastoral, authoritarian – and when we do, these are too often reduced to neoliberalism, as is the case in discussions of new forms of surveillance. Important national variations being obscured as neolib eralism interconnects with, and sometimes contradicts, older political traditions from both left (social dem ocracy) and right (neo conservatism) of the political spectrum. There are also other political discourses at play in contemporary cities. The discourses of environ mentalism – energy efficiency, sustainable development, land remediation – cannot be easily reduced to market based agendas. Even the rise of recent accounts of neo liberal natures show that market ways of thinking about the natural world articulate with longer standing debates about sustainable cities. Why don't we hear more about these articulations and tensions in the literatures on new forms of urban governance?

Second, the continued emphasis on political economy approaches also means that we don't know nearly enough about how people experience cities on an every day basis. In contemporary cities there is a whole raft of emotions and experiences which may or may not have anything to do with neoliberalism. How are cities lived? What is the 'affectual economy' of different cities? How do mundane practices and technologies shape urban environments? How do the rhythms of the city vary according to gender, sexuality, and class? And, with a few notable exceptions such as the work of Katharyne Mitchell on Vancouver, there is definitely not nearly enough about racialization and transnationalism in the literatures examined above. Relatedly, too many of these accounts of the neoliberal city are premised on the experiences of particular Western cities. It is striking that the emerging literature on South African, European, and Latin American cities often frames itself with reference to other literatures and debates. Do these cities also experience urban neoliber alism? With what consequences, and for whom?

Finally, and perhaps most importantly, why are cities now the dominant analytical lens for geographical ac counts of neoliberalism? Why is it that commentators such as myself, who have never formally worked on cities, increasingly find ourselves drawn into debates in and around urban geography? Why are economic geography and urban geography increasingly blurring? What does this say about the status of the city as an object of ana lytical enquiry? Is this because cities have been at the forefront of neoliberalism experimentation? Is it because of the growing importance of cities in the global economy? Or is it simply that cities are major concen trations of population and therefore register the impacts of neoliberal processes whose origins are actually else where? Is there actually anything distinctive about urban neoliberalism?

See also: Gentrification; Marxism/Marxist Geography I; Neighborhoods and Community; Urban Policy.

Further Reading

Brenner, N. and Theodore, N. (2002). *Spaces of Neoliberalism: Urban Restructuring in North America and Western Europe*. Oxford: Blackwell.

Castells, M. (1977). *The Urban Question*. Cambridge, MA: MIT Press.

DeFilippis, J. (2004). *Unmaking Goliath: Community Control in the Face of Global Capitalism*. New York: Routledge.

Harvey, D. (1978). The urban process under capitalism: A framework for analysis. *International Journal of Urban and Regional Research* 2, 101 131.

Lees, L., Slater, T. and Wyley, E. (2007). *Gentrification*. New York: Routledge.

Mitchell, K. (2004). *Crossing the Neoliberal Line: Pacific Rim Migration and the Metropolis*. Philadelphia, PA: Temple University Press.

Ong, A. (2006). *Neoliberalism as Exception: Mutations in Citizenship and Sovereignty*. London: Duke University Press.

Rose, N. (2000). Governing cities, governing citizens. In Isin, E. (ed.) *Democracy, Citizenship and the Global City*, pp 95 109. New York: Routledge.

Sites, W. (2003). *Remaking New York: Primitive Globalization and the Politics of Urban Community*. Minneapolis, MN: University of Minnesota Press.

Smith, N. (1996). *The New Urban Frontier: Gentrification and the Revanchist City*. London: Routledge.

Wolch, J. (1990). *The Shadow State: Government and Voluntary Sector in Transition.* New York: The Foundation Centre.

Network Analysis

M. J. Kuby, Arizona State University, Tempe, AZ, USA
T. D. Roberts, University of Colorado at Boulder, Boulder, CO, USA
C. D. Upchurch, University of South Carolina, Columbia, USA
S. Tierney, University of Denver, Denver, CO, USA

Glossary

Arc A connection in a network, also known as link or edge.
Complex Network Theory The study of the structure and evolution of interaction networks in which the formation of new linkages is a probabilistic function of the degree of the node.
Connectivity Matrix A matrix with 1's indicating a direct link between nodes and 0's indicating no direct link.
Degree of a Node The number of links connecting to a node.
Graph Theory A set of methods for analyzing the connectivity of networks and nodes.
Mathematical Programming A family of techniques for solving constrained optimization problems.
Network A set of points connected by a set of linkages, also known as a graph.
Node A junction or start/end point in a network, also known as vertex.
Planar Network A network in which arcs cannot cross without intersecting.

Introduction

'Network' is a general term for physical infrastructure or patterns of interaction that can be represented as a set of points connected by a set of linkages. The points are known as nodes or vertices, and can represent origins, destinations, and junctions. The linkages are known as links, arcs, or edges, and represent connections of some kind among the points. Links may be defined as one way (in a directed network) or two way (undirected). The terms 'vertices' and 'edges' generally are used together while 'nodes' are paired with 'links' or 'arcs'.

Networks can be planar or nonplanar. In a planar network, with all arcs on the same plane, a node must be placed at all intersections because two arcs cannot cross without intersecting. In nonplanar networks such as air line or friendship networks, arcs may cross without intersecting at a node.

In the field of geography, the term 'network' is highly flexible and encompasses both physical infrastructures over which transportation and communication occurs, as well as patterns of interactions themselves. Physical in frastructure networks can be composed of road, rail, pipeline, inland waterway, telephone, and Internet point and line facilities. Interaction networks can include a wide variety of interactions such as social connections, business transactions, Internet hot links, and academic citations, as well as transport and communication flows. Note that the range of phenomena that can be analyzed as networks include those in which both nodes and arcs are physically fixed in space (e.g., roads), neither are fixed in space (e.g., friendships), or only the nodes are fixed in space (e.g., seaports).

This article focuses on three methods of network analysis: graph theory, complex network theory, and network optimization (mathematical programming and heuristic algorithms).

Graph Theory

William L. Garrison and Duane Marble introduced sci entific analysis of network structure to geography be ginning in 1960, although their work was preceded by Shimbel's in 1953. Garrison and Marble's work helped kick off the quantitative revolution in geography.

Graph theory analyzes the topology of networks. It is sometimes referred to as static graph theory to dis tinguish it from the more recent field of complex net works (see below), which focuses on how networks evolve. Graph theory measures and compares networks in two ways. First, it computes a variety of full network measures in order to understand the structural properties and adequacy of the overall network. Second, it math ematically manipulates matrix representations of net works to assess the internal spatial structure and properties of particular nodes. Very little has been added to the graph theory literature since the mid 1960s, but some of these basic measures of network structure are still used today in academic and practical settings.

Full Network Measures

Various measures have been developed to characterize the connectivity or density of an entire network as a whole. These methods provide a basis for comparing networks of different sizes, modes, types, times, and places.

Connectivity measures

A minimally connected network, also known as a tree network, contains no circuits. The removal of any edge would produce two unconnected subgraphs (possibly consisting of a lone vertex). The minimum number of edges is given by

$$e_{\min} = v \quad 1 \qquad [1]$$

where e represents the number of edges and v the number of vertices in the graph. Most of the measures that follow are based on the insight that each additional edge added to a tree network produces one additional circuit. In a nonplanar graph, every vertex can connect to every other vertex, and thus the maximum number of edges assuming two way edges is

$$e_{\max} = v(v \quad 1)/2 \qquad [2]$$

In a planar network, the most highly connected network consists of a series of connected triangles, leading to this formula for the maximum number of edges:

$$e_{\max} = 3(v \quad 2) \qquad [3]$$

The 'gamma index' uses eqns [2] or [3] to summarize how relatively connected a network is. Its value ranges from 0 to 1 (using, of course, the proper equation for maximally connected networks for planar and nonplanar networks, respectively):

$$\gamma = e/e_{\max} \qquad [4]$$

One of several similar measures is the 'beta index', which measures the average number of edges per vertex, and has a minimum value of zero:

$$\beta = e/v \qquad [5]$$

A limitation of these kinds of measures, however, is that they take into account topology only, not scale or density in relation to the people or area served.

Regional Measures

Another set of full network measures relates the network to the region being served. They are widely used by governments and economic development organizations such as the World Bank to assess infrastructure investment needs.

Network density (ND) is a measure of the total network length L (all links added together) relative to some base variable B:

$$\mathrm{ND} = L/B \qquad [6]$$

The numerator L is often measured in kilometers but can also be measured in lane km or track km to better account for the capacity. The base variable B can be the area of the region in question, the arable land area (since there is less need for transport infrastructure in sparsely populated regions), or the population. Another measure of the adequacy of a network is the Eta index:

$$\eta = \text{total network miles/total number of links} \qquad [7]$$

Eta measures the average link length, which in a planar road or rail network is a reasonable proxy for density.

Several methods measure the degree of circuity in a network by comparing the actual length of links to the geodetic (straight line over a curved surface) distance between the start and end of the link. The degree of circuity, also known as the route factor, is a ratio of the actual over the geodetic distance for each link. Ratios of 1.0 indicate a straight line, whereas ratios greater than 1.0 indicate circuity. The degree of circuity for the entire network can be computed either as the average of the route factor ratios for all links, or as the ratio of the sums. In either case, higher numbers generally result from poor connectivity and/or rugged terrain.

Finally, traffic density is an indicator of the usage of a network relative to its size. It can be measured in several ways. The theta index computes the average flow (e.g., passengers or tons) per node. Similarly, an index that Black (2003) refers to as the iota index computes the average passengers or tons per mile. Both indices, however, fail to consider the average trip length. A more commonly used measure of traffic density, TD is:

$$\mathrm{TD} = U/L \qquad [8]$$

where usage U is measured in tkm (ton km shipped) or pkm (passenger km traveled) and L is again the total length of the network. The resulting indicator is measured in tkm or pkm per km or per lane km, depending on how L was defined. Higher traffic density is often interpreted as efficient use of network infrastructure.

Nodal Measures

Another branch of graph theory quantifies the relative positions of each node within a network. The basic building block for this type of graph theory is a representation of the network as a connectivity matrix C – a square matrix with a row and column for each node. A direct link between two nodes i and j is represented by a cell value $c_{ij} = 1$, while all other cells are 0. Using this matrix, one can calculate the degree k_i of node i which is a simple count of how many links it has to other nodes:

$$k_i = \sum_{j \quad 1}^{n} c_{ij} \qquad [9]$$

The degree of a node, however, can be misleading, because it measures only direct linkages. More meaningful measures of nodal accessibility also incorporate other principles:

1. Direct and indirect linkages: methods should recognize that it is possible to move from one node to another, even if no direct link between them exists, by traveling indirectly via other nodes.
2. Multiple paths: methods should give credit for the existence of alternative routes.
3. Attenuation: methods should discount indirect routes relative to direct routes.
4. Redundancy: methods should correct for alternative routes containing meaningless back and forth loops.
5. Topological distance versus length: methods should be able to measure connectivity to other nodes in terms of the actual length of all the arcs in a route, or in terms of the number of arcs.
6. Circuity: this is similar to attenuation, but for length rather than number of links.
7. Unequal node size: methods should account for the size of nodes.

The first two desired improvements to node degree, namely indirect linkages and multiple paths, can be incorporated by computing the total accessibility, T, matrix. To get T, one first multiplies the connectivity matrix C (referred to here as C^1) by itself to get the C^2 matrix (**Figure 1**). In the C^2 matrix, the values for each i–j pair represent the number of paths from i to j using two links. Multiplying C^2 by C^1 yields C^3, which gives the number of paths from i to j using three links. Note that the method counts redundant, nonsensical paths such as A–B–A–B and A–C–A–B as three-link paths from A to B.

Then, to find the T matrix,

$$T = C^1 + C^2 + C^3 + \cdots + C^n \qquad [10]$$

where n is the topological diameter of the network (the maximum number of links between any two nodes). Summing across the rows indicates the number of ways node i can reach all nodes in the network, but does not correct for attenuation, redundancy, length, or node size.

Other extensions to this basic technique address the remaining principles [3]–[7]. Garrison introduced a scalar weighting system that accounts for attenuation by assigning lower weights to paths with more links. Shimbel eliminated both attenuation and redundancy by generating a D matrix that records only the number of links in the shortest topological path between any pair of nodes, ignoring all paths with more links. An L matrix of shortest path lengths between all pairs of nodes can be generated with another type of matrix manipulation; the end result is similar to the output of an all-to-all shortest path algorithm. Summing across the rows of the D and L matrices will yield the total number of links or the total length to move from a node i to all other nodes in the network. Kansky developed measures of circuity for individual nodes, based on the difference between the Euclidean distances and the network's shortest paths from node i to every other node j. Finally, the topological distances in D or lengths in L can be mathematically combined with variables representing the size of the nodes to generate measures of each node's accessibility to the distributed population in the other nodes of the network, following work by Hansen. By summing across all destination nodes, one can generate the population

C^1	A	B	C	D	E
A	0	1	0	0	0
B	1	0	1	1	0
C	0	1	0	0	0
D	0	1	0	0	1
E	0	0	0	1	0

C^2	A	B	C	D	E
A	1	0	1	1	0
B	0	3	0	0	1
C	1	0	1	1	0
D	1	0	1	2	0
E	0	1	0	0	1

C^3	A	B	C	D	E
A	0	3	0	0	1
B	3	0	3	4	0
C	0	3	0	0	1
D	0	4	0	0	2
E	1	0	1	2	0

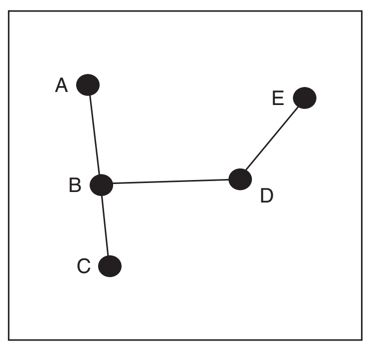

$c_{ij}^k = c_{i1}^1 c_{1j}^{k-1} + c_{i2}^1 c_{2j}^{k-1} + \cdots + c_{in}^1 c_{nj}^{k-1}$

for example

$c_{AC}^2 = c_{AA}^1 c_{AC}^1 + c_{AB}^1 c_{BC}^1 + \cdots + c_{AE}^1 c_{EC}^1$

or

$c_{AC}^2 = 0\times0+1\times1+0\times0+0\times0+0\times0$

Figure 1 The C^1, C^2, and C^3 matrices for the network shown. The value of cell (A, C) of C^2 is found by multiplying the A row of C^1 by the C column of C^1 as shown in the formula. The rows, columns, and cells in question are highlighted in gray. The multiplication shows that there is one and only one way to get from node A to node C using exactly two arcs. Note the 3 in cell (B, B) of C^2 indicates that there are three ways to get from node B to node B using exactly two arcs (they are B-A-B, B-C-B, and B-D-B).

potential of a node, given by

$$V_i = \sum_{j=1}^{n} \frac{P_j}{d_{ij}} \quad [11]$$

where V_i is the population potential, P_j is the population of node j, and d_{ij} can be either the topological distance in D or the shortest path distance in L between nodes i and j.

In addition to using graph theory measures to simply describe network structure and the positions of nodes within them, many researchers beginning with Garrison and Marble have empirically tested whether better connectivity, density, and accessibility are associated with improved economic performance at national, regional, and nodal scales.

Complex Network Theory

Complex network theory, the so called 'new science of networks', studies the structure of interaction networks and their evolution. Complex network theory assumes that new linkages are formed under a set of probabilistic governing principles based on the distribution of existing linkages. Scholars have found striking similarities in network topologies across a broad spectrum of interaction networks, from the Internet to academic citation webs, business networks, ecological networks, and sexual and disease networks (**Figure 2**).

The term 'interaction networks' is used here to refer to networks in which a linkage is defined by interaction of some kind between the two nodes, in contrast with infrastructure networks consisting of physical connections such as roads, rail, pipes, or wires.

Within the domain of complex networks lie both small world and scale free networks. Small world theory and scale free theory are not one and the same, although the terms are sometimes used interchangeably. The chief underpinnings of small world networks are (1) a small network diameter, as in the well known 'six degrees of separation' between any two people, and (2) a high degree of clustering. In scale free networks, on the other hand, there are many nodes with few links and few nodes with many links – in fact the probability distribution of nodes per link follows a power law functional form. In addition, in scale free networks, new links are more likely to attach to nodes that already have many links – the preferential attachment rule. These rules, it is thought, capture something fundamental in the way that interaction patterns drive network formation in the real world. The term 'scale free' implies that the topologies and rules seem to apply to networks of all sizes, and to subgraphs of larger networks.

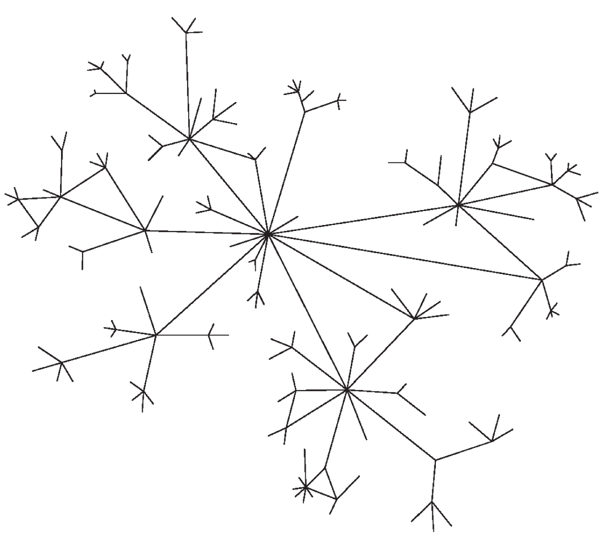

Figure 2 Hypothetical example of a complex network.

Network Topology Measures

Intrinsic in the structure of a complex network – as opposed to a random, lattice, or point to point network – is the power law distribution of node degree. In a static snapshot of a complex network, a few nodes have a great number of incoming links, while most nodes have few or perhaps only one. This probability distribution of node degree is described by a power law distribution:

$$P(k) \sim k^{-\gamma} \quad [12]$$

where k is the number of links and γ is the scaling exponent. The Barabási–Albert model gives the value of $\gamma = 2.9 \pm 1$, although networks considered to be scale free have ranged from 2.3 for movie actor networks to $\sim 5 \pm 1$ for sexual networks.

Small world networks are characterized by a high degree of clustering. The clustering coefficient (C) is related to the propensity for two linked nodes to have a common neighbor to which both are linked:

$$C = \frac{3(\text{number of triangles})}{\text{total possible } k \text{ in the network}} \quad [13]$$

where triangles are three nodes each connected to one another.

Perhaps the best known trademark of small world theory is the property of small network diameter. Network diameter is the path length between any two nodes in the network. Most studies of small world networks, dating back to Stanley Milgram's six degrees of separation, have used the average path length definition rather than the maximum. Graphs generated by a truly random process (known as Erdős–Rényi graphs) have an average diameter d that can be approximated by:

$$d \sim \ln N \quad [14]$$

where N represents the number of nodes in the network. Cohen and Havlin found that scale free networks display behavior that is consistent with

$$d \sim \ln(\ln N) \qquad [15]$$

for networks with degree distribution $P(k) \propto k^{-\gamma}$, where $2 \leq \gamma \leq 3$. The maximum path length can be found using a standard shortest path algorithm (see below). There is much debate about the significance of small diameter, which can also occur in many random networks of similar size.

Both the airline network and sections of the ocean shipping network display properties consistent with scale free topology. Guimerá *et al.* empirically illustrated that the international airline network is scale free and attributed that fact to multihub airline networks. Roberts found ocean shipping routes in the United States to be scale free with a degree distribution of $P(k) \sim 3 \pm 0.1$.

Network Growth

The power law distribution and degree of clustering are explained by the way scale free networks evolve. In scale free network theory, the probability of links being added to a node is proportional to the number of links the node already has. This is known as the preferential attachment rule:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \qquad [16]$$

where k_i is the number of links to node i and $\Pi(k_i)$ is the probability of a new link attaching to i. Networks that add links in proportion to how many links nodes already have tend to display clustering, a power law distribution, and small network diameter.

Vulnerability and Resiliency

Perhaps the most useful discovery in research about complex networks is that small world and scale free networks illustrate both great resiliency and vulnerability. Generally speaking, complex networks illustrate a higher resiliency to random attack and a higher vulnerability to a directed attack than many other networks. Barabási cited a number of instances of vulnerability when scale free networks, including regional power grids and online retailers, fell victim to a directed attack on highly connected nodes. The subsequent failures cascaded through the network, leading to near total temporary collapse of the systems. Other researchers have noted that, had the Federal Aviation Administration (FAA) not shut down the national airline system on September 11th, the nation's airline system would have ground to a halt anyway due to the closure of airports in New York and Washington, DC.

The root of this vulnerability to failure and attack is, of course, the degree distribution. Using simulation models, Albert *et al.* found that eliminating the top 5% of most connected nodes can double the network diameter. Removing the top nodes may also disconnect single nodes and smaller clusters of nodes from the rest of the network.

The same power law degree distribution, however, is behind the resilience of scale free networks to random attacks on nodes. Due to the skewed nature of the node degrees in the network, there is a far greater chance that nodes with very few links will be attacked or fail. Other research has shown that quick removal or quarantine of key nodes that are topologically close to the source of the attack is the best strategy to prevent catastrophic failure of system.

Network Optimization

Network optimization models are a class of prescriptive mathematical models solved using exact or heuristic methods. Many such models can be formulated and solved as linear programs (LPs) – an application of matrix algebra. An LP is composed of decision variables, an objective function, and constraints. The objective function represents a performance measure to be maximized or minimized. Constraints define which solutions are feasible and which are not. The optimal solution is the one with the best (max or min) objective function value that does not violate any of the constraints. LPs are the most basic form of problem in the larger family of mathematical programming problems, which also includes integer, mixed integer, multiobjective, nonlinear, and dynamic programming. Mathematical programming can solve difficult problems exactly, but may not be fast enough or solve large enough problems for some applications. In contrast, heuristic algorithms – such as Lagrangean relaxation, out of kilter, greedy, genetic, tabu search, or simulated annealing algorithms – often can find optimal or near optimal solutions to large problems efficiently, but are not guaranteed to find the exact optimum solution.

A wide variety of network optimization models have been developed and are widely used in industry, government, and academia. Here we introduce a few of the more basic and widely used models. Facility location problems, either on a network or a plane, are covered elsewhere in this volume.

Shortest Path Algorithm

The problem of finding the shortest path through a network underlies many network analysis problems. 'Shortest' can be defined in terms of distance, travel time, or cost. While it can be formulated as an integer program and solved as a relaxed linear program, heuristic algo rithms are more efficient and are known to find exact optimal solutions. The Dijkstra node labeling algorithm finds the closest node to the origin node, labels it as scanned, and then iteratively proceeds to find and label the next closest node, and so on, until the desired des tination node is reached. Floyd's algorithm uses a matrix approach to find the shortest path from all nodes to all other nodes.

Simple Transportation Problem

One of the first and most basic transport optimization problems was the simple transportation problem, which Hitchcock and Koopmans independently formulated and solved around 1941. In this problem, a set of supply nodes are given with amounts they can produce, as well as a set of demand nodes with amounts required. The costs of shipping one unit from each supply node to each demand node are known. The decision variables are how much to ship from each supply node to each demand node. The objective is to minimize the total shipping cost, while the constraints require that supply limits not be exceeded at supply nodes and demand be satisfied at all demand nodes. The model can be solved with a simple linear program or an exact heuristic algorithm in volving transportation tableaus. The Hitchcock–Koop mans problem is not, however, a network problem *per se.* The transport costs can be computed over a network, through Euclidean space, or according to other metrics.

Minimum Cost Network Flow Problem

The minimum cost network flow problem finds the least cost circulation flow through a network. In a circulation flow, the flows into each node exactly equal the flow out of each node. The decision variables in the model are the flows X_{ij} on each link i, j. The inputs are the unit trans port costs c_{ij}, lower bounds on the flow l_{ij}, and upper bounds u_{ij} of each link. The model can be formulated as a linear program:

$$\sum_i \sum_j c_{ij} X_{ij} \qquad [17]$$

subject to

$$\sum_i X_{ij} \quad \sum_k X_{jk} = 0 \quad \text{for all nodes } j \qquad [18]$$

$$X_{ij} \leq u_{ij} \quad \text{for all arcs } i, j \qquad [19]$$

$$X_{ij} \geq l_{ij} \quad \text{for all arcs } i, j \qquad [20]$$

Heuristic algorithms such as the out of kilter algorithm, however, can solve the network flow problem more effi ciently than linear programming software.

The minimum cost network flow problem differs from the previous two network optimization problems by virtue of having minimum and maximum arc flow cap acities. The problem is much more flexible, however, than it first appears. By adding artificial nodes and arcs, and by setting their costs and bounds properly, many other network problems, including the aforementioned shortest path and simple transportation problems, can be struc tured as minimum cost network flow models, allowing for them to be solved quickly and easily. For instance, **Figure 3** shows how flows from a set of sources (S1 and S2) to a set of sinks (D4, D5, and D6) over a capacitated transport network can be modeled. By connecting an artificial super source node (A7) to each origin by a zero cost arc with an upper bound equal to the source's supply limit, connecting each sink to an artificial super sink node (A8) with a zero cost arc with a lower bound equal to the sink's minimum requirement, and connecting the super sink to the super source by a zero cost, infinite capacity arc, the minimum cost circulation flow will be the cheapest way to route goods to satisfy demands without violating supply, demand, or flow constraints.

Traffic Assignment Problem

In the simple transportation and network flow problems above, there are amounts supplied and amounts de manded, but the model chooses which supply goes to which demand. In many real world problems, however, people must commute from particular zones to particular jobs, or trade must flow from particular seller to par ticular buyer. In the traffic assignment (aka network as signment) problem, a matrix called a trip table defines the required flow between each origin–destination pair. The problem is to find the routes for the given flows from origins to destinations over a capacitated network. There are two main versions of this problem. If the traffic flows are managed by a single entity, such as a national railway company, a prescriptive system optimal traffic assign ment problem is usually formulated as a cost minimiza tion model with 'brick wall' arc capacities. If independent network users (such as drivers) determine their own routes, a predictive user equilibrium traffic assignment problem is formulated with nonlinear congestion func tions representing arc capacities. The typical equilibrium condition, called a Wardrop equilibrium, is that no users can reduce their travel time/cost by switching routes, and the model is solved iteratively until convergence. User equilibrium traffic assignment problems are typi cally the final step in the four step urban transportation modeling process.

Network Design Problem

A network design problem decides which arcs, if any, to add to a network for efficient flow. Although there are

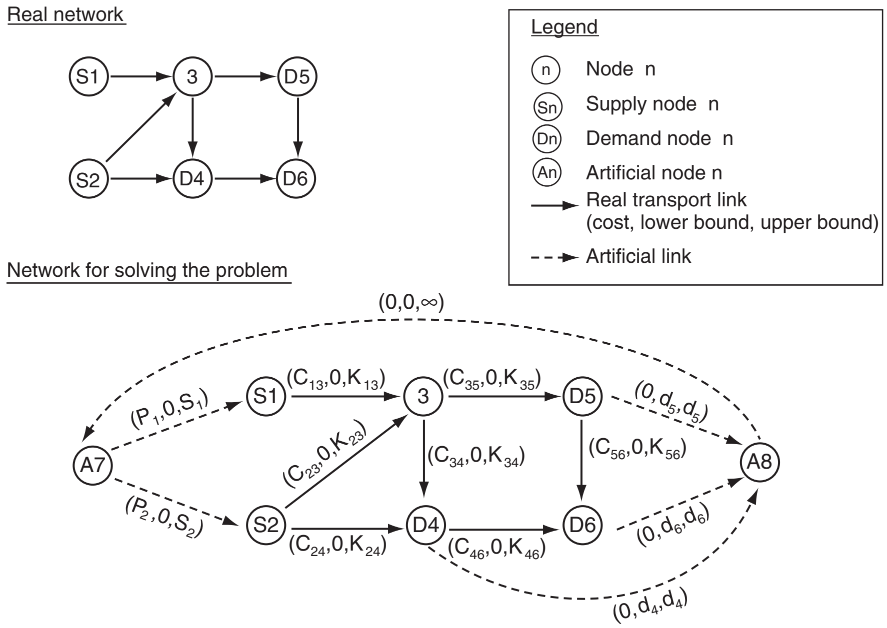

Figure 3 Flows from a set of sources (S1 and S2) to a set of sinks (D4, D5, and D6) over a capacitated transport network. Definitions: P_i supply cost of i; S_i supply capacity of i; d_j demand requirement of j; c_{ij} transport cost from i to j; K_{ij} transport capacity from i to j.

many variations of network design problems, they typically combine a traffic assignment problem with variables representing construction of new arcs. The objective function minimizes the sum of the costs of both expanding the network and using the network. The model trades off the fixed costs of adding new arcs with the savings on operating costs that result from using the new arcs. The network design problem is considerably more complex and computationally intensive than the traffic assignment problem because the decision variables representing whether or not to add an arc to the network must be binary (0, 1) variables.

Other Network Optimization Problems

A wide variety of other network optimization problems have been formulated and solved. A minimum spanning tree problem is to choose the subset of arcs of a network that have the shortest total length while providing a route between each pair of nodes. Hub network models are covered elsewhere in this encyclopedia. Scheduling problems for airline crews or pipeline flows have proven among the most difficult problems because of their continuous dynamic nature. The flow capturing problem is to locate p facilities on a network to capture the highest volume of passing flows (based on a trip table of flows and the shortest paths of each origin–destination pair). The problem assumes one facility anywhere on the path can intercept the flow. The flow refueling problem is an extension of the flow capturing problem where vehicles have a limited driving range and an adequately spaced combination of facilities may be necessary to refuel a given flow.

Other Network Analysis Methods

While graph theory, complex network theory, and network optimization are most likely to come to mind under the heading of network analysis, geographers use other methods to analyze networks. First and foremost, geographic information system (GIS) is widely used to analyze networks, and an entire field of GIS transportation (GIS T) has developed to store, capture, display, and analyze spatial network data.

Second, simulation models are dynamic models that mimic the process of a system and predict its changes through time. One type of simulation model used for transportation analysis is a stochastic, queuing model. At the most basic level, queuing type simulation models are built of entities, resources, and queues. While not all transport systems have queues, they are a very common feature, if only because most transport systems will, at some point, have more entities to service than they have available resources.

Agent based modeling (ABM) is a new simulation modeling paradigm that focuses on modeling disaggregated activities and decisions by autonomous agents, rather than modeling the system as a whole. Often called a bottom up approach, the crux of ABM is that a group of entities (the ‘agents’) behave according to certain rules in the simulation. Bonabeau asserted "ABM is a mindset more than a technology." The mindset of ABM is to model individual agents' behavior, and then let that be havior play out in a simulation that yields the aggregate results of their interactions. ABMs are well suited for studying evolutionary processes based on individual be haviors. Traffic flows, traffic congestion, and traffic acci dents are the net result of the interactions of vehicle operators on a network. In addition, agent based model ing is used to predict the co evolution of land use and transportation networks. Models such as UrbanSim are used by metropolitan planning organizations to forecast future urban growth.

Acknowledgments

This article is based on a NASA report by the authors, listed in the Further Reading section (NASA/CR 2005 213522). Permission has been granted by NASA for use of this material.

See also: Aviation; Garrison, W.; Hub Network Location; Internet, Economic Geography; Location Analysis; Networks; Quantitative Revolution; Railways; Simulation.

Further Reading

Ahuja, R. K., Magnanti, T. L. and Orlin, J. B. (1993). *Network Flows: Theory, Algorithms, and Applications*. Upper Saddle River, NJ: Prentice Hall.
Barabási, A. L. (2002). *Linked: The New Science of Networks*. New York: Plume.
Barabasi, A. L. (2002). Visual companion. http://www.nd.edu/~networks/Linked/photo.html (accessed 20 November 2008).
Barabási, A. L. and Bonabeau, E. (2003). Scale free methods. *Scientific American* May, 60 69.
Caldwell, C. K. (1995). Graph theory tutorials. http://www.utm.edu/cgi bin/caldwell/tutor/departments/math/graph/index.html (accessed on 20 November 2008).
Daskin, M. (1995). *Network and Discrete Location*. New York: Wiley.
Garrison, W. L. and Marble, D. F. (1961). *The Structure of Transportation Networks*. Washington, DC: US Department of Commerce, Office of Technical Services.
Guimerá, R., Mossa, S., Turtschi, A. and Amaral, L. A. N. (2004). *Structure and Efficiency of the World Wide Airport Network*. http://arxiv.org/abs/cond mat/0312535v1 (accessed on January 2008).
Kansky, K. J. (1963). *Structure of Transportation Networks: Relationships between Network Geometry and Regional Characteristics*. *Research Paper 84.* Chicago: Department of Geography, University of Chicago.
Kincaid, R. K. (2003). *Scale Free Graphs for General Aviation Flight Schedules*. Langley, VA: NASA Langley Research Center, NASA/CR 2003 212648. http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20040000900 2003137972.pdf (accessed on January 2008).
Kuby, M., Tierney, S., Roberts, T. and Upchurch, C. (2005). *A Comparison of Geographic Information Systems, Complex Networks, and other Models for Analyzing Transportation Network Topologies*. Langley, VA: NASA Langley Research Center, NASA/CR 2005 213522. http://techreports.larc.nasa.gov/ltrs/PDF/2005/cr/NASA 2005 cr213522.pdf (accessed on January 2008).
LeBlanc, L. J., Morlok, E. K. and Pierskalla, W. O. (1975). An efficient approach to solving the road network equilibrium traffic assignment problem. *Transportation Research* 9, 309 318.
Rodrigue, J. P., Comtois, C. and Slack, B. (2006). The geography of transport systems. http://people.hofstra.edu/geotrans (accessed 20 November 2008).
Taaffe, E. J., Gauthier, H. L. and O'Kelly, M. E. (1996). *Geography of Transportation* (3rd edn.). Upper Saddle River, NJ: Prentice Hall.
Watts, D. J. (2003). *Six Degrees: The Science of a Connected Age*. New York: W.W. Norton.

Relevant Websites

http://www.scienceofbetter.org
Institute for Operations Research and Management Science. Operations research: the science of better.
http://www.insna.org
International Network for Social Network Analysis.
http://oracleofbacon.org
University of Virginia, Computer Science Department, 1999. The oracle of Bacon at Virginia.

Network Regions

P. Cooke, Cardiff University, Cardiff, UK

Glossary

Associativeness The existence of strong networks of informal and formal associations in society that adopt a common discourse to promote the achievement of desired ends, neither entirely in the market nor in the state but seeking to influence both.

Discourse Language communicated in the form of a coherent paradigm articulating distinctive interests and members of important networks to achieve domination in a given field.

Innovation The commercialization of new knowledge.

Network A system of nodes and linkages among persons (and essential equipment) to evolve a discourse to achieve some purpose of consequence to the network's interests.

Paradigm A 'world' or 'perspective' of what is the case agreed upon by the powerful shapers of discourse in any field and accepted by all of consequence to that field, except the paradigm's critics.

Placeless Foodscapes Forms of food retailing or processing that discount space, either as a barrier to consumption of products on a global scale, as in supermarkets, or as forms of 'fusion' food that cannot be attributed to any particular place on the planet, as in restaurants.

Social Capital Exchange relations in society based on trust, reciprocity, and reputation resulting in favor, gift, and support actions largely outside market or state functions.

Introduction

The idea of a network region began with the notion of a 'network paradigm' for regional economic development which was first articulated in the early 1990s. A network region has many nodes and connections between institutions and organizations; it has high social capital and associativeness. However, the network concept had been bubbling up earlier than that, particularly in relation to innovation networks, but not especially 'regional' networks. However, it is clear that the Francophone European Research Group on Innovative Milieux (GREMI) group were approaching a discourse of regional development in 'network' regions but without explicitly utilizing that language (which in French is *résaux*) because of a preference for the rather vaguer notion of 'milieu'. Curiously, 'industrial' networks but not 'regional' networks were mentioned many times. What did 'milieu' signify? It means medium or atmosphere, such as that which sustains us by enabling us to breathe, or water for fish, which performs an equivalent life supporting function. More economically geographical, it is not that far from Marshall's notion of the key to or secrets of industry, or in this context innovation, being 'in the air', again a vague but irritatingly fascinating term that continues to be cited at least a century after he first articulated it. Marshall, of course, wrote about, and in a way discovered, the secret of nineteenth century capitalism, which is that its division of labor was highly selectively located in industrial districts that were, in modern parlance, suffused with network relations among entrepreneurs and employees. In this way, knowledge or, in Marshallian terms, 'information' was in the air because it was the content of communication about technology, specialization, labor processes, associated skills, and their price in these constrained geographical spaces and industrial communities. Neo Marshallian research had begun on highly networked, localized 'industrial districts' even before these 'milieu' dates.

The other main lineament of network regions, though regions were secondary to networks and innovation, arose in a much cited special issue of the journal *Research Policy*. Networks of Innovators were the objects of interest to which many economic geographers – such as Storper and Harrison, Saxenian, Scott, Lawton Smith, and Glasmeier – contributed. Perhaps only the second and fourth focused explicitly on regional networks as such. Freeman was more explicit, but less regional, in his review of 'networks of innovators', as were Bianchi and Bellini on policy support for such networks. Nevertheless, through these interventions and the book edited by Grabher published in 1993, such early awakenings to the notion of network interactions among innovators as a new understanding of the importance of geography to innovation, revealed economic geographers to be at the intellectual forefront of this burgeoning field. Economic geographers like Grabher himself explored the neglected social dimension in economic networks conceived by neoclassical economists as 'asocial', while in Grabher's book, Ash Amin sought to show how globalization was destroying regional networks, a peculiarly 'northern' perspective also later taken up by Coe and Townsend who stood on Amin's shoulders in seeking to 'debunk the myth of localized agglomeration', at least in London, something that now seems recklessly heroic. This is not least given, for example, Grabher's subsequent work on the advertising 'village' in Soho. Another author whose name would be much cited in the regional innovation and networks literature in subsequent years, Bengt Åke Lundvall, also appeared in Grabher's book writing about cooperation in innovation activities among firms. So this early 1990s era marks a break in the conceptual and real geography of innovation as smart entrepreneurial firms began running rings around large firms who, in turn, began outsourcing innovation to their supply chains, which, as Klaus Semlinger, yet another contributor noted, were often highly regionally networked, especially in his native Germany, but also elsewhere.

Integrating Regional Networks and Regional Innovation Policy

The enthusiasm for studying networks remained in a context of manifest decline in the coordinating capabilities of states and markets regarding leading edge research and innovation, which subsequent evidence shows set in from approximately 1991. But if the central state had become as debilitated as many large private corporations were to become regarding the lack of productivity from their large budgetary allocations to research and development (R&D), the 'regional state' seemed from empirical reportage of the kind discussed above to be on the rise. A parallel strand of research had evolved, which focused on regional innovation policy. Thus, the connecting concept of regional innovation systems (RISs) evolved from this even earlier thinking about 'regional innovation policy', in relation to 'regional innovation networks'. This happened in two publications, by Cooke (see Further Reading), the more widely cited one being less theoretically and empirically rich than the almost totally uncited one. The difference lies in the absence of any bibliographical influence from the 'innovation systems' literature in the 1992 paper, that thus has purer lineaments to economic geography. Contrariwise, the 1993 paper which shows the author had by then read Lundvall on 'innovation as an interactive process' and was also influenced by Johansson and Grabher in probably the first proper book on regional development from a 'network regions' perspective by Bergman, Maier, and Tödtling, also published in 1991.

It seemed necessary to place these distinctive 'network and policy' concepts in relation to each other in a layered model. So, the innovation policy dimension evolved conceptually into the idea of a subsystem, supporting with knowledge and resources the innovative firms in their networks. These formed a 'superstructural' subsystem dealing with actual innovation 'near market'. As we have seen, they had been spoken of as carrying out 'networking' with each other, not only laterally in alliances or partnerships and vertically in sometimes partly localized supply chains but also with the innovation policy and knowledge generation subsystem. So these also had subsystem characteristics related to the governance of innovation support. Each subsystem was also seen to interact with global, national, and other regional innovation actors, and even through technological or sectoral systems of innovation. Open systems ruled.

Over the years, the RIS framework has been analyzed in terms of many different 'varieties of innovation' relating to localized, networked, and hierarchical innovation 'governance' systems. Third Italy, Baden Wuerttemberg (BW), and French innovative regions exemplified each, respectively. Correspondingly, the 'exploitation' subsystem of firms, in the main, could be dominated by large firms or oligopolies – even foreign ones as with the Asian transplants to Wales in the 1980s and 1990s. Other regions such as Catalonia had a mix of large (SEAT) and small and medium sized enterprise (SME) 'district' type innovation relations, while other places might have innovation regimes in which only small, entrepreneurial firms predominated, as in places with observable 'industrial districts', not only Third Italy but also some newer technology 'clusters'. Later still, these, more entrepreneurial SME systems, living by venture capital and exploitation of public research from universities, could be differentiated further as 'entrepreneurial' (ERIS), market led systems, compared with those, especially in Europe, where they were more 'institutional' (IRIS) where state support was pronounced and 'entrepreneurship' was less advanced.

The First Candidate-Networked Regional Innovation System

RISs are not really 'implemented' by policy but rather they evolve through processes of incremental and sometimes even quite 'disruptive' institutional change by markets and the institutional innovation support system. Hence, an institutional change such as the founding of a regional development agency (e.g., Scotland in 1976) that evolved quite a powerful innovation support regime, does not – even today – have a well functioning 'innovation system'. Interinstitutional network gaps exist, and most SMEs do not function in fluent network relationships. Thus, having the innovation policy subsystem does not necessarily bring forth an innovative network firm subsystem, certainly not swiftly. By contrast, periodically Silicon Valley as an ERIS innovation model has shown excellent networking capabilities, not only among its manufacturing firms but also among its venture capitalists. But, until relatively recently, the governance of innovation displayed little meso level intervention of the kind to be discussed below. In the past decade an 'associative' organization called Joint Venture: Silicon Valley has formed to promote civic entrepreneurship around education and the environment. Hitherto there had been no civic association or representative governance body for this high tech complex other than Santa Clara county and the fragments of various cities into which it fell. The key networks were between firms, universities, and the US Federal mandates such as defense, energy, and healthcare.

Thus one reason why BW is a candidate as the first RIS is that its apparently modern style of network intervention by various institutions interacting with firms can actually be traced back to the nineteenth century Kingdom of Württemberg. Here, Ferdinand von Steinbeis established a pioneering vocational training system focused on the Esslingen locomotive works to train impoverished rural migrants in engineering because the regional economy of that time was too poor to sustain its population, and hundreds of thousands – as elsewhere in Europe – were migrating to America. This was a 'seed crystal' of the idea that 'catch up' with more industrialized (or today more knowledge intensive) economies can fail to occur due to 'market failure': in which case third party and, increasingly over time, state intervention to design new subnational institutional 'architectures', is justified. Normally, national governments took these actions, but following the Thirty Years' War and up to preunification Germany, sovereign governments were often scaled like regions or even city states (e.g., Hansa cities) rather than the multilevel governance systems present in contemporary Europe.

Assessments of outcomes in the BW case were that during the balmy post reunification economic times of the 1990s, financially stable and growing, its RIS worked well. In particular, new knowledge, especially from abroad, flowed through the institutional networks remarkably swiftly, not least because of the multitude of business associations, research institutes, intermediary associations (e.g., the Steinbeis Foundation), and chambers of commerce that held frequent seminars and workshops or keynote presentations for their members, who turned up, often in hundreds, to meetings on such topics as 'lean production', 'systemic innovation', and 'flexible labor markets'. Problem solving capacity, in which government was a partner, was accordingly swift, and innovation support measures or even innovative policy instruments were forthcoming. This seemed of the essence. But once the need for significant restructuring caused by sharp competition from Asia hit the institutional setup, it could not adjust, as a classic IRIS, as swiftly as ERISs like Silicon Valley. There, 400 000 electronics jobs disappeared almost overnight, many more than that were lost from the BW engineering platform, but by today Silicon Valley is well into its fourth or fifth disruptive innovation generation based on iPod and the semantic web. It is often said Google could easily have been envisioned in Europe – but it was not, and the reason why lies at the heart of Europe's innovation paradox. The European Union (EU) produces world class science but, often, minor league innovation.

The Well-Networked Regional System

A network region is a self organizing, open system which, in regard to innovation, translates 'exploration' knowledge from inside or outside the RIS, through a process of 'examination' of that knowledge to test and try it for dangerous performance under stress conditions, to the point where it can be successfully exposed to 'exploitation' knowledge in the form of a commercial innovation on the market embodying new knowledge, whether scientific, technological, or creative. This means the network region has appropriate institutions to cross fertilize ideas and bridge any gaps that need crossing between institutions and people. However, network regions are not confined to technological innovation. Arts based systems, though less studied, can also be innovation systems – as Montmartre, Montparnasse, and other points in '*fin de siècle*' Paris, networked to other nodes in France and other 'art cities' were. New York replaced Paris as the world's leading 'creative innovation system' in the late twentieth century, and in this century, London seems to have risen, through 'Britart' to challenge New York. In each case, systemic institutional support (academies and conservatories) and, especially, associativeness among artists and their networks to dealers or, in music or literature, impresarios and agents, led to outpourings of creative product. A network region functions through good transparency and openness between the knowledge generation or exploration subsystem and the knowledge exploitation subsystem, including nowadays what is perceived as practically a third, intervening, subsystem that translates the often tacit knowledge of the former, including from abroad, into primitive forms of codified knowledge for exploitation by the latter.

Increasingly, it can be seen that the 'innovation paradox' lies in this third 'intermediary subsystem' which in ERIS setups is occupied by knowledgeable attorneys or lawyers of various kinds, knowledge entrepreneurs who solve problems or seek solutions, financial engineers like venture capitalists, business angels and management accountants, and varieties of specialist consultants, incubator and accelerator managers, even well functioning knowledge transfer offices in universities. IRIS setups are slow and inflexible because markets for this talented kind of stratum are thin and largely absent so, through necessity, such functions are performed by unenterprising risk averse government officials who do their best but cannot be effective substitutes.

It is worth noting, in passing, a completely separate 'network geography' that built up around a different school of network thinking. The school is known as 'constructivists' and the approach is known as 'actor network' theory'. Strangely it also began from study of the worlds of scientists and innovators by the likes of Michel Callon and Bruno Latour. But it was also adopted by geographers, notably the late Jonathan Murdoch of Cardiff University, who utilized it successfully in rural studies, particularly in relation to the food industry. He and colleagues showed how regions like Tuscany, Wales, and California developed their high grade food and drink offers by the conscious initiative of networks and associations that promoted a discourse of excellence in local provenance in their diversified food products, both to tourists and quality food markets worldwide. The rise of organic food innovations, on the one hand, and functional food (biotechnological) can be explained by reference to the critical discourses articulated by actor networks. As, in a different way, can be the dominance of conventional food by the articulation of discourses from intensive farming, agri business, chemical fertilizer companies, agricultural machinery producers, supermarkets, banks, and governments that subsidize them.

Both these approaches have been criticized, but weakly and often off target. For the first, and larger, field of regional innovation networks, criticism comes in three ways. First, it is asserted that it excludes nonregional levels of analysis, a completely untrue claim. Second, that, in any case, only the state can influence the powers of multinational capital, again not true because, for example, the pharmaceuticals industry depends increasingly on university research and small firm innovation to survive. Third, that RISs are simply a microcosm of the theory of national innovation systems. This is also untrue since the first paper published on the subject reveals the author cited none of the national innovation systems literature, indicating it had not been read at the time of writing. Actor network theory has been criticized because in some variants, inanimate objects, like personal computers, are included as 'actors', something which seems strange, but given the importance of computerized equipment in laboratory research, there would often be no network without the machine. Another critique is that actor network theory does not, as it claims, address power structures in society. However, to the extent that power, and its critique, rely fundamentally upon discourse, then the revelation of that kind of 'false consciousness' performs an important fact in democratizing society.

RIS and Economic Transformation

Modern economies experience growth from successful competition against other economies and their constituent firms, especially when they demonstrate high absorptive capacities for usable, commercializable knowledge. These generate productivity increases and the largest part of firm productivity improvements, hence regional productivity growth, comes from innovation. But markets fail, knowledge is asymmetric, there are knowledge monopolies and oligopolies. Nevertheless, knowledge, even patented knowledge, is largely a publicly available good if not always (e.g., in the case of payment for a license to use someone else's patent) an actual public good. Because knowledge is public, asymmetric and monopolized it is incumbent on public and intermediary bodies to stimulate 'open systems' such as 'open science' and 'open source' software that in turn may facilitate 'open innovation' where large firms subcontract knowledge generation of various kinds to smart SMEs or laboratories, rather than conduct R&D in house. Countries are normally too large, complex, and metropolitan focused, in that, they generally favor asymmetric economic development in their capital cities and are ill equipped to deal with the knowledge complexity of assisting regional knowledge capabilities. In such cases of 'government failure' regions are more than justified, indeed they have the imperative to assist in the moderation of knowledge asymmetries in ways tailored to the regional economic mix, which is always different from that of the metropolis. Networked RISs perform precisely this function. They enable the translation, import, and diffusion of exploration knowledge that can enable innovation to be facilitated by those, mainly firms that are responsible for knowledge exploitation. Increasingly, the translation of such tacit or implicit and raw knowledge into more usable codified or explicit knowledge is conducted by intermediaries who are 'complicit' with the implicit knowledge originator and the explicit knowledge exploiter. Most accounts of top venture capitalists show them having precisely this complicit capability where they recombine (*à la* Schumpeter) tacit, technological knowledge and explicit market risk and innovation knowledge through being complicit with both. So the third intermediating RIS subsystem is the repository of, mainly, complicit knowledge.

RISs are equally important in developed and developing countries. Although few developing country regions have them, by no means all developed country regions do either. Because of asymmetric knowledge problems only a privileged few regions even in the developed world have them. But in developing countries the 'lighthouse' RISs are in Cuba's biopharmaceuticals innovation system whose epicenter is the West Havana Scientific Pole, India's in Bangalore and Hyderabad's software and bioinformatics based innovation systems, and Brazil's – mainly Sao Paolo's advanced bioethanol innovation system, that now even fuels the production of bioethanol with biofuels. China's economic takeoff was not unconnected with Deng Xiao Ping's decentralizing reforms after 1992 and China – like South Korea and Japan – pursues RIS type strategies to a greater extent than hitherto, precisely to escape centralist 'lock ins' that constrain regional innovation networks from evolving into systems. In developed countries, RISs are growing in importance even in industries that organizations such as the Organization of Economic Cooperation and Development (OECD) condemn as low tech, notably the agro food in novation systems combining biofoods (functional foods, nutraceuticals, or plain 'healthy foods'), with conventional but scientifically managed, processed, chilled and hygiene tested foods, and organic foods utilizing no chemicals or poisons in production. In regions like Skåne in southern Sweden, the innovation support system (schools, welfare, and other public institutions) adopted a policy to be organic by 2012. In Zealand, Denmark, the public food innovation consumption system targets at least 75% of the foodstuff used in the municipal kitchens and canteens to be organic before 2009. Both require that there is more attention given by firms to this demand before properly functioning agro food innovation systems are in place. Since, in Sweden as a whole, 80% of total food demand is from public institutions, and in Denmark not much less, such a change in the agro food innovation system profile will happen, or the con ventional food industry is likely to fail in the face of cheaper competition from elsewhere for its dairy and horticultural produce. Most likely, the conventional agro food pro duction subsystem will adapt, as the supermarkets have, to the growing public and private demand for healthy, pref erably organic food instead of the 'large footprint' produce that forms such 'placeless foodscapes'.

RIS and Policy

The most effective RIS operations can be influenced in coordinated markets like those of the Nordic countries by public procurement policy because the public sector accounts for such a large part of total demand. But it does not necessarily happen quickly. Thus the Skåne food cluster, generously funded by the Swedish Innovation Systems Development Agency, initially ignored organic food production and demand, even though its main municipality, Malmö, is committed to a 100% organic school meals policy. Skåne's Food Innovation at Inter faces initiative favors emphasizing the production of functional foods such as health drinks made from oats that purportedly improve the presence of 'good' intes tinal bacteria. Such is the absence of scientific evidence that consumption of such drinks has any health gain ef fect on consumers that a new category of 'health claim approved' was introduced by Sweden's food regulatory body to recognize that a 'claim' for health gain had been officially made, not that it had been scientifically valid ated. This 'innovation system' thus continues to favor biotech and conventional food over organic 'local food networks' despite well documented evidence that much conventional food sold in supermarkets produces nega tive health effects, like obesity, let alone possible avian flu, salmonella, *Escherichia coli* and 'mad cow' effects from its 'Fordist' mass production processes. Of course, these are by no means confined to Sweden; indeed, Sweden's food regulatory system is exemplary and much better, as with all Nordic countries, than in liberal markets such as US and UK. So maybe the 'spontaneity' of a tectonic shift in food markets away from possibly dangerous, con ventional food and toward both public and private pref erence for chemical free, organic food in combination will prove strong influences on RIS network integration and synchronization.

The most important policies to assist RIS develop ment are to reduce political conflicts, especially where they concern human well being, by establishment of an innovation culture that is inclusive, open, and transparent to all actors and institutions in any given region that interact in any way with innovation. In a process of problem identification and policy selection discourse related to innovation, a clear consensus must be elicited for a regional policy vision, supportive associativeness must be animated and leadership from appropriate quarters – not mainly policy – must be mobilized and incentivized to achieve agreed actions swiftly and ef fectively. Policy should inform itself of the newly dis covered importance to economic growth of industries that are neither too specialized (like the alloy golf club head cluster in the Californian desert extolled by Porter in 1998) nor too diversified (can there be much truly innovative discourse between the neighboring pig farmer and iPod manufacturer?), but characterized by 'related variety' where knowledge spillovers and absorptive cap acity between industries are high in the lateral dimen sion. The important task of regional innovation policy is to use the RIS to create platform policies for such inte grated, rapid innovation diffusing innovation 'platforms'.

The onset of Phase 2 Globalization in which not only routine production and services but also knowledge ex ploration, examination, and exploitation are increasingly outsourced to China, India, and other Asian countries more generally, means essentially one thing. With slight exceptions, these countries will remain for a long time at best incremental innovators but in the main, like Japan before them, mainly imitators. Schumpeter predicted this mimicry effect that 'swarms' in clusters as imitators pile into the market following the spawning of innovation. The West, especially the US, will remain a significant, possibly overwhelming source of disruptive, even radical, innov ation. The former is technology destroying, the latter competence destroying. Hence it is incumbent on Western countries to stimulate much 'related variety' in their economies, particularly of a thematic, rather than the old fashioned, sectoral kind – thus security software, bio energy, and human–computer interfaces with high grade networking capabilities rather than the stale old ICT, biotechnology, and nanotechnology trilogy highlighted in most national and sometimes regional science and technology prospectuses. Pervasive, platform based and flexibly specializable innovation is harder to copy than individual innovations like DVD and they are super value adding (i.e., the prospect of six innovations for the price of one). But there must be significant, especially increased private, investment in the predominating sources of complex knowledge, the universities, and their scientists must be better incentivized to innovate or network with innovators and complicit intermediaries, preferably themselves also as entrepreneurs if these RIS based opportunities are to be seized. Good quality, well resourced, and globally benchmarked centers of research excellence will be the key drivers of next generation innovation.

Countries and Regions Implementing RIS Network Strategies

At country level, Sweden, South Korea, and Norway in 2007 had explicit, networked RIS building strategies orchestrated for regions by national innovation agencies. China also, as noted, has had an RIS strategy since the regional reforms of the early 1990s. Indeed China's exponential recent growth cannot be divorced from its RIS experimentation strategy. Surprisingly, Japan has moved only a little in this direction, but will have to do more to escape its current condition of, in effect, economic stagnation. In Europe, the EU has tried to stimulate regional innovation strategies in over 100 regions but the results have been disappointing. Even Germany's BioRegio system building strategy has only been seriously sustainable in Munich, already the county's main biotechnology cluster. The most interesting experiments are bottom up as in Leuven, Belgium, where 20 000 high tech jobs were created in 1998–2006 by pursuit of a related variety platform strategy, and Lahti, Finland, which has independently evolved the same approach involving, for example, nanotechnologies in pervasive applications like energy, drug delivery, and nanobiotechnology bacterial applications. These could be called 'third generation' RIS models, based on bottom up experimentation, communication, and leadership. They are certainly more flexible and agile than first generation IRIS models and maybe even than sectorally specialized second generation ERIS models. In a study of 11 candidate EU IRIS models in 1996–98, only four were found to have all the main IRIS elements listed above – the Basque Country, Wales, BW, and Steiermark (Austria). Of those, the first has performed well in the subsequent decade; the second has gone backward in relative terms, though not absolutely; the third has stagnated; and the fourth augmented its economic performance. Steiermark, Flanders, and Southeast Finland may be the best current technology led RIS exemplars, while East Netherlands (centered on Wageningen), Skåne (despite its problems, it has great potential) along with cross border Copenhagen forming the Øresund region, and 'Food Valley', the cross border Basel Freiburg Strasbourg regional system, are probably Europe's leading agro food RISs. What such regions find is that 'network externalities' and their contribution to economies arise exponentially from a situation where if there is just one computer or 'knowledgeable' actor in a territory, they are zero, to a situation where if there are 1 million the externalities are incalculable. The more RISs there are to translate, transmit, and transceive commercializable knowledge the more knowledge spillovers there must be and the greater opportunity for lateral absorptive capacity among, hopefully, RIS supported innovative platforms. Perhaps the most crucial lesson brought about by the rise of networked regions and RIS type architectures is that more and more firms and policy officials now realize that mere 'learning' is not enough and 'To be innovative means no more copying'.

See also: Actor-Network Theory/Network Geographies; Innovation; Region; Rural Geography.

Further Reading

Aydalot, P. and Keeble, D. (eds.) (1988). *High Technology Industry and Innovative Environments: The European Experience*. London: Routledge.

Bergman, E., Maier, G. and Todtling, F. (eds.) (1991). *Regions Reconsidered: Economic Networks, Innovation and Local Development in Industrialized Countries*. London: Mansell.

Camagni, R. (ed.) (1991). *Innovation Networks: Spatial Perspectives*. London: Belhaven.

Cooke, P. (1985). Regional innovation policy: Problems and strategies in Britain and France. *Environment and Planning C: Government and Policy* 3, 253 267.

Cooke, P. (1992). Regional innovation systems: Competitive regulation in the new Europe. *Geoforum* 23, 365 382.

Cooke, P. (1993). Regional innovation systems: An evaluation of six European cases. In Getimis, P. & Kafkalas, G. (eds.) *Urban and Regional Development in the New Europe*. Athens: Topos New Series.

Cooke, P. (2004). Introduction: Regional innovation systems an evolutionary approach. In Cooke, P., Heidenreich, M. & Braczyk, H. (eds.) *Regional Innovation Systems*. London: Routledge.

Cooke, P. (2005). Regionally asymmetric knowledge capabilities and open innovation: Exploring 'Globalisation 2' a new model of industry organisation. *Research Policy* 34, 1128 1149.

Cooke, P. (2007). *Growth Cultures: The Globalisation of Bioregions*. London: Routledge.

Cooke, P., Boekholt, P. and Todtling, F. (2000). *The Governance of Innovation in Europe*. London: Pinter.

Cooke, P. and Morgan, K. (1993). The network paradigm: New departures in corporate and regional development. *Environment and Planning D: Society and Space* 11, 543 564.

Cooke, P. and Morgan, K. (1998). *The Associational Economy*. Oxford: Oxford University Press.

Grabher, G. (ed.) (1993). *The Embedded Firm: On the Socioeconomics of Industrial Networks*. London: Routledge.

Kenney, M. (ed.) (2000). *Understanding Silicon Valley: The Anatomy of an Entrepreneurial Region*. Stanford: Stanford University Press.

Morgan, K., Marsden, T. and Murdoch, J. (2006). *Worlds of Food*. Oxford: Oxford University Press.

Networks

G. Grabher, University of Bonn, Bonn, Germany

Reprinted by permission of Sage Publications Ltd from Gernot Grabher, Trading routes, bypasses, and risky intersections: mapping the travels of networks between economic sociology and economic geography, Progress in Human Geography 30(2): 1 27 © Sage Publications, 2006.

Glossary

Actor-Network Theory (ANT) This material-semiotic approach focuses on practices of mobilizing networks and conceives networks as hybrid collectifs of human and nonhuman actors.

Embeddedness This notion primarily counters idealized perceptions of markets as anonymous exchange and stresses the role of network relations (relational embeddedness) and social structure (structural embeddedness) for economic behavior.

Governance Approach Rather than with the formal configuration of networks and network positions of individual actors, this approach explores processes through which networks are initiated, coordinated, recombined, and terminated.

Social Network Analysis Behavior and social processes are interpreted with reference to the configuration of networks and the position of actors within networks rather than by the individual attributes of actors.

Structural Equivalence This notion of social network analysis refers to actors who occupy similar positions in a network by having identical ties with other members and predicts that structurally equivalent actors will use each other as a frame of reference.

Structural Holes This concept of social network analysis denotes the separation between nonrendundant contacts (i.e., disconnected clusters) that is expected to yield behavioral opportunities (*tertius gaudens*).

Translation In ANT this notion refers to the four processes (problematization, interéssement, enrolement, and mobilization) through which relationships between human actors, projects, interests, and naturally occurring entities are proposed and enacted.

Networks have turned into an almost universal principle of social organizing. In our current ‘network society’ they form the social grid of deterritorialized flows of information, capital, goods, symbols, and people; businesses are transformed from isolated organizations into nodes within networks of associated suppliers, competitors, and collaborators; technical networks such as those of the Internet surround us within an ever extending ecology of information and communication technologies; the increasing interpenetration of technical and social relations is predominantly conceived as an actor network. Networks, then, are not simply a theoretical concept. They have become metaphor, paradigm, method, and social practice alike.

The rather complex genealogy of networks can be traced back to Georg Simmel’s formal sociology and his concern for relational structures. For much of the second half of the twentieth century, research on networks was driven by the fields of sociology, social psychology, and anthropology. These fields produced a considerable corpus of research on the role of networks in mobilizing social support or reproducing social exclusion, and the key role of networks in the diffusion of information among others.

The enormous upsurge of theoretical interest in networks in the economy during the last few decades was mainly catalyzed by Mark Granovetter’s notion of the social embeddedness of economic action. By stressing the fundamental role of concrete personal relations and the networks of such relations, he provided a key concept for the study of institutional mechanisms by which networks are initiated, coordinated, monitored, recombined, and terminated. The particular Granovetterian understanding of embeddedness in fact turned into a key concept in the human geography of the so called ‘cultural turn’.

This contribution seeks to trace the selective engagement of human geography with three dominant network notions. The paper sets off by depicting the prevailing understanding of the network governance approach. Founded on Granovetter’s relational conception of embeddedness, this approach branches off into the study of different network forms with strategic, informal, and regional networks that are most relevant within human geography.

The paper subsequently briefly ventures into the terrain of the social network approach that so far remained largely uncharted by geography. This strand of network research attempts to explain behavior, rather than by attributes of actors like age, gender, or class, by their position within a network and the particular configuration of networks.

By moving way beyond the two dominant network traditions, the article in the final section moves on to the post structuralist approach of actor network theory (ANT). Although ANT has not reached the status of an alternative to the network governance approach dominant in geography, it has been employed as a conceptual pivot to shift beyond the established dualisms of structure/agency, subject/object, and human/nonhuman.

The Network Governance Approach

Governance, Embeddedness, and Network Forms

The governance approach is chiefly concerned with the institutional and organizational forms that govern exchange relations. Rather than with the formal structure of network configurations, the governance approach more specifically is interested in the institutional mechanisms through which networks are initiated, stabilized, and terminated. The goal of this line of inquiry is to account for the heterogeneity of organizational and institutional arrangements in contemporary economies – ranging from anonymous spot markets to long term collaborative agreements, from informal credit rotating schemes in ethnic enclaves to multinational banks, from tightly knit clusters to elaborate global franchise systems.

The stage for the more recent debate on the governance of exchange relations was set by Oliver Williamson's transaction cost framework in 1985. Williamson asked under what circumstances economic functions are performed within the boundaries of hierarchical firms rather than by market exchanges that cross these boundaries. His answer was that transactions involving high uncertainty about their outcome, that recur frequently, and require substantial 'transaction specific investments' – of money, time, or energy – are likely to take place within hierarchically organized firms. Exchanges that are straightforward and nonrepetitive, requiring few transaction specific investments, should take place across a market interface. This dichotomous view of firms as islands of hierarchical control in a sea of market relations attracted both followers and critiques.

The first most authoritative challenge of this framework was launched when Granovetter addressed the *Problem of Embeddedness*. In the transaction cost approach, under and oversocialized concepts of economic action complement one another. On the one hand, the market resembles the 'undersocialized' conception of an atomized and anonymous exchange, which neglects the role of social relations and local cultures that can provide order in economic life. On the other hand, an 'oversocialized' perception of hierarchical power within firms disregards the role of informal horizontal ties that often transcend vertical lines of hierarchy. In practice, Granovetter categorically stated, economic exchanges are embedded in personal relations and networks that are generating trust and malfeasance. From the early 1990s onwards, this fundamental understanding of the imbrication of the economy within personal relations and social structure was also increasingly taken up in economic geographic reasoning.

The second major line of critique on Williamson focused on the dichotomous view of markets and hierarchies. By conceiving of markets and hierarchies as 'pure' forms, the rapidly expanding spectrum of intermediate network forms ranging from complex multiparty alliances in high tech industries over elaborate subcontracting relations to vague mutual obligations or illegal informal agreements, were initially reduced to 'mongrel hybrids'. More and more, though, networks have come to be seen as a specific mode of governance in the ever extending range between the alleged 'ideal types'.

An approach to systematize networks in the governance tradition is to differentiate between their duration and governance. Different temporal dimensions have important consequences for the types of regulation of network relations. Long term networks, for example, are shaped by reciprocal ties of past experience and expectations emerging from the 'shadow of the future'. The governance of networks spans a broad spectrum from more hierarchical to more heterarchical. In hierarchical networks, like in the supplier pyramid of a car manufacturer, control is exerted by an identifiable center that regulates network practices and rules, such as the selection of network members, the allocation of resources, and the maintenance of network boundaries. In heterarchical networks, like regional networks, for example, the regulation of interactions and relations in contrast is distributed and associative and authority is decentralized. Within this relational space defined by governance on the one hand, duration of networks on the other, informal, strategic, and regional networks represent the most relevant network forms within human geography.

Strategic Networks

Strategic networks are more tightly orchestrated than either informal or regional networks with regard to the selection of network members, the allocation of resources and distribution of revenues, and definition of network boundaries. Ties in strategic networks are forged – in a much more distinct and manifest fashion than in other forms – by power. Strategic networks evolve both horizontally as inter organizational relations among competitors as well as vertically through the interpenetration of intra and inter organizational networks of global corporations.

By focusing on the inter organizational ties between large vertically disintegrating corporations and their suppliers, an earlier debate evolved around the different governance modes of supplier networks. The spectrum of the governance modes of supplier networks ranges from market driven adversarial transactions to collaborative ties of mutual learning. Cross country comparisons, in fact, have revealed that opportunism and trust are highly variable and dependent on specific national and regional institutional settings.

The global commodity chain framework of Gereffi and Korzeniewicz brought multiple geographical scales to the forefront of the analysis. At a time when geo graphical interest had turned to the study of localized production networks, this line of reasoning on com modity chains found its way into geography through Peter Dicken's *Global Shift*. Dicken's seminal piece paved the way for a strand of literature that began to explore the interdependencies between internal and external ties of global corporations and their host economies in more conceptual terms.

This literature increasingly crystallizes around the notion of the global production network that combines insights from the global value/commodity chain analysis with ideas derived from ANT and varieties of capitalism approaches. By drawing on these diverse antecedents, the global production network approach aims to push the conceptualization of territoriality beyond the highly ag gregated distinction of core and periphery rooted in Immanuel Wallerstein's world system framework and to appreciate more systematically the particular insti tutional conditions shaping (regional) development.

While global production networks primarily evolve around vertical ties, strategic alliances typically are confined to horizontal interfirm relations. Over the last two decades joint ventures have increasingly been re placed by nonequity partnerships, such as R&D pacts and joint development agreements that involve a lower de gree of organizational dependence and shorter time horizon. Presumably due to their nonlocal and ephemeral character, strategic alliances have hardly appeared in the focus of geographic analysis.

Informal Networks

Membership in an informal network is typically based on shared experience or the thick bonds of kinship and ethnicity that draw participants together. Given that such relationships emerge out of repeated exchanges, informal networks symptomatically involve long time horizons. Research on informal networks has portrayed this type of networks rather ambivalently. Whereas one strand of research pays tribute to the compensatory role and 'lubricating' effects of informal ties, a less prominent line of inquiry is primarily interested in exposing the po tentials of informal networks to disrupt markets and obstruct hierarchies.

The compensatory role of informal networks has been demonstrated in settings that appear to operate likepurely price driven markets, such as the financial markets of the Wall Street or the City of London, for example. Yet, traders in these market settings often es tablish continuing relationships based on reciprocity to validate information and exchange rumors and gossip. Human geography has significantly contributed to this line of reasoning, both conceptually and empirically.

Presumably, the most vivid contrast to the sterile image of spot contracting in pure market settings has been revealed in research on ethnic networks that pro vide the social infrastructure of enclaves and ethnic economies. Ethnic ties provide solidarity and trust that can be enforced in relatively closed networks. The amalgamation of these collective assets in the notion of 'social capital' subsequently tended to privilege the en abling attributes of informal networks. This focus on the short term benefits of social capital glanced over earlier accounts that illuminated how informal networks might turn into mobility traps over time.

Despite the imminent spatiality of ethnic networks and enclaves the geographical community initially has been less involved in the relevant conceptual debates. The more lively cross disciplinary exchange in recent times has moved beyond the emblematic genres of geography such as the idiosyncratic portrayals of the place based nature of local enclaves and ethnic econ omies to explore the transnational character of migration and 'transmigrants', ethnic networks, or global elite networks.

Informal networks, however, are also particularly well suited for practices ranging from conspiracy to organized crime in violent Mafia type networks or terrorist net works. In fact, the risks associated with crime make trust far more necessary among criminals than among busi nessmen. Geography largely remained absent from this strand of research and once more narrowed the focus to the more benevolent and functional dimensions of informality.

Like in market environments, informal ties in hier archical contexts fulfill multiple roles. On the one hand, informal networks can be mobilized to organize resist ance against hierarchical rule. On the other hand, they can compensate for the structural weaknesses of hier archies when they, for example, constitute 'communities of practice' that provide decentralized means of learning. The notion of communities of practice has rather rapidly migrated into a broad range of disciplinary contexts and, in the wake of the intertwined debates on the knowledge economy and learning, has been increasingly employed in economic geography.

Regional Networks

Regional networks typically evolve over even longer time periods but share the multiplex features of informal networks. They are, in other words, connected through a variety of intertwined heterarchical social and economic relations that make up their embeddedness. The early accounts on the amalgamation of family, community, polity, and business within tightly knit localities in the Third Italy provided most vivid embodiments of the notion of embeddedness. Although the Italian industrial districts, due to their dense familial and political bonds, represent an atypical case of collective governance, they provided an emblematic point of reference for the unexpected success of regional networks. These small firm networks obviously held the promise to combine economic imperatives of efficiency, innovativeness, and resilience with a sense of economic democracy and social fairness.

Subsequently, other thriving regions, most notably Baden Württemberg and Silicon Valley, were regarded as empirical proof that resilient regional networks can be based on a variety of mechanisms and relational configurations that foster reciprocal forms of exchange. Networks, however, were rarely theorized in an explicit fashion, but rather, envisioned mainly in a rather generic sense as a shorthand for all sorts of lasting ties that did not adhere to a straightforward market logic. Moreover, the reification of regional networks in a plethora of case studies tended to adhere to 'overterritorialized views on embeddedness'. Initially productive conceptions of territorial innovation models like innovative milieus, learning regions, regional innovation systems, and, above all, the cluster that had branched out from the district debate more and more seemed to constrict the view on regions to isolated 'islands of innovation'.

More recently, however, earlier attempts to break out of what increasingly appeared as the straitjacket imposed by Marshallian analysis gained momentum. The extent to which regional networks are integrated into global exchange relations has profound impacts on their internal governance, as well as their ability to tap into wider pools of information and relational resources. Trans regional relations, moreover, have been identified as rather robust pivots against the self reinforcing dynamics of closure and lock in.

The current reaffirmation of nonlocal ties has been mainly phrased in the conceptual terms drawn from Owen Smith and Powell's instructive distinction between 'local broadcasting' and 'global pipelines'. This conceptual step from the local into the nonlocal appears long overdue, yet it still seems rather distant from a topological understanding of space that allows an understanding of individual sites "as a place of trans scalar and non linear connections, and as a relay point of circulating knowledges that cannot be territorially attributed with any measure of certainty or fixity." (Amin and Cohendet 2004: 154). Most recent attempts, however, have begun to push beyond the static and problematic local–global duality in which different spatial scales are associated with distinct social relations, behavioral attitudes, and types of knowledge. Local ties, in other words, are no longer necessarily regarded as socially embedded relations bound up with tacit knowledge surrounded by a sparse array of disembedded ties confined to the exchange of explicit knowledge.

Some of the structural limitations of the regional network debate, above all its privileging of cohesion and benevolent trust based ties, reflect the willing engagement of economic geography with the governance approach. Human geographic imaginations of networks hardly appreciated the second, in fact older, tradition of social network analysis in a more systematic fashion. Presumably, human geography was deterred from venturing more deeply into this terrain by the austere and formalistic style of the social network approach that appeared to clash with the predilection for qualitative approaches of the 'cultural turn'.

Social Network Analysis

The hallmark of social network analysis is to account for "the behavior of network elements (i.e., the nodes) and of the system as a whole by appeal to specific features of the interconnections among the elements" (Laumann 1979: 349). The austere style of theorizing in social network analysis is deliberate and reflects the chief strategy to avoid the 'traps' of categorical thinking. Social network theory attempts to institute a relational mode of analysis and to challenge the unexamined construct of the person as an entity characterized by the typical list of variables of interest in social science – age, race, class, gender, etc. – that are conceived as sole causal factors of behavior.

Resonating with Simmel's program of a formal sociology of relations, networks represent sets of actors linked through specific types of connections. As a most basic representation of such affiliations, the formal language of the nodes and lines of the 'sociogram' became an emblematic representation of social network analysis. Over the last decades, the basic formal language of the sociogram or 'graph' has been translated into software tools (e.g., Inflow, KrackPlot, Pajek, UCINET) to analyze and depict features of relationships. In human geography, only a few elements of the conceptual toolkit of social network analysis have been employed while the explanatory power of others is yet to be more critically gauged.

Strength of Ties

The strength of ties is presumably the most frequently adopted concept of social network analysis within human geography. Granovetter's elegant simplicity of the 'strength of weak ties', in particular, has consistently been referred to in human geography and currently seems to attract some renewed interest. Put simply, this principle states that whatever is to be diffused can reach a larger number of network members, and traverse greater social distance (i.e., path length), when passed through weak rather than strong ties. Conversely, the information received in strong tie networks is likely to be stale information, already re ceived from other members of their own social group. Weak ties, in Granovetter's understanding, score rather low with regard to the amount of time, emotional intensity, intimacy (mutual confiding), and reciprocal services which characterize those relationships.

In exploring the significance of weak ties, Granovetter also refers to 'marginal' actors and 'outsiders' who seem to play a crucial role in the first phases of the diffusion of information and innovation. It seems to take only a small analytical step from Granovetter's sociological notion of the 'outsider' to Park's iconic 'marginal man' or Simmel's emblematic 'stranger'. These personifications of urbanity hardly played a noticeable role in the accounts of suc cessful regional networks. The local has rather been portrayed in terms of the social cohesion of a village than of the diversity of a city. Rephrased in Ferdinand Tön nies' terms, locality is seen as *Gemeinschaft* (community) rather than as *Gesellschaft* (society).

More recently, the strength of network ties appeared in the focus of economic geographic inquiry in the re appraisal of translocal ties. Although this appreciation of weak ties took geographic theorizing a crucial step for ward, the straightforward association of weak ties with the local scale has also caused severe objections that deny causality between spatial scale and density of ties.

Structural Equivalence

Although the notion of structural equivalence hardly has been employed explicitly in human geography, it reson ates with more recent debates on innovation in economic geography. The idea of structural equivalence challenges the more familiar line of reasoning that perceives the adoption of innovation as a result of frequent interactions and emphatic communication between network members. Structural equivalence, broadly conceived, occurs when two actors occupy similar positions in a social system by having identical ties with other network members.

Structural equivalence predicts that actors identically positioned in the flow of communication will use each other as a frame of reference even if they have no direct communication with each other. In other words, it is not through intense exchange between network members but rather through the perception of the proper action for an occupant of a specific network position that diffusion is primarily driven. Innovation in this perspective, in other words, is stimulated primarily by mimetic pressures of mutual comparison.

By revolving around dense webs of local linkages, geographical innovation models apparently tend to overemphasize the dynamics of cohesion at the expense of structural equivalence. More recently, a critical line of reasoning that resonates with the structural equivalence argument has been voiced pointing to the crucial role of mutual awareness and observation in stimulating regional innovation. Among the pioneers that explored this par ticular path in economic geography, Malmberg and Maskell in 2002 pointed to the relevance of observation and imitation more generally.

Tertius Gaudens and Structural Holes

A further key principle in social network analysis alerts to the vital importance of the network position of indi vidual actors for understanding entrepreneurial behavior. An actor that is positioned between two unconnected actors can leverage off a stable entrepreneurial decision as the *tertius gaudens* (the third who benefits) by divulging and brokering contradictions and tensions between pre viously unconnected actors. Burt in 1982 has further built on this classical strategy of 'divide and rule' in his dis cussion of how actors who connect two others previously unknown to each other bridge a 'structural hole'. These bridges represent unique, nonredundant ties between networks that otherwise would remain separated. By bridging a structural hole, an actor is able to look at a wider information screen and for that very reason be comes an even more attractive network contact to other actors thus providing new opportunities to expand net work contacts.

This understanding of entrepreneurship that flows from exploiting ambivalence is radically different from the incarnation of the Schumpeterian entrepreneur pre vailing in current economic geographic debates. While networks in economic geographic narratives are pre dominantly portrayed as webs of trustful ties to curb opportunism and to engender cooperation and innov ation, networks emerge from the *tertius gaudens* argument rather as vehicles to pursue opportunistic behavior and to produce competition. The iconographic economic geo graphic accounts are preoccupied with sharing, trust, and mutual learning; the *tertius gaudens* viewpoint focuses on dividing, arbitrage, and strategic games. Only few ac counts so far have begun to conceptualize the inter mingling of diverse social rationalities such as family, friendship, firm, church community, etc., within network relations.

Burt's interpretation of structural holes is presumably the most widely cited notion of social network analysis in human geography and, in fact, is regarded as the em blematic representation of this network tradition. Social network analysis, however, encompasses a diverse range of approaches that has advanced beyond 'structuralist determinism' to address agency in network structures. More significant steps beyond deterministic perceptions of network structure have been taken by Harrison White's seminal writings on identity, temporality, and narrative. Formerly a leading exponent of structural determinism, White began to push towards a recon ceptualization of the interrelation between network structures, culture, and agency.

Actor-Network Theory

From Network to Rhizome

Notwithstanding these major steps beyond determinism undertaken by 'structural constructivists' like White, this strand of social network analysis still continues along a structuralist tradition. A fundamentally different ontol ogy rooted in post structuralism is offered by ANT. ANT is a material semiotic approach developed by Michel Callon and Bruno Latour in Paris in collabor ation with John Law and others in the context of science and technology studies.

Both the network governance as well as the social network analysis adhere to a network construal rooted in a generic topographic imagination of social relationships that connect social actors. This grid of ties and nodes is contrasted in (later contributions to) ANT with the trope of the rhizome. The rhizome offers an alternative trope that perceives networks as multiple, intertwined, and branching roots with no central axis, no unified point of origin, and no given direction of growth. The botanical associations indeed seem intended: The metaphor of the rhizome foregrounds the transformative and processual dimension of networks; it deliberately departs from the static views of 'transport without transformation' in the dominant network approaches.

Relationality and Translation

This perception of multiple entanglements in ANT is intended to subvert our habitual juxtapositions of struc ture/agency, subject/object, and human/nonhuman. In the ontology of ANT, actor and network are inextricably enmeshed. An actor network is simultaneously an actor whose activity is networking heterogeneous elements and a network that is able to redefine and transform the constituent elements of which it is made of. The capacity to act and give meaning to action is neither solely em bodied in human actors nor localized in norms, values, and institutions that make up our familiar registers of 'social embeddedness'. In this perspective, we do not live in a society but in 'hybrid collectifs' that entangle human actors as well as nonhuman actors in multiple ways.

The point here is not to anthropomorphize insentient beings and things but to appreciate their particular role in collectifs: they stabilize. Whilst fixed in one sense, 'immutable mobiles' such as maps, texts, algorithms, or ganization charts or ships, canon balls, and navigation instruments are also re combinable. Recombination and alignment is achieved through the process of 'translation' that is a process in which sets of relationships between human actors, projects, tools, and naturally occurring entities are proposed and enacted. As Callon in 1986 elaborated, translation is achieved in four moments. Problematization involves the interdefinition of actors and the attempts of actors to establish 'obligatory passage points' in order to become indispensable network elem ents; interéssement confirms the validity of the prob lematization and the alliances it implies; enrolement describes the negotiations, trials, and tricks that enable enrolement to succeed; mobilization of allies, eventually, turns enrolment into active support.

ANT in Human Geography

Although the rhizome metaphor up to now has not dif fused into human geography on a similarly broad front as the tie and node imagery, it nevertheless exerts in creasing influence on human geographic imaginations of networks and space. First and more generally, ANT perforates the analytical distinction between practice and its scientific representation. The relational webs that constitute the economy, for example, not only comprise of the familiar catalogue of nodes such as firms, con sumers, and various institutions – but also of the economists who contribute through calculative practices and conceptual tools to the performance of the reality they describe: The economy is embedded in economics.

Second, the governance approach offered to human geography a model for conceiving (or at least implicitly assuming) the fabrics of socioeconomic life that could be assorted neatly onto different scalar levels (from local through regional to global); geographical notions of space themselves, though, remained largely untouched by this network metaphor. ANT radically breaks away from the Euclidean scalar understanding to a genuine relational perception of space as topological stratifications. This approach "opens up space time to the coexistence of multiple cross cutting networks of varied length and durability" (Whatmore and Thorne 1997: 302). ANT, in other words, follows the pleas against essentialist understandings of space and time. A topological geog raphy interweaves time and space with a heterogeneous network of actors that has been differentiated, for ex ample, into regions, networks, and fluid spaces.

Third, the methodological strategy of ANT helps to identify and embody the agents and networks of the global economy in a way that undermines the image of the "faceless juggernaut of globalization" (Dicken 2001: 106). By appreciating the multiplicity of interrelated processes in the constitution, stabilization, and reshaping of relational ties, ANT challenges the perception of a seamless and lucid global chain that links producer and consumers in a straightforward fashion. In contrast to the more functionalist understandings of global commodity chains, ANT inspired perspectives foreground the fibrous links that have no linear or bounded character and that continually evolve into new constellations in unforeseen ways. Instead of conceiving individuals, firms, industries, or nation states as coherent and unitary actors, ANT directs the focus to their constitution and reshaping via tracing their engagement within an array of actor net works. Studying a different type of global networks though from a similar point of view, Whatmore and Thorne in 1997 used ANT to guide their research on the making of durable 'fair trade' actor networks that link buyers in the North with producers in the South.

Fourth, the perception of networks that challenge the established demarcations between human/nonhuman opened up novel avenues to delve into 'hybrid geog raphies', in which nature, for example, is no longer per ceived as the traditional passive object but rather ascribed an acting role. ANT thus can perforate analytical de marcations that have become 'naturalized' in our pre vailing lines of reasoning in a productive fashion.

For this very reason, however, the underlying imagery of the rhizome can also turn into a trap when all too arbitrarily transplanted in any context. ANT inspired studies tend to privilege the relational dimensions of the web at the expense of considerations of the actors them selves. ANT seems to offer an invitation to glance past the differences between distinct types of actors in different domains and thereby also conceals uneven power relations. Although power (in its Foucauldian understanding) is by no means an alien concept to ANT in principle, in practice ANT misses that hierarchies are real.

Conclusion

This paper set out at tracing the engagement of human geography with the three most relevant network con ceptions of the network governance approach, social network analysis, and the post structuralist ideas of ANT.

Human geography, first and foremost, embarked pri marily on the route paved by the network governance approach. Far from being exhausted, the governance approach can carry human geographic reasoning on networks farther on. This seems a particularly fertile ground for cross disciplinary exchange with adjacent social sciences. With the increasing spatial awareness in the social sciences, as manifest in the debates on the resilience of national institutions, the '(re)discovery' of local clusters or the importance of micro geographies for learning in laboratories, the conditions, in principle, seem favorable. However, instead of confining the role of geography to attaching the proper spatial scale (along the familiar scale of local to global) to network structures, human geography might champion more challenging problematizations of space into the social sciences.

Second, apart from fairly loose interpretations of the strength of network ties and structural holes, human geography hardly engaged in the systematic inquiry of network structures and positions. In particular, more recent approaches of the structuralist constructivists like Harrison White have not (yet) been critically interro gated by human geography. Unlike in structuralist de terminism, networks in the constructivist perspective afford opportunities for action, though are but one antecedent of action. In this sense, human geography might benefit from examining the potentials of social network analysis. Human geography is strong on pro cessuality and contingency but weak on structure; social network analysis is strong on structure and, apparently, increasingly sensitive to processuality and contingency.

Third, human geography has already ventured into the terrain of the post structuralist strand of ANT. For human geography the proposition that action takes place in 'hybrid collectives' could imply, among others, to more systematically appreciate materiality of the social and the "diverse props … that sustain people's actions" (Strath ren 1996: 523). Studies of financial markets, for example, have exemplified the entanglements of actors with tools, instruments, and algorithms in an instructive fashion. Computer monitors that are used to 'screen' the markets are the very locations of markets on which trading is performed. Geography with its tradition in studying the material world seems in a privileged position to con tribute to a 'science of associations'.

See also: Commodity Chains; Embeddedness; Ethnic Economies; Food Networks; Global Production Networks; Industrial Districts; Poststructuralism/Poststructuralist Geographies; Regional Production Networks; Relational Economic Geography.

Further Reading

Allen, J. (2000). Power/economic knowledges: Symbolic and spatial formations. In Bryson, J., Daniels, P. W., Henry, N. & Pollard, J. (eds.) *Knowledge, Space, Economy*, pp 15 33. London: Routledge.
Amin, A. and Cohendet, P. (2004). *Architectures of Knowledge: Firms, Capabilities and Communities*. Oxford: Oxford University Press.
Amin, A. and Thrift, N. (1992). Neo Marshallian nodes in global networks. *International Journal of Urban and Regional Research* 16, 571 587.
Arlachi, P. (1986). *Mafia Business: The Mafia Ethic and the Spirit of Capitalism*. London: Verso.
Barbási, A. L. (2002). *Linked: The New Science of Networks*. Cambridge, MA: Perseus.
Bathelt, H., Maskell, P. and Malmberg, A. (2004). Clusters and knowledge: Local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography* 28, 31 56.

Becattini, G. (1978). The development of light industry in Tuscany: An interpretation. *Economic Notes* 2(3), 107 123.
Brown, J. S. and Duguid, P. (1991). Organizational learning and communities of practice. *Organization Science* 2, 40 57.
Bunnell, T. G. and Coe, N. M. (2001). Spaces and scales of innovation. *Progress in Human Geography* 25(4), 569 589.
Burt, R. S. (1982). *Structural Holes*. Cambridge, MA: Harvard University Press.
Burt, R. S. (1987). Social contagion and innovation: Cohesion versus structural equivalence. *American Journal of Sociology* 92, 1287 1335.
Burt, R. S. (1992). *Structural Holes*. Cambridge, MA: Harvard University Press.
Callon, M. (1986). Some elements of a sociology of translation: Domestication of the scallops and the fisherman of St. Brieuc Bay. In Law, J. (ed.) *Power, Action and Belief: A New Sociology of Knowledge?* pp 196 223. London: Routledge.
Callon, M. (1998). *The Laws of the Market*. Oxford: Blackwell.
Callon, M. and Law, J. (1995). Agency and the hybrid collectif. *The South Atlantic Quarterly* 94(2), 481 508.
Camagni, R. (1991). *Innovation Networks: Spatial Perspectives*. London: Belhaven Press.
Castells, M. (1996). *The Information Age: Economy, Society and Culture 1: The Rise of the Network Society*. Oxford: Blackwell.
Coe, N. M., Hess, M., Yeung, H. W. C., Dicken, P. and Henderson, J. (2004). Globalizing regional development: A global production networks perspective. *Transactions of the Institute of British Geographers* 29(4), 468 484.
Dicken, P. (2003). *Global Shift: Mapping the Changing Contours of the World Economy*. London: Sage.
Dicken, P., Kelly, P. P., Olds, K. and Yeung, H. W. C. (2001). Chains and networks, territories and scales: Towards a relational framework for analysing the global economy. *Global Networks* 1(2), 89 112.
Dicken, P. and Thrift, N. (1992). The organization of production and the production of organization. *Transactions Institute of British Geographer* 17, 279 291.
Dillon, M. (2002). Network society, network centric warfare and the state of emergency. *Theory, Culture and Society* 19(4), 71 79.
Emirbayer, M. and Goodwin, J. (1994). Network analysis, culture, and the problem of agency. *American Journal of Sociology* 99(6), 1411 1454.
Ettlinger, N. (2003). Cultural economic geography and a relational and microspace approach to trusts, rationalities, networks, and change in collaborative workplaces. *Journal of Economic Geography* 3, 145 171.
Faulconbridge, J. R. (2007). Relational networks of knowledge production in global law firms. *Geoforum* 38, 925 940.
Gereffi, G. and Korzenienwicz, M. (eds.) (1994). *Commodity Chains and Global Capitalism*. Westport, CT: Praeger.
Gertler, M. S. (2003). Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). *Journal of Economic Geography* 3, 75 99.
Grabher, G. (1993). Rediscovering the social in the economics of interfirm relations. In Grabher, G. (ed.) *The Embedded Firm*, pp 1 33. London: Routledge.
Grabher, G. (1993). The weakness of strong ties: The lock in of regional development in the Ruhr area. In Grabher, G. (ed.) *The Embedded Firm*, pp 255 277. London: Routledge.
Grabher, G. (2006). Trading routes, bypasses, and risky intersections: Mapping the travels of 'networks' between economic sociology and economic geography. *Progress in Human Geography* 30(2), 1 27.
Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology* 91, 481 501.
Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology* 78, 1360 1380.
Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology* 91, 481 501.
Haraway, D. (1997). *Modest Witness@Second Millenium. FemaleMan© Meets OncoMouse™*. London: Routledge.
Hedlund, G. (1986). The hypermodern MNC: A heterarchy? *Human Resource Management* 25(1), 9 35.
Herrigel, G. B. (1996). *Industrial Constructions: The Sources of German Industrial Power*. New York: Cambridge University Press.
Hess, M. (2004). 'Spatial' relationships? Towards a reconceptualization of embeddedness. *Progress in Human Geography* 28(2), 165 186.
Hess, M. and Yeung, H. W. C. (2006). Whither global production networks in economic geography? Past, present and future. *Environment and Planning A* 38(7), 1193 1204.
Knorr, C. K. and Bruegger, U. (2002). Global microstructures: The virtual societies of financial markets. *American Journal of Sociology* 107(4), 905 950.
Knox, H., Savage, M. and Harvey, P. (2006). Social networks and the study of relations: Networks as method, metaphor and form. *Economy and Society* 35(1), 113 140.
Latour, M. (1987). *Science in Action*. Cambridge, MA: Harvard University Press.
Latour, M. (1999). On recalling ANT. In Law, J. & Hassard, J. (eds.) *Actor Network Theory and After*, pp 5 15. Oxford: Blackwell.
Latour, M. (2005). *Reassembling the Social: An Introduction to Actor Network Theory*. Oxford: Oxford University Press.
Laumann, E. O. (1979). Network analysis in large social systems: Some theoretical and methodological problems. In Holland, P. W. & Leinhard, S. (eds.) *Perspectives in Social Network Research*, pp 379 402. New York: Academic Press.
Law, J. and Hassard, J. (1999). *Actor Network Theory and after*. Oxford: Blackwell.
Leitner, H., Pavlik, C. and Sheppard, E. (2002). Networks, governance, and politics of scale: Inter urban networks and the European Union. In Herod, M. & Wright, M. W. (eds.) *Geographies of Power, Placing Scale*, pp 274 303. Oxford: Blackwell.
Light, I. (1998). Afterword: Maturation of the ethnic economy paradigm. *Urban Geography* 19(6), 577 581.
Lorrain, F. and White, H. C. (1971). The structural equivalence of individuals in social networks. *Journal of Mathematical Sociology* 1, 49 80.
Malmberg, A. and Maskell, P. (2002). The elusive concept of localization economies: Towards a knowledge based theory of spatial clustering. *Environment and Planning A* 34, 429 449.
Massey, D., Allen, J. and Sarre, P. (1999). *Human Geography Today*. Cambridge: Polity Press.
Mol, A. and Law, J. (1994). Regions, networks and fluids: Anaemia and social topology. *Social Studies of Science* 24, 641 671.
Murdoch, J. (1998). The spaces of actor network theory. *Geoforum* 29, 357 374.
Nohria, N. and Ghoshal, S. (1997). *The Differentiated Network: Organizing Multinational Corporations for Value Creation*. San Francisco: Jossey Bass.
Obstfeld, D. (2005). Social networks, the tertius iungens orientation, and involvement in innovation. *Administrative Science Quarterly* 50, 100 130.
Owen Smith, J. and Powell, W. W. (2004). Knowledge networks as channels and conduits: The effects of spillovers in the Boston biotechnology community. *Organization Science* 15(1), 5 21.
Piore, M. and Sabel, C. (1984). *The Second Industrial Divide*. New York: Basic Books.
Podolny, J. M. and Page, K. L. (1998). Network forms of organization. *Annual Review of Sociology* 24, 57 76.
Powell, W. W. (1990). Neither market nor hierarchy: Network forms of organization. *Research in Organizational Behaviour* 12, 295 336.
Putnam, R. (1993). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, NJ: Princeton University Press.
Sako, M. and Helper, S. (1998). Determinants of trust in supplier relations. *Journal of Economic Behavior and Organization* 34, 387 417.
Sanders, J. M. and Nee, V. (1996). Immigrant self employment: The family as social capital and the value of human capital. *American Sociological Review* 61, 231 249.
Sassen, S. (2002). Global cities and diasporic networks: Microsites in global civil society. In Anheier, H., Glasius, M. & Kaldor, M. (eds.) *Global Civil Society*, pp 217 238. Oxford: Oxford University Press.
Saxenian, A. (1994). *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.

Scott, J. (2000). *Network Analysis: A Handbook* (3rd edn.). London: Sage.
Simmel, G. (1992). *Soziologie. Untersuchungen über die Formen der Vergesellschaftung (first publ. 1908)*. Leipzig: Duncker und Humblot.
Smith, D. P. and Bailey, A. J. (2004). Special Issue: Linking Transnational Migrants and Transnationalism. *Population, Space and Place* 10(5), 357 360.
Smith Doerr, L. and Powell, W. W. (2003). Networks in economic life. In Smelser, N. & Swedberg, R. (eds.) *The Handbook of Economic Sociology*, pp 379 402. New York: Russell Sage Foundation.
Strathren, M. (1996). Cutting the network. *Journal of the Royal Anthropological Institute* 2, 517 535.
Thrift, N. (2000). Rhizome. In Johnston, R., Gregory, D., Pratt, G. & Watts, M. (eds.) *The Dictionary of Human Geography*, pp 716 717. Oxford: Blackwell.
Tonnies, F. (1979). *Gemeinschaft und Gesellschaft: Grundbegriffe der reinen Soziologie* (8th edn.) (first publ. 1887). Darmstadt: Wissenschaftliche Buchgemeinschaft.
Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly* 42, 35 67.
Wasserman, S. and Faust, K. (1994). *Social Network Analysis*. Cambridge: Cambridge University Press.
Watts, D. J. (1999). *Small Worlds*. Princeton: Princeton University Press.
Wellmann, B. (1983). Network analysis: Some basic principles. In Collins, R. (ed.) *Sociological Theory*, pp 155 200. San Francisco: Jossey Bass.
Whatmore, S. (2002). *Hybrid Geographies. Natures, Cultures and Spaces*. London: Sage.
Whatmore, S. and Thorne, L. (1997). Nourishing networks: Alternative geographies of food. In Goodman, D. J. & Watts, M. J. (eds.) *Globalizing Food: Agrarian Questions and Global Restructuring*, pp 287 304. London: Routledge.
White, H. C. (1992). *Identity and Control: A Structural Theory of Social Action*. Princeton, NJ: Princeton University Press.
White, H. C. (2002). *Markets from Networks: Socioeconomic Models of Production*. Princeton, NJ: Princeton University Press.
Williamson, O. E. (1985). *The Economic Institutions of Capitalism*. New York: Free Press.
Yeung, H. W. C. (2005). Rethinking relational economic geography. *Transactions of the British Institute of Geographers* 30, 37 51.

Networks, Urban

N. Clarke, University of Southampton, Southampton, UK

Glossary

Heterophilic Network A spatial arrangement comprising nodes characterized by their difference from one another.

Homophilic Network A spatial arrangement comprising nodes characterized by their similarity to one another (see also glossary terms 'social capital' and 'weak tie').

Network A spatial arrangement comprising nodes connected by routes. Such nodes are usually arranged without linear direction (so as to distinguish networks from chains). They are sometimes arranged without one obvious center (so as to distinguish networks from webs).

Primary Contact A relationship between individuals (or the individuals themselves) exhibiting relative permanence, often founded in kinship, and usually broadly supportive.

Secondary Contact A relationship between individuals (or the individuals themselves) usually providing specialized support, and often mediated by money, calculation, and/or legislation.

Social Capital The value of social relations, evident in how they bond individuals together into communities of common interest (bonding social capital), and how they bond communities of common interest together into societies (bridging social capital).

Social Network A spatial arrangement comprising individuals connected by contacts, relationships, or ties, along which tend to flow resources and opportunities, and also norms, claims, demands, and expectations.

Strong Tie A relationship between individuals (or the individuals themselves) characterized by relative formality, and usually characterized by much overlap between the social networks of each individual (see also 'primary contact').

Weak Tie A relationship between individuals (or the individuals themselves) characterized by relative informality, and usually characterized by little overlap between the social networks of each individual, so that new information and opportunities often flow through the relationship.

Introduction

For much of the twentieth century, urban networks was a term used by urban sociologists and others to describe the networks of social support carried and translated into urban environments by rural–urban migrants, and constructed anew by metropolitans. These urban networks are discussed in the first section of this article. From sometime in the 1980s, urban networks became a term used by urban geographers and others to describe a number of more or less connected things, from archipelagos of world or global cities, through assemblages of social and technical actors that give urban life its apparent order/disorder, to new forms of social support arising in the first few years of the twenty first century. These urban networks are discussed in the second section of this article.

Urban Social Networks

From the very beginning, one of the central claims of sociology has been that social networks are important. Among other places, this claim was originally substantiated in Emile Durkheim's renowned study of suicide, completed in the last few years of the nineteenth century. Durkheim analyzed suicide rates in relation to various measures of alienation, race, heredity, climate, and imitation – the usual suspects of the time – but found little correlation. Where he did find correlation, however, was between suicide rates and various measures of social integration. He found that suicide rates were lower among Catholics, because Catholicism emphasized ties between the individual and the group (in a way that Protestantism did not). They were lower at times of crisis, because individuals pulled together at such times. In addition, they were lower among men, because women lacked the opportunities for collective existence afforded to men.

More recently, the claim that social networks are important has been substantiated by sociologists working predominantly in North America in the field of social capital (a term used to describe the value of social relations). For these analysts, social networks represent bonding social capital, since they bond individuals together into communities of interest. They also represent bridging social capital, since they bond communities of interest together into societies. In this literature, social networks, and the social capital they represent, explain variations in educational attainment and child welfare, because social networks are supportive of children, not least because they deliver financial and other resources. They explain variations in neighborhood safety and productivity, because they enable positive enforcement of standards among residents, and offer younger residents access to mentors and employment contacts. They explain variations in economic prosperity, because advice, information, techniques, recommendations, and the kind of generalized trust that leads to reduced transaction costs all circulate through social networks. Finally, in this literature, and also in broader public debate, social networks explain variations in democratic health: because they enable individuals to express their interests to, and to protect themselves from abuses by, governments; because individuals learn the practical skills necessary for public life through their involvement in social networks; and because social networks serve as forums for deliberation over public issues.

The claim, then, is that social networks are important for individuals, communities, and societies. In addition to this, however, the claim has also been that social networks are important for sociologists, anthropologists, psychologists, and others, in that social networks allow such analysts to move between the microlevel (of individuals, small groups, and interaction) and the macrolevel (of organizations, structures, and patterns). We can identify such a role for social networks in Alfred Radcliffe Brown's classical notion of social structure, which used the metaphors of fabric and web to describe social life. We can identify such a role in Jacob Moreno's influential methodology of sociometry, which constructed and represented the opportunities for, and constraints upon, personal psychological development, and their delivery through group relations. We can identify such a role for social networks in Clyde Mitchell's seminal theory of group organization – his concepts of personal order (the links an individual has with a set of people, and the links these people have among themselves), total network (the links among members of a community or organization, and also the links that cut across the boundaries of such a community or organization), and partial network (the links centered on an individual – as in the personal order – or the links centered on a particular social activity, such as kinship). On the back of these notions, methodologies, theories, and concepts, mostly formulated around the middle of the twentieth century, numerous studies from the last 60 or so years have taken social networks as their starting point and unit of analysis. These include Studies of disease transmission by general systems theorists; studies of group structure, solidarity, and norms by factory and community analysts; and studies of both African tribal societies and British small towns and villages by social anthropologists.

In this social network tradition, a prominent point of focus has been the relationship between urbanization and social networks. As early as 1887, Ferdinand Tonnies expressed his concern that rural social networks – founded in permanency of abode, characterized by sympathetic relationships between kinsfolk and old acquaintances – were threatened by urbanization, associated geographical mobility, and associated relationships mediated by money, calculation, and legislation. Similar concerns were common among urban sociologists of the early twentieth century, and especially Georg Simmel and Louis Wirth. Cities were defined by their size, heterogeneity, density, and consequent intense stimulation of the nervous system. Metropolitans were defined by their social technologies: clocks and watches, by which meetings took place; money and calculation, by which exchange took place. They were defined by their blasé attitudes and spatial segregation – both attempts to avoid internal atomization, given the overwhelming number and diversity of potential relationships, and associated claims and expectations, brought forth by the city. Most importantly, for the present article, they were defined by their poverty of primary contacts, and their abundance of secondary contacts. Though such a portfolio of contacts could be experienced as liberating by some people – liberating of individuality, for example – more often than not, it was assumed by urban sociologists to be debilitating for both persons and their communities.

We return to secondary contacts and liberating experiences below. Before that, we turn to the concerns of mid twentieth century urban sociologists, as they focused on the relationship between urbanization and social networks. More specifically, we consider two renowned studies of post war slum clearance and urban renewal, both of which raised questions about the position of Tonnies and others with regard to urbanization and social network dissolution/survival. In the late 1940s and early 1950s, Michael Young and Peter Willmott studied kinship networks in the London Borough of Bethnal Green. They found that such networks were very much alive in East London because, among other things, mothers and daughters continued to live in close proximity to one another, as they continued to share the experiences, interests, and associations of child rearing. Young and Willmott found that such networks had importance – they constituted a social asset – in that young women enjoyed support during childbirth and motherhood, grandparents enjoyed support in old age, relatives functioned as a bridge between the individual and the community, and people tended, therefore, to recognize faces in the crowd, to be recognized themselves, and to enjoy the interdependent senses of familiarity, belonging, and contentment. Finally, and significantly, Young and Willmott found that such networks were under threat from the policy of London Borough Council to move people away from places like Bethnal Green and onto housing estates at, or beyond, the city's boundaries. In short, with the concerns of Tonnies, Simmel, Wirth, and others in mind, Young and Willmott argued that traditional social networks had survived urbanization, at least in East London, but were now threatened by the next wave of geographical mobility: suburbanization.

The second study of post war slum clearance and urban renewal for consideration is an ethnography of native born Americans with Italian parentage living among other groups in West End, an inner city Boston neighborhood, undertaken by Herbert Gans between 1953, when the neighborhood was officially labeled a slum, and 1958, when demolition began under the Fed eral Renewal Program. Among other things, Gans found that West End was neither a slum nor an urban jungle – a low rent neighborhood that attracts criminals and the mentally ill, and is populated, therefore, by single men, pathological families, people in hiding, and individuals and groups providing services for such people. Rather, it was an urban village – a low rent neighborhood that attracts immigrants, and in which urban migrants attempt to adapt their nonurban institutions and cultures to the urban milieu.

In West End, such attempts led to what Gans identi fied as the peer group society – a society in which the strongest relationships were between people of the same gender, age, and life cycle status (e.g., siblings and cou sins), in which the rules of selection were based less on kinship ties and more on socioeconomic and cultural compatibility, and in which the mainstay was the family circle – a unit between the extended family of hunting or agricultural societies, and the nuclear family of urban societies. While this peer group society discouraged educational and occupational mobility, and discouraged close households, it encouraged sociability and mutual aid; it encouraged the exchange of news and advice, the planning of events, the exchange of memories, and the provision of entertainment and companionship. Most significantly, however, for Gans at least, the peer group society of West End inhibited collective action, since group members were preoccupied with constructing themselves as individuals through competition with each other. As a result of this, and the multiple concerns and actions of politicians and planners, the relatively harm less low rent district of West End was torn down between 1958 and 1962 and replaced by a luxury apartment complex, West Enders were dispersed across the metro politan area, and what Gans viewed as a functioning social system was destroyed.

This study by Gans, and that by Young and Willmott, and those by Tonnies, Simmel, Wirth, and others, can be seen as constituent parts of a tradition of urban sociology that views geographical mobility as threatening to social networks. The descriptor tradition is useful because it describes a conversation in which participants agree upon the topic of conversation, and this topic remains constant over time and space, while the positions of participants may differ in relation to place, time, and topic. Many commentators predicted the decline of traditional social networks with the rise of urbanization at the end of the nineteenth century. Around the middle of the twentieth century, some commentators found that traditional social networks in places like Bethnal Green had survived. Around the same time, other commentators found the adaptation of traditional social networks into new, but nevertheless functioning, social systems in places like West End. Many of these commentators predicted the decline of traditional or adapted social networks with the rise of suburbanization. One way to get beyond these different positions is to supplement ethnographic studies of particular neighborhoods with something like Robert Putnam's macrolevel study of trends in social capital across the US during the twentieth century. Using avail able survey data, Putnam examined trends in: political participation; civic participation; religious participation; connections in the workplace; informal social connections; altruism, volunteering, and philanthropy; and reciprocity, honesty, and trust. He found that levels rose for all of these variables between the beginning of the century and 1960, at which point they leveled off, before declining toward the end of the century. In other words, he found that levels of social capital rose during the first half of the century, a period of urbanization. And he found that levels of social capital stagnated, and then declined, during the second half of the century, a period of suburbanization. Having said that, it is worth noting that, for Putnam, while suburbanization is a minor culprit in this stagnation and decline, along with new time and money pressures, the major culprits are technology and mass media, and gen erational replacement (of the World War II generation by the baby boomer generation and Generation X).

It is also worth noting that Putnam found two or three exceptions to this general trend of stagnation and decline. He found that participation in small groups and mem bership of social movement organizations had both grown since 1960. And he glimpsed in the Internet, the potential to enhance communication, to include otherwise per ipheral participants, to flatten hierarchies, and to enhance democracy; though he also glimpsed an emerging digital divide, and emerging cyberbalkanization – the confine ment of communication to groups whose members share similar interests, leading to bonding within these groups, but not to bridging between such groups. These ex ceptions, identified by Putnam, point to some concluding comments about urban social networks.

The conversation about urban social networks among sociologists and others did not stop with studies of slum clearance and urban renewal in the mid twentieth cen tury. Since the 1960s, a group of social network analysts – first at Harvard under the leadership of Harrison White, then through the Toronto based International Network for Social Network Analysis, under the leadership of two former students of White, Barry Wellman and Steve Berkowitz – have turned their focus from the survival of traditional social networks in cities to the emergence of nontraditional social networks in contemporary society. Among other things, these analysts have focused on weak ties, directed by Mark Granovetter, who studied how people find work, and found that people acquire information about job opportunities through their in formal social contacts. So, weak ties are more than sources of alienation – the position of Louis Wirth and others in the early twentieth century. They help to dif fuse information, to increase opportunities, and to bridge social distance between cliques, leading to social co hesion, and to even the kind of collective action Herbert Gans found missing from inner city Boston. More broadly, these analysts have focused on multiple ties, their multiple geographies, and the multiple kinds of social support provided through these ties and geographies.

The work of Barry Wellman has been particularly influential in this regard. In Wellman's work, the oft conflated categories of primary ties, solidarity (shared sentiments), and locality are held apart. This enabled Wellman to observe social networks of the 1970s – a period of residential mobility, kinship dispersal, and urban diversity – while others were blind to them, having taken the traditional neighborhood as their primary ob ject of analysis. Wellman's observations led him to con clude that social networks neither died with nor survived the geographical mobility described by urbanization and later suburbanization. Rather, for Wellman, social net works were liberated by this geographical mobility – liberated from their local roots. So, primary ties have multiple bases, including kinship and friendship, and multiple geographies, including the neighborhood and the metropolitan area. They provide multiple kinds of social support, from emotional aid, through small and large services, to financial aid and beyond. And networks have multiple segments, one of which may contain immediate kin, whose relations are ascribed, stable, and broadly supportive, while another may contain work mates, for example, whose relations are achieved, adap tive, and specialized in support.

This multiplicity, identified by Wellman, leads to the second part of this article. As we have seen, for much of the twentieth century, urban networks was a term used by sociologists and others to describe social networks, their importance for bonding within communities and bridging between communities, and their relationship to the geo graphical mobility implied by urbanization and suburban ization. From sometime in the 1980s, however, usage of the term urban networks both multiplied and diversified across the social sciences in general, and particularly within geography. This was not least because urban geographers and others required a productive way out of the structure versus agency debates of the late 1970s and early 1980s, and networks had been helping sociologists and others to move between the micro and macrolevels since at least the 1960s. It was also because, while the relationship between geographical mobility and networks had been viewed as one of tension for much of the twentieth century, in the work of Wellman and others, during the 1970s, networks had become liberated from locality and looked more and more like appropriate units of analysis for an increasingly mobile and interconnected world.

Proliferating Networks

Across the social sciences, since the technological revo lution of the 1970s and the liberalization of international trade and investment, the network has become a popular metaphor – has become popularized by scholars like Manuel Castells and John Urry as a metaphor – to de scribe how people, ideas, and objects flow between nodes in a globalizing world. Within this broader context, the term urban networks has come to describe at least four things. First, it has come to describe archipelagos of world or global cities. Claims of the world or global cities literature include: that global production provides a new role for certain cities as command centers; and that such cities, with their management, finance, corporate service, and media jobs, and their catering, cleaning, and security jobs, are characterized by social polarization. An im portant claim, for the present article, is that centrality within the world city network – the network of cities with functions in the global economy – depends on the presence of other urban networks. These include net works of producer services that produce, in certain cities, a capability for global control. Saskia Sassen has written most extensively about these networks of producer ser vices, and has noted that, whereas such networks were concentrated in the center of cities during the 1980s, they are increasingly found stretched across suburban office complexes and edge cities – across anywhere rich in cyber routes, digital highways, and advanced telematics.

Indeed, if networks of producer services constitute one further network on which centrality within the world city network depends, then another is constituted by networks of infrastructure, including telecommunications – most importantly – and also transport, energy, water, streets, and so on. These networks – a second subject of the term urban networks in contemporary geography and the broader so cial sciences – appear most notably in the work of Steven Graham and Simon Marvin. They appear as something that integrated urban space between 1850 and 1960, during which time piecemeal provision became replaced by cen tralized and standardized systems, fragmented islands of infrastructure became integrated and consolidated, and services became more predictable and dependable across urban space. But, after 1960, these networks entered a period of crisis, in a context of fiscal problems that forced many nation states to explore transferring their infra structure operations to private operators (for the one off spoils of privatization), in a context of collapsed support for comprehensive urban planning among business leaders, feminists, environmentalists, and others, and in a context of continued urban sprawl. In the contemporary period, networked infrastructure has become unbundled or splintered into a myriad of individually financed and managed projects, since bankers have been reluctant to fund large scale and high risk developments. Infrastructure has enjoyed investment in prosperous or otherwise central areas, and suffered disinvestment in poor or otherwise marginal areas. Favored users and places have become increasingly connected by premium networks, while less favored users and places have become increasingly bypassed. As a result, cities have become increasingly fragmented and splintered.

Such a focus on networks of infrastructure – or networks of social support, or networks of service production for that matter – encourages an imagination of cities as crosscut by various networks that interweave and collide, that combine humans and nonhumans, and that help construct a sense of contemporary cities as, on the one hand, open and porous places of connections, juxtapositions, and encounters, and, on the other hand, ordered and regulated places of institutions, conventions, and technologies. Among others, Ash Amin and Nigel Thrift imagine cities in this way. Like Barry Wellman, they resist mourning the loss of traditional ties within coherent, stable, and proximate communities (while wondering if such communities ever existed anyway), just as they resist yearning for utopias, in the tradition of much writing about cities and their problems. Instead, they seek to acknowledge, surveil, and selectively support the small networks that do exist in the present and contribute to the maintenance of contemporary urban life – the urban actor networks, that describe friendship networks but also the technological unconscious of cities, by which is meant the clocks and watches that interested Georg Simmel at the beginning of the twentieth century, as much as the postcodes and software that interest urbanists at the beginning of the twenty first century.

So, at the beginning of the twenty first century, the term urban networks describes networks of cities with functions in the global economy, networks of infrastructure that have become unbundled in recent years, and networks of humans and nonhumans that help maintain urban life. The fourth subject of the term urban networks in contemporary geography and related disciplines takes the present article full circle and back to urban social networks. Analysts have continued to study these networks as they have changed over the last three decades – decades characterized by high levels of migration and new communications technologies. They have found that social networks are increasingly transnational, in that people, ideas, and objects move back and forth as a matter of routine, within migrant networks (both corporate and diasporic), and across the borders of cities and nation states. They have also found that social networks are increasingly institutionalized and self conscious. This is necessary given the instability and incoherence of contemporary cities. So, the word network has become a verb in recent times, the Internet contains social networking sites, and networking events take place, where people meet for short and intense exchanges of information, ordered in part by technologies of networking, such as business cards and databases. These urban networks are emerging at the present time. They are expected to attract much attention from geographers and others in the future.

See also: Actor-Network Theory/Network Geographies; Human Geography; Mobility; Network Analysis; Networks; Social Capital, Place and Health; Social Geography; Society-Space; Suburbanization; Transnational Ethnic Networks; Transnationalism; Transnationality; Urbanism; Urbanization; World/Global Cities.

Further Reading

Amin, A. and Thrift, N. (2002). *Cities: Reimagining the Urban*. Cambridge: Polity Press.

Castells, M. (1996). *The Rise of the Network Society*. Oxford: Blackwell.

Durkheim, E. (2002). *Suicide: A Study in Sociology* (first publ. 1897). London: Routledge.

Gans, H. (1962). *The Urban Villagers: Group and Class in the Life of Italian Americans*. New York: Free Press.

Graham, S. and Marvin, S. (2001). *Splintering Urbanism: Networked Infrastructures, Technological Mobilities and the Urban Condition*. London: Routledge.

Granovetter, M. (1974). The strength of weak ties. *American Journal of Sociology* 78, 1360 1380.

Putnam, R. (2000). *Bowling Alone: The Collapse and Revival of American community*. New York: Simon and Schuster.

Sassen, S. (1991). *The Global City: New York, London, Tokyo*. Princeton, NJ: Princeton University Press.

Scott, J. (1991). *Social Network Analysis: A Handbook*. London: Sage.

Simmel, G. (1950). The metropolis and mental life. In Wolff, K. (ed.) *The Sociology of Georg Simmel* (first publ. 1903), pp 409 424. New York: Free Press.

Smith, M. (2001). *Transnational Urbanism*. Oxford: Blackwell.

Taylor, P. (2004). *World City Network: A Global Urban Analysis*. London: Routledge.

Tonnies, F. (1955). *Community and Association* (first publ. 1887). London: Routledge and Kegan Paul.

Urry, J. (2000). *Sociology beyond Societies: Mobilities for the Twenty First Century*. London: Routledge.

Wellman, B. (1979). The community question: The intimate networks of East New Yorkers. *American Journal of Sociology* 84, 1201 1231.

Wirth, L. (1938). Urbanism as a way of life. *American Journal of Sociology* 46, 1 24.

Young, M. and Willmott, P. (1957). *Family and Kinship in East London*. New York: Free Press.

Relevant Websites

http://www.lboro.ac.uk/gawc
The official site of the Globalization and World Cities Study Group and Network hosted by the University of Loughborough.

http://www.insna.org
The official site of the Toronto based International Network for Social Network Analysis.

Neural Networks

I. Casas, University at Buffalo – SUNY, Buffalo, NY, USA

Glossary

Axons Single long fiber that makes part of a neuron and carries signals from the cell body out to other neurons. They constitute the transmission lines.

Cell Body It is the component of the neuron that processes the incoming signals from the dendrites.

Dendrites Tree-like receptive networks that make part of a neuron and carry electrical signals into the cell body.

Excitation of Neurons A state where the neuron has an increased probability of transmitting information to other neurons.

Inhibition of Neurons A state where the neuron has a decreased probability of transmitting information to other neurons.

Learning Rule A procedure to perform the learning process carried by neurons where weights and biases of a network are modified.

Neurons They are the structural constituents of the brain. They have three main components: dendrites, cell body, and axons.

Synapse Units that mediate the interaction between neurons.

Definition

Neural networks model the structure and functioning of the brain to perform a particular task of interest. They do so by simulating the human brain, and its basic components and processes, which include cell body (body of the neuron), dendrites, synaptic connections, axons, excitation and inhibition of neurons, neuron activity, and massive parallelism (see **Figure 1**). To represent these components, the neural network is composed of layers where each layer includes a set of nodes, connections, weights, a bias, a summer, and a transfer function. The nodes are the basic building blocks and constitute the input to the neuron. They are joined by connections to the summer which have a strength represented by a weight. These connections are analogous to synapses and store the knowledge necessary to solve specific problems. A bias constitutes another input to the summer and is similar to a weight but can be omitted if desired. The weights, the bias, and the node input go into the summer which produces as an output what is called the net input, which is then fed to a transfer function that produces the output of the neuron. The summer and the transfer function represent the cell body, while the output of the neuron represents the signal on the axon.

Neural networks are generally used to solve complex large problems of exploratory and explanatory nature. In geography, exploratory problems include mining of large spatial and nonspatial datasets, and pattern recognition and classification particularly in remote sensing. Explanatory problems include spatial interaction, choice modeling, network optimization, and statistical modeling.

Neural Network Model

Figure 2 shows the basic structure and operation of a neuron. The neuron has as input a set of attributes represented as nodes (p) that influence the output. These constitute the input layer also called an input vector. The inputs are multiplied by connection weights (w), which can be modified by a bias (b) (correction scalar, usually 1), to produce a net input. The net input is then passed through a transfer function (f) that results in an output that can be the input to the neurons in subsequent

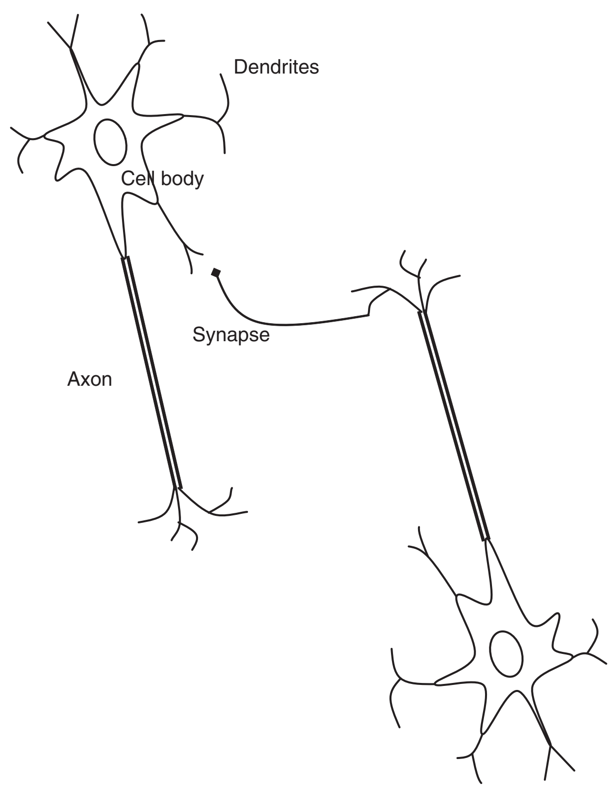

Figure 1 Human neuron.

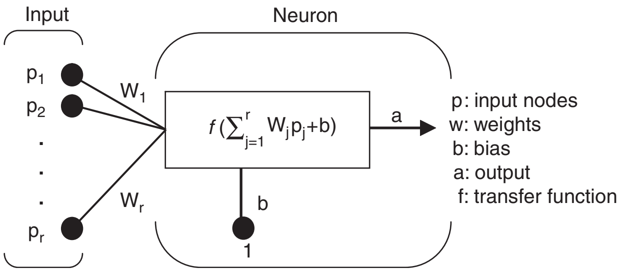

Figure 2 Simple network structure with multiple inputs.

intermediate/hidden layers or the final output. Transfer functions can be of different types (i.e., linear, sigmoidal, etc.) and define the threshold, which determines the state of a neuron (either on or off). When multiple neurons are combined to work in parallel, they form a structure called a layer. A neural network model can be composed of multiple layers of neurons, where the output of one layer constitutes the input to the next layer.

The number of inputs, intermediate layers, outputs, and the transfer function to be used defines the archi tecture of a network. The architecture determines the topology of the network. Three different types of net work architecture can be identified:

1. single layer networks: consist of an input vector con nected to a layer of *S* neurons;
2. multilayer networks: are characterized by one or more layers where the intermediate layers are called hidden; and
3. recurrent networks: are networks with feedback, where some of its outputs are connected to its inputs.

Learning

Given that the goal of neural networks is to model the way in which the brain performs a particular task, the process of acquiring knowledge is of primary importance. Knowledge is the result of a learning process via a learning rule. A learning rule is a procedure for modi fying the weights and biases of a network until a certain task is achieved. Learning rules fall into one of three categories: supervised learning, unsupervised learning, and reinforcement learning.

Supervised Learning

Supervised neural networks are characterized by being able to map a set of inputs to a set of outputs. The neural network uses the observed/input data to learn how to relate it to a given output. The input data are called the training data and their corresponding outputs are called the target. By means of a learning rule the weights and biases of the network are updated to make the network outputs closer to the targets. This step is performed iteratively until a statistical measure determines that the data have been successfully emulated. Once the training data have been learned, the network is ready to operate by itself in a different dataset called the test data. Supervised neural networks' basic process follows a series of steps:

1. obtain dataset: the data has to be divided into two separate sets, one for training and one for testing;
2. transform data to values in the 0–1 range;
3. identify network architecture: input variables, number of hidden layers, number of neurons per layer, and type of network to use;
4. train the network: choose learning algorithm and its corresponding parameters; and
5. test the model on the test data.

Supervised neural networks have advantages that make them very attractive when compared to more con ventional statistical tools. These include capability to handle nonlinear functions, no need for a model or functional form of the data distribution, capability to handle noisy data, ability to generalize data, learn from unknown relationships, and computational efficiency. However, they also suffer from some disadvantages. These include a limited modeling capacity, training is sort of an *ad hoc* process (i.e., trial and error), there is no standard methodology to determine the best architecture for a given problem/dataset, and testing and validation of their performance are not statistically robust.

Examples of supervised learning networks include the perceptron network, the backpropagation algorithm, and Widrow–Huff learning.

Unsupervised Learning

Unsupervised neural networks are characterized by having the ability to self organize to obtain a result. They teach themselves and identify patterns in the data. No target values associated to the inputs are required for training. This can be considered impractical given that it is counterintuitive to train a network when it is not known what it is supposed to do.

Unsupervised neural networks can be viewed as data classifiers. Their purpose is to partition the input data into clusters or groups with similar cases. Different types of data classifiers exist (i.e., cluster analysis, factor an alysis, and principal components); however, they are not sound when used with large datasets. Unsupervised neural networks provide an alternative due to their flexibility, degree of automation, quality of results, and type of technology.

When comparing unsupervised neural networks with traditional classifiers, advantages and disadvantages can be identified. Among the advantages, unsupervised neural networks do not require a class taxonomy, they do not get stuck in local suboptima, they do not make assumptions about the data's distribution (i.e., normal), and the analyst does not have to worry about spatial autocorrelation issues. Among the disadvantages, unsupervised neural networks require large datasets, the process by which results are obtained are not as easy to interpret as in traditional methods, the learning rate used in the network can affect its ability to properly cluster the inputs and in turn the learning ability, and the networks might exhibit instability when clusters are closer together.

Examples of unsupervised neural networks include associative networks, Hebbian networks, competitive networks, and Kohonen self organizing maps which have been often used in geographic applications.

Application of Neural Networks in Human Geography

Neural networks have been applied in the field of geography since the early 1990s. However, in human geography there are particular areas that have adopted the technology more rapidly with a level of success. Applications can be identified based on the use given to the neural networks models as either exploratory or explanatory.

Exploratory

Applications of an exploratory nature use neural networks as a data mining tool to identify either patterns in space and/or time in the data, classify large datasets, or propose taxonomy classifications. Examples include classifying census data, classification of arctic cloud and sea ice features, and recognition of cartographic area features. The area of remote sensing has also made ample use of neural networks for pattern and image recognition and classification.

Explanatory

Neural network models have been used as a replacement of statistical and optimization models in geography. They have been commonly used to predict and forecast flows in the area of telecommunications, trade, and transport. In transportation besides traffic flow neural networks have also been used in the areas of travel behavior and traffic management. For geography, the section on travel behavior is most relevant. Topics include understanding travel behavior, analysis of daily travel, commuting behavior, travel demand, activity scheduling, predicting travel times, and understanding route and mode choice. The potential applications of neural networks in freight transport have also been discussed.

Other examples of explanatory nature include predicting consumer spatial behavior, medical geography (i.e., forecasting the acquired immunodeficiency syndrome (AIDS) pandemic), and forecasting regional labor markets.

Example

To get an idea of how neural networks work, an example in the area of mode choice is presented. When making a trip, individuals are required to make a choice between the modes they have available for travel. Assume there are two modes: car and train. In order to model a network that can forecast these choices, an architecture composed of a set of inputs, a hidden layer, and a set of outputs is selected. This architecture is the result of a trial and error process. There is full connectivity between the inputs, the outputs of the intermediate layer, and the final outputs. The neurons make use of a sigmoidal or logistic transfer function. This function has a set of characteristics (i.e., continuity, increasing monotonicity, differentiable at all points) that makes it appealing for this kind of problem. The backpropagation learning algorithm is applied. The output of the neural network is the probability of an individual choosing the car as their travel mode. The inputs correspond to the characteristics of each mode: distance, time, and cost. The number of neurons in the hidden layer and the parameters of the learning algorithm are determined by running different experiments using a training dataset (this is the result of splitting the complete dataset into a training set and a test set). Based on the predictive capabilities of the models resulting, a choice is made as to what are the best parameters and architecture to use. At this point the network is ready for forecasting, using the test dataset. The neural network model will predict the probability of an individual choosing a car as their travel mode.

Where to Go from Here

Neural networks are very promising modeling tools. They possess certain elements that make them desirable such as faster speed in comparison to traditional statistical models, ability to learn, and ability to perform any task provided an appropriate architecture is selected. Neural networks also have benefits when compared to traditional models, some of which have already been outlined. In general, neural networks introduce a significant difference in performance; they offer solutions to hard type problems making them appealing to geographers.

Neurocomputing in geography has been introduced as a different paradigm for spatial analysis and modeling. Supervised neural networks are those more likely to be used and from the evidence found in the literature, this indeed seems to be the case. The success of this type of network and of the technology, in general, can be attributed to its benefits. These include their performance levels, lack of assumptions and knowledge when creating a model, lack of mathematical and statistical complexity, flexibility, ability to handle noisy data, and ability to handle spatial rich datasets. However, there are also potential problems that need to be considered when using neural networks. These include their reliance on data, difficulty in representing knowledge of the modeling task, variety in types (too many types to choose from), similarity to statistical models, unresolved issues like length of training, and difficulty in understanding the processes underlying them.

Regardless of their problems, neural networks are being increasingly applied to solve geographic problems. More recently self organizing maps have been adopted by geographers given their ability to address issues related to scale, space–time dependencies, nonlinear relationships, and data outliers. These are issues where statistical models typically fail to produce robust results.

There is enough evidence that shows neural networks should be and can be applied in geographic research successfully. In most applications they perform as well or better than their statistical counterpart. They can be applied to areas where statistical models have difficulties (e.g., spatial autocorrelation) and in theory could be applied to solve any geographic modeling problem. The neurocomputing paradigm is here to stay and its complexity and sophistication will continue to expand offering new alternatives to solve geographic problems.

See also: Simulation; Spatial Interaction Models.

Further Reading

Celik, H. M. (2004). Modeling freight distribution using artificial neural networks. *Journal of Transport Geography* 12, 141 148.

Dougherty, M. (1995). A review of neural networks applied to transport. *Transportation Research C* 3, 247 260.

Fischer, M. M. (1995). Fundamentals of neurocomputing. In Fischer M. M., Sikos, T. T. & Bassa, L. (eds.) *Recent Developments in Spatial Information Modeling and Processing*, pp 31 41. Budapest: Geomarket.

Fischer, M. M. (1998). Computational neural networks a paradigm for spatial analysis. *Environment and Planning A* 30, 1873 1892.

Fischer, M. M. and Gopal, S. (1993). Neurocomputing a new paradigm for geographic information processing. *Environment and Planning A* 25, 756 760.

Fischer, M. M. and Gopal, S. (1994). Artificial neural networks a new approach to modeling interregional telecommunication flows. *Journal of Regional Science* 34, 503 527.

Hagan, M. T., Demuth, H. B. and Beale, M. (1996). *Neural Network Design*. Boston, MA: PWS Publishing Company.

Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. Upper Saddle River, NJ: Prentice Hall.

Hewitson, B. C. and Crane, R. G. (eds.) (1994). *Neural Nets: Applications in Geography*. Dordrecht: Kluwer Academic Publishers.

Himanen, V., Nijkamp, P. and Reggiani, A. (eds.) (1998). *Neural Networks in Transport Applications*. Aldershot: Ashgate.

Nijkamp, P., Reggiani, A. and Tsang, W. F. (1997). Model and scenarios for European freight transport based on neural networks and logit analysis. *Research Memoranda* 35, 319 325. Amsterdam: Free University Amsterdam.

Openshaw, S. (1992). A review of the opportunities and problems in applying neurocomputing methods in marketing applications. *Journal of Targeting, Measurement and Analysis of Marketing* 1, 170 186.

Openshaw, S. (1993). Modelling spatial interaction using neural net. In Fischer, M. M. & Nijkamp, P. (eds.) *GIS: Spatial Modelling and Policy Evaluation*, pp 147 164. Berlin: Springer Verlag.

Openshaw, S. and Openshaw, C. (1997). *Artificial Intelligence in Geography*. Chichester: John Wiley.

Openshaw, S., Wymer, C. and Craft, A. (1991). Using neural nets to solve hard analysis problems. *Proceedings of EGIS'91* pp 797 807. Utrecht: EGIS Foundation.

Shmueli, D., Salomon, I. and Shefer, D. (1998). Neural network analysis of travel behavior. In Himanen, V., Nijkamp, P. & Reggiani, A. (eds.) *Neural Networks in Transport Applications*, pp 151 170. Aldershot: Ashgate.

Skupin, A. and Hagelman, R. (2005). Visualizing demographic trajectories with self organizing maps. *Geoinformatica* 9, 159 179.

Relevant Websites

http://www.doc.ic.ac.uk/~nd/surprise 96/journal/vol4/cs11/report.html

http://www.mathworks.com/products/neuralnet/
Accelerating the Pace of Engineering and Science.

http://lcn.epfl.ch/tutorial/english/index.html
Laboratory of Conceptual Neuroscience.

http://www.girardin.org/luc/cgv/www5/index.html
Mapping the Virtual Geography of the World Wide Web.

http://www.wolfram.com/products/applications/neuralnetworks/
Train and Analyze Neural Networks to Fit Data.

New Regionalism

G. MacLeod, University of Durham, Durham, UK

Glossary

Flexible Specialization A system of workplace organization and production featuring decentralized management (such as quality circles), flexible and versatile technologies and workforces geared for more specialized customers' demands in volatile markets, and signaling a contrast from Fordism's standardized mode of production for mass markets.

Industrial District Regional economies that are developed out of high-tech clusters or craft-based sectors often orchestrated through flexible specialization and close interfirm networking; in this sense, they draw some comparison with the nineteenth-century districts featuring an 'industrial atmosphere' identified by Alfred Marshall.

Institutional Thickness A concept which sees economic growth to be dependent on not only the 'hard' economic or physical assets but also the softer informational infrastructure, featuring a plethora of public, private, and voluntary organizations engaged in economic development, the collective representation of potentially roguish interests, and a mutual awareness of place so that regional actors come to perceive of a 'common agenda'.

Untraded Interdependencies A range of institutional forms such as locally derived rules, conventions, customs, values, labor market governance, as well as the collective goods provided by public and quasi-public organizations and agencies; they can be contrasted with the 'hard' traded input–output interfirm relations that form the cornerstone of transactions–cost economics.

Introduction

The new regionalism is a term used to describe the re search of certain scholars in economic geography and urban and regional planning highlighting how during the era of post Fordism and globalization, certain regional economies appear to be more capable than others of sustaining relatively successful economic growth. One crucial insight of this work has been to firmly rebuff the naïve view that globalization somehow signals the end of geography and to identify how the process of global ization is in actual fact choreographed through a con centration of economic power in certain regions and city regions. The influence of the new regionalism debate has also rippled out to the spheres of policy and politics with city mayors, regional politicians, influential commen tators, consultants, and even business gurus regularly canonizing the region as a functional scale for economic planning and political governance.

So why has the region come to assume such prom inence in recent academic debate and policy discourse? In responding to this question, the work of Michael Storper can prove to be most instructive. Storper, a stu dent from the so called Los Angeles School of geography and urban planning, surmises that during the middle decades of the twentieth century the region came to represent a residual category in the advanced developed countries. In the academic arena, regional science had attained ascendency: in spite of its title and in proffering a quantitative modeling approach to locational analysis and an unshakable quest to identify general laws to ex plain spatial behavior, regional science diverted the analytical lens away from regions as spaces of cultural and economic difference. The laws of regional science were given a concrete expression through Keynesian economic policies, with regions assigned a role as merely administrative spaces to integrate national economies and to help suppress regionalist ideologies, which were as sociated with the more obscure and violent fringes of the political arena.

This era can be contrasted with the period of the late twentieth and early twenty first centuries. For today the region is widely acknowledged to represent a funda mental basis of economic, social, and political life. We are said to be living in a regional world. Moreover, the concept has assumed a more profound ontological sig nificance as a meso level of analysis within which to examine the current regime of post Fordist reflexive capitalism. This article has three main aims: (1) to pro vide a discussion of the main arguments that characterize the new regionalism; (2) to offer a critical assessment of these arguments; and (3) to tender some thoughts about the potential future directions of new regionalist research in economic development.

Geographies of Post-Fordism: Industrial Districts, New Regional Economies, and Resurgent Rustbelts

Before going on to discuss the new regionalism as it re lates specifically to regional development, at the outset it is perhaps worth noting several alternative versions which more directly engage with questions of politics. First, the subdiscipline of international political economy is deploying a new regionalism to question the realist understanding of states as isolated entities: here the region is defined as a supranational space in a similar way to how international relations discourse habitually deploys the concept to describe a multinational landscape like the Middle East. A second version exists in political science and analyzes the territorial struggles that sometimes give rise to subnational spaces as significant politico administrative units. Another variant of an explicitly defined new regionalism has surfaced in American urban political science and relates to politically led endeavors to amalgamate and consolidate into a metropolitan form the diverse fragments of urban, suburban, and neighborhood government which have unfolded across the American landscape throughout the twentieth century.

Turning more specifically to debates in regional development, the new regionalism (hereafter NR) can, at least in part, be traced to some pioneering research in the late 1970s and early 1980s documenting the rise of a new mode of production named flexible specialization, which appeared to be emerging in certain parts of Italy and West Germany. At a time when the Western world's received wisdom centered around national models of development increasingly forced to tame footloose multilocational firms and an internationalizing spatial division of labor, the discovery of individual districts in north central Italy and Southern Germany featuring localized constellations of industrial firms interlinking and subcontracting with each other, was quite a revelation.

This flexible specialization thesis generated considerable excitement across the global academic division of labor as a miscellaneous coterie of economic sociologists, economic historians, urban and regional planners, and economic geographers variously heralded: (1) a reemergence of industrial districts such as Emilia Romagna in Italy premised around traditional sectors such as ceramics, furniture, and footwear and (2) regional economies such as Baden Württemberg in Germany propelled more by modern precision engineering, a post Fordist flexible production of luxury automobiles, and guided by a largely corporatist mode of regulation. Crucially, while each of these regions evolved through different historical traditions, their respective sectors were competing on quality rather than price, with the result that employees were often highly skilled, well remunerated, and unionized: again bucking a putative global trend of deskilling and union bashing.

Subsequent research throughout the 1980s and 1990s was to reveal distinctive territorial production complexes and entirely new industrial spaces such as Rhone Alpes in France, Silicon Valley in northern California, and Route 128 in Massachusetts, all quite removed from the older centers of industrial production: indeed Silicon Valley did not exist until the 1960s. The work by several economic geographers in the USA was especially valuable in identifying how such regions were literally being formed through the locational decisions of high technology firms and their buyers and suppliers, not least the spatial impact of their transaction costs. Such analysis was to offer powerful backing to the notion that regions could benefit from fostering clusters of networked firms specializing in key propulsive sectors.

Over in Europe, alongside a nascent political appetite for a 'Europe of the Regions', in depth qualitative research conducted by urban and regional scholars was beginning to disclose the specific character of economic and political governance in a range of high performing regions such as the Italian and German industrial districts, Catalonia in Spain, and Jutland in Denmark. Indeed, a consensus began to emerge that much of the long term economic success of regions such as Emilia Romagna and Baden Württemberg was down to a delicate balance between competition – firms and sectors obviously had to be nationally and globally competitive – and cooperation, featuring a rich institutional fabric of trade and business associations, influential chambers of commerce, innovation and technology transfer centers, development agencies, trade unions, and civic and voluntary organizations. All this was deemed to enable a high degree of trust based behavior among economic actors, not least through the provision of collective services for companies in spheres such as market forecasting, trade, training, and technology transfer and also a longer term horizon for strategic planning and investment. This was also being contrasted with the rugged individualism viewed to be dominant at a national scale in Anglo American societies.

A growing appreciation of these success stories precipitated a sense that regions could become proactive in fostering their own economic success. In part, this was fueled by research which was beginning to herald a new type of inward investment: one qualitatively different from the late Fordist branch plant whose external control and tendency to relocate to cheaper destinations left regions acutely vulnerable to minor market fluctuations. A particularly notable contribution in this debate was work on places such as the Great Lakes in the Midwest of the United States and Wales in the United Kingdom. These were former industrial heartlands whose economies had endured protracted deindustrialization throughout the 1970s and 1980s, but where a number of overseas Japanese investments were seemingly forming integrated transplant complexes and becoming a powerful vehicle for diffusing technology transfer and knowledge throughout the regions. The embedding of such post Fordist knowledge intensive corporate activity was thereby demonstrating how, alongside the provision of interrelated infrastructures in production, industrial governance, human capital, physical and digital communications, even older industrial regions such as the Great Lakes were showing the capability to adapt amid ever intensifying global competition. Hence, the references lists for university modules on economic geography and regional develop ment were seeing titles such as 'The rise of the rustbelt' being added to classics such as the 'Geography of deindustrialization'.

As we can see, then, a variety of perspectives began to punctuate the academic landscape of regional develop ment: some highlighting the location strategies of key firms, others emphasizing a supportive infrastructure for economic development. While each was drawing inspir ation from a diversity of theoretical foundations – insti tutional economics, economic sociology, evolutionary political economy, and economic geography – they also appeared to be united on at least one front: the enduring dynamism of certain feted high profile regional econ omies could not be explained by Ricardian notions of comparative advantage and certainly not in terms of raw natural endowments. Neither could it be fully accounted for by pure market forces or by the *dirigiste* policies and spatial Keynesianism that characterized much capitalism in the West during the post war era. Rather, the quali tatively informed approaches outlined above take ser iously Karl Polanyi's axiom of 'the economy as an instituted process' to contend that a major source of a region's competitive advantage is embedded in the op eration of its local civil society vis à vis a range of soft infrastructures relating to locally derived rules, norms, customs, habits, cultures, conventions, institutions, trust based interactions, and horizontal relations of reciprocity, which are seen to enhance the benefits of investments in physical and human capital.

By the late 1990s, a variety of concepts were being introduced to help encapsulate this institutional atmos phere of an increasingly reflexive capitalism orchestrated through local and regional strategies: regional innovation system, associational economy, innovative milieu, learn ing regions, institutional thickness, and untraded inter dependencies. All these became routinely deployed as part of the lexicon of the NR. The NR itself was to become immensely influential in debates on regional development as well as those relating to policy.

New Regionalism: A Critical Assessment

New regionalist thinking undoubtedly serves up some valuable conceptual insights and offers us an interesting map of global economic development: indeed, some scholars talk of a global mosaic of city regions and re gional economies. Much of the debate also provides persuasive evidence that the subnational region repre sents an increasingly crucial scale for territorially con vening the various organizations and actors which orchestrate the process of economic growth: ranging from hard infrastructure to soft info structure. Nonetheless, in recent years, each of these dimensions has also been subjected to considerable, at times biting, critique.

New Regionalism and Public Policy

Silicon Valley is being imitated all over the world. There is a valley clone in Bangalore, Russia, and China. Ireland has its own Silicon Bog and Scotland, a Silicon Glen. Of course, it is hardly surprising that elites and growth coalitions from cities and regions that have endured large scale and traumatic economic meltdown have looked outward and across to the experiences of cele brated prototypical success stories such as Silicon Valley and Emilia Romagna, and to seek to emulate their suc cess. Hence, the mania for cluster programs associated with the influential consultant Michael Porter, and also for accredited initiatives to foster 'hobbyist' forms of interfirm networking.

Relatedly, however, many of the perspectives associ ated with the new regionalism could invite economic development practitioners to assume that the ultimate endgame is to help occasion soft infrastructures with which to enable networking, untraded interdependencies, social conventions, and institutional thickness. If such policy prescriptions are to be ingested by a regional corpus whose vital organs are dominated by poorly performing industrial and postindustrial sectors all en cumbered by an outmoded institutional inheritance, in practical terms, the net result might offer little beyond euphemistic policy discourses and a public sector funding of business breakfasts for local bourgeois elites. The NR has courted much controversy in this regard. In some NR writings, there has also been a largely implicit assump tion of a correlation between the presence of a powerful regional government and regional economic prosperity, but any suggestion of a link between devolved regional government and regional economic success is evidence of association rather than causation.

Nonetheless, one ought to be careful when attributing responsibility in all of this. Indeed, far from necessarily acting as a Trojan horse for quick fix policy solutions, a close reading of certain new regionalist accounts will, quite ironically perhaps, alert us to the very dangers and shortcomings of such crude policy transfer. For concepts such as institutional thickness and untraded inter dependencies highlight locally specific assets that are noncodifiable and therefore extremely difficult to transfer across geoeconomic space, geopolitical boundaries, and oceanic divides. We are also informed how, in practical terms, soft infrastructures such as conventions, trust based interactions, and civic cultures are inherently esoteric and intricate, and much harder to magic than the more ostensibly visible and conventional public infra structures such as roadways, schools, training colleges, and, of course, the traditional photograph featuring a local politician shaking the hand of some senior executive of an overseas inward investor.

Conceptual Blind Spots: The State and the Social Geographies of Post-Fordism

Much discussion has also focused on certain blind spots in the NR's conceptual arsenal. It is increasingly being claimed that the most obvious analytical lacuna concerns the failure to connect contemporary analyses of regional economic development with the changing form and character of the state. Perhaps this is in part because the state seems much less involved in regional development than in the heyday of Fordism, what with its modernist commitment to state planning and its interventionist approach to shifting people and firms in order to balance growth. Indeed, perhaps this is the key point: regions are able to generate endogenous growth without regional aid from the center.

However, to think along these lines would blind us to the fact that the American gold standard regions of Silicon Valley and Route 128 have been in large part the product of vast amounts of spending on defense by the US federal government. They have also depended on the pumping of public investment into world class universities such as Stanford and MIT. Furthermore, new regional accounts rarely draw our attention to the ways in which distinctive locally derived institutions and vital infrastructures are mediated by national modes of regu lation vis à vis market rules and corporate governance or through policies relating to the labor market, education, and training. In thinking through these themes, perhaps we should also consider how and why many of the most successful localities and regions happen to be located within the most robust national economies such as the US and Germany. Moreover, the one dimensional focus on cooperative soft institutions can divert attention from the way that regions are increasingly inscribed into the ruthless global geographies of competitive capitalism. It is exactly the relational webs between various scales of government and governance which are lacking systematic investigation within much NR scholarship.

This neglect of the state, and in particular of the social relations between economic reflexivity and the changing character of the welfare state, also forecloses debate on what is considered to be the sociopolitical downside of post Fordist reflexive capitalism: a splintered urban–regional landscape, intensified social inequality, and an increasingly revanchist and authoritarian political regime that now almost seems to be mandatory for governing such a splintering society. Not least is this down to the NR's preoccupation with supply side innovation which leaves analysis bereft of demand led considerations and related questions of individual and collective consumption. Furthermore, critical research on sanctified regions like those in the Third Italy has begun to reveal their own particular sociological dark side, where cor porate dexterity is often intricately dependent on dual labor markets and hyperexploitation on the basis of class, ethnicity, gender, and other social divisions.

These blind spots are largely down to the fact that the NR analytically telescopes onto seemingly nonexploitative horizontal relations of networking and reciprocity within particular localities, simultaneously obscuring us from the vertical ordering and the hierarchical structuring of social relations across and between spatial scales of governance. In short, the asymmetries of economic and political power, prestige, and social privilege are conveniently sidelined, as are unpalatable themes which were once the stock in trade of critical urban and regional studies: uneven de velopment, exploitation, interterritorial competition, de valuation, and welfare retrenchment. All of which also begs searching questions about just how far economic and political democracy might extend to in the local and re gional worlds so regularly canonized in economic geo graphical discourse.

New Regionalism: Future Directions

Over the last two decades, the NR has had a significant influence in debates on regional development. With cities and regions looking, if anything, to become ever more significant as spaces of governance and arenas subject to global competition and associated pressures of place marketing and promotion, the debate looks set to con tinue. This brief final section sketches out several dir ections in which this debate might take.

The first concerns a series of conceptual challenges vis à vis the region. While undoubtedly central to aca demic and policy relevant analysis for over 100 years now, the region continues to be an elusive category. In general, and perhaps wisely in view of this, much NR research deploys an implicit reading of its spaces of in vestigation. However, it is argued that it is important to consider how much of the literature on industrial dis tricts refers to economically integrated areas such as Silicon Valley and Route 128 which do not readily cor respond to a regional sensibility either in a cultural or a political sense. On the other hand, Baden Württemberg's economic dynamism would appear to be quite closely intertwined with the formation of robust regional (Länder) political institutions, introduced as part of West Germany's federal reconstruction after World War II; thus, the boundaries which characterize the various re gions in the NR are at times incongruous. While con sidering this, perhaps too we should begin to avoid reifying cities and regions. Scholarly and policy dis courses regularly describe places as proactive, picking up their bootstraps: a plea is made for a fuller recognition of the fact that it is people and organizations working in and on behalf of these places that do the acting, not the spaces themselves.

Perhaps a more profound conceptual challenge emerges from those advocating a relational reading of the region. This approach contends that the major forces associated with globalization have been transforming the material and experiential character of cities and regions in recent decades – not least the routine trafficking of objects and people and the instantaneous circulation of information – and that these are surely disturbing any semblance of a world order made up of nested territorial formations. Viewed through this alternative lens, emer ging spatial configurations are no longer interpreted as territorial. Rather they are constituted through a kal eidoscopic web of networks and relational connections, which are not fixed or located in place but are constituted through various circulating entities, which in turn bring about relationality both within and between societies at multiple and varied distances. Such a perspective cer tainly opens up new ways of seeing the transnational circulation of capital vis à vis multinational corporations, money, and migration and could also enable us as re searchers to trace the global circulation of ideas, tech nologies, and knowledges which serve to inform the NR.

The relational approach also opens up some fresh ways to interpret the second theme that needs to be mentioned: that of environmental sustainability. Quite surprisingly, this is a topic that has largely been implicit or even omitted from many significant studies in the NR. However, the theme of environmental sustainability can hardly be avoided in the early twenty first century. Not least in that one of the NR's model spaces has been en during quite significant environmental challenges in re cent years. For, in Silicon Valley, housing is now so scarce that new arrivals and immigrants to the region often have to drive for 2 h to commute to work. As a result, traffic congestion is endemic and the physical environment under threat. While the higher education sector thrives, property taxes are so low that there is not enough money to fund a decent basic education system. Arguably, the most sensible way to solve these problems is through public initiatives, yet Silicon Valley currently lacks the political machinery to address these regional concerns. Similarly, the long term sustainability of German regions such as Baden Wurttemberg has also been thrown into some serious doubt: intensifying globalization is leaving its high wage and socially generous mode of regulations facing competition from low cost but high quality Far Eastern economies.

This leads to the final theme. As the discourse of post Fordism has tired and at least to some extent become eclipsed by one of globalization, so the analytical gaze in Anglo American writing has extended beyond Western Europe and North America to consider a 'global mosaic of regional economies'. Nonetheless, the debate thus far has been overwhelmingly concentrated in the continents of North America and Western Europe. This needs to change. With some Latin American countries looking toward the European Union as a model for regionally articulated growth, it surely behoves us to begin a much closer dialog with colleagues in Latin America, Africa, and Asia.

See also: City-Region; Economic Geography; Governance; Institutionalism/Institutional Geographies; Knowledge Economy; Public Policy; Region.

Further Reading

Amin, A. and Thrift, N. (eds.) (1994). *Globalization, Institutions, and Regional Development in Europe*. Oxford: Oxford University Press.

Cooke, P. (ed.) (1995). *The Rise of the Rustbelt*. London: UCL Press.

Cooke, P. and Morgan, K. (1998). *The Associational Economy: Firms, Regions and Innovation*. Oxford: Oxford University Press.

Hadjimichalis, C. (2006). Non economic factors in economic geography and in 'new regionalism': A sympathetic critique. *International Journal of Urban and Regional Research* 30, 690 704.

Harrison, J. (2006). Re reading the new regionalism: A sympathetic critique. *Space and Polity* 10, 21 46.

Keating, M. (1998). *The New Regionalism in Western Europe: Territorial Restructuring and Political Change*. Cheltenham: Edward Elgar.

Lovering, J. (1999). Theory led by policy: The inadequacies of the 'new regionalism' (illustrated from the case of Wales). *International Journal of Urban and Regional Research* 23, 379 395.

MacKinnon, D., Cumbers, A. and Chapman, K. (2002). Learning, innovation and regional development: A critical appraisal of recent debates. *Progress in Human Geography* 26, 293 311.

MacLeod, G. (2001). New regionalism reconsidered: Globalization and the remaking of political economic space. *International Journal of Urban and Regional Research* 25, 804 829.

Painter, J. (2008). Cartographic anxiety and the search for regionality. *Environment and Planning A* 40, 342 361.

Pike, A. and Tomaney, J. (2007). *Special Issue: Whither Regional Studies. Regional Studies* 41(10).

Pike, A., Rodriguez Pose, A. and Tomaney, J. (2006). *Local and Regional Development*. London: Routledge.

Sabel, C. (1994). Flexible specialization and the re emergence of regional economies. In Amin, A. (ed.) *Post Fordism: A Reader*, pp 101 156. Oxford: Blackwell.

Saxenian, A. L. (1994). *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Boston, CA: Harvard University Press.

Scott, A. J. (1998). *Regions and the World Economy: The Coming Shape of Global Production, Competition, and Political Order*. Oxford: Oxford University Press.

Storper, M. (1997). *The Regional World: Territorial Development in a Global Economy*. New York: Guilford.

New Towns

K. Kafkoula, Aristotle University of Thessaloniki, Thessaloniki, Greece

Glossary

Conurbation Coalition of two or more urban entities, that is, cities or towns, which by population increase and physical expansion merge to form one continuous developed area.

Neighborhood Unit Urban area of housing and everyday facilities (school, retail, etc.) conceived as a whole. Post-war New Towns comprised usually of 2000–4000 dwellings. In Eastern European planning, it was named as microrayon.

Population Overspill Projected population increase, which cannot be accommodated properly within the existing limits of a given urban entity.

Satellite Town New settlement created in proximity to an urban center, on which it depends for the employment of its inhabitants. A satellite town is distinguished from a dormitory suburb by the provision of retail premises, schools, and public utility bodies.

Introduction

New Towns are the result of powerful application of a state policy expressed by governmental decisions. In human history, founding of New Towns have repeatedly been applied as an orderly procedure against the other wise spontaneous, or haphazard evolution, in the Roman, medieval, and later times, or in the nineteenth century. Of all, the ideal communities of the industrial era were the offspring of a widely spread desire for social justice, expressed in philanthropic or communitarian action. Rooted in those movements were the garden cities, the forerunners of the New Towns, put into the planning scene by the Ebenezer Howard, of Britain, before the Great War. New Towns, the subject of this article, started off also in Britain during and after World War II, and incited similar policies elsewhere, to the point that they can now be found in entirely diverse planning environ ments in most parts of the world.

Definition

A New Town is a settlement with an official designation act issued at a precise moment in time. On the basis of an agreed policy, the authority decides on the purpose, role, location, and the physical characteristics of the new settlement, the implementation measures and funds, and the means to attract population, employment, and various other activities, the prerequisites of urbanity.

New Towns have never been a mere technical affair. Crucial for their creation in the aftermath of World War II was the conviction that a more orderly environment would address the shortcomings and malfunctioning of cities; and that new settlements, properly established and managed, could effect a redistribution of resources on a regional or at a national level.

Definition of the term New Town requires laying out a frame of reference, since in the twentieth century nu merous ones were built. The definition must not con centrate on the number of inhabitants, since there is no strict co relation between the size of the population and the urban quality expressed by the term 'town'. Also, one must ignore various recent developments of no negligible size, named in USA 'new communities', which sprang out of private initiative with no reference to a regional policy (Reston, Virginia, Irvine, California, etc.), or the so called 'edge cities', that is, outlying urban areas of office space, ample parking, and residential precincts popping up in the wish for real estate profits. The engagement of the state is critical, even when it is minimal, as in the case of some Asian countries. The capacity to grow steadily over a period of time after the original creation is an im portant factor, but is also insufficient if it is not coupled with the minimum requirements for a decent urban life.

Valid criteria should then be the preexistence of a general plan of regional impact, even if that plan undergoes substantial alterations on the way. Realization should match the provision of a house and a job, ideally not far from each other. This criterion (a heritage of the garden city ideas) originally distinguished the New Towns from satellite towns. However, with the devel opment of high efficiency transport means, the issue of proximity has somehow faded. The term New Town is increasingly being used for satellites also, or for towns that have been so for a considerable part of their life.

This article covers settlements that are named New Towns in their countries, discussing their performance in establishing a more rational relationship between people and their environment. Their assessment points out to the distance between the intended and the feasible.

The Origins: New Towns in Post-War Britain

New Towns in Britain originated as a response to the re corded ill balanced allocation of industry and population, in favor of London area and the South East in general. They were implemented by the enactment of the New Towns Act in 1946 and the post war planning machinery of the UK.

Eight New Towns were designated around London from 1946 to 1949, in order to accommodate projected overspill population: in chronological order Stevenage, Crawley, Hemel Hempstead, Harlow, Hatfield, Welwyn (originally a garden city), Basildon, and Bracknell. This was matched by New Towns in the provinces, Aycliffe (now Newton Aycliffe), Peterlee, and Corby in England; East Kilbride and Glenrothes in Scotland; and Cwmbran in Wales. Overspill population targets generated also Cumbernauld in Glasgow (1955) and the next mark of New Towns, from 1961 to 1964, Dawley, Redditch, Skelmersdale, and Runcorn in England, and Livingston in Scotland. Finally, Washington in North East England (1964) would induce regional development. For most of them, the intended population ranged from 30 000 to 80 000.

In view of the forecasts that the south of England would attract a new population of 3.5 million in the two decades 1961–81, three New Towns were planned as regional growth poles in that area: Milton Keynes (1967, 250 000), Peterborough (radical expansion, 1967, 185 000), and Northampton (expansion, 1968, 230 000). Parallel endeavors for decentralization elsewhere founded Warrington (1968, 202 000), Telford (1968, 250 000, the previous Dawley on a broader program), and Central Lancashire (1970, 500 000), all in England. In addition, Irvine was designated in Scotland (1966, again mainly for Glasgow overspill), Newtown in Wales for rural regeneration (1967), and four New Towns started in Ireland for the redistribution of industry and population (Craigavon 1965, Antrim 1966, Ballymena 1967, Londonderry 1969). On the whole 32 New Towns (**Figure 1**) were created in the United Kingdom till 1970.

State financed development corporations, one for each town, acquired the land, installed the infrastructure, constructed the houses and the retail premises, and managed the assets till the New Towns were incorporated in the system of local government. Today, they are home to over two and a half million people. The target of containing the population of older cities, that is decentralization within regions, was largely achieved. Amelioration of regional inequalities hardly progressed, since older cities, particularly London, a world metropolitan area, sustained their dynamism.

New Town Policies in Europe

In continental Europe, even distribution of population and employment were the two prevailing targets in the designation of New Towns. State land and public transport would facilitate the ultimate end of a balanced society in ecologically friendly environments.

In post war France, the need for emancipation from the dominance of Paris (20% of the population, 50% of the jobs) was translated in a twofold policy: to create metropolises in the provinces and to transform Paris into a decentralized metropolis. Therefore, in the 1970s a New Town was founded around each of the major centers Lille (Villeneuve d'Asq), Rouen (Le Vaudreuil), Lyon Grenoble (Isle d'Abeau), and Marseille (Étang de Berre) and five around Paris (Cergy Pontoise, St Quentin en Yvelines, Évry, Melun Sénart, and Marne la Vallée) – on the whole nine New Towns. The target population of the provincial cities was of 300 000, which would allow for a large spectrum of jobs and household incomes. However, it is the New Towns around Paris that have proceeded more successfully. State offices had the management, whereas design and construction were undertaken by private firms. In the eve of 2001, French New Towns had around 900 000 inhabitants.

The New Towns of the Netherlands were related to projects of land reclamation, in particular lands reclaimed from IJsselmeer (formerly Zuidersee), and thus state owned. By far the most important is Almere (1970s, projected 250 000 people), planned to restructure Amsterdam region. In Scandinavian countries, early post war, small size New Towns, Albertslund in Copenhagen area, Tapiola in Helsinki, and Vällingby, Farsta, and Skärholmen in Stockholm, became legendary for their modern design and, although not commercial in any way, they were a success in financial and environmental terms.

In what was once the communist world, New Towns were created centrally in support of industrial plants, and housing provided for rent or home ownership. From the 1950–51 New Towns of Hungary, the most important, Dunaujvaros (formerly Dunapantèle and later Sztalinvaros), served the largest iron complex, near Budapest. In Poland, the pride of the workers' New Towns, Nowa Huta was linked to a huge iron complex (today a part of Krakow, with over 200 000). In East Germany a New Town was connected to iron industry, Eisenhüttenstadt (initially Stalinstadt), near Frankfurt an der oder.

In ex USSR the program for New Towns conformed to two fundamental targets: to transfer population from the agrarian areas into industrial centers in an organized way and to create a balance in economic growth between the eastern (80% of resources) and the western regions (80% of the industries). According to records, albeit scarce, a total of 1100 New Towns were constructed till the late 1970s. Among the New Towns serving heavy industrial complexes, many were in the eastern provinces, the vast Siberian lands. Karaganda and Magnitogorsk near a huge steel plant in Irkutsk area were built before the war, and after the war Angarsk (petrochemicals) in the same region, Bratsk and Sumgait around the Caspian Sea (**Figure 2**), and Togliati to the north of Samara city. A small number of scientific centers

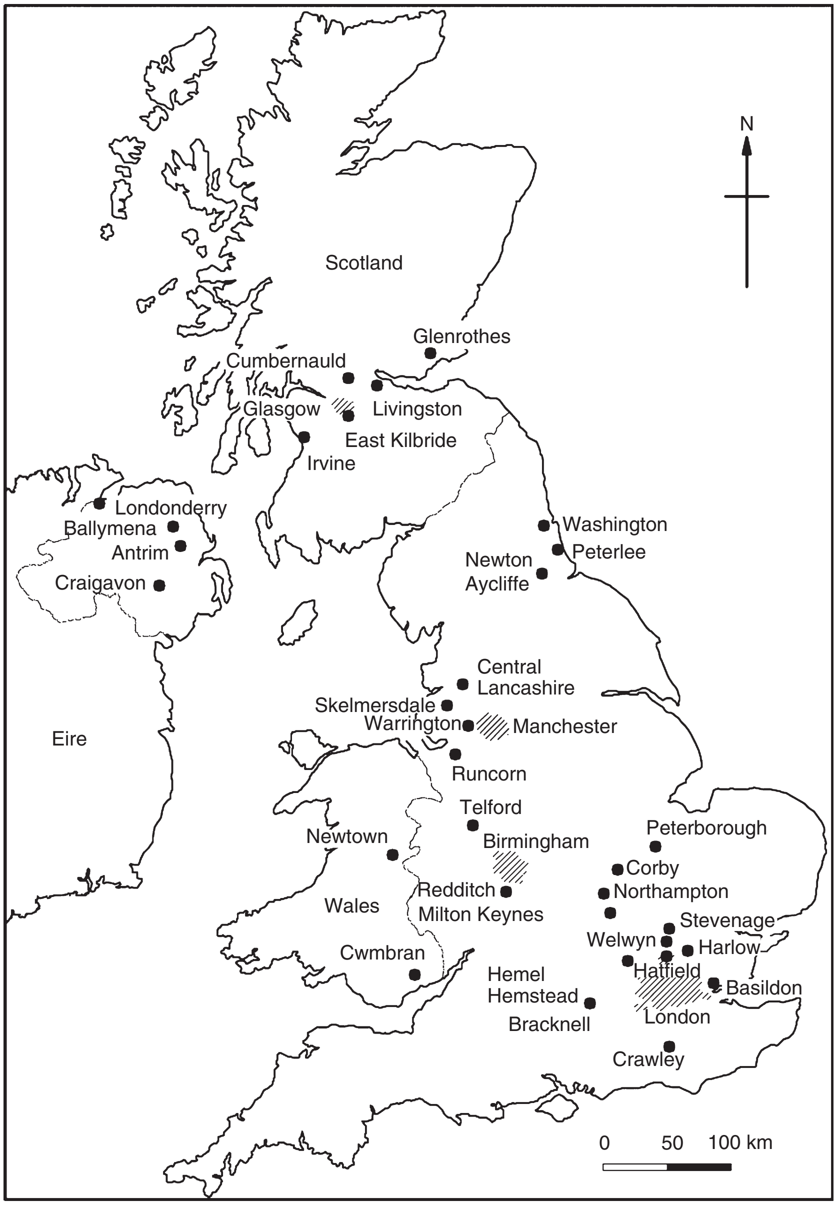

Figure 1 United Kingdom. Location of New Towns.

were established around metropolitan areas: Obninsk, in Moscow, and the celebrated Academgorodok in Novo sibirsk. Satellite towns for Moscow (a city with over 10 million inhabitants in the 1960s) were created on the lines of London New Towns, Zelenograd and Elektrostal outside the green belt (of 10 km width), Klimki and Krasnogorsk inside.

New Towns Policies in Other Parts of the World

In the 1960s and the 1970s, countries outside Europe took the lead in planning New Towns, responding to a strik ingly rapid rate of urbanization and, to a lesser extent, to regional inequalities. Problems are aggravated by huge refugee influxes, effected by wars, partitions, changes in legal status, shifts in the political orientation of govern ments, or immigration policies. As an indication, the population of Greater Shanghai and Greater Tokyo nearly doubled in the 20 years from 1960 to 1980 (from 7.7 to 15.0 million and from 10.7 to 20.0, respectively).

Japan, the most industrialized country in Asia, decided to provide housing for industrial workers and the influx of migrants into towns after the war. The first New Towns housed overspill population from Tokyo (Tama New Town), Nagoya (Kozoji New Town), and Osaka (Senri New Town). The second generation New Towns in the 1980s were meant to create 'technopoles' throughout the country, on the model of Tsukuba, the academic satellite

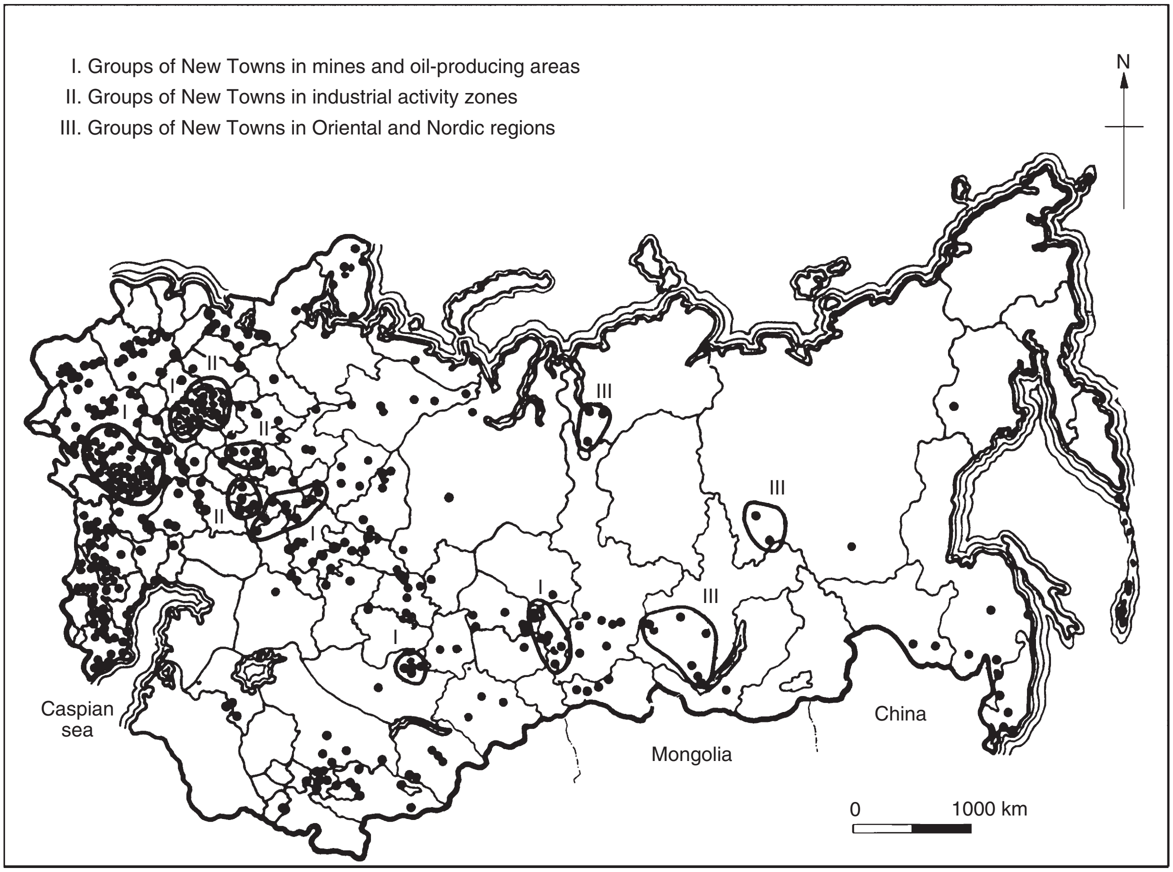

Figure 2 New Towns in ex-USSR. From Cahiers de l' I.A.U.R.P. No 38/ 1976. Permission kindly granted by the Cahiers.

outside Tokyo (1960s). From 26 sites designated around nearly all metropolitan areas, eight technopoles within 300 km from Tokyo developed as ground for modern high end technologies. The conurbation of Osaka–Kyoto–Kobe (named Keihanshin) is now emerging as a vast port industrial area of 16.8 million people, which hosts three New Towns, that is, two technopoles in the making and one scientific city, Kansai, an antipode of austere Tsukuba in terms of urban and housing program.

Singapore city state and Hong Kong former city state (since 1997 a region of China) had to house slum dwellers after the war, among which many were immi grants from Malaysia or China respectively (1.1 million squatters in Hong Kong in mid 1960s, 25% of its population). Both cities took up in earnest the con struction of New Towns (**Figures 3** and **4**) (formerly called satellites), places with economic and social life, a target that has not always been evenly managed. Public funds have guaranteed the completion of infrastructure and dwellings, and the provision of a variety of housing tenures. Construction has implied large scale physical transformation and land reclamation. In Singapore, after its independence from British rule in 1959 and the separation from the federation of the Malaysian states in 1965, the government created 23 New Towns, relatively self sufficient and interdependent entities of 250 000–300 000 people. The population of Singapore is estimated at 4.5 million (2006), of which 2.5 live in New Towns. The relevant figures for Hong Kong are 6.7 and 3.1 million, respectively (of the nine New Towns; Tsuen Wan alone has over 750 000 thousands). Notwithstanding the undeniable success of offering good affordable housing, and in many cases manufacturing and industrial activities, Hong Kong New Towns are criticized for their dependency on the central area for tertiary sector jobs. However, given that China has established Shenzhen 'special economic zone' across Hong Kong borders, the peripheral position of some New Towns may change.

China's urbanization is, above all, of tremendous scale. In the last two decades about 500 million people moved from agrarian to urban areas, a move mostly poorly planned, and in the next two decades nearly as many are expected to do the same. Shanghai, the nation's largest city, is now going to have seven new satellite or New Towns, according to the European press. Also Beijing (8.6 million in 2001) has 11 New Towns underway, in order to house an additional 5.7 million (they already have 2.5 million). The new city of Shenzhen, just outside Kong Kong, with 8.3 million in 2005, was a small village before 1989, when the first special economic zone was

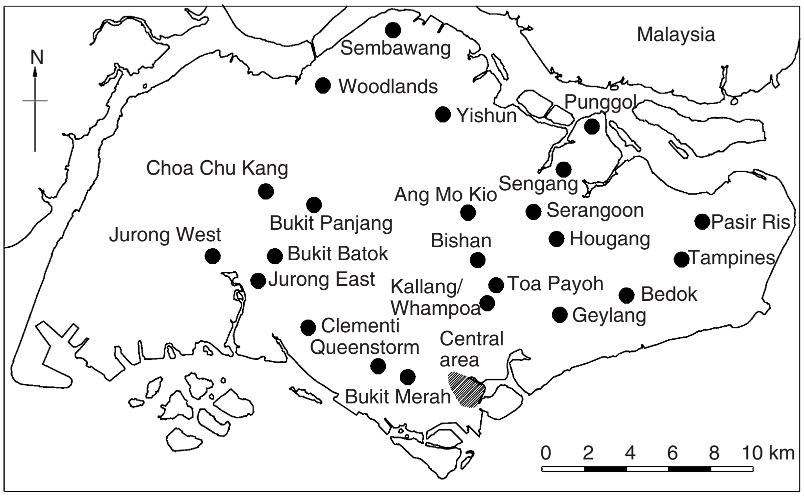

Figure 3 Singapore. Location of New Towns.

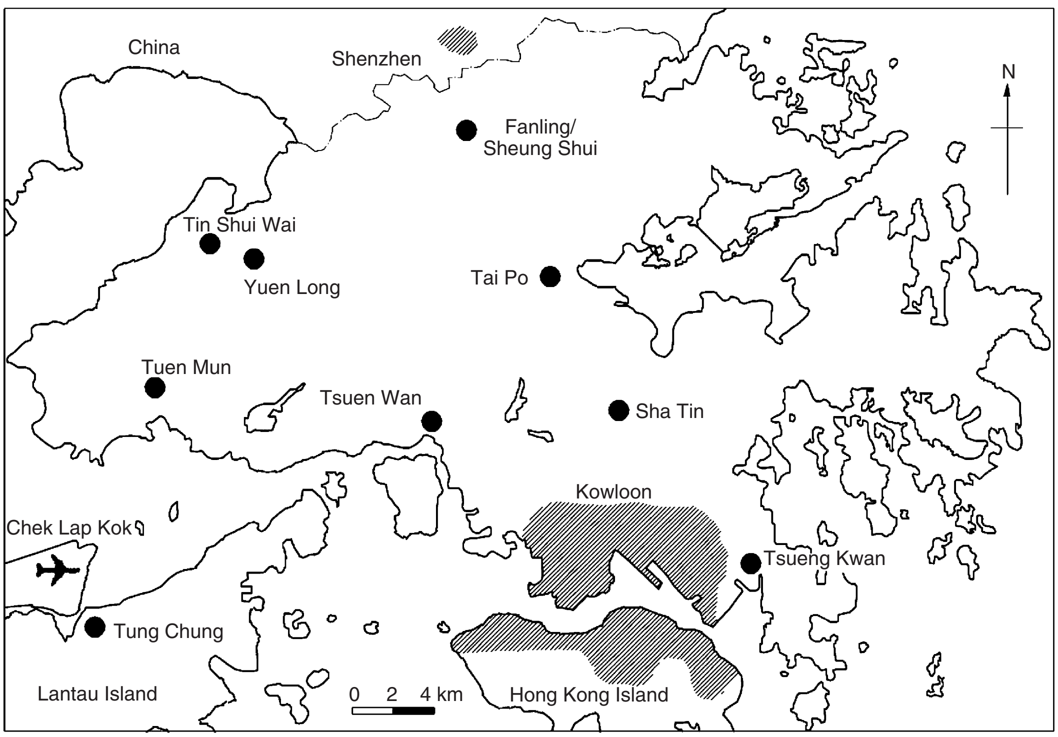

Figure 4 Hong Kong. Location of New Towns.

designated, having Shenzhen at the center. In the population figures are not included, the thousands who stay in temporary settlements outside the city, waiting or just hoping for a permit to establish there.

The creation of New Towns in post war India should be cast in the framework of the euphoria after in dependence (1947) and the search for equality and social justice. During Nehru's first government, decentralization was pursued as a means of balanced economic growth, through the installation of new nodes of production, education, and research. About 100 New Towns were planned, most of them near industrial plants, four state capitals (Chandigarh of Punjab State, Bhubaneshwar of Orissa, Gandhinagar of Guiarat, and later on Dispur of Assam), and also towns for refugees (**Figure 5**). Examples of planned urbanization include Durgapur in West Bengal State, Bhilai in Chhattisgarh, Bokaro in Jhark hand, all near large steel plants (of 500 000 each). Faridabad in Haryana acquired a next door New Town in 1947, in order to host refugees from Pakistan after the partition (2.2 million people for both in 2001). Navi Mumbai (New Bombay) in Maharashtra State was the brainchild of the autonomous Maharashtra Industrial Development Corporation (1970), which, among other progressive measures, established Bombay's counterpart, in order to relieve the historic city from the pressure of the tertiary sector (1.5 million for both in 2001).

In the oil producing Iran, the first New Towns of mid 1970s seemed in line with similar initiatives of their time. They were to house overspill population, but the program came to a halt when the ayatollahs took over in 1979. The warfare between Iran–Iraq (1980–88) pushed some 2.5 million people from the western and south western zone into safer areas. In addition, as the new regime abolished many of the Shah's family planning measures, rapid population growth aggravated extreme poverty and unemployment (urban to overall population in 1976 and 1986, 16/34 and 27/49 million, respectively). The need for nation wide measures generated the designation of 18 New Towns in 1985, around Tehran (four), Isfahan (four), and ten other cities. Their population was projected at 2 931 000. Under state coordination, semi private development companies were created, from which, only 12 have only made a first start (2004). Tehran's New Towns have 52 630 inhabitants, when over 900 000 city dwellers live in informal settlements, 300 000 of them in one informal town, Eslamshahr, which drags its population not only from rural areas but also from the capital.

The New Towns of Israel is a unique case. They are one of the tools toward the colonization of Palestine by the Israeli State, after its creation in 1948. Expelling the Arabs of the state, that is, two thirds of nearly 1.6 million people in 1946, was carried out between 1946 and 1948, before the incoming of Jews from Europe, Africa, and Asia, on the whole 884 800 people between 1948 and 1958. They were transferred to the New Towns or development towns (**Figure 6**) , over 30 till now (the ones on the map plus Azur, Bet Dagan, Or Yehuda, Rosh Ha'Ayin, Ramat, Tirat Karmel, Yishai, Yehud, Yoqne'am Illit). About 14 of them were established in 1949–50, mostly in towns that were evacuated from their Arab

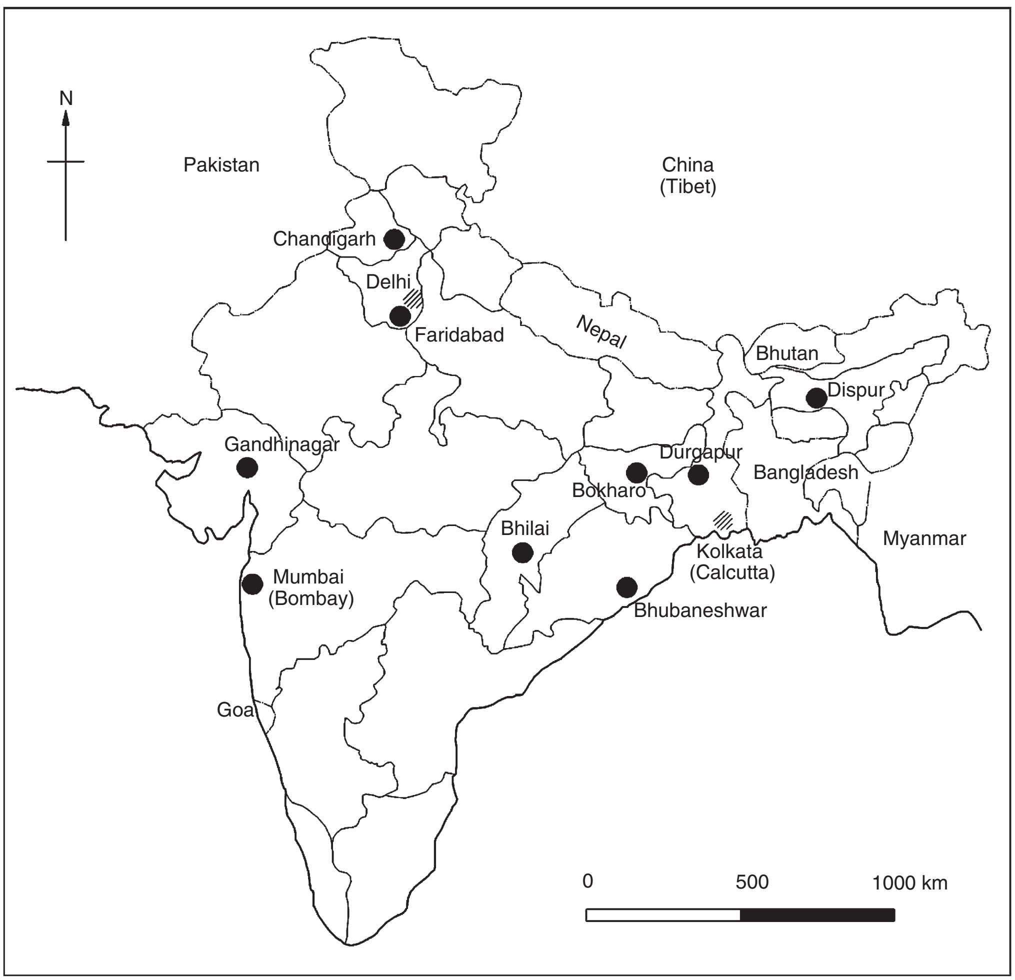

Figure 5 India. Location of New Towns.

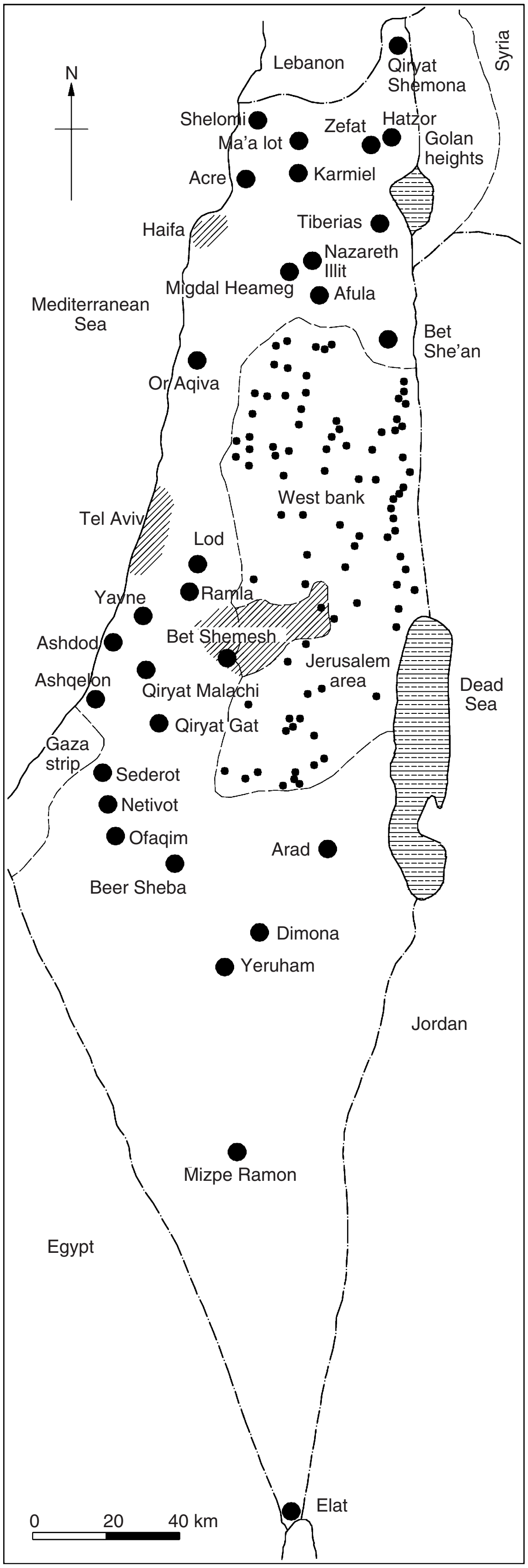

Figure 6 Israel. Location of New Towns.

population and were renamed and redeveloped within months (at least 400 Arab villages had been destroyed previously). A second wave of New Towns in 1951–56 headed at a Jewish majority in Negev and Galilee; finally, the last two settlements, Arad in Northern Negev (1961) and Karmiel in Galilee (1964), rehoused already established Jews and immigrants. The newest flow of 911 000 Jews from ex USSR in 1989–91 did not move to New Towns (save 130 000), but to most advantaged places, economically and socially. Available sources give the New Towns a total population of 800 000. The larger one is Beer Sheba in Negev (over 100 000, a university, etc.). However, the Judaization of the land apparently lags behind, and 30 more New Towns have been announced again in Negev and Galilee. In addition, since 1967 Israeli city making has expanded to the occupied territories of the West Bank and the Gaza strip (from which Israel withdrew recently). Jewish settlements in occupied lands are estimated at a number of 212 and 235 000 inhabitants, ensuring the control of about 60% of the land and specifically of Greater Jerusalem. Immigration is still encouraged. Israeli settlements are marked by thoroughly worked out racial segregation: series of electrified barbed wire fences, intervening minefields, highly sensitized TV–radar installations, night cameras, thermal imagers and sound sectors, armed guards, etc., 'protect' their Jewish population, usually immigrants from other parts of the world, against indigenous Palestinians.

Discussion

This, by no means exhaustive, presentation has focused on the most typical cases of a universally yet disparately applied tool for urban and regional restructuring. A lot more large cities have set New Towns in order to cope with future growth, Seoul, Kuala Lumpur, Cairo, Caracas, and others; sporadic installation of industrial towns includes widely different examples like Jubail and Yanbu New Towns in Saudi Arabia, near enormous industrial infrastructure (late 1970s), or Aspra Spitia in Greece for a single aluminum plant (1965). The most spectacular personalities of the New Town species are the new capital cities, created to fulfill national aspirations or governmental ambitions. Famous for its architecture, Brasilia is also known as the epitome of the distance between plans and reality. In the official city live 300 000 people, whereas nearly six times as many stay in 15 peripheral, unplanned settlements.

Creation and subsequent progress of New Towns are strongly connected to political will and planning tradition. The first British New Towns can be seen as the child of the end of war housing shortage and the victory of the Labour Party. They were committed to self containment, that is, balance of jobs to workers, and full social amenities, seeking social class balance. This was a state led version of the pre war garden city thinking, the physical characteristics of which ('a town in the country'), penetrated the formation of New Towns elsewhere, in France, Scandinavia, Israel, and even in Eastern Asia. A phenomenon of comparable strength, as far as political orientation and commitment is concerned, is the New Towns in postindependence India.

Organized installation of self contained entities was attracted to regional policies in Europe in the late 1960s and 1970s. However, the impact of the New Town strategy was minimal in Europe, where they are very seldom seen as new nodes operating within a national system of cities. Outside Europe, Siberia of former USSR and – perhaps – India may be regarded as the ground where regional policies were applied more thoroughly. It is in the field of urban containment and comprehensive housing supply that New Towns have performed well almost everywhere. Coordination of various policy instruments, such as land use planning, capital spending on infrastructure, and transportation planning, so as to balance housing, jobs, and facilities, was accomplished with considerable success, even if with delays in the first, 'heroic', times.

In the five decades after the establishment of most European New Towns, changes in administration or the public attitude, in technical knowledge or design values have affected various aspects of New Town making. In Britain during the 1960s, relocation of industries to areas (with a cheap labor supply made New Towns a favorite place for young working class families. During the Thatcher administrations, the public housing emphasis vanished, and upward social mobility influenced the community profile of several New Towns. In France, the need for rented accommodation gave way to the wish for ascent to private ownership. New Towns that did not adjust their policies in time suffered from population loss (as Evry). New Towns of Stockholm were less self contained than suburbs, due to an excellent rail system, which is now in line with the latest views about sustainable mobility. The opposite example is Milton Keynes, depending totally on private car.

Successful managing of land assets was a crucial issue in New Town economics, from the most committed European policies to the hesitant rise of social housing in early post war Japanese New Towns, or to the abortive case of Iran, which can only be regarded as a negative

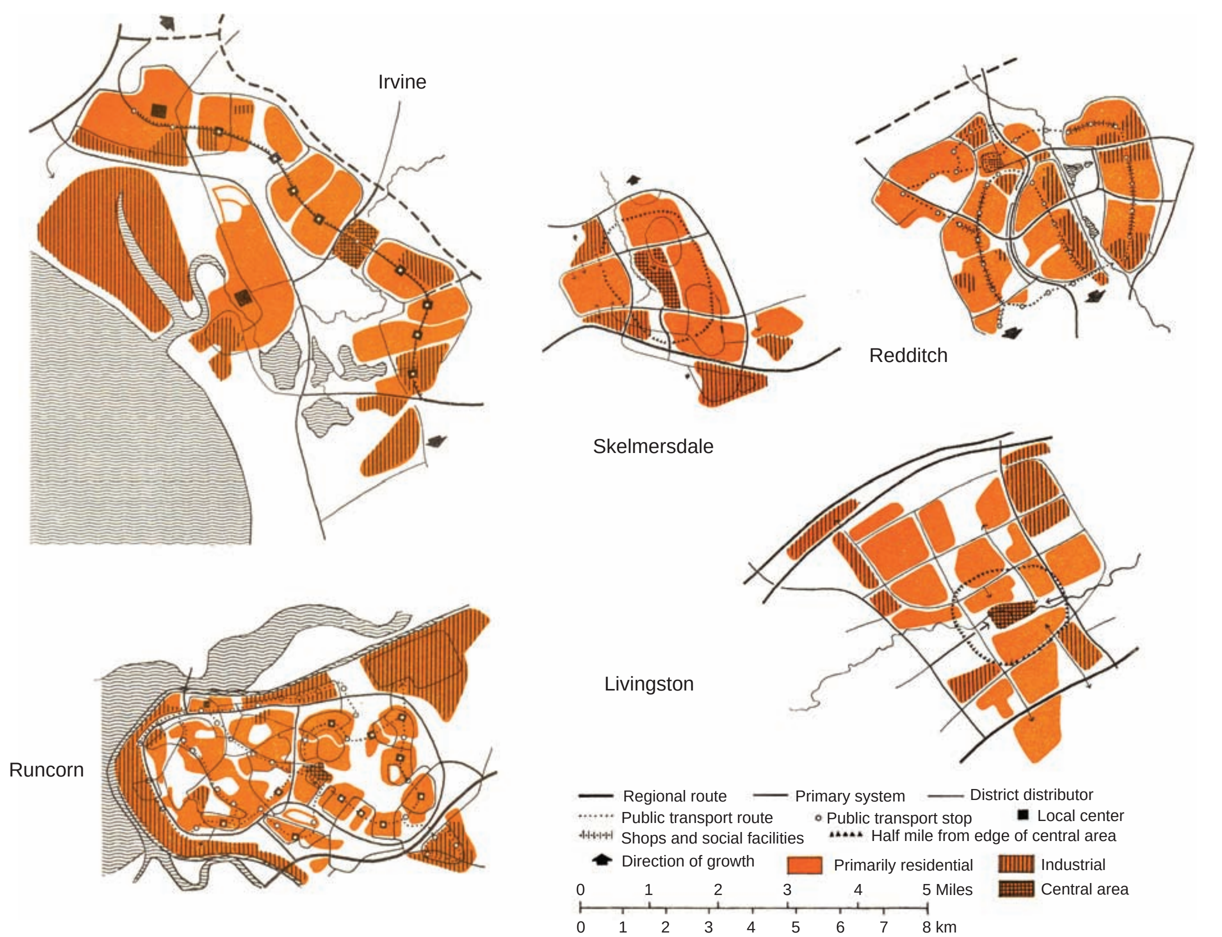

Figure 7 Plans for New Towns in the UK. From Town Planning Review No 3/1968. Permission kindly granted by the T.P.R.

Figures 8 Shopping center in Stevenage, UK.

Figure 9 Shopping center in Cergy-Pontoise, France (from a promotional poster).

example. Large scale land acquisition at favorable prices made the provision of housing and public utility services feasible, while keeping land for future needs (as is the case of the French New Towns). British New Towns were forced to wind up by the Margaret Thatcher administration in the 1990s, that is, sell out to private firms pockets of publicly owned land ripe for development. Thus, planning was frustrated from making the best of a public investment.

City building skills, urban architecture, planning, landscape design, and engineering proved important for the vigorousness of urban life. Plans for the first British New Towns conformed to the planning orthodoxy of the 1950s and 1960s, with their strict separation of uses and the application of the neighborhood unit concept **Figure 7**). They are now criticized for overly simplistic approach. In terms of architecture, brick built, pitched roof dwellings characterized British New Towns. The Scandinavian ones, with free blocks and towers, pursued the vision of a 'city in a forest', and were perhaps the prototype for Hong Kong or Singapore New Towns of 30 story blocs of flats. In France, first operations were felt as a fertile ground for pioneering architecture. In their enthusiasm, architects designed artificial and illegible environments, which were to be called science fiction utopias. Among planning innovations, open air shopping centers, a modest version of today's consumption temples, can be credited to New Towns (as in Stevenage (**Figure 8**), Cergy Pontoise (**Figure 9**), and others).

Today, ample land and the tradition of innovation make New Towns well equipped to respond to the challenge of more sustainable urbanism, and also of city marketing. Most European New Towns are now part of the local administration system. The discussion about them, while not overseeing pre mentioned issues, focuses on whether they are cherished by their people, that is, whether their urban, social, and cultural identity stimu late a sense of attachment to them. The ultimate target for New Towns is to become home towns.

See also: Edge Cities; Growth Poles, Growth Centers; Refugees and Displacement; Squatter Settlements; Urbanization.

Further Reading

Coon, A. (1992). *Town Planning under Military Occupation. An Examination of the Law and Practice of Town Planning in the Occupied West Bank*. Worcester: Dartmouth.

Hall, P. (1976). *Urban and Regional Planning*. Middlesex: Penguin.

Merlin, P. (1997). *Les Villes Nouvelles en France. Que Sais Je ?* Paris: Presses Universitaires de France.

Shaffer, F. (1970). *The New Town Story*. London: Paladin.

Sorensen, A. (1998). *Thematic Issue: Villes Nouvelles. Urbanisme* 301.

Sorensen, A. (2002). *The Making of Urban Japan*. London: Routledge.

Ward, C. (1993). *New Town, Home Town*. London: Caluste Gulbenkian Foundation.

Ward, S. (2004). *Planning and Urban Change*. London: Sage Publications.

Relevant Websites

http://www.cedd.gov.hk/eng/about/achievements/regional/index.htm
Civil Engineering and Development Department, Hong Kong.

New Urbanism

E. J. McCann, Simon Fraser University, Burnaby, BC, Canada

Glossary

Communitarianism A philosophy that advocates the good community as the foundation of society. Communitarians oppose the individualistic tendencies of liberalism and, instead, see human identity as relational – always embedded in a social context. Individuals' relationships within communities draw them into political participation, etc. which strengthens the community. Recent versions of this approach call for a return to participation in community organizations in order to build social ties, social capacity, or social capital.

HOPE VI An acronym for 'Housing Opportunities for People Everywhere' HOPE VI is a program of the US Department of Housing and Urban Development, focused on changing the form of public housing, empowering residents, and locating public housing tenants in mixed income contexts. Among other strategies, it has adopted New Urbanist design principles as a reaction against high-rise 'projects' of the past.

Smart Growth A political movement and planning strategy, originating in the US, which seeks to identify and overcome the negative impacts of urban sprawl. Focus tends to be on the preservation and reinvigoration of inner cities and inner suburbs as a way of discouraging development on the urban fringe. Advocates also try to preserve open space and farmland, frequently by using incentives to developers and businesses as much as regulatory limits to expansion.

Introduction

The New Urbanism is a movement intent on improving quality of life in cities by changing urban form. Its early iterations are most commonly associated with the work of a small group of US based architects in the 1980s. Yet, it is worth noting that early New Urbanists were partly influenced by the work of European architect, Leon Krier and his neotraditional town of Poundbury in the English county of Dorset. Indeed, the movement in general tends to hold certain European urban forms, such as Italian hill towns, in high regard. As a complex combination of influences distilled into a straightforward set of planning principles, New Urbanism has become a major force in planning practice in North America, Europe, and, to a lesser extent, other parts of the world.

New Urbanism's principles combine a critique of car oriented cities with a set of proposals on how urban built environments can be reformed to foster community and reduce negative human impacts on the environment. The ideal New Urban form is compact, features a mix of residential and tenure types, and includes a range of land uses, such as places of worship, schools, shops, and workplaces within walking distance of houses and apartments. Proponents believe that it is a form that encourages the development of community by emphasizing propinquity through compact housing, walkability, and abundant public space. In aesthetic terms, New Urban developments tend to exhibit an eclectic style, although one that is frequently marked by neotraditionalism – a nostalgic attraction to idealized architectural details, such as porches and picket fences. This is especially the case in the North American developments designed or influenced by Andres Duany and Elizabeth Plater Zyberk, although Krier's Poundbury has also been critiqued for its eclectic mix of traditional styles. As the principles of New Urbanism have seeped into planning practice across the world, however, traditionalism has become somewhat less common as attempts to encourage (a return to) compact city form have come to dominate neotraditional design.

Arguments for this idealized urban form are coupled with a strong political agenda which seeks to replace traditional decision making practices in urban planning with a bottom up model of participatory planning 'charrettes' – intense collaborative public meetings in which visions are articulated and where the details of designs are developed by architects, planners, developers, residents, and other interested citizens. The growing popularity of New Urbanist ideas reflects many political leaders' receptiveness to almost any well presented scheme that offers to mitigate the social and environmental problems of contemporary urbanization. For example, English cities were the focus of a government report in 1999, entitled *Towards an Urban Renaissance*. The report, which has informed subsequent government urban policy, echoed many New Urbanist ideas as it advocated for a sustainable urban form that fosters mixed uses, walkability, cycling, and transit. In the US, the federal Department of Housing and Urban Development has incorporated New Urbanist ideas into 'HOPE VI' public housing programs. These national level adoptions of New Urbanist principles have been accompanied by positive responses from mayors of cities across North America and increasingly in European cities such as Berlin.

New Urbanism's popularity also reflects the persuasive abilities of some of its leading lights. US based leaders of New Urbanism include San Francisco based Peter Calthorpe Associates, Miami based Andres Duany and Elizabeth Plater Zyberk and their Duany Plater Zyberk & Company (DPZ) architectural firm, and the RTKL firm, which originated in the Washington, DC area. Besides Leon Krier, the British engagement with New Urbanism has also been influenced by Richard Rogers who chaired the task force that produced the *Towards an Urban Renaissance* report and a subsequent update report, *Towards a Strong Urban Renaissance*, which reasserted a vision of compact well designed cities. These designers, and a number of allied writers like Peter Katz, James Kunstler, and Philip Langdon, have tirelessly promoted and popularized New Urbanist ideas and connected them to other reform agendas, such as those of the sustainable urbanism, compact city, and smart growth movements.

In the 1980s and early 1990s, New Urbanists found it difficult to convince planners in the US to approve the mixed use zoning needed to build their developments or to find developers and financiers willing to partner with them on new developments. Yet, today their ideas are commonplace in planning departments not only in the US but increasingly in other parts of the world. Furthermore, they are now seen as viable and indeed attractive investments by developers and lenders, having, for example, received a corporate 'stamp of approval' through their use by the Walt Disney Company in its Celebration development in Orlando, Florida.

There are now numerous companies designing in a New Urbanist style. A study published in 2001 noted that since the 1980s, over 400 New Urbanist developments had been started in the US alone. Perhaps more profoundly for the future of cities, numerous municipalities have adopted 'form based' (as opposed to standard 'use based') planning codes and regulations formulated by New Urbanists working as consultants. DPZ, for instance, is involved in both building and code development. The firm currently lists completed developments (like its original and iconic Seaside in Florida), ongoing plans, urban redevelopment projects, and a series of other projects in North America, South America, Europe, Asia, and Oceania. Calthorpe Associates list projects on four continents, and RTKL operates globally from offices in a number of North American, European, and Asian cities. (Although, it should be noted that the literature suggests that New Urbanist developments are largely concentrated in the cities of North America and Europe with little significant presence evident in the rapidly growing cities of Asia, for example.)

New Urbanism's foci on the urban built environment, public space, and questions of development and ecology make it an obvious object of geographic scholarship. To understand these foci, it is worthwhile briefly examining the strands of planning thought that have influenced or paralleled the movement's rise.

Sources of New Urbanist Ideas

New Urbanism is a somewhat diverse movement with a number of distinct, if connected, emphases – such as environmentalism and neotraditionalism. Similarly, its intellectual development has been characterized by a careful and selective appropriation of ideas from earlier periods of planning thought, often to the point of bringing together ideas that can be seen to be at odds with one another.

A central figure in the intellectual history of New Urbanism is Jane Jacobs, whose classic 1961 book, *The Death and Life of Great American Cities*, marked the beginning of the backlash against urban redevelopment and top down, car oriented planning in the US. Jacobs, like many New Urbanists, saw the vibrant urban neighborhood as the best model for planning. Her touchstone was the Greenwich Village of the 1950s, an exuberant, diverse, and organically growing palimpsest which developed in a dense grid of streets which were populated at most times of the day and night by an array of different characters and activities. From her experience living in the Village, she proposed four conditions for the good neighborhood: it must serve more than one economic function, it must have mostly short blocks on a grid, buildings must be of various ages, and it should contain a densely concentrated population. Three of these criteria have been central to the New Urbanist perspective but it has been difficult for the movement – which largely deals in newly built developments, whether in the suburbs or in central cities – to incorporate older buildings. Jacobs' ideas resonate in some ways with those of Kevin Lynch who, writing around the same time, argued for an urban form that was 'imageable' – easily read by its inhabitants. His work identified the importance of clearly defined paths, edges, districts, nodes, and landmarks in cities. Lynch's ideas have accompanied those of Jacobs into New Urbanist thought. Yet, these appropriations are selective and strategic. The newness and intricately planned nature of most New Urban developments, for instance, stand at odds with the ideas in Jacobs' book which tend to laud urban neighborhoods that have grown organically and somewhat haphazardly. After reading Jacobs' work, one might ask how possible it is to 'plan' diversity and vibrancy, as the New Urbanists hope to do?

Traces of the 'garden city movement' are also evident in New Urbanism. Originating in the work of Ebenezer Howard at the turn of the twentieth century, garden cities are a paradoxically anti urban idea. Developed through a critique of both the perils of life in the industrial city and of residential suburbs and rural areas, the garden city idea sought to construct a new type of city which incorporated the best of each – jobs, open air, community, etc. – through the careful planning of new communities. This utopian retreat from the unruly city distinguishes Howard's ideas from those of Jacob's. In fact Jacobs explicitly criticized both garden city ideas in general and specific built examples for their over planned anti urbanness, questioning, for example, the separation of cars and pedestrians onto different paths and roads in Radburn, New Jersey.

Nonetheless, the legacies of Jacobs and Howard have been somewhat improbably combined in New Urbanism. New Urbanist's ideal city form looks more like Jacobs' than Howard's, yet there is a strong desire in New Urbanism for order and beauty on the landscape, rather than the complex and not always aesthetically pleasing jumbles of urban life that inspired Jacobs. Nonetheless, New Urbanists have been able to accommodate these contradictions in their own minds and develop from them a strong utopian vision of the future of cities. Critiques of the inconsistencies and silences in their agenda remain, however.

Out with the Old and in with the Older? The New Urbanist Agenda

The New Urbanist critique of the contemporary urbanism is not without merit. Neither, for that matter, are some of the movement's proposed changes to urban form. Yet, questions remain as to whether their critiques and solutions – which are focused on urban form rather than on the processes that produce negative and unsustainable cities – will affect radical change or whether, instead, they are likely only to produce more convivial urban built environments for the wealthy few, fostering gentrification rather than more comprehensive forms of urban renaissance. Before engaging with these critiques, it is important to detail the New Urbanist agenda.

Clash of Utopias: The New Urbanist Critique of Post-War Modernist Urbanism

The New Urbanist critique of the way that cities have been planned, designed, and developed since the end of World War II is, in essence, a critique of a particular vein of planning thought. Planning in the post war period was part of a wider modernist movement and thus contained strains of utopian thought. These utopian impulses in planning envisioned, among other things, an urban future where the rational planning of land uses and infrastructure, coupled with particular building design principles, would produce socially just cities. The economy would operate efficiently and for the benefit of society through the development of more effective infrastructures; harmful land uses would be separated from population centers through horizontal functional zoning; and residents would have access to nature through various means, including urban parks, efficient automobile transport to the countryside, and large platt residential zoning which allowed substantial yards around detached single family homes. Yet, the modernist project was internally variegated and involved contradictions and unintended consequences. These are the contemporary urban problems that have been in the cross hairs of New Urbanist critique since the 1980s.

Urban form is, then, at the center of the New Urbanist agenda. From an original focus on US suburbs (at one point, an observer labeled the movement, 'the New Suburbanism') New Urbanists argue that suburban form promoted social distance and encouraged profligate energy use. They point out that standard zoning creates cities with largely similar housing types clustered together – large lot single family here, slightly more dense single family here, and multifamily housing (apartments, townhouses, etc.) over there. This form ensures a general separation of incomes, with all those who can afford a particular type of housing living together, their children attending the same schools, and so on. Social cohesion and a sense of community is, therefore, very difficult to maintain or promote in this fragmented urban geography, say the New Urbanists. Furthermore, they question the very basis for traditional forms of functional land use zoning in 'postindustrial' countries. They suggest that since cities (or at least the North American and European cities in which the movement is largely based) are now much less the centers of heavy, polluting industries, standard forms of single use functional land use zoning are now less tenable, as is post war planning's unease with mixed use zoning.

The (sub)urban form of the mid and late twentieth century is not only critiqued for its segregation of housing forms, classes, and presumably by extension, ethnicities. New Urbanists also focus on details of design and engineering and their social and environmental consequences. The contemporary city is, of course, car oriented. The standard post war model for roads in the suburbs eschews the compact grid pattern of many older urban centers in favor of wider, curvilinear streets that make driving more convenient, often at the expense of walking. This traffic model relies on a dendritic, or tree like, organization of streets and roads, where the smallest local streets in residential areas – ones which often end in cul de sacs or have other design features to discourage through traffic – branch into larger capacity collector streets, intended to carry traffic to local 'traffic generators' such as small shopping centers. Collector streets, in turn, feed into major arterial roads and highways, which service major 'traffic generators' like large workplaces, regional malls, sports stadia, etc., and eventually lead to freeways. This organization of roadways accompanies and necessitates a number of other landscape forms – such as large parking lots and ubiquitous median strips – while demanding certain lifestyles – for example, the need to drive via busy arterials in order to shop, go to school, or visit a park, or a tendency to be separated from neighbors as a result of traveling mostly by car rather than on foot.

New Urbanists point to the rise of increasingly anomic and placeless cities. They tie the blame for this situation to questions of urban form. (Indeed, supporters of the movement have gone so far as to link the 1999 Littleton, Colorado school shootings to the problems of suburban design and have proposed the New Urban form as a solution while the British urban task force discussed above focused on crime and urban design in their re ports.) They also critique the character of decision making in post war urban planning, suggesting that planning has become rigid and bureaucratic and has fallen under the spell of engineers and elitist avant garde architects who are itching to test their latest designs on an unsuspecting public. They argue that, in general, standard planning is closed to ordinary citizens, due to its use of specialist language and Byzantine zoning codes. Any merits of utopian modernism in planning and design have, they argue, been lost and a new optimistic view of urban form is necessary.

Strategies for Achieving the New Urban Form

New Urbanism advocates a change in zoning regulations. Zoning should focus less on use and more on urban form, they argue. It should, more specifically, be flexible enough to encourage mixed housing and tenure types. This, they suggest, will promote the social and economic mixing and community building that standard zoning militates against. Mixed use zoning codes also encourage compactness, walkability, and the 'human scale'. Build ings in which apartments are located above shops or restaurants should be encouraged, for instance. Similarly, zoning should allow for the development of mixed use nodes within walking distance to transit stops, in order to encourage alternatives to car travel – these transit oriented developments have been associated most strongly with Peter Calthorpe's brand of New Urbanism. Duany and Plater Zyberk, for their part, propose a zoning scheme, based on transects of land stretching from rural to urban areas, which emphasizes coherence of land use within specific areas. Apartment blocks are defined as an appropriate form for an area defined as very urban, ranch houses for other areas, for instance. The transect approach also indicates that New Urbanists have emerged from the suburbs in recent years and increas ingly deal with issues of inner city redevelopment which has, in turn, placed them at the center of the politics of gentrification.

New Urbanism's fastidious approach to urban plan ning is evident in the details of New Urbanist design. Model developments like Kentlands, Maryland feature single family homes, townhouses, and apartments in the same project, although it would be a stretch in most cases to define a significant number of these units as affordable housing. Neighborhood compactness is also central to this design approach. Houses are placed close together, at the expense of large yards and – ideally at least – shops, schools, workplaces, churches, community centers, and recreation facilities will be located among or within walkable proximity to the housing. Thus, cars will be less necessary and walking and cycling will become the common modes of travel. Streets in New Urbanist de velopments tend to disdain cul de sacs in favor of intersecting patterns, designed to encourage through traffic – cars, bikes, and pedestrians. This interconnection is intended to dispense with the need to travel long circuitous routes along arterial roads to reach stores, schools, etc. Road intersections are designed with rela tively sharp corners, rather than the wide, flowing cor ners of standard suburban streets. This reduces the width of the intersection and makes it easier for pedestrians to cross. This is part of the 'human scale' approach and, it is hoped, encourages walking. This detail oriented focus on improving the quality of pedestrians' lives also extends, for instance, to the details of sidewalk paving materials.

Design must also allow residents to clearly imagine the edges of their neighborhood, its main nodes, or community gathering points, etc. New Urbanists there fore pay great attention to planning these elements in order to enhance neighbors' sense of place. The social goal underlying these strategies is to promote com munity. New Urbanists believe that people will engage more with their neighbors if the form of their neigh borhood encourages them to be on the sidewalks and in local public spaces. The classic example of this is the use of front porches in many of these developments. In a nostalgic turn, the architects place porches close to the street with the hope that chance conversations will occur between homeowners and sidewalk strollers. For New Urbanists, these conversations will form the basis for building community and social capital.

The ideal of community is also central to New Urbanists' contribution to contemporary urban decision making processes. As an alternative to the top down bureaucratic planning they criticize, New Urbanists in voke the notion of the 'charrette', a term used in archi tectural practice to define an intensive, collaborative effort to develop ideas and designs for a project. This type of collaborative 'visioning' process is intended to break down barriers – between 'expert' planners, archi tects, engineers, or developers and residents or 'stake holders' while also building bonds among different groups of residents and stakeholders. Thus, the charrette process provides the opportunity for community building in a physical and a social sense. The outcome of this sort of process is, ideally, a consensus based plan with stake holder buy in, rather than a plan designed within the enclosures of the state and then presented to residents only for comment.

It is worth emphasizing once more that New Urbanist principles are ideals and that the strategies intended to achieve them are best practices. In reality, there is an uneven geography to New Urbanism. From the 'flagship' developments like Poundbury and Seaside to the various infill projects in a diverse array of cities in North America, Europe, and beyond, a great deal of variation exists in how New Urbanist ideas are used. As Jill Grant's book, *Planning the Good Community*, illustrates, the uptake of New Urbanist principles and their effectiveness in achieving all they promise differs depending on a range of social, historical, and geographical factors, even in 'showcase developments' such as Karow Nord in Berlin. Furthermore, New Urbanism's celebration of compact, central city living dovetails perfectly with the global phenomenon of gentrification. In this context, many of the movement's core principles serve as perfect mar keting tag lines for developers.

New Urbanism as a Social Movement

Before turning to geographers' and others' engagements with these ideas, it is worth emphasizing that New Urbanism is not merely a force within the design pro fessions. Rather, it should be thought of as a powerful and effective social movement that has been able to get its ideas listened to and its designs built and incorporated into urban policy. This has been achieved in a number of ways. First, New Urbanists engage in persuasion of other architects and planners and of politicians through en gagements in policy discussions, publishing books and scholarly papers, and organizing or participating in con ferences. The key players in the movement also offer their consultancy services to cities. They teach in various set tings, from classrooms to workshops and less formal gatherings, and through their almost religious zeal for their utopian vision, gain many converts who, in turn, have reshaped city policies. Through this process of en gagement, the leaders of the movement are often por trayed, intentionally or otherwise, as messianic figures – the saviors of the city. This image, whether accidental or purposively cultivated, has the potential to undermine the bottom up idealism of New Urbanism's approach to de cision making.

Perhaps with this tension in mind, New Urbanists have also organized themselves into a professional ad vocacy organization – the Congress for a New Urbanism (CNU) – with its own charter and set of written principles. As such, the movement presents itself as ser ious, firmly grounded, and broad based – not merely at the whim of a few 'starchitects'. The CNU provides weight and added persuasiveness to the New Urbanist cause. New Urbanists also act strategically to link their ideas to other like minded movements and current ideas that are influencing planners and politicians. For instance, while some New Urbanists have always been motivated primarily by the ecological consequences of urbanization, others identified the political importance of aligning the movement as a whole with the sustainable urbanism, antisprawl, and 'smart growth movements' in the 1990s.

Geographers' Engagement with New Urbanism

New Urbanism, with its claims about development, place, community, and human–environment interaction has been an object of study for geographers. While the discipline's journals occasionally feature articles that wholeheartedly accept or explicitly advocate for New Urbanism, most geographers accept some of New Urbanism's critiques of post war urbanism but tend to be somewhat skeptical of many of the movement's proposed solutions and their social, economic, and political implications.

Perhaps geographers' most common critique is that New Urbanism's approach to reforming contemporary urbanism is overly focused on questions of built form. It commits a 'physicalist fallacy' and indulges in a form of environmental determinism which holds that human behavior, social systems, and economic structures will change if the physical built environment in which they operate is changed. Arguments that neighbors will use their cars less and converse more when they live in houses with porches, built on walkable streets are re garded skeptically by geographers whose scholarship emphasizes the complex and multiscaled conditioning of everyday life. These conceptual objections have been accompanied by others from those who have empirically researched New Urbanist developments.

A second major critique of New Urbanism addresses the movement's claim that a better urban way of life will accompany increased 'community'. Both through their ideal urban form and through their decision making charrettes, New Urbanists invoke a notion of community building which mirrors longstanding communitarian concerns, which have recently gained new popularity through discussions of 'social capital'. Geographers have questioned this perspective in two ways. First, they note that there is a problematic equation of neighborhood and community in New Urbanist thought. Neighborhoods, as relatively small spatial units, neither define nor limit most communities, which are assemblages of social con nections stretching over large areas and various scales.

Certainly, there are benefits to the ideal of localizing certain social interactions (minimizing car travel, for example), yet it is misleading and simplistic to suggest that community can be defined and contained by a neighborhood. Second, 'community' for New Urbanists is seen to be wholly positive. Yet, geographers and many others have, through a great deal of empirical and con ceptual work, noted that communities are by no means necessarily progressive social forms. Community can often be exclusionary, with internal identity being de fined in terms of negative opinions of outsiders. Similarly, community can be internally oppressive and conserva tive, repressing difference and dampening critical thought in favor of conformity to certain social norms. Critics therefore question both the idealist and con servative tone of the New Urbanist agenda.

These questions are paralleled by critical geographers' understanding of New Urbanism as a powerful marketing scheme, as well as a social movement. They have examined the movement's marketing strategies, noting the ways in which particular stories about history, nature, culture, and place are 'cut and pasted' together into narratives that attract certain profitable niche markets of house buyers, as well as policymakers. These critiques of the inconsistencies, naiveties, profit motivation, and cul tural politics underlying the New Urbanist agenda should not lead us to ignore or dismiss the movement, however. Rather, New Urbanism should be taken seri ously by geographers and others precisely because of the way it sustains and draws sustenance from so many of the ideologies and practices that shape today's cities – from gentrification, to urban neoliberalism, to urban sustain ability. It is a powerful force in contemporary urbanism precisely because it proposes a number of broadly at tractive solutions to urban problems, even if some of those solutions do not stand up to critical scrutiny.

See also: Anti-Urbanism; Gentrification; Housing; Neighborhoods and Community; Planning, Urban; Public Spaces, Urban; Suburbanization; Urban Architecture; Urban Design.

Further Reading

Calthorpe, P. and Fulton, W. (2001). *The Regional City: Planning for the End of Sprawl*. Washington, DC: Island Press.

DETR Urban Task Force (1999). *Towards an Urban Renaissance: Final Report of the Urban Task Force*. London: Department of Transport, Environment, and the Regions.

DETR Urban Task Force (2005). Towards a stronger urban renaissance. http://www.urbantaskforce.org/UTF final report.pdf (accessed in Jul. 2008).

Duany, A. and Plater Zyberk, E. (1991). *Towns and Town Making Principles*. New York: Rizzoli.

Duany, A. and Plater Zyberk, E. (1992). The second coming of the American small town. *Wilson Quarterly* 16, 19 50.

Duany, A., Plater Zyberk, E. and Speck, J. (2000). *Suburban Nation: The Rise of Sprawl and the Decline of the American Dream*. New York: North Point Press.

Falconer Al Hindi, K. and Till, K. (eds.) (2001). *Special Issue: The New Urbanism and Neotraditional Town Planning, Urban Geography* 22(3), 189 286.

Ford, L. (1999). Lynch revisited: New urbanism and theories of good city form. *Cities* 16(4), 247 257.

Grant, J. (2005). *Planning the Good Community: New Urbanisms in Theory and Practice*. New York: Routledge.

Harvey, D. (1997). The new urbanism and the communitarian trap. *Harvard Design Magazine*, winter/spring. http://www.gsd.harvard.edu/research/publications/hdm/back/1harvey.pdf (accessed in Jul. 2008).

Kenny, J. T. and Zimmerman, J. (2003). Constructing the 'genuine American city': Neo traditionalism, new urbanism, and neo liberalism in the remaking of downtown Milwaukee. *Cultural Geographies* 10, 73 97.

Kleit, R. G. (2005). HOPE VI new communities: Neighborhood Relationships in mixed income housing. *Environment and Planning A* 37(8), 1413 1441.

McCann, E. J. (1995). Neotraditional developments: The anatomy of a new urban form. *Urban Geography* 16(3), 210 233.

Smith, N. (1993). Reasserting spatial difference. *Architecture New York* 1(1), 22 23.

Talen, E. (2000). New urbanism and the culture of criticism. *Urban Geography* 21(4), 318 341.

Till, K. (1993). Neotraditional towns and urban villages: The cultural production of a geography of otherness. *Environment and Planning D* 11, 709 732.

Veninga, C. (2004). Spatial prescriptions and social realities: New urbanism and the production of northwest landing. *Urban Geography* 25(5), 458 482.

Relevant Websites

http://www.smartgrowth.org/
Aims to Foster Attractive Communities with a Strong Sense of Place.

http://www.dpz.com/
Organisation Offering Planning and Architectural Services.

http://www.cnu.org/
Organisation Promoting Walking, Neighbourhood Based Development.

NIMBY

P. Hubbard, Leicestershire University, Leicestershire, UK

Glossary

Externality An unpaid for but experienced effect that results from an investment or land-use decision made by someone else. Externalities may be negative (e.g., the noise pollution caused by a new recycling facility) or positive (e.g., the convenience of having a recycling facility on one's doorstep). In economics, this phenomenon is often discussed in parallel with the idea of 'freeriding', whereby people benefit from a good or service without ever having paid for it: NIMBYism tends to be more concerned with perceived negative externalities.

Externality Field The area over which externalities are felt or experienced.

Gentrification The displacement of a lower-class group by upper-class homeowners and investors, often associated with the revaluation of specific census tracts or neighborhoods following years of decline.

Introduction

NIMBY – a widely used acronym for 'Not in My Back Yard' – is a term which summarizes the oppositional rhetoric of those residents concerned about the potential impacts of new developments on their neighborhood or locality. The term has hence been used to describe the arguments of those opposing the development of potentially hazardous facilities (landfill sites, power plants, incinerators, and mobile phone masts), as well as protests against 'undesirable' human service facilities (such as alcohol rehabilitation centers, prisons, and homeless hostels). Collectively, such facilities may be referred to as Locally Unwanted Land Uses (LULUs); however, NIMBYism may also be expressed against the development of facilities generally imbued with more positive connotations, with the development of new sporting stadia, supermarkets, casinos, and religious facilities sometimes provoking local campaigns of opposition. Yet perhaps the most frequent form of NIMBY protest is against new housing development, particularly where it is perceived higher densities of residential occupation will have a detrimental effect on neighborhood amenity and put pressure on local resources (as recent studies of opposition to 'town cramming' and new house building have demonstrated).

While it is clear that local opposition to urban development is not a recent phenomena (and, indeed, can be demonstrated to have occurred at least since the nineteenth century), the term NIMBY has a more recent provenance, reportedly first used in the late 1970s to describe the opposition of US suburbanites to the prospect of having hazardous landfill sites located in the vicinity of their homes. The term spread remarkably quickly, however, and by the 1980s was routinely used in political and media discourse in Europe, not least in the context of opposition to house building in rural areas of Britain, where (mainly) Conservative voters in urban fringe areas sought to oppose the relaxation of planning controls by Environment Secretary, Nicholas Ridley, who repeatedly characterized them as self interested NIMBYs.

This particular – and pejorative – use of the term to characterize protestors as motivated by selfish protectionism has tended to overshadow its use and value as a geographical concept, and there are many who have argued that it is in fact inappropriate for geographers to use the term given that it presupposes the motives of protestors. However, NIMBYism can be regarded as an explicitly geographical concept, in a sense that it is a term which, by definition, refers to 'local' opposition to development. Indeed, in most instances of NIMBYism, opponents begin by conceding that there is a need for such development to occur, yet contest the idea that it needs to be in the vicinity of their homes and families. NIMBY opposition is thus fuelled by an understanding that certain facilities will create 'negative externalities' or nuisances which will effect neighborhoods proximate to the development but have negligible impacts on those people and homes further away.

In this sense, NIMBYism can be viewed as involving a conflict between local/private and national/public interests. Exploring the ways such conflicts are played out has thus become an important tradition within urban social geography, with associated work in economics and urban property management seeking to quantify the real and imagined impacts that development has on particular communities. Demonstrating that locally unwanted land uses have different impacts on house prices at varied distances is thus an important geographical contribution to policy debates, as is the mapping of 'externality fields' which are of different intensity, effect, and distance–decay. In **Figure 1**, adapted from the work of geographer Michael Dear, the intensity of a negative externality (e.g., a measure of the noise generated by an incinerator) indicates the relative level of nuisance experienced at a given distance from a development, while the extent of the field indicates the total area within which some

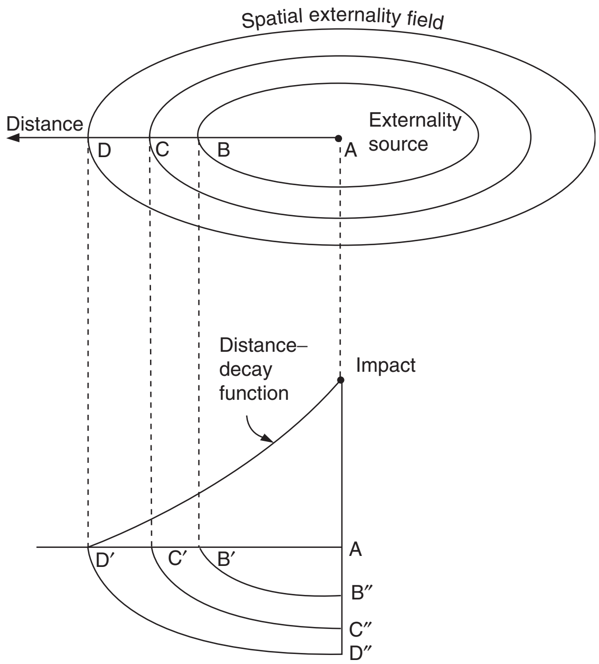

Figure 1 Dimensions of the spatial externality field.

nuisance exists. The distance–decay curve accordingly indicates how measured nuisances decline with increasing distance from their source.

Geographers' examination of locally unwanted land uses suggests these tend to be associated with overlapping externality fields whose intensity and extent varies according to the type of facility in question. For instance, a major facility like an airport may be associated with a larger externality field than is the case for a hospital or rehabilitation center; in the case of the former, noise impacts tend to generate a more extensive externality field than traffic and parking problems, which may be more localized. However, research exploring residents' perceptions and experiences of nuisance when living close to such facilities suggests that topography, predominant wind direction, and existing transport routes all conspire to ensure distance–decay rates are rarely uniform (in relation to airports, for example, it has been demonstrated that aircraft noise will impact more on communities in the take off and landing zone than on communities located at similar distances yet parallel to the main runways). Research also suggests that there are important diurnal and seasonal variations in the perceived levels of nuisance associated with many facilities so that, for example, aircraft noise might be experienced differently at night than during the day. Studies of the externality fields of football stadia likewise imply that nuisances of illegal parking, floodlight glare, littering, and noise may be very different when they are being used for pop concerts as opposed to football matches, irrespective of the number of spectators. Even so, Michael Dear (1992: 291) argues "there is one universal factor in all NIMBY conflicts: geographical proximity…the closer residents are to an unwanted facility, the more likely they are to oppose it."

Mapping externality fields thus offers a particular understanding of why residents might oppose specific types of development given it is entirely predictable that residents who perceive they will be negatively affected by externalities will act to protect their major asset (i.e., their home). In this regard, it is crucial to note that while homeowners can insure their house against fire, flood, or theft, the development of locally unwanted land uses represents an uninsurable risk. Of course, this perspective assumes that NIMBYs are owner occupiers, and suggests rentiers and leaseholders will be less motivated to join NIMBY protests because it is not in their interest to do so.

The idea that NIMBYs are motivated to protect property prices is thus a major conclusion of the literature on participation in local campaigns of opposition. As such, it is often suggested that NIMBY is best understood as an attempt by resource rich middle class homeowners to displace undesirable facilities toward transient and resource poor regions. In the context of urban land use, NIMBYism can hence be seen as one of the key 'push' factors that encourages the clustering of 'noxious' facilities in inner city areas away from wealthier, whiter census tracts. Alongside this oft noted centripetal movement, however, Stein also notes the tendency for NIMBYism to push unwanted developments toward the urban fringe, thus preventing 'efficient and sustainable' use of urban space; simultaneously, this process of displacement may create urban/rural conflicts as different political and planning interests seek to maintain a clear boundary between town and country. Hence, while often identified as a predominantly urban phenomenon, NIMBYism is also vociferously expressed in rural environments where NIMBY protestors may make great play of the potential despoilment of the countryside.

The Tactics of NIMBYism

By definition, all NIMBY campaigns are based on the argument that a particular facility, while necessary, could be better located elsewhere. However, the specific nuance of this argument varies from place to place, with campaigners citing a variety of reasons why a development is inappropriate in their locality. On occasion, protestors will emphasize the potentially deleterious impact of a development on their house prices, and emphasize the issues of blight that might be associated with living alongside a locally unwanted land use (especially when it is regarded as unsightly). More normally, however, attention is drawn to the potentially negative impacts of a development on the community's health and well being, whether these are related to the initial construction of a facility or its day to day operation. A common rhetorical strategy here is to emphasize the potential risk that proposed developments might pose to the health of children, who are frequently portrayed as 'innocent' parties caught up in a conflict not of their making. The idea that a new development might increase local traffic thus becomes depicted as a threat to the road safety of local children; aircraft noise prevents children from sleeping or studying properly; pylons and mobile phone masts are opposed because of the (unproven) allegation that they may be associated with childhood cancer and leukemia.

In relation to human service facilities (like prisons, hostels, and asylums), such arguments take a rather different inflection, with protestors typically suggesting it is the 'clientele' of the proposed facility that poses a threat. In many instances, it is the media stigmatization of such groups and communities that fuels such NIMBY urges. For instance, geographical investigation of NIMBY opposition to homes for those living with human immunodeficiency virus (HIV), needle exchange programs, homeless shelters, sex work projects, and housing for 'special needs' and disabled clients has suggested that protestors draw extensively on negative media stereotypes of such groups as a direct threat to families and property. This fear is most often expressed in relation to convicted offenders and recovering addicts but it is also expressed for the homeless and the mentally ill since their behavior is often regarded as unpredictable. The idea that such groups might participate in antisocial behavior such as loitering, begging, or drug dealing is an often cited fear. Images of uncleanliness and impurity may also be relevant in some instances, effectively stigmatizing some groups as a health risk, while a general concern that users of human service facilities are lazy and unemployable leads many to view them as a potential burden which a community will have to bear. Taking these concerns together, Takahashi and Dear thus identify a 'hierarchy of acceptance' whereby facilities for those with mental illnesses or addictions are regarded as less desirable than, for example, homes for elderly or infirm; in their survey of US suburban populations, it is actually homes for those living with HIV that are regarded as the least desirable community facility.

Prompted by fears of the impact of development on their homes and lifestyle, NIMBY protestors use a variety of means to articulate their anxieties. First, and perhaps most obviously, NIMBY opposition occurs through preexisting political channels, so that, for example, protestors make representations through processes of planning consultation, zoning hearings, or design review boards. Given most planning systems concern themselves solely with issues of environmental and residential amenity, and not social justice, NIMBY protestors are thus forced to shift the basis of conflict from questions of social or moral geography into the (apparently) asocial realms of environmental impact assessment. In practice, this means that NIMBY protestors seek to assemble credible evidence to highlight the negative externalities that will be associated with a development (noise, increased traffic volume, health risk, etc.), and lobby politicians and planners accordingly.

Another (successful) strategy employed by NIMBY protestors in this regard is to employ what Michael Dear refers to as a 'sophisticated' argument, whereby they disguise themselves as advocates for the cause but formulate their opposition from the perspective of the target clientele. For example, it might be argued by a NIMBY protestor that a homeless shelter or HIV clinic is an important and valuable facility, but that the neighborhood is not conducive to the rehabilitation of the client. A notable instance here is provided by NIMBY opposition to the development of asylum seeker accommodation in the UK: in the seaside town of Saltdean, near Brighton, opponents of a proposed center made repeated reference to the lack of suitable entertainment and leisure facilities in the area for such populations, as well as the lack of religious facilities catering for a diverse and multicultural community. Similarly, opponents of affordable housing might challenge its development with reference to the lack of local job opportunities or school facilities in the area, or point to the absence of public transport infrastructure. In such cases, NIMBY opponents seek to dismiss allegations they are acting in a parochial or selfish manner: rather, it is developers who are made to appear shortsighted by failing to locate a development in a more appropriate location.

As such, the formal processes which allow residents to make representations about proposed development tend to favor the most politically articulate, giving credence to the claim that NIMBY is a process in which resource rich residents displace unwanted land uses onto less articulate neighborhoods. This is born out by studies which note the pivotal role of gentrifiers in displacing groups and facilities from inner city areas where resident groups had previously been unable to mount credible campaigns of opposition. Nonetheless, it should be noted that NIMBY is not merely a phenomena occurring in the wealthiest suburbs, with class alliances sometimes emerging as groups coalesce around opposition to developments in deprived or struggling areas. Furthermore, those wealthy suburban dwellers who repeatedly oppose development gain a reputation among planners as serial NIMBYs, and that this ultimately undermines their credibility (and effectiveness). In contrast, forms of direct action (e.g., petitioning, pickets, sit ins, and squatting campaigns) may be more effective at halting development: while anti road building protests, squatting movements, and impromptu street parties are usually theorized by geographers as instances of resistance rather than NIMBYism, there is no doubt that they can bring together those who are opposed to development in a specific locale with 'eco rads' opposed to development in general. Indeed, studies of anti road building, anti mobile phone masts, and anti airport expansion movements stress that these protests often involve complex actor network spaces with national, as well as local dimensions: the definition of the 'local' is rarely clear cut in cases of NIMBY opposition.

Problematizing NIMBY

Though it remains prominent in media discourse, NIMBY is a term whose popularity among geographers has waned. One reason is a general uneasiness about the way the term is applied to antidevelopment protests, irrespective of the type of protestors involved. A key question here is whether NIMBYism is just that, or whether some NIMBY protestors are actually Not in Anybody's Yard (NIABY) or Build Absolutely Nothing Anywhere Near Anybody (BANANA). As Wolsink argues, the term NIMBY implies resistance to development in a protestor's own neighborhood, but assumes these same protestors would have no problems with the development of those facilities elsewhere. However, examination of NIMBY campaigns reveals that, in many instances, campaigners are not accepting of the need for development *per se*, and refuse to acknowledge that it is in the national (social) interest. For example, Wolsink has demonstrated that opposition to developments is often unrelated to proximity (challenging Michael Dear's assertion that opposition is dependent on locational proximity). In the case of rural wind farms, for example, he demonstrates that many 'NIMBY' opponents are not directly effected by the noise of wind farms, and may not even be able to see them: rather they are opposed to the idea of wind farms despoiling the countryside in general terms. Similarly, Hunter and Leyden found little evidence of NIMBYism in their studies of hazardous waste incineration. Rather than concern about local property values and aesthetics (usually regarded as characteristic NIMBY attitudes), they found that opposition related more to a lack of trust in the government's claim that this was an appropriate and safe way of disposing of waste.

In addition to the inappropriate labeling of some protests as instances of NIMBYism, Wolsink raises another concern about the language of NIMBYism being utilized by researchers. As he details, NIMBY is a term often employed by planners to emphasize the self interested and negative attitude displayed by some communities. Indeed, throughout the planning literature it is possible to find reference to NIMBYism as a major obstacle to the delivery of services and housing in a socially just and sustainable manner, with Pendall arguing that NIMBY "connotes a selfish desire to abdicate responsibility for important community facilities." For Wolsink and others, the labeling of protestors as NIMBYs precludes a sensitive understanding of their grievances and therefore prejudices their representation in planning and politics. Considering opposition to radioactive waste dumping, Kemp draws similar conclusion, noting the NIMBY concept may be "applied too readily, a convenient attribution of motive which disguises a more fundamental range of technical, environmental, and socio economic concerns…the NIMBY concept should therefore be rejected as distorting and unhelpful." In this sense, it is possible to concur with Burningham's view that when those concerned with researching, mediating, or resolving local disputes should steer clear of the language of NIMBY and instead engage with the 'diversity and complexity of 'local' concerns.

Taking NIMBY Forward

Notwithstanding the important caveats that should be applied when describing protests as NIMBY, it is clear that planning proposals continue to be hampered by protests, with certain facilities being displaced from some locales only for their negative externalities to be felt elsewhere. Given the frequency of such protests throughout the urban West, exploring the causes and consequences of local opposition remains an important task for geographers. In this respect, the fact NIMBY has principally been theorized as an instance of class conflict is problematic given the rigidities and inflexibilities which many associate with Marxist perspectives. Accordingly, the attempt to characterize NIMBY as a conflict between development capital and residents seeking to protect their commodified living place (as in the seminal work of Kevin Cox) has been rejected as simplistic in several respects. Noting this, Bob Lake's attempt to position NIMBYism as implicated in both the primary and secondary circuits of capital accumulation places NIMBY in a richer theoretical context; likewise Neil Smith's focus on revanchism and gentrification offers some useful cues for those seeking to understand NIMBY as a means of creating profitable opportunities for corporate investment at the same time it maintains the ambience of family oriented spaces.

Beyond this (long standing) focus on the mechanics and machinations of real estate markets, NIMBY studies have been enriched through geographers' more recent engagement with questions of cultural politics, and particularly the role of NIMBYism in the maintenance of Self/Other boundaries. There is hence a growing literature which explores not only the 'rational' economic basis on which homeowners oppose the development of controversial or noxious land uses, but also the instinctive reactions people have when faced with the prospect of living in the proximity of Other populations. In this sense, studies of NIMBYism have taken inspiration from more general literatures on sociospatial exclusion. Here, the work of David Sibley on sociospatial boundary making has proved highly influential. Drawing upon psychoanalytic ideas about the sorting of people and things into 'good' and 'bad' categories, Sibley has suggested the desire to create boundaries between Self and Other is inculcated in early infancy, being one of the ways the growing child seeks to situate itself in the social world and cope with anxieties about separation from the maternal. In Western societies that place considerable emphasis on order and predictability, these boundaries serve to exclude those who are depicted as disorderly, and hence as potentially polluting. Such perspectives imply people seek to defend their body, home, locality, or nation in response to the incursion of abject Others who threaten to overwhelm the boundaries of individual and collective identity. Problematically, such psychosocial perspectives also imply individuals are not able to fully articulate why they oppose particular developments or populations because exclusionary reactions are partially unconscious.

While such ideas have been mainly worked through in studies of reactions to human service facilities used by stigmatized groups, there are grounds for suggesting they are highly relevant in other contexts too. For instance, while opposition to affordable housing is often seen as being prompted by fears of declining house prices, Rose suggests those expressing NIMBY attitudes associate public housing with a failure to keep the local environment clean, which they interpret in terms of a lack of sense of responsibility on the part of the tenants. Of course, this stigmatization of the potential occupants of these developments as Other is often hidden in NIMBY discourse, and Rose insists it may only be after reeling off a series of environmental and esthetic concerns that interviewees mention what is really worrying them about a new housing development; namely, that the occupants will 'lower the tone' of the neighborhood.

Developing the case for considering fears of Otherness, Wolsink suggests anxieties about 'strangers' are not incidental even when considering opposition to infrastructure developments (such as landfill sites, phone masts, power plants, or wind farms). In each case, these developments are described using a terminology that stresses they are incongruent, with exclusionary metaphors of weeds, plagues, and bodily excretions frequently deployed to stress these developments would be 'out of place'. Such metaphors are also widespread in the oppositional discourse voiced toward many casino developments, with the potential users of such sites being identified as potentially indebted and/or criminal; likewise, debates about the 'negative secondary effects' surrounding adult businesses (sex shops, lap dancing clubs, and sex cinemas) revolve around the idea that such facilities attract criminally disposed clientele to the area. However, local authorities' attempts to define 'buffer zones' between adult businesses and residential areas imply that such fears of criminality are conjoined with more general moral anxieties about the public visibility of commercial sex, with NIMBY protests and practices of zoning combining to pushing sex work toward marginal locations.

Tacking between questions of discourse and subjectivity, such analyses have suggested the language of NIMBY serves to differentiate Self/Other, with protestors' rhetorical strategies revealing fundamental anxieties about individual and collective identities. Given discourses about color, disease, and nature feature predominantly in Western representations of Otherness, it is perhaps unsurprising such anxieties are often projected onto nonwhite groups, which are depicted as disproportionately reliant on human service facilities. Hence, within the literature on exclusionary reactions to proposed developments, increased attention has been devoted to the role of race, exposing the ways NIMBY actions often exclude people of color. Though protagonists contesting the development of human service facilities rarely see themselves as having malicious or racist intentions, Laura Pulido alleges that NIMBYism maintains white ethnic privilege by effectively constructing whiteness as the unnamed norm against which Otherness is gauged. Even when it is not characterized by hostile racism, NIMBYism may therefore be a key means by which white populations protect the benefits of their whiteness (such as the enhanced value of their homes and properties in relation to those in 'non white' areas). As such, studies of opposition to mosque building, public housing, and asylum centers flag up some important directions for geographers seeking to explore the intersection of class, race, and identity politics.

In sum, recent geographical perspectives suggest that NIMBY protests can be best understood as part of a suite of processes which serve to perpetuate geographies of social and spatial exclusion. The implication here is that instances of NIMBYism should not be treated as isolated instances of 'land use' conflict; rather, they always and inevitably involve questions of social difference and hence speak volumes about the social and social inequalities inscribed in everyday landscapes.

See also: Environmentalism; Neighborhood Change; Neighborhoods and Community; Planning, Urban; Public Good; Social Class; Social Justice, Urban; Social Movements.

Further Reading

Burningham, K. (2000). Using the language of NIMBY: A topic for research, not an activity for researchers. *Local Environment* 5(1), 55 67.

Cox, K. and McCarthy, J. (1982). Neighborhood activism as a politics of turf: A critical analysis. In Cox, K. & Johnston, R. J. (eds.) *Conflict, Politics, and the Urban Scene*. New York: St. Martin's Press.
Dear, M. (1992). Understanding and overcoming the NIMBY syndrome. *Journal of the American Planning Association* 58(3), 288 300.
Hunter, S. and Leyden, K. M. (1995). Beyond NIMBY: Explaining opposition to hazardous waste facilities. *Policy Studies Journal* 23(4), 601 619.
Lake, B. (1993). Rethinking NIMBY. *Journal of the American Planning Association* 59, 87 93.
Meyer, W. B. (1995). NIMBY then and now: Land use conflict in Worcester, Massachussetts, 1876 1900. *Professional Geographer* 47, 298 307.
Pulido, L. (2000). Rethinking environmental racism: White privilege and urban development in Southern California Annals. *Association of American Geographers* 90, 12 40.
Rose, D. (2004). Discourses and experiences of social mix in gentrifying neighbourhoods: A Montreal case study. *Canadian Journal of Urban Research* 13(2), 278 316.
Takahashi, L. (1997). The socio spatial stigmatisation of homelessness and HIV/AIDS: Towards an explanation of the NIMBY syndrome. *Social Science and Medicine* 45, 903 914.
Wilton, R. (2002). Colouring special needs: Locating whiteness in NIMBY conflicts. *Social and Cultural Geography* 3, 303 321.
Wolsink, M. (2006). Invalid theory impedes understanding: A critique on the persistence of the language of NIMBY. *Transactions, Institute of British Geographers* 31(1), 85 91.

Nongovernmental Organizations

Saraswati Raju, Jawaharlal Nehru University, New Delhi, India

Glossary

Rent-Seeking The term rent-seeking implies economic rent and refers to attempts by an individual, organization, or firm to gain money by economic and/or legal manipulation rather than by trade and production of wealth. Now, rent-seeking is more often associated with government regulation and misuse of governmental authority than with land rents.

Subaltern The exact meaning is disputed in current philosophical and critical usage. However, it commonly refers to people/perspectives/regions outside the hegemonic power structures.

Washington Consensus Term coined by John Williamson in 1989 to identify a set of ten economic policy prescriptions that would constitute a 'standard' reform package for poorer countries by institutions such as the International Monetary Fund (IMF), World Bank, and U.S. Treasury Department. The term has almost become synonymous with the reduced role of the state and expanding market forces associated with neoliberalism.

Nongovernmental organizations (NGOs) have always been present in some form across the world. However, the phenomenal increase in their numbers and the research attention paid to them in the recent past is symptomatic of several parallel processes, which include states' declining role in direct service provision in many parts of the world under the increasing significance of political and economic forces at the global level. Emergence of private players and withdrawal of the state also saw reconfiguring of NGOs' role and functioning and their interface with the state bringing to the fore several issues such as legitimacy, transparency, and accountability encapsulated in the notion of 'good governance'.

The discussion here is confined to development of NGOs on the premise that it is these NGOs that have received particular attention and brought these issues upfront. Moreover, scholars have argued that despite their locations in – arguably simplified divisions – Global North and South, NGOs now share specific discourses and practices largely because of the new managerial practices that lend a certain similarity to NGO issues in apparently very different contexts.

Definition of Nongovernmental Organization

The term 'NGO' was initially coined by the United Nations (UN) in 1945 at the San Francisco conference for use in its charter to differentiate between participation rights for intergovernmental specialized agencies and those for international private organizations. Thus, under Article 70, 'specialized agencies, established by intergovernmental agreement' could 'participate without a vote in its deliberations' while under Article 71 'NGOs' could have 'suitable arrangements for consultation'. Notably, the terms 'specialized agencies' and 'NGOs' essentially referred to international bodies within the UN context and there was no mention of maintaining cooperation with private bodies.

Prior to this, when 132 international bodies had decided to cooperate with each other in 1910 they did so under the banner of the Union of International Associations. The League of Nations officially referred to its 'liaison with private organizations', while many of these bodies at that time called themselves international institutes, international unions, or simply international organizations.

Over the years, more particularly in the late 1970s and early 1980s, the term NGO has become popular and widespread. However, the composition and the tenor of the organizations identified as NGOs have undergone so much change that not only an internationally acceptable definition is harder to come by, but there exists inconsistency in the use of the term as it carries different connotations in different contexts. In many ways, NGO has become almost an umbrella term to mean all sorts of organizations ranging from extremely influential transnational federations such as the Red Cross, Oxfam International, or World Wide Fund for Nature at one extreme to small neighborhood based citizens', church and community based groups at the other.

As there is no precise definition of 'an NGO', we cannot identify the characteristics shared by NGOs. Yet several basic attributes are generally agreed upon internationally, at least in principle, because these attributes match the conditions for the recognition of an NGO by a UN agency. That is, an NGO must be independent of direct control by any government; it must be voluntary and nonprofit making; it cannot be constituted as a political party; it cannot be a criminal group; and it has to be nonviolent. In reality, several of these conditions can blur, that is, NGOs can actually be part of a political organization explicitly or tacitly. Also, of late, 'voluntary' features of NGOs are being questioned in the wake of increasing professionalization of the sector.

Scholars have identified two major NGO roles: juridical and sociological. In juridical studies, NGOs' legal status takes precedence and the emphasis is placed on the legal status of the national and international NGOs and their implications for international laws, whereas in sociological works NGOs are seen as societal players and their composition and functioning get attention. For example, NGOs, in addition to their informal participation during all stages of the negotiation processes at global conferences, may also be part of formal official drafting committees. Amnesty International is a case in point. Greenpeace is another, in advancing international standards to protect the environment. At the Earth Summit in Rio in 1992, global NGOs played a critical role and were asked to sit on the Committee on Sustainable Development which was created. At the 1994 Cairo World Population Conference, increasing numbers of international NGOs took on the responsibility of setting the agenda for discussions, as they did at the Fourth World Conference on Women in Beijing and more recently in 2001 at the UN conference in Durban on antiracism. Despite the increasing involvement of NGOs in global politics, however, with a few exceptions, what legally characterizes NGOs remains unclear.

In the absence of a clear definition, terms such as 'private voluntary organization (PVO)', 'nonprofit organizations (NPO)', and 'voluntary associations' (VA) are often used synonymously as well as interchangeably with the term NGOs.

French scholars refer to the French equivalent 'organizations non governementales' or its abbreviation 'ONG', whereas PVOs and NGO are nearly co terminus in the United States. Although whether the two terms are equivalent or, alternatively, whether NGOs are a subset of PVOs, is a question that has not been fully explored. The situation gets further complicated, for example, in Germany the official translation of the 1945 charter text 'nicht staatliche Organisationen' is now seen as too broad a definition as it connotes other actors that are not NGOs, such as multinational companies or national liberation organizations. At times, NGOs are also referred as the 'third sector', 'voluntary sector', and also as 'civil society'. However, the ways in which each of these organizations has been defined are distinct from each other although the 'nonprofit' clause seems to be applicable across all of them. Since NGOs are essentially nonprofit organizations, they have often been seen as one of the many subsets of NPOs.

One of the much debated issues concerns the relationship between NGOs and civil society. Some suggest that the role of NGOs in countries with totalitarian regimes and/or politically instability or having fragile forms of representative government is that of a watchdog – a check, whereas in other situations they can be supportive or grow out of social movements. Thus, some argue that NGOs be posited within the framework of social movements as they are increasingly becoming far more relevant as forces for social change and they must intervene strategically in political processes so as to create pressure groups for better government and/or contribute to public debates about how government might work. Either ways, it is generally agreed that 'civil society' is a broader concept and cannot be synonymous with NGOs. Some have attributed the role of guardians of civil society to NGOs. However, scholars have been questioning such euphoric constructs of NGOs, essentially because of the subcontracting, monetary transactions, and fund flows that routinely influence NGOs' functioning.

Scholars often overcome this problem of ambivalence in defining NGOs by having their own definition in ways appropriate to their particular research agendas. NGOs themselves sometimes use different definitions. By and large, however, NGOs see themselves as offering alternative visions to dominant discourses, be it developmental paradigms, environmental issues, gender concerns, subaltern questions, and so on. By implication then, NGOs are particularly concerned with the marginalized and the poorer sections of the community and their livelihoods, governance, and political rights. However, some urban based organizations may cater to issues such as environmental ethics, conservation, and animal rights that may at times clash with the interests of the poor.

Classification of NGOs

Classifying NGOs is not merely an academic exercise as absence of any theoretical and conceptual framework impinges upon a proper understanding of the sector. Without a classification, it becomes rather difficult for those on the ground to effectively organize knowledge transfer and sharing of experiential learning across NGOs of different locations, stature, background, and expertise. Despite several attempts, the available classificatory schemes are not exhaustive, exclusive, or foolproof.

Composition – By definition, NGOs are outside the government ambit (third sector/voluntary sector) and self governing (autonomously managed), but in the contemporary environment of public–private participatory development, the relationship that is emerging between NGOs and the state has given rise to government organized NGOs (GONGOs) in poorer countries; quasi nongovernmental organizations (QUANGOS) largely based in the industrialized countries; and DONGOs (donor organized NGOs).

The preference of donors for NGOs in the 1980s and much of the 1990s led to a great increase in the numbers and size of NGOs. Because of NGOs' increasing pres ence, there have been considerable official funding flows from developed countries to developing countries. Some scholars distinguish these public service contractors (PSCs) from voluntary organizations.

Geographical location – NGOs have been classified on the basis of their geographical location – between northern NGOs (NNGOs) based in the industrial dem ocracies and southern NGOs (SNGOs) based in poorer countries – the idea being that these serve different constituencies. But this somewhat restricted categoriza tion has been recently questioned as wide ranging or ganizations have overlapping concerns across globe. More importantly, communication/information gateways have made it possible, at least for some NGOs, to build internationally dispersed coalitions. The 'Third World Network' of NGOs with its base in Malaysia and the International Federation of Human Rights in Paris with 89 human rights groups in 70 countries are prototypes of what has been called 'intermediary NGOs' acting as 'facilitators' for various purposes. Some such NGOs have acquired the status of global NGOs.

Size and organizational structure – measured in terms of number of staff, budgetary allocation, areal/population coverage – have been the more obvious differentiating characteristics in NGO classification. NGOs vary in their structures as there can be global hierarchies – a centrally placed key organization with transnational reach through local offices, or they can have independently functioning bodies under loosely defined federal arrange ments; alternatively, NGOs can be 'stand alone' – grassroots organizations. 'Autonomy' from funding/donor agencies – national or international – is another criterion. NGOs have been further classified on the basis of sour cing of their staffers – membership organizations, that is, from local community and nonmember organizations staffed by professionals from different social and ethnic backgrounds. NGOs can also be divided into service delivery, advocacy, and training, but once again such a neat classification is problematic as the roles can overlap and change.

Strategy and evolutionary trajectory – NGOs have also been designated on the basis of the strategies adopted and nature of their work. Thus, NGOs adopting the most basic first generation strategies are involved in direct delivery of relief and welfare services often in response to disasters or crises. Second generation strat egies prompt NGOs to engage in capacity building of people for self reliant local action. NGOs that follow third generation strategies reach beyond individual communities and address institutional and policy con cerns. In addition to these three generations of strategies, there is a fourth category which aims at broader social vision. These strategies can coexist and also mark the gradual evolution of NGOs from one stage to another.

Table 1 Alphabet soup: A symptom of classification confusion

BINGOs	Big international nongovernmental organizations
CBOs	Community-based organizations
CB-NGOs	Community-based nongovernmental organizations
DOs	Development organizations
DONGOs	Donor nongovernmental organizations
GONGOs	Government nongovernmental organizations
GROs	Grassroots organizations
GRSOs	Grassroots support organizations
IDCIs	International development cooperation institutions
INGOs	International nongovernmental organizations
NGDOs	Nongovernmental development organizations
NNGOs	Northern nongovernmental organizations
POs	People's organizations
PSCs	Public service contractors
QUANGOs	Quasi-nongovernmental organizations
SCOs	Social change organizations
SNGOs	Support nongovernmental organizations
WCOs	Welfare church organizations

Reproduced from Vakil A. C. (1997). Confronting the classification problem: Toward a taxonomy of NGOs. *World Development* 25(12), 2057 2070.

Gender concern – NGOs have also been classified in terms of work and policies in keeping with feminist concerns. Seven types of NGOs have been recognized on the basis of their inputs in assisting women achieve empowerment: (1) efforts initiated from outside, (2) small scale grassroots, (3) worker based, (4) via affiliation with a political party, (5) through provision of services, (6) research, and (7) coalitions. Out of these, outside initiated women's NGOs are seen as the weakest of the organizational types because of their top down approach, whereas grassroots NGOs are geared toward addressing feminist concerns most explicitly.

In sum, it can be concluded that a plethora of or ganizations can be classified as NGOs (**Table 1**) re flecting the definitional inadequacy of NGOs in the first place.

The Emergence of NGOs

The sudden growth and progressive presence of the NGO sector is strongly related to shifts in mainstream development ideologies and institutions. In contrast to an earlier emphasis on altruistic activities by passionate volunteers, NGOs' role changed significantly in the era of neoliberal development ideologies – embodied in structural adjustment programs (SAPs). SAPs were based on the opinion that states were often rent seeking, slow, and inefficient, acting as stumbling blocks to the market led development model put forward by the IMF and World Bank. This drew upon the logic of the intrinsic superiority of economic liberalism on the one hand and equating economic growth with enhanced personal economic welfare on the other. The protective regimes of the state were posited as slowing down development. The policies of the East Asian newly industrializing econ omies were articulated in such a manner that their rise was attributed to the requisite liberal norms coupled with the macroeconomic fundamentals.

The culmination of these processes may be seen as epitomized in the Washington Consensus imposed on particularly debt ridden poorer countries of the South by international financial and lending institutions in the 1980s, which in most general terms involved a commit ment to free markets, private property, and individual incentives as a development strategy with a circum scribed role for government.

By the 1990s, the euphoria that had accompanied the Washington Consensus had started dying down as free trade and unregulated markets did not deliver the promised benefits of market led growth, stabilization of financial markets, and political order. Markets privileged only a select few who could access the emerging op portunities, but rising prices and privatized social goods also meant that millions lost their access to livelihoods and regular earnings. Several studies pointed out that even as incomes were rising, disparities across regions, communities, and gender were increasing leading to general discontent.

A series of events followed: Mexico was hit by financial devastation in 1994; Thailand devalued its currency followed by the rest of East Asia, Brazil, and Russia in 1997 and 1998, and the world witnessed the so called 'Asian crises'. A number of critiques and events led to a reappraisal of the strict neoliberal orthodoxy that dominated mainstream development thinking and insti tutions over the 1980s and early 1990s and saw the World Bank and IMF revisiting the Washington Consensus to reconfigure the role of state vis à vis earlier free market ideology and unrestricted corporate control.

What came subsequently to be known as the post Washington Consensus thus talked about the state's mediating role through managerial and governance stra tegies whereby the state is replaced not with the market, but with a strong civil society represented by NGOs. In such a policy environment, as some scholars have put it, the state's role was reduced to steering rather than rowing. Along with NGOs' expanded roles, in particular in the developing world, the emerging vocabulary of public–private partnership for good governance meant that citizens' involvement in governance has to be en hanced and they also have to look after their own well being. These goals were to be attained through mutual trust, networking, and building up of social capital and NGOs were to be crucial players in cultivating and representing social capital and in improving public ser vices more generally.

The state started rolling back to gradually withdraw from the social sectors such as health, nutrition, edu cation, and water supply, and privatization of these sec tors in many parts of the developing world started taking place. In this context, NGOs were increasingly being seen as a cost effective service delivery mechanism in areas as wide ranging as disaster management, health and education, environment, political and human and gender rights, agricultural extension, and even banking and microcredit programs, to name a few. In fact, the vast literature on NGOs is privy to the potentials this sector has been entrusted with – from alleviating rural poverty and helping communities adapt to modernization and building vibrant civil societies at one end of the spectrum to affecting the state, international politics, and the for mation of an international civil society through NGO coalitions and networks at the other. In a nutshell, NGOs were entrusted with tasks that were conventionally the responsibility of state.

Although the post Washington Consensus redrew some contours for a more proactive state, according to some scholars they were more rhetorical and the post Washington Consensus continues to retain significant elements of the earlier neoliberal agenda in which the role of NGO sector gets further strengthened.

It is not to imply that NGOs had a small role before the late 1990s or prior to the Asian crises. However, scholars are of the opinion that their presence in the early 1980s, particularly within the mainstream devel opment ideologies and institutions concerned with long term outcomes, were sparingly, if ever, noted. At best, their roles were seen as confined to welfare and disaster management. Also, southern NGOs were nowhere on the scene in the official documents. This began to change rapidly, as the various World Development Reports suggest, over the 1980s and NGOs were often cited as examples of good practice, cost effectiveness, and in novation. By the 1990s NGOs ranging from service de livery to training and advocacy as well as transnational networks and alliances were acknowledged as significant players in development processes. As the following sec tion shows, this shift meant that the tenor and the functioning of NGOs would change significantly.

NGO and State

Since the presence of NGOs has become almost inevit able in the current international climate, the discourses around this sector are replete with yet another complex issue – state–NGO relationship. Scholars have critiqued the involvement of NGOs and their positioning as partners in crucial sectors of development as a tacit way of admitting that the state is no longer capable of playing its role. Some even suggest that the state's so called partnership with NGOs and by implication common citizens is a ploy to divert attention from real problems and uneven appropriation of resources by a select few.

In general, this relationship is fraught with tension – both sides viewing the other with mistrust, hostility, and often condescension furthered by the lack of procedural mechanisms for accreditation either by the state or NGOs and also probity and efficiency in state dealings with NGOs. On one hand, there are issues of NGOs getting co opted into state agendas and becoming mere extension arms in the delivery of services and programs and losing their critical edge. On the other, there is the danger of NGOs and states assuming competing roles.

Although the diversity of the NGO sector makes it rather difficult to be critical or supportive of all such organizations, there exist dissenting voices about NGOs. These critics argue that NGOs create alternative dependencies instead of self sufficiency among people with whom they work. Moreover, the almost mandatory nature of state–NGO partnerships in many developing countries and the resulting flows of national and international funds have led to a mushrooming of NGOs and their business like transformation into professional organizations.

The state–NGO partnership has brought forth the issue of legitimacy, more particularly in the contemporary context of development aid through which NGOs have become preferred vehicles for assistance. Legitimacy can be of two types: formal–procedural and substantive–purposive. Formal–procedural legitimacy is obtained by the state by virtue of popular consent of the governed (withholding the manner in which such 'popular consent' may have been acquired) to command and be obeyed. Substantive–purposive legitimacy is concerned with the actual outcomes of state authority, that is, achieving distributive justice or the welfare of its citizens beyond its formal–procedural presence.

As compared to the state, NGOs are not democratic in nature and cannot claim to be representative voices of the people except under authoritarian governments as their role as participants in democratic debate does not depend upon any claim to representative legitimacy. They may work with the state either as partners or as watchdogs for general welfare purposes, but these activities may have vested interests to protect. In principle, NGOs are non hierarchical and function through networks that are horizontal in nature. They run in informal and personal ways presumably in contrast with highly stratified, vertically structured state. However, NGOs may easily conceal internal power structures and control mechanisms. Research shows that in many developing countries, communal and parochial considerations decide the network exchanges and bargains. This raises normative questions concerning substantive–purposive criteria of legitimacy.

These concerns have foregrounded issues such as accountability, transparency, and effectiveness of the NGO sector. To whom should NGOs be accountable is a potentially vexed issue as NGOs may be precariously sandwiched between the (northern) donors and (southern) partners facing multiple accountabilities both to those placed downward – their partners and constituencies, and upward – their trustees, donors, and host governments. It has been suggested that weak accountability on the part of NGOs, in many instances, stems from the failure to prioritize such multiple accountabilities.

Transparency can include NGOs' openness to (1) share their decision making processes, (2) readiness in allowing peer and host/donor evaluation, and (3) honesty in revealing their real objectives rather than obscuring them. At times NGOs can camouflage their functioning in order to be more compliant with state priorities, or because of internally sensitive issues that are externally pushed by donors and therefore implicate access to funding, and/or real or perceived threats to state/national security. Funding and its disbursement and utilization are some of the crucial concerns in these matters.

Still states' relative failure in terms of reaching out to marginalized and poor sections of the population as compared to NGOs put the latter in an advantageous position, although much recent literature questions NGO effectiveness as agents of development as exaggerated or largely assumed rather than demonstrated. One of the reasons for such skepticism is because of the difficulties in assessing NGOs' effectiveness including sifting the contribution of NGO led input to a particular outcome from other factors external to NGOs' work. There are no standardized measures or tools for process or outcome evaluation. The commonest concern relates to that of scaling up and replicability. NGOs typically work with small groups in an intense, innovative, and responsive manner that is context specific. It has often been argued that their human and financial investments when posited against their efforts are not cost effective or viable, particularly in large scale developmental activities.

Conclusions

NGOs have increasingly become important stakeholders along with the state in development processes. However, these institutions have strengths and weaknesses – the state has the legitimacy and authority but by and large its actions are beyond public scrutiny. Despite several limitations NGOs can work as watchdogs and pressure groups to check the state, as well as effective providers complementing states suggesting that in the contemporary world both will have to work together in collaborative partnerships.

Given these ground realities and the inevitability of NGO–state partnerships, scholars argue that NGOs will necessarily have to plug into the larger political and socioeconomic processes in which the state continues to play a crucial role. Increasing the impact of even small NGO projects has nothing to do with scale as such, but rather with strategy bridging to larger structures.

See also: Aid; Civil Society; Empowerment; Governance, Good; Neoliberalism and Development; Participation; Rio Summit; Structural Adjustment.

Further Reading

Ahmed, S. and Potter, D. (2006). *NGOs in International Politics*. Bloomfield, CT: Kumarian Press.

Alikhan, F., Kyei, P. and Mawdsley, E. (2006). *NGOs and the State in the Twenty First Century Ghana and India*. Oxford, UK: INTRAC.

Bebbington, A. J. (1997). The role of NGOs: Charity and empowerment. *Annals of the American Academy of Political and Social Science* 554, 117 135.

Bebbington, A. J., Hickey, S. and Mitlin, D. C. (eds.) (2008). *Can NGOs Make a Difference? The Challenge of Development Alternatives*. Zed Books: London.

Edwards, M. and Hulme, D. (eds.) (1995). *Non Governmental Organizations Performance and Accountability: Beyond the Magic Bullet*. London: Earthscan.

Edwards, M. and Hulme, D. (1996). Too close for comfort? The impact of official aid on nongovernmental organizations. *World Development* 24, 961 973.

Gore, C. (2000). The rise and fall of the Washington Consensus as a paradigm for developing countries. *World Development* 28, 789 804.

Korten, D. C. (1987). Third generation NGO strategies: A key to people centred development. *World Development* 15(supplement), 145 159.

Korten, D. C. (1990). *Getting to the 21st Century: Voluntary Action and the Global Agenda*. West Hartford: Kumarian Press.

Laurie, N. and Liz Bondi (eds.) (2005). *Working the Spaces of Neoliberalism: Activism, Professionalization, and Incorporation*. Oxford: Blackwell.

Rahman, S. (2006). Development, democracy, and the NGO sector: Theory and evidence from Bangladesh. *Journal of Developing Societies* 22, 451 473.

Salamon, L. M. and Anheier, H. K. (eds.) (1997). *Defining the Nonprofit Sector*. Manchester: Manchester University Press.

Townsend, J. G. and Townsend, A. R. (2004). Accountability, motivation, and practice: NGOs North and South. *Social & Cultural Geography* 5, 271 284.

Vakil, A. C. (1997). Confronting the classification problem: Toward a taxonomy of NGOs. *World Development* 25(12), 2057 2070.

Relevant Websites

http://www.fidh.org/
International Federation of Human Rights.

http://www.jhu.edu/
Johns Hopkins Center for Civil Society Studies.

http://www.aidharmonisation.com
OECD 2005 Paris Declaration on Aid Effectiveness.

http://www.oneworld.net
One World Net.

http://www.twnside.org.sg
Third World Network.

http://www.worldbank.org/
World Bank Civil Society Pages.

Non-Representational Theory/Non-Representational Geographies

L. Cadman, University of Glasgow, Glasgow, UK

Glossary

Affect The pre-personal capacity for bodies to be affected (by other bodies) and, in turn, affect (other bodies). This capacity for affecting and being affected subsequently defines what a body is and can do.

Everyday Life The setting for the routine and mundane, but also improvised and transformative practices.

Immanence A concept which seeks to overcome all divisions, dualisms, and causal principles (associated with transcendence) by remaining committed to the virtual flow of life itself.

Nonrepresentational Thought A mode of thinking which seeks to immerse itself in everyday practice.

Performativity The processual and transformative nature of practice.

Practice Competences and (embodied) dispositions which precede and exceed contemplative thought and reflection.

Introduction

Geography and Representational Modes of Thought

Nonrepresentational geographies attend to both life and thought as practiced and, for this reason, 'in process' and 'open ended'. The genesis of the term nonrepresentational theory began in the early to mid 1990s through a series of books and articles written by the geographer, Nigel Thrift, although its philosophical heritage stems back much further. Thrift sought to challenge the dominant mode of representational thinking throughout human geography, in particular within cultural geography. The cultural turn, in particular, was deemed exemplary of the representational problematic in two senses. First, although it focused on everyday practices such as consumption, it tended to retreat from practice into the (cultural) politics of representation; creating deadening effects on an otherwise active world. Second, and consequently, by retaining contemplative and interpretative models of geographical thought and inquiry, much of the nonintentional, nondiscursive, and elusory nature of the everyday world was occluded from view. Contrary to this, Thrift sought to alert geographers to the embodied and performative nature of practice, much of which subsists prior to reflexive or cognitive thought. He equally sought to overcome epistemological models of geographical inquiry which maintain dualisms between theory and practice and thought and action.

Despite its title falling within the auspices of theory, nonrepresentational theory cannot be considered an epistemological approach, nor a concomitant social or cultural explanation; it equally resists solidification into a focus of enquiry which might be added to the canon of geographical thought (such as globalization or consumption, for example). At its most bold, nonrepresentational theory aims to overturn the very constitution of geographical knowledge production. Rather than creating an alternative systemic epistemological and ontological framework, however (which would reinstall the problem of representational thinking), it does this through a number of tenets which seek to engage and present (rather than represent) the undisclosed and sometimes undisclosable nature of everyday practice.

If nonrepresentational theory is to have a principle, then it is to configure geographical thought in the same way that it configures life: as a series of infinite 'ands' which add to the world rather than extract stable representations from it. It is for this reason that nonrepresentational geographies have claimed to be both ethical and political.

Philosophical Antecedents

The philosophical antecedents of nonrepresentational theory have been outlined by Thrift and are shown in **Figure 1**. It is notable that, in general, geographers have tended to focus on the continental philosophy tradition: in particular, Wittgenstein, Heidegger, and Merleau Ponty, through to Deleuze (including his collaborative writings with Guattari), and his influence on the work of Latour, Massumi, and Serres. While somewhat counterintuitive to the nonrepresentational style of thinking, it is possible to outline three dominant philosophical approaches which have influenced nonrepresentational geographies over the past few years. The first is broadly phenomenological (and includes philosophers such as Heidegger, Merleau Ponty, and Wittgenstein). Heidegger's phenomenological dwelling perspective provides a useful starting point for nonrepresentational geographies. Rather than identifying a ground for thought (in our minds, or in the world), Heidegger proposes that we are always already thrown into the world and inseparable from it. Our immersive practices of being in the world are, in themselves, disclosive and we must avoid turning to subjective or objective reasoning (or representational thought) to account

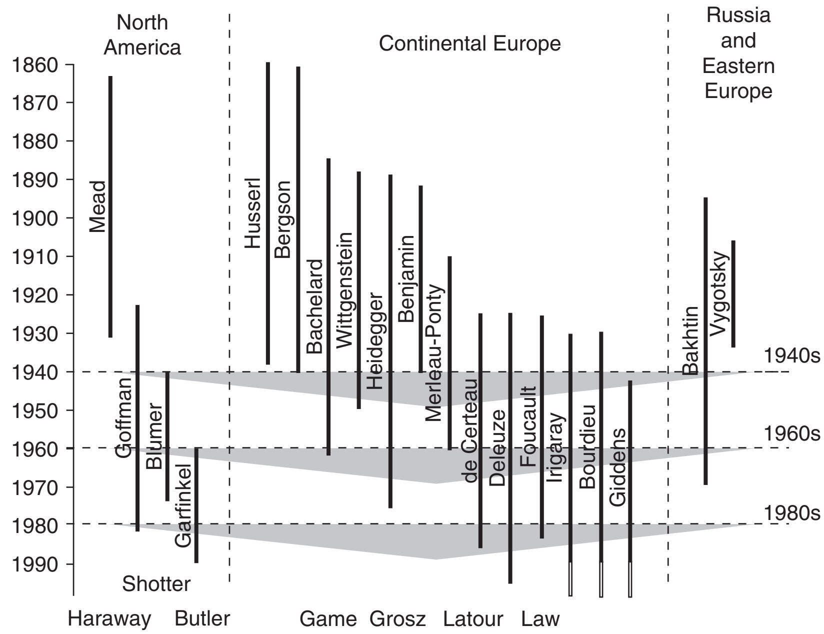

Figure 1 The life-time-lines of nonrepresentational theory. Courtesy of Prof. Doreen Massey.

for them. Merleau Ponty shifts things slightly through the notion of the 'lived body'. Here the disclosive nature of being in the world is available only through the body and our bodily competences. It is then down to the later Wittgenstein to account for the nonsystematic (it cannot be known in advance) and performative (its rules are only given in action) play of embodied practice. Taken together these authors have had a profound influence on non representational ways of thinking in the social sciences generally, from Bourdieu's 'habitus', and de Certeau's 'tactics', through to John Shotter's notion of 'know how'.

The second philosophical approach is neovitalist (and includes philosophers such as Spinoza, Nietzsche, Bergson, and their reworking through the collaborations of Deleuze and Guattari). The philosophy of Deleuze and Guattari in particular has offered tools for those geographers who want to escape phenomenology's largely human centered understanding of (embodied) practice and connect with the impersonal and transversal forces of the world. Like phe nomenology, Deleuze and Guattari engage with a register that preexists distinctions between subjects and objects, and things and appearances, but their philosophy does not begin or end with human experience or perception. Rather, they seek to overcome all perspectivism by engaging with a dynamic plane of immanence in which there are no dis tinctions between what things are and what they do. Instead of affections or perceptions residing in people or objects there are pre personal and continually differentiating af fects and percepts. It is for this reason that their philosophy connects nonrepresentational geographies with recent posthumanist and more than human geographies. Finally, a third philosophical antecedent ebbs toward post structur alism (bearing in mind that some of the most notorious post structuralist thinkers – such as Lyotard, Baudrillard, and Derrida – are absent from Thrift's original diagram). Post structuralism was arguably the theoretical flavor of the month for geography during the late 1980s and into the 1990s and, although somewhat sidelined by early non representational geographies, has been recently revisited.

Two elements of this (re)engagement stand out. First, and partly as a response to the perceived antirepresenta tional stance of nonrepresentational theory, post structural philosophers (such as Baudrillard) have been engaged to offer alternative or complementary critiques of presentist and representational thinking. Second, and partly through the influence of nonrepresentational theory, post structur alist philosophers (such as Derrida) have been repositioned in geography beyond the realm of texts, signification, and representation and into that of ethics, materiality, and force relations.

As a whole, the philosophical antecedents to non representational theory offer a critique of the bifurcation between perceivers and worlds found in post Cartesian models of idealism (belief that access to the world is provided through forms given in our minds) and realism (belief in an objective world which we should seek to align with our perceptions); that is, ways of thinking which have persisted into current representational modes of thought.

Interlocutions

Nonrepresentational theory has emerged amidst wider theoretical shifts in the humanities and social sciences generally. Three areas are worthy of specific attention.

Actor-Network Theory

Emerging in the 1980s, actor network theory (ANT) instigated its own critique of representational thought by focusing on the heterogeneous practices of association, enrolment, and translation, between humans and nonhumans, which together engineer worlds. More recently, ANT has introduced notions of alterity (otherness) to engage the performativity of these practices and overcome earlier somewhat flattening actor network accounts. The use of alterity here shares some similarities with nonrepresentational understandings of virtuality. Like nonrepresentational theory, ANT prioritizes mobile practices and shares an appreciation for the complexity of the social world; it equally aims to resist becoming pigeonholed into a form of social theory. While nonrepresentational theory draws much from ANT, there are also some important differences. In particular, as part of its intent to overcome the residual idealism of representational thought, ANT instigates a methodological symmetry between humans and nonhumans. Contrariwise, although taking an interest in materiality and the more than human, nonrepresentational geographies have tended to emphasize the expressive practices of human body subjects as most indicative of the flow and disruption of everyday life (a facet which has been disputed by proponents of ANT).

Performance Studies

The interdisciplinary arena of performance studies subsists at a productive interface between the performing arts, the use of performance in everyday life (from dance and multimedia technologies through to political rallies), and academic notions of performance and performativity. Nonrepresentational geographies have drawn much from the potentialities of this interface. For instance, the performing arts – in theater but also in forms of street art – provide a means for nonrepresentational geographers to witness and expand the realm of now time and conjure up its potential to create new forms of life. Here the very performativity of (scripted) artistic performance manifests the groundless uncertainty and hence potentiality of the present. Equally, through examples drawn from experimental and revolutionary theater (such as the performative social therapies of Fred Newman and Lois Holzman), nonrepresentational geographies are able to advance experimental ways of knowing that do not prescribe outcomes in advance.

The Body and Emotions

Since the 1980s the body has become a well established locus for study in the social sciences. Authors have sought to critique a prevailing Cartesian intellectualism which creates a dualism between mind (conscious intellect) and body (matter) and subsequently prioritizes the former. Calls to attend to the body have not escaped the geographical imagination: feminist, queer, and health geographies have all sought to explain how geographical knowledge is largely disembodied, and have repositioned the body as an important site of power knowledge in contemporary life. Further, and more recently, an 'emotional turn' in the social sciences has sought to overcome any residual Cartesianism found in rationalist ways of understanding human bodies – a facet which has given rise to an emerging subdiscipline of emotional geographies. With respect to the body, phenomenological approaches in nonrepresentational theory diverge from the construction or representation of bodies toward the very 'being' of bodies. Further, neovitalist approaches have shifted the focus from 'what is the body?' toward asking 'what can a body do?' (in its human and inhuman forms); and have done so by engaging the practices and technologies of bodily 'becomings'. These differences also permeate nonrepresentational geographies' preference for the term affect rather than emotion (although the two are sometimes used interchangeably). While neither nonrepresentational geographies nor emotional geographies seek to locate emotions in atomistic, or privatized individuals, and instead consider them as produced through relations, emotional geographies have tended to focus on personal narratives of human emotions (such as depression, fear, or love) whereas nonrepresentational geographies have positioned human and inhuman capacities for affecting and being affected as prior to any namable emotional states. This dualism between named emotions and impersonal affects is not a strict one however, and recent emotional geographies, particularly those influenced by the psychoanalytic tradition in geography, are engaging theories of practice (such as those found in the therapeutic encounter, for example) to reconfigure relations between affect and emotion.

Main Tenets

Practice

Nonrepresentational theory has also been referred to as a theory of mobile practices, and it is perhaps its attention to practice which offers the main challenge to representational modes of thought; notably, because if practice is primary then we cannot point to a world amenable to mental representation beyond those practices constitutive of it. Practice is not a new focus for geography: attempts to map everyday practices are evident in the time geography of Torsten Hägerstrand during the mid 1970s, and an intention to engage practice – understood as a primordial structure of the lifeworld which precedes scientific enquiry – preoccupied earlier phenomenological perspectives in humanistic geography, such as that of John Pickles in the 1980s. Further, an attention to practice can be allied to the general shift in social and cultural geography during the 1990s to take everyday mundane activities seriously.

One such recent example can be found in ethnomethodological approaches, which seek to engage with everyday 'talk in action'. The importance of practice has been taken, by nonrepresentational geographies, into five different directions. First, nonrepresentational geographies position non or precognitive practices as primary (and unlike ethnomethodology pay little attention to conversation analysis). Unlike earlier phenomenological accounts, however, they are not interested in eliciting a primordial reduction but in understanding practice through its historical (following Bourdieu) and spatial (following de Certeau) specificity. Second, practice is not conceived as the property of individuals with prior intentions but as dialogical and processual. This means that practices are responsive and often entail unpredictable, or unintentional, outcomes. Third, due to their processual nature, practices as such elude explanation. On the one hand, this means that nonrepresentational geographies refuse to reduce practice to a higher order interpretation; practice itself is much more attuned to the rhythm and flow of everyday life than any static understanding can offer. On the other hand, practice as a way of doing (or becoming) amidst the flow of everyday life will inevitably entail gaps in meaning (such that we may reflect on what we do, but we do not know what 'what we do' does). Practice is therefore closely tied to notions of perfomativity. Fourth, nonrepresentational geographies are increasingly recognizing that the expressive and experimental potential of practices, particularly body practices, are becoming increasingly valorized in contemporary life (e.g., through yoga and numerous behavioral therapies). Finally, an attention to practice not only urges geographers to attend to everyday life in its doing (and making) but it also pushes geographers to reflect on the practice of geographical knowledge production, in particular, on the status of theory.

Nonrepresentational theory as a whole displaces the term theory from its explanatory role as epistemology and pushes it toward a much more modest supplement to practice. While nonrepresentational approaches are notoriously intellectually rigorous, the use of theory is not sought to explain or represent but to provide a toolkit to engage and expand the world.

Everyday Life

Nonrepresentational geographies are concerned with the practices of everyday life: a popular term which has been adopted in three related senses. First, authors have focused on what is most commonly understood as everyday life – the mundane, oft routinized, humdrum of everyday living (such as listening to music, dancing, gardening, walking, and shopping). These activities manifest as habits which allow us to cope and go on in the world. Influenced by the likes of Bourdieu and de Certeau, nonrepresentational geographies have sought to grasp these taken for granted background practices as embodied dispositions. Where this work differs is that practices are not inevitably linked to symbolic orderings (such as taste in Bourdieu) nor is the potentiality of everyday life necessarily connected to tactical resistance amidst wider cultural forces (as with de Certeau). Indeed, once allied with a renewed interest in performativity, the taking place of everyday life instigates the routine and mundane but also improvization, play and, inevitably, change. This links with a second usage of everyday life in nonrepresentational geographies; one which seeks to counter the assumption that it is necessarily profane and ordinary. By thinking through the virtual realm of memory, the sacred, and, in its most concrete formations, bodily practices such as meditation or dance, nonrepresentational theory seeks to contribute to a (re)enchantment of everyday life (which runs counter to Weber's secular and routinized notion of disenchantment). Third, the spatiality of everyday life in nonrepresentational geographies is not simply personal, individual, or local. Taking cues from Lefebvre's notion of 'everydayness' through to Deleuze's 'virtualities', nonrepresentational geographies have sought to engage with the very life of everyday life; that is, a transversal force, or an excess, which constitutes the everyday rhythms of, for example, world cities.

Performance and Performativity

Attention to practice and everyday life is closely tied to notions of performance and performativity: terms which have had varied uses in geography since the 1990s. Whereas early work in geography tended to utilize notions of performance to describe the scripted routines of subjects in space, later work turned to Butler's usage of performativity to discuss the repetition of discursive scripts which precede and bring forth (gendered, sexed, and racialized) subjectivities. While nonrepresentational geographies have attended to the increasing use of performative knowledges in everyday life, methodologically they share most with the second perspective, in particular Butler's attention to the iterability of performance. However, they tend to question her emphasis on signs and signification to account for subjectification. Subjectification for nonrepresentational theory is not tied to relations between subjects and discourse, rather it is just one possible outcome in a series of irretrievable (i.e., they cannot be contained) and indeterminate (i.e., they cannot be wholly known) events. This understanding of performativity can be linked to two further facets of nonrepresentational geographies. First, performativity is constitutive of the ongoing nature of practice. Performativity is not an act in time, rather it is the spacing which allows the next moment; it enables the unexpected and transformative but also the mundane ability to simply go on. This sense of performativity allows nonrepresentational geographies to articulate embodied practice and yet retain its inherent openness amid the flow of the world. Second, once the world is considered as productive and processual, then, for academics to become worthy of the eventful nondiscursive world, it demands a refigured academic style; one which recognizes that academic praxis itself is performative. Nonrepresentational geographies tend to position academic research as partial, not in the vein of our situatedness, but in the very encounter or call to attend to the world. To be faithful to this encounter, nonrepresentational geographies do not aim to resemble it through academic representations, but to experiment with the thinking which occurs on the interstice between thought and practice.

Embodiment and the Body

There are, broadly, three ways that understandings of embodiment and the body have been engaged with in nonrepresentational geographies. These can be conceived as: sensuous and expressive; historical and invested; and capacitous and affective. First, drawing from Heidegger's attention to the disclosive capacities of being and, especially, Merleau Ponty's notion that the lived body is our vehicle for being in the world, phenomenological approaches to nonrepresentational theory have focused on the noncognitive and expressive nature of the human body subject. This work singles out the particular potentiality and capacity of human embodiment (as opposed to animals and other materialities) but has nevertheless sought to dismiss any residual humanism or intentionality by emphasizing that embodiment is infolded through joint body practices with other beings and other objects.

Furthermore, when combined with notions of performativity, being is bypassed in preference for the openness of becoming. Second, more historically attuned nonrepresentational geographies have sought to highlight that investment in the capacities of human bodies is part of an ongoing project of biopower which has shifted from Foucault's original formation – as the investment and management of the vital characteristics of human populations – toward the vital characteristics of life itself. Here bodies are not disciplined subjects but molecularized virtualities. Third, and sharing tenets with the second, nonrepresentational geographies have turned to the notion of affect as a way of expressing the force and capacities of (human) bodies. Although affect as a term traverses psychoanalysis, psychology, and psychiatry, the philosophy of Deleuze and Guattari gives nonrepresentational geographies an understanding of the body as defined by the affects of which it is capable. Through the notion of affect, bodily dispositions such as joy, boredom, and despair are wrenched from cognitive, atomistic, or strictly emotional understandings and into ethological encounters. Bodies here are not organs or functions (as in the causal analyses of etiology) but modes of speeds and slowness, and capacities for affecting and being affected, which never reside in things as such.

Virtuality and Multiplicity (Time and Space)

Nonrepresentational geographies are concerned with everyday practices, yet they seek to avoid any recuperation of these into strictly empiricist or modernist understandings – such as those accounts that conceive the goings on of the social world through a fixed viewpoint with the intention of describing pure movement. Following earlier calls for relational understandings of time and space, nonrepresentational geographies acknowledge the coexistence of a multiplicity of time–spaces. Where they make a distinct contribution to these ideas is that they engage the effective groundlessness of this multiplicity, in particular through adopting a quasi Deleuzian understanding of virtuality. The concept of virtuality does not entail a real world in which there are (virtual) possibilities, nor does it configure the possibilities which might manifest in a real world (in which case the real resembles the possible and does not consider the event of change). Virtualities are different from possibilities in that they are in every sense real but not always actualized. They do not settle in beings or things and can be described as the continuously differentiating relations between forces prior to any actualization. In one sense they are like a variable push or an outside that envelopes the forces which make worlds; and the combination of forces always changes. This notion of the virtual points to an intensive and formless multiplicity (through the groundless depth of virtuality) rather than simply an extensive multiplicity (through the depthless and timeless extension of relations), which is found in some network and relational accounts of space.

Implications

Nonrepresentational theory is both a methodology and a (political) practice (or, a practical poetics), inasmuch as it seeks to reconfigure both what it means to do research and to be political.

Research Methods

Nonrepresentational thinking tends toward an academic style which seeks to describe and present rather than diagnose and represent. Its ethos of generosity castes a shadow on much of the methodological toolkit available for geographical fieldwork (in particular: in depth inter views, focus groups, and participant observation which each accentuate contemplative and interpretive modes of thought). Still, there is no method as such for non representational research, nor has there been much sus tained engagement as to how nonrepresentational theory might reconfigure the collection of fieldwork. Indeed, it is questionable how far nonrepresentational geographies can really shift the inherent perspectivism and epi stemological commitment of field data collection tech niques. Nevertheless, nonrepresentational geographies do seek to harness and experiment with the mainstay of qualitative research methods (notably ethnography and participant observation but also diaries and in depth interviewing), developing approaches such as 'observant participation' and 'performative ethnography'. Further, the drive to acknowledge the research process itself as performative ranges from philosophical speculation on the particular call to witness the geographical world, through to, in its practical incarnation, an acknowledg ment of the co fabrication of the research encounter that inevitably creates something new (a facet shared with recent posthumanist and more than human geographies). Finally, the presentation of nonrepresentational research has adopted performative and experimental writing styles which seek to present either something of the ephemeral nature of everyday practice or the potential of per formative writing itself. The use of montage (which juxtaposes different research methods that inhabit dif ferent time spaces), for example, has been sought to address the virtual multiplicity of the nonrepresenta tional world.

Politics and Ethics

Nonrepresentational geographies have engaged with three senses of the political. First, they have positioned affect as central to individual and collective political dispositions, and as an important trope for contemporary forms of governing. Approaches here range from en gaging the affective circulation of fear during the War on Terror, to an awareness that a therapeutic ethos has be come an important tenet for governing in the West. Second, nonrepresentational geographies have advanced a politics of hope which seeks to retrieve hope from its grounding in utopianism and into the realm of the 'not yet'. Finally, nonrepresentational geographies have sought to rework what constitutes the political and what it means to be political. As this approach encompasses the previous two, it will be focused upon, here. Non representational geographies do not take the available spaces for political enunciation as given (such as that of pressure groups or policy relevant research); neither, however, do they retreat into restoring a 'true' ground for the political (e.g., the Greek *polis*). Rather, they are con cerned with cultivating political spaces through an awareness of the openness of the present time. This fol lows from the notion that political and ethical dis positions do not proceed from the cognitive facilities of human beings but exist beyond conscious thought (we rarely, for example, consult moral codes in order to act ethically). Rather than consider this an impediment to political action, nonrepresentational theory seeks to harness its productive potential. On the one hand, through a politics of disclosure, not for authenticity, but for the creative potentials of the precognitive realm (by engaging with numerous body practices and the per formance arts, for instance). On the other hand, and in its more Deleuzian incarnation, this disclosure moves toward a politics of witnessing; a form of attunement which seeks to open up geographical and political thought to the unfolding (and sometimes wonder and enchantment) of the world. These approaches are, no doubt, risky and experimental (i.e., one cannot prescribe in advance the outcome), but not unproductive (i.e., something happens which might create new forms of life). In brief, politics and ethics are pulled away from judgment and universals, not simply as a matter of principle but because political and ethical dispositions exist on multiple (affective, emotional, anticipatory, precognitive, technological, molecular) registers which have thus far been neglected in much of political theory.

Current Imbrications and Challenges

Reception and Critique

Although in principle nonrepresentational theory seeks to avoid becoming a subdiscipline, the project initiated by Thrift has certainly fashioned a niche in the academy for likeminded geographers, or fellow travelers, who are advancing and expanding its central tenets. At present a handful of (largely UK based) geographers might be la beled as nonrepresentational, and several others (inter ested in post structuralism, (post)phenomenology, and more than human geographies) strongly intersect with many of its themes. The key motif of nonrepresentational theory however – to overcome the 'dead' geographies of representation – has (perhaps unsurprisingly) also been received with ambivalence, critique, and, occasionally, outright annoyance. These engagements vary but can be summarized by three main concerns: first, a neglect of the importance of representation; second, the denial of power and power geometries; finally, and drawing on the previous concerns, the forgetting of gender and/or sexual difference.

The first critique suggests that, although representation occludes much from view, many people actively (seek to) represent. Further, as the history of subaltern politics tells us, representing is a very effective political tactic. This critique has emerged in response to a perceived tendency of nonrepresentational theory to set itself up as positive alternative (i.e., what we might have) to programmatic and representational politics (i.e., what we have got) which, in itself, is a programmatic gesture. Yet, the ethos of nonrepresentational theory also asserts that there is not really a choice between representation and nonrepresentation, there are only singular presentations. Indeed, even a seemingly representative political gesture (such as the broadcasting of a presidential speech) incorporates a whole array of presubjective and inhuman affects. Still, it is partly in light of this critique that nonrepresentational geographies have recently adopted the term 'more than representational'.

Second, authors have critiqued the retreat from discourse and power toward practice and affect. They have also questioned the emphasis on expressive forms of embodiment which makes it difficult to account for what appear as relatively stable, sometimes restrictive, bodily practices (especially those that are connected to wider imperatives, such as gender roles or nation building, for example). By way of response, the geographer, Derek McCormack, has offered a kind of Deleuzian inversion of Foucault – by which the forces to affect and to be affected precede, and exceed, any stratified formations of power knowledge. Here, the largely ineffable virtual realm can actualize into diagrams. These diagrams are not plans, nor ideas, but a spatiotemporal consistency that holds forces to affect and be affected together. Although compelling, this approach does not address the racialized, gendered, and sexualized dimensions of this affective stability, and is unlikely to satisfy those concerned with this occlusion.

Finally, feminist critiques have suggested that nonrepresentational theory has a tendency to invert the very dualisms it seeks to overcome – mind/body, representational/nonrepresentational, thought/practice – and it is somewhat ironically able to do this through a disembodied stance (preferring abstract discussions of affect over emotion, and being rather than sexed being, for example). Feminist geographies and nonrepresentational theory share some familiar territory in geography (in particular an interest in performativity, the body and, more recently, emotions and bodily biotechnologies) and it is perhaps this shared territory which has provoked most debate between the fields. While nonrepresentational theory has a certain indebtedness to feminist theory (and, if the above arguments are followed, perhaps even a debt to the feminine), and while feminist theorists (such as Colebrook, Braidotti, and Grosz) have engaged with nonrepresentational philosophers such as Deleuze, nonrepresentational theory in geography has yet to address issues of sexual difference. This is, in part, due to its desire to avoid discursive or psychical understandings of the body subject. It also points to nonrepresentational geographies' preferred Deleuzian approach to difference 'in itself' rather than feminist geographers' engagement and deconstruction of difference 'between' (male/female, masculine/feminine, man/woman). Indeed, whether issues of sexual difference can be incorporated into the nonrepresentational ethos as it currently stands in geography, or whether nonrepresentational geographies might reconfigure feminist geographies' understandings of gendered and sexed bodies is, thus far, a largely undiscovered terrain.

Passivity and the Nonrelational

Recent engagements, which have emerged from within nonrepresentational geographies, have sought to interrupt its propensity toward the active, expressive, and/or otherwise affective nature of (embodied) experience. Paul Harrison has introduced notions of passivity and the nonrelational to counter the imperative for nonrepresentational geographies to always instill an affirmative will to connect, relate, and become. He uses the example of mental or physical pain – a nonintentional state we can only bear or endure – to conceive of a passivity which cannot be subsumed into relation, but is simultaneously its condition of (im)possibility. Whether the nonrelational will be absorbed into the grammar of nonrepresentational theory or whether it offers an alternative (post)phenomenological path followed by authors such as Agamben, Blanchot, Derrida, Lacoue Labarthe, Levinas, and Nancy, is in the process of being explored.

See also: Actor-Network Theory/Network Geographies; Affect; Becoming; Body, The; Dwelling; Emotional Geographies; Nature, Performing; Performativity.

Further Reading

Anderson, B. (2004). Time stilled spaced slowed: How boredom matters. *Geoforum* 35, 739 754.

Dewsbury, J. D. (2000). Performativity and the event: Enacting a philosophy of difference. *Environment and Planning D: Society and Space* 18, 473 496.

Dewsbury, J. D., Harrison, P., Rose, M. and Wylie, J. (2002). Enacting geographies. *Geoforum* 33, 437 440.

Harrison, P. (2000). Making sense: Embodiment and the sensibilities of the everyday. *Environment and Planning D: Society and Space* 18, 497 518.

Harrison, P. (2007). 'How shall I say it … ?' Relating the non relational. *Environment and Planning A* 39, 590 608.

Latham, A. (2003). Research, performance, and doing human geography: Some reflections on the diary photograph, diary interview method. *Environment and Planning A* 35, 1993 2017.

Lorimer, H. (2005). Cultural geography: The busyness of being 'more than representational'. *Progress in Human Geography* 29, 83 94.

McCormack, D. (2005). Diagramming practice and performance. *Environment and Planning D: Society and Space* 23, 119 148.

Thrift, N. (1996). *Spatial Formations*. London: Sage.

Thrift, N. (1997). The still point: Resistance, expressive embodiment and dance. In Pile, S. & Keith, M. (eds.) *Geographies of Resistance*, pp 124 151. London: Routledge.

Thrift, N. (2000). Afterwords. *Environment and Planning D: Society and Space* 18, 213 255.

Thrift, N. (2004). Intensities of feeling: Towards a spatial politics of affect. *Geografiska Annaler* 86B, 57 78.

Thrift, N. (2004). Summoning life. In Cloke, P., Crang, P. & Goodwin, M. (eds.) *Envisioning Human Geographies*, pp 81 103. London: Arnold.

Thrift, N. and Dewsbury, J. D. (2000). Dead geographies and how to make them live. *Environment and Planning D: Society and Space* 18, 411 432.

Wylie, J. (2006). Depths and folds: On landscape and the gazing subject. *Environment and Planning D: Society and Space* 24, 519 535.

Nordic Geography

K. Simonsen, Roskilde University, Roskilde, Denmark

Introduction: In What Sense a 'Nordic' Geography?

In this age of excessive transnational exchange of knowledge, this article has to start from a consideration of the legitimacy of something called 'Nordic' geography. On the one hand, global circulation of information and ideas – with all its inequalities, restrictions, and deformations – renders problematic any essentializing claims and underlines the hybridity of all knowledge production. On the other hand, a widespread acknowledgment within contemporary human geography is about the contingent and contextual character of knowledge/s. As expressed in a notion such as Haraway's *Situated Knowledges*, knowledge/s has to be grasped in the place and time out of which it emerges. Even though situated knowledges is not necessarily local knowledges, each author is situated in many different ways, where the social and cultural contexts of the production of knowledge do matter. This is not a claim on social theory and knowledge being inescapably contexts bound, but rather an acknowledgment of the way in which they always carry some particularity and inherent assumptions from their place of origin. Cultural/academic traditions and institutional settings still situates them in time and space and facilitate some perspectives and ways of knowing at the expense of others. It is on these premises this article tries to draw a 'map' of Nordic human geography, seeing it as a discourse while at the same time intervening in 'international' debates and producing its own specificity/ies.

Now, the very designation 'Nordic' should, of course, also be used with caution. When a historical scholar such as Benedict Anderson ends up defining a nation as an 'imagined community', doubt must necessarily be cast upon calling anything inherently Nordic. National as well as 'Nordic' identities are phenomena of discourse, constructed at distinctive junctures in time, and for specific purposes. The invention of 'the Nordic' can, in fact, be traced back to nationalist Romantic movements of the nineteenth century. Alongside the spread of a nationalist discourse within the individual Nordic countries, a transnational ideology extolling a Scandinavian or Nordic spirit of communality arose. Interest in the uniquely Nordic past (e.g., the Vikings) was part of national, as well as Nordic, imaginations.

Acknowledging the Nordic as a construction – an imagined community – does not however mean that this construction does not work in a material sense, nor that particular connections and collaborations do not exist. Indeed, Nordic cooperation has a long tradition, especially when it comes to cultural collaboration between movements rooted in the civil society and between professions, scientists, and artists. Human geography is no exception to that. Also, even though the Nordic countries are not as different from the other European countries as ideology would sometimes have us to believe, certain distinctions do prevail. It is difficult to talk about a 'Nordic model' *per se*, but throughout the twentieth century the Nordic countries have probably undergone a more smooth modernization than have most other countries in Europe, and the welfare state has stood its ground.

If this renders possible the whole idea of writing about Nordic geography, the writing of such an article is still loaded with a touch of ambivalence. Writing a story of a 'local' geography from 'within' carries the danger of casting the writer as an unproblematic translator who – by way of a dual position between discourses – mediates the 'not easily' accessible to 'outsiders'. Like any other 'map' of an intellectual landscape, it is a social construction, a narrative construed by a writer who is ambiguously positioned in the very field one is trying to describe. This should not be seen as a confession or a way to disclaim the responsibility for the story to come, but rather as a problematization of the very presentation of local intellectual landscapes suggesting that, like all other practices of representation, they are necessarily situated, embodied, and partial.

Nordic Geography: Differences and Similarities

One constitutive contextual element for the development of human geography in the Nordic countries is the welfare state or what Esping Andersen calls "the social democratic welfare state regime." Some parts of the subject directly address the welfare state; for others, welfare issues are more or less conscious assumptions that permeate the theoretical and/or empirical formulations. This connection is a double edged one and the risk of importing into academia the kind of welfare state nationalism or 'Nordism', which has grown in popular discourse in the later decades in concurrence with challenges from immigration and European integration, is not negligible. At the same time, however, the prevalence of the welfare state ideology has carried a persistent clinging to egalitarian values, such as social justice, social equality, and politics of redistribution, as a background to academic discourse. In particular in Sweden, the marriage between human geography and the welfare state has resulted in a strong orientation toward planning and applied research. Even though this connection is less prevalent in the other Nordic coun tries, the orientation toward welfare and equality has resulted in a focus on issues such as regional develop ment, labor markets, planning, social reproduction, and the organization of everyday life, while, for instance, research into identity formation, representations, and imaginations (with a few exceptions) has been less prominently featured, and has only been embarked upon in more recent years.

Theoretically, Nordic human geography (like other social sciences in the Nordic countries) has been marked by a specific situatedness at the interstices of Anglo Saxon and continental intellectual movements. Both as regards geographical thinking and more general social theory, affiliations have – to a different degree – existed to, for instance, German, French, and Anglo American discourses. While these influences have, to a certain de gree, articulated with local features, different currents have also characterized the different Nordic countries. These differences are historically based and have emerged due to academic traditions and merits, insti tutional settings and organizational structures of the universities, and economic and political–ideological conditions in the different countries. In brief, in all countries the history of the subject has been marked by ideas and practices of utility, in relation to either training of teachers or forming a basis for planning. In the cases of Denmark, Finland, and Norway, training of teachers re mained the dominant function of geography until the 1970s. That is probably why (mainly German inspired) discourses on 'landscape' as a notion connecting physical and social circumstances of regions maintained a strong position in geography in these countries, and the subject repeatedly was facing discussions on the synthesis be tween physical and human geography, between science and social science.

Swedish human geography, on the contrary, both when it comes to institutional organization and subject matter, can be described as a social science since the 1930s. But it has been a particular kind of social science – an applied one based on a close connection between human geography and regional and local plan ning. This marriage opened up great opportunities for the Swedish human geographers when it came to a practical and methodological development of the subject, but it has as well caused problems with regard to the addressing of more philosophical and theoretical ques tions. The point here – maybe not that surprising – is that despite the relative similarities between these countries there is no 'single' Nordic human geography, no unitary focus or sole mainstream in the Nordic countries' geographical discourse. Differences in philosophical and methodological approaches are overlaid by the described differences in 'national' discourses in the making of a web of different interests and ideas. Specific academic developments have occurred, and they also influence the contemporary geographical discourse/s. Having said that, however, there is no doubt that the formation of networks across the Nordic countries, in particular be tween critically orientated geographers, as well as an increasing participation in 'international' discussions have leveled some of these differences and created cross national 'language games'. Also, in the postwar development of the subject it is possible to identify some distinctive moments that, even if they are differ ently based between the countries, have had a common appeal.

One of these is probably the most internationally well known contribution from Nordic geography, namely, 'time geography', introduced by the Swedish geographer Torsten Hägerstrand and his associates in the 1970s. It was a contribution that actually managed to combine applied research and association to planning with more philosophical reflections. Time geography was not supposed to be a theory, but rather an ontological con tribution focusing on how different phenomena are mu tually modified because they coexist in time and space. Hägerstrand himself, in a posthumously published paper, called it a 'topoecology' designated to grasp a society–nature–technology constellation. The basic element of time geography is the connections between continuous trajectories of individual entities in time–space. On that element, a set of descriptive concepts was developed, such as 'paths', 'stations', 'projects', 'prisms', 'time–space bun dles', and 'time–space domains'. They combine to a worldview on human condition stating that everybody is subject to a confinement in time–space within the limits formed by the bounded capacity of individuals to engage in more than one task at a time, by the speeds by which it is possible to move and assemble individuals, tools, and materials, and by regulations of access and modes of conduct within domains of local order. Most well known is the translation of this view into a nonlinguistic notation system representing 'possible' configurations in time–space. It proved a useful tool in planning and has been heavily used as a descriptive tool in Sweden as well as in other places. Also, time geography at a relatively early stage rendered everyday life a legitimate research object in Scandinavian geography.

While time geography was still embedded in a relative naturalist language, this was transcended in the second distinctive moment of postwar human geography, that is, the development of a 'critical/radical' geography that, like in other places, translated human geography into a genuine social science. A strong (cross disciplinarily based) neo Marxist geography developed during the 1970s and 1980s, having its main basis in Denmark, maybe due to a stronger affiliation to political movements in other parts of Europe. The development was performed in dialog with geographers from the Anglo American context, but it took its specific form by way of conjunction with other 'Marxisms' imported from France, Germany, and the then Eastern Europe. Empirically, uneven development on the urban, regional/national, and international scale was addressed, and geographers participated in public debates and activism in relation to planning and other issues. The new radical geography was discussed and challenged through joint Nordic conferences in critical geography, from those including particularities from different countries. From Norway came a 'local community' discourse, strongly embedded in a center/periphery imagination, and from Sweden a local 'action research' based in cooperation with local unions. Feminist geography developed as an offshoot from the radical discourse. Paradoxically, seen in relation to the relative strong position of the Nordic countries regarding issues of equal opportunities, it was a relative latecomer in Nordic geography. Even today, only Sweden can claim to have a well developed feminist geography.

Finally, one author, who is difficult to categorize, is worth mentioning for his role in Nordic and international geography, that is, Swedish geographer Gunnar Olsson whose role in Nordic geography has been less of 'making school' than of constantly challenging established truths and enforcing reflexivity. What, for Olsson, started as a critique of spatial analysis and connected ways of reasoning in planning and politics, developed into a philosophical project of an investigation into the 'cartography of thought' – into the invisible spaces of how ideas in the mind are shaped, combined, delimited, and hierarchialized. Olsson's work constantly focuses on the power and limits of language; it overturns all conventionalist forms of representation and mobilizes new conceptual geometries and modernist esthetics in order to 'think the unthinkable' and open new possibilities for human thought and action.

Dominant Current Contributions

It is of course impossible in the space available in this context to produce a satisfactory overview of contemporary Nordic human geography. Also, making delineations and categorizations of the field area dubious matter, which is deemed to fail by homogenizing the heterogeneous and violating the connections. Therefore, this article chooses another road and presents three issues in which I am prepared to claim that Nordic human geography has provided distinctive contributions.

Economy and Regional Development: A 'Scandinavian' School?

Nordic economic geography has a strong position both internally and externally, particularly, when it comes to analyses of innovation and post Fordist economies. Two background conditions – one contextual and one academic – might provide part of the reason for that. First, part of the readiness to grasp these issues might be seen as a consequence of some specific characteristics of the Nordic setting: the small size, general openness, industrial structures, and relatively high levels of prosperity of the economies of the Nordic countries. Small economies often differ from larger ones on issues, such as degree of openness, exposure to international competition, affinity toward corporatism, and networked relations across firms. It is possible to argue that these characteristics are a factor in explaining why Nordic scholars in economic geography have been particularly interested in the analysis of how networked relations in various industrial systems help create stable specialization patterns, and how innovation and competitiveness emerge from interactive processes with a markedly localized component. Second, the size of the discipline might have had an influence. Human geography in the Nordic countries ranks among the smallest social sciences. It has, in a sense, 'forced' geography to exploit its interdisciplinary potential as a synthetic discipline. Economic geographers have been rather proactive and positive to the contributions coming from other disciplines dealing with regional questions and to the necessity of establishing international networks. This has especially become evident in the close cooperation that exists with other heterodox (evolutionary and institutional) economists in the area of innovation studies.

Contemporary economic geography that, particularly in Sweden, takes a dominant position has been developed from two sources. One is the Marxist economic geography of the 1970s, which introduced subjects such as regional inequality, technological development in capitalist societies, and social consequences of industrial closures. The other, related, source is a Nordic variant of what, in Anglo American circles, is known as studies of the regional consequences of industrial restructuring. It was connected to the welfare state and involved a focus on regional policy and the options of local and regional measures in industrial policy. During the latest decades, keywords in the economic geography research have been 'innovation', 'knowledge', and 'learning', in different ways analyzing the importance of territorial agglomerations for the innovativeness and competitiveness of firms and regions in a globalizing economy. This research has been conducted in close cooperation with colleagues from other disciplines, together forming an internationally strong Nordic innovation research. Theoretical and empirical work has been performed on industrial districts, regional clusters and regional innovation systems. These are all concepts that see innovation as a socially embedded process, underlining the importance of 'noneconomic' factors, such as history, culture, and institutions for economic development. However, after having promoted these ideas of the importance of territorial agglomerations and 'learning regions' to economic growth, the continuous research has as well, generated reflexive questionings of established 'truths'. They particularly concern the understanding of learning, differentiating between various forms of knowledge, their interrelationships, and their relations to space and place.

While the new Nordic economic geography has been extraordinarily engaged in open discussions with international partners and with other economic disciplines, it as a subdiscipline has isolated itself more from the rest of human geography. The consequence is that the emphasis on noneconomic factors of regional development has missed the opportunity of mutual exchange with social and cultural parts of the discipline and that economic geography has lost some of the social engagement characterizing its roots in the critical geography of the 1970s.

Social and Cultural Geography

One characteristic feature is the existence in all the Nordic countries of a social geography that is concerned with the welfare state. The relationship is a dual one, combining explorations of the role of the welfare state in various aspects of social development and a more direct feeding of research results into welfare policy and planning. The main issue at stake in this group of works is social differentiation/polarization, often related to a theoretical discussion of urban development and urban policy in a globalizing world. This extends, more than anything else, to questions of housing and evolves on the classic urban social geography issue of segregation. Extensive empirical studies on 'social polarization' and 'segregation', often conducted in comparison and cooperation with other European countries, have been framed in the intersection between overarching and local explanations. In particular, the role of the welfare state in this process is addressed, leading to exposure of the paradox that the very pride of Scandinavian housing policy – the provision of social housing – has ended up being part and parcel of processes of 'ghettoization'. When it comes to a 'cultural turn', in particular, two countries have led on, each taking its own specific direction but both seeking a position in between constructionist and materialist perspectives.

Finland was the first Nordic country to open up to 'humanistic' approaches to geography. Eventually, however, the most prominent contribution from Finland is an approach working in the interstices between cultural and political geography and having 'territoriality' and 'identity' as keywords. Finnish geographers, maybe under the impression of the shifting geopolitical situation of the country, have performed an extensive theoretical and empirical work on the role of boundaries in the construction of territories and territoriality and subsequently theorized the changing meanings of territoriality in the changing social conditions. This development has progressed in critical dialog with 'international' theorization in the field. The question of identity has been a very important one in this stream of works. Originally developed by Anssi Paasi as a theoretical approach to the question of regional identity, it has been extended to a more general approach to space, scale, and identity. For example, it has found expression in a number of studies on nation building processes and national identity. This cultural–political approach has formed the background for a range of empirical studies, both in Finland and other parts of Europe.

In Denmark, social and cultural geography has most prominently been approached through the lenses of social theories of 'practice'. The authors involved in this exercise have taken off from a variety of sources, but what they have largely had in common is a view on subjectivity as primarily derived in practice. The subject's understanding of the world comes from lived experience and dialogical actions, bound together by mutual dispositions and shared understandings which they both take from and contribute to. The 'cultural turn' in Denmark, then, from the beginning took the form of a 'practice turn'. The concern has been with embodied knowledges and their formation in people's everyday lives, with the world of emotions, narratives, and imaginations, and with the multiplicity of encounters through which we make the world and are made by it in turn. An important issue in this stream of works has been the 'spatiality' and 'temporality' of practices. This has found expression in works on cities and urban everyday life, emphasizing the variety of temporalities and spatialities involved in the social construction of the city. Furthermore, different practices of mobility are analyzed in relation to constructions of place and identity. Finally, a research stream exists starting from a range of dissertations on the relationship between work practices, identity, and social change and developing to a more general engagement with the culture–economy interface, theorized as interconnected practices of innovation, social networking, and the formation of cultural identity.

Landscape and Environment

While the traditional approaches to landscape, in the Nordic countries, were partly influenced by the German *Landschaftsgeographie* and partly relied on intensive historical geographies of material cultural landscape, lineaments of cultural politics of landscape have come to the center in current renewals of the field. This work has grown in close dialog with 'international' discourses, and some authors argue that besides the British pictorial scenic approach informing the 'new cultural geography' and the layered 'blood from soil' approach originally developed in Germany, a third approach to landscape exists seeing it as 'land shaped by policy'. This third approach, characterized by a concern with history, cus tom/law, and language and culture as they work together forming a landscape polity and its geographic place, is argued to be strongly, but of course not exclusively, represented in the Nordic countries. Kenneth Olwig uses the term 'platial' to describe this approach, thereby constructing a shorthand for a spatiality in which the polity and its material fabric are constituted through ongoing (place oriented) social practices and ideologies of political representation, including customary law. Of course, these approaches to landscape are neither mu tually exclusive nor exhaustive. A creative example on their combination explores how a tension between such a 'platial' notion of landscape and a visual one, based on the possessive spatial rationality of the gazing individual, were pertinent to the protracted struggles over political representation in seventeenth century's the Netherlands. Also, in a more current context, Danish–Swedish regional identity politics are approached in the interface between three different concepts: landscape as a territory with a perceived internal coherence, landscape as scenery, and landscape as changing juxtapositions of assorted entities flowing in time and space.

Another, connected, line of thought more directly ad dresses the politicization of nature. Here, the culture–nature practices are critically evaluated as questions of domination and empowerment, both within local spheres of intersubjective discourse and in relation to industrial–technological change within local–global interconnections. A representative for this approach is the Finnish geog rapher Ari Lehtinen. He writes from the tradition of Nordic critical geography, and in his writings on Nordic forest communities, he combines ideas of political ecology with Foucault inspired thoughts of biopolitical power and an inclusion of the local traditions of community based politics.

Thus, what connects all the three groups of work, brought up to characterize current trends in con temporary Nordic human geography, is a combination of international orientation, particularly Nordic/North European developments and an intrinsic sensitivity to contextuality.

See also: Economic Geography; Innovation; Landscape; Radical Geography; Regional Development Theory; Situated Knowledge, Reflexivity; Territory and Territoriality; Time and Historical Geography.

Further Reading

Andersen, H. T. and van Kempen, R. (eds.) (2001). *Governing European Cities. Social Fragmentation, Social Exclusion and Urban Governance*. London: Ashgate.

Andersson, H., Joye, D., Jørgensen, G. and Ostendorf, W. (eds.) (2001). *Change and Stability in Urban Europe: Form, Quality and Governance*. Aldershot: Ashgate.

Asheim, B. T. (2006). Economic geography as (regional) contexts. In Bagchi Sen, S. & Smith, H. L. (eds.) *Past, Present and Future of Economic Geography*, pp 173 185. London: Routledge.

Buttimer, A. and Mels, T. (2006). *By Northern Lights: On the Making of Geography in Sweden*. Aldershot: Ashgate.

Hagerstrand, T. (1985). Time geography. Focus on the corporeality of man, society and environment. In Aida, S. (ed.) *The Science and Praxis of Complexity*, pp 193 216. Tokyo: The United Nations University.

Hakli, J. and Paasi, A. (2003). Geography, space and identity. In Ohman, J. & Simonsen, K. (eds.) *Voices from the North. New Trends in Nordic Human Geography*, pp 144 155. Aldershot: Ashgate.

Jones, M. and Olwig, K. (eds.) (2008). *Nordic Landscapes: Region and Belonging on the Northern Edge of Europe.* Minneapolis, MN: University of Minnesota Press.

Lehtinen, A. (2003). *Samhällsgeografi* and the politics of nature. In Ohman, J. & Simonsen, K. (eds.) *Voices from the North. New Trends in Nordic Human Geography*, pp 233 246. Aldershot: Ashgate.

Maskell, P., Eskelinen, H., Hannibalsson, I., Malmberg, A. and Vatne, E. (1998). *Competitiveness, Localised Learning and Regional Development. Specialisation and Prosperity in Small Open Economies*. London: Routledge.

Ohman, J. and Simonsen, K. (eds.) (2003). *Voices from the North. New Trends in Nordic Human Geography*. Aldershot: Ashgate.

Olsson, G. (1991). *Lines of Power/Limits of Language*. Minneapolis, MN: University of Minnesota Press.

Paasi, A. (1996). *Territories, Boundaries and Consciousness*. Chichester: Wiley.

Paasi, A. (2005). Between national and international pressures: Contextualizing the progress of Finnish social and cultural geography. *Social and Cultural Geography* 6(4), 607 631.

Simonsen, K. (2003). On being 'in between': Social and cultural geography in Denmark. *Social and Cultural Geography* 4(2), 255 269.

Simonsen, K. (2004). Temporality, spatiality and the construction of the city. In Bærenholdt, J. O. & Simonsen, K. (eds.) *Space Odysseys. Spatiality and Social Relations in the 21st Century, pp 43 63*. Aldershot: Ashgate.

Nordplan and Nordregio

G. Olsson, Uppsala University, Uppsala, Sweden

Introduction

From the 1960s to the present both Nordplan and Nordregio have played major, yet distinctly different, roles in the theory and practice of spatial and regional planning in the Nordic countries. Both institutes were established by the Nordic Council of Ministers, Nordplan (the Nordic Institute for Studies in Urban and Regional Planning) in 1968 and Nordregio (the Nordic Centre for Spatial Development) in 1997. As political children of their respective times, the shared roots reach deeply into the mixed soil of social welfare theory and social engineering practice. The differences between the two are nevertheless considerable, for whereas the emphasis of Nordplan was on graduate training and basic research, Nordregio's main mission is to serve as a clearinghouse for data collection and officially commissioned investigations. Common to both is the ideological mantra that intentional action should be doubly anchored; one fluke in the facts of scientific knowledge, the other in the norms of utilitarian ethics – in essence, a social democratic reformulation of the Marxian maxim that while the philosophers are trying to understand the world, the real point is to change it.

Nordplan

One-Year Program

In the early years of post war Europe there was a strong determination to build a new and better society, a democratic world free of dictatorial fascism, albeit not always of dictatorial communism. In that move from abstract ideas to concrete implementation, a key role was to be played by a cadre of bureaucrats with an ear sensitive to the political tunes of the day and a professional hand equipped for the task. From the very beginning, however, this ideal created its own problem, for there were simply not enough experts around to meet the growing demand. In addition, those who actually did get recruited into the professional ranks (primarily architects, engineers, economists, and geographers) often lacked the necessary experience and interdisciplinary background. The situation was essentially the same in all the Nordic countries (Denmark, Finland, Iceland, Norway, and Sweden) and since the numbers were too small for the creation of separate national schools, it was agreed that here was an ideal case for internordic cooperation.

Thus, after a decade of gestation, Nordplan was eventually born – an organization sprung from the Olympian heads of a small group of concerned architects and determined social engineers. Through the political connections of Lennart Holm – longtime chairman of the board – the institute was placed in Stockholm, not because the ideas were particularly Swedish, but merely because it had to be somewhere.

A number of teachers were recruited, including a troika of full professors, one in physical planning (Per Andersson, a Norwegian architect and consultant), one in transportation planning (Stig Nordqvist, a Swedish engineer and consultant, in addition the institute's first and longest serving director), and one in economic geography and planning (Tor Fr. Rasmussen, a Norwegian geographer with his doctorate from Lund). The students (really not students at all, but a group of 50 mid career bureaucrats) came from all of the five countries. The majority enrolled for a 1 year program of continuous education split into three intense periods of 2 months of lectures and seminars in Stockholm and 1 month of fieldwork in one of the other countries.

The languages of instruction were Danish, Norwegian, and Swedish, which after some initial difficulties could be understood by all. Since everyone (some Finns excluded) spoke and wrote his/her own mother tongue, the practice of conversing in a foreign language quickly changed from a deafening handicap into a whispering opportunity. This form of language training was in fact one of the major assets of the program, for it bore directly on the principles of teaching and learning. The reason is that even though the political, social, economic, and planning issues are essentially the same in all the five countries, the attempted solutions tend to be slightly different. In the numerous group meetings it was therefore natural that each participant initially argued for the particular approach into which he/she had been socialized. After a while, however, and often in confrontation with the beliefs and practices of the others, many began to question their own taken for granted. With those doubts as keys, the dialectical door of self conscious reevaluation was unlocked; indeed, it was normal that after the Nordplan experience a fair number left the planning profession for good. Although no formal degrees were offered, the accumulated impact was remarkable, individual lives profoundly changed in the process. In total about 1500 searching professionals went through the experience of going to the limits, by definition the only way to learn.

The months away were personally demanding and socially dangerous, for it was impossible for the practitioners not to experience their own activities as a power filled game of ontological transformations, a conception of action in which ideology turns to paragraphs, paragraphs to stone, stone to people. In addition, the teaching taught them that a basic tenet of utilitarianism is that the goodness of a given act should be judged in terms of its consequences, not in terms of the actor's intentions; tragedy in the making, reification, and alienation as well. But do the ideologues get lost because they are scientifically naïve or because they are morally wrong, obsessed by power for the sake of power? Little wonder that the discussions stretched into the nights, the trickling down tilting the bottoms up.

Over the years the course content and the ideological debates underwent considerable change, and it was often difficult to know whether the discussions said more about the times or whether the times revealed more about the discussions. The turns and twists were many, but while most of the early traffic was confined to the well paved highways of social engineering, the issues of the 1980s and 1990s were frequently found scattered around the swarming pathways of social movements. Throughout there was a penetrating focus on the relations between structure and event, large systems and human existence, power and counterpower. In the process, the role of planning got so seriously questioned that in the long run the institute could not survive. What nevertheless remains is a large invisible college, a tightly knit network of social and professional relations.

Doctoral Program

Although Nordplan began and ended as an institute devoted to continuous education, the founders realized that in the long run not even the most goal oriented teaching can do without free and basic research. The idea of a doctoral program was therefore explored and a series of seminars and examinations were conducted. In 1977 Gunnar Olsson was appointed as the new professor of economic geography and planning, and it was under his directorship (increasingly with the help of Kerstin Bohm, adjunct professor of sociology, and José L. Ramírez, teaching assistant) that the program found its characteristic form and content. Most crucial was the decision by the Nordic Council of Ministers to grant the institute the right to offer doctoral degrees in planning, an unprecedented event that effectively turned Nordplan into a miniscule but independent university.

The program was divided into two overlapping phases, the first organized as a 3 year long series of intensive seminars, readings, and excursions, the second of the writing and public defense of a dissertation. New students were accepted every third year and participation in the seminars was obligatory. Normally, each cohort consisted of about 16 members evenly divided between Denmark, Finland, Norway, and Sweden. The participants' professional background varied widely, but the majority were architects or social scientists, including a small number of geographers. Roughly, equal proportions were university teachers, centrally placed administrators, and active scholars from a range of national research institutes. No applicant with less than 10 years of professional experience was ever accepted, about half of them already holding the equivalent of a British or American PhD. Given these requirements, and since most participants kept their ordinary jobs throughout the study period, the average age was considerably higher than in conventional programs.

Many factors influenced the design of the program, but none was more important than this issue of age; every day in fact confirmed Aristotle's observation that "each man judges well the things he knows, and of these he is a good judge.… Hence a young man is not a proper hearer of lectures on political science; for he is inexperienced in the actions that occur in life." And so it is that comparing the dissertations of a 28 and a 60 year old involves a double insult, first to the young, whose potential is denied, and then to the old, whose experiences are ignored.

The heart of the program lay in the seminars, which met for a full week 7 times a year, approximately 20 times for each cohort (1978–80, 1981–83, 1984–86, and 1986–89). As a rule, three of the yearly meetings were held in Stockholm, one in Denmark, Finland, and Norway, respectively, and the seventh in a non Nordic country. Especially noteworthy among the latter events were the several meetings in St. Petersburg (Leningrad), Paris, Cambridge, Paris, Milan, Bologna, Berkeley, Los Angeles, Andalusia, Toledo, and Madrid. The pedagogic idea was to introduce the students to the various schools of thought in their original settings, thereby emphasizing the contextuality of everything taken for granted. The long list of lecturers reads as a veritable 'who is who in international social science'.

From the first seminars in 1978 to the last dissertation in 1995 the program was held together by the idea that knowledge is an exercise in translation, the rhetorical art of reformulation in practice. Telling the truth is consequently to say that something is something else and be believed when you do it – a paradigmatic formula so minimalistic that once it has been learned it can never be forgotten. No closure, though, for despite the fact that every serious investigator aims for perfect translation, the structure of the sign itself guarantees that perfect translation is impossible. The circumstance that at least four different languages were always spoken in the same seminar room provided ample proof that even though the Nordic cultures are closely related, they are at the same time drastically different. However, since what cannot be said can sometimes be shown, the restless feet and the pinching stomach gave clear warning that what is the normal here may well be the deviant there, the lo gical primitives of identity and difference by necessity separated by the emptiness of the excluded middle, that latter abyss serving as the well guarded home of power itself.

The conclusion came automatically: knowledge is as self referential as its own language, collective action always an encounter not only with Kenneth Arrow's impossibility theorem but also with logical paradox and moral predicament. As illustrative examples of the transdisciplinary nature of the program, the second seminar in Cambridge was built around half day sessions with David Bloor, Denis Cosgrove, Mary Douglas, Anthony Giddens, Derek Gregory, David Harvey, Gillian Rose, Nigel Thrift, and Raymond Williams. Along the same lines, the last Berkeley seminar – significantly en titled 'Languages of power, gender, and mind' – included meetings with James Clifford, René Girard, Donna Haraway, Robin Lakoff, Hanna Pitkin, Allan Pred, Paul Rabinow, Nancy Scheper Hughes, Richard Walker, and Michael Watts.

The radical message was constantly hammered home that the 'taken for granted' cannot, indeed must not, be taken for granted. It follows that the real challenge is not in understanding the other but in understanding oneself, albeit never directly but always via the other. But exactly as knowledge by definition is an exercise in translation, so politics is a struggle between us and them, master and slave, good and evil, here and there, now and then. For that reason – and given the participants' backgrounds and experiences – the case of Nordic planning was used as the central study object.

But what is planning? The answer is that planning is a special case of human action located in the taboo ridden interface between individual and society, a magic per formance of ontological transformations through which social meaning is turned into physical matter, use value into exchange value, signifier into signified. And vice versa. Since the overarching purpose of the teaching was to understand rather than interfere, the focus was on the development of theories 'about' planning, not on the establishment of rules 'for' planning, analysis instead of therapy. Since power is the name of the game, many were shocked to discover that it is normally more difficult to understand the world than to change it.

But how can one lay bare the techniques by which we become so obedient and so predictable? Perhaps by writing a drama in which H. C. Andersen's *Emperor* and Goethe's *Faust* are allowed to perform on the same stage. A dance of fragments, a choreography of micro and macro, rings of symbols ruling the waves. And for that reason every seminar week featured a visit to a theater or a museum, a meeting with a poet, painter, or sculptor.

Dissertations

Participating in a series of advanced seminars is a rare and exciting adventure. Writing and defending a doctoral dissertation, especially if it is done at a mature age, is even more demanding. The reason is – and herein lays the main difference between the two phases of the pro gram – that whereas the seminars were shared experi ences, the dissertation is a lonely work undertaken in the best of companies. For how many can spend years on their own self portrait?

In the Nordplan case surprisingly many. To be precise, about 70 curious individuals participated in the obligatory seminars, three quarters of them eventually subjecting themselves to one full day of oral examination. Even more remarkably, a total of 33 middle aged people (some in their late 50s) completed and defended their theses. Although on the surface the chosen topics might seem quite different, most texts shared a particular flavor of interdisciplinary research and philosophical awareness. As a way of guarding the quality – for a young and small institute a matter of crucial importance – the assessment committees were twice as large as in any of the Nordic systems.

Final Years

What has been politically created may be politically des troyed. And this is exactly what happened with Nordplan, which from the very beginning had been navigating be tween the political Scylla of social democracy and the professional Charybdis of neoliberalism; while the former forces clearly tried to fit the institute into its own agenda, the latter viewed its activities with great suspicion.

The ideological climate was nevertheless changing and in the wake of 1984 the critique of planning was gaining ground. In addition, the traditionally close ties between the Nordic countries began to weaken as more and more of the bureaucratic business was moving to Brussels. The resistance was strong, but in 1989 the edict was issued that no new doctoral students were to be ac cepted, albeit with the provision that those who were already enrolled had until the end of 1995 to finish.

With the doctoral program gone, the closure of the institute was a forgone conclusion. On the first of July 1997 Nordplan was no more. A glorious experiment sacrificed on the same altar on which three decades earlier it had been conceived, born, and blessed. A dia lectical case of power performed and power disrobed – a once dominant ideology on its deathbed.

A summary of the doctoral program, including per sonal testimonials from the participants, is in Gunnar Olsson's *Chimärerna: Porträtt från en forskarutbildning*. The title of that report is telling, for according to the dic tionary a 'chimæra' is (1) a fire breathing monster of Greek mythology, with a lion's head, a goat's body, and a serpent's tail, a creature that was killed by Bellerophon, himself the grandson of Sisyphus, the existentialist figure who in life was rewarded for his cunning slyness, in Hades punished for exactly the same traits; (2) an unreal creation of the imagination, a mere wild fantasy, an un founded conception; and (3) an incongruous union or medley. In addition, 'chimæra' is the botanist's term for a hybrid growth in which a branch of one genetic setup has been grafted onto the stem of another kind. Since the respective genetic setup of the two is not affected, it follows that even though the branch is given new strength, it is not fundamentally altered.

Such was the dream come true: an educational grafting. Not a factory for the production of predesigned copies, but the home of some unique originals. Not a group of brainwashed apparatchiks, but a set of indi viduals who knew that in order to change the world you must first understand yourself. Little wonder that by the year 2007 a total of 30 former doctoral students had been appointed to full professorships in a range of disciplines, seven in Denmark, one in Estonia, five in Finland, one in Germany, one in Ireland, eight in Norway, and seven in Sweden. The vast majority of the others are holding leading positions in (inter)national research, adminis tration, and politics.

Nordregio

Since power never sits still, it refuses to be caught in the net of logical categories and empirical observations. Deconstruction in practice.

Thus it was that when the Nordic Council of Ministers in 1997 concluded that the independence of Nordplan was not to their liking, they at the same time created Nordregio, a phoenix out of the ashes. According to one of its own websites:

> Nordregio is an institute for applied research and de velopment. Our fields of study include regional devel opment, urban policy and spatial planning, as well as the cross cutting aspects of spatial and regional policies. Sustainable development and territorial cohesion are among the most important policy issues in this respect. The geographical focus of the Institute is on the Nordic countries, the Baltic Sea region and the European in stitutional dimension. Our main clients are the Nordic Council of Ministers, the European Union, and the governments and regions of the Nordic countries. The institute is located in attractive surroundings in the City of Stockholm, Sweden.

How telling that the only thing to remain is the building. But the house that was once open for anyone to enter is now an edifice locked up behind a secret code, and yet, through the efforts of Hallgeir Aalbu, Nordregio's first director, the center has in fact become a most successful organization of flexible accumulation, an extension of systems thinking to policy making. The genuinely inter disciplinary staff consists of about 30 internationally re cruited research fellows all governed by a board whose members come from the Aaland Islands, Denmark, the Faeroe Islands, Finland, Greenland, Iceland, Norway, and Sweden.

Like the two institutes, these representatives are necessarily children of their own time, a typical instance of text and context intricately interwoven into a system without exits. For that reason it must be stressed that whereas Nordplan was financed directly over the Nordic budget, hence relying on relatively hard money with long range commitments, Nordregio is largely dependent on soft money from temporary research projects. Equally important is the fact that whereas the full professors of Nordplan were granted tenure by the Swedish state, the fellows of Nordregio are footloose hirelings on temporary and nonrenewable contracts. The conditions for basic research and fundamental critique are obviously distinctly different. To be precise, the focus is no longer on the growth of individual students but on the presumed needs of the political clients. Sustained development of another kind, the principle of self reference disguised.

Viewed together, Nordplan and Nordregio presents a fascinating case of organizational adjustment and ideo logical reorientation. As Søren Kierkegaard used to put it: "we live forwards and understand backwards."

See also: Nordic Geography.

Further Reading

Mariussen, Å. and Juhlin, Å. (eds.) (2006). *Transnational Practices: Systems Thinking in Policy Making*. Stockholm: Nordregio.

Olsson, G. (ed.) (1995). *Chimärerna: Porträtt från en forskarutbildning*. Stockholm: Nordplan.

Olsson, G. (ed.) (1996). *Poste Restante: En avslutningsbok*. Stockholm: Nordplan.

North–South

A. McGregor and D. Hill, University of Otago, Dunedin, New Zealand

Glossary

Keynesian-Style Economies A term originally used to designate the method of analysis developed by John Maynard Keynes as an explanation of the Great Depression. It later came to denote a state-interventionist style of economic policy. Keynesian-style economies have an emphasis on controlling aggregate demand and thus regulating fluctuations in the economy.

Kyoto Protocol An agreement undertaken under the United Nations Framework Convention of Climate Change. If this protocol is ratified, signatory countries are obliged to reduce their emissions of six greenhouse gases, including carbon dioxide, methane, nitrous oxide, and three groups of fluorinated gases. These signatory countries are able to maintain or increase their emissions of these gases only if they engage in emissions trading.

ODA Official Development Assistance: A term referring to funds transmitted from the countries of the Global North to the Global South as foreign aid, usually in the form of a bilateral agreement. The nature and purpose of ODA has shifted over time as new priorities have emerged in donor–recipient relationships.

Structural Adjustment Generally consists of both short-term stabilization and medium to long-term adjustment programes. The first priority of stabilization is to control rates of inflation and the budget deficits caused by both external and internal budgetary imbalances. The adjustment phase, intended to bring about longer-term structural changes to the economy, includes a wide variety of measures aimed at increasing economic competitiveness and efficiency, including increasing export performance.

Washington Consensus A term coined to describe the package of measures instituted in most countries of the Global South from the 1980s onward and reinforced through conditionalities attached to loans and aid. These measures are generally taken to include fiscal discipline, public expenditure priorities, tax reform, financial liberalization, exchange rates, trade liberalization, foreign direct investment, privatization, deregulation, and property rights.

Introduction

The terms north and south are commonly used to denote directions governed by the Earth's poles. North is the direction of the pole indicated by the pointer on a magnetic compass, while south is the direction of the opposite pole, or 180° from the zero point of a magnetic compass. It is conventional to have north at the top side of a map and south at the bottom, a historical reflection of northern cartographers' preferences. While such directions are important to all forms of geography, the combined and capitalized term North–South is used to refer to an imagined division between countries who share certain 'northern' or 'southern' traits. Northern countries (often referred to as the Global North) are typically assumed to be richer, more developed, and more powerful than those in the south (the Global South), which have become associated with underdevelopment and poverty. North–South also refers to the relationship between the two zones, with the Global North typically seen as exploiting or dominating the Global South through trade relationships and powerful multilateral institutions. The North–South imaginary has become a powerful concept for interpreting and organizing global processes with consequences for all scales of society.

History

North–South divisions began forming in colonial times when a rapidly industrializing Europe used its wealth and technology to extend its power and influence to other parts of the globe. Very few countries have not been colonialized by European powers at some stage of their history and even fewer could claim to have been unaffected by this period of European hegemony. The colonial experience differed across time and space, but was driven, at least in part, by a common desire to strengthen the power of the colonizing country by exploiting the labor and resources of the colonized. This was done through asserting political control, directly or indirectly, to ensure political acquiescence to European demands within particular, and often quite arbitrary, political units. Local economies were reoriented toward Europe's industrial needs by building new ports, road, and rail infrastructure to facilitate the trade of goods from mines, forests, and fields, for example, to European markets. While international trade predated colonialism, this period of change heralded the arrival of a truly global economy arranged to propel European economic growth. The economic and political divides of the colonial period were enhanced by cultural stereotyping in which northern colonizers generally considered themselves more civilized than their southern subjects, attempting to rearrange and 'improve' existing societies, sometimes referring to them as 'backward states'. While colonial, and subsequent postcolonial experiences varied widely throughout the world the geopolitical scripting and conceptual differentiation that accompanied this period of political, economic, and social domination set the foundations for global North–South divides today.

When direct European control of southern colonies weakened in the aftermath of World War II the need for foreign exchange meant that many exploitative trade relationships were maintained. However, a new period of political optimism embraced former colonial states who were able to pursue their own futures free from colonial oppression, albeit within the difficult geopolitical contours of an emerging East–West Cold War between the former USSR and the USA. The two superpowers competed to convert states to their ideologies of capitalism and communism through aid packages and trade deals as well as through covert operations and warmongering. While many ended up siding with one of the two major powers, there was also very outspoken opposition to Cold War geopolitics, stemming from strong nationalist and anticolonial agendas. This opposition contributed to the formation of particular institutions representing countries in the Global South, such as the Group of 77 in the United Nations and the Non Aligned Movement of over 100 states. Such institutions gave Southern states a distinct political identity within global geopolitics and emphasized their difference from Northern institutions, such as the North Atlantic Treaty Organization (NATO) and the Warsaw Treaty Organization. The Non Aligned Movement talked of a 'third way' and were soon known as 'Third World' nations, distinguishing them from the Western European, North American, and other 'First World' countries aligned with the USA and the Eastern European allies of the USSR in the 'Second World'.

While the world was aligning itself into separate political entities, the economic divisions inherited from colonialism continued to grow. Those in the so called First World have been the most economically successful with the period from World War II to the early 1970s sometimes referred to as the 'golden age of capitalism'. Their economic achievements have been institutionalized within powerful organizations like the Organisation of Economic Co operation and Development (OECD) and the G7 which dominate the world economy but have no Global South members. The Second World did not fare as well, with socialist economies struggling to match the capitalist efficiencies of the West. Most Third World countries looked toward the First World when building their economies, initially experimenting with Western Keynesian style approaches which propelled some (predominately Asian) economies forward, but experienced varying levels of success elsewhere. Expanding access to capital for Southern economies was facilitated by the reinvestment of petrodollars by the oil rich states of OPEC in Northern banks. However, a combination of bad loans and a global recession in the 1970s would eventually force many southern states into spiralling levels of debt. In the 1980s, as a solution to the emerging economic problems of the South, the Bretton Woods institutions (the International Monetary Fund (IMF), the World Bank, and the General Agreement on Tariffs and Trade – later to evolve into the World Trade Organization (WTO)) promoted neoliberal economic theories under what is known as the 'Washington Consensus'. Neoliberalism promotes closer North–South economic integration through free markets as the solution to growing global inequality. Opposing neoliberal approaches are neo Marxist theorists who believe that global inequality is sustained and necessitated by international economic interaction, with economic growth in the North reliant upon exploitative relationships with the South.

The growing gulf between First and Third World nations has seen the establishment of a development industry, ostensibly designed to assist allied states in other parts of the world to 'catch up' with First World standards. Development theorists have attempted to identify the underlying causes of Western success and apply these lessons to other countries. This has brought on a new way of imagining and categorizing the world according to their stages of development. While economic indicators such as gross domestic product (GDP) were originally the core focus of these analytical divisions, these have given way to a combination of social, political, and economic indicators best captured in the United Nations Development Programme's annual Human Development Reports. There has been impressive progress in many countries with falling infant mortality rates, improvements in education and social services, and longer life expectancies, all being examples of common achievements throughout the South. Ironically, however, despite making progress in diminishing North–South divides, development has simultaneously served to extenuate North–South differences in popular imaginations. Terms like 'developed' which is used for countries in the Global North, for example, and 'developing' or 'undeveloped' for most Global South countries, solidify North–South differences in geographic imaginaries. Postdevelopment writers go a step further to suggest that the terms 'developed' and 'developing' implicitly encourage people to believe that the developed North is a desirable endpoint for the developing South.

When the Cold War drew to a close in the late 1980s the political and economic divisions within the Global North lessened. The USSR and Eastern Europe collapsed and entered a period of liberal, democratic, economic, and political reforms, lessening the relevance of the

Second World concept. Russia, for example, has resolved many of its ideological differences with the USA, and is now a member of the powerful G8 (formerly the G7). With Northern tensions declining and unity growing within economic and political units the global power and influence of the North has been magnified. Consequently, North–South has replaced East–West as the key tool for interpreting global economic and geopolitical processes.

While ostensibly the global order has become less complex with the end of the Cold War, there remain questions about the extent to which previously used def initions remain appropriate. In particular, there is con tention about whether the countries of the ex communist bloc, sometimes referred to as transition economies, should be now classified in the same way as the erstwhile Third World countries or whether they require a different kind of characterization. Certainly, some transition econ omies now exhibit human development indicators that are of a similar level to those of the old Third World, and it can be argued that in a generic sense they face similar issues. Others suggest that the Asian and European tran sition economies are undergoing quite different kinds of transformations which demand that we analyze these re gions differently.

The Brandt Report

The first significant reference to North–South was in a report chaired by Willy Brandt for the Commission on International Development Issues in 1980. Entitled *North–South: A Programme for Survival*, but commonly known as the Brandt Report, it sought to explain global patterns of inequality, in which the Northern Hemi sphere countries were portrayed as much more economically and politically powerful than their South ern Hemisphere counterparts, and propose means for overcoming these divides. On the cover of the report a line was drawn, now known as the Brandt line, between countries included in the Global North and those in cluded in the Global South (see **Figure 1** which uses the Peters projection which some believe more accurately represents the relative size of tropical countries than the more common but Eurocentric Mercator projection). The report used social and economic inequality mea sures to describe the North–South divide, claiming, as examples, that 75% of the world's population lives in the South but they only earn 20% of the world's income; on average the life expectancy in the North is around 70 years, 20 years more than that in the South; almost everyone in the North receives secondary schooling, whereas 50% of the South are illiterate; 90% of the world's manufacturing is in the North and they control the vast majority of the world's patents and intellectual property rights. Using the analogy of a global community in which the stronger members have a responsibility to assist weaker members, and arguing it is ultimately in their benefit to do so, the report lists a wide range of recommendations in a proposed program of action.

An important component of the Brandt report was a rejection of some of the assumptions underpinning earlier geopolitical imaginaries regarding directions of growth. In a rejection of the linear pathways implicit in First/Third and developed/developing conceptual izations of progress the report writes

> We must not surrender to the idea that the whole world should copy the models of highly industrialised coun tries… Ideologies of growth in the North (and not just the West of the North) have had too little concern for the

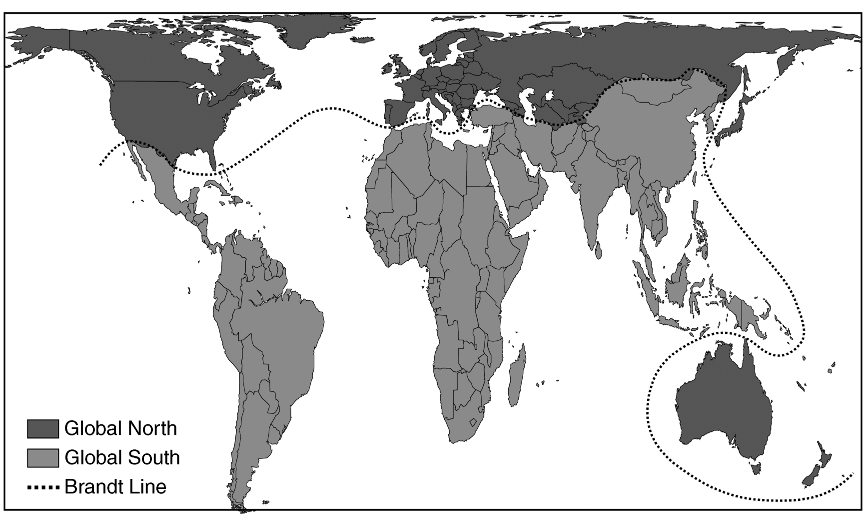

Figure 1 North South map (using Peter's projection) showing Brandt line. Modified from Brandt, W. (1980). *North-South, A Programme for Survival: Report of the Independent Commission on International Development Issues*. London, Pan Books.

> quality of growth… There is no uniform approach; there are different and appropriate answers depending on history and cultural heritage, religious traditions, human and economic resources, climatic and geographic con ditions and political patterns of nations… . (Willy Brandt, 1980: 23 24)

Unlike earlier development approaches that position Western countries as models which other countries should seek to emulate, the terms Global North and Global South are much more neutral and do not confer superiority to the North. Instead, they are used as ana lytical tools to identify trends and patterns of global in justice, exploitation, and inequality. In this sense, the term is less loaded than those from which it evolved, such as Third World, underdeveloped, and backward lands and as such has the potential to be a source of positive identity in renegotiating positionalities within emerging geopolitics and global economies.

Criticisms

A number of criticisms can be directed at the North–South conceptualization. There are geographical prob lems inherent in the original Brandt line, for example, not just in looping Australia and New Zealand into an im aginary North, but also by accidentally capturing a whole range of additional Pacific Island countries into the North in the process. Notwithstanding the sensitivity to difference intended by Brandt (quoted above), the North–South conceptualization can also be criticized for contributing to a lack of awareness about diversity. Lumping countries as diverse as Sierra Leone and Singapore into the same bloc, or China and Kiribati, risks blurring the very apparent distinctions between such different countries and deflects attention from inequality and exploitation within South–South or North–North relations. That the Asian Tiger economies and wealthy oil producing nations are still commonly grouped within the South, despite having comparable or superior in comes to those in the North, also raises concerns whether the Brandt line is now more reflective of global ethni cities and cultures than economies and politics.

Spatial divisions also risk diverting attention away from exploitative relations within countries, particularly class based relations which may be more pertinent cat egories for understanding power, exploitation, and wealth. Mobile wealthy classes from around the world, for example, share similar lifestyles irrespective of their spatial location, as do more spatially embedded industrial working classes. Further concerns regarding labor mo bility in the form of South–North legal and illegal mi gration and the vast sums of money sent back to the Global South in the forms of remittances also problematize simple spatial divisions. Finally, the global shift toward economic regionalism accentuates divides within North or South, such as the distinction between the European Union (EU) and the North American Free Trade Alliance (NAFTA), and blurs boundaries between North and South, with countries from both regions represented in organizations like the forum for Asia Pacific Economic Cooperation (APEC).

A key concern then is that the term Global South promotes an understanding that privileges the nation state as the main focus of analysis. This tends to limit understanding of the evolving economic interactions between different cities and regions. In this sense, the concentration on differences between countries risks overlooking concomitant changes taking place in the global economic system, as flows between nodes of the Global South increase and cities and regions become more intensively connected to each other. Instead, Taylor *et al.* argue that we can best understand this increasingly global economic interconnectivity by utilizing a re lational approach that focuses more on the interaction of different cities and regions than sovereign spaces.

Key North–South Issues

Despite such criticisms, North–South has become an important organizing concept for interpreting inter national economic and political relations, providing a useful lens for identifying unevenness and inequality at the global scale. The following sections explore how the North–South concept can be applied to reveal divisions and ongoing challenges in political, economic, socio cultural, and environmental spheres.

North–South Politics

At the global scale, many of the world's most powerful institutions either exclude Southern states from mem bership or marginalize them in decision making. The G8 and the OECD, for example, are comprised of countries that dominate the world's economy but have no repre sentatives from the Global South. Other institutions like the United Nations, with a mandate to prevent war, protect human rights, establish international laws, and promote progress and freedom, were consciously set up with principles of equity in mind, yet even there Northern states dominate key components. The UN General Assembly, which determines matters of policy and overviews other UN agencies, is the most democratic structure allocating one vote per state irrespective of population size or economic power. The more powerful UN Security Council, however, with a mandate to maintain international peace through sanctions and peacekeeping operations, is far less democratic. The

Security Council is made up of five permanent members, that is, the USA, France, Russia, the UK, and China, who have the right to veto any decision, and ten rotating members. China is the only permanent member from the Global South despite (or perhaps reflected in) the vast majority of Security Council Resolutions being directed toward Southern states. The current push by Japan, Germany, India, and Brazil for permanent representation only goes part way to rectifying this imbalance. Similarly, other powerful global institutions like the World Bank and the IMF are dominated by Northern states with voting rights allocated according to funds contributed. The North provides the bulk of funding to these insti tutions, thereby having a disproportionate amount of influence on policy despite the impacts of such policies being felt most intensely in the South.

North–South Economics

Troubling economic North–South disparities have been the focus of the development industry since World War II. While conditions have changed since the original Brandt report, much manufacturing has moved to the South, for example, and basic incomes have generally improved, Northern economies have generally out performed those in the South. A variety of theories have evolved to overcome these divisions, such as mod ernization, neo Marxist, grassroots, and neoliberal the ories, yet economic disparities, with a few important exceptions, continue to grow. Some of the contemporary economic North–South challenges include foreign aid, international debt, and agricultural subsidies.

Foreign aid

Foreign aid is one of the key mechanisms, along with foreign direct investment, through which the Global North transfers funds, knowledge, and expertise to the Global South. While such transfers may have underlying geopolitical motives, clearly seen in Walt Rostow's in fluential text on foreign aid *The Stages of Economic Growth* being subtitled *A Non Communist Manifesto*, they are os tensibly designed to lessen North–South economic and social divisions. Some common concerns, however, in clude that the Global North is not providing enough foreign aid, that the aid is not going to the right places, and that has become tied to particular undesirable con ditions. The criticism that the Global North is not pro viding enough aid derives from a UN General Assembly agreement from 1970 when Northern nations agreed to commit 0.7% of gross national income (GNI) to overseas development assistance (ODA). With a few notable ex ceptions – Denmark, Luxembourg, Netherlands, Norway, and Sweden (see **Figure 2**) – the vast majority of Northern countries have consistently reneged on this agreement with a country average, in 2004, of only 0.42% of GNI being committed to ODA. The direction of for eign aid is criticized for reflecting geopolitical interests rather than economic or social needs; strategically im portant Israel, for example, regularly receives more for eign aid from the USA than more impoverished Southern nations such as Sierra Leone. 'Tied aid' refers to the leverage Northern donors exert when allocating aid, re quiring Southern partners to agree to certain conditions, such as undergoing particular institutional reforms or purchasing certain Northern products and/or services with the aid money, before the aid program will go ahead.

International debt

One of the biggest obstacles to overcoming economic inequality is the international debt of Southern countries. Debt repayment has become a core expense, often at tracting more governmental financial support than social investment in education or healthcare. There are various reasons behind Southern debt, including bad loans for failed initiatives; 'odious debts' in which money was lent to corrupt elites who profited privately while their country inherited the debt; increasing costs associated with oil price rises in the 1970s; falling market prices particularly during global recessions; and the devaluation of currencies where some debts have been repaid in the local currency and not in the foreign currency in which it was borrowed. An international campaign by NGOs to provide debt relief to poor countries has succeeded in convincing the IMF to provide debt relief packages to a selection of heavily indebted poor countries (HIPCs). However, such relief packages are 'tied' to particular conditions, most notably that HIPC qualifying countries must undergo a structural adjustment program (based on an IMF/World Bank Poverty Reduction Strategy Paper) which encourages privatization, currency devaluation, and greater levels of foreign investment. The consequent privatization of basic commodities like water has caused suffering and widespread protests in some states, causing many to criticize the IMF's policies and practices as re flecting the investment concerns of the Global North, rather than the human rights of the Global South. Until debt issues are resolved in the South, debt repayments will continue to stifle social and economic investment in more productive areas of society.

Agricultural subsidies

A core conflict in North–South economic relations is borne out in the debates over agricultural subsidies. Southern countries, which are generally more eco nomically reliant upon the agricultural sector, are argu ing through the WTO, which is based on consensus decision making, that the Global North, particularly the EU and the USA, should abolish the subsidies they pay to their own farmers. Agricultural subsidies in the North are enormous, exceeding the amount they give as ODA,

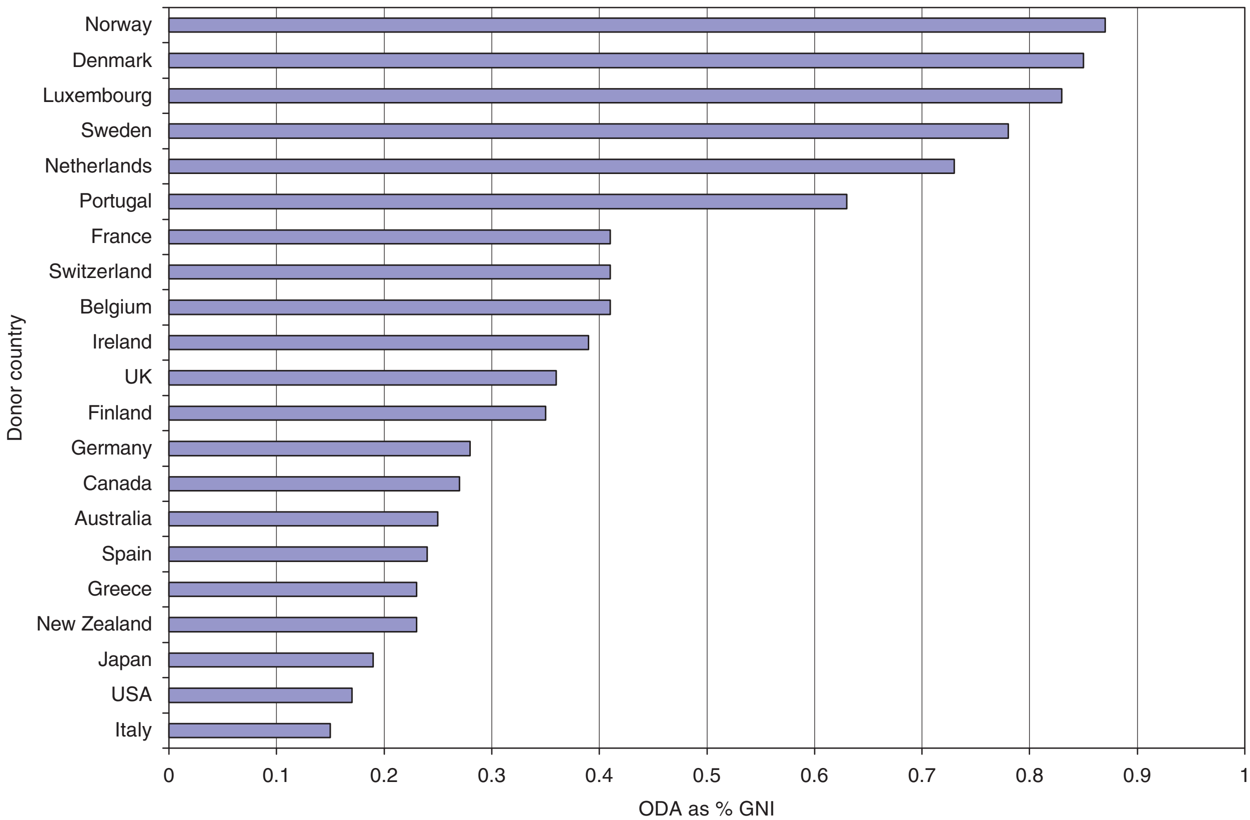

Figure 2 ODA as ratio of Gross National Income (GNI) of OECD donor countries. Source: Management Committee for the Reality of Aid (2006). *The Reality of Aid 2006: Focus on Conflict, Security and Development*. Quezon City: Ibon Books.

protecting uncompetitive Northern industries from Southern producers. Not only are Northern markets protected but the subsidies allow the North to export to the South at prices that local farmers cannot compete with. Such policies go against the neoliberal free market economic theories the North promotes in the South through aid and debt relief packages, adding to Southern frustrations. South–South unity has frozen negotiations within the Doha round of talks at the WTO but true agricultural reform within the North looks to be still some way off.

North–South Societies and Culture

Two key North–South sociocultural issues relate to international labor movements and debates about glob alization and glocalization.

International labor movements

International labor movements between North and South are carefully controlled by immigration and border protection agencies. Unlike capital, labor is not permitted to freely cross borders; instead, only carefully selected flows are allowed to take place. Due to the improved economic and social living standards in the North there are considerable pools of labor that would like to migrate South–North. The North chooses which of these workers will be allowed to migrate and under what conditions, ensuring only the most appropriate applicants succeed. The preference for skilled workers raises concerns for Southern nations who may have invested considerable capital into their education and training, only to lose them during their most productive years. Known as the 'brain drain', the Northern country inherits a skilled, educated worker for little or no investment. Some Southern countries, like the Philippines, for example, actively promote these international labor movements as the remittances sent home by migrants can contribute substantial amounts to domestic economies. The emi gration of the 'best and brightest', however, is generally thought to depress Southern development.

Unapproved or illegal South–North migration is also common. People may flee their home country for fear of persecution or conflict, while others may simply be at tracted to the higher wages of the North. Illegal migrants often fill important gaps in the labor market by finding low paying jobs most Northerners find undesirable. However, illegal immigrants risk their lives and freedom when crossing Northern borders, particularly when being smuggled in overloaded or unsafe boats. Many immi grants are caught and detained or returned to their country of origin where they can be imprisoned. Others, if escaping from persecution may apply for refugee status which, depending on the policies of the Northern state, can lead to confinement in refugee camps, sometimes for years at a time, or freedom to move and seek work while their claims are processed. As North–South inequalities spiral, it is likely that legal and illegal forms of immi gration will continue to grow.

Globalization and glocalization

A common argument today is that sociocultural North–South distinctions are lessening as a consequence of globalization. Across the world, technological develop ments are compressing both time and space and bringing geographically distant cultures into closer contact than has previously been possible. Globalization proponents believe this is leading to a dilution of cultural diversity as a homogeneous global culture begins to emerge. How ever, rather than this global culture reflecting the diver sity of societies that exist across the world, globalization favors those that have the most sophisticated and eco nomically powerful cultural industries – such as in film, media, fashion, literature, or sport. While there is evi dence of cultural traits from the South being represented in this global culture – particularly through food indus tries, for example – most cultural industries at the global scale are dominated by those in the North. Hence, there is concern that Southern cultures, in particular, are being lost as new generations are attracted to the lifestyles promoted by the Global North.

While evidence of globalization can be found throughout the world, a strong counterview known as glocalization has also emerged. This perspective agrees that time and space are being compressed, but, rather than locals mindlessly substituting global cultures for their own, they are reinterpreting and incorporating them in locally unique and useful ways. Glocalization suggests that hybrid cultures that are emerging integrate the global and the local in place specific ways contributing to greater vibrancy and heterogeneity rather than homo geneity. Such a viewpoint positions the Global South as active rather than passive agents selectively engaging in global trends and processes according to their own cul tural norms. The film industries of Bollywood in India and Nollywood in Nigeria are good examples of gloca lization where the Global South has engaged with Northern technologies in culturally unique and expres sive ways.

North–South Environments

North–South relationships extend beyond social con cerns into physical and biological systems. Two key areas of concern relate to global environment management as well as concepts of environmental justice.

Global environmental issues

Global environmental issues are those which are beyond the scope of a single state to address and instead require concerted action from the global community. Some ex amples derive from sharing a common global resource, such as global warming or ozone depletion, while others emerge from the cumulative impacts of individual state scale acts of degradation – such as global biodiversity loss, fisheries depletion, or deforestation. It is generally accepted that the Global North is causing more en vironmental degradation than the Global South due to the North's higher rates of consumption. The resources consumed by one urbanized North American, for ex ample, may be hundreds of times that of a rural family living in a remote African or Asian community. As Southern nations develop, however, consumption rates are increasing therefore adding to the pressure being put on the global environment. It remains to be seen, for example, whether the earth is capable of sustaining a global population that consumes resources at the rate of the Global North. Such concerns raise interesting di lemmas about how to reduce Northern consumption levels as well as the direction of Southern development. These concerns are borne out most vividly in con temporary debates regarding global warming with most Northern nations signing the Kyoto Protocol and making a commitment to reduce their emissions. States in the Global South, however, have been reluctant to make such commitments believing conservation prior ities are of secondary importance to economic development.

Environmental justice

Environmental justice, in the context of North–South, refers to concerns that environmentally damaging or polluting industries or processes are being transferred from North to South. Within increasingly globalized trading systems, more and more products that are con sumed in the North are being sourced from the South. This is partly a reflection of cheaper and more accessible land and labor in the South as well as increasingly stringent environmental regulations in the North. The majority of the world's paper, for example, is consumed by the North but much of it derives from forests in the South causing local environmental degradation. Simi larly, the surrounding and downstream pollution of waterways near mining sites in the South often derives from minerals destined for Northern markets. In effect, much of the environmental damage associated with Northern consumption patterns have been placed out of sight of consumers and relocated to Southern locations. Awareness of North–South environmental justice issues is growing through southern activism, consumer aware ness campaigns and labeling of eco friendly products, yet there is still much to be done.

Summary

The Brandt line dividing the world into North and South is very much an arbitrary division replete with problems, yet it has become a useful tool for visualizing, con ceptualizing, and analyzing global processes, injustices, and inequalities. The North–South imaginary reveals stark differences in the wealth and power of those in the Global North and those in the Global South. In doing so, it has provided a rallying point for Southern activists interested not necessarily in ‘catching up’ or becoming more like the North, but instead in ensuring that basic human standards are achieved in their own countries, enabling them to pursue their own paths into the future.

See also: Debt; Geopolitics; Liberalism.

Further Reading

Brandt, W. (1980). *North South, A Programme for Survival: Report of the Independent Commission on International Development Issues*. London: Pan Books.
Escobar, A. (1995). *Encountering Development: The Making and Unmaking of the Third World*. Princetown: Princetown University Press.
Management Committee for the Reality of Aid (2006). *The Reality of Aid 2006: Focus on Conflict, Security and Development*. Quezon City: Ibon Books.
Peet, R. (2003). *Unholy Trinity: The IMF, World Bank and WTO*. London: Zed Books.
Power, M. and Sidaway, J. (2004). The degeneration of tropical geography. *Annals of the Association of American Geographers* 94(3), 585 601.
Rostow, W. (1960). *The Stages of Economic Growth: A Non Communist Manifesto*. Cambridge: Cambridge University Press.
Shu Yun, Ma. (1998). Third world studies, development studies and post communist studies: Definitions, distance and dynamism. *Third World Quarterly* 19(3), 339 356.
Sidaway, J. (2000). Postcolonial geographies: An exploratory essay. *Progress in Human Geography* 24(2), 591 612.
Taylor, P., Derudder, B., Garcia, C. and Wiltox, F. (2006). From North South to global south: A geohistorical investigation using airline routes and travel, 1970 2005 GaWC Research Bulletin 209. http://www.lboro.ac.uk/gawc/rb/rb209.html (accessed June 2008).
UNDP (2006). *Human Development Report 2006: Water and Human Development*. Basingstoke: Palgrave MacMillan.
Willis, K. (2005). *Theories and Practices of Development*. London: Routledge.

Relevant Websites

http://www.globalpolicy.org
Global Policy Forum website Updated statistics relating to global economic and political inequality.
http://www.realityofaid.org
Reality of Aid website. Annual updates and current information on issues relating to the provision of foreign aid from OECD nations.
http://www.brettonwoodsproject.org
The Bretton Woods Project website review of critical perspectives on the World Bank, IMF and WTO policies.
http://www.undp.org
United Nations Development Programme website For annual Human Development Reports and updates on other development issues.
http://www.worldbank.org
World Bank website portal to one of the world’s most powerful and controversial organizations.

O

Oceania

W. E. Murray, Victoria University of Wellington, Wellington, New Zealand

Introduction

Oceania refers to the region comprised of the subregions of Melanesia, Polynesia, and Micronesia (see **Figure 1**). From time to time, Oceania is referred to as the Pacific Island region, or on occasion the South Pacific Islands. The former is too broad to be meaningful, and the latter is misleading as many of the nation states, territories, and other entities that comprise Oceania are, in fact, found north of the equator. Oceania sometimes also refers to the region of Australasia, defined here as Australia and New Zealand. Although New Zealand is sometimes considered an Oceanic nation, given that the indigenous Maori and Moriori are Polynesians, New Zealand is considered for purposes here to be part of the region of Australasia.

Here, Oceania is defined as comprised of 22 territories, of differing political status, and spanning the three regions of the Pacific. These include Melanesian territories of Fiji, New Caledonia, Papua New Guinea (PNG), Solomon Islands, and Vanuatu; the Polynesian territories of American Samoa, Cook Islands, French Polynesia, Niue, Pitcairn Islands, Samoa, Tokelau,

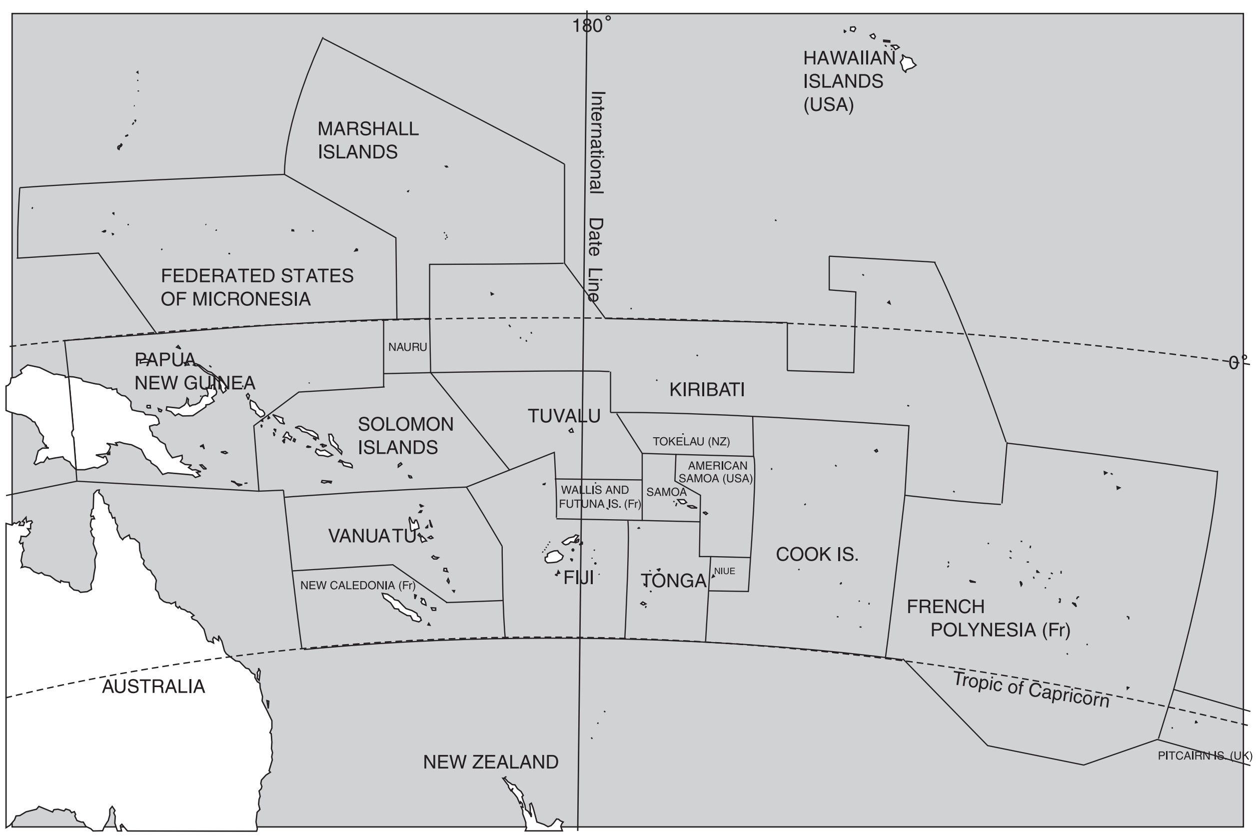

Figure 1 Oceania as referred to the region comprising the subregions of Melanesia, Polynesia, and Micronesia.

Tonga, Tuvalu, Wallis, and Futuna; and the Micronesian territories of Federated States of Micronesia, Guam, Kiribati, Marshall Islands, Nauru, Northern Mariana Islands, and Palau. Arguably, Hawaii, a full state in the USA, can also be included, as can Rapa Nui (Easter Island) most recently colonized by Chile.

Oceania is both the largest and the smallest region on the Earth depending on how it is measured. In terms of territorial area it stretches across much of the expanse of the Pacific Ocean. In terms of the landmass of those is lands that inhabit the region, it is small by global standards. The sea (measured in terms of the size of the relevant exclusive economic zone (EEZ)) to land ratio of the various territories and countries is as high as 29 000 in the case of Tokelau, for example, and most countries' territorial jurisdictions are over 99% water (see **Table 1**). This combination of small landmasses scattered across a massive ocean gives the region a very special geography, which yields unique and often challenging geographical problems in physical, economic, and political senses.

The total population of the 22 territories of Oceania as defined above was estimated as approximately 9.2 million in mid 2006 (see **Table 2**). Of the subregions, Melanesia is the most populous with close to 8 million inhabitants, the majority of these located in PNG with a population of approximately 6.2 million. Micronesia and Polynesia have estimated populations of 542 000 and 648 000, respectively. The figures hide significant vari ations ranging from the larger countries such as PNG, Fiji (830 000), Solomon Islands (490 000), and French Polynesia (260 000), to territories such as Niue (1600), Tokelau (1400), and Pitcairn (50). Population densities also vary widely, being relatively low in Melanesian territories, and generally very high by global standards in Micronesia and parts of Polynesia. Nauru and Tuvalu stand out in the latter regard with densities of 482 and 371 per km^2, respectively, in 2006. Population growth rates are highest in Melanesia due largely to high rates of natural increase, between 2006 and 2010, for example, the annual rate of growth is projected at 2.0%, representing a doubling time of approximately 35 years. In the more densely populated territories, growth is generally lower. This is partly due to out migration, although this is not a simple function of density pressure, and turns on the access of entry to Pacific Rim countries where Poly nesians and Micronesians might seek to move. In the case of Niue, Tokelau, and Cook Islands, for example, con stitutional links with New Zealand mean that migration is unrestricted. This is quite different in the case of the territories governed in association with the USA, where migration intake is closely rationed.

Despite its relatively small population and territorial area – or perhaps because of it – it is enormously complex in terms of human geography, history, economy, culture, and politics. For convenience it is often divided into the three subregions already mentioned – Micronesia, Polynesia, and Melanesia. These subregional classifi cations often hide more than they reveal, placing together a large range of cultures, histories, and societies. Although those in Polynesia, and to a lesser direct extent, Melanesia can be demonstrably linked in ethnic terms, this is much less than the case in Micronesia. Coined by a French navigator D'Urville in the early nineteenth century, together with the term Oceania, the subregional ideas are Western imaginaries used only since colonial times in order to classify and order these vast and diverse spaces.

Table 1 Land and sea areas in selected territories of Oceania

Territory	*Land mass (km^2)*	*EEZ (million km^2)*	*Sea to land ratio (%)*
Cook Islands	236	1.8	7267 (99.987%)
Fiji	18 272	1.3	71 (98.594%)
FSM	702	3	4274 (99.977%)
Kiribati	762	3.5	4821 (99.979%)
Marshall Islands	181	2.1	11 602 (99.991%)
Nauru	21	0.3	14 286 (99.993%)
Niue	259	0.39	1506 (99.934%)
Palau	488	0.629	1289 (99.922%)
PNG	462 840	3.1	7 (85.070%)
Samoa	2857	0.1	35 (97.143%)
Solomon Islands	29 785	0.6	20 (95%)
Tokelau	10	0.29	29 000 (99.997%)
Tonga	747	0.7	937 (99.893%)
Tuvalu	26	0.7	26 923 (99.996%)
Vanuatu	12 200	0.6	49 (97.9%)

FSM, Federated States of Micronesia; PNG, Papua New Guinea.
Sources: UNDP (2008).

Unique Geography and Distinctive Challenges

The physical geographical differences between the sub regions have partly determined cultural outcomes. Comparatively, elsewhere the region, Melanesia is com prised of relatively high volcanic islands with rugged terrain, leading to a situation where historically small groups of people settled in relative isolation from each other. Thus, the dispersal of migrants across the sub region was relatively slow. As such, Melanesian culture is significantly internally heterogeneous with, for example, hundreds of mutually unintelligible languages and dia lects. The Polynesian landscape in contrast is comprised of smaller volcanic islands and low lying atolls. The former tend to be fertile and have supported relatively large and prosperous cultures (as in Samoa and Tonga for example). In the case of the latter, however, relatively rapid out migrations in the face of scarcity appear to

Table 2 Population in Oceania, 2006 2015

Country/subregion	*Population 2006*	*Population density 2006*	*Population growth rate*	*Population projection 2015*
Melanesia	*7 965 061*	*15*	*2.0*	*9 496 691*
Fiji	831 263	45	0.7	891 729
New Caledonia	238 035	13	1.5	270 456
PNG	6 187 108	13	2.1	7 461 903
Solomon Islands	487 237	17	2.4	594 543
Vanuatu	221 417	18	2.6	278 059
Micronesia	*541 937*	*169*	*1.4*	*610 878*
FSM	110 218	157	0.5	116 071
Guam	167 371	309	1.1	183 609
Kiribati	93 706	116	1.9	110 499
Marshall Islands	55 981	309	1.0	60 149
Nauru	10 131	482	0.3	10 585
Northern Marianas	84 487	179	3.2	108 939
Palau	20 044	41	0.5	21 027
Polynesia	*648 073*	*81*	*0.9*	*700 435*
American Samoa	63 308	317	1.4	72 118
Cook Islands	13 572	57	2.0	11 543
French Polynesia	258 709	73	1.2	285 977
Niue	1591	6	2.3	1307
Pitcairn Islands	50	1		
Samoa	185 234	63	0.8	199 537
Tokelau	1398	117	0.5	1373
Tonga	99 298	153	0.3	102 834
Tuvalu	9652	371	0.1	9681
Wallis and Futuna	15 260	107	0.6	16 066
Total	9 155 070	17	1.9	10 808 004

Source: SPC (2007).

have characterized historic settlement patterns. Thus, Polynesian culture is much more internally homogeneous due to rapid dispersal made possible by astounding maritime prowess, given the low level of technology available. It is possible, for example, for a Rapa Nui speaker (i.e., from Easter Island) to understand New Zealand Maori, for example, despite the thousands of kilometers that separate their territories. The relationship between environment and culture in Oceania remains little understood, though most now agree that the environment,while not determining cultural outcomes, has always tended to set distinct limits on the possibilities.

The region's unique geography creates special challenges in terms of the interaction with the markets, polities, and societies of the world beyond. There are at least five geographical challenges that relate to the interface of the physical and human geography of the area.

Territorial

The territories of the Oceania region are isolated; this raises particular problems for economic development related to the frictional effects of distance. Although technology exists to make trade rapid and efficient, in such circumstances it is often expensive, and thus weighs against the small island territories. Most air traffic in the Oceania region flies over the islands, and there are, in a relative sense, very few direct flights in and out of the zone, compared to those that fly around and over it.

Environmental

The Oceania region is highly vulnerable to both extreme climatic conditions and climate change. In the case of the former, many of the territories lie within the cyclone belt and run the risk of being devastated by hurricane force weather systems. When such systems actually touch the ground they have the capacity to cause absolute economic havoc as was evidenced by the almost total destruction of national infrastructure and agricultural crops on the island of Niue following Hurricane Heta of 2004. In many ways the small island nations and territories of the Pacific are paying the ultimate price for the consumption activity of the West. Rising sea levels are causing potentially major problems across the region – especially on the low lying atoll island groups, such as Kiribati and Tokelau. Floods, salinization, and the increased incidence of cyclone activity are all consequences of increased sea temperature and sea level rise.

Political

Given the relatively small size of the national economies and polities, the Oceanic nation states and territories generally have low bargaining power internationally.

There have been numerous attempts at regional cooperation, such as the formation of the Pacific community in 1947 in order to get around this article. Consensus has been rare and even with success the total size of the bargaining unit would be low. Notwithstanding, the Oceanic nations do control vast amounts of territories and in some cases natural resources that have been used in order to win allegiance and political favor. High levels of aid from Taiwan and China recently can be explained in terms of access to fishing rights. Furthermore, given that Oceanic nations have voting rights in the United Nations and other for it has often been the case that the islands have been courted in order to win political concessions and votes – as evidenced in the recent diplomatic and aid assault undertaken by Japan in order to win 'yes' votes in the international whaling commission.

During the Cold War, the island nations had a particular geopolitical comparative advantage, and aid from the allies of the US was pronounced. Following the end of the Cold War, the US withdrew from the region largely, even pulling out the Peace Corp. In more recent times the larger neighbors on the rim have begun to engage politically with the island nations again, but this time in pursuit of free markets and democratic reform – although this has been occasionally dressed up as 'security' or 'counter terrorist' interaction. Notwithstanding this, the Oceanic territories have been policy takers rather than makers and this is direct result of the colonial legacy and diminutive size among other things.

Economic

Due to small economic size there are limited economies of scale and low diversity. In the case of the former this makes production, particularly industrial production, extremely difficult and there have been very few successful attempts at industrialization – the exception being Fiji in the 1980s during a phase marked by import substitution strategies in industrial agriculture and export orientation industries, including garments and fish canning. The industrial pathway to development has been largely abandoned on the islands due to this and other infrastructural and labor force factors, in favor of a mixture of agricultural/ocean based exports, tourism, aid, and remittances. As is discussed later, this varies from place to place in terms of its exact make up. There are very few examples of large scale agricultural production units that have been successful, with the most notable exception being the Fiji sugar industry – which has itself survived only due to favorable conditions granted under the Lomé and then Cotonou agreement as a favored postcolonial nation within the commonwealth. Mining has played an important role in certain economies including Fiji, PNG, and to a lesser extent, Solomon Islands. Fishing is important in most territories and of particular importance of late has been the sale of flags of convenience so that third parties might fish the waters, and this has often been linked to aid diplomacy. Tourism then has become the main productive economic activity (aid and remittances aside) rising to become the major earner in many of the territories and countries. Much of this tourism is foreign owned and the largest earner is resort based tourism, which although bringing employment and foreign earnings is often enclave in nature and the profits flow out to overseas companies, tax concessions are often very generous. Thus, the earnings that stay onshore can be negligible.

Factor Endowment

As already mentioned, there are limited land based resources on many of the islands. In Melanesia, which due to the volcanic geomorphic roots is largely comprised of large, raised islands and there are mineral deposits and significant land area for agricultural production. As such, mining and plantation agriculture has taken place in some of these areas. However, the outcomes of resource endowment of this nature have not always been positive, and particularly in the cases of PNG and Solomon Islands, with respect to gold and copper mining, it is possible to discern something of a resource curse in terms of social, economic, and environmental outcomes. In the low lying atoll islands the problem is most certainly the reverse in that there are no land based resources. In all cases there are significant ocean resources, although the technology does not always exist to exploit these.

While the constraints of territorial geography are being altered, and in some cases eroded, these geographical factors most certainly collude to present a significant challenge to the Oceania region. Some have argued that given these intrinsic factors the problems of the islands are insurmountable. Right wing commentators have also added the 'problems' of traditionalism and communalism to the list of Pacific characteristics, which they argue make economic development tantamount impossible. However, it is the region's role at the margins of the global political economy and the nature of its insertion into globalizing processes, and the conflict that this has often perpetuated that explains the poor relative position of the region in terms of welfare and development prospects. This peripheral role was cemented through colonialism and many of the processes that have taken place since have further marginalized the area. This is not to say that the source of all of the region's problems is external – this is far from the case, internal conflict and ineffective governance have also played a major role – but it is the case that the root of these problems can often be traced back to the marginal position that the region and its constituent territories occupy in the global political economy. It is the rela tionship of the region to the unfolding of globalization that is the major explanatory factor – an unashamedly structuralist view on the way that the region has fitted into the wider world over time and how this has con ditioned outcomes on the ground in the Oceania region. This is considered below in terms of the various waves of globalization that have washed across the region.

Historical Geography – From Indigenous Settlement to PostColonialism

Of all the regions on Earth, Oceania is one of the least understood. It is often not visible in atlases of the world and is left out in discussions concerning the 'Asia–Pacific' region. To many outsiders, then, it is often constructed as a paradise and has been 'orientalized' in the Western imagination. It is thus often assumed that the Pacific – derived from the Latin word for peace named by the explorer Magellan as he entered the unusually calm waters off the west coast of the Cape Horn – is a haven of tranquillity, little touched by the troubles of the outside world and processes that comprise economic and polit ical globalization. This could not be further from the truth – the region has been shaped by interaction with the 'outside' over centuries and its interaction with the outside world has altered it immeasurably.

The origins of human settlement in the Pacific region, and the subsequent patterns of dispersal are sources of considerable debate. It is known that a first cohort of non Austronesian speaking peoples arrived around 33 000 before the present (BP) in what is now Australia and PNG from Southeast Asia. A group of Austronesians are believed to have moved across to what is now known as Melanesia some 4000 BP, arriving in Fiji approximately 3000 BP. Around this time others joined those from further north on the Pacific Rim in the settlement of Micronesia – evidence suggests that parts of Micronesia were populated in part from the north around 5000 BP. Another group began the population of what became Polynesia, arriving in Rapa Nui around 2500 BP and Hawaii approximately 1500 BP. It is believed by some that Polynesians may have crossed to South America. Polynesians arrived in the Cook Islands around 1200 BP and Aotearoa/New Zealand approximately 1000 BP – their final destination. European contact was initiated in the early sixteenth century, but was sporadic. The period of direct interaction, involving missionaries, traders, and whalers, began in the late 1700s. European colonization and administration did not begin until after the mid 1800s. Although settlement by Europeans in numerical terms was relatively insignificant, they did bring rela tively large populations of other ethnicities to work in the islands, most notably indentured Indian labor to Fiji. As we explore below, however, the sociocultural, economic, and environmental impacts of the European influence were profound.

Waves of Globalization

The colonial history of the Oceania can be divided into two waves of globalization with a further more inward oriented period in between. In this sense globalization is understood to refer to the expansion of capitalism from the Western core. Arguably, the pre European waves of colonization by the Melanesians and Polynesians in particular represented precapitalist forms of global ization. Here, however, it is argued that it is the insertion of the region into capitalist circuits that has largely determined the broad political economic history of the region subsequently. As is the case everywhere, 'local' characteristics have interacted with wider global forces to produce a unique outcome. Notwithstanding, Pacific economic underdevelopment is best understood through an analysis of the way the island region fits into the wider, and increasingly internationalized and globalized, world:

1. colonial globalization – the first wave (1830s–1960s);
2. migration, aid, remittence, and government employ ment bureaucracy (MIRAB; 1960s–1980s); and
3. postcolonial globalization – the second wave.

Colonial globalization

Colonialism came relatively late to the Pacific Islands compared with much of Asia and Latin America, for example. Contact with Europeans in the years leading up to the early 1800s had been largely through missionaries, whalers, and traders. This initiated the roll out of certain Western concepts and early constructs of progress and civilization, including the use of money and Christianity. The South Pacific became an important geopolitical area given the competition among maritime powers toward the end of the 1700s and during the early 1800s. Rivalry was especially fierce between France and Great Britain and this was played out in the ocean. Early European involvement in the region then was based on some eco nomic possibilities but became strategic as time went on. Formal colonization began sometimes in the mid 1800s (see **Table 3**), as there were calls for law and order in the region. This was not a simple case of European power taking possession of well defined lands in terms of ownership; in some cases, different indigenous groups played each other off through their allegiances to the Europeans. In the case of Fiji, for example, Ratu Cakobau ceded the crown in order to gain protection from the French and in order to gain backing in his conflict with

Table 3 Colonization and independence in Oceania

Colonial power	*Incorporated into colonial power*[a]	*Dependent colony/ territory*	*Self-governing in free association*	*Independent*
Australia		Norfolk Island (c1788)		Nauru (c1920, i1968) Papua New Guinea (c1884, i1975)
Chile	Rapa Nui (Easter Island) (in 1888)			
France		French Polynesia (c1904) New Caledonia (c1853) Wallis and Futuna (c1842)		
New Zealand		Tokelau (c1916)	Cook Islands (c1901, sg1965) Niue (c1901, sg1974)	Samoa (c1900, i1962)
UK and France				Vanuatu (c1906, i1980)
United Kingdom		Pitcairn (c1838)		Fiji (c1874, i1970) Tonga (i1970)[a] Tuvalu (c1916, i1978)[b] Solomon Islands (c1920, i1978)[b] Kiribati (c1916, i1979)[b]
USA	Hawaii (c1898, in 1959)	American Samoa (c1899)	Marshall Is (c1885, sg1986)	
		Guam (c1898)	FSM (c1885, sg1986) Palau (c1885, sg1994)	

[a] Tonga was a protectorate of the United Kingdom from 1901, not a fully fledged colony.
[b] These countries were protectorates for approximately 20 years before formal colonialism. In many cases the colonial power changed hands once or twice (especially after WW1 when Germany lost its possessions) the latest colonial power is reported here.
Source: Adapted from Murray (2006). (i1970) = date of independence. (c1874) = date of colonization. (sg1986) = date of self government. (in 1888) = date of incorporation.

other chiefs of Fiji. Largely, the colonial territories were not settled extensively and were governed from capital cities in Europe. This first wave of globalization brought the Oceanic countries into the circuits of capital dominated by mercantilist forces as very much peripheral territories. Firth has argued:

> (G)lobalisation when combined with colonial rule, meant incorporation into the global economy on terms that suited the interests of the colonial powers…(T)he place of the tropical world in the first globalisation was to be subordinate to the temperate and developed world. (S. Firth, 2000: 182)

The colonial era lasted into the 1960s in the region – although many territories are either full colonies or under some other arrangement of governance at a distance. Independence did arrive later than in other peripheral regions for two reasons. First, the diminutive size of the territories and colonies that were created under colonialism made some potential countries economically nonviable, many were simply not willing to cut ties without assurances that transfers, protections, and rights that were afforded were maintained. Second, during the Cold War the region played an important strategic role, once again the interactions between the superpowers of the time. Thus, the US and its allies maintained the strategic location of bases and other military activities in the Pacific. The French carried out, from 1960 to 1996, almost 200 nuclear tests in their territory in Polynesia and explaining – among other reasons – their great reticence to cede control to a growing wave of calls from both inside and outside French Polynesia to allow independence.

The MIRAB phase

Following the independence of some of the Oceanic territories, the economic model that was first espoused was that of import substitution industrialization (ISI). This largely failed in all of the cases although it was only ever half heartedly applied. Aid from the former colonial powers often funded these attempts at modernization. A major social revolution occurred in the 1960s and 1970s in those countries that were newly independent, involving the migration of large numbers of Pacific Islanders to cities on the Pacific Rim – especially those in

New Zealand and Australia and also in the USA. This led to the establishment of proportionally highly significant flows of remittances from the new homes back to the islands. This combination of factors led to the develop ment of what have been termed MIRAB economies.

The acronym MIRAB was coined to refer to countries where migration (MI), aid (A), remittence (R), and gov ernment employment (bureaucracy) play a major role. It was first applied to the Oceanic region in study from 1985 which noted that living standards in the Pacific Island region were much higher than productive eco nomic measure predicted. It was argued that the post colonial migration of large numbers of islanders established transnational corporations of kin through which remittances flowed. These remittances, as sub sequently argued, show little tendency to deteriorate even in third or fourth generation migrants – as such the attachment to home (sometimes among people who had ever actually been to such territories) was considered large. Aid was also large in some economies (although the relative importance of aid and remittances was seen to vary between countries and subregions) and this was initially the case for strategic reason but latterly due to postcolonial obligations. The government sector often became the major employer and small island nations had inherited colonial bureaucracies, and there was a minimum efficient scale for governmental functions. The model varies in its applicability across the region and there is fierce debate as to whether it applies today.

Today, it is certainly the case that remittances and aid maintain their roles. In Niue, for example, over 50% of gross national produce (GNP) is in the form of aid. Remittences account for up to 40% of Samoan income, and in Fiji remittances are now the single largest foreign exchange earner, above both tourism and sugar. Indeed, there is little evidence for any long term downward trend in remittances – in fact during times of conflict and natural disasters, which arguably characterizes the pre sent, they have tended to rise. Notwithstanding, there is no doubt that until recently at least aid levels have been rationalized – and there has been a downward trend with a general shift to project, tied and targeted rather than budgetary aid. Migration controls, especially in the USA and also in Australia and New Zealand have tightened. The neoliberal models that regional trustees are seeking to apply to the region have insisted on governmental reform which reduces the bureaucracy (B) in MIRAB. Having said this, recent work has shown that the model still remains robust in many places although patterns continue to shift. Most notable, some have added a T for tourism to the acronym, making TOURAB. Pro ponents of the MIRAB analysis insist that it still repre sents a rational and sustainable method of obtaining development.

Postcolonial globalization

Across the region since the mid 1980s there has been a paradigmatic shift to neoliberalism – albeit applied in dif ferent ways in different places. It is most certainly the case that the financial donors and lenders to the Oceanic region do not see the MIRAB structure as something that is cap able of delivering self sustaining economic growth. Par ticularly since the end of the Cold War the nature of loans and, increasingly, aid has become conditional upon free market reforms as is the case elsewhere in the periphery. Thus, downsizing the public sector, outward orientation, and reduction in subsidies have been insisted upon.

Oceanic countries and territories have found it dif ficult to compete in the global economy since the end of the colonial period. The period of decolonization took place just before the global shift to neoliberal policy and although the Oceania region adopted such policy rela tively late, it has had a profound impact. It is highly debateable whether such policies suit the very particular political and economic condition that the Oceanic na tions find themselves as price and policy takers. Not withstanding, the adoption of pure free market reform has not been as extensively pursued in the region, as say in Latin America where reforms have been purer – albeit partially reversed over recent years.

The adoption of neoliberalism in the region was led by Fiji following the coup of 1987. Seeking international favor the government agreed to market reforms in order to gain finance, and therefore, export processing zones were established as the economy was opened to overseas investors, where it was formerly relatively closed and characterized by ISI. Recently, however, the pace of neoliberal reform has gathered speed and virtually all Oceanic nations are subject to structural adjustment type agreements for their aid or loan funds.

In terms of aid, for example, not only has its delivery become increasingly contingent on reforms, but also has increasingly been allocated to projects that seek to stimulate capitalistic development. Since the early 2000s, there has been a shift in aid focus by Oceania's neighbors, New Zealand and Australia, toward promoting good governance and reducing poverty – which indeed mirrors shifts in aid priorities across the world. Though osten sibly laudable, such goals have been criticized by some as merely delivering free market reform dressed up in other ways. Poverty reduction strategies are, more often than not, aimed at stimulating outward oriented development and good governance, and at stimulating the downsizing of the state and western style democracy. Since 2001, Australia in particular has been quite aggressive in pur suing neoliberal reform in the region arguing that such a shift will promote security and reduce potential terror ism. In the face of such pressure, some Oceanic nations are turning to China and other Asian nations for less conditional aid.

In adopting such models the Oceanic nations have had little choice. Firth argues that the Pacific Island countries (PICs) have little choice.

> Just as the place of the Pacific Islands in the first glob alization was to be sub ordinate to the temperate, de veloped world, so their place in the second globalization is also to be subordinate, this time to a set of inter national institutions that have set the rules of the global economy. The new globalization, now combined with independent sovereign rule, means incorporation into the global economy on terms that suit the interests of the financial markets, the aid donors, and those relatively few Pacific Islanders who are in a position to benefit from the new situation. (S. Firth, 2000: 186)

The impacts of the shift to free markets have been controversial. There is evidence to suggest deleterious consequences across a number of sectors and parts of society. Neoliberal thinkers argue that such things have come about precisely because of the failure of the region to grasp free market reform, wean themselves off aid, and adopt models of good governance (i.e., western de mocracy) as political systems, most of which are based in democracy to varying degrees, but which are perceived by right wing thinkers on the outside as corrupt and ineffective. Where western systems exist, they are often combined with chiefly and traditional systems, as is the case in Fiji. Some recent studies have found evidence of impacts in the following areas: the evolution of un sustainable agro export systems; rising urban unemploy ment and marginality due to the erosion of traditional networks; the reduced role of Oceanic countries them selves in making policy as outside powers became in creasingly deterministic; a sharp rise in poverty in urban and rural areas affecting women and children the most; scandals in the public sector; the development of enclave tourism economies; and pressures to introduce privatized communal land structures despite their enormous wel fare implications. Whether neoliberalism is directly re sponsible for this is unclear. What is clear, however, is that something is wrong in the region and neoliberal policies are not addressing this.

Contemporary Welfare and Development in Oceania

Living standards are relatively low in the Oceania region – although this varies considerably internally. Welfare measured in terms of the human development index (HDI) is particularly low in Melanesia and Micronesia (see **Table 4**). In the case of the former, recent political instability has greatly exacerbated this relative depri vation, and places such as Vanuatu, Solomon Islands, and

Table 4 The human development index (HDI) Oceania, 1998 and 2006

Country	*HDI 2006*	*HDI 1998*	*Performance*
Cook Islands	0.789	0.822	⇓
Palau	0.774	0.861	⇓
Niue	0.741	0.774	⇓
Samoa	0.731	0.590	⇑
Tonga	0.718	0.647	⇑
Tokelau	0.679		
FSM	0.678	0.569	⇑
Fiji	0.667	0.667	
Marshall Islands	0.665	0.563	⇑
Tuvalu	0.653	0.583	⇑
Nauru	0.652	0.663	⇓
Kiribati	0.592	0.515	⇑
Vanuatu	0.586	0.425	⇑
PNG	0.414	0.314	⇑
Solomon Islands	0.410	0.371	⇑

Source: Adapted from SPC 2007.

PNG lie at the bottom of the HDI ladder, although in all cases there has been some improvement between 1998 and 2006. Generally speaking, there is a negative correlation between political freedom and living standards at the re gional level and the highest HDI measures are found in territories in free association and other semi independent island groupings. The Cook Islands and Niue rank highly in the HDI measure, despite a marginal worsening be tween 1998 and 2006. This is largely due to the role of official transfers in aid that characterize such territories. In some full colonies, living standards measured in this way are higher still. French Polynesia, for example, has particularly high income per capita sustained in this way. Within the relatively better off territories, however, in come inequality is often pronounced with the indigenous often finding themselves at the bottom of the ladder partly explaining the active independence movements in the French territories, for example.

If one looks at progress in the millennium develop ment goals in the region since their implementation it is clear that the record is mixed. Poverty is rising, and primary school enrolment is generally good but quality is poor, there has been some progress in gender equality, and infant mortality is declining but is still high, espe cially in Melanesia (see **Table 5**) mothers in rural areas are still at risk, HIV is rising, environmental agreements have been developed but not implemented, and aid – although high – is declining, and has a mixed record of reducing import dependence. As **Table 5** shows, life expectancy is particularly low in Melanesia, falling as low as 53.7 in PNG. However, the lowest in the Pacific is Nauru, which, since the depletion of phosphate, has suffered dire developmental consequences. Infant mor tality in some Oceanic territories, such as Solomon Is lands, ranks among the worst in the world.

Partly driving this poor human development per formance is the relatively poor economic performance of

Table 5 Selected development indicators for Oceania 2006

Country/subregion	*Life expectancy (male)*	*Life expectancy (female)*	*Infant mortality*	*Urban population*	*Urban population growth*
Melanesia				*19*	*2.9*
Fiji	64.5	68.7	22.0	46	2.6
New Caledonia	69.9	77.6	6.9	60	2.8
PNG	53.7	54.8	64.0	13	2.8
Solomon Islands	60.6	61.6	66.0	16	4.3
Vanuatu	65.6	69.0	27.0	21	4.2
Micronesia				*65*	*1.6*
FSM	66.6	67.5	40.0	21	2.4
Guam	74.5	80.8	9.0	93	n.a.
Kiribati	61.2	66.9	44.0	43	5.2
Marshall Islands	65.7	69.4	37.0	65	1.6
Nauru	52.5	58.2	42.3	100	0.3
Northern Marianas	72.5	77.8	5.0	90	3.4
Palau	65.5	71.9	17.0	81	2.2
Polynesia				*41*	*1.4*
American Samoa	69.0	76.0	8.5	66	3.7
Cook Islands	68.0	74.3	21.0	68.0	1.9
French Polynesia	69.2	74.1	6.9	52	1.6
Niue	68.8	71.2	29.4	34	4.3
Pitcairn Islands					
Samoa	71.7	74.2	19.1	22	1.3
Tokelau	68.4	71.3	33.0	0	n.a.
Tonga	69.8	71.8	12.0	32	0.8
Tuvalu	61.7	65.1	35.0	47	1.4
Wallis and Futuna	70.2	74.3	7.4	0	0.0
Total				*23*	

Source: SPC (2007).

the Oceanic countries. It has been difficult for the small island nations to compete as independent economic in a globalizing world, which is evident in terms of growing trade deficits across the region. In 2005, for example, in 18 of the 22 territories, trade deficits grew, with many of these representing a significant proportion of total na tional income. In many cases, as already noted, these trade gaps are plugged by aid, remittances, or a combination of these, and to an extent, tourism. Gross domestic product (GDP) per capita varies across the region markedly as illustrated in **Table 6**, which ranks the territories in descending order. Guam, an un incorporated territory of the US receives extensive military transfers as well as substantial tourism receipts. New Caledonia and French Polynesia receive extensive transfers from mainland France. The independent nations, particularly of Mela nesia, fare the worst with levels of per capita GDP that place among the worst in the world. Economic growth over the last 10 years has been poor across the region.

Levels of poverty are high by international standards in the Oceania region (see **Table 7**). Assessing poverty is plagued by poor access to data and **Table 7** presents a snapshot of the trends in various parts of the region. It is, by any measure, high in the Melanesian countries with,

Table 6 Economic indicators for selected territories in Oceania 2006

	GDP Per capita (various years)	*Total GDP (millons of US$) (various years)*	*Total GDP per capita growth 1995 2004*
Guam	22 118	3420	
New Caledonia	17 436	3620	
French Polynesia	15 637	3777	
Cook Islands	8563	174	6.5
American Samoa	7821	444	
Niue	5854	10	
Palau	5808	120	
Nauru	3450	40	
Marshall Islands	2362	131	4.6
FSM	2113	228	0.6
Samoa	2108	374	3.1
Fiji	1926	1591	1.6
Tonga	1893	190	1.7
Tuvalu	1563	15	2.0
Vanuatu	1558	308	1.6
PNG	796	4060	
Kiribati	613	53	2.3
Solomon Islands	521	232	3.5

Source: SPC (2005) and UNDP (2008).

Table 7 Poverty Indicators for Oceania

Country	*% less than $1 a day*	*% below basic needs poverty line*	*% underweight children below age 5*	*Income share of poorest 20%*
Cook Islands		12	10	
Fiji	26	25.5	15	2
FSM		27.9	15	5.2
Kiribati	38	50	13	6
Marshall Islands	20	20	27	3.3
Niue				5
PNG	30	37.5	24.9	4.5
Samoa	6	20.3	1.9	6.2
Solomon Islands			21	
Tonga	4	22.3	2	4.9
Tuvalu	17	29.3		7.4
Vanuatu	26	40.6	12.1	2.1

Source: Collated from *UN and SPC, 2005*.

for example, approximately 40% of people in Vanuatu and PNG living below the basic needs measure of poverty. This figure is likely to be very high in the Solomon Islands given the conflict there, which in part explains why the figures are not available. It is also rising in Fiji given 10 years of unrest. Poverty is also particularly high in Kiribati and parts of Polynesia, including Samoa and Tonga. In dications are that levels have risen over the previous 10 years in most territories. As the final column in **Table 7** shows, income distribution is regressive across the region.

In the past, poverty, as measured in Western terms, has been argued to be offset by subsistence affluence – tra ditional access to plentiful coastal, or forest resources historically. This concept can be questioned on a number of fronts. It is debatable, for example, whether such abundance and security ever actually existed outside of an idealizing Western imaginary. Conditions in Oceania – though varying from place to place – have always been challenging with the prosect of natural dis asters – such as cyclones and tsunami, ever present. While it is certainly true that, in some parts of Oceania, resources for daily human consumption are plentiful, this has been largely a function of the scarcity of population and patterns of resource consumption that are relatively nonintensive. As populations expand – they are set to double over the next 30 years overall – and social con sumption patterns change, and as a consequence pressure mounts. Other practices associated with modernization are having deleterious impacts on subsistence and semi subsistence livelihoods; for example, coastal tourism has in many cases taken up fertile fishing grounds and has had negative implications for reef ecosystems, all of which reduces access to traditional foodstuffs. Forestry in the interior parts of some islands has also impacted dietary patterns. Particularly in the Melanesian coun tries, deforestation has become a major problem as for eign exchange is sought in return for precious hardwoods. In some cases, the drive for exports in fish stocks, or

Table 8 Urbanization in Oceania, 2006

Country/ subregion	*Urban population (%)*	*Urban population growth 2000 05*
Melanesia	*19*	*2.9*
Fiji	46	2.6
New Caledonia	60	2.8
PNG	13	2.8
Solomon Islands	16	4.3
Vanuatu	21	4.2
Micronesia	*65*	*1.6*
FSM	21	2.4
Guam	93	n.a.
Kiribati	43	5.2
Marshall Islands	65	1.6
Nauru	100	0.3
Northern Marianas	90	3.4
Palau	81	2.2
Polynesia	*41*	*1.4*
American Samoa	66	3.7
Cook Islands	68.0	1.9
French Polynesia	52	1.6
Niue	34	4.3
Pitcairn Islands		
Samoa	22	1.3
Tokelau	0	n.a.
Tonga	32	0.8
Tuvalu	47	1.4
Wallis and Futuna	0	0.0
Total	*23*	

Source: Adapted from SPC 2007.

sugarcane production, to use just two examples, has led to the monocultural exploitation of sea and land in ways that undermine biological and economic diversity, and thus exacerbate the vulnerability of subsistence and small scale commercial livelihoods. Despite the myths, economic and social processes associated with modernization and more latterly neoliberalism make daily living in Oceania as difficult as anywhere in the global periphery.

Urbanization is a process that is leading to significant environmental pressure in terms of access to traditional nutritional resources and therefore patterns of subsis tence affluence or otherwise. The Oceanic region is one of the most rapidly urbanizing areas on the Earth (**Table 8**), the pressure on peri urban land has – in some cases such as Suva, Fiji – spiralled out of control. The large rural populations of Vanuatu, the Solomon Islands, and PNG mean that overall statistics seem to suggest that most Oceanic dwellers are rural. In just under half of the territories in the region, more than 50% of people live in urban areas, and this is rising overall at a rate of over 3% per annum across the regional as a whole. Growth figures are highest in Kiribati, and in the Melanesian countries. In the past, urban areas were of sufficiently low density to be able to supply many residents with urban food sup plies on local communal land, in individual gardens, and in some cases, on state land. In the case of the former, such demand was relatively inconsequential and the authorities did not always discourage the use of land in this way. As population densities increase and absolute numbers swell in the cities people increasingly have reduced access to such gardens. As a consequence, many have to rely on imported – often inferior foodstuffs such as mutton flaps – and dietary health has been shown to suffer as a consequence. The future of the Oceania re gion is undoubtedly an urban one and design meeting the social, economic, and infrastructural needs that accom pany this will prove a major challenge. Already sights that are familiar in urban areas across the periphery, such as shantytown evolution and the rise of the informal sector, are becoming prevalent across the larger urban centers
of the Pacific, such as Suva, Honiara, Port Vila, South Tarawa, Nuku'alofa, and Apia.

One factor that sets apart Oceania from other regions in the global periphery is land tenure. In most territories, land is held in some form of communal ownership and this has bode well for access. Although income de struction is skewed and probably worsening access to land has remained relatively equal until recently. The traditional form of ownership has placed pressure in some areas which have come into conflict with more westernized pressure – as is evident in the case of Fiji's sugar cane land. Generally, however, it has helped pro tect welfare. As urbanization increases, however, access to land is reduced, and, in some cases, traditional lands are left uncultivated. In places such as Tonga, reduced access to land in the countryside partly drives urbanization. Given the hereditary nature of the land holding system there, as time goes on, sons become entitled to ever smaller plots of land, which in many cases are too small to farm. Maintaining communal ownership of land is an essential component of ensuring sustainable livelihoods across the region.

Political Conflict

Some have linked growing political conflict in the region (see **Table 9**) to the rise of neoliberalism. It is most certainly the case that the inequalities produced through new economic models have fueled conflict in Tonga, Vanuatu, and Fiji. Furthermore, the obsessive pursuit of export oriented growth has precipitated conflicts over natural resource in the Solomon Islands as well as PNG. To others outside the region, the escalation in conflict in the Pacific region is difficult to imagine, given its ideal ization. But hot spots of conflict have arisen across the region and show few signs of disappearing. The focus on hot conflict also greatly underestimates simmering unrest across the region – there is the potential for open conflict bubbling under the surface in French Polynesia, New Caledonia, Tonga, Samoa, and Vanuatu, based largely on issues of self determination as well as inequality and poverty. This has seen outside powers take a much greater interest in regional affairs, citing security con cerns, and playing a role in peacekeeping and other intervention activity. Recently where conflict has oc curred, outsiders have labeled it as evidence of failed or weak states.

Table 9 A selective chronology of recent 'open' conflicts in Oceania

Country	*Type of conflict*	*Year*
Fiji	Military coups (two)	1987
	Civilian coup and hostage crisis	2000
	Military mutiny	2001
	Political crisis	2004 06
	Military coup	2007
Solomon Islands	Civil conflict (c. 30 000 lives lost)	2000
	Armed civilian coup	2000
	Intervention of NZ and Australian military and police	2003
	Urban riots, Honiara	2005
Vanuatu	Urban land riots, Port Vila	1988
	Kidnap of President	1996
	State of emergency after riots in Port Vila	1998
PNG	Civil war (Bougainville) (c. 20 000 lives lost)	1988 97
	Defence force mutinies	1997, 2001, 2002
	Continuing transborder disputes with Indonesia and cessation conflict in West Papua (Irian Jaya)	
	Intermittent armed clashes Bougainville	2005 06
Tonga	Pro-democracy urban riots, Nuku'alofa, foreign police deployed	2006

As **Table 9** shows that most of the open conflict is found in Melanesia – forming an 'arc of instability' running from Fiji to East Timor (which as an observer at the Pacific Islands Forum could be considered part of Oceania). There has however been increasing unrest in Polynesia too. For example, years of frustration on the part of the pro democracy movement led to un precedented rioting when the King of Tonga died in 2006 and his son (and heir) refused to enact reform. The anger became focused on the immigrant Chinese community which plays a central role in the economic affairs of the capital, Nuku'alofa. The episode was portrayed by the global media as an interethnic affair. This has often been the case, the conflict in Fiji, for example, has often been cast by nonregional media as 'ethnic tension' between Indo Fijians and indigenous Fijians. In reality, it is far more complex than this and reflects pressure between different confederacy groupings of ethnic Fijians, tension over land access, and basic inequality and poverty. The recent conflict in the Solomon Islands was also portrayed as another example of Balkanization when in reality its roots were much more deep seated in a political and economic sense.

Given such misrepresentations, the intervention of outside power can sometimes illustrate very little understanding of the fundamental factors driving conflict in the region and in some cases can exacerbate it. The 2007 coup in Fiji has proven very difficult for New Zealand and Australian governments to grasp as its leader, Commander Bainimarama – currently the self appointed prime minister – himself an ethnic Fijian os tensibly undertook it in order to protect Indo Fijians from what he described as dangerous ethno nationalists. Whatever the particular reasons for the most recent military coup in the region, the pertinent issue here is that conflict is rarely simple and the intervention some times not well informed, despite generally good intentions.

Regionalism in Oceania

In order to tackle the problems associated with isol ation and diminutive political and economic power, as well as face recent issues of conflict, there have been numerous attempts since World War II, and especially since independence to foster regionalism. The South Pacific Commission was formed in 1947 and is today known as the Pacific Community. In 1971, this body held the first South Pacific Forum (now known as the Pacific Islands' Forum) in New Zealand, an event that has been held annually ever since. In the early days, it was preoccupied with issues of the transition to in dependence and collective diplomacy. Today it deals with a broad range of issues with trade, aid, and security at the forefront. Sixteen countries, including Australia and New Zealand, are members and New Caledonia and French Polynesia are associate members and Tokelau and east Timor are the observers. Con troversially, the current head of the Forum is an Australian, a move which some view cynically given the overriding concerns of the Howard administration to push regional antiterror reform and free market pol icies. It has also been argued that the political weight of the group's two largest economies is already great enough.

There have been two major declarations over the last decade made by the forum. The Bikitawa Declaration made in Kiribati in 2000 paved the way for the in volvement of peacekeeping forces in the crisis of Bougainville, Solomon Islands, and Tonga. The Auckland declaration of 2004 laid out the Pacific Plan, which is the most comprehensive statement yet toward the union of political and economic interests across the region with the vision statement that Oceania be a "region of peace, harmony, security and economic prosperity….in which its cultures traditions and religious beliefs are valued, honoured and developed" (PIF, 2004: 1). The four pillars of economic growth, sustainable development, good governance, and security formed the core of this docu ment that has paved the way for further integration. For example, The PACER (Pacific Agreement of Closer Economic Relations) and the PICTA (Pacific Island Free trade Agreement) build on the plan and are evolving to complement other trade based moves, such as the Melanesian Spearhead Group and SPARTECA (South Pacific Regional Trade and Economic Co operation Agreement). Aside from these essentially centralized agreements, there is also a range of decentralized re gional cooperative institutions, including the University of the South Pacific and South Pacific Regional En vironment Programme. But the course of regional co operation is not always smooth and the challenges facing Oceania will require increasing vigor in this area for their successful resolution.

The End of Isolation – Managing Internal External Challenges

The rolling out of modernization in Oceania has caused both economic and cultural tensions as Western systems clash with indigenous or traditional ones (however these are defined). Political governance systems, for example, based in some countries and territories on highly hier archical chiefly systems have come into conflict with democratic agendas. Resource management has shifted from the traditional to the modern leading to environ mental conflict and the free market has replaced the reciprocal systems of previous times. Cultures are shifting rapidly, especially in urban areas, where the young increasingly mimic European influences. In short, the coming of the West in Oceania has been highly problematic and the consequences of this are felt in the daily lives of the population.

In terms of interpreting the recent history of devel opment in Oceania, then, there exists a familiar divide between neoliberal and structuralist type accounts. The former tends to blame Oceanic peoples themselves, citing corruption and poor market performance. The latter tend to see the problem as inherited from colonialism relations of power, which have perpetuated inequality and underdevelopment. The truth lies somewhere between the two. What is certain is that isolation in the islands has now ended and solutions to the existing problems need to take into account both internal and external factors. It is impossible to turn back globalization in the ocean region, but it will need to be managed and resisted through collective international developmental efforts, the pursuit of inclusive democracy, regional collaboration, and the design of sustainable environmental, and economic sys tems if it is to deliver anything but increased fragility for these already vulnerable territories.

Acknowledgments

Many thanks to Teresia Teaiwa for comments on an earlier draft and Conway Pene for supplying the figure.

See also: Aid; Australasia; Neoliberalism; Postcolonialism/Postcolonial Geographies; Regionalism.

Further Reading

Bertram, I. G. and Watters, R. F. (1985). The MIRAB economy in South Pacific microstates. *Pacific Viewpoint* 23(3), 497 519.
Campbell, I. (1989). *A History of the Pacific Islands*. Christchurch: University of Canterbury Press.
Connell, J. (1999). Urban dilemmas. In Rapaport, M. (ed.) *The Pacific Islands: Environment and Society*, pp 326 335. Honolulu, HI: Bess Press.
Firth, S. (2000). The Pacific Islands and the globalization agenda. *The Contemporary Pacific* 12(1), 178 192.
Hempenstall, P. (1994). Imperial Manoeuvres. In Howe, K. R., Kiste, R. C. & Lal, B. V. (eds.) *Tides of History: The Pacific Islands in the Twentieth Century*, pp 29 39. St. Leonards, NSW: Allen and Unwin.
Henderson, J. (2003). The future of democracy in Melanesia: What role for outside powers? *Asia Pacific Viewpoint* 44(3), 225 242.
Murray, W. E. (2001). The second wave of globalisation and agrarian change in the Pacific Islands. *Journal of Rural Studies* 17(2), 135 148.
Murray, W. E. and Storey, D. (2003) *Post Colonial Transformations and Political Conflict in Oceania* (Special edition of *Asia Pacific Viewpoint*, 44/3). Melbourne: Blackwell.
Nicole, R. (2000). *The Word, the Pen and the Pistol: Orientalism in the Pacific*. New York: State University of New York Press.
Nunn, P. D. (2003). Revising ideas about environmental determinism: Human environment relations in the Pacific Islands. *Asia Pacific Viewpoint* 44(1), 63 72.
Overton, J. D. and Scheyvens, R. (eds.) (1999). *Strategies for Sustainable Development: Experiences from the Pacific*. London: Zed Press.
UNDP (United Nations Development Programme) (1999). *Pacific Human Development Report 1999*. Suva: UNDP.

Oceanographic Mapping

D. M. Lawrence, Virginia State University and Virginia Union University, Richmond, VA, USA

Glossary

Bathymetry The measurement of water depths and analysis of the undersea features those data reveal.

Contour Map A map that illustrates topography (or bathymetry) with the aid of lines that connect points of equal elevation (or depth).

Fathom A measure of the depth of water equivalent to 6 feet.

Fathometer An instrument that measures the depth of water by sound.

Panoramic Map A map that illustrates topography as viewed from above.

Physiography The study of the structure of the Earth's surface.

Sonar A technology used to detect ships or underwater obstacles through the use of sound waves that echo off the target.

Sounding Measurement of the depth of a body of water.

Introduction

The oceans are arguably the most challenging mapping environment on Earth, maybe even the Solar System. The land can be surveyed by teams on the surface, or mapped using imagery obtained from higher elevations, or via aerial or satellite based platforms. The surface of many planets, planetoids, and moons can be observed via Earth based telescopes, spacecraft, or satellites. The oceans, until recently, were quite another matter.

The first challenge to mapping the oceans is their size. Seventy percent of the Earth's surface is covered by the oceans. The second is the depth. The average depth of the oceans is 3.8 km – the maximum depth is more than 10.9 km.

Except in the shallowest areas, the bottoms of the oceans are impossible to observe directly without specialized equipment. Humans cannot breathe underwater, and must either transport the oxygen they need for survival in self contained underwater breathing apparatus (SCUBA), or else bring a breathable atmosphere with them in underwater vehicles such as submarines or submersibles.

Shallow areas of the oceans can be mapped with scuba divers using more or less traditional surveying techniques, albeit with equipment capable of surviving the harsh marine environment. Marine archaeologists do this fairly routinely when mapping shallow water wrecks. But over most of the oceans, anyone attempting to map the depths using scuba gear runs into another significant problem – pressure.

Water, unlike air, is not a compressible fluid; thus, pressure changes rapidly with depth – about 1 atm for every 10 m of depth. Thus, the pressure of the deepest areas of the oceans adds up to more than 1000 atm. Aside from wreaking havoc on human physiology, water pressure can be devastating to surveying and sampling equipment.

In traditional mapping, humans survey what they can see – this is impossible in most of the ocean basins, as sunlight penetrates only 100 or 200 m beneath the surface (sometimes less, depending on the clarity of the water). Some wavelengths, such as reds and yellows, are absorbed quickly, affecting the visible colors within the photic (sunlit) zone. Even when artificial illumination is available, only a limited area is visible.

Obtaining Data

Despite all the challenges, the oceans needed to be mapped. As maritime commerce developed in ancient times, mariners quickly found the need to record the locations of reefs, sandbars, currents, and other navigational hazards. Mariners needed to ensure that the water was deeper than the draft of their ships. They needed to know the characteristics of the bottom, such as whether it was muddy, sandy, or rocky. So how were the early maps of the ocean made? By feel.

Early soundings were obtained using poles or weights suspended from a rope. Stick the pole in the water, drop the weighted line; when the bottom is felt, or when the line halts its descent, measure the length of the pole or rope pulled out – that is the depth. This technology was not improved upon for thousands of years, and may still be in use today. The first evidence of such methods dates to about 3500 years ago to Egyptian bas reliefs and paintings depicting the use of both sounding poles and weighted lines.

The first written account of using a weighted line to determine water depth dates to about 2500 years ago to a description of sounding the Mediterranean near the mouth of the Nile by the Greek historian Herodotus. Herodotus noted that the bottom consisted of a yellowish mud similar to that deposited by the Nile along its banks. One of the deepest soundings obtained from the

Mediterranean during that time period was described by the Greek geographer Strabo – 1000 fathoms in the vicinity of Sardinia. That record sounding, obtained by Posidonious about 100 BC, was not equaled for nearly 2000 years.

Sounding is referred to at some length in the New Testament book *Acts of the Apostles* – in the passage that describes the shipwreck of St. Paul on the island of Malta. The ship was caught in a storm, and one night, as the sailors suspected the ship was near land, took soundings that confirmed the approaching shore. They anchored in 15 fathoms of water and survived the night, but grounded the ship on a shoal the next day.

As can be imagined, the depths that could be measured with a sounding pole were quite limited. Ideally, such poles should not be any longer than the hull itself. They should not be heavier that what could be readily handled at sea by a man.

Weighted ropes could be used to measure much greater depths, but they were not without challenges of their own. For one, it could be difficult to know whether or not the bottom was reached. Usually, mariners relied on some indication that the line has stopped or slowed its descent. The downward momentum created by the weight of the deployed line itself could pull more line overboard with little noticeable decrease in velocity.

To prove that the bottom had been reached, John Mercer Brooke, a midshipman working for Commander Matthew Fontaine Maury at the US Navy's Depot of Charts and Instruments, devised a system consisting of a sampling tube that passed through the center of a weight – usually a cannonball with a hole drilled through it. Upon contact with the bottom, a mechanism would release the weight; meanwhile, the sampling tube would be driven into the sediment. Thus, as the sounding line was retrieved, some of the benthic material could be brought up for examination.

Brooke's device was first tested on 7 July 1853, on a cruise of the USS *Dolphin*, commanded by Lieutenant Otway H. Berryman, in the North Atlantic. A midshipman, J.G. Mitchell, deployed the device from a small boat at 1:20 p.m. It took him 6 h to retrieve the device – from 2000 fathoms of water. The bottom had clearly been reached, the sampling tube containing sediments full of diatoms and foraminifera.

Sometimes the bottom was never reached. This could happen for a variety of reasons: maybe the weight attached to the line was insufficient to pull it to the bottom; maybe it was too much, causing the line to snap. On other occasions, the weight of the line alone proved too much, resulting in a break. Regardless of the problem, the line would pay out and pay out – drifting with subsurface currents – giving the impression of a bottomless abyss.

Even if a sounding effort went as planned, the ropes typically absorbed water, thus gaining more weight. This phenomenon, in the days before a steam winch was developed, meant more backbreaking work for the sailors cranking whatever device was being used to retrieve the line. Even after steam driven winches were developed, the weight of sounding lines posed a problem.

A logical next step was to use lines made of material that would not absorb water – namely lines or cables made of metal. The first recorded attempts to use wire for sounding the sea was on the United States Exploring Expedition (1838–42) led by Lieutenant Charles Wilkes. He chose copper wire, which did not work too well, and he reverted to using rope.

The US Navy, under the direction of Maury, experimented with piano wire on the USS *Taney* expedition, commanded by Lieutenant Joseph C. Walsh, to the North Atlantic in 1849–50. In order to get sufficient length, segments of wire were spliced together. The sounding wires repeatedly parted at the junctions. The use of wire or cable was eventually perfected – several decades later.

Soundings in the deep sea took a long time when muscle power was required to raise the sounding lines. Nevertheless, even with steam or diesel powered winches available, raising the sounding lines took a long time. For example, HMS *Challenger* was equipped with a steam powered winch on its famous voyage circumnavigating the world, but sounding and other deep water sampling took hours. From December 1872 through May 1876, the *Challenger* took little more than 350 soundings.

A more rapid way of obtaining depth data was needed. The answer was something known of at least as far back as Aristotle: sound. (Actually, humans have probably known that sound could be heard in water since the time they have been swimming, but Aristotle is the first to record that fact.) In more modern times, Leonardo da Vinci wrote of his successful attempts to listen to distant ships via a tube immersed in water.

The use of sound waves in water to determine distance and depth was proposed in 1807 by Dominique François Jean Arago, a physicist and future prime minister of France. Arago, a polymath, had many other interests – such as helping to develop the wave theory of light – and thus failed to follow up on his proposal. A number of researchers determined the speed of sound in various aquatic settings; others, such as Maury, proposed specific techniques to measure depth of water using sound, but inadequate listening devices rendered the proposals fruitless.

A number of inventors began working on electronic devices that measured depth of or distance in water by sending and receiving echoes at the beginning of the twentieth century. The effort was given even greater impetus after the sinking of the 'unsinkable' RMS *Titanic* following its collision with an iceberg on the night of April 14 and 15, 1912.

Reginald Fessenden, a Canadian inventor working for Submarine Signal Company, had been developing an oscillator that sent a clear, penetrating sound through water for the company, which had been seeking an underwater warning system to help ships avoid collisions or running aground. They implemented an effective bell type device on board a number of lightships, but eventually they wanted something with greater capacity to send and receive messages underwater. The bells were less than ideal.

Fessenden's oscillator, however, was promising. Fessenden and Submarine Signal – motivated by the *Titanic* disaster as well as a collision of the liners *Republic* and *Florida* off Nantucket in 1909 – refocused their efforts, seeking a way to detect obstacles, such as icebergs, as well as measure the depth of the bottom. Fessenden further improved his oscillator that sent out a clear 'ping' sound as well as a receiver that could detect the echo of the sound as it bounced off another object.

On 27 April 1914, a seasick Fessenden and a seasick crew on board the US Coast Guard cutter *Miami* finally caught enough of a break in the dreary, unsettled North Atlantic weather to test the efficacy of the device in detecting icebergs. A terse telegram sent to Boston announced the results: "Bottom one mile. Berg two miles."

Submarine Signal focused on marketing Fessenden's device as a means of underwater communications. Little thought of using it to measure depth on a routine basis occurred until 1922, when Submarine Signal asked engineers R. L. Williams and Herbert Grove Dorsey to work on improving it for depth measurement. Fessenden's oscillator required an operator to listen for echoes using headphones. Their contribution was in developing an apparatus that could detect echoes automatically. All the operator had to do was to read the depth on a lighted scale. Dorsey named the instrument the fathometer. Similar devices were called echo sounders.

Early versions of the fathometer were not capable of accurate depth measurement in very shallow waters, or while a ship was underway. While Williams and Dorsey continued Fessenden's efforts to improve the Submarine Signal device, others worked to develop similar instruments.

Paul Langevin, a French physicist, and Constantin Chilowski, an expatriate Russian electrical engineer living in Switzerland, collaborated on an instrument that used supersonic frequencies. Their echo sounder was the first to be used in a practical application – a survey of a route for an undersea cable crossing the Mediterranean from Marseilles to Philippeville, Algeria, in 1922. The British, under Robert Boyle, developed their own system – primarily for use in antisubmarine warfare.

Back in the United States, Harvey C. Hayes of the Naval Research Laboratory developed what he called a 'sonic depth finder' that could rapidly and accurately measure depth, even while a ship was steaming rapidly across the open ocean. Hayes tested his device, first on the battleship USS *Ohio* on a short run between Annapolis, Maryland, and New York City in February 1922.

Hayes' sonic depth finder was then transferred to the destroyer USS *Stewart* prior to its departure to the Philippines via Gibraltar and the Suez Canal the following June. The *Stewart*, departing from Providence, Rhode Island, acquired more than 900 soundings in the 9 days it took to reach Gibraltar. The achievement, in comparison with the 350 soundings made by the HMS *Challenger* in the 3 years it took to circumnavigate the world, was set to revolutionize oceanographic mapping. By 1924, it was already revealing errors in depth measurements made by traditional weighted line methods.

Despite the profusion of work in echo sounding the depths in the years after 1914, Fessenden was not forgotten. He was awarded Scientific American's Gold Medal in 1929 for his invention of the fathometer.

Fathometers were limited to measuring the depth of the ocean at a single point at a given time; thus, a ship could only record a single line of data as it traversed the oceans. An ideal next step would be to record data from several lines – beams – at once. Developments along those lines began with a failed attempt by General Instrument Corporation to develop an instrument for radar based mapping using the Mills Cross technique.

Named for a radio telescope in New South Wales, Australia, the Mills Cross technique uses a perpendicular array of instruments to collect data from a wide swath of space – in both the astronomical and geographical senses. For a radio telescope, the perpendicular arrangement of sensors allows precise determination of the location of a radio source. The General Instrument engineers hoped to adapt the technique to radar based mapping, by designing a system where an array of radar emitters was perpendicular to an array of radar receivers, allowing for multiple steered beams that could precisely determine distances to multiple points simultaneously with each ping.

While the effort to obtain the radar mapping contract failed, they contacted three engineers, Harold Farr, Paul Frelich, and Richard Curtis, in a sister division – the Harris Anti Submarine Warfare Division – to see if the sonar group could benefit from the idea. If it worked, it would allow high resolution images, analogous to contour maps, of the ocean floor. The idea worked. By 1963, the three had developed and installed a multibeam sonar system on the USNS *Compass Island*. By 1968, they had begun development of a commercial system called Sea Beam. The first operational Sea Beam system was installed on the French research vessel *Jean Charcot* and used on the *Vema II* cruise in 1977. Instead of just a single line, ships equipped with multibeam sonar can map thousands of square kilometers of ocean floor in a single cruise.

Even with multibeam sonar systems, the extent of the ocean floor mapped was limited to regions traversed by survey vessels. Another method was needed to map areas that had not been visited by ships. Just such a method arose from the study of the topography of the ocean surface.

Sea level is not level – even when the effects of wind and waves are taken into account. The sea surface is analogous to a rolling landscape consisting of hills and valleys deviating as much as 100 m from average sea level. The height anomalies are caused by the gravitational effects of the features of the ocean bottom. Geophysicists had long recognized this, and beginning in the nineteenth century had sought ways to measure the strength of the force of gravity at the ocean surface. At the time, gravity was measured with a variety of devices, such as a type of gravity barometer, but most methods were not of suf ficient precision to use in geophysical research. Pendu lum devices were best for measuring gravity, but ship motions on the surface rendered most such devices useless at sea.

Felix Andries Vening Meinesz devised a multiple pendulum device effective in measuring gravity in the soft, shifting soils of his native Netherlands, and adapted it for use at sea in the 1920s. He still needed a stable platform for taking the measurements at sea, and found what he needed in submarines. Vening Meinesz, a big man, squeezed himself into the small submarines of his day and, over a span of 15 years, from 1924 to 1939, traveled more than 125 000 miles and took more than 800 gravity observations at sea – nearly doubling the total at the time. After World War II, several thousand more gravity observations were made on submarine cruises until the mid 1950s when Anton Graf and Lucien LaCoste independently developed spring gravimeters accurate and stable enough to be used on surface ships.

The relationship between gravity and sea surface height is not what it may seem at first glance. Gravity is an attractive force which, according to Isaac Newton, is largely a function of the masses of the objects involved and the distance between them. (Gravity is better ex plained by the general theory of relativity than by Newton's 'laws' of gravity, but the 'laws' are still useful approximations in most cases.) Thus, it would seem that a massive object that rises from the ocean bottom to near the surface would create a depression in the ocean sur face; conversely, a depression in the ocean bottom, such as a deep sea trench, would create a rise in the ocean surface.

A seamount does pull water toward its slopes, but because the force of gravity acts perpendicular to the surfaces involved, the overall effect is to squeeze water toward the surface above the summit, creating a bulge instead of a dip. Conversely, the gravitational attraction along the walls of a depression in the sea floor pulls water away from the surface, creating a depression on the surface.

The work of Vening Meinesz and others established the relationship between gravity anomalies and ir regularities in the elevation of sea level. Surface based measurements, however, were limited in their coverage of the world's oceans. What was needed was a method of measuring gravity anomalies on a global scale.

The space race between the United States and the Soviet Union triggered the development of just such a method. The US National Aeronautics and Space Ad ministration began developing a series of Earth observing satellites in the 1960s and 1970s. Skylab, launched in 1972, was fitted with an early radar altimeter to measure the height of the ocean surface. A satellite, *Geodetic and Earth Observing Satellite 3* (*GEOS 3*; 1975–78), was equip ped with a better altimeter. *SEASAT* (1978) was specif ically designed for observing the Earth's oceans. Fitted with a radar altimeter that could measure the height of the ocean to within five centimeters, *SEASAT* held great promise, but its batteries failed within 3 months. Still, it logged more than 100 million km of tracks across the Earth's surface, far more than most ships could cover in a century. Newer satellites have since been launched to continue the work of *GEOS 3* and *SEASAT*.

William Maurice Ewing, a young geophysics professor at Lehigh University (and future founder of Lamont Geological Observatory, now Lamont Doherty Earth Observatory), further revolutionized oceanographic mapping by applying seismic techniques that he had learned in the oilfields of Texas to studies of the geology of the ocean floor.

Making Maps

Because of the dearth of depth measurements, much less accurate depth measurements, the first attempts to map ocean basins were doomed to be suspect. Matthew Fontaine Maury made the first serious attempt to map an ocean basin in 1854, producing the *Bathymetrical Map of the North Atlantic Basin with Contour Lines Drawn in at 1,000, 2,000, and 4,000 Fathoms*. The map, made with only about 200 soundings, proved that the ocean floor was far from a flat, featureless plain it was once thought to be, but it could not reveal the features that were there in much detail. The most prom inent feature he called the Dolphin Rise – an apparently broad plateau in the middle of the North Atlantic. The plateau is actually not so broad. It is a relatively narrow, often rugged mountain range now known as the Mid Atlantic Ridge.

The *Challenger* expedition generated so much data that original reports based on its results were published for decades after the ship returned to England in 1876. Sir John Murray, who began as a young naturalist on the expedition and who took over as editor of the exped ition's reports following the death of chief scientist Charles Wyville Thomson in 1882, made a fortune from mining phosphate rich guano deposits he discovered afterward on Christmas Island in the Indian Ocean. Much of the money he spent on scientific research – such as underwriting the publication of later volumes of the *Challenger* report, but some of it was spent on other en deavors, such as mapping the ocean basins using new data coming in every year. His maps show a progressive im provement in resolution, such that the Mid Atlantic Ridge looks like a ridge in his North Atlantic map of 1911.

Some of the best early maps of individual basins were prepared by Alexander Agassiz, son of naturalist Louis Agassiz, from data collected on three cruises of the US Coast and Geodetic Survey steamer *Blake* in the Gulf of Mexico, Caribbean, and the Western Atlantic between 1877 and 1880. Agassiz's maps, each based on hundreds of soundings, stand up fairly well today, even though they were published in 1888.

The next great advance in oceanographic mapping came with the advent of echo sounders. One of the premier expeditions in the early part of the twentieth century was that of the German research vessel *Meteor* from 1925 to 1927. The *Meteor* zigzagged across the Atlantic – primarily the South Atlantic – taking paired echo soundings at about 30 000 stations and following up with several hundred wire soundings for further verifi cation. Scientists on board *Meteor* also measured physical characteristics of the water such as temperature at depth. The ocean basin maps based on that data were a vast improvement over earlier work, including Murray's, in the Atlantic. For example, both sounding and tempera ture at depth data were used to prepare one map of the Atlantic Basin showing the Mid Atlantic Ridge snaking through the oceans roughly mid way between North America, Europe, and Africa. There is a hint of an ex tension into the Indian Ocean.

From Maury through the *Meteor*, maps of ocean basins had typically been drafted as contour maps. The vast amounts of data available made much better maps pos sible. But when oceanographer Bruce Charles Heezen and his geologist/cartographer partner Marie Tharp of what is now Lamont Doherty Earth Observatory em barked upon an effort to map the world's ocean floors in 1952, however, they faced a major problem. The US Navy, because of Cold War security concerns, forbade the use of their data in the preparation of a detailed bathymetric contour map.

Heezen and Tharp were only briefly stumped. Rather than quit, they decided to see if they could get away with making physiographic maps – illustrating what the ocean basins would look like if all the water had been drained – in the style of cartographer Armin Karl Lobek. While making transatlantic profiles working with the *Stewart*'s soundings, Tharp discovered a notch in the Mid Atlantic Ridge, a rift valley, which meant the Earth's crust was splitting along those cracks. The discovery, along with ensuing research tracing the location of the rift with a combined use of sounding data and earthquake epicenter locations, showed the mid ocean ridge systems extended for tens of thousands of kilometers around the surface of the Earth.

Heezen and Tharp began working on a series of ocean basin maps, beginning with that of the North Atlantic in 1957. They completed work on the South Atlantic in 1961. The pair originally planned to map the Medi terranean next, but the oncoming International Indian Ocean Expedition prompted them to map that basin next. All this time, Tharp, who plotted thousands and thou sands of measurements by hand, and Heezen, filled in areas in between sounding tracks by educated surmises as to what lay between. This method – the best anyone could do with the limited data of the time – caused problems in the Indian Ocean, however. No one had yet envisioned triple junctions – where three tectonic plates come together – so their map omits the triple junction now known to be there.

The Indian Ocean map did get the attention of the editors of *National Geographic Magazine*, however. Those editors teamed Heezen and Tharp with Heinrich Caesar Berann, an Austrian artist who pioneered the modern panoramic map, for a series of full color maps of the ocean basins. The team worked well together. After completing the series for *National Geographic*, they began preparing a world ocean floor map. That work, under written by the Office of Naval Research, was published in 1977. Both a map and a work of art, it remains an iconic presence in earth science departments three decades later (**Figure 1**).

Heezen, unfortunately, never saw the finished project. Shortly after approving the proofs and sending the map to the printer, he died of a heart attack while preparing for a dive on the Reykjanes Ridge aboard the US Navy's research submarine *NR 1* on 21 June 1977.

A problem that plagued Heezen, Tharp, and other mappers of the ocean floor was the uneven availability of sounding data. The Space Age, however, offered a solu tion in the form of eyes in the sky. William F. Haxby, a graduate student at Cornell University, was asked by his advisor, Donald L. Turcotte, to take a look at some *GEOS 3* radar altimetry data that Turcotte was given while attending a NASA meeting in 1976. Haxby could do little with that, but the 3 months of *SEASAT* data was another matter.

Haxby, intending to produce a map of gravitational anomalies, realized, as Felix Andries Vening Meinesz

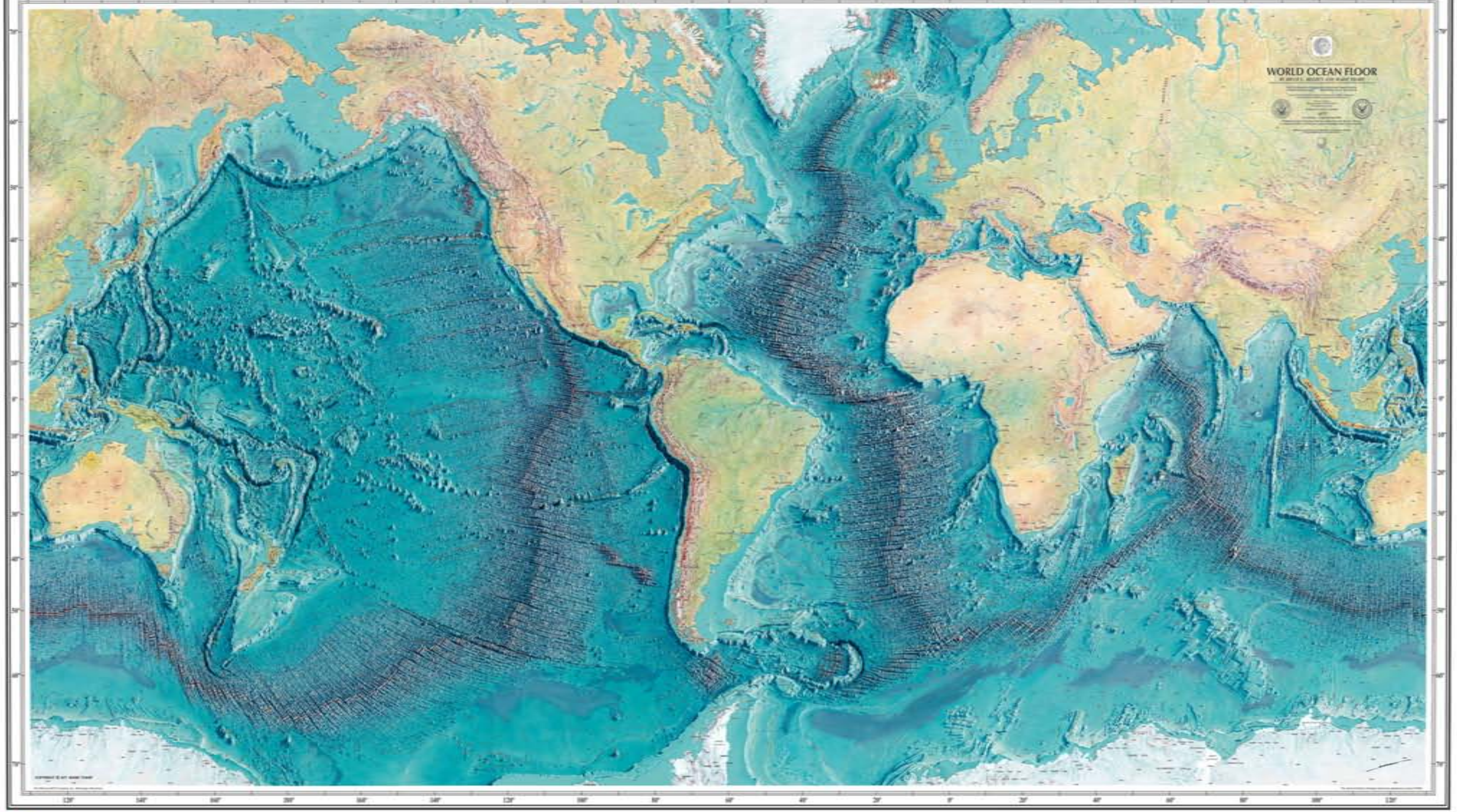

C M Y Black Process Black Type Black Grid

Figure 1 World Ocean Floor Panorama, Bruce C. Heezen and Marie Tharp, 1977. Copyright by Marie Tharp 1977/2003. Reproduced by permission of Marie Tharp Maps, LLC 8 Edward had, that the anomalies also revealed ocean floor topography. By 1982, he had produced a map of the world's ocean floor using satellite data alone. The map, 'Gravity Field of World's Ocean (Color Map)' was published by the US National Geophysical Data Center in 1985.

Even with the global reach of satellites, techniques such as Haxby's suffered from errors as well as insufficient resolution when applied to smaller areas or in areas of complex terrain. Walter H.F. Smith of the US National Oceanic and Atmospheric Administration and David T. Sandwell of Scripps Institution of Oceanography combined radar altimetry data with traditional sounding data where available. The result was a higher resolution, more accurate map of the world's ocean basins, published in 1997.

Accomplishments and Future Prospects

The efforts to map the world's ocean floors have revealed many secrets of the deep. Among the most important discoveries was the finding that the ocean floors were as topographically complex as any landscape above the ocean surface. The mapping effort revealed the composition and structure of tectonic plates as well as evidence for many of the processes involved in tectonic activity. The work has revealed many secrets of the Earth's past, and has led to the opening of vast areas of the Earth's surface to mineral exploitation.

As with all things, no single oceanographic mapping technique is adequate in all occasions. Modern hydrographers use every method at their disposal to get the quality data they need for their research. The focus is on mapping individual ocean basin and smaller regions in greater detail, although world ocean floor maps will continue to be improved as more data come in.

See also: Art and Cartography; Cold War; Exploration; Geodesy; Mapping, Topographic; Maps; Natural Resources; Oceania; Oceans; Surveying.

Further Reading

Adams, K. T. (1942). *Hydrographic Manual*. Washington, DC: US Government Printing Office.

Agassiz, A. (1888). *Three Cruises of the U.S. Coast and Geodetic Survey Steamer "Blake" in the Gulf of Mexico, in the Caribbean Sea, and along the Atlantic Coast of the United States, from 1877 to 1880, Vol. 1*. Boston, MA: Houghton Mifflin and Company.

Amato, I. (1998). *Pushing the Horizon: Seventy Five Years of High Stakes Science and Technology at the Naval Research Laboratory*. Washington, DC: Naval Research Laboratory.

Blake, R. F. (1914). Submarine signaling: The protection of shipping by a wall of sound and other uses of the submarine telegraph oscillator. *Transactions of the American Institute of Electrical Engineers* 33(2), 1549 1561.

Bruins, G. J. and Scholte, J. G. J. (1967). Felix Andries Vening Meinesz. 1887 1966. *Biographical Memoirs of Fellows of the Royal Society* 13, 294 308.

Dall, W. H. (1883). First use of wire in deep sea sounding. *Science* 1, 65.

Dorsey, H. G. (1933). Echoes give ocean depths. *Scientific Monthly* 36, 65 68.

Hall, S. S. (1992). *Mapping the Next Millennium: How Computer Driven Cartography Is Revolutionizing the Face of Science*. New York: Vintage.

Hall, S. S. (2006). The contrary map maker: Marie Tharp b. 1920. *New York Times Magazine December* 31, 45.

Hayes, H. C. (1924). The sonic depth finder. *Proceedings of the American Philosophical Society* 63, 134 151.

Lawrence, D. M. (1999). Mountains under the sea: Marie Tharp's maps of the ocean floor shed light on the theory of continental drift. *Mercator's World* 4(6), 36 43.

Lawrence, D. M. (2002). *Upheaval from the Abyss*. NewBrunswick, NJ: Rutgers University Press.

L 3 Communications SeaBeam Instruments (2000) Multibeam sonar: Theory of operation. Washington, DC: L 3 Communications SeaBeam Instruments.

Nelson, V. J. (2006). Marie Tharp, 86: Pioneering maps altered views on seafloor geology. *Los Angeles Times* September 4, B 15.

Raisz, E. (1931). The physiographic method of representing scenery on maps. *Geographical Review* 21, 297 304.

Smith, W. H. F. and Sandwell, D. T. (1997). Global sea floor topography from satellite altimetry and ship depth soundings. *Science* 277, 1956 1962.

Spiess, F. (1985). *The Meteor Expedition: Scientific Results of the German Atlantic Expedition, 1925 1927*. Emery, W. J. (trans.). New Delhi: Amerind Publishing Company.

Theberge, A. E. (1999). Sounding pole to sea beam. *Surveying and Cartography* 5, 334 346.

The New York Times (April 16, 1912) Many great liners paid toll of the sea. *The New York Times*.

Wilford, J. N. (2000). *The Mapmakers* (2nd edn.). New York: Alfred A Knopf.

Oceans

P. E. Steinberg, Florida State University, Tallahassee, FL, USA

Glossary

Containerization A series of practices that are associated with the use of shipping containers, large steel metal boxes that hold commodities on ships and that can seamlessly be moved to truck or railroad once a ship reaches its destination port.

Exclusive Economic Zone (EEZ) Established by the 1982 United Nations Convention on the Law of the Sea, a country's EEZ includes areas of the ocean that are between 12 and 200 nautical miles from its coastline and areas of the seabed between 12 nautical miles and the limits of the continental shelf (in instances where the continental shelf extends beyond 200 nautical miles from shore). A country retains exclusive rights to the living and nonliving resources of its EEZ, although for all other purpose high-seas freedoms prevail.

Flag of Convenience A national ship registry that is sought out because of low registration costs and/or lax safety and labor regulations. The owners of a ship flying under a flag of convenience typically have no material relationship with the registry state.

High Seas The area of the ocean that is beyond the territory of a coastal state. Although, according to the 1982 United Nations Convention on the Law of the Sea, the high seas begin 12 nautical miles from shore; in practice, many high-seas freedoms begin only at 200 nautical miles from shore because of the overlapping EEZ regime.

Manganese Nodules Potato-sized clumps of manganese, nickel, cobalt, copper, and other minerals that are found on the surface of the seabed in certain locations beyond the continental shelf.

Nautical Mile 1.15 statute miles, or 1.852 kilometers.

Territorial Waters The area of the ocean that is within the territory of a coastal state, defined in the 1982 United Nations Convention on the Law of the Sea as an area extending to 12 nautical miles from shore.

United Nations Convention on the Law of the Sea The document that emerged from the Third United Nations Conference on the Law of the Sea (1973–82) and that went into effect in 1994.

The Forgotten 71%

Oceans cover 71% of the Earth's surface and contain 97% of the planet's water. Over 20% of the world's petroleum is derived from offshore sources and 95% of world trade by weight, or two thirds by value, is carried by ship. Eighty percent of the world's fish catch comes from the ocean, supporting the livelihoods of 140 million people. Economists have calculated that the world's oceans provide services to humanity valued at US$21 trillion, as opposed to only US$12 trillion provided by land.

Geographers have long urged that attention be drawn to this space. Reflecting on Strabo's comment some 2000 years ago that "we are in a certain sense amphibious, not exclusively connected with the land, but with the sea as well," Ellen Churchill Semple noted in 1911, "Our school textbooks in geography present a deplorable hiatus, because they fail to make a definitive study of the oceans over which man explores and colonizes and trades, as well as the land on which he plants and builds and sleeps" (Semple, 1911). Forty years later, Richard Hartshorne (1953) declared that "the fundamental error in popular geographic thought" (Hartshorne, 1953: 382–393) was the tendency to view the ocean as simply a barrier rather than as a space of human society. Nonetheless, few geographers have heeded their call.

The ocean has failed to attract attention from more than a handful of human geographers for three likely reasons: first, despite the fact that geographers have long critiqued the idealization of the state as a naturally occurring, organic entity, geographers (and, more generally, social scientists) still have tended to conceive of the world as a universe of state territories. As a space that lies primarily external to the territory or sovereign authority of individual states, the ocean thus has appeared as a space that is external to the space creating processes of society.

Second, human geographers (and, again, social scientists more generally) have tended to view societies as occurring in place. Key social activities, such as production, reproduction, and consumption, as well as the cultural forms that support these activities, traditionally, have been associated with discrete places or territories. Movement typically has been viewed as a derivative activity that occurs simply because an individual or a commodity requires relocation from one society place to another, and little attention, therefore, has been directed toward the spaces across which this movement occurs. Thus, notwithstanding the ocean's substantial economic value as the space across which the bulk of the world's commerce flows, it has received little attention from human geographers.

Third, human geographers traditionally have viewed nature as ontologically distinct from society. Although geographers have long focused on the intersection between nature and society (examining, for instance, the way in which a place's nature and its society impact each other), this emphasis on the intersection between nature and society has tended to direct human geographers' attention away from spaces of nature that fail to display a clear human presence. Thus, as a space that is without permanent human habitation and that long was thought to be immune to human impact, the ocean typically has escaped the attention of human geographers who study nature–society relations.

A Sea Change in Epistemology

In recent decades, each of these barriers to human geographic research on the ocean has been challenged by broad reaching shifts in social thought. Turning first to the barrier of state centrism, since the 1970s the rise of globalization studies has led to the emergence of a number of perspectives that challenge the prevailing view of the world as a set of states that interact with each other. Instead, the world increasingly is being conceived of as one integrated (albeit differentiated) social system. This shift is significant for the ocean, because if, for example, one thinks of the world as consisting of 190 odd state societies, then the 71% of the planet's surface that is not in any of those 190 state societies is, by implication, beyond society. If, on the other hand, one thinks of the world as consisting of one society, with many constituent varieties of spaces, then it would be logical to include the ocean as one of those spaces of society. Hence, the ocean shifts from being conceived of as an external space between the world's societies to being conceived of as an integral space of society.

Similarly, the prevailing notion of society as the actions of people in place has been challenged by a range of perspectives associated with post structuralism, post colonialism, and transnationalism. These perspectives all attempt to destabilize the concept of essential, place based identity, instead directing attention to hybrid formations that emerge through interaction and movement. Border areas, including spaces of movement like the ocean, which formerly had been viewed as in between spaces that were not worthy of attention, are now recast as the spaces wherein social processes and identities are constructed (and are continually reconstructed), and they increasingly are foregrounded in social analysis.

Third, a growing number of human geographers have been interrogating the concept of nature. These geographers range from those who argue that nature is socially constructed to those who deny that there is any basis for making an ontological distinction between the social (or human) and the natural (or nonhuman). Once one subjects nature to a critique along these lines, it becomes impossible to dismiss the ocean as a space of nature that is beyond the concern of human geographers.

Emergent Geographies of the Sea

These epistemological shifts, in turn, have led to new areas of research in the human geography of the ocean. In the remainder of this section, four areas of human geography are reviewed: political, cultural, environmental, and economic geographies of the sea.

Political Geographies of the Sea

Historically, when human geographers have studied the ocean, they most often have looked to the sea as a space of politics and, in particular, as an arena for geopolitical conflict. The ocean has been viewed as a space in which the forces of land based states meet as each attempts to control crucial areas of the sea (e.g., choke points on shipping lanes or particularly productive fishing grounds) or to use the ocean as a platform from which power can be projected onto land. Increasingly, however, political geographers (especially those who associate themselves with the critical geopolitics movement) are criticizing this model of the world as one in which states with unambiguously bounded insides interact with each other on a preexisting (and presocial) spatial platform. Instead, the discursive bounding of states and the construction of the world as a universe of mutually exclusive territorial states is seen as an act of politics that itself is worthy of study. It follows from this critique that there is no *a priori* distinction between a state's 'inside' (its territory) and its 'outside' (space that is outside state territory). Thus, the ocean, historically viewed by political geographers as a space wherein political actors battle each other, is now recognized as a space in which politics (and political territories) are created.

Some of the work in this area queries the division of the sea into three distinct zones that is enshrined in the 1982 United Nations Convention on the Law of the Sea: the territorial sea (which extends to 12 nautical miles from shore), in which states have complete territorial authority except that they must allow free passage to foreign ships that are deemed to have 'innocent' intentions; the high seas (the area beyond 200 nautical miles from shore), in which coastal states have no authority and in which state authority is limited to the power that a state has over ships that are flying its flag; and the exclusive economic zone (EEZ; roughly the space between territorial waters and the high seas), in which high seas freedoms of the sea prevail except that coastal states have exclusive rights to marine and seabed resources. While the implementation of this regime can be traced to the power politics of global treaty negotiations, it can also be traced to structural contradictions within the spatiality of capitalism: the need to invent new investment sites (expressed in the territorial seas regime), the need to move capital and commodities at ever greater speeds (expressed in the high seas regime), and the need to steward nature so as to forestall capitalism's tendency to exploit it at an unsustainable rate (expressed in the EEZ regime).

The interplay of these forces not only constructs these three distinct legal zones but it also constructs other sociospatial patterns that reflect and reproduce the unevenness of political power and nature in the marine environment. Consider, for instance, the ease with which Japan (driven by domestic considerations) is able to use its economic leverage over cash poor but whale rich countries to win their votes in the International Whaling Commission. The resulting whaling regulations both reflect and transform relations of power and distributions of nature, at sea and on land. When viewed from this perspective, the ocean is seen not simply as an arena wherein political actors compete with each other but rather as a space within which the political structures of the world system are constructed.

In a similar vein, historical geographers of empire have begun turning to the sea in their attempts at understanding the relationship between the ordering of space and the ordering of peoples. If the ideology of ocean crossing has a special role in the discourses of global reach that have characterized modern empires (and perhaps ancient ones as well), then the ocean is itself a constructed space of politics. In particular, the British Empire with its strong ideological emphasis on sea power, has been subject to examination by Anglo American human geographers. Like the empire's colonies, the ocean was not only constructed as a space that was of the empire but that was also conceived of as one step removed from the essential space and character of the British (or English) nation. Thus, ocean space, like colonial space, not only in the British Empire, but also in others – particularly the Portuguese – was constructed through an ideology that combined elements of both universalism (or cosmopolitanism) and particularism (or chauvinism).

Finally, the recognition of the ocean as a space wherein politics is not only enacted but also constructed has led some to direct their attention to political actors in ocean space who directly challenge the norms of terrestrial political geography and the spatiality of the state system. From the eighteenth and nineteenth century pirate societies to the twentieth and twenty first century global shipping companies that fly under flags of convenience, these actors employ state signifiers and organize space in complex ways that simultaneously challenge, modify, and reproduce the land based system of sovereign, territorial nation states.

Cultural Geographies of the Sea

While contemporary studies of the political geography of the sea can draw upon a long history of political geographers studying maritime conflict, the study of marine issues is quite new in cultural geography. Of course, there always have been marine cultural geographies; seafaring and fishing communities invariably display distinct cultural formations that reflect and impact the surrounding marine environment. Historically, however, few geographers have devoted their attention to the cultures of fishing communities and even fewer have studied the cultures of societies engaged in uses of the deep sea (e.g., whalers, naval personnel, merchant mariners, oceanographic and fisheries researchers, or long distance fishers).

As has occurred throughout cultural geography since the 1980s, much of the impetus for new cultural geographic studies of the sea has come from outside the discipline, especially from anthropology, history, literature, and cultural studies. This turn to the sea – and to understanding the sea as a space of culture – perhaps emerged first in the discipline of history. Fernand Braudel's *The Mediterranean and the Mediterranean World in the Age of Philip II* demonstrated that the ocean, far from being an empty space between cultures, was in fact a space of interaction in which cultures (and natures) were formed and transformed. This perspective, which directly challenged prevailing ideas in cultural ecology (as well as in history and anthropology) about cultures being rooted in place, was soon taken up by other historians who started new ocean basin based organizations of history (most notably Atlantic history). In the 1990s, this work on the history of ocean regions began to be joined with work emerging from cultural studies (and, increasingly, cultural geography) on diasporas, hybrid identities, and transnationalism. Broadly, scholars associated with this school of thought stress the ways in which cultures are continually reproduced through movement and connection rather than through stasis in place. A key motivational book for integrating the ocean within this line of thinking was Paul Gilroy's *The Black Atlantic*. Although Gilroy only indirectly considers the ocean as a material space of society, the book, which has been highly influential in cultural studies, suggests the power of oceanic metaphors, and, by association, oceanic spaces, in interpreting cultural formations.

To ground Gilroy's metaphors, some geographers have turned their attention to shipboard life. Shipboard life provides a unique environment for geographic research, because it is simultaneously mobile and stable, it is both a workspace and a living space, it is both self contained and open ended, and a ship is both an insular community and one that typically brings people of many backgrounds and nationalities together. Again, work in this area has been aided by advances outside the discipline, in particular by social historians who have researched the role of the eighteenth and nineteenth century seamen, pirates, and whalers in constructing modern norms and disciplines of nation, class, gender, sexuality, and labor.

Within geography, the interdisciplinary Oceans Connect project, based at Duke University in the late 1990s, sought to reinfuse regional geography with an ocean basin based perspective that privileged interaction and transformation over insularity and stasis. These ocean basin based reconceptualizations of the region may be most effective when undertaken together with small scale, place based studies. The foregrounding of the ocean as a space of connection, after all, should not lead one to forget that, for many littoral dwellers, the ocean is perceived as fundamentally local: as a set of distinct, known sites whose integration into everyday life is conditioned by place based traditions. However, by foregrounding the ocean as a space of society and culture, one can come to appreciate that even these highly localized systems exist within contexts of connection and interaction. Thus, for instance, scholars of islands (which were long considered by biologists and anthropologists as paradigmatic laboratories of homogeneity and isolation) increasingly use the ocean to reposition islands as nodes amidst channels of movement.

In other words, the ocean is being turned to by two groups of cultural geographers. Some have turned to the ocean to expand their understanding of the spaces of local life worlds. Others, attracted by the ocean's character as a literal and metaphorical space of movement, have turned to the ocean in an attempt to decenter fixed ideas of nation, state, territory, and culture. These two perspectives, however, are by no means incompatible, and much future work in marine cultural geography will likely lie at their intersection.

Environmental Geographies of the Sea

Just as historically there has been little cultural geography of the sea, similarly few human geographers until recently have focused on the ocean as a space of nature. To some extent, this likely is because, until recently, it was believed that humans could do little to influence the condition of the marine environment or the status of marine resources. Prior to the series of United Nations conferences that culminated in the 1982 United Nations Convention on the Law of the Sea, the customary law of the sea that had guided state practice was based on the belief that the ocean was immune to overfishing as well as to environmental degradation from the inflow of pollutants. For the most part, the ocean also historically has been resistant to Western style agriculture. Thus, to the extent that environmental geographers, in the wake of the early twentieth century backlash against environmental determinism, were seeking to study the mutually impacting relationship between nature and society (i.e., 'man's' impact on the 'land' as well as the 'land's' impact on 'man'), the ocean was a space of little interest.

Recently, this situation has changed, for a number of reasons. Political ecologists increasingly have become attentive to the ways in which the Western model of constructing nature through taming it, controlling it, binding it into private property, and 'improving' it (i.e., the agricultural model) is just one way among many in which societies productively interact with nature. The turn to recognize alternative environmental management, common property resource management, and tenure systems has led a number of geographers to the ocean, where management systems frequently diverge greatly from the Western model.

Additionally, the marine environment has received increased attention because of technological developments: technologies both have revealed the myriad ways in which humans have long impacted marine nature, and they have allowed for this alteration to occur at an ever greater rate. The ocean has now joined environmental topics like atmospheric pollution, climate change, and the world's water supply as a transboundary problem that requires study and action on the global scale, and a series of multilateral treaties have been signed to address this issue.

Since marine nature is now recognized as a global issue, its study has come to intersect with the study of other global issues. In addition to its inclusion within the constellation of global environmental issues, geographers increasingly are finding that of this space of global nature must be studied in the context of global social phenomena, including the rise of global economic flows, the hegemony of transnational corporations, and the ascendancy of neoliberal ideologies that govern the stewardship of nature. Furthermore, because of the high cost of oceanographic research, work in this area almost always requires governmental funding, which raises further issues for human geographers interested in the relationships between states, science, and nature. Thus, issues in the understanding and management of the marine environment relate to bigger questions regarding the global management (or construction) of nature and the relationship between global economic–environmental processes and the ways in which nation states claim control over territory and scientific knowledge.

The role of the ocean as a space for integrating human geography with the sociology of science is further enhanced by its position between two apparently contradictory trends in knowledge creation. On the one hand, the ocean, now more than ever, is the domain of expert knowledge. On the other hand, the ocean is increasingly a space of Romanticism, as fewer and fewer individuals are directly experiencing the ocean as a space of everyday life and labor, and instead are turning to the sea as a space of relaxation, exploration, and escape. Thus, lying at the intersection of increasingly intense tendencies toward both scientific rationalism and nonrational Romanticism, the ocean serves as a laboratory for scholars who are attempting to explore the (in)compatibility of these two perspectives.

Economic Geographies of the Sea

Finally, economic geographers as well are beginning to direct their attention to the ocean. Countless local and national economies are driven by the tourism sector and, within the tourism sector, a wide range of activities involve tourists embracing the ocean, whether from the shore, a ship, the surface, or beneath the waves. Economic gain is also constructed in the ocean through the extraction of living and nonliving resources. The greatest source of economic value in the ocean, however, is derived from the function that it serves as a surface for trade. The tremendous growth in the quantity (and value) of goods shipped across ocean space (increasing from 996 million tons in 1959 to 6760 million tons in 2005) has been accompanied by the advent of containerization, a technological innovation that has had far reaching implications not just for the shipping industry but also for urban form and hierarchy. Containerization has led to port consolidation and the establishment of mechanized megaports that typically are in areas removed from historic harbor fronts. This transformation, in turn, has led to urban decay in some areas and opened up new opportunities for heritage tourism and/or gentrification in others, while the port management industry itself has attracted the attention of economic geographers who focus on the way in which this industry fuses investment in place with the development of transoceanic corporate networks.

Additionally, as economic geographers (and national economies) have turned away from the production and consumption of manufactured, tangible commodities to intangible services that produce value, trade (and the space across which trade occurs) has become more central to economic geographic research. This shift in focus among economic geographers has received additional impetus from the prominence of a group of economists who call themselves 'new economic geographers'. While many geographers have criticized the conceptualization of space utilized by the new economic geographers, a central message that the group has been bringing to their fellow economists is that the spaces within which goods are produced and transported are significant factors in economic processes and cannot simply be written off as transaction costs. It may be fitting that a favorite modeling device of the new economic geographers is the iceberg function, a distance coefficient that metaphorically refers to the temporality, nature, and spatial texturing of the marine environment within which so much economic activity takes place.

Oceans of Conflict and Imagination

Overriding these turns to the ocean across the various subdisciplines of human geography are two increasingly prevalent geographies of the ocean: the construction of the ocean as a site of conflict and its construction as a space of imagination.

The rise of the ocean as a site of conflict is a direct result of its dominant management paradigm, wherein specific zones of the ocean are dedicated to different functions. The viability of this paradigm, which has its origins in beliefs about the ocean's vastness and in a strict separation between coastal and deep sea ocean uses, is being questioned today. At the global scale, the frailty of this paradigm was evidenced during the debate in the 1970s and 1980s over the establishment of a regime for extracting manganese nodules from the seabed beneath the high seas, as the United Nations struggled to develop a universally acceptable set of institutions and norms for an activity that would require state like territoriality in a portion of the ocean that historically had been exempted from territorial control. At the meso scale, the limits of the paradigm can be seen in efforts to regulate the harvesting of fish that swim between national EEZs and between these zones and the high seas. At the local scale, numerous coastal communities are attempting to manage the activities of competing local users, including those engaged in subsistence fishing, commercial fishing, sport fishing, scuba diving, shipping, and mineral and petroleum extraction.

Amidst these conflicts (and in the wake of past attempts to solve them through the spatial fix of single use zonation), the ocean is a space that leads one to question the efficacy of rationalist planning paradigms. It likely is no coincidence that two of the works of geophilosophy most frequently cited by human geographers make reference to the ocean as a space of alternative sociospatial formations. In *One Thousand Plateaus*, Gilles Deleuze and Felix Guattari (1988) write, "The sea is a smooth space par excellence," (Deleuze and Guattari, 1988) while Michel Foucault (1986), in *Of Other Spaces*, writes, "The ship is the heterotopia par excellence" (Foucault, 1986: 22–27). In both cases, the allusions to the ocean as a space of alternate ordering are metaphorical, but, like all metaphors, they gain some of their power because they resonate with what is known about the material conditions of the entity being referenced.

Thus, for human geographers, the ocean, long ignored or, at best, viewed as an arena within which social actors encounter one another or nature, now is seen as a space of society, and, as a space of society, the ocean has come into its own in human geography as a space that one can 'think with' as one attempts to understand broader structures, processes, and potentialities – on and off shore.

See also: Cultural Geography; Economic Geography; Geopolitics; Hybridity; Land Change Science; Military Geographies; Mobility; Place; Political Ecology; Political Geography; Poststructuralism/Poststructuralist Geographies; Space II.

Further Reading

Borgese, E. M. (1998). *The Oceanic Circle: Governing the Seas as a Global Resource*. Tokyo: United Nations University Press.

Braudel, F. (1972). *The Mediterranean and the Mediterranean World in the Age of Philip II*. New York: Harper and Row.

Costanza, R., d'Arge, R., de Groot, R. *et al*. (1997). The value of the world's ecosystem services and natural capital. *Nature* 387, 253 260.

Creighton, M. S. and Norling, L. (eds.) (1996). *Iron Men, Wooden Women: Gender and Seafaring in the Atlantic World, 1700 1920*. Baltimore, MD: Johns Hopkins University Press.

Deleuze, G. and Guattari, F. (1988). *A Thousand Plateaus: Capitalism and Schizophrenia*. London: Athlone.

Foucault, M. (1986). Of other spaces. *Diacritics* 16, 22 27.

Gilroy, P. (1993). *The Black Atlantic: Modernity and Double Consciousness*. Cambridge, MA: Harvard University Press.

Hartshorne, R. (1953). Where in the world are we? Geographic understanding for political survival and progress. *Journal of Geography* 52, 382 393.

Lambert, D., Martins, L. and Ogborn, M. (eds.) (2006). *Special Issue: Historical Geographies of the Sea. Journal of Historical Geography* 32, 479 647.

Linebaugh, P. and Rediker, M. (2001). *The Many Headed Hydra: The Hidden History of the Revolutionary Atlantic*. Boston, MA: Beacon Press.

Mansfield, B. (2004). Rules of privatization: Contradictions in neoliberal regulation of North Pacific fisheries. *Annals of the Association of American Geographers* 94, 565 584.

Olivier, D. and Slack, B. (2006). Rethinking the port. *Environment and Planning A* 38, 1409 1427.

Psuty, N., Steinberg, P. E. and Wright, D. (2004). Coastal and marine geography. In Gaile, G. & Willmott, C. (eds.) *Geography in America at the Dawn of the 21st Century*, pp 314 325. Oxford: Oxford University Press.

Semple, E. C. (1911). *Influences of Geographic Environment*. New York: Henry Holt.

Sidaway, J. D. and Power, M. (2005). 'The tears of Portugal': Empire, identity, 'race', and destiny in Portuguese geopolitical narratives. *Environment and Planning D: Society and Space* 23, 527 554.

Steinberg, P. E. (ed.) (1999). *Special Issue: Geography of Ocean Space. The Professional Geographer* 51, 366 450.

Steinberg, P. E. (2001). *The Social Construction of the Ocean*. Cambridge: Cambridge University Press.

Strabo (1917). *Geography*. Cambridge, MA: Harvard University Press.

United Nations (1983). *The Law of the Sea: United Nations Convention on the Law of the Sea with Index and Final Act of the Third United Nations Conference on the Law of the Sea*. New York: United Nations.

Van Dyke, J. M., Zaelke, D. and Hewison, G. (eds.) (1993). *Freedom for the Seas in the 21st Century: Ocean Governance and Environmental Harmony*. Washington, DC: Island Press.

Wigen, K. and Harland Jacobs, J. (eds.) (1999). *Special Issue: Oceans Connect. The Geographical Review* 89, 161 313.

Relevant Websites

http://aag coma.hometead.com
Coastal and Marine Geography Specialty Group of the Association of American Geographers.

http://www.fao.org/fishery
Fisheries and Aquaculture Department of the Food and Agriculture Organization of the United Nations.

http://www.imo.org
International Maritime Organization.

http://www.unctad.org/Templates/Page.asp?intItemID=2618&ang=1
Review of Maritime Transport (annual publication), United Nations Conference on Trade and Development.

http://www.un.org/Depts/los
United Nations Office of Ocean Affairs and the Law of the Sea.

Olsson, G.

M. Gren, Hólar University College, Sauðárkrókur, Iceland

Glossary

Cartographical Reason To form and operate with concepts of abstraction in accordance with a geometric rationality and a logic of mapping.

Geography of the Invisible Unlike a traditional geography which has the representation of the visible Earth as original object, the aim here is to map phenomena without physical extensionality, like the world of human thought and imagination.

Human Interaction A broader term for interaction between humans resulting in common worlds than spatial interaction, which denotes interdependence between geographical areas.

Invisible Maps Relations in language that although they cannot be seen provide points, lines, and planes as orientations of and for thought.

Ontological Transformation To change what something is into something else, like when a physical object appears as a linguistic sign or a city plan turns into buildings and roads.

Taken-for-Granted Things and relations as unaware premises for thought and action.

Thought-and-Action A term that highlights that in the medium of language thought and action are internally and inseparably related.

Olsson, Gunnar (1935–)

Gunnar Olsson is a Swedish geographer and author of numerous books and articles (**Figure 1**). He was born in a small village in the county of Värmland (Sweden) where his mother was a local midwife and his father had a business in the timber and building trade. Since the mid 1970s he has been living in Uppsala (Sweden) with his wife Birgitta. Together they have two daughters, both with children on their own.

In 1957 Olsson began his academic studies at Uppsala University where, since 2000, he has been professor emeritus in the department of social and economic geography. In between he has been on a lifelong journey of self conscious reevaluations, pursuing his core theme of human interaction in search of its geographical es sences. Marked by a refusal to limit himself to the study of visible things, and a desire to be abstract enough, Olsson has continued to develop his geography of the human territory of language, meaning, power, and im agination – an original yet consistent contribution to one of the most fundamental domains of human geography (i.e., a theory of humans as geographical and carto graphical creatures).

In 1960 Olsson received his bachelor's degree in human geography, economic history, and political science and was awarded a 2 year fellowship by the Faculty of Social Sciences in Sweden, which would lead to several visits and stays at geography departments in the USA. Knowing that he wanted to use a quantitative approach, Olsson became deeply involved with theoretical devel opments within spatial science, especially attempts to translate earthly human behavior, such as migration, into the abstract language of distance and human interaction. Technicalities aside, his work with his licentiate and doctoral theses in the 1960s suggested that spatial interaction models designed to catch distance and migration in fact revealed more about the spatial distri bution of opportunities than about human interaction *per se*. At stake was the classic geographic inference problem between form and process, that is, how one should con ceive the relations between a spatial pattern and its

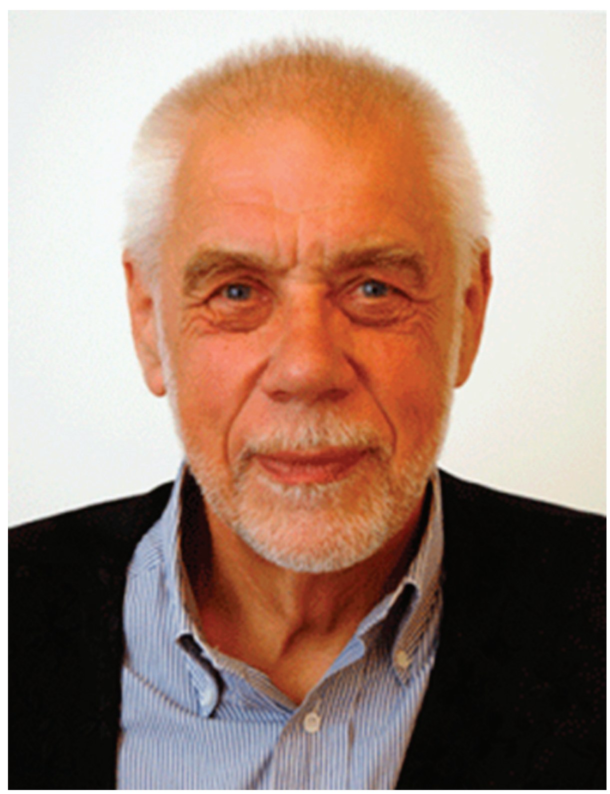

Figure 1 Gunnar Olsson.

generative processes. According to Olsson, both his own studies and spatial science more generally had shown that the same spatial form can be generated through drastically different processes. Even the most perfect description of a spatial pattern cannot be used as the basis for inferences about how and why it came about. Although it is sometimes possible to reason from process to form, the opposite is never appropriate. The conclusion was as inevitable for Olsson as it was devastating for human geography as spatial science: in principle it was not valid to draw inferences about human behavior from a mapped distribution, to reason from form to process.

For Olsson, applied geography (founded on the utopian vision that a better and more just society could be constructed on the basis of exact scientific knowledge of human spatial behavior) therefore involved a faulty reasoning. According to the principles of utilitarian ethics, purposeful action should be judged in terms of its consequences, but according to the principles of causal theory the consequences of actions based on observed autocorrelations cannot be foreseen. Of importance for Olsson were examples of planning in Sweden during the 1960s and 1970s that were based on a practical application of the gravity model but which, due to its internal language and contrary to political intention, conserved instead of alleviated existing inequalities in terms of access to welfare resources. For Olsson, this was an example of planning (a special case of human thought and action) as 'Greek tragedy'. For him, this was a case of good intentions turning into tragic results, but with no one to blame because during the journey all actors had merely remained faithful to their own internal reasoning.

In 1966 Olsson was appointed by the University of Michigan, USA, and was soon promoted to full professor (a sign of how his work in spatial science was valued because this was actually before he had completed his doctoral thesis in Sweden in 1968). In what was going to be 10 years at the department of geography at Ann Arbor, he continued to investigate the relations between theory, model, and reality through in depth engagements with theories of science, language, alternative logics, dialectics, epistemology, ontology, ethics, esthetics, and other philosophical issues. The seemingly objective methodologies of spatial science, including correlation and regression techniques, were in fact underpinned by a host of philosophical assumptions. Quite famously, he came to claim that allegedly objective analyses usually reveal more about the language they are written 'in' than the phenomena they are supposed to speak 'about'. For example, formal scientific language tends to privilege truth preservation, to make us assume that propositions are either true or false. However, in the language of practical reasoning the opposite is often the case: in this realm, meaning varies and changes with context, and we can make false what is true.

As a runaway spatial scientist in the 1970s, Olsson was one of the very first geographers to engage in depth with questions of language and thus made contributions to the development of humanistic geography. His early engagement with thinkers such as Jacques Derrida and Jacques Lacan was later to appear as deconstruction and post structuralism in the hands of a younger generation of (post)modern geographers. But Olsson, having reached the conclusion that there is no point of understanding outside of language, refused to side either with certainty or with ambiguity. Accepting that the human condition is a constant struggle between opposites and ultimates he became a practicing dialectician, seeking to place himself – like the God Janus – in the excluded middle where he wrestled with paradoxes and performed experiments with, on, and at the limits of language.

In 1977 Olsson moved from Ann Arbor to Stockholm, where he was appointed to the chair of economic geography and planning at the Nordic Institute in Urban and Regional Planning (NORDPLAN). During the 20 years he carried out the responsibility for an ambitious doctoral program, with students from the Nordic countries. As always, his work continued to be intimately related to his teaching and extensive lecturing. No surprises, then, that for a decade or so questions of ideology, socialization, politics, power, and planning came to the fore.

During the 1990s he made more explicit attempts to foreshadow a cartography of thought, to explore how humans manage to find their way not just in the landscapes of physical existences but of mental subsistence. Even though we automatically know how to speak and think, we do not know how to explicate the reasoning rules which govern that thinking and speaking. Hence, Olsson's oft repeated question: "How do I know the difference between you and me and how do we share our beliefs in the same?" The challenge was to be abstract enough and to draw maps of this invisible realm in such a way that they were not about something, but that something itself. Thus, Olsson's short answer: "we find our way in our world of language and imagination, always filled with power, by drawing on invisible maps of the invisible."

Among the premises that were to guide Olsson's mapping expeditions are that human beings are semiotic animals living in the border between the five senses of the body and the sixth sense of culture. In that abyss humans are held simultaneously together and apart by their usage of signs that enable and involve ontological transformations between material and immaterial realms. Expressed in a minimalist rendering characteristic of Olsson, human knowledge takes the translational form of $a = b$. Telling the truth is both to say that something is something else and to be believed by others. By using the epistemological operator of 'as if', humans are able to treat also their untouchable world of reasoning as if it was like the touchable. Their logical and rhetorical reasoning also involves a geometric bias, which means that they tend to privilege such reasoning.

So what finally unfolds, then, is geography as geometry with names and Olsson's understanding of modern reason as a form of cartographical reason (a term first coined by the Italian geographer Franco Farinelli). At least behind Western and European thinkers and philosophies, so Olsson claims, there has been a taken for granted geo metric coordinate net of lines, points, and planes which implicitly has enabled them to cartographically capture and create their worlds by merging mapping with stories about what it means to be human in particular ways. Hence his most recent work, a minimalist guide into what he imagines to be the landscape of Western culture. Starting with the key geographical coordinates – above, below, here, and there – he eventually constructs an atlas of crucial events where the boundaries of human territory have been created and delimited between the outer limits of mindscape and rockscape.

Within such later works are the old themes which Olsson never left (i.e., human interaction) but now in the form of a cartographical reason which is becoming in creasingly outdated. Yet, it is likely to continue to haunt us every now and then, here and there. Entrenched as this reason is in the practice of 'thingification', the Olssonian challenge for geographers remains: to be abstract enough.

See also: Applied Geography; Cartography, History of; Critical Geography; Culture; Deconstruction; Dialectical Reasoning and Dialectical Materialism; Geographically Weighted Regression; Humanism/Humanistic Geography; Mapping, Philosophy; Maps; Philosophy and Human Geography; Positivism/Positivist Geography; Postmodernism/Postmodern Geography; Poststructuralism/Poststructuralist Geographies; Quantitative Methodologies; Quantitative Revolution; Space-Time Modeling; Spatial Analysis, Critical; Spatial Autocorrelation; Spatial Interaction Models; Spatial Ontologies; Spatial Science; Translation.

Further Reading

Doel, M. (2003). Gunnar Olsson's transformers: The art and politics of rendering the co relation of society and space in monochrome and technicolor. *Antipode* 35, 140 167.
Farinelli, F. (2003). *Geografia: Uníntroduzione ai Modelli del Mondo*. Torino: Einaudi.
Gren, M. (1994). The invisible geography of Gunnar Olsson. In Gren, M. (ed.) *Earth Writing: Exploring Representation and Social Geography In between Meaning/Matter*. Gothenburg: Department of Human and Economic Geography, University of Gothenburg.
Olsson, G. (1965). *Distance and Human Interaction: A Review and Bibliography*. Philadelphia, PA: Regional Science Research Institute.
Olsson, G. (1968). *Distance, Human Interaction and Stochastic Processes: Essays on Geographic Model Building*. Uppsala: Avhandling for filosofie doktorsgraden, Uppsala universitet.
Olsson, G. (1980). *Birds in Egg/Eggs in Bird*. London: Pion.
Olsson, G. (1990). *Antipasti*. Goteborg: Korpen.
Olsson, G. (1991). *Lines of Power/Limits of Language*. Minneapolis, MN: University of Minnesota Press.
Olsson, G. (1993). Chiasm of thought and action. *Environment and Planning D: Society and Space* 11, 279 294.
Olsson, G. (2002). Glimpses. In Gould, P. & Pitts, F. (eds.) *Geographical Voices: Fourteen Autobiographical Essays*. New York: Syracuse University Press.
Olsson, G. (2007). *Abysmal: A Critique of Cartographic Reason*. Chicago: University of Chicago Press.
Philo, C. (1984). Reflections on Gunnar Olsson's contribution to the discourse in contemporary human geography. *Environment and Planning D: Society and Space* 2, 217 240.
Philo, C. (1994). Escaping flatland: A book review essay inspired by Gunnar Olsson's lines of power/limits of language. *Environment and Planning D: Society and Space* 12, 229 252.
Pickles, J. (2004). *A History of Spaces: Cartographic Reason, Mapping and the Geo Coded World*. London: Routledge.

Oral History

M. Boyle, National University of Ireland Maynooth, Co Kildare, Republic of Ireland

Glossary

Archive A repository holding information and data about past events and people but often organized, stylized, and presented according to the motivations, aspirations, resourcing, and politics of the archivist.

Life History A method of collecting, analyzing, and presenting self-reflective commentary on the biography of a person over his/her life course, prioritizing how that biography is characterized and accounted for.

Oral History A method of collecting, analyzing, and presenting first-hand testimonies from those who have lived through both dramatic historical events and more prosaic everyday pasts, with a view to enriching and enlivening knowledge produced by traditional historiographical methods.

Oral History Archive A repository holding transcripts, tapes, or films of oral history interviews largely unmediated by the archivist and often in itself a valuable 'raw' community resource.

Social Memory The idea that memory is far from a mental or cognitive process operating 'under the skull' and instead is better thought of as bearing the stamp of the social, political, economic, and cultural conditions in which recollection takes place.

Introduction

Geographical research is often interested in describing and explaining events which occurred in the past. But how can one come to know the past? In common with the discipline of history, much of the recent focus has been on the challenges surrounding the use of official documents organized, classified, and packaged in the form of public and private archives. But geographers make use of a number of other tools and methods of historical enquiry including chemical dating techniques, landscape interpretation, archeological and anthropological excavation, and photographs and other visual sources. Today the value of unearthing past events through the soliciting of oral histories from participants in those events is also being explored. Where these oral histories have focused upon the biographies of participants they have often been referred to as life histories. The data produced through the deployment of the oral history method has served to augment, enrich, consolidate, develop, and critique findings which would otherwise have been produced through orthodox techniques of historical retrieval.

This article seeks to introduce the oral history method to those who might be interested in collecting such histories to enliven geographical research. It is structured around two themes; the practice of oral history and oral history as a social practice. Oral history is often thought of as a valiant attempt to recover the voice of marginalized and powerless groups written out of 'professional' or 'official' histories. Its value lies in its ability to unearth histories 'from below'. In the first instance some illustrations of the ways in which oral histories have been used to better understand past events will be provided. Attention will be given to the issues researchers need to be aware of when 'doing' oral history. The status of the knowledge produced by the oral history method however has recently been open to question and the method has entered a new interpretive phase. In the final section, it will be contended that oral history produces as much as gathers memories; it is more interpretive than reconstructive. As such it needs to be thought of as a situated social practice.

The Practice of Oral History

The archive serves as a key method through which historical geographers secure knowledge about past events and processes. But the archive is clearly limited as a research tool. The range of voices it includes is out of necessity restricted and most commonly it is the powerful, the elite, the bourgeoisie, the dominant, and the privileged who get to speak. And of course it is the same powerful, elite, bourgeoisie, dominant, and privileged who get to classify, order, file, and control access to the official record. What stands as professional history then is in fact often a version of the past produced, circulated, policed, and consumed by and for select strata of society.

It is from this point of departure that advocates of the oral history method begin. As a method of data collection, oral history has a rich lineage within historiography and indeed more generally within the social sciences. While the basic approach can be found in nineteenth century historical manuscripts, and the phrase itself can be traced to professional historical research conducted in the United States in the 1940s, it has only been over the past 30 years that the method has taken flight. A cursory glance at its development reveals the breadth of historical phenomenon that oral histories are now being used to help illuminate. These include childhood, popular culture, religion, entertainment, leisure, education, the workplace, the home and domestic life, family and sexual relations, communities and change, politics, biographies, human and natural disasters, and migration.

Oral history, it is argued, has a role to play in prising open a more rounded social history of the past, re covering the unfolding of historical events from the vantage point of those normally excluded from the re cord. The collection of first hand testimonies from those who lived through and experienced both dramatic and mundane pasts adds a richness to historiography that has hitherto been lacking. Not only does oral history provide a 'history from below', it brings emotions, feelings, atti tudes, ideology, and opinions to bear on otherwise dry and arid historical narratives. Moreover it can scale memory to finer levels of resolution; the trials and tri bulations of everyday life can be excavated.

Consider for instance the case of research interested in understanding the migration of people from Pakistan to the United Kingdom in the 1960s. Public and private archives would clearly provide some insight into such a migration flow; the numbers of migrants could be counted and their socioeconomic characteristics profiled, the location of origin and destination chronicled, the welfare benefits, housing characteristics and domestic facilities accrued revealed, health inequalities mapped, the levels of assault, violent attack, and crime migrants are both subjected to and culpable off, and so on. But such a historiography would only capture what official records at a particular point in space and a particular point in time perceived as being worthy of noting.

Oral history provides a potential means of enriching understanding of this migration flow. Interviews with first and second generation Pakistani migrants who settled in the United Kingdom might provide insights into a whole new range of research themes. What conditions prevailed in Pakistan prior to departure? Why was flight necessary? What calculations were made in terms of where to go and how to get there? What did it feel like to leave? What happened during the transit? What fears, frustrations, hopes, and aspirations did migrants carry? How did they cope? What were their feelings toward the United Kingdom on arrival? How did they encounter the official systems? How easy was it for successful applicants to settle into the United Kingdom? Did settlement experi ence vary by age, gender, class, sexuality, and according to place of settlement? How do they now feel about home? What trans national networks did they establish and how did these evolve through time? What happened to those who were refused entry? How did their lives turn out? Where did they go?

Doubted for its rigor and often thought of as a poor relation to professional history, oral history has become preoccupied by questions of objectivity, accuracy, and representativeness. The practice of collecting, ana lyzing, and presenting oral histories has demanded that rigor be applied when deciding upon sampling strategies; compiling questionnaire; recruiting and preparing re spondents; conducting the interview; recording, tran scribing, and storing the interview; resolving ethical issues; and using the product:

1. Sampling strategies – Which sample is to be focused upon? What biases might the selection of this sample imply? Is breadth and representativeness to be pri oritized or is qualitative insight and understanding of greater import?
2. Questionnaire design – Is a questionnaire to be used at all? Which kinds of questions are appropriate? How might one avoid asking loaded questions? How open should interviews be and how much latitude should be given if discussion wanders off message?
3. Recruiting and preparing respondents – How is the sample to be approached and solicited? Who might broker this access? How can trust be built up with participants? How might access issues affect who is to be included in the sample? Are financial or other in ducements to be offered? How much should partici pants be told about the project in advance?
4. Conducting the interview – Where should interviews be held? Who should be present? What difference does location make to the conduct of interviews and the quality of insights generated? Should interviews always be held in the most comfortable environment for the interviewee?
5. Technology – What technology is best equipped to record, store, display, and transcribe testimonies? Should visual and oral technologies be used? How detailed should transcripts be and how are nonverbal communication prompts to be recorded? Where and how should interviews be stored?
6. Ethics – Oral histories tend to divulge more intimate and personal details than other methods. What ethical challenges does the method pose? How far should participants be invited to amend or review or develop existing transcripts? How far should transcripts and published accounts be annonymized?
7. Using the product – How should oral history testi monies be analyzed ? How should the data be pre sented? If deployed in film, radio, theatre, or textual outputs, ought one to adjust the method of analyses and presentation and if so how?

Oral History as a Social Practice

More recently, oral history has entered into a more in terpretive mode and new questions are being asked of it. How is the information generated through oral histories to be valued? What status is to be afforded it? Do oral histories really serve as a heroic means of recovering the past as lived and experienced by the powerless and the marginal? Are they of necessity biased, mediating the past by their very existence? Should we even be talking about the truth, validity, and representativeness of the testimonies solicited?

One response to this question is to lament the presence of 'faulty memory' in oral testimonies. It is unrealistic for people to recount a faithful representation of what happened on a particular day or week or month or year – especially many years after the event took place. Retrospective accounts will be plagued by re orderings of memory, lapses of memory, nostalgia, repression, and possibly even outright amnesia. Cognitive failure is inevitable with age and the truthfulness and therefore value of oral histories needs to be severely questioned. They are simply not objective enough to be taken seriously by professional historiographers.

The phenomenon of 'faulty memory' however requires more detailed investigation. Over the past decade, scholarship in a variety of disciplines has sought to critique orthodox conceptions of memory developed in mainstream psychology. In contrast to what might be referred to as 'cognitivist' approaches to memory, which privileges and indeed take for granted the concept of 'Cartesian memory', recent work has sought to draw attention to the social structuring of memory. Oral histories cannot be treated as an objective account of past events, not because cerebral failures and cognitive impairment operate under the skull to frustrate faithful recollection of the truth but because the oral history method is itself a situated social practice.

How can something so private as memory be conceived as a social activity? Recent developments in the conception of the social foundations of memory have pointed to the role of cultural repertoires – languages and more generally semiotic systems – in the appropriation of the past. Cultural repertoires furnish populations with a 'vocabulary' through which past events can be remembered and as such mediate between peoples and their past. Social memory in this context would refer to the ways in which populations, socialized into different cultures, possess different techniques of historical recovery. These techniques prove to be both enabling and constraining insofar as they furnish a competence to remember but only within certain conceptual schemes. From the most subtle everyday accounting practices to more organized commemorative practices, such as theatre, museums, folklore, music, poetry, literature, photography, marching, and parades, cultural repertoires present populations with a grid through which temporality can be tamed and understood.

But the limits imposed upon memory by different forms of commemorative practice is only part of the story. Maurice Halbwachs (1951), one of the earliest theorists of social memory, invites readers to log the number of memories recalled during a day which are the product of questions put by others, or preparations made by readers in anticipation of requests from others. In response to this rhetorical question, he argues that it is indeed primarily through negotiation with others that memories are affirmed (remembered), rejected (forgotten), or reshaped. Furthermore, and crucially, the shaping of memory through conversation needs, he argues, to be situated within its proper social context. In Halbwachs' study, for instance, memories are seen to be incited principally in relation to participation in certain social groups (those of kin, religion, and class). Since the bulk of memory work is of this sort – what he calls 'remembering together' – Halbwachs argues that it is meaningless to think of memory as being the possession of an individual as such. Instead, it is better conceived as a 'social practice.'

Memory can be said to be a social practice in so far as it performs certain kinds of functions in relation to present conditions. Questions about the status of knowledge produced through oral history interviews then must proceed by questioning what this particular act of memory gathering, negotiation, production, reproduction, and ordering accomplishes at different points in time and space. Here, the social relationships which exist between interviewer and interviewee, mediated by social class, gender, age, ethnicity, disability, cultural capital, and so on, would need teasing out. The location of the interview and the agenda behind the interview would need reflecting upon. The embeddedness of the practice within the university setting and the discipline of history or geography would emerge as important. And of course the prevailing political, economic, social, and cultural conditions in which the discussion is situated would need critical inspection.

Instead of enquiring into how accurate oral histories are or how faithful they recount past events, it is necessary to ask what function the practice of oral history performs, given the social context in which this practice takes place. Oral histories do not capture past events, rather they reveal the current platforms from which memory takes place. The oral history interview is generative – co authoring memories – rather than performing as a ventilator – neutrally bringing preexisting memories to the surface. It is at root a 'political practice'. The oral history method itself and not the memories, testimonies, or accounts it generates becomes the object of reflection, critique, and deconstruction. This commemorative practice tells us as much about the politics of how pasts are remembered as it does about the truth of those pasts.

To return to the example outlined above then, oral history interviews with Pakistani migrants who moved to the United Kingdom in the 1960s according to this framework cannot be thought of as accessing some truth about what it was actually like to make this move. Instead, these interviews need to be approached as a proactive sifting, sorting, filtering, and reconstitution of memories which bear the stamp of the moment. They tell us more about how the ongoing war on terror and tensions be tween the West and the Islamic world, the experiences of integration and socioeconomic advancement in different parts of the United Kingdom, the age, sex, ethnicity, religion, and nationality of the interviewer, the location of the interview, and the funding of research projects. In other words, prevailing social, cultural, economic, and political conditions cultivate certain ways of remem bering and forms of recollection. The interview is not a passive vessel innocently communicating memories but instead betrays the social, cultural, political, and eco nomic climate of the day.

See also: Archives; Historical Geography; Memory; Phenomenology/Phenomenological Geography.

Further Reading

Halbwachs, M. (1951). *The Collective Memory*. New York: Harper and Row.

In the United Kingdom, the journal Oral History produced by the Oral History Society has emerged as the standard outlet for research in the field. In the United States meanwhile Oral History Review published by the Oral History Association functions as a key reference point. More generally the International Oral History Association, formed to coordinate oral history research taking place around the world, publishes oral history research in its journal Words and Silences. Meanwhile, Thomson's (2000) The Voice of the Past, Perks and Thompson's (2006) The Oral History Reader, Lummis' (1987) Listening to history: The authenticity of oral evidence, and Ritchie's (2003) Doing Oral History: a practical guide are all useful starting points. For commentary on the use of oral history in Geography specifically see Riley and Harvey (2007) and Ni Laoire (2007). An interesting case study of the use of oral history in the study of the history of geography itself is provided by Jenkins and Ward (2001).

Jenkins, A. and Ward, A. (2001). Moving with the times: An oral history of a geography department. *Journal of Geography in Higher Education* 25, 191 208.

Lummis, T. (1987). *Listening to History: The Authenticity of Oral Evidence*. London: Hutchinson.

Ni Laoire, C. (2007). *A biographical and life course perspective on Ireland's recent return migration*. Critical Public Geographies Research Papers: University of Cork.

Perks, R. and Thompson, A. (eds.) (2006). *The Oral History Reader* (2nd edn.). London: Routledge.

Riley, M. and Harvey, D. (2007). Editorial: Talking landscapes: On oral history and the practice of geography. *Social and Cultural Geography* 8, 345 351.

Ritchie, D. A. (2003). *Doing Oral History: A Practical Guide*. Oxford: Oxford University Press.

Thomson, P. (2000). *The Voice of the Past* (3rd edn.). Oxford: Oxford University Press.

Oral History, Ecological

A. Nightingale, University of Edinburgh, Edinburgh, UK

Glossary

Ethnobotany The collection of local or traditional knowledge of plant names and qualities. Most often used in research projects in the tropics or other places where the flora is not well known. Researchers gather information on plants from local experts and compare it to scientific understandings.

Key Informants Research participants who are considered to have expert or in-depth knowledge on the subject of the research. Often they are community leaders or otherwise recognized as experts by other community members. Less commonly, key informants refer to research participants who have a particularly good relationship with the research team, and thus the quality of their interviews is considered very high.

Taxonomy The system of nomenclature used for the classification of plant and animal species. Scientific taxonomies use a standardized Latin systems, whereas ethno-botany seeks to produce new taxonomies based on local languages and knowledge.

Traditional Ecological Knowledge Knowledge of the natural world held within local traditions, including myths, daily practices, and beliefs. Often traditional ecological knowledge is based upon daily experience of living within or working with a particular ecosystem or species. Researchers working in less-developed contexts have become particularly interested in the knowledge local people can provide on their ecosystems that may not be known to science. While traditional ecological knowledge is often assumed to rest on historical observations and beliefs, it does not necessarily need to have a historical aspect.

Ecological Oral History

Ecological oral history is a type of oral history that seeks to document knowledge about the natural world that is held within people's memories and cultural beliefs. The methodology is used by researchers exploring traditional ecological knowledge (TEK), ethno botanists interested in local plant taxonomies, and by those who investigate competing claims about environmental change. Most often ecological oral histories are collected from key informants or local people considered to be experts about a particular species, habitat, or place. They have been used in geography, anthropology, and related subjects to document what are often seen to be threatened systems of knowledge.

Ecological oral histories are similar to other types of oral history in that they generally document knowledge and traditions that by definition are not written down. As a result, they have been used extensively in the developing world where local people are believed to carry nonwritten knowledge of their ecosystems, based upon long experience and observation of that place, rather than scientific investigation. Oral histories can be a powerful tool to gather information about these knowledge systems. The distinction between traditional knowledge and scientific knowledge is contested, however, as it implies that traditional or local ecological knowledge is not systematic or based on careful observation. On the contrary, many ecological oral histories have revealed local knowledge about the natural world that is highly systematic and differs only from scientific knowledge in the theoretical assumptions (or beliefs) behind them. Therefore, many studies that employ ecological oral histories are based upon the notion that local knowledges are of equal value to scientific knowledges. Importantly, many traditional systems of knowledge do not draw the same boundaries between human culture and the natural world. Australian Aborigines, for example, do not distinguish between Dream Time and the present, thus populating the landscape with ancestors and other spirits making their view of the environment radically different from scientific views. Thus, ecological oral histories can provide insight into local cultures as well as alternative ways of viewing and managing the environment. It is these insights that researchers hope to document through using oral history methods.

Local Systems of Ecological Knowledge

It is important to distinguish between ecological oral histories that seek to document knowledge of ecosystems that is considered to be general, and the ecological oral histories that seek to understand how a particular person or group of persons view their environment. While the distinction may seem subtle, in practice, the methodological and theoretical implications are significant. Projects that aim to document general systems of knowledge assume that it is something independent of the person conveying it. The researcher would therefore seek to gather information from the most knowledgeable people in the community. This kind of research in particular is often concerned with providing an alternative to scientific understandings of environments and to docu ment traditional systems before they are lost. The worry is that with increasing contact from the outside world, indigenous cultures are losing their traditional know ledge. Methodologically, the researcher would thus interview a few key informants and attempt to rectify any inconsistencies.

Oral histories that gather general knowledge are therefore rarely personalized. Instead, the traditions are ascribed to the entire community or group of people believed to hold those beliefs. There has been a great deal of concern over the past 30 years that many traditional knowledge systems are being lost as people come into contact with Western societies. There is evidence from all over the world that people are forgetting or dismissing traditional understandings of environments and as a consequence this local knowledge is no longer passed onto the next generation. Researchers have therefore used oral histories as a way to document and preserve these traditional knowledges. This type of ecological oral his tory is used both for general information about ecosystem processes and also to collect ethno botanical information. Ethno botanists document the local names and uses of plants with a focus on discovering new plants and new uses of plants not previously known. This kind of research assumes that a system of knowledge exists at the local level that is different from a standardized scientific one and importantly that the local knowledge can inform science.

Personalized Ecological Oral Histories

Research that seeks to gather more personalized know ledge, however, is less concerned with demonstrating the existence of a nonscientific logic to local knowledge, but rather to understand how different people interact with and view their environment. The theoretical assumption is that knowledge is produced in context and therefore an abstract, general system of knowledge will not necessarily be found. Most research of this kind does not refute the existence of local knowledges, but rather sees them as dynamic processes dependent on context. In these kinds of oral histories, it is often more important to explore why people might tell different narratives about their en vironments than it is to look for consistencies in them. Methodologically, in this type of project a researcher would gather oral histories from a wide variety of people and analyze them both for their overlap and their gaps. Key informants would not be sufficient in this kind of project since local disagreements over aspects of the en vironment are important. These kinds of research projects are less concerned with documenting a fading traditional knowledge base but rather want to understand the social politics within which knowledge is produced and contested. Ecological oral histories in this context can be very useful for understanding the basis of different claims to land and landscapes. They are also useful for under standing the rhetorical and symbolic ways that local people resist the modernizing influence of development.

Methods of Ecological Oral History

Ecological oral histories thus can be used for a wide variety of research projects. Data collection techniques often follow those used for other kinds of oral histories, although some novel methods are used as well. Many ecological oral histories follow techniques used in oral history and therefore use in depth, semistructured inter views. In this type of interview, the researcher develops a set of questions to be asked, but does not specify the order in which they will be asked. Rather, the researcher strives to create a conversation with the participant(s) and asks questions where relevant or as a prompt to begin or redirect the interview. Unlike oral histories, however, ecological oral histories often utilize more interactive methods. Interactive methods include ambulatory inter views, mapping, and discussion of photographs. In am bulatory interviews, the researcher and informant or a small group of informants walk through a piece of land and identify plants or discuss changes and other features of the landscape. These kinds of interviews are often very effective and yield information not only on the landscape, but also on the relationship between the participants and the environment (i.e., nature–society interactions). They are used extensively in studies seeking to document TEK and also in research seeking to understand the social–political dynamics of land use. Other forms of interactive interviews use mapping or photographs of places. Par ticipatory mapping is often considered a separate research technique in its own right and is used in nature–society studies or land use planning research projects in addition to ecological oral histories. In these interviews, re searchers ask people to draw a map of a place and then discuss what is included. Mapping provides a unique picture of landscape as seen from the perspective of the research participants. It can also illuminate important information on landscape features that may not be obvi ous to the researchers or the relative importance of par ticular resources. Photos of places are similarly used to invite discussion. Researchers either use photographs they have obtained or give participants cameras to take photographs of places or objects of importance to them. The researcher then discusses the photographs with the participants in order to lend a focus to the discussion. All these types of interviews can be done with groups as well as individuals. Group interviews are useful for obtaining accurate information on plant names and other aspects of shared knowledge of ecosystems, but they are equally useful for gathering information on the social politics of environments. A group context can help to illuminate which inconsistencies in oral histories are based on a lack of knowledge and which are based on competing claims. Members of the group correct each other or arrive at a consensus through discussion, helping to ensure accuracy as well as illuminating other dynamics that may be useful to the research.

Studies That Have Used Ecological Oral Histories

A number of studies in political ecology and related subdisciplines in geography have used ecological oral histories. For example, a study in western Nepal used ecological oral histories in order to explore past eco logical conditions and forest management practices. As such, the aim was both to gather TEK and to explore the social politics of forest management. Individual semi structured and ambulatory interviews were both im portant in generating a picture of the forest historically and in triangulating accounts on past management. Both individual and group oral histories were done and the data gathered from them was triangulated with other types of data in several different ways reflecting the aims. The oral histories thus asked questions both about con ditions in the forest past and present and about how it was managed. There are no other sources of information on the forest for that part of the country so the ecological oral histories provided one way of gaining some sense of forest change. Similarly, the accounts of past management were surprisingly consistent providing a rich and accur ate picture of historical practices.

In this study, the ecological oral histories were also used to help understand current management struggles. In most of the interviews, people compared past man agement to the present and evaluated how they were better or worse. For some people, the present system is better, whereas others lamented the loss of a system they thought was fairer. Analysis of group interviews was particularly useful for this research question, as people debated important aspects of both systems and argued about what had improved. Ambulatory interviews were an especially effective context for group interviews and helped expose areas of the forest that were contested and boundaries that were perhaps not as fixed as they ap peared on the official map.

Participatory mapping was tried in this study but was not successful due to the association between literacy, power, and the ability to draw. Illiterate participants asked others to draw maps for them, undermining the point of the exercise which was to collect people's dif ferent perceptions of the landscape. Other studies, how ever, have used mapping with excellent results. A study in Africa used mapping to understand land use patterns by having different groups of women draw maps in the dirt using sticks, stones, and other objects. The maps were used as a focus for discussion and also as evidence of different kinds of land use. In this context, the mapping exercise yielded rich insights and also provided a way to give back to the participants. The women found the process of making maps empowering and gave them a helpful way to think about their land use practices. In both of these studies, map making was used to under stand the landscape as well as the politics of knowledge about the landscape. The map making process revealed much about conflicts over particular pieces of land, knowledge of different species, and qualities of the landscape, as well as the power relations inherent in drawing a map. They are good examples of how geog raphers have used oral histories and participatory methods in research on nature–society questions.

See also: Environment, Historical Geography of; Oral History; Participatory Action Research; Triangulation.

Further Reading

Cancel, R. (1999). Narrating our pasts: The social construction of oral history. *Language in Society* 28, 107 110.
Davis, A. and Wagner, J. R. (2003). Who knows? On the importance of identifying 'experts' when researching local ecological knowledge. *Human Ecology* 31, 463 488.
Denzin, N. K. and Lincoln, Y. S. (eds.) (1998). *Collecting and Interpreting Qualitative Materials*. Thousand Oaks, CA: Sage Publications.
Finnegan, R. (1992). *Oral Traditions and the Verbal Arts: A Guide to Research Practices*. London: Routledge.
Harvey, D. and Riley, M. (2005). Country stories: The use of oral histories of the countryside to challenge the sciences of the past and future. *Interdisciplinary Science Reviews* 30, 19 32.
Johannes, R. E., Freeman, M. M. R. and Hamilton, R. J. (2000). Ignore fisher's knowledge and miss the boat. *Fish and Fisheries* 1, 257 271.
Krech, S. (2005). Reflections on conservation, sustainability, and environmentalism in indigenous North America. *American Anthropologist* 107, 78 86.
Nightingale, A. (2003). A Feminist in the forest: Situated knowledges and mixing methods in natural resource management. *ACME: An International E Journal for Critical Geographers* 2, 77 90.
Rocheleau, D. (1995). Maps, numbers, text and context: Mixing methods in feminist political ecology. *Professional Geographer* 47, 458 466.
Tonkin, E. (1995). *Narrating Our Pasts: The Social Construction of Oral History*. Cambridge: Cambridge University Press.
Williams, N. and Baines, G. (1993). *Traditional Ecological Knowledge*. Canberra: Australian National University.

Relevant Websites

http://www.ohs.org.uk
Oral History Society.
http://www.bl.uk
The British Library, The world's knowledge.

Orientalism

M. Haldrup, Roskilde University, Roskilde, Denmark
L. Koefoed, Roskilde University, Roskilde, Denmark

Introduction

Orientalism has three distinct meanings. First, it refers to a specific esthetic movement, especially in painting and also more generally within the eighteenth and nineteenth century artwork, which after the French conquest of Egypt sought inspiration in 'Oriental' motives and themes, often using them in highly eclectic and idiosyncratic ways (see **Figure 1**). Second, it refers more broadly to the study of countries and regions in 'the Orient'. During the period of colonial conquest, several European states set up institutions for dealing specifically with the Orient in trade and scientific study and by the nineteenth century 'Oriental Studies' was a well established discipline. Third, Orientalism refers to a particular discourse of knowledge about the Orient produced by the colonial powers of Europe (and North America) from the nineteenth century onward; a discourse still prevalent in popular understandings of 'oriental culture', foreign policy strategies, and interventions directed toward Muslim countries in the Middle East, and, more generally, cultural encounters with people and cultures originating from these countries. This meaning of the concept of Orientalism has been coined by the Palestinian American intellectual and cultural critic Edward Said (1935–2003; see **Figure 2**). Rooted in a post structuralist approach partly inspired by the work of French philosopher Michel Foucault (1926–84), Said coined the notion of Orientalism as a term comprising the whole of European (and later US) ideas, thoughts, cultural depictions, military reports, and claims to superiority over the Middle East, in particular the Arab/Muslim countries of Northern Africa and the Mediterranean. In this article, the concept of the Orient refers to the construction of people and places in the Middle East and North Africa; however, in a broader general parlance as well as within the concept of 'Oriental Studies', Central, South, and East Asia are also included. Later work by both Said as well as several authors engaged in geography, literature, cultural studies, anthropology, development studies, social science, and so on have sought to broaden the concept, applying it to a variety of other historical geographical contexts and phenomena.

It is almost impossible to underestimate the significance of Said's concept of Orientalism within human geography and this significance is reflected in the dominance of Said's concept in this article. The article begins

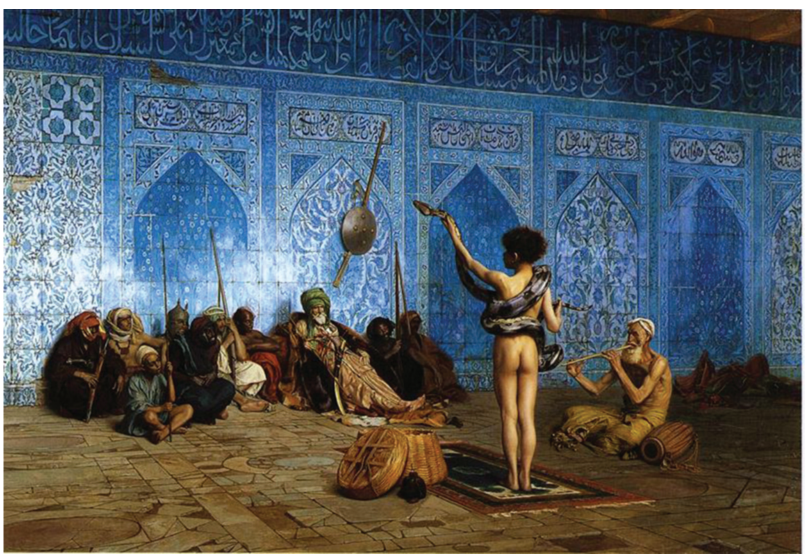

Figure 1 *The Snake Charmer* by Jéan-Léon Gerôme (1889) illustrates very well the idiosyncratic, eclectic, and Eurocentric mode of representation that characterized the Oriental movement. The picture, which also was on the front cover of the first edition of Said's book *Orientalism*, in 1978, depicts a young naked boy performing his act with a snake before an emblematic, oriental despot. In depicting this alleged typical oriental scene with all its exotic, seductive, and repulsive connotations, the painting mixes ornamental styles from India, the Ottoman Empire, and Arab culture to produce this particular imaginative geography of an Orient staged, through the intervention of the painter, for the Western colonial subject.

Figure 2 Edward Said (1935 2003) was born in Jerusalem but grew up with his family in Cairo and at various schools in the United States. Said was already an influential literature critic when he published *Orientalism* in 1978. Throughout his life, Said was active both as a cultural critic and a political activist combining his work as professor of comparative literature at Columbia University with a (often) tense political engagement in the Palestinian Liberation Organization. The two strands of his engagement as both a political activist and an academic provoked claims from the academic conservatives that he was unjustly politicizing in his writings on Orientalism. However, Said himself maintained that it was rather his moral and political obligations to the Palestinian cause that led him to pursue a deeper and broader understanding of Western culture and its colonial context, for example, in the novels of Joseph Conrad and Jane Austen. Courtesy of Palestinian Academic Society for the study of International Affairs.

by outlining the context for Edward Said's notion of Orientalism, especially emphasizing the central inspir ation he drew from post structuralism and especially the work of Michel Foucault and his concept of 'discourse'. It then summarizes some of the main features of Said's notion of Orientalism and how the concept is developed in his seminal book *Orientalism.* After this, the key notion of 'imaginative geographies' and the way this particular discourse produces and legitimizes Western hegemony is presented. Fourth, this article traces some of the most significant impacts the idea of Orientalism has had on studies of cultural encounters between the West and 'non Western' cultures, postcolonial theory, and human geography. It also briefly outlines some examples of how the basic ideas of Orientalism have been transferred to other empirical contexts. Fifth, some of the main cri tiques of the concept of Orientalism, as well as Said's own responses to these, are briefly examined. Sixth and finally, this article suggests some of the reasons why the concept of Orientalism has recently gained new relevance, espe cially in human geography.

Background: The Foucauldian Legacy

Orientalism refers to the construction of the Orient by European colonial powers in the nineteenth century and onward. As a Western means of dominating and gaining authority over the Orient it is, in Said's words, a style of 'thought' based on an ontological and epistemological distinction between the Orient and the Occident. Said describes how this style of thought, in literature and aca demic studies, produced the image of the Orient, for ex ample, the Middle East and Islam, as a threatening, inferior, and underdeveloped 'other'. The critical discussion of Orientalism and the distorted view of Islam in which, as Said says, the image of Arabs in the popular mind is that of camel riding terrorists, arose with the decolonization of the Middle East. Together with the critique of Western he gemony and the stereotypical Eurocentric idea that Europe is superior to its 'others', Abdel Malek, an Egyptian in tellectual, argued that the collapse of colonialism exposed the fallacies of Orientalism. By questioning the central idea of Orientalism, namely, that the Orient and Orientals could be objects of academic study stamped with their timeless otherness, Abdel Malek was, with a post structuralist ap proach, criticizing the essentialist conception of the Orient in Oriental Studies.

The linguistic turn in literature and social science marked the beginning of a new way of understanding the unequal relationship between the colonizer and the col onized. Inspired by Michel Foucault's analysis of know ledge and power, and the insight that language makes the world rather than mirroring it, Said suggested that while the Orient is an integral part of European material civ ilization, Orientalism expresses and represents that as pect culturally and ideologically, as a mode of discourse. Said drew his inspiration from Michel Foucault's notion of how the articulation of words, speech, and images produce discourse. Discourse, according to Foucault, arises through regimes of truth, for example, academic institutions that hold the power and the authority to be able to articulate truth, and the power to internalize this truth in the subject. Echoing Foucault, Said emphasizes that the Orient is not a free subject of action and thought, but rather, a created and imaginary geographical entity constructed in the context of colonialism and Western dominance.

Taking the late eighteenth century as a starting point, Said, in a Foucauldian framework of analysis, argued that the Western discursive system of dominance and authority in the unequal relationship between the Orient and the Occident can be understood as a regime of knowledge – a disciplined system of power – that not only describes, teaches, and rules, but also ‘produces’ the Orient. Said’s argument is that the complex of ideas, images, representations, and academic writings about the Orient actually ‘creates’, rather than describes, the Orient or, as he suggested, the Orient was ‘Orientalized’; it was ‘made’ oriental. This does not mean that Orientalism is an airy Western fantasy about the Orient, but rather a comprehensive archive of diverse texts, theories, practices, and knowledge about the Orient. In the framework of colonialism and Western hegemony, the complex Orient emerged as a suitable subject for study in the academy, in anthropology, sociology, economics, biology, and development studies, as well as in linguistic, racial, and historical theses. Said’s important point is that Orientalism is not a false knowledge construction, but a power relationship with political implications. Orientalism as discourse was not just an innocent attachment to colonialism; it was politically useful when colonial powers, such as Britain and France, were conquering their colonies. From this constructionist perspective, Said’s argument, then, is that Orientalism is a system of knowledge and a configuration of power that actively made colonization possible by legitimizing colonial rule.

Said does not follow Michel Foucault strictly, and it is not always consistent or clear whether Said is taking a constructive or a more realistic ontological position. In at least one sense, Said takes a more humanistic approach to discourse analysis. Unlike Foucault, he believed that the individual writer or text contributed toward the construction of the Orient, especially empirically, in the case of Orientalism. Said, therefore, employs a very close textual reading to reveal the dialectics between the author and the collective cultural apparatus, which, over time, created the powerful idea and image of the Orient.

Imaginative Geographies of the Orient

The concept of imaginative geography was first introduced by Edward Said in his critical discussion of Orientalism, in 1978. According to Said, the orientalist practices and constructions create a familiar space in our minds which is ‘ours’, and an unfamiliar space beyond ‘ours’ which is ‘theirs’. Imaginative geography works by producing contradictions between East and West, between ‘us’ and ‘them’, both in terms of mentality and territoriality. Inspired by Gaston Bachelard’s concept of the poetics of space, Said argues that space is not an objective phenomenon, but is converted into meaning in a poetic process with imaginative and figurative values related to naming and emotions. In the context of imaginative geography, the Orient is an idea which has a history and a tradition of imagery and vocabulary that has established a presence for it in the West.

Said describes how the Orient, as a field of knowledge produced in Europe and America, was represented as a kind of theatrical stage to the distant European audience. This oriental stage was peopled with a range of representative figures or tropes: stereotypical characters such as braggarts, misers, and gluttons and a whole cultural repertoire of monsters, devils, heroes, terrors, pleasures, and desires. Said’s argument is that the staged Orient was imagined as an enclosed space representing the whole of the East, and that it, in a very powerful way, shaped the perception and modes of encounter between East and West. Imaginative geography became a rigorous system of morality and a disciplined regime of truth, represented and produced through a network of corporate institutions. It was the institutionalized Western knowledge of the Orient that created the background for an almost mythical representation of the Orient, or as Said puts it, a latent Orientalism: a powerful ‘fantasy’ in which the Orient existed as a place isolated from European progress.

Imaginative geography works by dramatizing geographical distance and difference between what is close and what is far away. The artificial lines between East and West, between ‘us’ and ‘them’, are symbolic, and opposition is fabricated through the reproduction of a double discursive system. The ‘other’ (Orient) is often imagined in the negative as primitive, barbaric, and wild in contrast to the European self, which is pictured as civilized, advanced, rational, and modern. The Middle East and Islam, in particular, have always represented a trauma in the Western imagination as something that is religious, hysterical, strange, and dangerous. But imaginative geography also has a material foundation in colonialism, imperialism, and the Western interests in the Middle East. It works through spatialization, turning distance into difference, and producing two spaces by drawing a line between two continents. It has real and very visible material consequences. In Derek Gregory’s words, “Orientalism produces the effect that it names” (Gregory, 2004).

The configuration of power or force that Orientalism imposes on the Orient is, according to Said, a three way relationship between the Orient, the Orientalist, and the Western consumer. First of all, the Orient is penetrated, converted, and produced as something outside the boundaries of Europe or ‘our’ world. The Orient is orientalized and transformed from a free floating object into a unity of knowledge. Second, it forces and directs the Western reader to receive and accept the Western constructed image as being the true Orient. Third, it creates its own kind of truth in oriental studies, in a closed system that operates as learned judgment.

Table 1 In *Orientalism* Said argued that the differences between the West (Occident) and East (Orient) were produced by (rather than producing) a dichotomy; a binary discourse constituting a specific imaginative geography of western superiority over 'The Orient' and its people. Thus, it was the emergence of this binary discourse that made it possible for 'the West' to imagine itself as rational, democratic and mature as opposed to the irrational, despotic and childlike character of 'Orientals'

Binaries constituting the Occident versus the Orient	
West/occident	*East/orient*
Positive/presence	Negative/absence
Masculine	Feminine
Rational	Irrational
Democratic	Despotic
Progressive	Timeless

Source: Reproduced from Derek Gregory's entry in Dictionary of Human Geography 4th ed., p 566.

In many ways, imaginative geography and the inven tion of the Orient stabilized the unity and homogeneity of 'the West' by placing Europe in a hegemonic and su perior position in which European culture gained in strength and identity by contrasting itself to the Orient as a kind of surrogate or even underground self, as Said puts it. The manufactured self image of the West as the in verse of the Orient placed Europe in a historical position in which it could imagine itself to be progressive, civil ized, and rational (see **Table 1**).

Orientalism and Its 'Others'

Said's book immediately got widespread attention when it was first published and, together with other work demonstrating how national and territorial identities were invented and constructed, his book became central for subsequent studies of how non European cultures and territories, such as India and the Caribbean, were con structed and produced by the Western imagination and by Western colonial, cultural, and scholarly institutions. Said's ideas have also been translated to many other empirical settings to trace how processes of 'othering' and struggles for superiority and exclusion have been part of the internal colonialization of Europe – for example, in tracing the construction of Ireland to English colonial imaginations of Ireland as a colonial backyard for the British Empire. Many of these studies have had a textual orientation focusing predominantly on the discursive construction of places and people and not so much on the institutionalized power regimes installed and performed through such discourses. An example of the latter is T. Mitchell's *Colonising Egypt* which shows how Orientalist discourse was crucial to the colonial regime and its dealing with native inhabitants, organizations, and polit ical struggles in Egypt in detail. Mitchell's work has also, together with Said's notion of imaginative geography, been a key source of inspiration for the work of human geographer Derek Gregory in his interrogations of geo graphical imaginations and colonial domination.

Within human geography, Derek Gregory has insist ently shown how Orientalism still persists in disciplinary practices as well as being deeply embedded both in tourists' perceptions and performances of the Orient (historically as well as contemporary) and, more recently, in how the media, popular culture, and politics deal with the Muslim Middle East as part of the 'Global War on Terror' launched by the US after the attacks on New York and Pentagon in 2001. Thus, while Gregory's *Geo graphical Imagination*, in 1994, showed the persistence of Orientalism as a regime of knowledge, it continues to inform disciplinary work within geography (and other social sciences) *The Colonial Present*, in 2004, shows in great detail how Orientalist discourse informs how the peoples of the Middle East are being depicted and rep resented as potential enemies of the West, and how this discourse choreographs particular performances of dis tance, hostility, and warfare toward them.

Said's work is often seen as foundational to the group of writings in social and cultural studies often categorized as postcolonial theory. This rather heterogeneous group of writings comprising contributions from authors such as Fanon, Cesaire, Spivak, and Bhabha all point to the contingently unstable yet crucial and power ridden role of identity formation in the relations between Western and non Western cultures and people. What ties Said to this group of writers is, first and foremost, his insistence that the act of representation is not innocent but always contextual and political. Following this reading, cultural studies of postcolonialism inspired by Said have tended to pay more attention to difference in representation and the construction of identity, for example, by gendering accounts of the Orient by pointing out differences be tween female travelers to the Orient and Said's male subjects of the colonial apparatus. Similarly, the concept of Orientalism has also been extended to dealing with the 'internal' Orientalism at work in many Western societies following immigration flows and the emergence of non Western diasporas in these countries. Similarly, some writers have pointed to the necessity of extending the focus from textual and representational aspects to the embodied, nonrepresentational, and habitual perform ances of 'practical' Orientalism in everyday life encoun terd with multiculturalism.

In addition to these writings, other studies have sought to trace out the construction of 'the West' within non Western cultures. These studies of 'reverse' Orientalism or Occidentalism clearly draw their inspiration from Said; however, it must be emphasized that, in Said's perspective, Occidentalism and Orientalism cannot be conceived of as standing on an equal footing as it seems to be assumed in, for example, Buruma and Margalit's book *Occidentalism:*

The West in the Eye of Its Enemies. Abstracting the processes of representations of an 'other', from the networks of power which shape the colonial context for the production of such imaginative geographies reduces the postcolonial encounter to simply being cultural constructions, rather than elements in power struggles also over economic, political, and military hegemony and control.

Beyond Orientalism?

Many writers have sought to frame the problematic of Orientalism as a matter of moving 'beyond Orientalism'. While it is certainly important to reflect on the power relations reproduced through the Orientalist regime of knowledge, in order to move beyond the exercise of colonial power relations in academia, everyday life, and politics, the dismissal of Orientalism should, however, not be premature.

From the outset, Said's notion of Orientalism was welcomed as an important and necessary contribution to the study of Western colonial rule. However, it also gave rise to a number of friendly as well as not so friendly critiques. Among the latter was, not surprisingly, the response of the then (and now) leading Orientalist Bernard Lewis who was also the main target of Said's own critique of *Orientalism Now* in his book. He and other conservative academics working within the disciplinary traditions of Orientalist studies, by and large, dismissed Said's work and main points as overtly polemic, not empirical grounded, and based on an eclectic choice of sources neglecting vast parts of oriental studies. More sympathetic academics with a background in oriental studies pointed out that Said's polemics and political engagement seemed to weaken his strictly analytical points. Said's own response to that sort of critique has generally been that they avoid his point and that they exhibit a remarkable unwillingness to discuss the problems of Orientalism in the political or ethical or even epistemological contexts proper to it. Furthermore, he has argued that the points made in Orientalism should be read not only as an attack on Western culture but rather as a way of facilitating a fuller, more nuanced, interpretation of canonical literary work; a task Said himself embarked on in his essays on *Culture and Imperialism*, in 1993, which in significant ways generalizes and nuances the analysis of Orientalism begun in his 1978 book. *Culture and Imperialism* includes dense and detailed interpretations of the nineteenth and twentieth century European culture and literature, and in the book Said shows how Orientalism is not only an important backdrop to cultural works with obvious Orientalist themes, such as Guiseppe Verdi's opera *Aïda*, but is also, in more subtle ways, in play in literary works of European high culture, such as Jane Austen's *Mansfield Park* and in the formation of postcolonial literature in the twentieth century. In doing this, he shows how cultural and scientific institutions of Europe continuously reproduce the superiority of 'the West' over its colonial hinterlands.

Some critics of Said have pointed to his slightly ambivalent relation to Foucault and post structuralism, trying to fuse the post structuralist Foucauldian discourse analysis with a humanist standpoint. Furthermore, it has been argued that Said does not leave much room for the subaltern to speak back, but leaves the voices of 'the other' to be represented by Western intellectuals, academics, and bureaucrats, and that his focus on products of high culture and textual representations tends to underplay the significance of everyday life. Said's own response to these friendly critics has been, first, to emphasize his humanist standpoint, and the fact that he aims to integrate what he finds useful in Foucault's work into this perspective, and second, to point out that the neglect of everyday practices and 'oriental' representations is not a matter of principle but rather empirical focus. In addition to this, the publication of *Culture and Imperialism* surely makes up for the accused neglect of 'other' voices in Said's earlier work, and demonstrates that the value of the concept of Orientalism precisely lies in acknowledging its persistent imprint on works of culture, everyday life, and politics.

Orientalism Today

A key idea in Orientalism is that this particular regime of knowledge seems persistently to be recycled and reproduced on a multiplicity of scales. Thus, both contemporary and historical studies of tourism and leisure have pointed out how images of the exotic and seductive Orient continually fuel the imagination of the tourists as well as the marketing material of the global tourist industry applying the Orientalist mode of representation not only to Said's Middle East 'Orient' but also to the Far East, the Carribean, and so on. More dismal representations of the Orient's imaginative geography can be found in the worldviews underpinning the US led global war on terror. Thus it has been argued that media coverage and US (and its allies) politics in the Middle East indicate a continual 'colonial present' in contemporary international relations. In 1993, the controversial American political scientist Samuel Huntington revived the idea of a 'clash of civilizations' as a potential heir to the tensions of the Cold War. Huntington conceived of these civilizations as clearly (religiously rooted) cultural, organic entities with their own spatiality. The idea of an inevitable conflict, primarily between Muslim and Western culture was, however, pronounced 3 years earlier in an essay by none other than Said's fierce opponent Bernard Lewis on 'The roots of Muslim rage'. Later, Lewis developed his argument in *What Went Wrong? Western Impact and Middle Eastern Response*, in 2002. Here, he traces the roots of political Islam to the failed developments of science and culture during Ottoman rule. From a Saidian perspective, we may understand this postulate as yet another recycling of the idea of the Orient as an antithetical and internal homogeneous, irrational and potentially threatening 'other' to Western civilization. This is also the way Said himself addressed US politics toward the Middle East following 11 September 2001. Furthermore, Said's argument has been that precisely in emphasizing Muslim virtue and invoking the idea of an Arab Muslim community political Islam is not challenging the superiority of the West but merely mirroring the basic features of Orientalism: The Orient and the Occident as two internally homogeneous entities standing in dichotomous relation to each other without being able to bridge, bond, or blend.

Summary

For a cultural critique, Said and his concept of Orientalism has had an enormous impact within social and cultural studies. This is particularly due to the light his work puts on the construction of reality rather than the mirroring of it and, following from this, his continuous insistence on the importance of the power relations and institutional bodies that produce our conception of social 'reality'. In contrast to other proponents of constructionism, Said has insisted that the production of distorted conceptions, such as the ones Said treats in *Orientalism*, are a matter of moral, intellectual, and political choice. Here Said takes a more humanistic stance than most post structuralists, and instead of limiting his analysis, he consistently attempts to reveal the dialectics between the author and the collective cultural apparatus in his work. Among his most influential notions (not least within the discipline of geography) has been that of 'imaginative geography', a discursive construct that produces and legitimizes Western hegemony over 'the Orient'. This idea has since been applied to numerous other historical and contemporary settings. While most of Said's own work on Orientalism and the numerous studies inspired by his work have had a distinctly historical character, more contemporary work on the workings of Orientalism has begun to emerge. Thus, both the US led war on terror as well as current debates and struggles related to multiculturalism (especially in Europe) have provided room for new encounters in social science and geography with the notion of Orientalism. In this way, the concept of Orientalism as coined by Said in 1978 is, 30 years later, perhaps even more central and urgent in social analysis than ever before.

See also: Colonialism I; Colonialism II; Discourse Analysis; Humanism/Humanistic Geography; Multiculturalism; Postcolonialism/Postcolonial Geographies; Representation, Politics of.

Further Reading

Buruma, I. and Margalit, A. (2005). *Occidentalism: The West in the Eyes of Its Enemies*. New York: Penguin Books.
Clifford, J. (1988). On Orientalism, *In The Predicament of Culture*, pp 255 276. Cambridge, MA: Harvard University Press.
Fanon, F. (1967. Orig. 1952). *Black Skin*, *White Masks*. New York: Grove Press.
Gregory, D. (1995). Imaginative geographies. *Progress in Human Geography* 19, 447 485.
Gregory, D. (2004). *The Colonial Present*. Oxford: Blackwell.
Lewis, B. (2002). *What Went Wrong? Western Impact and Middle Eastern Response*. London: Phoenix.
Lockman, Z. (2004). *Contending Visions of the Middle East: The History and Politics of Orientalism*. Cambridge: Cambridge University Press.
Mitchell, T. (1988). *Colonizing Egypt*. Cambridge: Cambridge University Press.
Said, E. (1978). *Orientalism*. London: Penguin.
Said, E. (1985). Orientalism reconsidered. *Cultural Critique* 1, 89 107.
Spivak, G. C. (1988). Can the subaltern speak? In Nelson, C. & Grossberg, L. (eds.) *Marxism and the Interpretation of Culture*, pp 271 313. Urbana, IL: University of Illinois Press.

Other/Otherness

J.-F. Staszak, Université de Genève Geneva, Switzerland

Glossary

Ethnocentrism The propensity of a group (in-group) to consider its members and values as superior to the members and values of other groups (out-groups).
Exotic Belonging to a faraway, foreign country or civilization and thus demarcated from the norms established in and by the West.
Exoticism A taste for exotic objects/places/people.
Exotism Characteristic of exotic things/places/people.
In-Group A group to which the speaker, the person spoken of, etc., belongs.
Other Member of a dominated out-group, whose identity is considered lacking and who may be subject to discrimination by the in-group.
Othering Transforming a difference into otherness so as to create an in-group and an out-group.
Otherness Characteristic of the Other.
Out-Group A group to which the speaker, person spoken of, etc., does not belong.

A New Geographical Issue

The questions of the other and otherness took the geo graphical world by storm beginning in the 1980s. Of course, geographers were interested in the elsewhere before that date. Homer enchanted us with his de scription of faraway, dreamlike lands; Herodotus was fascinated by Persian society; Hippocrates sought to ex plain societal diversity through the environment's influ ence. Renaissance era explorers were amazed by the peculiarities of the civilizations they discovered. Starting with the end of the nineteeth century and the insti tutionalization of colonial geography in Europe, geog raphers sought to document the particularity of the physical environment and tropical societies.

All of these approaches seek to explain the spatial heterogeneity of societies. Although they claim to be more or less objective, they seek to demonstrate that Western civilization is superior to others and to explain why this is so.

Beginning with the development of radical geography and then feminist geography in the 1960s, geographers took an interest in minority groups who, here, distinguish themselves from the (white, male) norm. But this has more to do with denouncing systems of oppression than with inquiring into the otherness of these minority groups. It was not until the development of postmodern, postcolonial, and queer analyses that otherness became a geographical issue. In order to reach this point, geog raphers have had to ask questions about the diversity of groups in terms of socio discursive construction rather than in terms of supposed objective difference, as had been done until then.

Definitions

Otherness is the result of a discursive process by which a dominant in group ('Us', the Self) constructs one or many dominated out groups ('Them', Other) by stigma tizing a difference – real or imagined – presented as a negation of identity and thus a motive for potential dis crimination. To state it naïvely, difference belongs to the realm of fact and otherness belongs to the realm of dis course. Thus, biological sex is difference, whereas gender is otherness.

The creation of otherness (also called 'othering') consists of applying a principle that allows individuals to be classified into two hierarchical groups: them and us. The out group is only coherent as a group as a result of its opposition to the in group and its lack of identity. This lack is based upon stereotypes that are largely stigma tizing and obviously simplistic. The in group constructs one or more others, setting itself apart and giving itself an identity. Otherness and identity are two inseparable sides of the same coin. The Other only exists relative to the Self, and vice versa.

The asymmetry in power relationships is central to the construction of otherness. Only the dominant group is in a position to impose the value of its particularity (its iden tity) and to devalue the particularity of others (their otherness) while imposing corresponding discriminatory measures. Therefore, if the Other of Man is Woman, and if the Other of the White Man is the Black Man, the opposite is not true. Dominated out groups are Others precisely because they are subject to the categories and practices of the dominant in group and because they are unable to prescribe their own norms. Out groups cease to be Others when they manage to escape the oppression forced upon them by in groups, in other words, when they succeed in conferring upon themselves a positive, autonomous iden tity ('black is beautiful'), and in calling for discursive le gitimacy and a policy to establish norms, eventually constructing and devaluing their own out groups.

The power at stake is discursive: it depends on the ability of a discourse to impose its categories. But this ability does not depend solely upon the logical power of the discourse, and also upon the (political, social, and economic) power of those who speak it.

The West and Others

The ethnocentric bias that creates otherness is doubtlessly an anthropological constant. All groups tend to value themselves and distinguish themselves from others whom they devalue. For instance, according to Lévy Strauss, many autoethnonyms (such as Inuit or Bantu) refer to 'the people' or to 'the human beings', considering more or less the out groups as nonhuman.

On the other hand, the forms of this ethnocentrism are varied and have been constructed by discourse and practice throughout history. Certain constructs are specific to certain societies (such as the heterosexual/homosexual dichotomy), and others seem universal (such as the male/female dichotomy). All societies, then, create the self and the other with their own set of categories. Western society, however, stands out for two reasons.

First, otherness and identity are based on binary logic. Western thought, whose logic has been attached to the principle of identity, the law of noncontradiction, and the law of the excluded middle since the time of Aristotle, has produced a number of binaries that oppose a positively connoted term and a negatively connoted term and thus lends itself well to the construction of the self and the other. Many such dichotomies exist: male/female, Man/animal, believer/nonbeliever, healthy/ill, heterosexual/homosexual, White/Black, adult/child, etc.

Second, colonization allowed the West to export its values and have them acknowledged almost everywhere through more or less efficient processes of cultural integration. Western categories of identity and otherness, transmitted through the universalist claims of religion and science and forcibly imposed through colonization, have thus become pertinent far beyond the boundaries of the West.

Although this article only discusses Western constructions of otherness, this does not mean that other societies are unaware of the process. Rather, their particular categories of the self and the other have been less widely diffused than those of the West. Therefore, the system of races, although very recent and Western, has been imposed everywhere as a framework to conceive of human diversity, while the older caste system, belonging to the Indian world, has not.

Geographical Others

We, here, are the Self; they, there, are the Other. How and why do we think that those who are far away are radically different, to the point of being Others? How can otherness be essentially geographical?

In ancient Greece, the geographical form of otherness opposed Greek speakers and Barbarians. A Barbarian was a person who did not speak Greek and thus had not mastered the *logos* (and was not familiar with democracy). His culture was lacking and he belonged to another civilization. If this otherness comprises a geographical dimension, it is because cultural surfaces are divided into supposedly homogeneous spatial blocs (countries, zones, continents, etc.). This construction of otherness is based on a hierarchy of civilizations and requires the use of a universal criterion that allows their comparison. Language and political systems fulfilled this role until the advent of Christianity and Islam, and then religion replaced them to oppose Us (believers) and Them (nonbelievers). The Renaissance and the discovery of new civilizations, especially in America, brought the issue back to the forefront, paving the way for Westerners to search for the means to classify societies. The idea of universal progress, or a chronology that is valid for all societies, allows societies to be organized into a hierarchy from the most primitive (Hottentots, Kanaks, Bushmen, Pygmies, etc.) to the most civilized (Europeans). From the end of the nineteenth century onward, anthropology, ethnology, and geography (still poorly differentiated) sought to give scientific basis to a typology of peoples and societies, a typology that is more or less explicitly a hierarchy. Darwin's theory of evolution offers a coherent scientific framework to explain species' diversity through the diversity of the environments where natural selection takes place and through the relative isolation of these environments, which makes them favorable to the development of differentiated species. In order for it to 'justify' the otherness of man, Darwin's theory needs only to be transposed to human societies, with the development of a hierarchy of different environments and societies and the implementation of certain differences as principles for exclusion. Therefore, Europe's climate and natural resources would explain the fact that (one of?) the world's most advanced societies developed there, while the extreme climates and lack (or abundance, that works equally as well) of natural resources characteristic of all other parts of the world would lock humanity there into a prior and primitive evolutionary stage.

Obviously, thinking of civilizations as different like Others justifies the supremacy of Ours and legitimizes its propensity to dominate them. The Greeks must go to war with the Persians, believers with nonbelievers, Europeans with indigenous peoples. At worst, this is extermination or enslavement; at best, it spreads the Good Word, civilization, and progress.

The second geographical form of otherness does not oppose civilizations. Rather, it opposes civilized (meaning fully human) humanity and humanity still out in nature (or almost animal). It is the Savage, etymologically the 'man of the forest', opposed with man from cities and fields. This figure stigmatizes the Man who has not (yet) left his natural state. Folklore, if not European reality, is overflowing with these Woodsmen. Hairy and violent, they threaten villagers (especially the women of the village). But the figure of the Savage imposes itself as the descriptor of those who would constitute a lesser form of humanity during the Renaissance and the great explorations of Africa and, especially, America. It is thought that they go naked, cannot talk, engage in cannibalism, etc. They are even more worthy of extermination than the Barbarians because their ability to integrate into Humanity is called into question. This form of otherness has a spatial component because civilization is seen as being diffused from a central location (Jerusalem, the city, Europe), and savages are in faraway zones (Australia) or the interstices (our forests).

Carving humanity into races and the world into continents is the third and most recent template that Europe has used to create a spatial form of otherness. This template still uses the figures of the barbarian and the savage but puts a new criterion into place that allows us, White Men, to be opposed to them, Men of Color. Skin color and certain secondary signs that physical anthropology has helped to identify are used to distinguish White Men, the 'superior' phase of humanity, from 'inferior' races. Each race has a corresponding continent, a natural birthplace from which it can flourish. The anthropological fiction of races and the geographical fiction of continents allow these categories to be reified and naturalized by giving them a supposedly geographical legitimacy and a false sense of evidence ("it's obvious"). They feed off of each other to justify colonial policy and the domination of one race and continent over others. Looking beyond the many races, it is actually a binary form of otherness: the opposition of colonist/native or White/of Color.

Orientalism, as analyzed by E. Said, encompasses all of these components. The Oriental is characterized by his barbarity, his savageness, and his race. The Orient is the geographical fiction that gives him geographical basis. Orientalism is the discourse through which the West constructs the otherness of the Turks, Moroccans, Persians, Indians, Japanese, etc., all reduced to the same stigmatizing stereotypes, and thus gives itself an identity in opposition to them. The West thereby gains the right, if not the duty, to dominate the Orient, to save it from despotism, superstition, misery, vice, slavery, decadence, etc.

Spatial Organization

Not all forms of otherness are fully (or even mainly) geographical in nature. Women, homosexuals, and the insane, all major figures of otherness in the West, owe their stigmatization to something other than their location. They are, after all, found among the self. In addition, migrations (notably forced migrations in the context of trade) have resulted in making geographical Others, from elsewhere, come and live here among us.

The cohabitation of the Other and the Self in a common space is not a given. On the one hand, discriminatory policies toward Others are more difficult to implement if the two populations are intermixed. On the other hand, their cohabitation makes it more risky to maintain the particularities (real or imagined) and the stereotypes that distinguish the Self from the Other. Finally, it seems that symbolically, nothing can trump an otherness based on and justified by geography. It would be all the more clear if men came from Mars and women came from Venus. Also, sets of spatial constructs and practices are based on the discursive construction of otherness to separate the self from the other.

Segregation is the most evident of these constructs. Confining Blacks and Jews to ghettos prevents them from intermixing with and therefore contaminating Whites and Christians. Furthermore, confining Others to community life among themselves in a degraded ghetto – where the concentration of poverty and exclusion compounds their effects – creates favorable conditions for the development of visible misery and a specific culture. These serve *a posteriori* as justification for the stigmatization and isolation of the incriminated group and confirm the dominant group's sense of superiority. The ghetto creates otherness. In addition to 'pure' forms of ghettos created by law, there exist less clearly delimited forms of segregation that are maintained by the land market and/or the symbolic or material violence of the dominant groups. There again, the in group and the out group derive part of their identity and their otherness from the more or less stigmatizing space prescribed to them (e.g., the inner city or the suburbs).

On a smaller scale, the constructs confining the insane or the condemned in asylums and prisons fit the same logic. Their confinement sets them apart, worsens their condition, and confirms their particularity. They derive part of their identity – or rather, their otherness – from their prison. The domestic confinement of women can be analyzed in the same terms: by forbidding them access to public spaces and reducing the woman to her domestic role, patriarchal society creates and reproduces gender inequality. Female otherness is created not only by discourse but also by spatial practices and constructs.

Territorial constructions also fit the same logic, except that their effect is less to separate preexisting groups than to confer geographical identities on one another, creating an in group of those on this side of the border and an out group of those on the other side. This process works on all scales: from gangs who occupy different urban neighborhoods to nations separated by interstate borders. A material and symbolic affectation is added to linguistic, religious, ethnic, and other oppositions: people think that they owe their identity and superiority to those of their territory, and they ascribe to others the faults of their respective territories. B. Anderson has shown how discursive and spatial processes participate in the construction of national, imagined communities and, thus, in the construction of figures of otherness against which these communities define themselves.

Exotism

Exotism constitutes the most directly geographical form of otherness, in that it opposes the abnormality of elsewhere with the normality of here. Exotism is not, of course, an attribute of the exotic place, object, or person. It is the result of a discursive process that consists of superimposing symbolic and material distance, mixing the foreign and the foreigner, and it only makes sense from one, exterior, point of view. As a construction of otherness, exotism is characterized by the asymmetry of its power relationships: it is Westerners who, during the phases of exploration and then colonization, defined elsewhere and delimited exotism. The word exotic has become a synonym of tropical or even colonial. It is out of the question to describe Europe as exotic until minds and words are decolonized.

Exoticism is characterized by giving value to the other, contrary to ethnocentric bias. From Homer's fascination with faraway, more or less imaginary peoples to J. J. Rousseau's nostalgia for the noble savage, from the Romantic Orientalism of nineteenth-century writers and painters to the primitivism of a Gauguin painting, from curiosity for ethnic tourism to the recognition of specific rights for first peoples, the West celebrates the Other and even proclaims its superiority through multiple forms (that are not always unambiguous).

The taste for exotica was established in the eighteenth century when exotic turquerie, chinoiserie, japonaiserie, etc., came into fashion. It became commonplace in the nineteenth century with colonization and spread to the tropical world. Up to that point, it was essentially characterized by the import of exotic products, by their pastiches, and by travel books and then colonial literature. Only certain privileged persons, well-to-do aristocrats or explorers, traveled to experience the pleasures of exotic lands. The development of mass tourism in the 1960s leveled the playing field and made their exotism a major resource for many countries.

According to V. Segalen, exoticism is the pleasure of a sensation that, worn down by habit, is excited by novelty. But Segalen notes that from the end of the nineteenth century onward, everything has been seen, and radical exotism – the exotism experienced by the first explorers – is dead. In fact, T. Todorov demonstrates that it is paradoxical to value or desire something that is unknown. Exoticism consists more of showing enthusiasm for what has already been seen, said, or painted: what has been marked elsewhere as picturesque and been reproduced as such. The otherness of the exotic is not the brute and brutal otherness of the first encounter; it is the bland otherness, staged and transformed into merchandise, of the colonial world offered up as a spectacle, as in orientalist paintings, human zoos – and exotic dance. Exoticism is less the pleasure of confronting otherness than the pleasure of having the satisfaction of experiencing the sight of a reassuring version of this confrontation, true to our fantasies, that comforts us in our identity and superiority.

Conclusion

Geography is an remarkably effective producer of otherness. On the one hand, certain spatial patterns are very efficient, albeit discrete, in constructing and maintaining alterity. On the other hand, geography, like physical anthropology or history, has in the past proposed and continues to provide tales that form the basis of discursive constructions of a number of expressions of otherness. Yet these tales, rather than being taken for what they truly are – that is to say, fictions delegitimized by their links to colonial regimes – have acquired a veneer of inevitability by grounding themselves in apparent scientific rationality.

For instance, one of the most common arguments against Turkey's recent claim to being integrated into the European Union is that this country does not belong to the European continent. Legitimized by the so-called evidence of (physical) geography, this tautological statement seems not refutable, unless one acknowledges that continents are geographical fictions produced and used by colonial ideology.

Since alterity is consubstantial to relations of power and processes of oppression, geographers concerned by these must take personal responsibility in identifying and studying the spatial patterns that ground these. A critical and reflexive perspective thus necessarily implies identifying and deconstructing that various more or less learned and so-called scientific geographical representations that serve as discursive bedrock for oppression.

See also: Cultural Politics; Ghettos; Postcolonialism/Postcolonial Geographies; Psychoanalytic Theory/Psychoanalytic Geographies; Race; Segregation; Self-Other.

Further Reading

Anderson, B. (1983). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. New York: Verso.

Beauvoir, S.de. (1952). *The Second Sex*. New York: Alfred Knopf.
Duncan, J. (1993). Sites of representation. Place, time and the discourse of the other. In Duncan, J. & Ley, D. (eds.) *Place/Culture/Representation*, pp 39 56. London: Routledge.
Fanon, F. (1963). *Black Skin, White Masks*. Harmondsworth: Penguin.
Jervis, J. (1999). *Transgressing the Modern. Explorations in the Western Experience of Otherness*. London: Blackwell.
Lévy Strauss, C. l. (1961). *Race and History*. Paris: UNESCO.
Mason, P. (1998). *Infelicities. Representations of the Exotic*. Baltimore, MD: Johns Hopkins University Press.
Said, E. (1978). *Orientalism*. London: Routledge & Kegan Paul.
Segalen, V. (2001). *Essay on Exoticism: An Aesthetics of Diversity*. Durham: Duke University Press.
Sibley, D. (1995). *A Geography of Exclusion*. London: Routledge.
Todorov, T. (1994). *On Human Diversity: Nationalism, Racism, and Exoticism in French Thought*. Harvard: Harvard University Press.

Overlay (in GIS)

O. Ahlqvist, The Ohio State University, Columbus, OH, USA

Glossary

Boolean Binary (two-valued) system of variables and operations for logical operations developed by George Boole in the mid-nineteenth century.

Decision Support System An interactive, computer-based system that supports decision making.

Digitizing The process of converting analog spatial information from sources like paper maps to digital data.

Projection (Map) A method to transform the Earth's curved surface onto a plane.

Topology The geometric relationship between points, lines, and geometric forms that remains consistent throughout spatial operations in a digital mapping environment.

Introduction

Studying where things are and why they are there is at the heart of geographic analysis. Maps serve as one primary tool to collect and analyze geographic information. Maps are generally an organized collection of geographic features that emphasize location. Although we do not think of daily map use as an analytical process, the cartographer certainly was aware of that and wanted to design the map so that the information could be read by the intended user. If we look at a typical road map or online location service map (Google/Yahoo/MapQuest) we may find that cities are represented by points of varying size depending on their population. Similarly, roads may be found represented as lines of varying thickness depending on the type of road; interstate, state, or local route. Looking at the map we can easily figure out where cities and roads are located, which roads are connected to each other, and where a certain road leads to and from. Thus, through simultaneous display of very different real world features, for example, roads and cities, a good map delivers a whole that is far more than its constituent parts. This is the primary objective of overlay analysis.

Historically this more analytical aspect of map use followed naturally from geography's pre 1800s emphasis on exploration and mapping to document discoveries. With access to mapped information on vegetation, geology, climate, and cultures, scholars such as Alexander von Humboldt and Karl Ritter followed in the footsteps of Bernhard Varenius and started to systematize and look at the world as a whole, trying to integrate the separate and specialized studies of nature, humans, and society. During this 'classical period' of geography, spatial location became essential to compare and suggest relations between geographic phenomena, for example, the relationship between latitude, elevation, and vegetation portrayed in Berghaus' *Physicalischer Schul Atlas*. During the next 100 years or so there were several efforts at map based synthesis of geographic elements, albeit separated into either natural or social sciences, for example, Herbertson's work on 'The major natural regions: An essay in systematic geography'.

One of the first documented and analytically tractable overlay processes was carried out during the 1930s by Homer Hoyt at the Federal Housing Administration as part of the administration's effort to "guide the development of housing and the creation of a sound mortgage market" (Hoyt, 1939). The study *The structure and growth of residential neighborhoods in American cities* used detailed, city block maps of characteristics considered indicative of housing conditions such as average rental, percentage of owner occupied units, race, and so forth. The report demonstrated with a collection of map transparencies how these factors can be used to produce a spatial summary of factors that indicate, for example, poor housing conditions. **Figure 1** illustrates an overlay of four separate maps. Each map contains areas selected on the basis of some characteristics that were held to be indicative of poor housing conditions at the time. The four characteristics in this overlay are: population other than white exceed 60%, more than 25% of the buildings need major repairs, the majority of houses are 35 years or older, average rent is less than $20. The power of this new geographic technique is evidenced by the many indications that it was used up until the 1970s as a tool for lending institutions to 'redline' areas, and then to implement discriminatory practices by avoiding mortgage loans to homeowners in those neighborhoods.

Obviously many powerful technologies can be used for both good and bad purposes, and the overlay technique is no exception. With the arrival of computers in the 1950s and 1960s, the overlay process could turn into a digital process where data layers could be numerically integrated, offering a wider array of possibilities, both in terms of the amount of data that can be analyzed and the different logics that could be applied to the integration. During this period the landscape architect Ian McHarg, published his book *Design with nature* that brought back a more holistic view on the overlay analysis, in line with the work of Humboldt and Ritter. He proposed how

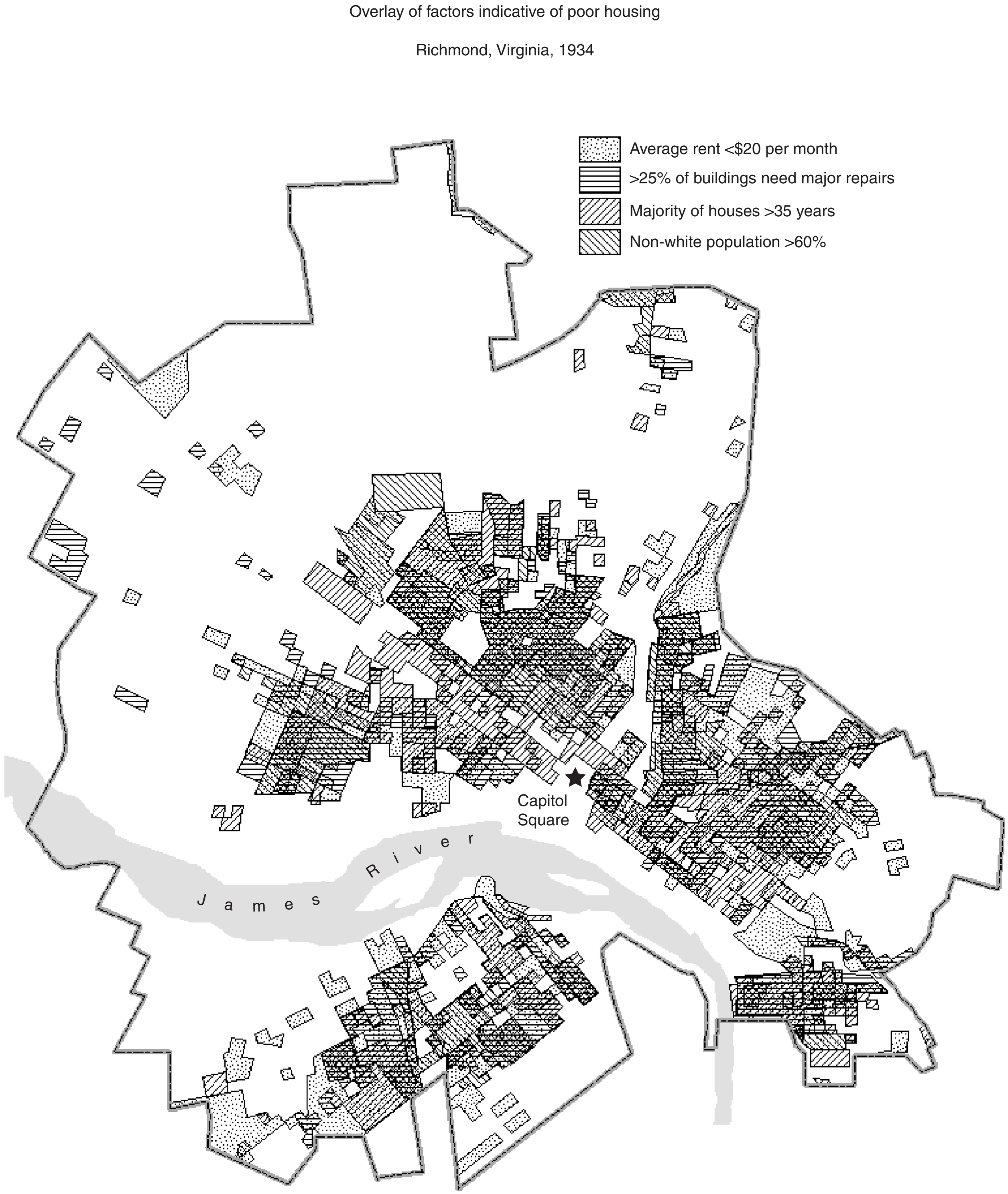

Figure 1 Overlay of factors indicative of poor housing in Richmond, Virginia, following the study by Hoyt (1939). The separate factors add a shade of gray and several co-occuring factors are percieved as darker shades. From Hoyt, H. (1939). *The structure and growth of residential neighborhoods in American cities.* Washington, DC: U.S. Federal Housing Administration.

carefully designed overlays could be used to find optimal solutions for future development that balances natural and social values. The technical implementation was still based on transparent map sheets but was soon to be followed by computer implementations where criteria and enumeration rules could be accurately defined and calculated. This strand of research has ultimately led to the development of contemporary, geographic information system (GIS) based spatial decision support systems.

General Procedure

Generally the overlay process uses a spatially indexed collection of different types of information that is then interpreted based on a question at hand. In its simplest form the overlay operation is merely a stack of map layers that show where things co occur, similar to the housing condition example described previously. In its most ad vanced form the overlay analysis is able to express causal relationships where temporal sequences also can be an integral part. Regardless of technical implementation – either undertaken as a manual overlay of transparent map sheets or through digital overlay of spatial data, a basic requirement is a common spatial frame of reference and accurate geocoding.

In the previous example of Richmond housing con ditions, it should be obvious that each layer needs to be produced at the same scale. Putting a world map on top of the Richmond map would not make much sense. Moreover, since we are dealing with map sheets there are issues related to the projection of a spherical Earth onto a flat plane. We cannot take a world population density map in the Mercator projection and overlay that with a birth rate map in the Robinson projection. Even at smaller scales a common projection is necessary to guarantee an accurate match between the layers. Lastly, each layer needs to have a high level of geocoding ac curacy, so that mismatches because of spatial errors are minimized. This issue can be demonstrated by the Richmond map overlay where we can see at a closer inspection that the separate layers do not line up per fectly with each other, creating small 'sliver' polygons that are an artifact caused by the manual digitizing of separate data sources (**Figure 2**).

This problem can be mitigated by using accurate cross checking techniques during data entry or through postoverlay processing that removes tiny polygons below a specified threshold. The occurrence of sliver polygons may also be due to difficulties in accurately outlining objects with inherently vague boundaries. In these cases the removal of sliver polygons is probably not an ap propriate practice since it removes areas that may be contested or otherwise uncertain.

GIS Implementation

Having illustrated the manual overlay process with transparent map sheets above, we will now look at the digital implementation of overlay analysis, typically performed in a GIS. Digital map data is commonly represented in a raster or vector data set. These are usually geocoded at a certain scale and resolution but can be modified numerically so that data collected at differ ent scales can be analyzed together. It is important to note here that although it is technically possible to overlay data produced at vastly different scales, this is typically not recommended.

In **Figure 3**, two data sets have been prepared for overlay, one representing an evaluation of flood risk and the other showing census block enumeration units. In the first geometric stage, the two maps are put on top of each

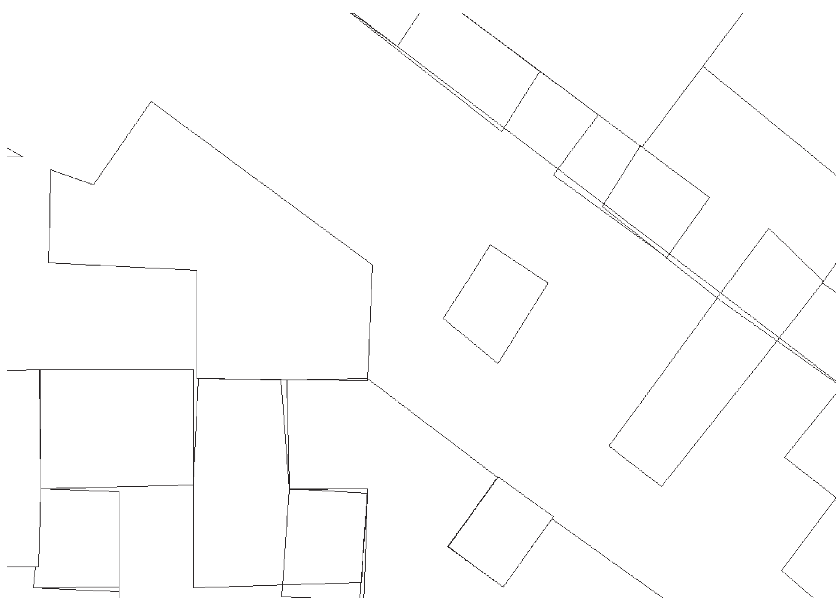

Figure 2 Close view of the Richmond housing data overlay reveals some of the small mismatches and sliver polygons generated by manually digitizing separate map sources.

other to form intersecting areas. The associated table illustrates how these new areas, N1–N5, are made up of combinations of the original areas and subareas. Information associated with the original areas can now be associated with the newly formed areas as well.

With raster data this geometric stage is really resolved in the preparatory geocoding stage where not just scale and projection is determined but also the pixel resolution is set for the data. Once that is completed, each raster pixel is its own spatial unit and all pixels should be perfectly aligned with the pixels in the other data layers. Again, when the geometric overlay is performed all spatially corresponding pixels will form a new pixel that can be associated with information from the source layers (**Figure 4**).

Analysis Alternatives

Once the geometric stage is complete various forms of analyses can be performed on the information that is associated with each new spatial unit. From the most simple to more complex tasks we can separate between descriptive, deductive, and inductive overlay analysis.

Descriptive Overlay

Many times the spatial combination of layers is of most interest, such as if and to what extent certain characteristics intersect or overlap in space. This is simply a description of the results from the first geometric stage. **Figure 5** illustrates three examples of such questions that can be answered by descriptive overlay analysis.

Figure 5a illustrates a situation where the user is interested in whether a certain neighborhood outlined by block 1 is at risk for flooding. In this case the flood risk area does not overlap with the block so it should be out of the risk area. This is an example of a simple topological query where focus is on the spatial relationship between two features. Another common task is to single out areas with certain combined characteristics. The example in **Figure 5b** shows how a subarea within block 3 and also

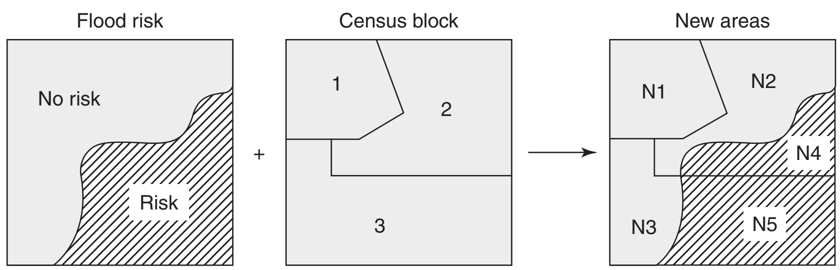

Flood risk	Census block	New area
Part of no risk	1	N1
Part of no risk	Part of 2	N2
Part of no risk	Part of 3	N3
Part of risk	Part of 2	N4
Part of risk	Part of 3	N5

Figure 3 The geometric stage of an overlay produces new areas that are a combination of the source layers.

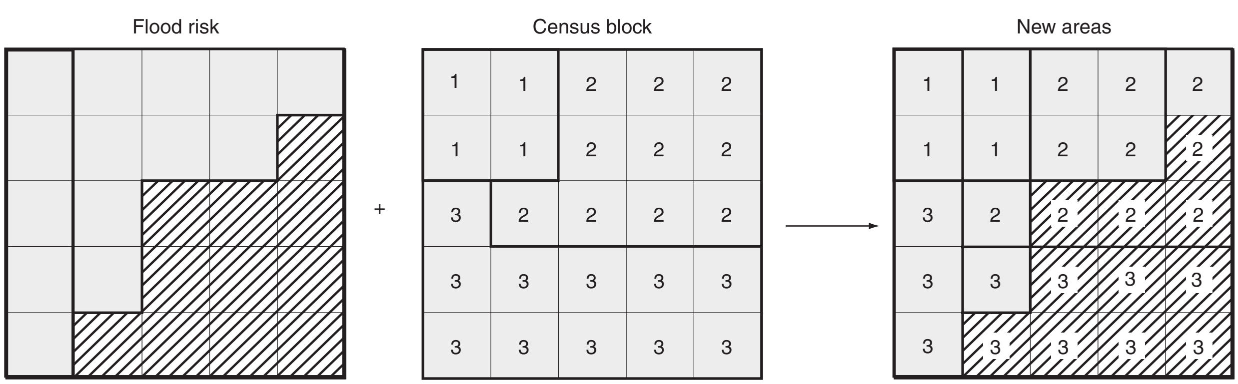

Figure 4 With raster data a common resolution and extent is usually set for each data layer before the actual overlay. Therefore no new areas are created in the geometric stage of the overlay.

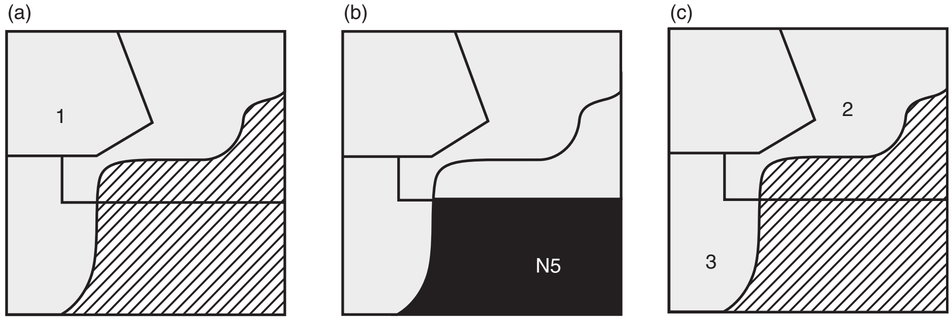

(a) Topological query – Is block 1 at risk of flooding?
(b) Boolean query – Area covered by block 3 and flood risk
(c) Spatial query – What population is affected by flood risk?

Figure 5 Three examples of descriptive analysis performed after overlay of a census block data layer and a flood risk layer.

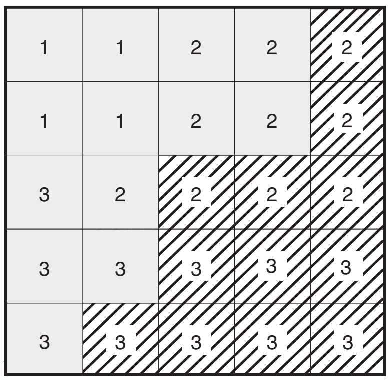

Contingency matrix

Flood risk	Census block		
	1	2	3
No risk	4	5	4
Risk	0	5	7

Figure 6 A contingency matrix (left) can summarize information from the separate area combinations or contingencies. In this example a simple count of pixels for each contingency is stored in the matrix.

within the flood risk area is identified using a Boolean query. In the final example in **Figure 5c** we use the overlay to extract information about the demographics of areas affected by flood risk. In this case we use a topo logical relationship between objects in one layer to find attribute information for overlapping areas in another layer.

The last type of spatial query in **Figure 5c** can also be used to effectively summarize attribute information for the whole area in a contingency matrix. In this matrix we can collect some type of attribute summary, such as minimum, maximum, mean, or sum, for each combin ation, or contingency, of the combined spatial units.

The contingency matrix summary is a general and important descriptive tool that allows for some specific uses. For example, **Figure 6** illustrates a raster version of the flooding and block data after overlay. The associated contingency matrix summarizes the number of pixels for each contingency. There are, for example, 4 pixels in census block 1 and they are all outside the flood risk area. Block two on the other hand is split so that half, 5 pixels, are in the 'no risk' area and the other half, 5 pixels, are in the 'risk' area. Using the spatial reference information about the layers the number of pixels can easily be recast into spatial area units so that an areal estimate is given in the contingency matrix instead of raw pixel counts. This technique is also commonly used to produce spatial summaries of combined factors, or for looking at changes over time in an area. For example, making an overlay with data from different times, t1 and t2, will produce a change image and the associated contingency matrix can now be termed a change matrix (**Figure 7**).

In the change image, each pixel can hold information on what change happened from time t1 to t2 and the change matrix may then summarize the amount of change from one category to another, so that it is possible to say that, for example, 3 pixels remained in category A but 1 pixel changed from A to B, 3 pixels changed from B to C, and 1 pixel changed from C to B. Again, appli cations often translate this into area units so the outcome could be interpreted as, for example, 15 sq mi changed from agricultural to residential land use, given that each pixel is 5 sq mi, B is agriculture and C is residential. The major diagonal has a certain meaning since all those cells summarize areas of no change. Applications of this technique can be found in various land use, population,

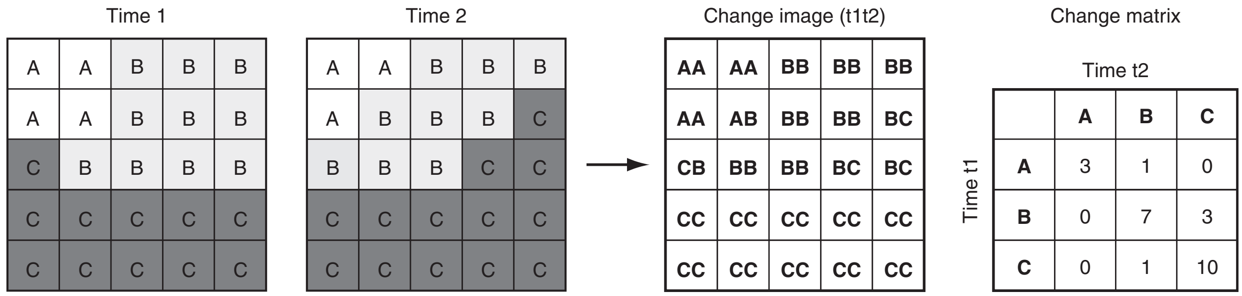

Figure 7 Using similar data from different times produces a change image and a contingency matrix that summarize change from one time to another by counting the number of instances of change from one category to another.

and climate change analyses. A variation of this technique is also a standard procedure to make accuracy assessment of map and satellite data in which one layer is the map estimate and another layer contains a carefully measured reference. The contingency matrix then is interpreted as an error matrix where off diagonal elements are indications of map error.

Deductive Overlay

The Boolean overlay illustrated in **Figure 5b** above is actually a simple example of a larger group of deductive analyses that can be performed as an overlay process. Deductive reasoning is a type of inference where a conclusion is reached based on a set of facts and an inference rule or logic to interpret those facts. We could, for example, say that all residential areas within a flood prone area are at a certain type of risk and should be selected for emergency planning. This translates into a Boolean logic: if a location in layer 1 is 'residential' and the same location in layer 2 is 'flood risk' then that location needs 'emergency planning', and the logic will ultimately select the area highlighted in **Figure 5b**. It is easy to see that we could include more than two layers, or criteria, to determine the outcome, and such spatial multicriteria overlay processes came to be almost synonymous with overlay analysis in early GIS applications. They were commonly used as a method to find, for example, an appropriate location for an activity using rules that followed expert knowledge or opinions. **Figure 8** illustrates how a simple suitability analysis can be performed based on two criteria found in layers 1 (census block demography) and 2 (flood risk). In the first layer the three census blocks are assigned values between 0 and 10 to indicate the degree of suitability based on, for example, the population density in each block. In the second layer the flood risk is rated in a similar way as to its suitability for site selection. A summary operation is then performed based on the geometric overlay of the value layers. This produces a resulting site suitability map where the highest value could correspond to the most suitable location.

There are obviously many ways that factors can be summarized and the choice of what logic to follow all depends on the data and question at hand. One type of logic is essentially to use a dominance rule on the data so that the most frequent, largest, or smallest value is preserved. Examples of some dominance rules are given in **Figure 9** for binary and interval data.

The maximum operation applied to binary data can also be thought of as a set union operation and a binary minimum operation would be equal to a set intersection operation (**Figure 9**, left). This logic is sometimes referred to as exclusionary screening and depending on how data is coded these operations make possible successive binary overlays with, for example, binary ratings of disqualifying factors to remove areas unsuitable for any further analysis. We can think of the Richmond housing condition study in terms of a manual version of the binary minimum operation to single out the poorest housing areas (**Figure 1**). With ordinal, interval, or ratio numerical data it is possible to single out the highest or lowest values in any of the layers. This may be a suitable logic if any of the several independent factors have a direct influence on the outcome.

A second type of logic uses some form of arithmetic to summarize the separate layers into a result. **Figure 8** above was one example of this logic where a simple sum was used to add factor suitability scores. Basically any standard arithmetic operation such as addition, subtraction, division, and multiplication, can be performed as an overlay operation given that the level of the measurement in the data supports the operation, see examples in **Figure 10**, left. It would, for example, not be appropriate to divide one ordinal number with another ordinal number. Furthermore, division and multiplication rules are very sensitive to zero values and multiplication tends to magnify combinations of moderately high values. For example, while $4+4$ and $7+1$ give the same result for a simple sum, $4*4$ is more than twice as much as $7*1$.

Following this line of reasoning we could, for example, modify the Richmond housing condition analysis so that each layer holds degrees of how, for example, average rent

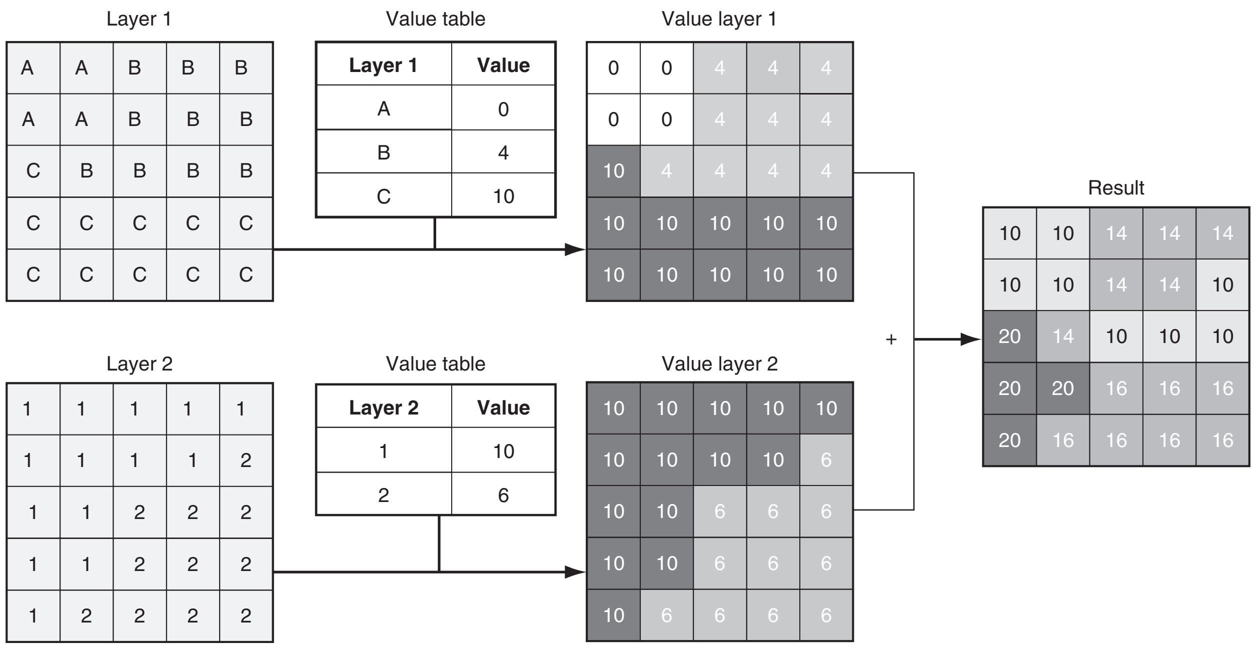

Figure 8 A simple multicriteria overlay analysis using two layers with census information (Layer 1, categories A, B, and C) and flood risk data (Layer 2, categories 1 and 2). Value tables translate factor categories into suitability ratings (0 10) that are overlaid and summed together to produce a summary suitability score for each location.

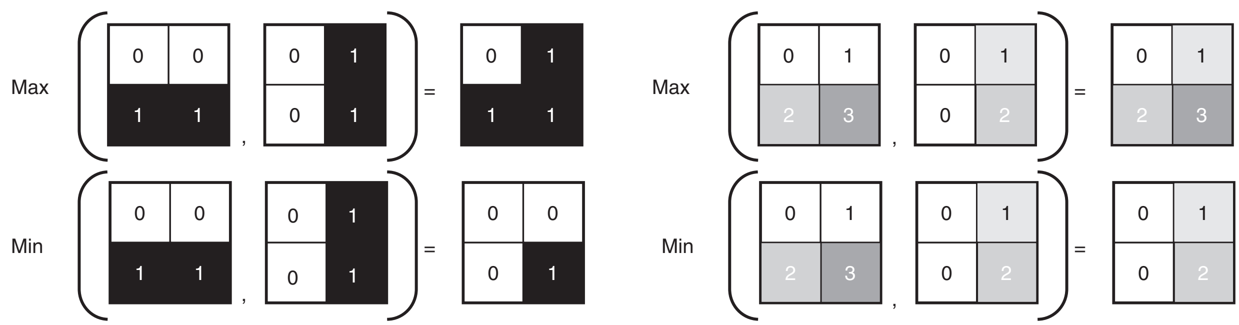

Figure 9 Example of dominance overlay rules applied to binary (left) and interval (right) data.

or age of buildings indicates poor housing. Instead of using just binary classes such as average rent more or less than $15 per month, we can use an ordinal rating of several classes, for example, using the highest rating for the lowest rents and the lowest rating for the highest rents. If all layers employ the same rating scale, for example, 1, 2, and 3 to denote slight, moderate, and severe (poor conditions), it is reasonable to use some form of summary arithmetic to evaluate the total severity of the housing conditions as a combined indicator of poor housing. This was also the logic that Ian McHarg introduced with his example evaluations of land suitability. In his case it was implemented manually by creating transparent overlays with different shades of gray representing grades of suitability. Put together the stack of transparencies showed the added result of different factors as darker or lighter areas. In a digital environment it is obviously easier to control and manipulate the aggregation of ratings through the chosen arithmetic and more refined versions of such arithmetic overlays as averages and weighted sums. These weighted summaries often preserve the original measurement scale in the result as opposed to a simple sum or product. This has the added benefit of being more intuitive so that any number of factors scored, for example, 1–3 will render a combined rating in the same range 1–3, see examples in **Figure 10** right.

While a mean operation simply gives equal importance to each layer, the weighted sum gives users an opportunity to assign different importance to the separate layers. In the Richmond housing condition analysis we could argue that average rent is a more important indicator of poor housing than race, and consequently give the rent factor a higher weight than the race factor. These methods and their successive refinements are commonly

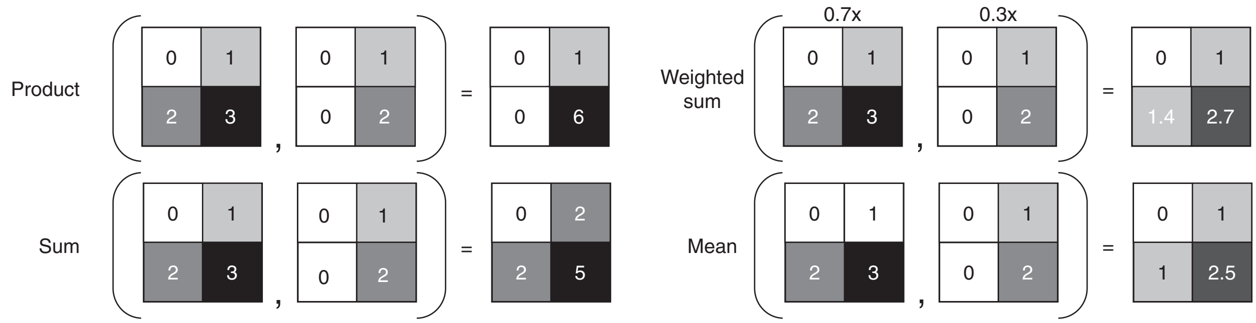

Figure 10 Example of simple (left) and weighted summary (right) arithmetic overlay rules applied to interval data.

referred to as multicriteria evaluation and as much as they are commonly undertaken in GIS overlay analysis they have also been criticized for being just a weighting and rating game. The most important issue is probably that of rescaling original factors, for example, monthly rent and building age, into standardized ratings, for example, 0–10, to indicate poor housing. Is 10 twice as bad as 5? Is $5+5$ (two moderate indicators) really the same as $0+10$ (two opposite extremes)? An important prerequisite for any multicriteria evaluation is to engage enough expert and stakeholder knowledge into the rating and evaluation process. This can be achieved through the practice of public participation GIS. Recent developments have also suggested that the use of fuzzy set based measures toge ther with a general logic for ordered weighted averages can help with the rescaling of factors in a way that makes aggregation effects more transparent to the user.

Inductive Overlay

A third major type of overlay analysis is to use the geo metrically combined overlay as a starting point for more exploratory and inductive tasks. Once the geometric stage is completed we can consider the new map as area ob servations for a multivariate data set. For each area we can have a large number of attribute values such as population density, type of land use, flood risk, etc. Using this in formation we can look for relationships in the data – for example, to see if it is possible to build a regression model based on the attributes so that one dependent attribute can be found or estimated using one or more other in dependent attributes. If we had data on some instances of poor housing conditions in the Richmond data, we could use those observations as regression data points to build a formula for the weighted sum of factors following standard univariate or multivariate regression techniques. Recently, geographically weighted regression has dem onstrated more refined versions of that same idea. Ex ploratory spatial data analysis is another example where inductive overlay analysis can be an integral part.

Application Summary

As a descriptive tool, overlay is an important part of most digital mapping applications since it enables users to put different data layers on top of each other to visually evaluate the totality of a place. The overlay process is also very often a starting point for many types of quantitative analyses and it is usually found as just one component of more elaborate spatial analysis tasks performed in a GIS. The analytical framework of map algebra, introduced by Dana Tomlin, integrates operations on zones and neigh borhoods such as creating buffers around roads or sites with overlay analysis to evaluate spatial adjacency, inter action, and other core geographic questions.

See also: Exploratory Spatial Data Analysis; Fuzzy Set and Fuzzy Logic; Geographically Weighted Regression; Georeferencing, Geocoding; GIS, Public Participation; Location Analysis; Projections; Regression, Linear and Nonlinear; Scale; Spatial Data Models; Uncertainty.

Further Readings

Carver, S. J. (1991). Integrating multi criteria evaluation with geographical information systems. *International Journal of Geographic Information Systems* 5, 321 339.

Chrisman, N. (2002). *Exploring geographic information systems* (2nd edn.). New York: Wiley.

Eastman, J. R. (2005). Multi criteria evaluation and GIS. In Longley, P. A., Goodchild, M. F., Maguire, D. J. & Rhind, D. W. (eds.) *Geographical information systems principles and technical issues* (2nd edn.). Hoboken, NJ: Wiley.

Herbertson, A. J. (1905). The major natural regions: An essay in systematic geography. *The Geographical Journal* 25, 300 310.

Hoyt, H. (1939). *The structure and growth of residential neighborhoods in American cities*. Washington, DC: US Federal Housing Administration.

McHarg, I. (1965). *Design with nature*. Garden City, NY: Natural History Press.

Tomlin, D. (1990). *Geographic information systems and cartographic modelling*. Engelwood Cliffs, NJ: Prentice Hall.

Tomlinson, R. (1967). *An introduction to geographic information system of the Canada land information inventory*. Ottawa: Department of Forestry and Rural Development.

Oxbridge Geographies

E. Pawson, University of Canterbury, Christchurch, New Zealand

Introduction

The geography departments of Oxford and Cambridge have distinguished histories. It was to both universities that the Royal Geographical Society (RGS) turned in the late nineteenth century when it sought to secure the institutionalization of the discipline. This story is well known, and often told in terms that imply a prominent degree of Oxbridge exceptionalism: that geographers at the two universities led where others subsequently fol lowed. In reality, the development of geography at each place was more uneven, and for long more vulnerable, than this portrayal suggests. It has also often been along divergent paths, so that it is appropriate to speak, in the plural, of 'Oxbridge geographies'.

Recent scholarship highlights the significance of contests for resources and the contingency of wider networks of support, from within and beyond institutions, in shaping the evolution of academic disciplines. This article approaches its subjects from this viewpoint. It also highlights some of the similarities between the two uni versities. For example, the provision of college based teaching systems has lent quite distinctive characteristics to student experiences of 'Oxbridge', a term which emerged in the twentieth century to distinguish the two old English universities from the host of later municipal, Redbrick, and 1960s campuses.

Establishment

The RGS Initiative

Since the end of the sixteenth century, there had been a continuous tradition of geographical teaching at both Oxford and Cambridge. This occurred in various guises, notably as 'descriptive' and 'mathematical geography', the roots respectively of regional geography and of cartog raphy. One of its best known exponents was the six teenth century Christ Church writer Richard Hakluyt. By the late nineteenth century, as the universities moved away from their essentially humanist conceptions of education and the curriculum, geography existed as an adjunct to the historical disciplines of classics, geology, and history itself. Assumptions that it was somehow institutionally secure because of associations with im perialism may hold for mainland Europe. They are not true for Britain where geography was poorly taught in schools, usually by nonspecialists, and lacked an in dependent presence in the universities, a necessary condition for the development and maintenance of an academic discipline in the way in which that concept is now been understood.

This was the context in which the Council of the RGS commissioned a report on geographical education from Scott Keltie, its secretary. It was published in 1886. Keltie observed that there were at least 45 chairs of geography in continental universities. His report paved the way for successful proposals that the RGS help fund the estab lishment of a readership in geography at Oxford and a lectureship at Cambridge. This occurred in 1887 and 1888, respectively, although any guarantee of disciplinary continuity was a long time in the making. Lectures in geography continued to contribute only toward papers in other subjects. Specific diplomas in the subject were not first awarded in Oxford until 1901 and in Cambridge even later, in 1907.

Although Oxford gave the title of professor to A. J. Herbertson, in 1910, it was a personal chair that died with him in 1915. A permanent appointment at this level was not made until 1932, one year after the move in Cambridge. This gradual cementing of disciplinary reproductive cap acity reflected the rapid growth in the popularity of the subject in the meantime in schools, driven in part by its practical and field based focus. Membership of the Geo graphical Association (GA), established to support geog raphy in the schools, rose from around 100 at the turn of the century to over 4000 by 1921. Yet these figures disguise the effort that went in to achieving a hard won place for geography at university level.

Imaginative Advocacy

In the universities, teachers of geography faced ongoing contests for respectability when, as now, a clear discip linary rationale could be difficult to articulate with con viction. In these early years, most teachers had degrees in subjects with long established cachet, such as history, law, geology, or chemistry. Geography could only make a name for itself through imaginative advocacy. Two advocates bestrode the stage, both holders of the Oxford Readership. One was Herbertson, the second reader; the other was the first occupant of that position, Halford Mackinder. While Mackinder is nowadays best known for a range of edu cational and political attainments probably not matched by a geographer since, the quiet and persistent groundwork laid by Herbertson was also of great importance in se curing a wider profile for an emergent discipline.

The New Geography

Mackinder came to notice promulgating what he called 'the new geography' in lectures for the Oxford extension movement, crystallized in a famous paper entitled 'On the scope and methods of geography' which he delivered to the RGS in 1887. Herbertson's name would be particularly associated with a paper he published in 1905 called 'The natural regions of the world'. What then was this new geography?

In the 1887 paper, Mackinder posed a rhetorical question about geography: "Can it become a discipline instead of a mere body of information?" He urged drawing together the then common division between physical and political geographies into one whole, defined as "the science whose main function is to trace the interaction of man in society, and so much of his environment as varies locally." Mackinder's geography is usually remembered for his defining work in geopolitics, with its imperial and nationalistic backdrop. But he was also driven by a humane desire to understand human life in relation to the environment. Herbertson picked up this theme, arguing in 1905 that, "Until we have a systematic classification of environments it is obviously impossible to study with any precision the effect upon the organism," in the sense of "a right understanding of social, political and historical problems."

Both Mackinder and Herbertson knew the importance of carrying the new geography into schools through the agency of trained teachers. Mackinder was instrumental in setting up the GA, the founding meeting of which took place in his college, Christ Church, in 1893. Herbertson was its honorary secretary from 1900 to 1915. Their work in codifying geography at both university and school levels came together in their vacation courses for teachers held at the Oxford School of Geography.

Oxford Ascendent

By 1920, 12 out of 35 recognized teachers of geography in British universities or university colleges were former students of the Oxford School, which had both the profile and, through this, the better students. Consequently, the contrast between the early history of geography at Oxford and in Cambridge is marked. Oxford's success was, though, largely dependent on Mackinder and Herbertson. Mackinder retained the readership until 1905, while also becoming the principal of the new university extension college at Reading in 1892, a founding lecturer at the London School of Economics (LSE) in 1895, and its director in 1903. Herbertson, in the years before his death in 1915, was delivering between two thirds and three quarters of all the Oxford lectures, in contrast to the situation in Cambridge where, from 1908, such responsibilities were more evenly spread between three university lecturers.

Consolidation

Tripos and Honours

If Oxford had outshone Cambridge geography before World War I, the situation was reversed thereafter, not least as the consequence of Oxford's long reliance on two individuals. After Herbertson's death, the readership was not filled until 1919, and despite the desire of the RGS to see an Honours School established in Oxford this did not happen for some time. It did occur in Cambridge where a two part Tripos was introduced, also in 1919. With the requirements of the Tripos, and greatly increased numbers of students post war (113 attending in 1919–20), the scale of teaching changed dramatically, and the first Cambridge degrees in geography were awarded in 1921. This was at the same time as Leeds, and a little later than Liverpool, Aberystwyth, and London.

The focus of the Cambridge Tripos was a mixture of systematic and regional geography. The six Part I compulsory papers were in physical geography; political and economic geography; cartography; history of geography; anthropogeography; and regional geography. The five Part II papers, of which a candidate was required to offer either two or three, were in surveying; geomorphology; oceanography and climatology; historical and political geography; and economic and commercial geography.

Oxford, meanwhile, not only retained a more regional focus, but was also stuck with its Diploma alone: and in 1930, only seven of these were awarded. Because applicants for university posts were by then required to have Honours degrees, this outlet for Oxford trained geographers had, in contrast to earlier years, dried up. After Honours had been introduced at Durham in 1930, only Oxford and St. Andrews among British universities teaching geography were left without the equivalent of an Honours School.

New Chairs

The Oxford situation was rectified in 1931, the same year that Frank Debenham became the first professor of geography at Cambridge. In 1932, Kenneth Mason became the first holder of the new chair at Oxford. Both appointments reflected the lingering influence of the RGS, whose pertinacity, through ongoing representation and provision of funding, had ensured that the subject had survived in both universities. Even so, this was after the establishment of redbrick geography, the first chair having been set up at Liverpool in 1917.

Mason was born in England but worked for much of his career for the survey of India and had explored and climbed in the Himalayas. Debenham was an Australian polar scientist, who had gone south with Scott on the *Terra Nova* expedition of 1910–13. Debenham had been appointed to a Cambridge lectureship in 1919, and held a BA in arts and a BSc in geology from Sydney, and a Cambridge BA and MA. Mason, by contrast, was trained at the Royal Military Academy, Woolwich, and the School of Military Engineering, Chatham, and had no previous university experience. Perhaps not surprisingly, he encountered ongoing difficulties in the reception accorded to geography in Oxford.

By the late 1930s, however, Mason had done much to restore its position in the university, despite not being an obvious choice for a department seeking to reclaim its academic credentials. In 1939, there were 39 Oxford Honours finalists. Cambridge, meanwhile, was gaining a considerable reputation for both geography and polar research, thanks to Debenham, who was also instrumental in the establishment of the Scott Polar Research Institute, which opened in 1934. However, it was during World War II that both departments undertook work that has since been seen as 'a monument in geographical writing'. This was the production of the Naval Intelligence Handbooks for the Admiralty.

Naval Intelligence Handbooks

Between 1941 and 1946, two teams of authors based in Oxford and Cambridge produced 31 regional titles (in 58 volumes), a massive undertaking to supply up to date accounts of countries that naval ships were likely to visit. In 1940, the Director of Naval Intelligence had asked Mason and his colleagues in Oxford to prepare a number of reports of immediate significance for the conduct of the war (such as the Norway landings). The *Handbooks* team in Oxford was also led by Mason, and included most lecturing staff in geography, as well as contributors from other disciplines. With its expertise in regional geography, and Mason's willingness, the task was largely handled 'in house'. In Cambridge, Debenham was not prepared to release his colleagues for such work, and it was instead coordinated by an editor in chief, H. C. Darby. Darby assembled a more institutionally dispersed team, including several geographers from the London colleges.

It is tempting to see the compendious handbooks as the acme of an 'Oxbridge geography' as it had developed over the preceding 60 years. The description may not be inaccurate for Oxford. It seems less so for Cambridge, which was establishing a reputation for field based studies in both physical and historical geography. Two examples are James Steers' *The Coastline of England and Wales*, first published in 1946, and Darby's *The Draining of the Fens*. Steers had been awarded a first in the first examination for Part II of the Cambridge Tripos in 1921, and in 1949 was to become professor and head at Cambridge. In 1966, he was in turn succeeded by Darby, who had been awarded the first Cambridge geography PhD, for his Fenland work, in 1931.

Diverging Fortunes

Post-War Oxford

The history of Oxford geography after World War II is not writ large like that of Cambridge. Its faculty mostly lacked the imaginative geographers that Cambridge fostered into legend during the next 30 years. It had relinquished to the other place the torch of disciplinary reproduction that it had carried after the turn of the century. Mason's entry in the *Oxford Dictionary of National Biography* observes that the staff that he appointed, during his tenure of the Oxford chair, which lasted until 1953, were almost wholly 'products of the Oxford school with little commitment to research and publication'.

Oxford did nonetheless produce graduates who went on to make notable careers in the discipline, although this was usually via doctoral research elsewhere. For example, Chauncy D Harris, who became an urban geographer of renown, returned to the US to a PhD at Chicago after earning his Oxford BA on a Rhodes scholarship in 1936. Richard Chorley, whose work subsequently revolutionized British physical geography, undertook his PhD with Arthur Strahler at Columbia after graduating from Oxford in 1951. Chorley's contemporary, Yi Fu Tuan, later a widely admired cultural geographer, went to Berkeley for his PhD. But in his autobiography, in 1999, Tuan's only observation on his Oxford geography training was that in his day the professors at both Oxford and Cambridge were explorers.

Tuan revisited Oxford in the mid 1960s and gave a memorable course of lectures on arid lands. This was a time when in house lectures were accorded little status, certainly in comparison to the main focus of the colleges: tutorial teaching. Tutorials (known in Cambridge as supervisions) emphasized wide reading and frequent writing, often taking the form of discussion of weekly essays between student (or students) and don (the Oxbridge term for an academic). There were some outstanding tutors, such as J. N. L. Baker of Jesus College who taught a generation of Oxford geographers before and after the war. His successor at Jesus, Ernest Paget, gave stimulating tutorials, particularly in social geography that were later to be marked in a *festschrift* assembled by former students who became notable figures in this field. They included David Ley at the University of British Columbia and Ceri Peach and Colin Clarke at Oxford.

Two of the more research oriented Oxford geographers both had connections with Cambridge. Marjorie Sweeting, the only Cambridge trained geographer on the Oxford staff, moved there in 1951. As a specialist physical geographer, she became a world authority on karst landforms. Robert Beckinsale, a University of London graduate, joined the staff after the war, having worked on the *Handbooks*. He was the more typical Oxford geographer of that time, with a range of interests spanning historical, regional, and physical geography. He attained note in collaboration with Chorley, his former student. Together they produced three magisterial volumes on *The History of the Study of Landforms*, the first of which appeared in 1964.

Cambridge Ascendent

Cambridge meanwhile had become very much the center of the British geographical universe. As in Oxford, a limited number of colleges were known for geography, notably St. Catharine's, St. John's, and, later, Sidney Sussex. 'Catz' under Steers had been a mecca for aspiring geographers in the inter war period, drawing in undergraduates such as Darby, Oskar Spate (later to be foundation professor at the Australian National University), and A. A. L. Caesar. It was Caesar who was to become a renowned tutor at St. Catharine's in the post war years, cementing its position as the home of an ever extending network of geographers. In 1983, Steers estimated that about 80 university posts, including 20 chairs, were or had been held by 'our men'. They included Michael Chisholm and Gerald Manners, who edited a volume published in 1971 to endow a college prize in honor of Caesar, their old tutor. It contained essays from, among others, Peter Hall, Peter Haggett, David Keeble, and Kenneth Warren.

Chorley and Haggett's Cambridge

Steers, who has been called an 'Olympian' head of department, presided over a place that was a ferment of innovation, which included Peter Haggett as student and, from 1957 to 1966, lecturer; Richard Chorley, as lecturer, reader and, from 1974 to 1994, professor; and David Harvey, a postgraduate student at St. John's, who went to lecture at Bristol in 1961. Haggett, in an obituary of Chorley published in 2002, described as 'benign' the environment that Steers provided in which Chorley's 'experiments could take place'. These were based on "a quantitative model based paradigm with an emphasis on General Systems Theory and numerical modelling" (Haggett, 2002). This was in direct contrast to the situation in Oxford, where the only geomorphic topic then offered for finals was the 'Geographical cycle of erosion'.

In contrast, Chorley had recalled, in an essay entitled 'Haggett's Cambridge', how Haggett had been called into Steers' office after a presentation to the RGS – of what was to become a landmark paper using quantitative analysis in a regional context – to be told that he was bringing the discipline into disrepute. This may have been the fear of geography's older establishment, but Haggett was soon to publish *Locational Analysis in Human Geography*, the book that did more than any other to promote the quantitative revolution and to position the subject in the excitement of university expansion in the 1960s. In it, Haggett formulated a model for human geography that was closely linked to recent developments in geomorphology. With Chorley, he worked to ensure the reach of the new approach by reviving the summer schools for teachers that had been popular years before in Oxford. These took place at Madingley, outside Cambridge, from 1963 until 1978. The early ones provided an audience that acted as a sounding board for the ideas that became *Frontiers in Geographical Teaching* and *Models in Geography*.

The Cambridge–Bristol Axis

These volumes encapsulated an approach to geography that had gained traction in the United States in the 1950s. Excepting Beckinsale's contribution to *Frontiers*, they are notable for their lack of Oxford contributors. In contrast to many of their Cambridge counterparts, most Oxford students had to discover such books for themselves. Taught to a curriculum that might have been recognizable to a student of 50 years before, some were even told in the late 1960s – in an echo of Steers' warning – not to include these new ideas in their finals papers. But there were a few Oxford lecturers, such as Ceri Peach and Richard Jackson, who did provide a more contemporary perspective for students.

There was, in fact, anxiety in some quarters in both Oxford and Cambridge (and in many other places) as to whether geography was to be captured by the new positivistic approaches, or if it was to uphold what was seen as the heritage of Mackinder and Herbertson. A fervent supporter of this heritage was the Oxford lecturer W. H. Parker, who in 1958 had responded to an attack on systematic and regional geography that appeared in the *Universities Quarterly*. The irony of such a position was that Chorley and Haggett worked hard to establish links between the natural and the behavioral sciences. Soon it was not uncommon to refer to the 'Cambridge–Bristol axis' of innovation in British geography. (Haggett had moved to Bristol in 1966, to a department chaired by R. F. Peel, another former Catz geographer.) This axis was widely understood to refer to the British nodes of the quantitative revolution and also to bypass Oxford, its geographic midpoint.

Oxbridge Geographies since the 1970s

Progress

There was a keen sense of anticipation in Oxford when the celebrated French geographer, Jean Gottmann, who had worked for many years in America, was appointed to the chair in 1968. Gottmann's brief was to revive the fortunes of Oxford geography. He had made his name with the publication of a compelling regional geography about the new urban form emerging on the Eastern Seaboard of the United States, *Megalopolis*. If this term penetrated more deeply into public consciousness than anything more directly associated with the Cambridge–Bristol axis, Gottmann nonetheless lacked connections with the proponents of the new geographies diffusing from that axis.

He did initiate the reorganization of the Oxford syllabus and backed this by breaking with the internally focused appointments policy of his predecessors, Kenneth Mason and Edmund Gilbert (1953–67). Some of his earliest appointments were of Cambridge trained geographers: Andrew Goudie (who succeeded him as professor in 1984), John Patten, and Kenneth Warren. Such moves began to re energize the Oxford School, while at the same time strengthening traditional interests of Oxbridge geographies in geomorphology and historical geography. When a second chair, in the name of Halford Mackinder, was introduced, the first holder (from 1974 to 1983) was John House, a Jesus man, previously at Newcastle. His field of regional economic problems was again a not uncomfortable fit with Oxford interests. At the same time, Ceri Peach was attracting a line of doctoral students in social geography, including Peter Jackson and Susan Smith, who have gone on to forge distinctive contributions to the cultural turn in human geography.

The new geographies characteristic of Cambridge were meanwhile being widely promulgated with the establishment in 1969 of the annual *Progress in Geography* series. It counted Richard Chorley, Peter Haggett, and David Stoddart among its founding editors. By 1976 it had divided into the two ‘piglets’, the serials *Progress in Physical Geography* and *Progress in Human Geography*. These remain some of the highest impact factor journals in the discipline.

Growing Diversity

Cambridge geography in the 1980s was marked by vigorous internal debate during the headship of Michael Chisholm, professor from 1974 to 1994 and a keen proponent of the positivistic, model based approaches rooted in the 1960s. This was a time when younger dons, such as Derek Gregory, Ron Martin, and Graham Smith, were working to convey to undergraduates “a sense that geography had much to contribute to the other social sciences but that recent advances, particularly in critical theory, were pivotal to making sense of our geographical world.” They were supported by people like Stoddart, a distinguished physical geographer and acolyte of Steers, but openly sympathetic to such interpretations. This atmosphere helped to disseminate a more engaged, theoretically informed, and critical human geography, as Cambridge renewed its role as a center of innovation.

A moment in this new engagement is captured in a photograph of student and senior members of the Vidal de la Blache Society, the undergraduate geographical society at Sidney Sussex College, in May 1983 (**Figure 1**). This is the college of which Gregory and Smith as well as Chorley were fellows; they were joined on this occasion by Haggett. Although all the senior members were male, there are more women students in the photograph than there are men. Students pictured who went on to university chairs as critical human geographers include Nick Fyfe, Gillian Rose, Chris Philo, and Tony Bebbington. Both Gregory and Stoddart moved to chairs in North America, at the University of British Columbia, and at the University of California, Berkeley, respectively. Just as St. Catharine had played a key role in disseminating new geographical orthodoxies in the 1950s and 1960s, so too did Sidney Sussex both with and for a later generation.

In response to this latest reorientation of human geography, in 1987 Oxford again selected a celebrated geographer, electing David Harvey as second holder of the Mackinder chair. Harvey has recorded how many college dons “treated me as if I had just arrived from Cambridge,” regarding the intervening years, including nearly two decades spent at Johns Hopkins in Baltimore, as “some unwanted exile.” Although the best known social theorist in human geography, Harvey has recorded his gratitude to the Cambridge trio of Chorley, Haggett, and E. A. Wrigley for the ‘liveliness’ that they had imparted to geography during his Cambridge years and for many thereafter. In Oxford, stung by the weakness of intellectual responses to Thatcherism, he wrote *The Condition of Postmodernity* and began the work that became *Justice, Nature and the Geography of Difference*.

Harvey stayed in Oxford for six years, an inspiration to students, and, adding his influence to appointments and the syllabus, furthering the process already underway of prising open the department to a greater diversity of contemporary geographies and geographers. The Oxford School today maintains ‘Spaces of globalization’ as one of its research themes. But the real focus of growth in the 1990s and beyond has been in a wide range of environmental research, spearheaded by Andrew Goudie. It now has a large faculty drawn from a broader range of backgrounds than at any previous time. The Environmental Change Institute was founded as an allied interdisciplinary unit in 1991; now both the Institute and the School are the keystones of the Oxford University Centre for the Environment. In all, the Centre today has ten professors, four of whom are women, reflecting a shift away from the masculinist tradition of Oxford geography.

The Cambridge department also has a number of environmentally related research clusters, linking society,

Figure 1 Members of the Vidal de la Blache Society, Sidney Sussex College, May 1983. Students in the photograph include Nick Fyfe and Gillian Rose, at each end of the back row, Chris Philo, fourth from the left in the middle row, and Tony Bebbington, in the center of the front row. To the left and right of Bebbington are Derek Gregory and Graham Smith, while forming the 'bookends' of the front row are Peter Haggett (left) and Richard Chorley (right). The illustration is reproduced from plate 18 in Haggett, P. (1990). *The Geographer's Art*. Oxford: Basil Blackwell.

environment, and development. There are 11 professors, only one of whom is female. It is Cambridge rather than Oxford that now retains a high proportion of home grown lecturing staff. Traditional strengths remain ap parent in a focus on polar research, on environmental processes, and on 'spaces of economy and society', a cluster which includes scholars such as Bob Bennett and Ron Martin. It also maintains a commitment to its legacy of historical geography, including the demographic work of the Cambridge Group for the History of Population and Social Structure (one of whose founders was E. A. Wrigley).

Oxbridge Today

In the first three Research Assessment Exercises, between 1986 and 1992, Oxford and Cambridge were both in the top category. In 1996 and 2001, Cambridge outscored Oxford. This, however, is only one way of positioning Oxbridge geographies. Equally if not more pertinent for the future of the discipline is the greatly enhanced size of the graduate schools in each place, compared to earlier times. This is in terms of both numbers of PhD students and those enrolled in that currently popular form of teaching, specialist masters courses. Cambridge lists over 80 PhDs and has about 20–30 a year on taught masters courses. Numbers in the Oxford Centre for the En vironment reflect its growth in standing: about 100 re search (mostly D Phil) students and 120 in taught masters courses.

Nonetheless, the majority of Oxbridge geographers are, as has always been the case, undergraduates. The continuing appeal of Oxbridge geographies to them lies only partly in relative standings. There are many other attractions: a college based, and convivial, education; the opportunity to learn not just within one's subject but to engage actively with many others; and the means that each university provides for exercising the imagination in many different spheres. It was after all the imagination that Mackinder identified as being crucial in the edu cation of a geographer.

Acknowledgments

The author is grateful to Robin Flowerdew, Pip Forer, Wendy Lawson, John Overton, Peter Perry and Simon Swaffield for discussing their experiences of Oxbridge geographies with me; to them, as well as to Ross Barnett, Andrew Goudie, Ron Johnston, David Ley and James Sidaway, for comments on an earlier draft; and to Nick Fyfe, Peter Haggett and Chris Philo for help with the photograph.

See also: Cultural Turn; Gottmann, J.; Harvey, D.; Historical Geography; Mackinder, H. J.; Quantitative Revolution; Redbrick University Geography in Britain; Regional Geography I; Social Geography.

Further Reading

Balchin, W. G. V. (1993). *The Geographical Association: The First Hundred Years*. Sheffield: The Geographical Association.

Barnes, T. J. (2002). Performing economic geography: Two men, two books, and a cast of thousands. *Environment and Planning A* 34, 487 512.

Chisholm, M. and Manners, G. (eds.) (1971). *Spatial Policy Problems of the British Economy*. Cambridge, UK: Cambridge University Press.
Chorley, R. J. (1995). Haggett's Cambridge: 1957 1966. In Cliff, A. D., Gould, P. R., Hoare, A. G. & Thrift, N. J. (eds.) *Diffusing Geography: Essays for Peter Haggett*, pp 355 374. Cambridge, MA: Blackwell.
Clarke, C., Ley, D. and Peach, C. (eds.) (1984). *Geography and Ethnic Pluralism*. London: George Allen & Unwin.
Clout, H. M. and Gosme, C. (2003). The Naval Intelligence Handbooks: A monument in geographical writing. *Progress in Human Geography* 27, 153 173.
Coones, P. and Stoddart, D. R. (1989). A hundred years of geography at Oxford and Cambridge. *The Geographical Journal* 155, 13 32.
Gregory, D., Martin, R. and Smith, G. (eds.) (1994). *Human Geography. Society, Space and Social Science*. Minneapolis, MN: University of Minnesota Press.
Haggett, P. (1990). *The Geographer's Art*. Oxford: Basil Blackwell.
Haggett, P. (2002). Cambridge geographer behind a revolution in landform study. *The Independent*, Saturday, 18 May.
Harvey, D. (2002). Memories and desires. In Gould, P. & Pitts, F. R. (eds.) *Geographical Voices: Fourteen Autobiographical Essays*, pp 149 188. Syracuse, NY: Syracuse University Press.
Herbertson, A. J. (1905). The natural regions of the world. *Geographical Teacher* III, 104 113.
Johnston, R. J. and Williams, M. (eds.) (2003). *A Century of British Geography*. London: The British Academy.
Mackinder, H. J. (1887). On the scope and methods of geography. *Proceedings of the Royal Geographical Society NS* 9, 141 160.
Parker, W. H. (1958). Geography defended. *Universities Quarterly* 13, 34 44.
Scargill, D. I. (1976). The RGS and the foundations of geography at Oxford. *The Geographical Journal* 142, 438 461.
Steel, R. W. (ed.) (1987). *British Geography 1918 1945*. Cambridge, UK: Cambridge University Press.
Steers, J. A. (1983). Geography at St Catharine's College. *St Catharine's College Society Magazine* 20 23.
Stoddart, D. (1986). *On Geography and Its History*. Oxford: Basil Blackwell.
Tuan, Y. F. (1999). *Who Am I? An Autobiography of Emotion, Mind and Spirit*. Madison, WI: University of Wisconsin Press.
Withers, C. W. J. and Mayhew, R. J. (2002). Re thinking 'disciplinary' history in British universities, *c*. 1580 1887. *Transactions of the Institute of British Geographers NS* 27, 11 29.

Relevant Websites

http://www.geog.cam.ac.uk
Department of Geography, University of Cambridge.
http://www.oxforddnb.com
Oxford Dictionary of National Biography.
http://www.ouce.ox.ac.uk
Oxford University Centre for the Environment.

P

Paasi, A.

R. Jones, Aberystwyth University, UK

Glossary

Institutionalization of Region The contingent process through which regions may assume social and spatial coherence.

Lifepaths The way in which an individual's biography or life history is intertwined with their life geography.

Anssi Paasi, a professor of geography at the University of Oulu in Finland, is one of the world's foremost political geographers. He was born on 14 September 1955 and was subsequently educated at the University of Joensuu in Eastern Finland (at master's, licentiate, and doctoral levels) in the period between 1979 and 1986. Following his doctorate, he was appointed to the post of docent in the University of Joensuu and was subsequently pro moted to a chair in geography in the University of Oulu in 1990. He still remains a docent of human geography in the University of Joensuu and was also elected as an honorary professor of human geography in the Uni versity of Wales, Aberystwyth, in 2006. He has, add itionally, carried out research as a visiting scholar in the University of Nijmegen, Loughborough University, and UCLA. To date, he has published over 150 papers, reviews, and chapters, as well as 19 books and monographs in both Finnish and English. His work has centered on understanding certain key concepts within political geography, most notably those relating to the region, nationalism, and terrriories/boundaries. The empirical focus for Paasi's research throughout his career has been on Finland and its relationship with a broader Baltic region.

In many respects, the fact that Paasi is a Finnish geographer has had a profound effect on his career. Per haps most prosaically, Paasi's nationality has meant that he has conducted the vast majority of his empirical re search on Finland. Similarly, his nationality has also meant that he has published a large proportion of his research in Finnish language publications, especially during the early stages of his career. But it is clear that Paasi's Finnish identity has also impacted on his research career in more subtle, yet equally far reaching ways. In the first place, the fact that Paasi is a Finnish geographer has meant that he has always worked on the relative 'margins' of human and political geography that is dom inated, in large measure, by geographers living and working within an Anglo American academy. Anssi Paasi is thus to be congratulated for being able to lead academic debates within contemporary political geography, despite this relative marginalization. At the same time, it is arguable that some of Paasi's potential impact on human and political geography in an international context may well have been dissipated as a result of being a voice speaking from the 'margin'. The clearest indication of this relative loss of academic impact, of couse, is the fact that a large proportion of Paasi's research has been published through the medium of Finnish and has not, therefore, become part of an international academic debate con ducted, for better or worse, in English.

In order to understand a second subtle influence of Paasi's Finnish identity on his academic career, one needs to appreciate the history of human geography within Finland. Paasi has noted that the character of human geography within Finland during the 1970s was affected by its academic affiliation to faculties of science within the vast majority of universities and by the dominance of an older generation of human geographers who were proponents of quantitative and positivist approaches. These twin influences made it difficult for a new generation of critical human geographers, such as Anssi Paasi, to gain a foothold within academic geography in Finland, especially during the 1970s and 1980s. Paasi's early work on regions, for instance, was deemed by certain individuals within an older generation of Finnish human geographers to be marginal to Finnish human geography and, indeed, he became labeled as a 'geosociologist' rather than a human geographer. Paasi, in this regard, has considered himself to be a geographer who was in the 'wrong place'. It was also perhaps as a result of this relative antipathy to his work in Finland, at least during the 1980s, that Paasi increasingly sought to publish his research through the medium of English.

In a second context, one must also appreciate Paasi's specific lifepath 'within' Finland (see **Figure 1**). As noted above, Paasi received his formal geographic education in the University of Joensuu and this possesses two important implications for his career. In the first place, Paasi has maintained that the Department of Geography in the University of Joensuu acted as a key location for the 'social scientification' of the discipline of geography within Finland. "Close connections with fields such as history, economics, sociology and social policy," (Paasi, 2005: 608) according to Paasi, enabled human geographers within the University of Joensuu to engage more fully with some of the important philosophical currents that were affecting the humanities and social sciences internationally from the 1970s onward. But of course, the 'social scientification' of human geography taking place in the University of Joensuu during this period also contributed to the relative disenfranchisement of Anssi Paasi and other researchers at a broader Finnish scale, as discussed above. It is also important to note that the University of Joensuu lies in Finnish North Karelia and, in this respect, Paasi's time within the university can be seen to have contributed in a direct way to some of his most notable and innovative work: much of Paasi's research to date has centered around developing conceptual and empirical understandings of the Finnish territory and border, most notably in the context of the disputed borderlands of Karelia. Indeed, Paasi has made significant contributions to at least three debates in political geography: the institutionalization of regions; the sociospatial processes that help to reproduce nationalism; and the broader significance of borders and territories as a way of articulating a sociospatial consciousness.

Figure 1 Finland, showing key locations that have contributed to Paasi's career.

It is Anssi Paasi's work on the institutionalization of regions that has, arguably, cemented his reputation as one of the world's foremost contemporary political geographers. He has admitted that the inspiration for this work derived from academic debates concerning social constructionism, which were emerging during the late 1970s, as well as a series of more specific contributions by individuals such as Allan Pred and Nigel Thrift, which were beginning to interrogate the meaning of regions in more theoretically sophisticated ways. Paasi believed, nonetheless, that these various contributions failed to explain fully how 'regions became what they were'. It was with this context in mind that Paasi sought to provide a new theoretical and methodological way of comprehending how regions – understood as a sociospatial unit and a version of a 'higher scale' history into which individuals are socialized – are institutionalized. Four, mutually constituting, stages are said to be key to the process of institutionalizing regions. The first relates to the 'territorial shape' of a region or, in other words, the degree to which the territorial extent of a region becomes embedded in individual or collective consciousness in various social practices. Second, Paasi draws attention to the importance of the 'symbolic shape' of a region, which refers to the process through which particular symbols come to represent the region in question. Third, a region may derive 'institutional shape' from the broad range of institutions, that give some formal structure to social and spatial life within the region. Finally, Anssi Paasi describes the way in which regions may become 'established' within the sociospatial consciousness and may assume "the material expression of the ends to which state power is applied" (Paasi, 1991: 247). Taken together, therefore, Paasi's work is concerned with explicating the way in which regions are in a constant process of 'becoming' in a variety of different contexts. Regions may emerge and become consolidated or 'established' within society. By the same token, they may become marginal to people's lives and, indeed, they may even disappear. Equally as significant as Paasi's efforts to devise a conceptual and methodological framework for understanding regions, though far less evident in an international and specifically English medium context, has been his empirical study of the institutionalization of four Finnish regions: Lapland, Southern Ostro Bothnia, Northern Karelia, and Uusimaa. This project used a range of archival material, including administrative documents, maps and atlases, regional newspapers, and school textbooks, along with 2600 questionnaires in order to explicate the extent to which the regions in question had become institutionalized as part of the regional structure and within the sociospatial consciousness of the inhabitants of the four regions.

Anssi Paasi's other two key areas of academic intervention have followed on, in many ways, from his earlier work on the institutionalization of regions. Some of his more recent work has examined the reproduction of nations and national territories – or, alternatively, the institutionalization of the nation – specifically in the context of Finland. This work, in conceptual terms, has attempted to show how nations, national territories, and their boundaries are reproduced through a combination of a sociospatial consciousness – centered on the activities of the state's agents at the national scale – and a social representation, which lies within the domain of more ordinary people within localities. Equally as interesting within this work was Paasi's empirical focus on Finnish national identity, particularly in the context of the role played by the border region of Karelia in reflecting and shaping the sociospatial consciousness and social representation associated with Finnish nationalism. In this way, Anssi Paasi has been able to show how the reproduction of a Finnish nation and national territory has come about as a result of a combination of formal state building projects and the more mundane and day to day practices of ordinary people. Paasi's concern with the role of the Finnish–Russian border in defining the Finnish nation has also encouraged him to engage in a second new area of research in recent years, namely, his attempt to think about the broader significance of boundaries in a contemporary globalizing world. This work has been part of a broader academic project to deliver a more conceptual attuned understanding of borders and boundaries. Part of this project has revolved around the need to view boundaries and borders as socially produced. Geographers such as Anssi Paasi have drawn inspiration from social and cultural theory to show how boundaries should be viewed as contingent processes in motion. They are, moreover, geographical entities that are produced as a result of a mixture of processes; some are undoubtedly state driven but other actors within civil society are also implicated in the process of 'narrating boundaries'. Moreover, globalization has not undermined the significance of boundaries but has merely made the various processes that are associated with the reproduction of contemporary borders more complicated and contingent.

Anssi Paasi's work, when viewed as a whole, has formed a remarkably coherent academic project over a number of years. He has drawn inspiration from new developments in social and cultural theory in order to rethink the significance of key concepts in political geography, most notably those relating to regions, nations/territories, and boundaries. The impact of his work has been widespread in political geography and further afield and is highly likely to remain so in future years.

See also: Nationalism; Political Geography; Region.

Further Reading

Berg, L. and Kearns, R. (1998). America unlimited. *Environment and Planning D: Society and Space* 16, 128–132.

Daniels, S. and Nash, C. (2004). Lifepaths: Geography and biography. *Journal of Historical Geography* 30, 449–458.

Newman, D. and Paasi, A. (1998). Fences and neighbours in a postmodern world: Boundary narratives in political geography. *Progress in Human Geography* 22, 186–207.

Paasi, A. (1986). *The institutionalization of regions: Theory and case studies*. Publications in Social Sciences no. 9. University of Joensuu, Joensuu.

Paasi, A. (1991). Deconstructing regions: Notes on the scales of human life. *Environment and Planning A* 23, 239–256.

Paasi, A. (1996). *Territories, Boundaries and Consciousness*. Chichester: Wiley.

Paasi, A. (2003). Boundaries in a globalizing world. In Anderson, K., Domosh, M., Pile, S. & Thrift, N. (eds.) *Handbook of Cultural Geography*, pp 462–472. London: Sage.

Paasi, A. (2005). Between national and international pressures: Contextualizing the progress of Finnish social and cultural geography. *Social and Cultural Geography* 6, 607–615.

Relevant Websites

http://www.oulu.fi
Anssi Paasi's homepage.

Pacific Rim

X. Chen, Trinity College, Hartford, CT, USA

Glossary

De-Bordering A dynamic and complex process in which the barrier role of borders is shifting to the role of bridges, facilitated more by cross-border economic flows.

Formal Regionalism State-led multilateral initiatives for regional economic integration and cooperation through intergovernmental negotiations and signed agreements between two or more sovereign nations.

Informal Regionalism Initiatives of combined local (municipal) state policies and private sector activities for promoting cross-border (sub)regional economic transactions and development, generally without national governments' involvement.

Re-Bordering A dynamic and complex process in which the bridging role of borders is shifting back to the role of barriers, brought about more by state action.

Transborder Subregion A relatively large transnational spatial-economic entity consisting of contiguous and/or adjacent subnational and local border areas of multiple countries.

Introduction

The Pacific Rim has a general geographic reference to the countries located on, along, and within the physical boundaries of the Pacific Ocean. However, the Pacific Rim is much more than a geographic concept. It denotes a broad and differentiated geographic landscape imbued with multiple and layered political, economic, and cultural dynamics and meanings. As these dynamics unfold over time to yield a temporarily shifting view of the Pacific Rim as a geographic structure, the meanings of the Pacific Rim evolve with it. With rapid and continued economic growth in the main geographic parts of the Pacific Rim – East and Southeast Asia – through the last quarter of the twentieth century, the heralding of the twenty first century as the 'Pacific Century' started to spread globally even against and through the temporarily dampened optimism during the Asian financial crisis of 1997–98. While the twenty first century is still young, it provides a temporal vantage point to take stock of the Pacific Rim as a dynamic geo graphic construct and its associated politicoeconomic and sociocultural dimensions.

A more conventional geographic treatment of the Pacific Rim is likely to follow the established regional divisions of East Asia, Southeast Asia, South Pacific, and whatever other countries that are not neatly labeled categorically. It may bring in the west coast of the United States, South America, and Australia but may not deal with them in depth and detail. This conventional view would see the Pacific Rim as consisting of several spatial blocks that organize different sets of political and eco nomic relations between countries. The latter are taken as constituent units of these geographic groupings and also constitute the larger set of bilateral relations span ning the entire Pacific Rim. While this approach may look for subnational dimensions and variations of these units and ties, it is premised on a state centric perspective on the Pacific Rim. The combination of conventional regional divisions and state centric re lational account has its advantages in that it not only establishes the broad geographic composition of the Pacific Rim but also identifies uneven relations that bind together clearly bounded countries. But this treatment can no longer fully capture the complex realities of the Pacific Rim.

The Pacific Rim today calls for an alternative con ceptualization because it differs sharply from the pre 1980 Rim. Over the last quarter century or so, a multitude of economic and geographic forces have unfolded in both convergent and divergent manners across the dual axes of regions and borders, reshaping the Pacific Rim in the process. The regional axis features the complementary and divergent trend of formal regionalism versus informal regionalism. The axis of borders exhibits the simultaneous processes of de bordering and re bordering. As the two axes intersect, they have altered the geographic contour of the Pacific Rim, created new political and economic networks and alliances, and activated old and/or en gendered new social and cultural ties and meanings that invoke and may constitute real or imagined regional communities.

Formal Regionalism

Formal or top down regionalism or regional integration has an obvious powerful impact on the Pacific Rim through strengthening various region wide or subregional sets of economic and political linkages, while weakening few others by excluding them. Unlike the European Union (EU), which typifies formal regionalism to the tilt, formal regionalism in the Pacific Rim or any Asian context has been weak in terms of producing anything that approaches the EU in form and function. The closest entity approxi mating Pacific Rim wide integration is Asia Pacific

Economic Cooperation (APEC), which consists of 21 member economies on both sides of the Pacific Rim. A relatively loose organization, or cooperative process more precisely, APEC holds an annual forum to promote intergovernmental, multilateral initiatives for regional economic integration and cooperation. As this process continues, with improved official economic relations and limited real economic benefits, it has been accompanied by other subregional trade promotion deals such as AFTA (the Association of Southeast Asian Nations (ASEAN) Free Trade Area) and ASEAN + 3 (China, Japan, Korea). The latter has recently entertained a proposal to become 'ASEAN + 6' by including India, Australia, and New Zealand. With 3.1 billion people, or half of the world's total population, and one third of the world's total GDP, ASEAN + 6, if materialized, can truly be the third pole of the world economy vis à vis the EU and NAFTA.

To the extent that the large scale realignment and regrouping of major Asian economies enhances the collective economic strength of the Pacific Rim, primarily its western side or half, it also is producing a more layered and jumbled structure of formal regionalism. To complicate this structure further, the East and Southeast Asian governments have developed specialized mechanisms for cooperating on regional monetary and financial issues. For example, as a response to the Asian financial crisis in 1997–98, ASEAN + 3 launched the finance minister process called the Economic Review and Policy Dialogue (ERPD), which promotes regional financial cooperation through information exchange, policy discussions, and peer pressure. In May 2005, ASEAN + 3 finance ministers agreed to strengthen the Chiang Mai Initiative (CMI), introduced in 2000, by increasing bilateral currency swap size and linking CMI to the regional economic surveillance of ERPD. Given its both more general and specialized institutional arrangements, formal Asian regionalism has been characterized by such muddled metaphors as 'a patchwork of hubs and spokes in a spaghetti or noodle bowl', with noodles adapted for describing Asia.

Informal Regionalism

While the formal regionalist view on the Pacific Rim sheds a broad light on its politicoeconomic 'moving parts' and reconfiguration, it reveals little about the shifting geography and its implications for seeing other new and distinctive aspects of the Pacific Rim. To balance against the more prevalent formal regionalist perspective, it is time to formally introduce an informal or bottom up regional lens that can illuminate the forces that impinge on the Pacific Rim from below, with the potential to reshape the bulk of the Rim. Looming clearly from this view is a myriad of established and emerging cities and areas in the geographic core and along the spatial margins of both maritime and mainland Asia that play an increasingly important role in reconfiguring Pacific Asia. Although these localities are connected in a general way not that dissimilar from the 'spaghetti bowl' of criss crossing interstate trade ties, the subnational and local units are embedded in more extended and tiered spatial and economic networks. More importantly, unlike aggregate trade flows among pairs of national economies, the networks of cities anchor global–local production networks and cross border trade and tourism of varied scales that involve diverse actors including municipal governments, multinational firms, local entrepreneurs, and border crossing migrants. These networked cities spread and spill economic and logistic influences into their hinterlands within and beyond national boundaries, creating extended spaces of translocal or interlocal flows that contribute to the formation of transborder subregions and global or globalizing city regions (see **Figure 1**).

Figure 1 gives us a very different picture of the western or Asia Pacific portion of the Pacific Rim than provided by formal regionalism. Through this lens, instead of seeing a broad region of totally bounded countries and state territories, we visualize a mosaic of transborder subregions that cuts across the multiple and adjacent boundaries of countries. This alternative re mapping of a conventional map is more than an arbitrary cartographic exercise. It instead is based on a very different theoretical premise and methodological orientation.

A formal regionalist analysis would focus on explaining the level of trade integration and benefits for the parties involved, while the typical explanatory variables include national goals, size of regional groupings (number of member countries) and power differentials among them, perceived costs and benefits regarding included versus excluded trading partners, and extraregional factors such as global economic influence or superpower interference. Large developmental and power disparities threaten but do not necessarily lead to failure in regional trade integration, as exemplified by rich, globalized Singapore and poor and repressive Myanmar in ASEAN. In addition, selective and targeted protection of certain domestic sectors of countries in existing and attempted trade deals creates barriers to true openness and integration. Instead of being nation state led and formally institutionalized, the units in the informal regionalist frame – transborder subregions and global(izing) city regions – are driven by a combination of local (municipal) state initiatives and private sector activities such as direct investment by multinational firms. This fundamental difference shifts theoretical analysis from questions and concerns associated with trade and international relations theories to a more complex theoretical challenge that needs to draw analytical assumptions and insights from research on global and secondary cities, global value

Figure 1 The Western Pacific Rim in transborder subregions. Adapted from Chen, X. M. (2005). *As Borders Bend: Transnational Spaces on the Pacific Rim*. Lanham, MD: Rowman & Littlefield, Map 1.1.

chains, regional development, and network analysis, as well as on the restructuring and rescaling of the state that favors the rise and developmental behavior of the local state.

De-Bordering and Re-Bordering

The turning or rotating of the Pacific Rim along and around the formal–informal regional axis or dimension is critical to a geographic reappraisal of where it stands now and may be headed. This cannot be accomplished based on its logic alone, however. The axis intersecting and reinforcing informal regionalism is 'de bordering' and 're bordering', which describes a pair of processes in which borders are considered mutating spaces rather than fixed lines. As economic interdependence between Asia Pacific countries and between them and the global economy widens and deepens, more sections of and points on their borders are open to more intensive interactions with the forces and actors of the neighboring countries. While this process represents de bordering by making borders more porous and more difficult to control, it can prompt central and local state authorities to reassert border control, resulting in re bordering. When more open borders bring about nasty unintended consequences such as increased cross border drug trafficking, threat of epidemics, or terrorist activities, states are forced to tighten their borders as a result. Examples include heightened border control in response to 9/11 in 2001 and SARS in 2003.

Despite some re bordering that features renewed border control, de bordering appears to be stronger and more widespread at and along many land and sea borders in the Pacific Rim. It is characterized by the shift of borders' role from barrier to bridge, especially across the previously closed and heavily guarded borders. These borders used to function as political barriers that blocked economic and cultural exchanges to differing degrees. They also existed as military frontiers that received heavy defense spending but little economic assistance. As a result, the zones around these borders became marginalized. Due to changes in a variety of conditions, these borders have begun to bridge separate and isolated local and regional economies. The emerging transnational spaces provide new opportunities for global–local links, generating either self sustaining and/or externally linked local development.

The bridging role of borders has either unleashed new informal regionalist tendencies or strengthened existing ones. Bridging creates a cooperative space between binational or trinational economies. Broadly viewed, bridging generates economic growth and transactions that can spread beyond the immediate border zones to other parts of the transborder subregions. On the other hand, bridging may turn borders and border zones into new contested terrains for illegal migration, spillover pollution, and other undesirable outcomes. States may respond to these threats by reimposing border control, thus moving back toward the barrier role of borders and thus re bordering again.

Within the Pacific Rim over the last two decades or so, informal regionalism has gained strength because de bordering has occurred on an increasing scale, while formal regional integration has proceeded apace through a variety of intergovernment cooperative mechanisms such as ASEAN + 3. Formal regionalism tends to reinforce state centric interactions and relations, which may lead to more cross border trade and investment. However, its reliance on formal institutional arrangements and involvement of entire states work differently from informal regionalism, which stems from de bordering of boundaries connecting contiguous and adjacent local and subnational units and thus exerts a strong and extensive bottom up pressure on the relational structure of the Pacific Rim. In the second half of this essay, several cases are used to illustrate the coupled force of informal regionalism and de bordering in reconfiguring the Pacific Rim.

The Various Facets of Informal Regionalism and Debordering

The first illustrative case is a transborder subregion that consists of Hong Kong, Taiwan, Macau, and China's Guangdong and Fujian provinces, primarily the Pearl River Delta region in Guangdong (labeled the Greater Southeast China Subregion (GSCS) in **Figure 1**). New economic linkages across the China–Hong Kong border emerged in 1979 when the establishment of four special economic zones in southeastern China triggered the initial movement of Hong Kong's labor intensive assembly operations over the border. Taiwanese investors and tourists began to flood Guangdong and Fujian provinces in the 1980s. Two and a half decades later, a massive (re)division of labor between Hong Kong and Taiwan with southern China has been firmly entrenched. By the end of 2005, over 10 million jobs were created by over 60 000 Hong Kong and Taiwan invested factories in the Pearl River Delta alone.

The blurring of the Guangdong–Hong Kong border has been intimately linked with the 'miraculous' growth of Shenzhen from a tiny fishing town of less than 100 000 bordering Hong Kong to a sprawling industrial city of over 10 million. Shenzhen and Hong Kong have literally grown into each other and formed an extensive cross border metropolitan region. Today huge flows of people and goods go through several rail and road checkpoints along the Shenzhen–Hong Kong border, and the Luohu crossing point has become China's busiest land port for human traffic. On weekends and holidays, thousands of Hong Kong residents ride the train and then walk through the crossing points to shop, dine, and enjoy entertainment in Shenzhen where these things are much cheaper. This pattern has reversed the flow of money and wealth in the old days when Hong Kong residents would bring over consumer goods such as household electronics and daily necessities to their relatives and friends north of the border. Just as Hong Kong consumers spend money in Shenzhen, the increasingly wealthy residents in Shenzhen actually spend more money in Hong Kong. However, by buying a large number of cheaper homes in Shenzhen as residence or investment, Hong Kong residents have raised the property prices and cost of living in Shenzhen to be among the highest in China now.

The second case is what is generally known as the Greater Mekong Subregion (GMS), which comprises Thailand, Vietnam, Laos, Myanmar (formerly Burma), Cambodia, and China's Yunnan province (see **Figure 1**). Given its composition and some intergovernment cooperative arrangements, the GMS crosses over the formal–informal divide of regional integration. However, despite growing international trade among the GMS or Mekong countries, border trade has constituted the most visible and significant cross border economic tie among the key border regions within the GMS. The 1990s and beyond saw the return of Yunnan province's historical crucial role in border trade with its southern neighbors. In the first 6 months of 2004, Yunnan province handled $2.5 billion in border trade, a 25.4% rise over the same period of 2003. In the first 6 months of 2004, Vietnam registered the highest growth rate of 37% among Yunnan's border trading partners. Yunnan's six border prefectures, which are geographically contiguous with Vietnam, Laos, and Myanmar, accounted for just about all the provincial border trade. Key Chinese and Myanmar border towns are intensive spots for border trade. The end of Chinese support before the collapse of the Burmese Communist Party in 1989 facilitated the opening of more border trading posts between the two countries. The Myanmar border town of Muse became open for border trade in 1988, while Yunnan's border city of Ruili created the Jiagao Border Economic Development Zone in 1991. Myanmar's border trade with Yunnan not only has enlivened the town of Muse but also has revived cities away from the border.

Border trade across the Mekong countries, however, has been constrained by the lack of physical infrastructure, prompting renewed efforts of the region's national and local governments to improve it. The authorities of Mekong Delta provinces and cities are working hard to perfect the 'one stop' mechanism for administrative procedures in order to meet investors' requirements. They attach priority to road and waterway systems, airports, and ports to facilitate goods transportation. Export processing zones and industrial zones are being built alongside with efforts to develop services in water and power supply, telecom networks, hospitals, and banks. In December 2006, Thailand and Laos opened the second bridge across the Mekong River, which is their common border, linking Mukdaharn province in northeastern Thailand to Laos' southern province Suvannakhet. On 1 January 2007, the governments of Myanmar and Laos began to charge transit fees for cargo boats, mostly Chinese owned and carrying Chinese agricultural exports to Thailand, using their stretches of the Mekong River. Some environmentalists are alarmed by Chinese projects to improve the navigability of the Mekong River by widening its banks and removing islands and rapids.

The third case illustrates both the complementary and conflicting sides of the transborder region in Southeast Asia known as the Indonesia (Riau)–Malaysia (Johor)–Singapore Growth Triangle (IMS–GT) (see **Figure 1**). The IMS–GT demonstrates a complementary mix of resources among the three units. In contrast to Singapore's high labor and land (very limited) costs, the Johor state of Malaysia just north of Singapore and Riau islands of Indonesia – 20 km away from Singapore by boat – have medium level and low land and labor costs. Transborder economic complementarity in terms of differential factors of production in the IMS–GT has fostered growing internal links, especially between Singapore and Johor and Singapore and Riau.

Yet the IMS–GT has not been without intraregional conflict. Following its ban on the export to Singapore of sea sand used for land reclamation in February 2003, Indonesia in January 2007 banned the export of sand, a basic construction material, to Singapore just as the latter's construction sector rebounded on strong demand for high end apartments and the building of two multi billion dollar casino complexes. The Indonesian government worries that massive sand quarrying from Riau islands could lead to the disappearance of small outlying islands, and as the maritime borders erode, Indonesia's territory and exclusive economic zone could also shrink. Indonesia imposed this ban knowing that it could trigger economic dislocation as thousands of local residents depend on sand quarrying for a living.

Finally, new transborder economic space has opened up in South Asia, extending informal regionalist and de bordering tendencies in the Pacific Rim to its most western edge (see **Figure 1**). On 6 July 2006, India and China (re)opened the Nathu La Pass (4545 m above sea level) between Tibet and the tiny state of Sikkim in northeastern India, 44 years after a brutal frontier war in 1962 shut down the ancient route. Until now the bulk of trade between China and India was transited by sea, and via Tianjin – a port city nearly 4400 km from Lhasa, which is only 1200 km by land from Kolkata (formerly Calcutta), a major Indian port city. Thousands of Indian pilgrims used to make the annual 15 day journey to Tibet's Mount Kailash, revered by Hindus as the home of Lord Shiva. With the opening of the border pass, the pilgrimage will be just a 2 day drive from Nathu La.

While giant warehouses have been constructed on both sides of the Nathu La Pass for its opening ceremony, trade at China's border mart is currently very small, while Indian exports are slightly higher, both far less than expected. However, truckers on the route expect to earn large annual revenues at present freight rates, even if the traffic volume were to be as low as 100 trucks a day. China does not impose any restrictions on cross border trade except for illegal items, and hostels have been built to accommodate Indian traders, but Chinese traders cannot spend the night at the Indian mart. India is concerned about low price Chinese products flooding the Indian market. Other barriers include the lack of infrastructure (poor roads and communications), banking facilities, container depots, and the need to expand the basket of permitted trade items.

Conclusions: Remaking the Pacific Rim

What would the Pacific Rim look like if its entire western half and far western margin that blends into the Indian

Ocean were spatially reconfigured around the eight transborder subregions plus the Greater Shanghai Region (see **Figure 1**) instead of the current nation states? These transborder subregions tempt us to imagine a recon figured Pacific Rim in which the bulk of the population and territory falls into the boundaries of the transborder subregions. The forces that facilitated the rise of the transborder subregions were embedded in a much earlier history and became more visible and active only recently. Yunnan province in China provided the historical must go through pathway on the 'Southern Silk Road' for trade with India and Burma dating back to China's Han dyn asty. Trade through the Nathu La Pass accounted for the lion's share of total cross border trade between China and India in the early 1900s. The disappearance of the geopolitical bipolarity with the end of the Cold War, coupled with a series of resolved border disputes, has activated and augmented these historical economic ties.

While the distinctive geographic position and physical boundaries of the units that make up the transborder subregions are stressed, there are crossover ties and spillover influence between some pairs of these regions. South Korean companies have invested heavily in both the Yanbian region of Jilin province in the Greater Tumen Subregion (GTS) and the coastal region of Shandong province that belongs to the Bohai/Yellow Sea Subregion (BYSS) (see **Figure 1**). The major ports of South Korea and Japan have developed close and fre quent shipping links to Shanghai. The informal bound aries of the transborder subregions are sufficiently porous to allow cross border links to extend or spill. Some powerful cities in these transborder subregions – Hong Kong, Taipei, Singapore, and Seoul – serve the different nodal functions for their respective subregional hinter lands. Since most of the local level units in the trans border subregions are smaller and less influential border cities and towns, other dominant cities can draw multiple links their way, even though they are not located within the boundary of any transborder subregion.

While informal regionalism and de bordering, which have fostered the transborder subregions, are remaking the Pacific Rim, their interface and relationship with formal regional schemes like APEC remain relatively weak and elusive due to their different logics of organ ization and operation. If APEC really opposes the idea of an inward looking trading bloc and instead pursues global free trade, as its leaders have repeatedly em phasized, it would be compatible with the open nature of the transborder subregions. At a lower level, there ap pears to be some connections and mutual influences between the ASEAN/AFTA and the transborder sub regions. To the extent that the Southeast Asian growth triangles are driven by market and private sector dy namics, they are largely consistent with the open and free trade orientation of ASEAN/AFTA. Although the geographic overlap between ASEAN + 3 and the parts of the transborder subregions in the countries involved is substantial (all constituent units of the GMS are included in ASEAN + 3), it remains to be seen if the more formal cooperative activities of ASEAN + 3 will be compatible with the informal cross border trade and investment flows within the transborder subregions.

To sum up, informal regionalism and transborder regions have done more than formal regionalism to alter the basic spatial ground and parameters for rethinking and reimagining the Pacific Rim of the early twenty first century, a Pacific Rim that is also moving west to envelope the rising economic power of India. Seeing the Pacific Rim through the dual lens of informal regionalism and de bordering also brings into clear view the geoeconomic, historical, and sociocultural conditions such as ethnic connections that foster the transborder subregions. This integrated treatment via a broad comparative perspective yields a richer and more nuanced picture of a new Pacific Rim that will continue to evolve and thus challenge our capacity to conceptualize it geographically and otherwise.

See also: Global Production Networks; Industrial Location; Regional Production Networks.

Further Reading

Aikman, D. (1986). *Pacific Rim: Area of Change, Area of Opportunity*. Boston, MA: Little, Brown.

Borthwick, M. (1992). *Pacific Century: The Emergence of Modern Pacific Asia*. Boulder, CO: Westview Press.

Chen, X. M. (2000). Both glue and lubricant: Transnational ethnic social capital as a source of Asia Pacific subregionalism. *Policy Sciences* 33, 269 287.

Chen, X. M. (2005). *As Borders Bend: Transnational Spaces on the Pacific Rim*. Lanham, MD: Rowman & Littlefield.

Dirlik, A. (ed.) (1998). *What Is in a Rim? Critical Perspectives on the Pacific Region Idea*. Lanham, MD: Rowman & Littlefield.

Linder, S. B. (1986). *The Pacific Century: Economic and Political Consequences of Asia Pacific Dynamism*. Stanford, CA: Stanford University Press.

Pempel, T. J. (ed.) (2005). *Remapping East Asia: The Construction of a Region*. Ithaca, NY: Cornell University Press.

Ravenhill, J. (2001). *APEC and the Construction of Pacific Rim Regionalism*. New York: Cambridge University Press.

Segal, G. (1990). *Rethinking the Pacific*. New York: Clarendon Press.

Sparke, M., Sidaway, J. D., Bunnell, B. and Grundy Warr, C. (2004). Triangulating the borderless world: Geographies of power in the Indonesia Malaysia Singapore growth triangle. *Transactions of the Institute of British Geographers* 29, 485 498.

Relevant Websites

http://www.prac.org/
Pacific Rim Advisory Council.

http://www.thegprd.com/index.asp
The Greater Pearl River Delta.

http://www.pacificrim.usfca.edu/
University of San Francisco, Center for the Pacific Rim.

Parenting/Motherhood/Fatherhood

S. C. Aitken, San Diego State University, San Diego, CA, USA, and Norwegian University of Science and Technology, Trondheim, Norway

Glossary

Communitarianism Belief that values exist in public space and community is an important part of moral development.

Factory Acts, The A series of nineteenth- and twentieth-century state legislations in the UK and US that excluded women and children from work in factories and mines.

Family Wage, The The notion of single salary (usually that of a man) that could provide for the economic needs of a family and the basis of much organized labor struggles in the twentieth century.

Neoliberalism Particular forms of economic and political relations that favor self-governance and a free-market ethic applied to all sectors of society.

Nuclear Family, The A functional unit, comprising two adults in a socially acceptable sexual relationship, one of whom (usually the female) stays at home looking after one or more children.

Patriarchy A system of domination that maintains power in men's hands.

Structuration Theory Anthony Giddens' theory that related social structures and individual agency.

Suburban Spatial Entrapment The sequestering of woman and children in suburbs and away from waged labor during the twentieth century.

The work of parenting as an emotional and day to day practice that is embedded in larger social and spatial contexts is a hugely debated topic. The institutions of motherhood and fatherhood are equally debated through contested ideological structures that are often seen as at odds with the work of mothering and fathering. To take one example, feminist critiques of the social science literature on parenting uncover an ideology that suggests the belief that active participation by mothers in the daily care of children is obligatory, whereas nurturing and care taking by fathers is discretionary. Through the externalization of women's roles and their increased paid employment it may be argued that men increasingly escape from so called 'good provider' responsibilities. And so, in an important and problematic way, fatherhood today assumes a voluntary dimension and fathers can either retreat from responsibility or participate more fully in the family. In addition, good enough practices of mothering and fathering are too often aligned to an idealized myth of parenthood within a nuclear family, and this myth is exacerbated as state institutions roll back responsibilities of childcare to communities and families. Assigned this neoliberal burden and caught up in a rhetoric of individuated child raising responsibilities, mothers and fathers exercise complex coping strategies that often end in what is akin to a familial implosion. Of late, geographic discussion engages the globalized contexts of parenting with increased concern for the roles of mothers and fathers, gender power relations, and the legitimization of parental authority and responsibility in increased neoliberalism and cosmopolitanism.

The formal academic study of parenting is a relatively recent endeavor, coming first to conscious attention in the late nineteenth century when rapid social and economic transformation precipitated dramatic familial changes. Some theorists maintain that parenting is a stable union of mothers and fathers as a natural unit of biological and social reproduction that needs no sense of history of geography to uphold and perpetuate its *raison d'etre*. In the early twentieth century, anthropologist Bronislow Malinowski argued that the conjugal family, comprising a mother and a father, was unassailable because it fulfilled universal needs. In the late 1940s, sociologist George Peter Murdock coined the term 'nuclear family' as a functional rather than a biological unit, which comprised two adults in a socially acceptable sexual relationship, one of whom (usually the female) stayed at home looking after one or more children. He later conceded that the configuration of the conjugal family varied greatly, being most often embedded and extended in a larger domestic unit or community. Marxist theoretical perspectives rejected biological determinism and eschewed Malinowskian functionalism in favor of fathers' and mothers' struggles over the means of subsistence and production. These perspectives speak to 'family adjustments' and 'family strategies' for coping within a capitalist waged labor system.

Academic discussion on parenting began in geography in the 1980s with feminist concern to redress certain research imbalances. Feminist geographers argued that focusing on the optimal location of firms and the rank size rules of cities, for example, was favored in geographic research over any serious work on women's issues such as childcare, school access and affordability, and parenting. Since that time there has been an outpouring of work on mothering in particular and, to a lesser extent, fathering. This work may be categorized broadly into theoretical work that frames spaces of parenting and ideologies of parenthood, and empirical and ethnographic studies that focus on the work of parenting and into the construction of mothering and fathering identities.

Of late, there is growing concern for the burden placed on parents with legislation favoring state rollback of welfare responsibilities to communities and families. Within these categories, specific spatial topics such as the access and affordability of childcare, schooling, trans portation, and the ways parents juggle paid labor, the emotional work of parenting, and domestic responsi bilities are of paramount importance. Diversity concerns highlight different styles of parenting amidst hegemonic representations of parenthood that problematically raise normative Anglo American, heterosexual white, middle class representations. Other issues relate to the ways parents strategize their productive and reproductive activities within spatial systems (institutions, neighbor hoods, and cities), and how these strategies play out in an age of neoliberal globalization.

Parenting as a 'Woman's Issue'

In a 1974 *Antipode* article that presaged the feminist call for geographers to include studies of families and women's work, Allison Hayford notes that women were functionally and emotionally central to the establishment of kinship systems in precapitalist times and with the emergence of capitalism it became important for them to be subjugated to parenting roles and control within the domestic sphere. Susan MacKenzie and Damaris Rose spatialize Hayford's ideas with a critical appraisal of the domestic economy and home life through the industrial revolution. They outline the historical origins of the complex relations among social production, the circu lation of commodities, and the reproduction of labor as they relate to spatial changes in parenting practices and family life. MacKenzie and Rose were the first to probe the importance of the spatial separation of the public sphere of production (work) and the private sphere of reproduction (home) as a series of struggles rather than a *fait accompli* of the emerging capitalist system. Indus trialization threatened to destroy the family with ex panded work hours for men, women, and children engaged in waged employment because increased time spent in the productive sector left almost no time for the domestic sphere. Long work hours and low wages in combination with cramped and unhealthy living spaces resulted in the creation of a separate domestic sphere where people pooled their wages to maintain themselves and where they lived. The household ceased to be the center of production although it remained the center of family life. Discontent with the lack of concern by factory and mine owners for the health and housing of families led in part to struggles for a separate domestic sphere. Philanthropists and local governments began to intervene on behalf of the family to provide education and medical care. Through a series of Factory Acts first in the UK and USA and then copied elsewhere, women and children were excluded from factories for the sake of their health, and their place was squarely recognized as within the 'safe haven' of the domestic sphere. Later, organized labor unions agitated for the 'family wage', based in part on a desire to keep married women in the home so that they might take care of the male wage earners' children.

The separation of home and work resulted also in the separation of fathers from their children, at least during working hours. From the nineteenth century onward another geographic solution to industrial urban squalor – the suburb – further separated fathers from children and spatially entrapped women with children. Ironically, contemporaneous reformers saw the city, and particularly working class neighborhoods, as reflecting and reinforcing the erosion of the family. Although the creation of suburbs in the nineteenth and twentieth century in dustrializing nations has a complex history they are, like family wages, a large part of the construction of the modern nuclear family norm.

The spatial entrapment of mothers thesis was de veloped by geographers in the mid 1980s through analyses of large data sets from North American and European urban populations. With the discovery that most women's work trips were shorter than those of men, researchers concluded that peripheral suburbs were built for a commuting husband with mothers fettered primarily to the domestic sphere and raising children. Studies in the United States, Canada, Sweden, and France demonstrated that mothering had a major impact on women's travel. A celebrated 1993 article on 'Sub urban pink collar ghettos' in the *Annals of the Association of American Geographers* by Kim England suggests more complexity to the spatial entrapment thesis. England found that mothers in Columbus, Ohio who would normally be thought of as spatially entrapped actually traveled further than never married women without children. England concludes that spatial entrapment theory is contingent upon historical patterns of the 1960s and 1970s and does not reflect a move toward neoliberal economics of the 1980s and 1990s that often require full time employment of fathers and mothers. Her study utilized both quantitative and qualitative methods, the latter of which were becoming increasingly important for place based studies of mothering and fathering identities.

In the 1990s, geographers argued that local places often provided meaningful opportunities for mothers to express relationships with the public sphere. Isabel Dyck used ethnographic interviews in conjunction with Anthony Giddens' structuration theory to trace the space–time connections of suburban mothers in Van couver. With a concern for women's agency, she shows that the notion of motherhood and the sets of practices making up mothering are interpreted, negotiated, and transformed as women respond to the social and eco nomic structure of particular places. She shows ways that women's interpretation and management of mother–child relations are linked to local contexts, as well as larger structural forces, arguing that women are not passive agents with regard to the place based conditions that contextualize their mothering, but rather they "modify these conditions and negotiate 'good' mothering practices though the recurrent practices of their everyday lives." In the preface of a book introducing a wide ranging series of geographic studies on North American working mothers, Kim England argues that the study of "childcare across a variety of scales is important because disparities and inequities often become more apparent at finer scales" (England, 1996: 4).

Sarah Holloway extends the issue of webs of meaning and childcare solutions into local cultures of mothering and their relations to spatial justice in a 1998 article published in *Gender, Place and Culture*. She uses the ex ample of British local education authority playgroups to suggest that locally embedded childcare cultures shape mothers' attitudes toward their children's educational development and their strategies for accessing childcare. Also negotiated through these local contexts are moral geographies of mothering that relate to the social con struction of what it is to be a good mother. Territorial justice is concerned with the public sphere (i.e., the delivery of childcare as conceptualized in spatial terms) whereas childcare cultures are concerned with inter twining public and private spheres with a contextual (i.e., local) understanding of need. Put another way, it seems that – at least in the British middle class context – homes and neighborhoods remain a major locus of reproductive activities. Arguments of this kind paved the way for neoliberal communitarianism such as Anthony Giddens' *The third way*, which proved hugely influential in British rollback millennium politics.

The Awkward Spaces of Fathering

What emerged through the twentieth century, and con tinues today, is a division of gender roles that was pre carious from the start because it gave fathers a choice that is apparently not afforded to mothers. The last two decades' interest in fathering has been fueled to a large degree by increased public employment for women and by unprecedented demographic changes in Western families. That said, elsewhere in the world, and in the context of low income families in the West, women have always engaged hugely in productive activities. The assumed role of men as breadwinners and guardians of families as a traditional key denominator for male identity is increasingly challenged. Questions that arise from these changes relate to family politics and changes in authority, the roles of fathers in the lives of infants and young children, and the performance of fathering identities.

A review of historical trends in Western notions of fatherhood suggests that 'the father' has evolved from the distant breadwinner of the nineteenth century, through the genial dad and sex role model of most of the twen tieth century to today's father as equal co parent. How ever, the practice of fathering remains vague and inscrutable. As the role of women changes in Western society, leaving new spaces for male identities, some geographers argue that younger generations of men do not easily fit into 'traditional' categories of father and breadwinner. At the same time, the creation of new social norms and values has not automatically led to an aban donment of old patriarchal norms, thus representing significant fields of conflict for the development of fathering identities. Much of 'crisis of masculinity' lit erature in the 1990s and contemporary critical assess ments of masculinities has no sustained discussion of fathering but focuses instead on public arenas such as employment, the workplace, and politics. As a con sequence of this academic and political neglect, father hood is regarded as a backwater to the dominant discourses of men in public power.

Contemporary wisdom suggests that the spatial entrapment of mothers and children in suburbia is con trasted with an ever more distant father who maintained control of productive, public space. With the rise of neoliberal processes of globalization, a paradox arises because this emotional separation and individuation through the late nineteenth and into the twentieth cen tury does not reflect day to day life in the same way that it buttresses society's mythic ideals of social repro duction. There are some interesting consequences to this inconsistency. As mothers enter the paid labor force and some decide not to become mothers, they are blamed either for maternal depravation or they are chastised for following masculinist roles. In critical indictments of research on fathers' roles, some feminists suggest that social science research is constructing fatherhood as a panacea for the problems mothers create and cannot control. One example is child development re search, which often assumes that a father's presence is almost always a good thing. Feminists also point to 'in stitutionalized father love' that serves to establish fathers' claims to authority over children whether or not they have actually participated in the work of caring for them. Mothers, on the other hand, are left not only with the emotions but also the physical labor that has been historically placed on their shoulders.

It seems that fathering as a practice and fatherhood as an institution are difficult to define without recourse to mothering and motherhood. The Western media is replete with stories of caring, nurturing, domesticated dads who are fighting the strictures of the workplace to spend the same amount of time with their children as mothers, and a burgeoning social policy literature is focusing on men's co parenting skills. However, in a study of a Lone Fathers Association in Newcastle, Australia, Hilary Winchester notes a reification of hegemonic forms of masculinity among members to the extent that her interviewees' views were "profoundly misogynistic." The most significant issues for these fathers concerned access to, and custody arrangements and maintenance payments for, their children. Most felt that the courts and their ex wives treated them unfairly. Winchester's interviews raised important questions about inequitable access to legal aid and the vulnerability of all separated fathers to unsubstantiated allegations of child physical and sexual abuse. Such contexts, real or alleged, are an important part of the hidden spaces of fathering. Winchester points out further that the fathers in her study were frustrated by legal, institutional, and familial structures predicated upon some very traditional ideas of fathers as the breadwinner of the family. Empirical focus in recent years has moved away from Winchester's lone and disgruntled fathers to the extent that different perspectives on fathering are embraced. These studies demonstrate how fathering identities are contingent on place and time.

Although a growing literature espousing the value of the domesticated father, a powerful patriarchal 'idea' continues to influence the work and spaces of fathering. Research on 'the new father' tends to be uncritical and positive in its analysis of the contemporary culture of fatherhood while missing what is implied by the work of fathering, with all its emotional trappings. Ironically, a monolithic concept of 'the new patently domesticated father' was originally justified and supported by a femi nist ideology that advocated men's involvement in par enting. The new, domesticated 'father ideal' sidesteps the point that insidious patriarchal hegemony forces may create more obstacles than bridges for fathers. The image of the new father also assumes homogeneity in fathering roles and situational contexts that deny the importance of complex forms of fathering practiced differently in dif ferent spatial contexts.

Powerful Emotions

Family relations cannot exist without parental authority, and some argue that these relations remain the best for enabling children to develop a capacity as adults to give allegiance to the principles and presumptions of a demographic society. Recent geographic research looks at how family authority is constituted and how gendered authority relates to raising children. By identifying discourse forms, as well as routines and practices in which family members engage on a regular basis, it is possible to understand how discourses constitute and organize social and power relations in families. Families are political in the sense that the activities and predi lections of members often are superimposed and/or sanctioned by subordinate members. However, those predilections are constantly changing. As political sites, through actions and practices, families are continually engaged in the reproduction of discourses, meanings, and subjectivities. It may be argued that because of the sup posed hegemony afforded to individual families in private life, there is greater potential for altering hegemonic discursive fields. This notwithstanding, deviations from accepted gender performances are transgressions which pose a threat to the continued hegemony of normative gender norms. The private relational nature of power in families offers opportunities for resistance, negotiation, and change for mothers, fathers, and children.

The emotional work of parenting focuses upon the value of parenting as labor that is emotionally charged as opposed to biologically determined. The 'biological facts' of motherhood and fatherhood are not 'given' but come into being as science progresses and identities assume cultural significance through political struggle. If we em brace feminism as a powerful de naturalizing force then we must conclude that gender is a socially imposed division of the sexes, and this includes women's relations to the birth process. Proponents of greater rights for fathers argue that no fact – such as the ability to bear and nurse children – entails or excludes a moral right or commitment.

All relationships are emotionally charged, and the emotional spaces between mothers and fathers are no exception. That said, familial spaces are forged over extended periods of time and they are gelled with com mitment and obligation. The biological connections between parents and their children are not just about sperm, gestation, labor, birth, transferred DNA, and nursing. Each of these facts is loaded with centuries of cultural evolution co mingled with biological evolution. The fact of motherhood is not just about the reality of giving birth, it is also about the labor that goes into making connections and appropriating the fetus and baby into the mother's moral and emotional space. Without diminishing the importance of the embodied connection between mother and child, there is clearly an emotional geography at work here for fathers and mothers that transcends, transforms, and reifies physical bonds.

Neo-liberal Processes of Globalization and Geographies of Care

Some important recent geographic work on parenting focuses on neoliberal processes of globalization. In large part, discussion critiques the rollback of state monies for children's welfare, with increased pressure on com munities and families to provide support. Much of this work looks at the impact of global economic restructuring on local family contexts. For example, in a 2001 article in *Antipode*, Elsbeth Robson shows how the AIDS crisis in Zimbabwe has resulted in children leaving school and caring for their ailing parents. In a 2006 article in *Area*, Joel Jennings and his colleagues point out that children in Tijuana as young as 12 can earn more than their parents. The notion of children as primary bread winners and caregivers in their families opens up inter esting notions of what precisely constitutes parenting.

In a recent study of Nicaraguan mothers, Julie Cupples finds new forms of mothering in a postrevolu tionary state that resist and transform the neoliberal present. State transitions to neoliberalism in countries in the majority global South create spaces in which nor mative understandings of femininity, masculinity, and parenting are renegotiated in ways not seen elsewhere. She argues that the nexus of work, single mothers, and men de mobilized from the recent revolutionary conflict in Nicaragua creates a context through which women take on both the emotional care and the economic care of children (and other dependents such as grandparents). The idea that marriage and motherhood means an end to paid work is identified as a discourse but rarely sub scribed to by some mothers in Nicaragua as is the idea that biological mothers should raise their children. Other studies in Latin America and Asia document numbers of poor mothers who leave their children for months and sometimes years to migrate to places where waged labor enables them earning power to send home family remittances.

With the hardship fomented through the excesses of neoliberal economic restructuring – reductions in ser vices and welfare provision, the failing of certain sectors of employment, and increased burdens on low income parents – studies in Africa, Asia, and Latin America suggest that earlier geographic studies of changes in parenting during Western industrialization tended to rely too much on stereotypes of motherhood and fatherhood, and concluded too readily that women and children were subjugated and marginalized. This past work focuses too narrowly on exploitation and fails to adequately theorize the noneconomic work, emotional behaviors, and iden tity restructurings of men, women, and children. A recent shift is away from a political economy perspective to a post structuralist perspective that emphasizes multiply constituted forms of mothering, fathering, and being a child/grandparent and engages critically hegemonic forms of motherhood, fatherhood, and domesticity. This, of course, is the contested and problematic 'illness and remedy' that constitutes contemporary neoliberal glob alization. Economies and systems of governing and edu cation are not the only things to be restructured under neoliberalism: motherhood and fatherhood undergo forms of restructuring of their own that unleash a number of complex, contradictory, and unexpected twists. Glob alizing processes are facilitating at the local level by new gender relations and identities, and new intersections between the work of mothering and fathering.

In her 1999 book *Globalizing care*, Fiona Robinson argues that when geographies of care do not respond well to distance they become utterly territorial, and that we need to move to a notion of relational responsibility that is about difference rather than exclusion, that is embodied, and that is not restricted to the very local. This strategy recognizes that the global is wholly implicated in the local work of mothering and fathering because parental identities are forged through embodied relations that are extended geographically, as well as historically.

See also: Care/Caregiving; Community; Feminism/ Feminist Geography; Neoliberalism; Patriarchy.

Further Reading

Aitken, S. C. (1998). *Family Fantasies and Community Space*. New Jersey: Rutgers University Press.

Aitken, S. C. (inpress). *The Awkward Spaces of Fathering*. Aldershot: Ashgate Press Forthcoming.

Cupples, J. (2005). Love and money in an age of neoliberalism: Gender, work, and single motherhood in postrevolutionary Nicaragua. *Environment and Planning A* 37, 305 322.

Dyck, I. (1990). Space, time and renegotiating motherhood: An exploration of the domestic workplace. *Society and Space* 8, 459 483.

England, K. (1993). Suburban pink colar ghettos The spatial entrapment of women? *Annals of the Association of American Geographers* 83(2), 225 242.

England, K. (1996). *Who Will Mind the Baby? Geographies of Childcare and Working Mothers*. London: Routledge.

Mackenzie, S. and Rose, D. (1983). Industrial change, the domestic economy and home life. In Anderson, J., Duncan, S. & Hudson, R. (eds.) *Redundant Spaces in Cities and Regions: Studies in Industrial Decline and Social Change*, pp 155 200. London: Academic Press.

Robinson, F. (1999). *Globalizing care*. Boulder, CO: Westview Press.

Winchester, H. (1999). Interviews and questionnaires as mixed methods in population geography: the case of the lone fathers in Newcastle, Australia. *The Professional Geographer* 51(1), 60 68.

Participant Observation

K. Walsh, University of Sussex, Brighton, UK

Glossary

Covert Participant Observation The practice of conducting participant observation without informing the researched of the researcher's identity or purpose.
Field The space in which research is carried out.
Field Notes Data recorded by the participant observer, often in the form of a field diary.
Gatekeeper The person through whom access to a particular group or space is negotiated.
Going Native The act of becoming involved in the community of participation to such an extent that observation becomes problematic.
Insider The position, not fully possible or desirable, of becoming one of the group being researched.
Positionality The subject position of the researcher defined through social categorizations, for example, gender or class, personal beliefs, experiences, and attitudes.

Introduction

Participant observation is a research method originally associated with anthropological field work, but increasingly used by human geographers, usually as part of an ethnographic approach. The terms ethnography and participant observation are often used synonymously but, while they are closely related in research methodology, they do have distinct meanings. Participant observation is a key method used by ethnographers, but ethnography refers to a more holistic research approach which might also include, for example, quantitative methods or interviewing techniques. Nevertheless, when human geographers talk about the strength of ethnographic research methods, they are usually implicitly, and sometimes explicitly, arguing that it is the addition of participant observation that has changed their understanding of a particular social or cultural phenomenon. This is because the term refers to a set of research practices thought to generate unusually in depth material.

In participant observation, the researcher attempts to learn about a particular socio cultural space and those who inhabit it by taking part and continually reflecting on what is happening. Unlike other qualitative methods, participant observation involves examining what people do, rather than only what they say or what they say they do. However, in contrast to simply observing, the active participation by the researcher is thought to maintain the naturalness of the research encounter and limit the disruption otherwise generated by their presence in the field. Participant observation is, therefore, different from 'pure observation', a method in which researchers try to exclude themselves as far as possible from the context of their research by using video cameras or tape recorders to collect date for later analysis. Likewise, participant observation is different to 'full participation'. This might occur accidentally during participant observation, but is considered undesirable as it refers to the loss of the identity of investigator and, by implication, a reduction in analytic possibility.

Participant observation is further differentiated from other research methods by the sensual, emotional, and embodied experience of the researcher being incorporated as a significant part of the research. The participant observer is figured as a fully 'human' researcher, rather than a detached and neutral scientist. This has led to recognition of the inevitably 'messy' nature of fieldwork. Accounts by human geographers of their experiences in the field resonate with those of anthropologists who emphasize the ongoing and subjective process of coming to understand social and cultural practices. Rather than seeing this as problematic, however, those conducting participant observation have worked hard to acknowledge the implications of their field corporeality through practices of critical reflexivity.

In traditional anthropological fieldwork, the boundary between home and field was thought to be clearly defined. The discipline established itself through fieldwork on the tribal so the field was always situated away from home, usually in another country. However, influenced by postcolonial theory and the increasing recognition of the power laden dynamics of fieldwork, the focus of some researchers began to alter. Participant observation is now just as likely to be carried out 'at home', within the society of the researcher, as it is elsewhere. So, the binary oppositions previously set up between home and field, here and there, have been complicated and sometimes collapsed. This is certainly the case for human geographers who have increasingly adopted participant observation to investigate subcultural groups and spaces within their own society or argued that they have felt a sense of belonging in the field.

Despite its central significance in anthropological fieldwork, participant observation has traditionally been a peripheral methodology in human geographical enquiry. The increasing importance of participant observation in contemporary research by human geographers can be linked to attempts to research everyday life and practice. Indeed, the influence of nonrepresentational theory on human geography has been enormously significant. Human geographers who have generated nonrepresenta tional theory argue that social performances are never limited to words, but also lived through corporeal and emotional experience. This argument has had an import ant impact on both the topics and methods of analysis being explored by contemporary human geographers. Within anthropology, participant observation has long been celebrated as a method which incorporates social and cultural meanings beyond those which are verbally expressed and this is also its appeal to many of the con temporary human geographers who have since adopted it as a central method in their research. Participant obser vation is considered a useful method to go beyond dis cursive and representational geographies.

The use of participant observation by human geog raphers has also increased with a wider recognition of the importance of qualitative methods within the discipline. Arguably, human geographers developed increasingly sophisticated qualitative research skills generating, and responding to, the social and cultural turns of the 1980s and 1990s. Since the mid 1990s, many human geographers have embraced participant observation, using it to investigate a wide range of social and cultural spaces, practices, experiences, and identities. For in stance, human geographers have used participant obser vation to analyze geographies of illness and disability, migration, home, work, activism, tourism, consumption, human–animal relationships, leisure practices, and café culture. Some of this research will be drawn upon to illustrate the remaining sections of this article. The initial focus is on 'doing' participant observation and the methodological debates surrounding participant obser vation as a research practice are then discussed.

Doing Participant Observation

In the 1970s, how to handbooks on participant obser vation were virtually absent and students in anthro pology, about to embark on their first research visits, were directed instead to classic accounts of life in the field, such as those of Bronislaw Malinowski, Margaret Mead, and Claude Levi Strauss. Fortunately, since then, an thropologists have begun to address this absence and human geographers have followed suit. Nevertheless, many still infer that doing participant observation is a matter of innate personal skills rather than something that can be learnt. This might well be related to the difficulty of anticipating all eventualities in the field, something that is thought to necessitate the flexibility and resourcefulness of the researcher. Also, the centrality of the researcher as a participant demands particular social skills, such as the ability to 'fit in' with people who may be perceived as belonging to a different (sub)culture. Yet, in accounts which argue that participant observation is a research skill, certain key practices are identified repeatedly. These include negotiating access, adopting covert or overt strategies, and writing field notes, each discussed here in turn.

Negotiating Access

It is generally accepted that participant observation has the potential to be used in the majority of field settings, but the ease of gaining and maintaining access will vary considerably. Accounts of geographical fieldwork suggest that negotiating access might be fraught with difficulties or surprisingly easy, but many emphasize the necessity of being flexible and persistent. Often the researcher will need to gain access to a particular site or event. This might be relatively straightforward if the research in volves public spaces. Examples here would include cafes, shops and malls, parks, street festivals, sports stadiums, clubs, and bars, although hanging out for a long time in such spaces might still attract attention. However, when it is private spaces that need to be accessed, the researcher will often seek permission or introduction through a 'gatekeeper'. Examples here would include research in the home, as well as workplaces, hospitals, social groups, schools, and many other organizations. A further possi bility is that the researcher will work with particular individuals, accompanying them as they go about their daily routines in a range of public and private spaces (access being gained to the latter through invitation). This requires recruitment of key respondents willing to be shadowed or share their lives with the researcher on a temporary basis. The difficulties of gaining access have led to many human geographers choosing to conduct participant observation in places and with groups or in dividuals they already had access to prior to the start of the project.

Covert Strategies

When access is very difficult or the researcher's presence is expected to be particularly disruptive or controversial, covert participation observation is sometimes used. In covert participant observation, the researcher does not inform the researched of their motives for being in the field but instead presents themselves as just an other participant. Among those human geographers who practice covert participant observation, a diversity of approaches is evident and these can be understood as tactical responses to variations between field spaces. For example, some researchers will be putting themselves in situations where, if found out, they might be at risk of physical harm or legislative retribution. They may, therefore, feel forced to lie in order to keep their position secret and complete the research at all. Other researchers are able to be completely open about their position as a researcher, but feel the need to hide the focus of their analysis from all or some of the researched. Meanwhile, there are other researchers who initially set out to research one theme, but find themselves compelled to respond dynamically to the material they are collecting as a result of their presence. Deciding whether to reveal or conceal their purpose and identity, raises both practical and ethical questions for the researcher. While covert participant observation may be regarded by some as dishonest, many researchers argue it is a compromise they must make to achieve access to socio cultural spaces that they would otherwise be denied. It is often considered necessary by researchers who require access to more powerful individuals and institutions responsible for the social marginalization or exclusion of minority groups. A political commitment to the well being of the less powerful, something that guides many researchers inter ested in social justice issues, therefore results in ongoing ethical debates, but covert participant observation has enabled many human geographers to produce sensitive and/or critical geographies of the world we live in.

The work of the feminist geographer Hester Parr is useful as an example through which to explore the ethical issues surrounding covert participant observation strategies. In her research on people with mental health problems, Parr adopts both overt and covert strategies in a range of spatial settings. To access hospitals and homeless hostels, Parr negotiated official access to con duct her research. However, in the public spaces of streets and parks she used covert participant observation to make contact with people with mental health problems who avoided institutional spaces. Furthermore, Parr deliberately used her body in similar actions to those people she was observing (e.g., smoking, foot tapping, and adopting certain postures) in an effort to make herself unthreatening and, therefore, to enable her to participate in verbal conversation and 'body talk'. Such corporeal presentation, as Parr herself recognizes, raises significant ethical issues, particularly in research on vulnerable in dividuals and groups. Yet, in this case, covert participant observation was also an undeniably necessary part of producing empathetic and useful knowledge about the everyday experiences of people living with mental health problems. Many more accounts of covert participant observation by human geographers could also be used to illustrate how the ethical issues surrounding such strat egies are rarely straightforward.

Writing Field Notes

The issue of how to write field notes has been relatively neglected in the methodological literature on participant observation, in stark contrast to the noticeable concern across disciplines with the politics and poetics of repre senting ethnographic material at the 'writing up' stage. Yet, field notes are a distinguishing feature of participant observation and, as the experiential and observational material (or data) that the researcher will continually analyze to explore their research questions, they are an extremely significant part of the research process. One concern that might explain this reticence is that field notes in their 'raw' form can be perceived as too messy; they might be full of emotional response, incomplete analysis, unresolved questions, interpretations that are not yet fully thought through, or even admissions of mistakes. By the time field notes are to be presented or published, they are reordered and reworded, their significance selected and revalued for an audience.

A further problem is that there is seemingly no con sensus regarding how exactly to write field notes. Cer tainly, there are no published examples of raw field notes against which to compare first attempts. However, some human geographers have used extensive extracts from their field notes to illustrate their research which, although undoubtedly polished for publication, offer clues about what sort of information might be recorded and reassurance that a personal style is inevitable. Others provide lists of things that might be included, such as: reflexive accounts of access issues, power relations, emotional experiences, as well as more pragmatic logging of who you talk to, where you go, and when. The matter of when to write field notes is also contentious: some argue that memory distorts field notes written at the end of the day, while others claim that a notebook can be disruptive as it reminds the researched of your motives for being there.

Methodological Issues and Debates

A Scientific Method?

A persistent critique directed at human geographical enquiry using participant observation is that such re search is considered inherently unscientific. This claim is based on questions about the reliability of the material gathered, as a result of the high significance placed on interpretation in participant observation techniques and the significance of the researcher themselves in the field. A further concern raised by participant observation cen ters on the difficulty of generalizing from empirical re search based on small sample sizes. Sustained interaction with a particular group or site is time consuming and, therefore, limits the possibilities of including as many research participants as surveys, or even interviewing methods, routinely managed. This raises questions about the representativeness of the sample and the applicability of the interpretations to other sites or groups.

While such critiques are recognized as serious concerns for the human geographer, many have argued that the taken for granted scientific objectivity of knowledge is actually socially constructed and, therefore, must be challenged whichever research methods are being used, because all research involves interpretation. Likewise, it is argued that the problems related to small sample sizes in participant observation are far outweighed by the in depth nature of such research and that empirical findings can often be generalized when the researcher chooses appropriate respondents and/or sites of analysis. Furthermore, for many, the researcher's subjectivity is an advantage in participant observation: it is their emotional and embodied experience of learning that offers insight into the social and cultural meanings they set out to explore. These arguments are so widely accepted in contemporary human geography that explicit response to such critiques is now deemed for the main part unnecessary (although it is important to note that many students on degree programs in human geography still come across resistance to their desire to use qualitative research methods of any kind in projects and theses).

Critical Reflexivity

For human geographers using participant observation, critical reflexivity is often a hugely significant research practice. In the work of feminist geographers, the term refers to the practice of continual critical introspection on the researcher's fieldwork practices and relationships, particularly with respect to the power involved. In adopting the technique of participant observation, the researcher must be sensitive to their centrality in the research process and consider the implications carefully. The researcher is often figured in participant observation as the 'vehicle' or 'tool' of the research method and many human geographers argue that with this position comes a heightened responsibility to engage with questions raised by their own positionality. The multiple subject positions of the researcher will impact on the research in ways that can never be fully traced. Factors such as gender, class, ethnicity, age, sexuality, education, and the personal beliefs and perceptions of the researcher will influence every stage of the research, including the research questions they choose, their interactions with the researched, and their interpretations of the material they collect. No research is without such dilemmas, but the close personal relationships over a long duration which are involved in participant observation potentially increase the importance of critical reflexivity.

Ethical arguments for using critical reflexivity are also heightened in research on marginalized or vulnerable groups. This is most obviously a concern when those associated with institutions in the West travel to conduct their research 'elsewhere' in postcolonial settings and development contexts. However, concerns about power relations do not disappear for human geographers conducting research in their own society, especially when they are researching across generational, class, and ethnic divisions, or examining geographies of illness and disability.

Insider/Outsider

Increasingly, the false binary of insider/outsider has been deconstructed in accounts of participant observation, including those written by human geographers. Initially, it was argued that it might be more appropriate to consider a range of possible positions along a continuum from insider to outsider. More recently, however, fieldwork accounts have been produced that recognize multiple shifts between various positions, or the possibility of holding several positions simultaneously. Nevertheless, anxieties persist about the potential susceptibility of the participant observer 'going native'. This is a term which refers to the researcher becoming too involved in the space being studied, by accident or by choice. This is considered by many human geographers to limit the ability of the researcher to distance themselves from the events and people around him/herself and, therefore, to maintain a sufficiently critical analysis.

However, in accounts of the researcher as activist, claims to remain detached have been superseded by an acknowledgment of the researcher's various identities and by efforts to be transparent about their effects on the research. Indeed, some human geographers have been actively working to disrupt the separation of the academy from something that lies beyond. Participatory approaches to research that are activist led, challenge these boundaries by positioning the academic as a facilitator and the role of the participants as co researcher or co activist. In this way, participants are thought to become more empowered than they might through more traditional approaches and the researcher becomes an active part of the community, albeit in an unusual position. For many, this has become a way of doing human geographical research which offers the potential to increase the relevance of their working lives and make a difference to the world outside their university.

Such ongoing debates about being an insider/outsider are also highlighted by some specific contexts in which participant observation might be carried out. For instance, the process of researching sexuality is thought to raise particularly challenging dilemmas for human geographers using participant observation and qualitative methods more generally. Feminist geographers Alison Bain and Catherine Nash, for example, have conducted participant observation at a queer bathhouse event. For them, as for many other human geographers, it is not enough to simply state their positionality through social labels, for instance, lesbian women. Instead, Bain and Nash question what it means to place their lesbian bodies and identities into the sexualized space of the field, detailing their experiences of conducting participant observation (dressing, positioning, and interacting with their bodies) in order to interrogate the partiality of their knowledge construction. In doing so, they explore their position as both insiders and outsiders and the multiple ways in which they negotiate this. For instance, they chose to dress in streetwear to communicate that they were not available to participate in sexual activities, but acknowledge their simultaneous inclusion as desiring bodies in voyeuristic practices.

Field Relationships

Participant observation demands the establishment and ongoing negotiation of social relationships for eliciting findings, and the quality of these relationships is understood to be critical to the success of the research. At the same time, the researcher is often away from home and/or working with their informants over a relatively long time period. This kind of immersion in the field site generates particular ethical dilemmas for the participant observer around notions of intimacy, while also potentially exaggerating many of the ethical dilemmas facing qualitative researchers more generally.

Intimacy is central to participant observation, evoked in the language of rapport, trust, and friendship. The participant observer is advised to get close to the researched in order to become accepted and unobtrusive. This will usually involve distinct, deliberate efforts on the part of the participant observer and sometimes 'learning to act' in certain ways to fit in. Many human geographers discussing their field experiences have commented, therefore, on the difficulties of negotiating such significant, yet unusually strategic, relationships. On the one hand, such relationships can feel deceptive; this is often the case when the participant observer considers that they would not normally associate with the researched due to their different cultural and political practices or moral principles. On the other hand, the participant observer may feel that their relationship with the researched goes beyond the research project itself. Indeed, through long term participant observation they may begin to feel as part of their lives. When relationships feel more 'natural' or genuine, however, ethical dilemmas do not simply disappear. In each case, the researcher draws boundaries around what is research and what is not, and must decide if the researched is to have any input in this process.

Crucially, however, participant observation does not just rely on being able to establish and maintain relationships, but also to detach from them. While the researcher is called upon to provide evidence of their presence, they must also demonstrate their critical distance. Detachment is necessary both in the field, during daily observations and writing of field notes, and when leaving the field at the end of a period of fieldwork. For this reason, integration must never be complete and relationships (except where fieldwork is done nearby) temporary, or at least they must be adapted to virtual or transnational connections. For the researcher who has lived fully immersed in the field, therefore, the development of field intimacies often raises further questions about the emotional geographies of home and belonging. This tension between the personal and professional identities of the participant observer is evident in many accounts written by human geographers.

See also: Ethnography; Fieldwork.

Further Reading

Amit, V. (ed.) (2000). *Constructing the Field: Ethnographic Fieldwork in the Contemporary World*. London: Routledge.

Bain, A. and Nash, C. (2006). Undressing the researcher: Feminism, embodiment and sexuality at a queer bathhouse event. *Area* 38, 99 106.

Bell, D., Caplan, P. and Karim, W. (eds.) (1993). *Gendered Fields: Women, Men and Ethnography*. London: Routledge.

Carter, J. L. and Hill, G. J. (2007). Critiquing environmental management in indigenous Australia: Two case studies. *Area* 39, 43 54.

Clifford, J. and Marcus, G. E. (eds.) (1986). *Writing Culture: The Poetics and Politics of Ethnography*. Los Angeles and Berkeley: University of California Press.

Cook, I. (1997). Participant observation. In Flowerdew, R. & Martin, D. (eds.) *Methods in Human Geography*, pp 127 149. Prentice Hall: London.

Crang, P. (1994). It's showtime: On the workplace geographies of display in a restaurant in southeast England. *Environment and Planning D: Society and Space* 12, 675 704.

Dewalt, K. M. and Dewalt, B. R. (2002). *Participant Observation: A Guide for Fieldworkers*. Oxford: Rowman and Littlefield Publishers.

Emerson, R. M., Fretz, R. I. and Shaw, L. L. (1995). *Writing Ethnographic Fieldnotes*. London: The University of Chicago Press.

Fuller, D. (1999). Part of the action, or 'going native'? Learning to cope with the 'politics of integration'. *Area* 31, 221 227.

Fuller, D. and Kitchin, R. (2004). *Radical Theory/Critical Praxis: Making a Difference beyond the Academy?* Kelowna, BC: Praxis E Press.

Herbert, S. (2000). For ethnography. *Progress in Human Geography* 24, 550 568.

Parr, H. (2001). Feeling, reading and making bodies in space. *The Geographical Review* 91, 158 167.

Walsh, K. (2007). It got very debauched, very Dubai! Heterosexual intimacy amongst single British expatriates. *Social and Cultural Geography* 8, 507 533.

Wilton, R. (2000). 'Sometimes it's OK to be a spy': Ethics and politics in geographies of disability. *Ethics, Place and Environment* 3, 90 96.

Participation

S. Hickey and U. Kothari, University of Manchester, Manchester, UK

Glossary

Stakeholder An individual or group that is affected by or impacts upon an action or intervention.

What Is Participation in Development?

In development discourse, the term participation is generally used to describe the type and level of stakeholder or beneficiary involvement in development planning, projects, and practices. The ostensible aim of participatory approaches is to make 'people' central to development by encouraging the involvement of socially and economically marginalized people to influence the policies and practices that affect them. In this way, it is expected that development processes will become more relevant, efficient, and sustainable, and that unequal power relations will be challenged in ways that empower previously subordinate actors. Originally promoted as a strategy for 'alternative development', the 1980s and 1990s in particular saw the widespread mainstreaming of participatory approaches within the aid industry. As a result, participatory methodologies – often referred to as Participatory Rural Appraisal (PRA) or Participatory Learning and Action (PLA) – have become commonplace across virtually all sectors of development. Importantly, they have also been scaled up from the localized level of project intervention to form an ideological and instrumental cornerstone of contemporary development policy, as illustrated by the World Bank studies on the 'Voices of The Poor' and its promotion of participation within flagship policy initiatives such as the 'poverty reduction strategy' process. We focus here on the implications of this trajectory, identifying some of the problems associated with participatory approaches and recent efforts to deal with these, and draw attention to the broader ways in which 'participation' has been associated with development.

The Emergence and Benefits of Participation in Development

'Participation' has been a clarion call for a variety of ideological and institutional interests in development, from the colonial project of 'community development' through to the focus on social movements by those espousing post Marxism and postdevelopment alike. There have been a number of different approaches within development theory and practice over the past century that have taken participation as a key element of their overall project (see **Table 1**). Here it is instructive to distinguish between the role that participation is envisaged to play in terms of development as a historical process (e.g., capitalism, political development, social change), and the more immediate world of development as a specific set of interventions (e.g., projects). For example, approaches to participatory action research during the 1970s encouraged participants to analyze wider societal problems and then sought to integrate analysis and action in iterative cycles, with action against oppression often taking a political form (e.g., within political parties). More broadly, popular participation has clearly underpinned developmental and political action in many developing countries without external intervention. Against this, the later approach of PRA sought to transform relations within specific development encounters – between development professionals and project 'recipients' – and to generate knowledge not only for empowerment but also for development planning.

It is this version of 'participation in projects' that dominates in international development. Influenced by key pioneers such as Michael Cernea (a strategically placed sociologist at the World Bank) and Robert Chambers, participatory approaches were progressively globalized over the late 1980s and early 1990s through training workshops, specialist journals, and word of mouth. A range of committed nongovernmental organizations and practitioners from both the developed and developing world were critical in the expansion and mainstreaming of such approaches within the larger bilateral and multilateral agencies.

There are a number of broader reasons for the emergence of participatory approaches and their rapid and widespread adoption by development agencies and practitioners in the 1980s. The ineffectiveness of externally imposed and expert oriented forms of research and planning had become increasingly evident, and participatory approaches emerged primarily as a response to this. The shift from blueprint to process approaches to development planning that began in the early 1980s led to a concomitant shift from expert and professional knowledge to people's knowledge implying a new professionalism and a change in balance of power of stakeholders with greater recognition and support for 'local' people. Furthermore, it became increasingly clear that beneficiaries could be empowered not only as a consequence of the development intervention but also through the methods of enquiry. The evolution of these methods also took place in connection with a shift in the social sciences, away from quantitative, and toward more

Table 1 Participation in development theory and practice: A selective history

Era	*Approach*	*Institutional and intellectual influences*	*Development theory: approach to development as historical process and development as intervention*	*Approach to citizenship*	*Locus/level of engagement*
1940s–1950s	Community development (colonial)	United Kingdom Colonial Office 1944 Report on Mass Education in Africa	*Development as historical process* (Re)produce stable rural communities to counteract processes of urbanization and sociopolitical change, including radical nationalist and leftist movements *Development as intervention* Development requires participation and self-reliance; cost-sharing. *Animation rurale*, adult literacy, and extension education, institution building, leadership training, and local development projects	Participation as an obligation of citizenship; citizenship formed in homogeneous communities	Community
1960s–1970s	Community development (postcolonial)	Postcolonial governments (Social Welfare or specialized departments)	*Development as historical process* As above; also development of state hegemony, moral economy of state penetration, etc. *Development as intervention* As above; also health, education, etc.		
1960s	Political participation	North American political science	*Development as historical process* Political development dimension of modernization theory. Participation as securing stability, legitimacy for new states, and strengthening the political system *Development as intervention* Voter education; support for political parties	Participation (e.g., voting, campaigning, political party membership) as a right and an obligation of citizenship	Political system and constituent parts; citizens
1960s–1970s	Emancipatory participation	Radical 'southern' researchers/ educationalists. Friere, Fals Borda, Rahman	*Development as historical process* Analyze and confront structures of 'oppression' within existing forms of economic development, state formation, political rule, and social differentiation	Participation as a right of citizenship; participatory citizenship as a means of challenging subordination and marginalization	Economic and civic spheres; communities; citizens

(*Continued*)

Table 1 Cont nued

Era	*Approach*	*Institutional and intellectual influences*	*Development theory: approach to development as historical process and development as intervention*	*Approach to citizenship*	*Locus/level of engagement*
	L berat on theo ogy	2nd Vat can Counc , Lat n Amer can Catho c pr ests. Gu tt erez, Sobr no	*Development as intervention* EP: Part c patory act on research (PAR), consc ent zat on, popu ar educat on, support for popu ar organ zat ons LT: Form base Chr st an commun t es, tra n ng for transformat on, popu ar educat on, etc.		
1970s–1990s	A ternat ve deve opment'	Dag Hammarskjo d Conference 1974. Deve opment D a oge, FAD Doss er Nerfin, Fr edmann	*Development as historical process* Cr t que of ma nstream' deve opment as exc us onary, mpover z ng and homogen z ng; proposa of a ternat ves based around terr tor a sm, cu tura p ura sm, and susta nab ty *Development as intervention* Popu ar educat on; strengthen soc a movements and se f-he p groups	Part c pat on as a r ght of c t zensh p; c t zensh p as a key object ve of a ternat ve deve opment, to be rea zed n mu t - eve ed po t ca commun t es	n t a y focused on commun t es and c v c soc ety, atter y the state through nc us ve governance'
1980s–present	Popu st/Part c pat on n deve opment	Deve opment profess ona s, NGOs (e.g., MYRADA, ED) Wor d Bank Part c pat on Learn ng Group, NGDOs, UN agenc es Chambers	*Development as historical process* L tt e d rect engagement; mp c t cr t que of modern zat on *Development as intervention* Fa ure of top-down projects and p ann ng; part c pat on requ red to empower peop e, capture nd genous peop e's know edge, ensure susta nab ty and effic ency of ntervent ons Part c patory: rura /urban appra sa , earn ng and act on, mon tor ng and eva uat on; NGO projects.	Focus on part c pat on n projects rather than n broader po t ca commun t es	Deve opment profess ona s and agenc es; oca part c pants
M d-1990s–present	Soc a cap ta	Wor d Bank Soc a Cap ta and C v	*Development as historical process*	Part c pat on as a r ght and ob gat on of c t zensh p	C v c assoc at ons

		Soc ety Work ng Group Putnam, Bourd eu, Narayan	Soc a cap ta promoted as a bas s for econom c growth *Development as intervention* Loca nst tut on bu d ng, support part c pat on n networks and assoc at ons		
Late 1990s–present	Part c patory governance and c t zensh p part c pat on	Part c patory Research and Act on (De h), nst tute for Deve opment Stud es, Br ghton (Part c pat on Group).	*Development as historical process* Deve opment requ res bera or soc a democracy, w th a respons ve state and strong c v soc ety. Some focus on soc a just ce *Development as intervention* Convergence of soc a ' and po t ca ' part c pat on, sca ng-up of part c patory methods, state–c v c partnersh ps, decentra zat on, part c patory budget ng, c t zens hear ngs, part c patory poverty assessments, PRSP consu tat ons	Part c pat on as pr mar y a r ght of c t zensh p	C t zens, c v soc ety, state agenc es and nst tut ons

qualitative, research methodologies which were seen to have stronger explanatory power. In historical terms, the emergence and use of participation has also been linked to the pursuit of political objectives at different moments. These have included efforts by colonial powers to co opt and stabilize rural communities into particular forms of state building and social engineering, and more recent attempts to promote higher levels of state accountability and legitimacy by involving citizens in evermore influential processes of policymaking and governance.

Arriving at the time of the 'impasse' in development studies, participatory approaches held the promise of renewal for a field of theory and action beset by internal soul searching and external critique. In this context, it is important to note that the rise of participation gained ideological support from alternative/radical perspectives that emphasized popular empowerment and participatory forms of democracy, but also from within the growing neoliberal hegemony, whereby individuals (rather than the state or collective movements) could now be cast as the key agents of development.

There are numerous benefits of participatory development, not least that such analysis is seen to be better informed, because the research and data on which it is based has been generated collectively between development practitioner and potential beneficiary. Furthermore, participatory research builds a sense of commitment, and allows local knowledge to contribute to how the project is implemented while participatory evaluation gives beneficiaries the chance to comment on the effectiveness of a given development intervention.

Another advantage, perceived by those who see participation as an end in itself, is that the process of involving intended beneficiaries in decisions, planning, implementation, and evaluation of projects, from which they had traditionally been excluded, can itself be empowering. Moreover, participation can enable poor and marginalized people to become aware of the structural reasons for their exclusion and how their own views of the world and thought processes can or cannot challenge these.

Participatory approaches are seen as producing information that is closer to the 'truth' than other less participative, top down methods of enquiry and knowledge accumulation. These claims to acquiring 'truer' knowledge and, for some practitioners, to empowering participants through their involvement in the process, have encouraged the overwhelming adoption of participatory techniques within development policy and practice.

Participatory Methods: A Focus on Research

Although participatory development incorporates a range of processes and practices, tools, and techniques, it is most commonly associated with PRA. PRA can be used to identify and assess community needs and priorities for, and feasibility of, development activities as well as for monitoring and evaluating the impact of these. It can also inform continuous readjustment of the program as a consequence of the information gathered. A significant characteristic of PRA methods is the emphasis on visual methods of collecting and analyzing data, primarily for the purpose of ensuring that less literate people can be fully involved in the process. These methods include, diagramming (seasonal calendars, time lines, Venn diagrams, etc.), mental and social mapping and modeling, transects and historical timelines, ranking and scoring preferences, observation, focus groups, and role play.

Participatory training provides concrete skills in these methods as well as highlighting the potentials and limitations of these. Importantly, research training manuals emphasize the necessity of practitioners to adopt particular sensibilities to empower community members to express, share, enhance, and analyze their knowledge, including, for example, showing respect for local people and interest in what they know.

In this approach, informants are not objects of study, but are participants in the research process and therefore problems need to be understood from their point of view. A priority within participatory approaches is to limit the separation between the collection and analysis of information. Thus research is not a mechanical procedure of information gathering where data is collected in one place and then analyzed 'back home' but an iterative and flexible process where information is collected and analyzed in the 'field' and issues that arise feedback into the process. Although indigenous knowledge is prioritized, triangulation is a key component of the research process whereby information sources, informants, and methods are cross checked to incorporate different people's perspectives and different methods. Indeed, one of the key principles of PRA is the offsetting of biases not only among informants but also among research facilitators. Participatory practitioners are encouraged to consider and address the different understandings of interviewer/interviewee about the purpose of the enquiry and in what ways their relationship shapes the kinds of data gathered. Furthermore, they need to be aware of the dynamics of group activities and to ensure that information generated by different social groups with potentially conflicting viewpoints have been adequately represented and to be cognizant of any groups that have been left out of the process and the power dynamics that shape different people's participation and interests. This has resulted in a version of participation geared toward knowledge production as a means of catalyzing social change and challenging dominant relations of power.

Critical Challenges

The mainstreaming of participation has been accompanied by a rising tide of critical debate concerning the extent to which it has actually lived up to its often ambitious claims. Criticisms have come both from within the 'participatory movement' and without, and generally take two main forms: those that focus primarily on definitional differences, debates over the objectives of participation and the applicability and appropriateness of the methodologies and techniques and, those that pay more attention to the theoretical, political, and conceptual limitations of participation.

Technical limitations of the approach are primarily articulated by those who stress the need for self critical reflexivity and a reexamination of the methodological tools used. Within the discourse there is an espousal of ongoing self critical epistemological awareness that is considered an essential component of participatory ideology and practice. This generates a continuous dialog between practitioners on the quality, validity, and ethics of what they are doing in order to prevent poor practice or exploitation of the people involved. This reflexivity does not represent a critique of participatory methodology *per se*, however, but is seen more as an intrinsic facet of the approach itself. In this way, it is envisaged that the methodological and practical problems of the approach are continuously highlighted and revisited. However, the often pressurized daily realities of development work do not always allow for such reflexivity to emerge in practice.

There are other critiques of participation, particularly focused on PRA, that foreground issues of diversity and differentiation. Use of the homogenizing term 'community', for example, has been challenged with questions raised about representations of 'communities' as static and harmonious units within which people share common interests and needs. Critics argue that this understanding of the notion of 'community' conceals power relations within 'communities' and further masks biases in interests and needs based on, for example, age, class, caste, ethnicity, religion, and gender. Frequently, simplistic notions of 'community' employed in PRA expose its weakness in exploring social and political complexity, which may be better understood using nonparticipatory methods. There is also an assumption within participatory practice that 'consensus' is possible and/or desirable in a 'community', whereas, it may represent co option of the process in the interests of existing elites. The question therefore arises as to whether participatory research can ever take place outside of local power relations, let alone challenge these. Further dangers can arise when the data produced by PRAs are interpreted out of the context in which they were produced. This context includes informants' perceptions of the purpose of the PRA, and of the facilitators, and the problem is compounded when there is a scarcity of reflection and reporting on the PRA process, including the role of the 'practitioners'.

Such methodological dangers are particularly apparent where efforts are made to scale up participatory research exercises in order to inform policymaking at the national level. For example, 'Participatory Poverty Assessments' have involved efforts to aggregate PRA findings from a wide variety of different sites, despite the fact that the particularly flexible process of participatory research (unlike more standardized survey work) rarely unfolds in a similar enough way to allow for genuine comparisons and aggregation to occur. It has also proved difficult to determine which issues should be prioritized from this mass of rich oral and visual data, with long and unproductive wish lists sometimes being the unsatisfactory outcome.

Other critics move beyond an identification of limitations of, and adjustments to, the methodology to more fundamental critiques of the discourse of participation, albeit recognizing that some of these do indeed emerge out of technocratic concerns. These critiques provide important insights into the participatory development discourse, through a conceptual and ideological examination of its theories, methods, and practices. They focus on the extent to which participatory facilitators override existing legitimate decision making processes, how group dynamics can lead to decisions that reinforce the interests of the already powerful, and how certain kinds of knowledge are deemed less important than others. They highlight potential barriers in the public arena for the expression of certain views and views of certain marginalized people who would not normally be expected even to be present in formal public encounters and who are unaccustomed to expressing themselves in terms suitable for a public audience. A further concern is related to the inexorable rise and prioritization of participatory methods in development that is seen to have driven out other methods of enquiry that have advantages participation cannot provide such as in depth ethnographic and more theory based research.

Others challenge the conceptual and political limitations of the approach, demonstrating how participation can lead to political co option and mask continued centralization of development policy and planning in the name of decentralization and participation. For example, the 'local knowledge' shared in the process can be used to justify or legitimize a predetermined project agenda rather than reshaping or influencing the project objectives. These problems arise, in part, due to the perceived tension of incorporating and matching participants' interests with existing donor priorities and thus the potential for manipulating people's needs and interests so that they fit into those of the funding organization.

Some critics examine the complex micro politics and social practices of knowledge production uncovering differences in the constitution of knowledge and the diverse ways of knowing. Culturally, socially, and politically produced, knowledge is a powerful normative construct formed out of an accumulation of social norms, rituals, and practices which, far from being constructed in isolation from power relations, is embedded in, or against, them. Within much participatory discourse, however, 'people's knowledge' or 'local knowledge' is seen as a fixed commodity that people intrinsically have and own, as individual experience, preference, or choice. It is not seen to embody power relations since it appears to be articulated and believed in by all.

Participatory research has also been criticized for simplifying complex processes and realities so that they are manageable for the planner. It 'cleans up' local knowledge through mapping and codification, and marginalizes or discards that which might challenge the status quo or is messy or unmanageable. Techniques such as seasonal calendars and wealth ranking require a purification of knowledge and the exclusion of experiences that do not fit the structured representations implied by the tools.

These critiques from within and outside the participatory movement are clearly related. The mainstreaming of participation during a particularly neoliberal and technocratic moment of development theory and policy is considered by some to have drained the approach of its radical potential. Participatory approaches become a technical method of project work designed to 'change attitudes' rather than a political methodology for challenging structural exclusions at a range of scales. The failure to look beyond 'the primacy of the personal', and to examine the underlying causes of exclusion, rather than those related to development interventions alone, has lent participatory approaches a methodological individualism that is inevitably depoliticizing, and which obscures deeper understandings of what makes participation difficult for marginal groups in the first place, particularly in relation to processes of state formation, social stratification, and political economy. Significantly, participation lacks a coherent theory of social change, such that it is not clear how transformations within development interventions relate to broader social change.

Participatory approaches can mask, and indeed sustain, inequalities and injustices, by co opting and thus reducing potential spaces of conflict and dissent. Yet, intended participants can, and do, resist participation in a number of ways. These include simply refusing to participate, rejecting projections about their lives, retaining information, and presenting themselves in a variety of diversionary and conflictual ways. Furthermore, given the long, and often unhappy, history of efforts to co opt popular agency in many developing countries, let alone the time costs of participating, noninvolvement may well be an understandable response.

Moving Participation Forward: Dealing with Power and Politics?

Despite the veracity of these critical challenges, they have notably failed to halt the spread of participation as a development concept and strategy. Apparently undeterred, and increasingly underwritten by policy and funding support from major development agencies, the participatory turn has become expressed more deeply and diversely within development theory and practice over recent years. Within this continued upward trajectory it is possible to discern two types of response to the above critiques, particularly the charge that participatory approaches fail to adequately address issues of politics and power that the rhetoric of empowerment raises.

The first has been a largely practical and methodological response, whereby problems of power and politics are addressed by levering participatory approaches into ever higher levels of decision making. Examples here include not only the scaling up of PRA methods into national research exercises (as with participatory poverty assessments) but also a wider push for participation to form the basis of development policymaking at national levels. The key exemplar of this has been the insistence by the World Bank that governments should undertake broad based consultations with citizens as part of formulating, implementing, and evaluating their debt related poverty reduction strategies. Here participation is charged with its usual roles of securing relevance and legitimacy but also of the broader goal of 'ownership'. Such moves are relatively new, although there is evidence to suggest that participatory processes within such processes have led to some institutional innovations that have brought civil society organizations into the policy process, and helped ensure a stronger pro poor focus within policymaking. However, numerous problems have also emerged, including the exclusion of key actors from consultative processes (particularly unions and political parties); the timing and constraints on genuine participation; the discretionary and patronage based character of these forms of 'invited' consultation; and the failure to open up key aspects of government and donor policy to participatory scrutiny (most notably macroeconomic policy).

Linked but not confined to this shift has been the broader emphasis placed on participatory forms of democracy and governance. Promoted by donors, but often originating within national political projects, many countries have sought to implement democratic forms of decentralization, participatory planning within municipal governments, and even participatory budgeting, whereby citizens play a direct role in setting budget priorities and monitor expenditure patterns. Although some successes have been apparent, it is not always clear that such ap proaches are more successful or effective than centralized approaches, or that they have transformed the tensions between participatory and more representative forms of democracy and governance into more progressive synergies.

The second set of responses address the more fun damental criticisms of participation by reconceptualizing participation into a more politicized form of theory and (by implication) action, and derive from a belief that participation remains a critical aspect of an emancipatory and progressive form of politics. For example, some have suggested that what is required is to focus on developing the 'political capabilities' of poor and marginal groups. Here, the notion of political capabilities can 'repoliticize' the theory and practice of participation in development and governance by drawing attention to long term pro cesses of political learning; the ways in which ideas, identities, and collective self awareness constitute valu able political resources; and the importance of under standing empowerment in relation to the relationship between the state, the organizational resources of the poor and relevant policy areas. As such, it appears to offer new strategies for promoting political inclusiveness.

Others have sought to relocate 'participation' within citizenship analysis. This situates it in a broader range of sociopolitical practices through which people extend their status and rights as members of particular political communities, thereby increasing their control over socioeconomic resources. Participatory notions of citi zenship are particularly relevant to marginalized groups by offering the prospect that citizenship can be claimed 'from below' through their own efforts in organized struggles rather than waiting for it to be conferred 'from above'. The question for participatory interventions becomes how the 'competency' of participants to project their agency beyond specific interventions into broader arenas can be enhanced, thereby progressively altering the 'immanent' processes of inclusion and exclusion. In theory, then, citizenship offers a means of covering the convergence between participatory development and participatory governance, and of linking participation to rights based approaches, thus providing a stronger pol itical, legal, and moral imperative for focusing on people's agency within development than is currently the case. It draws attention to development as both intervention and social change, by situating participation within a broader political, social, and historical perspective that draws attention to the politics of inclusion and exclusion that shape popular agency beyond particular interventions.

While debates concerning the legitimacy and trans formative effects of participation continue, the recent period of intense critique and response has led to more realistic claims being made by both critics and proponents of participatory approaches, although vested interests and ideological commitments remain strong. It is now increasingly recognized that participatory discourse and practice need to engage more seriously with issues around power and politics.

See also: Aid; Citizenship; Civil Society; Debt; Empowerment; Indigenous Knowledges; Neoliberalism; Neoliberalism and Development; Nongovernmental Organizations; Participatory Action Research; Participatory Video; Poverty.

Further Reading

Blackburn, J. and Holland, J. (eds.) (1998). *Who Changes? Institutionalising Participation in Development*. London: Intermediate Technology Publications.

Chambers, R. (1994). The origins and practice of participatory rural appraisal. *World Development* 22(7), 953 969.

Cooke, B. and Kothari, U. (eds.) (2001). *Participation: The New Tyranny?* London: Zed.

Cornwall, A. and Pratt, G. (eds.) (2003). *Pathways to Participation: Reflections on PRA*. Londo: Intermediate Technology Publications.

Freire, P. (1972). *Pedagogy of the Oppressed*. Harmondsworth: Penguin.

Gujit, I. and Shah, M. K. (1998). *The Myth of Community: Gender Issues in Participatory Development*. London: Intermediate Technology Publications.

Heller, P. (2001). Moving the state: The politics of democratic decentralization in Kerala, South Africa, and Porto Alegre. *Politics and Society* 29(1), 1 28.

Hickey, S. and Mohan, G. (eds.) (2004). *From Tyranny to Transformation? Exploring New Approaches to Participation*. London: Zed Books.

IDS Bulletin (2002). *Special Issue: Making Rights Real: Exploring Citizenship, Participation and Accountability. IDS Bulletin* 33(2).

Kapoor, I. (2002). The devil's in the theory: A critical assessment of Robert Chambers' work on participatory development. *Third World Quarterly* 23(1), 101 117.

McGee, R. (2001). Participation in development. In Kothari, U. & Minogue, M. (eds.) *Development Theory and Practice: Critical Perspectives*, pp 92 116. Basingstoke: Palgrave.

Mohan, G. and Stokke, K. (2000). Participatory development and empowerment. *Third World Quarterly* 21(2), 266 280.

Mosse, D. (1994). Authority, gender and knowledge: Theoretical reflections on PRA. *Development and Change* 25(3), 497 526.

Naryan, D., Chambers, R., Shah, M. K. and Petesch, A. (2000). *Voices of the Poor: Crying Out for Change*. New York: Oxford University Press.

Whitehead, L. and Gray Molina, G. (2003). Political capabilities over the long run. In Houtzager, P. & Moore, M. (eds.) *Changing Paths: International Development and the Politics of Inclusion*, pp 32 57. Ann Arbor, MI: University of Michigan Press.

Relevant Websites

http://www.worldbank.org
Participation and Civic Engagement, The World Bank.

http://www.un.org
Participation in CSD 17 and its Intergovernmental Preparatory Meeting New York, 23 28 February 2009 and 4 15 May 2009, United Nations Division for Sustainable Development.

http://www.ids.ac.uk
Participation Team Overview, Institute of Development Studies.

http://www.drc citizenship.org
The Development Research Centre on Citizenship, Participation and Accountability.

Participatory Action Research

S. Kindon, Victoria University of Wellington, Wellington, New Zealand
R. Pain, University of Durham, Durham, UK
M. Kesby, University of St. Andrews, St. Andrews, UK

Glossary

Discourse A specific series of representations, practices, and performances through which meanings are produced, connected into networks, and legitimized.

Emancipation A term used to describe the process of freeing someone (usually a disenfranchised group or individual) from various political and other restrictions, or the state of being free from those restrictions.

Empowerment A process of developing the power and capacity to overcome internalized oppression, to negotiate and influence the nature of relationships, and/or to act collectively to change political structures and local conditions.

Paradigm The assumptions, procedures, and findings generally accepted by a group of scholars, which together define a stable pattern of their research orientation and activity.

Participatory Development An approach to international community development which believes that people are important actors capable of shaping their own future. It uses various approaches, which involve local decision making in the design, planning, and implementation of any intervention.

Introduction

Participatory approaches to research offer a radical challenge to 'how' data are collected, 'what sort' of new knowledges result, and 'what impacts' they have; as well as to 'who' directs and benefits from research. Participatory action research (PAR) – the most popular of these approaches in human geography to date – is a collaborative process of research, education, and action explicitly oriented toward social change. It involves academic researchers (usually full time and paid) and nonacademic co researchers and participants (usually part time on the project and not paid) working together to examine a problematic situation in order to change it for the better on participants' own terms.

This kind of research is certainly not new (in general or to geography). For over 70 years, advocates of such approaches have sought to challenge the traditional model of research (which might be characterized as imperial, hierarchical, and extractive). They have also sought to promote alternative modes of investigation, which enable participants' empowerment in and through the research process and in which the benefits of research accrue more directly to communities involved. However, notwithstanding this long history, a recent sharp resurgence of interest has meant that participatory approaches are only now becoming a leading paradigm within contemporary social and environmental sciences.

We believe that PAR presents some of the most radical, creative, exciting, and challenging approaches to knowledge production in human geography, and clearly our voices are dominant within this article. However, in keeping with the value orientation of PAR, which works to embrace diverse knowledges and destabilize hierarchical relationships between researchers and participants, this article also includes voices from nonacademic colleagues reflecting on their priorities and experiences with PAR.

Origins and Epistemological Underpinnings

When engaging with participatory approaches, it is important to avoid simply adopting the many novel methods into 'research as usual'. First, it is vital to understand the theoretical reasoning that enables innovative techniques to be deployed within genuinely alternative modes of research that are action oriented.

Historically, the term 'action research' was initially coined in the USA by Kurt Lewin (1946), who introduced the idea of an iterative (or repeated) cycle of action and reflection in research, something he called 'spiral science'. Similarly, Sol Tax and William Foote Whyte developed an approach to research that enabled local people to directly voice their concerns without mediation by outside experts. However, it was the South American emancipatory educator Paulo Freire's community based approach to research in 1960s' Brazil that most famously inspired researchers to involve their participants directly in knowledge production and turn research into a means of social transformation. Freire developed an epistemology in which investigation was grounded in people's struggles and local knowledges, and he saw research as an opportunity for *conscientização* (conscientization), a process through which poor and marginalized people could develop a heightened awareness of the forces affecting their lives, and on which they could base social and political action. His ideas connected with others in the majority world who were dissatisfied with the ongoing legacies of colonization, modernistic development inter ventions and positivist research paradigms promoted by university based researchers, and the early 1970s par ticipatory approaches proliferated across Latin America, Africa, and India (where they resonated with the earlier social movements led by Mahatma Gandhi). In the years that followed, the foundations for PAR were laid, among others, by Marja Liisa Swantz in Tanzania, Rajesh Tandon in India, and Orlando Fals Borda in Colombia. Then in the 1980s, a 'second wave' of PAR emerged in community led international development and partici patory development.

As a result of this innovation, a plethora of approaches to action oriented research and development now exists. Of these, we favor PAR because it signals a commitment to directly engage participants in collaborative processes as well as to engage in research for action. Further more, within the diverse and overlapping 'schools' that compile contemporary PAR, we particularly value those informed by feminist, post structural, and postcolonial perspectives. We discuss these briefly here and return to them later in more detail.

Feminist theories offer insights into the nature and depth of social inequality, highlight the masculinist nature of 'research as usual', and speak directly to the need for collaborative, participatory research and alter native methods. Feminism also has a longstanding com mitment to activism within and outside of the university. Janet Townsend's early geographic work with women in Latin America exemplifies this kind of participatory feminist approach and illustrates how participatory practice can be extended into collaborative research outputs, a point we return to later.

Post structuralist theories have been helpful in pro viding alternative and relational understandings of power and the workings of discourse, which both constrain and enable participatory processes. Working on a post structuralist vein, geographers like Jenny Cameron and Kathy Gibson have demonstrated the politically powerful effects of paying attention to language, identity, and rep resentation in PAR projects in Australia. Caitlin Cahill's work also draws productively on post structuralist think ing to explore how subjectivities of young women change throughout their engagement as co researchers on a PAR project in New York.

Postcolonial theories have been invaluable in calling explicit attention to the ongoing effects of colonization in present day institutional arrangements, images, and representations, and interpersonal relationships. This is critically important in the context of PAR's focus on working with marginalized and traditionally oppressed groups. Sara Kindon's work with Geoff Hume Cook and Te Iwi o Ngaati Hauiti – a Maaori tribe in Aotearoa, New Zealand – has used participatory video to destabilize the masculinist and colonial gaze of mainstream media representations of Maaori, and has provided a vehicle for self representation and community building.

Clearly, as an approach, PAR is diverse because it is always shaped by a commitment to locally appropriate public engagement. Nevertheless, it is also always marked by a commitment to theorize and produce knowledge through mutual respect, dialog, and inclusive methods, and as such a number of general characteristics of PAR, and those that facilitate it, are distinguishable, which provide a useful guide to the politics, practice, and pro fessional identity of those involved (**Box 1**).

The Participatory 'Turn' in Geography

Three forces have driven the recent explosion of interest in participatory approaches in human geography: first, a

Box 1 Key characteristics of PAR and PAR researchers

Participatory action research

1. Aims to change practices, social structures, and media which maintain irrationality, injustice, and unsatisfying forms of existence
2. Treats participants as competent and reflexive agents capable of participating in all aspects of the research process
3. Is context-bound and addresses real-life problems
4. Integrates values and beliefs that are indigenous to the community into the central core of interventions and outcome variables
5. Involves participants and researchers in collaborative processes for generating knowledge
6. Treats diverse experiences within a community as an opportunity to enrich the research process
7. Leads to the construction of new meanings through reflections on action
8. Measures the credibility/validity of knowledge derived from the process according to whether the resulting action solves problems for the people involved and increases community self-determination

Participatory action researchers are generally

1. Hybrids of scholar/activist where neither is privileged
2. Interdisciplinary
3. Mavericks/heretics
4. Patient
5. Optimistic, believe in the possibility of change
6. Sociable and collaborative
7. Practical and concerned with achieving real outcomes with real people
8. Able to be flexible and accommodate chaos, uncertainty, and messiness; able to tolerate paradoxes and puzzles and sense their beauty and humor
9. Attracted to complex, multidimensional, intractable, dynamic problems that can only be partially addressed and partially resolved
10. Engaged in embodied and emotional intellectual practice

series of conceptual shifts over the last decade in philosophical critique, in economic and environmental policy, and in international geopolitics; second, the wider uptake of PAR in other social and environmental sciences; and third, increasing critical reflection among those employing such approaches.

Ethically, PAR is increasingly viewed as one answer to the recent (re)questioning of relevance and to renewed calls for more morally aware, less hierarchical research practice which challenges and acts upon inequality. Intellectually, PAR is well suited to research that explores people's relations with, and accounts of, space, place, and environment because it is context specific, forefronts local conditions and knowledge, and produces situated, rich, and layered accounts. It also offers a concrete means of attempting to connect the multiple relations between issues and processes at different scales.

The 'participatory turn' in geography is, itself, stimulating a 'geographical turn' in participatory approaches as advocates become aware that PAR reflects and affects the spaces within which it takes place. Thus, the geographies of PAR are vital to how it plays out to effect or to stymie change. Informed by these spatialities and relationalities, PAR leads to new or modified theories of its materialities and objects of study, such as environmental degradation, the circulation of social capital, or the impacts of crime in rural areas. PAR's grounded and relational orientation toward knowledge construction challenges the hegemonic norms of insular academic production and enables more diverse and contextually rich theorizations and actions to emerge.

For example, in Rachel Pain's current PAR art project with locally born and refugee young people in Newcastle upon Tyne UK, Gaby Kitoko, a refugee community activist and research collaborator reflects that one benefit of the research has been that it has worked as a space where young white and black children have been able to integrate, often for the first time (**Box 2**).

It has also worked to produce knowledge and expressions of feelings about being a young refugee, which were used to try and effect change in wider arenas (**Figure 1**).

The relationships between the participatory and other 'turns' in geographical enquiry are worthy of further exploration. For example, while some viewed the cultural turn of the early 1990s as a move away from forms of radical research focused on practical social change, nevertheless, its focus on the voices and histories of marginalized social groups and emphasis on qualitative methods, produced the conditions in which PAR's older critique of conventional research could flourish.

More recent resonances have been with the 'emotional turn', which insists on the emotional and embodied nature of knowledge production, and the 'material turn', which has refocused attention upon how the materialities of objects within particular cultures and landscapes might be mobilized to challenge injustice and ground political struggles. Finally, while the 'nonrepresentational turn', has largely ignored the legacy and potential of participatory approaches, its call for methods that attend to the range of sensory experiences that inform knowledge and action, and which reach beyond the limits of the textual, verbal, and material to engage new media, lived experiences, performative and haptic knowledges, might find some useful responses within PAR.

Box 2 Grounding theoretical 'turns' through participatory action research

… black and white kids were not mixing very much. But I'm happy because we start to have those activities where they can come together and let the community know about […] things that are important to them to try to break this ice between those communities.

Gaby Kitoko, African Community Advice North East

Figure 1 Role play: young people wrote and acted out a drama in a public space to show how to deal with bullying, in Newcastle-upon-Tyne. From Pain, R., Askins, K. and Kitoko, G. (2007) *Connecting Places, Connected Lives: An Antibullying Art Project.* Durham: Durham University.

Critiques of Participatory Approaches: Power and Empowerment

Naturally, participatory approaches (including PAR) are not without their critics. Many being influenced by the recent post structuralist and postcolonial 'turns' mentioned above, associated with the work of theorists like Michel Foucault, Gayatri Spivak, and Edward Said. While those sympathetic to participatory approaches worry that their terminology and techniques are being appropriated and commodified within policy and research initiatives (e.g., in international development) that essentially remain unchanged (i.e., top down and extractive), a more hostile set of critiques propose that participatory approaches are just another form of power, producing various negative effects (see **Box 2**). Most damningly, these critics suggest that rather than facilitating the empowerment of participants, participatory approaches actually reinforce existing hierarchical power relations, reproducing the very inequalities they seek to challenge (**Box 3**).

Rather than abandoning participatory approaches however; we respond to these critiques by suggesting that the post structuralist, postcolonial, and participatory 'turns' need not necessarily lead in different directions. Yes, participation is a form of power, but then there is no escape from power. Indeed, it is helpful to retheorize the empowerment sought and facilitated through participatory research along post structuralist and postcolonial lines, viewing it as an 'effect' of the deployment of the powerful resources of participation. Such an understanding helps to explain why empowerment and decolonization can be difficult to sustain over time and space, why it is insufficient to create the conditions for empowerment and decolonization within the arenas of participation, and highlights why researchers need to identify which resources will 'travel' and can be 'distanciated', so as to sustain empowered and decolonized performances in contexts that would normally facilitate the subjection or marginalization of certain people and groups.

Our interpretation of post structuralist and postcolonial critiques is not absolutist, but rather is refracted through an older feminist lens that perceives the ambivalent role of strategic essentialism. Thus, we would argue that it is legitimate, as well as practically necessary, to deploy forms of governance like participatory research, which although they shape and direct people's conduct, at least do so in ways that enable them to challenge and change other, more domineering powers and relations in their lives.

Box 3 Some potential negative power effects of participatory approaches

- Delegitimization of research methods that are 'not' participatory
- Production of participants as subjects 'requiring' research/development
- Production of suitably disciplined subjects as 'participants' expected to perform appropriately within participatory processes
- Retention of researchers' control while presenting them as benign arbiters of neutral or benevolent processes
- Reauthorization of researchers as experts 'in' participatory approaches
- Romanticization or marginalization of local knowledge produced through participatory processes
- Reinforcement of preexisting power hierarchies among participating communities
- Legitimization of elite local knowledge simply 'because' it is produced through participatory processes
- Legitimization of neoliberal programs and institutions (such as the World Bank) that also deploy participatory approaches and/or techniques

The key of course, is to try and maintain a self reflexive appreciation that rather than being a privileged, power free mode of research, participatory approaches, and PAR, are situated, limited works in progress that are improvable, contestable, and ultimately, replaceable. In this regard we see participation, post structuralism, and postcolonialism as mutually informative and not inherently antagonistic. Furthermore, we believe that the people friendly, accessible, and flexible methodologies of participatory approaches actually offer useful resources through which the critical social reimaginations promoted by post structuralist and postcolonial scholars might be distanciated beyond the academy.

Research Processes, Methods, Outputs, and Outcomes

We have placed a discussion of participatory methods toward the end of our article in order to underline the key point made earlier – conducting participatory research requires far more than the use of participatory methods or techniques in data collection. It requires an epistemological and practical shift in how we do research, how we relate to our participants, and what we think of as knowledge. The ethics of participation involves explicit attention to the negotiation of ownership at each stage of the research from scoping to publication.

As suggested earlier, the general methodology of PAR is structured as a cyclical process of action and reflection. Researchers and participants identify an issue or situation in need of change or identify strengths and attributes they wish to enhance in their communities; they then initiate research that draws on capabilities and assets to precipitate relevant action. Both researchers and participants reflect on, and learn from, this action and proceed to a new cycle of research/action/reflection.

While a considerable catalog of techniques have already been developed and recorded, the key is that researchers and participants should together develop context specific methods that facilitate the iterative cycles of action and reflection in their own specific projects (see **Box 4**).

Significantly, most of the approaches and methods used within PAR projects are variations of standard methods used in conventional research (e.g., mapping, interviewing, questionnaire surveys, and focus groups, as well as approaches involving technologies like photography, video, and geographic information systems (GIS)). What is different in PAR is that these methods

Box 4 Methods used in participatory action research

- Art and media
- Diagramming and drawing
- Dialog
- Educational camps
- Exchange programs
- Group work and discussions
- Interviewing
- Learning by doing
- Participant observation
- Political action and advocacy
- Mapping (paper and using GIS)
- Ranking and scoring
- Role plays and theater
- Secondary data analysis
- Shared analysis, writing, and presentations
- Storytelling
- Surveys
- Transect walks

and approaches are deployed within a participatory framework committed to genuinely democratic and noncoercive forms of research 'with' and 'for', rather than 'on' participants. It is this framework and value orientation that fundamentally changes the work done, the knowledge generated, and outputs produced.

PAR researchers (and others, notably many feminist researchers) question the impact of outputs such as academic journal articles on the issue being studied, as well as whose goals, interpretations, and perspectives they primarily represent. Therefore, a range of outputs typically characterize PAR projects and commonly include: diagrams, maps, reports, websites, videos, photography, stories, poetry, performances (drama or dance), public art, poster or sticker campaigns, awareness raising workshops, and formal presentations.

A key principle is that these outputs are woven into the research's outcomes or material effects, that is, they are geared toward influencing social change or transformation. As such, they may aim to improve interpersonal or interinstitutional relations, or increase the potential for self help within communities. In most cases, they may seek to enhance the personal growth and level of critical awareness of the individuals involved so that they leave behind enhanced capacities for future research, and social change. A useful illustration of this orientation is reflected in the comment from a nonacademic project member working with Sarah Elwood on a participatory GIS project in Chicago (**Box 5**).

The key point here is that while people's participation in research is important (and an ethical necessity), unless it translates into action that somehow improves their lives on their own terms, it serves to perpetuate (perhaps even exacerbate) existing inequalities.

Box 5 The importance of action-oriented outputs

If you do research, it has to be action-oriented […], or it will be thrown away and never really used. […] The GIS has become part of what we do in our organization […] and is one of the most important things that will help us prepare a community plan.

Staff Member, Near Northwest Neighborhood Network, Chicago

PAR projects are varied in their action orientation. Many work with or for institutions to affect change in service delivery or to policies, conventions, and laws affecting the groups involved. Other projects are more overtly radical and seek to further advocacy, resistance, or activism. In what they call 'solidarity action research', Paul Chatterton, Duncan Fuller, and Paul Routledge argue that PAR should more fully embrace a spirit of solidarity between academic researchers and participants, which involves active critique and disagreement, and does not romanticize or defer to the 'authentic' voice of those marginalized. This way of working returns us to an explicit engagement with power, emotions, and the spaces in which they occur, in our efforts to practice research that matters beyond the academy – engagements which human geographers are well placed to facilitate.

Conclusion

PAR and other participatory approaches present radical challenges to what might be, and should be researched in human geography, how, and by whom. While participatory approaches seek socially and environmentally just processes and outcomes, they nevertheless constitute a form of power and advocate the need to be alive to the fact that they can sometimes reproduce the very inequalities they seek to challenge. However, PAR is one means of repoliticizing research and practicing relevance, emphasizing dialogic engagement with co researchers, and the development and implementation of context appropriate strategies oriented toward empowerment and transformation at a variety of scales. While not a panacea for all the ills of contemporary research and development, there is much radical potential in PAR, as long as we maintain a critical and honest awareness of its challenges and dangers. By recognizing the roles of power, emotions, space, and place within PAR, and PAR's embeddedness within multiple scales, we can be more explicit about the cartographies of our engagements and ensure that all three elements – participation, action, and research – combine productively to effect positive political change.

See also: Activism; Auto-Photography; Community; Critical Geography; Cross-Cultural Research; Empowerment; Ethical Issues in Research; Focus Groups; GIS, Public Participation; Participation; Participatory Video; Postdevelopment; Situated Knowledge, Reflexivity.

Further Reading

Brydon Miller, M., Maguire, P. and McIntyre, A. (eds.) (2004). *Travelling Companions: Feminism, Teaching and Action Research*. Westport, CT: Praeger.

Cahill, C. (2007). The personal is political: Developing new subjectivities in a participatory action research process. *Gender, Place, and Culture* 14(3), 267 292.

Cameron, J. and Gibson, K. (2005). Participatory action research in a poststructuralist vein. *Geoforum* 36(3), 315 331.

Chatterton, P., Fuller, D. and Routledge, P. (2007). Relating action to activism: Theoretical and methodological reflections. In Kindon, S., Pain, R. & Kesby, M. (eds.) *Participatory Action Research Approaches and Methods: Connecting People, Participation and Place*, pp 216 222. London: Routledge.

Cooke, B. and Kothari, U. (eds.) (2001). *Participation: The New Tyranny?* London: Zed Books.

Elwood, S., Feliciano, R., Gems, K. *et al*. (2007). Participatory GIS: The Humboldt/West Humboldt Park Community GIS Project, Chicago, USA. In Kindon, S., Pain, R. & Kesby, M. (eds.) *Participatory Action Research Approaches and Methods: Connecting People, Participation and Place*, pp 170 178. London: Routledge.

Freire, P. (1972). *Pedagogy of the Oppressed*. Harmondsworth: Penguin.

Kapoor, I. (2005). Participatory development, complicity and desire. *Third World Quarterly* 26(8), 1203 1220.

Kesby, M. (2007). Spatialising participatory approaches: The contribution of geography to a mature debate. *Environment and Planning A* 39, http://www.envplan.com/abstract.cgi?id = a38326 (accessed in June 2008).

Kindon, S., Pain, R. and Kesby, M. (eds.) (2007). *Participatory Action Research Approaches and Methods: Connecting People, Participation and Place*. London: Routledge.

Lewin, K. (1946). Action research and minority problems. *Journal of Social Issues* 1 2, 34 36.

Maguire, P. (1987). *Doing Participatory Research: A Feminist Approach*. Amherst, MA: Centre for International Education, University of Massachusetts.

McTaggart, R. (ed.) (1997). *Participatory Action Research: International Contexts and Consequences*. New York: State of New York University Press.

Reason, P. and Bradbury, H. (eds.) (2006). *Handbook of Action Research*. London: Sage.

Townsend, J., Arrevillaga, U., Bain, J., Cancino, S., Frenk, S., Pacheo, S. and Crez, E. (1995). *Women's Voices from the Rainforest*. London: Routledge.

Relevant Websites

http://www.alara.net.au
Action Research, Action Learning Association (ALARA) Home page.

http://www.fed up honeys.org
Fed up Honeys Home page.

http://www.pygywg.org
Participatory Geographies Working Group (PyGyWG) of the Royal Geographical Society, UK Home page.

http://www.goshen.edu
Web Links to Participatory Action Research Sites, Anthropology/Sociology, Goshen College, Indiana, USA.

Participatory Video

S. Kindon, Victoria University of Wellington, Wellington, New Zealand

Glossary

Paradigm The assumptions, procedures, and findings generally accepted by a group of scholars, which together define a stable pattern of their research orientation and activity.

Participatory Development An approach to international community development which believes that people are important actors capable of shaping their own future. It uses various approaches, which involve local decision making in the design, planning, and implementation of any intervention.

Video A shorthand term for a range of audiovisual media of different formats, lengths, and purpose, from feature documentaries and short films to video clips, public service announcements, and raw footage.

Introduction

Participatory video (PV) has its origins in a desire to effect change. From its radical origins in Canada nearly 50 years ago, the use of video for participatory development, communication, and action has become more possible with changes in technology. Along the way, it has taken many forms, which have had different effects. In this article, the history and evolution of PV is reviewed and then its current uses in development, research, and advocacy was explored. Its process is summarized and various issues related to power, resources, ethics, and audience are attended to. Its benefits are identified and it is suggested that human geography has much to gain from engaging this particular methodology.

History and Evolution

The first participatory use of video occurred on Fogo Island, Newfoundland, Canada in the late 1960s. Widely referred to as the Fogo Experiment, the pioneering interactive use of film and video in remote locations was part of the government's Challenge for Change program. The Challenge for Change program was revolutionary in its desire to document social problems. It also sought to use the audio visual documentation process as a means of raising awareness and information sharing within and beyond the communities involved to achieve social change and community empowerment. It aimed at producing three kinds of films. First, films for government departments and the general public which explained a problem; second, films for social workers and change agents involved in the problem; and third, films planned and produced by the people affected by the problem.

In terms of current practices of PV, it was the last type of film that was most radical and influential. By putting community members behind as well as in front of the camera and by facilitating a process of community feedback (or dialog) on the films produced, the filmmaker Colin Low and academic researcher Donald Snowden demonstrated that the 'process' of production was often more important than the actual films produced. Through the dialogic process, research participants became image and meaning makers who explored and worked to change their own 'realities' through the production and analysis of video products.

In this respect, their process enabled participants' technical empowerment by increasing their knowledge and confidence using video and editing equipment. It also facilitated participants' personal and community empowerment by showing their lives on screen and facilitating discussions about possible actions for change. The Fogo Process enabled individuals in isolated communities to connect with each other and distant others through the collective representations produced, and the resulting discussions about them and the lives they represented. These connections fostered greater senses of both imagined and virtual communities, which in turn promoted greater social cohesion and action on the ground.

The Fogo Experiment lasted for 18 months then moved elsewhere in Canada through the Challenge for Change program for almost a decade. In 1969, the first indigenously authored film was produced dealing with treaty rights along the USA–Canadian border. By the mid 1970s and early 1980s, the process was exported to the USA as well as parts of Africa and Asia. Alongside the Fogo Process, Martha Stuart in the USA developed the Village Video Network of New York, which traveled to India in 1984 to work with the Self Employed Women's Association (SEWA) in Ahmedabad. Since its initial training, SEWA has gone on to be one of the most well known proponents of PV. They have involved hundreds of women in the process and produced more than 400 films. Through both the process and the products, they have achieved some major changes associated with labor and women's rights at various scales from household gender relations through to Indian state legislation.

Alongside these development oriented uses of video, it (and film before it) was being used in disciplines like anthropology and geography for research purposes. Using realist documentary conventions, academics sought to 'capture the reality' of people and their activities in naturalistic settings for subsequent analysis, or representation to others. In the 1960s and 1970s, however, in a radical departure from previous work, the technology was handed over to participants to do their own filming. One of the most famous examples of this approach is perhaps the work of Sol Worth and John Adair. In 1966, they taught a group of six inexperienced Navajo in Arizona to make and edit silent 16mm films. However, ultimately the purpose of the process remained the same: for academics to achieve a better (i.e., more objective and 'truthful') understanding of their subjects' lives through the products of their own labor. Their work did not try to change or improve the lives of the Navajo through the process.

In Australia in the mid 1970s, Eric Michaels took this approach a step further by integrating an action orientation to his work with rural aborigines. He didn't just want to understand their engagement with the technology, but worked actively to enable them to use it to tell their own stories on their own terms. Since Michaels' pioneering work, it has been mainly feminist researchers in anthropology and sociology who have sought to use video as an empowering and democratizing technology, and these uses are not dissimilar to PV within development, but maintain more of a research focus.

In some respect, the distinction between the use of PV in development or PV in research is less clear cut than the distinction between the emphasis on the process or the product. Ultimately, the use of PV depends on the purpose of people's involvement. If the purpose is to facilitate interaction, enable self expression, and achieve specific localized goals, then the process associated with making and engaging with the video is paramount, and the videotape produced may have little life outside of its immediate context. If however, the goal is to reach an external audience for educational purposes or to exert pressure to achieve a defined goal of change, then the videotape becomes more important than the process of its production. The different goals of the people involved as well as either the researcher or development facilitator will influence the precise form and orientation taken.

Current Uses of Participatory Video

Over the last 20 years, two parallel shifts have resulted in an increased use of PV around the world. The first relates to the rise of alternative paradigms to mainstream economic development and positivistic social science respectively. Increasing dissatisfaction with top down, hierarchical, and extractive interventions into people's lives by policy makers, practitioners, and academics have produced a range of different approaches to development and research. These more bottom up and people centered approaches seek to involve people as active agents of their own analysis and change processes emphasizing participation and representation. While varied, these approaches generally fall into the categories of participatory development and participatory action research. They seek to facilitate research and development by, for, and with the people concerned, not on or about them. For video, this has meant a shift in the locus of power and control from outside technical experts to processes that enable people to research and produce their own representations, either to facilitate self and group empowerment or community development.

The second shift relates to reductions in the size and cost of the technology involved. Since the 1990s, the increasing availability and affordability of digital video, computer, and Internet technology has fueled the growing use of PV within community development, academic research, and social justice activism. Prior to this, making films and videos was largely confined to Western academics, wealthy elites, and commercial filmmakers working with expensive and cumbersome equipment. Now, the relative ubiquity of video cameras (even in mobile phones) and laptop computer based editing software, plus increasing access to Internet distribution mechanisms mean that more people than ever before are becoming producers rather than consumers of their own video products.

PV is currently used in three main ways: as a tool to facilitate community development, as an action oriented research methodology, and as an advocacy or resistance tool.

Participatory Video (for Development)

The main use of PV builds on the experience of the Fogo Experiment discussed in the previous section. Also known as process video or the 'video as dialog process', PV is used as a tool to support community education, cultural identity and preservation, organization, and political participation.

Generally, participatory or process video does not involve commercial production. Rather, it involves 'ordinary' people as video directors, camera operators, interviewers, and subjects. The products are primarily oriented to those involved in the production and may not conform to standard Western filmmaking conventions. This doesn't matter however; as in many cases, it is the discussion of, and immediate feedback on, the video products, along with the sense of community and cooperation fostered through making the video that is considered most important. Where conventions may become more important, the goal is to use the video products within advocacy (also see below) to communicate key issues to people in positions of power as a vehicle for social or political change.

Participatory Video (for) Research

As within anthropology, video has been used within geographic research. Often it has been used to record interviews for the academics' subsequent analysis, or to record people's movements and interactions in particular spaces and landscapes – as a form of naturalistic inquiry. With a more interactive orientation, people have been supported to produce video diaries or to carry out video tours as part of wider projects with academic researchers.

More recently, with the rise of the so called 'participatory turn' in human geography since the late 1990s, and falling costs of technology, there is now a growing body of work in geography and allied fields engaging more participatory uses of video. Caitlin Cahill, for example, is facilitating a participatory research video documentary project with young people from Salt Lake City in the USA on undocumented students' experiences of racism at high school. Kay Haw in the United Kingdom has been exploring how PV can provide an alternative voice on youth risk and resilience associated with crime. In the Caribbean, Pamela Richardson has been using PV to understand more about sugar and agribusiness networks, and, in Uganda, Louise Waite and Cath Conn have been using PV in their work with young women around issues of sexual health. In Western Australia, Guy Singleton and his colleagues from the Ngalia people are exploring the role of PV and other ICT to develop effective cultural heritage management strategies to protect their sacred places/sites and develop sustainable livelihood options for their community. And finally, the author's own work with Geoff Hume Cook and members of Te Iwi o Ngaati Hauiti (a Maaori tribe in the central North Island of Aotearoa, New Zealand) represents perhaps the first discussions and analyses of PV's application in geographic research.

Video Advocacy

The third way in which video is commonly used is as a tool for advocacy because video can counter stereotypes and build intercultural understanding and solidarity. Advocates argue that it can circumvent issues associated with literacy and reach diverse audiences to mobilize people into action. The mobilization occurs because video elicits powerful emotional responses and connects audiences to the personal stories of those being represented. A number of organizations around the world now work effectively with this use of video (see **Box 1**).

Where the use of video advocacy differs from the other uses of participatory video is in the locus of decision making and control. In the majority of advocacy videos, production is discussed with the people involved and their consent sought to film. However, many also take place covertly because of the desire to expose human or environmental rights abuses. In addition, the subjects of advocacy videos are not usually the producers or editors. Rather, trained professionals ensure that the video products can reach and be understood by their intended audiences to effect change, a point the author returns to later. Of course, the subjects of the video may be empowered through the telling of their stories to change the conditions of their own lives, but this is often an indirect outcome of the work.

Box 1 Some organizations using video in their advocacy work

- Appalshop: United States of America
- CEFREC: Bolivia
- Chiapas Media Project: Mexico
- Drishti Media Collective: India
- INSIST: Indonesia
- Labor News Production: South Korea
- Self-Employed Women's Association: India
- Undercurrents: United Kingdom

Outside of these organizations, there is also a worldwide movement of radical alternative video journalism. Organizations like Altermedia International, the Guerrilla News Network, and Guerrilla Video Productions encourage people to shoot, edit, and share footage on the Internet as a form of activist resistance to the control of information and news by mainstream media conglomerates. These independent digital video producers or 'diguerrillas' tend to follow the ethic of: 'use less money, less time, fewer tools, and new rules', and are more anarchic in their orientation than the video advocacy organizations discussed above. They are focused on the video product rather than the process of the video's production. And, while they do often desire to effect wider social change, they do not work with a personal or community development orientation.

Process

In this section, the focus is more explicitly on the use of PV for research and/or development, rather than on video advocacy or guerilla video, to further explore the forms it can take. **Table 1** locates these different forms or applications within a participation continuum. Thinking more specifically about participation is helpful to consider the issues of power, ownership, and representation, which are at the heart of this work.

While not intended to be a hierarchy, the continuum is helpful to illustrate the different positions that can be occupied by community members, outside video producers (or researchers), and various audiences within any

Table 1 Participation models continuum and participatory video process

'Model' of participation	*Applicable video process*
Co-option	Outside video producers meet with key community members, decide on salient issues, and make their own program. Local community may see the final product, which is aimed at a national 'majority' or international audience.
Compliance	Community members conduct research. Outside video producers decide on salient issues and make their own program, with some community members in on-camera roles. Local community may see the final product but the target audience is not local.
Consultative	Community members share their thoughts and opinions about important issues, which outside video producers consider when developing their program. Local community may see the final product and be aware that it is meant to be shared with other communities.
Cooperative	Community members and outside video producers develop concept together. The outsiders make the video, generally with some community members in minor behind-the-scenes and/or on-camera roles. Local audiences generally see finished product but are not the target audience.
Co-learning	Community members and outside video producers develop concept and 'script' together. The production is made jointly with an outsider directing or facilitating the process. Community members are involved in a wide range of roles. Local audiences will see, and have access to, or a degree of control of, the product. It is shared with other communities.
Collective action	Community members determine the important issues, develop the story/script, and make the video themselves (any outsider involvement is within community-determined parameters). The purpose of the video is for use by the community to raise the status of issue(s) and to advocate for change within the community, or between the community and others that impact on it.

Adapted from DeNegri *et al.* (1998:4) with input from Geoff Hume Cook, pers.comm 2007.

Box 2 Generic outline of a participatory video process

Conceptualization and research

- Contact and establishment of relationship between outsiders (academic and/or video producers) and community members
- Negotiation of purpose, process, and outcomes (this may include a Memorandum of Understanding covering copyright/ownership of media produced)

Preproduction

- Negotiation of access to equipment, resources, funding
- Identification of, and negotiation with, participants to be trained
- Production timeline and planning
- Training of community members increased familiarity and confidence with equipment, concepts, and techniques

Production

- Filming of video footage by community members

Postproduction

- Training of community members
- Editing of video footage with and by community members

Screening, Dialog, and Feedback

- Footage screened (to participants and/or to wider community) and discussed
- Learning and exchange facilitated
- Possible change or action as a result of discussion
- Management of video products (dissemination, storage, access)

Then more production, postproduction, screening etc. Or preproduction

- Motivation to create more projects of benefit to the community
- Liaison with outsiders about future and ongoing involvement and so forth…

participatory use of video. Such positions reflect how power circulates. They enable researchers, practitioners, and community members to be more realistic about their relationships and the outcomes that these will both enable and constrain.

In general, PV follows a process similar to any video production process. The main difference is the time and energy invested in the development of relationships with the people who will become the community producers (and/or editors) and their training (**Box 2**).

For PV processes to be successful, sufficient time and energy needs to be invested in preproduction and post production stages. The production stage of the process is relatively short. Within participatory development applications of PV, the cycle of stages in the field is often shorter than represented in **Box 2**, as much of the initial conceptualization and preproduction work takes place between development agencies and the outside video facilitators. Within research applications however, the process in **Box 2** is more typical.

Issues

PV as a process for research and development is not without its critics. Four issues are discussed here, which require more analysis and theorization if their power effects are to be more adequately understood.

Tyranny

Video as a method or even as a methodology rarely stands alone. It is used as part of larger participatory research or development initiatives. As such it is susceptible to the criticisms about the tyranny of participation associated with the role of facilitators, the disciplining effects of participatory methods, and the desire for group consensus. In addition, within PV projects facilitators must also manage technical aspects associated with camera operation, sound quality, editing, and storytelling. Not surprisingly, PV is only as good as its facilitation, and PV's facilitation can be demanding.

Resources

Video can also be expensive to use. While the costs of camcorders are decreasing, the other essential equipment associated with sound recording, editing, and screening means that initially, at least, a considerable investment in technology is needed. Furthermore, within PV where the goal is often to affect some kind of sustained change, adequate time is needed to develop the capacity for community members to film and edit their own footage. If the aim is to produce short videos for immediate feedback and discussion, this may limit community members' participation in hands on editing. If the aim is a larger scale documentary project, then ways to sustain and recompense community members' time and energy are essential. As with any project, the level of investment in resources, and the ownership and control over those resources can have huge effects on the power relations at work and the outcomes possible.

Ethics

Working with video involves layers of ethical considerations. Most of these are not yet adequately addressed by university ethics review boards, which predominantly concern verbal and written research approaches. In PV, questions arise about the ethics of engagement (i.e., the levels of consent required at different stages in the process) and the ethics of representation (i.e., the facilitator/researcher's and/or community members' skills at using visual images ethically and artfully). Visual images are open to different interpretations and uses. With the proliferation of web based domains such as on YouTube, their original intent or message once screened, may be re edited and re presented in ways that create new ethical issues over which the original producers have no control.

Audience

PV is only meaningful if there are audiences for the products produced. But how audiences are defined will influence what they are expected to do as a result of watching. Within video advocacy approaches, audiences are usually external to the community involved in filming or being filmed. The products must therefore conform, to some extent, to the dominant forms of cultural and visual literacy if they are to be 'read' appropriately. Within PV however, participants themselves or their surrounding community(ies) may be the intended audiences and issues of cultural and visual literacy may be less important than the process of production and discussion.

In either case, questions arise about the relationship between the communicational intent of the video product(s) and its attendant production process, and the ability of the intended audience(s) to frame the images and messages appropriately so as to access their intended meanings. Video doesn't capture or record reality; rather, it constructs or produces partial realities mediated by culture, ideology, and subjectivity. Aspects of the process of production, its representation, and translation are intertwined and informed by the discourses within which audiences are positioned at the time of viewing.

Benefits

While there are these ongoing issues associated with the use of PV, it has been applied successfully across a wide range of cultures and contexts. It has also been especially powerful in communities where there has been less exposure to commercial mainstream audio visual media, and where participant groups have been marginalized or disenfranchised. In the final part of this article are outlined three reasons why, within the contexts of human geography, PV is worthy of closer scrutiny.

Critical Consciousness

PV and the act of putting someone behind, instead of in front of, a camera momentarily detaches them from their immediate surroundings and effectively provides them with a different way of viewing and engaging their environment. Such an experience increases their ability in Paulo Friere's terms to 'read the world' by developing a more critical consciousness. Their development and application of critical thinking about representations within mainstream media associated with their community expands the notion of literacy. Participants develop the ability to express their understandings of 'reality' in ways that go beyond the written word and which produce new meanings and resist others.

Voice

PV changes the voice of research and development. By emphasizing more democratic and horizontal forms of communication, it can legitimize people's voices and enable them to be heard and taken seriously by others in positions of institutional authority. In addition, its aim is to make texts (and the knowledge they convey) more accessible, and challenge hegemonic discourses by opening up space for alternatives. PV communicates information that is not always easily translatable into the verbal, and the combination of image and text can be more subversive than text alone. Its political effects can provide new opportunities for theory building and action engaging the emotions, which in itself often challenges the *status quo*.

Social Construction of Knowledge

Using video makes the construction of knowledge more explicit. Once positioned behind a camera, a person be comes acutely aware of what s/he chooses to frame and privilege. The often taken for granted connection be tween 'ways of seeing' and 'ways of knowing' is called into question. When editing, the social and cultural codes drawn on to produce meaningful images become more evident through negotiations about the length and dur ation of shots, their sequencing, and associated sound track or voice over. In addition, cultural and moral dimensions of knowledge are negotiated with respect to the ethics of representation and issues of translation required by the audience. The whole process increases the practice of self reflexivity, which in turn forms an iterative cycle of action and reflection and deepens participants' critical consciousness and ability to affect change in their lives.

Conclusion

PV is an exciting vehicle for communication, connection, and change. Varied and adaptable in its application, it has been successful in informing constructive community research and development in a range of cultural contexts. While questions remain about the intersections and political effects of its process, representation, and trans lation for diverse audiences, PV serves as a useful tool with which to explore the interplay of visual, verbal, and written knowledges. It also enables a more explicit engagement with the ethics of research and development and the practice of self reflexivity by those involved.

See also: Activism; Auto-Photography; Community; Critical Geography; Cross-Cultural Research; Empowerment; Ethical Issues in Research; Movies and Films, Analysis of; Participation; Participatory Action Research; Postdevelopment; Representation, Politics of; Visuality.

Further Reading

Chavez, V., Israel, B., Allen, A. *et al*. (2004). A bridge between communities: Video making using principles of community based participatory research. *Health Promotion Practice* 5(4), 395 403.
Falzone, P. (2004). Transcendent ethnography: Designing an action research approach to ethnographic film within cultures of conflict. *Action Research* 2, 326 344.
Gregory, S., Caldwell, G., Avni, R. and Harding, T. (eds.) (2005). *Video for Change: A Guide for Advocacy and Activism*. London: Pluto Press (in association with Witness).
Holliday, R. (2000). We've been framed: Visual methodology. *Sociological Review* 48(4), 503 522.
Hume Cook, G., Curtis, T., Potaka, J. *et al*. (2007). Uniting people with place through participatory video in Aotearoa New Zealand: A Ngaati Hauiti journey. In Kindon, S., Pain, R. & Kesby, M. (eds.) *Participatory Action Research Approaches and Methods: Connecting People, Participation and Place*, pp 160 169. London: Routledge.
Kindon, S. (2003). Participatory video in geographic research: A feminist practice of looking? *Area* 35(2), 142 153.
Lunch, N. and Lunch, C. (2006). *Insights into Participatory Video: A Handbook for the Field*. http://www.insightshare.org/training book.html (accessed in June 2008).
Mayer, V. (2000). Capturing cultural identity/creating community: A grassroots video project in San Antonio, Texas. *International Journal of Cultural Studies* 3(1), 57 78.
Parr, H. (2007). Collaborative film making as process, method and text in mental health research. *Cultural Geographies* 14(1), 114 138.
Pink, S. (2007). *Doing Visual Ethnography* (2nd edn.). London: Sage.
Robertson, C. and Shaw, J. (1997). *Participatory Video: A Practical Approach to Using Video Creatively in Group Developmental Work*. London: Routledge.
White, S. (ed.) (2003). *Participatory Video: Images That Transform and Empower*. London: Sage.

Relevant Websites

http://www.c4c.org
Communication for Change (C4C).
http://www.guerillavideoproductions.com
Guerilla Video Productions.
http://insightshare.org
Insight Video.
http://drc citizenship.org
Participatory Video: A Useful Lens for Research?

Patriarchy

C. J. Nash, Brock University, St. Catharines, ON, Canada

Glossary

Second-Wave Feminism A historical moment in Western feminism in the 1970s and 1980s when activism and scholarship focused on politicizing systemic/structural inequalities based on gender differences.

Third-Wave Feminism A historical moment in Western feminism in 1980s and 1990s where activism and scholarship turned to postmodern and post-structuralist ideas to deconstruct identity categories.

Introduction

Patriarchy, traditionally defined, refers to a system of social relations in which there is gender inequality between socially defined men and women. This system of inequitable social relations is embedded in the political, social, cultural, and economic institutions of a society as well as in, and through, private domestic relations. In these systems of gender inequality, women find themselves disadvantaged in various ways within social relations and organization. In patriarchal societies, women are collectively excluded from political, social, and economic positions of power; women find themselves paid less well for work of equal value; and women are more likely to experience poverty and unequal access to resources, goods, and services. While individual women may experience success in various spheres, women as a distinctive social group are generally disadvantaged in these ways.

Earlier sociological and political definitions of patriarchy focused on domestic social relations organized around the law of the father and the social control that men, as heads of households, have over their wives and daughters. Within feminist scholarship, patriarchy has been understood more broadly as the system in which men as a group are constructed as superior to women as a group and as such have authority over them. In some instances, scholars have distinguished between 'paternal' patriarchy to refer to forms of private, household organization run by the father and 'fraternal' patriarchy to refer to the domination of women in public civil society. For geographers, this distinction mirrors the division between private domestic space and public social space. More generally, patriarchy is conceived of as a broadly based system of social, legal, economic, political, and cultural structures and practices, which position men as the dominant social group and as able to marginalize, and exploit women.

Scholarly research suggests that current Western forms of patriarchal relations are constituted within Western notions of modern science that emerged during the Enlightenment. The notion that science is a universal, value free system for producing knowledge and acting as a liberating force for humanity is critiqued as a system that excluded women from that emancipatory project. Modernist notions about rational thinking and objectivity require a dichotomy between men and women, nature and nurture, and the rational and emotional, in ways that support or operate in conjunction with masculine and scientific domination over nature and women. Modern science is portrayed as a consciously gendered patriarchal activity and project, grounded on the subjugation of women and nature. Scholars link these patriarchal representations of knowledge, rationality, and the objective subject to the social and political needs of the newly emergent industrial capitalist system in Western societies. What such thinking highlights is how patriarchal relations can be conceptualized as differently interwoven into social systems and structures depending on specific historical and cultural eras.

The term 'patriarchy' first gained prominence in feminist scholarship in the 1960s and 1970s, when feminists developed the notion of gender roles and gender relations as categories of analysis. Gender as an analytic category is conceived of as distinct and separate from biological sex and as a concept that could undermine the claims of sexual differences between men and women in Western society. Gendered behaviors, expectations, practices, and roles are seen as reflective of social and cultural processes and are not biological 'givens'. The binary categories of man/woman; masculinity/femininity; and male/female are seen as key factors in the organization of social life. These categories are ordered hierarchically with women constructed as inferior to men, and with those attributes seen as 'feminine' undervalued. These constructed, binary differences that are embedded in Western systems of thought and practice, as well as our social structures and institutions and social relations, provide powerful explanatory tools for analyzing women's inequality across distinctive historical and cultural eras.

Patriarchy and Capitalism

Within feminist theorizing, the concept of patriarchy surfaced in 1960s socialist and Marxist feminism theorizing the linkages between patriarchy and capitalism in advanced industrial societies. Feminists explored the interrelationship between gender and class within the spheres of production and reproduction. Explanations for changes in gender roles and relations in the modern era were located in major structural changes that occurred due to the shift to industrial capitalism. Feminist geographers, influenced by socialist/Marxist feminism, argued that women's inequality could be explained as a result of the social and spatial separation of production and reproduction and women's containment in the private, domestic sphere. However, feminists disagreed over the nature of the relationship and linkages between Western industrial capitalism and patriarchy. Some argued that the separation of the public and private spheres of work and home was driven by capitalism's need for this particular social order and patriarchy was a secondary structure arising out of capitalist social relations. For others, capitalism and patriarchy were conceptualized as distinctive but entwined systems of structural relations. Developing a coherent theoretical framework is still a work in progress.

For feminist geographers, the work of Sylvia Walby has been particularly influential for thinking through the various formulations of patriarchal relations. Walby argues that patriarchal relations in advanced industrial societies operate through, and are maintained within, six analytical structures in which men dominate and exploit women. These six structures are interconnected and operate in different ways in different places and circumstances. The six structures are household production; patriarchal relations in waged work; patriarchal relations in the state; male violence against women; patriarchal relations and sexuality; and patriarchal relations and cultural institutions. Scholarship also acknowledges that patriarchy takes different forms in different places and contexts and that its operation is dependent upon multiple and interrelated understandings of age, class, sexuality, ethnicity, and other markers of identity. Within patriarchal relations, various formulations of power are in circulation such that some individuals are complicit in their oppression and may find pleasure and security in the normative expression of gender roles, behaviors, and practices.

In contrast to Walby, Robert Connell argued that the notion of patriarchy proposed forms of male power that were too monolith and coercive. He proposed the concept of a 'gender regime' which some scholars regard as providing a more nuanced conceptualization of how dominant regimes not only dominate and oppress but can also extract the willing consent of those marginalized groups. Further, he argues that dominant gender regimes, although seemingly stable over time, are actually made up of a number of conflicting regimes that challenge dominant assumptions and bring about change and, potentially, relief from domination. Scholars argue that this is a more optimistic conceptualization of gendered social relations, allowing for a recognition of the pleasure and satisfaction some people experience in particular gender regimes. In addition, it opens up the possibility for women's agency in resisting and negotiating patriarchal relations and thereby creating opportunities for women's greater participation in civil society.

The concept of patriarchy has been criticized for being too universal and overarching in its conceptualization. Early work on patriarchy positioned it as an all encompassing feature of relations between men and women and as an ahistorical and acultural structured set of inequalities, seeming inflexible and resistant to change. Given its myriad formulations, patriarchy is perhaps more helpfully understood as neither unitary nor self contained and as incapable of being separated, analytically, from the evolving social structures in which it is embedded.

Currently, arguments positioned within a postmodern intellectual framework (advanced by so called third wave feminists) critique universal structural explanations contained in concepts such as patriarchy, arguing that it lacks a nuanced and complex appreciation for the diverse forms that oppression may take. Postmodernist arguments suggest that, in deconstructing the categories of men and women in order to demonstrate their fragmented and unstable constitution, the concept of patriarchy fails when the binary categories on which it depends are dismantled. Despite these more recent arguments cogently critiquing patriarchy as an analytical and conceptual tool, patriarchy continues to be a useful theoretical perspective for highlighting structured inequalities in gendered spatial relations and social organization.

Patriarchy and Geography

Western feminist geographers have found the concept of patriarchy useful for conceptualizing gender relations across a variety of scales. The concept of patriarchy has been used to conceptualize the heterogendered and heterosexual regulation of women's bodies, the division between public and private spheres, women's access to the public spheres of waged work and their presence in civil society. Patriarchal relations operate at a variety of interlocking social and material scales.

Patriarchy and Bodies

Feminist geographers argue that the body, as a place, is the site and locus of the material and embodied individual whose subjective understanding of the self is constituted through social relations. Within patriarchal relations, individuals understood as biologically female have normalized expectations of dress, manner, gestures, behaviors, interests, and desires such that women come to understand themselves as gendered and sexualized sub jects. Within patriarchal power relations, women find themselves subject to bodily constraints not only through expectations of physical behavior but also constraints imposed through expectations about what that body is capable of and where that body belongs, both literally and figuratively.

Because of women's ability to give birth and breast feed, scholars argue that women are then constituted as best suited to raise children and care for the family and are thereby seen as most properly found in the private, domestic spaces of the home. Historically and culturally specific political and social institutions such as marriage, welfare provisions, income tax laws, and daycare policies arguably work to ensure women's normalized restriction to the domestic sphere and the patriarchal authority of the male head of the household. Women's bodies are constituted as being monogamously available for one man, ensuring the legitimacy of children for inheritance purposes. Women's bodies are also points of contestation in debates about abortion and national issues around fertility rates and population reproduction.

Women's bodies are also constituted as out of place in the public sphere of waged work and civil service. Employment opportunities are gendered with some forms of work seen as being more suitable to women rather than men. In addition, institutions such as the military and, more particularly, combat positions, are constructed as expressly masculine pursuits and beyond the ability of women. Arguments are often made that women are not physically or emotionally capable of killing; are in need of defending by the nation's men; and are, in their embodied presence, a dangerous distraction to men waging war.

Public and Private Spaces

Feminist geographers use the concept of patriarchy to theorize the connections between women's household or domestic labor in the private sphere and the public sphere of waged work within advanced industrial capitalism. The distinction between productive and reproductive work is portrayed as essential to the smooth operation of capitalism. Women's reproductive labor in the home reproduces and maintains the health and well being of the productive labor force on both a daily and a generational basis. Within capitalist relations, women are exploited both by the capitalist system (collectively benefiting men) and by individual men who are able to appropriate women's labor for their benefit. Others argue that in advanced industrial capitalist societies, women's labor in the home helps to stabilize patriarchal structures through their fulfilling of social roles – defined as wife and mother. Women help stabilize the economy through their role as both producers and as low waged laborers themselves in the labor market. Patriarchal gender roles ideologically confined women to the private sphere where the gendered characteristics of nurturing, car egiving, domesticity, and child rearing are seen to belong. With the family increasingly targeted as a market for mass produced goods, women as consumers of household goods and services made the 'home' a unit of con sumption, as well as one of reproduction. The spatial dimensions of this capitalist system of social relations ensure that women's domestic labor is confined to the private sphere of the home and is considered outside of the purview of public civil society.

This confining of women to the private sphere of domesticity through combinations of capitalist and patriarchal relations has led feminist geographers to focus on the spatial constraints imposed on women through these social relations. This includes research on the emergence of the residential subdivision and the single family dwelling as dominant residential land forms in Western cities. Women living in these spatial arrange ments often find themselves limited in their ability to travel beyond the confines of these neighborhoods due to restrictive public transit options, limited child care op tions, and few opportunities for waged work outside the home.

Feminist geographers argue that not only do women find themselves constrained within the private sphere, but patriarchal social relations also construct and maintain gender and gender relations in public spaces. Patriarchal social relations are seen as operative in the constitution of public spaces through both formal and informal discip linary systems imposing expectations on the way women dress, behave, and act in public spaces. Women are often expected to use certain urban spaces only during certain times of the day and to limit their occupation of public space in the nighttime unless accompanied by male escorts. Certain types of bars, restaurants, taverns, and sporting and leisure events are also often seen as places more suitable for men. Patriarchal relations also operate through the institutional regulatory regimes of city planning, administration, and management which con stitute public spaces as 'safe' or appropriate for women through street lighting, landscaping, and various forms of surveillance. More recent work has considered how contemporary public surveillance by video and still camera is experienced differently by men and women. While women may paradoxically feel safer due to the various forms of surveillance, others argue that sur veillance in urban space is mistrusted by women who are uncertain that they can rely on those behind the camera because of the reproduction of patriarchal power through surveillance.

Research on the urban sex industry also demonstrates the contradictory ways in which women experience public space. For women, public visibility subjects women to the male gaze and objectification that often results in a commodification of women's public embodiment. However, scholars argue that sex work can be seen as a new form of labor whereby women can claim full sexual citizenship and the opportunity for sexual emancipation in ways that resist male domination. Public visibility and asserted expressions of female sexuality challenge oppressive heterosexual norms and women's visibility and public space. Scholars such as Nancy Duncan claim that sex workers' use of public space is a form of resistance similar in kind to that undertaken by other sexual minorities, who use visible public presence to promote sexual diversity. Through other activities such as Take Back the Night Marches, women are able to challenge the patriarchal nature of public space by using that space to make claims for sexual and gender rights.

Citizenship

The concept of patriarchy has also been used to explore the nature of the relationship between gendering and the state. Scholars argue that the state is patriarchal in structure and organization, given its support of male privilege through its various institutional structures. The question of the relationship between patriarchy and capitalism also arises with some scholars arguing that the state is principally capitalist and secondarily patriarchal. Others assert that capitalism and patriarchy are dual and overlapping systems operative around and through the state. Claims that the state is patriarchal, arguably, presents the state as uniform and monolithic when the state can perhaps more properly be described as a series of arenas in which patriarchal relations may be in circulation in particular ways.

Feminist geographers argue that men's and women's knowledge and experiences of urban spaces and their ability to occupy public spaces is differently constituted. In terms of being a 'citizen', scholars argue that having meaningful civil, political, and social rights and participation has a spatial dimension. The ability to participate in the affairs of the state and in the activities of citizenship require access to the material and symbolic public sphere, and if men and women have access to and experience these spaces differently, then being a citizen is differently constituted as well. Given that citizenship is defined by either inclusion in or exclusion from the national space, spatial accessibility is fundamental to notions of citizenship and equality. Arguably, patriarchal spatial divisions between men and women, operative through the notions of the public and private sphere, work to symbolically and materially exclude or limit women from participating in the public sphere of citizenship and civil society.

Nationalism and the State

At the scale of the nation and the nation state, feminist scholarship demonstrates how nationalism is not only gendered but racialized and sexualized as well. Scholarship on colonialism and imperialism, for example, demonstrates how English national identity in the nineteenth century was grounded in the ideal of a white patriarchal family. This ideal served to distinguish and distance the white, colonial explorer from the distinctive, matrilineal, extended kinship systems and unstable gender relations often in operation in those lands brought under colonial rule. Through, imperialism and colonialism, scholars argue that the British, white patriarchal ideal about family, sex, and gender roles circulated globally and imposed this normalized standard on colonial groups often replacing women centered social structures. Feminist scholarship demonstrates that in various colonial locations, culturally specific, patriarchal structures incorporated Western colonial notions of patriarchy, and thus require careful historical research to chart the emerging distinctions between places.

Feminist research has also focused on how nationalistic and patriotic narratives draw on various forms of patriarchal social relations in constituting national mythologies. These include representations of the nation as female and deserving of male protection through military might and warfare; the need to preserve the 'purity' of the nation through protection of women from the foreigner; and the construction of the military as a singularly male institution engaged in the protection of the feminized nation.

Globalization

While patriarchy has been critiqued by some feminist scholars for being too universal and overarching as a conceptual tool, patriarchy has remained a fruitful analytic for many non Western feminist scholars. In a number of developing countries, grassroots feminist movements and national and international development organizations have increased the focus on women's roles in international economic development. With increased financial and organizational resources, these various organizations have struggled to further women's causes and improve the lot of women within distinctive social, political, and cultural systems. Research has demonstrated that simply adding women to existing institutions and systems is problematic when these institutions themselves are constituted within and through specific patriarchal social relations. Scholars argue that further work needs to be done on the particular historical, political–economic relationships between capitalism and patriarchy, within the exploitative colonial relations underpinning current processes of globalization and the international division of labor. For some scholars, capitalism and patriarchy are theorized as inseparable with capitalism underpinning historically and culturally specific patriarchal social relations in advanced industrial societies. Both capitalism and patriarchy, arguably, need to be interrogated as an integrated system working to maintain women's exploitation and oppression. In these contexts, patriarchy remains a powerful analytical tool for conceptualizing particular historical and societal dimensions of women's oppression. Such a perspective demonstrates how distinctive patriarchal systems emerge and develop differently in particular times and in different geographical regions. Seen this way, capitalism can be conceptualized as the latest manifestation of patriarchal relations. In some developing countries, this perspective supports the argument that women's experiences in contemporary globalization cannot be explained by reference to some precapitalist system of patriarchal social relations that has been transformed by contemporary capitalism. On the contrary, scholars argue that as capitalism struggles for dominance, its various transformations both rework and sustain patriarchal social relations in new and unforeseen ways.

Within the context of globalization and international development, Third Wave feminism emerged as a powerful voice in the late 1970s and early 1980s. Third Wave feminists argued that 'empowerment' was central to women's emancipation. Transforming economic, social, and political structures at national and international levels was seen as being crucial for achieving women's equality. Scholars argued that women's empowerment required systemic changes to the specific social and cultural forms of patriarchy in different places and locales. To effect change, an analysis of the place specific family, community, market, and state structures that supported and stabilized patriarchal norms was required. Women needed to gain access not only to economic and material resources, but also needed to work for the dismantling of the broader social and political structures that supported patriarchal hierarchies. From this perspective, many mainstream and state led developmental interventions now tend to focus on local gender relations and the roles that women, in particular, play in certain societies. As a result, women's access to resources, and their active and visible participation in development processes, becomes a central plank in development efforts. Despite the modest success of these approaches, scholars also need to consider how women's obligations to family and community are constituted within patriarchal relations in ways that constrain women's ability to effect change. Women's participation in the public sphere of work may cause considerable stress and difficulty in the private domestic sphere, in ways that may be dangerous or debilitating. In working towards women's participation in development processes, attention must be paid not only to the broader economic structures but to the embedded patriarchal relations that order and regulate women's participation and possible emancipation.

Men and Patriarchy

Other scholars have also considered the conceptual intersections of patriarchal social relations and men and masculinity. Arguably, if patriarchy is conceptualized as a set of interrelated structures supporting men's exploitation and oppression of women, then masculinism, as Brittan argues is "the ideology that justifies and naturalizes male domination and as such is the ideology that justifies and naturalizes patriarchy." Formulations of masculinity are conceptualized as working to ensure that men are constituted within and are the beneficiaries of Western patriarchal systems that assert male privilege. Yet despite men's subjective constitution and positioning within patriarchal systems, scholars contend that men are in a position to understand and challenge the essentialist and narrow possibilities provided for the masculine subject within patriarchy. Men need to become aware and acknowledge male privilege and to find ways to work against and potentially dismantle supporting political, social, and economic structures.

See also: Body, The; Emigration; Masculinities; Parenting/Motherhood/Fatherhood; Public Space.

Further Reading

Connell, R. (1995). *Masculinities*. Cambridge: Polity Press.
Duncan, N. (ed.) (1996). *Bodyspace*. London: Routledge.
Foord, J. and Gregson, N. (1986). Patriarchy: Towards a reconceptualization. *Antipode* 18(2), 186 211.
Garry, A. and Pearsall, M. (1996). *Women, knowledge and reality: Explorations in feminist philosophy* (2nd edn.). London: Routledge.
Gordon, A. A. (1996). *Transforming capitalism and patriarchy: Gender and development in Africa*. Boulder, CO: Rienner.
Herbert, S. (2001). 'Hard charger' or Station Queen: Policing and the masculinist state. *Gender, Place and Culture* 8(1), 55 71.
Mackenzie, S. and Rose, D. (1983). Industrial change, the domestic economy and home life. In Anderson, J., Duncan, S. & Hudson, R. (eds.) *Redundant spaces? Social change and industrial decline in cities and regions*, pp 155 200. London: Academic Press.
McDowell, L. (1999). *Gender, identity and place: Understanding feminist geographies*. Minneapolis, MS: University of Minnesota Press.
Mies, M. (1986). *Patriarchy and accumulation on a world scale*. London: Zed Books.
Moss, P. (ed.) (2002). *Feminist geography in practice*. Oxford: Blackwell.
Nelson, L. and Seager, J. (eds.) (2005). *A companion to feminist geography*. Oxford: Blackwell.
Pateman, C. (1989). *The disorder of women*. Standford, CA: Standford University Press.
Townsend, J., Zapata, E., Rowlands, J., Alberti, P. and Mercado, M. (1999). *Women and power: Fighting patriarchies and poverty.* London: Zed Books.
Walby, S. (1990). *Theorizing patriarchy*. Oxford: Blackwell.

Relevant Websites

http://www.africabib.org
African Women's Bibliographic Database.

http://ec.europa.eu
European Union, European Commission, Employment, social affairs & equal opportunities, gender equality.

http://www.emporia.edu
Gender in geography bibliography.

http://www.un.org/womenwatch
United Nations Inter agency network on women and gender equality.

Peasant Agriculture

J. D. van der Ploeg, Wageningen University, Wageningen, The Netherlands

Glossary

Modernization Process The multidimensional restructuring of agriculture toward highly specialized, large-scale, intensive, and market-oriented production.

Objects of Labor The organic resources that farmers work with and reshape.

Peasant Condition The overall position of peasants within a particular society.

Repeasantization The twofold process through which the number of peasant units increases and/or the peasant nature of existing units is augmented.

Technical Efficiency The ratio between resources employed and productive output.

Modernization and the Demise of Peasant Agriculture

Since the large modernization programs of the 1950–80 period, which had a wide ranging impact on agricultural production systems all over the world, peasant agriculture has been increasingly regarded as irrelevant and/or due for extinction. Many programs have promoted depeasantization. These include the Mansholt reforms in the former European Economic Community (now the European Union) and the Green Revolution schemes applied in Africa, Asia, and Latin America. Land reform processes, credit schemes, and other state induced changes that eliminated typical 'peasant crops' (e.g., coffee in Brazil) to make room for other crops (like soybeans) that are produced by larger enterprises, all contributed to this process. These programs have redefined, reshaped, and reordered the technicalities of farming. They have also aimed to create a new professional identity. It was assumed that the peasantry would be transformed into or replaced by new agricultural entrepreneurs and corporate farmers. This would move agriculture beyond the backwardness and stagnation assumed to typify peasant agriculture. Peasant farming was seen as having no future because of the technical constraints inherent in its resources. Consequently, external interventions were seen as necessary in 'getting agriculture moving', to borrow the title of a well known manual from the 1960s. Farming strategies oriented toward self subsistence, survival, and equilibrium were to be replaced by new strategies that centered on profit maximization and expansion. There equally was to be a shift from partial to full market integration, and a change from local cultural repertoires to rational decision making. All these changes were seen as important cornerstones of the transformation of agriculture.

The modernization of agriculture (always characterized by a central role of the state) was strongly influenced and ordered by modernization theories which then held great sway (and sometimes by Marxist approaches that equally saw the peasantry as a main obstacle for development). The close intertwining of theory and practices (the latter being inspired and informed by the former, the former increasingly a reflection of the latter) resulted in a remarkable path dependency that ultimately backfired on both theory and practice. The many indications that the newly emerging agricultural entrepreneurs were far from textbook applications of *Homo Economicus* as stipulated in agrarian economics and the 'early adoptor', specified in the diffusion of innovations tradition, did not feed back into theory; they were perceived as being just anomalies. The same occurred with the findings that modernization, as an empirical phenomenon, was far less unilinear and complete than should be the case, according to reigning theories.

In short, modernization theories heralded a transformation that implied, both theoretically and empirically, a reallocation of the peasantry and peasant agriculture to remote times and places – to the past and to the periphery. In the modern world, there was no place for peasants and peasant agriculture – theoretically at least.

There is little point in discussing the validity of modernization theory, or whether it represented a true picture of reality. What is crucial is that it was made to come true. Yet, modernization has been a highly differentiated process that has resulted in unequal, and often unexpected, outcomes. Three of these are of particular interest. First, many expressions of modernization (as, e.g., land reform processes and irrigation schemes) turned out to be broken promises. They did not function as assumed and often impacted in a negative way upon the affected agrarian sectors, as well as on wider society. Second, it is proved to be increasingly difficult (if not impossible) to explain a highly modernized agriculture, such as those existing within Western Europe, and the farmers who operate within it, solely in terms of agrarian entrepreneurship, that is, as a definitive adieu to peasants and peasant agriculture. Third, among an economically active population of 2.6 billion worldwide, a little more than 50% (i.e., 1.35 billion) are engaged in farming, and a large majority of them are peasants. Together, these observations raise new questions that the (still dominant) modernization paradigm is barely able to address. For instance, in many areas in the world, farming is a major employer and this is one of its principal functions, second only to the need to produce good food and maintain the resource base. Despite this, the policies that draw on modernization theory and its assumptions about entrepreneurship still look to cut farm labor to the bone – and then cut it again.

The Peasantry of the Third Millennium

Although the relative weight and the different interrelations vary considerably, nearly all agricultural systems in today's world consist of three different politico economic arrangements (see **Figure 1**). These are capitalist or corporate farming (in which the wage–labor relationship is central), entrepreneurial farming, and peasant farming. The main difference between the latter two forms is that peasant farming is strongly based on available ecological capital (especially living nature), while entrepreneurial farming is consistently moving away from nature. Artificial growth factors replace natural resources and the availability of (or dependence upon) financial capital becomes a main feature. This allows for economies of scale and rapid (albeit often partial) increases in productivity. Peasant farming also needs to be distinguished from other rural activities. Compared to the landless, peasants maintain a self controlled resource base, although there are evident variations in its dimensions. Compared to other rural professionals, peasants are special in that they are involved in co production, that is, the ongoing interaction of man and nature and the associated transformation of both natural and social resources.

These interfaces highlight and define the distinctiveness of peasant agriculture, which is grounded on the use of a self controlled resource base in which nature is central and is used for co production. The results obtained belong directly to those involved in production, although gender, age, and ethnicity might make for inequitable internal distributions.

Different politico economic arrangements imply different modes of farming. That is, they translate into different material configurations, for example, different breeds and selection strategies, a different layout of the fields, and a different coloring of the landscape. The different modes of farming also translate into different relations with rural society and the rural economy. The associated development trajectories contrast considerably. A proper understanding of the multiple differences that delineate peasant agriculture from the corporate and entrepreneurial forms requires a more detailed specification of the peasant condition.

The Peasant Condition

The overall position of peasants in wider society can be summarized in the concept of peasant condition. Peasant agriculture (or, the peasant mode of farming) originates from and is embedded in this condition. The peasant condition induces and reproduces the central characteristics of peasant agriculture. Without the former, the specificities of the latter cannot be properly understood.

The peasant condition consists of a context characterized by dependency relations, marginalization, and deprivation, which is countered through a struggle for autonomy (see **Figure 2**). Despite the many differences that exist between agriculture in the developed and developing worlds, it is important to note that both are subject to high levels of dependency. In developed countries, this occurs through the squeeze on agriculture, regulatory schemes, and the power of agribusiness. The forms and mechanisms of dependency, and the associated deprivation, marginalization, and insecurity,

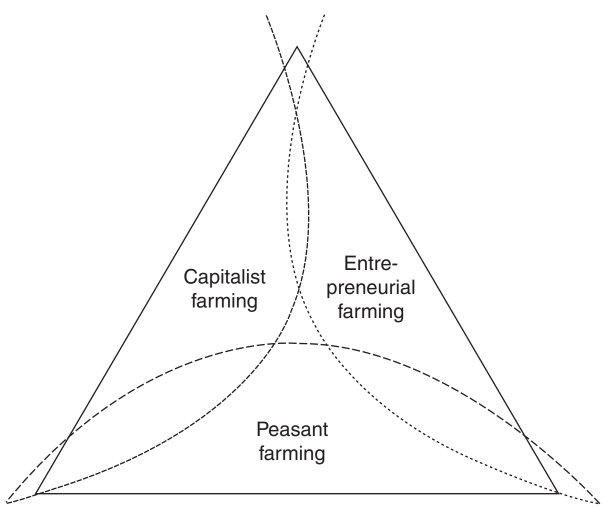

Figure 1 The basic ingredients of agricultural systems.

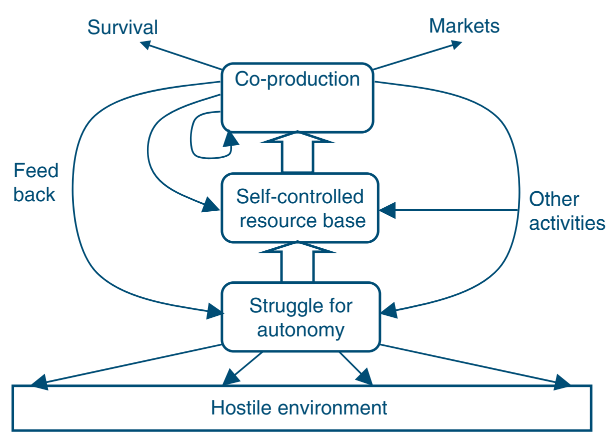

Figure 2 Choreography of the peasant condition.

might differ but farmers in both locations are confronted with a hostile environment. The resulting struggle for autonomy aims at, and materializes into, the creation and development of a self controlled resource base. This resource base embraces both social and natural resources (knowledge, networks, labor force, land, cattle, irrigation channels, terraces, manure, crops, etc.). Land forms a central pillar of this resource base, from a material as well as a symbolic point of view. It represents the basis of a certain degree of independence. It is, as it were, the rock from which the hostile world is to be faced and coun tered, hence, the centrality of land in the many peasant struggles of the past and present. In turn, this resource base allows for different forms of co production between man and living nature. Co production (i.e., the process of agricultural production) is molded in order to fit, as much as possible, the interests and prospects of the peasant family and, in turn, co production interacts with the market. A variable part of production will be sold, and another part used for the reproduction of the farm and the peasant family. Thus, co production allows, directly and indirectly, for survival and for future prospects, and also feeds back into, and strengthens, the resource base, thus improving the process of co production. This mostly occurs through qualitative improvements: making land more fertile, breeding more productive cows, selecting the best potato seedlings, making better storage facilities, enlarging knowledge, making roughage production meet the needs of the herd, etc. Such qualitative improvements feed back positively into co production; they also (might) translate into enlarged autonomy and thus reduce dependency. Depending on the particularities of the pre vailing socioeconomic conjuncture, survival and the de velopment of the self controlled resource base might be strengthened through engagement in other nonagrarian activities. Taken together, these relations are arranged into a strategically ordered flow of activities through time.

The pattern of the interrelations between peasant agriculture and markets plays an important role within this flow. In this respect, it is crucial to distinguish between input markets and those for outputs.

Market Linkages and Distantiation

From an analytical point of view, farming consists of three interrelated and mutually adapted processes (**Figure 3**). These are the mobilization of resources, the conversion of resources into (end) products, and the marketing and reuse of the end products. The first and the third process, and increasingly the second one assume, and *de facto* imply, relations with the markets; however, these can have completely different patterns.

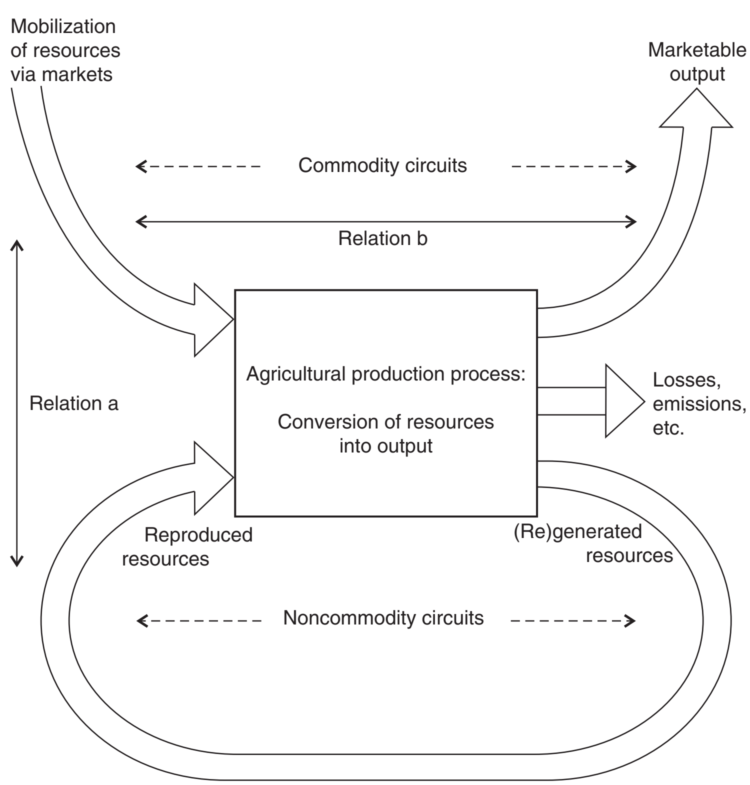

Figure 3 The basic flows entailed in farming.

Resources can be mobilized through different markets or, equally, can be produced and reproduced within the farm. This applies to all social and material resources, whether cows, feed and fodder, fertilizers, seed, labor, knowledge, working capital, or buildings. They might be obtained through market transactions and enter the production process as commodities. They might equally be produced and reproduced within the farm unit itself, or be obtained through socially regulated exchange. Those resources that cannot be produced physically on the farm (like heavy machinery) can be acquired in different ways: either with one's own resources (e.g., savings) or through creating dependency by borrowing to raise the money. Thus, the specific social history of objects can make a very real difference.

Peasant agriculture tends to be mainly based upon noncommoditized resource flows. This is reflected in (and might be measured through) relations a and b in **Figure 3**. If, however, commodity circuits play central roles in the mobilization of resources, farming falls within the entrepreneurial (and/or corporate) agricultural constellations. Thus, the 'degree of peasantness' becomes central to the analysis of agriculture. This varies through space and time. Peasant agriculture is less dependent on markets for factors of production and inputs, while an increase in such dependency will move agriculture toward the entrepreneurial or capitalist modes of farming.

Peasant Agriculture: Six Decisive Features

The specific position of the peasantry within wider society (i.e., the peasant condition) has important implications for the way in which peasant agriculture is structured. The first and probably the most important one is that peasant agriculture is geared to producing as much value added as is possible under the given circumstances, and that its development aims, above all, to increase value added within the farm household. This focus on the creation and enlargement of value added mirrors the peasant condition: the hostile environment is faced through independently generating income in the short, medium, and long run. Although the centrality of value added production within the framework of peasant agriculture might seem self evident, this feature distinguishes peasant agriculture from other types of farming. The entrepreneurial mode is equally oriented toward the takeover of opportunities to produce value added as it is to the direct creation of value added. This process occurs through the acquisition of resources previously owned by others. Capitalist agriculture centers on the production of profits, even if this implies a reduction in the total value added.

The environment in which agriculture is embedded significantly influences the levels of value added and how they unfold over time. Peasant agriculture, in particular, requires space to fulfill its potentials. If such space is not available, due to negative interactions between peasant agriculture and the society in which it is embedded, its ability to realize these potentials will be blocked. Thus, peasant struggles are a reflection of the multifaceted nature of interactions between peasant agriculture and society at large.

A second distinctive feature of peasant agriculture is that the resource base available to each unit of production and consumption is limited and coming under increased pressure. This is partly due to internal mechanics such as inheritance practices that mostly imply a distribution of available resources among a growing number of households. It is also due to the external pressures on resources as, for example, climatic change and/or usurpation of resources by large export oriented corporate interests. Peasants will not seek to compensate these downward pressures by expanding their resource base through establishing substantial and enduring dependency relations with markets for factors of production, as this runs counter to the search for autonomy and would also imply high transaction costs. The (relative) scarcity of available resources increases the importance of improving technical efficiency. In peasant agriculture, this implies realizing maximum output with the given resources and without a deterioration in the quality of these resources.

A third important feature regards the quantitative composition of the resource base: labor will often be relatively abundant, while the objects of labor (land, animals, etc.) will be relatively scarce. In combination with the first characteristic, this implies that peasant production tends to be intensive: production per object of labor will be relatively high and the development trajectory will be shaped as an ongoing process of labor based intensification.

The qualitative nature of the interrelations within the resource base is also important. This brings out a fourth characteristic: the resource base is not separated into opposed and contradictory elements (e.g., labor vs. capital or manual vs. mental labor) but, instead, the available social and material resources represent an organic unity that is owned and controlled by those directly involved in the labor process. The rules governing the interrelations between the involved actors (and defining their relations with the resources) are typically derived from (and embedded in) local cultural repertoires, including gender relations. Chayanovian types of internal balance (e.g., those between drudgery and satisfaction) also play an important role.

A fifth characteristic (that follows on from the previous ones) concerns the centrality of labor: the productivity and future development of a peasant farm critically depend upon the quantity and quality of labor. Associated aspects of this include the importance of labor investments (terraces, irrigation systems, buildings, im proved and carefully selected cattle, etc.), the nature of applied technologies (skill oriented as opposed to mechanical), and peasant innovativeness.

In the sixth place, reference needs to be made to the specificity of the relations established between the peasant unit of production and the markets. Peasant agriculture is typically grounded upon (and simul taneously embraces) relatively autonomous, historically guaranteed, reproduction. Noncommodity flows and circuits dominate commodity flows and circuits. Each cycle of production builds upon the resources produced and reproduced during previous cycles (see **Figure 3**). Thus, they enter the process of production as use values, as labor objects and instruments (in short, as non commodities) that are used to produce commodities, and at the same time to reproduce the unit of production. This pattern starkly contrasts with market dependent reproduction, in which most or all of the resources are mobilized through markets, entering the production process as commodities. Thus, commodity relations penetrate into the heart of the labor and production processes.

From a neoclassical point of view, the differences between the situation of actively constructed self provisioning (i.e., a relatively autonomous and historic ally guaranteed reproduction) and the one characterized by high market dependency is irrelevant. Seen from the neoinstitutional perspective, they are both clear examples of the basic dilemma: to make or to buy? The typical peasant response to this dilemma is as relevant in developed countries as it is in developing ones.

The characteristics elaborated above flow together in the distinctive, albeit often misunderstood and materially distorted, nature of peasant agriculture, which is pri marily oriented toward the search for, and the subsequent creation of, value added and productive employment. In the capitalist and entrepreneurial modes of farming, profits and levels of income can be increased through reducing labor input. Both modes progress through, and contribute to, continuing outflows of labor from agri culture. This does not so readily occur in peasant units and, if it does, it is a step backward. In peasant farming, emancipation (successfully facing the hostile environ ment) necessarily coincides with the enlargement of total value added per unit of production (i.e., per farm). This occurs through a slow but persistent unfolding of the resource base, and/or through improving the technical efficiency. Mostly these two movements are combined and intertwined and are self reenforcing. In peasant agriculture, the ongoing increase of value added per farm is brought in line with simultaneous increases at two interconnected levels: that of the peasant community as a whole and that of individual actors engaged in the pro cess of production.

At the level of the peasant community as a whole, possession of a specific resource base by a specific family is generally recognized. Within the prevailing cultural repertoires (or moral economies), the takeover of adjacent plots or possessions is definitely not viewed as progress; for the peasant community as a whole this would be tantamount to self destruction. Hence, indi vidual peasant families strive to progress (albeit with different rhythms and different degrees of success) within and through their own efforts and using their own re sources. This adds to the overall growth of value added at the level of the community (or regional economy). Within the pattern typical for capitalist and/or entre preneurial farming, growth at the level of individual enterprises often associates with a stagnation or even decrease of the total amount of value added at higher levels of aggregation. A peasant economy excludes the occurrence of such a pattern. This also explains why, throughout history, agricultural sectors worldwide showed an ongoing increase in the absolute size of the total agricultural labor force. It also explains why demographic growth translated into further agricultural growth. It is only from the 1950s onward that such interrelations have been interrupted and replaced by new patterns.

Processes of Repeasantization

The world is currently witnessing new and often un expected processes of repeasantization. In Brazil, more than 1 million new peasant units have been forged through the struggles of the *Movimento dos Sem Terra* (MST) – the movement through which the dispossessed from large metropolis are returning to the countryside in order to create an existence as peasants. Similar, albeit less visible, processes are occurring throughout Latin America. In Eastern Europe, many people entered the peasant condition after the collapse of the Soviet Bloc. And, within the European Union, the widespread, en dogenous process known as farmer led rural develop ment is explicitly enlarging the peasantness of its agriculture. Expressions of this process include entering into on farm processing, developing direct retailing, of fering agro tourist facilities, engaging in the management of landscape and nature, creating new forms of energy production, developing new forms of cost reduction by regrounding farming on nature, and strengthening organic agriculture. These expressions of repeasantiza tion are intended to counter the squeeze exerted upon agriculture through liberalization and globalization and to enlarge autonomy at the farm level. At the same time, this repeasantization results in the creation of new and positive interrelations between the farming sector and society at large.

See also: Agriculture, Sustainable; Food Networks, Alternative; Green Revolution; Modernization Theory; Post-Productivist and Multifunctional Agriculture; Regional Development, Endogenous.

Further Reading

Altieri, M. A. (1990). *Agro Ecology and Small Farm Development*. Ann Arbor, MI: CRC Press.
Boserup, E. (1970). *Evolution Agraire et Pression Demographique*. Paris: Flammarion.
Bray, F. (1986). *The Rice Economies: Technology and Development in Asian Societies*. Oxford: Blackwell.
Chayanov, A. V. (1966). In Thorner, D., Kerblay, B. & Smith, R. E. F. (eds.) *The Theory of Peasant Economy*. Manchester: Manchester University Press.
Ellis, F. (1988). *Peasant Economics: Farm Households and Agrarian Development, Wye Studies in Agricultural and Rural Development*. Cambridge: Cambridge University Press.
Mendras, H. (1970). *The Vanishing Peasant: Innovation and Change in French Agriculture*. Cambridge: Cambridge University Press.
Mosher, A. T. (1966). *Getting Agriculture Moving: Essentials for Development and Modernization*. New York: Praeger.
Ploeg, J. D. van der. (2003). *The Virtual Farmer: Past, Present and Future of the Dutch Peasantry*. Assen: Royal Van Gorcum.
Scott, J. C. (1976). *The Moral Economy of the Peasant*. New Haven, CT: Yale University Press.
Shanin, T. (1990). *Defining Peasants*. Oxford: Basil Blackwell.
Thiesenhuisen, W. C. (1995). *Broken Promises: Agrarian Reform and the Latin American Campesino*. Boulder, CO: Westview Press.
Toledo, V. M. (1990). The ecological rationality of peasant production. In Altieri, M. & Hecht, S. (eds.) *Agroecology and Small Farm Development*, pp 53 60. New York: CRC Press.
Tudge, C. (2004). *So Shall We Reap: What's Gone Wrong with the World's Food and How to Fix It*. New York: Penguin Books.

Relevant Websites

www.viacampesina.org
La Via Campesina: International Peasant Movement.
www.rural impact.net
The Socio Economic Impact of Rural Development.
www.jandouwevanderploeg.com/en/empire peasantry and resistance
Rural Development Challenges and Interlinkages.

Peet, R.

A. Kobayashi, Queen's University, Kingston, ON, Canada

Glossary

Critical Modernism a term used by Dick Peet and Elaine Hardwick to describe an approach to contemporary political economy that views underdevelopment as a result of a series powerfully structured historical modes of production, with attendant changes in consciousness and forms of representation.

Peet, Richard (1940–)

The influence of a scholar depends not only on his or her written words, but also on the time in which those words are written. Richard Peet's voice has been influential in geography both because of the depth of his scholarship and because his words have provided intellectual and political leadership at key moments in the history of the discipline (**Figure 1**). He is both an academic and an academic activist.

Born in Southport (England) in 1940, Richard Peet received his formal education from the London School of Economics (BSc, economics, 1961), the University of British Columbia (MA, geography, 1963), and the University of California, Berkeley (PhD, geography, 1968). He has been teaching geography at Clark University since 1967.

Figure 1 Richard Peet (1940–)

Peet's doctoral research in traditional economic geography applied von Thunen's theories to the global expansion of commercial agriculture in the nineteenth century, showing a fascination with global spatial and structural systems, and their connections that has survived to the present. However, Peet was also a product of the socially tumultuous late 1960s in America and became a strong advocate of 'social relevancy' at a time of the anti war and the Civil Rights movements. In 1970, he became the editor of *Antipode*, the first geographical journal dedicated to a radical perspective that encouraged geographers to address urgent social issues and to use geographical knowledge to promote change. In the first issue, he outlined an agenda for teaching an undergraduate course in radical geography.

Peet's early publications take aim at poverty in the United States, stressing inequalities in American society and geography's indifferent or complicit role in fostering class differences, racism, war, and environmental degradation. He documented poverty as a geographical phenomenon that varies according to location, and in specific social, economic, and physical environments, including rural environments and racialized inner cities. This early writing has a pragmatic caste, emphasizing the need for practical solutions forged in public policy and planning.

Throughout the 1970s, Peet concentrated on economic theories to understand the structural conditions that foster inequality. As the pages of *Antipode* increasingly demonstrated the inadequacies of conventional geographical models to bring about social change, Peet's became one of the most articulate voices in the breakthrough to Marxism. He focused on two distinct, but interrelated, themes: the role of ideology in defining geographers' interpretations of the world, and the ideological frameworks that support and sustain economic systems.

During the 1980s and 1990s, his emphasis shifted from poverty in the developed world to international underdevelopment. With that shift came a deepening of his interest in structural understanding of social processes and modes of production causing underdevelopment, deindustrialization, spatial shifts in production, environmental degradation, and the expansion of new forms of economic consciousness. These issues have remained at the core of Peet's work, expressed in some of the most incisive critiques in the discipline of the transition from Keynesian to neoliberal economic regimes, and of the real differences in the material lives of ordinary people that that transition has hailed.

The depth of his intellectual contribution has been most apparent at key moments of transition (or 'paradigm shifts') within the discipline. During the 1990s, as the discipline was negotiating the jagged edges of the post structural turn, he became deeply involved in contem plating the relationship between economy and culture. Some scholars on the geographical left had criticized Peet for maintaining a 'binary' between the material and ideal. The result of this conversation has been an extended and fruitful debate, published in widely read journals such as the *Annals of the Association of American Geographers* and *Economic Geography*, over his staunch refusal to give ground on the philosophical question of how theory making relates to external reality. When Trevor Barnes, for example, wrote a critique of the use of metaphors in economic geography theories, claiming that the act of theorizing by all geographers, including Marxists, is a constitutive – and therefore contingent – one, Peet responded with characteristic pugnacity, maintaining his faith in the search for a truth beyond ideology, and giving no quarter toward work he views as relativist, amoral, or socially detached. And he gives no quarter toward economic geographers, including those writing from the left, in whose work he sees idealist tendencies.

His critique of idealism is extended in *Modern geo graphical thought*, widely regarded as his most important intellectual contribution, a large scale mapping of the discipline, from positivist, to radical, to post structuralist, and postmodern perspectives. It tacks between lucid descriptions of intellectual developments and charged critiques of the political and ideological implications of geographical ideas. From those critiques flows his concept of 'critical modernism', developed with Elaine Hardwick in *Theories of development* to support his enduring contention that underdevelopment is a result of powerfully structured historical successions of modes of production, in which changing forms of consciousness have played a role that he increasingly ascribes to acts of representation. While this development shows the extent to which he has taken seriously the work of post structural theorists such as Foucault, always he maintains a distinction between semiotic struggle and discursive constitution of reality, whether with respect to the natural environment or the industrial landscape.

Peet's more recent work addresses power in creating and sustaining both the ideology of capitalist relations and the instruments of state and economic control. Peet's thinking remains strongly influenced by Gramsci's understanding of ideology in forming state–society relations, but he exceeds Gramsci's emphasis on discursive meaning as an instrument of power. This emphasis gives his ongoing work a powerful basis for understanding the strategies of international actors such as the World Bank and the International Monetary Fund.

Peet's work shows a number of recurrent themes. Geographers need to have a radical conscience, to take responsibility both for understanding geographies of inequality and for using scholarship as a basis for social change. A geographical conscience implies a political position. Most important of all, perhaps, are his insistence on the politics of engagement and his faith in the pos sibility of emancipatory and liberation economics. As a scholar, a teacher, and an activist, Dick Peet is most aptly described as 'engaged' with the world that he has helped to change.

See also: Marxism/Marxist Geography I; Marxism/Marxist Geography II; Radical Geography.

Further Reading

Jones, J. P. III. (2001). Review of Richard Peet, modern geographical thought Richard Peet. *Annals of the Association of American Geographers* 91(4), 762 763.

Peet, R. (1970). Outline for a second year course on the socioeconomic geography of American poverty. *Antipode* 1(1), 1 34.

Peet, R. (1972). *Antipode monographs in social geography 1: Geographical perspectives on American poverty*. New York: Antipode.

Peet, R. (1977). *Radical geography*. Chicago: Maaroufa.

Peet, R. (1985). The social origins of environmental determinism. *Annals of the AAG* 75(3), 309 333.

Peet, R. (1991). *Global capitalism: Theories of societal development*. London: Routledge.

Peet, R. (1991). *Modern geographical thought*. Oxford: Blackwell.

Peet, R. (1996). Discursive idealism in the 'landscape as text' School. *The Professional Geographer* 48(1), 96 98.

Peet, R. and Hartwick, E. (1999). *Theories of development*. New York: Guilford.

Peet, R. and Watts, M. (2003). *Unholy trinity: The IMF, World Bank and WTO*. London: Zed Books.

Peet, R. (2007). *Geography of power: The making of global economic policy*. London: Zed Press.

People's Geography

D. Mitchell, Syracuse University, Syracuse, NY, USA

Introduction

People's geography describes an orientation toward the production of geographical knowledge. People's geographies draw on the theoretical advances achieved over 40 years of radical geography to create knowledge that is useful in the struggle for social and economic justice. This knowledge may take the forms of pointed scholarly analyses, popularizations of existing knowledge, or the translation of critical theories into political practice.

Roots of People's Geography

In a 1984 'Manifesto' David Harvey declared that "the geography we make must be a *people's geography*" that depicted the world as it really was rather than as we hoped it might be. He argued that a historical–materialist approach to human geography required focusing on the historically developed structures that shaped and determined people's lives and against which they struggled for change. Fifteen years later a group of American geographers met in New York City to found the People's Geography Project (PGP). The PGP is dedicated to the proposition that in the struggle for social and economic justice it is vital to understand how everyday life is structured through relations of power that are fundamentally geographical. It seeks to popularize radical geographical analysis so it becomes a stronger tool in the fight for a more just world.

The notion of a people's geography has many roots. Among US geographers, Howard Zinn's influential *People's History of the United States* is one. This book outlines the social and political history of the United States from the perspective of working people, minorities, and the dispossessed, and seeks thereby to shed critical light on the ideological claims of freedom, right, and goodness that support the American imperial project at home and abroad. The book has long been, and remains, a lodestar to activist groups seeking progressive social change. Similarly, the more general social history movement of the 1960s and 1970s, culminating perhaps in the work of the American Social History Project in the 1980s and 1990s, sought to show how ordinary people make their own histories – if always in highly constrained ways – and thus are a powerful social force. This general orientation guides people's geographies.

Following the pioneering work of the geographer, Allan Pred, and taking inspiration from theories of the production of space developed by Henri Lefebvre and his followers in Anglo American geography, geographers seeking to develop a people's geography understand that any production of history requires a concomitant production of geography and thus the taking and making of space is always a contested process. Pred's particular project was to show how geographies – spatial relationships and produced spaces – always constrained and enabled social life and thus had to be a keen focus of political struggle. Indeed, they always have been such a focus, but that this was the case was not sufficiently understood. In turn, this lack of understanding hobbled social movements in their attempts at radical – or even everyday – social change. The goal of any people's geography has to be to help foster the realization of the fundamental importance of space and spatiality to social life and social justice.

Another crucial influence on the development of people's geographies in general and the PGP in particular, is the Detroit Geographical Expedition and Institute (DGEI) and its offshoots in Toronto, Vancouver, London, and elsewhere. Developed in collaboration with neighborhood residents and activists (especially teenagers) by pioneering geographer, Bill Bunge, after he moved to Wayne State University in the mid 1960s, the DGEI sought (1) to uncover and map the structural forces that shaped the unjust geographies of everyday life in the city; (2) to bring geographical knowledge to 'the people' and thereby to transform the nature of geographical knowledge production and learning; and (3) in the process to transform higher education (especially geography) to make it more accountable to community people. Working with community members and activists, the DGEI sought both to diagnose the geographies of injustice within which (particularly) African Americans lived and to transform those geographies. Mapping was critical to this project – 'expeditions' mapped everything from incidents of broken glass, to rat bites, to where drivers typically sped (Bunge was particularly interested in how unjust geographies threatened the survival of children) – but so too was analysis. For example, the DGEI hypothesized that schooling for African American children would be better if it were under 'sympathetic authority', that is, if local community members, and members of the same race, had control of local schools. Analysis thus focused on how to change the geographies of authority – through electoral campaigns, advocacy movements, etc. – within the Detroit school district. In other words, the DGEI sought not only to uncover the reality of people's geographies, but because it understood that these geographies shaped – even determined – the structures of life and death, it sought to radically alter them.

A final critical determinant of the rise of interest in people's geographies in the 1990s and early twenty first century was the general scaling back of leftist geographers' ambitions as they retreated from constructing a 'radical' project to an interest in a 'critical' one. Radical geography's early years were suffused with a sense of urgency and filled with the potential that the world really could be changed. Forged in the fires of the global uprisings of the 1960s, radical geography was radical both in that it sought and expected real social transformation (and understood that geographic knowledge was critical to such change), and in the sense that it was creating something radically new in what had been a fairly hidebound discipline. If the so called quantitative revolution had shaken geography's old guard but nonetheless seemed to be aligned with the political status quo (it was largely a technocratic development), then radical geography hoped to upend both the discipline and the world of which it was part. Radical geography was thus infused with experimentation: theoretical experimentation, as well as experimentation in modes of exposition. It sought to be directly relevant in ordinary and oppressed people's lives in ways that academic knowledge production rarely was.

However, with the defeat of the global left in the 1970s, radical geography turned inward. Much of this was salutary, leading as it did to a consolidation and development of theoretical gains in Marxism, feminism, and other oppositional modes of theorizing. Academic debate in left leaning geography during the 1970s, 1980s, and into the 1990s was sharp. New conceptualizations of scale, social reproduction, the relationship between structure and agency in everyday life, the social production of nature; new projects focusing on the fate of places in the midst of industrial restructuring, the emerging geography of sexuality, the restructuring of local and national states; new theories imported from French post structuralism, cultural feminism, cultural studies, and elsewhere: all these built on, while also strongly critiquing, the foundation of radical geography forged in mid twentieth century Marxism and socialist feminism. They radically transformed radical geography. Indeed, so great had been the change, and so thorough the disengagement with the social movements and political ferment that had marked the birth of radical geography, that by the early 1990s numerous left leaning geographers, especially but not exclusively in Britain, had begun calling for the dismantling of those institutions that had sustained the radical movement, such as the Socialist Geography Specialty Group of the Association of American Geographers (which itself had succeeded the far less staid Union of Socialist Geographers), and their replacement with institutions that would instead promote 'critical geography'. (And in fact, by the start of the twenty first century, this move had been successful with the Socialist Geography Specialty Group rebranded as the Socialist and Critical Geography Specialty Group and a new international formation of left leaning geographers that came together under the banner of the International Critical Geography Group.)

By the 1990s, 'critical geography' had become the preferred term for all that scholarship that was (1) oriented toward theoretical development and explication; and (2) did not overtly hew to a corporate, military, or elite class, sexist, racist, or homophobic line. Critical geographies encompassed everything from studies of fashion in men's magazines and everyday shopping behavior, to abstruse arguments about how creative people use commodities creatively, to polemics about how capitalism was nothing more than a discourse that could be combated by creating alternative discourses. Critical geography, in many ways, was the academic echo of 'third way' soft social democracy. To be sure, there was still much strongly oppositional, directly political work – such as the turn to Lefebvre already noted, the theoretical and empirical development of labor geography, the rise of queer geographies, and the close, critical examination of Clinton–Blair workfare policies that paved the way for a concerted theorization of, and attack on, neoliberalism in the 2000s – but such work was scattered, poorly coordinated, often attacked for being 'totalizing', and generally swamped by a larger celebration in critical geography of geographers' own theoretical sophistication and cleverness.

Such disciplinary contexts – the history of intellectual and political ferment, and the retreat into the carapace of 'critical' scholarship – provided the immediate context for those who met in New York City in 1999 to revive an interest in people's geography and a return to the arguments Harvey had made in the 1980s for the need to develop a "mundane enterprise that reflects earthly interests, and claims, that confronts ideologies and prejudice as they really are, that faithfully mirrors the complex weave of competition, struggle, and cooperation within the shifting social and physical landscapes" of the contemporary world. Such disciplinary contexts made clear the need to reorient the progressive geographical enterprise back toward its roots in understanding directly the ways in which geographies of social and economic injustice were produced, what they meant for those who lived them, and thus how they could be thoroughly transformed.

The People's Geography Project

The PGP was founded to address this need. Founded in late 1999, the PGP brought together radical geographers from around the US with the goal of taking what was best from the impressive intellectual ferment that had marked geography since the 1960s – the rapid theoretical advancement, the lasting, if now muted, commitment to social justice – and putting it to work 'for' people. The PGP's founding goal was to popularize radical geography (and radicalize popular geography) so that the geographic insights necessary to progressive social change could be ready at hand for both activists and those who simply wanted to better understand the worlds of which they were part. The PGP was, therefore, as much an intellectual project as an activist one. Taking its cue from the radical popularizations of Zinn and the social history movement, members of the PGP outlined a series of projects that ranged from creating 'guerilla geographies' (pamphlets addressing critical issues of the moment), to developing teaching materials for both university and school teachers, to producing a comic book of radical geography (modeled on the famous 'For Beginners' series), to writing a large scale *People's Geography of the United States*. In addition, the PGP was active in organizing special sessions at conferences of the International Critical Geography Group, the (American) Working Class Studies Association, the American Studies Association, Rethinking Marxism, and elsewhere, in an effort to draw others into the project of constructing peoples' geographies.

Over the next year work began on each of these initiatives. With the terrorist attacks in the US on 11 September 2001, however, the work of the PGP shifted. Working with a team of students from Syracuse University (where the PGP director was employed), the PGP rapidly mobilized to become a clearinghouse for teachers seeking alternative materials for teaching about the attacks, American imperialism, the geographies of terrorism, and related issues. A later, similar clearinghouse was developed in the run up to the war against Iraq. Such efforts monopolized the time of PGP members during this period and momentum was lost on other planned projects, such at the *People's Geography of the United States* book which still has not been written.

Simultaneously, PGP faculty and students at Syracuse University turned their attention to pressing local concerns: hunger, economic injustice, racism, and so forth. Beginning first with the Syracuse Hunger Project and developing into a Syracuse Community Geography Project, geographers, local social service providers, activists, and others, began a concerted campaign to map and analyze the landscape of food injustice, racism and incarceration, the structural determinants of teen pregnancy, and other critical forces. Such work, while in the spirit of the PGP (and drawing as long as it lasted on PGP funding), detracted from the intellectual and larger scale popularizing projects the PGP had outlined as primary goals.

Indeed, the PGP has not achieved most of its original goals. It has, however, maintained an active website with resources for scholars and students seeking to develop their own people's geographies; it has been an active presence at conferences; and work under its banner continues to be published in journals. More importantly, others have picked up the banner.

Other Projects

In Los Angeles, Laura Pulido and others have developed maps and walking tours of people's histories and geographies. In Chicago, Carrie Breitbach and the Chicago Center for Working Class Studies have published a labor history map. And in the American South, Derek Alderman has used funds from the PGP to produce maps and brochures on the geographic legacy of Martin Luther King, Jr. A course on the People's Geography of American Empire is offered at Evergreen State University in the state of Washington; its instructor, Zoltan Grossman, is working with Joseph Nevins of Vassar College to write a book of the same name. Nik Heynen of the University of Georgia is reworking people's geographies as geographies of survival.

There is a renewed interest in producing public geographies, which, though not identical to people's geography, share many overlapping concerns. Katharyne Mitchell at the University of Washington used her term as the Simpson Professor of the Public Humanities to promote the role of public intellectuals in geography and beyond. At Birmingham University, Ian Cook and his colleagues have developed a public geographies project that possesses many of the same goals as the PGP. Scholars and activists in Australia have adopted the people's geography title for a website that links activist scholarship with progressive politics. Activist students and faculty in Hong Kong, Hungary, Italy, Japan, Taiwan, and elsewhere have begun to develop the people's geography idea in ways appropriate to their own contexts. And, as witnessed by the 2007 Summer Institute on Geographies of Justice (cosponsored by the PGP), a new generation of geographers, no longer content to be merely 'critical', are reclaiming the radical mantle.

In other words, the 'idea' of people's geographies continues, and continues to change, even if the specific goals of the PGP have gone unmet.

See also: Critical Theory (After Habermas); GIS, Public Participation; Historical-Geographical Materialism; Marxism/Marxist Geography I; Marxism/Marxist

Geography II; Participatory Action Research; Radical Geography.

Further Reading

American Social History Project (1989, 1992). *Who Built America* (2 Volumes). New York: Pantheon.

Bunge, W. (1971). *Fitzgerald: Geography of a Revolution*. Cambridge, MA: Schenkman.

Bunge, W. and Bordessa, R. (1975). *Geographical Monographs No. 2: The Canadian Alternative: Survival, Expeditions, and Urban Change*. Toronto: York University Department of Geography.

Castree, N. (2000). Professionalism, activism, and the academy: Whither critical geography. *Environment and Planning A* 32, 955 970.

Fuller, D. and Kitchin, R. (eds.) (2004). *Radical Theory/Critical Praxis: Making a Difference Beyond the Academy?* Victoria, BC: Praxis (e)Press.

Hague, E. (2002). Antipode, Inc.? *Antipode* 34, 655 661.

Harvey, D. (2001). On the history and present condition of geography: A historical materialist manifesto. In *Spaces of Capital* (first publ., 1984), pp 108 120. New York: Routledge.

Heyman, R. (2007). 'Who's going to man the factories and be the sexual slaves if we all get PhDs?' Democratizing knowledge production, pedagogy, and the Detroit Geographical Expedition and Institute. *Antipode* 39, 99 120.

Heynen, N. (2006). But it's alright, ma, it's life, and life only: Radicalism as survival. *Antipode* 38, 916 929.

Mitchell, D. (2006). The people's geography project: Popularizing radical geography. In Mizuuchi, T. (ed.) *URP Research Paper No. 1: Critical and Radical Geographies of the Social, the Spatial, and the Political*, pp 10 21. Osaka: Urban Research Plaza, Osaka City University.

Mitchell, D. (2007). Community geography. *Newsletter of the Association of American Geographers* 42(3), 18 19.

Peet, R. (ed.) (1977). *Radical Geography*. Chicago, IL: Maaroufa Press.

Pred, A. (1984). Place as historically contingent process: Structuration theory and the time geography of becoming places. *Annals of the Association of American Geographers* 74, 279 297.

Zinn, H. (2003). *A People's History of the United States*. New York: Harper Perennial.

Relevant Websites

http://www.econgeog.misc.hit u.ac.jp/icgg
International Critical Geography Group.

http://www.academic.evergreen.edu
People's Geography of American Empire, Evergreen.

http://www.peoplesgeography.com
People's Geography/Reclaiming Space.

http://www.communitygeography.com
Syracuse Community Geography.

http://www.peoplesgeographyproject.org
The People's Geography Project.

Performance, Research as

F. Morton, Royal Scottish Academy of Music and Drama, Glasgow, UK

Glossary

Nonrepresentational Theory A body of thought developed by Nigel Thrift endorsing many key strands of post-structuralist theory, but extends to consider how space is made in ways outside of representation. It considers social practices that cannot be represented through words and discourse, and amplifies the importance of these in terms of related affect, emotions, and spaces.

Post-Structuralism A branch of philosophy, a subsection of which was developed by French and Continental thinkers such as Gilles Deleuze and Felix Guattari, and which moves away from strict binaries that form the core of structuralism, toward the notion of multiple realities and possibilities and an opening out of the world.

Research as Performance: Introduction and Context

Performance, although often researched as a field in its own right can be used as a research tool, and/or a metaphor to think and act through when researching performance and social practice. Within human geography, this consideration has emerged alongside a body of work attentive to post structuralist philosophy (most notable in the work of Deleuze and Guattari and other Continental philosophers), and the applied nature of performance in all its manifestations – from the social practices of everyday life to performing arts such as dance, drama, music, theater, and so on. This body of work is called nonrepresentational theory and has provided an opening for performance research within human geography. It claims to supplement the understanding of practice and performance, rather than write a general prescriptive theoretical approach. It can be argued that Thrift is the source of inspiration for this particular strand of performance research.

Thrift considers the problems that geographers researching performance frequently face – nonverbal, noncognitive bodily practices that often resist interpretation, representation, description, and definition. The nonrepresentational approach attempts to explore some of these issues, and indeed, to put forward some suggestions for actually carrying out fieldwork into these social practices. Although the roots of nonrepresentational thought can be traced to wider philosophical and theoretical movements, the more evident, consolidated, geographical interest in performance, performance theory, performance studies, and an emphasis on social practice emerged in the late 1990s and early twenty first century, filtering through to methodological explorations.

The Methodological Shift

A distinct dissatisfaction with conventional qualitative techniques for performance research began to manifest within geography at the start of the current decade. For example, the special edition of *Environment and Planning D* in 2000 addressed specific facets of performance (of both everyday life and the arts), and how to research it, as did the *Possibilities of Performance* special edition of *Environment and Planning A* in 2003.

Doctoral students at the School of Geographical Sciences at Bristol University around this time were also embarking on methodological innovations in relation to much interest in post structuralist and nonrepresentational approaches. A research cluster emerged around issues of performance and practice, and more importantly, asking how to actually employ methods that would be able to get at these issues. The then Bristol based research cluster (comprising, JD Dewsbury, John Wylie, Paul Harrison, Derek McCormack, Mark Paterson, Claire Pearson, Emma Roe, and others), created a core of practice based approaches that would then unravel to inform research methods using performance as a metaphor. Within human geography these methods originally emerged as a response to various needs in the fields of performance research. A summary of some typical key problems with traditional methods are outlined here:

1. Existing qualitative methods within the social sciences have frequently relied on interpretation and social construction, and therefore lack the capacity to grasp the more expressive, emotive, and nonverbal, noncognitive aspects of social practices and performances that resist interpretation.
2. There is a lack of suitable vocabulary within current geographical and social scientific methodology to talk about, present, or document practices and performances.
3. Existing methods are not attuned to the sensuousness of performance and practice, and the accompanying temporalities, for example, events in real time.
4. Existing methods rely heavily on words to describe and explain rather than to experience or evoke

practice in a much more relevant way, through other means such as film, photography, sound, etc.

5. Existing methods do not accommodate the unpredictability of performance, or the significance of events occurring in real time.
6. Existing methods do not allow us to get close to, or get within, performance/practice.

Performance Methods

There are two main ways of dealing with the problems that conventional methods create for performance research.

Reworking Existing Methods

Methods have the potential to engage with, and go with, the flow of embodied practices, and their ways of being in the world, and so there should be more use of techniques that instill conventional research methods with the sense of the creative and the performative. Explorations into using methods in such ways, and adapting methods to suit performances being researched, have been taking place within geography. In particular, the performing arts such as dance, drama, and music, are apposite examples because they have performative and embodied articulations and negotiations of space and social life at the very core of their existence. They are embodied, affective, and expressive, and they involve acts of doing in real time – in the now. Some examples of reworked methods are given here:

Diary–photo diary–interview

Alan Latham's diary–photo diary–interview requires respondents to keep a diary of their sociable practices for a certain period of time, which comprises a log of places they went, at what time, with whom, and why. An interview based on the diaries is then undertaken. The use of photography by the participants is also encouraged (see the section titled 'Knowledge of practice/knowledge in practice' below). The benefit of this approach is the variety of narrative accounts, accompanied with pictures and additional interview material, which offers a wide range of possibilities for presentation of research materials. It may also help to embody and convey the sense of performance and movement within the actual research (see **Figure 1**).

Reworked ethnography

Ethnographies of use

Loretta Lees, in her study of the new public library building in Vancouver, endorsed a technique that she calls ethnographies of use. She acknowledges and presents the everyday engagement and use of the library

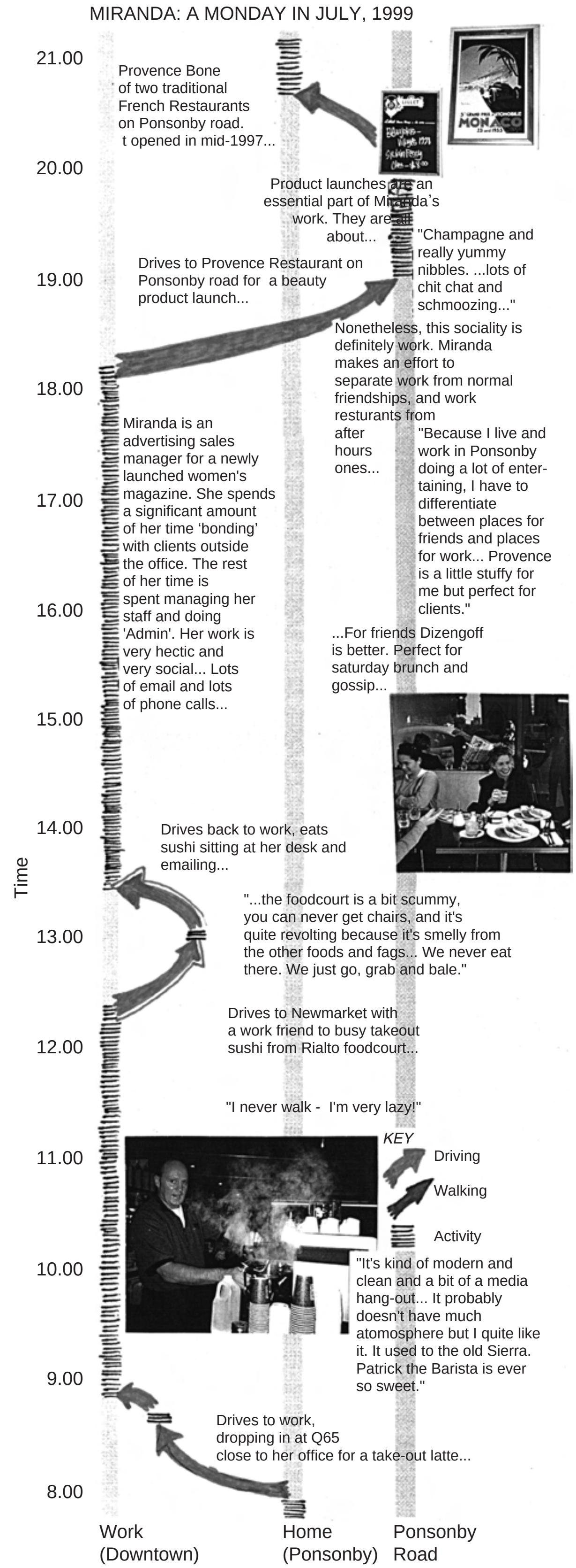

Figure 1

building through a series of vignettes, based on her own experiences of being in the building.

Audio ethnography

By piloting research that prioritizes sound rather than sight, Susan Smith reveals that existing qualitative methods are too engrossed in what can be seen, rather than what can be heard. Her proposal for an audio ethnography involves an adaptation of existing traditional methods in order to account for the sound environment, in addition to crafting new procedures to access and to present the sound world.

Musical ethnography and ethnomusicological influences

Anthropologist, Sara Cohen, claims that musical ethnography should accentuate the particular complexities that take place in musical performance. Similar innovations, but within the field of ethnomusicology, have occurred via a conceptual shift from classification, description, and explanation of music structures, toward an understanding of music as culture and music as a lived out social practice. This shift is especially evident within the methodology of contemporary ethnomusicology research practices.

Participant observation and observant participation

New ways of researching music by active participation and engagement rather than observation and description, and by using the musical skills and knowledge of the researcher, are evident in much contemporary ethnographic work. Active and personal participation in social and often skilled practice can have innovative results, for example, narrative accounts of playing the piano. In geography exclusively, much more attention is being paid to taking part in actual live events of musical performance, and how we can engage with/in them, as opposed to the textual analysis of musical scores, or song lyrics. Indeed, Thrift's emphasis upon observant participation is fitting to this sort of approach. Wood *et al.* also identify some key elements of this type of performance research, for example, 'participant sensing' being particularly relevant here, whereby the researcher experiences the performance at the time it occurs, while note taking at the same time.

Performative ethnography

Performative ethnography is an approach related to the specific practice of Irish traditional music sessions in Galway City. It differs from standard ethnography due to its emphasis on the multiplicity and fluidity of performing and improvising in the now – how people move through different spaces and experience them differently. It also differs from conventional ethnographies, which observe particular cultural groups and only write about them from the outside because this research performs within the social practice in question, and so experiences the performance. This approach moves away from rigid and standardized research questions and structures, and toward capturing the sense of the social practice as something always in process. The techniques employed within this performative ethnography are outlined below in **Table 1**.

This approach has some crossover with techniques of participant sensing, photo elicitation, recording, and styles of writing to embody performance that musical geographers are pioneering.

New Methodological Approaches to Performance Using the Influence of the Performing Arts and Performance Studies

Performative writing

Performative writing is a creative and useful tool for any type of performance based research and should be an attempt to creatively act out the emotional and affective forces of performance. Performative writing should be evocative and bring the reader into contact with and conjure up the various aspects of performances, as exemplified in the excerpts below (**Boxes 1** and **2**).

Multimedia and creative presentation

Using media like photography and video have also emerged as additional ways of evoking performances, and another way of imbuing our research methodologies with a sense of practice. For example, research by Emma Roe uses video methodologies for mediating the practices of food. Such an approach can work in synergy with performative writing, to evoke place and time (see **Figure 1** as a way of evoking bodily movement when making place). For example, using video, photography, sound recordings, etc., makes it possible to gain more of a sense of performance, and what the spaces and times involved are like (see **Figure 2**).

Knowledge of practice/knowledge in practice

Knowledge of specific arts and disciplines can be considered specialist because these have often been practiced, developed, and refined over a long period of time. If, for example, the researcher does not have specialist knowledge of the practice in question, performance research is still useful. It is still possible to get close to practice at different levels in order to get at different things. For example, one geographer has narrated the experiences of learning French folk music and dancing and the ability to engage with music to understand the affective and embodied elements of folk practice.

Cultural geographies in practice

Examples of new research styles using performance are evident in the accounts in 'Cultural geographies in

Table 1 Performative ethnography

Technique	*Participants*	*Recommendation*
Spoken diaries	All musicians	Each musician was given a tape and tape recorder and asked to record their experiences of each session they played in or went to. They were given some rough guidelines to cover, such as who was playing, who was in the crowd, how they felt during and after the session, did they enjoy it or not.
Audio recording	Researcher	Performances were recorded to make a CD of the music in the research. This provides an aural presentation of research materials.
Participatory interviews	Researcher and musicians	Each interview differed according to the musician. The aim was to talk about the sessions, and eke out the issues they did not address explicitly during the spoken diaries. These interviews resembled conversations about the performances and general thoughts on Irish music and Galway City rather than interviews.
Photography	Researcher and musical assistant	Use of digital photography allowed images of the city, and memorable performances to be captured. This was for the purpose of more visual presentation of the research alongside written material.
Video	Researcher and musical assistant	Use of video allowed DVDs to be made from the recorded performances, and the recordings of Galway City. This adds to the creative visual presentation of the research without using words.
Academic interviews	Academics in music departments	These informal interviews with academics in Ireland allowed development of the theoretical and conceptual side of the research, and provided an ongoing review process for the research material accumulating.
Performative writing	Researcher using musicians' and personal accounts	This formed a second phase of research, where writing performed and embodied the expressive and affective nature of the research. It also involved writing evocative vignettes from a research diary for inclusion in the thesis.

Box 1 Excerpt from Morton, 2005a

My eyes are closed, and all I can feel is rhythm pulsating through my body. I can feel the shape of the tune, the patterns it makes as the melody rises and falls. The wood of the flute is buzzing through my hands, as the feeling runs lightly along the volar ridges of my finger tips, and the soft pads vibrate. I open my eyes and the rich pine green color of the walls in Taaffe's, lined with pint-bearing listeners, seeps through the wispy layers of smoke, and I look toward Lorenzo as I can feel that the end of the tune is near, and something is going to happen, it is going to change: he nods at Barry, then at me, all the time still playing the pipes, and Barry, who is playing the box, his body jolting as he stamps his foot in time with the rhythm, gives a slight 'hup' and then the tune has changed. We all flow from the last to the next, to this, and I close my eyes again to feel the rhythms of not only the music, but of the sounds and smells as they close in on me, and I disappear into the world of music.

Box 2 Excerpt from Phelan, P. (2000). Love's geography. *Performance Research* 5(3), 86–89.

Dearest,

If I told you I love this city, love the way it opens its dirty limbs to me, love the way it absorbs me without noticing me, love the way it gives rhythm to my walk, gives nuance to my skin, gives purpose to my plots, would you be jealous? I know you refuse to be jealous of people, but I wonder how you would feel if I swear my utter devotion to a place.

Figure 2

practice' – a methodological forum in the journal *Cultural Geographies*. The emphasis here is on alternative presentations of research, and experimenting with ways of writing up research. This is related to performative writing, and similar work can be found in performance studies.

Ethics and Problems in Performance Research

Taking part in performance research means intensive participation. Blurring of boundaries and roles between being a researcher and a performer is inevitable, however, can also highlight ethical considerations for the social practice and the dynamics of the performance that is being researched. For the researchers with specialist knowledge of a particular practice, this may be less of an issue, as they are more likely to be aware of etiquette and acceptable levels of interruption. Accepting that the presence of a researcher may alter events slightly is also inevitable, and can be problematic, however, there are also benefits of this, as the researcher becomes a per former in the process, alongside the research participants, and the issue of positionality becomes a less power filled problematic.

Contributions and Benefits of Performance Research to Geography

Inclusive Spaces

Performance related methods make it possible to nego tiate access to spaces that are created in the now. Al though these spaces only exist for the duration of performance, and are therefore temporary and transient, they are given more weight in geographical research. This is because they make sense, if not necessarily of fering rational explanation or recognizable meaning. These spaces of the now constitute a sense of belonging and understanding for many people who participate in their making.

Geography can now begin to be more inclusive toward ways of knowing and sensing the world that exist in practices such as sign language, music, dance, and art therapy, to name a few, rather than remain focused on verbal and/or textual ways of talking and writing about space. Despite the degree of need for the scriptural within geography, there must be space made to consider the alternative advantages that the nonverbal can bring to the discipline, especially in terms of including groups and practices that do not use words or texts.

Politics of Research

Using performance in research advances a bottom up methodological politics, which, rather than follow a prescriptive theory, respects the practice in question. In this sense, practice becomes the means through which theory is created. It becomes an active methodology – a theory of practice is actively being written as perform ances take place. Research is a more democratic process – power is diffused from the researcher's assumed central position, toward participants who are reconstituted as co researchers, rather than 'subjects' of research. The per formance based methodology also endorses a reworking of the dynamics between the researcher and the re searched by emphasis on co production of events and materials to be used in the presentation and execution of the findings. Experiencing the strangeness, the imma terial forces and unsettling aspects of performance, and the complexity of researching all this puts the researcher in a different position whereby he/she is susceptible to the effects of performance rather than only selecting events that fit into a research program. As a result, it is possible to feel, hear, and experience the difference that nonrepresentational approaches can make to both theory and methodology. Overall, this type of research endorses a different kind of research politics than conventional methodologies do.

From Positionality to Positioning

Positionality within the field and among the people in volved in the research has been a central problematic in traditional and contemporary ethnography because of the suggestion that our position affects research by making it more subjective. Within performance research there is a need to move from one particular position (that of re searcher) to an awareness of the fluid and changing ways that people become positioned within practices (posi tionings), especially researchers and participants. This provides potential to produce relational knowledge and a relational appreciation of positionings – people can move from being performer (playing music, undertaking ob servant participation) to researcher (actively researching within a given situation) and from researcher to per former (performing inside and outside of the research).

Interdisciplinarity

The kinds of spaces that are accessed by performing have rarely been presented to a geographical audience before, despite their longstanding contribution to other discip lines and interdisciplinary research such as performance studies; ethnomusicology; cultural studies; drama, and other forms of the arts such as writing, poetry, etc. The method of performing research within geography dares to move away from more Euclidean, geometric, and static concepts of space, and toward more live, active, em bodied, and dynamic ever changing spaces.

See also: Affect; Embodied Knowing; Ethnography; Performativity; Space I.

Further Reading

Cohen, S. (1998). Sounding out the city: Music and the sensuous production of place. In Leyshon, A., Matless, D. & Revill, G. (eds.) *The Place of Music*, pp 269 290. New York: The Guilford Press.

Jones, S. H. (1998). Kaleidoscope Notes: Writing Women's Music and Organizational Culture. *Qualitative Enquiry* 4(2), 148 177.

Latham, A. (2003). Research, performance, and doing human geography: Some reflections on the diary photo diary interview method. *Environment and Planning A*. *Special Edition*: *Possibilities of Performance* 1993 2017.

Latham, A. (2006). Sociality and the cosmopolitan imagination: National, cosmopolitan and local imaginaries in Auckland, New Zealand. In Binny, J., Holloway, J., Millington, S. & Young, C. (eds.) *Cosmopolitan Urbanism*. London: Routledge.

Lees, L. (2001). Towards a critical geography of architecture: The case of an ersatz colosseum. *Ecumene* 8(1), 51 86.

Morton, F. (2005b). Performing ethnography: Irish traditional music sessions and new methodological spaces. *Social and Cultural Geography* 6, 661 676.

Phelan, P. (2000). Love's geography. *Performance Research* 5(3), 86 89.

Pollock, D. (1998). Performing writing. In Phelan, P. & Lane, J. (eds.) *The Ends of Performance*, pp 73 103. New York and London: New York University Press.

Smith, S. J. (2000a). Performing the (sound) world. *Environment and Planning D: Society and Space* 18, 615 637.

Smith, S. J. (2001). Doing qualitative research: From interpretation to action. In Limb, M. & Dwyer, C. (eds.) *Qualitative Methodologies for Geographers: Issues and Debates*, pp 23 41. London: Arnold.

Sudnow, D. (2001). *Ways of the Hand. A Rewritten Account*. Cambridge, MA: The MIT Press.

Thrift, N. J. (2000a). Non representational theory. In Johnston, R. J., Gregory, D., Pratt, G. & Watts, M. (eds.) *The Dictionary of Human Geography* (4th edn.), p 556. Oxford: Blackwell Publishers Ltd.

Thrift, N. J. (2000b). Performance. In Johnston, R. J., Gregory, D., Pratt, G. & Watts, M. (eds.) *The Dictionary of Human Geography* (4th edn.), p 577. Oxford: Blackwell Publishers Ltd.

Thrift, N. J. (2000c). Afterwords. *Environment and Planning D: Society and Space* 18, 411 432.

Wood, N., Duffy, M. and Smith, S. (in press) The art of doing (Geographies of) Music. *Environment & Planning D*: *Society & Space.*

Performative and Embodied Mapping

C. Perkins, University of Manchester, Manchester, UK

Glossary

Cartography Practices, processes, and discipline concerned with transforming and representing a real world in a mapped form.
Corporeal Appertaining to the body as against the mind.
Ethnography Qualitative methodology based on in-depth field-based contextual research.
Performance Doing action and work.
Performativity A nonfoundational approach to the production of identity, grounded in action and practice.

Introduction

People have always made maps and employed them to make sense of their worlds. Their maps communicate information in symbolic forms, which we read and interpret. What we make of them varies. For Western society, the map now stands mostly as a mirror like representation of the real, a factual tool to help us to navigate, plan, and control the world out there. Employed to chart explorations, administer cities, foster trade, bound nations, regulate property transfer, locate people, places, or events, and to link us to the world, the map as artifact is a practical device – an outcome of a problem solving technology. In our time the artifact carries with it an aura of authority: what else would one expect from a functional mechanism?

The processes that wrap around the object have been regarded as being less significant than the map itself. Until recently, the history of cartography was the history of mapmaking, in which accuracy constantly improved, technologies advanced, and innovation took place and from which the discipline of cartography emerged after World War II. Cartographic research mostly focused on a very narrowly defined set of practices: the goal was a normative improvement of map design, in isolation of any social context. The scientific maps produced by cartographers were the maps that mattered, and these were taken for granted; all other mapping was ignored or demonized. Social processes underpinning this progressive narrative went largely unnoticed until the last decade of the twentieth century, and the rich diversity of everyday mapping was marginalized.

Since the late 1980s, however, there has been a major shift toward considering mapping processes. Some are technical, within the ambit of traditional cartographic research. Maps are surveyed, derived, drafted, copied, and scanned; they may be a source for other mapmakers, digitized, and rescaled; information may be selected for inclusion, generalized, and symbolized. An artifact may be produced from these practices, and then a whole series of other things happen around it. The map might be folded, or rolled; converted to another file format; and embedded with other media. It might be packaged, marketed, sold, bought, used, stored, collected, reused, thrown away, or recycled. The cultural uses of mapping are particularly complex: it is read in different ways in different contexts, but the same map may mean different things. It might fire the imagination, inspire a revolution, plan a journey, make money, or teach moral values. Meaning is constituted in the actions that mapping processes call into being. So the performative and embodied mapping characterized in this article is both dynamic and enacted. The argument focuses upon the ways that people deploy mapping behavior and artifacts, and how these might be related to more abstract theoretical constructs such as Butler's notion of performativity. The cultural shift toward practice and performance forms its chief focus.

The Turn toward Practice

In the last decade of the twentieth century, there has been a shift in the humanities and social sciences from considering texts as the bearers of culture, toward performative ways of knowing the world, in which the dynamic aspects of culture matter. New ways of understanding the world have emerged from theater studies, anthropology, environmental psychology, the philosophy of language and linguistics, sociology, science studies, visual culture, and art history, and these have increasingly been employed to interpret mapping. The philosophical shift away from representation, toward action, has been paralleled by a greater contextual concern. Relations and flows in a network are more important in this approach than entities and their properties. Contextual approaches to mapping have increasingly emphasized the interplay between history, geography, and social or literary studies. The changing ways in which mapping has 'worked' in different contexts have shifted the emphasis from a study of artifacts, to a much more ethnographic concern with practice and the relations of mapping performances to other actors and actants.

During this change, fields of study have been reshaped. Mapping is being rethought in interdisciplinary endeavors, which reject long accepted binary distinctions. New concepts have emerged, with mapping theorized as performative, emergent, narrated, and affectual, or as networked flow. These changing ways of thinking about mapping focus around performance and embodied ways of understanding processes and a shift toward ethnographic ways of understanding.

Judith Butler's notion of performativity offers an important way of understanding mapping practices. Butler argued that identities are constituted through what people do, rather than who they are, and that bodily practice is central in this process. She recognized that unspoken psychoanalytical factors contribute to the formation of subjects, but argued that social norms and historical contexts also strongly affect action, through the operation of discursive power and past performative utterances. We never quite ever conform to the ideal norms around which performativity acts, so individual agency and resistance are still possible. A performative approach sees mapping as not only taking place in time and space, but also capable of constituting both. New worlds are made every time a map is deployed. Del Casino and Hanna argue that the practices of mapping might be understood using this construct. They suggest that mapping is always being made and consumed and that, following Butler, ethnographic work should move beyond the binaries author/reader, body/society, and subject/object. Instead, they argue that spaces and mapping co constitute one another, produced by, and productive, myriad corporeal practices. Tourists consume a map of historic Fredericksburg in conjunction with other texts, with other tourists and workers on the site, and through bodily practice as they move around. Meaning depends on action. They imagine the place before they go or recall their visit after; their gaze not only deploys the map during the visit, but also produces new geographical imaginings. The map is never complete; like identity it may be temporarily fixed, but it is always fluid and contested, deployed in new ways. A hybrid mapping emerges from this process, at once produced and consumed, not only cultured but also natured.

Other theorists also argue that mapping practices call spaces into being and that maps themselves are ontogenetic, rather than ontic. Instead of an emphasis upon the hybridity, practice is understood as always changing. Maps are always being called into being, to perform particular tasks, only existing when they are deployed, of the moment, dependant on a local context. Focusing upon the transformational qualities of practice, Kitchin and Dodge draw on the concept of transduction to explain how we constantly solve relational problems during all phases of mapping behavior. Like Hanna and Del Casino they stress that all mapping involves remapping: making and use can never be separated. They then deploy the notion of technicity, to explain mapping practice as an event involving people and technology, with processual outcomes, always emergent, in which representation is mutable.

For some, this relational and contextual turn has been theorized as Deleuzian. This kind of approach also reflects a philosophical shift toward mobility, and away from essence, in which mapping is imagined as a suite of cultural practices involving action and affects. James Corner recognizes that this bestows an agency on mapping, embodying it with creative power. Instead of a mere tracing of the world, mapping becomes rhizomatic. It becomes a powerful, open ended, indeterminate process that unfolds and makes new places and actions.

Some researchers have used this kind of approach to focus on the significance of movement in the construction of identity and used this to critique a representational approach to mapping and way finding. A more than representational approach to mapping becomes possible. For example, Tim Ingold repositions mapping as storytelling about journeys, in which the map becomes an inscription, that possibly emerges from a reenactment of mobility, rather than reflecting the existence of any preexisting cognitive structure. Others have stressed the affectual qualities of mapping practices focusing on the power of practice to move, and on the emotional geographies that emerge from performing mapping and geovisualization.

Mapping is also mobile and networked, an outcome of a technology that allows cultural messages to be transformed and shared. The development of 'actor network theory' (ANT) in science studies encouraged a consideration of the transformative potential of technologies and also focused attention on empirical consideration of the performance of mapping. Flows between different actors and actants can be traced, and a thick description of the context can lead to a relational understanding of mapping practice. Researchers not only must trace the inscriptions left behind by different actors, but also need to carry out ethnographic work in the field. Research can then chart the translations, through which mapping works, to reveal the ambiguities of mapping practice.

Performative Methods

Scientific practice seeks to control empirical investigation, simplifying complexity and establishing carefully regulated experiments into narrowly defined aspects of mapping, with a predominantly quantitative methodological emphasis. Discursive approaches, in contrast, focus largely upon analysis of the map as a textual source, amenable to critical approaches, many of which draw inspiration from qualitative ways of knowing the world. In contrast, the turn toward performance implies a change in research strategies. Research is starting to

Figure 1 Mapping as a social activity.

employ ethnographic approaches to carry out empirical investigations of everyday mapping practices in the field. An ethnographic approach sees mapping as a social ac tivity, rather than an individual response. By observing and participating in the performances around mapping we can explore its relations to identity, how different spaces are co constructed, and the ways in which people behave when carrying out mapping tasks. Ethno methodological work is beginning to yield rich de scriptions of the actions preformed in, and around, the mapping process. Brown and Laurier, for example, chart the social use of mapping in everyday car and tourist navigation activities, in which the mapping process is cast as a complex negotiation of identity in a social context. Map reading depends upon the context of the task: ethnomethodological observation reveals much more at stake than individual cognitive map reading (**Figure 1**). The previously taken for granted social actions and interactions in the field are recorded, revealing complex everyday behavior. Film, video, or photo diaries can be coded, translated, and interpreted, and a rich empirical exploration of actions becomes possible, if the researcher is deeply immersed in the mapping context. Situated and locally contingent ways of approaching mapping also depend on auto ethnographic methods and a reflexive appreciation of the role of the researcher in practice. Mapping narratives are there to be told if mapping be comes a set of social actions.

The Performative Tradition

In the multivolume *History of Cartography* project, a catholic definition of mapping has been popularized, which goes well beyond the Western metrical tradition of the map as an object. Instead, Woodward and Lewis argue that mapping comprises three different but interrelated traditions: one grounded in material culture leading to the creation of real maps and which previously domin ated research; a cognitive tradition in which mapping is imagined, deployed as a mental image to help make sense of the world; and a social tradition in which mapping is performed, by telling a story, recalling a dream, per forming a dance, singing a song, or enacting a ritual. Spatial and social meaning in this tradition is com municated through movement, gesture, words, music, narrative, and action. Performative mapping may make ephemeral traces, such as chalked map on a wall or a route mapped in soft mud, but is often strongly embed ded in cultural practice. So the performative tradition reinforces a more symbolic role for mapping. Mapping comes to embody cultural values and reinforces par ticular practices. It becomes a powerful agent, formu lating social cohesion or difference, influential in how we live in the world and form a sense of place.

This performative tradition is most obviously enacted in indigenous mapping, in recent Western technological interactivity, and in artistic practice.

The Indigenous Tradition

Most research in this performative tradition emphasizes cultures outside of existing Western orthodoxies. Map ping traditions either reside in evidence from the past or are uncovered by anthropologists engaging with sur viving marginalized practice. Evidence of nonmaterial mapping is less likely to have survived or is harder to establish. It is the artifactual maps that have been pre served and institutionalized.

Nevertheless, an impressive amount of evidence sur vives about the history of performative cartography. Im portant comparative studies have been carried out, and we now appreciate the rich performative tradition from first peoples in the Americas, Africa, Asia, Australasia, the Arctic, and the Pacific. A holistic and unified worldview is often revealed in these performative indigenous map pings, which avoids the dualisms of modern material maps. Art and science merge in practice; myth and ritual make the mapping, but are also made by it; landscape and the event are not separated; time and space become relative and culturally enacted; and religious and secular beliefs combine.

For example, aboriginal mapping in Australia operates with different conceptions of time and space to those that govern Western maps. Song lines, or the tracks made by the ancestors across the continent in the Dreaming, are celebrated and perpetuated in art, dance, and story telling. They serve as a complex and practical oral tradition, not only facilitating way finding, but also symbolizing an individual's obligation and reciprocal relationship to the land. Other performative traditions also serve practical and more abstract cultural roles, without recourse to direct representation of the world. For example, Melanesian navigators were able to travel vast distances across the Pacific, identifying and locating tiny atolls. In the Marshall Islands stick charts served as part of a performative and heuristic navigational tradi tion. The charts were at once indexical and egocentric: they served as a framework studied before a voyage, placing an individual in relation to an abstract grid, and allowing navigators to learn about swell patterns that might be encountered on a voyage. Way finding on the journey depended upon an embodied experience, a combination of sensing the refraction of swells in re sponse to changing bathymetry, and also celebrating memory.

Encounters between local performative traditions and modernist colonial representational mapping have also been well documented. In many cases it is only through these encounters that evidence has survived of other traditions. Of course the performative, oral tradition may only be documented through a Western written heritage, so historians have had to be cautious in their interpret ations. Contemporary ethnographic research into per formative indigenous cartographies is increasingly considering non navigational and metaphorical relations between maps as objects, and more performative tradi tions, such as the ways in which stories about the world might be related to embodied spatial practice.

Everyday Western Mapping Performance

Mapping in Western societies is nowadays a ubiquitous and everyday part of people's lives. Maps of the real world vie with imaginary worlds mapped in fantasy novels; media mapping grounds news narratives; journeys are planned; wars fought; and cities imagined. Space is described and knowable. Maps are everywhere in visual culture. The vast majority of mapping experiences in developed Western societies are, however, nowadays enacted through an electronic medium, distributed and shared over the web, or viewed on a television or on stand alone personal computers. Technology not only allows much greater control but also makes complex performance the norm. Mapping is now interacted with, and frequently changed, by readers, who are also pro ducers. Printed maps are increasingly customized to meet particular desires, and the published hard copy is in creasingly being replaced by the ephemeral display. Animated mapping is now commonplace. People zoom in, pan around a map, and perform simple GIS operations such as selecting a layer. Mass collaborative cartography is performed by a growing number of activists, creating mashups based on Google and other map portals, linking their own location based data to web served backdrops (**Figure 1**). People employ mapping in social situations where it would not have appeared even 10 years ago.

In the past maps largely supported work, rather than play. Specialist printed leisure mapping met particular market needs: a tourist map for holidays, a canal map, a distribution graphic for birders, etc. The printed map formed a backdrop for some board games, around which activities could take place. Or the map jigsaw might be reassembled. Only orienteering mapping was created by users. Playing with maps has, however, changed funda mentally, because of digital interactivity, networking, increasing real income, and new media with which to play. Computer gaming encourages new mapped worlds to be designed, and shared between players, and the design of mapping has itself become a game for many. New outdoor pastimes are emerging around this theme, such as geo caching, a kind of GPS based treasure hunt, and geo graphing, collaborative web served collecting of photographs of grid intersections.

These shifts toward mass availability, collaboration, interactivity, changing design and content, ephemerality, and play have led to increases in the number of people mapping. These people are more likely to be part of the map, instead of being its subject. The technology allows greater control and surveillance, but the mapping per formances it supports have also encouraged emancipation and individual agency.

Embodied Mapping

In the seventeenth century heyday of decorative Euro pean cartography artists and mapmakers shared similar concerns. The subjective world of the artist employed the map as a scientific device to poke fun at reason: im aginary maps in seventeenth century novels, the ubi quitous use of mapping metaphors in Shakespearian drama, and satirical images such as the Foolscap World Map critiqued the all knowing myth of professional scientific mapping (see **Figure 2**). These critiques are clearly revealed in allegorical mapping, and especially in the genre of embodied mapping.

Embodied maps draw analogies between corporeal form and cartographic depiction of geographic shapes. People came to represent and personify places. Rulers literally appear as the map of their nation, the coastlines recognizably morphing into familiar human forms. The political messages in these embodied maps depend on their context. Elizabethan personification of the Queen as the nation reinforces the power of monarchy; Victorian satirical embodied maps trade on derogatory national stereotypes, supporting imperial discursive constructions of nationality, that speak to Anderson's notion of

Figure 2 The Foolscap Map.

imagined communities. The genre continues to be popular among modern artists employing mapping in their work. The forms of the body are sometimes represented in a map like fashion, contoured or gridded, cartographic forms juxtaposed with the subject, in a corporeal contrast. Or else the body may be warped, transformed, and projected, as in the work of Lilla LoCurto and Bill Outcault, where it frames questions about visibility and representation.

Mapping is also deployed on the body: map tattoos move as muscles ripple, activating the apparently static landscape. Artists working with the visual form have also deployed mapping as clothing to drape around the body, and clothing itself depicts the cartographic form, T shirts selling places through their outline.

The body also appears on decorative maps of the early modern period, often juxtaposed to landscapes laid bare for colonial or European conquest. Here the naked, native, female form is often presented as eroticized adornment in the cartouche, reinforcing the narrative and gendered power of the map as a male discourse, naturalized and juxtaposed to clothed male explorers, or symbols of European commercial power.

(E)motional Mapping

The increasing academic interest in affect has a parallel in the genre of emotional maps, which chart human feelings onto a cartographical landscape. These maps move their readers, transporting them across terrain, guiding them on an emotional journey through life. The landscape in these maps often serves a moral purpose, mapping out accepted or deviant behavior. The most studied example of the genre, Madame de Scudéry's Carte de Tendre (**Figure 3**), for example, maps a landscape displaying relationships and feelings, which acts as a metaphor not only for an imaginary novel, but also for experience. Artists continue to make emotional maps, which depict our lives but also question how we live. For example, Kathy Prendagast's work highlights questions of identity and its links to landscape, by engaging with the political significance of place names. Recently, technology has allowed an even more performative mapping to be enacted within this genre. The World of Experience web site allows users to devise and customize their own emotional landscape, choosing what kinds of thoughts or experiences and feelings or passions to map, save, and perhaps purchase, through a simple Flash based application. Corporate versions of this approach are also marketed from www.companymap.com, allowing the quest for profits, to be mapped and customized.

Performative Artistic Mapping

Since the 1980s avant garde artists have increasingly employed mapping in their works, as a device to draw attention to the ways in which we relate to the world. Mapping is now more socially ubiquitous than ever before. It is hardly surprising that artists have deployed these images to say things about our society and its values. Surrealists, pop artists, situationists, land artists, conceptual artists, community artists, digital media artists, and live artists have all employed maps in their work, encouraging a performative encounter. Common techniques include: fragmenting known maps and rearranging them in novel ways; juxtaposing far with near; distorting space into a relative or egocentric form; emphasizing networks and encounters; changing orientation;

Figure 3 Carte de Tendre.

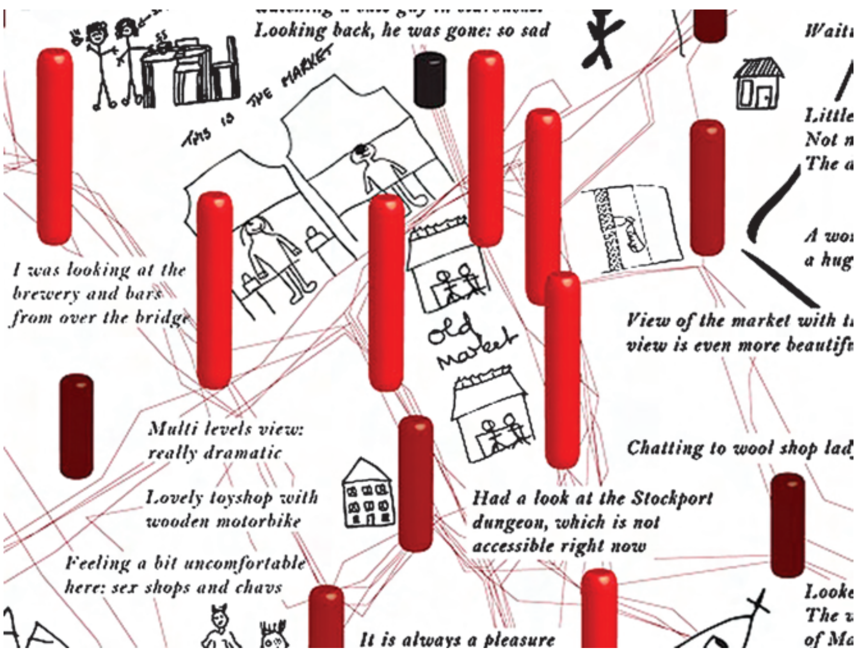

Figure 4 Emotion map of Stockport.

manipulating projection, scale, and generalization to infringe accepted mapping standards; drawing on standard cartographic tropes such as the border, or naming to question social norms; abstracting and overcoding a known form; employing recognizable country shapes in new ways; shifting novel conceptual frames onto familiar icons such as the globe or tube map; and mapping onto different media so as to ask questions about the world or our identities.

Many of these artistic encounters with mapping explicitly focus on the performative potential of the medium. The Situationists in the 1950s and the 1960s enacted a strongly oppositional cartography, moving art from the galleries to the streets of Paris. Guy Debord proposed psychogeographic encounters with the city and encouraged people to take part in dérives, or drifts, walking across urban space, enacting random but creative encounters, that subverted the controlled modernist dream of the city. Situationist maps were published as collages, fragmented, multimediated alternatives to the all knowing pictorial image on which they were based. This performative cartography has continued to inspire urban opposition ever since. Artists enact and perform many different mapping encounters. For example, Jake Barton creates public installations, which employ mapping as a way of promoting social interaction and collaborative storytelling. His maps in digital archives, street kiosks, and museum installations lure people and persuade them to tell each other about their own experiences, and in so doing destabilize accepted spatial and social categories. They serve as digital and collaborative memory maps. Live artists also enact mapping, for example, by layering and reenacting one set of alien events against a familiar mapped place. Others devise mapping puzzles to encourage urban exploration, casting the user as a player in an evolving work.

Locative art is emerging as particularly performative. Digital and satellite mapping technologies supporting networked, portable, location aware computing devices allow user led mapping, social networking, and artistic interventions, to be enacted on the canvas of the real world. People can track themselves, across a landscape. Traces can then be overlain in novel ways. In skywriting, for example, the old certainties of land art are subverted: participants walk the shape of letters or the outlines of animals, and, instead of the landscape itself being shaped as art, unseen messages are overlain onto mapping. Others deploy technologies to chart changing feelings in space: Christian Nold, for example, has devised biomapping technologies, merging GPS and biosensors, to produce collaborative emotion mapping (**Figure 4**). The worlds of art and science are once again converging through mapping.

Conclusion

Performative and embodied mapping not only depends on local contexts, but also makes these places. It suggests an optimistic possibility for creating new futures, in which human agency is recovered. Technologies become empowering instead of controlling. The visual can be deployed in new embodied ways. On the one hand, all mapping can be seen as performative, enacted, and creative: called into being to meet particular human needs, flowing from action, instead of being grounded in power; networked and relational, capable of moving us all. Dead representation becomes alive. On the other hand, some mapping is more likely to be performative. Indigenous traditions reveal a rich diversity of mapping practice; networked technology led interactive mapping is facilitating newly performative mapping, and artistic practices also reveal this performative potential, sometimes in explicitly performative actions and sometimes in the creation of artifacts that encourage action or change.

Embodied and performative mapping opens possibilities for a coming together of action and theory in critical praxis. This kind of approach to mapping brings the academic world closer to those of the creative artist and activist. It shows how mapping can be at once social and empowering.

See also: Art and Cartography; Embodied Knowing; Ethnography; Mapping, Philosophy; Performativity.

Further Reading

Corner, J. (1998). The agency of mapping: Speculation, critique, invention. In Cosgrove, D. (ed.) *Mappings*, pp 213 252. London: Reaktion Books.

Cosgrove, D. (ed.) (1998). *Mappings*. London: Reaktion Books.

Del Casino, V. and Hanna, S. P. (2006). Beyond the ‘binaries’: A methodological intervention for interrogating maps as representational practices. *ACME: An International E Journal for Critical Geographies* 4(1), 34 56. 35. http://www.acme journal.org/vol4/VDCSPH.pdf (accessed Jul. 2008).

Harmon, K. (2004). *Wish You Were Here: Personal Geographies and Other Maps of the Imagination*. New York: Princeton Architectural Press.

Ingold, T. (2001). *The Perception of the Environment*. London: Routledge.

Kitchin, R. and Dodge, M. (2007). Rethinking maps. *Progress in Human Geography* 31(3), 331 344.

Kwan, M. P. (2007). Affecting geospatial technologies: Toward a feminist politics of emotion. *The Professional Geographer* 59(1), 22 34.

Perkins, C. (2004). Cultures of mapping: Power in practice. *Progress in Human Geography* 28, 381 391.

Perkins, C. (2006). Mapping. In Douglas, I., Huggett, R. & Perkins, C. (eds.) *Companion Encyclopaedia of Geography* (2nd edn.), pp 555 571. London: Routledge.

Pickles, J. (2004). *A History of Spaces: Mapping Cartographic Reason and the Over Coded World*. London: Routledge.

Wood, D. and Krygier, J. (eds.) (2006). *Special Issue: Art and Mapping Theme Issue. Cartographic Perspectives* 53.

Woodward, D. and Lewis, G. M. (eds.) (1998). *The History of Cartography, vol. 2, Book 3: Cartography in the Traditional African, American, Arctic, Australian, and Pacific Societies*. Chicago, IL: University of Chicago Press.

Relevant Websites

http://www.acme journal.org
ACME: An International E Journal for Critical Geographies.

http://www.companymap.com
CompanyMAP®.

http://www.worldofexperience.com
World of Experience.

Performativity

D. P. McCormack, Oxford University Centre for the Environment, Oxford, UK

Glossary

Discourse A set of representational practices and performances through which meaning is given a degree of stability.

Identity The recognizable and shared meaning ascribed to an object, person, place, or event.

Performativity The quality of practices that means they are generative of the world rather than merely reflective.

Representation Something (typically an image or symbol) which mediates our experience of the world by standing in for the objects and events which constitute this world.

Subject The center of human agency and intentionality, often understood to be located within the mind.

Introduction

The importance of the concept of performativity within human geography can be situated within the emergence, during the 1980s and 1990s, of an interest in questions of representation and its role in shaping the relation between space and culture. Responding in part to what they saw as the limits of the kind of cultural geography associated with the work and influence of Carl Sauer, a generation of geographers began to foreground the importance of representations in understandings of space and place. Such geographers began to pay particular attention to the role that representations played in shaping the meanings and 'identity' of places and their inhabitants of those places. At the same time, the authority and stability of representations were also questioned: rather than fixed things through which the world was reflected, representations were understood increasingly to be unstable, partial, and contestable, and within the context of wider discursive formations. A corollary was that the identity of place and people was also unstable and contested.

From the early 1990s, such efforts found an important source of conceptual support in the work of the North American philosopher and cultural theorist Judith Butler, and specifically her ideas about performativity. For geographers, particularly feminist geographers, this concept provided a crucial way of understanding the contested and politicized relations between the identity of bodies and the identity of place. Yet while Butler's ideas about performativity continue to be influential within the discipline, they can be and have been challenged, particularly insofar as they seem to place too much emphasis on the discursive construction of difference and identity. Thus while important, any discussion of the concept of performativity within human geography must also attend to the other traditions of thinking through which the concept is articulated. In this respect geographers have drawn upon the work of philosophers such as Gilles Deleuze in order to produce an understanding of the performative dimensions of space and spatiality that modifies the discursive emphasis of many interpretations of the work of Butler.

Performativity Matters

Given the influence of her work, it is worth beginning by outlining the key elements of Butler's writing on performativity. Butler's discussion of this concept can be understood as a key element of her critical genealogy of the identity of the human subject, particularly in relation to questions of sex/gender. Much of Western thinking and philosophy can be understood as an attempt to establish a relatively fixed conception of the subject as the basis for proper political action and thought. This subject is presumed to be a rational, coherent, self aware actor. It is also a subject with intentionality – the subject can exert a strong degree of agency over its action. Put another way, the individual subject is a kind of sovereign power. Crucially, the identity of this subject is fixed: it has an essence that is not contingent upon the changing circumstances (cultural, social, and political) within which it finds itself. While this conception of the subject has become taken for granted – to a degree at least – its apparent naturalization actually conceals its dependence upon a certain set of exclusions. The rational, mindful subject is defined in opposition to that which it is not: irrational, emotional, and bodily. Furthermore, it has come to be defined as implicitly masculine, and many of the qualities it excludes have tended to be identified as feminine.

Seen in this way, the apparent stability of the identity of the subject can be unsettled by resituating it as something dependent for its meaning upon a complex set of relations with other identities. It is exactly such unsettling of the apparent fixity and identity of the subject that we find at the heart of Butler's work. In developing her critique of the subject, Butler draws upon a range of influences, including the dialectical philosophy of Hegel;

psychoanalytical theories of desire; and post structuralist theories of discourse and language. Her critical geneal ogy of the subject owes a particular debt to the work of Michel Foucault, for whom discourse was not just a system of neutral representations but was a series of re lated statements whose repetition is generative of certain stabilities of meaning across a range of practices. Understood in these terms, the identity of the subject is not borne during a kind of foundational moment: it does not preexist discursive practices but is the effect of such practices. Butler is particularly interested in the impli cations of such ideas for how we think of gender identity. Again, rather than the essential quality of an individual possessed from birth, gender identity emerges through the individual being addressed (as 'boy' or 'girl') from a very early age. Here Butler draws also upon the idea (derived from elements of the work of thinkers such as Austin and Derrida) that speech itself is performative – it is generative of the thing it merely purports to name.

An important upshot of this claim is that the identity of the body is neither fixed nor given. Rather than a mere thing, the body comes to matter through the repeated stylization of its features or gestures – through ways of behaving, talking, gesturing, dressing, etc. For Butler this also means that it makes little sense to make a sharp distinction between sex and gender: indeed, she is critical of those feminists who make this distinction. Any identity, including sex, emerges through the repetition of cita tional practices within discourses. The important point is that even if Butler seems to make room for the possibility of subverting such identities, it is not the individual who is performative, but the discursive practices within which the individual finds her identity articulated: discourses 'perform' the subject. Such ideas about the peformativity of gender identities could also be employed to think about questions of race, class, ethnicity: rather than naming a set of distinct and fixed set of identities, the contested meaning of each category is located within a set of performative discourses.

Performativity, Identity, and Space

For geographers, Butler's work provided a way of thinking about the relations between the performativity of identity and questions of space. Indeed, early readings of Butler work stressed the need for geographers to 'spatialize' performativity. An early and controversial example of this was a piece by David Bell, John Binnie, Julia Cream, and Gill Valentine in which the authors explored the rela tionship between sexuality, space, and identity. Butler's work has also allowed geographers to think about the relation between discourse and economic processes. Ra ther than identifying real underling structural forces and mechanisms, the 'economy' can be seen as a category whose ongoing citation stabilizes the identity of a diverse set of forces. Furthermore, with this economy, work often involves becoming implicated in highly stylized bodily performances differentiated in highly gendered terms through modes of presentation and adornment. Butler's work has also been used to raise important question about the production of geographical knowledge itself. As Gil lian Rose argued in a key article about the politics of doing research in the wake of feminist critiques of the subject and identity, it became difficult for geographers to employ the tactic of self reflexivity and the disclosure of positionality when any appeal to knowledge of the 'self' seemed to rely upon the kinds of subject positions about which Butler's work is so critical.

While highly influential, the use of Butler's work by geographers has been the subject of various critiques and debates. For instance, Lise Nelson argued that Butler's account of performativity – and its adoption by geog raphers – leaves little scope for any conception of transformative agency. Thus, by equating agency with a kind of abstract, ideal, and masculine subject, Butler seems unable to account for the possibility of active, intentional reflection from which various forms of re sistance to hegemonic discourses might emerge. This critique also complicates Butler's reading of various kinds of apparently subversive forms of identity performance through, for instance, dressing across genders – to what extent are such performances to be explained by ap pealing to a kind of intentional subject which itself is the target of Butler's critique? In turn, Nelson's critical commentary has been challenged insofar as it mistakenly reads Butler as rejecting any appeal to agency. Thus, for Catherine Nash, Butler encourages us to develop a more sophisticated account of the relation between agency, identity, and discourse.

To such debates we might add two further substantive critiques of Butler's conception of performativity. First, by equating agency with subjectivity Butler's work seems unable to account for the more distributed understanding of agency we find for instance in recent research in human geography, particularly work influenced by that set of ideas known as actor network theory. Such work suggests that it is possible to hold on to some sense of agency or actancy without necessarily appealing to in tentionality as the sovereign source of such agency. Then, and second, by emphasizing repeatedly the danger of appealing to the pre discursive in critical accounts of subjectivity, Butler arguably limits the range of practices and processes upon which geographers and others might draw in articulating new ways of thinking about the performative or political aspects of space and spatiality. How, for instance, might we make sense of spatiality of affective processes which seem to have some kind of tangible, visceral force before they are ever named ex plicitly through language?

Making More of Performative Movements

One way to focus the increasingly diverse conceptions of performativity circulating within the discipline is through the example of specific kinds of performance practices. Dance has figured strongly here. For instance, as part of his elaboration of nonrepresentational theory, Nigel Thrift uses dance as a practice that serves to exemplify the performative dimensions of bodily space and spacing. For Thrift, it is precisely the way in which dance works upon the precognitive or nonrepresentational dimensions of bodily thinking that gives it a performative quality. Such claims seemed to move against the grain of geographical work influenced by Butler, insofar as they decouple performativity from representation, and the critical/political purchase a focus on the latter term has provided geographers. Hence, therefore, the concern expressed by Catherine Nash in a review of how ideas about performativity had been put to work by Thrift in thinking through practices such as dance. For Nash, Thrift seemed to be using dance to appeal to the possibility of escaping the kinds of discursive practices that had been rendered more tangible through Butler's work. As such he seemed to run the risk of eroding the very political gains that feminist and cultural geographers had worked hard to make in the wake of various readings of Butler's work. Furthermore, as Donna Houston and Laura Pulido argued, Thrift's apparent focus on the ephemerality of practices such as dance seemed to work against and affirmation of more obviously political performances through which the performative identity of space and place was contested and subverted.

It would be a mistake however to posit any neat distinction between an emphasis on performance as the escape from real politics and a more politically attuned conception of performativity derived from the work of Butler. Thrift's argument, refined in more recent writing, is that there are more ways of making performativity (and performance) political than that offered by Butler. At the same time Thrift has drawn attention to some important limitations in Butler's work. These limitations include the fact that Butler is primarily a theorist of the symbolic registers of cultural life; that her work tends to equate political change with active interventions made by 'marginalized' subject positions; that it says little about space; and that it remains inattentive to the potential positivity of certain practices from which can emerge various kinds of distributed agencies and forces for change.

Thus, while Butler provides one set of conceptual apprehensions for thinking through the question of performativity, there are others. These include, for instance, the tradition of ethnomethodological approaches to the performance of everyday life, exemplified in the pioneering work of Erving Goffmann. But in working up an account of the nonrepresentational dimensions of the performativity of space and spacing, geographers have also drawn upon a much wider tradition of thinking, making particular use of the work of the philosopher Gilles Deleuze. J. D. Dewsbury elaborates upon the contribution Deleuze's thinking makes to any conception of performativity. In contrast to Butler's conception of performativity as a function of the repetition of certain kinds of representational practices, Deleuze draws attention to a conception of difference that does not depend upon such repetition. For Dewsbury, following Deleuze, thinking the performativity of space – or spacing – requires us to come to grips with a conception of difference as ongoing differentiation without representations. It also means that the event of performativity is always excessive of attempts to capture it in representational form. In developing this discussion, Dewsbury problematizes any neat distinction between the concept of performance and performativity, and particularly any attempt to define performance as a set of practices presupposing an intentional subject.

So, to return to dance, while such a practice can still be understood in terms of its articulation through discursive processes, it also always opens onto registers of experience that are more or less than representational. Furthermore, the performance of dance is never simply a matter of individual creation – it can be and is emergent from the relations between bodies, texts, technologies in ways that complicate any subject centered account of agency. Thinking through dancing bodies does not so much mean ignoring questions of gender: it encourages us to think of the moving body as a site through which multiple processes of differentiation at a range of durations and intensities are engendered. Some of these differentiated engenderings may, through their recognition and articulation through webs of discursive signification, coalesce into recognizable identities. Others may not, continuing to exceed the terms of representational politics by facilitating various forms of relations within and between bodies.

An important point that needs to be made here is that nonrepresentational accounts of performativity do not just expand the range of the processes and practices to which geographers attend. They also encourage us to make more of the performative character of the very process of this knowledge production. On one level, this means thinking of many of the more familiar techniques of knowledge production (maps, diagrams, and images) less as devices of representational capture than as performative operators – things that do work rather than arrest the 'doing'. It also means making more of the key registers through which knowledge is presented – writing and speaking. Indeed insofar as writing is performative, one of its key tasks might be to elicit those qualities of difference which are always excessive of textual strategies. One way to do this might be to fold the performative intensity of various practices – such as dance – into the style of writing and presentation. Such writing would never be about capturing the movement of dance, but about developing ways of allowing the qualities of movement to animate the logics of geographical thinking.

And More to Come…

In many ways, the concept of performativity is still working its way through the discipline. It continues to offer ways of agitating and altering theorizations of space and place, and beyond the obviously cultural corners of the discipline. For instance, political geographers have drawn upon the concept in order to rethink the ways in which the performative articulation of certain imaginative geographies as part of the ongoing 'war on terror' actually produce the effects they purport merely to name. At the same time, feminist geographers are continuing to refine Butler's critique of materiality as part of a wider attempt to rework conceptions of the lively matter of bodies. Elsewhere, the concept is being used to rethink the way in which various forms of code and inscription contribute to the operation and organization of everyday space times. Here, drawing upon insights about the dynamic character of technological systems, geographers such as Rob Kitchen and Martin Dodge have argued that mapping actively transforms both itself and the places it purports to 'represent' in a performatively iterative process.

Cutting across these various employments and reworkings of performativity is an ongoing affirmation of a dynamic conception of space. This is perhaps the key contribution made by the concept to human geography – an enlivening of the very matter and meaning of spatiality. Rather than a static container within which things happen, spatiality has, in part because of the influence of notions of performativity, come to be seen as an ongoing contingent and contextual process, emergent through a range of practices and processes. That is not to say that ideas of performativity mean that space is always up for grabs at any given moment. To say that space is performative is not to claim that anything can happen: rather, it is to become open to the possibility that we do not quite know what might happen, where, how, or when. In these terms, the concept of performativity remains an important and lively participant within the practices of which the discipline(s) of human geography consists, even if the precise ways in which it participates are always matters in transformation.

See also: Becoming; Fluidity-Fixity; Identity Politics.

Further Reading

Bell, D., Binnie, J., Cream, J. and Valentine, G. (1994). All hyped up and no place to. *Gender, Place and Culture* 1, 31 47.

Butler, J. (1993). *Bodies That Matter: On the Discursive Limits of 'Sex'*. London: Routledge.

Colls, R. (2007). Materialising bodily matter: Intra action and the embodiment of 'fat'. *Geoforum* 38, 353 365.

Dewsbury, J. D. (2000). Performativity and the event: Enacting a philosophy of difference. *Environment and Planning D: Society and Space* 18(4), 473 496.

Dodge, M. and Kitchin, R. (2005). Code and the transduction of space. *Annals of the Association of American Geographers* 95(1), 162 180.

Gregson, N. (2000). Taking butler elsewhere: Performativities, spatialities and subjectivities. *Environment and Planning D: Society and Space* 18, 433 452.

Houston, D. and Pulido, L. (2002). The work of performativity: Staging social justice at the University of Southern California. *Environment and Planning D: Society and Space* 20, 401 424.

McCormack, D. P. (2005). Diagramming practice and performance. *Environment and Planning D: Society and Space* 23(1), 119 147.

McDowell, L. and Court, G. (1994). Performing work: Bodily representations in merchant banks. *Environment and Planning D: Society and Space* 12, 727 750.

Nash, C. (2000). Performativity in practice: Some recent work in cultural geography. *Progress in Human Geography* 24(4), 653 664.

Nelson, L. (1999). Bodies (and spaces) do matter: The limits of performativity. *Gender, Place and Culture* 6(4), 331 353.

Rose, G. (1997). Situating knowledges: Positionality, reflexivities and other tactics. *Progress in Human Geography* 21, 305 320.

Thrift, N. (1997). The still point: Resistance, expressive embodiment and dance. In Pile, S. & Keith, M. (eds.) *Geographies of Resistance*, pp 124 151. London: Routledge.

Thrift, N. and Dewsbury, J. D. (2000). Dead geographies and how to make them live. *Environment and Planning D: Society and Space* 18, 411 432.

Phenomenology/Phenomenological Geography

G. Backhaus, Loyola College in Maryland, Baltimore, MD, USA

Glossary

Back to the Things Themselves The Husserlian slogan of phenomenology that indicates the process of bracketing (putting-out-of-play) all theories, models, philosophies, etc. in the attempt to faithfully describe things just as they present themselves (appearances) in experience (intuition).

Eidetic Phenomenology A specific form of phenomenology that relies on imaginative variation in order to remove inessential characteristics leading subsequently to the apprehension of the essence of an object. It is to be contrasted with empirical science, which must remain faithful to perceptual manifestations of phenomena and thus is limited to empirical generalizations.

Essences or Eide Material *a priori* principles that are necessarily present in order for something to be that which it is; morphological essences are the correlates of phenomenological description. These are contrasted with exact essences, or the idealities of mathematics, which are components of quantitative explanatory science.

Geographical Lifeworld According to Seamon, it is comprised of the taken-for-granted level of meanings, experiences, behaviors, and events in relation to an environment, space, place, and landscape.

Hermeneutic Phenomenology Heidegger's examination of the ways that Dasein, the being for whom its own 'being' is at issue, interprets 'being'. According to Mugerauer, hermeneutics allows for the expansion of phenomenological description into interpretation by taking into consideration the historical context of meaning constitution, which can be applied to landscapes, environments, places, and architectural elements.

Implacement According to Casey, the collusion of the lived-body and place constitutive of the emergent, precognitive, spatial meaning-structures of qualitative dimensionality and directionality.

Lived-Body Intentionality According to Merleau-Ponty, it is the precognitive relational field of meanings autochthonously arising from the experiential body in its situation/milieu, manifesting as gestural/postural enactments of meaning.

Neo-Kantian Formal *a Priori* Forms imposed on phenomena (inchoate contents) in an explanatory-nomological science serving to construct a knowable reality.

Objectivism The viewpoint of scientific detachment that encompasses both the materialist reduction to quantifiable abstractions and the positivist restriction of knowledge to observable facts, atomistic propositions, and empirical reality.

Phenomenological Material *a Priori* The eidetic (nonempirical) structures of phenomena constituted through the evidencing procedures of descriptive-phenomenological science.

Place According to Relph, the interconnection of natural and human processes constituting a person's immediate experiences and centralizing a person's dealings with the world.

Introduction: Humanistic Geography and Phenomenology

Humanistic geography is concerned with geographical experience – what the spatiality of the earth/world means to those whose lives are part of it – in contrast to those who study geography as externalized, objective data. Geography is not merely a dimension of human life, but rather pervades it, and thus the core argument is that geography is humanism. Phenomenology offers a radical methodology for description that opens the field of meaning constitution to investigation, displaying the evidence upon which meaningful and valuable entities, events, and the limits of the experiential horizon appear within experience. Thus, phenomenology offers a methodology that can open up the full spectrum of the spatial inscriptions of human life, the fundamental concern of humanistic geographers. Geographers advocating a humanistic turn developed an interest in phenomenology in the late 1960s, followed by intense productivity in the 1970s and early 1980s. Publications in humanistic geography can be placed in roughly six categories: articles wrestling with theoretical issues relating humanistic geography with phenomenology (e.g., those by Buttimer, Entrikin, Mercer and Powell, Relph, Samuels, and Tuan); articles identifying a quasi phenomenological turn in geographical literature of predecessors and contemporaries (e.g., Sauer, Lowenthal, Gregory, Kirk, and Wright); textbooks and encyclopedias devoting space to humanistic geographers' focus on phenomenology (e.g., Ley and Samuels); hybrid treatises (e.g., Warf's synthesis of humanist and structuralist perspectives, and Schatzky's synthesis of Heideggerian spatial ontology with the spatial causality of explanatory science); works that employ humanistic geography as a component (e.g., Porteous' landscapes of the imagination); and full length monographs representative of the position (see readings).

The renewed interest in lived geography of the 1990s revolved around several substantive issues such as de veloping a phenomenological environmentalism, for example, David Seamon's humanistic geography/archite cture of built space, Robert Mugerauer's various meth odologies for interpreting environments, and Edward S. Casey's phenomenology of place. In his 1996 text, *Sensuous Geographies*, Paul Rodaway develops the cross fertilization between humanistic and postmodern geographies. A re cent trend, enactive geographies, incorporates Maurice Merleau Ponty's existential phenomenological behavior ism as one of its influences. Editors Paul C. Adams, Steven D. Hoelscher, and Karen E. Till developed a critical hu manistic geography in their 2001 volume, *Textures of Place: Exploring Humanist Geographies*, which investigates global izing forces in light of place. The further development in methodological pluralism situates phenomenology along with humanistic geography as one approach among many in the study of a particular phenomenon (see, for example, Martin Selby's 2004 volume, *Understanding Urban Tourism*, especially ch. 6, 'The experience of urban tourism'). Lastly, in a more diffuse way, the spirit of humanistic geography pervades works that do not directly make reference to it, but call on the same phenomenological sources and ac complish the same kind of work (see, for example, Charlie Hailey's 2008 work, *Campsite: Architectures of Duration and Place*).

Humanistic geography emerges as both a reaction and a response to the hegemonic development of objectiv istic/naturalistic geographical science. Geographers committed to the humanistic turn, such as Yi Fu Tuan, Edward Relph, and Anne Buttimer, had been dis appointed by the objectification of human phenomena and they took up the project to develop the geography of lived experience. Objectivistic science alleges that human life is only accessible to scientific examination from the perspective of detachment. Objectivism's discounting of geographical experience is a vestige of the Cartesian double world theory in which the world as lived is taken to be merely subjective; the real, mechanistic, quantifiable world can be reached only through thought, not through life. The empirical version of this objectivist dualism is positivism, which emphasizes perceptible/observational facts, accounted for through causal explanation, versus nonobservable, unscientifically verifiable subjective experience.

Humanistic geographers were drawn to phenomen ology because it provided a philosophically articulated alternative to objectivist science, critically challenging its ontological dualism and scientific monism, offering methodologies congenial and adaptable to their geo graphical concern for the experiential dimension. Four difficulties immediately faced implementation. First, humanistic geographers had to learn phenomenology without background in epistemology, ontology, and the history of philosophy. Second, they studied phenomen ology within the mind frame that philosophical acumen is irrelevant to the goal of the human sciences. Third, they appropriated phenomenology to their respective humanistic bents, that is, their primary goals were not to become phenomenologists, but rather to forge their identities as humanistic geographers. Lastly, they had to learn to apply phenomenology to their subject matter in order to uncover the phenomenological structures of geographical experience. As with any new area, the introduction of phenomenology to geography brought about self reflective, critical discussions and accusations of misinterpretations and misunderstandings within their own group, for example, Relph's contentions with Tuan and Buttimer concerning the significance of humanism and a humanistic geography, and John Pickles' insistence on a rigorously orthodox Husserlian geography against other interpretations.

Phenomenology's Historical Context: Neo-Kantian Explanatory Science versus Husserlian Descriptive Science

Phenomenology arose amidst the conflicts of the nine teenth century. Just as soon as Immanuel Kant's critical philosophy had eliminated metaphysical speculation and had provided a compromise between rationalism and empiricism, the excess of metaphysical extremes emerged once again in idealism and materialism. On the one hand, G. W. F. Hegel created an 'absolute idealism' on the basis of pure speculation: reason viewed as the process of concretizing transcendental subjectivity in history. The uneven progression of 'spirit' is exhibited in Hegel's infamous geography of continents and nation states based on the relation of physical geography to culture. On the other hand, metaphysical materialism provided for scientific reductionism – all of being is re ducible to quantifiable, isolatable bits of matter and mechanistic laws. Naturalism, its less extreme cousin, reduces all of reality to natural objects, operational concepts, and explanatory principles. Natural science had been grounded in Kant's critical philosophy – the *a priori* principles of possible experience are at the same time the universal laws of nature – and thus is viewed as avoiding metaphysical commitment. Eschewing the extremes of nineteenth century idealism and materialism, the goal of the neo Kantians was to ground the newly forming human sciences (psychology, sociology, anthropology, etc.) on a Kantian basis. Edmund Husserl, the progenitor of phenomenology, shared this anti metaphysical con cern for scientific foundation. However, Husserl's phenomenological science profoundly differs from Kantianism – by its rejection of Kant's formal *a priori* through the development of the material *a priori* based on the apprehended evidence of phenomenological analysis. This formal/material distinction is the key to properly grasping the source for, and the difference between, the explanatory (natural) science adopted by the empirical (human) sciences, which the humanistic geographers reacted against, and the descriptive (phenomenological) science that they embraced. The neo Kantians maintained that reality as given is inchoate; in order for it to be known, the inchoate must be structured (formed) by cognition. These constructions within sociohistorical settings in turn remain scientifically unknowable until formed through the explanatory methodologies (model formations) of science. It is this hegemony of neo Kantian naturalism (all of being, *qua* knowable, is to be treated as an object, a natural entity) that Husserl (in his doctrine of regional ontologies, being is differentiated: eg., physical thing, pure consciousness) and the humanistic geographers found unsatisfactory.

For Kant, *a priori* synthetic judgments are possible on the basis of constructing concepts through their application to sensible intuition. Based on Kant's Copernican Revolution, that objects conform to the mind, the *a priori* structures of the mind impose form on reality through the construction of concepts on the basis of the given. Husserl calls this Kantian doctrine '*a priori* constructionism' and Kant's 'faculty psychologism' – mental functions constructing appearances. Husserl's battle cry 'back to the things themselves' and his various ways into phenomenology are attempts to apprehend the things by avoiding any admixture of Kantian constructionism. The neo Kantians' development of the human sciences advocates Kant's constructionism. Husserlian phenomenology instead opens up the field of the material *a priori* by which objects are constituted (not constructed) on the basis of intuition, the givenness of evidence within the intentional structure of consciousness. Kant's 'poverty' of 12 categories of the understanding and the pure sensible intuitions of space and time are replaced by Husserl with his doctrine of intentionality whereby a rich plethora of intentional acts are inseparably connected to a rich plethora of possible objects. Each mode of intentional act is related to its mode of intended object – an act of perception/aperceived object, an act of imagination/an imagined object. Each intentional mode exhibits its own eidetic (nonempirical or pure necessary *a priori*) structures. This opening of the field of phenomenological study is the most crucial point for the human sciences (and lived geography): either an epistemological path is cleared for a science of description, or science is limited to the construction of explanatory (causal) models. Without philosophical justification, the humanistic turn to lived geography (experience) would remain equivalent to lighting candles (speculative idealism), holding hands (soft minded preliminary to real science), and talking about feelings (solipsism); in other words, the generic positivistic criticism, along with those made by geographers such as David Harvey and Richard Morrill concerning the nonscientific character of phenomenology would stand. Accordingly, without phenomenology's rejection of objectivism, lived experience would remain at the merely subjective level of what Kant calls judgments of perceptions (what it feels like to me), whereas by contrast, naturalistic science for Kant involves judgments of experience (formed by application of *a priori* categories, e.g., cause and effect) and thus involves permanent and universal cognition, the necessary conditions for *scientia*. However, if the things themselves exhibit (material) *a priori* principles, *eidé* or essences, as Husserl maintains, then rigorous descriptions based on lived experience are a most proper realm of scientific enquiry – the apprehension of evidence through phenomenological analysis is a legitimate form of scientific investigation. Phenomenological description is to be contrasted with description within the naturalistic attitude that remains naively object focused. Husserl's phenomenological reduction (in its Cartesian mode) leaves consciousness as its residue through which the researcher remains solely occupied with appearings. Through the phenomenological *epoché*, which suspends all judgments concerning the spatiotemporal existence of things assumed to be 'out there', opens up the field of meaning constitution – how the world comes to appear.

Both Martin Heidegger and Maurice Merleau Ponty point out that the Husserlian doctrine of empty and fulfilled intentions, especially concerning categorial intuition, is the break with neo Kantian constructionism that opened the phenomenological field of meanings. For Kant, judgment forms are constructed by the *a priori* categories of the understanding. For Husserl, judgments are empty intentions that are to be filled through the intuitive givenness of categorial objects as presented in experience. Husserl's first work in static phenomenology dealt with eidetic description of an act and its object (act psychology). As act psychology is limited in scope, in that it does not consider context/horizon of both world and subjectivity, Husserl developed constitutive phenomenology: lived experience is 'thick', consisting of more than merely an act and object. The subjective context of experience (core fringe of attention, act complexes, etc.) and the objective context (a milieu with the world as horizon) are brought into phenomenological investigation. At the same time, Husserl elaborated eidetic phenomenology (all phenomenology is eidetic, but eidetic phenomenology is object focused), which involves the realm of the possible (apprehending essences vs. the empirical generalizations of actuality – positivistic matters of fact) through the methodology of imaginative variation. Husserl also extended his investigations to create a genetic phenomenology that accounts for the stream of consciousness (the temporal nature of con sciousness itself), which allows for the uncovering of temporal depth – sedimentations of meaning built up over time. Husserl later worked out the phenomenology of the lifeworld, the pre given structures of the world of experience prior to any of our theoretical/scientific (re)constructions – the world that is always already there. He offers a methodology of reflective history in his critical work on the crisis of European science. Husserl forges each of his ways into phenomenology with the goal of apprehending the things themselves through meaning constitution, carefully avoiding naturalistic scientific theoretical construction (the primacy of intuition over construction). The phenomenological slogan, back to the things themselves, involves bracketing the natural atti tude that informs the empirical sciences, and training eidetic seeing that opens the phenomenological field of meanings.

Other Major Phenomenologists That Influenced Geographers

Alfred Schutz developed social phenomenology on a Husserlian foundation in order to clear up ambiguities in Max Weber's notion of 'verstehen', distinguishing meanings that the social world has for a social actor (insider) from the meaning interpretations of an outside observer. Schutz puts forth a constitutive phenomenology of the natural attitude, the world of the everyday as it is socially structured. Schutz describes eidetic social structures: we relationship, thou orientation, intersub jectivity, anonymity, zones of operation, system of rele vancies, stock of knowledge, the taken for granted, etc. In terms of human/social science, Schutz was not so much interested in personal testimonies; he was interested in the formation of ideal types, second order typified social puppets that could be hypothetically situated by the human/social scientist and typically described in terms of characterology, course of actions, etc. The world of the social puppets must coincide with the typifications employed by social actors (insiders) living within the context of the actual social milieu. The social (actor) insider/(observer) outsider distinction greatly influenced geographers Relph, Buttimer, and Seamon, even though there was some confusion/conflation of the insider's perspective with Husserlian egology. Transcendental phenomenology and Schutz's constitutive phenomen ology of the natural attitude are to remain conceptually distinct – unfortunately, the confusions play into the positivist's critique of phenomenology as merely the subjective description of an insider (a story, not science).

Heidegger creates a hermeneutic phenomenology for the purpose of raising anew the question of 'being'. But this first requires the necessity of investigating the being for who its own 'being' (Dasein) is necessarily implicated. Critical of the metaphysical tradition, Heidegger claims that 'being' is not itself a being. The fundamental struc ture of Dasein is being in the world, which preempts the modern tradition of the subject–object dichotomy. It is the spatial ontology of worldhood, a constituency of being in the world that most influences humanistic geographers, for it articulates a nonphysicalist notion of space and provides the possibility for uncovering the existential significance of place/environs/umwelt. Posi tivism's ontological commitment to discrete things and isolable facts ('being' as beings) does not know what to make of Heidegger's holism whereby context or world horizon provides significance and the objects and facts so dear to positivism appear only in a secondary or de rivative mode of knowing, rather than in the primary mode of circumspection (existential engagement). Hei degger's later philosophy opens up the hermeneutic on the basis of language as the house of 'being'. Through an ontological poetics, Heidegger explores the root mean ings of dwelling and building and examines the con trasting essences of traditional and modern technologies.

Merleau Ponty, on the basis of Husserl's later cor rection of his earlier intellectualism, is led to the notion of lived body intentionality. In *Experience and Judgment*, Husserl develops a genealogy of logic where logical forms (the source for Kant's transcendental deduction of the formal *a priori* categories) are shown to be already present (nascent) in the 'cognitive basements' of per ceptual consciousness, thus further destroying Kant's formalism/constructionism. Merleau Ponty radicalizes this Husserlian doctrine of pre predicative experience into a matter pregnant with form doctrine whereby meanings are autochthonously birthed in the structures of embodiment. Lived body intentionality is precognitive and involves an ambiguous field of meanings that is re ducible neither to transcendental subjectivity nor to natural thing. Subject–object roles are reversible between the co constitutive poles of body sensibility and milieu. Merleau Ponty describes the senses as essentially spatial, a spatiality that enacts meanings in the lived body's en gagements and negotiations with its milieu. Space, then, is neither a Kantian *a priori* form to be constructed nor an objective container to be measured. Meanings emerge through spatial morphologies as they are manifest through the body schema. The body schema reflects the meaning of the milieu on the basis of the lived body's postures. Spatiality is an existential process of être au monde. Thus, behavior is not reducible to naturalistic accounts (human bodies are not merely 'natural objects'), which has huge implications for geographical study, both in developing research in behavioral geography and in critically chipping away at the alleged adequacy of the positivistic objectification of behavior.

Edward S. Casey has published two volumes con cerning the philosophy of place, *Getting Back into Place* and *The Fate of Place* that have influenced and stimulated the work of both environmental philosophers and geog raphers. Casey develops the notion of implacement, the collusion of the lived body and place. Casey's approach is highly informed by Merleau Ponty's intentionality of the lived body but further develops the spatial structures in the meaningful enactment constitutive of place. Investi gating the fundamental structures of dimensionalities and directionalities, providing a Merleau Pontian perspective on the notion of dwelling, and developing a phenome nology of wild places, Casey links up concerns of lived geography with environmentalism.

Other phenomenologists and existential phenomen ologists, for example, Bachelard, Jaspers, Scheler, Buber, and Sartre, have had some influence for example, Jean Paul Sartre's influence on Marwyn S. Samuels, Gaston Bachelard's influence on Buttimer. References to these thinkers appear in various seminal articles, but their in fluence is too diffuse to consider here.

Geographers Influenced by Phenomenology

Edward Relph's critical discussions concerning human ism led him to argue that humanism is an attitude, rather than a substantive parameter or subfield of geography. Influenced by Heidegger, Relph's phenomenology examines the lived experiences of people engaged in their environments. He sees the method of phenomeno logical description as an optimal way of apprehending the plurality of spatial worlds of lived experience. His work on place and placelessness in effect offers a spatial in terpretation of the care structure of Dasein, for he sees place as the spatial rooting of authentic existence. Place constitutes the meaningful center of one's circumspectful dealings with the world. Inauthenticity involves place lessness, a weakening of identification with place. A fundamental sociogeographical structure discussed by Relph is the insider–outsider continuum. Specific struc tures are empathic insidedness, existential insidedness, existential outsidedness, and objective outsidedness. Relph also develops a way into phenomenological geography: seeing, thinking, describing. He extrapolates John Ruskin's patient and careful manner of perceiving, Heidegger's reflective (vs. calculative) way of thinking, and the later Ludwig Wittgenstein's sense of showing. He defends himself against charges of egocentrism, idio syncrasy, and mere impressions, but does not use rigorous phenomenological methodology to preempt such criti cism. By 1981, in *Rational Landscapes and Humanistic Geography*, Relph eschews humanism because he sees it historically developing into modes of domination and exploitation based on scientistic arrogance, which follows from Heidegger's rejection of humanism. He advances a new attitude of geographical humility, which leads him to focus on notions such as genius loci. Remaining re spectful of genius loci requires responsible modes of attunement and careful sensitivity in place making. Relph is able to uncover the essence of modern building: the instant environment machine and its construction of commodity landscapes. Disregarding genius loci, the in stant environment machine constructs ersatz places, now based on the marketing mythologies of postmodern capitalism.

Yi Fu Tuan's works are quasi phenomenological, for he rigorously adheres to only one of its principles: the phenomenological reduction that brackets the use of the natural and cultural sciences – all theories of the natural attitude. Tuan considers the interrelation of human and world to be fundamentally both geographical and mean ingful and sees phenomenology as the study of 'man in the world'. In his best known work, *Topophilia*, Tuan avoids a technical or methodological presentation and this sets his *modus operandi* for his subsequent monographs. Presented as if a surface level phenomenological de scription, his insights, however, exhibit much deeper phenomenological analysis, which is recognizable to a trained phenomenologist. *Topophilia* is informed by phe nomenology due to its focus on the subject pole in the experience of the environment/setting. More specifically, he describes the structures concerning perception of, at titudes of, values of, feelings of, topos, concerning the human species *per se*, the cultural group, and the indi vidual. Tuan's overarching goal is to phenomenologically describe topophilia, the affective bond between people and place. In *Cosmos and Hearth*, Tuan presents hearth/home–cosmos/world as a fundamental eidetic structure of the lifeworld as it is spatially/geographically experienced. He develops a genetic phenomenology that describes the buildup of meaning in experience as it progressively ex pands from home toward world. Tuan unveils the ten sional relation between hearth and cosmos to be eidetic: neither is satisfactory of itself and both together are contradictory. This tension is illustrated in his two tele ological–historical reflections concerning China and the United States. Tuan questions the sedimentations in the genetic buildup of sense, much like Husserl's presentation of the Galilean style of science in the *Crisis of European Sciences*. Teleological genesis that forgets its roots in the hearth involves a crisis unless the hearth can somehow be brought back into experience. The cosmopolitan hearth gives us a way back to home/hearth from the uprooting tendency of the cosmopolitan point of view. This 'solu tion' is analogous to Husserl's strategy for overcoming Galilean science's abstractive (objectivist–naturalist) uprootedness from experience by re rooting science in the lifeworld.

John Pickles is the most rigorously Husserlian of those geographers influenced by phenomenology. Pick les elaborates his geography on the Husserlian model of a regional ontology. He seeks to investigate the eidetic structures of world, worldhood, space, human spatiality, place, etc. Unlike many of the humanistic geographers, and especially theoretically opposed to Buttimer, Pick les is not so much interested in the study of insiders' testimonies or the everyday lifeworld. Better than any other geographer, Pickles understands Husserl's phe nomenology to be a rigorous science that, in terms of its substantive exploration as eidetic science, would be complementary to the empirical sciences. Pickles em phasizes eidetic phenomenology, focusing on the *a priori* structure of geographical phenomena; it is not about experiences, even though it is ultimately founded on transcendental subjectivity. Pickles realizes that Hus serlian phenomenology offers the fundamental ground for intellectual responsibility, a rigor that precludes the naivete of objective science. Objects are not simply there as they are naively taken to be from the standpoint of the natural attitude, and the sciences built on that attitude; phenomenology examines how human consciousness becomes aware of objectivities and constitutes their meanings. The phenomenological field of constitution accounts for the appearance of objectivities in scientific cognition, the 'readymade already there' objects of ob jective science. The explanations of objective science are based on dominance and control over reality, without attaining to the fundamental and radical responsibility of accounting for scientific life, which involves the epistemological recognition that the subject–object re lation is primary. Ignoring this primacy is an aspect of the crisis of the European sciences. Unfortunately though, after making much of the nature of transcen dental subjectivity, Pickles then turns to a Heideggerian ontology of space and related themes, without sensing a tension between the two orientations. Heidegger's her meneutic of Dasein does not capture the Husserlian sense of a regional ontology nor does it take the ap proach of a Husserlian analysis of intentionality. It is unclear, given Pickle's emphasis on regional ontology, why he would not attempt an eidetic methodology fo cused on the phenomenon of space itself. It is unclear why in his later interest in maps he does not carry out a Husserlian genetic phenomenology concerning cogni tive mapping, given his earlier emphasis on transcen dental subjectivity.

David Ley's *A Social Geography of the City* draws on the social phenomenology of Alfred Schutz through making thematic the eidetic sociogeographical structures of the urban environment. Ley's sociogeography is recognized by the titles of the parts comprising the book: 'The geography of everyday life', 'The social basis of urban life', and 'The city and social structure'. He is interested in the spatial structure of everyday life – how social structures are spatialized in the urban setting. Ley is highly conscious of his experiential turn, critical of the limits of other approaches, for example, behaviorism/ objectivism, and aware of the pitfalls that need to be avoided in engaging 'philosophies of meaning', such as an appeal to subjective perspectives/impressions. Yet, his phenomenology is embedded in a social science approach of citing the findings of empirical studies, rather than carrying out eidetic phenomenological description. To illustrate, phenomena such as access, spatial mobility and immobility, restrictive environ ments, place opportunities for social contact, etc., are nowhere given formal eidetic description in the Schutzian style – as fundamental structures of sociogeo graphical phenomenology. Thus, the Schutzian com ponent in his work is not displayed in its distinctive approach and thus the phenomenology that he extols figures as dispensable, except as a contributor to the overall sociogeographic orientation. Perhaps this is why there have been no further major developments in the phenomenology of sociogeography, for in this work it appears to be overshadowed by empirical case studies.

Anne Buttimer, in *Values in Geography*, maintains a fundamental humanist position that her science or pro fession is not compartmentalized from her life and that the horizon or values toward which she is directed rests on commitment and care. She argues for her position by showing that to promote value free science is in fact an expression of a value system. Thus, the geographer is not a neutral observer, but rather the concepts and meth odologies employed for study already contextualize the geographer in value commitments. In sociogeographical study, the sense of any place is revealed through values that manifest in the choices and actions of its constitu ency. Values and meanings allow the geographer to engage in the study of lived geography. Buttimer's ap proach also owes much to Bachelard's notion of topoa nalysis, which is a kind of phenomenologically informed psychoanalysis that examines the interconnections of places and identities – cultures and individuals. She explores such subtopics as the interrelation of place and creativity. Buttimer eidetically distinguishes the insiders' experiences as fundamentally of process from the out siders' experiences as fundamentally of images. Her ap proach then is to mediate between them, envisioning a veritable science that transcends the insufficiencies of the subjective/insider and the objective/outsider points of views.

David Seamon continues the work of Buttimer, but in addition owes much to Merleau Ponty's notion of lived body intentionality. He later develops a richer position influenced by architectural theory by placing more emphasis on the co constitutivity of the lived body in relation to the built environment. He envisions his work as offering a reflexive perspective in behavioral geography, providing a humanistic–phenomenological alternative to behaviorism. In his earlier work Seamon develops such notions as the body subject, body ballet, and place ballet; the latter is illustrated in a description capturing the time–space routines of a marketplace in Sweden. Seamon explores the world of the everyday through three of its components: movement, rest, and encounter. In more recent interests, Seamon embraces the architectural work of Bill Hillier. Hillier examines the relationship between the structural patterns of design and human interactions. What is interesting here is that Hillier is usually viewed as an empirical scientist, whereas it can be argued that his apprehension of morphological structures in relation to human inter action really offers an eidetic phenomenology. Seamon, then, complements Hillier's work with his own work concerning the precognitive level of lived body–milieu, emphasizing the co constitutive aspect of the field. What is gained is how difference in patterns sets par ameters for the taking place of events. Thus, Seamon adds to what Merleau Ponty did not fully articulate with his lived body intentionality, the contribution of the milieu itself to the meaningful field of spatial enactment.

Trained in Heidegger's phenomenology, Robert Mugerauer provides exemplars of phenomenological description through his analyses of yards and porches. Through his sensitivity to the hermeneutic component of description, which he envisions as an expansion of the phenomenological project, Mugerauer becomes aware that meaning genesis in the context of historical time is a fundamental moment of interpretation. He reasons that if landscapes/environments are addressed as texts, then interpretations of those texts are a function of changes in assumptions and considerations over time, over historical eras. Mugerauer masters interpretive methodologies across the spectrum of traditional/historical, phenomenological/hermeneutic, and decon structive/postmodern in his interpretations of environ ments, place, and architectural elements. His descriptive studies are didactic by articulating the essential dis tinguishing features of each methodological perspective and by showing the benefits of multiple interpretive perspectives, especially for practical application. He critically focuses on contemporary environmental problems such as the loss of place and the loss of the sense of place, as well as the problematic role of tech nology. His goal has been to inform geography and the design professions with deep theoretical insights while opening paths of practical applications toward solving environmental problems.

Concluding Remarks

Phenomenology has entered into the various fields of human science exhibiting varying trajectories and mo ments of success. Within geography it was a humanistic turn in reaction to objective science through which phenomenology came to be introduced to the field. The experiential turn of phenomenology informed geography immediately set up a reactionary moment and an ad versarial atmosphere in relation to 'acceptable' geo graphical paradigms. But it is really only the vestige of a metaphysical commitment to naturalism that disallows the possibility for overcoming these scientific conten tions. It has long been this author's view that eidetic science and empirical science are complementary, which he dramatizes by the following (perhaps hyperbolic, but heuristic) statement: "Eidetic science without empirical science is empty, empirical science without eidetic sci ence is blind." He puts forth that such a position would promote the highest possible excellence in geographical science and would overcome the polarizations. Also of consideration is that, the momentum that had been built up in phenomenology oriented geography had been weakened by new and exciting postmodern forms of exploration. Yet, this 'distraction' has not destroyed the movement, for there is some overlap and interrelation between approaches. The fact that there are not many phenomenology minded geographers occupying chairs in geography departments in recent times has not des troyed the movement, either. But what is needed perhaps for its revitalization is one or two newly published de finitive works explicating its methodologies in rigorous, but readable, terms for geographers and one or two volumes of excellent case studies exhibiting the eidetic/empirical complementary approach in order to assuage its earlier confrontational history.

See also: Behavioral Geography; Existentialism/Existential Geography; Humanism/Humanistic Geography; Positivism/Positivist Geography; Post-Phenomenology/Post-Phenomenological Geographies.

Further Reading

Buttimer, A. (1974). *Values in Geography*. Washington, DC: Association of American Geographers Commission on College Geography, Resource Paper 24.
Buttimer, A. and Seamon, D. (eds.) (1980). *The Human Experience of Space and Place*. London: Croom Helm.
Gregory, D. (1978). *Ideology, Science and Human Geography*. London: Hutchinson.
Ley, D. (1983). *A Social Geography of the City*. New York: HarperCollins.
Ley, D. and Samuels, M. (eds.) (1978). *Humanistic Geography: Prospects and Problems*. Chicago, IL: Maaroufa Press.
Mugerauer, R. (1996). *Environmental Interpretations: Tradition, Deconstruction, Hermeneutics*. Austin, TX: University of Texas Press.

Peet, R. (1998). *Modern Geographical Thought*. New York: Blackwell Publishing.
Pickles, J. (1985). *Phenomenology, Science and Geography: Spatiality and the Human Sciences*. London: Cambridge University Press.
Relph, E. (1976). *Place and Placelessness*. London: Pion.
Relph, E. (1981). *Rationalistic Landscapes and Humanistic Geography*. Beckenham: Croom Helm.
Seamon, D. (1979). *A Geography of the Life World*. New York: St. Martin's Press.
Seamon, D. and Mugerauer, R. (1985). *Dwelling, Place, and Environment: A Phenomenology of Person and Place*. Dordrecht: Martinus Nijhoff.
Tuan, Y. F. (1974). *Topophilia: A Study of Environmental Perceptions, Attitudes, and Values*. Englewood Cliffs, NJ: Prentice Hall.
Tuan, Y. F. (1999). *Cosmos and Hearth*. Minneapolis, MN: University of Minnesota Press.

Relevant Websites

http://www.phenomenologycenter.org
CARP: Center for Advanced Research in Phenomenology.
http://www.arch.ksu.edu
EAP: Environmental and Architectural Phenomenology Newsletter, The College of Architecture, Planning and Design.
http://www.towson.edu
IASESP: International Association for the Study of Environment, Space, and Place.
http://www.o p o.net
OPO: Organization of Phenomenological Organizations.
http://www.spep.org
SPEP: Society for Phenomenology and Existential Philosophy.
http://www.phenomenology.org
WPI: The World Phenomenology Institute.

Philosophy and Human Geography

S. Elden, Durham University, Durham, UK

Introduction

Geography has always had a relation to philosophy. Many of the great thinkers of the Western philosophical tradition have written explicitly on many of the key issues of concern to geographers. These have included discussions within metaphysics and physics concerning nature, space, time, place, and the human; logic and epistemology on the nature of knowledge and scientific method; and ethical questions about behavior and its spatial variation, political organization, cultural difference, and our role in the world.

Philosophy, traditionally, included a far broader range of aspects than the university discipline which goes under this title today. Philosophy is, literally, the love (*philos*) of wisdom (*sophia*), and can be argued to provide the foundation for core arguments across the human and physical sciences. The field of physics, for instance, used to be known as natural philosophy, as the term physics derives from the Greek *physis*, nature. One of the key insights of philosophy is that our assumed ideas are dependent on prior argument, and tracing the history of fields of human knowledge can provide insight into how we came to think as we did. Philosophy can be a forbidding area, since it often depends both on conceptual rigor and knowledge of earlier thinkers and their ideas, but its potential impact is substantial. This extends across the range of concerns of geography, for many great philosophers, such as Gottfried Leibniz, René Descartes, or Aristotle were important scientists in their own right, and many of the figures of social/spatial theory who have been so influential in contemporary human geography were trained in philosophy.

Philosophy tends to abstract from particulars, whether through initial observation and generalizing from patterns to produce explanatory theory, or through beginning with reasoned argument which is then used to make sense of the world. Its impact has been felt throughout the history of the discipline. In modern times, both positivism or the spatial science of geography, or the reaction against this in humanism or postpositivist theory have all been uses of philosophical arguments.

The Philosophical Tradition

Early Greek Thought

Within the Western philosophical tradition – the key thinkers and texts since its inception – there are a number of philosophers that have explicitly thought about geographical concerns. From the very beginning of preserved Western thought, in the fragments of the texts of the pre Socratics (so called because they preceded Socrates), philosophers have been concerned with the nature of the world. Heraclitus, for example, likened change in the world to a child playing a game, suggesting the elements of chance within rules that shape its destiny.

It was, however, a later group of Greek thinkers that defined much of the way in which philosophy came to think, and the terms within which it did so. Socrates, whose ideas were presented in dialogs by his student Plato, provided a model of philosophical inquiry as well as many of the key questions themselves. Plato's writings, which cover the full range of human inquiry, often do not have specific topics, and issues of interest to geographers appear throughout. His dialog *Timaeus* attempts to provide an explanation for the origin of the universe, with the idea of the creator and the laws governing it. It is also important for its discussion of the *khora*, a term sometimes misleadingly translated as 'space', but which is closer to a notion of place, a location, or a receptacle for something.

Aristotle

Plato's most brilliant student was Aristotle, whose extant works are largely notes for lecture courses, and therefore, tend to be more systematic and less literary than those of his predecessors.

Aristotle wrote on almost all fields of human knowledge, from zoology and biology to reflections on theater and rhetoric. He made a number of analyses of questions in language and logic, and studied political and social organizations in a comparative manner. His *Nicomachean Ethics* contains reflections and arguments not just on ethics, justice, and virtue, but also discussions of the nature of knowledge and the division between practical and philosophical wisdom. His *Rhetoric* offers more than simply an analysis of speech, but discussions of moods and the collective community. In his *Physics* he analyzes a range of concerns around place, nature, and movement, and the *Metaphysics* – so named because it came after (*meta*) the *Physics* – is a collection of theoretical treatises on being, matter, and mathematics.

Aristotle's work demonstrates an early version of what might be called the scientific method, with a combination of abstract reflection with practical observation. He deduced that the Earth had a spherical shape because he observed lunar eclipses and reasoned this from the shadow cast on the moon's surface. His work on biology and comparative anatomy set the tone for much future analysis. His work on motion in the *Physics* is tied to his understanding of place. For Aristotle, place is an inherent quality of beings, it is their capacity to be present. Yet, the physical extension of an object is insufficient to under stand place. Objects have, for Aristotle, their proper place, and therefore, motion is tied to this understanding of place. Heavy things tend downward, and air and fire tend upward. Celestial bodies move in a circular motion. These are natural motions. Aristotle believed that the speed of falling was proportional to both the weight of the object and the density of the medium it is falling through. In Aristotle's theory, place does not move, and only what is movable is in a place. Motion is not spon taneous, and therefore, there must always be a mover, giving rise to the idea of God as the prime mover – the being that initially sets things motion.

Medieval Philosophy

The Medieval period is not straightforwardly character ized, but tends to be understood through its relation to theology. Certainly, many of the key thinkers were theologians as well as philosophers – Augustine and Thomas Aquinas among them. In these writers, there can be a tendency to retreat from the understanding of this world in favor of thinking about other matters. Aristotle's work was hugely influential in this period, and he was often referred to simply as 'the philosopher'. Many of the key developments in this period were the product of influences from Arab thinkers, such as al Idrisi, whose reflections on geography and cartography were important for several centuries.

Descartes and Modern Philosophy

Modern philosophy emerged out of the Renaissance and the coming back into general circulation of a number of texts of the ancients. Somewhat crudely, philosophy of the seventeenth and eighteenth centuries can be seen as an extended debate between the Rationalists and the Empiricists. For the Rationalists, which would include Descartes, Leibniz, and Baruch Spinoza, the world is understandable through the abstract processes of reason. Descartes is the paradigmatic example.

In Descartes' work, the material and mental worlds are strictly divided. Descartes suggested that in order to get knowledge on a firm footing, he would doubt everything of which he could not be certain. He argued that he could not doubt that he was doubting, and therefore, could be sure that he was thinking. Because he was thinking he could be sure that he existed, therefore the famous phrase *cogito ergo sum* – I think, therefore I am. This meant, for Descartes, that there were certain things that could be characterized as *res cogitans*, thinking things. Everything else had a physical manifestation, material or matter. These he characterized as *res extensa*, extended things, because he reasoned that extension in three dimensions was their central, fundamental characteristic.

Other characteristics of material things – such as their weight, color, and shape – were secondary. Descartes' example for this is to take a piece of wax and make a number of observations of it. Heating the wax would cause almost everything to change – its color, shape, and smell, for example – but not that it occupied space, that it was extended. Descartes, therefore, defined a material body as bounded, enclosed in a place, and filling up space. The last part of this is particularly important – space is a container, itself extended in three dimensions, of which objects exclusively filled up a part of it, excluding other objects from that same space. Things can be in the same place – different pairs of shoes in a cupboard, for instance – but not in the same space. Space is bounded and ex clusive. Descartes, therefore, understands motion as movement through space, position as the location vari ously shaped things possess in relation to each other, and shape and its change as dependent on the limit of the extension.

This gives rise to a particular way of seeing and conceiving the material world as a whole. Extension in three dimensions – length, breadth, and depth – is the central characteristic of nature, and geometry is the sci ence that allows us best access to it. In other words, there is a mathematical determination of the material world. Descartes' work on geometry is extremely important in its own right, since he develops an analytic or coordinate geometry which allows geometrical problems to be re duced to numbers – equations, or the length (i.e., the quantity) of lines. Many contemporary geographers have criticized this abstraction and mathematicization, in cluding feminists who see it as masculinist.

For the Empiricists, who include John Locke, Bishop Berkeley, and David Hume, observation of the physical world and generalization of rules from this observation is the correct way to proceed. Prediction may be difficult, but it is not impossible and is based on generalization of experience. Hume famously suggested that all works should be evaluated on two simple criteria – whether they contained "abstract reasoning concerning quantity or number" or "experimental reasoning concerning matter of fact and existence." If the answer was no to both, then he suggested that we "commit it to the flames for it can contain nothing but sophistry and illusion." Empiricism has had a significant impact on what might be seen as a commonsense view of geography, whereby gathering data leads to descriptions of the world and its features and inhabitants.

Kant

Immanuel Kant is generally seen as the most important philosopher since Aristotle. One of his many merits is that he provided a way of reconciling rationalist and empiricist positions within a unified model of thought. He also thought that there were some things that were matters outside the realm of rational thought, such as the existence of God. He convincingly demolished the arguments for or proofs of God's existence, suggesting that this was rather a matter of faith.

In his most important work, the *Critique of Pure Reason* in 1781, Kant tries to bring together experience and reason. He suggests that there are a number of kinds of statements we can make about the world. Statements can be either analytic or synthetic – meaning that they explicate the meanings of terms or add something additional. Analytic statements would be of the form 'all fathers are male' or 'green things are colored' – statements whose negation would be absurd, and which do not need to validated. Synthetic statements would be of the nature of 'Kant was a philosopher' or 'copper conducts electricity' – things that are not inherently self evident. Statements can also be *a priori* or *a posteriori*. There are some things we know prior to experience and some we know after it. These divisions of statements are not unrelated – it is clear that all analytic statements are *a priori* and that all *a posteriori* statements are synthetic.

This leaves the question of whether synthetic *a priori* statements are possible. Hume had denied that they were, whereas Kant thinks they are. For Kant, they are combinations of concepts and particulars, of reason and experience. While the Empiricists suggest all our knowledge must conform to experience, Kant says all experience must conform to knowledge. What this means is that we can only experience within our conceptual frame. Kant, therefore, suggests that experience is a 'necessary' condition for knowledge (thereby criticizing the Rationalists), but it is not a 'sufficient' condition for knowledge (thereby criticizing the Empiricists). Kant's famous phrase is "thoughts without content are empty, intuitions without concepts, blind."

The conceptual frame from within which we experience the world means that, for Kant, our minds are constructed in a particular way. He suggests, crucially for geographers, that space and time are *a priori* concepts – that is they precede experience – which serve to structure the world we perceive. He calls these the transcendental esthetic, from the Greek *aisthesis*, perception. We cannot think without using these concepts – they are like seeing through a pair of irremovable, tinted glasses. Space and time for Kant are thus part of our perceptual apparatus, the way in which we perceive the world, not that which we perceive, which means that we experience within space and time rather than experience space and time themselves. Space and time are, therefore, determined in advance in Kant's thought, the material world is understood before it is experienced, and therefore, what we experience is conditioned by our prior knowledge. The limits of the system, thus, become the limits of the experience; the composition of the thought of nature becomes the limits of nature itself. Nature ceases to be what it might be except in terms of the system through which it is understood.

Kant wrote a number of other important works, including two further 'critiques' – *The Critique of Practical Reason*, in 1788, and *The Critique of the Power of Judgment*, in 1790, dealing with ethics, esthetics, and teleology. He also lectured on anthropology and physical geography for a number of years. While he edited the former into a book, *Anthropology from a Pragmatic Point of View*, in 1798, the *Physical Geography* exists only as a compilation from lecture transcripts. Kant thought that, together, anthropology and geography provided knowledge of the world, which served as empirical grounds for his thought. Both of these were taught for their pragmatic dimension – what they could provide as guides for our moral and practical life. This leads Kant to what he calls 'moral geography,' which looks at the customs and characters of different peoples and some outdated and discredited views of race. Kant provides extensive discussions of contemporary thinking on the Earth and its terrain, earthquakes, climate, rivers and water, flora, fauna, and minerals. The final part of the book is a series of descriptions of particular regions and places in the world. While much of the detail may be outdated and therefore of merely historical interest, Kant's way of structuring geographical knowledge and its relation to his thought as a whole is of enduring importance. This importance lies both in the way he understands geography as a counterbalance to history, and in terms of the organization of knowledge. For Kant, all perceived things are located in logical classifications, such as those of Linnaeus; and in space and time. Logic deals with the first; physics with space and time, and of these, geography deals with space; history with time. Geography, therefore, allows us access to the ordering and categorizing of the world. Indeed, Kant distinguishes geography, as the description of the whole world, from topography, as the description of single places, and chorography, as that of regions.

Kant's actual impact in geography has been erratic, with the schema for the relation between time and space and geography and history more important than the empirical and analytical details. Yet his wider philosophical thinking of the categories of space and time have been profoundly important in shaping the way subsequent philosophies of this topic have engaged with these topics. Georg Wilhelm Hegel's philosophy, and in particular, his political philosophy of the state as an organic whole were influential for German geopolitics of the late nineteenth and early twentieth century, in particular in the work of Friedrich Ratzel.

After Kant, philosophy becomes much less of a unified discipline. Indeed, the two dominant paradigms of Western philosophy can often barely agree on who counts as a philosopher after this date. The continental European tradition would include Hegel and the German romantics, Karl Marx, Friedrich Nietzsche, the phenomenology movement best exemplified by Edmund Husserl, Martin Heidegger, and then a range of post World War II thinkers in France, such as Jean Paul Sartre, Michel Foucault, and Jacques Derrida. Analytic, or Anglo American, philosophy on the other hand tends to take its influence from other sources, such as Gottlob Frege and Ludwig Wittgenstein, the American pragmatists, and the ordinary language philosophers who analyzed how language works instead of abstract reason. Both of these traditions, and the debates within them, have much to say about aspects of geographical concern.

Positivism

Positivism within geography can be described as the use of modern scientific method. Although undoubtedly more influential within physical geography, this has had a significant impact within human geography as well.

Like Hume, positivism suggests that there two kinds of meaningful statement. Those that are logically true, such as those concerning the nature of mathematics or language; and those that are verifiable through empirical observation. Anything else is metaphysical and, ultimately, meaningless. This leads to an emphasis on observation, which is based only on what is directly seen rather than inferred, and the primacy of this over theoretical propositions. Models and theories should be developed from their basis in this empirical observation. Observation could be guided by a prior hypothesis, so that rational thought structured the process of inquiry, but this hypothesis was open to testing, and therefore capable of being disproved. In this, the positivists were influenced by early twentieth century philosophers, particularly those of the Vienna Circle. One of their number, Karl Popper, proposed that the criteria for validity in science should not be that there are examples which prove it, but that there could be those which would disprove it. In other words, the strength of a theory is not that it is sufficiently general that it appears to explain, but that it is sufficiently narrow and focused that if a prediction proved untrue the theory would be exposed. This is the criterion of falsifiability. For Popper, and others, this meant that the explanatory claims of much social science were not scientific at all.

In social science and human geography more specifically, this led to the development of a number of quantitative and statistical techniques. The aim was to ground geography as a spatial science, with the removal of value judgments and the utilization of scientific methods. This is sometimes known as the quantitative revolution in geography. David Harvey's 1969 book *Explanation in Geography* was a key text in this field, but it also extends to some forms of cartography and locational analysis. It also provides much of the conceptual foundation to the claims of Geographic Information Systems (GISs), which in practice is strongly associated with Empiricism.

Humanism and Marxism

There have been various reactions to positivism within geography. Humanist geography emphasizes the individual experience of place, and suggests that general theories neglect the unique perceptions of people and groups. Some elements within this strand of geography adopted ideas from phenomenology and existentialism, stressing the sense of place as opposed to the abstract geometries of space. Yi Fu Tuan's *Topophilia*, in 1974, and *Space and Place: The Perspective of Experience*, in 1977, were important texts. For others, the insights of hermeneutics allowed human geography to be situated within the humanities rather than the social sciences.

Marxist thought was relatively late in terms of its influence within geography, compared to other social sciences and humanities disciplines. It can be seen as part of a wider trajectory of radical geography. Marxist geography has both introduced the arguments of Marxism, particularly political economy, into geographical analysis, but also provided the introduction of a spatial element in Marxism. In the first, there is an emphasis on questions of value, rent, and appropriation. The second is potentially more challenging, since it attempts to redress the imbalance of historical materialism, with an emphasis on temporal chance, with a historical and geographical materialism, which takes space as a crucial determining context to all social interaction and struggle.

David Harvey's *Social Justice and the City*, in 1973, was a crucial mediating text in introducing Marxist ideas into geography, but it was arguably his *The Limits to Capital*, in 1982, which had a more profound impact. This study looks to fill the 'black boxes' of Marxist thought from a geographical perspective. The French Marxist philosopher Henri Lefebvre's *The Production of Space*, in 1974, first translated in 1991, also had an important impact. In this work, Lefebvre provides a summary of philosophical theories of space, and proposes an understanding of space through a combination of physical, material attributes; mentally conceived space; and the lived spaces of social interaction. He proposes a historical investigation of these different spaces, simultaneously challenging traditional Marxist accounts of time and space.

A range of other Marxist thinkers have proved in fluential within the discipline. Althusser's work on ideology and power in the modern state and Gramsci's work on Fordist production, hegemony, and the structure of political action are two other particularly important examples, but traces of a Marxist approach and politics are widespread.

Radical geography also makes use of other ap proaches, such as feminist philosophy and anarchist thought. Its key contribution has perhaps been to raise the political question within geography again. Feminist thought in geography draws on a range of thinkers, many of whom – such as Julia Kristeva and Luce Irigaray – engage with a feminist re reading of the Western philo sophical tradition.

Postpositivism in Contemporary Geography

Besides the alternatives offered by humanism and radical geography, positivism has also been challenged by more scientifically minded approaches. The arguments of philosophers of science, such as Thomas Kuhn and Paul Feyerabend, have proved influential in rethinking the development of scientific knowledge, which has helped some to think about geographical advances. Kuhn, in *The Structure of Scientific Revolutions*, in 1962, proposed understanding scientific development through the notion of paradigms, which provide a framework of under standing within which science takes place. Most science is therefore 'normal science', operating within previously agreed rules and laws. Only some science truly chal lenges, opening up the possibility of a paradigm shift. Feyerabend's *Against Method*, in 1975, argued that there was no single model of approaching science and that to impose one would hinder scientific progress. This theoretical pluralism is, he believed, less restricting on science.

In human geography, a range of contemporary phil osophers have been influential with geography. Many of these have been appropriated under the rubric of social/spatial or sociospatial theory, but their insights can also, and perhaps better, be understood as philosophical. Michel Foucault, Jacques Derrida, and Gilles Deleuze have all had an impact in a broadly conceived, post structuralist geography. This has been concerned with challenging the disciplinary range of what geography is, the binaries that traditional geography has tended to think with, and to emphasis the emotional, cultural, and nonsystematic elements of the interaction between humans and the world.

To turn the relation the other way round – to ask of the geography of philosophy – is to raise a wide range of issues. While some approaches of this kind would smack of a geographical determinism, locating particular kinds of thoughts within a particular place or tradition, there can be more productive work done. Philosophers regu larly speak of contemporary French thought, German romanticism or idealism, English empiricism, and the Scottish enlightenment, for instance, and there is a long running debate about whether there is such a thing as a distinctively American philosophy. More productive work has been done by geographers of thought and sci ence, such as Trevor Barnes, Anne Godlewska, David Livingstone, and Charles Withers. In their last book together, *What Is Philosophy?*, Deleuze and Félix Guattari proposed a concept of 'geophilosophy', which they claimed was founded by Nietzsche. But perhaps the most interesting geographical issue is the way that most dis cussions of philosophy – including this account – is of Western philosophy, that which can be derived from a lineage back to ancient Greece. This, of course, neglects important thought that has its roots in India, China, or other non Western locations. This is rarely taught in philosophy programs, and has had much less impact on philosophical thinking in geography.

In summary, it is clear that many of the issues of concern to geographers have also been analyzed by philosophers. Geography has always been based on philosophical positions, but until relatively recently these were largely unexamined.

See also: Critical Geography; Enlightenment Geography; Human Geography; Mapping, Philosophy; Phenomenology/Phenomenological Geography; Positivism/Positivist Geography; Quantitative Revolution; Radical Geography.

Further Reading

Casey, E. S. (1997). *The Fate of Place: A Philosophical History*. Berkeley, CA: University of California Press.

Cloke, P. and Johnston, R. (eds.) (2005). *Spaces of Geographical Thought*. London: Sage.

Deleuze, G. and Guattari, F. (1994). *What Is Philosophy?* London: Verso.

Elden, S. (2009). Reassessing Kant's Geography. *Journal of Historical Geography*, Vol. 35.

Glacken, C. J. (1967). *Traces on the Rhodian Shore: Nature and Culture in Western Thought from Ancient Times to the End of the Eighteenth Century*. Berkeley, CA: University of California Press.

Gregory, D. (1978). *Ideology, Science and Human Geography*. London: Hutchinson.

Gregory, D. (1994). *Geographical Imaginations*. Oxford: Blackwell.

Hartshorne, R. (1958). The concept of geography as a science of space, from Kant and Humboldt to Hettner. *Annals of the Association of American Geographers* 48(2), 97 108.

Harvey, D. (1969). *Explanation in Geography*. Oxford: Blackwell.

Hubbard, P., Kitchin, R., Bartley, B. and Fuller, D. (2002). *Thinking Geographically: Space, Theory, and Contemporary Human Geography*. London: Continuum.

Livingstone, D. (1992). *The Geographical Tradition: Episodes in the History of a Contested Enterprise*. Oxford: Blackwell.

Martin, G. J. (2005). *All Possible Worlds: A History of Geographical Ideas*. New York: Oxford University Press.
May, J. A. (1970). *Kant's Concept of Geography and Its Relation to Recent Geographical Thought*. Toronto, ON: University of Toronto Press.
Peet, R. (1998). *Modern Geographical Thought*. Oxford: Blackwell.
Romm, J. S. (1992). *The Edges of the Earth in Ancient Thought: Geography, Exploration, and Fiction*. Princeton, NJ: Princeton University Press.
Soja, E. W. (1989). *Postmodern Geographies*. London: Verso.
Tuan, Y. F. (1977). *Space and Place: The Perspective of Experience*. Minneapolis, MN: University of Minnesota Press.

Photogrammetry/Aerial Photography

P. Collier, University of Portsmouth, Portsmouth, UK

Glossary

Color Infrared Photography A color film in which one layer of the emulsion has been made sensitive to electro-magnetic radiation in the near infrared, just outside of the visible spectrum (0.7–0.9 μm). Vegetation has its maximum reflectance in the near infrared, making this type of photography particularly suitable for monitoring vegetation. As it is a magenta dyed layer that is sensitive to the near infrared, vegetation has a characteristic red appearance on this type of photography.

High-Resolution Photographs Photographs that show fine detail on the ground, usually produced by using long focal length lenses and fine-grained emulsions.

Land Systems Analysis An aerial photography-based system of carrying out resource inventories. The approach was developed by Christian and Stewart at the Commonwealth Scientific and Industrial Research Organisation to meet the needs for rapid and cost-effective mapping of resources.

LiDAR Light Direction and Ranging, an active system, which uses a laser beam to measure the direction and distance of ground points relative to the system, which is normally mounted in an airplane. Used in conjunction with GPS and an inertial system, it can be used to create very accurate terrain models.

Orthophotograph An aerial photograph, which has all the errors in position and scale due to height distortions and tilt removed. An orthophotograph has the same geometric properties as an orthogonal map, but with the higher information content of the original photographs.

Photomosaic A composite image made up from several aerial photographs, which have been correctly positioned to give the appearance of being a single photograph.

Aerial Photography

Aerial photography can mean both the process of taking photographs from some kind of aerial platform, and the resultant imagery. The photography can be taken with the camera pointed vertically down toward the ground (the normal case), or tilted off vertical, so called oblique photography. Aerial photography is normally taken using a specially designed large format film camera producing a negative measuring 230 mm × 230 mm (9 inches × 9 inches).

The Early Days of Aerial Photography

In 1858, Felix Tournachon (Nadar) took the first aerial photograph over Paris from a captive balloon. However, Tournachon was not the first to suggest the possibility of aerial photography, or to see its benefits as a way of collecting information for map making. The earliest known suggestion was as a joke in a French lithograph from 1840. Despite being able to take photographs from a captive balloon, practical limitations of the photographic process meant that Tournachon had to take the darkroom up in the balloon as well as the camera. This was because the photography of his time used the so called wet plate process, where the glass negative was coated with the emulsion, exposed, and processed while the emulsion was still wet. This also meant that only tethered balloons could be used, limiting the range of targets that could be photographed.

To make aerial photography a viable technology, it was necessary to have a stable dry plate emulsion, portable cameras, and an aerial platform capable of being positioned at will over the required target. Technical advances in emulsions led to suitable photographic media being available by the 1880s. Advances in lenses meant that cameras suitable for aerial photography were available from about 1890. All that was now needed was a suitable platform. Experiments were carried out using free balloons, kites, and even pigeons, but it was the development of the aeroplane that made aerial photography a practical proposition.

Within a few years of the Wright Brothers' first flight, a number of aerial photographs had been taken. These usually involved nothing more complex than pointing a camera over the side of the aeroplane and taking a picture. While interesting as images, they were of little practical value. The first aerial photographs to be used for a practical purpose were those taken by Captain Piazza of the Italian Army during the Italo Turkish War in 1912. These photographs were taken for military purposes, and the development of aerial photography was to be largely dominated by military applications until after World War I. The one notable exception to this was the United States, where the US Geological Survey (USGS) carried out experiments in civilian mapping from aerial photography prior to the United States joining the war in 1917.

World War I

Prior to the outbreak of World War I aerial photography had been taken experimentally by most of the combatant nations. In Britain, for example, aerial photography had been taken during fleet maneuvers in the summer of 1913. However, when the war started, aeroplanes were initially used as if they were a modern form of cavalry, scouting ahead of the army and the aircrew providing written descriptions of what had been observed. As far as is known, no combatant army was equipped for aerial photography. However, some airmen took their own cameras with them to record what had been seen, and the value of the photography was quickly recognized by the intelligence services. Once the front line had stabilized and trench warfare had become the norm, aerial pho tography became the only way in which intelligence could be gathered on the movement, deployment, and defenses of the opposing armies. In addition to providing intelligence, aerial photography quickly became the most important source of information for mapping.

There were major improvements in cameras during the war. Anti aircraft fire forced aeroplanes to fly higher to avoid being hit, and this necessitated the development of longer focal length lenses to ensure high resolution photographs for interpretation. Although there were developments in film technology, most of the cameras used during World War I used glass plates, with the camera operator needing to change magazines in flight.

On the Western front, aerial photography was mainly used to revise and update existing maps, to plot the positions of trench lines, artillery positions, barbed wire, supplies dumps, railroads, and any other potential mili tary targets. In the Middle East, it was used to produce entirely new mapping of previously unmapped areas.

The experiences gained during World War I were to have a major impact on mapping policies after the war. Perhaps even more important, however, was the con sequence of exposing so many soldiers to this new data source. Many soldiers were to take the knowledge gained with them on their return to civilian life. After the war new applications of aerial photography were quickly explored. Among the disciplines to benefit were archae ology, geology, physical geography, biology, and forestry.

The Inter-War Period

During the inter war period, there were a number of technical advances in aerial photography with the de velopment of improved cameras, which used film rather than glass plates and had improved lenses and shutters. Infrared and color film were introduced in the 1930s, but had hardly begun to be exploited before the outbreak of World War II. In practice, aerial photography was to remain dominated by black and white photography until the 1970s.

Archaeological uses of aerial photography received a great boost when Britain's Ordnance Survey appointed O.G.S. Crawford to its staff. Crawford had become con vinced of the value of aerial photography and demon strated it through a series of publications for the Ordnance Survey and in *Wessex from the Air*. Aerial archaeology has continued to be an important area of work, but benefiting from an increase in interest in the late twentieth century through its use in conjunction with Geographic Information Systems (GIS) for historic landscape reconstruction. Despite civil applications of aerial photography becoming important during the inter war period, in Europe, at least, military requirement continued to be an important driving force behind technical developments. This became increasingly im portant with rearmament in the mid 1930s. It was really only in the United States that aerial photography was able to develop within an almost entire civilian context. This civilian context also encouraged the growth of commercial aerial photography companies. In Europe it was much harder for commercial companies to become established and to grow. One of the driving forces behind the civil developments in the United States was the relatively poor coverage of the country by the mapping of the national mapping agency. This encouraged the local administrations not only to adopt newer mapping tech niques to create maps of the rapidly growing urban areas, but also to use aerial photography as a planning tool. The Federal Government also played its part in encouraging the development of aerial photography. Both the USGS and the Coasts and Geodetic Survey were keen advocates of the use of aerial photography.

However, in common with many government agen cies, they could not develop as quickly as they would have wished due to government financial stringency.

Probably, the biggest single drivers in the development of aerial photography during the inter war period were the New Deal initiatives, such as the Agricultural Adjustment Administration and the Tennessee Valley Authority. Both initiatives relied very heavily on aerial photography, leading to technical advances in the field, developing a large cadre of trained personnel, and giving much needed work to private aerial photography companies.

Lee and Joerg had discussed the possible use of aerial photography for studying urban areas, but the first serious attempts to use aerial photography for urban analysis did not take place until 10 years later with the work of Finch and Hudson. In practice, however, it was not until after World War II that anyone attempted systematic civilian studies of urban areas using aerial photography.

World War II

During World War I, the primary use of aerial photog raphy was the updating of maps and the identification of targets for shelling. In World War II these remained important applications, but the war spawned a wide range of other geographic applications that were to have important uses in the post war era. The huge distances over which the war was fought, and the heavy reliance on bombing of strategic targets, meant that the aerial photography was heavily used both as a reconnaissance tool and for target identification. In the early years of the war, the Germans made the most successful use of aerial photography, but by 1941 this lead was eroded and by the end of the war the Western allies had the largest and most efficient system for photographic analysis.

A large part of the effort throughout the war was dedicated to target analysis, to identify potential targets for bombing, and then for postbombing analysis. Specialist teams were trained to identify particular types of targets, such as aircraft production facilities, or oil refineries, and to estimate the output potential of the identified facilities. Detailed studies were carried out of urban areas, which involved identifying functional areas within the town, and also the susceptibility to incendiary bombing of the buildings within the areas.

Aerial photography was also used for the detailed mapping of defenses prior to amphibious assaults, whether in northwest Europe or islands in the Pacific. In addition to studying the beach defenses, photography was also used to assess beach gradients and beach sediments. In the Pacific campaign, color aerial photography was used to determine the depth of water over reefs prior to amphibious assaults.

In 1942, Kodak introduced the first color infrared photography. Originally intended for camouflage detection, it was only rarely used in the immediate post war era, as it was expensive, relatively unstable, and difficult to handle. However, from the 1960s, it was increasingly used for vegetation mapping and monitoring, becoming the standard tool for that purpose until challenged by airborne scanner systems in the late 1990s.

As with World War I, the most important legacy of World War II was not the technical advances that occurred during the war, but rather it was the huge numbers of personnel trained and experienced in the use of aerial photography by the war's end.

Post-War Reconstruction and Development

The use of aerial photography as a geographical tool was still largely experimental before World War II, but in the post war world its use was to become standard. In part, that was due to the large number of service personnel, who had become familiar with aerial photography during the war, returning to their peacetime occupations aware of what aerial photography could offer. In addition, the victorious powers had large numbers of surplus aircraft and cameras that could be put to use in collecting aerial photography.

Among the first to starting making extensive use of aerial photography as a tool were the planners charged with the reconstruction of the devastated cities of Europe and Japan. The absence of up to date maps meant that photography was the only way in which a picture could be obtained of the situation on the ground prior to reconstruction. Even in countries like the United States, huge changes had taken place in urban areas as a result of the expansion of war related industries. Not just factories, but new roads and housing had been constructed, but not shown on existing mapping. In part, the inadequacies of the existing mapping were also due to the diversion of mapping staff to war related activities, leading to the neglect of the normal work of official mapping agencies. Planners quickly learnt that the photomosaics that had proved so useful as planning tools for military operations could equally well be used to show what already existed and what needed to be done. Within 3 years of the end of the war, the first seminal work had appeared, *Aerial Photography in Urban Planning and Research*, by Melville C. Branch, who had trained as a photointerpreter in the US Navy.

In Australia, World War II had demonstrated the vulnerability of the poorly populated northern part of the country. The war was followed by a drive to develop those vulnerable parts to encourage their settlement. As there was an almost complete absence of mapping, the Commonwealth Scientific and Industrial Research Organisation developed the method of land systems analysis to provide the necessary information on natural resources. The approach was entirely dependent on the use of aerial photography to identify terrain units, and was subsequently adopted in a number of countries both for general resource surveys and for specific development measures, such as road construction.

Prior to World War II, the development of most of Africa, even in the 'settler colonies', had been neglected due to a combination of lack of will and resources. Few colonies were mapped at more than a superficial level and there had been little in the way of systematic surveys of natural resources. Following the war there was a concerted attempt to develop the colonies, in some cases so that they could provide much needed raw materials and food to the metropolitan powers, in others to prepare them for eventual independence. The 'groundnut scheme' in Tanganyika was the most notorious scheme to provide foodstuffs for the metropolitan power, but there were similar smaller scale projects in a number of colonies. Using military aircraft and cameras, huge tracts of Africa were systematically photographed in the post war period. This photography was used to create the first proper topographic mapping, or in the form of photomosaics as a basis for resource inventories.

As African colonies were prepared for independence, aerial photography was used to carry out population censuses. Initially this was carried out using crude counts of buildings likely to be in domestic use. A scaling factor based on average occupancy rates was then applied to arrive at population totals. Subsequently, the net residential area approach was developed to provide more accurate population estimates. The use of aerial photography for population estimation is still widely used in Africa, Asia, and Latin America, both at national and at local level, particularly to monitor the growth of unplanned squatter settlements on the edges of cities.

One of the most significant applications of aerial photography has been its use for land use and land cover surveys, particularly in the United States by the USGS. After considerable preliminary work, the USGS formulated a classification scheme specifically for use with remotely sensed data. In practice, for all but the smallest mapping scales, this involved the use of aerial photography. Subsequently, many of the classification schemes used elsewhere in the world were based on the USGS approach. More recently, the Food and Agriculture Organisation developed a new approach, but there is, as yet, little reported use of the system.

Aerial photography continued to be important as a tool for military surveillance in the post war period, only being replaced by digital scanners toward the end of the century. The most famous military aerial photographic missions were those flown by the Lockheed U2 operated by the CIA. Designed to penetrate Soviet airspace above the range of air defences, it played a key role in the Cuban missile crisis of 1962. From the 1960s, aerial photography was supplemented by satellite data from programs such as Corona and Keyhole.

The early environmental satellites, such as Landsat, did not challenge the role of aerial photography as the dominant data source for land use or land cover studies. However, the launch of high resolution satellites at the end of the twentieth century has started to erode that dominance. The pricing policy in the United States has started to favor satellite imagery, especially as LiDAR is also challenging aerial photography as the usual data source for digital terrain modeling. Digital cameras designed to provide imagery for mapping are also starting to challenge traditional film cameras, but the large number of traditional film cameras in service are unlikely to be phased out in the immediate future.

Photogrammetry

Photogrammetry is the science and technology of making measurements using photographs. Although aerial photographs appear to be similar to maps in the way they present a plan view of the Earth, all aerial photographs will contain some positional errors, which means that they cannot be used directly to make accurate measurements of distances, areas, or directions. The photogrammetric process is designed either to derive accurate information from aerial photographs, or to create a photographic image from which all errors have been removed.

Laussedat first proposed the use of photographs for surveying and map making in the 1840s, but the photographic technology of the time was unsuitable for the purpose. Following the early work of Laussedat, which used a camera lucida, the main focus of developments was on architectural applications. Practical photogrammetry for geographic applications started in the 1880s using photographs taken from the ground, so called terrestrial photogrammetry. In the main, these early geographical applications were confined to mountainous areas where the camera could be used to provide images of areas that would be difficult or dangerous to survey on the ground.

Deville was responsible for the earliest attempts to use photography for mapping in North America when he used the technique for mapping during the construction of the Canadian Pacific Railroad, on boundary surveys and in the Rockies.

The first successful attempts at photogrammetry usually involved graphical techniques to plot the detail, but in the early years of the twentieth century independent researchers in Europe and South Africa developed instrumental methods that increased the accuracy and the speed of plotting. These early instrumental methods were only suitable for use with terrestrial photography; the first really successful instruments designed for use with aerial photography were not introduced until the late 1920s. This meant that the huge expansion in the use of aerial photography for map making and map revision during World War I was reliant on simple graphical techniques or the use of optical projection systems, such as the epidiascope. Nonetheless, aerial photography had become a standard mapping tool by the end of World War I.

The instruments first developed in the late 1920s, such as the Zeiss C4, the Wild A5, and the Multiplex, and those developed from these early models, were responsible for instrumental photogrammetry becoming accepted as the standard methodology for most mapping. Their development also led to the huge extension in the coverage of topographic mapping available. Prior to World War II, few countries in Africa, Asia, or South America had complete mapping cover at medium scale (in the range 1:25 000 to 1:100 000). While it can be argued that the Cold War was the reason for so much of the World being mapped in the decades after World War II, it was photogrammetry that made the mapping possible. Initially, a graphical technique, called radial line plotting, was also in widespread use, but by the late 1960s it had fallen into disuse.

In the 1970s, a new type of photogrammetric instru ment started to have an impact; this was the analytical plotter. The first generation of instruments, usually re ferred to as analog plotters, had relied on the optical reprojection of photographic images to create a theore tically correct model, or had a mechanical projection system, which mimicked the optical system. These in struments needed to be manufactured to very exacting standards, and required frequent calibration to ensure accurate measurements. This made the instruments ex pensive to manufacture and to maintain. In addition, the types of photography that could be used in them were limited and there was little scope for automation of the data collection process. Analytical instruments, which used a mathematical model instead of a physical model as a basis for measurement, offered significant advantages. The instruments could be used with almost any pho tography for which calibration data were available, or could be obtained; manufacture of the engineered parts was simple and relatively cheap, and they offered the possibility of automation. An additional important ad vantage of analytical instruments was that the operators did not require the same level of skill as that needed to achieve good results using analog instruments.

Analytical instruments were initially more expensive than analog ones, but this was largely a function of the high cost of the computing power necessary to compute model solutions in real time. However, as the cost of computing power fell in the 1980s, analytical instruments rapidly became far cheaper than analog ones, and manufacturers ceased to produce them. However, be cause analog instruments had been so widely adopted, they continued to be in use in many parts of the world.

The dominance of analytical instruments did not last long. In the early 1990s, the dominance was challenged by digital, or softcopy, photogrammetric solutions. In this approach, hardcopy photographs are replaced by scanned aerial photographs, or by scanner data. This permits the use of automated image matching (or correlation) tech niques to carry out the largely manual functions of the earlier solutions. In addition to speeding up a number of the processes, it also allowed the use of staff with a lower knowledge and skill base than earlier techniques. For the first time it opened up the possibility of desk top solu tions for nonphotogrammetrists.

There has also been a significant shift in the appli cations to which photogrammetry has been applied. Traditionally, it has been largely associated with topo graphic mapping at medium to large scales, and this, together with military applications, has been the main driving factor behind its development. In the 1950s, a new type of photogrammetry emerged, orthophoto graphy. The aim in orthophotography is to produce an image, which has the geometric properties of the con ventional line map, with the information content of the original aerial photograph or scanned image. Air photo based maps were produced from the 1930s, but these lacked the geometric properties of conventional line maps. A number of early photogrammetrists had at tempted to create orthophotographs; these had been unsuccessful due to technical limitations in the equip ment. The first important breakthrough was made at the USGS, where Russell K. Bean demonstrated a prototype plotter in 1955.

Until the early 1990s, orthophotography was primarily used as a substitute for line maps where a conventional map would give an inadequate representation of the landscape. Around the 1990s, two independent develop ments combinedly resulted in an increase in demand for orthophotographs, while making the images cheaper to produce. The increase in demand was generated by the increased use of GIS. Users of GIS began to realize that line maps were often inadequate backgrounds for GIS operations, and that the higher information content of orthophotographs made them much more suitable for this purpose. The development of digital photogram metric systems also led to a dramatic reduction in the cost of orthophoto production as they could be generated with very little operator input. A by product of ortho photo production is a digital terrain model (DTM), thus leading to more cost effective production of terrain visualisations and 'fly throughs'.

During the last years of the 'Cold War', the weapons programs of the US government created a demand for DTMs for missile targeting. Digital photogrammetry was, in part, a response to that need. The ending of the Cold War meant that technologies previously considered to be strategically sensitive could be derestricted and made more widely available. One of these technologies was high resolution satellite imaging. Previously, this technology had been restricted to so called spy satellites, but in the more relaxed atmosphere of the 1990s the US government sanctioned the launch of the first generation of high resolution civilian satellites. The data from these satellites, together with the ease of use of digital photo grammetry, and its relatively low cost, have led to its major impact on the use of photogrammetry in fields such as planning. Photogrammetry is now widely used to produce simulations of new developments by combining terrain visualizations with CAD drawings of proposed developments.

In the early twenty first century, the first digital cameras were introduced by the major manufacturers, which were designed specifically for aerial survey oper ations. These new cameras allowed mapping organiza tions to have a completely digital production flow from image through to final maps. This offers considerable cost advantages over the use of film cameras, where the film needs to be scanned to create the digital image for use in softcopy photogrammetric systems. The high initial cost of the cameras, and the existing capital investment in more traditional approaches, has deterred many organ izations from going entirely digital, but the introduction of a low cost camera by Vexcel is likely to lead to a more rapid adoption of the technology.

For much of the twentieth century, aerial photography for geographic analysis and aerial photography for photogrammetry were able to coexist quite happily. In optimal circumstances, the same imagery could be used for both purposes, reducing the cost of each accordingly. Early digital cameras designed specifically for photo grammetry threatened this relationship since they did not offer the possibility of stereoscopic viewing. The most recent generation of cameras allows for both photo grammetric and geographic users, restoring the mutually beneficial relationship.

See also: Cold War; Mapping, Topographic; Remote Sensing.

Further Reading

Babington Smith, C. (1957). *Evidence in Camera*. London: Chatto and Windus.

Blachut, T. J. and Burkhardt, R. (1989). *Historical Development of Photogrammetric Methods and Instruments*. Falls Church, VA: American Society of Photogrammetry and Remote Sensing.

Branch, M. C. (1948). *Harvard City Planning Studies, 14: Aerial Photography in Urban Planning and Research*. Cambridge, MA: Harvard University Press.

Branch, M. C. (1971). *Harvard City Planning Studies, 17: City Planning and Aerial Information*. Cambridge, MA: Harvard University Press.

Collier, P. (2002). The impact of technological change in data capture on topographic mapping between 1900 and 1939. *Cartography and Geographic Information Science* 29, 155 174.

Collier, P. (2006). The work of the British Government's Air Survey Committee and its impact on mapping in the Second World War. *Photogrammetric Record* 21, 100 109.

Collier, P. and Inkpen, R. (2003). Photogrammetry in the Ordnance Survey from Close to MacLeod. *Photogrammetric Record* 18, 224 243.

Colwell, R. N. (ed.) (1960). *Manual of Photographic Interpretation*. Washington, DC: American Society of Photogrammetry.

Crawford, O. G. S. and Keiller, A. (1928). *Wessex from the Air*. Oxford: Clarendon Press.

Finch, V. C. (1933). *The Geographic Society of Chicago, Bulletin No. 9: Geographic Surveying*. Chicago, IL: University of Chicago Press.

Hudson, G. D. (1936). The unit area method of land classification. *Annals of American Geographers* 26, 99 112.

Joerg, W. L. G. (1923). The use of airplane photography in city geography. *Annals of American Geographers* 13, 211.

Lee, W. T. (1920). Airplanes in geography. *Geographical Review* 10, 310 325.

McGlone, J. C. (ed.) (2004). *Manual of Photogrammetry* (5th edn.). Bethesda, MD: American Society of Photogrammetry and Remote Sensing.

Monmonier, M. (2002). Aerial photography at the Agricultural Adjustment Administration: Acreage controls, conservation benefits, and overhead surveillance in the 1930s. *Photogrammetric Engineering and Remote Sensing* 68, 1257 1261.

Walker, F. (1953). *Geography from the Air*. London: Methuen.

Relevant Websites

http://www.p.igp.ethz.ch
Chair of Photogrammetry and Remote Sensing, Geodetic Sciences at ETH Zurich.

http://www.ifp.uni stuttgart.de
Institut fur Photogrammetrie, Universitat Stuttgart.

http://www.photo.verm.tu muenchen.de
Institute of Photogrammetry and Cartography, Universitat Munchen.

http://www.isprs.org
International Society for Photogrammetry and Remote Sensing providing links to their online archives.

http://www.isprs.org
ISPRS Journal of Photogrammetry and Remote Sensing.

http://www.asprs.org
Photogrammetric Engineering and Remote Sensing.

http://www.rspsoc.org
Remote Sensing and Photogrammetry Society providing links to other relevant sites.

http://www.sli.unimelb.edu.au
University of Melbourne.

Photographs

R. Sanders, Temple University, Philadelphia, PA, USA

Glossary

Daguerreotype The daguerreotype is a direct-positive process (without the use of a negative) that creates a highly detailed image on a sheet of copper coated with a thin coat of silver.

Digital Image A representation that takes its form by means of discrete digits and encodes the data mathematically. It involves coding the information in bits and assigning each a mathematical value. Digital photography records images as binary data. This makes it possible to store and edit on computer and also gives the photographer the ability to manipulate and delete unsuccessful images immediately.

Ground Truthing A method used to confirm or validate directly from information or data derived indirectly.

Phenomenology How we react bodily and emotionally as well as intellectually to the world around us.

Photocontext Internal and external cues that enable the photograph to be trusted. Photocontext provides limitations to be assessed; it is multi-faceted, reflecting the academic discipline, research paradigm and theoretical framework, and the disparity between the photographer's culture, race/ethnicity, religion, gender, class and values, and the object of the photograph. No photograph is without context.

Photographic Truth The ability of photographs to be regarded as unmediated copies of reality. The myth of photographic truth suggests that photographs are understood to be evidence of actual people, events, and objects of the past, even though they are highly subjective and can be manipulated. The truth value of the photograph has been the subject of ongoing debate particularly with the increasing popularity of highly modifiable digital images.

Picture Theory It investigates differences and relations between images and words.

Realism The depiction of subjects as they appear in everyday life, without embellishment or interpretation.

Representation The act of portraying, depicting, symbolizing, or presenting the likeness of something. Much like maps, photographs function to depict and symbolize aspects of the real world. Representations re-present some aspect of the real.

The Gaze A term used to describe the act of looking particularly as it is complicit with desire; for example, the gaze can be motivated by a desire for control over its object. Theories of the gaze have explored the complex power relations that are a part of the acts of looking and being looked at. The gaze is intimately linked to imagination and fantasy.

Visuality The quality or state of being visual. Typically, visuality does not concern itself with photographs, which are designed to be seen, but rather with the ordinary, the quotidian, and the everyday.

Overview/Introduction

Much of geography, both human and physical, is based on the looking and observing the landscape. In human geography, many of the key ideas – areal differentiation, urbanization, environmental determinism, friction of distance, gentrification, globalization, mental maps, patriarchy, spatial interaction, and structure/agency – ultimately revolve around a consideration of how ideas ground themselves in space and place. As such they, or their effects, can be seen, processed, and monitored. Photography has been particularly useful in this regard. Despite its complexities, it is one of geography's most trusted scientific methods and has played an essential role in the formation of ideas about the meaning and meaningfulness of space and place. Indeed, photography is one of the ways in which geographers undertake analysis of what they encounter in the field, and simultaneously for many geographers, photography is how they do what they do. The image captured in a photograph provides a touchstone for theory as well as a way of establishing ground level realities. Because of the decisive place it holds in shaping notions of what is actual, it also lends itself to a discussion of what is featured beyond the frame and allows us to delve into complex issues such as representation, meaning, epistemology, and ontology.

Photography shares much with geography. Geography was catapulted into the world stage because of its contribution to exploration and colonization. So too was photography. Both helped usher in and shape the modern world and the very idea of what it means to be modern. Maps in particular provided an image of how the world was imagined and like photography, they also provided a world view. Both afforded new ways to collect, document, and display information and, in doing so, opened up new vistas. Metaphorically, geography and photography made the past a part of the present and reshaped the world by making the distant familiar and knowable. Because they share similar intentions – to impart a way of seeing in order to know the world and gain knowledge about people and place that is simultaneously scientific, rational, and objective; and subjective, experiential, and artful – both have the potential to bridge the separation between science, technology, and art/humanities and ultimately shed light on the human condition.

Quite simply, photography might be thought of as a technological process that yields a product. Accordingly, there are several ways of making sense of the process and the product. One way to think about the product is that it is an image recorded by a camera, fixed on a negative, and printed on paper. It is a physical object, a print, and picture. What is pictured is exactly what is there. Thus, photography captures what actually exists.

Geographers, influenced by recent work in decon structionism and interpretative analysis, have begun to interrogate the practice of photography and inquire into its objectivity. Is photography and its product a truthful instrument of simple observation or, are both process and product invested with intentionality and purposefulness? How is what is photographed different from the actual? How much is a photograph a product of technology versus a product of the mind? Thus, another way to think about photography is through the questions it spawns.

For example, consider when we read words on a page. They create an idea which we can see or visualize in some way. That the words on a page produce an image imbued with meaning suggests that our subconscious, in processing characters or images on a page, summons a level of mental acumen that signals our brain's perceptual apparatus to give a 'spin' to what we see. That is, what we chose to attend to and how we frame what we see and do is not entirely random. Unbeknownst to us, deep in the interstices of our brains are structures that function as repositories of signs, texts, symbols, meanings, and intentions. What the photograph is showing does not mirror it. The mind elaborates, fills in voids, embellishes, and proposes. The interpretation we bring is not void of context, *a priori* social construction, and purposiveness.

Still another way to think about photography is that it is a scientific endeavor, a means of collecting and classifying information and communicating it to popular, elite, and scientific audiences. In this way, the process is equivalent to collecting any sort of data and the product or what is represented in a photograph is not much different from a number or a specimen in a jar. It is theoretically informed and it allows us to understand a physical, historical, and social phenomenon that is implanted in our minds or on a landscape.

Photography might also be thought of institutionally, the product of technological know how at a given time. Within that technical establishment, photography is a means of capturing a discrete moment in the production and circulation of cultural meaning. In this way, it is once again a literal description of what exists – in the culture at that time; but without the available technology the description could not occur. When all is said and done, photography is a complex and far reaching communi cative medium – empirical, informational, and expressive all at once.

History

Photography first appeared in Europe in the early nineteenth century during a particular moment when concepts of positivist science held sway. It relied on the technical instruments available at the time and thus was seen as scientific and reliable.

The immediate predecessor of the photograph was the daguerreotype developed by the French inventor, Louis Daguerre. The daguerreotype, though cumbersome and unwieldy, was an improvement over previous image making techniques that required vast amounts of time and effort to yield a final product. The daguerreotype was also revolutionary. It gave an illusion of integrity and truth that could not be achieved by pen and ink and import antly was commercially viable. Despite this, the process and instruments associated with the daguerreotype were still burdensome, awkward, and time consuming. Pho tography and the photograph provided the perfect solu tion. By comparison, the process and instruments were manageable, easy, and convenient. More importantly, photographic truth was not sacrificed.

The invention of photography and the use of the photograph also coincided with the rise of the modern nation state. In geography, photography was used as an aid in establishing nation and empire. It facilitated travel and discovery, and furthered landscape description and state administration. Schwartz and Ryan (2003) draw on a quote by William Lake Price to describe how geog raphers perceived early photography.

> In a multiplicity of ways, photography has already added and will increasingly contribute to the knowledge and happiness of mankind: by its means the aspect of our globe from the tropics to the poles its inhabitants, … its productions, animal and vegetable, the aspect of its cities, … are made familiar to us. (William Lake Price, 1868 cited in Schwartz and Ryan (2003))

The photograph also played an integral part in the scientific professions and the regulation of social behavior by bureaucratic institutions of the state. One of the earliest uses of the photograph in the social sciences was classification. Anthropologists and sociologists in particular used photographs to gather and report detailed information that would aid in classification and categor ization of human populations in order to distinguish the 'normal' from the 'abnormal' and further specific social agendas. The photograph was highly valued for its ability to document and record ethnographic data and was in strumental in furthering claims of racial superiority and classifying and cataloging racial groups and racial char acteristics. All of these geographical, sociological, and anthropological endeavors were highly politicized, thus raised concerns about photography and the photograph.

Ways of Thinking about the Photograph/ Photography

At the heart of the photograph is its ability to construct knowledge (documentary, evidentiary, identification, ideo logical, social control, and forensic). This singular trait underlies its acceptance. Ultimately, it can be viewed as a piece of data, not unlike those used in more traditional analyses. Like all data, the photograph is produced in a context dictated by material and social infrastructure, disciplinary paradigms, and thematic/theoretical and stylistic norms/codes. The discipline trained picture taker has intimate knowledge of both the subject matter and the media and draws on specific discourses to produce nor mative images or photographs that make sense. In other words, the discipline in which photographic activity occurs triggers a system of codes, conventions, and a set of ex pectations. Just as researchers are partisan in their reading of numerical data and apply conventions of statistical an alysis and discipline based theories, when taking a photograph, social scientists are attuned to look at and for the essential elements that their training tells them are important.

Beyond academic training, the orientation of the re searcher in time and place, the agenda of the fact gath erers, and the frame of reference of the audience are also important. Together they conspire to affect and constrain meaning making. Geographic fieldwork is a two way reflexive relationship that draws on the researcher's skills of hearing, observing, listening, touching, and feeling. It is not an emotionless undertaking, void of human inter pretation, and human intervention. To its credit, pho tography acknowledges this.

Proponents of photography maintain that photographs are 'better' than numbers in that they capture realities in a way that numbers cannot. Numbers may be statistically significant on paper, but the reality they represent can only be understood by seeing how it grounds itself on the landscape. The 'hopelessness and desolation' associated with deindustrialization or the 'sense of danger' an unaccompanied woman feels on a deserted city street simply cannot be understood with numbers alone. The link between theory and ground level reality becomes clearer using a photograph that can freeze an idea in a time and space that is shared by those who witness what is pictured and those who take the picture. This is the power of photography.

It is precisely this power or presumed truthfulness of photography that has attracted the attention of geog raphers and prompted powerful critiques ranging from how photographs are understood to issues of interpret ation. These concerns have less to do with the technology and what can be produced using that technology and more to do with how we understand what is pictured. Thus, it is the discourse around photography – what it reveals and what it conceals that has received most attention. So, in historical geography, for example, the purpose of photography is only secondarily to allow twenty first century eyes to look back in time. The pri mary purpose is to allow the viewer to understand and appreciate nineteenth century ideas about space, iden tity, and power.

Contemporary Research Practices

Some of the earliest and most frequent uses of photog raphy in geography were documentary photography and landscape photography. The aim of documentary pho tography is objectivity – to produce a piece of data that accurately records what it intends to capture. Geog raphers use documentary photography to illustrate ideas, to give visual support to claims and assertions, and to provide the reader with an image of a site or structure referred to in the text. Landscape photography, on the other hand, has been used by geographers to detail fea tures of a place or region. In research, these are literal representations used to provide a survey for purposes of documentation. In other instances, landscape photog raphy has esthetic purposes: working with light, com position, and orientation to achieve just the right result.

Photography is also the basis of much of the inter pretative and deconstructive analysis that is undertaken in geography. Interpretative analysis addresses issues of power and cultural meaning. It is based on the assump tion that visual imagery is never innocent and thus untangling the relationship between the picture taker, the place/person, and the audience is a necessary starting point in photographic method. Understanding motive, intent, and codes is also important. There are no hard and fast rules for deciding what to decode and what signs to interpret, but deconstruction and interpretation pro vide the basis for addressing concerns not tackled in other photographic methodologies.

'Photo elicitation' is another photographic method used by geographers. It takes many different forms but is used primarily in geographic education with school children. It is carried out by interviewing groups or individuals who respond to visual, rather than lexical, prompts. It is similar to the open ended interview in that it actively involves the subject in the research process. The interviewee and the interviewer discuss pictures together in a nonstructured format. Successful photo elicitation requires some foreknowledge of the respondents group's use of photographs so that it can be assessed within the context of the informant's shared meanings. A single photo or a set of photographs are assembled by the researcher on the basis of prior analysis with the assumption that they have significance for interviewees. They are then shown to individuals to explore their knowledge, values, beliefs, and attitudes. The photographs may also be used to trigger reactions and explore how individuals and groups engage with a place. Photo elicitation often leads to further questions by unraveling individual semiotic constructions and untangling the connection between how people understand and experience place and scientific theories that explain what happens in place.

'Reflexive photography' has been used more in research undertaken by visual sociologists than geographers, although a few geographers have begun to explore its possibilities. Reflexive photography draws on standpoint theory and shifts the point of view away from the interviewer and places the interviewee in the center of the meaning making. It consists of the test person taking a photograph independently of the interviewer and later, in the interview situation commenting on what it mean. The interviewee then becomes the expert on his/her own photographs and the meaning they hold. As with photo elicitation, in geography reflexive photography is most commonly used in a classroom context with school children.

Both physical and human geographers use 'repeat photography' to monitor changes on the landscape. Repeat photography consists of five steps: (1) selection of the phenomenon – this could be a place (a neighborhood, community, street corner, worksite, or park) or a social group (a gang or family); (2) identification of the visual indices of change, informed by theoretical understandings; (3) collection or acquisition of the base line data; (4) setting the time frame or 'bracketing' of the work – the time interval might be the course of 1 day, several months, or several years; and (5) follow up in which the original sites, people, or archives are relocated and rephotographed. When going back to the original sites, it is ideal to use the same vantage point and angle of view if possible. **Figures 1a–1c** are pictures of Cecil B. Moore Avenue in Philadelphia taken over 8 years. Seeing the street in this way provides a visible record of shifts in circumstances. Not unlike the ground truthing that occurs in remote sensing, repeat photography confirms or validates the ground level reality based on information or data derived from other sources. Repeat photography can be used to ask questions that arise only by returning to the exact place where the photograph was taken. The photographs then act as a catalyst to generate new ways of thinking about the landscape, but as important it provides a way of gauging how ideas ground themselves on the landscape over time.

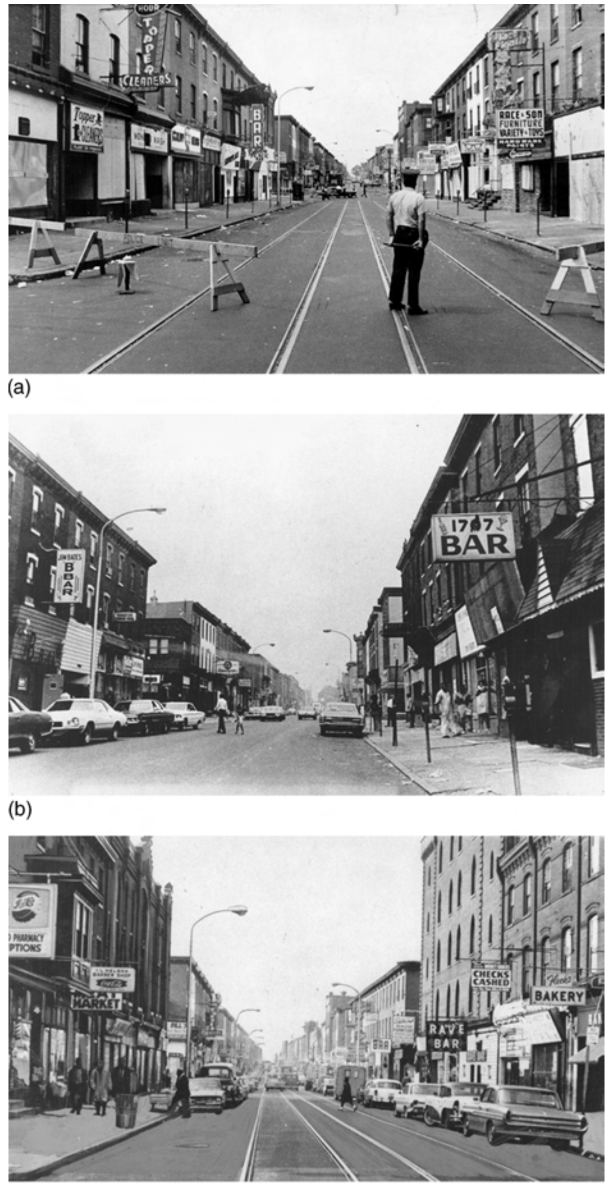

(a)
(b)
(c)

Figure 1 Pictures of Cecil B. Moore Street in Philadelphia: (a) 1961; (b) 1965; and (c) 1969. (a c) Courtesy of Temple University Library Urban Archives.

'Paired comparisons' employ duos of photographs focused on a common theme to arrive at macrolevel generalizations. In sociology and anthropology, paired comparisons explore the juncture between personally meaningful, context based experiences and large scale, theoretically informed understandings of social phenomena. In geography, paired comparisons typically explore the differences in how identical phenomena impress themselves in different spatial contexts, for example, how ideas around power look in capitalist and

Figure 2 Food sources: (a) Bryn Mawr, PA; (b) North Philadelphia. (a, b) Courtesy of David Colangelo.

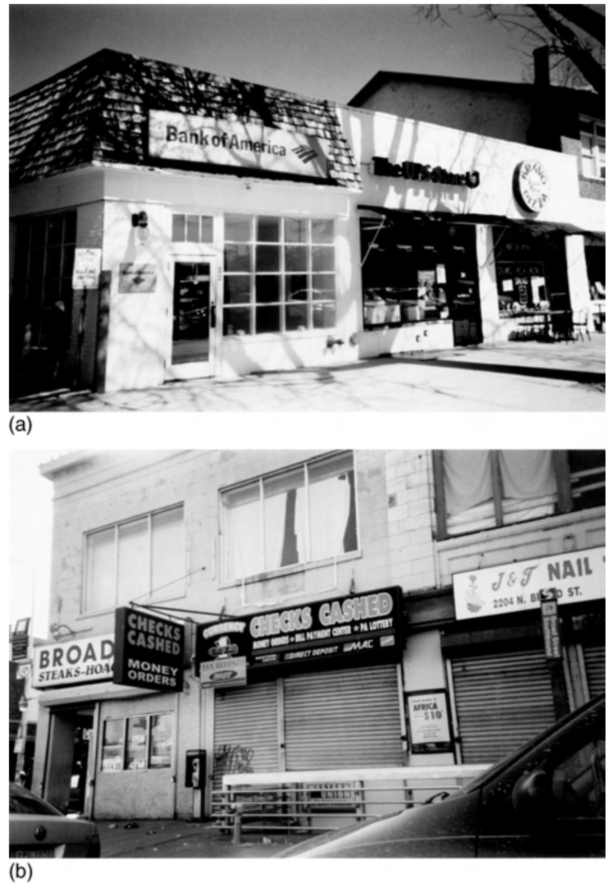

Figure 3 Financial institutions: (a) Bryn Mawr, PA; (b) North Philadelphia. (a, b) Courtesy of David Colangelo.

socialist countries or how low income housing looks in the European and American context. **Figures 2a** and **2b**, and **3a** and **3b** contrast food markets and financial institutions in two neighborhoods of Philadelphia – Bryn Mawr (a predominantly white, upper income area) and North Philadelphia (a predominantly black, lower / middle income area).

Reactions/Critique

The photograph as a method of conducting research has raised numerous issues. In historical geography for instance, the photograph has been criticized for being a servant of empire and for favoring the efforts of imperialism. In the early days of European expansionism, geographers frequently used the photograph to assist government in its efforts to extend territorial control, establish boundaries, inventory colonial holdings, and document the success or failure of development initiatives writ large. This complicity with imperialism has made geographers suspicious of photography's claims to knowledge and has prompted the perception that the photograph is far from objective but rather serves the interests of those who make them, those who commission them, those who publish them, and the audiences who consume them. Because of photography's complicity with an agenda – both overt and covert – it should be regarded as performing ideological work.

Another issue is interpretation. A photograph carries both informative as well as affective power. Its informative function is grounded in empiricism in that it represents the real world by substituting the name of an attribute for that of the intended subject and standing in for an object or event in its spatial and/or temporal boundaries. In other words, a photograph of a man's face is not a man. It stands for a man and perhaps even for a class of men. A photograph of an inner city housing project is not just a housing project. It contains an idea and a feeling about housing projects in general. A photograph is also deeply personal and privately engaged. It is a representation, an abstraction extracted from the totality of an experience and foregrounded into consciousness. Visual anthropologists are quick to point out that through a photograph, we share a feeling, but not necessarily the same one the photographer had for the object. It might not even be the feeling that the photographer intended us to have. S/he recognized the suggestive power of the image to direct us into a specific and known feeling, state of being, or place. Distant people and places are made intelligible in pictures. But the perceptions of these people are not alien. They have been suggested to us before in some way. Thus, a photograph is "an idea that fills the mind and satisfies one's previously conceived expectations."

Digital photography raises yet other issues. The line between retouching an image and reconstructing an image grows thinner and thinner as technology stretches the limits of what is possible. Minute changes may be just that or they may be the beginning of a slippery slope. Digital photography raises issues regarding permanence and accessibility. The images stored on disk, USB ports, and hard drives are seen as being more vulnerable than images printed on paper and stored in family albums, in galleries, or even in shoeboxes. More so, the digital technologies can become obsolete and computers can crash and the images stores on them can disappear.

Due to the cultural turn, the photograph enjoys a special place in geography because of the discussion it spawns regarding power and meaning, but aside from the discourse photography is indeed a mechanism of dis covery in its own right.

See also: Critical Geography; Cultural Turn; Culture; Film; Historical Geography; Positivism/Positivist Geography; Subjectivity; Travel and Travel-Writing; Visuality.

Further Reading

Banks, M. (2001). *Visual Methods in Social Research*. Thousand Oaks, CA: Sage Publications.

Barthes, R. (1982). *Camera Lucida: Reflections on Photography*. Howard, R. (trans.). New York: Farrar, Strauss, and Giroux.

Bunge, W. (1971). *Fitzgerald: A Geography of a Revolution*. Cambridge: Schenkwen Publishers.

Burgin, V. (ed.) (1982). *Thinking Photography*. London: Macmillan.

Collier, J. and Collier, M. (1986). *Visual Anthropology: Photography as a Research Method*. Albuquerque: University of New Mexico Press.

Crang, M. (1997). Picturing practices: Research through the tourist gaze. *Progress in Human Geography* 21, 359 373.

Crang, M. (2003). Qualitative methods: Touchy, feely, look see? *Progress in Human Geography* 27, 494 504.

Harper, D. (2002). Talking about pictures: A case for photo elicitation. *Visual Studies* 1, 13 26.

Hay, I. (2000). *Qualitative Research Methods in Human Geography*. London: Oxford University Press.

Latham, A. (2003). Research, performance, and doing human geography: Some reflections on the diary photo diary interview method. *Environment and Planning A* 35, 1993 2017.

Lutz, C. A. and Collins, J. L. (1993). *Reading National Geographic*. Chicago, IL: University of Chicago Press.

Prosser, J. (ed.) (1998). *Image Based Research: A Sourcebook for Qualitative Researchers*. London: Routledge Falmer.

Rose, G. (2000). Practicing photography: An archive, a study, some photographs and a researcher. *Journal of Historical Geography* 26, 96 101.

Rose, G. (2001). *Visual Methodologies*. Thousand Oaks, CA: Sage Publications.

Schwartz, J. and Ryan, J. (eds.) (2003). *Picturing Place: Photography and the Geographical Imagination*. London: I.B. Tauris.

Shurmer Smith, P. (2002). *Doing Cultural Geography*. Thousand Oaks, CA: Sage Publications.

Sidaway, J. D. (2002). Photography as geographical fieldwork. *Journal of Geography in Higher Education* 26, 95 103.

Sturken, M. and Cartwright, L. (2001). *Practices of Looking: An Introduction to Visual Culture*. Oxford: Oxford University Press.

Wells, L. (ed.) (1997). *Photography: A Critical Introduction*. New York: Routledge.

Relevant Websites

http://www.geog.leeds.ac.uk/images
UNI of Leeds, School of Geography, Images.

Physical Geography and Human Geography

S. Harrison, University of Exeter, Penryn, UK

Introduction

Traditionally, geography is seen as the subject charged with describing and analyzing the spatial and temporal variation of phenomena over the Earth's surface. As such, it incorporates perspectives from a very wide range of diverse subjects from the social sciences (including history, sociology, politics, economics, and psychology) and from the natural sciences (including physics, biology, ecology, and geology). To some, this diversity has been regarded as a sign of weakness, suggesting that a subject with such a broad set of research fields must treat those in a superficial manner. Others have seen it as a sign of strength, arguing that geography has avoided the intellectual trap of increased specialization and held on to a holistic view of Earth phenomena. While this particular debate continues, there is a general feeling that geographers have moved considerably beyond Pattison's suggestion, in 1964, that geography embraces four main areas of interest: the spatial tradition; area studies; man–land; Earth sciences.

However, within these four traditions and at various times throughout geography's history as an academic subject, its two main strands, human and physical geography, have interacted loosely, joined forces to provide hybrid geographies and evolved and worked separately. This discussion of the nature and history of the links between human and physical geography aims to examine a number of important themes and milestones within the development of the subdisciplines.

Many of the issues we face – ranging from globalization to migration, geopolitics of natural resources to climate change, racism to soil erosion – require a broader perspective than that offered by many of the other natural and social sciences. A successful assessment of soil erosion, for instance, requires an appreciation of a range of implicated factors including soil structure, farming methods, climatology, and land use change. The interdisciplinarity needed is therefore seen as a crucial approach for subjects seeking to understand many contemporary geographical issues and geography as a subject is unique among academic disciplines in its ability to offer a range of perspectives on these. However, this recent call for 'inter and postdisciplinarity' should not be taken to signify that geographers in the past have neglected such interactions. It can be argued that academic geography in North America and much of Europe began as an observational subject; concerned mainly with issues of regions and territory. In many ways, viewing human activity and the natural environment (within which this activity was pursued and which imposed constraints and provided opportunities) as an interrelated subject of study, can be traced back to the beginnings of academic geography in the eighteenth and nineteenth centuries. This point will be developed by looking at the works of geographers in North America and continental Europe.

A History

Contrary to contemporary geographic traditions where the distinction between human and physical geographers is regularly made, in nineteenth century America, geography was seen as an adjunct of geology. Perhaps, the most influential American geographer of this time was William Morris Davis (1850–1934). He started work as an assistant to Nathaniel Shayler at Harvard in 1876 and became an instructor of physical geography in the department of geology at Harvard in 1878. He is best remembered for his qualitative analysis of the ways in which landscapes develop. His view was that landscapes develop through time over a cycle of erosion. Time became elevated to the status of process and landscapes were seen as evolving slowly through the stages of 'youth' and 'maturity' to the endpoint of 'old age', after which tectonic activity and/or climate change could result in their rejuvenation. This theory formed a very seductive research and teaching tool, combining simplicity and elegance with the concepts of evolution and uniformitarianism; two dominant intellectual concerns of the nineteenth century.

The Davisian approach to research was built on three main characteristics. First, his assessments of landscape change were made following detailed and objective field observations, and the products of this approach are the accurate field sketches by which he is justifiably remembered. Second, Davis was at pains to highlight the importance of understanding the processes which drove landscape change, although his inability to measure these empirically was one of the main criticisms leveled at him in attacks on his work during the 1950s and 1960s by the new breed of process geomorphologists. Third, he had a deep appreciation of the interrelationships between human and natural factors in understanding landscape change. As Davis wrote:

> …any statement is of geographical quality if it contains a reasonable relation between some organic element of the earth on which we live, acting as a control, and some

> elements of the growth or behaviour or distribution of the earth's organic inhabitants, serving as a response. (Davis, 1909)

This theme of the symbiotic nature of life and landscape was also highlighted by the influential American geographer Ellen Churchill Semple (1863–1932) who brought many of Friedrich Ratzell's ideas to an English speaking audience. She is credited with developing a geographical philosophy based upon environmental determinism.

Criticism followed this work; many of her ideas about the ways in which the natural environment determined the development of human society and attributes were dismissed because of their potentially racist connotations. Critics argued that describing human societies as purely a product of their natural environment was a shortcut to justifying imperial and colonial actions. Criticism of such environmental determinism also followed because the concept failed to recognize that humans were able to modify the landscape deliberately and inadvertently.

However, this may be a partial misreading of Semple's position because, like Davis, she too embraced the idea that humans both affected, and were affected by, the landscape. She wrote in the beginning of her seminal book *Influences of Geographic Environment*:

> Man's relations to his environment are infinitely more numerous and complex than those of the most highly organized plant or animal. So complex are they that they constitute a legitimate and necessary object of special study [….] man has been so noisy about the way that he has 'conquered nature', and Nature has been so silent in her persistent influence over man, that the geographic factor in the equation of human development has been overlooked.

Outside of North America, the dominant schools of geographic thought were based in France and Germany. In France, a tradition of regional geography developed that took the region as a fundamental unit of study, and used understandings of physical and human influences on the region as its methodology. The major exponent of this approach was Paul Vidal de la Blache (1845–1918).

His approach to geography was to stress the interconnectedness of the environment and the human influence on it and, in doing this, he developed a dualistic way of seeing these interactions – the *pays* and the *genre de vie*. A *pays* is a natural region with defined boundaries (such as a mountain region, or an island) where Vidal believed the landscape to have played an important role in driving the development of human activity. This activity is constituted through the *genre de vie*, a cultural context enveloping language, traditions, institutions, food, etc. As a result, a landscape may possess both physical and cultural sets of characteristics.

Vidal's work can also be seen as a response to environmental determinism. He argued that the landscape set the paradigm within which human activity was played out – creating opportunities as well as setting limitations on human endeavor. This possibilism came to be seen as a much more sophisticated treatment of the ways in which natural and human systems interact than the simplistic causality espoused by determinism.

By the middle of the twentieth century, geography could be seen to have developed along twin tracks; a strong regional tradition in France and among some North American geographers, and a distinct systematic focus, largely driven by German geographers. Up to now, geographers would have defined themselves on the basis of their object of study. However, an important philosophical program had developed among European mathematicians, physicists, and philosophers in the 1920s and 1930s which was to radically alter the future direction of geography by introducing the idea that unifying the different strands of geographical practice could be achieved on 'methodological' grounds. The Vienna Circle was an enormously influential group of mainly German and Austrian thinkers which included philosophers such as Moritz Schlick and Otto Neurath and the mathematicians and logicians Kurt Gödel, Rudolf Carnap, and Ernst Mach. They developed a methodology based upon empiricist logical analysis which allowed them to produce a manifesto setting out the ways in which discrimination could be made between scientific and nonscientific analyses based upon metaphysics. By constructing a philosophy developed from logical positivism, they argued that assertions about the world are only scientific and meaningful when based upon empirical and verifiable observations. The impact this group had on many subjects was very considerable as it presented a methodological rationale for differentiating between scientific and nonscientific subjects. While the German geographic tradition emphasized systematic study, geography in North America and Britain had a strong regional focus and this concentration on what was seen as primarily, and perhaps unfairly, an idiographic philosophy (one based upon the study of unique phenomena) was viewed as being metaphysical by the logical positivists.

At this stage, there was a clear future roadmap for academic subjects. Those that wanted it to be seen as a science (with all the sociological cachet that this brings) had to embrace the methodology of the logical positivists. This initiated a vigorous debate between geographers concerned with the future direction and status of the subject. For physical geographers, accepting the logical positivist movement was a natural step toward placing models of landform evolution onto a more mathematical and quantitative footing. This meant that many of the qualitative narratives used by Davis and his followers to describe landform evolution were now seen as out of date. This new physical geography was dominated by a focus on mathematical, physical descriptions of process, and the construction of mathematical models with pre dictive capabilities, an approach which could be seen as a vindication of the pioneering research on hydrology by the American geologist Grove Karl Gilbert (1843–1918), a contemporary of W M Davis.

For human geographers too, the language of positiv ism ushered in a new direction for research. One of the important ways in which this new methodology was in tegrated into human geography was via spatial analysis. Spatial analysis had a long tradition in geography, but with the development of new methods of analysis and observation such as computers and aerial mapping, the geographers' concern with locational variables was rejuvenated. In the 1950s and 1960s, geographers such as Peter Haggett in the British Isles and Edward Ulman in the United States pushed forward this new field of geography. By integrating it with geographic information systems, physical geographers also used the tools of spatial analysis to map distributions of soils, ecologies, landforms, and climates. Both human and physical geography at this time had therefore adopted the meth odologies of scientific method to study the complexities of the natural and human worlds, thus placing them firmly in the camp of other complex sciences such as geology and ecology.

However, this changed direction for human geography was not without its critics, and, over time, various dis senting voices such as the British geographer David Harvey and the American geomorphologist turned phe nomenologist Yi Fu Tuan, came to be heard increasingly loudly. In the rapidly changing political world of the 1960s, many human geographers began to reimagine different spatial forms of economics and social structures which would begin to address some of the problems of inequality and poverty. Associated with this was the development of different ideas around what constitutes legitimate politics, and these analyses introduced issues of race, gender, and different forms of capitalism into the intellectual mix. This period is often considered to represent a turn toward a more critical theory or philosophy in human geography, where many of the mathematical structures employed by the spatial analysts were discarded as research tools in favor of more narrative, qualitative, and personal assess ments. Looking back at this time, however, it is clear that the cultural turn toward Marxist, humanist, and behavioral geographies was not a revolution in the subject with one catalyst, more an intellectual movement com prising a proliferation of (sometimes competing) ideas.

With the benefit of hindsight, we can now see that changing methodologies of analysis and observation gave the space for an increasing concern for theory and method. In certain ways, and over time, this allowed the subsequent split between human geography and physical geography to develop, as human geography began to institute a number of reimagined forms of theory as a means to becoming increasingly politicized. As a result, with this movement toward a more relativistic cultural geography, it became clear that many of the methodo logical links between human and physical geography were no longer in place, and we can argue that it is in geography that the development of CP Snow's two distinct cultures of 'art and science' came to become manifested most strongly perhaps in the academe. Additionally, much of physical geography had become increasingly specialized, following in the footsteps of the rest of science where the information explosion created by the advent of computers and the growth of science had made it extremely difficult for any individual to have an expert understanding of more than a small part of a discipline. As a result, physical geographers tended to see themselves as scientists in a specific field (geomorph ology, ecology, climatology, hydrology) who happened to work in geography departments, rather than as 'physical geographers' and this is borne out by the specialist nature of the journals in which many physical geographers concentrate their publications. Clearly, the relationship between human and physical geographers suffered as a result, with many on the physical sciences side, seeing human geography's increasing concern with issues such as queer theory, Marxism, feminism, and relativism as irrelevant to understanding the nature of environ mental systems. Many human geographers, too, held the alternative view that physical geographers were largely in thrall to logical positivism – a philosophy which many on the cultural side saw as an increasingly irrelevant and flawed philosophy. This split in the subject is seen in the changing pattern of publications in *Transactions of the Institute of British Geographers*, the leading journal of British geography. In the 1960s and 1970s, the journal was seen as an appropriate outlet for much physical geography research. By the 1990s, the journal was dominated by increasingly theory laden work by human geographers. This separation between the subdisciplines (and more generally between physical scientists and cultural the orists) perhaps reached its apotheosis in the Sokal affair of the late 1990s.

Alan Sokal, a quantum physicist, had become increasingly frustrated by the relativist philosophy movement underpinning much of cultural theory, which argued, among other things, that all knowledge was value laden and that, therefore, science was only one of a number of equally valid ways of looking at the world. This relativism was embraced by large parts of human geography and used a whole range of concepts from physics, mathematics, and logic to describe and explore issues in cultural theory. It was (and is) clear that these concepts were being used with little appreciation of their meaning or physical significance; indeed, many physical scientists were of the opinion that much of this appropriation was done to impart an aura of intellectual sophistication to subjects otherwise lacking in this. Sokal wanted to expose this, what he considered to be an intellectual fraud. As a physicist who was politically left wing, he also railed against the hegemonic nature of the paradigm within which large parts of the social sciences were operating, arguing that it stifled debate by suppressing dissenting voices, especially in the US and among young academics without tenure.

To expose the shallowness of the relativist movement, Sokal wrote a paper called 'Transgressing the boundaries: Towards a transformative hermeneutics of quantum gravity' and submitted it to the leading cultural studies journal *Social Text*, which was published in 1996. The paper was a hoax and this was revealed by Sokal in the same year in the journal *Lingua Franca*. While some cultural theorists and philosophers of science argued (and still argue) that Sokal had misunderstood the nature of the cultural theory program, suggesting that even specialized language must be able to have multiple meanings, it is clear that enormous damage was done to the intellectual credibility of postmodernism. While debate was rather muted in geography, the damage to the external (and perhaps internal) credibility of parts of human geography reinforced the views of many physical geographers that they stood to gain very little academically or intellectually by working with their human geography colleagues.

However, while the Sokal hoax certainly reinforced the damaging suspicion among physical geographers that the cultural turn in human geography had driven the two subdisciplines apart, substantial parts of human geography, such as economic geography and spatial theory, had remained essentially empiricist and rigorous throughout this period. As a result, it would be a mistake to believe that the cultural turn was the paradigm shift for human geography that the quantitative revolution was for physical geography.

At the beginning of the twenty first century, two other issues damaged the credibility of postmodernism. It has been argued that the first was the attack on the World Trade Center in New York in September 2001; the second was the realization that climate change was likely to become enormously significant and damaging to natural and human systems. Both of these called for a new pragmatism in politics and environmental stewardship, and both questioned the credibility of postmodern philosophies that had argued that 'nothing was real'. As a result, over the last few years there has been a move toward redeveloping the links between human and physical geography. This has been partly driven by renewed philosophical arguments between the subdisciplines, such as about the status of realism as a unifying philosophy of geography, and partly by pragmatic requirements. The latter can be illustrated with reference to three different debates and concerns: around nature and culture, applied geography, and climate change.

The Nature–Culture Debate

The idea that humans affect their natural environment, and that the natural environment plays a cultural role is, as we have seen, clearly not a new one and this view has been recently rejuvenated by human geographers such as Sarah Whatmore. However, in physical geography, the impact that humans have had on natural systems in the past has received renewed impetus by the development of improved dating techniques since the 1950s or so. In many regions, there is considerable and growing evidence that humans have been one of the main drivers of landscape change. For instance, in the uplands of the British Isles, periods of landscape instability from the middle of the Holocene coincide partially with phases of human occupation and consequent forest clearance and other land use changes. Landscape change in such areas is manifested by the construction of alluvial fans and river terraces, but there is debate about the extent to which this is driven by climate change, human activity, or a combination of these. Physical geographers have used a variety of techniques such as pollen analysis (palynology), analysis of tree rings (dendrochronology and dendroclimatology), and radiocarbon dating to assess the extent to which landscape change is forced by natural or human factors. This research area has also attracted archaeologists and human geographers who have added new insights into how humans colonized landscapes in Europe.

One of the major debates concerns the rapid nature of the landscape changes that have occurred since about 5000 BP (before present). At this time, there is a marked decline in many upland regions of certain tree species (such as birch and lime), probably associated with episodes of selective forest clearance by early agricultural communities, and with an abrupt deterioration in slope stability. Many physical and human geographers have been tempted by these temporal coincidences to develop models of landform development which stress the roles humans play in changing their natural environment.

Interdisciplinary research in the uplands has also been encouraged between archaeologists and cultural geographers and physical geographers from geomorphology and quaternary science backgrounds. For instance, in many of the upland moorlands of Europe, there exists a well developed periglacial (and therefore 'natural') landscape which has been partially modified by Neolithic and later human activity. This has created a landscape complex comprising elements which can be interpreted as being the result of geomorphological processes, and elements which form a ritual and cultural landscape employed, and developed, by humans. Discrimination between these is problematic, and requires the successful collaboration between human and physical geographers. In this context, cultural geographers, archaeologists, and anthropologists have been especially concerned with phenomenological issues whereby natural elements of landscapes are employed by humans to form a sense of place or 'home'. Other work has examined the different ways in which 'natural' landscapes can be constituted and what it is to be 'natural' in a world which has undergone enormous transformation by human agency.

Applied Geography

Geographers are increasingly aware of the political importance of academic subjects becoming more relevant in economic or societal terms. This movement comes with increased recognition of the costs incurred by society in developing and maintaining healthy higher education sectors. At the same time, many problems that societies face require integrated approaches to research. For instance, successful infrastructure planning and de velopment in geologically or climatically unstable regions calls for a risk mapping approach bringing together haz ard assessments from a physical and human perspective.

Climate Change

Perhaps the biggest issue which has allowed a rejuven ation of interactive research by human and physical geographers is climate change. There is now over whelming scientific consensus that human activity has changed the global climate, and will continue to do so in the foreseeable future, with potentially enormous implications for all natural systems on Earth. Taking a broad view of scientific understandings of climate change allows us to identify three distinct stages in the development of the concept. The first was the realm of physical scientists, such as atmospheric and computa tional physicists, and quantum chemists who developed models of atmospheric energy balance and radiative transfer. In the second phase, research began to look at the effects of climate change on a broad range of physical systems, and physical geography in all its manifestations has played a leading role in this program. For instance, glaciologists and cold climate geomorphologists have assessed the stability of ice sheets, mountain glaciers, permafrost, and related systems to predicted climate change. Fluvial geomorphologists have developed flood models to better predict where inundation will occur during flood events; and quaternary scientists have reconstructed past climate variability and estimates of climate sensitivity which have been used to put con temporary climate change into its historical context. The third stage has involved many of the social sciences working in collaboration with the physical sciences to explore the political, cultural, social, and economic responses to climate change and the policy implications of these. This last phase provides a set of major oppor tunities for human and physical geographers to integrate further their research agendas.

Human Geography and Physical Geography: The Future?

Judging the state of human and physical geography near the start of a new century depends on the viewpoint of the reader. It seems clear that there are two competing and opposing forces at work in the discipline. The first, and currently the one where research is actively being carried out, sees integration as an important program, starting with a series of conversations between re searchers across the disciplinary divide. This program argues that geography is stronger as a discipline when integrated. Recent movements in physical geography (especially perhaps in biogeography and geomorphology) toward understanding of the nature of emergence and interacting elements as important characteristics of nat ural systems, is seen by some as a welcome departure from the more reductionist process form modeling undertaken elsewhere in the discipline, and offers several ways of linking with philosophical and practical concerns in human geography. Indeed, addressing and under standing the complexities of climate change on Earth systems may well require such a 'holistic' approach, and hence there may be signs that the reductionist program has run its course. Such work has the potential to play an important role in the development of public policy toward issues such as climate change.

The alternative perspective stresses the increasing 'depth' of science, and argues that it is simply not possible to undertake scientific research without increasing specialization. Physical geography, viewed like this, has therefore become a loosely linked set of specialisms (geomorphology, climatology, hydrology, biogeography, and the like with their own specialist journals) whose researchers rarely interact 'within' physical geography, let alone between physical geography and human geog raphy. As a result, the future evolution of the subject may be seen to be one where increasing fragmentation rather than interaction is the norm. This latter position is cer tainly the case in many British universities where joint research and teaching between human and physical geographers is still rare.

Judging between these two outcomes is difficult, but it seems clear that the major geographical issues that we face will require human geographers and physical geographers to increase rather than reduce the research conversations and interactions between them.

See also: Climate Change; Environment; Landscape; Nature; Vidal de la Blache, P.; Wilderness.

Further Reading

Castree, N. (2005). *Nature*. Oxford: Routledge.
Castree, N., Rogers, A. and Sherman, D. (2005). *Questioning Geography*. Oxford: Blackwell Publishing.
Couper, P. R. (2007). Fluvial geomorphology and semiotics: A Wittgensteinian perspective of the 'divide' between human and physical geography. *Transaction of the Institute of British Geographers* 32, 279 294.
Harrison, S., Pile, S. and Thrift, N. J. (2004). *Patterned Ground*. London: Reaktion Books.
James, P. E. and Martin, G. J. (1981). *All Possible Worlds: A History of Geographical Ideas*. New York: Wiley.
Johnston, R. J. (2003). Geography: A different sort of discipline? *Transaction of the Institute of British Geographers* 28, 133 141.
Roberts, N. (1998). *The Holocene*. Oxford: Blackwell Publishing.
Sokal, A. and Bricmont, J. (1998). *Intellectual Impostures*. London: Profile Books.
Whatmore, S. (2002). *Hybrid Geographies*. London: Sage.

Place

T. Cresswell, Royal Holloway, University of London, Egham, UK

What Is Place?

Place lies at the center of geography's interests. In a commonsense way geography is about places. But the commonsense uses of the word place belie its conceptual complexity. While the word 'place' has been used as long as geography has been written, it is only since the 1970s that it has been conceptualized as a particular location that has acquired a set of meanings and attachments. Place is a meaningful site that combines location, locale, and sense of place. Location refers to an absolute point in space with a specific set of coordinates and measurable distances from other locations. Location refers to the 'where' of place. Locale refers to the material setting for social relations – the way a place looks. Locale includes the buildings, streets, parks, and other visible and tan gible aspects of a place. Sense of place refers to the more nebulous meanings associated with a place: the feelings and emotions a place evokes. These meanings can be individual and based on personal biography or they can be shared. Shared senses of place are based on mediation and representation. When we write 'Calcutta' or 'Rio' or 'Manchester' for instance, even those of us who have not been to these places have some sense of them – sets of meanings produced in films, literature, advertising, and other forms of mediation.

Consider the location 33.325° 44.422°. This location in abstract space marks the city of Baghdad in Iraq. While its location tells us where Baghdad is and enables us to locate it on a map or program it into a Global Positioning System, it does not really tell us much else. Baghdad is also a locale. It has mosques, homes, markets, barricades, and the Green Zone. It has a material structure that, in part, makes it a place. And finally Baghdad has senses of place. Some of the meanings associated with Baghdad are personal and vary according to whether you are an occupying soldier, a Sunni or Shi ite Muslim, someone who is trying to make a living, or a tourist who visited in the 1970s. But other meanings are shared and are not dependent on having been there. As Baghdad appears on our TV screens almost nightly, many of us in the Western world only know Baghdad as a war zone and a place of danger. Baghdad, like all places, has a location, a locale, and senses of place.

In any given place we encounter a combination of materiality, meaning, and practice. Most obviously, perhaps, places have a material structure. New York has its skyscrapers, Paris its boulevards, Los Angeles its freeways, and Sao Paulo its shanties. Places are often recognized in terms of their material structures which come to stand for place. Think of the Eiffel Tower. Think of the Sydney Opera House. On a more everyday level, towns and city neighborhoods across the world have their material form – libraries, shops, places of worship, streets, and sidewalks. In addition, places have all the material things that pass through them – commodities, vehicles, waste, and people. Even a totally imaginary place has an imaginary form in order to make it place like. The sense of place evoked by fantasy novels, for instance, is usually based on a description of the material environs. Think of the hobbit holes of *The Lord of the Rings* or the magic staircases of *Hogwarts*.

The idea of meaning has been central to notions of place since the 1970s in Human Geography. Location became place when it became meaningful. Meaning marks the most obvious difference between 33.325° 44.422° (a mere location) and 'Baghdad' – the place that occupies that location. Cruise missiles can be loaded with information like 33.325° 44.422° but not with 'Baghdad' and all the meanings that that place implies. Meanings can be very personal and connected to indi viduals and their personal biographies – places where we fell in love, or where loved ones are buried, or where we went to school. But meanings are also shared and, in some important ways, social. The Twin Towers of the World Trade Center in New York for instance had many shared meanings as they projected American power, the importance of capitalism, phallic masculinity, and so on. It is at least in part for this reason that they were attacked. While crashing four planes into empty fields would no doubt have been a disaster, it would have had a completely different impact from the destruction of such well known places. It was not just an attack on material structures but an attack on place – on meaning. Now of course, the site of the Twin Towers is acquiring new meanings which are still being contested. Will it be a quiet site of remembrance or a new site of American power projected across the world? While meanings are shared they are never fixed once and for all, and always open to counter meanings produced through other representations.

Finally, places are practiced. People do things in place. What they do, in part, is responsible for the meanings that a place might have. Most obviously places are left with the imprint of notable events such as battles and signings of treaties. A field in Belgium becomes a dif ferent place once we learn of the battles that were fought there in World War I. But more mundane practices are, perhaps, a more significant ingredient in place. Places are continuously enacted as people go about their everyday lives – going to work, doing the shopping, spending leisure time, and hanging out on street corners. The sense we get of a place is heavily dependent on practice and, particularly, the reiteration of practice on a regular basis. Space becomes a place when it is used and lived. Experience is at the heart of what place means.

Materiality, meaning, and practice are all linked. The material topography of place is made by people doing things according to the meanings they might wish a place to evoke. Meanings gain a measure of persistence when they are inscribed into the material landscape but are open to contestation by practices that do not conform to the expectations that come with place. Practices often do conform to some sense of what is appropriate in a particular place and are limited by the affordances particular material structures offer. While it is possible to skateboard on park benches, it is not possible to walk through walls.

While we most often think of place as the kinds of places mentioned so far – Baghdad, New York, Sydney – places can in fact exist at many scales. The corner of a favorite room is a place for a child who has little say in the constitution of the wider world. At the opposite extreme, the whole earth is a place when seen from outer space. Astronauts often commented on how the earth looks like home when it is seen from afar.

Chora and Topos

While it is the case that place was only formally conceptualized as a meaningful segment of geographical space by humanistic geographers in the 1970s, it is not true to say that the concept of place was invented by geographers. The origins of a philosophy of place can be seen in classical Greek philosophy and particularly in the writings of Plato and Aristotle. Plato (428–348 BC) developed the loosely defined notions of 'chora' and 'topos' in the context of an account of the origins of existence and the process of 'becoming'. Becoming, in Plato's terms, is a process that involves three elements – that which becomes, that which is the model for becoming, and the place or setting for becoming. This final element is 'chora', a term which implies both extent in space and the thing in that space that is in the process of 'becoming'. It is often translated as a receptacle and differs from the void of 'kenon' (abstract space) in that it always refers to a thing within it – it is not empty. Topos is often used interchangeably with chora in Plato but is usually more specific. While chora most often referred to a place in the process of becoming, topos would refer to an achieved place. Later Aristotle would use chora to describe a country while topos would describe a particular region or place within it. Both chora and topos would become part of geographical language through the notion of chorology (study of regions) and topography (the shape of the land surface). Both chora and topos are different from the notion of kenon (the void) in that they refer to something more particular – more like place than space. While kenon is limitless space chora and topos are finite and contain things.

If anything, Plato's student, Aristotle (384–322 BC) had even more fundamental things to say about place. To Aristotle, place was a necessary starting point from which it is possible to understand both space (the infinite, the void) and movement and change. Place, he wrote "takes precedence over all other things" (Casey, 1997: 71). To understand change and motion, for instance, it was first necessary to acknowledge that the "most general and basic kind [of] change is change in respect of place, which we call locomotion" (Casey, 1997: 51). The geographical question of 'where' is absolutely fundamental to Aristotle for everything that exists must be somewhere "because what is not is nowhere – where for instance is a goat stag or a sphinx?" (Aristotle in Casey, 1997; 51). To Aristotle, place comes first because everything that exists has to have a place – has to be located. Thus "that without which nothing else can exist, while it can exist without the others, must needs be first" (Casey, 1997: 52). So in Aristotle then we have a very powerful philosophy of place as the starting point for all other forms of existence. Philosophically, Aristotle marked a high point for thinking about place as philosophers turned to the seemingly more profound notion of space. Indeed, it was not until the late nineteenth and early twentieth centuries that place reemerged as a central philosophical concept – particularly in the work of Martin Heidegger.

Being There and Dwelling

Place was not completely forgotten about in the long centuries between Aristotle and Heidegger. In the thirteenth century, for instance, Albertus Magnus, a German Dominican scholar who taught Thomas Aquinas, studied the work of Aristotle (when many versions were banned by the church) and developed theories of the nature of places suggesting, for instance, that things (people, plants, stones) work best when they are in the place in which they belong and weaken the further they are removed from it. In the Arab world, geographers such as Ibn Batuta (1304–68) and Ibn Khaldun (1332–1406) made exhaustive travels throughout north Africa and as far afield as China describing the landscapes and customs of different places in great detail. They did, not however, develop general approaches to place in the process. Geographers such as Varanius (1622–50) argued strongly for the importance of a 'general geography' that concerned the whole world and its mathematics rather than the 'special geography' that focused on the particularities of regions and places. It was the works of Martin Heidegger (1889–1976), however, that proved to be particularly influential on the work of humanistic geographers who were to develop the notion of place in the 1970s. Heidegger, throughout his career, had struggled with the nature of 'being'. To Heidegger to be was to be 'somewhere'. The word he used to describe this was 'dasein' – or 'being there'. Note this was not simply being in some abstract sense, as if in a vacuum, but being 'there'. Human existence is existence 'in the world'. This idea of being in the world was developed in his notion of 'dwelling'. A way of being in the world was to build a world. Dwelling in this sense does not mean simply to dwell in (and build) a house, but to dwell in and build a whole world to which we are attached. Dwelling describes the way we exist in the world – the way we make the world meaningful, or place like. Most famously, Heidegger used the image of a cabin in the Black Forest to describe both building and dwelling.

> Let us think for a while of a farmhouse in the Black Forest, which was built some two hundred years ago by the dwelling of peasants. Here the self sufficiency of the power to let earth and heaven, divinities and mortals enter *in simple oneness* into things, ordered the house. It placed the farm on the wind sheltered mountain slope looking south, among the meadows close to the spring. It gave it the wide overhanging shingle roof whose proper slope bears up under the burden of snow, and which, reaching deep down, shields the chambers against the storms of the long winter nights. It did not forget the altar corner behind the community table; it made room in its chamber for the hallowed places of childbed and the "tree of the dead" for that is what they call a coffin there: the *Totenbaum* and in this way it designed for the different generations under one roof the character of their journey through time. A craft which, itself sprung from dwelling, still uses its tools and frames as things, built the farmhouse. (Heidegger, 1993: 300)

In this cabin everything seemed to have its place and the cabin sat almost organically in the natural world, linking the cosmological to the everyday. Here was the model kind of building and dwelling – a model kind of being in the world. Such a model, however, makes little sense to the modern, urban, hyperconnected life many of us lead. As a model it seems a little regressive and romantic. Indeed, this image has been connected by some to Heidegger's well known connections with the Nazi Party which often espoused a very similar ideology of rootedness in the deep soils of the Black Forest. The corollary of such a belief was that some kinds of people were unconnected to these deep soils and this manner of dwelling. Jews, gypsies, and urbanites, in general, were all seen as leading rootless and inauthentic existences. Jews, for instance, were symbolized as snakes crawling over the desert sands and the desert became a space where it was impossible to make roots. Authentic dwelling, Nazi ideology insisted, could not happen in the desert or the city.

Place and Humanistic Geography

Despite its central role in human geography, place was not self consciously written about until the 1970s and the advent of humanistic geography. The 1960s and early 1970s had been marked by spatial science, the quantita tive revolution, and logical positivism, all of which had looked at the world and the people in it as objects rather than subjects. People were most often thought of as rational actors in a rational world. The lexicon of spatial science included terms like 'location', 'spatial patterns', 'distance', and 'space'. Large parts of human geography had operated more or less as a pseudoscience. The focus on place was a central part of a humanistic critique of this way of thinking about the world and the human inhabitation of it. This critique is made clear in the final pages of Yi Fu Tuan's *Space and Place* (1977) where he compares the richness of an experiential perspective to more scientific approaches.

> What we cannot say in an acceptable scientific language we tend to deny or forget. A geographer speaks as though his knowledge of space and place were derived ex clusively from books, maps, aerial photographs, and structured field surveys. He writes as though people were endowed with mind and vision but no other sense with which to apprehend the world and find meaning in it. He and the architect planner tend to assume familiarity the fact that we are oriented in space and home in place rather than describe and try to understand what "being in the world" is truly like. (Yi Fu Tuan, 1977: 200 201)

And later, "The simple being, a convenient postulate of science and deliberate paper figure of propaganda, is only too easy for the man in the street – that is, most of us – to accept" (Yi Fu Tuan, 1997: 203). The term 'simple being' brings to mind the 'rational man' of spatial science and economics: the man who weighs up all options before making a rational choice about what to do next. Such a view of humanity has no place for meaning and is thus not concerned with place but only with 'location' or 'distance'.

Even though the humanistic engagement with place was the first attempt to define and conceptualize place as we know it today, it did not arise out of thin air.

In addition to the work of continental philosophies of meaning, such as Heidegger's, these geographers looked to what might broadly be called the 'regional geography' of the early decades of the twentieth century that has seen geographers attempting to produce a science of chorology that focused on the way various variables were interlinked in unique ways in space. First, the soil and the climate, then the natural landscape, then the cultural landscape and finally habits, customs, and beliefs. Place as such was rarely the focus of such work, the related concept of 'region' was more central. Of particular significance was the work of Paul Vidal de la Blache (1845–1918) and his followers in France. Vidal's deeply human geography had sought to examine the regions of France in terms of their distinctive ways of life. Region (pays) in Vidal's geography had been about seeing everything together *in situ*. He connected the physical environment with the cultural way of life (*genre de vie*). He noted how regions were marked by particular natural environments and cultural forms (clothing, food, archi tecture, institutions, etc.) that together formed specific regional life worlds. To put it most simply, you would know when you were in the Massif Central that you were in a different kind of place than if you were in northern France because of the unique combination of physical and cultural attributes that marked one place off from another. The physical landscape looks different, people wear different clothes, and people eat different food.

Inspired by philosophies of meaning such as phenomenology, the work of Vidal and elements of the earlier cultural geography of Carl Sauer, humanistic geographers insisted that geographers needed to think about people as knowing and feeling subjects rather than either objects or simply rational beings. To make geog raphy fully human, they argued, geographers needed to be more aware of the ways in which we inhabit and experience the world. Central to this awareness is the concept of place. This conception of place describes a way of relating to the world. It insists that people have the burden of making their own meaning in the world through their own actions. Key here is the idea of 'experience'. It is this notion of experience that lies at the heart of the humanistic approach to place. Ideas such as 'experience' were not in the vocabulary of human geographers in the early 1970s. Spatial scientists were not very interested in how people related to the world through experience. People as objects or rational beings are not 'experiencing' the world and geographers studying them are certainly not interested in how they experience the world. To focus on experience, therefore, was quite revolutionary. While the spatial scientists wanted to understand the world and treated people as part of that world (just like rocks, or cars, or ice but with the magic ingredient of rationality added), humanistic geographers focused on the relationship 'between' people and the world through the realm of 'experience'. Tuan writes "[t]he given cannot be known in itself. What can be known is a reality that is a construct of experience, a creation of feeling and thought" (Tuan, 1977: 9). The new focus on place, therefore, attends to how we, as humans, are in the world – how we relate to our environment and make it into place.

What experience does is transform a scientific notion of space into a relatively lived and meaningful notion of place. While space was the favored object of the spatial scientist (and is still the favored object of social theorists), it is the way space becomes endowed with human meaning and is transformed into place that lies at the heart of humanistic geography.

> What begins as undifferentiated space becomes place as we get to know it better and endow it with value … the ideas "space" and "place" require each other for defin ition. From the security and stability of place we are aware of the openness, freedom and threat of space, and vice versa. Furthermore, if we think of space as that which allows movement, then place is pause; each pause in movement makes it possible for location to be trans formed into place. (Tuan, 1977: 6)

This is the most important contribution of humanistic geography to the discipline – the distinction between an abstract realm of space and an experienced and felt world of place.

The humanistic division between space and place became a taken for granted distinction in human geog raphy by the early 1980s. This distinction is still taught to students across the English speaking world. Geographers have continued to develop broadly humanistic notions of place. Robert Sack, for instance, has considered the ways place as a center of human meaning ties together worlds that are normally held apart – the worlds of nature, meaning, and society. In addition, he has developed an ethical and moral framework for human action based on a humanistic notion of place. More importantly, perhaps, many of Tuan's ideas have traveled into geographies which are not explicitly humanistic and indeed to other disciplines entirely. The notion of place, for instance, has been developed by philosophers such as Edward Casey and J. E. Malpas.

The Social Construction of Place

While humanistic geographers developed a rich idea of place as experienced, felt, and sensed, they did not, on the whole, have much to say about how power is implicated in the construction, reproduction, and con testation of places and their meanings. The humans in 'humanism' had a tendency to either remain determinedly individual or, alternatively, as but instances of a ‘universal’ humanist subject. Geographers inspired by both humanism and the radical approaches of Marxism, feminism, and post structuralism began, in the 1980s, to develop a critical approach to place which sought to rectify this problem.

The kind of organic, rooted, and bounded place evoked by Heidegger’s notion of dwelling began to be seen as limiting and exclusionary. It was also perceived as a very ecological way of thinking about place as though human places were natural, authentic – the way they were supposed to be. Marxists in particular began to point to the social processes (particularly under capitalism) that are involved in the construction of places. Places, they argue, may seem natural but are in fact anything but that. The material structure of a place, even a log cabin in the Black Forest, is often the result of decisions made by the very powerful to serve their ends. Most of us, after all, only get to build places on a relatively small (but nonetheless important) scale. The meanings associated with these places, insofar as they are shared, are also more likely than not to be meanings assigned to place by people with the power to do so – the people who build the buildings and monuments and inscribe texts on to the material fabric of place. All of these involved choices that exclude people and the meanings they represent. It is observations such as this that led David Harvey to write that: “The first step down the road is to insist that place in whatever guise, is like space and time, a social construct. The only interesting question that can be asked is, by what social process(es) is place constructed?” (Harvey, 1993: 5). While humanists sought to show how place was a fundamental and universal ingredient in the ‘being in the world’ in a way that transcended particular social identities and divisions, Harvey insisted that place was often used in quite regressive and reactionary ways. He points to the rise of gated communities in the United States and other defensive place based definitions of community (such as emergent nationalism in the Balkans at the time) that are, more often than not, based on some threatening outside that is being kept out. This, then, is the dark side of Heideggerean notions of place inherited by humanistic geography.

These tensions can be best illustrated by the notion of home. To humanistic geographers home is a particularly ideal kind of place – the location where meanings and attachments are most intense. Heidegger’s cabin is one example of this kind of home place – where everything is where it should be and all is right with the world. Home place is a center of meaning and field of care – a place where (in an ideal world) we feel safe, secure, and loved. Home is frequently used by humanistic geographers as a symbol for universal attachment. Home, like place, can exist on many scales from our individual abodes to the whole earth. Critical human geographers concerned with the way power operated geographically had a very different view of home. To feminist geographers home frequently features as a site of patriarchal authority often associated with extremes of abuse, boredom, and backbreaking labor. To others, home is a place associated with violence against and abuse of children. Home is a place of order where even the children that inhabit it are symptoms of disorder. Consider the following from David Sibley:

> Inside the home and the immediate locality, social and spatial order may be obvious and enduring characteristics of the environment. For those who do not fit, either children whose conceptions of space and time are at variance with those of controlling adults or the homeless, nomadic, or black in a homogeneously white, middle class space, such environments may be inherently exclusionary. (Sibley, 1995: 99)

Here home as place is symptomatic of the ways in which the social production of place as a site of belonging can reinforce social relations of systematically asymmetrical power relations.

It was questions such as these that led critical cultural geographers to explore how places and their associated meanings have been implicated in processes of exclusion. The mapping of particular meanings, practices, and identities on to place, they have argued, leads to the construction of normative places where it is possible to be either ‘in place’ or ‘out of place’. Things, practices, and people labeled out of place are said to have transgressed often invisible boundaries that define what is appropriate and what is inappropriate. We all know that we are not supposed to shout in a library or walk naked down a public highway. These unspoken rules exist in the world of commonsense. It is this very commonsense nature of place based norms that make them so powerful an ideological tool. To take one example, the ex Mayor of New York City, Mayor Koch, responded to homeless people in Grand Central Station by asking the police to remove them. When this course of action was thrown out by the Courts he declared the following:

> These homeless people, you can tell who they are. They’re sitting on the floor, occasionally defecating, urinating, talking to themselves….We thought it would be reasonable for the authorities to say, “you can’t stay here unless you’re here for transportation.” Reasonable, rational people would come to that conclusion, right? (Mayor Koch in Cresswell, 1996: 4)

Here, Koch links the place, Grand Central Station, to particular practices and meanings as though such connections were natural and obvious and any other practices and meanings were simply out of place. In so doing he disconnects the station from the surrounding context of New York City in the 1980s, a place marked by high levels of homelessness contrasting with expensive urban development projects and the intense gentrification of areas such as the Lower East Side where people on low incomes could no longer afford to live.

This process of identifying how normative con structions of place exclude 'others' both physically and existentially has been identified across a whole range of identities including class, race, sexuality, gender, and physical (dis)ability. Geographers and others have also revealed how these social constructions of place are constantly contested, transgressed, and resisted by the excluded. Young people gather on street corners or skateboard on street furniture; the homeless finds ways to live in inhospitable places; artists redecorate well known monuments to invert their established meanings; gay, lesbian, and bisexual people hold kiss ins in public space. Whatever kinds of places are constructed they are never truly finished and always open to question and transformation.

Placelessness and Nonplace

In 1976, the humanistic geographer Edward Relph wrote *Place and Placelessness*. In this book he argued that places were becoming placeless. He gives many reasons for this – mass production, an increasingly mobile world, an emphasis on 'disneyfied' and 'museumified' places that were fake copies of more worthy originals. He described these places as 'inauthentic'. They are inauthentic, he argues, because it is impossible to be an existential insider in Disneyworld, or McDonalds or on a mass produced housing estate. It is also impossible, he argues, to make significant attachments to place if we move about too much, never stopping to linger and create roots. Variations of this argument are now quite commonplace and have been variously described as 'McDonaldization' or 'Americanization'. More recently, the French anthro pologist Marc Auge has used the term 'nonplace' to refer to sites such as motorway service stations, airports, and places of transit which never actually go anywhere but endlessly refer to other places indirectly. Auge does not believe such 'nonplaces' are inauthentic but simply a condition of the way we lead our lives now. Auge con siders conventional conceptions of place as bounded, rooted, organic, and associated with modes of dwelling that are embedded in history. As an anthropologist Auge notes how his discipline has traditionally thought of its purpose as a kind of deep description, excavating all the particularities of a culture *in situ*. What, he asks, could such a discipline make of a mobile world in which all these characteristics (bounded, rooted, etc.) are absent or weak? A world of shopping malls, motorways, airports, and service stations, he suggests, demands a different kind of anthropology – an anthropology of nonplace. The world of 'supermodernity', he argues is marked by three new characteristics. First, the world is marked by a speeding up of communications and information flows that leads to a bombardment of images of spaces and times other than the one a person may be immediately located in. Second, a shrinking of the planet due to time–space compression and third, increased individualism as people, exposed to so much, withdraw into themselves and fail to make sustained social relationships. All of these characteristics of supermodernity, he suggests, can be found in 'nonplace'.

Nonplace is very different from a traditional human istic conception of place as a location full of meaning to which people become profoundly attached. Nonplaces are marked by a lack of attachment, by constant circu lation, communication, and consumption that act against developing social bonds and bonds between people and the world. These nonplaces are marked by a plethora of texts, screens, and signs which facilitate mediated relationships between people and places rather than direct ones.

One of the alleged prime causes of placelessness and nonplace is increased mobility. This mobility has led some to go as far as to declare the 'end of geography' as though mobility itself was not geographical. While it is clearly the case that we live (in the West at least) in a way that is increasingly mobile and uprooted it is surely not the case that place is no longer important. Even Auge remarks that nonplace in pure form does not exist. Rather it is at the end of a continuum with the traditional notion of place at the other end. Real world locations contain varying degrees of both. Even airports have their inhabitants, people that work there, the homeless or frequent fliers who see the same people, at the same time, every workday. The link between the characteristics of nonplace and what Auge calls 'super modernity" is also tenuous as many of these characteristics have been commented upon for a very long time in reference to the effects of phenomena such as the stagecoach, the railroad, and the telegraph among others.

Place, Process, and Mobility

The concepts of nonplace and placelessness raise the issue of the relations between place and mobilities. Traditionally, place has been quite a static concept. The focus on borders, rootedness, and singular identities has mitigated against notions of dynamism and process. Writers such as Tuan and Relph have frequently argued that too much mobility mitigates against senses of place. Nevertheless, since the advent of humanistic geography there have been attempts to think through the ways place is in process and how process makes place. Indeed, in Plato's and Aristotle's musings on chora and topos place was very much about the process of becoming – how things come to be. In the 1970s and 1980s, David Seamon, in a series of books and papers, developed the ideas of the phenomenological philosopher Maurice Merleau Ponty to describe place as the product of everyday habitual mobilities. Merleau Ponty is best known for his account of bodily intentionality. Perhaps the most fundamental insight of phenomenology is the notion proposed by the psychologist Franz Brentano and developed by phil osopher Edmund Husserl that consciousness is always consciousness of something. Consciousness, in other words, includes within it some object that we are con scious of. There can be no judgment, for instance, without something being judged, no love without some thing being loved, no desire without something being desired (to paraphrase Brenato). Intentionality, therefore, describes 'aboutness'. This 'aboutness' refers to a relation between consciousness and the world and this relation between consciousness and the world is key to both humanistic geography and to the notion of place. Place, and particularly sense of place, differs from location (for instance) because of its insistence on the centrality of human consciousness and experience. To return to Merleau Ponty, he developed intentionality by moving it away from the idea of a knowing subject where con sciousness was purely a mental process. His innovation was to describe this process of aboutness as bodily and habitual – below the radar of consciousness. To be in the world to Merleau Ponty was not primarily about the mind but the body. He described the 'body subject' as a body that has a kind of bodily intentionality toward the world. He described everyday actions, such as typing, as remarkable bodily knowledges. In fact the more we think about such things the less able we become. The body in the world is, itself, knowing. The body does not just receive data but experiences the world through a certain openness to the outside – a sense of being in touch with the world. In this view, the movements of the human body constitute the most basic form of intentionality. So what does this say about place?

David Seamon developed Merleau Ponty's ideas to provide a humanistic account of place which was full of movement. He used the term 'body ballet' to refer to how the body moves habitually as it is performing some task such as driving, typing, or cooking. When these body ballets are strung together through a day they produce what Seamon called a time–space routine. Individuals, he argued, followed 'time–space routines' throughout the day. Often these routines are habitual. We drive to work, walk to the train, or go shopping on an almost daily basis. They are not often the product of great thought. This is what Seamon (following Merleau Ponty) means by 'habitual'. When these individual 'time–space routines' coalesce they form a 'place ballet'. In the town center or city square there are hundreds of people conducting their individual 'time–space routines' but they do so collect ively in such a way that recognizable and regular patterns of practice emerge. Examples would be rush hour at a busy interchange, the 'school run' in the middle of the afternoon, or outside a workplace at the day's end. Seamon's argument is that places exhibit a kind of unchoreographed yet ordered practice that makes the place just as much as the place's more static and bounded qualities do. Indeed the meaning of a place may arise out of the constant reiteration of practices that are simultaneously individual and social. Places in this sense are intensely embodied and dramatic. While such a view shares many things with other humanistic geographers it is also different in its emphasis on bodily subjectivity and constant process.

Geographers inspired by structuration theory were also keen to develop a more process oriented view of place. Allen Pred, for instance, argued that human geographers had tended to see places as objects that were essentially static.

> Until recently, places and regions have usually been treated in ways that emphasize certain measurable or visible attributes of an area during some arbitrary period of observation. Thus, whether presented as elements of a spatial distribution, as unique assemblages of physical facts and human artifacts, or as localized spatial forms, places and regions have been portrayed as little more than frozen scenes for human activity. Even the "new humanist" geographers, who see place as an object for a subject, as a center of individually felt values and meanings, or as a locality of emotional attachment and felt significance, in essence conceive of place as an inert, experienced scene. (Pred, 1984: 279)

What Pred proposes is a view of place as process where the activities of people and institutions produce and are produced by social structures that are saturated with power. Place is produced through action and action is produced in place through a constant reiterative process. In a way that mirrors Seamon's discussion of time–space routines and body ballets Pred introduces the notion of 'paths' from time geography. Paths describe the way people and objects move though space and time over a given period of time. These paths produce human and object biographies that coalesce to produce places. "Place is therefore a process whereby the reproduction of social and cultural forms, the formation of biographies, and the transformation of nature ceaselessly become one another at the same time that time space specific activities and power relations ceaselessly become one another" (Pred, 1984: 282). Where Pred differs from

Seamon is in his attentiveness to the operations of power. Pred's paths exist within the context of social structures that provide the context for objects and humans in place. At the same time these social structures are produced through and by these paths. Pred refers to 'institutional projects' that have the power to direct and construct the paths of individuals. Some of these institutional projects have more of an impact than others on "the daily paths and life paths of specific people and, therefore, upon the details of individual consciousness development and socialization" (Pred, 1984: 282). These dominant institutional projects override both other such projects and individual paths that fall outside of the remit of any institutional project. Pred, then, thinks of place as produced through process but in a way which is at least partly structured by social relations that are systematically asymmetrical. Process and mobility in place is not just a matter of habit but a matter of power. So while Seamon saw place as produced through an unchoreographed coming together of mobile intentional body subjects, Pred is keen to show how these body subjects are differentiated by social divisions that allow some mobilities and force impediment on others. Men may be able to move through place in ways that women cannot for instance. Black people are often stopped while driving through cities in the United States on suspicion of having committed some crime. This has been called 'driving while black'. People who appear to be of middle eastern origin have to think twice before using public transport or taking a flight as they are frequently stopped and treated with suspicion. Young people in 'hoodies' are similarly frowned upon when gathered at a street corner.

A Progressive Sense of Place

In both Seamon and Pred's work the key to understanding place is through its relation to mobilities – to the dynamism and flow of objects and people through them. But what about the relationships between particular places and the mobilities of the wider world? Thinking about the world in terms of deeply rooted, fixed places with clear boundaries and stable associated identities can be characterized as a sedentarist metaphysics. Once the world is thought of as a world of such places then other ways of thinking tend to follow. People who lead mobile lives, either through choice or through compulsion are seen as necessarily threatening to this view of place. Thus, the homeless, refugees, gypsy travelers, traveling salesmen, and others who are perceived as mobile are labeled as a threat to place and the moral values associated with it. The more clearly the world is ordered into discrete places the more people and things that exist outside of these places are likely to be labeled as disorder – as out of place. The production of order is simultaneously the production of disorder and deviance.

To some geographers the humanistic conception of place, which has been the predominant understanding of place since the 1970s, is simply too fixed, too bounded, and too rooted in the distant past. As a consequence of these notions of fixity, boundedness, and rootedness, place too often becomes the locus of exclusionary practices. People connect a place with a particular identity and proceed to defend it against the threatening outside with its different identities.

Rather than thinking of mobility as a threat to place Doreen Massey has argued that places are actively constituted by mobility – particularly the movement of people but also commodities and ideas. Places to Massey are not clearly bounded, rooted in place, or connected to single homogeneous identities but produced through connections to the rest of the world and therefore are more about routes than roots. They are sites of heterogeneous, not homogeneous, identities. This conception of place, she variously calls, a 'progressive sense of place', a 'global sense of place', and an 'extrovert sense of place'. Here there is no longer a clear inside and outside and therefore it is much harder to make judgments about insiders and outsiders. To illustrate this notion she describes Kilburn High Road in north London. On this street she encounters Irish pubs, Indian sari shops, and a Moslem newspaper seller. Planes from Heathrow fly overhead. The street is thoroughly constituted by its connections to the wider world. This progressive sense of place, which describes place constituted through mobility, can be contrasted with ideas of nonplace and placelessness which pitch mobility against place – as a threat to place. It also calls into question ideas of place which pitch insiders against outsiders. It is hard to exclude the outside when there is no clear outside but, instead, a set of constitutive connections. So while ideas of place as mobile and process oriented have focused on the mobilities internal to place and the way in which they constitute a sense of place, Massey expands this to consider the connections between place and wider world and this opens up place to a more global sensibility.

Conclusion

While place is clearly central to human geography as well as to everyday life, it is equally clearly a changing and contested concept. Places range in scale from the corner of a room to the whole planet. They are, in the broadest sense, locations imbued with meaning that are sites of everyday practice. Beyond that simple definition there is considerable debate about the nature of place. While some think of place as an essential and foundational fact of human existence others choose to focus on the social processes that produce places in particular ways that serve the interests of some over others. The relationship between place and mobility is also marked by disagreements between those who see mobility and process as antagonistic to place and those who think of place as created by both internal and external mobilities and processes. At the extreme there are those who argue that such are the processes of mobility and communication in the modern world that place is becoming insignificant in a world of placelessness and nonplace. It seems likely that these debates will continue and veer off in new directions.

See also: Humanism/Humanistic Geography; Landscape; Migration; Mobility; Space I; Structuration Theory; Structurationist Geography; Time Geography.

Further Reading

Adams, P., Hoelscher, S. and Till, K. (eds.) (2001). *Textures of Place: Exploring Humanist Geographies*. Minneapolis, MN: University of Minnesota Press.

Auge, M. (1995). *Non Place: Introduction to an Anthropology of Supermodernity*. London: Verso.

Buttimer, A. and Seamon, D. (eds.) (1980). *The Human Experience of Space and Place*. New York: St. Martin's Press.

Casey, E. (1997). *The Fate of Place: A Philosophical History*. Berkeley, CA: University of California Press.

Cresswell, T. (1996). *In Place/Out of Place: Geography, Ideology and Transgression*. Minneapolis, MN: University of Minnesota Press.

Cresswell, T. (2004). *Place: A Short Introduction*. Oxford: Blackwell.

Harvey, D. (1993). From space to place and back again. In Bird, J., Curtis, B., Putnam, T., Robertson, G. & Tickner, L. (eds.) *Mapping the Futures: Local Cultures, Global Change*, pp 3 29. London: Routledge.

Heidegger, M. (1993). Building, dwelling, thinking. In Ferrell Krell, D. (ed.) *Martin Heidegger: Basic Writings*, pp 361 362. London: Routledge.

Malkki, L. (1992). National geographic: The rooting of peoples and the territorialization of national identity among scholars and refugees. *Cultural Anthropology* 7, 24 44.

Massey, D. (1993). Power geometry and a progressive sense of place. In Bird, J., Curtis, B., Putnam, T., Robertson, G. & Tickner, L. (eds.) *Mapping the Futures: Local Cultures, Global Change*, pp 59 69. London: Routledge.

Merriman, P. (2004). Driving places: Marc Auge, non places and the geographies of England's M1 motorway. *Theory, Culture and Society* 21, 145 167.

Pred, A. (1984). Place as historically contingent process: Structuration and the time geography of becoming places. *Annals of the Association of American Geographers* 74(2), 279 297.

Relph, E. (1976). *Place and Placelessness*. London: Pion.

Sack, R. (1997). *Homo Geographicus*. Baltimore, MD: Johns Hopkins University Press.

Seamon, D. (1980). Body subject, time space routines, and place ballets. In Buttimer, A. & Seamon, D. (eds.) *The Human Experience of Space and Place*, pp 148 165. New York: St. Martin's Press.

Sibley, D. (1995). *Geographies of Exclusion: Society and Difference in the West*. London: Routledge.

Tuan, Y. F. (1974). Space and place: Humanistic perspective. *Progress in Geography* 6, 211 252.

Tuan, Y. F. (1977). *Space and Place: The Perspective of Experience*. Minneapolis, MN: University of Minnesota Press.

Relevant Websites

http://pegasus.cc.ucf.edu/~janzb/place/
Research on Place and Space Bruce B. Janz.

Place Names

V. R. Savage, National University of Singapore, Singapore

Glossary

Toponymics or Toponyms The Greek words *topos* or place and *onomia* or name refer to the study of place names. The study of place names (i.e., countries, regions, provinces, towns, cities, rivers, mountains, etc.) is a well-developed field of enquiry not only among geographers but also among historians, cartographers, anthropologists, sociologists, folklorists, linguists, and the informed public. The study of toponyms involves essentially queries into three areas: (1) the earliest date of the place name; (2) the meaning of the place name; and (3) how the place name was derived or who bestowed the place name.

Toponymics: Unfolding the Geographies of Landscape Symbolism, Territorialism, and Societal Histories

Introduction: Discovering Place Names

The study of place names or toponyms (*topos*: place and *onomia*: name) has been a long research tradition among historical and cultural geographers. However, this is by no means the territorial academic preserve of geographers alone. Archaeologists, anthropologists, historians, folklorists, cartographers, linguists, sociologists, planners, poets, as well as the informed public have all been interested in place names research. The fascination with place names is because it provides invaluable keys and insights into landscape histories, settlement origins and patterns, physical geographies of places, sequent occupance, ethnic and political changes, nationalistic sentiments, human activities, and cultural diffusion processes.

Given that place names are embedded in and products of languages both past and present their investigations reflect etymological–linguistic exercises, onomastic (relating to names) inquiries, and word meaning investigations. Reflecting the diversity of dialects and languages spoken in the past, older place names thus become difficult to trace in terms of their origins. Putting together the comprehensive listing of Aboriginal place names in Australia was no easy task given that 'old Australia' had at least 500 languages or distinct dialects. Margaret Gelling notes on the other hand how painfully difficult it is to trace English place names because it reflects pre Celtic, Pictish, Greek, Roman, Celtic, Latin, Germanic, Welsh, Old English, Norman French, Cornish, Gaelic, and Gallo Brittonic languages and dialects. Accepting the multiplicity of meanings of past words, she cautions nonphilologists in jumping to hasty conclusions around sets of similar looking words which may have totally different meanings. Given that linguistic theory and sociolinguistics have been neglected fields in cultural geography the onomastic interpretation of place names remains a fertile field in the discipline.

Geographers note specifically that place names provide evidence of environmental settlement and social conditions at the time the name was coined and hence the utility of place names is limited to periods in which names were being created. Place names are products of only two processes: some names 'evolved' from the 'folk mind' through an 'unpersonalized linguistic development' and become crystallized over time. Other names were 'bestowed' in which an individual consciously declared a place name which was documented and preserved. Whether evolved or bestowed, place names become important signs, signals, and symbolisms of evolving cultural landscapes that may take on multiple or different meanings by communities within a period or at various points of time. They also provide a sense of cultural and social continuity between the past and present. Place names remain for many residents the mental maps of meaning involving codified knowledge, historical legacies, and symbolic representations. Given their long history over cultural changes, place names can at times be ambiguous. What is fascinating about the study of place names is why certain names persist historically (over centuries), while other names might be erased, modified, and changed over time. To some extent languages can act to preserve certain place names. For example, folk etymology as underscored by the process of assimilation and phonetic transfer has helped to preserve Indian, English, French, Spanish, and Portuguese names in America. Many Spanish names are easily absorbed into English by phonetic transfer and remain today as place names in America: San Francisco, San Diego, Los Angeles, Sierra Nevada, Colorado River, and El Cerrito.

On the other hand one might seek to find out the variation and density of place names spatially. Several factors explain the density of place names. These include the physical environment and topography: flat areas with no significant physical landmarks would lend themselves to less place names compared to a variegated topography. Population density bears a somewhat proportional relationship to the density of population because of the proliferation of habitation names. Other factors that affect the density of place names are length of human settlement in an area (new names replacing or added on to old names) and the cultural pattern of people inhabiting an area. Stewart argues that Americans and their relatively unpopulated landscapes saw things in a 'larger way' and hence names covered large areas of the landscape (representing miles of an area) compared to Europeans who filled their densely populated landscapes with a density of place names in Europe.

Indigenous Landscape Signifiers

The value of place names is that they reflect specific localities, landmarks, and identities in otherwise amorphous spaces and environments. In nonurban landscapes, the imprint of indigenous place names by inhabitants provides significant symbolic signifiers (sacred sites, legendary places, tabooed locations) in a large landscape. With its wide expanse of open space and pristine environments, Australian Aboriginal place names reflect specific aspects of physical geography (landscape morphology, watering places, flora, fauna, insects, etc.), folktales and mythologies, and ancestor experiences. Australian Aborigines embed their place names in their 'Dreaming' which connects important sites and pathways for each community's survival in their desert environment. Unlike most societies where place names are part of public and societal knowledge, among the Australian Aborigines the knowledge of their environment and its identity remains a multilevel secret within a tribal community and kept away from non Aboriginals and other communities. In New Zealand, Maori place names generally underscore two themes: (1) they stem from myths and early history based on the names from the homeland of Hawaiki that have been transplanted to Aotearoa (New Zealand) and (2) they reflect very earthy and natural landscape themes (water, rivers, soils, vegetation, fauna, and landforms). Historic English place names on the other hand define overwhelmingly topographical, geomorphological, and biogeographical landmarks (rivers, marshes, hills, forests, springs, valleys, stones, trees) and cultural landscape features (villages, forts, towns, fortified dwelling sites, agricultural fields, burial sites, farms, churches). Place names in China mirror to a large extent cardinal directions (Peking – northern capital; Kiangsi – river west; Shantung – east of the mountains) and physical geographical features (Chinkiang – river market; Hupeh – north of the lake; Szechuan – four streams).

In Australia and Southeast Asia, indigenous cosmologies have bearings on specific geographic landscapes and environments, by signifying places of identity, social experiences, and cultural memory. James Fox notes that cosmologies are unveiled in the form of 'topogeny' or "the recitation of an ordered sequence of place names." Such topogenies may reveal the journey of an ancestor or the migration route of a group. In some cases, Fox notes topogenies may refer to movement of objects as in the case of the Rotinese (Indonesian) story of two rocks that settled in the north coast of their island. In the case of Australian Aborigines in northern Australia, indigenous sea cosmologies reflect journeys of ancestral 'Dreaming beings' across the seas. In the process, the journeys, sites, islands and places of these Dreamings serve as a way of legitimizing claims to lands, resource ownership, and perceptions of territoriality.

Cartographic Representations

One important repository of place names and their changing references to villages, towns, physical landscapes, areas, kingdoms, countries, and regions lies in maps. Maps provide an almost encyclopedic reference point of place names over time from various cultural vantage points. But by far the Western mapping tradition since the fifteenth century had the greatest impact in documenting place names across an expanding *oikoumenē* or in Jane Jacobs' view the "over determined signifier of the spatiality of the imperial imagination." In Southeast Asia, two historical map books by Thomas Suárez and David Parry provide a monumental testimony to the frenzied Western interest in mapping this region complete with all its indigenous and Western baptized place names.

For historical geographers and cartographers, one of the great exercises has been to try and give place name meanings and current spatial identification to the numerous names of ports, towns, rivers, mountains, islands, and seas of the Old World (Europe and Asia) in Claudius Ptolemy's (*c.* AD 90–168) *Guide to Geography*. Given the supposed reference to a treasure trove in Southeast Asia, many adventurers and scholars have tried to identify Ptolemy's *Aurea Khersonese* (Golden Peninsular). Paul Wheatley's masterly work, *The Golden Khersonese* is one example of how painstakingly a historical geographer has tried to piece together all the place names in Ptolemy's *Geography* on the Southeast Asian region by looking at indigenous, Arabic, Indian, and Chinese names and the natural produce and trading activities of the past. Yet, the debate on Ptolemy's names and places never ends. Even Wheatley's thesis, that the Golden Peninsular refers to what is currently West Malaysia (or Peninsula Malaya), is challenged by W. J. van der Meulen in his equally meticulous reconstruction of vernacular names, past trading activities, and physical geography of the region.

In the case of the celebrated French historian, Fernand Braudel in his monumental work, *The Identity of France*, the study of place names is not only a guide to the history of places and regions but can take scholars back even further in time to the prehistory of areas. The use of dialects and their many alterations in place names in France he notes provide "many chronological land marks," that can send "beams of light into the darkest corners of our past." In a way, place names make history visible because it perpetually elicits historical and cul tural curiosity. Braudel saw the study of place names of *pays*, villages, *bourg*, *départements*, and territorial units as revealing France's past diversities, a mosaic of landscapes changed and reshaped that informs us of past forces that shaped a "series of different Frances."

Place names have been important in registering the changing political geographies of areas and regions. What Jane Jacobs calls the "promiscuous geographies of dwelling in places" (Self and Others, past and present), cultural landscapes and their place names often recall spatial struggles, the binary struggles of colonial powers and local populations, or what Edward Said refers to as the "geography which struggles." In Latin America, Robinson argues that when "regimes fall, empires col lapse, or one local elite replaces another, very often a name changing process is initiated in particular places." In Southeast Asia, many names of places reflect the places of defeats and victories of past battles: *Siem Reap* or 'the defeat of Siam' (in Cambodia) is the place where King Ang Chan of Cambodia defeated the Thai king, Mahachakrapat. On the other hand the adoption of foreign names as place names by local polities can help elevate and legitimize local rulers to a wider civilizing power. Sheldon Pollock argues that the rulers of Angkor by the "wholesale appropriation of its toponymy" became part of the "circulatory space" of the Sanskrit cosmopolis from South Asia. In Taiwan, Hsieh analyzed place names to construct the sequent occupance of colonizing immi grant groups. More recently (May 2007) it is rival political parties that are replacing old place names and creating politically biased place names in Taipei: the then ruling party, the Democratic Progressive Party (DPP) stripped the name of the once revered Kuomintang (KMT) leader, Chiang Kai shek from a Taipei monu ment and renamed park 'Taiwan Democracy Memorial Park'. In retaliation, the KMT elected mayor of Taipei renamed the road outside DPP elected President, Chen Shui bian's office, the 'Anti Corruption Democracy Plaza' as a jab at the corruption tainted President.

Colonial Impacts

At a global scale, one can see how colonialism created a new landscape of place names Christianized by European names and Christian relationships: America (after Amerigo Vespucci), Veracruz (*Villa Rica de la Vera Cruz* or Rich Town of the True Cross), San Salvador (Holy Savior), La Paz, Bolivia (after *Nuestra Senora de la Paz* or Our Lady of the Peace), the Philippines (after Prince Philip of Spain), Easter Island (discovered on Easter Sunday), Christmas Island (discovered in 1643 but named Christmas Island by Captain Cook on 24 December 1777), and Rhodesia (after Cecil Rhodes). In Africa, the Portuguese trading in the coastal areas of West Africa referred to each area based on its exports hence the names: Grain Coast, Gold Coast, Slave Coast, Ivory Coast, and Diamond Coast.

In a broader sense, the expanding European *oikoumenē* was signified by Eurocentric place names. Many places around the world were seen as 'new' Western discoveries hence the 'new' added to place names. America was the *Novus Mundus* (New World) and other 'new' places fol lowed in America and around the world: Newfoundland, Nova Scotia (New Scotland), New England, New Guinea, New Zealand, New South Wales. For Indo nesians, the country's name Indonesia (meaning Indo: Indian, *nesos*: islands) came from Westerners in the region, but Indonesians refer to their country as *Tanah Air* or 'Earth and water'. Similarly, the Chinese refer to their country or kingdom variously as T'ienhsia (under heaven), Chung Yuan (central origin or original center) and more recently as Chung Kuo (central country of middle kingdom) while the Europeans referred to the empire as China, derived from the Chin dynastic term, Ch'in Kuo (or Chin kingdom).

In colonial cities, street names were meant to char acterize European towns reflecting their Portuguese, Spanish, British, Dutch, Belgian, or French metropolitan masters. Colonies were not viewed merely as abstracted or conceptual spaces but existential social spaces and Europeanized place names reflect best Doreen Massey's notion of the "ever shifting social geometry of power and significance." Each of the colonies carried with them succeeding place names reflecting their colonial masters. In India, it was Portuguese, French, and English; in Sri Lanka and Malacca it was Portuguese, Dutch, and English; and in the Philippines it was Spanish followed by American. In Australia, Jane Jacobs argues that the dominant colonial power in the name of the imperial project evaluated space, drove territorializations, created homely environments, established ordered grids of occupation, mapped areas of their occupation, and marginalized the earlier geographies of Aborigines. In colonial Singapore, Brenda Yeoh observes that competing vernacular representations (Malay, Chinese, Tamil) of street names served as competing representations of the colonial enterprise of governance and surveillance. The Asian populations' use of their own place names signified a diversity of sociocultural influences, competing repre sentations of the landscape, as well as a syncretic character of the city. Ironically, while colonial govern ments were creating new place names to imprint their authority over their colonized landscapes, many colonial administrator scholars were also equally interested in the meanings of vernacular place names. Indeed without their documentation in published records, much of the older vernacular place names would have been lost in many developing countries.

Postcolonial Changes: National Identities

The postcolonial rise of nationalism in many developing countries has seen a reassertion of indigenous place names. On independence, many of the island states in the Pacific reverted back to their indigenous names: Tuvalu (from Ellice Islands), Vanuatu (from New Hebrides), and Kiribati (from Gilbert Islands). In the continent of Africa, the reversion to indigenous country names was widespread after independence from the colonial masters: Benin (from Dahomey), Burkina Faso (from Upper Volta), Ghana (from Gold Coast), Djibouti (from French Somaliland), and Rhodesia (named after Cecil Rhodes) became Zimbabwe. In Burma, the military junta opted for Myanmar (comes from *myan* or strong and *maa* meaning hardy) as the country's new name, its capital of Rangoon was changed to Yangon, and its main river the Irrawaddy became the Ayeyarwady. Similarly, Ceylon became Sri Lanka (1972), British North Borneo became Sabah (1963) and in India, the roll call of city name changes since independence has not stopped. Every few years, cities named by the British are being renamed: Trivandrum to Thiruvananthapuram (1991), Bombay to Mumbai (1995), Madras to Chennai (1996), Calcutta to Kolkata (2001), and the proposed Bangalore to Bengalooru (2007). Even in the noncolonized country of Siam, the country's rulers changed the Western derived name to reflect its true national spirit and renamed the country, Thailand (1949), the 'land of the free'. These name changes in the contemporary world reflect in part the complex processes of deterritorialization and reterritorialization of the histories of connection and disconnection and what Doreen Massey calls the "progressive sense of place."

Every war of nationalism has also been signified with an endorsed place name to reflect the indigenous politics: East Pakistan to Bangladesh; Saigon after its fall to North Vietnam's communist masters to become Ho Chi Minh City (HCMC). Like the city of Brasilia (new capital of Brazil), in Malaysia a new capital was established under the former Prime Minister, Dr. Mahathir Mohammad and was named Putrajaya (or Putra's success) after the name of Malaysia's first Prime Minister, Tunku Abdul Rahman Putra Al Haj. More recently, the military government in Myanmar also chose to build a new capital called Pyinmana Naypyidaw (Naypyidaw means 'the abode of kings'). The new capital's place name seems to reinforce Myanmar's illustrious kingdoms of the past and the ruling junta's belief in astrological predictions. The above postcolonial naming exercises of national figures like Ho Chi Minh and Tunku Abdul Rahman can be interpreted as indicators of nationalistic fervor or community identity.

The nationalistic unfolding from formerly developed 'imagined communities' has given way to more nation states in the developing and communist worlds. The most recent dramatic examples of new nation states are East Timor, the fragmentation of the former Soviet Union, and Yugoslavia. Yugoslavia ('Land of the South Slavs') has been Balkanized into numerous nations: Serbia, Croatia, Slovenia, Bosnia, Hercegovina, Macedonia, and Montenegro. After an interlude with communism, Russian leaders decided to change the name of Leningrad (after Lenin) to the city's more legendary name of Saint Petersburg and Stalingrad (after Stalin) to Volgograd, a reflection of deep historical nationalism. Singapore's street names reflect the contradictions and swings of policies in nation building in which street names at times underscored nationalist ideologies (e.g., Malayan, dialect and *pinyin* names) to construct the 'nation' and at times names were erased in line with the reshaping of national imaginary. In Great Britain, the Council for Name Studies in Great Britain and Ireland saw the importance of place names as a national endeavor and hence encouraged its own members in March 1965 to compile for the first time a comprehensive and systematic list of place names covering England, Scotland, and Wales. The result was the book by Margaret Gelling, W. F. H. Nicolaisen, and Melville Richards, *The Names of Towns and Cities in Britain.*

At the national level, it seems evident that place names provide communities a sense of social belonging, territoriality, and national ownership. As Claude Lévi Strauss observed "space is a society of named places." A place name gives identity to people as in country names. We refer to citizens by their country or nationality: English, Russians, French, Mexicans, Japanese, Indians, and Australians. Names of nation states carry a sense of deep pride for their citizens. In some ways, country names have become important brand names for food, sports, history, holidays, and other human activities. We think of Thai and French foods, of Brazilian soccer or football, of Italian and French fashion designs, and German and Swiss precision. Place names also carry with them idyllic, Edenic, and relaxing associations: Hawaii, Bali, Tahiti, the Caribbean, and the Maldives.

For the Westerner, the terms, 'Orient', 'the East', and 'Asia' are more than cardinal reference points. The geographical remoteness and esthetic distance of Asia evokes in Westerners exotic ideas, romantic echoes, fantastic overtones, and a byword for the marvelous. Asia to the Persians and Europe to the Greeks became counterpart entities and divisions, a product of names developed in blundering fashion but probably one of the first triumphs of scientific geography. Similarly the term 'Indochina', its mélange of Indian and Chinese elements stirs French imaginary scaffoldings on which Indochina as a fictional place, a *lieu de memoire*, and a vision of an exotic utopia were erected and against which the fantasies of artists, writers, filmmakers, architects, political figures, and others emerged. On the other hand, place names like Casablanca, the French Riviera, and Monaco distill romantic imaginations especially for the Western world, a product no doubt of the movie industry. In France, the names of *pays* are identified with specific types of cheese (Brie, Normandy) and liquors (wines, brandy, champagne: Cognac, Champagne, Burgundy, Medoc). In other cases, place names of pilgrimage cities reveal religious and spiritual connotations: Mecca, Varanasi, Bethlehem, Fatima, and Lourdes.

Urban Cues

For many countries, cities, and places around the world, their names might carry important mythical, magical, and legendary stories. In the Pacific Islands, place names are seen as products of mythologies that the original immigrants carried with them. The process of '*land nama*' or 'land naming' or 'land taking' referred to newly entered lands assimilated by an immigrant people to its imported heritage of myth. Place names also reflect myths developed from specific places. Singapore (after the Sanskrit name *Singhapura* or Lion City) derived its name when an indigenous ruler, Sri Tri Buana in the fourteenth century saw a supposed 'lion' in the island and decided to call its thalassic kingdom, *Singapura* (Lion City) from then onward.

Given that cities are the most culturally imprinted and extensively engineered landscapes of human habitation, it is not surprising that cities have the greatest concentration of place names compared to any other areas of human habitation. Cities are defined by their street and road names and less so by their physical landscape, a testimony to the human and cultural transformation of the urban landscape. As Lewis Mumford reminds us, the city is not only the symbol of human civilization it is the ultimate artifact of human engineering, planning, and culture. However, while odology, the science or study of roads and journeys is pertinent to transportation and urban geographers, street names are equally important to historical and cultural geographers. This underscores the fact that roads are no longer pathways to places, they are places. Hence their place names give identity in an otherwise dense and chaotic urban landscape. In urban areas especially, place and street names often provide the most veritable historical geography of a city.

In the streets of Singapore's old Chinatown, the resident and tourist alike have an invaluable guide to Singapore's fascinating cultural and historic district. The Chinese street names of Chinatown provide the cultural, religious, social, and leisure activities of Chinese who lived there in the past. Indeed, the very reference to Chinatown as a Chinese district is defined and encased by a street's name: 'Bullock Cart Water' or *Kreta Ayer* (Malay), *Gu Chia Chwee* (Fujian or Hokkein), *Ngow Chay Shui* (Cantonese), or *Niu Che Shui* (Mandarin). Ask any taxi driver to take you to 'Chinatown' and he will look at you with a blank stare but state you want to go to *Kreta Ayer* or *Gu Chia Chwee* and he knows you want to go to Chinatown. At a time when over 60% of Singapore's population lived in less than 2% of the island's total area, streets were the working, recreational, social, food, and open spaces of people. The streets of colonial Singapore contribute to the 'imageability' (meaning, expressiveness, stimulus, and sensuous delight) of Singapore in foreign eyes.

Since the early development of cites, societies have embellished their cities with thousands of place and street names. Place names provide urban dwellers the cues, landmarks, and legibility of the city. Place and street names help the millions of city inhabitants move around effortlessly. More important, they give city denizens a sense of place and identity in an otherwise large, impersonal, and amorphous environment. It is through street and place names that people grow attached to their hearths, their affection for their neighborhoods, and their sense of nostalgia for their city. Street and place names provide invaluable social bonding and place relationships for urban citizens living in a fast paced, rapidly changing urban environment. The experiences in and attachments to places reflect Yi Fu Tuan's topophilia defined as the affective bond between people and place or setting. For an older generation of urban citizens, the city's places and street names are their security blanket in mobility, their cultural beacons and their nostalgic reflections of personal experiences. Older people need more landscape cues to navigate their urban spaces. The personal geography of the elderly requires a coping mechanism within their environment to enhance their lives. While the change of street names have become publicly contested issues, unfortunately, as rapidly developing cities expand and are renewed, replanned, and rejuvenated, streets continue to be expunged, open spaces and landmarks are destroyed, and these dramatic landscape changes inevitably make the city less legible to their long term residents.

Street, ethnic areas, and urban district names are colorful landscape anchors that provide insights into many past activities, trades, personalities, ethnic groups, landmarks, buildings, leisure, and cultural activities of cities. Street names *inter alia* unveil localized trades and tradespeople. Indeed one might say that the study of urban street names can lend invaluable insights into the city's history, its evolving cultural landscapes, its social fabric, and its economic relationships. In her analysis of 'Old Singapore', Maya Jayapal used a whole litany of street and place names to bring out Singapore's social, cultural, and economic landscape at the turn of the twentieth century. Street names provide evidence of the diverse peoples, personalities, and renowned persons who have helped shape the development of their cities. Hence, street names contribute to our future because the memories of the streets belong to all the children of a particular city.

Research Issues on Toponymics

Notwithstanding its cultural, social, economic, and political insights and perspectives, the study of place and street names are best pursued by scholars with a histor ical geography inclination and interest. Given that the uncovering of the origins and meaning of place names is a painstaking and tedious process, it is not an exercise that can be rapidly undertaken. It requires patience, a gift for languages, and academic serendipity. Place name research requires work on archival records, old paintings and photographs, antique maps and charts, interviews with pertinent people, on site field work, myths and legends, and published materials (newspapers, journals). Given the long, meticulous, in depth knowledge and historical research of countries, cities, and places, most researchers on place names have little mobility in the spatial outreach of their research. Indeed, most place name researchers tend to spend a lifetime of research within familiar, circumscribed areas they live in (cities, countries).

Despite all the evidence available, the veracity of a place name poses several problems: (1) the earliest recording of a place name may have occurred centuries (with linguistic transformations) before its coining and available records; (2) the interpretation of particular elements from which the place name is derived can be disputed and hence remains uncertain – often there are several possible interpretations for a place name; (3) in dating the original coining of a place name it is important to know when it reflects a feature of landscape or society that is long term (more information and hence reliable) or ephemeral (less informative); and (4) the place name might not be a settlement, field, or feature it now rep resents, given that mobile villages were common in the Dark Ages in Europe and hence the name of a village may be a feature of some other place it was located formerly. For these above problems, the study of place names by varied researchers remains a slow investigative process in which the end goal is the satisfaction of verifying data and finding the truth to the origins and meaning of a place or street name.

See also: Belonging; Neighborhood Change; Postcolonialism/Postcolonial Geographies; Sense of Place.

Further Reading

Braudel, F. (1988). *The Identity of France Volume One: History and Environment*. New York: Harper & Row Publishers.
Braudel, F. (1990). *The Identity of France Volume Two: People and Production*. New York: Harper & Row Publishers.
Campbell, J. (1987). *Primitive Mythology: The Masks of God*. New York: Penguin Books.
Crawfurd, J. (1828). *Journal of an Embassy from the Governor General of India to the Couts of Siam and Cochin, China, etc*. London: Henry Colburn.
Ekwall, E. (1954). *Street names of the City of London*. Oxford: Clarendon Press.
Firmstone, H. W. (1905). Chinese names of streets and places in Singapore and the Malay Peninsula. *Journal of the Straits Branch of the Royal Asiatic Society* 42, 53 208.
Fox, J. J. (1997). Place and landscape in comparative Austronesian perspective. In Fox, J. J. (ed.) *The Poetic Power of Place: Comparative Perspectives in Austronesian Ideas of Locality*, pp 1 22. Canberra: Research School Pacific and Asian Studies, Australian National University.
Gelling, M. (1977). *Place names in the Landscape*. London: Dent.
Gelling, M. (1978). *Signposts to the Past: Place names and the History of England*. London: Dent.
Gelling, M., Nicolaisen, W. F. H. and Richards, M. (eds.) (compl.) (1970). *The Names of Towns and Cities in Britain*. London: Batsford.
Gerini, G. E. (1909). Researchers on Ptolemy's geography of Eastern Asia (further India and Indo Malay Archipelago). *Asiatic Society Monographs No. 1*, London: Royal Asiatic Society, Royal Geographical Society.
Gibson Hill, C. A. (1954). Singapore: Notes on the history of the old strait; 1580 1850. *Journal of the Malayan Branch of the Royal Asiatic Society* 27, 162 214.
Harley, J. B. and Woodward, D. (eds.) (1987 2005). *The History of Cartography*, 3 vols. Chicago and London: The University of Chicago Press.
Harris, C. M. (2006). *What's in a Name?* Harrow: Capital History.
Harrison, T. (1949). Gold and Indian influences in West Borneo. *Journal of the Malayan Branch of the Royal Asiatic Society* 22, 33 104.
Haughton, H. T. (1889). Notes on the names of places in the island of Singapore and its vicinity. *Journal of the Straits Branch of the Royal Asiatic Society* 20, 75 82.
Haughton, H. T. (1982). Native names of streets in Singapore. In Sheppard, M. (ed.) *Singapore 150 Years*, pp 208 219. Singapore: Times Books International (first published in 1891).
Hsieh, C. M. (1980). Sequent occupance and place names. In Knapp, R. G. (ed.) *China's Island Frontier: Studies in the Historical Geography of Taiwan*, pp 107 114. Honolulu: University Press of Hawaii.
Jackson, J. B. (1997). *Landscape in Sight: Looking at America*. New Haven: Yale University Press.
Jacobs, J. M. (1996). *Edge of Empire: Postcolonialism and the City*. London: Routledge.
Jayapal, M. (1992). *Old Singapore*. New York: Oxford University Press.
Johnston, R. J., Gregory, D., Pratt, G. and Watts, M. (2000). *The Dictionary of Human Geography*. Oxford: Blackwell Publishers Ltd.
Kalina, R. (2007). The dream of aboriginal art. *Art in America,* April, 96 103.
Khoo, S. N. (1994). *Streets of Georgetown Penang: An Illustrated Guide to Penang's City Streets and Historic Attractions* (2nd edn.). Penang: Janus Pring and Resources.

Kong, L. and Yeoh, B. S. A. (2003). *The Politics of Landscapes in Singapore: Construction of Nation*. New York: Syracuse University Press.

Lévi Strauss, C. (1996). *The Savage Mind*. Chicago: University of Chicago Press.

Linehan, W. (1951). The indentification of some of Ptolemy's place names in the Golden Khersonese'. *Journal of the Malayan Branch of the Royal Asiatic Society* 24, 86 98.

Lynch, K. (1960). *The Image of the City*. Cambridge: The MIT Press.

Massey, D. (1994). *Space, Place and Gender*. Cambridge: Polity.

McCrindle, J. W. (1885). *Ancient India as Described by Ptolemy*. London.

Mumford, L. (1961). *The City in History*. New York: Harcourt Brace Jovanovich.

Norindr, P. (1996). *Phantasmatic Indochina*. Durham: Duke University Press.

Pannell, S. (2007). Of gods and monsters: Indigenous sea cosmologies, promiscuous geographies and the depths of local sovereignty. In Boomgaard, P. (ed.) *A World of Water: Ram, Rivers and Seas in Southeast Asia*, pp 71 102. Singapore: University Singapore Press.

Parry, D. E. (2005). *The Cartography of the East Indian Islands: Insulae Indiae Orientalis*. London: Countrywide Editions.

Pollock, S. (2006). *The Language of the Gods in the World of Men: Sanskrit, Culture, and Power in Premodern India*. Berkeley: University of California Press.

Raja Singam, S. D. (1939). *Malayan Street Names: What They Mean and Whom They Commemorate*. Ipoh: Mercantile Press.

Reed, A. W. (1988). *Aboriginal Place Names* (compl.). Artarmon, NSW: Reed Books Pte Ltd.

Reed, A. W. (1983). *A Dictionary of Maori Place Names*. Wellington: A H & A W Reed Ltd.

Robinson, D. (1989). The language and significance of place in Latin America. In Agnew, J. A. & Duncan, J. S. (eds.) *The Power of Place: Bringing Teacher Geographical and Sociological Imagination*, pp 157 184. Boston: Unwin Hyman.

Rogers, A. (1995). Cinco de Mayo and the 15th January: Contrasting situations in a mixed ethnic neighbourhoods. In Rogers, A. & Vertovec, S. (eds.) *The Urban Context: Ethnicity, Social Networks and Situational Analysis*, pp 117 140. London: Berg.

Said, E. (1993). *Culture and Imperialism*. London: Chatto and Windus.

Savage, V. R. (1984). *Western Impressions of Nature and Landscape in Southeast Asia*. Singapore: Singapore University Press.

Savage, V. R. (1992). Street culture in colonial Singapore. In Chua, B. H. & Edwards, N. (eds.) *Public Space: Design, Use and Management*, pp 11 23. Singapore: Singapore University Press.

Savage, V. R. and Yeoh, B. S. A. (2004). *Toponymics: A Study of Singapore Street Names* (2nd edn.). Singapore: Eastern Universities Press by Marshall Cavendish.

Schama, S. (1995). *Landscape and Memory*. London: Harper Collins.

Spencer, J. E. (1941). Chinese place names and the appreciation of geographic realities. *Geographical Review* 31, 71 94.

Steadman, J. M. (1969). *The Myth of Asia*. New York, New York: Simon and Schuster.

Stewart, G. R. (1975). *Names on the Globe*. New York: Oxford University Press.

Stump, R. W. (1975). Toponymics commemoration of national figures: The case of Kennedy and King. *Names* 36, 203 216.

Suárez, T. (1999). *Early Mapping of Southeast Asia*. Hong Kong: Periplus.

Tamisari, F. (1998). Body, vision and movement: In the footprints of the ancestors. *Oceanie* 68, 249 271.

Tan, S. (1990). *Streets of Old Chinatown Singapore*. Singapore: Page Media Publication.

Teo, P., Yeoh, B. S. A., Ooi, G. L. and Lai, K. P. Y. (2004). *Changing Landscapes of Singapore*. Singapore: McGraw Hill Education (Asia).

Tuan, Y. F. (1974). *Topophilia: A Study of Environmental Perception, Attitudes, and Values*. New Jersey: Prentice Hall.

van der Meulen, W. J. (1974). Suvarnadvipa and the Chryse Chersonesos. *Indonesia* 18, 1 40.

van der Meulen, W. J. (1975). Ptolemy's geography of mainland Southeast Asia and Borneo. *Indonesia* 19, 1 32.

Wheatley, P. (1961). *The Golden Khersonese*. Kuala Lumpur: University of Malaya Press.

Weiner, J. (1991). *The Empty Place: Poetry, Space and Being Among the Foi of Papur New Guinea*. Bloomington: Indiana University Press.

Yeoh, B. S. A. (1992). Street names in colonial Singapore. *The Geographical Review* 82, 313 322.

Yeoh, B. S. A. (1996a). *Contesting Space: Power Relations and the Urban Built Environment in Colonial Singapore*. Kuala Lumpur: Oxford University Press.

Yeoh, B. S. A. (1996b). Street naming and nation building: Toponymics inscriptions on nationhood in Singapore. *Area* 28, 298 307.

Yeoh, B. S. A. and Kong, L. (eds.) (1995). *Portraits of Places: History, Community and Identity in Singapore*. Singapore: Times Editions.

Zelinsky, W. (1983). Nationalism in the American place name cover. *Names* 31, 1 28.

Place, Politics of

L. A. Staeheli, University of Edinburgh, Edinburgh, UK
D. Mitchell, Syracuse University, Syracuse, NY, USA

Glossary

Demos The people governed by a democratic state; refers historically to Greek city-states, but is now applied more broadly.

Dualism A conceptual division and usually an opposition.

Externalities Spillovers from a land use or activity that affects other properties or other properties. In economic terms, externalities are not generally reflected in the price of the property that creates the externality, even though it affects surrounding property values – either positively or negatively.

NIMBY Stands for 'not in my back yard'. Usually associated with political struggles to keep facilities and noxious land uses out of neighborhoods.

Political Opportunity Structure A mixture of the institutions, ideas, political agents, and networks that condition the kinds of political activities possible in a place.

Politics of Turf Often associated with NIMBYism, these politics are often concerned to preserve or to protect a local area, neighborhood, or even a particular parcel of land. It implies a defensive strategy to forestall change.

Rational Choice Theory A normative theory of individual decision-making. It posits that decisions will be made on the basis of an evaluation of self-interest or of what will provide the greatest benefit with least cost to the individual.

Standpoint Epistemology A theory of knowledge that holds that knowledge is positioned, in the sense that what one knows of the world depends on the social positioning of the knower. Usually associated with an argument that members of marginalized groups – e.g., women, racial minorities, subalterns – have a more complete understanding of the world because they must learn how to negotiate the power relations that members of the dominant group can ignore.

Subaltern A marginalized group that views the dominant order 'from below'.

Transgression Most simply, involves crossing boundaries that, from a normative perspective, should not be crossed. It involves a violation of rules, norms, and/or spaces.

Voluntary Service Organization Often associated with social and human service delivery, these organizations are part of the nonprofit, or nongovernmental, sector. The term 'voluntary' is a misnomer in most contemporary societies. It reflects a tradition and time in which many of these organizations were organized on the basis of volunteer labor and voluntary contributions. Many organizations now, however, are highly professionalized and receive significant government and private sector funding.

Introduction

The politics of place are contested. They are contested, in an empirical sense, as conflicts over ownership, the activities in a place, and the meanings or significance of place are waged through public debates, policy, law, and acts of transgression. These conflicts are important not simply because they make and remake spaces, but also because they are implicated in governing: in making a public governed and in constituting the demos. The politics of place are also contested in terms that are more theoretical, if no less political. Here, the debates are about whether the politics of place and space are regressive or progressive, and therefore whether the public that is formed through the politics of place is inclusive and democratic or whether it is riven with power relations that divide, marginalize, and oppress. Some critical social theorists have proclaimed space to be static, fixed, and therefore better suited as a tool for domination and control than for progressive, radical, or democratic change. Political struggles that are locally waged or that focus on localities have been tagged by these theorists as parochial, limited, and reactionary because they do not contest that larger spatial context. Against this perspective are a variety of theorists – including liberal, communitarian, and radical democratic theorists – who point to the progressive potential for a politics of place, as long as place and locality are understood as standing *in relation* to other places and as formed out of the operations of broader structural pro cesses and forces. Rather than being static, fixed, and parochial, they argue the politics of place link the struggles of the everyday and of the everyplace to the broad processes that shape our world, thereby providing an entry by which struggles from the margins can in fluence power relationships at the center, forming the dominant order. From this perspective, the politics of place are about affecting social and political change broadly understood. Theorists from both sides of this debate, however, understand the politics of place as being manifested in a host of key issues considered within political geography, including nationalism, territoriality, transgression, resistance, place making, and processes of inclusion and exclusion.

In the preceding paragraph, the progressive and reactionary potential of the politics of place are framed as a dichotomy – as being either one or the other, and significantly, as being either 'good' or bad'. The politics of place are thus located in the convergence of what seem to be dualistic relationships. They are argued to be particularistic, even as they are shaped by putatively homogenizing processes such as those associated with globalization and modernization. They often involve attempts to create fixity (rather than reflecting it) in a changing, mobile world. They are struggles that are material, as well as ideological and emotional. It is in the tensions between these seeming oppositions that place – as object, as a resource, and as an expression of relationality – becomes political. Indeed, the tensions between these apparent oppositions mean that neither 'place' nor the 'politics of place' can be easily or meaningfully discussed in the abstract; rather, the politics of place must be understood as grounded in empirical contexts and struggles. Through struggle, the apparent dualisms of place converge to create political opportunity structures that constantly mutate and create openings and closings for political projects of many sorts.

Politics of Placemaking and Place Identity

The perspective just outlined has not been the primary way in which the politics of place have been described. Instead, much of the geographical scholarship has been directed to empirical analyses that address the ways in which ideas about place and its significance are mobilized in various struggles. Sometimes these struggles zero in on particular places and the politics are focused on the people in a place, and the place's use, history and future. In much of this research, the focus is on politics *about* place and can often be thought of in terms of place making, whereby political action and debate are intended to make (or remake) the place. In other cases, however, geographers have focused on place as a tool – commonly a signifying tool – in broader political struggles. Very often, the emphasis is on the characteristics of the place that can be used to promote a particular cause. In these cases, the politics of place often emphasize place identity in an attempt to promote a particular vision or goal. In most cases, however, placemaking and place identity are intertwined in political mobilizations, and so authors focus on the resources and opportunities in and of place. Their concern is with the ways that institutional structures, 'idea sets', political relationships, and networks within and beyond the place in question shape the possibilities and potential for political action.

Geographical research on the politics of place has included a range of specific topics, including the politics of turf and NIMBYism (not in my backyard), the location of public services, neighborhood and contextual effects in elections, economic development priorities, public space, border studies and boundary making, and to some degree, nationalism. In this research, attention is given to the qualities, characteristics and contexts of a given place and how these affect the particular political topic under study. Attention is often directed to the values, economic conditions, and social relations that may characterize different areas and the ways these influence political goals and the outcomes of political processes. Studies of public service delivery in the context of devolution and neoliberalism, for instance, often consider the varying needs of neighborhoods across the city as well as the capacity of neighborhood based voluntary service organizations to meet those needs. They may also analyze the ability of local activists to draw attention to their needs, based on their political and economic connections and their ability to mobilize neighborhood residents and others from beyond the local area to draw attention to their plight.

As these examples suggest, political struggles over place are not 'merely' local struggles – as the invocation of place might imply – but rather, occur at multiple scales. For example, nationalist movements often use particular sites or monuments as an emblem of broader claims on territory as a nation. Similarly, what might appear to be a localized struggle related to a polluting factory in one neighborhood might be networked into a national or international organization fighting toxins or working toward environmental justice. This is one aspect of the networked, relational nature of the politics of place.

Through political struggles, there is often a hope on the part of participants that the qualities of a place can be changed by new users of the place and of adjacent places, when 'users' can be thought of as people, land uses, or activities. This belief suggests that the politics of place are implicated in the making and remaking of places. Sometimes the remaking of place is intentional; in other cases, it is perhaps not an intended consequence of action, but an outcome nevertheless. In still other cases, change in the qualities of a place – or perhaps the fear of change – is a reaction to actions that remake other places, but that have spillovers or that carry externalities. For example, economic development politics – whether at the level of the building, block, neighborhood, city, or region – entail the remaking of place at their core. This remaking is often discussed in terms of regeneration or revitalization, or may be even simply 'sprucing up' a place. Because places are interlinked, changes in one place will have effects on other places. Sometimes that enables neighboring places to 'free ride' off the benefits of redevelopment, but sometimes it means that other places have to carry the costs. Redevelopment efforts that attempt to move 'incompatible' land uses out of a renewal area – for instance, that move services for low income and homeless people out of areas to be redeveloped in high income housing – often simply force the land uses into other neighborhoods; rarely do they address the underlying reasons those services are required. In such cases, those services migrate to other neighborhoods, with attendant consequences. The ripple effects created through these efforts to change the characteristics of the redevelopment area can set in motion other efforts to 'protect' the areas to which services move; these secondary political efforts are often discussed in terms of a politics of turf that reflect a NIMBY attitude and are characterized as reactive and negative. The politics of placemaking, it seems, are characterized by the same power relations as society at large; they often reinforce those power relations, rather than challenge them.

The politics of placemaking often entail a struggle over the 'important' or defining' qualities of a place. These politics, in turn, rest on a normative vision of place, addressing not simply who or what *is* in a place, but who or what *should be* in it. Nationalist movements, for instance, often argue that certain people – usually ethnic or religious groups – have a right to occupy a place and that their culture and identity have a preeminent or even primordial claim to it. Such claims have, in turn, been used as justification for delimiting and strictly marking boundaries and for other acts of exclusion directed toward people or groups who do not 'belong' to the place. Sometimes these can be violent acts of exclusion, such as in ethnic cleansing; at other times, these acts of exclusion can be more subtle, even if relying on violence of a different type. Efforts to move homeless and street people out of redevelopment districts and spaces of consumption, for instance, involve a normative vision of who belongs in those newly made places. Curfews that limit the access of teenagers to shopping malls on evenings or weekends similarly reflect beliefs (and the power to enforce those beliefs) about who qbelongs in the mall and under what conditions. Laws and police tactics that regulate protest and political speech in certain areas and at certain times are similar statements of who and what belongs in a place. Each of these acts involves the identification and legitimation of the people and activities that can and should be in particular places by promoting – and often valorizing – place identity. Nationalism often mobilizes a place identity and takes extraordinary efforts to preen and maintain that identity. Less spectacularly, but perhaps no less effectively, the 'small' or mundane acts of regulating, policing, and managing a place rests on an identification of and with the place and an ability to act on beliefs about who and what belongs in place and, by extension, who and what is out of place. Because the practices that maintain place identity often involve exclusion, they give rise to concerns that the politics of place – and particularly politics that mobilize place identity – are reactionary, parochial, and in extreme cases, construct spaces of hatred and violence.

Geographers have demonstrated, however, that the efforts to claim space and identity are never so complete, finished, or clear cut. Resistance is also part of the politics of place and may also entail a normative vision or identity for place. So, for example, when countries build and maintain border fortifications to keep people from the 'other side' out, such as was done by the United States on its border with Mexico and by the Israeli government in the West Bank, the governments are claiming that people belong on one side of the border or the other, but not in both places; the basis of belonging is identity as citizen of a place, an identity that is exclusive and that 'others' do not share. The identity of a people, then, is important to the identity of the place. Yet, barrier fences are often quickly 'tagged' by graffiti artists, art installations, and political posters that promote a different identity for the area, identities that are accompanied by different understandings of who should be given access to the place, who belongs to the place, and under what terms. Similarly, acts of trespass, squatting, and other forms of transgression are often attempts to take land and to remake place. Community gardens in New York City, for example, were often built on squatted land – land that was formally owned by the City and held as a site for housing redevelopment, but that was claimed by the gardeners as a site of community formation; the gardens provide the basis for political mobilization to challenge what the gardeners believed were sweeping injustices in the distribution of social resources. In other cases, transgressions do not involve taking land, so much as changing the norms as to what is acceptable in a given place. Movements to allow breast feeding in public can be seen as attempts to transform what counts as a 'public' activity, shifting and perhaps making more permeable the metaphorical boundary between public and private places. Decisions on the part of Muslim women to wear the veil in public places are also sometimes motivated by a desire to change the boundaries between public and private, to advance a public recognition of the legitimacy of an identity that might otherwise be 'located' in private.

While some might romanticize these acts as progressive or as indicative of the ability of the weak or marginalized to challenge powerful institutions and ideologies, it should be recognized that resistance is not the exclusive domain of the oppressed and that people with more resources and power can also resist what they see as inappropriate uses and users of place.

Such resistances can be seen when police and private security agencies patrol neighborhoods or when business districts 'resist' the persistence of loiterers, vagrants, or people who do not 'belong' in a place. And the putatively progressive or reactionary nature of given resistances is not always clear. When indigenous groups confront governments and corporations that attempt to exploit the resources on indigenous lands, they often claim property rights the land and their culture. These acts are exclusionary at the core, relying on rights that are often assumed to be the domain of the powerful. As this example suggests, acts of resistance and transgression are often paradoxical. Such acts draw on deeply held ideas about what a place should be like and seek to remake place to be compatible with those goals and visions. They are political acts that are dependent on place, but the political goals are not pre given. Whether these are inward looking and regressive or outward looking and progressive is similarly contingent.

Place and Politics

While the meaning and signifying use of place is pliable and indeterminate, it is nevertheless the case that it is invoked in a variety of political conflicts and processes. What is it, then, about place that gives it such purchase? Why is it so frequently invoked – as an inspiration for mobilization and as a site – and why is it useful for analyzing political struggle? In answering these questions, it is helpful to suggest a definition of place – of what we appeal to when talking about place – developed by John Agnew in his influential 1987 book *Place and Politics*. Agnew defines place as constituted by three intertwined elements:

> *locale*, the settings in which social relations are constituted (these can be informal or institutional); *location*, the geographical area encompassing the settings for social interaction as defined by social and economic processes operating at a wider scale; and *sense of place*, the local 'structure of feeling'. (Agnew, 1987: 28)

These elements of place each plays a role in the politics of place, but in ways that shift or that differ from struggle to struggle. Their real significance is seen when they are combined, but it is useful to discuss them individually at first.

The first element – locale – is significant because it provides a focus for politics, because it grounds politics. For geographers, it is a truism that politics occur in place; it is therefore important to consider the character of place and the relationships in place that shape the ways in which politics play out. The locale, then, may provide a specific example or political issue around which agents can mobilize. The nature of the relationships within a locale – say for example, within civil society or between leaders of local industries and the local government – will shape how those politics play out. Values, goals, idea sets, and the qualities of place discussed previously are included in this concept of the locale.

Places, or locales, do not exist in isolation, however, but are connected to other places through the operations of processes that extend beyond a given place. They are also connected through the networked relationships between individuals and institutions in different places. 'Location' is the term Agnew uses to describe these externally oriented relationships and a given place's position within them. Many recent theorists of place use the language of actor network theory (ANT), instead, to talk about place as being a node in a network of economic, political, and cultural relationships through which power flows; these networks can be mobilized or activated in political struggle. The language used by Agnew and in ANT is slightly different, but they share a conceptualization of places existing in relation to each other and as being constituted through those external relationships in addition to the internal relationships of locale. In other words, they refuse to separate place from the networks and relationships that operate within and beyond the locale. These relationships and networks create what Doreen Massey refers to as power geometries in which political agents form identities and engage in political projects.

Finally, 'sense of place' stands for a shared set of interests that can (but that do not always) link people, even in the face of obvious differences. For example, a shared fate that comes from living together in a place can form the basis for an emergent sense of being a community, a public. While this commonality may not lead to agreement on all issues, it might be the basis for a willingness to work together, at least with respect to some issues. This sense of shared purpose may be vital to a particular political project, but it is also important for the creation of a polity in which diverse individuals form a collective that is capable of working together and capable of being governed. This communal spirit is as important to grassroots social movements as it is to nationalist movements. It is made possible by the social relationships that operate in places, big and small. Its importance, in terms of the meaning of place and in terms of how it may be used in political projects, is conditioned by the same power geometries mentioned above.

So how might these three elements of place come together? A host of studies have combined them in different ways, as scholars debate their relative importance, whether they are likely to lead to progressive or reactionary political movements, and so forth. One recent contribution to these debates is provided by Kevin Cox who discusses the 'spaces of engagement' and the 'spaces of dependence' that shape political struggles. On thinking particularly about urban politics, Cox argues that certain spaces – or to use his term, 'scales' – construct the issues or dependencies that are often the focus of local political mobilizations. These might be the broader political and economic relationships and networks that are part of the location of politics (to revert to Agnew's terms). These spaces are constituted through different relation ships than are the spaces of engagement, the spaces where political agents can address the issues that affect their lives and are the focus of political mobilizations. So, for instance, a neighborhood within a city may have suffered disinvestment due to decisions taken either in the central business district in the city, or more likely, due to decisions that are taken by people who are networked into the central business district, but also to places far away from the particular neighborhood and city. These far away places might include national capitals and corporate headquarters and might be conditioned by stock markets around the world, by resource availability, and by the emergence of new labor forces in other neighborhoods and cities. Together, these are the spaces of dependence. Yet, people in the newly impoverished neighborhood may not have access to those spaces of dependence. Instead, the spaces in which they can engage in politics might be those of the local community center, churches, or perhaps city hall; in other words, the locale. Yet people in neighborhoods even very poor neighborhoods – often can access networks that may transcend the neighborhood, and that may extend beyond the locale. Their political location, then, shapes the spaces of engagement, or to use the language of social movement theory, the political opportunity structures they can access. These are also part of the power geometries that shape the strategies of various political agents and the directions that political struggles follow.

Conceptualized this way, the politics of place are not inherently local, but rather, are connected to broader relationships, networks, geometries that converge, that link places, and that condition the kinds of resources that various agents may be able to mobilize. Similarly, the politics of place are not inherently inward looking. The elements of place combine to mean that the politics of place are simultaneously introverted and extroverted, thus often giving rise to the paradoxes mentioned earlier. By keeping the introverted and extroverted nature of these relationships in tension, some theorists suggest that we should rejig our language to talk about politics 'in' place, rather than the politics of place. Such a shift *in* language would recognize the role *of* specific features of place – the locale and perhaps the sense of place – in shaping the ways that political struggles are enacted, but still acknowledge the importance of the networks and relationships that locate politics in multiple places and spaces.

Tensions in and of Place

Introversion and extroversion would self evidently seem to be in a dualistic relationship, to form a dichotomy. Yet as we have argued, many theorists of place do not see a necessary opposition between them because of the dif ferentiated nature of place or the multiple networks and processes that construct place, and hold it relation to other places. These theorists see not opposition so much as tension defining the pulls and forces that simul taneously ground relationships and link places to each other. This refusal to adopt dichotomous thinking sug gests that other dichotomies that have pervaded debates about the politics of place might usefully be reevaluated. A host of geographers have taken up this task and, in so doing, have addressed some of the core assumptions and ontologies related to place and politics that have underlain social science research in this arena.

A logical consequence of understanding place as simultaneously extroverted and introverted is a reex amination of the assumption that place is particular and unique, whereas space is abstract and therefore capable of being generalized as a basis for either research or politics. Distinctions between space and place for instance, are called into question as geographers attempt to unravel the ways that the characteristics of place – what is often discussed in terms of context – are the basis of connections and networks. As networks link disparate places and struggles, they may also provide access to the resources, audiences, and political imaginations that can form the basis for broader political movements. For example, the so called 'anti globalization movement' is probably best understood as a series of movements that link very specific struggles – such as indigenous land rights campaigns, struggles to preserve old growth for ests, labor union campaigns against sweatshops – to each other and to other struggles in the effort to build social justice. Each of these movements has its own tensions – tensions that arise from any effort to work through the complex conditions, networks, and relationships that shape particular conflicts, as well as broader processes of exploitation and resistance. The point is not to sweep away these tensions but to show how place matters to the specific forms they take and to broader efforts to combat injustice. In so doing, the unique and particular (often assumed to be place) and the general and generalizable (often assumed to be space) are reconceptualized as co constituted, rather than distinct.

This approach to place and politics also entails a rethinking of the dichotomy between the material and the ideological or emotional, or the objective and sub jective. This can be seen in Agnew's insistence on in cluding sense of place as one of the three elements of place. Agnew roots his concept of sense of place in Raymond William's influential arguments about 'structure of feeling'. Williams developed this concept as a means to explain the lived experiences of people in place. 'Structure' is a key word because such experiences are shaped by the long, historically sedimented social relations that define a place, relations that are always both locally and extralocally derived. Such social relations are not made anew every day and in every encounter but are structured through ongoing – and struggled over – social, economic, and ideological practices. These practices are not invariant, but neither are they necessarily easy to shift: they are 'structured'. 'Feeling' is crucial, for Williams, because it is real humans with real lives who experience these structured practices. The emotional or subjective aspects of peoples' lives are critical to how they live these practices, what they do and do not do with them, and thus, how they may be restructured. For Williams, therefore, the objective (structures) and the subjective (feelings) are inseparable: each is a necessary component of the other. In this regard, sense of place is the affective understanding of structures of feeling, the ways in which people understand and experience the practices and meanings that are built into place; it conditions what those practices and meanings entail, are interpreted, why they are important, and how they make (and do not make) a place somehow valuable. As such, understanding sense of place is critical to understanding how place shapes political processes and debates. Returning to indigenous land claims, what are often claimed is more than property rights and the material stuff that is on some extent of land. Rather, these claims are often based on a different accounting of what the place is, which is the same as saying they are based on a different understanding of what the place means. Furthermore, many of these claims are cultural claims that are inextricably bound up with the differing lived experiences of the land in dispute. The sense of what place is, therefore, matters deeply to how, for example, rights claims of indigenous peoples or transnational corporations are put forward. The material and ideological components of place, in other words, are not dichotomous; each is necessary for the other.

Understood in this way, it becomes clear that place is inherently political, rather than prepolitical. For example, if place is understood as being constituted only in the realm of the material, then it can be mobilized in attempts to naturalize place, and thereby to erase politics or at least to neutralize them. At the same time, understanding that places are constructed, rather than pregiven may lead to a different kind of politics. Such an understanding is necessary to a recognition that place can be remade in ways that might redress inequality or that reflect different political values and goals. As such, refusing the dichotomy between the material and the ideological is a critical underpinning for analyzing the politics in and of place.

From this perspective, place provides an opportunity for addressing political issues in ways that might challenge common, settled, understandings of political issues. Returning to the discussion of sense of place, we noted that place is sometimes used to create a sense of common purpose that can be useful in attempting to mobilize people around particular political causes. This is certainly the case in nationalist mobilizations, but also in environmental politics and in the community power agenda from the 1960s and 1970s in North American and European cities. These examples (the first often cast as reactionary and the others as progressive and even radical) demonstrate that place is used in political strategizing, but the ends to which those strategies are put can differ dramatically. These examples, though, also demonstrate that apparent de politicization can also be part of the political uses to which place can be put. Nationalism often presents the claim to territory – to place – as natural, as unchanging, and in that way, as uncontestable. While that may be a strategy utilized by nationalist movements, moving away from the dualism between the material and the ideological exposes the strategy as ineluctably political.

A final dualism in the ways that place is sometimes conceptualized poses the static, fixed nature of place as being in opposition to mobility. In an increasingly mobile world, characterized by movement and flows rather than stasis, the opposition between movement and fixity can seem to make place and the politics surrounding it either irrelevant or reactionary. Geographers who emphasize the networked, relational basis of place, however, have argued that this opposition does not hold. If one assumes that places are networked, then the people, ideas, and activities that occur in one place will have implications for other places – and not simply in the sense of externalities discussed previously. Rather, analysts emphasize the ways in which people move through the world, through networks of places, physically and virtually. In this movement, people carry many of the expectations, values, and understandings of the world developed in their 'home' place or in one place, but each of these may change by virtue of movement. The result, scholars argue, is the need to reconceptualize sense of place to incorporate mobility. This new sense of place combines both stability and movement in ways that deny their opposition, and instead leads to a sense of place that is simultaneously grounded and mobile – that is, a mobile sense of place.

Other Knowledges, Other Politics

The preceding comments suggest an instability with regard to place and politics that reflects the networked, relational quality of place. This language, however, implies sterility and a dehumanized understanding of place. In emphasizing the positionality of place, it sometimes seems as though the ‘people’ within that place have lost their positionality, and perhaps their humanness. Recent theoretical work in cultural politics, however, has argued that it matters who is in these networked, relational places.

The study of politics has often applied a western sensibility rooted in the Enlightenment. The emphasis is on a universalistic definition of rationality, sometimes described as the ‘view from nowhere’. This view dismisses as irrational any political claims that do not fit into its own philosophy of knowledge. It is one source of the denial of the emotional and ideological, the concomitant elevation of the material, and the continuation of the dualism between those. It has perhaps reached its apotheosis in rational choice theory. This view and definition of rationality, however, is not universal, and might better be understood as a dominant view that is western, masculine, and bourgeois. It is therefore located within the relations that structure societies and that privilege certain kinds of agents and their knowledges. Yet, there are other perspectives, other positions, other knowledges that give rise to other rationalities and other politics. These views have been developed from a number of subaltern positions, including those associated with feminism, antiracist scholarship, queer theory, postcolonial studies, and indigenous studies. They sometimes give rise to arguments that people who come from marginalized subject positions have a better understanding of the world and of politics, because they must recognize and be able to negotiate dominant power structures in order to survive and engage in their political struggles. This is an argument often associated with standpoint epistemology. In other cases, these views have been presented as giving rise to different kinds of politics that are based on the knowledges that come from different subject positions without privileging the margins as a site of knowledge formation.

The political struggles surrounding indigeneity provide clear examples of what these arguments entail for politics and place. Struggles over land and the ability to maintain ways of living and being in place have become common, as corporations and governments seek to gain access to resources in areas long occupied by indigenous groups. As an important tool in these efforts, governments, corporations, and sometimes development agencies have mapped the areas, indicating where resources are abundant and accessible; these maps have often focused on resources that are of economic interest in the broader economy. In resisting these attempts, some indigenous groups have drawn their own maps, with different accountings of the resources in the area. These maps may involve different conceptualizations of what a resource is, identify resources in terms of cultural properties and meanings, and draw from indigenous knowledges about the significance – economic and otherwise – of what is in the place. While one might expect these struggles to remain fragmented and disconnected from each other, indigenous land movements in several places have joined together, sharing strategies and connecting their struggles to other struggles such as the global social justice movements. In so doing, they have sometimes reframed their arguments by couching them in the same languages and arguments about property that are commonplace in capitalist economies. As such, these seemingly local and particularistic struggles are networked into a broader set of struggles and political relationships such that local and global politics converge in particular places. There is no guarantee that such networks and convergences will lead to any greater likelihood of success, however that might be defined. Rather, the significance of these networks lies in the ways that different knowledges and understandings – knowledges and understandings that are often thought to be marginal to international politics and economics – can be connected and become part of the framework for debate in other places. In this way, the local knowledge that is the basis of these local politics moves beyond a given place, even as both are simultaneously grounded in place. The knowledge – no less than the politics – is thereby located.

Politics in Place

The politics in and of place are a struggle over how lives are to be lived, within the context of a globally and unevenly connected world. The politics *in* place – the politics rooted or grounded in particular locales, locations, and senses of place – are never only local, but are inextricably bound with struggles over the fates of people and places both near and far. The politics *of* place – the political struggles over who does and does not belong, over how the resources and rights of the locale are to be distributed, over which individuals and groups are to be granted the power to determine futures – are shaped both by locally rooted structures of feeling and by larger scale processes over which any one place may have very little control.

Places are storehouses of value and values. Places are crucibles within which meanings are forged and ways of life are shaped. As such they exist in dialectical tension with nationally, regionally, and globally scaled practices of economy, culture, and politics. Understanding this, David Harvey argued in his 1989 book, *The Condition of Postmodernity,* that a simple rule could describe the politics of place. He argued that those who command space – through ownership, regulation, or other forms of power – also control the politics of place, even as command over space takes some degree of control over a particular place in the first instance. But, as the foregoing has suggested, this simple rule needs to be amended. Recent work on the politics of and in place has made it clear, rather, that those who command space must always contend with – as different from controlling – the politics of place. Command is never settled or guaranteed, just as politics in and of place are never settled. In short, the politics of place are always contested.

See also: Actor-Network Theory/Network Geographies; Borderlands; Citizenship; Devolution; Emotional Geographies; Feminism/Feminist Geography; Gentrification; Harvey, D.; Homelessness; Mapping, Race and Ethnicity; Massey, D.; Nationalism; Neoliberalism; Postcolonialism/Postcolonial Geographies; Public Space; Queer Theory/Queer Geographies; Regionalisations, Everyday; Resistance; Social Movements.

Further Reading

Agnew, J. (1987). *Place and Politics: The Geographical Mediation of State and Society*. Winchester, MA: Allen and Unwin.

Amin, A. (2002). Spatialities of globalization. *Environment and Planning A* 34, 385 399.

Castree, N. (2004). Differential geographies: Place, indigenous rights and 'local' resources. *Political Geography* 23, 133 167.

Cresswell, T. (1996). *In Place/Out of Place*. Minneapolis, MN: University of Minnesota Press.

Cresswell, T. (2004). *Place: A Short Introduction*. Oxford: Blackwell.

Cox, K. (1998). Spaces of dependence, spaces of engagement and the politics of scale, or: Looking for local politics. *Political Geography* 17, 1 23.

Dikec, M. (2005). Space, politics, and the political. *Environment and Planning D: Society and Space* 23, 171 188.

Harvey, D. (1989). *The Condition of Postmodernity*. Oxford: Blackwell.

Harvey, D. (1993). From space to place and back again. In Bird, J., Curtis, B., Putnam, T., Robertson, G. & Tickner, L. (eds.) *Mapping the Futures: Local Culture, Global Change*, pp 3 39. London: Routledge.

Massey, D. (1994). *Space, Place and Gender*. Minneapolis, MN: University of Minnesota Press.

Massey, D. (2005). *For Space*. London: Sage.

Pred, A. (1984). Place as historically contingent process: Structuration and the time geography of becoming places. *Annals of the Association of American Geographers* 72, 279 297.

Routledge, P. (2003). Convergence space: Process geographies of grassroots globalization networks. *Transactions, Institute of British Geographers* 28, 333 349.

Staeheli, L. (2003). Place. In Agnew, J., Mitchell, K. & Toal, G. (eds.) *A Companion to Political Geography*, pp 158 170. Oxford: Blackwell.

Williams, R. (1978). *Marxism and Literature*. Oxford: Oxford University Press.

Relevant Websites

http://www.chej.org
Center for Health, Environment, and Justice, an anti toxins and environmental justice social movement.

http://www.forestpeoples.org
Forest People's Programme, organizes indigenous land rights movements.

http://www.greenguerillas.org
Green Guerillas, organizes gardens in low income neighborhoods in New York City.

http://ipsm.nativeweb.org
Indigenous People's Solidarity Movement, organizes indigenous groups making claims to land and self determination.

http://www.cobbmedia.com
New York City Environmental Justice Alliance, advocates for low income neighborhoods and environmental justice.

Planning, Urban

M. Huxley, University of Sheffield, Sheffield, UK

Glossary

Development Control The process of considering and granting or refusing permission for development.

Land–Use Planning The regulation and control of land uses through legal and statutory measures, such as zoning or development control.

Planning Theory General propositions about the nature of various forms of spatial planning: planning theory includes normative considerations of how good planning should be done.

Spatial Planning A comprehensive approach to coordinating policies and dealing with competing demands on the use of land at the regional or national scales.

Urban Design Three-dimensional esthetic and functional design of city buildings and spaces, including the master planning of specific large-scale projects.

Urban Planning State-related policies and programs for neighborhood, local and metropolitan areas, aiming to: effect broad-scale allocation of land uses to areas; order boundaries between them; manage ongoing uses of land, the spatial aspects of economic and social activities and connections between them; and ensure the optimal functioning of urban economic processes and social interactions.

Zoning The regulation of land uses through designation of permitted and prohibited uses in given areas delineated on zoning maps.

Introduction

Many textbooks on urban and regional planning begin with a discussion of different dictionary definitions and derivatives of the verb 'to plan', and draw out the connections between 'planning' as a general application of forethought to any form of future action, and 'plans' as designs or representations of physical space or built form. Town and country, urban and regional, city and rural, and physical and spatial planning are accordingly seen as specific instances of general human intentional capacities, that combine both 'planning' and 'plans' to bring about spatial distributions of built forms and social and economic activities.

While such discussions have the merit of attempting to clarify, at a conceptual level, the nature and purposes of spatial regulation and policy, they tend to homogenize what are culturally specific and historically contingent developments of regulatory systems. But rather than being specifically spatial instances of some universal, human future oriented mode of thought, urban planning, and related forms of spatial regulation may be seen to arise from diverse contexts and identifications of, and attempts to solve, disparate problems relating to space, populations, individuals, and activities occurring in particular trajectories of urbanisation. The ways problems are identified, and processes set in place to solve them, are not necessarily the result of the functional social needs, rational decision making, technical procedures or even participatory deliberation that some versions of planning history or theory would suggest.

Urban planning is temporally and culturally embedded in particular discursive practices that embody and express relations of power, which do not stand outside politics as neutral technical exercises in resource distribution or improvements to a general 'quality of life'. However, this is not to say that such attempts to bring about spatial order or create spatial connections have had no positive outcomes: but it does suggest that the aims and effects of urban planning need to be examined in relation to the particular historically situated circumstances of their institution.

This overview of urban planning therefore discusses, first, planning as expressed in systems of spatial regulation and the knowledge bases drawn on in relation to these. Second, three of the main strands of theorizing within the discipline are examined; and third, some of the intersecting critical debates about the objects and effects of planning are highlighted. In conclusion, convergences and divergences between the disciplines of urban planning and urban geography are examined to draw out the positive potentials for research and practice.

Urban Planning

Urban planning is one of a number of designations for forms of spatial planning that encompass ways in which land, land use, spatial morphologies, resource distributions, and social interactions may be planned and managed. Spatial planning involves attempts to plan socioeconomic processes to bring about certain ends, together with drawing up plans, maps, or diagrams that indicate where these activities should take place. Urban planning more specifically has as the objects of its concerns, metropolitan areas, towns, and localities (e.g., residential neighborhoods).

The very proliferation of qualifying terms for 'planning' indicates the diffuse scope, multiple purposes, and changing remits of spatial planning. Historically, the term 'town planning' appeared in English around the 1890s (although related terms and practices appeared earlier in Germany), and soon became widespread. Subsequently, town or urban or city planning became linked to a number of other 'plannings' – for instance, land use planning, country or rural, regional or spatial, social, and, more recently, environmental planning.

The remits of these planning foci vary in different times and places: in some planning systems, for instance, they include housing policy; in others, the focus is on urban design; in some, on environmental and ecological science; and in still others, planning includes what elsewhere might be considered building regulation. However, in general a distinction can be made between regulatory land use planning – variously known as development control, zoning, statutory planning – and broad level policies seeking to distribute social and economic activities in metropolitan, regional or national space ('spatial planning'). In policy analysis terms, detailed land use regulation is often considered to be a vehicle for the implementation of broader policies, but this is not universally the case: in parts of the US and of Australia, for instance, zoning regulations can operate separately from broader urban policies and are justified as bringing about good spatial and social order in and of themselves.

Urban planning is also variably positioned in relation to cognate knowledges and disciplines, not only in terms of theory and research, but also in indicating the objects of planning practice. In the UK, for instance, early planning texts in the first part of the twentieth century described planning as an art and a craft – an extension of architecture and design – applied in regulating physical structures and the esthetics of the built environment. But such texts also acknowledged the importance of applied sciences such as sanitary, civil and transport engineering, and technical skills in cartography, demography, and statistics. Over time, these connections and planning specialisms have been augmented by, among other fields, cybernetics and modelling; decision theory and policy analysis; physical and human geography and GIS; conservation and heritage studies; anthropology and sociology; environmental science and sustainability; and by perceptions of planning as a locus of place based conflicts requiring communication and mediation skills to resolve or manage.

This array of fields may reflect the complexity of the objects of study and practice. And yet, in general, the development of planning theories has been more concerned with the processes of planning and how they should best be carried out – data gathering and analysis, decision making, policy formulation, participatory practices, or professional ethics – than with questioning formulations of the problems that planning sets out to solve, or with assessments of its outcomes. The next section, therefore, provides a brief and necessarily simplified overview of some of the main strands of planning theory that illustrate this attention to planning process and normative prescription.

Planning Theories

Planning theories can be considered under three headings: theories that see urban planning as the production of urban design and/or large scale master plans; theories that see planning as a process of decision making or as providing policy advice for political decision makers; and theories that see planning as an arena for participatory democracy. These are not mutually exclusive and coexist in many forms, and the emphasis on each varies with time and place. (It should be noted that the trajectory traced here mainly relates to developments in British planning theorization. Since urban planning is culturally and historically specific, a comprehensive coverage of planning and planning theories, systems, and practice across all English speaking and European polities – let alone the rest of the world – is not possible within the constraints of this essay.)

Physical Planning, Urban Design, and Master Plans

Planning as urban design and master planning or neighborhoods or whole towns shares the assumptions underpinning Ebenezer Howard's ideas for Garden Cities as antidotes to untrammelled Victorian urbanisation. The idea that the environment of whole communities could be designed to facilitate not only health, but also productive economic activities and moral uplift, informed both the institution of planning legislation and the production of texts on the art and science of planning. Up until the 1960s, especially in the UK, planning texts focused on the technical provision of services and facilities, increasingly calculated through standardized measures of minimum requirements and spatial distributions – for light, air, street layout, traffic circulation, open space, schools, hospitals, shops, and so on. Planning theory was concerned with setting out general principles for bringing about desired physical configurations and social outcomes which could be applied in particular localities. Alongside these considerations were esthetic prescriptions for the production of appropriate styles of housing and village like neighborhoods, reflecting an anti urban standpoint that resurfaces in 'new urbanism', 'urban villages', and the generally renewed interest in urban design and master planning.

Systems Theories and Rational Comprehensive Planning

In contrast to concerns with the physical and esthetic qualities of urban development, the core concerns of the systems theories and procedural planning theories which rose to prominence in the 1960s and 1970s saw physical design and planning as too limited in scope to solve wider urban problems. Rather, solutions lay in improving the way decisions could be made about the objects of planning.

For systems theory, based on cybernetic conceptions of the maintenance of homeostatic systems, information about urban processes was the key to successful planning and the nature of that information was fairly un problematic. In principle, there was no limit to the pos sible range of information that could be rendered into quantitative form and manipulated by computer. These data would inform planners about the state of the urban system, thus allowing disturbances to systemic equi librium to be adjusted by suitable means – the provision of more services, adequate roads, better located industry, and the like.

At the same time, procedural approaches argued that planning was a form of rational decision making in which 'rational comprehensive' or even 'disjointed incremental' procedures for arriving at decisions were as important, if not more so, than the substantive matters about which decisions were to be made. In this, procedural theory was aligned with a US literature on general government policy formulation and analysis, concerned with im proving policy by ensuring that relevant aspects of a problem were assessed and weighted according to rational criteria before decisions were taken.

However, both these forms of planning theory en countered difficulties in dealing with questions of imple mentation and with questions of the relation between 'rational' planning, politics, and power. 'Optimal' solutions, based on the scientific 'correctness' of procedures and the 'truth' of the results of measurement and quantification, were increasingly challenged by protests over road building and high rise housing development and similar large scale reconstructions and population displacements in impoverished inner cities in the 1960s and 1970s. Government concerns about resistance to urban devel opment and policy implementation led to public con sultation becoming a required part of planning procedures, and planning practitioners found themselves increasingly involved in public meetings and community forums.

Communicative and Collaborative Planning

Communicative and collaborative planning approaches respond to the questions of implementation, politics, and power by exploring the involvement of 'the public' or 'the community' in planning decisions about local issues. Communicative planning theories envision planning's role as extending beyond managed consultation to mediating conflicts around land use and development decisions and fostering participatory inclusive local democracy. Ideas of 'multicultural' and 'insurgent' plan ning similarly connect planning's aims for the good city in the good society with political themes of resistance to power and transformations to direct democracy. This strand of planning theory has been particularly promin ent in the US, where there is a long history of local community political engagement and semi autonomous local government (in which land use zoning is an integral part of local 'police power').

A further development of planning theory and the critique of value free rationality currently takes the form of pragmatism, particularly in discussions of ethical planning and the collective or socially derived nature of truths and moral decisions. Pragmatist perspectives on the aims, purposes, and practices of planning raise questions about the status of truth, knowledge, and action in the world, some of which parallel (but are by no means identical with) post structuralist concerns (see below).

Collaborative planning theory, in contrast, is generally more applicable where issues of 'governance' are related to reconfigurations of the state, such as in UK or European contexts. Collaborative planning thus places emphasis on general processes of bargaining among diverse 'stakeholders' over spatial issues, in which plan ning plays a mediating role in the shaping of places.

Communicative, collaborative, and pragmatist per spectives on planning act as critiques of technical plan ning expertise and assumptions about the possibility of 'value free', 'rational' decision making, and seek inclusive participation by diverse social groups. Yet, while ques tioning the technical rationality of planning, these approaches also position planners as having skills and legitimacy in conflict management, mediation, and situ ated decision making that can be exercised in arenas where the effects of politics, conflict, and unequal power relations (including that of planners) – if not removed – can be identified and muted, restrained, or negotiated.

These three main strands of planning theory – design and master planning, systems theory and rational pro cedural decision theories, and communicative and col laborative theories – have in common a focus on the processes of planning rather than the objects or aims of planning; and share a normative prescriptive view of the purpose of theory as being the improvement of planning practice, especially planning as the practice of a profession. Planning is seen to be oriented to taking action, and to have practice at the heart of its theory and education, so although these theories may not be grounded in historical specificities of planning regulation or in actually existing planning practices (such as development control and application processing), they tend to support claims to professional legitimacy.

Until relatively recently, much less attention has been paid to how, and in what ways, planning practice relates to the improvement of cities and the lives of citizens. There are a number of studies that have attempted to unravel the difference urban planning might have made, and there are imaginings of what a good city could be like. But by and large, although planning education draws on a range of related disciplines that might provide bases for evaluations of planning, planning theories have tended not to problematize planning itself. Here, perhaps, urban geography and planning studies might fruitfully interact.

Critical Debates

While in the heyday of positivist quantitative modeling planning drew heavily on urban geography's developments of location theory, central place theory, and Chicago School urban ecology, the links between planning and geography have become more tenuous, with perhaps the most direct and currently being the application of GIS techniques. However, over the last half century, much of the critical debate about urban planning and its ability to produce the outcomes for which it aims has been propelled by work in the fields of urban studies and urban geography, particularly evident in critiques of positivist assumptions about the operations of spatial and behavioral processes.

Planning and the State

So for instance, in the 1970s and 1980s, Marxian derived critiques of planning as an instrument of the state argued that urban and regional planning could not hope to overcome the uneven spatial development inherent in capitalist processes, and could only act to smooth out disjunctures between interests in property markets, creating abstract homogenizing landscapes of capital and acting as an ideological smoke screen for the interests of capital against those of the working classes. Similar dark views of the state's role in producing visible spaces can be noted in some versions of surveillance studies that trace the implication of planning in the management of urban conducts.

These debates are pursued in planning literature, but they are difficult for planning academics and practitioners to assimilate into notions of professional practice. Views of the state as an arena for struggle around class or (Weberian) status interests, along with democratic theories of liberal government that offer at least some possibility of altering unequal distributions of power and resources, find more purchase.

Gender, Sexuality, and the City

In like manner, many feminist studies of urban inequalities faced by women have demonstrated how zoning and the separation of (paid) work places from residential areas, commuter oriented transport, and inadequately supplied or poorly located services directly affect the lives of women, reinforcing the difficulties of domestic work and creating barriers to paid employment. Like other structuralist social theories, radical feminism that sees the state and the planning profession as expressions of structurally inherent patriarchal power, implies that, without fundamental social change, planning cannot hope to achieve its aims. Liberal feminist views on the needs for equal political and institutional opportunities, on the other hand, have had significant influence on planning education and organizations.

Both of these bodies of critique of the gendered implications of planning policies, regulations, and practices have given rise to a substantial literature: but despite policy guidance and some regulatory change, these analyses have not been taken up in policies that might make substantial differences to the spatial arrangements of cities. Similarly, in many policy arenas the spatial and design barriers for individuals with differing physical capacities tend to fall, like gender considerations, under the remit of 'social planning', which historically has been a marginal concern in conventional planning systems.

Queer theory offers corresponding challenges to the heteronormative assumptions underpinning the planning of residential areas, the regulation of public spaces, and the interconnections of place and sexual identity. By making gay and lesbian identities and differences visible, queer theories unsettle the 'planning imaginary' that underpins spatial regulation.

Indications of planning's role in the creating and perpetuating spatial and social inequalities in the city are, of course, most starkly illustrated in the case of racial segregation.

Segregated Cities

The study of racially divided urban and regional spaces has long been a concern of urban geography, urban studies, and planning research. The roles of zoning, land use regulation, urban design, and urban policy in reinforcing and perpetuating segregated cities have been examined in almost all countries with formalized planning and housing systems, from the US and UK to South Africa under apartheid and Palestinian Arab settlements under Israeli planning control.

Here again, assessment of whether planning's implication in these social and spatial divisions is redeemable through better planning education, strategic knowledge, commitment to change, and willingness to take action, or whether such systems require complete social transformation, has depended on whether or not race is seen as a structuring factor in social relations.

Seen historically in the UK at least, and in concert with other assumptions about individual capacities and identities in relation to the creation of spaces and environments, much of planning's implication in racial segregation can be traced to nineteenth century projects of sanitary reform of the city. As historical geographers have shown, poor urban conditions – dirt, grime, human waste, lack of light and air – were seen as common causal factors for both disease and moral degradation, which were to be solved in schemes to separate healthful and morally uplifting living areas from industry. Reform of the physical conditions of the working classes sought to neutralize the medical and political threats posed by the 'rookeries', and to 'civilize' the inhabitants into acceptable modes of living. These conceptions coupled those of class (and gender) difference to racial difference, and informed equally, 'domestic' reformatory projects and the control and 'civilisation' of colonial populations.

Colonial Planning

Questions of the racial implications of planning's segre gationist impulses are present in planning studies of the circulation of settlement patterns across the British Empire and later, the Commonwealth. But while aware of the inequalities and dominations enacted through colonial planning, such studies have tended toward somewhat unproblematized views of the diffusion of planning practices from 'originators' to 'imitators'. Nevertheless, there have been a few notable examin ations of how local practices, legal and land ownership systems, or building styles have produced the particu larities of different planning systems in different places; how these differences circulate to become incorporated into the practices of planning elsewhere; and in par ticular, the significance of colonial urban experimentation for policies, plans, and practice in Europe and the UK.

Post colonial studies of urbanization in, for example, India and South East Asia, unsettle notions of uni directional colonial impositions of sanitary regulation, urban form, housing provision, or racial segregation. Further developments of such historically informed analyses could provide bases for conceptions of planning attentive to differences that 'parochialize' planning systems of the North/West.

Towards Poststructuralist Planning

Most of the foregoing critical examinations of the aims and effects of planning point to the way in which plans, regulations, and spatial policies seek to create order, fixity, and functional divisions of space, and to ensure transparent visibility of subjects and actions. Aspirations to overcome this 'dark side' of planning hope to work through participatory, inclusive, multicultural forums, and practices to engender democratic decision making. Within the planning literature, a little explored inter section with 'community development' studies and practices in the 'global South' is a field that could provide material for further reflection on the importance of theorising historical and cultural specificities.

But from other perspectives, such as Foucauldian approaches for example, participatory exercises are seen to foster and incite subjects to enact forms of liberal subjectivity and to conform to the rules of the game of participation. Therefore, while participation conducted in the name of 'planning' is productive of certain forms of liberal subjectivity, it is unlikely to produce the transparent, democratic deliberation between autono mous subjects to which some theories aspire.

Nevertheless, and at the same time, Foucauldian and post structuralist approaches emphasize the instability of discourses and the malleable, performative interrelations of subjects and spaces. Power is not unidirectional; 'the social' is an arena of indeterminacy; multiple discourses intersect, reinforce, contradict, and overturn one another; rather than consensus, 'agonism' and struggle are the stuff of politics; and subjectivity is open ended 'becoming'. The implications of post structuralist approaches to 'the urban' – space, place, and subjectivity – are being developed in planning contexts, where they present further provocation to policies in search of technical, rational 'solutions', and open the possibility that planning might not necessarily be tied to the production of static identities in fixed spaces.

Critical debates in and about planning parallel and draw on those in human and cultural geography, and while the connections between urban geography and planning are no longer as direct as they once were, their objects of research and analysis inevitably intersect. In so doing, they raise questions of significance for the normative aspirations of planning theory and practice.

Conclusion

Urban planning as the preparation of plans for, and regulation and management of, neighborhoods, localities, towns, cities, and metropolitan regions involves attempts to organize social and economic relations in space through different levels of government and 'governance'; and to deal with the political and social consequences of delineating spatial boundaries and influencing spatial allocations of development and resources. Thus, research in both urban geography and planning into issues such as gentrification, urban renewal, and regeneration, the creation of urban villages, social exclusion, the nature of the state, property investment and speculation, sustain able urban development, or analyses of policy discourses, are of central importance to understanding the objects of planning. In this, the mundane particularities of zoning, development control, and land use regulation should not be ignored by critical studies in either geography or planning.

Increasingly also, debates about space, place, post coloniality, 'the social', and 'the subject' taking place in geography more generally illuminate questions about how present configurations of space and social relations have come about, what might or might not need to change, and what urban planning might or might not contribute in any particular instance.

But planning considerations of such issues involve added dimensions of seeking guides to action, and a de sire for the convergence of theory, analysis, and practice. The emphasis on normative and prescriptive relevance has influenced how research and theorization about planning is conducted in the discipline, with attention being focused on the processes of planning – whether at a theoretical level or as required by the government of the day. Concerns with what is needed to improve planning processes, especially those relating to participatory involvement, tend to presuppose that these procedures will unproblematically result in improved environmental, spatial outcomes.

Geographical research into urban issues might be brought more directly to bear on critical examinations of planning processes, especially highlighting the situated contingency of planning practices, and thus further pro ductively modulating the (global North/West) under pinnings of planning's normative aspirations. Similarly, geography's concerns with the interconnections of space, place, power, subjects, and identities could well include closer attention to the aims and effects of planning practice and regulation in discursive and material struggles in specific localities and the creation of places.

See also: City-Region; Gender in the City; Governance, Urban; Historical Geographies, Urban; Imperial Cities; Lesbian Geographies; Place, Politics of; Postcolonial Cities; Postmodern City; Queer Theory/Queer Geographies; Regional Planning and Development Theories; Segregation, Urban; Urban Design; Urban Order.

Further Reading

Arnstein, S. (2000). A ladder of citizen participation. In LeGates, R. & Stout, F. (eds.) *The City Reader*, pp 240 253. New York: Routledge (reprinted from *Journal of the American Institute of Planners* 1969, 35(3), 216 225).

Dear, M. and Scott, A. (eds.) (1981). *Urbanization and Urban Planning in Capitalist Society*. London: Methuen.

Driver, F. (1988). Moral geographies: Social science and the urban environment in mid nineteenth century England. *Transactions of the Institute of British Geographers NS* 13, 275 287.

Fainstein, S. and Campbell, S. (eds.) (2003). *Readings in Planning Theory* (2nd edn.). Oxford: Blackwell.

Fainstein, S. and Servon, L. (eds.) (2005). *Gender and Planning: A Reader*. New Brunswick, New Jersey: Rutgers University Press.

Faludi, A. (ed.) (1973). *A Reader in Planning Theory*. Oxford: Pergamon.

Flyvbjerg, B. (1988). *Rationality and Power: Democracy in Practice*. Chicago, IL: University of Chicago Press.

Forester, J. (1989). *Planning in the Face of Power*. Berkeley, CA: University of California Press.

Friedmann, J. (1987). *Planning in the Public Domain: From Knowledge to Action*. Princeton, NJ: Princeton University Press.

Greed, C. (ed.) (1999). *Social Town Planning*. London: Routledge.

Hall, P. (2002). *Cities of Tomorrow: An Intellectual History of Urban Planning and Design in the Twentieth Century* (3rd edn.). Oxford: Blackwell.

Hall, P., Gracey, H., Drewett, R. and Thomas, R. (1973). *The Containment of Urban England*, 2 vols. London: Allen and Unwin.

Healey, P. (1997). *Collaborative Planning: Shaping Places in Fragmented Societies*. Basingstoke: Macmillan.

Home, R. (1997). *Of Planting and Planning: The Making of British Colonial Cities*. London: E and FN Spon.

Jacobs, J. (1962). *The Death and Life of Great American Cities*. London: Jonathan Cape.

Khakee, A., Somma, P. and Thomas, H. (eds.) (1999). *Urban Renewal, Ethnicity and Social Exclusion in Europe*. Aldershot: Avebury.

Keeble, L. (1952). *Principles and Practice of Town and Country Planning* (1st edn.). London: The Estates Gazette.

McLoughlin, J. B. (1969). *Urban and Regional Planning: A Systems Approach*. London: Faber and Faber.

Pile, S., Brook, C. and Mooney, G. (eds.) (1999). *Unruly Cities? Order/Disorder*. London: The Open University/Routledge.

Sandercock, L. (1998). *Towards Cosmopolis: Planning for Multicultural Cities*. Chichester: Wiley.

Taylor, N. (1998). *Urban Planning Theory Since 1945*. London: Sage.

Ward, S. (2002). *Planning the Twentieth Century City: The Advanced Capitalist World*. Chichester: Wiley.

Watson, S. and Gibson, K. (1995). Postmodern politics and planning: A postscript. In Watson, S. & Gibson, S. (eds.) *Postmodern Cities and Spaces*, pp 254 265. Oxford: Blackwell.

Relevant Websites

http://www.planning.org
American Planning Association.

http://www.acsp.org
Association of Collegiate Schools of Planning (US).

http://www.aesop planning.com
Association of European Schools of Planning.

http://www.inura.org
International Network for Urban Research and Action.

http://www.plannersnetwork.org
Planners Network North America.

http://www.pnuk.wikispaces.com
Planners Network UK.

www.whatisplanning.org
Planning Institute of Australia and Department of Environmental Planning, Griffith University.

http://www.planum.net
Planum: the European Journal of Planning on line.

http://www.rtpi.org.uk
Royal Town Planning Institute UK.

http://www.planningportal.gov.uk
UK Government Planning Portal.

Plant Geographies

A. Franklin, University of Tasmania, Hobart, TAS, Australia

Glossary

Actor-Network Theory (ANT) According to the Institute for the Studies of Science, Technology and Innovation, ANT focuses around an actor, sociotechnical entity, or technology, conceptualized as an emerging and increasingly stabilized network of associations between diverse material and nonmaterial elements – artifacts, humans, texts, symbols, concepts, etc. ANT usually follows the network-building strategies of a central actor. It stresses the mutual constitution and transformation of elements in the process, and the generation of social phenomena – agency, knowledges, institutions, power – as effects of network building.

City Forests The sum total of trees and plants in a city while fragmented by space, different plantings and management nonetheless constitute an environment similar to most other forests. They have very specific ecosystemic characteristics and can be manipulated and managed to achieve specific ecological tasks.

Dwelling This concept identifies that human cultures as well as landscapes and natural entities are co-constituted by their specific ties across time and space in specific landscape where they co-dwell and evolve. Dwelling prevents us from treating these entities as separable and explicable outside of their place and time and the other entities they share it with.

Posthumanism This perspective attempts to span the 'great divide' between science and social science by arguing that we cannot view humanity as simply 'humans among themselves' and seeks to produce accounts of relations between humans and nonhumans in which neither forms the controlling center and where agency is widely distributed.

The relationship between plants and humanity is a relatively recent field of study, with a number of important works emerging from the 1990s onward. It is prompted by a broader project to bridge what C. P. Snow called 'the great divide' between the natural sciences and the humanities in recognition of the multiple ways in which the natural and human worlds interrelate. Such a view recognizes there are fields and objects of that interrelation that are routinely missed by the institutional separation of the sciences and humanities.

In this emerging field, topics to date include tree symbolism and culture, trees and place, the agency of plants, gardens and gardening, city forests, bushfires and wildfires, modern forest dwelling, forests, national identity, and memory. While concern about environmental and ecological issues and the broader issue of understanding modern practices and engagements with the natural world underwrite much of this work, there are also important theoretical and methodological dimensions of this relationship that are explored through this research. Again, this has generally resulted in the problematization of the humanism of the humanities and the antihumanism of the sciences or what Nigel Clark calls the "now routine insistence on the porosity of the nature/culture binary." It has also renewed interest in nature–culture assemblages and concepts, such as Sharma's work on the forests of Europe, Cloke and Jones' work on orchards, Ingold's discussions of landscape and dwelling, Franklin's identification of the agency of trees in the life of bush suburbs, and Godefroid and Koedam's discussion on 'city forests'.

It is arguable that the emergence of our ecological and environmental sensibility only came into being through the work of early phytogeographers such as Alexander von Humboldt. Not only through his unusually broad perspective but also through his experience as an intrepid and prodigious plant collector (von Humboldt, for example, collected 6300 new plants during one epic field trip across the Americas between 1798 and 1904), he was able perceive something that had not come into view from previously laboratory based and intensive study of specimens: discrete natural environments. His global classification of vegetation zones became the foundation for new models and language for the natural world that would eventually become separable into discrete environments as systems of interconnected and coevolved natural communities. Subsequently, new subdisciplines of geography emerged around ecological perspectives and, more broadly, biogrography.

Humanity has proved problematic not only to ecological models but also to ecological politics. Quite where can humanity be properly and ideally fitted into ecological communities and 'the environment' has never been certain, even if most identified modern humanity as the source of most of the problems. At the same time, it was recognized paradoxically that modern humanity had also developed a great esthetic and moral regard for natures of many kinds. It was largely in relation to this paradox and almost no systematic empirical investigations of it that a new human geography of plants (and animals and other nonhumans) began.

It is not only through vegetation zones that plant life has become evocative and constitutive of place. Jones and Cloke endorse a view of place as "'local distinctiveness'; as 'some form of physical/imaginative space such as a village, urban district, park, wood, forest or region' and having some internal cohesion distinct from that around it" (Jones and Cloke, 2002: 9). Their book *Tree Cultures* focuses on the "place of trees and trees in their place" and documents how trees are central to understanding such places as cemeteries, heritage trails, English urban squares, and orchards – in general terms and in their local specificity. Trees and other plants form very specific backcloths to the design, planning, and symbolism of human spatial arrangements. But plant geography is far more than plant semiotics. People live and work with plants (and other nonhumans) as they all form parts of a sensual, embodied chain of 'dwelling' in specific landscapes. In the 'dwelling' perspective, landscapes become accretions and productions of what Haraway calls naturecultures. In 'places' the boundary between nature and culture cannot be seen because natural and cultural life is never lived in a separable way, nor are their specific plant and cultural forms cultivated in isolation from one another. One has only to think of how hedges in England articulate as important cultural and natural objects – simultaneously esthetic and economic, legal and archaeological, work and leisure, regional and ethnic, habitat and floricultural, boundaries/enclosure, and pathways/freedom. While we can make theoretical points with hedges in general, they are only ever specific hedges in time and space in practice, a point that Schama develops with respect to forests in his book *Landscape and Memory*.

As an antidote to the omnipresent tendency to generalize Western views of nature and environment (e.g., Hay's notion of 'Western environmental thought'), Schama's analysis of forests in Poland, Germany, and England demonstrates how quite specific and even opposed meanings and practices can be associated with trees and forests, even at the national level and even among nations with strongly overlapping cultural and economic traditions. In his analysis of the mythic foundations of forests in the Polish imaginary, it is clear that they emerge as the historic refuge of a nation in trouble from powerful neighbors. His analysis of *Pan Tadeuz* by the Romantic writer Mickiewicz demonstrates that "the truly heroic historians of the drama are trees. Their antiquity gives them an authority that spans the generations of Polish history, and they shelter within their woodland recesses the values that keep Lithuania – an idea as much as a place – alive" (Schama, 1995: 56–7). Mickiewicz saw the forest as the mythical origin of the Polish Lithuanian nation and a fortress place that would protect and preserve them. To that end, they were aided by 'native wood fauns' or *Leshy* who would deceive and punish any enemies who pursued them into its depths.

Schama concludes that "landscapes are culture before they are nature; constructs of the imagination projected onto wood, water and rock…. But…once a certain idea of landscape, a myth, a vision, establishes itself as an actual place, it has a peculiar way of muddling categories, of making metaphors more real than their referents; of becoming, in fact, part of the scenery" (Schama, 1995: 56–7). This is tragically true for forest myths of Germany.

By contrast to the Polish Lithuanians, the German people were 'formed' by their forests, and it was their specific, unique life and culture as forest dwellers that distinguished them from all other ethnic groups. The original idea of an ethnic purity and a strength that came from that purity is shown by Schama to originate from a backhanded compliment from the Roman historian Tacitus in his *Germania* written in AD 98. After 210 unsuccessful years of trying to subdue and bring the Germanic tribes under Roman rule, Tacitus referenced their tenacity, strength, and ability to succeed in a place of 'foul bogs' and 'bristling forest' as a means of chastizing what he took to be a languishing and corrupt Roman culture. In other words, he penned the original noble savage. Schama shows how *Germania* – the ethnography – had an important after life as *Germania* – the constitution – suggesting strongly not only the notion of a pure nation but also one gained through association with the forest. By the late eighteenth century a forest cult, the Romantic *Grove League*, emerged under the symbol of the oak tree committed to 'rejuvenate their Fatherland'. It produced an enduring suspicion of urban modernity as unhealthy and un German, sentiments that combined, portentously, with Riehl's view that Jews dominated urban and commercial life. Following the humiliating defeat of 1918, the Tacitus myth was reworked by archaeologists and prehistorians into an ethnic blueprint for an expansive and exclusive warrior culture. This was eagerly used by the Nazi leadership who built themselves gothic lairs in the woods and unified a nation with blood and soil slogans.

In the English Greenwood mythology, forests and trees play yet another important role, this time as the seat of liberty and the bastion of ancient English communal values. During the period of Anglo Saxon kings, the highly worked woodlands had been a critical part of the common lands – a nourishing resource base for an entire rural culture. After the Norman conquest, vast areas of 'forest', a very different undomesticated woodland, was created for the exclusive hunts of the Norman kings and their elite. This barbaric act not only created hardship and poverty but also symbolized in a ubiquitous place from the values that had been lost but that would be recovered. It also had a number of lives and manifestations in the Robin Hood stories, in Shakespeare, and in the romantic novels. "It is a social idyll: …in the sylvan habitat of Merrie England… it is forever green, always summer. The nightingales sing, the ale is heady, and masters and men are brought together in fellowship by the lord of the jest: Robin Hood" (Schama, 1995: 141).

Schama's examples show just how significant plants are as symbols of social identity, but his analysis goes beyond a purely representational geography to suggest that practice and dwelling are constitutive of, or neces sary for, the myth themes themselves. Rival's edited collection *The Social Life of Trees* takes this further. The contributors, all social anthropologists, agree that "the long tradition of interpretation of natural symbols as internal representations of external reality has come to a close." Ever since Mary Douglas disowned her own representational analysis of natural symbolism, anthro pologists have followed her lead: identification between nature and social life works both ways. Inspired by Frazer's *Golden Bough* which enumerated 38 cultures with tree cults, these authors aimed to find, simultaneously, the local theories and use based practices that give rise to meanings of trees "as similar to those characterising the human life cycle, or the continued existence of social groups." The collection concludes that there is something both universal and local and theoretical and practical about tree symbolism. Critically, it seems to suggest the significance of a new research methodology: we should no longer merely ask what plants 'mean to humans', we should also ask what they 'do with humans'.

The idea that identification between plants and humans operates 'both ways' is illustrated by Philippe Descola, who develops the idea of 'overarching formulae' that emerge from the experience of interacting with nature. Descola cites the work of Haudricourt on the radical nature of the Neolithic revolution on relations between humanity and the natural world: "human beings were led to establish bonds of affective coexistence with the species they had domesticated, bonds of the same order as those prevailing within the social sphere." However, the new pastoral technologies in turn also produced the 'social' figure of the shepherd, who became enshrined in Western cultures as the good leader of men or sovereign. The shepherd is a hybrid form, never existing in nature or society before but arising from the experience of the new technology and naturecultures. A similar dialectic example is given of the very great care and attention given to yam growing in New Caledonia. The emphasis on friendship and respect and the avoid ance of violence is paralleled by the political institutions that developed around the possibility of their surplus production and the associated but subsequent patronage of 'big men' politics.

Similar new directions have influenced the way geographers and sociologists have investigated relations between plant and human cultures. Influenced by the actor network theory (ANT) which seeks to unearth and situate the agency of nonhumans (as texts, narratives, rocks, plants, etc.) in heterogeneous configurations and assemblages with humans, Jones and Cloke show how a variety of trees (as orchards, cemetery plantings, heritage, or urban squares) are co constitutive of place and dwelling. As they argue, "this approach offers a way to deal with the 'richness' of places, where the ecological and the cultural, the human and the non human, the local and the global, and the real and the imaginary all become bound together in particular formations in particular places" (James and Cloke, 2002: 9). Critically, they demonstrate that without an understanding of what trees 'do', such places and cultures cannot be understood either as a formation or as an entity charged with ethical and moral content.

It is not only the tree–place relatings that have interested recent scholars. Others working from within the related posthumanist perspective have asked: "what have we always missed from the disciplinary separation of science (say botany) and the humanities (say, soci ology)? If their separation is truly foolhardy, an act that blinds as much as it enables us to see, then what has been missed as we think about trees, plants and humans in different moments, as belonging to different times and processes?"

Franklin's work on 'eucalypts' and Australia illustrates how entities that at first sight seem supremely distant (primordial climax vegetation and modern city cultures) are intimately related, so much so that they can best be described as dance partners in a 'choreography of fire'. His analysis shows that the fire orientated (or needing) evolutionary development of the eucalypts was first en hanced and completed by the arrival of aborigines and their fire torch technologies. A relatively minor and spatially confined disturbance orientated group of plants broke free of their ecological constraints as aboriginal firing intensified and encompassed the continent. The domination of eucalypts in the Australian landscape was only made possible by the geographical spread of humanity, but equally, the spread of aborigines and their manipulation of the environment for hunting was intimately related to the fire agency of the eucalypts. Franklin also shows how the modernizing footprints of the colonial and postcolonial urban settler societies sin gularly failed to achieve a humanist domination of nature and in fact became enrolled as new dance partners for the eucalypts. A belated romanticism for Australian nature which developed into an econationalism ensured that modern Australians built their suburbs into those areas of bush that were unsuitable for farming, often into hills and gullies. Small townships, rural towns, and the leading edge of the sprawling metropolis were all intertwined with these deadly trees. Every new generation built their homes on this fire front, and every generation became fire orientated as they experienced the tension of every 'fire season' and the excitement and thrill of the unpredictable but inevitably regular inferno. Franklin shows not only how preparation and fighting of fires provided a unique ordering of Australian society but also how the increasing numbers of deliberately lit serious fires relate to the positive generation of individual and community ties, especially in those areas that remain relatively eventless for most of the time. Both plant and human community, regeneration arises from the ashes of this fiery dance.

A posthumanist approach to relations between culture and plants in the Australian Rangelands by geographers Kay Anderson and Nicholas Gill reveal yet more surprises. Interviews with pastoralists suggest that their close observations of the land allowed them to restructure a pastoral imprint in very arid areas. Against environmentalist claims that the heavy footprints of cattle destroy delicate soils, the pastoralists observed that the craters formed by cattle create a "cradle for new seeds to develop, protected from the wind on the open plains and held little pockets of water when it rained." Far from destroying native grasslands and herbage, many landscapes greened up after stocking. As one pastoralist respondent put it: "a lot of country in Central Australia is no good until it's stocked…it's just like ploughing the land" (Gill and Anderson, 2005: 3).

The hybrid forms that are produced through the interaction of humans and plants are nowhere better illustrated than through the emergence of modern gardening. Franklin argues that modern gardening provided the means by which modern cultures 'renaturalized' themselves after the sudden and violent urbanizations of the early nineteenth century. Such was the force of the new esthetic for the stock of globalized garden plants that it began to influence not only the new forms that cities were to take (garden suburbs being a twentieth century architype for suburbanization everywhere) but also how cities expanded (very often along railways or underground railway lines that offered city dwellers an opportunity to live in 'the country', enjoy their garden, but work in town). Recent research by Mark Bhatti and Andrew Church demonstrates how horticultural engagements with plants in everyday garden spaces provide powerful and commonplace engagements with the natural world, which may have far wider ramifications for human–natural relations than the mere production of flowers, vegetables, or esthetic effects. As they argue, there are important and complex elements about relationships established in urban gardens that are frequently overlooked, namely, the mixing together of domestic social relations with domesticated natural species. Quite what gardeners and plants do, together, has been the subject of Hitchings' enquiries into gardens in London using the ANT. Following Latour's command to 'follow the actors' Hitchings set himself the difficult task of not only following the gardeners into the garden but also somehow following what the plants were doing. In talking with the gardeners they clearly thought the garden was their creation; that 'they' were the designers and that plants were merely passive objects under their control. However, by focusing on what the plants were doing, he discovered "their individual liveliness, beauty and unpredictability. The plants performed themselves into existence as discrete entities such that they became almost considered as similar to people. And this was something that the gardeners enjoyed." The monolithic humanism that normally characterizes discourses of gardening is thus challenged by Hitchings, but in so doing it exposes the play of power and enrolment that is going on between human and plant actors, not only in gardens but also more or less everywhere.

See also: Biopolitics; Culture/Natures; Nature, Social.

Further Reading

Bhatti, M. and Church, A. (2001). Cultivating natures: Homes and gardens in late modernity. *Sociology* 35(2), 365–383.

Cittadino, E. (2003). Ecology. In Heilbron, J. L. (ed.) *The Oxford Companion to the History of Modern Science*, pp 158–159. New York: Oxford University Press.

Clark, N. (2002). The demon seed: Bio invasion as the unsettling of environmental cosmopolitanism. *Special Issue on Cosmopolis. Theory Culture and Society* 19(1 2), 101–126.

Coke, P. and Jones, O. (2001). Dwelling, place and landscape: An orchard in somerset. *Environment and Planning A* 33, 649–666.

Davis, M. (1999). *Ecology of Fear*. London: Picador.

Descola, P. (1992). Societies of nature and the nature of societies. In Kuper, A. (ed.) *Conceptualising Society*, pp 107–126. London: Routledge.

Franklin, A. S. (2002). *Nature and Social Theory*. London: Sage.

Franklin, A. S. (2006). Burning cities: A posthumanist account of Australians and eucalypts. *Society and Space* 24(4), 555–576.

Franklin, A. S. (2007). A choreography of fire: A posthumanist account of Australians and forests. In Pickering, A. & Guzik, K. (eds.) *The Mangle in Practice: Science, Society and Becoming*, pp 17–45. Durham, NC: Duke University Press.

Gill, N. and Anderson, K. (2005). Improvement in the inland: Culture and nature in the Australian rangelands. *Australian Humanities Review* 34, 31–48.

Godefroid, S. and Koedam, N. (2003). Distribution and pattern of the flora in a peri urban forest: An effect of the city forest ecotone. *Landscape and Urban Planning* 65, 169–185.

Hitching, R. (2003). People, plants and performance: On actor network theory and the material pleasures of the private garden. *Social and Cultural Geography* 4(1), 99–114.

Hitching, R. (2004). At home with someone nonhuman. *Home Cultures* 1(2), 169–186.

Jones, O. and Cloke, P. (2002). *Tree Cultures: The Place of Trees and Trees in Their Place*. Oxford: Berg.

Macnaghten, P. and Urry, J. (2001). Bodies in the woods. In Macnaghten, P. & Urry, J. (eds.) *Bodies of Nature*, pp 166–182. London: Sage.

McMahon, P. (2004). Improving Australianness: The native plant breeding project in the Melbourne Botanic Gardens, 1949–1957.

British Australian Studies Association Biennial Conference, ‘Projecting Australia’ Cardiff 2 5 September 2004.
Pickering, A. (2007). New ontologies. In Pickering, A. & Guzik, K. (eds.) *The Mangle in Practice: Science, Society and Becoming*, pp 1 16. Durham, NC: Duke University Press.
Pyenson, L. and Sheets Pyenson, S. (1999). *Servants of Nature: A History of Scientific Institutions, Enterprises and Sensibilities*. New York: W.W. Norton.
Pyne, S. J. (1992). *Burning Bush: A Fire History of Australia*. North Sydney: Allen and Unwin.
Rival, L. (1998). *The Social Life of Trees*. Oxford: Berg.
Schama, S. (1995). *Landscape and Memory*. London: HarperCollins.

Point Pattern Analysis

E. Delmelle, University of North Carolina at Charlotte, NC, USA

Glossary

Delaunay Tessellation A geometrical triangulation technique based on Thiessen polygons connecting nearby points, resulting in triangles. This method is extremely useful in determining the neighbors in the direct vicinity of an event.

Monte Carlo Simulation A process by which a set of events is simulated randomly several times. The ultimate goal is to derive a sampling distribution via simulation of the null hypothesis. Monte Carlo simulations create several realizations of a phenomenon as it potentially varies over space.

Moving Window A filtering process, in which the value at the center of the moving window is computed as a function of the surrounding cells, or events located within the window.

Poisson Process A nondeterministic process used to model the locations of random incidents in geographic space. Random incidents are independently occurring. The Poisson parameter λ stands for the intensity of the process.

Smoothing A process (or result of a process) where values are averaged based on surrounding values, following a moving window for instance. Smoothing also refers to procedures that amend an existing grid.

Basic Visual Exploration of Point Patterns

A spatial point pattern refers to data in the form of points, where a point denotes the location of an event. In geography, it is usually desirable to analyze whether these particular events, such as crimes, car accidents, fires, emergency calls and diseases for instance, exhibit a spatial pattern (e.g., hot spot), in the hope to better understand the underlying process that generated the events. Such events are considered discrete, because they occur at specific locations. A visual inspection of a map showing the locations of those events (e.g., scatter plot) may not always bring a correct interpretation of the true pattern, especially when events occur repeatedly at the same location. True clusters may go unnoticed. Events that occur repeatedly in time at the same location can be represented using a bubble plot, where the size of the bubble is a function of the event frequency.

General Descriptive Methods

A wide variety of descriptive statistical techniques exist to describe the geographical characteristics of a point pattern, such as its central tendency (the 'center' of the point pattern) or its dispersion (the degree of separation or clustering among the points). To measure central tendency, a first step consists of computing the mean center of a set of events. The mean center is the central or average location of a set of points, computed as the mean x and y coordinate values for all the events in the study region

$$x_{\mathrm{mc}} = \frac{\sum_{i=1}^{n} x_i}{n}, \; y_{\mathrm{mc}} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad [1]$$

where $\bar{x}_{\mathrm{mc}}$ and $\bar{y}_{\mathrm{mc}}$ are the coordinates of the spatial mean, and x_i and y_i are the coordinates of an event i, and n is the number of events. The mean center is significantly affected by the presence of outliers as well as the frequency of occurrence in the case an incident happens more than once at the same location. The introduction of weights in eqn [1] can reflect the importance of some events, which extends eqn [1] to the notion of weighted spatial mean. The weight is an interval or ratio value associated with a feature attribute. The weighted mean center is given by

$$x_{\mathrm{wmc}} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}, \; y_{\mathrm{wmc}} = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i} \qquad [2]$$

where the subscript 'wmc' stands for weighted mean center. The addition of weight is particularly useful for areal data since some regions can be larger, or simply more important than others. The weighted mean center is an important indicator, especially in locating facilities, which will serve the entire population. The location of a new emergency service is an excellent example since we try to minimize the distance to the population.

Dispersion of Point Distribution

The spatial mean indicates the central location of a set of events, but does not reflect the dispersion of the point distribution. Indeed, two point patterns may have the same mean center, yet one may be very clustered around that mean center, whereas the other may have a highly dispersed, uniform pattern. The standard distance *SD* indicates the dispersion of point events from the spatial mean.

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - x_{\mathrm{mc}})^2 + \sum_{i=1}^{n}(y_i - y_{\mathrm{mc}})^2}{n}} \qquad [3]$$

When point events are very dispersed, the *SD* value will be higher than when incidents are clustered around the spatial mean.

Orientation of Point Distribution

In which direction do events tend to cluster? To illustrate the spatial spread of a set of point locations, the use of a standard deviational ellipse is a common and effective visualization tool, because it captures the directional bias in a point distribution. For instance, events occurring along a road network, such as bike accidents, will exhibit a linear pattern. The deviational ellipse is characterized by three parameters: the angle of rotation θ, deviation along major axis a, and the deviation along the minor axis b. As a result, we obtain the directions of maximum and minimum spread. **Figure 1** illustrates the mean center and directional ellipse for two distinct point patterns. The point pattern represents the origins of patients being treated at a city hospital in Cali, Colombia for the months of September and July 2004. A larger ellipse denotes that patients are clearly more dispersed in September; the north northeast (NNE) direction of the ellipse is affected by the presence of patients in the northern and northwestern part of the city. In July, however, the ellipse is compacted (closer to a circle), and points in the east northeast (ENE) direction. As a result, the mean center is closer to the hospital.

The absolute coordinate values of event i (x_i, y_i) are transformed to relative values (x_i', y_i') based on the location of the mean center. The center of the transformed coordinate becomes (0, 0).

$$x_i' = x_i - x_{\mathrm{mc}}$$
$$y_i' = y_i - y_{\mathrm{mc}} \qquad [4]$$

The angle of rotation θ is given by:

$$\tan\theta = \frac{\left(\sum_{i=1}^{n} x_i'^2 - \sum_{i=1}^{n} y_i'^2\right) + \sqrt{\left(\sum_{i=1}^{n} x_i'^2 - \sum_{i=1}^{n} y_i'^2\right)^2 + 4\left(\sum_{i=1}^{n} x_i' y_i'\right)^2}}{2\sum_{i=1}^{n} x_i' y_i'} \qquad [5]$$

where a negative tangent implies a rotation of the angle counterclockwise, and a positive tangent implies a clockwise rotation, respectively. Finally, to reconstruct the ellipses, the deviations along the x and y axes must be known:

$$\delta_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i' \cos\theta - y_i' \sin\theta)^2}{n}},$$
$$\delta_y = \sqrt{\frac{\sum_{i=1}^{n}(x_i' \sin\theta + y_i' \cos\theta)^2}{n}} \qquad [6]$$

General Clustering Methods

Quadrat Analysis

The general descriptive techniques discussed in the previous section do not inform on the clustering level of

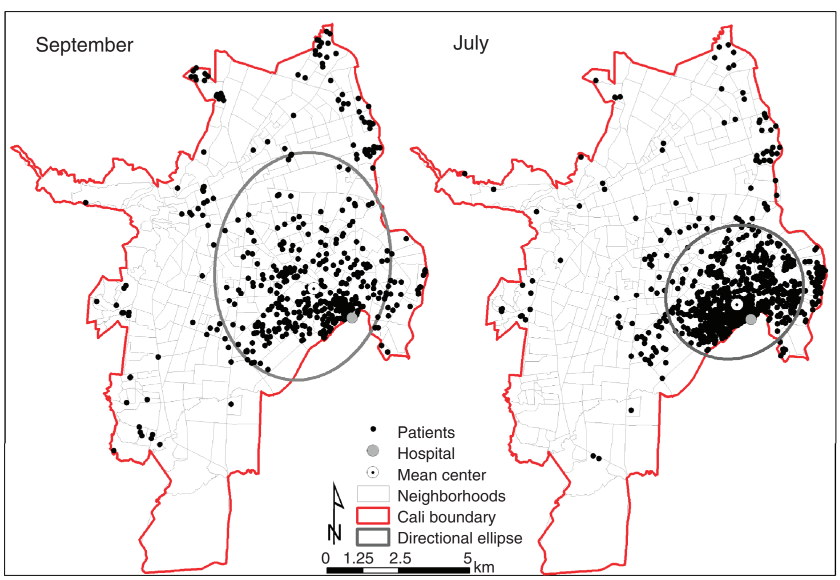

Figure 1 Patients being treated at a city hospital in Cali, Colombia, for the months of July and September 2004.

the data; rather they indicate the potential spread and orientation of the dataset. One spatial statistical method to measure potential clustering among events is a quadrat analysis. It consists of overlaying a uniform grid, which is made out of equal sized quadrats, onto the map containing point incidents. The number of events within each cell (or quadrat) is determined. Essentially, the analysis compares the actual count of events within each quadrat with the expected frequency of occurrence if the process generating those events was random in nature (known as complete spatial randomness or CSR). Complete spatial random processes, also called Poisson processes, are characterized by the fact that each point is likely to occur at any location. In general, we are interested in the hypothesis that the current phenomenon does not follow a Poisson process – in other words, that points are more clustered or more dispersed than would be expected under a Poisson process.

Define $P(X=x)$ as the probability of finding x events within a specific quadrat. Assuming a Poisson process, this probability can be calculated as follows:

$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!} \qquad [7]$$

where λ is the intensity of the process. The true value of the term λ is usually unknown, but can be computed as the average number of events per quadrat. For instance, if a grid of 100 cells is imposed over a study area, and 67 events occur within this area, a good estimate of intensity is $\hat{\lambda} = .67$. The probability of finding no event within a cell is $P(X=0)=.5117$, one event $P(X=1)=.3428$, and two events $P(X=2)=.1149$.

Determining the presence or absence of clustering

The variance mean ratio (VMR) is a formula for measuring the degree of clustering within a study region based on quadrat analysis.

$$\mathrm{VMR} = \frac{\dfrac{\sum_{i=1}^{m}(x_i - \lambda)^2}{m-1}}{\lambda} \qquad [8]$$

where m is the number of cells of equal size, λ the mean number of points per cell, and x_i the number of cells containing i events. The numerator stands for the variance of the frequency of cells containing i events, and the VMR index is then used to standardize this variance relative to the mean cell frequency. A VMR value less than 1 characterizes patterns that exhibit a tendency toward uniformity. However, when the VMR value is greater than 1, cells are characterized by either a much greater or a much lower number of events than expected which indicates a highly clustered spatial pattern. A VMR value close to 1 is typical of a random, Poisson process.

Example

Figure 2 illustrates a simulated point process of 100 events. Half of those points are generated in a random fashion, while the remaining half is explicitly clustered within three blocks, two of size 20 by 20, and one of size 7 by 7. The size of the study region is 100 by 100 units. If we divide the area in four rows and four columns for instance, it creates 16 squared cells of 25 side units each. The corresponding VMR ratio is 2.54, characterizing a pattern of strong spatial clustering, since there is a substantially greater and lower number of events in each of these 16 cells than expected.

Advantages and limitations of the technique

Besides the relative easiness to implement the quadrat analysis and the quick results it provides, the technique has some drawbacks, mostly related to cell size, and the ability of the method to differentiate different spatial patterns of points.

The size of the quadrat is of paramount importance. A cell too small in size may cause a high variability in quadrat counts, which will result in several empty cells and consequently clustering may go unnoticed. When the cell size is too coarse on the other hand, the within cell patterns will be missed. **Figure 3** shows the variation of the VMR value as a function of the cell size. As could be expected, when the cell size is too small, the spatial pattern is unnoticed. It is usually desirable to fix the cell size to obtain an average of 1.6 to 2 events per cell. As opposed to the nearest neighbor approach (see below), quadrat analysis does not look at the interseparation distance between events. Additionally, the technique solely relies on frequency counts and the spatial arrangement of the events is not explicitly

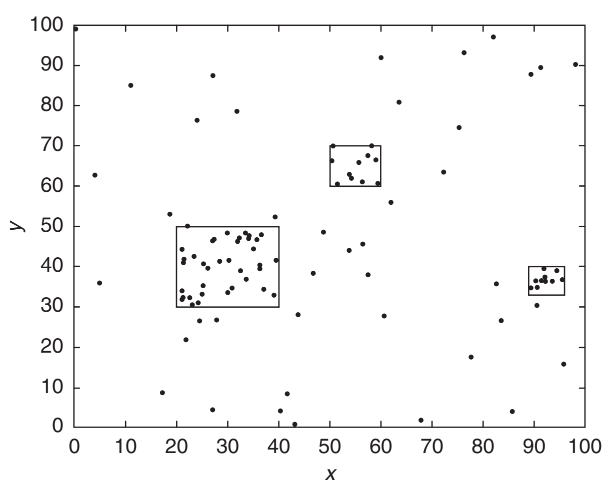

Figure 2 Simulated point pattern where half of the points are generated at random, and the other half is clustered within the three boxes.

considered. In **Figure 4**, two distinct patterns of incidents have been simulated. However, their VMR value is exactly the same. In the right **Figure 4b**, a clear clustering occurs in the northwestern part of the study area, while on the left **Figure 4a**, incidents seem to be spread out.

Nearest neighbor

The nearest neighbor technique avoids the problem of quadrat size determination. Essentially, the nearest neighbor statistic is used to test whether a set of incidents are closer together than would be expected by random distribution. The statistic computed R is the ratio between the observed average distance between all events and the expected value if the events were distributed in a random fashion:

$$R = \frac{R_o}{R_e} = \frac{\bar{d}}{\dfrac{1}{2\sqrt{\lambda}}} \qquad [9]$$

where $\bar{d}$ is the mean of the distances from incidents to their nearest neighbor and λ is the total number of incidents in the area. The term R_o denotes the observed average distance between incidents and their nearest neighbors, while R_e is the expected distance between incidents under random circumstances. When the near est neighbor statistic R is equal to 0, all points are in one location. When $R_o = R_e$, we have a perfectly random pattern. For a square area, when $R = 2.14$, we have a perfectly uniform pattern where incidents are spread out evenly.

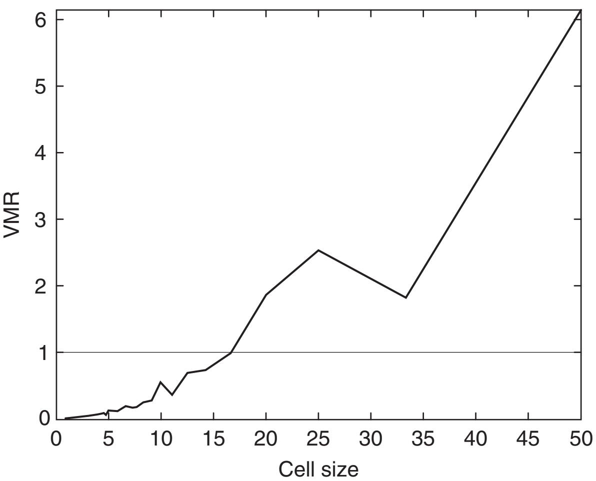

Figure 3 Change in VMR with varying cell size. Uniform patterns are observed when the quadrat size is small, while clustering is obvious at coarser scale (cell size greater than 17), and maximum for a cell size of 50 (study area divided into 4 quadrats).

Example

The point pattern from **Figure 2** exhibits a R_o value of 3.186, a R_e value of 4.896, and an R value of 0.651, which confirms that the pattern is clustered. Confirming the presence of clustering can also be accomplished using Monte Carlo methods, by simulating a random pattern many times, and obtaining the R_e value based on those random patterns. For instance, following 1000 Monte Carlo simulations, an average R_e value of 3.731 was obtained, and R was equal to 0.854, which confirms that the point pattern was clustered. The Monte Carlo approach is especially useful when the study area is irregularly shaped.

Advantages and limitations of the technique

The shape of the study area will decrease the nearest neighbor statistic, especially if it is narrow, because events will necessarily be next to one another. Another problem commonly noted in the literature is that

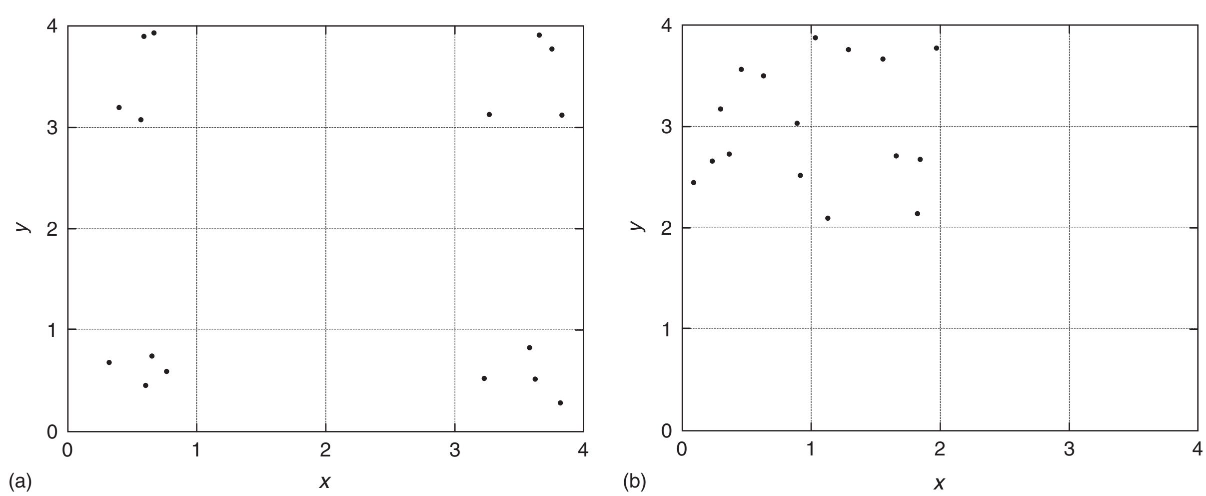

Figure 4 Two very distinct spatial patterns exhibiting a similar VMR value of 3.2 (cell size is of unit 1).

clustering may only be detected on a relatively small scale. In that case, it is possible to extend the statistic to higher orders (second or third nearest neighbor). It should be noted that it is possible to compute the cumulative distribution of separating distances, which in turn sheds light on the spatial scale of clustering. In **Figure 5**, a thick line denotes the cumulative distribution of the clustered dataset mapped in **Figure 2**. Not surprisingly, the curve increases sharply close to the origin since many events are located next to one another, which suggests clustering due to inter event attraction within certain parts of the study area. The curve flattens out beyond a separating distance of 100 units, indicating an absence of spatial clustering at that scale. The dotted line is the cumulative distribution of a random point pattern. At short separating distances, this pattern exhibits a greater amount of pairs of points than in the case of a symmetric pattern. Finally, the last curve represents the cumulative distribution following a perfectly symmetric (regular) point pattern. Since the events are very spread out from one another, the potentiality for a spatial clustering pattern is always below a random point pattern.

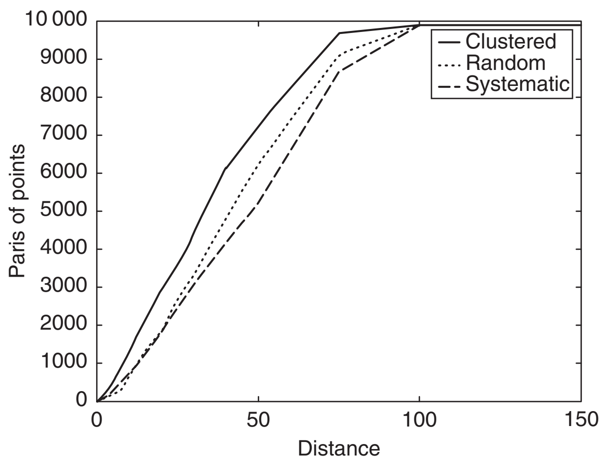

Figure 5 The cumulative distribution of pairs of points as a function of the separating distance (since n 100, there are n^2 possibilities).

The nearest neighbor technique solely relies on the distances between events, not whether there is a high concentration of events next to each other. In **Figure 6** for instance, two distinct patterns of incidents have been simulated. Although their nearest neighbor values are exactly the same, their spatial patterns are totally different. Patterns may vary when calculating the neighbor statistic at different orders of distance (second or third nearest neighbor for instance).

Most geographic information systems (GISs) offer the flexibility to compute this statistic. Since a complete search for the nearest neighbor of each point is carried out, the procedure can be time consuming when n is large. It is however possible to construct a Delaunay tessellation of the n events and search for nearest neighbor distances within the tessellation.

K-Function

The K Function provides an alternative to the nearest neighbor statistic as a technique aimed at determining the amount of clustering at a wider range of scales. To compute the statistic, a circle of a specified radius (h) is placed over each point in the set (i). Points within this circle (j) are counted and the circle is then moved to the next point to begin the process again. When this has been done for all existing points n, the radius is then expanded and the process is repeated until a specified maximum radius is reached.

$$K(h) = \frac{A}{n^2} \sum_{i \neq j} \sum I_h(d_{ij}) \qquad [10]$$

where d_{ij} is the distance between two events i and j within the study region, A the size of the study area, and $I_h(d_{ij})$ an

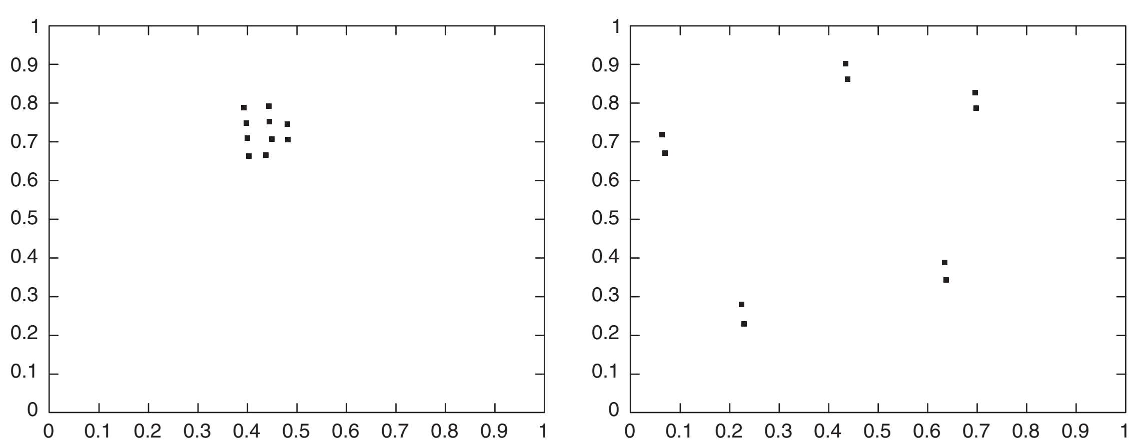

Figure 6 Two very distinct spatial patterns exhibiting a similar nearest-neighbor value.

indicator function defined as:

$$I_h(d_{ij}) = \begin{cases} 1 & \text{if} \quad d_{ij} \leq h, \\ 0 & o.w. \end{cases} \tag{11}$$

In general, a high value of $K(h)$ denotes clustering. The value of $K(h)$ can be graphed against the distance (h) to show the scale(s) at which the point pattern exhibits randomness, clustering, or dispersion. These values can be tested using Monte Carlo methods in which K functions are computed for a large number of randomly generated point patterns. For each h value, the 5% highest and 5% lowest K values for these random patterns form the upper and lower significance envelopes under the null hypothesis of CSR. **Figure 7** illustrates this concept, where the dotted lines represent the 5% and 95% confidence intervals, and the thick line represents the observed K function values (for different h values) for the dataset of **Figure 2**. When the observed value for a given h is between the upper and lower envelopes, we conclude that the point pattern is random at that spatial scale (h). When the K function is above the envelopes, we observe a clustered pattern, whereas the pattern is spread out when the K function is below the envelopes. The K function value increases as h increases since more observations are counted within the radius h.

Advantages and limitations of the technique

The K function is superior to the nearest neighbor technique in that it analyzes possible point patterns at different scales, while the nearest neighbor approach uses distances only to the closest events, hence nearest neighbor only considers the smallest scales of patterns. The K function is however relatively time consuming to compute, and it is recommended to use specific software, such as CrimeStat.

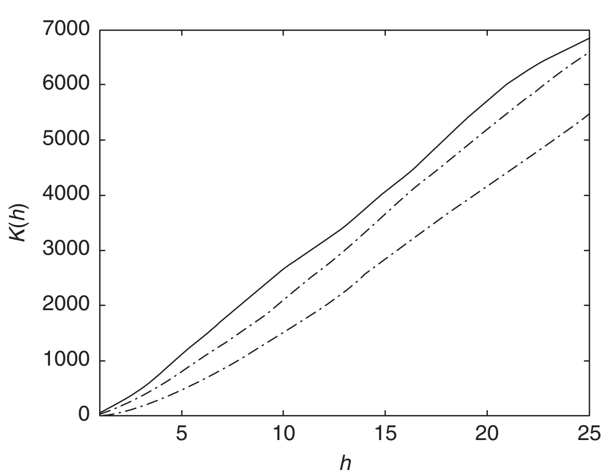

Figure 7 The *K*-function for the clustered dataset of **Figure 2**. Note the lower and upper envelopes, which were obtained after 100 Monte Carlo simulations.

Kernel Density Estimation

Neither the nearest neighbor, nor the K function, nor the quadrat analysis, identify locations of clusters; rather, they determine an overall tendency toward clustering, randomness, or dispersion. Kernel density mapping provides a means for visually identifying clustered areas. An advantage of the density mapping is that most GIS softwares support the method.

To calculate this density, the entire region is divided into a grid and a search radius is drawn around each grid point g in a similar way to a moving window average. A suitable window – defined by the radius size τ – is moved over the fine grid of locations, and the intensity at each grid point is estimated from the event count per unit area within the window centered on that grid point. The total number of events i that fall within the search radius is divided by the size of the window resulting in a density value for each grid point. Kernel density mapping extends this method by assigning weights when performing the search so that points closer to the center of the window receive a higher weight than those further away. Mathematically, the kernel density at a grid point is denoted $\hat{\lambda}_\tau(g)$ and can be estimated as follows:

$$\hat{\lambda}_\tau(g) = \sum_{h_i \leq \tau} \frac{3}{\pi\tau^2}\left(1 \quad \frac{h_i^2}{\tau^2}\right) \tag{12}$$

with h_i the separating distance between an event i and the grid point g. The bandwidth τ determines the amount of smoothing. The size of the bandwidth τ will affect the outcome of the map as a smaller search area will result in more distinct events to be highlighted. On the other hand, a larger radius can identify broad zones where a high number of incidents exist. A bandwidth that is too large will stretch the kernel and the surface will appear flat. The choice of the bandwidth may depend on the purpose of the study.

Example

Figures 8 and **9** represent the surface following a kernel density passed on the events from **Figure 2**, where bandwidths of 5 units and 10 units were used, respectively. The peaks denote regions where there is a strong concentration of events, which coincide with the three blocks highlighted in **Figure 2**. As expected, the intensity of the peaks is higher when the bandwidth is smaller, and a greater smoothing occurs when the bandwidth increases. Interestingly enough, many small peaks appear on **Figure 8** when the bandwidth is $\tau = 5$, which suggests that this bandwidth was not appropriate, as those peaks correspond to a single event. The Kernel map appears more flat in **Figure 9**, because a greater bandwidth ($\tau = 10$) smoothes out the concentration of events.

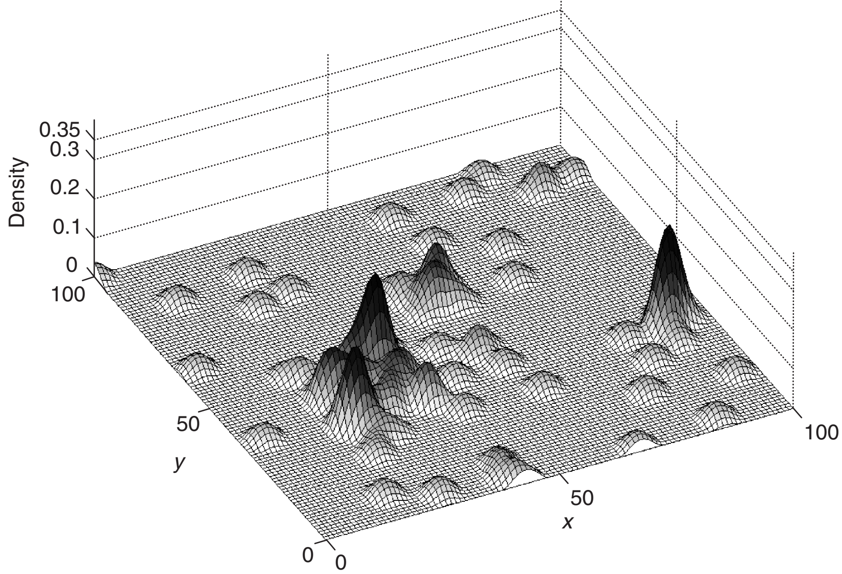

Figure 8 The kernel density surface for the clustered point dataset of **Figure 2**, calculated with a bandwidth τ 5.

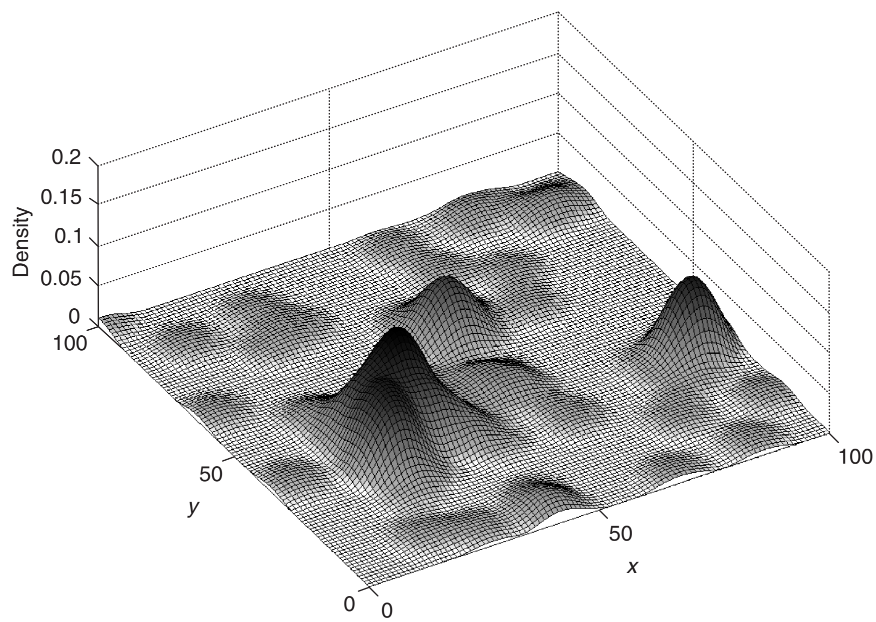

Figure 9 The kernel density surface for the clustered point dataset of **Figure 2**, calculated with a bandwidth τ 10.

Spatial Patterns on a Network

One drawback to using these aforementioned spatial clustering methods is that they assume that events can be located anywhere in the study area. However, many point events occur on a network, such as car crashes or pedestrian accidents. Such events are most often restricted to the existing road network; so ideally, the analysis should focus solely on the places where it is possible for accidents to occur. Recently, network based point pattern analysis methods have been developed that provide a more accurate computation of accident clustering. However, such methods remain computationally intensive and to this date they are not supported by any GIS.

Conclusion

The techniques presented in this article are excellent tools to analyze spatial point patterns, and remain relatively easy to perform, either within a GIS or by programming. However, one must remain careful, and consider the various limitations of each technique. For instance, the nearest neighbor distance analysis is prone to errors associated with irregular boundaries, while quadrat analysis is prone to errors associated with edge effects and variability within quadrat cells. Attention must also be paid to a few key issues such as spatial scale, edge effects, and events occurring on a network. It is usually desirable to conduct a test for clustering at different scales to evaluate the magnitude of this clustering. These issues are being researched now; however, most of them are not available yet in commercial GIS.

See also: Edge Effects; Monte Carlo Simulation; Scale Analytical; Spatial Autocorrelation; Spatial Clustering, Detection and Analysis of; Spatial Data Mining, Cluster and Pattern Recognition; Spatial Data Mining, Geovisualization; Spatial Filtering/Kernel Density Estimation.

Further Reading

Bailey, T. C. and Gatrell, A. C. (1995). *Interactive Spatial Data Analysis*. New York: John Wiley.

Chou, Y. H. (1996). *Exploring Spatial Analysis in GIS*. Santa Fe, NM: Onword Press.

Diggle, P. (1983). *Statistical Analysis of Spatial Point Patterns*. London: Academic Press.

Fotheringham, S., Brunsdon, C. and Charlton, M. (2004). *Quantitative Geography*. London: Sage Publications.

Levine, N., Kim, K. and Nitz, L. (1995). Spatial analysis of Honolulu motor vehicle crashes: I. Spatial patterns. *Accident Analysis and Prevention* 27, 663 674.

Okabe, A. and Yamada, I. (2001). The K function method on a network and its computational implementation. *Geographical Analysis* 33, 271 290.

Okabe, A., Yomono, H. and Kitamura, M. (1995). Statistical analysis of the distribution of points on a network. *Geographical Analysis* 27(2), 152 175.

Rogerson, P. A. (2001). *Statistical Methods for Geography*. London: Sage Publications.

Wong, D. and Lee, J. (2005). *Statistical Analysis of Geographic Information*. New York: Wiley.

Yamada, I. and Rogerson, P. A. (2003). An empirical comparison of edge effect correction methods applied to K function analysis. *Geographical Analysis* 35, 97 109.

Yamada, I. and Thill, J. C. (2004). Comparison of planar and network K functions in traffic accident analysis. *Journal of Transport Geography* 12, 149 158.

Relevant Websites

http://www.icpsr.umich.edu
Inter University Consortium for Political and Social Research (ICPSR): CrimeStat.

Policing

N. R. Fyfe, University of Dundee, Dundee, UK

Glossary

Antisocial Behavior A broad term to capture a complex of thoughtless, inconsiderate, or malicious activity.

Closed-Circuit Television Used to refer to a network of cameras linked to a centralized control room where images can be monitored remotely.

Territoriality A strategy used by individuals, groups, or organizations to exercise power over space.

Zero Tolerance A term used to refer to a form of policing which focuses on relatively minor crimes in the belief that this will prevent more serious types of crime from occurring.

Introduction: From the Geography of Crime to the Spatiality of Police Work

The origins of geographical interest in policing can be traced back to studies of the geography of crime. Although early work mapping patterns of crime and deviance rarely considered the significance of the police other than as the source of 'official' crime statistics, by the early 1980s a more critical engagement with the inter connections between policing and the study of crimin ality began to develop. For some geographers this involved recognition that the mapping of official crime statistics needed to acknowledge the significant impact of peoples' behavior in terms of their decisions to report crime to the police as well as the effect of police recording practices and procedures on what crimes appear in official statistics. For others, notably the Canadian geographer, John Lowman, it meant a more radical attempt at developing a 'geography of social control' that rejected the analytical separation of crime from the control of crime. For Lowman, crime could not be adequately theorized in isolation from the control system which defines and processes deviance, and thus, a concern with police practices needed to be a central element in any explanation of patterns of crime. Further impetus for geographers to engage with policing came in the work of Susan J. Smith. Determined to move beyond traditional geographies of deviance and the concerns of conventional geographies of crime, Smith focused on the experiences of victims, exploring the importance of fear and examining the role of the police and other agencies in tackling the problem of crime. Moreover, much of her research was conducted against the backdrop of violent clashes between the police and black communities in several UK cities, and thus a concern with many of the strategic policing questions that emerged from these conflicts feature prominently in her work. For example, how is it possible to resolve the tensions between central government and local public interests in the determin ation of local policing policies? And how is it possible to secure police participation in local multiagency approaches to law and order?

Building on these early contributions, more focused and in depth research into policing began to emerge in the 1990s. Nicholas Fyfe examined some of the key interrelationships between the police, space, and society revealed by research in the rapidly expanding field of police studies. Different spatial patterns of police deployment at a local level, for example, can impact significantly on the distribution of crime, fear of crime, and community relations. In addition, ethnographic research into police subculture (the so called 'cop culture') highlights how police offers develop complex mental maps in order to make sense of the places they police. In other contributions, Fyfe drew on his ethno graphic research with London's Metropolitan Police Service to show how Hagerstrand's time geography can help understand the spatialities of police work. Using the notions of 'capability', 'coupling', and 'steering' constraints to examine the time–space choreography of routine police work, Fyfe developed a framework that could provide insights into the interplay between the time–space movements of individual officers and the contexts or structures of 'work', 'community', and 'law' which shape police actions.

A second major contribution to understanding the spatiality of police work was Steve Herbert's *Policing space*. Like Fyfe, Herbert's key concern was to make sense of the interplay between structure and action in the analysis of everyday police behavior, paying particular attention to questions of police territoriality. Drawing on a period of participant observation with the Los Angeles Police Department, Herbert identified a set of 'normative orders' (the rules and practices that structure action) which inform the ways in which police officers 'make and mark space' and thus shape police territoriality. These include the law (which underpins a distinction between public and private space, and therefore, affects where the police can and cannot act); bureaucratic control (which determines the influence of the hierarchical police command structure on local policing activity); a love of adventure (which is important in terms of officers seeking opportunities to demonstrate courage and physical strength); and a concern with morality (which informs how the police attempt to construct boundaries between 'good' and 'evil'). Employing this analytical framework provided important insights into a range of territorial police actions, from creating and enforcing boundaries to regulating peoples' movements in space.

However, while these studies of the police in the 1990s were enormously important in countering what had, until then, been the conspicuous absence of policing from the landscapes of human geography, the research agenda in this field also had three important limitations. First, the analytical gaze of human geographers was firmly fixed on the public police rather than on the broader concept of policing which embraces a multiplicity of agencies and activities that use surveillance and the threat of sanctions in order to ensure security. This conflation of 'the police' with 'policing' thus created an unnecessarily narrow research focus, one which overlooked the important roles played by a wide range of public, private, and voluntary actors in regulating conduct and maintaining social order. A second and related limitation of these early studies was that, through their focus on the spatiality of police work, there was a tendency to marginalize consideration of the wider social and political landscapes within which policing takes place. As a result, the potential links between studies of policing and broader debates within social and political geography were underdeveloped. A third limitation of these early geographical studies of the police was that they were largely focused on the local scale and on urban environments. This meant that policing at broader geographical scales (such as the global or transnational level) and the policing of other environments (like cyberspace) were not addressed. Against this background, the following sections consider some of the ways in which the research agenda around policing in geography has expanded in recent years – driven both by theoretical and policy concerns – and has begun to address some of the limitations identified above.

Policing, Urban Restructuring, and the Politics of Space

An important stimulus in shifting the attention of geographers from 'the police' to 'policing' came with the plethora of research carried out on urban restructuring in the 1990s. Amidst the various claims and counterclaims that contemporary urban restructuring had spawned a new form of urbanism – the postmodern city – the issue of policing loomed large in vivid descriptions of gated communities, fortified buildings, and intensely surveilled public spaces, shopping malls, and theme parks which are viewed as crucial components of postmodern urban structure. Indeed, for many researchers, one of the defining characteristics of the postmodern city is that it is a place where 'form follows fear' – a 'carceral city' in which a complex mix of public police, private security guards, and electronic surveillance are used to ensure the security of high income gentrified enclaves and corporate citadels.

Although more recent contributions to analyzing contemporary urban restructuring have shifted the theoretical spotlight from postmodernism to the politics of neoliberalism, questions of policing continue to feature prominently. A key challenge for neoliberal urbanism is the question of how best to secure the commercial success of downtown areas which have witnessed significant capital investment in the bid to create vibrant, entrepreneurial cities. Echoing the insights of postmodernists, analysts of neoliberal urbanism have made much about how the economic vitality of downtown areas is highly dependent on a costly system of surveillance performed through a blend of architectural design, private security, and a technological infrastructure of closed circuit television (CCTV) cameras as well as the presence of the public police. Indeed, reflecting more general trends in neoliberal forms of governance, the policing of entrepreneurial cities increasingly involves complex local 'security networks' comprising the public police, municipal policing (such as city wardens provided by local urban authorities), civilian or voluntary policing (such as membership of neighborhood watch or block watch schemes), and the rapid expansion of commercial or private policing. Such local security networks, it is argued, are being actively encouraged under neoliberalism as part of a wider political project which stresses the importance of partnerships between the public and private sectors and emphasizes the need to reconfigure responsibility for social welfare in ways which rely less on the state and more on individuals and communities. Moreover, it is claimed that under the aegis of urban entrepreneurialism the priorities of such security networks are focused on reclaiming the streets for the 'consumer citizen' and enforcing patterns of behavior commensurate with the free flow of commerce and the new urban esthetics.

It is not just the downtown, commercial core that is witnessing a reconfiguration of policing. The political spotlight has also been turned on suburban, residential communities, where the drive to tackle 'antisocial behavior' and bring back a sense of 'respect' has also resulted in the emergence of new forms of social control as governments focus on the politics of conduct. This is most clearly evident in the UK where the Blair Government's 2006 Respect Action Plan has introduced a range of mechanisms to tackle the problems of antisocial neighbors and incivility. At the heart of this strategy are the antisocial behavior orders (ASBOs), which can be used against anyone causing harassment or distress in their neighborhoods and involve prohibitions on individuals entering certain areas or carrying out specified acts. Through the use of ASBOs, the government hopes to encourage 'governance through community' by directly involving communities in the surveillance and regulation of antisocial conduct. While this new battle to confront antisocial behavior is still in its early stages, for geographers concerned with understanding the changing nature of local policing, the introduction of ASBOs, alongside other mechanisms for defining the required civil conduct and disciplining of incivility, open up important opportunities for research. Indeed, when these developments in the policing of residential communities are taken, together with the changing nature of policing in downtown areas, it is clear that a fundamental restructuring of social control in cities is underway in which the public police are now working in complex partnership with a new coalition of individual citizens, local communities, and the private and voluntary sectors.

Policing Difference

The contemporary restructuring of the policing of downtown areas and residential communities, clearly, has significant implications for the social geography of cities. Indeed, a core concern for many geographers is the extent to which the public spaces of cities now appear to be under threat from new forms of policing. Responding to warnings that a consequence of the crusade to secure the city is the destruction of any truly democratic urban space, geographers have drawn attention to the way different forms of policing are implicated in strategies aimed at the 'purification' of public space. One example is the so called zero tolerance policing (ZTP) which was introduced in New York City in the 1990s and has subsequently been adopted in several other US and UK cities. Focused on 'quality of life' offences, including drunkenness, begging, public urination, and vandalism, ZTP aims to reclaim public spaces for 'respectable citizens' and prevent neighborhoods from entering a downward spiral, whereby minor offences lead to more serious types of crime. Many urban geographers, however, have been quick to condemn ZTP and its impact on the social landscapes of late modern cities, arguing that police targeting of the homeless, panhandlers, graffiti artists, and squeegee cleaners, crushes street spontaneity and vibrancy. Indeed, according to Neil Smith, writing in the context of New York City, ZTP provides yet further evidence of a revanchist urbanism underpinning contemporary urban policy in which the wealthy middle classes in a divided city are increasingly vicious in defending their privileges from the poorest communities.

Similar concerns about the social consequences of new forms of policing have been articulated by geographers in relation to the increasingly widespread deployment of CCTV surveillance systems. The CCTV research literature is replete with allusions to Orwell's 1984 and Foucault's discussion of Bentham's panopticon, drawing attention to the role of CCTV surveillance in processes of social exclusion and in circumscribing the rights of individuals and groups to the public spaces of cities. Indeed, studies of the operation of CCTV surveillance have shown how networks of cameras are routinely used to target the homeless, street traders, young people, and ethnic minorities – all of whom are viewed as potential threats to the 'orderly city' and the economic success of urban regeneration initiatives. As a consequence, geographers have argued that under the gaze of CCTV surveillance cameras, any claims that streets symbolize public life with all its human contact, conflict, and tolerance are difficult to sustain. Increasingly, CCTV surveillance is portrayed as an instrument of social control geared towards preserving the public spaces of the city for the consumer citizen, while effectively excluding those whose spending power is low. Some social geographers, however, are more cautious in their assessment of the implications of public space CCTV surveillance, drawing attention to the ambivalent emotional geographies created by surveillance and in particular the ways in which CCTV cameras can make people feel both more secure and more fearful.

It is important to recognize, however, that there are important parallels between the operation of CCTV in public space and the surveillance activities carried out by private security guards responsible for policing areas of mass private property such as shopping malls and office space. Research by geographers on shopping malls, for example, has shown how the use of surveillance by private security teams is particularly aimed at making people who do not 'belong' (i.e., those of low spending power) feel uncomfortable. Moreover, any conduct that is viewed as discouraging the custom of other visitors – from illegal trading to begging and loitering – becomes a cause for intervention, with the threat of exclusion for those who persist in transgressing strict behavioral rules of shopping malls.

The approaches to policing social difference revealed by studies of ZTP, CCTV surveillance, and the actions of private security guards in shopping malls raise important questions about the accountability and responsiveness of police organizations to wider audiences. In the case of the private policing of shopping malls, the situation is relatively straightforward in the sense that security personnel are hired by mall owners and have a right to exclude whomever they wish from their property. With regard to the public police, however, the situation is much more complex and raises intriguing geographical questions about the extent to which local policing can reflect the needs and wishes of local communities. Drawing on a case study of West Seattle, Herbert's *Citizens, cops, and power* probes this issue in some detail. Idealized versions of community policing, Herbert contends, would have the police subservient to the community, but in reality, it is important for the police to retain some autonomy with respect to local demands, in part, to protect the interests of minorities against an unjust authoritarianism. Moreover, the dynamics of police subcultures and the fact that communities themselves are often divided over what they want from the police, significantly weaken the community policing project. For Herbert, this means that it is now time to abandon policies of community policing and look to other mechanisms to ensure that the public police are responsive to local concerns.

Beyond Local Policing: From Transnationalism to Cyberspace

As this review reveals, in a relatively short period, geographical interest in policing has developed rapidly and has forged important connections to wider debates within social and political geography. Nevertheless, the focus of geographical interest remains firmly on the local scale and, with only a handful of exceptions, the policing of urban environments. In this final section, therefore, it is important to signpost possible future directions for the geography of policing. One area of potential interest is transnational policing. Against a background of growing concerns about international terrorism, drugs trafficking, corporate fraud, and environmental crime, significant elements of policing can no longer be considered as simply a domestic function of national governments but as part of a realm of transnational practices. Such transnational policing is not, of course, a new phenomenon. Concerns about tackling cross border criminality were raised early in the twentieth century with the establishment of the International Criminal Police Commission or Interpol, although strictly speaking, this was an international rather than transnational initiative given that it was based on promoting mutual assistance between police organizations in separate countries. What has given significant impetus to transnational policing in recent years, however, is the convergence of several major global restructurings, including the opening up of global markets, the information revolution, and the end of the Cold War. Exemplifying the significance of these and other developments is the policing of the European Union. Terrorist activity in the 1960s and 1970s and the dissolution of internal border controls in the 1980s have all been important in creating structures for practical police cooperation, culminating in the establishment of Europol, in 1992, as a supranational police organization. Viewed as a potent symbol of the 'political virility' of the European Union because it embraces the identity of Europe as a single 'security community', the development of Europol is also inextricably bound up with fears that immigration and asylum, terrorism, and organized crime, pose a threat to the common interests and 'way of life' of the European Union. For geographers with interests in geopolitics, the example of Europol and other similar developments in transnational policing offer rich opportunities to examine the structures and processes of globalization from a fresh perspective.

A second area where geographers could potentially make a significant research contribution is in examining the policing of cyberspace and the problems of tackling cybercrime. The importance of this topic was illustrated in dramatic fashion in May 2004 when a Phillipino computer science student released a computer virus, known as the Love Bug, on to the Internet. Within 2 hours, it had spread around the world destroying the computer files and data of 45 million users in 20 countries at an estimated cost of $10 billion. Although the offender was identified by the FBI, the investigation was thwarted because, at the time, the Philippines had no cybercrime laws and the US were unable to extradite him because a crime has to be recognized in both the extraditing nation and the nation seeking extradition. For geographers, cases like this raise intriguing questions about the divergent legal responses to cybercrime in different countries and the implications this has for law enforcement in cyberspace. A further issue raised by such cybercrimes concerns the different roles and responsibilities of individuals and organizations in patrolling cyberspace. In addition to specialist units within the public police, there are Internet service providers (ISPs) who can draw up codes of conduct and remove offending material; private security firms with a role in protecting corporate data; grass roots activists, like Cyberangels who provide advice to protect people from becoming victims of cybercrime; and individual citizens who, as internet users, can report the presence of harmful material on websites or evict offensive participants from chat rooms. Mapping the contours and evaluating the impacts and implications of these different forms of surveillance in cyberspace have intriguing parallels with attempts by geographers to understand the local security networks responsible for policing terrestrial space.

See also: Postmodernism/Postmodern Geography; Public Space; Surveillance.

Further Reading

Flint, J. and Nixon, J. (2006). Governing neighbours: Anti social behaviour orders and the new forms of regulating conduct. *Urban Studies* 43, 939 956.

Fyfe, N. R. (1991). The Police, space and society: The geography of policing. *Progress in Human Geography* 15, 267 269.

Fyfe, N. R. (1992). Space, time and policing: Towards a contextual understanding of police work. *Environment and Planning D, Society and Space* 10, 469 481.

Fyfe, N. R. (2004). Zero tolerance, maximum surveillance? Deviance, difference and crime control in the late modern city. In Lees, L. (ed.) *The emancipatory city? Paradoxes and possibilities*, pp 40 56. London: Sage.

Fyfe, N. R. and Bannister, J. (1996). City watching: Closed circuit television surveillance in public spaces. *Area* 28, 37 46.

Herbert, S. (1997). *Policing space: Territoriality and the Los Angeles police department*. Minneapolis, MN: University of Minnesota Press.

Herbert, S. (2006). *Citizens, cops and power: Recognizing the limits of community.* Chicago, IL: University of Chicago Press.

Lowman, J. (1986). Conceptual issues in the geography of crime: Towards a geography of social control. *Annals of the Association of American Geographers* 76, 81 94.

MacLeod, G. (2002). From urban entrepreneurialism to a revanchist city? On the spatial injustices of Glasgow's renaissance. *Antipode* 34, 602 624.

Sheptycki, J. W. E. (1998). Policing, postmodernism and transnationalization. *British Journal of Criminology* 38, 485 503.

Smith, N. (1999). Which new urbanism? New York city and the revanchist 1990s. In Beauregard, R. A. & Body Gendrot, S. (eds.) *The urban moment: Cosmopolitan essays on the late 20th century city*, pp 185 208. Thousand Oaks, CA: Sage.

Smith, S. J. (1986). *Crime, space and society*. Cambridge: Cambridge University Press.

Wall, D. (2002). Insecurity and the policing of cyberspace. In Crawford, A. (ed.) *Crime and insecurity*, pp 40 56 and 186 209, Cullompton, UK: Willan.

Relevant Websites

http://www.police.homeoffice.gov.uk
Home Office Police.

http://www.worldcongresscriminology.com
International Society of Criminology.

Political Boundaries

A. Paasi, University of Oulu, Linnanmaa, Finland

Glossary

Artificial Boundary This concept is related to the idea that boundaries made by human beings are artificial if they do not correspond to natural boundaries (see below 'natural boundary'). The distinction between natural and artificial boundaries has been largely rejected by political geographers since World War II.

Border This denotes the adjacent areas lining boundaries. In the current interdisciplinary literature, this word is often used in a parallel sense to boundary.

Borderland The transition zone within which a boundary (line) is located.

Boundary This is one of the traditional keywords in political geography. Boundaries have been defined by tradition as physical or imaginary lines of contact between the key power containers, that is, states. Nowadays they are understood more broadly, often as social institutions and symbols that exist in social practices and discourses.

Boundary-Producing Practices An idea coined by representatives of critical geopolitics, practices which suggest that instead of mere empirical analysis of concrete boundaries it is important to study the social and political practices in which boundaries and their meanings as instruments of distinction are produced and reproduced. This challenges the views on states and their boundaries as fixed fulfillments of historical destiny.

De-territorialization This expression normally refers to the de-bordering of the world purported to be taking place mainly as a consequence of economic and cultural globalization (see below 're-territorialization').

Frontier A category that usually refers to political divisions (zones) between states or to divisions between settled and uninhabited areas within a state.

Natural Boundary This concept is related to the notion that nature dictates the 'correct' locations for boundaries and it is the task of the state to strive to achieve such boundaries. In a more practical sense, a natural border is used to denote physical elements such as rivers or mountain ridges.

Re-territorialization The process of remaking boundaries. A much-used example of such processes has been the violent boundary making in the formerly dispersed Eastern European states. This expression is often used together with de-territorialization.

Introduction

'Boundary' has been one of the keywords in political geography since the institutionalization of this subfield at the end of the nineteenth century. Although the words used in various languages to denote the idea of boundary (e.g., German *Grenze*, French *frontière*, Swedish *gräns*, or Finnish *raja*) may have slightly different connotations, this category can in general be regarded as the most important single 'mediating' keyword in the tradition of political geography. It is a mediating category in the sense that so many other key categories of political geography – and sets of social practices and ideologies – are implicitly or explicitly related to boundaries, for example, fundamental categories such as state, nation, nationalism, territory, territoriality, governance, or sovereignty. Therefore, the meanings associated with the concept of boundary crucially shape not only how these other keywords and their relations are understood but also how we understand the 'boundedness' of the contexts of politics and social life in modern states.

The importance and sensitive role of boundaries for human societies can easily be recognized when we think of the traditional nationalist ideal and its famous condensation by Ernest Gellner into the principle that political and national units should be congruent. The fact that the current world harbors some 200 states but perhaps 600–800 groups of people identifying themselves as 'nations' clearly displays the vagueness of this ideal and undermines the concept of a neat one to one relation between state and nation. A number of human geographers, political scientists, and anthropologists have challenged this ideal both theoretically and empirically, especially since the 1990s.

A particularly insightful theoretical critique was provided by the political geographer John Agnew, who scrutinized the key dimensions of what he called the 'territorial trap', that is, how state centric thinking establishes a dividing line between inside and outside. The territorial trap rests on three assumptions that the sovereignty, security, and political life of the modern state require a bounded territorial space, that there is a fundamental opposition between the internal and external affairs of a state, and that the territorial state functions as a geographical container for modern society, that is, state boundaries are coincident with the boundaries formed by political and social processes. This way of thinking suggests that the world is composed of firmly delimited, mutually exclusive territories all having their own collective identity.

One important ideological manifestation of the territorial trap is that national cultures and identities are often presented as homogeneous, coherent phenomena, so that territory and mutual exclusivity become naturalized elements in the definitions of culture. The identities of population groups are not 'natural', however, but are always created consciously for particular purposes. This holds good for smaller regional communities too, but it is most obvious in the case of states, which usually have developed complex institutions (military forces, an education system, national symbols) that produce and reproduce nationalism and national identity narratives. Political boundaries between states often manifest nationalist feelings, symbolism, or iconographies and performance. This suggests that boundaries are not merely neutral geographical lines but crucial elements in what is generally understood as a national unit. The number of perpetually ongoing more or less open border disputes shows that such lines, both inside and between states, may be deeply contested (**Figure 1**).

All political boundaries, therefore, require some selection and are inevitably the results of negotiation and reflect power relations. Practically, all boundaries made by human beings – boundaries that in some sense or other divide the communal living space on various spatial scales – are political, giving expression to power relations. It is precisely due to this fact that boundaries are open to dispute. The importance of power relations is reinforced by the fact that boundaries involve the politics of delimitation, representation, and identity. They keep things apart, their meanings are expressed in particular terms, and they allow certain expressions of identity to exist, while often blocking others.

Recent empirical objections against taking state–nation relations for granted often accentuate the challenges raised by the radical increase in border crossings and transgressions of all kinds, whether by human beings (tourists, immigrants, refugees), ideas, cultural influences, or capital that occur on all spatial scales as a consequence of cultural, political, and economic globalization. Political boundaries are therefore seen in an integrating world as being increasingly multifaceted and complex. This fact will doubtlessly arouse both contextual sensitivity and broader theoretical understanding on the part of human geographers.

This article looks at the tradition of boundary studies in political geography, introduces current perspectives, and discusses the future challenges raised by political geographers during recent decades. Political boundaries exist on all spatial scales, from local electoral or administrative units to states and suprastate arrangements like the European Union (EU). Nevertheless, it is state boundaries that have received most attention in political geography, and these will be scrutinized here. The current world harbors some 300 land boundaries between states, and also more than 40 international sea boundaries. Due to the great variety of boundaries and the strong empirical tradition, political geographers such as Victor Prescott have suggested that all boundaries are unique and therefore the development of general theories on political boundaries has failed. In spite of such comments, and in spite of popular ideas on the rise of a 'borderless world' and the emerging relational critiques in human geography that challenge all kinds of bounded spaces, boundaries have become notable objects of research since the early 1990s. One aim of this article is

Figure 1 Banal, border-related nationalism performed in front of the national audiences on the India-Pakistan border. Photo by Anssi Paasi.

therefore to show why and how boundaries are recognized as perpetually important theoretical and empirical categories by human geographers. It is suggested that the connotations and functions of boundaries can be understood meaningfully only in relation to broader categories such as state, nation, and territory, and that this conceptual constellation can also help us to understand the persistence of boundaries in the current phase of globalization. It will also be shown that boundaries are significant not only for political geographers but also for many scholars working in fields such as regional, cultural, and economic geography.

The Tradition of Boundary Studies in Human Geography

While boundaries seem to mediate the relations between a number of important categories in political geography, boundary studies have for their part sometimes been labeled as one of the most torpid areas of inquiry in this field. This is based on the fact that such studies have often had a very strong empirical, even empiricist focus, and have thus been rather descriptive, atheoretical investigations into concrete, material border landscapes.

Since boundaries have been by tradition understood mainly as empirical manifestations of state sovereignty, political geography has been rich in morphological, empirical, and generic studies of boundaries. Early studies in particular concentrated on constructing typologies and classifying boundaries into types. Prescott's profound review shows that boundaries have been regarded for a long time as unique phenomena, generalizations about them have not been regarded as valuable, and the continual search for means of classifying them has not been seen to be profitable. While some scholars such as Julian Minghi were already claiming during the 1960s that it is important to search for generalizations in boundary studies, it would be no exaggeration to suggest that the empirical orientation prevailed from the beginning of the twentieth century until the 1980s. Such an orientation was originally linked to the more general understanding of geography as an empirical field of research and also to the fact that boundary scholars often had very practical motives for their research.

The impact of boundaries upon landscapes marked a further step away from simple description and categorization of borders. Studies on the functional roles of borders and transboundary interactions expanded the geographical approaches. Much current work – especially applied research – still follows this logic. The cultural roles of borders received more attention in geography only at the beginning of the 1990s, a couple of decades later than in cultural studies, but political geographers were still mostly concerned with concrete border landscapes and interaction across borders.

The rest of this section looks at the roots of boundary studies in political geography, the debates on 'natural boundaries', the emergence of a specific language in these studies, and finally the resurgence of boundaries as objects of geographical research that has taken place since the end of the Cold War.

The Roots of Boundary Studies

Boundary studies have a long tradition in geography. Friedrich Ratzel's *Politische Geographie*, the first major book in political geography, already included a long and profound analysis of the versatile functions and dimensions of boundaries, in which he discussed themes such as 'natural' and 'artificial' boundaries, boundaries as results of mobility and the meanings, and roles of boundaries for human societies. Ratzel regarded the margins of the territory of a state as important elements in the same way as the central regions: the strongest states showed, he suggested, close ties between the border and the core. Boundaries were for him an expression and measure of state power and as such he regarded them as dynamic rather than static elements. Ratzel believed in his organistic thinking that all 'vigorous states' try to expand in spatial terms while declining states contract to physically easily defensible land contours. Boundaries are in this sense merely a 'stop' in mobility. In Ratzelian thinking, the border fringe (*Grenzsaum*) is the reality and the border line (*Grenzlinie*) is its abstraction. The border line as such is a supporter of the imagination (*Vorstellung*) that helps our thinking. He also recognized that there is something symbolic in all boundaries. They can be symbols of peace, for example, or symbols of historical events.

Even before Ratzel, much practical and political discussion has taken place on the role of boundaries and what should be the ideal locations for them. Historically, one of the most influential philosophical notions emphasizing the connection between a territory and its natural environment and culture (in contradistinction to other territories) has been the French concept of *les limites naturelles*, according to which every state has its 'natural boundaries' which is natural and justifiable for that state to pursue and to establish in concrete form. As shown by Norman Pounds, the idea of natural boundaries was originally an appeal to historical arguments, and it was only later that recourse was had to a higher authority, 'natural law'. Where the historical arguments were usually of a specific character, however, and were connected with one given boundary, natural law affirmed the obligation to honor natural boundaries in principle without providing any actual guidance on how these should be identified. The significance of the 'will' of nature in this abstract sense, the notion that it 'wants' to regulate the actions of nations and to prevent them from becoming mixed (while the state *tries* to do this by means of violence or cunning) attains prominence in the *Eternal Peace* of the philosopher Immanuel Kant, published in 1795, but it is found equally well in the arguments of Karl Haushofer, the major propagandist of German geopolitics between the two World Wars.

A comparable case in the field of culture would be the emphasis placed on 'language' as a 'natural boundary'. Again Kant, in the work mentioned above, is among those who cites language alongside religion as a manifestation of 'the will of nature' to keep the peoples apart and prevent them from mixing. A typical example of a definition analogous to a natural law, although undoubtedly setting out from the people (*Volk*), is the notion put forward in the early nineteenth century in connection with the strengthening of German nationalism, that language constitutes the only true natural boundary. This is based on the idea that God created the differences between languages and therefore a common language and culture constitute a natural law of a higher order than natural boundaries formed by rivers or mountains. This emphasis on language and *Volk* as a major starting point for determining boundaries continued to enjoy prominence in Germany throughout the nineteenth century, and the cartographical presentation of linguistic boundaries gained in popularity from the 1840s onwards. This adoption of cartographic methods in effect drew attention to the organic connection between cartography and power, a matter which has been widely emphasized in geographical circles since the late 1980s.

Such debates were thus not only academic but were also crucially associated with questions of war and peace. Natural boundaries were often discussed in relation to what were labeled as 'artificial boundaries', boundaries made by human beings. As this division implies, natural boundaries were regarded as something genuine and original, whereas artificial boundaries were regarded as temporary and haphazard. Such ideas were raised in many countries not only in the geographical literature but also in the sphere of practical geopolitics before and during World War II, in Germany, for example, and in Finland, which was allied with Germany during the years 1941–1944. A number of Finnish geographers and other scientists were involved in efforts generated by statecraft to show that the existing Finnish Soviet border was not a natural one and that the real, natural border was to be found much deeper into Soviet territory. Since Finland had occupied a slice of Soviet territory at that time, these claims were strengthened by the extensive fieldwork carried out in the occupied areas by a number of scholars ranging from natural scientists to linguists. Some of them were engaged in mapping the *Lebensraum* (living space) of Finland (and also used this terminology in their scientific publications), thus adopting the language developed originally by Ratzel and later exploited by the Nazis. After the war, the occupied areas and almost 10% of the previous Finnish territory were ceded to the Soviet Union and such geopolitical speculations came to an end.

World War II and the unscrupulous use of geopolitics finally altered the situation elsewhere as well, and both geopolitics and the expansionist boundary terminology were largely abandoned. Since the war, most political geographers have rejected ideas on natural boundaries and have suggested that all political boundaries are human, social constructs in the sense that they are based on human motivations, decisions, and power relations. This also holds good for boundaries that follow rivers, mountain ridges, and coastal areas, as there is nothing natural in them as political boundaries and instruments of social control. In principle, boundaries are therefore nothing more than lines or vertical levels drawn to demarcate state sovereignty, and it is irrelevant what types of terrain or sea areas they pass through. This does not mean that the location of boundaries is as such irrelevant, since these considerations can be important for the fortunes of a state in military conflicts, which often take place in border areas.

Ratzel's discussion showed effectively that boundaries are often associated with another key category of political geography: territory. Indeed, Robert David Sack has defined territory and the associated practice of territoriality partly through the concept of boundary. For him, territoriality is a strategy for affecting, influencing, or controlling resources and people, and this takes place by controlling an area. Circumscribing things in space, or on a map, as when a geographer delimits an area to illustrate where corn is grown or where industry is concentrated, identifies places, areas or regions in an ordinary sense, but it does not itself create a territory. Sack suggests that this delimitation becomes a territory only when its boundaries are used to affect behavior by controlling access. Territoriality is therefore a strategy, not something natural, primordial, or unchanging. And as a strategy it can be turned on and off.

Boundary Terminology and Its Problems

The first half of the twentieth century was significant for the development and establishment of the terminology of border studies in political geography. This language was partly associated with the emergence of the state as the key institution of modernity and the fact that the institutionalization of geography as an academic discipline was crucially related to this process. It has been noted that the key ideological motives for the institutionalization of geography as an academic discipline were nationalism and imperialism. Such ideologies are very closely linked to the processes of dividing space and to power relations. Political geographers have observed that it was at the beginning of the twentieth century that the all encompassing nation state system finally transformed a world that had been characterized by more or less flexible and permeable frontier zones into a world of relatively fixed boundaries, a world that gradually manifested itself in the emergence of guarded boundaries, passports, and other instruments that were needed for controlling state territory. This also manifested itself in maps and atlases, which became increasingly important elements in the state centric national socialization that took place in schools and other places of education. In a word, boundaries were 'drawn' on the Earth's surface, on maps, and in people's minds and social practices in order to produce and reproduce the territorial trap.

The key categories of the emerging modernist language in political geography were border, boundary, and frontier. These categories are still in use today, although their meanings seem to vary between academic fields. In political geography, Victor Prescott in particular struggled to define these categories in relation to each other. He has defined a boundary as a line of physical contact which affords a basis for both discord and opportunities in cooperation between states. A border is a slightly broader category in this terminology and denotes the adjacent areas which line a boundary, while a frontier is a zone that is associated either with a political division between states or with a division between the settled and uninhabited areas within a state. One more important category is Prescott's term 'borderland', which is normally understood as a transition zone within which a boundary line lies. We can perhaps add to this list the concept of a 'border landscape', which in a way condenses the many elements of traditional empirical approaches: a local perspective, an empirical approach, and attention paid to visual elements. Border landscapes have also been objects of research in many boundary studies focusing on ordinary people's perceptions of a border.

It is important to recognize that this division between the concepts is by no means the only possible one. Hence the current literature published in anthropology or political science – and even political geography – displays quite free usage of these terms, so that the scholars may be talking about borders at one time and about frontiers or boundaries in a more or less overlapping sense at other times.

This ambivalent use of concepts is related to the fact that since the collapse of the Cold War divide between the Western and Eastern world, boundary studies have been mushrooming not only in political geography but also in many other fields, so that scholars engaged in research in international relations, anthropologists, historians, sociologists, and literary theorists, for instance, have all been looking at the concrete, symbolic, and metaphorical meanings of boundaries. One problem in the current boom has been that researchers working in different 'academic territories' often operate inside their own territory, without noticing that more or less similar work is being done beyond the 'boundaries' of these territories. This crossing of disciplinary boundaries poses a perpetual challenge for border scholars.

The existence of divergent views on the roles of boundaries gives expression to the fact that scholars have a wide variety of theoretical frames of reference available for their work. At the same time, boundaries themselves serve a great diversity of functions: they are administrative tools in the international system of nation states, they are instruments of the political and territorial control exercised by governments, and they also act as structural factors in existing patterns of social identities and/or challenge these patterns. In practice, such functions are in operation simultaneously.

The meanings attached to boundaries are always contextual and historically conditional. They are not only infused with ideological significance, but also have material practices and consequences associated with them. Most state boundaries are emotionally loaded with bloody memories and recollections of moments of glory that have been achieved or have been left behind. Such memories and related performances are effectively exploited in nationalist ideologies. Like the various forms of nationalism, boundaries are loaded with implications of power, politics, and culture. They are part of the means by which people try to create order in the world on different spatial scales. It is for this reason that boundaries are frequently matters of dispute, so that actors – both individual and corporate – imbued with a multiplicity of passions will attempt to contest their existence.

The Resurgence of Boundary Studies after the Cold War

The post World War II situation in Europe in particular and the period following the Cold War very much fixed the meanings of boundaries in this context, that in which much of the language and practice of border studies had originally been created. A certain status quo as to boundaries was regarded as useful during the Cold War period for reasons of security policy. This claim is still implicit in the EU, which has required from new members that they should not have unsolved border disputes. The EU has now taken more active steps to reduce the impact of boundaries, for example, through a cross border regional policy. While this is taking place inside the EU, its external border has concomitantly become more exclusive.

After World War II new political boundaries emerged, particularly in the continent of Africa, where a number of new states were established after the end of the colonial period. These tendencies did not arouse much theoretical motivation among border scholars, however, for although they have always recognized the crucial link between boundaries and state power, they have not by tradition been apt to theorize over the relations between the state, the nation, nationalism, and boundaries in terms of social and cultural theory. Such theorization represents a relatively recent phenomenon that has emerged since the 1990s.

It was therefore not until the collapse of the Cold War geopolitical order and the disappearance of the dividing line between West and East at the beginning of the 1990s that a new wave of border research was experienced in political geography and many other fields. The resurgence of boundary studies in this new situation manifested itself not only in academia and in the increasing number of boundary related publications, but also in the rapid establishment of border research institutes in many countries around the world. Contrary to purely academic border studies, such institutes have often had concrete aims, for example, those of creating and maintaining information on various border cases.

The end of the Cold War both created new boundaries and removed some very important existing ones. The dispersal of the Soviet Union and Yugoslavia in particular gave rise to a number of new boundaries and territories that claimed sovereignty. In contrast, the end of the long divide between the two Germanys removed the key symbol of the Cold War, the Berlin Wall, and transformed it – or what was left of it – into a new relict boundary that was soon to become a tourist attraction (**Figure 2**).

Alongside its many other local consequences, the removal of the East West dichotomy meant in the broader constellation of international relations that both the West and the East lost the 'Other' by contrast with which they had built up their own identity. 'Security' has traditionally been linked with national political identities, which have been perceived as dependent on boundaries defining the Other. In this new situation, the foreign policy élites of many countries developed a need to redefine the core content of their territoriality, their foreign, and defense policies – and naturally also their identity. This became part of a wider rescaling of the state in the globalizing world. In Europe the changing landscapes of security led many of the former Eastern European states to join NATO. The international landscape then experienced another major upheaval in the form of the terrorist attacks on New York on 11 September 2001. New lines of integration and separation were established and new images of threat were exploited in order to do this. Such new lines and Others have often been drawn on the basis of Samuel Huntington's famous but much criticized ideas on the clash between 'civilizations', especially in the case of the line between the Western and Islamic cultures.

Examples of New Themes and Debates

The ongoing social transformations and new interdisciplinary theoretical orientations have fuelled new debates on boundaries. The mushrooming of discussion over the impacts of globalization, for example, has created whole new fields of discourse. The most significant as far as the future roles of international boundaries are concerned has been the debate on the disappearance, or at least retreat, of the nation state and the related ideas on a new 'borderless world' that were put forward after the mid 1990s. The borderless world thesis is associated with the name of the Japanese business guru Kenichi Ohmae, who looked on the networks of 'region states', often emerging on the border regions of larger states (e.g., HongKong, Tijuana San Diego) as the potential motors for the new global economy. Another important figure motivating the debates on the new roles of political boundaries, de/rebordering, and de/re territorialization has been Manuel Castells, whose works on networks and on the space of flows that will challenge the space of places have forced researchers to rethink the contemporary and future roles of national states. Castells has nevertheless perpetually reaffirmed his recognition of the power of territorial identities in social life and in the struggles that occur on various spatial scales.

While the borderless world was perhaps more a metaphor of neoliberalist economic thinking than something that was intended as a universal characterization of the current globalizing world, Ohmae was on the right track when recognizing the importance of cross border regions. It is no exaggeration to argue that questions regarding border crossings form one of the key areas in boundary studies today. Contrary to many other themes in political geography, boundary studies concentrating on the theory and practice of border crossings are practiced actively outside the Anglophone world, especially on the continent of Europe, where integration and the enlargement of the EU, together with the new cross border policies developed in this context, have made the political boundaries between states very important objects of research and practical politics. Cross border cooperation and new forms of regionalization are leading the EU to an increasing degree into a situation in which, to use the expressions of Oscar Martinez (1994: 1–15), many "alienated borderlands" are turning into "interdependent borderlands", and may eventually integrate to form a single region. More than 150 cross border regions have been created in the EU with the aim of lowering the boundaries between states and promoting the economic and cultural life of border

Figure 2 The Berlin Wall was the key symbol of the Cold War period in Europe. After the collapse of the socialist Eastern Europe the remnants of this border became a famous tourist attraction. Photos: Anssi Paasi.

areas. Some of these units exist between two states and others serve to integrate a number of states.

In spite of these tendencies and the change of focus toward European borders, perhaps the most iconic single cross border region in the world is nevertheless still the US Mexico border, which has been studied intensively by geographers, political scientists, and even literary theorists. The terrorist attack in 2001 has raised the importance of this border to new heights, since it perpetually symbolizes many problems and research questions that are normally associated with international political borders: a major gap in the standard of living between two states and the resulting exploitation of the labor force on the poorer side, questions of immigration, smuggling, prostitution, or strict surveillance. Similar problems are increasingly being experienced on the new external boundaries of the EU, however, causing the expression 'Fortress Europe' to be adopted to refer to the efforts by the EU to keep non EU nationals, goods, or businesses out of the member states. Although the connection was perhaps unintentional, the use of this expression is grotesque in the sense that the Nazi leader Heinrich Himmler frequently referred in his days to 'the fortress of Europe'.

It was thought in the 1990s that the rise of new information technologies, the Internet, and cyberspace would dramatically change the nature of bounded spaces and somehow reduce their power. In many cases such technologies have doubtlessly helped, for example, in enabling members of diasporic communities to maintain connections across borders. Similarly, social movements based on gender questions, sexuality, environment, or globalization/antiglobalization have certainly benefited from the new technologies. But at the same time, it is obvious that they have become effective instruments of boundary control and maintenance around the globe. This tendency toward surveillance and security in daily life has led to a situation where boundaries – in the form of control – actually exist not only in border areas but are also distributed all over our societies: in airports, shopping precincts, or the streets of our city centers. This has taken place especially after the 11 September 2001.

Human geographers and related scholars have also developed other conceptual ideas in the context of de bates on globalization that imply borders or border crossings in the spheres of politics, economy, or culture, examples of which might be 'ecopolitics' in international relations (IR) or 'annihilation of space' in economic geography. Ecopolitics reminds us of the fact that many environmental problems (e.g., the Chernobyl catastrophe in 1986) do not respect the boundaries drawn by polit icians and military officers. Annihilation of space, for its part, points to the fact that capital accumulation perpetually strives to annihilate space (and boundaries) and to reduce turnover time.

A further, more theoretical, inspiration for new aca demic debates on borders in the 1990s was the purported 'postmodern condition' and the post structuralist phil osophy that both challenged the existence of permanent truths and at times accentuated the socially constructed character of the world and its nature as a constellation of all kinds of networks and interactions. Such perspectives questioned the whole idea of fixed boundaries and power relations and emphasized the fragmentary, hybrid, and transforming character of the world. This way of thinking is still extant in a modified form in human geography, where the normative pronouncements of the 'relational thinkers' have challenged all bounded spaces as re gressive and put the stress on hybridity, fluidity, and networks. Similar thoughts are often advanced by border scholars who draw on cosmopolitan thinking.

It is clear that in this new situation different opinions exist on the alternative futures for boundaries. Many researchers suggest that national states and their boundaries have lost much of their significance in the world, in which the new order dictated by geoeconomics and information economics will give priority to mobility, speed, flows of various kinds, and an entirely new type of economic region. Some other scholars maintain that national states and their boundaries will continue to be significant instruments of governance in the international system, but that this will occur in new, less static ways. A third opinion emphasizes that both the nation and the state will continue to be of significance but the nation state itself will be in trouble. This idea is based on an observation that has become more obvious as a con sequence of ethnonationalist activism, immigration, and displacement, namely, that a 'nation' is not inevitably restricted to a given territory but can exist on both sides of a border or require bounds of its own that are not congruent with existing state boundaries at all. The 'nation' itself is recognized as a social and ideological construct and a product of nationalism that often draws on ideas of bounded territories to maintain the 'territorial trap'.

From Boundaries to Boundary-Producing Practices

In this ambivalent situation characterized by social transformations and new theoretical ideas, political geographers have been paying increasing attention to boundary producing practices as an alternative to the previously dominant descriptive approach and the analysis of unique boundaries, border landscapes, and cross border activities. This line of thought was set in motion especially by the representatives of critical geo politics who were following the simultaneous debates in the field of IR. The IR scholars had challenged the assumptions hidden in the division between the internal and external aspects of states during the late 1980s. Simon Dalby and Gearoid O'Tuathail suggested that critical geopolitics should pay particular attention to the boundary drawing practices and performances that characterize the everyday life of states. They pointed out that, in contrast to conventional geography and geopolitics, this approach should investigate both the material borders at the edges of states and the conceptual borders that designate material boundaries between a secure interior and an anarchic exterior.

The processes of 'Othering' and how such processes become a part of banal or mundane nationalism have been scrutinized in many IR and geographical studies. Paasi, for instance, examined how the Finnish Russian border has been exploited ideologically in Finland in the nation building process as a line separating 'us' from the Other (the Soviet Union/Russia) and how this ideology was used effectively – through the media and the national education system – for the purposes of 'spatial social ization'. This is the process through which individuals and collectives are socialized as members of specific territorially bounded spatial entities and through which people more or less actively internalize collective identity narratives and shared traditions and memories.

New theoretical approaches have motivated political geographers to become increasingly interested in social and cultural theory when developing new interpretations and approaches to boundary studies. Newman and Paasi looked at the theoretical literature and empirical studies on boundaries and identified four interdisciplinary themes in border studies in which human geographers also had an active interest. First, the emerging debates on postmodernism and post structuralism and the related debates on de territorialization and re territorialization on various spatial scales inspired many geographers in their research. Second, the roles of diverging borders in the construction of sociospatial identities and in the creation of distinctions between 'us' and the Other were studied not only in the field of IR but also by political geographers. Third, increasing attention was paid to boundary narratives and discourses that were identified as significant in the construction of spatial identities and in the national socialization process, that is, in the process whereby citizens adopt the narratives of national identities. Correspondingly, geographers have studied various mechanisms that generate boundedness and have identified the important role of national education systems and the media in such processes. Fourth, political geographers became interested in the roles of various spatial scales in the construction of national borders. This theme shows the importance of boundaries as a meeting place for local, national, and international processes, and in fact displays the problematic character of such universally accepted scales in social life.

Indeed, due to the complexity of the spatialities of the globalizing world and the new theoretical orientations in boundary studies, it is possible to widen the horizon on boundaries and argue that it is not only the subfield of political geography that deals with them but also certain other areas of human geography. The boundaries displayed in **Figure 3** are ones that are implicitly or explicitly present in many other areas of geographical inquiry in addition to political geography, most notably in economic, cultural, and regional geography. The figure brings together some traditional and more recent 'key words' in these fields, words associated with a bunch of social practices that imply bordering or border crossings, or else simply challenge the notion of bounded spaces. Unfortunately, economic, political, cultural, and regional geographers tend to work on these border related questions without knowing very much of what scholars in other subfields of the discipline are doing. This is clearly a challenge for border research within human geography.

This complexity means that instead of taking any of the boundary related categories for granted, one important task for border scholars is to conceptualize the categories in relation to each concrete research setting and to create abstractions that will help researchers to reveal relevant social, political, and experiential dimensions of *bordering* in such a context. This will require the analysis of relevant modalities as to how the boundary in question exists, what are its social and emotional meanings, and how such meanings have been constituted in social life.

Coda–Boundaries Are Everywhere

This article shows that human geographers have studied political boundaries intensively since the institutionalization of academic geography and that their interests

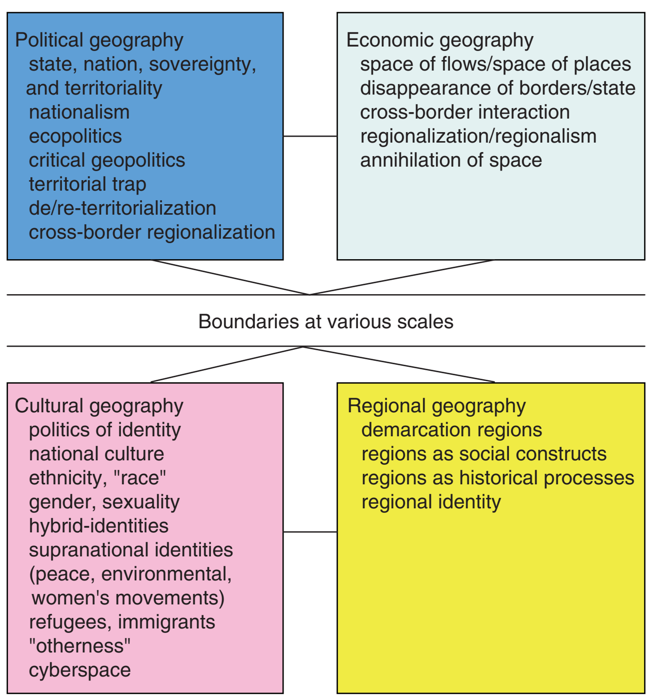

Figure 3 Some keywords in political, economic, cultural, and regional geography that are related to boundaries on various scales. Reproduced from Paasi, A. (2005). Generations and the 'development' of border studies. *Geopolitics* 10, 663 671.

have gradually moved from empirical studies of concrete cases and border landscapes to broader themes, for example, the mapping of divergent boundary producing practices and the roles of boundaries as social symbols and institutions that are used to make social distinctions within societies. Political geographers understand boundaries nowadays in a much broader manner than as mere physical lines that separate social entities or power containers, especially states. More often than not boundaries are conceived as discursive formations and processes that have grown up to form parts of social and cultural practices that constitute and are constitutive of territories.

Such broader perspectives mean that notions such as territory, territoriality, sovereignty, nation and nationalism, nationality, and identity are crucial for understanding the meanings and functions of boundaries and the power relations associated with their formation and reproduction. Boundaries can therefore be understood as part of the process by which territories and their identities and meanings are formed and renewed. In the course of this process, which may be labeled as the institutionalization of territories, effective use is made of boundaries, symbols, and the institutions that maintain them in order to produce and reproduce a territory.

Despite the fact that concrete boundaries and physical border landscapes are often loaded with national symbolism, we can think that boundaries exist at the same time in the innumerable practices and discourses that constitute the sociospatial delimitations that place some individuals inside a certain boundary and others outside it. This broader perspective means that boundaries are manifested – either visibly or covertly – in politics, governance, economics, culture, the construction of ethnic relations, the media, educational practices, and other forms of national socialization. It is these practices that lay the foundation for our awareness of the sovereignty of states, of 'us' as an abstract, imagined community with a specific identity, and for our understanding of the relations between 'us' and 'them'. Such practices intersect and become fused across spatial scales. Some practices may be local, some national, and others international, which means that both the origins and manifestations of a boundary may exist simultaneously on several spatial scales.

This article also shows that political boundaries are not just neutral lines drawn on maps or through concrete border landscapes, but are actually spread all over state territories, and they can even manifest themselves at times outside states. One example of such an extension is the fact that the UK and French passport controls for the Eurostar trains are juxtaposed in Paris and London, respectively. This extension, like the increased monitoring and surveillance at national airports, in shopping precincts or in city streets, is part of the 'technical' landscapes of social control that have been created in the name of the often abstract discourse of security.

There is another way, too, in which boundaries exist more broadly within society. This is based on the fact that they are at the same time both symbols and institutions through which social groups generate distinctions and are themselves created through such distinctions. In this sense boundaries are a part of what can be called the 'discursive' landscape of social power, which is manifested in social practices. All states lean on such landscapes, drawing on both the past heritage, current symbolisms, and future societal expectations, when they try to produce and reproduce loyal citizens who will identify themselves with the nation, or at least with the political system.

Both the technical landscapes of social control and the discursive landscapes of social power makes it easier to understand the persistence of boundaries in the networking world and to comprehend why boundaries will not necessarily disappear even though certain social practices, for example, the economy, may produce alterations in the ways in which boundaries function. As Peter Taylor has shown, territoriality may have many shapes in the operation of nation states. Instead of one territoriality we have numerous territorialities that may have divergent functions. This suggests that human geographers working with political boundaries will continue to face a number of interesting theoretical and empirical challenges in the future.

See also: Borderlands; Cold War; Maps and the State; Nation; Sovereignty; State; Symbolism, Iconography; Territory and Territoriality.

Further Reading

Agnew, J. (1994). The territorial trap: The geographical assumptions of international relations theory. *Review of International Political Economy* 1, 53 80.

Anderson, M. (1996). *Frontiers: Territory and State Formation in the Modern World*. Cambridge: Polity.

Biger, G. (1995). *The Encyclopedia of International Boundaries*. New York: Facts on File.

Castells, M. (2004). *The Power of Identity*. Oxford: Blacwell.

Gellner, E. (1983). *Nations and Nationalism*. Oxford: Blackwell.

Martinez, O. J. (1994). The dynamics of border interaction: New approaches to border analysis. In Schofield, C. H. (ed.) *Global Boundaries: World Boundaries Series*, vol.1, pp 1 15. London: Routledge.

Minghi, J. (1963). Boundary studies in political geography. *Annals of the Association of American Geographers* 53, 407 428.

Newman, D. (2006). The lines that continue to separate us: Borders in our 'borderless' world. *Progress in Human Geography* 30, 143 161.

Newman, D. and Paasi, A. (1998). Fences and neighbours in the postmodern world: Boundary narratives in political geography. *Progress in Human Geography* 22, 186 207.

O'Tuathail, G. and Dalby, S. (1998). Introduction: Rethinking geopolitics: Towards a critical geopolitics. In O'Tuathail, G. & Dalby, S. (eds.) *Rethinking Geopolitics*, pp 1 15. London: Routledge.

Paasi, A. (1996). *Territories, Boundaries and Consciousness: The Changing Geographies of the Finnish Russian Border*. Chichester: Wiley.
Paasi, A. (2003). Boundaries in a globalizing world. In Anderson, K., Domosh, M., Pile, S. & Thrift, N. (eds.) *Handbook of Cultural Geography*, pp 272 462. London: Sage.
Paasi, A. (2005). Generations and the 'development' of border studies. *Geopolitics* 10, 663 671.
Pounds, N. J. G. (1951). The origin of the idea of natural frontiers in France. *Annals of the Association of American Geographers* 41, 146 157.
Prescott, J. R. V. (1965). *The Geography of Frontiers and Boundaries*. Chicago, IL: Aldine.
Ratzel, F. (1925). *Politische Geographie* (orig. publ. 1897). Munchen und Berlin: Von R. Oldenbourg (Dritte Auflage).
Rumford, C. (ed.) (2008). *Special issue: Citizens and Borderwork in Contemporary Europe. Space and Polity* 12, 1 156.
Rumley, D. and Minghi, J. (eds.) (1991). *The Geography of Border Landscapes*. London: Routledge.
Sack, R. D. (1986). *Human Territoriality, Its Theory and History*. Cambridge: Cambridge Univesity Press.
Taylor, P. J. (1994). The state as container: Territoriality in the modern world system. *Progress in Human Geography* 18, 151 162.

Relevant Websites

http://www.dur.ac.uk
International Boundaries Research Unit, Durham University.

Political Ecology

R. P. Neumann, Florida International University, Miami, FL, USA

Glossary

Environmental Degradation A catchall term for a range of ecological problems such as desertification, soil erosion, deforestation, and species loss.

Environmental Orthodoxies Highly generalized and simplified explanatory narratives of ecological change that are widely accepted and persist over time despite the accumulation of empirical evidence that would overturn or modify them.

Homeostatic Systems Systems that are self-maintained and self-regulated through complex control mechanisms, energy pathways, and feedback loops in a way that resists change and results in a stable equilibrium state.

Indigenous Technical Knowledge (ITK) Indigenous technical knowledge, a concept emerging out of participatory development initiatives of the 1970s that stressed the importance and relevance of local environmental knowledge and land-management techniques to development planning.

Neo-Malthusian The extension of Thomas Malthus's theory of the relationship between population growth and food scarcity to contemporary environmental problems, positing that overpopulation is the sole or primary cause of environmental degradation.

Social Relations of Production A term from Marx's theory of historical materialism that refers to the types of societal structures and relationships that develop from particular production processes, including socioeconomic stratification, property ownership and control, and patterns of socioeconomic dependency.

Introduction

Political ecology is a relatively new interdisciplinary field that developed initially through the merger of cultural ecology with political economy. Cultural ecology, with its emphasis on ethnographic fieldwork in small, agrarian communities shaped political ecology's early emphasis on rural, Third World settings. The central concern of cultural ecology in the 1950s and 1960s was to introduce the theoretical constructs of biological ecology, particularly the ecosystem concept, into the study of nature–society relations. A key assumption was that ecosystems are characterized by homeostasis and equilibrium. The homeostatic model of ecosystems held great appeal to anthropologists and geographers interested in demonstrating that cultural adaptation was the key to understanding complex nature–society interactions. Culture was theorized as playing the same role in human society that genetics plays for nonhuman species in evolution and adaptation to the environment.

Scholars began to question the conceptual and philosophical foundations of cultural ecology, noting that even the most remote communities did not function as independent, autarkic systems. A series of studies in the 1970s and 1980s refocused cultural ecology toward an analysis of how agrarian communities dealt with the social and ecological demands of an external capitalist economy. Critics explained that human behavior, reframed in cultural ecology as 'adaptation', is rendered crudely functionalist and teleological. Nature–society relations, it was argued, are better understood through political economy, with a focus on the social relations of production under capitalism. This reorientation toward investigating the social relations of production was guided by broadly Marxist political economic scholarship in geography, social history, and agrarian studies.

Development studies provided another root of political ecology. Specifically, within development studies in the 1960s and 1970s, there was a major focus on Third World environmental degradation. In a critical response to neo Malthusian explanations of environmental degradation, scholars interrogated their empirical and ideological bases. Neo Malthusianism, scholars argued, is overly simplistic, empirically unsubstantiated, and theoretically barren, ignoring the social relations, economic constraints, and political power structures that shaped land and resource use. Their alternative explanation was founded on the principles and concepts of political economy, including questions about wealth distribution, social patterns of accumulation, interclass relations, the role of the state, patterns of land ownership, and control over access to natural resources.

Out of this political economy approach to environmental degradation emerged a set of methods and questions that provided the foundation for a new field. Although the term had been used in other contexts, two geographers, Harold Brookfield and Piers Blaikie, are widely recognized for initially defining political ecology as linking ecology with political economy in their 1987 book on land degradation. A central premise was that ecological problems were at their core social and political problems, not technical or managerial, and therefore demanded a theoretical foundation for analyzing the complex social, economic, and political relations in which environmental change is embedded. Because of the complexity of the phenomena under investigation and the interdisciplinary character of the field, there is no single methodology or set of theoretical concepts that could be used to define political ecology research. Nonetheless, political ecology research does have distinguishable features. Multiscalar analysis has been a hallmark of political ecology, as have political economic analysis, historical analysis, ethnography, discourse analysis, and ecological field studies.

Theoretical Developments

Political ecology emerged and matured as a distinct field in a period of revolutionary theoretical shifts in both the social and biophysical sciences, which challenged many widely accepted ideas of nature–society interactions. Post structural social theory and nonequilibrium ecology have initiated a reexamination of the way we conceive of nature–society interactions, throwing into question much of what environmental managers and scientists previously had taken for granted. As an interdisciplinary field, political ecology has been at the forefront of attempts to integrate the advances in post structural social theory and nonequilibrium ecology into nature–society research.

Science and Nature in Post-Structural Social Theory

Since its coalescence in the 1980s, the field of political ecology has adopted, challenged, or been altered by new developments in social theory, especially those addressing the social construction of nature. The idea of the social construction of nature is a term commonly employed to stress the role of representation, discourse, and imagery in defining and framing our knowledge of the natural world and ecological change. The key debate is between an extreme relativist view of nature and an extreme realist one. Political ecologists have staked out the middle ground, stressing that the idea of the social construction of nature does not mean that nature exists only in our collective imaginations. The prevailing position within political ecology accepts the existence of a material world independent of human consciousness and sensory perception, while at the same time recognizes that our knowledge of that world is always situated, contingent, and mediated. Attempts to link the insights of social constructivist concerns with a scientific understanding of nature have adopted a philosophical position known as critical realism. Critical realism starts from the premise that the world exists independently of our knowledge of it and that its very independence means that human knowledge is not reality, but a representation of it. Nature exists apart as a reality, but scientific efforts to describe or give order to it should not be confused with that reality.

Another important source for the constructivist approaches to nature, feminist theory, must be highlighted here. Feminist post structuralists, of whom Donna Haraway is perhaps the most prominent, present a critique that challenges the assumptions of rationality and objectivity in science. Feminists' rejection of the prevailing concept of objectivity in science is based on a strong skepticism of scientists' claims that their observations can somehow be disembodied (removed from all context) to produce an objective and universal empirical truth. Haraway suggested an alternative understanding labeled situated knowledge; the idea that all observation is located, embodied, and partial, and that by recognizing and accepting this partiality, science can better be positioned to understand the limits of knowledge claims. Harraway's work, along with Bruno Latour's, also has been important to a branch of political ecology that incorporates science and technology studies' concepts of hybridity and networks, the former idea being increasingly important for theorizing the social construction of nature and the latter being viewed as corrective to Blaikie and Brookfield's 'chain of explanation'.

Post structuralism also introduced to political ecology an emphasis on new social movements based on socially constructed identities of race, gender, and ethnicity. In particular, it has introduced the idea of discourse analysis in political ecology research and the importance of exploring and revealing the ways in which the environment and environmental problems are discursively constructed. While a great deal of debate about the precise role of discourse analysis in the field continues, over the past decade or so, it has become widely recognized that material analyses in political ecology cannot be conducted in the absence of or separately from discursive analyses. In political ecology's engagement with discourse analysis, emphasis is placed on a critical perspective toward modernist notions of objectivity and rationality, on interrogating the relationship between power and scientific knowledge, and the recognition of the existence of multiple, culturally constructed ideas of the environment and environmental problems.

Nonequilibrium Ecology

The idea that equilibrium, order, and permanence characterize nature has long held sway in the science of ecology. For most of the twentieth century, the dominant view of environmental change was that ecological communities progressed through a linear series of stages, which ultimately culminated in a final, stable climax stage. Any force, natural or anthropogenic, that upset the normal equilibrium state was characterized as disturbance. It was further theorized that following a disturbance, the community would progress stage by stage, returning ultimately to the climax stage of stable equilibrium.

In the late twentieth century, the equilibrium view of ecology came under increasing scrutiny as more and more empirical findings were in contradiction with it. A new ecology – nonequilibrium as opposed to equilibrium ecology – subsequently emerged. This perspective suggests that stability is not the norm for many ecological communities and that revised, if not revolutionary, models of nature are required. The new ecology replaces assumptions of equilibrium, predictability, and permanence with instability, disequilibria, nonlinearity, and chaotic fluctuation. Disturbance is no longer assumed to be anomalous, or an event from which a system recovers, but rather is considered sometimes to be integral to the system's functioning. The development of chaos theory has also influenced the introduction of the notion of nonlinear change in ecology. Chaos theory highlights system complexity, recognizing that chance is inherent in natural systems. Thus, what appears as chaotic change is actually ordered, though not linear or predictable. Furthermore, chaos theory demonstrates that many systems display high sensitivity to initial conditions. That is, small differences in initial conditions magnify into great differences in final conditions.

A final feature to emphasize in nonequilibrium ecology is the importance of spatial and temporal scale in the analysis of ecological change. First, a linear or regular relationship between the intensity of ecological impact and distance in time or space cannot be assumed. Second, nonequilibrium ecology raises the question of how large a spatial scale is necessary before we can confidently identify a system in balance. Third, it further suggests that in complex natural systems, different rates of change, propelled by different abiotic and biotic processes, are occurring simultaneously at different spatial and temporal scales. An important focus, then, is the coupling of subsystems operating under different periodicities and at different spatial scales.

In the wake of these theoretical developments, there have been increasing attempts within political ecology to identify convergences in nonequilibrium ecology and social science that can provide the basis for a synthetic treatment of nature–society research. Among these efforts at synthesis, four themes emerge repeatedly. The first concerns the production and utilization of environmental knowledge in conventional science. Much of this work focuses on the critical evaluation of taken for granted scientific claims about the direction and causes of environmental change. Researchers are combining the lessons of nonequilibrium ecology and critical social theory to demonstrate the weakness, or in some cases, the complete absence, of scientific evidence to support widely accepted claims about human–environment interactions. The second theme emphasizes how incorporating the insights of nonequilibrium ecology can bring a renewed focus on nature's agency in shaping human–environment interactions.

The third theme highlights the importance of temporal and spatial scales in both nonequilibrium ecology and social theory. The analysis of human–environment interactions on a temporal scale underscores the nonlinear, noncyclical, and chaotic character of environmental change identified in nonequilibrium ecology. The critical evaluation of the production of knowledge on a temporal scale helps elucidate the politics and relations of power from which environmental orthodoxies emerge and become entrenched as scientific truth. Ideas of spatial heterogeneity, interscalar connectivity, and scale dependent processes in nonequilibrium ecology are combined with a social theoretical approach that treats scale as socially produced rather than ontologically given. The fourth and final theme concerns the critical evaluation of the institutions, policies, and management practices in light of the advances in social theory and nonequilibrium ecology. Allowing for multiple perspectives in society and unpredictability and indeterminacy in nature has brought new appreciation for local management systems, demonstrating the roles of local institutional arrangements and cultural practices that emerge in an ecological context defined not by stability, but by extreme spatial and temporal variability.

Critical Approaches to Environment and Development

Political ecology's approach to nature–society relations has explicitly linked capitalist development with ecological change across multiple temporal and spatial scales. The field has been an important source of critical analyses of the social and ecological effects of development, focusing particularly on the material and discursive aspects of property rights. Specifically, analysis has focused on the way that rights are defined, negotiated, and struggled over among different social groups – class, gender, racial, or ethnic groupings – as a means to explore patterns and linkages between economic development and ecological change. These same analytical tools have been brought to bear on territorially based biodiversity conservation strategies (i.e., national parks and other protected areas), the mainstream political response to ecological crises under capitalist development.

Development for Whom?

To a significant degree, political ecology has operated as a critique of the discourse and practice of mainstream development thought, from modernization theory to sustainable development. A key approach in early political ecology was to evaluate development initiatives from the bottom up, emphasizing the importance of local and indigenous technical knowledge (ITK) about the environment. The role of ITK in environmental man agement and regional development has since been widely debated within political ecology. The argument that ITK should provide the basis for development from below depended on a fairly limited conceptualization of knowledge. ITK literature tended to overgeneralize about technological modernization's destruction of local economies and environments and gave too much weight to ITK as a necessary and sufficient foundation for improving livelihoods. What were overlooked are the other types of knowledge that rural producers may not possess, such as market conditions, price variability, and distant institutional constraints, that may be necessary for the improvement of local livelihoods.

In their analyses of development, political ecologists have borrowed from and made important contributions to common property theory and our understanding of tenure relations more generally. In political ecology, property rights are viewed as an expression of the social relations of production, forged in specific places in spe cific political economic contexts and embedded in locally generated meanings of land and resources. Development interventions that alter ownership, control, and man agement of land and resources also raise questions regarding labor, production, and surplus appropriation. Several studies have demonstrated that – contra World Bank policy rationale – privatization is neither necessary nor sufficient for the promotion of rural development. In analyzing the consequences of privatization and state appropriation of the commons, the fundamental question in political ecology studies has been, then, which prop erty rights are being secured for whom.

Development in the mainstream literature is con ceptualized as the solution to rather than the cause of backwardness, poverty, and environmental degradation in rural Third World communities. For some political ecologists, however, the post World War II project of Third World development is an exercise in domination – of the North over the South and the state over local communities – and therefore should be abandoned. Decades of development interventions, it is argued, have failed to relieve either poverty or environmental deg radation while succeeding in strengthening the power and authority of the state and international actors over local communities. For these scholars, development is conceptualized as a discourse, constructed not to liberate the Third World people from poverty, but to subjugate extensive regions in the Global South to the needs of a Northern based capitalist world economy. There is a lively, ongoing debate within political ecology con cerning the treatment of development as discourse and the nature of local encounters with development. Based on ethnographic case studies, some have argued that local community resistance politics is not merely a rejection of development, but rather culturally embedded demands for development as a right or entitlement from the state. Several political ecologists have argued that many local social movements are concerned not with halting development interventions, but with managing modernization processes on their own terms and forcing the state to address problems of underdevelopment and inequality.

Who Pays for Biodiversity Conservation?

The twenty first century biodiversity conservation strategy is structured around the idea that the estab lishment of national parks and protected areas is the best and indeed the only way to insure the survival of wild species. Critically referred to as the fortress model, this strategy is based on the idea that biodiversity conser vation requires large territories where ecosystems are allowed to function undisturbed by human activities. The 1970s saw the greatest expansion of protected areas in history, with the number established between 1970 and 1990 exceeding the number created in all the previous decades. Most of the increase has occurred in tropical Third World countries in the form of the fortress model.

Biodiversity conservation is a territorially based strategy that is anchored in the proprietary claims of the state. As such, it raises questions that are at the very core of political ecology, such as how the relationship between nature and society is conceptualized, how access to land and resources is controlled, and how environmental costs and benefits are distributed. Fortress style nature conservation has been highly politicized because the state has claimed ownership over wilderness areas that others considered to be their homeland. Political ecology has offered a powerful analytic to assess both the ecological efficacy of the territorial approach to biodiversity pro tection and the politics that produce and are produced by it.

Political ecology studies have demonstrated that the establishment of protected areas often functions as a form of enclosure. Forests, bushes, and pastures that were held as communal property for community members to hunt, gather, and graze have been enclosed and con verted to state ownership for the conservation of wild flora and fauna and their touristic potential. Historically, this has been conducted with minimum involvement of the people most affected by enclosure. State agents have often resorted to force in order to curtail com munity access to the commons and to evacuate residents reluctant to relocate. As a consequence, those com munities located on the boundaries – more often than not having been displaced in the process of park establishment – are commonly antagonistic toward state directed conservation policies. Political ecology is concerned with the issues of political conflict, social justice, and ecological efficacy that surround state directed biodiversity conservation.

Of particular interest in political ecology are the competing demands for material access to biodiversity (in all of its forms) and the various symbolic meanings that the state, international conservationists, and segments of civil society attach to parks and protected areas. Political ecologists have pursued several areas of inquiry related to fortress style biodiversity conservation. For example, what has been the history of human use and occupation of protected areas and how has it influenced biodiversity? Advocates of fortress style parks have tended to disregard or downplay historic human occupation and the role of human use and management on the ecology and landscape targeted for preservation. Recent scholarship, however, details how various peoples had occupied, used, and significantly shaped the ecology of what is purported to be primeval nature. Second, what is the relationship of mandatory evacuations of protected areas to state directed segregations, resettlement, and reservation schemes? Political ecology research has shown that governments have often combined conservation enclosures with other state policies of social control and segregation, including various state schemes of relocation, native reservation, and settlement concentration. Finally, how do ideas of national parks and equivalent reserves relate to the social construction of collective identities? Several political ecology studies have demonstrated the important role of parks and protected areas in the construction of national identities. Others have investigated the ways in which protected area policies and practices intersect with social constructions of class, ethnicity, and race hierarchies.

Political ecology's primary concern with the fortress approach is that it is deeply flawed for both ecological and political reasons. In ecological terms, global biodiversity losses have accelerated during the same period in which the number of parks and equivalent reserves increased exponentially. In political terms, the obsession with the wilderness ideal has meant that communities are displaced and the local commons enclosed on principle. In the absence of reliable data on human activities and ecological change, the standard procedure has been to evict first and ask questions later, if ever. Local communities thus have borne the costs of protected area establishment but have received few of the benefits.

Partly because of the contentious politics produced through protected area establishment, new approaches to biodiversity conservation have been proposed, emphasizing participatory and emancipatory community based initiatives. Sometimes referred to as nature–society hybrids, these are new conservation territories that, in contrast to fortress conservation, integrate human habitation and resource use with biodiversity conservation goals. Nature–society hybrids go by a number of generic labels, including buffer zones, indigenous reserves, integrated conservation and development, and community based natural resources management. Political ecologists have suggested incorporating both the environmental politics of progressive social movements and the theoretical advances of nonequilibrium ecological science into new kinds of conservation planning. This would take the form of conservation territories with boundaries that are flexible, multiscalar, and overlapping so as to accommodate the spatial and temporal heterogeneity of ecological processes as well as the extralocal geography of social and political movements.

Recent Trends and Future Directions

Political ecology is a young field that has rapidly expanded beyond rural, Third World concerns into new research contexts and new areas of theoretical and philosophical inquiry. For example, a strong urban political ecology focus has developed in recent years, partly in recognition of the rapid urbanization of large regions of the Third World and the accelerating expansion of global trade that is intensifying and multiplying the linkages among far flung places around the world. Urban political ecology characterizes urban environments as the products of the intermingling of social and ecological forces, which results in a hybrid form requiring vast inputs of capital and labor, yet still linked to and dependent upon biophysical processes.

Another new focus of analysis is environmental security. Political ecologists have developed one of the most potent critical assessments of environmental security thinking and a cogent argument for an alternative analysis of the relationship between environmental change and war and armed conflict. The key criticisms of environmental security include, first, that it relies heavily on discredited Malthusian ideas of resource scarcity. Second, in the environmental security literature, resource scarcity is conceptualized in such a way as to encompass very different social and environmental processes and conditions, resulting in a confused and muddled set of variables and causal relationships. The third area of criticism is that environmental security thinking treats conflict as a problem internal to 'groups' or 'societies' with little or no analysis of interactions with the international political economy. Emerging from this critique are several new political ecology studies that explore the relationships among violence, justice, and the environment.

A third important new direction in political ecology features the incorporation of philosophies of ethics. Recent developments within and outside of the academy suggest that an engagement with formal ethics will provide an important trajectory in the evolution of political ecology. For example, the rise of new social movements for environmental justice in the 1980s and 1990s pushed political ecology to explore the linkages between environmental concerns and political empowerment and emancipation, and thus closer to an explicit engagement with ethics. Environmental justice movements have challenged technological and managerial approaches to environmental problems by asserting the legitimacy of moral and ethical concerns. One expanding area of concern is the mixing of human and nonhuman genes in bioengineered organisms, raising fundamental concerns about the rights and moral standing of such hybrid creatures. New studies in the field are now probing the ethics of nature–society relations from a variety of philosophical perspectives. New areas of exploration include the possibilities for developing ethical relations in consumer behavior and ethics in the practice of political ecology, both in terms of research strategies and classroom pedagogy.

Finally, debates over indigeneity, property rights, and stewardship have sparked a range of fascinating new works in political ecology that are beginning to probe more deeply the relationship of identity and environment, raising new questions about the way we theorize culture and nature, and about the politics of conservation and resource management. Future political ecology studies will need to be aware of the instabilities and cultural constructedness of group self identification while at the same time remaining sensitive to the validity of local historical narratives, practices, meanings, and attachment to place. The theoretical complexity and empirical and ethnographic detail of many of the studies constituting a new wave of attention to identity in political ecology have set a high standard and suggest a range of inquiries for future work.

See also: Chaos and Complexity; Conservation and Ecology; Development I; Development II; Environment, Historical Geography of; Environmental Hazards; Indigenous Knowledges; Nature, Social; Natures, Gendered; Peasant Agriculture; Poststructuralism/Poststructuralist Geographies; Sustainable Development.

Further Reading

Adams, W. (2001). *Green Development: Environment and Sustainability in the Third World* (2nd edn.). London: Routledge.
Blaikie, P. and Brookfield, H. (1987). *Land Degradation and Society*. New York: Methuen.
Braun, B. and Castree, N. (eds.) (1998). *Remaking Reality: Nature at the Millennium*. London: Routledge.
Escobar, A. (1995). *Encountering Development: The Making and the Unmaking of the Third World*. Princeton, NJ: Princeton University Press.
Forsyth, T. (2003). *Critical Political Ecology: The Politics of Environmental Science*. London: Routledge.
Harvey, D. (1996). *Justice, Nature and the Geography of Difference*. Oxford: Blackwell.
Neumann, R. P. (2005). *Making Political Ecology*. London: Hodder Arnold.
Peet, R. and Watts, M. (eds.) (2004). *Liberation Ecologies: Environment, Development, Social Movements* (2nd edn.). London: Routledge.
Peluso, N. (1992). *Rich Forests, Poor People. Resource Control and Resistance in Java*. Berkeley, CA: University of California Press.
Peluso, N. and Watts, M. (eds.) (2001). *Violent Environments*. Ithaca, NY: Cornell University Press.
Robbins, P. (2004). *Political Ecology: A Critical Introduction*. Oxford: Blackwell.
Rocheleau, D., Thomas Slayter, B. and Wangari, E. (eds.) (1996). *Feminist Political Ecology: Global Issues and Local Experiences*. New York: Routledge.
Schroeder, R. (1999). *Shady Practices: Agroforestry and Gender Politics in the Gambia*. Berkeley, CA: University of California Press.
Stonich, S. (1993). *I am Destroying the Land! : The Political Ecology of Poverty and Environmental Destruction in Honduras*. Boulder, CO: Westview.
Zimmerer, K. and Bassett, T. (eds.) (2003). *Political Ecology: An Integrative Approach to Geography and Environment Development Studies*. New York: Guildford.

Relevant Websites

http://globetrotter.berkeley.edu
Berkeley Workshop on Environmental Politics, Institute for International Studies, University of California, Berkeley.
http://www.centerforpoliticalecology.org
Center for Political Ecology.
http://www.stetson.edu
Cultural and Political Ecology Specialty Group (CAPE) within the Association of American Geographers, Stetson University.
http://jpe.library.arizona.edu
Journal of Political Ecology.
http://www.polisproject.org
POLIS Project on Ecological Governance, Law and Political Ecology.
http://www.geog.cam.ac.uk
Political Ecology of Development Group, University of Cambridge, Department of Geography.
http://www.thecornerhouse.org.uk
The Corner House.

Political Economy, Geographical

R. Le Heron, The University of Auckland, Auckland, New Zealand

Glossary

Geographical Imaginaries The concept recognizes that implicit and explicit geographical dimensions are integral to all thoughts and practices and that in strategic and operational terms geographical imaginaries are integral to governing initiatives.

Methodological Individualism A knowledge production strategy that confines inquiry to discrete entities or their surrogates and assumes that aggregated information (the sum of the parts) pertaining to the entities is an adequate strategy.

Methodology The how part of producing knowledge. Decision steps, procedures and practices, explicitly or implicitly derived from assembled knowledge, that are followed in order to construct answers to questions or hypotheses of conceptual, theoretical, methodological, or substantive kinds.

Performative or Nonrepresentational Knowledge This knowledge aims to provide more politically and ethically informed capabilities to reflect while involved in making decisions.

Political Economy A field of knowledge that seeks to reveal economic processes in their institutional and organizational settings. Developed over several centuries, it has provided abstract and context-specific understandings of key transformations in social conditions of production and consumption, the state at all levels and forms, interactions among state–economy–civil society (community). It offers broad schema that reveals who gets access to what, under what basis, and with attention to wider implications.

Relational Ontology An understanding of human and nonhuman processes in terms of relationships rather than things (e.g., individual entities). Particular relations and interactions constituted through the relations are explored to delineate processes and their outcomes and patterns.

Political Projects Speculative imaginaries (of particular interests) involving rationalities that are intelligible, spaces that are seen as potentially governable, governable subjects, translation strategies to link to immediate circumstances and practices, and technologies that facilitate action and assist with knowing the world.

Representational Knowledge Term used to signify knowledge that purports to represent the world (in various ways) through learned (cognitive) processes. In Western experience observation and the visual has been privileged, at the expense of the breadth of senses. Most knowledge systems portray the world but offer little about how to change the world. Other senses are increasingly considered to be important in developing new strategies for working on the world.

Social Relations This term summarizes the social relationships of production and consumption. In geographic political economy the term has mainly meant capitalist social relations. Recent work shows that a diverse set of social and spatial relations characterizes human existence.

Introduction

In 1981, the Association of American Geographers held its annual meeting in the new Bonaventure Hotel, downtown Los Angeles (LA). The meeting has stuck in author's memory because it was in the economic geography sessions that he saw for the first time the sharp edge of knowledge politics. He witnessed differences in interpretation over the word structure, in 'structural change' and 'changing structures'. Heated argument in the corridors exposed lines of deeply felt difference. First was the politics of what economic geography was about. Structural change meant alterations in the makeup of the economy, the mix of industries and firms as distributed across space, usually measured by employment because that data was easy to obtain and full employment was the approved goal in the post 1945 long boom. Changing structures was code for structures of social relationships especially the class based relationships that distinguished capitalism where production was organized for profit. A second dimension involved methodological choices; evidence of structural change was produced using analytical techniques such as location quotients and shift and share analysis, while changing structures were summoned into existence by theorizing qualitative shifts in underlying (capitalist) relationships. Third, each group named different authorities in support of their arguments: economics and Keynes, political economy, and mostly Marx. Who counts as friends and from whom inspiration is drawn means a great deal. Finally, looking back, there were hidden ontological tensions. In a challenging way the meeting had a spectre of intellectual alienation. Being asked to defend the premises relating to why and how one sought to produce economic geography knowledge, and not being able to do so when confronted to do so, was akin to being stripped of the means of producing knowledge as one knew it. What commitments lay ahead for those who chose to engage with the burgeoning literature that was surfacing in economic geography under the name of political economy? Indeed, would openness to influences from within and outside geography be a feature of the yet to be solidified new economic geography practices or would challenge from any quarter be perceived as too threatening and dismissed?

The author's anecdote cut into a particular institutional and geographical context: an ascendant Anglo American geography, past the high point of regional science that informed much economic geography curriculum and research practice, a US conference, a world disturbed by oil shocks, rising unemployment in the flagship industries of the major capitalist economies, a time when many in economic geography were finding their existing frameworks to be underperforming in terms of providing answers about a perplexing and changing world. The methodological individualism of most economic geography could not cope with structural forces that seemed to lie behind developments. Perceptions of economic geography's inadequacies were reinforced by the critique of capitalism that was gathering momentum in the social sciences. That anxiety and confusion as well as new resolve and commitment should reign in LA was hardly surprising.

The reflection on an economic geography encounter nearly 30 years ago raises issues that must be confronted when boldly declaring by label as an encyclopedia entry (the author resists using the abbreviation GPE, it empties the label of its descriptive content) that geographic political economy is an identifiable and distinctive thread of economic geography research and scholarship. A Google search, however, of combinations of words around geographic political economy produces only a handful of links, including ironically for a field with an unrepentant bias toward the urban and industrial one entry from the rural political economy stable and a couple of hits on university research programs, at Bristol and Manchester. This article has to argue the field identifier into existence!

Several richly detailed, penetrating, and geographically contained stock takes of Anglo American political economic geography practice are available. These sources are must reads for anyone wishing to learn how early movers and shakers of the field dealt with the trajectory they helped forge and how they worked with powerful crosscurrents arising from preoccupations in the social sciences. Richard Peet and Nigel Thrift wrote in 1989 of a perspective barely 20 years old making its mark. Allen Scott in 2000 offered an upbeat account of 'The great half century' of economic geography aka geography and political economy, followed by a more muted 'Perspective' 2 years later. Peet and Thrift concluded that original contribution of the political economy approach is its interventionary ability, while Scott presciently lays down a gauntlet when he contends that the great testing ground for economic geography in the future will be the central question of development, as a normative project of global significance.

Aims

This article highlights the knowledge production practices of those who have tried to privilege geography into economic geography inquiry in diverse ways, in combination with inspiration from political economy. This aim diverges from earlier retrospectives and field summaries by its singling out of guiding geographic ideas. Through this focus it is hoped the possibilities and extent of political work and ethical engagements rising from the field will be seen in a new light.

First, a short primer on interpretations adopted. Economy is understood in its broadest sense as social economy, or way of life, founded in production and consumption. Social production and consumption are viewed not as a neutral acts by neutral agents but as political acts carried out by social groupings (including classes), with a variety of, especially, managerial and elite interests, especially in investment and related choices. Geographic political economy seen in this way has intersecting imaginaries at play. These geographic imaginaries emanate from a changing internationally linked community of knowledge workers – academics, students, practitioners, and other supportive actors – who are doing two main things: representing the world through conceptualizations, theories, models, analyses, methodologies, metaphors, and narratives, and other approaches and presenting these representations through a wide variety of political and ethical (personal) strategies in different contexts. This conception prioritizes what is available, accessible, and acceptable for assembly and use by economic geography practitioners. It is also concerned to make explicit examples of change when some aspect of a geography–political economy relation could be said to have been understood, in some context, in quite different terms. To consider geographic political economy in this way is to reflect on the extent to which political economic geographers have come to grips with a key post structural insight, stated two decades ago for the audience of the time by Peet and Thrift, as a deep commitment to changing one's self as well as planning out programs to change social structures. The message has been more powerfully restated recently by J. K. Gibson Graham as, if we change the world, we change ourselves and if we change ourselves, we change the world. The poignancy of this statement is the subject of the final section of the entry.

Very obviously, geographic political economy as a whole offers significant resources for the current times. This view was arrived from the mapping of intellectual assets of practice that can be broadly associated into the field. **Figure 1** provides a selective but illustrative summary of this mapping. It is a geographically embedded trace of developments in the field showing the gradual appearance of contributions from new nodes. Dividing the figure loosely into three parts portrays the idea of key moments in development and a sense of a trajectory of accumulated practice. The nature of the entries varies in an effort to give something of the different mix of objectives being addressed. The metaphorical and emotive content of some entries in the figure is very noticeable. It reflects practice. The ordering is an approximation at best of a chronology, dates for the most part having been omitted because the timing of publication and uptake of ideas from publication varies enormously. Names are left out so as to maintain a social or more than one identity to the field. Where particular practices originated is signaled only to show the increasingly international endeavor of the field. As with any conceptualization there are problems. One is that **Figure 1** unavoidably 'others' a lot of geographic political economy and the rest of economic geography into invisibility. Another is the international movement of economic geographers. These said entries have all been important energizers within the field, though their geographic reach varies. Although no direct reference is made to the figure in the rest of the article, it should be kept in mind throughout because it reveals a remarkable spread and depth of actual and implied practices.

In the Background

Before exploring how practices have been developed from nearly 40 years, three short sections touch on influences that often intersect with the argued trajectory of geographic political economy.

The Influence of Regional Science

What regional science conferred upon economic geography by the 1970s and its subsequent legacy was an

Beginnings: (geographic) differentiating and positioning work around political economy

Regions (in question) UK
Regions (in crisis) UK
In what sense a regional problem? UK

Elaborating geographic processes under capitalism

Anatomy of job loss (the how, why, where, and when of employment decline) UK
Urbanization and urban planning (in capitalist societies) USA
Spatial divisions of labor (social structures and the geography of production) UK
How industries shape regions USA
Regional specialization and growth USA
Geography matters UK
New industrial spaces (flexible production organization) USA
Theaters of accumulation Can/NZ
Geography of economic restructuring Aust
Geographical industrialization USA
Localities UK

Decentring, re-contextualizing and reconceptualizing geographic political economy

Taking the rest of the world seriously UK
A global sense of place UK
Global shift (the internationalization of economic activity) UK
Geographies of accumulation Aust/NZ
Power geometries UK
Re-territorialization institutions and innovation systems Italy, Sweden, UK
Spaces of globalization USA
Global shift (transforming the world economy)
Globalizing era UK
Regional worlds (territorial development in a global economy) USA
Global economic geography conference (Singapore 2000, Beijing 2007) Singapore
Disclosing diverse economies Aust/US
Global shift (mapping the contours of the world economy) UK
Globalizing economic geography in the context of globalizing higher education NZ

Figure 1 Some ways of coming to know geographic political economy.

overwhelming interest in explaining the location and behavior of firms, the geographical organization and pat terning of the (national) space economy, and the sources of regional growth as read through a geographic data matrix of economic activities, often elaborated with reference to the input–output relationships of the economy. This was a census using field, often adding data from surveys. While regional science has less steadily declined in influence something of a look alike has reappeared under the label of evolutionary economic geography, which stresses rou tines and path dependence, the co evolution of firms, technologies, and territorial institutions and concern with convergence/divergence in spatial systems.

Incorporating and Extending Social Theory

A significant impulse of social theoretic writings taken up by political economy geographers has been the com mitment to a relational ontology. Key is the notion that the spatial is not merely a reflection of the social (e.g., the landscape is a mirror of capitalism) but constitutive of what the social is, in multiple ways. The phrasing, 'the spatial is socially constructed and the social is spatially constructed', was normalized by the late 1980s. When connected to critical realism project especially by Andrew Sayer space was seen as making a difference to whether the causal powers of particular social relations are activated and the forms which the social relations take. More recently, theorizing involving associating political economy with post structural understandings has addressed how such activation might be co constituted with others and performed consciously and reflexively. For some economic geographers, this is a distinctly different basis to the political and the economic, edging toward a geographic post structural political economy.

Reading Geographic Political Economy in a Changing World

Significant changes in world conditions (e.g., oil crises in the 1970s, Middle East conflicts, German unification, the dissolution of the USSR, the Internet, neoliberal politics, the Asian economic crisis, 9/11, climate change) impact on the usefulness of current ideas and trigger new the oretical and methodological directions. This often leads to splintering of effort, sometimes around core prop ositions and other times around contentious initiatives.

Framing a Field – Opening Narratives

Those who became disenchanted with the approaches of mainstream economic geography and regional science attempted to explain to their peers why they were ex ploring new directions. This involved two main threads, both motivated by concern about the inability of the existing repertoire of practices to provide answers to wider political economic questions. The first target was location theory which provided little or no purchase on the social. Attention quickly turned to opening the black box of the region. Doreen Massey, in a brilliant play with words that foreshadowed the post structural turn, invited a complete rethink of the nature of the uneven geography of economic activities. 'In what sense a regional problem?' is a classic. Her skillfully presented argument socialized the region, not in the manner of the multiregion frame work of regional science, but in a manner that made visible the theorizable interconnections of any region with other regions – through the competitive process. Regions should not be seen as problems in their own right (epitomized in the idea of the regional problem, reinforced by regional growth theory's emphasis on atomistic models such as export base or growth pole theory). This prompted supportive critiques, including Charles Gore's *Regions in Question*, which systematically laid down the logical flaws in different facets of regional growth theory that could be identified from some knowledge of capitalist processes.

The early writings thus utilized political economy to reapproach traditional problems. The result was a recasting of the economic world's gain/loss, profits/ wages, employment/unemployment realities as socially constructed in its geographic patterning, and particular location opportunities. Perhaps the exemplar of this epistemological break was *The Anatomy of Job Loss*. This methodologically innovative book (it combined data analysis with graphs, excerpts from press coverage, quotes from interviews, theoretical discussion, and empirical analytics) contains a much overlooked figure that goes to the heart of political economy of whatever ilk. **Figure 2** should be on the wall of every local economic develop ment practitioner. It shows how growth initiatives in one setting are conditional upon what is happening elsewhere. The wider context of a regional establishment is no longer hidden.

Geographical Imagination in (Largely Nation-State) Political Economy

The initial concern with instating the field around a re visioning of region rapidly morphed into a much wider research agenda. A series of thought experiments in geographic political economy were implied by the directions followed. These included:

1. Examination of the underlying spatiality that both enabled and resulted from capitalist development at different times and over long periods of time. These included the processes of industrialization, urban ization, and suburbanization.

Forms of production reorganization \ *Output change*	*Falling*	*Stable average growth*	*Rising*
Intensification	Employment losses *in situ*		–
Rationalization	Employment losses *in situ* Plant closure Possibility of relatively small local gains as a result of concentration of capacity. No new locations		–
Technical change	(1) Employment cuts *in situ* or (2) New locations (i.e., some local gains) and closures	Any combination of *in situ* employment losses, new locations, and plant closures to meet capacity requirements	

Figure 2 Spatial implications of forms of production reorganization.

2. The study of restructuring processes from a range of perspectives. A widely read special issue of *Economic Geography* published in 1982 that sprang from the lively political economy debates at Clark University was a marked departure from the journal's usual content.
3. What, it was asked, if the dynamics of industries were carefully traced with territory, emerging spatiality, and constitutive effects being from the outset of the guiding propositions of inquiry? The pioneering work in this regard has strong roots in two rather different Northern Californian areas, Silicon Valley (more accurately Santa Clara County and environs) and the Salinas Valley. More than two decades of contribution has come from Anna Lee Saxenian's theoretically informed research on the microelectronics industry dynamics, at first US centered, then incorporating offshore assembly, overseas state assisted competitor startups, increasingly with interlinked labor mobility and intellectual capital dimensions with Taiwan, China, and India. The Salinas Valley formed the focus of Margaret FitzSimmon's strategy of investigating the ramifications of an emerging increasingly capitalist agricultural region. Her focus on regional specialization and the growth of agribusiness showed how capitalist agriculture was an expansionary process that quickly led to large scale impacts locally and extra-regionally. The two examples from the wider San Francisco Bay area illustrate the intensity and transformative nature of capitalist dynamics.
4. In an ambitious project Richard Walker and Michael Storper formulated a model of geographical industrialization. The special merit of this strategy was that some of the futility of competition could be visualized within a theoretical framing. The diagrams they use in *The Capitalist Imperative* to depict geographies of industry emergence (heavily influenced by the microelectronics industry) can be read as showing how success is spatially contingent – a great deal depends on what is going on elsewhere!
5. A contrasting agenda characterized the UK scene. Would fruitful exchanges come from setting side by side conventional economic geography and political economy approaches? *Politics and Method* does just this, effectively demonstrating that when the disciplinary and wider social politics of research questions are opened to scrutiny, it takes some effort to design and align methodologies that actually provide answers to questions posed.
6. A vein of research examined capitalist processes in terms of responses by unions, in a range of settings including North America (John Holmes), the UK (David Sadler), and Australia (Bob Fagan).

Exploring the Political – in Geographic Political Economy

Where the going got tough was when political economy geographers attempted to frame strategies of intervention using their new knowledge production practices. Two contrasting experiences expose some of the difficulties.

In the UK, the Economic and Social Research Council (ESRC) funded Changing Urban and Regional Systems or Localities project sought to examine local restructuring experiences with special attention to critical realist propositions. Considerable skepticism soon mounted as the project began to be perceived as yet another series of detailed inductive studies, albeit enriched by commentary on local politics. The belief that too much empirical was taboo meant analysis would swiftly become detached from experience. This impasse sprang from competing intellectual ideas (empirical and critical realist positions in the context of regulation theoretical framings). However, the enor mity of industrial restructuring in the British context, the spectre of the politics of Thatcherism, and the distance of geographers from levers of power must be recognized. Looked at through contemporary lenses the challenges of politicizing the knowledge of difference and the implications of difference were hardly addressed and with that the political potential of the project slipped away.

During roughly the same period Australasia's eco nomic geography was also energized – by equally revo lutionary politics (though with quite different conditions and with people seeing the world in very different terms) and by the 1988 International Geographical Union (IGU) in Sydney and the meeting of the IGU Commission on Industrial Geography in Tasmania. A research project, dubbed GOER (Geography of economic restructuring) by Bob Fagan at Macquarie University, dovetailed with efforts by the late Steve Britten and Richard Le Heron in New Zealand to articulate a geography of political economy perspective. The New Zealand initiative, assisted by the New Zealand Geographical Society, produced two books, *Changing Places: A Geography of Restructuring in New Zealand* (dealing with the spatial impacts of state led restructuring) and *Changing Places: New Zealand in the Nineties*, focusing on new interactions by actors in changed circumstances. Each book was a collaborative (49 and 72 contributors, respectively) among geographers in and outside the academy. Although widely read in and beyond geography and the academy in New Zealand, the book lacked demonstration of techniques to engage with people. Actual dilemmas experienced by people in the new regime and the nature of their political conundrums were absent.

Toward a Globalizing Geographic Political Economy

Since the early 1990s political economy geographers have had to come to terms with working in and on a globalizing world. Indeed, while Nigel Thrift exhorted British economic geographers in 1985 to take the rest of the world seriously, it was not until 1991 that Doreen Massey again provided inspiration with her short note on 'A global sense of place'. Viewed in conjunction with Peter Dicken's textbook *Global Shift*, then subtitled, 'The internationalisation of economic activity', geographic political economy seemed doubly empowered. An over view and a lived example of globalizing processes was available for scrutiny.

In this context a plethora of cross cutting ideas, from the social sciences and more recently the sciences, have swirled through geographic political economy over the past 15 years or so. These have strained at the underlying premises and met varied resistance on the one hand and reinvigorated and stimulated existing and new avenues of research on the other. Again, for brevity, only indicative points are highlighted for each of a number of stra tegically valuable influences. Some added mostly to the representational capabilities of the field. Others are considered to be more active, denoting deeper political dimensions, with a focus on developing knowledge about emerging possibilities. The interpenetration and in some instances co constitutive aspects reflects the vigor of the field and the continuing reassessment that comes from fresh cohorts of researchers addressing new substantive issues. The explicit international vision of the IGU commission framework is also considered.

Representational Contributions

Regulation theory (and flexible accumulation)

This body of knowledge forced a re look at the state and led to greater respect for and awareness of the com plexities of institutional change during what became to be known as the era of globalization. The multiplicity and geographical variability of capitalist processes was picked up in the ideas of geographies of economies (which marked a shift from industrial to economic geography as the focus of the field and was reflected in the title of a book edited by Roger Lee and Jane Wills) and geographies of accumulation and more recently in the concept of global production networks. The European institutional thickness literature (e.g., Anders Malmberg, Bjorn Asheim) derived from wide interest in Third Italy had a geographic political economy emphasis but did not embrace the social capital thesis. Another direction was work on financialization by David Harvey, Andy Leyshon, Adam Tickell, Gordon Clark, and Nigel Thrift. Notably it led to little political engagement or development in political practice, especially as it was soon swept up in efforts to understand neoliberalism.

Cultural economy

The cultural turn exposed the Cinderella status of con sumption activities as an object of study in economic geography at large, the growing importance of con sumption inspired infrastructure investment on massive scales and the changing profile of work with the expan sion of female participation in the labor force and busi ness and other services (e.g., shopping malls, dockland revitalization, gentrification, coastal international devel opment, sporting professions, international consultancies in accountancy, law, management, architecture, con struction, and so on). However, a consumption focus has yet to be mainstreamed in geographic political economy despite efforts by researchers such as Louise Crewe, Louise Johnson, Peter Jackson, Nigel Thrift, and Neil Wrigley. As a carrier of post structural technique geo graphic political economy can be thankful for cultural economy. The capacity to frame problems has been immeasurably enhanced and this lies behind some of the latest developments in the field.

City-regions in a world of regions

In what some would contend is slippage toward an apolitical representation of capitalist processes the field has re embraced many conventional economic ideas. This has seen an explosion of research around the new economy, new regionalism, and more recently city regions. The lengthy catalog of accomplishments includes the topics of regions as sources of competitive and creative advantage, built up from work on informal cultures, conventions, and untraded interdependencies that sustain the functioning of localized production; in stitutions of social regulation and their effects on regional economic performance and the structuring of labor markets; and learning, innovation and creative processes and the way they shape institutional change. More recently, this work has begun to consider sustainability issues. What is less clear is the political content attached to these initiatives. Noticeable lacunae are the growing patterns of industry concentration and the intensifying influence of oligopolistic transnational cor porations (TNCs).

An Epistemological Break-In: Positioning Performative Knowledge

Deconstructing and reconstructing

Until the mid 1990s the call by feminist economic geographers to recognize, in geographic political econ omy, the over privileging of capitalist processes and the press of patriarchy went unheeded. The publication of J. K. Gibson Graham's *The End of Capitalism* (*As We Knew It*) rocked the boat, as it was intended, but to the con sternation of the authors, it also polarized the political economy geographers' community. As a demolition job on the field there are few comparisons. Its real contri bution, however, lay in the different questions placed on the table that could redeploy insight from the field in new ways. What would the world of political economy begin to look like if it was framed to include the full variety of social relations that go into constituting the economy? This trajectory has gradually leveraged more interest, especially from agri food and development studies researchers.

Rethinking agri-food commodity chaining

The feminist critique enabled significant developments in the agri food sphere of geographic political economy. These have confronted how value is appropriated in the roundabout of connections from field, paddock, and farm to the table, fork, and finger in the globalizing food economy. This has several foci; understanding the power relations and implications of value appropriation in conventional agri food activities (by a diverse field of researchers including Bob Fagan, Bill Pritchard, Niels Fold, Charles Mather, Terry Marsden, Jonathan Murdoch), the Berkeley and Santa Cruz tradition around Michael Watts, Margaret FitzSimmons, David Goodman, Bill Friedland, and more recently Julie Guthman, attempts to understand the potentialities and limits of alternatives (e.g., fair trade, organic, ethical production) especially contributions by Alex Hughes, Michael Goodman, and Sarah Whatmore, and links to political ecology through writers such as James McCarthy, Paul Robbins, Scott Prudham, Karen Bakker, and Becky Mansfield. The research has cross continental dimen sions, exploring particularly the Latin America–North America and Africa–European connections.

Following the commodity

A related line of research has focused on following the commodity and re presenting the encounters that come from such a research methodology (e.g., bananas, beans, chicken, coffee beans, cut flowers, papaya, and tomatoes). This approach involves much international field work, often over many years and has been spearheaded by geographers who reside in the space between the socio cultural and economic (e.g., Suzanne Freidberg and Ian Cook). These studies, many using the actor network approach, give new meaning to the geographies of global commodity chains and throw new light on political possibilities.

Firms (organizations) as temporary coalitions

A special issue of *Economic Geography* (2000) reopened the investigation of the firm. The enterprise had been rejected in early geography of political economy as a black box and a misspecification of social relations but defended by Peter Dicken and Nigel Thrift because of the role of the firm in the organization of production and the production of organization. Recent work by Mike Taylor (who has promoted the idea) and John Bryson on theoretical issues and manufacturing firms blends with initiatives by Erica Schoenberger and Linda McDowell on corporate investigations, Gerhner Grabher on project based studies and Peter Dicken and Henry Yeung on TNCs. Interpreted more generally the idea of temporary coalitions of people and resources facilitates a re look at the politics as well as the power relations of segmentation among small and medium enterprises (SMEs), multinational corporations (MNCs), and TNCs, Roger Hayter's studies of large firms as big firms locally, the structuring of governmental and policy frameworks and public–private partnerships and joint ventures.

Building diverse economies

Using the orthodox classification of the economy as transactions, labor, and enterprise J. K. Gibson Graham showed that the lived range of content for these cat egories is substantial. Thus, what is discernible are mar ket, alternative market, and nonmarket; wage, alternative wage, and unpaid; and capitalist, alternative capitalist, and noncapitalist social arrangements. This conceptual work opens up new avenues for geographic political economy.

Governing at a distance

A notable feature of neoliberal experience has been the fashioning of systems of comparison. These include techniques of benchmarking, best practice, standards, audit, certification, and so on. The investigation of these phenomena through the lens of post structural political economy has revealed to political economy geographers how spatialities, temporalities, subjectivities, and social ities linking places and their economies, and therefore the accountabilities and responsibilities that could ensue, can be explored using the governmentality literature.

Making industries and co-constituting political projects

Recent work on the New Zealand designer fashion in dustry (led by Wendy Larner) takes the post structural political economy direction further. The study traces the making of New Zealand designer fashion as an industry, and relates this process to four political projects of the state in which designer fashion is understood to have different possibilities – globalization, the knowledge economy, creative cities, and social development. It shows that the industry is a space in which political projects are mobilized and harnessed co constitutively.

Other Voices

The economic geography community has been actively supported by the commission structure of the IGU. Commissions on Industrial Change, Organisation of Industrial Space, Organisation of Economic Space and Dynamics of Economic Spaces have, for example, pro vided dedicated annual residential conferences in which to explore themes as diverse as 'Local development: Issues of competition, collaboration and territoriality' and 'Enterprising worlds: Entrepreneurship and ethics for sustainable futures'. The various Commission chairs have come from the UK, Australia, Italy, Korea, and New Zealand and the annual meetings have enjoyed considerable support from German, Dutch, Scandi navian, Korean, Japanese, Chinese, South African, and Australasian economic geographers. The Commission's strong publication series with Ashgate is illustrative of the coexistence of a range of related research agenda in economic geography. A new summer institute for emer ging researchers under the auspices of the World Uni versity Network has recently provided new impetus. The Commission meetings have been an outlet for sharing and subjecting empirical evidence to interrogation from a variety of frameworks, invaluable in maintaining the identity of economic geography as a field and for trans lating geographic political economy ideas.

Critical Reflections and the Politics of Possibilities

It is not an overstatement to say that the small comple ment of economic geographers committed to the field has been extraordinarily productive in moving forward understandings about the emergence of capitalism. The Marxist political economy tradition has been integral to this advance. Indeed, recent efforts to highlight the growing crisis of reproduction and human rights vio lations, as exposed by the Iraq situation and the aftermath of Hurricane Katrina in New Orleans, reconfirms the insight of this tradition. The field, however, is presently at a watershed. One is left with the impression of a rather splintered field, with tensions akin to those in the early 1980s. The insight of three decades is being cycled into the latest textbooks. Many still conceive of an economic geography wedded only to a political economy of capitalism. This constrains what politics is possible because emergence is circumscribed by assumptions about the immutability of capitalism. An economic geography where the interpretation of capitalism is less hierarchical and instead plural offers at least the promise and prospects of creating conditions for expression of possibilities however they are imagined in different sites/places. This will give new content and enlarge interest in the field. Those exploring the performative directions have mostly graduated since the early 1990s and influ enced by the post structural and cultural turns.

The watershed is exemplified by two internationally visible events in the 2000s. The first, the inaugural global conference on economic geography held in Singapore in 2000 and supported by the National University of Singapore, attended mostly by economic geographers from wealthier countries and universities, led to critical reflection on how global the conference (and the field) really was. The follow up in Beijing in 2007 (co organized by Henry Yeung and Weidong Liu) shifted the center of gravity of international economic geography participation even further. The second event was a workshop organized at the 2004 Association of American Geographers meeting on economic geography pedagogy. Articles published in the *Journal of Geography and Higher Education* attracted a critical assessment that identified several concerns: that it was unclear how far the North American geographic political economy is thinking beyond the nation state, that there was little appreciation that globalizing economic geography is caught up in globalizing higher education, and that there is much ambivalence to embracing post structural in sights. Depending on how we try to frame globalizing geographic political economy at the moment, the futures we make in geographic political economy, and the futures of geographic political economy's interactions with other communities of knowledge production inside and outside the academy could look very different. With this might come new identities and community in economic geog raphy. Recognizing the current condition of geographic political economy is critical to identifying priorities. Thinking geographically reveals our inherited and prized frameworks for what they are…governmentalities.

See also: Capitalism; Competitiveness; Consumption; Cultural Economy; Global Commodity Chains; Neoliberal Economic Strategies; Poststructuralism/Poststructuralist Geographies; Regional Science.

Further Reading

Barnes, T. (2004). The rise (and decline) of American regional science: Lessons for the new economic geography. *Journal of Economic Geography* 4, 107 129.

Cloke, P., Le Heron, R. and Roche, M. (1990). Towards a geography of political economy perspective on rural change: The example of New Zealand. *Geografiska Annaler* 72B, 13 25.

Coe, N. and Yeung, H. (2006). Revitalizing economic geography through teaching excellence: Some pedagogic reflections. *Journal of Geography in Higher Education* 30, 389 404.

Dicken, P. (2007). *Global Shift. Mapping the Changing Contours of the World Economy*. London: Sage.

Gibson Graham, J. K. (1996). *The End of Capitalism (As We Knew It): A Feminist Critique of Political Economy*. Oxford: Blackwell.

Gibson Graham, J. K. (2006). *A Postcapitalist Politics*. Minneapolis, MN: University of Minnesota Press.

Heynen, N., McCarthy, J., Prudham, S. and Robbins, P. (eds.) (2007). *Neoliberal Environments. False Promises and Unnatural Consequences*. Oxford: Routledge.

Hughes, A. and Reimer, S. (eds.) (2004). *Geographies of Commodity Chains*. Oxford: Routledge.

Le Heron, R. (2007). Globalisation, governance and post structural political economy: Perspectives from Australasia. *Asia Pacific Viewpoint* 48(1), 26 40.

Le Heron, R. and Lewis, N. (2007). Globalising economic geographies in the context of globalising higher education. *Journal of Geography in Higher Education* 31, 5 11.

Lewis, N., Larner, W. and Le Heron, R. (2008). The New Zealand designer fashion industry: Making industries and co constituting political projects. *Transactions of the Institute of British Geographers* 33(1), 42 59.

Peck, J. and Yeung, H. (eds.) (2003). *Remaking the Global Economy*. London: Sage.

Peet, R. and Thrift, N. (1999). Political economy and human geography. In Thrift, N. & Peet, R. (eds.) *New Models in Human Geography*, pp 3 29. Oxford: Blackwell.

Pritchard, B. and Fold, N. (eds.) (2005). *The Dynamics of Cross Continental Food Commodity Chain Systems*. London: Routledge.

Scott, A. (2000). Economic geography: The great half century. In Clark, G., Feldman, M. & Gertler, M. (eds.) *The Oxford Handbook of Economic Geography*. Oxford: Oxford University Press.

Scott, A. (2004). A perspective of economic geography. *Journal of Economic Geography* 4, 479 499.

Political Geography

D. Storey, University of Worcester, Worcester, UK

Introduction

Political geography as a systematic branch of human geography has a long, but not always distinguished, history. Like other branches of the discipline a precise definition is elusive with the nature of political geography, the issues explored, the approaches adopted, and the methods utilized displaying considerable breadth and variety. The nature of the subdiscipline has changed over time and its fortunes have waxed and waned for a variety of reasons. While being mindful of oversimplifying, it could be said that political geography is concerned with the interface between politics and geography. To be more precise, there is a focus on the spatial dimensions of power and with political phenomena and relationships at a range of spatial scales from the global down to the local. Another way of viewing this is to see it as revolving around the intersections of key geographical concerns of space, place, and territory on the one hand and issues of politics, power, and policy on the other. From this it follows that contemporary political geography encompasses a wide variety of themes. Rather than rigidly defining political geography it is perhaps best to think in terms of geographical approaches being brought to bear on a wide range of political issues. For some the study of spatial political units is central, for others there is an emphasis on major processes such as colonialism, while for still others it is concepts such as territory, state, or nation that are key. The diversity of issues and approaches means that it is more meaningful to talk of political geographies rather than a single unidimensional political geography.

Changes in themes, approaches, and methods have occurred in response to intellectual and methodological developments within the broader discipline and in academia more generally. However, they have also reflected broader social, economic, political, cultural, and environmental changes. The blurring of boundaries within the subdisciplines of human geography (and indeed the blurring of disciplinary boundaries more generally) means that some research that might be regarded as falling within the ambit of political geography is carried out by people who would not necessarily describe themselves as political geographers. Indeed they might not even regard themselves as geographers at all.

Evolution

In the late nineteenth century political geography was effectively synonymous with human geography. While physical geography was concerned with delineating regions on the basis of climate and topographical features, human geography was concerned with political divisions. Hence we still have a basic distinction between maps indicating physical features and those indicating 'political' features. However, in the late nineteenth and early twentieth century as the discipline evolved political geography became a more clearly defined arm of human geography running alongside commercial and colonial geography. Subsequently, and reflecting divisions within the broader social sciences, we can see political geography coexisting alongside social and economic geography.

Over the years the main preoccupations of political geography have changed. One hundred years ago key practitioners such as Halford Mackinder in the United Kingdom were concerned with international relations in what was then a colonial world. During the mid part of the twentieth century, the key focus became the state and its associated geographical characteristics most notably territory and borders. To a considerable extent this phase was strong on description of political phenomenon but weak on analysis. The last few decades of the twentieth century, however, saw something of a rebirth of the subdiscipline as a consequence of the introduction of more radical and politically engaged perspectives. More recently still, elements of social theory have been incorporated both deepening and broadening its conceptual base. There is now a much more critical approach and a more diverse one being brought to bear on a subject matter that extends well beyond the realms of the state. While much contemporary political geography maintains a focus on what might be seen as 'big' politics (states, governments, etc.), there has been an increased concern with 'small' politics (local issues, gender, ethnicity, social identities). In this way, a whole swathe of topics have been added to the more traditional interest in territory, borders, and states.

Ideas and Developments

The German geographer Friedrich Ratzel has been attributed with the first major work to include political geography in its title – *Politische geographie*, published in 1897. Ratzel likened the state to an organism which needed further space or *lebensraum* in order to expand. This organic theory of the state could be said to reflect two main characteristics of much late nineteenth century and early twentieth century political geography. First, it employed ideas derived from biology and other physical sciences and applied these to politics and society. Second, it reflected the view of Ratzel and others that geography and other branches of academia should be put into the service of the state. Ratzel was a supporter of German imperialism and his theories and his politics went hand in hand.

In the UK perhaps the most famous political geographer of the era was Halford Mackinder. He has been credited with attempting to shape the nature of geography in Britain in the early twentieth century and argued that the discipline should have objectives which were not just academic and educational but which should also serve the needs of politicians and business people. This vision of geography was a reflection of his own political views and he was at one time an MP and served as a British government advisor at the Versailles conference at the end of World War I. Mackinder is most remembered for his heartland theory or geographical pivot of history in which he suggested that, while sea power had served Britain well, technological changes, most notably the expansion of railways, meant that control over major land masses was of crucial importance. He argued the Eurasian landmass (essentially Russia) was of crucial significance and containing Russian expansion was of the utmost importance for Britain in its wish to maintain political prominence. Mackinder, like Ratzel, wrote from what can be seen as an imperial point of view. For all these 'imperial' political geographers their ideas were bound up with the influence of both physical features, such as mountains and rivers, and human geographical considerations on politics and their role in affecting the strategies of states. Their task, as they saw it, was the devising of practical geostrategies which could be utilized by political leaders.

Like his European counterparts, the US geographer Isaiah Bowman was an adviser to the government. Following World War I, he played a role in the redrawing of European borders in his capacity as a member of the US delegation at Versailles. While Bowman tended to see his contributions as detached and objective, his view of the world was mediated through the lens of US geostrategic interests. With his anticommunist views he endeavored to fashion a political geography reflective of US interests. At about the same time the German geographer Karl Haushofer utilized and developed Ratzel's ideas of *lebensraum*. Hitler's subsequent use of the concept as justification for German territorial expansion meant that both Haushofer and his ideas were heavily criticized. The subsequent significant decline of geopolitics has been at least partly attributable to this tainted episode in its history.

In the 1930s and 1940s the American geographer Richard Hartshorne attempted to delineate the field of geography. More specifically he also sought to outline the sphere of political geography as a subdiscipline. He saw political areas, most obviously the state, as a central concern, which dovetailed with his broader view of geography as an idiographic discipline in which the region was the central object of study. The discrediting of geopolitics and the retreat of geography into a largely descriptive regionalizing phase meant that for much of the mid part of the twentieth century political geography became preoccupied with the study of states and their borders. Issues of natural and artificial borders and theories of state evolution which likened these human creations to 'natural' phenomena were devised. While it would be simplistic to dismiss all of the work produced at this time, it has been seen as largely descriptive and politically conservative in its 'naturalizing' of the role of the state.

Recent decades have seen a significant rebirth of political geography stimulated by a number of methodological and theoretical developments. The dawn of a more quantitative approach in the 1960s saw political geographers work much more extensively with large volumes of data. This is most obviously reflected in the development of an electoral geography which cast light on such things as the importance of place and locality in voting patterns and the intersections of national issues with more local concerns. In exploring voting patterns attention was drawn to the ways in which neighborhood effects cross cut broader political issues to produce particular spatial patterns.

Geography's radical revolution brought about two key political geographic changes. Firstly it saw the introduction of more overtly structuralist perspectives into political geography and secondly it resulted in a very politicized geography. In the first instance a range of human geographical phenomena were analyzed within a broader framework. The impact of overarching social, economic, and political processes on people and places was emphasized. An example of this is the application of Wallerstein's world systems theory in the work of Peter Taylor. Here, rather than seeing states in isolation, there is an emphasis on the complexities of an interstate system and the connections between long term cycles of economic change and state stability and state structure.

The radical revolution within geography saw an intense politicization of the discipline. The 1960s saw the promulgation of the idea that everything is political and hence all areas of geographic inquiry were infused with political connotations. Radical geographers began asking intensely political questions about the distribution of wealth and power, control over resources, issues of discrimination, development and so on, while advocating radical political change in order to eradicate inequality in its various guises. Territorial social justice became a key concern, while the idea of academic objectivity and neu trality was eschewed in favor of a politically committed geography designed to radically transform the unjust structures which resulted in gross inequities. This devel opment reflected broader social and political turmoil in an era of US involvement in Vietnam, struggles against racism in various parts of the world, and the growth of the women's movement and environmental concerns. The resurgence of interest in political geography and the emergence of politicized geographers were reflected in the launch of the journal *Political Geography Quarterly* in 1982. This has subsequently increased its annual output and is now known simply as *Political Geography*.

More recently the influence of social theory in its broadest sense can be detected in some branches of research. Some geographers have explored aspects of everyday life and the ways in which political processes impact on people and inform their sense of identity. It is clear that politics affects us all in a myriad of ways. While older political geographies tended to focus on the world of formal politics and happenings in the 'corridors of power', more recent versions explore the everyday consequences of power and challenges to it. There has been a growing emphasis on the politics of identity, whether framed in terms of nationality, gender, ethnicity, sexuality, or other divisions. Alongside this, and closely connected to it, a postmodern influence has focused attention on the discourses employed by different groups in relation to political space. There is an emphasis on diversity and allowing different voices to be heard. This has helped bolster earlier concerns with those at the receiving end of unequal power relationships. Similarly, using the ideas of Foucault and others, there have been explorations of the ways in which power operates and diffuses across geographic scales.

Major Themes

While political geography now embraces a wide range of concerns some core themes can be identified. Chief among these are:

- Territory and territoriality.
- State.
- Geopolitics.
- Nation.
- Identity and citizenship.
- Electoral geography.
- Environment.

These themes are not discrete and they cross cut the more general geographic themes of space and place. As noted earlier the themes examined and the approaches adopted owe a lot to shifting paradigms within the broader discipline. The regionalist, empiricist, structuralist, and postmodern perspectives in particular are all apparent within the work produced within the past 50 or so years. By the same token, the scales of analyses have varied from the global through the national and regional down to the local. More recently, the inter connections between different scales of analysis have been emphasized. The local impacts of more national or global political processes or policies are one example.

Territory and Territoriality

Territory refers to a bounded geographic space with territoriality referring to the attempts to control that space. Traditionally 'formal' political territories (most obviously states) were the main concern of political geographers but more recently territory at a more in formal level has come in for scrutiny. Thus, we can think not only of territory defined by the borders between countries but also reflected in everyday life when we are confronted with signs saying 'authorized personnel only', 'keep out', 'no tresspassing', 'strictly no admittance', and so on. The idea of the home as private space and terri torial divisions in the home based on divisions between adult and child space or on gendered divisions of labor are examples of more micro scale territoriality.

Some strands of thought have 'naturalized' territorial behavior as something innate in humans. In this view the claiming of geographic space is seen as natural and defense of territory is seen as a biological urge. Others have viewed this in more social terms and, utilizing the ideas of Robert Sack, have seen territory as a key organizing device and territorial behavior as a means of retaining or resisting political power. From a political geographic perspective, territorial behavior might be viewed as a geographic and political strategy designed to achieve particular ends. In this way, the control of geographic space can be used to assert or to maintain power, or to resist the power of a dominant group. It follows from this that territories, whether, states, regions, counties, workplaces, the home, are not natural entities but the outcomes of a variety of social practices and processes in which space and society are linked. It is easy to regard territories as spatial con tainers but they are much more than that. They serve to convey messages of authority, power, and control. Issues of territory and territoriality are seen to be key underpin nings for many political geographic issues.

Geopolitics

Geopolitics is sometimes defined as the geographical dimensions of power (and hence seen as synonymous with political geography). It has tended to be concerned with international relations and in particular with the geostrategic concerns of major powers. The work of Ratzel and Mackinder are early examples of geopolitical writing with an emphasis on the geographic bases of political strategy. Different geopolitical eras can be identified characterized by changes in hegemonic power from the colonial era dominance of the likes of Britain and France, through the USA–USSR rivalry of the Cold War, to the current military dominance of the United States. As well as examining the geopolitical strategies of states and political leaders, geographers have pointed to the role of maps and mapping, together with ideas of environmental determinism (closely linked to scientific racism) as mechanisms through which colonialism was facilitated and justified. Within the Cold War there are clearly demonstrable geostrategic elements within US thinking. These included the invoking of the Monroe Doctrine (with the Americas as the US sphere of influence), the propounding of domino theory (that once one country fell to communism others would follow), and territorial strategies of containment (prevention of communism spreading beyond its existing 'borders').

Much traditional geopolitical writing tended not just to be written from the perspective of a particular geographical and ideological standpoint but also from an assumption of inevitability. The development of a more critical geopolitics through the work of people like Gearóid Ó Tuathail and Simon Dalby has created a more politically challenging perspective on global political issues. Here the emphasis on *realpolitik* is challenged by a view which refuses to accept the inevitability of current power relations and which affords the possibility (and desirability) of radical change. A strand of this is exploring the geopolitical discourses used by governments and political leaders. The ways in which places and regions are reduced to their strategic significance within global power politics is reflected through the use of particular terminology. For example in the 1980s former US President Ronald Reagan referred to the then USSR as an 'evil empire'. In a similar vein, the current Bush administration in the US conjured up an 'axis of evil'. These constructions and others such as some Islamic fundamentalist characterizations of a 'decadent' West can be read as geopolitical discourses designed to frame events in particular ways. The current 'New World Order' has been viewed in various ways ranging from Samuel Huntingdon's famous (and somewhat reductionist) exposition of a 'clash of civilizations' to David Harvey's political economy perspective in which current US geopolitical strategy is seen to be driven by territorial and capitalist imperatives and closely linked to resource control.

As well as the more obvious discourses emanating from politicians, geopolitical discourse has also been explored through film and other media. The ways in which events such as the Vietnam War have been framed by filmmakers or the worldview refracted through publications such as *Reader's Digest* are examples of the ways in which people, places, and events are represented and through which particular ideologies (often using simplistic representations of 'good guys' and 'bad guys' and particular constructions of concepts such as 'freedom') are disseminated and reproduced.

State

The division of the world into bounded political units, commonly referred to as states, is the best known example of formalized territories and of political–territorial organization. As a consequence, the state has long been a central element in political geography with a focus on various facets of the state including their origins, spatial development, key properties, roles, and functions. Traditional approaches in political geography have tended to take the state for granted. However, while it may be the dominant form of political territorial organization in the contemporary world, it has not always been so. Nevertheless, the state has been naturalized in much political geography. Geographers such as Ratzel developed 'organic' theories of the state which likened it to a natural entity which needed living space. Others devised theories of state growth in which it was argued states evolved through phases from youth to maturity, similar to rivers. These ideas can be seen as part of the broader trend of devising theories of political behavior analogous to natural processes. Ideas of state stability or instability were often linked to state size in terms of either land area or population and the extent of internal regional differences, whether physical, economic, or cultural. These centrifugal or centripetal pressures would help to determine state stability or instability.

Conflicts between states, especially border disputes, have also been a focus of attention. While earlier consideration of borders tended to explore distinctions between natural boundaries (such as rivers) and artificial boundaries (lines of latitude or longitude, for example), more recent perspectives have broadened to examine borders, not just as lines dividing territories, but as social and discursive constructs which can have important ramifications in people's everyday lives. They may have a profound impact on people's ability to travel and on a whole range of 'ordinary' activities. For some, such as nomadic groups, borders may be irritants that disrupt their social practices. Some borders are more significant than others; the French–Spanish border is less significant than the Polish–Russian one as the former separates two member states of the European Union (EU) while the latter marks the Union's eastern boundary. The collapse of communism led to a weakening of the borders of the Eastern European states many of which are now incorporated into the (theoretically) borderless EU. The Italian–Slovenian border became relatively open with the accession of Slovenia to the EU, an event symbolized by the removal of the border fence separating the Italian town of Gorizia from its Slovenian counterpart Novo Gorica (**Figure 1**). Alongside the easing of border controls within the EU there has been an increased hardening of its external perimeter. Various measures have been implemented making it more difficult for migrants to get into the EU, particularly those attempting to enter from African countries. Similarly, while the US–Mexico border is relatively permeable for those traveling South, it remains a sizeable barrier for Mexicans heading North.

Geographers have also been instrumental in exploring the role and functions of the state. While some have concentrated on the visible role of the state as regulator and service provider, others have delved into political theory to explain its functions. Questions about the relationship between state, society, and the role of the state within a capitalist economy have been addressed. Ideas of the pluralist state as a neutral arbiter between various competing interests have been challenged by those who see its role intimately bound up with a capitalist system. Gramsci's ideas of hegemony have been used by some to indicate the role the state plays in reproducing dominant ideologies.

The various processes sometimes conflated under the generic heading of globalization have led some to herald the end of the state as a meaningful political entity as a world of demarcated political spaces is replaced by increased flows of capital, labor, information, etc. It is suggested that the state's role as the sovereign authority over its own territory is much diminished. Contemporary political trends have focused attention on the ways in which sovereignty is asserted and contested whether through global processes, secessionist nationalism or suprastate institutions such as the EU leading some to predict the end of the state as a viable political–territorial entity. However, others argue that, far from disappearing, the state will continue to play a key role. At the time of writing, the US (a state) has recently invaded Afghanistan, Iraq and, appears to be contemplating the invasion of another, Iran, with a view to reshaping these in a way that suits its geopolitical ambitions. This hardly signals the demise of the state as a territorial phenomenon.

Recent decades have also seen the 'rolling back' of the state in many countries as public services have been increasingly privatized. In countries such as the United Kingdom, this has been accompanied by an emphasis on partnership arrangements between various statutory and nonstatutory bodies, alongside an apparent attempt to involve community groups in decision making. This reflects what some see as the fragmentation of authority from the central state and has led to a shift in emphasis in research from government to governance reflecting the increasing range of organizations involved to a greater or lesser extent in service delivery and decision making. While this has resulted in more diffuse patterns of service provision it does not in itself diminish the role of the state as an arena of political socialization or as a regulator of economic (as well as social and cultural) activity.

While the state remains a central object of study within political geography, it is now recognized that states are historically contingent, they are dynamic, and they reflect processes through which territory (and those living in that territory) are controlled. More recently there has been a move away from the more traditional

Figure 1 Wire fence separating Gorizia and Nova Gorica.

'taken for granted' view of the state. More structuralist approaches have focused on the broader context in which states exist. Here the incorporation of world systems theory by Peter Taylor has focused attention on an interstate system and the interactions of states. The complex web of power relationships within which states are embedded led some such as John Agnew to be wary of state centered thinking and he cautions against falling into what he has termed the 'territorial trap'.

Nation

Alongside a focus on the state, geographers have also explored ideas of the nation, national identity, and the political–territorial ideology of nationalism. A nation can be seen as a collection of people bound together by some sense of solidarity, common culture, shared history, and an attachment to a particular territory or national homeland. While history (whether actual or 'invented') is central to the nation's being, its right to exist usually rests on claims to a particular national space and, within this, particular places and landscapes often assume a symbolic importance. In exploring the connections between place and nation, three different levels can be identified. First, we can see the connections in allusions to the 'generic' territory of the nation. References to the national soil, the area of land seen to belong to the national 'imagined community', abound within such discourses. In this way fighting for, or even dying for, the land are seen as supreme acts of patriotism, ensuring that the land does not fall into 'foreign' hands.

The second territorial element is the importance attached to generic features which acquire huge symbolic significance. In this way particular landscape features such as rivers, lakes, or mountains, take on a much deeper meaning. Otherwise 'ordinary' landscapes are imbued with huge symbolic meanings to the extent that they come to be seen as emblematic of the nation.

A third territorial component of relevance to discussions surrounding the nation is the significance attached to particular places, not just generic features. In this way, the White Cliffs of Dover come to symbolize England (and, by extension, Britain) thereby acquiring a meaning which extends well beyond their geographical location. In some instances this can have quite serious ramifications with particular places seen as worth defending due to their national symbolism. For example, Kosovo has assumed almost mythical status for many Serbs as it is seen as a region central to Serbian territory and identity. In this way its loss would be likened to a rupturing of Serb identity. The breakup of Yugoslavia in the 1990s witnessed the phenomenon of ethnic cleansing justified under the guise of purging 'others' from land seen as being 'Serb' or 'Croat'. This can be seen as an attempt to 'purify' places and territories of those not seen as possessing the appropriate ethnonational identity.

The territorialized manifestations of national conflict are not just about the nation's macro territory but also about its micro level outcomes. In this way the geographies of ethnic division, such as the religious divides which run through Northern Ireland, become important objects of study along with the territorial markers which signify those divisions. The proliferation of 'peace walls' and of wall murals in the religious and politically divided city of Belfast signifies territorial defense and separation from the 'other side' (**Figure 2**). Equally support for secessionist nationalism is reflected in the landscape through political graffiti in places such as the Basque country where slogans and posters in the Basque language call for independence from Spain (**Figure 3**).

Identity and Citizenship

Geographers have also been interested in changing notions of citizenship and the relationships between individuals and the state. Established ideas of rights (granted by the state) and duties (obligations to the state) have been extended to encompass broader questions such as relationships, duties, and obligations to those beyond the borders of the state (distant others) or to the environment. Similarly questions have been raised, not just about individual rights, but collective rights in respect of particular groups (such as ethnic minorities, people with disabilities, sexual minorities) and attention has been drawn to the ways in which political structures, processes, and policies impact on such groups. Rather than the traditional notion of the relationship between the individual and the state, geographers and others have begun to explore the contested spaces of citizenship.

Overt discrimination against particular groups denies them 'real' citizenship and these divisions in terms of social identity are often manifested spatially. Some forms of discrimination are overtly obvious such as apartheid in South Africa from the 1950s through to the early 1990s with its designation of particular spaces for racially defined groups. Some lead to extreme outcomes as with ethnic cleansing in the former Yugoslavia designed to secure territory for specific ethnonational groups or the massacres of Rwandan Tutsis by the majority Hutu in 1994. Religious identity (closely intertwined with political affiliation) in Northern Ireland has created highly segregated social spaces with well defined residential divisions separating Protestants from Catholics (see above). The *banlieues* of Paris are both physically separate and socially distant from the heart of the city, cut off by a ring road (the *périphérique*) from the center and isolated from each other. Urban riots in the poorer *banlieues* on the edge of Paris in 2005 spread to other French cities. In these areas of high unemployment younger residents,

Figure 2 ‘Peace wall’, Belfast.

Figure 3 Graffiti calling for Basque independence, Bilbao.

many of North African immigrant origin, gave violent expression to their feelings of social and spatial alienation.

A concern with social identities and with minority groups has intersected with broader political questions of the relationship between the state and those who live within its borders. If citizenship relates to ideas of democracy and participation then legitimate questions can be asked about the barriers or constraints placed on some groups limiting the extent to which they can participate. Such concerns overlap with a heightened interest in questions of identity, social movements, and diversity. Cultural geographies, influenced by a postmodern emphasis on difference and diversity, have focused attention on both the multitude of social identities and the unstable, contingent, and relational nature of these identities. Groups have arisen with the aim of highlighting concerns over the attainment of particular goals linked to equality and esteem. Quite often these are identities which have been suppressed or largely ignored whether sexual, ethnic, or religious. The ways in which groups negotiate the political landscape or create political spaces are of interest. The emergence of gay spaces such as the Castro district in San Francisco or West Hollywood in Los Angeles has both enhanced the political power of sexual minorities through place based politics and may also have served to diminish the negative perspectives of others toward these groups. That the emergence of such gay spaces is also associated with commercial factors (and forms of gentrification) may have mixed benefits.

Within this there is a risk of promulgating essentialized or fixed ideas of identity rather than a recognition of multiple or many layered identities. A person may be French, black, gay, and the relative importance of these to the individual may vary from time to time and from place to place, depending on the circumstances in which they find themselves.

Ideas of more active citizenship have been encouraged by government through what might be seen as 'approved' channels, such as Neighbourhood Watch schemes in the UK or through involvement in community groups. The emphasis placed on community development, partnership working, and local capacity building reflects this. However, other versions of active citizenship may be less welcome by governments. Political protests in the form of such things as anti war marches, anticapitalist protest, and attempts at reclaiming private space from commercial development are less likely to enjoy government support (**Figure 4**).

Increasingly, citizenship is seen as multilevel and as not just national but both supranational (through attempts to create a sense of Europeanness, for example) and international (such as promotion of ideals of global citizenship with responsibilities for those beyond our own borders). Rather than a state centered view of citizenship, there are attempts to view it at spatial scales both below and beyond the state.

Electoral Geography

Probably the strand within political geography which most overtly followed a quantitative approach has been electoral geography. Within the geographical study of elections three broad concerns can be identified. These are the geographies of elections (forms of transfer of power), of representation (types of electoral system), and of voting (spatial patterns of voting behavior). The first explores the ways in which power is transferred, whether through elections or through more 'irregular' processes such as a military coup. Geographies of representation examine such things as the nature of electoral systems (e.g., first past the post or proportional representation), the existence of centralized or federalized systems, and the construction of electoral boundaries and associated issues such as gerrymandering. Ideally electoral boundaries should be constructed in such a way as to ensure roughly equal numbers of voters. However, population changes may render this difficult and political parties may have vested interests in maintaining levels of over or under representation in certain cases.

It is perhaps the study of voting behavior where geographers have raised the most interesting questions. Spatial analysis of voting patterns has revealed that

Figure 4 Anti-war march, London.

voting behavior is not just a function of party allegiance and ideological disposition. There are neighborhood or locality influences which may reinforce or override other factors. Events such as the closure of a local industry or the reputation of a candidate and her family in an area may be strong determinants of voting behavior. Ron Johnston and colleagues have examined the shifting geographies of voting in the United Kingdom where broader national trends intersect with more local factors. In Worcestershire in England, the election of an independent MP campaigning against the downgrading of a local hospital in UK elections in 2001 and 2005 is an example of the potential impact of local factors. More broadly the extent to which some parties have a broad national appeal while others have a more local or regional support is also of interest. This is particularly evident in countries such as Italy (with a north–south split) and Belgium (with an entrenched linguistic division) where some political parties have an explicit subnational focus.

Environmental Politics

Growing environmental concerns of recent decades have prompted interest among some political geographers in a variety of themes. These include issues of conflicts over resources and the role that resources such as oil or indeed water may play in generating political conflict whether at regional or global scales. A concern with global issues such as climate change and global warming has emerged alongside a consideration of such things as campaigns against road building programs in parts of Britain. While these latter might be seen as local place specific issues they can also be viewed as part of broader environmental campaigns bringing together issues of pollution, landscape esthetics, and the preservation of biodiversity.

The growth of green politics has led to the study of environmental organizations, their nature, composition and tactics, and the environmental discourses used by various groups (including both campaigners and political parties). Other areas of interest include political events such as the Earth Summits and the subsequent filtering down of environmental policies to regional and local authorities, the general rise of green political parties, and the incorporation of green concerns into the policies of mainstream parties. Policies such as those of the French state in carrying out nuclear tests in its Pacific territories rather than in mainland France combine a set of issues linked to environment, place, and politics.

A good example of an issue in which a range of political arguments can be seen to coalesce and which links places together and is multiscalar in nature concerns plans by the Shell Oil Company to bring natural gas onshore at Rossport in the west of Ireland for processing via a high pressure pipeline. The plans, which enjoy the support of the Irish government, have met with popular local resistance with overlapping concerns related to land ownership (some local farmers have been faced with compulsory purchase orders by the state), health, and control of resources. Local opposition is linked to broader national arguments within Ireland over who benefits from the exploitation of national resources (Shell is a multinational company and the Norwegian state company Statoil has a stake in the gas field) and to environmental campaigns outside the country. In addition, links have been forged with Nigerian activists where Shell has been a focus of opposition due to its activities in Ogoniland and elsewhere in the Niger delta (referred to below). In this example issues of local and national politics intersect with questions of place, resources, and environment.

Politics of Geography

Another take on political geography is to examine the interventions made by geographers into ‘live’ political issues. It should be obvious from what has already been said that much political geography is far removed from the model of detached academic neutrality much touted in various circles. Instead, the imperialist leanings and political biases of the likes of Mackinder, Ratzel, and Bowman have been obvious. The ways in which political geographers have interacted with those in the world of ‘real’ politics have varied over time. It could be said that many geographers, particularly in the past, saw their role as supporters of those in power and their work as serving clear political ends. More recently many geographers have seen their role more as one of a critical observation and a questioning of accepted orthodoxies rather than a willing subservience to power. Among other things this has involved focusing attention on the complicity of geography in colonialism, for example, and casting light on the political agendas underpinning some geographical research and the ideological nature of much ‘objective’ research.

Here are three relatively recent examples of geography and politics intersecting in a very direct way:

Shell in Nigeria

Throughout the Niger Delta region the activities of Shell have proved controversial in terms of allegations of environmental damage and human rights abuses. In the region of Ogoniland local activists and environmental campaigners have argued that Shell brought little if any benefit to local people and that their activities were detrimental to the environment. The execution of nine Ogoni activists (found guilty of hotly disputed murder charges) by the Nigerian government in 1995 precipitated a debate among members of the Royal

Geographical Society – Institute of British Geographers (the body to which many professional geographers belong). Shell was a corporate sponsor of the organization and debate centered on whether a learned society or professional body should continue to accept sponsorship from Shell. Those opposed felt that the company's environmental record and its questionable benefit for the people of Ogoniland rendered it inappropriate to retain such links with the company. In the event the sponsorship was retained but many academics left the society as a consequence. The episode precipitated debates over (among other things) the links between activism and the academy, the implications of commercial sponsorship for academic integrity, and the moral and ethical values held by geographers.

Academic Boycotts

The issue of an academic boycott of Israel has recently exercised the minds of some geographers. Calls for a full academic boycott of Israeli academic institutions have been circulating for some time. These have emanated both from Palestinian groups and from academics in other countries. There have been repeated calls in some quarters for a series of sanctions (of which an academic boycott would be one component) designed to isolate Israel as a consequence of its policies on the Palestinian question. The argument rests on the view that Israeli academic institutions have done little if anything to question the policies of their governments in relation to the Palestinian question. Some point to the earlier boycott of South Africa during the apartheid era as a basis for such a tactic in trying to bring about political change.

More specifically the academic boycott issue intruded directly into the policies and workings of the journal *Political Geography* in 2005. Controversy over whether or not to accept an article by Israeli based academics promoted a flurry of media coverage and an exchange within the pages of the journal itself. On the one hand there are arguments surrounding academic freedom, the primacy of free speech, and the need to engage in open dialog, while on the other there are calls for a recognition of the legitimacy of boycott as a tactic to try and bring about pressures for change. Debate continues over the issues of the morality, equity, practicality, and effectiveness of such actions.

Elsevier and Arms Fairs

A recent political issue with which some geographers wrestled concerned the links between Elsevier (publishers of this encyclopedia) and the arms trade through two associated companies which, among other things, organize arms fairs. Elsevier are a major international academic publishing company which produces a range of books and academic journals including *Political Geography* and many saw their connections with arms fairs as a very unsavory activity. A number of academics signed a petition calling on the company to sever its connections with this trade while others (many geographers among them) called for a boycott on publishing in, or subscribing to, Elsevier publications (including this one). Following this pressure and disquiet from the public and company employees, the firm has agreed to withdraw from this activity. In engaging with these and similar issues, geographers are dealing with the messy realities of 'real' politics and tensions that emerge involving issues of personal and career motivations to publish and disseminate ideas, political beliefs, and matters of conscience, ethics, idealism, and pragmatism.

Summary

Political geography is a diverse and ever changing field of geographic enquiry. As such it defies easy definition. It has moved from being an account of the distribution and arrangement of power at different (though overlapping and interdependent) geographical scales to a consideration of how power diffuses across different scales. The workings of political networks, the drugs trade, terrorist organizations, environmental campaigns, and many other political issues operate across geographical scales. While this account has perhaps given the impression that political geography has evolved in a linear fashion, that is not the case. People like Elise Reclus and Peter Kropotkin wrote geographies heavily influenced by their anarchist beliefs and critical of big power politics as long ago as the late nineteenth century.

A key criticism of traditional political geography has been its state centeredness. The uncritical assumption of the state as a 'natural' political unit has been replaced by attempts at a deeper exploration of the nature of the state and its reasons for existing. Another criticism is linked to the focus on a traditional politics centered on the state and ignoring other political issues. The state is now seen as only one of a number of actors (albeit a very powerful and resilient one) to which attention needs to be directed.

Another key change within political geography (and human geography more generally) has been the shift from the supposedly objective 'view from nowhere' to the clear acceptance of the idea of positionality on the part of the academic. While many may accept the 'situated knowledge' approach, it is not universal with some academics continuing to promulgate an academic objectivity that many others see as unsustainable.

Political geography has been shaped by a variety of influences including intellectual currents within the broader discipline and political events and practices beyond the academy. It continues to evolve and displays considerable vibrancy both in terms of the range of issues now considered and the variety of approaches brought to bear.

See also: Apartheid/Post-Apartheid; Bowman, I.; Community; Difference/Politics of Difference; Hartshorne, R.; Harvey, D.; Identity Politics; Kropotkin, P.; Mackinder, H. J.; Maps and the State; Nation; National Spatialities; Nationalism; Neoliberalism; Political Boundaries; Postmodernism/Postmodern Geography; Radical Geography; Reclus, E.

Further Reading

Agnew, J. (1997). *Political Geography: A Reader*. London: Arnold.
Agnew, J. (2002). *Making Political Geography*. London: Arnold.
Agnew, J. and Corbridge, S. (1995). *Mastering Space: Hegemony, Territory and International Political Economy*. London: Routledge.
Blacksell, M. (2006). *Political Geography*. London: Routledge.
Cox, K. (2002). *Political Geography: Territory, State and Society*. Oxford: Blackwell.
Dodds, K. (2000). *Geopolitics in a Changing World*. Harlow: Prentice Hall.
Gregory, D. (2004). *The Colonial Present: Afghanistan, Palestine, Iraq*. Malden: Blackwell.
Johnston, R. J., Shelley, F. M. and Taylor, P. J. (eds.) (1990). *Developments in Electoral Geography*. London: Routledge.
Jones, M., Jones, R. and Woods, M. (2004). *An Introduction to Political Geography: Space, Place and Politics*. London: Routledge.
Newman, D. and Paasi, A. (1998). Fences and neighbours in the postmodern world: Boundary narratives in political geography. *Progress in Human Geography* 22(2), 186 207.
O'Tuathail, G., Dalby, S. and Routledge, P. (eds.) (2006). *The Geopolitics Reader* (2nd edn.). London: Routledge.
Painter, J. (1995). *Politics, Geography and 'Political Geography': A Critical Approach*. London: Arnold.
Sack, R. (1986). *Human Territoriality: Its Theory and History*. Cambridge: Cambridge University Press.
Storey, D. (2001). *Territory: The Claiming of Space*. Harlow: Prentice Hall.
Taylor, P. J. and Flint, C. (2000). *Political Geography: World Economy, Nation State and Locality* (4th edn.). Harlow: Prentice Hall.

Political Representation

B. Forest, McGill University, Montréal, QC, Canada

Glossary

Descriptive Representation A principle of political representation in which the social and/or demographic composition of an elected body is similar to that of the population or voting constituency.

Gerrymandering The manipulation of electoral district boundaries intended to affect the outcome of elections.

Independence Theory of Representation A theory of political representation in which the representative is free to act independent from, and even contrary to, the preferences of her/his constituents.

Mandate Theory of Representation A theory of political representation in which the representative if bound to act strictly in accordance with the preferences of her/his constituents.

Proportional Representation An electoral system in which voters cast ballots for parties rather than candidates, and representatives are selected from lists generated by parties in accordance with the proportion of votes received by each party.

Territorial Representation An electoral system in which candidates are elected from geographically defined constituencies.

Substantive Representation A principle of political representation where the interests of social, demographic, and/or political groups are represented in an elected body.

Vote–Seat Ratio The ratio between the proportion of votes a party wins and the proportion of seats the party receives.

The Meanings of Political Representation

Nearly 40 years ago, political theorist Hannah Pitkin argued that only a constellation of meanings could adequately describe representation, and political representation in particular. The concept is so complex because representation is an apparent absurdity: to make present something that is not literally present. This would be difficult enough if the 'something' in question was a simple object in the world, but an act of political representation more typically calls the something into being. In other words, an election or the formation of a legislature constitutes the 'object' being represented, for example, the people, the will of the people, the nation, the national interest, etc. At the same time, political representation (at least in functioning democracies) must also be a practical activity providing the means of governance. The relationships among representation and geography involve both the constitutive and the practical dimensions of representation.

All democratic political representation relies on a kind of political fiction, the existence of 'the people' as a sovereign holding, but abstract entity. That is, the people hold power, but it is never possible to actually identify a particular set of the population as the people. Indeed, every less than unanimous vote reveals the fictitious nature of the people because any resulting representative or policy will have the support of (at best) a majority of voters. Democratic political representation presents a particular dilemma because the people are simultaneously rulers and subjects. As rulers, they are free to act according to their political will, but as subjects, they are bound by their own decisions. Political representation, and its practice in electoral systems, resolves this dilemma by delegating power to representatives who carry out the activities of governance. These representatives are typically restrained by constitutional provisions and previously established laws – as well as by future elections.

The creative or constitutive nature of political representation means that all systems of political representation have both actual and imaginary geographies. That is, geographical studies of political representation involve issues such as voting patterns and the division of state powers, but also issues such as the right to vote, a question that fundamentally defines the nature of the state or nation.

Issues of political representation have attracted increasing interest in political and electoral geography in the last 20 years. The roots of this renewed interest lie in historical, technical, and disciplinary developments. First, the fall of communism in Central and Eastern Europe (1989–90), the breakup of the Soviet Union (1991), and the end of apartheid in South Africa (1991–94) led to a renewed 'wave' of democratization that often required careful negotiation over both the institutional forms of electoral systems and the nature of the political communities to be represented. Second, the increased power and sophistication of geographic information systems (GISs) have brought new analytic power to the practice of political districting and the analysis of voting patterns. Finally, electoral geography is also moving beyond the analysis of spatial patterns of voting, instead focusing more on place specific or 'neighborhood' effects on voting preferences. These developments mean that political geography in general, and electoral geography in particular, has had to reexamine its assumptions about political representation and related concepts.

The Mandate–Independence Debate

Debate over the role of the representative, the so called mandate–independence debate, is one of the central issues in political representation. Under the mandate theory, a representative must always act in accordance with the will of her/his constituents, and indeed, ceases to represent them if s/he goes against their wishes. Strictly speaking, the mandate theory holds that a representative is simply a delegate, acting as a conduit to convey the actions and interests of voters to an assembly. In contrast, the independence theory characterizes the representative as free to act according to her/his own judgment and will, recognizing representatives as having unique skills and expertise unavailable to a typical constituent. This perspective would recognize a person as a representative even if s/he acts contrary to the expressed wishes of the constituency. The common meaning of democratic political representation lies somewhere between these two extremes, and relies on a reciprocal relationship between the represented and the representative.

Pitkin in particular argues that a proper understanding of political representation requires elements of both theories. The represented must be seen as capable of understanding and expressing their own interests, rather than simply being the subjects of caretaking (like children or the mentally infirm). There should not typically be conflict between the wishes of constituents and the actions of the representative, but such divergence must be possible. When differences arise, the representative must be able to explain and justify them. Democratic states institutionalize this reciprocal relationship through periodic elections, where voters can (in principle) replace inadequate representatives.

Descriptive and Substantive Representation

Another major debate involves the relationship between the identity of the representative and the composition of her/his constituency. Political theorists typically characterize this as a conflict between descriptive and substantive representation, sometimes described as the politics of presence versus the politics of interests or ideas. (Such discussions occur most often in the context of gender, racial, ethnic, or religious minorities, but are not necessarily limited to the role of such groups.) Descriptive representation or the politics of presence requires the physical presence of women, ethnic minorities, etc. in a representative assembly. In a sense, a body is representative when groups are present in the assembly in the same proportion as in the population. Some argue that descriptive representation has a value in and of itself, and/or that it helps confer legitimacy to an elected government, particularly in societies with significant inequality and stratification. More sophisticated defenses of descriptive representation suggest that the presence of women and minorities actually changes the quality and nature of deliberation in legislative bodies. A number of representative democracies, such as India, have institutionalized the politics of presence by instituting quotas for women and minority groups in legislative bodies. The US also imposed gender quotas for elections in Iraq, although it eschews them at home.

In contrast, substantive representation regards the identity of the delegates as irrelevant because representatives are seen as agents protecting the 'interests' of constituents. While voters may consider the identity of candidates while voting, the demographic composition of the legislature *per se* is not important as long as representatives are chosen in a way that allows the effective expression of the interests of voters. A male, for instance, would be seen as representative of females if he consistently supported women's interests. The concept of such objectively defined interests, however, can lead to an apparent absurdity. If interests exist objectively, substantive representation suggests that a constituency need not have a role in electing their representative, as with the caretaker of a child or incapacitated adult. Consequently, in its extreme form, substantive representation poses problems that are similar to those found in the independence theory of representation. This issue was one of the major disputes during the American Revolution, where the colonists objected to their 'virtual representation' in Parliament by representatives for whom they had not voted.

As a practical matter, democratic representation seldom involves either descriptive or substantive representation exclusively. Even among voters who might reject the idea of legislative or party list quotas, the gender or ethnic identity of candidates can be a concern if they feel that a member of their own group is better able to understand and represent their interests. Much like the mandate–independence debate, the extreme version of each form of representation violates the principle that – within limits – both representatives and constituents must be able to act freely and to exercise judgment.

Critics of Democratic Representation

Marxist critics in particular argue that such theoretical debates are irrelevant because 'democratic representation' is impossible under a capitalist system. The interests of capital too easily determine elections, and debates between parties and candidates simply serve to disguise how the state acts in the interests of capital. The idea of candidates and parties competing with each other is simply a political version of free market economic ideology. For these critics, structural forces rather than individual merit or effort determine the winners and losers in each realm. Insofar as political debate takes place through the media, major corporations can easily control the content and structure of political discussion and identify only favorable candidates as legitimate and 'serious'. Similarly, the funding required for political campaigns means that wealthy individuals and organ izations exercise a disproportionate impact on elections. Moreover, elections themselves may be largely irrelevant because those with the most influence over represen tatives – lobbyists, institutions, corporations, and so forth – are not affected by elections. The interests of capital are represented regardless of the form of 'democratic representation'.

Electoral Systems and Political Representation

Disputes over political representation generally take place in conflicts over electoral systems, rather than in abstract policy debates. An electoral system is the mechanism through which votes are translated (or not) into political power. Within liberal democracies an enormous variety of electoral systems exist in pure or mixed forms, although most are variations of either proportional representation (PR) or territorial plurality representation. Each system translates votes into power in slightly different ways. Each offers particular tradeoffs in the relationship between representatives and con stituents, and in the stability and responsiveness of the political system. The choice of an electoral system reflects the institutionalized political values of a society – or at least the values of those in a position to choose the system.

Proportional Representation

PR is the most common electoral system in the world. It attempts to translate the proportion of votes for a particular party into the same proportion of seats in the elected assembly. Typically, parties establish a list of candidates who win seats in proportion to the number of votes the party receives. For example, in a parliament with 200 open seats, a party receiving 50% of the vote would place the first 100 candidates on its list in the assembly. All proportional systems, however, have limi tations. The total number of contested seats determines the minimum percentage of votes required to elect a representative. In a 200 seat parliament, a party would need at least 0.5% (1/200) of the vote to place one candidate. In practice, however, most jurisdictions using PR establish a minimum threshold for electing candidates (typically around 5%) regardless of the theoretical minimum. Any party receiving less than the threshold cannot place candidates in the assembly, and the re maining seats are allocated to the other parties. **Table 1** shows the results with five parties contesting a 200 member assembly with a 5% minimum threshold for election.

In this example the remaining seats are allocated proportionally, rounding up to the nearest whole number, but other methods can be used as well. The example illustrates that proportional systems may not produce proportionality; the top three vote getting parties receive extra seats, and the two smallest parties, representing 8% of the voters, are completely shut out. This problem of disproportionality is worse in elections where a large number of small parties fail to meet the minimum threshold, or where the minimum threshold is very high. States typically require minimum thresholds to limit political fragmentation by forcing small constituencies to find common ground in one party or another. In parlia mentary systems, very low thresholds can give excessive influence to small parties who hold the balance of power in coalition governments.

Territorial-Plurality Representation

The other common electoral system is territorial plurality representation. In the simplest version of such a system, two or more candidates compete for a single seat from a specific territorial unit (districts, ridings, boroughs, etc.). In such single member, first past the post systems, the candidate with a plurality (or majority) of votes wins the election. The major advantages of such systems are (in principle) greater accountability and closer

Table 1 Allocating seats in a proportional system with 5% minimum threshold in a 200-member assembly

Party	*Percentage of the vote*	*Initial seats (% Vote X 200)*	*Remaining seats (% Vote X (200 – 184))*	*Total seats*	*Percentage of seats*
A	42	84	7	91	45.5
B	30	60	5	65	32.5
C	20	40	4	44	22
D	4				
E	4				
Total	100	184	16	200	100

Table 2 Territorial-plurality representation in a five-member assembly

District	*% Votes for candidates from party A*	*% Votes for candidates from party B*	*Winning candidate*
1	51	49	A
2	52	48	A
3	51	49	A
4	52	48	A
5	10	90	B
Total votes	43.2	56.8	
Total seats	80 (4 seats)	20 (1 seat)	

ties between a representative and her/his constituents because representation is tied to a specific geographic region and to a specific territorial constituency. All territorial plurality systems, however, suffer from the so called 'vote–seat' problem: the proportion of votes gained by a party as a whole may not correspond closely to the proportion of seats they receive in the legislature. **Table 2** illustrates how territorial representation can lead to such disproportionality in a five member assembly representing five districts with equal numbers of voters.

In this example, candidates from Party A win close elections in four of the five districts but the party loses the fifth district badly. The four narrow margins mean, however, that Party A wins 80% of the seats in the assembly with just over 40% of the total vote. (Under PR, Party A would win just two of the five seats or 40%.) Departures from proportionality can be even more severe if three or more parties contest each district and winning candidates only achieve pluralities rather than majorities. Perhaps the best known recent example of the vote–seat problem was the 2000 US presidential election, in which George W. Bush received fewer votes than Vice President Al Gore, but won the election with narrow victories in several states. Although the vote–seat problem is endemic to all territorial systems of representation, it is of particular concern when it leads to the exclusion of political (gender, ethnic, or religious) minorities.

Territorial systems are also highly vulnerable to gerrymandering, a technique that influences the vote–seat ratio by manipulating the boundaries of electoral districts, boroughs, or ridings.

Variations of the single member, first past the post system can either ameliorate or exaggerate the vote–seat mismatch. The voting system in multimember or at large districts (where candidates run for several seats simultaneously) can either completely exclude minority parties and candidates, or can provide a mechanism for proportionality. In the former case, election rules might require candidates to run for particular seats, and require voters to cast one and only one vote for each seat. A cohesive political plurality could then elect all representatives. Alternatively, candidates might all run against each other simultaneously with the top vote getters winning election. Such an arrangement prevents the exclusion of a political minority, but allows the election of candidates who have a relatively narrow base of support. If the top candidate attracts 70% of the vote, for example, the remaining representatives would be elected with less than 30% support among the electorate. Nor do such systems guarantee proportionality since (as in the above example) a single candidate might win the lion's share of the vote.

Numerous electoral mechanisms have been proposed and employed to overcome the seat–vote mismatch and to address the problem of minority exclusion. The single transferable vote (STV) (called instant run off voting or IRV when used to elect candidates to a single seat) identifies the candidates with the broadest support by having voters rank them in order of preference. Ballots that would otherwise be wasted (either because the first choice candidate received more votes than needed for election, or too few to be competitive) are transferred to the voter's second (third, etc.) ranked candidate until all seats are filled. STV attempts to ensure a closer match in proportionality between votes and seats while still retaining a territorially based system of representation.

Mixed Electoral Systems

States and other jurisdictions use a wide array of proportional and territorial electoral systems, and scholars have proposed an even greater variety to address the practical and theoretical shortcomings of the existing ones. PR and territorial plurality systems are not necessarily mutually exclusive, however, and many jurisdictions use some combination of them. Some seats in a legislature, for example, might be elected from territorial units while the remaining seats are assigned by proportion of the total vote. Proportional systems may also be used within a set of districts, where parties designate separate lists for each territorial unit. The effect on proportionality in these cases depends on the particular design of the electoral system and on the distribution of votes in any given election. Mixed systems often reflect practical compromises among political factions who would benefit disproportionally from either PR or territorial representation. As such, they embody mixed principles and concepts of representation.

Principles of Representation in Electoral Systems

Arguably, PR can promote both descriptive and substantive representation more easily than territorial representation because it is simpler to regulate the composition of party lists than the composition of territorial electorates. It is easier, for example, to require that women or minorities constitute some minimum percentage of the candidates on party lists, than to create electoral districts where the same percentage of women or minorities are likely to be elected. PR used with such regulated party lists can thus easily guarantee descriptive representation in an elected body. The profusion of parties in a PR system may also increase the chances that a particular constituency or interest group will have representatives in the assembly.

PR (especially without regulated party lists) suggests a bias toward the independent model and substantive representation because representatives generally have weaker ties to a particular constituency – even one that is concentrated geographically. Representatives are not personally tied to a set of voters because voters cast ballots for a party rather than individuals. Representatives thus embody the interests and positions of the party, rather than a specific constituency *per se*, and are consequently freer to act according to their own judgment (within the limits of party discipline) and are (arguably) less accountable to voters.

Although territorial representation is typically less effective in creating proportionality and descriptive representation, district systems ensure a strong tie between representatives and their constituencies, and in principle increase accountability. Such a system encourages a mandate model because voters can remove specific incumbents in favor of candidates who will more accurately represent their wishes. More generally, territorial systems allow for a distinct concept of the object of representation: the geographic community. Rather than representing the agglomerated will of individual voters (as in PR), a representative comes from and embodies a particular, identifiable community that is – or is imagined to be – more than the sum of its individual members.

Responsiveness and Stability in Electoral Systems

Electoral systems must also balance responsiveness and stability. These refer to the relationship between a change in voting patterns and the consequent change in the elected assembly. PR systems are (in theory) perfectly responsive; a 5% increase (or decrease) in the votes for a party will translate into a 5% increase (or decrease) in the number of seats held by that party. In practice, changes in vote totals may not translate perfectly into changes in seat totals because of threshold requirements and limited assembly size.

The balance between stability and responsiveness is more complex in territorial representation because the vote–seat ratio of a particular set of electoral districts can produce outcomes that are either too stable or too responsive. In the former case, even large changes in the overall vote total do not change the composition of the elected assembly, whereas in the latter case, small changes in the vote produce huge swings in the legislature. Overly stable systems result when parties have either large majorities or small minorities in districts. Even relatively large losses or gains by one party or another in total votes are not sufficient to change the majorities or minorities in many districts, so the composition of the assembly remains unchanged. In contrast, a set of districts with very narrow margins may be overly responsive because a tiny shift in the electorate may cause one party to lose every seat. A particular set of electoral districts may also be asymmetrically responsive: a large loss of votes by one party may not produce any change in the assembly, but a small gain in votes may result in significant gains.

Stability and responsiveness raise final dilemma of territorial representation. The principle of accountability holds that voters should be able to remove an incumbent from office, and elected officials are presumably more responsive to their constituents if they are in danger of losing the next election. At the same time, legislatures benefit from some degree of political continuity and stability. Thus a situation that is desirable in an individual district – accountability ensured by narrow election margins – may produce excessive responsiveness when applied in every district.

Political Representation and Political Communities

Concepts and principles of political representation are fundamental not only to the structure of electoral systems, but also to the relationships among the franchise, the polity, and citizenship. In nation states, the right to vote is one criterion of membership in the nation, but in all states, this right is a measure of full citizenship. Full citizenship, including the franchise, confers membership in 'the people'. A full citizen therefore carries both the right of sovereign power and political interests that require representation. In short, only full citizens are full participants in the political life of a state. In this way, acts of political representation create political communities rather than simply reflecting the will of preexisting ones. Consequently, analyses of political representation ultimately involve issues of citizenship, exclusion, and power. While systems of territorial representation present clear geographic questions, all systems of political rep resentation – even pure PR – raise significant issues for political geographers.

See also: Citizenship; Democracy; Electoral Cartography; Electoral Districts; Electoral Geography; Gerrymandering; Governance; Identity Politics; Private/Public Divide.

Further Reading

Forest, B. (2001). Mapping democracy: Racial identity and the quandary of political representation. *Annals of the Association of American Geographers* 91, 143 166.

Johnston, R. (2002). Manipulating maps and winning elections: Measuring the impact of malapportionment and gerrymandering. *Political Geography* 21, 1 31.

Johnston, R., Pattie, C., Dorling, D. and Rossiter, D. (2001). *From votes to seats: The operation of the UK electoral system since 1945*. Manchester: Manchester University Press.

Keyssar, A. (2000). *The right to vote: The contested history of democracy in the United States*. New York: Basic Books.

Lijphart, A. (1999). *Patterns of democracy: Government forms and performance in thirty six countries*. New Haven: Yale University Press.

Lublin, D. (1997). *The paradox of representation: Racial gerrymandering and minority interests in Congress*. Princeton, NJ: Princeton University Press.

Marston, S. A. (1990). Who are the people?: Gender, citizenship, and the making of the American nation. *Environment and Planning D: Society and Space* 8, 449 458.

Monmonier, M. S. (2001). *Bushmanders and bullwinkles: How politicians manipulate electronic maps and census data to win elections*. Chicago: University of Chicago Press.

Morgan, E. S. (1988). *Inventing the people: The rise of popular sovereignty in England and America*. New York: W. W. Norton and Company.

Phillips, A. (1995). *The politics of presence*. Oxford: Clarendon Press.

Pitkin, H. (1967). *The concept of representation*. Berkeley: University of California Press.

Relevant Websites

http://www.fairvote.org/
A site with many election related resources, sponsored by the Center for Voting and Democracy.

http://www.electoralgeography.com/new/en/
A site with many worldwide election results and electroal maps, especially of recent elections.

http://www.redistrictinggame.org/
An interactive redistricting site from the Annenberg Center at the University of Southern California.

Polycentricity

P. Hall, University College London, London, UK

Glossary

Edge City A 'new downtown' at the edge of an urban area; generally American.

Functional Polycentricity Polycentricity in terms of the functions performed by an individual urban unit in a system of cities; such a unit can, at least in part, perform independently of the First City or cities in that system.

Geographical Polycentricity Polycentricity in terms of the existence of independent and separate urban units in a system of cities; they may also exhibit a flat hierarchy, that is, there is no dominant city or cities.

Mega-City Region A region including a number of separate cities or urban units which are functionally related.

Morphological Polycentricity See Geographical Polycentricity.

Multiple Nuclei A city with more than one commercial nucleus, over and above the Central Business District (CBD).

Polycentricity An urban system (city; city region) with more than one constituent unit.

Space of Flows The pattern of flows of information between units or agents in an urban system.

Introduction

The term 'polycentricity' is widely used in urban academic literature, and in Europe it has become embodied as a policy objective in the 1999 European Spatial Development Perspective (ESDP). But there is considerable confusion about both its meaning and its policy significance. This article seeks to unravel the concept, distinguishing first, different 'scales' of polycentricity: local (urban), regional, national, and continental; second, different 'kinds' of polycentricity – geographical or morphological, and functional.

Scales of Polycentricity: Local, Regional, National, and Continental

It is important to appreciate that the concept of polycentricity may operate at different spatial or geographical scales, and that polycentricity at one scale may mean monocentricity at another. Polycentricity was first used at the 'intra urban' or 'local' scale: in the work of Chicago sociologists and land economists, from the 1920s onward, as an analytical device to capture the structure of urban areas, it referred to the possibility of additional urban centers or subcenters which might supplement or even compete with the role of the dominant central business district, whether for retail or commercial office development. But the early descriptive models of urban activities and land uses, like the ring or annular model developed by the sociologists Ernest Burgess and Robert Park or the sectoral model developed by the land economist Homer Hoyt, all assumed a fundamentally monocentric pattern dominated by a single strong central business district (CBD); the sole exception was the interesting multiple nuclei model developed in 1945 by the Chicago geographers Chauncy Harris and Edward Ullman. Since then, though several basic urban models have been developed using a monocentric basis (such as those of William Alonso and Richard Muth, in the 1960s), observers like Kevin Lynch and Catherine Bauer Wurster simultaneously noticed that cities could develop, and indeed were developing, a more polycentric structure with the emergence of multiple business centers (in London: the City, Westminster/West End and Canary Wharf; in Paris: the Étoile Champs Elysées and La Défense; in New York: Downtown, Midtown, and Battery Park City; and in Tokyo: Otemachi Marunouchi and Shinjuku). Indeed, as early as the 1960s it was suggested that western American cities like Los Angeles or Phoenix, which had grown up largely in the age of mass automobile ownership, were developing a distinctly polycentric urban structure with many subcenters rather than a single dominant center. More recently the American journalist Joel Garreau has observed the phenomenon of such new downtowns, or Edge Cities, developing in suburbs such as Tysons Corner outside Washington DC.

Simultaneously, from the 1960s, geographers and planners noticed that polycentricity could also occur at a 'regional' scale: regions such as Randstad Holland in the western Netherlands, or Rhine Ruhr in Germany, exhibited a structure wherein there was no dominant city, but rather a number of cities of roughly equal size and importance (Amsterdam, Rotterdam, the Hague and Utrecht; Bonn, Cologne, Düsseldorf, Duisburg, Essen, Bottrop, and Dortmund); further, such an arrangement could have real advantages in urban functioning, since it avoided excessive concentration of activities and associated congestion, while allowing different cities to specialize within a spatial division of labor. (This represented a considerable distortion, even contradiction, of the central place system first identified by the German geographer Walter Christaller in 1933, which posited a regular spatial distribution of a hierarchy of smaller cities around a larger central city.) At the same time, urbanists noted that the same principle applied to the 'national' urban structure of these countries: the Netherlands had never had a single dominant city like London or Paris, while Berlin in Germany had lost much of its importance due to the post war division of both the nation and the city. Even after reunification and the return of the capital from Bonn to Berlin, the German urban system remains characterized by strong provincial capital cities (Hamburg, Bremen, Hannover, Frankfurt, Stuttgart, Munich, Leipzig, Dresden, and others) reflecting the fact that the nation emerged only in 1871 through the reunification of many smaller principalities. Again, this pattern has been seen as offering decided advantages in avoiding excessive congestion and, equally important, in the diffusion of metropolitan features (culture, the media) among a number of cities rather than concen tration in the capital.

Even more recently, within Europe – where the question has aroused the most intense policy discussion – polycentricity has been defined at a 'European' or 'continental' scale: 'polycentricism' in the 1999 ESDP meant actively promoting alternative centers, outside the so called central Pentagon bounded by Birmingham, Paris, Milan, Hamburg, and Amsterdam, into Gateway cities elsewhere in Europe, many of which are national, political, or commercial capitals serving wide territories in Ireland, the Iberian peninsula, Scandinavia, and east ern Central Europe. But paradoxically, this may help promote greater 'monocentricity' at the 'national' scale in these more peripheral European countries, as capital and labor are drawn to leading cities, thus creating regional imbalances between core and periphery within each country: a situation observable in and around Dublin, Lisbon, and Madrid in the 1980s and 1990s, and around eastern Central European cities like Budapest, Prague, Warsaw, and Tallinn in the 2000s. As a further compli cation, in the largest such national capital regions – especially in the Pentagon – population and economic activity may decentralize at a 'regional' scale from the First City (the largest city in each region, a term chosen to avoid the term 'primate') to other urban places, as evident in the formation of polycentric mega city regions (MCRs) like Southeast England, Central Belgium, Rhine Main, or Northern Switzerland. Such regions, first recognized in the mid 1990s in Pacific Asia, and later in the work of the geographer Allen Scott, essentially consist of a number of discrete cities and towns, phys ically distinct but networked together. In this process, though decentralization of population may commonly entail actual physical migration, economic activities decentralize much more through a process of differential growth and/or decline: empirical observation shows that even when First City employment grows, the rate of growth in other cities is much higher, partly because much of it takes the form of growth of residentiary services to cater for population growth. In addition, however, the evidence suggests the progressive emer gence of a spatial division of labor within the region: higher level services, particularly those networked into global systems of information exchange, continue to grow in the First City because of its unique advantages for face to face exchange, while lower level service func tions (generally with a more regional orientation) dis perse out from the highest order central city or cities to lower order cities. Thus, an extremely complex regional structure of cities may emerge: in the extreme case noticed in Europe, Southeast England, no less than 51 cities and towns with their surrounding urban com muter fields form an MCR, as depicted in **Figure 1**, with some 19 million people stretching up to 160 km (100 miles) from Central London.

Two Kinds of Polycentricity: Geographical (Morphological) and Functional

Geographical or Morphological Polycentricity

A polycentric region, or nation, or continent, by defin ition, is a geographical unit consisting of more than one urban place. Polycentricity thus may simply be 'geo graphical' or 'morphological', in the sense that physically separate urban units exist. As noted, a region such as the Dutch Randstad and the Rhine Ruhr area of Germany, or a country such as Germany, were identified as poly centric in this sense from the 1960s. But the clear im plication was, and is, that such regions are characterized not merely by a number of different cities, but also by a weak or flat hierarchy: there is no single dominant city, but rather a set of cities within a fairly restricted size range. One well known way of measuring this is the rank–size rule (see **Figure 2**): if cities and towns forming part of a system (which can be a continent, a nation, or a region) are arrayed on double logarithmic paper, by rank on the x axis and by (population) size on the y axis, then a truly polycentric system would produce a distribution along a line at 45° to both axes, the largest place having a population double that of the next, and so on, without evidence of primacy at the top of the distribution. Analyzed in this way, the city systems of countries like Germany and the Netherlands have indeed been found to be polycentric in comparison with those of the United Kingdom or France. More recently, analysis of 'polycentric' city regions in Northwest Europe has confirmed that some (the Paris region, Dublin) are

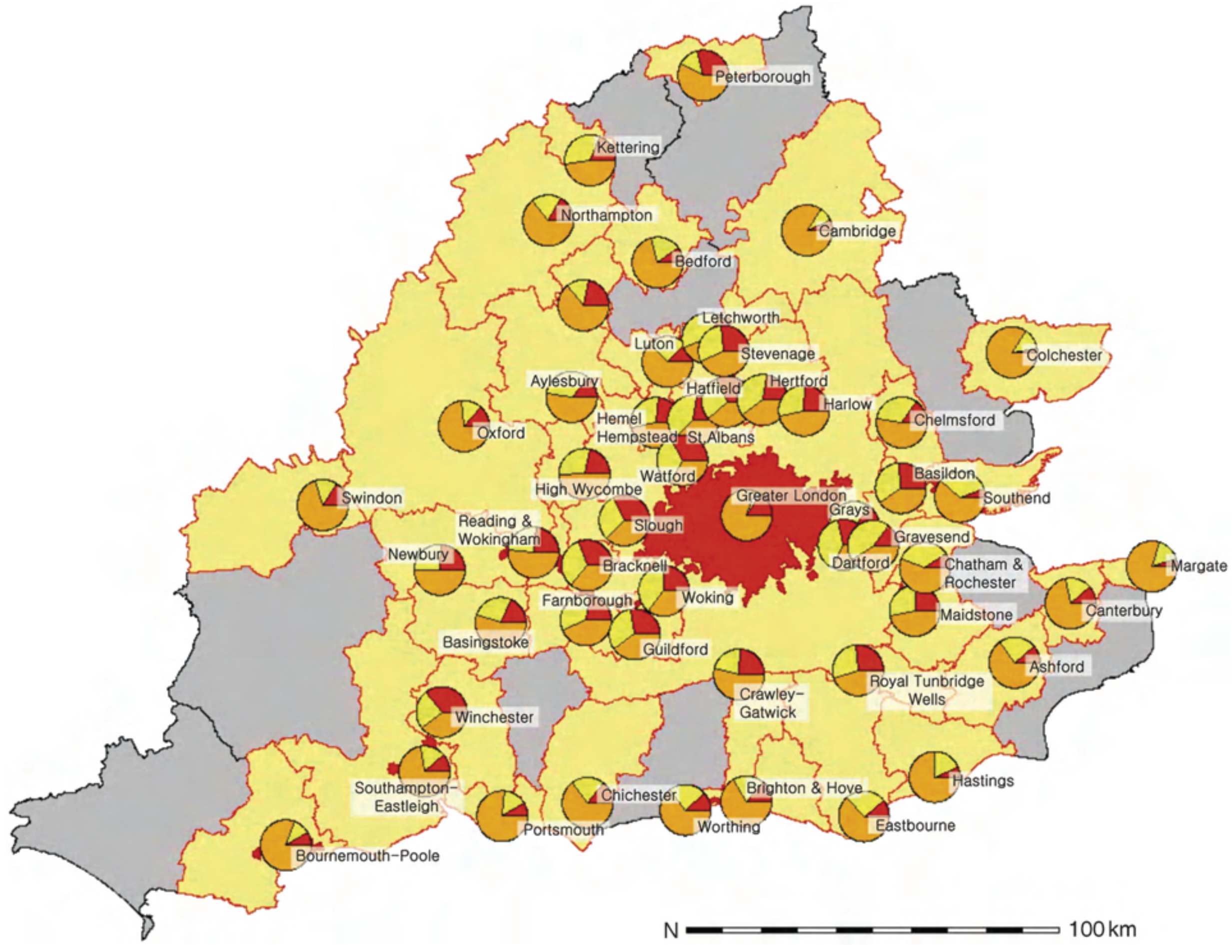

Figure 1 Southeast England mega-city region, 2001: Constituent functional urban regions (FURs) and degree of self-containment. The region, with 51 constituent functional urban regions (FURs) all, apart from London, medium-sized or small shows a very high degree of self-containment beyond a radius approximately 70 km (40 miles) from central London: here, FURs are 75 85% self-contained, in that the great majority live and work within the same local commuter region. In-commuting, out-commuting, live & work. *Source*: Hall, P. and Pain, K. (2006). *The polycentric metropolis: Learning from mega-city-regions in Europe.* London: Earthscan.

relatively monocentric as compared with the Randstad or Rhine Ruhr; however, even the Randstad (together with Rhine Main and Northern Switzerland) is more accurately described as semiprimate, whereby one or more dominant cities are superimposed on a predominant lognormal distribution.

Functional Polycentricity

This however relates purely to the physical morphology of a national or regional system of cities, as measured superficially by size; it says nothing about how the constituent urban units relate to each other. It is more significant to seek to identify 'functional' polycentricity, whereby such units interrelate in complex ways, through exchanges of people and goods and information. Measuring these relationships obviously depends on the quantity and quality of the available information to be measured.

One of the commonest ways of measuring functional polycentricity, because data are readily available for most countries and most times, is through patterns of 'daily commuting'. Since the 1950s, American urban researchers have used and successively refined a concept now known as the metropolitan statistical area (MSA). This is essentially based on a central city (or cities), plus a surrounding ring from which significant numbers of commuters (typically 15% or more of the resident workforce) travel daily to that city. Such a concept which uses functional urban regions (FURs) as the basic statistical building blocks has been used widely in the UK and the European urban studies since 1970.

Two recent European studies of polycentric city systems have adopted semi arbitrary definitions. In the first, ESPON, potential urban strategic horizons (PUSHs) were defined on the basis of contiguous 45 min isochrones from the center of the leading city; in the second, POLYNET, a Polycentric MCR was defined simply on

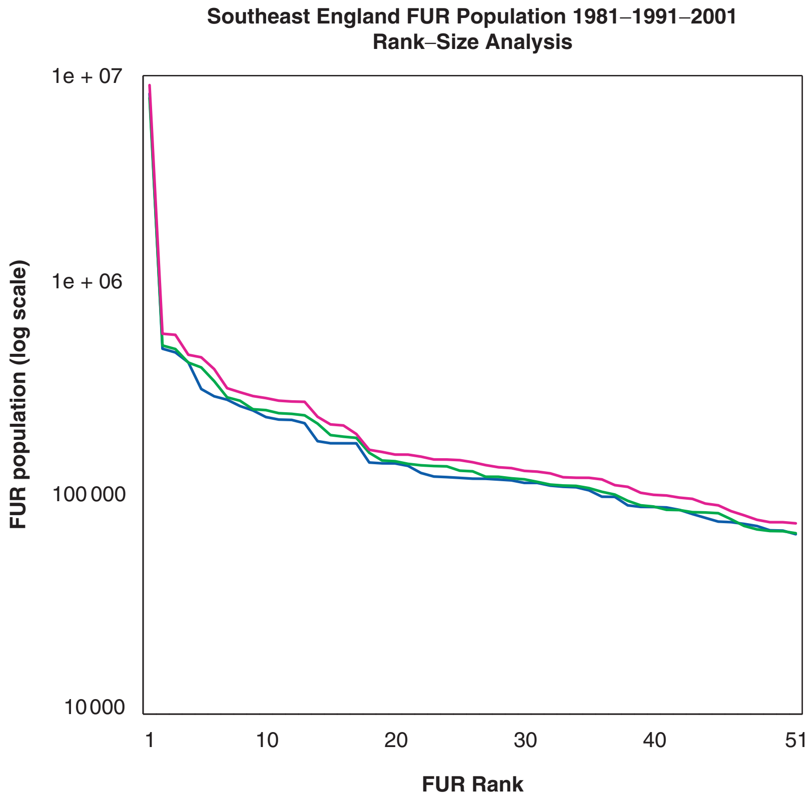

Figure 2 Rank -size plots, Southeast England and Randstad Holland, 2000-02. comparison of the graphs of two MCRs in Northwest Europe, in which population size of constituent FURs is plotted against the rank order on double-logarithmic paper, shows two distinctively different patterns: in the Randstad, an almost pure rank size relationship with points arrayed in a regular sequence; in Southeast England, London dominance superimposed on a rank size relationship at lower points in the hierarchy. FUR populations in 1981 ■, in 1991 ■, and in 2001. ■ *Source*: Hall, P. and Pain, K. (2006). *The polycentric metropolis: Learning from mega-city-regions in Europe*. London: Earthscan.

the basis of contiguous FURs, without any evidence (positive or contrary) of relationships between these FURs, in order then to permit statistical analysis of commuting data to identify possible functional relationships. First, commuter flows were mapped and inspected to identify strong relationships that 'bypassed' the First City or cities. Such flows do appear in some regions, sometimes unexpectedly. In Southeast England, where London attracts remarkably strong commuter movements from quite distant places up to 140 km (75 miles) away, there is a west–east split: the area west of London shows a strong superimposition of criss cross flows bypassing the capital, while the area to the east (which admittedly is bisected by the Thames estuary) shows no such pattern. In the Dutch Randstad, there are relatively weak flows between the northern 'wing' (Haarlem–Amsterdam–Utrecht) and the southern 'wing' (Leiden–Den Haag–Rotterdam–Dordrecht), and also between areas in Brabant south of the Randstad and the Randstad itself. Second, the study analyzed the relative self containment of the individual FURs. It found that, beyond some critical commuting distance from the First City in each region – typically about 120 km (75 miles) – FURs became quite strongly self contained, with some 75–85% of working people living and working in the same FUR. The study then used commuting data to produce a 'general polycentricity index', defined as follows:

The formula for 'special functional polycentricity' is

$$P_{SF}(N) = (1 \quad \sigma_\partial/\sigma_{\partial\max}) \cdot \Delta$$

where P_{SF} is special functional polycentricity for a function F within network N; σ_∂ is the standard deviation of nodal degree; $\sigma_{\partial\max}$ is the standard deviation of the nodal degree of a two node network (n_1,n_2) derived from N, where $d_{n1} = 0$; and $d_{n2} =$ value of the node with highest value in N. Δ is the density of the network.

The expression for 'general functional polycentricity' is:

$$P_{GF}(N) = [(1 \quad \sigma_\partial/\sigma_{\partial\max}) \cdot \Delta]/n$$

where P_{GF} $(N_1, N_2, \ldots, N_n)$ is general functional polycentricity for functional networks N_1, N_2, $\ldots$, N_n; P_{SF} is the sum taken over all; and n is the number of networks.

Remarkably, it found very weak polycentricity in all eight regions under study: the index varied from only 0.02 for the traditionally ‘monocentric’ Paris region, to 0.15–0.20 for the traditionally ‘polycentric’ regions of the Randstad and Rhine Ruhr. But it noted that any figure higher than that might imply a pattern of cross commuting so extensive that the region would become operationally dysfunctional and environmentally unsustainable. Indeed, previous Dutch research as well as German and British comparisons suggest that polycentric urban regions in Northwest Europe may exhibit features that conflict with ESDP sustainability objectives. This conclusion is significant, because the ESDP (1999) has as a central policy objective, the active ‘encouragement’ of polycentricity, or ‘polycentrism’.

Measuring the ‘Space of Flows’

To measure what Castells calls the ‘space of flows’, however, different data and different methods are required. One is to adopt the quantitative method developed by the Globalization and World Cities (GaWC) Study Group at Loughborough University to measure the connectivity between cities through study of cross border service networks across multiple city based offices worldwide. This, further, is capable of measuring connectivity at four different geographical scales: regional, national, European, and global. But, it is important to stress, it cannot measure the strength or importance of the ‘actual’ flows of information between the towns and cities of a polycentric region, or between one such region and another. Such information – virtually by e mail, or telephone, or in people’s heads through business travel – requires the direct capture of vital primary data on the actual flows of information passing within and between the MCRs. But this proves difficult: email traffic in particular is increasingly infected by spam and phishing; and it proves extremely difficult to obtain statistically significant samples of either e or personal traffic, either directly (for reasons of commercial confidentiality) or by asking respondents to maintain records. More significantly still, however, the problem is that such surveys of this type can never hope to convey the ‘quality’ of the information that is exchanged, either in e traffic or in personal business traffic; quantitative measurement has definite limits, both of practicability and of the quality of the output.

Given these limitations, the limited research on information flows in such MCRs does suggest that some are functionally more polycentric than others: in them, a proportion of all information exchanges (telephone calls, e mails, business trips) bypass the First City of the region, suggesting that other centers have a degree of functional independence. Thus, in Europe, Southeast England, Central Belgium, Randstad Holland, and Rhine Ruhr appear to have greater functional polycentricity than the Paris region, Greater Dublin, or the Rhine Main region. This finding, albeit based on somewhat fragmentary data, may be significant for strategic planning of such regions because it suggests that perhaps such independence may be induced.

This research concerns information flows within the advanced producer services that increasingly dominate the urban economies of nations and cities in the developed world, and that have underpinned polycentric urban structures recently researched in Europe (ESPON, POLYNET) and in the United States. It may however, provide an equally important clue to the interpretation of economic structures in the even larger MCRs of Pacific Asia, where the phenomenon was first recognized, best exemplified by the Pearl River Delta (Hong Kong–Guangzhou) and Yangtze River Delta (Shanghai–Suzhou–Nanjing–Hangzhou) regions of China, which function as complex networked systems of manufacturing. Similar regions in Latin America, such as the Golden Triangle of Brazil (Rio de Janeiro–São Paulo–Belo Horizonte), may represent a mixture of networked manufacturing and producer services. The underlying principles may be the same: production, whether of goods or of services, is organized through the First City which operates as a global gateway to and from the rest of the world; the other cities are in a sense, tributary to it, but may over time acquire a degree of functional independence from it.

See also: Space I; Space II; Space-Time.

Further Reading

Davoudi, S. (2003). Polycentricity in European spatial planning: From an analytical tool to a normative agenda. *European Planning Studies* 11, 979 999.

ESPON (2005). *Espon 1.1.1: Potentials for polycentric development in Europe: Project report*. Stockholm: Nordregio.

European Commission (1999). *ESDP: European spatial development perspective: Towards balanced and sustainable development of the territory of the European Union*. Brussels: European Commission.

Garreau, J. (1991). *Edge city: Life on the new frontier*. New York: Doubleday.

Hall, P. and Pain, K. (2006). *The polycentric metropolis: Learning from mega city regions in Europe*. London: Earthscan.

Harris, C. D. and Ullman, E. L. (1945). The nature of cities. *Annals of the American Academy of Political and Economic Science* 242, 7 17.

Lang, R. E. and Dhavale, D. (2005). America’s megalopolitan areas. *Land Lines: Newsletter of the Lincoln Institute of Land Policy* 17(3), 1 4.

McGee, T. G. (1995). Metrofitting the emerging mega urban regions of ASEAN. In McGee, T. G. & Robinson, I. (eds.) *The mega urban regions of southeast Asia*. Vancouver: University of British Columbia Press.

Scott, A. J. (ed.) (2001). *Global city regions: Trends, theory, policy*. Oxford: Oxford University Press.

Polyvocality

Saraswati Raju, Jawaharlal Nehru University, New Delhi, India

Glossary

Gaze The gaze refers to how an audience views the visuals. It was made popular by French intellectuals, for example, Michel Foucault around the 1960s, in postmodern philosophy and social theory, particularly in feminist scholarship; however, it is not simple viewing as the gaze is seen as imbued with socially constructed meanings and biases.

Homo Economicus It refers to man as economic man who is rational, informed, and interested in the self alone, whose yearnings are for wealth, who avoids unnecessary labor, and who is able to make judgments toward those ends.

Positionality It is a person's position not as an autonomous self but as shaped by multiple as also overlapping social processes embedded in historical-geographical locations.

Situated Knowledge As against the concept of universalistic knowledge, situated knowledge acknowledges social, relational, and contextual specificities of knowledge formation. Thus, no particular knowledge takes precedence over the other.

Defining Polyvocality

'Poly' is usually a prefix meaning 'much', 'many', 'more than one', or 'of many kinds or parts'. 'Vocal' refers to 'uttered or produced by the voice', 'having a voice, capable of speaking or making oral sounds', 'full of voice or voices', or 'speaking freely'.

Interestingly, 'poly' in combination with quite a few words occupies several columns in any standard dictionary and yet there is no such term cited as 'poly vocal' in these works. Similarly, the term vocality does not appear in the *Webster's Dictionary* of the American English nor does it find place in the latest edition of the *Cambridge Dictionary*. *The Dictionary of Human Geography* also does not carry an entry on 'polyvocality' as is also the case with *A Feminist Glossary of Human Geography*, although the latter has an entry on 'polyphonic'. It is not surprising, therefore, that the 'google' search for the term 'polyvocality' *per se* yields no better result.

Simply put, 'polyvocality' means multiple voices representing multiple viewpoints and discourses coex isting without privileging one voice over the other. This is contrary to 'monovocal' style of the narration – often seen as single voice of omniscience – wherein the singular voice not only subsumes but also overrides and subordinates other voices; it also becomes hegemonic, justified, and legitimized in lending authority to a par ticular viewpoint as dominant, concealing, and often manipulating, in the process, diverse and conflicting voices and realities.

Acknowledging Polyvocality

It may be said that the polyvocality essentially emerged and sustained on the terrains of various modes of thought and intellectual practices of postmodern philosophy, and the subsequent spatial turns in social theory that inform much of the contemporary geographical thoughts. A brief on postmodernism philosophy is, therefore, in order.

One of the precepts of what is understood as modern worldview was the historical moment called the En lightenment in the postmedieval era of darkness. Being modern stood by and large for having rational and objective viewpoints – the 'enlightened' way of seeing the world as orderly, organized, and causally generalized. The central concern therein has been the public use of reason to change human society and to demystify the world as relatively homogeneous with quite precise temporal definition – in a way, a world where the *Homo economicus* responds to certain impulses as a rational human being.

In such pursuits, modernist scholars tried to transfer physiological properties of organized matter, the forms and conditions of existence of life which usually follow definitive evolutionary track to the realm of social. They did so in terms of looking for spatial localization of function, hierarchy, regulation, and norm in society similar to that found in the body. This was, in part, a product of what has been called as particular 'practices of reason'. It was in this genre of scholarship that one can place Brian Berry's 'equifinality' that he so infamously conceptualized while discussing about Chicago and Calcutta acquiring over time identical profiles in terms of status based residential ordering, among other characteristics.

It was against such overarching meta narratives, the 'grand theory', the 'universal truth', sometimes called 'god's eye view' that postmodernism arose, not as replacement, but as reaction, skepticism, and resistance, initially in the domain of architecture and planning, linguistics, literature, and philosophy and subsequently spreading to theology, the arts, and medical and social sciences, including geography.

Although postmodern positions are highly contested and disputed, there is a general consensus that post modernism is not only about questioning meta narratives or the grand theories, but also about questioning the construct of human as rational being independent of social and cultural embeddedness, completely over whelmed by hegemonic structures.

Since modernism saw the world as essentially rational and orderly, production of 'truth' entailed use of methods through overarching theorization which could 'prove' that rationality and order without providing for any biases and/or persuasions of those seeing the reality. It was not surprising therefore that objectively produced knowledge that could stand scientific rigor, scrutiny, and validation took precedence over other forms of indige nously produced knowledge. That such objectivity is not only impossible but it may in fact be unnecessary and problematic as well was never a question. A certain hierarchy in knowledge production was inevitable. The so called 'representative scientific knowledge' thus becomes the knowledge of a 'chosen few', the privileged gaze of white Anglo Americans, essentially males – given the overwhelmingly androcentric nature of 'scientific' research – researchers with a view – what can be termed also as 'ethnocentric universality'.

Scholars began to point out that the so called objective statistics are essentially produced within the framework of preconceived categories in the minds of researchers, and however dispassionately this viewing of the field is done, it still remains a viewing from a vantage point. Representing the viewing is thus not free from the cultural baggage the researchers carry with them. Therefore, texts and landscapes which get represented are constructed, authored, and only partial and may not even be true.

The postmodern critique of the modernist view thus concerned the latter's singularity as well as its privileged position – the fact that it allowed for only one truth and one story, whereas there can be many truths/stories and even one truth/story can be told in myriad ways. That is, there can be no hegemonic or monolithic text; no objective truth to discover, but many voices – articulated individually as well as collectively – to hear.

Polyvocality and Its Implications for Research

As the overarching theories were essentially architected on the basis of masculine views as the norm, much of the initial discourses on polyvocality in geography were particularly informed by feminist scholarship who, in fluenced by concurrent debates in other disciplines, have been questioning the androcentrism of geographical enquiries and the links between theory, power, and knowledge, among other issues.

Acknowledging the presence of polyvocality also warranted researchers' sensitivity and responsibility toward opening discursive spaces to 'others', especially those whose vulnerabilities were such that their voices were routinely marginalized or suppressed both in thought and in social practices by elite compromises. It also meant celebrating differences, bringing upfront multilocalities as well – struggles, experiences and voices of women, people of color, queers, subalterns, and so on. Importantly, it was also pointed out that representational texts may be read variedly, what Spivak called 'the pleasure of the bottomless', in the process of which the perceived certainties are decentered and the authority by which all knowledge claims are made is questioned.

Scientific research looking for neatness and order had devalued personal experiences and histories as legitimate sources of informed enquiry, which stripped the 'per sonal' from writing and research. As against such presumably detached objective authors 'at a distance' dispassionately capturing the truth, polyvocality neces sitated active engagement with the subjects – blurring the typically unequal power relations/boundaries between researchers and the researched. However, interlinked were the issues regarding authenticity or inauthenticity of 'representational' voices – who 'speaks to' or 'speaks for' and who 'represents' and 'interprets' these voices – these were brought forth and intensely debated. It was pointed out, for example, that the presumable com monalities shared among women (and men) researchers – often from academia – can get fractured because of their privileged positionality vis à vis the researched in the field resulting in 'otherness', which may be as important as their gendered identities.

Polyvocal voices necessarily implied polylocalities as well, spatialities of experiences – 'speaking from' and 'speaking of' acquired new meanings. Since 'represen tational' voices would necessarily be contextually an chored, it was emphasized that the 'personal' into public discourse be admitted. This required that the authors make their personal, intellectual, social, and political locations explicitly known. That is, the author could no longer be absent from the text, but makes the authorial presence felt, in order to understand the author's gaze. Discourse analysis techniques that expose knowledge and power by deconstructing texts, institutions, and landscapes have also become commonplace. Polyvocality brought a new vocabulary such as the 'other', situated ness of knowledge, positionality, diversity and differ ences, and the deconstructions became the hallmark of all discourses. What is often called 'emancipatory geographies' no longer refers to doing research about others but to doing it in a participatory mode with the subjects.

With the myth of absolute objectivity – the 'gaze from nowhere' – busted, it was increasingly recognized that the 'truths' do not need validation and legitimization through data, mostly quantitative in nature, of the sort as in hard sciences, and that they can sustain on localized mix of cultural politics, knowledge, and power and can be obtained through equally legitimate means of qualitative or ethnographic research methodologies. These meth odologies include detailed interviews, case studies, focus group discussions, oral histories and personal narratives, and participants' observations, etc., besides collaborating with participants in the development of research ques tions, the interpretation of data at both the descriptive and interpretive levels, and the writing of research reports.

Critical Interrogation of Polyvocality

Polyvocality questions universalization of theories and essentialization of meta narratives and thus rejects pri orities built around flawed conceptual categories and politics of representation. It has also generated critical debates about norms that constitute dominant referential universes. Yet the concern for polyvocality has its share of critiques. It is argued that polyvocality has created considerable anxiety among scholars about not being 'authentic' in trading alien territories to the extent that it has even created an occasional 'apology' by researchers for being 'the other' because they cannot be 'authentic' voices. Some have pointed out that such self reflexive exercises amount to mere 'navel gazing' intended to gain legitimacy as 'authentic researchers' and that such pressures and fears have led to an impasse increasingly reflected in avoidance of fieldwork by researchers.

Polyvocality deals with the question of representing voices, particularly marginal and subaltern voices, and yet serious researchers have become painfully aware that true or authentic voices cannot be retrieved, primarily on two accounts. First, experiences of the researched at an existential level are mostly private vis à vis even the most sensitively inclined researchers' theoretical frameworks, methods, and concepts that are primarily addressed to public domains. To that extent, the original narratives do get transformed into the researchers' overarching agenda. How then the researchers treat their own interpretative voices – as one among many or would they enjoy a privileged position? How does one differ entiate between the author's and the participants' voices? Even as the possibility of multiple voicing is accepted, each individual participant can be polyvocal. There is an issue therefore as to which and whose voice is authentic? Second, even as researchers recognize differences and try to include the multiple voices of subject and object within the text, their own identities continue to be a part of the very structures of power that are being deconstructed.

It has been argued that polyvocality can make it appear as if other voices had been allowed a say whereas, even in pluralistic narratives, the author/researcher always retains control over other voices, however sub consciously it may be. It is quite likely that in order to replace a modernist singularity with a multiplicity, the dominant continues to occupy the center without adequately recognizing the reality of multiplicity of others with their own trajectories.

Speaking for others may often be value laden and epistemologically violent – a way also to overcome guilt for being privileged. The most ideal situation is that the researched should speak for themselves, but what if they cannot speak in the language that can be heard and responded to by those who matter? Do the privileged remain silent even if their speaking, however tinted and biased their voices might be, makes a difference?

Polyvocality insofar as it mutates collectives in stand alone singularity of perspective has been severely cri tiqued, especially by scholars working with people of color or in non Western contexts as politically debili tating fragmentation. Apart from the fact that the prob lem of voice is a vexed problem of multiplicity as well as a problem of representation, the major argument is how it undercuts any basis for collective action. Rooting for polyvocal postmodernism, these scholars argue that in stead of playing out identity politics drawn on differences and multiple voices, the differences can be selectively appropriated. This is to build strategic alliances resulting in multiplicity of resistance toward the desired political end in recognition that certain of the differences reflect and reinforce broad structures of social domination that need constant contestation and overturning rather than one grand dismissal of singular voice. Thus, subsequent discourses on multiplicity of subjects and voices tend not to seek deconstruction of subject along differences as much as reconstruction of a critical subject.

See also: Embeddedness; Eurocentrism; Feminist Methodologies; Postmodernism/Postmodern Geography; Subaltern.

Further Reading

Arber, R. (2000). Defining positioning within politics of difference: Negotiating spaces 'in between'. *Race Ethnicity and Education* 3(1), 45 65.

Bell, D., Binnie, J., Cream, J. and Valentine, G. (1994). All hyped up and no place to go. *Gender, Place and Culture* 1, 31 47.

Christopherson, S. (1989). On being outside 'the project'. *Antipode* 21, 83 89.

Clifford, J. (1988). *The Predicament of Culture*. Cambridge: Harvard University Press.

Demeritt, D. (1996). Social theory and the reconstruction of science and geography. *Transaction of the Institute of British Geographers, New Series* 21, 484 503.

Gregory, D. (1994). *Geographical Imagination*. Oxford: Basil Blackwell.

Haraway, D. (1985). A manifesto for Cyborgs: Science, technology and socialist feminism in the 1980s. *Socialist Review* 80, 84 105. (Reprinted in Nicholson, L. (ed.) (1990). *Feminism/Postmodernism*, pp 190 233. London: Routledge.)

Harding, S. (1986). *The Science Question in Feminism*. Ithaca, NY: Cornell University Press.

Johnston, R. J., Gregory, D., Pratt, G., Watts, M. and Smith, D. M. (2000). *The Dictionary of Human Geography*. London: Blackwell Publishing.

McDowell, L. (1992). Multiple voices: Speaking from inside and outside 'the project'. *Antipode* 24, 56 72.

McDowell, L. and Sharp, J. P. (1999). *A Feminist Glossary of Human Geography*. London: Arnold.

Mohanty, T. C. (1984). Under Western eyes: Feminist scholarship and colonial discourses. *Boundary 2,* 12(3), 333 358.

Rodman, M. C. (1992). Empowering place: Multilocality and multivocality. *American Anthropologist, New Series* 94(3), 640 656.

Sutherland, E. (2004). In the field politics: Multiple voices of outworkers and researcher. *Australian Geographer* 35(2), 161 168.

Popular Culture

J. Dittmer, University College London, London, UK

Glossary

Geographical Imagination The way in which an individual imagines the world to be, both in terms of the relative positioning of places and also the characteristics and normative values associated with places.

Globalization The process by which global flows of information, goods, and people are accelerated by the time–space compression associated with changes in transportation technology, information technology, and regimes of trade and governance.

Hegemony A theory developed by Antonio Gramsci in which the capitalist state is protected from overthrow by the institution of middle-class culture and values as commonsense, thereby alienating the working class from their own culture and identity.

Material Culture Any tangible product that is rooted in cultural preferences.

Pop Art An Anglo-American visual artistic movement from the 1950s and 1960s that linked visual popular culture of the time with the abstract expressionist movement.

Post-structuralism A broad term for a shift in philosophy and cultural theory beginning in the 1960s that denied the existence of meaning outside of culturally specific frames.

Structuralism A broad term for the intellectual approach that seeks fundamental structures on which culture and society are built, and through which meaning can be derived.

Meanings and Origins

Popular culture is a contested term that is defined by commentators in myriad ways. In general, however, most definitions include within the term popular culture the whole spectrum of activities involving mass consumption. Thus, everything from cinema and television to the Internet and fashion magazines, as well as the advertisements promoting these other manifestations of popular culture, can be included in the definition of popular culture itself. To trace the genealogy of the term it is necessary to go back to the origins of 'popular culture', in which the term was used to contrast (derisively) the culture enjoyed by the working classes with the 'high culture' enjoyed by social elites. This distinction is critical to understanding the often denigrating use of the term popular culture in mainstream cultural critiques and also often, unfortunately, in academic discourse. However, the line between 'high' and 'popular' culture has become increasingly blurred with the general undermining of cultural authority structures throughout North America and Europe since the 1960s and 1970s. Indeed, now cultural authority is more often associated with those occupying roles in the popular culture industries, such as Hollywood, MTV, and *Rolling Stone* magazine than with 'high culture' journals such as *The Paris Review* or professors of film studies.

Of course, the relationship between 'popular' and 'high' culture is socially contested at all times, as the categories themselves are social constructions. In practice, the border between these forms of culture is highly malleable, with each pulling from the other for inspiration (most famously in the Pop Art movement of the 1950s and 1960s). Particularly difficult to categorize are cultural artifacts occupying the same media, such as 'art' films and Hollywood cinema, or romance novels and 'literature'. Despite the fact that categorization of cultural artifacts is ostensibly about the 'quality' of the artifact, categorization is usually made via the market mechanism: the more popular the artifact (in terms of sales), the lower the artifact is perceived on the spectrum of 'popular' and 'high' culture. This illustrates a central point to be made in any discussion of popular culture, which is that debates about popular culture have more to do with identity and the subject than they do with the actual cultural artifact itself.

Popular culture as a source of modern debate and anxiety can be traced to the rise of the printing press and other forms of mass media. Because of this multiplicity of media and messages, elites became concerned about the effects of media on the masses; soon they would learn to co opt those media and messages to their own purposes. Nevertheless, with the end of World War II, post war affluence in North America (and later in Western Europe) led to the rise of the consumption society, in which popular culture figured particularly strongly in young peoples' processes of identity formation. With allowances and other sources of income empowering young people, the market adapted to their preferences, providing entertainment deemed unacceptable by many parents who associated 'high' culture with their socioeconomic hopes and aspirations for their children and 'popular' culture with the underclass, ignorance, and crime.

Geographies of Popular Culture

Popular cultures vary over time and space, especially in their emphases on individualist traditional forms of culture (literature and visual art), mass media (such as cinema, television, radio, advertising, and periodicals), and participatory culture (sports, fashion, and speech). The scale of production for each of these popular cultures varies, from the corporate macroscale of the mass media, to the artisanal scale of the traditional forms of culture, to the embodied microscale of participatory culture. These scalar differences, however, are not absolute, with each form of popular culture being influenced by the others (e.g., embodied fashion is also influenced by corporate decision making and vice versa). These influences wax and wane over time, as the capitalist nature of most popular cultures leads to a constant need to rearticulate the popular and shift fashions to expand markets. Thus, popular cultures are often defined by their transience. Some locations emerge as privileged sites of this rearticulation, for example, Manhattan and Los Angeles in the United States, London in the United Kingdom, and Mumbai in India. These sites emerge out of historically contingent processes to dominate the macroscaled cultural production associated with the mass media. However, as described below, these sites of popular culture production often draw on culture from less privileged sites, as in the phenomena of denim jeans (from the American West) and reggae music (from the Caribbean).

Technological innovations over the past 100 years have enhanced the dissemination of popular cultures, both in terms of distance from their hearths and also in terms of the permeation of popular culture into society. This can be traced back to the invention of the printing press, which then enabled the creation of vernacular Bibles and the consequent changes in Christianity and increases in literacy rates. Since then, other forms of mass media have emerged, with similar effects. For example, periodicals, radio, television, and their associated advertising industries have revolutionized public and private spaces through the seemingly inescapable commercialization of everyday spaces. With technological advances carving out new media to be colonized by the popular culture industry, concerns over the content of popular culture have increased even as the possibility of separating oneself from popular culture has declined. The rise of Internet culture on the backs of the various technological platforms created by the telecommunications/entertainment complex (such as instant messaging, 'video to phone' capability, etc.) has created a new host of concerns for parents and cultural commentators because the technology has often excluded older, less technologically savvy citizens from full participation. The following two sections titled 'Popular culture in studies of globalization' and 'Audiences and cultural studies' will discuss how academic geography has traditionally dealt (or not dealt) with the perceived crisis in society over popular culture.

Popular Culture in Studies of Globalization

Globalization is theorized most commonly in geography as the intensification of the social relations in the global production complex, resulting from, among other reasons, the transportation revolution and its ensuing round of time–space compression. Among general populations, this is most perceptible in the circulation of goods and services from all over the globe to their localities. Goods from all over the world are available for purchase, and people from all over the world can be found in what had previously been isolated, seemingly homogenous, places. Alongside increasing cultural heterogeneity comes increasing difference in both goods desired and goods consumed. Popular culture has come to reflect this increasingly global outlook, incorporating elements of disparate cultures into the products sold by the culture industry. Some view this as a normative good, as increased interaction with 'the other' will necessarily increase understanding and decrease conflict and bigotry. Examples often used in this perspective include the Taoist symbolism of yin and yang, the spiritual/physical exercise of yoga, etc. Others, however, view this as the co opting of folk culture for the service of the profit imperative, stripping context, authenticity, and locality out of culture. Examples often used in this perspective include the American version of Chinese food and 'world' music. The common denominator for both of these views of popular culture is that these artifacts come out of supposedly 'unmodern' contexts and are incorporated into a cosmopolitan modernity.

This idea has been critiqued by many as constructing a normative geography that scripts the West as advanced, rational, and progressive while scripting the 'ethnic' parts of the globe as backward, sensual, and timeless. Globalization is deemed an attempt to advance a universal Western culture that is accented by cultural difference as proof of its inclusive nature, but only on the terms of Western consumers. This charge is often associated with the United States, the dominant force behind the structural changes that enhance the effect of globalization, such as the promotion of free trade. When associated with one country in this way, or with a limited set of countries, this accusation is known as cultural imperialism. Cultural imperialism is seen as negative by disempowered groups because it seeks to establish norms of consumption that are favorable to the exporter of culture and unfavorable to local producers of culture. Because cultural consumption is so closely linked to identity, this is seen as threatening the very underpinnings of society.

Thus, the idea of 'popular' culture can be seen not just in opposition to 'high' culture but also to 'folk' culture, which is defined as the culture of a relatively unchanging locality dominated by so called 'traditional' beliefs, building styles, cuisine, modes of production, familial relations, etc. The contrast here is different than that of 'high' and 'popular' culture, which are differentiated normatively. Rather, it is a contrast of temporality, as folk cultures are perceived to be pockets of underdevelopment within the capitalist world system. The contrast of 'folk' and 'popular' cultures is one perspective that has dominated geographic literature, particularly in the subfield of cultural geography. The conflict between local folk culture and the seemingly inexorable advance of Western popular culture around the world as embodied in goods such as Hollywood films, Coca Cola, McDonald's, and the 't shirt' has been a common leitmotif for both academic geographers and journalists such as Thomas Friedman. Among many cultural critics, this would lead to the very destruction of geography itself as social difference becomes flattened over time with American market dominance leading to cultural homogeneity.

Several flaws have been identified in this perspective. Most flow from its initial assumption: market domination leads to cultural homogeneity. The first critique argues that culture and commodities are not entirely linked; you are not necessarily what you buy. The assumption identified in this critique limits the definition of culture to material goods, excluding rituals, dialect, religion, and all else from the equation. The second critique questions the assumption that using identical forms of material culture (i.e., getting coffee at a Starbucks in Seattle vs. in Jakarta) equates to identical cultural practice. This critique differs from the first in that it deals specifically with the meanings associated with particular manifestations of popular culture. To draw the example further for the purposes of explanation, it is hypothetically possible that Starbucks consumers in Seattle derive feelings of cosmopolitanism and elite social status from their purchase, while Starbucks consumers in Jakarta purchase coffee in solidarity with rural coffee workers in Sumatra. Hence, the meanings associated with popular culture are not necessarily being Americanized even if the goods originate in the United States.

Thus, the globalization discourse that has been so significant in academic geography for the past 20 years has led to an impoverished view of popular culture as material goods that circulate around the world. Within cultural and political geography, however, a different but related concern has existed alongside questions of globalization: how does popular culture create or influence ideas of place and citizenship?

Audiences and Cultural Studies

As described above, there is much anxiety voiced in society about the effects of popular culture on audiences. This dates back to World War I and the widespread use of propaganda as a method of waging war. The ability of mass media (such as radio or leaflets) to undermine elites' preferred geopolitical visions was startling to these elites and sparked the first academic studies of audiences. These studies assumed that the public was essentially passive and vulnerable to the power of suggestion. These concerns were amplified by the rise of fascism in Europe, which was perceived to be the result of a combination of charismatic leadership and the growth of the mass media as a multiplier of that charisma. Following World War II, fear of communist subversion motivated the same types of studies that had sought to understand the rise of fascism prior to the war. Of particular importance to these studies was the work of the Frankfurt School, especially Theodor Adorno and Max Horkheimer, who together shaped the idea of the 'culture industry', a massive complex of entertainment industries that claim to respond to the consumers' desires but rather shape those multiple and complex desires into a standardized (and therefore profitable) set of goods. The effect of this culture industry is that consumers become sated and docile, placated by the popular culture that captivates them. This has sometimes been referred to as the 'bread and circuses' effect, dating from the Roman Empire's policies of free wheat and entertainment to prevent uprisings.

The 'bread and circuses' critique of popular culture dovetails nicely with a Marxist conception of social relations, and particularly with the work of Antonio Gramsci, who has been influential with many political geographers because of the explicit connection he made between culture and the state. In particular, Gramsci was concerned with the ways in which capitalism had become so entrenched in the state despite the teleological view of most Marxists that revolution should be forthcoming. He argued that something very similar to Adorno and Horkheimer's culture industry had constructed a hegemonic ideological culture that promoted bourgeois values as superior. Thus, Gramsci's theory of hegemony and that of the Frankfurt School both hold that popular culture is like a pill that is provided by elites to the masses as a vehicle for inculcating preferred values. The only difference between them is that Gramsci allowed for some degree of accommodation; elites have to incorporate some (but not all) bourgeois and working class culture to maintain hegemony. For Gramsci, it was necessary for revolutionaries to first create a hegemonic culture of the working class that would create an obvious fault line between the 'popular culture' of the working class and the bourgeois culture of the middle class and elites. In this formulation, the denigration of 'popular' culture by advocates of 'high' culture, as described at the beginning of this essay, can be seen as an ideological discourse promoted to keep the working class engaged with the status quo.

Over the past few decades the field of study known as 'cultural studies' has taken root as an interdisciplinary project. Including sociologists, scholars of literature, anthropologists, geographers, and others, it has taken as its object of study popular culture and the ways in which audiences impact, or are impacted by, it. Theoretical shifts within cultural studies have largely been away from the structuralism of the Frankfurt School toward poststructuralism. In these post structuralist accounts, the audience is retheorized as an active agent in the constitution of popular culture. The work of de Certeau seeks to recenter the production of popular culture in the audience by arguing that meanings are not fixed through authorial intent, and instead are often read differently by various audiences. This critique of structuralist theorizations of popular culture is similar to the example of Starbucks coffee given in the above discussion of cultural imperialism. The use of popular culture for ends not intended by the creators of culture in the culture industry effectively subverts the entire paradigm of structuralist cultural theory, as audiences can be shown empirically to not react in ways that are consistent with the goals of producers. Classic examples of this re reading can be found in gay and lesbian culture, where cultural artifacts originally intended as tacitly heteronormative such as the television show *Xena: Warrior Princess* and the comic book relationship between Batman and Robin are effectively queered by audiences, often in the face of authorial complaints.

This post structuralist view of popular culture is also critical of the dichotomy between folk culture and popular culture described above, although for different reasons. Whereas traditional cultural geography perceived folk culture as normatively good in that its extinction at the hands of a globalizing Western/American culture led to a loss of cultural diversity, poststructuralist cultural theory, especially the work of de Certeau, argues against this sense of nostalgia for folk culture. Instead, he posits that (post)modern culture is no less 'cultural' than folk culture, and indeed that because interpretation of popular culture is critical to the constitution of its meaning, the expansion of popular culture artifacts and forms from the United States (or the West) do not lead to cultural homogeneity, but instead to a diversifying number of interpretations of that artifact or form as it is reinterpreted by different communities.

This perspective offers new theoretical spaces for geographers to map, as it ties in well with much of the state centric work that has traditionally been done in political geography. The role of national identity in shaping interpretation of popular culture is a fertile ground for research. However, national identity is but one form of identity through which interpretive communities can crystallize, as other identities such as ethnicity, gender, and sexual orientation may be more salient for various groups. Interpretive communities are based on a critique of post structuralist theory in which the infinite potential readings assumed by a purely poststructuralist view is deemed unlikely. Instead, audiences make sense of texts using the cultural resources that they have at their disposal. Through the regular arrangement of linguistic and other cultural references, common meanings are established within interpretive communities; for effective communication to exist within a community a system of norms must exist that effectively narrows the field of possible readings inherent to any text.

Thus, the communal myths of national identity help shape the ways in which new artifacts of popular culture are consumed. In the critical geopolitics literature this phenomenon, when applied to the news or other forms of new knowledge, is known as geopolitical scripting; in literary theory it is often referred to as a narrative frame. Both terms refer to the way in which context helps shape interpretation of events. Geographies of popular culture can easily be seen to connect to geographies of nationalism, such as in the work of Benedict Anderson. Anderson argues that the rise of nationalism over the past several centuries is directly tied to the rise of mass media (i.e., the printing press) in conjunction with the overarching social movements associated with the decline of religious authority and other authorities derived from religion. Thus, 'imagined communities' come into being as the result of a common experience of consuming printed material. This problematizes interpretive communities because Anderson claims that interpretive communities arise from consumption of popular culture, while in general interpretive communities are seen as predating the popular culture that they consume.

This tension between the impact of popular culture on identity and behavior and the role of the audience in providing meaning is the subject of ongoing academic debate. However, in the popular press there remains very little consideration of the interpretive process. This is most visible in recent debates over phenomena such as rap music (deemed misogynistic by many) and video games (likewise deemed ultraviolent).

See also: Art and Cartography; Children/Childhood; Critical Geography; Critical Geopolitics; Critical Theory (After Habermas); Cultural Politics; Cultural Studies and Human Geography; Cultural Turn; Culture; Cyberspace/Cyberculture; Film; Globalization, Cultural; Identity Politics; Imperialism, Cultural; Literature; Material Culture; Material, The; Media; National Spatialities; Orientalism;

Poststructuralism/Poststructuralist Geographies; Soundscapes; Structuralism/Structuralist Geography; Visuality; Youth/Youth Cultures.

Further Reading

Adorno, T. and Horkheimer, M. (2002). *Dialectic of enlightenment*: *Philosophical fragments*. Jephcott, E. (trans.). Palo Alto, CA: Stanford University Press.

Anderson, B. (1991). *Imagined communities: Reflections on the origin and spread of nationalism*. New York: Verso.

Billig, M. (1995). *Banal nationalism*. Thousand Oaks, CA: SAGE Publications.

Brooker, W. (2001). *Batman unmasked: Analyzing a cultural icon*. New York: Continuum.

Crothers, L. (2006). *Globalization and American popular culture*. Lanham, MD: Rowman and Littlefield Publishers.

De Certeau, M. (2002). *The practice of everyday life*. Berkeley, CA: University of California Press.

Dodds, K. (2003). Licensed to stereotype: Popular geopolitics, James Bond and the spectre of Balkanism. *Geopolitics* 8, 125 157.

Gans, H. (1999). *Popular culture and high culture: An analysis and evaluation of taste*. New York, NY: Basic Books.

Gramsci, A. (1971). *Selections from the prison notebooks*. New York: International Publishers.

Kitchin, R. and Kneale, J. (eds.) (2005). *Lost in space: Geographies of science fiction*. London: Continuum.

Norcliffe, G. and Rendace, O. (2003). New geographies of comic book production in North America: The new artisan, distancing, and the periodic social economy. *Economic Geography* 79, 241 264.

Sharp, J. (2000). *Condensing the Cold War: Reader's digest and American identity*. Minneapolis, MN: University of Minnesota Press.

Storey, J. (2003). *Inventing popular culture: From folklore to globalization*. Oxford, UK: Blackwell.

Wertham, F. (1954). *Seduction of the innocent*. Mattituck, NY: Amereon Ltd [reprinted 1996].

Relevant Websites

http://www.americanpopularculture.com
Americana, The Institute for the Study of American Popular Culture.

http://www.horowitzfreedomcenter.org
David Horowitz Freedom Center.

http://www.popularculture.org
Popular Culture Association/American Culture Association.

http://www.popcultures.com
Sarah Zupko's Cultural Studies Center.

http://www.ncge.org
The National Council of Geographic Education.

Population Geography

A. J. Bailey, University of Leeds, Leeds, UK

Introduction

Population geography is the subdiscipline of human geography that has concerned itself with describing, analyzing, and reflecting upon the geographical orga nization and growth of human populations in their environmental and social settings. It thus distinguishes itself from demography which, as the mathematical sci ence of the human population, focuses on the processes of fertility, mortality, and migration and how these create predictable populations. Put another way, population geography emphasizes how populations and population processes appear and change across space, while demo graphy emphasizes change over time. This article explores the disciplinary basis of knowledge, reviews three major approaches to knowledge creation, and closes with a brief resumé of outstanding debates. It argues that while spatial science approaches have lent credibility and vitality to intellectual debates within and beyond the field, and structural approaches have offered important insights about inequalities, a further integration of these with emerging cultural approaches is needed to tackle contemporary knowledge challenges.

Disciplinarity

Like other clusters of knowledge in geography, popu lation geography frequently presents itself, and is criti cized, as an institutionalized subdiscipline of geography. The advantage of such a 'disciplinary' view is to draw attention to the structures, norms, and intellectual bases of such an enterprise, including the extent to which de velopments outside geography (notably in demography, but also economics, sociology, mathematics, and cultural studies) and within geography transform the field. The potential disadvantage with this organizing view is to divorce the activities of the field from their broader social and political context – that is, we must pay heed to how the activities of population geographers always reflect, and sometimes influence, events in the 'real world'. Indeed, some of the outputs of the field are closely allied to political agendas, such as nineteenth century popu lation maps that showed the national distribution of racial and ethnic concentrations and helped legitimize the idea of a natural community which was contained within the bounds of the nation state. Certainly, taking the long view, human societies have consistently had a use for knowledge about their people, how they occupy the surface of the earth, and what such patterns of organization might mean for political and civic organization, military security, commerce, inequality, environmental depletion, resource use, and well being. Yet, while population geography has cross cultural and interdisciplinary roots, it is still domin ated by its Euro American intellectual pedigree, which is not to say this will not nor should not change.

It is the conventional view of the disciplinary devel opment of the field that the comments of the then president of the Association of American Geographers, Glenn Trewartha, at the Cleveland annual conference in 1953, provided a nucleus around which some key concepts began to cluster. Trewartha believed in a uni tary and synthetic geography, and spoke of a holistic vision to highlight the importance of studying areal dif ferentiation and geographic diversity. Trewartha wrote: "Physical and cultural, systematic and regional, general and special, are dualisms which appear at times to fog the oneness. As geographers I believe we are committed to the study of earth regions" (Trewartha, 1953: 71–97). Rejecting dualistic thinking within geography (human–physical, physical–cultural, production–consumption), his approach placed population at the apex of geo graphy's concerns, buttressed by analyses of the cultural earth and the physical earth: "The here suggested tri nomial organisation results in giving man his deservedly explicit and important position within the unitary geo graphic structure. The only *final* value is human life or human living, and this being the case it is difficult to understand why geographers should judge the creation of man, and the environment out of which he creates them, relatively more important than man himself" (Trewartha, 1953: 71–97). Students of this new subdiscipline would study regional variations in population distributions and look at explanations for diversity in terms of dem ography, the potential for livelihood, and settlement systems. Furthermore, historical population geography would serve to place regional variations in their historic context. The key characteristics of population included the gross patterns of numbers, fertility and mortality dynamics, measures of under and over population, density of population and settlement and migration patterns. They would also study the physical characteristics of populations. Informed by mid and late nineteenth century anthropological and socio logical classifications, these included body type, race, nationality, sex, age, and health status. Relevant socioeconomic characteristics included religion, education, occupation, family structure, urban/rural residence, development level, and customs.

This rather detailed vision was to have an enduring legacy for how population geography developed. Although discussions by Preston James and David Hooson also adopted a unitary view of geography that emphasized the importance of looking at both human and environmental explanations, developments elsewhere in geography and sciences led researchers to focus on that part of Trewartha's work that said that population processes can be understood from description and analysis of population patterns. In short, this inspired the field to cross disciplinary boundaries in a quest for ideas, concepts, and theories that could help explain, or at least bring some sense of order to, the population diversity they saw. Crucially, this was an overt foundation of the spatial science approach that was to exert a powerful influence on human geography during the middle part of the twentieth century.

Spatial Science Approaches

In its purest form, a spatial science approach to population concerned itself with three questions. First, research would describe the distribution and composition of the population. Thus, an opening chapter of the Peters and Larkin text (one of the most widely circulated texts in the field which reached its eighth edition in 2005) includes discussion of: patterns of global distribution of population broken into the ecumene (livelihood supporting) and nonecumene; patterns of population density as represented by arithmetic and nutritional density measures; national and international variations in sex structure; local, national, and international variations in age structure, including the baby boomers and elderly in the US; the distribution of race and ethnicity in the US including a discussion of measures of segregation. Second, research would uncover processes of change, both inductively and deductively, paying close attention to spatial variations and spatial contexts in their accounts. According to demographic analysis, populations changed as a direct result of the changing balance between fertility, mortality, and migration; by appreciating the spatial patterns and trends in these determinants, analysts could comment upon aggregate patterns, distributions, and compositions. Third, having commented upon causes of population change, research would examine the consequences of population structures, distributions, and dynamics. Here, spatial science approaches exhibit a diversity of focus, but links between population and resources, food, ecological change, pollution, economic development, socioeconomic standing and poverty, urbanization, and planning (national and local government but particularly private sector and commercial) continue to be explored.

Such spatial science approaches dominated population geography scholarship during the second part of the twentieth century to the extent they became synonymous with the field. By 1966, Wilbur Zelinsky enthused that population geography defined itself as the systematic study of: (1) the simple description of the location of population numbers and characteristics, (2) the explanation of the spatial configuration of these numbers and characteristics, and (3) the geographic analysis of population phenomena (the interrelations among areal differences in population with those in all or certain other elements within the geographic study area). The spatial science approach thus "deals with the ways in which the geographic character of places is formed by, and in turn reacts upon, a set of population phenomena that vary within it through both space and time as they follow their own behavioral laws, interacting one with another and with numerous nondemographic phenomena" (Zelinsky, 1966: 5). The somewhat hierarchically organized methods of inquiry – pattern, process, consequence – both clarified the field's position with respect to an internal debate within geography about focus, and fitted with a wider view in society that social science should break the big social issues of the day into their smaller parts and build specialist understanding from these components.

While population geographers drew on the techniques of geography to describe and classify population diversity, they increasingly looked to other disciplines to understand and explain population patterns. For example, from demography, the insights of stable population theory helped researchers build credible population projections and form a picture of how many people of which ages they would expect to find in a national population. Differentiating closed systems (societies where population change over time is only caused by differences in fertility and mortality) from open systems (which also include the impacts of population migration), a simplified 'model' population structure, with predictable numbers of persons at different ages, can be derived by assuming that fertility and mortality inputs remained relatively constant over the long horizon. Widespread use is still made of the fruits of this framework for calculating life expectancy and applying life tables and actuarial methods, synthetic measures of fertility including the total fertility rate, regional demographic accounting models, and in population projections, as applied to future populations and 'backward' to historic reconstructions. Population geographers and demographers began what proved to be a long and productive relationship.

The search for order and regularity in a world of complexity and messiness increasingly turned to a series of influential categories of study and analysis. Indeed, urban sociology had by now established a rich tradition of describing and explaining patterns of neighbourhood change and succession in urban areas by applying ideas and metaphors from botany and ecology to their study of population. In turn, other relevant characteristics to study were inspired by the categorizations of populations by skin color and racial phenotype. What were once regarded as common sense categories (white, black, male, female, young, old, etc.) quickly became an established – and largely unquestioned – part of the field's toolbox. In a similar manner, the way in which researchers took environmental and social context into account became increasingly formalized. Bailey argues, for example, that while broader scholarship on populations and their geographies have long seen context through multiple lenses of place, environment, and space, at this time the field focused exclusively on spatial ideas. Thus, space could be a container of areal differentiation; a way to build synthesis, such as used in the regionalizations that inspired the demographic transition theory; a surface where concentration and proximity foregrounded and prioritized some explanations over others; a barrier to interaction and diffusion that slowed down some processes and sped up others.

What did this approach add? Many spatial science contributions took – and continue to take shape as descriptive accounts concerned with patterns of distribution. Frère de Montizan's population distribution of France in 1830 and Harness's population density map of Ireland in 1837 are examples of how population distribution maps helped political states project territorial integrity. Ideas from social physics helped the field describe the density and hierarchical organization of populations. This includes concentration (including the gini coefficient), distribution (e.g., population centroid, or center of gravity), hierarchy (including the rank size rule), and interaction (including the family of gravity models, and intervening opportunities framework). Depictions of a population's notional 'center of gravity' were originally used as snapshots of complex underlying processes but now have their own cultural notoriety as places of pilgrimage and web discussion. John Snow's map of cholera cases around a water pump in London's Soho district went further by suggesting a cause (dirty pump) for an effect (cholera outbreak) and foreshadowed spatial epidemiological investigations of disease clusters.

While descriptive accounts of spatial variations in populations and their characteristics continue to provide a naturalized starting point for much work (as indeed Hettner had foreseen in 1927), it was the field's ability to contribute to and lead theoretical development that attracted funding, new students, and practitioners, and built confidence. Major contributions followed from the development of process understandings and theory. As Zelinsky (1966) noted, "If a regularity or pattern is suspected in certain geographic aspects of population, and if this suspicion survives trial as a working hypothesis to blossom into a full blown theory that seems to fit all known causes, then such a formulation could be extremely valuable." Continuing work on the demographic transition illustrates this well. The belief that the onset and diffusion of modernization triggers a sequenced mortality and fertility decline remains at the heart of the widely circulated classic demographic transition model. The field's sensitivity to space and spatial analysis has critiqued and influenced the development of the classic demographic transition model in at least three ways. First, the technique of regionalizing areas (countries) with similar mortality and fertility dynamics lay behind Landry's 1934 *La Révolution Demographique*, which proceeded to link regions with low population growth rates to the characteristics of public health interventions and fertility reductions attributable to the widespread practice of birth limitation. Regionalization continues to be used to infer causality and to refute hypotheses in both contemporary and historical settings of population change. Population geographers have also explored the extent to which social processes underlying transition could be modeled in a diffusion framework. An appreciation of the rapid and intense spread of public health norms and, more recently, family planning prerogatives (often undertaken as part of structural adjustment programs) to countries with large rural and impoverished populations has enabled the universal nature of the model to be questioned. Similarly, a greater appreciation of how particular forms of fertility limitation became legitimized and used in particular communities at specific historical moments has provided another critique of applying the experience of a few Western European countries in the nineteenth and early twentieth century to the contemporary majority world. This has also shifted attention away from the individual as an autonomous and rational actor and onto broader social structures including community and, for John Caldwell, tropes of Westernization and culture more generally. More generally, diverse work continues on the fertility transition, the mortality (epidemiological) transition, the gap between the two (sometimes discussed as a form of demographic convergence, the mobility transition, and the plurality of household types as part of a second demographic transition. The influence of cross-disciplinarity is in evidence, with neoclassical (marginal) economics permeating not only individual level rational choice models of fertility determinants and mortality differentials but also macro analyses linking population growth and development and economic development and demographic transition.

Such concerns with actively theorizing how both demographic and nondemographic elements were part of a broader system of processes driving demographic transition led some spatial analysts to adopt a systems (dynamics) approach. The work by another demographer, Kingsley Davis, proved pivotal. He noted:

> the process of demographic change and response is not only continuous but also reflexive and behavioural reflexive in the sense that a change in one component is eventually altered by the change it has induced in other components; behavioural in the sense that the process involves human decisions in the pursuit of goals with varying means and conditions. (Davis, 1963: 345)

While analysis had long recognized the importance of urbanization and the urban milieu for changing the relative benefits and costs associated with raising children and for affecting the status of women in the household, Davis's multiphasic (or, change and response) framework inspired greater attention to the diversity of family for mations and population age structures as reflexive elements that changed behavioral strategies. Van de Kaa's formulation of a 'second demographic transition', atten tion to the relationship between demographic aging and transition theory, and elaborations of the roles of migration and ethnicity take this tradition forward.

Systems approaches borrowed strength from the popularity of the framework within human geography in the 1970s, and sparked several applications to the study of human migration. One landmark application was the Brown and Moore model of residential mobility. This behavioral model separated out the decision to move from the decision about where to move, and attributed the first to factors including characteristics of the immediate environment (encroachment, blight, change in ethnic composition, change in relative location) and characteristics of the household (including job change, family size, marital status, retirement). Once a decision to move, in principle, had been made, the model focused on how individuals within households acquired information about alternative opportunities, assuming that the quan tity and quality of information decayed with increasing distance. A lack of appropriate or affordable options either triggered a series of '*in situ*' changes to an existing premises, a change in aspirations, or a new search; like wise, decisions to move affected the composition of neighborhoods at the origin and new destination and fed back into the decisions other households would be making about mobility. Another influential systems application, also published in 1970, was Mabongunje's depiction of rural to urban migration. This recognized positive feedback effects between urban areas and potential migrants and unpacked the environments of the urbanite (including wages, consumer preferences), the urban control subsystem (including communi cations), the rural control subsystem (including government policies and agricultural practices), and the potential migrant (including education, social develop ment). As Mabogunje (1970: 241–260) wrote "one of the major attractions of this approach is that it enables a consideration of rural urban migration as no longer a linear…cause effect movement, but as a circular, inter dependent, progressively complex, and self modifying system." Approaches to migration that specify such fac tors of cumulative causation have an important legacy in sociology and economics.

Of the three demographic elements to drive popu lation change, migration has received the field's greatest attention. Like mortality, morbidity, and fertility, the study of migration has proceeded with the development of a large number of measures and categories. Many of these can be traced to the seminal work published by Ravenstein at the close of the nineteenth century. Such measures of migration help classify diverse flows on the basis of their direction (e.g., in and out migration, and return migration), the scale over which they occur (e.g., within urban areas/intra urban, regional, and international), and on the balance of flows between an origin and a destination (e.g., net migration and gross migration). Ravenstein's so called 'laws of migration' combined the empirical observations that most moves (in to the rural to urban population redistribution that Britain had experienced with the intensification of industrialization in the late nineteenth century) occurred over short distances and up the urban hierarchy with social physics views about forces of attraction and re pulsion to stimulate further empirical research, promote a push–pull view of migration, and spark the develop ment of the family of gravity (spatial interaction) models.

A wide array of data and models help build forecasts on migration trends. This is important because in many countries with low fertility and mortality rates (including many in Europe and Oceania) migration is the key driver of population change, particularly at the regional and local levels. The involvement of such countries in global systems of international migration means that demo graphic destinies are increasingly interdependent issues. Much empirical work summarizing new international migration regimes appeared in origin countries of the former Soviet orbit, Asia, Africa, and Latin America, and the transformation of some countries of emigration into countries of rapid immigration. Elsewhere, burgeoning work on diasporas and transnational communities suggests that many migrants expect (and are expected) to live outside their home areas for indefinite periods of time and adapt particular patterns of settlement, integration, and economic activity accordingly. At a finer scale, research also describes trends in urbanization, rural depopulation, regional shifts in population concentrations (e.g., Sunbelt growth in the US), sub urbanization, counterurbanization and nonmetropolitan growth, neighborhood growth and decline within cities, and gentrification.

As with the study of mortality and morbidity, migration scholarship has become increasingly specialized. For example, broader links between migration and health and migration and HIV AIDS have recently formed new poles of activity. Such specialization has tended (although not exclusively) to follow the lines of inquiry prompted by early typologies of migration. For example, Peters and Larkin identify research on: primitive migration (groups overwhelmed by unfavorable natural forces); mass migration (groups larger than families moving *en masse*); individual voluntary migration (those willing and able to move); restricted migration (individual migrants somehow constrained or guided by institutional context); and forced migration (those impelled by social and political institutions to relocate). As with any typology, while these ideal types may help to clarify what is an extremely complex picture, they can also obscure some processes and realities. This has led Skeldon to reinvoke the migration systems approach to understanding dynamics across the contemporary Asia and Pacific region and examine settler migration, contract labor migration, skilled migration, student migration, and asylum seeker migration.

The study of migration differentials explores how some individuals are more likely to migrate than others. Migration selectivity has been described at particular ages (for example, around the time young adults leave childhood homes), when workers change jobs, when couples marry and divorce, and so on. Such insights spur the development of particular theories. For example, the human capital theory argues that individuals choose migration as an investment that can lead to better opportunities in the future, and uses as supporting evidence the observation that individuals in their twenties often have the highest propensity to migrate at a time when they have the longest time horizon over which to recoup the initial investment cost. Gendered migration theory notes that, in different settings, men and women have different rates of moving overseas. In many countries, including Mexico, cultural norms are associated with the numerical dominance of young men in migration flows to the US in the 1950s and 1960s. In other cases, including the Philippines, women have dominated out migration flows, leading Jim Tyner and other researchers to focus on how government development policies might be linked to such sex selectivity.

Migration theory building exploited cross disciplinary insights. For social physics and human capital approaches, the search for process became a search for those socioeconomic and environmental variables that could capture the costs and benefits of migration to rational individuals. But while human capital theory seemed to be supported by what was known about the age selectivity of migration, a growing feminist critique that the category 'sex' should not only be visible in these analyses but also be reconsidered through sensitivity to gender and gender relations has changed migration theory. Scholarship on family migration illustrates this development well. Family migration scholarship is rooted in the aggregation of individual human capital theory to the household context, with Jacob Mincer arguing that partners moved wherever their joint expected future earnings could be maximized, even if this meant one partner lost out or 'was sacrificed' for the common good. Because women, historically, had had less access to higher education, and some gender roles restricted their access to paid employment, many researchers found that wives sacrificed more than husbands from moving, at least in the short term. However, evidence accumulated that place and historical context matters, with factors including the direction of migration (particularly if it was to a new area or a returning move), the set of resources available in the destination, the nature of the gender roles within the household, and the ways in which women and men negotiate 'the trauma of migration' all identified as important.

Growing interest in the diversity and pluralization of household types in many societies of the global North and South led population geographers to connect together formerly separate research on family migration, employment and educational strategies, housing choices, travel and commuting behavior, and health experiences. Much of this work in turn sees households as important formations where complex social processes collide, are negotiated, and help restructure the social and cultural norms of society. For example, lifecourse analysts are interested in the connections between 'linked lives' of household and extended family members, in the sequence and timing of key events in these lives, in the effects of previous experiences upon current and future activities (path dependency), and in the interactions between the biographies of individual lives, the transformation of institutional rule sets and norms, and the slower passage of generations, formations, and epochs. Calling up Hagerstrand's earlier time–space geographies, and making use of biographical and longitudinal data and new techniques of analysis including microsimulation, studies have explored the links between single social processes and a demographic outcome (e.g., how does marriage affect migration?), the deeper interdependencies between individual characteristics and lifecourse changes (e.g., what is the likelihood of being trapped in poverty because of early life experiences?), and the wider implications of economic changes for social systems (e.g., how do changing energy costs affect patterns of household organization and feed into housing markets and economic growth?). The impulse to better understand the changing dynamics of the contemporary household is partly linked to high profile debates about social policies on flexible working, work–life balance, child care, medical insurance, and social welfare in general which have changed the context within which household formations operate. In turn, this has profound implications for populations, including social diversity, patterns of low fertility and childlessness, and postponement, adult–child proximity, and migration decisions. Yet, as Buzar *et al.*, (2005: 413–436) note "the conceptual importance of household demography in understanding contemporary patterns of urban transformation in the developed world…has often been marginalized in the geographical mainstream…we argue that a deeper understanding of these dynamics can improve the quality of social, cultural, and economic theorization within geography."

While research on migration and household demography increasingly stresses the interdependence of demographic and nondemographic factors and integrates cause and consequence in ways consistent with Zelinsky's earlier reading, spatial analyses that link population to environmental and resource issues have been more formulaic. For example, the well known 'IPAT' equation, which proposes that environmental Impacts like pollution are a function of population size and growth, population Affluence and consumption activities, and the Technologies used to support this system of production, is an outshoot of such thinking. The 'resource region' proposed by Ackerman in 1967 is one attempt to add a richer understanding of spatial variations in context to this debate through regionalizing the population–resource–technology relationship and through highlighting spatial interdependencies like trade. While the systems inspired 'limits to growth' model further explored feedback effects, it generally underplayed the role of policy and policy changes in influencing causal relationships. Similarly, approaches to the development consequences of population have been criticized for not adopting more integrated analyses.

Where the field's research has gained traction is in the assessment and critique of public policies that aimed to, or did impact upon population patterns or processes. Mapping spatial distributions of variables of interest proves a quick and effective way of demonstrating how one size fits all public policy may have very different local and regional repercussions. Examples relate to sex ratios that show the prevalence of female infanticide and demonstrate that cultural norms have an important translation effect upon fertility limitation policy, abortion ratios that show how religion can mediate policy translation, and low birth weights that show how isolation and segregation continue to drive significant ethnic variations in health outcomes. Analyses of diffusion of diseases including influenza, measles, HIV AIDS, and foot and mouth have informed the development of, and critique of, public health policies. Political transitions in the former Soviet Union have produced morbidity and mortality conditions that undermine the epidemiological transition framework. Rural to urban migration in China has long been affected by the *hukou* (registered permanent residence) system that undermined the ability of urban migrants to be legally resident and which now contributes to planned patterns of 'return' migration to rural areas.

As an example of more applied scholarship, spatial analysts have lobbied to improve the quantity, quality, and reliability of the data used in demographic analysis (particularly those data that can be georeferenced). Data and census systems are seen as a basis for the democratic functioning of institutions within society, including those concerned with taxation, representation, and drawing attention to human suffering. The eleventh century Domesday Book collated information on the size, location, and wealth of the English population. Snapshot type counts (censuses), retrospective (event history) surveys, and panel (continuous) data form the basis of analysis, and this has spawned and been promoted by developments in measurement, recording, storage, processing, and analytical methodologies (longitudinal data analyses, maximum likelihood estimators, agent based modeling, microsimulation, and GIS). Such data and models provide important applied inputs for local planning processes and are used by geodemographic and other commercial institutions for marketing, product placement, and locational analysis.

Using newly available population data, showing intellectual innovation and delivering policy advice that informed housing, transportation, and health decision making, spatial science approaches helped lend respectability to the field and human geography more generally. However, one of the concerns associated with a data intensive approach is the reliance on information that may have a short half life and may suffer from reliability (and thus inference) problems. Indeed, White and Jackson's now infamous critique of population geography calls attention to an over reliance on quantitative data, with a tendency to accept its categories at face value. Several traditional census categories – including race, ethnicity, sex, and age – attempt to capture deeper concepts (including identity, gender, generation, and so on) that are socially constructed, cross cutting, and varying between places and over time spans. Taken alongside the questioning of the behavioral and positivist approaches in human geography and the rise of social and political movements in the 1960s, 1970s, and early 1980s that questioned Cold War, interventionist, and establishment ideologies, concerns such as those articulated by White and Jackson arose with a flourishing of what Allan Findlay calls a more pluralistic population geography. In particular, this gave more elbow room to other approaches – some structural and some cultural – that were already present in more out of the way parts of the field.

Structural Approaches

Structural accounts emphasize how population matters are an intrinsic part of the economic and social inequalities that characterize contemporary patterns of capitalist accumulation. More widely, within human geography, structural accounts have been particularly interested in how inequality and unevenness arise, appear within, and are passed on through the spaces and scales of capitalism. While the geographical materialism of David Harvey traces its lineage to Marxian views of the labor theory of value, a diverse panoply of interdisciplinary contributions on dependency relations and world systems theory, uneven regional development, differential patterns of urbanization, global systems, and the articulation of network society touch population geography. Other neo Marxist critiques have exposed the ideological assumptions behind neo Malthusian readings of population crises and resource issues. Radical feminist scholarship drew attention to the gender relations inherent in patriarchal society and how these gender systems intersect with class systems to circulate inequality.

Population appeared in these accounts in different ways. Population phenomena like the overurbanization of many cities in the majority world may be read as 'symptoms' of deeper contradictions within capitalism: for example, the systematic overproduction of workers by both natural increase and migration that has the effect of driving wages down. Increasingly, however, structural accounts conceptualize population activities as 'social processes' that reproduce structural inequality in interdependent ways. For example, flows of skilled migrants benefit metropolitan economies in the global North by assembling a class of innovators and future leaders of the economy, but have negative impacts on sending countries of the global South which cannot realize the potential of the innovative class (despite having been responsible for training) and fall further behind developed economies. Examples are widely drawn from nursing and health care but also extend to engineers, educators, and aid workers.

Emphasizing how population processes have been an important component of the growth of the world economy and globalization offered the field an opportunity to focus attention on systematic differences in material conditions that were widely documented between, for example, the majority world of the global South and the metropolitan economies of the global North, between rural and urban areas in the majority world, and across classes, ethnic and 'racial' groups. Some structural accounts thus gained prominence as critiques of individual centered neoclassical economic renditions of migration, while others built on feminist theory to situate accounts of uneven spatial divisions of labor that perpetuated 'glass ceilings' and entrapment in the suburbs in readings of the social division of labor. As contemporary research acknowledged that individuals engaged in activities that added and circulated capital directly (e.g., by working) and indirectly (through a range of social reproduction activities, such as raising and socializing future members of society), a parallel with an earlier period of intellectual development in the former Soviet Union demonstrates the nonlinearity in knowledge cycles. That is, early population geographers drew on Marx's historical materialism to see population distribution as reflective of the territorial basis of economic production. But internal migration, urbanization, and the sectoral and spatial labor shortages of the Soviet 1950s compelled Pokshishevskiy to remark that "any cluster of population creates a node of consumption, therefore the geography of population discloses places of considerable concentration of separate branches of economy which belong to the consumer's side of social production process."

Structural approaches in population geography can begin to be differentiated along a continuum that puts population as a symptom of inequality at one end and population as part of a matrix of social processes leading to inequality at the other end. The four sets of contributions summarized below represent an indicative rather than exhaustive account of such scholarship. First, a significant corpus of research shows how practices of population mobility stem from inequalities in the distribution of opportunities, and help perpetuate structural divides. Theorists had long been aware of the negative impacts of brain drain type migration. For example, work on colonially inspired patterns of labor mobility within Africa tied the manipulation of labor flows to "the penetration of the domestic modes of production by the external money economies" (Swindell, 1979: 248) through measures like agrarian reform, legal restrictions, and forced labour regimes. Today, the economic and ideological resources of globalized labor markets which enable the free flow of elite migrants suggests that this form of migration has become an entrenched part of globalization. Similarly, accounts suggest that skilled, unskilled, and forced migration has been produced internationally to supply cheap labor to metropolitan hubs, with dependency relations further extended through reliance upon such transfers as economic remittances.

Second, and drawing on economic and sociological labor market theories, analysis focuses upon the links between migration and global cities. Part of a broader agenda to understand the contours of economic globalization, migration is theorized as an accessory to the bifurcated and polarized labor markets that have been observed by Friedmann and Sassen in the economic control and coordination nodes like London, Paris, and Tokyo. The rise of producer services, the demand for cheap, flexible, and expendable labor, rising levels of female labor force participation, and unwillingness of many native born workers to fill menial jobs creates a structural demand for immigration. Much of this demand is sector specific and, combined with the emigration policies of sending countries like the Philippines, creates the impetus for the arrival of female target migrants in activities including sweatshops, domestic work, health care services, retail, and prostitution. The gender and class dimensions of these flows in most metropolitan areas of Australia, the US, and Europe turns on economic and ideological principles that associate such work with women of particular class and nativity backgrounds, and leads to segmentation and entrenchment. In a similar vein, and in deference to neoclassical and household economic theory, Bauder argues that rather than labor markets driving migration, it is migration that structures the trajectories of particular labor markets.

Third, research explores how decisions about how to organize and care for households deepens social inequalities. Many middle income households in the global North hire live in nannies and domestic workers to balance their own work–life priorities. Permissive government regulations (often amounting to turning a blind eye to this unregulated sector), nonavailability of native workers, and the segmentation of foreign born workers in need of jobs has led to the staffing of such work by young immigrant women. Moreover, such work obliges many of these immigrant women to maintain split households of their own and endure not only separation from their own children, but added pressure to remit and support extended family in distant origin regions. Such transnational families arise as complex global care chains connect working families of the global North and South together in new, but unequal ways. Further examples concern the organization of elder care, of increasing significance in aging societies where fertility rates fell dramatically over the past century. Again, working families (often with their own children still at home) are obliged to move closer to one or more ailing parent and not only forego economic opportunity but also take on additional care demands. These fall disproportionately on the working class and, because of gender ideologies, women, further deepening social divides in the absence of social support from the state. The suggestion made by Gibson Graham (1996: 261) that "nonmarket transactions (both within and outside the household) account for a substantial portion of transactions, and that therefore what we have blithely called the capitalist economy…is certainly not wholly or even predominantly a market economy, perhaps we can look within and behind the market to see the differences concealed there" places tasks of social reproduction – caring activities – as key social processes that intersect and circulate inequality.

Fourth, political economy and political ecology work has highlighted how state policies on population perpetuates inequality. For example, research on the context of family planning policies in China and Indonesia reveals how patriarchal and imperial relations inform the model and techniques of policy delivery, variously resulting in female infanticide and the nonsanctioning of abortion for unwed women, and perpetuating some women's low status in their respective communities. Reproductive rights frameworks have been discussed as ways to address the issues identified in this research.

The continuation of patterns of inequality – and in some cases the widening of poverty gaps and demographic divides – lends credibility and legitimacy to structurally based accounts of population. However, as scholarship has integrated economic, political, social, and cultural process, views of particularity – of how context matters, of contingency, and of space and place – have become increasingly fragile. One response has been to unpack the ways in which culture (often, but erroneously, associated with agency and institutions acting from the ground up), and political economy (structure, again erroneously conflated with top down) influence each other. The merits of the structurationist approach of Anthony Giddens are often debated within the field but the approach has proved difficult to operationalize. Another response, from within human geography, considers how culturally inspired forms of difference arise and circulate, and this forms the core of a range of cultural studies research.

Cultural Studies Approaches

A key focus of an emerging cultural studies tradition concerns the diverse ways population circulates power and difference within society. While rarely recognized (or intended) as a discrete and fixed terrain of inquiry, those inspired by this framework tend to acknowledge and integrate an appreciation of geographic diversity that stems from spatial analytical research and to work toward a narrative that takes account of the systematic inequities and ideological content articulated by structural accounts. At the same time, research aspires to avoid functional and essentialist treatments of categories that are acknowledged to be socially constructed and to probe the meanings, and potential for transformation, of categories. Appreciating the contingencies of place has emerged as a key focus, with accounts stressing richly textured and nuanced treatments of space and most of all its relational nature.

Such approaches do have a faint but discernable 'culturalist' legacy within the field itself, as well as being more directly inspired by developments in human geography and beyond. Within human geography, and by the close of the 1980s, two influential texts (Harvey's *Condition of Postmodernity* and Soja's *Postmodern Geographies*) were signalling an intention to fold in the 'trialectics of space, time, and social being' to accounts of modernity. Despite possessing arguably limited spatial imaginaries, these interventions, along with works like Jackson's *Maps of Meaning*, did seed what Elspeth Graham has recently termed "a flowering of spatial imaginations within geography," although one might add that parts of the geographic garden were in bloom sometime before the others. Population geography was one such late bloomer, although connections with earlier intellectual sympathies are apparent. Indeed, for Bruhnes it was only through 'the cultural landscapes of place' that population issues were worthy of study: "It is in the connection with the house, the village, and the city that the questions of the distribution of the population must be examined – under its real and logical aspect." Treating 'place' as materially and hermeneutically constituted is part of the spirit, if not the letter, of those scholars who worked within *geographie humaine* and who felt that particular 'ways of life' (*genres de vie*) were worthy, and indeed in tegrative objects of attention. In turn, Paul Vidal de la Blache and Jacqueline Beaujeu Garnier were among those to attach central (but not singular) import to matters of culture and population in accounting for in equality in society. Perhaps paradoxically, the influence of French theorists continues through the attentions given to authors such as Foucault, Lefebvre, Bourdieu, and Deleuze.

One way to begin to appreciate the scope of this current article is to loosely distinguish those analyses concerned with governmentality and disciplinary power, those concerned with population and the production of space, and those interested in performance. As links be tween classification and power became acknowledged, views on how populations were made and maintained, and for what purposes, have greatly expanded. Foucault's ideas about 'biopower' as a political technology of the state centers on the inscription of the human body as a machine (to do specific types of work in optimal ways) and on the disciplining of population groups (like refu gees and undocumented immigrants) to accomplish a broad range of tasks of production and reproduction. The use by states of certain forms of discourses – that is, the construction and circulation of forms of knowledge that separate out some groups from others on the basis of seemingly objective and naturalized categories (including race, ethnicity, nationality) – has sparked research on constructions of outsiders and others and the use of medical facilities by refugees.

While the not inconsiderable literature in population geography on transnationalism exhibits diversity of approach and focus, a good deal is concerned with how identity and institutions transform each other under globalization. For example, research on remittance flows and transnational connections between Mexico and the US points to 'transnational social fields' (most generally, communities stretched over national borders) that: both extend the reach of the origin state (Mexican nationals exert increasing influence in many US cities) and the destination state (increased dependence of many com munities in Mexico on the availability of US jobs); transcend the reach and legitimacy of these states (by changing the support for traditional opposition parties and new political movements, and by vesting more influence with family networks and commercial service providers, including coyotes and parasitic economy establishments); transform the social institutions that facilitate daily life (split families, new faith based com munities); and precipitate new cultural constructions of what is experienced as proximity (network space), affection (often mediated through the economy of gift giving), and anxiety (heightened and exploited by immigration authorities to discipline potential workers) with the promulgation of new meanings of space and time. These transformations may become codified in language, food, popular culture, and other 'soft' cultural formations. Research continues to explore how contin gency affects the production and circulation of difference in these new and networked spaces and times.

A third pole of activity combines lifecourse insights on how biographies unfold across linked lives with as pects of nonrepresentational theory that explore how repeated activities define social and cultural norms. Both approaches come together around ideas of relationality, arguing that all social processes are fluid and in various states of becoming, and critiquing ideas of stable cat egories. The profound implications for such insights on age based rites of passage and the definitions of gener ation and childhood, gender, and disability have been discussed. Moreover, relational thinking opens up new questions about mobility, migrancy, and sedentarism, and how a range of institutions (including government im migration laws, and educational curricula) help regulate cultural norms that in turn define deviance, difference, and power. For Rosi Braidotti (2006), "the differences in degrees, types, kinds and modes of mobility and – even more significantly – of non mobility need to be mapped out with precision and sensitivity…Grounded, histor icized accounts for the multiply positioned subjects of postmodernity are needed for people who are situated in one of the many poly located centres that weave together the global economy." More work considers population events and processes like mothering, dating, dyng, and racialization.

While it remains too soon to establish clear research agendas that will produce sustained trails of knowledge, it is already apparent that much interdisciplinary and some intrafield boundary crossing is in evidence (studies of transnationalism, and childhood, for example, are increasingly informed by hybrids of the above collations). Yet, despite trading on conceptions of difference, diver sity, and inequality, the gap between cultural approaches and structural/spatial analytical readings remains rather wide.

Ongoing Debates

A confluence of internal and external forces has given the field's practitioners continued opportunities to reflect on collective directions forward. These include the launching of a specialized journal (*International Journal of Population Geography, Now Population, Space and Place*), the advent of the fiftieth anniversary of 'founding father' Trewartha's rallying cry for more attention to be devoted to population geography (see the special issue of *Popu lation Space and Place* in 2003), the institution of a biannual international population geography conference series, and public intellectual concerns around questions of security, vulnerability, ecological catastrophe, food security, economic collapse, and so on. One reading of the current state of the subdiscipline is a field charac terized by diverse approaches with some practitioners opting for continuity of approach, some turning to more critical frameworks, many mixing and matching in a celebration of plurality. Other (and not oppositional) readings are that the field needs to pay more attention to pressing societal concerns such as famine and vul nerability and sustainable development. Communicating results more effectively to diverse audiences and par ticipating more equally with diverse populations are part of this. For example, relatively few population geog raphers use the power of visualization to communicate their findings, with US geographers Jim Allen and Eugene Turner and British geographers Danny Dorling and Bethan Thomas being recent exceptions.

Like all subdisciplines, the field takes stock of its border crossings. Rather less traffic has passed through the economic border of late and more with social and urban geography, while links with health geography, population studies, and demography remain strong. Given the waxing and waning of the field's preoccupation with the production and consumption sides of popu lation, and the current ascendancy in human geography of integrated cultural economy approaches, future border crossings may be influenced by how the field interfaces with cultural theory. Early signs of how it will interface with the widely observed cultural turn that has influ enced the direction of travel within the closed in sub disciplines have yielded relatively disparate knowledge projects within population geography. Reconnecting debates within population and economic geographies through cultural lenses can provide capacity to tackle the big issues of the day. However, any such endeavor has to acknowledge and address the continuing paucity of scholarship from the majority world.

See also: Census Geography; Demography; Health Geography; Migration.

Further Reading

Ackerman, E. A. (1967). Population, natural resources, and technology. *Annals of the American Academy of Political and Social Science* 369, 84 97.

Allen, J. and Turner, E. (1988). *We the People: An Atlas of America's Ethnic Diversity*. New York: Macmillan.

Bailey, A. J. (2005). *Making Population Geography*. London: Hodder.

Beaujeu Garnier, J. (1956 1958). *Géographie de la Population*, 2 vols. Paris: Presses Universitaires de France.

Bonney, N. and Love, J. (1991). Gender and migration: Geographical mobility and the wife's sacrifice. *Sociological Review* 39, 335 348.

Boyle, P., Halfacree, K. and Robinson, V. (1998). *Exploring Contemporary Migration*. Harlow: Longman.

Braidotti, R. (2006). *Transpositions. On Nomadic Ethics*. Cambridge: Polity Press.

Brown, L. and Moore, E. (1970). The intra urban migration process: A perspective. *Geografiska Annaler* 52B, 1 13.

Buzar, S., Ogden, P. E. and Hall, R. (2005). Households matter: The quiet demography of urban transformation. *Progress in Human Geography* 29(4), 413 436.

Clarke, J. I. (1965). *Population Geography*. Oxford: Pergamon.

Clark, W. A. V. (1998). *The California Cauldron*. New York: Guilford.

Davies Withers, S. and Clark, W. A. V. (2006). Housing costs and the geography of family migration outcomes. *Population Space and Place* 12, 273 289.

Dorling, D. and Thomas, B. (2004). *People and Places*. Bristol: Policy Press.

Findlay, A. M. (2005). Editorial: Vulnerable spatialities. *Population Space and Place* 11, 429 439.

Gibson Graham, J. K. (1996). *The End of Capitalism (As We Knew It): A Feminist Critique of Political Economy*. Oxford: Blackwell.

Jackson, P., Crang, P. and Dwyer, C. (2003). *Transnational Spaces*. London: Routledge.

Jones, H. R. (1981). *A Population Geography*. London: Harper & Row.

Kalipeni, E., Craddock, S., Oppong, J. R. and Ghosh, J. (eds.) (2004). *HIV and AIDS in Africa: Beyond Epidemiology*. Malden, MA: Blackwell.

Lawson, V. (2000). Arguments within geographies of movement: The theoretical potential of migrant's stories. *Progress in Human Geography* 24, 173 189.

Mabogunje, A. L. (1970). Systems approach to a theory of rural urban migration. *Geographical Analysis* 2, 241 260.

Mattingly, D. (2001). The home and the world: Domestic service and international networks of caring labor. *Annals of the Association of American Geographers* 91(2), 370 386.

Philo, C. (2001). Accumulation populations. *International Journal of Population Geography* 7, 473 490.

Plane, D. A. and Rogerson, P. A. (1994). *The Geographical Analysis of Population. With Applications to Planning and Business*. New York: Wiley.

Pratt, G. (1999). From registered nurse to registered nanny: Discursive geographies of Filipina domestic workers in Vancouver, B. *C. Economic Geography* 75, 215 236.

Samers, M. (1997). The production of diaspora: Algerian emigration from colonialism to neo colonialism (1840 1970). *Antipode* 29, 32 53.

Sassen, S. (1991). *The Global City*. Princeton, NJ: Princeton University Press.

Silvey, R. (2004). Power, difference, and mobility: Feminist advances in migration studies. *Progress in Human Geography* 28, 490 506.
Skeldon, R. (1992). International migration flows within and from the East and South East Asia region: A review essay. *Asian and Pacific Migration Journal* 1, 19 63.
Skeldon, R. (1995). The challenge facing migration research: A case for greater awareness. *Progress in Human Geography* 19, 91 96.
Skeldon, R. (1997). *Migration and Development*. Harlow: Longman.
Smith, M. P. (2001). *Transnational Urbanism*. Oxford: Blackwell.
Trewartha, G. (1953). The case for population geography. *Annals of the Association of American Geographers* 43, 71 97.
Tyner, J. (1994). The social construction of gendered migration from the Philippines. *Asia and Pacific Migration Journal* 3, 589 612.
Van de Kaa, D. J. (1987). Europe's second demographic transition. *Population Bulletin* 42, 1 57.
Walton Roberts, M. (2004). Transnational migration theory in population geography: Gendered practises in networks linking Canada and India. *Population, Space and Place* 10(5), 361 373.
White, P. and Jackson, P. (1995). (Re)theorising population geography. *International Journal of Population Geography* 1, 111 123.
Wong, T., Yeoh, B. S. A. and Graham, E. (2003). Spaces of silence: Single parenthood and the 'normal family' in Singapore. *Population Space and Place* 10, 45 58.
Woods, R. I. (1982). *Theoretical Population Geography*. London: Longman.
Zelinsky, W. (1966). *A Prologue to Population Geography*. Englewood Cliffs, NJ: Prentice Hall.
Zelinsky, W. (1971). The hypothesis of the mobility transition. *Geographical Review* 61, 219 249.

Relevant Websites

www.igu net.org/uk/what is igu/commissions.html
International Geographical Union Commission on Population and Vulnerability.
www.prb.org
Population Reference Bureau.
www3.interscience.wiley.com/journal/106562735/home?CRETRY=@SRETRY=0
Population Space and Place (leading journal).
www.unfpa.org
UN Population Fund.

Port-Industrial Complexes

M. Dunford, University of Sussex, Falmer, Brighton, UK
G. Yeung, National University of Singapore, Singapore

Glossary

Containerization A multimodal system of freight and cargo transport involving the use of International Standards Organization (ISO) containers that can be packed, sealed, and loaded onto container ships, airplanes, railway rolling stock, and trucks using standardized equipment.

Entrepôt The word for a warehouse, in French, used to refer to a trading center or to an individual warehouse into which merchandise can be imported and from which it can be exported without the payment of duties as the trading center is simply an intermediate point in the export–import chain.

Growth Poles Originally conceived by Perroux as a dominant industry or firm (existing in abstract economic space, understood as a field of economic forces) that has propulsive economic effects due to the existence of a combination of high degree of innovation (in the sense of Schumpeter) and relatively strong input–output linkages and external economies. Whereas Perroux's concepts were designed for the study of economic growth (centered on the existence of a historical succession of poles which generate strong centrifugal and centripetal forces and strong multiplier effects), Boudeville and a number of regional planners identified growth poles and growth centers as geographical clusters or geographical concentrations of industries that exert strong multiplier effects (due to shared infrastructures, proximity of customers and suppliers, more diversified labor markets, and more rapid diffusion of innovations). These ideas opened the way to the active development of growth poles and growth centers as places in which propulsive industries should be located with a view to promoting and subsequently diffusing development.

Hub-and-Spoke Network A hub is a central transport node through which the movements of people and goods are routed, while spokes are the routes to the destinations that ships, airplanes, and vehicles serve from that center. Each destination is connected directly to the hub. A transport operator may establish a number of hubs.

Logistic Chain The full set of activities involved in planning, implementing, monitoring, and managing the flow and storage of goods, services, and related information from their places of origin to the place at which the goods, services, and information so produced are used or consumed.

Maritime Industrial Development Area (MIDA) The concept of combining large-scale industries (steel, oil, and chemicals, in particular) with new port facilities capable of handling large and specialized bulk carriers and well-endowed with pipeline and other modern transport infrastructures emerged in the 1960s as a result of a coming together of the desire to develop vertically integrated coastal steel, oil, and chemical complexes, a related restructuring of maritime transport and seaport systems, and state strategies of national industrial development and the creation of growth poles in economically less developed areas.

Seaport A gateway with facilities for receiving and (intermodal) transferring passengers and cargo between water and land transport. A port connects a hinterland with a foreland. Its evolution depends on the nature and speed of economic development of the areas it serves and its capacity to adapt in the light of the evolution of maritime (including in the recent past the widespread development of bulk terminals, container ports, and roll-on/roll-off methods of cargo handling) and other transport technologies and logistic systems.

Twenty-Foot Equivalent Unit (TEU) A measurement standard used to indicate the capacities of container ships and terminals. One 20-foot ISO container (20 feet in length, 8 feet wide, and 9 feet high) equals 1 TEU. One 40-foot ISO container equals 2 TEUs.

Introduction

A port is essentially a gateway with facilities for receiving and transferring passengers and cargo between water and land transport. As such it is also a place where related industrial activities are found giving rise to port industrial complexes. To identify the nature and development of these complexes, attention must first be paid to the nature of ports and the economic activities associated with them.

Analyzing Ports and Their Industries

A port is an intermediate location in the global flow of passengers and freight (on which we shall concentrate). These freight flows can be construed as steps in commodity, value, or supply chains and as elements of the logistic systems put in place to manage them. A com modity chain is a functionally integrated set of production, consumption, distribution, and exchange activities extending from the acquisition of raw materials through to the sale of finished goods and services. At each stage in a chain, there is a pressure to drive down costs and add value. The organization and management of these sequential processes involve the development of complex logistic systems which embrace all of the activities involved in the physical movement and hand ling of raw materials, goods, and services and the related information flows: ordering, scheduling, loading, move ment, transshipment, warehousing and storage, inventory management, and so on. As in the case of commodity or value chains as a whole, these logistic systems (whose costs amount to some 10–15% of world 'Gross Domestic Product' (GDP)) are subject to strong pressures to secure time economies, eliminate inventories (in a context of just in time supply chain management), integrate activities (to permit door to door freight distribution), and ensure reliability, among other things. As shipments by sea account for about 80% of international trade by volume, seaports are vital nodes in global supply/logistic chains, as well as in supranational and regional trade.

In the framework of these ever changing economic and logistic systems, each port serves and/or competes for a share of the traffic generated in its own hinterland and in the market area made up of ports and port hin terlands throughout the world with which it is connected. To survive in market societies, ports must engage in a competitive struggle to expand their market shares. The outcome depends on the capacity/quality of the ports and their infrastructures, of the road, rail, and water communications infrastructures linking port with their hinterlands (land access), of their maritime interface (length of berths and quays), and of their conditions of maritime access (tidal range, and channel and berth depth, protection from wind, waves, and storm surges). These capacities/qualities are themselves the result of massive investments, and involve frequent redevelopment and relocation of the port itself, as is indicated by suc cessive phases in the development of Europe's major port of Rotterdam (**Figure 1**). As a result of this expansion, Rotterdam's port (which concentrates on trade in oil and chemicals, containers, iron ore, coal, food, and metals) and its industrial complex cover 10 500 ha and stretch 40 km from the city to the Maasvlakte along the Nieuwe Waterweg canal.

The development of individual ports also depends on a range of technological factors, geopolitical develop ments, trends in regional/world trade and passenger movements (reflecting the degree of global economic integration), and shifts in the relative economic strength of the areas ports serve: shifts in the center of gravity of world economic development (measured by GDP *per capita* of market areas and hinterlands) have a profound influence on the evolution of ports. The development of the port of Yantian (**Figure 2**) in southern China reflects the outsourcing, from the 1980s onward, of manufacturing, processing, and assembly operations to China. Opened in 1994, a second phase of development was completed in 2000, a third in 2006, and a further expansion project is due for completion in 2010. By the end of the first three phases, Yantian had been trans formed from a small fishing village to the fourth largest container port in the world by throughput. At the root of its vertiginous expansion lay the way in which the growth of processing activities in Mainland China, the increasing development of Mainland Chinese port infrastructures, and rising handling costs in Hong Kong's container ports (it costs about US$400 more to ship one twenty foot equivalent unit (TEU), the size of a standard 20 foot container, from Hong Kong to the US or Europe than from Yantian) are seeing more and more major transna tional corporations (TNCs) ship their goods directly from their subcontractors in southern China to Yantian for loading into container vessels and export to major European and North American destinations. To cater for the expected growth in demand, a logistics park 3 km from the port of Yantian is under construction.

In general what has resulted is the development of regional port systems and associated industries com prising a major hub port, several feeder ports and their hinterlands, and a strong process of concentration of traffic on a small number of major gateways (due to the existence of high fixed capital costs, scale economies, and

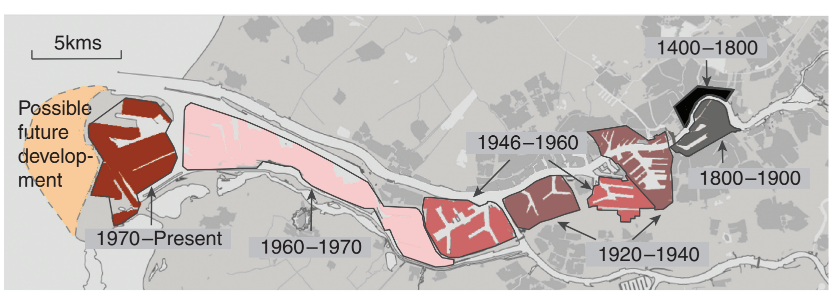

Figure 1 The phases of development of the port of Rotterdam.

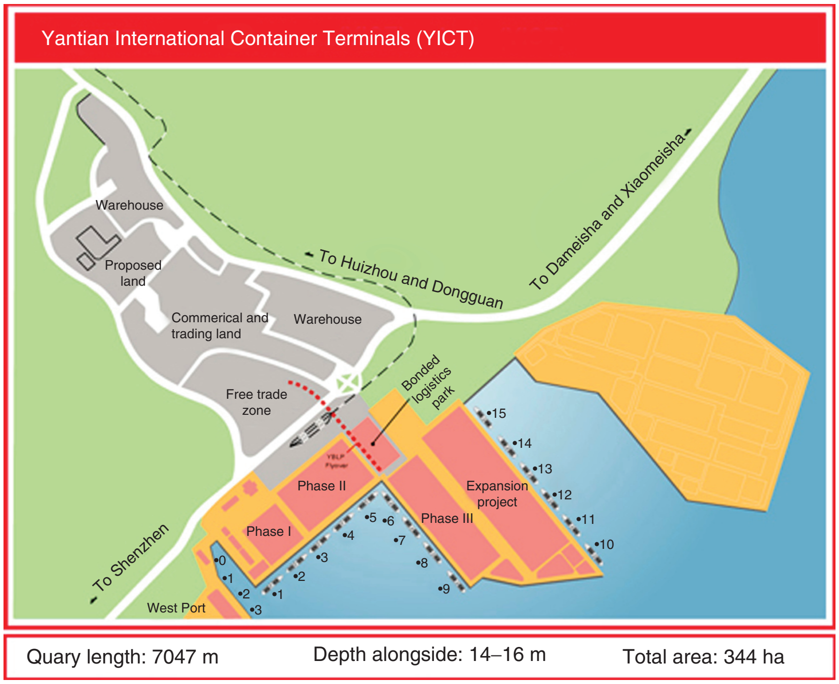

Figure 2 The phases of development of the port of Yantian.

high entry costs in port and related industrial sectors). **Table 1** records the 2004 cargo and container traffic of the world's most important ports, while **Table 2** reports the strong increases in container traffic of the largest gateway ports. Among other things, these tables also indicate the remarkable share of Chinese and East Asian ports, which is largely due to the growth of manu facturing industries in these areas. Alongside these developments, there have been important changes in the structure and governance of port, transport, and logistic systems with, in particular, movements in the direction of de regulation and a withdrawal of governments and public authorities from the ownership and management of ports.

All ports are also focal points for economic activities and for associated industrial development. This com bination of activities is sometimes formally recognized in the designation of 'maritime industrial development areas' (MIDAs). The economic activities associated with ports fall into several groups. The first includes cargo and passenger handling and storage and distribution activities directly related to the port function, ship repair, and a host of transport related services located in the port itself and in its city centers. A second group com prises a set of processing industries that transform im ported materials before their onward shipment/re export taking advantage of the intermodal, transshipment, and break of bulk functions of ports. A third group of industries located in port industrial complexes are those whose inputs comprise bulk commodities imported through the port. Examples include oil refineries and related chemical industries, iron and steel mills, and sugar refineries.

Alongside the direct attraction of port related indus tries, two more general mechanisms also shape the development of port industrial complexes. The first is the existence of external economies: the availability of shared infrastructures generates external economies that can attract other economic activities and encourage processes of agglomeration, although ports and their associated industrial areas also generate a series of negative externalities in the shape of noise, pollution, and visual blight that can also deter further economic diversification and development. The second relates to the way in which the concentration of port dependent economic activities can itself attract customers and suppliers. Indeed the existence of backward and forward linkages was often seen in growth pole theories as creating the possibility of cumulative processes of industrial development. In practice, the degree of downstream diversification actually achieved was quite limited, making employment in port industrial com plexes particularly dependent on the fortunes of a few industries.

Monopolvilles: The Development of Maritime Industrial Development Areas

The development of port industrial complexes was a striking feature of economically advanced and rapidly industrializing economies in the 1960s and 1970s.

Table 1 World port ranking, 2004[a]

Total cargo volume, metric tons (000s)					*Container traffic (TEUs, 000s)*			
Rank	*Port*	*Country*	*Measure*	*Tons*	*Rank*	*Port*	*Country*	*TEUs*
1	Singapore	Singapore	FT	393 418	1	Hong Kong	China	21 984
2	Shanghai	China	MT	378 962	2	Singapore	Singapore	21 329
3	Rotterdam	Netherlands	MT	352 563	3	Shanghai	China	14 557
4	Ningbo	China	MT	225 850	4	Shenzhen	China	13 615
5	Hong Kong	China	MT	220 879	5	Busan	South Korea	11 430
6	Busan	South Korea	RT	219 760	6	Kaohsiung	Taiwan	9714
7	Guangzhou	China	MT	215 190	7	Rotterdam	Netherlands	8281
8	Tianjin	China	MT	206 161	8	Los Angeles	United States	7321
9	South Louisiana, LA	United States	MT	203 517	9	Hamburg	Germany	7003
10	Houston, TX	United States	MT	183 419	10	Dubai	United Arab Emirates	6429
11	Nagoya	Japan	FT	182 289	11	Antwerp	Belgium	6064
12	Chiba	Japan	FT	169 254	12	Long Beach	United States	5780
13	Kwangyang	South Korea	RT	165 875	13	Port Kalang	Malaysia	5244
14	Qingdao	China	MT	161 650	14	Qingdao	China	5140
15	Ulsan	South Korea	RT	156 517	15	New York/New Jersey	United States	4478
16	Kaohsiung	Taiwan	MT	152 468	16	Tanjung Pelepas	Malaysia	4020
17	Antwerp	Belgium	MT	152 327	17	Ningbo	China	4006
18	Qinhuangdao	China	MT	150 320	18	Tianjin	China	3814
19	Dalian	China	MT	145 162	19	Tanjug Priok	Indonesia	3597
20	New York/New Jersey	United States	MT	138 328	20	Laem Chabang	Thailand	3529
21	Shenzhen	China	MT	135 246	21	Bremen/Bremerhafen	Germany	3469
22	Yokohama	Japan	FT	126 960	22	Tokyo	Japan	3358
23	Hamburg	Germany	MT	114 484	23	Guangzhou	China	3308
24	Incheon	South Korea	RT	113 073	24	Gioia Tauro	Italy	3261
25	Port Hedland	Australia	MT	108 500	25	Algeciras	Spain	2937

[a]The cargo rankings based on tonnage should be interpreted with caution since these measures are not directly comparable and cannot be converted to a single, standardized unit.

MT metric ton; HT harbor ton; FT freight ton; RT revenue ton.

From American Association of Port Authorities (2006) http://www.aapa ports.org.

Table 2 Container throughput in TEU (twenty-foot equivalent units)

	Container throughput in TEU					*Rank*	*Rank*
	2005	*2004*	*2003*	*2002*	*2001*	*2005*	*2004*
Singapore	23 192 000	21 329 000	18 410 500	16 940 900	15 571 100	1	2
Hong Kong	22 427 000	21 984 000	20 449 000	19 144 000	17 826 000	2	1
Shanghai	18 084 000	14 557 200	11 281 000	8 620 000	6 334 400	3	5
Shenzhen	16 197 173	13 655 484	10 649 900	7 613 754	5 076 000	4	8
Pusan	11 840 445	11 491 968	10 407 809	9 453 356	8 072 814	5	3
Kaohsiung	9 470 000	9 714 115	8 843 365	8 493 052	7 540 525	6	4
Rotterdam	92 86 757	8 280 787	7 143 920	6 506 311	6 096 142	7	6
Hamburg	8 087 545	7 003 479	6 137 926	5 373 999	4 688 669	8	9
Dubai	7 619 222	6 428 883	5 151 955	4 194 264	3 501 820	9	13
Los Angeles	7 484 624	7 321 433	7 178 940	6 105 857	5 183 520	10	7
Long Beach	6 709 818	5 779 852	4 658 124	4 526 365	4 462 967	11	10
Antwerp	6 488 029	6 063 746	5 445 437	4 777 151	4 218 176	12	11
Qingdao	6 307 000	5 139 700	4 238 000	3 410 000	2 638 500	13	16
Port Kelang	5 543 527	5 243 593	4 841 235	4 533 212	3 759 512	14	12
Ningbo	5 208 000	4 005 500	2 772 200	1 860 000	1 213 100	15	20
Tianjin	4 801 000	3 814 000	3 015 000	2 408 000	2 059 420	16	18
New York	4 792 922	4 478 480	4 067 811	3 749 014	3 316 275	17	14
Tanjung Pelapas	4 169 177	4 020 421	3 487 320	2 660 000	2 049 000	18	19
Laem Chabang	3 765 967	3 624 000	3 181 050	2 656 651	2 312 439	19	17
Tokyo	3 594 000	3 580 000	3 313 647	3 028 000	2 968 284	20	15

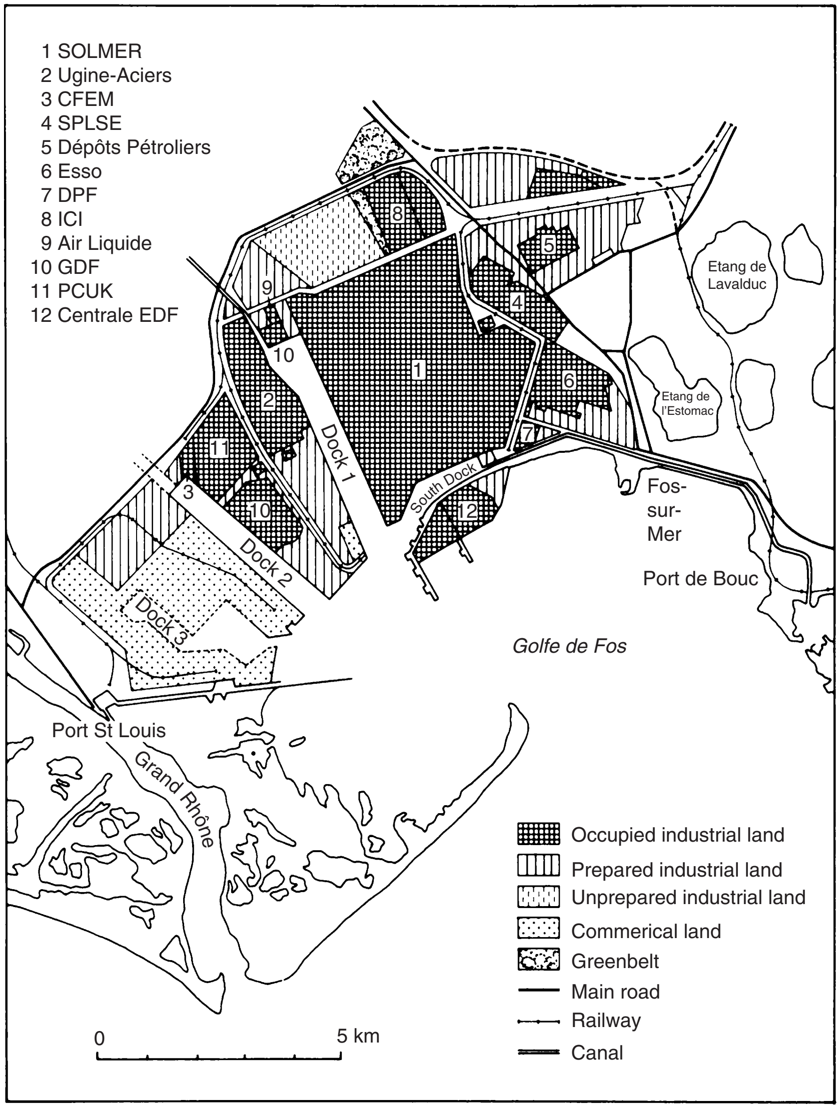

Figure 3 The structure of the Fos-sur-Mer MIDA in the late 1980s. Adapted from Dunford, M. (1988). *Capital, the State and Regional Development*. London: Pion.

European examples included Teeside in the UK, Dunkirk and Fos sur Mer in France (**Figure 3**), Taranto, Brindisi, and Gioia Tauro in Italy, Huelva and Cadiz in Spain, and the never completed Pylos port and industrial project in the Peloponnese in Greece. In the Americas, the development of a coastal chemical complex in Puerto Rico was the subject of a classic study in regional science by Isard, Schooler, and Vietorisz.

At that time, projects for the development of MIDAs were a result of a coming together of developments in industries producing intermediate goods such as steel and chemicals and a restructuring of maritime transport and seaport systems on the one hand and state strategies of support for national champions (in different industrial sectors) and (often) for the development of economically disadvantaged areas on the other.

In a context of increasing demand, intermediate goods industries experienced strong processes of concentration and centralization of capital. The emergence of a smaller number of larger groups was designed to permit the devalorization of outdated plants and the creation of a smaller number of large integrated complexes capable of reducing handling and transport costs, speeding up the turnover of capital, and realizing scale economies. As these new complexes drew on imported raw materials (for reasons related to the availability or the comparative cheapness and quality of imported materials), coastal sites with undeveloped flat land and natural deep water channels were sought out. The creation of these plants encouraged, in turn, the growth of upstream industries providing energy and gas and downstream metal and chemical using industries, although the degree of diversification and of long term employment creation were often limited, with some zones remaining largely mono industrial.

In the maritime transport and port sectors, a related quest for scale economies and a competitive struggle to reduce the time and costs of circulation and to increase the speed of turnover of capital saw a series of related developments: a growing use of extremely large and specialized vessels whose time at sea had to be maximized to secure a better rate of return on the large amounts of capital locked up in them, and the development of ports capable of receiving them and of ensuring rapid access and fast turn around times, so that the amount of time they spent in the port was minimized. To meet these needs, ports required deep access channels and berths, specialized handling equipment, and large amounts of space. The associated reorganization of port operations, in turn, provided the occasion for an often conflictual recomposition of the dock workforce.

These projects also involved significant state intervention. MIDA projects were, in many cases, major components of corporatist national or sectoral economic development plans. At the same time, many MIDA projects were promoted as growth pole projects for depressed or less developed areas. As a result, these projects were often co planned by regional planning agencies, and were financially supported by regional development funds: state funds financed infrastructural investments receiving less than the normal market rate of return and/or provided grants and low interest loans to companies located in these zones.

MIDA projects diminished in importance after the mid 1970s' economic crisis. The stagnation of demand for intermediate goods along with the creation of new capacity for refining oil and for producing primary chemicals and steel in less developed and newly industrialized countries saw the rationalization of the industries that underpinned MIDA development in the advanced world and a movement of investment into other areas of economic life. As a result, port development projects were refocused, at the same time as increased globalization saw large increases in the volume of international trade and in international freight (and passenger) movements.

Supply-Chain Management and Port-Industrial Complexes

In the last five decades, two other waves of significant innovation in sea freight transport have had profound implications for the supply chain management and for the development of port industrial complexes.

Containerization was the first wave of innovation. The idea of a container was conceived in 1956, when an American trucking magnate (Malcom McLean) calculated that the cost of loading his 'cargo contained metal box' onto a ship was less than US$0.16, compared with the US$5.83 for loading cargo piece by piece. The widespread adoption of containers in cargo transport led to the containerization of significant parts of the logistic network. The deployment of supporting infrastructure, such as gantry cranes, warehouses, and rail, motorway, and canal/riverine connections promoted new industrial sectors dedicated to port development. In addition, the introduction of dedicated container vessels in the 1960s, in turn, demanded deep water container ports. In the pursuit of economies of scale, the capacity of container vessels has increased significantly over time. In the late 1960s, the largest container vessel had a capacity of 1000 TEUs. By the early 2000s, vessels with the capacity of 7000–9000 TEUs were the norm. It is estimated that the per container operating costs of a 6000 TEU vessel are some 20% less than those of a 4000 TEU vessel. These developments facilitate the deployment of information and communications technologies (ICT) to improve the efficiency of logistic management, maximizing the throughput of ports and minimizing the time that containers are sitting idle in vessels or on quaysides.

The second wave of innovation involved the application of ICT in the operation of ports. As well as places for transshipment and storage, established global ports are cities with ICT based logistics management hubs, the city hub, with Singapore being a typical example. As a result of ICT advances, customs declarations and transactions are conducted electronically, speeding up transactions and shortening container storage time. It is estimated that more than 30 000 container trucks travel each day between Kwai Chung container terminal in Hong Kong and various locations in China (as in contrast to most global hubs, there is no direct rail link between Kwai Chung and major Chinese cities). The temporary storage and trucking cost savings can be substantial.

Gioia Tauro: From Planning Disaster to Container Gateway

In March 1970, Gioa Tauro (**Figure 4**) in Southern Italy (an area of rich citrus groves) was selected as a future MIDA and more specifically as the site for state owned Italsider's fifth integrated shore based steel plant. By the late 1970s, 7500 jobs were to be created in an area of high un and underemployment and very limited industrialization. State regional development funds were devoted to port construction and site preparation which started in 1975–76. Although the construction of a port capable of receiving giant ore and coal carriers was completed, after several revised plans the steel project was quietly

Figure 4 Gioia Tauro port.

dropped. The reason lay in the steel overproduction crises of the 1970s and their devastating financial con sequences for Italsider. A subsequent electrical power station project designed to make use of Gioa Tauro's notorious white elephant was dropped for environmental reasons. In the mid 1990s, however, this large and still modern port finally found a use as a container port and as a pole for the transshipment of containers from ocean going to Mediterranean vessels (a classic hub and spoke network), not least due to its situation along the route connecting the Suez Canal to Gibraltar. Starting from 16 034 TEUs and docking facilities for 50 ships in 1995, its operations expanded at breathtaking speed to more than 3 million TEUs in 2004 and 3060 ships, making it the largest container transshipment terminal in the Mediterranean.

Industrial Agglomeration and Port Hinterlands

After containerization, the size of the hinterland over which agglomeration effects were exercised increased, in part, as successful ports themselves occupied more space, and, in part, due to reductions in transport costs, creating extended and discontinuous port industrial complexes. Agglomeration economies derived from the clustering of manufacturing sectors, which were, in turn, attracted by accessibility to overseas markets. John Dunning and the UNCTAD both recently argued that spatial clusters of (foreign financed) firms in comple mentary sectors profited from agglomeration/positive external economies, including improved access to shared infrastructures, factor inputs (especially skilled labor), established distribution networks, and regional markets. One of the most successful examples of port development (from a transshipment port to the city hub of Pacific Asia) with agglomeration economies in its hinterland is Hong Kong.

Because of its strategic location and its ice free 12.5 m deep harbor, Hong Kong has long been the leading gateway in East Asia, especially to southern China. Hong Kong developed from an entrepôt in the early to mid 1900s to become the largest container port by throughput in the world by 1987. In addition to being a significant node in regional distribution networks in East Asia (in particular for cargo originating from and destined to southern China), Hong Kong was made into a vital logistic center for TNCs in Pacific Asia.

Hong Kong was a relative latecomer to container ization compared with two of its major, Asian com petitors, Japan and Singapore (which started to handle containers in the late 1960s). In contrast to Singapore, where the Port of Singapore Authority (PSA), which directed port operations and development, is an au tonomous government agency, incorporated as a com mercial port operator in 1997 directed the development and operation of port, the Hong Kong government only provided the land and certain infrastructures (mainly motorway) for the development of a dedicated container port in Kwai Chung, allowing private operators to run it, of which the most important are Modern Terminals Ltd (MTL) and Hong Kong International Terminals (HIT).

After China implemented economic reforms and opened up to foreign investors in 1979, there was a massive relocation of manufacturing industry from Hong

Kong and other Southeast Asian countries into four special economic zones and 14 coastal open cities in southern and eastern China. As the only developed container port that shared land physically with Mainland China, and the transshipment port for China–Taiwan trade (there were no direct links between China and Taiwan until 1997), Hong Kong enjoyed annual double digit growth in throughput and developed as a regional hub for the (foreign financed and) export oriented manufacturing firms in the Pearl River Delta. In the mid 1990s, the government estimated that 'port related' activities (broadly defined) accounted for about 21% of employment in Hong Kong.

Hong Kong, along with Singapore, Busan in South Korea, and Kaohsiung in Taiwan accordingly came to dominate container handling in East Asia. More recently, Hong Kong's dominance as a regional hub has been challenged by the growth of dedicated container terminals in Mainland China. One of the mainline operators, Global Alliance, started to call at Yantian in Shenzhen directly in 1995. With a draft at 14 m at the quayside and a direct connection to the Beijing–Kowloon railway, Yantian in the Pearl River Delta is well equipped to handle container vessels with capacities of more than 7000 TEUs. As mentioned earlier, it is thought that it costs US$300 less to ship a TEU from Yantian to West Europe or North America than via Hong Kong. As mainline operators are increasingly calling at ports in Mainland China to handle China's booming exports, especially Yantian and Shanghai, it is expected that the annual, double digit growth rates registered in Chinese ports since the late 1990s will lead them to overtake ports in other East Asian countries in the near future. In 2005, the throughput of container terminals in Yantian reached 16.1 million TEUs (up from 5.1 million in 2001), making it the fourth largest in the world, just after Singapore, Hong Kong, and Shanghai.

The extraordinary growth of investment and output of China's export oriented manufacturing sectors contributed to the double digit growth in ports' throughput. The investment in port development in China has exceeded that of all other countries in the world combined since 1985. By 2010, it is estimated that the combined throughput of Chinese ports will double to about 140 million TEUs. One of the instances of significant investment in port capacity is Shanghai. To overcome the maximum draft restriction of 11 m in the existing container terminals and to facilitate economic growth in the Yangtze Delta, the Shanghai city government is investing US$17 billion in a 20 year project to build one of the largest deep water container ports (with a maximum draft restriction of 15 m and a total of 50 berths with a capacity of 25 million TEUs) in the Yangshan islands. Interestingly, both of the Hong Kong hubs' major operators, HIT and MTL, have been investing heavily in port development in Mainland China. An example is Hutchison Whampoa, which owns Hutchison Port Holdings (HPH). HPH operates 18 container terminals worldwide and accounts for 10% of global container throughput. In addition, HPH holds either significant or dominant shares in the major ports in Hong Kong, Yantian, Zhuhai, Nanhai, Shantou, and Jiangmen in southern China, and Shanghai and Xiamen in eastern China. Equally interesting is the fact that Chinese TNCs and European port operators are investing heavily in European ports to facilitate the rapid growth of cargo throughput originating from China. For instance, COSCO Pacific holds 25% of the equity in Antwerp Gateway, the operator of a Deurganckdock container terminal in Antwerp. Meanwhile, the Antwerp Port Authority has trained more than 2000 Chinese workers to work in the shipping industry and set up a Chinese language school for the children of Chinese expatriates. A Chinese website dedicated to companies in China is also planned.

Some of the most aggressive investors in port development are located in the Middle East. For instance, United Arab Emirates owned port operator, DP World, bought Peninsular & Oriental Steam Navigation (P&O) for US$6.8 billion in February 2006. DP World was subsequently forced to agree to sell the British company's interests in six US ports after US lawmakers threatened to block the takeover on the grounds that it threatened US national security. In 2005, DP World ports handled 17 million TEUs, 19% more than in 2004. Combined with that of P&O, its throughput reached some 37 million TEUs. By comparison, Singapore owned PSA International, the world's second largest port operator, handled 41.2 million TEUs in 2005.

Competition and Port Development

The development of leading port and industrial complexes should not divert attention away from the other side of the coin: the rationalization and run down of smaller and less favored ports with all of their consequences for port and industrial decline, job losses, dereliction, and attempted regeneration. A striking example relates to the recent development of Franco British ports not least for the differences it reveals in the scope for public action.

In 1994, on the French side of what was until the arrival of the Channel Tunnel the main Paris–London rail route, the Dieppe Chamber of Commerce and Industry put a large amount of investment into a new cross channel passenger and freight port capable of handling new generations of larger and faster ferries and larger cargo ships. The success of this investment depended, however, on complementary investments in

Newhaven on the other side of the channel in comparable facilities. On the cargo side, for example, James Fisher's modern cold store facilities in Newhaven was constantly losing contracts due to insurance problems associated with the restricted depth of the Newhaven port. On the United Kingdom side, however, the former British Rail ports (Folkestone, Harwich Parkeston Quay, Newhaven, Heysham, Fishguard, Holyhead, and Stranraer) and fleet had been privatized in 1984. A Conservative government sold all of British Rail's ports and ferries to Bermuda based Sea Containers for £66 million. Six years later, most of the ferry operations acquired from British Rail were sold to Stena Line for £259 million. (Stena Line had been a party to a hostile takeover bid for Sea Containers.) Sea Containers was left with the port facilities, its Hoverspeed operation, and a massive privatization windfall, as well as its containers activities and its property development, hotel, and publishing interests.

As the owner of the Newhaven port, Sea Containers did secure planning permission for port modernization and the construction of a new outer harbor (complicated by the inclusion of a controversial housing development in the submitted plans). To support these plans, East Sussex County Council secured approval for £6.8 million of capital challenge credit and £1 million of European Union INTERREG funding for a new road connecting the planned outport with the national trunk road network. The situation was, however, that the local authorities would not construct the road (or the modified road agreed to when Sea Containers suggested a revised interim plan of deepening the existing harbor) until Sea Containers made a firm commitment to port development, while Stena Line (soon to be involved in a long drawn out merger with P&O) would not make a long term commitment to the route, and Sea Containers itself showed no inclination to proceed with the port development.

In the late 1980s and early 1990s, there was relatively rapid investment and growth in North Sea and cross Channel ferry operations after the negative impact of the Herald of Free Enterprise disaster of 1987, as the ferry operators prepared to confront the opening (finally in 1994) of the Channel Tunnel. One reason was the positive impact on traffic and passenger movements of the internal European market. Another was that the ferry operators anticipated a cost advantage due to the partial amortization of their fleets and the high debts of the Channel Tunnel operators.

Early in 1999, however, the new P&O Stena Line joint venture withdrew from the loss making (in part, due to its diversion of traffic onto its Dover–Calais route) Newhaven–Dieppe route. For several years, Hoverspeed ran a summer passenger service. Traffic on the route collapsed with negative consequences for the port of Dieppe itself and threats of relocation by local transport operators in the Dieppe area. Two years later, in 2001, the Conseil Général de Seine Maritime (a French departmental council) responded by purchasing the port of Newhaven (via a Société d'Economie Mixte). The aim was to reduce their dependence on private companies, and to reinstate passenger and freight ferry services between Newhaven and Dieppe. Additional money was spent to improve Newhaven's dilapidated facilities; a cooperation agreement was made with United Kingdom local authorities and, together with three French Chambers of Commerce, the Conseil Général established Transmanche Ferries and restarted the cross Channel service. In spite of many difficulties on the Newhaven side, it carries some 450 000 passengers per year, and, in 2006, planned to introduce two new ferries specifically designed to operate on the route. In 2007, however, the ferry operation was transferred to a private operator (Louis Dreyfus Armateurs) after the intervention of the European Commission. As this example indicates, a complex set of interests and economic and political mechanisms come into play to shape port development and port rationalization and decline.

See also: Agglomeration; Cassa per il Mezzogiorno; East Asian Miracle; Export Processing Zones; Global Commodity Chains; Growth Poles, Growth Centers; Trade, International.

Further Reading

Airriess, C. A. (2001). Regional production, information communication technology, and the developmental state: The rise of Singapore as a global container hub. *Geoforum* 32(2), 235 254.

Beynon, H., Hudson, R. and Sadler, D. (1994). *A Place Called Teesside: A Locality in a Global Economy*. Edinburgh: Edinburgh University Press.

Bleitrach, D. and Chenu, A. (1975). L'aménagement: Régulation ou approfondissemnt des contradictions sociales? Un exemple: fos sur Mer et l'aire métropolitaine marseillaise'. *Environment and Planning A* 7, 367 391.

Castells, M. and Godard, F. (1974). *Monopolville. Analyse des rapports entre l'entreprise, l'Etat et l'urbain*. The Hague and Paris: Mouton.

Cullinane, K., Wang, T. F. and Cullinane, S. (2004). Container terminal development in Mainland China and its impact on the competitiveness of the port of Hong Kong. *Transport Reviews* 24(1), 33 56.

Dunford, M. (1988). *Capital, the State and Regional Development*. London: Pion.

Dunford, M. and Greco, L. (2006). *After the Three Italies*. Oxford: Blackwells.

Dunning, J. (1998). Location and the multinational enterprise: A neglected factor? *Journal of International Business Studies* 19, 45 66.

European Sea Ports Organization Organisation des Ports Maritimes Européens (2005). Factual report on the European port sector. *Report of the European Sea Ports Organization Organisation des Ports Maritimes Européens*. Brussels: European Sea Ports Organisation (http://www.espo.be).

Hadjimichalis, C. and Vaiou Hadjimichalis, D. (1980). Penetration of multinational capital into backward regions: A policy analysis in Greece. *Antipode* 12(2), 17 24.

Isard, W., Schooler, E. W. and Vietorisz, T. (1959). *Industrial Complex Analysis and Regional Development: A Case Study of Refinery Petrochemical Synthetic Fiber Complexes and Puerto Rico*. Cambridge, MA: The MIT Press.

Loo, B. P. Y. and Hook, B. (2002). Interplay of international, national and local factors in shaping container port development: A case study of Hong Kong. *Transport Reviews* 22(2), 219 245.

Robinson, R. (1976). Modelling the port as an operational system. *Economic Geography* 52, 71 86.

UNCTAD (1998). World investment report 1998: Trends and determinants. *Report of the UNCTAD.* New York: United Nations.

Wang, J. J. (1998). A container load center with a developing hinterland: A case study of Hong Kong. *Journal of Transport Geography* 6(3), 187 201.

Relevant Websites

http://www.pdc.gov.hk
Hong Kong Port Development Council.

http://www.mpa.gov.sg
Maritime and Port Authority of Singapore.

http://www.marseille port.fr
Port of Marseille Authority.

http://www.portofrotterdam.com
Port of Rotterdam Authority.

http://www.port technology.com
Port Technology.

http://www.yict.com.cn
Yantian International Container Terminals.

Positivism/Positivist Geography

D. Bennett, Carleton University, Ottawa, ON, Canada

Glossary

A Posteriori With respect to propositions and to knowledge, deriving their authority from the evidence of experience.

A Priori With respect to propositions and knowledge, truths which are known prior to experience: not warranted by experience.

Empiricism The philosophical doctrine that our knowledge comes only from experiences which arrive in the mind which is otherwise blank, a *tabula rasa*, through the five senses.

Epistemology The philosophical theory of knowledge, it is concerned with criteria for distinguishing between our beliefs and our knowledge claims.

Idealism A metaphysical theory about reality: it holds that what can be said to be real is restricted to the contents of our own, individual, minds.

Materialism A metaphysical doctrine that everything is made of matter, or more weakly, is somehow instantiated in matter. 'Mind' is a problematic notion here, as is consciousness.

Metaphysics That part of philosophy which ruminates on questions of being, of what there is, and why there is something rather than nothing. It is opposed by logical positivists, by analytical philosophers, and by postmodernists and deconstructionists, separately and on very different grounds.

Ontology Has many meanings in philosophy: it is used here to refer to the set of things that exist for a particular theory.

Proposition In modern logic, it is what is asserted when an indicative or declarative sentence states that something is true (or false).

Rationalism A blanket term for views in philosophy that base knowledge on reason and which usually involve intuition rather than experiences, or senses, as underpinning warranted belief, or knowledge.

Introduction

Positivism, positivist, positivistic, positivist science, positivistic science, positivist geography, positivistic geography – these words have been used widely and frequently in human geography over the last 30 or so years, and hardly ever in a positive sense. Usually, they are used so that some preferred philosophical, epistemological, ideological, or political stance can be propagated as positivism's negative, somewhat similar to the Canadian construction of identity as 'not American'. The tone is often pejorative.

Positivism's home territory is philosophy, specifically philosophy of science. So, that is where we start, by describing its historical, intellectual, and social contexts, where, when, how, why, and by whose actions it came to be, what its claims, concepts, and necessary doctrines were, what its problems and failings were, and what it left behind. One of geography's myopias is the view that geography changes, philosophy changes, but science stays the same. In fact, science as theorized and as practiced is not the same in the times of Comte (1798–1857), of the logical positivists (1920s and 1930s), of positivist geography (1953–73), and of postpositivist geographies (mid 1970s to late 1990s: there hasn't been a new one for a while). Much has changed since Newton developed the first fully fledged scientific metatheory. In his time there was one science, natural philosophy. Newton's universe was deterministic and certain. Chemistry, biology, geology, psychology, and the social sciences have since brought increasing levels of complexity, to which the relatively simple analysis appropriate to Newtonian physics may be of limited direct applicability. The biggest change to the theorization of science has been the adoption of explanation based on probabilities, and the acceptance that there are immutable limits to the processes of producing knowledge. By the time the Vienna Circle officially dissolved in 1938, physics in particular and science in general had started to come to grips with the fallout of the Einsteinian revolution, quantum theory, Heisenberg's indeterminacy principle and Godell's incompleteness theorems, all scientific products, all pointing to the impossibility of ever completing a science whose endpoint is certainty.

The biggest change to the practice of science has been one of scale, its industrialization and corporatization. After World War II, the goal directed research and development model honed in the Manhattan Project continued in the post Sputnik arms race and NASA's project to put a man on the moon before the end of the 1960s. This was a rearrangement of the relationship between science and the state which led to the situation that now seems 'natural'. In *Scientific Knowledge and Its Social Problems*, Ravetz says that applied science had become the basic means of production in a modern economy. Science and technology are now mass activities. The process of discovery is itself now industrialized, carried out within large, hierarchical institutions which exist to do this kind of work, orientated more and more toward applications which are defined first, with the research necessary to bring them about following on.

It might reasonably be supposed that positivist geog raphy would be any and all geographies which are underlain by the epistemological principles and meth odological practices of positivist science. In the same vein, positivist science would be any and all science underlain by the epistemological principles and meth odological practices of positivism, logical or otherwise. Thus, positivist geography would connect to positivist philosophy. This is, in fact, not so. Following the path from philosophy into geographical usage, we see that positivist geography was not particularly connected to positivism. Hardly any of the aspects of science which are described as positivist by geographers can be traced, even indirectly, to logical positivism. The core of logical positivism is an incredibly parsimonious and extreme empiricism which can be summed up in one sentence. 'The meaning of a proposition is the method of its verification'. That is it. The putatively objectionable aspects of science – such as objectivity/disinterest, value freedom, and rationality – described by unsympathetic social scientists are neither necessary parts of positivism, nor necessary consequences of positivism for science.

Logical Positivism in Philosophy

Aims of the Positivist Project

Logical positivism asks the question: what does it mean to say that a proposition is correct? Essentially, logical positivism is empiricism pushed to the extreme, absolutely as far as it can go. It is antimetaphysical, anti idealist, and convinced that science alone can provide knowledge. Knowledge exists only in the form of empirically confirmed propositions. All else is belief or opinion, and belongs to the enormous category of statements which are deemed to be nonsense or meaningless, in the sense that Wittgenstein defined meaning (see below).

Logical positivism is a robust and vigorous defense of science as the Vienna Circle (Weiner Kreis) understood science to be conceptualized and practiced in the late 1800s and early 1900s, a defense mounted explicitly to stymie the influence of Germanic idealism which the positivists held to be deeply irrationalist and necessarily prone to making unexaminable knowledge claims based on an intuitionist epistemology. There was a small group of like minded activist philosopher scientists in Berlin, led by Hans Reichenbach (1891–1953). They held the same anti idealist views and attitudes toward science as did the Vienna positivists but the two groups remained distinctive in membership. The Wiener Kreis was by far the more influential. From now on this account treats logical positivism as synonymous with the Wiener Kreis.

The logical positivists wanted to establish that only empirical, publicly examinable evidence could furnish statements of belief with the warrant necessary for them to become truths. In the view of the positivists, Germanic idealism had all kinds of potential for social, cultural, and political intolerances, based on private and putatively superior claims to have access to or create knowledge. The positivists lay in direct line of descent from the Enlightenment and British empiricism, which is perhaps why they had more impact in Anglo America than they did in continental Europe. Their values were such that they would have been quite at home in England or America of that time, so much so that they probably would never have existed as a distinctive movement in either of those countries, as there was no threatening public acceptance of idealism as a practical mode of thinking for them to fight against. Idealism was confined harmlessly to the arcane world of the universities, where it dominated, rather than being the dominant component of public discourse, as it was in the German culture realm.

Contexts

In 'replacing positivism in medical geography', Bennett argues that a fuller understanding and appreciation of logical positivism may come from trying to see the world that the positivists saw, by trying to envisage their place in the broad intellectual contexts and more local and specific geopolitical situations within which they lived and worked. He suggests that positivism can be under stood as the constructed philosophical architecture arising out of an engagement with the perennial issues of empiricism versus rationalism as played out in the immediate social, cultural, and political circumstances in interwar Vienna.

Philosophy and Knowledge

The broadest context for positivism is the history of post Medieval Western philosophy which consists of conflicts between idealism in its various forms and materialism as ontologies, and rationalism and empiri cism as epistemologies.

Idealism is a metaphysical theory about the nature of reality. It holds that what is real is in some way confined to or at least related to the contents of our own minds. There are three major forms of idealism: (1) the sub jective idealism of Bishop Berkeley, (2) the objective idealism of Schelling and Hegel, and (3) the transcen dental idealism of Kant. They have in common the fundamental view that there is no access to reality apart from what is provided by the mind, and further that the mind can provide and reveal to us only what it contains. Note that this imagery suggests that 'us' are outside of mind, looking in/on. This is a problem. Nevertheless, reality consists in ideas. This is not the same as saying that reality is known via ideas: it is much stronger than that. Reality is ideas: ideas are reality. The extreme form is solipsism.

A materialist or physicalist ontology stands in utter opposition to idealism. It holds that the only thing that really exists is matter in its various states and configurations. Naïve materialism is an extreme form of robust commonsense which has a superficially reassuring, but unsustainable, hardheadedness, rooted in classical physics. However, as concepts of matter evolve post Einstein, crispness segues into fuzziness. And common sense turns out to be culturally specific.

Rationalist epistemology holds that knowledge can come only from the exercise of pure reason; the method is to set up a proposition that is beyond doubt – or about which it is possible to be 'virtually' certain – then see what further propositions this can be said to justify. Knowledge is said to be *a priori*, meaning that it is a category known to the mind without benefit of sense experience, which clearly resonates with idealism. It is a category about which sense experience can tell us nothing, and to which sense experience could not have led us. It is, therefore, an idea inherent to mind, and as such is something that mind 'imposes on' experience rather than being something which is known 'because of' experience. Space, time, and causality are examples of things which – say the rationalists – the mind knows intuitively, and which once known can be used to organize and interpret sense data. To the rationalist, no amount of sensory information could lead to the mind forming the ideas of, say, causality or determinism – neither can be directly experienced.

To the empiricist, the basic credo is Locke's (1632–1704) *Nihil in intellectu quod non prius in sensu* – there is nothing in the mind but what entered it by way of the senses. Empiricism holds that our immediate knowledge, or at least the propositional materials from which our knowledge is constructed, are delivered to the mind through our five bodily senses. It explicitly rules out any sixth sense, intuition, or extrasensory access to the mind. Naïve empiricism is the notion that all we are ever entitled to know about the world is what the world cares to tell us, on the world's terms. The mind is a *tabula rasa*, a blank tablet. In this view, the apprehender is completely passive and cannot (indeed, should not attempt to) influence the delivery of data via the senses. The empiricist tradition runs from Bacon, through Locke, Hume, J. S. Mill, and into the logical positivists.

Empiricists argue that the truths of mathematics are analytical and, therefore, cannot tell us anything about the world; they mostly agree that these truths are indeed *a priori* but they are not synthetic; that is, they do not define anything else but their own terms. They are an empty calculus, devoid of content about things in the real world. Empiricists claim that the things which science knows are synthetic – they tell us about more than just the terms in which they may be expressed – but they are based on experience, and so are synthetic *a posteriori.*

Neither naïve empiricism nor Cartesian rationalism – the extreme versions of each epistemology – is entirely coherent or persuasive, but no intermediate position is any easier to occupy. Coupled with the observation that deductive reasoning – favored by rationalists – can only reveal what is contained in its premises, and inductive reasoning is logically insecure, it is clear that knowledge – if defined as complete and certain – is not possible. Yet individually and collectively humans claim to have it.

Running throughout the post Copernican period is the related conflict between philosophy – as metaphysics – and science – as empiricism – over which of them addresses the really important questions, and which deserves the greater respect as a provider of deeper explanations, and as a repository of truth. The metaphysicians are the builders of great all encompassing philosophical systems, primarily exploring questions of 'being', God, mind, and freewill, self evidently important topics, worthy of the attention of only the most powerful minds. They are contemptuous of the analysts as nihilistic technicians, incapable of capturing the grand sweep of speculative thought as it seeks to find humanity's essence, to preserve the soul. Thinkers such as Plato, Descartes, Spinoza, Hegel, and Sartre fall into this group. Their opponents are the critical or analytical philosophers who regard the propositions of metaphysics as neither misguided nor wrong, but as fundamentally meaningless. Among the analytical philosophers are Moore, Russell, Wisdom, Wittgenstein (of the Tractatus) and Ayer, and, of course, the logical positivists. Analysts regard metaphysical puzzles as illusory. From their viewpoint concreteness is something that we bestow on these topics by our failure to see that they arise from the use of inadequately specified language. They are not 'real' problems at all but meaningless puzzles which exist only because we are hopelessly entangled in our own creation, language. Analysts, the logical positivists among them, analyze metaphysics out of existence. Naturally, metaphysicians object that this is some kind of trick, a solution achieved by sleight of mind.

The metaphysician makes no claim to be able to establish truth as it relates to fact. He or she sees no need for empirical confirmation. A proposition takes its place in the edifice of thought only because it fits with the rest of the structure and not because it conforms to an external 'fact'. The metaphysician's test is one of 'coherence'. The critical or analytical philosopher, on the other hand, sees no reason to accept any statement unless it is either empirically confirmable or is an analytical (that is tautological) truth. That is, (s)he favors a 'correspondence' theory of truth. These styles of phil osophizing are incommensurable.

Political and Social Contexts

Logical positivism was constructed in the discussions and publications of the Vienna Circle between 1907 and 1938. Although membership changed over that period, Schlick, Neurath, and Carnap were central. Many of the Circle were Jewish, Marxist, or both. After World War I, Vienna was a socialist city, almost a city state, within a territo rially much diminished Austria, a fragment of the multi ethnic Austro Hungarian Empire that had entered the war. The city was in the hands of the socialists between 1919 and 1934, whereas the federal state was dominated by a coalition of reactionary pan Germans and Christian socialists who kept alive traditional anti Semitism and ideas of Aryan superiority which became fertile ground for Nazism. In the 1930s, Europe seemed to be gearing up for an ideological showdown. Fascism was on the rise in Germany, Italy, and Britain, and Marxist totalitarian ism was firmly established in the Soviet Union. Right and left in Austria had private armies, streetfighting militias ready for civil war.

Uniquely in German speaking Europe, philosophy had long informed much of public life, through the formation and expression in actions and institutions of shared values rooted in a culturally and linguistically distinctive form of idealism. German metaphysical idealism – as expressed in the writings of Kant, Hegel, Schelling, and Fichte – had as its goal the vindication of belief in an *a priori*, extra empirical knowledge of time, space, God or the absolute, and the advancement of the *Volk* and the interests of the ethnically homogenous nation state. Prior to the logical positivists the only German language philosopher to criticize idealism was Schopenhauer (1788–1860), whose assessments of Fichte, Schelling, and Hegel as 'sophists', 'charlatans', 'shams', and debauchers of language and logic is unsurpassed for vitriolic invective of one eminent philosopher against another. Schopenhauer was also the first philosopher to speak out against his profession's penchant for obfuscation and pseudo profundity.

Logical positivism was heir to Schopenhauer's position. It developed to defend science against the repressive potential of idealist claims to have privileged access to knowledge, intuitively and therefore unexaminably known.

Although the logical positivists were not overtly politically motivated, by setting themselves directly against German romantic idealism they had inevitably entered political territory. They had taken themselves into the broader politico cultural arena in which idealist claims of epistemological superiority supported theories of social and ethnic superiority. The Vienna Circle came to be seen as dangerously subversive.

Neurath accepted this and urged the Circle to act in the manner of a political party whose platform was the destruction of traditional metaphysics, which he con sidered to be socially and politically reactionary and oppressive. Ironically, this is the current conventional perception of positivism among the leftist establishment in the humanities and social sciences.

Science of the Day

The members of the Vienna Circle were either scientists or mathematicians with philosophical interests, or phil osophers with interests in understanding how the natural sciences seemed to be able to produce knowledge, which is defined as warranted belief. Neurath was the only so cial scientist. Most of the scientists were physicists. At the theoretical level, it was hard to distinguish whether an individual was primarily a physicist or a mathematician. When the logical positivists referred to science they meant experimental and theoretical physics as practiced in the late 1800s and early 1900s. The science of their day was craft science, involving a relatively small inter national community, where it was possible for any one practitioner to know, or to at least know of, just about all the other members. Communication was personal, slow, and often involved traveling to meet. The role of the state was minimal. The technology was largely mechanical. It was unique and home made. Scientists were for the most part either amateurs or university based experimenters who really did share many of the values such as dis interest, neutrality, a belief in objectivity, supranationalist openness, and sharing of information which contempo rary postpositivist critics of science mock.

Logical positivism can be seen as the endgame of the philosophy of pre Einsteinian science. The science that the members of the Vienna Circle knew and believed in was the Baconian–Newtonian view that laws of nature were available to be discovered, which science could do by compiling observations in an unconceptualized and dis interested way, and in these data patterns could be dis cerned, materials for the construction of an edifice of correct, universal, reliable, and certain knowledge. The materiality of the Industrial Revolution and the ability of laws and theories to predict gave practical support to this view, although this merely avoided and did not solve the problem of inductive inference. So, the positivists sought certainty, led no doubt by their attachment to the deductive certainty seemingly offered by mathematics and logic, underpinned by the tangible successes in physics which led some to predict that science would soon come to an end.

Logical Positivism: The Vienna Circle

In 1902 Hans Hahn (mathematician), Philipp Frank (physicist), and Otto Neurath (sociologist and political economist) started an informal study group focusing on Ernst Mach's work. They met in coffee houses in Vienna. Mach was a radical empiricist who believed that what science had to say was anchored in and restricted to its observations which are known only as direct and immediate sensations, delivered by our bodily senses. He made significant contributions to mechanics, optics, acoustics, cosmology, and the philosophy of science. He opposed obscurantism in all fields.

Frank left to take a chair in theoretical physics in Prague in 1912, and Hahn was away from Vienna between 1915 and 1921. Mach died in 1916. In 1922 Moritz Schlick was invited to take Mach's old chair in Inductive Philosophy at the University of Vienna. Schlick established the Wiener Kreis in 1924: the Circle's official name was the Verein Ernst Mach (the Ernst Mach Society). Hahn, Frank, and Neurath joined the Circle, along with Waissman, Zilsel, Feigel, and Kurt Godel. The Circle never had more than 20 members, a core group and a penumbra who attended for a time, left, sometimes joined again. In 1926 Rudolf Carnap joined the Circle: he was effective in disseminating the group's ideas, but not as effective as A. J. Ayer. The positivists invited Ayer to sit in as a guest during his visit to Vienna on his first honeymoon during the winter of 1931–32 and out of that experience came *Language, Truth and Logic*, published in 1936 when Ayer was 26 years old. It is not an original philosophical work, yet it established Ayer's reputation and career as an analytical philosopher. Ayer's contri bution was as a spokesman: he introduced the logical positivists' ideas to the English speaking world. In a paper published in 1931, Blumberg and Feigel had given the Circle the name by which it has been known ever since, the logical positivists.

Between 1931 and 1938 the Circle gradually declined as members died, moved away to Britain or America, or as in the case of Schlick were murdered – shot on the steps of the university by a student variously described as 'deranged' or as 'a Nazi sympathizer'. Many German philosophers, including Heidegger, pressed for re strictions on who could teach subjects, such as philosophy or history of science, which had any bearing on Germanic culture. Only pure Germans, they said, should be allowed to teach German youth. This was part of an official antipathy to the circle, which gradually disintegrated. In 1938 the Ernst Mach Society dissolved itself, and in that same year its publications were banned from sale in Germany and Austria.

Influences

The logical positivists' project was similar to Comte's in that both made universalist claims for 'science' and both are called positivism. However, there is no evidence that the logical positivists were even aware of Comte's work. The logical positivists had little knowledge of and no use at all for traditional philosophy except insofar as they saw themselves as continuing the tradition of classical British empiricism in the style of Hume. More immediate and strong influences extended into the Vienna and Berlin groups from the work of Moore, Russell (of the *Principia Mathematica* period), and Wittgenstein at Cambridge, and of Gottlob Frege at the University of Jena in mathematics and logic.

Russell continued the empiricist tradition of Hume. He shared Hume's view of causality, basically favored the primacy of evidence for our beliefs, and gave favored status to the deliverances of our senses in adjudicating between beliefs. He had a degree of natural sympathy for the logical positivists, as he shared their goal of banishing all talk of superior worlds and metaphysics, but he never joined or even visited the Circle. Russell and Frege created modern logic, and Russell influenced Wittgenstein. Collectively they influenced the logical positivists, who in turn had contact with and influence on, first the analysts who turned their attention to the use of language – as in the Ordinary Language school at Oxford – and second, critical rationalism as created by Popper as an attempt to deal with the problems of verifiability and induction which stymied the logical positivists.

The positivists were impressed also by Einstein's relativity theories, particularly the way in which they built on the Michelson–Morley experiment (1885) by taking an experimental result which was at odds with Newtonian physics to be correct, which meant that Newtonian theory was incorrect, or, at least, insufficiently accurate. When observations of the solar eclipses of 1919 and 1922 confirmed that light rays were indeed bent in the vicinity of a massive object (the sun), and that the bending was by the amount predicted by Einstein's theories, they took this to be vindication of their faith in the empirical and empiricist basis of knowledge. As a bonus Einstein showed that Kant's assertion that certain basic concepts, such as time and space, are known in tuitively is undermined by science's demonstration that seemingly intuitive Newtonian notions of time and space are wrong.

Wittgenstein

The strongest influence on the Vienna Circle was Ludwig Wittgenstein, specifically his *Tractatus Logic Philosophicus*. Wittgenstein was born into one of the wealthiest families in Vienna in 1889. An unexceptional student at school, he did well enough to train as an engineer in Germany before going, at the age of 19, to Manchester, England, to carry out experimental work in the infant subject of aeronautics. From the applied mathematical problem of designing a propeller he was drawn into pure mathematics, into logic, and eventually into the philo sophical foundations of mathematics. Gottlob Frege suggested that he contacted Bertrand Russell at Cam bridge. Wittgenstein did not formally enter Cambridge University. He simply moved to Cambridge in 1911 and attached himself to Russell, rapidly earning Russell's respect and friendship.

When he arrived in Cambridge, Wittgenstein entered a distinctive philosophical world. Since Newton, Cam bridge had been a center of scientific and mathematical studies, but in the early years of the twentieth century a markedly new style of philosophy was being created, the Cambridge school of analytical philosophy. This was aggressively opposed to all other forms of philosophy which were lumped under the label of 'traditional' philosophy. The main concern at Cambridge was to analyze the meaning of statements, especially scientific statements, through a commonsense prism. It is difficult to appreciate now, but before Russell philosophers had for 2000 years used language without ever being par ticularly self conscious about it. Language was held to be transparent. Russell freely acknowledged that initially he treated language as transparent, but in this he was con tinuing the mainstream of philosophy: it was from his studies in mathematical logic and behavioristic accounts of knowledge that he came to give greater emphasis to linguistic aspects of epistemology.

After 2 years, Wittgenstein reckoned he had learned all that Russell could teach him, so he left Cambridge, went to Norway, and built a hut in the mountains where he stayed, thinking, until World War I broke out. He volunteered for service in the Austrian army and served as an artillery officer until he was captured by the Italians. Throughout his captivity he worked on his philosophical ideas which he wrote down in exercise books which he carried in a knapsack.

The 'exercise books' were published as *Logisch Philo sophische Abhandlung* in 1921 in the periodical *Annalen der Natur Philosophie* and as the *Tractatus Logico Philosophicus* in English in 1922. Russell contributed an admiring introduction. The Tractatus was written under great physical duress, an extraordinary feat of sustained con centration. It is a radical break with philosophy as it then existed. Wittgenstein was largely unconstrained by pre cedent and authority, having initially been educated outside of traditional philosophy. In the preface to the Tractatus, he says, "I do not wish to judge how far my efforts coincide with those of other philosophers. Indeed, what I have written here makes no claim to novelty in detail, and the reason why I give no sources is that it is a matter of indifference to me whether the thoughts I have had have been anticipated by someone else" (Wittgenstein, 1974: 3).

Wittgenstein thought that he had resolved the essen tial problems of philosophy by showing that they lay in the insufficiently precise use of language as the vehicle of thought – that is, they were pseudo problems – so after his release from captivity he turned his back on philosophy, gave away his family fortune, and worked as a village schoolmaster in Austria from 1920 to 1926.

As the Tractatus became more widely known, attempts were made to persuade him to rejoin the academic world, and in 1929 he returned to Cambridge, to a fellowship in Trinity College. In that same year he submitted the Tractatus as his thesis for the PhD degree at Cambridge. His examiners were Russell and Moore. Moore's report is said to have been: "It is my personal opinion that Mr. Wittgenstein's thesis is a work of genius; but, be that as it may, it is certainly well up to the standard required for the Cambridge degree of Doctor of Philosophy." Until 1935 he taught a small group of students whom he repeatedly exhorted to find something more useful to do than philosophy. Wittgenstein felt compelled to do philosophy, but he seems to have been quite skeptical about what it could achieve. He ends the preface to the Tractatus by saying: "I therefore believe myself to have found, on all essential points, the final solution of the problems. And if I am not mistaken in this belief then the second thing in which the value of this work consists is that it shows how little is achieved when these problems are solved" (Wittgenstein, 1974: 4).

His post Tractatus ideas were more widely circulated in the form of notes made by these students, without his approval. The 'bootleg' notes became the 'Blue Book' and the 'Brown Book'. The ideas of the post 1929 period stand in sharp contrast to those in the Tractatus. In a sense there are two Wittgensteins. 'Wittgenstein the Younger' influenced the positivists. 'Wittgenstein the Elder' is favored by those who are fascinated by the ambiguities in natural languages as they are used. In 1939 he was appointed to a chair of philosophy at Cambridge which he resigned in 1947. He died of cancer on 29 April 1951. *Philosophical Investigations* was published posthumously in 1953.

The Tractatus

Wittgenstein produced two major and contradictory works, the Tractatus and the Investigations. They have in common a concern with language. They start with the almost banal assertion that thinking involves language: indeed, thinking takes place in language or, at least, the external expression of thought does. Wittgenstein radically shifted the course of English language phil osophy onto the explicit analysis of what philosophers said (or wrote): thus, philosophy became the activity of analyzing rather than the mastering of a body or tradition of knowledge.

The Tractatus is not a book one 'reads' in the conventional sense. It is a dense, deductively argued structure of numbered propositions, a mere 74 pages in the 1974 Routledge edition, which requires a sophisti cated knowledge of symbolic logic in places, and a level of concentration in the reader well beyond the ordinary. Even then, much of it is likely to evade one's grasp. At the end it is truly enigmatic.

However, the main themes can be stated roughly. First, Wittgenstein developed a pictorial theory of language in which word elements correspond to elements in the facts which make up reality. Thus, "Proposition 2.021: Objects make up the substance of the world" (and) "Proposition 2.02: Objects are simple." Word elements can be put together to form significant sentences, but the structural principles of the language cannot be expressed in language. They can only be 'shown'. For two people to understand the same thing by what is said is for them to see the same picture. The 'simples' correspond to elements of the world ("Proposition 1: The world is all that is the case") but in language they must 'cohere'.

Second, Wittgenstein created the picture theory of language so as to draw a line between meaningful and nonsensical statements. Thus, the following propositions:

4 A thought is a proposition with a sense.
4.001 The totality of propositions is language.
4.002 Man possesses the ability to construct languages capable of expressing every sense, without having any idea how each word has meaning or what its meaning is – just as people speak without knowing how the individual sounds are produced.

Everyday language is a part of the human organism and is no less complicated than that.

His third theme is that philosophy is an activity, a process, not a body of knowledge at all, and certainly not a science. It is the job of philosophy (now rid of meta physics) to elucidate and clarify…although how this is possible without making interpretive statements which are nonsensical has never been clear! Thus:

4.1 Propositions represent the existence and non existence of states of affairs.
4.11 The totality of true propositions is the whole of natural science (or the whole corpus of the natural sciences).
4.111 Philosophy is not one of the natural sciences.
4.112 Philosophy aims at the logical clarification of thoughts. Philosophy is not a body of doctrine but an activity. A philosophical work consists essentially of elucidations. Philosophy does not result in 'philosophical propositions', but rather in the clarification of propositions.

In this the logical positivists saw the authority to authenticate their predisposition to valorize 'science' and denigrate philosophy. To them the boundary was clearly drawn: only science could manufacture true propositions, and it could do so because it had access to empirical verification. All else was nonsense, save tautological truths. The positivists followed Wittgenstein in holding that mathematics was a universe of tautological truths in which there was no meaning at all: it was seen as an empty calculus, whose truthfulness came not from tap ping into some essential current of higher and purer truth, but rather from our having created it and having 'bestowed' this property on it.

The argument that only an ideal language modeled on the statements of logicians and scientists permits us to make meaningful statements means that Wittgenstein could not expect the Tractatus itself to be accepted as containing truth. Russell saw this immediately: in his introduction to the Tractatus he wrote that he had some "… hesitation in accepting Mr. Wittgenstein's position, in spite of the very powerful arguments he brings to its support. What causes hesitation is the fact that, after all, Mr. Wittgenstein manages to say a great deal about what cannot be said, thus suggesting to the skeptical reader that possibly there may be some loophole through a hierarchy of languages, or by some other exit" (Wittgenstein, 1974: XXI). Wittgenstein had put forward a 'verbal' argument that meaningfulness depended on a 'pictorial', that is nonverbal theory of correspondence.

We cannot stand outside of our own minds in order to examine our minds: there is no place to stand. Witt genstein pointed out that we cannot think of what lies outside thought, and that we cannot speak of what lies outside of language. That is, we must be silent, but not about anything. According to Ayer, the Cambridge philosopher Ramsey cheerfully pointed out "What you can't say, you can't say, and you can't whistle it either." Wittgenstein confronted the problem of distinguishing between statements that are genuinely meaningful and those which are nonsense. No doubt much of what is said inside and outside of philosophy is nonsense, and it was an act of great courage to say so and then to go on to try to work out a method for making the distinction. He failed. Everyone does.

Wittgenstein was well aware that people would not stop using everyday language just because it did not fit the mold of cognitive truth. People learn language and use it to communicate emotional, ethical, or other exhortations, which can be claimed to be truths, but which cannot enjoy empirical verification. Wittgenstein was also well aware that almost all human discourse takes place in this world about which nothing meaningful can be said. Wittgenstein obviously lived a life in which he routinely said things that were not in his terms meaningful, as we all must. He had a strong mystical streak which breaks through periodically in the Tractatus.

By 1929 when he returned to Cambridge, Wittgen stein had decided that his picture theory of language was wrong. He abandoned his view that language had a level of essential meaning. He then sought to study language as it is used by human beings to communicate, rather than concentrating on the logic of the statements that philosophers made. He argued that the purpose and context in which something is said is vital to appreciating what it is that is being said. He also reached the conclusion – very dismaying for positivists – that it was not possible to clarify a statement by translating it into another statement in the same language. His extreme reluctance to publish his thoughts or see others discuss them in print lest they be misunderstood or misconstrued eventually brought him to repudiate much of what the positivists made of his one written statement, the *Tractatus Logico Philosophicus* (1921).

The method of proceeding suggested by Wittgenstein was for philosophers to devise language situations or games where the ways in which words were used could be studied to see how confusion arose by the same word being used in slightly different ways. Russell came to disagree profoundly with the later Wittgenstein and he thought that the exploration of language for its own sake was trivial.

Logical Positivism: Concepts and Problems

Schlick alone of the logical positivists considered what they were doing to be philosophy. The others agreed with Mach's claim that he was not constructing a philosophy but trying to unify science as something different from philosophy. Carnap was adamant that there could be no accommodation with philosophy. Philosophy (metaphysics) had to be destroyed. In this they were reflecting back the hostility that Germanic idealism directed toward science. The basic problem facing the logical positivists – which neither they nor anyone else has ever solved in any final sense – was how to fashion a conceptual tool which would cleanly separate scientific knowledge from metaphysics without destroying science.

The positivists rejected metaphysics because its pronouncements such as 'The absolute is beyond time' could not possibly be confirmed by anything we could experience. As an example (quoted by Quinton, 1982: 168) of the sort of thing they declared meaningless, consider the following, from Heidegger's *Hegel's Concept of Experience* "…the absoluteness of the Absolute – an absolution that being absolvent absolves itself – is the labour of unconditional self certainty grasping itself." It is hard to tell, but the statement seems to be grammatically correct, but what does it tell us? Meaning does not necessarily follow from grammar. To dismiss the statements of metaphysics was not novel: they had often been described as useless, empty, or unscientific, but the positivists raised the acrimony level to a new high by adopting Wittgenstein's view that the statements were meaningless.

Likewise, they dismissed conventional epistemology because no number of experiences could confirm the presence or absence of an external world. Realism and idealism were considered to be equally meaningless. The existence of an external world was a necessary 'assumption', not an empirical conclusion.

They saw in Wittgenstein's work the legitimization of their predispositions to throw out all that was not empirically knowable. They took from his work the view that accepting a proposition means the set of experiences which together are equivalent to the proposition's being true. They also believed that Wittgenstein had given a satisfactory empirical account of mathematics and logic by showing that they are analytical – that is, tautological truths. This was pleasing to them because the traditional rationalist view had been that mathematical truths are of a higher order than those based on experience. The positivists did not want there to be any other source of truth, certainly not a superior one, so they embraced Wittgenstein's demonstration that logical and mathematical truths are independent of experience only because they are empty of meaning. This did, however, come to cause them problems in connection with verifiability and the more general problem of inductive reasoning.

They thought they had found what they needed in the Tractatus. At their weekly meetings held around a rectangular table in a room at an institute near the university, they read out loud and discussed the entire Tractatus, line by line. Wittgenstein had, by this time, concluded that the ideas in the Tractatus were wrong, and he was headed in the opposite direction. Hints of this possibility are found toward the end of the Tractatus where Wittgenstein makes some enigmatic and somewhat mystical comments which indicate that he believed that all the important things in life lay in the realm of what could not be 'said' – that is, they lay outside of cognitively meaningful statements. This ran counter to the positivists requirements, so they simply ignored it.

They accepted the Kantian division of *a priori* and *a posteriori*, which when cross classified with analytical and synthetic yielded a four category taxonomy of all logically possible knowledge. They were persuaded by Russell and Wittgenstein that analytical *a priori* statements were true because they were tautologies. Analytical *a posteriori* statements were logically impossible. That left the Kantian synthetic *a priori* and the synthetic *a posteriori*. The synthetic *a priori* is what they wanted to get rid of by denying its possibility in the first place. In a philosophical version of last man standing, that would leave only the synthetic *a posteriori*, the category of empirical truths, those of science.

To demarcate the synthetic *a priori* from the synthetic *a posteriori*, they proclaimed the essential notion which defines positivism, the verifiability principle, justification for which they found in the Tractatus. The verifiability principle is the key to the positivist doctrine that "the meaning of a proposition is the method of its verification." Only those assertions which can be verified by experience can have meaning.

To their chagrin the positivists found that the verifiability principle did indeed hit metaphysics, but it was more a shotgun blast than a rifle shot and it hit science too. In trying to drive a wedge between metaphysics and science, they would have to rule out as meaningless the very scientific laws they had set out to defend. They realized that scientific laws are not conclusively verifiable, for there is no set of experiences such that having these experiences is equivalent to the truth of a scientific law. The laws of science could not themselves be verified empirically. They were subject, as are all empirical truths, to the limits set by the general problem of inductive or ampliative inference. This limit is intractable in logic. Popper found a way of making it go away, rather than solving it in the conventional sense, by developing a philosophy centered on falsificationism with all this implied for the end of certainty as an achievable goal of science.

Another set of problems arose out of the questions of what sorts of things would qualify as verifiers. In pursuing their goal of unifying all knowledge, the positivists were anxious to remove the division between external and internal worlds which seemed to set psychology apart from the physical sciences and they tried to do this by saying that in both instances the raw 'stuff' of a person's construction of the world consisted in experiences. But how can one show that one person's experiences are identical with another's? If one cannot, then an experience based science is fundamentally subjective. The ultimate content of science, lying in experiences, is forever and in principle private. That seems to be no improvement over idealism, as public observation is logically necessary to counter exclusive and unexaminable claims to have access to truth.

During the meetings of the Circle, Schlick and Neurath sat at opposite ends of the table. In all the time that Ayer sat in as a guest, discussion centered week after week on *Protokolsatze*, basic perceptions which Schlick insisted are delivered by sense experiences, about which the knowing subject could not be mistaken. Neurath argued that one must start (and stay) at the level of physical objects, that no beliefs were sacrosanct, not even those which seemed to be delivered by one's own senses. He was eventually forced to conclude that experiences are, ultimately, private so they could not be the verifiers of statements. Only a proposition can verify a proposition: a statement can only be significantly compared with another statement. Not only is this sterile and circular, it drove Neurath to uphold a coherence theory of truth: positivism was supposed to do away with that touchstone of the metaphysicians' coherence – because it permitted the most elaborate structures to be built metaphorically in midair if only the parts of the structure fitted together. Neurath could not justify a correspondence theory of truth except by invoking criteria of coherence.

But it was the realization that the verifiability principle was flawed that proved fatal to the positivists' project. The positivists held that it was not necessary to be able to specify how a proposition could be subjected to the court of experience right now, but it was essential that it be amenable in principle to being so tested. The technology might not exist, or the necessary experiment might be of a scale, cost, or complexity such as to make it unlikely or currently impractical to carry it out. But that was seen as quite different from assertions that were in principle and forever beyond the reach of empirical verification. Unfortunately, the verifiability principle seemed to be just such an assertion. The positivists tried several ways to salvage the principle. But whether they called it a 'proposal', a 'recommendation', or an 'explication of a concept' they could not wriggle around the mortifying realization that the principle was being weakened from being an imperative to something that a metaphysician could simply choose to ignore.

They realized, then came to accept, that there were no conceivable experiences which would lend empirical security to the verifiability principle. They concluded that the verifiability principle was by their own criterion not a meaningful statement. It was itself a metaphysical statement. In crafting a wedge to cleave meaningful from meaningless statements, the positivists found that in being logically consistent they had to exclude statements such as the abstract statements of science which they wished to protect. Hoist on its own petard, positivism began its long fade out.

By the end of World War II, all the surviving original positivists – now scattered to Britain and America – had abandoned the more extreme positions of the doctrine as being logically untenable.

Positivist Geography

Positivist geography is characterized by the central and explicit role given to the relationship between science and human geography in the context of the production of scientific knowledge. Although 'positivist geography' has a clear meaning, established by prolonged and consistent usage, it is nevertheless a considerable misnomer. It is generally used to signify the kinds of geography that emerged – first in the US and then in Britain – between the early 1950s and the early 1970s. This was a period of theoretical and methodological exuberance committed to transforming human geography into a nomothetic spatial science, urban and economic rather than cultural, social, or historical. Deeper philosophical correlates of the methods and techniques adopted remained largely un examined until critiques of so called positivist geography emerged in the early to mid 1970s.

The Positivist Era: An Overview

Between the publication of Hartshorne's *The Nature of Geography* in 1939 and Harvey's *Explanation in Geography* in 1969 the dominant project in Anglo American human geography was bringing science into geographic practice along with the adoption and development of spatial theory. Human geography was recast as spatial science.

Change in this period can be made to look orderly, rational, and inevitable when interpreted with the benefit of hindsight, but David Livingstone in *The Geographic Tradition* warns against writing history backwards, con structing it from a presentist starting point – so called 'Whiggish history.' At the time events seemed unguided by any conscious shared purpose or plan. Statistical techniques introduced at the research level filtered down into the undergraduate curriculum. Statistical analysis – univariate and descriptive to begin with – was aided by the increasing availability of large data sets, a by product of the technocratic state's monitoring of society as part of large scale social engineering. Efficiency criteria gave way to issues of equity, mostly racial in the US, social class in Britain. Early statistical work was computation ally constrained; mainframe computers (large, slow, expensive, difficult to use by today's standards) only became widely accessible in universities in the 1960s. Nevertheless, the cumulative effect of essentially tech nical innovations was to propel human geography into acting as if it were a science. Or, more accurately, acting in a way that the geographic revolutionaries thought was proper for a science.

There is an extensive literature on whether or not the positivist and postpositivist episodes should be inter preted as paradigms in the Kuhnian sense, or should be seen more as a proliferation of chaotic conceptions. Kuhn's notion of paradigm refers to changes along a single pathway – in which successive paradigms swallow previous paradigms – whereas geography is pluralistic, maintaining multiple coexisting parallel pathways which have quite different objectives. This suggests that a paradigm based historical interpretation is inappropriate.

Whether driven by a perceived need for intellectual respect, or as part of a spontaneous intergenerational movement to oust the old guard and to usurp power within the discipline – a sociological interpretation suggested by Peter Taylor in 1976 – or propelled by a profound dissatisfaction with the forms and standards of explanation employed in geography – a point of view often expressed by Peter Gould – human geography changed drastically in the positivist years. The change was initially referred to as the Quantitative Revolution. Only later, and in critique, was it labeled 'positivist geography'. In 1963 Burton proclaimed that the quanti tative and theoretical revolutions were over. Actually, resistance to quantification is still with us over 40 years later. Peter Gould argued that in concentrating on quantification, the traditionalists failed to see that the essence of the new geography lay not in the numbers but in a new way of looking at the world, seeking to make law like generalizations rather than see only particulars.

Until the 1950s geography identified itself with the idea of the region. Geography's claim to distinctiveness lay in conceptually integrating layers of phenomena so as to construct regions, the geographical equivalent of the historian's period. Starting with the geological base, layers were added – surficial geomorphology and landforms, drainage, vegetation, settlements, economic activities, and so on – as the geographer worked through a checklist of ingredients. "… and when you got to the bottom of the list you were finished. So nearly was the field as an intellectual discipline…" said Peter Gould (1979: 139–151). The epistemology of geography was naïve realism – not that geographers of that time talked about such things as epistemology or ontology. The world was considered to be simply and essentially what it was. It awaited discovery. It was considered to be discoverable because it was unproblematically accessible via our senses, which were assumed to add nothing to what was sensed.

The philosophy of geography resided in one book, Richard Hartshorne's *The Nature of Geography*. Hartshorne insisted that geography should have no truck with change, time, sequence, and dynamics. Geography should concern itself with a chorological approach, addressing areal differentiation as it existed at one instant in time. However, physical geography moved on from morpho logical concepts to processes and made contact with cognate disciplines, such as hydrology and meteorology. Ever since, the physical geographies have aligned themselves with science with scant questioning about its nature, positivist or otherwise.

In 1953 Fred Schaeffer challenged Hartshorne's nor mative grip on geography by setting out an anti excep tionalist philosophy drawing heavily on logical positivist ideas, as transmitted by Bergman, one of the Vienna Circle's participants who had fled to the USA. Schaeffer and Bergman were colleagues at the University of Iowa. Schaeffer argued that geography could share the meth odology of the sciences – or that it should – because uniqueness is not unique to geography. All knowledge deals with events that are unique in their manifest detail, but all events can be regarded as members of classes about which general statements can be made if we are prepared to stand back a little and see them as related in some relevant ways. The only thing one can do with an event we choose to see as unique is contemplate its uniqueness. Schaeffer argued that geography should generalize about its phenomena, should seek explanations of a law like character and he saw no necessary impediment to so doing in the concern with the particular, because scientists apply to each individual concrete situation all the generalizations that they believe are relevant. Without generalization, language itself would be impossible. A law is simply a descriptive generalization of a high degree of reliability: it need not be perfect, it may be probabilistic, and it may or may not need theoretical amplification. Schaeffer died just before his paper was published and it fell to Bergman to shepherd it through to publication.

The scholarly as distinct from the power relations aspect of the dispute centered on the idea of uniqueness, but this was confused with the idea of singularity. A singularity is an event to which the generalizations of science do not apply. For example, inside the event horizon of a black hole lies a singularity: there are many such phenomena in the universe so any one of them is not unique, it belongs to the class 'singularity'. In *Theoretical Geography* Bill Bunge argued along similar lines in asserting that locations are not unique. There is a contemporary echo in the claim that law like generalizations about human beings are not possible because each human is unique. This claim is only correct if each human is singular. Being unique in possessing a particular combination of general characteristics does not prevent an individual from being treated as a member of classes defined by criteria held in common but in different degrees and distinctive mixtures. The argument really is about the bases of classes such as gender, race, ethnicity, sexuality, etc.

James Bird has gone so far as to say that the only truly revolutionary change brought on by the new geography was the desire to make generalizations which enjoyed a high degree of empirical support, and that arguments over theory are about which theory – not theory versus no theory – and arguments over quantification are arguments over forms of evidence – not evidence versus no evidence. According to Bird, this revolution happened in 1966.

Although Schaeffer and Bergman were locally influential – Les King says that Bergman was *de rigeur* reading for graduate students at Iowa – only in a whiggish history does the Hartshorne/Schaeffer face off become the spark from which positivist spatial science grew. So called positivist geography actually came about quite independently of this specific debate.

Positivist geography in the form of spatial science developed in the USA during a time when science enjoyed great public trust and respect, largely because as applied through technology it was socially, economically, and politically fruitful. In World War II, science had produced radar, the jet engine, rocketry, plastic surgery, the industrialization of antibiotic manufacture, plastics, rudimentary computers, aerial photography, atomic bombs and nuclear power, the mass production of planes, ships, trucks, cars, and so on. American prosperity and Cold War security relied on science. In Europe, recovery involved rebuilding entire cities, industries, and transportation systems. Science provided cheap, plentiful medications, foremost the range of antibiotics and vaccinations which took much of the risk and fear out of childhood in North America and Western Europe, and high tech health care at first seemed to be capable of miracles. Mankind seemed to have entered an age of dominance over nature in which infectious diseases would one by one succumb to the power of science – or so it was thought. It turned out to be short lived and hubristic. By the early 1970s science's honeymoon was over: social and racial inequity in the USA persisted, spilling over into annual summer unrest culminating in the National Guard patrolling the streets of Detroit in tanks in 1969, echoes of Budapest in 1956. The Vietnam war was seen as an immoral high tech corruption of science in the service of the military industrial complex that President Eisenhower had warned about in his farewell address back in January 1961. Only one disease (smallpox) was a serious candidate for eradication, and science's intrusion into daily life seemed to be becoming less benign, altogether more sinister. Public skepticism became the norm, and in geography spatial science was equated with the inability to speak to social and political problems, and the whole lot was hitched to positivist science and thus to positivism. The critiques of so called positivism in the 1970s are very much children of their sociopolitical times.

Spatial Science

Human geography as spatial science began in multiple centers of innovation, more a post war sea change in intellectual objectives and expectations than an organized, coordinated, or planned movement. It began in America, spread to Britain, and made common cause with a similar movement that had started quite independently in Sweden. In the USA, the main centers of innovation were the extraordinary coterie of graduate students working under Bill Garrison at the University of Washington in the late 1950s who came to be known as the space cadets, and smaller and less influential groups at the University of Iowa, and Northwestern University in Evanston, Illinois.

The revolutionaries were open to theory from outside of geography: they had to be, geography had no theory of its own. Ever since, geography has been a net importer – one might say, a plunderer – of theory: spatial, behavioral, Marxist, humanist, social, structurationist, neo Marxist, feminist, postmodern, post structural, postcolonial, and so on, none of them homegrown.

There are several comprehensive interpretive accounts of this period, so the following is only a sketch, thematic rather than chronological.

The spatialists were influenced by the new discipline of regional science that came out of Walter Isard's pioneering work at the University of Pennsylvania, particularly *Location and the Space Economy* in 1956, and *Methods of Regional Analysis* in 1960. Isard was dissatisfied with traditional economics' blindness toward the role of space. He drew attention to the ideas of Losch, von Thunen, and Walter Christaller, who all worked within the traditional deductive framework of rationalist reasoning in a world simplified by sweeping assumptions, aided by mathematical and statistical techniques. Spatialist geographers adopted the regional scientists' isotropic surface as the starting point for model construction. The model actor was economic man, ungendered, rational, a maximizer, infinitely sensitive to changes in the information environment, omnisciently surveillant, computationally perfect, and unencumbered by family or emotion, and above all one who acts alone.

Pace Christaller, geographers – Brian Berry the foremost – found ranges and thresholds of goods and services, described and predicted settlement sizes, understood the distribution of settlements, calculated basic/nonbasic employment ratios, found hexagons – perfect and distorted – in the landscape, especially in the Great Plains of America and the prairies of Canada where life and landscape most closely imitated theory.

Settlements could be nodes at one scale, areas at another. If cities formed systems, then within cities there were systems too. Models of organization based on social and ethnic cohesion and exclusion, played out against a backdrop of land use controlled by bid rent curves, gave us Burgess's concentric rings, Hoyt's sectors, or Harris and Ullman's multiple nuclei model. The whole world's urban areas were seen through spectacles made in and optimized for Chicago, metanarrative rampant, until studies of Rome and Cairo showed that Shevky Bell Social Area analysis was culturally specific and did not travel too well. Descriptive models of social differentiation over space leapt forward computationally if not conceptually when statistical techniques met secondary data, and factorial ecology became fashionable.

Peter Gould experimented with game theory. He and his students Rodney White and Roger Downs brought perception into the positivist fold by conceptualizing and operationalizing mental maps, adding the isopercept – the line of equal perception – to spatialism's toolbox. Some spatial models were analogies drawn from classical physics. John Q. Stewart developed a new version of social physics: the astronomer Edmund Halley is thought to be the originator of the term social physics, and August Comte used it in connection with his newly coined term 'Sociology', the scientific study of society. William Warntz, educated as a geographer before World War II, trained as a meteorologist during the war, modeled income fronts sweeping across North America. Zipf adopted Newton in the general gravity model, applied to flows between nodes, flows of phone calls, migration, passenger and freight traffic. Stouffer added intervening opportunities. Geographers talked of spatial processes, then of social, economic, political, etc. processes in space – space being a container – which led eventually to contemporary ideas of space as constructed and produced by processes.

In Sweden, a strong planning tradition and a superb population register database supported a group of mathematical geographers of whom the best known is Torsten Hagerstrand. Hagerstrand experimented with logarithmic transformations of space to create maps showing the inverse relationship between interactions and distance. Richard Morrill from Washington spent a year in Sweden: he brought the spatialist conceptual machinery to Swedish data to build a simulation of the Swedish settlement system incorporating such conceptual novelties as post diction, the prediction of states earlier than now – so they can really be known and used as verifiers – from even earlier states.

In the 1960s, spatial science and quantification crossed the Atlantic, carried by Peter Haggett from his summer school visits to America, and by Richard Chorley, the Cambridge geomorphologist who did his graduate work in the US. What sympathizers and converts saw as the diffusion of exciting innovation, many of the old guard, such as Haggett's head of department at Cambridge, regarded as being more akin to a process of contamination. The ideas in *Models in Geography* co edited with Chorley, *Network Analysis in Geography* written by Haggett and Chorley, and *Locational Analysis in Geography* written by Haggett on his own galvanized British geography. David Harvey had finished his Ph.D. (on the geography of the nineteenth century Kentish hops industry!) at Cambridge in 1961 and was a teaching colleague of Haggett and Chorley before moving to Bristol where he wrote *Explanation in Geography* in 1968, publishing it in 1969. The climatologist/meteorologist Stan Gregory had introduced statistics to undergraduates at the University of Liverpool in 1963, the year he published *Statistical Methods and the Geographer* and the year I began to read for my own bachelor's degree in geography in that department. We students couldn't see why we should have to do statistics: most of us were 'artsie' mathophobes. By the time I started my postgraduate research at Liverpool in 1967, we had changed our minds, embracing the new ideas, methods and techniques, enjoying the discomfort they caused some of our more traditional professors, young and not so young. Simulations, Markov chains, factor analysis, principal components analysis, canonical correlation, general systems theory, numerical taxonomy, and allometry – from the work of Von Bertalanfy and D'Arcy Wentworth Thompson, via D. Michael Ray – 'quantifiers' took them all on board.

Assume a landscape, assume a decision maker, start the model, watch the outcomes. Androgenous match stick people move over bleak landscapes, guided by superb reasoning and information gathering powers, bent on finding optimal solutions to highly constrained prob lems, only occasionally diverted by a troublesome cost curve or resource constraint. This conceptual emptying of the world and its repopulation with an army of robotic actors was repugnant to many: it seemed inhuman, and unfruitful to boot. If this was science, if this was posi tivism in action, if you embrace positivism, then this is the world you get, probably the world you deserve.

In a macro sense the only possibilities to preserve spatialism lay in reducing the gap between the world as it manifestly can be described empirically and the world as predicted from models, by either making the assumed world or the spatial actor more realistic. Behavioral geography attempted to humanize economic man. Gradually, the two elements in this conceptual model, the world and the actors, became more sophisticated. The world became more plural and actors embraced new empiricisms, informed by more realistically complex domains of experiences – gendered, aged, sexualized, oppressed, etc.

Julian Wolpert investigated the decision making behavior of a sample of farmers in southern Sweden, focusing on rationality and optimality. He found that they actually behaved in a rationally suboptimal fashion, choosing a strategy which Wolpert called "satisficing" – settling for "enough" with a reduced risk of failure, rather than going for "the most" with an elevated risk of losing. In *Behavioural aspects of the decision to migrate* Wolpert brought concepts from social psychology to bear on the quintessential spatial process, human mobility. Here, decision making based on calculating and choosing among place utilities is rational with respect to the in formation available to the decision maker, largely a function of location within formal and informal infor mation fields.

Positivism in Geography

By the end of the 1960s, theory, method, and technique had made their ways into human geography. Philosophy was notably absent, despite it being widely recognized that the new methods at least implicitly brought along a relationship with science. Michael Hill in *Positivism: a hidden philosophy in geography*, argued that positivism entered geography by piggybacking on quantification, and that it needed to be 'exposed'. Hill even developed a question and answer self diagnosis test so that students could find out if they had become unwittingly infected with positivism.

Human geography as spatial science reached its zenith in Abler, Adams, and Gould's *Spatial Organisation: The Geographer's View of the World*, published in 1971. The first full and explicit account of the methodology underpin ning spatial science – David Harvey's 'Explanation' – had been published 2 years before.

By the end of the 1960s, many geographers had come to treat science and positivism as synonymous. The image which human geographers hold of science rarely comes from close acquaintance with original ideas in or about science and philosophy. It emerges from a few guidebooks, attempts to make coherent stories out of issues in science, philosophy, and geography. So, to some extent it was inevitable that these interpretations became opinion leaders. Geographers were grateful that someone had carried out the task of interpretation and summary for them, but it put them in a relatively dependent position, unequipped to question and challenge the new orthodoxy.

What do geographers believe science to be? Science, positivism, and logical positivism are spoken of in the same breath, and while they are related they are far from being the same thing. The images of science and positivism in geography were established early and once established they endured. By now they are largely taken for granted, part of the furniture. They were largely created by two works – Harvey's 1969 *Explanation in Geography*, a leader, and Johnston's 1986 *Philosophy in Human Geography*, a chronicler.

Harvey's 'Explanation' was the fullest statement of how the new geography should be done. It addressed the scientific method while categorically eschewing any need to consider explicitly the objects, orientations, values or domains of geography. David Smith has described this indelibly influential book as at once a milestone and a tombstone. It summarized a way of looking at the world that was coming to an end even as it received its first thorough expostulation. It was the starting gun of the race to postpositivism.

Harvey felt the individual personal need to write *Explanation in Geography* in much the way that quantitative geography collectively evolved beyond an infatuation with specific techniques. Starting with statistical analysis – which he cheerfully acknowledges he some times inadvertently misapplied, as did Ron Johnston, David M. Smith, and Peter Gould – Harvey came to see that although a philosophy of scientific method was implicit in quantification, it was not a necessary accom paniment to quantification. As he said, "Although it was no accident that quantification was forcing us to up grade our standards of argument, we could, if we so wished, up grade those standards without any mention of quantification. The issue of quantification per se there fore faded into the background and I became much more interested in the general issue of the standards and norms of logical argument and inference which geog raphers ought to accept in the course of research" (Harvey, 1969: vii).

Harvey's main concern was similar to Gould's, not a desire to see geography necessarily become quantified but to see it put onto a decent intellectual foundation shared with other disciplines, rather than continuing to claim to have a privileged exceptionalist epistemology which not only separated it from other intellectual enterprises but set itself up as superior. Adopting the scientific method meant at least the setting up and observing of decent intellectual standards for rational argument. He claimed that he was concerned solely with method, not with philosophy. Harvey equated philosophy with metaphysics, with value judgments and speculation, with inwardly directed criteria of knowledge. Basically he equated philosophy with rationalism.

Harvey claimed that what we believe to be the objects and objectives of geographic study cannot be contested on logical grounds, because this belief reflects our *a priori* judgments about what is and what is not worthwhile. This in turn constitutes our personal philosophy, and so he gave the label "philosophy of geography" to ruminations about the nature of geography, its objects and its goals. These, he claimed, cannot be analyzed away, but they can be clarified, and their coherence can be challenged and (possibly) improved. The aim of the geographer is to describe or explain the things that his or her philosophy designates as being geographical and worthwhile. Harvey saw little difference between description and explanation. But he saw a great deal of difference between 'phil osophy' (as he used the word) and method: explanations arise from method and are not value based. They are essentially logical, are at least standardized, and can be contested by comparison with 'objective' criteria. He claimed that methodology is value free, and is amenable to be judged to be correct in some final sense. The touchstone for establishing correctness is 'the' scientific method.

Harvey discussed hypotheses, laws, models, theories, and explanations. He outlined inductive and deductive forms of inference as they relate to the logic of con firmation. He described two routes to knowledge, the Baconian way of working back from the myriad events in the real world toward generalizations based on pattern recognition, and the method of hypothesis formulation and testing based on *a priori* knowledge. And he described some major model forms which explanations can take. Overall, he argued that science produces relatively cer tain knowledge, that this knowledge is most elegantly captured by high level abstractions which we call theory, which rest in turn on layers of laws, hypotheses, and facts, the whole having an internal coherence which comes from having a methodology in which things can be shown – empirically – to be so. Harvey's clear preference was for the deductive nomological form of explanation wherein the conclusions which follow necessarily from premisses must be correct if the premisses are also cor rect. He briefly acknowledged – but did not develop further – the difficulty that the information content of premisses (which basically are about the real world) arises from our believing that we already 'know' some things to be true, before we embark on the process of deducing other things that we then believe to be demonstrably, logically true. The foundation can only be provided by inductive inferences about the world, based on sense data, and to this extent they cannot fit with the standards by which truth can then be established. This sort of circularity is unavoidable in our reasoning.

Harvey further separated explanation from the 'human' acquisition of knowledge by concentrating only on the logical 'form' of explanation. The price of this is that the 'human' activities which lead to knowledge are ignored, and the social processes of getting and communicating knowledge are shunted to one side. The behavioral aspects of science as an organized activity, the product of a community with rules and practices that are neither self evidently appropriate nor universally unquestioningly shared, remain unexamined.

The kinds of geography that are possible in the world according to early Harvey have been criticized for being (1) mechanistic, (2) unable to account for human actions because the motivations, perceptions, and intentions of actors are not included, (3) based on unsustainable ideas of value freedom or objectivity, (4) unable to 'really' account for social and political deprivations and injustices, (5) supportive of the *status quo*, and so on. All these sins have been lumped together in geography as being the necessary outcomes of positivism, the brand of epistemology allegedly put forward by Harvey.

In the early 1970s it became easier and easier to question the social appropriateness of technology, of science, of the epistemological respect given to em pirically verifiable knowledge. In geography the ques tions grew out of the relevance debate – a response in part to the persistent and deepening social and eco nomic divides within affluent nations – and focused on an examination of positivism as geographers had understood it. The annual meeting of the A. A. G. in Boston in the spring of 1971 was a watershed. By 1973 Harvey, in an about face of Wittgensteinian surprise and courage, repudiated Explanation, and has never returned to it.

Explanation in Geography is generally seen as the definitive positivist statement in geography. Yet nowhere in its index 'positivism', 'Comte', and 'logical positivism' are found. The bibliography shows Carnap on prob ability, Bergman on the philosophy of science, three papers and one book by Hempel (one of the Berlin group), and three publications by Reichenbach (the leader of the Berlin group). On page 8, the expressions 'logical positivists' and 'logical positivism' make their only appearances. If positivism is there, it is implicit.

Harvey's book was original in that it brought many intellectually 'foreign' ideas into geography as part of a long and coherent argument concerned with changing the way in which we should 'do' geography. It was a formidable attempt to persuade us to change our minds. The second book which played a major part in forming geographers' collective image of science is *Philosophy and Human Geography*, which passed rapidly into a second edition, an indication of both the pace of change and of the demand for guidance in this difficult area. Johnston drew the more philosophical threads out of his earlier book on *Geography and Geographers* and attempted to survey and interpret the enormous surge of material in post explanation geography, and to link this to the trends in the philosophy of the social sciences. This is a self described 'primer', a survey and interpretation rather than an extended original argument, but its influence was probably all the broader as a result.

According to Johnston, the origins of positivism lie in the works of August Comte, published between 1830 and 1854. Comte believed that science was the only valid method of investigation and that science developed natural laws, the understanding of which gave scientists the means to manipulate phenomena because they were able to intervene in natural systems with confidence that the effects of their interventions would be predictable. Comte claimed that the same capabilities were, potentially, available in the social sciences; that is, the model of investigation should be the same even when the subject matter was markedly different, and the goal of uncovering laws as a sound basis for action was similarly ubiquitous. This sort of sweeping confidence was not uncommon in the Victorian era, and it reflected in part the feeling in the natural sciences at the time that the correct view of the universe was Newtonian, deterministic and mechanistic, and that the natural sciences were drawing to a close as the finishing touches were being applied to the church of reason.

Johnston stated that although the emerging social sciences were influenced by Comtean positivism, the major impact of positivism came from the logical positivists working in Vienna between the World Wars. "The logical positivists very largely accepted Comte's scheme of a positive science. They codified it with regard to current practice and focused most of their attention on how scientific enquiry might be conducted so as to lead to firmly established generalizations." In fact, as we have seen, the connections between Comtean and logical positivism are neither as direct nor as explicit as Johnston suggests.

Johnston (1986: vii) suggested that "…three competing philosophies (termed approaches here because each embraces a variety of related viewpoints) dominate contemporary human geography." They are positivist approaches, humanistic approaches, and structuralist approaches. In the positivist approach Johnston mixed together Comtean positivism, logical positivism, the "scientific method", the "positivist conception of science", the "standard view of science", and the "positivistic attitude".

By his own admission Johnston constructed his account of positivism entirely from secondary sources: there is no direct engagement with the ideas, writings or history of the logical positivists, positivism's inventors. There are no positivists in the bibliography. Popper is there, and Ayer, neither of whom were members of the Circle. Johnston relied heavily on the views of Russell Keat, whose opinions on *Social Theory as Science* (the title of the book he coauthored with John Urry) were guaranteed to put him at ideological odds with positivism, somewhat like asking the fox for advice on how to defend the hen house.

Where Harvey provided the first substantial account of scientific method for geographers, all too glibly labeled positivism, Derek Gregory developed the first substantial discussion of possible alternatives to positivism. Gregory focused almost exclusively on Comtean positivism, not logical positivism. He chose to locate his case for a critical social theory in a detailed critique of positivism in which he gave primacy to the Comtean rather than the logical variety. Gregory worked through a closely argued examination of the theory of scientific knowledge developed by August Comte in the 1850s, which made claims for a unified science of the natural and social worlds and which clearly conceived of mankind's passage through phases of knowledge in terms of evolutionary, unidirectional maturity. He linked Comtean positivism to logical positivism which he clearly considered to be derivative and of lesser importance, so gave it scant attention. He wrongly identified Wittgenstein as the dominant figure of the logical positivist movement, and gave the logical positivists credit for developing the distinction between analytical and synthetic statements when it is actually Kantian. *Ideology, Science, and Human Geography* is critique: a certain image of positivism is necessary to launch and justify Gregory's championing of radical social theory.

Gregory's early exploration of science as ideology has grown into one of modern geography's major intellectual explorations, of social theory, of the geographic imagination, of writing and representation, so it is perhaps a little churlish to point out that it started out by giving undeserved prominence to a form of positivism with faults and limitations quite different from those of the more relevant positivist variety. But Gregory seems to be a little uncomfortable with his own framing of positivism: he warns that his account is a "caricature" of positivism's more striking features, and that such a "casual dismissal of one position" hardly justifies "the adoption of another" (Gregory, 1978: 48).

What's Next?

Harvey gave us a treatise on method with scarcely a mention of positivism, Johnston said it was all second hand, Gregory said his account was caricature. So, why did we buy into the standard version of positivism – as Bennett describes it in *Replacing positivism in medical geography* – as the blending of Comtean positivism, logical positivism, science, empiricism, science politics, scientism, quantification, and so on into one unholy mix where the subtext is that positivism is shallow, reactionary, oppressive and on the side of the *status quo*?

In *The Geographic Tradition*, Livingstone (1992: 321) concludes that "…the label "positivist" within the geographic tradition has simply been used as a convenient term of reference under which to subsume its modern scientific aspirations, namely as a designator for a parsimonious orientation toward a mathematized account of the distribution of observable entities. Geography rarely engaged in any profound way the positivist epistemological program." Perhaps it is past time that it did so engage.

Positivism is a philosophy of meaning, of what it means when you say that something is true, when you offer it as something which others 'must' also accept because it is, according to some allegedly neutral standard, true. It seems that the reaction in geography to positivism was misplaced, and was really a reaction to the particular sorts of work being done in the name of science. Spatialism and positivism became linked in the geographer's mind, but they do not necessarily imply each other at all. The real question is: is the kind of placeless, abstract, and morally uninformed geography which was done under the heading of spatial science the 'only' kind that 'could' have been done under the banner of positivism? Is it the only kind that 'can' be done? No, positivism as a theory of knowledge does not 'necessarily' require the abstract geometrical world that spatialism produced. As a theory of knowledge which gives primacy to a standardized ability to interrogate the world as both a source and a confirmation of our 'ideas' it can refocus geography and the social sciences on the role of evidence, without unduly constraining the form that the evidence takes.

The logical positivists simultaneously abhorred, and were afraid of, the social, ethnic, and intellectual intolerance rooted in intuitionist claims to 'just know' which permeated German culture and philosophy. Individuals can claim to know anything they like as long as they claim to know it only for themselves. There is little chance of harm in this. But to claim to know how things really are based only on intuition – which by definition is private – is potentially dangerous because it is, as the positivists clearly understood, a claim that is social, that others must contend with. History is a sorry chronicle of miseries inflicted by those who know on those who know something different. Pluralism in the sense of epistemological richness is no guarantee of tolerance. Positivism says we should beware of those – geographers among them – who just know, and who proclaim rather than demonstrate the propositional truthfulness of what they know. Claims to know are coercive: intuitionist claims to know are dangerously coercive.

In *A Short History of Philosophy* (1996: 276), Solomon and Higgins pithily put the case for engaging directly and explicitly with positivism. "The positivists were not… narrow minded science fanatics and logic choppers. They were… champions of sanity in a world going insane…There is more than enough nonsense in the world, claimed the positivists and it is the job of philosophers, to the best of their abilities, to make sure there is no more of it."

In philosophy, a cyclical rather than a unidirectional interpretation of change suggests that a swing away from the current dominance of epistemological constructivism back toward some form of positivist empiricism is not out of the question. So perhaps it is fair to ask what can come next in a geography currently infatuated with a host of 'posts, postcolonialism, postpositivism, poststructuralism, and postmodernism'. What can be post the 'posts'? Perhaps many extremes are being reached, and only some form of retreat affords a way forward.

See also: Critical Rationalism (After Popper); Idealism/Idealist Human Geography; Quantitative Revolution; Regional Science; Scientific Method; Social Studies of Scientific Knowledge; Spatial Science.

Further Reading

Ayer, A. J. (1977). *Part of My Life*. London: Collins.

Ayer, A. J. (1984). *More of My Life*. London: Collins.

Bennett, D. (2005). Replacing positivism in medical geography. *Social Science and Medicine* 60, 2685 2695.

Bird, J. (1993). *The Changing Worlds of Geography: A Critical Guide to Concepts and Methods* (2nd edn.). Oxford: Clarendon Press.

Castree, N. (2005). Is geography a science? In Castree, N., Rogers, A. & Sherman, D. (eds.) *Questioning Geography: Fundamental Debates*, pp 57 79. Oxford: Blackwell Publishing.

Friedman, M. (1999). *Reconsidering Logical Positivism*. Cambridge: Cambridge University Press.

Gould, P. R. (1979). Geography 1957 1977: The Augean period. *Annals of the Association of American Geographers* 69(1), 139 151.

Gould, P. R. and Strohmayer, U. (2004). Geographical visions: The evolution of human geographic thought in the twentieth century. In Benko, G. & Strohmayer, U. (eds.) *Human Geography: A History for the Twenty First Century*, pp 1 24. London: Arnold.

Guelke, L. (1978). Geography and logical positivism. In Herbert, D. T. & Johnston, R. J. (eds.) *Geography and the Urban Environment*, vol. 1, pp 35 61. New York: Wiley.

Hay, A. (1985). Scientific method in geography. In Johnston, R. J. (ed.) *The Future of Geography*, pp 129 142. London: Methuen.

Hill, M. R. (1981). Positivism: A “hidden” philosophy in geography. In Harvey, M. E. & Holly, B. P. (eds.) *Themes in Geographic Thought*, pp 38 60. New York: St. Martin’s Press.

Honderich, T. (ed.) (1995). *The Oxford Companion to Philosophy*. Oxford: Oxford University Press.

Johnston, R. J. and Sidaway, D. (2004). *Geography and Geographers: Anglo American Human Geography Since 1945*. London: Arnold.

Livingstone, D. N. (1992). *The Geographical Tradition: Episodes in the History of a Contested Enterprise,* ch. 9. Oxford: Blackwell.

Magee, B. (1997). *Confessions of a Philosopher: A Journey through Western Philosophy,* ch. 3. New York: Random House.

Marshall, J. U. (1985). Geography as a scientific enterprise. *The Future of Geography*, pp 113 128. London: Methuen.

Monk, R. (1990). *Ludwig Wittgenstein: The Duty of Genius*. London: Vintage.

Quinton, A. (1982). *Thoughts and Thinkers*. London: Duckworth.

Ravetz, J. K. (1971). *Scientific Knowledge and Its Social Problems*. New York: Oxford University Press.

Solomon, R. C. and Higgins, K. M. (1996). *A Short History of Philosophy*. sec. IV, pp 243 305. New York: Oxford University Press.

Wisdom, J. O. (1975). *Philosophy and Its Place in our Culture*. New York: Gordon and Breach.

Wittgenstein, L. (1974). *Tractatus Logico Philosophicus*. London: Routledge and Kegan Paul.

Wittgenstein, L. (1953). *Philosophical Investigations*, Anscombe, G. E. M. (trans.). Oxford: Basil Blackwell.

Possibilism

V. Berdoulay, CNRS and Université de Pau, Pau, France

Glossary

Determinism Broadly speaking, the view that all events and actions result from knowable laws and processes.

Environmental Determinism The view that culture or human action is determined by the characteristics of the environment.

Genre de vie A relatively stable way of life which relates environmental and cultural resources.

Neo-Kantian Conventionalism A philosophy of science which insists on the conventional aspect of scientific statements.

Possibilism A theoretical approach to the study of the human relationships to the environment whereby human initiative is recognized as significant.

Introduction

As an explicitly formulated theoretical approach, possibilism was first forged by French geographer, Paul Vidal de la Blache at the end of the nineteenth and early twentieth century. However, as it encompasses, or cuts across, many of the various ideas which have been developed about culture/nature relationships throughout history, the term is often used in a loose sense, that of an approach which sets forth human freedom of action vis à vis the environment. Although too loose a definition, it refers to a rich repository of ideas which extends back to ancient times. The issue of environmental influences on culture was already interestingly formulated in the Hippocratic corpus and it continuously inspired students of history and places. As shown by C. J. Glacken in his 1967 masterpiece, *Traces on the Rhodian Shore*, three major themes have been used throughout the history of Western thought on culture/environment relationships. The first one is based on the idea of a designed Earth, whether for human purpose or not; interestingly enough, determinism of human action is not always implied by this theme, because the issue of the human responsibility for caring or managing the Earth is often invoked as a consequence of living on a well organized Earth the functioning of which should not be disturbed. The second widespread theme is that of environmental determinism on human life, and it is more precisely the theme which possibilism takes an issue with. It also builds on the third theme, that of the human transformation of the Earth, taking into account the ancient, but growing, impact of culture on its aspects and functioning. In other words, environmental determinism has rarely been considered as the sole explanation of human differences on the surface of the Earth, even among authors who are cited for having made significant use in their writings of this type of argument (e.g., Hippocrates, Herodotus, Machiavel, Botero, Bodin, Montesquieu, Ratzel, Semple, Huntington, Toynbee, etc.). There were always arguments in favor of other explanatory determinisms, whether spiritual, social, or economic.

Even though the attitude which demonstrated reluctance for giving primacy to environmental determinism had always been widespread, it was rarely formulated within a theoretical approach. However, this issue strongly came to the forefront of scientific thought in the nineteenth century, when the various strands of evolutionism attracted attention toward the environment as a key factor in the becoming of species and cultures. If the environment is the motor of historical and social evolution, can we establish a geographic science outside of an environmental deterministic approach? This is why possibilism as conceived by Vidal de la Blache took on such importance in the history of geographic thought and must be focused upon first, in order to give perspective to subsequent interpretations by historians and geographers and to contemporary attempts to address similar issues. What is most interesting with Vidalian possibilism is the grounding of the culture/environment debate on firm epistemological foundations.

Vidalian Possibilism: Its Neo-Kantian Conventionalist Foundations

Paul Vidal de la Blache (1845–1918) was a prominent figure of French intellectual life at the end of the nineteenth and beginning of the twentieth century. While founding what came to be known as the French school of geography and institutionalizing this discipline in the university system, he was influential in societal and political issues (such as education, colonization, regionalization, city planning, World War I peace negotiations, etc.). His key positions held first at the Ecole normale supérieure and then at the Sorbonne, as well as his numerous journeys abroad, especially in Germany, placed him at the crossroads of many philosophical and scientific currents. Those provided many of the elements of the possibilism which he put forth, and they facilitated his formulation of an approach which had the originality of combining up to date scientific breakthroughs with a well informed philosophy of science.

The importance of epistemological issues for studying culture/environment relationships is well illustrated by the deterministic trap set by a rhetoric of the scientific law. It has frequently been observed that environmental determinism is present in texts by authors who definitely do not have it as their general philosophy of culture/nature relationships. This is a methodologically induced determinism, resting on a positivistic conception of scientific laws, that is, universal and necessary relations which explain cause and effect among phenomena and allow prediction; a phenomenon is thus explained when it can be conceptualized as an instance of a more general, lawful relationship. In this positivistic perspective, an observed influence of the environment on culture is methodologically treated as significant insofar as its explanation can be couched within a lawful statement. Thus, the positivistic concept of a scientific law as a universal, necessary relationship contributes to environ mental deterministic discourse in geography. It also induces a fragmentation of the field into restricted, sep arated branches, where it is easier to use or find laws bearing on the studied phenomenon. The diverging of physical geography (resting on natural laws) and human geography (resting on economic or other laws) can often be related to this rhetoric of laws. This positivistic view of science could not give epistemological foun dations to Vidalian possibilism.

In the 1870s, when historian trained Vidal de la Blache was beginning to build his conception of geog raphy, a new philosophical milieu was being brought by the 'return to Kant' which affected France (especially through the philosophical essays of Lachelier, Renouvier, Boutroux), as well as other countries, including Germany, with specific styles and concerns in each national context. Although it was primarily intended to provide modern political and moral bases for the new French republic, it brought a significant revision of the positivist philosophy of science. Its emphasis on the cognitive capacities of the human mind made neo Kantianism provide a framework in which the autonomy of the mind, and thus human freedom, could be enhanced at the same time as science would not be depreciated. This approach led to 'con ventionalism', as heralded by scientists and philosophers such as Henri Poincaré, Pierre Duhem, Paul Tannery, Edouard Le Roy in France, or Ernst Mach in Germany. For them, especially for the best known Poincaré and Duhem, scientific theories were like pigeonholes for classifying facts, for storing observed information – the main criterion for the appropriateness of a particular pigeonhole being simplicity. Building on Kant's idea that the mind imposes order on the world, the general view was that the scientist gives to his thoughts the form of a hypothesis about the reality of things, while at the same time it was accepted that there was some real, essential agreement between nature and the mind. This epi stemological orientation showed the relativity and arbi trariness of scientific theories with respect to reality; it also integrated the creative role of the scientist; con sequently, it emphasized the hypothetical and heuristic character of all scientific generalizations.

It has been shown that Vidal belonged to the same ideological position in French society as the philosophical, scientific, and political promoters of neo Kantianism, and that an intellectual affinity was uniting them. The contention that Vidalian geography was consistent with French neo Kantian conventionalism is illustrated by comparing Poincaré's definition of science given in his 1905 book on *La Valeur de la Science*: "Now what is science? (…) it is before all a classification, a manner of bringing together facts which appearances separate, though they were bound together by some natural and hidden kinship. Science, in other words, is a system of relations," with Vidal's definition of geography in his 1913 paper on the '*caractères distinctifs*' of the discipline: "What geography can bring to the common fund, in exchange for the help it receives from other sciences, is the ability not to break apart what nature has brought together, to under stand the correspondence and correlation of phenomena, be it in the all encompassing terrestrial environment or in the regional milieus where these phenomena are located."

Vidal's conventionalist attitude is further reflected in his writings. He did not claim an absolute reality for the groups of phenomena which he identified and investi gated, as he was imbued with the idea of the relativity of concepts and theories, which were viewed as heuristic devices to approach the study of relationships. For example, he took care of talking of only the 'notion', or the 'idea', of milieu. He also referred to milieu as a 'guiding principle', that is, according to the terminology of the time, a higher order convention, a convention which has not been disproved. Vidal talked also of the Darwinian 'theory' formulated by Moritz Wagner on the migrations of organisms and of Ratzel's 'theory' of the growth of States. This attitude is clear in Vidalian regional geography. The concept of region represented the reality of relations existing within a certain set of phenomena, but the region thus distinguished was not given an absolute reality. Instead, several types were distinguished according to the type of relations focused upon: the natural region, the historical region, the eco nomic region. A similar conventionalist attitude is most clearly revealed in Vidal's discussion of the principle on which to found a general geographical approach: it is what he termed the 'idea of terrestrial unity', and referred to as a 'principle' on which a certain point of view and related methods could rest. It bred the notion of 'terrestrial organism', which always remained for Vidal – he made it clear – an idea, a sort of validated working hypothesis.

Then, how was explanation achieved? How was the environment included in the search for causal relationships? Most noteworthy, if the words cause and effect are found in Vidal's writings, the expressions which appear most frequently are 'series of phenomena' and 'enchainment'. This stress on causal series and causal successions or sequences was a basic tenet of Vidalian methodology. By bringing geography into close contact with the natural sciences, Vidal hoped to contribute to the social scientific endeavor, as it was well stated in his 1904 lecture on the 'relationships between sociology and geography': "The study of human societies will certainly acquire new insights if it situates itself in the line of perspective of physical and biological phenomena, upon which and by means of which man's intelligence acts," as claimed. In short, human geography derives its explanatory power from its close association to physical geography. However, explanation of a phenomenon comes from placing it in a sequence, rather than from striving to subsume it under a universal law: In Vidal's 1905 paper on 'La conception actuelle de l'enseignement de la géographie', he made it clear that the geographer "will have given facts their explanation, or at least what one may designate by this word in the human sciences" when "he will have assigned to facts the place which belongs to them in the sequences of which they are a part," as claimed. This view presented the advantage of providing an explanation for a singular, unique phenomenon, which was often called an 'individuality'. Most of all, it allowed Vidal to dissociate the idea of law from that of causal explanation, that is, necessity from causality. Possibilism could be scientific without rejecting determinism.

However, a problem then arises. A phenomenon results from complex relationships, as underlined by Vidal in his 1913 paper 'Des caractères distinctifs de la géographie': "The eminent thinker Henri Poincaré, very attentive to things geographical, expressed himself thus in one of his last writings: 'The state of the world, and even of a very small part of the world, is something extremely complex and depends on a great number of elements'." As it is clear that the combination of all the causes which produce a certain effect can never be repeated, how are causal explanations possible? Should not some role be given to chance? In fact, the debates among French neo Kantian philosophers focussed on contingency, its definition, and its extent in nature. Suffice it to say here that contingency was introduced through the conventional, thus approximate, character of scientific laws and concepts which could not fully grasp the complex and changing character of nature. For many, like Poincaré, probability calculus offered a methodological solution: although no series of phenomena that the scientist identifies in nature is quite identical with any other series, they can be classified according to their similarity, that is, in terms of probability; consequently, some degree of determinism, consciously conceived by the mind, exists even if no cause and effect relationships are absolutely repeated. Thus probability calculus, which deals with contingency, permits the search for causal relationships, but it did not imply nor require the idea of necessity. This widely held view among conventionalists about the significance of contingency can also be found in Vidal's writings, as it permitted a reconciliation of science, with its deterministic bias, with the philosophical demands in favor of free will. Vidal's own view of milieu and of human history was replete with contingency. However, his practical methodology did not extend up to the use of probability calculus, no more than other formal mathematics. He recommended a 'descriptive method', which did not mean that one should stop short of explanation. He simply meant an empirical method with a strong narrative component, which alone could respect the variety of nature's aspects, reveal the connections among phenomena, and arrive at explanation. Obviously, with such a methodological framework, the distinction between nomothetic (systematic, law seeking) and idiographic (regional, exceptionalist) approaches was irrelevant for Vidal, whereas it was significant for those who took inspiration either from positivism or from its counter philosophies.

Neo-Lamarckian Aspects of Vidalian Possibilism

The theoretical propositions of Vidalian possibilism echoed the neo Kantian solutions given to epistemology: the emphasis on the delicate interplay between empirically given nature and the ability of the mind to conceive it inspired the view that relative freedom and creativity could be noticed in all other ways of human activities. According to Vidal's conception of science, geography selects certain phenomena out of the mass presented by nature, and it studies their interrelationships from its own disciplinary viewpoint. Similarly, Vidalian possibilism stated that human beings select uses of the environment out of the many possibilities presented by nature, that is, human activity is not bound to an optimum, determined use of the Earth. The type of use depends on culture, which is both inherited and continuously elaborated through the interaction with the environment. Thus, different uses – all of them economically valid – could be made of the same portion of the Earth's surface. In other words, according to particular culture, and associated technological knowledge, human action selects certain elements (i.e., possibilities) in nature and relates them to each other to create new forms of social, economic, or political organization to serve its needs. As Vidal put it in his *Tableau de la*

Géographie de la France, published in 1903, "a geographical individuality does not result simply from geological and climatic conditions. It is not something delivered complete from the hands of Nature. It is man who reveals a country's individuality by molding it to his own use. He establishes a connection between unrelated features, substituting for the random effects of local circumstances a systematic cooperation of forces." This human structuring of nature operates at all scales. For instance, in his 1911 important papers on 'the *genres de vie* in human geography', Vidal writes that "a field, a pasture, a plantation are typical examples of associations created to suit mankind." Other examples are regions, *pays*, countries, cities, and most importantly, 'ways of life' (*genres de vie*) which were to become one of the privileged concepts in Vidalian possibilism. Noticeably, all these notions or concepts are both constructs and mediations: they are constructed through human action from a mix of material and social elements, and they mediate the human relationship to environmentally given nature.

Vidalian possibilism did not rule out environmental determinism any more than scientific determinism. For Vidal, the search for environmental influences in human activities always remained a relevant question, however among others and within a possibilist framework which encouraged a focus on mediations rather than on direct influences. In addition, human freedom to organize the Earth's surface according to the perceived needs is bound by environmental limitations and associations which derive from the deterministic laws of nature. Furthermore, human activity had the power to cause or originate 'enchainments' which then necessarily obeyed scientific determinism, because, as Vidal wrote in his 1889 review of Ratzel's work on political geography, "[man] triumphs over nature only by means of the strategy which it imposes on him and with the weapons which it provides him." In other words, possibilism was set within an ecological viewpoint, and this was facilitated by Vidal's effort to take advantage of natural sciences. He was particularly interested in the self organizing power of 'milieu', insisting in his book *Principes de Géographie Humaine*, published posthumously in 1922, on the idea of "a composite milieu, endowed with a power capable of grouping and holding together heterogeneous beings in cohabitation and reciprocal correlation."

The originality of possibilism was that it approached the contribution of natural sciences through the lenses of neo Kantian epistemology and philosophy. The most significant outcome of this approach was the way neo Lamarckism informed Vidal's theoretical outlook on evolution. Although he refused to debate the various nineteenth century evolutionary theories, he necessarily leaned toward aspects of them when he had to borrow from natural scientific theories in order to deal with the role of milieu in the evolution of species and human activities. In fact, although the opposition of Darwinian and Lamarckian theories structured part of the debates, Vidal chose to borrow from both of them. Actually, they often were not perceived as contradictory and, above all, it was much more neo Lamarckism which inspired the framework that plant ecologists were elaborating for this new science. Vidal took inspiration from it in order to give scientific legitimacy both to his overall ecological view of human geography and to his emphasis on human initiative in relation to the environment. He took from Darwin the idea of a milieu which is complex, unstable, often fragile, and always subjected to chance. However, he demonstrated a definite reluctance to make use of the mechanistic concept of environmental selection (survival of the fittest). In line with the neo Lamarckians, he briefly mentioned the struggle for existence but definitely preferred to focus on adaptive processes. For them, natural selection – to the extent that it played a role – explained the elimination of species but not their creation. Vidal stressed the fundamental idea of neo Lamarckism that adaptation is a key process which is not mechanical but rather depends on the initiative taken by the organism. Through this active view of adaptation, neo Lamarckism reinforced the neo Kantian emphasis of the autonomy of the subject.

Thus moral neo Kantian considerations could find an echo in the theoretical framework fit for neo Lamarckism. It insisted on effort, and therefore will, which could be determinant in the interaction with the environment. The primary issue was not natural selection, but rather adaptation. Furthermore, this active adaptation process usually relies on cooperation in order for a group to meet its needs. Again, neo Lamarckism echoed neo Kantian moral considerations on society, as the latter promoted the ideology of solidarity as the major social, economic, and political process to be reckoned with. Moreover, adaptation is a reciprocal process: a species or a culture makes efforts in order to adapt to its environment, but, at the same time, the former transforms the latter. These are major tenets of possibilism: reciprocal interaction, the construction of new mediations, and their moral dimension. It is both interactionist and constructivist in its conception and outlook.

Neo Lamarckians tried to offer new perspectives on evolution by stressing the role of habits of life in causing new body structures to emerge. Although this aspect of neo Lamarckism was submitted to severe and growing scientific attacks at the end of the nineteenth century, the very idea – whether proved or not – that acquired characteristics could be inherited just reinforced already existing and widely held opinion on the force of tradition in human history. In Vidal's view, this mix of philosophical and neo Lamarckian views converged toward the 'force of habit' which presides over the relative stability of 'ways of life', once they have emerged from the adaptive process to new environmental conditions. Nevertheless, the most significant aspect of possibilism is much more related to human initiative than to a static view of human culture. This is because both neo Lamarckism and neo Kantianism converged in the effort for taking into account the apparent finality which is at work in adaptation. However, this is a finality which is, in Kant's terms, internal to the beings considered. Thus, in *Principes de Géographie Humaine*, Vidal spoke of 'intention' in order to show mankind's general will to transform nature in certain directions, as "one may observe, despite the variety of materials provided by nature, a resemblance among the processes of transfor mation employed … This reveals that it is intention which presides over the transformation of matter, and that it is through his inventiveness that man makes his mark;" and he adds, in a very neo Kantian way: "The unity of the human spirit is sufficient for its manifest ations to be nearly similar." Vidal stressed, by associating them, the notions of intention, initiative, will, and artistic sense, showing that the human transformation of nature is analogous to "what sustains the artist in his struggle against matter, in his effort to communicate to it his inner feeling." (*ibid.*) Vidal noted that art was an example of intentionality in the human interaction with the en vironment. The idea of a significant role devoted to human initiative in the evolutionary process became in creasingly important among followers of Vidal, as views on 'creative evolution' were being developed in the early twentieth century by Bergson and various thinkers.

In other words, neo Lamarckism served the purpose of giving scientific legitimacy to the possibilist theoretical approach formulated by Vidal de la Blache. It allowed him to stress the role of milieu, *genres de vie*, and all other mediations between nature and culture. At the same time, making use of neo Kantian epistemology and moral philosophy, he could acknowledge the role of contingency, while enhancing free will, initiative, and cooperation. Clearly, Vidalian possibilism was not anti deterministic; rather, it provided a balanced theoretical approach to the study of the nature–culture relationships. Paradoxically, possibilism has often become associated with human freedom of action vis à vis the environment and sometimes quite synonymous with antideterminism. How did such a semantic diversion in favor of a radical possibilism occur? There were several sources for this paradoxical shift.

Historians of the Annales and the Great Diversion

An important element in the becoming of possibilism was its promotion in intellectual and academic life by French historians. The impact of this interest is paradoxical, in that it gave great diffusion and respect ability to the approach among the human and social scientists at the same time as it restricted its significance for the development of geography.

Possibilism was heralded by Lucien Febvre (1878–1956) in his book on *La Terre et l'Évolution Humaine. Introduction Géographique à l'Histoire*, which contributed foundations of what came to be known as the Annales school of history, and it set the types of relations that this discipline planned to establish with other human and social sciences. Febvre's book was a plea for Vidalian possibilism as the best way for approaching human interactions with the environment, an approach fit to define the contribution of geographers in relation to other human and social sciences. His enthusiastic glori fication of Vidal's contribution had a two fold con sequence. On the one hand, it drew historians closer to the contribution of Vidalian geographers, and later on, favored exchanges among them. However, on the other hand, it framed geography in a reduced intellectual role vis à vis the study of society: this is because the role of geography is viewed as that of identifying the deterministic influences of the physical environment when they are important, that is, mostly at the beginning of the historical processes responsible for the various forms of society. This recurrent difficulty in defining the relative role of geography and history was detri mental to the development of possibilism by geographers.

Vidal had been careful in avoiding this pitfall, espe cially as he was a historian by training and made sure to develop rich intellectual relationships with them. When he was asked to write a *Tableau de la Géographie de la France* as the first volume of historian Ernest Lavisse's *Histoire de France*, he was trekking in the tradition that many nine teenth century historians had to clarify the stage on which history develops. He especially had a famous precursor, Jules Michelet, who wrote a well known *Tableau de la France* in 1833 as part of his *Histoire de France*. He was convinced of the importance of the 'geographical factor' in the historical evolution of France, that is, in the early Middle Ages, when the various 'races' which had migrated into former Gaule had intermixed and settled permanently. This is why the *Tableau de la France* con stituted only the second volume of the *Histoire de France*, the first volume stressing the early ethnic history. Finally, starting with the late Middle Ages, institutions came to the fore as a determining factor. Thus Michelet's idea was that ethnic and environmental determinisms were important only in the early evolution of a country. Then, freedom (liberty), which manifests itself in laws and social institutions, overcomes fatality: "Freedom is strong in civilized ages, nature in barbarous times, when local determinism is all powerful, and mere geography is history." Even if possibilism did not exclude the search for environmental influences which were thought to be more perceptible in isolated, primitive groups, it could not follow Michelet on such a deterministic slope for characterizing the task of geographical research. Michelet's enthusiastic emphasis on freedom had led him into radical possibilism, whereas Vidal's acute interest for the then developing natural sciences made him gain a sense for the complexity of environmental conditions.

Febvre was well aware of this, and his direct knowledge of Vidal's teachings made him want to develop a new conception of history, whereby this discipline should also adopt a more complex view of society (including environmental factors) than focus on mere political facts. Vidal's broad outlook on society, with all its material and spiritual complexity, was more fundamental in the emergence of the Annales school than the possibilism which the latter celebrated. However, in the historians' effort to transform their discipline into a full human and social science, the issue for them was again to characterize the role of environmental conditions. Febvre, like Vidal and others, was then attentive to the relative role of determinism and chance in human history. They all demonstrated interest for the attempts to give theoretical and methodological significance to contingency and probability. Picking up on the contribution of Henri Berr, one of the historians who gave him inspiration, Febvre insisted on developing a broad view of contingency in order to give Vidalian possibilism more significance in this respect. Berr distinguished clearly chance from another form of contingency – 'individuality' – which was viewed as the result of a fortuitous combination of conditions, that is, a relatively stable outcome of the intersection of independent causal series. His discussion was especially relevant for geographical issues, and he gave as examples of individualities the '*pays*' and the 'region': they result from the intersection of causal series of the natural and human realms – the *pays* being identified by the inhabitants themselves, the region by the geographer.

Especially through the concept of individuality, the theory of contingency became a way to capture the idea of creative evolution scientifically. In this perspective, another example of individuality was stressed upon by Febvre: the '*genre de vie*'. For him, this conceptualization was a way to integrate human creative evolution into a scientific framework. He suggested just this when he stated that human creativity can be given a place in science in the same way that vitalists gave place in biology to 'spontaneity', the "faculty to create something new." Refering to Berr, as well as to Bergson (known for his treaty on *L'Évolution Créatrice* published in 1907), Febvre tried to show that this place for creativity can be handled as a contingent fact, that is, in terms of probabilities, and he related it to the biological theory of preadaptation: "Under the cover of this theory, there reappeared in scientific studies a vitalism which revived, to the great scandal of a number of scientists brought up in the blind and uncompromising mechanistic faith, the fruitful and by no means unscientific idea of chance." Thus, for Febvre, "was not the theory of ways of life, in geography, as Vidal de la Blache formulated it, also a fitting translation of intellectual needs of just the same, or at any rate of a similar character, whether its author was fully aware of it or not?"

In addition to the epistemological extension given by Febvre to possibilism, his illustration of it as encapsulating what is significant in the geographers' contribution to knowledge had several consequences detrimental to geography in the long run. Clearly, geographers first benefited from the prestige of an alliance with leading historians of the future Annales school; they particularly avoided the strains of fighting for their contribution against the sustained attacks made by the newer and ambitious school of sociologists founded by Durkheim, since most of the violent critiques written by Febvre in his book were primarily addressed to them. However, the quality and polemical force of Febvre's arguments weighed on the geographers' conception of their own discipline. As they worked on themes which were not fully recognized in Febvre's defense of possibilism, their contributions became marginalized, somewhat at odds with what had given them respectability and good relations with historians. As far as the presentation of possibilism was concerned, the emphasis placed by Febvre on contingency and detailed regional monographies gave it significance within an idiographic approach: a search for general processes and statements on the human relationships to the environment was illusory, that is geography gets scientific significance only from local, detailed monographies. In addition, Febvre's violent rebuttal of the Durkheimians made worse the relative scientific distance between geography and sociology. Needless to say, the Vidalian interest for environmental influences disappeared from Febvre's presentation of possibilism. Furthermore, in spite of the philosophical extension it received from Febvre, possibilism saw its scope reduced to certain moments in history. The historians' tendency was again to consider that the environment was significant for the establishment of human arrangements or settlements; when these are established, other processes – be they economic or social – become dominant. In other words, this view of possibilism tended to restrict environmental considerations – and thus the geographer's contribution – to initial conditions. The restricted scope left to geography was thus reinforced, and the geographers' early concern for material, daily life, or for cognitive representations and 'mentalities' was easily taken over by historians of the Annales.

An analogous critique can be made about Braudel's glorification of Vidal's geographical contribution to knowledge. Fernand Braudel was a leading figure during post World War II in history and human and social sciences. His masterpieces on the sixteenth century Mediterranean world, on the development of a world economy, or on the identity of France, all concur to demonstrate his appreciation of the significance of geography for history. However, by only glorifying the regional possibilist geography of Vidal and his early disciples, he made it clear that his contemporaries did not write significant contributions. At any rate, in Braudel's work, geography is confined to an ancillary role in relation to a more encompassing discipline, which is history. And again, this authoritative and prestigious positive discourse in favor of geography actually played against the development of a renewed geographic con tribution. Possibilism was trapped in the double bind set by historians to geographers: their contribution is highly valuable as far as they restrict it to partial, initial con ditions of phenomena, that is, as far as they focus on … environmental deterministic relations.

A Diverse Heritage Among Geographers

While historian, Febvre was heralding possibilism at the cost of making it evolve far from its Vidalian original concerns, geographer, Jean Brunhes (1869–1930) was the disciple of Vidal who most developed the possibilist ap proach. Although his position at the Collège de France where a new chair of human geography had been created for him gave a wide audience to his views, it did not favor the continuation of his research themes because the Collège was not a doctorate granting institution (as was the university). Pierre Deffontaines was among the few geographers who best carried on some of the many themes opened by Brunhes.

In contradistinction with other Vidalian geographers indulging in Febvre's interpretations, Brunhes delib erately chose not to privilege regional studies. For him, the possibilist approach could best be cultivated through theme oriented research, these themes being provided by the various terrestrial forms of interaction between culture and nature. Starting from the study of these 'geographical facts', one could arrive at more general statements about the humanistic significance of the human transformation of the Earth. This humanistic concern ranged from applied, practical recommendations to quasi metaphysical meditation on the becoming of the human being on Earth. Among the best known concerns that he developed from a geographical point of view was the issue of the increasing realm of thought on the surface of the Earth, which Deffontaines referred to by using the terminology created by Vernadsky and promoted by Teilhard de Chardin for dealing with the emergence of this 'thinking layer' on the globe: a 'noo sphere'. However, his observation of humanly operated environmental destructions and his support for the conservation pleas in his time brought a worried, if not dark, note to this view and made him much less optimistic than the other Vidalians about human eco logical future.

A key aspect of possibilism as promoted by Brunhes was the importance of the 'idea' that each people held about the world or elements of it – ideas which enhance the human autonomy vis à vis the environment. In his 1913 paper *Du caractère propre et du caractère complexe des faits de géographie humaine*, Brunhes claimed that "Man carries in his mind's eye a certain representation of the universe which depends partly on himself." On this representation depends the human selective use of the Earth's possibilities. This 'psychological factor' (to use Brunhes' term from *La Géographie Humaine*, first published in 1910) ties and mediates human beings and environ ment. This is why Brunhes, clearly expressing a Vidalian position, rejected the division of human geography into 'passive or static' and 'active or dynamic' branches, whereby the former would study the 'action of nature on man' and the latter the 'reaction or action of man on nature'. He argued that human geographical phenomena result from a constant interplay between nature and culture, and, consequently, that possibilism should nurture a holistic approach. The important role given to representations was to be seen in this perspective. Although the concern for them already existed among Vidalians and others, Brunhes' clear cut contribution on this set possibilism in the incipient cognitive tradition of geography.

His fundamental concept of the 'psychological factor' had a very Bergsonian flavor. His colleague at the Collège de France, philosopher, Henri Bergson, held the view of a creative evolution responsible for a state of perpetual change, adjustment, and redefinition, which was similar to Brunhes' 'principle of activity'. As he put it, "the human psychological factor, which is thus at the origin of geographic fact, the inevitable intermediary between nature and man, might be called, according to an expression dear to Henri Bergson, 《the direction of the attention》." (*ibid.*) Echoes of this Bergsonian concern for creativity and flux can also be found in the writings of a later member of the French school, Jean Gottmann (1915–94), who was attentive to the emergence of new forms of settlements and of social, economic, and polit ical spatial organization. However, there was a diffuse and diluted form of Bergsonism which seemed to have spread among most French geographers who did not get their training at the time of Vidal. They overemphasized contingency, shying away from strong causal schemes. Worse, many of them misinterpreted Bergson's emphasis on the importance of intuition for knowledge; they took it as the legitimation of their relative disdain for formal, analytical, or mathematical analyses. Added to the biases introduced by Febvre, this tendency encouraged possibilism to slide away from the strong framework provided by Vidal and Brunhes toward idiographic case studies.

Basically, they juxtaposed a vague rhetorical rejection of environmental determinism and a nonsystematic search for diverse causal connections between the local environment and social characteristics. Obviously, ex ceptions existed in relation to this broad trend which started occurring in the inter World Wars period and became dominant after. For instance, Max. Sorre (1880–1962) implemented part of the research program set by Vidalian possibilism when he focussed on the biological aspects of the human relation to the environment, giving geography its full ecological significance, and establishing new thematic orientations such as medical geography. Other examples include Raoul Blanchard (1877–1965) who tried to apply the possibilist approach to the study of cities, albeit with mixed results in this respect, or Maurice Le Lannou (1906–92) who criticized most studies of *genres de vie* for neglecting their social dimen sion. Mention should also be made of Pierre Gourou (1900–99) who progressively changed his explanatory emphasis from environmental constraints to sociocultural organization. Nevertheless, the dominant Febvrian in terpretation of possibilism stopped geographers short from structuring their argument within a theoretical argument; they were left studying local correlations be tween environmental and social characteristics, with the risk of inducing some deterministic outlook on society. This was the case with tropical countries where too many geographers who correlated climatic conditions with underdevelopement unwittingly slid into environmental determinism. It clearly begged for reactions, the most strident coming from Marxists geographers, but leading to other (economic and political) determinisms. By the 1970s, Vidalian possibilism had thus progressively dis appeared from most French research, as a result of its interpretation as a weak theory.

In other countries, the possibilist orientation inspired many works and nurtured numerous debates. Sometimes, there were strong convergences and collaborations be tween Vidal himself and a few of his contemporaries. For instance, the Russian geographer, Alexander Voiekov showed close to Vidal's conceptions, contributing in 1901 a well documented article for the French school's periodical, the *Annales de Géographie*, in which he focussed on the force of the human impact on the Earth through his power over movable bodies. Cast in more general terms, the idea of the human population as an important transformer of nature laid a receptive ground for Vidalian possibilism. Another source of this receptivity was the felt necessity to combat the strength of environmental deterministic biases in geographical practice. This was often the case in Anglo American geography, and for a long time (actually up to the 1960s) there were sustained debates on the determinism/possibilism theme. However, the tendency to frame the issue as an 'either/or' option made those debates miss the Vidalian approach which was neither derteministic nor radically possibilist.

If at times the Vidalian inspiration was direct, espe cially in countries which modeled their institutional ization of geography on the French innovations, in many cases it was intermingled with the widespread neo Kantianism which pervaded philosophical and social scientific currents of the beginning of the twentieth century. A case in point comes from English Canada, with the influential work of Harold Innis on the eco nomic formation of his country. But it certainly is the Berkeley (or Carl Sauer's) school which best extended abroad some implications of the possibilist theoretical approach. This was achieved through the full recognition that the study of culture is the key to understand the complex and multifaceted results of the human inter action with the environment. The cultural geography which it inspired gave possibilism its full cognitive dimension, opening it to other research along this line and to new themes relating to environmental perception and behavior.

Thus, possibilism has enticed a whole range of research themes, but, somehow, most of the philosophical and theoretical Vidalian kernel has long ceased to exist. However, its central issue – determinism vis à vis relative human freedom – has come back to the fore and called for a return to the original possibilist discussion.

Now that neo Kantian and classical contributions to the history of geographical thought are less depreciated than before, and that the concern for place is no more rejected as antiscientific even by its formerly most out spoken enemies, the epistemological and theoretical positions entailed by Vidalian possibilism can be revisited in light of the recent attempts at encompassing the relative roles of voluntarism, contingency, and deter minism in human affairs. It is noteworthy that key issues on the current geographical agenda include the relationship between different forms of explanation (e.g., neopositivistic and narrative), the reemerging interests in vitalism and what the life sciences might contribute back to human geography, the relationships between 'agency and structure' or 'individual and soci ety', and the pressing challenge of sustainability. These issues are all in effect returning to concerns first articu lated by Vidal and coded into his vision for a possibilist human geography.

See also: Determinism/Environmental Determinism; Exploration; Lamarck(ian)ism; Philosophy and Human Geography; Probabilism; Structurationist Geography; Vidal de la Blache, P.

Further Reading

Berdoulay, V. (1988). *Des mots et des lieux. La dynamique du discours géographie*. Paris: Ed. CNRS.

Berdoulay, V. (2008). *La formation de l'école française de géographie (1870 1914)* (3rd edn.). Paris: Ed. CTHS.

Berdoulay, V. and Soubeyran, O. (1991). Lamarck, Darwin et Vidal: Aux fondements naturalistes de la géographie humaine. *Annales de Géographie* 561/562, 617 634.

Brunhes, J. (1910). *La géographie humaine* (transl. *Human Geography*, 1920). Paris: Alcan. Chicago, IL: Rand McNally.

Febvre, L. (1922). *La Terre et l'évolution humaine* (transl. A Geographical Introduction to History, 1925). Paris: La Renaissance du Livre. New York: Knopf.

Gregory, D. (1981). Human agency and human geography. *Transactions of the Institue of British Geographers, N.S* 6, 1 18.

Sack, R. D. (1997). *Homo Geographicus: A Framework for Action, Awareness, and Moral Concern*. Baltimore, MD: Johns Hopkins University Press.

Sauer, C. O. (1967). *Land and Life*. Berkeley, CA: University of California Press.

Sorre, M. (1943 1952). *Les fondements de la géographie humaine*, 3 vols. Paris: Armand Colin.

Soubeyran, O. (1997). *Imaginaire, Science et Discipline*. Paris: L'Harmattan.

Tatham, G. (1951). Environmentalism and possibilism. In: Taylor, G. (ed.) *Geography in the Twentieth Century*, pp 128 162. New York: Methuen.

Vidal de la Blache, P. (1902). Les conditions géographiques de faits sociaux. *Annales de Géographie* 11, 13 23.

Vidal de la Blache, P. (1903). La géographie humaine, ses rapports avec la géographie de la vie. *Revue de Synthèse Historique* 7, 219 240.

Vidal de la Blache, P. (1911). Les genres de vie dans la géographie humaine. *Annales de Géographie* 20, 193 212 and 289 304.

Vidal de la Blache, P. (1922). *Principes de géographie humaine* (transl. *Principles of Human Geography*, 1926, New York: Holt). Paris: Armand Colin.

Postcolonial Cities

A. D. King, State University of New York Binghamton, Binghamton, NY, USA

Definitions

In everyday usage, postcolonial cities refer to those cities (frequently capitals) in what were previously colonial societies. In addition to this temporal or historical use of the term, however, postcolonial may also imply a particular critique, one which not only emphasizes the distinctive impact which colonialism has had on the economy, society, culture, spatial form, and architecture of the city but also on the way the city itself is understood and represented. Thus, some postcolonial critics would reject or deny represen tations of cities as postcolonial, arguing that this privileges a particular (usually 'Western') interpretation of the city at the expense of a more indigenous one. As the terms im perialism and colonialism are often used interchangeably, in what follows, imperialism refers to the imposition of the power of one state over the territories of another, frequently by military force. Where imperialism originates in the metropole, what happens in the colonies resulting from economic, political, and cultural control and domination is colonialism, even though this can take various forms. While the term postcolonial can be used to apply to all historical and geographical instances of colonialism, whether of the Romans in Britain, the Japanese in Korea, or the English in Ireland, this article focuses on European industrial colonialism and its aftermath in the non European world.

In a less frequent usage, the term postcolonial cities refers to those capitals or major cities of the one time imperial metropoles, such as Paris, London, Birmingham, or Amsterdam, not only because of the large postcolonial populations they have attracted following the formal end of empire and independence of what were previously col onies, but also because for migrants from these ex colonies, for example, Paris or London is a postcolonial city. The term can also be used more metaphorically to refer, for example, to cities in Eastern Europe, previously under the hegemonic control of the Soviet Union. Irrespective of the different uses of the term, understanding the meaning of the postcolonial city presupposes familiarity with the inextricably connected concept of the colonial city.

Origins of Interest in Postcolonial Cities in Human Geography and the Social Sciences

In the 30 years following World War II, over 70 states, previously European colonies, gained their political if not economic independence. From the mid 1950s, a number of mainly North American and European social scientists (geographers, sociologists, and anthropologists), studying aspects of urbanization in Asia, Africa, and the Middle East, began to identify distinctive social, spatial, and cultural characteristics of towns and cities in recently decolonized countries where European traders, settlers, and colonial officials had lived (or were still living) and whose urban development they had largely controlled. With the departure of the former colonial power, these colonial towns and cities had now been inherited by the newly liberated subjects of the independent nation. What these studies described were dual cities, ethnically, socially, and spatially segregated between the 'European' town and 'native' settlement; the first, modern, spacious, low density, well maintained through the use of town planning, and culturally different from the surrounding environment; the second, usually separated from the first by parks, railway lines, or open space, invariably more densely settled, with traditional housing, social and cultural buildings overcrowded, and lacking in services and infrastructural provision. A powerful condemnation of this phenomenon from French Algeria that simul taneously conveys the sense of oppression and injustice experienced by members of the colonized society is that of the Martiniquan nationalist writer, Frantz Fanon, in his oft cited critique of colonialism, *Les Damné de la Terre*:

> The colonial world is a world divided into compartments … of native quarters and European quarters, of schools for natives and schools for Europeans; in the same way we need not recall apartheid in South Africa … The colonial world is a world cut in two. The dividing line, the frontiers, are shown by barracks and police stations. In the colonies it is the policeman and the soldier who are the official, instituted go betweens, the spokesman of the settler and his rule of oppression … The settler's town is a strongly built town, all made of stone and steel … The town belonging to the colonized people, or at least the native town, the Negro village, the medina, the reservation, is a place of ill fame … It is a world without spaciousness, men live on top of each other … The native town is a crouching village, a town on its knees, a town wallowing in the mire. (Frantz Fanon, 1961: 37 39)

While this dual cities phenomenon, though varying greatly between different places, was widely recognized, it is evident from the mid 1960s' comments of American sociologist, Janet Abu Lughod, that, despite the occasional observations mentioned above, the subject had not yet attracted serious theoretical or policy related attention, either from local or foreign observers:

> The major metropolis in almost every newly industrialising country is not a single unified city, but, in fact, two quite different cities, physically juxtaposed but architecturally and socially distinct … These dual cities have usually been a legacy from the colonial past. It is remarkable that such a phenomenon has remained almost unstudied. (Janet Abu Lughod, 1965: 420)

The years which followed witnessed a number of serious attempts to remedy this situation. The interest in human geography and other disciplines came from two sources. One was the growing dissatisfaction among geographers, sociologists, and others, increasingly familiar with cities outside the West, of the inadequate state of comparative urban theory at this time. This was seen as largely based on Western industrial experience with theories of urban and city growth dominated by the work of the Chicago School. For these scholars, serious thinking about the concept of the colonial city and theories of colonial urban development were to take place between the late 1960s and mid 1970s.

The second impetus came from urban professionals concerned with policy related issues of urban development planning following the end of colonial rule. It was in these two, often interrelated, contexts that a more theoretically informed understanding of the notion of both the colonial and postcolonial city developed. It was an understanding which recognized that, despite political independence, the persistence of colonial structures – represented among others by the grossly unequal social and spatial divisions of major cities – was an affront to the democratic aspirations of newly independent nations. This 'colonial space' of the European part of the city and its distance from the native settlement was not just based on the standards and specifications of the so called modern city in the West; in many cases, in order to satisfy the political and cultural demands of the colonial elite, it also represented a gross inflation of these standards. In many postcolonial cities, decades after independence, while the native area of the city crumbled under pressure from rapid migration, indigenous elites occupied the spaces vacated by their foreign predecessors.

In addition to these real spatial inequalities, however, the newly independent nation was faced with additional symbolic contradictions in the city. In different colonial towns and capital cities, the architecture and urban design of the imperial power, whether modest or grandiose, had been consciously conceived to convey cultural as well as political authority. For the political elite of the new postcolonial nation, removing monuments and replacing street names with those of their own folk heroes was relatively simple; however, endowing the city with a totally new national identity, one that draws on its own vernacular cultures, representational spaces, and modes of signification, was not only a much longer term project but also, because of competing regional, ethnic, linguistic, and other tensions in newly established nation states, a deeply contentious one. Moreover, more immediate priorities had to be addressed, such as political stability, land redistribution, economic development, or providing shelter for the burgeoning population in squatter settlements. In this context, transforming the image of the city, projecting a new identity, had usually taken second place on a new nation's agenda. In India, for example, it took almost 40 years after independence for the names of India's one time colonial port cities, such as Bombay, Madras, Calcutta, to be vernacularized as Mumbai, Chennai, and Kolkatta, respectively.

Representing the City

As with the term postcolonial itself, it can be argued that the category of the postcolonial city is more of an outsider's than insider's label. It privileges a representation of the city which foregrounds its colonial past, rather than the city's present or future. In this sense, postcolonial can carry the meaning of the failure of decolonization. Is the once colonial city destined to be forever postcolonial or can it, like Hong Kong or Singapore, gradually metamorphose into a world or global city? The mutability of categories applies equally to other terms used over the last 50 years with reference to other cities in Asia, Africa, or Latin America. Where world space is couched within a developmentalist framework, rather than one of colonialism and postcolonialism, the same cities are referred to as 'Third World Cities', a phrase coming into use following the invention of the 'Three Worlds' paradigm in the early 1950s. In this conceptual framework, they are represented not with regard to issues of cultural identity, heritage, and questions of representation, but rather in relation to inadequate water and shelter provision, unemployment, and overburdened infrastructure. As all such representational categories emanate from a globally dominant anglophone positionality and a discourse originating in the West, it follows that they are either contested or rarely used by the inhabitants of the city itself, whether urban professionals or local residents. However, since the late twentieth century, new research and scholarship on the colonial and postcolonial city by scholars native to the city has resulted in original and interesting perspectives on both.

Revisiting the Colonial and Postcolonial City

While the earlier studies of colonial and postcolonial cities (principally in English, though also in French and Spanish) produced during the first three or four decades following the end of colonial rule might be described as the first critical accounts of such cities, they have generally located the topic within a European or Euro American frame of reference. Moreover, as they were generally based on European archives and language sources, the agency for the cities' development has frequently been primarily accorded to their European inhabitants, with less attention paid to that of the in digenous population. Since the late twentieth century, however, studies by scholars native to the city or country, drawing on previously unused sources, often in local languages, have begun to transform this perspective. Accounts of the development of Singapore, for instance, based on a detailed study of local archives, have chal lenged the notion of the colonial city as simply the outcome of dominant (Western) forces. They represent the city as a product of indigenous, local agency where the colonized have maintained substantial control over their lives.

Other recent studies of the colonial city have chal lenged one of the dominant themes of these earlier Euro American studies, namely, the ethnic, racial, social, and spatial divide between the indigenous city and the European colonial settlement. Detailed historical studies of Delhi, for example, drawing on a variety of different local archives, have questioned the use of dualistic categories of colonizer/colonized, traditional/modern, European/native, old/new, and shown that social and spatial divisions in the city were not nearly so clear cut as has been represented. Instead, there were charged interconnections between the two spaces. The new co lonial settlement offered opportunities for some Indian residents to move between cultures and spaces, con structing new identities, identifying with the new by rejecting the old, and creating indigenous modernities. A similar study of Calcutta, making use of local lan guages, challenges an oft repeated colonial represen tation of the city as a space marked by a Manichean division between the 'European city' and the 'black town'. While such labels certainly exist on European maps and in the anglophone colonial records, helping to confirm European perceptions of the city, they (and similar terms) are absent in local Bengali accounts of the city, and hence, do not inhibit the movement or activities of local inhabitants. For them, the meaning of the city is not constructed through a perception of colonial spatial divisions but rather as a cradle of Bengal nationalism.

These recent accounts, postcolonial in both senses of the term, show that the space of 'the colonial city' is both ambiguous and contested. Most importantly, attitudes have changed over time. The same is true of the post colonial city, a concept not easy to define. What has happened in the decades since it first came into use is that its meaning has fluctuated, not least because the conditions in which the postcolonial city exists have also changed. This is most clearly seen in relation to the postcolonial Indonesian capital of Jakarta. Here, fol lowing the end of Dutch colonial rule in the mid twen tieth century, social and political oppression not only returned during the three decades (1966–98) of President Suharto's dictatorial 'New Order' regime, but also their severity and scope increased. In this context, the urban, housing, and architectural developments undertaken by the Dutch have, by comparison, been seen by an Indo nesian scholar as relatively enlightened, a 'colonial gift' inherited by the postcolonial society and state. Clearly, however, not all postcolonial accounts of colonial cities are so benign.

In this context, how is the postcolonial city recog nized? This is not easily answered as cities are not necessarily postcolonial in the same way. The most drastic solution to transforming the colonial city and, in the process, changing the public's sense of commitment to the nation and promoting a sense of national pride, is to shift the geographical location of the capital, not only building a new city on a new site, but also using spec tacular vernacular architecture and an indigenous name to do so, a strategy adopted by Sri Lanka in shifting its capital from Colombo to Sri Jayawardhanapura Kotte. Less dramatic is the construction of a new capitol or parliament building, a strategy adopted by Bangladesh, Papua New Guinea, and Kuwait, among others. Else where, as in India's capital city, the lavish spatial layout of the central areas of New Delhi has, over the half century since independence, been re densified by the addition of numerous government buildings. Toponymic reinscrip tion is a common way of reclaiming the cultural space of city streets. The urban and architectural attributes of the one time colonial city can also be treated as part of the city's inheritance and used for the promotion of tourism. Yet such policies are invariably contested. For some, but not for others, half a century after independ ence, the symbolic significance of colonial buildings has lost its oppressive meaning and become an apolitical part of the nation's heritage. While such a policy may be enthusiastically endorsed by UNESCO's International Committee on Monuments and Sites, it can be bitterly resisted by local groups.

In other cities, such as Brazilia, Chandigarh, or Islamabad, modernist architecture has been combined with nationalist rhetoric to create new images for the nation. From the late twentieth century, constructing the world's highest building, as in Malaysia's Kuala Lumpur or Taiwan's capital, Taipei, has become an essentially competitive strategy for postcolonial nations to 'put themselves on the map' and make claims on others' definitions of modernity. In all these strategies, the aim is to convey to the city's inhabitants a new sense of national citizenship, a new collective consciousness.

Yet, irrespective of these changes to the physical environment, inherent structures of power, inherited from the colonial regime and institutionalized in the centralizing and authoritarian practices of city and state bureaucracies leave an indelible scar on the urban landscape, not least in the post apartheid cities of South Africa.

Cities: Postcolonial or Postimperial?

So far, it has been assumed that postcolonial cities refer primarily to those cities existing in formerly colonized societies. Yet the distinction between postcolonial and postimperial cities can be ambivalent, depending on the position and location of the speaker. Labeling Paris or London as (technically) postimperial cities foregrounds their earlier imperial role without necessarily invoking imperial contexts. Yet postcolonial is increasingly used to describe them, not only because of the ethnic and racial composition of their populations but also because of their public culture and civil society, including matters of education, politics, religion, and culture, among others. As world space is simultaneously global, postcolonial, and colonial (among other categories), postcolonial histories of migration not only distinguish the population, politics, and culture of one postcolonial (or postimperial) city from another but also from other 'world' or 'global cities', such as Frankfurt, Chicago, or Zurich. For example, of the roughly one third of population of New York that is foreign born, over half are from the Caribbean and Central America, with significant proportions from Europe, South America, and South and Southeast Asia. Any explanation for the presence of a substantial portion of that population, particularly the English speaking migrants from the Caribbean, South and Southeast Asia, has to take account of colonial and postcolonial histories. This is equally the case with the Spanish speaking migrants from the Caribbean and Central and South America. In London, of the roughly one fifth of the population that is foreign born, the large majority is principally from postcolonial countries, particularly South and Southeast Asia, Ireland, East, West and South Africa, the Caribbean, North America, and Australia, as well as from continental Europe. The same is true of Paris, where about one fifth of the population is foreign born; of these, a substantial proportion are primarily from postcolonial North Africa (Algeria, Morocco, and Tunisia), or Mauritius. Of the legal immigrants in Lisbon, many are from Portuguese ex colonies such as Brazil, Angola, Cape Verde islands, and Macau.

These distinctively postcolonial migrations have major influences on the economy, society, culture, religion, politics, spatial and built environments, and also security, of the city. They bring to the global city a variety of vibrant postcolonial cosmopolitanisms specific to the city (and the state) where they exist. Numerically or economically powerful postcolonial minorities can bring their influence to bear on government policies, both domestic (in regard to immigration, education, or welfare) or foreign (in relation to foreign policy, international disputes, and disaster relief) or in regard to public culture, the arts, dress codes, and media. They shape the definition and nature of multiculturalism and create new forms of hybridity in different cultural realms, whether in regard to language, religion, literature, or the arts. Understood in this sense, the postcolonial city can be said to generate not only multiple temporalities but also multiple spatialities. These extend the real and the virtual space of the city and its inhabitants to other urban and rural locations worldwide.

The significance of these real and virtual spaces can be seen in the growing outsourcing of employment (especially in call centers) from cities in North America, Europe, and the UK to the large anglophone labor markets in South and Southeast Asian cities, an aspect of globalization that cannot be understood without reference to a postcolonial urban frame. Employment outsourcing has generated both extensive employment opportunities and also office building in major metropolitan Indian cities such as Bangalore, Hyderabad, and Delhi. In Paris, most postcolonial minorities (mainly North African) are in the *banlieues* which, from the late twentieth century, have concentrated the socially and economically marginalized in the peripheral areas of French cities. Structurally equivalent to British and American inner city areas, and often referred to as ghettos, the *banlieues* provide a space to develop a separate cultural agenda where the inhabitants reterritorialize housing projects that were previously socially anonymous. In Britain, the urban landscapes of (especially) eastern, western, and southern inner suburbs of the postcolonial global city of London have been regenerated and transformed by South Asians, a large proportion of whom, while being united by and sharing the same religion, are nonetheless divided along national and ethnic lines. In the urban enclaves and terraced housing of Britain's second major postcolonial city of Birmingham, the largest group of South Asian Muslims are from Pakistan, comprising 7% of the city's population. The east Midlands, manufacturing city of Leicester, with almost 30% of its 280 000 population from South Asia, occasionally described as 'Europe's largest Indian city' must also be the most postcolonial of this continent. In the words of Singapore geographer, Brenda Yeoh, not only are the colonial city and imperial city "umbilically connected in terms of economic link ages as well as cultural hybridization but their 'post equivalents' … need to be analyzed within a single 'postcolonial' framework of intertwining histories and relations" (Yeoh, 2001: 456–68). In this context, policies of colonial urbanism in the one time colonies can also be seen as being related to the development of racial and class divisions in major Western cities. However, it has to be underlined that individual urban authorities in different countries have their own distinctive policies, whether in regard to questions of housing, planning, education, or other spheres. In multicultural societies, it can be expected that members of particular communities, particularly if they are of recent settlement origin, stay together, with access to their own shops, social centers, and places of worship (as, for example, was also the case with British and other European and American com munities abroad).

Long after the formal end of empire, postcolonial memories and associations continue to affect the use of spaces in the one time imperial city. Debates over his toric sites in the City of London, for example, remem bered as the economic and financial center of the Empire, have been used to influence subsequent decisions about urban design. Memories of empire have been kept alive through maintaining the historical importance of par ticular places. Yet, postcolonial as well as postimperial memories can also be celebratory for the multicultural population of the city. This consciousness points to a changing, vibrant future, a new kind of intellectual milieu created by new ethnicities, hybridized identities, and diasporas that create new and distinctive cultures in each unique, geographically and culturally specific postcolonial city. Whether in the cities of the one time metropole or the one time colony, postcolonialism creates both the split, as well as the suture, between traditional and modern identities. It is inextricably linked to the creation of plural societies and transnational cultures.

The more metaphorical (rather than literal) use of the terms, colonial/postcolonial, have also been deployed to refer to urban processes in contemporary Europe. With migrant labor, legal as well as illegal, arriving from all over the world, including postcommunist Eastern Eur ope, filling the lowest paid slots in an ever increasingly globalized economy, there are new colonized populations in Western cities. In this context, features that to a greater or lesser extent distinguish postcolonial populations, such as a language shared with the host society, some famil iarity with the laws, culture, and social practices of the metropolitan society, a shared, if contested, history, the presence of long established communities, are features which, among others, differentiate postcolonial com munities and migrants from those of nonpostcolonial origin. In the context of these criteria, multicultural Berlin or Zurich differs from multicultural London or Paris.

Various scholars have shown how the postcolonial paradigm has been expanded to cover many historical and geographical instances. Though not sufficient in themselves, postcolonial histories, sociologies, and geog raphies are nonetheless key to the understanding of a plethora of issues, not just in the multicultural, post colonial, and postimperial global cities, such as Hong Kong, Toronto, or Amsterdam, but in many other cities worldwide. The growth of a postcolonial urban per spective in the last two decades is a result of globalization and migration as well as the extension of virtual worlds through the spread of information and communications technology. It is an essentially comparative, cross cultural, and cross temporal perspective. It results not only from observations of postcolonially mobile authors and their postcolonial mobile readers, but also from more sophisticated interpretations of globalization, from the viewpoint of countries and religious cultures outside the West.

The nature of what is postcolonial about the post colonial city can be radically different in different places and at different times. What is meant by the phrase must always be elicited on the basis of the specific historical, cultural, and political circumstances in which each postcolonial city exists.

See also: Colonialism I; Imperial Cities; Multicultural City; National Spatialities; Race; Urban Architecture; Urban Design; World/Global Cities.

Further Reading

Abu Lughod, J. (1965). Tale of two cities; the origins of modern Cairo. *Comparative Studies in Society and History* 7, 429 457.

Bishop, R., Phillips, J. and Yeo, W. W. (eds.) (2003). *Postcolonial Urbanism: Southeast Asian Cities and Global Processes*. London: Routledge.

Chattopadhyay, S. (2005). *Representing Calcutta: Modernity, Nationalism and the Colonial Uncanny*. London: Routledge.

Drakakis Smith, D. W. (1987). *The Third World City*. London: Methuen.

Fanon, F. (1961). Les Damnés de la Terre, pp 37 39. Paris: François Maspero (*The Wretched of the Earth* (1968). New York: Grove. Transl. Farrington, C.).

Horvath, R. V. (1969). In search of a theory of urbanization: Notes on the colonial city. *East Lakes Geographer* 5, 69 82.

Hosagrahar, J. (2005). *Indigenous Modernities: Negotiating Urban Form*. London: Routledge.

Jacobs, J. (1996). *Edge of Empire: Postcolonialism and the City*. London: Routledge.

Kincaid, A. (2006). *Postcolonial Dublin: Imperial Legacies and the Built Environment*. Minneapolis, MN: University of Minnesota Press.

King, A. D. (1976). The post colonial city. In *Colonial Urban Development: Culture, Social Power and Environment*, ch. 11, pp 282 288. London and Boston: Routledge and Kegan & Paul.

King, A. D. (1990). *Urbanism, Colonialism and the World Economy: Cultural and Spatial Foundations of the World Urban System*. London: Routledge.

King, A. D. (2004). *Spaces of Global Cultures: Architecture, Urbanism, Identity*. London: Routledge.
Kusno, A. (2000). *Behind the Postcolonial: Architecture, Urban Space and Political Cultures in Indonesia*. London: Routledge.
Legg, S. (2007). *Spaces of Colonialism: Delhi's Urban Governmentalities*. Oxford: Blackwell.
Mabogunje, A. (1968). *Urbanization in Nigeria*. London: University of London.
McGee, T. G. (1967). *The Southeast Asian City*. London: Bell.
Perera, N. (1998). *Decolonising Ceylon: Colonialism, Nationalism, and the Politics of Space in Sri Lanka*. New Delhi: Oxford University Press.
Rayfield, J. R. (1974). Theories of urbanization and the colonial city in West Africa. *Africa* 44, 163 185.
Ross, R. and Telkamp, G. J. (eds.) (1985). *Colonial Cities: Essays on Urbanism in a Colonial Context*. Dordrecht: Martinus Nijhoff.
Sidaway, J. D. (2000). Postcolonial geographies: An exploratory essay. *Progress in Human Geography* 24, 591 612.
Yeoh, B. S. A. (2001). Postcolonial cities. *Progress in Human Geography* 25(3), 456 468.

Relevant Websites

http://www.iniva.org
Paul Gilroy: London: Postcolonial City, CelloidCities at Iniva.

Postcolonialism/Postcolonial Geographies

C. McEwan, Durham University, Durham, UK

Glossary

Ambivalence The coexistence of two opposed and contradictory emotions, such as simultaneous attraction and repulsion.
Colonialism The extension of a nation-state's sovereignty over territory beyond its borders through the creation of either settler colonies or administrative dependencies, in which indigenous populations are ruled directly or displaced.
Discourse Social practices through which the world is made meaningful and intelligible, including narratives, concepts, ideologies, and signifying practices.
Hybridity One of the most disputed terms in postcolonial studies, which can also be used in other contexts (e.g., hybrid natures). In relation to postcolonialism, it can be defined as the creation of new transcultural forms (cultural, political, and linguistic) within zones of contact produced by colonialism.
Neocolonialism New forms of economic and political control, in contrast to direct rule, of the powerful states of developed economies over the former colonies, which appear to be independent but in reality are not.
Orientalism Western constructions of the East as an opposite, working ideologically through a series of binary oppositions and underpinning imperialism.
Othering A way of defining and securing one's own positive identity through the stigmatization of an other; the self-identity of the colonizer and its culture is defined as superior through the identification of an inferior, opposite, colonized other.
Post-Structuralism A term coined to describe intellectual developments, primarily within French philosophy, that have had significant impacts within the social sciences and humanities. It encompasses diverse approaches and many theorists reject the post-structuralist label, but central elements are: emphasis on critical theory and the centrality of culture in understanding social organization; emphasis on texts and the significance of systems of knowledge; rejection of claims to truth and critique of dominant Western philosophy and culture.

Geography and Postcolonial Studies

Engagement by geographers with postcolonial studies can be traced back to the early 1990s. This engagement took different forms in different areas of the discipline: feminist geographers responded to criticisms of Western feminism by black feminists and women in the global South with a renewed focus on the politics of difference; feminist and development geographers began to explore issues of positionality, representation, and power relations that arise through cross cultural research; historical geographers renewed their interest in the spatialities and critical geographies of empire and colonialism (Blaut's *The Colonizer's Model of the World* was an early example); and there was an outpouring of critical historiographies that highlighted the closely bound nature of the history of modern human geography with the history of colonialism (Edward Said's *Orientalism* was a particular influence within geography). Running through these disparate engagements was a concern with Western centrism – that the previously unrecognised complicities between geography and colonialism had introduced a series of unacknowledged and unexamined assumptions into the very heart of the Anglo American discipline that are profoundly insensitive to the meanings, values, and practices of other cultures. Geographers have thus drawn on various aspects of postcolonial studies in an effort to 'de colonize' the discipline.

Definitions

Postcolonialism is a difficult and contested term. This is because it is sometimes assumed to refer to 'after colonialism' or 'after independence', describing the wide range of social, cultural, and political events arising specifically from the decline and fall of European colonialism that took place after World War II. However, while European colonialism may no longer exist as before, it is far from clear that colonialism has been relegated to the past. Not only do colonies and colonial powers still exist (Britain and France, for example, still possess scattered territories around the world and Indonesia – in relation to disputed territories in East Timor, Irian Jaya, Sipadan, Ligitan, and Batam – and China – in relation to Tibet – for example, are putative colonial powers) but critics argue that developed nations continue to enjoy economic and political advantages that accrue to them because of the alleged abuses of globalized capitalism. Thus, neocolonialism is seen as a continuation of the domination and exploitation of formerly colonized countries but through different means, primarily inequitable international trade and geopolitical relations.

The meaning of postcolonialism is not simply an epochal term and is not limited to 'after colonialism'. Rather, it refers to ways of criticizing the material and discursive legacies of colonialism that are still apparent in the world today and still shape geopolitical and economic relations between the global North and South. As a metaphysical, ethical, and political theory, it addresses issues such as identity, race, ethnicity and gender, the challenges of developing postcolonial national identities, and relationships between power and knowledge, both in terms of how colonial powers produced and used knowledge of colonized peoples in their own interests and in how these knowledges continue to structure inequitable relations between formerly colonized and colonizers. As a literary theory it deals with literature produced both by authors in colonial countries and by colonized peoples responding to colonial legacies by 'writing back', or challenging colonial cultural attitudes through literature. Thus postcolonial theory addresses both the ways in which the literature of the colonial powers has been used to justify colonialism through the perpetuation of repre sentations of colonized people as inferior, and the ways in which writers from (de)colonized countries have attempted to articulate, reclaim, and celebrate their cultural identities.

There is no single definition of postcolonialism and some critics reject the use of the term on grounds that it presupposes a continued sociocultural dominance by former colonial powers. However, broadly speaking, postcolonial approaches and perspectives can be said to be anticolonial. They have become increasingly important across a range of disciplines over the last 20 years, including both geography and development studies.

Origins and Debates

A number of different strands of thought have contrib uted to what we now understand as postcolonialism. Just as definitions are complex and contested, there is no single origin of postcolonialism, rather a number of different responses to colonialism and de colonization that have inspired both liberation struggles and academic inquiry. Many of these responses predate the period when the term postcolonial began to gain currency and are now claimed retrospectively as continuous or con tiguous with postcolonial politics and modes of cultural practice. They include the writings of African American W. E. B. Du Bois and South African Sol Plaatje in the late nineteenth/early twentieth centuries, the Harlem Renaissance, the Négritude Movement of the 1940s and 1950s, the writings of authors and critics as different as Trinidadian C. L. R. James, Nigerian Chinua Achebe, Colombian Gabriel García Márquez, Argentinian Jorge Luis Borges, and many 'Commonwealth' writers of the 1960s and 1970s, who are now seen as the precursors of postcolonial criticism. However, major influences on postcolonial theory since the 1980s can be traced to a number of key writers.

Fanon's Psychology of Colonialism and Racism

Frantz Fanon was perhaps the preeminent thinker of the twentieth century on the issue of de colonization and the psychopathology of colonization. Born in Martinique and active in anticolonial activism in French North Africa, his works inspired both anticolonial liberation movements throughout the world and postcolonial theorists (notably Edward Said and Homi Bhabha). His first book, *Black Skin, White Masks* (1952) analyzed the impact of colonial subjugation on the black psyche. Based on Fanon's personal experience as a black intellectual in a white dominated world, the book elaborates the ways in which the colonizer/colonized relationship is normalized. As a colonized subject and recipient of French education, Fanon conceived of himself as French, but felt profound disorientation after encountering French racism, which decisively shaped his psychological theories about culture. Fanon develops an understanding of racism as generating harmful psychological constructs that both blind the black man to his subjection to a universalized white norm and alienate his consciousness (it should be noted that Fanon's account has been criticized by feminists for failing to take into account the experience of subjugation of black women).

Fanon's explanation of the effects of being colonized by language was also influential: speaking the language of the colonizer meant that the colonized accepted, or were coerced into accepting, the collective consciousness of the colonial power, which identified blackness with inferiority, evil, and sin. In an attempt to escape the association of blackness with evil, the black man wears a white mask and internalizes white cultural values, cre ating a fundamental disjuncture between consciousness and body. This focus on alienation reveals the Marxist roots of Fanon's writings. Fanon's insistence that the category 'white' depends for its stability on its negation, 'black', that notions of white superiority cannot exist without an opposite (black inferiority), and that the origins of these notions were formed at the moment of imperial conquest, have been central in postcolonial critiques of colonialism and racism. His subsequent book, *The Wretched of the Earth* proposed the overcoming of the binary system in which black is bad and white is good through peasant revolution. This faith in the African peasantry and emphasis on language anticipates the work of Ngũgĩ Wa Thiong'o.

Ngũgĩ Wa Thiong'o and 'Writing Back'

Ngũgĩ, a Kenyan author and the first East African to publish a novel in English, embraced Fanonist Marxism in his third novel, *A Grain of Wheat*. He subsequently renounced English, Christianity, and the name James Ngugi as colonialist. He changed his name to Ngũgĩ wa Thiong'o and began to write solely in Gĩkũyũ and Swahili. While in exile following a period of imprison ment, he wrote his most influential book, *Decolonizing the Mind*, in which he argues for Africans to write in their native languages, rather than European languages, in order to overcome the suppressions of colonialism and to build an authentic African literature. Contemporary writers have drawn on Ngũgĩ's notions of the power of language; some appropriate the language of the colon izers to 'write back' from the margins and reclaim post colonial cultural identities (Salman Rushdie's novel *Midnight's Children* is proclaimed as one example, its liberal mix of English and Indian languages being a departure from conventional Indian English writing).

Subaltern Studies: Postcolonial History, Nationalism, and Identity

The Subaltern Studies Group emerged in the 1980s as a collective of scholars (including Ranajit Guha, Gyan Prakash, Dipesh Chakrabarty, Gayatri Spivak, and Partha Chatterjee) interested in the postcolonial societies of South Asia in particular and the developing world in general. Originally the collective adopted an approach, inspired primarily by Gramscian Marxism, of history from below, focusing on the lives of peasants – subalterns – rather than on elites. This was intended to counter colonialist versions of (primarily Indian) history and the elitism of bourgeois nationalists in the historiography of Indian nationalism, both of which had been apparent in debates in the 1970s between Cambridge University scholars and orthodox Marxist historians in India.

The focus on subaltern histories of insurgency and the role of nonelites as agents of political and social change, together with an engagement with complex processes of domination and subordination meant that the Subaltern Studies Group played a vanguard role in postcolonial theory. The collective made particularly influential contributions in studies of the active participation and intellectual creativity of subaltern classes in processes of nation state formation, in elaborating the discourses and rhetoric of emerging political and social movements (in contrast to highly visible actions like demonstrations and uprisings), and in attempting to recover the silenced voices of the formerly colonized without losing sight of the structural inequalities between the dominant and the subjugated. Tensions eventually grew within the collective between those who subscribed more closely to Marxist approaches and those who were engaging with post structuralist critiques. However, it was arguably these tensions that inspired debates around issues such as knowledge production, historicism, and the (ir)retriev ability of subaltern voices that have inspired much of postcolonial theory since the 1980s.

Postcolonial Theory: Said, Bhabha, and Spivak

While origins of postcolonialism can be traced to a number of different and disparate sources and while, in theoretical terms, it has been greatly influenced by Marxist political economy and post structuralist cultural and linguistic analysis, it was arguably Edward Said and subsequently Homi Bhabha and Gayatri Spivak who came to have the greatest influence within postcolonial theory. Said is best known for his book *Orientalism*, which is often considered the founding text of contemporary postcolonial theory. In this, Said critiqued the constel lation of false assumptions underlying Western attitudes towards Asia and the Middle East and decried the subtle and persistent Eurocentric prejudices against Arabs and Islamic peoples and their cultures. He drew implicitly on Fanon's ideas about binaries (what he called 'othering') to suggest that a long tradition of false and romanticized images had served as an implicit justification for Europe and the USA's colonial and imperial ambitions. He also denounced Arab elites for internalizing Orientalist ideas of Arabic culture. Although criticized for homogenizing 'the West', Said's use of Foucauldian theory to argue that colonial discourse has material effects on those being represented has been particularly influential.

Like Said, Bhabha challenges the tendency to treat postcolonial countries as a homogeneous category about which stereotypes can be perpetuated but, in addition to drawing on Fanon, his ideas are more heavily influenced by post structuralism, most notably the writings of Derrida, Lacan, and Foucault. His work is noteworthy for identifying ambivalence in colonial dominance and thus producing a more nuanced understanding of colonial power. In *Location of Culture*, Bhabha deploys concepts such as mimicry and hybridity to challenge the colonial production of binary oppositions (center/margin, civil ized/savage, enlightened/ignorant), suggesting that cul tures interact, transgress, and transform each other in a much more complex manner than binary oppositions allow. Thus hybridity has the potential to intervene and dislocate processes of domination through reinterpreting and redeploying dominant discourses; the spaces where differences meet become important. Although in cele brating cultural heterogeneity Bhabha has been criticized for relying on notions of original and distinct cultures, his emphasis on the subversive effects of hybridization has been influential.

Spivak has perhaps made the most innovative and substantial contributions to postcolonial forms of cultural analysis. She has been referred to as a 'Marxist feminist deconstructionist', since she sees each of these approaches as necessary but insufficient by themselves, always in tension yet productive together. Her overriding ethicopolitical concern has been the tendency of institutional and cultural discourses and practices to exclude and marginalize the subaltern, especially sub altern women. Her most influential (and controversial) essay, 'Can the subaltern speak?' explores whether sub alterns can speak for themselves or are condemned only to be represented and spoken for in a necessarily distorted fashion. Spivak locates the problem not with the inability of the subaltern to speak but with the unwillingness of the culturally dominant to listen. In *A Critique of Postcolonial Reason*, Spivak explores how European philosophers not only tend to exclude the subaltern, but actively prevent non Europeans from occupying positions as fully human subjects. She also explores the links between discourse and the inequities of global capitalism, in which women in developing coun tries are positioned as 'superexploited' labor, and coins the term 'strategic essentialism' to refer to a means of overcoming difference in temporary solidarity for the purpose of social action, a notion of particular signifi cance for feminists and women activists.

Spivak's writings on positionality have also been particularly influential, specifically on the need to 'un learn one's privilege as one's loss'. This unlearning requires recognition that one's race, ethnicity, class, gender, and nationality create relative privilege, which conditions certain prejudices and learned responses, limiting knowledge by preventing access to other knowledges. Unlike Western rationalism which, since the Enlightenment, has deemed that the world is knowable through observation, Spivak argues that there are certain knowledges, experiences, and existences that are closed off from the privileged view. This is only resolvable by 'learning to learn', which opens up the possibility of gaining knowledge of others. It involves an ethical rela tionship with the other that opens up spaces where those others can have a voice and where they can be taken seriously. The implication of Spivak's arguments is a need for continual questioning and reflexivity about one's relative position and vigilance of the potential negative effects of our attempts to create knowledge about and to represent others.

Critiques of Postcolonial Theory

Despite its significance, postcolonialism has been subject to a great deal of criticism. It has been accused of be coming institutionalized, representing the interests of a Western based intellectual elite (Bhabha, Said, and Spivak were employed at universities in Europe and/or the US when they made their most influential interventions, for example) who speak the language of the contemporary Western academy, perpetuating the exclusion of the colonized and oppressed. Some critics suggest that greater theoretical sophistication has created greater obfuscation; postcolonialism is too theoretical and not rooted enough in material concerns. Emphasis on dis course is accused of detracting from an assessment of material ways in which colonial power relations persist; not enough consideration is given to the relationship between postcolonialism and global capitalism; economic relations and their effects elude representation in much of postcolonial studies. Some critics berate postcolonial theory for ignoring urgent life or death questions, in cluding questions of inequality of power and control of resources, human rights, global exploitation of labor, child prostitution, and genocide. Postcolonial theory cannot easily be translated into action on the ground and its oppositional stance has not had much impact on the power imbalances between North and South. Critics accuse it of being preoccupied with the past, while new forms of Orientalism continue to disadvantage large parts of the world. While some of these criticisms are arguably somewhat overstated, especially when considering the role that Marxism has played in much of postcolonial thinking and the fact that its origins were primarily in modes of cultural analysis, the intersections between postcolonialism and development studies are one area where these tensions have proved productive.

Postcolonialism and Development

Until recently, there was little dialog between post colonialism and development, reflecting differences in disciplinary traditions, politics, wariness over motives, and divergences in the languages and concepts used to articulate core issues. Critics allege that development studies tend not to listen to subalterns and postcolonial studies neglect material issues such as poverty. The politics of postcolonialism diverges sharply from other discourses and, although it shares similarities with dependency theories, its radicalism rejects established agendas and accustomed ways of seeing. This means that postcolonialism is a powerful critique of development (as an idea and as a practice), fuelling mutual distrust between the two approaches. However, the possibility of producing a truly de colonized, postcolonial knowledge in development studies (including development geog raphies) became a subject of considerable debate during the 1990s. Postcolonialism is an increasingly important challenge to dominant ways of apprehending North–South relations and has begun to pose significant questions for development studies, particularly issues concerning the material effects of discourse and notions of agency and power. This has perhaps been most marked within gender and development and also in the rise of participatory approaches to development research.

Critiquing Discourses of Development

Postcolonialism critiques development discourses as unconsciously ethnocentric, rooted in European cultures and reflective of a dominant Western worldview and challenges the very meaning of development as rooted in colonial discourse, depicting the North as advanced and progressive and the South as backward, degenerate, and primitive. Early postcolonial writers, such as South African Laurens van der Post, challenged this assumption by referring to hunter gathers as the first affluent peoples. Postcolonialism has prompted questions about whether such indigenous systems of equity, reciprocity, and communalism are more advantageous to peoples of the South than the pursuit of capitalism, with its emphasis on individual wealth and incorporation into the global economy. Postcolonialism problematizes the ways in which the world is known, challenging the unacknowledged and unexamined assumptions at the heart of Western disciplines that are profoundly insensitive to the meanings, values, and practices of other cultures. The superiority of modern industrialization and technological progress is increasingly questioned, creating alternative knowledges to reshape views of non Western societies and their environments.

Some critics have suggested that to subject development to postcolonial critique might be considered a form of intellectual faddism and that while ever there are pressing material issues such as poverty in the world, concerns with the language of development are esoteric. However, language is fundamental to the way in which those interventions are ordered, understood, and justified and postcolonialism thus offers ways of understanding what development is and does, and why it is so difficult to think beyond it. The texts of development are written in a representational language, using metaphors, images, allusion, fantasy, and rhetoric to create imagined worlds that postcolonial critics argue bear little resemblance to the real world. Consequently, development writing often produces and reproduces misrepresentation. Postcolonialism seeks to remove negative stereotypes about people and places from such discourses, requiring that categories such as ‘Third World’ or ‘developing world’ are problematized and rethought and an understanding of how location, economic role, social dimensions of identity, and the global political economy differentiate between groups and their opportunities for development.

Knowledge and Power

Postcolonial critiques challenge the experiences of speaking and writing by which dominant discourses come into being. For example, terms such as ‘developing’ or ‘Third World’ homogenize peoples and countries and have other associations – economic backwardness, the failure to develop economic and political order, and connotations of a binary contest between ‘us’ and ‘them’, ‘self’ and ‘other’ – which are often inscribed in development writings. These practices of naming are not innocent. Rather they are part of the process of what Spivak calls ‘worlding’, or setting apart certain parts of the world from others. Said demonstrates how knowledge is a form of power, and by implication violence; it gives authority to possessors of knowledge. Knowledge has been, and to large extent still is, controlled and produced in the West. The power to name, represent, and theorize is still located there, a fact which postcolonialism seeks to disrupt.

Postcolonial theorists suggest that the texts of development are both strategic and tactical, promoting and justifying certain interventions and delegitimizing and excluding others. Power relations are clearly implied in this process, with certain forms of knowledge dominant and others excluded. The texts of development contain silences; postcolonial approaches interrogate who is silenced in and by these discourses, and why? Ideas about development are not produced in a social, institutional, or literary vacuum, but are part of a development industry that is implicated in relationships of power and domination. A postcolonial approach to development literature, therefore, can say a great deal about the apparatuses of power and domination within which those texts are produced, circulated, and consumed.

Development discourse promotes and justifies very real interventions with very real consequences. Contrary to critics of postcolonialism, such approaches do not simply represent a ‘descent into discourse’, where the effects of development languages are confined to the text and the possibilities of effecting change are denied. Rather, postcolonialism enables an exploration of the links between the words, practices, and institutional expressions of development, and between the relations of power that order the world and the words and images that represent the world.

Agency in Development

Postcolonialism invokes an explicit critique of the spatial metaphors and temporality employed in development discourses. Whereas previous designations of developing countries signaled both spatial and temporal distance – ‘out there’ and ‘back there’ – postcolonial perspectives insist that the ‘other’ world is ‘in here’. The ‘developing’ world is integral to Western notions of ‘modernity’ and ‘progress’, contributing directly to the economic wealth of Western countries through labor and resources. In addition, the modalities and aesthetics of the former colonies have partially constituted Western languages and cultures. Postcolonialism, therefore, attempts to rewrite the hegemonic accounting of time (history) and the spatial distribution of knowledge (power) that constructs the developing world. It also attempts to recover the lost historical and contemporary voices of the marginalized, the oppressed, and the dominated, through a radical reconstruction of history and knowledge production that seeks to recover the agency and resistance of peoples subjugated by both colonialism and neocolonialism.

Postcolonialism challenges the notion of a single path to development and demands acknowledgement of a diversity of perspectives and priorities. The politics of defining and satisfying needs is a crucial dimension of current development thought, to which the concept of agency is central. Postcolonialism prioritizes questions such as: who voices the development concern, what power relations are played out, how do participants' identities and structural roles in local and global societies shape their priorities and which voices are excluded as a result? Postcolonial approaches attempt to overcome inequality by opening up spaces for the agency of non Western peoples. Of course, poverty and unequal access to technology make this increasingly difficult; non Western academics, for example, rarely have the same access to books and technologies of communication as their Western counterparts. However, despite this, the work of academics in formerly colonized countries has led to a questioning of authorization and authority.

Significant questions have been raised for scholars and practitioners in development. By what right and on whose authority does one claim to speak on behalf of others? On whose terms is space created in which they are allowed to speak? Is this merely an attempt to incorporate and subsume non Western voices into Western canons? It is no longer feasible to represent formerly colonized peoples as passive, helpless victims. Their voices are now being heard, and their ideas are increasingly being incorporated into grassroots development policies, if not at the level of national governments and international development agencies and financial institutions. This is particularly the case in gender and development and in participatory approaches. In the former, postcolonial approaches have forced a move away from totalizing discourses and a singular feminism (based upon the vantage point of white, middle class Western feminists, which failed to acknowledge the differences between women) toward the creation of spaces where the voices of black women and women from the South can be heard. Postcolonial feminisms allow for competing and disparate voices among women, rather than reproducing colonialist power relations where knowledge is produced and received in the West, and white, middle class women have the power to speak for their 'silenced sisters' in the South. There are also some overlaps between feminist approaches and the rise of participatory methods, which, although still often problematic and not always particularly successful, are at the very least notable in their attempt to engage with postcolonial critiques concerning agency.

Postcolonial Geographies

In addition to its potential in development studies (including development geographies), postcolonialism is beginning to have significant impacts in human geography more broadly. Postcolonial theory demonstrates how the production of Western knowledge forms is inseparable from the exercise of Western power, attempts to loosen the power of Western knowledge, and reasserts the value of alternative experiences and ways of knowing. It articulates clearly some difficult questions about writing the history of geography, about imperialist representations and discourses surrounding 'the Third World' and about the institutional practices of disciplines. Postcolonialism has an expansive understanding of the potentialities of agency and shares a social optimism with other discourses, such as gender and sexuality in Western countries, where rethinking has helped generate substantial changes in political practice. Therefore, despite the seeming impossibility of transforming North–South relations by politics of difference and agency alone, postcolonialism is a much needed corrective to the Eurocentrism of much writing about non Western worlds and is becoming increasingly important in understanding and critiquing contemporary geopolitics.

While postcolonial approaches are viewed as constructive in many parts of the discipline, no coherent project of postcolonializing geography has emerged. The use of postcolonialism in geography must always be provisional and combined with political geography, feminism, and understandings of globalization and transnationalism. However, human geography risks irrelevance if it ignores the lessons of postcolonialism. The concerns of the ascendant (if not the most powerful) majority of the world's population are very different to those of the powerful Western minority; unless the intellectual preoccupations of the West acknowledge this shift they are likely to become ever more out of step with political and economic dynamics of the majority of the world's population. Anglo American geography has long had a tendency to universalize the parochial: economic geography is almost exclusively concerned with advanced economies; urban geography focuses on global and other Western cities and ignores the nature and effects of rapid urbanization taking place in developing countries; cultural geography's focus on identity and cultural politics is overwhelmingly Western centric; political geography, with its theorizations of geopolitics, territoriality, and the state, remains largely a view from the West.

Development geography is arguably one part of the discipline that has not abandoned area studies in favor of a view from nowhere and is thus perhaps more attuned to a postcolonial sensibility. However, recent years have seen a productive engagement across the discipline with postcolonialism. Some economic geographers are beginning to explore diverse forms of economy and to 'theorize back from' the South. Attempts are being made to 'de colonize' and provincialize urban geographies by challenging Eurocentric classification systems that rank 'world cities' and ignore diverse urbanisms around the world (Robinson's *Ordinary Cities* is one such example). Urban and cultural geographers are exploring how the imperial experience helped shape and continues to shape the economic, social, political, and cultural landscapes of the imperial powers themselves, not just those places they colonized (for example, Jacobs' *Edge of Empire*). Today's global cities are celebrated not only as sites of carnival, hybridity, and liberation, but also explored critically as containing very real vestiges of the past in terms of polarization, segregation, cultural hegemony, racialization, and nationalism, and thus as sites of ever evolving, competing, and complex identity and cultural politics. Political geographers are drawing on post colonial theory to examine new forms of Orientalism and their contemporary geopolitical consequences (e.g., Gregory's critique of Western responses to Iraq and Afghanistan in *The Colonial Present*).

Geographers are also becoming more attuned to the ethical dimensions of postcolonialism. Feminist and development geographers have been particularly inspired by Spivak's writings on ethical relationships between researcher and researched, developing participatory methodologies that negate some of the inequalities in power relationships and creating spaces in which the researched are able to speak for themselves. This relates not only to the subjects of contemporary research, but also to the recovery of historical agency and resistance (e.g., in Clayton's study of indigenous peoples of Vancouver Island). Historical and cultural geographers have challenged notions of unidirectional diffusion of discourses, knowledges, and cultures from metropoles to peripheries by exploring historical and contemporary interconnectedness (examples include analyses of colonial governmentality and ethical/fair trade). Postcolonialism is also inspiring thinking around the politics and ethics of teaching, with geographers drawing on the likes of Said and Spivak to challenge the negative portrayals of non Western places. Central to this has been a concern with the affective elements of pedagogy – recognition that negative stereotypes can produce disaffection and disconnection (e.g., representing Africa as a 'basket case' induces a lack of concern on the grounds that nothing can be done to improve conditions), and a need to retain hope, optimism, and an ethic of generosity in our critical geographies (e.g., by focusing on the remarkable advances that some African countries have made in less than 50 years of independence).

Geographers are responding to the postcolonial challenge of finding ways of comprehending other worlds, including our relations with them and our responsibilities toward them, without at the same time being invasive, colonizing, and violent. Postcolonialism thus holds out the possibility for both a new agenda in geography teaching and research and continued vitality and relevance of geography in a rapidly changing world.

See also: Eurocentrism; Imperialism, Cultural; Neocolonialism; Orientalism; Race; Subalternity; Third World.

Further Reading

Ashcroft, B., Griffiths, G. and Tiffin, H. (eds.) (1995). *The Postcolonial Studies Reader*. London: Routledge.
Bhabha, H. (1994). *The Location of Culture*. London: Routledge.
Blunt, A. and McEwan, C. (eds.) (2002). *Postcolonial Geographies*. London: Continuum.
Chakrabarty, D. (2000). *Provincializing Europe: Postcolonial Thought and Historical Difference*. Oxford: Princeton University Press.
Clayton, D. (2000). *Islands of Truth: The Imperial Fashioning of Vancouver Island*. Vancouver: UBC Press.
Crush, J. (ed.) (1995). *Power of Development*. London: Routledge.
Fanon, F. (1967). *Black Skin, White Masks* (trans. Markmann, C. L.). New York: Grove Press (first published 1952).
Gregory, D. (2004). *The Colonial Present*. Oxford: Blackwell.
Guha, R. (ed.) (1982). *Subaltern Studies*. New Delhi: Oxford University Press.
Jacobs, J. M. (1996). *Edge of Empire: Postcolonialism and the City*. London: Routledge.
Loomba, A. (1998). *Colonialism/Postcolonialism*. London: Routledge.
McEwan, C. (2003). Material geographies and postcolonialism. *Singapore Journal of Tropical Geography* 24(3), 340 355.
Robinson, J. (2003). Postcolonializing geography: Tactics and pitfalls. *Singapore Journal of Tropical Geography* 24(3), 273 289.
Robinson, J. (2005). *Ordinary Cities: Between Modernity and Development*. London: Routledge.
Said, E. (1978). *Orientalism*. London: Peregrine.
Sharp, J. and Briggs, J. (2006). Special Issue: Postcolonialism and Development: New Dialogues. *The Geographical Journal* 172(1), 6 77.
Spivak, G. (1988). Can the subaltern speak? In Nelson, C. & Grossberg, L. (eds.) *Marxism and the Interpretation of Culture*, pp 271 313. Urbana, IL: University of Illinois Press.
Spivak, G. (1999). *A Critique of Postcolonial Reason*. Cambridge, MA: Harvard University Press.
Thiong'o, N. W. (1986). *Decolonizing the Mind*. London: James Curry.

Relevant Websites

http://www.lib.virginia.edu/area studies/subaltern/ssmap.htm
Bibliography of the writings of the Subaltern Studies group, 1988 2000.
http://www.heinemann.co.uk
Link to Heinemann publishers and their African Writers Series and Caribbean Writers Series.
http://www.ipcs.org.au/
The Australian based Institute of Postcolonial Studies; contains links to various sites connected with the study of postcolonialism, journals, and reading lists.

Postconflict Geographies

J. O'Loughlin, University of Colorado, Boulder, CO, USA

Glossary

Consociationalism A constitutional arrangement in a country with endemic conflict between ethnic groups or regional interests to guarantee each some representation in parliamentary bodies. It is designed to end violence and promote democratic contests and voice for the diverse groups.

Partition A border cut through at least one community's national homeland, creating at least two separate political units under different sovereigns or authorities.

Proxy Wars Contests between states, often major powers, through which they arm and support their allies in their wars against each other. They were particularly common in the Second Cold War (1980–86) between the United States and the Soviet Union.

Secession A declaration of independence for part of a state's territory by a group present there; most secessionist movements are based on ethnic differences with the rest of the population.

Introduction

Trends in conflict, both civil and international, are turning down over the past 20 years. While there is still a disproportionate ratio of civil to international wars (about eight to one over the past two decades), the absolute numbers of both types currently underway are probably at an all time low. (Lack of reliable information before 1600 requires the insertion of uncertainty into this claim.) According to the Stockholm International Peace Research Institute (SIPRI), in 2005–06, there were 17 major armed conflicts in 16 locations. These were all civil wars since no interstate conflicts were active.

However, the relative scarcity of contemporary conflicts (from 1997 to 2006, there were 34 different major conflicts) belies their long term effects. Wars have long term effects not only on the populations in the conflict zones, but also on populations beyond these territories through the flight of refugees and by the economic impacts of the significant disruption in war zones and the externality effects on regional economies. The landscapes of war zones are well marked, sometimes indelibly, by the legacy of war in the form of minefields, damaged buildings and infrastructure, destroyed land scape icons such as churches and mosques, ravaged cities, abandoned lands and during post war rebuilding, by war memorials, and altered place and street names. Much of the postconflict research focus is on the economic con sequences of civil wars, especially capital flight, negative economic feedbacks, destruction of capital stock, and diversion of governmental resources to military use. Growing attention is now being given to the number and distribution of refugees and forced migrants near conflict zones and their potential effect in destabilizing neigh boring regions. Human geographic factors such as popu lation distributions, especially migration and refugee flows, have also been correlated to increased conflict risk.

Although the relationships between altered demo graphic circumstances and further conflict are under theorized, large and diverse populations in poor countries can be involved in a struggle over power, representation, and resources (education, healthcare, food, and employ ment). Ethnically concentrated populations may be mobilized to perpetuate conflict beyond a ceasefire if an ethnic/grievance agenda dominates their political discourse. Diasporas, especially those from a minority region in rebellion against the state, may also assist in support for disparate groups and hence, prolong conflict. The overseas support in the twentieth century for Irish, Armenian, Zionist (Israeli), Cuban, Ukrainian, Croatian, Polish, and Palestinian separatist or independent move ments kept them going at times when they were under severe threat in their home bases. Thus, wars in one country tend to have significant spillover effects on adjoining states, especially if border controls are weak, and more extensively on regional peace. If the conflict draws in major outside powers or involves use of a global resource, such as oil or diamonds, the global implications can be dramatic, long after the shooting stops.

Of course, not all conflicts follow routinized procedures and involve groups with uniforms, political agendas, and territorial aims. Recently, rebel groups are becoming relatively more involved in conflicts with objectives that are often economic, religious, or ideo logical. Further, attacks on unarmed civilians (terrorism) have risen dramatically in the past 20 years, especially in the Middle East, and the targets have both the objectives of destabilizing governments by undermining their popular support and also to wreck the social fabrics of communities and societies through generating repressive government responses and ethnic strife. The long term impacts of these kinds of conflicts are as yet unclear and while terrorism is not a new phenomenon, its presence and impact have only reached global consciousness in the past decade.

How the Causes and Nature of Wars Affect the War Outcomes

Few civil wars remain confined within the borders of the country at war, especially due to refugees displacing to neighboring states. Furthermore, rebel groups try to gain power by utilizing resources that they can access (e.g., diamonds, criminal networks, drug dealing, and informal taxes), and thus, conflicts tend to diffuse, making border zones prime territorial targets for governments and rebels. There is a considerable literature that separately examines the onset, escalation, and termina tion of civil wars, each taken as separate phases, disconnected from one another. At the same time, most of the literature has also looked at civil wars as self contained and homogenous phenomena, ignoring the connections of civil wars to both regional and local conditions. Grievance is thought to be the root cause of civil conflict; civil wars are more common in societies with low incomes. Yet, poverty itself cannot provide an adequate explanation of why groups resort to violent conflict. Economic misery and poverty are ubiquitous, but only a few countries experience violent conflict.

Geographic work stressing the diffusion of conflict and the locational attributes of civil war has implied that certain contextual circumstances influence the onset and proliferation of war. Particular geographical locations in the periphery of the world economy suffered dis proportionately as de colonization, superpower proxy wars, and Third World conditions created an environ ment of endemic poverty, and poor governance, in the 1980s especially fostered discontent and violent conflict. Strategic geopolitics – the relationships fostered by resource and strategic location dependence – have continued to create ‘shatterbelt’ regions well past the Cold War; these locations are likely to remain as hot zones. Shatterbelt regions, such as the Caspian Sea–Central Asia, are defined as areas with a precious natural resource, ethnic diversity, external interventions, and a history of conflict.

Another approach to the study of civil war ignores the question of why they start and focuses instead on how they continue. In general, factors affecting the onset of violent conflict are also thought to influence the duration of conflict and patterns of escalation. This line of research has given us additional insights about civil war dynamics, in particular the conditions under which peaceful settlement may be feasible. However, these approaches have also been almost exclusively focused on country level aggregates. The geographic location of a civil war within any particular country is fundamental for understanding conflict dynamics. Civil wars that take place in the periphery of countries tend to last much longer than those occurring close to national capitals, for example. Although disaggregated data presently do not yet exist in a manner that easily allows cross national comparisons, it is already clear that a disaggregated, spatial perspective on civil wars will augment our understanding.

War Settlements: Partitions and Secessions

To the victors go the spoils, and the winners typically set the post war political arrangements which include territorial changes, population movements (both forced and voluntary), constitutional arrangements for post war governments, and the installation into power of favored leaderships or parties. Massive territorial changes and post war migrations occurred in Europe after World Wars I and II, and resentment at the Versailles Peace Agreement in 1919 helped to spur the Nazi rise to power in 1930s Germany. A powerful picture of the effects of war related border shifts can be seen in Michel Foucher’s *Fragments d’Europe* atlas. The European map is littered heavily with lines representing previous borders and though most of these relict borders are no longer visible in the landscape or the nonmaterial representations, their identification clearly indicates the impermanency of territorial divisions and the frequency of post war adaptations by the peoples of war zones.

Two geographic options invariably come into play to try to resolve territorial disputes, especially around questions of ethnic homelands and control. Secession is a declaration of independence for part of a state’s territory by a group present there; most secessionist movements are based on ethnic differences with the population of the rump or metropolitan state. Many states have been created through this method, including the United States, India, Ireland, many Latin American countries, and Norway, and more recently, Slovakia, and the countries emerging from the former Soviet Union and former Yugoslavia. Declarations of secession are often followed by a civil war or guerilla conflict as the metropole tries to prevent a successful breakaway.

The second geographic division strategy is partition, defined by O’Leary “as a fresh border cut through at least one community’s national homeland, creating at least two separate political units under different sovereigns or authorities” (O’Leary, 2001: 54). But he goes further to remind us that reactions to a political partition are always subjective, though in a systematic manner – “there will be proponents, opponents and the indifferent, who are always with us” (O’Leary, 2007: 887). Tir’s review shows little evidence for the argument that partitioning leads to failure since partitioned countries tend to be ethnically diverse. Using a comparative description, O’Leary writes that “secessions are proposed, and opposed, with the same analogical battery – of cuts, tears, slashes and rips. … in fact secession is promoted and opposed as "unfastening", dividing along a previously established line of division. Secessionists usually have an established claim to a unit, either in recent or older history" (O'Leary, 2007: 889). Secessions often evolve into par titions when an intractable conflict shows no sign of victory for the rump state. In the twentieth century, O'Leary identifies six partitions that ended conflict but produced a legacy of hate and recrimination that resulted in further hostility, suspicion, and conflict. The partition of Ireland (1920) stimulated the Irish civil war and the future 'troubles' in Northern Ireland; the partition of Hungary by the treaty of Trianon (1920) left a large proportion of the Hungarian people outside the new state; the partition of Kurdistan (1920–23) into three countries (Iran, Iraq, and Turkey with smaller numbers in Syria and the Soviet Union) violated a promise of the Treaty of Sèvres (1920) to create a Kurdish state; Radcliffe's commission created by the British government worked on the partition of India (1947) between pre dominantly Muslim and Hindu states but events over took its deliberations and ended in the flight of millions to adjust to the new borders of Pakistan and India and long term hostilities between these states; partition plans for Palestine between Jewish and Muslim territories were developed in 1937 but again events overtook deliber ations and the eventual partition of 1948 created massive refugee outflows and further hostilities between Israel and its neighbors; and the Turkish army partition of Cyprus in 1974 after their invasion of the island created a 'pseudo state' (Turkish Republic of Northern Cyprus) for the predominantly Turkish minority region of the north. Post war partitions among the victorious Allies divided Germany and Austria into zones of occupation but eventually these were erased (after 45 years and 10 years, respectively). Korea is still divided along the 38th parallel as a result of the 1953 ceasefire in its civil conflict.

Pseudo-States

A peculiar institution has emerged in a number of post war situations in the 1990s where a permanent settlement has not been agreed and neither party is strong enough to impose its own territorial settlement – a reintegration or independence. Though the nomen clature is not yet agreed for the political units that sit in this gray zone, the term 'pseudo state' is gaining adherents to describe self proclaimed independent ter ritories that have emerged from the former Soviet Union. All have experienced conflicts triggered by the attempt by a minority in the newly independent state to carve out a territorial unit not subject to the control of the majority. The four most well known of these pseudo states have formed a 'secessionist international', all holding referenda at the end of 2006 'certifying' through popular mandates their *de facto* independence since 1991. Transniestria (a mixed ethnic slice of Russians, Ukrainians, and Moldovans in eastern Moldova), Abkhazia (previously mixed ethnically between Abkhaz and Georgians) in Northwest Georgia, South Ossetia (populated by Ossetians and Georgians), and Nagorno Karabakh (a mostly Armenian region of Azerbaijan) have successfully resisted international efforts to end their secession through proposals for devolved power to the ethnically minority regions. These conflicts are 'frozen' in the sense that there is little movement toward any permanent settlement and the continuance of the pseudo state status is supported by Russia. In contrast to Kosovo (under United Nations (UN) mandate and gaining independence from Serbia in February 2008), most pseudo states have rarely hit the news headlines, though accounts of the methods by which their regimes finance the secession (mostly through smuggling and other illegal contraband) have generally portrayed the regimes as thuggish and resistant to compromise. With a powerful benefactor, like Russia, such breakaway regions can survive despite the strongest opposition of the state from which they separated. This post war development, initially appearing as a temporary ceasefire arrangement, is increasingly taking on a more permanent appearance.

Economic Impacts of Conflicts

Though political geographers have focused on territorial alignments emerging from wars, most of the research on war outcomes has been completed by economists and political scientists. Macroeconomic evidence from Paul Collier and his colleagues in a large World Bank project has demonstrated that economic growth rate is reduced by 2% yearly during civil wars, the dominant type of modern conflict. Furthermore, the impact is not just in the country that experienced war but it is even more profound in neighboring states that get dragged into civil wars by the basing strategies of rebels, by the placement of refugee camps on their territory, and by the spillover of conflict across the borders. Modern conflicts generate clandestine political economies so that, for example, the study of post war reconstruction in Bosnia in the after math of the 1992–1995 war cannot be explained adequately without taking into account the critical role of smuggling practices and quasi public criminal com batants. A recent Oxfam report (2007) summarized the effects of civil wars in sub Saharan Africa in a few powerful statistics. The total cost of post Cold War (since about 1990) Africa's wars has been over $300 billion equivalent to all the foreign aid that the continent has received in that time and over 10 million deaths. Compared to African countries without conflicts, those with civil wars have had 50% more deaths, 15% more undernourished people, while life expectancy is reduced on average by 5 years. Further, indirect deaths for civilians, mostly through healthcare cuts to fund the militaries, are 14 times higher than deaths in combat. Overall, conflict shrinks economies by 15% on average and it takes up to a decade to recover, though obviously the length of time is a function of the severity and extent of the conflict. Capital flight, exploitation, and export without controls of 'lootable resources' such as diamonds, the exodus of educated individuals for better economic opportunities, and damaged infrastructure from fighting all contribute to extending the effects of conflict well beyond the end of fighting and reduce the capacity of the state apparatus to engage in post war rebuilding.

Refugees and Health

Refugee flows are believed to exacerbate the risk and proliferation of conflict through a process of diffusion, wherein refugees alter the balance of power (through ethnic populations or resource exploitation), or escal ation, an occurrence when refugee flows bring new combatants into a contentious situation. The impact of a refugee flow is contingent upon the political and military cohesion among the refugees, the stability of the host state, and the extent of external intervention.

Typical of wars, large movements of people change the ethnic ratios in specific regions and higher rates of disease and infections result. Despite the efforts of aid agencies, generally poorer nutrition, difficulties in finding reliable supplies of clean water, and inaccessibility to healthcare result in higher mortalities. Political scientists have clearly demonstrated these powerful indirect effects by showing that 'civil wars kill and maim people – long after the shooting stops'. By taking one sample year 1999, the additional burden of death and disability from the indirect and lingering effects of civil wars in the years 1991–97, was approximately equal to that incurred dir ectly and immediate from all wars in 1999. Using World Health Organization data, they could show the disability adjusted life years (DALYS) missing due to specific dis eases and conditions and how conflicts' long term effects disproportionately fall on women and children. In dis aggregating these global results by country, it is evident that the burden falls disproportionately on the sub Saharan African states though regional concentrations are also evident in South and Southeast Asia as well.

The subject of refugees combines traditional geo graphic interest in migration and in the geography of conflicts. Much of the work has focused on Africa, one of the world's concentrations of refugees from the numerous civil wars fought over the past half century. Mapping of refugee locations has been well served by the growth of satellite monitoring for international agencies so that high resolution (less than 1 m) imagery is now used to count tents and to estimate populations in camps in inaccessible areas, such as the current conflict zone of Darfur (Sudan). Temporal remote sensing monitoring allows a documentation of population movements, animal husbandry, land use intensities, and infrastruc tural development. While the UN tries to protect and manage refugee zones, the nature of conflict does not always allow such protection and careful monitoring is needed to estimate the level of protection needed. The African refugee crisis, in particular, shows no signs of abating as conflict zones remain too dangerous for resettlement. The geography of the refugee crisis exacerbates the crisis in the delivery of health services as refugees frequently move to better environ ments to try to eke out a living, and camps tend to be breeding grounds for disease. Refugees have been shown to be an important vector of transnational violence of civil war so that refugees are not only a consequence, but also a cause of violence. Their presence often pro vokes a violent response from locals affected by their arrival – and, refugees expand rebel social networks. Although the vast majority of refugees never directly engages in violence, refugee flows may facilitate the transnational spread of arms, combatants, and ideologies conducive to conflict; they alter the ethnic composition of the state; and they can exacerbate economic competition in localities.

The politics of managing refugees, both in war zones and in destination countries where refugees seek to move, has its own geography. Most refugees from conflicts remain in the countries adjacent to the war zones – for example, in Iran and Pakistan from the 1990s Taliban war in Afghanistan. Other countries generally expect any refugees that manage to traverse the distances separating them from the war zones to remain only temporarily – as was the case for Bosnian refugees in Germany during the 1992–95 war. While the notion of a refugee is typically one that implies a temporary stay, the situation in most parts of the world is now a protracted one with refugees in temporary camps for decades, as is the case of three generations of Palestinians in the countries surrounding Israel. The refusal of Israel to accept any of them as returnees in a peace settlement prolongs both their precarious existence and the difficulty of a permanent peace settlement.

The Continuing Problem of Landmines

United Nations International Children's Fund (UNICEF) estimates that in total, there are an estimated 110 million mines and other unexploded ordnance (UXO) scattered in 64 of the countries worldwide, remnants of wars from the early twentieth century to the present. Africa alone has 37 million landmines in at least 19 countries, with Angola by far the most densely affected zone with 15 million land mines and an amputee population of 70 000, the highest rate in the world. Chechnya (the southern republic in Russia that has been the scene of a civil war for over a decade) proportionately has more injuries per population than any other war zone. The mine problem has received a lot of attention due to the award of the Nobel Peace Prize to the ICBL (International Campaign to Ban Landmines) in 1997 and the involvement of movie stars and other prominent persons, such as Princess Diana, in the cam paign. Despite this belated awareness, the removal of mines proceeds at a glacial pace due to the danger, cost, and lack of international agreement on priorities, between \$300 and \$1000 per mine. Efforts to prohibit the future use of mines by militaries have foundered on their cheap (\$3–\$10) and easy accessibility, and their effectiveness in military campaigns. Geographers have argued that in mine ridden countries, such as Mozambique and Angola (legacies of protracted civil wars after independence in 1975), any notion of sustainable development is illusory as agriculture, transportation, resource access, and general international investment are hindered. In Bosnia, only about 4% of the mines laid over a decade ago have been removed and at present rates of removal, it will take over a century to complete the process. Modern geographic tools such as geographic information systems and remote sensing have proved to be valuable in monitoring, map ping, and explaining to stakeholders (usually local farmers) the various options for removal and reclaiming lands currently too dangerous to use.

Conclusions

Most of the academic work on post war developments have centered on the institutional arrangements that have been agreed to, or imposed on, the warring parties. In particular, the promise of consociational procedures has drawn on valuable experiences in Northern Ireland (after 1998) and South Africa (after 1990). However, each war context is different and though useful lessons can be learned from earlier implementations of constitutional provisions, there is little agreement on what constitutes a 'best practice'. The current conflict in Iraq and the apparent intractability of internecine hostilities between Kurds, Shias, and Sunnis illustrate the difficulties of simple application of paper provisions.

While much of the political geographic work on postconflict environments has concentrated on the changes in the visual landscape, much of it has a de scriptive character that documents place name changes (e.g., from Danzig to Gdansk after World War II), preferences for street names (e.g., from Communist to nationalist heroes in the states of the former Soviet Union), the destruction of iconographic landscape elements of a previous regimes (historical memorials, churches or mosques, cemeteries, etc.), and shifts in border markers and controls. However, as indicated in this brief review, the economic, health, and environ mental difficulties of people in war zones are both long term and very difficult to resolve because most wars do not end definitively and tend to have consequences that last for generations. Recriminations, rather than restitution, respect, and conciliation, are the most likely consequences of most conflicts.

See also: Borderlands; Development I; Development II; Nation; Political Boundaries; Political Ecology; Refugees and Displacement; State; Sustainable Development.

Further Reading

Collier, P. *et al*. (2003). *Breaking the Conflict Trap: Civil War and Development Policy*. New York: Oxford University Press and the World Bank.
Dahlman, C. and Tuathail, G. O. (2005). Broken Bosnia: The localized geopolitics of displacement and return in two Bosnian places. *Annals of the Association of American Geographers* 95, 644 662.
Ghobarah, H. A., Huth, P. and Russett, B. (2003). Civil wars kill and maim people long after the shooting stops. *American Political Science Review* 97, 189 202.
Kolsto, P. (2006). The sustainability and future of unrecognized quasi states. *Journal of Peace Research* 43, 723 740.
Kalipeni, E. and Oppong, J. (1998). The refugee crisis in Africa and implications for health and disease: A political ecology approach. *Social Science and Medicine* 46, 1637 1653.
Mather, C. (2002). Maps, measurements, and landmines: The global landmines crisis and the politics of development. *Environment and Planning A* 34, 239 250.
O'Leary, B. (2001). Elements of right sizing and right peopling the state. In O'Leary, B., Lustick, I. S. & Callaghy, T. (eds.) *Right Sizing the State: The Politics of Moving Borders*, pp 15 73. Oxford: Oxford University Press.
O'Leary, B. (2007). Analyzing partition: Definition, classification and explanation. *Political Geography* 26 (in press).
Salehyan, I. (2007). Refugees and the study of civil war. *Civil Wars* 91, 27 141.
Tir, J. (2005). Dividing countries to promote peace: Prospects for long term success of partitions. *Journal of Peace Research* 42, 545 562.
Unruh, J. D., Heynen, N. C. and Hossler, P. (2003). The political ecology of recovery from armed conflict: The case of landmines in Mozambique. *Political Geography* 22, 841 861.

Relevant Websites

www.icbl.org
International Campaign to Ban Landmines.
www.stabilisationunit.gov.uk
The Stabilisation Unit (targeted assistance projects to countries after violent conflict).
web.worldbank.org
The World Bank.

Postdevelopment

J. Nederveen Pieterse, University of Illinois, Urbana-Champaign, IL, USA

Glossary

Alternative Development Critical, bottom-up oriented approaches to development that are centered on society (rather than on the state or market).

Deep Ecology A radical approach to ecology that rejects anthropocentrism.

Discourse Analysis The scrutiny of language and text to show structures of thought.

Easternization International East Asian and Asian influences.

Hermeneutics Rules of interpretation.

Luddism Violent nineteenth-century reactions against mechanization (such as the sabotage and 'machine breaking' by English weavers of mechanical spinning wheels).

NGOs Nongovernmental organizations.

Along with 'antidevelopment' and 'beyond development', postdevelopment is a radical reaction to the dilemmas of economic and social development in the developing countries. Perplexity and extreme dissatisfaction with business as usual and conventional development rhetoric and practice are keynotes of this perspective. Mainstream development policies are rejected because they don't work; they impose science as power, and inflict cultural Westernization and bring environmental destruction. They are rejected not merely because of their results but because of their intentions, worldview, and mindset. Economism implies a reductionist take on existence. Thus from the viewpoint of postdevelopment it is not the failure of development that is feared but its success.

Postdevelopment starts out from the realization that attaining a middle class lifestyle for the majority of the world population is impossible. In this view, the 'development decades' have been an irresponsible experiment that, in the experience of the world majority, has failed. In time this has led to a position of total rejection of development. Antidevelopment thinkers such as Gustavo Esteva in Mexico find that development stinks. Postdevelopment overlaps with Western critiques of modernity and technoscientific progress. It stands to development as 'deep ecology' does to environmental management. There are different strands to this way of looking at development. 'Antidevelopment' is inspired by anger with development business as usual. 'Beyond development' (*au delà de développement*) combines this aversion with scanning the horizon for local alternatives. 'Postdevelopment' combines these views with a Foucauldian methodology of discourse analysis and a theory and politics inspired by post structuralism. These positions are not all consistent, yet they sufficiently overlap to group them together under the heading of postdevelopment.

Development is the management of a promise – and what if the promise does not deliver? Living in Chiapas or other oppressed and poor areas, chances are that development is a bad joke. Postdevelopment is not alone in looking at the shadow of development policies; all critical approaches to development deal with its dark sides. Dependency theory raises the question of global inequality. Alternative development focuses on the lack of popular participation. Human development insists on the need to invest in people. Postdevelopment focuses on the underlying premises of development and what sets it apart from other critical approaches is its rejection of development.

Since the 1980s activists and intellectuals in Latin America, Asia, Africa, and Europe have taken up these views, which have become prominent since they coalesce with ecological critiques and through bestsellers such as Wolfgang Sachs's *The development dictionary*. Postdevelopment critics typically problematize poverty, view development as Westernization, and question modernism and science.

Economists speak too readily of poverty. One can distinguish between 'frugality', as in subsistence economies, 'destitution', which arises when subsistence economies are weakened through growth strategies, and 'scarcity', when the need for commodities takes over. A policy that follows is to implement growth strategies with caution, building on frugal lifestyles with appropriate technology and maximum use of local resources. However, rejecting economic growth or social development does not necessarily follow.

Viewing poverty simply as a deficit is to adopt the perspective of economism. Approaches such as conscientization and human scale development acknowledge the agency of the poor. Majid Rahnema criticizes the way poverty alleviation campaigns present their case and give the public distorted impressions of how the world's poor live. They are often presented as incapable of acting in their own interest and as preventing development experts from helping them. As counterpoints Rahnema draws attention to 'vernacular universes' that provide hope and strength, to the spiritual dimension and to 'voluntary or moral poverty'. In this view development economism is truly pauperizing. Valid as these considerations are, a consequence is that poverty alleviation – for what such efforts are worth – slips off the map. Another problem is that less market participation does not necessarily imply more social participation – unless we romanticize poverty and equate poverty with purity (and the indigenous and local with the original and authentic). The step from a statistical universe to a moral universe is worth taking, but a moral universe also involves action, and which action follows?

According to Serge Latouche, Arturo Escobar, and other postdevelopment thinkers, development is Westernization, an external approach based on the model of the industrialized world. What are needed instead, they argue, are endogenous approaches and discourses. Development, in this view, is a successor to colonialism. With decolonization came critiques of cultural imperialism, Coca Colonization and later, McDonaldization. According to the cultural homogenization thesis, Western multinationals, media, and consumerism induce cultural uniformity. Yet, in effect, this denies the agency of the global South. It denies the extent to which the South also owns and promotes development. Several development perspectives, such as dependency theory, alternative development, and human development, originate mainly in the South. Furthermore, what about 'Easternization', as in the East Asian model, touted by the World Bank as a 'development miracle'? What about the influence of Japanese management techniques and Toyotism? What about the fast rising influence of China? Besides, in view of steep historical differences between Europe and North America 'Westernization' is a lumping concept. This argument also overlooks more complex assessments of globalization. If we adopt a polycentric approach, the rejoinder to Eurocentrism is the recognition that multiple centers, also in the global South, shape development discourses.

Antimodernism

Part of anti Western sentiments is antimodernism. No doubt development policy suffers from a condition of 'psychological modernism', has erected monuments to modernism and vast infrastructures, and big dams that place technological progress over human development. The worship of progress is not reserved to the West. Developing countries have used science as an instrument of power, creating 'laboratory states', with high tech modernization drives from China to India. The nuclear tests in South Asia are another rendezvous of science and *raison d'état*.

Aversion to modernism also exists in the West. There are many overlaps between critical theory and counterculture in the West and antimodernism in the South. The alternative economist Schumacher ('small is beautiful') found inspiration in Buddhist economics and Fritjof Capra in Eastern mysticism, while the critical Indian thinker Ashis Nandy's outlook has been shaped by Freud, the Frankfurt School, and Californian psychology along with Gandhi and Tagore. Part of the critique of modernism is the critique of science. A leitmotif, also in ecological thinking, is to view science as power. 'Science' here means Enlightenment thinking and positivism, an instrument in achieving mastery over nature. Critique of Enlightenment science runs through the work of Vandana Shiva. However, this is not a simple argument. Vandana Shiva was originally trained as a physicist. Science has been renewing itself, for example, in quantum physics and chaos theory, and undergoing paradigm shifts leading to 'new science'. Besides, there are countertrends within science. In social science, positivism is no longer the dominant temperament; increasingly the common sense in social science is constructivism. Positivism prevails in economics, but is also under attack. A distinction should be made between critique of science and antiscience. Acknowledging the limitations of science, the role of power/knowledge, and the uses made of scientific knowledge does not necessarily mean being antiscience. Critique of science is a defining feature of new social movements, North and South. Yet ecological movements use scientific methods to monitor energy use, pollution, and climate changes, as in 'green accounting'. Antidevelopment at times sounds like Luddism, longer on rhetoric than on analysis. Also in the global South being 'antiscience' is not a viable option. It is more appropriate to view modernism as a complex historical trend, which is partly at odds with simple modernization.

Development discourse, according to Arturo Escobar, has been a mechanism for producing and managing the Third World, just as the Western scholarship of the Orient – the Orientalism analyzed by Edward Said – produced and managed 'Oriental societies' during colonial times. Discourse analysis, the careful scrutiny of language and text as a framework of structures of thought, contributes to understanding colonialism as an epistemological regime and serves to analyze development policy talk. Discourse analysis has become a critical genre in development studies. Discourse analysis is not specific to postdevelopment; what is distinctive for postdevelopment is that from a methodology discourse analysis has been turned into an ideological platform. According to Escobar, World Bank studies and documents "all repeat the same story." However, this ignores the tremendous discontinuities in the Bank's discourse over time (redistribution with growth in the 1970s, structural adjustment in the 1980s, poverty alleviation and social liberalism in the 1990s, and good governance, anticorruption, and social risk mitigation in the 2000s).

Many concerns of postdevelopment are not new and are shared by other critical approaches to development.

Postdevelopment parallels dependency theory in seeking autonomy from external dependency, but takes this further to development as a power/knowledge regime. Postdevelopment faith in the endogenous parallels the emphasis on self reliance in dependency theory and alternative development. While dependency thinking focuses on the nation state, postdevelopment is concerned with local and grassroots autonomy. Alternative development shares with postdevelopment the radical critiques of mainstream development but retains belief in and redefines development. The record of development is mixed and does include achievements, as noted below, so what is the point of rejecting it *in toto*?

Antimanagerialism

Adherents of postdevelopment are not interested in alternative development approaches, but rather in alternatives to development. Alternative development is rejected because it reflects the same worldview that has produced the mainstream concept of science and development. Alternative development is even more odious than mainstream development because of its friendly exterior. According to Gustavo Esteva, alternative development is a deodorant to cover the stench of development.

Antimanagerialism and dichotomic thinking may explain the antagonism to alternative development. Development thinking is steeped in social engineering and the ambition to shape economies and societies, which makes it an interventionist and managerialist discipline. It involves telling other people what to do in the name of progress, modernization, nation building, mobilization, sustainable development, human rights, poverty alleviation, empowerment, and participation (participatory management). Through postdevelopment runs an anti authoritarian sensibility, an aversion to control, and perhaps an anarchist streak. The post structuralism inspired by Michel Foucault also involves an 'antipolitical' sensibility, as a late modern skepticism. If the public sphere is constructed through discourse and if any discourse is a claim to truth and thus a claim to power, what may follow is political agnosticism. This also arises from the preoccupation with autonomy, the problem of representation, and the indignity of representing 'others'.

Another fundamental objection to mainstream development is that it is inherently antidemocratic. Viewing economic and social development policies through the lens of democratization is pertinent enough, in relation to foreign aid and the Asian authoritarian developmental states. Development managerialism not only involves states but also international institutions, donors, and the 'new managerialism' of nongovernmental organizations (NGOs), which all lack humility. In postdevelopment there is suspicion of alternative development as an 'alternative managerialism', which makes sense in view of the record of many NGOs and their donor driven interventions in developing countries. So what to do? One response is nothing. However, doing nothing comes down to an endorsement of the status quo. Under the heading of 'post' thinking, this is actually profoundly conservative.

An Assessment

Postdevelopment thinking is uneven. For all the concern with discourse analysis, the use of language is often sloppy. In proposing to do away with development, postdevelopment essentializes 'development' and implies that it is possible to arrive at an unequivocal definition. Essentializing 'development' is necessary to arrive at the radical repudiation of development, and without the antidevelopment pathos the postdevelopment perspective loses its foundation.

At times one has the impression that postdevelopment turns on a language game rather than an analytic. Some of the claims of postdevelopment are misleading and misrepresent the history of development. *The development dictionary* refers to the 1940s as the beginning of the development era. However, the application of development to the South started with colonial economics; besides, development has an older history in the latecomers to industrialization in Central and Eastern Europe and in Soviet economic planning. Dichotomic thinking, pro and antidevelopment, underrates the dialectics, the zigzag moves, and the complexity of motives and motions in modernity and development. Even though at given points particular constellations of thinking and policy seem to present a solid façade, the actual course of development theory and policy shows constant changes of direction and ongoing improvisations. Postdevelopment's take on real existing development is narrow. Many reflections are general and do not discuss cases. The instances cited in postdevelopment literature mainly concern Africa, Latin America, and South Asia. The experience of East Asia, Southeast Asia, China, and newly industrializing countries is typically not discussed, even though they are the current trendsetters of development.

Fine points of theory aside, what is to be done? Postdevelopment does make positive claims and affirms Gandhian frugality, conviviality, grassroots movements and local struggles, indigenous knowledge, and cultural diversity. However, none of these are specific to postdevelopment nor do they necessarily add up to rejecting development. In response to the commonsense question what do we do now, the tenor in most sources is to stop at critique, which means an endorsement of the status quo and, in effect, more of the same.

If we consider neoliberal critiques of the develop mental state side by side with postdevelopment critiques of development power, the parallels are striking. Both agree on state failure, though for different reasons. According to neoliberal critics, states fail because of rent seeking and inefficiency; postdevelopment critics object to authoritarian social engineering and the faith in pro gress. Yet the net political effect turns out to be much the same. In other words, there is an elective affinity bet ween the development agnosticism of neoliberalism and postdevelopment.

The development dictionary features critiques of the market, state, production, basic needs, etc., which are historically informed but overstate their case, offer no alternatives, and ultimately fall flat. According to Escobar, many of today's social movements in the global South are involved in or mediated by antidevelopment discourses. However, social movements in the South are much too diverse to be captured under a single heading. Many popular organizations are concerned with 'access' to development programs, with inclusion and participation, while others are concerned with renegotiating develop ment, decentralization, or alternative political action. 'Antidevelopment' is too simple and rhetorical a de scription for the views of the 'victims of development' and 'victims of development' is too simple a label. Early alternative development views also underestimated the desire for and appeal of development and engaged in 'island politics'.

The political horizon of postdevelopment is one of resistance and local struggles *à la* Michel Foucault rather than emancipation. Postdevelopment's use of discourse analysis as an ideological tool becomes apparent when it comes to politics. Foucault's imagination of power is an imagination without exit. Accordingly, postdevelopment takes critique of development to the point of retreat. Retreat from business as usual can be a creative position from which an alternative practice can grow. Thus critical theory, though pessimistic in outlook, served as a point of reference and inspiration. However, the imaginary of power that inspires postdevelopment leaves little room for forward politics. Postdevelopment's quasi revolutionary posturing reflects a hunger for a new era and a romantic nostalgia politics of glorifying the local, the grassroots, and community. Postdevelopment politics range from the 'romance of resistance' to a convention ally compartmentalized world.

Marx, and Joseph Schumpeter after him, referred to modern capitalism as a process of 'creative destruction'. What happens in postdevelopment is that of 'creative destruction' only destruction remains. What remains of the power of development is only the destructive power of social engineering. Gone is the recognition of the achievements of development policies, the improvements in healthcare, the reduction of infant mortality, the enormous reduction in poverty levels particularly in East Asia and China, and the creativity of developmental changes. Instead postdevelopment offers a cultural cri tique of development and a cultural politics in which development is a stand in for modernity, and the real issue is modernity.

Postdevelopment is caught in rhetorical gridlock. Using discourse analysis as an ideological platform in vites political quietism. In the end postdevelopment offers no politics besides the self organizing capacity of the poor, which lets the development responsibility of states and international institutions off the hook. Post development arrives at development agnosticism by a different route but shares the abdication of development with neoliberalism. Since most insights in postdevelop ment sources are not specific to postdevelopment (and are often confused with alternative development), what is distinctive is the rejection of development. Yet the rejection of development does not arise from post development insights as a necessary conclusion, that is, one can share its observations without arriving at this conclusion, so there is no compelling logic to post development arguments.

Postdevelopment belongs to the era of the 'post' as in post structuralism, postmodernism, postcolonialism, and post Marxism. It is premised on an awareness of endings, the 'end of modernity' and the 'crisis of the future'. Postdevelopment parallels postmodernism both in its acute intuitions and in being directionless in the end, because of the refusal to translate, or lack of interest in translating, critique into construction. It also matches the neotraditionalist reaction to modernity. There are romantic and nostalgic strands to postdevelopment and its reverence for community and the traditional and its attitude toward science and technology shows elements of neo Luddism.

Postdevelopment is based on a paradox. While it is part of the broad critical stream in development, it shows no regard for the progressive potential and dialectics of modernity, for democratization, soft power technologies, and reflexivity. Social movements and civic activism, North and South, cannot be simply captured under the label 'anti'. Postdevelopment's strength is a hermeneutics of suspicion, an antiauthoritarian sensibility, and hence a suspicion of 'alternative managerialism'. However, since it fails to translate this sensibility into a constructive pos ition what remains is whistling in the dark. What is the point of declaring development a 'hoax' without pro posing an alternative?

Alternative development thinking looks at develop ment from the point of view of the disempowered, from the bottom up, and combines this with a perspective on the role of the state; in brief, a strong civil society needs a strong state. Postdevelopment adopts a wider angle in looking at development through the lens of the problematic of modernity. Yet, though its angle is wide, its optics is not sophisticated and its focus is blurred. Its take on modernity is one dimensional and ignores different options for problematizing modernity, such as modernities in the plural, new or reflexive modernity, and its corollary in relation to development, reflexive development. Postdevelopment and 'alternatives to development' are flawed premises, flawed not as sensibilities but as positions. The problem is not the critiques, which one can easily enough sympathize with, but the accompanying rhetoric. 'Alternatives to development' is a misnomer because no alternatives to development are offered. There is critique but no construction, resistance but no transformation. 'Postdevelopment' is misconceived because it attributes to 'development' a single and consistent meaning which does not match either theory or policy and thus replicates the rhetoric of development rather than penetrating its polysemic realities. It echoes the 'myth of development' rather than leaving it behind. Postdevelopment makes engaging contributions to collective conversation and reflexivity about development and as such contributes to philosophies of change, but its contribution to politics of change is meager. While the shift toward cultural sensibilities is a welcome move, the plea for 'people's culture' or indigenous culture can lead, if not to ethnochauvinism and 'reverse orientalism', to reification of both culture and locality or people.

See also: Desertification; Development I; Development II; Imperialism, Cultural; Indigenous Knowledges; Modernity; Orientalism; Participation; Poststructuralism/Poststructuralist Geographies; Poverty.

Further Reading

Escobar, A. (1992). Reflections on 'development': Grassroots approaches and alternative politics in the Third World. *Futures* 24, 411 436.

Escobar, A. (1995). *Encountering development: The making and unmaking of the Third World*. Princeton, NJ: Princeton University Press.

Latouche, S. (1993). *In the wake of the affluent society: An exploration of post development*. London: Zed.

Nanda, M. (2003). *Prophets facing backwards: Postmodern critiques of science and Hindu nationalism in India*. New Brunswick, NJ: Rutgers University Press.

Nederveen Pieterse, J. (2001). *Development theory: Deconstructions/reconstructions*. London: Sage.

Rahnema, M. and Bawtree, V. (eds.) (1997). *The post development reader*. London: Zed.

Rist, G. (1998). *A history of development: From Western origins to global faith*. London: Zed.

Sachs, W. (ed.) (1992). *The development dictionary: A guide to knowledge as power*. London: Zed.

Said, E. W. (1985). *Orientalism* (original edn., 1978). Harmondsworth: Penguin.

Posthumanism/Posthumanistic Geographies

J. Lorimer, King's College, London, UK

Glossary

Affect It refers to both the material, ecological properties of a body and the forces and processes that link them together. It describes prediscursive, embodied experiences that are subsequently codified into subjective emotions.

Becoming It describes the state of bodies under a vitalist ontology where nothing is fixed and all entities are in a constant state of flux. A politics and philosophy of becomings differs radically from one concerned with fixed and transcendent beings.

Immanence It refers to the ontological condition in which present forms are understood as one actualization of a series of unlimited virtual potentialities. In contrast to transcendence, immanence describes a world without immutable essences and universal archetypes that is more open to contingent differentiation. It is most commonly associated within the philosophy of Deleuze and his engagement with Spinoza.

More-Than-Human An approach to geography and the social sciences more generally that is open to the agency of nonhumans and recognizes the material and affective interlinkages that cross between humans and nonhumans. More-than-human has been suggested as an alternative label to posthuman to describe recent work in cultural geography.

Vitalism A tradition in philosophy that believes in the existence of a life force in material assemblages that cannot be explained through mechanistic approaches. As an ontology, it argues for the immanent emergence of form, rather than the existence of transcendent archetypes. Vitalism is often associated with the philosophy of Henri Bergson.

Introduction

Posthumanism is an ambiguous term that suffers from similar semantic uncertainties as postmodernism. Its 'post' prefix hints at the arrival of a new epoch – as 'post'humanism it claims to identify a new mode of being in the world that departs significantly from the conditions of humanism. As post 'humanism', the 'post' radicalizes its suffix to suggest a new model of politics and philosophy; however, it does so in a fashion that echoes its antecedent's critical ethos. Posthumanism thus refers to the populist diagnosis of a new era, a new mode of critical enquiry that defines itself in relation to humanism, and a working through of the latter's critical tradition.

Humanistic geography puts the (singular) figure of the human front and center and examines the role of human agency, creativity, and consciousness in making sense of the world. It understands the human in binary opposition to the nonhuman. In contrast and in different ways, posthumanist geographers cast their analytical net a little wider to challenge a singular model of the human subject and to blur and enliven the boundaries between the human and the nonhuman. Empirically, they push in two interwoven directions – within the category of the human and along its margins – to formulate new spatial philosophies and politics.

Posthumanism has been greeted with ambivalence by human geographers and other social scientists. Some have welcomed the fresh perspective it provides on our contemporary molten and cosmopolitan condition, where it is very difficult to separate humans and nonhumans, and the singular model of the human subject has been thoroughly critiqued. Others are concerned about a turn away from humanism and the critical resources this provides, while a third strand questions whether we have ever been human (in a formal sense) and doubts the need for such an epochal theoretical turn. Posthumanist geography is still an emerging, open, and contested field.

The broader movement toward posthumanism in social theory and the social sciences has been driven by at least two forces. First, there is a growing public awareness in the West of the empirical impossibility of separating nature from society and a questioning of the political advantages that supposedly accompany such a division. This emerging posthuman sensibility has been triggered by concerns and debates over innovations, events, and crises in a number of diverse arenas. These include recent development in technoscience (e.g., genetic modification, artificial intelligence, nanotechnology, and reproductive technologies), the food system (like the risk of 'mad cow disease' and the rise of the organic movement), public health (in the form of zoonoses such as avian influenza), and climate science (where the diagnosis of climate change has demonstrated the global reach of anthropogenic activity).

Popular reactions to these developments have varied widely but many share a sense that we have crossed a threshold and are entering a new era. For some technocratic optimists this brave new world is full of the emancipatory promise of modern science. It presents new opportunities for human enhancement, individual self realization, and significant new arenas for (capitalist) economic growth and profit. For more conservative and environmentalist critics, this novel posthuman condition is apocalyptic and threatens our human identity and the future habitability of the planet. The stakes are high in these debates and the battle lines are drawn and fiercely contested.

Second, posthumanism arrived within the academy on the back of a growing intellectual dissatisfaction with humanism. This critical mode of posthumanism was driven in part by the popular reappraisal of the human, and has challenged both the constitution of the human category itself and the possibility of drawing neat boundaries. Feminist, post structuralist, postcolonial, and queer theorists have all questioned the possibility of identifying universal and authentic models of human consciousness and agency. They have drawn attention to the exclusionary work done by adhering to a singular model of human identity. This historicizing of humanism found a home in a reinvigorated cultural geography, which maintains humanism's critical imperatives and has sought to map the vast diversity of existing human subjectivities and to deconstruct the processes through which they are constructed and contested.

Furthermore, on the margins of the category human, vitalist philosophers, ethologists, neuroscientists, and sociologists of science have highlighted the material and immaterial interlinkages that blur the human–nonhuman boundary. This vitalist, material turn has raised the possibility of a more than human geography that would run alongside a posthumanist model of identity and difference. This current maintains the humanist attention to the body and embodied practice but is more open to the materialities and affective forces that flow between humans, organisms, and objects, cutting across modern ontological divides. This critical posthumanism seeks to tread a path between the apocalyptic and hyperbolic extremes of popular posthumanisms.

Modalities of Posthumanism

Working through the advent of posthumanism it is possible to identify four different modalities in which it has been received and discussed. These could be termed hyperbolic, apocalyptic, deconstructive, and vitalist posthumanisms. The first two represent different modes of popular posthumanism – contrasting discourses circulating in public debate. The latter two account for two different ways in which posthumanism has been debated and developed in social theory and social science. This article discusses each in turn, making reference where appropriate to their reception in human geography, before identifying some key themes.

Hyperbolic Posthumanism

This is the posthumanism of technocratic optimists and some science fiction writers. It underpins the research agendas of many biotechnology and technoscience corporations working around issues such as human enhancement and genetic engineering and informs the philosophy of the Extropians, who advocate a proactionary principle toward scientific innovation. This modality of posthumanism celebrates the arrival of a new epoch in which humans, technologies, and animals can be morphed together into novel collections of hybrid forms. It is full of hyperbole about the potential of scientific developments and technology to enhance the human condition and to forge new social and spatial formations.

Hyperbolic posthumanists argue that we are moving into a new stage in the evolution of life. Here, through human agency we can take control of our own destinies and of those around us, rearranging molecular structures to produce new and enhanced posthumans or downloading consciousness into artificial life forms. This hyperbolic position also informs the discourses of techno optimists in relation to the problems posed by climate change, resource scarcities, and biodiversity loss. Putting their faith in modern science they argue that we will be able to engineer our way out of peril, (re)creating new life forms, synthesizing food stuffs, sequestering emissions, or – if the worst comes to the worst – emigrating to other planets.

In this popular posthumanism the future is a space for the realization of a narrowly defined individuality and posthuman futures shape up in an apolitical world of free play and unbridled opportunity. The political naivety that underpins this approach is founded on a (neo)liberal conception of the sovereign individual and advances a fundamentally capitalist modality that has visions of the posthuman condition opening up vast new arenas for investment and economic growth. Culturally, this liberal position bleeds into popular postmodernist science fiction, which highlights the possibilities for the formation of new individual identities through deliberate surgical, pharmacological, and technological interventions. Ultimately however, these liberal (supposedly) postmodern accounts only deviate slightly from a thoroughly modern imperative that seeks to harness the potential of posthuman developments for emancipatory and/or profitable ends.

Apocalyptic Posthumanism

This is the response to the posthuman condition of many deep green environmentalists, religious and social conservatives, and defenders of traditional liberal humanism. It includes the work of popular writers such as Bill McKibben, Francis Fukuyama, and Jurgen Habermas. With differing degrees of alarm and vitriol they shout enough at the apocalyptic changes they see being wrought by capitalist modernity. Environmentalists bemoan the end of nature and challenge the ethics and necessity of genetic engineering and nanotechnology – innovations they argue are out of sync with natural space times. Similarly, social conservatives opposed to stem cell research and human genetic modification, for example, feel that such developments threaten the very essences of humanity and nature. They seek to reign in technoscience, harking back to a more stable and binary models of human–nonhuman interaction – where humans were human and nature was natural.

Like hyperbolic posthumanism, apocalyptic post humanism is a popular position whose advocates claim to have diagnosed the arrival of a new epoch. However, here humans have crossed some dangerous critical threshold. In contrast to the former's optimism, for the harbingers of the posthuman apocalypse the arrival of this new condition is fraught with danger. The solution is not to plough on but to retreat. This results in a reactionary and either religious or neohumanist politics of fixed identities framed around an eschatological tra jectory of decline and doom. This politics reaches back for an idealized time and a naturalized set of sociospatial and socio environmental relations, often through a pro cess that precludes democratic deliberation.

In this way apocalyptic posthumanism reinvigorates narrow models of the human and nature – purified through a dualistic ontology – and advocates policies toward their protection or reintroduction. It is thus grounded in a static geography and temporality of transcendent forms, outside of social and ecological history, or at least evolving at geological timescales. Identifying these singular ideal types is always prob lematic and is generally reliant on the selective use of statistics, normal curves, and standard deviations, which efface the differences on the margins.

Deconstructive Posthumanism

This is the posthumanism of a set of post structuralist, feminist, postcolonial, and queer theory scholars who draw on a diverse collection of philosophical resources to deconstruct the single figures of the human and the nonhuman at the heart of humanism and humanist epistemology. They seek to draw attention to the vio lence that results from an adherence to such identity thinking. This body of work is informed by the phil osophies of Michel Foucault, Jacques Derrida, Judith Butler, and Giorgio Agamben, among many others.

While humanism argued for the existence of a core and authentic sense of human subjectivity and con sciousness that could and should be shared by all people, deconstructive posthumanists argue that all subject pos itions stem from and are related to the social matrix in which we find ourselves. In its most radical manifest ations this critique is antihumanist and argues that there is nothing universal or natural about any singular model of human identity. Post structuralists like Foucault have deconstructed the different ways in which subject positions are created and enforced through the social distribution of discourses. They sketched out theoretical resources to allow people to decode, transgress, and resist subjectification and to forge new identities.

Instead of fixed identities deconstructive post humanists emphasize cultural difference – cut along in finitesimal axes including gender, ethnicity, age, sexuality, (dis)ability, and a host of more finely grained categories. Others draw on psychoanalysis to explore how identities are often formed through a binary process of othering. Here the positive characteristics of the same are defined in relation to a negative projection of the other. This recognition of the widespread abjection toward difference opens analysis to human desires and anxieties outwith the understandings of human agency and consciousness at the heart of humanism. Here the subject is at times unknowing and unknowable and is always defined in relation to another in situations that vary by geographical context.

Social theorists have traced the incidence of a range of self–other dualisms including man/woman, western/oriental, and gay/straight. Perhaps the most important manifestation of this tendency in relation to post humanism is the humanist dualism that establishes the fundamental division between the human and the non human (or inhuman). Giorgio Agamben terms this an anthropological machine which works to separate the two both within the human – to establish the unique human conscious subject, divorced from its animal body – and on the margins of the human – to create a lumpy class of animals. Writing in the shadow of the Holocaust, Agamben documents how these divisions bleed together through the tendency to elide the human other with the inhuman or to relegate it to a subhuman status alongside or even subservient to the nonhuman.

This model of deconstructive posthumanism was readily appropriated by a new and reinvigorated cultural geography in the 1990s. Alongside their feminist col leagues, cultural geographers used textual and qualitative methodologies to map the moral geographies of cultural difference and normalization. In this linguistic turn they explored how identities are territorialized and deterri torialized through discursive and signifying practices, which spatialized a self/other binary. They sought to give literal and theoretical space to a radical postmodern politics of difference based around the creation and expression of hybrid and fluid identities. Empirically, this body of work has been exceedingly heterogeneous and has encompassed arenas varying from food to music and from natural history to housing. This challenge by cultural geographers to degrading self/other dualisms continues the humanist critique of the singular figure of rational man, which lies at the heart of most modern economic and political theory.

Deconstructive posthumanism also challenged the narrow epistemology of humanist social science and opened a complex and contested politics of knowledge. Multiple and splintered subject positions allowed for the expression of heterogeneous vernacular epistemologies. The epistemological challenge posed by deconstructive posthumanists was also developed in the Sociology of Scientific Knowledge (SSK) by authors such as David Bloor and Bruno Latour, whose work was to have a significant influence on human geography. Cultural geographers' dabblings with vernacular epistemologies raised the specter of epistemological relativism and provoked alarm among the physical and environmental scientists with whom they share their discipline.

This deconstructive ethos has also been extended out from a concern with the contents of the human category to examine what happens to those on the margins – the nonhumans denigrated by the anthropological machine – and critical animal geographers have examined animals as another other. Drawing on Derrida, who highlighted the tendency to lump together the great diversity of the nonhuman realm under the undifferentiated label animal, they have begun to map and document the animal spaces of modern life. This move has brought animal geography into contact with ongoing work on animal rights and animal welfare along with wider currents in environ mental ethics.

Deconstructive posthumanism has had an important influence on human geography and has broadened and deepened the critical focus of humanist approaches by opening analytical space for multiple subject positions and forms of spatial connection and performance, out with the narrow confines of a narrowly defined human subject. However, and in reaction to the narrow grounds of humanism, deconstructive posthumanism has re mained anti ontological in its theorizing. With its preoccupations with textual representation and the for mation of identities it has been unwilling to discuss the material basis to human and nonhuman life. The lively agencies of bodies, technologies, and places are glaringly absent from these accounts.

Vitalist Posthumanism

This is the posthumanism of a growing collection of philosophers, feminist theorists, and nonrepresentational and more than human geographers, who together pick up on the critical impulse of deconstructive post humanism but seek to ground their analyses of the human condition in a nonessentialist, vitalist ontology. While deconstructive posthumanism aims its critical energies at the figures of the human and the nonhuman, vitalist posthumanism addresses the ontological politics of their making and remaking. This final strand to post humanism is informed by an engagement with the vitalist philosophies of figures such as Gilles Deleuze, Bruno Latour, and Isabelle Stengers and their reworkings of antecedents such as Spinoza, Bergson, and Whitehead. This engagement underpins a material turn in geography and the social sciences, which moves off from the linguistic turn associated with deconstruction.

Some of the earliest work in this area emerged from science studies and the innovative work of Bruno Latour and his fellow actor network theorists. Latour and his colleagues argued that we have never been modern and challenged the ontological hygiene that accompanies the modernist dualisms of nature–society (human–nonhuman), subject–object, and the mind–the body. In place of these purifying dualisms they drew attention to the vital agencies of nonhuman technologies and the role of embodied skills in the performance of everyday life. Latour was particularly interested in the processes of representation. He took forward the epistemological agnosticism of deconstructive posthumanists, but through a series of ethnographic studies, attended to the embodied practices and skilled negotiations through which scientists translate the world into representations. His work conveys the reality of scientific practice and documents the important agencies of scientific instru ments and the skilled bodies of scientists themselves in producing and maintaining representations. These are often effaced from their final published accounts.

Latour's concern with the material grounds to everyday life echoes an interwoven strand of vitalist posthumanism which draws attention to the creative role of the human body in making sense of space and pushing back against different modes of subjectification. This concern with the body has been developed in different ways by feminist and postcolonial scholars and those in nonrepresentational geography. Together, these thinkers have returned to and reworked earlier concerns with the body as a sense making vessel, as expressed in the phe nomenology of Merleau Ponty and Heidegger. To this strand they have added insights from the lively ontology of Giles Deleuze. In contrast to the static, certain, and transcendent archetypes of humanism, Deleuze argued for an ontology of intensities, producing forms that are in a constant and unstable state of becoming. Deleuze eschewed identities and dualisms for a world of differ ence and immanent differentiation. His difference is not defined by the relationship between self/other or human/nonhuman; it is infinitesimal and emergent.

A growing collection of nonrepresentational geog raphers have engaged in such theoretical alchemy and have developed concepts from theorists of practice and ethology to explore the embodied, affective, and nondiscursive basis to much social life. Inspired largely by the work of Nigel Thrift, nonrepresentational geog raphers are critical of the linguistic excesses and anti ontological stance of much new cultural geography, where identities, difference, and meaning are reduced to social constructions. Instead, they seek to document the skills, senses, and emotions that underpin human social interactions and examine the ways in which these act according to and depart from wider attempts to prescribe identity. Taking a performative understanding of social life they draw attention to the vast array of precognitive forms of knowledge that frame our engagement with the world and the open ended and contingent nature of any interaction. Epistemologically, they trace how the pro cesses of representation involve learning to be affected by the material world.

This work has informed and interwoven with a bur geoning strand of emotional geography, which develops the concerns of humanistic geography with subjective experience and broadens earlier engagements between geography and psychoanalysis. The rise of emotional geography can also be understood as part of a wider interest in emotion in political science, neurobiology, philosophy, and cultural studies. Emotional geographers have sought to look beyond individual, internalized experiences of emotion to map broader social and relational landscapes of affect. Nonrepresentational geographers differentiate between emotion and affect, where emotion is understood as the cognitive experience of the individual subject of a particular state – anger, joy, envy, etc. – while affect describes prediscursive, embodied sensations, yet to be encoded into emotions. Affect provides an alternative mode of shared intelli gence to the material world that operates outwith the individual subject. Posthumanist geographies of affect therefore take forward psychoanalytical critiques to get at the more than/other than rational energies that underpin social life, operating beyond the around the individual, conscious subject of humanistic approaches.

Feminist scholars, like Elizabeth Grosz, have taken on this concern with the visceral and contingent nature of social life and have explored what the material specifi cities of the female body bring to our understandings of social order and control. In contrast to earlier under standings of race and gender as social constructions, these corporeal feminists have sought to reontologize such categories and to engage with the ontological politics of their performance and contestation. This approach moves off from an understanding of the body as a passive con tainer of social identity to the body as a living assemblage of biological materials and processes. They have traced the material consequences of efforts to cut this non essentialist world of differentiating entities into neat identities and categories that prescribe geographies, power relations, and modes of behavior. Similar lines of enquiry have recently been pursued by postcolonial geographers who acknowledge the materiality of race and aim to reontologize racial difference along nonessentialist lines.

Moving out from the vitalism of the human body, more than human geographers have acknowledged the lively agencies and hybrid ontologies of the nonhuman realm and have begun to construct accounts of human–nonhuman interaction which do not privilege human agency and consciousness. This endeavor has been strongly influenced by the work of Deleuze and Latour and seeks to develop a critical posthumanist politics for our contemporary condition that avoids the twin poles of complacent hyperbole and reactionary nostalgia. At present, it has two divergent strands. The first has grown out of Marxist political ecology and employs a weak model of dialectics to explore the metabolic interactions of nature and society under capitalism and the production of hybrid natures under different social conditions. The second model is more firmly tied to the immanent ontology and ethology of Deleuze and cir cumvents the scalar politics and binaries of the nature–society dialectic. Instead, it emphasizes the contingencies and interminglings that occur between human and non human life and highlights the difficulties and dangers that accompany the construction of meta narratives to explain human–nonhuman interaction. Debate is ongoing between these two approaches as to which best explains our contemporary condition and provides the optimum mode of intervention.

Geographers have also traced the important biopoli tical role of material technologies in governing everyday life. They have documented how technologies sort, classify, and order behavior and provide the often unconsidered background to normal daily practices. On the material boundaries of the human–nonhuman, more than human geographers have explored the inter corporeal interminglings of different bodies in the visceral arenas of food networks, transgenic organisms, and viral epidemiology. This work has highlighted the permeability of the human body and its material dependence on and susceptibility to the nonhuman environment. These embodied geographies sketch out a lively and promiscuous ontology that undermines the solid units of the human–nonhuman binary.

Animal geographers have also begun to engage with this vitalist ontology to explore the beastly places of nonhuman lives, in and among the geographies of human habitation. They have mapped a series of networked geographies and lively topologies that unsettle simple modern mappings of nature – like urban rural, the nature reserve, the nation state, and the sanctity of human body. This engagement has been explicitly ethical and has explored the potential mismatch between the humanist discourse of species identity and the immanent potential within any organism and ecological complex for differ entiation. Attending to this nonhuman difference, vitalist posthumanists have outlined the parameters of a relational or ethological ethics that avoids the anthro pocentrism of much animal rights discourse and the utilitarianism of most modern regimes of environmental governance.

Various authors have also sought to push forward the positive political potential of this vitalist ontology to rethink difference and to forge a critical posthumanist politics. Perhaps the most famous examples of this type of endeavor can be found in the work of Donna Haraway, who has given us the fertile concepts of the cyborg and the companion species as metaphors for forging a more than human politics. A similar task can be found in the ongoing work of Isabelle Stengers, whose hybrid and immanent 'cosmopolitics' is starting to attract the atten tion of several cultural geographers as a way of forging a more than human politics and epistemology. The political challenge this poses is discussed in greater detail below.

Posthuman Themes

This brief overview of the four different modalities of posthumanism gives a sketch of the popular and theo retical landscapes encompassed by the term. The next section to this article highlights five important themes that run through the recent interest in posthumanism and frame the key debates amongst its advocates. These relate to the 'ontologies', 'epistemologies', 'methodologies', 'ethics', and 'politics' of posthumanism and posthumanist geographies.

Posthumanist Ontologies

There are a range of different ontological positions in play in the four different modalities of contemporary posthumanism outlined above. Deconstructive post humanists are avowedly anti ontological in their work. They contest all claims to truth that relate to an essential human or nonhuman nature and are reluctant to ground their interventions; their identities are constructed in the heady realms of discourse and representation amid swirling networks of language and signification. In con trast, both hyperbolic and apocalyptic posthumanism are grounded in a naïve epochal ontology that diagnoses the demise of a pure state in which nature and society were separate, and the arrival of a new posthuman condition, which is characterized by proliferating hybrids. They differ in their judgment of this millennial event but their ontology is consistent. It is the final vitalist modality that advances the most coherent ontology of the four and it is this that has stimulated the greatest interest among posthumanist philosophers and geographers. The ontology of vitalist posthumanism challenges both deconstructive and apocalyptic positions and seeks to harness and guide the hyperbolic excesses of unbridled modern technoscience.

Latour's amodern ontology, in which the world is comprised of hybrids, quasi objects, and quasi subjects, historicizes the epochal imaginings of both apocalyptic and hyperbolic posthumanists and contests their claims that our contemporary condition comprises a new ontology of blurred boundaries. Latour argues that these boundaries have always been blurred. While the con temporary blurrings might be novel there is no such thing as a pure human or a pure nonhuman, there never has been and there never will be. Deleuze's ontology of immanence takes these critiques a step further to chal lenge the transcendent forms of nature and natural bodies beloved of conservative apocalypse harbingers. Like Marx, Deleuze argues that nothing is static, everything is a product of historical conditions, and all that is solid can melt into air. Such an understanding undermines any politics of nostalgic retreat that uses objective science to naturalize any past form or set of social relations.

The ontology of vitalist posthumanism thus takes us beyond the anti ontological stance of deconstructive posthumanism without lapsing into the essentialist for mations of which the latter is so critical. However, the flat ontology of Latour and other actor network theorists has been criticized by geographers and other social scientists for its generalized symmetry. This reduces humans and nonhumans to the same ontological condition, and leaves theorists unable to differentiate specific human com petencies, like the ability to reflect and react to classifi cation. Critical theorists argue that this forgoes too many of the explanatory and political resources of humanism.

In place of Latour's flat ontology, Deleuze presents a nonessentialist ontology of difference. He draws on Spinoza and ethology to outline a relational ontology of affects. Affects are the material properties and forces of a body that condition the nature of any interaction in a particular spatial context. These affects are not immut able essences but are immanent from the assemblage in which a body is encountered and are open to the play of future differentiation. With this emphasis on the imma nent potential of any body to become otherwise, Deleuze maintains the postmodern focus on difference, without reducing difference to the play of textual signifiers. Nonrepresentational geographers argue that he thus outlines a relational ontology that helps move beyond the realism–relativism impasse that plagued cultural geog raphy and the sociology of science in the 1990s. They have developed Deleuze's ontology in the context of their work on embodiment, affect, skill, and intelligence.

In many ways it is possible to map overlaps between the exuberance for immanence of some vitalist post humanists and the celebration of innovation among hyperbolic posthumanists. Although the former see the contemporary condition as a radical departure, both relish the experimental and creative ethos of modern technoscience and welcome the possibilities for social reinvention and individual self expression. Indeed, as feminist and Marxist critics of Deleuze have pointed out, his groundless ontology struggles to provide criteria for differentiating between different becomings – between, for example, the genetic screening of human embryos and the emergence of virulent zoonoses. On occasions the radical defense of immanence and flow among some vitalist, posthumanist geographers comes close to the complacent hyperbole of some of the more extreme advocates for unbridled technoscience. The political conundrums this poses is discussed below.

Finally, each of these four modalities of posthumanism performs their own spatial ontology, with differing con sequences for posthumanist geographies. Apocalyptic posthumanism implies a geography of retreat – a return to spatial units and imagined pasts with clearly drawn divisions – the wilderness or nature reserve, the purified body or preindustrial society. This modality imagines an impossible regional topology of neat territories. Decon structive posthumanism implies a networked spatial ontology in which human subjects are caught up in textual webs of significance strung between different places and spanning different eras. These networks hold in place distinct regimes of governmentality that create and police territorial identities.

Vitalist posthumanism encourages a livelier and more fluid set of topologies in which place, and the nonhumans who help to make it up, have an active agency in the performance of social life. This happens over many dif ferent intersecting space times, from the geological to the parasitical and weaves together humans and non humans through complex and often unseen foldings and interminglings. The networked or fluid spatial ontologies of posthumanist geographies operate without a strict or hierarchical notion of scale and, in their most extreme manifestations, lead to spatial imaginaries of unbridled openness and absolute free flows and becomings. This geography without scale has been the subject of much recent debate and has been critiqued by political economists and ecologists, who defend scale as a con ceptual tool for political analysis.

Posthumanist Epistemologies

Deconstructive posthumanists like Foucault and Derrida and their acolytes in cultural geography have challenged the epistemology of humanist social science on the grounds that it is founded on an ideological construct – the single human subject, which they argue is an exclusive model that is generally white, male, hetero sexual, and Western. Claims to truth that stem from and relate to this subject efface the vast diversity of other voices and modes of subjectivity that exist in a world of difference. In place of this universalized view from nowhere – or god trick – feminist writers like Donna Haraway and Sandra Harding, argue for an epistemology grounded in situated knowledges. Here the author reflects on the context of their subject position both geographically and in relation to the power relations that surround their claims to representation. This reflexive process foregrounds an author's own standpoint, which they claim develops a stronger form of objectivity than that emerging from an idealized humanist subject.

Cultural geographers have welcomed this postmodern epistemological challenge and have documented an extensive variety of vernacular or folk knowledges and knowledge spaces hitherto marginalized from the Whiggish histories of modern science and geography. Deconstructive posthumanists have also drawn attention to the primacy of the gaze in Western epistemologies. The view from nowhere can be understood as a reflection of the broader dualism established between the mind and the body – where the mind, as the locus of thought, receives a direct and mediated representation of the world through the panoptic gaze. The body is absent from such an epistemology. Developing the method ologies and concerns of humanistic geography, cultural geographers have mapped the consequences of this ocular centric epistemology and have drawn attention to the exclusive ways of seeing it sustains. By and large this work has focused on deconstructing textual representations.

In line with their materialist ontologies, vitalist posthumanists have taken forward the deconstructive critiques of modern (social) scientific epistemologies but have sought to ground them in explorations of the embodied practices of knowing and representation. More than human and nonrepresentational geographers have drawn in particular on ideas from phenomenology, performance studies, and the sociology of science to trace the skilled practices, technologies, and affective energies required to tune in to the world and to construct tenta tive representations of what is witnessed. This produces an alternative, modest, and embodied epistemology based open learning to be affected by the world in partial and situated contexts. It is grounded in the relational ontology of Latour and Deleuze and moves on from the epistemological relativism offered up by Derrida and Foucault that had troubled progressive social scientists and environmentalists.

This vitalist approach has opened analytical space for the consideration of a host of other senses beyond the visual. Nonrepresentational cultural geographers have documented the sensory geographies of taste, touch, smell, and sound, engaging with and witnessing the spaces experienced by differently abled communities.

Out in the field, sociologists of science have explored the embodied interactions that occur between humans and nonhumans on an affective level and have drawn attention to the vernacular expertise performed by lay constituencies. This democratization of expertise broadens the standard epistemologies of natural science and gives political space to hitherto marginalized groups.

Posthumanist Methodologies

The linguistic turn in cultural geography and the subsequent material or performative turn by more than human and nonrepresentational geographers have lead to a significant diversification in the methodologies and methods employed by human geographers. This trend has significantly increased the modes of intervention available and has resulted in a more catholic understanding of what counts as data within the discipline.

Deconstructive posthumanism in cultural geography drew on a range of methods for textual analysis. Historical and contemporary texts – including film and broadcasts – were read against the grain to unravel their latent discourses and disguised power relations. Such qualitative methods were enhanced with other techniques designed to generate talk – and thus text – from research subjects. This has led to a significant increase in the use of interviews and focus groups among geographers. These methodologies involve the transcription and coding of textual data, which is then subjected to similar forms of discourse analysis to those used on published material.

Nonrepresentational geographers are critical of this textual focus, which they argue misses the life of the events being recorded and fails to witness affects, skills, and everyday practices. Instead, they have drawn on and refined a collection of ethnographic, visual, and performative methods to try and get closer to objects and the practices of people in real time. These ethnographic methods involve the geographer acting as an observant participant, experiencing places along with their research subjects through similar modes of embodied engagement. Geographers have employed video and creative writing techniques to witness and record their experiences for later analysis. This type of engaged and affective fieldwork moves toward a participatory model of research, in which the social scientist actively intervenes in the research situation – rather than seeking to remain a detached observer. This model challenges the figure of the academic as impartial and above the fray. Instead, it advocates getting involved in any emerging politics in the research context.

To explore the agency of nonhumans – both organic and inorganic – more than human geographers have borrowed ethnographic methods from science studies and anthropology – in particular from actor network theory. Following the object or tracing the commodity through multisited ethnographies has become a popular method for examining how different nonhumans link together peoples and places and for revealing the hidden connections they perform. Attending to the specific skills and practices required to translate an object into a mobile representation has proved a powerful mechanism for exploring the pragmatics of representation and for unraveling the geographies of power operating through distributed networks.

Posthumanist Ethics

The different modalities of posthumanism outlined in the first section to this article challenge, in their different way, several of the core principles of humanist ethics. In particular, deconstructive and vitalist posthumanisms question the primacy afforded reason and human consciousness in ethical decision making and the binary (human–nonhuman) formation of who or what has ethical status.

More than human geographers have sought to outline a relational ethics that goes beyond the humanist privileging of the human subject as the single container of value. They aim to broaden the ethical franchise in line with the insights of a hybrid and vitalist ontology. The relational ethics they advocate also avoids the moral extensionism of neohumanist animal rights theorizing, which elevates a small subset of species to ethical parity with humans and ignores the nonhuman majority. A relational ethics attends to the specific ecologies and immanent potentialities of an ecological complex and is founded on a nonessentialist, ethological ethics that aims to respect the difference of the nonhuman. To date, the normative articulations of these ethics have (necessarily) remained vague and largely comprise appeals for less violence toward animals and to tread lightly on the earth.

This relational model of ethics therefore claims to avoid the transcendent and static ethical codes advocated by apocalyptic posthumanists. Here the nonhuman realm – as nature – is reduced to an ontologically pure domain to be held *in situ*. This type of ontological conservatism can be found in deep ecology, wilderness movements, and many organizations campaigning to save a specific model of the human. In place of this ethical ontology more than human, relational geographers have argued for a postenvironmentalist ethics that would embrace the contingent and hybrid ontology that characterizes our amodern condition and would avoid the tendency among environment ethicists to fix the nonhuman in time and place.

Moving off from the humanist understanding of rationalism as the imperative for ethical action, nonrepresentational and emotional geographers have also explored the affective underpinnings to ethical action.

This more than rational approach to ethics develops the ideas of postmodern ethicists, such as Zygmunt Bauman, who argue after Foucault that ethics requires a sensibility as well as a code. It is this sensibility that impels people to act in relation to a code and an ethical sensibility rarely stems from the act of rational contemplation. Instead, an ethical sensibility emerges from the practices and affections that constitute everyday life, performed in the material world in a noncontemplative fashion. This precognitive basis produces an ethological ethics that is grounded in the messy and lively ontology of our posthuman condition. Geographers and social theorists have drawn attention to the importance of enchantment, wonder, and curiosity as the foundation for a generous disposition toward the nonhuman realm. They are also starting to develop the ethical implications of abjection, phobias, fears, and more 'negative' affective responses.

Posthumanist Politics

The popular posthumanisms of hyperbole and apocalypse represent two different extremes along a spectrum of political responses to the contemporary posthuman condition. In apocalyptic posthumanism we are offered a politics of retreat, a nostalgic return to a mythical and pure natural past or an essentialized human archetype. These fixed forms are often determined by experts through an autocratic model of science that shortcuts democratic politics. In contrast hyperbolic posthumanism envisages a realm beyond politics in which individual, self interested actors can seek transcendence through promiscuous interminglings. This complacent posthumanism offers few critical resources to address inequities in the development of technoscience or to manage the ecological and physiological risks that accompany sociomaterial innovations.

In their different way, deconstructive and vitalist posthumanism offer two very different models of politics, which try and thread a path between these two extremes. Neither of them is explicitly antimodern but they do seek to reign and guide the excesses of complacent posthumanism. Deconstructive posthumanism has largely concerned itself with challenging the political formations and modes of governance that are associated with modern thought. Its advocates have been less interested in advocating a positive politics and those who have taken on their ideas drift toward a model of antistatist anarchism or a form of postmodern identity politics that widens the types of difference recognized by progressive politicians, out from traditional concerns with socioeconomic class and gender.

Vitalist posthumanism seeks to ground its politics in the stuff of life in a nonessentialist and immanent fashion. It argues that we cannot start at a fixed beginning – the human/nature – but instead must begin in the middle, in among the processes of becoming. What is human and what is natural is always the effect, rather than the grounds for politics; forms and identities emerge from the interaction of lots of different actants, rather than providing the starting point for these interactions. Accordingly, vitalist posthumanists argue that we need a politics that does not seek to conserve existing beings but to produce new forms. This is an immanent model of politics that contests appeals to natural forces and the unchanging existence of political economic structures. Instead it gives space to flows, becomings, and deterritorializations of regimes of power and subjectification.

The politics of vitalist posthumanism therefore differs markedly from the politics of scale that has dominated much critical geography over the last 20 years. Scalar politics argues that the world is ordered hierarchically, with more extensive global forces and networks interpenetrating with local spaces to determine outcomes. It defends the existence of universal global forces and entities – like capital and capitalism. Instead a Deleuzian or Latourian approach would see these networks as contingent and fragile and, most importantly, continuously performed in a multitude of local settings underdetermined by a scalar hierarchy. Several different models of positive posthumanist political theory exist that are informed by this ontology. These include Donna Haraway's metaphorical use of the cyborg and the companion species, Isabelle Stengers' and Bruno Latour's cosmopolitics, or Hardt and Negri's concept of the multitude. In different ways, geographers are now beginning to develop and spatialize these models.

Marxist and feminist critics have argued that the radical emphasis on immanence and flexibility in the work of Deleuze and Guattari and those who have developed their ideas veers too far toward a hyperbolic and complacent posthumanism. They posit that the monstrous couplings envisaged by Deleuze and Guattari and the forms of subjectivity they enable too closely approximate those posthumans being created by capitalist technoscience. In celebrating immanence and fluidity orientated around a flat ontology, Marxist critics argue that vitalist posthumanist struggles to maintain a critical space from the imperatives of global capitalism for flexibility and creative destruction. Similarly, feminist critiques of hyperbolic posthumanism argue that this is masculine fantasy of wild spaces, new frontiers, and sublime encounters.

This tension between fixity and fluidity and the relative merits of scalar analysis has provoked much debate within the discipline. This centers on the challenge for a vitalist approach to move off from the certainties of both political economy and neoliberalism and to forge a critical, material posthumanism that can thread a path through the reactionary desire for fixity and the complacent celebration of fluidity.

Posthumanism and 'Human' Geography

The rise of a posthumanist geography within the sub discipline of human geography raises obvious questions about the suitability and the future of the human–phys ical divide within the discipline. In recent years, the ontological hygiene implied by this division has come under critical scrutiny from a host of geographers seeking some form of interdisciplinary (or intradisciplinary) rapprochement. Posthumanism – especially in its vitalist modality that appreciates the liveliness of the nonhuman world – can be understood as one strand of this broader questioning. Indeed more than human geographers have begun to collaborate with natural scientists – both geographers and others – to develop new methodologies to allow interdisciplinary collaboration.

In some ways this could be seen as a return to an older model of cultural geography, as practiced by the Berkeley School, where the materialities of cultures and natures were considered in a more symmetrical fashion. How ever, this move toward a reunion also aims to forge a more critical, deliberative, and democratic model of sci ence than that envisaged by Carl Sauer. Here, geog raphers like Sarah Whatmore have drawn on the work of Isabelle Stengers, who outlines a methodology for interdisciplinary collaboration and the redistribution of expertise in environmental decision making. Stengers advocates the formation of competency groups for deliberation in matters that concern social and natural scientists, organic nonhumans, and different public constituencies.

Conclusions: The Latent Humanism in Posthumanism

On many levels posthumanism is therefore something of a misnomer. In ontological terms we have never been human, in the strict sense of the term, and cannot now be beyond something which cannot and did not exist. Popular claims for the passing of the human era and arrival of a new posthuman condition are misplaced and hyperbolic. Boundaries between the human and the nonhuman have been crossed throughout the entire history of the human species. However, perhaps popular posthumanism is actually more concerned that the pace of these ontological interminglings has been stepped up, that we have entered into a new temporality of imma nence and hybridization that is inherently both more risky and more promising. It would be fair to say that the current temporality of innovation, interspecies exchange, and environmental degradation are unprecedented and therefore demand critical attention.

It is here that the mis appellation of those activities grouped under the labels deconstructive and vitalist posthumanism is perhaps most clear. In many ways their concern with the failings of humanism and with the character of emerging posthuman futures situates their activities firmly within the tradition of progressive critical theory. Even as they seek to move away from their humanist predecessors, deconstructive and vitalist posthumanists remain firmly tied to the Enlightenment ethos of critical enquiry. To be sure, the human is now much less assured, much more caught within the play of forces and the discursive networks of other actants. It is also much less self evident to itself (the psychoanalytic critique), but the point about the collision of the terms 'post' and 'human(istic)' is not that the latter – the human, and indeed the humanistic philosophies which feasted upon it – are so much entirely erased as thor oughly problematized (while still leaving flickering traces in what remains). Posthumanism and humanism remain thoroughly interwoven; even as the Human (sing) leaves, critically reflexive humans (pl.) are not going to disappear anytime soon.

So what are we post? What does it mean to be post human? Is posthumanism just another faddish post 'ism' or does it make a valuable contribution to the geo graphical enquiry? In many ways these are early days and the answers to these questions are still being worked through; the jury is still out. However, it is clear from this brief overview that posthumanist theories have posed a strong challenge to the unitary figure of the human in humanism and to the boundary work this figure does in defining the nonhuman. Posthumanist geographers are in the vanguard in forging a critical mode of post humanist enquiry that provides fresh perspectives on established themes of geographical research. A critical posthumanist geography is well placed to make timely interventions into contemporary hot topics as the char acter and consequences of the current molten moment work themselves through.

See also: Actor-Network Theory/Network Geographies; Affect; Becoming; Dwelling; Fluidity-Fixity; Humanism/Humanistic Geography; Human-Nonhuman; Interdisciplinarity; Material, The; Nature; Nature-Culture; Non-Representational Theory/Non-Representational Geographies; Other/Otherness; Postcolonialism/Postcolonial Geographies; Postmodernism/Postmodern Geography; Poststructuralism/Poststructuralist Geographies; Psychoanalytic Theory/Psychoanalytic Geographies; Resource and Environmental Economics.

Further Reading

Agamben, G. (2004). *The Open: Man and Animal*. Stanford: Stanford University Press.

Badmington, N. (2000). *Posthumanism*. Basingstoke: Palgrave Macmillan.

Bennett, J. (2001). *The Enchantment of Modern Life: Attachments, Crossings, and Ethics*. Princeton, NJ: Princeton University Press.
Castree, N. and Nash, C. (2004). Mapping posthumanism: An exchange. *Environment and Planning A* 36, 1341 1363.
Deleuze, G. and Guattari, F. (1987). *A Thousand Plateaus: Capitalism and Schizophrenia*. Minneapolis, MN: University of Minnesota Press.
Derrida, J. (2003). And say the animal responded. In Wolfe, C. (ed.) *Zoontologies: The Question of the Animal*, pp 121 146. Minneapolis, MN: University of Minnesota Press.
Fukuyama, F. (2003). *Our Posthuman Future: Consequences of the Biotechnology Revolution*. London: Profile.
Grosz, E. (2004). *The Nick of Time: Politics, Evolution, and the Untimely*. Durham, NC: Duke University Press.
Haraway, D. (1991). *Simians, Cyborgs and Women: The Reinvention of Nature*. New York: Routledge.
Hayles, N. K. (1999). *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature and Informatics*. Chicago, IL: University of Chicago Press.
Ingold, T. (2000). *The Perception of the Environment: Essays on Livelihood, Dwelling and Skill*. London: Routledge.
Latour, B. (1993). *We Have Never Been Modern*. Cambridge, MA: Harvard University Press.
Marston, S., Jones, J. P. III. and Woodward, K. (2005). Human geography without scale. *Transactions of the Institute of British Geographers* 30, 416 432.
McKibben, B. (2004). *Enough: Genetic Engineering and the End of Human Nature*. London: Bloomsbury.
Philo, C. and Wilbert, C. (2000). *Animal Spaces, Beastly Places: New Geographies of Human Animal Relations*. London: Routledge.
Thrift, N. (2005). From born to made: Technology, biology, space. *Transactions of the Institute of British Geographers* 30, 463 476.
Whatmore, S. (2002). *Hybrid Geographies: Natures, Cultures, Spaces*. London: Sage.
Wolfe, C. (2003). *Animal Rites: American Culture, the Discourse of Species and Posthumanism*. Chicago, IL: University of Chicago Press.

Relevant Websites

http://www.posthumanism.org.uk/
Ethics, technology, culture.
http://www.nature.com/
Nature.
http://www.opendemocracy.net/
Open democracy, free thinking for the world.
http://www.prospect magazine.co.uk/
Prospect Magazine.

Postmodern City

S. Warren, Eastern Washington University, Cheney, WA, USA

Glossary

Deconstructionism A philosophical approach to knowledge connected with postmodernism that attempts to interpret meaning by stripping away layers of embedded assumptions, biases, and ideas in order to understand the ways in which meaning is socially, historically, and contextually constructed whether the object is a text or a building.

Disneyfication The appearance of themed architecture, sanitized leisure spaces, the privatization of public space, and other theme park-like elements in urban space; usually used in a derogatory sense.

Epistemology A branch of philosophy concerned with systems of knowledge, and how knowledge is formed and validated.

Neologism The coining of a new word or phrase to express a concept that cannot be adequately represented with existing language.

New Urbanism Postmodern urban planning and design movement whose goal is the creation of neighborhood-oriented urban communities based on appealing and livable architecture.

Post-Fordism An economic formation associated with the postmodern era, in which capital moves beyond the mass production, assembly-line model of Henry Ford and instead shifts to a model of consumer-driven flexible accumulation where labor and production are globalized.

Postmodernism A catch-all term that has been applied to literary criticism, art and architecture, and the social sciences to describe theory and practice that actively rejects the scientific certainty and rationality of modernism, and instead engages linguistic and cultural relativism, subjectivity, and heterogeneity, and in some cases links them to underlying social and economic transformation.

Post-Structuralism A variety of postmodern methods and theories such as deconstructionism that approach knowledge from the philosophical standpoint that meaning is socially constructed and always shifting; a denial of the binary opposition of structuralism.

Definitions

The postmodern city as a concept virtually defies a single definition. Since the time of its entry into common urban lexicon, there have arisen so many competing definitions of the term 'postmodern' it has been likened to an intellectual Rorschach test. The term appears to invite debate regarding its actual historical moment, its cultural milieu, and its politics. It is impossible to pin down historically, with various authorities claiming its beginnings anywhere from the late nineteenth century to the 1970s (the phrase postmodern itself dates back to 1870, and was used in an architectural context as early as 1945). It is variously treated as an architectural esthetic, an urban social condition, an epistemological stance, and an intellectual subculture. It has been asso ciated, albeit in different ways, with political stances on both the left and the right. Some of the most prominent figures held up as superlative examples of postmodernist work steadfastly refuse to be publicly associated with that term. Given this intellectual cacophony, the term 'postmodern city' can best be approached as a collective (but not necessarily cohesive) attempt to define and shape the city that emerges out of twentieth century modernism.

Introduction

Beginning in the 1970s and early 1980s, a new type of architectural style began to appear sporadically in the skyline of North American and European cities. Sometimes the distinguishing features were subtle, such as the Chippendale flourish atop the otherwise modernist AT&T building in New York City, and sometimes the features were boldly unique, such as Frank Gehry's irregularly angular house in Santa Monica, clothed in rough asphalt and chain link fencing. Such buildings attracted the attention of architectural critics who declared them postmodern, that is, conscious efforts to reject the sleek and coldly efficient tenets of architectural modernism. At the same time, urban critics noted a sea change in the broader cultural, social, and economic conditions of the city as indus trialism waned, inner city poverty intensified, service sector oriented leisure and consumption prevailed, and new global economic relations underpinned the urban fabric. These critics also identified these changes with the term 'postmodern' and incorporated related philo sophical developments such as post structuralism and deconstructionism into their theoretical understanding of these changes. Thus, by the mid 1980s, the post modern city had been rhetorically constructed as an architectural style, an urban condition, and a theoretical perspective.

Emergence of Postmodern Perspectives on the City: Questioning the Modern

One consistent thread across all postmodern urban thought is a conscious distancing from modernism. While some postulate a dramatic reworking of modernist tendencies and others see a complete break from them, all agree that the signature characteristics of the modern city – an efficient machine for living crisscrossed by superhighways and planned from a rational scientific standpoint – no longer apply to contemporary urban (and suburban) space. Modernism, like postmodernism, also is a multifaceted construct, and from the postmodern perspective its underlying tenets come under attack from all sides. Postmodern architects question modernism's definition of beauty as stark functionalism, as epitomized by Mies van der Rohe's famous statement "less is more." Postmodern planners question the top down elitist process associated with the development of the twentieth century city illustrated by figures such as Robert Moses. Postmodern geographers set their gaze upon the underlying theories used to interpret the modern city, questioning a legacy of modernist geographic theory from positivist scientific approaches popularized in the 1960s back to the urban ecology introduced by the Chicago School in the 1920s. Some, following the lead of European critical social theorists, reach back to modernist theories as they were developed in the hands of Marx, Weber, and others in the nineteenth century to remap a theoretical path that explicitly theorizes the spatial. The work of French theorists such as Jean Baudrillard, Michel Foucault, and especially Henri Lefebvre has been particularly influential in the development of postmodern geographic thought.

While all postmodern perspectives may share the same motivation to uproot assumptions about modern urban space, the different theoretical and disciplinary starting points shape divergent interpretations of the nature of postmodern urban space. Two distinct conceptual poles have emerged: one that celebrates a playful, pedestrian friendly, and historically sensitive landscape where healthy communities are nurtured; and the other, its diametric opposite, a sinister hyperspace that is the active embodiment of the military industrial complex of late capitalism/globalization. The former is more often associated with postmodern architecture and planning, while the latter is more common in postmodern urban geography. Some of the more provocative interpretations are situated somewhere between the two poles, investigating the hegemonic possibilities of postmodern urbanity's complex and contradictory impulses.

Building, Planning, and Understanding the Postmodern City

However one chooses to label contemporary urban forms, they are actively being designed and built by architects who align themselves with postmodern styles, planned by urban planners who embrace a postmodern communicative model of development (and sometimes even before the dust has settled), interpreted by geographers and urban theorists who invoke the phrase postmodern to explain what they see. The postmodern city has extraordinary material presence that borders on visual branding – specific buildings and entire cities are indelibly associated with the term. Architects, planners, geographers, and other urban theorists each play a different role in constructing the powerful imagery, as well as the concrete landscapes, of the postmodern city.

Designing the Postmodern City: Postmodernism as Architectural Style

Some of the earliest references to the postmodern city are made in connection with specific architects and their designs. By the time Philip Johnson had completed the aforementioned AT&T Building in 1984 (**Figure 1**), perhaps the most famous postmodern commission of its time, several prototypical works insistently announced

Figure 1 Philip Johnson, AT&T Building (now Sony Building), New York City, 1984.

Figure 2 Charles Moore, Piazza d'Italia, New Orleans, LA, 1978.

the arrival of postmodern styles in the city and set the parameters of debate. Designs such as Charles Moore's Piazza d'Italia in 1978 (**Figure 2**) with its shimmering juxtaposition of free standing classical columns and colorful neon lights in a decaying inner city neighbor hood, and Michael Graves' playful Portland Building in 1982 that housed government workers within a 15 story structure adorned with what appeared to be giant holiday ribbons, introduced architectural elements that brazenly rejected the modern. They introduced symbolism and decoration, historical referents, pedestrian scale, and humor. Postmodern architecture was also linked with tourism and consumption early on; James Rouse's festival marketplaces arose out of malingering industrial sites to combine decorative historical charm with cleverly marketed shopping opportunities, and epitomized (some might say obscured) underlying postmodern land use transitions. Seminal texts such as Robert Venturi, Denise Scott Brown, and Steven Izenour's 1972 *Learning from Las Vegas* and Charles Jenck's 1971 *Language of Post modern Architecture* helped formalize and popularize the move ment. In 1985, the Disney Company chose to engage well known postmodern architects to carry out ambitious expansion efforts on their Florida and California prop erties, and, a few years later, in France. As a result, high profile commissions were awarded to numerous postmodern architects, ensuring that literally millions of tourists each year would now visit some of the most representative postmodern architecture as part of their vacation. Architectural esthetic merged seamlessly with tourist destination; Michael Graves' Swan and Dolphin hotels (with a total of over 2000 rooms), for example, became the nightly endpoint after a long day of play (**Figure 3**); Robert Stern's cartoonish Disney Casting Center, a sight out the left side of the tour bus, and Frank Gehry's Festival Disney in Paris, an ironically symbolic shopping mall for European visitors. Michael Eisner, then the CEO of the Disney Company, referred to his postmodern vision as 'entertainment architecture' a provocative combination of urban functions not lost on urban theorists.

Planning the Postmodern City: Postmodernism as Planning Discourse

The original goals of postmodern architecture – playfully decorative styles, sensitivity to the local, and the fostering of open and engaged community settings – continue to be of great importance to those planning the city. Postmodern planning discourses focus on ways to achieve these goals by looking beyond just one building at a time and instead grappling with urban development at neighborhood and regional scales. Just as architects bemoaned the perceived stranglehold that modernist principles held over their discipline, a growing number of planners felt equally straight jacketed by the preferred scientific rational methods of large scale, bulldozer driven approaches of the day. Jane Jacobs was one of the first to criticize the *status quo* in her scathing 1961 assessment *The Death and Life of American Cities*; other planners noted with some bemusement that while traditional cities like Detroit and Cleveland imploded upon themselves, the artificial townscape of Disneyland (1955) was in fact a far more successful model of a city

Figure 3 Michael Graves, Walt Disney World Swan Resort, Orlando, FL, 1989.

that actually worked. Early Disney proponents such as James Rouse and Peter Blake noted that Disney contributed not just a visually, historically charming architectural lexicon, but a blueprint for an urban infrastructure that efficiently provided power, transportation, sanitation, and somewhat intangibly, happiness to tens of thousands of people at a time. Translating these achievements to the real world has been an important goal for many postmodern planners. Probably the most recognizable effort has been labeled 'new urbanism', a movement that some have noted owes its success as much to sympathetic journalists as to architects and planners. Most strongly associated with the design team of Andres Duany and Elizabeth Plater Zyberk and their works such as Seaside, Florida (1981), new urbanism refers to collective efforts to reintroduce the wisdom of the nineteenth century walking city into twenty first century urban and suburban spaces thought to have been dehumanized by decades of modernist plans. New urbanism brings with it a very tactile agenda – an architecture based on traditional wood and stone building materials, brick lined streets (to slow traffic), front porches (to encourage neighborliness), shade trees for comfort, water features to sooth, and diversity – visual, social, and economic. The neighborhood, not the individual building, remains the key object of design; new urbanists plan to ensure a usable, walking city lined with attractive homes, shops, parks, offices, and schools. As a visual iconography for urban design, new urbanism has attracted considerable attention both within and outside the architectural profession. The professional organization Congress for New Urbanism, founded in 1993, 12 years later claimed over 2000 members in 20 countries and over 200 new urbanist developments in the United States alone. New urbanism has been influential in shaping United States public housing policy: The US Housing and Urban Development (HUD) HOPE VI grant program, also initiated in 1993, incorporated new urbanist design principles into their program guidelines. And by 1994, perhaps somewhat inevitably, new urbanism came full circle to again be linked with the Disney Company. When Disney announced plans for a new urbanist styled new town development Celebration (**Figure 4**) adjacent to their Walt Disney World property in Florida, more than 5000 people lined up to compete in a lottery for just 341 properties.

New urbanist developments, though they may number in the hundreds, represent only isolated pieces of the urban fabric. Other postmodern planners are concerned with developing comprehensive discursive practices that can rewrite the entire planning process itself, and thus affect space at regional and national levels. Alternatively known as the communicative planning model, the collaborative model, or the consensus model, the top priority is the reevaluation of the planner's role in the development process. No longer is the planner seen as the trained scientific expert who informs the public what is best for them. Such a model, it was argued, led to fundamentally undemocratic results, ensuring that those in power maintained their power by replicating a formative mechanism of oppression. Instead, the postmodern planner should be envisioned as a facilitator, a patient listener who hears the stories told by the various stakeholders, distills the points of connection, and helps

Figure 4 Walt Disney Company, Master-Planned Community, Celebration, FL, 1994.

build consensus among the various parties. These theories are predicated upon a body of postmodern work that sees all knowledge as socially constructed, accepts different forms of knowledge as equally valid, and, drawing from Jurgen Habermas' theory of com municative action, posits that consensus achieved through communication can be the foundation for rational action. Following Habermas, the role of language itself is key – it is through talking and listening, contesting or accepting others' claims, that pragmatic mutual understanding is reached. Proponents of the communicative models further note that recent research indicates that even in the heyday of modernist planning, the decision making stakeholders never paid as much attention to so called expert knowledge as was assumed. Postmodern planning approaches, thus, are meant to more realistically model the nonlinear and unpredictable paths by which decisions are reached.

Figure 5 John Portman, Bonaventure Hotel, Los Angeles, 1977.

Theorizing the Postmodern City: Postmodernism as Interpretive Framework

By the 1980s, geographers and other urban observers began to inquire whether the new esthetic emerging in the city was not just a new architectural style or the result of different planning procedures but a break from mod ern urban conditions, and not necessarily a positive one at that. The notion that the postmodern city represents fundamentally different cultural, social, and economic relations has become one of the most active themes in urban research. This version of the postmodern city looks beyond both the material landscape and its related planning processes to ask what economic, social, and cultural relations underlie contemporary urban form. Often the initial motivation to pursue these questions is triggered by specific buildings or places – the new venues are symbolic reminders of the increasing difficulty in navigating the terrain with existing theory (or, for that matter, existing cartography). Buildings like Portman's Bonaventure in Los Angeles (**Figure 5**), cities like Sea side, Florida, or the artificial fabric of Disneyland have inspired urban theorists to interrogate specific places using methods derived from postmodern (post structural) philosophical thought, and to reevaluate the epistemo logical bases of their own urban theories as well. Perhaps no city has been as enigmatic a puzzle, and hence as influential a site, as Los Angeles, and no group of scholars as closely associated with postmodern urban theory than the so called LA school.

The LA school emerged somewhat fitfully in the late 1980s, bringing together an otherwise divergent set of scholars who all shared the conviction that the processes that converged to produce the seemingly unique con tours of Los Angeles were in fact the same processes occurring at national and global scales. Los Angeles has been identified as prototypical of significant trends in cluding economic restructuring, demographic upheaval, the death of public space, the disneyfication of urban space, and environmental degradation at unprecedented levels. A growing collection of postmodern tours guide the reader through Los Angeles' extremes: Fredric Jameson, for example, takes readers through the dizzying privatized hyperspaces of the Bonaventure Hotel in order to illustrate the cultural expression of postmodern capital; Edward Soja places readers in the pilot's seat of an imaginary airplane to fly over the discombobulated metropolis of ethnically, economically, and environmentally balkanized spaces rigidly enclosed by an outer ring of strategically placed military complexes; and Mike Davis gives a postmodern weather report where the reader can experience Los Angeles' imminent destruction from fire, flood, wind, and geological fault lines, often due to the unintended consequences of recent urban development.

The LA school is also unified in the conviction that conventional urban theory cannot explain these new formations, and that profound reconceptualization is needed. Some embrace a critical post Marxist perspective, some a perspective that borrows from other postmodern social theories such as feminism and postcolonialism, and some a perspective more clearly drawn from post structural/deconstructionist philosophies. All rethink the epistemological grounds upon which they are based, and reject totalizing master narratives in favor of systems of socially constructed knowledge based on bringing in alternative voices and divergent perspectives. Often, though not always, these postmodern geographers explore new languages of presentation, from inventing new words (neologisms) to disrupting the expected structure of discourse.

A significant result has been what is referred to as the spatialization of theory; that is, a conscious effort to theorize the powerful role space plays in shaping urban conditions. Working from the assertion that most theory (either modern or postmodern) privileges the way social conditions shape urban space and not vice versa, this revitalized urban theory attempts to carve out a dynamic and formative role for the spatial. Three of most influential approaches have been introduced by Michael Dear, Edward Soja (both working out of Los Angeles), and David Harvey. Michael Dear offers what is on the surface the most playful of the approaches, building an alternative model of urban space consisting of, in his words, dreamscapes, carcereal cities, privatopias, ethnoburbs, and the like, representing (again in his words) the keno capitalism of the new world bipolar disorder. Beneath the somewhat mischievous vocabulary, the point is that new economic relations of power are manifest in radically new spatial form. Edward Soja works extensively with Henri Lefebvre's Marxian theories on the production of space to develop a concept he calls 'thirdspace', a deconstructive possibility where perceived space, conceived space, and lived space are all simultaneously constructed and mediated. David Harvey – though working thousands of miles away from Los Angeles – has developed a more explicitly neo Marxist theoretical perspective that has been similarly influential for postmodern urban theorists. Harvey revisits Marx to excavate the spatial undercurrents from his writing in light of twentieth and twenty first century waves of economic restructuring and globalization. His theories provide a direct link between the broad processes of de industrialization, post Fordist flexible accumulation, global immigration, and other material patterns and the shiny commodified streetscapes of postmodernity. Harvey also explicitly states what is implicit in much postmodern urban writing: that the production and reproduction of urban space is an inherently utopian process, in that it is a continually unfolding discourse about how the world should look and how it should operate. Indeed, postmodern space is often interpreted in utopian terms, whether it is a Disney like utopian charmscape or a Jameson esque dystopian hyperspace. Harvey helps locate the utopian impulse by theorizing it within contemporary spatial form.

Postmodern Critiques of the City/ Critiques of the Postmodern City

There are almost as many criticisms of postmodernism as there are definitions; even when just considering the postmodern city, numerous formidable critiques have been voiced. They fall broadly into two camps – first, critiques regarding the desirability of a world filled with postmodern urban spaces, and second, critiques attacking the theoretical and epistemological underpinnings of postmodern urban models. Adherents of the first camp agree that postmodern urban changes have occurred, but lament what they see as the degradation of urban space. Adherents of the second camp question the underlying assertion that the models being offered are indeed adequate to interpret contemporary urban space, or whether in fact a postmodern break has occurred at all.

Gated communities, skyscrapers built as fortresses, surveillance cameras, fabricated shopping streets, skin deep historical facades, and other hallmarks of postmodern form have drawn harsh criticism from many observers. The architectural forms of the postmodern city have been dismissed as culturally vacuous, or worse, treated with great suspicion as the material embodiment of a neoliberal agenda that foreshadows an increasingly balkanized city where, despite postmodern rhetoric promoting diversity and heterogeneity, actual urban spaces become unprecedentedly polarized along socio economic fault lines. Critics fear that the postmodern city with its commodified theme park environments will seductively distract the citizenry from seeing what is really going on – those citizens, that is, who were not barred entrance by gates, surveillance cameras, and other militaristic means. Further, a planning process that encourages planners to assume the role of facilitators is identified as a key element of the problem. In their attempts to build consensus, critics argue, planners sim ply legitimate the interests of corporate capital as they inadvertently become mouthpieces for the most powerful stakeholders at the table. Thus the postmodern planning model is seen to potentially replicate the hierarchies of power that created geographically uneven development in the first place.

Other critics of the postmodern city go further and question the underlying tenets upon which postmodern urban theory is based. A common argument centers on whether a radical postmodern break has indeed occurred, and if so when. With no postmodern break, critics assert, traditional urban theories still serve well. Even if a break has occurred, others question, does proposed theory ex plain it? Postmodern urban theory is often accused (as is most postmodern theory in general) of going too far in rejecting master narrative and letting in competing voices, to the point of allowing interpretation to dissolve into an incommensurate pastiche of observation or un verifiable personal anecdote. A related concern focuses on the ways in which culture is implicitly theorized within postmodern theory – does cultural expression inevitably and deterministically follow other social changes? Recent scholarship in cultural studies suggests cultural form, while entwined with social and economic production, is not so easy to predict. The image painted by Jameson of cultural zombies wandering helplessly through the Bonaventure hotel unable to construct the cognitive maps that would lead them to the exits is theoretically unsettling to some. Ironically, for a body of work that rejects totalizing theory, much of postmodern thought ends up treating culture as monolithic and static. Culture, critics argue, is hegemonically more creative than that.

These general concerns have led some to examine actual postmodern spaces more closely, and to track how they change once they are lived in and part of the material landscape. Quite interesting observations have been made that encourage more investigation into the significance of postmodern urban form. Frank Gehry's famous Santa Monica house (where he lived), for example, after several years underwent a more con ventional remodel much to the horror of Gehry's post modern fans. The motivation, however, was not for any inherent esthetic reason but reportedly to better accommodate the needs of his two growing boys when they became teenagers. Charles Moore's Plaza d'Italia, upheld as an icon of postmodern space both publicly accessible and architecturally stunning, was subsequently taken over by a nearby luxury hotel which now claims an exclusive catering contract for the space. And the pinnacle of postmodernity itself, Disneyland – both embraced as charming, livable urban space and reviled as darkly sinister embodiment of post Fordist commodifi cation – is itself no longer the white, middle class, themed environment Walt Disney originally envisioned. Now over 50 years old and situated in a predominantly Latino region, Disneyland has undergone remarkable and unexpected transformations at the hands of new gener ations of visitors (though postmodern advocates can rest assured the same transformations have not occurred to the same degree at the Walt Disney World property in Florida). In the end, postmodern urban space – whether it formally exists or not – is a fluid and exceptionally revealing representation of the ways urban forms are constituted, reconstituted, and discussed in specific geo graphic contexts.

See also: Cultural Geography; Feminism/Feminist Geography; Fordism, Post-Fordism and Flexible Specialization; Los Angeles School of Post-Modern Urbanism; Modern City; New Urbanism; Planning, Urban; Postmodernism/Postmodern Geography; Privatization; Public Space; Quantitative Revolution; Urban Architecture.

Further Reading

Davis, M. (1990). *City of Quartz: Excavating the Future in Los Angeles*. London: Verso.
Debord, G. (1999). *Ecology of Fear: Los Angeles and the Imagination of Disaster*. New York: Vintage.
Dear, M. J. (2000). *The Postmodern Urban Condition*. Oxford: Blackwell.
Dear, M. J. (ed.) (2002). *From Chicago to LA: Making Sense of Urban Theory*. London: Thousand Oaks; New Delhi: Sage Publications.
Duany, A. and Plater Zyberk, E. (eds.) (1992). *Towns and Town Making Principles*. New York: Rizzoli.
Ellin, N. (1995). *Postmodern Urbanism*. Oxford: Blackwell.
Fainstein, S. S. (2000). New directions in planning theory. *Urban Affairs Review* 35, 451 478.
Harvey, D. (1989). *The Condition of Postmodernity: An Inquiry into the Origins of Cultural Change*. Oxford: Blackwell.
Harvey, D. (2000). *Spaces of Hope*. Berkeley: The University of California Press.
Jacobs, J. (1961). *The Death and Life of Great American Cities*. New York: Random House.
Jameson, F. (1991). *Postmodernism, or the Cultural Logic of Late Capitalism*. Durham: Duke University Press.
Jencks, C. (1991). *The Language of Post Modern Architecture* (6th edn.). New York: Rizzoli.
LeFebvre, H. (1974). *The Production of Space*. Oxford: Blackwell.
Soja, E. W. (1989). *Postmodern Geographies: The Reassertion of Space in Critical Social Theory*. London: Verso.
Soja, E. W. (1996). *Thirdspace: Journeys to Los Angeles and Other Real and Imagined Places*. Oxford: Blackwell.
Sorkin, M. (ed.) (1992). *Variations on a Theme Park: The New American City and the End of Public Space*. New York: The Noonday Press.
Venturi, R., Brown, D. S. and Izenour, S. (1972). *Learning from Las Vegas: The Forgotten Symbolism of Architectural Form*. Cambridge, MA: The MIT Press.

Watson, S. and Gibson, K. (eds.) (1995). *Postmodern Cities and Spaces*. Cambridge, MA: Blackwell.

Relevant Websites

http://www.cnu.org
Congress for new urbanism.

http://www.ihabhassan.com
Ihab Hassan, a biography.

http://www.egs.edu
Jean Baudrillard, a professor of philosophy of culture and media criticism.

http://www.sprawl busters.com
Sprawl Busters international clearinghouse for anti sprawl.

http://www.dpz.com
Website for duany plater zyberk & company.

Postmodernism/Postmodern Geography

C. Minca, Royal Holloway, University of London, Egham, UK

Introduction

In last few decades, the postmodern debate has literally transformed the cultural and intellectual scene across a whole range of disciplines, from literary criticism, the visual arts, and architecture, to large portions of philosophy and the social sciences, all of which witnessed an unprecedented series of theoretical innovations and critical 'streams', commonly brought together under the 'postmodern' umbrella.

Radical relativism, the fragmentation and instability of subjects and objects, a loss of confidence in comprehensive theoretical frameworks, and a rejection of the grand narratives that laid the foundations of mainstream, 'rational' understandings of the modern tradition: these were just some of the characteristics of this proliferation of critical approaches. But it is these same characteristics that also render particularly difficult (and unstable) any attempt to trace the perimeters of what, at least in some Western academic circles, is considered as one of the most relevant intellectual developments of the late twentieth century.

Human geography, especially in the English speaking world, was deeply marked by this revolutionary set of events. On the one hand, the generalized 'rediscovery' of the spatial dimension in social theory and philosophy brought by the postmodern turn drew new attention to geographical thought and the geographical tradition; on the other hand (and according to many commentators), the postmodern contributed in an important way to the 'return' of geography to the mainstream of the social sciences, paving the way to a new and highly prolific interdisciplinary dialog.

If there was one common denominator that characterized the diverse 'postmodern' streams of thought, it was the focus on the so called 'crisis of the representation'. Geography, as a discipline that had always been concerned with the ways in which the world and its spatial relations were represented, could not exempt itself from this radical questioning of the very principles that guided the production of modern Western knowledge. According to George Benko and Ulf Strohmayer, human geography came away from the encounter with the postmodern "with a renewed sense of mission, vindication, even of pride." This new centrality assigned to the discipline of human geography was visible in a number of ways: the sheer amount of space assigned to geographers in volumes dedicated to postmodernism; the frequent reference to geographical theories on the part of other social scientists as well as those working in cultural studies; and even more pervasively, the new fashion for spatial (and geographical) metaphors in contemporary social theory.

In the last decades, the postmodern and postmodern geography (however we may define it) have become an established presence in the discipline. It is a presence evidenced not only by the vast production of articles that, from 1986 onward, have appeared on the pages of some of the most prestigious geographical journals, and the publication of numerous books explicitly informed by a 'postmodern perspective'; the key place of the 'postmodern' in geography today is also supported by the attention afforded to it by some of the most influential editorial initiatives that have attempted to map post war Anglo American human geography. By and large, the postmodern wave in geography is recognized as a phase – particularly intense between the mid 1980s and the mid 1990s, but which also retained part of its vibrancy in the years that followed – that contributed in a fundamental way to the intellectual atmosphere of the disciplinary debates of those years; a phase during which human geography gained, in many ways, newfound prominence within the social sciences.

Writing about the postmodern or, better yet, writing after the postmodern turn, is fraught with difficulties; any attempt at tracing a cognitive mapping of the postmodern itself (or of 'postmodern geography'), is necessarily a partial and subjective endeavor. In other words, any attempt at defining the state of the art of postmodern geography is doomed to fail; if only for the very fact that the introduction of the postmodern to geography – as to other fields of enquiry – has implied the breaking down of any paradigmatic logic and skepticism toward any linear reconstruction of the history of the discipline. The aim of this article, therefore, is not to describe and define the contents and the borders of a putative 'postmodern geography'. On the contrary, its objective can only be that of identifying, in a partial and entirely subjective fashion, some of the intellectual trajectories that marked one of the most important moments in recent English language geographical debates, together with the cultural and intellectual atmosphere that made them possible.

A singular and coherent body of 'postmodern theory' has never existed, possibly due to the very nature of postmodern thought. Equally, there has always been a lack of consensus regarding what the postmodern really consists of, and, as a consequence, its historical roots and relative evolution. Some commentators have envisaged the postmodern as enabling new forms of cultural and political resistance; conversely, others have seen within it the very end to any form of progressive politics and a tacit legitimation of the *status quo*. The postmodern attitude has been often depicted as anti Marxist, and yet it has been adopted by many 'post Marxists', and even considered by some as an unintended offspring of Marxist theory; it has represented a radical revolt against modern rationality and its consequences, but its reflections are rooted, in many ways, in critical modern theory; it has been the subject of attacks from both the Right and the Left, but in its own 'relativism', it is still struggling to define what a genuinely postmodern polit ical stance might be; it has fundamentally challenged all forms of intellectual authority, and yet its impact is precisely the product of the academic prestige of its early advocates. In order to compose at least a partial history of the postmodern turn, the next section will attempt to 'excavate' some of its most influential trajectories.

Postmodern 'Archeologies'

The exact origins of the term 'the postmodern' (or 'postmodernity') are uncertain. In Best and Kellner's reconstruction of the postmodern's genealogies, besides a few isolated (and relatively obscure) episodes in the late 1800s and early 1900s, when the term was used to define a certain artistic or intellectual avant garde, the idea of the 'postmodern' emerged only in the years immediately following World War II. Best and Kellner trace the first popular usage of the term to 1947, in D. C. Sommervell's summary of Arnold Toynbee's epic *A Study of History*. Sommervell identifies a presumed 'rupture' with the modern era and the birth of a new, 'postmodern' epoch: Toynbee himself will later adopt the term as well, identifying the 'breaking point' between the modern and postmodern epochs in the year 1875, a historical moment characterized by widespread anarchy and disorder and an accompanying cultural relativism.

From the 1950s onward, the term begins to gain currency, especially in the United States, where it appears in several volumes. Albeit lacking a precise definition, it is almost always used with a historical connotation: an epoch that follows the modern one; an epiphenomenon marking the end of modernity or at least of some of its elements/aspects. Probably the best known example of this early usage can be found in C. Wright Mills' *The Sociological Imagination* (1959); in Mills' peri odization of Western civilization, the current epoch signals the beginning of a new, 'postmodern' era. None theless, the early usage of the term represents, at most, a series of relatively isolated episodes, for they do not ignite a genuinely open debate on the putative emer gence of a 'postmodern epoch'; it is thus difficult to link these early works with the theoretical developments of the decades that follow.

American cultural theory begins to adopt the term with more frequency in the 1960s and 1970s, but with an entirely different meaning: the 'postmodern' becomes one of the code words for a broader antimodernist stance that finds its clearest expression in the visual arts and, more generally, the emerging pop culture. The very beginnings of a 'postmodern debate' can thus probably be traced back to these early stirrings (although, again, the term was not that widely diffused). Since a reliable and comprehensive reconstruction of the multidisciplinary debates of these early years is virtually impossible, Hans Bertens suggests looking at the history of American lit erary criticism in order to begin to grasp the emergence and the evolution of the idea of the postmodern, for it is in this disciplinary realm that the postmodern debate takes its clearest shape, at least initially. Literary criticism is, according to Bertens, probably the only institutional ized field of enquiry where a certain antimodernist stance can be identified with a certain degree of continuity from the 1950s onward. Despite the fact that the revolutionary impact of the postmodern would, at the end, be much more pronounced in the 'arts', it is the location of post modernism's beginnings within American literary criti cism that will be more influential on later interpretations: as various scholars have suggested, much of the confusion surrounding the origins and nature of postmodern thought can be attributed precisely to the fact that it emerged within this domain, a domain characterized by a very specific conception of the 'modern'.

What is most important to note, however, is that for many of its advocates, this first postmodern wave repre sented above all an escape from – and a rejection of – the grand narratives and epistemologies of modernity, toge ther with their associated metaphysics of representation. It is within just such an intellectual atmosphere that the postmodern idea began to affirm itself as a specific understanding and codification of the emerging new forms of expression in the arts and the humanities more broadly. Set against the rationality, the 'seriousness' and the paradigmatic structuration of modern art, the new, 'postmodern' experiments professed an ironic, playful lightness of being, an eclecticism marked by pastiche and, in their more extreme expressions, even a certain degree of nihilism.

Following these American beginnings, the 1970s witnessed a shift in the postmodern debate, as it becomes gradually absorbed into the orbit of French post structuralism. The political and cultural events of the late 1960s, in France in particular, are a defining moment. These, together with the rapid transformations occurring in the global political economy, exact a wholesale revision and reconsideration of the structuralist episteme that had dominated the cultural scene in France up to then. Until then, social analysis had focused largely on the identification of underlying, 'universal' rules; it was aimed at the production of a knowledge that sought to be objective, coherent, rigorous, and, above all, 'true'. The social sciences were particularly concerned with maintaining at least a veneer of 'scientificity' and 'objectivity' in order to grant legitimacy to their theories.

The post structuralist 'revolution' consisted, above all, in an attack upon the 'scientific' claims of dominant paradigms, claims that laid the basis for the construction and the adoption of rational models of society (and social analysis); claims whose ontological foundations were based in Enlightenment idea(l)s of truth and objectivity. Setting itself in opposition to such universal and abstract models, post structuralism placed emphasis, rather, on a recovery of everyday experience and 'contextual' interpretations of the social; on the need for an awareness of the political dimension of every form of knowledge, and the instability of any language; it called for a new understating of the 'cultural', and of the inescapably Eurocentric nature (and origin) of all modern theory.

Many post structuralists were influenced by the writings of Jacques Derrida and, in particular, Derrida's call for a radically new philosophical praxis. Indeed, the initial encounter between post structuralism and the postmodern was centered largely on an adoption of the deconstructionist approaches pioneered by Derrida and Roland Barthes; it is only later that other post structuralist influences would gain importance, such as the theorizations of Michel Foucault on power and discourse and the Lacanian reinterpretation of Freudian thought.

The 'deconstructionist' postmodernism of the early years was strongly marked by a 'linguistic textual' approach and focused on contesting the essentializing nature of all representations. Many of these early works were devoted, indeed, to an analysis of the functions of language and, especially, the impossibility of language to represent that which transcended its own boundaries, with a great deal of attention to the deconstruction of texts and to intertextuality. Since deconstructing representations was to be conceived as an inherently 'political' act, the postmodern celebrated first the supposed 'death of the Subject' and, subsequently, that of the Author as well. Paradoxically enough, the putative demise of the metaphysics of representation rendered the questions of subjectivity and authorship even more important – precisely because when faced with the absence of absolute truths, what now mattered most in the process of the transmission of knowledge was who was speaking (or writing), why, and for whom.

Another key post structuralist influence on the postmodern turn was the preoccupation with the functioning of power and the constitution of subjectivity, in great part inspired by the work of Michel Foucault. The production of knowledge that, in positivist terms, was often described as potentially objective and neutral, was now seen as inherently imbued with power – and thus became an object of analysis, as did its practices and protagonists. While postmodern analysis did not entirely embrace Foucault's theoretical edifice, it adopted in large part his guiding claim that knowledge and language were inseparable from the exercise of power. It is for this reason that many postmodern scholars focused their writings on 'unveiling' the implicit political dimension of all representations and on combating the hegemony of dominant discourses that inevitably silenced other, submerged voices; this, in part, explains the postmodern's insistence on the recognition and valorization of difference, polyvocality, and pluralism. Much attention was devoted, indeed, to the analysis of the condition of the Other – that is, the condition of those normally excluded by hegemonic, 'normal', and normalizing discourses and practices: ethnic, racial and sexual minorities, the differently abled, etc. In particular, great importance was granted to understanding the processes of the constitution of these various Others, and the role of representations, discourse and everyday language in this very constitution.

The translation into English of Jean Francois Lyotard's *La Condition Postmoderne* in 1984 is an important turning point: here, a prominent post structuralist theorist adopts the 'postmodern' term, marking the definitive confluence of the two schools of thought – American postmodernism and French post structuralism. Just like post structuralism, this 'new' postmodernism radically rejects the idea that language can faithfully represent reality; that representations can provide the key to 'true' knowledge; and that the things of this world are somehow 'reflected' by our discourses. Rather, it argues, it is language that constitutes our world; knowledge, therefore, can only be partial, inevitably influenced by social, political, and cultural circumstances and the specific historical context within which it is articulated. What is important to note here is that post structuralism was not only a key source of inspiration for the postmodern turn; at certain moments, one and the other became essentially indistinguishable (at least in their North American variants). In other words, the postmodern debate not only absorbed the post structuralist critique of structuralist approaches, but also this latter's search for new modes of enquiry, new languages, and new forms of subjectivity. The sense of radical rupture with the past that this break represented came to be seen either as a long awaited liberation from the shackles of the modern episteme or, more negatively, as the definitive loss of traditional Enlightenment ideals and values.

But not everyone conceived the postmodern turn in this philosophical–cultural–linguistic optic. For some analysts, it was the world itself that was changing in such rapid fashion that new understandings, new cognitive tools were necessary to be able to decipher its transfor mations. The globalization of many economic processes, the IT revolution, the supposed planetary diffusion of certain consumption models, the increasing fragmen tation of lifestyles and cultures, the emergence of a myriad of new political and cultural subjects, the growing concern for the environment, the reemergence of regionalisms, localisms, and a variety of fundamentalisms, the triumph of a capitalist regime of flexible accumu lation, the retreat (if not the 'end') of the nation state, and the progressive privatization of many realms of social life – these were only some of the expressions of the epochal shifts occurring across the globe; shifts interpreted by some as the closing of an era, as the confirmation of the end of modernity and the emergence of a new, 'postmodern' society and culture.

Frederic Jameson, in his agenda setting article 'Post modernism, or the cultural logic of late capitalism' that came out on *New Left Review* in 1984, on the one hand depicts the postmodern as a product of the transforma tions of global capitalism; on the other, as a radical shift in our spatio temporal coordinates, calling for a radically new set of analytical tools. Jameson's piece was to prove crucial in bringing the postmodern debate to geography and geographers. His focus on the built environment and on architecture, as well as the particular emphasis gran ted to space and spatial processes as key to understanding and codifying the shifts in the organization of advanced global capitalism, could not but capture geographers' attention.

Envisioning the emergence of a 'postmodern society' and of concrete, architectural expressions that pre sumably marked its birth, meant in a sense legitimizing a linear vision of historical and social evolution. And yet it is this understanding of the postmodern that was to initially inspire geographers and other social scientists. Certainly, the fact that the principal early protagonists of the postmodern turn in geography (and other disciplines) shared a Marxist background was important: Jameson's materialist interpretation of the transformations of global capitalism was, in many ways, articulated in very familiar terms. What is more, Jameson's was the first widely accepted analysis of the postmodern in the social sciences and in their attempt to bring the discipline back to the heart of the intellectual debates of the day, geographers could not but be influenced by it.

Up to that moment, the social sciences had, for the most part, generally ignored the debates on the post modern that were developing in the meantime in philosophy, the arts, architecture, and the humanities, more broadly. It was only in the mid 1980s, right around the time of the publication of Jameson's piece, that sociologists, in Britain especially, began to turn their attentions to the postmodern. The initial foci of their debates were the more clearly sociological aspects of Jameson's theorizations, as well as those of French soci ologist Jean Baudrillard: in particular, the relations between the postmodern and new forms of production and consumption. This theoretical perspective charac terizes a significant proportion of English language social scientific production of the subsequent years, focused especially on the emergence of new, post Fordist production regimes and the broader transformations of global capitalism as a response to the systemic crises of the 1970s. Jameson's piece was to prove particularly useful here, offering social scientists a novel reading of these transformations; a reading able to link the new, as yet unexplored phase of advanced capitalism to the astounding series of shifts visible in the social and cultural spheres. What is more, it proposed an entirely new theoretical vocabulary and analytical apparatus, able to describe the spatialities of late capitalism.

The publication of Jameson's article marks, in fact, the upsurge of a much wider interest in the postmodern throughout the 1980s. Numerous books and articles appear in these years, as does a journal that will soon become an important forum for the postmodern debate: *Theory, Culture and Society*, edited by Mike Featherstone. In 1988, the journal dedicated an entire issue to the postmodern turn, in many ways setting the agenda for the years to come. It is also these years that witness the publication of some of the key texts of the postmodern debate: Foster's 1985 *Postmodern Culture*; Hassan's 1987 *The Postmodern Turn: Essays in Postmodern Theory and Culture*; Featherstone's 1991 *Consumer Culture and Postmodernism*; Lash's 1990 *Sociology of Postmodernism*; Bauman's 1992 *Intimations of Postmodernity*; Hutcheon's 1989 *The Politics of Postmodernism*; and Jencks' 1992 *The Post Modern Reader*. Geographers add their own voices to this debate: in particular, David Harvey's *The Condition of Postmodernity* and Edward Soja's *Postmodern Geographies*.

We can thus trace three possible routes through the 'archaeology of the postmodern', routes that will produce three distinct interpretations of the postmodern in geography. The first interpretation draws upon that which was identified here as the initial, post structuralist phase of the postmodern debate. Strongly influenced by the work of Jacques Derrida and, in part, that of Roland Barthes, it focuses on the deconstruction of geographical discourses and texts, in order to unveil the 'taken for granted' that disciplines such discourses and texts and grants them their 'normalizing' power.

The second interpretation, on the other hand, is largely inspired by the theorizations of Michel Foucault and the other authors whose work fueled the second phase of the postmodern debate. It is an interpretation that places particular emphasis on the exercise and functioning of power and, in particular, the relationship between power, discourse, and space. Key here are the influences of feminist scholars, as well as the emerging literature in postcolonial studies. This is a particularly rich and innovative strand of the postmodern debate, containing a great variety of diverse perspectives and approaches, and thus difficult to capture within a single mapping.

The third and final interpretation is that most closely inspired by Jameson's work and the early debates in the social sciences, and directly emergent from Marxist and post Marxist traditions. This strand of postmodern geography is often focused on studies of the postmodern city and, more generally, postmodern spaces, as well as analyses of social, cultural, and economic processes in a postmodern 'era'.

This brief archeology of the postmodern is certainly partial and incomplete. Many other authors and influences could have potentially been taken into consideration in order to offer a more nuanced and complex understanding of a phenomenon that has always evaded a strict mapping. For instance, the influence of not only Pierre Bourdieu, Guy Debord, and Richard Rorty, but also that of Gilles Deleuze and Felix Guattari, is cited in most reconstructions of the postmodern's early archeologies. To this list of names, we could also add those of James Clifford, Clifford Geertz, Ernesto Laclau and Chantal Mouffe, Henri Lefebvre, as well as Edward Said – all individuals whose work has been enormously influential, also in geography.

Postmodern Geographies …

What has the postmodern turn meant for human geography, then? How did geographers respond to it? How did it change our way of conceiving – and of doing – geography? The end of the 1980s witnessed a whole series of important shifts in human geography that can be understood both as a reaction to developments within the discipline itself and also as a response to a variety of external stimuli. Many 'critical' Anglo American geographers began, in fact, to turn against positivist interpretations and, in particular, those associated with spatial analysis. At the same time, however, those years also witnessed a relative disenchantment with radical and Marxist approaches on the part of some progressive geographers: at the heart of this disenchantment lay the fact that some Marxist theorists still confided in the power of 'grand explanations' and in the 'progress' of (social)scientific analysis, while at the same time devaluing the importance of the subjective dimension. Some Marxist geography was accused, in particular, of advancing 'essentialist' theses: in having tried to universalize the concept of social class, it had essentially forgotten the importance of context (and thus the role of diverse places) as well as the importance of individual identities. Such critiques, however accurate they may have been, were key in opening geographical reflection to the postmodern debate that appeared to offer, at least in part, an alternative to the structuralism and 'essentialism' of Marxist perspectives, through its attention to – and valorization of – difference in its myriad expressions.

The growing interest in this new set of approaches was also spurred on by the rapid transformations occurring in all realms of social life – from politics, to popular culture, to the economy – transformations that necessitated a new conceptual apparatus and new analytical tools from geography and geographers. Nonetheless, it is important to note that geographers were relative latecomers to the postmodern debate, and it is indeed debatable whether the discipline really ever entered a 'postmodern phase'. So while it is true that most post structuralist theories today make use of spatial metaphors to one extent or another (if not of geographical concepts tout court), it is equally true that while the postmodern debate was raging in other disciplines at the end of the 1970s and beginning of the 1980s, 'critical human geography' was still struggling with its Marxist legacy and a variety of post Marxist critiques of structuralism, such as structuration theory.

The fact that post structuralist approaches would come to Anglo American geography mainly through Marxism and the writings of Marxist/post Marxist geographers would have important repercussions on the development of the postmodern debate within the discipline. Indeed, many of the ideas that we consider today as integral part of the postmodern tradition in geography can only be understood as the fruit of the encounter of postmodern theories with notions coming from historical materialism. Again, Jameson's 1984 piece was key in this encounter, but so was a much broader body of critical social theory of the early 1980s. What is more, successive Marxist critiques of postmodern ideas – or even the wholesale rejection of postmodern perspectives on the part of some Marxist geographers – came to be considered, paradoxically enough, as some of the most influential works on 'postmodern geography'. Perhaps the best known case is that of David Harvey's 'The condition of postmodernity', a work that was explicitly anti postmodern but that came to be considered by many as a foundational text of this new wave of 'postmodern geography'. The Marxist influence on 'postmodern geography' will begin to lose its strength only in the following decades, with the growing importance of post structuralist, postcolonial, and feminist theory.

The early years of the postmodern in geography were also marked by a distinctive set of research foci. Key to these was the question of the emergence of the 'postmodern city'. Rejecting wholesale the principles of the Chicago School, these early works called for the elaboration of new understandings in the spirit of a 'postmodern urbanism', conceiving the city as a complex socio spatial phenomenon that could not possibly be captured within a single exhaustive definition, within a singular logic. Alongside this new interest in the city and its transformations, 'postmodern' geographers turned their attentions to the broader changes occurring in society, hypothesizing the emergence of a 'postmodern epoch', an epoch characterized by a new cultural logic, by post Fordist political economies, by a radical reconfiguration of the spaces of the political, and by the presumed triumph of representation over reality. Postmodernity was, therefore, not seen as a singular manifestation, awaiting discovery and description by geographers: it was envisioned, rather, as a complex of diverse perspectives, processes, and objects, a confused and convoluted collage of people and places, un decipherable with the analytical tools of the past. Following the writings of Scott Lash and John Urry on the end of 'organized capitalism', the work of Mike Featherstone on the globalization of culture, and the theorizations of Pierre Bourdieu describing the new division of society into distinct classes of consumers, some geographers began to turn their attention to the role of the mass media and of representations, as well as the changing role of consumption. Some of this research focused its attention on the places of consumption, as well as on the consumption of places themselves. Many writings investigated the relationships between postmodern lifestyles and tastes and the transformation of specific urban and regional contexts, as well as the role of 'cultural capital' in radically transforming entire city neighborhoods.

The postmodern turn thus brought with it a number of important consequences for human geography. As in other disciplines, it was welcomed with enthusiasm by a significant number of geographers who found its innov ations (and provocations) inspirational. But the post modern also encountered considerable resistance, if not outright hostility, on the part of others who perceived its arrival as a menace to the discipline's identity. Yet, neither the initial misunderstandings between the two 'factions' nor the relative indifference with which some treated it at the outset, prevented the postmodern from blossoming in geography and becoming, in the decades that followed, a key part of the disciplinary 'tradition'.

Postmodern Geographers…

Most accounts identify two crucial passages, two crucial moments in the emergence of a 'postmodern geography'. The first is the period between 1986 and 1988, when a series of articles appeared, all making explicit reference to postmodern analysis. The journal *Environment and Planning D: Society and Space*, first published in 1983, rapidly became the forum of choice for cutting edge reflection on social theory in geography and, from 1986 on, the place where a great part of the debates on the postmodern were played out. It is *Society and Space*, in fact, that in 1986 published the first articles explicitly dedicated to the postmodern: Michael Dear's reflection on 'postmodern' urban planning and Edward Soja's exploration of Los Angeles' 'postmodern' spatialities.

Other articles followed in rapid succession: in 1987, David Harvey's piece on postmodernism and the American city in *Antipode*, and David Ley's analysis of postmodern urbanism in Vancouver in the *Journal of Historical Geography*. A year later, Michael Dear published 'The postmodern challenge: reconstructing human geography' on the pages of *Transactions of the Institute of British Geographers*. Dear's article was particu larly important as both a first attempt at mapping the postmodern turn in geography, as well as a call to arms to the discipline as a whole: the postmodern offers a unique opportunity, Dear argued, to bring the discipline back to the mainstream of the social sciences. Equally influential was Derek Gregory's 1989 essay on 'Areal differentiation and postmodern geography' (part of the edited volume *Horizons in Human Geography*), sug gesting a curious parallelism between postmodernism and geography. Geography's long standing interest in 'difference', Gregory argued, rendered it a potentially more welcoming field to the relativist claims of the postmodern agenda. The postmodern thus emerged from its relatively marginal position in the early 1980s, to affirm itself as 'the' expression of cutting edge reflection in geography. Between 1986 and the end of 1993, numerous papers, articles, and essays dedicated to the postmodern and postmodernity appeared, giving life to a series of different 'strands' of reflection, associated with the different positions in the disciplinary debate.

In some ways, this wealth of reflection can be seen as simply the definitive codification of a much longer dis ciplinary 'ferment'. Already during the preceding decade, geographers had begun to explore the epistemological and ontological foundations of the discipline in ways that would strongly influence the 'postmodern' debates in the late 1980s and all through the 1990s. We can note, among others, Soja's writings on the socio spatial dialectic from the early 1980s (a reflection that would culminate with the publication of his seminal book *Postmodern Geographies* 9 years later), as well as Gunnar Olsson's considerations on the nature of language, dating to more or less the same period. However we may choose to locate these influ ences, what is most important to note in this brief his torical reconstruction is the flux of new, provocative ideas into the discipline that, by the end of the 1980s, have fundamentally transformed its theoretical debates and its role in the social sciences.

1989 represented, in fact, a turning point, with the appearance of two books entirely dedicated to the postmodern in geography. The first was Edward Soja's aforementioned *Postmodern Geographies*, a collection of essays on the postmodern read through a Marxist 'lens' and inspired, in particular, by the work of Henri Lefebvre (and, to an extent, also by Michel Foucault). In this by now pathbreaking book, Soja sets out a call for a new 'geographical materialism', able to marry the latest developments in social theory together with a new, privileged attention for space, something that had been traditionally sacrificed by historical materialism. Apart from its contribution to 'bringing the spatial back in', the collection contained also two 'empirical' chapters in which Soja attempted a 'postmodern' reading of the Los Angeles spaces, launching an entirely new way of doing urban geography. Soja's engaging prose and even his choice of topics revolutionized traditional ways of writing about the city and *Postmodern Geographies* would prove a fundamental influence for not only the next generation of geographers, but also urban scholars in other disciplines.

The second volume to appear that year and to leave an equally important mark on the development of post modern reflection in the discipline was David Harvey's 'The condition of postmodernity', both for the notoriety of its author, and its decidedly antipostmodern stance. In his critical appraisal of contemporary post Fordism, Harvey provided a scathing analysis of the new global regime of flexible accumulation, as well as of the 'post modern culture' dictated by the market, a culture based on a problematic ellision between the profit imperative and artistic production, between fiction and reality. Harvey's critique is quite distinct from other postmodern interpretations: the contemporary historical phase was, in his view, to be located (and understood) within a broader periodization of capitalism, characterized by ever increasing time–space compression. 'The condition of postmodernity' thus follows the spirit of Marxist critiques of the postmodern, of which the author would become one of the most authoritative interpreters.

The publication of these two volumes marked the definitive legitimation of postmodern reflection in Anglo American geography but not only, given its growing in fluence on other national geographical communities. From that moment on, research adopting the postmodern as theoretical and/or methodological perspective began to proliferate, from studies of cities, to those on land scape, on the spectacularization and consumption of culture, on the spatial effects of flexible accumulation regimes – as well as on geography and the production of geographical knowledge. In the meantime, a new gen eration of geographers started to reconsider the contri butions of geography's postmodern 'pioneers'. For some, the post Marxist legacy was seen as actually part of the problem raised by postmodern reflection in the social sciences, rather than a solution or set of viable answers. The postmodern debate in geography thus began to engage with a new series of approaches, this time around influenced much more directly by post structuralist thought. Some of the most fruitful terrains of analysis where this occurred are the various deconstructionist approaches to the study of geographical maps and texts; the reflection on hermeneutics and geography (strongly influenced by the writings of Derrida and Barthes); the deep engagement with the work of Foucault and Lefebvre; and the new attention for the body, inspired by Lacanian psychoanalysis. The post structuralist turn also took on other forms, however: we can mention here the important role of feminist critiques of the postmodern, the ongoing dialog with cultural studies and postcolonial theory, or even the birth of critical geopolitics (though the latter inspired more by the post structuralist turn in international relations than in geography itself).

Giving some order to all of these diverse currents of thought is a next to impossible task, not only because of the diversity of the various perspectives and topics of research and the inevitable contaminations between them, but also because many of the protagonists of the post structuralist debates in geography would most likely not consider their work as part of the 'postmodern' canon, albeit (in many cases) recognizing its decisive influence. This is perhaps unsurprising: the 'postmodern' perspective rejects, *a priori*, any paradigmatic closure and is therefore more usefully conceived as a particular way of understanding geographical thought, rather than a specific, alternative methodological apparatus. For this reason, the very proliferation of approaches in con temporary human geography can be considered in itself a direct consequence of the postmodern turn.

Postmodern Cities and Spaces?

The analysis of the transformations of the contemporary city represented, in many ways, the very heart of the geographical reflection on the advent of a new post modern era and the codification of new postmodern spaces. Los Angeles was, for long, at the center of this reflection, identified as the postmodern city *par excellence* – a codification that has not been without its detractors. Some have even gone as far as identifying a distinct Los Angeles School of urban scholarship as the successor to the Chicago School and, like it, destined to influence all future geographical analyses of the city. However accurate this assessment may be, it does mark the enor mous influence of the Big Orange on the collective imaginary of its (postmodern) scribes. What is more, the idea that the birth of a postmodern epoch or, certainly, the end of modernity, corresponds to a distinct set of cultural transformations that can best be read off the changing spaces of the 'postmodern' city found its ideal laboratory in LA, whose 'lessons' were later applied to other urban contexts. Nonetheless, it is difficult to reduce even the Los Angeles school to one common denomin ator, and to trace a clear, singular trajectory from Soja's first explorations in the 1980s, to his and others' more recent travels through the Los Angelene 'postmetropolis' and its edge cities, theme parks, exopolis, and the affirmation of that which Dear would term the post modern urban condition.

The great wealth of writings of geographers, urbanists, sociologists, and anthropologists on the postmodern city can be roughly divided into two distinct camps, reflecting in many ways the two dominant (and seemingly anti thetical) interpretations of the postmodern. The first, as was already noted, considers the postmodern as a distinct era or epoch, as a distinct phase or stage of advanced capitalism; the second, on the other hand, attempts to codify a specific postmodern way of thinking and knowing. In the first interpretation, the object of analysis is the changing 'urban reality', its built environment, and the political and cultural economies of the post Fordist city; in the second, it is more a question of adopting/elaborating a new, 'postmodern' approach to the study of cities. In other words, the postmodern city can be approached either by trying to codify the momentous changes occurring in many European and North American urban regions (and not only) and the emergence of what this literature will term a 'postmodern urban condition' – or, rather, by trying to understand how the transformations in the contemporary city can be best conceived and theorized. Postmodern urbanism can thus be interpreted both as the emergence of new forms of urban development, new urban spaces, as well as a new lens, a new logic within which to interpret the urban.

It is perhaps useful at this point to take a step back ward. The new architectural styles that we usually refer to as 'postmodern' appeared, in fact, already in the early 1970s, with the intention (as Edward Relph suggests in his *The Modern Urban Landscape* in 1987) of revitalizing the old city centers by means of heritage preservation and by means of a whole series of new approaches to urban design and planning, including a new attention to what becomes termed 'community planning'. According to Relph, 'postmodern architecture' was not merely every thing that came after modernism; it was better under stood as an aware revival of elements of previous styles and forms. It is such a 'revival' that guides the various attempts at urban regeneration, heritage preservation, and the redefinition of urban design. It would therefore be correct to say that the postmodern refers not only to a new architectural style, but more importantly a new at titude toward architecture and planning and its role in the rapidly transforming contemporary city.

In his influential edited collection *Variations on a Theme Park* that came out in 1992, Michael Sorkin ex plored some of these new urban 'fragments', proclaiming the end of the traditional, modern city, made up of or derly squares, streets, and boulevards, and the emergence of the postmodern city, linked by fiber optic cables and made up of self referential, securitized enclaves, where public space is privatized and the traditional center dis appears. For Sorkin, postmodern cities were marked by three characteristics: the first is the disappearance of all stable material and cultural relations that, in the past, linked communities to specific places; the second is the obsession with security and the growing connections between the virtual world of the Internet, television and closed circuit cameras, and everyday life; the third is the increasing transformation of urban spaces into 'themed' enclaves. The postmodern city described by Sorkin thus not only transcends traditional urban categories and divisions, but also regiments urban spaces within a new geography of segregated segments, 'public' in appearance only.

A few years later, Michael Dear and Steven Flusty attempted to map these new geographies by tracing some of the most recent developments in Southern California (that the authors consider, nonetheless, indicative of broader urban trends), to then move on to a description of a possible 'postmodern urbanism' able to interpret the cities of today, transformed by what they term 'keno capitalism'. The reconfiguration of the city they described evokes, in many ways, the urban fragments of the Los Angeles region identified by Soja: exopolis, flexcities, cosmopolis, splintered labyrinth, carceral city, and simcities. The variable geometries of the LA urban region were, moreover, associated with a variety of shifts in the global political economy that express themselves 'on the ground' in the form of the technopoles described by Allen Scott in 1993, Fishman's techoburbs in 1987, but also the 'landscapes of despair' of the homeless and the new poor, identified by Wolch and Dear in 1993. The great variety and originality in ter minology is just one way in which the Los Angeles School tried to put some order in the urban magma that fills the spaces of Southern California. The theoretical framework proposed by Dear and Flusty in order to interpret this urban space 'gone mad' was a series of fragments and fragment(ing) processes within which to contain some of the abovementioned manifestations: denominated with a series of (very postmodern) neo logisms such as the 'global latifundia', 'cybergeoisie', 'protosurps', 'memetic contagion', 'citistat', 'com mudities', 'cyburbia', 'citidel', 'cyberia', 'pollyanarchy', 'disinformation superhighway', 'holsteinization', 'praeda torianism', and 'protosurps'.

The Los Angeles School did not only try to make sense of what was happening in Southern California, but also continued to experiment with new readings of urban space that could be more widely applicable – with the assumption that LA, ideally at least, was destined to be come the capital of the twenty first century. It is not by chance that another volume edited by Allan Scott and Edward Soja in 1996 was entitled *The City: Los Angeles and Urban Theory at the End of the Twentieth Century*. It was a voluminous work that brought together the contributions of the principal protagonists of the LA School (together with the editors, Mike Davis, Michael Dear, Charles Jencks, and Jennifer Wolch, among others) and that aimed to provide a definitive assessment of the Los Angelene 'urban condition' through a series of diverse readings, ranging from more structured and structuralist approaches, to narratives that are more 'postmodern' in both language and style.

Certainly, Los Angeles was not the only site of reflection on the postmodern in those years. Katherine Gibson and Sophie Watson's 1995 edited collection *Postmodern Cities and Spaces* was another important con tribution to the debate, bringing together a series of di verse contributions that spanned, as the editors note, two quite antithetical understandings of the postmodern: the first, considering it an era or distinctly new socio cultural condition; the second, a specific way of knowing and understanding the changing urban context. The structure of the book in many ways reflected this divide, although the editors are careful to note in their 'Introduction' that while the tension between these two perspectives may lend itself to a strict mapping, the volume aimed, rather, to offer readers a 'route map' along two complementary itineraries through the postmodern that at times run parallel, at times converge or diverge. Indeed, the most interesting aspect of this collection is the attempt that it makes at bringing together quite diverse writings on the postmodern, from accounts of the postmodern city, its fragmentation and new political subjectivities, to a series of trajectories through the city, inspired by Foucault's work on heterotopia, by the concept of chora, by Walter Benjamin's reflections on the figure of the flâneur, or by feminist reflection on private and public spaces (and bodies). It should be noted that the tension between the two diverse understandings of the postmodern high lighted by Gibson and Watson's volume – that between the postmodern as 'object', a novel, multiple, and frac tured reality, and postmodern(ism) as an 'attitude', or approach – had been already identified by another key text a few years ealier – Cloke, Philo, and Sadler's 1991 *Approaching Human Geography*.

Conclusions

The uncertain trajectories of the postmodern that we have tried to identify in the preceding paragraphs do not exhaust themselves at the end of the 1990s of course. The theoretical reflection on the postmodern in geography has, indeed, continued with great vigor, as evidenced by the publication of a number of 'overview' volumes in the decade that followed: to Benko and Strohmayer's *Space and Social Theory: Interpreting Modernity and Postmodernity* in 1997, Doel's *Poststructuralist Geographies: The Diabolical Art of Spatial Science* in 1999, and Dear and Flusty's *The Spaces of Postmodernity* in 2001, to cite but a few.

A book that deserves special mention is Edward Soja's *Thirdspace*, published in 1996. Soja's approach in this volume is quite distinct from that adopted in *Postmodern Geographies*. This new work explores the interstices be tween postcolonial theory and postmodern geography, using the idea of 'third space' to 'open up' geographers' explorations of the spatialities of life, adopting what Soja terms a 'critical postmodernism', distinct from both the celebrations of the end to all certainties pronounced by certain relativists, as well as from a 'critical modernism' that presents a caricature of the postmodern as nihilistic and self celebratory. Besides the work of Homi Bhabha and other writers who inspired contemporary post colonial reflection such as Gayatri Spivak and Edward Said, Soja's analysis also draws upon the work of Foucault and Lefebvre, as well as the spatial feminism of Gillian Rose and the postcolonial feminist writings of Gloria Anzaldua.

Alongside the above cited volumes, the late 1990s as well as the decade that follows also witness the publi cation of new works dedicated to the postmodern city; innovative analyses of urban spatialities that, in many cases, also pioneer new theoretical and methodological tools, adding to the broader reflection on the postmodern in geography. Key in this respect are Soja's *Postmetropolis*, published in 2000, and Michael Dear's *The Postmodern Urban Condition* in 2000: volumes authored by the two geographers who opened up the reflection on the post modern city 14 years earlier. One of the last books explicitly referring to the 'postmodern' was Claudio Minca's 2001 edited collection *Postmodern Geography: Theory and Practice*. The volume highlights some of the tensions and controversies that have accompanied the advent of the postmodern in geography, tensions that have since consolidated into a distinct set of positions, at least within English speaking geography. In this optic, the comparison with perspectives from other academic contexts is particularly revealing: the contributions of Italian and Scandinavian geographers included in the volume offer, in fact, a very different view on the legacy of the postmodern debate but also its possible futures.

Overall, reexamining these somewhat convoluted disciplinary trajectories in retrospect, at least one thing is clear: the postmodern has become integral part of the geographical tradition. Not only: it is, by now, widely accepted that the advent of postmodern reflection in geography significantly contributed to revitalizing disciplinary debates of the past 20 years and opening geographers' eyes to a whole series of new research avenues and approaches.

See also: Deconstruction; Feminism/Feminist Geography; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Poststructuralism/Poststructuralist Geographies; Structuralism/Structuralist Geography.

Further Reading

Benko, G. and Strohmayer, U. (eds.) (1997). *Space and Social Theory: Interpreting Modernity and Postmodernity*. Oxford: Blackwell.
Bertens, H. (1995). *The Idea of the Postmodern*. London: Routledge.
Best, S. and Kellner, D. (1991). *Postmodern Theory*. New York: The Guilford Press.
Dear, M. (1988). The postmodern challenge: Reconstructing human geography. *Transactions, Institute of British Geographers* 13, 262 274.
Dear, M. (2000). *The Postmodern Urban Condition*. Oxford: Blackwell.
Dear, M. and Flusty, S. (1998). Postmodern urbanism. *Annals of the Association of American Geographers* 88(1), 50 72.
Gregory, D. (1994). *Geographical Imaginations*. Oxford: Blackwell.
Harvey, D. (1989). *The Condition of Postmodernity: An Enquiry into the Origins of Cultural Change*. Oxford: Blackwell.
Huyssen, A. (1984). Mapping the post modern. *New German Critique* 33, 5 52.
Jameson, F. (1984). Postmodernism, or the cultural logic of late capitalism. *New Left Review* 146, 53 92.
Ley, D. (1987). Styles of the times: Liberal and neo conservative landscapes in inner Vancouver, 1968 1986. *Journal of Historical Geography* 13, 40 56.
Soja, E. W. (1980). The sociospatial dialectic. *Annals of the Association of American Geographers* 70, 207 225.
Soja, E. W. (1989). *Postmodern Geographies*. London: Verso.
Soja, E. W. (1996). *Thirdspace*. Oxford: Blackwell.
Soja, E. W. (2000). *Postmetropolis*. Oxford: Blackwell.

Post-Phenomenology/Post-Phenomenological Geographies

J. Lea, Lancaster University, Lancaster, UK

Glossary

Experience The unit of measure of the post-phenomenological. Experience is not necessarily grounded in the human.
More-Than-Human The realms beyond the human (inhuman, nonhuman) that co-constitute practiced engagement.
Post-Phenomenological A rereading of phenomenology as phrased in the most part through post-structural theories.
Practice The bodily relation to the world (encompassing movement, sensation, perception, and action).

Introduction

Post-Phenomenological Geographies

This article gathers together 'post phenomenological' geographies: an emergent and, thus far, fragmented body of work. The wider context of post phenomenological writing is limited, and although only a handful of geog raphers have explicitly identified their approach as 'post phenomenological', there is value in addressing these geographical dialogs between phenomenology and post structuralist theories. The fact that this growing area of research was name checked in a recent 'Progress in human geography' report on cultural and nonrepresentational geographies suggests that it might be an area of future growth. This encyclopedia article gathers these geog raphers together with a range of others whose work explicitly or implicitly references, or measures itself against, phenomenological theorists (and at the same time shows that these areas of research may have similar pre occupations, but have not achieved a measure of consist ency with each other). Phenomenology focused upon the experiencing subject, and, to humanistic geographers in the 1970s, offered an attractive conceptual frame to turn away from the spatial scientific focus that predominated in geography, and to 'get back' to the human experience. Rather than the aggregation of Human beings into stat istics, figures, or mappings, phenomenological approaches were perceived to focus back onto the human as an indi vidual figure; the medium and recipient of experiences, sensations, and fleshy embodiment.

Geographers such as Edward Relph emphasized the deep significance of place (albeit relying on a particular imagination of place) to human being, folding ideas of authentic existence and alienation into a mode of being in place. In brief, phenomenology offered a vehicle to take note of (human) experience, the impact of sensation through the flesh of the body, and the ways in which this experience is constituted by, and of, place. Phenomeno logically influenced human geographies subsequently became less popular as geographers such as Derek Gregory and David Harvey critiqued the ideas as not offering the political purchase required for the devel opment of a critical human geography. For Harvey, dwelling did not offer the resources to deal with the occurrence of societal change, and the 'universal' body offered by phenomenology made Gregory look to the theoretical resources of post structuralism, feminist theorists, and postcolonial theorists instead.

However, in the context of a more general 'practice turn', phenomenology made a return, alongside other theorists such as Bourdieu, Wittgenstein, Garfinkel, and Deleuze. Particularly relevant for the subsequent dir ections taken by post phenomenological geographies was the use of phenomenological ideas by Nigel Thrift in the development of nonrepresentational theory. This signaled a shift from the predominantly humanistic interest in transcendental phenomenologists such as Husserl and the constitutive phenomenologies of Schutz, to the existential forms of phenomenology embodied by writers such as Merleau Ponty.

Transcendental phenomenologists work toward a universal conception of being: their study is not of a particular person in a particular environment; rather it is concerned with a more impersonal, less located form of experience. The interest is in the transcendental ego: a universal source of all experience (from which individual experiences are personalized according to context). Constitutive phenomenologists lend more of a socio logical and geographical conscience, focusing as they do upon the role of society in molding our intersubjectively located understandings of the world; and creating shared communities of meaning that exist within the life worlds of particular societies or cultures.

Existential phenomenologies, having contributed most to post phenomenological geographies, reject the uni versal form of experience and take less of a meaning based approach, arguing instead for the importance of the study of the particularity of experience. Instead of being created through meaning, the life world is constituted through experience and interpretation of this experience. Rather than locating experience in the body, or in intersubjective space, post phenomenological geographies extend experience trans humanly. This allows the analytic focus to be decentered from and extended beyond the human body, taking in the inhuman (including, e.g., the agency of animals or objects) and the nonhuman (the 'pull' of the world). Experience does not just occur within the body, confirming our boundaries and corporeality, but is a creative force distributed across bodies and worlds (all manner of 'things' – objects, events, places, spaces – matter here, having performative and constitutive force).

Crucially, none of these post phenomenological accounts 'leave behind' the phenomenological (as a realm of experience), but rather refigure what experience might be, where it might be located, how it comes about, and how we, as social scientists, might account for it. These post phenomenological modes of working revisit the resources offered by phenomenology, critically returning to concepts such as 'dwelling', and working them through current 'phenomenological' theorists such as Alphonso Lingis and Michel Taussig, and theorists such as Deleuze and Levinas, whose intellectual trajectories developed in post World War II France in critical relation to phenomenological traditions. In this way, these post phenomenological geographies have resonances with other 'post phenomenological' moves across the humanities and social sciences over the last 15 years. Although a diverse and contested field encompassing, for example, Don Idhe's sociologies of science and technology, and more recently a journal special issue edited by Suzi Adams concerned with the ways in which the world is produced and placed, all these post phenomenological writings investigate exactly what it means to transform, rather than abandon, phenomenologies, and in doing, continue to push the boundaries of what it means to be in the world.

Experience

Post phenomenological geographies have made several distinct contributions to the area of geographies of experience. Phenomenological frameworks have been used in conjunction with participant observation and in depth interviews in order to offer a nuanced understanding of the lived world of humans. These accounts of phenomenological experience have overlapped and been mutually constitutive with such areas of research as emotional geographies and embodied health geographies. Phenomenological writings have been used to understand the interrelations between self and other as lived, and the spaces between body, self, and other. Beginning with perceptions, which fill the gaps between perceiver and perceived, phenomenology gives insights into the ways in which space is lived, and how lived space is integrally related to our sense of self. This spatially specific understanding of who we are lends analytical purchase to Joyce Davidson, who uses Merleau Ponty's understanding of being at home in the world (or constructing a 'homely' space) as integral for mental health, to construct a phenomenological understanding of the lived experiences of being agoraphobic. This emphasizes the need of individuals to reinforce their human selves as centers, in order to counter the depersonalizing tendencies that are incipient in every sensation. In allowing an understanding of what happens when the raw 'power' of sensation overwhelms and threatens an individual's embodied sense of self, phenomenology offers opportunities to both describe and explain embodied geographies of health and illness.

The second area of post phenomenological research into experience turns against the remit of phenomenology more explicitly, in that it troubles the category of experience itself, questioning the phenomenological location of experience in the figure of the human. This constitutes a desire to move beyond the reductive act of constituting experience through the contours of the human, and the act of sensemaking (feeling and naming a sensation). A number of different conceptual devices and empirical strategies have been employed to extend accounts beyond the cartographies of human embodiment.

Materializing Corporeality

The first mode of extending experience is through engagement with the materialities of bodies (and not just those parts of it's materiality that we can grasp through the sensing capacities of our bodies). Emma Roe has done precisely this in her research on food and eating. What post phenomenological geographies do here is expand the realms of what the (experiential) human is, what counts as the empirical field (and how we go about evidencing this), and also what geography is (beginning to employ the knowledges amassed in, e.g., cognitive science or biomedical disciplines). Thinking in these terms has the potential to refigure our understanding of the relation of the body to the world. For example, thinking in terms of stimuli and neuronal firing patterns, or biochemical reactions that occur across the epidermal boundaries of the body offers a different perspective on experience. Rather than understandings based upon objects or individuals, these are system based understandings that delocate and distribute actions transhumanly. Experience in these terms can be understood in terms of flows and feedback and expanded through a consideration of what these non cognitive processes contribute to (human) experience and sensation. Methodologically, this kind of research draws upon cognitive scientists such as Francisco Varela, supplementing personalized first person narrative frames with third person accounts that document muscular reactions, micro scale fleshy movements, and those parts of the body that ordinarily lie outside our sensible frame. This demands accounts that are never solely located in the personal, requiring innovation in research documentation with the use of technologies (such as video) that constitute new forms of visibility.

Affect

Affect offers an increasingly important language for geographers to speak to corporeal changes precipitated by transsubjective experiences (linking material processes to immaterial registers). Affect is a pre personal force; an intensity or experience that is not necessarily tied to the human, but which can become a charged 'background' catalyzing bodily performances and actions. The main theoretical lineage of affect for post phenomenological geographies has been Deleuze, who explicitly rejected phenomenological claims for some form of foundation of sense, on the grounds that the unfolding nature of the world continually unpicks any foundation. Attending to this register (which has become a shorthand for the precognitive and nonrational) shifts (place based) experience into a register less centered on the human, thus enabling geographers to take account of the role of affect in precipitating corporeal formations and subjective emergences. Alongside percepts (another Deleuzian concept), affects are used to account for the constitutive relation between body and world, inside and outside, and self and landscape by John Wylie. What interests him is the ways in which affect and percept fold into the performative creation of self and landscape. This happens through double movements of solicitation and response which give emergent form to subjects, bodies, and specific temporalities or spatialities. Affects and percepts speak to the emergent becoming of the world (in contrast to personally owned perceptions and affections, which serve to situate us in the world, solidifying our being). Mark Paterson similarly uses affect in constructing a 'felt' phenomenology, concerned with touch. He maps the relation between the sensory experiencing body and how this manifests in emotional encounters, tracing both intra and intercorporeal intensities.

Witnessing

Somewhat different suggestions for expanding the phenomenological realm of experience arise when we refigure the representational habits and strategies amassed by the discipline, and the categories geographers use to divide and understand the world. Taking apart the phenomenologically understood space of immediate bodily experience, J. D. Dewsbury argues that the link between the sensible and the intelligible has been intellectually habitualized, but this offers only one way of thinking experience. Dewsbury finds the immaterial realm interesting; in particular those things such as emotions, passions and desires, and spirit, belief, and faith that are both located in the phenomenological and also strain at the edges of human experience. This requires a particular kind of accountability that moves beyond the human and the phenomenologically experiencing subject, and opens up questions such as what is the relation between sense and sensation, how does the sensible come into being, and at what point is sense made? The stance that Dewsbury suggests might answer these questions is that of witnessing; a stance that at once is located in the subject, and unpins from it, continually querying the human located phenomenological horizon. Witnessing is a form of being called into being by the event of the world, and evidencing this through all of the senses. Witnessing never amasses a final picture of the world, but constitutes experience through the almost intangible, the uncertain, and the haunting.

The Spectral

As much as Dewsbury points toward the edges and uncertainties of experience, perhaps the furthest 'outside' of experience that post phenomenological geographers, such as John Wylie, have pushed is that of the spectral or ghostly. The spectral, as the Derridean return of a presence, unsettles certainties (place, temporality, and the certainty of the way we evidence the world through sensation). This offers another, as yet less developed, route for post phenomenological geographies to trouble the being in the world of the phenomenological self.

The Unexperiencible

Another, quite distinct, route offered by post phenomenological geographies is a shift away from the emphasis phenomenology places (mirrored in other geographies of 'practice') upon intentionality and purposive action. Intentionality is central to phenomenology, but constricts the possibilities of a relation between body and world to those parts of the body that are involved in the making of sense (thus requiring the subject as a starting point). As argued by Paul Harrison, this enacts a lack of concern for passivity, and the ways in which the subject might be held in relations of passive exposure – those that bind us in unchosen relations. Harrison queries what happens when you take a different starting point and grant a phenomenological role to the unwilled or the sensuous (as a realm that is not attached to any activity). Suffering is the state of extreme passivity that allows Harrison purchase upon the limits of experience, hinting towards the realm of the unexperiencible (which might itself take on many different dimensions). In this the limits of the phenomenological (as realm of experience), and the limits of phenomenology (in accounting for these realms of experience), are tested, opening up questions such as what might a (post)phenomenology of the unwilled look like, or how might we do a geography of bodily states that do not have any phenomenal content? These questions raise methodological challenges, given the focus of many of the post phenomenological accounts on movement and the moving body, and raises the question of how to research passivity, and how geographers might challenge the (sociocultural) value attached to movement.

Landscape, Corporeality, and Subjectivity

Another area, overlapping with that of experience, in which post phenomenological geographies have been most clearly articulated is in geographies of landscape. Forms of phenomenological understanding situate landscape through the experiencing body, and have thus offered alternatives to the work that has situated landscape as a cultural image to be read in terms of multiple and contested meanings. Post phenomenological approaches have re problematized landscape, returning in some way to ideas of practice and experience but dislocating these from the human body. John Wylie and Mitch Rose suggest that landscape is tension (of presence and absence, performance and perception, proximity and distance, immersion and detachment). They problematize the classical phenomenological move of locating the self in the body and the body in the landscape, asking instead whether the subject is in, of, or on the world, and how we might be all three of these things at once. In this, they trouble notions of locating and embedding, and instead think about landscape in terms of a folding, or an implication of body and landscape. This fleshy form of mutual constitution allows post phenomenological geographers to think about the challenge of landscape to corporeal borders and coherences. A number of geographers have made related post phenomenological interventions around landscape.

Dwelling

Dwelling, often drawn from a Heideggerian lineage via social anthropologist Tim Ingold, offers one way for geographers to negotiate landscape and place. Beginning neither in the subject nor the world, dwelling references togethernesses, foldings, and involvements of human and nonhuman. As such, this return to dwelling has been staged in reference to the desire to negotiate nonhuman agency in the practiced, lived construction of place. It also partly negotiates the perceived inadequacies of actor network theory in dealing with spatially specific locatedness. Dwelling offers one way to negotiate how place is a time deepened ensemble of things and beings (human and nonhuman) together, while at the same time recognizing the more fluid, ephemeral, or dynamic 'becomings' that come to make up place. A notable use of dwelling is that of Paul Cloke and Owain Jones in their study of the nonhuman agency of trees in the performance of an orchard as place. They focus upon the ways in which repeated performances in, and of, landscape concretize place, and work through the ways in which modernity might be destructive to (or merely change) practices of dwelling.

Hayden Lorimer develops ideas of dwelling in slightly different ways from Cloke and Jones, using it to frame his questions about living landscapes. He experiments with the possibilities of narrating the moving geographies of nonhuman animals, and how these might add to our understandings of landscape. In this, he invokes the roles of knowledges on the ground and grounded knowledges, in performative, more than human, creations of place. Weaving historical sources with topographical empirical encounters, Lorimer maps place through energies, forces, and flows; way marking routes through the material and immaterial, permanent, and impermanent. This nods toward an eclectic range of influences, including the psychogeographies of Iain Sinclair and the performance archaeologies of Mike Pearson. As such he suggests ways for post phenomenological geographies to negotiate the tactile and practical and to recognize the importance of vernacular and folk knowledges, and the learned, intimate skill of folding with landscape. Caitlin DeSilvey echoes this approach to the temporalities of dwelling, as she assembles and disassembles the material artifacts and fragmentary objects implicated in the ambiguous processes and practices of dwelling on a Montana homestead. DeSilvey acts as curator, detailing the interdependent lives of individuals, animals, objects, and buildings, and from this, salvaging the near histories of dwelling on the margin. In both these cases, dwelling opens up the doing of landscape beyond the human, and questions processes of inhabiting the world.

Paul Harrison has staged a more theoretical (and less empirical) engagement with dwelling. Here he strips dwelling down to a mode of naming the relation between one and the other, and the event of that relation taking place (through a movement of spacing). Drawing conceptual resources from Levinas, Harrison addresses an erasure within Heidegger's conceptualization of dwelling; that of alterity, or relation to the other. As well documented, otherness made Heidegger uneasy and as such he suppressed all forms of otherness by subsuming difference within sameness. In making this highly political move, Heidegger excludes the ways in which difference interrupts, reveals, and places into question an individual's form of being in the world. Harrison argues that it is necessary to take account of this action of difference in a way that does not simply cast it as an alien threat to the possibility of dwelling, to make dwelling a viable ethico political concept. Thinking the event of the re lation, or the space between individuals in this way has valuable consequences (and indeed is imperative) for thinking otherness and processes of othering (on cultural, ethnic, religious, etc., grounds), and thus for encountering identity, community, and subjectivity. Exploration of this space between us (and how then it might be accounted for) constitutes another route that post phenomenological geographies might take.

Postcolonial Landscapes

The use of postcolonial theories in conjunction with phenomenological theories of vision, embodiment, and locatedness is another line of post phenomenological enquiry. This weaves a passage between lived experience, cultural representation, and the performance of ideolo gies. Marking a move away from the text–context–text analytical loop of representational politics that has situ ated much postcolonial analysis in geography, this instead connects sensory surfaces with processes of meaning making, while looking at practices of colonial spatial inhabitation. In particular, Jessica Dubow has pursued the historical situation of empire as bound up with bodies, experiences, and expressions. Thus colonial figures are corporeal bodies who experience place bound sensations, and become colonizers through their embodied relation to the land. Whereas many postcolonial geographies focus upon the colonial scopic regime as a point of the view on the world, what phenomenological writings do is locate the colonial gaze and body in the landscape. The use of the metaphor of phenomenological groundedness to locate the space based experience of colonial bodies draws critical attention to the ways in which colonial spaces were produced.

Landscapes of Meaning

One line of critique leveled at post phenomenological geographies is developed by Tim Cresswell, who suggests that the separation of bodily experience from con sciousness does little to move on from the Cartesian separation of mind from body. Characterizing the body solely as a realm of ‘nonrepresentation’, or even focusing too heavily upon this nonrepresentational realm, con cretes separations (both conceptual and empirical) be tween representation and practice, both of which he argues are implicated in human movement and experi ence. In claiming that human mobility is both repre sentational and practical, Cresswell expands the realm of the post phenomenological to consider just how experiences and practice are wrapped up in psychosocial processes of meaning making, and are never separate from cognitive processing. This explicit requirement to consider the ways in which bodily practice is tied up with ideological practice (and thus power) is something met in part by Mitch Rose, whose approach asks how subjectivity and meaning become layered with landscape. In asking how subject formations are directed in par ticular ways through particular relations with landscape, Rose considers the ways in which landscape is not encoded with meaning as such, but is rather a surplus of meaningfulness. Rather than tying landscape down to a framework of meaning, Rose understands meanings as affectively and emotionally initiating and provocative. Advocating a suspension of the action of cultural theory in relation to conventions of meaning encoding and decoding, Rose asks what the world can possibly tell us. How do particular subject formations occur, and what is the act through which we become separate enough from landscape in order to understand ourselves as an ‘I’?

Future Directions: Doing Post-Phenomenological Geographies

While post phenomenological geographies are presently not particularly cohesive, this leaves open a wider variety of paths for future research. This might be in terms of empirical focus, theoretical exposition, and methodological focus. The methodological challenge is using methods that clearly and habitually situate the human at the center, and adapt them so they allow access to the inhuman forces that are the subject of post phenomenological geographies. There have been a number of strategies in this.

Experiential Research

The first of these is the empirical focus upon spaces and practices in which the action of these forces is heightened, and thus more accessible. Thus, these forms of research have dwelt upon body–body contact (through practices such as massage or Reiki for instance), bodily movement (e.g., focusing upon walking, dancing, or cyc ling), and body–object relations (through practices such as eating, or becoming intoxicated). Written documen tation (in accordance to the demands of the academy) is important, and the question of how to narrate these geographies underlines the questions of subjectivity and corporeality that the post phenomenological geog raphers documented here have raised. An interesting tension exists in much of this research because it works from individual ethnographic experience in order to speak to these transhuman forces and impulses. Experiential research and personal testimony are used to appeal directly to the forces of the inhuman, and the ways in which the unfolding of the world draws the subject in ways not immediately of their own choosing.

Current worries about this mode of accounting for the world have arisen due to the lack of attention paid to the more traditional modes of accounting for the subject (class, gender, identity, etc.). What this research does however, is to try to understand how the more than human parts of the world contribute to the performance, maintenance, and solidification of sociocultural lines of gender, class, and identity, rather than taking these as 'natural' starting points of analysis. They also attempt to underline that the narrative does not come from an unaffected, unaffecting coherent subject. Rather, a point of view as narrator is an outcome (rather than a preexistent position from which to speak) of the practice being researched. Other possibilities have been explored, such as Lorimer's imaginative mobilizations of historical sources, or Paterson's use of the body as a recording surface upon and through which the traces of the research practice are written.

Documentation

The second is experimenting with the documentation of these post phenomenological geographies. Roe, in her research concerned with the moving material nonhuman, questions this becoming post phenomenal through the realm of movement. Practical activities documented through videoing are envisioned as a methodological resource by Roe, as it is in movement and practice (rather than meaning) that sensation meets flesh, and the realm of experience is located. Roe is particularly interested in the event as an ontological device (the event as bodies coming together, where 'bodies' are not just restricted to the human), and what kinds of experience we might excavate in the post phenomenological spaces of the body (i.e., those spaces and processes that are beyond our sensate cognition – biochemical reactions, etc.). Video has been a strategy also adopted by Eric Laurier, who, although positioning his approach as ethnomethodological rather than post phenomenological, has experimented with the use of video to document the production of everyday social conventions, without necessarily grounding these in consciousness, or locating them in the human body. His distributed attention has so far been focused upon such mundane (but vital) practices as dog walking, drinking coffee, and driving cars.

These methodological directions, combined with the already flourishing empirical and theoretical post phenomenological geographies outlined above, suggest a number of fruitful spaces for future engagement. The diverse directions which have been outlined, are united by the phenomenological concern with the embodied self in the world negotiated in relation to the fragmented and fragmenting subject at issue in post structuralist theories, giving geographers insights into themes such as subjectivity, corporeality, landscape, and place.

See also: Affect; Body, The; Dwelling; Ethnomethodology/Ethnomethodological Geography; Non-Representational Theory/Non-Representational Geographies; Performance, Research as; Phenomenology/Phenomenological Geography; Place; Posthumanism/Posthumanistic Geographies.

Further Reading

Cloke, P. and Jones, O. (2003). Dwelling, place and landscape: An orchard in Somerset. *Environment and Planning A* 33, 649 666.
Davidson, J. (2000). A phenomenology of fear: Merleau Ponty and agoraphobic life worlds. *Sociology of Health and Illness* 22, 640 660.
DeSilvey, C. (2007). Salvage memory: Constellating material histories on a hardscrabble homestead. *Cultural Geographies* 14, 401 424.
Dewsbury, J. D. C. (2003). Witnessing space: Knowledge without contemplation. *Environment and Planning A* 35, 1907 1932.
Dubow, J. (2000). From a view on the world to a point of view in it: Rethinking sight, space and the colonial subject. *Interventions: International Journal of Postcolonial Studies* 2, 87 102.
Harrison, P. (2007). The space between us: Opening remarks on the concept of dwelling. Environment and Planning D 25, 625 647.
Harrison, P. (2008). Corporeal remains: Vulnerability, proximity and living on after the end of the world. *Environment and Planning A* 40(2), 423 445.
Ingold, T. (2000). *The Perception of the Environment: Essays in Livelihood, Dwelling and Skill*. London: Routledge.
Lorimer, H. (2006). Herding memories of humans and animals. *Environment and Planning D* 24, 497 518.
Paterson, M. (2006). Affecting touch: Towards a 'felt' phenomenology of therapeutic touch. In Davidson, J., Bondi, L. & Smith, M. (eds.) *Emotional Geographies*, pp 161 176. London: Ashgate.
Roe, E. (2006). An aesthetic methodology: More than visual readings of video based data. Available from author School of Geography, University of Southampton, Highfield, Southampton, SO17 1BJ, UK.
Rose, M. (2006). Gathering 'dreams of presence': A project for the cultural landscape. *Environment and Planning D* 24, 537 554.
Rose, M. and Wylie, J. (2006). Guest editorial: Animating landscape. *Environment and Planning D* 24, 475 479.
Wylie, J. (2006). Depths and folds: On landscape and the gazing subject. *Environment and Planning D: Society and Space* 24, 519 535.
Wylie, J. (2007). *Landscape*. London: Routledge.

Post-Productivist and Multifunctional Agriculture

G. A. Wilson, University of Plymouth, Plymouth, UK

Glossary

Diversification Process associated with agricultural actors moving away from a sole focus on agricultural production toward recognition of multiple uses of agricultural and rural resources.

Multifunctionality Territory bounded by agricultural/rural productivist and nonproductivist action and thought characterized by multiple pathway opportunities.

Multifunctionality Spectrum A spectrum of decision making that explains action and thought of rural actors in the context of productivist and nonproductivist pathways.

Nonproductivism The opposite of productivism in the multifunctionality spectrum of decision making; action and thought associated with agricultural practices where maximization of food and fiber production is 'not' important.

Pluriactivity Farm households engaging in several activities to differentiate income streams.

Productivism Agricultural action and thought associated with maximum food and fiber production based on Fordist modes of production and industrial–agricultural systems.

Post-Productivism The era 'after productivism' associated with agricultural and rural spaces characterized by reduced emphasis on maximizing food and fiber production where nonagricultural uses of rural space assume greater importance.

Introduction

Two key arenas of concern have influenced recent re search and thinking on conceptualizations of agricultural and rural change. The first relates to the proposed transition from a 'productivist' to a 'post productivist' agricultural regime (the p/pp transition), and the second revolves around the notion of 'multifunctional agri culture'. Although there is growing criticism of linear concepts surrounding the postulated p/pp transition, it is still assumed by many that agricultural regimes have moved toward a post productivist era. Debates on this transition have brought together a wide array of researchers from different disciplinary backgrounds, including rural studies, history, sociology, economics, or social psychology, with human geographers being at the forefront of debates.

This article reviews current conceptualizations of post productivist and multifunctional agriculture and provides a critical evaluation of both terms. Debates on the p/pp transition are closely intertwined with recent conceptualizations of the notion of multifunctional agriculture and none of these conceptual constructs can be understood without reference to the other. Although some continue to advocate the utility of post producti vism, many are now criticizing this transitional concept and argue that the notion of nonproductivism may pro vide a better heuristic opposite to productivism. Under standing productivist/nonproductivist action and thought as a 'spectrum' of multiple decision making pathways enables the conceptualization of multifunctionality as the conceptual territory bounded by the two extreme agri cultural transitional pathways of productivism and non productivism – a space that, compared to the p/pp transition model, better encapsulates the temporal non linearity, spatial heterogeneity, global complexity, and structure–agency inconsistency that characterizes agri cultural and rural decision making across the globe.

The Transition from Productivist to Post-Productivist Agriculture

Agriculture in Crisis

Over the last 50 years, global agriculture has witnessed profound changes with regard to technological inno vations and food and fiber production techniques, actor spaces, policy frameworks, food regime structures, ideologies, and impacts on the environment. Food pro duction has increased dramatically so that today more than 6 billion people can, in theory, be adequately fed. However, many agricultural systems have also seen dra matic structural upheavals, including the rapid decline in agricultural workforce in most advanced economies with concurrent destabilization and restructuring of rural communities. This has been accompanied by the substi tution of subsistence farming patterns with cash crop production for external markets in many regions of the developing world, and the dismantling of the often sus tainable traditional farming systems. In many areas, this has occurred simultaneously with an increase in the severity and extent of environmental degradation linked to over intensive agricultural practices and production techniques.

In recent years, much theoretical and conceptual work has attempted to understand these changes in modern agricultural regimes and rural areas, both in developed and in developing countries. One of the most powerful theoretical concepts to emerge from these debates has been the notion that modern agricultural regimes have moved from a 'productivist' to a 'post productivist' era.

The Transition to Post-Productivism

In recent years debates on the p/pp transition have added an interesting new conceptual dimension to agricultural/rural research – a field that was, for a long time, seen as having relatively static theorizations of agricultural and rural change. Agricultural productivism and post productivism are relatively recent terms, with debates emerging in academic circles in the UK in the early 1990s. Discussions were spearheaded by some of the key thinkers on agricultural change at the time including Terry Marsden, Philip Lowe, Paul Cloke, Richard Munton, Keith Halfacree, Jonathan Murdoch, and Neil Ward. These authors set the scene for a burgeoning debate on the p/pp transition that has continued unabated to the present day.

In advanced economies, it has been argued that the transition to a post productivist era began in the 1980s and has lasted to the present day. Post productivism has been generally seen as the mirror image of productivism – in other words, post productivism has been defined in the negative as what it is not, rather than as what it may be. Post productivist agriculture is characterized by a reduction in the intensity of farming through extensification, diversification, and dispersion of agricultural production with an associated move away from agricultural production toward consumption of the countryside. This has been associated with the loss of the central position of agriculture in a society characterized by contested countrysides, a widening of the agricultural community to include formerly marginal actors at the core of the policy making process, and a weakening of the state role in policy making with a more inclusive model of governance that also includes grassroots actors. Farming techniques in the post productivist era are seen to be more in tune with environmental protection through reduced application (or total abandonment) of external inputs (e.g., organic farming), while during the post productivist era the main threats to the countryside are generally perceived to be agriculture itself rather than urban or industrial development. Although conceptualizations of the p/pp transition have been largely restricted to the situation of the UK, some studies have attempted to assess how applicable this transition is elsewhere, in particular, in Australia, New Zealand, and the European Union. As a result, reference to the p/pp transition is now well established in the agenda of rural geography, prompting some to suggest that it has become one among the accepted orthodoxies for conceptualizing recent agricultural change.

For developing countries, meanwhile, work on famine has led to increased nuancing and locally grounded definitions of food security. By the end of the 1990s a post productivist view may have emerged for some developing countries that has emphasized the multiple and shifting ways with which individuals and households achieve food security. Partly reflecting notions of the p/pp transition in the developed world, parallel shifts can be recognized in debates surrounding agricultural change in developing countries, in that food security in the poorer South is seen to have moved from the global and the national to the household and the individual, from a food to a livelihood's perspective, from objective indicators to subjective perception, and from production to consumption of the countryside.

Deconstructing Post-Productivism

Several years after the first widely publicized discussions of a shift toward post productivism, and at a time when further profound changes in agricultural/rural arenas at the local, national, and global scale are taking place, many argue that it is time to reevaluate existing conceptualizations. In particular, there is an inherent danger in uncritically accepting seemingly straightforward and linear concepts of societal change as embodied by the p/pp transition (**Figure 1**). There is increasing concern that the p/pp model is not robust enough to explain agricultural change – be it from the perspective of developed or developing countries. Indeed, echoing many scientific critiques of other ism/post ism debates

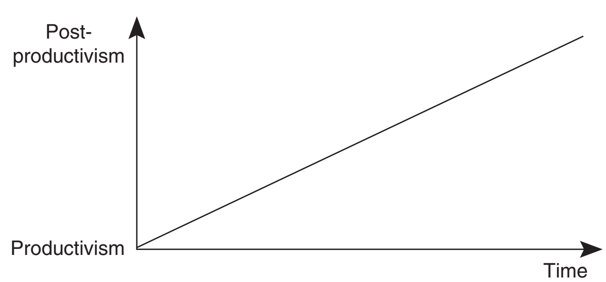

Figure 1 Temporal linearity and the productivist/post-productivist transition. Source: Wilson, G. A. (2007). *Multifunctional Agriculture: A Transition Theory Perspective*. 111p. Wallingford, CT: CAB International.

(e.g., transition to post Fordism, post modernism or post socialism), evidence increasingly suggests that the so called transition toward post productivism is characterized by temporal nonlinearity, spatial heterogeneity, global complexity, and structure agency inconsistency. These recent debates, therefore, make the notion of a shift toward post productivism increasingly difficult to substantiate. Some argue, therefore, that – marred by issues of UK centrism, presentism, an overemphasis of political economy conceptualizations, and confusion about linearity and homogeneity – the p/pp transition no longer stands up to critical scrutiny as a model adequately explaining contemporary transitional processes in agriculture.

The problem of the implied linearity of the p/pp concept has partly been linked to the fact that those who conceptualized the term post productivism largely failed (until recently at least) to take into account the wealth of actor oriented and behaviorally grounded research. An inclusion of such research points toward considerable differences in productivist and post productivist action and thought at different levels of actor spaces. Grassroots actors, for example, have remained largely productivist in their agricultural occupational identities (i.e., not all stakeholder groups are moving along the p/pp transition at the same pace and in the same manner as **Figure 1** would imply). Some argue, therefore, that we witness a spectrum of different views rather than two easily definable and separate conceptual entities. Acknowledging findings from actor oriented and behaviorally grounded studies suggest that the breakdown of productivist attitudes and ideologies have 'not' yet occurred, and that agriculture should be seen as an increasingly 'bifurcated' process, comprising 'both' productivist and post productivist action and thought.

The problem also partly stems from the fact that the p/pp transition has been conceptualized from a UK centric perspective that has largely failed to discuss whether, and to what extent, the concept has a wider applicability elsewhere. In particular, there is growing evidence of a time lag and 'spatial' inconsistencies in the adoption of post productivist action and thought across the world, as different localities are positioned at different points in the p/pp transition. This 'territorialization' of productivist and post productivist actor spaces highlights the wide ranging diversity that exists within agricultural transitional processes. That such productivist and post productivist action and thought occur in 'multidimensional coexistence', therefore, has led researchers to increasingly question the implied directionality of the p/pp model. This means that post productivism has 'not' replaced productivism. Indeed, it appears that both processes 'coexist' temporally, spatially, and structurally. The postulated transition to post productivism, therefore, should not imply that productivist institutional forms, networks, ideologies, and norms have been superseded, suggesting that post productivism has not been radical, but rather incremental and accommodationist to productivist action and thought.

Some have, therefore, rightly asked whether the notion of post productivism may only be a conceptual construct that describes patterns at the macro structural level, or that it may even be an academic myth. Like any terminology beginning with post, the notion of post productivism indicates a hesitance by those who propelled the term into the academic terminology to create a new term linked to problems of dualistic theorizations of societal transition. From a theoretical perspective, critics, therefore, increasingly encourage the search for new tools to explore the nature and dynamics of the conceptual ground in between the extreme pathways of productivism and post productivism. Yet, the main conceptual problem rests less with the notion of productivism, but with the concept of 'post' productivism (especially with the prefix post) and the suggestion of an irreversible transition from one organizational state to another. A reconceptualization of the p/pp transition model has long been overdue if this model (or parts thereof) is to have wider scientific credibility in the future.

Multifunctional Agriculture and the Productivism/Nonproductivism Spectrum

Conceptualizing 'Nonproductivism'

The relative failure of the reductionist p/pp transition model to explain contemporary agricultural change means that the continued use of the term post productivist is untenable. Post should imply something following after something else, and the mounting evidence does not support the assumption that a coherent and conceptually well defined era has followed on from the productivist era. There are, therefore, increasing calls that agricultural action and thought should be viewed as occurring along a temporally nonlinear and spatially heterogeneous spectrum of decision making possibilities, which negates the need to define distinctive boundaries and acknowledges the fluidity and flexibility of agricultural action and thought. Recent work suggests that the notion of 'nonproductivism' provides a better conceptual term to describe the opposite end of productivism (referred to as the 'productivist/nonproductivist transition model'; hereafter the p/np transition model). Nonproductivist action and thought contains many of the indicators that have been used to conceptualize post productivism in the past (see above), but without the underlying assumption that there is necessarily a temporally linear movement toward such action and thought over time.

One of the key advantages of the notion of non productivism is that, as opposed to post productivism, it is a neutral term, simply denoting action and thought that is not productivist. This helps us get away from the danger of subjectivity that has hitherto branded pro ductivism as the bad and post productivism as the 'good' era. Embedding both productivism and nonproductivism as part of a spectrum allows recognition that there is no action and thought affecting agricultural and rural areas that is entirely productivist or nonproductivist, but that each process may contain elements of both. Transition should be seen as occurring along and within different pathways of constraints and opportunities within the p/np spectrum over time, rather than being conceptually restricted to a transition from one set of seemingly structured coherences (i.e., productivism) to another (i.e., post productivism).

Multifunctional Agriculture

In its simplest definition, the notion of 'multifunctional agriculture' implies an agriculture that has multiple functions, many of which go well beyond the production of food and fiber as its central objective. Multifunctionality is the antithesis of monofunctionality, where agriculture serves relatively narrow functions largely focused on food and fiber production. Yet, many debates on multi functionality have narrowly focused on economistic, pol icy based, and structuralist interpretations that place great emphasis on economics and the state (i.e., the political economy) as key drivers of multifunctional transitions. In part, this is a legacy of structuralist interpretations that dominated conceptualizations of agricultural and rural change during the 1990s. It is only in the last few years that wider issues linked to culture, the consumption of countryside, and society's changing views of agriculture have been incorporated into conceptualizations of multi functionality, and discussions on the importance of multifunctional identities of grassroots actors and their aspirations have only recently become the focus of at tention. The predominance of structuralist interpretations has led to a relatively narrow view of multifunctionality that focuses more on processes, rather than acknowledging that multifunctionality should be understood as an over arching concept that can describe 'and' explain con temporary agricultural change.

Multifunctional agriculture discourses emerged syn chronously within both policy based and academic discourses, referred to as the 'institutionalization of multifunctionality'. This has possibly negated the urgency for thorough scientific theorization of the term due to its early acceptance by policy making circles as a notion with seemingly 'real' policy relevance. Until recently, academic discourse has thus been influenced by political discourse rather than itself influencing the politics of multifunctionality. As a result, multifunctionality has been used as a pretext and smokescreen to defend pro tectionist policies (especially in the EU), and one of the most important attributes of the multifunctionality concept has, therefore, been its malleability in the face of policy considerations.

Until recently, this politicization of multifunctionality has led to a theoretically weak concept. As a result, the largely European based notion of multifunctionality has, so far, not required a holistic, broad based, and theoretically/academically well informed vision. Instead, multifunctionality has largely been relegated to being a short term and pragmatic tool and a mechanism for shallow policy decision making – in other words, a quick fix notion that lives an uneasy existence between political practicality and academic unease as a term that has not yet convinced at the conceptual and theoretical level. Further, contemporary conceptualizations of multi functionality have also been marred by both sectoral theoretical interests within relatively narrow disciplinary confines (also involving human geographers) and by the predominance of structuralist political economy inter pretations of the term that have largely neglected post structuralist and cultural turn related emphases on grassroots stakeholder actions and thought. Some have argued, therefore, that the notion of multifunctional agriculture can only be fully understood by taking into account existing debates and theorizations of agricultural change and transition. Some of these debates suggest that only by theoretically anchoring the concept of multi functional agriculture in transitional frameworks, such as the p/np spectrum outlined above, is it possible to pro vide a holistic and normative understanding of what multifunctionality should be about.

Multifunctional Agriculture and the Productivism/Nonproductivism Model

Until recently, few debates have shed light into what the notion of multifunctionality implies, who the bene ficiaries should be, and how it ought to be implemented into practice – in other words, the notion of multi functionality has remained undertheorized, process oriented, and poorly linked to wider debates in the social sciences. This has led some to call for improved theorization of the notion of multifunctionality, in par ticular, in the context of extensive debates on agricultural transition. Recent work suggests that the conceptual territory of multifunctionality can be theoretically anchored in the p/np model (see above), and that multifunctionality occupies the (conceptual and real) space between the extreme pathways of productivism and nonproductivism – referred to as the 'multi functionality spectrum' (**Figure 2**). Contrasting most

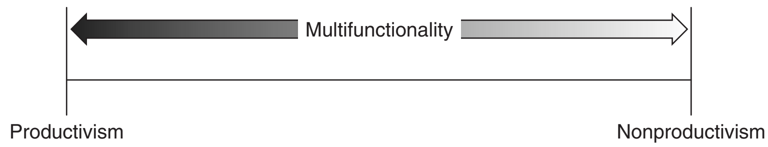

Figure 2 The multifunctionality spectrum bounded by productivist and nonproductivist action and thought. Source: Wilson, G. A. (2007). *Multifunctional Agriculture: A Transition Theory Perspective*. 214p. Wallingford, CT: CAB International.

current conceptualizations of multifunctionality as a static state, this new view of multifunctionality sees it as a dynamic transitional process occurring within the boundaries of the productivist and nonproductivist decision making spectrum. Contrary to the binary notions of productivism and post productivism, the multi functionality spectrum, therefore, enables the under standing of multiple actions and processes simultaneously in a nonlinear and spatially heterogeneous way.

The underlying assumption of this new understanding of multifunctionality is that almost any agricultural/rural action and thought will contain elements of 'both' pro ductivism and nonproductivism – in other words, that few actions, thoughts, and development pathways can be described as solely productivist or nonproductivist. The advantages of this view are that it allows the exact positioning of any actor, actor space, or process along a spectrum of decision making based on the various dimensions making up productivist and nonproductivist spaces. Conceptualizing multifunctionality as part of a spectrum acknowledges the complexity and multi causality underpinning agricultural change. This new concept can be applied in any location around the world, as it sees policy change as only one of many components of multifunctionality, and is, therefore, largely independent of the history of policy making trajectories specific to a given territory (e.g., the EU) that has hitherto marred the exporting of the European multifunctionality concept to non European geo graphical contexts. The new concept can also be invoked to explain processes in any farming system (both with productivist and with nonproductivist tendencies) and is not limited to traditional forms of agriculture. Most importantly, the multifunctionality spectrum no longer needs to be associated with a Euro or Northern centric policy discourse but can be applied equally to both the developed and the developing world.

A Normative View of Multifunctionality

The new multifunctionality concept implies normative judgments about the value and direction of multi functionality. In other words, multifunctionality can and should not be a neutral term, but has to be accorded a value and positionality along the p/np decision making spectrum. Defining multifunctionality is, therefore, about accepting that 'priorities' need to be set with regard to specific actions and processes influencing agricultural (and rural) development pathways. Recent commentators have suggested that the multifunctionality spectrum is characterized by varying quality based on 'weak', 'mod erate', and 'strong' multifunctionality. Weak multi functionality is the worst type of multifunctionality close to the productivist decision making pathway, while strong multifunctionality is the ideal model that all societies should be striving for. In this view, strong multifunctionality engenders synergistic mutual benefits between individual actions and processes (e.g., landscape protection and food quality, direct marketing and rural community embeddedness, etc.) and, therefore, is good for the environment, farmers, rural social relationships and governance structures, food quality, and agriculture–society interactions in general. Strong multifunctionality is, therefore, predicated on strong natural, social, moral, and cultural capital.

Figure 3 shows how this concept of weak, moderate, and strong multifunctionality can be applied to describe individual development pathways of individual farms or entire agricultural regimes where multifunctional quality may fluctuate over time (pathways a–h in the figure). Strongly multifunctional transitional pathways are characterized by action and thought located close to the nonproductivist pathway boundary of the multi functionality spectrum, with a strong emphasis on public good characteristics and strong social, economic, cultural, moral, and environmental capital. High environmental sustainability plays a key role in the strong multi functional model, based on highly diverse assemblages of crops and livestock, often akin to palaeotechnic types of agricultural production. Actors in the strongly multi functional agricultural regime will show strong tenden cies for local and regional embeddedness, in particular, through horizontally integrated rural/farming com munities with very close interaction between local communities through reciprocal rural–agricultural rela tionships characterized by strong lateral actor linkages and newly empowered stakeholder groups. This reson ates closely with human geography conceptualizations of spaces of hope as alternative economic spaces, whereby newly assembled (or rediscovered) horizontal platforms of action are (re)created out of local agricultural/rural systems. Such embedded social networks particularly include activities that will help provide new income and employment opportunities for the agricultural sector

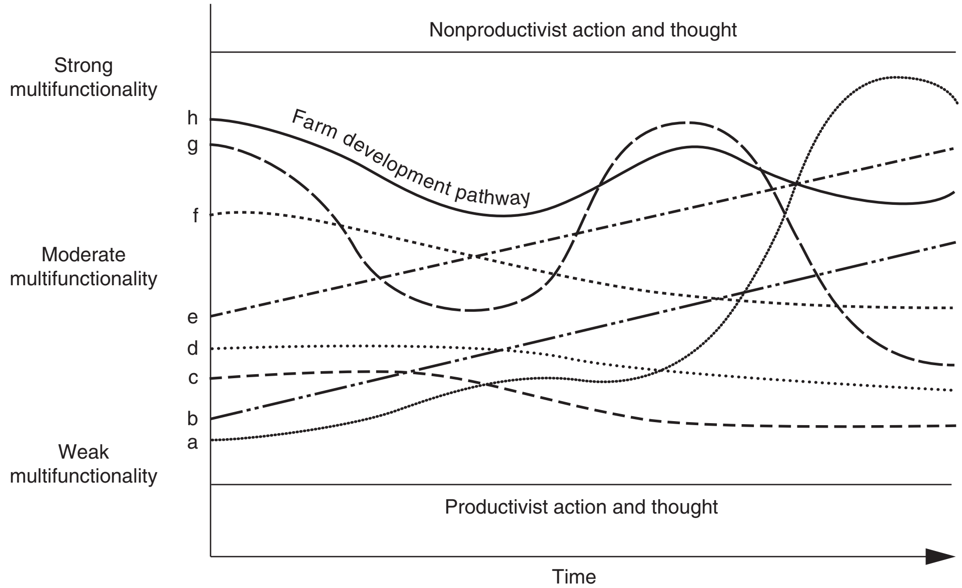

Figure 3 Farm development pathways and the multifunctionality spectrum. Source: Wilson, G. A. (2007). *Multifunctional Agriculture: A Transition Theory Perspective*. 284p. Wallingford, CT: CAB International.

(i.e., new or strengthened relationships between the agricultural sector and rural society), as well as so called 'associational interfaces' that are both informal but highly significant in establishing trust, common understandings, working patterns, and different forms of cooperation and cooptation between different stakeholder groups in the food supply chain. In countries with ethnic minorities living in rural areas (e.g., Australia, New Zealand, Canada) local embeddedness in strongly multifunctional agricultural regimes will also mean empowerment and inclusion of 'ethnic groups' (e.g., Australian Aborigines) in agricultural/rural decision making processes.

Strongly multifunctional systems will also be charac terized by short food chains and high(er) food quality associated with more differentiated food demand by consumers (the quality turn), the capacity to resocialize or respatialize food, a demand for food products with high (often regionally based) symbolic characteristics, the creation of additional value for rural regions, and enlightened visions about the links between food and health focusing on the interlinkages between both human and environmental health. Strongly multifunctional food consumption shows a shift toward healthy diets (e.g., Mediterranean foods) predicated on food quality and good nutritional balance. Strongly multifunctional systems will also have low farming intensity and productivity, in most cases characterized by a reluctance to use Green Revolution or genetically modified crops. There will also be high degrees of diversification and pluriactivity, in particular, away from traditional pro ductivist food and fiber production, leading to a revaluation of existing farm household knowledge (e.g., of women and young people) and the need to develop new skills and professional abilities. Strongly multifunctional farming systems are also likely to embark on diversifi cation pathways that move the farm enterprise away from a singular focus on food and fiber production. Some also suggest that strongly multifunctional rural spaces are more likely to be only weakly integrated into the global capitalist market, as only partial or complete disengage ment from global capitalist (productivist) networks and agriculture liberalization processes will enable on farm implementation of strong multifunctionality pathways. Further, strong multifunctionality will also imply that substantial mental changes are taking place among vari ous stakeholder groups, in particular, through open minded farming and rural populations who see farming and agriculture as processes that go well beyond pro ductivist food and fiber production, mass production, and profit maximization. A strongly multifunctional agriculture can also be seen to act as a bridge between agricultural practices and the wider community (in par ticular, disadvantaged groups in rural society). Finally, strongly multifunctional systems will also be closely associated with open minded societies who accept that the nature of farming and agriculture is in the process of change. Overall, the strongest level of multifunctionality can be achieved if all of the above processes and activities occur simultaneously (which will rarely be the case), based on self reinforcing multiplier effects of multi functional activities.

Commentators nonetheless suggest caution about not falling into the trap of romanticizing one multi functionality model over another. Historically, for example, it would be wrong to argue that for most of the time since the inception of farming, agricultural regimes operated on the basis of the strong multifunctionality model. Indeed, the question has been asked whether the strong multifunctionality model in one area may be predicated on the coexistence (temporally and spatially) of moderate and weak multifunctional agricultural regimes in other areas. In other words, it is the aim to achieve strong multifunctionality possible and based on a win–win situation, or is it possible to conceive of a 'zero sum game' situation in which the existence of one multifunctionality regime is predicated on the existence of another? It may well be that in the long term eco nomic efficiency, the survival of farming systems, and the ability to feed the world may be predicated on pulling all agricultural systems increasingly toward strongly multifunctionality pathways.

Conclusions

This discussion has shed light on the notions of agri cultural productivism and post productivism and has highlighted how commentators have increasingly called for a deconstruction of the notion of a linear p/pp transition. Instead, studies increasingly suggest that a revised model based on a spectrum of decision making bounded by productivist and nonproductivist pathways better encapsulates the nonlinearity, heterogeneity, and complexity that characterize agricultural transitions. Although the p/pp transition model has made a vital contribution to current debates on agricultural transition, it is suggested that post productivism cannot, and should not, be seen as the end point of contemporary agricultural change. Instead, a modified transitional model based on the notion of a productivist/non productivist spectrum of decision making forms a more appropriate nonlinear conceptual model of agricultural change that also acknowledges spatially heterogeneous processes and structure–agency inconsistencies in agri cultural change.

Debates on this p/np transition are closely inter twined with the notion of multifunctional agriculture. While many existing understandings of multifunctional agriculture see it largely as an economic or policy based process that only provides a limited basis with which to understand agricultural change, recent commentators are arguing that multifunctionality should be understood as an overarching normative concept anchored in a revised productivist/nonproductivist spectrum of de cision making. Thus, discussions about multifunctionality have increasingly moved away from positive to normative theory, implying the need for subjective value judgments about the quality of multifunctional decision making pathways. Conceptualizations of weak, moderate, and strong multifunctionality pathways suggest that a value judgment can be taken with regard to 'better' and 'worse' agricultural development pathways, and that the strong multifunctionality model, closely associated with non productivist action and thought, is qualitatively and morally the best pathway to follow and implement.

See also: Agriculture, Sustainable; Agri-Environmentalism and Rural Change; Conservation and Ecology; Environmental Policy; Food Networks; Food Regimes; Fordism; Global Commodity Chains; Rurality and Post-Rurality.

Further Reading

Brouwer, F. (ed.) (2004). *Sustaining Agriculture and the Rural Environment: Governance, Policy and Multifunctionality*. Cheltenham: Edward Elgar.
Burton, R. J. (2004). Seeing through the 'good farmer's' eyes: Towards developing an understanding of the social symbolic value of 'productivist' behaviour. *Sociologia Ruralis* 44(2), 195 215.
Clark, J. (2006). The institutional limits to multifunctional agriculture: Subnational governance and regional systems of innovation. *Environment and Planning C: Government and Policy* 24, 331 349.
Evans, N., Morris, C. and Winter, M. (2002). Conceptualizing agriculture: A critique of post productivism as the new orthodoxy. *Progress in Human Geography* 26(3), 313 332.
Goodman, D. (2004). Rural Europe redux? Reflections on alternative agro food networks and paradigm change. *Sociologia Ruralis* 44, 3 16.
Halfacree, K. (1997). Contrasting roles for the post productivist countryside: A postmodern perspective on counterurbanisation. In Cloke, P. & Little, J. (eds.) *Contested Countryside Cultures: Otherness, Marginalisation and Rurality*, pp 70 91. London: Routledge.
Holmes, J. (2002). Diversity and change in Australia's rangelands: A post productivist transition with a difference? *Transactions of the Institute of British Geographers* 27, 362 384.
Ilbery, B. and Bowler, I. (1998). From agricultural productivism to post productivism. In Ilbery, B. (ed.) *The Geography of Rural Change*, pp 57 84. Harlow: Longman.
Marsden, T. (2003). *The Condition of Rural Sustainability*. Assen (NL): Van Gorcum.
Marsden, T., Murdoch, J., Lowe, P., Munton, R. and Flynn, A. (1993). *Constructing the Countryside*. London: UCL Press.
Mather, A. S., Hill, G. and Nijnik, M. (2006). Post productivism and rural land use: Cul de sac or challenge for theorization? *Journal of Rural Studies* 22, 441 455.
McCarthy, J. (2005). Rural geography: Multifunctional rural geographies reactionary or radical? *Progress in Human Geography* 29(6), 773 782.
Potter, C. and Tilzey, M. (2005). Agricultural policy discourses in the European post Fordist transition: Neoliberalism, neomercantilism and multifunctionality. *Progress in Human Geography* 29(5), 1 20.
Pretty, J. N. (1995). *Regenerating Agriculture: Policies and Practice for Sustainability and Self Reliance*. London: Earthscan.
Robinson, G. (2004). *Geographies of Agriculture: Globalisation, Restructuring and Sustainability*. Harlow: Pearson.
Van Huylenbroek, G. and Durand, G. (eds.) (2003). *Multifunctional Agriculture: A New Paradigm for European Agriculture and Rural Development*. Aldershot: Ashgate.
Walford, N. (2003). Productivism is allegedly dead, long live productivism: Evidence of continued productivist attitudes and decision making in south east England. *Journal of Rural Studies* 19, 491 502.

Ward, N. (1993). The agricultural treadmill and the rural environment in the post productivist era. *Sociologia Ruralis* 33, 348 364.
Wilson, G. A. (2001). From productivism to post productivism … and back again? Exploring the (un)changed natural and mental landscapes of European agriculture. *Transactions of the Institute of British Geographers* 26, 77 102.
Wilson, G. A. (2007). *Multifunctional Agriculture: A Transition Theory Perspective*. Wallingford: CAB International.
Wilson, G. A. and Rigg, J. D. (2003). 'Post productivist' agricultural regimes and the South: Discordant concepts? *Progress in Human Geography* 27, 681 707.

Post-Socialist Cities

L. Sýkora, Charles University, Prague, Czech Republic

Glossary

Centrally Planned Economy An economy in which the allocation of resources and investments is decided by the hierarchically organized decision-making system with the main power held by the central administration (alternative term is 'command economy').

Post-Socialist City Specific, temporarily, existing type of transitional city characterized by the adaptation of inherited socialist urban environment to new political and economic conditions given by the political, economic, and cultural transformations to capitalism.

Post-Socialist Transition A broad, complex, and lengthy process of societal change starting with the refusal of communist regimes and central planning, leading to democratic political regimes and a market economy.

'State Socialism' or 'Communism' A societal system characterized by the common ownership of the means of production and the administrative allocation of resources. Because of the totalitarian nature of communist regimes, the term 'socialism' is sometimes replaced by the terms state 'socialism' or 'communism'.

Socialist City A specific type of city significantly shaped by socialist political and economic regimes.

Spontaneous Transformations Transformations of economic, social, and cultural environments that develop spontaneously within the context of democratic policymaking and international market economy.

Systemic Transformations Government-controlled transformations of political and economic system, such as the establishment of democratically elected political representations, institutional de-regulation, price liberalization, privatization of state assets, and liberalization of foreign trade.

Transformation Processes Transition consists of multiple transformation processes that are related in certain time sequence, address universal or more specific areas, and include both the political application of normative concepts as well as spontaneously proceeding societal changes within (re)established market environment.

Introduction

The relatively gradual processes of urban change are sometimes interrupted by periods of more radical urban restructuring influenced by turbulent developments in society. The principles on which societies are founded can be substantially changed and a massive remodeling of social and spatial organization follows. A recent example of such changes are the urban transformations in post socialist countries conditioned by the transition from communist totalitarian to democratic political regimes and from centrally planned (command) to market economies.

Even though the principles behind the production of urban environments change quickly, the pace of changes in the built environment, land use patterns, and residential differentiation are much slower. The key aspect for the understanding of post socialist urban change is the distinction between the short term period when the basic principles of political and economic organization are changed, the medium term period when peoples' behavior, habits, and cultural norms are adopted to new environment and transformations in a number of spheres begin to effect broader societal change, and the long term period in which more stable patterns of urban morphology, land use, and residential segregation are reshaped.

The post socialist transition is a broad, complex, and lengthy process of societal change. It starts with the refusal of a communist regime and central planning. A society with democratic political regime and market economy is generally accepted as a desired destination. This general societal transition is realized through multiple transformation processes that are highly contingent upon specific historic, cultural, as well as place specific contexts. Despite certain common principles, the starting conditions, paths, and processes of change as well as its outcomes vary and any account of transition from socialism to capitalism must be aware of both commonalities and dissimilarities. The post socialist urban transformations substantially differ between groups of post socialist countries as well as between cities within individual countries due to their position within an urban hierarchy and regional economic geographies.

This article mainly builds on the experience and situation in major capital cities of post socialist Eastern Central Europe, since these are the cities that have been studied the most. The generalization based on these cities may also be relevant for other post socialist cities.

Socialist City

Cities in former socialist countries were shaped by socialist principles adapted by the ideology of leading communist parties. The nature of socialist, political, and economic systems was radically distinct from capitalism. As a consequence, the impact on the built environment, land use, and residential differentiation yielded distinct patterns which contrast with capitalist urban develop ment. Socialism was implanted by a radical reform to the preexisting social and cultural milieu as well as to the built environment. Most cites developed prior to socialism under feudal and capitalist conditions. The exceptions were new towns erected under the communist regime and that were numerous especially in the former Soviet Union. The urban environment itself differed because individual socialist countries were in strikingly different phases of industrialization and urbanization process. Furthermore, socialism itself was not uniform. It varied across regions and so did its influence on urban development.

The main political aim of the socialist system was to create a socially just society. This was supposed to be achieved by eliminating capitalist mechanisms, thought to be responsible for producing social inequalities and replacing them with a new, socially just system of pro duction and distribution of societal resources. The key aspects of the socialist system were the common own ership of the means of production and the administrative allocation of resources governed by the state and the ruling communist party (consequently, some authors prefer to use the term 'state socialism' or 'communism'). Socialist society was considered as a temporary phase in the transition to a classless communist society where everyone would contribute with her/his abilities and receive according to her/his needs.

The common ownership of almost all means of pro duction was achieved through nationalization. National ization brought enterprises from private ownership to the hands of the state, and thus the exploitation of workers by capitalists was removed. To alleviate inherited injustices in housing, dwellings were redistributed from bourgeois to working class families. However, the system highly favored communist party members and officials. Socialist ownership (state and cooperative) received ex plicit preference over private ownership. Furthermore, a distinction was made between personal ownership (given by the amount appropriate for the use by household) and private ownership (property in excess of what would be considered as appropriate to possess at household level thus providing means for extra income). For instance, while single family houses were considered personal ownership, apartment buildings with dwellings for lease demonstrated private ownership, as they presented a capitalist mean that provided rent to the owner.

The distribution of revenue among employees, com panies, and governments was regulated by the system of central planning. Central planning integrated political control and economic coordination of the state and the communist party. The main objective was to achieve economic growth that would guarantee an increasing quality of life and socially equitable distribution of scarce resources. Priority was given to production and strategic (mostly heavy) industries in certain regions. A significant portion of resources were allocated to the military complex.

On the individual and family level, the allocation of resources was governed by needs declared by the state. The administrative allocation set standards and regulated price levels. Support was provided for basic needs, not luxury goods. For instance, there was a normatively given relation between family size and the size of dwelling corresponding to it. An excess space was charged with higher rent. Public housing was heavily subsidized and real property transactions were governed by the state. Land rent in terms of the mechanism of spatial allocation of functions did not exist. The spatial allocation of in vestments was determined by a complex hierarchically organized structure of decision making within the system of central planning. Priorities in the redistribution of scarce resources yielded new economic, social, and territorial disparities. The administrative allocation resulted in new inequalities specific to socialism.

Socialism created specific features in cities. Since many legacies of previous capitalist and feudal systems could not be easily removed or eliminated, the socialist cities incorporated a significant share of pre socialist areas and the inherited built environment importantly determined the development of land use patterns under socialism. The new construction concentrated (beside infill developments) right at the city edge in high density residential areas of multistory apartment buildings (see **Figure 1**), thus maintaining a compact urban form that contrasted with the sprawling North American and spreading West European cities. Although capitalist urban landscapes were affected by similar modernist planning ideas, the scale of housing estates constructed for several thousands of inhabitants, and the con sequently high proportion of multiple family housing noticeably distinguished the socialist city from the cap italist city. The inner city decay that could be similarly observed on both sides of Iron Curtain was caused by the disinvestment of the state as the main landlord, whose prime concern was to build new socialist areas. Renewal was applied in some places beginning in the 1970s, leading to brutal reformulations of pre socialist inner city landscapes. Reconstruction and regeneration of old structures were used in an experimental form only during the 1980s.

The lower share of services and higher share of industry in terms of both employment and land use stem from the nature of the socialist political economy including its strategy to keep manual workers as the revolutionary force in major cities. Thus, massive industrial areas spread from old inner city locations to

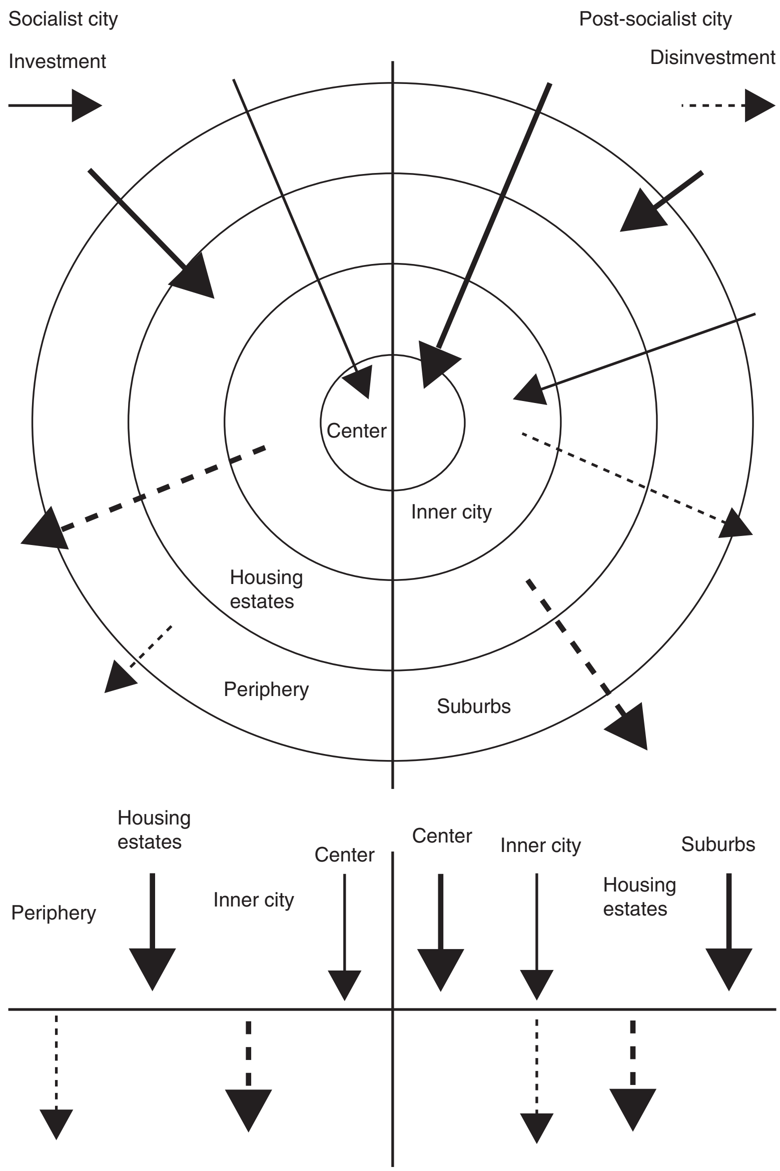

Figure 1 The pattern of investment and disinvestment in socialist and post-socialist city.

newly established outer city zones. Services were distributed in a hierarchical structure of city, sector, and residential district centers, in order to guarantee economies of scale as well as even accessibility of population from residential areas to individual types of services. The commuting between places of residence, work, services, and leisure activities was assured by a dense network of public transportation. Central areas retained a significantly high share of residential function. The political role of city centers was accentuated through the location of government and communist party buildings there. Centers were prime locations of culture with the presence of numerous cultural amenities, such as museums or theaters. Although central areas concentrated on retail and other consumer services, they did not resemble the picture of capitalist urban downtowns dominated by commerce and corresponding business driven architecture.

In socialist cities, population in urban space was primarily differentiated by demographic characteristics. This was caused by spatially concentrated housing construction, housing allocation to young generations, and a low level of migration. Consequently, particular generations of housing estates corresponded with certain age cohorts and their positions in family life cycle. There was a lower level of residential socioeconomic differentiation compared to the capitalist cities. However, residential socioeconomic inequalities still existed. This was not only because of the uneven nature of the residential environment and housing stock but also due to the

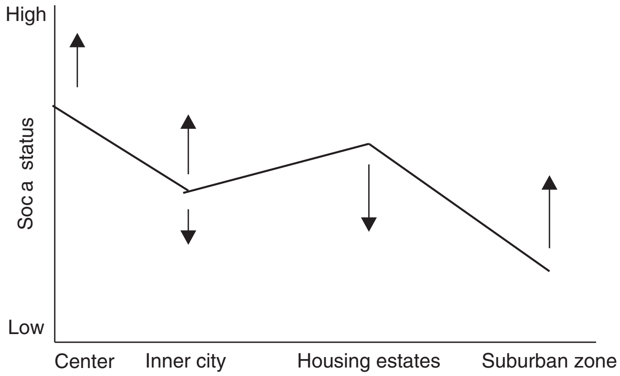

Figure 2 Social status of population in socialist city and its changes during post-socialist transformations.

mechanisms of housing allocation under socialism. Certain groups of people received better housing based on higher societal or communist party merits. This resulted in a pattern of large areas characterized by a social mix of households, less segregated and more evenly distributed lower social status population and a concentration of elite in exclusive neighborhoods. On the macro level, social status declined with the growing distance from city center, with sharper decline in some inner city sectors and stronger social status of some well located new housing estates (see **Figure 2**). On the micro level, a mosaic of various socially mixed residential areas reflected the combined effects of past and present as well as the less socially deterministic nature of the administrative allocation of housing. In the cities within the Soviet Union, the sociospatial pattern was even further complicated by ethnic differences caused by migration movements between Soviet Republics within the whole country territory.

As socialism emerged as a fundamental alternative to capitalism, the socialist city was an alternative to the capitalist city. This was not only for the normative goals based on socialist principles but also due to the outcomes of the really existing socialism. Similar to capitalist cities, the socialist city was shaped by cultural modernity, industrialization, technological progress, and advancing divisions of labor. However, the particular character of the socialist, political, and economic organization steered these general processes of societal development through its backwaters. Within the common background of a modern city of the twentieth century that was shaped by industrialization and physical planning, socialist cities represented the major alternative to capitalist urbanization.

Societal Transformations and Their Impact on Urban Change

Post socialist transition does not proceed in vacuum as the replacement of one reality with another. It is a highly complex process that reformulates existing and creates new structures. Transition consists of multiple transformation processes that are related in certain time sequence, address universal or more specific areas, and include both the political application of normative concepts as well as spontaneously proceeding societal changes within (re)established market environment.

In the first years of the transition, systemic changes in political and economic systems played the leading role. Their realization was a necessary precondition for spontaneous transformations to take place. The systemic government controlled reforms aimed to establish a capitalist system based on a pluralist democracy and a market economy, and to integrate countries into international political and economic systems. In most post socialist countries, the basic reforms of the political system were achieved in the first months and years after the collapse of communism. They included democratic elections at all government levels and the decentralization of power to local governments. Then the new democratic governments focused on transformations of economic system, the main pillars of which included the privatization of state assets, the liberalization of prices, and free foreign trade relations. The main outcomes of these systemic government controlled transformations were new societal rules based on democratic policy making and market economy principles, a vast number of new private actors and owners, and an openness of local economic systems to international economic forces.

The market principles of resource allocation and the growing exposure to an international economy formed conditions for the development of spontaneous societal transformations of economic, social, and cultural environments. Urban change has been especially influenced by internationalization and globalization, economic restructuring in terms of de industrialization and the growth of producer services, increasing social differentiation, new modes of postmodern culture, and neoliberal political practices.

International companies and foreign direct invest ments started to play the decisive role in reshaping local economies and economic geographies and in determining the position of post socialist countries, regions, and cities in a global economic order. In major cities, inter nationalization was strongly present in producer services that formed the soft infrastructure for capitalist system expansion to new markets and later in consumer services and industrial production. Foreign managers and high salary employees of foreign companies formed a specific segment of the demand on the high income housing market. In some countries, immigrants formed the bottom tier of the socioeconomic hierarchy. Foreign investment in real estate (office, retail, industrial, as well as residential) began to massively reshape urban land scapes and skylines. A highly uneven balance between Western and domestic actors represented by capital and access to Western resources was further exacerbated with policies and measures favoring capital strong for eign enterprises.

Economic restructuring involves de industrialization and tertiarization. Economic decline that was caused by de industrialization and the development of consumer services, particularly retail which offers low paid jobs affected most cities. The growth in advanced producer services provided fortunes only to major command and control centers, usually capital cities. This left the remaining urban places to struggle for foreign direct investments in manufacturing that would bring reindus trialization and jobs for local population. Local economic systems were integrated into the global economy in a highly uneven manner with capital cities ranking well among major global centers, successful regions providing cheaper and well skilled labor to routine production within the global assembly line, and per ipheries waiting with the mixture of hope and depression. De industrialization left behind extensive brownfields posing both potential for redevelopment as well as threat of further spreading decay. Producer services concen trated in expanding city cores of major centers and contributed to their rapid commercialization. Many of these services later decentralized to out of center locations and business parks at the city outskirts. Retail and tourist facilities in attractive places brought new bright consumption landscapes to city centers followed by later massive expansion of malls and big boxes in outer city areas.

Economic change mediated through labor markets was expressed in growing wage and income disparities. Various national forms of social security systems miti gated some of the social hardships of economic restructuring. However, they could not change the gen erally accepted move toward higher income differen tiation that contributed to social polarization among households and the formation of new territorial disparities in well being. The income inequality was reflected in the reemergence of pre socialist patterns of residential differentiation, the establishment of new enclaves of affluent population, as well as segregated districts of socially excluded. Social and physical upgrading has developed via incumbent upgrading and gentrification of existing neighborhoods and construction of new districts of condominiums for sale in inner city areas, and growth of new suburban communities of family housing for new wealthy population. In some post socialist cities, the escape to rural suburban areas presented the survival strategy of low income population, which otherwise remains trapped in some decaying inner city neighbor hoods and less desirable postwar housing estates.

New conditions allowed for the development of a greater plurality of values, a tendency toward indi vidualism and promotion of self interests. Values and preferences of a younger generation shaped by new op portunities as well as some social hardships brought about by economic reforms have resulted in a rapid decline of family formation. An exaggerated second demographic transition has impacted residential choices with consequences for urban housing. More importantly, consumption has become the key life target for popu lation masses. Globalization has brought a homogeniza tion of consumption via the same goods offered by transnational corporations. But not everyone enjoys these opportunities. Consumption by most people is limited by their income level and preferences shaped by advertise ments tailored to the standard mass consumer behavior. Shopping centers use visual features and various forms of entertainment to attract consumers, and transnational corporations highlight their presence in the urban structure. Private firms use cultural strategies to sell themselves, strengthen their influence and competitive ness, and demonstrate their pride and power. The post modern esthetic in post socialist city utilized and intensified the collage of the past and present, local and global, standard and specific, and real and virtual. The socialist propaganda of collective will was replaced by individualistic choice in consumption oriented cap italist city.

National as well as urban governments often pursued political strategies influenced by neoliberal political discourse. However, the real pragmatic politics mixed neoliberal ideology with attempts to keep social peace through retaining some socialist regulations in operation. The decisions of both the central government as well as local politicians were grounded in a neoliberal ideology, which saw free, unregulated market as the mechanism of allocation of resources that would generate a wealthy, economically efficient, and socially just society. Poli ticians perceived the state as the root of principal harms to society and the economy in particular. Urban planning was perceived as contradictory to the market.

Clientelism prevailed and *ad hoc* decisions were preferred to the creation of basic rules of the game embedded in a long term plan, strategy, or vision of urban development. The internal urban transformations have often been left to the operation of free market still partially bound within the framework of traditional, rigid, physical planning instruments. However, after the first decade of transition, many urban governments learned new techniques of urban management and governance. They started to apply more sophisticated tools, such as stra tegic planning. The application of European Union programming documents in accession and new member states of European Union further helped to consolidate urban government measures toward more coordinated and complex policies to tackle urban problems. The power of the planner who governed allocation of in vestments in the socialist city was replaced by investors who steer politicians and planners to facilitate urban development in a manner favorable to capital.

The new mechanisms that shape urban spatial or ganization of post socialist city developed through a double transition that includes the local post socialist transition to market economy and the global transitions conditioned by economic globalization and its influence on local political, economic, social, and cultural re structuring. These two transitions are mutually inter related. The local transition was necessary to link local societies to the capitalist world. 'Western winds of change' could then start to play their decisive role in local market led transformations. Despite the strong role of local socialist legacies, power to shape future lies in the hands of actors operating in the global market place. Precisely, this double transition is the key feature of post socialist change, which distinguishes it from other types of transitions.

Restructuring Urban Spatial Organization

The post socialist transition is shaping former socialist societies in similar directions. However, the starting conditions varied, transformation paths and the speed of change differ, and the outcomes exhibit both similarities and differences. Despite similar destinations, market reforms were not uniform. For instance, the use or refusal of restitution (property return to previous owners) cre ated strikingly different conditions within urban property markets. In Prague and Tallinn, restitution provided a significant supply of real estate in inner city property markets leading to turbulent land use changes in attractive locations; this was in contrast with Budapest, where restitution was not applied. The character of the built environment has shaped possibilities for current investment. For example, the conservation of historic city core in Prague entailed strikingly different conditions compared to central Warsaw that has been remodeled with the dramatic construction of skyscrapers. The memory of the pre socialist era and the burden of the socialist legacy, the progress of country reforms and their institutional affiliation with West, the size of market, and the geographic and cultural proximity – all these have affected the inflow of foreign direct investments that play such a key role in local post socialist economic development. While the global economic system em braces the whole planet, individual countries, regions, cities, and even neighborhoods are integrated into it in a highly uneven manner. The changing geography of post socialist city reflects place specific local features (of urban environment, institutional regulations, and cultural values) as well as generally similar, but still in ternally differentiating impacts of the transition to global capitalism.

Most urban growth in post socialist cities is concen trated in the attractive localities of the city center, some adjacent nodes and zones in inner city, and in numerous suburban locations (see **Figure 1**). The main transfor mations in the spatial pattern of former socialist cities and their metropolitan areas included the commercialization and expansion of city centers, the dynamic revitali zation of some areas within the overall stagnation in inner cities, and the radical transformation of outer cities and urban hinterland through commercial and residential suburbanization (see **Figure 3**). City centers and suburban areas are the territories with the most radical urban change. First decade of transition was characterized by investment inflow to city centers causing their com mercialization and decline in residential function, albeit substantial physical upgrading. Decentralization occurred later with investments flowing to both out of center and suburban locations. Central and inner city urban re structuring involved the replacement of existing activities with new and economically more efficient uses, taking the form of commercialization, gentrification, construction of new condominiums, brownfield regeneration, and the establishment of new secondary commercial centers and out of center office clusters. In the course of time, sub urbanization has become the most dynamic process changing the landscapes of metropolitan regions of post socialist cities. It is bringing a complete reformulation of metropolitan morphology, land use patterns, and socio spatial structures.

Besides its positive economic and physical effects, city center commercialization also brings decline in residential function, conflicts with historical heritage, and overutilization accompanied with congestion. The value of city center scenery has been appropriated by busi nesses, showing the uneven juxtaposition of international private capital over the local public. The socially and spatially selective gentrification that is strongly associated with foreign business elite proceeds unchallenged since it

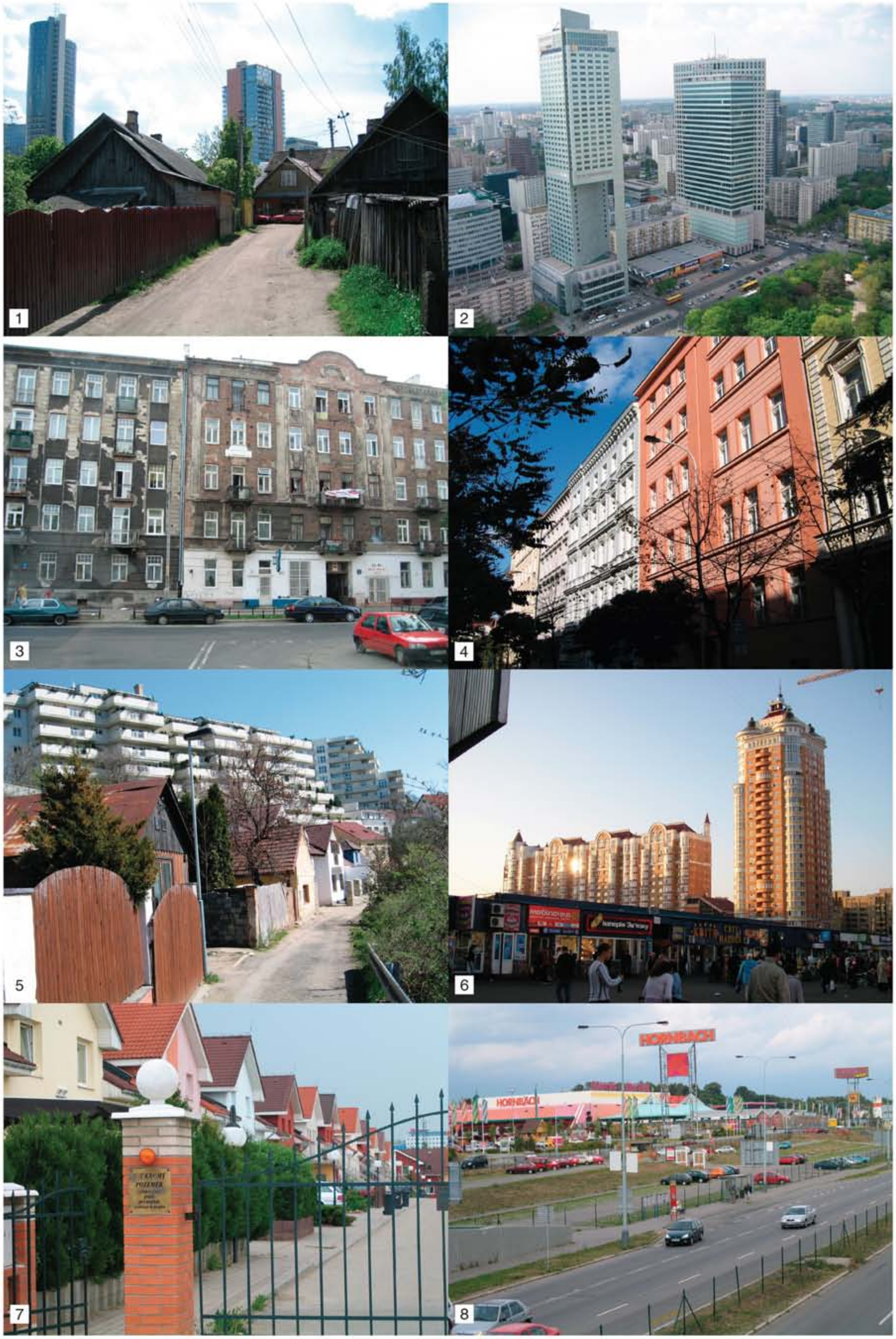

Figure 3 Post-socialist landscapes: (1) contrast of new international and old wooden vernacular housing (Vilnius), (2) manhattanization of central Warsaw, (3) inner-city stagnation in Praga (Warsaw) contrasting with (4) gentrification at Vinohrady (Prague), (5) new condos overlooking urban village (Prague) and (6) housing estate (Kiev), (7) gated communities, and (8) ribbon retail development in suburbs (Prague) (photo: Michael Gentile (6), Luděk Sýkora (1–5, 7–8)).

is thought to be a natural process that brings bourgeois neighborhoods back to their former glory. Revitalization, whether based on residential or nonresidential functions, is now directed by capital, not governments. It has brought investment to some decaying areas, leaving many other places unattended in the sea of stagnation. The formation of brownfields and the physical and social decline of some housing estates constructed during socialism present major problems in the post socialist city. Most controversy is linked to suburbanization. It is praised for the freedom of housing choice under free market conditions, and blamed for social, economic, and environmental insustainability.

The compact physical morphology of former socialist city is changing under the impact of activities sprawling into the hinterland. Property development brings densi fication in central city landscapes. Brownfields formed by de industrialization present new inner spaces of opportunities whose former structure can be completely remodeled provided there is the interest of investors. City centers are gaining the spirit of business and in some instances tourist districts, keeping the presence of government and some cultural functions. Tertiarization also brought fortunes of growth to outer areas, forming secondary nodes of employment in advanced services concentrated in office districts and new landscapes of consumption in shopping and entertainment complexes offering services to mass population as well as low paid jobs to less skilled, mostly female, workforce. High concentrations of retail facilities are being strategically located in new nodes between existing residential areas with their original small neighborhood centers, and between the compact city of mass consumption power and new booming wealthy suburbs. The original hier archically organized system of services with the domin ance of the central city is transformed with layers of new centers to a more polycentric structure. The new func tional–spatial division of roles between new nodes and specialized areas affects the life and death of smaller centers that with their limited scale and scope of services on offer cannot compete with retail giants. Sociospatial patterns are also changing (see **Figure 2**). The strengthening of the existing disparities between areas within cities is counterbalanced with a social recompo sition of gentrified districts, declining housing estates, and suburbanization of the wealthy. Surprisingly, the level of sociospatial disparities has not increased during the first decades of capitalism. Segregation processes operate in their initial phases against the socialist heritage, bringing some higher social status neighborhoods (housing estates) as well as lower social status zones (suburbanization of rural hinterland) closer to the city average. However, the new capitalist principles set the city in motion toward new outcomes that will in the course of time illuminate themselves more boldly on the watermark image of socialist city. This is in the meantime signaled by con trasting pictures of ghettoizing areas and gated com munities as examples of extreme developments in segregation patterns.

Post-Socialist City: Socialist Legacy, Drive to Capitalism, and (In)definite Future

Despite approaching two decades since the political turnover and the beginning of political and economic transformations, cities in former socialist countries can not be described by a set of key stable attributes. Their key definitional characteristics are transformations – processes of change. The principles and mechanisms that shape urban spatial organization have already been largely changed. Urban growth, decline, and restructuring are conditioned by market principles and regulations based on democratic policymaking. The spatial structure of post socialist cities contains new urban landscapes, which emerged under capitalist con ditions. However, large urban sections still resemble socialist cities. There have been important urban changes; nevertheless, they have not completely altered the former spatial patterns yet. While the core transfor mations of political and economic systems were accomplished within few years, changes in settlement structures will take several decades.

Post socialist cities are cities in transition. They are characterized by the dynamic processes of change rather than static patterns. The urban environment formed under previous system is being adapted and remodeled to new conditions given by the political, economic, and cultural transition to capitalist society. Many features of socialist city suddenly appeared in contradiction with the principles that govern capitalism. These contradictions between capitalist rules of the game and socialist urban environment led to the restructuring of existing urban areas. In the course of time, new capitalist urban devel opments are gaining more important impacts on the overall urban organization. The post socialist develop ments bring the reemergence of some pre socialist pat terns, transformations in some areas from socialist times, and creation of new post socialist urban landscapes.

Post socialist city is a temporarily existing phenom enon. Its transition from socialist to another type of city is definite in time. Post socialism is the time period between the change in the rules of the game and the completion of corresponding transformations in built and social environments. When will such transformations be accomplished and cities will become one of the many variants of urban places under capitalism? This question cannot be answered yet. Post socialist cities are still on the road of their transformations.

See also: Capitalism; Communist and Post-Communist Geographies; De-Industrialization; Gentrification; Globalization and Transnational Corporations; Segregation; Socialism; Transitional Economies; World/Global Cities.

Further Reading

Altrock, U., Gunter, S., Huning, S. and Peters, D. (eds.) (2006). *Spatial Planning and Urban Development in the New EU Member States: from Adjustment to Reinvention*. Aldershot: Ashgate.

Andrusz, G., Harloe, M. and Szelenyi, I. (eds.) (1996). *Cities after Socialism: Urban and Regional Change and Conflict in Post Socialist Societies*. Oxford: Blackwell.

Borén, T. and Gentile, M. (2007). Metropolitan processes in post communist states: An introduction. *Geografiska Annaler B: Human Geography* 89, 95 110.

Enyedi, G. (ed.) (1998). *Social Change and Urban Restructuring in Central Europe*. Budapest: Akadémiai Kiadó.

French, R. A. and Hamilton, F. E. I. (eds.) (1979). *The Socialist City: Spatial Structure and Urban Policy*. Chichester: John Willey & Sons.

Geografiska Annaler B: Human Geography (2007), *Special Issue* 89(2).

GeoJournal (1999), *Special Issue* 49(1).

Hamilton, F. E. I., Dimitrowska, A. K. and Pichler Milanović, N. (eds.) (2005). *Transformation of Cities in Central and Eastern Europe: Towards Globalization*. Tokyo: United Nations University Press.

International Review of Sociology (2006), *Special Issue* 16(2).

Kovács, Z. and Wiessner, R. (eds.) (1998). *Procese und Perspektiven der Stadtentvicklung in Ostmitteleuropa, Münchener Geographische Hefte 76*. Passau: L. I. S. Verlag.

Stanilov, K. (ed.) (2007). The Post Socialist City: *Urban Form and Space Transformations in Central and Eastern Europe after Socialism. Dordrecht:* Springer.

Sýkora, L. (2000). Post communist city. In Jazdzewska, I. (ed.) *XII Konwersatorium Wiedzy o Mieście. Miasto Postsocjalistyczne Organizacja Przestrzeni Miejskiej i jej Przemiany*, pp 41 45. Łódź: Universytet Łódzky, Komisja Geografii Osadnictwa i Ludności PTG, Łódzkie Towarzystwo Naukowe.

Szelenyi, I. (1983). *Urban Inequalities under State Socialism*. Oxford: Oxford University Press.

Tsenkova, S. and Nedović Budić, Z. (eds.) (2006). *The Urban Mosaic of Post Socialist Europe: Space, Institutions and Policy*. Heidelberg: Physica Verlag.

Poststructuralism/Poststructuralist Geographies

K. Woodward, University of Exeter, Exeter, UK
D. P. Dixon, The University of Wales, Aberystwyth, UK
J. P. Jones, III, University of Arizona, Tucson, AZ, USA

Glossary

Deconstruction Post-structuralist method that renders contextual or uncertain a conceptual system by illustrating that system's codependence on other, presumably independent concepts or systems.
Epistemology Study of how we know and understand the world; typically described by key binary forms such as objectivity/subjectivity, interpretation/explanation, discrete/relational, etc.
Genealogy Post-structuralist approach to both historical analysis and the historian that substitutes origins, linearity, and truth, for multiplicity, dispersion, and power/knowledge.
Ontology Study of what the world is like, or must be like in order for us to know it; typically described by key binary forms such as individual/society, nature/culture, space/time, etc.

Introduction

Over the course of the twentieth century geography underwent a series of conceptual revolutions, inspired as much by tumultuous on the ground events, as well as the crossover of ideas and concepts developed within other disciplines. Though the arrival of post structuralism can be considered simply another 'turn' within the discipline, what renders this approach distinctive is its rigorous interrogation of those core concepts – such as objectivity and subjectivity, center and margin, materialism and idealism, truth and fiction – that underpin much of modern day academia, including the majority of geo graphic thought and practice. To be specific, post structuralism brought to the field of geography in the late 1980s and 1990s a critique that unsettled both the epistemological (i.e., theories on how we know the world) and ontological (theories on what that world consists of and how it works) moorings of the then dominant theoretical frameworks: spatial science, critical realism and Marxism, and humanism.

In these early years post structuralism did not offer a clear counter ontology to these frameworks. Rather, by claiming that any ontology is always already an outcome of epistemology, of our socially constructed ways of knowing, post structuralists asked that we reflect not only on how we know, but also on how elements of ontology – such as space, place, nature, culture, individual, and society – become framed in thought in the first instance. In posing such questions, post structuralism found, and continues to find, a productive moment in examining how social relations of power fix the meaning and significance of social practices, objects, and events, determining some to be self evident, given, natural, and enduring. In regard to geography, this requires an analysis of why some objects – landscapes, regions, space, place, etc. – rather than others are taken to be central to geographic inquiry, as well as an analysis of how those objects are understood to exist and relate to one another.

This sort of post structuralism, then, is epistemo logically centered. Certainly, it came to prominence under the banner of a more widespread 'linguistic turn' in the humanities and social sciences, which emphasized the production of meaning and the 'social construction' of reality, and made much of the emphasis placed upon the construction of meanings in the work of Derrida and Foucault. For this reason, critics of post structuralism offered a rejoinder that touted the sometimes brute for ces of materiality: of class exploitation and the uneven distribution of wealth and poverty, of the forces of nature and the concrete effects of environmental conditions, and of gender relations and the facts of biological repro duction, to name a few.

Criticisms that post structuralists have been con cerned only with discourse and representation, as opposed to the 'real' material conditions within which these meanings were considered to be embedded, had a profound impact on geographic debate during the 1990s. Marxists and some feminists were often found accusing post structuralists of a simplistic idealism and a relativist, even nihilist, politics. Post structuralists responded by claiming that no objects were outside the systems of representation, and that any claim to know them in an unmediated way was no more than an exercise of power whereby one theoretical stance was privileged above all others as both accurate and truthful.

As these debates waged they also waned to the point of tiresomeness. It has been in the latter half of the move ment that post structuralist geographers have come to reassert their claims over ontology, largely by rethinking difference and representation in more explicitly materi alist terms. In doing so, these geographers have interro gated more closely the ontological ramifications of the work of Derrida and Foucault, but have also explored the work of Deleuze and Latour, whose contributions are discussed here. At the same time, what was a somewhat knee jerk critique of other epistemological stances, for example, 'Marxism', as a 'meta narrative' that promised to singularly explain real world conditions, has been tempered by a more sensitive appraisal of the diversity of ideas beneath such a label. Indeed, it is this more nuanced approach to academic discourse that resonates with a post structural emphasis on difference understood not only as a rejection of sameness and the status quo, but also as a receptivity toward the experimental and the new.

Taken together, both dimensions of post structuralist thought – the epistemological and the ontological – pivot around a set of fundamental questions. These include: If meaning and representation are indeterminate and contextual, and if, as a consequence, the 'real' world is 'constructed' as an ontological fact, then how does power work to produce its truths? And, if difference in the world is not a residual from or a bad copy of a singular Identity, but is rather the immanent force characteristic of all materialities, including imaginings, emotions, words, and meanings (as well as those elements more usually thought of as material, such as organisms and the landscapes they inhabit), then how, in the shift in thought that moves from 'being' to 'becoming', do we go forth in the world to think and speak in terms of things and their qualities?

Before we begin to address these questions, it seems worthwhile to emphasize that though its intellectual roots are in Continental philosophy and literary theory, post structuralism knows no boundaries when it comes to objects of analysis. So, though its impact has been most strongly felt in cultural geography, where it has not only invigorated research questions but has also led to the identification of new objects of analysis (e.g., films and other texts), its critical stance toward simplistic forms of truth, representation, materiality, and politics have become points of engagement between it and other geographic subfields, including economic geography; geopolitics and the state; urban and rural geography; cartography and geographic information system (GIS); social geographies of gender, 'race', and the body; postcolonial geographies; and nature–society relations. And, it is through this process of de stabilization that post structuralism has wide ranging implications for geography at large.

Poststructuralism

Structuralism

Part of understanding the relationship between poststructuralism and structuralism is highlighted in the prefix 'post', as opposed to 'anti'. The former negates its suffix, but it does so relationally and in ways that carry structuralism with it – as in beyond rather than against. For this reason, it is impossible to write about poststructuralism without first coming to some understanding of structuralism.

The literature associated with structuralism is complex and wide ranging, but in all its forms it holds that all manner of processes, objects, events, and meanings (let us call these POEMs for short), are taken to exist not as discrete entities, but as parts relationally embedded within, and constituted by, underlying wholes, or structures. It is not unusual to see structuralism rendered as an inflexible and static framework, but that would be a misunderstanding. A structure is not an external architecture upon which POEMs are hung, for such a view implies that a structure exists independently of the parts it embeds; instead, structures are constituted solely from the relations among their constitutive elements, or parts. And, since they do not exist independently of POEMs, structures are dynamic and spatially differentiated fields of relations. Finally, structures do not have material form nor do they have the ability to act; they are not visible in the empirical realm but, inasmuch as they systematize the relations, and therefore the causal efficacy, of POEMs, they are presumed to operate.

The most important structuralist thinker for the development of post structuralism was the Swiss linguist, Ferdinand de Saussure (1857–1913). His goal was to understand the abstract structures behind all forms of social communication, from painting and religious rituals to chess games and the rules of courtship. As a linguist, Saussure applied his theory of semiotics – that is, the science of signs – to the study of language. In doing so, he rejected the traditional, historical approach to the study of language, a philological endeavor focusing on detailed descriptions of the historical evolution of particular languages and language families. He also rejected the positivist line of research dominant in his day, which sought to understand language through the analysis of sounds and their impact on the nervous system. For Saussure, elements of language gain their currency according to the structure that they create and within which they are embedded. A particular language, therefore, must be studied as a systematized collection of sounds and inscriptions, each of which, as in structuralism more generally, can only be assigned a value or meaning when thought of in relation to the remainder of the elements.

But how does language work to transmit meanings from one person to another? In analyzing the relations among these elements, Saussure struck an analytic distinction between the 'signified', which is the mental construct, or idea, of a particular phenomenon, and the 'signifier', which takes the form of a distinguishable 'mark', such as a sound, inscription, or special body movement. Within a language, signifieds are associated with particular signifiers to form a 'sign'; in consequence, when people communicate they use particular signs to convey and understand meaning. In addition, because signifiers are considered to exist within the realm of the symbolic, that is, as abstract representations that refer to real world phenomena, the systems of communication within which they are embedded can be thought of as relatively autonomous from any real world referent. Given that there is no necessary relationship between the signifier and the signified, the actual choice of signifier is arbitrary. This is why various languages can have different words (signifiers) for the same object (the signified). Indeed, the signifier only has value when it can be differentiated from other signifiers and used to convey a particular signified again and again. All languages, then, depend upon the fact that we learn to recognize this difference between signifiers.

Now the very fact that communication can occur through signifiers that are fundamentally arbitrary implied for Saussure that a system, or set of rules, must exist by which people can indeed be taught to differentiate between signifiers, and to which all must subscribe if communication is to proceed unhindered. Just as chess and courtship (both systems of signs) are built around certain rules of the game (the moves of the knight, the lingering glance), so all languages are founded upon abstract regulations that shape the ways in which they are played, or manifested in practice. Within this conception, the underlying structure that allows communication to take place is called langue, while the actual practices by which communication takes place Saussure called parole. To sum up, for Saussure the elements of language constitute interrelated signs, whose mark or signifier is embedded in a structure of langue, which itself may be transformed through the practice of parole.

That Saussure's model could be applied to any number of sign systems in any language and across myriad communication systems accounts in part for its popularity well into the 1960s in a variety of disciplines, including literary theory and philosophy. Freudian psychoanalysis, in particular the analyses of dreams, was rooted in structuralism. So too was anthropologist, Lévi Strauss's search for the organizing principles of culture. And, in some versions of Marxism, structuralism underwrote attempts to explain many aspects of social life as determined by the underlying mode of production. It was with these and other forms of structuralism that post structuralism took issue.

Language and Discourse

Though elements of post structuralism can be found in the work of philosophers such as Friedrich Nietzsche and Martin Heidegger, its formal recognition as a body of theory can be traced to a host of more contemporary social, cultural, and literary theorists. Here, under the heading of language and discourse, both of which speak to underlying issues concerning epistemology, we discuss the work of two of the most important theorists, Jacques Derrida (1930–2004) and Michel Foucault (1926–84). Later, in the section titled 'Materiality and difference', which addresses more avowedly ontological concerns, we take the opportunity to discuss the work of two other prominent post structuralists, Gilles Deleuze (1925–95) (and colleague Félix Guattari, 1930–92) and Bruno Latour (1947–).

It was Derrida who, at a 1966 conference on structuralism in the city of Baltimore, introduced post structuralist thought to an international audience through the presentation of a paper titled, 'Structure, sign, and play in the discourse of the human sciences'. The major goal of the conference was to stimulate innovation in structuralist thought across a wide variety of disciplines. Yet, Derrida's paper (published in 1970 and reissued in 1972) critiqued the very notion of structure by analyzing the process of 'centering' upon which diverse forms of structuralist thought were constructed. Tracing back particularly significant manifestations of this centering process in Western thought, Derrida suggested that what seemed to be ontological securities, such as presence, essence, existence, cause, origin, substance, subject, truth, God, and 'man', were merely epistemological constructs handed down through generations of philosophers and scientists.

Specifically, Derrida noted that in the process of producing a structure's parts, all those elements that seemingly do not have some form of relation with its center must be excluded. Thus while the center is related to all of those elements within the structure, it is also held to be beyond the excluded elements, and therefore fixed and inviolable, at least with respect to those 'other' elements. However, Derrida argued that the structure could not exist without the accompanying exclusion, and this meant that the center was both within (i.e., a presence) and outside (i.e., an absent presence) at the same time, implying that a center is not really a center after all, but a contradiction, a force of desire or power rather than an ontological foundation. Another way of viewing his contribution is from the perspective of within: if centers rely on the exclusion of outside elements to produce structures, then they and their associated structures are dependent upon the 'outside other', or 'constitutive outside'. Derrida would go on to show how we can uncover the productive 'trace' of this 'other' within centers and structures so as to 'deconstruct' the effrontery of their claims to independence. In addition, Derrida explained how a structure paradoxically provides both the grounds and constraints for what he called the freeplay of the structure. This freeplay – that is, the movement and change that determine how a structure unfolds – is both made possible and limited by the 'structurality' or centering of a structure, which contra dictorily hinges on its fundamental immobility and reassuring certitude. As Derrida went on to note, such centering is the product of a binary – an either/or – epistemology. Such a way of thinking about the world stabilizes not only the meaning of one term, such as truth (a center relying on such 'parts' such as objectivity, facts shorn of ideology, etc.), but also, through the assignation of a periphery, defines an 'other' that falls outside of its purview, fiction. In a similar vein, binary thought pro duces sharp contrasts between essence and ephemerality, cause/effect and contingence/serendipity, substance and chimera, God and idol, and 'man' and nature.

Post structuralists drew three important conclusions from this mode of thought. First, they maintain that in all of these binary systems, what appears to be the foun dation for a system of thought is but a hypothetical construct, one that reveals more about the society that produced it than the supposed character of the real world. In this case, post structuralists turn their attention to the production of centers, margins, and the boundaries that demarcate them. In recognizing centers and margins as products of an either/or mode of thinking rather than the natural state of affairs, post structuralists are drawn to several key questions. On the one hand, they ask who has the social power to draw the boundary between a center and margin; on the other hand, they question to what end is any such system of differentiation directed. In recognizing categorizations as the product of social relations of power, attention turns to which social groups have the discursive resources to construct categories; that is, who has the ability to name the world? Thus, a major component of post structuralist research involves in quiries into the categories that frame reality according to either/or binaries.

Second, binaries presume a totalizing epistemology, so termed because either/or thought can only posit a world in which everything either 'is' or 'is not'. Epistemo logically speaking, the effect of binary thought is to constrain what can be conceived about the world. Now, in some instances binary thought can be productive, as in, for example, the 'freeplay' set loose in the formulation of computer languages that operate on an underlying sys tem of ones and zeroes. However, in other instances binaries so stricture what knowledge is possible that they unduly limit what can be conceived in the world. In this way, the binary epistemology ultimately infuses ontological concepts (or what we presume the world to consist of, e.g., the individual vs. society, local vs. global, conscious vs. insensate, subject vs. object, chaos vs. order). Consider, for example, our understanding of phenomena as either natural or cultural. Such binary thinking will ultimately organize virtually all questions researchers might want to ask about social or physical systems. These questions, however, can ultimately be exposed as circular in character, for, though researchers may think they are posing questions about 'real' categories, they are by default investigating the products of their own binary epistemology.

A third complaint about structuralist thinking is that it is not, in fact, as fully relational as structuralism claims it to be. For example, while Saussure's model assumes that language comprises an arbitrary system of signifiers whose elements become meaningful through their relation to each other (the word 'cat' does not sound like 'dog' and thus permits us to understand the differ ence), for him the concept of a feline, four legged mammal (the signified) becomes the agreed upon, or correct, re presentation (see below) of the real world animal, or referent, independent of the existence of its canine variant. Using Derrida's critique of Saussure, however, we could argue that the mental construct of a feline is not grounded in the one to one relationship between it and the referent, but is definable only in relation to all other concepts that give feline its dis tinction by referring to what feline is not. Thus, feline is negatively defined in relation to a host of other concepts such as canine, leonine, equine, lupine, and bovine. Moreover, and this is the important point, all of these concepts, from feline to bovine, are themselves produced within a myriad of other relational fields of meaning, popular, as well as scientific. Rather than to assume a uniformity of meaning in the face of such complexity, post structuralists point instead to contradiction, juxta position, bricolage, and imbrication. In this manner, post structuralism throws doubt onto all certainties regarding researchers' ability to accurately represent reality, for our concepts do not simply re present that reality, in the sense of mirroring their referent, but represent reality within a fully relational system of understanding that does not require the referent to be cognized in the same manner by all.

A second influential theorist of post structuralism was the French philosopher historian, Michel Foucault. Whereas Derrida focused on the dualistic presumptions of Western philosophy, Foucault undertook to problem atize the production of modernist forms of knowledge, noting how 'Western' scientific ideas and practices since the eighteenth century produced a series of non normal others, such as the insane, the abnormal, and the sexually perverted. In doing so, he drew attention to how this modernist undertaking has been underwritten by a par ticular conception of 'Man' (as Foucault invariably put it) as a unique being, capable of describing, explaining, and mastering the operation of body and mind, as well as society and nature. Hence, within these scientific analyses, the human subject is not only the object of her/his own understanding, s/he is also understood to orchestrate the social and physical realms within which s/he lives.

In placing humans within these contexts, Foucault argues, modernist forms of knowledge necessarily establish a series of insurmountable paradoxes. First, s/he appears as an object to be studied empirically alongside other objects, but is also posited as the transcendental source from which the possibility of all knowledge can flow. Second, in determining the domain of conscious thought, s/he has also framed the unconscious. And yet, in presenting her/himself as the source of intelligibility, s/he must attempt to explain this latter realm; that is, to think the unthinkable. And third, humans conceive of themselves as the product of history, and yet posit themselves as the source of that very history. For Foucault, modernist sciences of 'Man' simply cannot produce a comprehensive account of their subject/object, and so must disintegrate under the weight of their own contradictions.

Foucault's own historical analyses can be considered a commentary on this same condition. Among the most important of his concepts is that of discursive practice. Put briefly, a discursive practice is a regularity that emerges in the very act of articulation. As such, it should not be thought of as a set of meanings that are somehow imprinted onto real world phenomena. (To think this way is to rely on an idealist understanding of the mind as the source of knowledge, and presume an unwarranted distinction between mind and body, self and social.) For Foucault, by contrast, each articulation establishes the conditions of possibility for thought and action; that is, it posits what is appropriate and reasonable to be thought and practiced. As such, an articulation is more than mere communication – it is an active intervention in the social and physical realms. From this position, Foucault derived two analytic projects.

First, he noted that each articulation is produced and understood within a given context, such that it is afforded meaning. The kinds of articulations Foucault was interested in were those that had gained sufficient authority such that they were deemed to be valid even when they were taken out of context. That is, they had gained the status of Truth. Hidden in previous analyses of communication, argued Foucault, were the means by which these particular articulations gained distinction. Within a discourse, he maintained, a disciplining process takes place within and between strategies of power, which are all those techniques by which a statement is accepted as valid and appropriate, and by which that statement could not but be articulated in the way it was. In regard to social research, for example, these techniques would include empirical confirmation, dialectical argument, and phenomenological bracketing. Each of these allows for the privileging of some articulations over and above others as valid claims concerning the nature of 'reality'. For Foucault, power is considered within this context to operate through discourse, and to be complicit with the production of specific forms of knowledge that not only claim to provide insight into how the world works, but which are also deployed in the active management of that world. Key to this process is the emergence of a specialized cadre of experts, such as scientists and educators, who draw on these bodies of knowledge to further enhance their own status by ensuring the diffusion of particular ideas and concepts through society. Importantly, this legitimacy ensues not from their 'personal' character, but from the positions they hold within an institutional framework, as well as within a given set of social relations. A discourse, then, is not something that is simply produced and received by people; rather, it is tied into a discursive site, such as a school, church, office, scientific laboratory, and so on, where knowledge is actively produced and disseminated.

Second, and following Saussure's focus on semiotics as a science of signs, Foucault interpreted the term 'discourse' far beyond speech to include the inscription of social relations (and thereby the exercise of power) on and through the body itself. The complex interplay of social relations of power both enables and constrains the body in certain ways: that is, the capacity of the body to be shaped and to act. Foucault refers to the emergence of what he termed 'technologies of the self' – disciplinary actions that have become taken for granted. These range from new standards of punctuality to the self regulation of dress and hygiene. In making this argument Foucault's aim was not to reiterate the imposition of coercive power over individuals, but to show the tendency for modern day power to be depersonalized, diffused, relational, and anonymous. Power is not held by one particular group, but rather is exercised through a series of everyday activities. For some critics, this position denotes a hopeless pessimism, in that power is understood to discipline and normalize more and more dimensions of everyday life. For Foucault, however, the means of resisting relations of power lie in the disruption of this daily performance. It is at this level, the site of the body itself, that resistance takes place.

The case studies Foucault chose to pursue, consisting of penal, education, and medical systems, focus accordingly on the ways in which the self is constructed through discursive practices. The ensuing histories are also, however, illustrative of Foucault's attempt to produce a body of work that does not operate according to modernist modes of interpretation. For Foucault, there is no necessary trajectory to history, nor is there a definitive causal mechanism, such as human agency, that lies at the heart of social change. Rather, in representing history, each mode of analysis – or genealogy – must be considered as conceived and articulated around present day issues and concerns, such that succeeding analyses of the same topic must necessarily rewrite the past from the perspective of the present. Nevertheless, his own project was to bring to light how particular clusters of discourse/power worked to produce, fill, and maintain the cat egories by which we claim to know the world, such as culture, nature, history, geography, individual, society, modern, primitive, objective, subjective, and so on. All of these terms have an archaeology to them, by which Foucault meant a series of discursive formations, or epistemic spaces, within which these knowledges have been produced.

In sum, for Foucault, social researchers cannot secure, and therefore should not pursue, truth, at least when interpreted as a category with the status of a universal, timeless quality (that is, Truth with a big 'T'). This is because each society has its own regime of truth, the specifics of which are fashioned by: the types of dis courses deployed (these can be legal, moral, rational, and teleological, among others); the techniques and pro cedures used to distinguish between true and false statements; and the status of those who are charged with saying what counts as true. It is at this point that we can see an affinity between Foucault's account of discursive practices and Derrida's concern to de stabilize those centers – such as origin, subject, essence, and Truth – by which explanatory modes of analysis operate. And, as we shall note below, it is their shared concern to work against accepted ways of thinking, researching, and writing about the world that inspired an emergent cadre of post structural geographers.

Materiality and Difference

As we mentioned in our introduction, a post structuralist focus on language and discourse puts the point of emphasis on epistemology rather than ontology. How ever, it would be disingenuous, if not downright anti post structuralist, to assert a clean and neat division between the two, and indeed, as Chris Philo has intim ated in a 1989 essay, it would be incorrect to locate a philosopher like Foucault in one camp or another. In his case, Philo argues, discourse is understood to possess a distinctive ontological status. Hence, Foucault's work can be read as an analysis of how the ontological is consti tuted in part by the myriad effects produced by dis course. Nevertheless, it is safe to suggest that for many geographers inspired by their work, both Derrida and Foucault were crucial in shaping their own analyses not of the way the world is (ontology), but with the power orientations and slippages at work in how we understand our worlds (epistemology). By contrast, Gilles Deleuze's explorations of the ontological status of pure difference have been understood as offering an influential coun terpoint to the discussion so far.

Prior to Deleuze, accepted approaches to ontology preceded almost exclusively from the Platonic tradition, a strategy describing what is in the world through formal, ideal categories. Such an ontology operates by describing orders of similitude, where the world is understood by isolating and organizing objects supposed to be of the same type into exclusive families. For instance, while for Plato there supposedly exists a perfect (though imma terial) chair form of which all actual chairs are but imperfect copies, chairs are nevertheless identified as members of that family by virtue of the fundamental characteristics that constitute their similarities to that form, their 'chair ness'. The point behind the Platonic strategy is to find a way to account for everything that is in the world within our own finite and limited language and representation. Thus, he would suggest that there is a form to which every category of 'thing' corresponds: chairs, certainly, but also people, statues, dogs, and so on. This ontology can be termed idealist insofar as the forms are thought to be transcendent; that is, they precede their material incarnation, representing – within the world of forms or concepts – the totality of possible forms that substances might take. Thus, when a substance does take form, according to Plato, we identify what it essentially is by virtue of its similarity to a set of possible forms. This systematic or categorical approach has continuously dogged Western philosophy, insisting upon a structuring analytic that predetermines and limits what can be expected to be in the world.

Deleuze's establishment of an ontology of pure dif ference works against the popular tradition initiated by Plato. Inspired by several strands of minoritarian philosophical thought (encompassing work by Spinoza, Nietzsche, and Bergson, to name a few), Deleuzean ontology describes not a world of similarities shared by static objects, but rather one in which all of materiality is continuously moving, mutating, and transforming, differentiating even from itself in a constant process of becoming. And yet, in working to avoid thinking about the world in terms of similitude, a problem arises in the apparent impossibility of ontological description, both in terms of: (1) the reliance of language itself upon similarities and categories, and; (2) the sheer impossibility of accounting for the ever growing infinity of differen tiations that are continuously unfolding. By contrast, Deleuze presents a world encountered from the per spectives of movement and force relations: rather than structured, whole objects (the human, the subject, what have you), there are continuously interconnecting multi tudes of partial objects affecting and being affected by other partial objects, constituting – if only for a moment – assemblages that appear to cohere by working together or initiating processes that are specific to that relation.

Although he dismissed any such endeavor to produce titles or 'isms', Deleuze is often included among the ranks of post structuralists because his anti essentialist, assemblage based approach to ontology is concerned first and foremost with the introduction of newness and variation in thought and life. While many post structuralist thinkers satisfy themselves by leveling critiques against the systematic and oppressive structurings that pervade social life, Deleuze sought strategies for thinking, living, and relating beyond these. Thought turned toward difference readies itself for encounters with random, alternative arrangements and events emerging through the dynamic, interactive encounters of materiality.

In this sense, Deleuze's philosophy should be considered an active one. Thinking about difference is a matter of asking what bodies can do, and from this springs a literalist conceptualization of ethics as a form of inquiry into 'what is to be done'. Certainly, we are inundated with any number of different norms prescribing to us what our bodies should be doing, through which, as Foucault suggested, we discipline ourselves and police our subjectivity. And, with Deleuze, we are explicitly challenged to interrogate such disciplinarity by experimenting with variations in thought and embodiment that might aid in uncovering what else can be done. Such experiments need not be immense or completely transformative, but should be predicated upon the understanding that a life presents a virtually infinite number of experimentations continuously conditioned by its orientations to and relations with other bodies.

Orienting inquiry toward what things do corresponds with another line of thought that challenges structuralist preconceptions of social life and science, especially spearheaded by Bruno Latour and broadly known as actor network theory (ANT). This work addresses epistemological and ontological problems in the production of scientific knowledge by interrogating how such knowledge is negotiated and by challenging the limited, anthropocentric domain of agency in accounts of materiality. If Foucault was the historian of circulating discourses of power, Latour might be considered their ethnographer. By observing the generation and negotiation of scientific knowledge as produced through everyday lab life, Latour challenges modernist conceptions of truth and transparency in scientific fact by witnessing the ways that scientists, in the process of discovery, actually participate in the creative production of those truths. Rather than being something that preexists and is slowly uncovered through repeated experimentation (a modernist conception), a scientific truth is something that is invented and reworked through situated exchanges by members of the community. Thus, scientific knowledge is always an emergent discourse dependent upon not only other circulating (and equally constructed) knowledges, but also upon the varying positionalities and privileges of the scientists themselves, the vogue of scientific topoi, access to resources and funding, and so on.

At the same time, Latour draws upon something similar to Deleuze's notion of assemblage ontologies to extend the place of agency beyond the human to nonhuman objects that, through various connectivities, bring about certain events or results. Thus, just as we cannot deny the participation of the scientist in the production of scientific knowledge, we likewise cannot ignore the role of various objects (tools, for example) in the accomplishment of the same. To return to Deleuze, we might say that while our concern should still be with what a body can do, we must always keep in mind that no body acts in a vacuum. Rather, as Latour has explained, objects are always immersed in a complex network that is established through the interactions between and among its members. Because there are any number of activities and processes unfolding across this network, it becomes impossible to grant certain participants at specific sites within it (that is, humans) a causal agency over and against that of other participants (that is, objects) that play just as crucial a role in carrying out a desired end.

Such a recasting of ontology, whereby human beings are located alongside a host of nonhuman entities within a broad and complex network, or assemblage, does raise the question of whether or not ANT has taken a step back, philosophically speaking, by returning to the Platonian notion of similitude. Indeed, some have commented on the fact that ANT analyses do run the risk of treating all elements within a network in the same manner, wherein anything and everything has agency in the sense of having the capacity to act or intervene within a situation so as to produce a particular affect. And yet, it would be more in keeping with Latour's examination of agency to see this rather as a necessary entry point into a much more complex form of interrogation, wherein thought is given to how people and things exist in the world such that they are constituted with particular capacities for action, whether it be through focused and deliberate intent, the movement of certain parts, or even a passivity in regard to the actions of other entities. All of these are a form of agency in the sense that they can work to produce particular affects, but their specific manifestation and mode of operation is very much context dependent. And, we would suggest, it is this attentiveness to context that not only augurs a particular role for geographers, long concerned with this topic, in the continuing development of post structuralist thought and practice, but positively demands it.

Post-Structuralist Geographies

Representation and Space

In the following discussion of post structuralist geographies we echo the preceding sections, tackling first epistemic issues and subsequently taking on more avowedly ontological ones. We begin, then, with a dis cussion of representation as a key concept for geog raphers interested in drawing upon, and contributing to, post structuralist debates. Representation can be theo retically distinguished from re presentation by reserving the latter's meaning as implying the impossible, namely, capturing and reflecting – as in confining and mirroring – a real world referent in thought, language, and visual media. Representation, by contrast, refers to the social mediation of the real world through ever present pro cesses of signification.

By illustration, we can return to the relationships among signifiers, signifieds, and real world referents. As should be clear, within post structuralism such relations are not taken for granted; rather, they become foci of analysis in their own right. This is because signifieds are not presumed to stand alone, but are considered to be defined in relation to each other through the socially constructed discourses that give them their definition and character. As a result of the embeddedness of signifieds in discourse, no signifier can be presumed to stand in a one to one relationship with a real world referent. And finally, since referents themselves are not understood as having essential qualities but are instead defined and refined within relational systems of difference (for example, not 'I, the cat…' but 'I, not dog, not bird, not child, etc.'), even the bedrock of real things are, under post structuralism, no longer thought of as things in themselves. In light of this reconceptualization, post structuralists refer to representational processes instead of re presentation, and they direct their investigations toward an understanding of the mediating role of dis courses in representational processes.

There are a number of implications to this theoretical position. First, post structuralists take note of and cri tique forms of thought that distinguish between the real world and its mere re presentation in communication, whether conceived in terms of language, sensory per ceptions, or electronic media. As Deborah Dixon and John Paul Jones note in their 1998 essay, this critique is a source of continuing conceptual confusion between post structuralists and other theorists, such as spatial scientists, since much traditional social thought is predicated on the very distinction between the real world of POEMs and their re presentation in thought and language. Conflated with real/representation are other binary formulations such as materiality/ideology and concrete/abstract. For non post structuralists this distinction, and the impulse to resolve it, implies a faith in the possibility of unmediated re presentation, wherein researchers might actually get it right. For post structuralists, by contrast, there is no Truth lurking behind real world objects. This is not to reject the existence of the world *per se*, but rather to maintain that the world can never be known in a manner that is not somehow already socially mediated. What is more, any claim to know can emerge as complicit with authoritarian forms of power in which a particular group names and frames reality for all. And, it is this noninnocent character of constructs that points to the importance of all the other constructs and to the entire social context within which their interdependencies become fixed and stabilized.

Second, in taking into account these interdependen cies, post structuralists take note of both context and intertext. The former refers to the temporary stabilization of meanings drawn together in the articulation of a discourse that communicates those meanings in a sensible form by establishing differences among them. Context, then, fixes the relational field of meaning, but it does so only by drawing upon previous contexts that are themselves embedded in still other contexts. This inter contextual character of the relations among constructs is intertext, a term specifying how one context is related to others, but also how they might be transformed. As used in Kolson Schlosser's article on the interplay between geopolitics and medical discourse, intertext is the re lational field – of flows, imbrications, and concatenations – for the production of new contexts. To give an example, in reading a book, context might be temporarily established by an author who draws upon meanings established within a genre. The act of reading, however, involves the production of a new field of meaning by the reader within which those meanings are de stabilized and restabilized yet again. In this view, and in virtue of the intertextual character of all communication, meanings cannot be permanently fixed according to: the intent behind their production; their content, genre, or mode of dissemin ation; or the perspectives of the reader. Thus, the mean ings that adhere to signifieds cannot be presumed as fixed or fully present, but are always in process, awaiting their deployment in new contexts.

While the above points to the open and unfinished character of representation, it also suggests a problem for reflexive social analysts – that is, those willing to judge their own work within the purview of representation. Put succinctly, if representation within post structuralism denies the disclosure of Truth, and if the subject is no longer speaking with the security and advantage of Identity, then how are we to trust our and others' ana lyses? Feminists have grappled with these issues at some length. One widely accepted response is to reflect on one's social positionality (white, homosexual, male) vis à vis researched participants and texts, recognizing that the outcome of those inquiries are influenced by the different standpoints that infuse the research process. However, others have argued, following post structuralism, that the register of social experience that gives these stand points their presumed stability (however temporary) is no guarantee of Identity. Moreover, in the wake of the 'death' (or dissemination) of the author, the postmodernist claim to recover marginalized voices through a sensitivity to multiple knowledges becomes suspect. The differences that postmodernist inquiry seeks to activate will not be fully relational if inquiry is driven by an underlying faith in the researcher's ability to demarcate center from margin, or, for that matter, to fix the relationship between the real and the represented.

In considering these claims, some critics have responded with the argument that what is lost in the poststructuralist crisis of representation is any possibility of strong evaluation; in particular, post structuralism is charged with relativism and nihilism. But while it is certainly the case that this approach eschews the notion of an external vantage point from which judgments concerning accuracy and ethicality can be made, this need not lead directly to either relativism or nihilism. This is because, first, the de stabilization of centers – what has been termed in deconstruction as the 'cracking of nutshells' – is very much a political project in the way it points to the constructedness and, hence, the contingency, of all kinds of authoritative claims. To choose to undertake post structuralist analysis is, therefore, already a form of evaluation and intervention. Second, we can acknowledge that post structuralism holds that all evaluative ethical projects, including those undertaken in the name of liberty, community, and democracy, can only be 'evaluated' within the particular spatial–historical contexts of their articulation. However, even though they exist as discursive constructs without guarantees – how could they be otherwise? – this does not imply the end of politics or of evaluation. This is because whereas in structuralist forms of inquiry the researcher unproblematically takes on the certainties invested in the roles of arbiter, analyst, and de coder, in poststructuralism the researcher explicitly interprets, activates, and transmutes meanings and their contexts. Indeed, rather than presume to work within the domain of similitude, researchers understand and articulate their role as proliferating difference by acting out the multiplicities of a mobile researcher.

With these comments on representation in mind, let us now turn to a more explicit discussion of the spatialities of representation by outlining the emergence of a twofold agenda for geographic research: (1) to investigate the spatial character of discourses through an investigation of the geographic meanings embedded in particular representations and discursive sites, and (2) to understand the representational character of space itself. Regarding the first, we can assert not only that representations signify spatially, say by invoking particular places and stamping them (by placing them within a particular context) with specific meanings, but also that any signification or discourse is 'always already' spatial. How can this strong view of spatiality within representation be maintained?

An answer is to be found in the dialectic of space and social power elaborated by writers such as Michel Foucault and those geographers who have been influenced by his work. Foucault's work points to the indivisibility of space and social power – from the ways that social relations are constituted in and unfold through spatial distributions, built environments, and spatial significations, to the ways that space itself is socially produced through relations of social power. In this view, it is untenable to conceive of social relations of class, gender, race, or nation as falling outside of the purview of the spatiality through which they are practiced and reproduced in everyday life. Even the discourses of progress, morality, and reason (to name just a few) emanating from these relations are spatial – marking sociospatial constructs such as public/private, global/local, and chaotic/orderly. As Foucault's own empirical work has revealed, these discourses also carry with them spatial concatenations, attenuations, and disjunctures; that is, they mark the other spaces of immorality, insanity, and the abnormal.

The second moment in this agenda is to consider the representational character of space itself. It is worthwhile first to point to what can be termed a spatial epistemology; that is, our ways of knowing space. Under the aegis of Western reason, this epistemology has been suffused with ocularcentrism, Cartesian perspectivalism, and lineation. Whereas ocularcentrism privileges vision over other forms of knowing and is one basis for social power through surveillance, perspectivalism coordinates vision by situating the viewer from a vantage point above; lineation, meanwhile, is the basis for gridding social space – the Cartesian counterpart to categorization more generally. Any number of histories can be written about the imposition of reason's grid epistemology, from the configuration of social space in the projects of colonialism and nation building, to the carving of towns out of nature and bringing them together through systems of transport, to the policing and self disciplining of bodies in the gendered microspaces of everyday life. Even our concept of scale – cascading as it does through the above examples – can be thought of as an outcome of the grid epistemology.

A relevant example of the differing material transformations wrought by successive articulations of a grid epistemology can be found in Foucault's commentary on eighteenth century France, wherein a crucial shift occurred in regard to how, and by whom, space was conceived and discussed. Previously, space had been the domain of the architects, who envisioned the governed city as a metaphor for the governed territory. The primary spatial trope was one of penetration, whereby all of the city, and by projection, all of the state, was laid open to the regulatory surveillance and practices of the police. As the century progressed, new cadres of experts, including engineers and builders, emerged as the authoritative sources of knowledge on transport linkages and planning. The associated discourses addressed space in terms of speed, mobility, and networks, and entailed a revisioning of the links between the exercise of political power and the lived space of both the city and the territory.

From this theorization of the ways in which spaces are already invested with epistemology we can proceed to rethink some of geography's key objects of analysis, such as landscape. Whereas previous theorizations understood landscapes to be the imprint on nature of a culture, or the effect of social process such as capitalism, post structuralism has pointed to their status as a complex of significations and discourses that are intertextually bound with a host of other landscapes and discourses. The landscape as text metaphor thus sees place as intersecting with an infinite number of other texts and contexts, such that we cannot demarcate where one starts and another begins. These multiple sites of discursive propagation open a circuit beyond the earthiness of landscapes, seeping into other representational media such as film, television, cyberspace, the body, political discourse and other forms of speech, and written texts of all kinds, including maps.

Difference and Space

Much of geography's engagement with post structuralism has revolved around issues of representation, as manifest in a burgeoning series of deconstructive analyses: in these, the seeming authority of particular discursive sites, such as institutions, built environments, and the media, to fix the meaning of objects, peoples, and places, and the relations between them, is undermined by reference to the roles of context and intertext, which allows for their (temporary) production, as well as their reinterpretation. Less attention has been paid to issues of ontology, which, it must be stated, has more often than not been analytically sidelined as a mere product of epistemology. This absence has led to the charge that deconstructive analyses are ontologically agnostic, caring little for the form and content of the material conditions that allow for discourses to be manifested (from pen and paper to landscape itself) and that allow for their dissemination.

In part, this absence can be laid at the door of the discipline more broadly, as geographers have commonly conflated these material conditions with the notion of a preexisting or 'given' Euclidean surface across which all manner of phenomena are arranged and rearranged, and across which social and environmental relations are played out. This presumption of a Euclidean surface pervades the discipline, as when we locate, map, track, and correlate various phenomena, and plays a fundamental role in the operation of the grid epistemology, as described above. Despite post structural efforts to point to the discursive, and thereby fundamentally arbitrary, construction of just such a space it has remained fairly impervious to critique and, hence, to analysis.

And yet, it would be wrong to suggest that post structuralism is necessarily characterized by such agnosticism. In recent years the discipline has been witness to an emerging interest in how the 'nature' of materiality can be productively reinterpreted from just such a perspective. For the most part, this has taken the form of a decoupling of the link between the notion of materiality and that of Euclidean space and the simultaneous articulation of multitudinous spaces, continually produced and transformed through the work of people and things. That is, post structural geographers such as Marcus Doel have begun to articulate these spaces – or spacings – as continually emergent rather than a static dimensionality. These spacings (and their associated timings, for the two can no longer be considered as separate, Euclidean dimensions) are constituted from a multitude of pre personal and affective encounters and connections between and among human/nonhuman assemblages. As geographers such as Nigel Thrift note, many of these encounters confound and/or exceed those acts of cognition and description that are associated with deconstructive modes of analyses, which, as outlined above, tend to focus on the fixing and unfixing of meaning.

In formulating this notion of multitudinous spacings, as opposed to space, some geographers have turned to Deleuze's work on immanence, which is usually described as the reverse of the more traditional notion of transcendence. Transcendence finds its origins in theological, as well as philosophical discourses, in that it points on the one hand to a divine being that stands outside of and beyond the world, and on the other to the more abstract notion of a unifying principle – which may be divine but which may also relate to an idealist notion of the formative power of the human mind to imagine – that underpins all of the phenomena that make up that world. If we refer back to Deleuze's emphasis on difference, then we are reminded that this work is very much an effort to confront a Platonic, Western philosophy of representation, wherein all phenomena are considered but the contingent manifestation of a universal model or standard. For Deleuze, difference does not mean different in resemblance; rather, it refers to a radical alterity, wherein each encounter proceeds to alter the world around us and within us, producing something new rather than more of the same.

Because such changes are wrought without reference to an outside power, or universal reference, Deleuze's ontology can be called one of immanence, where the distinctions that haunt transcendental thought (mind and body, God and matter, interiority and exteriority, and so on) are erased or flattened. Here, there are no preeminent forms, subjects, or structures, nor is there the development or organization of these; instead, we find haecceities (blocks of uniqueness) which, if they can be said to hold a property at all (a term that usually implies having something in common), hold the property that makes each individuation unique. This form of difference is continually renewed as haecceities become tendencies, forming complex assem blages of forces, particles, connections, affects, and becomings, relations of movement and rest, speed and slowness. They unfold within and through what Deleuze terms the plane of immanence, each discovering its own kinetic relations, as well as the relative limits of those relations; in effect, haecceities work to map their environment at the same time that both the environment and themselves are transformed through each and every encounter.

It is in this self organizing immanence that we find Deleuzean and Latourian inspired rejections of funda mental spatial concepts that rely on transcendence. Most prominent among these is that of scale. In its usual, vertical/hierarchical form, scalar levels – whether fixed or in more contemporary theories as socially produced – are an invitation to think POEMs as nested from the local to the global, and everywhere in between (body, household, neighborhood, region, nation). In this layered imaginary we find ample opportunity to sort social processes, such as those driving economies and culture, into their respective levels, as if somehow culture might work at one rung (say, urban), while economy at another (say, nation). This segmentation, which belies the con crete and immanent material contexts within which things cultural and economic reside and do their work is, in turn, overcome by a further idealist maneuver: con necting the processes at one level with those at another, with clever opportunities for horizontal flows followed by displaced touchdowns, and for scalar jumping to different levels. However, from a post structural perspective, the problem with such theorizing is precisely its reliance on a transcendental spatial imaginary, one that is detached from the actual sites in which concrete material objects and practices and their discursive counterparts are lodged – literally so, in the sense of being inserted into, and immanent with, dense unfolding of sites of social activity and material objects.

For example, to write as one might do of Wal Mart's corporate strategies in terms of a global entity operating within an ever expanding international political econ omy, and to frame locational decisions as an example of the struggle between economic globalizations and the preservation of local cultures, is to miss the point that the former are just as local – just as embodied – as the latter. They unfold differentially in sites and depend on the dispersed but nonetheless situated and materially con nected traders who track the firm's stock exchange price signals in virtual reality (New York Stock Exchange (NYSE): Wal Mart Stores (WMT)); on the janitors, furniture, toilets, geographic information system (GIS) experts, computer keyboards, secretaries, and calendars that are put to work in organizing the boardrooms that bring together the decision makers; on the last minute deals cut between the property owner, the real estate developer, the city councilwoman, the transportation planner, and the president of the watchful neighborhood group; and all of them, in their turn, on the existence of that simple, Latourian object, the shipping container, invented only in 1956, and without which none of these dense objects and networks could have mobilized to enable the corporation to bring cheap goods from the social sites (factories) of China to the US in the first place. Globalization is not scaled; nor does it flow, untethered, like the airline route map. To the extent that we can call it anything meaningful after having rejected transcendence, globalization is a process in which materials, practices, and meanings travel long distances (or are taken by others) only to nestle up to one another in place. Even when these connectivities are at a distance – as when these discourses are linked by cell phones, the Internet, or the press – their site specific materialities can always be traced: to the phone's network of transmission towers, to the servers that enable Internet transmissions, and to the materials employed in the printing and distribution of newspapers.

This understanding of radical alterity, or pure dif ference, stands in stark contrast to transcendent versions of globalization that our own cultured, categorized thought tends to turn toward in its search for order and sameness. One of the symptoms of this acculturation is a tendency to assume that what Deleuze refers to as the 'actual', that is a material world characterized by a multitude of groupings such as species, kinds, properties, spatiotemporal locations and subjects, constitutes all that is real and, moreover, is naturally grouped and sorted in this manner. Rather, Deleuze offers, this actual, material realm is but the contingent realization of what he calls the virtual, which is formed not from what is, but rather what can be. Here, the virtual is perhaps best considered a potential that, often hidden from our gaze through a cultural emphasis on order, can occasionally be glimpsed during moments of systemic change or bifurcation. In an ontology of immanence, it is the form of the relation between the actual and the virtual that forms the heart of an analysis. Hence, Deleuze's emphasis on experimen tation can be understood as an attempt to further actualize the virtual realm, as expected identities are troubled by the occurrence of new sensations that defy identification.

Conclusion

As we have shown in this essay, one side of post structuralism is dedicated to unsettling routine modes of social inquiry relying on handed down concepts that purport to contain either essential truths (e.g., progress, reason) or that fixes the meaning of worldly objects and events (e.g., parts of nature, the definition of resistance). These structurings and the centering effects they pro duce are taken as significant objects of inquiry in and of themselves; accordingly, the aim of both Derridian deconstructive and Foucaultian genealogical analysis is to understand how, and with what effect, such concepts operate, and to whose benefit. This tradition of thought has been extremely helpful in rethinking many of the constituent elements of human geography. Our nodding acquiescence to repetitive invocations of scores of nor malized categories has been profoundly shaken under post structuralism. Terms such as community, nature, public/private, identity, scale, experience, space/place, culture, animals, development, history, justice, agency, authenticity, borders, citizen, technology, gender/sexu ality, transgression/resistance, memory, and nation – and more – all have new, if unfixed, meanings. Even what it means to 'post' does not escape the critical gaze of post structuralism! The movement has unhinged these con cepts from their earlier on securities, tossing them into a differential space of relational meanings buttressed by wide sociospatial–historical contexts and everyday social articulations; peering into that space we can examine their stabilizations and de stabilizations, their inexact certitude, and their exacting uncertainties.

Meanwhile, a productive, political moment is seized upon in the ever unfolding and complex differentiation of social and natural assemblages. Post structuralism rejects the notion that there is an ordered trajectory to the emergence and development of socionatural worlds. Concomitantly, it rejects any notion of continuity or order in history. Far from terrifying, our capacity for hope is meant to be expanded from this recognition, for it enables us to invite into play new worlds whose potential is always emergent in the transition from the virtual to the actual. It is this pulsing, vibrating, crumpled, and folded space, rather than the stable, transparent, and bidimensional one, that is both product and recipient of that transition. In the future, making sense of this new view of the possible will require us to revise the analytic metaphors necessary to describing difference, moving us from fixities to fluidities, from rootedness to mobilities, and from hierarchies to networks. And yet, as a 'post' rather than an 'anti', this body of thought will always be very much embedded in a web of relations with respect to its other, structuralism, which is both its trace and its becoming difference.

See also: Actor-Network Theory/Network Geographies; Deconstruction; Foucauldianism; Local-Global; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Posthumanism/Posthumanistic Geographies; Structural Marxism.

Further Reading

Bonta, M. and Protevi, J. (2004). *Deleuze and Geophilosophy: A Guide and Glossary*. Edinburgh: Edinburgh University Press.
Butler, J. (1990). *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge.
Carter, S. and McCormack, D. P. (2006). Film, geopolitics and the affective logics of intervention. *Political Geography* 25(2), 228 245.
Crampton, J. and Elden, S. (eds.) (2007). *Space, Knowledge and Power: Foucault and Geography*. Aldershot: Ashgate.
Dixon, D. P. and Grimes, J. (2004). Capitalism, masculinity and whiteness in the dialectical landscape: The case of Tarzan and the Tycoon. *GeoJournal* 59(4), 265 275.
Dixon, D. P. and Jones, J. P. III. (1998). My dinner with Derrida, or spatial analysis and poststructuralism do lunch. *Environment and Planning A* 30, 247 260.
Doel, M. (2000). *Poststructuralist Geographies: The Diabolical Art of Spatial Science*. Edinburgh: University of Edinburgh Press.
Doel, M. (2005). Deconstruction and geography: Settling the account. *Antipode* 37(2), 246 249.
Foucault, M. (1980). *Power/Knowledge: Selected Interviews and Other Writings, 1972 1977*. (Gordon, C., ed.). New York: Pantheon.
Latour, B. (2005). *Reassembling the Social: An Introduction to Actor Network Theory*. Oxford: Oxford University Press.
Lulka, D. (2004). Stabilizing the herd: Fixing the identity of nonhumans. *Environment and Planning D: Society and Space* 22(3), 439 463.
Marston, S. A., Jones, J. P. III. and Woodward, K. (2005). Human geography without scale. *Transactions of the Institute of British Geographers* 30, 416 432.
Natter, W. and Jones, J. P. (1997). Identity, space, and other uncertainties. In Benko, G. & Strohmayer, U. (eds.) *Space and Social Theory: Interpreting Modernity and Postmodernity*, pp 141 161. Oxford: Blackwell.
Philo, C. (1989). Thoughts, words and creative locational acts. In Boal, F. W. & Livingstone, D. N. (eds.) *The Behavioural Environment Revisited*, pp 205 234. Routledge: London.
Pickles, J. (2004). *A History of Spaces: Cartographic Reason, Mapping and the Geo coded World*. New York: Routledge.
Schein, R. (1997). The place of landscape: A conceptual framework for interpreting an American scene. *Annals of the Association of American Geographers* 87, 660 680.
Schlosser, K. (2007). The bio politics of bodies politic: Nature and intertextuality in classic US geopolitical discourse. *GeoJournal* 69(3), 199 210.
Thrift, N. (2007). *Non Representational Theory: Space, Politics, Affect*. New York: Routledge.
Whatmore, S. (2002). *Hybrid Geographies: Natures, Cultures, Spaces.* Sage: London.

Poverty

A. O. Chimhowu, University of Manchester, Manchester, UK

Defining and Measuring Poverty

In 2001, the United Nations Committee on Economic, Social, and Cultural Rights defined poverty as a "sustained or chronic deprivation of the resources, capabilities, choices, security and power necessary for the enjoyment of an adequate standard of living and other civil, cultural, economic, political and social rights." This definition crystallizes more than a century of research on poverty and shows how far understandings of poverty have come from the early pioneering works of Charles Booth at the close of the nineteenth century. Although poverty remains one of the most prominent concepts in contemporary economic and social development discourses, defining it is notoriously difficult. This is due in part to the politics that usually surround the mere existence of poverty and what to do about it. Similarly, the multiplicity of disciplines involved in its study and the plurality of perspectives this brings along does not simplify things. Cognate disciplines that include economics, sociology, geography, planning, and political science all have some focus on understanding poverty from different angles. More recently, the emergence of development studies as a discipline has provided poverty studies with a focused discipline bringing together the cognate disciplines identified earlier. One thing that is clear from any contemporary reading of the poverty literature is the changing policy discourses and practice as new development paradigms emerge to influence how we look at poverty. With such changes also come shifts in how poverty is defined and the nature of solutions proposed.

Money-Metric Definition of Poverty

In general, contemporary discourses on poverty recognize a range of definitions, although it is more common to distinguish between narrow and broader definitions. Narrow definitions simply consider poverty in income (or the related expenditure or consumption) terms. Early systematic studies on poverty like the 1889 publication of Charles Booth's groundbreaking *Poverty Maps of London*, the 1899 work by Seebohm Rowntree on York, and Robert Hunter's 1904 study of poverty in America established an early trend to focus on income what has variously been termed 'primary' or 'subsistence' poverty – what contemporary discourse generally calls 'absolute' poverty. In these early works, a household was deemed as living in poverty if its income did not allow it to access what Rowntree termed "minimum necessities for the maintenance of physical efficiency." Clearly, under this narrow conceptualization poverty is understood as a simple lack of sufficient income to perform productive (read labor output) and reproductive roles. Households whose incomes fall below this poverty threshold are more likely to be on social welfare assistance programs in countries where such schemes exist.

Since the 1970s and following the work of Peter Townsend, a distinction is often drawn between absolute and relative poverty. Absolute poverty refers to those people living below an officially determined poverty threshold, which generally varies by size of family unit and composition. For example, a household living on less than a US dollar a day in developing countries might be considered to be living in absolute poverty. In the UK, this measure is not used much; although in political hyperbole, households with less than half the inflation adjusted average 1997 incomes are seen as living in absolute poverty.

The concept of relative poverty is commonly used to refer to the extent to which household income is below the average household income in a given country. It is, therefore, often seen as indicative of the level of inequality in society. While absolute poverty's poverty threshold is usually independent of time (only adjusted occasionally through political decisions), the relative poverty threshold tends to be adaptable and changeable as local income levels and standards of living change. For example, in the European Union, households whose income resources are below 60% of contemporary median net disposable income of that country are considered as living in poverty. This median value changes as national economic fortunes adjust over time. Relative poverty is often used as an indicator of how unequal a society is. In general, except for the United States where both absolute and relative poverty concepts are used officially, most developed countries tend to use relative poverty, while less developed economies use the absolute poverty concept.

Broader Definition of Poverty

While money metric defined poverty makes it easier to 'objectively' determine who is poor and who is not, the premise upon which it is based has often been criticized as misleading. For a start, at household level, defining poverty this way ignores welfare variations among household members. Some have argued that the multidimensional nature of poverty means that defining it in income terms fails to take into account the welfare characteristics of households living in poverty. Specifically, in the late 1970s, it was argued that raising incomes does not necessarily increase household welfare if the incomes cannot provide basic needs. It has therefore been suggested to focus on basic households entitlements. This implies that a household without access to basic entitlements such as decent shelter, food, clean drinking water, sanitation, health, and education be deemed as living in poverty. Following on from the groundbreaking works of Amartya Sen in the 1990s the term deprivation, is used more commonly in poverty discourses to draw focus away from income only, but rather on what people are able to do or be. This implies a focus not on the utility but on the welfare benefits of an income. Poverty in this context is then defined broadly as lack of capabilities rather than income. For example, being educated, well fed, and being free to exercise choice gives an individual a better living standard than being wealthy but in ill health. In contemporary poverty discourses, this new generation of poverty definitions coexists alongside the money metric ones, and is seen as complementary rather than replacing the former. Those working with this broader definition tend to use the term capability poverty to distinguish it from conventional money metric measures.

In developed economies, adoption of this capability poverty approach has been subsumed into new discourses on 'social exclusion'. This acknowledges that those living in poverty are unable to participate in institutions used by mainstream society. Contemporary poverty research in developed economies therefore tends to take place under this social exclusion label, although the term is also used in developing contexts.

Chronic/Transitory Poverty

In addition to understanding the scale and depth of poverty, contemporary concern is also with the duration of poverty. A distinction is now often made between those living in long term or chronic absolute poverty, and transitory poverty (those who are occasionally or sometimes poor). There are variations about the length of time that qualifies an individual as living in chronic poverty and in the main this is arbitrary and often depends on data availability and the cycles of surveys. Emerging consensus, however, suggests that ultimately it is the data availability that can tell us which households have remained in poverty between censuses and therefore in chronic poverty. The upper limit for chronic poverty is often a lifetime. At the turn of the millennium estimates suggested that out of the estimated 1 billion people living in extreme poverty (defined using the 'international' US$1/day line), up to 420 million experience chronic poverty.

The concept of chronic poverty allows us to distinguish households that have always been poor (have never been above the poverty threshold) and those that are usually poor (may occasionally be above the line but on balance are below the poverty threshold most of the time). The reason for this distinction is largely to do with the nature of policy responses to the different poverties. Transitorily poor households include those whose fortunes fluctuate around the poverty threshold. In agrarian societies, this might include households that emerge out of poverty during a good agricultural season, but fall below the line with any adverse event to their livelihoods. Also included in the category of the transitorily poor are households that have experienced poverty before but on balance have living standards above the poverty threshold. For example, a household whose breadwinner loses a job may decline into deprivation till the means of livelihood are restored, at which point the household may live above the poverty threshold.

Classifying poverty in this way enables policymakers to develop better targeting mechanisms. Chronic poverty or hard core poverty is more entrenched and therefore more difficult to deal with when compared with transitory poverty.

Measuring Poverty

A key question for those trying to formulate policies to fight poverty relates to how poverty is measured. Several measures or indicators of poverty are currently in use. Depending on the definition adopted, a distinction is made between 'objective' measures and those of a subjective nature. When poverty is defined in income terms, the poverty threshold is often used to measure the magnitude and depth of poverty. Since the 1960s, following the work of Mollie Orshansky of the US Social Security Administration, a poverty threshold based on the minimum cost of a nutritional diet for differently sized households is used as the basis for poverty lines. Different countries tend to compute the poverty thresholds differently and have also come up with different types of poverty lines. For example, like many countries, Zimbabwe draws a distinction between the food poverty line (FPL) as the minimum cost of a nutritional diet for an average size household and the total consumption poverty line (TCPL), which refers to the minimum cost of the acceptable minimum living standard (including food, clothing, education, and health). A family experiencing food poverty is seen as being more vulnerable than one living below the TCPL. Setting the poverty line is often political as decisions have to be taken about what constitutes acceptable living standards, and this often raises controversies. While in the past, professionals have often taken a lead in compiling the poverty lines, we now know from recent work on 'voices' of the poor that poverty lines can only be indicative rather than definitive when it comes to setting the boundaries of what minimum consumption standards are.

Once the poverty threshold has been set, some suggested that the use of key indicators for poverty can be used to analyze patterns, depth, and trends of poverty. Following on the seminal works of Forster, Greer, and Thorbecke (FGT) in the 1980s, three measures (the headcount index, the poverty gap, and poverty severity), generally in common use, are generally called the FGT class of poverty indicators. The three poverty indicators are computed in relation to the 'poverty line' and are the most commonly used. The simple head count index (HCI) can tell us the proportion of population with incomes or consumption levels falling below the poverty line. Expressed as a percentage, this gives us an idea of what share of the population lives in poverty and can be monitored over time to give an idea of the progress made in reducing poverty. Based on the poverty line, it is also possible to calculate the poverty gap ratio, which gives an idea of the depth of poverty or how far the poor are below the poverty line. Households with incomes or consumption above the poverty line are considered as being nonpoor. The third FGT indicator is the poverty severity index, which gives us an idea of inequality among the poor. This is calculated as the average squared income or consumption shortfall of the poor households (as a percentage of the poverty line). This measure assigns more weight to households with incomes far below the poverty line and is often seen as being a response to Amatya Sen's view that income transfers from a poor household (already below the poverty line) to an even poorer one should be reflected in an improvement of how we measure income poverty. In general, a higher index reflects a high number of very poor households among those living in income or consumption poverty.

Although widely used, the FGT indicators carry the inherent weakness that they capture poverty at a moment in time and fail to take account of fluctuations in mean income or consumption. Further, collecting income data through surveys is notoriously difficult and can often be misleading. Apart from the problems of individual recall, there is evidence to suggest poverty can be over or understated as we may not capture the full portfolio of resources available to a household. It is, therefore, generally the case that household consumption data are preferred to income data as a measure of poverty. While sounding objective, it should be borne in mind that it still involves value judgments on what goods and services to include, what to do about it, and who makes that decision.

Increasingly, it is also the case that people may 'feel' they are living in poverty even if official lines may suggest otherwise. A subjective assessment of poverty (e.g., the Gallup Polls in the USA) tracks the 'gut' feeling of individuals and what they consider to be the poverty threshold. Following on from the works of Robert Chambers in the 1990s and the subsequent publication of *Voices of the Poor* in 2001 by the World Bank, there is a general consensus that subjective measures provide credible 'lived' experiences of poverty that can inform policy responses when used in conjunction with or to triangulate other 'objective' measures. In most developed countries, experience with 'subjective measurements' shows households with incomes below the median consider themselves as living in poverty. This is usually a much higher threshold than official poverty lines.

In some more developed countries, a simple record of the number of people on state income support or welfare benefits is also used as a proxy measure of poverty. It is, however, also widely accepted that not all people living on income support live in deprivation and also that not all who live in deprivation are able to access income support.

Comparing Poverty Levels across Countries

Although there are some international benchmarks on minimum living standards (e.g., daily per capita calorie intake differentiated by age and sex), understanding poverty in relative terms means there are variations in what could be defined as poverty across countries. This makes comparisons across countries difficult. A way out of this has been the adoption of an absolute poverty line to define populations living in poverty and on this basis to set an internationally comparable measure of income or consumption that is deemed adequate for households to avoid deprivation. The World Bank suggested that individuals living on less than US$1 per day adjusted for purchasing power parity (PPP) based on 1985 prices be classified as living in extreme poverty. The PPP is a method used to take account of the relative purchasing power of different countries' currencies over the same types of goods and services. Its main function is to allow comparisons in living standards to be made across countries. Till 1993, the PPP was based on the Penn World Tables. Since then, consumer PPP based on World Bank estimates is used. This change has meant that the poverty threshold set in 1985 is actually US$1.08 when we recalculate the 1993 PPP. For simplicity, the figure of global poverty threshold for developing countries is set at US$1/per day/person. Using this simple measure of absolute poverty, it is possible to estimate the number of people living in severe poverty. **Table 1** shows the regional shares of the people living in severe poverty and the changes over time. This clearly shows that sub Saharan Africa and South Asia have more people living in severe poverty than any other region.

Table 1 Percentage of population living on less than a 1US $ a day by region[a]

Region	*1981 (%)*	*1990 (%)*	*2002 (%)*
East Asia and Pacific	57.7	29.6	11.6
China	63.8	33.0	16.6
Europe and Central Asia	0.7	0.5	2.1
Latin America and the Caribbean	9.7	11.3	8.9
Middle East and North Africa	5.1	2.3	1.6
South Asia	51.5	41.3	31.2
Sub-Saharan Africa	41.6	44.6	44.0
Global	40.4	27.9	21.1

[a] NB Data for East Asia excludes China.
Based on the data from: World Bank. (2004). Millennium Development Goals Data. Web URL http://ddp ext.worldbank.org/ext/GMIS/gdmis.do?siteId = 2&goalId = 5&targetId = 15&menuId = LNAV01GOAL1SUB1 (accessed in Aug. 2008).

Although in 1990, some 1.2 billion people (27.9% of global population) lived in poverty (less than US$1/day), by 2004 estimates suggested that 19% of the world's 6.6 billion people lived on less than a dollar a day. This indicates some progress in reducing global poverty rates, although it should be mentioned that most of the gains have been in Asia. The decline in the number of people living in poverty has largely been attributed to the global initiatives to reduce poverty in the framework of the Millennium Development Goals (MDGs). Some 189 countries have signed up to eight goals meant to address key social development problems. The first of the goals is the reduction of extreme poverty and hunger with a target to "reduce by half the proportion of people living on less than a dollar a day" using 1990 as a base year.

Comparing Capability Poverty across Countries

Apart from these income/consumption based measures, since 1997 those working with a capability poverty definition also make use of a human poverty index (HPI) to measure poverty. Disparities in global living standards meant it was necessary to construct an index for developing countries (HPI 1) and another for more developed economies (HPI 2). This is a composite value based on four main parameters known to influence human capabilities. For HPI 1, these are: probability of not living to 40 years, adult literacy rate, average population without clean water, and the percentage of children born under weight. For HPI 2, these are the percentage of people likely to die before the age of 60, the percentage of people whose reading is far from adequate, proportion of the population whose disposable incomes is less than 50% of median income, and proportion in long term unemployment (more than 12 months). **Table 2** shows the top ten and bottom ten ranked countries on the HDI index and their performance on the HPI index. The contrast between the top ten and the bottom ten is clear to see when we take a look at some of the individual indicators used to compute the HPI. Differences in the quality of life clearly show the North–South divide.

Spatial and Social Distribution of Poverty

North–South Divide in Poverty Distribution

Globally, most of the people living in poverty are found in what the United Nations in 1971 termed less developed countries. These are economically challenged and facing institutional and human resource inadequacies that are often compounded by geography (location in space) and nature and human induced constraints. There are several ways of classification, but the two main ones emanate from the Bretton Woods Institutions (The International Bank for Reconstruction and Development (IBRD) and the International Monetary Fund (IMF)) and the United Nations. The IBRD classification is based on per capita gross national product (GNP). Countries with GNP per capita less than US$750 are seen as low income countries or less developed countries (LDCs). Currently, most countries in the Northern hemisphere have high GNP per capita, while those in the Southern hemisphere have low GNP per capita. This North–South divide often leads to generalizations that there are more poor people in the Southern hemisphere than there are in the Northern hemisphere. This, however, does not imply the absence of poverty in the more developed countries, but merely indicates in relative terms where people are more likely to be living in poverty.

Since 1990, the UN has come up with an alternative classification system based on a composite index they call the human development index (HDI). This brings together an income measure based on gross national income (GNI). A country with a GNI per capita less than US$750 would be classified as an LDC. The HDI also recognizes human resources constraints using a human assets index (HAI). This considers human resource issues such as literacy rates, nutrition, health, and education, in general. In addition, an economic vulnerability index (EVI) that takes account of: instability of agricultural production, instability of exports of goods and services, share of manufacturing and modern services in GDP, merchandise export concentration, and population displaced by natural disasters. Based on these, the HDI is constructed triennially by the UN Committee on Development Policy. In 2007, there were 50 countries (up from 43 in 1996) classified as LDC. Of these, 34 are in Africa. In these countries, a majority of the people are likely to be living in poverty.

Table 2 Human poverty ndex va ues for top ten (HP -1) and bottom ten (HP -2) countr es on the human deve opment ndex

HDI Rank	*Country*	*HPI-1 (%)*	*HPI-2 (%)*	*Probability at birth of not living till 40 2000–05*	*Probability at birth of not surviving to age 60 2000–05*	*Adult Illiteracy 1995–05 % aged over 15 years*	*Adults lacking functional literacy skills 1994–03 % aged 16–65*	*Population living on less than US$1/ day 1990–05*	*Population below 50% of median income 2000–04*	*Rate of long-term unemployment 2006*
1	ce and	na	–	na	5.9	na	–	na	–	0.2
2	Norway	na	2	na	7.9	na	7.9	na	6.4	0.5
3	Austra a	na	13	na	7.3	na	17	na	12.2	0.9
4	Canada	na	8	na	8.1	na	14.6	na	11.4	0.5
5	re and	na	18	na	8.7	na	22.6	na	16.2	1.5
6	Sweden	na	1	na	6.7	na	7.5	na	6.5	1.1
7	Sw tzer and	na	7	na	7.2	na	15.9	na	7.6	1.5
8	Japan	na	12	na	6.9	na	–	na	11.8	1.3
9	Nether ands	na	3	na	8.3	na	10.5	na	7.3	1.8
10	France	na	11	na	8.9	na	–	na	7.3	4.1
167	Burund	37.6		38.2	na	40.7	na	54.6	na	na
168	Congo (Democrat c Repub c of the)	39.3	na	41.1	na	32.8	na	–	na	na
169	Eth op a	54.9	na	33.3	na	64.1	na	23	na	na
170	Chad	56.9	na	32.9	na	74.3	na	–	na	na
171	Centra Afr can Repub c	43.6	na	46.2	na	51.4	na	66.6	na	na
172	Mozamb que	50.6	na	45	na	61.3	na	36.2	na	na
173	Ma	56.4	na	30.4	na	76	na	36.1	na	na
174	N ger	54.7	na	28.7	na	71.3	na	60.6	na	na
175	Gu nea-B ssau	44.8	na	40.5	na	–	na	–	na	na
176	Burk na Faso	55.8	na	26.5	na	76.4	na	27.2	na	na
177	S erra Leone	51.7	na	45.6	na	65.2	na	57	na	na

Complied from UNDP (2007) *Human Development Report 2007/8* Hampshire and New York Palgrave/Macmillan

Rural and Urban Divide

Apart from the North–South divide, there is also a rural–urban divide in the spatial distribution of poverty. This is particularly evident in the largely agrarian societies of the South where poverty tends to be largely a rural phenomenon. Current World Bank estimates suggest that up to 70% of those living in poverty are found in rural areas of the South, and make a living mainly from agri culture related activities. It is also known that most of the rural dwellers in LDCs are poor because of lack of land and poor returns for their produce, including labor. It has also been suggested that part of the poverty in the rural areas of the South can be explained by 'urban bias' in policymaking. This bias has seen rural areas lacking in basic social service provision and infrastructure, as urban based policymakers tend to provide better facil ities in urban areas. Through increased urbanization, however, estimates by the United Nations at the turn of the century suggest that by 2025, there will be more poor people living in urban zones of developing countries than in the rural areas. This urbanization of poverty is often seen as being driven by rural to urban migration under conditions of poor economic growth that characterizes the developing countries.

Social Dimensions of Poverty

Poverty discourses also recognize that within populations in both North and South, poverty is borne by certain social groups more than others. The Fourth World Con ference on Women held in Beijing in 1995 brought to global prominence knowledge that poverty is not gender blind and that households headed by women are more likely to live in poverty than those male headed ones. The term feminization of poverty has been used mostly since the 1970s to describe this socially constructed predis position to poverty. A key causal factor is gender bias built on male dominance in politics, society, and economy. This is often seen as creating conditions to disadvantage the female gender economically. For example, in 2000, the United Nations estimated on average that women earn slightly more than 50% of what men earn. This often creates economic disadvantages that often drive female headed households into poverty. Apart from gender, eth nicity, social class, age, and caste are often seen as factors that predispose an individual to poverty.

Understanding and Explaining Poverty: Theories of Poverty

If defining and measuring poverty are difficult, the con temporary discourses on what causes poverty show an even deeper level of plurality. This is in part a reflection of the diverse nature of disciplines engaged in the study of poverty. The literature distinguishes between what we can call geographical and socioeconomic theories of poverty. Here, a cursory look at these is in order to contextualize the responses to poverty.

Geographical Theories of Poverty

In general, the geographical theories of poverty proffer explanations for the spatial distribution of poverty. In trying to answer why some places become poverty traps and others do not, the theories bifurcate into determin istic and possibilistic perspectives. The former tend to focus on the influence of the physical landscape in causing poverty, while the latter's focus is on 'spatiality' or the interaction between the physical conditions and human processes to give poverty outcomes. Determinists usually seek to identify the physical characteristics that result in 'spatial poverty traps'. Early works of writers such as American geographer Ellsworth Huntington tended to credit the physical environment, especially climatic conditions, with determining development outcomes. Thus, polar and tropical regions were seen as being predisposed to poor economic development out comes, while the mid latitudes were considered as more stimulating and hence offering better prospects for eco nomic development. Possibilistic thinking suggests that the physical environment merely influences rather than determines development outcomes, and hence poverty. For example, marginality and remoteness of a place is only seen as influencing the co production of under development and spatial poverty traps when we take account of the specific 'spatiality' of the area. This means, rurality need not mean poverty in developing countries, unless certain specific historical, economic, and social conditions create such conditions. Rural areas only be come poverty traps because of the nature of development decisions that are taken and the role ascribed to the rural space economy in the development processes.

Socioeconomic Theories of Poverty

Two main sociological theories offer some ideas on why an individual or a household lives in poverty. The first, often called the individual pathologies theory, blames those living in poverty for their condition. Usually based on observations of those living in poverty, this view suggests that there are individual inadequacies that pre dispose those living in poverty to their situation. Being lazy, addiction to the welfare system, low intelligence, or some mental and physical disability, for example, are some attributes identified in this thinking to explain why an individual cannot take opportunities that have allowed the nonpoor to thrive. In the 1990s, Hernstein and Murray's infamous bell curve theory is a typical example of such works. This thinking has been criticized for failing to appreciate that an individual is not necessarily a free agent with all livelihood options available. Rather there are structural barriers that inhibit them from taking advantage of opportunities out of poverty. Policy responses in this frame of thinking tend to focus on correcting the individual pathologies through, for example, weaning the individual from welfare benefits ('can work, must work' policies) by offering targeted assistance to the individual.

An alternative to this individualization of poverty thinking is focuses on the family and the community as the cause of poverty. This is also often called the culture of poverty thinking. Its main argument is that poverty is rooted in socially generated belief systems. Through socialization (transmission of values, beliefs, and skills) poverty reproduces itself through intergenerationally reinforced behaviors. For example, a child growing up on welfare benefits is often seen as being inducted into this subculture and will likely grow up to live in poverty as an adult. While offering an alternative narrative, this way of thinking has been criticized for racial stereotyping and for failing to understand the structural factors influencing the emergence of subcultures.

Based on a critique of the individual pathologies and the culture of poverty narratives, respectively, an alternative explanation centered on the structural causes of poverty emerged. This argues that poverty is caused by structural barriers within the capitalist economy, and in some cases political systems can produce poverty.

Theories of Development and Poverty

Two main theories of economic development that emerged after World War II proffer different interpretations of lack of development and ultimately poverty. On the one hand is 'modernization' thinking that uses cultural and technological differences among societies to explain underdevelopment and ultimately poverty. According to this theory, affluence rather than poverty is what needs explanation and the answer revolves around a modern industrial economy. Technology is seen as the lynchpin of economic development. Certain cultural environments are seen as being inimical to technology and progress, and traditionalism is the biggest barrier to development and ultimately opportunities for wealth creation. Over time, as countries' economic productivity grows, wealth is generated and the process of modernization spreads (initially from some core economic hubs) across the space economy removing underdevelopment (read backwardness). According to modernization thinking, poverty therefore can be solved by modernizing traditional societies. Capitalism is seen often as the solution rather than the cause of poverty and developed countries are seen as playing a more positive role in reducing poverty in the LDCs. In contemporary poverty discourses, analysis of initiatives like the Poverty Reduction Strategy Papers linked to the highly indebted poor countries (HIPC) suggests that these are modern variants of modernization thinking meant to kick start local economic development. Critiques of this theory have mostly focused on the fact that it uses the history of European development to try and influence development in parts of the world with different historical trajectories and their own value systems. Further modernization is simply seen as a way of extending markets for goods produced in the more developed countries.

On the other hand is the 'dependency' theory which suggests that at a global level, poverty in LDCs can be explained in part by the exploitation of poor countries by the rich ones, owing to the structure of the world economy that privileges some countries at the expense of others. The key argument is that poor countries in the South are kept in poverty by producing raw materials that are exported cheaply and processed in the more developed countries and sold back to them as manufactured or value added goods. For most post colonies, the roots of contemporary poverty are often still traced to a colonial experience that shaped specific economic relations between the more developed and LDCs. This theory suggests that underdevelopment is an ongoing process rooted in the 'dependent' relations created by the structure of the global capitalist economy. Initially, solutions to this 'structurally induced' poverty lay in delinking from this exploitative global economic system, substituting imports with locally produced goods, value adding, and formation of mutually beneficial regional trading blocks. More recently, the 'trade dependency development pathway' as a solution to poverty has emerged. This suggests that countries can trade themselves out of poverty through fair and ethical trade. Fair trade has become one of the most prominent populist antipoverty campaigns with a two pronged strategy targeting removal of trade barriers (and giving preferential access to lucrative markets) for poor countries, on the one hand, and campaigning for more ethical value chain structure that pays fair prices to the commodity producers in developing countries, on the other.

See also: Dependency; Determinism/Environmental Determinism; Fair Trade; Livelihoods; Modernization Theory; Possibilism.

Further Reading

Chambers, R. (1997). *Whose Reality Counts. Putting the first Last*. London: Intermediate Technology.

Chronic Poverty Research Centre. (2005). The Chronic Poverty Report 2004 05, CPRC Manchester, http://www.chronicpoverty.org/pdfs/CPR1%20FINAL/CPRfinCOMPLETE.pdf (accessed in Aug. 2008).
Hulme, D. and Shepherd, A. (2003). Conceptualizing Chronic Poverty. *World Development* 31(3), 403 424.
Lipton, M. (1977). *Why Poor People Stay Poor: A Study of Urban Bias in World Development*. London: Temple Smith.
Narayan, D. (2000). *Voices of the Poor. Can Anyone Hear Us*, vol. 1. Washington: World Bank Publications.
O'Brien, D. *et al.* (1997). Poverty and Social Exclusion in North and South. IDS Working Paper 55, University of Sussex, http://www.ids.ac.uk/ids/bookshop/wp/Wp55.pdf (accessed in Aug. 2008).
Sen, A. (1993). Capability and Well Being. In Nussbaum, M. & Sen, A. (eds.) *The Quality of Life*. Oxford: Clarendon Press.
Townsend, P. (1979). *Poverty in the United Kingdom*. London: Penguin.
United Nations. (2006). *Millennium Development Goals Report*. New York: UN.
World Bank. (2004). Millennium Development Goals Data. Web URL http://ddp ext.worldbank.org/ext/GMIS/gdmis.do?siteId = 2&goalId = 5&targetId = 15&menuId = LNAV01GOAL1SUB1 (accessed in Aug. 2008).

Relevant Websites

http://www.chronicpoverty.org
Chronic Poverty Research Centre: Chronic Poverty Report 2008 2009.

http://www.sed.manchester.ac.uk
Institute for Development Policy and Management (School of Environment and Development The University of Manchester).

http://www.imf.org
International Monetary Fund: Poverty Reduction Strategy Papers (PRSP).

http://www.worldbank.org
The Worldbank: PovertyNet.

http://www.undp povertycentre.org
United Nations Development Programme (UNDP): International Poverty Centre (IPC).

http://www.undp.org
United Nations Development Programme (UNDP): Poverty Reduction Programme.

Poverty, Rural

A. R. Tickamyer, Ohio University, Athens, OH, USA

Glossary

Core Powerful and affluent nations characterized by industrial and postindustrial economies that dominate international affairs.

Culture of Poverty Theory that poverty is the result of dysfunctional cultural traits that condone behavior and attitudes antithetical to economic advancement.

Development Processes of economic growth, poverty alleviation, and related social and political changes that are the goals of poor developing nations of the periphery.

Globalization The process of increasing interdependence and exchange of capital, goods, and labor in the world system.

Global North The same as core nations.

Global South The same as peripheral nations.

Modernization Theory of development that sees economic growth as a series of preset, progressive stages that all nations must experience to achieve prosperity.

Parastatal Semi-independent agency used to carry out central government policies related to economic growth and globalization.

Periphery Poor developing nations and regions, often, but not always, former colonies of core nations, sometimes divided into periphery and 'semiperiphery' or nations that have advanced economically beyond the least developed countries.

Poverty Threshold A measure of poverty that indicates the income below which a person is classified as poor.

Social Exclusion Social isolation and marginalization as an outcome of poverty.

Underdevelopment Impoverishment and dependency as the outcome of exploitation of the periphery by core nations, elites, and organizations.

World System The global division of labor that is characterized by unequal terms of exchange between states and corporations of the core and those of the periphery.

Introduction

Across the globe, the world's population is in the midst of a historic shift from residence in predominantly rural regions to majority urban. The growth of mega cities and the constant stream of migration from rural to urban locations focus attention on urban issues and problems, including the ravages of poverty in densely populated and often crime ridden urban slums. Nevertheless, poverty remains disproportionately located in rural places. Currently three quarters of the world's poor are estimated to reside in rural regions, and this figure will decline only slowly, with rural poverty remaining the dominant form of deprivation for many years to come. Rural poverty is not relegated to the poorest regions and nations, but is also disproportionately found in many core industrial and postindustrial nations. In the United States, rural poverty rates rival or exceed those of urban inner cities, and while statistics for other nations of the Global North are not always organized by spatial categories, large, sometimes hidden, pockets of rural poverty remain in the most economically advanced and powerful countries of the globe.

The rural poor are not just a reservoir of misery and deprivation for millions (currently almost a billion persons according to the International Food and Agricultural Development Agency (IFAD)), but they also form a link between the problems of rural and urban centers and developing and developed regions. Rural poverty fuels a constant stream of migration between rural and urban areas as economic and political refugees, guest workers, and migrant labors seek better opportunities in urban centers within their nations of origin and across national boundaries. As part of an accelerated flow of the transnational exchange of labors, goods, and capital known as globalization, the rural poor provide an enduring source of cheap labor for the world's fields and factories, often exchanging the hardships of rural poverty for the problems of the urban proletariat and underclass. Conversely, when economic recession or political pressures make urban life unsupportable, rural areas may act as a refuge for return migrants, exacerbating the already severe problems of these areas.

Although the causes, consequences, and issues facing the rural poor share many similarities regardless of regional status and location, spatial context plays an important part in shaping forms of rural poverty, and significant differences between the Global North and South dictate separate consideration. In the poor and developing countries of the Global South, rural poverty is still by far the most dominant form, and it is closely affiliated with marginal agricultural and natural resource livelihoods and landscapes. In affluent nations of the core, such as the United States, Western Europe, and parts of Asia, rural poverty tends to be hidden and ignored, rendered invisible by the concentration of population and attention to urban centers. Nevertheless, the existence of reliable, regular data collection on eco nomic conditions in core nations provides relatively good information on the degree of deprivation experienced by rural populations. The relative neglect and failure to acknowledge serious rural poverty in core nations make this phenomenon especially pernicious. This article will first consider important conceptual and analytic distinctions necessary to understand rural poverty and then will briefly review the topic in peripheral nations, in core nations, and the connections between the two.

Conceptual and Analytic Issues

The study of rural poverty is a multidisciplinary enter prise, spanning the social sciences and represented in disciplinary based literatures and interdisciplinary policy analyses. Disciplines most often represented in rural poverty studies include economics, especially agricultural economics, sociology and rural sociology, geography and regional science, anthropology, political science, international relations, development studies, and policy analysis. Numerous journals, both disciplinary based and interdisciplinary analyze rural poverty, most often with a specific country or regional focus. Inter national and parastatal organizations such as the United Nations and its many agencies, the World Bank, and regional development agencies and associations conduct poverty assessments as part of their mission. Finally, many nongovernmental agencies such as the Ford Foundation sponsor research, fund programs, and pro duce reports on the topic. The wide variety of disciplines, agencies, and political entities whose agenda at least touch on rural poverty create the need to establish common understanding of key concepts and issues. These include definitions of both rural and poverty, spatial versus demographic approaches to poverty, and rela tionships between poverty, development, and inequality. Variations in definitions and approaches mean that assessments of rural poverty inevitably run into problems of comparability across time and space.

What Is Rural?

Definitions of rural vary by jurisdiction. Although the idea of rural evokes images of sparsely populated agrarian landscapes, not all rural residences or live lihoods are agricultural or even natural resource based, and not all rural populations are sparsely distributed. Furthermore, political and administrative considerations are always present in official definitions of rural places, and these vary widely from nation to nation and within nations over time. Definitions of what is classified as rural by different nations vary from a population of less than 2000 persons per settlement to more than 10 times that amount. Administrative definitions and boundaries may give alternative statistical accounts even within a particular nation. For example, in the United States, the official definition of rural is a place with less than 2500 persons, but the measure most often used to collect, analyze, and disseminate information on rural affairs is the county classification of nonmetropolitan, a proxy for rural that measures lack of urban development rather than rurality directly. Nonmetropolitan counties often contain settlements that exceed the 2500 person limit and are officially defined as urban. Similarly, metropolitan counties that contain world cities may have sparsely settled agricultural or natural resource areas that fit the most stringent definition of rural. Currently, there is no standard definition of rural that spans all nations, and it is safe to conclude that the lowest common denominator for defining rural remains not urban.

What Is Poverty?

The definition of poverty also varies greatly across time and place, with implications for the accuracy and com parability of the assessment of the extent and forms across regions and nations and, even within nations, across time. Although, poverty is variously defined in economic, psychological, and even philosophical terms, most measures are based on some variant of capacity for material consumption. For international comparisons among developing countries, the World Bank poverty lines of purchasing power of less than the equivalent of \$1 and \$2 per day are standard. Increasingly, these measurures are supplemented with multidimensional indices such as the United Nations Development Program (UNDP) Human Development Index (HDI) and Human Poverty Index (HPI) that provide more extensive information on different facets of poverty and deprivation such as life expectancy, literacy, and access to safe water and healthcare. Countrywide studies and reports by international organizations such as the United Nations and the World Bank provide a wealth of infor mation on rural poverty for developing countries, always recognizing the limits posed by different definitions of both rural and poverty and limited capacities to collect statistical data of the countries involved.

More affluent nations that possess far more efficient data collection apparatus and procedures are divided between those that favor absolute versus relative meas ures of poverty. For example, the official US poverty line is calculated by the government on the basis of the income necessary to maintain a minimal subsistence level determined by the cost of a market basket of consumer goods that is adjusted for family size and inflation. This provides the means to determine how many and what proportion of different populations fall below an absolute poverty threshold, either over time or across space. Alternatively, countries that are part of the European Union (EU) typically take what is called a relative approach to measuring poverty by drawing the poverty line at income less than half the national median, calculating the proportion of persons falling below that threshold. The difference is between a measure that purports to calculate absolute levels of deprivation regardless of how many or how few people fall into that category versus one that is sensitive to the distribution of income and thus is as much a measure of inequality as of poverty. Finally, many more affluent countries, especially in the EU, also focus on the effects of poverty with measures of social exclusion, that is, the degree to which impoverished individuals and groups experience social isolation and marginalization. Regardless of type of measure, to provide meaningful data and comparisons, it is necessary to collect information on numbers of poor persons of varying social and demographic characteristics by spatial location. It is not possible to accurately count rural poverty without good measures of both poverty and rural.

Poverty of Persons and Poverty of Place

Poverty can be conceptualized demographically or spatially. Studies of poverty of persons focus on the social, economic, and even psychological characteristics of the poor; studies of poverty of place focus on institutional and structural characteristics of places with large numbers of poor people or disproportional rates of poverty. Of course, these are analytic distinctions, the two are closely connected, yet they represent different conceptual approaches. Rural poverty is by definition spatially defined, and much research examines the unique characteristics, causes, and consequences of poor rural places. At the same time, within rural regions, sociodemographic characteristics of poor people may be examined for their own informational value and to determine if they differ from other types of locales. Similarly, the demography of poverty may be compared across spatial boundaries to determine the relative situation of different population groups across spatial boundaries.

Poverty, Inequality, and Development

Poverty can be viewed as an isolated state, but it is more productively understood as a relational condition, usually accompanying serious inequalities in resource access and distribution within and across societies. Hypothetically, it is possible to conceive of a population characterized by a perfectly equal distribution of deprivation, but in practice, poverty is intimately linked to inequality, and the dynamics of unequal exchange are implicated in processes of poverty causation. Thus, the Nobel award winning economist and political philosopher, Amartya Sen, demonstrated that famine, long a symbol of the most extreme form of poverty, is less a matter of the absolute lack of food, but rather symptomatic of a crisis of entitlement created by social inequality. Poverty and inequality are both causes and correlates of development problems, with pockets of poverty in both the peripheral and core regions related to exploitive relations of inequality that characterize underdevelopment.

Rural Poverty in the Periphery

Poor rural places in the world's developing nations are often marginal agricultural or natural resource based economies with populations whose livelihoods depend on undependable subsistence activities. Lacking either a favorable resource base or a political climate conducive to successful investment and development of resources for the benefit of local populations, the rural poor are subject to the vagaries of natural and man made hardship. Factors implicated in rural poverty range from bad climate to poor infrastructure. Vulnerable groups such as female headed households, children, and race, ethnic, religious, or national minorities are especially likely to be disadvantaged. Studies of rural poverty in the Global South, either in individual countries or across regions, find a laundry list of causes and correlates of rural poverty including marginal land, pollution, lack of access to technology, transportation, and credit, remote location, corruption, natural disaster, disease, conflict and war, resource overuse and degradation, colonization and exploitation, and generally the conditions identified as underdevelopment – exploitation of peoples and extraction of resources by economic and political powers under terms favorable to the exploiter and unfavorable to indigenous populations.

The depredations of underdevelopment underscore the links between peripheral and core regions and their implications for rural poverty. While there are very few remaining isolated human groups whose subsistence economies operate almost entirely outside the global economy, for the most part, the poor rural peoples of the world are linked to the larger world system. Thus, unlike earlier culture of poverty and modernization theories that attributed poverty to the deficiencies of individuals or the backwardness of societies, most prevailing explanations of rural poverty in the periphery understand it as the legacy of historical and contemporary unequal relationships that create dependency and disadvantage in the global economy.

The development paradigm subsumes a number of theories that explain poverty in terms of global power differentials and relations of exploitation, whether historically or in the present. Variants include world systems theory, various brands of dependency theory, and most recently globalization. They differ in their theo retical origins from those rooted in Marxist and neo Marxist approaches to more state based orientations, focusing on political rather than economic actors. They vary in emphasis on formative experiences from the plunder of colonization to the exploitive investments of multinational corporations and the unequal terms of exchange forced on poor places by the powerful trans national organizations of the core, a process that has accelerated with increasing globalization. Regardless of specific formulation, however, they share the view cap tured in Andre Gunder Frank's phrase, "the development of underdevelopment," that poverty is constructed through the outcomes of exploitive power differentials between core and periphery.

Of particular importance in rural areas are the un equal terms of exchange that characterize agricultural production and natural resource extraction in the global economy. Farming for local consumption increasingly is displaced by income generating export based agriculture requiring more capital intensive forms of cultivation that favor large land holders and rich investors. Small land holders and landless labor are squeezed off their land and employment, creating conditions conducive to migration to urban areas in search of work. At the same time, even the largest agricultural producers in poor countries have difficulty competing in a global marketplace where rich nations protect and subsidize agricultural production to the competitive disadvantage of poor nations that lack resources and infrastructure. Similar lopsided conditions afflict natural resource extraction, where the necessary capital investments and infrastructure for value added production are unavailable, leading to distorted terms of trade. In this manner, countries rich in natural resources fail to benefit from their assets, and poor peoples within these nations remain impoverished or become even poorer.

Rural Poverty in the Core

The geography of poverty in the core affluent nations also over represents rural areas. For example, in the US in 2003, the poverty rate in nonmetropolitan counties was 14.2% compared to 12.1% in metro areas. Although the gap between urban and rural poverty rates has decreased over time, the pattern of higher rates in rural areas has persisted for almost 50 years since poverty rates were officially established and tracked. Many remote rural areas have been marked by high (more than 20% of the population) poverty rates for decades, leading to desig nation as sites of persistent rural poverty that bear many similarities to the rural poverty found in poor developing countries. In particular, rural poverty in the core often is the result of underdevelopment processes of exploitation, corruption, and neglect that characterize peripheral re gions of the Global South.

This becomes evident when examining the spatial distribution of persistent poverty in the US. Rural pov erty is not evenly distributed, but is disproportionately clustered in counties in the South and West, especially in the former slave owning states of the plantation South, Appalachian coal fields, Indian reservations, and colonias of the Southwest, areas noted for their histories of sub jugation of subordinate populations by powerful local and national elites. Historical accounts and case studies trace patterns of rural impoverishment, demonstrating practices of discrimination, exploitation, exclusion, and corruption that turn formerly wealthy places such as resource rich areas of Appalachia into the persistently impoverished places they subsequently have become. Theories of underdevelopment constructed to explain the situation of nations of the periphery have found equal use applied to peripheral regions of the core.

Of course, not all rural poverty in the US is predi cated on longstanding historical patterns of exploitation and resource extraction. More recent trends of economic restructuring, deindustrialization, and accelerated appli cations of capital intensive technologies in all forms of production from agriculture and mining to manu facturing have had an especially deleterious effect on the less diversified economies of rural areas. The loss of jobs and livelihoods, whether in farming, mining, or manu facturing, in areas that depend on one or two industries creates new hardship and exacerbates existing economic vulnerabilities. The specific form and pattern of eco nomic loss varies by region and type of economy. The virtually complete dominance of corporate agriculture has made the family farm obsolete. Deindustrialization in the upper Midwest rustbelt communities and Southern mill towns results in the loss of rural manufacturing plants to offshore locations with cheaper labor. Depletion of natural resources and increased environmental regu lation in timber and fisheries devastate whole com munities dependent on these industries. The growth of low level service sector jobs such as in retail sails, care work, and tourism does not make up for the loss of higher wage industries in these areas.

These trends are widespread across the industrial world, and the US is not the only example of a core nation with severe problems of deep and persistent rural poverty. Studies of rural poverty in Britain find similar patterns, and it can be speculated that other EU nations follow suit. However, it is difficult to gauge the extent of rural poverty in many core nations, since they do not always publish poverty data by rural–urban location, but rather by region or by occupation, making it difficult to determine uniquely rural poverty rates or to distinguish between agricultural and other rural economies. Furthermore, at least in many European countries, rural and agricultural subsidy and protection, as well as more extensive welfare states provide universal access to social welfare provision that buffers rural residents from some of the harshest aspects of poverty. This corresponds to a greater concern with poverty as a source and form of social exclusion than found in the US and thus, the prevalence of studies that assess forms and degrees of isolation and marginalization rather than absolute deprivation.

Rural Poverty across National and International Divides

Rural poverty in the core and periphery share many similarities. Poverty results 'from' lack of jobs and viable livelihoods, at least in part the result of accelerated globalization processes that put more isolated rural areas and their residents at a disadvantage in the global division of labor. Rural poverty results 'in' deprivation and isolation, with rural residents more likely to lack access to resources, including adequate income, food, water, healthcare, and education. Women, children, and minorities are especially vulnerable to processes both internal to specific societies and to broad global patterns of patriarchy and discrimination that create disadvantage for women and their children, as well as other vulnerable populations.

That said, there also are important differences. Deprivation in core nations may not always equal the depths of poverty in the poorest countries. However inadequate, unlike many developing nations, most core nations have some sort of safety net that provides assistance to their poorest residents, even if access is more limited in rural places. Finally, rural regions in core nations, however poor, exploited, and underdeveloped, also are part of larger entities that may be implicated in the exploitation of the rural poor in the periphery. Although powerless themselves, the rural poor of the core collectively are associated with the policies of their governments.

More notably, there are important links between the core and periphery, urban and rural regions beyond the fact that they are part of a global division of labor that increasingly draws all persons and places into its orbit. As noted earlier, economic migration and return migration provide steady streams between regions and across boundaries. The earnings of migrants to urban centers and core countries provide remittances that support those left behind. The policies adopted by bilateral and multilateral organizations whether free trade agreements such as North American Free Trade Agreement (NAFTA) and General Agreement on Tariffs and Trade (GATT) or structural adjustment policies dictated by the International Monetary Fund (IMF) and World Bank rearrange economic relations in ways that affect the poor of both rich and impoverished regions. The rural poor, even in the most isolated and remote places, bear the brunt of these policies.

See also: Counterurbanization; Historical Geographies, Rural.

Further Reading

Billings, D. and Blee, K. (2000). *The Road to Poverty: The Making of Wealth and Hardship in Appalachia*. New York: Cambridge University Press.
Escobar, A. (1995). *Encountering Development: The Making and Unmaking of the Third World*. Princeton, NJ: Princeton University Press.
Frank, A. G. (1966). The development of underdevelopment. *Monthly Review* 18(4), 17 31.
IFAD (International Fund for Agriculture Development) (2001). *IFAD Rural Poverty Report 2001*: *The Challenge of Ending Rural Poverty*. http://www.ifad.org/poverty/index.htm (accessed in June 2008).
Kerbo, H. (2006). *World Poverty: Global Inequality and the Modern World System*. Boston, MA: McGraw Hill.
McMichael, P. (2004). *Development and Social Change* (3rd edn.). Thousand Oaks, CA: Pine Forge.
Milbourne, P. (2004). *Rural Poverty: Marginalization and Exclusion in Britain and the United States*. London: Routledge.
Partridge, M. and Rickman, D. (2006). *The Geography of American Poverty: Is There a Need for Place Based Policies?* Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.
Sen, A. (1999). *Development as Freedom*. New York: Anchor Books.
Tickamyer, A. R. (2006). Rural poverty. In Cloke, P., Marsden, T. & Mooney, P. (eds.) *Handbook of Rural Studies*, pp 11 427. London: Sage.
US Census Bureau Housing and Household Economic Statistics Division (2006). *Poverty*. http://www.census.gov/hhes/www/poverty/poverty.html (accessed in June 2008).
US Department of Agriculture, Economic Research Service (2006). *Briefing Rooms: Rural Income, Poverty, and Welfare*. http://www.ers.usda.gov/Briefing/incomepovertywelfare/RuralPoverty (accessed in June 2008).

Relevant Websites

http://www.ifad.org
Enabling Poor Rural People to Overcome Poverty.

Pragmatism/Pragmatist Geographies

S. J. Smith, University of Durham, Durham, UK

Glossary

Pragmatism A philosophical orientation which pays attention to the practical consequences rather than the metaphysical origins of ideas.

Pragmatist Geographies Geographies adopting a pragmatist orientation which emphasizes how the search for 'truth' or knowledge should be grounded in the everyday practices of people/places under study.

Pragmatism: Origins and Character

Pragmatism was developed in North America, in the aftermath of the Civil War, by scholars at the universities of Chicago and Harvard. A kernel of the idea formed around a small group of philosophers – perhaps a single 'metaphysical club' – in the 1870s. Described by Louis Menand as 'a knot in the tapestry' of turn of the century life, pragmatism set itself the task of reconciling metaphysical disputes by focusing on the practical effects of one view or another being true. Pragmatism eschewed the abstractions of 'armchair' philosophy, rejected Enlightenment dualisms (between mind and body, idealism and materialism, subject and object, and so on), had limited faith in positivist scientific method, was suspicious of rationalism's appeal to reason, and looked for radically new – even 'hybrid' – ways to occupy a precarious and uncertain world whose future remains to be made.

The best known of the early, or classical, pragmatists include John Dewey, William James, George Herbert Mead, and Charles Sanders Peirce. Other key figures include Oliver Wendell Homes (who was a lawyer) and Charles Beard (a historian), as well as logician and ethicist, Clarence Irving Lewis. Philanthropist and activist, Jane Addams – previously overlooked – is increasingly recognized as a key participant in this movement.

The earliest influential accounts of pragmatism were published in 1878 (in essays on 'How to make our ideas clear' by Charles Peirce, and on 'Spencer's definition of mind as correspondence' by William James). However, the label 'pragmatism' did not receive public recognition until 1898 when James coined it during a lecture on 'Philosophical conceptions and practical results' at Berkeley. It became an immensely popular idea, and it is not surprising to find that pragmatisms and pragmatists come in diverse styles and traditions. Some authors draw a distinction, for example, between James and Dewey, who may be thought of as 'social nominalists' (regarding society as the sum of the individuals comprising it), and Mead and Peirce, with their leaning toward 'social realism' (believing that society amounts to more than the properties of its constituent parts). Others distinguish James's focus on individuals and their emotions from Mead's more fully social view of the self, and also from Dewey's wider concerns with sociality and the public realm. Several commentators draw attention to Peirce's distinctively logical, highly systematized style of pragmatism which – for better or worse (and scholars are divided on this) – stands well apart from the rest.

There are, nevertheless, some common themes, which – as befits a philosophy whose signature trait is the social character of knowledge – merit attention. There are many ways of labeling these core ideas, but most lists make reference to pluralism, instrumentalism, fallibilism, and experimentalism, all of which are rooted in a distinctive conception of society, and contain a firm ethical steer.

Pragmatism is, first and foremost characterized by a deep seated 'pluralism': it is in every way an anti foundationalist, anti essentialist, 'more than representational', enterprise whose core metaphysical claim is that the future is – to a significant degree – open: plastic, negotiable, alterable. In this radically open view of the world, 'reality' is neither mirrored in nor mined by ideas, but is manufactured through their application. In this action orientated, and therefore performative, space, theories are always provisional and 'truth' is an act, not a fact: it is something that happens to an idea. Ideas are nothing in themselves: they are (or are not) 'made true' by events. Pragmatism then is very much a philosophy of practice.

To that end, a further definitive characteristic of pragmatism is its 'instrumentalism'. Ideas are not truths waiting to be discovered but social experiments whose veracity is judged by their effects. Ideas (which are social rather than individual creations) are instruments for achieving particular purposes: their function is to help societies adapt to a changing but unpredictable world. In a *New York Times* interview published in 1907, James argued that pragmatism's primary interest is in its (instrumental) doctrine of truth: an idea is 'true' if it 'works' and does not clash with logic or experience. It follows (for James) that the role of the mind is to be creative not reflective; its work is not to apprehend but to transform: 'the *use* of most of our thinking is to help us to *change* the world'. And this is not so much 'a fact' as an ethical impulse, as might be anticipated from a philosophy invented precisely to resolve the dilemma of determinism.

This imperative to act rests on experimentation, not least because a corollary of pragmatism's pluralism is its 'fallibilism'. Pragmatism is deeply aware of the fragility of knowledge, recognizing that all matters of 'fact' are liable to modification in light of experience. Such modification is not done by individuals, but by a critical community of inquiry. To inform this community, pragmatism espouses the virtues of 'experimentalism'. There is (largely thanks to Dewey) a Darwinian impulse driving pragmatism: a conviction that change is inevitable but progress is not. So pragmatism, a philosophy of action, leans toward experimentation as a way of transforming unknown – but not unknowable – environments so as better to meet human needs. Experience here is not about passively registering the pre given; it is an experimental resource – a way of learning through doing; a way of continually testing, evaluating, and modifying the meanings and materials that make up the world. Experimentalism is therefore both empirical and radical. William James in fact coined the term 'radical empiricism' (a 'philo sophic attitude' which he distinguished from, and later preferred to, pragmatism) to describe this method of settling metaphysical disputes by appeal to experience. Such disputes are only worth having if there are practical consequences of one notion rather than another being true. In assigning action priority over contem plation, pragmatists engage with a world shot through with contingency in order to reduce or manage its uncertainties.

Finally – a point implicit in all the above – Pragmatism occupies 'a social and an ethical world': a space created by a mix of meanings, materials, and moralities which is characterized more by process than by form, and which exists in and through communication. Pragmatism indeed signals the sheer diversity of sites and situations, of spaces and places, in and through which knowledges are enac ted, tested, revised, and practiced. The pragmatic mes sage in a world where change is inevitable but progress is not is that human agency has the potential – and the ethical obligation – to intervene. But how and to what ends? Dewey's work is especially apposite here. He was deeply concerned with the impossibility of achieving consensus as to either desirable outcomes or the means of achieving them. He recognized that there is a fragmented geography of pragmatic knowledge communities as befits a world whose hallmark is diversity: even at a specific site, or in the context of a particular setting, there could never be a single appropriate prescription for change, much less an agreed mandate to implement it. Dewey's pragmatic solution – a form of deliberative democracy – was the ongoing process of communication, negotiation, and experimentation required continually to test, revise, and come to a compromise over appropriate interventions (including their political forms) in the light of newly emerging situations.

Neopragmatism

Although highly popular in the 1920s and 1930s, prag matism was eclipsed in the 1940s by a turn to analytical philosophy with its faith in logical positivism and its foundational theory of truth. However, for social and political theorists, pragmatism left traces in the work of Émile Durkheim, Friedrich Neitzsche, Max Weber, Ludwig Wittgenstein; it was in parts embraced by Jürgen Habermas; and it proved, in many ways, central to the inspiration of Gilles Deleuze. Pragmatism was more self consciously 'reinvented' in the 1980s following the pub lication, in 1979, of Richard Rorty's *Philosophy and the Mirror of Nature*. This was an attempt by Rorty to lead analytical philosophy out of the corner it had – by then – painted itself into. This work might not have had such impact had he not followed it up 3 years later with a collection of essays on *The Consequences of Pragmatism*. Its import was also assured by the fact that established figures such as Richard Bernstein had already primed the field, and because others – like Hilary Putnam and Richard Shusterman – were ready to enter the debate. Neopragmatism also appealed to feminist thinkers and civil rights activists such as Cornel West and Nancy Fraser. Today, this movement continues to gather momentum, while remaining as diverse and contentious as it was a century ago.

One key achievement of the neopragmatists is to turn attention to the role of language not as a way of repre senting the world, but as a mode of communication or interaction; a way of deliberating problems and agreeing solutions collectively. Language, or more particularly, 'conversation', has thus been placed at the heart of the transactional, interactional, and wholly social view of knowledge and experience that the classical pragmatists espoused. Another avenue opened up by this new gen eration of thinkers concerns the physicality of esthetic and ethical experience: the art of living is about practice, as well as talk, and this turns attention to the physical, emotional, and instrumental workings of the personal and the social body. These ideas were all – sometimes in a different style using different linguistic conventions – entertained by the first wave of pragmatist thinking, and it is not surprising that the pragmatic revival has been very much a celebration of philosophical roots. Deweyan ideas have been particularly widely embraced. They were, for example, Rorty's main inspiration, albeit in controversial ways. Today, however, a clutch of social science disciplines – including human geography – are engaging with, and building upon, a much broader pragmatist legacy.

Pragmatism and Geography

Pragmatism is a wide ranging philosophy which has left its mark on the law, in theology, across the arts, and in education, as well as in social and economic research. Any role it might have had in the shaping of geographical traditions has still to be established, though the early wave of pragmatism almost certainly had a bearing on the geomorphological legacy of G. K. Gilbert and W. M. Davis, as well as on the environmental geography of Gilbert F. White. Although Robert Park and perhaps Carl Sauer were also exposed to, and possibly influenced by, this body of thought, there is no clear evidence that it filtered directly through to academic human geography at the time: certainly there is little to suggest that the University of Chicago's geography department, founded in 1903, took on the pragmatists' mantle. In fact, notwithstanding such seeds as were sown in the early decades of the twentieth century, those more recent geographers most engaged with philosophical and methodological debate have hitherto prioritized Con tinental (European) philosophy and have hence had rather little to say about the role and relevance of (North American) pragmatist ideas.

That said, there are exceptions; and increasingly so. They can be hard to pin down, not least because those geographers whose work is inspired by, or infused with, a pragmatist ethos tend to be at pains to emphasize that theirs is an intellectually modest project. The corollary of regarding 'realities' as inexhaustible is a recognition that there is unlikely to be consensus on any aspect of how the world works. So geographers engage with pragmatism as an attitude to inquiry: an approach which does not *a priori* define the objects or subjects of study, nor dictate a particular set of methods, knowledges, or practices; an approach, grounded in experience, to the problem of living. So pragmatism's legacy is, above all, and as Dewey recognized, 'a metaphysical ground map'. Exploring this territory, human geography's project has been enlarged in several key ways.

Perhaps the earliest account of pragmatism as an explicit possible philosophical inspiration for geography is contained in an essay published in 1981 by John W. Frazier. He prized in particular the pragmatic maxim that 'reality' is a mix of knowledge and error – a mix which can be altered to produce a range of different outcomes. Pragmatic methods of experience and experimentation thus constitute practical problem solving devices which could – in this case by 'invigorating applied geography' – help promote human welfare.

Frazier's essay is the forerunner to what has become the most sustained early impact of pragmatism on the work of human geographers, which is 'methodological' and at the same time 'ethical'. This practice centered philosophy recognizes the 'actancy' of both people and things, and emphasizes the importance of experience in mediating between intellect and common sense. It may seem ironic that as a way of doing research it not only inspired the qualitative methods that put the 'human' at the center of (humanistic) geography in the early 1980s, but also that – being so profoundly nonrepresentational – it spoke to the more material (and quite deliberately 'more than human') turn toward the current millennium. The fact that pragmatism inspired both these shifts tes tifies not to its malleability as a doctrine, but rather to the relational middle ground it occupies between antihuman determinisms and antimaterial idealisms.

On the one hand, therefore, the interpretative humanistic tradition of urban sociology and anthropology known as symbolic interactionism was self consciously inspired by pragmatism. Nurtured for a time at the 'Chicago school', this interactionist tradition – with its emphasis on encounter and engagement – became a methodological hallmark of urban social geography for at least a decade. On the other hand, and more recently, parallels have been drawn between some key pragmatic and neopragmatic ideas and the nonrepresentational styles of working (with complex networks of people and things, with the artifacts and organisms of the human and nonhuman world) that have become so central to human geography today. In particular, methods of knowing the world through practice, means of creating the future through performance, all illuminate the relational, transactional, emergent qualities of human beings. It is not surprising that some key figures in the formation of nonrepresentational geographies and material sociologies – Michel Callon, Bruno Latour, Nigel Thrift – have all found it helpful to refer to the pragmatic tradition in their work.

Next, there is the possibility of linking the legacy of classical pragmatism with the energies of its later exponents in order to 'reinvigorate some key sub disciplinary areas of human geography'. Take, for ex ample, economic geography. Here, it might be imagined that Peirce's logical pragmatism would appeal. However, while Peircean pragmatism may have been highly influential in the development of American quarternary geology and geomorphology, it is the later pragmatists (generally those most indebted to Dewey) who have most inspired economic geography. Trevor Barnes, for example, has used Rorty's ideas to unpack the gravity model, while Peter Sunley has looked to Bernstein and Putnam for progress in appreciating the importance of context for economic analysis. Likewise in the flourishing field of urban geography, Gary Bridge, for example, has traced the ideas of Dewey into and beyond the work of Jürgen Habermas, in order to rethink the nature of the city and the public realm as a transactional, performative, esthetic, and experimental affair.

Additionally, there is a place for pragmatism in 'the recovery of emotional geographies'; recognizing the sensory body, the social self, the affective content of everyday life, and, to an extent, the psychoanalytical qualities of human beings. Mead's ideas, for example, have been used in geography to link understandings of the psyche with the practicalities of the body, and to grapple with the relationality and sociality of the self. Likewise, James, with his radical take on social psych ology, wrote extensively on the emotions, which he regarded as both a condition of knowing (building on the pragmatist idea that only that which can be sensed can be known) and a practical resource: a means of action that would otherwise be inhibited. Pragmatism was, as much as anything else, a reaction to philosophy's longstanding suspicion of the emotions; thus, it is unsurprising that these ideas have helped reposition emotional affairs in the heart of human geography.

Pragmatism further speaks to a new 'preoccupation with uncertainty and risk'. The possibilities here were laid out by James Westcoat over 15 years ago in a thought provoking discussion of common themes in the work of Dewey and Gilbert F. White (who was pre occupied with hazards and resources). This Deweyan take on uncertainty has been recovered and extended in more recent years in a series of papers by health geog rapher, Macolm Cutchin, which contain a prescription for handling context and contingency in a radically pragmatic world. Geographers have also looked to pragmatism to flesh out 'key disciplinary concepts' like space, place, and situation. In the early days of this, pragmatism – with its focus on engagements, attachment, relationalism, and materiality – tended to be used as a counter to prevailing impressionistic, literary, reflective, and representational accounts of place. The most recent flurry of published papers, however, are more forward looking, noting the importance of anti essentialist thinking, as well as the centrality of critique, but stressing too the value of using geographical sites instrumentally, experimentally, ethically, and indeed normatively – as a tactic for creating futures that are better than the past.

Pragmatism's Critics

Pragmatism is not without its critics, the earliest of which came from within its own ranks when Peirce distanced himself from his peers by coining the term 'pragmaticism' to distinguish 'his' system of logical rules from 'their' metaphysical theory, as a means of clarifying ideas. More generally, pragmatism has been criticized as at best overly incrementalist, and at worst an 'anything goes' phil osophy which is anti intellectual, lacks radical edge, and is too easily hijacked by a liberal – indeed commercial and imperial – worldview. Perhaps the most enduring critique, however, concerns pragmatism's lack of atten tion to the systematically uneven distribution of power or to the structures of inequality that such unevenness sustains.

It is not entirely correct to say that power is absent from the world of pragmatism. On the contrary, pragmatism has a lot to say about the relationship be tween knowledge and power, and is committed to using 'knowledge as power' to recognize and resolve conflict. But it is true that, beyond this, power is undertheorized – a condition which is hardly surprising for a philosophy that has always resisted formalism, rarely dwelling long enough on structures to recognize the enduring in equalities they express. Recently, though, John Allen has sketched the bones of what might become a pragmatist account of power: this would be an expedient affair that is antifoundational in character and energized through situated practices. The intriguing suggestion here is that, by adding an account of the difference that space (and spatiality) make to how power is practiced, there is also a chance for geography not just to be inspired by prag matist ideas, but actively to give something back; to shape the future of the single major philosophy most open to such change.

See also: Chicago School; Non-Representational Theory/Non-Representational Geographies; Performativity; Positivism/Positivist Geography; Psychoanalytic Theory/Psychoanalytic Geographies; Symbolic Interactionism.

Further Reading

Allen, J. (in press). Pragmatism and power, or the power to make a difference in a radically contingent world. *Geoforum.*

Baker, V. (1996). The pragmatic roots of American quaternary geology and geomorphology. *Geomorphology* 16, 197 215.

Barnes, T. (1991). Conversations and metaphors in economic geography. Richard Rorty and the gravity model. *Geografisker Annaler B* 73, 11 20.

Barnes, T. (in press). American pragmatism: Towards a geographical introduction. *Geoforum.*

Bridge, G. (2005). *Reason in the City of Difference: Pragmatism, Communicative Action and Contemporary Urbanism*. London: Routledge.

Cutchin, M. P. (2004). A Deweyan case for the study of uncertainty in health geography. *Health and Place* 10, 203 213.

Cutchin, M. P. (in press). John Dewey's metaphysical ground map and its implications for geographical inquiry. *Geoforum.*

Frazier, J. W. (1981). Pragmatism: Geography and the real world. In Harvey, M. E. & Holly, B. P. (eds.) *Themes in Geographic Thought*, pp 61 72. London: Croom Helm.

Jackson, P. and Smith, S. J. (1984). *Exploring Social Geography.* ch. 4. London: George Allen and Unwin.

Jones, O. (in press). Stepping from the wreckage: Geography, pragmatism and anti representational theory. *Geoforum.*

Menand, L. (2001). *The Metaphysical Club*. New York: Harper Collins.
Smith, S. J. (1984). Practising humanistic geography. *Annals, Association of American Geographers* 74, 353 374.
Smith, S. J. (1988). Constructing local knowledge: The analysis of self in every day life. In Eyles, J. & Smith, D. (eds.) *Qualitative Methods in Geography*. Cambridge: Polity Press.
Sunley, P. (1996). Context in economic geography: The relevance of pragmatism. *Progress in Human Geography* 20, 338 355.
Westcoat, J. L. (1992). Common themes in the work of Gilbert White and John Dewey: A pragmatic appraisal. *Annals, Association of American Geographers* 82, 587 607.
Wood, N. and Smith, S. J. (eds.) (in press). Pragmatism and geography. *Geoforum* (theme issue).

Pred, A.

B. Warf, University of Kansas, Lawrence, KS, USA

A child of the Bronx, Allan Pred (1936–2007; **Figure 1**) was one of geography's premier leading theoreticians and practitioners of urban geography and social theory. He entered Antioch College at the age of 16, graduated in 1957, and subsequently completed an MA at Pennsylvania State University and a PhD at the University of Chicago in 1962. For 44 years, he was a professor of geography at the University of California, Berkeley, where he arrived as an assistant professor at the age of 25. He was long married to his Swedish wife, Hjőrdis, with whom he had two children, and divided his time between Berkeley and Sweden for much of his life. In 2007 he died of lung cancer.

Pred's academic output is legendary, including 22 books and more than 70 journal articles and book chapters. His awards, too numerous to detail, included the Anders Retzius medal from the Swedish Society for Geography and Anthropology, honors from the Association of American Geographers and the Polish Academy of Sciences, an honorary doctorate from Uppsala, and multiple visiting appointments. Pred's career took him through diverse conceptual approaches in a series of shifts that both reflected and contributed to geography's multiple intellectual transformations. He drew for inspiration upon authors and genres as diverse as urban growth theory, innovation diffusion, Walter Benjamin, feminism, structuration theory, linguistics, and poststructuralism to inform his work.

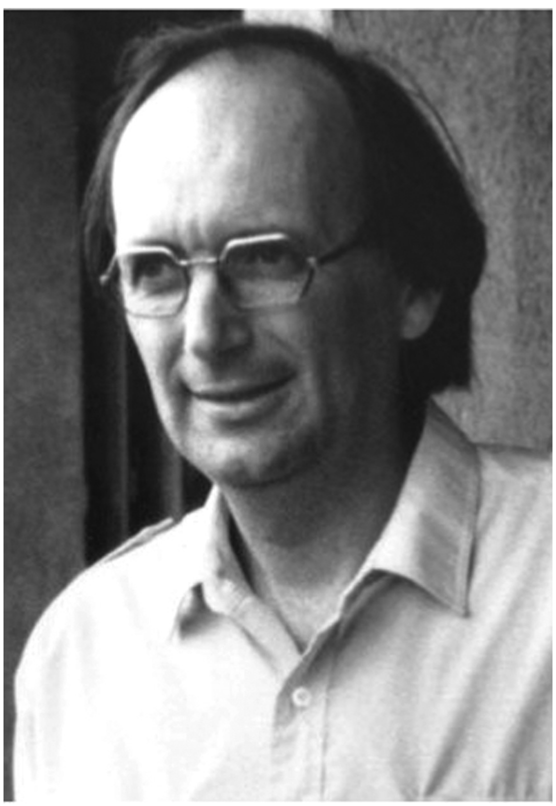

Figure 1 Allan Pred. From the 2008 Annals of the Association of American Geographers, and article entitled "Allan Richard Pred, 1936 2007: Reflections on a Life" by Michael Watts and Richard Walker, v. 98 (2), article on pp. 487 493, Figure on p. 487.

In the 1960s, his early work centered on the historical geography of American manufacturing and its linkages to the country's urban growth patterns. From the start, his perspective was resolutely historical, emphasizing the contextualized nature of urban growth rather than abstract processes independent of historical time. Increasingly, the early focus on manufacturing was supplemented by a growing interest in the role of information and its relations to the differential growth of cities. Much of this phenomenon was theorized in terms of circular and cumulative feedback loops. In particular, he noted the powerful role of the telegraph in the long distance transmission of information in the nineteenth century, a technology that greatly reduced uncertainty and accelerated the penetration of Midwestern regions by East coast capital. The crowning achievement of this phase of Pred's career was the highly influential volume *City Systems in Advanced Economies.*

In the 1970s, Pred turned toward the time geography initiated by Torsten Hagerstrand, and began to apply this framework to the US context. Places, he maintained, are always and everywhere times, that is, always coming into being, and geography can never be fruitfully portrayed as a static landscape frozen at one temporal moment. This move both led him on the one hand to delve more deeply into issues of individual consciousness, including the notion of sense of place advocated by humanistic geographers, and to integrate individual life trajectories with broader questions of social reproduction on the other hand.

Pred became a fierce advocate of Anthony Giddens' notion of structuration, and worked assiduously to integrate it with geographers' notions of place and locale. In this reading, the socialization of the individual and the reproduction of society and place are two sides of the same coin; that is, the macrostructures of social relations are interlaced with the microstructures of everyday life. People reproduce the world, largely unintentionally, in their everyday lives, and in turn, the world reproduces them through socialization. In forming their biographies everyday, people reproduce and transform their social worlds primarily without meaning to do so; individuals are both produced by, and producers of, history and geography. Because people are endowed with the capacity to resist hegemonic power relations, the production of places, in short, is always contingent. One summary of this perspective was offered as a presidential address to the Regional Science Association. Empirically, Pred offered applications to Boston, examining the lives of merchants in the nineteenth century as the city was catapulted through waves of mercantile capitalism, industrialization, and a rapidly changing world economy.

One facet of this line of thought led Pred to a sustained concern for the role of language in the making of place and people. As a set of representations of the world that simultaneously shape consciousness and define the meanings given to place, language came to assume a central role in his analysis of social reproduction and local politics. Pred exhibited a sensitivity to the gendered nature of language that few geographers appreciated at the time.

Historical context, the world economy, time geography, structuration, and language were all parts of his ambitious analysis of nineteenth century Stockholm. Having spent considerable time in Sweden, with access to its rich census data, Pred meticulously traced the contingent reconstruction of the city's streets and neighborhoods as an emerging bourgeois identity was firmly imposed on its landscapes, both annihilating older, preindustrial meanings and simultaneously defeating alternative, working class ones. This project lay at the core of a broader agenda to uncover the historical roots and implications of modernity in general.

Pred's fascination with Sweden included a growing critical strain. Even in the ostensibly most tolerant of societies, he argued, racism was widespread. Pred thus became a pioneer in geography's belated examination of ethnicity and race, which, under the impact of post structural critiques, revealed race to be a set of economic, political, cultural, and discursive relations that shape individual and collective life chances and identity. By contextualizing race as an ideology – that is, as a socially constructed relation, not a pre given biologically determined 'given' – Pred illustrated the complex ways in which race is constructed, experienced, and negotiated in everyday existence; how it intersects with class and gender; and the manner in which racial relations of domination and subordination play out over space and time. Like gender, racial and ethnic identity is fundamentally a power relation that typically reflects, reinforces, and naturalizes existing lines of inequality. This body of work was extended to include a critical examination of the 'war on terror' and its profoundly political repercussions.

Pred's long and productive career have established him as a scholar who is both sensitive to the changing intellectual climates of the discipline and as one of its most insightful practitioners. His work exhibits a sustained concern for issues such as power and identity, everyday life and social relations, uneven development, and the interpenetration of the global and local.

See also: Language and Research; Race; Structuration Theory; Transnationalism and Technological Transfer.

Further Reading

Pred, A. (1973). *Urban Growth and the Circulation of Information: The United States System of Cities, 1790 1840*. Cambridge: Harvard University Press.

Pred, A. (1976). Interurban transmission of growth in advanced economies Empirical findings versus regional planning assumptions. *Regional Studies* 10, 151 171.

Pred, A. (1977). Choreography of existence comments on Hagerstrand's time geography and its usefulness. *Economic Geography* 53, 207 221.

Pred, A. (1977). *City Systems in Advanced Economies*. New York: Wiley.

Pred, A. (1978). Impact of technological and institutional innovations on life content some time geographic observations. *Geographical Analysis* 10, 345 372.

Pred, A. (1980). *Urban Growth and City Systems in the United States, 1840 1860*. Cambridge: Harvard University Press.

Pred, A. (1981). Social reproduction and the time geography of everyday life. *Geografiska Annaler* 63, 5 22.

Pred, A. (1983). Structuration and place on the becoming of sense of place and structure of feeling. *Journal for the Theory of Social Behaviour* 13, 45 68.

Pred, A. (1984). Place as historically contingent process: Structuration and the time geography of becoming places. *Annals of the Association of American Geographers* 74, 279 297.

Pred, A. (1984). Structuration, biography formation and knowledge observations on port growth during the late mercantile period. *Environment and Planning D Society and Space* 2, 251 275.

Pred, A. (1985). Interpenetrating processes human agency and the becoming of regional spatial and social structures. *Papers of the Regional Science Association* 57, 7 17.

Pred, A. (1989). The locally spoken word and local struggles. *Environment and Planning D: Society and Space* 7, 211 233.

Pred, A. (1990). In other wor(l)ds fragmented and integrated observations on gendered languages, gendered spaces and local transformation. *Antipode* 22, 33 52.

Pred, A. (1990). *Lost Words and Lost Worlds: Modernity and the Language of Everyday Life in Late Nineteenth Century Stockholm*. Cambridge: Cambridge University Press.

Pred, A. (1990). *Making Histories, Constructing Human Geographies*. Boulder, CO: Westview Press.

Pred, A. (1995). *Recognizing European Modernities: A Montage of the Present*. London: Routledge.

Pred, A. (1996). Interfusions: Consumption, identity and the practices and power relations of everyday life. *Environment and Planning A* 28, 11 24.

Pred, A. (1997). Somebody else, somewhere else: Racisms, racialized spaces and the popular geographical imagination in Sweden. *Antipode* 29, 383 416.

Pred, A. (1998). Memory and the cultural reworking of crisis: Racisms and the current moment of danger in Sweden, or wanting it like

before. *Environment and Planning D: Society and Space* 16, 635 664.
Pred, A. (2000). *Even in Sweden: Racisms, Racialized Spaces and the Popular Geographical Imagination*. Berkeley, CA: University of California Press.
Pred, A. (2004). *The Past Is not Dead: Facts, Fictions and Enduring Racial Stereotypes*. Minneapolis, MN: University of Minnesota Press.
Pred, A. and Gregory, D. (eds.) (2007). *Spaces of Terror and Violence*. London: Routledge.
Pred, A. and Watts, M. (1992). *Reworking Modernity: Capitalisms and Symbolic Discontent*. New Brunswick, NJ: Rutgers University Press.
Watts, M. and Walker, R. (2008). Allan Richard Pred, 1936 2007: Reflections on a life. *Annals of the Association of American Geographers* 98, 487 493.

Pregnancy and Childbirth

E. Chacko, The George Washington University, Washington, DC, USA

Glossary

Abortion The natural or induced expulsion of the fetus from the womb.

Contraception A device, substance, or method that prevents pregnancy.

Ectopic Pregnancy A pregnancy in which the embryo develops outside the uterus, usually in a fallopian tube or sometimes, in the ovary, cervix, or abdominal area.

***In Vitro* Fertilization** A method of assisted reproduction in which a man's sperm and a woman's egg are fertilized in the laboratory and then transferred to the uterus of woman for implantation and development.

Life Cycle Approach An approach to human health that refers to the entire span of a person's life from birth to death, rather than a shorter time period of ill health.

Maternal Mortality Ratio The number of maternal deaths per 100 000 live births.

Obstetric Fistulae Tears between the vagina and either the bladder or the rectum, causing incontinence as urine and feces leak through the vaginal opening.

Obstetric Morbidity Illnesses and disabilities associated with pregnancy, delivery, and the postnatal period.

Safe Motherhood Initiative (SMI) A global initiative put in place in 1987 by the World Health Organization (WHO) with the aim of reducing the health risks of pregnancy and childbirth.

Stillbirth When a baby is born dead at or after 20 weeks of gestation.

Surrogate Pregnancy A pregnancy contracted between the woman carrying the baby and the persons who will raise it, with the intention that the pregnant woman will give up the baby after it is born.

Introduction

A woman's reproductive or fecund years, which begin with menarche and end with menopause (broadly considered the period between 15 and 45 years of age), are when she is capable of becoming pregnant and giving birth. In humans, the gestation period (the length of pregnancy) typically lasts 9 months. During this time, the baby grows in its mother's uterus or womb, receiving oxygen and nourishment and expelling wastes via the umbilical cord, which is attached to the placenta. The latter is an organ that develops during pregnancy to facilitate the exchanges of oxygen, food, and waste between mother and baby. An ectopic pregnancy occurs when the embryo develops outside the uterus, usually within a fallopian tube or more rarely, in the ovary, cervix, or abdominal area. The embryo cannot grow or survive outside the uterus. One in every 40–100 pregnancies is ectopic.

At the close of gestation, labor pains caused by uterine contractions usually signal the onset of the birth process. During labor, which is regarded as the first stage of natural childbirth, contractions strengthen and accelerate in frequency. The duration of labor varies considerably, but averages 10–12 h for women giving birth to their first child (primiparae) and 8 h for women who have already given birth. In the second stage of childbirth, known as delivery, the baby is pushed out through the birth canal (vagina). The third and final stage of childbirth occurs when the uterus expels the placenta as afterbirth. Childbirth, which is also called birth or parturition, is therefore the natural conclusion of a human pregnancy during which the fetus is at least 20 weeks old and occurs when the newborn leaves the mother's womb.

Traditionally, births took place at home with midwives and female relatives in attendance. This practice continued until the twentieth century. Thereafter, the hospital became the site for giving birth for the majority of women in industrialized countries and this is increasingly the case in developing countries. After a hospital birth, the mother and child are often separated and cared for independently. In developed countries, to counter this trend of the 'medicalization of childbirth' and treating pregnancy and childbirth as illnesses, a small number of women are opting for homebirths supervised by trained midwives. Such women wish to fully experience giving birth in the familiar and safe settings of their homes without the use of drugs such as epidurals. An additional incentive is that homebirths tend to be less expensive and that delivery by cesarian section is rare in births taking place in nonhospital settings.

While bringing a child into the world is normally a happy occasion, childbirth can be a stressful event for both mother and infant. Some women experience postpartum (postnatal) depression, a condition that is considered biomedical in its origins due to hormonal imbalances. However, depression is compounded by distress stemming from inadequate emotional, economic, and social support. Contrary to the assumption that sociocultural contexts associated with childbirth in non Western societies protect mothers from depression, factors unique to the culture including gender preference and low involvement of fathers in childcare can be a cause of stress to mothers and exacerbate postnatal depression.

Abortions

Not all pregnancies result in a live birth. Fetal death can occur either before 20 weeks of gestation (spontaneous abortion or miscarriage), or at or after 20 weeks of ges tation (stillbirth). Spontaneous abortion occurs in about 10–15% of all pregnancies. Induced abortion, on the other hand, is caused deliberately so that the pregnancy is not completed. Because of risks to the woman's health, most induced abortions are performed during the first 3 months of pregnancy, except in special cases. When performed by medical doctors in a sanitary environment, abortions are safe and relatively simple procedures. However, those performed by untrained persons in nonsterile environments can be extremely dangerous, resulting in serious problems including infertility and even the death of the woman from infection and bleeding.

Induced abortions that are unsafe create a major health hazard for women in developing countries; of an estimated 19 million such abortions practiced worldwide in 2000, nearly all (18.5 million) occurred in developing countries. Approximately 10.5 million of the unsafe abortions were performed in Asia, 4.2 million in Africa, and 3.8 million in Latin America and the Caribbean. Practically all of the estimated 68 000 women who die from the complications of unsafe abortion each year live in developing countries of Asia (34 000), Africa (30 000), and Latin America and the Caribbean (4000).

Governments vary widely in their attitudes to abor tion, which is reflected in variations in ease of access to safe abortions performed by medically trained persons in different countries. The US Supreme Court's landmark decision in Roe versus Wade in 1973 allowed American women to seek an abortion for any reason through the end of the 13th week of pregnancy. Abortion as a form of population or birth control was practiced in many countries of Eastern Europe and Japan and continues to be used as an option in developing countries like China and India which wish to stabilize their populations.

New methods, such as testing the amniotic fluid to detect the sex of the fetus earlier in the pregnancy, have made it possible to abort the fetus if it is not of the desired sex. Such sex selection is almost always in favor of male children. In one study of 8000 abortions in India, 7999 were female fetuses. In China, where the one child policy was strictly enforced, preference for males resulted in an imbalanced sex ratio of 120 males per 100 females in 2000. Today, tests to determine the sex of the fetus are illegal in both India and China.

Advocates of legalized abortion cite the dangers to women of unsafe abortion, and the problems of over population and unwanted children. Those opposed to abortion feel that it is akin to taking a human life. The moral versus legal distinction between late abortion and infanticide continues to be a matter of heated legal de bate in countries such as the United States of America.

Infertility and Its Solutions

If unwanted pregnancies are a problem, so too is the inability to have children. About 10% of couples of reproductive age are infertile. Women with partners who are infertile or subfertile may turn to artificial insemin ation by donor. *In vitro* fertilization is a method of assisted reproduction in which a man's sperm and a woman's egg are fertilized in the laboratory and then transferred to the uterus of woman for implantation and development. Usually, two to six embryos are placed in the uterus at any one time. The procedure was developed to bypass blocked or missing fallopian tubes of infertile women.

If the women's reproductive organs are not func tioning properly or if pregnancy may endanger her life, prospective parents may seek a surrogate mother. In surrogate pregnancy, or a contracted pregnancy, a woman carries a child for other persons with the intention of giving up that child when it is born. The surrogate mother is often also the baby's biological mother, although she may be implanted with another woman's fertilized egg. While most surrogate pregnancies occur in Western countries, they are on the rise in developing countries.

Pregnancy and childbirth have many meanings for women, couples, families, communities, and countries. The psychological dimensions of motherhood and bringing new life into the world are as important as the physical aspects of being pregnant and giving birth. The role of culture, economics, social expectations, and healthcare policy are also important factors in the well being of pregnant women and mothers.

Preventing Pregnancy

Any method of avoiding pregnancy is called 'birth con trol'. The term was coined in 1914–15 by Margaret Sanger, founder of the birth control movement in the United States and an international leader in the field. Reliable methods to prevent pregnancy were developed in the twentieth century as a consequence of scienti fic advancements and better knowledge regarding the functioning of the human body. Birth control methods can be either temporary or permanent. Today, modern contraception and sterilization techniques allow greater control over fertility, providing women with reproductive freedoms not possible earlier. However, even before the advent of modern medicine and contraception, fertility had begun to drop in industrialized countries, probably due to later marriage and 'natural' methods of birth control based on the woman's menstrual cycle, abstin ence, or withdrawal. At the turn of the nineteenth cen tury, American women bore an average of seven children. By 1900, white women in the United States had halved their fertility to 3.56 children.

There are both chemical as well as mechanical ways of preventing pregnancy, methods that prevent sperm and egg from coming together to form an embryo. Birth control devices and methods have increased in number and sophistication; the vast majority is female centered.

Most methods of birth control are temporary or reversible. The birth control pill was first used in the 1960s. 'The pill' as it is popularly known, contains hormones (progestin and estrogen) that interrupt the normal cycle of ovulation by creating an imitation pregnancy. Highly effective if used correctly, its side effects are generally mild: nausea, headaches, and weight gain. However, the pill is not advised for women over the age of 35 years. Diaphragms, caps, and shields are barriers made of thin latex or silicone that are inserted into the vagina to cover the cervix and thereby block the opening of the uterus. These devices are usually used with a spermicidal cream or jelly. A flexible contraceptive ring inserted in the vagina prevents pregnancy. The ring is usually left in place for 3 weeks and taken out for a week. An intrauterine device (IUD) is a small T shaped device usually made of plastic that is inserted in the uterus by trained medical personnel. An IUD prevents a fertilized egg from being implanted in the uterine walls, a necessary condition for the growth of the embryo. IUDs can be used to prevent pregnancy for 5–12 years and can be removed when the woman wishes to become pregnant. However, these devices can cause cramps, bleeding and even infection, and are not considered suitable for all women. Pregnancy can be avoided by inserting an implant containing the hormone progestin under the skin of a woman's upper arm to prevent ovulation. A plastic patch that releases progestin and estrogen can be placed on a woman's skin to protect against pregnancy for up to 1 month. An injection containing progestin prevents pregnancy for up to 12 weeks.

Condoms are contraceptives made of thin rubber and are traditionally used by males. Male condoms function as birth control devices by preventing sperm from es caping and coming into contact with an egg cell. Less popular is the female condom, a thin sheath of poly urethane that is inserted into the vagina, thereby lining it. Condoms are also used to prevent the transfer of sexually transmitted disease (STD) between partners. Emergency contraception in the form of pills containing high doses of progestin and estrogen can be used after unprotected sex to prevent pregnancy.

Unlike the methods listed above, surgical sterilization is a permanent method of birth control. The surgery can be performed on either a man or a woman. Vasectomy in males prevents the sperm from being ejaculated with the semen. In women, ligation of the fallopian tubes (tying the tubes) keeps eggs from reaching the uterus. This procedure too must be done in a medical facility. Sometimes, a woman becomes pregnant even after undergoing tubal ligation. Such women face a high risk (60%) of an ectopic pregnancy.

Fertility and Population

National governments are cognizant of the critical role of fertility in population stabilization and development. Countries are concerned about the social and economic ramifications of births within their boundaries, particu larly if they perceive fertility to be either too high or too low. Total fertility rate (roughly the number of children a woman has during her reproductive years), has dropped by over one third in both developing and developed countries since 1960. In developing regions, the highest fertility rates are found in Africa, with Uganda heading the list in 2006 with a total fertility rate of 6.9. Africa on the whole has a total fertility rate of 5.1, Latin America and the Caribbean 2.5, Western Asia 3.4, Southern and Central Asia 3.1, and Southeast Asia 2.5.

Many less developed countries that consider their birthrates too high have official government policies to reduce birth rates. On the other hand, most developed countries, especially those with fertility levels below the replacement level of 2.1 are concerned about the challenges they will face if their populations continue to shrink. Many of these countries including Japan, South Korea, and most European nations have adopted prona talist policies that encourage more births. In 2006, South Korea and Taiwan each had a total fertility rate of 1.1. The rates in some European countries were only mar ginally higher. Total fertility rate was 1.2 in Belarus, Ukraine, Bosnia Herzegovina, San Marino, and Slovenia; while Germany, the Czech Republic, Bulgaria, Hungary, Greece, Italy, Spain, Slovakia, Russia, and Moldova had a slightly higher rate of 1.3. With the exception of Iceland which is at replacement level fertility, all European countries have birthrates well below 2.1.

Experiences and Hazards of Pregnancy and Birth

As health professionals and societies continue to define pregnancy as an abnormal condition, pregnant women's bodies are often seen as sick, misshapen, ugly, fat, and 'out of place'. Geographers in particular have investigated the mutually constitutive relationship between women's bodies and their use and experience of space/place, providing insights into the nexus between social relationships, power, and the female body. They have found that as pregnancy progresses, women may curtail their use of public space and reduce or stop activities that could draw attention to their growing bodies. In comparison to men, women's identities are constituted more relationally and their bodily boundaries are often more fluid. Although women may consider their pregnancy intensely personal, societal norms that favor a view of the pregnant woman as primarily a container for the growing fetus that needs to be monitored and controlled often results in an infringement of women's bodily autonomy. Some common encroachments on the highly personal space of the body as it is subjected to the public gaze include touching the pregnant woman's belly and offering unsolicited advice on diet, exercise, and other practices aimed to enhance the health of the unborn child.

The geographical spaces in which birth takes place and the practices that occur there are also socially constructed, varied, changing, and sometimes, politically charged. The shift from the home to the hospital as the locale for giving birth in the early twentieth century was with the intention of making birth safer for mother and child. However, feminists point to this trend as another example of the control of patriarchal structures over women's lives. Giving birth at home in a familiar and comfortable place allows women greater control over the process and makes birth a part of the continuity of familial life. The home and the hospital continue to be contested sites of childbirth particularly in developed countries. Greater efforts are made to create more comfortable and home like birthing spaces in hospitals, while conversely, homebirths are encouraged only if the pregnancy is normal. Women with medical complications usually give birth in hospitals.

Hazards of Pregnancy

Around 80% of pregnancies are low risk. However, obstetric morbidity, defined as illnesses that are associated with pregnancy, delivery, and the postpartum period can occur in some cases, while the lingering effects of complications during these events can result in ailments of the reproductive tract and disabilities for the woman. Factors that may make a birth high risk in the developed world include prematurity, a mother who is at the lower or higher end of her reproductive years, high blood pressure, diabetes, and a previous cesarean section. In developing parts of the world in addition to these factors, women often face additional dangers due to poor nutrition exacerbated by cultural food taboos and poverty. Moreover, getting pregnant at an early age, continuous cycles of pregnancy with short birth intervals, infection, and hard physical work compound the problem, leaving women in a state of chronic maternal nutritional depletion. The fetus can be harmed by the pregnant woman's behaviors including the use of teratogenic drugs. Teratogens halt the pregnancy or produce a congenital malformation (a birth defect). Classes of teratogens include radiation, maternal infections, and chemicals and drugs such as alcohol, cocaine, tetracyclines, and thalidomide.

Pregnancy and delivery related illnesses cause suffering to millions of women. In the mid 1990s, the World Health Organization (WHO) estimated that on average there are 16 acute morbidities (illnesses) for every maternal death. However, individual studies point to higher figures reaching a possible 100 cases of morbidity per maternal death.

In 2000, the number of deaths worldwide attributed to childbirth and problems related to pregnancy were estimated to be 529 000. The bulk of maternal mortality occurred in Africa and Asia, each accounting for over 250 000 maternal deaths. Latin America and the Caribbean accounted for 4% of the deaths, while less than 1% of global maternal deaths took place in the more developed regions. The vast difference between pregnancy/birth related risk in the developing and developed worlds is reflected in the lifetime risk of dying from maternal causes (a measure that combines maternal mortality ratios (MMRs) with fertility rates), which is 1 in 61 in developing regions and only 1 in 2800 in developed regions. Worldwide, the MMR or the number of maternal deaths per 100 000 live births is 400. The MMR in Africa is more than double at 830, followed by Asia with 330, Oceania with 240, Latin America and the Caribbean with 190, and the developed countries with 20 deaths per 100 000 live births.

Overall, pregnancy in South Asia, particularly in the rural areas, is a hazardous experience. Within the region, India has the highest estimated number of maternal deaths (136 000), followed by Pakistan (26 000), Afghanistan (20 000), and Bangladesh (16 000). A number of African countries also rank high in total estimated maternal deaths: Nigeria (37 000), the Democratic Republic of Congo and Ethiopia (24 000 each), the United Republic of Tanzania (21 000), Angola and Kenya (11 000 each), and Uganda (10 000). These countries together with China (11 000) and Indonesia (10 000) account for 67% of all maternal deaths.

Factors Affecting Maternal Morbidity and Mortality

The risk associated with childbearing is higher among women in their teens or younger and those over the age of 35. Teenage mothers face a high risk of serious preg nancy related complications and the likelihood of preg nant teenagers dying from obstetric problems is at least 20% higher than for women in their twenties. The risks are amplified for pregnant women less than 15 years of age. Obstetric problems are much greater for women who are in the later years of their reproductive period: high risk pregnancies usually occur in girls less than 18 years of age, women who are older than 35 years, among women with more than four births, and those who give birth less than 2 years apart.

Childbearing among women less than 20 years of age is of particular concern in both the developing and de veloped worlds. Also sometimes referred to as teenage pregnancy, this phenomenon has both physical and social dimensions. From a biological standpoint, a teenager is not ready for pregnancy. Pregnancy associated hyper tension is greatest in early and first pregnancies. The disproportion between the mother's still underdeveloped pelvis and the full term fetus is greatest in females less than 18 years of age. The risk of experiencing obstructed labor, which occurs when the birth canal is too narrow or deformed for the passage of the infant, is directly related to maternal age, developmental stage, and stature. Ob structed labor is a leading cause of maternal mortality and morbidity among adolescent mothers in the de veloping world. Adolescents are less likely to seek pre natal care and also to seek unsafe and late abortions which can have a detrimental effect on their health. In 2000, approximately 59% of all unsafe abortions in Africa are performed on young women aged 15–24 years. In the developing world, the highest teenage birth rates of 143 per 1000 women aged 15–19 years are found in sub Saharan Africa.

Adolescent pregnancies in developed countries usu ally occur outside of marriage and teenage motherhood is associated with lower educational levels, higher rates of poverty, and poorer life outcomes for both the teenage mothers and their offspring. Hence, many governments of industrialized countries are working on reducing the number of adolescent pregnancies. In the developed world, teen birthrates are highest in the United States, approximating nearly 50 per 1000 women aged 16–19 in 2000, although the country has seen a fall in teenage births over the years. Within Europe, the United King dom has the highest teen birthrate of about 30 per 1000 women.

Women over 35 years of age also face added health risks when pregnant or giving birth. Many more women in industrialized countries delay having children, ap parently in response to the needs of their careers. Older mothers are more likely to miscarry, develop compli cations such as diabetes and hypertension (high blood pressure) during pregnancy, and their offspring are more likely to suffer from congenital defects.

Clinical Causes of Maternal Morbidity and Mortality

Maternal morbidity or illness during pregnancy and delivery can be induced by a number of clinical causes. The most common of these are hypertension during pregnancy, hemorrhage, obstructed labor, anemia, sepsis, and ruptured uterus. In the mid 2000s, WHO estimated that globally over 300 million women suffered from the short or long term complications of pregnancy and childbirth, with 20 million new cases being added every year. More than half a million women die of compli cations from childbirth annually, making this a leading cause of death globally for adult women of reproductive age. Hemorrhage, eclampsia, and sepsis are particularly dangerous signs of morbidity, but also account for three fourths of all maternal deaths worldwide.

'Hemorrhage' or bleeding during pregnancy is widely regarded as dangerous to the successful completion of pregnancy. Bleeding early in the pregnancy can lead to a miscarriage. Heavy bleeding during or after childbirth can lead to shock, organ failure, and even death if the blood loss is not quickly stemmed through blood trans fusion, medical drugs, or surgery. Hemorrhage is a major cause of maternal deaths worldwide.

Hypertensive disease is potentially dangerous as it can lead to eclampsia, a primary clinical cause of maternal illness and death. Swelling, especially of the face and hands, is widely regarded in the medical community as a signal of possible preeclampsia. The regular monitoring of the pregnant woman's blood pressure, checking her urine for traces of protein and her face and hands for swelling, all signs of possible preeclampsia allows early detection and treatment.

'Obstructed labor' can result in vesicovaginal and rectovaginal fistulae (tears between the vagina and either the bladder or the rectum) causing incontinence as urine and feces leak through the vaginal opening. Most women who develop untreated fistula have given birth at home with no assistance from a skilled birth attendant. In Nigeria, it is estimated that 1 in 10 women suffers from the condition following childbirth. In addition, because of the stigma linked to their condition, women and girls with obstetric fistulae are often abandoned by their husbands and excluded by families and communities. In India and Pakistan, some 70% of women with fistula have been abandoned or divorced. Many are forced to live lives of isolation and abject poverty. Obstetric fistulae account for 8% of maternal deaths worldwide.

'Infections' such as STDs, especially chlamydia, if not diagnosed and treated can also lead to pelvic inflamma tory disease (PID), which has severe and irrevocable effects: infertility, ectopic pregnancy, chronic pelvic pain, and recurrent infection. Infections that occur during pregnancy are called congenital, while perinatal infections occur around the time of birth. Congenital syphilis and congenital HIV infection can be passed from mother to child. Babies born to mothers with chlamydia or gonorrhea or both can develop an eye infection that may lead to blindness if not treated at or shortly after birth. HIV infected mothers' risk of transmitting HIV to their infants range from 13% to 45%. Syphilis and genital herpes appear to be associated with both intra uterine growth retardation and premature delivery, while gonorrhea, chlamydia, and trichomoniasis affect uterine growth and produce premature delivery.

Infections can also occur if aseptic procedures are not followed during the delivery. Frequent vaginal examin ations and prolonged or obstructed labor are risk factors for infection. Fever and inflammation are symptoms of infection and can be minimized by maintaining sanitary conditions. Among the long term consequences of in fections during childbirth are chronic problems such as pelvic inflammatory disease and secondary infertility as well as complications in succeeding births.

Societal Causes of Morbidity and Mortality

Every society has cultural practices and social norms that influence reproductive health, pregnancy, and childbirth. Religious and personal beliefs and behaviors; social in stitutions and customs, economic means and constraints, and government policies also affect women's health during pregnancy and childbirth. Female morbidity and mortality are not just functions of physical differences between the sexes, but are also rooted in differences in roles and status in the societies in which they live. In the developing world, gender biases often result in the neg lect and undervaluation of girls and women, as seen in their low social status and the discriminatory traditional practices they may face. These factors compounded by poor environmental conditions often result in health problems for women during their reproductive years, particularly during pregnancy and childbirth. Among the most important factors affecting mortality in childbirth are adequate nutrition and access to quality medical care, where 'access' is affected both by the cost of available care, and distance from health services.

Undernutrition leads to high rates of maternal mor bidity and mortality. It also increases the risk of giving birth to a low birthweight infant (i.e., an infant weighing less than 2500 g). If the pregnant woman is well nourished and in good health, complications of pregnancy and childbirth can be reduced. A deficit of adequate amounts of iron, which in its extreme form can cause anemia, increases the risk of both maternal mortality and mor bidity. Extremely high prevalence rates of anemia of between 30 and 50% are reported for low income women in tropical countries. In South Asia, over 60% of pregnant women and nearly as high a percentage of nonpregnant women in their reproductive years were found to be anemic. Factors that contribute to the high level of anemia among pregnant women in South Asia include poor dietary practices during pregnancy, infection, malabsorption of nutrients, and malaria. Iodine deficiency has severe negative consequences for mothers and infants, causing impaired mental function, and in creased rates of spontaneous abortions, still births, and infant deaths. Iodine deficiency exists in pockets throughout the world, most often occurring in inland mountainous regions or floodplains where the topsoil is leached of iodine by rain, flooding, glaciation, or snow.

Cultural practices such as female genital cutting (FGC) also called female circumcision or female genital mutilation have grave effects on the well being of women during their reproductive years. The procedure involves cutting or removing the external female genitalia for nontherapeutic reasons. FGC is traditionally practiced among societies mainly located in a band of African countries north of the equator, from Mauritania and Guinea in the west to Somalia and Egypt in the east. It is also found in some parts of South Asia in less extreme forms. It is most often carried out by traditional prac titioners on girls that range in age from a few days old to 20 years old. FGC can cause pain, severe bleeding, in fections, and even death. It makes childbirth very painful and the woman more prone to infection. FGC is banned in most Western countries, but has accompanied immi grants from areas where it is traditionally practiced. Families who continue to observe the practice may view genital cutting as both culturally and religiously pre scribed, making it difficult to eradicate. There is concern that legislation against FGC may be counter productive, forcing it underground where it is performed in unhygienic settings. Recently, African countries such as Eritrea and Egypt have banned FGC: it has the approval of state governments only in Sudan and Somalia.

Policies for Improving Maternal Health

Special care for pregnant women through public health services was a relatively late development. It was only in the 1930s, about 30 years after the formalization of labor and delivery care that women were offered regular check up during pregnancy as an integral part of maternal care even in a developed country like the United Kingdom.

During the United Nations Decade for Women (1976–85), as the magnitude and scale of the problems associated with poor maternal health became increas ingly apparent, international agencies and government programs placed an emphasis on the health of women during the childbearing period of their lives, but par ticularly during pregnancy and childbirth. The World

Health Organization's (WHO's) Safe Motherhood Ini tiative (SMI) in 1987 was the first to focus specifically on the health of women. In cooperation with other international institutions, the WHO aimed to reduce the health risks of pregnancy and childbirth worldwide through this initiative. The actions proposed included better community based healthcare, referral facilities especially in cases of emergencies, and transport systems. The SMI's target was to reduce maternal mortality by at least half by the end of the twentieth century. The pro gram has been successful in some countries like Egypt, where the national MMR was halved between 1992–93 and 2000. However, it has not been uniformly successful in all developing countries.

Today, particularly in the developing world, the majority of maternal deaths is preventable through measures such as screening to detect women at high risk (such as very young women, older mothers, women with STDs, and other reproductive tract infections); referring women with complicated pregnancies to higher level care and encouraging them to deliver in health facilities, and providing tetanus toxoid immunizations, iron/folate supplements, and where necessary, malaria prophylaxis to pregnant women. Other measures include ensuring that all pregnant women receive pelvic examin ations and monitoring of their weight and blood pressure.

High priorities for care at delivery include provision of hygienic supplies, the training of birth attendants (both traditional and public health staff), and the maintenance of a referral system for complications. The averting of maternal mortality through the combined effects of antenatal (prenatal) care and safe delivery especially for high risk pregnancies are among the most cost effective interventions for adults, and have been highlighted and adopted as specific goals at all key international confer ences since the 1990s that focused on the health of women and children. Such major global conferences include the World Summit for Children in 1990, the International Conference on Population and Development in 1994, the Fourth World Conference on Women in 1995 and their 5 year follow up evaluations, and the United Nations General Assembly Special Session on Children in 2002.

The political momentum of the policies formulated at these conferences was echoed in the Millennium Development Goals (MDGs). Goal 5 of the MDGs is to 'improve maternal health'. Its target is to reduce the maternal mortality ratio by three quarters between 1990 and 2015. The goals and targets coupled with a consensus about how to tackle maternal mortality technically, greatly improves the prospect of reducing women's death and disability rates due to pregnancy and birth. International efforts to reduce maternal mortality are also looking to assist states deliver the right services to women through their health systems, particularly in rural areas in developing countries.

Advocacy for improved maternal health and women's health overall, as articulated in the conferences in Cairo and Beijing, are grounded in the notion of women's rights. There is also greater recognition that women's reproductive rights, such as choosing when, if, and how to give birth should be safeguarded and promoted. Pressure to restrict and rescind access to abortion, forced steril ization of minority women such as the Romani in gov ernment run health facilities in the Czech Republic and Slovakia not only during the Communist era, but also as recently as 2003; and the restriction of access to an array of contraceptives are all regarded as infringements on women's reproductive rights.

The idea that women's health status is a result of the accumulation of conditions that began earlier in life has led to looking at maternal issues within the context of the life cycle approach. The life cycle refers to the entire span of an individual's life from birth to death. Organ izations such as the Women's Global Network for Reproductive Rights (WGNRR) underscore the need to address a wide range of issues concerning maternal health, mortality, and childbearing; not just fertility control. Currently, even programs that promote women's reproductive health are steering away from a narrow focus on the biological aspects of fertility such as preg nancy and birth, and looking to improve sociocultural aspects that impinge on the lives and health of girls and women over the entire life cycle. Recognizing that all of society is implicated in women's health, the organiza tional frameworks of many reproductive and maternal health programs also look to incorporate men and indeed, all of society in the process.

See also: Body, The; Chicago School; Health Geography; Medical Geography.

Further Reading

Abel, S. and Kearns, R. A. (1991). Birth places: A geographical perspective on planned home birth in New Zealand. *Social Science & Medicine* 33(7), 825 834.

Buor, D. and Bream, K. (2004). An analysis of the determinants of maternal mortality in sub Saharan Africa. *Journal of Women's Health* 13, 926 938.

Cook, R. J., Dickens, B. M. and Fathalla, M. F. (2003). *Reproductive Health and Human Rights: Integrating Medicine, Ethics and Law*. Oxford: Oxford University Press.

Dixon Mueller, R. and Dagg, P. K. B. (2002). *Abortion and Common Sense*. Philadelphia, PA: Xlibris Corporation.

Dyck, I., Lewis, N. and McLafferty, S. (eds.) (2001). *Geographies of Women's Health*. New York: Routledge.

Fannin, M. (2003). Domesticating birth in the hospital: 'Family centered' birth and the emergence of 'homelike' birthing rooms. *Antipode* 35(3), 512 535.

Koblinksy, M. A. (ed.) (2003). *Reducing Maternal Mortality: Leaning from Bolivia, China, Egypt, Honduras, Indonesia, Jamaica and Zimbabwe*. Washington, DC: International Bank for Reconstruction and Development/World Bank.

Longhurst, R. (2001). *Bodies: Exploring Fluid Boundaries*. (Critical Geographies). London: Routledge.

Maternal Mortality in 2000 (2004). *Estimates Developed by WHO, UNICEF and UNFPA*. Geneva, Switzerland: WHO.

Maynard, R. A. (ed.) (1997). *Kids Having Kids: Economic Costs and Social Consequences of Teen Pregnancy*. Washington, DC: The Urban Institute.

Teather, E. A. (ed.) (1999). *Embodied Geographies: Space, Bodies and Rites of Passage*. London: Routledge.

The World Health Report (2005). *Make Every Mother and Child Count (2005)*. Geneva, Switzerland: WHO.

Relevant Websites

http://www.who.int/reproductive health/publications/maternal mortality 2005/index.html
WHO UNICEF UNFPA. 2007. Maternal mortality in 2005 Estimates Developed by WHO, UNICEF and UNFPA.

http://www.who.int/reproductive health/publications/antenatal care/index.html
WHO UNICEF. 2003. Antenatal care in developing countries: Promises, achievements and missed opportunities.

http://www.who.int/reproductive health/publications/ifc/index.html
WHO. 2003. Working with Individuals, Families and Communities to Improve Maternal and Newborn Health.

http://www.who.int/reproductive health/publications/RHR 01 5 advancing safe motherhood/index.html
WHO. 2001. Advancing Safe Motherhood through Human Rights.

Private/Public Divide

D. Collins, University of Auckland, Auckland, New Zealand

Glossary

Dichotomy A division or distinction that divides a phenomenon, or set of phenomena, into two mutually exclusive, opposed, or contradictory groups.

Introduction

The private/public divide is a central concept in political and legal debate, as well as much everyday conversation. Because of this prominence, as well as the frequency with which the concept is articulated in spatial terms (referring either literally, or metaphorically, to the distinction between public and private spaces), it is also of considerable interest to many human geographers, particularly those concerned with issues such as gender, power, and property. While the private/public divide is a common entry point into debates about the organization of space and society, the meanings of the terms private and public are often unclear. There is a lack of consensus on how they should be used, and what they refer to. This ambiguity of meaning appears to add the rhetorical power of the divide by lending it flexibility: in many arguments, opposing sides can and do invoke the distinction to support radically different views.

The private/public divide has been characterized as one of the grand dichotomies of Western thought. It is a conceptual opposition that divides the world into two mutually exclusive categories, which together appear to account for all elements of human life and experience. This distinction is not only all encompassing, but also hegemonic in that it tends to subsume other social distinctions (e.g., politics/economics and work/home). While the private and the public are understood as opposites, they are also inseparable: to speak of one particular sphere of life, such as health care, as being public suggests that there is a private alternative. Similarly, public space stops where private space begins (and vice versa). Because private and public are generally thought to account for the entirety of the human world (although geographers have identified alternative ways of characterizing some spaces), it follows that the extension of one sphere (or space) will invariably reduce the scope or power of the other.

Expansions and contractions of public and private influences, and spaces, typically prompt normative judgements. In neoliberal discourse, for example, governmental intervention into the free market is often portrayed as a worrying, inefficient, and unjustified intrusion of public authority into the private world of freely contracting individuals. Many people, of course, contest this assessment – but typically do so using the same broad vocabulary of public and private spheres. This points to the pervasiveness of the assumption that all elements of the human world can, and should, be labeled as either public or private (although fair minded individuals can and do disagree as to how this labeling should proceed).

Historical Distinctions

In discussing public and private, we are dealing with categories of Greek origin. In the Greek city state, Hannah Arendt writes, the private/public distinction mapped onto a distinction between the *polis*, which was common to all free citizens, and the *oikos*, or household. The public sphere of the *polis* was constituted by political speech and common action, the latter including the waging of war and competition in athletic games. This realm existed by virtue of a patrimonial slave economy centered on the private sphere of the *oikos*. The labor that was performed in this sphere by women, slaves, children, and foreigners provided household masters with the foundation and autonomy that were prerequisites for engaging in public life. While private life was a realm of necessity (both economic and biological), the public sphere was envisioned as a realm of freedom. The *polis* provided a setting in which citizens could distinguish and immortalize themselves: while citizens interacted as equals, each sought to excel. Arendt recommends this public sphere as a model for refocusing politics on the public good – although in so doing arguably paints herself as an apologist for a world in which only a handful of males were able to participate in public life, with the great majority of the population slaving (often literally) to satisfy their private needs under constant threat of violence.

If a clear distinction between public and private was characteristic of the Greek city state, the same could not be said for feudal Europe. Under feudalism, royal and ecclesiastical authority, the apparatus of the state, the economic world of production and consumption, and domestic life of the family were all part of a unitary hierarchy. This was a system based upon a web of personal ties and dependencies, and all aspects of life were at once personal and political. Change was brought by a variety of factors, including the emergence of liberal political thought, which argued for a private realm within which every rational individual was free to exercise his own judgement and make choices (particularly in matters of economic exchange and religious belief). This personal freedom came to be expressed in the form of individual rights which determined the boundaries of state power.

The right to religious freedom secured one of the first spheres of private autonomy, promising that the individual (and, by extension, the family unit) was not accountable to external authority in matters of belief. In particular, the authority of the state to coerce individuals to accept or reject particular conscientious beliefs was circumscribed. While this foundational liberal principle is well known, its application continues to be highly contested: especially within the spaces of public institutions such as schools, and occasionally parks, court houses, and legislatures.

With the ascendancy of this school of thought, and the rise of a mercantile economy, social and economic relations began to lose their directly political character: a civil society of privately contracting individuals began to stand in opposition to the public authority of the state. Jürgen Habermas provides a highly influential account of this process, which culminated in the emergence of a bourgeois public sphere, in which relatively prosperous individuals came together to debate questions of public interest in a so called rational manner. He goes on to describe and evaluate this sphere in a largely aspatial fashion, presenting it as a suite of institutions and activities that mediated between a private sphere founded on the free market (often referred to as civil society) and a formal public sphere of state authority. Through reasoned written communication, as well as face to face debate, Habermas claims, the emergent European bourgeoisie created public opinion. While aware of the exclusion inherent in this sphere (e.g., of the working class and, in many instances, of all women irrespective of socioeconomic status), Habermas praises its normative potential as an ideal for participatory political action.

Contemporary Definitions

Defining the private/public divide today is no easy task, in part because of its lengthy and much debated heritage. One useful starting point is the observation that it does not refer to a single dichotomy, but rather a set of interrelated binary distinctions. While unifying or common themes may be identified among these alternative definitions, four of which are considered below, they also differ in important ways: they point to different phenomena, rest on different understandings of the social world, and are underpinned by different ideologies. Identifying and analyzing these alternative definitions can become so complex a task that one loses sight of the social meaningfulness of the terms involved. Although no single, unproblematic definition of the private/public distinction can be offered, and although scholars sometimes advance incompatible conceptualizations, the words private and public still appear to mean a great deal in everyday life. To speak of public space, for example, is to suggest certain expectations about access, ownership, and use of that space – even though we may not be certain who owns that space, or for what purpose.

Some understandings, or conceptualizations, of the private/public divide are particularly prominent in academic, legal, and popular discourse, and four are considered here, following the work of Jeff Weintraub. The first approach is strongly associated with liberalism, and prevails within much economic debate and public policy analysis. It equates the difference between public and private with that between the state and the market economy. It is often invoked within discussions over whether a particular good or service (again, health care is a pertinent example) should be supplied by the market, or subject to some form of governmental intervention up to and including complete state control. Such arguments often center on where the line between the public (as the state) and the private (as the market) should be drawn: they seldom ask whether it should be drawn at all, or whether it should be conceived of in altogether different terms.

A second understanding of the distinction is based in classical and republican philosophy, and understands the term public to refer to active political life, which is envisioned as analytically distinct from both the private realm of the market and administrative matters of government. At the heart of the public realm, in this model, is discussion, deliberation, and decision making about issues of legitimate collective concern. It is this understanding which is typically invoked when people speak of participating in public life, or contributing to public debate. The work of both Hannah Arendt and Jürgen Habermas, referred to above, sought to characterize and theorize the public life of active citizens (in Ancient Greece and nineteenth century Western Europe, respectively). Habermas, for example, identified a public sphere which was a site for the production and circulation of rational debate, and the communication of common concerns. It was conceptually distinct from both the formal apparatus of government, and the market economy.

A third distinction centers on different forms of social activity, counterposing the intimate relationships found in the family home with the social interaction of diverse individuals and groups found elsewhere. According to this view, public life is not about politics (however this term might be understood) but about social interaction with strangers and casual acquaintances. This understanding of public life is often invoked in considerations of the extent to which urban public spaces promote (or inhibit) social activity – a central concern of new urbanism, for example.

Finally, feminist thought conceives of the difference between private and public in terms of the distinction between the family and the wider economic and political order. In general terms, the family home is seen to represent the private realm, as is the case for those theorists primarily concerned with different forms of sociability. The public sphere is not defined solely in terms of social interaction, however, but also in terms of the state, the official economy of paid employment, and arenas of public discourse. A central concern is the relationship between the public sphere of formal economic production (and paid labor) and a private sphere of domestic reproduction (and unpaid labor). Feminist scholars have developed powerful critiques of this distinction, noting not only the exclusionary nature of the public realm, and the exploitation which can occur in the private realm, but also the connectedness of the private and the public, a point considered in the following section.

In reflecting on these conceptions of the private/public divide, and others, it is possible to follow authors such as Weintraub, and Benn and Gaus, in identifying multiple axes of differentiation:

1. What is political versus what is nonpolitical.
2. What is personal versus what is societal.
3. What is hidden versus what is open, revealed, or accessible.
4. What is individual, or pertaining only to a distinct group of individuals, versus what is collective, or of concern to the whole community.

Each of these axes is expressed as a dichotomy. Yet it is possible to conceive of public and private as existing on a continuum, which allows identification of degrees of publicness and privateness. Thus, a space in public ownership, such as a remote park or beach, might be perceived as relatively private when visited by very few people, but grow to be more public as it becomes more widely known, popular, and/or accessible. While such a continuum is appealing, it cannot alter the fact that there are separate sets of criteria that may be used to define public and private in the first instance (in this case, the third axis identified above is most pertinent).

Clearly, the private/public distinction is complex, and somewhat treacherous. This is true in both theory, and political practice – where those spheres of life often identified by those on the left of the spectrum as properly public are consistently characterized by many on the right as properly private (and vice versa). For example, left wing thought conventionally recognizes a legitimate public role in limiting and regulating the economic power of property owners, while right wing thought often emphasizes the need to protect the privacy of such power. Conversely, many in the left support greater privacy in matters of civil liberties and reproductive choices, while many of those on the right advocate public regulation and stronger state control in these areas. The private freedoms recognized by left and right are often opposed, sometimes diametrically, although once again we can see at least one commonality: both sides generally believe in the normative and analytic value of labeling particular phenomena, public or private.

Maintaining the Distinction

The liberal distinction between the private sphere of economic activity and the public sphere of government action is today the most commonly invoked private/public distinction. It is powerful for a number of reasons. First, it maps on to a spatial distinction between real property owned by private individuals and corporations, for their own use and benefit (i.e., private space), and that owned by local government and the state, typically for the use and benefit of the wider public (i.e., public space). Second, it is policed by the legal system, which invests considerable energy in maintaining the boundary between the private realm of individual actions and the public sphere of state action. Moreover, the compartmentalization of human affairs and concerns into public and private spheres is often presented as preordained and politically neutral. Yet, as the discussion of left and right wing political positions above suggested, maintaining a meaningful distinction between matters of state concern, and those of concern only to private actors, is a difficult and ultimately ideological process. Indeed, a number of scholars have suggested that the distinction is arbitrary and ultimately unsupportable.

One critique notes that public authority is thoroughly caught up in activities deemed private – not least of all because it is the state that defines, protects, and lawfully abridges the rights of individuals (including the rights of private property owners). This school of thought points to the implausibility of a private economic realm, when the free market is in fact a political choice, and one that is made possible by virtue of public enforcement of private property and contractual rights, especially within the legal system. Such arguments connect to a long tradition of socialist thought which presents the liberal private/public distinction as serving a primarily ideological purpose. Far from reflecting natural or innate differences between the state and the market, it functions primarily to depoliticize the inequalities that result from private ownership and control of property. If one accepts this line of reasoning, the relevant question for political debate becomes not whether the state ought to intervene in private affairs, but on whose behalf it is already intervening.

A related critique, advanced by many progressive political groups, seeks to publicize, and politicize, issues traditionally deemed private. They have highlighted, among other issues, racial and gender discrimination, exploitation in the workplace, violence against women and children, and the impacts of private decision making on the natural environment. One consistent theme of struggles against oppressive social relations has been to change their designation from private and nonpolitical concerns to legitimate issues of public interest and political intervention. In this they have achieved considerable success. However, it is not only the political left that can reconfigure the private/public divide: right wing thought has responded to the rise of progressive politics with forthright reassertions of individualism and private property rights (often envisioned in absolute terms).

Additional critiques of the private/public divide can be found within feminist thought. Feminist analyses of the distinction have taken at least three forms: some have deployed it as an analytical tool for examining prevailing social relations, others have sought to alter it by challenging existing designations of particular arrangements as public or private, while a third approach has sought to collapse the distinction altogether as part of a broader critique of the dichotomies that pervade Enlightenment thought. One school of thought within the third of these approaches emphasizes what women's unpaid domestic labor reproduces (both physically and ideologically) the paid labor that the capitalist economy requires. Taken to its logical conclusion, this argument subverts the private/public distinction: it can be argued that the only real difference between a woman raising children in the private home, and a man laboring for a public company, is that the man receives financial remuneration, while the woman does not.

Regardless of the approach adopted, feminist scholarship has been characterized by a strong concern for the ways in which private and public are gendered, and by a keen interest in social consequences of labeling family life, and domestic space, as private. These designations are seen to reflect, and reinforce, inequality between men and women, with the latter often confined to subservient (if sometimes idealized) roles. Clear links have been drawn between this subordination, and classical thought which suggested women and women's bodies represented the private, the particular, and the natural: everything that should be excluded from the public sphere.

Classifying Spaces

The characterization of material spaces as either public or private typically refers to the distinction between parcels of real property owned by state authorities (most obviously the street), and those owned by private individuals (the archetype being the private home). This basic geographical manifestation of the private/public distinction plays an important role in shaping social behaviors and expectations. Most children are taught, from a relatively young age, about the meanings of property boundaries, and why they cannot enter privately owned spaces without permission. The benefits of private property ownership have been understood to encompass the right to exclude not only other members of the public, but also the public authority of the state (although today the latter right is significantly conscribed by an array of legislation). Public space, by contrast, is ideally open to all.

At first glance, public and private seem to exhaust the possibilities of property ownership. On closer inspection, however, the situation is more complex. Many geographers have reflected on the extent to which much of the public life of cities now occurs in privately owned spaces, especially spaces of consumption such as shopping malls. These spaces are clearly open to the public, on whose custom they depend, and have become sites of public sociality – often to the detriment of publicly owned streets, parks, and squares in downtown areas. At the same time, they are privately owned and controlled, with management free to invoke the law of trespass to exclude unwanted guests for almost any reason (or, indeed, for no reason at all). Accordingly, shopping malls, retail tunnels and walkways below and above cities, and privately owned festival sites have been labeled quasi public spaces.

The rise of quasi public spaces, especially in suburban locations, has prompted a variety of responses in downtown areas concerned about real or potential decline. One such response has been to remake the downtown's public owned spaces in the image of private commercial spaces. This has involved increased regulation and policing of public spaces with a goal of attracting middle and upper class consumers back to the city. In many instances, this has entailed an increasing role for local private property owners and major corporate tenants in governance, greater levels of private commercial activity, and a prominent role for private security in enforcing rules of public deportment. The net effect, it has been suggested, is to create quasi private spaces, which are formally owned by the state, in the name of the public, but are subject to control by private interests, often to the detriment of the poor and homeless. For many authors, this privatization combined with increasingly intense and exclusionary regulation, constitutes the death of public space.

In some respects, this characterization is not unreasonable – the principles of public stewardship and open access are severely eroded, while the power of private actors is significantly increased. Yet, a privately controlled and public space without members of the public present, preferably in large numbers, would be a failure, even by the relatively narrow terms of its regulators and promoters. What such transformations often entail are partnerships between public (governmental) and private (commercial) interests to redevelop a space so as to make it more attractive to a relatively prosperous, consumption oriented, security conscious public, and to foster bourgeois forms of public sociality, at the expense of more marginal publics previously present in that space. In the process, some public (collective) interests, such as the desirability of a relatively safe, economically productive downtown, may be advanced, while others, such as social justice, and right to exercise free expression, may be circumscribed.

In considering such issues, the multiple meanings of the private/public divide are once again apparent. With respect to material space, it is clear that merely identifying the owners of a particular site as either public or private actors may go only part way to describing how private and public interests are weighted at that site. Moreover, while the terms public and private are almost invariably used in opposition, it is clear that many spaces cannot be neatly characterized as one or the other. There are many sites at which public and private interests converge. In contemporary Western cities, the shopping mall is often a particularly clear example of this convergence. More generally, however, we can also observe that state authorities have a long standing interest in regulation and surveillance of the family home, and the rights of private property owners to use space as they see fit are often extensively conscribed in the public interest. In some less developed countries, activities which appear deeply private to many Western eyes – such as personal cooking and cleaning, as well as small scale commercial activity – occur in the public space of the street. Thus, the geographies of state regulation, as well as those of everyday social life, often bear only a partial resemblance to formal distinctions between private and public space.

Summary

There are many options available when seeking to draw a line between private and public. Yet, whichever option is adopted in a particular case, the result can be socially and politically meaningful. In particular, as both feminist and socialist critics of the divide have noted, the characterization of certain issues and spaces as private typically has the effect of shielding them, at least partially, from critical analysis in the public world of debate and democratic decision making. This is not necessarily to condemn the private/public distinction, nor to advocate its demise; it is simply to recognize its social power. Indeed, abandoning the distinction could also produce socially and politically deleterious results: it would undermine the right to privacy, and potentially expose every aspect of human life and activity to greater external scrutiny, particularly scrutiny from impersonal institutions. For some commentators, this is the recipe for a totalitarian nightmare.

Early theorists of the divide often suggested that a clear and unproblematic boundary can and should be drawn between private and public: for both Arendt and Habermas the public sphere of political action was founded upon a private sphere devoted to production and reproduction, yet properly distinct from it. Much contemporary debate is more critical. Geographers have contributed to this critique in a number of ways, in particular by illustrating that private and public may be more tangled than often thought. While it is commonplace to characterize the owners of spaces as either private or public, this distinction is formalistic, and belies the extent to which private and public interests and concerns often overlap in the same space.

See also: Central Business District; Feminism/Feminist Geography; Home; Liberalism; Malls/Retail Parks; Neoliberalism; New Urbanism; Philosophy and Human Geography; Public Space; Public Spaces, Urban.

Further Reading

Arendt, H. (1958). *The human condition*. Chicago: University of Chicago Press.

Benn, S. I. and Gaus, G. F. (1983). *Public and private in social life*. London: Croom Helm.

Blomley, N. (2005). Flowers in the bathtub: Boundary crossing at the public private divide. *Geoforum* 36, 281 296.

Bobbio, N. (1989). *Democracy and dictatorship: The nature and limits of state power*. Cambridge, UK: Polity Press.

Calhoun, C. (1992). *Habermas and the public sphere*. Cambridge, MA: MIT Press.

Collins, D. (2006). Culture, religion and curriculum: Lessons from the 'three books' controversy in Surrey, BC. *Canadian Geographer* 50, 342 357.

Drummond, L. B. W. (2000). Street scenes: Practices of public and private space in urban Vietnam. *Urban Studies* 37, 2377 2391.

Elshtain, J. B. (1981). *Public man, private woman*. Princeton: Princeton University Press.

Habermas, J. (1989). *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. Cambridge, MA: MIT Press.

Hopkins, J. (1993). A balancing act? Trespass amendments for public/private places. *Urban Geography* 14, 114 118.

Mitchell, D. and Staeheli, L. A. (2006). Clean and safe? Redevelopment, public space, and homelessness in downtown San Diego. In Low, S. & Smith, N. (eds.) *The politics of public space*, pp 143 175. New York: Routledge.

Pateman, C. (1989). *The disorder of women: Democracy, feminism and political theory*. Cambridge, UK: Polity Press.

Weintraub, J. (1997). The theory and politics of the public/private distinction. In Weintraub, J. & Kumar, K. (eds.) *Public and private in thought and practice*, pp 1 42. Chicago: University of Chicago Press.

Privatization

P. M. O'Neill, University of Western Sydney, NSW, Australia

Glossary

Deregulation It refers to the shift in societal organization away from government-led regulations and procedures toward market-based systems of allocations alongside increased private responsibility for social and economic outcomes. The term 'market liberalization' is often sued as a synonym for deregulation. Market liberalization refers to the removal of government regulations controlling the nature and extent of market transactions.

Keynesian Keynesianism refers to the practices of macroeconomic management based on the writings of British economist John Maynard Keynes (1883-1946). Keynesian asserts that the level of economic activity in a domestic economy is based on the level of aggregate demand - due principally to private consumption, investment spending, exporting, and government expenditure. Economic management based on Keynesianism principles assumes that government expenditure can be manipulated to a society's advantage to compensate for the relative performance in other forms of demand. Keynesianism thus asserts a prime role for governments in the positive management of modern economies. Keynesian was the predominant form of economic management in industrial economies in the decades after World War II.

Governance It refers to the assemblage of laws, regulations, and conventions that control and steer the operations of a society. This assemblage also includes the institutions that build and reproduce these laws, regulations, and conventions.

Introduction

Privatization is the transfer of publicly owned or publicly operated means of production to private ownership or operation. The argument for this transfer is usually that privately run enterprises are subject to the discipline of the market and therefore they will be more efficient. A related argument for privatization is that enterprises that are privately owned are more valued and better maintained. Both arguments support a view that privatization maximizes public benefit and welfare.

The nature and extent to which states should be involved in the affairs of the market are as old as the idea of the state itself. Urgings for privatization are equally historical. The discussion here, though, focuses on the contemporary period and its antecedents in the years following World War II.

Types of Privatization

Privatization is a complex thing. It is the subject of intense political and economic debate, yet often with indistinct or confusing categorizations. In order to provide some clarification, four distinct forms of privatization are discussed in the following.

Sale of Public Assets

The most common form of privatization is the sale of publicly owned enterprises, also called state owned enterprises or SOEs, into private ownership. In developed nations this form of privatization has involved the widespread sell off of SOEs and utilities that were established or nationalized in the two decades following World War II. During this period it was common for governments to take control of the production of goods and services in certain sectors for specific public interest reasons. One reason was to protect the public from unreasonable behavior by a producer that might have had a natural monopoly for the supply of a good or service giving it unreasonable power over price, delivery conditions, or quality. Examples of natural monopolies in the post war period included postal and telecommunications services; public health and education services; water, electricity, and gas supply; police and most public transport services.

Another common reason why states expanded public ownership was to build their capacity to generate national economic growth and prosperity. Common post war nationalizations and SOE startups involved mines, steelworks, coastal shipping lines, airlines, banks, and insurance agencies. Government enthusiasm for public ownership often over rided strict economic management criteria, however, with public ownership pursued for reasons of national sentiment rather than purely in the pursuit of efficiency or fairness. Owning a national airline, for example, became a common practice.

Public ownership of productive enterprises was even more pronounced in the communist bloc nations, now rebadged as the transition economies. These are the economies emerging from long periods of state socialism and communism, especially the former USSR and the Eastern European nations. The sale of state owned assets and enterprises has been a key component of privatization in these nations in their transformation from command to market economies, although the actual nature of privatization has varied markedly from one nation to the next. Privatization in Russia has involved a variety of means including the distribution of ownership shares to Russian citizens through voucher schemes, share auctions, and outright sales of SOEs to both domestic and foreign interests. In contrast, the Czech and Slovak Republics have relied heavily on voucher privatizations while Hungary's privatizations have in volved a high concentration of direct foreign sales.

A major impediment to the success of the privatization programs in the transition economies has been the slowness in implementing appropriate regulatory systems to govern the operation of private markets and to determine the conditions of private ownership including rights to ongoing transfer. Another has been the tardiness of the banking system across the transition economies to meet the demands of the newly privatized entities. A further issue, not yet adequately addressed, and one which has been most manifest in the large utilities, has involved the problem of labor displacement as newly privatized or marketized enterprises seek productivity gains by drastically cutting staffing levels.

Public ownership of productive enterprises has been a more recent phenomenon in developing countries. Typically, these nations have had only a short period of independence following their emergence from European colonialism in the mid and late twentieth century. Consequently, national control and operation of public enterprises in developing nations have been limited. Nevertheless, the reversal of public ownership in these nations has been as prominent as in developed and transition economies. According to the United Nations Conference on Trade and Development (UNCTAD), a majority of the 50 largest privatizations in the world during the 1987–99 period were in developing countries. The most common form of privatization of SOEs in the developing world has been by direct sale to both domestic and foreign interests.

Some of the most significant privatizations of SOEs around the world have been British Airways (1987), British Steel (1988), Deutshe Telekom (1996), the Italian gas and energy producer ENI (1996), the French oil and gas producer Total (1996), the Chilean telecommunica tions company ENTEL (1986), the Commonwealth Bank of Australia (1991–96), and the Canadian electricity utility Ontario Hydro (1999).

Shift from Provision to Procurement

Privatization also refers to a shift in responsibility for the provision of basic and community services. A com mon expectation around the world is that nation states will organize the provision of basic services to their citizens. Many of these services fall into a category called public goods, which are goods or services that are created though universal provision. This means that most, if not all, of the benefit of the good or service is severely cur tailed should it only be offered as a private or commercial commodity. A suburban road is one example since its purpose is to give all residents of a neighborhood access to a common thoroughfare for movement to and from their homes. The personal and private provision of roads in urban areas makes no sense. A viable suburban road can only exist as a public good. Likewise, a public health service is a public good since the general health of any community depends critically on the maintenance of minimum levels of health among all individuals such as by the control of communicable diseases through immunization and vaccination programs. So too, day to day life requires a universal police service to ensure the orderly and safe movement of people and goods including the protection of property.

Yet many governments, while retaining responsibility for ensuring that basic, or public, goods and services are available, have privatized their provision. This means they have moved to a model of procurement and away from one of direct provision. In education, health, or child care, for example, this has involved direct payments by governments to private providers to ensure citizens have access to services without governments being dir ectly responsible for their provision. Sometimes this type of privatization is accompanied to cover part of the cost with exemptions provided to some users on equity or need criteria. Many of the world's public health systems now rely on private providers with a contributory or co payment required of users. Other examples of the shift in government role from provision to procurement include privately operated toll roads and railways, contracted garbage services, private water supply companies, the substitution of private security for public policing ser vices such as at football matches and rock concerts and for protection of more exclusive private property, and even military functions such as supply, ordinance, and frontline roles in the Iraq war (2003–).

The shift to procurement can also involve enlisting agencies from the nonprofit sector. This has become increasingly common in the provision of welfare and employment assistance services with governments en gaging charity and religious based organizations, for example, to manage high risk groups on a contract basis. Commonly, not for profit groups are seen as effective ways to target those householders seemingly beyond the service capabilities of government agencies and private (for profit) providers.

Shift to Market-Based Regulatory Systems

Privatization also refers to the shift in regulatory roles away from centralized determination by governments and government agencies. This type of privatization is also a form of deregulation. Some of the impetus for regulatory shift comes from globalization. Examples include the removal of controls over the exchange rates of national currencies and the interest rates chargeable by private lenders, the setting of global standards in a wide variety of industrial sectors by the privately run International Standards Office, and the monitoring of global health standards by the partially privatized World Health Organization. At a national level, examples of the shift away from state led regulatory systems include industry based regulatory bodies, sometimes called self regulation systems, such as involving the media sector, public advertising or product safety; shifts away from centralized, state administered labor regulations to work arrangements based on employer–employee contractual relations; and the replacement of authorities with responsibility for the conduct of social and sporting events with bodies drawn from the voluntary and community sectors. The shift in regulatory responsibility away from the state, or the abandonment of regulations altogether, are often described by reformers as market liberalization policies.

Transformation of the Modern State

Privatization in its broader sense refers to the contemporary political transformation of the role of the state. This transformation elevates a preference for *laissez faire* liberalism above state centered collectivism as society's principal organizational rationale. A regime of privatization is seen to maximize efficiency and enhance individual choice and freedom through a reliance on private property rights and market forces. State led collectivism, on the other hand, is seen to produce heavy handed regulation, inefficient bureaucracy, and a brake on progress. In this broad sense, then, privatization is a mode of societal and economic organization within the ideology of neoliberalism which has specific preferences for private ownership, monetized ways of measuring value, market based regulation and control arrangements, and individual, merit based reward and advancement systems. Privatization, according to proponents of neoliberalism, is seen as more capable of delivering a prosperous, efficient, and equitable modern society.

If this were indisputably true, however, then the embrace of privatization around the world would be incontrovertible. Yet privatization is a contested and controversial process. Importantly, theoretical arguments over how best to allocate economic responsibilities between markets and the state are enduring. Markets are preferred when they operate under certain conditions. Usually these are that there are no significant externalities in production or consumption, that the product is not a public good, that the market structure is not monopolistic, and that information transfer costs are relatively cheap. When some or more of these conditions are absent, markets have a tendency to fail on efficiency, effectiveness, and fairness grounds requiring some form of state led organizational vehicle, SOE, or utility to take a lead role. When market failure is likely, then, government intervention or takeover, to a greater or lesser extent, is desirable.

For these reasons, privatization, in any of its forms, has not overwhelmed the organization and provision of goods and services around the world. While privatization has shifted significant resources and means of production across to the private sector, the state still remains a key producer of public goods, the prime regulator of economy and society, and the most significant entity for the advancement of social well being in sustainable and equitable ways.

History of Privatization

In the developed world, the focus of privatization has been the dismantling or diminution of the powerful and pervasive type of nation state that was assembled in the post World War II period, especially in the UK, Western Europe, Australia and New Zealand, and to a lesser extent the USA. The presence of the state rose in four major ways: as a hands on manager of a nation's macroeconomic conditions; as an active regulator of society's industrial, social, and environmental affairs; as a major producer of social and public goods and services; and as a producer of producer and consumer goods and services for commercial sale.

The motivation for this rise was threefold: (1) rehabilitation following the war, (2) economic insurance following the 1930s depression, and (3) Cold War competition between the developed world and the industrially advancing communist nations of Eastern Europe. In the UK, for instance, there was extensive nationalization of mines and industrial production such as iron and steel production, major investment in state owned transportation systems and utilities, and the rollout of comprehensive national health and social welfare systems. Likewise, post war French governments nationalized and developed air transportation, banks, insurance, mines, utilities, armaments and aircraft manufacturing, and even the Renault motor vehicle operations. As early as 1949, the French state owned over one fifth of the nation's industrial capacity.

The substantial levels of economic growth that delivered widespread benefits across the Western nations also ensured that the mix of public and private involvement in economic production and management remained relatively stable in the 1950s and 1960s. Moreover, this mix became a model for economic growth and development for newly independent developing nations. Prolonged worldwide economic stagnation in the 1970s and 1980s, however, saw the capacity and role of the Western nation state attacked vigorously. Privatiza tion became a significant part of a conservative reform agenda in a conscious move to shift control and owner ship of productive resources to private sector entities. At the same time, those activities left in state control and responsibility were subject to reforms which have been described by some as "the marketization of the state." This involves the application of performance measures and operational principles that require government agencies to perform as if they were operating in a prices led market system.

Britain's privatization program was particularly aggressive following the election of a Conservative government in 1979 under the leadership of Prime Minister Margaret Thatcher (1979–90). By the time of the defeat of John Major's (1990–97) Conservative government in 1997, UK privatization sales are estimated to have totaled £65 billion and the contribution of nationalized industries to gross domestic product (GDP) estimated to have fallen from a 9% share to just 2%. Privatization was adopted for similar reasons elsewhere in Europe though at a more measured rate. In the US, where nationalized industries were historically rare compared to Europe, privatization activity during this period, especially under the leadership of President Ronald Regan (1980–88), was directed more at dimin ishing the role of the state as a macroeconomic manager and as a regulator.

The collapse of communism in Eastern Europe in the late 1990s and early 1990s saw the onset of massive privatization measures in that region. As we have seen, the rapid shift to capitalism in what are now known as the transition economies involved widespread dismantling of systems of state planned and state controlled develop ment and the sell off of SOEs. The pattern of privatiza tion in the transition economies of Eastern Europe drew heavily on what became widely known as Thatcherism.

Concerted privatization programs in the developing countries also date from this period. The initial impetus for the sell off of SOEs came from rising public in debtedness in a number of Latin American, east Asian, and sub Saharan African nations in the mid and late 1980s and, related, a desperate need to raise public finance. Shedding SOEs and a preference for market liberalization and privatization became critical parts of the International Monetary Fund's (IMF) structural adjustment programs for developing nations in need of assistance. An IMF led structural adjustment program typically involved a retreat from Keynesian based fiscal strategies in favor of fiscal austerity and balanced budgets, a reliance on liberalized markets for the allocation of productive resources and for international trade, the promotion of private property as the dominant form of ownership, and a concentration on good governance – especially through law and order, and anticorruption campaigns – rather than active economic management as the prime duty of the state. Support for market liberal ization and privatization in the developing nations came also from multilateral donor and regulatory agencies which added to the compelling nature of the liberal ization and privatization agenda.

China, as ever these days, is a special case. Privat ization of the Chinese economy commenced in the late 1970s with the drive for national economic reform. Privatization commenced first in the agricultural sector with a massive transfer of land ownership into private household management under the household responsi bility scheme, collective ownership falling from 93.5% of agricultural land in 1978 to just 7.9% in 1985. In the nonagricultural sector, privatization has been advanced primarily by the encouragement of the growth of pri vately owned small and medium sized enterprises with only limited sale to date of China's SOEs. Nevertheless, the penetration of the Chinese economy by private enterprise in the 1980s and 1990s was very significant with nonstate enterprises contributing 71.8% of industry output in 1998 compared with 21.3% in 1980. Similar transformations have occurred in the retail, transport, and construction sectors.

Impacts of Privatization

There is a substantial literature which measures and analyzes the impacts of privatization. This literature finds that privatization has, in general, led to higher rates of growth in national income and wealth creation but these have come at the expense of increased levels of social and regional inequality.

While the benefits of privatization are, arguably, widely available due to the general boost to national income that is generated, the costs of privatization are invariably borne disproportionately by certain groups and regions. This follows especially from the labor shedding that typically follows privatization programs. Regions that host SOEs or utilities targeted for privat ization thereby suffer higher than average unemployment rates especially where these regions have specialized historically in SOE or utility activity. Examples of such vulnerable regions include coal and hydro regions that have spawned electricity generating authorities, and specialist industrial cities and regions that are dominated by SOEs.

Paralleling these negative regional experiences of privatization is the impact on specialized occupational groups. Increased efficiencies from a privatized SOE or utility can result from labor intensification, the substitution of capital for labor, outsourcing to other private sector firms, and from technical change and in novation. All of these will reduce employment demand for workers with traditional, inflexible skills working in jobs designed specifically for a particular SOE or utility. These workers will find transfer to new jobs difficult especially if they are older and located in vulnerable regions.

Beyond workers and regions, assessing whether any particular nation is worse or better off following its experience of a substantial privatization program is a very difficult task given the highly complex reasons for suc cessful economic performance. Yet strong, consistent growth in the developed world, in many newly indus trialized countries, and in China and India during the 1990s and the first decade of the twenty first century, has coincided with the implementation of liberalist economic management policies, including privatization. Not surprisingly, then, the world's general experience of privatization has changed the way we think about the state and about the economy in profound ways.

The state is now a much changed entity from the immediate post war years, seen as having different re sponsibilities, significantly different operating domains and territories, and much altered capacities. These flow directly from an acceptance of a reform package with privatization at its center, including floating exchange rates; a preference for monetary policies over activist, or Keynesian, macroeconomic tools in national econo mic management; a shift to individual over collective responsibility in social and industrial settings; and, in general, a trust in market procedures rather than cen tralized planning for the distribution of productive resources and outputs.

The Political Debate over Privatization

Placing the market as the central organizing device for the organization of society derives directly from neoliberalist arguments about the value of markets as effective organ izational devices. There are various rationales for the primacy of the market and therefore for privatization in its various guises. The main one draws on arguments from the writings of Frederick Hayek (1899–1992) which advanced the view that privately run markets are capable of processing an infinite number of wishes and preferences as compared to government bodies which can only con sciously consider a handful of pieces of information relating to an issue. Hayek considered the market to be both a more efficient regulator of human endeavor as well as a more democratic arbiter for the allocation of resources to households with competing needs and wants. A provider that fails to match the standards desired by a population will go out of business, according to Hayekian thought, while all sorts of inefficient, corrupt, and undesirable practices are seen as being cultivated within government run regulatory regimes especially because of the inability of the regulator to process all relevant information.

Hayek was the prime motivator for a package of reforms – trust in markets, the adoption of privatization, the shift away from a central role for the state in eco nomic management – devised in the US in the mid 1970s through the aggressive political activity of the Chicago School led by Nobel Prize winner Milton Friedman (1912–2006). Labeled neoliberalism, the market based philosophy underpinned Thatcher's social and economic reforms in the UK in the 1980s, and then elsewhere, in the ensuing decades. Market based neoliberalism has now displaced social collectivism as the dominant mode of social organization in most Western societies. This can be seen in the expectation that individuals are primarily responsible for a household's economic well being, that privately organized production systems outperform those organized by the public sector, and that market failure itself is caused primarily by inappropriate intervention or regulation measures by the state and its agencies, an argument articulated most prominently by public choice theorists led by Chicago School economist George Stigler (1911–91).

Privatization, then, as a core component of neoliber alism, is associated strongly with the world's experience of early twenty first century economic growth and prosperity. Born from economic crisis in the 1970s, how well liberal market approaches to economic and societal management can survive or even assuage substantial economic crisis is as yet unknown. Supporters point to the efficiency of a market based system to produce rapid adjustments to economic shock and so prevent crisis. Skeptics point to the vulnerability of a society, especially its less powerful sections, to respond to shock or crisis when its collaborative resources and capacities have been largely eroded by privatization.

See also: Neoliberalism; Regulation.

Further Reading

Bakker, K. (2004). *An Uncooperative Commodity: Privatizing Water in England and Wales*. Oxford: Oxford University Press.

Chai, J. C. H. (2003). Privatization in China. In Parker, D. & Saal, D. (eds.) *International Handbook on Privatization*, pp 235 261. Northampton, MA: Edward Elgar.

Hayek, F. (1960). *The Constitution of Liberty.* Chicago, IL: University of Chicago Press.

Heynen, N., McCarthy, J., Prudham, W. S. and Robbins, P. (eds.) (2007). *Neoliberal Environments: False Promises and Unnatural Consequences*. London: Routledge.

Judt, T. (2005). *Postwar: A History of Europe since 1945*. London: William Heinemann.

Kohl, B. (2004). Privatization and regulation: A cautionary tale from Bolivia. *International Journal of Urban and Regional Research* 28, 893 908.
Mudambi, R. (ed.) (2003). *Privatization and Globalization: The Changing Role of the State in Business*. Cheltenham: Edward Elgar.
O'Neill, P. M. (1997). Bringing the qualitative state into economic geography. In Lee, R. & Wills, J. (eds.) *Geographies of Economies*, pp 290 301. London: Edward Arnold.
Parker, D. and Saal, D. (eds.) (2003). *International Handbook on Privatization*. Northampton, MA: Edward Elgar.
Starr, P. (1988). The meaning of privatization. *Yale Law and Policy Review* 6, 6 41.
Wright, V. and Perrotti, L. (eds.) (2000). *Privatization and Public Policy*, Vols. I, II, and III. Cheltenham: Edward Elgar.

Relevant Websites

http://www.privatization.org
A program of Reason Foundation devoted to providing information on privatization, government reform, contracting out, and public/private partnerships.
http://www.itcilo.org
International Training Centre.
http://www.reason.org
Reason Foundation: free minds and free markets.
http://www.worldbank.org
The World Bank.

Probabilism

R. Flowerdew, University of St. Andrews, St. Andrews, UK

Determinism and Possibilism

Since the birth of geography as an academic subject, there have been problems in arguing that physical geography and human geography really were parts of the same discipline. They were tied together, however, in regional study. Geographers would study a region by discussing its structure and relief, its climate, and its vegetation and soils, and then go on to discuss its economic activities, usually concentrating on primary industry. The justification for this approach was the idea that the physical attributes of the region were in some sense causes of the human attributes. When most people were engaged in farming, mining, or resource based manufacturing, this made some sense, although modern technology and subsidy systems may overcome a hostile environment, and tertiary and quaternary industries do not have obvious environmental controls.

Nevertheless, the idea that the environment deter mined the success or failure of human cultures was perhaps a useful balance to the intellectual life of the late nineteenth and early twentieth century. It was interesting to note that areas with similar climates often tended to have similar economies, such as the relationship between citrus fruits and Mediterranean climate (as found in California, central Chile, South Africa, south west Australia, as well as the Mediterranean region itself). Less happily, racist ideas about national character could be explained by environment – Africans were lazy because of the heat; Britain was a successful colonial power because of the 'stimulating' climate. This idea that environment determined human activities was one of the few theoretical statements generated in geography at this time. Environmental determinism (sometimes 'environmentalism') was adopted among others by Friedrich Ratzel, Ellen Semple, Ellsworth Huntington, and Griffith Taylor. Even the young Carl Sauer was heavily influenced: "Geology is indeed a subject of every day importance in Tennessee, for upon it depend the wealth or poverty, progress or stagnation, even the political convictions of its sections to an extraordinary degree."

Although possibilism is regarded as the antithesis of determinism, they have in common the idea that environment affects human actions; they just differ according to the rigidity of the effect. Possibilists see the environment as offering a set of possibilities from which human groups can choose. It was championed by the French school, led by Paul Vidal de la Blache and Jean Brunhes.

This debate had begun in the early twentieth century but was still active in the 1950s, featuring articles by A. F. Martin and Emrys Jones, although, for the first time, geographers were beginning to take an interest in how people affect the environment instead of how the environment affects people. Very influential in this respect was the publication in 1956 of *Man's Role in Changing the Face of the Earth*, a collection of essays edited by William L. Thomas.

Spate and Probabilism

O. H. K. Spate (1911–2000) was primarily a regional geographer, best known for his work on the Indian sub continent. Starting out as a historical geographer at Cambridge, he worked in Rangoon and became an expert on the geography of the Indian subcontinent. After working at Bedford College, London and the London School of Economics, in 1951 he became professor of geography in the Research School of Pacific Studies at the Australian National University in Canberra, Australia. In addition to his regional geography publi cations, Spate was interested in methodological issues, such as the relationship between geography and history and the determinism/possibilism debate. He first de scribed the concept of probabilism (originally spelt 'probablism') in a paper largely concerned with evalu ation of the work of historian Arnold Toynbee and determinist geographer Ellsworth Huntington, read to the Royal Geographical Society and published in the *Geographical Journal* in 1952.

Spate felt that "neither environmentalism nor possi bilism is adequate without large qualification." His rejection of the former (environmental determinism) seems obvious today – clearly different people have adjusted to similar environments in different ways. His rejection of possibilism is less convincing. He uses an example of the close correlation of voting maps with physical maps of the American South, which he claims shows a very strong influence of mountains on people, concluding that "possibilism simply will not do." Prob abilism is presented as a middle position. Clearly it is distinguished from determinism by the admission that cultural and social factors are important in how people adjust to their environments. It is distinguished from possibilism by the acceptance that environment may have an impact on this adjustment. The possibilist Febvre claimed "there are no necessities but everywhere possibilities." Spate argues that these possibilities are not all equally likely, and the physical environment influences how probable they are.

In 1956, Spate was invited to give a lecture in Cambridge, and he took the opportunity to address the old debate of 'possibilism versus determinism'. He argued that ignoring the question was effectively empirical description without understanding or rational explanation, and this would reduce the influence of geography on other subjects. He believed that geography "of its nature … must constantly regard both sides of the equation, man [*sic*] and land." If geography is to be a science, it must seek explanation and geography can only be distinguished from economics or sociology by geography's "preoccupation with the factor of land." Possibilists are criticized for "minimizing the land factor to such an extent that they are writing sociology, without sociological techniques."

The lecture is relatively kind to determinists. Spate cites Griffith Taylor's remark that possibilists "for the most part come from western Europe and the United States, where industrial technology is most massively deployed and has its most dazzling achievements." Those who have lived in more extreme climates, such as Australia, like Taylor and Spate himself, "might be less certain that geographical controls hardly count."

Thus Taylor, a late determinist, had argued that most of the interior of Australia was determined by the environment to be uninhabitable. It was therefore a waste of time to encourage immigration in order to populate the empty areas of the continent.

Spate points out that more recent determinists have moved to a weaker position. As Tatham argued a few years earlier, "The terms they use are more moderate, 'control' is replaced by 'influence' and 'influence' by 'response' or 'adjustment'." Spate also quotes possibilists like Tatham and Brunhes who clearly appreciate that the physical environment is important for humanity. He sees "a tendency towards the narrowing of the gap" between determinists and possibilists, and tries to define common ground which would reconcile the two schools and allow geographers to get on with the job of solid geographical interpretation.

Spate's attempt at reconciliation is based on three principles. First, "there is an irreducible minimum of influence by the physical environment which, however, acts negatively and permissively, not by positive mandate." Second, "that this is but part of the complex of causation which results in the observed facts," the cultural environment also being of great importance. Third, "that the problem is often not so much a matter of an all or nothing choice or compulsion, but a balance of probabilities."

Evaluation of Probabilism

The concept of probabilism has some appeal as a compromise between the extremes of determinism and possibilism, but it is questionable whether it is really necessary. No possibilist would deny that some possibilities are more likely to be adopted than others, or that environmental factors may influence the probabilities of choice. As Crowe pointed out in discussing Spate's original paper, probabilism is really just a variant of possibilism, dressed up to be more acceptable to determinists.

The reaction to the idea of probabilism was fairly muted, largely because the entire debate about environmental determinism was running out of steam. Advancing technology and the decline in agricultural employment made it harder to believe in environmental controls, and many major engineering projects transformed the natural environment. If anything, the possibilists won the debate, but even they put far more stress on environmental controls than geographers would do nowadays. The decline of interest in the debate was partly because of growing concern that human activity was influencing the environment to its detriment. Human effects on the environment began to seem more important to study than the environment's effect on humanity.

It could also be argued that the determinism/possibilism debate was upstaged in the late 1950s and 1960s by the spread of the quantitative revolution and the redirection of geographical research that it implied. An important theme in quantitative geography has been the use of probability models to represent geographical processes, by Les Curry, Michael Dacey, and others. It may be tempting to see this approach as linked to Spate's ideas, which could be seen as a precursor to the spread of statistical methods based on probability models. However, Spate himself was an opponent of the quantitative revolution, and early quantitative work owes much to economics and plant ecology, and little to Spate's concept of probabilism – although Curry claims to have become interested in probability through studying the relationship of rainfall to grass growth.

The type of models developed in the quantitative revolution also served to reduce the impact of the whole debate to which Spate's concept of probabilism contributed. Environment could have little effect on people if activities take place on an isotropic plain, as assumed in agricultural location and central place theory. There was a change in how geography itself was defined, away from the human/environment focus toward the study of spatial processes and relationships. Even though environmental studies have become so important subsequently, few, if any, geographers have returned to the theme of how environment constrains human activities.

See also: Possibilism; Postdevelopment; Quantitative Revolution.

Further Reading

Crowe, P. R. (1952). Comment on Spate. *Geographical Journal* 118, 426 427.

Curry, L. (2002). A random walk in *terra incognita*. In Gould, P. R. & Pitts, F. R. (eds.) *Geographical Voices: Fourteen Autobiographical Essays*, pp 81 98. Syracuse, NY: Syracuse University Press.

Jones, E. (1956). Cause and effect in human geography. *Annals of the Association of American Geographers* 46, 369 377.

Martin, A. F. (1951). The necessity for determinism. *Transactions, Institute of British Geographers* 17, 1 12.

Sauer, C. O. (1918). Geography and the gerrymander. *American Political Science Review* 12, 403 426.

Spate, O. H. K. (1952). Toynbee and Huntington: A study of determinism. *Geographical Journal* 118, 406 424.

Spate, O. H. K. (1957). How determined is possibilism? *Geographical Studies* 4, 3 12.

Spate, O. H. K. (1960). Quantity and quality in geography. *Annals of the Association of American Geographers* 50, 477 494.

Tatham, G. (1951). Environmentalism and possibilism. In Taylor, G. (ed.) *Geography in the Twentieth Century*, pp 128 162. London: Methuen.

Taylor, G. (1940). *Australia: A Study of Warm Environments and Their Effect on British Settlement*. London: Methuen.

Probability Models

R. Flowerdew, University of St Andrews, Fife, UK

Glossary

Autocorrelation correlation of one variable with itself as recorded at previous times; spatial autocorrelation occurs when the variable is correlated with its values at nearby places.

Bayesian, Frequentist and Likelihood Concepts of Probability varying attitudes to how probability is thought of. Frequentist approaches argue that the only valid basis for assigning probabilities to events is the frequency with which those events have been observed in the past. Likelihood inference assumes that data fit a probability model whose structure is known but whose parameters are unknown, and estmates the most likely parameter values to generate the observed data. Bayesians point out that we seldom analyse data that we know nothing about, but we go in with our own subjective probabilities. Bayesians say we should approach an analysis with a prior probabiility in mind; the new analysis is able to take account of this prior information, weighted appropriately, to derive a posterior probability.

Goodness of Fit a measure of how well the predictions from a model fit the observed data. For example, a high R squared value is often used as evidence that an Ordinary Least Squares regression model is successful; the deviance statistic is the usual measure.for Poisson models.

Poisson Distribution a probability distributiion applicable to count data. For a given parameter it is possible to work out the probability that the count is zero; that it's 1; that it's 2, and so on. It has a positively skewed distribution and has the property that the variance is equal to the mean. For high values of the parameter, the Poisson distribution converges to the normal.

The Quantitative Revolution

The quantitative revolution in geography transformed the discipline in the 1960s. It had several components – in addition to improved measurement techniques, a stronger critical sense in evaluating arguments, and a concern for general principles rather than special cases, there was an explicit attempt to position human geography as a science, and to adopt the goals, methods, and priorities of science. Although some human geographers, like Douglas Amedeo and Reg Golledge, attempted to formulate general laws of geography, it was more common to recognize that this process was impracticable, for the obvious reason that events in human geography were seldom if ever perfectly predictable – though some outcomes were more likely than others. The term 'law' has deterministic overtones, and most geographers might be happier talking about 'tendencies' rather than 'laws'. Probability itself is defined in different ways by different schools of thought. Bayesian, frequentist, or likelihood concepts of probability have different implications for determining the best form of analysis. In any case, many situations could be fruitfully studied using concepts of probability, which could in turn be approached using the hypothetico deductive method of scientific reasoning. An approach to geography based on development of theory, statement of hypotheses, development of suitable probability models, and testing using appropriate statistical techniques was developed.

Today most quantitative work in geography uses, explicitly or not, probability models as a way of understanding geographical reality. The statistical testing of these models is a route to evaluating their truth or credibility. Many of the biggest recent successes in quantitative geography relate to the development of new statistical methods to test these models (multilevel modeling is a good example). Much of the frustration of this approach stems from the wide gap between processes we know or suspect are important and the limited range of statistical models which often fail to reflect these processes adequately.

Among the first geographers to draw attention to probability models in geography was Les Curry in his work on climatic change and on the 'random spatial economy' and Michael Dacey in his probability models for the location of houses and settlements. David Harvey, in *Explanation in Geography*, provided a pragmatic discussion of these approaches, relating them to the then standard hypothetico deductive model of scientific enquiry. A problem is that different probability models require expertise in different aspects of mathematics, often beyond the competence of those trained as geographers. However, mathematicians, engineers, economists, and others have developed the methods to test ideas of interest in human geography.

Some of the simplest models are concerned with the probabilities of transition from one state to another. For example, somebody may or may not have a particular disease. There may be two states, health and disease, and it may be assumed, for example, that in one time period there is a probability p that an individual will make a transition from health to disease, and a probability $1 - p$ that the individual remains healthy. Likewise, there is a probability q that there will be a transition from sickness to health and a probability $1 - q$ that the individual will remain sick. It is possible to project the future incidence of the disease in this system, depending on the values of p and q.

The system can be made more complex by intro ducing other states. Thus it may be possible to move from the state of sickness to the state of death. In this case, the probability of moving from death to either health or sickness is zero. Models of this general type are used in demography and have been adapted to include systems where the states are regions and probabilities of moving from one region to another are used to investigate the long term consequences if current trends continue. If the probability of moving from one state to another is con stant in the sense that it is independent of how long an individual has been in the origin state, the system is said to have the Markov property. Such systems are called Markov chains.

Normal Probability Models

Most of the body of statistical theory adopted by geog raphers in the 1960s was based on the normal distri bution, which was the basis of the standard forms of correlation and regression analysis. A standard (Ordinary Least Squares or OLS) multiple regression analysis (perhaps the most widely used statistical tech nique) can be regarded as fitting a series of probability models. The response (or dependent) variable Y is regarded as having a normal distribution whose mean is a linear combination of certain explanatory (or independ ent) variables X_1, X_2, etc. (and whose variance is the same for each case in the data set). A series of models is fitted, usually starting with one explanatory variable and step by step incorporating other explanatory variables. A model can be stated:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$$

where Y_i is the value of the response variable for case i, X_{1i} is the value of explanatory variable X_1 for case i, β_0 and β_1 are parameters calculated in the model to opti mize goodness of fit, and ε_i is an error term whose value is regarded as being taken from a normal distribution with mean zero and standard deviation constant for all cases. For example, Y_i might be life expectancy in country i and X_{1i} might be country i's gross national product (GNP). If the model is fitted to appropriate data, it is possible to assess whether GNP could have an effect on life expectancy, to estimate the direction and approximate size of the effect, and the overall success of the model in estimating life expectancy across the data set. Note that GNP does not tell you exactly what life expectancy is in any country; it just gives an indication of the probable range of values, subject to influence from other factors not in the model, including random variation.

Having fitted a probability model which states that life expectancy follows a normal distribution around a mean equal to a linear function of GNP, further models can be fitted using other variables which may be related to life expectancy. If X_{2i} is the literacy rate of country i, X_{3i} is the level of urbanization in country i, and so on, it is possible to fit any of a set of models, differing in which of the available X variables are used, for calculating the mean of the normal distribution. Models can be assessed using a goodness of fit statistic (R^2, the coefficient of determination, is commonly used in OLS regression). This gives some insight into whether the model could reasonably be supposed to have generated the data and which of the X variables have the greatest influence on Y (controlling for the other X variables in the model).

Statistical significance tests can be used to evaluate both the model as a whole and each of the X variables. It is customary to use type 1 error (the probability of rejecting a null hypothesis that is true) as a criterion for deciding whether a particular X variable is significant or not in modeling the distribution of the Y variable. By convention, this value is set at 0.05 (sometimes 0.01) and there has been controversy over the implications of this rather arbitrary choice.

Interpretation of goodness of fit is not always very straightforward. Even a very high R^2 value does not mean that the model reveals the actual process generating the data for Y; the X variables and Y may both be influenced by some other factor. Nor does a very poor R^2 mean that the X variables have no effect on Y – a real effect may be masked by failure to include a very important X variable, or by a very large error term. Or, of course, there may be a strong but nonlinear relationship.

Issues in Probability Models

Unfortunately the normal distribution applies only to continuously distributed variables, and hence is not applicable to categorical data (appropriately modeled with binomial or multinomial probability models) or count data (appropriately modeled with the Poisson probability model). Even continuously distributed variables are not necessarily normal, but in many situations they can be transformed to approximate normality. For example, the distance decay relationship commonly found in studies of human movement can be modeled as a negative exponential function, which can in turn be transformed to normality using a logarithmic transformation. Of course, many geographical processes cannot be represented by relatively simple probability models like this, and there is plenty of scope for developing and applying a range of more complicated and realistic models.

The role of time in probability models must also be considered – attempts to test the geometrically elegant central place theory fell foul of the absence of a time dimension to the processes regarded as responsible for the geometric pattern studied by Walter Christaller and August Losch. In contrast, other geographical processes, such as spatial diffusion, had an explicit time dimension and were more amenable to probabilistic models. It is sometimes argued that the study of spatial variation cannot be viable unless time is considered too. While probability models can be formulated without reference to time, such models will not be very realistic unless a time element is included to reflect the historical devel opment of the system under study.

Another important issue is spatial dependence, where the value of a variable at one place is (in part) a function of its value at other places. This could result from a contagion or diffusion process. Even without such a process taking place, many geographical distributions are spatially autocorrelated, simply because they are affected by some other variable not included in the model. Ideally, one might attempt to remove or reduce apparent spatial dependence by including such a variable or variables. In practice, this may be difficult or impossible. From the classic work of Andrew Cliff and Keith Ord on spatial autocorrelation to more recent work on local indicators of spatial association and on geographically weighted regression, efforts have been made to develop probability models to take account of this factor.

Examples

Today the use of probability models tends to be embedded within sub fields of geography, rather than attempting to relate to the whole of human geography. Studies of population geography and medical geography offer good examples of how probability models are used today. For example, the study of spatial interaction shows an increasing sophistication in the use of probability models. The early attempts by the so called 'social physics' group and by Torsten Hagerstrand and Gunnar Olsson regarded the amount of interaction between two places as generated by a normal or lognormal distribution whose mean is a function of origin and destination size and distance (and perhaps other variables). This allows the use of OLS multiple regression as outlined earlier. In this context, it is possible to test whether incorporating the effect of distance (or, more successfully, the log of distance) on interaction improves on the model without a distance variable, assessable either in terms of the sig nificance of the relevant β parameter or the improvement in goodness of fit.

The study of human interaction was revolutionized by Alan Wilson whose work on entropy maximization changed the probability model from the normal distri bution to the multinomial distribution (approximated by the Poisson). This was then restated by Robin Flowerdew and Murray Aitkin in terms of Poisson regression analysis, allowing a similar set of probability models to be fitted as in the normal distribution case. An advantage of the Poisson model is that the Poisson mean is equal to the variance, so that goodness of fit can be calculated unambiguously without a potentially large variance term obscuring the issue. Fitting a series of probability models via the Poisson distribution usually leads, however, to such conclusions as 'the model fits better if variable X is included, but it is not good enough to be an adequate explanation of the data'. One reason why the Poisson distribution often fails to account for all the variation in the data is that it assumes individuals behave independ ently, which is clearly unlikely in many types of human interaction, such as migration. Probability models can be fitted based on relaxation of this assumption; these are likely to be modifications of the Poisson distribution such as the negative binomial model.

The binomial logistic model has also been used extensively in human geography in recent years. It is appropriate for situations where an event does or does not happen in a particular case. A common example in medical geography concerns whether or not a person has contracted a specific disease. The simplest probability model might be that each person has an equal probability of having the disease, but it is often the case that other factors may alter the probabilities. Again, there may be a set of potential explanatory variables and a series of probability models can be fitted using these variables, either singly or in appropriate combinations. In this case, the explanatory variables can be used to estimate the logistic (log odds ratio) function, and the probability of the observed data assessed, given the model results. Another important application of binomial or multi nomial models is in discrete choice theory, where somebody's choice of one alternative over one or more others is modeled as a function of a set of other variables. Logistic, Poisson, OLS, and other statistical models have been grouped together as 'generalized linear models', which have a common structure but differ according to the probability models employed.

An important extension to probability models is multilevel modeling, which is based on the idea that in some situations there may be some important influences operating at different scales – for example, the edu cational achievement of children may be modeled in terms of individual level and school level factors, and perhaps even a regional level effect. A multilevel model incorporates error terms at all these scales.

Assessment of Probability Models in Geography

It may often be thought that there are many factors affecting processes in human geography – indeed, it is a truism that most explanations in human geography are multifactorial. Further, many possible influences on geographical outcomes may be unrecorded in a data set, or even impossible to measure. This suggests that it may be impossible to construct adequate probability models. Indeed, the failure of probability models to reflect such factors (people's emotions and values, for example) was one of the main reasons for dissatisfaction with the quantitative revolution in the 1970s and 1980s.

Failures of another kind occur too. There may be strong theoretical reasons for expecting a particular model to be a significant improvement, but the data may fail to support this result. Of course, the theory may be wrong, but it is also possible that the data are insufficient to establish a true theory. In terms of classic hypothetico deductive reasoning, a null hypothesis, that the simpler model could have generated the observed data, can be rejected if the evidence is sufficient, but the alternative hypothesis, that the additional *X* variable improves the model, cannot be – though a failure to affect a goodness of fit statistic is a fairly clear signal to change tack.

The use of probability models may be frustrating in another way. Often probability theory is limited in its technical development, so a probability modeler may have to choose between a model that is imperfect in its assumptions or no model at all. It is a matter of judgment, and perhaps of the modeler's quantitative abilities, which direction to take.

In fitting models of the types discussed above, we are operating according to the standard hypothetico deductive method, with the difference that we accept that there will be counterexamples and falsifying a hypothesis depends on the number or proportion of counterexamples, not just the existence of one. Most of human geography is based on very complicated situ ations, and any theory can only match the main points of the relationships under study. The variation in the values of a variable *Y* can be split into 'explained' and 'unexplained' variation. Unexplained variation can come from three sources: omission of important explanatory variables from the model (perhaps because these vari ables have not been measured or cannot be measured); mis specification of the probability model; and random variation, whose magnitude depends on the variance of *Y*. The composition of the random variation may be thought of as a combination of chance factors and a potentially large number of minor effects which are too small to be significant in the regression model.

See also: Choice Modeling; Probabilism; Regression, Linear and Nonlinear.

Further Reading

Amedeo, D. W. and Golledge, R. G. (1975). *An Introduction to Scientific Reasoning in Geography*. New York: Wiley.
Chorley, R. J. and Haggett, P. (eds.) (1967). *Models in Geography*. London: Methuen.
Curry, L. (1964). The random spatial economy: An exploration in settlement theory. *Annals, Association of American Geographers* 54, 138 146.
Curry, L. (1967). Chance and landscape. In House, J. W. (ed.) *Northern Geographical Essays: In Honour of G.H.J. Daysh*, pp 40 55. Newcastle: Department of Geography, University of Newcastle upon Tyne.
Dacey, M. F. (1966). A county seat model for the areal pattern of an urban system. *Geographical Review* 56, 527 542.
Flowerdew, R. and Aitkin, M. (1982). A method of fitting the gravity model based on the Poisson distribution. *Journal of Regional Science* 22, 191 202.
Fotheringham, A. S. and O'Kelly, M. E. (1989). *Spatial Interaction Models: Formulations and Applications*. Dordrecht: Kluwer.
Hagerstrand, T. (1957). Migration and area. In Hannerberg, D., Hagerstrand, T. & Odeving, B. (eds.). *Migration in Sweden: A Symposium*. Lund Studies in Geography, Series B, No. 13; C. W. K. Gleerup: Lund.
Harvey, D. (1969). *Explanation in Geography*. London: Arnold.
Jones, K. (1991). The specification and estimation of multi level models. *Transactions, Institute of British Geographers* 16, 148 160.
Olsson, G. (1965). Distance and human interaction: A migration study. *Geografiska Annaler B* 47, 3 43.
Robinson, G. M. (1998). *Methods and Techniques in Human Geography*. Chichester: Wiley.
Senior, M. L. (1987). The establishment of family planning clinics in south west Nigeria by 1970: Analyses using logit and Poisson regression. *Area* 19, 237 245.
Wilson, A. G. (1970). *Entropy in Urban and Regional Modelling*. London: Pion.

Projections

F. C. Kessler, Frostburg State University, Frostburg, MD, USA

Glossary

Azimuth The horizontal direction of a line measured clockwise from a line of reference, which is usually true north.
Chronometer A portable timekeeping device capable of showing time with considerable precision and accuracy.
Developable Surface A surface that can be flattened to form a plane without compressing or tearing any part of the original surface. The three commonly used developed surfaces for map projections include a cylinder, cone, and plane.
Graticule The combination of latitude and longitude.
Isocols A symbolization method designed to graphically illustrate the amounts of angular or areal distortion on a map projection.
Map Projection The systematic transformation of the Earth's curved two-dimensional latitude and longitude values onto a planimetric surface.
Rectangular Grid System A plane coordinate system based on a map projection in order that latitude and longitude positions can be easily converted into Cartesian coordinates.
Reference Globe A conceptual model of the Earth that is reduced to a specific stated scale corresponding to the intended map scale.
Rhumb Lines A line on the Earth's surface intersecting all meridians at the same angle.
Scale Factor The ratio of an infinitesimally small distance on the map at a given point to the true distance found on the reference globe.

Introduction

Whether discussing petroglyphs, stick charts, topographic maps, or maps produced online, map projections are an essential element of any mapping effort. A map projection transforms the Earth's spherical two dimensional coordinate system called latitude and longitude to the plane or planimetric map surface. Throughout history, map projections have been developed to meet various mapping purposes that satisfy societal needs. Early purposes were philosophical in nature such as cosmography, where there was a spiritual need to map the relationship between the heavens and the Earth, or for the quest of geographic knowledge, where scholars and explorers attempted to illustrate the known world. Other societal needs emerged as sea going technology improved – better ship designs facilitated ocean going navigation and thus propelled exploration, colonization, and extended political power. Modern activities such as air travel, delineation of borders, space travel, and search and rescue operations rely upon map projections that accurately map the Earth's coordinate system. Of course, cultural differences in the way societies have viewed their geographic surroundings have led to various incarnations of projections and their applications. Moreover, societies have used projections in such a way as to advance their own ideological agendas, often with unconscionable results. Progress in both mathematics and computer technology has facilitated the development of hundreds of map projections for a variety of applications.

Earth's Coordinate System and Shape

Creating a map projection begins by considering the Earth's coordinate system. For thousands of years, scholars, scientists, and even lay people have labored to develop coordinate systems that would provide the ability to accurately locate positions on the Earth's surface. The spherical two dimensional coordinate system widely accepted today is credited to Hipparchus (2nd C. BC) who arranged lines of latitude and longitude in a systematic fashion by tying their locations to astronomical observations. The angular degree values used to describe a location's latitude and longitude are in a sexagesimal base 60 system credited to the Babylonians about 1900 BC. Collectively, latitude and longitude are often referred to as the Earth's graticule.

The ancient Greeks, Chinese, and undoubtedly other civilizations knew that the Sun's direct rays made a predictable annual progress across the Earth's surface signaling the solstices and equinoxes. These astronomical occurrences were used to delineate the first recognized lines of latitude. Other lines of latitude were arbitrarily assigned to important cities or delineated climate zones. Today, we understand lines of latitude running east to west across the Earth's surface, but we use them to measure locations north or south of the Equator (an assigned origin of 0° latitude). Since lines of latitude are parallel on the Earth's surface, they are also called parallels. A location's latitude value is the angle between a point on the Equator and another point located either north or south of the Equator. The vertex of the angle is at the Earth's center. Values of latitude range from 90° N to 90° S – the North and South Poles, respectively (see **Figure 1a**).

Lines of longitude were established differently and with far greater effort. Lines of longitude, also known as meridians, are north–south lines connecting the Earth's two poles. The term meridian comes from the noon day episode when the sun is the highest in the sky marking the mid day meridian. Since longitude is related to the duration of the Earth's rotation, computing accurate longitude location was difficult until John Harrison invented the chronometer in the 1760s. Even after Har rison's invention, longitude remained problematic in that, unlike latitude where the Equator is a natural origin, no natural origin for longitude existed. Lacking a natural origin resulted in individual countries adopting different longitude origins until an internationally agreed upon prime meridian at Greenwich, England, was adopted at the International Meridian Conference in 1884. We now accept that lines of longitude run north–south converging at the poles and measure locations east or west of the prime meridian (an assigned origin of 0° longitude). A location's longitude is the angle formed between two planes passing through both poles where one of the planes is aligned with the prime meridian and the other plane is located either east or west of the prime meridian. The vertex of the angle is at the Earth's center. Longitude values range from 0° to 180° – the prime meridian and the international dateline, respectively (see **Figure 1b**).

Various ancient societies speculated on the Earth's shape through philosophical or religious means. There was a general consensus of a spherical Earth, although the exact nature of its spherical shape was largely unknown. Other Earth shapes were set in accordance with a religious view. For instance, a Hindu Earth envisioned a hemispheric dome supported by elephants standing on the back of a turtle, while a Christian view saw the Earth as a flat disc surrounded by an all encompassing ocean called a T in O map (see **Figure 2**). In 1687, Issac Newton published *Principia Mathematica Philosophiae Naturalis* in which he presented his theory on planetary motion, centrifugal force, and how these ideas influenced the shape of a rotating planetary body. According to him, a planetary body should take on an oblate ellipsoid – flattened at the poles but bulging at the Equator. Con trary to Newton's proposition, in the early 1700s, Jean Dominique Cassini and later his son Jacques promoted the Earth's shape to be a prolate ellipsoid – flattened along the Equator and bulging at the poles. The intel lectual debate that ensued was another extension of the long standing antagonism between England and France. Not until the 1750s was the Earth shape issue laid to rest

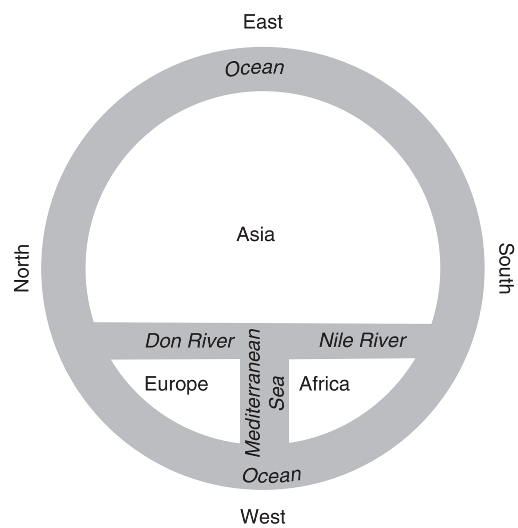

Figure 2 Stylized representation of a T-in-O map.

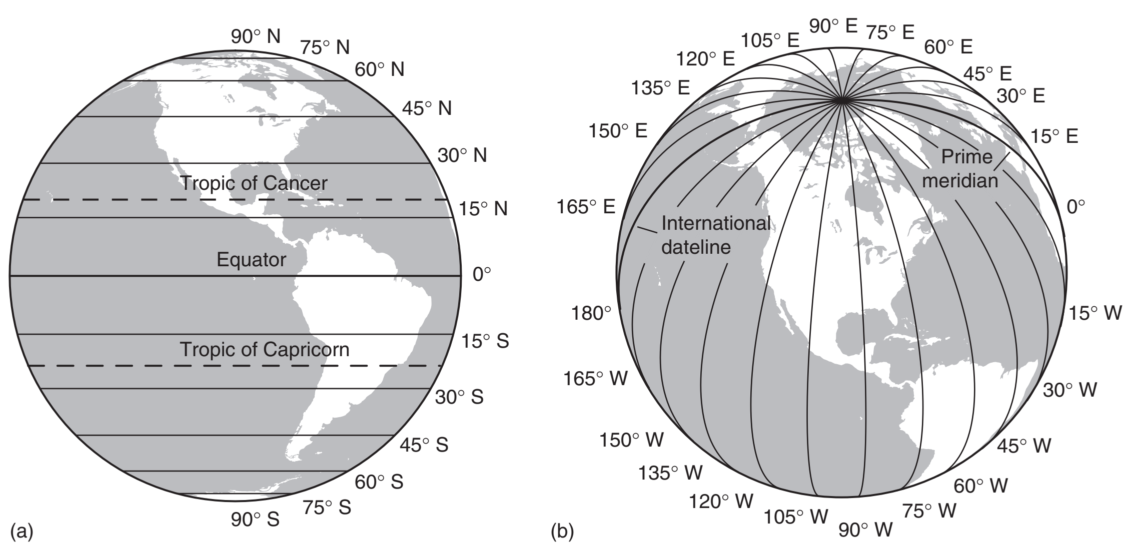

Figure 1 Lines of latitude (a) and longitude (b) on the Earth's surface.

when the results of two independent surveys measuring 1° latitude along a meridian near Peru (modern day Ecuador) and a meridian near the Arctic Circle in Lapland were reported. The surveys proved that Newton was correct. Today, satellite measurements reveal the Equator's diameter to be approximately 1/298 longer than the polar diameter. This diameter difference is significant in that spherical or ellipsoidal Earth models inherently vary in their ability to represent the Earth's shape more accurately and thus impact various mapping goals. In general terms, spherical Earth models will result in a less accurate representation of the Earth's shape, but are suitable for small scale mapping purposes, such as thematic maps of a country. On the other hand, ellipsoidal Earth models more closely reflect the obl ateness and are appropriate when more accurate repre sentations are needed, such as in topographic mapping.

Conceptual Development of Map Projection

The graticule rests upon a two dimensional spherical Earth. However, maps and the projections upon which they are based are generally thought of as existing on a two dimensional plane or a map. Thus, how the spherical graticule is projected onto a flat plane is the problem addressed and solved by projections. A map projection transforms all coordinate values on the Earth's surface (latitude, longitude, coastline, borders, etc.) onto a two dimensional plane through a systematic process, usually involving mathematical manipulation. Prior to the advent of computers, mathematical computations were carried out manually. In today's GIS environment, however, the projection process is handled quickly through trans parent mathematical equations that transform the Earth's spherical coordinate values into x and y Cartesian coordinates.

The projection process will be presented here in a conceptual framework involving four steps. The first step conceptually scales the Earth's enormous size to the same scale intended for the final map. In other words, if a map is to be shown at a scale of 1:1 000 000, then the Earth is conceptually scaled to a reference globe whose scale is also 1:1 000 000, where 1 inch on the reference globe represents 1000 000 inches on the actual Earth. The sec ond step places a developable surface in contact with the reference globe (see **Figure 3**). A developable surface is a surface such as a plane, cylinder, or cone that can be unrolled into a flat surface (e.g., a map) without any tearing or shearing. Depending on the kind of developable surface used, contact with the reference globe is made tangent at one point, tangent along one line, or secant along two lines, producing a standard point, standard line, or standard lines, respectively. No distortion is found on the map at the standard point or line(s), and the scale factor is equal to 1.0. At all other locations on the projection, the scale factor will be greater or less than 1.0 suggesting distortion is present. Knowing how the scale factor varies across a projection is important in under standing distortion patterns and ultimately for what mapping purpose the projection is suited.

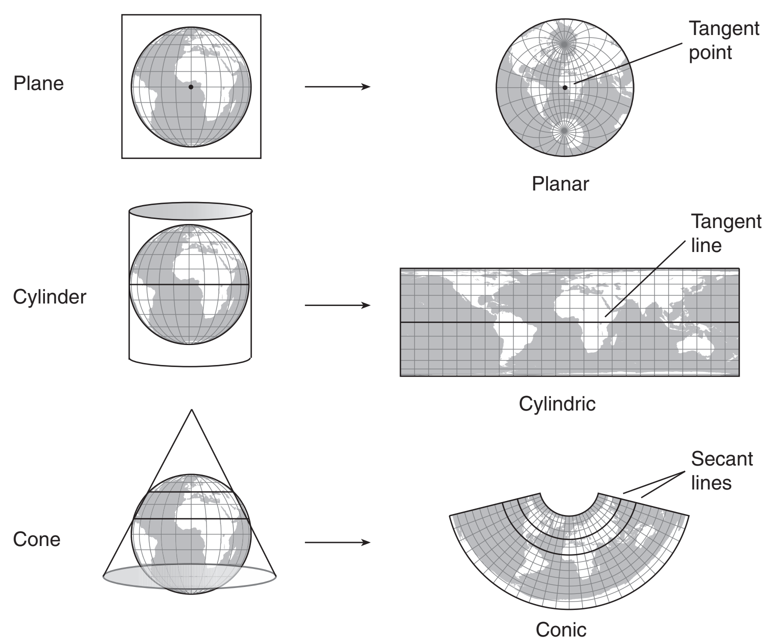

Figure 3 Positioning the plane, cylinder, and cone developable surface to the reference globe. Then, by unrolling the developable surface the planar, cylindric, and conic class of map projections are created.

How the developable surface is positioned with re spect to the reference globe results in a specific aspect of a projection. The aspect allows a specific geographic area to be brought to the projection's center. There are three general aspects of a projection: equatorial, polar, and oblique (see **Figure 4**). A projection whose aspect is equatorial has the projection's center positioned some where along the Equator, whereas a polar centered pro jection has its center placed at one of the two poles, and an obliquely centered projection has the center located somewhere between one of the poles and the Equator. Once a developable surface has been positioned in a particular aspect with respect to the reference globe, the landmasses and graticule can be projected onto the developable surface.

The third conceptual step involves 'projecting' the graticule, coastline, and any other lines on the reference globe onto the developable surface. The fourth step in volves cutting along the long axis of the cylinder or cone to unroll the developable surface, thus displaying the map. Since the planar developable surface is already flat, it does not require this fourth step. The developable surface of a plane, cylinder, or cone produces a projection with a distinctive graticule appearance that can be used to classify projections according to the graticule arrangement. Map projections developed from a plane, cylinder, and cone are commonly called planar, cylindric, and conic projections, respectively (see **Figure 3**). In the polar aspect, the planar class arranges the parallels as concentric circles centered about one of the poles while the meridians radiate outward from one of the poles as straight lines. In the equatorial aspect, the cylindric class arranges the parallels as straight parallel lines while meridians are straight, equally spaced lines that intersect the parallels at right angles. The conic class arranges the parallels as circular arcs centered about a pole, with meridians that radiate as straight lines from the pole. Other projection classes are possible that are a variation of the plane, cylinder, or cone. For example, pseudo cyindric projections represent the parallels as straight lines similar to the cylindric class, but the meridians are curved (see **Figure 5a**). Polyconic projections con ceptually involve the use of many cones and result in curved parallels similar to conic projections, but the meridians are curved rather than straight lines (see **Figure 5b**).

Map Projection Properties and Distortion

Measuring areas, angles, distances, or azimuths on maps can produce different results than if those measurements were taken directly from the Earth's surface, a con sequence of distortion introduced during the projection process. However, projections can be developed allowing specific measurements to be preserved so that when maps are used to measure, for example areas, those areal measurements will accurately reflect the corresponding areas on the Earth's surface. Projections can also be created that specifically maintain angles, distances, or azimuths.

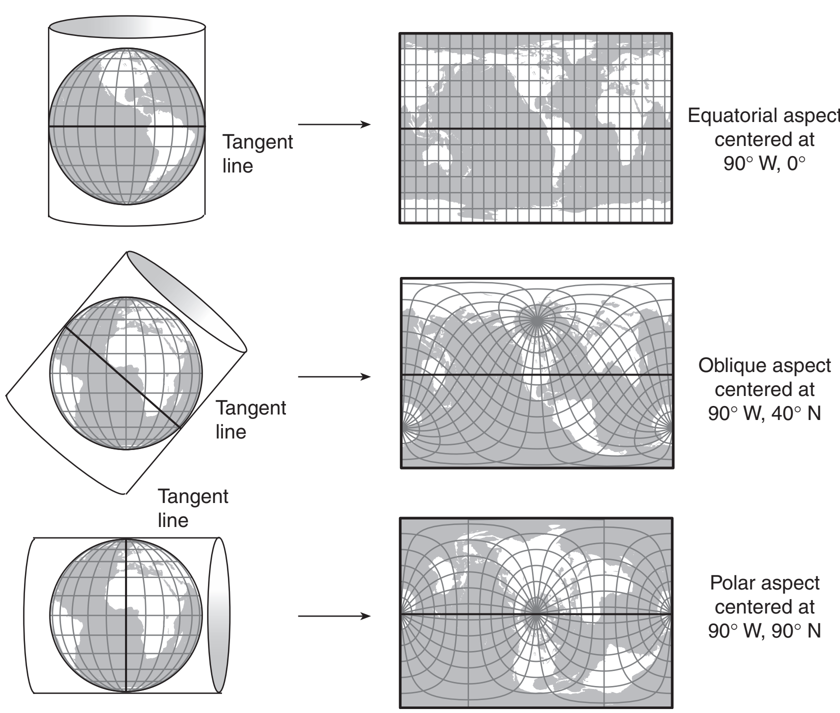

Figure 4 The equatorial, oblique, and polar aspect of a cylindrical projection.

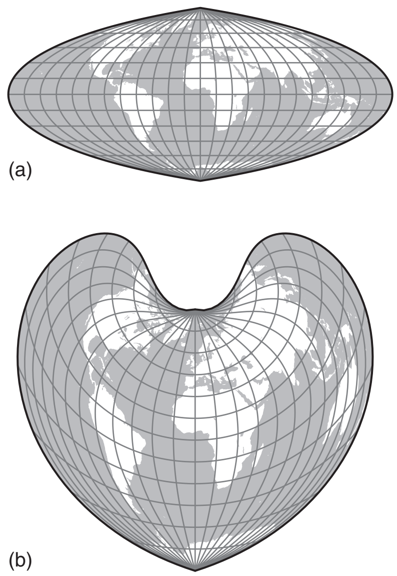

Figure 5 The authalic pseudocylindric (a) and Bonne polyconic (b) projections.

Map Projection Properties

Map projections generally preserve one spatial relationship while sacrificing all others. Preserving a spatial relationship or property for a specific map purpose is one of the fundamental reasons why several hundred projections have been developed. Map projections that maintain areas on a map in true relation to the Earth's surface are called 'equivalent'. As an example, consider a projection showing Japan. If the projection is equivalent, then the map user will be able to measure the total land area of Japan on the map and obtain 374 744 km^2, which is the correct area of Japan. Map projections that preserve angles are called 'conformal'. Assume a surveyor measures several angles on the Earth's surface as part of a boundary survey. Once the survey is complete, the surveyor may want to produce a map that preserves those angles that comprise this boundary; a conformal projection will correctly show the angles on a map. Distances and azimuths are two other spatial relationships that can be preserved on a projection, but in limited ways. Distances on the Earth can be preserved on equidistant projections, usually along all meridians or from a straight line drawn from the center of the projection to any point on the map. Calculating azimuths is a common practice used in navigational applications. As with the equidistant projection, azimuthal projections limit how true directions can be represented on a map. Azimuthal projections permit azimuths to be correctly represented as an angle from a reference line (usually true north) to a line extending from the center to some point on the map. It is mathematically impossible for distances or azimuths to be preserved throughout a projection in all directions.

Not all projections preserve a specific property. Compromise projections do not preserve any specific property, but rather reduce the amount of overall distortion compared to a projection that preserves a single property. For example, equivalent projections contain considerable angular distortion while conformal projections possess high amounts of areal distortion. Often this scenario reduces the desirability of using one of these properties for a mapping application, especially at smaller scales. In general, compromise projections result in maps that represent landmasses closer to their appearance on the reference globe than other projections where a single property is preserved.

Map Projection Distortion

Map projections manipulate scale factors across a projection to preserve a particular property. Where this property is not maintained some level of distortion will exist. In order to understand the rate at which the scale factors change on a projection and ultimately the use for which a projection is suited, distortion can be quantified and visualized. Quantifying distortion relies upon mathematics to determine the amount of distortion in areas, angles, distance, and azimuths that results from the projection process. While quantifying distortion provides one avenue of analysis, it is often more advantageous to graphically illustrate distortion on a projection using a symbolization method. One such approach is to use isocols to map the amount of angular or areal distortion across the projection. Doing so creates very distinctive distortion patterns that are usually tied to the developable surface and location of the standard point, line, or lines For example, isocols on the Behrmann cylindric equivalent projection are parallel to the standard line which, in this case, coincides with the Equator (see **Figure 6**). Another popular method of showing distortion on a projection is to utilize Tissot's indicatrix. In the 1860s, Nicolas Tissot presented his indicatrix as a quantitative and graphical assessment of how scale factors and thus distortion change across a projection. Recall, that the projection process results in a point, one line, or two lines where the scale factor is 1.0 implying no distortion is present. At all other locations on the projection the scale factors will be less or greater than 1.0, implying some level of distortion. Where the scale factor is 1.0 on a projection, the indicatrix is represented as a unit circle. At all other locations the indicatrix will take on various sizes and or shapes (see **Figure 7**). If the indicatrix appears as a circle either larger or smaller than the unit circle, the scale factor has been uniformly altered in all directions at each point, but angular relationships are maintained, as they would be on a conformal projection (see **Figure 7a**). If the Indicatrix appears as an ellipse of the same size as the unit circle, then the scale factors have been

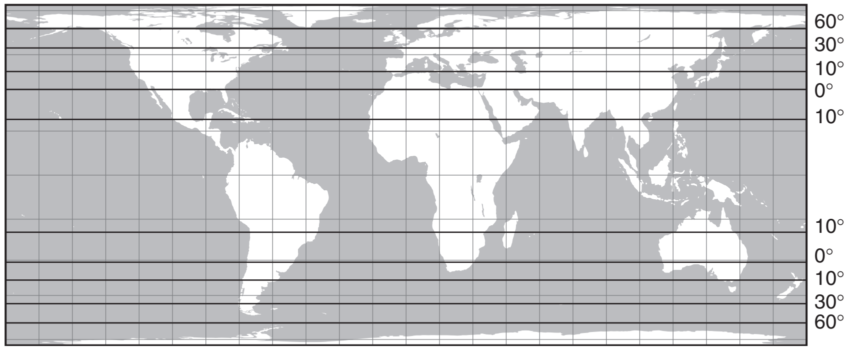

Figure 6 Isocols illustrating equal angular distortion (in degrees) on the Behrmann cylindric equivalent projection.

manipulated so as to preserve areas, and so we have an equivalent projection (see **Figure 7b**). If the indicatrix appears as an ellipse and is not the same size as the unit circle, then the scale factors have not been manipulated so as to preserve angles or areas (see **Figure 7c**).

Mathematics and Map Projections

There is a clear connection between the development of projections and the evolution of mathematics. Mathematical ideas, such as trigonometry, developed some 5000 years ago, provide the foundation for modern projections; yet, the first projections developed during the time of Thales (600 BC) and Ptolemy (CE 100) up to and including the period of the Renaissance did not explicitly rely upon mathematical equations. Rather, geometrical construction utilizing the compass, dividers, and a ruler in concert with knowledge of the characteristics of the Earth's graticule were the guiding principles used to create projections. These techniques resulted in crude representations of the graticule. Parallels were often drawn as circular arcs rather than mathematically correct curves, making it impossible to create a projection that maintained, for example, correct areal relations across the map. Familiar projections such as the equirectangular, gnomonic, orthographic, and stereographic were originally developed through such geometrical construction techniques. The corresponding mathematical equations for these projections would not be developed until after the Renaissance.

By the end of the Renaissance, geometrical construction techniques were gradually being replaced by mathematical rigor as new ideas about mathematics were being formalized. Several mathematical developments greatly facilitated the development of projections: tables of logarithms, the Cartesian coordinate system, and calculus. Tables of logarithms, credited to John Napier (1550–1617), reduced the computational burden in computing projection formulas and allowed two or three decimal place accuracy. In the 1630s, Descartes developed analytical geometry, which combined the idea of algebra (numbers) with geometry (lines and shapes) to produce the familiar x and y Cartesian coordinate system. Analytical geometry showed how to translate problems in geometry into problems in algebra by using a coordinate system to visualize the problem and the result. Calculus, developed in the early 1700s, is credited to both Issac Newton and Gottfried Leibniz. Newton synthesized many of the existing mathematical principles of his time, while Leibniz developed many of the symbols routinely used in calculus (e.g., ${}_a\int^b f(x)\,\mathrm{d}x$). Calculus has many applications in the field of map projections. For example, calculus can be used to compute distortion values on a projection by measuring infinitesimal changes in the scale factor in several different directions or is useful when the cartographer wishes to produce a map projection exhibiting a particular pattern of distortion.

In concert with these mathematical developments, new definitions of the Earth's shape were evolving along with advances in surveying techniques, which spurred the need for projections capable of mapping more accurate coordinates. In the late 1700s, John Lambert developed several projections, including the well known conformal conic (see **Figure 8a**), cylindric equivalent (see **Figure 8b**), azimuthal equivalent (see **Figure 8c**), and a lesser known equivalent conic (see **Figure 8d**). Lambert's approach to creating these projections was novel in that he was the first to use specific mathematical criteria, such as the preservation of angles and equality of area, in generating the projections. He also developed the concept for the Earth to be represented in a circle while maintaining conformality, which would later be formalized by Joseph LaGrange, whose name is credited to the projection (see **Figure 9a**). Lambert also arrived at formulas for projecting the polar aspect of the Mercator projection in a spherical context. Patrick Murdoch (1700–74) was likely the first to generate a table of distances of parallels along a meridian for the Mercator projection on an ellipsoid Earth model rather than a spherical model, which precipitated the advent of ellipsoidal mapping.

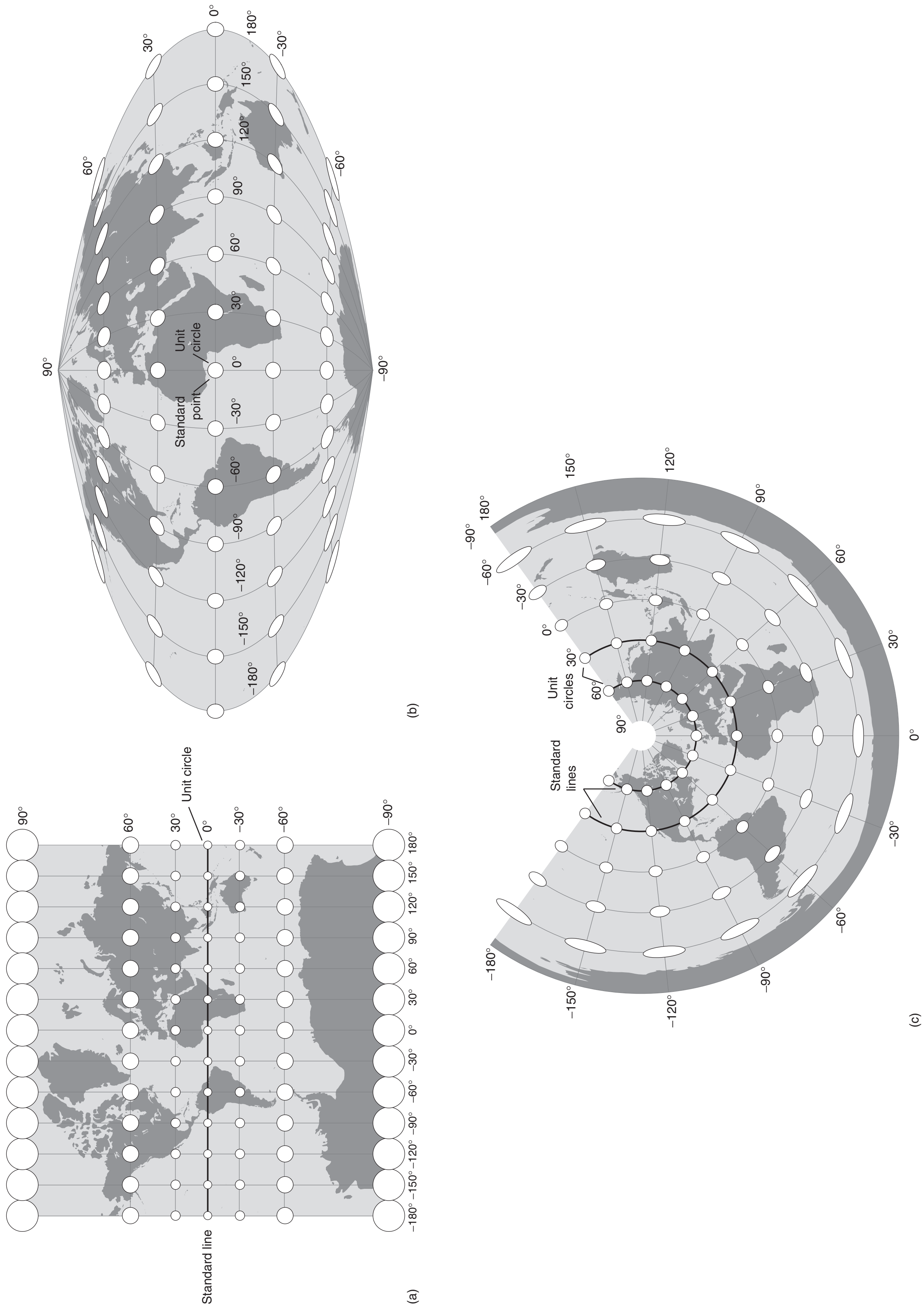

Figure 7 T ssot's nd catr x on the Mercator conforma (a), Eckert-Gre fendorff equ va ent (b), and the Eu er equ d stant (c) project ons.

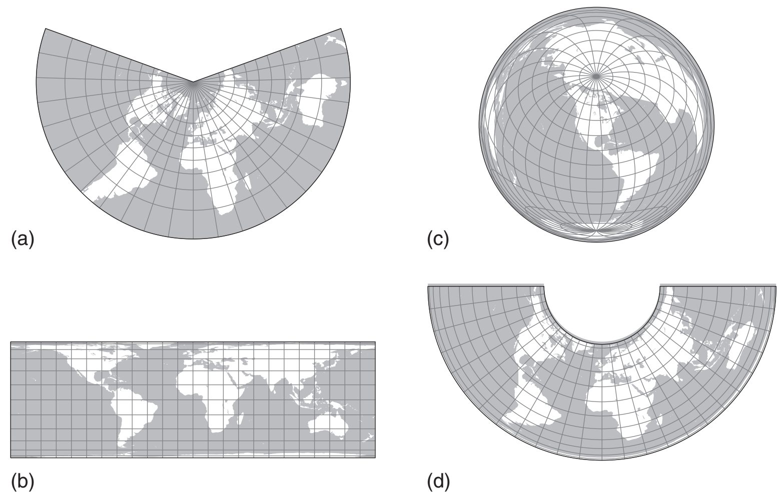

Figure 8 Four projections developed by Johann Lambert including the conformal conic (a), cylindric equivalent (b), azimuthal equivalent (c), and conic equivalent (d).

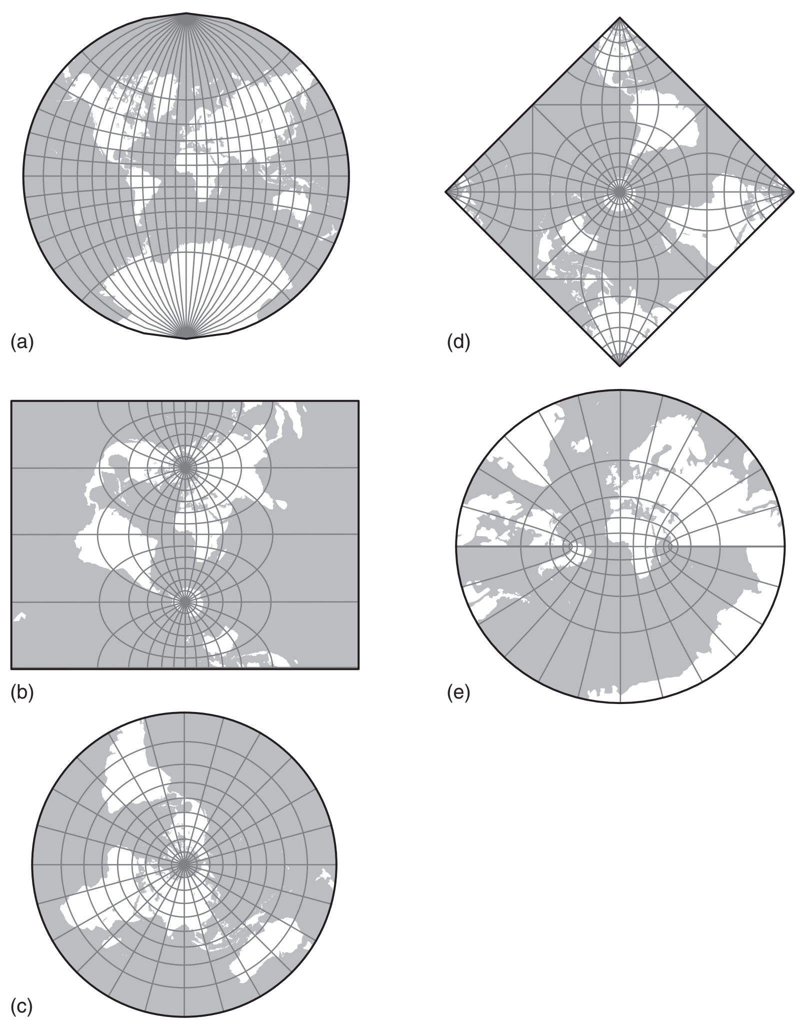

Figure 9 The LaGrange conformal (a), transverse Mercator conformal (b), Airy azimuthal (c), Pierce conformal (d), and Littrow conformal (e) projections.

Several mathematical accomplishments relating to projection development during the 1800s include: el lipsoidal/conformal mappings, least squares, and ellip tical functions. Carl Friedrich Gauss (1777–1855) described mathematically the conformal transformation of an ellipsoid model of the Earth to the plane and is credited with the term conformality. Gauss applied his ellipsoidal based conformal mapping concept to the Mercator projection creating the familiar transverse Mercator projection (see **Figure 9b**). This projection would later become the standard for many countries' large scale topographic mapping endeavors. Least squares is a mathematical principle that attempts to find, for example, a 'line of best fit' to a set of data by min imizing the sum of the squares of the differences called residuals between the fitted line and the data. George Airy (1801–92) applied this principle to minimize the sum of squares of errors in scale, which led to the de velopment of his minimum error azimuthal projection (see **Figure 9c**). Airy's minimum error projection is neither equivalent nor conformal, but rather balances the errors creating moderate distortion throughout the pro jection. Elliptical functions were also investigated during the 1800s. Historically, elliptical functions were associ ated with the problem of solving the arc length of an ellipse. The conformal projections resulting from a der ivation of elliptical functions are generally obscure and have limited applications, but some, like the Pierce (see **Figure 9c**) and Littrow (see **Figure 9d**), have found their way into navigational applications.

The field of map projections saw continued math ematical progress in the 1900s. A notable development during this time includes logarithmic projections in which scale does not change in a linear fashion but rather logarithmically. In 1957, Torsten Hägerstrand used a special tapered azimuthal projection (see **Figure 10**) to highlight the center portion of the map specifically to show migration patterns near Lund, Sweden. Other magnifying glass azimuthal projections, such as those developed by John Snyder in the 1980s, give the ap pearance of using a magnifying glass to view the Earth at a distance from outer space. The 1900s also saw the application of projection formulas to cartograms. Carto grams can be used to show demographic data (see **Fig- ure 11**); for example, the voting districts within a state can be scaled to the number of popular votes a candidate receives. Other cartograms show the distance from the

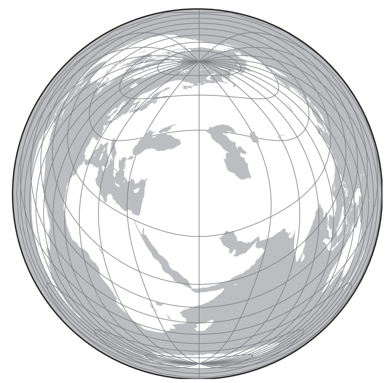

Figure 10 The Hägerstrand logarithmic projection centered on the Middle Eastern region.

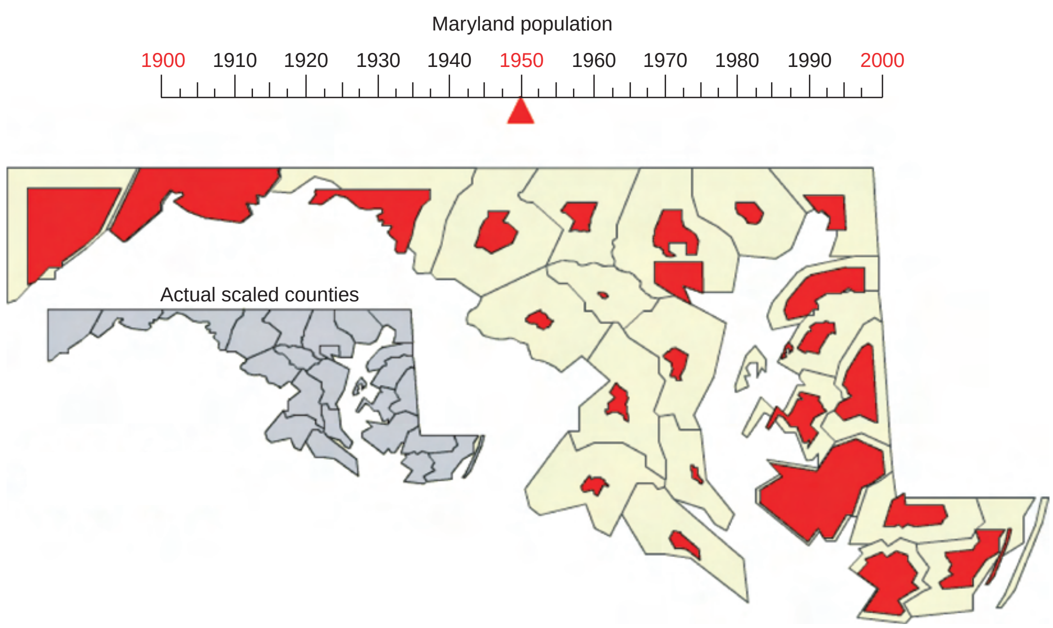

Figure 11 A cartogram of Maryland's population in 1950.

center of the map to be in proportion to some metric, such as travel time or shipping costs. Although cartograms first appeared in the mid 1800s, it was Erwin Raisz who, beginning in the 1930s, discussed the application and construction of cartograms in his *General Cartography*. Building on Raisz's initiative, in 1963, Waldo Tobler developed the mathematical basis around which area based cartograms could be constructed by adapting formulas for equivalent projections.

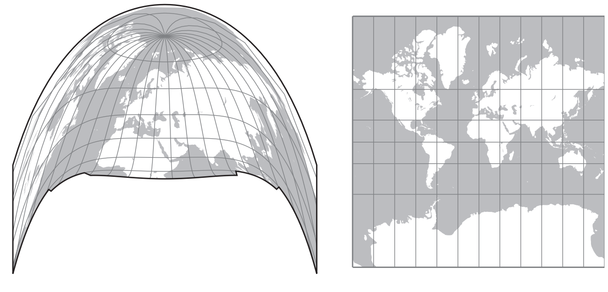

Figure 12 The Craig retroazimuthal map projection centered on Mecca 21° 25′ 19″N and 39° 49′ 34″ E (a) and the Mercator conformal projection (b).

Map Projection Applications

Throughout history, map projections have been applied to a variety of mapping purposes to suit societal needs. To begin, the proliferation of map projection development that began in Europe during the 1600s took considerable time before permeating into and being adopted by non Western societies. Thus, for purposes of discussion, projection applications in non Western and indigenous societies will be treated first, followed by Western projection applications after the 1600s.

Non-Western and Indigenous Societies' Application of Map Projections

Evidence suggests that projection applications in the non Western and indigenous societies shared common purposes with their Western counterparts: navigation, star charts, gazetteer maps, cosmography, and administrative oversight. However, implementing projections for these purposes took on different approaches according to specific societal beliefs and practices. With indigenous societies, maps were produced to supplement information that was already known about the local or broader landscape. Indigenous maps usually were not the inventories of information to which the Western world is accustomed, but presented a judicious selection of information on the map designed for a specific purpose. Interaction with the local environment played a key role in how space was organized in map form. For example, the Oceanians developed maps of their environment that were organized according to their experiences while navigating at sea. Knowledge from their mental maps of island locations, ocean currents, and swells was organized spatially into a map form known as 'stick charts'. With stick charts, known islands were symbolized by shells arranged in correct topological position based on their relationship to the constellations. The shells were tied to sticks of bamboo that represented the various ocean currents and swells. This map form was used as teaching instruments for young navigators or as reference devices examined prior to sailing.

Map projections were also developed to meet religious requirements. One interesting projection application to come out of the Islamic tradition allows Muslim worshipers to determine the Qibla – the direction to the Ka'ba in Mecca. Qibla projections have the useful property of being able to find the correct azimuth from any point on the map to the projection's center (Mecca). This property is called retroazimuthal and is the opposite of the azimuthal property, where the directions computed from the center of the projection to any point on the map are true. The Craig retroazimuthal projection (see **Figure 12a**), developed in the early 1900s, is an example of a projection in which the Qibla can be determined.

Asian societies organized space in unique ways. For example, in traditional Chinese society, the map and the landscape it represented were viewed in combination with language to communicate information about a specific place. In other respects, early Chinese maps relied upon conventions of direct and indirect measurements of ground distance and direction in attempts to draw maps to scale. The rectangular grid pattern formed a kind of projection that served as a measurement tool for distances and areas but not as a locational device. It would not be until the early 1600s that the influence of Western map projections would find their way into Chinese cartography.

Application of Map Projections after the 1600s

Map projection applications after the 1600s fall into one of five broad categories: cartometric activities, atlas mapping, topographic mapping, research, and miscellaneous investigations. Cartometric activities involve the accurate measurement of areas, angles, distances, and azimuths and rely upon projections that preserve these spatial relationships. Atlas mapping is associated with thematic maps, where the preservation of areas is important, and with general reference maps, where the appearance of the landmasses is not excessively distorted. Topographic mapping requires an ellipsoidal Earth model, as well as a conformal projection for large scale mapping. Research also provides for experimentation and usually results in projections that are not widely used or known. Miscellaneous investigations highlight map projections' versatility as being able to satisfy a variety of interesting and unique mapping purposes.

Cartometric Applications

In the 1400s, advances in ship building design and other nautical instrumentation technologies were making long distance exploration into uncharted waters more feasible. This expanding frontier greatly increased geographic knowledge and the need for maps to represent these new discoveries. In addition, ocean going commerce between nations was becoming an important activity. The portolan chart was one of the first navigational aids that enabled sailors to lay out a sailing course on a map. While these maps were generally produced without showing the graticule, contemporary observation suggests that the equirectangular projection with its square graticule was at least a likely foundation. However, the eventual watershed projection for navigation was the Mercator projection (see **Figure 12b**), developed by Gerardus Mercator in 1569, and probably one of the most recognizable projections. A conformal cylindric projection, the Mercator, was never intended to be displayed as a world map, although many atlases of the nineteenth and twentieth centuries relied heavily upon this projection for their world maps. Rather, the Mercator projection was designed to permit rhumb lines drawn on the projection to be straight lines. This characteristic was particularly useful to navigators in that they could lay out a rhumb line on the map and then follow its compass direction from the origin to destination.

Significant leaps in transportation occurred in the twentieth century, which ushered in the 'air age'. Projections were called upon that highlighted a view of the Earth from space and allowed, for instance, airline routes to be correctly displayed. The gnomonic projection (see **Figure 13a**), known for centuries, was re introduced since it shows all great circles and therefore the shortest routes as straight lines. Other projections utilized during this time included the azimuthal equidistant and orthographic. For instance, the azimuthal equidistant projection (see **Figure 13b**) shows the entire world in a circle and allows the distance from the center of the projection to any point to be accurately computed. On the other hand, Richard Edes Harrison's *Look at the World: The Fortune Atlas for World Strategy* relied heavily upon the orthographic projection (see **Figure 13c**) to show the spatial association of various landmasses around the world.

The launch of satellites and the ability to track their paths on the Earth required the creation of special projections. In the early 1970s the United States launched Landsat, a government funded satellite program. Developing a projection that displayed the ground track of the satellite's swath path at correct scale that was also conformal proved to be a problem. John Snyder, a

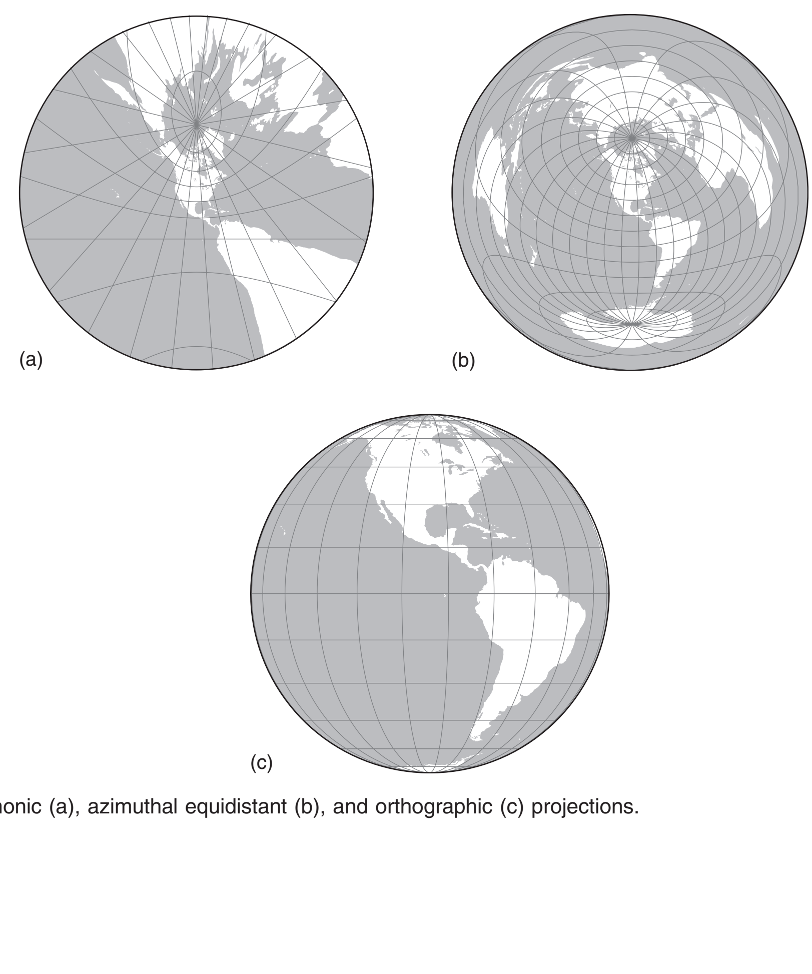

Figure 13 The gnomonic (a), azimuthal equidistant (b), and orthographic (c) projections.

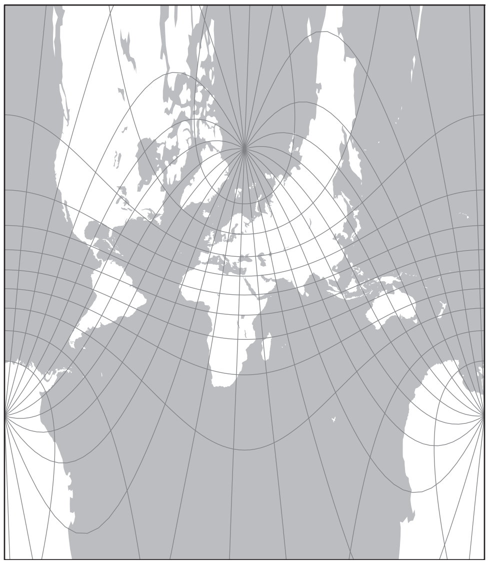

Figure 14 The satellite tracking projection developed by John Snyder.

chemical engineer, developed the space oblique Mercator projection as a solution to the problem. Later, in 1977, a cylindric projection was developed by Snyder that displayed the ground tracks of the Landsat satellite as straight lines (see **Figure 14**). Today, satellite ground tracking maps rely upon the equirectangular projection, with curved rather than straight orbital paths being shown.

Topographic Applications

Beginning in the 1700s, ellipsoidal Earth models, advancements in surveying instruments coupled with the ability to determine longitude, and the political leverage that was gained from accurate countrywide surveys necessitated that coordinates be mapped with greater accuracy than ever before. From these great surveys, the topographic map was born. Cassini de Thury developed the Cassini projection (see **Figure 15**), which was used for the official topographic map of France. The Cassini projection has a polar aspect of the plate carrée with the central meridian passing through Paris and drawn to correct scale for the ellipsoid. Other ellipsoidal based projections that today have association with large scale topographic mapping and were mathematically described during the 1700s include the Lambert conformal conic and the transverse Mercator conformal projection, which are both heavily used by various US government agencies. The transverse Mercator projection is also known as the Gauss–Krüger projection (see **Figure 9b**) in various parts of the world since Carl Friedrich Gauss (1777–1855) and Louis Krüger were responsible for articulating the mathematical principles for the ellipsoidal form of this projection. This projection is in use in over 80% of

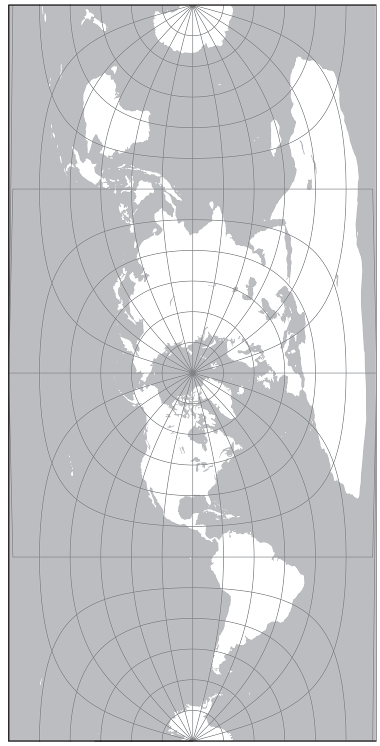

Figure 15 The Cassini equidistant projection.

the world's large scale topographic mapping programs. In the 1800s, due to its ease of construction, the equivalent Bonne projection replaced the Cassini projection for large scale topographic mapping and was subsequently adopted by numerous European countries for their topographic mapping. During the 1820s, Ferdinand Hassler of the United States Coast and Geodetic Survey recognized that a new projection was warranted during the geodetic survey of the United States and developed the American polyconic projection (see **Figure 16a**), which was subsequently employed as the projection for large scale topographic mapping by the United States Geological Survey until the 1950s.

Topographic maps will often include a rectangular grid system superimposed on top of a map, which assists in computing distances on the map. Rectangular grid systems are not new. In fact, the administration of land holdings was prominent in early societies. For instance, Roman cadastral maps implemented a rectangular grid system as a way to systematically inventory their land holdings. In modern times, the ellipsoidal version of the transverse Mercator projection is also the basis for grid systems, such as the Universal Transverse Mercator (UTM) and the British National Grid. In fact, most countries have implemented some kind of a gird system as part of their national mapping programs. For instance, the United States has adopted the ellipsoidal version of the polar aspect of the stereographic projection for the Universal Polar Stereographic (UPS) grid system of the

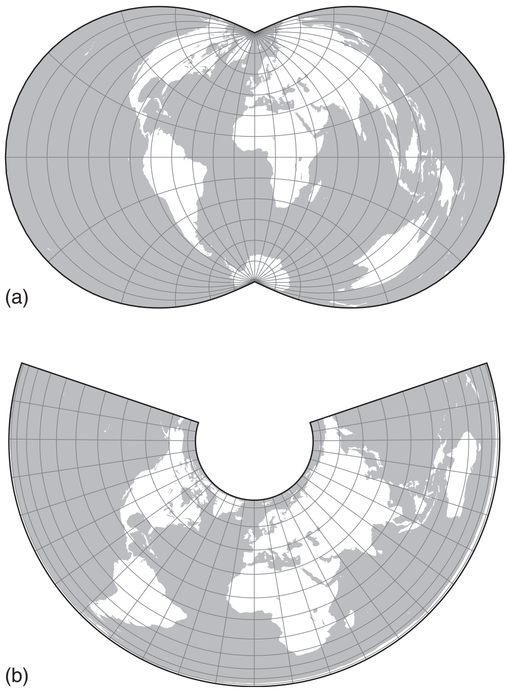

Figure 16 Hassler's American polyconic (a) and Albers equivalent (b) projections.

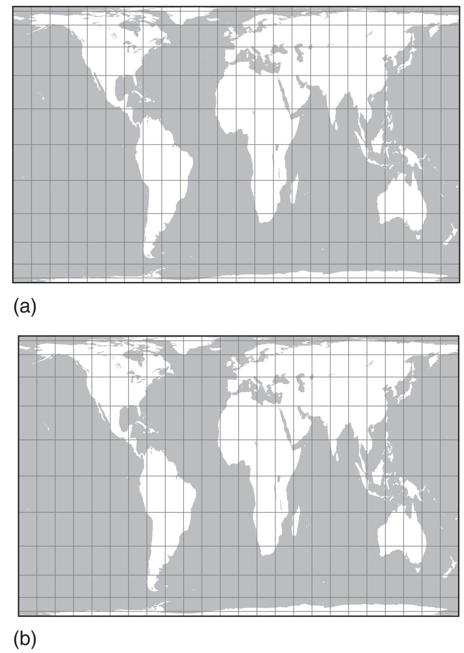

Figure 17 The Peters (a) and Gallorthographic (b) equivalent projections.

polar areas as a complement to the UTM, which extends between 84° N and 80° S. The transverse Mercator and Lambert conformal conic are two projections that serve as the foundation for the United States' State Plane Coordinate System (SPCS), a grid system designed for state level mapping purposes.

Atlas Mapping Applications

Atlases had their beginnings in the later Middle Ages and focused on mapping the known world, but it was during the 1800s that atlas mapping, with special emphasis on thematic treatments and regional depiction, began to be practiced in earnest. Prior to the 1800s, atlases utilized projections that represented the world, hemispheres, and continents. Hemisphere maps relied upon one of the several globular and planar projections. The simple conic, Bonne and sinusoidal projections were commonly applied to continental or country maps as they depicted geographic regions in correct area relations and were aesthetically pleasing with curved meridians or parallels. Two other notable projections that eventually found their way into atlases were Hassler's polyconic projection (see **Figure 16a**) and Albers equivalent conic (see **Figure 16b**). Heinrich Albers (1773–1833) developed his conic projection for a general map of Europe. In the early 1900s, Oscar Adams of the US Coast and Geodetic Survey derived the ellipsoidal form of the Albers equivalent conic projection as the standard projection for the United States with two standard parallels at 29° 30′ N and 45° 30′ N. Since its development, the Albers projection has seen significant usage in commercial atlases throughout the world.

The Mercator projection found its way into the classroom as wall maps and in atlases in the twentieth century. Objections to the extreme distortion on the Mercator projection led to a host of other projections being proposed. James Gall, Carl Braun, Alphonos van der Grinten, Osborn Miller, and Arno Peters were some of the individuals who proposed alternatives to what was seen as the considerable distortion found on the Mercator projection when used for world maps. Probably the most controversial of these projections was the Peters cylindric (see **Figure 17a**). Presented in Germany in 1974, the Peters projection is equivalent and visually resembles the Gall orthographic (see **Figure 17b**), which was developed by the Reverend James Gall in 1855. In 1983, Peters also published an atlas called *Die Neue Kartographie*, which relied on the Peters projection almost exclusively. Arno Peters championed the claim that his projection, which stretched the Equatorial region in a north–south fashion, gave a fairer representation to the less developed countries than did the Mercator, which greatly exaggerates landmasses in the mid and upper latitudes. Peters' promotion was not lost to cartographers who strongly rebuked his claims of the projection's excellence and out of their angst various publications resulted countering Peters' claim. The scholar's objections to the Peters projection were not heeded by more socially based organizations such as the Vatican and UN that have adopted the projection in some fashion. The significance of the Peters projection rests not on what agency adopted the projection, but rather that projections no longer were viewed in a mathematical vacuum or simply as a by product of mathematical quest for knowledge. Instead, projections, just as maps, can convey hidden messages and a particular point of view that might not be immediately obvious.

The twentieth century also witnessed the develop ment of many new projections designed for atlas mapping. Chief among these are the pseudocylindric projections. Max Eckert developed six pseudocylindric projections during the 1920s; of these six, the Eckert IV and VI (see **Figures 18a** and **18b**) are both equivalent and have been used considerably for atlas mapping. Another well known pseudocylindric projection was de veloped by John Goode in 1925. After experimenting with the sinusoidal and Mollweide projections, Goode fused the two, using the sinusoidal to map latitudes between $\pm 40^\circ\ 44'\ 11.8''$ N and S, and the Mollweide outside this range. The resulting Goode projection (see **Figure 18c**) is pseudocylindric, equivalent, and interrupted in order that landmasses can be shown with reduced distortion. Beginning with the fourth edition, the Goode projection became the standard projection for Rand McNally's *Goode's Atlas*. Possibly, one of the more recognizable psuedcylindric projections is the Robinson. Developed in the 1960s, Arthur Robinson was approached by Rand McNally to develop a world pro jection that was, among other criteria, uninterrupted, contained the least possible shearing, showed landmasses with a low amount of area distortion, possessed a simple graticule, and would be suitable for readers of all ages. After examining existing projections, Robinson crafted his own orthophanic (right appearing) projection (see **Figure 18d**). Instead of developing specific mathematical functions to generate the graticule, Robinson chose to geometrically construct the gratiucle to look right and thus preserve the appearance of the landmasses and configuration of the graticule as they appeared on a globe. Mathematical formulas were subsequently developed for computer based plotting of the Robinson projection, and it was later adopted by National Geo graphic Society in 1988 for use in their world maps.

Map Projection Research

Research in map projections has become an activity that has generated considerable interest in the scientific community. Some recent research activities include determining the appropriate projection for global raster based data sets, selecting appropriate projections, and designing the optimal projection. As more and more satellites are orbiting the Earth, a greater number of global raster based data sets are available. To work with these data sets in a GIS environment, a projection has to be applied. Research has suggested that, depending on the cell resolution of the raster data set, not all equivalent projections produce the same results – some distort the cells to a greater extent than other equivalent projections. Another research front focuses on assisting in selecting map projections. One of the research approaches involves developing an expert system that generates a list of recommended projections according to various input parameters supplied by the user. Another approach is to develop an interactive online wizard, such as the United States Geological Survey's Decision Support System to help select a map projection. There also have been at tempts to create optimal projections that fulfill a specific criterion. Frank Canters, Aribert Peters, and John Snyder each have embarked on their own attempts to create projections that exhibit low error. For example, Snyder utilized a technique described by Chebyshev in the 1850s

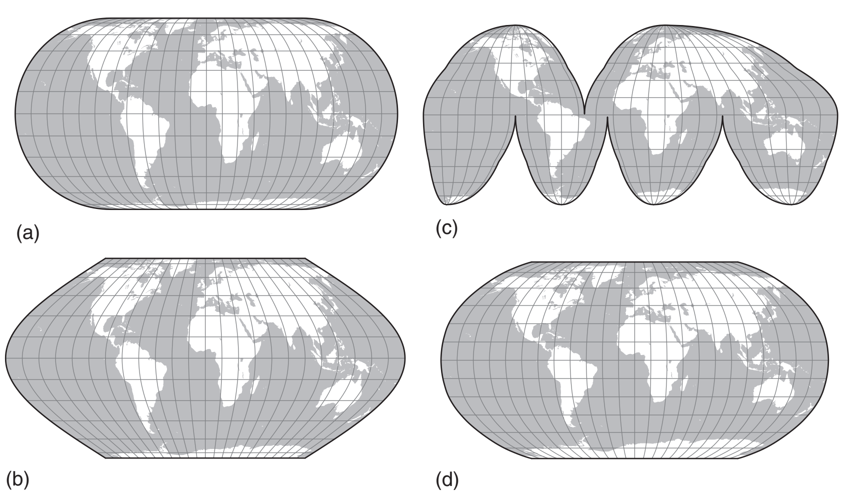

Figure 18 The Eckert IV (a) and VI (b) equivalent, Goode homolosine equivalent (c), and Robinson compromise (d) pseudocylindric projections.

where a region can be shown conformally with low error bounded by lines of equal distortion (isocols). The isocols can be circular, oval, regular polygons, or rectangular in shape. Chebyshev's technique inspired Snyder to create the GS50 projection. This is a low error conformal projection of the United States using isocols that bound the outline of the United States.

Miscellaneous Investigations

Space exploration and the subsequent mapping of planets have provided an interesting arena for the application of projections. Projecting other planetary bodies in our solar system has generally relied upon familiar conformal projections. Since planetary applications usually concentrate on topographic features, conformal projections depict round craters as round features, which are expected by most map users. In planet wide mapping series, the Mercator projection (see **Figure 12b**) is normally used for equatorial regions, the Lambert conformal conic projection (see **Figure 8a**) for the mid latitudes, and the stereographic conformal projection (see **Figure 19a**) for polar areas. For instance, NASA's Cassini–Huygens mission to Saturn has produced maps of various planetary bodies using the stereographic planar, equidistant cylindric projection. As a variation of this rule, the National Geographic Society published a global map of Mars using the Winkel Tripel projection (see **Figure 19b**), which is a compromise projection.

Map projections have also found their way into the popular realm through company logos and artistic endeavors. Global Van Lines, a worldwide moving company, incorporates a likeness of the Mollweide pseudocylindric projection as part of their logo. Various airlines such as Pan Am and Continental have included planar projections as part of their logos reinforcing the 'air age'. The American Association of Geographers adopted the Berghuas star projection showing the world in a five pointed star. Even postal organizations around the world have routinely included maps as part of postage stamps. In the early 1900s, the United State Post Office published a stamp that used the Werner projection, which is shaped like a heart and is thus appropriate for Valentine's Day letters (see **Figure 20**).

Selecting Map Projections

One of the more problematic issues when dealing with map projections is how to select one for a given mapping situation. Part of the problem stems from the fact that there are hundreds of map projections that have been developed. Moreover, selecting a map projection often involves balancing the map's purpose with several competing map projection parameters (e.g., selecting where to place the standard parallels) and characteristics (e.g., which projection property is necessary). To help in selecting a map projection, various selection guidelines have been developed in both nonautomated and automated environments. Several selection guidelines will be briefly summarized. Then, a selection guideline will be described in detail that offers some guidance on how to select a projection.

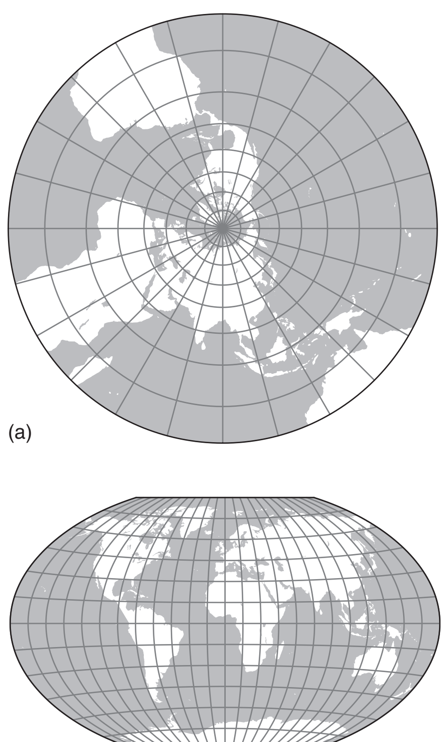

Figure 19 The stereographic conformal (a) and Winkel Triple (b) projection.

Figure 20 The Werner 'heart-shaped' projection.

Nonautomated Selection Guidelines

An elementary nonautomated selection guideline links the latitude range of the geographic area under consideration to the map projection class. In this guideline, the planar class is suitable for polar regions (60–90°), conic class is recommended for mid latitude areas (30–60°), while the cylindric class is appropriate for locations situated near the Equator (0–30°). This particular guideline places a close relationship between the latitude extent of the geographic area to be mapped and each class' ability to select the standard point or line(s) that coincides with that geographic area. For instance, Mongolia is located at approximately 50° N latitude and accordingly can be represented by a conic projection with two standard lines.

Taking a different approach, Derek Maling developed a selection guideline that included the amount and kind of distortion present on a map projection as well as the intended use of the projection. He viewed certain mapping situations associated with logical map projection choices where the choices are determined by the kinds and amounts of distortion present on the map projection. For instance, those mapping situations involving navigation require conformal projections while thematic mapping requires equivalent projections. This dichotomy is based on the concept that conformal projections, while preserving angular relations, do not maintain areal relations whereas equivalent projections, while preserving areal relations, do not maintain angular relations. All other properties and thus map uses require projections that compromise on amounts of angular and areal distortion. He frames his selection guideline as a continuum where conformal and equivalent projections are at opposite ends of and are useful for navigation/topographic and thematic maps, respectively. Other amounts and types of distortion (small areal distortions), as positioned along this continuum, are associated with a given mapping situation (e.g., atlas mapping).

Arthur Robinson and his colleagues describe a selection guideline that closely associates the projection properties with the map's purpose. The first issue they discuss is to consider the projection property. Maps that require the measurement of angular relationships such as navigational activities or recording movement (e.g., meteorological monitoring) require a conformal projection. Conformal projections which they recommend include the Mercator, transverse Mercator, and Lambert's conformal conic. Small scale general reference maps that involve the comparison of areas require equivalent projections. This requirement is based on the fact that some symbolization techniques (such as dot mapping) require equivalency in order to portray relative densities correctly. Recommended equivalent projections include Albers conic and Lambert's azimuthal. Mapping the entire world often poses the greatest challenges when choosing a projection because the amount of distortion varies widely on most world map projections. As Robinson and his colleagues mention, some projections such as the Goode's equivalent interrupted projection (a combination of the sinusoidal and Mollweide) displays the world with lower overall distortion than other noninterrupted maps. Planar projections such as the azimuthal equidistant, orthographic, and gnomonic have steadily increased in popularity since the advent of air travel as these projections show hemispheric views of the world. For instance, Google Earth has relied upon the general vertical perspective projection (an azimuthal projection) to display its maps. Robinson and his colleagues also note that some projections have been designed to meet specific purposes. For example, the space oblique Mercator projection was developed to monitor the ground track of satellites.

Automated Selection Environments

While these and other nonautomated selection guidelines are useful tools when considering a map projection, caution must be exercised, especially for those without a background knowledge in map projections. In most cases, to successfully work through a nonautomated guideline, the user is expected to have a certain level of map projection knowledge. Another problem is that these nonautomated guidelines will often recommend several candidate map projections, which leaves novice users ill equipped to select a single projection from even a small group. For instance, returning to the elementary nonautomated selection guideline discussed earlier, when considering a map projection for a mid latitude region, the conic class is recommended. However, there are dozens of conic projections from which to choose and without further guidance, the user is left with an arduous task of selecting among these conic projections. To alleviate this issue, several attempts have been made to develop automated selection guidelines where users interact with computer software (e.g., expert systems) to select a projection. Using expert systems or 'artificial intelligence', specific procedural rules can be developed that link the requirements of a mapping situation to the appropriate projection parameters and characteristics. The result of developing these procedural rules is to help users with limited projection knowledge to choose a projection that matches their specific requirements. This process is accomplished through interaction with a graphical user interface where the expert system presents the rules to the user in a transparent fashion, thus minimizing the level of projection knowledge required by the user.

There have been a handful of attempts at automating the process of selecting a map projection. Two selection guidelines that have been developed within the context of an expert system are John Snyder's EMPSS

(Expert Map Projection Selection System) and Timothy Nyerges and Pitor Jankowski's MaPKBS (Map Projection Knowledge Based System). With both systems, the selection process is guided by a series of questions that queries the user about issues such as the geographic area under consideration, map purpose, preferred pro jection properties, and final map scale. Users of both systems view general questions first. Next, based upon users' responses to these general questions, more specific questions are posed to which users respond. During the entire selection session, the more specific questions are tailored to the users' particular mapping situation. Ultimately, both systems return a single named pro jection that matches the requirements set forth by the user. While this expert systems approach to selecting map projections seems advantageous, unfortunately, neither systems were developed beyond the prototype stage.

Two other automated approaches to the selection process have also been developed. William De Genst and Frank Canters created their automated approach to selecting map projections. In their approach, they assume that the importance of selecting a projection is based upon the map's purpose and the mapping situation largely determines the features that the map projection must include. Thus, their approach interacts with a user to determine the various map projection features that should be retained on a map. Some of the important features to be considered include the outline shape of the map (e.g., rectangular), representation of the pole (e.g., a line), shape of the parallels and meridians (e.g., both lines curved), and the spacing of the parallels and meridians (e.g., both equally spaced). In addition, any special fea tures of the map such as the desired property or aspect are also considered. Once these features are obtained from the user, a specific projection is recommended. In another attempt, the USGS has created an online map projection selection guideline. This guideline begins by asking users to choose the geographic region of interest (global, continental, and regional). Once this region is identified, users are prompted other questions about the nature of the mapping situation such as desired property and type of data. An interesting aspect of this portion of the selection guideline is interaction level. For example, if users choose continental, then a world map appears permitting users to point and click with the mouse on the desired continent. Having supplied all requested input, a specific named projection is returned along with a graphic illustrating the projection's appearance.

Recommendations on How to Select a Map Projection

One of the more comprehensive and accessible guide lines was developed by John Snyder. His approach casts the selection process in a hierarchical series of questions that ultimately leads to a recommended projection. To begin, his guideline asks about the spatial location of the geographic area to be mapped. In his approach, the Earth is categorized into three geographical areas: World; hemisphere; and continent, ocean, or smaller area. Under each geographic area he provides a series of increasingly detailed questions that steps a user through the guideline ultimately leading to a recommended projection. Under the world category (see **Table 1**), the

Table 1 An adaptation of Snyder's nonautomated projection selection guideline for maps of the world

Property	*Characteristic*	*Named projection*
Conformal	Constant scale along Equator	Mercator
	Constant scale along a meridian	Transverse Mercator
	Constant scale along a great circle	Oblique Mercator
	Constant scale not needed	Lagrange, August, or Eisenlohr
Equivalent	uninterrupted	Mollweide, Eckert IV or VI, McBryde or McBryde-Thomas Boggs eumorphic, Sinusoidal, Hammer (a modified azimuthal) Other miscellaneous pseudocylindrics
	Interrupted	Any of the above except Hammer Goode's homolsine
	Oblique aspect to group continents	Briesemeister, Oblique Mollweide
Equidistant	Centered on a pole or a city	Polar or oblique azimuthal equidistant, respectively
Straight rhumb lines		Mercator
Compromise distortion		Miller cylindrical or Robinson pseudocylindric

Table 2 An adaptation of Snyder's nonautomated projection selection guideline for maps of hemisphere

Property	*Named projection*
Conformal	Stereographic (any aspect)
Equivalent	Lamberts azimuthal (any aspect)
Equidistant	Azimuthal equidistant (any aspect)
Global look	Orthographic (any aspect)

Table 3 An adaptation of Snyder's nonautomated projection selection guideline for maps of a continent ocean, or smaller area

Directional extent	*Location*	*Property*	*Named projection*
East West	Along Equator	Conformal	Mercator
		Equivalent	Cylindric equivalent
	Away from the Equator	Conformal	Lambert conformal conic
		Equivalent	Albers equivalent conic
North South	Anywhere	Conformal	Transverse Mercator
		equivalent	Transverse cylindric equivalent
Oblique	Anywhere	Conformal	Oblique Mercator
		equivalent	Oblique cylindric equivalent
Equal extent	Anywhere	Conformal	Stereographic
		equivalent	Lambert azimuthal equivalent
		Straight rhumb lines	Mercator
		Straight great circle routes	Gnomonic
		Correct scale along meridians	Center at pole: azimuthal equidistant
			Center along Equator: plate carrée
			Center away from pole or Equator: equidistant conic

next question determines the desired projection property (e.g., conformal, equivalent, equidistant, straight rhumb lines, or compromise). Once the projection property is identified, then the characteristic is chosen. With con formal projections, the characteristic determines how and in what fashion scale should change across a projection. The Mercator, transverse Mercator, and oblique Merca tor are some of the projections that satisfy this charac teristic. For equivalent projections the characteristic focuses on whether an interrupted or noninterrupted projection should be used, or if an oblique aspect is needed. Various pseudocylindric projections are offered. The characteristic for equidistant projections is where to position the projection's center. Straight rhumb lines and compromise distortion are two additional properties that can also be considered. For maps of a hemisphere (see **Table 2**), the deciding factor is the required property. Conformal, equivalent, equidistant, and global look can be satisfied by the stereographic, Lambert azimuthal equivalent, azimuthal equidistant, and ortho graphic projection, respectively. Maps showing a con tinent, ocean, or smaller geographic area (see **Table 3**) include several criteria. The first consideration looks at primary directional extent of the geographic region. For instance, regions that extend east to west include Russia or Canada, north–south include Argentina or Vietnam, while equal extents include Antarctica or the Indian Ocean. Once the directional extent of the land mass is determined, the center of the geographic area of interest is selected. Specifying the center of the geo graphic area (e.g., near or away form the Equator) defines the projection's aspect. The decision rests whether the mapping situation calls for a conformal or equivalent property. In some cases, a special property is needed such as straight rhumb lines, straight great circle routes, or correct scale along meridians.

The Future of Map Projections

Map projections solve the problem of transforming the Earth's two dimensional spherical coordinate system onto a flat surface. The utility of projections on maps and mapping software to accurately locate points on the Earth's surface cannot be understated. The quest for developing the most accurate map representation for each given application will likely continue to fuel the field of map projections. However, computer environ ments have begun to create situations in which map projections are no longer as important as they once were. For instance, virtual worlds are now possible in which the user can interact in a true three dimensional environ ment, thus removing the necessary step of projecting the Earth's spherical two dimensional coordinates onto a planimeteric surface. While this technology provides an interesting solution to handling the Earth's coordinate system in a true three dimensional environment, every day GIS users still struggle with map projection issues such as coordinate conversions, datum transformations, and re projection. These issues are not likely to vanish in the short term. In fact, map projections will continue to be an important body of knowledge for those in the mapping fields to understand for the foreseeable future.

See also: Cartography, History of; Geodesy; GIScience and Systems; Indigenous Geographies; Indigenous Knowledges; Indigenous Mapping; Mapping, Non-Western; Mapping, Topographic; Photogrammetry/Aerial Photography.

Further Reading

Bugayevskiy, L. M. and Snyder, J. P. (1995). *Map Projections, a Reference Manual*. London: Taylor and Francis.

Canters, F. and Decleir, H. (1989). *The World in Perspective: A Directory of World Map Projections*. West Sussex, UK: Wiley.

Dana, P. H. (1995). Geographer's Craft: Map Projections Overview. http://www.colorado.edu/geography/gcraft/notes/mapproj/map proj f.html (accessed April 2008).

De Genst, W. and Canters, F. (1996). Development and implementation of a procedure for automated map projection selection. *Cartography and Geographic Information Systems* 23, 145 171.

Feeman, T. G. (2002). *Portraits of the Earth: A Mathematician Looks at Maps*. Providence, RI: American Mathematical Society.

Harrison, R. E. (1969). *Look at the World: The Fortune Atlas for World Strategy*. New York: Knopf.

Lewis, M. G. (1998). *Cartographic Encounters: Perspectives on Native American Mapmaking and Map Use*. Chicago, IL: The University of Chicago Press.

Maling, D. H. (1992). *Coordinate Systems and Map Projections* (2nd edn.). Oxford: Pergamon Press.

Monmonier, M. (2004). *Rhumb Line and Map Wars: A Social History of the Mercator Projection*. Chicago, IL: University of Chicago Press.

Nyerges, T. L. and Jankowski, P. (1989). A knowledge base for map projection selection. *The American Cartographer* 16, 29 38.

Peters, A. (1983). *The New Cartography*. New York: Friendship Press.

Robinson, A. H. (1985). Arno Peters and his new cartography. *American Cartographer* 12, 103 111.

Robinson, A. H., Morrison, J. L., Muehrcke, P. C., Kimerling, A. J. and Guptil, S. C. (1995). *Elements of Cartography* (6th edn.). New York: Wiley.

Snyder, J. P. (1987). *Map Projections A Working Manual*. Washington, DC: US Geological Survey.

Snyder, J. P. (1993). *Flattening the Earth: Two Thousand Years of Map Projections*. Chicago, IL: The University of Chicago Press.

Snyder, J. P. and Voxland, P. M. (1989). *An Album of Map Projections*. Washington, DC: US Geological Survey.

Tobler, W. (2002). Qibla, and related, map projections. *Cartography and Geographic Information Science* 29, 17 23.

Van Sickle, J. (2004). *Basic GIS Coordinates*. Boca Raton, FL: CRC Press.

Wilford, J. N. (1983). *The Mapmakers*. New York: Random House.

Relevant Websites

http://carto research.er.usgs.gov/projection/index.html
Methods to Achieve Accurate Projection of Regional and Global Raster Databases, United States Geological Survey's Cartographic Research on Map Projections.

http://mcmcweb.er.usgs.gov/DSS/
United States Geological Survey's Decision Support System for Small Scale Map Projection Selection.

Protest, Rural

M. Woods, Aberystwyth University, UK

Glossary

Convergence Space A contingent and context-dependent space (which may be physical or virtual) in which various social formations may combine temporarily to achieve a greater capacity to act with respect to particular issues of shared importance.

Direct Action Participatory political activism that seeks to challenge power centers directly rather than through elected representatives, often using confrontational tactics.

Land Occupation The seizure of unproductive land, often owned by a large absentee landowner, by movements of landless people, aimed at achieving a more equitable distribution of land in developing countries.

Peasant One who lives and works on the land, following traditional, small-scale and sustainable agricultural practices. Used as a positive affirmation of identity by many rural social movements.

Policy Community A closed policymaking network involving a limited number of selected participants with a high degree of consensus and stability.

Politics of the Rural Political debate and conflict which concerns what it means to be 'rural' and the regulation and development of rural space.

Social Movement A network of individuals and/or groups engaged in political or cultural activity aimed at social and/or political transformation. 'New social movements' are characterized by loose structures, identity-driven politics, and the use of diverse tactics, including protests.

Introduction

Rural protests have become an increasingly commonplace feature of contemporary politics in many countries. The form that rural protests and campaigns take varies, as do the issues that they address and the constituencies that are mobilized. Farm incomes and changes in agricultural policy; global trade relations and regulations; hunting and animal welfare; land rights and social justice; resource management; the development of rural land; and the rationalization of rural services, have all generated protests. These seemingly disparate concerns are connected by their discursive positioning as core elements in the social construction of rurality adhered to by particular groups, such that protests are represented by their protagonists as not just about farm incomes or the right to hunt, but as a defense of a rural culture and way of life against external interference. In other circumstances, protests may reflect a more aspirational ruralism, seeking to promote an imagined rural idyll by, for example, opposing industrial development in the countryside. In both cases, contemporary rural protests in the developed world may be regarded as expressions of a new 'politics of the rural' in which the meaning and regulation of rurality itself is the core issue of debate.

Contemporary rural protests in the developed world have also engaged new participants in political activity. Indeed, although there is a historical tradition of rural protest in most countries, the twentieth century saw the rise of corporatist and insider strategies of rural political representation and policymaking, militating against rural political protest. The emergence of a new wave of rural protests at the end of the twentieth century may therefore be presented as a reaction to the failure of conventional rural politics to adequately respond to the disruption of rural restructuring. New organizations have appeared in the vanguard of rural protests, including the Countryside Alliance and Farmers for Action in Britain and the Confédération Paysanne in France, and have often adopted new political tactics and strategies, including direct action. As such, it has been argued that contemporary rural protests in the developed world might be considered as a new social movement, recognizing the underlying emphasis on rural identity, the reliance on participatory involvement, the loose and fluid organizational structures with an absence of centralized leadership, the eschewing of traditional party politics, and the conscious use of the media to convey protest messages, with some 'stunts' aimed primarily at achieving media coverage. However, this proposition remains controversial with critics pointing to the organizational and ideological incoherence of rural protest groups, their largely middle class composition, and the prominence of traditional rural elites in leadership positions.

A stronger and longer tradition of rural protests can be found in the developing world, where rural social movements such as the Movimento dos Trabalhadores Rurais Sem Terra (MST) in Brazil and the Bharatiya Kisan Union and the Karnataka State Farmers' Union in India have campaigned on issues including land rights, social justice, and resistance to externally driven agricultural 'modernization'. While these concerns may appear to be very different to those of rural protesters in the developed world, they share an underlying sense of rural identity and commitment to protecting a rural way of life. Moreover, the integration of agrarian economies in the Global North and South under globalization has promoted a common agenda for rural campaigners in contesting the liberalization of agricultural trade, expressed through the formation of an international farmers' coalition, Via Campesina, and protests at meetings of the WTO.

Also, there are notable geographical differences in the extent, nature, and significance of rural protests within both the developed and developing worlds. Rural protests have, for example, achieved a far higher profile in Europe in recent years than in North America or Australia. Similarly, there are marked differences in the social and political forms of rural social movements in Africa compared with Latin America and Asia. Understanding these variations, their causes, and their consequences is one way in which geographers have a particular contribution to make to the analysis of rural protests and social movements. Geographical analysis can also illuminate understanding of the spatial strategies implicit in rural protests, from blockades and land occupations to the organizational strategies involved in mobilizing a dispersed rural population to the transposition of rural protests to external urban and international locations. Finally, geographers have an interest in rural protests because they revolve around an essentially geographical struggle to resist homogenization and integration, to protect local difference, and to defend distinctive forms of social, economic, and cultural activity that are perceived to constitute a rural identity associated with the particular, if ill defined, territory of 'the countryside'.

Historical Rural Protests

Rural protests have a long history. In preindustrial Europe, rural areas not only accommodated the majority of the population, but also were the areas of weakest state control. Extreme poverty and the authoritarian rule of local landowners also contributed to tensions that sparked rebellions such as the Peasants Revolt in fourteenth century England and the jacqueries uprising in fourteenth century France. Additionally, rural areas attracted radical campaigners such as the Levellers and Diggers in seventeenth century England, who challenged existing property and political structures and sought to build new utopian settlements. While these protests appear to have little in common with contemporary rural protests they are, significantly, often appealed to by modern protesters who seek to gain legitimacy by positioning their actions as part of a long historical tradition.

Outside Europe, rebellions and protests in rural regions of Asia, Africa, and Latin America formed part of the struggle against colonialism, and in particular, resistance to the forcible appropriation of land, labor, and resources for imperial economic activity. In North America, farmers' uprisings that had started as anticolonial agitation became interventions in the political development of the early United States, defending settlement rights and opposing perceived excessive taxation through protests including Shays's Rebellion of 1786 and the Whisky Insurrection of the 1790s. A strain of rural radicalism continued in America into the early twentieth century, branching on the one side into the ill fated populist and agrarian socialist movements, and on the other into vigilantism, but with both sides combining electoral politics with public protests and occasional direct action.

Industrialization provoked a new wave of rural protests, including both antitechnology movements such as the English Luddities and working class mobilizations such as the Australian shearers' strike in 1891. Yet, industrialization also contributed to the petering out of rural protests in much of Europe, North America, Australia, and New Zealand. De population reduced the potential pool of protesters, while the rise of resource capitalism promoted a new political settlement for rural regions that militated against protests as a political strategy. The new political settlement rested on three core elements. First, capitalist agriculture (or other forms of resource capitalism such as forestry) was positioned at the core of the rural economy and society. Farmers were encouraged to think of themselves as business owners, and workers were encouraged to identify with the interests of agriculture as an economic sector. Second, agricultural exceptionalism meant that agriculture was given preferential treatment in government policy, including price supports, subsidies, and protectionist trade policies. Third, rural (or, more particularly agricultural) interests were given privileged access to government through close knit policy communities that incorporated the major farm unions and, in some countries, rural based political parties that exploited coalition politics to monopolize agricultural ministries – most notably the Country Party in Australia (later renamed the National Party). The smooth functioning of the system generally meant that rural protests became unnecessary, occurring only occasionally either to reinforce the negotiating position of farm unions, or as the result of conflict between rival farm unions (as in France).

Modern Farm Protests

The political stability described in the section above held as long as insider groups such as farm unions and rural political parties continued to deliver results to rural constituents. By the late twentieth century, however, the political settlement was coming under pressure from several directions and its ability to deliver became compromised. This happened in the United States in the 1970s when the agricultural policy community weakened just as farmers faced a mounting debt crisis. Disaffected farmers formed the American Agricultural Movement, a network of loose cells with no formal leadership that organized a farm strike and two 'tractorcade' protests in Washington, DC. The first tractorcade, in January 1978, attracted 3000 farmers to the capital, while the second, in February 1979, caused traffic congestion with a 40 km long line of tractors. In Australia, high interest rates and weakened farm union influence with a newly elected Labor government prompted 25 000 farmers to partici pate in a demonstration in Canberra in July 1985; while in New Zealand, farmers protested against the radical de regulation of agriculture proposed by another newly elected Labour government, despite support for the policy by some in the Farmers Federation. In all three cases the farmers' protests ultimately proved to be un successful and subsequent policy reforms and structural adjustments have reduced the capacity for rural protest in these three countries by drastically altering the structure of the agricultural sector.

European farmers have arguably been more successful in resisting, or at least delaying, attempts to reform the Common Agricultural Policy (CAP). Early reforms in 1982 were abandoned after 100 000 French farmers marched through Paris in protest undermining the French government position, and farmers in Britain and elsewhere protested against the introduction of 'milk quotas' in 1984. Increasing farmer militancy in Europe has also been associated with disaffection with estab lished farm unions. In France, discontent in the early 1990s with the handling of renewed CAP reform efforts by the Fédération Nationale des Syndicats d'Exploitants Agricoles (FNSEA) union led to defections to breakaway groups including Co ordination Rurale and the Con fédération Paysanne, which adopted more confronta tional direct action tactics such as road blocks. In Britain, growing farmer discontent at the handling of the bovine spongiform encephalopathy (BSE) crisis in 1996 and falling farm incomes in 1997 provoked a series of protests by dissident farmers including port blockades, town center rallies, and pickets at supermarkets and food pro cessing plants. Over 100 protests are calculated to have been organized by farmers in Britain between summer 1996 and March 2000. The loose network of activists which had organized these protests formalized themselves as a new organization, Farmers for Action, in May 2000. In September 2000, Farmers for Action played a key role in the blockade of fuel depots across Britain in protest at increased fuel taxes, copying farmer led fuel protests that had already been successful in Belgium, France, Ger many, Italy, Norway, the Netherlands, Spain, and Sweden.

Rural Restructuring and the Politics of the Rural

Structural adjustment in agriculture has been just one dimension of a wider process of social and economic restructuring in rural regions of the developed world over the last quarter of a century. This has encompassed economic restructuring, including a shift in economic significance and employment from traditional primary industries such as farming and forestry to consumption based economic activities such as service industries and tourism; in migration to rural areas as part of a trend of counterurbanization, and associated patterns of second home buying, gentrification, and rising property prices; growing environmentalism and increased regulations to protect rural landscapes, plants, and animals; and the increased mobility and prosperity of rural residents producing new lifestyle patterns that have, combined with neoliberal economic policies, stimulated the ra tionalization of rural service provision. These processes of rural restructuring have not only disrupted traditional rural political structures, creating opportunities for new political groups and practices to emerge, but have also undermined the discursive stability of the 'countryside' as an imagined space. It has hence been suggested that the late twentieth century witnessed a transition from 'rural politics' – concerned with the management of rural in dustries and land – to a 'politics of the rural', in which the meaning and regulation of rurality itself is the primary focus of conflict and debate.

One of the first people to observe the consequences of rural restructuring for the politics of rural areas was a Belgian rural sociologist, Marc Mormont. Writing in the 1980s, Mormont argued that rural spaces are now the focus of competing ideas about their form, function, and development, giving rise to new 'rural struggles' in which emergent rural social movements fight to promote their particular vision of rurality. Mormont cited as examples of such rural struggles conflicts over a tourism develop ment and over a village school, thus directing attention to the local scale. In the two decades since Mormont's art icle, local rural conflicts of this type have become com monplace in rural localities across the developed world. They range from very locally focused campaigns against specific building developments or the closure of a local post office, to more significant protests against major developments such as new roads, airports, and quarries or to defend certain cultural practices. Local rural protests are usually grassroots led, engage new participants in political activity, and employ a range of different tactics including rallies and demonstrations, petitions and letter writing campaigns, interventions in local elections and, on occasion, direct action. Often, local rural protests have been initiated by in migrants seeking to defend their financial and emotional investment in the rural idyll, yet it is erroneous to reduce rural conflicts to local incomer tensions, and rural protests can mobilize participants from all sectors of the rural population, for a variety of motivations.

Many local campaigns have been successful in achieving their goals, but many have been frustrated in their attempts to affect real change due to the limited power of rural local government. This has led to the banding together of pluri local coalitions and the refocusing of attention at the regional or national scale. The rescaling of the politics of the rural at the national level has also been prompted in some countries by leg islation that has been perceived to threaten rural com munities or cultures, including notably the National Competition Policy in Australia (which de regulated elements of agricultural trade and service infrastructure provision), and legislation on hunting in Britain and France.

Indeed, hunting has sparked some of the most pas sionate and high profile rural protests of recent years, most notably in Britain. It is also interesting because unlike many rural protests it concerns not a physical site, or a set of economic interests, but the perceived right of rural people to participate in an activity that they regard as central to their culture and way of life. In Britain, ‘hunting’ commonly refers to the pursuit of wild mam mals by hounds and has been the focus of political conflict for several decades. The election of a Labour government in 1997, however, was considered by hunting supporters to pose a significance threat to their sport. They responded by forming a new pressure group, the Countryside Alliance, from three smaller pro hunting organizations. Between 1997 and 2005 the Country side Alliance organized three major demonstrations in London – the Countryside Rally (July 1997, with 120 000 participants), Countryside March (March 1998, 250 000 participants), and Liberty and Livelihood March (September 2002, 408 000 participants) – as well as a series of smaller long distance marches, regional rallies, demonstrations outside parliament and party confer ences, town center events, and publicity stunts. Yet ten sions also developed within the Countryside Alliance over strategy and particularly the balance between pro tests and conventional political lobbying, leading to a number of breakaway factions. These splinter groups, including the Countryside Action Network, the Rural Rebels and the ‘Real Countryside Alliance’, tended toward more direct action tactics, including road blocks and the appropriation of prominent landmarks to convey pro hunting messages. However, while these protests undoubtedly delayed hunting legislation, they failed to prevent a ban on hunting with hounds being introduced in February 2005.

Hunting has also been the focus of rural protests in a number of other European countries, as well as in some parts of the United States, albeit with a lower profile than in Britain. In France, proposals to limit the hunting season for migratory birds prompted both demonstrations and the formation of a new political party, Chasse, Pêche, Nature et Tradition (CPNT), which polled 12% of the vote in the 1999 European elections. In Belgium, the establishment of a Flemish Environmental Network of protected wetlands in which hunting would be prohibited led to a protest march by 20 000 hunting supporters in Gent in May 2003, organized by Platform Buitengebeid (‘Countryside Platform’) under the slogan “March for a Liveable Countryside.”

Significantly, several of the larger rural campaign groups have explicitly made connections between dif ferent issues affecting rural areas. The Countryside Alliance, for example, was deliberately named to associ ate the core issue of hunting with wider concerns about the future of the British countryside and has over time developed a broader campaigning platform, including agricultural issues, affordable housing, and rural service provision.

Rural Protests in the Developing World

As the trajectory of rural protests in the Global North has been shaped by eras of industrialization and rural restructuring, so the trajectory of rural protests in the Global South has been shaped by colonialism, de colonization, political struggle and revolution, modern ization and uneven development. Rural social movements emerged in the 1960s and 1970s, building on the heritage of rural resistance to colonialism to mobilize protests opposing externally driven agricultural ‘modernization’ programs that threatened peasant livelihoods, or re pressive state action against rural communities (or sometimes both). More recently, rural protests have challenged agrarian and economic reforms associated with neoliberal policies, including those imposed on developing countries by global institutions such as the World Bank. Some commentators suggest that these contemporary protests represent a ‘new wave’ of rural movements, characterized by a mixed social base of rural peasants and urban proletarians, leadership from ‘peasant intellectuals’, an ‘antipolitical’ strategy and use of direct action tactics, an internationalist vision and the ideo logical fusing of Marxism and ethnic politics. Others, however, contend that this model is subscribed to as an ideal that finds only partial actual articulation in rural social movements.

Contemporary rural protests in the developing world have commonly addressed the interlinked issues of land reform, agricultural modernization, and political and social justice. The MST in Brazil, for example, founded in 1984 to campaign for a more equal distribution of land, expressly positions itself as part of a wider struggle for social and political reform. The MST has adopted a range of tactics, including a national protest march in 1997, but its primary tactic has been land occupations – appropriating unproductive land for the settlement of landless families. Between 1985 and 2000, the MST averaged 345 land occupations a year, claiming 25 598 ha of land and resettling over half a million families. Land occupations have been a feature of rural protests in a number of other developing countries, including Ecuador, Malawi, Paraguay, the Philippines, South Africa, and Zimbabwe, often with state support or complicity. Elsewhere, protests over land rights have opposed the sequestration of property from peasant farmers, including resistance to urbanization in China, where it has been reported that an average of 230 rural protests a day were held in 2005, mostly against land seizures.

Peasant movements have also been active in resisting neoliberal agrarian reforms, opposing the spread of genetic modification technologies sponsored by global agri food corporations, and promoting alternative, sustainable, forms of agriculture. The Karnataka State Farmer's Union in India, for example, successfully opposed the promotion of genetically modified (GM) cottonseeds and field trials by Cargill and Monsanto. Moreover, in some regions, peasant protests have fused with wider political struggles, most famously in the Chiapas region of Mexico, where the Zapatista movement (EZLN) has mobilized a peasant base in an armed uprising that has combined land and agrarian questions with indigenous rights and calls for regional autonomy.

Rural protests in the developing world have therefore addressed a range of issues, and have often championed concerns that are not necessarily exclusively rural, yet they share a common emphasis on peasant identity and a commitment to the social, economic, and environmental sustainability of rural communities. Groups such as the Bharatiya Kisan Union in India have staged protests over a range of interconnected rural issues, including electricity supply, sanitation and education, as well as agricultural prices, presented as part of a rural–urban struggle.

Globalization and Rural Protest

Rural protests in the Global North and the Global South may have developed along different trajectories, but globalization means that rural campaigners in the North and South are increasingly facing shared, or at least, interlinked, challenges. Globalization can spark rural protests in a number of ways. Increased global mobility of migrants, for example, forms a background to protests by migrant rural workers (such as the farmworkers movement in the United States), as well as more reactionary protests by rural residents against illegal immigration or asylum processing centers. Equally, the global dissemination of 'universal' environmental and animal welfare values has informed protests over resource exploitation, animal husbandry, and hunting, often involving international pressure groups. Conflicts such as protests against the commercial logging of 'old growth' forest in British Columbia have achieved global publicity and support through these networks.

It is, however, economic globalization and particularly the liberalization of global trade in agricultural commodities that has become a key focus for rural protests in both the Global North and South. Farmer groups in Mexico, for example, combined to protest against the North American Free Trade Agreement in 2003 under the slogan "El Campo no Aguanta Más" (The Countryside Can't Take It Anymore). Anti free trade protests have also been staged by farmers in Bolivia, Colombia, Ecuador, South Korea, and Thailand.

The expansion of free trade areas, and concern at imports, is equally a focus for rural protests in the Global North. Vehicles carrying food imports have been targeted by farm protesters in Britain and France; in Poland, farm groups have mounted roadblocks in protest at imports and at accession to the European Union (EU); while the 'Resurrecting Rural and Regional Australia' campaign in 2004 united various rural protest groups in campaigning against agricultural de regulation, a free trade agreement with the United States, and foreign takeovers of companies. Yet, unlike rural movements in the developing world, these protests have tended to be insular and protectionist and have not looked toward global solidarity.

A notable exception is the Confédération Paysanne in France. The Confédération Paysanne has been a key protagonist in rural protests since the 1980s, but achieved global attention in August 1999, when José Bové led a protest dismantling a McDonald's restaurant in Millau, sparked by increased US import tariffs on Roquefort cheese imposed in retaliation to an EU ban on US hormone treated beef. Bové has targeted McDonalds' as a symbol of globalization and in his trial in June 2000 he further confronted globalization, calling as witnesses rural campaigners from Poland, the Philippines, Senegal, Brazil, and the United States. A free antiglobalization festival, attended by an estimated 100 000 supporters outside the court, reinforced the message.

The desire to build global solidarity between rural protest movements led to the formation of Via Campesia in 1993, an international farmers' coalition that now represents over 140 organizations in over 60 countries. Significantly, these include not only the MST, the Bharatiya Kisan Union, and other peasant movements in Latin America, Africa, and Asia, but radical European farmers unions including the Confédération Paysanne, as well as the National Farmers Union of Canada and the National Family Farm Coalition in the United States. Via Campesina has organized farmers' protests at meetings of the WTO, including at Cancun in September 2003 where Korean farmers' leader Lee Kyoung hae publicly committed suicide to highlight the effect of trade liber alization on Korean rural communities. The demon strations at WTO meetings, along with other events such as transnational 'caravans' of farmers from Latin America and Asia through Europe and North America, form 'convergence spaces' through which the disparate inter ests of rural groups can be melded into a singular protest and the barriers of space and scale can be overcome.

See also: Citizenship and Governmentality, Rural; Peasant Agriculture; Resistance; Social Movements.

Further Reading

Bové, J. and Dufour, F. (2001). *The World Is Not for Sale: Farmers against Junk Food*. London: Verso.
Desmarais, A. A. (2008). The power of peasants: Reflections on the meanings of La Vía Campesina. *Journal of Rural Studies* 24, 138 149.
Dibden, J. and Cocklin, C. (2007). Contesting competition: Governance and farmer resistance in Australia. In Cheshire, L., Higgins, V. & Lawrence, G. (eds.) *Rural Governance: International Perspectives*, pp 175 190. Abingdon: Routledge.
Mormont, M. (1987). The emergence of rural struggles and their ideological effects. *International Journal of Urban and Regional Research* 7, 559 578.
Moyo, S. and Yeros, P. (eds.) (2005). *Reclaiming the Land: The Resurgence of Rural Movements in Africa, Asia and Latin America*. London: Zed Books.
Naylor, E. L. (1994). Unionism, peasant protest and the reform of French agriculture. *Journal of Rural Studies* 10, 263 273.
Routledge, P. (2003). Convergence space: Process geographies of grassroots globalization networks. *Transactions of the Institute of British Geographers* 28, 333 349.
Stock, C. M. (1996). *Rural Radicals: Righteous Rage in the American Grain*. Ithaca: Cornell University Press.
Woods, M. (2003). Deconstructing rural protest: The emergence of a new social movement. *Journal of Rural Studies* 19, 309 325.
Woods, M. (2005). *Contesting Rurality: Politics in the British Countryside*. Aldershot: Ashgate.

Relevant Websites

www.confederationpaysanne.fr
Confédération Paysanne.
www.countryside alliance.org
Countryside Alliance.
www.mst.org.br/mst/home.php
Movimento dos Trabalhadores Rurais Sem Terra (MST).
www.viacampesina.org
Via Campesina.

Psychoanalysis

P. T. Kingsbury, Simon Fraser University, Burnaby, BC, Canada

Glossary

Analysand The patient in psychoanalysis.

Ego The part of the mind that concerns the sense of self and mediates between the unconscious and conscious, as well as the Id, Superego, and social reality.

Hermeneutics A branch of knowledge that concerns the art, science, and philosophy of interpretation.

Id An unconscious part of the mind that is associated with immediate gratification and the primary instinctual impulses of aggression and sexuality.

Inductive Reasoning The process through which general conclusions are reached through the evaluation of particular events and observations.

Method A systematic means or particular procedure that is used to collect data and is guided by methodology.

Methodology The systematic theoretical analysis of methods, as well as the study of the directions and implications of empirical practices, techniques, and the research process.

Superego The part of the mind associated with moral agency, self-critical conscience, as well as social standards and imperatives.

Transference The redirection to a substitute, typically the analyst, of former infantile feelings toward a significant person – typically one's parents.

Unconscious, The A partially inaccessible and dynamic part of the mind that mainly consists of repressed thoughts, injunctions, and feelings that profoundly influence our conscious lives.

Introduction

Psychoanalysis is characterized by ongoing critical reassessments of its clinical methods. Although psycho analysis is composed of different and often contrary theorizations about the human mind, all psychoanalytic 'schools' are united by a reliance on the methods of talking, listening, and interpretation. Much of the literature on psychoanalytic methods is underpinned by enduring debates about psychoanalysis's relation ship to science. On the one hand, Sigmund Freud (1856–1939), who founded psychoanalysis during the late nineteenth century, always strived to establish psychoanalysis as a scientific project. On the other hand, Carl Jung (1875–1961), who in 1913, famously ushered in the first major break from the Freudian paradigm, was unconcerned about whether psycho analysis was scientific or not. Beyond psychoanalytic circles, numerous scholars working in the sciences, social sciences, and humanities have passionately lauded or vehemently condemned psychoanalytic methods vis à vis questions about scientific validity.

While debates about the (un)scientific status of psychoanalysis are obviously bound up with different understandings of science and psychoanalysis, the fol lowing questions have been regularly raised ever since the early twentieth century: First, to what extent can psychoanalytic objects, processes, and hypotheses such as the unconscious, libido, repression, transference, the death drive, and the Oedipus complex be empirically observable, testable, and causally explained? Second, how can psychoanalytic concepts become usefully and rigor ously integrated into non psychoanalytic methodological practices? Third, to what extent can clinically based and individually centered psychoanalytic methods be used for the investigation of social contexts by social scientists and texts by scholars working in the humanities?

In response to the above questions, some commen tators have argued that psychoanalysis is nothing but a glorified sectarian dalliance composed of fairy tales that demand absurd, that is, unscientific conceptual leaps of faith. Other commentators, however, dismiss such con demnations because of their single mindedness, lack of rigor, and narrow understandings of what constitutes science, scientific inquiry, and psychoanalysis. To be sure, whenever psychoanalysts make theoretical propositions about psychical life, they are often at pains to avoid falling into the traps of reductive reasoning and totalizing knowledge claims. Whenever one reads a psychoanalytic text (especially Freud's), one is sure to encounter as much doubt and hesitation as certainty and assuredness. For the most part, psychoanalysis relies on inductive reasoning whereby general theories of the mind are de rived from and based on particular and repeated empir ical observations in clinical settings.

Given the psychoanalytic goal of explaining seem ingly 'irrational' behavior, the alleged idiosyncratic even 'crackpot' psychoanalytic concepts such as the unconscious, the death drive, and Oedipus complex are used not so much as an attempt to faithfully capture or mirror psychical 'reality', but rather as a conceptual strategy in which to explain or take into account the iterations of enigmatic and hitherto fore inexplicable psychical phenomena. It is on the question of how to explain the psychical dimensions of seemingly irrational social and spatial behavior that human geographers increasingly began to turn to psychoanalytic theories during the mid 1990s. As a result, psychoanalytic geographies have become an important subfield in human geography. And yet, extensive critical assessments of the validity, value, and potentiality of psychoanalytic methods in human geography are rare. Put simply, geographers have found psychoanalytic concepts much more valuable than psychoanalytic methods. In addition, the use of psychoanalytic methods in geography usually involves adding psychoanalytic concepts to already existing methods and methodologies.

There are at least three main reasons for this situation: First, psychoanalysis has been frequently viewed by a significant number of geographers with a cautiousness that borders on suspicion that even borders on paranoia. Geographers' paradigmatic commitments to social constructionism, historicism, and cultural studies have meant that we are easily disturbed by the allegedly a cultural and a historical psychoanalytic concepts such as the unconscious, the drive, and desire. While many geographers are correct to assert that these concepts cannot be entirely reduced to the effects of discourse, the social, history, culture, and biology, most geographers find the theoretical use of an abstract psychoanalytic concept much more palatable and legitimate than the prospect of methodologically enacting psychoanalytic concepts and methods in the lived spaces of research. Second, because geographers have, for the most part, selected concepts and literatures that conform to and emerge from the above dominant paradigms in human geography, much of the radicality and thus the very points of psychoanalytic epistemologies and ontologies have been elided. Thus psychoanalysis is widely viewed in geography as both an extravagant and redundant theoretical detour: psychoanalysis spends too much theoretical time on problems that could be more easily tackled with other, thriftier theoretical frameworks. And so, geographers have parried or simply foreclosed discussions about the commensurability of psychoanalytic methods with existing methods in human geography. Third (and related to the first reason), the methodological tenets shared by psychoanalytic and psychotherapeutic practices have convinced some geographers, most notably Stuart Oliver, that psychoanalytic methods are simply best confined to clinical rather than social settings.

Despite this lack of engagement and uncertainty about psychoanalytic methods in human geography, there are a number of ostensibly psychoanalytic methods in geography. Before turning to geographers' rendering of psychoanalytic methods, let us first briefly elucidate some key psychoanalytic methods.

Key Psychoanalytic Methods

In a psychoanalytic session, it is the patient, rather than the analyst, who does most of the work. This is why in psychoanalysis, the patient is often referred to as an 'analysand' – the gerund term highlights the active role of the person under therapy. Here, 'work' refers to analyst's job of facilitating the analysand's task of discovering and 'working through' unconscious material that has, for example, been repressed (expelled from consciousness into the unconscious), foreclosed (expelled from consciousness and the unconscious), or disavowed (unconsciously denied yet acknowledged). During the early years of psychoanalysis, Freud noted how difficult it was for analysands – despite their best intentions and efforts – to recuperate the most important and painful unconscious memories. Freud's notion of 'resistance' refers to these various struggles and blockages. According to Freud, resistance mainly consisted of defense mechanisms that included rationalization, repression, regression, displacement, and reaction formation. The 'defenses' helped to protect the conscious mind from aspects of reality deemed unacceptable or dangerous. In order to overcome resistance and access unconscious material, Freud originally used the method of hypnosis and the power of suggestion. Freud, however, soon abandoned the method of hypnosis because it was unpredictable and not many patients were able to enter a hypnotic state. Freud soon turned his attention on fine tuning numerous methods, including three methods that have become extremely important for nearly all practicing psychoanalysts. It is to the key psychoanalytic methods of free association, transference, and analytic listening that we now turn.

Free Association

Having abandoned hypnosis, Freud began to develop the method he called 'free association' that basically involves encouraging analysands to say whatever is in or comes into their minds no matter how silly, embarrassing, or confusing those thoughts may be. One of the main reasons analysand's lie on comfortable couches that face away from the analyst is to facilitate the process of free association. The primary goal of free association is to build a substantial amount of material in a way that unfolds an associative network of the analysand's unconscious. Freud also used free association as a method to enable the analysand to recall or accidentally come across a crucial, perhaps even pivotal for the analytic treatment, forgotten memory. The main point of free association, however, was not to discover or unearth the analysand's 'big secret', but rather, to elucidate in speech and over a long period of time the analysand's psychical conflicts that kept at bay certain memories and wishes from the conscious mind.

Transference

Much of psychoanalysis relies heavily on the 'cathartic method of treatment', that is, the task to make patients feel the emotions of previously repressed or forgotten early childhood memories. Now, all this free associating talk about infantile memories means that the relationship between the analyst and analysand is quite a peculiar one. A common outcome of this relationship involves analysands' reenactment of intense infantile emotions, demands, experiences, and identifications in the analytic session by projecting them onto the analyst. Freud called this process 'transference'. When Freud first encountered transference, he initially categorized it as another form of resistance. Later, Freud regarded transference as a unique way to facilitate analytic progress. For example, the analysand may begin to transfer powerful feelings such as love and hate onto the analyst that they once directed toward their mother or father. Through transference, the analyst becomes a substitute for the analysand's original parent. Thus, Freud believed that transference usefully revealed the analysand's history (especially their childhood experiences) because it involved the repetition of previously learnt behavior. Freud also encountered the process of countertransference wherein the analyst may transfer their unconscious feelings back onto the analysand. Notably, there is much debate in psychoanalytic literature about the extent to which transference and countertransference aid or restrict psychoanalytic treatment.

Analytic Listening

The French philosopher Paul Ricoeur (1913–2005) once likened psychoanalysis to a 'hermeneutics of suspicion'. Ricoeur's comment brings to the fore how psychoanalysis is preoccupied with developing interpretative techniques that discern meaning in subtexts, disguises, and multiplicities. Exemplary here is the method of analytic listening. For Freud, analytic listening consisted of taking a position that adopted a meditative state of 'evenly hovering attention'. Put differently, analytic listening is open and attentive to new observations and alternative meanings in what the analysand says. In order to achieve this openness, analysts had to avoid the following: attributing a single meaning to the analysand's speech; succumbing to countertransferential suggestion; holding onto particular memories of what the anaysand had said previously; and cultivating desire and judgment when listening to the analysand. Furthermore, analytic listening requires that the analyst listens in an engaged manner and is attentive to the numerous psychical dimensions and registers of the analysand's speech. When an analysand speaks, analysts are trained to listen to not just one voice, but to many voices, that is, to the polyvocality of speech. Here, psychoanalysis attends to three hermeneutic 'horizons' or temporalities: the time of the analysand's discourse, the time of the analytic session, and the time of the patient's life. For Anna Freud (1895–1982), analytic listening meant that the analyst position themselves 'equidistant between Ego, Id, and Superego'.

Obviously, given the psychoanalytic premise that the unconscious exists and has effects, analysts themselves will always import unconscious judgments, biases, desires, etc. Methodological conundrums and inevitabilities such as these are central to the vast psychoanalytic literature on methods and one of the reasons why Freud famously dubbed psychoanalysis 'the impossible profession'.

Psychoanalytic Methods in Geography

Psychoanalytic geographers have yet to use the key psychoanalytic methods of free association, transference, and analytic listening. As we shall see, however, the methods that *are* used by psychoanalytic geographers import many of the methodological commitments outlined above. To be sure, much of critical human geographical research is replete with processes that are comparable to free association, transference, and analytic listening. Think, for example, of the free association like meandering pronunciations and spontaneous ramblings voiced during unstructured interviews. Think, for example, of the powerful transferential bonds that often develop between the researcher(s) and researched during interviews, ethnographies, and focus groups. Think, for example, of the degree to which geographers, especially feminist and poststructuralist geographers, attempt to foster data collection techniques and conditions that are comparable to analytic listening: attendance to self reflexivity, mindful of multiple viewpoints and meanings, recognition of the situatedness and partiality of knowledge, vigilance toward power dynamics, and appreciative of the ineluctable instability of insider/outsider and researched/researcher distinctions.

It is important to recognize that psychoanalysis and its methods have at once significantly and subtly influenced how researchers in social sciences such as geography conduct and conceptualize research processes and subjects. How so? Put simply, it is difficult to underestimate the impact psychoanalysis has had on the development of the critical and social theories that infuse so much of human geography's methodologies. It is perhaps more accurate to state, then, that geographers have yet to systematically or explicitly use the aforementioned key psychoanalytic methods. Geographers' tentativeness toward psychoanalytic methods is neither an inevitable outcome nor has it anything to do with the intrinsic properties of psychoanalytic methods. In the discipline of sociology, for example, free association has been used to explore the emotional dynamics of small social groups such as families. Geographers' tentativeness toward psychoanalytic methods is best understood as the result of a more general hesitation toward psychoanalytic epistemologies and ontologies. Such tentativeness is marked by the inverted commas that often surround psychoanalytic concepts in geography such as the unconscious, and, as we shall see, in some of the following methods that geographers have adopted under the aegis of psychoanalytic geographical research.

The following section is not exhaustive, but rather representative of the most recent and extensive examples of how geographers have engaged with psychoanalytic methods. The first subsection exemplifies how geographers have used psychoanalysis to theoretically evaluate current methods and methodologies in geography. The next subsection illustrates how a psychoanalytic method is concretely used in a research setting. The subsequent subsection reveals how psychoanalytic theory has been used to inform interpretative and explanatory frameworks that try and make sense of complex socio spatial processes.

Empathy and Identification

To date, Liz Bondi's work on empathy and identification is one of the most incisive and extensive theoretical assessments of the potential benefits of using psychoanalytic methods in geography. Bondi asserts that psychoanalytic understandings of identification and empathy provide 'conceptual resources' that can enhance feminist understandings of fieldwork, specifically, questions about how positionality and power infuse the dynamics between the researcher and the researched. It is important to note that Bondi does not wish to forge a distinctive form of therapeutic or psychoanalytic fieldwork, but rather, to explicate some psychoanalytic insights that can theoretically enhance already existing research methods.

Bondi primarily draws on an 'object relations' psychoanalytic approach. Briefly, proponents of object relations such as Melanie Klein (1882–1960), Michael Balint (1896–1970), and British psychoanalysts, including Donald W. Winnicott (1896–1971), Ronald Fairbairn (1889–1964), and Harry Guntrip (1901–1975) emphasize how (following Freud's notions of an 'object' and 'object relation') the psyche's interactive and intersubjective world is constituted by and through relationships 'and' objects. From this perspective, 'relations' are to be understood as interrelationships wherein the psyche not only constitutes its objects, but is also constituted by, composed of, and relates to itself through objects. 'Object' refers to how something, typically a person, body part, or a character role can become the medium, procurer, and/or locus of the drive. Very briefly, the drive is a concept that refers to how human sexuality is not entirely reducible to a biological instinct or a cultural construction. An object can be a 'part' or 'whole' object, a 'good' or 'bad' object, and an 'internal' or 'external' object. Many object relations theorists focus on how much of people's lives involve the forging of and dependence on psycho social bonds, that is, interrelationships with other people. In so doing, object relations theorists (especially the British psychoanalysts as opposed to the 'Kleinians') de emphasize the role of the drive by considering psychical life not so much in terms of the accomplishment of pleasure, that is, the reduction of psychical tension, but rather, in terms of the unconscious representations of objects.

For Bondi, object relations theories, with their focus on the 'richly peopled' space of the psyche, can enhance understandings of the interpersonal and dynamic environment of fieldwork because they provide a rich vocabulary in which to engage the topics of power and positionality in research practices. In addition, for Bondi, the concepts of identification and empathy can provide researchers with unique theoretical guidance on how to tackle the tricky balancing act of understanding and making distinctions between the interviewees' feelings and one's own feelings.

How so? Put simply, Bondi argues that research practices are replete with the processes of identification and empathy. The former term (following Melanie Klein) consists of, on the one hand, the projection or unconscious expulsion of aggressive impulses and 'bad' objects onto other people and, on the other hand, unconscious processes of introjection, that is, the internalization of 'good' objects. Empathy primarily involves the capacity to understand the emotional condition of another person through oneself, especially through acts of listening. Bondi addresses how the interview process (like many other social interactions) is structured by conscious and unconscious divisions and exchanges of labor exemplified by the process of one person talking while another listens. Researchers, then, are often tested on their ability to receive, process, and then give back to the researched unconscious material. As Bondi notes, the process of empathy is quite unlike identification because through empathy the researcher offers, rather than expels, unconscious material in a way that can be easily handled and 'owned' by the interviewee. Given the extent to which interviews can involve oscillatory processes of observation and participation, as well as intensely affective exchanges of anger, confusion, and sadness, Bondi makes a convincing case that the concept of empathy is extremely useful to reflect on, if not orient, the very practices of qualitative interviews.

'Playing'

Beginning with the object relations thesis that adults' sensations, perceptions, and experiences of landscape are profoundly influenced by their infantile experiences of the Self with the Other (mother/primary caregiver), Amanda Bingley's (an experienced practicing psycho therapist) research tackles the key psychoanalytic ques tion of how to access and interpret elusive unconscious processes. For Bingley, any methodological gap pre venting the import of practices of therapy into academic research is ultimately a negligible one. In order to tackle the question of the relationship between the unconscious and adults' experiences of landscape, Bingley enlists two concepts from Winnicott: the first, 'potential space' or 'intermediate space', is a space through which meditative, fluid, transitional, and exploratory demarcations of Self and Other (e.g., person, people, landscape, object, etc.) take place during infancy. Such a space is replete with the interconnective capabilities of 'objects' (see above). The second concept is the 'transitional object' which typically includes the breast, food, a toy, and blanket. Such tran sitional objects are mediators and travelers between permeable inner and outer psychical worlds which, through acts of 'playing' (inverted commas in original), help to establish a potential space. In addition, transi tional objects usher in 'unintegrated' states (between the capture of reality and threat of integration) for their users, meditate between fantasy and the real world, and become more sophisticated as we grow older. For Win nicott, potential space creatively infuses existing and possible cultural worlds of art, music, literature, and stories.

Bingley's research involves the creation and maxi mization of the conditions of a potential space and states of 'unintegration'. How is this achieved? Bingley's research involves practical workshops composed of 12 adult participants. Key here is how Bingley's research is informed by Jacquie Burgess's use of 'small group' ana lyses that draws on the work of S.H. Foulkes (1898–1976). Notably, Bingley uses a wide range of methods of in cluding in depth interviews, verbal feedback, and artistic three dimensional modeling in group workshops as a way in which to address sensory experiences from the in formants' earliest childhood to the present day. Bingley's main research method, however, is the art therapy and Jungian technique of sandplay. Through sandplay, research participants are able to become part of a potential space. Sandplay, which involves participants playing (often with their eyes closed) with wet and dry sand with their hands, feet, and tools such as small buckets, brings to the fore tactile rather than visual ex periences of the landscape. Thus, research participants can 'get in touch with touch', and, in so doing, attend to not only an overlooked sense, but also their over looked emotions. In this way, sand acts as a transitional object because it is a blank canvass or screen onto and through which participants can materially enact alter native psychical and sensory registers. Sandplay opened up new sensory landscapes for the participants and prompted numerous childhood memories and hitherto fore unarticulated associations that straddled a psycho phenomenology of being 'in here' and 'out there', identifications with sand figures, precipitation of free association.

Bingley notes that much of the entire project worked through what Winnicott calls a 'facilitating environment' or 'holding environment': a space that thrives on and fosters trust, empathy, and knowledge building. For example, Bingley traveled to potential participants' homes in order to ensure they felt secure and thus hopefully willing to be involved in the research. In addition, like the analyst, Bingley did not seek to impress her judgments or even observations on the results of the participants' sandplay. By creating a facilitating space and 'holding' the research group, Bingley not only brings to our attention the relevance of various psychoanalytic concepts for methodological discussions, as well as our questions about society and space more generally, she also provides one of the most focused methodological accounts of how to conduct psychoanalytic methods in geography.

'Tracking' and 'Mapping'

The final two psychoanalytic methods discussed in this article, 'tracking' and 'mapping' (inverted commas in the originals), are exemplary of how geographers adopt a psychoanalytic framework in order to interpret socio spatial phenomena. Mapping and tracking, used by Heidi Nast and Steve Pile respectively, are both used to elab orate socio spatial analyses, rather than subject based research.

In *Real Cities*, Steve Pile toys with our assumptions about what is 'real' in urban life. Pile's central thesis (following the sociologist Robert Park) is that a city's state of mind or personality is as important, that is, as real as its built environment. Inspired by the rebellious methods of the Situationist International and several contemporary novelists, Pile aims to reveal why urban imaginations, fantasies, and emotions matter because they are thoroughly material and political. Pile's notion of the real city brings to the fore how cities are unsettling and overdetermined, that is, the outcome of multiple social, psychical, and political forces. To illustrate this thesis, Pile uses the Freudian–Marxian analytic categories of 'city work', 'magic work', 'dream work', 'emotional work', 'time work', 'blood work', 'grief work', and 'space work' to show how emotions and fantasies do ideological work in city life.

The main method in Pile's empirical analyses is what he calls 'tracking' which aims to 'get at the circulation of urban imaginaries' and things such as postcards, bill boards, advertisements, architecture, shop windows, cer eal boxes, graphic novels, (installation) art, tourist leaflets, newspapers, theater sets, buildings, graffiti, fantasies, information, stories, and ideas that circulate in cities, including London, Singapore, New York, and New Orleans. Tracking, then, is a method that is open to the contingency and multiplicity of possible objects, scenes, events, and processes. As a result of 'tracking', Pile's interpretations and narrative, echoing Freud's dream analyses (a topic that Pile has previously explored), focus on the spaces of short circuits, disconnects, and coincidences.

Pile posits that it is not sufficient to simply 'ground' our theories vis à vis the method of grounded theory. Put differently, Pile rejects the idea that when we do research we must bring into concrete realms our abstract concepts and theses. An adequate understanding of the material, Pile asserts, demands an adequate understand ing of the immaterial. Alongside extensive archival research, Pile uses 82 black and white illustrations that are as sparkling and flat as the urban spaces they underscore. By addressing the visual, Pile's maneuver (in contrast to Bingley's) echoes Gillian Rose's work on methodology and her use of psychoanalytic concepts such as 'lack' and the 'gaze' in order to examine the visual.

As mentioned earlier, a key motive for geographers to adopt a psychoanalytic theoretical framework is because of its explanatory power of seemingly irrational behavior. Exemplary here is Heidi Nast's work on the segregated spaces of racism and racial fear in the context of US cities and the fears of black men raping white women during and after transatlantic slavery. For Nast, the causes of racist violence such as lynching are not entirely caught up in the social. Rather, racist violence emerges precisely because there are limits to the social: not everything can be socially articulated or collectively put into words and collectively acknowledged.

Underpinning Nast's research is the psychoanalytic idea that in order to explain phenomena, one focuses not so much on the links between the Particular (e.g., diverse locales) and the Universal (e.g., global forces), but rather, on the links between the Singular and Universality, that is, on how a detailed phenomenon marked by excess, ex ception, and intangibility reveals a universal logic. For Nast, there are three interrelated singular events vis à vis the lynching of black men in post Reconstruction settings. First, of the thousands of black men who were lynched, many of them were castrated. Second, many lynchings were celebrated publicly with hundreds and thousands of white children and family members in attendance. Third, many lynchings of black men were justified as a response to the alleged rape of a white woman. Nast argues that these three singularities are indicative of the universal logic of the Oedipal family *qua* a hegemonic mode of socio spatial organization that legitimates racist violence.

In order to explicate these points, Nast adopts the method of what she calls 'mapping' – a mélange of the oretical and empirical investigation that is sensitive to historical geographic contexts and highly informed by psychoanalytic theories. Like Pile's method of tracking, Nast's mappings focuses on specific material objects that include white colonial mother dolls, frontispiece illus trations, films, poems, and a magazine cover. Unlike Pile's 'tracking' which predominantly embraces the magical (sur)realism of space, Nast's mapping is primarily an interpretative strategy through which to explain the causes that infuse the irrational violence of racist land scapes. In so doing, Nast addresses questions about the repression of geopolitical mechanisms and forces and the circulation of desire. Nast also takes her cue and expands upon two methodological sources: First, Fredric Jame son's notion of the 'political unconscious' that seeks to methodologically avoid the alleged apolitical indi vidualism in psychoanalysis. Second, Gilles Deleuze and Félix Guattari's critique of what they see as psycho analysis's all too routine uncritical acceptance of the Oedipal family as a unit that structures and defines de sire. Much of Nast's method of mapping is informed by what she calls 'nodal thinking' which involves attending to spaces not so much as parts of larger story, but rather as spaces through which stories are (un)told and struc tures can take hold.

Conclusion

In the beginning of this article, the author argued that one of the reasons for geographers' hesitations about using psychoanalytic methods is the idea that psycho analysis and its methods are best left lying on the couch, rather than brought out into social scientific research arenas. There is a certain irony here of course, because Freud did not want psychoanalysis and its methods to be left to a cadre of expert physicians. What mattered to Freud was good training and experience rather than professional diplomas. This is not to say that it is high time geographers began to psychoanalyze their research subjects, nor is it to dismiss the profound ethical ques tions about the use of methods that may usher in forms of therapy. Rather, the question of psychoanalytic methods in geography prompts us to remember that much of the value of research coincides with the extent to which the research embraces new questions and theoretical frameworks.

It is only very recently, via the work of Mary Thomas and other geographers, that the long held critical axiom that research questions, theoretical frameworks, and methods are fundamentally inseparable is finally being considered in geography from a psychoanalytic perspective. This work is extremely important not only because it promises to enrich how human geography is 'done', but also further explicate the methodological distinctions between data generation and data analysis. In so doing, geographers may begin to formulate exciting and truly radical research projects wherein psychoanalytic methods are as equally relevant and valuable as the psychoanalytic concepts that they ultimately depend on.

See also: Affect; Critical Geography; Feminist Methodologies; Grounded Theory; Interdisciplinarity; Language and Research; Other/Otherness; Polyvocality; Psychoanalytic Theory/Psychoanalytic Geographies; Psychotherapy/Psychotherapeutic Geographies; Reliability and Validity; Self-Other; Situated Knowledge, Reflexivity; Situationism/Situationist Geography; Subjectivity; Surrealism/Surrealist Geographies; Uncertainty.

Further Reading

Bingley, A. (2003). In here and out there: Sensations between self and landscape. *Social and Cultural Geography* 4, 329 345.

Bondi, L. (1999). Stages on journeys: Some remarks about human geography and psychotherapeutic practice. *Professional Geographer* 51, 11 24.

Bondi, L. (2003). Empathy and identification: Conceptual resources for feminist fieldwork. *ACME: An International E Journal for Critical Geographies* 2, 64 76.

Burgess, J., Limb, M. and Harrison, C. M. (1988). Exploring environmental values through the medium of small groups. Part 1: Theory and practice. *Environment and Planning A* 20, 309 326.

Callard, F. (2003). The taming of psychoanalysis in geography. *Social and Cultural Geography* 4, 295 312.

Kingsbury, P. (2003). Psychoanalysis, a gay spatial science? *Social and Cultural Geography* 4, 347 367.

Kingsbury, P. (2004). Psychoanalytic approaches. In Duncan, J. S., Johnson, N. C. & Schein, R. H. (eds.) *A Companion to Cultural Geography*, pp 108 120. Oxford: Blackwell.

Nast, H. (2000). Mapping the 'unconscious': Racism and the oedipal family. *Annals of the Association of American Geographers* 90, 215 255.

Nast, H. (2001). Nodal thinking. *Historical Geography* 29, 74 76.

Oliver, S. (2003). Geography's difficult engagement with the psychological therapies. *Social and Cultural Geography* 4, 313 321.

Pile, S. (1991). Practising interpretative geography. *Transactions of the Institute of British Geographers* 16, 458 469.

Pile, S. (1996). *The body and the city: Psychoanalysis, space and subjectivity*. New York: Routledge.

Pile, S. (2005). *Real cities*. New York: Sage.

Rose, G. (2001). *Visual methodologies: An introduction to the interpretation of visual materials*. New York: Sage.

Thomas, M. (2007). The implications of psychoanalysis for qualitative methodology: The case of narrative data analysis. *The Professional Geographer* 59, 537 546.

Psychoanalytic Theory/Psychoanalytic Geographies

P. T. Kingsbury, Simon Fraser University, Burnaby, BC, Canada

Glossary

Analysand The patient in psychoanalysis.

Ego The part of the mind that registers the sense of Self and mediates between the unconscious and conscious, as well as the Id, Superego, and social reality.

Id An unconscious part of the mind that is associated with immediate gratification and the primary instinctual impulses of aggression and sexuality.

Identification A process, primarily involving the Ego and the Superego, through which the human subject psychically adopts and incorporates an attribute or attributes (e.g., values, images, words, mannerisms, social roles) from another person or persons.

Jouissance Often translated as enjoyment, this Lacanian term refers to an intensely painful pleasure that is at once fascinating and threatening to psychical well-being and social stability.

Repression The process by which disturbing thoughts, wishes, and memories are expelled from consciousness into the unconscious.

Superego The part of the mind associated with moral agency, self-critical conscience, as well as social standards and imperatives.

Uncanny, the The English translation of the German word *unheimliche* which literally means 'un homely' and originally used by Freud to designate a mild and specific form of anxiety that is derived from encounters with familiar objects that have become suddenly strange.

Unconscious, the A partially inaccessible and dynamic part of the mind that consists of repressed thoughts, injunctions, and feelings that profoundly influence our conscious lives.

Introduction: What Is Psychoanalytic Theory?

Critical evaluations of a theory (from the Greek word *theōria* meaning 'contemplation' and 'speculation') seldom proceed by asking whether a theory is true or false. Theories are primarily hypothetical propositions that attempt to explain and analyze phenomena. Theories also help to guide speculative thinking, coordinate research questions, inform methods and methodologies, as well as provide interpretative frameworks through which we can make sense of complex data. Strictly speaking, then, theories provide neither 'right' or 'wrong' results, nor 'good' or 'bad' answers. Most critical evaluations of a theory usually pose the following questions: What are the epistemological assumptions, ontological implications, and methodological ramifications of a theory? How is a theory geographically and historically situated? How, why, and where has a theory been contested and revised? In what ways and to what extent is a theory compatible or incompatible with other theories? What is the relation between a theory and practice? Mindful of the above questions, let us turn to some fundamental theoretical maneuvers in psychoanalysis.

Psychoanalysis mainly consists of clinical 'practices', techniques, methods, and theories that seek to reduce psychical suffering. Psychoanalysis was first developed by Sigmund Freud (1856–1939) during the late nineteenth century and has been continually revised, disputed, and deformed within and without psychoanalysis. Like the psychoanalytic subject that is traversed by numerous conflicts and splits, for example, between conscious and unconscious life, as well as between biology and cultural significations, psychoanalysis is composed of unruly and sometimes unkind paradigmatic divisions. Let us briefly illustrate this point with three oft cited examples. First, Freud once vehemently demanded that his former 'crown prince of psychoanalysis', Carl Jung (1875–1961), stopped using the term 'psycho analysis' because of Jung's theoretical differences. Second, in 1953, Jacques Lacan (1901–81) was famously expelled from Inter national Psychoanalytical Association because of his idiosyncratic clinical methodology that rejected the rule that analytic sessions end after 50 minutes. And finally, during the early 1940s, Melanie Klein's (1882–1960) emphasis on aggression and the early development of the Superego in very young children prompted fierce and prolonged controversies with orthodox Freudians lead by Anna Freud (1895–1982) and the British Psycho Analytical Society.

In addition to its renowned theoretical volatilities, psychoanalysis is arguably a hybrid theory *par excellence* because its numerous theoretical approaches incorporate ideas from sources as diverse as literature, philosophy, philology, mythology, astrology, political economy, neuro science, religion, biology, anthropology, mathematics, and topology. Now, what makes psychoanalysis different from other psychological and psychotherapeutic approaches? Much of psychoanalysis is concerned with the interpret ation of and clinical intervention of unconscious thoughts, memories, and desires. While psychoanalysis is composed of many and oftentimes contrary theoretical approaches, all psychoanalytic 'schools' of thought are united by their theoretical attempts to explain the suffering that ensues from ostensibly superfluous and inexplicable feelings such as chronic guilt, remorse, jealousy, powerlessness, persecution, self loving, and self loathing. Taking it as axiomatic that the unconscious is an ineluctable mode of psychical apprehension, rather than simply a sociohistorical construct, psychoanalytic theory makes the radical claim that people are ultimately cut off from and cannot know what they want. Such a theoretical assumption ushers in two key psychoanalytic premises about subjectivity: people are not biologically or in stinctually predisposed to desire or love certain objects. This is why, for example, Freud asserted that people are no less predisposed to forge heterosexual than homo sexual or bisexual relationships.

Second, psychoanalysis does not take it for granted that people will necessarily pursue their own self interests. Crucial here, is the psychoanalytic theoretical emphasis on what Freud called the 'drive' (*Trieb*). Very briefly, the drive is a concept that concerns how human subjectivity and sexuality are not entirely determined by biological and cultural forces. Freud argued that human pleasure and procreation do not coincide easily and cannot be understood in terms of 'natural' or 'normal' behavior. In psychoanalysis, perversion is always the norm. In contrast to the innate attribute of instinct, Freud argued that the psychical and somatic montage of drives coordinated human sexuality. By insisting on the exist ence of the dynamic unconscious and the importance of the drive, psychoanalysis theorizes a series of deceptively simple, pertinent, and troubling questions. For example, psychoanalysis asks: Where does psychical pain come from? Why do we repeatedly put ourselves in harmful situations? Why and how do certain stultifying thoughts get 'into' our minds given the contexts and contingencies of our lives? What happens when we fall in and out of love? What is the relationship between ethics and desire?

In formulating answers to these questions, psycho analytic theory asserts the importance of real and imagined events that take place during the formative years of our early childhood. Why? During the first 5 or so years of our lives, we spend a great deal of our time acquiring language, learning cultural codes, and forming role models. Moreover, this period of our lives involves the never to be entirely reconciled trauma of psychically separating ourselves from our mother or primary care taker. Psychoanalysis contends that these events are critical in the formation of sexual difference and gender roles through unconscious formations of desire, defenses, identifications, and fantasies. Critical here are psycho analysis's various retheorizations of Freud's notion of the 'Oedipus complex', that is, the shifting psychical battlefields and pathways through which boys and girls organize their desires for and against their mothers and fathers. For most psychoanalysts, the Oedipus complex takes place once the infant has been through the 'oral', 'anal', and 'phallic' stages of psychosexual development.

Although Freud once dubbed psychoanalysis a 'depth psychology', psychoanalysis does not contend (as it is usually supposed) that the unconscious is a dark and primal thing that hides and gurgles away in the shadowy depths of people's minds. Rather, psychoanalysis asserts that the unconscious takes place in the heart of people's lives and everyday processes such as speaking, thinking, and dreaming. Furthermore, psychoanalysis suggests that the unconscious can speak and be interpreted like a ciphered message. For most psychoanalysts, the unconscious is the accumulation of other people's spoken words, that is, the aggregated subtle din and babble of people's desires, demands, compliments, scolds, as well as cultural codes, more generally. This is why Lacan famously called the unconscious "the discourse of the Other."

The clinical task of getting the unconscious to speak and then make sense of what it says is highly complex. To begin with, psychoanalysis takes a lot of time: on average, analysis takes place several times a week and usually lasts about 6 years. Analysis is never a question of simply bringing the unconscious into consciousness. Rather, the main psychoanalytic task is to relationally understand by obliquely mapping and partially unknot ting the tangled web of unconscious thoughts, wishes, fears, and so on. When analysts listen to what their analysands say, they do not simply focus on the content or meaning of their speech. A psychoanalyst focuses on the 'form' of what the analysand says, that is, 'how' the analysand talks: their stutters, slips, hesitations, repetitions, silences, and deviations. Here, psychoanalytic theory privileges the role of social discourses over the individual because psychoanalysis asserts that language always trumps our intentions: when we say something, we always say something more than we meant to say; when we speak, we always leave elements of doubt. Psycho analytic theory also proposes that language is material, that is, words and signifiers have material effects. For example, words can wound, deliver emotional stings, incite actions, prompt blushes, define bodies, create feelings, forge memories, and cinder pain. Psycho analysis, then, makes the following key theoretical wager: talking – putting things into words – can not only illu minate unconscious thoughts, desires, and goals, etc., but can also alleviate the pain of the speaker's symptoms.

However, why does psychoanalytic theory focus so heavily on the unconscious? Because psychoanalysis theoretically adheres to the 'principle of sufficient rea son', that is, the belief that when something happens, it does so because of a specific reason. On this point, psychoanalysis makes three important and somewhat unorthodox theoretical maneuvers vis à vis causality and psychological phenomena: first, psychoanalysis contends that mental events such as dreams, symptoms, and fantasies are composed and result from numerous rather than single causes. In Freud's words, all psychical phenomena are 'overdetermined'. Second, psychoanalysis theorizes cause not in terms of lawful and linear determinations, but rather, in terms of the indefinite and contingent. According to psychoanalysis, cause takes place when something unexpectedly fails or goes awry. Third, psychoanalysis posits that cause and effect do not and cannot coexist on the same empirical or phenomenal field. Psychoanalysis argues that it is a theoretical error to conflate cause and realization. From a psychoanalytic perspective the primary causes of people's behavior cannot be entirely present in, immanent to, or coincident with the realms of their consciousness. This is why psychoanalysis insists on 'going beyond' the conscious parts of people's psychical lives and attending to what is unthinkable and unknowable, that is, the unconscious parts of people's minds. The theoretical strategy here is not so much an attempt to faithfully capture or mirror psychical 'reality', but rather to explain or take into account the iterations of enigmatic and inexplicable psychical phenomena. In so doing, psychoanalytic theory is extremely pragmatic: it prefers working hypotheses to metaphysical guarantees.

One of the most pragmatic tenets of psychoanalysis is its theoretical investments in thinking about the complex geographies of psychical life. Taking seriously the idea that space is dynamic and expressive, psychoanalysis argues that our feelings and the causes of their feelings do not simply take place as a result of forces that move back and forth between the 'insides' and the 'outsides' of our heads. According to psychoanalysis, feelings and subjectivity itself are inherently spatial because they take place through psychical, corporeal, and discursive processes that involve, for example, fragmentation, displacement, condensation, introjection, and splitting. It is important to note that these key psychoanalytic categories are not spatial metaphors or conceptual allegories. These are categories through which psychoanalytic theory tries to answer questions about how our psyches and social worlds 'really' work.

Given the extent to which psychoanalysis attends to the spatialities of psychical phenomena, it should be hardly surprising that psychoanalytic theory has become a major theoretical approach in human geography. Moreover, Steve Pile asserts that psychoanalysis is a 'spatial discipline'. How so? A quick glance of any psychoanalytic text will reveal how psychical life is theorized in terms of the interactions and interdependencies between different places or locales within, outside, and betwixt people's minds, words, and worlds. Thus, psychoanalysis provides a rich theoretical vocabulary in which to 'map' the inter- and intrasubjective dimensions of space and society. However, psychoanalytic geographies are not simply the inevitable result of psychoanalysis's theoretical commitments or sensitivities toward space. It is important to consider geography's disciplinary contexts through which psychoanalysis has become relevant and legible to many human geographers. It is to a partial and brief examination of how and why psychoanalytic theory was first adopted by geographers that we now turn.

Psychoanalytic Geography

Geography's 'Discovery' of the Freudian Unconscious

Like geography, psychoanalysis has an abiding interest in the worldly. Psychoanalysts frequently use inductive reasoning whereby theories derive from the space–time singularities of people's lives, that is, the uniquely topsy-turvy worlds of people's broken hearts, shattered dreams, and recurrent tears. To be sure, psychoanalysis is spelt 'psych*o*analysis' not 'psych*e*analysis' or 'psych*ic*analysis'. This quiet, yet indelible 'o' that is inscribed between 'psych' and 'analysis' is the mark of an orb – an Earth writing or 'geo-graphia' of sorts – wherein the 'o' also signals 'of the world'. How so? According to psychoanalysis, the clamorous and delicate dramas of the world are inescapably psychical and people's psychical torments and jitters are inescapably worldly. Psychoanalysis provides an extremely rich theoretical vocabulary in which to explore the volatile reservations and subtle bonds that (de)structure, traverse, and often catch unawares subjects and groups. This is why, during the early 1930s, Albert Einstein, in responding to the League of Nations' invitation to publish a series of open letters that addressed pressing social issues, wrote to Freud and asked him how to prevent the 'urgent and absorbing problem' of the growing 'menace of war'. Thus, on questions about interpretation, passion, and expression, many disciplines in the arts and humanities find theoretical engagements with psychoanalysis indispensable, if not unavoidable. But when and why did geographers start to engage with psychoanalytic theory? What was it about psychoanalysis that first incited geographers' interest? And how have the numerous theoretical approaches that compose psychoanalysis been used by geographers?

Pile has written extensively on the disciplinary conditions and questions through which geographers came to 'discover' the Freudian unconscious. Briefly, during the 1970s, many geographers worked in the subfield of 'behavioral geography'. Behavioral geographers, rejecting the discipline's 'quantitative revolution' that reduced humans and human relations to statistical phenomena such as dots on a grid, examined the spatial dimensions of reasoning, decision making, place attachment, mental images, behavior, environmental perception, and cognition. Pile notes that behavioral geography was shackled with a fundamental theoretical problem: it did not possess an adequate theorization of how the mind worked vis à vis the phenomenal environment. Many behavioral geographers worked with a troublesome 'black box' model of the mind which conceptualized the mind as an ultimately chaotic, obscure, and unknowable entity. Furthermore, by the late 1970s and early 1980s, behavioral geography was increasingly critiqued by the 'humanistic' and 'radical' geographers. These two groups critiqued behavioral geographers for neglecting important dimensions of human subjectivity and agency. On the one hand, humanistic geographers, influenced by phenomenology and qualitative methodologies, brought to the fore the importance of cultural symbolic meanings. On the other hand, radical geographers, drawing on Marxist theories, tackled issues of cultural power structures. Theoretical questions about how the mind worked, however, still lingered. Specifically, humanistic geographers' reliance on the vague analytic category of 'experience' and radical geographers' heavy reliance on the binary of structure versus agency both proved insufficient to address the complex ways through which cultural meanings and spaces were contested and (re)produced.

By the end of the 1980s, geography underwent a significant 'cultural turn'. Partly a result of the increasing importance of the 'cultural' dimensions of the political events in the New World Order (e.g., the fall of the Berlin Wall, the Tiananmen Square protests, growing white nationalism in the United Kingdom and France), and partly the result of other 'cultural turns' across the social sciences and the humanities, many geographers (initially those working in the UK) began to focus on the interpretation of the social spaces of power, resistance, and struggles 'alongside' the contingent and contextual spaces of identity, subjectivity, and representation. Among these 'new cultural' geographers, Pile asserted that psychoanalysis offered rich theoretical and methodological frameworks in which to practice and engage with the intersubjective dimensions of 'interpretative geography'. Published in 1991, Pile's intervention was crucial in the inauguration of a subfield that would soon become referred to as 'psychoanalytic geography'.

One recent and valuable assessment of psychoanalytic geography is Felicity Callard's (2003) article, 'The taming of psychoanalysis in geography'. According to Callard, much of the taming involves geographers' dissolution of psychoanalytic concepts into psychological concepts. How so? Geographers often rely on interdisciplinary and assimilatory secondary literatures that promote historicist, culturalist, and social constructionist rather than psychoanalytic understanding of subjectivity and spaces. In addition, in contrast to psychoanalysis, many critical human geographers uncritically adhere to progressive theorizations about subjectivity that laud and take for granted people's capacity and willingness for political resistance and transformation. In addition to reevaluations of psychoanalytic theorizations in geography is the emergence of geographers' engagement with psychotherapies. This work, pursued by Liz Bondi and others, is important because it brings to the fore related and alternative approaches to the *sine non qua* of psychoanalysis: the reduction of somatic and psychical suffering.

Theories and Research Themes in Psychoanalytic Geography

Like psychoanalytic theory, psychoanalytic geography is difficult to categorize. Put differently, it is difficult to discern or impose a neat paradigmatic boundary around what is and what is not psychoanalytic geography. In what follows, the contours of three important theoretical approaches that compose psychoanalytic geographies are elaborated – Freudian, Lacanian, and object relations theory approaches – and their attendant theoretical questions and research agendas. What follows is an overview of keyworks and research themes, rather than an exhaustive account of what constitutes psychoanalytic geography. Readers are thus encouraged to read and read beyond the references cited below, as well as the cross references.

Dreams, the Unconscious, and the Uncanny

Pile contends that a major reason behind the enduring critical attention devoted to Freud's groundbreaking book, *The Interpretation of Dreams*, is because the work demonstrates the incisiveness of Freud's 'spatial thinking'. Freud attempts to understand the shifting and elusive meanings of dreams and how they 'work' by drawing on spatial analytic categories such as 'condensation', 'displacement', and 'associative paths'. Notably, Pile illustrates how Freud's 'dream spaces' are not only caught up in the revisions and reversals of personal desires, but also the illuminations and reflections of social things. For these reasons, Pile theorizes urban spaces as dream spaces. Theoretically, Pile uses the Freudian–Marxian analytic categories of 'city work', 'magic work', 'dream work', 'emotional work', 'time work', 'blood work', 'grief work', and 'space work' to show how emotions and fantasies do ideological work in city life. Psychoanalysis enables Pile to show how an effective 'grounding' of theories (whether psychoanalytic or not) not only requires an adequate understanding of materiality and space, but also an understanding of the immaterial.

To illustrate this theoretical point, Pile draws on the psychoanalytically inflected social theories of Walter

Benjamin in order to interpret cities as urban 'phantas magorias', that is, alluring space–time processions of optical illusions, secret desires, irrational anxieties, imaginary figures, moody misdeeds, and fantastic stories. Pile's psychoanalytic geography also explicates how dreams are a key part of everyday political geographies. How so? Dreams involve politics because they incite struggles: not everyone shares the same dreams and/or nightmares. For Pile, dreams are also political because our ability to shape and intervene in the world is partially determined by how we are gripped by the world of dreams.

One of the most extensive and influential investi gations in human geography of how the Freudian un conscious works in sociospatial formations is Heidi Nast's research on 'mapping the unconscious'. Like Pile's assessment of dreams, Nast asserts that the unconscious plays a significant role in everyday life and politics. Nast provides insight into how the unconscious plays a con stitutive role in the spatial organization of violence, injustice, and exploitation in US racist landscapes in the context of southern plantation, post Reconstruction set tings, and the educational policies and urban renewal programs in 1950s Chicago. Underpinning Nast's investigations are two psychoanalytic maneuvers. The first idea concerns how a geographical explanation of a social phenomenon can proceed not so much by empirically mapping the links between the particular (e.g., the local) and the universal (e.g., the global), but rather by dialectically connecting the universal to the singular, that is, the exceptional. How does this distinctly Freudian (and indeed Hegelian) theoretical mode of analysis play out in Nast's interpretations? For Nast, there are three interrelated singular events. First, thousands of black men were not only lynched, they were also castrated. Second, many lynchings were not clan destine or secret, but rather public celebrations con sisting of hundreds even thousands of white family members. Third, numerous lynchings of black men were frequently libidinized insofar as they were typically the direct response to the alleged rape of a white woman. For Nast, the singularities of the excessive or seemingly useless act of castrating a dead body, the proximity of the family unit to the scene of lynching, and the frequency of the alleged rape of a white woman can provide clues to explain the senseless or irrational vio lence of racism.

The second psychoanalytic idea that underpins Nast's complex, yet extremely rewarding paper concerns the theoretical premise that it is a mistake to conflate cause and realization. For Nast, the motives and causes of racist violence such as lynching are not and cannot be entirely caught up in the social. Rather, such violence emerges precisely because there are limits to the social: not everything can be socially articulated or collectively put into words and acknowledged. Furthermore, as Nast argues (following Freud), the constitution and ostensibly normal functioning of sociospatial relations actually requires certain things to be rendered unspeakable or unthinkable, that is, sociability requires the repression of specific dangers and threats. In Nast's paper, in the context of societies dominated by racist white Oedipal (father, mother, and son) families, the 'repressed bestial being' that is made 'legitimately secret' is an incestuous wish fulfillment between the mother and the son. In Nast's 'mappings' or case studies, incest is racially en coded as blackness and symbolically aligned with young black males or 'boys'. Young black males, then, are un consciously produced as threats toward white women *qua* mothers and thus become 'repositories' of colonial and racist violence.

From a Freudian psychoanalytic perspective, social relations are ultimately compromise formations that are borne out of, require, 'and' continually fail to gentrify the repression of an underlying and antagonistic trauma. From a psychoanalytic perspective, social and cultural realities are not simply contingent and constructed; they are also extremely volatile and vulnerable to the dictates of aggression. In theorizing this relationship between so cial space and traumatic fissuring, several geographers have drawn on Freud's notion of the uncanny. For example, Rob Wilton has examined the uncanny effects of an HIV/AIDS hospice in Los Angeles's suburban land scapes and Derek Hook has examined the ideological roles and uncanniness of monuments in Pretoria, South Africa. Notably, Laura Cameron has examined the over laps between Freud's theories and Arthur G. Tansley's work on plant ecology. In addition, Mary Thomas has considered Freud's notion of the unconscious to rethink qualitative methodology in human geography using the example of narrative data analysis.

Imaginary, Symbolic, and Real Spaces

Jacques Lacan is widely cited as the most significant psychoanalytic theorist since Freud. It is difficult to underestimate the influence of Lacanian theory in the humanities, cultural studies, gender and sexuality studies, and increasingly, via the extensive work of Slavoj Žižek (1949–), the social sciences. Lacan's thinking stretches over five decades and is extremely complex. Lacanian theory is united by the theory of the 'registers'. During the 1970s, Lacan examined knots from the point of view of their topological properties. Lacan devoted much theoretical attention toward the Borromean knot which is composed of three interlocking and constitutive rings. For Lacan, the Borromean knot aptly illustrates the structural interdependence of three registers that com pose psychical life: the Imaginary, Real, and Symbolic. Lacan theorized these three registers not as discrete entities but rather, as the musical connotation suggests, in terms of proximate relations of discordance and harmony. What do these registers mean? Very briefly, the Imaginary (the focus of Lacan's work in the late 1930s and 1940s) refers to dual or binary relations that do not involve the imagination *per se*, but rather concern the Ego and its illusions of totality, liability to fragmentation and lures, narcissism, rivalry, and aggressiveness. The Symbolic (the focus of Lacan's work in the 1950s and early 1960s) is a noun that primarily refers to language, cultural rules and conventions, exchange, measurement, lack, death, and the unconscious. The Symbolic also shields people from the disturbances and chimerical effects of the Real. The Real (the focus of Lacan's work in the late 1960s and 1970s) is probably the most misunderstood and complex of Lacan's registers. In a nutshell, the Real takes place through sublime, contingent, and terrifying phenomena. Crucially, we cannot directly confront, obtain, or grasp the Real. Rather, the Real involves near misses or barely elided encounters that subsist outside, disturb within, and are unassailable to Symbolic signification and Imaginary wholeness. Spatially, the Real concerns phenomena that are, on the one hand, anamorphic *qua* forever elusive limits and, on the other hand, obdurate *qua* disturbances that are always cropping up and tripping up things. For example, if the discursive constructions of race concerns the Symbolic and racial identification concerns the Imaginary, then racism concerns the Real. Interestingly, for Steve Pile, Lacan's three registers can be roughly mapped onto and through Henri Lefebvre's spatial triad of 'lived space' (or, representational spaces), 'conceived space' (representations of space), and 'perceived space' (spatial practices).

Lacan's theory of the three registers, as well as the works of Freud and Jacques Derrida (1930–2004), inform much of the theories, or better still the *écriture féminine* (feminine writing) of three influential French feminists: Luce Irigaray (1930–), Hélène Cixous (1937–), and Julia Kristeva (1941–). Numerous feminist geographers have drawn on the writings of Irigaray, Cixous, and Kristeva to examine the patriarchal biases that infuse cultural practices, knowledge, representations, as well as corporeal and visual spaces. For example, Gillian Rose has mainly drawn on Irigaray's understanding of the Imaginary to question masculinist demarcations between 'real' and 'nonreal' spaces. Jenny Robinson has used Kristeva's concepts of the Symbolic, abjection, and the semiotic to emphasize the ineluctable vulnerabilities and failures of sociocultural borders and distinctions. Given that Lacan was highly and openly skeptical of feminism, many geographers have questioned the value and viability of Lacanian psychoanalysis. And yet, feminist geographers such as Virginia Blum, Heidi Nast, and Liz Bondi have acknowledged the relevance and potential of Lacan's theories to challenge heteropatriarchy and spatial oppression. One way this potential is being realized is through Lacan's oft misunderstood category of the Real. Paul Kingsbury, for example, has drawn on Žižek's political readings of Lacan's category of *jouissance* that basically means an antagonistic and excessive libidinal charge in order to theorize the links between the social bonds of exploitation and domination in the context of tourism and consumption. In so doing, Kingsbury's work builds on Felicity Callard's argument that psychoanalysis is often theoretically 'tamed' in geography.

Other notable Lacanian approaches in human geography include Robert Wilton's work on physical disability, Ben Page and Jon Goss's research on the politics of production and commodification, and Stephen Healy's examination of the ethics and social fantasies that reside in informal caregiving and healthcare delivery in the US.

Boundaries of Selves and Others

The third main theoretical approach in psychoanalytic geography is informed by object relations theory. Briefly, proponents of object relations such as Melanie Klein (1882–1960), Michael Balint (1896–1970), and British psychoanalysts, including Donald W. Winnicott (1896–1971), Ronald Fairbairn (1889–1964), and Harry Guntrip (1901–75) emphasize how (following Freud's notions of an 'object' and 'object relation') the psyche's interactive and intersubjective world is made by and through relationships 'and' objects. From this perspective, 'relations' are first and foremost 'inter'relationships wherein the psyche not only constitutes its objects, but is also constituted by, composed of, and relates to itself through objects. 'Object' refers to how something, typically a person, body parts, and somatic sensations, or a character role can become the medium, procurer, and/or locus of the drive.

According to object relations theory, infants begin to develop a sense of self through a growing awareness about how their bodies are separate from their mother's or primary caretaker's body. As boundaries between the infant's self and the external world become increasingly demarcated, the infant realizes the extent of its vulnerability. The child's subsequent experiences of loss, separation, abandonment, desire, fear of dissolution, and anger produce an ambivalent response to the mother. When the mother provides food, warmth, love, and comfort, she is a 'good' object. However, if the mother leaves the child on her or his own and does not provide gratification when demanded, she becomes a 'bad' object. In this way, the mother is loved for providing love and hated for the withdrawal of love. The child is faced with a dilemma because it cannot comfortably hate or express anger because it fears punishment and thus the child must control these emotions. This ambivalence pervades the child's development and structures its responses to the qualities its mother once provided (loss, love, comfort, etc.) later in life.

David Sibley's work, informed by object relations theory, has influenced a great deal of social and cultural geography. Asserting that the delimitation of communities and social groups involves the same difficulties as constituting oneself and others, Sibley examines how dominant social groups shore up their power by enforcing boundaries that secure dominant groups' sense of purity and sameness through the exclusion and marginalization of other groups that are typically codified in terms of impurity, defilement, and shame. Sibley shows us the extent to which these psychosocial processes rely heavily on modern media representations of threats and vulnerability that are configured along axes such as race, gender, sexuality, class, and age. Sibley elaborates how these representations such as newspaper photographs labeling gypsies as 'criminal types' are a crucial element in the social construction of 'good selves' and 'bad others'.

Stuart Aitken and Thomas Herman have drawn on two object relations concepts. The first, 'potential space' or 'transitional space', is a space through which meditative, fluid, transitional, and exploratory demarcations of Self and Other (e.g., person, people, landscape, object, etc.) take place during infancy. Such a space is replete with the interconnective capabilities of 'objects'. The second concept is the 'transitional object' which typically includes the breast, food, a toy, and blanket. Such transitional objects are mediators and travelers between permeable inner and outer psychical worlds which, through acts of 'playing' (inverted commas in original), help to establish a potential space. According to Aitken and Herman, these two concepts (especially potential spaces) provide useful and 'refreshing' ways, to understand how in Western contexts children develop and form identities through acts of play that disrupt and resist adults' masculinist and patriarchal imperatives of control, order, and reason. Another important example of object relations theory in human geography is Amanda Bingley's development of a Winnicottian methodology that enhances our access to unconscious relations between the Self and landscape.

Also engaging with landscape via object relations theory is Robert Wilton. Drawing on Kristeva's concept of abjection, Wilton examines the social anxieties that surround the planned construction of a small AIDS hospice in a Los Angeles neighborhood. Wilton theoretically and empirically shows us how people's anxieties and intolerance of someone and/or something that is different are bound up with fragile yet intransigent psychosocial spatial relations of embodiment, defilement, and purification. A significant psychosocial process here is projection wherein people expel and find in other persons and objects disturbing psychical material. Wilton's work is important because it demonstrates the explanatory power of psychoanalytic theory to show why people experience anxiety vis à vis social differences and how they contribute to the reproduction of social and spatial divisions.

Concluding Remarks

Psychoanalysis typifies the convulsive beauty and liveliness of doing theory because it illustrates how theories can incite building, bashing, mongering, tailoring, spinning, bigotry, and blindness. Psychoanalytic theory is one of the most valuable and controversial theories in human geography. Psychoanalytic theory shares numerous axioms and goals with other important theoretical approaches in geography. For example, with historical materialism and Marxism, psychoanalysis not only emphasizes how modes of living are inescapably historical and social, psychoanalysis also strives – through 'praxis' – to ameliorate the condition of people's lives. Alongside feminist theories, psychoanalysis brings to the fore questions about people's gendered/sexed relations, enactments, embodiments, conflicts, confusions, and expectations. With poststructuralist theories, psychoanalysis posits that much of the happenings in the world and people's lives take place through and because of the dynamisms, fixations, and uncertainties of language. And with non representational theories, psychoanalysis focuses on how affect and imperceptible yet material forces are situated within complex sets of unfolding intersubjective relations and representations.

And yet, on numerous points that have yet to be fully explicated by geographers, as well as this small encyclopedia article, in many ways psychoanalytic geography is quite different from the above theoretical approaches in geography. How so? In a nutshell, the *sine qua non* of psychoanalytic theory is its contention that the human subject is fundamentally and forever radically split and conflicted: on the one hand, our minds are divided in terms of consciousness and the unconscious. On the other hand, our bodies are forcibly carved by the dictates of biology and language or cultural significations. Taking into account these two divisions, much of psychoanalytic theory is concerned with understanding how the world must be organized and work in ways that people can hold onto something that resembles 'subjectivity' and 'reality'. Who can doubt the importance of how the geographies of these 'things' – subjectivity and reality – will always inform the world that we live in, as well as all our impassioned theorizations?

See also: Affect; Behavioral Geography; Cognitive Geography; Critical Geography; Critical Theory (After Habermas); Feminism/Feminist Geography; Human Geography; Humanism/Humanistic Geography; Interdisciplinarity; Landscape; Language and Research;

Non-Representational Theory/Non-Representational Geographies; Other/Otherness; Place; Poststructuralism/Poststructuralist Geographies; Psychoanalysis; Psychotherapy/Psychotherapeutic Geographies; Self-Other; Situationism/Situationist Geography; Subjectivity; Surrealism/Surrealist Geographies.

Further Reading

Bondi, L. (1999). Stages on journeys: Some remarks about human geography and psychotherapeutic practice. *The Professional Geographer* 51, 11 24.

Blum, V. (1998). Ladies and gentlemen: Train rides and other Oedipal stories. In Nast, H. J. & Pile, S. (eds.) *Places Through the Body*, pp 263 280. New York: Routledge.

Blum, V. and Nast, H. (2000). Jacques Lacan's two dimensional subjectivity. In Crang, M. & Thrift, N. J. (eds.) *Thinking Space*, pp 183 204. New York: Routledge.

Callard, F. (2003). The taming of psychoanalysis in geography. *Social and Cultural Geography* 4, 295 312.

Cameron, L. (2001). Oral history in the Freud archives: Incidents, ethics and relations. *Historical Geography* 29, 38 44.

Hook, D. (2005). Monumental space and the uncanny. *Geoforum* 36, 688 704.

Kingsbury, P. (2004). Psychoanalytic approaches. In Duncan, J. S., Johnson, N. C. & Schein, R. H. (eds.) *A Companion to Cultural Geography*, pp 108 120. Oxford: Blackwell.

Kingsbury, P. (2005). Jamaican tourism and the politics of enjoyment. *Geoforum* 36, 113 132.

Kingsbury, P. (2007). The extimacy of space. *Social and Cultural Geography* 8(2), 235 258.

Lefebvre, H. (1991). *The Production of Space*. Malden, MA: Blackwell.

Nast, H. (2000). Mapping the 'unconscious': Racism and the Oedipal family. *Annals of the Association of American Geographers* 90, 215 255.

Philo, C. and Parr, H. (2003). Introducing psychoanalytic geographies. *Social and Cultural Geography* 4, 283 293.

Pile, S. (1991). Practising interpretative geography. *Transactions of the Institute of British Geographers* 16, 458 469.

Pile, S. (1996). *The Body and the City: Psychoanalysis, Space and Subjectivity*. New York: Routledge.

Pile, S. (2005). *Real Cities*. New York: Sage.

Robinson, J. (2000). Feminism and the spaces of transformation. *Transactions of the Institute of British Geographers* 25, 285 301.

Rose, G. (1996). As if the mirrors had bled: Masculine dwelling, masculinist theory and feminist masquerade. In Duncan, N. (ed.) *Bodyspace: Destabilizing Geographies of Gender and Sexuality*, pp 56 74. New York: Routledge.

Sibley, D. (1995). *Geographies of Exclusion*. New York: Routledge.

Wilton, R. (1998). The constitution of difference: Space and psyche in landscapes of exclusion. *Geoforum* 29, 173 185.

Žižek, S. (1989). *The Sublime Object of Ideology.* New York: Verso.

Relevant Websites

http://www.p e p.org/
Psychoanalytic Electronic Publishing.

Psychotherapy/Psychotherapeutic Geographies

L. Bondi, The University of Edinburgh, Edinburgh, UK

Glossary

Autoethnography A method of qualitative research in which the researcher treats his or her own experience as the subject of ethnographic investigation. This approach rejects the traditional separation between researcher and research subject and entails a form of ethnography conducted by the researcher on the researcher.

Cartesian Relating to the seventeenth-century French philosopher René Descartes, the term Cartesian is often used for key ideas from his work, including his framing of mind and matter as a dualism, that is, as two mutually exclusive and exhaustive categories.

Nondualistic Concepts and approaches that break with Cartesian dualisms and use both–and, rather than either–or, formulations.

Psyche An aspect of human being and vitality. From the Greek concept of self, psyche may be translated as mind, soul, or spirit. It is sometimes linked with soma, as in psyche-soma, to refer to an integrated, holistic view of human being.

Psychogeography A branch of psychological speculation or investigation and a political movement concerned with the effects on the psyche of the geographical environment.

Psychotherapeutics Methods of treating disease or alleviating suffering by nonphysical, psychic means including spiritual healing as well as psychotherapy.

Psychotherapy A way of responding to illness, distress, or suffering by nonphysical or psychic means. Psychotherapy has its origins in psychoanalysis and is sometimes referred to as a psychological therapy or a talking therapy. Many different forms of psychotherapy are available.

Situationism A revolutionary international political and artistic movement active during the 1960s that sought to engage the disruptive and creative possibilities of ordinary everyday events and contexts.

Soma The material body, typically in contrast with the soul, mind, or psyche. It is sometimes linked with psyche, as in psyche-soma, to refer to an integrated, holistic view of a human being.

Subjectivity Qualities of a human being that characterize human subjects' situated self-experience. Subjectivity is often associated with feelings, perceptions, and beliefs of human individuals. However, a sense of individuality is itself an aspect of subjectivity and is not necessarily inevitable or universal.

Transference A concept used in psychoanalysis to describe how a person's patterns or habits of relating are transferred from one context (typically that of early life with parental figures) to others. Psychoanalysts use the idea of transference to describe key aspects of their experience of how their patients relate to them, which they theorize in terms of early relationships with caregivers.

Introduction

What Is Psychotherapy?

The words psychotherapy and psychotherapeutic come from the Greek and are subject to a variety of translations and interpretations. As a therapy, psychotherapy aims to treat, heal, cure, or ameliorate illness, distress, or suffering. This therapy may be understood as a medical intervention, as an alternative or complement to medical interventions, or as a wholly different activity or practice, perhaps construed as personal development or as a philosophy of life. Definitions of the psyche on which psychotherapy focuses include soul, spirit, and mind. These share reference to something other than, or at least not confined to, physical form. Under the influence of the Cartesian distinction between mind and body, psycho therapy may be understood to entail the treatment of the (nonphysical) mind or psyche, as opposed to the (physical) body or soma. The terms soul and spirit are preferred to mind by those who reject Cartesian dualism in favor of a more holistic and nondualistic philosophy. They also connect psychotherapy more closely to tradi tions of faith healing than to scientific medicine.

According to the *Oxford English Dictionary*, isolated usage of the word psychotherapy can be traced to the middle of the nineteenth century, with more common occurrence initiated in the 1890s. The appearance of the word in Sigmund Freud's work helped to ensure its rapid spread and its increasing association with forms of therapy derived in some way from the so called talking cure of psychoanalysis.

The term psychotherapeutic is used as both adjective and noun. Apart from pertaining to psychotherapy, it also refers to the treatment of disease by psychic means, including hypnotism and spiritual healing as well as psychotherapy. In this sense its meaning has tended to remain somewhat broader than psychotherapy's associ ation with the so called talking cure. As a noun, it is almost invariably plural – psychotherapeutics – and refers to the various methods of treatment involved in treating the mind, soul, or spirit by psychic means.

This article focuses primarily on psychotherapy as a term that refers to a group of practices, including and derived from or inspired by psychoanalysis. Some versions involve talk. Others use music, art, and other modes of communication. Some involve meetings between one practitioner and one person seeking help. Others involve couples, families, and groups who may work with one or more practitioners. The theories of personhood, psyche, illness, suffering, and healing on which psychotherapists draw vary widely, including three main groups, namely psychoanalytic theory (including its various post Freudian branches); scientific, cognitive, and behavioral psychology; and a diverse array of humanistic philosophies. These different theories generate numerous debates and conflicts over the nature of subjective life, the sources of psychic suffering, and more specific or practical issues such as diagnosis, therapeutic techniques, and training requirements. Different positions within these debates are associated with brand names or schools of psychotherapy such as Lacanian, Kleinian, Jungian, object relations, relational, cognitive behavioral, rational emotive, person centered, existential, gestalt, transactional analysis, transpersonal, drama therapy, and many others. What unites different approaches to psychotherapy across these different traditions and schools is their status as practices (not solely theories) and convergence in relation to key aspects of their theories of practice, notably the central importance accorded to the therapeutic relationship. This article leaves aside theoretical and practical differences and debates between traditions and schools of psychotherapy. Instead, it discusses geographical themes pertaining to psychotherapy as a diverse but nevertheless related family of practices.

The Influence of Psychotherapy

Psychotherapy (including such practices as psychoanalysis, counseling, counseling psychology, clinical psychology, analytic psychology, arts therapies, family therapy, relationship therapy, and hypnotherapy) has become increasingly widespread and influential since the early twentieth century. To some extent psychotherapy has replaced older ways of responding to distress, suffering, and problems in living, such as spiritual healing, pastoral care, the Catholic confessional, retreats, and asylum. However, the growth of psychotherapy signals more than a replacement of traditional with modern responses to illness, suffering, and distress. A wide panoply of psychological ideas and practices, including psychotherapy, have transformed human societies, and, in some ways, human subjectivity itself. Broadly psychotherapeutic approaches permeate many aspects of modern societies and recruit people into an expansive sense of individual subjective life. In so doing, psychotherapy contributes to ways of conceptualizing persons as embodying interior lives: it imputes to human beings private lives of the mind and internal subjectivities that actively open up what might be thought of as interior geographies. In this context it is hardly surprising that people have sought and been guided to forms of assistance, support, or treatment that claim to address subjective, emotional, and psychological distress. Arguably, culture, the state, and modern modes of governance are themselves saturated with the psychotherapeutic, simultaneously conceptualizing people as psychological consumers, citizens, and subjects, and calling people to mobilize and attend to the subjective interiors constructed through such conceptualizations. Hence suggestions that we live in a therapeutic culture governed by a therapeutic state and subject to therapeutic modes of governance. Psychotherapeutic geographies are an expression of and response to these same trends.

Geographies of Psychotherapy

Geographers have studied the spatial distributions and forms of numerous phenomena, some of which relate to psychotherapy, including, for example, the geographical incidence of mental illnesses, the historical geography of madness, the availability and accessibility of psychiatric services, and the spatial experiences associated with specific mental disorders. However, rather limited attention has been paid to the spatial distribution and forms of psychotherapy and psychotherapeutic practices. This article therefore suggests an agenda for such research, which would help to delineate geographies of psychotherapy.

Historical Geographies of Psychotherapy

Where are which forms of psychotherapy and psychotherapeutics found and how have these distributions evolved? Geographers have begun to approach these questions through explorations of certain strands of psychological, psychiatric, and psychoanalytic thinking, such as the establishment of child psychology in late nineteenth century New York and case studies of the application of psychoanalytic and psychiatric ideas in experimental schools and mental institutions, work well represented by a special issue of *Environment and Planning D: Society and Space* edited by Elizabeth Gagen and Denis Linehan. Alongside such studies, some other aspects of the historical geography of psychoanalysis are well known. Sigmund Freud's position as founding father is widely accepted, linking the origins of psychoanalysis to late nineteenth century Vienna. His early patients, collaborators, and followers were also located in Central Europe, drawn primarily from affluent Jewish communities, with the entry of Carl Jung into the field welcomed by Freud partly as a signal that psychoanalysis might be taken up by other ethnic and social groups. By the early twentieth century Freud's ideas had arrived in North America, and by the 1930s, under the shadow of Nazism, many central European psychoanalysts were themselves on the move, migrating to numerous countries around the world.

We also know that Freud's ideas were received and developed differently in different places. For example, in the USA, psychoanalysis rapidly became the exclusive preserve of the medical profession and was effectively incorporated into medical–psychiatric practice. By contrast, in Europe, psychoanalytic training has always remained open to those who do not hold medical qualifications, which, in the early twentieth century, enabled women to gain access at a time when their entry to medical training was nonexistent or very limited. Highly influential lay (i.e., nonmedical) psychoanalysts include Melanie Klein. While some European psychoanalysts were and are medically trained, the relationship between psychiatry and psychoanalysis developed rather differently in Europe from North America. Non psychoanalytic approaches to psychotherapy have traveled too. Key figures in the mid twentieth century development of humanistic psychotherapy include the American Carl Rogers, the German Fritz Perls, and the Canadian born, US trained Eric Berne. Among these and other influential figures, some initially trained as psychoanalysts but subsequently broke with traditional psychoanalytic approaches, while others never immersed themselves in the psychoanalytic tradition. Many trained, practiced, and taught in different countries, posing interesting questions about the interrelationships between places and the development of particular psychotherapeutic practices. The impression on local landscapes of these psychotherapeutic practices has begun to be explored by geographers and is a theme to which this article returns in due course.

The tradition of the training analysis generates psychoanalytic genealogies in which today's psychoanalysts can trace their lineage to foundational figures through a small number of steps. While other approaches to psychotherapy do not depend so strongly on the training analysis, relationships between trainees and their psychotherapists, trainers, and clinical supervisors are still accorded considerable importance, with practitioners often describing their credentials through such lineages. This personalized approach to the development of practitioners offers enormous scope to those interested in tracing genealogies of practice, and a historical geography lens provides an important way of situating such lineages in specific cultural contexts.

In his fascinating cultural history of American psychoanalytic psychotherapy, Philip Cushman has argued that a mutually reinforcing relationship developed between particular psychotherapeutic ideas about inner lives and the rise of a consumption based economy. Thus, the possibility of enriching one's interior sense of self through the acquisition of consumer goods and services was propagated, if not actively fostered, through psychotherapy, which also found itself responding to, and consumed by, some of those for whom such a project led to emotional distress and illness. While suggestive, this account pays minimal attention to place, implying a homogeneous view of America. However, the images of psychotherapy associated with, say, New York and California are sufficiently different that we might pose questions that open up more complex and differentiated accounts of the role of historically and geographically contextualized traditions of psychotherapy in relation to specific cultural landscapes.

Geographies of Psychotherapeutic Provision and Inclusivity

Turning from practitioners to those accessing psychotherapeutic services, we find extraordinarily complex patterns of provision. In the US and the UK, for example, services are located in the public, private, and third or voluntary sectors. Exactly what is available where, to whom, and on what basis varies enormously. Research on mental health services has demonstrated that patterns of referral to specialist psychological and psychiatric services are patterned by race, class, and gender. However, we know remarkably little about geographical dimensions of these various forms of unevenness and selectivity. In primary or community care, unevenness appears to be the norm. For example, in the UK, some primary care health services include counseling and others do not. Where such services are available, some general practitioners (family doctors) refer their patients while others do not: it appears that some regard counseling as a useful and relevant component of primary care while others regard it as irrelevant and perhaps seek to resist the rise of psychotherapeutic modes of response to distress and disease. The unevenness of primary care services is magnified in the voluntary sector, which depends heavily on local initiative. Moreover in this sector, many services are designed for specific sociodemographic groups (young people, women, ethnic minorities, etc.), and others focus on particular problems (relationship difficulties, addictions, survivors of sexual abuse, etc.), generating highly complex patchworks of provision. Psychotherapy provision has also been expanding in workplaces, with some, but not all, employers providing employee assistance programs or in house provision. Provision within health insurance schemes is also highly variable. In all of these cases, very little is known about the ensuing geographies of provision.

Alongside questions about the geographies of provision are questions about the accessibility and inclusivity of what is available. Services may be perceived in socially exclusive ways, despite apparent intentions to the contrary. For example, in the UK relationship counseling services have struggled to shake off a long standing reputation for being intrinsically middle class, in terms of the social backgrounds from which both their counselors and their service users are drawn.

The geography of provision and inclusivity of psychotherapy services is further complicated by matters of confidentiality, and fears about visual exposure and stigmatization. For example, some people from minority ethnic groups prefer to see a practitioner from any ethnic background other than their own; some people refuse to use services where their attendance might be observed by people they know; some service users prefer to travel considerable distances rather than using otherwise more convenient services in their own residential localities. Services respond with locational strategies, such as sitting in multi use buildings, and by resisting restrictions on access based on tightly drawn geographical boundaries.

Micro-Geographies and Experiential Spatialities of Psychotherapy

Different forms of psychotherapy are often highly prescriptive in their configuration of bodies, furniture, furnishings, décor, doors, and windows, generating distinctive micro geographies. They generally require two or more people meet in a room that is private and free from external impingements. Other people do not peer in, conversations cannot be overheard, telephones do not ring, and external noises are not intrusive. Rooms are generally not highly personalized in the sense of containing images and artifacts that reveal much about the personal life of the psychotherapist, but they are nevertheless welcoming and comfortable rather than austere and impersonal. Within such rooms the spatial organization of bodies and furniture varies. In classic psychoanalysis and psychoanalytic psychotherapy, the analysand or service user lies on a couch, with the psychotherapist out of their line of vision. Other forms of psychotherapy position participants face to face seated in comfortable chairs, placed at a (culturally specific) respectful distance, without other furniture in between. These arrangements are not arbitrary but are grounded in ideas about subjectivity and intersubjectivity. The use of the couch is informed by the idea that reducing visual other orientation and encouraging physical repose supports analysands to engage with and focus on their own subjective experience. The use of chairs facing one another is informed by egalitarianism – reducing the symbolic differences between participants – and by the idea that it is useful to engage more explicitly with intersubjective interaction. Thus, underlying ideas are expressed, represented, and/or performed through spatial configurations.

The subjective spatialities underlying psychotherapeutic practices are elaborated in many other ways within the concepts on which practitioners draw. For example, practitioners' descriptions of psychotherapy utilize ideas about the subjectively felt positions of practitioners and service users. The authority attaching to practitioners is often troubled or actively disavowed through accounts that make use of numerous spatial concepts, such as the notion of the practitioner being at the same level as, or alongside, the service user, or accompanying them on their journey. The idea that subjective experience is layered, with one set of feelings underneath another, offers another form of spatiality. So too do appeals to depth, as in ideas about going into things in depth, touching on deep feelings, or meeting at relational depth. Ideas about boundaries are also prominent in psychotherapeutic thinking and practice. Boundaries may refer to limits of time and space in which participants in psychotherapy meet, and to the forms of interaction that are or are not permissible. Sexual contact is strictly prohibited, while views on other forms of touching vary. The concept of boundaries is also used to delimit roles, with interactions between participants outside the psychotherapeutic relationship generally being proscribed or minimized. These ideas produce imaginative geographies of subjectivity and intersubjectivity, and illustrate the scope for further exploration of the spatial forms of subjectivity promulgated within psychotherapeutic practices.

Psychotherapeutic Possibilities of Geography

Alongside questions about the geographies and spatialities of psychotherapy are questions about the ways in which geography might be psychotherapeutic or might actively mobilize psychotherapeutic concepts. Such psychotherapeutic possibilities stem from attributes of the world on which geographical study focuses and from ways of studying that world.

Environments, Places, and Landscapes as Psychotherapeutic

That particular environments might impact upon the human psyche has long been recognized. Within the geographical tradition, the notion that the physical environment might mold or even control human personality, constitution, and culture has a long history, epitomized by the development of environmental determinism in the early twentieth century. A directly causal, deterministic framing of the relationship between physical environments and human life subsequently fell into disrepute. Other ways of approaching and conceptualizing the psychic impacts of specific environments (including streetscapes, buildings, gardens, and wilderness) have subsequently developed. For example, the situationist movement of the 1960s sought to transform everyday life through experimental interventions designed to foster playful, idiosyncratic, and creative engagement with urban surroundings, in place of normatively controlled behavior in these spaces. This has been taken up in a modern psychogeography movement, which advocates nonpurposeful wandering and observation, especially in urban spaces. It assumes that geographical environments impact upon people's emotions and behavior, but that the potential (and potentially transformative) pleasure and creativity of these impacts are screened out as a result of the overwhelmingly functional, purposeful character of movement.

If environments impact on subjective experience, then these effects may be psychotherapeutic in the sense of alleviating suffering or distress by acting on the human psyche. The idea that particular environments, places, and landscapes might have restorative, curative, or healing properties has led to a body of research on therapeutic landscapes. As the term indicates, this research is not organized according to whether healing is understood to work via the psyche or the soma. Rather it adopts a holistic, nondualistic approach to well being, health, and illness. Sources of healing are often attributed to nature, including spring water, fresh air, forests, flower meadows, rural environments, wilderness, and animals. But these elements are themselves saturated with cultural meanings and actions, resisting any binary distinction between nature and culture. How the modes of healing associated with therapeutic landscapes work (if at all) is the subject of debate. While waters rich in minerals, sunshine, and exercise may be understood to work on the body, numerous appeals are made to spiritual aspects of healing, often including the journeys or pilgrimages undertaken by those in search of relief from suffering. Thus, psychotherapeutics would appear to be important in therapeutic landscapes.

Overlapping with geographical study of therapeutic landscapes is work in environmental psychology. As well as examining how environments influence people, environmental psychology informs the design of built environments, urban parks, nature reserves, and other places with a view to enhancing health and well being. Again, holistic concepts of health and well being are deployed, implicitly invoking psychotherapeutic dimensions to this way of approaching people–environment relations.

If environments of particular kinds, or designed in particular ways, can contribute to healing by psychic means, this suggests that some kinds of people–environment relations may be psychically damaging or unhealthy. For example, for people who suffer from phobias of animals like snakes or phenomena like thunderstorms, these elements of environments generate intense suffering. In vertigo and agoraphobia, relationships to space are the source of powerful distress. In these examples, a variety of ways of reconceptualizing, reordering, or otherwise reframing the relationship between sufferer and source of phobic fear may help to alleviate the intense emotional response to these environmental stimuli. The environment that induces terror is thereby recruited into a new, more psychotherapeutic, role. A related perspective is taken up in the idea of ecotherapy, which attributes mental, emotional, and spiritual disease to a supposed loss of healthy connection to the natural world. Ecotherapy claims to heal by helping people to reattune themselves with, and to love, the natural world. Although seeking to reintegrate human beings and the natural worlds, it relies on a notion of pristine nature extensively criticized by human geographers.

Geographical Study as Psychotherapeutic

Through its focus on the world in which we live, geography and geographers seek the betterment of human life. Insofar as this entails alleviating suffering, geography can reasonably claim to be therapeutic. However, in many cases, this involves using geography and insights from geographical research to raise awareness and to inform policy and practice in ways that have little to do with psychotherapy. Nevertheless, geographical study may work on the student or researcher in ways that entail a form of psychotherapeutic action. This is perhaps most obvious in the case of geographers who are aware that the choices they make about the focus of their work reflect preexisting personal preoccupations. In such circumstances, the research may be undertaken as a way of expressing and working through concerns, and maybe suffering, by psychic means. Examples include explorations of experiences shared by the researcher as well as the research subjects whether pursued through autoethnography, auto/biography, or other means. In other cases, researchers' motivations may not be linked in any overt way to matters of personal concern. However, the psychoanalytic concept of the unconscious suggests that researchers may be drawn to particular topics or approaches for reasons beyond their conscious awareness. Perhaps without knowing it, we sometimes use our research to work through unconscious dynamics. In this sense too, therefore, doing geography may be psychotherapeutic.

A different kind of psychotherapeutic possibility for geographical study entails the use of psychotherapeutic ideas to inform the concepts and practices of geography. There are three fields in which such possibilities have been explored. First, psychoanalytic geography is informed primarily by psychoanalysis but primarily as a theory rather than a practice. In some contexts, such as postcolonial geography, theoretical ideas are mobilized in ways that actively avoid engaging psychotherapeutic possibilities in ways that might be construed as psychoanalytic subjectification. However, elsewhere the difference between theory and practice is more actively blurred, for example, in this author's own attempt to offer a psychoanalytically inflected autobiographical account of stages on journeys between human geography and psychotherapeutic practice.

Second, arguments have been made for the relevance of psychotherapeutic ideas to the field of emotional geographies. In this context it is argued that psychotherapeutic concepts provide useful resources for emphasizing and exploring the relationality, betweenness, and intersubjective mobility of emotion. Psychotherapeutic ideas also hold the potential to contribute to performative approaches to emotion, in which the articulation, expression, and performance of emotion are understood not as representing interior emotional states but as doing emotion or affect. In so doing, psychotherapeutic ideas contribute to rapprochement between geographical work that explores affect from the perspective of nonrepresentational theories and emotional geographies that focus on narrative accounts of personal experience.

Third, reworking and developing Steven Pile's earlier call to consider the relevance of the psychoanalytic concepts of transference and counter transference for interpretative human geography, psychotherapeutic ideas have been used in relation to debates about methodology, including aspects of data generation and analysis. Examples include the use of nonverbal methods to access preconscious or unconscious experiences, and consideration of transferential dynamics between interviewers and interviewees in the conduct and analysis of research encounters.

See also: Agoraphobia; Autoethnography; Dialectical Reasoning and Dialectical Materialism; Emotional Geographies; Emotional Knowing; Feminism/Feminist Geography; Health Systems and Health Services; Healthcare Accessibility; Historical-Geographical Materialism; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Mental Health; Nature; Psychoanalysis; Psychoanalytic Theory/Psychoanalytic Geographies; Self-Other; Situationism/Situationist Geography; Space II; Subjectivity; Therapeutic Landscapes; Wellbeing.

Further Reading

Bingley, A. (2003). In here and out there: Sensations between self and landscape. *Social and Cultural Geography* 4, 329 345.
Bondi, L. (1999). Stages on journeys: Some remarks about human geography and psychotherapeutic practice. *The Professional Geographer* 51, 11 24.
Bondi, L. (2003). Empathy and identification: Conceptual resources for feminist fieldwork. *ACME: International Journal of Critical Geography* 2, 64 76.
Bondi, L. (2005). Making connections and thinking through emotions: Between geography and psychotherapy. *Transactions of the Institute of British Geographers* 30, 433 448.
Bondi, L. and Fewell, J. (2003). "Unlocking the cage door": The spatiality of counselling. *Social and Cultural Geography* 4, 527 547.
Cushman, P. (1995). *Constructing the Self, Constructing America. A Cultural History of Psychotherapy*. Cambridge, MA: Perseus Press.
Davidson, J. (2003). *Phobic Geographies: The Phenomenology and Spatiality of Identity*. Aldershot: Ashgate.
Gagen, E. and Linehan, D. (2006). Guest editorial. From soul to psyche: Historical geographies of psychology and space. *Environment and Planning D: Society and Space* 24, 791 797.
Oliver, S. (2003). Geography's difficult engagement with the psychological therapies. *Social and Cultural Geography* 4, 313 321.
Philo, C. (2004). *A Geographical History of Institutional Provision for the Insane. From Medieval Times to the 1860s in England and Wales*. Lampeter: Edwin Mellen Press.
Philo, C. and Parr, H. (2003). Introducing psychoanalytic geographies. *Social and Cultural Geography* 4, 283 293.
Pile, S. (1991). Practising interpretative geography. *Transactions of the Institute of British Geographers* 16, 458 469.
Robin, K. and Gesler, W. (eds.) (1998). *Putting Health into Place: Landscape, Identity and Well Being*. Syracuse, NY: Syracuse University Press.
Rose, N. (1990). *Governing the Soul*. London: Routledge.
Williams, A. (ed.) (1999). *Therapeutic Landscapes*. Lanham, MD: University of America Press.

Relevant Websites

http://www.allaboutpsychotherapy.com
All about psychotherapy.
http://www.ahpweb.org
Association for Humanistic Psychology.
http://counsellingresource.com
CounsellingResource.com.
http://www.freud.org.uk
Freud Museum.
http://www.iarpp.org
International Association of Relational Psychoanalysis and Psychotherapy.
http://www.ipa.org.uk
International Psychoanalytical Association.
http://www.psychotherapyarena.com
Psychotherapy and Counselling Arena.
http://www.freudarchives.org
Sigmund Freud Archives.
http://www.psychotherapyresearch.org
Society for Psychotherapy Research.
http://www.pce world.org
World Association for Person Centred and Experiential Psychotherapy and Counselling.
http://www.worldpsyche.org
World Council for Psychotherapy.

Public Good

K. R. Cox, The Ohio State University, Columbus, OH, USA

Glossary

Commodification The process of transforming some use value into an exchange value. Uncongested streets are a use value and once appropriate technology is installed in cars, access to them is something that can be sold.

Externalities (positive, negative) Sometimes known as unintended side-effects of production or consumption, the producer or consumer receives neither compensation for them from those who benefit (e.g., the landowners who benefit from the enhanced accessibility resulting from the construction of a new subway station); nor do they compensate those who suffer from them (e.g., the adverse health effects of air pollution).

Locational Conflict Resistance to some impending location, public or private, as a result of the negative external effects the land use is likely to produce.

Market Failure A situation in which there is demand for some product or service but suppliers fail to emerge since provision would be unprofitable. This is typically owing to some degree of publicness in their consumption.

Public Goods Goods and services which, when produced, enjoy publicness in their consumption. Their consumption cannot be restricted to those who have agreed to pay for them.

Welfare Geography A geography of well-being typically constituted by variations in state provision, including education, and in the presence or absence of various externalities like air pollution and the congestion of public facilities; income variations are sometimes included.

Definitions

The idea of goods or services as 'public' rather than 'private' emerged in the context of attempts to understand, on the one hand, what was called 'market failure' in the provision of goods and services; and on the other, to make sense of that range of goods and services that tend to be provided by the state. Under conditions of market failure, markets, literally, do not form and so there is a problem of provision. This can be particularly acute where the production of private goods or services depends on that provision; that is, where goods and services that can be and are produced for markets depend for their production on goods and services that 'cannot' be produced for markets. Accordingly, much of what the state does is providing conditions of production that capital cannot provide itself. In order to understand the idea of market failure, consider the following examples:

- In many American cities, there is a demand for light rail service. Its attractions lie in its capacity for reducing commuting times and costs and congestion. However, it is likely to be underprovided by private firms since not all the benefits it provides are ones for which it can collect revenue. Consequently, profitability is less than it might be. Stations provide enhanced access to those living close by and conceivably this will influence what people are willing to pay for the service. However, landowners also benefit and can extract increased rents from land in the vicinity of stations. Yet this is not a benefit that the light rail company can charge them for. So revenues are less than they might otherwise be and the incentives to providing light rail service appropriately reduced, perhaps to the point that no service is provided at all.
- Consider now the case of air pollution from heavy industry. There is a good deal of evidence that it adversely affects the health of people living in the vicinity. Yet there is no market in clean air. The firms in question enter into markets in order to secure the labor power they need, along with the raw materials and instruments of labor. But there is no market in air as a receptacle for their effluent. One result is that their production costs are reduced and their profitability increased. So clean air has certain public good characteristics; if it is provided, its consumption, even where that consumption degrades it, as in the case of the smoking chimney, cannot be restricted, for example, by charging for it. Like light rail, markets fail to emerge for the provision of clean air, and state intervention to control pollution, subsidize light rail, or whatever, is warranted.

In both cases, there are what are called 'externalities' at work. In the light rail case, the putative private company would provide benefits to landowners that it could not internalize: these are what are known as positive externalities. The benefits do not appear as revenues in the firm's accounts and so profitability is less than it might otherwise be. In the air pollution case, the industrial firms are imposing costs on people, in the form of health costs, which they do not compensate them for: costs that, unlike labor costs, raw material costs, do not appear in their accounts. These are negative externalities.

Again, in both cases, geography enters into our understanding. Air pollution has a geography, and so too, by definition, does access to light rail stations. There will also be various knock on effects of a broader geographic scope. In each instance, housing markets over larger areas of the city are likely to be affected as the overall geog raphy of residential desirability is altered. In other cases of public goods, however, or rather of their degradation, the effects may be less limited in their geographic scope. The obvious case here is global warming where market failure in the sale of the atmosphere for the purpose of absorbing carbon will have effects on sea level that are world wide. This has implications for the geographic scale at which regulation should occur.

There are also, just as clearly, different degrees of publicness in the consumption of various 'goods' and 'services'. The benefits of light rail transit can obviously be internalized by the provider to a substantial degree; and if they were able to purchase the land around the (planned) stations prior to construction, or prior to news of construction leaking out, then the enhanced rents could also be appropriated, allowing more of the benefits provided to be internalized. The same goes for something like education. It has some features of a private good in that its benefits can be internalized in the form of enhanced career prospects and therefore wages. But some might argue that it also has some public features through instilling civic virtue and through the enhanced taxability of the wealthier. There is, in consequence, a public interest in providing education.

The carbon content of the atmosphere, on the other hand, with all its implications for the earth's energy budget and therefore for global warming, is as pure a public good as it is possible to imagine. This is because the privately appropriable benefits that accrue to a firm that engages in reduction of carbon emissions are so in finitesimally small. Rather the benefits are almost entirely public and beyond internalization by the firm in question. Yet recent attempts to create a market in rights to use the atmosphere as a repository for carbon through the idea of carbon credits, suggest that this might be a premature conclusion; rather that the degree to which a good is public or private depends on the degree to which it has been commodified so that it can enter into market transactions. Some goods might seem to be more easily commodifiable than others. Capitalist societies, however, have proven extraordinarily adept in pushing back that particular frontier. Among other examples, one thinks of the way in which recording instruments in automobiles can connect with stationary monitoring equipment, so allowing a market in urban space to emerge with the consequence of forcing motorists to internalize the costs of the congestion that they impose on other motorists. So it would seem that whether a good is more public than private or vice versa is far from set in stone. This is not just a matter of technology. Institutions are also import ant as in the attempt to create a world market in carbon credits.

One thing that is already clear is that the distinction between the public and the private is a very slippery one indeed. This is not just because what seemed to be public can become private, or at least, more private than it was. It is also the fact that the provision of what we regard as private goods depends upon the provision of ones that are more clearly public. If private goods are to be produced, then there have to be institutions to protect private property and enforce contracts, and that provision cannot be private since it is itself subject to market failure. Typically, of course, this is where the state steps in, though we should be careful not to reduce public provision to that particular mode. The state has not been alone in providing those conditions for capitalist production that capital cannot provide itself. Historically the reproduction of labor power has depended sub stantially on the institution of the family as well as on mutual aid mediated by wider circles of kin and friend ship. Without the fact of trust in the relations of firms one with another, with their workers, and consumers, an economy based on exchange would break down entirely. The recent literature on governance also underlines the danger of a simple state market polarity in talking about public goods.

Even so, the state in the advanced capitalist societies is hugely important in the provision of a wide range of goods and services, including highways, education, financial support for the unemployed, public health and in some instances housing, mass transit, airports, public parks, to mention only a few. As recent rhetoric around neoliberalism and pushing back the boundaries of the state reminds us, however, the identification of what is public and therefore appropriate for the state as opposed to private provision is highly contested. This is not just a matter of changing technology and insti tutional possibilities as might have been implied by the discussion here of urban congestion and carbon emissions. Rather it depends a great deal on just how desperate capital is to widen its opportunities for profit making, and this is something that varies over time in accordance with the rhythms of capital accumulation on a global scale.

Applications in Human Geography

As far as human geography is concerned, the idea of public goods and the drawing out of their geographic implications makes its historic entry early in the 1970s. The background to this was the interest in social rele vance. This was in part and in turn a critique of the spatial quantitative work of the 1960s and its inability to cope with issues of power or the state; but also, and quite indissolubly, it was a response to the sense of an urban problem subsequent to a heightened awareness of inequalities in the city brought about by the civil rights movement. Human geography was not alone in this. The field of urban economics underwent vigorous growth during that period, and from virtually nothing, and while it was less radical in its approach than what ensued in human geography, particularly in its public choice incarnations, its emphasis on externalities, public provision, and the proper role of the state was an important catalyst.

With respect to that period and the enlivening of work on a more socially relevant human geography through the introduction of ideas about public goods, state provision, and externalities, two names stand out: David Harvey and Julian Wolpert. Harvey's direct contribution was more fleeting, corresponding to what he termed in his book *Social Justice and the City*, his liberal phase. Of particular interest is his paper on urban form and the redistribution of real income which appeared as Chapter 2 in his book. It testifies to a quite profound immersion in the urban economics literature of the time and an ability to draw out its implications for the politics of geographic change in the city. It was largely through Harvey that the idea of externalities entered the geographic literature.

Like Harvey, Wolpert had been a central figure in the spatial quantitative revolution, but something of a maverick, calling into question some of the underlying social assumptions of that work. Initially that had to do with the impossibility in a geographically differentiated world of the assumption of perfect information. The *post hoc* predictions of location theory just did not seem to work. By the early 1970s, and in the same vein, he had turned to the question of how power relations condition location and it was through him that the important idea of locational conflict entered the literature.

Through the work of Wolpert and his students, like Michael Dear and Jennifer Wolch, locational conflict became one of the more enduring ways in which the idea of externalities and therefore public goods assumed a role as part of the conceptual apparatus of human geography. The idea was a simple one but with important ramifications. Wolpert's attention had been originally drawn to the politics surrounding the urban renewal and highway projects of the 1960s. People were impacted by them in their residential neighborhoods and this often resulted in opposition, and then attempts on the part of the state to co opt and defuse. Apart from the fact that the idea was a nice rebuttal of the frictionless location process implied by location theory, on reflection it was found to have substantive applications of considerable breadth. Elaine and Julian Wolpert, for example, went on to show how it could be applied to the concentration of halfway houses for those released from mental institutions. Still later, it was to resonate with what became known as NIMBY ism and was developed by Robert Lake in new ways; Lake's contribution was to show that conflict over the location of land uses regarded as nuisances depended on the degree to which those land uses were needed, and therefore on the 'production' of garbage, a population of the homeless, and so on: in short, productions of external effects that were subject to social control.

A second development in the 1970s that clearly followed on from the introduction of ideas of public goods was that of welfare geography. This idea was already implicit in Harvey's essay on urban form and the redistribution of real income. Externality fields had obvious and divergent effects on people's well being. Market forces, particularly the real estate market, mediated access to those externalities and tended to reinforce the inequalities resulting from the distribution of the product among profits, rents, and most variably, wages.

An additional layer was added by variations in public provision: variations, that is, in the provision of education, public safety, and other services like street lighting and recreational facilities. In the American metropolitan area these variations could be considerable as a result of a quite distinctive and intense jurisdictional fragmentation. Again, the real estate market played a part in access to these public goods: public in the sense that they were not provided through the market, there were difficulties in such provision, and the state had assumed responsibility. Housing values were bid up where the tax rate/public provision combination was more favorable and bid down, so allowing purchase by poorer families, where it wasn't. Accordingly geography entered into an understanding of welfare geographies not simply through the fact of differentiation but also through the movements, particularly residential, that mediated access.

Much of this literature, however, was of a highly descriptive nature. Seemingly the major objective was an identification of the variations, and how poor people were the ones who ended up with the least favorable bundles of public goods and in those neighborhoods most exposed to the negative externalities of air pollution, traffic congestion, crime, and so forth. The ideas of territorial justice and the importance of introducing some redistributional measures, if only implicit, were typically present. In short, welfare geography was stuck in what one might call, not too unkindly, a liberal rut. To the extent that there was analysis, it was in terms of the dynamics of a market economy and the role that competing local governments and neighborhood organizations played in attracting land uses with positive externalities and pushing the negative elsewhere. This was useful and certainly helped shed light on Harvey's old problem of the distribution of real income. It also endures in work on environmental justice. But it had quite severe limits and assumed more than it should have done, as shown in this article.

A third and more recent application has been to the understanding of the politics of local and regional development. This returns us to the complementarity between state and capital; how the state depends on capital accumulation for its revenues, but in turn, without the state, capital lacks some of the conditions necessary for that selfsame accumulation process. It also moves the analysis away from the concern with living place issues that was apparent in the work of Harvey and Wolpert, and indeed, in that on welfare geography. The literature in this area has grown rapidly since the 1980s. A central motif has been the competition for inward investment and the role that 'business climate' plays in attracting it in. The idea of public goods and externalities as essential features of that business climate and something that local government and whatever local governance mechanisms there are should attend to has been more implicit than critically scrutinized, however.

Finally, in this brief survey of how ideas of publicness in provision have affected human geography, we should make reference to works on the state and its territorial structure: not just the fact that the state enjoys constitutionally and statutorily given powers with respect to a bounded area, but that it also operates, to some degree, through what might be called a territorially defined scalar division of labor – local, provincial, and central branches, for example. The fundamental premises here are, first, that states provide goods and services which, in virtue of their public effects, can be provided through markets only with difficulty; and second that the externalities that occur in their provision have geographies that recommend some sort of internalization – via the state – of a similarly geographically restricted form. This is the conception advanced by one of the most frequently cited writers on the topic in geographic literature, the sociologist Michael Mann. Interestingly his understanding is quite close to the neoclassically inspired public choice understandings of the state, as in the work of Gordon Tullock. Mann made no reference to the state's territorially defined scalar division of labor, but the public choice literature certainly did, drawing attention to differences in the externality fields of different activities; so (e.g.) there is a logic to providing local government with powers over the regulation of land use where externality fields are spatially very limited, but not over the provision of limited access highways of national scope. On the other hand, ideas about the implications of jurisdictional fragmentation for facilitating the realization of diverse preferences for public goods, ideas originally attributed to Tiebout, have also played a role in arguments about the appropriate allocation of powers and responsibilities to different levels of the state's territorial hierarchy.

Public Goods, Externalities, and Their Necessary Conditions

Finally in this article, how might we develop a more critical position on the question of public goods, particularly as they relate to human geography? What are the underlying assumptions that are typically taken for granted and left unexamined? And how might public goods, externalities, and the practices surrounding them, relate to broader, more all embracing conceptions of the social, or more accurately the 'sociospatial', process? Several points can be made here.

1. The first is the sociohistoric specificity of public and private goods and the distinction between them. They assume a world of commodification, one in which commodities are produced with commodities, but in which commodification encounters barriers. Not all goods and services can be commodified. Yet it is on the provision of some of these in some form or other that the commodification of other goods and services depends. Initially these would have included a legal system through which to enforce private property rights and contracts. Later, some regulation of the money supply that was above the imperatives of the market place would have been necessary. Historically, we have seen, this provision of goods and services in the noncommodity form has tended to be through the state, though not exclusively.
2. The second point has to do with how we should interpret externalities. The standard view is simply one of market imperfections that can be eliminated. This can be achieved through the provision of appropriate monetary or regulatory incentives or, where positive externalities are of such a magnitude as to preclude private provision, assigning production to the state. There are alternative understandings. The one which is explored here is a reinterpretation in terms of contradictions within the capitalist mode of production. The capitalist mode of development is characterized by an unprecedented socialization of production. There is an extensive and intensive interdependence mediated, among other things, by the division of labor and by means of production that are worked on or with, collectively. It is through that socialization that productivity, and so production, are enhanced. This social character, however, the development of all round interdependence, enters into contradiction with the way in which the product gets appropriated, that is, privately. The structurally given goal of the capitalist is not to facilitate the net revenues of others but to enhance his/her own. To the extent that production generates negative externalities, then they are to be dumped on others rather than internalized, and regardless of how this affects the

social character of the conditions on which individual production depends, such as the reproduction of some shared resource. Where the use values produced cannot be sold or sold only with difficulty, but they are nevertheless necessary conditions of production, then, and assuming a logic of individual profit maxi mization, they will not get produced.

Examples are not difficult to find. They would include:

- Issues in the reproduction of labor power: These are a mixture of negative externalities like pollution and health and safety in the workplace, and market failures in the provision of use values; use values that, because they benefit other capitalists do not get produced. Capitalists systematically under invest in the training of workers given the fact that workers are free to leave, take their talents with them, and so provide other capitalists with a free good. Likewise, capitalists do not support workers when they are unemployed. They want workers to stay around until business picks up again, but there is no guarantee that former workers who they might well have otherwise supported will not be rehired by other firms.
- The commons: The overpumping of aquifers, over fishing of the oceans, and overuse of the air as an absorber of effluents are well known examples. Since in these instances an essential condition of production is free, profitability is higher, encouraging increased investment and use, and so the erosion of the common resource base itself. In other cases, common resources can result in under investment as in neighborhoods and disincentives to rehabilitation of properties.

Just as clearly, this contradiction between the social character of production and the private character of ap propriation tends to get suspended. Shared conditions of production subject to market failure in their provision do tend to get provided. There are regulatory initiatives designed to limit the erosion of the commons. The state is of major significance in all these instances, though the role of various forms of inter firm governance should also be recognized.

But the important qualifiers are 'tend' and 'suspend'. The public and the private, the sphere of the production of goods and services in their public form and the private economy, exist in a relation of considerable tension. While capital requires state intervention if it is to accu mulate, it also resists it, typically producing the de graded, inadequately funded public sector of which Galbraith complained so many years ago in *The Affluent Society*, but also in more class specific forms than he would have admitted. The general support of the labor movement for greater state intervention, and historically for the planning that promised to internalize externalities by eliminating the hostile relations between firms, and between firms and labor, is not coincidental. Even so, the discursive environment has tended to work to the advantage of capital. The powerful notion of money as the measure of all things – if contested by, among others, the environmental movement – has been an essential part of the debate about state regulation and provision. The emphases and silences of cost–benefit analysis are eloquent testimony to this and can be turned to capital's advantage when it demands state intervention as well as shedding doubt in those instances where it is resisting it.

Furthermore, the contradiction is never abolished; it is merely suspended, only to be displaced somewhere else, some other time, and in some other form. Shifting pro vision into the public realm does not eliminate negative externalities. State agencies have their own agendas and their different policies often work counter to one another, as Theodore Lowi demonstrated. And the very fact that the state must obey capital's laws of motion if it is to tax and so obtain the revenues necessary to carrying out its functions means that internalization in one place will be accompanied by externalization onto another. Through land use planning, local governments can regulate land use so that synergies are exploited and the incompat ibilities that would lower land values are avoided. But local governments have no incentive to make their plans compatible with those of their neighbors.

See also: Neighborhood Change; NIMBY; State; Planning, Urban; Welfare Geography.

Further Reading

Coates, B. E., Johnston, R. J. and Knox, P. L. (1977). *Geography and Inequality*. New York: Oxford University Press.
Cox, K. R. (1979). *Location and Public Problems*. Chicago: Maaroufa Press.
Cox, K. R. and Johnston, R. J. (eds.) (1982). *Conflict, Politics, and the Urban Scene*. Harlow: Longman.
Davis, O. and Whinston, A. (1965). Economic problems in urban renewal. In Phelps, E. S. (ed.) *Private Wants and Public Needs*, pp 40 153. New York: W. W. Norton.
Harvey, D. (1973). *Social Justice and the City*, ch. 2. Baltimore: Johns Hopkins University Press.
Lake, R. (1992). Structural constraints and pluralist contradictions in hazardous waste regulation. *Environment and Planning* A 24, 663 681.
Mann, M. (1984). The autonomous power of the state. *European Journal of Sociology* 25, 185 213.
Smith, D. M. (1973). *The Geography of Social Well Being*. New York: McGraw Hill.
Tullock, G. (1970). *Private Wants, Public Means*. New York: Basic Books.
Wolpert, J. (1970). Departures from the usual environment in locational analysis. *Annals of the Association of American Geographers* 50, 220 229.
Wolpert, E. and Wolpert, J. (1974). From asylum to ghetto. *Antipode* 6(3), 63 76.

Public Policy

M. Perry, Massey University (Wellington), Wellington, New Zealand

Glossary

Applied Geography The application of geographical knowledge and skills to the resolution of real world social, economic, and environmental problems.
Deep Policy Analysis Analysis that questions the assumptions on which policy intervention is based and that is more likely to be used by policy makers to understand policy impacts than to help design new policies.
Policy-Driven Research Research instigated by policy makers to understand particular public policies.
Policy Geography Geographical enquiry for the specific purpose of addressing a policy problem identified by policy makers.
Policy Maker Politicians and government agency employees who have command of the institutions that give approval to public policy actions.
Policy-Relevant Research Activity decided upon by the researchers themselves that may involve critiques of public policy and support to groups and individuals seeking to change the existing public policy agenda.
Public Geography Geographical enquiry that involves engagement with individuals and groups that are seeking to influence the policy made by public agencies.
Public Policy Action which employs the authority of a government agency to commit resources in support of a preferred value or activity.
Shallow Policy Analysis Analysis that accepts the aims and assumptions of policy makers.

Introduction

Public policy is ultimately the responsibility of government agencies that make choices to do something or to do nothing affecting the community or some section of it. Groups outside of government contribute to public policy by recommending courses of action for public agencies to act upon and possibly by campaigning or otherwise seeking action on their recommendations. Whether the recommended actions are actually taken up is a measure of the influence obtained but the absence of implementation does not preclude the ability to claim relevance to public policy. Consequently geography may be judged to contribute to policy when it offers a solution to a problem requiring the attention of government irrespective of whether that solution is acted upon. Over the long term, the public policy ambition could be said to have failed if there was a consistent inability to influence policy decisions. Over the short term, policy related geography can be distinguished from other forms of geography in going beyond the understanding of a phenomenon of social concern to guidance on what action should be taken in relation to that issue.

The application of geography to policy is generally viewed as desirable and something that geographers should actively seek to encourage. Nonetheless the relationship between geography and policy is subject to much debate partly as there is no agreement over what counts as a contribution to public policy. Three audiences have been recognized:

- At the top of the policy making hierarchy are the politicians and officials linked to central government agencies whose actions potentially have national and international implications. As well as serving these decision makers and policy designers, geographers may assist outside groups vying for influence over 'high level' policy.
- Public policy making is also a function of local and regional government agencies that operate within structures set by national government. A trend in many countries has been for more policy to be devolved to second and third tier levels of administration. Geographers may direct their policy guidance to these agencies and the associated interest groups focussed at this tier of activity.
- A 'domestic' audience for public policy has been envisaged in terms of the institutions and individuals associated with the places where geographers themselves work. Addressing this internal audience can be justified as a case of 'putting one's house in order' before addressing others. This argument believes, for example, that geographers who recommend greater wage equality in their advice to policy makers in central government might have more credibility if they have first addressed this issue in their own workplace.

Some evaluations have criticised geographers' contribution to public policy for giving too much. On the other hand, if policy influence is restricted to the lower tiers of policy making, there can be an implication that it is of lesser significance to that directed toward national policy. That implication is even greater if only internal influence is sought and consequently national policy makers are generally seen to be at least one of the audiences to aim for either directly or by supporting others who aspire to change national policy settings.

Reasons for and against Policy Connection

Gaining influence over policy making substantiates the claim that human geography is a critical social science generating insight into economic, environmental, and social issues. In this context, engagement with policy is viewed as a way of making sure that geography remains relevant to society at large rather than being a subject of interest only to other geographers. Some geographers go further and argue that policy related research is valuable in its own right as compared with it being viewed as a secondary or ‘grey’ area of activity of lesser status than ‘theorizing’. As well as geographers’ motivation to be policy orientated, the demand from policy makers also affects participation in policy related geography. Since the 1990s, policy makers in many countries have espoused increased interest in so called ‘evidence based’ policy and this implies an increased interest in basing policy decisions on research data. More generally, governments tend to expect universities to make a contribution to the community, and engagement in public policy development is one way that academics including geographers can meet this expectation.

At the same time, it is important to recognize that some geographers say policy related research should be avoided. One concern is that it risks geography being hijacked or subverted by the organizations that sponsor or utilize the work. This might arise where policy related research is commissioned by agencies that limit the scope of investigations over what geographers might have preferred and which then control the release of findings and the researcher’s ability to comment on them. Such a situation can be mitigated where there is a diversity of policy agencies sponsoring research. For some ‘radical’ geographers such protection is insufficient as policy agencies tend to support the status quo, whereas geographers should work with a wider agenda including the fundamental transformation of society. This requirement was most famously made in a landmark article by the influential geographer David Harvey who encouraged geographers to ask “what kind of geography for what kind of public policy.” Harvey’s critique has been influential, encouraging many geographers to prefer to absorb new philosophical approaches and debates from social and cultural theory rather than directing attention to policy questions within the scope of existing agencies.

Geographers aiming to reform rather than transform the fundamentals of society have at least four ways to make public policy contributions. First, they may apply existing geographical theory or techniques to help understand and solve ‘real world’ social, economic, or environmental problems. Second, geographers may undertake research that evaluates public policy initiatives or that tests the assumptions guiding policy interventions. Third, geographers may be contracted by public agencies to investigate an issue of concern to policy makers. Such a relationship may give direct opportunity to influence a matter of immediate relevance to policy makers, as well as encouraging wider dialog between public officials and geographers. Four, policy makers may be professional geographers, as when academics are appointed to the boards of public bodies, or have completed a geography qualification before joining a government agency.

Despite multiple channels for participation, many geographers who want to see geography influence public policy believe that the discipline’s contribution has fallen short of what it should be. One obstacle is that exerting influence over policy requires understanding of the policy making process and this draws on skills and behaviors that are more likely to be possessed by political actors than geographers. This challenge is one that geographers potentially share with academics in other disciplines too and so it can provide only part of the explanation. Two main reasons can explain why geography has had a relatively minor influence over public policy: policy makers may view geographical theories and techniques as less helpful than those from other disciplines or geographers may not make sufficient appropriate effort to make their discipline relevant to public policy. Some mix of these explanations may also explain the impact obtained but it is helpful to examine each possible explanation in turn.

Policy Makers Exclude Geography

A preference for guidance from other academic disciplines than geography is one reason policy makers tend to sideline the discipline. Among academic disciplines, public policy draws most on economics for guidance and especially welfare economics. Welfare economics is based on the notion that individuals, through market mechanisms, should be relied upon to make most social decisions. The methods of analysis used by welfare economists recognize instances where markets cannot be relied upon to distribute resources efficiently. To use economists’ language, there are times when markets cannot aggregate individual utility maximizing behavior (meaning decisions to purchase or nor purchase some good or service) so as to optimize overall social welfare. When this occurs, markets are said to fail and there is a justification for government actions to supplement or replace markets.

Welfare economists recognize many sources of market failure, but there are a number of archetypal situations that are used to justify a need for public policy.

- Natural monopoly: in industries with large capital requirements and large economies of scale, such as railway networks and power generation, it can be difficult for new businesses to compete with estab lished ones. Where one or a few firms dominate an industry, the lack of competition can reduce the ability of individuals to make welfare maximizing decisions.
- Imperfect information: to maximize welfare, indi vidual decisions must be made on the basis of ad equate information but there are instances where either producers have no incentive to reveal infor mation about their product or service or consumers do not have the expertise to evaluate the information given and instances where both failures arise.
- Externalities: these arise in relation to those costs of production that are not paid for by the producer and in relation to those benefits that are not paid for by buyers of the producer's output. Externalities are an incentive to overproduce some things, where costs are avoided, and underproduce things where benefits are not fully remunerated.
- Common property goods: resources that are open for anyone to draw upon risk being used unsustainably where individual decision makers maximize their current use of the resource without regard to its long term future. This context is known as the tragedy of the commons.
- Destructive competition: competition between enter prises that causes negative side effects for employees and society may be seen as a market failure where inadequate wages and workplace investment produce living and working conditions with high social costs.

As well as guidance on when policy makers may need to act, welfare economics is appreciated for its ability to inform the selection of an appropriate response. A further great advantage of welfare economics is that it can appear to offer a neutral and dispassionate justification for policy action. Policy makers are recommended to act because a market is failing and society will work more efficiently if corrective action is taken. This does not mean that wel fare economics reduces public policy to a technical process. It is frequently difficult if not impossible to convert real world situations into the data needed by welfare economists to run their forms of analysis and even where they can it should not be imagined that policy makers are 'won over' simply by evidence of a market failure. Policy choices are bound by political in stitutions and made by political actors, often in response to political pressures. Neither should it be overlooked that there is much internal debate among welfare economists about the recognition of market failures and what, if anything, should be the response. Nonetheless, the market offers a benchmark and a rationalizing logic justifying policy intervention. In contrast, geographers tend to believe that policy recommendations must be justified by explicit statements of the researchers' normative values such as their position on human rights, equity, and local democracy.

Of course, geographers are free to draw on economics and frequently do so but they tend to be uncomfortable with the assumptions associated with welfare economics. For geographers, there is a preference to see markets as socially constructed phenomena created by actors whose perceptions and resources are a product of particular circumstances. Geographers devote more of their effort to understanding how the environment has constructed a market rather than analyzing whether the market is failing and how it should be corrected. Policy makers, on the other hand, tend to work with markets as they exist and from their perspective questioning of how a market has come about and how it might be constituted differ ently is not always helpful to their cause.

Shallow and Deep Policy Analysis

The scope of potential advice has been discussed in terms of the difference between shallow and deep policy an alysis. Shallow policy research involves evaluating the impact of policy interventions and can be completed using standardized impact assessment methodologies. The aims and objectives of the policy research are pro scribed by the agency seeking the evaluation leaving little scope for exploration of the issue outside the agency's narrow interests. Deep policy geography, on the other hand, is not restricted to the parameters and exclusions set by a policy agency and recognizes that the policy process itself is a worthy target of investigation, as well as policy outcomes. Whereas deep policy analysis poten tially offers the greater insight, much of the demand for policy research is concerned only to obtain a shallow analysis.

Some geographers feel that other disciplines are more willing to engage in shallow policy research than are geographers. Geographers are, for example, frequently preoccupied with demonstrating the local embeddedness and path dependency of economic and social phenom enon. Compared with economists they are also more likely to wish to collect their own data and to do this locally. Deep policy analysis may generate unique in sights into the 'real' nature of social and economic pro cesses. It can also be viewed as knowledge that is more relevant to policy implementation issues and program outcomes than to the 'front end' of the policy process where policy is first conceived and designed.

Geographers Exclude Policy

The dominance of economics presents a challenge but does not exclude geography extending some influence over public policy. Indeed there is a long tradition of practical application and policy relevance, going back to the involvement in cartography and mapmaking. From mapmaking has grown geography's contribution to the development of remote sensing and geographical information systems that permit increasingly sophisticated and analytical ways of collecting and interpreting spatial information. Such work forms part of 'applied geography' and can be contrasted with 'pure geography' that is focused on the development of theories, ideas, concepts, and methods. A synergistic relationship between these two sides of the discipline, with the development of theory and practice informing each other, would help to maximize the contribution of geography to policy making. In practice, the two sides of geography have often existed as alternative endeavors rather than operating as a smooth continuum of activity. This lack of connection has provided one context in which the geographical profession itself has responsibility for the influence it exerts over public policy.

An idealized relationship between pure and applied geography envisages two way interactions. Pure researchers develop principles and methods that are picked up by applied geographers to provide the framework for asking questions about an issue of policy relevance and to provide a normative standard against which current and future social conditions may be judged. In the opposite direction, applied research that tests theories and methods in the 'real world', arguably the ultimate proving ground for new ideas, offers a further means for refining theory.

Opportunities for mutual learning exist and are sometimes acted upon but practical considerations and philosophical objections have acted against ongoing engagement between pure and applied geography.

Practically, there tends to be cycles in the relative importance of pure and applied geography rather than an ongoing interaction. At any point in time, activity has concentrated on either the pure or applied forms of geography leaving a neglected cousin whose fortunes change more from perceived over indulgence of the other side of the family than through the transfer of insight and skills. Starting in the late nineteenth century up to the 1990s, Michael Pacione drawing on a schema first development by Peter Taylor identified three periods when applied geography was in the ascendancy (the late nineteenth century, the period from the end of World War I to the start of World War II and the mid 1980s) separated by two periods when pure geography was dominant (the early twentieth century and from 1945 to the 1970s) and speculated that post 1980 the discipline had shifted back to pure geography. That speculation seems to have been confirmed by a growth of comment since the end of the 1990s on the need for geography to demonstrate its policy importance.

Cycles in the focus of academic enquiry have a number of potential origins. One interpretation is that they link to the state of the world economy. Thus during the 1930s and 1980s, at times of economic recession, applied geography flourished in fields such as land use planning and regional development. During periods of national economic expansion it appears that geographers may have felt less obligation to work in applied areas or had more opportunity not to do so. Such connections are interesting to explore but other explanations should not be overlooked as well. When one side of the discipline attains a measure of dominance there is perhaps an inevitable reaction in demands to redirect activity. This certainly seems to have occurred in the debate about the relevance of geography that took hold in the 1970s. Since the 1960s, leading edge human geography was dominated by increasingly abstract forms of spatial analysis, such as point pattern analysis, nearest neighbors, and factorial ecology. Critics of this form of work pointed to a growing disparity between these kinds of sophisticated techniques and understanding of how real places and social groups were changing. The call to make geography 'relevant' was responded to with a surge of interest in themes such as territorial social indicators, territorial social justice, and the need to match the provision of public facilities with need.

The calls to make geography relevant illustrated another source of the separation of pure and applied geography. During the 1960s, the development of computers facilitated more complex statistical modeling than was previously possible and provided researchers with many opportunities for novel investigations. Disciplines tend to develop through such waves of intellectual ferment rather than in a steady incremental fashion. During a period when some exciting new area of theory, ideology, or method is opening up it is perhaps inevitable that the applied role of geography assumes lesser status. The decade from the mid 1990s was another time of intellectual fervor as human geographers were drawn into debates around postmodernity and new cultural perspectives, once again leaving relatively few geographers to progress policy related studies.

A lack of integration risks pure research delivering theories and concepts that are too underdeveloped for practical application. An influential criticism of the performance of geographers has been made on this basis by Ann Markusen. She has argued that much contemporary economic geography works with 'fuzzy concepts', defined as entities, phenomena, or processes that possess two or more alternative meanings. Reliable identification or application by different readers or researchers is difficult when dealing with fuzzy concepts. New concepts still under development may inevitably have a degree of fuzziness. They can be deliberately vague so as to appeal to a range of different audiences. For Markusen, of more significance is that a lack of desire to guide action has allowed standards for admissible evidence to drop too low. A tolerance of under researched fuzzy concepts commences a downward spiral of declining precision and evidential confirmation and then increasing distance from the policy community. Markusen's judgment that the key concepts exciting economic geography, such as flexible specialization and its associated spatial tendency of reagglomeration, have remained fuzzy is controversial but she does identify a genuine dilemma that geographers have faced. Geographers must be explicit about the values and priorities guiding recommended courses of action if they are to address policy. On the other hand, declarations of political preference can be seen to compromise the claims to be conducting 'scientific' investigation. Rather than deal with this potential conflict, Markusen argues that many geographers stay away from close engagement with the policy community.

Public Geography and Policy Geography

While there is much support for the idea that geography should produce useful knowledge, there are different perceptions as to the form that it may take. Drawing on the ideas of the sociologist Michael Burawoy, the UK geographer Kevin Ward has made a distinction between policy geography and public geography. Policy geography is conceived as a comparatively narrow activity directed to specific ends that are determined from the outset, whereas public geography is a more diffuse activity distinguished by the extent to which academics seek to engage with the community outside universities. This engagement may be through expanding the range of outlets for their intellectual products, such as writing in magazines, or by working with area based or other interest groups where the process of mutual learning may be the outcome as much as or instead of any particular campaign objective. Whereas policy geography is judged on its effectiveness to policy makers in helping design or evaluate an individual intervention, public geography is judged by its relevance or usefulness to groups outside the government's policy making agencies.

The public versus policy geography divide overlaps with a distinction between policy driven and policy relevant research. The former is targeted on specific government policies and has a purely instrumental motive related to the introduction or modification of a particular policy stance. Activity that is both policy driven and policy directed involves geographers only after it has first been identified as a policy issue by others. In this work, geographers may have little control over the agenda particularly if their involvement is as contractors to the agenda setters.

Policy relevant work may involve critical evaluation of public policy and activity supporting groups that are challenging public policy agencies such as trade unions, unemployed, or think tanks working to a different agenda than government agencies. With policy relevant geography, the decision to pursue an investigation or advocate on its behalf is made by geographers themselves and for this reason alone is more likely to characterize the work of university based geographers than policy driven research. Policy relevant activity can include the engagement geographers have with campaign groups linked to their area of geographical enquiry. For example, a geographer interested in the location of the defense industry and the fate of communities dependent on defense industry funding may be said to be engaged in policy relevant activity when supporting peace groups or giving public talks on the issue of converting defense based business to alternative nonmilitary markets. Geography can claim a larger engagement with public policy than is frequently recognized when including this type of contribution.

See also: Applied Geography; Cultural Geography; Embeddedness.

Further Reading

Coppock, J. T. (1974). Geography and public policy: Challenges, opportunities and implications. *Transactions of the Institute of British Geographers* 63, 1 16.

Harvey, D. (1974). What kind of geography for what kind of public policy? *Transactions of the Institute of British Geographers* 63, 18 24.

Henry, N., Pollard, J. and Sidaway, J. (2001). Beyond the margins of economics: Geographers, economists, and policy relevance. *Antipode* 33, 200 207.

Herbert, D. and Matthews, J. (2004). Introduction. In Matthews, J. & Herbert, D. (eds.) *Unifying geography common heritage, shared future*, pp 259 265. London: Routledge.

Markusen, A. (1999). Fuzzy concepts, scanty evidence and policy distance: The case for rigour and policy relevance in critical regional studies. *Regional Studies* 33, 869 886.

Martin, R. (2001). Geography and public policy: The case of the missing agenda. *Progress in Human Geography* 25, 189 210.

Pacione, M. (1999). Applied geography: In pursuit of useful knowledge. *Applied Geography* 19, 1 12.

Peck, J. (1999). Grey geography. *Transactions of the Institute of British Geographers. New Series* 24, 131 136.

Pollard, J., Henry, N., Bryson, J. and Daniels, P. (2000). Shades of grey? Geographers and policy. *Transactions of the Institute of British Geographers. New Series* 24, 243 248.

Ward, K. (2006). Geography and public policy: Towards public geographies. *Progress in Human Geography* 30, 495 503.

Relevant Websites

www.igu gapp.org
IGU Commission on Geography and Public Policy.

www.psi.org.uk
Policy Studies Institute is one of Britain's leading independent research institutes.

Public Space

D. Mitchell, Syracuse University, Syracuse, NY, USA
L. A. Staeheli, University of Edinburgh, Edinburgh, UK

Glossary

Business Improvement District A private association of business or property owners with the power to levy fees for use in property maintenance, security, and promotional activities.

Gentrification The 'upgrading' of property, the in-migration of people of higher classes, and the subsequent displacement of people of lower classes; homelessness is often understood as both a result of and a threat to gentrification.

Heteronormativity The *a priori* ideological and social expectation that a space and its users are 'straight', or heterosexual.

Neoliberalism A system of governance and economy that promotes private interests, including property, over public interests, and that is suspicious of state involvement in the economy. This does not mean that state discipline over marginalized groups is necessarily moderated.

Neosituationists Activist organizations dedicated to the 'anarchic' decommodification and transformation of the streets and other public spaces; ideas and techniques are based on the influential actions and theories of the Situationist International that was active in 1950s–1960s Europe.

Patriarchy A social system based on masculine dominance.

Public Sphere The realm within that discourses circulate and public opinion is formulated.

Subaltern Counterpublic Sphere The alternative sphere of discourse and organizing created by marginalized groups both as a means of identification and as a challenge to the hegemonic public sphere.

Transgression The breaking of social rules, norms, and expectations.

Introduction

In general, public space is property that is open to public use, including streets, sidewalks, parks, plazas, malls, cafes, interior courtyards, and so forth. Much public space research equates public space with public property, but such an equivalence is historically and practically inaccurate. A more accurate definition would be that public space is that space where 'the public' is formed and thus social and cultural rules governing public behavior predominate. In this sense public space is definable in part only through comparison with more private spaces, or the spaces over which individual or small group sovereignty reigns. The public space of a restaurant is different from the private space of its kitchen; the public spaces of the city are defined in relation to the surrounding private spaces such as homes and offices. By extension 'public space' is often used metaphorically to describe a realm of practice where some sort of public interaction is possible. Scholars often speak, in this sense, of discourse constructing a public space. By this they mean that an opportunity for debate and discussion – a public 'sphere' – has been created. Whether used metaphorically or in reference to physical spaces, public space is understood to be essential to contemporary urban life.

Geographic Research on Public Space

Given its central role in contemporary urban life, public space is seen by many as under threat, and this sense of threat is what drives much research on public space in geography. The threat comes from multiple directions.

Early geographic research on public space – taking its cue from urban sociology, the work of great mid twentieth century urbanists like Jane Jacobs and William Whyte, and phenomenology – understood public space physically as the space between buildings in a city, and socially as both a space of community formation and of 'strangers' (tourists, visitors, outcasts). Geographers sought to understand the relative vibrancy of urban public spaces, analyzing what some called the 'place ballet' that developed through the relatively unscripted, but nonetheless norm structured interactions of people going about their business, hanging out, and moving through. For researchers in this tradition, the threat to public space comes from the homogenization of everyday life through modernist urban design and planning, suburbanization (which is also a triumph of private over public interest), or just plain neglect as people forsake the life of the street for the charms of television.

In many cities, the perceived abandonment, or dereliction, of public space was thought to leave it open for 'colonization' by marginalized peoples – the homeless, 'bums', delinquent youth, etc. Geographers have sought to trace the ideologies behind this sentiment, the processes at work (especially in the wake of the massive rise of street homelessness in the United States in the early 1980s), the roles that public space plays in the lives of people who have little recourse to private space, and the ways that city elites have responded, which is often by radically transforming, and sometimes privatizing, public space. Research has therefore examined the structural transformation of Skid Row and the ways in which gentrification 'produces' homelessness, even as it is while seemingly being 'threatened by' the presence of homeless people in public space. It has also focused on graffiti and other means by which marginalized groups mark a public space as 'theirs'. Urban planners, commentators, and many city residents see homelessness, graffiti, gangs of teens on street corners, and so forth as evidence of the decline of public space.

Critical urban geographers, by contrast, often see the campaigns against these groups' use of public space as itself a threat to its public nature. In this sense public space is seen not simply as the publicly accessible spaces of a city, within which a public forms and interacts, but a key battleground defined by struggles over inclusion and exclusion, visibility and violence, order and 'anarchy'. The threat to public space perceived here in geographic research is the threat posed by dominant groups' abilities to keep out those users and functions they do not like, coupled with urban planners' and others' desires to create new public spaces that are seen as safe for middle class residents, tourists, and, especially, inward capital investment. These are spaces such as festival marketplaces, malls, or rebuilt urban parks but with control vested in private conservancies or local property owners. From this perspective, orderly public space is seen as a precondition for urban redevelopment or gentrification. Herein, however, lies a contradiction: planners and cities try to sanitize spaces so that gentrification can proceed, but in the process, often create the very disorder they wish to eliminate, or at least make invisible.

As a space of struggle, public space is unavoidably a political space. The exclusion of homeless people, racial minorities, and others from public space is thus often interpreted as a political act that has deeply political effects: it transforms who is and who can be part of the public. Writing in the wake of the uprisings in Eastern Europe, and especially the Tiananmen massacre of 1989, geographers have sought to understand how urban public space is a foundation for political life, a site where publics represent themselves to their members and to others. Public spaces – Tiananmen Square, Hyde Park, or the Plaza Mayor in Argentina – are staging grounds for protest and dissent. As public spaces are transformed – and all too often privatized – by urban redevelopment, gentrification, and the demands of global capital, many geographers argue, the space for dissent shrinks. At the same time, legal and policing practices – ranging from requiring permits for protest to creating large no protest zones around significant national or international events to simply banning individuals and groups from even being in a space – have been analyzed as a closing down of the political functions of public space, and especially as a reversal of rights to public space won over a long history of struggle from before the women's suffrage movements to after the civil rights movement. In this analysis, public space is made public through struggle (it is not *a priori* public in any meaningful sense), and it is made less public when dominant forces of order and control 'win'.

In such analyses, public space is often presented as an ideal of openness toward which social groups struggle. However, other research shows that public space is always, even in the wake of such struggle, exclusionary. It is not infrequently a violent space. This violence and exclusion is gendered. Public space has often been a male space. In part this is an ideological scripting – male is public, female is private – and one that does not necessarily match reality (women have always worked in public spaces). In part it is a function of women's fears of public space where sexual or other assault is a possibility – it is both a 'random' occurrence and a very real practice of exclusion – even as most violence against women occurs within the private space of the domestic sphere. In part it is because public spaces are made unwelcoming in other more 'subtle' ways (e.g., through sexualized imagery). And in part, despite the necessity of women's labor in public settings, it is a function of patriarchal labor markets. Much feminist research on public space, therefore, has focused on the nature of gendered exclusions. This research has been joined by research on the normatively heterosexual nature of much urban public space.

Similarly, geographers have sought to understand how women, queers, people of color, the disabled, and others who are not normatively part of the public have differentially struggled for inclusion, while at the same time struggling to change the very nature of public space – to make it into something different than it had been. These struggles can be class based. The interests of middle and upper class women in creating a space for promenading and display in nineteenth century New York, for example, might be very different from the interest of Filipino domestic workers creating a space for recreation on their days off in twenty first century Hong Kong. The interests of the working class gay men in the 1920s American city may be quite different from their middle class counterparts after the Gay Liberation Movement of the 1970s. The struggles are different for the abled and the disabled, for dominant and subordinate racial formations, and so forth. Here the threat to public space is both that under some conditions it is too open, and under others it is not open enough. It is too open in that it threatens violence and expulsion through the 'normal' workings of racism, patriarchy, and heteronormativity; it is not open enough in that it will never really be public until struggles for greater inclusion are won.

The geography of public space, therefore, can best be defined as the relationship between the physical materiality of specific kinds of (generally) publicly accessible spaces, the processes that structure inclusion and exclusion, and the struggles to change (or maintain) both the structures and form of public space. The interaction of these elements produces public space as (actually existing) public space.

Public Sphere/Public Space

If early public space research was rooted in urban sociology, phenomenology, and urban critique, the focus on structures of inclusion and exclusion has realigned it. It is now more frequently rooted in political economy, feminism, and political theory, especially as these relate to analyses of the public sphere and its role in democracy. The public sphere can be defined as the realm within which discourses circulate and public opinion is formulated. In Jürgen Habermas's influential analysis, the public sphere was a realm apart from the state in which civil society developed. The early modern bourgeois public sphere of eighteenth and nineteenth century Europe comprised coffee house debates, relatively independent newspapers, journals and pamphlets, and other places where ideas germinated and were discussed. From this, the public sphere has been conceptualized as a sphere of rational debate, and as such there were norms of behavior that structured how people interacted.

Separate from the sphere of the state, it was also different from the private sphere. The home was the preeminent marker of the private sphere – a place of privacy, retreat, and shelter from the hurly burly of public life. The private sphere was not necessarily a sphere of rationality, as Habermas defined it. For Habermas, the public sphere comprised private individuals (people with access to the private sphere) who ventured into the public – engaged in public discourse – precisely for the purpose of forming 'a public'. Habermas was most concerned with the structural conditions that allowed for the formation, and shaped, this public sphere. Consequently he sought to analyze the public sphere's transformation as structural conditions changed – as technology created new modes of communication, as the public sphere was democratized by the demands of 'the masses' seeking political representation and power, and especially as the 'private' interests of the capitalist economy came to dominate public life.

Criticisms of this model of the public sphere are many: Habermas underplayed the role of gender relations in structuring the bourgeois public sphere, both glossing over the degree to which the private sphere was scripted as the sphere of bourgeois women and that women and other subordinated people had always been in the public sphere (as workers, for example) even if not of it (as recognized members of the polity); he too readily assumed that the Enlightenment bourgeois public sphere was a sphere to which all people aspired (that it was a universal ideal rather than a very specific construction); he theorized the public sphere as unitary, rather than highly fractured; and his concerns about the public sphere's structural transformation were undemocratic, since he seemed to take the highly restricted sphere of the late eighteenth century as almost a golden age and the democratic struggles of the nineteenth century as a reason for that age's demise. Not all these criticisms were entirely valid, but they have had the effect of sparking a lively debate about what the public sphere is, who has access to it (and under what conditions), and how race, gender, nationality, and other social relations are key components of the public sphere's structure and the struggles over it.

Most prominently two revisions of Habermas's model have become influential in geographic studies of public space. First, Nancy Fraser and others have developed the concept of subaltern counterpublic spheres. The argument here is that subaltern groups, with differential access to resources and power, 'carve out' separated spheres of discourse in which they create what Fraser calls 'alternative publics'. Subaltern counterpublic spheres are, for her, "parallel discursive arenas where members of subordinated social groups invent and circulate counter discourses to formulate alternative interpretations of their identities, interests, and needs." The public sphere is not unitary; it is multiple and complex, although this revision does allow, as the term 'subaltern' indicates, for an understanding that some public spheres are hegemonic. Second, many scholars have questioned the relatively easy separation between public and private that underlies Habermas's model. There is, they argue, no bright line between public and private. Rather, the boundary between the two is always socially constructed under specific conditions, and therefore a site of significant struggle. In this regard it is useful to understand actions as relatively public or relatively private – placed, through historical contingency, social struggle, the relative success of subaltern counterpublics in making their claims, and so forth, along a continuum from most private to most public. It is hard to see the private sphere of the home as entirely private (the realm of private individuals), for example, given the way that the state, public opinion, and so forth actively structure what goes on there; and it is hard to see the public realm as fully public given the power of private actors (e.g., corporations monopolizing the media) to shape public action.

There are important implications of these two critiques for public space research. In part taking its cue from the second critique of the Habermasian model, a general assumption in geography is that public and private spaces, like public and private actions, exist along a continuum. Yet too much research on public space has quickly assumed that the continuum of space runs parallel to the continuum of actions: that public actions take place in public space, and private actions take place in private space. When this parallelism is broken, the result is a transgression. This assumption often holds even in studies founded in sophisticated arguments about the social construction of boundaries between public and private. But as Lynn Staeheli has shown, it makes much more sense to see the continua of spaces and actions as orthogonal to each other (**Figure 1**). Many actions scripted as 'public' take place in spaces understood as 'private' and vice versa – and not all of them are transgressive. Indeed, the ability to create a (relatively) private space for political organizing and alternative discourse is essential for the formation of alternative – or subaltern counter – publics. Study of the spaces of political action shows that they are not 'parallel discursive arenas' as Fraser puts it, but, precisely, 'alternative spaces'.

The upshot of this analysis has been to introduce a third critique into the discourse on the structural conditions of the public sphere, namely, to show that physical space (and its structuring) matters to the kinds of publics and public spheres that can form. Little public sphere theory has carefully considered the material spaces that do and must form the foundation of the public sphere. The public sphere is written about almost as if it is simply an amorphous 'realm' or a 'medium' within which we all live – something like water for a fish or air for a mammal. Public space research, on the other hand, has sought to show that the shape and structure of the spaces available, how they are bounded and policed, who is included and excluded from them, all make a difference to what the public sphere is.

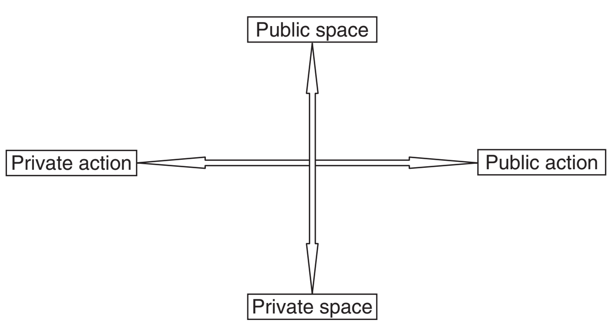

Figure 1 The continua of spaces and actions.

As a result, in geography and broader urban theory, there is a strong focus on the production of different kinds of public spaces, and on the processes and struggles through which specific spaces are made relatively more or relatively less 'public'. The sense of threat that drives much research in this vein is that despite the fact that public actions cannot be directly mapped onto public spaces, nonetheless, the kinds of public spaces socially available significantly shape possibilities for public action and for subaltern groups to be seen in and as part of the public. Public space research is thus increasingly focused on analyses of the myriad struggles over access to urban space, and, therefore, over the forms of exclusion that structure public space.

Public Space/Exclusionary Space/ Contested Space

While the ideal of public space as politically, socially, and culturally open space has been a powerful force, both in shaping struggles over urban processes and in urban theory, the fact is that all social space is at some level exclusionary. Inspired by the arguments of Mike Davis in *City of Quartz* (which examined the urban restructuring of Los Angeles and showed how this restructuring was 'protected' by closed circuit television, private security forces, the fortification of urban space, and the clearance of homeless people from downtown) and by the contributors to Michael Sorkin's *Variations on a Theme Park* (which explored the end of public space in the new American City), geographers in the 1990s debated the relationship between changes in governance in the wake of the Reagan/Thatcher revolutions, urban form, and the means by which certain groups were excluded from publicly accessible spaces.

Research on laws governing public space showed that public space in America and Britain (the focus of most of the research) was never 'open' in any full sense, but always highly regulated. But new laws in the 1980s and 1990s, focusing especially on homeless people, made the acts of exclusion both more blatant and, seemingly, more popular to the general public. Many people want and expect the homeless – and other suspect people like presumed gang members – to be excluded from public space. Under the banner of restoring 'quality of life', cities thus enacted laws that banned sitting on sidewalks, aggressive panhandling, sleeping in public areas, offering services like windshield washing, loitering in 'gang colors', and so forth. All of these were attempts to keep homeless people, young people, and some people of color out of public space. Besides laws, cities experimented with softer means of exclusion including removing public benches or changing their design to discourage sleeping;

randomly turning on sprinklers (especially at night) on lawns that homeless people sat or slept on; removing lawns altogether and replacing them with prickly plants; massively expanding closed circuit televisions systems; and requiring business owners to fence off and lock up their garbage dumpsters.

Many cities also encouraged the formation of business improvement districts (BIDs) to maintain and police public space. BIDs are typically chartered by the city to cover a certain area (in New York City, for example, there is a BID in the Times Square area, another around Grand Central Station, another in the Wall Street district, etc.). The BID in each area assesses a fee – above and beyond normal taxes – from local business or property owners and uses the revenue generated to provide services, such as street sweeping and washing, general maintenance, and, especially, security. BIDs hire private security guards (in New York and San Diego they are often recruited from the ranks of the formerly homeless) who act both as 'ambassadors' (handing out tourist literature, providing directions) and as 'enforcers' (moving the homeless along, breaking up loitering youth). The latter function is highly controversial, and there have been several instances of abuse by BID security agents, who have greater latitude for moving people along than the police precisely because they are not publicly accountable in the same way the police are. But the important point for public space is that the combination of new taxing structures, little public oversight, and privatized maintenance and policing hands control – ownership, in a sense – of public space from publicly accountable government agencies to private interests (nearby property owners and businesses). In the process the modes of exclusion are transformed, and the nature of what Henri Lefebvre called "the right to the city" – the right to be actively involved in the production of the public sphere – is reworked.

Against this sort of exclusionary practice, as well as against what much research sees as a 'corporatization' of the city (the monopolization of visual public space by corporate advertising, the appearance of the same retail chain stores in every city and neighborhood, the corporate sponsorship of public squares, parks, and even streets), significant opposition has developed and public space has become a zone of contestation. Homeless people's organizations, for example, have squatted on abandoned or derelict land and created self governing encampments. These encampments are also a reworking of public space; they are no doubt exclusionary zones, but the modes and reasons for exclusion are quite different than those practiced by the BIDs. They exclude so as to create a space where an excluded people may live. In New York, similarly, community organizations and other groups squatted abandoned land that had reverted to city ownership when owners failed to pay taxes in the wake of the fiscal crisis of the 1970s. Many abandoned lots were turned into community gardens that added green space to the city, provided fresh vegetables and flowers in poor communities, made a space for community or political organizing, and encouraged revitalization of red lined neighborhoods. Though often fenced, and jealously guarded by their developers, such gardens created a new kind of public space. In addition, they were so successful in sparking community revival that by the 1990s the city sought to reclaim many of them to sell to private developers. Gardeners fought back and eventually many of the lots were transferred from the city to land trusts, changing the nature of public ownership and the modes of exclusion that shape the space.

On another track, corporate control of public space has been challenged by neosituationist activists such as Reclaim the Streets or Space Hijkackers. Using street theater, temporary occupations, and other strategies, activists take over public space for parties and festivals, usually without permission, and often against the wishes of public authorities. Their goals are to de commodify public space and, especially, to bring a sense of fun and play back into the city: to reclaim public space as a space of and especially for the people. Finally, despite often cumbersome permit processes and restrictive 'time, place, and manner' rules, public space still hosts political demonstrations, parades, and the like. Geographic research on the contestations over and in public space shows that predictions of its 'end' are premature; rather, rules of, and restrictions on, access to public space are constantly struggled over and transformed. In the process, what public space is, is transformed.

One of the ways the nature of public space is being transformed is through changing property regimes. Geographers have focused mostly on public space as space, without considering the ways in which that space is owned. Some publicly accessible space (malls, festival marketplaces) is privately owned. Some remains publicly owned but is privately controlled, as with BIDs – such public space can be labeled pseudo private property. New York's community gardens were owned by the city, and were threatened; many are now owned by land trusts (which are private) and seem more secure as public spaces – even as their governance has now been transferred more firmly into private hands. Such changing relations of property can be crucial to the kinds of exclusions that are possible and therefore to the type of public sphere that may form. Relations of property structure – for good and for ill – the possibilities for the formation of subaltern counterpublics. Analysis of property (in rural, as well as urban settings) is becoming a key focus of geographic research on publicly accessible space.

Future Directions

Research on urban public space in Anglo American human geography has primarily focused on the English speaking world. Most case studies are from Britain, the United States, and Canada, although there is a significant minority of case studies focusing on non Anglo Europe. There is, however, comparatively little work on the non Euro American world, although public space is of obvious importance there, and in some ways is structured differ ently. Studies are now appearing of women's use of public space in Istanbul, Jakarta, Hong Kong, and a suite of Indian cities; on quality of life initiatives in Mexico City; on homelessness and public space in Germany, Japan, and Scandinavia; and the on plaza life and urban restructuring in Latin America. In addition, there is some cross pollin ation between studies of changing urban property regimes and the changing politics of property in rural areas in both the developed and the developing world. This work has the potential to de provincialize Anglo American analyses of public space in the neoliberal era. At the same time there is a great need for more historical–geographic studies of Euro American public space to examine more carefully the forms of inclusion and exclusion that served as the foundation for the contemporary public sphere, the sorts of counterpublic spheres that have historically developed (and their effects), and, especially, the ways in which urban public spaces have historically been used by urban residents as they exercised – often against high odds – their right to the city.

See also: Homelessness.

Further Reading

Bell, D. and Valentine, G. (eds.) (1995). *Mapping desire*. London: Routledge.

Calhoun, C. (ed.) (1992). *Habermas and the public sphere*. Cambridge, MA: MIT Press.

Chauncey, G. (1995). *Gay New York: Gender, culture, and the making of the gay male world 1890 1940*. New York: Basic Books.

Davis, M. (1990). *City of quartz: Excavating the future in Los Angeles*. London: Verso.

Fraser, N. (1990). Rethinking the public sphere: A contribution to the critique of actually existing democracy. *Social Text* 25/26, 56 80.

Fyfe, N. (ed.) (1998). *Images of the street: Planning, identity, and control in public space*. London: Routledge.

Habermas, J. (1989). *The structural transformation of the public sphere*. Cambridge, MA: MIT Press.

Jacobs, J. (1961). *The death and life of great American cities*. New York: Random House.

Lefebvre, H. (1991). *The production of space*, Nicholson Smith, D. (trans.). Oxford: Blackwell.

Lefebvre, H. (1996). The right to the city. In Kofman, E. & Lebas, E. (eds. and trans.) *Writing on cities*, pp 63 181. Oxford: Blackwell.

Light, A. and Smith, J. (eds.) (1998). *Philosophy and geography II: The production of public space*. Lanham, MD: Rowman and Littlefield.

Mitchell, D. (2003). *The right to the city: Social justice and the fight for public space*. New York: Guilford.

Sorkin, M. (ed.) (1992). *Variations on a theme park: The new American city and the end of public space*. New York: Hill and Wang.

Staeheli, L. (1996). Publicity, privacy, and women's political action. *Environment and Planning D: Society and Space* 14, 601 619.

Staeheli, L. and Mitchell, D. (2008). *The people's property? Power, politics, and the public*. New York: Routledge.

Whyte, W. (2001). *The social life of small urban spaces*. New York: Project for Public Space.

Zukin, S. (1995). *Cultures of cities*. Oxford: Blackwell.

Relevant Websites

http://www.grandcentralpartnership.org
Grand Central Partnership, a pioneering BID.

http://www.nlchp.org
National Law Center on Homelessness and Poverty, tracks legal changes to public space that affect homeless and other poor people.

http://www.justiceonline.org
Partnership for Civil Justice, a public interest legal organizations that tracks the policing of protest in public spaces.

http://www.pps.org
Project for Public Spaces, an advocacy group that seeks to create and maintain urban public space (often through private means).

http://rts.gn.apc.org
Reclaim the Streets London, a neo situationist organization dedicated to making spaces for the people.

http://www.spacehijackers.org
Space Hijackers, a neo situationist group of "anarchitects".

Public Spaces, Urban

D. Collins, University of Alberta, Alberta, Canada
B.-M. Shantz, University of Auckland, Auckland, New Zealand

Glossary

Revanchism Attempts to reclaim lost territory via aggressive acts. In the context of contemporary urban politics, it refers, in particular, to efforts on behalf of the middle and upper classes to recapture urban space from the poor and disenfranchised.

Introduction

Urban public spaces are at the center of a great deal of literature concerned with the public life of citizens. The material spaces most often referred to are the street (including the sidewalk or footpath), the park, and the civic square. Each of these types of space has its own historical legacy, but in broad terms they became increasingly standardized features of Western cities from the mid nineteenth century, with the rise of urban planning and increased investment by city governments in infrastructure for the general use and benefit of the public. Since this time, ownership of, and responsibility for, streets, parks, squares, and other such communal spaces has generally been invested in public authorities (i.e., branches of government), and all members of the public have been guaranteed access, typically without financial charge. Accordingly, they have been spaces in which city residents can encounter the broad spectrum of socioeconomic difference present in urban centers.

Since the early 1980s, the adoption of policies intended to exclude some members of the public from these spaces, as well as privatization of various aspects of their management and ownership, has been seen by many geographers and other commentators to compromise their public character. This has given rise to theories of 'the end of public space', which point to increasing intolerance, surveillance, regulation, and exclusion in parts of the urban environment previously envisioned as open and universally accessible. The ongoing rise of privately owned, commercial alternatives to conventional public spaces – most notably the shopping mall – is seen as a further threat to the existence of truly common spaces in the city. Accordingly, the actual 'publicness' of contemporary urban public spaces is highly differentiated.

The erosion of principles of public (governmental) ownership and control of these spaces, and of the ideal of universal access, is potentially a matter of concern for a number of reasons, not least of which is the importance of streets, squares, parks, and similar spaces for popular political action. One of the enduring uses of public spaces has been for marches, parades, and protests seeking to advance political goals, large and small. While much political organizing can occur in private and virtual spaces, the visibility associated with taking to the streets remains a powerful tool for achieving recognition, and sometimes, for effecting change. Put another way, many venues exist for communication among strangers and casual acquaintances, yet there is an enduring value associated with face to face interaction in, and occupation of, material public space. It is a way to express concerns, gain attention, and sometimes, garner respect. However, the political functions of urban public spaces are often compromised, particularly in the name of its commercial functions – attracting and encouraging consumption, appealing to the mobile upper and middle classes, and acting as a thoroughfare for individual movement.

Conceptualizing Urban Public Spaces

Urban public spaces take a multitude of forms, and serve a wide variety of purposes, and thus defy simple definition. Nevertheless, it is possible to distinguish between broad types of public space. One early distinction, offered by Michael Walzer, pointed to the difference between close minded types of space intended for only one type of use, and those which were open minded, where a variety of uses (and users) were tolerated. Many modern urban spaces – shopping centers, government complexes, highways, and dormitory suburbs – fell into the former category, and sought to preclude social mixing and political activity. By contrast, many of the more established outdoor public spaces of the traditional inner city – the street, piazza, and park – were open minded and offered opportunities for political action and civic discourse. Importantly, however, the unpredictability of open minded spaces was often perceived as threatening, and in practice, many developed at least informal exclusionary practices.

Notions of exclusion and access are also at the fore in Kurt Iveson's examination of four models of public space; models which refer both to particular organizations of material space, and to different ideals promoted by theorists of public life. First, the ceremonial model of public space centers on large scale civic sites – most notably city squares and plazas – in which people gather, particularly to mark significant events in the life of the nation or city. Historically, these spaces have been important sites for the display of state and church power, while more recently they have been subject to increasing corporate influence. By virtue of public ownership, central location, and large scale, they have also been key sites for political gathering and protest.

The second model centers not on particular publicly owned sites, but on the social mixing and sense of com munity promoted by certain traditional elements of urban design, such as high densities, mixed uses, pedestrian access, and shared facilities. This particular organization of urban space, and associated sociality, characterized many older central city areas in North America and Europe, especially those established prior to the advent of the automobile dependency. This community model of public space – where a vibrant public life traverses both publicly and privately owned urban spaces – has recently inspired New Urbanist alternatives to the privatism of conventional suburban design, as well as new forms of inner city property development. For example, since the 1980s cities have expressed renewed interest in mixed use complexes consisting of offices, shopping areas, and entertainment venues. This trend has been linked to gentrification, and the emergence of consumption driven urban lifestyles. Accordingly, urban public spaces are often (re)designed with the express intention of feeding this consumer culture and attracting members of the 'spending public'. The ethos behind the management and maintenance of urban public spaces is increasingly oriented toward the consumption patterns of a respectable middle class 'community'.

By contrast, the third model of public space holds as its ideal not a sense of community, but notions of openness and accessibility. Reflecting liberal concerns for universal rights of participation in public life and affairs, it emphasizes the need for urban public spaces to which all residents can be admitted, irrespective of social pos ition. Such thinking drove much urban planning prior to World War II, and exerted particular influence over the parks movement in the late nineteenth and early twen tieth centuries. While contributing to improvements in urban quality of life, public spaces in the liberal model seldom, if ever, fully realized inclusionary ideals: while social difference was theoretically irrelevant, in reality formal and informal expectations about acceptable behaviors and uses of space strongly reflected bourgeois norms, especially around class and gender. Thus, the preeminent park designer Frederick Law Olmsted im agined that urban parks would play a moralizing role: exerting a tempering, soothing effect on the industrial working class, and encouraging the virtues of countryside living in inner city environments. Women, for their part, had a highly tenuous and contested place in emergent liberal public spaces. As city life was popularly associated with freedom from patriarchal authority, and sexual adventure and danger, women's presence in urban space was often perceived as disorderly. Broad sets of concerns for social order (spanning prostitution, violence, crowds, and the imperative of reproducing the private sphere) underpinned efforts to exclude women from the public life of the city.

The fourth model of public space identified does not seek to privilege ceremony or community, nor to ignore difference, but to enable interaction between multiple publics. In this way, it is suggested, the true diversity of the public can be experienced: members of different groups can mingle and overlap in shared spaces, but need not become homogeneous. In a practical sense, embracing diversity in this way requires relaxing zoning and the segregation of different land uses, as well as regulatory controls over public space, so as to allow a broad spectrum of social expression. However, such imperatives appear to run contrary to many con temporary trends in the development and control of public spaces, as is noted below.

Political and Economic Significance

Urban public spaces are often envisioned as sites of political engagement. They are bastions for conveying political messages to large, diverse groups of people, which help to maintain democratic forms of citizenship. For example, parades, protests, and demonstrations are uses of public space that demand the witness of the surrounding public. Demonstrators from across the pol itical spectrum routinely organize and join parades to express their views in the public arena. In doing so, they tend to draw upon a common set of symbols (e.g., ban ners, effigies, and flags) and actions (e.g., chanting and marching). Nowhere is this more visible than when parades are followed (or met) by counter parades, which use the same styles but promote opposing agendas.

Critically, the political significance of urban public spaces is appreciated not only by protesters and paraders, but also by corporate and government elites. This was evident in Baron Haussmann's reconstruction of Paris in the mid nineteenth century, which included an extensive set of public works, and the replacement of narrow streets with broad avenues and boulevards. These were intended to promote public order by inhibiting the ability of protest movements to establish blockades, and improving the ability of troops to move through the city.

Today, managers of urban space typically link the social and esthetic qualities of the public environment to city image and prosperity. This has important ramifi cations, as can often been seen in cities hosting hallmark events, such as economic summits and major sporting contests. Preparation for these events, and the publicity they bring, often entails a series of 'improvements': streets are repaved, historic buildings restored, and security greatly enhanced – sometimes to the point of outright militarization during the event itself. The homeless are frequently removed (e.g., to alternative accommodation provided in more peripheral locations), and street trading and informal economic activity prohibited. The emphasis is on the creation of a city that is safe, clean, and apparently devoid of economic and social problems. The perspectives of those displaced by such efforts are seldom sought out; instead, official and media discourses portray the poor and disenfranchised as an illegitimate presence in public space. Importantly, these same trends are evi dent in more enduring efforts to create tourist districts and historical precincts cleansed of physical and social discomforts, and presented as wholly nonconfrontational and harmonious.

The public spaces of the inner city are also routinely connected to economic prosperity. Following large scale deindustrialization in Western cities, and the increasing mobility of capital, many downtown areas experienced significant economic decline, and an associated deterior ation of the built environment. One response has been large scale redevelopment projects oriented around ser vice sector employment and consumption activities. Thus, derelict industrial sites have been transformed into landscapes of consumption: festival waterfronts, enter tainment complexes, landscaped corporate plazas, shop ping centers, large scale office high rises, residential developments, and recreation areas. Distinctive, and typically sanitized, landscapes of consumption are deemed essential for success both in attracting the local spending public (from the suburbs), and in interurban competition for investors and tourists.

The forms of exclusion that often accompany contemporary consumption oriented reformulations of urban public space run contrary to notions of urban public spaces as sites of freedom, anonymity, and excitement. This said, exclusion from public space is not without precedent. Concerns for moral order, and for the temptations and dangers associated with public spaces in the city, have long motivated formal and informal exclusionary practices. Imposing social and legal order on the city has often meant excluding women and children (at least as individuals, as opposed to as members of family groups) along with other potentially disruptive elements – the working class, the poor, and some visible minorities. This exclusion has been justified in terms of both protecting those targeted from insults and dangers, and protecting the broader society from the disorder associated with their presence.

These twin imperatives can be found, for example, in efforts to constrain the occupation of public spaces by migrant female domestic workers in cities such as Singapore and Hong Kong. Specifically, it has been claimed that these women should be sheltered from the moral and physical dangers that lurk outside the private home/workplace, and that the visual and social order of public space needs to be protected from mass occupation by disruptive 'others' marked out by gender, class pos ition, language, and ethnicity. Similar justifications often underpin the exclusion of unaccompanied children, and older youths, from urban centers: it is claimed that young people are both vulnerable in public space (e.g., to vio lence and vehicular injury), and hostile to its values.

Ownership, Order, and Surveillance

Contemporary reorganizations of urban public spaces commonly involve significant roles for private (especially corporate) actors in aspects of design, surveillance, regulation, and maintenance. This trend toward privat ization of public space is particularly marked in the United States, where private business improvement dis tricts (BIDs) have taken responsibility for the mainten ance of many public spaces. They replace traditional (and at least theoretically democratic) forms of control with privatized sets of rules and exclusionary policies. In broad terms, BIDs are voluntary associations of local businesses that tax themselves and pool funds to 'improve' urban public spaces through the provision of street lighting, closed circuit television (CCTV), security staff, cleaning, and street furniture. They seek to deliver cleaner, safer public spaces amenable to consumption oriented activ ities and the redevelopment of adjacent private property. BIDs are driven by profit seeking objectives, and are not accountable to local electors, but seek to forge urban public spaces that compete with other cities for tourists and investment, and with suburban malls for preferred (i.e., consuming) members of the public. Interestingly, the mall (together with other commercial landscapes) also contributes to privatization, by inducing the relocation of much public life onto private property.

An influential US literature in the 1990s characterized these trends as 'the end of public space', as public access and control of communal spaces had been thoroughly debased by development processes emphasizing private profit and the creation of bourgeois landscapes of con sumption. These trends were linked to issues of gentri fication, the neoliberal turn in urban policy, and the emergence of tourism and consumption led development as economic strategies. Downtowns now feature tourist 'bubbles' intended to shelter visitors from certain realities of city life, and fantasy landscapes oriented toward sanitized leisure and entertainment. These spaces in clude, for example, festival marketplaces, and new pedestrian systems (skyways, tunnels, etc.) that promise safer ways to travel through urban space, in part, through physical separation from the traditional public domain. In this environment, public spaces have not necessarily been 'lost'; instead there is greater regulatory control over access to, and behavior in, public space. Often, access is contingent upon appearing to be a member of the consuming public.

The new found emphasis on consumption is not as straightforward as it may appear. The focus is on promoting formal types of consumption consistent with images of prosperity: thus, the consumption of alcohol in expensive bars and restaurants may be considered legitimate, and encouraged, in a way that the consumption of alcohol in down market taverns, or in a nearby park, is not. Some types of informal trading are similarly problematic: Van Deusen's recent account of the redevelopment of a civic square noted that hot dog vendors were deemed an illegitimate presence, as they did not sell the 'right' kind of product, and did not attract the 'right' kind of public to the space. This rhetoric led to the vendors being legislated out of the space. By contrast, up market cafes are often sought after tenants for such redevelopments, in part, because they offer a more affluent, refined form of consumption, which attracts more desirable (and more prosperous) members of the public. The messages sent by the presence of particular types of services about preferred uses, and users, of public space may be reinforced by more explicit messages in the form of signs, warnings, and quasi legal imperatives (e.g., do not litter, do not drink, do not sit, and do not skateboard).

Concerns around the increased exclusion from, and privatization of, public space are often more complicated than they initially appear. One issue is the language of privatization itself: public (state) agencies often lead the way in efforts to curtail public access, and while this is frequently done in collaboration with private actors, it complicates any simple picture of public versus private interests. Public policymakers have, for example, enacted laws to make life increasingly difficult for the urban homeless, by rendering their behaviors in public space illegal. In so doing, they have transformed the meaning of the term 'public' so as to exclude the homeless from membership: they have been redefined as symbols of all that is wrong with public space, and on this basis, their exclusion from it is deemed both appropriate and necessary.

These policies may be understood as part of a broader program of revanchism – that is, an aggressive attempt to recapture urban space from the poor on the part of the upper and middle classes. Revanchism was particularly evident in New York City under former Mayor Rudolph Giuliani, whose zero tolerance policies targeted low level public order offences and incivilities (e.g., pan handling, window washing, and graffiti) in the name of 'getting tough on crime' and 'cleaning up' the city. Revanchist policies seek to expunge unfavorable and unsightly activities/people from city streets and parks in order to increase the image and marketability of the city, and to establish new social norms around legitimate membership of the public. In pursuing this program, public and private actors have diminished (or ignored) the political significance of public space in order to promote its role in serving commerce. In particular, they act in the belief that the relatively prosperous do not wish to confront social difference, or disorderly people and behaviors, while participating in retailing, leisure, and recreation. In this atmosphere, consumption is the highest ideal, and nonconsumption is seen as deviant.

One of the most popular methods for 'securing' public space is CCTV. It attempts to increase the visibility of crime, disorderliness, and antisocial behaviors, partly so as to improve police responses, but also to encourage greater self regulation. A routine part of urban regeneration projects and efforts to promote the moral order of public spaces, CCTV is justified primarily on the grounds that it deters crime and increases safety. As with revanchist policies, local businesses often push for the installation of CCTV in public spaces, contending that it encourages a greater flow of potential consumers. Moreover, the technology is used as a drawcard for enticing investment in the city center. There is considerable debate over the effectiveness of CCTV, and one outcome may be the displacement of targeted individuals and behaviors to peripheral locations.

While such regulatory and technological changes have received much attention, there is a longer established, if prosaic, threat to the vitality and openness of urban public spaces: the automobile. The rise of the car degraded urban public spaces through direct demolition to facilitate the construction of highways and related infrastructure, but also through congestion, air and noise pollution, and creating physical dangers and inconveniences. Streets with multiple purposes became roads designed entirely for cars; open spaces used for social gatherings or children's play became parking lots. As older forms of public mobility (buses, trains, walking, and cycling) have been replaced by privatized transport, many public spaces have been reduced to antisocial thoroughfares for private commuters. In this respect, the car undermines the public life of the city: as journeys between private spaces have become automobilized, they have also been privatized, and many traditional forms of social interaction and communication with strangers rendered unnecessary.

Contemporary trends in the control, management, ownership, surveillance, and (re)development of public space often appear to be overwhelmingly hostile to notions of inclusiveness, and to realizing the political potential of public space, especially for individuals and groups who are socially or economically marginal. Yet the picture is more complicated than this. First, exclusionary practices are not necessarily as rigid and steadfast as they seem; forms of resistance and transgression remain possible. At a basic level, a significant gap often remains between regulatory expectations for public space, and actual practices: for example, bylaws against begging, sleeping, or skateboarding in public seldom result in the outright disappearance of those behaviors. Protests and parades continue, with or without official sanction, and continue to seek the power to define public memory, and promote particular ethical and political agendas. The intended order of public space is routinely subverted, both intentionally and unintentionally.

Second, exclusion is not the sole intent behind the trends. Rather, efforts to 'improve' or 'redevelop' public space typically seek to encourage greater access and use by selected elements of the public, albeit at the expense of others. In this respect, it is useful to understand 'the public' as fractured rather than unitary. Thus, for example, increased privatization, surveillance, and policing in an urban park may make the space more appealing and accessible for members of the public who might previously have avoided it (e.g., white collar workers), while simultaneously discouraging or actively inhibiting use by others considered less desirable. Moreover, many new developments incorporate and integrate public spaces, in self conscious attempts to foster appropriate forms of social activity, often with reference to 'successful' antecedents such as the piazza. The common goal is to attract preferred members of the public, and to encourage them to linger, in part, by excluding those others who pose real or imagined threats to their comfort and safety.

Efforts to transform the socioeconomic character of public space in this way raise important questions about the rights of urban residents to use, and shape, their environment. They run contrary to Henri Lefebvre's notion of the 'the right to the city' – an idea which suggests, at minimum, that, residents have the right to co create the city by virtue of living it. This co creation encompasses the right to participate in the production of urban spaces, and the right to appropriate them. Such notions fly in the face of much contemporary regulation, which is organized by, and for, social and economic elites, in a manner that reinforces the power of capital, especially in the form of private property rights. Indeed, some contemporary efforts to 'improve' urban public spaces are intended primarily to promote or jumpstart the development of adjacent private property.

Summary

Public spaces are a defining feature of Western urbanism, and have acquired considerable social, political, and economic importance, especially since the mid nineteenth century. While often thought of as universally accessible, in practice, use and access are often extensively regulated, in part because 'open minded' public spaces are unpredictable, and may threaten elite interests (e.g., in maintaining order). Moreover, questions about how public spaces are best organized extend beyond questions of access to encompass broader issues relating to design, sociality, and understanding of the term 'public'.

While public spaces may be conceptualized in many ways, the broad trend toward increasing regulation and surveillance, much of it undertaken by, or on behalf of, private commercial interests, is thought to be compromising its public character. The political value of public spaces as sites of protest and demonstration is seen by many to be under significant threat, as is the social value attached to encountering difference. What is being prioritized, in many instances, is the economic value of public spaces: their potential to facilitate consumer activity, attract tourists and investors, and encourage private investment.

For some commentators, these trends signal 'the end of public space'. The material environments in question are no longer provided for the benefit of the public, but are ordered around the imperatives of commerce (i.e., legitimate, orderly production and consumption). To achieve this, local institutions, both public and private, seek to impose a politics of vigilance that minimizes the risk of disorder. Others, however, suggest that urban public spaces retain some potential for political action and participatory democracy. From this perspective, city streets, parks, and squares continue to provide valuable opportunities for individuals and publics to represent their needs, hopes, and demands, and to (re)assert their 'right to the city'. Nevertheless, the exercising of rights in public space is often disorderly, and frequently projects images that are not conducive to consumption and investment. The challenge, perhaps, is to ensure a balance between two public goods: safety and social justice. Public spaces must not be allowed to descend into continuous violence or uproar (which will ensure they are inaccessible to many), but nor must they be reduced to mere tools for urban elites to advance their aims, often at the cost of further marginalizing those who are already vulnerable.

See also: Central Business District; Citizenship; City Marketing; Gentrification; Homelessness; Malls/Retail Parks; Neoliberalism, Urban; New Urbanism; Planning, Urban; Policing; Private/Public Divide; Public Space; Surveillance; Urban Design; Utopian Cities; Waterfront Development.

Further Reading

Davis, M. (1990). *City of Quartz: Excavating the Future in Los Angeles*. London: Verso.

Iveson, K. (1998). Putting the public back into public space. *Urban Policy and Research* 16, 21–33.
Jacobs, J. (1961). *The Death and Life of Great American Cities*. New York: Vintage Books.
Lees, L. (ed.) (2004). *The Emancipatory City? Paradoxes and Possibilities*. London: Sage.
Lefebvre, H. (1991). *The Production of Space,* Nicholson Smith, D. (trans.). Maldesn: Blackwell.
Lofland, L. (1998). *The Public Realm: Exploring the City's Quintessential Social Territory*. Hawthorne, NY: Aldine de Gruyter.
Mitchell, D. (1997). The annihilation of space by law: The roots and implications of anti-homeless laws in the United States. *Antipode* 29, 303–335.
Mitchell, D. (2005). The SUV model of citizenship: Floating bubbles, buffer zones, and the rise of the 'purely atomic' individual. *Political Geography* 24, 77–100.
Sennett, R. (1976). *The Fall of Public Man*. New York: Knopf.
Smith, N. (1996). *The New Urban Frontier: Gentrification and the Revanchist City*. London: Routledge.
Sorkin, M. (ed.) (1992). *Variations on a Theme Park: The New American City and the End of Public Space*. New York: Hill and Wang.
Van Deusen, R. (2003). Public space design as class warfare: Urban design, the 'right to the city' and the production of Clinton Square, Syracuse, NY. *GeoJournal* 58, 149–158.
Walzer, M. (1986). Pleasures and costs of urbanity. *Dissent* (Fall), 470–475.
Young, I. (1990). *Justice and the Politics of Difference*. Princeton, NJ: Princeton University Press.
Zukin, S. (1995). *The Culture of Cities*. Cambridge: Blackwell.

Q

Q Method/Analysis

P. Robbins, University of Arizona, Tucson, AZ, USA

Glossary

Concourse A range of opinions, claims, representations, or subject positions surrounding a specific topic.

Condition of Instruction The statement, which respondents in a Q study use to organize their rank-ordering of items; for example, 'most agree/most disagree' or 'most like myself/least like myself'.

Factor Analysis A statistical procedure to simplify complex data sets and uncover relationships among many variables.

Operant Subjectivity A model of subjectivity that assumes individual viewpoints are self-referent, contextual, and performed/behaved.

Q Set A subsample of the concourse, made up of individual items that are rank-ordered by the study respondent.

Introduction

Q method (hereafter: Q) is an empirical technique for eliciting subjective beliefs, opinions, and discourses, in a way that allows controlled comparison between individuals and groups, but that also enables an exploration of the relationships between ideas, claims, and concepts within people's subjective points of view. At its essence, Q is a procedure in which respondents rank order their agreement/disagreement, like/dislike, or affinity/repulsion to a carefully selected set of independent claims or items. From a collection of these responses, the investigator derives abstracted typical arrangements of the items, or subject positions in common, against which individual responses can be evaluated.

The technique has not traditionally been used extensively in human geography, in part because unnecessarily obscure jargon is sometimes used to describe and explain it, but is increasing in application as interest grows in discourse and subjectivity and new methods (beyond questionnaires) have been sought to better elicit and evaluate them. The procedure in Q requires both interpretive and statistical techniques, making Q difficult to define as either 'quantitative' or 'qualitative'. Similarly, Q is both 'objective', in the sense that specific observed responses are recorded, and subjective, in that the selection of items, the interpretation of results, and the responses themselves are derived from personal and situated knowledges. Despite some important limitations, Q is an excellent tool for use in projects seeking to understand shared and divergent points of view, a cornerstone for critical human geography, humanist geography, environmental perception, feminist geography, political ecology, and a host of other subdisciplinary approaches.

Traditional factor analysis, a technique to simplify complex sets of data correlations, typically seeks to take a body of scores or measurements from a population of people (e.g., test scores in math, English, history, Latin, geography, biology, music) and create abstract grouped factors that summarize the sets of scores that tend to correlate with one another. Q analysis was originally introduced in the 1930s, when its progenitor, psychologist, and physicist, William Stephenson, described a modification of this traditional approach to factor analysis in a letter to the editor of *Nature*. Here, he asserted that by inverting the usual analysis, it would be possible to derive factors that described groups of individuals or test subjects, rather than groups of variables used to characterize them. Stephenson went further, however, proposing a 'Q method', where the tests of individuals specifically took the form of ranking matters of opinion, belief, or subjective preference. The ranking of these items together, for any person, was understood to represent their point of view or feeling on a subject, and the factors derived would describe common, abstracted arrangements, with each additional factor being significantly different from the last, and so forming a set of ideal types, or diverse common structures.

For social science the method has been used extensively in the 70 years since, particularly in the wake of one of its senior proponents Stephen Brown and his book *Poltical Subjectivity*, and especially in political science, psychology, and the health sciences. Usually, the method is used to characterize differing perspectives, con structions, opinions, discourses, and beliefs. These toge ther, in the language of Q methodology, all represent subjectivity. This view, common to most Q practitioners, holds that subjectivity is 'operant', meaning that it is not essential to the subject but instead contextual and be haved. For research in fields like human geography, this view allows research into the socially contingent, cultur ally coherent points of view held by differing individuals in differing economic, political, and social contexts.

The results of a Q method analysis typically take the form of a set of factors, which are composed of arrange ments of a set of stimuli, typically taking the form of questions, quotes, or photographs. These are understood to reflect real and existing types of points of view. Along with these, the analysis provides a set of scores that shows the loading (or degree of affiliation) of any individual respondent to each of these factors. This research product is considerably different from that obtained through more traditional surveys and provides evidence of something rather different than such surveys usually do, though the two techniques are by no means incomparable.

Q Method versus Traditional Survey Techniques

A traditional survey is usually designed to determine the proportion of a population that does, says, or thinks *X* (where *X* might be the use of a detergent, a view of a policy, or the level of agreement with a controversial claim). These results are typically used to examine cor relations between responses and sociodemographic characteristics. Usually, such a survey requires a random sample of the population in question, large enough to allow inference that the proportions of respondents likely represents that of the larger population and the socio demographic groups of interest. By way of example, a survey instrument might query the degree to which people think lawn chemicals are dangerous (e.g., 'very dangerous', 'sort of dangerous', 'not dangerous'). The results can be descriptive (e.g., 45% of those surveyed claim that lawn chemicals are 'very dangerous') or used analytically (e.g., women are disproportionately likely to claim that lawn chemicals are dangerous), showing how a population thinks or behaves on average and across previously defined groups.

Like surveys, Q requires individuals to respond to an instrument and report their opinions. Very much unlike traditional survey techniques (hereafter: R techniques), however, Q does not generate an assessment of view points across a representative population, but instead provides an empirically derived set of commonly asso ciated responses to a set of claims, ideas, or stimuli – showing how they relate to one another within view points of respondents. By way of parallel example, a Q method approach to the same question might assemble a large number of claims about lawns, lawn chemicals, safety, and risk, and have individual respondents sort them in order of agreement. The results will provide factors or types of complex assemblages (e.g., those that internally associate lawn chemical use with good stew ardship, importance of traditional marriage, and neigh borliness vs. those that associate it with property values, economic opportunity, and smart urban growth policy). Each respondent from the survey, moreover, is statistic ally more or less associated, positively and negatively, with all of the factors.

Clearly, each of these results provides evidence of very different things. While surveys are excellent for evalu ating the relationship of opinions to known groups and types of people, with the added possibility of correlating certain claims, ideas, or points of view, Q is specifically designed to query the relationship between certain kinds of claims, deriving groups of people in the process.

The results of each type of study can be used itera tively of course. Q might be used in an exploratory way to derive patterns of opinion that might be tested on a larger population using a survey. Conversely, a survey may suggest certain patterns of opinion in a population, which might be followed more intensively with a Q method analysis of the relationship of these opinions to other ideational, discursive, or symbolic associations. In either case, these methods do not replace one another, but in stead stake out differing terrains for social explanation.

Doing Q Research

Q research is normally conducted in a series of standard steps, including:

1. obtaining and recording a concourse and a Q set.
2. Q sorting (ranking) by respondents, including dis cussion and follow up.
3. performing correlation and factor analysis (including possible rotation).
4. interpreting and iterating with respondents.

Obtaining a Q Set

The initial step begins by determining the area of re search interest and the kinds of issues, topics, and prob lems of interest in the study. Topically, these topics are the ones selected because they tend to produce complex and divergent responses from different people, but this is not necessary. Topics where the method has been applied in the field of geography have included environmental priorities in Indian comanaged forests, views of science in US nonpoint source water policy, and perceptions of wilderness and property in Yellowstone, among others.

Once the area of concern has been determined, a concourse can be established. A concourse is a set of quotes, pictures, objects, or other stimuli that reflect a range of variation surrounding the topic. Typically, this takes the form of a series of statements or claims. The source of these claims and their selection is a matter of central importance to the execution of the study, since these claims, and their relative ranking positions, provide the raw input for the later analysis. Like survey questions, the phrasing, length, and connotations of the elements of the concourse will be important as respondents implement the Q sort survey, and so the selection of the concourse should be practically as well as theoretically informed, and should depend directly on the research questions involved.

In many cases, the concourse is formed by purposively sampling direct quotes from previous interviews in the area and communities under study. The advantage of such a 'naturalistic' sample is that it is context sensitive, phrased in an understandable fashion, and usually contains a range of divergent or contrasting materials from which to build meaningful factors.

It is also possible, however, to draw the raw claims of the concourse from secondary sources, whether those are contextual (e.g., local newspaper stories) or not (e.g., statements by Gandhi and Stalin). So too, it is possible to mix primary and secondary materials to span a range of concepts and views. There are some 'standardized' Q sets, including the so called 'California Q set', which was designed to identify personality types for use in certain schools of psychology. For the most part, however, the concourse is derived and designed to fit the specific nature of the study and research question. As noted previously, it is also possible to use nontextual materials, most typically including photographs, though even colors and odors have been used in Q method, again directed to specific questions about perception.

The final Q set derived from the concourse is typically a subset of the much larger array of potential sortable materials, but it varies in its size. Concerns in creating the Q set include length of time required for respondents, level of redundancy between claims, and range/variety of types of statements based on theoretical issues. Some studies have been performed with as few as ten items and some with an excess of 100, though neither is typical, with most Q sets having between 30 and 60 items.

Q Sorting by Respondents

The Q set is then placed in front of respondents for sorting. To facilitate this, the items are usually transferred to a physical mechanism, like a set of cards, each containing one quote, and each individually numbered. Other techniques for arranging the Q sort exist, including web based interfaces that allow respondents to sort at the computer.

In any case, the procedure remains the same; respondents are given a condition of instruction and told to rank order the Q set based on that instruction. This condition may vary from study to study. While it is most typical to ask respondents to 'rank order the items from those you most strongly agree with to those you most strongly disagree with', other options exist. It is possible, for example, to have respondents rank order based on the way they believe another person (e.g., their life partner) might, or to order them from 'most useful to least useful' or 'most like to most dislike'. The approach stresses relative ranking of items, so while it is possible to have respondents simply assign numerical values to all items, it is generally preferred that the respondent explicitly rank the relative position of each item. This is important to the goal in Q, which is to establish the relationship 'between' the items, 'within' any given person.

The items are typically arranged along a normal distribution (**Figure 1**), with a majority of items being more indifferent to respondents, and several key items defining the most important far ends of the curve. This is sometimes facilitated by providing a grid into which respondents may place items/cards. The normal distribution is not essential to the success of a Q study, however, and there is some controversy in the literature on its desirability (since it is 'forced' rather than 'free'), though most Q methodologists have concluded that as long as variability is pronounced, the normality of the distribution affects the outcome little. Similarly, the number of categories across which the items are sorted (e.g., $+3$ to -3 or $+6$ to -6) has also been demonstrated to matter little to the results.

Typically, each respondent provides a single unique sort, though on occasion the same respondent might sort

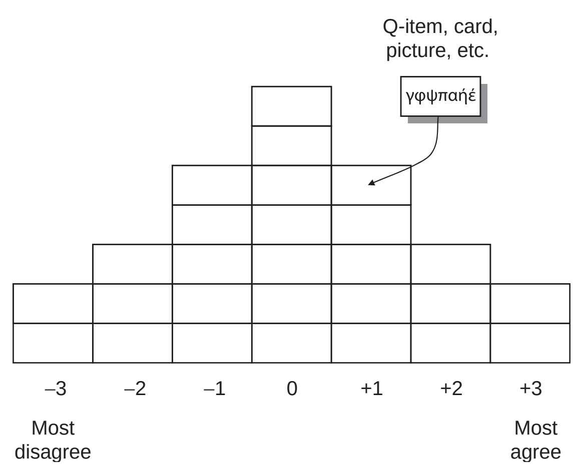

Figure 1 Normalized Q sort structure for 27 items.

multiple or varying sets of items. The investigator then records the relative position associated with each numbered item, to be used in later analysis. Ideally, during and after the sorting, the investigator has an opportunity to query the respondent on his/her logic, rationale, and considerations in the relative positioning of items. This nonquantitative reflection is essential for later interpretation of results.

Correlation and Factor Analysis

Following the sorting of the items, the investigator enters the results into statistical software (several packages exist for managing Q method analysis). This simply requires that for each sort, the unique Q set item is assigned a value (e.g., +2, −3, 0, etc.) along the ranking. Once the data are entered and each item associated with a statistical ranking, analysis proceeds in two phases, including the creation of factors and their interpretation.

Creating factors from sorts, once they are coded into software, involves several staged procedures, though the main goal is to produce a few coherent factors that represent different kinds of sorts. The main statistical procedures are (1) producing a correlation matrix, (2) deriving and selecting factors, and (3) rotating the factors.

The statistical procedure begins by computing correlation coefficients for every pair of Q sorts in the study, resulting in a matrix of each of these intercorrelations. The goal of statistical factor analysis of the matrix, as McKeown and Thomas (1988: 50) simply put it, is to: "lend statistical clarity to the behavioral order implicit in the matrix by virtue of similarly (or dissimilarly) performed sorts." To accomplish this, the correlation matrix is subjected to further analysis to produce factors. These factors are typically derived using centroid analysis (CA) or principal components analysis (PCA). The result is a set of factors composed of the *z* scores of all of the items sorted, with each factor representing a significantly different and common way of arranging the items. The number of factors that result can be selected by the analyst, though the typical procedure is to produce many (e.g., as many as seven or eight) and eliminate less meaningful ones later. The final number of factors from each study will vary greatly and depend on the variability of responses by respondents and upon the tolerance and theoretical questions of the investigator.

As an aid in determining which factors to retain and how many factors are significant, each factor has its own eigenvalue and percentage of explained variance, and each Q set item has a normalized *z* score for each factor. What this means is that each factor has its own explanatory power and represents some body of individual respondent sorts, and that each is composed of certain specific items whose rank position is important for composing that factor. A crude measure is to exclude all factors with eigenvalues less than 1 but theoretical demands may explore a great many more than this or include factors that explain less variance but that represent an important response type.

Once the factors to be retained are determined, each factor is aligned with the original individual Q sorts to determine the relationship of individual respondent sorts to the 'types' reflected in the factors. To align each factor with the Q sorts, the factors are rotated in multidimensional space. This process can be automated, using statistical algorithms such as varimax rotation, with the goal being to capture the greatest explanation of the largest number of variates using the fewest factors. On the other hand, it can be conducted 'manually' by the investigator, to test hypotheses and explore relationships. For example, if one respondent is deemed as an 'ideal' or 'typical' response, because it is important to the study in one way or another, a factor axis can be aligned toward its sort, allowing the investigator to see which other sorts are similarly aligned as a result. After rotation, whether the process is automated or done by hand, each individual respondent Q sort will therefore also have a statistic associating it to a greater or lesser degree with each factor.

Given the availability of readymade software for this procedure, much of this part of analysis can be 'black boxed' for those less familiar with statistical methods. For interpretation, however, it is usually desirable to have a familiarity with the procedures of factor analysis and the resulting statistical measures, including especially the eigenvalues of factors, the normalized *z* scores for items within each factor, and the loadings of individuals onto factors.

Interpretation and Iteration

Once the statistical analysis is performed, the investigator will have one or more factors composed of, and typified by, certain key items ranked positively and negatively. These must be interpreted to understand and project their meaning. Where statements have been used as the concourse and Q set, it is usually a matter of reading the statements together, as a sort of narrative metastatement. Where photos and other materials have been used, interpretation is considerably more difficult, and will likely rely heavily on the knowledge the investigator brings to the relationships between images and concepts, provided by previous experience and by the in depth questions asked during the Q sort (as noted above).

So too, each factor – now a common story, narrative, or concept set – has a group of individual respondents who are most strongly associated with it. At this phase it is possible to examine and analyze what makes these individuals most closely associated with these views, and to consider other contextual data about the respondents (e.g., affiliation, gender, age, employment, social power) that might help explain this commonality. It is also quite possible at this stage that individual respondents may not sit easily on a single factor, but may straddle two or more, their specific response diverging from the ideal factor type and blurring the differences between abstract sub jective positions.

This moment in analysis provides an opportunity to return to the respondents themselves, describe the results and the specific associations of their sorts relative to the abstract or common factors. Do they recognize the results as being their own views? Did the investigator get it 'right'? Why does their sort diverge from the abstract type? What are the implications for their point of view relative to that of other people? These further questions and analyses open the door onto fundamental geographic questions con cerning perception, values, and the politics of knowledge.

Example: Q Study in Yellowstone

Consider the problem of competing points of view about wildlife management around the Greater Yellowstone Ecosystem in Montana, USA. Here, differences of opin ion over wildlife management are rife between state agents, local producers, hunters, outfitters, and environ mentalists. But what is the structure of opinion in the minds of those debating the point?

A Q approach began with a large number of in depth interviews with key advocacy players from a range of interest groups. This produced a number of quotes pro pounding different ideas regarding management and suggested certain recurrent themes. These quotes were abstracted into a concourse and reduced to a Q set of 20 statements, printed on laminated cards. These state ments included claims like:

> Around Yellowstone, cattle are playing the ecological role that bison used to; intense cattle grazing, for short periods, improves wildlife habitat by removing decadent growth and encouraging important forage.
> Yellowstone is no more a natural ecosystem than the public and private lands outside the park.
> Access to wildlife is a collective good that cannot and should not be privatized.

These were given to 30 respondents, including some individuals previously interviewed along with others new to the study, responding to the condition of instruction, 'sort the provided quotes along a range from those with which you most agree to those with which you most disagree'.

The resulting factor analysis of the data produced two key factors with high eigenvalues. The first factor was typified by agreement with claims which stressed that access to nature is important and that the environment should be a free, collective good and disagreement with claims that nature can be privatized or enclosed. The second factor was typified by agreement with claims that stressed that the environment is amenable to and some times requires human intervention and maintenance and disagreement that nature is self regulating. This result is itself a useful finding, since it demonstrates the terms with which debate in the region occurs, which in some ways contradicts simplistic notions about some observers being more 'green' or environmentally oriented than others, stressing instead the question of the role of humans, rights, and property in environmental management.

Along each factor, moreover, differing individuals loaded differently, from strong association with that fac tor, indifference, or negative loading, representing an tipathy to this position (**Figure 2**). This revealed that while some commonly identified interest groups clus tered together in perception and opinion, many indi viduals did not fit into a group norm (especially state managers), and some groups (like environmentalists and hunting advocates) clustered surprisingly closely. Later iterative responses concerning these results from re spondents resulted in some surprise, especially in the many elements of the factors over which there was little or no disagreement, especially between historically opposed political communities. These results open the door for further interpretation, follow up study, and re consideration of political coalitions in the region.

Applications, Limitations, and Cautions

The potential applications for research using Q in human geography are practically limitless, and its employment allows certain ways of thinking about individuals as subjects to be operationalized more easily. In critical geographic scholarship over the last several years, and owing to theoretical and empirical work on socialized human experience, it has become more difficult to im agine or think about an 'essential' subject that is either universal or inherently tied to its class, race, or gender identity. At the same time, it is increasingly clear that social context in part produces subjective experience as individuals perform and reproduce their own identity.

Typical survey research design makes it harder to access this experiential reality or express it, since, as explained previously, survey responses can only be as sessed and statistically described in terms relative to pre given social categories or measures (e.g., income, age, gender). Q method, conversely, seeks out the way indi viduals construct complex claims into coherent positions and assesses groups in terms of the convergences of these operant subjectivities. Such an approach is well suited to a range of research questions in environmental per ception, feminist geography, political ecology, and pol itical geography, among other subfields. How are discourses assembled and mobilized in the constitution of

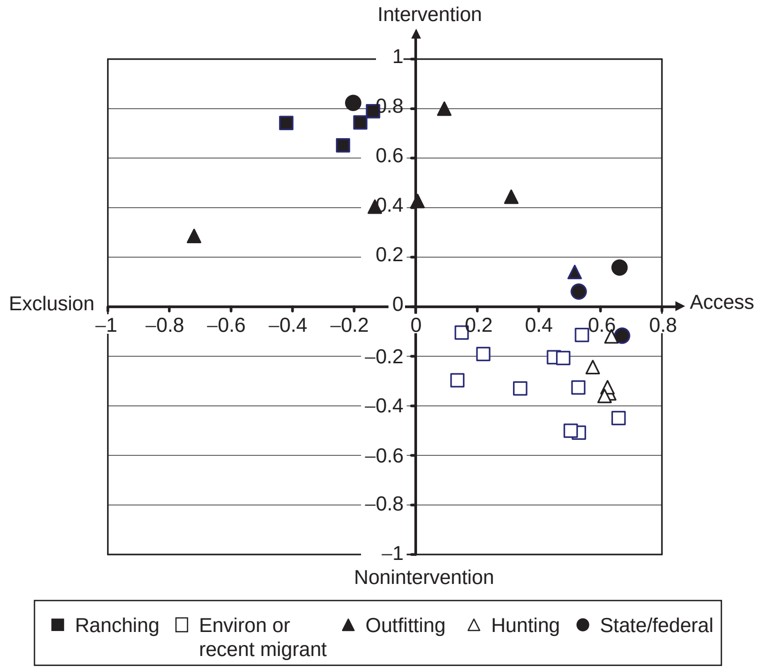

Figure 2 Example of loadings of individuals on Q factors. Reprinted with permission from Robbins, P. (2006) The politics of barstool biology: Environmental knowledge and power in greater Northern Yellowstone. *Geoforum* 37(2), 185 199.

the subject? Where do concept worlds converge and under what conditions are they hegemonic? Along what material and identity conditions do divergent perceptions and experiences occur? Q method renders these theore tical questions partly empirical.

Having said this, there is nothing about the technique, which amounts to little more than an interview, survey, and a statistical manipulation, which can allow un mediated access to the experience of other people. All of the cautions and caveats that apply to other forms of research apply doubly to Q method; therefore, the stat istical products of the technique can be taken to imply concrete, objective truth about subjectivity. At every stage of research, however, the investigator's conceptions intrude upon and frame the responses of the interviewee: in topic selection, concourse construction, Q set re duction, and factor interpretation. But at every moment, Q encourages and enforces a relationship, whereas the investigator cannot do that framing by himself; instead, he constantly iterates with the respondent, whose feed back comes in providing concourse materials, sorting the Q set, and iteratively interpreting the results.

The elegance of Q, therefore, is not that it 'gets around' the mediation of representation inherent to re search, since the investigator is by definition one that represents the respondent, with all of the implied power relations such a relationship entails. Instead what Q method can do is enable an iterative and reflexive process between the researcher and the respondent in which each provides the other with a series of explicit tracings of what they hear one another say, rigorously comparing and ab stracting differing claims. As such, it offers a unique tool for approaching crucial questions in human geography.

See also: Factor Analysis and Principal-Components Analysis.

Further Reading

Addams, H. and Proops, J. (eds.) (2001). *Social Discourse and Environmental Policy: An Application of Q Methodology*. Cheltenham, UK: Edward Elgar.

Brown, S. (1980). *Political Subjectivity: Applications of Q Methodology in Political Science*. London: Yale University Press.

Eden, S., Donaldson, A. and Walker, G. P. (2005). Structuring subjectivities? Using Q methodology in human geography. *Area* 37(4), 413 422.

Kline, P. (1994). *An Easy Guide to Factor Analysis*. New York: Routledge.

McKeown, B. and Thomas, D. (1988). *Q Methodology*. Newbury Park, CA: Sage Publications.

Robbins, P. (2006). The politics of barstool biology: Environmental knowledge and power in greater Northern Yellowstone. *Geoforum* 37(2), 185 199.

Robbins, P. and Kreuger, R. (2000). Beyond bias: The promise and limits of Q methodology in human geography. *Professional Geographer* 52(4), 636 649.

Relevant Websites

http://www.pcqsoft.com/
Michael Stricklin's PCQ program Web site.

http://www.qmethod.org/
Q Methodology: A Method for modern research.

Qualitative Geographic Information Systems

M. Cope, University of Vermont, Burlington, VT, USA
J.-K. Jung, University of North Dakota, Grand Forks, ND

Glossary

Counter-Mapping Using cartography to critique existing, mainstream assumptions about the world and how it functions, often employed in participatory research with goals of empowerment of oppressed social groups.

Epistemology A way of knowing, a set of ideas about what constitutes 'truth'.

Mash-Up Primarily web-based efforts by the general public to combine spatially referenced data with information of interest.

Qualitative Data Generally non-numerical data, usually the result of qualitative research.

Qualitative GIS Diverse forms of integrating qualitative and quantitative data through mixed-methods approaches using GIS.

Qualitative Research Scholarly inquiry that aims to understand processes, context, and relations; usually performed through use of a wide range of methods involving intensive exploration such as interviews, life histories, focus groups, journaling, participatory research, archival analysis, participant observation, art, and photography.

Visualization Working with computer-generated maps and other views of geographic information, including charts and tables, 3D scenes, time-based representations, interactive maps, and schematic views of relationships.

Introduction

Qualitative geographic information systems (GIS) is a relatively new term describing a set of developments, practices, and ideas that have emerged from multiple directions. Qualitative GIS is centered on some form of integration of qualitative research and/or qualitative data with geographic information systems (GIS, traditionally – but erroneously – assumed to be limited to 'quantitative' data processing). This integration can be accomplished in many ways, from many different approaches, but the three most common avenues identifiable in the present literature are: (1) qualitative research projects that benefit from the data analysis and geographical representation strengths of GIS; (2) quantitative projects using GIS that benefit from the contextualizing and explanatory strengths of incorporating some qualitative research and data; and (3) research projects that were planned from the start using mixed methods, integrating the strengths of various methods, and allowing them to complement, contradict, or confirm each other. Rapid technology advances are fostering ever quicker combinations of spatially referenced data with other forms and sources of information and different software systems are increasingly becoming compatible, all of which generate new possibilities and practices for integrating qualitative data and GIS.

On the practical side, qualitative GIS has emerged from diverse practitioners and researchers, including geographers, human ecologists, planners, community development organizations, and many others trying to blend qualitative data and research with GIS. Some of the key developments have come from those who want to incorporate qualitative data into a GIS for their representation purposes, such as adding photos, audio or video clips, or other non numerical data to computer generated maps. Another common approach to mixing these techniques employs a recursive strategy, such as using GIS to analyze and display quantitative data (e.g., census patterns), which then raises new questions to be pursued through qualitative methods of research (e.g., interviews or participant observation), which in turn may require a return to GIS for new layers of mapping and spatial analysis, integrating insights from both qualitative and quantitative data. Further, although they may not have been explicitly aimed at combining qualitative work with GIS, many public participation GIS (PPGIS) projects have fostered innovative techniques for incorporating community input with planning processes, to the extent of training local residents and organizations to use GIS in their development efforts. Through PPGIS projects and scholar practitioner initiatives, the fresh perspectives and the multiple needs of local groups have often led to new and different types of geographic representations that integrate diverse sources and types of data and research.

Finally, as GPS units and mapping capabilities have become more and more accessible through new technologies and faster, more complete internet connectivity, the lay public has increasingly played a role in pushing the boundaries of combining computer cartography with qualitative information. Some examples include the geotagged photographs of Flickr, Wikimapia, OpenStreetMap, and the use of Google's 'my maps' function that allows users with little to no cartographic background to add their own notes, photos, links, arrows, routes, and points of interest to base maps. A different kind of example is the rapid increase in 'mash ups' available on the web that incorporate various forms of data into digital maps, including unique representations such as the various smells encountered by passengers in New York City subway stations. Despite their sometimes bizarre content, we take these vernacular forays into ethnography and cartography as a serious indication of the saturation of spatially referenced, popularly informed data displays; they suggest we are only at the beginning of a new era of qualitative GIS.

On the more abstract and theoretical side, qualitative GIS has been informed by several streams of debate, scholarship, and praxis, such as PPGIS, critical GIS, and feminist uses and critiques of GIS. GIS has been critiqued for being implicated with the agendas of the elite, capitalism, militarism, and patriarchy, in part because its roots lie in government and marketing applications and its software has been prohibitively expensive and complicated. However, defenders of GIS point out that as a set of tools and techniques, GIS can be used to further many agendas and can align with many differing epistemologies. Toward this end, there has been an impressive amount of work done in the area of social justice research that employs GIS techniques with its advantages of persuasive displays and clear pattern identification. An example of this is the Syracuse Hunger Project in which hunger is mapped against food pantries and soup kitchens to document pockets of severe disadvantage in the city of Syracuse, New York. Further examples include the many community organizations who have been engaged with learning and using GIS in their efforts to improve the lives of local residents.

Notwithstanding progressive social accomplishments using GIS, the existing software tends to privilege a particular meaning and representation of space that is inherently characteristic of the 'West' or the Global North, such as defining 'places' through points, nodes, or polygons with little opportunity to include vernacular 'sense of place' data. Further, due to its very architecture, current GIS technologies also tend to rigidify and overdetermine the boundaries or edges that in reality may be subtle, nuanced, gradual, and pervious. In response, there have been recent attempts by indigenous people, such as some groups of Native Americans, to explore the possibilities of reconciling the Cartesian system of spatial representation (x–y coordinates) with their own perspectives of space. In a similar vein, some feminists have argued that GIS is intrinsically male centered in both its construction and practice, due in part to its roots and in part to its architecture and privileging of some views over others. However, other feminists and critical researchers contend that GIS is a useful technique for identifying and analyzing patriarchy and other forms of oppression, along the lines of socially progressive uses of GIS mentioned above, and can be usefully employed in conveying a sense of 'situated knowledge', multiple stories or truths, and relational spatial processes.

Indeed, feminists have been particularly active in raising several important issues in the area of qualitative GIS. Although GIS does not have a particular epistemology ingrained within it *per se*, it has, in the past, been accused of serving as a tool of positivism and empiricism. Feminists are critical of positivism, seeing it as a masculinist project that frames the goal of 'science' as uncovering the Grand Truth through neutral, value free, empirical observation and law building. Rather, feminists argue that there are multiple 'truths' that are equally valid depending on one's positionality, and that empirical experiences of the world are always informed by multiple subjectivities; thus, it is impossible to be a 'neutral' observer. However, feminist scholars have pointed out that, just like statistical analysis, GIS as a tool for quantitative research can be used for socially relevant purposes, including documenting and questioning the oppression of marginalized social groups (a central goal of feminism). That is, GIS can be practiced in ways that are epistemologically in line with feminism. In fact, GIS's capacity for analyzing and displaying multiple versions of the world, the inherently reflexive processes of exploratory data analysis (EDA) that it supports, and its strength in incorporating mixed methods research make GIS an important tool in the feminist's toolbox.

Overall, there are many forms and meanings of qualitative GIS and together they hold a great deal of potential for both mainstream approaches and critical social applications. Although the forms of qualitative GIS may be as numerous and diverse as its practitioners, there are several trends that can be identified and categorized. Four types of practical strategies for integrating qualitative research and data with GIS are reviewed here: inserting qualitative data into GIS, visualization, mixed methods research with GIS, and modifying GIS.

Inserting Qualitative Data into GIS

First, at the most basic level, qualitative data can be directly input to GIS. Using various display techniques (such as mouse over functions), photographs, audio and video clips, music, noises recorded in a particular location, fieldnotes, public commentaries, drawings, and other forms of non numeric data can be easily added to map layers. These richly textured displays can be used to represent multiple social meanings, different perspectives of various social groups, and can incorporate change over time. Such strategies have also been used in interactive forums, such as on the Internet or in public kiosks, to vastly expand both the sources of input and the extent of outreach of the maps. For example, people can submit photos, commentaries, or other digital media to be added to locational data.

Second, at a somewhat more complicated level, the 'hyperlinks' function in GIS can be used to link users directly to various materials, particularly multimedia, which may be stored either locally or on the Internet. Incorporating hyperlinks into a GIS has benefits in allowing a wide variety of nonspatial data to be con nected to spatial reference points (or lines, polygons, or features). There are also benefits to both the designers and the users because large digital files (e.g., video) do not necessarily have to be saved within the GIS database, but rather are called up from their storage location when the user clicks on the link. This also means that as ma terials are updated, the link automatically connects to updated material. However, the downside to having data stored elsewhere is that if material is removed or other wise discontinued from the local or internet host, the hot links in the GIS will no longer work, and, further, that it is more difficult to include hyperlinked data in explora tory analysis within the GIS database if the data are not fully integrated.

Visualization

The power of 'visualization' has been well theorized by critical geographers concerned with the social impli cations of GIS, as well as critical cartographers who are interested in the concept of the socially constructed map and making socially produced maps. More broadly, critical visual methodologies include the understanding that images are made and used in all sorts of ways by different people for different reasons.

First, at a fairly basic level, qualitative data can be visualized in GIS by quantifying the data. For example, the way households engage in particular economic sur vival strategies in different areas of a city can be cat egorized and counted and then matched with other relevant spatial data in a cartographic display. Then these visualized qualitative data can be overlaid with other types of demographic data. This, in turn, helps the re searcher visualize other patterns and disjunctures in the data, fostering new insights into underlying processes. In this way, the visualization power of GIS allows multiple forms and sources of data to be explored analytically.

Second, visualization can be helpful in the iterative processes of mixed methods research, for example, GIS's capacity for exploratory analysis helps researchers who are trying to identify spatial characteristics in both qualitative and quantitative data. Being able to visualize the phenomena being studied, in combination with other contextual spatially referenced data, can create new understanding (or at least new questions to be explored).

Further, GIS generated maps can be used in the re search process itself, such as to stimulate focus group discussions or solicit community input for future (re) development. Researchers in PPGIS and participatory GIS (PGIS) have shown how the visualization impacts of GIS are influential in producing local knowledge and can be included as an important facet of participatory re search. Not only can GIS stimulate initial discussions, it can also be used to recursively visualize spatial narratives and represent local knowledge, adding sequential layers of richness to participatory research.

Finally, the concept of 'counter mapping' expands the visualization power in GIS further. Counter mapping generally refers to the efforts of local residents and op pressed social groups to contest or undermine power relations and asymmetries in relation to maps or map making processes. GIS can be used to produce powerful counter mapping strategies, which can dramatically in crease the power of people to control the representations of themselves and their environments. Counter mapping strategies have frequently been used in conservation mapping, where they originated, but have also generated new possibilities for expressing the critical consciousness of indigenous peoples, low income urban communities, and socially marginalized residents by enabling them to present their local knowledges, which have traditionally been omitted or ignored by more powerful actors and institutions.

Mixed-Methods Research

Hyperlinks and visualization techniques can be useful as parts of mixed methods research, both at the level of integrating different forms of data and different forms of research practice. The term 'mixed methods' refers not merely to research projects employing 'multiple' meth ods, data sources, or techniques, but rather connotes a further level of blending, reflection, and critical inte gration. In addition, mixed methods research does not merely suggest conducting parallel qualitative and quantitative inquiries on the same topic, but means that the results from different data collection and analytical tasks are taken into consideration with and in contrast to each other. In some cases, results from different techni ques may verify each other, in others they provide con text or explanation for each other, and in yet others they provide differing or even incompatible results, which can itself lead to interesting follow up questions. Qualitative GIS is in inherently 'mixed methods' because integrating qualitative and quantitative research through the medium of GIS requires some level of reflexive consideration of both the methods and the outcomes.

Practitioners and scholars doing PPGIS and PGIS have often been at the forefront of using mixed methods through GIS, as well as challenging the epistemological assumptions of positivism by embracing the idea of multiple perspectives and subjectivities. The interventions made in PPGIS and PGIS have generally been to foster broader public participation in the use of GIS (in project design, data collection, spatial analysis, and decision making), to the extent of training community members in using GIS as an empowering medium for counter mapping.

A Hypothetical Qualitative GIS Research Project

What does a mixed methods research project using qualitative research and GIS look like? While there is no definitive formula, a common pattern is for an initial research question to be raised by a scholar or a local community – for instance, unusual rates of a particular phenomenon such as cancer, hate crimes, housing discrimination, or deforestation – and then a set of data collection and visualization methods is set in motion. On the qualitative side, there may be focus groups, interviews, or other intensive techniques aimed at understanding the occurrence and impacts of the phenomenon. On the quantitative side, there may be statistical analysis to see whether certain variables stand out or if particular correlations can be identified. GIS (itself not specifically quantitative or qualitative) is often used to display relevant spatially referenced data, such as the patterns created by the incidence of the phenomenon. All of these initial forays into explanation are considered with regard to each other. For example, members of a focus group identify a few racist landlords based on their own house hunting experiences, the GIS map shows that indeed there are few racial minorities in those census tracts, and statistics can rule out other factors such as class. After looking at the maps, the community group identifies other areas in the neighborhood that are of concern, and a new round of data collection – both qualitative and quantitative – is initiated. These routes of data collection and analysis are not just simultaneous but integrated, reflexive, and mutually informing. Increasingly, with the rise of participatory research strategies, these practices are not only engaging the public as participants, but are in fact responding to subjects' own interests in order to value the knowledge production of the relevant social actors themselves, and in so doing, shifting scholars out of the role of 'expert' toward that of 'facilitator'.

Modifying GIS

To work around inherent limitations of GIS packages, some researchers have modified the programs for specific research purposes as a way to incorporate more analytical power for qualitative data. There have been some fruitful approaches for qualitative GIS in attempts to alter GIS itself, such as to enable incorporation of diverse representations of spatial knowledge. General purpose, 'off the shelf' GIS products have many features that are useful in one realm (e.g., marketing) but less so in another (e.g., service delivery), giving users models or functions that are not ideally designed to meet the particular needs of a given research project. While this dilemma is not unique to participatory work, this is an issue commonly faced by these researchers.

Some examples of recent innovations include modifying the GIS database for the incorporation of local knowledge or in order to help analyze data. Another commonly used way to modify GIS is to develop new extensions (such as ArcGIS and ArcView extensions). In one project the researchers then packaged the extension with the main program for a community organization to continue its work with new custom functions such as cluster analysis, an incidence rate calculator, and an areal interpolator. In another instance, a custom extension helped to integrate multiple (previously incompatible) sources of data for visualization purposes at the neighborhood level. These approaches to creating new extensions in GIS are significant because their efforts are deeply rooted in the specific goals and the nature of the research project.

Beyond modifying the database and creating extensions (i.e., working 'within the system'), there are also moves toward completely rebuilding the foundations of GIS to create new forms of geographic information systems. Some examples include GIS/2, which refers to a new medium to convey local people's ideas. In this, linking multimedia files to point data on the maps serves as a way to express people's ideas in GIS, but instead of incorporating embedded tools in current GIS such as a 'hyperlinks' tool, there is a process of de constituting and rebuilding GIS procedures by using extensible markup language (XML) to link local knowledge and spatial data without using GIS programs. Other approaches to this kind of deconstructing and reordering include public participatory visualization (PPVis) and 'alternative GIS'. Other new directions include developing one interface that encompasses web based systems, data archives, parallel and counter texts in diverse media, standard maps and datasets, and sketch maps and field notes. These examples have made GIS more than just a visualization tool and each of them offers new opportunities for integrating qualitative, quantitative, and spatially referenced data through qualitative GIS.

There are several ironies here, though. First, the goal of much PPGIS has been to be more inclusive of those whose perspectives have not informed the original architecture of commercial GIS, yet in order to customize and change the system to incorporate these perspectives, the technology must be learned by those traditionally left out of the process. While some attempts to train community members have been successful, the situation remains ironic – the oppressed must once again use the tools of the powerful to gain access to respect, knowledge, and resources. The second irony is that as soon as the community members become savvy in the workings of GIS, they are folded into the very 'tech nocracy' that has been critiqued, their input becomes 'information' for more public use, and their innovative extensions become part of the marketable framework for future versions of software.

Future Directions

The concept of integrating qualitative research with GIS has fairly recently developed from practitioners seeking better ways of understanding the world, and as such is a newly emerging topic with a somewhat scattered litera ture and few critical debates to bind it together. However, the simultaneous growth of scholar practitioner inno vations in qualitative GIS and the rapid expansion of spatially referenced data along with technological methods for integrating multiple sources of data through computer cartography suggest that we can expect to see a significant rise in these methods in the near future. Several new publications are aiding in the endeavor to critically examine the theory and practice of qualitative GIS, including special issues of journals and a forth coming book. These reflect rising interest within the discipline of geography in mixed methods research and in critical GIS perspectives. However, they also demon strate a solid commitment to digging beneath patterns in the search for processes in order to answer questions of not only where something happens, but 'why'.

See also: Critical GIS; Feminism/Feminist Geography; GIS and Society.

Further Reading

Cieri, M. (2003). Between being and looking: Queer tourism promotion and lesbian social space in Greater Philadelphia. *ACME: An International E Journal for Critical Geographies* 2(2), 147 166.
Cope, M. and Elwood, S. (2009). Qualitative GIS. London: Sage Publications.
Crampton, J. W. (2001). Maps as social constructions: Power, communication and visualization. *ACME: An International E Journal for Critical Geographies* 25(2), 235 252.
Dennis, S. F. (2006). Prospects for qualitative GIS at the intersection of youth development and participatory urban planning. *Environment and Planning A* 38(11), 2039 2054.
Elwood, S. (2006). Negotiating knowledge production: The everyday inclusions, exclusions, and contradictions of participatory GIS research. *Professional Geographer* 58(2), 197 208.
Elwood, S. and Leitner, H. (1998). GIS and community based planning: Exploring the diversity of neighborhood perspectives and needs. *Cartography and Geographic Information Systems* 25(2), 77 88.
Gilbert, M. R. and Masucci, M. (2006). The implications of including women's daily lives in a feminist GIScience. *Transactions in GIS* 10(5), 751 762.
Harris, T. and Weiner, D. (1998). Empowerment, marginalization and community integrated GIS. *Cartography and Geographic Information Systems* 25(2), 67 76.
Jiang, H. (2003). Stories remote sensing can tell: Integrating remote sensing analysis with ethnographic research in the study of cultural landscapes. *Human Ecology* 31(2), 215 232.
Knigge, L. and Cope, M. (2006). Grounded visualization: Integrating the analysis of qualitative and quantitative data through grounded theory and visualization. *Environment and Planning A* 38(11), 2021 2037.
Kwan, M. P. (2002). Feminist visualization: Re envisioning GIS as a method in feminist geographic research. *Annals of the Association of American Geographers* 92(4), 645 661.
Kwan, M. P. and Knigge, L. (2006). Doing qualitative research using GIS: An oxymoronic endeavor? *Environment and Planning A* 38(11), 1999 2002.
Matthews, S. A., Detwiler, J. E. and Burton, L. M. (2005). Geo ethnography: Coupling geographic information analysis techniques with ethnographic methods in urban research. *Cartographica* 40(4), 75 90.
McLafferty, S. L. (2005). Geographic information and women's empowerment: A breast cancer example. In Nelson, L. & Seager, J. (eds.) *Companion to Feminist Geography*, pp 486 495. Malden, MA: Blackwell Publishers.
Nightingale, A. (2003). A feminist in the forest: Situated knowledges and mixing methods in natural resource management. *ACME: An International E Journal for Critical Geographies* 2(1), 77 90.
Obermeyer, N. J. (1995). The hidden GIS technocracy. *Cartography and Geographic Information Systems* 22(1), 78 83.
Pain, R., MacFarlane, R., Turner, K. and Gill, S. (2006). 'When, where, if, and but': Qualifying GIS and the effect of streetlighting on crime and fear. *Environment and Planning A* 38(11), 2055 2074.
Pavlovskaya, M. (2006). Theorizing with GIS: A tool for critical geographies? *Environment and Planning A* 38(11), 2003 2020.
Pavlovskaya, M. E. (2002). Mapping urban change and changing GIS: Other views of economic restructuring. *Gender, Place and Culture* 9(3), 281 289.
Schuurman, N. (2000). Trouble in the heartland: GIS and its critics in the 1990s. *Progress in Human Geography* 24(4), 569 590.
Sieber, R. (2004). Rewiring for a GIS/2. *Cartographica* 39(1), 25 39.

Relevent Websites

http://www.maxwell.syr.edu
Maxwell School of Syracuse University.
http://gawker.com
New York City Subway Smell Map.

Qualitative Spatial Reasoning

R. Billen, University of Liege, Liege, Belgium
N. Van de Weghe, Ghent University, Gent, Belgium

Glossary

Allen's Interval Logic It is a calculus for temporal reasoning that defines possible relations between time intervals (e.g., '*X* starts with *Y*' or '*X* takes place before *Y*') and provides a composition table that can be used as a basis for reasoning about temporal descriptions of events.

Artificial Intelligence It is the study and design of intelligent agents where an intelligent agent is a system that perceives its environment and takes actions which maximize its chances of success.

Convex Hull In mathematics, the convex hull or convex envelope for a set of points *X* in a real vector space *V* is the minimal convex set containing *X*.

Homeomorphism In the mathematical field of topology, a homeomorphism is a special isomorphism between topological spaces which respects topological properties. Roughly speaking, a homeomorphism is a continuous stretching and bending of an object into a new shape which respects topological properties (e.g., a cube may be stretched into a ball, but cannot be stretched into a donut).

Map Algebra It is a set-based algebra (primitive operators) to manipulate geographic data. Depending on the spatial neighborhood, operators are categorized into four groups: local, focal, zonal, and incremental. The input and output for each operator being map, the operators can be combined into a procedure to perform complex tasks.

Mereotopology It combines mereology and topology. It is a first-order theory of the topological relationships among wholes, parts, parts of parts, and the boundaries between parts.

Introduction

> Saying that Alaska is 1 518 800 km^2 is sufficiently exact quantitative information about size and distances in Alaska but very likely it is not meaningful in relation to the spatial knowledge of the average listener. On the other hand saying that Alaska alone is bigger than all the states of the East coast from Maine to Florida is cognitively more immediate. (Clementini *et al.*, 1997: 318)

Spatial reasoning can be seen as any deduction of knowledge from a situation having spatial properties. It goes from simple calculation of spatial attributes (e.g., the area of a region) and the description of spatial relations (e.g., two polygons are disjoint) to all operations of map algebra. More complex tasks, such as planning of routes, answering to shortest path problems, and allocation of resources, can be categorized in the broad area of spatial reasoning as well. Reasoning can be performed on quantitative, as well as on qualitative information. Typically when working with quantitative information, a predefined unit of a quantity is used. For example, one could say that the distance between Ghent and Liège is 150 km. In the qualitative approach, continuous information is being quantized or qualitatively discretized by landmarks, depending on their relevance to the current research context. In this respect and in Belgian context, Liege is far from Gent (as opposed to 'close to'), Brussels is closer from Ghent than Liege, Brussels is between Gent and Liege, Liege is east of Gent, etc.

Spatial reasoning in our everyday interaction with the physical world is mostly driven by qualitative abstractions of the quantitative space. Of course, numerical representations may be well suited, in particular, where precise spatial information of a definite situation is available, and if the output required from the system is itself primarily numerical. However, such quantitative information is often too precise for the given spatial context. For example, if we show a person the way to get to the post office, we do not need to be more precise than the streets he has to follow. Inferring new knowledge from this qualitative spatial information is called qualitative spatial reasoning (QSR).

QSR is a branch of artificial intelligence (AI), linked to spatial cognition, highly related to geographical information science (GIScience) and by extension to most aspects of geography. It is a particular (through its spatial component) subfield of qualitative reasoning (QR) and is often associated with spatiotemporal (qualitative) reasoning when time is considered. In short, QR is an approach for dealing with commonsense knowledge of the physical world without numerical computation. Its main goal is to make this knowledge explicit, so that given appropriate reasoning techniques, a computer could make predictions, diagnose, and explain the behavior of physical systems in a qualitative manner. QSR is the part of QR dealing with spatial concepts. Due to space complexity, most work in QSR has focused on single aspects of space. The most important aspects of space are topology, orientation, and distance. As shown in psy chological studies, this is also the order in which children acquire spatial notions. Other aspects of space include size, shape, morphology, and spatial change (motion). Beside geography, other domains interested in qualitative spatial reasoning are, for example, robotics, linguistics, experimental psychology, and qualitative physics.

In this article, we present fundamentals of QSR, and some of the major current approaches and models in the domain. Afterward, we point out some of the most promising research directions.

Fundamentals

First, it is necessary to choose the basic entities on which qualitative representation and reasoning will be based. This ontological choice in the QSR research community is often reduced to two possibilities: points or regions (being extended spatial entities). While it is easier to deal with points rather than with regions in a computational framework, taking regions as the basic entities is certainly more adequate for commonsense reasoning. Further more, if points are required, they can be generated from regions. Another ontological distinction is the nature of the embedding space. The most common notion of space is n dimensional continuous space (R^n). But, there are also approaches which consider discrete or finite space.

Much of the work on QSR has concentrated on rep resentational aspects, for example, how to handle or characterize region's topological properties and relation ships, how to characterize region's shape, etc. Talking about relationships, they could refer to two regions (binary relationships), three regions or two regions together with an external frame of reference (ternary relationships), etc. Examples of qualitative spatial representations are pre sented in the next section.

However, the biggest challenge is to provide calculi, allowing reasoning about objects in multidimensional space without the use of traditional quantitative know ledge and techniques. The most popular reasoning methods used in qualitative spatial reasoning are con straint based methods. Shortly, they are based on con straint satisfaction which is the process of finding a solution to a set of constraints, known as the constraint satisfaction problem (CSP). To be applied to spatial qualitative relationships based on particular aspects of space such as topology, direction, and distance such re lationships must have the characteristic to be jointly exhaustive and pairwise disjoint (JEPD). Qualitative binary relations are jointly exhaustive and pairwise dis joint if between any two spatial entities exactly one of the basic relations holds. The same goes for n ary relations, but between n spatial entities. Being JEPD, the set of all possible relations is then the set of all possible unions of the basic relations.

The most prevalent form of qualitative reasoning is based on the composition table (CT). The underlying idea of a CT (also known as inference table, formerly known as transitivity table) is to compose a finite set of new facts and rules from existing ones, that is, if two existing relations ($R_1(a, b)$ and $R_2(b, c)$) share a common object (b) they can be composed into a new relation set $R_3(a, c)$. Such a composition is also called an inference (here symbolized by $\otimes$). Thus,

$$R_1(a,b) \otimes R_2(b,c) = R_3(a,c)$$

Compositions of relations are usually precomputed and stored in the CT. CTs originate from the domain of temporal reasoning. An example of a temporal com position is

$$\begin{aligned}(\text{interval}_1 \ \text{before} \ \text{interval}_2) \otimes (\text{interval}_2 \ \text{meets} \ \text{interval}_3) \\ = (\text{interval}_1 \ \text{before} \ \text{interval}_3)\end{aligned}$$

Going beyond the domain of purely static spatial calculi, it is possible to reason about spatial changes. Spatial changes may occur on a specific entity (change of shape, position, etc.) and may have some impacts on its rela tionships with the environment. Assuming that changes are continuous, another major qualitative reasoning technique is the conceptual neighborhood diagram (CND). Continuous means that things change smoothly. A simple consequence of continuity, respected by all systems of qualitative physics, is that, in changing, a quantity must pass through all intermediate values. That is, if $A < B$ at time t_1, then it cannot be the case that at some later time t_2 $A > B$ holds, unless there was some time t_3 between t_1 and t_2 such that $A = B$. Following this continuity assumption, we could say that two relations between pairs of events are conceptual neighbors, if they can be directly transformed into one another by con tinuously deforming (i.e., shortening, lengthening, mov ing) the events (in a topological sense) and that a set of relations between pairs of events forms a conceptual neighborhood if its elements are path connected through conceptual neighbor relations. CNDs describe all the physical possible transitions between relations for a given model. Such diagrams allow inferring information; for example, if relation R_2 is a conceptual neighbor of re lation R_1 and R_3 and if R_1 and R_3 are not conceptual neighbors, then evolving from relation R_1 between en tities A and B to relation R_3 is only possible through an intermediate stage R_2 between entities A and B.

Keeping this framework, from representation to rea soning methods, in mind, we may move on to major approaches and models in QSR.

Major Approaches and Models

Topology

The properties of geometric shapes and geometric relations between spatial objects that are invariant under homeomorphisms such as translation, rotation, and scaling are respectively topological properties and topological relations. Thus, a first group of stretching, bending, twisting, or compressing a figure are transformations which do not affect topological properties and relations. However, a second group of tearing, puncturing, or inducing self intersection do affect topological properties and relations. In other words, topological properties of a drawing on a rubber sheet are preserved when the rubber sheet is deformed according to the above mentioned first group of transformations. For this reason, topology has also been called the geometry of the rubber sheet.

Generally, topology can be seen as the most developed area of qualitative spatial representation and reasoning, resulting in a variety of topological approaches that can be found in literature. We describe two standard works: the region connection calculus (RCC) and the 9 intersection model. Both theories were developed independently from each other at the beginning of the 1990s. Although the 9 intersection model originates from the domain of database theory and RCC is from the field of qualitative reasoning related to AI, both have as a major conclusion that there are eight topological relations between two regions in 2 D space.

The region connection calculus (RCC)

At the end of the 1980s and the beginning of the 1990s, Randell, Cui, and Cohn transformed Clarke's interval logic for reasoning about space and space–time, which is the spatial counterpart of the interval calculus, into RCC.

The basic idea is a primitive binary relation $C(x, y)$, read as x connects with y, which is defined on non null regions, being the space that physical entities occupy at any given time. Using the relation $C(x, y)$, a set of eight JEPD relations has been worked out, resulting in RCC 8. These relations describe differing degrees of connection between regions: is disconnected from (DC), is externally connected to (EC), partially overlaps (PO), is identical with (EQ), is a tangential proper part of (TPP), and is a nontangential proper part of (NTPP). All except the last two are self inverse. Taking the inverse of the last two results in the inverse of a tangential proper part of (TPPI) and the inverse of a nontangential proper part of (NTPPI, see **Figures 1** and **2** and **Table 1**).

The set of eight RCC 8 relations can be reduced to five by taking the disjunction of DC and EC (resulting in DR, meaning x is discrete from y), the disjunction of TPP and NTPP (resulting in PP), and the disjunction of TPPI

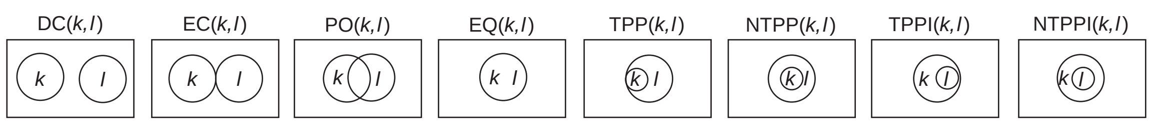

Figure 1 Basic relations of RCC-8.

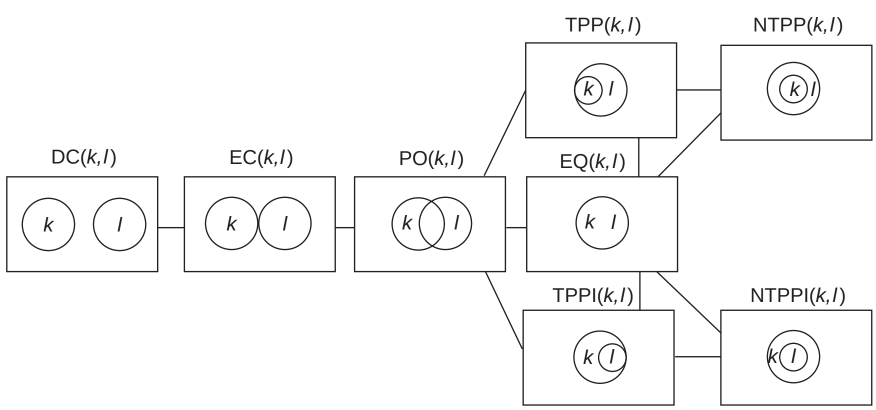

Figure 2 Conceptual neighborhood diagram of RCC-8.

Table 1 Composition table of RCC-8

$R_1 \otimes R_2$	*DC*	*EC*	*PQ*	*TPP*	*NTPP*	*TPPI*	*NTPPI*	*EQ*
DC	No info	DR, PO, PP	DR, PO, PP	DR, PO, PP	DR, PO, PP	DC	DC	DC
EC	DR, PO, PPI	DR, PO, TPP, TPI	DR, PO, PP	EC, PO, PP	PO, PP	DR	DC	EC
PO	DR, PO, PPI	DR, PO, PPI	No info	PO, PP	PO, PP	DR, PO, PPI	DR, PO, PPI	PO
TPP	DC	DR	DR, PO, PP	PP	NTPP	DR, PO, TPP, TPI	DR, PO, PPI	TPP
NTPP	DC	DC	DR, PO, PP	NTPP	NTPP	DR, PO, PP	No info	NTPP
TPPI	DR, PO, PPI	EC, PO, PPI	PO, PPI	PO, TPP, TPI	PO, PP	PPI	NTPPI	TPPI
NTPPI	DR, PO, PPI	PO, PPI	PO, PPI	PO, PPI	PO	NTPPI	NTPPI	NTPPI
EQ	DC	EC	EC	TPP	NTPP	TPPI	NTPPI	EQ

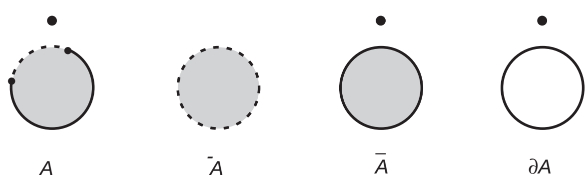

Figure 3 Object *A* and its interior, closure, and boundary.

$$\begin{bmatrix} \partial A \cap \partial B & \partial A \cap \mathring{B} & \partial A \cap \overline{B} \\ \mathring{A} \cap \partial B & \mathring{A} \cap \mathring{B} & \mathring{A} \cap \overline{B} \\ \overline{A} \cap \partial B & \overline{A} \cap \mathring{B} & \overline{A} \cap \overline{B} \end{bmatrix}$$

Figure 4 The 9-intersection matrix.

and NTPPI (resulting in PPI). This set of five relations is known as RCC 5. RCC 23 is a result of the introduction of an additional primitive function *conv*(r_1): the convex hull of r_1. By determining whether a region r_1 is inside (I), partially overlaps with (P), or is outside (O) the convex hull of another region r_2, EC and DC are re placed by more specialized relations, resulting in a set of 23 base relations: RCC 23. Ignoring the difference be tween DC and EC results in a set of 15 base relations: RCC 15.

The 9-intersection model

Egenhofer and Franzosa present the same set of relations as those presented in RCC, but these were generated in a different manner. Originally, the 4 intersection model was developed. Here, two sets of points are associated with every region in R^2, its interior and its boundary. Only regions without holes, meaning 2 D connected objects with connected boundaries, are considered.

- The interior of *A* is defined to be the union of all open sets that are contained in *A*, that is, the interior of *A* is the largest open set contained in *A*.
- The closure of *A* is defined to be the intersection of all closed sets that contain *A*, that is, the closure of *A* is the smallest closed set containing *A*.
- The boundary of *A*, is the intersection of the closure of *Y* and the closure of the complement of *Y*. The boundary is a closed set (see **Figure 3**).

Considering empty and nonempty as the values of the intersections between the boundaries and interiors of two sets, a total of 16 topological relations can be character ized, each relation represented by a 2×2 matrix. If the sets are restricted to spatial regions in R^2, then the number of relations is reduced to eight: disjoint, overlaps, meets, equals, inside, contains, covers, and covered by.

The 4 intersection model was extended to the 9 intersection model, in which a third set of points is associated with every region, being the region's com plement. This approach resulted in 512 (2^9) matrices having three columns and three rows (see **Figure 4**). However, after taking into account all assumptions about the nature of regions in 2 D space, only eight matrices remain. These correspond once more to the RCC re lations, with: 'disjoint' equals is disconnected from (DC), 'overlaps' partially overlap (PO), 'meets' is externally connected to (EC), 'equals' is identical with (EQ), 'inside' is a nontangential proper part of (NTPP), 'contains' is the inverse of a nontangential proper part of (NTPPI), 'covers' is a tangential proper part of (TPP), and 'cov ered by' is the inverse of a tangential proper part of (TPPI).

The 9 intersection model and also the calculus based method (CBM) of Clementini, which is not presented here, are both Open GIS Consortium (OGC) standards models for topological relationships. Such models have been widely extended to other geographical primitives; points, lines, and bodies (object with a 3 D extension) in

R^2 and in R^3. Extensions have been made based on the dimension of objects intersections, number of inter sections, etc. Other interesting work is the comparison of topological relations in a continuous (vector) space of R^2 and a discrete (raster) space of $\mathbb{Z}^2$, applying the 9 intersection model in a raster model.

Orientation

Topology may be regarded as the most abstract spatial structure or the weakest geometry, which makes clear that there are many situations where topological infor mation alone is insufficient. This section explores an important complementary qualitative relationship, namely orientation: orientation relations describe where the objects are placed relative to one another and can be defined in terms of three basic concepts: the primary object, the reference object, and the frame of reference. The orientation of the primary object is then expressed with respect to the reference object as it is determined by the frame of reference. Thus, unlike the formerly pre sented topological relations, orientation is not a binary relation, that is, if we want to specify the orientation of a primary object with respect to a reference object, a frame of reference is necessary. The choice of reference system depends on the calculus one works in. Also due to the context of the spatial situation, one reference system can be more appropriate than another. When there is a sudden change in context, it appears to be preferable to use multiple frames of reference. An example situation occurs when a driver gets off a highway to enter a local street network. This is reflected in the use of maps having a different scale.

There are two main groups of qualitative orientation calculi: the first one having an extrinsic frame of refer ence and the second one having an explicit triadic re lation. In the first approach, the orientation of objects is obtained by using orthogonal or nonorthogonal pro jections of objects onto external axes, and by subsequent 1 D reasoning mostly using Allen's temporal logic. An example of this is Guesgen's approach. The major the ories of the second group are Frank's approach (cardinal direction calculus), Hernandez' approach, and the Fres ka's double cross calculus. Frank discusses the use of cardinal directions for spatial representation and rea soning. In order to determine each orientation relation, 2 D space is segmented around the reference object, by the use of the eight well known standard directional symbols (north, northeast, east, southeast, south, south west, west, and northwest) resulting in the cone shaped and the projection based models (see **Figure 5**). The eight cardinal directions of the cone shaped approach are extended with the qualitative value 0, representing an undetermined direction between points that are too close to each other to be able to give a cardinal direction. The eight cardinal directions of the projection based model are also extended with the qualitative value 0, standing for the so called neutral zone. These extensions greatly increase the power of the calculus, which is exemplified by typical qualitative reasoning methods such as the creation of CTs. This approach could be combined with qualitative distances, such as far and farthest.

Other Developments

Although topology and orientation are the most covered domains, there are other models and approaches in QSR. Without being exhaustive, we wish to show some of them to demonstrate the wide range of the field and its flour ishing research activities.

First of all, regions have been so far considered as having crisp boundaries. Introducing some vagueness in an object's boundary definition, some of the previously mentioned models have been extended to explicitly represent and reason about uncertain information. For instance, RCC postulates the existence of noncrisp re gions in addition to crisp regions and then adds another binary relation meaning that '*x* is crisper than region *y*'. This extension is known as the 'egg yolk' calculus, which essentially models regions with intermediate boundaries

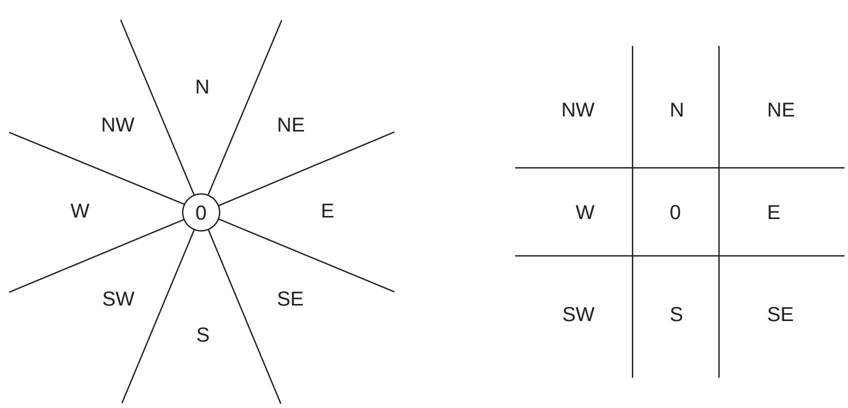

Figure 5 Cone-shaped and the projection-based orientation.

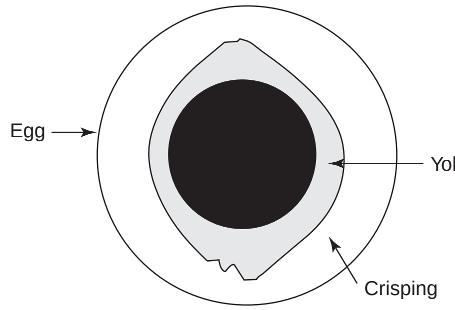

Figure 6 Crisp and broad boundaries: the 'egg-yolk' model.

as a pair of regions (see **Figure 6**). A rather similar ex tension of the 9 intersection model has also been de veloped, and refers to regions with broad boundaries. This calculus can also be specialized to cover other kinds of regions including convex hulls, minimum bounding rectangles, buffer zones, and rasters.

Distance is one of the most important aspects of space. When communicating about distances, people usually use qualitative categories like '*A* is close to *B*' or quali tative distance comparatives like '*A* is closer to *B* than to *C*'. If quantitative information can also be used when considering distance between two entities (absolute dis tance) like '*A* is about 1 m away from *B*', comparative distances are purely qualitative. Most approaches to qualitative distance consider points as the basic entities. Absolute distance relations are obtained, for example, by dividing the real line into several sectors such as 'very close', 'close', 'far', etc. depending on the chosen level of granularity. Relative distance can be obtained by com paring the distance to a given reference distance. How ever, such concepts lead to several reasoning difficulties, for example, transitivity of relationships may depend on the orientation, position of the points. This is why dis tance is often combined with orientation to get positional information. Several models have been developed, for example, one combining a cone based orientation ap proach with absolute distance relations (see **Figure 7**).

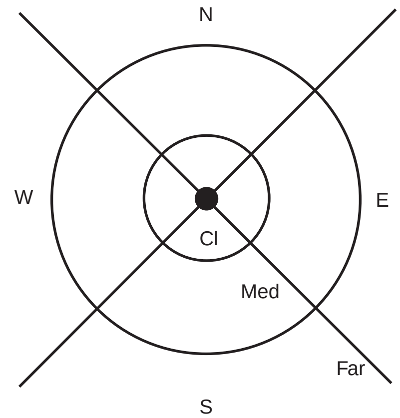

Figure 7 Cone-based orientation combined with absolute distance relations.

As it has been said when introducing orientation calculus, topological information is not always sufficient to deal with all the qualitative spatial situations. Between topological space and metric space, there are other mathematical spaces which can be used in QSR. One of them is the projective space, by Clementini and Billen, whose main invariant is collinearity (three collinear points remain collinear under projective transformation). Such property is the base of another type of relation ships; ternary projective relationships. Contrary to tern ary relationships mentioned in the orientation calculus section, these ternary relationships do not need an ex ternal frame of reference, but only a primary object and two reference objects. The space is partitioned into five zones based on the two reference object shapes. The relative position of the primary object against the two reference objects is determined by the primary object's intersections with the zones. Such approach is another way to model concepts such as being on the left, on the right, between, for example, an object *A* is left of objects *B* and *C* if *A* intersects the zone left (*B*, *C*).

Finally, we wish to mention time as a complementary concept of space. QSR is strongly linked with temporal reasoning, resulting in the development of spatio temporal reasoning. Recently, spatiotemporal calculi have been designed. For example, the spatio temporal con straint calculus (STCC) by Gerevini and Nebel com bines Allen's interval algebra with RCC 8. Moreover, the qualitative trajectory calculus (QTC) by Van de Weghe *et al.* allows for reasoning about moving objects.

Conclusions

QSR and reasoning is so close to human spatial cognition processes that it should take a bigger place in geography in the future. It is (should be) of prime importance in naïve geography, mental mapping, landscape analysis, and also in analyzing human behavior (with time, allowing analysis of spatiotemporal history), etc.

However, it is a new domain, and there are still a lot of things to be developed; some of the space aspects still have to be identified, formalized, and if possible associ ated to reasoning processes. Furthermore, such models, even if they are based on strong mathematical concepts, can be meaningless if not related to some human cog nitive perceptions. Cognitive validity of QSR models is one of the biggest challenges of this promising scientific domain.

See also: Categorical Data Analysis; Cognitive Geography; Complexity Theory, Nonlinear Dynamic Spatial Systems; Critical GIS; Mapping, Topographic; Qualitative Geographic Information Systems.

Further Reading

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM* 26(11), 832 843.
Clementini, E. and Billen, R. (2006). Modeling and computing ternary projective relations between regions. *IEEE Transactions on Knowledge and Data Engineering* 18(6), 799 814.
Clementini, E. and Di Felice, P. (1997). A global framework for qualitative shape description. *Geoinformatica* 1(1), 11 27.
Clementini, E., Di Felice, P. and Hernandez, D. (1997). Qualitative representation of positional information. *Artificial Intelligence* 95(2), 317 356.
Cohn, A. G., Bennett, B., Gooday, J. and Gotts, N. M. (1997). RCC: A calculus for region based qualitative spatial reasoning. *Geoinformatica* 1, 275 316.
Cohn, A. G. and Gotts, N. M. (1996). The 'egg yolk' representation of regions with indeterminate boundaries. In Burrough, P. & Frank, A. (eds.) *Proceedings of GISDATA*, pp 171 187. Taylor and Francis.
Cohn, A. G. and Hazarika, S. M. (2001). Qualitative spatial representation and reasoning: An overview. *Fundamenta Informaticae* 46(1 2), 1 29.
Cohn, A. G. and Montello, D. R. (2007). *Spatial Cognition and Computation*. http://www.sis.pitt.edu/∼scc (accessed April 2008).
Egenhofer, M. and Franzosa, R. (1991). Point set topological spatial relations. *International Journal of Geographical Information Systems* 5(2), 161 174.
Frank, A. U. (1992). Qualitative spatial reasoning about distances and directions in geographic space. *Journal of Visual Languages and Computing* 3, 343 371.
Frank, A. U. (1996). Qualitative spatial reasoning: Cardinal directions as an example. *International Journal of Geographical Information Science* 10(3), 269 290.
Freksa, C. (1992). Temporal reasoning based on semi intervals. *Artificial Intelligence* 54, 199 227.
Galton, A. (2000). *Qualitative Spatial Change*. Oxford: University Press.
Golledge, R. G. and Stimson, R. J. (1997). *Spatial Behavior: A Geographic Perspective*. New York: Guilford Press.
Guesgen, H. W. (1989). Spatial reasoning based on Allen's Temporal Logic, *Technical Report 89 049*, International Computer Science Institute, Berkeley, USA.
Hernandez, D. (1994). Qualitative representation of spatial knowledge. *Lecture Notes in Artificial Intelligence*. Springer Verlag (804).
Kitchin, R. M. and Blades, M. (2001). *The Cognition of Geographic Space*. London: IB Taurus.
Randell, D., Cui, Z. and Cohn, A. G. (1992). A spatial logic based on regions and connection. In Nebel, B., Swartout, W. & Rich, C. (eds.) *Proceedings of the 3rd International Conference on Knowledge Representation and Reasoning (KR)*, pp 165 176. San Mateo, CA: Morgan Kaufmann.
Renz, J. and Nebel, B. (2006). Qualitative spatial reasoning using constraint calculi. In Aiello, M., Pratt Hartmann, I. & van Benthem, J. (eds.) *The Logic of Space*, pp 161 215. Dordrecht: Kluwer Academic Publishers.
Stock, O. (ed.) (1997). *Spatial and Temporal Reasoning*. Dordrecht: Kluwer Academic Publishers.
Van de Weghe, N., Cohn, A. G., De Tré, B. and De Maeyer, P. (2006). A Qualitative Trajectory Calculus as a basis for representing moving objects in Geographical Information Systems. *Control and Cybernetics* 35(1), 97 120.

Relevant Websites

http://www.cosy.informatik.uni bremen.de
Cognitive System: Scientific Staff, Freksa's Research Group.
http://www.informatik.uni trier.de
University of Trier: DBLP Bibliography, The Spatial Cognition Conference.

Quantitative Data

M. Charlton, National University of Ireland Maynooth, Maynooth, Republic of Ireland

We outline some of the characteristics of both quantitative data and spatial data. We then consider the initial exploration of such data, one variable at a time with appropriate statistical and graphical summaries – inference is introduced in this section. This leads naturally into a consideration of multivariate techniques, including regression, data reduction, and classification methods. We outline some of the characteristics of spatial data and discuss some of the problems which arise in spatial data analysis. We briefly consider some methods for local analysis before spending some time examining methods for modeling and regionalization with spatial interaction data. An extended section deals with regression based methods for spatial data including spatial lag and spatial error models and geographically weighted regression. Finally, we look at some methods for spatial interpolation and conclude with a look at some software for quantitative analysis.

Quantitative Data

In many studies, quantitative data exists as a data matrix in which the rows represent the units of observation (residents of a town, counties, or streets within a city) and the columns represent measurement of some items of interest and relevance to the study in hand. The nature of the research question should have already defined the appropriate units of observation and the items to be measured. Items to be measured are sometimes known as attributes (particularly in geographic information systems (GIS)) or variables in a statistical study. The data may have been collected by the researcher in person, perhaps with the aid of a team of interviewers, or they may have been obtained from another source, for example, data about the residents of a set of counties in a population census.

Variables will contain either discrete or continuous measurements – discrete data have a finite number of categories, whereas continuous data can vary by infinitesimally small amounts. Discrete data are sometimes referred to as categorical data. The level of measurement of each variable is important in determining what methods are appropriate in the analysis of the data which has been collected.

Categorical data can be either nominal or ordinal: the categories in nominal data have no ordering (names of villages, gender, ethnic group), whereas there is an ordering to the categories in ordinal data (age cohorts, income groups, ranks). Discrete data can exist as interval or ratio level: interval data have no natural zero (zero on the centigrade scale does not imply an absence of temperature), whereas ratio level data have a well defined zero (age in years, pulse rate, unemployment rate, population count). The level of measurement helps determine which methods will be appropriate in an analysis.

Quantitative methods also include a body of techniques which deal with measures of location in the form of spatial references. The positional information might take the form coordinates of the object itself, if it can be represented spatially as a point, or the coordinates of some central location, for example, center of gravity, in an area. The measurements may be in degrees of longitude and latitude, or they may be in one or more projected coordinate systems. The analyst, faced with measurements in degrees, may convert these to projected coordinates before beginning the analysis.

It is also necessary to consider the data representions in GIS because there is a close connection between quantitative methods and GIS. Discrete objects in the real world can be expressed in a GIS data base as points, lines, or polygons – the vector model. This representation consists of geometric information indicating the position of the object and associated attribute information which describes its characteristics. Continuous phenonomena such as population density or elevation may be represented as a grid or raster where the study area is divided into a regular lattice (usually with square cells), each cell containing the value of the phenomenon of interest. The GIS representation may be convenient for analysis: adjacency matrices are easily created from polygon data, which are used in a range of quantitative techniques – GIS software packages often provide an extensive range of spatial analysis and modeling methods.

Exploration

Most analyses begin with some exploratory analysis – this helps in understanding the nature of the distributions of the variables which are to be used later in the study. For continuous data parametric statistics, such as the mean, variance, skewness, and kurtosis, provide useful summaries. Nonparametric indicators including the median, quartiles, and interquartile range are conveniently computed. Such statistical summaries may also be accompanied by plots of the distribution shapes – histograms and boxplots are useful displays for examining the shape of the distribution (is it symmetrical, is it skew, are there outliers?).

For projected coordinate data, summary measures exist – these are known as centrographic statistics and include the mean and median centers, as well as measures of spatial dispersion. There is also a body of statistical theory and practice which deals with directional data – the 'average' of a set of bearings is known as the directional mean and the variance is known as the circular variances. The theory of directional statistics provides a framework for inference, tests for uniformity, and goodness of fit to theoretical distributions.

Categorical data may be explored by tabulating the frequencies in the individual categories. Visual methods include histograms – of interest is the modality of the distribution (has it one or more?) and the class sizes. One may also be interested to determine whether the distribution follows some theoretical distribution – might the distribution have arisen from a Poisson process?

Tukey has suggested a range of exploratory methods – these include both numerical and visual summaries, some of which may be available in existing software. As analysis in the context of research in human geography, maps are convenient for summarizing the spatial pattern of both continuous and categorical data.

There are a number of methods of examining bivariate relationships: which one is used depends on the research question and the nature of the data. Pearson and Spearman correlation coefficients are useful for measuring the strength of the association between two continuous variables. A chi squared statistic can assist in determining the relationship between two categorical variables. The relationship between categorical and continuous variables might involve computing the means and variances of the continuous variable in each category of the other variable.

Exploratory visual analysis of bivariate relationships may employ a number of tools. The scatterplot displays the relationship between a pair of continuous variables – with more than two a scatterplot matrix is sometimes helpful. Boxplots may be useful for exploring variation of a continuous variable within the categories of a nominal or ordinal variable.

Given the values of two means computed from two samples, it is reasonable to ask the question whether there is sufficient evidence for the analyst to conclude that the difference between them is large enough to conclude that it is not simply due to chance. Statistical inference deals with selecting and using the appropriate tests to address these kinds of questions. If the two means are drawn from sufficiently large samples, the analyst would use a z test to determine whether the difference is not zero ('significant' in statistical parlance); if the sample sizes are small, a t test would be appropriate. A nonparametric version of the test is the Mann–Whitney test. If there are several groups, means analysis of variance may be used to determine whether any differences between the means are small enough to be due to chance variation. The underlying theory of such tests assumes that the observations under analysis are independent – this is not always the case with spatial data. However, in spite of this, inferential testing is widely used with spatial and nonspatial data.

Methods for Multivariate Data

With multivariate data, there are a wide range of tools for exploration and analysis. Regression techniques relate the variation in one variable (the y , dependent, or response variable) to the variation in one or more independent (x or predictor variables). The most commonly encountered form is ordinary least squares (OLS) regression in which the y variable is modeled as a linear combination of the independent variables, with an error term which is normally distributed with zero mean. If the dependent variable is one of counts (say cases of a disease) then Poisson regression is appropriate, and if the dependent variable is binary, the logistic (binary logit) regression is a useful model. Generalized linear models extend the regression framework by providing a wide range of options for modeling the error terms and the relationships among variables. A typical linear regression equation (model) with n variables might take the form

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n + \varepsilon$$

y is the dependent variable, $x_1, \ldots, x_n$ are the independent variables, ε is the error term, and $\beta_0, \ldots, \beta_n$ are the parameters which are to be estimated. Estimation using OLS attempts to find the combination of parameters that minimizes the sum of the squared differences between the observed values of y (dependent variable) and the predicted values of y computed from the equation. The dependent variable is a linear combination of the independent variables: multiplying each variable by its coefficient and summing up the resulting products yields the predicted values.

An important part of regression is the interpretation of the coefficients in the model, in particular, whether they are sufficiently different from zero (a variable with a zero valued coefficient will not contribute to the variation in the dependent variable). It is also important to examine the residuals from the model (these are the differences between the actual and the predicted (fitted) values). The residuals should have a zero mean, and should exhibit homoscedasticity (i.e., any samples taken from the residuals should have the same variance). With spatial data it is also necessary to examine the residuals for any spatial pattern – if there is significant autocorrelation, the coefficients will be unreliable because the assumptions on which the model has been estimated will have been violated.

There are many data reduction methods available to the analyst. 'Principal components analysis' is used to create a number of new variables, known as components, from the given data. Components are linear combinations of the original variables. While there will be as many components as the original variables, typically, only a subset of these is used. The component loadings may be interpreted (per haps to create a deprivation measure), and the component scores (the values of the new variables for each observation) may be useful for subsequent analysis. Relationships be tween two sets of variables can be explored using pairwise correlations, or canonical correlation analysis.

To group observations together into related families or clusters, there is a wide range of 'cluster analysis' techniques – these take the values for a set of variables and attempt to create clusters of cases which have similar characteristics. Agglomerative methods will successively fuse cases together into a single group – fusion is usually stopped earlier to create the desired number of groups. Less frequently encountered are divisive techniques which start with a single group and attempt to split it successively into the component cases. Among the non hierarchical techniques, *k* means is commonly used – the user must specify the number of desired groups in ad vance of the analysis; so typically, a series of cluster analyses will be taken and compared before the most appropriate one is chosen. Mapping the results of a cluster analysis can yield useful insights. The cluster characteristics can be obtained by computing the means of the variables in each cluster – visualization techniques such as parallel coordinate plots may be helpful in examining the individual cluster structures.

Should the analyst have an existing classification into a number of groups of the objects under study, linear 'dis criminant analysis' is useful for relating the classification to a set of independent variables. Linear combinations of the independent variables give rise to discriminant func tions which discriminate between the groups. Examining the discriminant functions can reveal which variables are important in distinguishing between the groups. One of the outputs is a prediction of the group membership for each case, given the values of variables for that case. This is often cross classified with the input classification to create a confusion matrix – cases whose group member ship has been correctly classified will lie on the leading diagonal. Misclassified cases warrant further investigation into the reasons for the incorrect assignment.

Problems with Spatial Data

Spatial data can provide the analyst with some awkward problems. Secondary data, perhaps from a census, are often aggregated to a set of areal units for reporting purposes. It is common to use such data to compute means, variances correlations, fit regression models, and carry out more complex tasks without thinking too deeply about the implications of using such data.

It has been known since the 1930s that aggregating data from smaller to larger areal units alters the variance structure of the data – two sociologists, Charles Gehlke and Katherine Biehl, noticed that aggregating data to larger areal units leads to larger estimates of the correl ation coefficient. This phenomenon has been labeled the modifiable areal unit problem (MAUP) – there are two related issues to the MAUP. The aggregation or 'scale effect' leads to the behavior observed by Gehkle and Biehl. There can be a large number of ways of aggregating a set of n smaller areal units into m larger areal units where $m < n$ – this 'zoning effect' will also influence the value of statistics derived from spatially aggregated data.

A second problem which impinges on the use of quantitative approaches in human geography is that of spatial autocorrelation – the association between data values over space. Positive spatial autocorrelation occurs when objects have similar values of a variable as their neighbors; negative spatial autocorrelation occurs when neighboring values are dissimilar. Spatial autocorrelation is a problem – positive autocorrelation leads to an under estimate of the standard error of the mean, negative autocorrelation leads to an overestimate, with con sequences for any inference that might be made. There are several diagnostics for examining the extent to which spatial autocorrelation is present – the Getis Ord G statistic and Moran's I are two widely used global meas ures of spatial autocorrelation. There are also a group of local indicators of spatial association, in particular a vari ant of Moran's I which can be used to determine whether such association varies locally across the study area.

Spatial Pattern

The analysis of spatial pattern attempts to determine the underlying processes which lead to such patterns in points, lines, and areas. The search for evidence that some observed pattern has not arisen from some chance process leads into a range of techniques. Many of these deal with point pattern analyses: summary measures in clude the mean center and standard distances; inferential techniques include quadrat and nearest neighbor analy sis. More recent activity has been concerned with mod eling such processes, in particular determining whether points in some study area exhibit clustering. Kernel density estimation is used to determine the variation in the intensity of the point process in the study area, and a useful technique for characterizing any clustering of the points is through *K* functions.

Computing the density of a spatial pattern – perhaps the locations of residential burglaries or road accidents – usually involves GIS software. Where the spatial pattern exists within a spatially clustered base population, the spatial variation in the base population needs to be controlled. Possible approaches here include Openshaw's 'geographical analysis machine' (GAM) – the original intent of this was to identify clusters of disease cases in a spatially inhomogenous population. While GAM received some criticism, it was one of the first examples of a scan statistic (a statistic that scans an area to search for local clusters), its output was directly mappable, and it stimulated the development of further methods, notably those of Besag and Newell, and Bithell and Stone. Interest in methods for identifying localized clustering has extended outside geography, notably into crime pattern analysis. Many of the methods used are incorporated into the CrimeStat software.

Spatial Interaction

Spatial interaction deals with movement between locations. Movement here refers to people, objects, or information, and for any single transfer, there will be an origin and a destination. The types of movements which are frequently dealt with include journey to work, migration, freight flows, and telephone traffic. One particular type of information which is often collected as part of census of population enumeration is journey to work data. The forms of this may vary but commonly the analyst may be presented with the numbers of individuals moving between pairs of zones. These flows could be represented in matrix form, with the origin zones forming the rows and the destination zones forming the columns – the cell at the intersection of a particular row/column pair is the flow of workers between the corresponding pair of zones. Flows are then of two types: interzonal, the origin and destination zones are different, and intrazonal, the flow from residences to workplaces is wholly contained within the zone. The intrazonal flows are those along the leading diagonal of the flow matrix.

Depending on the data supplier, there might be some disaggregations of the total flows – by gender, by age group, by mode of transport, by industry type, or by socioeconomic group.

It is not unreasonable to wish to model the variation in such flows – more journeys may well be made between zones which are closer together than zones which are farther apart. Affluent residents may travel farther than less well off ones. Trips made by public transport may be shorter on average than those made by car. The attributes of the residential zones may well influence the number and type of trips which originate from them; equally the attributes of the destination zones may well influence the numbers of workers attracted. Early modeling efforts employed analogs from physics to create 'gravity models'; Wilson's attempts to place spatial interaction modeling in a sounder theoretical framework led to the development of 'entropy maximizing techniques', and Fotheringham's 'competing destinations model' represents further refinement.

Models for dealing with such data fall into a broad group known as spatial interaction models. Generally, such models attempt to predict the flows between zone pairs, using information on the travel costs between them, and the different characteristics of the zones as origins and destinations. Fitting these models to the observed flow data entails a process known as calibration – this is often computer intensive. Deriving the characteristics of the zones as origins and destinations requires collating a wide variety of data (e.g., estimation of zonal floor space and money turnover of workplaces; and computation of the travel cost based on road distances and types). The flows are integers, so Poisson models are commonly used to avoid the prediction of negative flows.

Flow data need not be confined to journey to work flows. Census enumeration often leads to detailed migration flow data, and a series of research questions on the drivers of migration. The outcome of these exercises to understand the causes of migration may lead to the formulation of policies for influencing it.

Regionalization

A related activity using flow data is the construction of functional regions. Most methods restrict themselves to the information in the flow matrix. The idea is to combine zones into larger areas – when commuting data are used for this purpose, these larger areas are variously known as travel to work areas, local labor market areas, or functional regions. There are various algorithms – generally they start by combining the two zones with the largest interaction (the flow data are usually transformed). The two rows and two columns representing these zones are combined and the interactions recomputed for the smaller matrix. The largest interaction is again identified and the zone pair fused. This continues until all zones have been fused together into the desired number of regions. There are many variations on this basic approach, some of which require the analyst to supply a set of parameters to control the fusion process.

Specialized software is usually required for regionalization. The analyst may be required to code the methods from descriptions in the published literature. Software implementing the 'intramax' procedure is downloadable. The output is usually a list of the original zones and the identifier of the region to which each has been allocated. Further analysis will generally involve GIS software to create the boundary files for mapping and data re or ganization. The analyst should be aware of any effects of the modifiable areal unit problem when undertaking further analysis.

Spatial Regression

OLS regression models are widely used in quantitative analyses. However, for the geographer they can be problematic because, among a range of assumptions, they are required to be independent of the residuals. It has been observed that dependencies often exist in spatial data which lead to a violation of one or more of the assumptions of the OLS model – typically, we find that the residuals are spatially autocorrelated. Spatial re gression models extend OLS models in an attempt to deal with spatial autocorrelation.

There are a number of model forms which add terms to the OLS model to deal with spatial dependence. 'Spatial autoregressive' (spatial lag) models consider that the level of the response variable is dependent not only on the levels of the independent variables in the model, but also on neighboring values of the response variable itself. An additional term in the model is used to repre sent this:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n + \rho W y + \varepsilon$$

where W is a spatial weights matrix and ρ is an additional coefficient to be estimated which measures the auto correlation. The spatial weights matrix contains 0 where zones are not contiguous; if there are m neighboring zones, the cells in the row corresponding to the adja cencies contain the value $1/m$. This ensures that the additional term contains the mean of the y variable in the neighboring zones. The means may be weighted using a Euclidean or non Euclidean distance such that nearer zones are given more influence in the computation of the adjacent mean value.

An alternative form considers autocorrelation in the error term. The model structure takes the form:

$$\begin{aligned} y &= \mathbf{X}\beta + \varepsilon \\ \varepsilon &= \lambda W \varepsilon + u \end{aligned}$$

The error term ε has a spatially weighted component $\lambda W \varepsilon$ and a random error term u.

Such models are referred to as 'spatial error' models but they are also known as 'spatial moving average' models. Downloadable software, GeoDa, is available for estimating a range of models including the spatial lag and the spatial error models.

One of the assumptions with conventional regression models is that the regression equation is identical in form for every observation. If the process being modeled is not spatially homogenous, then any spatial structure will be revealed in the residuals from the model: it will be spa tially autocorrelated. However, one of the basic as sumptions of regression is that the residuals are uncorrelated, so any autocorrelation is undesirable.

One approach to dealing with the problems of mod eling with data concerning individuals and their neigh borhoods (or different levels in a spatial hierarchy) is 'multilevel modeling'. This modeling framework allows the incorporation of data concerning individuals simul taneously with information about the neighborhoods in which they live or work. Specialized software is required to calibrate multilevel models.

A simple technique to incorporate spatial effects into a model is to include place specific dummy variables, perhaps to deal with broad regional effects where the regions cover several of the underlying data cases. Dummy variables are those which take the value 0 or 1, often representing yes/no or presence/absence. The ef fect of the dummy variables, if the parameter estimates are significant, is to move the regression line up and down the y axis – they modify the value of the intercept term for each region. Interaction terms (in which a dependent variable is multiplied by a regional dummy variable) will allow the value of that variable's parameter to be modified for a particular region – the combination of regional dummy variables and interaction allows the estimation of different regression structures in different regions. However, the multiplicity of terms in the model rapidly leads to difficulties in interpretation.

A number of approaches have been suggested for allowing the model structure to vary across the study area to deal with this problem. Among the earlier attempts is the 'expansion method' developed by Casetti and Jones. In these models, the parameters are functions of location (with coordinates u_i,v_i)

$$y_i = (\alpha_0 + \alpha_1 u_i + \alpha_2 v_i) + (\beta_0 + \beta_1 u_i + \beta_2 v_i) x_i + \varepsilon_i$$

The linear expansions of the parameters allow their values to 'drift' over the study area. In the example above, if α_1 and α_2 were positive, the value of the intercept would increase toward the northeast of the study area; if β_1 and β_2 were negative, the value of the parameter for the variable x would increase toward the southwest of the study area. Other forms for the expansion are possible, for example, quadratic or cubic. The analyst must specify the form of the expansions (linear, quadratic, cubic) prior to analysis, so if the model structure changes in some manner other than that dictated by the expansion func tions, there may still be spatial structure present in the residuals.

A more recently developed technique is 'geographic ally weighted regression' (GWR) – this generalizes the concepts of spatial drift. This model was proposed by

Fotheringham, Brunsdon, and Charlton, who have written widely on GWR. The model can provide local parameter estimates at the locations where the observations have been sampled or at any other location in the study area. GWR models have the general form

$$y(u) = \beta_0(u) + \beta_1(u)x + \cdots + \beta_n(u)x + \varepsilon$$

where u is some location in the study area – a set of parameters are estimated for every observation in the dataset. The coordinates for zones are usually taken at some central point in the zone. To understand the nature of the geographical weighting we must consider the estimator:

$$\hat{\beta}(u) = (XW^{T}(u)X)^{-1}X^{T}W(u)y$$

The geographical weighting is provided by the matrix W which is specific to a particular location. Observations which are closer to the location u are given greater weighting in the estimation and observations which are further away receive a lower weight. The weighting functions typically have a Gaussian like shape, and the influence of the distance between any observation and the location u is controlled by a parameter known as the bandwidth – this is supplied by the analyst and can be estimated from the data.

The parameter estimates, and their associated standard errors, can be mapped. Also mappable are the residuals and local goodness of fit measures. Significance tests on the individual parameters are possible, as is a significance test to determine whether the spatial variation in the parameter estimates for a single variable differs from random.

The GWR approach is quite flexible. Typically, the parameter estimates are made at the locations at which data are available. However, parameter estimates can be made at other locations in the study area, and if independent variables exist for those locations, estimates of the dependent variable can be made at these locations.

GWR is a model of spatial heterogeneity. Interesting developments are in the area of semi parametric GWR models – in these models, some variables have spatially fixed parameters and others have spatially varying parameters. Model selection – particularly in deciding which parameters should be fixed and which should vary – becomes an important issue. Estimating such models requires special software: Fotheringham, Brunsdon and Charlton have written suitable software; there is a GWR package for R, and Gaussian GWR models can be estimated in ArcGIS 9.3.

For mapping parameter estimates, GWR output can be imported into suitable mapping software. For mapping convenience, a model can be estimated at the sample points to determine a bandwidth, and then estimates made at the mesh points of a regular grid to create a parameter 'surface'.

Spatial Interpolation

A related activity is spatial interpolation. Typically, the analyst has values for a variable which have been collected at one set of locations and requires estimates of values for another set of locations in the study area – usually the mesh points of a grid. These techniques require the presence of spatial coordinates for each observation. Spatial interpolation techniques are widely used to create grid based digital elevation models from irregularly spaced measurements of elevation, but they can also be used in human geography research.

A commonly used technique is 'inverse distance weighting'. This provides a distance weighted mean estimate of the measurements. Given the values of a variable of interest, an estimate of the mean at a location u is provided by

$$\hat{x}(u) = \frac{\sum_{i=1}^{n} d_i^{-\alpha}(u)x_i}{\sum_{i=1}^{n} d_i^{-\alpha}(u)}$$

where $d_i(u)$ is the distance from the ith observation to the location u, and α is an exponent supplied by the analyst. A typical value for the exponent is 2. Enhancements sometimes include restricting the local sample size – only the closest m points to u (where $m < n$) are used.

Another approach is provided by 'kriging'. Kriging is associated with the field of geostatistics. The estimates are derived from a distance weighted mean, but with kriging, the weights are computed from a theoretical variogram whose parameters, the nugget, sill, and range, are estimated from the observed data. This is usually a two stage process – first, the analyst computes an empirical semi variogram from the data which measures the spatial covariance of the data at various distances; to this is fitted a theoretical variogram by choosing suitable values for the parameters. As well as estimates of the mean of the variable being kriged, standard errors on the estimates are also available.

We can think of spatial data to be interpolated using kriging as being composed of a deterministic component and a stochastic component. The deterministic component may be a mean or a more complex spatial drift term – residuals are left after the deterministic component has been removed from the data. The kriging process attempts to estimate the value of the residuals at a particular location, given the values of the residuals in the neighborhood. To generate the residuals, the analyst provides a suitable mean model: in 'simple kriging' the analyst supplies a value for the mean of the data; in 'ordinary kriging' this is computed from the data. In 'universal kriging' a more complex model is used to remove spatial trend, or 'drift', from the data – typical regression models used here include linear and quadratic forms. More complex models may be used – a study of house price variation might model the trend as a hedonic regression. Once the re siduals have been computed from the sample data, the errors can be estimated at other locations, and the mean or trend added back in to provide the final estimate. Kriging requires specialist software – Pebesma's gstat package and ESRI's geostatistical analyst provide powerful tools.

Software for Quantitative Analysis

There is a range of computer software which is available to the analyst. These include comprehensive collections of statistical methods running on personal computers or workstations – notable examples are SPSS and SAS. The analyst is required to prepare a file of data which is then converted by the program to an internal format. The analysis involves making choices from logically organized menu structures, and interpreting the output correctly. While such programs are relatively easy to use, they will not prevent mistakes being made, such as creating de scriptive statistics intended for continuous data from suitably coded categorical data. Other software may be more specialized and less widely used by geographers, for example, Statistica and Stata.

Some geographers have created their own software for spatial analysis: Luc Anselin's GeoDA, and GeoVISTA created by Pennsylvania State University Geographic Visualization Science, Technology, and Applications Center are examples. GeoVISTA is easily modified to create new tools in emerging areas such as geovisual analytics. Other examples of more task specific software include Crimestat (aimed at crime analysts) and Clus terseer, Boundaryseer, and SpaceStat which are marketed by TerraSeer Inc.

At the other end of the spectrum is R, a widely used statistical computing environment which is freeware. R has a rich scripting language which the analyst must use to perform the analyses which are required – statements in the language are typed into a text window. This in volves remembering the names of a wide range of stat istical functions, their syntax, and the syntax of the scripting language itself. While R is powerful and flex ible, it is not always easy to use. There is a large library of add on packages which have been written by R users from a wide range of disciplines: these include several libraries for spatial analytic methods, including kriging and geographically weighted regression.

The Future

Many of the techniques described here have been de veloped in the last 30 years. The development of GIS has been among the influences that have driven researchers to seek solutions to some of the awkward problems which are provided by spatial data. Much parallel activity has also taken place outside geography, notably in spatial epidemiology and spatial economics. Great progress has been made by statisticians in the field of spatial statistics – there is a rich variety of spatial models available for the analyst when dealing with spatial problems. There is a strong future for quantitative approaches in human geographical research.

See also: Categorical Data Analysis; Ecological Fallacy; Exploratory Spatial Data Analysis; Factor Analysis and Principal-Components Analysis; First Law of Geography; Geocomputation; Geographically Weighted Regression; Kriging and Variogram Models; Longitudinal Methods (Cohort Analysis, Life Tables); Markov Chain Analysis; Modifiable Areal Unit Problem; Monte Carlo Simulation; Movies and Films, Analysis of; Multidimensional Scaling; Naturalistic Testing; Neighborhood Effects; Network Analysis; Neural Networks; Oral History; Oral History, Ecological; Overlay (in GIS); Participant Observation; Participatory Action Research; Participatory Video; Performance, Research as; Photographs; Point Pattern Analysis; Polyvocality; Psychoanalysis; Q Method/ Analysis; Qualitative Geographic Information Systems; Quantitative Data; Quantitative Methodologies; Questionnaire Survey; Regionalization/Zoning Systems; Regression, Linear and Nonlinear; Reliability and Validity; Remote Sensing; Representation and Re-presentation; Sampling; Scale Analytical; Scientific Method; Segregation Indices; Selection Bias; Shift-Share Analysis; Simulation; Situated Knowledge, Reflexivity; Sound and Music; Space-Time Modeling; Spatial Analysis, Critical; Spatial Autocorrelation; Spatial Clustering, Detection and Analysis of; Spatial Data Mining, Cluster and Pattern Recognition; Spatial Data Mining, Geovisualization; Spatial Data Models; Spatial Expansion Method; Spatial Filtering/Kernel Density Estimation; Spatial Interaction Models; Spatial Interpolation; Spatially Autoregressive Models; Statistics, Descriptive; Statistics, Inferential; Statistics, Overview; Statistics, Spatial; Structural Equations Models; Subalternity; Subjectivity; Text, Textual Analysis; Thiessen Polygon; Time Geographic Analysis; Time Series Analysis; Time-Space Diaries; Transcripts (Coding and Analysis); Translation; Trend Surface Models.

Further Reading

Bailey, T. C. and Gatrell, A. C. (1995). *Interactive Spatial Data Analysis*. Harlow: Longman.

Bivand, R., Pebesma, E. J. and Gómez Rubio (2008). *Applied Spatial Data Analysis with R*. New York: Springer.

Cressie, N. A. C. (1993). *Statistics for Spatial Data* (revised edn.). New York: Wiley.

Fischer, M. M. and Getis, A. (eds.) (1997). *Recent Developments in Spatial Analysis: Spatial Statistics, Behavioural Modeling and Computation Intelligence*. Berlin: Springer.
Fotheringham, A. S., Brunsdon, C. and Charlton, M. (2000). *Quantitative Geography: Perspectives on Spatial Data Analysis*. London: Sage.
Fotheringham, A. S., Brunsdon, C. and Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: Wiley.
Haggett, P. (1965). *Locational Analysis in Human Geography*. London: Edward Arnold.
Mardia, K. V. and Jupp, P. E. (1999). *Directional Statistics*. Chichester: Wiley.
Schabenberger, O. and Gotway, C. A. (2005). *Statistical Methods for Spatial Data Analysis*. Boca Raton, FL: Chapman and Hall.
Thomas, R. W. and Huggett, R. J. (1980). *Modelling in Geography: A Mathematical Approach*. London: Methuen.
Upton, G. and Fingleton, B. (1985). *Spatial Data Analysis by Example 1 Point Pattern and Quantitative Data*. Chichester: Wiley.
Upton, G. and Fingleton, B. (1989). *Spatial Data Analysis by Example 2 Categorical and Directional Data*. Chichester: Wiley.
Waller, L. A. and Gotway, C. A. (2004). *Applied Spatial Statistics for Public Health Data*. Hoboken, NJ: Wiley.
Wrigley, N. (ed.) (1979). *Statistical Applications in the Spatial Sciences*. London: Pion.

Quantitative Methodologies

M. F. Goodchild, University of California, Santa Barbara, CA, USA

Glossary

Central Place Theory A body of theory about the locations, sizes, and offerings of settlements in agricultural landscapes.

Data Mining The application of computational methods to large volumes of data in an effort to detect pattern.

Data Model A template or format for data.

Geographic Information System Software for the analysis of geographic data.

Google Earth A web-based service for displaying information about the surface of the Earth.

Interval Data Data values that can be subtracted to establish differences.

Least Squares A principle used to fit a mathematical function to data.

Linear Regression The fitting of a linear relationship between two variables in a sample.

Neural Network A form of analysis originating in artificial intelligence.

Nominal Data Data values that serve only to differentiate.

Null Hypothesis A statistical proposition concerning a sample's relationship to its parent population.

Ordinal Data Data values that establish order or rank.

Ratio Data Data values that can be divided to establish ratios.

Self-Organizing Map A form of analysis originating in artificial intelligence.

Social Physics The application of concepts from physics to social systems.

Spatial Interaction Model A model of the interaction between an origin and a destination that includes the impeding effect of intervening distance.

Tobler's First Law The assertion that "all things are related but nearby things are more related than distant things."

Introduction

Ever since the development of counting and measurement systems, humans have recognized that quantities are more useful in many respects than qualities. The administration of land in the Fertile Crescent and the Nile Delta required measurements of length and area, and basic principles of geometry. Navigation required the ability to measure direction and position on the Earth's surface, and led to the science of map projections. It is clearly important for a retailer to be able to predict how many shoppers will patronize a store, and how much they will spend. The knowledge that an earthquake will strike is of far less value than the knowledge of precisely where and when.

In human geography, however, there is far less agreement on the value of quantification, and many of these issues are far from settled. Concern is often raised over whether humans can ever be predictable, and whether measurement is not of its very nature inhumane. Measurements of humans have been used for purposes that are now discredited, such as the nineteenth century eugenics of Galton. Thus, the urge to quantify has waxed and waned in human geography over the years. Social physics, or the adaptation of principles from physics to society, has occasionally caught the imagination, as it did in the 1950s with the work of Zipf, Stewart, Warntz, and others, who sought to apply Newton's law of gravitation and related principles to social phenomena. One of the more successful threads of this research led eventually to the models of spatial interaction that today provide invaluable predictions of traffic counts, retail store patronage, telephone traffic, and many other phenomena.

The heyday of quantification occurred in the 1960s during geography's Quantitative Revolution, when a generation of young scholars, many trained at the University of Washington, argued for a focus on theory and its empirical verification using quantitative statistical techniques. The spirit of the time is best captured in Bunge's *Theoretical Geography* and Harvey's *Explanation in Geography*. Quantification has made enormous strides since then, and today quantitative methods and scientific reasoning are as fundamental to human geography as they have always been to physical geography.

While the distinguishing characteristics of quantitative methodologies – a reliance on quantities, in the form of counts or measurements – may seem straightforward, on closer examination it is possible to distinguish several somewhat independent threads. The following sections discuss each of them in turn.

Qualitative and Quantitative

Theories of data identify four basic forms: nominal, ordinal, interval, and ratio (directional data is sometimes identified as a fifth). Data are said to be nominal if their purpose is solely to distinguish one instance from another. Thus names of people are nominal, as are classes of land use. Nominal data may be numeric, telephone numbers and social security numbers being obvious examples, but in such cases it makes no sense to perform arithmetic operations. The telephone number 8938049 is not more or less than the number 8933146. Data are said to be ordinal if their purpose is to express order, or to rank. Thus first may be better or more than second, and class 1 land may be better than class 2 land. Again, ordinal data may or may not be numeric: the colors of the rainbow provide an order, for example. Interval data are numeric, and differences between them are meaningful. Data expressed using the Celsius scale of temperature have interval properties, for example, and it makes sense to perform such arithmetic operations as averaging. Finally, both differences and ratios are meaningful for ratio data. It makes sense, for example, to say that a weight of 20 kg is twice a weight of 10 kg.

In this framework nominal and ordinal data are qualitative, and interval and ratio data are quantitative. This distinction is precise, but it hardly explains the very strong cleavage that exists between proponents of quantitative and qualitative geographic methodologies, and the tendency for researchers to align themselves with one side or the other at an early career stage. After all, it is easy to make quantitative measurements of the area occupied by class 1 land, an ordinal and therefore qualitative property, or to develop quantitative models that attempt to predict such qualitative properties as whether or not an individual will migrate in a given time period.

Instead, the labels have come to characterize methodological differences that are far more profound and nuanced than the grouping of four data types into two categories would suggest. First, quantitative methodologies are strongly associated with the sciences. There is a willingness to believe in the process of statistical inference, which attempts to extract truths of general applicability from the study of limited samples, using rigorous reasoning. There is a belief in replicability, that investigators should reach the same conclusions from the same experiments; and in the need for universally accepted terms with rigorous definitions. In these respects quantitative human geography resembles physical geography and other environmental, life, and physical sciences in its fundamental assumptions and goals. Qualitative methodologists on the other hand may question many of these assumptions, and employ a wider range of techniques that stretch them to varying degrees. They may express doubts about the value of conventional philosophies of science when applied to social phenomena, and be willing to embrace knowledge that is not necessarily replicable; and will likely view their methods as uniquely applicable to the social sciences.

The Role of Theory

Quantitative approaches are firmly grounded in the notion of theory, or the possibility of general propositions about the domain of human geography. Much of the fuel for the quantitative revolution derived from Central Place Theory, a body of propositions developed by Christaller and Lösch, addresses the patterns of settlements that could be expected to develop on an agricultural plain. Under fairly stringent assumptions about the behavior of consumers and entrepreneurs, Christaller and Lösch concluded that settlements should form a discontinuous hierarchy, with the smallest centers offering only limited ranges of goods and the most expensive goods being offered only in the largest centers; and that settlements should position themselves precisely at the nodes of a hexagonal network.

In the best scientific tradition, efforts were made to find confirmation of the theory by comparing its predictions to actual patterns of settlements, using areas such as Iowa that could be assumed to approximate a uniform agricultural plain. Efforts were also made to adjust the theory to specific circumstances, when for example the distribution of rural consumers was not uniform, due perhaps to spatial variation in agricultural productivity and the value of crops. Today, geographers think of 'central place theory' as a set of ideas that can be used to structure investigations of settlement patterns; but have largely rejected the notion that it might provide a precise model of the social world. In that sense the paradigm has come to resemble the one dominant in economics: propositions about the social world that are general but not necessarily in agreement with reality, but that nevertheless form a framework for understanding. Thus, quantitative human geographers still believe in an interplay between theory and empiricism, in the best traditions of experimental science. The process of induction draws on exploration of the social world, suggesting general principles that accumulate into a body of testable theory; and the process of deduction draws on theorizing to suggest new principles that can be tested against reality.

If theory offers precise predictions, then it follows that quantitative methods will be needed to test them. Theory is necessarily general, so the methods used to test theory must involve large numbers of samples, and formal investigations of whether the samples confirm or deny the theory. Moreover, unlike theories about the physical world, it seems inevitable that theories about the social world must be less than perfect in their predictions – that the goal of perfect prediction is fundamentally unachievable, if only because humans are free to contradict predictions about their own behavior. Predictions that are less than perfect require large numbers of samples to confirm them, unlike perfect predictions which a single counter example can refute. Thus, there are many reasons why a human geography that is concerned with the discovery of general, testable truths should align itself with quantitative methodologies.

Statistical Inference

Statistical inference originated in the life and physical sciences, and in concern over what could be concluded from experiments involving limited numbers of samples. For example, a field might be sown with two types of seeds, using 100 seeds of each type, and while one type appears to result in larger plants, there is overlap in the results – some plants from the better seed are smaller than some plants from the poorer seed. Is the apparent improvement a result of chance, or does it indicate a real superiority of one seed type over the other? A counter or null hypothesis is posed, in this case that the two seed types are equally productive, and the probability determined that the actual experimental result could occur if the null hypothesis is true. If this probability falls below some threshold, typically 5%, the effect is said to be statistically significant and the null hypothesis is rejected.

Several broad classes of inferential tests can be identified, depending on the nature of the null hypothesis. This example is a two sample test, for which the null hypothesis is always that both samples were drawn from the same population. In other instances a single sample is compared to some theoretical proposition, such as a population with a specific mean value, and the null hypothesis proposes that the sample was drawn from the theorized population. Statistical inference can also be used to evaluate a numerical relationship between two variables, in which case the null hypothesis proposes that the sample was drawn from a population in which the variables are statistically independent. Of particular interest to geographers are tests of spatial pattern against null hypotheses of spatial randomness.

The practice of statistical inference always carries risk. If a null hypothesis is rejected at the 5% level of significance then the researcher accepts a 5% chance of being wrong, in other words the null hypothesis is in fact false. On the other hand, if it is accepted, it is possible that a weak effect is nevertheless present, and a larger sample would show the opposite result. These two outcomes are described as type I and type II statistical errors, respectively.

Since its inception in the nineteenth century, the practice of statistical inference has grown to become an almost essential part of any empirical research. In human geography, most experiments are not controlled, as in the planting of two types of seed, but natural, in the sense that the conditions of the experiment are outside the researcher's direct control. Thus, the two types of seed might correspond to two study areas, perhaps two cities or two ethnic groups. Under such circumstances, the assumption that each sample is randomly and independently chosen tends to be untenable, particularly if samples are taken from locations close together in space.

Moreover, statistical tests in human geography are often plagued with the problems associated with drawing multiple inferences from the same sample. Consider a pattern of points denoting instances of a disease, and suppose that the researcher wishes to determine if clusters occur – if areas that appear to have higher density do indeed have anomalous properties. It is possible to test any area against a null hypothesis of uniform density, but how should the researcher decide which areas to test? Inevitably, the choices will be determined by the distribution of the very phenomenon one is proposing to test – in other words, the null hypothesis is untenable *a priori.*

Reference has already been made to the fundamental complexity of human behavior, and the impossibility of perfect prediction. In the world of human geography, one suspects that virtually any effect is detectable given enough data – in statistical terms, that with enough data one can almost always refute a null hypothesis, and that failure to reject is almost always a type II statistical error. Moreover, the acceptance or rejection of a null hypothesis confounds the strength of any effect with the size of the sample. Nevertheless, establishing the significance of an effect has become a primary goal in much of the quantitative literature. Peter Gould's 1970 paper 'Is *statistix inferens* the geographical name for a wild goose?' is a compelling review of these arguments for and against statistical inference.

Statistical Models

Another thread of the quantitative revolution concerned the ability to infer generalities from pattern, through the analysis of patterns of features on the Earth's surface. Certain general principles, it was argued, would leave characteristic footprints, such as the hexagonal network of settlements of central place theory, and analysis of pattern could therefore lead to inference about theory. But given the complexity of the human world, any such patterns could be expected to show erratic distortions, which might be modeled using statistical principles. Accordingly, much energy was expended in the 1960s in efforts to fit statistical models to the distributions of settlements, in order to show that while perfect hexagons did not exist, some degree of hexagonality or other kind of order could still be detected. Geographers were able to show that the numbers of settlements in each cell of a grid laid over Iowa were indicative of a pattern that was not random, but tended toward a uniform spacing of settlements; and similar methods were used to detect clustering in patterns of disease, ethnicity, and many other phenomena.

In such cases, however, the set of statistical models against which real patterns can be compared is extremely limited. More recently, geographers have become interested in various kinds of simulation models that exploit the power of computers to represent complex behaviors and interactions, and to compute the patterns that they produce on the geographic landscape. In effect, such simulation models replace the simplistic null hypotheses of traditional statistical analysis with more realistic hypotheses, and success results when a simulation matches observation and the hypotheses are accepted, in contrast to the double negative statistical tradition of rejecting a null.

Geographers are particularly interested in the kinds of patterns that result when values are assigned to reporting zones such as counties or census tracts. For example, such patterns arise in the distribution of income or population density, and several hundred statistics are produced for every reporting zone in the aftermath of every decennial census. The characteristic known as positive spatial dependence – a tendency for nearby areas to have similar values – is almost universally observed in such data, and has led to the assertion known as Tobler's first law. Specialized statistical tests for the presence of spatial dependence have been developed, and are widely used by geographers, based on rejection of a null hypothesis of independence.

Curve Fitting

Consider a quantitative relationship of the form $y = f(x)$ where y is some dependent variable, f is a function, and, x is one or more independent variables. Such a relationship might describe the effects of income x on economic deprivation y (in this case f would be expected to be inverse, in other words an increase in x would result in a decrease in y). In another example y might represent average income in a county, and x years of education. In a classic study of this nature, Openshaw and Taylor examined the relationship between percent Republican voters and percent 65 and over in the 99 counties of Iowa. The model $y = f(x)$ might include spatial or temporal lags, if it is hypothesized that y depends on levels of education in previous time periods, or in adjacent counties.

The simplest possible relationship between two variables is linear, and in the absence of other knowledge or hypotheses a researcher might evaluate the relationship $y = a + bx$, where a and b are constants. Linear regression is a time honored technique for determining a and b from real data, in other words of fitting the curve f. It proceeds by a process known as least squares, by choosing a and b to minimize the sum of squared deviations between observed values of y and their values predicted from the model. Linear regression has spawned a host of related techniques for handling more complex models, more specific concepts regarding deviations from the model, and more complex hierarchical structures. Of particular interest to human geographers are versions of the model that incorporate spatial lags.

Figure 1 shows the relationship between median value of housing and percent black by county from the 1990 US census. **Figure 1a** shows the counties of California, and indicates a clear quantitative trend, counties with higher percent black having higher median value. **Figure 1b** shows the counties of Alabama, and indicates the opposite trend. In California the counties with the highest percent black are urban and comparatively wealthy, whereas in Alabama the counties with the highest percent black are rural and comparatively poor.

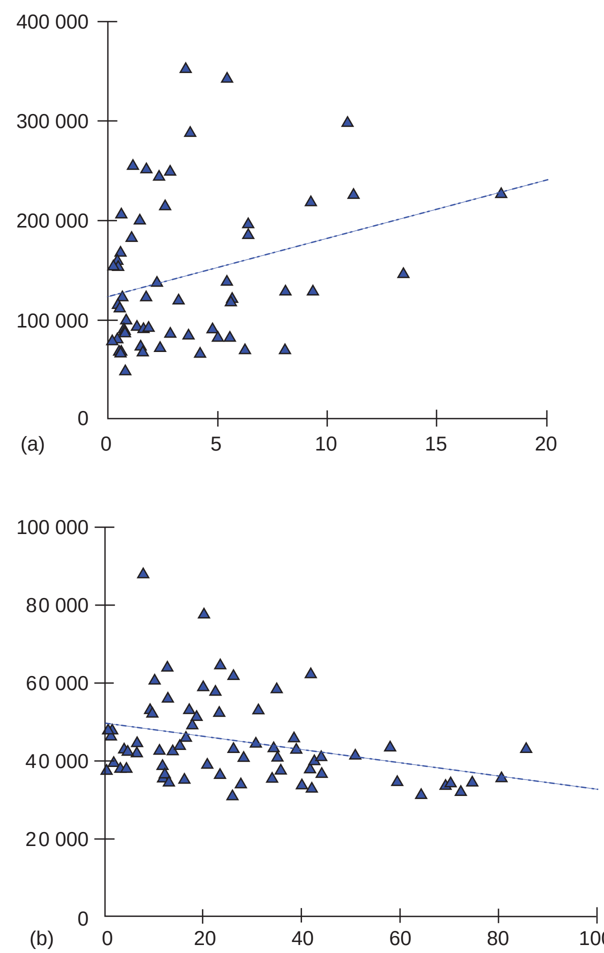

Figure 1 Scatterplots of the counties of (a) California and (b) Alabama, showing quantitative relationships between median value of housing and percent black. The dashed lines represent fitted linear trends.

Curve fitting is invaluable in prediction, in making estimates of the value of some outcome y based on scenarios involving one or more inputs x. One of the more successful applications in human geography has been to the spatial interaction model, which attempts to predict interactions or flows between an origin and a destination based on properties of both and of the trip or separation between them. The classic spatial interation model has the form $I_{ij} = E_i A_j f(D_{ij})$, where I_{ij} denotes the interaction between origin area i and destination area j, E_i is a factor characterizing the origin area's propensity to generate interaction (its emissivity), A_j is a factor char acterizing the destination area's propensity to attract interaction (its attractiveness), and d_{ij} is a measure of the separation of i and j. The function f will be a decreasing function of distance or separation corresponding to the impedance associated with geographic separation.

The spatial interaction model has been applied to a host of phenomena from journeys to work to migration and social interaction. Data from these areas have been used to calibrate the model, in other words, to determine the values of various unknowns, such as the form of the impedance function f, by comparing the model's pre dictions to real data. Various elaborations of the model have been described, and there are several theoretical justifications for its functional form. It is routinely used to predict shopping behavior under scenarios involving new development, closure, and modifications to the transportation network.

Number Crunching

One of the most fruitful areas for development of quantitative methodologies in recent years has been in the area somewhat pejoratively described as number crunching – the mining of large amounts of data in search of pattern and ultimately hypotheses, in a decidedly in ductive spirit that is largely independent of theory. One of the strongest proponents of this paradigm in human geography is Openshaw, whose series of geographical analysis machines are designed to submit data to large numbers of exploratory hypotheses, many of which may not make any immediate sense in any recognized theoretical framework.

An early version of this approach was popular in the 1970s, particularly at the University of Chicago, a tradi tional center of quantitative social science. Factor analysis was devised in the 1930s as a tool for examining large matrices of data in a search for fundamental but hidden dimensions. For example, one might submit the results of a large number of psychological tests to such an analysis in an attempt to identify what one might claim to be the underlying dimensions of personality. A similar approach was adopted in examining large amounts of census data in an effort to discover the underlying dimensions of geographic variation in human society. Studies were conducted on many cities, and two dimensions consist ently emerged: a composite of various indicators of wealth, and another of indicators of life cycle stage. Critics pointed to the unknown effects of the variables chosen by the census for tabulation, the unknown effects of the reporting zone boundaries, and the arbitrarily linear nature of the analysis.

With the phenomenal growth of computing power and data availability over the past two decades, such inductive methods have become increasingly popular. Neural networks began as an effort to provide a crude model of how the brain might operate, but have been adopted as theory neutral tools for the analysis of large datasets, with some success in the general area of prediction. Self organizing maps are another product of research in ar tificial intelligence that have appealed to geographers as methods for discovering pattern, and perhaps hypotheses, in large data sets. The term data mining has been popu larized in this context.

Optimization

The quantitative revolution's interest in central place theory stemmed largely from its potential as an explanation of settlement patterns – of why settlements appeared on agricultural landscapes in the observed locations and sizes, and offering particular combinations of goods. From time to time, the same theory has been used for a quite different purpose, as a basis for planning new landscapes, when decisions on locations, sizes, and perhaps offerings of goods are in the hands of planners. For example, planners were required to make decisions about the locations of settlements during the draining of the Dutch polders in the mid twentieth century.

Similar concern for design has underlain many other applications of quantitative methods in geography over the past half century. Geographers have contributed to the literature on the optimal location of linear facilities such as highways, pipelines, and power lines; point fa cilities such as schools, fire stations, and retail stores; and area facilities such as nature preserves and voting dis tricts. Many of the methods fall under the general heading of operations research, a subdiscipline that is also found in transportation, industrial engineering, and management science.

The use of optimization for design purposes has been termed normative geography, to distinguish it from the more traditional interaction between theory and experi ment that one might term positive geography (in line with the prevailing positivist philosophy of science). But the boundary between these two paradigms is often blurred. Consider, for example, an application of central place theory to design. While the planner might deter mine the locations and sizes of settlements, the success or failure of an offering of a particular good at a particular settlement will be determined in the real world, as will the choices consumers make about which settlements to visit and how much to spend. But successful planning clearly requires that these aspects be predictable, even though they are not controlled. Thus, positive geography must be invoked to provide the predictions, in order to ensure the success of the normative plan.

Data Modeling and Software

The early quantitative geographers worked with pen and paper, aided from time to time by tables of logarithms and slide rules. But the advent of powerful computing machines, which became available on university cam puses in the early 1960s, undoubtedly added momentum, and today computing and quantitative methods are in separable. Gone are the days when instructors could insist that students perform tests by hand, to ensure that they understood the mechanics, before being unleashed on computing environments. Today, statistical software is universally available, and geographers are aided by spe cialized applications such as geographic information systems (GISs), software for exploratory spatial data an alysis, and web based services such as Google Earth.

Within the world of statistical computing, the dom inant data model is the table. Cases, observations, or samples are arrayed in the rows, and the columns hold the various measures and counts associated with each row. Tables are the primary format of data distribution through institutions such as the US's Interuniversity Consortium for Political and Social Research (ICPSR) or the UK's Essex Data Archive, and are the primary data model for statistical software.

Although tables are a powerful mode of represen tation, they are unsatisfactory in several respects for re search in human geography. If the data refer to points, lines, or areas, then it is likely that the locations and geometric forms of the individual cases will be captured, as an aid both to cartographic display and to analysis. While the locations of points can be represented as pairs of coordinates, lines and areas require more complicated structures that do not fit neatly into tables. Handling such complex geographic features has become the domain of GIS, and a common format known as the shapefile has become the *de facto* standard, despite its origins in the products of a single software vendor.

See also: Exploratory Spatial Data Analysis; Hypothesis Testing; Scientific Method; Spatial Autocorrelation; Spatial Interaction Models; Statistics, Overview.

Further Reading

Bunge, W. (1966). *Theoretical Geography* (2nd edn). Lund: Gleerup.
Christaller, W. (1966). *Central Places in Southern Germany*, Baskin, C. W. (trans.). Englewood Cliffs, NJ: Prentice Hall.
Clark, W. A. V. and Hosking, P. L. (1986). *Statistical Methods for Geographers*. New York: Wiley.
Fotheringham, A. S., Brunsdon, C. and Charlton, M. (2000). *Quantitative Geography: Perspectives on Spatial Data Analysis*. Thousand Oaks, CA: Sage.
Ghosh, A. and Rushton, G. (eds.) (1987). *Spatial Analysis and Location Allocation Models*. New York: Van Nostrand Reinhold.
Gould, P. R. (1970). Is *statistix inferens* the geographical name for a wild goose? *Economic Geography* 46, 439 448.
Griffith, D. A. and Amrhein, C. G. (1991). *Statistical Analysis for Geographers*. Englewood Cliffs, NJ: Prentice Hall.
Haining, R. P. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University Press.
Harvey, D. (1970). *Explanation in Geography*. New York: St. Martin's Press.
Losch, A. (1954). *The Economics of Location*, and Woglom, W. H. and Stopler, W. F. (trans.). New Haven, CT: Yale University Press.
Maguire, D. J., Batty, M. and Goodchild, M. F. (2005). *GIS, Spatial Analysis, and Modeling*. Redlands, CA: ESRI Press.
Miller, H. J. and Han, J. (eds.) (2001). *Geographic Data Mining and Knowledge Discovery*. New York: Taylor and Francis.
O'Sullivan, D. and Unwin, D. J. (2002). *Geographic Information Analysis*. New York: Wiley.

Quantitative Revolution

T. J. Barnes, University of British Columbia, Vancouver, BC, Canada

Glossary

Areal Differentiation It is a focus on spatial differences across human and physical phenomena on the Earth's surface, and especially as found among different regions.

Correlation and Regression It is a set of formal mathematical techniques developed in statistics to assess the existence and strength of a relationship among different variables.

Exceptionalism It is a term coined by Fred Schaefer to denote Richard Hartshorne's view that geography was different from other disciplines, methodologically unique.

Idiographic Concepts and approaches directed toward the elucidation of the unique and the particular.

Neoclassical Economics The orthodox school of economics defined by: the use of the rationality postulate, a belief in market equilibrium and optimality, and the deployment of formal theory and rigorous techniques of empirical testing.

Nomothetic It is the study and discovery of general traits within phenomena with the ultimate end of making universal law-like statements.

Scientific Laws These are statements of an invariant universal relationship holding between either different phenomena or different states of the same phenomena.

Spatial Interaction It defines the relationship of one place to another measured by the flows of phenomena (e.g., people, goods, and ideas) between them.

Theory It is defined as a set of abstract, logically connected statements for the purpose of explanation.

Introduction

The quantitative revolution represented a profound transformation in Anglo American human geography from the mid 1950s and defined by the systematic application of scientific forms of theorizing and rigorous statistical techniques of analysis and description. In the process, the previous idiographic regional geography concerned with describing, cataloging, and delineating areal differentiation was pushed aside, replaced by a nomothetic 'new geography' directed toward explaining, scientifically proving, and abstractly theorizing spatial phenomena and relations. Geography was no longer rote memorization of place names, mountains chains, and population sizes, but was a science – spatial science.

Antecedents: Regional Geography, World War II, and the Cold War

Anglo American geography before World War II was dominated by a regional approach, fixated on inventorying the peculiar characteristics of particular places, and carried out using one typological scheme or another. Put the St. Lawrence under rivers; the Rockies under mountains; Montreal and Toronto under cities; and barley, wheat, and corn under agricultural produce. By comparing the contents of the typological grid for different regions, the uniqueness of each place was manifest.

Such practices were given a splash of intellectual respectability by the American geographer Richard Hartshorne (1899–1992) when in 1939 he wrote the most systematic and philosophically sophisticated monograph yet published in English on the history and definition of the discipline, *The Nature of Geography*. Drawing especially on German writers, he argued that the building blocks of geographical knowledge were regions, defined as complex combinations of hard facts and specific causal relations, both of which were capable of objective disclosure. In turn, this combinatorial character ensured that each region was unique, not found anywhere else. Consequently, traditional scientific forms of explanation based on general laws did not apply. Natural sciences explain by asserting general (generic) relationships between homogeneous classes of phenomena: if class of phenomena A, then class of phenomena B. But Hartshorne postulated that geographical regions were never homogeneous; each region was unique. Traditional scientific explanation in which instances of broader classes of phenomena were related by a law like statement had no purchase. Regions are not categorizable as homogeneous classes. They could be described, but neither scientifically explained nor predicted.

Just before *The Nature* was published, Europe was pitched into war, and 2 years later, America too was embroiled. Hartshorne was recruited as an administrator into what became the Office of Strategic Services (OSS), forerunner of the Central Intelligence Agency (CIA). The OSS, and especially its Research and Analysis Branch, was charged with collecting and analyzing information that bore on national security. To do so, a large number of American academics as well as foreign *émigré* scholars were recruited, and made to engage one another across their disciplinary affiliations. The result was a new kind of social scientific inquiry that emphasized science much more than the social. It was team based, mixing disciplines, instrumental, model and even machine driven, quantitative, and drawing upon abstract theory. Further, following science, it emphasized explanation rather than description, generality rather than uniqueness. Of course, this kind of pursuit was antithetical to Hartshorne's view of geography. Insofar as he could, he prevented geographers employed at OSS from undertaking this form of investigation. But the genie was out of the bottle.

The new approach fomented within OSS influenced a range of social sciences once World War II was over. Economics and psychology became paradigm examples and, to a lesser degree, sociology and political science, and even some humanities such as philosophy. Human geography became just another (albeit late) addition to the list.

Giving momentum to the change was also the Cold War. Believing that science and the kind of social science pioneered at OSS had won World War II, the US government through military organizations such as the Office of Naval Research (ONR) poured inordinate amounts of money into science and social science to win the Cold War. The result was yet further impetus to what Carl Schorske, a former OSS alumnus, called the 'new rigorism' within post war US human sciences. As a force, it was difficult to resist. Certainly, a number of younger US geographers, some of whom had served at OSS, such as Edward Ackerman, Chauncey Harris, and Edward Ullman, could not resist. Once demobilized and back in the academy, they began hesitantly to practice a different kind of geography: a scientific kind. It was one in which regions were still important, but now conceived quite differently to Hartshorne, as explanatory, theoretical, and instrumental, a tool to achieve functional objectives. Geography's quantitative revolution was about to begin.

Centers of Geographical Calculation

It took almost a decade to realize, however. It was only in the mid 1950s that the first real pangs of conversion from a dull, older regional geography to a shining new spatial science were registered in the discipline as a whole. At first, the change was highly localized, confined initially in the United States to two centers: the University of Washington, Seattle, and the University of Iowa, Iowa City.

Key at the University of Washington were two faculty members: Edward Ullman, formerly in the OSS; and William Garrison, who, during the war, served on Army Air Force bombers in the Pacific Theater and after his demobilization, he completed a PhD in geography at Northwestern University. In the Air Force, Garrison was trained in statistics and mathematical methods. None of that knowledge was useful, though, once he entered Northwestern as a graduate student, where he was mired within the Hartshorne tradition. He later said in an autobiographical essay that his resulting regional geography based PhD dissertation involved "just a lot of walking around, and mere description." In contrast, he was interested in knowing "what's the theory, what's the theory, what's the theory" (personal communication).

He began to answer that question once he was hired at the University of Washington in 1950, assembling around him others interested in the same question. *Annus mirablis* was 1955, when in the autumn of that year a remarkable group of graduate students serendipitously arrived under the supervision of Ullman and Garrison. Later labeled the 'space cadets', they were the revolutionaries of geography's quantitative revolution. In the first term, they were exposed to the first advanced course in statistics ever given in a US geography department, Geography 426, quantitative methods in geography and offered by Garrison. Richard Morrill, in that first class, says "it wasn't just the introduction to beginning statistics but the whole gamut from beginning to all that was known in those days. So, it was a ferocious baptism" (personal communication). But it was not only numbers to which they were exposed, but also to machines. There were the large, thudding Frieden calculators, but even more important was the recently acquired, even larger, computer. In an early advertisement for the Washington department, the Head, Donald Hudson, boasted about the departmental use of an IBM 604 digital computer, a national first. The programming technique of the so called patch wiring involving plugging wires into a circuit board was crude and inefficient, but it helped define and consolidate the scientific vision of the discipline.

Then there was theory. Without any indigenous tradition, theories were frequently begged, borrowed, and stolen from other fields. Ullman's seminar on urban location theory, also given in 1955–56, provided the basics of Walter Christaller's and August Lösch's central place theory. Garrison offered the cadets a seminar in economic theory using the regional scientist Walter Isard's just published *Location and Space Economy.* Regional science was a kindred movement to the quantitative revolution, and early on at least there was a symbiotic relationship. Always more economic than economic geography, regional science was steeped in neoclassical economics. Isard's book defined by differential equations, simplified abstract assumptions, and hardheaded rigor and logic left Garrison's students with no doubts about "what's the theory."

At the other center, the University of Iowa, the work of Harold McCarty was important. As at the University of Washington, McCarty funded his students from grants provided by ONR. He used the money to carry out studies of industrial linkage and represented formally using inferential statistics. In 1953, McCarty was the first human geographer ever to use correlation and regression techniques.

A cantankerous colleague of McCarty's at Iowa, Fred Schaefer, was a second Iowa contributor. Schaefer provided the larger philosophical rationale for the revolution: positivism. A political refugee from Nazi Germany, Schaefer published, in 1953, in the flagship journal of American geography a pointed philosophical critique of Hartshorne's idiographic justification of regional geography that he labeled 'exceptionalism'. Schaefer's alternative was positivism – a philosophy based on the canons of natural scientific practice emphasizing explanation, logical deduction, stringent empirical testing, and the virtues of a formal, universal mathematical vocabulary. It was everything that Hartshorne's regional geography was not. Schaefer's positivism gave those not only at the University of Iowa, but also at other later centers of geographical calculation, an intellectual rationale grounded in a body of seemingly unassailable analytical philosophical writings.

Outside of this North American frame, two other centers require mention. The first was the department of geography at Cambridge University. In autumn 1958, Richard Chorley and Peter Haggett (with David Harvey as a teaching assistant) began lecturing first year undergraduates – for the first time in the history of the department – on statistical methods, matrices, set theory, trend surface analysis, and network analysis. Chorley and Haggett were later dubbed 'the terrible twins' of British geography for their pioneering work that later changed fundamentally the practices of both economic geography (Haggett's specialty) and geomorphology (Chorley's field). Further, that Chorley and Haggett came from different halves of the discipline, but could speak the same language demonstrated the centrality of scientific reasoning for potentially uniting physical and human geography – a long held dream, but hitherto never realized.

The second is Lund, Sweden, and associated with the iconoclast, Torsten Hägerstrand. Virtually on his own back, Hägerstrand developed and deployed during the early 1950s a set of statistical and theoretical techniques to understand innovation diffusion across the Swedish space economy. It was sophisticated and original, resulting in Hägerstrand visiting the University of Washington in the late 1950s, influencing several of the cadets.

By 1960, all the elements necessary for the quantitative revolution in human geography were assembled: statistical and mathematical techniques, machines, theory, and an underlying philosophy. They needed to be put together, though, and diffused to a wider audience. Exactly this process unfolded during the ensuing decade. Partly, the diffusion occurred through the bodies of the revolutionaries themselves, as, for example, the space cadets left Seattle for new jobs throughout America. As they moved, they spread the word of revolution to new sites such as the University of Chicago, University of Michigan, Northwestern University, and later, Ohio State, SUNY Buffalo, Toronto, and Bristol. Partly, the diffusion occurred through seminars, special sessions at conferences, training camps in quantitative methods for the uninitiated (the first was held at Northwestern in 1961), and dedicated meetings such as the Madingley Hall lectures that Chorley and Haggett arranged at Cambridge for secondary school geography teachers in England and Wales (the first was held in July 1963, and subsequently published as a series entitled *Frontiers in Geographical Teaching*). Finally, diffusion occurred through the circulation of purple mimeographed papers, various discussion paper series (the first originated at the University of Washington in March 1958), and later formal journals such as *Geographical Analysis* founded in 1969.

That said, not everyone was convinced, not every department of geography went quantitative, and not every subdiscipline turned to numbers. American cultural geography was a hold out, with Carl Sauer, its unofficial leader, writing in 1967, "I am saddened by model builders and system builders and piddlers with formulas for imaginary universals" (quoted in Sauer, 1987: xv). British geography was generally slower on the quantitative uptake than its US counterpart. Urban and social geography, and especially regional geography, did not experience change until the second half of the 1960s, if they experienced change at all. In some cases, resistance was by an old guard who was bent on holding on to the reins of power (which was the case of the University of California, Los Angeles (UCLA) regional geographer, Joe Spencer, who as the editor of the important publication *Annals of the Association American Geographers* had the means to retard the revolution), but in other cases, it was an intellectual and political antipathy against the movement. The University of Minnesota's Fred Lukermann, for example, was a critic as early as the late 1950s, and the University of Minnesota where he was chair and later dean never became a 'center of calculation'.

Geography after the Revolution

For this reason, Ian Burton's 1963 announcement of the completion of geography's quantitative revolution, precipitating within the discipline "a radical transformation in spirit and purpose" (Burton, 1963: 151–162), was premature and exaggerated. But something was changing, and throughout the 1960s, increasing numbers of geographers and geography departments were pulled in. Five characteristics of the revolution are particularly germane.

First, the discipline was characterized by the use of a particular form of theory, one found in the physical sciences, especially physics, and later mimicked by economics. For this reason, the quantitative revolution was perhaps a misnomer. What excited revolutionaries was less numbers *per se* (although enthused by some), but abstract theoretical explanations couched as scientific hypotheses, mathematical models, and, the pinnacle, laws. This form of theory derived from the philosophy of science, and in particular positivism. Here, theories were conceived as formal statements, expressed in the vocabulary of mathematics, positing causal connections among classes of phenomena and capable of empirical verification.

The problem for quantitative revolutionaries, though, was the historical absence of theory of this type within geography, indeed, of any theory. Apart from a discredited dalliance with the theory of environmental determinism, human geography previously prided itself on its atheoreticism. Theoretical inspiration, therefore, necessarily needed to come from outside. There were several sources. A recouped tradition of German location theory that included Johann von Thünen's (1783–1850) formulations of agricultural land use and rent, Alfred Weber's (1869–1958) model of industrial location, and August Lösch's (1906–45) general theory of the space economy provided an initial set of core concepts. Added to these from orthodox neoclassical economics, and regional science, were rational choice theory as well as general and partial models of market equilibrium. Later from sociology came urban land use models, urban factorial ecology, and the rank size rule. Geometry provided network and graph theory as well as the mathematics of topological forms and used to explain spatial patterns such as transportation routes and trip patterns. Finally, physics offered gravity, potential, and later entropy models that, in turn, were used to theorize spatial interaction. These last models likely represented human geography's finest theoretical hour. Thousands of papers were published on spatial interaction, producing (at least for some) the ultimate: a geographical law. Waldo Tobler (1969) formulated the 'first law of geography': "everything is related to everything else, but near things are more related than distant things."

While initially theory would be taken from others, the eventual aim was to construct home grown versions. In what became the methodological and philosophical bible of the quantitative revolution, *Explanation in Geography*, David Harvey (1969) famously finished with the rallying cry, "By our theories you shall know us."

Second, there was the quantitative part of the quantitative revolution. Quantitative did not mean simply numbers. Geographers were numerate from the beginning. Between 1852 and 1871, the American Geographical (AGS) and Statistical societies were officially twinned, and even when they went their separate ways, the mandate of the AGS remained the collection, classification, and scientific arrangement of statistics and their results. 'Quantitative' for the quantitative revolutionaries, rather, meant formal statistical techniques and used to represent numerical data, and to draw scientific inferences. The first forays were in descriptive statistics, but inferential statistics, that is, drawing conclusions about larger populations from samples, quickly followed, at least in the United States.

It was a slightly different story in Britain, where the quantitative revolution was slower in disseminating. In 1964, at the Institute of British Geographers' annual meeting, Peter Haggett showed a multiple regression equation. The next day he was summoned by his head of department at Cambridge, Alfred Steers, who was in the audience. "This kind of thing has got to stop," Steers told him, "You are bringing the subject into disrepute." Of course, it did not stop, nor could it stop. In fact, 2 years later, Haggett was promoted to professor of urban and regional geography at Bristol University, and in part because of his multiple regression equation. More generally, Haggett, along with other British geographers, throughout the second half of the 1960s and early 1970s developed and refined a new body of statistical methods designed to accommodate the peculiarities of geographical data, and which included novel spatial sampling methods, spatial autoregression techniques, and work on the areal modifiable unit problem.

Third, to cope with the numbers, and to calculate statistical formulae, required machines. Initially, 'computing rooms' meant places in which were found banks of hand calculators, either of the mechanical kind like the Monroe, or the electric kind like the Frieden. But as computations became more complex and fraught because of the use of multivariate statistical techniques, and increasingly large datasets, even the trusty Monroe and Frieden balked. Something bigger and better was needed. Fortunately, the military–industrial–academic complex came to the rescue. The first computer in the US, the Electronic Numerical Integrator and Computer (ENIAC), was first used to make calculations for testing ordnance at the Aberdeen weapons site, Maryland. Within 10 years, it carried out more calculations than had been completed during the whole of previous human history. This was exactly the kind of machine that the quantitative revolution needed. In fact, it got something even better. By the mid 1950s when ENIAC was decommissioned, computer development was already transformed (in large part because of the Cold War imperatives and money). In 1954, IBM, which was contracted by the military as its computer manufacturer, began selling computers commercially to universities, the first going to Columbia. Using the computer meant bootstrap learning, though. There was no online help and not even a language for programming. Instructions were sent by plugging and unplugging electrical circuits.

Waldo Tobler, one of the Washington space cadets, remembers his early computing experiences in Seattle:

> We had to go up to the attic of the Chemistry building at 2 am so we could run the computer by ourselves. They didn't have any computer operators in those days, and that was before computer languages like FORTRAN. … To cover programming on the [IBM] 650 you had to pick up two bytes of information on one rotation of the drum. It had a 2K memory which rotated real fast. And if you were clever, you could pick up two pieces of information in one rotation. (personal communication)

By today's standards, the IBM 650 on which Tobler and the other space cadets worked during the early hours was a lumbering dinosaur, but even then it could perform calculations with a speed, consistency, and stamina that no human could match, and vital to the success of the quantitative revolution.

Fourth, the quantitative revolutionaries appealed to a larger philosophical project – positivism – to justify and legitimate. Positivism as a philosophy existed since Auguste Comte's formulations in the first part of the nineteenth century. During the twentieth century, it was taken up and reworked as logical positivism by a group of philosophers, mathematicians, and physical scientists in Vienna (the Vienna Circle). They argued there were only two kinds of true, meaningful (scientific) statements, each of which could be precisely defined and delineated. First, analytic statements were true by definition, and the basis of the formal sciences, logic, and mathematics. Second, synthetic statements were empirically verifiable, that is, statements that could be unambiguously proven true or false by comparing them to real world observations. They were the basis of the substantive sciences.

Positivism as a philosophy was made for post war American social science, emphasizing the importance of a mathematical vocabulary in which to express theory and the importance of quantitative empirical verification. Spatial scientists were certainly attracted. Fred Schaefer's paper was the opening shot. A colleague of Schaefer's in the philosophy department at Iowa, Gustav Bergmann, and with whom he discussed his paper, was a former member of the Vienna Circle. Further, it was Bergmann who finalized Schaefer's paper for publication after Schaefer died from a heart attack in an Iowa City cinema. It was also Bergmann who taught a philosophy of science course that all incoming geography graduate students at Iowa were compelled to attend (Douglas Amedeo and Reginald Golledge's 1975 textbook *An Introduction to Scientific Reasoning in Geography* was written as a result of them taking that seminar, and in this sense, it is the heir to Schaefer's 1953 paper). If Schaefer's paper was the opening shot, then David Harvey's *Explanation in Geography* was the closing one. Much more than Schaefer ever did, Harvey provided systematic accounts of both the analytic statements geographers could make in mathematics, logic, and especially geometry, as well as their synthetic statements, particularly methods of empirical testing using statistical techniques.

Finally, there was a repositioning within the discipline's social hierarchy, albeit not without contestation. Young, male, very ambitious, very able graduate students and junior faculty primarily forged the quantitative revolution, and as it succeeded, they became the new top dogs in geography. In some cases, their advance was rapid. Brian Berry completed both his MA and PhD at Washington in only 3 years, and by fall, 1958, he was an assistant professor at the University of Chicago. Within 7 years, he was full professor (at the age of 31). Again, Haggett was appointed at Cambridge in 1957 as a demonstrator (i.e., below an assistant lecturer). Less than a decade later, he was professor of urban and regional geography at Bristol. Peter Taylor has argued that the quantitative revolution was propelled, in part by internal sociological reasons, to secure precisely a rapid rise in career status.

Perhaps the most marked social characteristic of the revolutionaries was their masculinism. All of the early quantitative geographers were male. Moreover, they often acted typically male, tending to be competitive, pushy, playing practical jokes, and being sometimes boastful and arrogant. Michael Dacey, one of the believers at Washington, says "we were very aggressive, very ambitious, and very appreciative. … We were full of missionary fervor, and I imagine we were unlikable brats. … In retrospect we must have been very disorientating to the establishment" (personal communication). Some of these characteristics come out in their writing, which could be occasionally brash and combative. It also made it difficult for women to participate. Susan Hanson, when she enrolled at Northwestern University in 1967, says, "We knew very well that we were entering male turf." Ten years later, not much seemed to have changed. Pat Burnett (personal communication) who was a faculty member at Northwestern in the late 1970s sued the department for its 'climate of sex discrimination'.

The broader point was that the quantitative revolution did not happen at arm's length, as if theory, numbers, machines, and a positivist philosophy entered the discipline autonomously on their own volition. But they were connected to a set of social processes and bodily inscriptions. It was never spatial science in the abstract, but spatial science incarnate.

Revolting against the Revolution

The revolution could not last. It was out of synch with the discipline's own intellectual past that emphasized grounded context and geographical singularity; it was out of synch with its own historical and political moment of especially the late 1960s that began to react against the violence of abstraction, technology, formal ordering, and centralized authority; and it was even out of synch with scientific logic as assorted logical contradictions, in consistencies, and *aporias* within the work of the quanti tative revolutionaries revealed.

Harvey was a central figure in the subsequent un raveling. Even before he finished *Explanation*, he had doubts. In 1971, at the annual meeting of the Association of American Geographers in Boston, those doubts erupted. He announced there that:

> [Geography's] quantitative revolution has run its course and diminishing marginal returns are apparently setting in as … [it] serve[s] to tell us less and less about anything of great relevance…. There is a clear disparity between the sophisticated theoretical and methodological frame work which we are using and our ability to say anything really meaningful about events as they unfold around us. … In short, our paradigm is not coping well. (Published as Revolutionary and counter revolutionary theory in geography and the problem of ghetto formation, *Antipode* 4, page 6, 1972.) (Harvey, 1972: 6)

The rest of the 1970s was a decade in which various elements of the quantitative revolution were in turn held up for scrutiny and found wanting. Harvey began by at tacking the usefulness of the theory and statistical tech niques, portraying them as at best irrelevant and worst politically regressive ('counter revolutionary'). Gunnar Olsson, disgruntled with the spatial interaction models he earlier triumphed, argued that their very formal rea soning undid any claim to empirical veracity, and when they were applied they made bad worlds and not good ones, worsening the lot of humans and not improving it. In similar vein, Robert Sack argued that the very idea of a separate science of space was logically inconsistent with the scientific principles invoked. Finally, special disdain was directed toward positivism, thus becoming the hap less sparring partner for a series of postpositivist ap proaches, including Marxism, humanistic geography, critical realism, and critical theory.

In spite of this critique, there have remained some striking continuities. Two stand out. First, the quantita tive revolution never really died, but morphed into new forms, especially in the practice and discussions around geographic information system (GIS) from the 1980s. Indeed, two of the original Seattle space cadets, Art Getis and Duane Marble, were pioneers in the later GIS revolution. Further, some of the debates in the quanti tative revolution found echoes in discussions in GIS around visualization, datasets, values, and political rele vance. The second is the persistence of a theoretical vocabulary and which was introduced by the quantitative revolution. Although rarely recognized, the upholding of a theoretical discourse represents a watershed, separating the last 50 years of the discipline's history from the 50 years before that. The meaning of theory has altered, but the constancy of a theoretical vocabulary has proven more important in subsequently shaping the discipline than any rupture. We are more the product of some defunct theoretical spatial scientist than perhaps we realize.

See also: Cold War; Masculinities; Quantitative Methodologies; Scientific Method; Spatial Science.

Further Reading

Barnes, T. J. (2004). Placing ideas: Genius loci, heterotopia, and geography's quantitative revolution. *Progress in Human Geography* 28, 565 595.

Barnes, T. J. (2004). The rise (and decline) of American regional science: Lessons for the new economic geography? *Journal of Economic Geography* 4, 107 129.

Barnes, T. J. and Farish, M. (2006). Between regions: Science, militarism, and American geography from World War to Cold War. *Annals of the Association of American Geographers* 96, 807 826.

Billinge, M., Gregory, D. and Martin, R. (eds.) (1984). *Recollections of a Revolution: Geography as Spatial Science*. London: Macmillan.

Burton, I. (1963). The quantitative revolution and theoretical geography. *The Canadian Geographer* 7, 151 162.

Gould, P. (1978). The Augean period. *Annals of the Association of American Geographers* 69, 139 151.

Gould, P. and Pitts, F. R. (eds.) (2002). *Geographical Voices: Fourteen Autobiographical Essays*. Syracuse, NY: Syracuse University Press.

Haggett, P. (1965). *Locational Analysis in Human Geography*. London: Edward Arnold.

Harvey, D. (1969). *Explanation in Human Geography*. London: Edward Arnold.

Johnston, R. J. (2003). Order in space: Geography a discipline in distance. In Johnston, R. & Williams, M. (eds.) *A Century of British Geography*, pp 303 345. Oxford: Oxford University Press.

Mirowski, P. (2002). *Machine Dreams: Economics Becomes a Cyborg Science*. Cambridge: Cambridge University Press.

Olsson, G. (1975). *Geographical publications, Number 15: Birds in Egg*. Ann Arbor, MI: Department of Geography, University of Michigan.

Schaefer, F. K. (1953). Exceptionalism in geography: A methodological introduction. *Annals of the Association of American Geographers* 43, 226 249.

Schorske, C. E. (1997). The new rigorism in the human sciences, 1940 60. *Daedalus* 126, 289 309.

Taylor, P. J. (1976). An interpretation of the quantification debate in British geography. *Transactions of the Institute of British Geographers NS* 1, 129 142.

Queer Theory/Queer Geographies

K. Browne, University of Brighton, Brighton, UK

Glossary

Heteronormativity Is the normalization of forms and practices of heterosexuality within particular structures of relationships, gender, ethnicity, and ability. The practices and ideologies that normalize heterosexuality in these ways render other forms of sexuality (as well as other ways of doing heterosexuality) deviant and 'other'.

Homonormativity Is the normalization of forms of homosexuality, currently in terms of relationship structure, economic wealth, and sexual practices. These practices and structures are often read as mimicking heterosexual norms, rather than contesting the normativity of, for example, marriage and capitalist accumulation.

Normativity This is an ideal state that is strived for but can never be achieved. It is however often used as a measure of 'normalcy' in comparison to 'deviancy'. It is not static or fixed across time and space.

Queer Theory Is a term that is diversely deployed. It critically examines the creation of normativities and looks to deconstruct these usually, but not solely, by investigating the normative links between gender, sex, sexuality, desire, primary relationships, and sexual practices. Queer theory has critically addressed homonormativity and heteronormativity using thinking that is nonessential, relational, and contingent.

Introduction

'Queer' has a plethora of definitions and is deployed in numerous ways. Queer here will not simply, or only, be taken as synonymous with lesbian and gay. Instead, along with the assertion that 'queer' encompasses sexual and gender 'dissidence', it will be explained as a mode of thinking, a way of knowing specifically a critical stance that seeks to challenge normalizations and explore taken for granted assumptions, particularly (but not exclusively), regarding categories of gender, sex, and sexuality.

Queer can be thought of as post structural, anti essentialist, and relational, yet these terms have historical and contextual meanings and because of this can be decidedly 'anti queer'. Defining queer theory is necessarily difficult as it is elusive and unfixed. It was denied by Teresa De Lauetis, the academic who first coined the term, shortly after she used it in 1990, because it was, in her view, co opted by the mainstream. This is indicative of what queer theory 'is': it can be described as an anti normative impulse that is mobilized to consider the world in antinormative ways, as such to describe queer, can be construed as normalizing it, 'in this way and not that' and for some this is antiqueer. It is with this *proviso* that this article begins and attempts to introduce the reader to queer theory followed by a discussion of the geographies of queer and queer geographies, and finally, the article will outline some of the trouble with queer troublings. Reducing, mainstreaming, and homogenizing queer theories means that this project is in and of itself 'unqueer'.

Queer Theory

There are disciplinary traditions closely associated with the diversity of uses of the term queer. These disciplinary divergences and alternative histories have meant that these are used differently and can offer competing histories as well as theories of 'queer'. Thus to write a genealogy of queer is impossible as queer has many genealogies! It is, for some, possible to separate queer theory (that is, humanities based and read as 'theoretical' and based in 'texts') from queer studies, which can be seen as being more engaged in community and activist research and practices. However, both strands often draw from similar theorists and authors. There remains a contested use of the term where lesbian and gay can be unproblematically used interchangeably with the term 'queer'. This section, will discuss some of the histories of the use of 'queer', offering a glimpse of the reclaiming of the term before working through the academic understandings that underpin many deployments of queer theory.

Queer as 'strange', 'weird', or 'unusual' has altered its meaning significantly throughout the twentieth century. Though there continues to be an abusive message in the jibe 'you queer', there has been a positive reworking of such negative attachments that has sought not to assimilate, but to challenge, contest, and dispute the naturalness of heterosexuality. In reclaiming from negative connotations and abusive uses, in one guise queer came to represent an inclusive identity, one that encompassed lesbians and gay men where those outside of normative heterosexuality (heteronormativity) could claim an incoherent 'queer' identity. In this usage, all who are lesbian, gay, bisexual, trans, intersex, or genderqueer could be 'unified' into this term. The use of the term 'queer' in this all encompassing way has enabled those who seek an easy answer to the proliferating identities associated with gender and sexual diversity. The demand for these dif ferences to be recognized asked for a term that moved beyond the limited identifications of 'lesbian' and 'gay'. It also testifies to the increasing demand for rights and visibilities of bisexual, trans, and gender questioning individuals.

What we now understand as queer theory was de veloped in literary theory and humanities in the US. Teresa De Lauretis was the first to officially coin this term in 1990. From these beginnings, authors such as Judith Butler and Eve Sedgwick have contested the naturalization of binaries of heterosexual/homosexual, de stablizing categories of sexualities and gender, and questioning the normativities of these categories. This contrasts with sexual theorises that view sexualities and gender identities as either socially constructed or innate/ essentialist traits. These uses of 'queer' and conceptual izations of bodies, identities, and desires move beyond gay/lesbian as fixed identities and recognize that these are fluid through time and across spaces.

Judith Butler's thesis regarding the heterosexual matrix, has been highly instrumental in queer theoriza tions of gender and sexualities. She posited that rather than being an underlying structure or an innate drive, heterosexuality is an ideal that can only be approximated. Within this ideal, the subject positions of man/woman emerged as opposites. This has been used to explore how individuals live between man and woman and to question the solidifying of sexed bodies using gendered norms. In questioning heterosexuality, understanding of fixed gen ders as reflecting (or resisting) sexed bodies becomes uncertain and rather than sex leading to gendered acts, gendered acts are read as (re)constituting sexed bodies. In rendering sex fluid the often solid positioning of man/ woman becomes questionable. Butler, and others, have used drag queens and kings to argue that gender is an act that is not necessarily related to genital sex. However, in her later work, and following extensive critiques, Butler is clear to note the distinction between performance and performativity. Performance as a one off act does not necessarily constitute bodies, identities, or spaces, rather these are (re)created through performativity, which is a set of repeated acts that are actively policed and self policed through dominant codes and commonsense norms.

Two things should be noted from this discussion, first 'queer' is not limited to deconstructing sexualities, but fundamentally incorporates, gender, race, and other as pects of the Oedipal subject and Oedipal subjectivities. Second, this use of queer aims not simply to challenge the links between these categories but also to explore and challenge how oppositional binaries such as hetero sexual/homosexual and man/woman are mutually constituted. By contesting and challenging normative femininity/masculinity and sexuality it is possible to understand that heterosexuality itself is performative. Using discussions of performativity gender was/is defined as:

> The stylization of the body, a set of repeated acts within a highly rigid regulatory frame that congeals over time to produce the appearance of substance (Butler, 1990: 33).

Butler contended that the heterosexual matrix created a regulatory frame in which men and women as opposites were meant to come together within a heterosexual frame for the purposes of reproduction. In this way, sexuality and sex are mutually constituted. Where gender is questioned, the stability of sexual identities, such as lesbian, gay, and heterosexual, is also dubious and there is a contestation of the stability of the homosexual/het erosexual binary. Consequently, this matrix is used to understand not only how heterosexuality is normalized but also how this normalization is constructed through particular gendered, racial, able bodied, and other social relations. The term 'heteronormativity' was thus de veloped to understand how the Oedipal subject position is re created. Although the Oedipal subject position is an ideal that can only ever be approximated, it informs the performances of gender and sexualities that create man/ woman as binary opposites and the dichotomy of het erosexuality and homosexuality.

Crucial to the discussions of the heterosexual/ homosexual dichotomy is the 'othering' of homosexuality. Historical theorists have argued that heterosexuality emerged through delimiting homosexuality and thus is dependant on it. Butler contended that homosexuality is the origin of heterosexuality as heterosexuality was originally, and continues to be, constituted through its 'otherness' to homosexuality. This 'otherness' establishes a 'normal', from which the 'deviant' can be discussed. Therefore, definitions of sexualities are fluid varying temporally and, as geographers have illustrated, spatially. Many queer theorists have explored the fluid possibilities of sexualities, questioning the assumed links between sex–gender–sexuality and the object of desire.

In destablizing these normativities, the argument is that normality is not simply constructed, it needs to be continually reproduced and always has the possibility of not being reproduced in the same way. Authors have explored the many ways genders and sexualities fail to be re performed 'correctly'. They have also documented how this failure can result in overt and subtle forms of policing that restores a plethora of possible disruptive moments back to normative outcomes. Exploring the possibilities of immanence and not becoming, as well as becoming, in each moment has offered fruitful avenues of enquiry for other possibilities and other desires. This has also rendered visible the normative frames, commonsense norms, and expectations as well as the policing of sexed and sexualized acts that can recuperate potentially dis ruptive or queer moments, actions, politics, and identities.

Queer theorizations do not simply pertain to sexu alities. Theorists of race, class, and disability have also developed queer theorization that explores the raciali zation, classing, and abling of particular bodies. These theories and theorists have explored how normative bodies are not simply sexualized but also classed, raced, and abled in particular ways. Sara Ahmed, and others have addressed the mutual (re)constitution of gender, sexualized, and racialized bodies, through the con struction of 'others' and 'strangers'. Queer disabled work has looked to the intersections between queer studies and theories of disabilities, as well as addressed the able bodied assumptions inherent to heteronormativity. In addition to rendering certain 'heterosexual' practices, bodies, identities, and lives 'queer', these forms of en gagement can question the assumed antinormativity of particular homosexualities. This has lead to discussions of 'homonormativities' where particular bodies, practices, and identities have been read as normative and therefore, afforded particular privileges such as being able to buy into particular sexual economies that demand tolerance of lesbian and gay identities. This normativity can also be seen as hierarchizing particular relationship forms, that are usually monogamous, long term, and wealthy relationships. For both heteronormativity and homo normativity, it can be argued that it is the privileges that are afforded to those who are 'normative' that are being critically interrogated rather than heterosexual or homosexual practices, lives, or experiences *per se*. This is not to suggest that queer practices are not possible as these can be associated with particular political actions, ways of living, and/or sexual acts that continue to be non normative, such as bondage, dominance, and sado masochism (BDSM).

Similar to heteronormativity then, conceptualizations of homonormativity take many divergent forms. It should be noted that the universalizing tendencies of these theoretical positions are coming into question. There are challenges to the associations of homonormativity where particularly white, middle class men argue that despite their seemingly hegemonic status, this subject position is not uniform or necessarily powerful. As sumptions regarding the status of homonormativity are also problematic given 'queer' histories where individuals still 'bear the scars' of earlier confrontations with nor mativities. Furthermore, questioning the assumptions of heteronormative power, studies have pointed to the moral economies of prostitution and noted that hetero sexuality is not a homogenous category.

The temporalities of normativities, whereby the dominance and margins are not static but continually negotiated and re defined means that the critical stance on 'the dominant' can never be fixed. Consequently, 'queer' is never fully static, it is geographically as well as temporally contingent. (The contextual dependency of this hegemony is something geographers have pointed to and will be returned to below.) In other words, as he gemonies shift to incorporate what was once marginal – the critical stance that is 'queer' has the potential to come to question the normativity of something that was once 'queer'.

Finally queer has also reconfigured the ways we can understand pleasure and desire, questioning the as sumptions that fixed identities (man/woman, hetero sexual/homosexual) link seamlessly with sexual practices and the opening up of new possibilities of desire that do not necessarily equate to sexual intercourse. 'Queer' contestations of normative sexual practices are often experimental, creative, and designed to be fun. With gatherings such as Queeruptions sex parties, non nor mative sexual practices are created by who participates (in terms of how many, how they identify, and even what counts as 'participation'), as well as, what (acts, props and toys), and where.

Geographies of Queer: Queer Globalizations and Contestations

What we now call Queer Theory was largely developed by academics working in the Humanities department in US universities – this geography has meant that it is often viewed as an Americanized set of theoretical ideals that have little applicability beyond the 'developed' United States. Particularly problematic for the spatiality of sexualities has been the use of queer as an all en compassing identity and theoretical frame with hom ogenizing diversities between various manifestations of non normative sexual identities, desires, and practices. 'Queer' as synonymous with 'lesbian' and 'gay' is likened to the Americanization phenomena of globalization processes. These include, specifically, overarching as sumptions of liberation and 'gay rights' that are un questioningly assumed and often cast countries as either 'forward' or 'backward'. Such a globalizing queer identity and agenda does not recognize the cultural nuances and diversity of sexualities beyond minority ('developed') world contexts.

A number of studies have shown that as sexualities and genders have geographies, the various formations of normative and non normative sexualities and genders are not uniform across the globe. Thus, for example in Indonesia, whilst 'lesbians' are viewed as deviant and 'not Indonesian', women can live in same sex relationships as long as one partner adopts a traditionally masculine/ butch role. Similarly, men who are identified as 'gay' after exposure to global media can still retain an expectation of marriage. By contesting the normative (Anglo American) conceptualizations of sexualities, these practices and identifications could be described as 'queer' as they question the boundaries of sexualities and gender. Simultaneously, however, they also contest an understanding of a globalized universal 'queer' identity. Thus, it can be argued that lesbian and gay 'liberation' may be based on particular understanding of sexuality and gender that does not account for cultural nuances and practices. Hegemonic assumptions that contend that 'lesbian' and 'gay' identities that are 'out' are markers of tolerance, acceptance, and liberation can be critically examined illustrating the spatial contingency of identities and norms (including examining supposed 'liberal' countries and 'who' is accepted and tolerated and 'where').

Queer Geographies

Building on the queer approach that suggested identities and bodies were not fixed but continually being re made, geographers have contended that space is not simply the container in which things happen, but is actively produced through the actions that take place. In 1994, David Bell and colleagues suggested that space was not simply heterosexual but heterosexualized through particular acts and norms. From this beginning, geographers have contended that sexualities are fluid, contextually enacted, and, importantly, spatially contingent. Works on sexuality in this vein are challenging how space is always already assumed to be 'heterosexual space', or rather, how the quotidian repetition of heterosexual relations becomes normalized, such that everyday space is assumed not to be sexual. In this way, taken for granted heterosexuality becomes invisible. These geographies examined the spatial practices of disrupting such heterosexualised space, including lesbian and gay Pride events, where city streets are temporally 'queered' and the normativity of heterosexual uses exposed. This was particularly apparent in the early 1990s where Pride marches sought equality and deliberately aimed to re appropriate space for same sex acts in otherwise hostile cities. It could be argued that initial investigations of sexualities in geographies queered the taken for granted heterosexual norm of the discipline. In even introducing sex, fucking, and sexual 'deviants' into geography, the sanctity of this heteronormative discipline was contested and the boundaries of geographies challenged. Yet as queer theorizing has developed, the critically deconstruction of sexual normativities has emerged as distinct from the identity politics that were central to early geographies of sexualities.

As queer geographies continue to develop their divergences from geographies of sexualities that focus on sexual identities, subjectivities, and lives, queer geographies as a critical approach to normativities becomes apparent. Although queer theory can be traced back to its American humanist roots, these diverse sets of ideas were first taken up by British, rather than American geographers. Knopp has suggested that this was partly a reflection of the more conservative outlook of American geography departments, and also a measure of the different political and cultural traditions of gay 'communities' on either side of the Atlantic – a British 'cultural politics of resistance' versus American attempts to spatially consolidate gay economic and political power. Similar to the divergences between lesbian and gay studies (or sexuality studies, more broadly) and queer theory, then, it can be argued that there are now differences (as well as overlaps) between geographies of sexualities and queer geographies. Whilst these interact and overlap they are not synonymous and can offer critiques of each other. This position is contested and some argue that geographies of sexualities are always queer as long as they contest a heteronormative discipline. This does not account for queers deconstructive possibilities including the requirement of queer to question categories such as 'gay' and 'lesbian' which geographers require to map and identify gay ghettos and territories. It also denies the potentials of queer thinking to challenge the assumptions inherent in geographies of sexualities, and epistemologies and methodologies.

Geographers, similar to other queer theorists, have recently begun to address homonormativity. In exploring how spaces come into being, at the start of the twenty first century, there were investigations of how spaces come to be sexualized through multiple and interlocking social differences. This most recent queer geographical work looks not just at the hegemonies of heterosexuality, but also at the multiple diversities between those who identified as gay, lesbian, and/or queer. Taking the Pride example, the recent commercial interest in Pride events and the embracing by certain cities of Pride celebrations has arguably challenged Pride as an antinormative practice. Thus, these events, whilst remaining lesbian and gay, can and have been critiqued using a critical antinormative frame.

Yet there have been some contestations of homonormativity with geographies. Those in seemingly 'hegemonic' positions/positionalities have argued that the assumption of privilege based on a homogenous subject position is problematic, particularly as the discipline, the academy, and universities remain strongly within heteronormative frames. Moreover, they have contended that spatialities need to be introduced to discussions of homonormativity so that atemporal or aspatial assertions regarding privileged subject positions are not produced. Yet, questioning presumptions of homonormativity that homogenise experiences and lived realities, should also recognise and explore privilege.

'Queer', in geographies, is usually explored in relation to sexual marginalization/'deviancy', nevertheless gender and gender transgressions are beginning to be addressed. Within feminist geographies the continued and often unquestioning use of categories of man/woman suggests that these have yet to be extensively problematized. Thus, whilst geographies of sexualities are engaging with queer theory, gender and feminist geographers have yet to explore the questioning of the categories of man/woman. Thus, there is a dearth of research and theorization of gender disruptions in geographic research. There has been some discussion of 'camp' men and more recently women who are mistaken for men. This research builds on queer theories discussion of the instability of sex and its need to be reiterated, augmenting it by arguing that space and sex are mutually interdependent. These explore the spatial contingency of gendered enactments, that is, how gender is both done and policed differently in different places. It then goes on to argue that sexed embodiments can both create sexed spaces (often within the dichotomy man/woman) and are made by such (male/female) spaces.

The most recent work in queer geographies has explored immanence and potentialities, particularly how sexualities and genders are continually being re created in this way, rather than that. Potentialities and the realization of particular events have been used not only to render sexualities, genders, and spaces fluid, but also to discuss their contingent re formation. Drawing particularly on Deluze and Guittari, these geographies are refocusing attentions onto that which can never be as well as what comes to be normalized. This allows for disruptions of taken for granted norms, in favour of what might have been and what still could (potentially) be.

Geographies of sexualities, which often address sexual dissidents, are represented in many courses across the discipline, and sexualities are increasingly common in research and teaching texts. It could then be argued that certain forms of geographies of sexualities are now normalized, or at least, have a certain level of acceptance in the discipline. This contention would question whether geographical investigations of sexualities are necessarily non normative or critically deconstructive. Queer geographies is beginning to question the orthodoxies and what is taken for granted or assumed by geographies of sexualities, as well as in the broader discipline and academy. Thus, it is possible to contend that geographies of sexualities now coexist with queer geographies.

The Trouble with Queer Troublings

As outlined above, 'queer' as an all encompassing term has been critiqued for its lack of understanding of geographical nuances. However, queer has also been contested for its questioning of identity categories. This is because these identity categories are used as a basis for identity politics and in addressing the marginalization of particular individuals and groups. Categories such as woman, lesbian, and gay can be seen as increasingly gaining some 'equality' with those who are understood as having particular societal privileges (men, heterosexuals). Queers questioning of normative identities could undermine the basis of identity politics as a platform to argue for social justice for previously marginalized groups. Consequently, queer can be read as not creating productive change and in its multiplicity, losing the power of collective action. It has been argued that 'queer' is a dangerous and premature phenomenon for groups that have yet to gain equality.

There have been a number of feminists who have argued that 'queer', as an overarching category, erases significant gender differences. Lesbian feminists in particular have argued that using the term 'queer' to represent all those who are lesbian, gay, bisexual, or trans or even those who are not normatively heterosexualized does not address the many (gendered) divergences in this grouping. Some have asserted that using the overarching term 'queer' rather than the gendered specific term 'lesbian', means that just as with the use of 'man' to refer to humans, women once again disappeared under this universalizing signifier. They, and others, contend that 'queer' can encompass too much, such that the term, and its usefulness, comes into question.

Conversely, there has also been a questioning of queer theory as fragmenting particular categories that continue to be salient to everyday lives, such as man/woman. Recognizing that identity categories are not homogenous and yet, working within those, in particular, disempowered categories (such as woman) to affect legislative or other forms of social change has been termed 'strategic essentialism'. This has included a move to reinstate the place of the category 'woman' (amongst other categories). This is based on a backlash to what is perceived as dangerous and premature postmodern deconstruction of the category 'woman' just as women began to gain particular power. Authors in this vein posit that categories of woman are necessary for collaboration in light of the continuing patriarchy and inequity between men and women. These feminist authors are not alone in arguing for coherent identity categories as the basis for fighting for equality and recognition of the needs for particular gender and sexual rights. Lesbian and gay activists have also contended that queer has prematurely explored hegemonies within lesbian and gay rights movements and therefore weakened an already disenfranchised grouping. Gathering collectives (for some 'communities') for specific causes or to address particular social inequities does not necessarily disregard differences between groupings. Rather, the aim of strategic essentialism is to focus on that which is the 'same', to affect social change. In this way, queer concepts such as heteronormativity, and potentially homonormativity, can be deployed to challenge inequitable practices, yet this can have normative impulses which could be read as 'antiqueer'.

The final trouble with 'queer', addressed here, is the aspect of the debate regarding the place and importance of 'queer' in its emergence through humanities and the lack of 'real world applicability' 'queer' has beyond deconstructing cultural texts. Where power relations and particularly inequality is central to discussions within feminism and lesbian and gay studies, it has been argued that 'queer's' theoretical focus is distanced from the struggles in the 'real world'. Is there a place for deconstructing identities where individuals continue to be maimed, beaten, and killed as a result of their 'queerness' (be that around sexuality or gender)? Similarly whilst identities may be theoretically rendered unstable and fluid, the policing of these flows can have material consequences, in forms such as lack of employment opportunities and social support as well as experiences of violence, abuse, and harassment. These policings do not simply need theoretical responses, but they also necessitate emotional and physical support in numerous forms from the provision of safe space through to legislative equalities. (However, this argument neglects the material anarchist and non normalised ways of living that are practiced at queer events such as queeruption, as well as the place of queer protests in anti war and other forms of demonstration.)

'Queer' as a critical stance can therefore be considered as dangerous and as not effecting lasting change, moving from one critique of normalization to the next. The language used is often seen as alienating, elitist, and exclusionary as queer theorists grapple with difficult concepts and ideas. These are rarely translated into accessible language, and often this may not be possible. To argue that this form of theorizing should not occur is however problematic as queer theorists have informed the development of thinking in many areas beyond sexualities and genders including recent theoretical impulses in geographies which include postrepresentational theory.

Conclusion

From the beginning 'queer' has been a contested concept, its originator disowning it when she felt it had become normative. 'Queer' can be read as an all encompassing term for sexual dissidents; however, queer theory has contested such identity categories and developed as a critical engagement with sexual normalizations. Queer theory is useful for continually questioning how hegemonies come into being as well as dismantling the very structure that prejudices rely on. Its antinormative impulses and the engagement with non normativities contests the assumption that sexual and gender identities are fixed, stable, or coherent and questions their formation beyond other social differences such as race, class, and disability. Questioning power, hegemony, and hierarchies within and between sexual dissidents, queer theory has investigated homonormativity, and the normalization of particular lesbian, gay, bisexual, and transgender (LGBT) lives, identities, and subject positionings. Queer theorists have also challenged the aspatiality of identity categories and the globalization of particular forms of gay and lesbian 'rights' without recognizing the spatial and temporal contingencies of such normalizing processes.

Engaging with the performative (re)formation of identities and embodiments, queer geographies have contended that not only is space sexualized through specific normative processes, it also, in part, sexualizes and genders bodies. The divergence between queer theory and sexuality studies resonates with recent geographical work that seeks to distinguish between geographies of sexualities and queer theories.

Queer theory is controversial in its engagement with potentially disenfranchised and marginalized groups. By querying the (re)creation of identity categories, identity politics come into question and civil rights and social justice movements could arguably be undermined, yet it can be contended that there is also a need for other forms of inquiry, activism, and critique. These could recognize and use sexual and gender categories as a basis to fight such prejudices as patriarchy and homophobia. There are those who dismiss the queer project as irrelevant, unnecessary, and elitist – this denies the relevance of multiple forms of methodologies, epistemologies, and ontologies in the investigation of geographies, lives, and texts.

See also: Embodied Knowing; Feminism/Feminist Geography; Gay Geographies; Heteronormativity; Identity Politics; Masculinism; Performativity; Postcolonialism/Postcolonial Geographies; Sexuality.

Further Reading

Ahmed, S. (2004). *The cultural politics of emotion*. Edinburgh: Edinburgh University Press.

Bell, D. (2001). Fragments for a queer city. In Bell, D., Binnie, J., Holliday, R., Longhurst, R. & Peace, R. (eds.) *Pleasure zones: Bodies, cities, spaces*, pp 84 102. New York: Syracuse University Press.

Bell, D. and Binnie, J. (2000). *The sexual citizen: Queer politics and beyond*. Cambridge: Polity Press.

Bell, D., Binnie, J., Cream, J. and Valentine, G. (1994). All hyped up and no place to go. *Gender, Place, and Culture* 1(1), 31 47.

Bell, D. and Valentine, G. (1995). *Mapping desires: Geographies of sexualities*. London: Routledge.

Brown, G. (2007). Automony, affinity and play in the spaces of radical queer activism. In Browne, K., Lim, J. & Brown, G. (eds.) *Geographies of sexualities*, pp 195 207. London: Ashgate.
Browne, K. (2004). Genderism and the bathroom problem: (Re)materialising sexed sites, (re)creating sexed bodies. *Gender, Place, and Culture* 11(3), 331 346.
Browne, K. (2006). Challenging 'queer' geographies. *Antipode* 38(5), 885 893.
Browne, K., Lim, J. and Brown, G. (2007). *Geographies of sexualities: Theories, practices and politics*. London: Ashgate.
Butler, J. (1990). *Gender trouble*. London: Routledge.
Butler, J. (1997). *The psychic life of power: Theories in subjection*. Stanford: Stanford University Press.
Halberstam, J. (2005). *In a queer time and place: Transgender bodies, subcultural lives*. New York: New York University Press.
Hubbard, P. (1999). *Sex and the city: Geographies of prostitution in the urban west*. London: Ashgate.
Knopp, L. (1998). Sexuality and urban space: Gay male identity politics in the United States, the United Kingdom, and Australia. In Fincher, R. & Jacobs, J. (eds.) *Cities of difference*, pp 149 176. London: Guilford.
Oswin, N. (2008). Critical geographies and the uses of sexuality: Deconstructing queer space. *Progress in Human Geography* (Forthcoming).

Relevant Websites

http://www.theory.org.uk/ctr quee.htm
Media/identity/resources and projects University of Westminster.

Questionnaire Survey

V. Preston, York University, Toronto, ON, Canada

Glossary

Anonymity Researchers are unable to associate responses with specific respondents.

Closed-Ended Question Responses that are chosen to make sense, to be mutually exclusive, and to be exhaustive.

Confidentiality The responses of each respondent will not be identified publicly even though the researcher may be able to identify them.

Inferential Statistics They are used to determine whether and how patterns observed in information from a sample can be generalized to the entire population.

Nonprobability Sample They are samples in which members of a population have different probabilities of being chosen.

Open-Ended Question Any response is valid and they are often recorded verbatim.

Pilot Study A few respondents are selected and surveyed according to all the procedures that will be used to administer the questionnaire.

Population All units of analysis that share one or more characteristics.

Pretest Friends and colleagues complete the questionnaire to provide comments on wording, layout, order, and length.

Questionnaire Survey A technique for gathering statistical information about the attributes, attitudes, or actions of a population by administering standardized questions to some or all of its members.

Random Sample Each member of a population has an equal probability of being chosen ensuring that information from the sample is representative of the underlying population.

Reliability It describes whether the same information would be obtained by drawing repeated samples from the same population.

Response Rate The proportion of people in the sample who complete the questionnaire.

Sampling Frame An enumeration of all members of a population from which a sample can be selected.

Validity It refers to the correspondence between the questions and the concepts in which the researcher is interested.

Questionnaire Survey

Many of us are familiar with questionnaire surveys in the form of political polls and marketing surveys usually delivered by telephone and invariably at inconvenient times. The national census in many industrialized countries such as the United States, Australia, and Canada is also a questionnaire survey. "A technique for gathering statistical information about the attributes, attitudes or actions of a population by administering standardized questions to some or all of its members" (Buckingham and Saunders, 2004: 13), the ubiquity of questionnaire surveys attests to their utility and ease of use. Mainly useful for descriptive research, surveys provide broad coverage of populations enabling us to explore the extent and nature of spatial and social variations in people's attributes, attitudes, and actions. Generally, the aim is to obtain information suitable for statistical analysis. As a result, attention is paid to how respondents are selected, the extent to which questions relate to underlying concepts, and completion rates.

Questionnaire surveys are administered in many different ways. Paper forms are still completed in face to face interviews with an interviewer asking each question and recording each answer. Paper forms are also used in self administered questionnaires that are completed without any intervention from the researcher or his or her surrogate, the interviewer. Self administered questionnaires may be delivered to people's homes, sent by mail, or through the Internet. Growing numbers of surveys are being conducted over the Internet. Remarkably inexpensive, self administered Internet surveys allow access to small and dispersed populations. To save money and time, many surveys are now administered over the telephone. An interviewer asks questions as they appear on a computer screen and enters the responses directly into an electronic database. Computer assisted interviews on the telephone, through the Internet, and in face to face interviews allow for more sophisticated skips among questions so that the responses to early questions alter the content and order of subsequent questions. Direct entry of information also improves the quality of data while shortening the time required for data preparation.

When Is a Questionnaire Appropriate?

A questionnaire survey is only appropriate for certain research questions. Its suitability depends on the types of information needed to answer a research question and the people from whom the researcher wants to elicit information. Questionnaires are inappropriate for collecting information about sensitive topics such as sexual orientation and illicit activities. People will rarely talk about actions that put them at legal risk. The structured nature of the questions and the brevity and superficiality of the social encounter between researcher and respondent do not encourage the intimacy and trust that are prerequisites for people to reveal behaviors, beliefs, and attributes that might be unsanctioned.

Questionnaires that rely on people's abilities to convey information accurately are also often ineffective for learning about the past. Over time, memories evolve so that responses about the past are often incomplete. Panel studies in which the same respondents participate in a series of surveys are a good method for learning about past actions and attitudes; however, they are expensive and difficult to implement. Panel studies also pose unique geographical difficulties. To reduce costs, samples are often small so they provide little information about populations in specific places. For example, in Canada and Australia, panel studies of immigrants are rich sources of information about settlement processes. Unfortunately, the small samples mean information is available only for the largest metropolitan areas in each country.

Finally, a questionnaire survey is effective only when respondents have knowledge of the topic and they are competent to answer the questions. Questions must be relevant to respondents, and respondents must have the information and the ability to answer. Surveys of households usually exclude children as potential respondents because they often lack the information needed to answer the questions, the issues under study are often not relevant to them, and young children may lack the cognitive ability to answer complex questions.

Despite these limitations, numerous research questions and topics may be addressed using information collected with questionnaire surveys. Geographers have used questionnaire surveys to learn about the attributes of many different populations, including employees and owners of firms, neighborhood residents, the homeless, antiglobalization activists, politicians, and refugees. Equally varied behaviors ranging from international migrants' remittances, men's and women's working conditions, and their travel patterns to food practices and access to health services have been investigated using questionnaire surveys. Survey information has also been collected about residential preferences, mental maps, regional identities, and political opinions. In all of these cases, the success of the survey depended on questionnaire design and administration.

Designing a Questionnaire Survey

The information obtained from questionnaire surveys is constructed through the process of designing and administering the questionnaire and compiling the data that result. The design of the questionnaire reflects the research question, the type and size of the sample, and the available resources.

The Research Question

The research question dictates the topics of questions. As a structured set of questions, the content of the questionnaire is designed to provide relevant information. Clear and detailed research objectives are crucial for designing a valid questionnaire. Validity refers to the correspondence between the questions and the concepts in which the researcher is interested. Detailed research objectives clarify the meanings of the concepts around which questions are developed.

Detailed research objectives also help researchers avoid irrelevant questions. Researchers often draw on previous questionnaires for suggestions about topics, wording, and order of questions. Although this strategy has many advantages, it may encourage the inclusion of irrelevant questions. For example, questionnaires often include questions about the sex and age of each respondent; however, this information is useful only if it is relevant to the research objectives.

Samples

A clear statement of the research objectives also facilitates identification of the population under study. The population refers to all of the units of analysis being considered in the study. Questionnaires were developed mainly to learn about individuals. Geographers who examine collectives such as households, firms, and organizations need to identify who will speak on behalf of the collective. For example, in a household, any person living in the dwelling who is over the age of 18 is often asked questions about residential history.

Researchers rarely survey all members of a population. Rather, a questionnaire survey is administered to a sample of the population, a subset of the population under study. Depending on the way the sample is selected and its size, sample information may represent the population. Ideally, samples are selected randomly and with replacement so that each member of the population has an equal probability of being chosen. Information from a simple random sample is considered to be representative of the population and reliable. Reliability describes whether the same information would be obtained by drawing repeated samples from the same population. Error is measured by comparing a sample estimate with the equivalent value in the population. For example, average annual income per person in the population of a metropolitan area may be US$40 000. One sample may provide an estimate of US$39 995 and another of US$40 002. In these instances, the error is small so the samples would usually be considered reliable.

Many surveys use stratified samples in which each member of a population subgroup has a roughly equal chance of being selected. Stratification ensures that key characteristics of the units of analysis are considered. Geographers often stratify samples on the basis of lo cation, so that the number of respondents from each lo cation is proportional to their number in the total population. When the sample is small, stratification en sures all groups and locations are represented in the sample. Alternatively, the researcher may want to over represent a location or a social group. If the samples from the strata are analyzed separately and never combined or compared, each sample is representative. To combine stratified samples, weights that represent the fraction of each underlying population in each sample must be calculated and applied to correct the estimates for the total population.

Nonprobability samples are often selected to save time and money. Researchers often lack a sampling frame, an enumeration of all members of a population from which a sample can be selected, and compiling a sampling frame is often expensive and time consuming. For these reasons, many geographical research projects use systematic samples, that is, knocking on every fifth door on the street. Quota sampling is a more sophisti cated version of nonprobability sampling in which knowledge of selected characteristics in the population is used to set quotas that ensure that all types of people and places are included in the sample. For example, the population consists equally of workers from two different cities. The target number of respondents for each city is set at 50. The questionnaire is administered until there are 50 completed questionnaires from each city. Quota sampling is fast and cheap. The questionnaire is ad ministered to whoever meets the criterion until enough completed questionnaires are obtained. On the other hand, a sampling frame is not needed. Unfortunately, systematic and quota samples are not representative of the underlying populations. As a result, inferential stat istics are invalid and valid response rates cannot be cal culated, a problem that will be discussed shortly.

Sample size influences the reliability of the infor mation obtained from a questionnaire survey. Large samples are more likely to represent a population re liably. For example, a simple random sample of 400 provides estimates of central tendency that are accurate within a 95% confidence limit or 19 times out of 20. As sample size increases, the reliability of the information improves. Compared with the sample of 400, a simple random sample of 10 000 drawn from the same popu lation will provide estimates that are accurate within a 99% confidence limit or 99 times out of 100. Stratifi cation improves the reliability of sample estimates by ensuring that all groups in the population are included in the sample.

The Constraints of Time and Money

The design of all questionnaire surveys reflects the availability of resources. To save money, the number of questions measuring a concept is reduced, sometimes reducing the validity of the information. Simple random samples are often replaced by nonprobability samples to limit costs. Compromises are inevitable, but researchers need to ensure there is sufficient time and money to undertake a valid and reliable questionnaire survey. Limited resources often require rethinking the research question to narrow the topic and the population of interest.

Constructing the Questionnaire

The design of the questionnaire is influenced by the research objectives and the way that the questionnaire will be administered. The format, wording, and order of questions differ for questionnaires delivered by different methods. For example, questions to be asked in face to face and telephone surveys may benefit from the inter vention of an interviewer who can clarify questions and probe for answers. In contrast, the questions used in self administered surveys must be meaningful and clear to the respondents without the benefit of an intermediary.

Question Structure and Response Options

In geographical research, there is much debate about question structure with the most heated debates relating to the choice of closed ended and open ended questions. In closed ended questions, the response categories are spe cified. The categories are chosen to make sense, to be mutually exclusive, and to be exhaustive. For open ended questions, any response is valid. Closed ended questions are appropriate under three circumstances: when the number of possible responses is small, when the response is a specific number, and when researchers want to force respondents to reply with a limited set of categories. Open ended questions are appropriate for exploratory research to learn what people say about a topic, perhaps to narrow the possible range of responses. The method of delivery of the questionnaire also affects the utility of each format. Open ended questions should be avoided for self administered questionnaires. They are more appropriate for face to face and telephone interviews that allow the interviewer to explain, prompt, and probe for answers.

Questions may also have simple yes and no responses that are converted to indexes. The number of affirmative responses is calculated as a proportion of all responses to create an index. To construct a valid index, the individual questions, known as items, should be of equivalent value. Several techniques including correlation analysis are used to indicate how strongly the items are related to each other and whether they comprise a valid index.

Although indices can be very useful, more sophisti cated information can be obtained from scales that assess the intensity of people's actions and attitudes. Although there are many different types of scales, two are used often in geographical research: Likert and semantic dif ferential scales. In Likert scales, respondents indicate the intensity of their agreement or disagreement with an opinion. For semantic differential questions, people rate their feelings on pairs of adjectives. The individual items thought to comprise a scale must be assessed for their interrelations and coherence.

Each type of question has its strengths and weak nesses. For example, closed ended questions are more reliable than open ended questions whereas open ended questions have greater validity than closed ended ques tions. The research questions and existing knowledge of the research topic are the two main considerations when deciding question structure.

Writing Questions

Successful questions are short, clear, and meaningful. Brevity and clarity help respondents understand the question and retain their interest in the questionnaire. Successful questions also refer to concrete situations ask ing people how many times they drove to work in the past week rather than how often they usually drive to work. Simple wording also reduces bias in questions that en courages respondents to answer in a particular way. Bias arises from many sources, including the choice of words in a question and the identification of an attitude or action with a powerful person or institution. Negation often creates bias by confusing respondents who are uncertain whether they should agree or disagree with a negative statement. Double barreled questions in which a person may agree with one part of the question, but disagree with another part are another potential source of bias. Leading questions such as 'How often do you beat your wife?' may encourage respondents to reveal behaviors and attitudes that they might otherwise conceal, however, this type of question may also confuse respondents who are generally reluctant to answer negatively.

Questions used in previous research are often in corporated into subsequent questionnaires. With clear wording, previously used questions and their instructions may enhance a questionnaire. Consistent wording also creates opportunities for comparative research that may offset the disadvantages of a small sample. However, questions need to be borrowed judiciously and only if they are related to a study's research objectives.

Order and Layout

The order, layout, and length of a questionnaire will influence the willingness of respondents to answer questions. Questions need to be ordered so that the answers to early questions do not affect subsequent answers. The way the questionnaire is delivered also af fects the order of questions. In face to face and telephone interviews, the interviewer builds rapport by asking questions about personal attributes at the beginning of the survey. In self administered questionnaires, the most interesting questions begin the questionnaire and per sonal questions are asked at the end of the survey.

Many questions are relevant only for selected re spondents. For example, questions about working con ditions are only asked if the person is employed. Contingency questions in which a question is asked de pending on responses to the first question in a series allow researchers to ensure the relevance of questions. They also create complicated sequences of questions that require clear and detailed instructions for the respondent and/or the interviewer.

The ideal length of a questionnaire is a vexing issue. Researchers must balance the desire to include every possible question with the additional cost of each ques tion and the unwillingness of respondents to answer lengthy questionnaires. Detailed research objectives are needed to guide the inevitable pruning of superfluous questions. Difficult editorial decisions should be based on consideration of the research objectives and appreciation for the time and effort being asked of respondents.

A pretest is invaluable for evaluating the length of a questionnaire, the wording, the order, and the layout. A pretest occurs when respondents complete the ques tionnaire, not just read it. A pretest can be completed by friends and colleagues who will identify questions and response categories that are ambiguous and boring, confusing instructions, and questions that are out of order. A pretest will also indicate whether the question naire is too long. A pretest is often followed by a pilot study in which a few respondents are selected and sur veyed. Unlike the information from a pretest, information from a pilot study may be included in the survey.

Administering the Questionnaire

Once the questionnaire has been designed, the sample selected, and a pretest and pilot study completed, the survey is ready to 'enter the field'. Those who agree to participate in the survey need to be tracked from the time they are identified in the sample to the final stage of compiling and analyzing the data from the questionnaire survey. Respondents' participation can be increased by judicious use of cover letters and other forms of pre notification. An adequate response rate is important to ensuring the reliability of the survey.

Informed Consent: The Ethics of Social Research

Social research is guided by several important ethical principles. They include asking people to participate only in worthwhile research that has some academic or social value. Researchers must also ensure that partici pants will not come to any harm. If there is any possi bility of harm, potential participants must be informed about the possibility and advised about appropriate sources of help. Respondents must be given all relevant information about the research including their right to withdraw at any time.

The last ethical principle concerns the protection of respondents' identities. Confidentiality assures each re spondent that his or her responses will not be identified publicly even though the researcher may be able to identify individual respondents. Confidentiality is different from anonymity which occurs when researchers are un able to associate responses with specific respondents. Confidentiality is achieved by separating information that could identify a respondent such as his or her name or telephone number from all other questionnaire infor mation. Respondents are identified only by identification numbers in written and electronic records. Anonymity requires that completed questionnaires are numbered in dependently of any tracking scheme for respondents and usually in the order that the questionnaires are completed. Since courts have sometimes subpoenaed completed questionnaires and other research materials, researchers need to think carefully about how they will ensure re spondents' confidentiality or anonymity.

Tracking Respondents

As the sample is drawn, a mutually exclusive identifi cation number should be assigned to each potential re spondent and a system for tracking each respondent should be developed. The system will indicate how many times and the dates that the potential respondent was contacted and the outcome of each contact. People who can never be contacted should be distinguished from those who refuse to participate in the survey.

Depending on the way that the questionnaire is de livered, certain additional information will also be re corded. For face to face and telephone interviews, the date, time, and duration of the interview, and the name of the interviewer are often recorded. Information about mail questionnaires usually includes the dates that the initial letter introducing the research and any follow up contacts were sent. Computer assisted surveys on the telephone, Internet, and face to face often collect this administrative information automatically. The importance of identification numbers cannot be overemphasized. They are the first step to maintaining respondents' privacy.

Cover Letters and Other Forms of Prenotification

Despite its brevity and superficiality, successful ques tionnaire surveys require establishing a social relation ship between the respondent and the researcher or interviewer. To convince the respondent to participate, the social and academic value of the research must be conveyed immediately upon initial contact. In face to face and telephone interviews, the interviewer introduces the survey by outlining the purpose of the research, the topics of the questions, the expected length of the interview, and the rights of the respondent.

Self administered questionnaires are usually intro duced with a printed version of the same information. A convincing introduction accompanied by an attractive and well laid out questionnaire with clear instructions and engaging questions and a stamped, addressed en velope in the case of a mail questionnaire will increase completion and return rates. Follow up contacts that often consist of an encouraging message in the case of an Internet survey or a letter in the case of a mail survey may increase the completion of self administered ques tionnaires. It is often effective to send a series of follow ups that begin with a brief reminder about the survey and a subsequent letter or electronic message accompanied by an additional copy of the questionnaire to make the task of completing the questionnaire as easy as possible.

Response Rates

To achieve a representative sample, the response rate, the proportion of people in the sample who complete the questionnaire, needs to be evaluated. People who cannot be contacted, those who refuse to participate in the survey, and those who do not finish the survey are likely to come from specific groups within the sample, so their exclusion will bias the information from the questionnaires. Keeping careful track of the history of contact with each person in the sample is essential for calculating response rates.

In general, response rates for surveys are declining. Many factors contribute to the decline, including the growing proportions of unlisted telephone numbers, in creasing reliance on answering machines to screen tele phone calls, the increased pace of people's lives, and growing suspicion about requests for information from strangers. In response, many researchers are investigating the best ways to calculate response rates, techniques for improving response rates, and the calculation of weights to compensate for low response rates from specific groups in the sample. The increasing unwillingness of people to participate in questionnaire surveys highlights the importance of all aspects of survey design that may help to engage reluctant respondents.

Analysis, Interpretation, and Presentation

Most questionnaires are designed to provide numerical information for statistical analysis. There are two paths to evaluating the questionnaire data. One relies on statistical inference to assess the statistical quality and significance of the information. The second relies on substantive interpretation of the information. Both require the initial compilation of the answers usually in electronic databases.

Maintaining Anonymity and Confidentiality

Among the many issues that arise when compiling questionnaire information, one of the most important concerns the responsibility to ensure that respondents cannot be identified individually. In most databases, each respondent is identified only by an identification number and information that might identify the respondent is removed. For geographers, maintaining confidentiality often requires thinking about how to store geographical information that could identify a respondent while still having the ability to link geographical information with information about the attributes, attitudes, and actions of each respondent.

Interpreting the Responses

Questionnaire surveys are usually designed to provide information for statistical analysis. The emphasis placed on random samples, high response rates, and clear and unambiguous questions reflects the need for reliable and valid information that can be analyzed statistically for descriptive and inferential purposes. Inferential statistics are used to determine whether and how patterns observed in information from a sample can be generalized to the entire population. Statistical significance measures the reliability of a statistical estimate such as average income and the relationships among different variables such as income and support for recycling. Many questionnaire surveys do not meet the requirements for statistical inference. In these instances, questionnaire information can be used to describe the sample itself, but it cannot be generalized beyond the sample.

The interpretation of information from questionnaire surveys is often based on two considerations. Statistical significance is used to measure the reliability of estimates and the strength of relationships. Significant findings must also have substantive significance, the observed estimates and relationships must be interpretable and meaningful. This is the only criterion that can be used to assess questionnaire information that is unsuitable for statistical inference.

Presentation of Findings

Despite the appearance of objectivity that is often associated with numerical data, information from questionnaire surveys is interpreted by the researcher. Common practices for presenting findings based on information from questionnaire surveys help the reader evaluate the researcher's interpretations. In general, sampling strategies, the composition and size of the sample, and response rates are described. In some instances, the wording and format of questions used to elicit answers may also have to be discussed. The aim is to inform readers about any aspects of the questionnaire design that may influence the validity and reliability of the information.

Questionnaire surveys provide descriptive information about populations that is useful for examining the nature and extent of social and spatial variations in actions, attitudes, and attributes. The information complements more detailed information gained from semi structured interviews and other qualitative methods. Survey information can also be evaluated relative to specific guidelines about sampling and response rates. Following the guidelines does not ensure successful research, but it does enhance the chances that the research will contribute to geographical knowledge.

See also: Content Analysis; Fieldwork; Focus Groups; Interviews: In-Depth, Semi-Structured; Mixed and Multiple Methods; Quantitative Data; Quantitative Methodologies; Statistics, Descriptive; Statistics, Inferential.

Further Reading

Bradburn, N. M., Sudman, S. and Wansink, B. (2004). *Asking Questions: The Definitive Guide to Questionnaire Design: For Market Research, Political Polls, and Social and Health Questionnaires*. San Francisco, CA: Jossey Bass.

Buckingham, A. and Saunders, P. (2004). *The Survey Methods Workbook: From Design to Analysis*. Cambridge: Polity Press.

Dillman, D. (2007). *Mail and Internet Surveys: The Tailored Design Method*. Hoboken, NJ: Wiley.

Flowerdew, R. and Martin, D. (2005). *Methods in Human Geography: A Guide for Students Doing a Research Project* (2nd edn.). Harlow: Prentice Hall.

Hoggart, K., Lees, L. and Davies, A. (2002). *Researching Human Geography*. London: Arnold.

Presser, S., Rothgeb, J. M., Couper, M. P. *et al*. (eds.) (2004). *Methods for Testing and Evaluating Survey Questionnaires*. Hoboken, NJ: Wiley Interscience.

Schuman, H. and Presser, S. (1996). *Questions & Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. Thousand Oaks, CA: Sage Publications.

Statistics Canada (2003). *Survey Methods and Practices*. Ottawa: Statistics Canada, Social Survey Methods Division.

Relevant Websites

http://www.iser.essex.ac.uk
Institute for Social and Economic Research, Essex University.

http://www.icpsr.umich.edu
International Consortium for Political Science Research, University of Michigan.

http://www2.norc.org
National Opinion Research Center, University of Chicago.

http://www.statcan.ca
Statistics Canada.

R

Race

J. Winders, Syracuse University, Syracuse, NY, USA

Glossary

Colonial Discourse Texts, practices, and ideas concerning the dynamics between colonizer and colonized, particularly vis-à-vis the unequal power relations and constructions of difference associated with colonialism.

Difference The parceling out of bodies, peoples, and spaces along lines of race, class, and other aspects of identity, often in binary fashion.

Ethnicity A socially constructed category linked to a presumed shared history and cultural heritage.

Overdetermination The idea that categories, such as race and class, always work in relation to one another to such an extent that separating each category's effects is largely impossible.

Racialized Landscapes Places where racial formations are inscribed into the landscape, which subsequently informs ideas about race, history, and a collective future.

Racism Ideas or practices that, through a power-difference coupling, essentialize linguistic, cultural, or national differences and use them to structure unequal social, economic, political, and cultural relations across space and scales.

Social Construction The idea that identity categories such as race are not natural or essential but produced through social actions, daily life, and broader structural forces.

White Privilege The accrual of social, economic, political, and cultural privileges by white individuals solely by virtue of being white; requires neither acting agents nor intentional discrimination.

The Power of Race

Race is not easy to define. At times, it has been conflated with biology and treated as a 'natural' category that predicts intellectual capacity, work ethic, and physical prowess. At other times, race has been confused with ethnicity and framed as something that groups and individuals can 'overcome' with the right amount of effort. Today, it goes without saying that race is a social construction but, as such, has material consequences that can be matters of life and death. Race has been central to spectacular campaigns, from the Civil Rights Movement in the US to anticolonial struggles across Africa. It has also been key to everyday acts, from where people shop to what jobs they fill. Although there is nothing 'natural' about organizing people into categories defined by skin color, there is nothing arbitrary about race's workings vis à vis colonialism, nation building, industrialization, and other processes. Understandings of race change over time; and it has been studied through different theoretical, empirical, and political approaches. What has remained constant across framings of race, however, is the coupling of power and difference that gives race its analytic purchase on the world around us and racism its deadly ability to fix difference and produce inequalities across scales.

What, though, is race? As a starting point, race can be understood as a chaotic, yet powerful, collection of ideas and practices through which peoples and places have been organized and ranked across time and space, according to a dynamic set of embodied and social characteristics often linked to skin color and always structured by unequal power relations. Where, however, does race come from? Is race geographical? Is geography racialized? How does race affect – and how is it affected by – social, economic, political, and cultural relations? How is race gendered, classed, sexed, bordered, and vice versa? Where, why, and how does race matter in everyday lives and wider social structures? What does racism look like, and how can it be undone?

This article examines these questions, and the links between race and space. Overdetermined by axes of difference from class to gender to nationality, race acts as both a set of 'constructs' produced across scales and a set of 'practices' informing and informed by daily life. Race is both a purportedly fixed category whose content is so obvious as to require no explanation and an ostensibly fluid social force whose form is malleable across time and space. A contradiction in Enlightenment humanism, a justification for colonial exploitation, and a component in geographies of uneven development, race possesses chameleon like characteristics that allow it to work in seeming paradoxical ways. Race, for instance, has not only been operationalized to underwrite an assortment of exclusionary projects but has also acted as a means to challenge exclusionary agendas and create community.

Race works geographically, parceling out bodies and spaces, and is itself responsive to geographic scales and borders. Whether understood as that which explains human differences or that which itself needs explanation as a social and political category, race links present and past, here and there, large socioeconomic forces and small everyday acts. Recursive across space and time, race acts as both pretext and context for social interactions. We live our lives across places and scales; and racial lessons learned in one place inform social actions in another and affect how we understand, and act upon, race itself. How can race be and do all these things?

Given race's slippery nature, much work on the topic focuses on what race does, rather than on what it is. Enumerating where and how race matters in people's lives, how it structures the state and civil society, and how it informs economic forces, social organization, and cultural meaning, scholars have examined race's role as an organizing grammar of difference, perhaps most profoundly in colonial discourses and practices. Intimately connected to the conflation of 'European', 'Christian', and 'civilized', race was a, if not the, structuring concept (and justification) for the colonial world and power matrix. With roots in medieval religious practices, it has supported social stratifications since the sixteenth century, although most scholars agree that race gained strongest currency in the early eighteenth century shift from theological power toward scientific authority, exploration, and colonial exploitation. In this era, race played a key role in the emerging map of politically and culturally defined nations, as race and nation merged in scientific and popular reflections on color and civilization. Informed by natural science, underwritten by colonial power, and defined through geography, a racialized hierarchy came to define a European, white, masculine self against a colonized, nonwhite, feminized other. Although the term racism did not circulate in this era, it informed this spatialized system that privileged white European men over nonwhite colonized men and women.

Colonial discourses of race were also tightly bound to gender and sexuality. White women in British colonial territories, for example, were often positioned as keepers and transmitters of 'civilized' white identity, charged with both racial uplift of the colonial subjects with whom they interacted and the material and cultural reproduction of the white race. This latter task in particular was frequently a key rationale for the strict policing of spatial and social lines between colonizer and colonized. White masculinity, too, was defined vis à vis colonial practice and geographies as Victorian ideas of civilization, manhood, and race positioned white men as preeminent examples of masculinity and virility 'at home' in Europe and 'abroad' in the colonies. Colonial discourses defining white 'manliness' worked against both the image of the white European woman discussed above and the figure of the colonized man, sometimes feminized and sometimes hypersexualized but always lacking reason and other 'civilities'. For colonized subjects, links among colonialism, race, and gender had dire consequences, as poignantly discussed by postcolonial writers. In his analysis of colonialism, for instance, Franz Fanon detailed colonialism's denial of black male subjectivity on its own terms. Although Fanon failed to engage the gendered nature of his arguments, he rightly stressed that the plight of always being black in relation to whiteness made postcolonial struggles about not only political independence but also subjectivity, showing the centrality of race to decolonization efforts.

Equations among race, colonialism, and difference also permeated the nineteenth century, as racial concepts first developed in the colonial encounter returned with brute force to Europe to link race, progress, and human evolution. Through social Darwinism and eugenics, social traits and biological hereditary were conflated, and the meaning of race was transformed. Racial characteristics previously defined by place or culture and thought to be relatively mutable became fixed essences defined by biology and nature and made to account for physical and cultural differences in human behavior and bodies. Kay Anderson's recent project on British colonial encounters in Australia has sharpened understandings of this mid nineteenth century shift toward racial determinism, showing the need to qualify general narratives of colonial dominance. Her work also joins a growing body of geographic work that critically engages the links among race, nature, and humanness itself.

Drawing a more rigid definition of race into constructions of nation and citizenship, nineteenth century political discourse in Europe, North America, and Latin America used practices such as eugenics as scientific and social justification to define the racial boundaries for the body politic. In many cases, the desired citizenry was defined as white, evident in past immigration policies privileging white immigrants in countries from the US and Canada to Australia and Argentina. Links among race, nation, and state, however, also operate within national borders. Kay Anderson's earlier work on Vancouver's 'Chinatown', for example, shows the state's role in constructing racial categories, such as 'Chinese', through spatial forms, such as 'Chinatown'. Nicola Foote's analysis of the intertwining of nation, state, and race in early twentieth century Ecuador shows how black and indigenous Ecuadorians were differently drawn into nationalist agendas, with indigenous groups becoming objects of education and 'uplift' and blacks excluded altogether from national constructs. Political geographies, too, are often grounded in race. Many political districts, for example, are constructed on assumptions that minority groups occupy, and assert the right to representation through, claims to particular spaces.

Just as race is deployed to carve out political spaces and districts, it is also used to create political constituents and groups. Through technologies like the national census, racial categories are generated and subsequently endowed with political power, a process evident in the evolution of the category 'Hispanic'. Invoked in the late nineteenth century American Southwest to insert class and nationality into the category 'Mexican', Hispanic became a way for more privileged persons of Latin American descent to claim Spanish heritage separate from poorer Mexicans. Legally identified as white but socially othered, Hispanics did not occupy the social, political, or economic status of whiteness, nor did legal whiteness halt discrimination against them throughout the twentieth century. In 1980, Hispanic became even more politically charged, and entered a wider sphere of circulation, when it appeared on the US national census. Although technically an ethnic category within whiteness, Hispanic continued to function as a racial category that engulfed all groups with Latin American origins. Straddling the American color line, Hispanic itself became a contested term, when its link to a Spanish heritage of conquest and its elision of indigenous and non Spanish speaking Latin Americans were contested through the newer category *Latino/a*.

As the complexities of Hispanic as an ethnic but profoundly racialized category show, race and ethnicity are linked but not interchangeable. Although early conceptions of race conflated cultural, phenotypic, social, and environmental characteristics and nineteenth century political theorists argued that both racial and ethnic identification would dissolve through modernization and industrialization, race and ethnicity were pulled apart conceptually in the early 1920s. Through 'modern' scientific approaches to measuring urban space and race relations, Chicago School sociologists offered ethnicity as a way to challenge both racist hysteria in the US and elsewhere and biologistic understandings of race, by highlighting race's social nature. Attention to ethnicity over race continued through the mid twentieth century, as scholars used white European immigrants' experiences to present ethnic assimilation as a universal model in which cultural differences faded over time. Groups that could not get 'beyond' ethnic identity – namely, communities of color – were criticized for continuing to embody and mobilize a group identity that should have disappeared. Although these approaches to ethnicity were grounded in efforts to challenge race as biological, they could not account for structural racism and, instead, blamed racialized groups for failing to 'improve' their lot. Thus, subsequent critical work has paid more attention to the differences that race makes in immigration, identity, and community politics and reception.

As this last example shows, race is, and has been, a powerful force in determining life chances and economic outcomes. Racial segregation in neighborhoods, schools, and other social spaces, for example, is a determining factor in graduation rates, job placement, health, toxic exposure, and other factors that influence life and death. In part because of race's power to inform life outcomes, it has been a node around which political projects from slave insurrections to white supremacy movements have been mobilized across time and space. Race as an organizing principle, though, can be a double edged sword, as it simultaneously highlights similarities among individuals and groups embracing or placed in a racial category and masks the myriad experiences of gender, sexuality, and so on that overdetermine race. For many political and social movements, then, the question has been, how can differences within racial categories be acknowledged without losing sight of race's power as a rallying point? In response, some scholars and activists have questioned race's efficacy as a lone organizing identity, arguing instead for an approach that places struggles and campaigns at the interface of systems of difference.

The Place of Race

Through the mutual constitution of race and place, race has been a fundamental aspect of how the modern, and postmodern, world has taken form socially, economically, politically, and culturally. Human geography's engagement with race, however, has been uneven across paradigms and subdisciplines. From the early twentieth century through the early 1970s, for example, geographers primarily mapped and measured 'race' through fixed categories that were assumed to be hermetically separate and rank ordered under whiteness. Although such studies used race to explain spatial differences in livelihood practices, life chances, and social features, geography as a whole was largely silent on questions of race, partly because of its invidious link to debunked approaches such as environmental determinism. By the late 1960s, however, new mapping and statistical techniques were entering human geography; and new attention was turning to race and urban issues. Responding to factors from new immigration policies across settler societies to civil rights and decolonization movements around the world, geographers returned to the topic of race, particularly in urban and social geography.

Even with this new attention, race continued to be treated as a homogeneous category whose content was obvious and easily defined. Although geographic work in the 1970s focused on race relations, for example, racial groups were understood to exist prior to social interactions, rather than to be produced through them. Methodologically, race was quantified through mappings that read social interaction from spatial distance and statistical manipulations that estimated group characteristics and trends. In the mid 1980s, however, geographers, like other social scientists, began to pay more attention to the production and contestation of racialized categories and meanings. Drawing from interdisciplinary debates about culture, geographers critiqued approaches to race as preexisting the social, political, and economic systems it was understood to contour. Instead, social constructivist works drew on racial formations theory, which emphasized the anti essentialist nature of race, the historical flexibility of racial categories, and the contested aspect of racial dynamics.

Particularly through this approach, historical geographers began to blend analyses of imperialism, colonialism, and landscape with critical attention to race. Framing racial formations as essential to the production and politics of landscape and everyday geographies, geographers such as Rich Schein examined racialized landscapes, those places where dominant understandings of race assume material form in the landscape, which subsequently informs ideas of race and a collective past, present, and future. Historical and contemporary racialized landscapes, through their 'everydayness', were understood as central to processes of racialization, the material and ideological ways that differences among people and places are translated into racial differences with lasting social, political, cultural, and economic meaning.

Although these shifts in how race has been examined in human geography have powerfully reworked understandings of race as a geographic process, they did not occur sequentially. Instead, race has been, and is, approached through a number of theoretical and methodological frameworks, including post structural and postcolonial theory, critical realism and Marxism, and positivism and humanism, as well as quantitative and qualitative methodologies. Across these perspectives, geographic scholarship tends either to 'document' racial patterns, differences, and inequalities or 'analyze' race's meanings, contradictions, and practices, with little interaction across this divide. This separation limits not only the policy and activist contributions of critical race scholarship but also the analytic purchase of descriptive studies of race and merits more attention from geographers pushing the boundaries of race scholarship in new thematic and geographic contexts.

As several geographers have stressed, geographies of race are not always easily mapped. Multicultural neighborhoods or worksites, for example, do not necessarily imply social interactions across racial boundaries; and spatial and social proximity are not interchangeable. In late nineteenth and early twentieth century southern US cities, for example, black and white residents often lived cheek by jowl but occupied separate social worlds. Geographies of race are also temporal. Because people live across social spaces (work, home, school, etc.), geographies of race ebb and flow with work and leisure activities. The geography of race in a gated community, for example, may look quite different during the day, when immigrant workers enter neighborhoods to care for property, children, and families, and resident families head to schools, downtown worksites, and consumption spaces.

In thinking about the place of race, scale, too, must be considered, as how race is understood is intimately linked to the scale at which it is considered. The clarity of race relations and categories, for example, may fade when analyses shift from broad brush claims about group interactions to everyday experiences of men and women in neighborhoods and other social spaces. Along the same lines, thinking the place of race entails thinking race across places. Whether people move across international borders or across town, the ways they are racialized and themselves understand race often shift, as they encounter new understandings and practices of race while not necessarily abandoning former systems of racialization.

Because of these mixings, some scholars suggest that border crossings produce diasporic or hybrid racial formations. Even when groups share a diasporic history, however, they may not share understandings of race. Blacks in the US and Cuba, for example, while both connected to the black Caribbean, have different experiences of racialization. In the US, blackness represents not only a structure of exploitation but also a site of resistance. In pre Revolution Cuba, however, blackness was framed as a trait to be downplayed, by not talking about race, or transcended altogether, by subsuming racial identity within national sentiment. Negotiations of race and borders become even more complex across generations, when immigrants and their children not only understand their racial identities differently but also are differently racialized vis à vis receiving societies.

As this last point shows, people transgress racial boundaries across their life paths and through their daily activities. Mixed race encounters and households, for example, problematize hermetically sealed racial categories and challenge academic preoccupations with separation and segregation. Even as racial categories become more fluid through intermarriage, multiracial census categories, and global cultural flows, however, racial boundaries may fade unevenly. In the US, for example, a black–white binary may be blurring; but a black/not black binary is not. Racial identities can also change through employment. In industrial America, European (male) immigrants initially framed as foreign and nonwhite became white as they became workers, intertwining race, nation, and masculinity and lifting immigrant workers toward whiteness on the backs of black co workers and neighbors. In Ecuador, urban em ployment can be a pathway toward whiteness for rural women and a *mestizo* identity for rural men, showing the difference that gender and place make in the links be tween race and work.

The Problems of Race Analysis

Although race is a powerful analytic for understanding the world, it is not unproblematic. Because of both the hegemony of urban North America and Britain in geo graphic scholarship and a black–white binary in framings of race, race's workings in other places and racial systems are difficult to assess on their own terms. Even as 'new' sites are brought into geographic discussions of race, such inclusion can veer toward voyeuristic focus on 'the other', leaving a normative center unexamined and reifying the link between race and 'other'. Along the same lines, at tention to 'diversity' can commodify difference from a white norm, as the perspectives, practices, and voices of persons of color 'flavor', rather than destabilize, hege monic whiteness. On a basic level, although race is central to human geographies across scales and time periods, *a priori* centering of race as the primary category of analysis or oppression can overshadow its interaction with other systems of difference. Instead of assuming, then, that race immanently orders space and inequality or that racial groups and spaces are 'naively given', scholars must deconstruct the ways that race itself becomes a key category for understanding inclusion and exclusion, in terrogating, rather than reproducing, that 'common sense' status.

Conceptually, a recurring problem of race analyses has been determining its relation to political economic forces. Although largely absent in geography, race–class debates have spanned disciplines and have been theore tically and empirically contentious, particularly vis à vis Marxism and critical race theory. Empirically, scholars have wrangled with explaining interactions between labor and race. Particularly since industrialization, em ployers have used race to divide the labor force and control workers. Studies of such divisions have reached no conclusion, however, concerning job competition, with some arguing that intraethnic competition out weighs interethnic competition and others suggesting that mutually reinforcing political, economic, and ethnic transitions push less educated workers of color away from employment opportunities. Theoretically, scholars, particularly labor historians, have queried how and why workers organized along race or class lines, arguing ei ther for race's centrality in determining economic and labor processes or political economic forces' centrality in producing racial practices.

Most importantly, although much geographic work has examined racial categories, formations, and politics as sociospatial forces, deconstructing race is not the same as challenging racism, which, unlike race, possesses a sin gularity in its ability to exploit and renew fatal power difference complexes. The term racism came into use in the 1930s to describe Nazi theories and became an ex plicit object of attack in independence movements across Africa and Asia. Racism is propped up by state structures but does not require state support. It draws upon an ideology of biological difference but can operate without it. Interpersonal, structural, and institutional discrimin ation are central to racism's workings; but discrimination and racism are not synonymous or mutually dependent. Racism can be driven by the synergy of desire and dis taste but is not simply disliking the 'Other'.

As a working definition, then, racism is present when differences rooted in linguistic, cultural, or national practices are treated as if immanent, unchanging, and embodied and used to structure various inequalities. Across space and time, racism works as a scavenger ideology in and through colonialism, industrialization, modernization, and other forces. Its insidiousness comes from the foreclosure of similarities or movement across racialized boundaries positioned as natural and eternal. One of the most tenacious forces in human interaction, racism has survived alongside the Enlightenment notions of human equality, modern ideas of progress, and multicultural celebrations of diversity. One of the most powerful forms of spatial regimentation, it has enforced where bodies may, and may not, go and who may dwell in what spaces.

Although racism works by riveting difference to fixed hierarchies that structure social, economic, political, and cultural relations across space and scales, it is a highly labile phenomenon. One of the ways that racism main tains this flexibility is through white privilege. Requiring neither acting agents nor intentional discrimination, white privilege is white individuals' accrual of social, economic, political, and cultural privileges solely by virtue of being white. Although white privilege has not received sustained attention in geography, whiteness it self has been central to geography's engagement with race. Linked to discourses of normalcy, freedom, and morality, whiteness assigns white bodies political, eco nomic, and spatial privileges, as well as the certainty of occupying the center of a racially marked system of stratification. Academic attention to whiteness surfaced in the late 1980s, although people of color have been aware of its workings since the colonial encounter. Driven by work in labor history, film studies, and literary theory in particular, whiteness's popularity as a theme is also tied to reflexive turns across the social sciences, as well as white scholars' desire to locate themselves and their own racial experiences within their scholarship.

The Possibilities of Race

If whiteness creates structural advantages and privileges for a white identity and an unacknowledged central point from which to view self, others, and society, geography, as a set of disciplinary practices and a collection of scholars, has yet to come to terms with its own institutional history of whiteness. Human geography may be a subject concerned with the diverse experiences of the world's population, but the structural racism that reproduces its whiteness creates epistemological and pedagogical blinders concerning those 'diverse experiences'. Addressing this situation, as several geographers of color have stressed, requires increasing geography's relevance for scholars and students of color, as well as integrating, without co opting, their voices into geographic work. Doing so can change not only the kinds of questions geographers ask but also the ways they understand the world. It also, however, entails a new kind of responsibility to maintaining a critical edge vis à vis understandings of race, racism, and antiracist agendas as part of broader struggles to create a more equitable world. Although unmasking the workings of whiteness can be an initial step in antiracist politics, it cannot be an end goal; and geographers must take subsequent steps to change the discipline structurally and institutionally.

Linda Peake and Audrey Kobayashi have recently argued for greater attention to geographies of racism, more antiracist work in institutional practices and teaching, and increased participation of geographers of color. Although geography has yet to reach its potential as a contributor to critical race scholarship and has ground to cover vis à vis racism and antiracist agendas, geographers are well poised, as Ruth Wilson Gilmore suggests, to interrogate race as a condition of existence and a category of analysis by examining race and racism's inner workings. Through comparative work that examines race across geographic and temporal contexts, through more attention to critical race theory and its methodological and political contributions, and through ongoing focus on how socially constructed racial categories are always embodied with serious material consequences, human geography holds much potential for future critical studies of race.

See also: Apartheid/Post-Apartheid; Difference/Politics of Difference; Ethnicity; Identity Politics; Other/Otherness; Racism and Antiracism; Whiteness.

Further Reading

Anderson, K. (1988). Cultural hegemony and the race definition process in Chinatown, Vancouver: 1880 1980. *Environment and Planning D: Society and Space* 6, 127 149.
Anderson, K. (2007). *Race and the Crisis of Humanism*. London: Routledge.
Barrett, J. and Roediger, D. (1997). Inbetween peoples: Race, nationality, and the 'new immigrant' working class. *Journal of American Ethnic History* 16, 3 44.
Bonnett, A. (1996). Constructions of 'race', place and discipline: Geographies of 'racial' identity and racism. *Ethnic and Racial Studies* 19, 864 883.
Brubaker, R. (2002). Ethnicity without groups. *Archives Europeenes de Sociologie* 43, 163 189.
Eze, E. (ed.) (1997). *Race and the Enlightenment: A Reader*. Malden, MA: Blackwell.
Fanon, F. (1967). *Black Skin, White Masks*. Markmann, C. (trans.). New York: Grove Press.
Foley, N. (1997). *The White Scourge: Mexicans, Blacks, and Poor Whites in Texas Cotton Culture*. Berkeley, CA: University of California Press.
Foote, N. (2006). Race, state and nation in early twentieth century Ecuador. *Nations and Nationalism* 12, 261 278.
Fredrickson, G. (2002). *Racism: A Short History*. Princeton, NJ: Princeton University Press.
Hartigan, J. (1999). *Racial Situations: Class Predicaments of Whiteness in Detroit*. Princeton, NJ: Princeton University Press.
Hooks, B. (1992). *Black Looks: Race and Representation*. Boston, MA: South End Press.
Mahtani, M. (2006). Challenging the ivory tower: Proposing anti racist geographies within the academy. *Gender, Place, and Culture* 13(1), 21 25.
Marable, M. (2000). The problematics of ethnic studies. In Marable, M. (ed.) *Dispatches from the Ebony Tower: Intellectuals Confront the African American Experience*, pp 243 264. New York: Columbia University Press.
Mignolo, W. (2003). The markers of race: Knowledge and the differential/colonial accumulation of meaning. *Neohelicon: Acta Comparationis Litterarum Universarum* 30(1), 89 102.
Omi, M. and Winant, H. (1994). *Racial Formations in the United States: From the 1960s to the 1990s* (2nd edn.). New York: Routledge.
Pulido, L. (2000). Rethinking environmental racism: White privilege and urban development in Southern California. *Annals of the Association of American Geographers* 90, 12 40.
Radcliffe, S. (1999). Embodying national identities: *Mestizo* men and white women in Ecuadorian racial national imaginaries. *Transactions of the Institute of British Geographers NS* 24, 213 225.
Schein, R. (2002). Race, racism, and geography. *The Professional Geographer* 54(1), 1 5.
Schein, R. (2003). Normative dimensions of landscape. In Wilson, C. & Groth, P. (eds.) *Everyday America: Cultural Landscape Studies after J.B. Jackson*, pp 199 218. Berkeley, CA: University of California Press.

Racism and Antiracism

A. Nayak, University of Newcastle, Newcastle, UK

Glossary

Antiracism A political challenge and intervention in racist policy and practice.
Race A fictitious social construct used to divide up the human population.
Racism The practice of discrimination based on cultural stereotypes and prejudice toward nation-states, ethnic groups, or individuals.
Semiotics A research method that sees signs, symbols, and images as cultural texts that carry meaning.
Whiteness A term used to describe contingent forms of race privilege that may be linked to skin color, occupation, consumption, social class status, wealth, or home ownership, among many other indicators.

Race is arguably one of the most controversial concepts in human geography. For these reasons it is often parenthesized in inverted commas, omitted from discussion, or conflated with other seemingly less contentious social terms such as 'ethnicity' or 'culture'.

This short review begins by outlining what is meant and understood by the concept of race and how it relates to racism. To understand geography's relationship to racism attention is drawn to the discipline's historical roots and the shadow cast by its imperial constitution. The article then turns to cultural expressions of racism which have been the focus of investigation in work on race and place. These geographies form part of a paradigm shift from looking at race as a knowable object to focusing upon racism as a set of social relations inscribed with power. The review then investigates the overlapping discourses of racism and religious discrimination. Following this the article critically reflects upon the enduring whiteness of contemporary human geography and its links to race privilege, before addressing the way in which antiracist geographies can be utilized to transform the theories, practice, and overall constitution of the discipline.

Introduction: Race and Racism

Broadly speaking, race is premised upon a belief in innate human difference. This derives from an abiding conviction that biological distinctions exist between social groups, delineating, for example, among 'Europeans', 'Africans', or 'Asians'.

This simplistic and highly contentious assertion further implies a level of homogeneity within these communities suggesting that race is a timeless and enduring set of characteristics passed down through our genes to successive generations. There is, however, little scientific evidence to support the idea of essential racial characteristics let alone anything approaching the existence of definitive or unique species within humankind. For this reason race can be seen as a modern myth – a human invention used to divide up and make sense of a complicated social world. In this vein, contemporary human geographers prefer to speak about the 'idea of race' rather than ascribe any particular qualities to this social category. Deconstruction remains a powerful device for imploding race categories as it is concerned with unpacking the ways in which the concept of race is made meaningful. In doing so race and ethnicity are no longer identifiable 'objects', a fixed locus for classification, but become intrinsically connected to relations of power. This entails a radical shift in geographical investigation from a focus on race toward more politicized and critical understandings of 'racism'.

Biological forms of racism were a recurring motif in imperial expansion and remain a feature of modern day genocides. The Nazi holocaust witnessed the mass execution of Jews, premised on the racist belief that they were inferior beings whose blood could only weaken the Aryan race. In Rwanda, long standing conflict between the Hutus and Tutsis – tribal communities who share many ethnic characteristics in terms of language, diet, and customs – has often centered upon blood, origins, and soil. When Belgian colonialists arrived in central Africa in 1916, they considered the Tutsis to be biologically superior, an attribute that Tutsis were keen to embrace as they achieved better jobs and privileges over the next two decades. What had once been a mixed society gave way to sporadic conflict and growing resentment which eventually culminated in the slaughter of nearly 20 000 Tutsis by Hutus in 1959 with many fleeing to the neighboring countries of Burundi, Tanzania, and Uganda. Three years later, Rwandan independence was granted when Belgium withdrew from Africa, but more recent genocide has demonstrated the power of racial discrimination and the haunting return of the colonial legacy. Similar world examples can be seen with the breakup of the former Yugoslavia, where conflict between Serbs and Croats resulted in 'ethnic cleansing' and the attempt to purge 'impure' people from the body of the nation state.

Geography and the Shadow of Empire

Over the last two centuries, the degree to which the discipline of geography has been wedded to the ideology of Empire has been nothing less than startling. So firm has been the interlacing of geographical knowledge with the imperial drive that the configuration has led scholars to rethink the 'geographical tradition' through its colonial heritage. However, one would be mistaken in believing that nineteenth century proponents of the discipline were at all coy about the imperial roots of their subject. Instead, many luminaries viewed geography as a practical subject that was intimately connected to empire building, the acquisition of territories, the extrapolation of re sources, and the exploitation of native peoples. Un doubtedly, this enabled the ideological 'mastery' of white Europeans to be secured at the expense of colonized nations.

Moreover, geography had become a prime contributor to models of racial classification. The Royal Geographical Society established in 1830 was content to sponsor a series of presentations in which the concept of racial difference was legitimated. Moreover, a number of nineteenth century geographers who subscribed to the evolutionary theories of Charles Darwin were convinced that climate had a deterministic influence upon the racial disposition of native groups. Research on 'ethno climat ology' was to make spurious links between the environ ment and populations. In alluding to the deterministic effects of temperature, whether light and physical land scapes, this work helped instigate various pseudo scien tific reports proclaiming that many native peoples were 'backward', morally unrestrained, sexually libidinous, intellectually inferior, and altogether uncivilized when contrasted with white Western Europeans. It appeared that the natives had long succumbed to the heat of the tropics. These antiquated ideas are recognized to be formidable racist constructs whose appeal ceaselessly echoes throughout the modern age. Contemporary rep resentations of Africans, Asians, aboriginals, or Arabs are frequently underpinned by discourses of danger, fear, excess, and desire.

Cultural Racisms

The shifting contours of racism witnessed in the last quarter of a century have seen a move away from bio logical forms of race superiority toward 'cultural' forms of racism. As we saw above, essentialist beliefs in race have not been entirely eradicated but the focus of popular racism in late modernity is more frequently upon cultural difference and the incompatibility of 'dif ferent cultures'. These accounts treat culture as a her metically sealed and bounded category, a marker of authenticity, rather than the more troubling, open ended, and promiscuous process it actually is.

Some of the most incisive and theoretically adept investigations of cultural racism can be traced in richly nuanced geographical accounts of race and place. For example, the spatialized representations of race and their material, symbolic, and historical effects are captured in detailed geographical media accounts of British urban uprisings in the 1980s which show how cultural racism is enacted upon the space of the inner city. In turning away from the 'object' of race, cultural geographers developed a fascination for the communication of racism through the semiotic production of race signs, where the use of particular words and phrases – 'riots', 'ghettoes', and a 'flood' of immigrants that allegedly 'swamps the nation' – is seen to inform and speak back to a broader repertoire of race thinking and race making devices. These re searchers argue that any understanding of space and place should engage with the metaphoric, metonymic, and syntagmatic aspects of cultural racism that render the category of race to life.

In investigating the contingent and competing ways in which race signs and cultural racism operate, geog raphers reveal how place and community are arbitrary and incomplete constructions. These visions have the power to include and exclude. For example, many urban areas in Britain, Australia, South Africa, and the US are deemed 'no go' areas by the police, outsiders, and the popular media, yet equally these may be safe spaces and 'home' to various ethnic minority communities. Racist constructions of these sites frequently make invidious connections between people and place, casting them as the conveyors of crime, drugs, violence, and prostitution. Such cultural forms of racism rarely consider the con nections between crime and social deprivation and the material effects that unemployment, poverty, homeless ness, or low grade living has on choice and action. In this respect, the 'ghetto' becomes a trope through which numerous racist fantasies are projected. Through this particular geographical imagination the relatively afflu ent, mainly white suburbs, are confirmed as 'ordinary', 'civilized', 'family orientated', and 'safe' spaces to inhabit.

Racism has also featured in coastal areas where im migration and asylum have become part and parcel of modern global diasporas. Racist conflict on Sydney bea ches in Australia between white Australian and Middle Eastern youth of predominantly Lebanese heritage re cently revealed the intractability of cultural racism. Al legations were made that Lebanese youth had sexually harassed white women, beaten up an Australian coast guard who tried to intervene and had generally 'taken liberties', if the participants in the regular radio 'shock jock' programs were to be believed. Rumors, scare mongering, drunken anger, and latent racism culminated in rioting with some 5000 white males seeking to 'reclaim the beach', deploying territorial slogans such as 'We grew here, you flew here'. Although historical press analysis would reveal long standing conflict between youth in the Cronulla suburb of Sydney, it is the manner in which particular incidents become racialized and 'scaled up' that is strikingly apparent. These 'out of scale' images present a monstrous Arab Other: violent, sexually un restrained, and incompatible with the 'decent' white citizens of the Australian state who have 'welcomed' these strangers to the country.

Orientalist constructions of racialized Others feature in a number of geographical studies located throughout Chinatown districts across the globe. The Orientalist production of these areas in magazines, television pro grams, and tourist guides tends to conceive of these spaces as an exotic, ornamental bauble within easy reach of white metropolitan consumers who desire a 'taste of spice'. The power of racism is particularly pervasive when it is linked to neoliberal forms of urban govern mentality. This has seen the decimation of numerous aboriginal neighborhoods in the name of progress, sig nified in the hegemonic language of regeneration, gen trification, upward mobility, business, and corporate capital. State sponsored racism of this kind operates through white norms where whiteness is made manifest in neoliberal ideas pertaining to consumption, property ownership, 'good' schools, individual choice, and family life. At the heart of these disputes is the nationalist presumption and imaginary construction of Australia as a white space. These constructions enable indigenous communities to be excised from the landscape. Cultural geography studies of crime and racism in Canada, land scape use and symbolism in Australia, neighborhood gangs, graffiti, and territory in the US, and the specter of racism in Sweden illustrate how new international geographies of race and place are locally and globally materializing.

Racism and Religion

Recent work on urban racism is extending to explore the intersections between geography, race, and religion. Racialized divisions between Catholics and Protestants in Northern Ireland, Jews and Muslims in the Middle East, or Hindus and Muslims in Kashmir are indicative of the role that religious fervor and fear can play in the making of 'community'. These issues have become especially pertinent post 9/11 as terror and geopolitical divisions have become displaced from Communist countries to Islamic nation states. Following disturbances in the English multiethnic textile towns of Bradford, Burnley, and Oldham in 2001, a number of geographers have critically discussed recent policy centered upon 'parallel lives'. This concept suggests that Muslim communities may be involved in practices of self segregation through speaking different languages and choosing to worship, live, work, and attend separate faith schools. While these issues are hotly debated, possibilities for complex geog raphies of race, class, and religion are evident. What human geographers can demonstrate is that where ethnic enclaves do exist they are rarely a matter of 'choice' but a consequence of social exclusion, material deprivation, and a response to the racism of white communities, the police, and the politics of nation state. These empirical accounts challenge the more open ended, 'global sense of place' envisioned in some of the cosmopolitan, cele bratory descriptions of the metropolis.

Qualitative research by geographers has attempted to move beyond casual policy depictions and media repre sentations of Islam by engaging with the lived experience of faith communities. In doing so this work presents a more complex picture, showing, for example, how young women from British Asian Muslim backgrounds combine Western and Islamic dress to present different, hybrid subject positions that subtly rework fixed notions of culture as bounded and complete. This research is complemented by work undertaken in the Scottish cities of Glasgow and Edinburgh on young Muslim men which reveals how different masculine subject positions are taken up in response to identity issues and can read through a lens of local, national, and global race conflict that presents Islam as Other at each of these intersecting scales. Elsewhere research in North America has shown how Muslim women were forced to alter their patterns of daily travel post 9/11 as risk and anxiety fed back into their everyday lives. It is increasingly likely that as war, risk, and global terror become salient themes in late modernity, questions pertaining to ethnicity, religious difference, and citizenship will come to the fore in new cultural geographies of racism.

Racism and the Geographies of Whiteness

If studies of racism have in large part tended to focus upon visible minorities more recent work is considering the overlapping interplay of race, place, and whiteness. Influenced by such approaches geographers have engaged in different place specific analyses of whiteness. Where the inner city has been the primary site for racial enquiry one account uses ethnographic insights to examine the geography of racist violence in suburban districts. This work demonstrates how an English skinhead gang is able to transform their neighborhood into a 'white space'. At a macro level this is partially achieved through economic transformations in the region, 'white flight' from the city and local authority housing policies aimed at dispersing ethnic minorities throughout the conurbation. However, the constitution of the estate as 'white' is further secured at a local micro political level through intimidation, harassment, and the perpetuation of racist graffiti which frequently results in complaints by minorities to public housing authorities who relocate them back within the multiethnic inner city. Another suburban case study includes a press analysis of the shootings that occurred in Columbine High School in Littleton, Colorado, USA, to reveal how the neighborhood is portrayed in the geographical imagination as a 'safe' (white) space when juxtaposed to the vividly racialized inner city 'hood' and ghetto.

The geographies of racism are also evident in South African townships where historically segregated areas, patterns, and practices are not immediately elided in the post apartheid era. The color based distinctions made between 'black', 'brown', 'white', and 'colored' peoples are informative of the ways in which race signs are readily assigned to bodies and places and the material consequences of such racial divide. A subtle reworking of this process can be seen in Sydney, Australia, where an area known as the Block came to be either ignored by estate agents and promoters of the 2000 Sydney Olympics or was discursively represented as a 'black ghetto' or 'no go zone' by the Australian media and white population. The mass gentrification that accompanied the Olympics witnessed a new version of Sydney urbanity that sought to write out indigenous communities in an attempt to present the city as akin to a sleek, modernist, white Manhattan. In this way one may argue that whiteness, gentrification, and corporate capital have the effect of producing areas like the Block as 'dead spaces' in which indigenous communities can appear 'abject bodies', haunting the contemporary postcolonial landscape. However, whiteness is a social and mutable category that is open to more critical interpretations. In a geographical study of whiteness in Southeast England another geographer has remarked how these 'escalator regions' are frequently regarded as bland, anomic 'nonplace spaces' marooned from the cosmopolitan energies of the multiethnic inner city. Here, whiteness is represented as 'lack' – a lifeless absence associated with boredom and homogeneity.

The theme of whiteness as 'lack' features in an account of 30 white women living in Southern California, close to the Mexican border. If the testimonies of these women are anything to go by whiteness often comes into being in relation to racial Otherness. Geographical studies of whiteness in nineteenth century Barbados in the new postslavery moment indicate that newly 'freedmen' and rural poor whites were cast as transient, liminal figures whose in between status meant they could not easily be absorbed into a rigid, inflexible, and simplistic black/white racial binary. The gradations of whiteness are further exemplified in other international accounts of white identity in places as diverse as China, Japan, Venezuela, and Brazil. Here, whiteness is frequently linked to money, consumption, and beauty, as paraded in American global cinema and advertising. Youth ethnographies undertaken in the postindustrial district of Northeast England further illuminate how whiteness operates through a complex circuit of local–global relations, where it is interwoven within the discursive and material histories of place relations, and comes to be embodied by different subjects in different ways. Indeed, neighborhood accounts of shopping, media, and popular culture around the world reveal how whiteness is threaded through production and consumption, wherein global commodities can come to signify whiteness and Western modernity. These dynamics can be seen in geographical research undertaken in the Indian coastal resort of Goa revealing a racially demarcated landscape in which Western rave tourists could deploy the privileges of whiteness and cosmopolitan citizenship at the expense of local tourists and inhabitants.

Geography as a White Discipline

Despite the attempts of cultural geographers to include the voices of racialized minorities into their accounts, geography has been and continues to remain, an ostensibly white discipline. Whiteness is embedded in its historical foundations, its institutional apparatus, and of course the networks of university students, teachers, and researchers that comprise the 'geographical community'. It would, then, amount to an academic conceit to presume that just because the discipline of geography is now interrogating its imperial legacy, that it can simply emerge unblinking and untarnished from the shadow of race and empire. Where feminist geographers have long contested and subverted the masculine basis of the discipline, the whiteness of geography has only recently been subject to remark. Reflecting on its inherent whiteness one geographer elegantly compares the discipline to the starched white institution of the golf club or the US Supreme Court, while others maintain that the lack of ethnic minority participants could be a reflection of the imperial legacy. The power of whiteness is then augmented in the constitution of editorial boards, promotion committees, funding bodies, grant assessors, and other such disciplinary gatekeepers. But what are the cultural effects of this overbearing whiteness, and how do they shape geographical thinking today?

We might contend that the whiteness of geography enables certain truths and norms to operate as the dominant *modus operandi*. This in turn means that other truths and knowledges are obscured or subordinated as it becomes increasingly difficult to challenge the institution from the outside as most of the teachers, students, and academics within it are white. This means that these acts of cultural knowledge production are sited within a very particular and highly partial realm of experience. The recursive privileging of white racial norms, effectively lends itself to exclusionary practices of intellectual production. This is particularly problematic for a subject that claims to be the 'world's discipline'. However, the seated position of privilege that whiteness holds does not mean that a more reflexive, antiracist politics is impossible. While some human geographers are increasingly reflexive about their whiteness others campaign to make geography a more inclusive and thereby less starkly white discipline. This suggests that though white geographers may be implicit in the production of racial norms they are active agents who can also resist and overturn these ideas. To gain further insight into how this is being achieved we can turn to some of the activist work that constitute what is becoming known as 'antiracist geographies'.

Antiracist Geographies

Antiracist geographies in their many diverse forms, seek to make a meaningful political intervention in race debates. Broadly speaking, antiracism refers to activities and practices which seek to challenge, extinguish, or ameliorate racism. Within academia this could include, for example, the establishment of antiracist policies across universities; the inclusion of work by black writers and antiracist activists on course syllabi; attempts to recruit black and minority ethnic students and teachers into the discipline; the organization of campaigns against race discrimination; the development of race awareness across the campus, and so forth. Tackling the culture of institutional racism in the workplace is a difficult but pivotal aspect of antiracist practice. For example, recent reports by the Policy Studies Institute and the Association of University Teachers (AUT) in the UK have found ethnic minority lecturers are more likely to be employed on casual fixed term contracts, have heavier teaching loads, lower paid salaries, and less chance of gaining promotion or financial increments compared to white colleagues across the University sector. In US geography departments, the disciplinary picture is equally bleak where the number of ethnic minority tenured and senior staff remains well below those figures in congruent fields such as history, sociology, and anthropology.

However, antiracist geographies are not an entirely new or novel aspect of the human geography tradition. Rather, antiracist thought and practice is part of a longer struggle with racism that calls into question the purpose of geography in the postcolonial period. An explicit example of antiracism within the discipline was seen at the Association of American Geographers (AAG) annual conference in 1971, arguably the largest and most prestigious public arena for geographical debate. Here, a number of geographers called for the US to withdraw its troops from Vietnam, a struggle which was viewed as an act of imperial conquest. As Marxist ideas spread within the discipline many US campuses became sites for political protest, which included a number of geography staff and students giving their support to the Civil Rights Movements. This antiracist zeal is captured in lambasting remarks aimed at 'armchair geographers' of this period who appeared seemingly untouched by political events and continued to restrict their analyses to the rarefied confines of academic ivory towers. What was being advocated instead was a type of 'geography for all' in which local people and indigenous communities become part of the solution, as opposed to being the objects of geographical study.

To speak of antiracist geography belies the fact that it is a marginalized subfield of human geography, held together by a tiny cluster of feminist, antiracist, and minority ethnic actors. Notwithstanding these remarks, antiracist geography operates as a live wire whose current passes through, but reaches beyond traditional academic scholarship on race. For these agents direct action through political involvement is one of the ways that geography can be used to change the everyday world in which we live. This can include involvement in campaigns that may challenge the political representations of famine, poverty, development, pollution, water scarcity, and so forth.

Antiracist geographies may also operate at different scales and in different ways. For example, the 2006 AAG conference in Chicago included a debating stream on 9/11 where left wing activists entered into a productive and thought provoking dialogue with post structuralist political geographers. Many antiracist geographers, aware of continued conflict in Iraq, also withheld their contributions from this volume until the publishers Reed Elsevier withdrew their support from arms conventions and military organizations. In this way antiracist geographers can help effect change in small but significant ways. They seek to integrate theoretical knowledge on critical race studies with a transformative agenda capable of improving race relations within and beyond the discipline. The need for critical race geographers to engage in antiracist practices can then be part of a mutually enriching process where a series of dialogues and activities are undertaken.

One area where antiracist ideas are trying to effect change is in the student curricula. When it comes to teaching geography in schools, colleges, and universities unreflective statements may actually reinforce race thinking. This can be seen in at least some of the standard geography text books where uncritical discussions of development, immigration, and mobility can be found.

Work published on behalf of the Geographical Association in the UK has attempted to redress this imbalance and provide teachers with a practical guide from which to develop meaningful antiracist geographies. This is especially important in a discipline where international fieldtrips and other 'expeditions' are seldom placed within a postcolonial framework from which an antiracist geography could transpire. Engendering an antiracist awareness has the benefits of facilitating alternative geographical imaginations and developing more critical and reflexive understandings. In schools and within higher and further education the voices of non white scholars are rarely detected. Thus, antiracist geographers seek to address racism in teaching environments and facilitate the participation of non white geographers with a view to transforming the norms of the discipline. Developing antiracist geographical teaching can challenge the whiteness of the discipline from within and any attending racist assumptions. This may entail a critical reflexivity on behalf of teachers and the production of new texts and resources that explicitly subvert the colonial legacy in the classroom.

Although antiracist geography is very much a niche area comprising a few willing campaigners and marginalized minority scholars it remains a theoretically and politically challenging space to inhabit. It reminds postcolonial geographers of the need to connect their objects of study with everyday institutional practices. It also demonstrates how a number of geographical issues concerning the environment, population growth, military conflict, diasporas, and global media can be further developed through critical understandings of race and racism rooted in notions of inequality. In this way, antiracist geography is important not only because it speaks back to social theory in new ways – to make us rethink what we do and why – but also because it informs us of how we might do geography differently in the future.

See also: Cultural Geography; East/West; Ethnicity; Hybridity; Imperialistic Geographies; Multiculturalism; Race; Self-Other; Semiotics.

Further Reading

Anderson, K. (2007). *Race and the Crisis of Humanism*. New York: Routledge.

Bonnett, A. and Nayak, A. (2003). Cultural geographies of racialization: The territory of race. In Anderson, K., Domosh, M., Pile, S. & Thrift, N. (eds.) *Handbook of Cultural Geography*, pp 300 312. London: Sage.

Jackson, P. (ed.) (1995). Languages of racism. *Maps of Meaning: An Introduction to Cultural Geography*, pp 132 154. London: Routledge.

Livingston, D. (1992). *The Geographical Tradition*. Oxford: Blackwell.

Nayak, A. (2003). *Race, Place and Globalization: Youth Cultures in a Changing World*. Oxford: Berg.

Peake, L. and Kobayashi, A. (2002). Policies and practice for an antiracist geography at the millennium. *The Professional Geographer* 54(1), 50 61.

Pred, A. (2004). *The Past Is Not Dead: Facts, Fictions, and Enduring Racial Stereotypes*. Minneapolis, MN: University of Minnesota Press.

Schein, R. (ed.) (2006). *Landscape and Race in the United States*. New York: Routledge.

Shaw, W. (2007). *Cities of Whiteness*. Carlton, VIC: Blackwell.

Relevant Websites

http://www.citizenship pieces.org.uk
Citizenship Pieces: Anti Racism, websites and teaching resources.

http://www.runnymedetrust.org/
The major UK nongovernmental organization on race and ethnicity debates.

http://www.sociosite.net/
Worldwide academic race research centres, anti Semitic and antiracism websites.

Radical Environmentalism

I. G. Simmons, University of Durham, Durham, UK

Glossary

Biomes The major divisions comprised of climate–soil–vegetation–animal communities, such as tropical forests, deciduous forests, deserts, etc. Many of these are now radically modified by human action.
Biotechnology An all-encompassing term for intervention in living matter at the molecular and submolecular levels, as in genetic engineering and modification, *in vitro* fertilization (IVF), and nanotechnology.
Dualism The separation of mind and matter in much Western philosophy. It latches onto a very deep human desire to separate everything into binary pairs.
Deep Ecology The foundation epistemology of radical environmentalism in which humans and the rest of the Earth are of equal intrinsic value.
Holism Everything is linked to everything else and must be studied and acted upon as such.
Teleology The notion (most prevalent in theology) that everything is destined toward a particular end. Anathema to Darwinists, who insist upon an open-ended destinty for living species.
Wilderness Lands where the human impress is absent or at the very least not detectable visibly.

The Concept

Because of its reliance on the findings in the field of the natural sciences labeed ecology, albeit an ecology which includes the human animal *Homo sapiens*, the notion is sometimes called radical ecology and in a term from one of its philosopher champions, deep ecology. So, in this account, radical environmentalism will be taken as co terminous with radical ecology and deep ecology.

The first principle is a metaphysics which places humans in nature and not above it. In some senses, this is also true of a shallower ecology which aims to curtail environmental degradation by reforming socioeconomic practices of a kind now usually called unsustainable. Deep ecology goes much further by asserting that people have an obligation to maintain the integrity of the eco sphere rather than use it as a set of resources, giving every species the right to exist and evolve. The formu lation of the foundations of deep ecology took place in the 1980s based especially upon the work of the Nor wegian philosopher Arne Naess (1912–). His outlook embraces a philosophy of total nondualism: the aim is the flourishing of all things, since all life is bound together and therefore the word 'life' means rivers, mountains, and ecosystems as well as human and other living organisms. Thus, radical environmentalism is a human construct, as it must necessarily be, but one which tries to move well beyond the 'shallow ecology' of most environmental re form policies and movements, in effect the internalizing of nature as Other. It aims to diminish the all pervasive and conquering role of human thought (notably mod ernity) and action that is especially characteristic of the industrial and postindustrial eras of the last 250 years, but which had their roots in the earliest centuries of the Holocene. The broad aims described above have been elaborated by many adherents, most notably by Bill Devall and George Sessions in a rapid response to Naess' philosophy and later by Carolyn Merchant.

Starting with inputs from ecological science, there is equal value to all components of ecosystems. Just because some are called dominant, it is not the case that others are less significant. In a tropical forest, for example, the trees are dominant but they could not flourish without the mycorrhizal soil fungi that recycle nutrients, and none of this would happen at all without regular rainfall. However, this system is seen as having a value in itself (an intrinsic value), independent of the human use of the natural world. So the richness and diversity of life forms is part of that intrinsic value, and humans have no right to reduce those qualities except to satisfy vital needs, just as a tiger temporarily diminishes the genetic variety of a deer population by eating one of its members. This type of need is sharply differentiated from the twenty first century consumerism which addresses itself to demands which can never be satisfied, since a whole industry is devoted to creating new desires, especially through positional goods. Yet, no discussion of human demands upon the rest of the planet can ignore the number of humans. Though it is customary to point out the 80:20 division of the Third and First World components of total populations, radical environmentalists point out the disproportionate demands made by the rich, as well as the environmental degradation caused by the poor in the absence of other choices. However, they depart from many feminists and social commentators in commending programs of reduction of population growth rates and indeed the controlled reduction of some human numbers. The alternative, it is argued, is not only an attack on the intrinsic value of nature but also a vulnerability to dis ease, famine, and warfare – a basically neo Malthusian position which in radical environmentalism mandates a smaller human population. This leads directly to the realization that current levels of human manipulation of the nonhuman world transgress both the rules for eco logical resilience and those for the proper valuation of nature. Changes in attitudes and policies will doubtlessly affect the structures, agencies, and performativities of much of the world economy: not for nothing is the term 'radical' used. The thrust would be to replace growth with notions of stasis and equilibrium (here there is a link with some of the ideas of sustainability), of quantity with quality, and of living on the Earth with living with the Earth and caring for it. Inevitably, suggestions, for ex ample, zones of intensive use, low intensity use, and human absence conflict with many development policies as now applied. Lastly, Naess and his exegetes command that believers should proselytize by personal direct ac tion, even where this opens the activist to accusations of inconsistency of lifestyle, for example.

Precursors and Inputs of Radical Environmentalism

No idea springs from nowhere: its appearance on the scene may be the first appearance in print, for instance, of thoughts prevalent among a small band of people, or when a society adopts literacy as its main transmission mode for culture instead of oral communication. In this section, some of the main strands of precursor thought are considered in their historical context, that is, various times before the 1980s.

Distant Ancestors

In some of the more romantic discussions of environ mentalism, preindustrial societies are regarded very fa vorably. They are praised for their low impact upon the natural environment, their intimate knowledge of fauna and flora, and their general simplicity of lifestyles, with special reference to low levels of material possessions. Hunter gatherer communities are especially thought to have lived in a spiritualized world of virtual continuity between hunters and prey, where salmon, for instance, might be regarded as fellow humans who just happened to have put on salmon coats temporarily. Further, this spirit world was easily accessed by a local shaman whose knowledge was at the service of the community's survival. Just how far these interpretations of preindustrial life ways are 'real' and how much they are a feature of recent longings is difficult to discern. Some may indeed come from simply low population densities: many Plains In dians were wasteful of buffalo but the animals' numbers were such that only with the advent of horse and rifle were they dented and only with European levels of en joyment of extirpation (of other humans, too) were they made virtually extinct. Environmental tenderness does not seem to have prevented many precontact groups to have been very cruel to each other when resources were scarce. It is especially important to remember that pre sent knowledge of such groups is mostly filtered through early European contacts such as traders, missionaries, or colonial administrators; even anthropologists have been known to find what they were looking for.

A second category of distant voices, but preserved in their entirety unlike aboriginal humans groups, are writ ings which postulate entirely new and usually well con trolled sets of human uses of the Earth. One strand in this is fictitious, where a central protagonist is cast into a different society or environment: Daniel Defoe's *Robinson Crusoe* is generally kind toward a would be slave trader who lives very simply and declines to kill known cannibals who have not, as it happens, tried to kill him. More complex constructions also exist: in Thomas More's *Uto pia*, there is likewise an emphasis on simple living, though in this case administrators are given the best food, as in today's universities. There is a wider range of thought in Francis Bacon's *New Atlantis* of 1627, where methods of dealing with inventions, foreign ideas, and goods were strictly controlled by the state with the aim of improving the material standards of the population. Scientific en deavor was praised, but every new 'invention' had to go through a strict process of scrutiny before it was adopted.

> We have three that bend themselves, looking into the experiments of their fellows, and cast about how to draw out of them things of use and practise for man's life, and knowledge, as well for works as for plain demonstration of causes, means of natural divinations, and the easy and clear discovery of the virtues and parts of bodies. These we call Dowry men or Benefactors.
> Then after divers meetings and consults of our whole number, to consider of the former labours and col lections, we have three that take care, out of them, to direct new experiments, of a higher light, more pene trating into nature than the former. These we call Lamps.
> (Francis Bacon, 1627)

A number of other utopias have been written but few come at all close to describing features of nature or so ciety that seem to bear upon today's concerns. In fact, many clearly shy away from such matters and are more simply satirical of social manners, invent creatures that somehow missed the chance to evolve, or simply give some people a good (less often bad, though Jules Verne even made Captain Nemo disappear eventually as a re sult of volcanic eruption) time. However, the concept of using an invented society and environment to convey strong opinions about humanity and its attitudes to the Other has persisted: William Shakespeare's *The Tempest* (dating probably about 1603–11) has something to say to modern environmentalists of many varieties.

In terms of influence carried forward, however, the account of his sojourn at Walden Pond (near Concord MA) by Henry David Thoreau (1817–62) is unmissable. In 1845, Thoreau went to live in a small hut in the second growth woods about 3 km from the town. This was not a wilderness, for he could hear the trains from the nearby railroad; he did though travel to northern Maine, which was far less developed and where he needed an Indian guide. He stayed for 2 years at Walden and published his philosophical–naturalist–memoir book in 1854. This has become a touchstone for North American environmentalist thought of all shades, since it extols the virtues of simplicity of living and harmony with the land and its creatures in an environment where not only were they disappearing but also the extent of controls and taxes imposed on the individual was fast increasing. He advocated civil disobedience to unjust impositions, thus forming a link to the obligations of radical environmentalists as set out by Arne Naess. His insistence (rather diluted nowadays by the knowledge of his proximity to the grocery store) on the unity of things was a feature of his fellow transcendentalist philosopher Ralph Waldo Emerson (1803–32) whose poem 'Blight' carries resonance in any geography field class:

> But these young scholars who invade our hills,
> Bold as the engineer who fells the wood,
> And travelling often in the cut he makes,
> Love not the flower they pluck, and know it not,
> And all their botany is Latin names.

More Recent Antecedents

The period after 1945 is taken as 'recent' in the present context. It can be divided into two main subdivisions: a 'coalescing period' of the 1950s and 1960s when some ideas and actions came together in unitary frameworks or particular places, and the key period of the 1970s and 1980s when the actual formulation described in the early paragraphs of this article took place.

A link with the 'distant ancestors' is provided by the name of Walden. This was the key element in the title, *Walden Two*, of a successor to the utopias of the past written by a behavioral psychologist, B. F. Skinner (1904–90). It relates the methods and outcomes of a community where everybody is happy. As with Thoreau, low consumption, happy and deep social relationships, satisfying work, and abundant leisure are central. Skinner goes further and suggests forms of behavioral conditioning which bring about such conditions. He believed that environment and genetics determined behavior, but that they might be modified by positive reinforcement of the desirable traits. Its relevance to environmentalism comes in the means by which human societies might migrate from the mythology of everlasting plenty based on fossil fuels to a leaner and more thrifty set of ways: should this come from government imposed restrictions (like taxes) or are there ways in which a society can 'green' itself as it were from within? Not surprisingly, Skinner's work remains controversial. The continued importance of fiction in the history of radical environmentalism is shown by the utopia created by Aldous Huxley (1894–1963) in his last novel, *Island*. In contrast to his dystopian *Brave New World*, his island paradise of Pala (a bit like a combination of Bali and Sri Lanka, perhaps) has all the virtues. There is low impact living, happy relationships, and electricity from hydro schemes, along with no harm drugs, easy contraception, and lots of recreational sex. It all falls apart with the discovery and exploitation of oil. Given an often negative attitude to technology, it is refreshing to remember Nigel Calder's vision of a world in which the economy and the people were kept under plexiglass domes and the ecology flourished more or less untouched outside. Algal tanks and cell cultures (only the fillet steak and not the horns) provided nutrition.

During these years, most professional philosophers were not concerned with environmental matters and the one who came into focus only in the years well after World War II, Martin Heidegger, was German and partly because it transpired that he had been a member (and for a short while a keen one) of the Nazi Party. In fact, the work of Martin Heidegger (1889–1976) is in a sense *post hoc* if the 1980s were the time of major crystallizations of radical environmentalism, with Michael Zimmerman's highly relevant assessment coming in 1990. But it makes more sense to consider it at this point as part of a developing *Zeitgeist* that led to the formulations of the 1980s. Heidegger was a highly romantic German who wanted only to live in the Black Forest and who was deeply suspicious of industry, speed, and the drive to control. He argued that philosophy's abstract trends ever since Plato had produced an all pervasive anthropocentrism, that Western societies had ignored meditative trends in Asian thought which wanted to let things alone (as in Taoism's *wu wei*), and that we were isolated from living with alertness to natural processes in the 'natural' region of our birth and upbringing. All things, therefore, are disclosed to us as resources for use rather than as allowed to manifest themselves according to their own authentic and evolutionary possibilities. By the time Heidegger's manipulation of language (in which he was something of a precursor of French postmodernists like Jacques Derrida) has found its way into English, it becomes easier to point to Nazi Party membership in the 1930s as a reason for ignoring his potential. 'He joined a fascist party and his work is somewhat obscure, therefore he and radical environmentalists which take up his ideas are fascist', is a thought process sometimes employed, though deeper thinkers find problems with the work itself. But Heidegger's rejection of modernity and the adoption of older and apparently softer ways cannot be ignored as an input to unfolding radical ideas in which humans' main role is to care for the Earth.

Interestingly, there is a less contentious (if less well known) parallel in Naess' thinking, which highlights the role of the wild and undeveloped in a way which has a long tradition in Norwegian thought and culture. Such a set of attitudes found its echo in a seminal piece of legislation in Lyndon Johnson's presidency of the USA, the Wilderness Act of 1964 (Pub. L. 88–577), which recognized the undeveloped nature of many areas of public land and the value of keeping watersheds intact, protecting rare species, and providing for a limited amount of public access for recreation. At the time, these areas were seen as untouched by human hand, since there was a general ignorance of the role of native peoples in manipulating forests and grasslands, especially with the use of fire. In spite of depredations sanctioned by more instrumentalist governments, the national system includes about 428 000 km^2 of federal land and water. Most recreations are forbidden to use the internal combustion engine, and so these areas conform to notions of environmentalism that wish to sequester fossil fuel driven systems from the rest of the planet.

The synonym of deep ecology is a reminder of perhaps the most important component of all, namely, the science of ecology. The term itself belongs to the later nineteenth century and was initially focused largely on the relations of an individual species of plant or animal with its environment. It developed a greater community sense as time went on and in the early 1950s was much organized around ideas of chronological change of communities of plants and animals (now designated by the term 'ecosystem') from relatively simple colonizing communities (as of bare rock following glaciation or open fresh water bodies) to complex mature systems, called climaxes which were deemed to be in equilibrium with the regional climate and to fit into a worldwide system of large scale biomes, such as temperate forests, desert, and taiga. (A few atlases still print maps of 'world vegetation' of that kind, thus perpetuating the role of fiction in environmental writing.) The more subtle ecologists perceived that far from being untouched by humans, many ecosystems had their temporal paths diverted by humans actions, were stopped short of their potential climax state, or even totally replaced by another community, as a forest might give way to a cornfield. The latter was still an ecosystem, for it fixed solar energy via photosynthesis, took up plant nutrients and water from the soil, and housed animals. So the economic concept of productivity might be applied in terms of the amount of unit product per unit area per unit time, in this case, expressed in SI units as dry weight of plant or animal tissue from a given area over a given time, as in g m^2 a (grammes per square metre per year). In general, climax communities had a higher productivity than early stage ones and were thought to be more resilient to perturbation.

As detailed ecological research progressed through the later decades of the twentieth century, refinements were made to these foundational ideas. An ecosystem might at any stage, for instance, have a key species whose removal would cause a catastrophic shift in the nature of the system; not all systems proceeded smoothly to a climax, and in any case, since Holocene, climates had constantly been shifting, and the concept of equilibrium was something of a chimera. There was, too, chance and contingency within systems: two comparable systems starting from an initially similar state might easily develop along different paths. Modern texts in ecology tend to stress the different levels of scale which must be scrutinized. But in spite of all this untidiness, it looked as if everything was connected to everything else, one way or another. Even though difficult to handle in scientific work, the principle of holism, as defined by Jan Smuts in 1926 as the trend to form wholes that are greater than the sum of the individual parts seemed to apply. If true, then ecosystems might exhibit holistic behavior, and so the classic scientific methodology of studying ever smaller and simpler component systems was at worst epistemologically flawed and at best very definitely limited in its findings and applicability.

In one direction, the merging holistic trend in ecology led to a controversial model of the whole planet and its evolution, called the Gaia hypothesis by its founder, the analytical chemist James Lovelock (1919–). A barebones summary must include the ideas that the whole planet functions as a superorganism and achieves a level of homeostasis through complicated feedback loops. Without life, however, the planet would be too hot for organic forms as they are known: the whole system has in the course of time evolved to support life, including but never confined to human life. Lovelock argues that current levels of human impact, including forest clearance, intensive agriculture, and the emission of 'greenhouse gases' threaten key feedbacks in the Gaian system and one result may well be an Earth unsuitable for the human species or at the best a very small number of them. For him, a rapid move from fossil fuel use is absolutely crucial. Detractors were never in short supply, especially those who thought that Darwinian selection only took place at the species level, and those who saw in the model a kind of teleological progress which is incompatible with Darwinian ideas. The discovery of the biochemistry of some of the feedback mechanisms (such as ocean temperatures, phytoplankton production, aerosol density, cloud formation, and downwind rainfall) has much strengthened the case and a thrifty version of the hypothesis (it has attracted a mystical and New Age following) is not usually contested now by scientists.

Lovelock’s warning that only a small percentage of the human species might survive rapid global warning harks back to the eighteenth century prophet Thomas Malthus, whose basic proposition that rates of human reproduction could always outrun their subsistence base has often been disapproved (most notably by industrialization) but never disproved. Its revival, labeled ‘neo-Malthusianism’ was strong in the 1950s and 1960s when world population was increasing by about 2.0% per annum, which meant a doubling time of around 30 years. So ecologists were happy to support moves at many scales to allow family size to be limited, and Paul Ehrlich gained the status of a global guru on the basis of one paperback book. Initial emphases on Third World populations were complemented by the highlighting of the fact that it was the consumption patterns of the First World that drove a great deal of environmental change, especially in a post 1950 globalizing world economy. The idea that ‘growth’ and ‘progress’ were synonymous was challenged by the publication of the early computer models in this field, which were interpreted by their makers in 1972 as showing ‘limits to growth’ beyond which the Malthusian constraints were likely to impinge. They challenged much conventional thinking, especially that of economists, and have never quite disappeared from the intellectual scene, and might indeed be about due for a rival in the light of current carbon dioxide (CO_2) modeling.

Just as the influence of Norwegian culture can be noted, there was a concatenation of culture and people in the development of radical environmentalism, in this case on the west coast of North America. A series of seminars on resources was held at the University of California at Berkeley and published in 1967. It came to the notice of activists in the Sierra Club, then a mountaineering cum campaigning against development group in the mountains of the Western Cordillera and took off into a major antigrowth movement, which was a brave thing to do in boomerist California at that time, except that the ‘alternative’ people were much attracted to the ideas and left for rural communes in large numbers, leaving the urban middle classes to make occasional trips to what were then voluntarily manned weekend only recycling facilities in otherwise unused parking lots. Whenever a non-instrumentalist viewpoint surfaces in today’s North America, then it is the grandchild of those processes and that group, whose lasting printed views can be found in the publications of Daniel Luten.

Contraries

What emerges from the crystallizations of the 1970s and 1980s is a pluralistic set of ideas, drawing on many sources: indeed the term ‘interdisciplinary’ could be properly applied. But given its variety, it has been easy for those of different views to mount a considerable weight of opposition. Intellectual resistance has tended to focus on particular components: on the models of future population–resource–environment relations, for example, or on the apparent abandonment of the conquest of nature implied by attribution of intrinsic value. Because parts at least of radical environmentalism can be identified with more recent concepts of ‘sustainability’, those who think of that term as a mere piece of rhetoric tend to argue against its apparent parentage as well. Sustainability can be seen as basically derived from concepts of equilibrium in ecology and population ecology, which have been shown to have a shaky foundation.

In the world of agency and action, however, the most important counterarguments have simply been pragmatic: even if we wanted to live at such a low impact level, how could we get from A to B, let alone Ω? Are not the politics of radical environmentalism at best elitist and designed to pull up the ladder after the rich have got to an acceptable (to them) level of wealth? At worst, is it not a basically fascist program, since its demands could only be achieved by coercion, in the areas of resource consumption (energy is the obvious example) and population size, for example? Given the diversity of the world and its people, is this not a ‘one size fits all’ approach which nobody much will adopt in a postindustrial era where individualism is supreme, to the point of eclipsing most communal approaches to any issue?

Perhaps history may be no guide to the future. Hunter gatherers at 10 000 BC would have said they were in a sustainable relation with nature, yet within 1000 years they were well on the way to the margins in all senses; solar powered agriculturalist in AD 1750 would have said much the same, yet within 100 years their relations with nature had mostly changed irreversibly. If from those changes it is deduced that technology is the main driver of human–environmental relations, then who knows the outcome of the biotechnologies and the nanotechnology (to name but two instances) under development today? History also provides a stance which wraps up many of the objections to radical environmentalism: that it aims at a utopian condition. Isaiah Berlin pointed out that those who espoused utopias were liable to tread all over everybody who opposed them and that it was better to muddle through in a series of temporary compromises and fudges of which radical environmentalism is certainly not one.

Outward Influence

No idea of consequence is, as it were, an island unto itself. It resonates through both the world of ideas and dissemination in academia, in think tanks, and among special advisers: there are arcane papers in specialist journals as demanded by Research Assessments and there are cobbled together TV programs on channels with high numbers. If it has legs, so to speak, then it or some version of it may percolate into the world of action, most often not only at various levels of government but also in corporate policy. They are rarely separate in the way they have to be written about, but more like temporary crystallizations in the flow of energy, matter, and ideas that encompass this whole concept.

Influences of a political kind

John Maynard Keynes' famous dictum:

> … the ideas of economists and political philosophers, both when they are right and when they are wrong, are more powerful than is commonly understood. Indeed the world is ruled by little else. Practical men, who believe themselves to be quite exempt from any intellectual in fluence, are usually the slaves of some defunct economist. (John Maynard Keynes, 1936)

The idea is likely to be transferable knowledge. It is difficult to trace any transition of the core ideas of radical environmentalism into today's governmental policies, for example, though much less so with the thrust of non governmental organization (NGO) advocacy in organ izations such as Greenpeace, FoE or EarthFirst! It seems incontestable that reform environmentalism of the kind that diminishes toxic pollutants on a regional scale, or discourages landfills by installing recycling plants, or tries to protect a rare species by declaring nature reserves is almost universal and has part of its origin in an ecology based environmentalism or else in a romantic scenery preservation outlook, but the infinitely more thorough reforms demanded by 'deep green' thinking seems a long way off and likely buried in the landfill with a lot of other unread and unrecycled papers. However, if some recent investigations of a scientific kind allied to reports from influential economists and placed in an ecosystem con cept are a true example, then their very universality and large (though not complete) measure of agreement on certain social actions is a step toward radical environ mentalism which would not have been foreseen 20 years ago. The cockpit is of course global climatic change and its likely origin in the emissions due to the industrial phase of the human use of the Earth. The start can be seen in the lead up to the Montreal Protocol of 1989 which concerned the thinning of polar ozone layers by chlorofluorocarbons (CFCs) and then the setting up of the Intergovernmental Panel on Climatic Change (IPCC) and its influential reports, especially those of 2001 and 2007. The predominantly science based tenor of this work was supplemented by an economist's assessment of the twenty first century trends and the whole taken up into a foregrounding of the concept of the ecological footprint as an integrating measure of how much people's 'pressure' on the Earth's ecosystems exceeded (or not, with some poor countries) the likely capacity of those systems to sustain the processes to which they were subjected. The calculation of footprint in general terms segued quite quickly into 'carbon footprint' as the role of CO_2 from all kinds of human activities became more explicit: the role of the Internet in disseminating views and providing personal carbon calculators in the first decade of the twenty first century will be a useful PhD topic in the future.

Few commentators foresaw the rapid explosion of interest in carbon content reduction after about 2006. It can scarcely be said to stem directly from teaching of human geography at tertiary level in the preceding few decades, where the remains of the post structuralist schools' view that any metaview of the world was an illusion was still a topic of debate. The preparation of the ground for acceptance of the radical notion that the West and large developing nations such as China and India ought to use a lot less energy can be credited to teaching in climate dynamics, climate history, and environmental sciences at all ages from 2–22 years. There was in a sense a repetition of the situation in the 1970s when there was an immense surge of concern about 'the environment', mostly at a reform rather than a radical level. But geography as a discipline was much more concerned with numbers and the possibility of laws (a kind of physics envy) and largely turned its back in spite of attempts by the present author among others to bring it to the pro fession's attention. This time, the back may not be turned but the input into the thinking has been limited. The current concern is underlain by the work of the atmos pheric modelers, chemists, and economists, and the geographers are largely wondering about the local and regional impacts that may be forced by climatic change or wandering in a free form aerosol of detached theory reminiscent of Anthony Gormley's *Blind Light* at Tate Modern in 2007. Human geography can be found in the complexities of many interactions but look at the TV after the next IPCC report is published: who gets inter viewed? Perhaps we are looking in the wrong place for the wrong thing: geography's contribution may come in the form of a viewpoint which goes beyond the present divisions into a more synthesized realm.

Influences of an Intellectual Kind

In this tome, the relations with recent human geography must dominate. Though geographers in the English using world talk a great deal about environment, and although the combination of the natural and human worlds might seem to be central to their discipline's epistemology, the signs of engagement with radical environmentalism as discussed above are not very thick on the ground. That is to say, that they [we] are careful to include the place of nature in any extended discussion at research level, and will indeed likely emphasize the necessity of careful treatment of environment in any broader methodological stance. Yet the noninstrumental, intrinsic equality standpoint as set out by Naess and his followers has made little progress. This theme was explored in depth in the 1990s by Peter Beaumont and Chris Philo, and picks up on more narrowly focused studies which relate to wider concerns about the global environment (as it affects humans, of course) and the way global concerns might be better studied within the academy, given that almost everybody has a contribution to make toward a better *modus vivendi* between peoples and their surroundings. From North American geographical sources, however, the emphasis is on emplacing the discipline of geography more firmly in an institutional sense, whereas British empiricism of the David Stoddart kind is less concerned with what is said, than to whom it might be said, in a practical sense. The notion of human–nature hybridity, which may entwine nature, culture, and space (in its formulation by Sarah Whatmore) is a further stage toward the integration that so many have found elusive. Yet in most outworkings of such a view, then the dominant gene, as it were, is bound to be the cultural one except in the case of the most resistant pieces of natural world, as with the 2004 tsunami in Asia. Further, the metaphor carries with it a commitment to latitudinal studies (for hybrids needs to undergo genetic change at every new interfacing) and may also need rethinking in the light of new knowledge of the genome and the way it can be influenced during the lifetime of its carrier organism, the field known as epigenetics. This perhaps takes us further toward the idea of networks which embrace all entities as actors with agencies (a notion not confined to the organic) that Bruno Latour dangled in front of the social sciences and even maybe to the edge of the 'Jewel Net of Indra' conceived by Hua yen Buddhism, where every jewel is reflected in every other jewel at every other intersection in the universe and so the precious Western notion of $A \rightarrow B$ causality is no longer relevant. In geography, as everywhere else in academia, needless to say, the first filter is the human mind and the models carried with it.

Environmental ethics took off in the 1980s and has become an established field centering around the question of whether humans ought to alter the rest of the planet. It looks at questions of ontology – what there is (that is, 'real') as distinct from what there is not – and epistemology – how knowledge about the reality is constructed. Its relevance to geography cannot therefore be overlooked. Most physical geographers have looked for a language which can unite their constructions of nature with the socially focused constructions of the human geographers, with many of the latter regarding any tendency to binary oppositions as an artifact of categories such as scientific and not scientific, for example, or interpretive and noninterpretive. Thus, with Gillian Rose, the first of each of those binary pairs relates to a masculine and emotion free set of approaches, with the opposite construed as essentially female. Rather than trying to adjudicate between any two sets of approaches, James Proctor wants to retain the tensions inherent in a discipline that covers so much ground, both phenomenologically and epistemologically, to keep alive the search for an authentic way for humans to live as Earth dwellers. Readers of (or more likely, about) Heidegger will recognize some of the language that is being used, though even he would not have been prepared for the debates in the 1990s and early 2000s in New Zealand and Austria about giving full legal rights to nonhuman primates such as gorillas and chimpanzees.

The other social and natural sciences must, unhappily, get short shrift here. Without doubt, all of them have taken the environment seriously and have developed perspectives upon it. It is though too much to expect anthropology, economics, and sociology, let alone political science, to adopt a viewpoint which places humanity in an egalitarian relation with the rest of the cosmos. The humanities have greater freedom in this regard, where many of these questions (and indeed the language in which they are framed) is in religious metaphysics. Human domination is not a given at that level of epistemology though it is easy to find examples of texts which confirm it or historical processes egged on by the clerisy and their writings. This is especially so if belief systems such as Leninism–Marxism are included, which took to extremes the tenet that manipulation of the nonhuman world was fundamentally necessary, including the Maoist slogan that with the *Little Red Book* the cadres would conquer nature. This type of approach seems to devolve around questions of authority and trust, which might on the one hand be pulled between ecology science, and religious/political authority figures on the other, as Proctor has suggested for the USA. There might of course be a third way, in which the small screen of TV and PC relay authority in the roles held in the past by print media and men in black frocks.

Summary

It would be rash to assert that radical environmentalism has become emplaced via the anxieties of climatic change. In most of the material offered for public consumption, there is an emphasis on how life will be difficult for humans, with a sideways nod toward the effects on biodiversity and the fate of rare species, and occasionally of familiar ones. Thus, the myths of finding a technological solution are a place of frequent resort in popular media. There are deeper undercurrents never theless which relate to some of the themes of environ mentalism. One major motif is the holism implicit in calls for major changes in the human–environment relation. The bulk of the emissions may come from only a few countries, but the effects will be found in all: the sports utility vehicle (SUV) in Iowa helps to drown the home of an Andaman islander. Yet, this holism and the need for a concerted international effort to change, with mutual coercion mutually agreed upon, and a set of common understandings, is in fierce tension with today's cultural outworkings of individualism, in which choices are wherever possible shifted to the individual. The ability of the individual then to screen out anything unpleasant becomes paramount: 'my news' may well not include the extra fires, floods, and hurricanes. In such a cultural dy namic, relativism seems to be dominant and perhaps above all that belief is inimical to the fundamental no tions at the heart of radical environmentalism.

See also: Anarchism/Anarchist Geography; Ecology; Nature-Culture; Sustainability.

Further Reading

Beaumont, P. and Philo, C. (2004). Environmentalism and geography. The great debate? In Mathews, J. A. & Herbert, D. T. (eds.) *Unifying Geography: Common Heritage, Shared Future*. London: Routledge.

Bruggemeier, F. J., Cioc, M. and Zeller, T. (eds.) (2005). *How Green Were the Nazis? Nature*, *Environment and Nation in the Third Reich*, pp 171 203. Athens, OH: Ohio University Press.

Calder, N. (1967). *The Environment Game*. London: Secker and Warburg.

Cook, F. H. (1977). *Hua yen Buddhism. The Jewel Net of Indra*. University Park, PA: Pennsylvania State University Press.

Devall, W. and Sessions, G. (eds.) (1985). *Deep Ecology. Living as if Nature Mattered*. Salt Lake City, UT: Peregrine Smith Books.

Ehrlich, P. (1968). *The Population Bomb*. New York: Ballantyne Books.

Huxley, A. (1962). *Island*. London: Chatto and Windus.

Latour, B. (1993). *We Have Never Been Modern* Porter, C (transl.). New York: Harvester Wheatsheaf.

Liverman, D. M. (1999). Geography and the global environment. *Annals of the Association of American Geographers* 89, 107 120.

Lovelock, J. (2005). *Gaia: A New Look at Life on Earth*. Oxford: Oxford University Press.

Lovelock, J. (2006). *The Revenge of Gaia: Earth's Climate in Crisis and the Fate of Humanity*. New York: Basic Books.

Luten, D. B. (1986). In Vale, T. R. (ed.) *Progress against Growth: D. B. Luten on the American Landscape*. New York: Guilford Press.

Manguel, A. and Guadulpi, G. (1999). *The Dictionary of Imaginary Places*. London: Bloomsbury.

Meadows, D. H., Meadows, D. L., Randers, D. L. and Behrens, W. W. (1972). *The Limits to Growth*. New York: Potomac Associates.

Meadows, D. H., Meadows, D. L. and Randers, D. L. (1992). *Beyond the Limits*. Mills, VT: Chelsea Green Publications.

Merchant, C. (2005). *Radical Ecology. The Search For a Livable World* (2nd edn.). London: Routledge.

Naess, A. (1973). The shallow and the deep, long range ecology movements. *Inquiry* 16, 95 100.

Naess, A. (1989). *Ecology, Community and Lifestyle: Outline of an Ecosophy* Rothenburg, D. (ed.) (transl.). Cambridge: Cambridge University Press.

Parsons, J. J. and Ciriacy Wantrup, S. V. (eds.) (1967). *Natural Resources: Quality and Quantity*. Berkeley, CA: University of California Press.

Proctor, J. D. (1998). Geography, paradox and environmental ethics. *Progress in Human Geography* 22, 234 255.

Proctor, J. D. (2006). Religion as trust in authority: Theocracy and ecology in the United States. *Annals of the Association of American Geographers* 96, 188 196.

Reed, P. and Rothenberg, D. (eds.) (1993). *Wisdom in the Open Air: The Norwegian Roots of Deep Ecology*. Minneapolis, MN: University of Minnesota Press.

Ridley, M. (2003). *Nature via Nuture. Genes, Experience and What Makes Us Human*. London: HarperCollins.

Rohkramer, T. (2005). Martin Heidegger, national socialism and environmentalism. In *How Green Were the Nazis? Nature, Environment and Nation in the Third Reich*, pp 171 203. Athens, OH: Ohio University Press.

Rose, G. (1993). *Feminism and Geography: The Limits of Geographical Knowledge*. Minneapolis, MN: University of Minnesota Press.

Simmons, I. G. (1991). No rush to grow green? *Area* 22, 384 387.

Simmons, I. G. (1995). Green geography: An evolving recipe. *Geography* 80, 1 9.

Skinner, B. F. (1948). *Walden Two*. New York: Macmillan (Reprinted with a new introduction by Macmillan in 1976).

Stoddart, D. R. (1987). To claim the higher ground: Geography for the end of the century. *Transactions of the Institute of British Geographers* 12, 327 336.

Thoreau, H. D. (1854). *Walden*; *or*, *Life in the Woods* (Available in many modern editions and online, for example, at http://www.gutenberg.org/etext/205 or http://publicliterature.org/books/walden/xaa.php. Accessed in Aug. 2008).

Turner, B. L. (2000). Contested identities: Human environment geography and disciplinary implications in a restructuring academy. *Annals of the Association of American Geographers* 90, 52 74.

Whatmore, S. (2002). *Hybrid Geographies: Natures, Cultures, Spaces*. London: Sage.

Zimmerer, K. S. (2000). The reworking of conservation geographies: Nonequilibrium landscapes and nature society hybrids. *Annals of the Association of American Geographers* 90, 365 369.

Zimmerman, M. E. (1990). *Heidegger's Confrontation with Modernity: Technology, Politics, Art*. Indianapolis, IN: University of Indiana Press.

Zimmerman, M. E. (1994). *Contesting Earth's Future. Radical Ecology and Postmodernity*. Berkeley, CA: University of California Press.

Relevant Websites

http://www.resurgence.org
Resurgence.

Radical Geography

R. Peet, Clark University, Worcester, MA, USA

Glossary

Antipode Radical geography journal started in 1969 at Clark University in Massachusetts to publish socially relevant and critical leftist articles.

Marxist Geography Analysis of spatial and environmental questions informed by historical materialism.

Political Ecology Study of environmental issues informed by political economic concepts.

Political Economy Analysis using economic concepts that are openly political rather than scientifically neutral.

Production of Nature The notion that nature has been profoundly altered by human activity.

Spatial Dialectics Conception of space that sees activities relating in contradictory ways.

Spatial Fetishism Notion that space in and of itself is a causal agent criticized in the early years of radical geography to make way for alternative ideas about social causation.

Uneven Development The idea that the activities embedded in space develop unevenly rather than evenly because of contradictions among them.

Introduction

In the late 1960s radical movements were formed by political activists in several social sciences, including geography, as parts of a broader response to inequality, racism, sexism, environmental destruction, and the Vietnam War. The dramatic events of the time, such as the burning of large sections of cities in the Western world, student–worker uprisings in Paris in 1968, massive anti war protest actions, and the formation of a radical ('hippie') culture, made the concerns of the traditional discipline of geography, for all its quantitative update, and even its humanistic concern, seem socially and politically irrelevant, especially to radicalized students and junior faculty members. Up to that time, the discipline of geography had dealt with safe, obscure, even quaint topics, like barn types, wine districts, or shopping behavior. However, starting in the mid to late 1960s, articles dealing with more 'socially relevant' geographical topics began to appear in the discipline's mainstream journals. In 1969, *Antipode: A Radical Journal of Geography* was founded at Clark University in Worcester, Massachusetts, specifically to publish the new kind of work. The term 'a radical journal' was meant to indicate a broad array of viewpoints toward a transformed geography. Thus early issues of the journal dealt with urban and regional poverty, discrimination against women and minority groups, unequal access to social services, advocacy planning, Third World underdevelopment, and other, similarly political, topics. Only gradually, and with reluctance, did socially relevant radical geography turn into Marxist geography.

Social Relevance

Early work in radical geography accepted the view that the discipline was a science of space and environment, but argued for a change in the more particular topics of concern in the direction of urgent social issues. The language used by radical geography in its early years tended to be confrontational, optimistic, and anarchistic, and the message in *Antipode* was often phrased informally, as drawings, cartoons, and posters, as well as the usual articles and reviews. The idea prevalent at the time was that radical geographers were different kinds of intellectuals. The radicals were often from working class origins, rather than the middle and upper middle class backgrounds that had prevailed in earlier academia. Their political commitments were held fiercely, with passion and even anger. They were different kinds of agents practicing their daily lives, thinking, doing research, and teaching in new ways. Earlier scholars, it was thought, had entered academic life for different, more genteel, and scholarly reasons than the new generation of radicals, who thought hard, and taught well, in an effort to transform society, and were uncomfortable with the usual trappings of intellectual power, like being called professor or doctor, or being addressed by their last names. Meetings of the Association of American Geographers (AAG) were (and still are) held at prestigious hotels in major cities. These hotels always have a central hallway packed with socializing 'academic conventioneers'. Young radical geographers saw this as a status driven activity, the discipline's power structure expressed spatially as a kind of central place hierarchy, with power/space patterns of who talks to whom on the convention floor; gods talking with gods, disciples hovering, and graduate students pecking up occasional crumbs of condescending attention. Radicals vowed never to be like that. How could they know that within a decade they would be the new gods of the geographic intellect, invited all over the world to give polite talks on revolution?

How could they know that structure would overcome intention in replicating only a slightly more relaxed power/space hierarchy?

The late 1960s and early 1970s were years of political involvement and intellectual excitement unique in the modern history of geography. Looking back from the future, when gains made then have become commonplace, even *passé*, it is difficult to realize how dramatic the shift was between the kinds of interests and level of dialog predominating in the geography of the mid-1960s, and those coming to the fore by the mid 1970s. 'Social relevance', the key phrase of the time, was conceived not in a vague way, as 'feeling sorry for one's fellow human beings', but as taking the side of the oppressed, advocating their causes, and pressing for fundamental social change. Even those openly professing liberal values, like the quantitative geographer Richard Morrill, who criticized the revolutionary premises of the New Left, still maintained that the academic's role was "to help bring about a more just, equal, and peaceful society" and "search for more 'radical' ways and means to achieve change." The result of work like this was indeed a more socially relevant geography but, it was later concluded, one tied still to a philosophy of science, a set of theories, and a methodology compatible with the existing academic and disciplinary power structures.

Anarchic Leanings

In the early days of radical geography the tendency was to draw freely on a variety of political theories, including anarchism. There are structural similarities between anarchism and geography in terms of anarchism's naturalistic interpretation of the human character (as essentially cooperative), and the decentralized spatial form of its vision of an alternative society. Leading anarchist theoreticians, such as Elisee Reclus and Peter Kropotkin, had been geographers. With such connections radical geography made numerous attempts in developing an anarchist base with a distinct political perspective on alternative forms of society drawing from Kropotkin's writings.

Kropotkin thought that we should learn from the great sweep of history in building an alternative to capitalism. For long time periods humans lived in groups organized around principles of cooperation and mutual support, for it was found that cooperation and altruism were the only lasting bases of social life. Kropotkin believed that natural cooperativism forms the basis of peoples' ethical system. Yet competition also functions for individual survival in (relatively short) periods of struggle. Raising competitivism, a short term reactive emotion, into a long term social organizing function makes capitalism contradictory. Kropotkin thought that we must return to societies based on cooperation and mutual aid, principles that have continued to be practiced (e.g., in the family) and are propagated still via an underground people's history. Kropotkin thought that producers are so mutually interdependent that all labor is social in character, all people are equal inheritors from the past, and therefore scarce goods should be distributed according to need.

Based on such principles, the anarchist geographer Myna Breitbart advocated 'integrated labor' rather than the division of labor as the socioeconomic basis for the organization of production. People should perform several different kinds of tasks in free associations with the means of production and products held in common by a decentralized society. Essentially self sufficient regions would become 'integrated cells' prepared to equally interchange ideas and products with other regions. From anarchist principles such as these, Breitbart derived a 'location theory' to guide the active construction of a new landscape. Anarchist economics, she said, sacrifices traditional forms of productive efficiency in favor of alternative social goals like equity, need and creative work: these would form their own kinds of cooperative efficiency based on producing what is necessary in the most humanly satisfying way, with the time freed from work available for leisure, inventive or artistic pursuits. Anarcho–communists believe in the gradual disappearance of the dichotomy between rural and urban land uses, the decline of large cities, and the formation of egalitarian spaces. Production decisions should be democratically made at the grassroots level by confronting needs with available resources, including alternative uses for labor. Centralized hierarchies should be replaced by directly connecting (horizontal) networks. Work places and living places should be nearer together, allowing greater integration of the various spaces in which life is lived. Such rudiments of people's location theory, Breitbart suggested, might be considered a prelude to the making of landscapes which benefit everyone living in them. Believing deeply in such ideals many radical geographers have long understood communism to be a radically democratic, decentralized society, where the people directly control the making of their own spaces, the very opposite of the state authoritarian 'communism' practiced in the former Soviet Union. No wonder that when red baited for being 'authoritarian Marxist–Leninist communists' radical geographers reacted with amused wonder.

Radical Geographical Practice

The early days of radical geography were characterized also by extensions of the political activism typical of the anti war and civil rights movements into more obviously 'geographical' arenas. In 1968, William Bunge, previously a leading quantitative–theoretical geographer, founded the Society for Human Exploration. This called for re discovering geographic skills of exploration (this being meant in an ironic sense) and using them for new, pol itical purposes that reversed their previous intention. For Bunge "the tyranny of fact compels that geographers go into a state of rationally controlled frenzy about the ex ploration of the human condition" by forming 'exped itions' to the poorest areas of cities, contributing to, rather than taking from, communities, planning with, rather than for, people. Geographers should become people of the regions they explore, should discover the kinds of research people need doing, and address them selves energetically to these problems. Local people should be trained in geographic skills so they could be come part of the solution, rather than being mere passive objects of study. Thus, in response to a request from a black community organization, the Detroit Geographical Expedition prepared a report on school desegregation which maximized black control over black children's education. A report on Trumbull, a working class inner Detroit neighborhood, countered plans for expansion by Wayne State University. Furthermore, as part of the training program of the Expedition, courses were or ganized at Michigan State University that enrolled 670 students before expeditionary principles such as com munity control and free tuition led to the educational project being terminated by officials of the university. Also, by 1973, the Detroit Geographical Expedition ceased to function beset by problems such as the 'en forced mobility' (to Canada) of its founder and student members, and the refusal of nearby universities to grant tenure to supportive faculty members. Expeditionary movements were carried to Toronto (Canada), Sydney (Australia), and London (England) in the resulting dias pora, where they survived into the late 1970s.

Within the discipline of geography, a Union of So cialist Geographers (USG), founded in 1974, organized a couple of hundred members in North America in its first year, and a hundred more members in Britain soon after. Its most successful union local (at Simon Frazier Uni versity in Burnaby, British Columbia) enrolled 50 student members and participated in several expeditionary and advocacy activities in the city of Vancouver. Members of the USG organized special sessions at AAG conventions and Institute of British Geographers (IBG) annual meetings on topics like imperialism, women, and revo lutionary theory, much to the dismay of the old guard in the discipline of geography ('is this geography?'). As an active force the USG lasted until the early 1980s when its place was taken by socialist specialty groups within the dominant disciplinary institutions – a sign that what was once oppositional had been 'disciplined'.

The Radical Critique

In the early 1970s, self consciously radical geographers began to seriously question mainstream ideas of the dominant stream of quantitative–spatial geography. This critique went beyond the social relevancy issue in its depth, political intensity, and transformational intent. The new criticism centered on geography's pervasive distinction between the spatial dimension to human ac tivity, and the full range of social activities concerned with the reproduction of existence, the key critical term being mainstream geography's 'spatial fetishism'. Thus in the case of industrial location theory, quantitative geo graphy's prize product, Doreen Massey found treatments of 'the spatial' as a closed system, analyzed via an au tonomous (location) theory, to be incoherent – because space is not a separate dimension to human life, but is immersed in social activity. She argued that the location theory being used is not scientifically neutral, but is based on neoclassical economics, a form of analysis that is ideological, in that it directs attention to the profits of industrialists rather than the interests of working class people, and in general justifies the existing economic system. As a result of these deficiencies, a gap widens between location theory and the formation of real world spatial problems (or, rather, socioeconomic problems with spatial manifestations). Ideology, she said, makes con ventional science ineffective.

Likewise, in a critique of development theory, David Slater claimed that spatial patterns could be described in the most minute mathematical detail without disclosing anything significant about causal social processes – for Slater geography was caught in a methodological trap in which the collection and manipulation of spatial data was preferred over the elaboration of social theory. For both Massey and Slater geographical location theory un critically draw their basic frameworks of understanding from the bourgeois social sciences.

Perhaps the most sophisticated version of the notion of conventional science as ideology was elaborated by David Harvey. Harvey found the argument that the 'scientific method' guarantees objectivity to be an ideo logical claim. For, the (normative) principles of sci entific method can be justified and validated by appeal only to something external to science – metaphysics, religion, ethics, convention – realms that scientists would agree are penetrated by ideological consider ations. This does not convert science into mere personal opinion. But it forces the further concession that 'sci entific' enquiry expresses social ideas and conveys social meanings.

By comparison the only method capable of dealing with geographical issues, such as the population–re sources relation, in an integrated, dynamic way, must be founded on Marx's dialectical materialism.

From Radical to Marxist Geography

This proposal for employing Marxist ideas was instrumental in bringing about a transformation in radical geography. Radical geography had begun as a quest for social relevance at a time of contradiction and crisis in capitalist society. But social relevance produced a contradiction within radical geography, between political objectives that were virtually unlimited (i.e., aimed at the transformation of society), and analytical capacities that were claustrophobically constrained by the techniques, methodologies, and paradigmatic boundaries of the conventional scientific concepts employed in radical geographic analysis. Thus radical geography was radical in topic and politics, but not in theory, or method of analysis. Truly radical explanations, it came to be believed, had to trace spatial issues to their origins in societal structures. As the anti war movement escalated into political insurrection, and as many radical geographers became dedicated revolutionaries, the need for a politicized social ontology tying environmental space to society became ever more compelling. The urgent need for an alternative mode of understanding became clear even to people reluctant to completely abandon positivism, with its quantitative methods that geographers had slaved over in a desperate attempt at becoming white coated 'scientists'. Again, the transformation of radical into Marxist geography was typified by the trajectory of David Harvey.

Harvey began to explore how ideas usually regarded as distinct from the philosophy of science, notions of social and moral philosophy, could be related to geographical inquiry. Harvey wanted to breach the gap between what appeared to be two irreconcilable modes of analysis, the ethical and the scientific. The emerging objective social conditions and geographers' inability to cope with them made necessary a further revolution in geographic thought. The question, for Harvey, was how to accomplish this? For Harvey, there were several possible paths. One was to reject the positivist basis of the quantitative movement in favor of phenomenology – that is 'the concept of man as a being in constant sensuous interaction with the social and natural realities which surround us'. But phenomenological approaches could lead to idealism or naive empiricism – the so called 'behavioral revolution' in geography had shown this. A more fruitful strategy was to explore a Marxist understanding in which positivism, materialism, and phenomenology overlapped to provide interpretations adequate to social reality. Hence, like positivism, Marxism is a materialist analysis, but whereas positivism seeks merely to understand, Marxism understands with a purpose in mind – changing the world. Put differently, positivism draws its categories and concepts from existing reality in a straightforward and uncritical way, but the Marxist dialectical method incorporates contradictions and paradoxes, and points to their resolutions.

Harvey examined the notion of rent as a central factor in urban land use allocation. For Marx rent was a manifestation of surplus value produced by the exploitation of labor. Marx provided a discussion of origins missing in positivistic accounts that covered everything except the most basic characteristics of capitalist market economies, and devised all manner of solutions except those challenging the continuation of the capitalist form of economy. Such limited discussions and superficial "solutions," Harvey said, made geographic researchers look foolish. Hence the need for a revolutionary approach to theory that went beyond the mapping and measuring of social problems to become "nothing more nor less than the self conscious and aware construction of a new paradigm for social and geographic thought through a deep and profound critique of our existing theoretical constructs" and the production of "a superior system" that, when judged against the realities requiring explanation, would make "all opposition to that system of thought look ludicrous."

This led Harvey to an alternative Marxian theory dealing with land use within cities and cities in space. As opposed to the neoclassical economic notion of rent as a positivistic rationing device, sorting land uses into locations, Harvey counterposed a Marxian theory of rent as payment to the monopolistic power of private property. He also proposed a general theory of urbanism that linked cities to modes of production; that is, cities are formed through the geographic concentration of the social surplus product and function to stabilize socioeconomic contradictions. Marxism, he found, was the only method capable of grappling with issues such as urbanism in this general, system connecting way.

In essence Harvey reached (slightly in advance) a position similar to that realized by many other radicalized geographers – that it was impossible to do radical geography without a truly radical social theory. This theory was increasingly found in Marxism. In many geography departments groups of committed radical geographers began, or joined, Marxist reading groups, particularly useful for getting over the infamously difficult chapter 1 in volume 1 of Marx's *Capital*, but more generally effective in building rapport as theoretical understanding interacted with the development of a revolutionary politics. From these reading groups emerged a new generation of geographers that moved toward intellectual maturity in the mid 1970s, a group that retained its committed radicalism into the more conservative years of the 1980s and beyond. As many people began to read and understand Marx, in part stimulated by Harvey's work, geographical thought was transformed in content and philosophical depth. Simply put, the level of prevailing discourse changed from an obsession with the empirical and the quantitative to an equally committed obsession with the political–epistemological and the political–ontological, with philosophy, social theory, and with geography's place within it.

Marxist Geography

Fired by ideas like these, of a different order than those exercising previous generations, radical geographers started in the mid 1970s to teach, research, and write using consciously Marxist approaches and methodolo gies. The difference between early socially relevant radical geography and radical Marxist geography resides in the depth and style of intellectual practice, the far greater contact made with other social theories via the corpus of Marxist works, and the sense of long term mission in doing research. Revolutionary theorists are committed to developing the most insightful theories ever known, and are determined to use the power that comes with that persuasive ability to change society completely and for ever. If the radical geography of the late 1960s was exhilarating for its adherents, Marxist geography in the mid to late 1970s provided a deep sense of committed satisfaction, tempered by the occa sional realization that acts of intellectually elitist arro gance were occurring – in that they believed Marxism to be better than all other forms of social analysis. For, the interaction between Marxism and knowledge of en vironment and space (hardly the least significant of human understandings) produced powerful theoretical insights into profound issues of human existence, and these in turn produced disciplinary power in many guises.

For much of the 1970s, Marxist geographers in the Anglo American tradition were preoccupied with de riving analytical concepts and categories from the corpus of Marx's works and applying them to practical issues of immediate concern. As understanding matured into a fully Marxist geography, exegesis was gradually, although never completely, replaced by the creation of specifically Marxist geographic theories, mainly through interactions between the new insights and the old preoccupations with nature and space. Neil Smith's work can be taken as symptomatic of the best of Marxist geography at the time.

The Production of Nature

For Smith, the various conceptions of nature are organ ized into an essential 'dualism': on the one side, nature is external, a realm of extra human objects and processes; on the other side, nature, including human nature, is universal. The historical roots of this dualism, in the Western philosophical tradition, can be traced to Kant, for whom external nature was the physical environment, internal nature was human passion, and the knowing mind (rational consciousness) was means of overcoming the dualism. Smith thought that this Kantian dualism permeates Western thought to the level of intuitive correctness. Dualism flows through the scientific and literary traditions. For Smith dualism is an ideologically necessary contradiction in contemporary thought. As externality, nature legitimates subjugation. As universal ity, nature invests certain forms of human behavior, like competition, with normative status. Dualism however has disastrous consequences in reducing nature to being a mere 'external factor' in capitalist economies and in conventional forms of knowledge like neoclassical economics.

By contrast, in his early work, Marx insisted on the 'unity' of nature. For Marx, society was internal to nature, the mediating processes literally passing through 'me tabolism' (i.e., the process of building nutritive material into living matter) or 'metabolic interaction' with labor as active force and unified science as mode of understand ing. For explication Smith turned to the main theorist of the Frankfurt School of Marxism who thought about natural relations: Alfred Schmidt detects in Marx's notion of metabolism as eternal, nature imposed necessity, a 'negative ontology' of the purposive human will's tri umph over nature, with technology as its emancipating, utopian force. In the Frankfurt School, by comparison, technology is seen as a dominating force. Smith found this interpretation of Marx to be a figment of Schmidt's imagination. Indeed Smith found the Frankfurt School's theme of the domination of nature replacing the inter pretation he preferred, a Marxian emphasis on class conflict; Smith thought that this shift in emphasis, from class to a generalized domination, leads to a critical politics centered on the 'human condition' rather than a Marxism centered on social relations. This inserts a flawed, dualistic conception of nature into the environ mental movement.

In place of the Frankfurt school's 'domination' of na ture Smith substituted a more complex notion of the 'production' of nature. People are originally natural beings, with their consciousness directly interwoven with material and natural activity. But this is a differentiated unity of people and nature: humans distinguish them selves from animals by producing their means of sub sistence (the implication being purposive and deliberate production), a process that eventually yields permanent surpluses of products, the material bases of class for mation. With this, interactions with nature come to be mediated by unequal social relations. For Marx this is a contradictory movement: social surplus begins a process of emancipation from natural constraint; emancipation is accompanied by internal social differentiation together with institutions like the state, patriarchy, and the market; the precise resolution of the contradiction between emancipation from nature and social domination forms the specific kind of society.

A further transformation in human nature occurs when, with development of the productive forces, and with division of labor and commodity exchange, the individual is alienated from work, product, and nature. When production comes to be primarily for exchange, 'pristine' or 'first' nature is separated from a 'second' socially produced nature. Natural relations come to be primarily characterized by exchange, with nature relegated to the position of just one more commodity. The production of second nature speeds up emancipation but sharpens social differentiation; this twin process (emancipation with differentiation) is key to the bourgeois revolution, the formative social process of capitalism, a class society of competition, and accumulation (economic growth) for the sake only of further accumulation. Under the abstract dictatorship of accumulation, capital must expand continuously, stalking the Earth in search of raw materials, appending nature in its totality to the production process. This recast Marxist argument leads Smith to his ultimate conclusion – his 'production of nature' thesis: "Human beings have produced whatever nature became accessible to them" and "No God given stone is left unturned, no original relation with nature unaltered, no living thing unaffected." This suggests, for Smith, the 'natural inevitability' of social transformation within the possibility of the 'revenge of nature' (e.g., the greenhouse effect). Smith thinks that this analysis achieves what Schmidt's conception of nature could never accomplish, a politics of social and environmental action: "Truly human, social control over the production of nature . . . is the realizable dream of socialism," he concludes.

Political Ecology

Responding to the environmental activism of the 1970s and early 1980s an important offshoot from Marxist inspired theories of society–nature relations emerged as a more pragmatic 'political ecology'. This involved, even more than with Smith's work, a sustained critique of previous geographic notions of human–nature relations. Thus conventional geographic work carried out under an earlier 'human ecology' perspective had seen natural hazards as problems of mistaken perception, or imperfect knowledge in the human adjustment to natural events, as flaws remediable by environmental policies, rather than as basic and necessary components of the reproduction of capitalist social formations. In place of this deficient approach, ecologically concerned geographers attempted to weld together issues of how communities are integrated into the global economy, with issues of local resource management and environmental regulation and stability. Forged in the crucible of neo Marxian development theory, this new political ecology was inspired by peasant and agrarian societies caught in the throes of complex forms of capitalist transition. This relation led to an increase in the incidence of catastrophes which came to be seen as resulting as much from social vulnerability, as from natural vulnerability. In this context, a series of works wedding political economy to a critique of cultural ecology appeared in the early to mid 1980s.

Michael Watts used a version of the articulation (combination) of modes of production approach, that allowed for specificity and contingency, to look at environmental relations. Watts began his analysis with an account of the precapitalist mode of production in the Hausaland region of northern Nigeria. He found that Hausa farmers traditionally possessed an adaptive flexibility that could accommodate climatic risks at the levels of the household, community, and region; droughts were expected but there was an indigenous relief system integrated into the traditional political–economic system. This system was eroded through articulation with an expanding colonial capitalism, which incorporated commodity production (groundnut, cotton) into the international economy, but left subsistence production (food) to reorganize itself. Yet export cropping substantially altered the food production system, the culture of reproduction, and the role of the state: the moral economy became a money economy. Hence integration into capitalism dissolved many of the response systems that had buffered peasant households from the vagaries of a semiarid environment, making them vulnerable to droughts that had long been occurring. In other words, peasant households are not intrinsically pathological, but are constrained in their ability to manage threats, disturbances, and perturbations.

The most renowned and best argued political ecology to appear at the time came from Piers Blaikie. Blaikie argued that soil erosion in Third World countries was conventionally seen as an environmental problem caused by irrational peasant land uses in the context of overpopulation, with the solution found in involving peasants in market (capitalist) economies. Blaikie developed a critique of this position. He found the physical, natural context of erosion important, but chose to emphasize instead the political–economic context, especially the social relations of production determining the nature of land use. Following a bottom up approach, Blaikie began with the smallest decision making unit, usually the (peasant) household, operating under constraints of limited assets. The household is immersed in two kinds of social relations: local relations of production; and exchange relations with the world market. In both spheres "surplus is extracted from cultivators who then in turn are forced to extract 'surpluses' . . . from the environment . . . and this in time and under certain physical circumstances leads to degradation and/or erosion" (Blaikie, 1985).

Through incorporation into the global capitalist system peasants are spatially marginalized, crowded onto lands of limited agricultural potential, where they adopt desperate survival strategies. Hence political–economic processes are interrelated with natural processes; steeply sloped areas, with high natural propensities to soil erosion, are deteriorated by desperate farmers; soil erosion then exacerbates their predicament, reenforcing or intensifying the original process. For Blaikie, underdevelopment, rather than 'false perception' or 'peasant ignorance' lay at the heart of what is better phrased as a poverty–nature syndrome. Small producers cause soil erosion under conditions of threat to livelihood, making cooperation difficult, and state intervention ineffective, while the available conventional ideas about soil erosion ignore the economic imperatives that compel peasant actions.

Spatial Relations

Developments in the derivation of a Marxist theory of the production of nature were part of a more general process of exegesis, rethinking, and theoretical creation that focused even more intensively on an evolving conception of the production of space. During the 1970s a series of articles, mainly appearing in *Antipode*, drew out the spatial implications of works by Marxist theorists in a kind of 'capital logic' school of radical geographic thought. That is, the movement of theory was from contradictions in the historical dynamic of the accumulation of capital to manifestations of these contradictions in space, and then back again, as spatial effects (uneven development) became contributory aspects to basic societal contradictions. A main component of this position was the notion of the build up of crises internal to regions expressed in outward movements in space – hence interregional social relations were theorized as spatial components of a more general internal–external dialectic. Two versions of this position appeared: from a decade or more of Marxist work, conducted mainly by Harvey and his students (Lata Chatterjee, Richard Walker, Neil Smith, and others at The Johns Hopkins University in Baltimore) emerged a theory of the expression of contradiction in uneven development; from Henri Lefebvre and those influenced by him, such as Edward Soja, came theories of the complex production of space and sociospatial dialectics.

For Harvey, the Marxist theory of growth under capitalism puts capital accumulation at the center of a dynamic and inevitably expansionary mode of production. This dynamism encounters barriers in labor, means of production, and markets, that precipitate crises, such as chronic unemployment, or realization crises (inabilities to get rid of products), with the eventual result of a shift in the accumulation process to a higher plane – new social wants, new markets, etc. Newly 'rational' locational patterns are an inevitable, necessary part of capital accumulation. But there is a spatial aspect to contradiction: in overcoming spatial barriers and 'annihilating space with time' spatial structures are created that ultimately act as barriers to further accumulation. This is particularly the case when capitalism comes to rely on immobile, fixed capital (i.e., means of production fixed in space) rather than more mobile variable capital (i.e., labor). Capital then has to negotiate a path between: preserving the value of past investments in the built environment (e.g., the urban landscape); and destroying these investments to make new room for accumulation. The virtue of Marx's location theory, dispersed though it may be in textual fragments, lies in the way space can be integrated into 'fundamental insights into the production of value and the dynamics of accumulation'.

A series of theorists in the Marxist tradition, such as Lenin, Luxembourg, Baran, and Frank, connect capitalism's internal contradictions with its external solutions of the violent penetration of precapitalist societies and the production of a center–periphery spatial organization in the resulting global economy. For Harvey imperialism is the main geographical necessity deriving from the expansionary dynamic of accumulation. The specific economic form of imperialism – be it mainly a search for markets, places to invest surplus capital in, or new sources of raw materials – varies over time, and is a matter for empirical, historical analysis. In terms of geographic theory this dynamic theory of contradictions contrasts with bourgeois location theory's equilibrium analysis of optimal spatial configurations. "The Marxian theory . . . commences with the dynamics of accumulation and seeks to derive out of this analysis certain necessities with respect to geographical structures. The landscape which capitalism creates is also seen as the locus of contradiction and tension, rather than as an expression of harmonious equilibrium. . . . The Marxian theory teaches us how to relate theoretically, accumulation and the transformation of spatial structures and ultimately . . . the reciprocal relationships between geography and history." Thus the internal and external dimensions of space are linked and related to capital accumulation. Revolutions in the productive forces, increasing scale of production, and the concentration and centralization of capital are paralleled by urban agglomeration in a widening international capitalist space.

Neil Smith too attempted to systematize a synthesis of Marxist notions about space. As we have seen, Smith argued that different conceptions (of space in this case) are produced by different types of human activities (in space); hence there is a history of theories of space forming part of the history of human experience. The development of second out of first nature, with advance of the forces of production, is the historical–material basis of a theoretical bifurcation of absolute, natural, and physical space (the world of physical and natural phenomena) from relative and social space (the humanly constituted field of social events). For Smith, conventional conceptual divisions like these accept a dualism between space and society. In an equally conventional economic geography, this dualism is no problem – space is the given substratum, society uses it, the two realms interact. But when the uprisings of the 1960s provoked an interest in urban social space, criticisms began to appear of the dualism of space and society: hence humanistic geography saw space not as objective structure but as social experience; and the Marxist tradition saw space and society as dialectically intertwined. For Smith: "The notion that space and society 'interact' or that spatial patterns 'reflect' social structure is not just crude and mechanical in its construction, but also prohibits further insights concerning geographical space; at root this . . . view of the relation between space and society remains tied to the absolute conception of space. Two things can only interact or reflect each other if they are defined in the first place as separate." Smith substitutes the "production of space" as means of demonstrating this unity so that "human practice and space are integrated at the level of the concept of space itself." For Smith, the development of the forces of production transforms absolute (natural) space into relative (social) space in a contradictory movement. Development emancipates society from the constraints of natural space leading toward 'equalization' but only by sinking capital into certain spaces producing 'differentiation' and relativized space. These contradictory movements determine the specific forms of capitalist space as synthetic outcome: "Space is neither leveled out of existence nor infinitely differentiated. Rather the pattern which results is one of uneven development." Integrating the geographical notion of 'scales' (urban, national, and global) as levels of specificity, with temporal notions of rhythms, cycles, and long waves, Smith proposes a 'see saw' theory of uneven development: "capital attempts to see saw from a developed to an underdeveloped area, then at a later point back to the first area which is by now underdeveloped."

Pioneering work by the French sociologist–philosopher Henri Lefebvre (1901–91) lay beneath much of this discussion. Lefebvre had been writing on topics like uneven development, urbanism, everyday life, and the production of space since the 1930s. But it was only with the translation of his book *The Production of Space* that Lefebvre's ideas became widely available in English. Themes of praxis and everyday life, urbanism and space, pervade Lefebvre's writings. Space, Lefebvre argued, has become a "mental thing," deeply separated from physical or social reality, in mathematics and philosophy, but also in post structural social theory. Hence Lefebvre proposes an analysis of the actual 'production' of space to bring the various kinds of space and the modalities of their genesis within a single theory. Furthermore, while Lefebvre does not accede to the (post structural) priority of language, believing that material activity probably precedes it, he thinks that space can be 'read' or 'decoded', for space implies a process of signification as well as material practice – codes are spatial languages, parts of practical relationships between subjects and their surroundings. Lefebvre's thesis, simply put, is that social space is a social product. Every society, every mode of production, produces its own space. But the production of social space is not like the production of commodities, because space subsumes many different things, is both outcome and means (of fresh action), and is both product (made by repetitious labor) and work (i.e., something unique and original). It consists of objects (natural and social) and their relations (networks and pathways). Space contains things, yet is not a material object; it is a set of relations between things. Hence we are confronted with many, interpenetrated social spaces, superimposed one on the other, a 'hyper complexity' in which each fragment of space masks not one social relationship but many. For Lefebvre the 'problematic' of space has displaced that of industrialization. He therefore proposes an analysis of space itself to uncover the social relations embedded in it using a triad of spatial concepts:

1. Spatial practice – production and reproduction secrete a society's space, forming the particular locations and spatial sets (the perceived space) characteristic of each social formation. Spatial practice ensures continuity and some degree of cohesion for social formations and guarantees a level of competence and performance to people as actors. (Social) spatial practice, for Lefebvre, is practical (material reproduction), employs accumulated knowledge, and involves also signifying processes.
2. Representations of space – abstract conceptions of space using a system of verbal and graphic signs, the celestial sphere in antiquity, maps, and plans in the modern period. Representations of space, shot through with knowledge (mixture of understanding and ideology), always relative and changing, subordinate people to a logic and play a dominant role in social and political (spatial) practice.
3. Representational space – space as directly lived through associated images and symbols by inhabitants and users but also writers and artists who describe it. Representational space overlays physical space, with symbolic use made of natural objects. Representational space tends toward systems of nonverbal symbols and signs; it obeys no rules of consistency or cohesiveness. Redolent with imaginary and symbolic elements, it has its source in the history of people, and is therefore studied by ethnologists, anthropologists,

> and psychoanalysts. Representational space has an affective center of everyday life (ego, bedroom, house); it embraces the loci of passion, action, lived situations, and thus implies time. It may be qualified in various ways, as directional, situational, or relational, because it is essentially qualitative, fluid and dynamic.

The three moments of the triad, perceived, conceived, and lived, are held together by dialectical relationships. The triad contributes to the production of space in different combinations according to the mode of production and the historical period. Relations between the three moments are never simple or stable, nor are they entirely conscious.

Echoes of these Lefebvrian positions can be heard in the Anglo American literature of the 1970s and early 1980s. Edward Soja argued that Lefebvre was interested in explaining the survival of capitalism through the attenuation of its internal contradictions through the production of space. Soja compared this with the work of the Belgian Marxist, Ernest Mandel who argued that the unequal development of regions was as fundamental to capitalism as the direct exploitation of labor power. Mandel's basic point is that the historical survival of capitalism depends on the differentiation of space into over – and underdeveloped regions, the principal role of underdevelopment being to furnish reserves of labor and compensatory markets to relieve pressures in a spasmodic, contradictory capitalist development. This global capitalist system is maintained and intensified by the geographical transfer of value through the extraction of super profits from rural peripheries under the cover of imperialism. Soja placed Lefebvre and Mandel at one extreme in a Marxist debate on space: they give the structure of spatial relations a transformational capability comparable with class struggle. At the other extreme lies the conventional Marxist approach, centered on traditional class analysis. In between, Soja says, are those like Harvey who seems to adopt the Lefebvre–Mandel position but, when pushed to make an explicit choice, maintain the preeminence of aspatial class definitions. This is part of a wider syndrome within Marxist analyses of space, what Soja calls a rigidifying pattern that weakens analysis because it incorporates "an unnecessarily limited and inappropriate conceptualization of space and spatial relations in an effort to escape spatial fetishism." This is ironic for the primary source of misunderstanding is a failure of Marxist analysis to appreciate the dialectical character of social–spatial relations. What Soja calls the "socio–spatial dialectic" fits neither alternative: "The structure of organized space is not a separate structure with its own autonomous laws of construction and transformation, nor is it simply an expression of the class structure emerging from the social (i.e., a spatial) relations of production. It represents, instead, a dialectically defined component of the general relations of production, relations which are simultaneously social and spatial." For Soja, a homology (correspondence) exists between class and spatial relations in the division of organized space into dominant centers and subordinate peripheries, a structure captured by the concept of geographically uneven development. The two sets of relations are not only homologous, in that both arise from the same origins in the mode of production, but also are dialectically intertwined. This dialectical homology was suggested by Marx and Engels, but a hundred years of scholarship failed to develop it as Western Marxism evolved without a strong spatial perspective. For Soja a flurry of debate over a new urban sociology, new geography, etc. signifies something broader, a dialectical materialism simultaneously historical and spatial.

Critique of Radical Marxist Geography

These were some of the main positions taken by radical Marxist geographers in the 1970s and early 1980s. Many of these positions survive into the present. But already, by the mid 1970s while Marxism remained at the frontier of geographic interest and adherence by progressive intellectuals, significant criticisms appeared of Marx's original thought, and of Marxist geography as derivation. These critiques centered on what was seen to be Marx's functional, mechanical, and economistic understanding of history and geography, as well as the totalitarian, repressive nature of the supposedly Marxist inspired 'communist' countries. For example, the sociologist Anthony Giddens (1981) thought that there is much in Marx that is "mistaken, ambiguous or inconsistent … in many respects Marx's writings exemplify features of nineteenth century thought which are plainly defective when looked at from the perspective of our century." Basically his position is that historical materialism, understood as the progressive growth of the forces of production, as the history of class struggle, or as a schema of historical evolution, should be rejected, leaving only the more abstract elements of a theory of human practice. He proposes instead that societal transformation occurs through episodes that are not mechanically inevitable, while different kinds of society coexist along 'time–space edges'. In general Giddens (1981) rejected Marx's argument for the primacy of production over other aspects of social life: "It certainly does not follow that, because material production is necessary to sustain human existence, the social organisation of production is more fundamental to explaining either the persistence of, or change in, societies than any other institutional forms." Giddens offered instead a theory of 'structuration' that portrays time–space relations as constitutive (i.e., formative) features of social systems.

The late 1970s and early 1980s also saw a sudden escalation in the ongoing dialog between humanistic and Marxist geography. From the side of humanistic geography came a critique apparently of structural Marxism, but actually of the whole Marxist enterprise in geography and social science. James Duncan and David Ley critically assessed the "ambitious radical tradition [in human geography] derived from structural Marxism." Their critique finds Marxian analysis in geography inclined toward mystification and theoretical stasis; indeed they think that the Marxist movement in geography foundered when theory encountered geographical circumstance. For them, this impasse derives from a philosophy of holism, totality, and essence, derived originally from Hegel, but present too in Marx, which frustrates empirical analysis. Their critical focus therefore lies on the 'problematic relations' between philosophical holism and geographical reality.

Marxism and Marxist geography was also criticized from positions established by post structural and postmodern social theorists like Foucault, Derrida, Deleuze, and Guattari. Virtually all post structural theorists use spatial metaphors, if not geographical concepts, as media of expression. Geography came late onto the post structural scene and the main route of diffusion of post structural ideas into Anglo American geography passed through Marxism, rather than directly through geographical readings even of the more obvious, spatialized versions of the originals. Such was the dominance of Marxism and its offshoots in critical human geography that many ideas in what is supposedly post structural and postmodern geography turned out as hybrids derived from blending these with historical materialist notions. Additionally, the years which saw the spreading influence of post structural ideas were also noticeable for the growth of a feminist movement in and around the discipline of geography. Many postmodernists are also feminists, feminist geography has materialist (socialist feminist) and postmodern strains, the postmodern/feminist politics of difference draws from all schools of thought. This confluence of ideas produced a post structural–postmodern period in geographical thought that distanced itself from radical geography. By the mid 1980s radical geography as a coherent, political, and intellectual project centered on Marxism was more or less over. Work continued under the broad rubric of 'critical geographies', including various post structural, postmodern, feminist, left liberal, and Marxist and neo Marxist views, with many theorists taking hybrid positions drawing on several of these views. Hence the main alternative geographical conference is organized by the International Critical Geographers Group. The Marxist project continued, often blended with other ideas drawn from post structuralism and feminism.

See also: Dialectical Reasoning and Dialectical Materialism; Feminism/Feminist Geography; Historical-Geographical Materialism; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Nature; Space I.

Further Reading

Blaikie, P. (1985). *The Political Economy of Soil Erosion in Developing Countries*. New York: Longman.
Breitbart, M. (1975). Impressions of an anarchist landscape. *Antipode* 7(2), 44 49.
Bunge, W. (1969). The first years of the Detroit geographical expedition: A personal report. *Field Notes Discussion Paper No. 1*. Detroit, MI: Society for Human Exploration.
Duncan, J. and Ley, D. (1982). Structural Marxism and human geography: A critical assessment. *Annals Association of American Geographers* 72, 30 59.
Giddens, A. (1981). *A Contemporary Critique of Historical Materialism*. Berkeley, CA: University of California Press.
Harvey, D. (1973). *Social Justice and the City.* Baltimore, MD: The Johns Hopkins University Press.
Harvey, D. (1974). Population, resources and the ideology of science. *Economic Geography* 50, 256 277.
Harvey, D. (1975). The geography of capitalist accumulation: A reconstruction of the Marxian theory. *Antipode* 7(2), 9 21.
Harvey, D. (1982). *The Limits to Capital*. Chicago, IL: University of Chicago Press.
Harvey, D. (1989). *The Condition of Postmodernity: An Enquiry into the Origins of Cultural Change*. Oxford: Basil Blackwell.
Kropotkin, P. (1902). *Mutual Aid: A Factor of Evolution*. London: Heinemann.
Lefebvre, H. (1991). *The Production of Space*. Nicholson Smith, D. (trans.). Oxford: Basil Blackwell.
Massey, D. (1973). Towards a critique of industrial location theory. *Antipode* 5(3), 33 39.
Massey, D. (1984). *Spatial Divisions of Labor: Social Structures and the Geography of Production*. New York: Methuen.
Morrill, R. (1969 1970). Geography and the transformation of society. *Antipode* 1(1), 6 9 and 2(1), 4 10.
Peet, R. (1977). *Radical Geography.* Chicago, IL: Maaroufa Press.
Peet, R. (1981). Spatial dialectics and Marxist geography. *Progress in Human Geography* 5, 105 110.
Slater, D. (1975). The poverty of geographical inquiry. *Pacific Viewpoint* 16, 159 176.
Smith, N. (1979). Geography, science and post positivist modes of explanation. *Progress in Human Geography* 3, 356 383.
Smith, N. (1984). *Uneven Development: Nature, Capital and the Production of Space*. Oxford: Basil Blackwell.
Soja, E. (1980). The socio spatial dialectic. *Annals, Association of American Geographers* 70, 207 225.
Soja, E. (1989). *Postmodern Geographies: The Reassertion of Space in Social Theory.* London: Verso.
Watts, M. (1983). *Silent Violence: Food, Famine and Peasantry in Northern Nigeria*. Berkeley, CA: University of California Press.

Relevant Websites

http://www.blackwellpublishing.com
Blackwell Publishing: *Antipode*.
http://econgeog.misc.hit u.ac.jp
International Critical Geography Group.
http://www.peoplesgeography.org
The Peoples Geography Project.

Radical Political Economy

F. Mizuoka, Hitotsubashi University, Tokyo, Japan

Glossary

Economics A branch of social science that analyzes the production, distribution, and consumption processes of goods in division of labor theoretically and conceptually. Its development as a systematic theoretical body started since a British classical economist Adam Smith published *The Wealth of Nations* in 1776.

Fetishism A state of intangible social relation or human spirit reified to a physical substance. An awesome mountain, for example, is a pure geological phenomenon, yet it can be regarded by the locals as having mythical power to generate some unusual happenings, reflecting the spirit of the locals. When a social relation, in contrast to spirit, is fetishized, it is not possible to eliminate the fetishism unless the very social relation is eliminated.

Labor Theory of Value First discovered by a British classical economist David Ricardo, the labor theory of value asserts that goods are exchanged in accordance with the 'value', which is the amount of labor input embodied in the products. Based on this Ricardian concept, Marx discovered that it is fetishism of commodity that social division of labor imputes value to each commodity.

Logical Ascent Also called 'successive approximation', this term describes the method of Marxian economics to approach from abstract theory to reality. It takes the successive logical steps called ascent to approach the actual reality. At the opening of *Capital*, the most fundamental concept of market economy, commodity, was introduced, and stepwise new logical elements were added to explain more complicated economic relations. Labor power as a special kind of commodity in deducing surplus value concept, and the profit concept that forms an element to explain the law of the tendential fall in the profit rate are the examples for this logical ascent. In this process, the abstract theory of radical political economy is successively brought up to 'middle-range theories', one of which is economic geography.

New International Division of Labor The term, often abbreviated as NIDL, was first put forward by German economists Fröbel F. *et al*. In conventional international division of labor, commodities traded were mostly final goods ready for consumption. After mid-1970s, however, three important changes took place in global economy: (1) new production technologies that facilitated vertical disintegration in manufacturing process; (2) innovation in transport and communication; and (3) opening up of the pools of cheap labor in developing and former socialist countries. These factors facilitated the multinationals to unfold its production network across the globe, which created peripheral Fordism in some successful macroeconomies especially in East Asia.

Organic Composition of Capital The constant capital or the portion of capital applied to procure equipments and materials divided by the variable capital or portion of capital applied to procure labor. A capital-intensive industry has high organic composition of capital, whereas labor-intensive industry has low organic composition. Technological innovation tends to make organic composition higher, because labor is typically replaced with automated machines.

Subsumption (to Subsume) At dawn of capitalism, capitalists subsumed skills that were embodied in the artisans. Since it took a long time and cost to train alternative artisans, they often acted as barriers to further capital accumulation and domination of capital over labor. Marx termed the relation causing these conflicts formal subsumption of labor. To transcend this conflict, capitalists invented machineries and disembodied the skill out of the laborers. This process of transforming physical nature, called real subsumption of labor, transcended the contradiction. This concept can be extended to deal in relation to society with other physical substances, as space or physical environment.

The Luddites An early nineteenth-century social movement in Britain, the Luddites believed that such evils of modern industrialization as unemployment and sweatshops came from the technological progress in physical means of production (i.e., machines) rather than the exploitative capitalist social relation. In this belief, they physically destroyed cotton and wool mills, and were later summarily arrested and some executed.

Variable Capital Abbreviated as '*v*', it is a portion of capital invested by a capitalist to purchase labor power. This compares with constant capital, or a portion of capital to purchase production equipment and materials, which is abbreviated as '*c*'.

Introduction

Classical economics, often called 'political economy', took pains to analyze the social relations behind the division of labor.

Radical political economy emerged out of this context. In the era of the industrial revolution, social activists denounced social evils contrary to universal justice and moral in capitalism. Radical political economy attempted to sublimate those critics on moral grounds into those backed with more robust logic and scientific theories and direct the proper path to solve these evils.

For example, in early nineteenth century, the Lud dites, claiming that the machines in the factory were the cause of the capitalist evils, set out to destroy the ma chines. The utopian socialists blamed self centered and evil mentality as being the cause of these evils, and in its place attempted to create a harmonious utopian environment.

Karl Marx, on the other hand, took the conceptual and theoretical path. With belief that it was the capitalist imperative that coerces the capitalists to create those evils, he set out to discover what was hidden behind this imperative through formulation of the critical theory of capitalist economy. To prove this, Marx wrote *Capital*, which was the comprehensive and theoretically robust analysis of contradiction ridden social relation behind capitalist mode of production. Following the tradition of Marx, radical political economy treats economic system as a social organization, and attempts to analyze critically division of labor that gave rise to market and capital. As such, it has strong interest in the social conflicts em bedded in those relations, and in the ways to transcend these conflicts.

In contrast, neoclassical economics presupposes con flict free economic persons, regards market foundation of economy as eternal, and often sees social problems as mismatch between various economic variables.

Three Most Original Contributions of Marx

Marx's three volume *Capital* has been offering conceptual base to radical political economy until now. In *Capital*, Marx made three important conceptions: (1) fetishism of commodity, (2) surplus value, and (3) the law of ten dential fall in the profit rate. We now explain these concepts in turn.

It is important to note that Marx explicitly dis tinguished market economy from capitalism. The market economy is more universal and forms a social base of historically more specific capitalism. The first contri bution of Marx, 'fetishism of commodity', analyzes the secret behind market economy. It explains why a com modity, a physical substance with utility for others ('use value' by Marx), has price (which Marx regarded as the surface appearance of 'value') in the market. Marx ex plained the concept of labor theory of value developed by David Ricardo as reification of division of labor organized as market economy. A commodity with certain use value (e.g., 1 pound of iron) has power of being exchanged with another commodity (e.g., 1 yard of linen) that requires the same amount of labor input (e.g., 1 h). The countless repetitions of such exchanges in the market eventually generate interpretation among the market participants that 1 yard of linen has a power of being exchanged with other commodities that socially require 1 h to produce it. Over time, this imputes a 'equivalent form of value' to 1 yard of linen, which is a 'mirror' of other commodities to represent the value, the real substance of 'exchange value'. Eventually, the division of labor imputes 'general form of value' to a unique physical substance (gold, for example) that functions as 'universal equivalent' to represent the value of all other commodities. This is the origin of money. The making fetish of a physical substance is, therefore, not the outcome of mere intersubjectivity, but reification of the tangible social relation. The social structure assumed here is the same as that of neoclassical economics: everyone is atomic and equal, and no dom ination of one over another.

'Surplus value', another important conceptual dis covery, is directly related to capitalism. Marx proved how egalitarian market structure transforms itself into the vertical social relation of domination and subordination. The key element in this proof is the comodification of labor power. Its value augments while it is consumed, because the costs of reproducing a laborer him /herself plus raising subsequent generation of laborers are always less than the value that the laborer creates. The capitalist obtains the value that a laborer produces by selling the products in the market, and pays him/her only the amount to cover his/her reproduction costs as wage. The difference between what the labor produces and what labor receives is surplus value which is exploited by the capitalist. As labor has been deprived of all the means of production in the process of primitive accumulation and subsequent process of social reproduction, s/he can do nothing but keeps selling his/her own labor capacity in order to survive. Marx thus succeeded to explain how social inequality emerges out of egalitarian market society.

Marx thought that the exploitation of one class by another must end sometime in the future, through transcendence from capitalism to a communal mode of production (which Marx called communism, the term later tarnished by Stalin). Besides class struggle, Marx proved a self destructive 'time bomb' is built in endogenously within capitalist economy. This logic is called the 'law of the tendential fall in the profit rate'.

According to Marx, profit rate p' is given in the fol lowing formula:

$$p' = \frac{s}{c+v} = \frac{s/v}{c/v+1}$$

where s is surplus value, c constant capital and v variable capital. Dividing both denominator and numerator by v, we get the rightmost fraction. Here s/v is 'rate of exploitation', and c/v the 'organic composition of capital'. The higher technology level becomes, the less labor and more constant capital shall be deployed. Provided that s/v or rate of exploitation remains constant, the profit rate p' tendentially falls over time, because the value of the denominator increases. Since the motive of the capitalist mode of production is to get profit, the incentive for continuing production shall wane over time, which eventually triggers crisis. This is the economic proof (as opposed to political proof which is class struggle) of the ephemerality of capitalist mode of production.

Not as ingenious as above, Marx's tenets on joint stock company is nevertheless of considerable import in understanding contemporary capitalism. He pointed out that joint stock company enabled planned large scale capital investments, which were in fact transcendence of *laissez faire* market economy within the frame of capitalism. Stock is a form of fictitious capital, which converted the capitalists into managers of capital in part owned by others, and the capital owner into money owner seeking interest in form of dividend, as well as capital gains.

Marxian Economics after Marx

Marx died in 1883, leaving much of his works unfinished. The draft for volumes 2 and 3 of *Capital* was edited by his collaborator, Frederick Engels. Debates took place after Marx's death on the agenda that Marx failed to give definitive solution for and on the structural changes of capitalism into oligopoly and monopoly.

Hilferding gave a comprehensive theoretical frame for changes of capitalism after Marx. Setting out from the concept of money as fetish by Marx, he elaborated the concepts of joint stock company that Marx had put forward, and developed the concept of financial capital, which is the lynchpin of monopoly capitalism, out of Marx's conception of credit and interest. Hilferding analyzed the stock exchange in detail and pointed out the speculative nature of fictitious capital traded in stock market. He also examined the role of the banks in the era of monopoly capitalism as the institution controlling and dominating over the management of industrial capital through its leverage of capital credit. He called it 'finance capital'. It sought investment abroad in pursuit of higher rate of profit.

Lenin followed Hilferding's concept of finance capital and developed the concept to explain global politics of imperialist countries under monopoly capitalism. Imperialism emerged, according to Lenin, when partition of the globe by the countries exporting capital failed to expand colonial frontier once the globe had been filled with boundaries. All that remained possible was redrawing of the boundaries only to be achieved through imperialist wars. This tenet of Lenin has direct connection to the tenet of political geography: from frontier to boundary.

In radical political economy, crisis theory is of foremost import, because it is supposed to prove the ephemeral nature of capitalism. Given this significance of crisis theory, it has attracted significant debates among radical political economists.

The contradiction that eventually gives rise to crisis is already embedded in the market economy itself, where the commodities in dialectical unity of use value and value are traded. The market agents supply use value, with the ultimate aim of obtaining exchange value. Yet, there is no prearranged match between supply and demand.

Capitalists have incessant desire to accumulate by injecting commodities for sale into the market, which is possible for a while because 'selling' and 'buying' are split in time and space in the market economy. Yet it will eventually hit the limit of consumption. Overproduction and underconsumption are thus inevitable in capitalism. They are both sides of the same coin, yet the overproduction theorem emphasizes the imperative of capitalism to accumulate without limit, whereas the underconsumption theorem focuses on the propensity of the capitalist to exploit laborers to the end. Rosa Luxemburg associated this aspect of crisis theory to inevitability of imperialism, claiming that overproduction could only be solved with fresh market that had yet been subsumed by the capitalist mode of production. American radical political economists active until 1960s subscribed to underconsumption theory, which tended to view capitalism as the state of secular stagnation, contrary to what happened in most of the macroeconomies in the US, Western Europe, and Japan after WWII.

Another version of crisis theory, attempting to come to terms with the economic prosperity of post war capitalism, called on 'profit squeeze' theorem. The more capital accumulates, the demand on labor increases and the wage rises. Capital accumulation should eventually hit a point where further accumulation of capital brings about negative marginal profit. As soon as capitalism hits this labor bottleneck, the capitalist abandons further capital accumulation. Additional demand on fresh production equipment ceases, thus a group of industrial sectors that produce means of production ('department 1' by Marx) no longer finds the market to exchange their products with consumption goods produced by a group of industries called by Marx as 'department 2'. This leads to disequilibrium and eventual downturn of macroeconomy. The 'profit squeeze' theorem is therefore more fir for explanation of the business cycle.

To explain crisis as the development of internal contradiction of the law of tendential fall in the profit rate is another theorem. The organic composition of capital rises as skilled laborers come to be replaced with machines. According to Braverman, skilled labor shall eventually be replaced with computerized high tech machines, and only simple labor shall remain that even a gorilla can perform. However, there is an obvious flaw in this argument, as someone must keep developing automated machines and write sophisticated computer programs that make the deployment of deskilled labor possible. Someone must be in charge of the shop floor and corporate managements as well. Gillman statistically proved that organic composition of capital did not increase in the US before WWI. The likely outcome is segmentation of the labor market into primary or skilled and secondary or unskilled, rather than general degradation. In the current neoliberalist labor market this segmentation is increasingly entrenched, with the latter segment mostly carried out by those in the informal sector. The deployment of the pool of low wage laborers in the periphery is but a global manifestation of this segmented labor market.

Macroeconomic management in socialism was another arena of debate. Stalin, who seized power of the Soviet Union after Lenin's death through struggle with Trotsky, adopted a centralized system of organizing division of labor similar to equipment deployment and mobilization scheme similar to military operation. Alienation of the laborers and concomitant sense of decommitment among laborers ensued in the USSR. This led to the stagnation of production and eventual collapse of Stalinist socialism itself. Many people in the West were disillusioned in socialism, and some even in radical political economy itself. Many conscious radical political economists are thus seriously investigating alternatives to the Stalinist type of socialism in materializing the ideal of a communal type of economy that Marx put forward in his final years.

The Movement for Radical Political Economics in Late Twentieth Century

In countries where left wing political movements remained strong, as France and Japan, the radical heritage in economics survived well into 1980s, whereas in the US and West Germany, where the Cold War played considerable negative effect over the politics and academia, the heritage failed to survive.

Amidst this political ambience, many left wing scholastic movements emerged or re emerged in the United States. In geography, Richard Peet of Clark University and a group of graduate students there published in 1969 the inaugural issue of *Antipode*, the journal of geography ever published in English that specially catered to radical geography. It quickly grew into the forum of radical geography in the Anglophone world, and functioned as the incubator for the first generation critical geographers.

The economics students were instigated through engagement in protests against the Vietnam War and relatively abundant government research funding provided by the US government under the 'war against poverty'. According to Attewell, the research agenda for this 'second generation' radical political economy are the following: theories of economic segmentation and poverty, theories of the labor process and the firm, crisis theory, imperialism and dependency, theories of state and culture and ideology. Radical political economists also engaged in critical appraisal of neoclassical economics, which comes increasingly to dominate many departments of economics across the world.

In 1968, the students and concerned faculty members from University of Michigan and other universities founded the Union of Radical Political Economics (URPE). They eventually got teaching positions at the Department of Economics, University of Massachusetts at Amherst, which took a leading role in the researches of this group. It has now grown into a major organization of left scholarship in the United States.

The URPE does not maintain steady relations with labor movements, while it associates loosely with various social movements, including the World Social Forum and other forms of political praxis. It issues the professional journal *Review of Radical Political Economy* and offers teaching resources of economics in alternative orientation for both universities and secondary schools.

Radical political economy attempts to give theoretical explanations to the evils of capitalist society, using various versions of Marxist economics, not limited to the 'orthodox', endorsed by the communist parties. Nevertheless, the hallmark of a robust social science is the capability of demonstrating causal relationship. It is important for radical political economics to pursue the single frame of reference where the base or economic instance ultimately determines such superstructure as politics and culture.

This contrasts with postmodernism, which sometimes uses critical or radical narratives similar to radical political economy, yet rejects general frame of reference, especially those recognizing economics or capitalist class relations at its base, by blaming it as reductionism or 'grand narratives'. Move for 'pluralist renovation', claiming that there is no one way causal relationship between economy and society, shares at some point similarity with postmodernism. They therefore have the inherent danger of depriving radical political economy of causal basis of robust social science, leading to fragmented, multiple frames of reference which are not related much to one another. Radical political economics shall thereby lose its competitive edge against the op ponent, neoclassical economics, which has been taking pains to demonstrate this causal robustness using so phisticated mathematics.

Another important task for radical political economy is to develop ideal types of society alternatives to market economy. In this respect, a radical political economist needs to admit Sayer's claim that division of labor is in itself transhistorical, and there is no way for us to get back to primitive Stone Age society where a family produces everything that they use self sufficiently. Yet market is not the sole social system to organize the division of labor, and the fetishism of commodity is not something of eternal universality over history. As Polanyi, a Hungarian economic anthropologist, claimed, the social system called reciprocity does exist as the alternative to market system in organizing division of labor, albeit the details remain to be elaborated further. This point cannot be emphasized too much, because after the collapse of Stalinist type of socialism in almost all East bloc coun tries, many people were drawn into Margaret Thatcher's axiom: "There is no alternative (TINA)" to market economy.

This point is demonstrated by using the Cartesian surface as **Scheme 1**, on which types of individuals and society are shown. Along the vertical axis the spectrum of the types of individuals from egocentric to altruistic is shown; and along the horizontal axis that of society from monetary and impersonal to communal and personal.

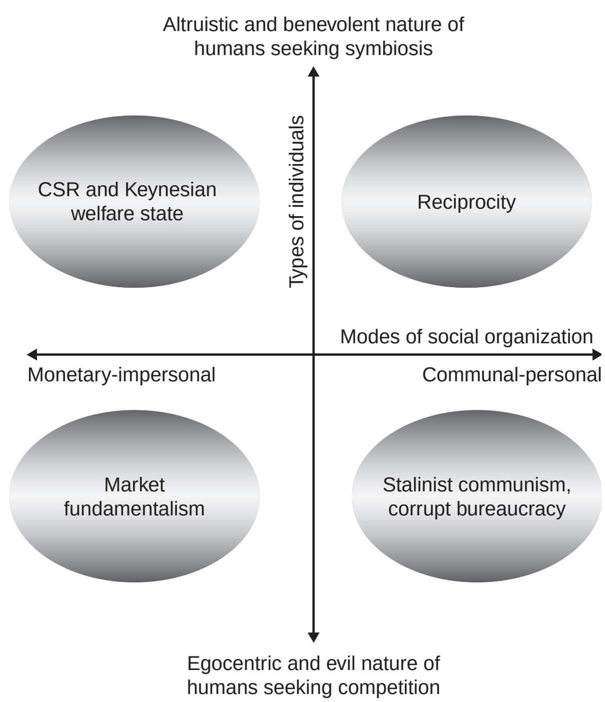

Scheme 1 Spectrum of typologies of societies and personalities from Mizuoka, F. (2006). *Globalism.* Tokyo: Hassaku-sha.

Then reciprocity comes in clear contraposition to market fundamentalism. Other possible modes of economy can be placed on the surface as shown. This scheme shows that reciprocity is the ultimate goal of the radical movements away from market economy in pursuit of alternative modes of socioeconomic organization.

Radical Political Economy and Critical Economic Geography

As Harvey invited geographers to learn, physical pristine space has two attributes: absolute and relative. Initially, they are formally subsumed. Although Smith asked Peet to 'leave aside', pristine absolute space is indeed essential to geography, as it provides the economic agents and substances with operational space, while its contagious ness associates anything in space physically, which may contradict with the requirement of a mode of production to separate something from some other things. This contiguity thus needs to be bounded to the dictate of the mode of production, so that society and economy operate without spatial interference. This process of producing space, which can be understood in the frame of Marx's concept as 'real subsumption', creates boundary and concomitantly a layer of fragmented territories. Spatial fix is possible because of this bounding and fragmented macroeconomies are overlapped with layers of higher order territories.

Pristine relative space, on the other hand, provides points of interaction. Formally subsumed, different points are separated with physical distance over two dimensional spatial expanse, which contradicts with the requirement of society and economy to become an in tegrated whole. This separation is to be overcome by the process of real subsumption, with spatial integration, or production of a physical network of transport and com munication consisting of one dimensional lines. The bounded and fragmented spaces or the territories and interconnected and integrated space came out of the process of interaction between thus subsumed pristine space and economy/society.

Uneven geographical development is the outcome of the processes of space subsumption. Economic geog raphers have much interest in global economy, mainly through research of the multinational corporations. Conventional economic geographers have deployed lo cation theories and regional economics, most of which have classical or neoclassical base. It has assumptions of atomic and rationally behaving economic men devoid of classes achieving the conflict free equilibrium. It asserts that this rationality holds eternally, at any historical conjuncture. It can become more critical by incorpor ating the agenda of radical political economy: crisis theories, neocolonialism, and dependency, as well as the global capital–labor relations. Critical economic geography, sharing the same theoretical base as radical political economy, attempts to construct 'middle range theory' which bridges abstract radical political economy and the realities on the Earth's unevenly produced surface.

On the other side of the bridge, there are radical political economists. They share the same interest with economic geographers in uneven development of global economy; yet they mainly draw upon dependency theory. They view underdevelopment as horizontally transposed capital–labor relations between countries: the metropole exploiting the satellite. Emmanuel attempted further to prove 'unequal exchange' between developed and underdeveloped countries in drawing upon production price concept that Marx had put forward in volume 3 of *Capital*. The advanced capitalist countries are the 'global capitalists', and the countries suffering from underdevelopment are 'the global proletariats'. Viewing the capitalist class relation sideways, those radical political economists proposed for the solution for underdevelopment issues to disassociate from the global capitalist system, and to develop instead endogenously on their own resources and market. They also expected 'global revolution' where countries suffering from underdevelopment seize triumph over the advanced capitalist countries and invert the global relation of production altogether. The diplomatic orientation of the People's Republic of China in 1960s was along this line.

Ironically, dependency theorists tend to view each macroeconomy as homogenous. In reality, a country with a large city of global importance may indeed have in the countryside poverty stricken villages. A global city may have slums or ethnic areas within itself where the people from the poor periphery flow in. Scott showed through his empirical study of Los Angeles that physical distance and fragmentation of space have a profound impact on labor market segmentation. The different social groups, positioned in different parts of the city, have different powers to overcome distance. Those engaged in the primary labor market live in more upmarket urban districts, and are ready to commute over longer distances. The same applies to women and minorities who are placed in disadvantaged spatial positions.

This fallacy of dependency theory extended, the radical political economists make light of domestic class relations and spatial segregations, and end up supporting the native capitalists exploiting local labor often more harshly than the capitalists at the core. The emancipation of the masses in the developing countries and the urban poor in the developed countries often drop out of their sight.

Critical economic geography was, in its earlier stage of development, also influenced by dependency theories, and regarded the hierarchical spatial relations just as a surrogate of the capitalist class relation. Smith was enthusiastic in redressing this fallacy in his rebutting against Peet's 'spatial dialectics' which treated space just as a set of containers.

To discover real interaction between 'vertical' class relation and 'horizontal' spatial relation is not an easy task as one might imagine.

To begin with, economic geography gives a more elaborated logic and the empirical patterns for the virtically integrated territories of the action space of multinational production and financial corporations. They produce the hierarchy of the layers of territories spreading across the globe, not necessarily those of countries, in their process to maximise their revenue.

In the decolonization process after World War II, former hegemony of colonial powers waned and was transformed to economic hegemony. The NIDL played a crucial role in this transition, NIDL after the Vietnam War. Low wage labor that the multinational capital sought functioned to counter 'profit squeeze' crisis at core countries. Certain developing economies thereby managed to grasp opportunities for development under 'peripheral Fordism'. The NIDL also functioned to counter the law of tendential fall in profit rate, as higher exploitation in periphery worked as a countervailing force.

As the NIDL develops, the capital has been strengthening its grasp on the labor. The privilege that labor in the 'core' countries enjoyed during the Fordist era has been eroded away by the threat from entrepreneurs that they will relocate their plants overseas where wages are cheaper, due to bounded labor mobility unless the labor accepts less favorable contracts. As Harvey suggested, space thus becomes a powerful weapon for the capital to counter class struggle waged in a particular locality. Overseas capital investment of the multinationals thus creates its own uneven spatial configurations both within and across macroeconomies in order to carry out the global exploitation of labor.

Soja also attempted to redress the dependency fallacy in geography. He stressed that space is in itself "political, strategic and ideological," maintaining dialectical relationships with society. He put forward that the spatial relation should not be interpreted homologically using the concept on social relations. To identify space as an active moment in building up and perpetuating the capitalist mode of production has become an important agenda of critical economic geography in the 1980s.

Harvey demonstrated a trenchant theory demonstrating how economic geography is intertwined with crisis prone processes of capitalist economy. He divided the crisis theory into three 'cuts', and incorporated the geographical elements to all of the three. In the 'first cut', he pointed out that value of commodity, a fetishized market relation, is sustainable only when the commodity fills social demand and thereby the commodity circulates through market transaction. Lacking demand, there shall be devaluation or decline in value embedded in the commodity, and restructuring and redevelopment of the built environment shall ensue. The 'second cut' crisis is generated through speculation in financial sector, where excess money capital is converted into fictitious capital. Speculation severely disturbs the claimed market rationality in resource allocation. This includes disturbance in land use coordination that ground rent will perform in the property market, resulting in disordered configuration of urban built environment. The 'third cut' is to do with transfer of crisis situations across different bounded territories and concomitant uneven spatial development, which is, interestingly, somewhat akin to Peet's 'spatial dialectics'. Therefore, the class holding control over the scale of higher order territories has power to dominate in society, because they can organize a 'spatial fix', or dumping of the crisis elements to other greenfield territories, at their own will.

Harvey also identified 'secondary circuit of capital', where the built environment plays a crucial role to abate crisis and make capitalism more 'sustainable'. The production of the built environment necessitates huge amounts of investment without short run profitability. The fund for investment is supplied from overaccumulated capital collected publicly and privately through tax, equity, and banking systems and postpones the outbreak of crisis.

The overproduction theorem of crisis intertwined with the study of multinational corporations can explain why neoliberalist structural adjustment made at every corner of the world does not bring about 'global convergence'. When the capacity of production has already hit the ceiling of demand, multinationals have no need to press their spatial frontier forward. Overaccumulation of capital thus limits multinationals' NIDL frontier, beyond which poverty stricken people and debt laden governments in least developed countries groan. The uneven development between the least developed countries and developing countries that succeeded in hosting some multinationals will widen, rather than converge.

As the crisis of overaccumulation becomes the norm, 'second cut' crisis becomes ever more apparent. It moves across the globe under the dictate of multinational finance and banking capital, with their bases at a very limited number of global cities. These multinational financial capitals play the game of 'second cum third cut' crisis, by hopping from one speculation venue to another to maximize their capital gain. It is totally detached of stable and real production processes, and as the Asian economic crisis of 1997 well demonstrated, it has enough power to wreck havoc of macroeconomy in developing and former socialist countries at its own whim. The repeated crisis of different 'cuts' shall lead to building up of global crisis, as Harvey pointed out.

These conceptual developments of critical scholarship in geography after 1980s in particular have successfully saved economic geography from former second rehash of radical political economy, and firmly established themselves within the scholarship of radical critical economy with their own academic merit. Yet, these contributions could be put together in a more systematic frame of critical economic geography explaining the space–society relationship more systematically. This could be done, in drawing upon Marx's dialectical concept of formal and real subsumptions, which constitute a step of the logical ascent in Marxian economics.

Conclusion

Radical political economy has existed in parallel with conventional economics for more than one and a half centuries. It has taken pains to sublimate the moral critique of capitalism into the robust theoretical bodies of economics critically analyzing the fundamental contradiction of market economy and capitalism, with their eventual transcendence and establishing a communal alternative in organizing the division of labor in view. Critical economic geographers have made considerable contribution to enrich and elaborate the theoretical import of radical political economy; while economic geographers indeed need more trenchant study of these theoretical bodies in order to conduct research on the issues of contemporary spatial configurations and their uneven development on critical perspective, with robust and causal explanatory power, as Harvey demonstrated to us.

See also: Anarchism/Anarchist Geography; Capital and Space; Capitalism; Capitalism and Division of Labor; Core-Periphery Models; Critical Geography; Critical Geopolitics; Dependency; Dialectical Reasoning and Dialectical Materialism; Economic Geography; Economies, Alternative; Gated Communities/Privatopias; Globalization and Transnational Corporations; Harvey, D.; Japanese Geography; Marxism/Marxist Geography I; Moral Economies; Neocolonialism; Public Space; Radical Geography; Regional Inequalities; Segregation; Social Class; Socialism; Space I; Structural Marxism; Territory and Territoriality; Transnational Corporations in Developing Countries; Uneven Development; Uneven Regional Development.

Further Reading

Attewell, P. (1984). *Radical political economy since the sixties: A sociology of knowledge analysis*. New Brunswick: Rutgers University Press.

Braverman, H. (1974). *Labor and monopoly capital: The degradation of work in the twentieth century*. New York: Monthly Review.

Emmanuel, A. (1972). *Unequal exchange: A study of the imperialism of trade*. New York: Monthly Review.
Frobel, F., Heinrichs, J. and Krey, O. (1980). *The new international division of labour*. Cambridge, UK: Cambridge University Press.
Gillman, J. R. (1957). *The falling rate of profit*. London: Dennis Dobson.
Harvey, D. (1982). *The limits to capital*. Chicago: The University of Chicago Press.
Harvey, D. (1985). *The urbanization of capital*. Baltimore: The Johns Hopkins University Press.
Hilferding, R. (1981). *Finance capital: A study of latest development of capitalist development. Translated by T. Bottomore*. London: Routledge and Kegan Paul.
Lenin, V. I. (1969). *Imperialism, the highest stage of capitalism: A popular outline*. Beijing: Foreign Languages.
Luxemburg, R. (1973). *The accumulation of capital: An anti critique*. New York: Monthly Review.
Marx, K. (1977/1981). *Capital, 3 vols. Fowkes, B. and Fernbach, D. (trans)*. New York: Vintage.
Mizuoka, F. (2006). *Globalism*. Tokyo: Hassaku sha.
Peet, R. (1981). Spatial dialectics and Marxist geography. *Progress in Human Geography* 5(1), 105 110.
Polanyi, K. (1944). *The great transformation: The political and economic origins of our time*. Boston: Beacon.
Sayer, A. (1995). *Radical political economy: A critique*. Oxford: Blackwell.
Scott, A. (1988). *Metropolis: From the division of labor to urban form*. Berkeley: University of California Press.
Smith, N. (1981). Degeneracy in theory and practice: Spatial interactionism and radical eclecticism. *Progress in Human Geography* 5(1), 110 116.
Soja, E. (1980). 'Socio spatial dialectic'. Annals of the Association of American Geographers 70(2).

Relevant Websites

http://www.urpe.org
Union for Radical Political Economics.

Railways

J. Shaw, University of Plymouth, Plymouth, UK
I. Docherty, University of Glasgow, Glasgow, UK

Glossary

Differential Collapse in Time–Space Suggests that the process of time–space convergence is geographically uneven, meaning in practice that improvements in transport do not reduce journey times between all places by the same proportion.

EU 15 The 15 members of the European Union before its recent enlargements first to 25 member states in 2004 and then to 27 member states in 2007.

Multimodal Transport Strategy An overarching approach to transport which uses the major transport modes – train, bus, car, air, walking and cycling – together, preferably working to each other's strengths, rather than favoring one single mode such as the car.

New Mobilities Paradigm A recent and very marked upturn in the significance of the study of mobility – for example, how, why, and where people move – in the social sciences.

Railway Mania A period of British railway history in the mid-1800s where the construction of new railway lines burgeoned.

Severance The splitting into two of an existing place by the construction of new transport infrastructure.

Time–Space Convergence The idea that places effectively move closer together as improving transport links cut both the travel time and cost of travel between them.

Transit-Oriented Development The investment led by the construction of a public transport route through an area. TOD is particularly evident around station sites.

Travel Space The interior of the vehicle in which a person travels, or the area(s) occupied within that vehicle.

Introduction

Railways, like transport more generally, are inherently geographical, although this is not necessarily a conclusion one might draw from scanning the contents of geography books and journals. While the new mobilities paradigm has 'sexed up' the topic in recent years, transport has all too frequently been regarded by mainstream human geographers as a marginal, if not faintly tedious, line of inquiry. Although there are notable exceptions, transport issues are often conspicuously absent from the analyses of economic, political, and social geographers. Transport geographers themselves have been partly to blame for this. Again there were exceptions, but many remained silent in significant theoretical debates and continued to work within the generally empirical tradition associated with the quantitative revolution and the cognate discipline of transport studies.

Transport networks connect places together by enabling movement across space, and almost all realms of human activity depend on this mobility to some extent or another. Like the activities they underpin, transport networks are borne of the societies they serve, and so it is no surprise that they reflect the social and economic values of these societies in both the geography of the networks themselves and their influence over unfolding geographies of social and economic development. Railways, especially given their fixity, are surely among the most significant of the transport networks, and with reference to the developed world we suggest that their relationship with geography might – to date, at least – be broken down into three broad stages.

Initially, whereas improvements in maritime transport played a significant role in promoting interactions 'between' places separated by the sea, the rapid development of the iron roads (as they are called in many languages) instigated massive change in the nature of interaction 'within' those places. Being capable of reaching hitherto unimaginable speeds, trains brought an increasing number of settlements closer together, and the advent of the railways represented the first real episode of time–space convergence using mechanized transport on land. As the geographical reach of the rail network grew, so the geography of many established activities was radically affected.

But such circumstances tended not to last, and the influence of railways waned as, ironically, the very spatial forms and patterns they helped to produce rendered them ill equipped to deal with new transport and communications needs. Increasingly complex patterns of human activity, in addition to related factors such as rising prosperity and changing sociopolitical expectations, often favored more flexible and quicker modes of transport such as the car and the airplane, and railway networks in some places shrank considerably. In the third stage, however, the potential for something of a railway renaissance has appeared as challenges to the hegemony of car and air transport have arisen at a range of spatial scales.

Our article considers these three main stages drawing mostly upon examples from Europe and the United States. It is worth pointing out that the nature and extent of the relationship stages we identify vary over time and across space, but the general pattern is well exemplified by a discussion of the birth, decline, and rebirth of European and American railways over the last 180 or so years.

The Rise of the Railway

Railways, or at least their forerunners, are actually a lot older than is often thought. Prepared rutways with a uniform gauge have been traced back as far as the ancient civilizations of the eastern Mediterranean and Mesopotamia, and grooved roads with passing loops existed in classical Greece. Wooden wagon ways and iron track ways – now with raised rails rather than grooves in the ground – became commonplace in European mines after the Renaissance, and in Britain particularly they extended from pits to wharfs to carry coal for shipment by boat. While such developments eased the extraction of minerals and helped them on their way to market, they were local in scale and did not link centers of population or facilitate the rapid movements of goods or people.

The first to do both of these things was that opened in 1830 between the English cities of Liverpool and Manchester (a former chairman of British Rail, Sir Peter Parker, remarked that "the world is a branch line of the pioneering Liverpool to Manchester railway"). Despite the death at the opening event of a parliamentarian who fell under a train, the route very quickly became successful. The subsequent geography of the network in England and indeed the British Isles more generally was determined by need, although this was both real, as in the example of the Liverpool to Manchester above, and perceived. Some promoters, eager to make their fortune through building a railway line, advanced outlandish schemes connecting small places unlikely ever to generate much in the way of business. As railway mania swept Britain the network eventually reached around 23 000 miles before contracting in the twentieth century. The British government was reluctant to intervene in the affairs of the railway – and certainly could not countenance producing anything like a strategic plan capable of ensuring a sensible and coherent network – as the dominant political philosophy was very much *laissez faire*.

In part, this explains why Britain's railway lines do not always take the most direct route between their origin and destination points. (The inescapable reality of the country's physical geography should also be taken into account here; a famous example of a 'diverted' line includes that between Exeter and Plymouth, in the southwest of England, which runs along the coast to avoid the difficult terrain of Dartmoor.) Each railway line required an Act of Parliament to authorize its construction and wealthy landowners whose property was to be taken by the new enterprises were seldom enthusiastic supporters of train travel, at least in the beginning before they realized they could extract a very high price for relatively small slivers of land. For example, surgeon Sir Astley Cooper who owned land in Berkhamstead through which the London and Birmingham railway was to pass, complained that the proposal was "preposterous in the extreme," and was at a loss as to why "our estates" had to be "cut up in all directions for the purpose of making an unnecessary road." Perhaps betraying his true motivations, he concluded that "if this sort of thing be permitted to go on, you will in a very few years destroy the *noblesse*."

Elsewhere in the world, railways were also appearing rapidly. Construction was proceeding at an astonishing rate in the United States, where by 1860 in characteristically free market conditions 30 626 miles of track had been completed – three times more than in Great Britain – in what is now the east and midwest of the country. (The American government subsequently played a significant role in helping to finance the great Transcontinental railways – the first of which was completed at Promontory Summit, Utah, on 10 May 1869 – through loans and land grants. This proved astute. Although a financial scandal in 1873 meant that loans were only ever provided to the companies meeting at Promontory Point, these were repaid in full with interest. Land close to the railways which the Federal government had retained was also sold at twice the usual $1.25 per acre.) Governments in mainland Europe tended to pursue a rather different approach, however, and the notion of centralized planning was very much more in evidence. In Belgium, recently split from the Netherlands, the network was completely planned by the state as a political expression of territorial unification and control. The Prussians, Russians, and Austrians planned for national welfare and military purposes, and in France companies were granted concessions to establish regional monopolies. Notwithstanding the virtually unregulated domestic market, railways in the British colonies were usually constructed in a centrally planned manner.

But it was not only that existing activity patterns within countries determined the geography of railway networks. Very quickly it became evident that the relationship also worked the other way around. Urban geography was very noticeably altered by the arrival of railways in at least three ways. First, entire new urban areas were created, sometimes as a result of the need for servicing or train building facilities (such as the railway works at Crewe), and sometimes as a result of railways bestowing new functions upon an existing small settlement (such as the siting of a dock for Atlantic cruise ships in Southampton). Second, they necessitated the demolition of large swathes of conurbations to allow for the construction of new lines heading toward city centers. In Glasgow, at least 20 000 people were displaced and the figure in London was closer to 120 000. Third, new suburbs were also created as lines extended beyond existing city boundaries and enabled commuting from former rural areas. An excellent example of this is Metro Land, the residential sprawl associated with the construction of Metropolitan Line north west out of London (**Figure 1**).

A new economic geography was also promoted as the railway changed established patterns of supply and demand. As the economy expanded in general, huge benefits were reaped by traders whose access to good railway links enabled them to obtain cheaper coal or transport their goods to new and larger markets. Such advantages were not confined to large scale industrial concerns transporting goods in bulk. Farmers able to transport livestock to market by train rather than on foot, for example, gained considerably because cows weighed more at the end of their journey. Early examples of footloose industry sprang up, as enterprises previously rooted to locations in the vicinity of their market – such as those producing perishable goods – could relocate as a result of greatly improved shipping speeds. At the same time, certain enterprises lost out. Particularly badly hit were those providing outmoded services such as staging post inns, or those firms now outcompeted by more efficient concerns in distant cities whose goods had previously not penetrated local markets. Others lacked easy access to a decent railway connection – a form of differential collapse in time–space – although the number in this position diminished as the railway network expanded.

In social terms, arriving as they often did along with the telegraph, the railways enhanced intercity and interregional communication. We have already seen that

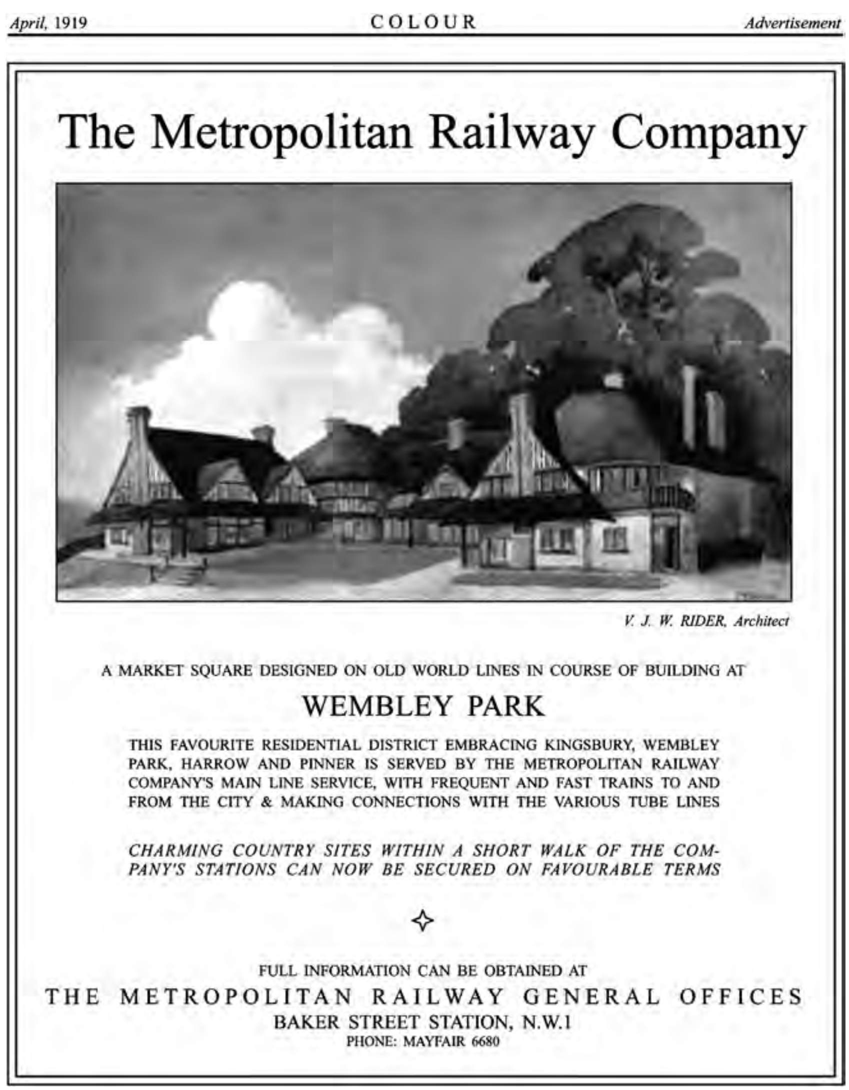

Figure 1 Advert for Metroland in 1919.

the Belgian government used railways as part of a nation building exercise (and the iron road would play a significant part in building other nations, not least the United States). Tourism also became more easily possible and resort towns flourished after the arrival of the railway and enabled nineteenth century city dwellers to luxuriate in what was being aggressively promoted as the healthful properties of the sea, the mountains, or the countryside in general. A final effect worth mentioning here is the role played by the railways in the standardization of time. Railway operations were often hampered by the significant variation of local times prevalent in places relatively close to one another. Noon in Chicago, for example, was 11.27 am in Omaha, 11.50 in St Louis, 12.09 in Louisville, 12.17 in Toledo, and 12.31 in Pittsburgh, so the American Railway Association divided the United States into four time zones from 18 November 1883. Although this initiative was generally accepted by the public it was not adopted by Congress until 1918.

A Mode in Decline?

The decline of railways' significance in society, a process which began in Britain and other developed states in the early to mid twentieth century, has varied from country to country. Generally speaking, railway systems remained important for longer in those places where private transport did not quickly assume prominence in transporting people and goods. For example, in many of the former socialist regimes of Central and Eastern Europe, which emphasized community and discouraged individual expression and entrepreneurialism, rail networks remained highly important as few private citizens of these countries owned or had access to a car. Often train services complemented urban metro, tram, and/or trolleybus networks to function as part of an integrated transport system (a concept to which we shall return). In the West, however, a political ideology based around capitalism and freedom, combined with associated factors such as growing prosperity, restructuring economies, rising expectations, and rapid technological innovation, served to promote the ascent to prominence of cars, lorries, and later, airplanes.

Road transport – so long the poor relation to the technologically superior rail system – flourished as the benefits of its flexibility could be exploited more and more to satisfy the ever expanding transport needs and desires. Very large road building projects were undertaken across Europe and North America, and car ownership grew as the relative cost of owning and running a car reduced; indeed, in some countries it is still falling (**Table 1**, **Figure 2**). Rail use, both as a proportion of all journeys made and in some cases in absolute terms, began to fall away and is now relatively stable at a low level (**Figures 3a** and **3b**). In those countries which came to value the neoliberal notion of individual liberty the most – usually the Anglophone world – the dominance of the private car relative to other transport modes became all the more acute. Because the car allows people mobility where and when they choose, politicians and indeed the public at large came to view it as a symbol of freedom and, by some in the United States, of ideological superiority over the Soviets. Over long distances air travel has all but replaced the train in countries such as the USA, Canada, Australia, and New Zealand.

The impact of all this on the geography of the rail network was significant. In the United States, passenger rail routes were abandoned in favor of freight only services on many corridors (rail freight remains attractive in America because of the long distances between origin and destination points). In Britain, the rail network shrank significantly from a peak of 23 000 miles to under 12 000 today, and only around 7% of passenger and freight journeys are now made by rail. Bus patronage also declined and the reduced availability of public transport in some areas has increased the social exclusion of groups such as the elderly and the financially disadvantaged. Former socialist countries are also now experiencing surges in car ownership, and the viability of elements of their rail networks may well be threatened in the future.

It is perhaps – especially with hindsight and from our cultural standpoint – entirely unsurprising that this course of events transpired given the inevitable march of 'progress', but it has had an interesting geographical consequence. Ironically, developments over the past few decades have led to the point where those spatial structures the railway helped to create now serve to undermine its erstwhile dominance in the transport market. Processes of large scale dispersal initially made possible by the railways – suburban living, footloose industries, and so on – have been reproduced on a much greater scale as a result of the flexibility afforded by cheap 'private' transport. The very complexity of modern day human existence effectively rendered obsolescent much of the inflexible and expensive 'public' railway (and more broadly, transport) network, and trains have retained significance only in a few markets such as large commuter networks and heavily used city center to city center corridors. Yet 'progress' – or at least its externalities – continues to influence the relationship between geography and the railways, and a clearly emerging third stage is now what we turn to discuss.

A New Railway Mania?

While cars and airplanes bestow many benefits, it has become evident that the volume of traffic in much of the world has risen to a level where these advantages are

Table 1 Increase in population, number of drivers, and vehicles in the United States, 1960 2003

Year	*Population (m)*	*Drivers (m)*	*Motor vehicles (m)[a]*	*Increase in vehicles (m)*	*% Growth*
1960	180	87	74	N/A	N/A
1961	183	89	76	2	+2.72
1962	186	91	79	2	+2.63
1963	188	94	83	3	+3.79
1964	191	95	86	3	+3.61
1965	194	99	90	4	+4.65
1966	196	101	94	4	+4.44
1967	197	103	97	3	+3.19
1968	199	105	101	4	+4.12
1969	201	108	105	4	+3.96
1970	204	112	108	3	+2.85
1971	207	114	113	5	+4.62
1972	209	118	119	6	+5.30
1973	211	122	126	7	+5.88
1974	213	125	130	4	+3.17
1975	215	130	133	3	+2.30
1976	218	134	139	6	+4.51
1977	220	138	142	3	+2.15
1978	222	141	148	6	+4.22
1979	225	143	152	4	+2.70
1980	227	145	156	4	+2.63
1981	230	147	158	2	+1.28
1982	232	150	160	2	+1.26
1983	234	154	164	4	+2.50
1984	236	155	166	2	+1.21
1985	239	157	172	6	+3.61
1986	241	159	176	4	+2.32
1987	243	161	179	3	+1.70
1988	246	163	184	5	+2.79
1989	248	166	187	3	+1.63
1990	248	167	189	2	+1.01
1991	252	169	188	1	0.52
1992	255	173	190	2	+1.06
1993	258	173	194	4	+2.10
1994	260	175	198	4	+2.06
1995	263	177	202	4	+2.02
1996	265	180	206	4	+1.98
1997	268	183	208	4	+1.94
1998	270	185	208	0	0
1999	273	187	216	8	+3.84
2000	281	191	218	2	+0.92
2001	281	191	226	8	+3.66
2002	288	195	230	4	+1.76
2003	291	196	231	1	+0.43

[a] Includes passenger cars, motorcycles, and other two axle, four tire vehicles.
Source: United States Department of Transportation.

being compromised by the environmental, economic, and social impacts of driving and flying. Such impacts manifest themselves, broadly speaking, across three spatial scales. At the local level, concerns about noise pollution and deteriorating air quality along main traffic corridors and close to airports have come to the fore. We have already referred to a decline in public transport services. And despite significant improvements, road safety remains a problem: on average 110 people per day are killed on the roads of the EU 15 as a result of vehicle crashes (**Figure 4**).

Regionally and nationally, economic efficiency is being hindered by worsening levels of congestion on key traffic routes. The main problem here is that time spent in a traffic jam imposes a financial cost on those affected. Business people take longer to get to meetings; goods take longer to deliver, so more lorries and more drivers are needed to maintain a delivery schedule; leisure travelers lose time from their holiday or miss some of the cricket match they are traveling to; and so on. The cost of congestion to the British economy alone has been estimated at over $30 billion annually. (The reliability of this

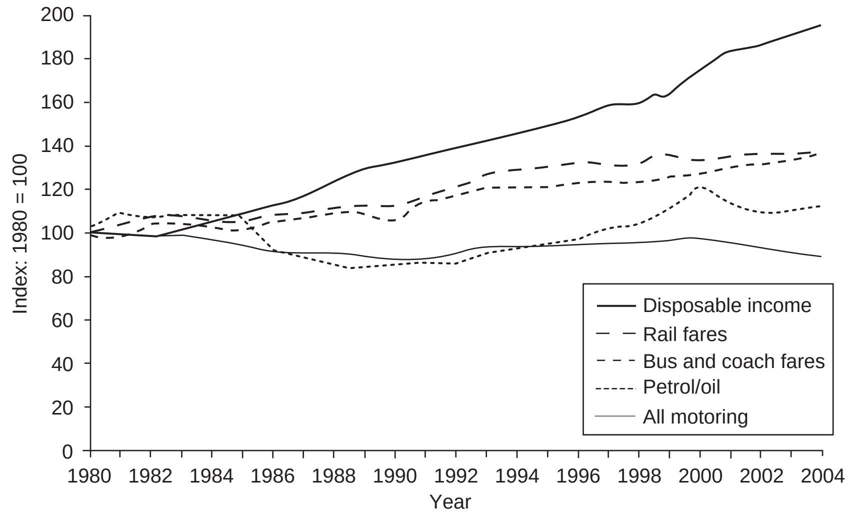

Figure 2 Relative cost of motoring and public transport in the UK, 1980 2004. Source: Office for National Statistics.

estimate is very much open to question.) At the same time, though, the amount of time people spend in transport related delays brings into focus the function of travel space. Those traveling by a mode of transport which does not require them to control their vehicle are free to use their travel space in a range of productive or enjoyable ways, such as a mobile office or leisure facility.

At the international scale the principal concern is the impact of current transport trends on global warming. Carbon dioxide from road vehicles (mainly private cars and trucks) accounts for the vast majority of transport's overall contribution to greenhouse gas emissions. Air travel is also significant. Although airplanes currently account for only a small proportion of carbon dioxide emissions, this figure is set to increase markedly as global demand for air travel continues to grow despite rising fuel prices and terrorist threats. Recent research also indicates that because aircraft emissions occur pre dominantly at altitude, their effect per unit is greater than at ground level. The cost to the global economy of damaging events associated with climate change (such as increased magnitude of violent storms, crop failure, sea level change, and so on) is likely to be extremely high.

All of this places the railways back at the heart of transport thinking because the train is among the most sustainable of all mechanized transport modes. Clearly there are negatives associated with railway operation and development – trains also take land, create noise, and produce carbon dioxide – but per passenger journey they are safer, more environmentally friendly and less land hungry than cars, and generate far less carbon dioxide at a much lower altitude than planes. They also help to reduce social exclusion and congestion, and can provide a more spacious and potentially productive travel space for business and leisure travelers alike. With the need for sustainable transport strategies asserting itself more strongly throughout the world, it seems as though the nature of the relationship between rail transport and geography is changing now. Will railway networks again expand, and can the railways once more drive spatial change on a significant scale?

Of course, current transport trends have built up over decades and will not be capable of radical change in the short term. This is for a variety of reasons including the difficulty of influencing travel behavior, the inability of existing transport infrastructure to accommodate large scale modal shift from car to public transport (although this is more of a problem in some countries than in others), and the effect of vested interests – car manu facturers, motorists, politicians, and so on – who might perceive as a threat moves to challenge the *status quo*. There are also institutional barriers to overcome, since many public transport operations have been privatized and/or deregulated in recent years. For a variety of reasons, drawing definitive conclusions from quantitative data on world railway trends is problematic, but recent developments indicate that a revival of rail transport might be taking place; three examples are given here.

High speed lines continue to be developed within Europe (**Table 2**), and in Japan a well established net work continues to thrive. Existing patterns of settlement and activity once again become of significance in deter mining the geography of the rail network – Europe and Japan are both densely populated with large cities rea sonably close together – but the geographical impact of new investment in high speed lines is also important. In Spain, for example, research on the Alta Velocita Espa ňola (AVE) lines shows that not only is the service cap able of effecting huge modal shift from air to rail, it also changes commuting and industrial location patterns. In France, the Train à Grande Vitesse (TGV) has produced

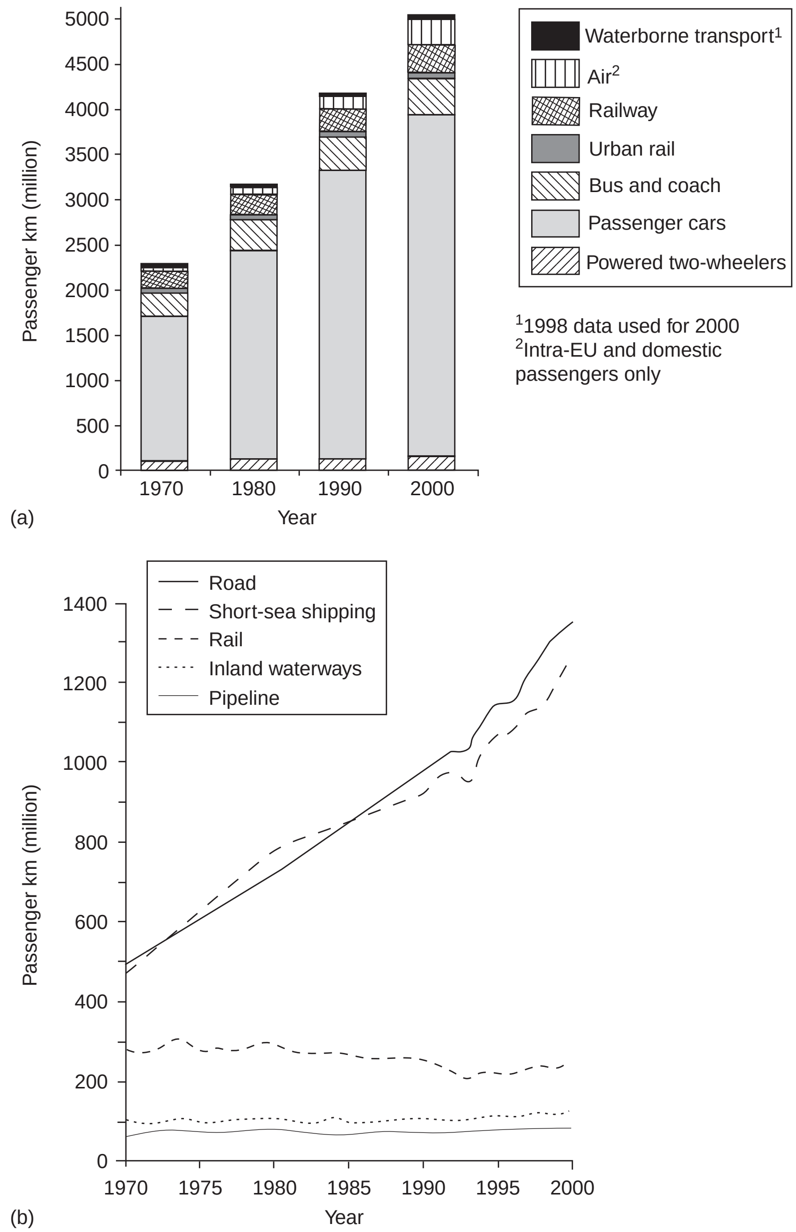

Figure 3 (a) Modal split for passenger transportation in the European Union, 1970 2001. Source: Eurostat. (b) Modal split for freight transportation in the European Union, 1970 2001. Source: Directorate General Energy and Transport.

similar results, with cities away from Paris – particularly Lille and Lyon – emerging as reinvigorated centers of commerce, leisure, and commuting. In Germany, the ICE system combines new lines with adapted existing ones both to improve rail communications between major cities in the west of the country and as a reaction to changing spatial imperatives (reflected in the need for better and faster transport to Berlin) since unification.

In the UK, where the private car has become some what more dominant than elsewhere in Europe, central and devolved governments have been attempting to pursue an integrated transport policy which revolves around a vibrant and expanding rail industry. The legacy of decades of underinvestment and an ineffectual and expensive privatization process – annual subsidy now runs at around $10–12 billion compared with $3.5 billion under British Rail – combined with ministers' fear of being viewed as 'anticar' has hampered the conversion of policy into practice, but rail patronage has still risen by 40% over the last 10 years and a high speed line con necting London to Paris and Brussels has just been completed. Despite such positives, UK government ministers remain wedded to a system of franchising de signed more to bring down expenditure than to secure large scale service improvements.

The link between spatial planning and rail transport is neatly exemplified by developments in Denver, long renowned for its urban sprawl associated with the private car. Voters have recently passed a ballot measure to build three light and three heavy rail lines by 2016, with the $5 billion cost being met from increased sales taxes. The impact of the new lines will result in a complete reorientation of established planning practice, with communities and developers already lining up to construct high density urban villages around new stations on the network. Such transit oriented development is a new concept in the Intermountain West but its impact on the geography of metropolitan Denver should be dramatic. In the words of one proponent, "the idea is that when you fly over Denver in 20 or 30 years from now, hopefully you will be able to see these dots of urban villages, and make out the skeleton of our transit system."

Thinking in terms of sustainable development, the biggest challenge will be to ensure that rail led development becomes the norm over the coming years. This challenge does not only apply to developed nations. As former socialist countries and emerging industrial powerhouses such as China and India raise their standard of living, the temptation is to follow Western models of development which privilege the private car and airplane over rail and bus transport in many cases. (China is in the process of investing considerably in high speed and other rail lines; a new Maglev line has opened in Shanghai with a second on the way between Shanghai and Hangzhou, and conventional high speed lines are under construction between Beijing and Tianjin and, subject to delay, Beijing and Shanghai.) This could have serious knock on effects for the often dense and well used rail networks already in existence. It would be a great shame to watch rail decline in such countries just as it enjoys a resurgence in others which are learning the hard way that losing a high quality rail network can prove a costly mistake.

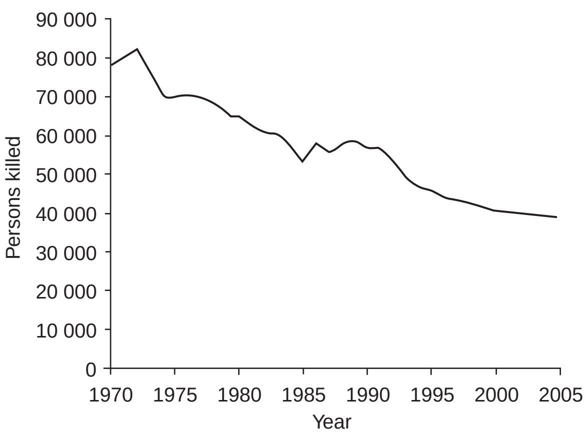

Figure 4 Number of persons killed in road accidents in the EU 15, 1970 2000. Source: Eurostat.

Conclusion

In the preceding sections we have discussed the development of railways from a geographical perspective. We have suggested that rail networks and services – and, indeed transport in general – are inherently geographical as they link together places by enabling movement across space. Our discussion has highlighted that the relationship between railways and geography is bidirectional: the geography of rail networks reflects spatial patterns in existence at the time of their construction, while at the same time those networks exert an influence over the spaces in which they operate. After a period of great significance and a subsequent decline in fortunes, the railway is once again emerging as a socially, economically, and environmentally desirable mode of transport in a range of different settings. Rail's ability to move large numbers of people in a safe and environmentally friendly way while tackling congestion and social exclusion is leading to new investment that is both expanding the geography of some networks and refocusing patterns of spatial development. Whether this trend continues to become worldwide in character remains to be seen, but so long as climate change and economic efficiency retain prominence in political discourse, rail's further ascendancy up the political agenda seems likely to continue.

Table 2 Length of the EU's high-speed rail network, 1990 2003 (km)

	Belgium	*Germany*	*Spain*	*France*	*Italy*	*EU-15*
1990				667		667
1995			377	1124		1501
1996	12	434	377	1152	237	2212
1997	71	434	377	1152	237	2271
1998	71	486	377	1147	237	2318
1999	73	491	377	1147	237	2325
2000	73	491	377	1147	237	2325
2001	73	510	377	1382	237	2579
2002	127	644	377	1400	237	2785
2003	140	645	377	1400	237	2799

A further 2500 km of lines are under construction or in planning. Source: Eurostat, 2007.

See also: Intermodality; Transport, Public; Transport and Deregulation.

Further Reading

Docherty, I. and Shaw, J. (2008). *Traffic Jam: 10 Years of 'Sustainable' Transport in the UK*. Bristol: The Policy Press.

Dyos, H. and Aldcroft, D. (1971). *British Transport: An Economic Survey from the Seventeenth Century to the Twentieth*. Leicester: Leicester University Press.

Eurostat Statistical Books (2007). *Panorama of Transport*. Luxembourg: Office for Official Publications of the European Communities.

Faith, N. (1994). *The World the Railways Made*. London: Pimlico.

HM Treasury (2006). *Stern review on the economics of climate change*. www.hm treasury.gov.uk/independent reviews/stern review economics climate change/sternreview index.cfm (accessed in May 2008).

Knowles, R. (2006). Transport shaping space: Differential collapse in time space. *Journal of Transport Geography* 14, 407 425.

Letherby, G. and Reynolds, G. (2005). *Train Tracks: Work, Play and Politics on the Railways*. Oxford: Berg.

O'Dell, A. and Richards, P. (1971). *Railways and Geography*. London: Hutchinson & Co.

Sheller, M. and Urry, J. (2006). The new mobilities paradigm. *Environment and Planning A* 38, 207 226.

Wolmar, C. (2007). *Fire and Steam*. London: Atlantic Press.

Rational Choice Theory (and Rational Choice Marxism)

G. Bridge, University of Bristol, Bristol, UK

Glossary

Bounded Rationality An understanding of rationality that takes computational difficulties, multivariate preferences, and the costs of obtaining information into account.

Cooperative Game Theory A game in which some pre-play negotiation and joint action is permitted.

Dialectics A process way of thinking in which what is apparent or manifest (in material terms – e.g., capitalism and capital) in the world has its negation (working-class formation) which opens it up for future resolution (through, for example, social revolution).

Evolutionary Game Theory Describes a dynamic process in which better strategies are more likely to be chosen in subsequent plays of the game. Strategies are initially embedded as types of players: they do not require strict assumptions about individual cognitive rationality.

Evolutionary Stable Strategy An equilibrium in which one particular strategy dominates over alternatives.

Game in Extensive Form A game tree on which all the possible sequences of choices that lead to the payoff are mapped out.

Game in Normal Form A game matrix that shows the payoffs for alternative actions.

Historical Materialism An idea of historical development that emphasizes the material basis of existence and the social relations and their tensions that are obtained in terms of the appropriation of that material base.

Imperfect Information Players cannot observe each others' moves.

Incomplete Information Players do not know each other's payoffs.

Mixed Strategy Courses of action will be chosen at random from a set of pure strategies with associated probabilities.

Nash Equilibrium A player's best response given what the other players do. A Nash Equilibrium is one in which all players are content not to change their choice of strategy given what the other players choose.

Noncooperative Game Theory A game in which no prior agreements or joint actions are enforceable.

Parametric Rationality Rational choice in conditions where the other players are nonintentional.

Payoffs The utility a player receives for a certain choice.

Pure Strategy A nonrandom course of action.

Strategic Rationality Rational choice in conditions of interdependent decision making.

Structuralist Marxism An approach that emphasizes the strict causal connections, from the material economic base of human existence to its social and cultural superstructure.

Zero-Sum Game Games in which one player's loss is another player's gain.

Conventional Rational Choice Theory

Rational choice theory (RCT) has been the single most influential model of human behavior in the social sciences but one that has had very little impact on geography. The discipline that has made the most of this approach is economics. Rational choice is the basis for microeconomic theory. Microeconomic theory starts with individual preferences as a way of explaining aggregate demand for goods in a market. In this way it focuses on preferences over goods that can be supplied in a market (from apples and oranges to various forms of healthcare, for instance). The basic assumptions of rational choice are that humans have preferences, that they are able to rank their preferences ('I prefer apples to oranges'); that their preferences do not overlap (no tied rankings). An individual's ranked preferences have different levels of utility for them. Acting rationally means acting in a consistent way – that is, to maximize utility. Individuals may have preferences for different things that are not immediately obtainable, for goods or states of being in the future. Because the future is inherently uncertain elements of chance or probability have to be introduced. Choosing about future goods or situations in a rational way means maximizing expected utility, that is, utility tempered by certain judgments (beliefs) about the likelihood of achieving/obtaining these things in the future. Traditionally economists deliberately do not take account of where people's tastes or preferences come from or how they were formed (very much the concern of sociologists). Preferences are assumed to be exogenous (simply treated as givens), although there is now a branch of economics that deals with endogenous preferences formation or how preferences are formed. In the model of preferences for goods utilities are used to measure what people want and how much they want it.

Considerations of utility maximization for individuals tie microeconomic theory closely to the philosophical traditions of utilitarianism. Utilitarianism is a philosophy originally derived for lawmakers as a framework of choice between different laws according to the maxim that the greatest happiness (utility) to the greatest number is the measure of right and wrong. This ties utility calculus to a moral theory, or alternatively moral questions are re solved through instrumental calculation.

In neoclassical economics the model of rationality is that of instrumental rationality. The equilibrium concept that attempts to explain human behavior is that derived from the marginal revolution which occurred at the end of the nineteenth century. It assumes that an actor will engage in an activity up to the point that the marginal cost of obtaining an additional unit of the good or service equals the marginal benefit or use to which that add itional unit can be put. This is Robinson Crusoe eco nomics involving isolated actors making self interested decisions. With the development of game theory since the 1940s a new form of equilibrium was developed, one that has much more purchase of social/geographical problems. Game theory models ideas of rational choice in conditions of interdependent decision making, that is, situations in which the decision of one actor is dependent on the decision of the other player(s) and vice versa. The fates of players are mutually implicated in this way. This can quickly lead to difficulties of coordination of action even when they are able to negotiate beforehand (co operative game theory) as much as when they cannot (noncooperative game theory).

Prisoner's Dilemma

Some of the dilemmas of assuming this parsimonious view of human nature is illustrated most easily referring to certain classic games. The most famous of these is Prisoner's Dilemma. This models a situation in which two people have been apprehended for a crime and are being held in separate cells. They cannot communicate with each other and there has been no prior agreement of what to do in such a situation. If prisoner A betrays prisoner B and prisoner B stays silent, A goes scot free and B gets 10 years imprisonment, and vice versa. If both stay silent they each get 6 months. If both betray each other they both get 5 years in goal. There are a number of aspects of this game that are deemed important. It is a non zero sum game. Zero sum games are where one person's gain is the other person's loss. In this game there are potential benefits to cooperation in that both would get only 6 months if they cooperated and both stayed silent. This situation is Pareto optimal, that is, it opti mizes the utility of both players without it being at the expense of the other player's utility. However it is not the choice that rational players will make. Both players are strict individual utility maximizers.

Looking at the choices available (**Table 1**) if I am prisoner A, if I assume prisoner B will stay silent, if I stay silent I will get 6 months and if I betray I will go free. Therefore assuming prisoner B stays silent my best outcome is to betray. On the other hand assuming that prisoner B betrays, if I stay silent I will get the maximum penalty of 10 years. If I betray I will get 5 years like my partner who has also betrayed. Assuming that prisoner B betrays it is better for me to serve 5 years than 10 so I will choose to betray. This means that whether prisoner B stays silent or betrays my best strategy is always to betray (and so for prisoner B). What my partner chooses to do does not affect my choice because in all scenarios my choice is dominated by the betray option. In this sense Prisoner's Dilemma is not truly strategic because my strategy choice (betray if B stays silent, betray if B be trays) is not affected by what my partner chooses. In that sense it has more in common with parametric situation in which the environment of choice can be treated as a given, or parameter. Prisoners A and B both choosing betrayal is a Nash Equilibrium (neither would want to change their strategy choice, see below) but it is not one that is based on interdependent decision making, and it is not Pareto optimal.

The point about rational self regard for individual welfare compromising collective welfare (the benefits of cooperation) has been used as an instructive idea for understanding the dilemmas of welfare economics and the use of natural resources. This connects RCT to wider concerns in political philosophy. It helps inform how social and political relations can be treated as forms of games and how interdependent human action can change if the games are repeated, (e.g., if betrayal was punished in repeated play of Prisoner's Dilemma in a form of tit for tat). Alternatively there are considerations of how insti tutional mechanisms can be forged that help change human behavior into more cooperative forms of social contract. One prominent example of this is how John Rawls's hugely influential arguments on liberal egali tarianism take assumptions about self interested rational actors and their risk averseness to explain how self re garding individuals might consent to a society that has an idea of justice as fairness. Along other routes ideas drawn from Prisoner's Dilemma involving incentives and op portunities for collective action in terms of, for example,

Table 1 Prisoner's dilemma

	Prisoner B silent	*Prisoner B betrays*
Prisoner A silent	Each serves 6 months	A serves 10 years B goes free
Prisoner A betrays	A goes free B serves 10 years	Each serves 5 years

social movements, have been pursued. These explorations reveal the difficulties of interdependent coordination.

The Nash Equilibrium

The equilibrium concept that resolved many of the indeterminacies that came from modeling interdependent choice is the Nash Equilibrium (named after John Nash the Nobel Prize winning mathematician and subject of the Hollywood movie *A Beautiful Mind*). A Nash Equilibrium is where each player is content with their choice of strategy given the choice of strategy of each of the other players in the game. Any other pair of strategy will not be an equilibrium because at least one player will wish to deviate from their choice of strategy. A strategy is a se quence of moves in a game that is dependent on how the player thinks the other players will play at each point. The simplest games in this strategic situation assume perfect information in which each player is able to ob serve the moves in sequence of each of the other players before making their next move. These situations normally lead to a single Nash Equilibrium (i.e., formalization of how individuals will behave interdependently in this situation). Game theorists also model (more realistic) situations in which actors have less than perfect infor mation informing their choice. Their information might be incomplete: where players do not know the payoffs of the other players (i.e., they don't know what type of player it is because they are ignorant of the payoffs that relate to preferences). Players might also have imperfect infor mation if they are not able to observe each other's choices. Situations like this often result in several Nash Equilibria. A classic example of this is the dating game (**Table 2**).

Phil and Leslie are partners trying to decide where to eat. All means of communication have broken down and so they must try to decide which restaurant to go to, based on what each thinks the other will choose, knowing that they are both trying to anticipate each other's choice. This is a game of complete but imperfect information. It is complete because each player knows the other's preferences (or payoffs). Phil's payoffs are represented by the first number (util) in each bracket and Leslie's by the second number. Phil prefers Italian to Chinese so gets the biggest payoff if they both meet at the Italian (2 utils to Leslie's 1 util). Leslie prefers Chinese to Italian so gets the biggest payoff if they both meet at the Chinese restaurant (2 utils, Leslie 1 util). However, both prefer to be together regardless of restaurant, such that if they fail to meet that both get their worst payoff (0, 0). However neither can observe what the other chooses (they have imperfect information). Just going for their own prefer ences is not sufficient here because they also want to meet. So what do they do? Rationality (game theory) tells them that there are two possible solutions to this game (2 Nash Equilibria) Italian/Italian or Chinese/Chinese, that is, neither will want to change their choice of strategy given what the other chooses. However ration ality cannot tell them which solution to choose. In situ ations where there are multiple equilibria something else must coordinate expectations. Game theorists refer to these coordinating devices as conventions or norms and see these collective dilemmas as situations ripe for the emergence of norms.

Table 2 The Dating game (after Luce and Raiffa, 1957)

		Leslie	
		Italian	Chinese
Phil	Italian	(2, 1)	(0, 0)
	Chinese	(0, 0)	(1, 2)

Social Norms and Power

Nash Equilibria have the characteristic that they are self enforcing. Where rationality itself cannot determine which equilibrium is chosen game theorists turn to other devices that can coordinate rational expectations. Prece dence (tradition) or some conspicuous feature in the environment can provide that hook on which future ex pectations are coordinated. It is possible to think of these as potential or proto norms. They need not be deter mined by the rightness of the norm, simply that they provide a way of coordinating when coordination is dif ficult to resolve rationally. What the Nash Equilibrium suggests is when any Nash solution is chosen it will quickly become established because of the rationally self reinforcing nature of this form of coordination. Thus, game theorists start to argue for the micro foundations of social and cultural norms. Norms are seen here as solu tions to coordination problems, that are reinforced by rationality once chosen. The way they are chosen is through a generalized expectation directly based on in dividual strategic action (unique Nash Equilibrium) or through a possible set of norms which tradition or some other focal point helps coordinate. Norms are seen as forms of rationally binding expectation that emerge from individual needs to coordinate action. From a rational choice perspective norms are coordinating devices – they may or may not be in themselves desirable (patriarchy has traditionally decided the dating game coordination, for example). It also suggests that norms by their nature have alternatives (a trivial case would be driving on the left or the right hand side of the road) and how norms are self enforcing through choice – in contrast to macro sociological approaches that see them as constraints or conditions of action. These considerations of coordination and the emergence of norms start to bring rational choice into conversation with more sociological ideas of intersubjectivity, a route that a number of political theorists have followed. Furthermore, RCT is not necessarily a theory that assumes voluntaristic choice in the sense that power can be mapped in various games through the constraints on choice or through ideas of strategic advantage for more powerful actors (first mover advantage or first call in the ultimatum game, for example).

As has been discussed conventional RCT relies on an idea of constrained optimization. There might be limits to choice (through unequal power relations, for instance) and the environment of choice might be uncertain or risky, but nevertheless RCT suggests that rational agents will optimize under these conditions. It was increasingly acknowledged that there might be limits to the computational abilities of agents in certain situations, as well as there being a cost associated with gathering information to make the choice. From this came the idea of 'bounded rationality' and that, given these difficulties, rational agents might be satisficers rather than optimizers. They might rely on simple heuristics (rules of thumb, emotional intelligence; analogs, imitation, and other cultural tools) to cut down on the informational and computational complexities. Interest in bounded rationality has grown significantly in recent years and starts to bring rational choice to consideration of practical rationality rather than exhaustive cognitive reasoning.

Relaxing the criteria of rationality is taken one big step further in evolutionary game theory. Here the underlying model is not physical mechanics and equilibrium but evolutionary biology. Evolutionary biologists used game theory to help explain how genetics (genotypes) led to a variety of behavioral types within and between species (phenotypes) and how the interaction between phenotypes led to certain behavioral types winning out over others. When one phenotype became dominant and could withstand the interactions of all other phenotypes it came to dominate in what was called an 'evolutionary stable strategy'. Economic game theorists borrowed back these insights from biology to try to explain the dominance of certain behavioral strategies over repeated interactions in a human population. This idea completely removes the need for an assumption of individual cognitive rationality. Strategies are embedded in people's minds (as a result of socialization, cultural norms, instilled experience) and the most successful strategies win out in interaction (with some possibilities of cognitive learning and adaptation by both successful and less successful players between rounds of interaction). Evolutionary game theory is a rapidly growing area of RCT.

Rational Choice Theory and Geography

Outside economics RCT has had considerable impact on political science and has formed a substream of enquiry in sociology (in the latter case considering such important issues as class differentials in educational outcomes). RCT has had limited impact in geography. There were some attempts to introduce game theory into the discipline in the early 1960s. One such was Peter Gould's argument in his 1963 paper on game theory and human/environment relations. Here Gould explored a range of situations in which farmers had to choose strategies over crop planting given the probability of wet or dry growing seasons. The examples suggested the best mix of crops to maximize utility (yield) over the medium to long term. This is a classic case of parametric rationality in which one player (the farmer) plays a limit or parameter (the environment) which can be treated as a nonintentional agent that presents them with a series of probabilities of actions (wet or dry seasons) which affect the calculus of the farmer in trying to maximize his/her expected utility.

Conventional RCT reached had its greatest influence in geography during the Quantitative Revolution of the 1960s and 1970s in the application in economic geography particularly in the heyday of location theory. Here general equilibrium theory (which tries to explain the relationship of supply and demand across a whole economy with different markets) was used to provide an approach to industrial location theory. Utility theory was incorporated into demand behavior by adapting Christaller's earlier 'central place theory' to suggest that in spatial terms, rational consumers work on the principle of least effort, or distance minimization to purchase/consume a good. These assumptions were relaxed somewhat in behavioral geography one branch of which, through the work of Allan Pred, Julian Wolpert, and others, looked at individual decision makers as satisficers, rather than optimizers. Interestingly although the influence of behavioral geography waned, as we have already noted applications of the ideas of bounded rationality, satisficing, and rules of thumb in decision making have grown considerably in economics in recent years.

Despite these important interventions, overall within geography RCT and its assumptions have usually been the subject of critique. The usual critiques are that RCT is too reductivist by focusing on utility maximizing individuals and that it ignores cultural and geographical context. It has been dismissed on any of the following grounds: it is a 'thin' theory of human motivation; it assumes selfish motivations in actors; it is individualistic to the point of being atomistic; it assumes that human beings are perfect calculating machines; it is sociologically naive in that it takes peoples' preferences for granted and does not inquire into how they got those preferences. RCT is also seen as part of the imperialism of economics in social science because it applies the assumptions of 'homo economicus' to life situations outside exchange in markets. In terms of treating rational coordination as the foundation for institutions the difficulty of achieving collectively rational outcomes from individual decisions however rational they may be has been outlined in a number of ways (through ideas of perverse or aggregation effects). As sociologists have long noticed repeated actions can lapse into habits and thus much of what passes for action or choice might indeed be made unconsciously, rather than in the full glare of clear reasoning and perfect information. Finally, by treating utility as a completely open category that can be treated tautologically (utility is just something you maximize) starts to suggest that you can be rational about anything (including things that are highly undesirable). By confining itself to ideas of social norms merely as forms of coordination RCT is criticized for paying too little attention to values and normative questions. There have been attempts to bring the idea of rationality into normative questions in terms of rationality as a form of communication and consensus over norms (especially through the work of Jurgen Habermas). On the other hand, post structuralists such as Michel Foucault see rationality as bad normativity in the way that it regulates life. It operates as a powerful discourse that categorizes and constructs subjectivities.

Some of the critiques of RCT are powerful and some seem misplaced. Many target the idea of homo economicus, but this is the rational agent of parametric rationality and marginal utility. The situations of interdependent choice studied by game theory and for which the Nash Equilibrium is the core concept are very different and in which the nature of the social situation of choice and of other agents is vital. The criticism that these agents seem to come ready made with preferences and clear reasoning and that this ignores the social and geographical context in the formation of subjectivity is well made. There may be a division of labor in as much as rational choice concerns itself with aspects of human behavior in social and cultural settings in which forms of future coordination are important. There is growing evidence of a conversation in political science at least between ideas of strategic rationality in game theory and treatments of rationality from critical theory. An earlier and rather striking juxtaposition of RCT and critical theory came in the form of rational choice Marxism.

Rational Choice Marxism

RCT assumes that the larger structures of society and economy can be explained by assuming the rational behavior of individuals or institutions. Rational choice Marxism applies some of these premises to Marxist thought. This approach had its heyday in the 1980s and consisted of subjecting Marxist ideas to the logical and parsimonious analysis that characterizes wider analytical philosophy. Analytical Marxism represents a departure in that a group of scholars applied the analytical approach sympathetically to Marxism in various ways. Rather than relying on the premises of dialectical or structuralist thinking rational choice Marxism assumed that the rationality postulate could weed out the inconsistent or unrigorous elements of Marxist thought and provide a mechanism of explanation for what remained. Thus historical materialism could be supported from the point of view of being a rational process of response to technological change and the rational impulse to maximize productivity for least effort. Equally the labor theory of value (value only being represented by congealed labor from production embedded in the commodity) could be abandoned. Exploitation arose not from surplus value extraction in the process of production but could occur between owners of capital who had different assets in exchange relations. Equally exploitation could be seen as a form of rational coordination in which high asset players structured the game such that withdrawal for players (e.g., workers) who had an alternative and an incentive to do so was made prohibitively costly. That is, the structure of the game made it rational to choose to stay in the game (in ways that coincide with the analysis of power in conventional rational choice through such mechanisms as first mover advantage, dominated strategies, and the ultimatum game).

In contrast to structural explanations that argued that there were economic structures that produced social relations, or dialectical relations that operated in ways independent of conscious thought, rational choice Marxists suggested how these social relations of production could come about through the conscious choices of individual actors, including the mass of those who were being exploited. In this way rational choice Marxism comes close to conventional rational choice in that the focus is on rational choices within constraints that result in nonrational aggregation effects or suboptimal outcomes across the population.

Rational choice Marxism has been subject to some of the same critiques as conventional rational choice. It is seen as having a thin view of human motivation and is overly reductivist by not explaining the social and geographical context of rational individuals. The argument has been that space and place unsettle the conclusions of the reductivist analytical approach in terms of its assumptions about historical change, the possibilities of collective action, and the spatial elements of asset distribution and exploitation. In addition to these standard critiques it has been argued that rational choice Marxism is simply outside the spirit of Marxism itself. Marxism is seen as a philosophy of praxis (of action) in contrast to the abstract or spectator view of affairs from analytical Marxism. Rational choice Marxism treats Marxist ideas not as tools to change the world but rather concepts that require dry and logical examination to see whether they pass analytical muster. In terms of geographical questions, like conventional RCT, analytical Marxism is seen as being out of step with the more contextualized concerns of place and interspatial relations that charac terize the accumulation of value in capitalism.

Although Marxist theory has been very influential in geography over the last 30 years, the rational choice version has had very little purchase on geographical thinking. That is in part because the broader ideas of analytical philosophy itself have had little impact on geographical ideas. Early Marxist interventions in geog raphy were in some senses conditioned in response to analytical egalitarian liberalism. Rather than concen trating on justice in distribution the emphasis was on injustice in the processes and social relations of pro duction which were seen as involving geographical, as well as historical elements. Emphasis was on the laws of motion of capital and how these conditioned human consciousness, as well as the dialectical relation between production, accumulation, and space. From a structuralist point of view the emphasis was on the geography of the capitalist economy and the way this ordered the super structure of society and culture. Thus when rational choice Marxism was establishing itself in the late 1970s dialectical and structural approaches were already in fluencing a new generation of geographers. In this cli mate rational choice Marxism was either seen as giving too much over to the approach of neoclassical economics, or to analytical philosophy of the liberal mode. In that sense rational choice Marxism never really got a foothold in the discipline. Marxist geographers were much more concerned with the role of space in the process of historical–geographical materialism. There were some significant attempts to apply rational choice Marxism to geographical questions. One such was Peter Sunley's paper in 1991 on rational choice Marxism and spatial divisions of labor in which he pointed to the inroads that could be made using this approach to understand labor market dynamics and labor relations, although he did conclude that the historical and geographical constitution of agents was important and was largely overlooked by this approach. Analytical Marxism as an intellectual project more widely had largely ended by the late 1990s. Its main proponents went on to more general consider ations of egalitarianism or explorations of democratic theory. It seems that the contrasts between the rational choice approach and this form of critical theory were just too great.

Conclusion: Rational Choice Approaches and Geography

Although the assumptions behind RCT and rational choice Marxism seemed too narrow for the sorts of complex and contextual questions of space and time that geographers address there are rapprochements now oc curring between rational choice and other approaches that have interesting implications for future geographical scholarship. First of all game theory and the idea of interdependent decision making is constantly pointing to the importance of social norms and the scales at which social cooperation can be achieved. Links are also being made between rational choice and the 'critical theory' of the Frankfurt School and between rational choice and philosophical pragmatism (a philosophy that is receiving much more attention in geography). There are inter esting connections being explored between rational choice and forms of public deliberation and radical democracy and in relation to the strategic action of global environmental groups. An attempt has been made to indicate how the idea of strategic rationality in RCT can be linked to ideas of communicative and indeed per formative rationality (from philosophical pragmatism) and the way that space, in this case the different spaces of the city, help materially constitute practical rationalities and the sorts of choices actors make and the style of reasoning pursued. This increasing focus on the contexts of rationality suggests at least the possibility of a revived and productive role for RCT in geography in the future.

See also: Behavioral Geography; Economic Geography; Industrial Location; Marxism/Marxist Geography I; Marxism/Marxist Geography II.

Further Reading

Binmore, K. (1992). *Fun and Games*. Lexington, MA: D C Heath.

Binmore, K. (1994). *Game Theory and the Social Contract, Volume 1: Playing Fair*. Boston: MIT Press.

Breen, R. and Goldthorpe, J. H. (1997). Explaining educational differentials: Towards a formal rational choice theory. *Rationality and Society* 9, 275 305.

Bridge, G. (2000). Rationality, ethics and space: On situated universalism and the self interested acknowledgement of difference. *Environment and Planning D: Society and Space* 18, 519 535.

Bridge, G. (2003). Everyday rationality and the emancipatory city. In Lees, L. (ed.) *The Emancipatory City: Paradoxes and Possibilities*. London: Sage.

Bridge, G. (2005). *Reason in the City of Difference: Pragmatism, Communicative Action and Contemporary Urbanism*. London: Routledge.

Carver, T. and Thomas, P. (eds.) (1995). *Rational Choice Marxism*. London: MacMillan.

Cohen, A. (2000a). *Karl Marx's Theory of History: A Defence (Expanded Edition)*. Oxford: Oxford University Press.

Dixit, A. and Nalebuff, B. (1991). *Thinking Strategically*. New York: Norton.

Elster, J. (1985). *Making Sense of Marx*. Cambridge: Cambridge University Press.

Elster, J. (1989). *The Cement of Society: A Study of Social Order*. Cambridge: Cambridge University Press.

Goldthorpe, J. H. (1998). Rational action theory for sociology. *British Journal of Sociology* 49, 167 192.

Gordon, D. (1991). *Resurrecting Marx: The Analytical Marxists on Freedom, Exploitation, and Justice*. Princeton, NJ: Transaction Publishers.

Gould, P. (1963). Man against his environment: A game theoretic framework. *Annals of the Association of American Geographers* 53, 290 297.
Haggett, P. (1965). *Locational Analysis in Human Geography*. London: Edward Arnold.
Hargreaves Heap, S. and Veroufakis, Y. (1995). *Game Theory: A Critical Introduction*. London: Routledge.
Hirst, P. (1985). G. A. Cohen's theory of history. In *Marxism and Historical Writing*. London: Routledge.
Johnson, J. (1991). Habermas on strategic and communicative action. *Political Theory* 19, 181 201.
Johnson, J. (1993). Is talk really cheap? Promoting conversation between critical theory and rational choice. *American Political Science Review* 87, 74 86.
Knight, J. and Johnson, J. (1999). Inquiry into democracy: What might a pragmatist make of rational choice theory. *American Journal of Political Science* 43, 566 589.
Levine, A. and Wright, E. O. (1980). Rationality and class struggle. *New Left Review* 123, 47 68.
Lewis, D. (1969). *Convention: A Philosophical Study*. Oxford: Basil Blackwell.
Luce, R. D. and Raiffa, H. (1957). *Games and Decisions: Introduction and Critical Survey*. New York: Dover.
Mayer, T. F. (1994). *Analytical Marxism*. Thousand Oaks, CA: Sage.
Parijs, P. (1993). *Marxism Recycled*. Cambridge: Cambridge University Press.
Pred, A. (1967). *Behaviour and Location Lund Series in Geography,* Series B, no 27.
Przeworski, A. (1985). *Capitalism and Social Democracy*. Cambridge: Cambridge University Press.
Roberts, M. (1996). *Analytical Marxism: A Critique*. London: Verso.
Roemer, J. (1982). *A General Theory of Exploitation and Class*. Cambridge, MA: Cambridge University Press.
Roemer, J. (ed.) (1986). *Analytical Marxism*. Cambridge: Cambridge University Press.
Schelling, T. (1960). *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.
Sheppard, E. and Barnes, T. J. (1990). *The Capitalist Space Economy: Geographical Analysis after Ricardo, Marx and Straffa*. London: Unwin Hyman.
Sugden, R. (1989). Spontaneous order. *Journal of Economic Perspectives* 3, 85 97.
Sunley, P. (1991). Game theories: Spatial divisions of labour and rational choice. *Marxism Antipode* 23(2), 229 239.
Ullmann Margalit, E. (1977). *The Emergence of Norms*. Oxford: Oxford University Press.
Wolpert, J. (1964). The decision process in spatial context. *Annals of the Association of American Geographers* 54, 537 558.
Wood, A. (2004). *Karl Marx*. New York: Routledge.
Wood, E. (1989). Rational choice Marxism: Is the game worth the candle? *New Left Review* 177, 41 88.
Wright, E. O. (1989). *The Debate on Classes*. London: Verso.

Relevant Websites

http://kuznets.fas.harvard.edu
Alvin Roth's game theory and experimental economics.
http://levine.sscnet.ucla.edu
David Levine's game theory: Papers, lecture notes.
http://else.econ.ucl.ac.uk
ELSE: the ESRC Centre for Economic Learning and Social Evolution.
http://plato.stanford.edu
Game Theory and Ethics (Stanford Encyclopedia of Philosophy.)
http://GameTheory.net
Game theory resources for educators and students.
http://Marxists.org
The most complete library of Marxism with content in over 40 languages.

Reclus, E.

J. P. Clark, Loyola University, New Orleans, LA, USA

Glossary

Anarchism A social and political theory based on the rejection of all forms of domination, which it identifies with such institutions as the state, capitalism, and patriarchy. It proposes a society based on voluntary, cooperative, and nonhierarchical forms of social organization.

Dialectic A theoretical perspective that emphasizes such principles as the following: reality and thought are in a constant process of self-transformation; development results from contradiction or opposition; phenomena can only be understood as parts of larger developing wholes; there is an interpenetration or mutual determination of opposing phenomena; the structure of a phenomenon embodies the history of that phenomenon; a phenomenon is determined or defined by that to which it is related.

Esperanto The most widely spoken constructed international language. Designed to serve as a universal second language, it was developed by Polish linguist L. L. Zamenhof, who published its first grammar in 1887.

Mutual Aid Cooperative, symbiotic behavior that is beneficial both to individual persons or organisms and to the human or natural communities of which they are parts. A key concept in anarchist and libertarian socialist theory, it was popularized by geographer Peter Kropotkin in *Mutual Aid*: *A Factor in Evolution*, in which he argued that such activity was central to natural selection and human social progress.

New Urbanism A movement in contemporary urban planning that focuses on the design of mixed use, mixed income, and human-scale communities. It opposes urban sprawl and invokes images of the traditional neighborhood and the concept of the urban village.

Paris Commune The working-class municipal government that ruled Paris briefly between March and May of 1871. It was crushed brutally by the central government and assumed legendary status for the international left and labor movements.

The First International (International Workingmen's Association) An international organization founded in London in 1864 to unite working-class political and labor movements from various countries based on the idea of class struggle and the abolition of capitalism. It dissolved in 1876, primarily as the result of conflicts between Marxist and anarchist tendencies.

Elisée Reclus was born on 15 March 1830, in Sainte Foy la Grande in Southwest France and grew up in the town of Orthez in the Pyrenees. He was educated at Protestant schools in Germany and France and briefly attended the University of Berlin, where he came under the influence of the noted geographer Carl Ritter. After opposing Napoleon III's *coup d'etat* of 1851 he was forced into exile for 5 years, most of which he spent in Louisiana and New Granada (Colombia). His travels in this period led to his first geographical writings, in addition to early political works that included a critique of slavery and other orms of social injustice.

During the 1860s, Reclus published many geo graphical essays and completed the first of his three major geographical projects, the two volume *La Terre*: *Description des Phénomènes de la vie du Globe* (translated as *The Earth: A Descriptive History of the Phenomena of the Life of the Globe* and *The Ocean*, *Atmosphere, and Life*). He also wrote *Histoire d'un Ruisseau* (*The Story of a Brook*), a popular work that became a classic of nature writing for young people.

In 1871, Reclus played an active role in the Paris Commune and helped defend the city as a member of the National Guard. After the city fell, he was captured, imprisoned, and sentenced to deportation to New Caledonia. Primarily because of his international reputation as a geographer his sentence was reduced to 10 years exile in Switzerland. There he met revolutionary leader Michael Bakunin and fellow geographer revolutionary Peter Kropotkin. He became active in revolutionary politics, soon emerging as a major proponent of anarchism. During this period he also completed most of his *Nouvelle Géographie Universelle* (translated as *The Earth and Its Inhabitants: The Universal Geography*). This monumental 18 000 page work appeared in 19 large volumes between 1876 and 1894 and has often been characterized as the most comprehensive geographical project ever undertaken by a single researcher.

In 1890, Reclus was allowed to return to France and in 1894 accepted an invitation to teach at the New University in Brussels, where he completed his last great project, *L'Homme et la Terre* (*Man and the Earth*), shortly before his death in 1905. This six volume, 3500 page work is a synthesis of his ideas in geography, social history, philosophy, natural science, political theory, sociology of religion, and anthropology. With it, Reclus' completed his life's work: the development of a global geography that was at once universal, historical, dialectical, and political.

Reclus began his great synthetic work *L'Homme et la Terre* with the epigraph: "Humanity is nature becoming self conscious." This statement is an expression of his project of tracing the course of human history within the context of a larger planetary history in which humanity and nature interact dialectically. In carrying out this project, Reclus established his place as an early prophet of the developing globalization of humanity. In the *New Universal Geography* he observed that the system of industrial technology was creating a kind of planetary 'nervous system' with universal standard time widely adopted and demand growing for a universal language (which he argued should be Esperanto). He also noted an expansion of the collective self consciousness of humanity as general knowledge of geography and history create a new global spatiality and temporality. Reclus believed that this universal consciousness would become a powerful social force to the extent that it would produce corresponding forms of cooperative global political and economic association. He hence saw the establishment of the International Workers Association in 1864 as an event of world historical importance in ushering a new age of such global solidarity.

Reclus' conception of planetary temporality and spatiality was expressed not only in his great geographical works but also in several practical proposals for popular geographical education. One project was for the construction of a great globe to be placed at the center of a large city or at a World's Fair. A winding staircase or some other means of locomotion would be constructed to allow the spectators close access, and a scale of 1:50 000 would have allowed even relatively small topographical details to be perceived. He also proposed the publication of a large world atlas with a uniform scale of 1:10 000 000. It would have consisted of 36 maps depicting the entire surface of the Earth. Additionally, he suggested that a new calendar be adopted that would take the first solar eclipse recorded in human history as its initial point. In beginning with an event of both terrestrial and astronomical significance, his calendar would relate the course of history to the place of humanity and the Earth in the larger cosmos.

Reclus' central problematic was the development of a historically based, dialectical social geography. He identifies three fundamental 'orders of facts' that would be central to the analysis of human society in such a geography: the class struggle (*la lutte des classes*); the quest for balance and stability (*la recherche de l'équilibre*); and the autonomous activity of individual agents (*la decision souverain de l'individu*). He argued that the interaction between the first two of these orders constitutes a dialectic between social conflict and social order, while the interaction between the first two orders and the third constitutes a dialectic between the macrolevel of social institutions and the microlevel of the individual. Reclus' social geography focused on the complex interrelationships between human society and the natural world, which he divided into three general spheres. These are: an external realm of nature that predated humanity, with which the 'human social milieu' always remains in intimate interrelationship; a 'static milieu' that included 'the natural conditions of life'; and a 'dynamic milieu' made up of 'the artificial sphere of existence'. The static milieu, though shaped by human culture, retains a certain degree of natural necessity, whereas the dynamic milieu can be reshaped more fundamentally by human transformative activity. The static milieu includes not only such environmental variables as soil, climate, and a large range of geographical conditions but also such 'primordial social facts' as modes of labor, dietary practices, kinship relations, and forms of community. On the other hand, he classifies such specific institutions as wage labor, private property and the state as 'secondary facts' since many societies have existed without them. Reclus saw these distinctions as crucial for developing an adequate understanding of the interconnections between nature, society, and social change. He contended that, on the one hand, the influence of external nature and of the 'static milieu' had been neglected in previous analyses and that consequently the power of these determinants has been underestimated. He attempted to redress this imbalance when, for example, he discussed 'survivals' of past practices and institutions that exert a force within the existing social system. He held that, on the other hand, the social relativity of the dynamic milieu had not been adequately comprehended. As a result, he argued that contingent social institutions had often been treated as if they were unalterable facts of nature, whereas they are in reality subject to far ranging intentional transformation.

Throughout his analyses, Reclus strongly emphasized tendencies in both nature and human society toward order and unification; however, he also stressed contradictory tendencies toward discord and imbalance. While he focused, much like Kropotkin, on the role of mutual aid in natural and social evolution, he was more willing than the latter to recognize the role of human struggle with the natural world. Reclus anticipated in some ways the later shift in ecological thought from a static 'balance of nature' model to a more dynamic 'balance of order and chaos' position.

Yet above all else, Reclus placed what he called 'the social question' – the historical problematic of liberatory social transformation – at the center of his geographical project. By integrating this 'social question' into geography he expanded the scope of the discipline far beyond the limits generally observed in his era, incorporating into it issues of economic class, race, gender, power, and social domination, organizational forms and scale, urbanization, technology, and ecology. Reclus' political geography included two major areas of inquiry. As a geography of social domination, it investigated the evolution of systems of domination since the beginnings of hierarchical society, and the spatial expansion of these systems as dominant states, cultures, and economies extended their power and hegemony across the globe. As a geography of social freedom, it traced the evolution of liberatory institutions as manifest in tribal societies, the Greek polis, Icelandic democracy, European free cities and communes, the Enlightenment, and modern revolutionary, labor, and liberation movements. It also delineates the spatial expansion of these currents into a global movement for the attainment of human equality and freedom.

Reclus' political geography also included a strong ecological dimension. He cited many cases throughout history in which humanity has established a cooperative relationship with the natural world, and used these models to help envision a future society that could achieve justice and freedom in harmony with the natural world. He also devotes extensive attention to the accelerating rate of damage to the biosphere through human activity, and relates this ecological destruction to the historical development of the state, capitalism, and industrial technology. As early as the 1860s, he presented trenchant analyses of ecological problems such as the destruction of ancient forests, defacement of landscapes, the spread of invasive species, and the abuses of industrialized agriculture. In addition, he advocated programs of ecological restoration as a response to the developing ecological crisis.

Reclus was also a pioneer in urban geography during an epoch in which cities were becoming an increasingly powerful determinant in all social and ecological spheres, but remained a neglected area of geographical analysis. His distinctive contribution followed in large part from his methodological premise that all aspects of 'the urban question' are inseparable from 'the social question'. He analyzed the city in accord with his general problematic, which focused on the constrains imposed on human development and the natural environment by various forms of domination and on the possibilities in a given era for human flourishing in harmony with nature. He argued that unless fundamental social inequities and the institutional context of social domination were addressed, seemingly progressive policies such as slum clearance and civic beautification would merely displace social evils rather than eliminate them. His analysis hence anticipated aspects of contemporary radical urban theory, such as the ongoing critique of the new urbanism associated with neoliberal urban process.

Reclus depicted the urban milieu as spatially and historically inseparable from the prevailing economic, political, and social systems and from the ecological and geographical regions in which it is embedded. He contends that the town or city must be comprehended as the product of the social and natural conditions that constitute its history, and which live on as significant determinants that are neglected in ahistorical, non developmental and non ecological analyses of urban conditions. His tradition of radical urbanism, with its holistic approach to the ecological region, its economic and political decentralism, and its focus on popular participation, lives on through his own modest influence on the practice of geography, as well as his more direct influence on some of the pioneers of town and country planning, notably, Patrick Geddes and Lewis Mumford.

See also: Anarchism/Anarchist Geography; Kropotkin, P.; New Urbanism; Political Geography; Radical Geography.

Further Reading

Becker, H., Bedos, P., Roy, D., Bolufer, J. and Colombo, M. (eds.) (1998). *Elisée Reclus,* special edition of *Itinéraire* 14 15.

Clark, J. and Martin, C. (eds. and trans.) (2004). *Anarchy*, *Geography*, *Modernity*: *The Radical Social Thought of Elisée Reclus*. Lanham, MD: Lexington Books.

Cornuault, J. (1995). *Elisée Reclus: Géographe et Poète*. Eglise Neuve d'Issac: Fédérop.
Dunbar, G. S. (1978). *Elisée Reclus: Historian of Nature*. Hamden, CT: Archon Books.
Fleming, M. (1988). *The Geography of Freedom: The Odyssey of Elisée Reclus*. Montréal, QC: Black Rose Books.
Lacoste, Y. (ed.) (1981). *Elisée Reclus: Un géographe libertaire,* special edition of *Hérodote* 22.
Reclus, E. (1869). *Histoire d'un Ruisseau*. Paris: Hachette.
Reclus, E. (1871). *The Earth: A Descriptive History of the Phenomena of the Life of the Globe*. New York: Harper.
Reclus, E. (1873). *The Ocean, Atmosphere, and Life*. New York: Harper.
Reclus, E. (1905 08). *L'Homme et la Terre*, (6 vols.). Paris: Librairie Universelle.
Reclus, E. (1882 95). *The Earth and Its Inhabitants: The Universal Geography*, (19 vols.). London: H. Virtue.

Relevant Websites

http://raforum.apinc.org
Reclus, Élisée (1830 1905). Geographer.

Redbrick University Geography in Britain

T. R. Slater, University of Birmingham, Birmingham, UK

Glossary

Campus University A tertiary educational institution where all of its buildings are established within a single park-like estate, usually in the suburbs of the city. Reading University (1926) was the first single-site campus in Britain but Birmingham had both a city-center site and a suburban campus (at Edgbaston) from its founding in 1900.

Geographical Association A British educational organization dedicated to the promotion of the teaching of geography in British schools at primary and secondary levels. Its headquarters is in Sheffield. It was founded in 1893 as an independent unincorporated association financed by its membership subscriptions.

Redbrick Universities The six British universities founded at the beginning of the twentieth century in the provincial towns of Birmingham, Bristol, Leeds, Liverpool, Manchester, and Sheffield. They were noncollegiate; they admitted students without reference to religious affiliation and concentrated on science, engineering, and training teachers. Their buildings were characteristically of red brick. The term was first used in a 1943 book on the University of Liverpool entitled *Redbrick university*.

Royal Geographical Society The oldest learned society and professional body for geography and geographers in the United Kingdom. It was established in 1830 to promote 'the advancement of geographical science'.

Introduction

The history of British academic geography before World War II is well known in outline. The anniversary celebrations of professional bodies such as the Royal Geographical Society (RGS), the Geographical Association (GA), and the Institute of British Geographers (IBG), and of early established departments such as Liverpool and Aberystwyth have provided a firm framework for this history. This can be supplemented from the memoirs of individuals active in these early years, biographies and obituaries of significant individuals, and some general surveys of the history of British geography.

The academic community of geographers consisted only of some 120 individuals in the period before 1945. They were scattered through all the universities and university colleges then existing, and few departments of geography consisted of more than half a dozen academic staff. The personal and institutional links within this academic community have hardly begun to be explored. The careers of some, though by no means all, of the founding heads of university departments have been charted but the achievements of many of those who taught in universities before 1945 are today forgotten.

The universities functioning in 1945 fell into a number of broad groups. The Scottish universities have much in common with each other, not least their different system of examination throughout the education system when compared with England. The university colleges of Wales provide another broad group, as do the university colleges of London. The university colleges still examined under the external regulations of London in 1945 (Exeter, Hull, Leicester, Nottingham, and Southampton) form a fourth coherent group which became independent universities in the 1950s. Finally, the so called 'civic' universities, which were founded as tertiary colleges in the second half of the nineteenth century, and which gained Royal Charters as independent degree awarding, noncollegiate universities and which did not demand allegiance to the Church of England form a fifth grouping. Five of them (Birmingham, Leeds, Liverpool, Manchester, and Sheffield) were further linked by the Joint Matriculation Board (JMB) of examiners. These five are the classic 'redbrick' universities of England, all founded in the first 5 years of the twentieth century, and it is upon them that this article will focus. The term 'redbrick university' was first used in 1943, in the title of a book about Liverpool University, and was quite quickly extended to the other four in the JMB group. There are two other civic universities founded at the beginning of the twentieth century, that at Bristol (1909) and Queen's University, Belfast (1908), neither of which was part of JMB but both of which are sometimes included in the 'redbrick' university grouping.

Science Colleges and Early Geography Teaching

The early development of geography teaching must be set in the wider structural context of educational development, on the one hand to meet the needs of a new industrial society and the commercial exploitation and management of the Empire, and on the other, to incorporate Darwinian and related scientific thinking. In the tertiary educational sector these factors were manifested in the development, in the mid Victorian period, of provincial science colleges in the major manufacturing and commercial cities of the Midlands and the North. The earliest teaching of geography went hand in hand with the founding of these science colleges, but it formed only a minor part of the curriculum, usually for commerce and arts students. Once the teaching of geography became fully established in elementary schools from the 1880s, the trainee school teachers were also taught geography at the colleges and it was that decade which saw the beginnings of the campaign to establish geographers in the universities.

The oldest established predecessor of the five redbrick universities was that at Manchester, where Owen's College had been founded in 1851 with a substantial endowment from a local businessman. It was another 20 years before science colleges came to be founded at Leeds, in 1874, and at Sheffield in 1875. At Birmingham, Mason Science College was established in 1880, and Liverpool University College was founded in 1883. None of these later colleges were as well endowed or as broadly conceived as Owen's College had been, with its wide spectrum of teaching in the arts and sciences, though Liverpool, funded by the city's prosperous mercantile community, came closest. Birmingham, Leeds, and Sheffield were much more narrowly conceived and were intended to provide a basic scientific education for the sons and daughters of the industrial middle class of these cities. Few students entered for degree level work but where they did it was for the external degrees of the University of London.

At Owen's College the earliest geography teaching was by geologists and historians. In 1892, it became the first of the five colleges to appoint a full time lecturer in 'political and commercial geography'. H. Yule Oldham was appointed and was succeeded 2 years later by A. J. Herbertson. He, too, held the post for 2 years. At Birmingham, geography teaching was firmly established in the geology department, which had been founded by Charles Lapworth, FRS, in 1881. The 'physiography' course included aspects of geomorphology, climatology, and oceanography, to use today's vocabulary, and was for first year students. In 1891, the appointment of W. S. Boulton enabled a second year course in political geography to be offered and this pattern continued up to the metamorphosis of the college into a full university. At Leeds, geography lectures were provided from 1884 for teachers in the expanding elementary school system and from 1886 geography formed part of a 2 year commercial course for prospective entrants to business careers. Again, these courses were given by teachers in the geology department. Courses for commerce and business students were also the first to include the earliest geography teaching at Liverpool, notably by E. C. K. Gonner, the lecturer (later professor) in economic and commercial science, but other courses were given by assistant lecturers in physics and chemistry. At Firth College, in Sheffield, there are no indications of early geography teaching. Given that it was the smallest of the five institutions, this is not surprising though eventually it too began to provide basic training for teachers.

From College to University

Within a period of 5 years at the beginning of the twentieth century all five institutions had become independent universities. The first to be created was Birmingham in 1900, primarily because of the drive of Joseph Chamberlain who determined that his city should have a university. The remaining four resulted from the disintegration of Victoria University, which had been established in the 1880s by combining the colleges at Manchester, Liverpool, and Leeds. For some time its cumbersome administrative mechanisms had proved unsatisfactory and, in May 1901, Liverpool College applied to the Privy Council for the establishment of an independent university in the city. It was the successful founding of Birmingham that probably gave the final push. As a result of the Liverpool application, Manchester, too decided that an independent existence was to be preferred and in February, 1903 the universities of Liverpool and Manchester were established. In Sheffield, Firth College had combined with the city medical school and technical institute to form a new University College in May, 1897. The intention had been to apply to join the Victoria University but the latter's disintegration saw overtures made to Leeds to form a collegiate Yorkshire University. These were rejected, and both Leeds and Sheffield therefore petitioned the Privy Council for independent university status. Leeds obtained its charter in 1904 and Sheffield in May 1905.

By the beginning of the twentieth century, therefore, the expanding demand for advanced education and the growing civic pride of Britain's provincial cities had converted the five science colleges into full universities. The demand for teaching and commercial and business qualifications gave geography an established place in the universities, as did their efforts to broaden their curriculum beyond the confines of the pure and applied sciences, to the arts and social sciences, so as to enhance and confirm their new status. However, for all of the new universities except Manchester there were three pressing problems. First, the teaching provision for arts subjects generally was still less than adequate in the former science colleges; second, buildings, laboratory provision, and libraries needed considerable enhancement; and third, the finance available was not sufficient to provide both additional buildings and staff and to build up the long term investments necessary for an independent institution. It is against this background that the development of geography teaching must be viewed. In particular, the employment of specialist geography staff was normally dependent upon grants being obtained to pay salaries.

Geography at the New Universities

Geography at Birmingham continued almost unchanged after the university had been founded. It was taught by geologists and was an optional half subject course for the commerce degree in the first and second years. University records show it not to have been a popular course with only two or three students enrolled each year. The courses were given by W. W. Watts and F. Raw. In 1904, in an effort to keep him at Birmingham, Watts was given a personal chair. Since there was already a professor of geology, and given Watts' responsibility for geography teaching, he was titled Professor of Geography. Herbertson considered this a substantial boost for the subject. However, it was a false dawn as Watts resigned in 1906 to take up the chair of geology at Imperial College.

The foundation of the new university at Manchester coincided with the establishment of a properly funded lectureship in 'economic and political geography' in 1903. J. McFarlane held this jointly with a lectureship in economics for 3 years, but in 1906 his responsibilities for economics teaching were reduced and he became 'lecturer in geography'. The reason for his appointment was to extend the university's commerce teaching and, by 1904, geography had become a compulsory subject for the commerce degree and higher commercial certificates, as well as an option for arts degree students.

At Liverpool it was commerce students who again formed the backbone of those attending geography courses and it was as an assistant lecturer in the economics department that P. M. Roxby was appointed to develop courses in geography in 1904. By 1906 his success was such that he became 'lecturer in regional geography' and 3 years later a separate department of geography was founded. As at Manchester, it became possible to offer geography for the MA degree, and a substantial array of courses was on offer for undergraduates in the arts and commerce faculties. By 1914 these were attracting a 100 or more students. Liverpool was easily the most successful of the incipient geography departments and it is no surprise that it was the first to offer an honours course in geography from 1917. It was in that year, too, that a prominent local businessman, John Rankin, endowed a chair of geography to which Roxby was appointed.

Geography teaching at Leeds was almost nonexistent before 1914. The small size of the university, its previous history as a science college, and the necessity of an extensive building program in the early years meant that there was little money for new staff. Geography continued to be taught in the geology and economics departments for intermediate students and it was in the latter department that Ll. Rodwell Jones was appointed lecturer in economic geography in 1913. The appointment encouraged the university to develop the place of geography in honours courses in both science and arts and, though Rodwell Jones departed for a post at the London School of Economics after the war, he was replaced by C. B. Fawcett for whom a readership and a separate department were created in 1919.

The first new department to be established after the University of Sheffield received its charter was for geography. The RGS was instrumental in its founding in the person of its then president, Sir George Goldie, who had been chief guest at a dinner held after the meeting of the university court in 1907. He spoke eloquently on the importance of the scientific study of geography for the trade and welfare of the Empire and the desirability of establishing a geography department at Sheffield. Within a few days, Edgar Allen, a prominent businessman in the city who had already given £10 000 to the university for a new library, offered to give £150 for 5 years to endow a lectureship in geography if the university would do likewise. His offer was accepted, and R. N. Rudmose Brown was appointed to the lectureship in 1908. Geography formed an optional subject in the BSc degree and other courses were given to honours students in history and economics. There were usually between 10 and 15 students in any 1 year.

Before World War I, therefore, some geography was being taught in all five universities. At Leeds and Birmingham this was minimal and there was no geography department, but at Liverpool, Manchester, and Sheffield lectureships had been established and degree level courses were being taught. The RGS had some influence in the establishment of geography at Manchester and Sheffield, but of greater significance was the need of all these new universities to widen their teaching in the arts and commerce fields. With the exception of Sheffield, it was commerce students and the training of geography teachers that provided the bulk of the students for these new geography departments and most of their students, it can be noted, came from the immediate city and region which they served.

The cessation of hostilities in 1919 marked a new start in some of the universities and a continuation of former patterns in others. Geography's professional organizations were united in their efforts to bring about changes in the school examination system so that geography could be recognized as a full subject. They were united, too, in their efforts to build upon pre war success in the universities and establish honours teaching in geography. The early efforts of the RGS at Oxford and Cambridge were to some extent surpassed by those of the GA and Section E of the British Association in the provincial universities after the war. The combined enthusiasm and talents of H. J. Fleure and P. M. Roxby at the GA proved especially important. Readerships were established at Manchester and Leeds in 1919. Assistant lecturers were appointed to the departments at Liverpool, again in 1919; at Leeds in 1922; at Manchester in 1925; and at Sheffield in 1928. All but the last of these lecturers were graduates of Roxby's honours school at Liverpool. Honours teaching began at Manchester in 1923 under W. H. Barker; at Sheffield in 1923 (unusually, in the science faculty to begin with), and at Leeds in 1928. The second chair to be created in the redbrick universities after that at Liverpool, was at Manchester, for H. J. Fleure in 1930, following Barker's premature death. However, the complexities of development in five departments through the 1920s are too great to trace in this article and so an extended treatment of only one is given to illustrate influences at work.

Birmingham Beginnings

Birmingham was the only one of the five universities under discussion where geography teaching declined after World War I. It continued to be taught by the staff of the geology department but only for first year students as an option in the BA course, and in Commerce. At Birmingham, as at all the former science colleges, the arts faculty was the poor relation in terms of internal status and finance, though it was recruiting as many students as science by 1920. This was exacerbated by its location in the outdated Mason College building in the city center, whereas most of the science departments were by then located on the new campus at Edgbaston (Britain's first purpose built university campus). It is no surprise to find, therefore, that the development of a new subject was not a high priority for the majority of arts faculty members when their own positions were so parlous.

To the university, however, the situation was an embarrassment and it was an embarrassment that was made good use of by the professional organizations. The organizer of the campaign to establish geography in all the universities was H. J. Fleure, the energetic secretary and editor of the GA and Professor of Geography at Aberystwyth, together with Roxby at Liverpool. They organized a continuing series of letters and pamphlets to those universities, such as Birmingham, which had not appointed professional geographers; continued the campaign by advocating that lecturers be advanced to readers and professors in charge of their own departments; and then proceeded to encourage the establishment of honours courses.

The way in which such campaigns were managed can be demonstrated for Birmingham through the archives of the RGS. The link between the GA and the RGS in this instance was J. L. Myers who had already been instrumental in the founding of geography at Liverpool. It was he who contacted A. R. Hinks, Secretary of the RGS, to suggest that their support for a memorandum which the GA was preparing to send to the council of the university would be appreciated or, better still, that they might send their own representations to Birmingham. Myres' letter also reveals that Sir Charles Grant Robertson, principal of the university, was actively encouraging such representations and was certainly not unsympathetic to the cause being pleaded. He, too, was an eminent Oxford historian who had lectured in historical geography at All Souls. In his annual report to the court in each of the 5 years after 1919 he made special reference to the urgent need to establish a geography department but, as in the other redbrick universities, the problem was one of finance since the building project at Edgbaston was a severe drain on resources.

Despite the location of geography teaching in the geology department at Birmingham, it was to historians that Hinks turned in an effort to discern the feelings of staff at the university. In particular to Professor Sir C. Raymond Beazley, head of the department of history, whose monumental and justly famed *The dawn of modern geography* had been published at the turn of the century while he was a Fellow of Merton College, Oxford. Beazley replied that he personally would "welcome an appeal such as you suggest … (and) would support it, if moderately and kindly expressed with mercy to our deficiencies,… in the Senate." He suggested, too, that it would be better if the RGS and the GA made separate representations.

Besides such representations to the university directly, the pressure was maintained in the geographical community of local school teachers by Fleure's forthright editorial comments in successive issues of the *Geographical Teacher*, where new appointments were carefully recorded and laggard universities such as Birmingham and Durham were reprimanded. The local branches of the GA were an important pressure group in their own right since they consisted largely of the secondary school teachers responsible for directing their pupils toward an appropriate university. At a time when the great majority of students in the provinces went to their local university the necessity to move elsewhere to study a subject of growing popularity was not a good advertisement for Birmingham in the West Midlands region. There can be no doubt that this was an important factor in the eyes of the university since the Birmingham GA local branch was large and flourishing. It had been founded in 1913 with the support of the staff of geology and local secondary school teachers and, by 1919, Fleure could commend it as "our largest and one of our most enthusiastic branches."

Robertson's annual appeals to the Court, together with all this external pressure, paid dividends in the early part of 1924 when an anonymous private charitable trust offered to provide £600 per year for 5 years as the stipend of a reader in geography. Two candidates were considered, both of them overseas at the time. The unsuccessful candidate was Dudley Stamp, then heading the geography department of Rangoon, Burma and well known to Fleure and Roxby because he had established a hugely successful overseas branch of the GA there. The successful candidate was Roxby's colleague at Liverpool, R. H. Kinvig. Kinvig's studies of the West Country's woollen industry had gained him an MA in 1919 but in 1924 he was in South America undertaking research for a never to be completed book and he had to return posthaste for the interview. A year later, geography had been approved as a principal subject for the BA degree and some 40 students enrolled for his classes. Honours teaching began in 1926 with three students registered for the course, but there were more than 80 registered for other courses and Robertson was able to report with satisfaction to the court "the demand already shown fully justifies the readership." Kinvig was consequently provided with his first assistant lecturer, Gordon Manley, a Manchester graduate who had spent two years at Cambridge under Debenham. He had then worked for a year at the Meteorological Office before going as a member of Sir James Wordie's geomorphological expedition to Greenland in 1925. He was therefore well equipped to relieve Kinvig of the teaching of physical geography.

Clearly, Kinvig's achievement was a substantial one. In only 2 years he had taken geography at Birmingham from a position of virtual nonexistence to an established department with two staff, a full honours course, and buoyant student demand. The same task was to take Fawcett 9 years at Leeds and had taken Rudmose Brown 16 years at Sheffield. These differences were less ones of ability, of course, than of external circumstances. The significance of the middle years of the 1920s was that the other great campaign of geography's professional organizations, to persuade the Board of Education to accept geography as an 'advanced subject' for study in secondary schools, reached fruition in 1925. As a result there was an accelerating demand for university trained geography teachers.

The earliest syllabus for the honours school at Birmingham was devised by Kinvig, no doubt drawing on his experience at Liverpool and having due regard to Manley's specialities. It included a substantial historical component. The first year covered 'the physical basis of geography': climatology, oceanography, and geomorphology (the latter taught in the geology department), together with cartography, European history, and a foreign language course. Human and regional geography composed the mainstay of the second year, together with practical work based on the local region, economic history and theory, and a second foreign language course. In the final year, courses were planned on the historical and regional geography of the British Isles; the lands of the North Atlantic; the monsoon lands and a series of options taken in other departments including plant geography (botany), the economics of transport (commerce), and the history of geographical discovery (history). By the time this syllabus was actually taught to the final year it had been somewhat modified to include four regional geography papers out of seven and 'a thesis on some region, or on a particular aspect of geography'. This emphasis on regional geography, which was taught with a strong human and cultural emphasis, was a reflection of the needs of geography teachers in the 1920s and 1930s.

The Post-1945 Story

After World War II geography departments in the redbrick universities expanded substantially in both staff and student numbers. Wartime experiences opened new avenues of research in regional specialisms and systematic branches of the subject. By the 1950s the first doctoral students were being trained and they then moved on to university and college teaching posts. In the 1960s the Robbins Report (1963) led to the expansion of the universities and geography departments shared in that expansion. It also saw new intellectual endeavors in the subject with the 'quantitative revolution' affecting both human and physical geography and the decline of regional geography. In the civic universities, the embeddedness of geography as a specific academic discipline and its popularity as an advanced subject for study in the secondary school system ensured that geography departments continued to expand in both student and staff numbers. In some civic universities geographers were able to position themselves to take advantage of their growing size in the governance of their institutions, but in others, their origins as colleges specializing in science and engineering left political and financial clout in these faculties, to the detriment of geography. The redbrick university geography departments remain among the largest departments in the country, both in terms of the number of students graduating each year and the number of staff employed in them. Some might argue that in terms of intellectual endeavor, however, that this very size militates against intellectual innovation and that leadership in the subject's development lies elsewhere.

Conclusions

The institutional history of British geography has thus far concentrated primarily upon the innovations in methods and the personalities involved in developing the subject in Oxford, Cambridge and, to a lesser extent, London. Rightly so, in that few could deny that it was here that the twentieth century expansion of the subject in all sectors of education was forged. However, once the initial battles had been fought, it was precisely the expansion of geo graphical teaching into the colleges and universities of the provinces which provided the subject with the broad base of enthusiastic support necessary to enable further developments to be carried through. It was these forces, under the organization banner of the GA and its prin cipal officers, which ultimately saw the establishment of geographers and geography departments in all the pro vincial universities.

The role played by the institutional development of the provincial universities themselves has been less often alluded to and the conclusions which are obtained here may seem obvious. The successive interaction between the growth of the Empire and colonial trade, the neces sity for better commercial education, the legislative intervention of government to redirect and enhance educational provision, the growth of tertiary education, and the development of geography as a relevant and necessary part of that growth may, with hindsight, seem natural and preordained. However, remembrance of the development of the colleges of advanced technology into universities in the 1960s, and the founding of new uni versities then, did not see a similar expansion of geog raphy departments though the former, in particular, were seeking to expand their arts and social science teaching in much the same way, and for much the same reasons, as the late nineteenth century science colleges when they became universities. Clearly expansion was not pre ordained; institutional developments provided a suitable context, but ultimately the expansion of academic geography depended upon innovative and enthusiastic geographers who believed in what they were doing. It is geography's good fortune that the pioneer geographers in the early established departments trained men and women of equal enthusiasm to take the subject firmly into the civic universities by the 1920s and on into the post 1945 expansion of higher education thereafter.

See also: Enlightenment Geography; Historical Geography; Knowledge and Education, Historical Geographies of; Oxbridge Geographies.

Further Reading

Baker, J. N. L. (1963). *The history of geography*. Oxford: Oxford University Press.

Fleure, H. J. (1953). Sixty years of geography and education: A retrospect of the Geographical Association. *Geography* 38, 231 264.

Freeman, T. W. (1980). *A history of modern British geography*. London: Longman.

Livingstone, D. (1992). *The geographical tradition: Episodes in the history of a contested episode*. Oxford: Blackwell.

Ploszajska, T. (1999). *Historical Geography Research Group monographs 33: Geographical education, empire and citizenship: Geographical teaching and learning in English schools, 1870 1944*. Liverpool: Historical Geography Reasearch Group.

Proctor, N. (1986). The pioneers of geography: A new and culturally relevant perspective. *Transactions of the Institute of British Geographers* NS 11, 75 85.

Slater, T. R. (1988). Redbrick academic geography. *Geographical Journal* 154, 169 180.

Steel, R. W. (1967). Geography at the University of Liverpool. In Steel, R. W. & Lawton, R. (eds.) *Liverpool essays in geography: A jubilee collection*, pp 1 24. London: Longman.

Stoddart, D. (1985). *On geography and its history*. Oxford: Blackwell.

Withers, C. W. J. and Mayhew, R. J. (2002). Rethinking 'disciplinary' history: Geography in British universities, *c*. 1580 1887. *Transactions of the Institute of British Geographers* NS 27, 11 29.

Relevant Websites

http://www.manchester.ac.uk
About Geography, The University of Manchester.

http://www.geos.ed.ac.uk
A Short History of Geography in Edinburgh: Looking forward, looking back, School of Geosciences, Institute of Geography, University of Edinburgh.

Redlining

M. B. Aalbers, Columbia University, New York, NY, USA

Introduction: Redlining as a Self-Fulfilling Prophecy

Redlining is the refusal of a mortgage lender to grant mortgage loans in particular neighborhoods. Many authors include uneven conditions, such as higher transaction costs and higher interest rates, in their definition of redlining. For matters of clarity, we define such uneven home mortgage finance conditions as 'yel lowlining' which is different from 'redlining' where a lender does not grant any mortgage loans at all. Redlining is a form of place based financial and social exclusion. Mortgage applicants are excluded 'from' obtaining housing by denying them mortgages in redlined neigh borhoods. Current homeowners are excluded 'through' housing because they are unable to sell their house, be coming trapped in their neighborhood.

Financial institutions reject mortgage applications for neighborhoods where they have no trust in the financial dependability of potential customers or in the (future) continued appreciation of the properties, expecting real estate prices to fall. The risk that a homeowner can only sell his home for a lower price than purchased, and as a result is unable to pay off the entire mortgage, is con sidered too big. Lenders assume that members of certain groups are on average less able to fulfill their financial commitments. If lenders will believe that minority ap plicants are more likely to default than white applicants with the same observed credit characteristics, this creates an economic incentive to discriminate against minority applicants or minority neighborhoods – whether these beliefs are founded on hard data or not. This behavior is illegal – a lender must base his or her decision on the observed credit characteristics of an applicant – but some lenders may respond more to economic incentives than to legal requirements. Redlining in the mortgage market strikes low income neighborhoods and ethnic minority neighborhoods in particular. Yet, redlining not only strikes low income families and ethnic minorities, but also everyone else applying for a mortgage in a redlined neighborhood.

According to the sociologist Robert Merton per ception alone is sufficient for people to act; prejudices may come into play when excluding certain subjected groups from mortgage capital. If the exclusion is based on aggregate data, it is a matter of statistical discrimination: individual members of a group are excluded simply by the fact that their group on average has a bad credit profile. By defining a situation where certain neighbor hoods come with a large depreciation risk or a large default risk, and are therefore denied mortgage capital, lenders evoke a chain of effects which eventually causes houses in these neighborhoods to actually lose value and which will subsequently increase default rates. Because mortgages are constrained, a majority of the potential homeowners cannot buy a house in these neighborhoods. Drop in demand leads to sinking real estate prices. The circle is closed. Real estate prices have dropped because it was impossible to get a home mortgage in redlined neighborhoods – and not so much because these neigh borhoods were high risk investments. They became high risk investments because they were 'perceived' high risk investments. This is a matter of a self fulfilling prophecy: because prices are expected to fall, people act accord ingly, which in fact causes prices to fall. The principle behind a self fulfilling prophecy is that people not only react to objective circumstances in a situation, but also, and sometimes primarily, to the meaning that the situ ation has for them. Just like Merton has analyzed, we see that public definitions of a situation, like expectations and prophecies, become an integral part of that situation and influence future developments.

So far, we have assumed that redlining is something that takes place at the level of the city, where mortgage lenders decide where to grant and where not to grant mortgages. And in fact, the actual decision to redline often, but not always, takes place at this level. Yet, the decision taken at the city level by mortgage lenders de pends on many factors, some of which are located at the city or neighborhood level, while others are located at other levels. The idea of redlining cannot exist without a notion of internal differentiation of geographical space: no internal differentiation – or to put it differently, no submarkets – no redlining. The question of scale is central to any relevant explanation. We could say that redlining is a product of the uneven development of capitalism, of the application of credit scoring models, of prejudiced urban managers and gatekeepers, of the structure and regulation of the mortgage market, of statistical discrimination, or of neighborhood decline and the devalorization of capital invested in the built en vironment. It is, of course, in some way, a product of all these forces. This article is not the place to connect these different forces and scales at length. Alternatively, we look at the origins of redlining in the US, the develop ment and evidence of redlining in the US, redlining outside the US, and redlining outside the mortgage market. In the final part of this article we return to the relevance of scale to the understanding of redlining.

The Origins of Redlining

The origins of redlining are to be found in the devel opment of modern mortgage markets in the US. Initially the US mortgage market was modeled after European mortgage markets. But in 1913 the US Federal Reserve Act allowed commercial banks to lend money on real estate, not permitted in most European countries until the second half of the twentieth century. The entrance of commercial banks into the mortgage market coupled with the economic optimism of the 1920s led to a rapid growth of US mortgage markets. The boom of the 1920s was followed by the Wall Street Crash of 1929 and the crisis of the 1930s. During this crisis homeowners lost their houses in two different ways. First, widespread un employment made it impossible for many homeowners to pay off their mortgage loans, resulting in foreclosures and tax sales. Second, people who retained their job were often unable to refinance their loan after 5 years because of the withdrawal of financial institutions from the mortgage market, and therefore also lost their homes. In the early 1930s the average number of foreclosed mort gage loans was 250 000 per year; half of all home mort gages in the US were in default. Also financial institutions failed and went bankrupt along with their borrowers, partly as a result of the massive withdrawal of savings.

As a reaction to the depression President Hoover organized a conference on home building and home ownership in 1931. The conference proceedings make four recommendations:

1. the creation of long term, amortized mortgages;
2. the encouragement of low interest rates;
3. the institution of government aid to private efforts to house low income families; and
4. the reduction of home construction costs.

Hardly ever have any policy conference proceedings been implemented as actual policies so quickly and com prehensively. This is not the place to go into details and we only focus on the creation of the Home Owners Loan Corporation (HOLC). The HOLC was designed to provide emergency relief to homeowners by refinancing or purchasing defaulted mortgages – in other words, to forestall foreclosures. Thanks to the HOLC tens of thousands of borrowers were kept from losing their homes in the mid and late 1930s and, in addition, it refinanced more than one million mortgages, all on relatively low interest rates. The HOLC introduced the long term, self amortizing mortgage with uniform pay ments spread over the life of the debt, thereby replacing the 5 years, nonamortizing mortgage with a balloon payment at the end of the loan period. It set the standards for mortgage lending till this day and systematized ap praisal methods across the US.

The HOLC was not only instrumental in developing and applying uniform, standardized appraisals, but it was also instrumental in implementing and institutionalizing redlining practices. The HOLC developed a neighbor hood rating system comprising of four colors corres ponding to four different numbers and four different letter codes: green, first grade or A referring to "homo geneous neighborhoods (…) hot spots (…) in demand as residential locations in good times and bad (…) Ameri can business and professional men"; blue, second grade or B referring to "stable (…) still good (…) still desirable" areas that had "reached their peak"; yellow, third grade or C referring to "definitely declining (…) hetero geneous" neighborhoods that attract "undesirable elem ent[s]" and are "infiltrated by a lower grade population; and lastly, red, fourth grade or D referring to neighbor hoods "in which the things taking place in C areas have already happened" as a result of "detrimental influences in a pronounced degree" and where houses have "little or no value today, having suffered a tremendous decline in values due to the colored element now controlling the district" (Jackson, 1985: 197–200; Hillier, 2005: 216–217). According to the historian Kenneth Jackson, Jewish neighborhoods, even the stable and affluent ones, would never be considered first grade, while black neighbor hoods were by nature considered unstable and declining or declined and habitually colored red. Crossney and Bartelt, however, claim that the connection between ethnicity and lending is less strong, even though they admit it is clear that black and mixed areas were more often redlined than homogeneous white areas.

Although the HOLC is often blamed for 'introducing' redlining policies and practices, it is important to note that the HOLC was simply following dominant ideas in the real estate and mortgage markets, already practiced at the local level. However, the HOLC did 'implement' and 'institutionalize' redlining practices. As a government institution it did institutionalize already existing practices of redlining into government policies by designing 'residential security maps' for more than 200 American cities. Fourth grade, red colored neighborhoods were no exception, but rarer on some city maps than on others. While the map of St. Louis County, for example, only shows a small number of neighborhoods colored red, the map of Newark is full of redlined neighborhoods and no single neighborhood in Newark is colored green. **Figure 1**, the map for the city of Baltimore, shows a classic pattern: inner city neighborhoods surrounding the central business district are colored red, the next ring of neighborhoods and the older suburbs are largely col ored yellow, while newer suburbs – albeit with a few exceptions – are colored blue or, less common, green. In addition, in a few years time, the number and size of redlined areas could dramatically increase as Hillier's analysis of the HOLC residential security maps for Philadelphia clearly shows: in two years time the share of city land redlined almost doubled to 34%.

Redlining practices were adopted not only by the HOLC and private mortgage lenders, but also by the Federal Housing Administration (FHA). The FHA, established in 1934 under the Roosevelt administration, was created to insure private mortgage loans. (The Veterans Administration (VA), created in 1930, started doing the same in the late 1930s.) A borrower pays a loan premium for an FHA or VA insured home mortgage loan; the premiums are used as reserves and would flow to the lender in case an insured borrower defaulted. Like the HOLC, the creation of the FHA changed the mortgage market and set standards and practices which are followed to this day. By 1972, the FHA had insured about 11 million loans for purchase and about 22 million loans for home improvement. FHA's standardization made it easier, less risky, and cheaper to buy a home, thereby fueling the development of the mortgage market. The FHA helped to encourage suburbanization, but also

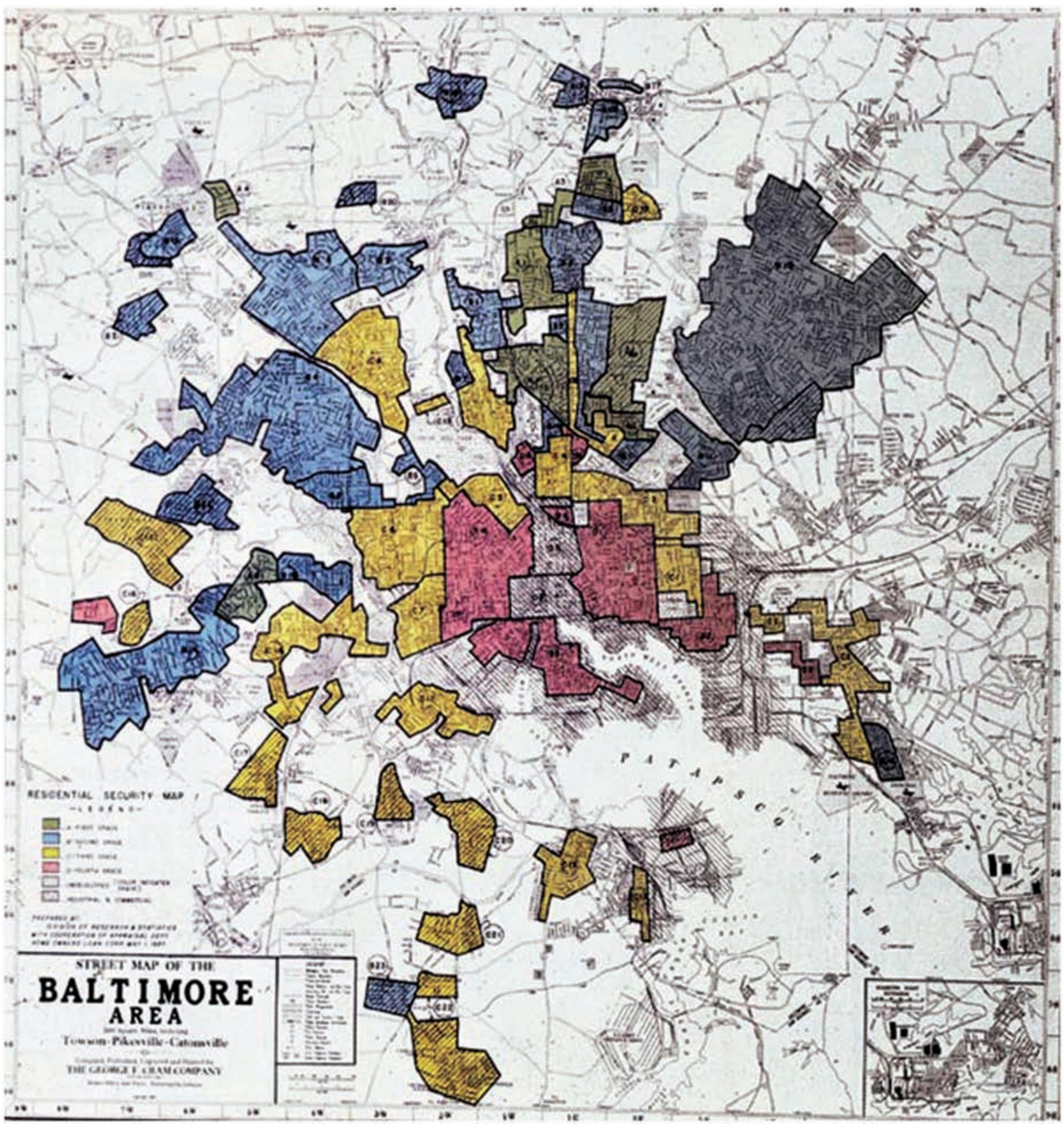

Figure 1 HOLC map of Baltimore, 1937. Source: National Archives.

"hastened the decay of inner city neighborhoods by stripping them of their middle class constituency", because "in practice, FHA insurance went to new residential developments on the edges of metropolitan areas, to the neglect of core cities" (Jackson, 1985: 206). The inner city areas were overlooked partly because they had lower appraised values for housing and FHA simply did not grant insurance in many of these areas. While FHA insurance was meant as a public back up to ensure the provision of mortgage loans, the FHA, like the HOLC, redlined areas in which private actors were also less likely to grant mortgages, or would only grant mortgages under less advantageous conditions, such as higher down payments and higher interest rates. Jackson concludes his groundbreaking work by claiming that the

> FHA also helped to turn the building industry against the minority and inner city housing market, and its policies supported the income and racial segregation of suburbia. (…) FHA exhorted segregation and enshrined it as public policy. Whole areas of cities were declined ineligible for loan guarantees; as late as 1966, for example, FHA did not have a mortgage on a single home in Camden or Paterson, New Jersey, both declining industrial cities. This withdrawal of financing often resulted in an inability to sell houses in a neighborhood, so that vacant units often stood empty for months, producing a steep decline in value. (Jackson, 1985: 213)

Gotham in his study of uneven development in Kansas city throughout the twentieth century arrives at similar conclusions arguing that the FHA's "insurance system and home ownership subsidies established a racially dual home financing market by refusing to insure mortgages in areas not covered with a racially restrictive covenant, thus denying mortgages to Black families, and channeling capital into suburban housing construction" (Gotham, 2002: 63). Indeed, discussions on redlining have been connected to debates on the causes of segregation and forms of racial discrimination: redlining is often seen as a form of (institutionalized) discrimination and as one of the causes of segregation. Together with many other mechanisms and policies, redlining is understood to be one of the causes of the erosion of cities, the decline of (inner city) neighborhoods and the mass exodus to the suburbs.

In the mid and late 1960s the FHA was forced to change its policies and make mortgage insurance available in formerly redlined and yellowlined areas. Today, FHA loans are actually responsible for increasing homeownership among black households. However, as research by Wyly and Holloway shows, even if controlled for income, black households significantly more often rely on FHA insured loans. The assumption is that black households rely on FHA insured loans because it is more difficult for them to acquire loans in the noninsured mortgage market.

Redlining in Post-War US

In the wake of the social rights movement, redlining returned to the political and research agenda's in the late 1960s and 1970s. Community based organizations in particular claimed that lenders were redlining large parts of American inner cities. In 1968, discrimination in housing – including mortgaging – became legally prohibited through the Fair Housing Act. In addition, the US government responded by implementing the Home Mortgage Disclosure Act (HMDA) in 1975, and the Community Reinvestment Act (CRA) in 1977. These acts require lenders to report granted loans by census tract and to provide credit to the local communities within the states in which they are active. Despite these acts and the related move of FHA to the inner city, research from the 1970s, 1980s, and early 1990s clearly shows the existence of redlining, mostly in inner city areas. Some of the earliest studies from the 1970s found evidence of redlining, but have been criticized for omitting important variables and for the use of aggregate level data. Studies from the 1980s have demonstrated that minority and inner city neighborhoods receive less credit, but these studies are susceptible to the criticism of preselection bias. Whatever the case may be, it is clear that community and legal struggles against redlining have been successful in diminishing redlining practices. The story of Atlanta, described superbly in an essay by Keating and colleagues, illustrates the potential impact which can be achieved through cooperation between community activism, university research (in this particular case an MA thesis in city planning), and media investigations and reporting (in this case the 1988 Pulitzer Prize winning articles by Dedman).

Research often uses HMDA data to demonstrate the continuation of redlining, but recent evidence of *de facto* redlining is scarce. But like earlier studies, these studies are also open to criticism. A recent study by Ross and Tootell finds evidence of redlining, but the probability of denial is not higher in ethnic minority neighborhoods, suggesting that there are other factors that are more important such as a high concentration of low income households. Yet, studies of discrimination in lending show that social groups as defined by race or area fare poorly in obtaining mortgage loans: for example, white neighborhoods receive three to four times more loans per 1000 mortgageable structures when compared to minority neighborhoods; even after controlling factors, such as income, a wide gap remains between different racial groups and between the neighborhoods they inhabit. Although this does not necessarily prove the existence of redlining, it does show that disparate treatment on the basis of both race and place is indeed taking place. Or, in other words, recent research carried out in the US shows that race based exclusion clearly takes place, but that redlining (place based exclusion) is often harder to prove and more contingent on the metropolitan area researched and on individual bank policies.

In conclusion, it is fair to say that while few would deny the existence of historic redlining in the US, recent evidence of redlining is scarce. Today, redlining in the US is less likely to take place because it is prohibited and banks have to make their lending data available for re search. In addition, changes in financial markets have made it more likely that lenders will charge higher interest rates and closing fees in high risk areas (yel lowlining) rather than that redlining will occur in these areas. Yet, we cannot simply conclude that redlining no longer exists in the US. Indeed, researchers who dem onstrate the existence of redlining practices can easily be accused of omitting variables. At the same time, re searchers who demonstrate the nonexistence of redlining can be criticized on the same grounds but also for not taking most of the lending process into account, for in stance by ignoring preapplication denials. Moreover, a great deal of research has overlooked the fact that lenders can easily adjust their spatial lending policies. Since redlining is measured on the district level, they can engage in cherry picking behavior by redlining part of a district as long as they grant mortgages in other parts.

More recently, authors have argued that redlining is replaced by predatory lending and subprime lending. This is possible due to the increased use of credit scoring. Credit scoring enables lenders (often through credit agencies) to map the social characteristics of households, and by coupling those to financial characteristics and past financial behavior, mortgage lenders try to predict the default risk of a borrower. Low risk prospective bor rowers can then be offered low priced loans and while high risk borrowers would have been denied a mortgage loan in the past, they can now get a high priced loan. The problem is not only that some loans become too high priced for households to handle, but also that several lenders in the subprime market (i.e., the mortgage market for nonstandardized and high risk loans) offer mortgage loans with high prices that no longer hold a real correl ation between risk and price. This segment of the sub prime mortgage market – often referred to as 'predatory lending' or 'reversed redlining', causes more harm to homeowners than it does benefit.

But how does this relate to redlining? It may seem that credit scoring systems have been so well developed to accurately predict borrower behavior that lenders no longer need such crude methods as redlining to limit risk. Although this may very well be the case, this does not necessarily mean that redlining has in fact become obsolete. First, redlining may take place next to credit scoring: this way formal application procedures that generate the input for credit scoring systems have to make no mention of 'zip code policies', while lenders can still enact redlining policies. Second, lenders may use geographical location as an additional criterion, in particular when reassessing 'review cases'. Third, be cause geographical location may be included in credit scoring systems, redlining may take place through those credit scoring systems, as geographical location may be the decisive factor in turning down an application, in particular if geographical location gets a relatively large weight in a credit scoring system. Even though geo graphical location will not be the only factor predicting exclusion or inclusion, the result nonetheless is *de facto* or actual redlining. Research from the Netherlands, for example, has demonstrated that redlining takes place after and next to credit scoring.

Redlining Outside the US and Outside the Mortgage Market

Redlining is well documented in the US mortgage mar ket, but does redlining also take place outside the US and outside the mortgage market? The answer to both these questions seems to be yes, it does. A small amount of research has focused on insurance redlining and an even smaller amount on retail redlining. In addition, evidence of redlining by postal services, delivery services, and mail order companies also exists, although most of it is of anecdotal nature or from news (paper) reports. Insurance redlining, in particular, has a potentially high impact on neighborhoods and their residents. The denial of prop erty insurance, for example, is not only problematic for residents, but may also withhold investment and may actually depress real estate values. It may also make it more difficult to get a mortgage loan. In addition, yellowlining is commonplace in almost all insurance markets, generally disadvantaging those in already dis advantaged situations. Sometimes, insurance in high risk areas may be two to five times as expensive, thereby ef fectively discouraging customers to apply. The effect of excessively priced insurance plans may be comparable to that of *de facto* insurance redlining.

Even though most research on redlining is carried out in the US, there are also a number of studies docu menting redlining processes in other, mostly Anglo phone, countries. Research in the UK, South Africa, Australia, Canada, and the Netherlands has demon strated the existence of redlining, whether in the 1970s and 1980s or in the 1990s and 2000s. In addition, recent research on Milan, Italy has demonstrated the existence of yellowing, albeit not redlining. Redlining is sometimes mentioned in other countries, but we are not aware of any redlining research in the Global South or elsewhere in Europe. Yet, this does of course not imply no redlining takes places outside the before mentioned countries. In general, the existence of redlining is denied up till (and often beyond) the point when redlining is demonstrated. For example, redlining was said not to take place in the Netherlands until recent research showed that it does. The most recent unambiguous example of redlining comes from the city of Rotterdam where at one point all the major banks (who together dominate the Dutch mortgage market) redlined large parts of the city. **Figure 2** shows a map of the southern part of Rotterdam, used in 1999 by the largest mortgage lender in Rotterdam. It is striking how similar the design of this map is to the US redlining maps of the 1930s. Typically, redlined neighborhoods in Rotterdam are areas with relatively high percentages of low income people, ethnic minorities, rental housing, and low priced real estate.

The Spatialization of Redlining

Redlining is an inherently geographical phenomenon, but geographers are relatively silent in the debate. Most empirical research on redlining is not carried out by geographers or sociologists, but by economists. Therefore, with a few exceptions, economists dominate not only the academic debate, but also the public and policy debates. Important exceptions include the work of Harvey, Reibel, Wyly, Holloway, Williams, Aalbers, and Smith. While the first six have done empirical studies of redlining in respectively the US, UK (Williams), and Netherlands (Aalbers), Neil Smith's work on gentrification includes an important discussion on the relation between redlining and gentrification. In the so called 'rent gap hypothesis', Smith argues that the growing gap between property values and underlying land values, partly as a result of redlining, in the inner city is the driving force behind gentrification. This has opened up a growing rent gap which has been exploited by the actions

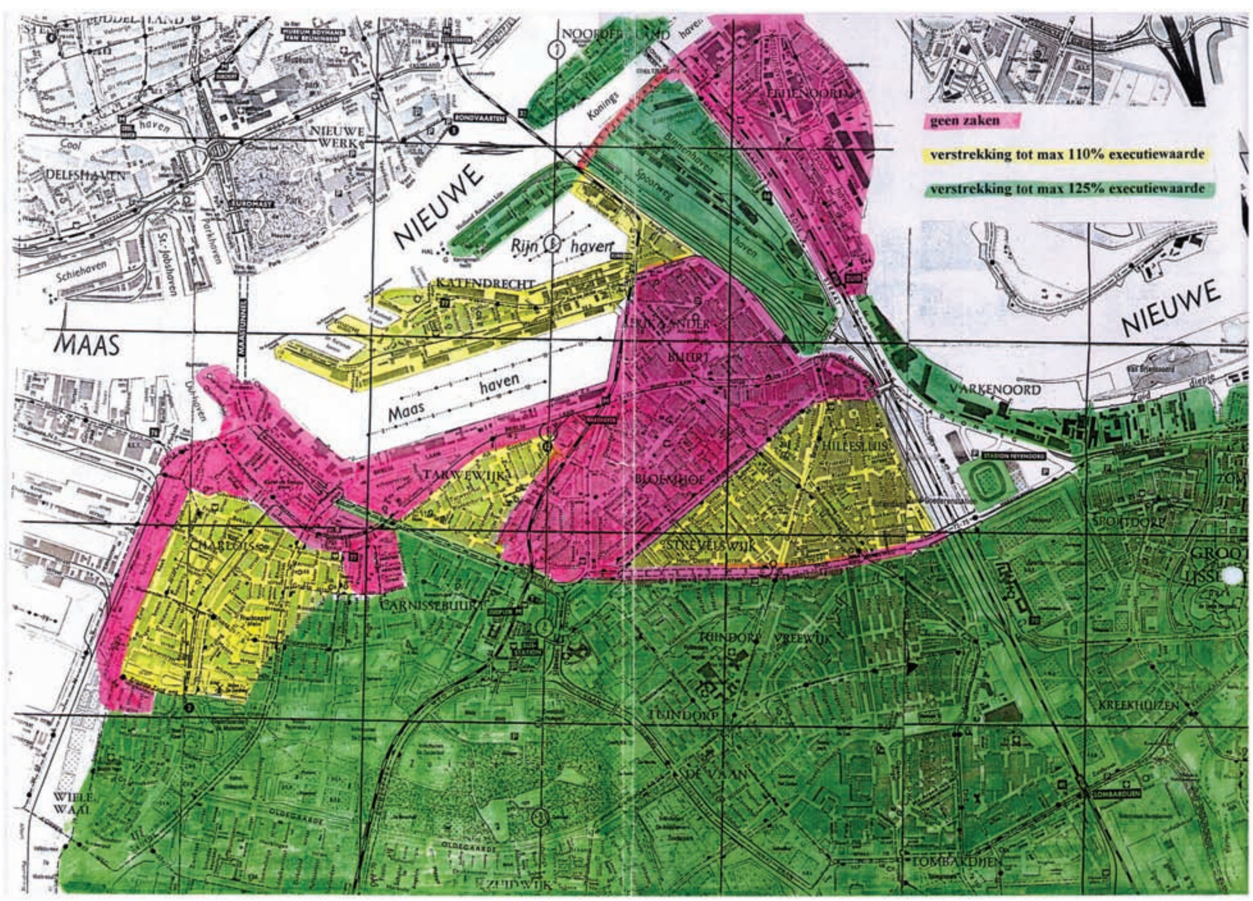

Figure 2 Lender's map of Rotterdam-South, 1999. From Aalbers, M. B. (2006). *Who's Afraid of Red, Yellow and Green? Geographies of Redlining and Exclusion in the Netherlands and Italy*. Ph.D. thesis, Amsterdam: University of Amsterdam. (http://dare.uva.nl/record/199435).

of property based capital, estate agents, developers, and the like, which has redeveloped or rehabilitated and gentrified undervalued inner city housing for profit.

David Harvey's work on 'capital switching' forms another inspiration for geographers working on redlining. In his work, Harvey not only stresses capital switching between the primary and secondary circuit (i.e., between different sectors of the economy), but also within sectors of the economy, between forms of property, and between places (geographical capital switching) and on different scales – from one neighborhood to another, from central city to suburbs, from periphery to 'prime city', or from one country to another – in order to exploit uneven development. The formation of submarkets in the built environment facilitates this process by creating a differ entiated rate of return which is necessary for the real ization of class monopoly rent. In Harvey's view, the built environment is shaped to meet the requirements of capital accumulation. As a result, financial institutions have a significant role in restructuring urban neighbor hoods, in residential differentiation, and in (re) creating housing submarkets. One of the things Harvey showed in his more empirical oriented work was that banks were redlining inner city areas. Contrary to the beliefs of neoclassical economics, he argues that geographical variations in mortgage lending are not merely a reflection of underlying demand side variations, because there are intrinsic contradictions in the structure of relations be tween different agents as well as between the individual agent and the structure of capital. This reinforces existing uneven geographies, but the existing geographies of uneven development do, of course, also influence the geography of mortgage finance and the segmentation of urban mortgage markets.

The work of Harvey, Smith, and other geographers shows us that we should not overlook the multilevel causation of social phenomena: the differences in lending behavior between neighborhoods depend not only on the differences among these neighborhoods within the city but also on processes taking place at different scales. The production of space on the local level is not just dependent on local processes but is a result of the interpenetration of global, national, regional, and local processes. Redlining at the level of the bank is not just a function of neighborhood variables such as income, eth nicity, and housing value; it is also mediated through processes at other levels, such as the institutional frame work at the national and city level. One cannot conclude that a high concentration of ethnic minorities or a low average income per household 'explains' redlining. It is important to see how lending behavior interacts with other processes at different levels and to show why red lining does not take place in neighborhoods of other cities.

The geographical analysis of redlining has demon strated that it is important to recognize that the occurrence of redlining is not of a structural nature, nor of an occasional nature, but rather of an endemic nature: it occurs under certain circumstances and conditions. Indeed, redlining is partly of a cyclical nature: it is in fluenced by the economic trends in the housing and mortgage markets as well as by the applicability of deliberate institutional control measures, whether they are laws or institutional structures such as private or public mortgage insurance. A decision to redline an area is made by lenders on a city level, but their decision is framed not only by the differences between neighbor hoods in the city, but also by processes and institutions at all scales. Future research would benefit not only from focusing on the different geographical levels or scales through which redlining is constituted but also from fo cusing on the search for actual or *de facto* redlining, be cause most of the recent literature has neglected this basic issue. It is incorrect to conclude that redlining does not exist as such merely because a local study has not found any evidence of redlining, just like it would be incorrect to suggest that redlining occurs everywhere because it occurs in one city. For the last two decades neoclassical economists have dominated redlining re search; it is up to human geographers to replace and resocialize the inherently spatial and social nature of redlining.

See also: Anti-Urbanism; Capital and Space; Financial Exclusion; Financial Knowledge; Financial Risks and Management; Gentrification; Ghettos; Harvey, D.; Housing; Land Rent Theory; Mapping, Race and Ethnicity; Neighborhood Change; Racism and Antiracism; Scale; Segregation; Segregation, Urban; Slums; Smith, N.; Suburbanization; Technological Change; Uneven Development.

Further Reading

Aalbers, M. B. (2006). *Who's Afraid of Red, Yellow and Green? Geographies of Redlining and Exclusion in the Netherlands and Italy*. Ph.D. thesis, Amsterdam: University of Amsterdam. (http://dare.uva.nl/record/199435).

Aalbers, M. B. (2007). Place based and race based exclusion from mortgage loans: Evidence from three cities in the Netherlands. *Journal of Urban Affairs* 29(1), 1 29.

Crossney, K. B. and Bartelt, D. W. (2005). Residential security, risk, and race: The Home Owners' Loan Corporation and mortgage access in two cities. *Urban Geography* 26(8), 707 736.

Dedman, B. (1988). The color of money. *Atlanta Journal Constitution*, May 1 4, http://powerreporting.com/color.

D'Rozario, D. and Williams, J. D. (2005). Retail redlining: Definition, typology, and measurement. *Journal of Macromarketing* 25(2), 175 186.

Eisenhauer, E. (2001). In poor health: Supermarket redlining and urban nutrition. *GeoJournal* 53, 125 133.

Gotham, K. F. (2002). Beyond invasion and succession: School segregation, real estate blockbusting, and the political economy of neighborhood racial transition. *City and Community* 1, 83 111.

Harvey, D. (1985). *The Urbanization of Capital. Studies in the History and Theory of Capitalist Urbanization*. Oxford: Blackwell.
Harvey, D. and Chatterjee, L. (1974). Absolute rent and the structuring of space by governmental and financial institutions. *Antipode* 6, 22 36.
Hillier, A. E. (2005). Residential security maps and neighbourhood appraisals. The Home Owners' Loan Corporation and the case of Philadelphia. *Social Science History* 29(2), 207 233.
Immergluck, D. (2004). *Credit to the Community. Community Reinvestment and Fair Lending Policy in the United States*. Armonk, NY: Sharpe.
Jackson, K. T. (1985). *Crabgrass Frontier. The Suburbanization of the United States*. New York: Oxford University Press.
Keating, L. E., Brazen, L. M. and Fitterman, S. F. (1992). Reluctant response to community pressure in Atlanta. In Squires, S. D. (ed.) *From Redlining to Reinvestment: Community Response to Urban Disinvestment*, pp 170 193. Philadelphia, PA: Temple University Press.
Merton, R. K. (1968). The self fulfilling prophecy. In Merton, R. K. (ed.) *Social Theory and Social Structure*, pp 475 490. New York: The Free Press.
Reibel, M. (2000). Geographic variation in mortgage discrimination: Evidence from Los Angeles. *Urban Geography* 21(1), 45 60.
Ross, S. L. and Tootell, G. M. B. (2004). Redlining, the Community Reinvestment Act, and private mortgage insurance. *Journal of Urban Economics* 55, 278 297.
Ross, S. L. and Yinger, J. (2002). *The Color of Credit: Mortgage Discrimination, Research Methodology, and Fair Lending Enforcement*. Cambridge, MA: MIT Press.
Smith, N. (1979). Toward a theory of gentrification: A back to the city movement by capital not people. *Journal of the American Planning Association* 45, 538 548.
Squires, G. D. (1992). *From Redlining to Reinvestment: Community Response to Urban Disinvestment*. Philadelphia, PA: Temple University Press.
Squires, G. D. (ed.) (1997). *Insurance Redlining: Disinvestment, Reinvestment, and the Evolving Role of Financial Institutions*. Washington, DC: Urban Institute Press.
Stuart, G. (2003). *Discriminating Risk: The U.S. Mortgage Lending Industry in the Twentieth Century*. Ithaca, NY: Cornell University Press.
Williams, P. (1978). Building societies and the inner city. *Transactions of the Institute of British Geographers* 3, 23 34.
Wyly, E. K. and Holloway, S. R. (1999). The color of money revisited. Racial lending patterns in Atlanta's neighborhoods. *Housing Policy Debate* 10, 555 600.

Relevant Websites

http://cml.upenn.edu
Cartographic Modeling Lab specializes in spatial analysis using Geographic Information Systems (GIS) technology.
http://dare.uva.nl
Digital Academic Repository of the University of Amsterdam.
http://www.urban.org/publications
Nonpartisan Economic and SocialPolicy Research Policymaking and the New President.
http://www.nhi.org
Predatory Lending: Redlining in Reverse.
http://powerreporting.com
Power Reporting Columbia Journalism Review.

Refugees and Displacement

R. Black, University of Sussex, Brighton, UK

Glossary

Asylum Seeker A person who has requested protection and refugee status from a nation-state other than his own.

Displacement Population movement that is forced, whether within a nation-state, or beyond.

Refugee A person who is forced to leave his home country, owing to a well-founded fear of persecution.

> It is racist to call American citizens refugees.
> Rev. Jesse L. Jackson, speaking after Hurricane Katrina displaced over 500 000 people from New Orleans, August 2005

> The people we are talking about are not refugees. They are Americans.
> President George W. Bush, also after Hurricane Katrina

America's Largest Displacement

In a rare display of public agreement, President George W. Bush and Jesse Jackson were both legally correct in asserting that America's largest forced displacement since the Civil War did not produce any refugees – even if some 30 000 households were still living in temporary trailer homes provided by the Federal Emergency Management Agency (FEMA) at the start of 2007, a year and a half after the storm. Yet the example of Hurricane Katrina highlights the ambiguity of the term refugee. The experience of Katrina victims also draws attention to the complex and often long term consequences of displace ment for individuals, communities, and states. In practice, there are at least three distinct ways – legal, analytical, and popular – in which the term refugee can be defined, each reflecting the context in which that definition takes place, and each producing widely varying estimates of the number and distribution of refugees. At the same time, there are also significant overlaps between the no tion of a refugee and other categories of individuals, in cluding displaced persons, migrants, and disaster victims.

Who Is a Refugee? Legal definitions

The most obvious starting point for any discussion of refugees is the 1951 Geneva Convention Relating to the Status of Refugees, which defines a refugee as someone who:

> … owing to well founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group or political opinion, is outside the country of his nationality and is unable or, owing to such fear, is unwilling to avail himself of the protection of that country; or who, not having a nationality and being outside the country of his former habitual residence as a result of such events, is unable or, owing to such fear, is unwilling to return to it. (Article 1 A (2) Convention Relating to the Status of Refugees, 28 July 1951)

According to the Geneva Convention definition of refugees, there were officially some 8.7 million refugees worldwide at the end of 2005 according to the United Nations High Commissioner for Refugees (UNHCR), with the largest numbers living in Africa and Asia. This was the lowest figure recorded since 1980, with numbers falling by one third over the previous 5 years.

Yet this Convention definition of a refugee is very much a product of its time. First, until extended by the 1967 New York Protocol, it encompassed only people displaced within Europe, as a result of events occurring before 1951; in other words, it was aimed at providing political solution for people displaced during and after World War II, and who found themselves with no legal protection. The Convention's Cold War origins also show through in its concern with individual persecution – a focus that has remained, even if regional variations have extended legal understanding of refugee status to include groups displaced by "external aggression, occupation, foreign domination or events seriously disturbing public order" (Organization of African Unity Convention, 1969) or "generalized violence, foreign aggression, internal conflicts, massive violation of human rights or other circumstances which have seriously disturbed public order" (Organization of American States, in the Carta gena Declaration, 1984).

Critical to all legal definitions of refugees are the notions that individuals must cross (or have crossed) an international boundary in order to be considered a refugee; that the causes of their displacement should be fundamentally political, rather than economic, socio cultural, or environmental; and that they must have had access to an official refugee determination procedure. These limitations have frequently been criticized – not ably the requirement to cross an international border. Indeed, calls for legal recognition and assistance to internally displaced persons (IDPs) who have not crossed an international frontier led to publication by the United Nations of a set of 'Guiding Principles on Internal Displacement' in 1998, which have been widely disseminated and implemented. However, despite a commitment on the part of UNHCR in 2007 to provide assistance to IDPs as a matter of global policy wherever they have government permission to do so, there have not, to date, been any revisions of either national or international law to extend the legal definition of a refugee.

One reason for this derives from a clear legal understanding of what it means to be a refugee, and why a special legal category is necessary at all. Legally, refugees are not simply people who have been displaced, or people in need of assistance; rather, they are people who fall outside the nation state system – or better, people for whom the relationship between the individual and the state has broken down – and who are entitled by law to protection from another state. By this line of reasoning, the nation state system is a universal one, in which states have a duty to protect all those considered to be citizens by virtue of birth or naturalization, and where all human beings can call on the protection of at least one state. In circumstances where a state is unwilling or unable to protect some of its citizens – indeed, where an individual has a reasonable fear they may be persecuted rather than protected – the category of refugee is designed to provide appropriate protection by another state, until such time as the fear of persecution is removed. However, to gain such protection, refugees must 'seek' it – hence the category 'asylum seeker', meaning someone who has requested protection from another state, but has not yet received it – and the principle that those who remain within their own country are not, *de facto*, seeking the protection from another state (and so cannot be called refugees).

A further important element of the legal definition of a refugee that follows from the above discussion is the principle that in international law at least, refugee status is temporary – something that is conferred by states for the period that a well founded fear of persecution remains. However, in practice, many states – especially those in Europe and North America – have chosen to grant refugee status not as a temporary, but as a permanent measure, with refugees often being entitled to apply for full citizenship after a period of time. Indeed, as some states have grown weary of granting permanent refugee status, they have tended to introduce a series of new categories of temporary protected status or temporary leave to remain, rather than reasserting refugee status as a temporary concession.

The fact that receiving the protection of another state is an entitlement under international law for someone who fits the legal definition of a refugee, and the *de facto* permanence of refugee status in many European and North American states, has both practical and statistical consequences. On the one hand, the granting of permanent residence rights to refugees makes becoming a refugee one of a limited number of ways to obtain the substantial privileges accorded to European and North American citizens and residents. As such, it undoubtedly encourages some without a legitimate claim to demand refugee status – although many with a genuine claim may also be rejected, or unable to access the determination procedure in the first place.

On the other hand, it is clear that the real number of refugees, even using a legal definition, is not the same as the official number of refugees listed by UNHCR. This is because (1) circumstances change in home country, so that individuals may no longer fit the legal definition, even though they retain refugee status; but also (2) some people lose refugee status when they obtain full citizenship of their new host nation, but continue to fit the legal definition of a refugee, since the circumstances in their home country which led them to flee remain unchanged.

Who Is a Refugee? Analytical Definitions

If legal definitions of a refugee are sometimes confusing, this is even more so if we consider refugees or displaced people from the point of view of explanation in social science. For here, our purpose in defining a refugee is not to establish legal entitlement to protection by a specific state, but to set out a clear and theoretically grounded category of people who are analytically distinguishable from other categories of people. In such a context, it may well be appropriate to consider refugees and IDPs as part of the same analytical category – both may, for example, have left their place of habitual residence for the same reason; both may have suffered physical losses, social dislocation, and psychological trauma in doing so; and both may have similar material needs and circumstances.

In such circumstances, the fact of having crossed, or not crossed, a perhaps arbitrary international political boundary may be irrelevant to their objective circumstances; both are examples of displacement, rather than migration, an analytically different phenomenon. Such parallels are explicitly recognized by UNHCR too, who provide protection and assistance to so called people of concern, regardless of their formal legal status. At the end of 2005, over 20 million people of concern to UNHCR were noted in their annual statistical review, a figure that rose by 6% during 2005, but remained largely unchanged since 1998 when records began.

Yet there remain significant problems in placing boundaries around the category of a refugee in analytical terms, even if the incentive to misallocate between categories, produced by legal entitlements, is removed. For example, we might conventionally distinguish displacement as involving a lack of choice in whether individuals move or not, or situations in which choice is severely constrained, from migration, where choice does exist. This distinction may have analytical importance for social scientists, since it is an indicator of an individual's independence of action, and may influence their attitude to other actions, such as their integration in the host society, or their interest to return.

However, such a distinction between refugee/migrant, or displacement/migration (or forced/voluntary migrant/migration) hides the fact that the degree of choice faced by individuals is highly varied, with hard and fast distinctions between choice and no choice very difficult to maintain. Thus, political activists who flee persecution as a result of their political activities may have exercised a choice to carry out these activities, even if they are ultimately left with no choice but to leave. Equally, an individual may choose to leave at a particular time, or in a particular direction, responding to opportunities to make the best of a difficult situation – as with the political activist who chooses exile during a foreign visit. Yet those choosing to leave their country to seek a better life in Europe or North America may also find their choices severely limited. In this sense, to characterize a refugee or forced migrant as someone who does not exercise choice, or whose choices are more constrained than voluntary migrants may be a massive oversimplification.

Another possible distinction is between refugees, whose motives for leaving are fundamentally political; and migrants, whose motives for leaving are fundamentally economic (or sociocultural, or environmental). Such a division may be relevant analytically since some of the insights, for example, of the 'New Economics of Labor Migration' (NELM) – a theoretical perspective that helps to explain among other things why migrants send remittances – rely on the assumption that peoples' motivations for migration are fundamentally economic. Yet such distinctions are highly problematic too, not least in the sense that most people who move juggle multiple motivations, which increase or decrease over time. Thus, those who fear persecution because of their ethnicity, or membership of a particular social group, may also find themselves excluded economically in their country of citizenship, providing powerful economic as well as political motives for them to leave. Moreover, regardless of the proximate cause of a particular movement, the behavior of refugees and migrants may ultimately remain essentially economic (or political, or social, depending on the theory in question).

In practice, in most or all of the dimensions where refugees might be distinguished from migrants, we find a continuum of circumstances rather than a simple dichotomy. For example, some have argued that it is important to distinguish environmental refugees – either as a legal or an analytical category. Yet environmental degradation or disaster may also reflect neglect of areas inhabited by politically isolated or vulnerable groups – or worse, deliberate manipulation of or damage to an environment that is intended to lead to displacement of particular groups. In such cases, the environment is arguably simply a route through which political persecution takes place, rather than an autonomous cause of movement in its own right.

The lack of a clear distinction between displacement and migration is also as much an analytical opportunity, as a limitation, since it implies that many of the theories and concepts that are applied to migration could equally be applied to refugees. One example is the emerging field of transnationalism – the notion that individuals are increasingly connected across international borders through family, social, business, and other networks, thus inhabiting a series of overlapping 'spaces' that are unconstrained by simple geographical location. This is a field of study that, like the NELM, has been largely developed in the context of voluntary, economic migrants, but where the forced transnationalism of refugees may share numerous similarities with that of other migrants.

Similarly, there is an increasing recognition that, like other migrants, refugees may face discrimination in job markets, be socially excluded, and pose challenges for community cohesion; but where they may also simultaneously contribute to national economies, remit money to their home countries, and aspire to return to invest their savings, or live out their retirement. This suggests the need for analytical flexibility in definitions, rather than the delimitation of rigid boundaries that characterizes legal discourses.

Popular Definitions: Why It Matters to be a Refugee

Refugees are distinguished from other migrants for both legal and analytical reasons. However, there is also a more popular sense in which a refugee is different. As Zetter noted in a seminal contribution to the new field of refugee studies in 1988, "the label 'refugee' both stereotypes and institutionalizes a status." In other words, it denotes a category of people who are victims, institutionalizing a dependence on others; it acts as a humanitarian category, designating people who are poor, and in need of assistance; it also highlights a group of people facing extraordinary changes in to their lives. Yet this reflects perhaps the most problematic aspect of the use of the term 'refugee' – one that has neither legal nor analytical clarity, one where the reasons for application of the label may be as analytically interesting as the label itself.

For example, it is here that the concerns of Jesse Jackson, or President Bush, noted above, really lie: Americans fleeing Hurricane Katrina cannot be refugees not because they have not fled persecution or moved to another nation state, but because of the negative stereotyping associated with the word refugee – that they are poor, helpless, and dependent. For the US government in particular, there may even be a case for denying that displaced people are refugees, in order to minimize the visibility of the displacement, the numbers involved, and the extent of associated poverty, since all of these appeared to reflect quite badly on government (in)action during the crisis.

Yet interestingly, for those arguing for the opposite position – the introduction of the term 'environmental refugee' into international law and/or popular discourse (among whom Katrina evacuees would likely have been included) the purpose is equally less to provide analytical or legal clarity, and more to make the ideological point, that people displaced by environmental degradation or climate change are as deserving of public attention as those displaced by political violence or persecution. This is arguably not unconnected with the very large estimates of the numbers of environmental refugees worldwide – potentially over 200 million according to some sources, even if these are highly disputed.

Also important is that the labeling of refugees denotes not only rights or entitlements in a legal sense, but also often highly prejudicial attitudes and institutional forms of behavior that draw on such attitudes. Concretely, a humanitarian industry has emerged that is charged with securing the rights and entitlements of refugees not only to protection from persecution, but also to minimum standards of assistance, yet which has frequently been accused of failing to meet such standards because it stereotypes refugees as vulnerable or dependent.

Most important among the practical standards and guidelines that humanitarian assistance is expected to meet is the 'Humanitarian Charter and Minimum Standards in Disaster Response' developed by the Sphere Project, a collaboration of international nongovernmental organizations (NGOs) launched in 1997. Sphere now encompasses a handbook, training guidelines, and networking designed to ensure accountability among refugee assistance agencies not only to international donors who provide financial support, but also to refugees themselves who are the recipients of aid. Legal entitlement is not as much the key factor as is the fact of institutionalized assistance, and the commitment of participating agencies to ensure that those assisted are able to have some input into the nature of this assistance, and how it is delivered.

Such commitments to international standards also come out of deep criticism of international agencies for failing to see refugees as actors, with both knowledge and experience, capable of participating in the search for solutions to their situation, and frequently let down by those supporting them.

Geography and Refugees

What then is the role of geography in studying a group of people, particularly where legal and popular definitions of the group appear more certain and powerful than analytical definitions, and where even the size and distribution of the population is open to question, with different estimates spanning two orders of magnitude?

There are various ways in which geographical insights have been applied both to describe and to understand the situation of refugees. At a more technical level, organizations such as UNHCR have increasingly turned to geographic information system (GIS) technologies to map and analyze refugee movements; this has included contributions to camp planning, the compilation of detailed demographic analysis of camp populations (which in turn spawned an international campaign on behalf of warehoused refugees, trapped in protracted refugee situations – defined as living in camps for five years or more), spatial analysis of potential impact of refugee on sensitive environments such as forests or endangered species, as well as practical mapping of infrastructure and services of importance in the efficiency of refugee assistance.

In contrast, within the academy, the study of refugees from a geographical perspective has been more diffuse. Within the UK, and some other northern countries, it has included analysis of the patterns of choice among asylum seekers and refugees in terms of where they have sought asylum, where they settle within cities and countries, and whether they intend to return; attempts to identify the sources of information on which such decisions are made; and analysis of the experiences of refugees after arrival and settlement. In developing countries, there has also been a significant amount of geographical attention directed to refugee livelihoods, and how the socioeconomic strategies and outcomes intersect with humanitarian aid policy.

In much of this research, there is little to distinguish geographical work on refugee mobility from that on migration more generally. For example, like this wider literature, there has been a focus on how age, gender, ethnicity, or educational or occupational background interact with other geographical factors to influence outcomes. A growing literature on gender and refugees has analyzed not only the relative experiences of men and women in exile, but also how exile influences social relations between men and women, and the consequent transformation of gender roles. Attention has also been paid to the broader connections between refugees and development, or between refugee integration and social cohesion in cities.

At the same time, geographers have also explored the geopolitics of displacement and refugee assistance, casting a critical eye on the claims of states and international agencies to act impartially or merely 'technically' in the face of humanitarian need, and highlighting the deeply entrenched political motivations of policy whether at a national or international level. For example, whichever the source and magnitude of refugee statistics, it is fairly clear that the majority of refugees – however defined – remain in poorer countries of the world. Yet to read popular discourse on the subject, and to see the actions of governments seeking so called regional solutions that would avoid the necessity for asylum claiming in the richer north, we might easily conclude otherwise. Geographers have also contributed to understanding how legal, social, and political discourses and practices impact on the experiences of refugees and asylum seekers, often increasing their marginalization from the economy and wider society.

Refugees: Theory and Public Policy

The study of refugees and displacement is still in its infancy, at least in terms of its institutionalization within academia, and the production of a coordinated and coherent body of knowledge. One of the reasons for this is perhaps the strong public policy interest in refugees, which has simultaneously ensured a supply of funds for research, but encouraged this research to focus on specific public policy objectives, rather than contributing to wider theoretical debates within social science. Thus, refugees have often been treated somewhat as exceptions to broader social science paradigms, reflecting the exceptional nature of policy toward them, whether in the form of special immigration rules, or humanitarian rather than development policy. Even research that has critiqued policy positions has tended to be designed more for advocacy than for theoretical reflection.

Yet in many ways, the study of refugees and other displaced people also provides a potential insight into processes of social, economic, and political transformation, as refugee situations are ones in which such transformations are often rapid and observable. Indeed, the sometimes turbulent changes experienced by refugees could even be seen as the key to societal transformation, rather than the more gradual change over time on which much geographical and social science research concentrates. There is a parallel here with debates in physical geography over the relative importance of gradual change and extreme events in processes of land degradation or climate change. In this sense, refugee researchers arguably not only have the opportunity to highlight the experiences of refugees, but also stand in a unique position to contribute to understanding more of the human condition in general.

See also: Biopolitics; Geopolitics; Migration; War.

Further Reading

Bailey, A. J., Wright, R. A., Mountz, A. and Miyares, I. M. (2002). (Re)producing salvadoran transnational geographies. *Annals of the Association of American Geographers* 92, 125 144.

Black, R. (2001). Fifty years of refugee studies: From theory to policy. *International Migration Review* 35, 55 76.

Castles, S., Crawley, H. and Loughna, S. (2003). *States of Conflict: Causes and Patterns of Forced Migration to the EU and Policy Responses*. London: Institute for Public Policy Research.

Crawley, H. (2001). *Refugees and Gender: Law and Process*. Bristol: Jordan Publishing.

Hyndman, J. (2000). *Managing Displacement: Refugees and the Politics of Humanitarianism*. Minneapolis, MI: University of Minnesota Press.

Nash, A. (ed.) (2002). *Special Issue: Geography and Refugees*. *Geoforum* 56.

Nolin, C. (2006). *Transnational Ruptures: Gender and Forced Migration*. Aldershot: Ashgate.

Stewart, E. (2005). Exploring the vulnerability of asylum seekers in the UK. *Population, Space and Place* 11, 499 512.

UNHCR. (2006). *The State of the World's Refugees: Humanitarian Displacement in the New Millennium*. Oxford: Oxford University Press.

Zetter, R. (1988). Refugees: A label and an agenda. *Journal of Refugee Studies* 1, 1 6.

Relevant Websites

http://www.state.gov
Bureau of Population, Refugees and Migration, US Government.

http://www.blackwellpublishing.com
Disasters: the Journal of Disaster Studies, Policy and Management, Blackwell.

http://www.forcedmigration.org
Forced Migration Online.

http://www.odihpn.org
Humanitarian Practice Network.

http://www.icar.org.uk
Information Centre on Asylum and Refugees (UK).

http://jrs.oxfordjournals.org
Journal of Refugee Studies.

http://www.reliefweb.int
Relief Web.

http://www.sphereproject.org
Sphere Project.

http://www.unhcr.org
United Nations High Commissioner for Refugees.

http://www.refugees.org
US Committee for Refugees and Immigrants.

Regeneration to Renaissance

T. Butler, King's College London, London, UK

Glossary

Collective Consumption A term used by Manuel Castells to refer to those publicly provided goods and services necessary to ensure the reproduction of labor power.

Commodity The process whereby an object is imbued with monetary value as a result of the labor invested in its production.

Modernity The period after the Enlightenment and associated with individualism, universalism, and 'affective neutrality'.

Multiple Deprivation Refers to the institutional situation whereby individuals who are deprived on one measure are often found to be similarly deprived on a number of dimensions.

Urban Regeneration to Sustainable Cities

Urban regeneration has emerged over the last 15 years as a key field in urban geography and is the clearest link between the academic and policy communities. Like other terms (e.g., restructuring), regeneration did not become part of the academic lexicon until the late 1980s. Indeed, a search through one of the new computerized bibliographic databases in the early 1990s would only have thrown up medical references. The policy interest in remaking cities, however, has a long and continuous history although the terminology used to describe this has undergone regular revision as Lees (2003) has argued. What started off in post war (or more accurately wartime) Britain as reconstruction has gone through a series of reincarnations as Lees points out: reconstruction, redevelopment, regeneration, and renaissance all describe ways in which the 'take' on what to do with our cities has been subject to political and policy change over the past 60 years. In recent times, the focus has now changed to building sustainable cities, or more frustratingly given the definitional imprecision of both terms, sustainable communities. Each time the policy change can be traced, on the one hand, to a failure to deal with the problem of urban stress and its manifestations in terms of crime, social dislocation, dysfunctionality, and design and, on the other, to a new emergent understanding of how urban and social problems have become entwined in new ways. Each time, this has implied a new approach to social inequality (or, more recently, social exclusion) that would lead to a better, more harmonious yet economically more competitive society. This has become the holy grail of urban regeneration.

Following a brief account of how an understanding of urban development (urbanization) has always been contextualized and confused within the overall development of capitalist and industrial society, the remainder of the article focuses on the experience of the last 60 years drawing mainly (but not exclusively) on the British experience. Britain has probably had a more developed 'urban policy' than any other comparable nation yet paradoxically remains one of the least planned economies in Europe. The article concludes with a brief account of the apparent change in urban policy from a concern with growth and competition to cohesion and sustainability and questions how fundamental a change this might be.

Background

The urban has always been a problem for social scientists. The founding fathers of sociology discussed industrialization, capitalism, and urbanism often in the same breath. Friedrich Engels' *The Condition of the Working Class in England* (1971) is the classic account of capitalist industrialization and of how Manchester was partitioned according to the needs of the capitalist production system. The city was, to use a term later coined by Mumford (1961) in his classic work on the city since antiquity, the 'container' for capitalist industrialism. Many sociologists have continued to conflate capitalism with industrialization and urbanization although, as Savage *et al.* (2003) point out, it was modernity rather than capitalism that was caught up in the changes that occurred to the city in the nineteenth century. They argue that many of the classic statements about the city (e.g., Louis Wirth's (1938) famous article 'Urbanism as a way of life') are essentially more about modernity and how social life changes than about the capitalist organization of production. The truth of this is perhaps to be found in the resurgent postindustrial city of today in which the decline of industrial production (in the global North) has, if anything, led to the intensification of urbanity and density rather than its dilution, which was more associated with the suburbanizing tendency of its highpoint in the middle of the twentieth century. Thus the linkages between the city, social life, and economic organization are, at best, fuzzy and this may account for the continued failure of urban policy in the guise of urban regeneration to deal with what were essentially the problems of capitalist industrialism (and now postindustrialism). Policy was aiming at the wrong target and therefore – not sur prisingly – missing it.

The Urban Question

In the 1970s, Marxists were struggling to reinvent a Marxism that was both non Stalinist and reflected the changed priorities of capitalism following the student led revolts of 1968. In what became the manifesto for a new urban sociology, Castells' *The Urban Question* (1977 – originally published in French in 1972) argued that the definition of urban was essentially social as opposed to spatial. He argued that there were two components to the production of value in capitalist economies: the pro duction of commodities by labor power and the repro duction of labor power itself. Marxism had focused on the former and tended to ignore the latter. Capital, for its part, would like to ignore the latter and transfer its re sponsibilities for social reproduction onto workers and their families. In the context of post World War II ad vanced industrial economies and, as a consequence of successful struggles by trade unions and other workers' organizations, the state had taken over from capital the responsibility for organizing what Castells termed 'col lective consumption'. This comprised issues of social reproduction such as healthcare, housing, education, and public transport. Cities, he argued, were now as much 'factories for producing labour power' as they were for producing commodities. In this context, the state as the provider of collective consumption goods, became the site for struggles over the nature of collective con sumption – what was sometimes referred to as the social wage; increasingly 'social movements' could be seen as playing the same role in regard to the provision of col lective consumption as trade unions played in relation to wages and conditions of work in the sphere of production. This led Castells to argue that the definition of the urban now lay in the manner and extent to which collective consumption was organized. This new urban sociology meshed with urban policy and the argument that urban regeneration was as much about 'soft' as 'hard' issues.

From a more specifically geographical perspective, Castells' work was paralleled by the development of a Marxist turn in geography that is associated with the work of David Harvey (1973, 1982), Doreen Massey (1984), and Ed Soja (1989). These writers focused on the massive changes in the management of the space econ omies in advanced industrial economies from the late 1970s onward (a process they termed 'restructuring') and the ways of explaining them. Cities were becoming de populated and were subject to a process of disinvestment by capital. However, having been devalued, some sub sequently acquired a new investment status as post industrial service centers but often in so doing found themselves with the 'wrong sort' of populations. It fell to the state to re engineer these social processes because they threatened the new investment; not only were the old populations perceived as threatening but they also did not offer the requisite skills required by the new industries which were now at their economic cores.

There were early signs of these discontents in the United States in the so called 'long hot summers' of the 1960s in which many of the downtown districts of major US cities such as Detroit and Los Angeles were the scenes of sustained rioting by African Americans. This resulted in a massive flight from the cities from which many of them never recovered – Detroit's inner districts remain largely abandoned to this day. The response, in an early and atypical (for the US) exercise in state led urban regeneration, was the so called American War on Poverty (AWOP) declared by President, Lyndon Johnson, in 1964 some of whose programs were imaginative and relatively successful, others less so. The relative success of 'soft' regeneration projects such as Operation Headstart is in stark contrast to the long history of failure of 'hard' re generation projects like the Pruitt Igoe public housing project in St. Louis which was eventually demolished.

Developments in Britain paralleled those in North America – albeit less dramatically. The Notting Hill riots and those in the St. Anne's district of Nottingham in 1958 were allegedly sparked by the sight of white women associating with black men but in both cases reflected much deeper changes in urban social structure. Formerly working class areas were undergoing huge changes in their social, economic, and cultural character. However, it was another 10 years before this notion of the black threat to British urban culture took root in the political culture. In 1968, the Conservative shadow Cabinet minister, Enoch Powell, made his infamous 'rivers of blood speech' in which he posed the future where 'our cities' would be populated by people of an 'alien race' leading to a situ ation which he claimed would lead to untold strife and communal conflict. Some London dockworkers marched to Parliament in support of Powell. In response, the La bour Home Secretary, Roy Jenkins, set up in 1969 the Community Development Projects (CDPs) in some of the most stressed inner city. Things, however, did not go ac cording to plan and the workers in most of these projects argued that there needed to be a far more dramatic intervention to deal with job loss and housing deterior ation than the government envisaged. It was argued that simply identifying the dimensions of the so called 'mul tiple deprivation' being experienced in these inner city working class communities, in which jobs were dis appearing and where housing conditions were sub standard, would not solve the problem of long term urban decline. The Home Office eventually closed the projects down in 1976 but not before they had defined the agenda of dissent with regard to Britain's urban program. Their critique of urban policy intervention was that the kind of economic restructuring and urban change that was to become increasingly familiar in the 1980s required intervention in housing and welfare services (Castells' collective consumption) and – crucially – an economic development program that specifically addressed the issue of employment. Without such policies urban policy would be bound to fail. It was precisely such policies that were anathema to the 1979 incoming Conservative government led by Mrs. Thatcher. Her response was an experiment in market led regeneration, which drew on the private sector, with government playing a minimal enabling role. This was epitomized by the London Docklands Development Corporation (LDDC), which was set up in 1979 to lead the regeneration of the former Docklands areas in East London. Ultimately, the bankruptcy of the largest developer (Olympia and York), the need for infrastructural investment, and to maintain social consent led to a resurgence of interest in more traditional forms of state led regeneration. Increasingly, as we see below, this has emphasized 'soft regeneration' policies focusing on education, social inclusion, and training rather than physical infrastructure and buildings.

A Short History of Regeneration

This section gives a short history of urban regeneration in Britain and follows the model outlined by Lees (2003) from reconstruction to renaissance to which it adds a further stage of sustainability.

Post-war reconstruction

The prime task facing the incoming Labour government in 1945 with the war having been won and millions of servicemen coming home with high expectations (in contrast to what happened after World War I) was that it would be a land 'fit for heroes to come home to'. Hundreds of thousands of houses were needed as a result of the destruction of Britain's cities by the Luftwaffe in the blitz and many more thousands were in a terrible state of repair following years of neglect. At the same time, the new industries sparked by the war effort in electronics would need to be built afresh in new locations. The Abercrombie Plan, hatched during World War II to plan for the post war future, argued for the creation of a green belt around London and other major cities and the development of 'new towns' where new housing standards, services, and employment opportunities could be offered. At the same time, the architects working for the London County Council and others were drawing up ambitious plans not just for new housing using the latest technologies but also new schools and hospitals such as were envisaged in that other great wartime report (The Beveridge Report), which promised to deal with the five 'great evils' of 'want', 'disease', 'ignorance', 'squalor', and 'idleness'. There was a major program of reconstruction, which aimed by the early 1950s to achieve 250 000 new homes a year through a mixture of private and state provision. It also resulted in the building of many of the new towns envisaged by Abercrombie and enshrined in the New Towns Act and the associated Town and Country Planning Act. These were built mainly around London (Stevenage, Harlow, Basildon, Billericay, etc.) but also around northern cities (Skelmersdale, Cumbernauld, and Washington) and in Scotland (Livingstone) and Wales (Pontypool, Cwmbran). Other bombed out cities were comprehensively redesigned from the center out around North American precepts about the car (Plymouth, Southampton, and Coventry, for example). By the late 1950s, the immediate phase of reconstruction was completed and Harold Macmillan could move from being minister of housing, having met the 250 000 target, to becoming prime minister secure in the claim that 'you have never had it so good.'

Comprehensive redevelopment

During the 1959 general election campaign, Macmillan further, and more controversially, claimed that 'we are all middle class now' and, questionable though this claim was, it indicated the great change that was beginning to occur in British society with the decline of industrial employment and, probably more importantly at that time, working class culture. Cities were beginning to lose their middle and increasingly their working class residents and had to compete with the burgeoning suburbs as sites for consumption. The state at local and central government level began to plan ambitious redevelopment projects that would both rid them of much of the still substandard housing and build North American style malls to lure the newly affluent working and middle classes into their emporias of consumption. This was the era of regional development policies as the old steel, shipping, and coal towns of the north, Wales, and Scotland began to plan a future based on new industries and a newly affluent working class. It also brought together machine politics, state policy, and big business all looking to advantage themselves, so the architect, John Poulson; T. Dan Smith, the Labour leader in the North East; Reginald Maudling, the Conservative politician; and Bovis founded a marriage of convenience to regenerate Newcastle (and other cities) with a massive development plan using system built housing much of which has subsequently been knocked down. Poulson was eventually jailed for corruption and this did much to discredit the urban redevelopment of the period that was meant to improve working class lives but, in many people's minds, led to the enrichment of a few and the incarceration of the many in the new high rise slums. Dunleavy's (1981) study of the workings of this system in the London

Borough of Newham points to the ways in which the state and private sector were inextricably linked during this period. The collapse of Ronan Point (one of Newham's many system built high rises) early one morning in 1968 did more than anything else to seal the fate of this era of redevelopment. Matching this large scale public housing was the commercial redevelopment of city centers, which came to be symbolized by the ubiquitous Arndale centers beginning with one built in the old ship building town of Jarrow in 1961 and spreading across the UK. These de velopments provided an outlet for the increasing funds available for investment by the expanding pension funds that had come to dominate London's stock market. These city center redevelopments were based around car use and roads came to dominate their city centers – nowhere more so than in Birmingham, the car capital of the UK.

Urban regeneration

By the 1970s with rising unemployment, factory closures, and the social problems referred to above to which the CDP initiative was intended to respond, these forms of hard engineering solutions were seen to be increasingly ineffective and indeed counterproductive. This was par ticularly acute in port cities whose economies were being undermined by the shift to containerization, ro ro (roll on, roll off) ferry traffic, and nonunion labor – nowhere more than Liverpool and London. In Liverpool, the Toxteth riots of 1981 and the acute unemployment of East London hastened the incoming Conservative gov ernment to try a bold experiment in hands off regener ation where the state facilitated development rather than organizing and financing it. The legislation enabled urban development corporations (UDCs) to take charge of the regeneration process in selected urban areas. In practice, the only major UDC was the LDDC that was charged to develop the riversides of the three dockland boroughs of Newham, Tower Hamlets, and Southwark. The LDDC was given a limited life of 18 years at the end of which the planning powers would be handed back to the re spective boroughs. It was this control of planning and the ability to trade the land owned by nationalized industries (Port of London Authority, British Rail, and Gas Board, primarily) that were the main assets of the LDDC. Under its plain speaking first chief executive, Reg Ward, the LDDC used these powers to the full to get the area rebuilt: social cohesion was very much a secondary consideration to persuading investors to invest. The planning powers given to the LDDC enabled them to create an Enterprise Zone which meant that investors paid no taxes for the first decade and there were few of the usual restrictions on use or style. Despite this, it was a slow process persuading private investors to put their money in an area that was previously associated with communist shop stewards and to which there was no public transport or much prospect of any. By 1987, it was felt sufficient progress had been made on the con struction front that some attention could be paid to issues of social inclusion which were beginning to threaten the success of the project and a new phase under a new chief executive took place. This was all but killed off by the major recession beginning in 1988 which hit London and the south east of UK particularly hard and the project plunged back into crisis. The major developers of Canary Wharf – Olympia and York – owned by the Toronto based Reichmann brothers, went bankrupt and a coalition of the government and the banks had to provide the fi nance for the Jubilee Line Extension and finish off the building of 1 Canada Square (the iconic building often known as 'Canary Wharf'). The recession, together with the growing alienation of many living in Britain's inner cities (which became apparent in the loss of Conservative seats in most inner city constituencies in the 1987 elec tion), forced a new turn toward socially inclusive urban policies. Outside the LDDC area, the Single Regener ation Budget (SRB) funding was developed; cities were able to bid for regeneration money for not only hard but increasingly soft projects. This policy was continued under New Labour when it came to power in 1997. Key to both governments' approach was that local areas had to compete and they had to build into their proposals evaluation procedures which showed that they had met their key targets and objectives. By the time the LDDC was wound up in 1997, it was widely felt to have been a success in developing what was once the most run down and degenerated part of London – albeit at the cost of continued social exclusion.

Urban renaissance

When Tony Blair came to power in 1997, he was deter mined to make cities places that people wanted to live in and he had listened long and hard to the architecture lobby that argued you could create livable cities through a combination of good design practices, dense communities, and social policies designed to maximize social inclusion. Barcelona was often pointed to as 'the' example of how this could be done. Richard Rogers was asked to chair an urban taskforce and in due course it produced its report (Urban Taskforce 1999) which argued this case forcibly. The emphasis on good design led many to argue that gentri fication not regeneration was the real agenda: to attract the new class of professional and managerial workers into the city as part of an overall economic strategy not only to develop traditional service industries in finance and legal services but also in the new cultural and creative indus tries. In many respects, this policy has been wildly suc cessful not only in London but also in many of Britain's major cities which have seen 24/7 communities develop in old city center commercial and industrial buildings. However, it has run up against the same old problem of social division that has plagued all previous attempts to resolve the problem of urban change where a new economy comes up against the remnants of the workforces of the old urban economy. Nevertheless, some policies developed by New Labour such as New Deal for Communities (NDC) have enabled many of the most notorious rundown estates (such as the Holly Street estate in Hackney) to be regenerated and have raised the quality of life of many of the country's most deprived populations.

Sustainable development

While the urban renaissance approach taken by New Labour has had undoubted successes despite leaving many in still unregenerated housing estates and increasingly in 'buy to let' private sector properties with no rights of tenure, regeneration policy has moved on. In London and the south of UK the problem is recognized as an insufficiency of affordable housing and elsewhere – arguably – an oversupply leading in both cases for government to argue that there is market failure. In the south east of UK the planning system prevents sufficient new houses being built in the right places thus threatening London's economic viability by imposing huge costs on its key workers. Elsewhere, the provision of too much housing of inappropriate quality has been seen as standing in the way of attracting the 'right social mix'. The solution in the south of UK has been to find new places to build and elsewhere – in part at least – to reduce the housing stock in order to create demand for appropriate housing. In the southeast, of UK sustainability has now become a dangerously double edged word in regeneration discourses. In order to sustain London's economy, it is necessary to guarantee the supply of reasonably priced housing with good access to the city; increasingly this has meant building on London's flood plain, especially in the Thames Gateway to the east of the capital which is now planned to become a dense example of new urban design. Given the importance of London to the UK economy, this is seen as a necessary contradiction. Increasingly, regeneration policy now has to take into account these multiple and often contradictory appeals to sustainability.

Comparative Perspectives

This has been a largely British focused article whereas urban regeneration (like its sibling gentrification) has now become an increasingly global phenomenon. This is not simply a function of Anglocentricity on the part of the author but reflects the unique way in which de industrialization, the state, and urban policy came together in the context of the need to rebuild Britain's cities in the aftermath of their destruction during World War II. That reconstruction, as we have recounted, rapidly became overshadowed by wider events: the declining competitiveness of Britain's industrial production compared to other countries in Western Europe and Japan, the developing economic crisis of the 1970s, and the growth of urban deprivation and social exclusion. Britain has had an urban policy throughout this period – at no time more so than during the Blair government – but this is a comparatively recent development elsewhere in Europe where, for the most part, urban regeneration has coincided with the rolling out of the European Union and neoliberal hegemony. In North America the focus has been on the private sector and fiscal incentives which have, in effect, made gentrification the chosen tool to turn around the many failing cities. The most recent, and perhaps dramatic instance of this approach, has been in the rebuilding of New Orleans in the aftermath of Hurricane Katrina.

There are now striking examples of urban regeneration across Europe, often drawing on European Union structural funds to deal with the consequences of industrial decline. For the most part, this has relied on cities, city regions, and regional governments working in close partnership; Barcelona and Bilbao provide examples of cities which successfully leveraged funds out of both the central government and the European Union for ambitious schemes to transform themselves into centers of postindustrial production. Nevertheless, only France and the Netherlands have had longstanding national policies on urban regeneration and – like Britain – both were largely a response to large scale immigration into what were previously ethnically cohesive cities. In France, the construction of *grands ensembles* on the outskirts of major cities to replace the *bidonvilles* originally thrown up by first generation immigrants from North Africa rapidly became ethnic ghettoes of the dispossessed. Many became the focus for the riots that threatened, in the autumn of 2005, a rerun of 1968 in which the protagonists were ethnic minority youth rather than disaffected students. French urban policy has struggled to regenerate these areas particularly in recent years when successive interior ministers have proclaimed a willingness to 'get tough' so as not to be outflanked on the right by the fascist Front National led by Jean Marie Le Pen which notoriously came second in the presidential election of 2002. In the Netherlands the state has enabled developers to enter in agreements with city authorities to develop land on a long term basis that has been characterized as being 'state led gentrification'. The line between urban regeneration and gentrification has become increasingly blurred across Europe. Smith's claim that gentrification is now the preferred model for 'rolled out' neoliberalism would appear increasingly justified.

Emerging Themes

In conclusion, it can be argued that there is an increasing convergence of both hard (physical) and soft (social) regeneration agendas around a common theme of sus tainability in which issues of environmentalism, social density, and social cohesion are being brought together. There appears to be an increasing rapprochement here between British models of state led regeneration, American ones of private developer led gentrification, and the Dutch tradition of state inspired gentrification in developing a notion of sustainable regeneration. This is furthest developed in the UK with a 'private–public partnership' model in which the state underwrites long term investment by the private sector in an increasingly wide range of projects involving not only infrastructure but also the provision of services. This new conventional wisdom has become the latest way of resolving the in herent conflict between competition and cohesion that has riven urban regeneration policy.

See also: Gentrification; Neoliberalism, Urban; Sustainability, Urban; Urban Policy.

Further Reading

Buck, N., Gordon, I., Harding, A. and Turok, I. (eds.) (1995). *City Matters*. Basingstoke: Palgrave.
Castells, M. (1977). *The Urban Question*. London: Edward Arnold.
Dunleavy, P. (1981). *The Politics of Mass Housing in Britain, 1945 1975: A Study of Corporate Power and Professional Influence in the Welfare State*. Oxford: Clarendon Press.
Engels, F. (1971). *The Condition of the Working Class in England*. Oxford: B. Blackwell.
Florida, R. (2002). *The Rise of the Creative Class and How It's Transforming Work, Leisure, Community and Everyday Life*. New York: Basic Books.
Foster, J. (1999). *Docklands: Cultures in Conflict, Worlds in Collision*. London: UCL Press.
Harvey, D. (1973). *Social Justice and the City*. London: Edward Arnold.
Harvey, D. (1982). *The Limits to Growth*. Oxford: Basil Blackwell.
Lambert, C. and Boddy, M. (2002). *Transforming the City: Post Recession Gentrification and Re urbanisation*. Bristol: ESRC Centre for Neighbourhood Research.
Massey, D. (1984). *Spatial Divisions of Labour: Social Structures in the Geography of Production*. Basingstoke: Macmillan.
Mumford, L. (1961). *The City in History: Its Origins, Its Transformations and Its Prospects*. New York: Harcourt Brace.
Raco, M. (2007). Building Sustainable Communities: Spatial Policy and Labour Mobility in Post war Britian. Bristol: Policy Press.
Savage, M., Warde, A. and Ward, K. (2003). *Urban Sociology, Capitalism and Modernity*. Basingstoke: Palgrave.
Smith, N. (2002). New globalism, new urbanism: Gentrification as global urban strategy. *Antipode* 34(3), 427 450.
Soja, E. (1989). *Postmodern Geographies and the Reassertion of Space in Critical Social Theory*. London: Verso.
Urban Taskforce (1999). *Towards an Urban Renaissance Final Report of the Urban Task Force Chaired by Lord Rogers of Riverside*. London: Department of Environment Transport and the Regions.
Wirth, L. (1938). Urbanism as a way of life. *American Journal of Sociology* XLIV, 1 24.

Relevant Websites

http://www.wcml.org.uk
Community Development Projects 1969 77, Working Class Movement Library.

Region

J. Tomaney, Newcastle University, Newcastle upon Tyne, UK, and Monash University, Melbourne, VIC, Australia

Introduction

Region is a key concept in human geography. A concern with the region is bound up with beginnings of the discipline in the modern era. The concept of the region has also interested anthropologists, ecologists, economists, historians, planners, political scientists, and sociologists, among others. If geography has considered itself an integrating discipline, then the region has been a conceptual and empirical meeting ground. Regional geographers placed the study of areal delineation at the heart of the discipline. In short, it is difficult to imagine a concept more central to geographical inquiry. More recently, within the discipline of geography, the region has been a key site for debate and dispute, involving quantitative, humanistic, Marxist, and post structuralist approaches.

The scientific controversies that surround the region tend to mirror the ebb and flow of its fortunes as a political and cultural category. In the modern era of the nation state, the region often has been associated with particularism. This association can be traced back to the birth of the modernity, for instance, to the rebellion in La Vendée in 1793–96 against the revolutionary government in France. The brutal repression of the insurgency by the '*colonnes infernales*' devastated the region in an episode that has been described as genocidal. However, Karl Marx, in his *Eighteenth Brumaire of Louis Bonaparte* in 1852, used the term Vendée to signify reaction in the face of progress. Thus, there is a recurring idea of the region, especially as a cultural or political category, as symbolizing backwardness in the face of integrative nation building projects. As a cultural and political mobilizing force the region continually confronts the ideological power of nationalism. By contrast, the region as a statistical, administrative, or economic category has grown in importance in many states throughout the world. To varying extents, political compromises and controversies surrounding the region manifest themselves in the arrangement of subnational administration at level of the region, despite the power of nationalist doctrines. Although there are important differences between terms such as region, province, and state, collectively they suggest the persistence of the region as a scientific and political category at the subnational level. From the 1980s, the European Union (EU) witnessed the emergence of 'new regionalism' in which regions/states gained a greater political prominence and a new economic status. The region/province/state represents an important political division in Canada, Russia, China, Indonesia, India, Mexico, and Brazil, among others. Some geographers went so far as to identify a 'resurgence of region' at the end of twentieth century as an economic, cultural, and political entity and as a subject of study.

The etymology of region in English is complex. It appears in the early fourteenth century from the Latin word *regio* meaning direction, boundary, or district, linked to *regere*, meaning to direct or rule. Although its use has been cloaked in ambiguity, it evolved to mean a definite part of a territory or an administered area, which is part of a larger whole. In the era of the nation state, it has tended to mean a subordinate part of a larger state, most obviously seen in the use of the adjective regional. In the nineteenth century, regionalism emerged as term to describe incomplete integration or centralization of a state. The terms region and regionalism have subsequently been adopted by those willing valorize their territory and its interests. In cultural terms, the adjective regional has been used to indicate something subordinate or inferior and, according to Raymond Williams, signaling a 'limiting judgement', connected to the metropolitan provincial distinction in national life. The term can be used to describe geographical phenomena at various spatial scales (Europe, the Mediterranean, and Iberia), but its use in contemporary human geography has tended to refer primarily to territories within nation states.

Evolution of the Idea

The earliest forms of geographical inquiry showed a concern with identifying and describing regions. The origins of Chinese geography can be traced to the period of the Warring States (481–221 BC) and focused on describing the physical and human attributes of the nine provinces. Detailed descriptions of regions can be identified in the literature of the Mediterranean civilization, including that of Herodotus (525–484 BC), Pliny the Elder (AD 23–79, Ptolemy (AD 90–168), and in Strabo's (63/4 BC–AD 24) opinionated *Geography*. Strabo offered a naturalistic account of the differences between places. Islamic geography drew on this tradition. Abu Abd Allah Muhammad al Idrisi (1100–65) sought to divide the known world into regions. Similar regional descriptions can be found in the sources remaining from the European Age of Discovery. Travel literature can be found in many cultures after this period. For instance, Matsuo Bashō's *Oku no Hosomichi* or *Narrow Road to the Interior* described in prose and haiku poetry the remote and dangerous northern reaches of Honshū during Edo Japan. Although these early attempts at regional description helped to describe a world where places are situated in relation to one another, they did not, typically, provide a framework to explain these differences in a systematic way nor did they achieve a satisfactory categorization of regions. In the European case, it was the scientific advances of the Renaissance and, subsequently, in the Age of the Enlightenment, as well as the rediscovery of the work of Ptolemy and other ancient geographers that led to a more systematic concern with analyzing and explaining regional differences. But at the same time the growth of a self consciously formal science placed an emphasis on the universality of space rather than the particularity of place. Varenius, in the early seventeenth century, was one of the first to make the distinction between a general and a specific geography in his *Geographia Generalis* (first published by Elsevier in 1650). At the same time, advances founded on improvements in cartography and the growth of natural sciences and related developments in land surveying gave new possibilities for more exact accounts of regions. Very often they had the biggest impact where existing or new administrative authorities sought to define their territories, such as Thomas Jefferson's book *Notes on the State of Virginia*. In Europe, processes of nation building, the emergence of a modern economy and society, the rapid growth urbanization, and the changing nature of the countryside formed the backdrop to a concern with describing the region such as in Daniel Defoe's *A Tour Thro' the Whole Island of Great Britain* or William Cobbett's *Rural Rides* through southern England.

The growth of geography as a scientific discipline was both cause and effect of the interest in the region as evidenced by the regional sensibilities present in the work of von Humboldt and Ritter in the first part of the nineteenth century, which represented the first modern attempts to develop a systematic geography of the world's regions. Regional geography in different forms became influential in many countries by the turn of the twentieth century. In Germany, geographers saw the study of regions as their main task. Hettner, who saw himself writing in a tradition that included Ptolemy, Strabo, Kant, Humboldt, and Ritter, outlined a chorological approach in which the purpose of inquiry was to focus on the causal relations between assemblages of phenomena which lend individuality to particular places. For Hettner, geography was the science of the Earth surfaces, studied according to their local differences, be they continents, lands, regions, or localities. The comparative study of these differences (*Länderkunde*) was the means by which a general geography would be created. Hettner's approach had an influence beyond Germany, notably in the United States through the interpretive role of Hartshorne.

Paul Vidal de la Blache, the founder of French regional geography, writing at the beginning of the twentieth century, was influenced by these developments in German geography. Vidal de la Blache was concerned primarily with chorology (from the Greek *choros* meaning space) and with how groups made use of their environment, based on a view of the Earth as a living organism. In this method, a region can be regarded in relation to its physical characteristics (*pays*) or with regard to its human characteristics (*genre de vie*). A *pays* is a natural region with some homogeneous physical characteristics. A *genre de vie* is a way of living or local culture. The environment not only sets limitations and but also offers potentials for personal and cultural development. Hence, this view of the region is associated with the doctrine of 'possibilism' as an alternative to environmental determinism, although Vidal de la Blache never used the term. From this perspective regions were viewed as embodying a 'personality' and given a value which disturbed the French republican tradition of political centralization. In practice, French geography developed a tradition of identifying bonds between society and its geographical milieu, typically reflected in the production of detailed studies of particular *pays* and a strong emphasis on fieldwork. Moreover, this approach later influenced Sauer's notion of the 'cultural landscape', with culture as the agent and nature as the medium in the production of the region.

French and German debates on the region were typically bound up with nation building exercises. However, a concern with the region in the work of Peter Kropotkin and Élisée Reclus connected to themes of self sufficiency and autonomy drawn from anarchist theory. Politically, this found a manifestation in some of the regionalist movements that emerged during the Spanish Civil War and continued to influence politics in Spain thereafter.

French and German debates about the region found echoes in British geography in the early part of the twentieth century, particularly in the work of Herbertson, and also in the intellectual milieu centered on Sir Patrick Geddes, sociologist and planner, whose work strongly influenced a generation of geographers, as well as many others, including Lewis Mumford. Strongly influenced by the French sociologist Frederic Le Play – who was also an influence on Vidal de la Blache – Geddes identified the trinity of 'place–work–folk' (*Lieu Travail Famille*) as the underpinning of concern with the region as a focus for science and social action. Geddes was a biologist by training and, with a strong interest in evolutionary theories, viewed the region as a living organism. The *leitmotif* of his approach was the Valley Section designed to demonstrate interrelationship of landscape and life. Geddes emphasis on the region as a focus for social action accessible to public influence was especially significant and inspired a regionalist movement in planning in Britain and beyond, one example of which was the scheme for the regionalization of England proposed by C. B. Fawcett in 1919, but much influenced by contemporary French thinking about regions.

In 1905, Herbertson identified a series of 'major natural regions' at the world scale defined according to relief, climate, and vegetation. This approach not only stimulated many efforts to identify natural regions in Britain, but also led to innovative efforts to emphasize the human role on the process of regional formation. H. J. Fleure proposed 'human regions' defined by their anthropological characteristics and their 'lived experience'. Fleure placed particular emphasis on boundaries and border zones where ideas and beliefs were shared and social progress made. Herbertson himself later emphasized the importance of 'mental and spiritual' dimensions of region.

Taken together, these approaches – centered on area differentiation and chorology – which privileged the region as the site for geographical enquiry contributed to the emergence of 'regional geography' as the dominating paradigm in the discipline in the early twentieth century, especially in the English speaking world. Underlying this approach were assumptions about the exceptional nature of geography as a discipline with a unique empirical focus and methodology. This approach was perhaps best exemplified in Hartshorne's influential monograph *The Nature of Geography* in 1959, developing an earlier essay of 1939. By the end of the 1950s, however, this approach to the study of the region was falling into disrepute as a result of a number of more or less integrated conceptual developments. First, the notion of chorology was criticized as the unity and isolation of the *pay* was destroyed by industrialization, urbanization, and migration. Regional geography was accused of giving a static account of the region, while the pace of social change was quickening. In addition, the emergence of a mass industrial society, accelerated by the onset of total war, led to a growing demand for geographical research that could produce generalizations that could inform the policies of national planners. The focus on the region was deemed to be anachronistic and left geography as a discipline isolated from important developments during social science in this period. Moreover, the region was seen as an ambiguous concept, on the one hand, a theoretical entity – like 'an atom or a neutron', which could not be seen but whose character could be inferred from its effects – and, on the other hand, a heuristic device for the organization of geographic data.

In practice, some of the criticism leveled at the traditional conception of the region overlooked the sophistication of geographers such as Ritter and Vidal de la Blache who were not merely concerned with providing descriptions of isolated regions. They had a multiscalar understanding of the region and saw study of the region as a means to build a global geography. These insights were largely disregarded as the region became a battleground in the philosophical and methodical controversies that marked the discipline from the 1960s. Geography came under the influence of the positivist approaches that were influencing social science at that time. The privileging of the region as a unique combination of phenomena, together with geography's claims to be an integrating discipline, were attacked by those who saw the discipline's task as producing verifiable general laws. This conflict between the value of idiographic and nomothetic approaches to explanation – symbolized in the debate between Hartshorne and Schaefer – was an old one in geography, but the status of the region now lay at the heart of it. Although regional geography never entirely disappeared, by the 1960s the emphasis had shifted away from a focus on the uniqueness of regions to the search for more 'systematic' accounts of geographical phenomena.

The most obvious manifestation of this shift was the 'quantitative revolution' in geography. While affecting many aspects of the discipline, the application of quantitative approaches had particular implications for geographical understandings of the region. The proponents of this new understanding eschewed a particular focus on the region, and instead sought to apply mathematical approaches and theories to the study of geographical patterns, with the aim of formulating generalizations about space and recasting geography as a 'spatial science' capable of generating verifiable laws. This approach chimed with changes occurring in state and society in industrial countries, including the growth of the state planning functions and increased interregional and international flows of people, commodities, and money. The emphasis in the new approaches was on modeling these flows and relationships, drawing on the work of geographers such as Christaller and Lösch in location theory. Alongside and overlapping with geography, a new discipline, that of regional science, developed. Its concepts and methodologies were rooted in neoclassical economics, but its approaches became influential within geography. The moving force in these developments was not a geographer but an economist, Walter Isard, and research in this field became embodied in a new Regional Science Association, although it proved influential within geography. The growth of regional science drew on and reflected the growing importance of economics as a discipline and in the policy world, as well as the growth of computational forms of analysis and, more broadly, the pervasive influence of logical positivism. In addition, it spoke to the needs of rapidly growing industrial economies in the postwar period providing data to aid planning. A concern with the region did not vanish as a result of the quantitative revolution and the rise of regional science, but the problem was reformulated. Defining the region became a special type of classification problem. While classification of populations into mutually exclusive categories is a vital element of quantitative techniques in social science, the need to define contiguous areas meant that they were not necessarily as homogeneous as other classified groups. Thus, the development of complex mathematical models and theories became central to the definition of regions.

Despite its success in establishing a scientific basis for the discipline and apparently holding out the possibility of making itself useful to policymakers, both quantitative geography and regional science became targets of the general critique of positivism which developed in geography from the 1970s. Regional science entered a period of decline, although never entirely disappearing. A central difficulty with regional science was the ever increasing ambition of its models, which, in the end, had dubious practical relevance and became fixated on elaborating spatial patterns, rather than understanding the social processes that they concerned. Moreover, during the 1970s, equilibrium based approaches drawn from economics struggled to account for widening regional inequalities and dramatic changes in the spatial structure of the economy. These large scale changes inspired an interest in structural approaches, which offered an account of the radical nature of the transformations occurring in the order of regions in industrial societies. Critics maintained that the heavy and exclusive emphasis on quantitative methods frequently resulted in a sophisticated form of mathematized description and that the task of research should be to understand the spatial organization of social relations rather than simply describe visible geographical distributions.

Marxist approaches, in particular, situated the region within the context of class relations of capitalist political economy. In this perspective, regional change is not just an outcome of economic and social change but is integral to it. The region is a manifestation of the capitalist imperative for a spatial fix as a means of extracting value. Early Marxist analysis emphasized the power of capital to determine the fate of regions through investment/divestment and valorization/devalorization and focused on the region as a local effect of general processes of capitalist accumulation, competition, and technological change. Within the contemporary Marxist tradition there are numerous disputes and differences, but in addition to focusing how capital produces regions through, notably, investments in production and the built environment, there has tended to an emphasis on the way the region is constituted through class practices and class conflict. The Marxist perspective on the region, though, is distinguished by its emphasis on the way these regions remain embedded within the logic of capitalist accumulation.

Humanistic geographers, as their name suggests, drew on perspectives from the humanities to emphasize the importance of human agency in the production of regions. Humanistic geography not only has emerged as a contemporary critique of the positivism of the quantitative revolution, but has also criticized Marxist geography for its neglect of the experiential dimension of the region. Drawing on techniques from history and literary and art criticism, it was concerned with the meanings that people attached to particular regions. This led to a concern with particularity of regions that echoed traditional geography, but was very often informed by cultural theory and characterized by the use of methods from ethnography and historiography. This tradition has been most prepared to rejuvenate and adapt older concepts, such as Carl Sauer's notion of cultural landscape.

Post structuralism represents a third important contemporary perspective on the importance of the region. Here the focus is on language and the degree to which meaning is produced within rather than reflected through language and by focusing the relationship between signifier and signified. Taken to its fullest conclusion, in the post structural view the identification of the region as a territorial entity is eschewed and, instead, the region is viewed as an object of discourse with no necessary or pregiven status. Post structuralists then are concerned with deconstructing the discourse surrounding the region in order to undo any claims concerning its coherence.

On the face of it, these theories present quite different versions of the region. Indeed, the conceptualization of the region has been decidedly disputatious throughout the history of geographical thought – no more so than in recent times. At the same time, despite disagreements, it is also possible to find overlaps between these perspectives. It is evident that by the 1980s geographers were rediscovering the region, although in ways that differed from the approaches of the early twentieth century, notably insofar as the new approaches were informed explicitly by new developments in social theory. Indeed, during the 1980s and 1990s, claims about a 'new regional geography' began to appear in the literature. Typically, geographers returned to the region as a means to make a distinctive contribution to debates about social change. In part, the rediscovery of the region was a response to developments in the world at large. Processes of economic restructuring associated with the end of the long postwar period of growth, deindustrialization, and, later, new forms of economic growth and waves of international economic and political integration, appeared to have a markedly uneven regional geography. In addition, the region appeared to be reemerging as a focus for cultural and political identification, albeit formed around new and complex intersubjective understandings. Region specific networks of social interaction provide the basis for the reproduction of regional differentiation. The region came to be understood as a medium for social interaction – as a process by which institutions and individuals interact in space and time, albeit these interactions are understood as being conditioned by the operation of structures and power.

The new regional geography observes the persistence of regional diversity. Despite the advance of modern ization and apparent standardization of economic and cultural processes, distinctive regions – with diverse economic, cultural, and political characteristics – remain visible. Their existence draws our attention to the persistence of diversity and inequality in society. Understanding the region then is a means to under standing society itself. The region is a medium and an outcome of social processes. The region is a historically contingent process. The coherence of a region comes from the power of particular groups or interest to embed their values in ways which engender the solidarity ne cessary to specify a distinctive area. The region is ana lyzed now as a social structure in constant evolution.

These insights prompt a renewed interest in accounts of the region as historically contingent social constructions rather than physical entities. This theoretically informed new regional geography pays attention to the relationships between regions. It demonstrates a renewed focus on the uniqueness of regions. It suggests that the region matters in ways which generalized models of explanation are not able to demonstrate or analyze because social processes play out differently in different social context. Moreover, regions are not merely receptacles of social change, but have effects of their own on social processes. With this context in mind, it is possible to identify a series of treatments of the region as a cartographic, economic, cultural, political, and ecological phenomenon, drawing on a range of different perspectives.

Cartographic Region

Almost all states collect statistical data on a regional basis to aid planning and administration. Attempts to collect and analyze regional data often comprise a major part of the activities of government statistical offices. Data can be collected from within a predefined region, or region can be defined in relation to patterns derived from data. Some of the issues that surround the creation of statistical re gions are exemplified in the efforts of the EU create a consistent method of defining regions. The EU's stat istical agency, Eurostat, considered two ways of defining regions, using formal criteria (e.g., 'forested/not forested') or in terms of administrative boundaries. In the 1970s, faced with the need for comparable regional data across Europe, Eurostat opted for the second alternative as the only way of ensuring that the regions thus defined could be used to collect data on a range of indicators. This led to the creation of the Nomenclature of Territorial Units for Statistics (or NUTS in French acronym) classification. In each member state, a three level hierarchy of regions was established on the basis of 'existing' administrative regions or groupings of these. Regions at NUTS level 1 typically are either nation states (Latvia or Malta) or large subnational units (such as Scotland or Bavaria), each of which usually comprises a number of NUTS 2 regions (examples of this level include the Autonomous Communities in Spain or the 'regions' in France). In turn, these are made up of NUTS 3 regions (such as the 'Kreis' in Germany). Although broadly very stable over time in a number of countries, the NUTS classification has been amended several times, most recently in 1995, 1999, and 2003.

Notwithstanding the marginalization of quantitative treatments of the region, geographers have continued to examine quantitative data as means of defining regions that move beyond administrative convenience. The idea of the functional region captures the idea of a territory characterized by spatially related human activities. The US census pioneered attempts to go beyond adminis trative areas in its data presentations in the 1950s. This effort evolved into the concept of the daily urban system (DUS). This approach initially defined metropolitan re gions based on their employment core together with their commuter hinterlands. In essence, a DUS describes the pattern of labor market activity within a functional re gion on a typical working day, but has been extended, on occasion, to include housing, retail, and social patterns. This approach to defining labor market regions has been adopted in most other countries such as the UK, where Travel to Work Areas have become an important plan ning instrument.

Contemporary patterns of economic restructuring in many countries have been associated with the widening of functional regions, so that they spill over adminis trative boundaries as commuting and other flows take place over longer distances. The extension of the func tional region raises questions of how they are best gov erned if existing regional structures do not match them. In the United States, the debate about metropolitan regionalism addresses these issues. In Europe, the con temporary focus on city regions has developed in re sponse to the widening economic footprint of cities, although the notion of the city region can be traced back in the geographical literature at least 60 years. The growing extent of the functional region is seen most obviously in the case of 'global cities', in which are concentrated the command and control functions of major corporations and financial institutions. The func tional regions surrounding London, New York, and Tokyo extend many miles. In the case of New York, its economic footprint extends into the states of New York, New Jersey, and Connecticut. The London functional region extends across a large area of southern England (**Figure 1**). In one perspective, northwestern Europe can be seen as dominated by a serious of economically powerful city regions (**Figure 2**). Similar processes can be identified in emergent global cities: the Seoul

Figure 1 Functional Labour Market Regions in Great Britain. Source: Champion, T, Coombes, M. G, Raybould, S. and Wymer, C. (2007). *Migration and socio-economic change: a 2001 Census analysis of Britain's larger cities* Policy Press, Bristol.

Figure 2 Functional Regions in North West Europe. Source: Cheshire, P. and Gornastaeva, D. (2003). Cities and Regions: Comparable measures require comparable territories, Cahiers de l'Iaurif n°135, pp 12 32.

functional region contains 40% of the Korean population. Smaller provincial cities have begun to conceive of themselves as city regions in the context of heightening territorial competition, although, generally, these devel opments have been associated with widening regional inequalities between regions.

Economic Region

Contemporary patterns of economic restructuring have led geographers (and other social scientists) to renew their attention in the region as an economic category. Neoclassical economics sees the market as self regulating and, typically, has viewed regional problems as the spatial manifestation of the failure of the factors of production to adjust to new market signals. In the Keynesian per spective the economy is characterized by disequilibrium processes. Although Keynes himself paid little attention to regional questions – despite their importance in 1930s' Britain – his essential insights that governments should intervene in the economy to manage disequilibrium processes, underpinned the forms of 'spatial Keynesian ism' that marked many postwar industrial economies. Governments intervened to redistribute investment from faster to slower growing regions, including rural regions, through systems of incentives and controls.

The economic position of the region in industrial societies has been analyzed through the prism of 'spatial divisions of labor'. Patterns of industrialization in Europe and North America, and latterly in Asia and Latin America typically produced distinct regional economies based on spatial sectoral specialization, agglomeration economies, and increasing returns. Alfred Marshall's original account of these 'industrial districts' in England emphasized also how the knowledge and understandings that underpinned these industries 'were in the air'. Classically, this manifested itself in the growth of what are now known 'traditional industrial regions', such as coal mining, steelmaking, shipbuilding, and engineering regions, like the *Ruhrgebiet* in Germany or the Pittsburg/western Pennsylvania region in the United States. The regions entered periods of restructuring and relative decline during the 1970s and 1980s, there the problems being attributed to overspecialization in outmoded sec tors. Patterns of regional development during the 1960s and 1970s were affected by the growth of new industries. Highly capitalized, mobile firms were able to separate their activities by function. Thus, sectoral spatial spe cialization was replaced by functional spatial separation. Thus, firms typically located routine production activ ities requiring semiskilled workers in peripheral or lower cost regions, including declining industrial regions, while headquarters, administration, and research and development were located in metropolitan regions. Dealing with the problems of traditional regions became a central task of regional policy in Europe and North America in the 1980s and 1990s and the location de cisions of mobile firms were often influenced by the availability of government financial incentives and con trols. New spatial divisions of labor typically overlaid older ones. It was the combination of these layers which produced distinctive regional economies.

By the 1990s, however, geographers (and other social scientists) were identifying a new logic of economic re structuring with the region at its center. New forms of competitiveness and economic organization based on networks, 'untraded interdependencies', cultures of interfirm cooperation, and the rediscovery of agglomer ation economies were associated with rapid growth of some regions in Europe, the United States, and Asia. Often these regions were home to industries of the new knowledge economy – such as Silicon Valley in Cali fornia or Cambridge in England. But some researchers identified the resurgence of regions making more tradi tional products using network forms of organization, such as the industrial districts of the central and northern Italy (the Third Italy), where local firms produced premium leather goods, ceramics, and agricultural machinery, or Baden Württemberg in southern Germany where net works of middle sized firms (*MIttelstand*) produced car parts and machine tools. In some cases, geographers pointed to the distinctive role of public authorities in underpinning regional production systems with support for training, marketing, and innovation. This led some geographers to proclaim the rebirth of the region as a central concept in the paradigm of a network economy. Policymakers sought to ape what they believed to be 'best practice' which these regions were deemed to exhibit. In practice, there were as many differences as similarities in these 'regional success stories'. Moreover, the economy remains dominated by large firms, which are able to switch production between regions. The debate about the status of regions as distinct (resurgent) economic entities is still beset by the question of whether the examples which have dominated the geographical literature (such as Silicon Valley) represent the 'leading edge' of a 'new paradigm' of regional development, or whether they represent exceptional 'successes' which will be difficult for other regions to replicate. Such economic arguments often are mobilized to justify regionalization efforts, for instance, in the case of the proposed *doshu sei* reforms in Japan in the 2000s. Whatever is the answer to these questions, significant regional inequalities remain the order of the day in many countries. The EU is charac terized by regional inequalities both within and between countries, the overall scales of which have been increased with the end of Communism and the accession of new member states in Central and Eastern Europe (**Figure 3**). Similarly, the rapid industrialization and growth of the Chinese economy has been accompanied by a large in crease in regional inequalities (**Figure 4**).

Cultural Region

The idea of the region as a cultural entity was noted in the earliest statements of regional geography. The idea of a *genre de vie* was another way of describing the region as a cultural entity. Discussion of regional culture quickly be comes embroiled in debates about particularism. In Eng lish, the term 'Balkanization' has been used to describe the fragmentation of national territories into competing sub national territories: it is almost always used pejoratively and carries a note of condescension. In this context, for much of the twentieth century, regional culture was seen at something on the margins of processes of standard ization, modernization, and nation building. In Europe, from the 1960s onward, regions began to assert claims to their cultural identity, often based on language issues. In some cases these transformed from regionalist claims into nationalist ones. Among the prominent examples are Wales and Scotland, Brittany and Corsica, Catalunya and the Basque Country. In Europe, rapid proliferation of re gionalisms followed the demise of the Cold War. Typically, these combined closely linked cultural and political claims. From Transdnestr in the East to Galicia in the West, regional and semisecessionist claims became visible in many European countries. But regionalist claims also surfaced in Latin America, Africa, and Asia. The Indonesia archipelago is fractured by multiple complex claims for regional autonomy, based on economic, ethnic, and lan guage cleavages. Similarly, China is characterized by the complex interplay of ethnic and provincial loyalties, the political importance of which have ebbed and flowed in the history of Chinese civilization.

Cultural processes are central to the institutional ization of regions; that is, how regions emerge, are transformed, and disappear. The formation of regions and processes of regional transformation is connected to local, state, and international economic and political practices. Regional formation involves the constitution of a terri torial area, its symbolic representation, its coalesc ing into institutional forms and its embedding in the re gional system and social consciousness. The formation of the region is the outcome of strategies of actors inside and outside the region. The region is reproduced by the practices of actors in the administrative, economic, po litical, and cultural spheres. The cultural symbols cement an image of the region in the consciousness of its popu lation. Territorial symbols provide a focus for group solidarity. Such symbols compete against the symbols of national identity and the relative strength of these is critical to strength of the processes by which a region is

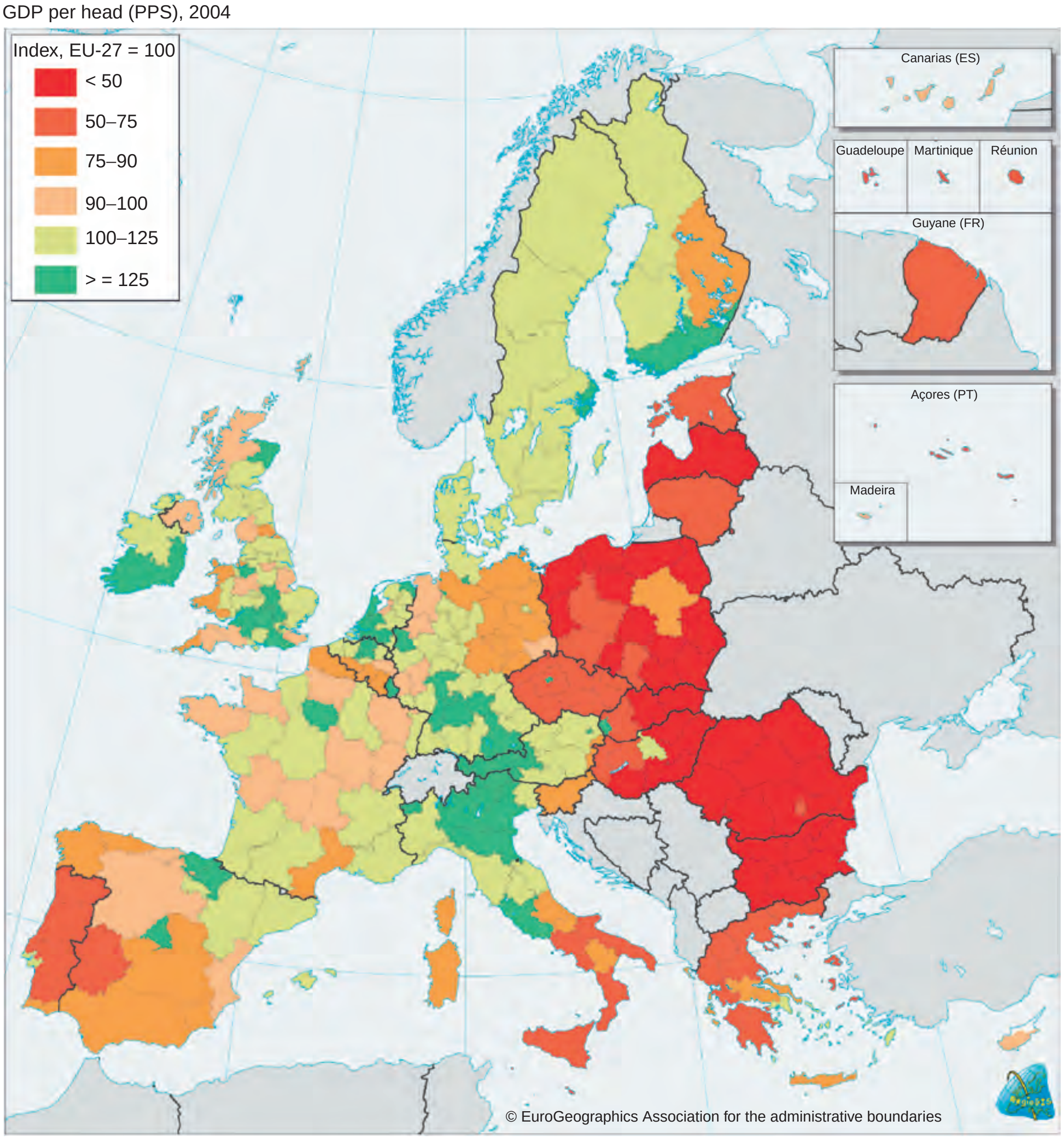

Figure 3 GDP per head (PPS) in Member States of the European Union and regional extremes, 2004. * In these regions, the GDP per head figure tends to be overestimated because of commuter flows. Source: Eurostat.

institutionalized, either in the form of formal administrative or other structures or distinctive practices that are understood within and outside the region. Once the region has become established, either administratively or discursively, it can become a focus for place marketing efforts or a rallying point for conflicts over resources and power (**Figure 5**). Regional actors produce practices and discourses which give meaning to bounded material and symbolic worlds in order to create intersubjective meanings. Regions must be identified and that identification must be shared. Memories shared by their constituents give regions (and societies in general) their (moral) particularity, while the revalorization of regional culture is central to the creation of a modern regional identity. This is evident in the assertion of Scottish and Welsh identity in the UK or Catalan identity in Spain, or Quebecois identity in Canada, which embody national claims or Scanian identity in Sweden or Silesian identity in Poland where, for instance, the revival of cultural distinction in art and literature, the valorization of dialect, and the emergence of regional media systems has been an important contributor to the production of symbols which provide cultural contexts for political action.

Political Region

The apparent growth and integration of the region in functional, economic, and cultural terms underpins its

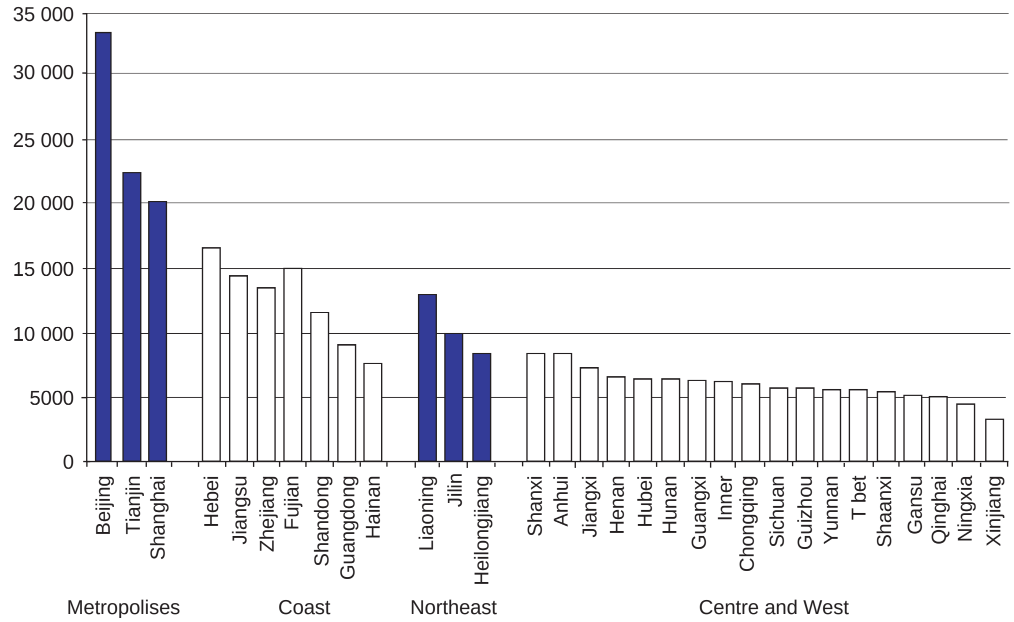

Figure 4 Per Capita GDP of Chinese Provinces in RMB in 2002. Source: National Bureau of Statistics of China.

Figure 5 Cultural identity and place marketing: Catalanism in New York. Source: http://www.llull.cat/llull/estatic/ny2007/eng/jornades/graella.shtm (Accessed: 10 June 2007).

emergence as a focus of political mobilization across the world in the contemporary era. The rise of the region as political arena in Europe has generated much comment and analysis. Through the 1970s and 1980s, across Western Europe, regions became increasingly assertive political actors. National governments showed a preparedness to devolve political power to regional governments. As new member states acceded to the EU they generally created new structures of regional government. The idea of a 'Europe of Regions' (in implicit contrast to a Europe of Nations) became widely discussed. But this was not merely a European phenomenon. Most Organization for Economic Cooperation and Development (OECD) states had created structures of regional government by the mid 1990s, to the extent that some observers have identified a 'new regionalism' (**Figure 6**).

New trends toward the decentralization of government are discernible in different political contexts, including Europe, Australia, Latin America, North America, and Africa. For instance, the transition to democracy in South Africa was accompanied by a radical devolution of power to newly constituted provisional governments (**Figure 7**). Although the structures of regional government vary significantly between societies, some writers have identified the emergence of a new regionalism linked to a newly attached importance for subnational government by international organizations, such as the OECD and World Bank. The trend toward decentralization of government is driven to a large extent by political factors at the national and subnational level. In countries such as Spain and Brazil, the creation or strengthening of structures of subnational governance occurred alongside the restoration of democracy following long periods of authoritarian government. In the UK, devolution to Scotland and Wales was a response to the rise of nationalist sentiments in a multinational state. Elsewhere, in Europe, 'historical regions' have asserted their cultural identities, while 'new', 'economic regions' have also claimed the right to act directly in their own political interest. In the United States, the rise of 'metropolitan regionalism' is a response to the rapid growth of suburbs, the creation of satellite cities, new modes of transportation, increased mobility of citizens, and easier forms communication, which require new structures between the city, county, and state. More generally, the trend to devolution may reflect a rejection of centralized forms of politics and government at the national level and declining faith in the nation state as an instrument for solving social and economic problems. But it may also be a means for national governments to devolve welfare responsibilities to subnational authorities in order to

	Population, millions, 1997	Number of regions	Population per region, millions	Notes
States with a full system of elected regional government				
USA	266.8	50	5.3	Fifty federated states
Japan	126.2	8/59	15.8/2.1	Japan has eight regions, sub-divided into 59 prefectures (47) and cities (12) with the 3 same status as prefectures
				Thirty-two federates states
Mexico	94.2	32	2.9	Sixteen lander
Germany	82.1	16	5.1	Eighty provinces
Turkey	63.7	80	0.8	Twenty-two regions in
France	58.6	22	2.7	metropolitan France (excludes overseas possessions)
				Twenty regions
				Eight provinces
Italy	56.9	20	2.8	Nine federated states
Korea	46.0	8	5.8	Seventeen autonomous
South Africa	43.4	9	4.8	communities
Spain	39.3	17	2.3	Eight provinces
				Thirteen federated states (10)
Poland	38.7	8	4.8	and territories (3)
Canada	30.3	13	2.3	Seven federated states
				Twelve province
Australia	18.5	7	2.6	Fifty-one prefectures
Netherlands	15.6	12	1.3	Eight regions
Greece	10.5	51	0.2	Flanders, Wallonia, and Brussels
Czech Rep	10.3	8	1.3	Forty counties and municipal
Belgium	10.2	3	3.4	counties, including Budapest
				Eight regions
Hungary	10.2	40	0.3	Nine bundeslaender
				Twenty-six cantons
				Fourteen countries
Sweden	8.8	8	1.1	Nineteen regions
Austria	8.1	9	0.9	
Switzerland	7.1	26	0.3	
Denmark	5.3	14	0.4	
Finland	5.1	19	0.3	
States with no, or limited, elected regional government				
Portugal	10.0			Principle of regional government rejected in a referendum
UK	59.0			Elected regional governments in London, Scotland, Wales and – possibly – Northern Ireland

Figure 6 Regional government in OECD countries* and in South Africa. * With populations exceeding five millions. Source: Author's research.

relieve fiscal burdens. Regional structures of governance, then, are a feature of an increasing number of states.

A number of factors explain the rise of the regional politics. First, the rise of subnation political identities is a powerful driver of demands for devolution and the framing of the region as a political space. Second, regionalists invoke the suggestion that regional political structures are closer to the citizen and therefore enhance democratic accountability. In Europe, this argument is underpinned by the doctrine of subsidiarity. Third, the idea that the complexity of modern states and the services they perform mean that central governments have become overloaded and ill equipped to deal with regional diversity: central governments are too big to undertake some tasks, while local government is too small. The regional level is often a level at which important public services such as health and further education is managed. Finally, and crucially, the case for the political regions rests on economic foundations, based on claims that regional authorities are more nimble economic actors in an age of intensified global trade and increasing trade specialization by region.

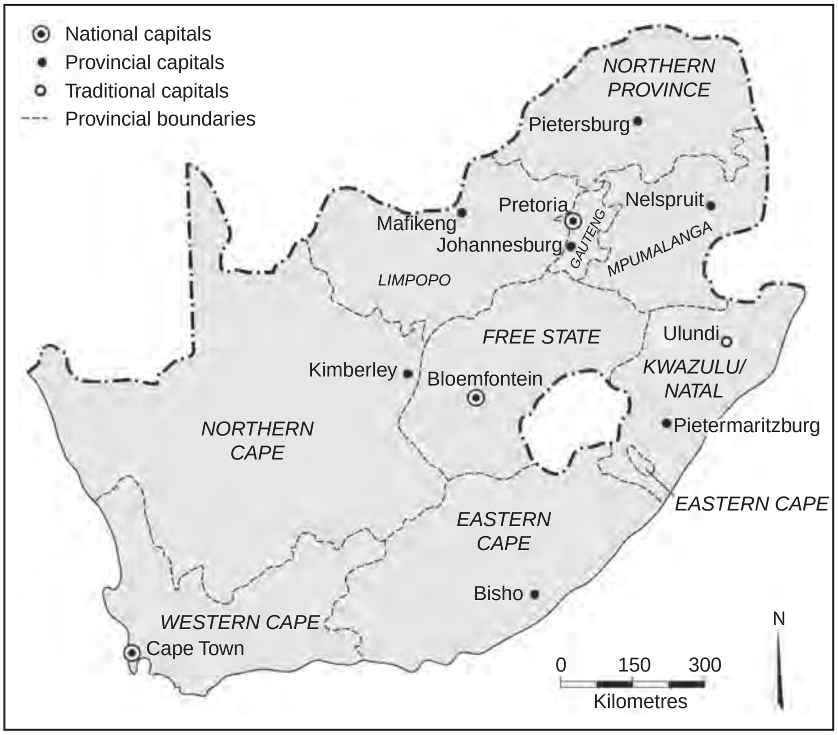

Figure 7 Provincial governments in South Africa.

The growth of regional government across the world reflects a widely held belief that the regional scale rep resents a particularly salient scale at which to organize policy interventions in the economy. Indeed, effective governance has been seen as a factor in underpinning the performance of successful regions by providing support for the formation of skills, technological change, and the nurturing of the natural and built environment and through the promotion of 'untraded interdependencies' that include labor markets, public institutions, rules of action, understandings, and values. It is important, how ever, not to assume an easy causal relationship between good governance and economic success. Structures of governance undoubtedly help to shape patterns of eco nomic development but effective forms of governance are often underpinned by strong economic performance. In some countries, strongly performing regions lead de mands for increased autonomy.

Regional institutions, therefore, are seen as a neces sary ingredient of bottom up forms of regional policy in an era in which localities and regions are more directly exposed to the international economy. This changing context makes regions agents of their own development. The role of the devolved state has shifted from one of local management of the welfare state aimed at ensuring redistribution to one based on 'entrepreneurialism', whereby regions directly compete with each other. Such entrepreneurialism needs to be examined at a variety of spatial scales, including the region.

The new regional politics tends to be underpinned by development coalitions which comprise a cross class, place based alliance of social and political actors dedi cated to economic growth in a specific location, the composition of such coalitions varying from one place to another. Such coalitions tend to be embodied in regional institutions which provide public goods and foster dialog and communication among economic actors in the re gion. Regional political leaders become important figures in shaping development coalitions. Political structures provide legitimacy for regional governance arrange ments, but even within Europe, for instance, there is great diversity in the forms of political structures and the powers and responsibilities of regional governments re flecting different national debates and histories.

Rescaling of the state in pursuit of regional gover nance can in some instances involve the transgression of national frontiers. A good example of this type of de velopment is the creation of the cross border Øresund region which links the region of Skåne in southern Sweden and Zealand in Denmark. Of symbolic and practical significance in the construction of this cross border region was the opening of a bridge, which crosses the Øresund strait, funded by the Danish and Swedish governments, in order to link Malmö and Copenhagen. The region is governed by the Øresund Committee – a gathering of political representatives from the local and regional authorities on both sides of the Øresund. The key institutions are the Copenhagen regional authority and the region of Skåne, which was created in 1999 as one of a number of 'experimental regions' within Sweden, which took over responsibilities from the central government in the field of economic development.

One danger of the debate about novel forms of regional governance, though, is that they detract attention from the enduring role of the nation state and national governments in the framing of public policy. It has been an assumption of some writing on the new regionalism that that the combined consequences of globalization and devolution have rendered the nation state redundant. Yet such an assumption is not borne out in reality. Despite processes of globalization, privatization, and declining trust in political institutions in some societies, there are strong reasons to believe that national governments retain an important role in shaping social and economic structures, including at the regional level. Typically, regions operate within systems of multilevel governance in concert with governments at other spatial scales. Moreover, without national (or increasingly supranational) forms of regulation there is a danger that the new regionalism may accelerate wasteful forms of territorial competition between regions seeking to promote their own development. Nevertheless, the strengthening of regional governance can add to the capacity of public authorities to make effective interventions to promote economic development.

Ecological Region

Despite the focus on the region as a medium and outcome of social processes, the physical environment continues to place constraints on the human populations that occupy it. The implications of this are very obvious in rural societies, where forms of agriculture are dependent in a more or less direct way on the physical geography of territory. Regions can be characterized in relation to geographically distinct assemblages of natural communities and species. Human activity occurs in relation to flows of matter and energy. Relief, climatic conditions, and water catchments affect the biodiversity of flora and fauna and continue to place constraints on patterns of development, even if, simultaneously, they are transformed by human activity. While there is no universally valid classification of the topographic frameworks within which human activity occurs, the notion of the ecoregion has been developed to indicate these relationships. Numerous efforts have been made to describe and characterize ecoregions. Terrestrial ecoregions are relatively large units of land containing distinct assemblages of natural communities and species, with boundaries that approximate the original extent of natural communities prior to major land use change. Below the level of ecoregion, water catchments – or drainage basins – often have delimited natural regions, with watersheds, for instance, providing natural boundaries. Although rivers such as the Mississippi mark borders between subnational territories, watersheds also frequently represent territorial boundaries. The border between Scotland and England is marked in places by a watershed. Such considerations figured prominently in the work of earlier regional geographers. In 1919, geographer C. B. Fawcett devised a system for the regionalization of England, in which watersheds played a critical role. There are signs that some of these sensibilities are returning to the fore.

The significance of drainage basins in the hydrological cycle means that they loom large as a focus for regional planning and resource management. They frequently cross administrative boundaries in ways which can create political tensions. In the early 2000s the prolonged drought in south east Australia led to political conflict over the use of water from Murray Darling Basin between the states of Victoria, New South Wales, and Queensland and the Federal government.

The Murray Darling Basin represents a natural region, albeit one which has been transformed by human intervention (**Figure 8**). Its management has been controversial since the federation of Australia in 1901, with different colonies and then states arguing over its agricultural and potable use. Aboriginal claims to the river and its management have added to the political complexity. By the 1990s, human intervention – through irrigation – was consuming 77% of its flows. While there was wide agreement that this situation was unsustainable, the states could not agree on a solution. By the early 2000s, the territorial politics of water in south eastern Australia remained intractable and controversial.

Rising ecological awareness has generated a new interest in ecoregions and the politics of bioregionalism. This raises the prospect of redrawing the boundaries of political regions to match those which demarcate ecological, or bioregions. There is evidence of a growing desire on the part of populations and public authorities in some countries to identify the unique ecology of particular bioregions. The growing interest in (post) industrial societies of the provenance of food and the desire to reduce food miles and to regionalize food chains in more sustainable ways is linked to debates about biodiversity and the desire to conserve native flora and fauna. Within architecture the renewed preference for using local materials in the production of built environment speaks of the same ecological concerns. The notion of natural regions has been regarded as somewhat anachronistic in the new regional geography, which has tended to focus on the social construction of regions. But the effects of climate change and increasing pressures on natural resources have redrawn attention to the ecological foundations of the region. Among other things,

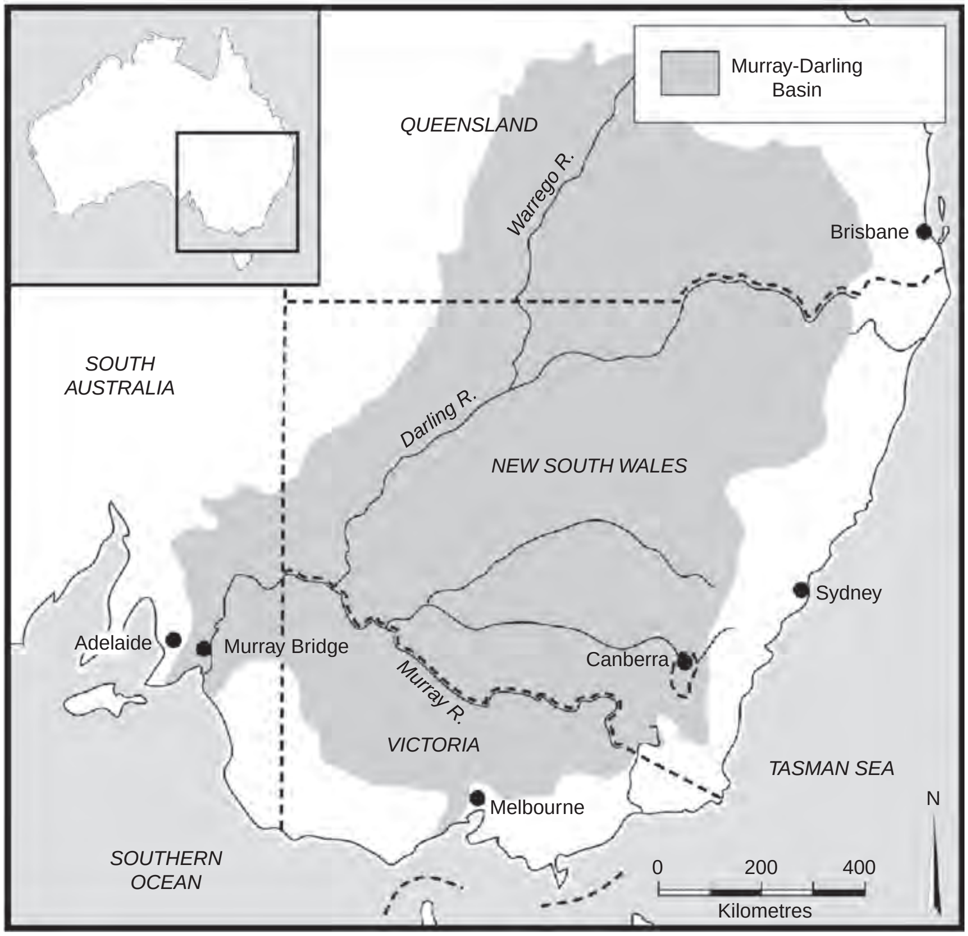

Figure 8 The Murray-Darling Basin and political boundaries.

this suggests that traditional regional geography may yet contain insights that are exceptionally useful in an era of heightened environmental concerns.

Region Redux?

Arguments about the region in human geography have been central to the development of the discipline. But the debate in the international literature remains dominated by European and North American voices to date, despite the importance of the region as a cultural, political, and ecological entity in every continent of the globe. Indeed, the most fertile ground for study of the region is likely to be provided by China or Indonesia or Malaysia: countries which are experiencing rapid growth in regional in equalities and the assertion of subnational territorial identities.

Some commentators link the (re)birth of the region to the impacts of globalization. On the one hand, it is ar gued, globalization reshapes material and discursive flows and undermines the political position of the nation state. The region is an important cultural, political, and ecological category in countries across the world. How ever, there is great diversity in the formation and trans formation of regions. There is no unfolding logic of regionalization. The new regional geography, broadly defined, is recognition and response – and perhaps stimulus – to developments outside the academy.

On the other hand, according to some writers, the intensification of global flows is undermining the notion of territorial political identities. In these accounts, re gions are understood as open and discontinuous spaces, constituted by social relationships and material, dis cursive, and symbolic forms stretched across and through regions. It adopts a topological rather than topographical approach to space and ultimately rejects the notion that regions are territorially fixed, bounded, or closed. 'Re gions' are defined by their place in networks and by their relationship within and beyond predefined territorial boundaries. Regions are 'unbound'. Ideas of stable and clearly demarcated hierarchies of scale are eschewed in this perspective. Indeed, the notion of the region is re placed by the idea of a 'regional assemblage' or, a mul tiactor, multiscalar topological geometry. This relational view usefully draws attention to the way in which (post)modern life is embedded in socially constituted and spatially stretched networks. But, in this perspective, there is a danger in overlooking the persistence of ter ritoriality and the setting of boundaries and scales as expressions of social practice, discourse, and power. Moreover, as noted previously, the return to the region reflects very real concerns about environmental sustain ability; that is, with the life giving qualities of land and water. These questions, although differing in detail, are as important in principle – and increasingly in practice – to porous and unbound global city regions as they are to rural regions of Central America. For these reasons, and for all the others described here, human geographers will continue to study the region.

See also: Culture; Governance; Identity Politics; Regional Geography I; Regional Geography II; Regionalism.

Further Reading

Claval, P. (1993). *An Introduction to Regional Geography*. Oxford: Blackwell.
Fleure, H. J. (1919). Human regions. *Scottish Geographical Magazine* 35, 94 105.
Dickinson, R. E. (1947). *City Region and Regionalism: A Geographical Contribution to Human Ecology*. London: Kegan Paul, Trench, Trubner & Co.
Gilbert, A. (1988). The new regional geography in English and French speaking countries. *Progress in Human Geography* 12(2), 209 228.
Hartshorne, R. (1939). *The Nature of Geography*. Lancaster, PA: Association of American Geographers.
Herbertson, A. J. (1905). The major natural regions: An essay in systematic geography. *The Geographical Journal* 25(3), 300 310.
Kimble, G. (1951). The inadequacy of the regional concept. In Stamp, L. D. & Wooldridge, S. W. (eds.) *London Essays in Geography*, pp 151 174. London: London School of Economics/Longmans, Green & Co.
Markusen, A. (1987). *Regions: The Economics and Politics of Territory*. Totowa, NJ: Rowman & Littlefield.
Massey, D. (1978). In what sense a regional problem? *Regional Studies* 13(2), 233 243.
Paasi, A. (1991). Deconstructing regions: Notes on the scales of spatial life. *Environment and Planning A* 23(2), 235 256.
Tomaney, J. (2008). Regional culture. In Higgins, M., Smith, C. & Storey, J. (eds.) *Cambridge Companion to Modern British Culture*. Cambridge: Cambridge University Press.
Vidal de la Blache, P. (1928). *The Personality of France* (Bretnall, H. C. (trans.). London: Christophers.
Williams, R. (1988). *Keywords: A Vocabulary of Culture and Society,* (revised edn.). London: Fontana.

Relevant Websites

http://www.cor.europa.eu/
Committee of the Regions.
http://ec.europa.eu/dgs/regional policy/index en.htm
European Commission Directorate General for Regional Policy.

Regional Actors

M. Tewdwr-Jones, University College London, London, UK
A. Lord, University of Liverpool, Liverpool, UK

Glossary

Federal System of Government A political system where certain aspects of state power are geographically diffused to regional legislatures.

Organizational Learning Collective learning that becomes embedded within an organization *en masse*, informing its ongoing operation as opposed to being limited to a small number of core personnel.

Parochial Capital An expression used by original growth machine theorists to describe businesses tied to a particular location due to vested, geographically specific, interests which may accordingly seek political representation to influence policies in that area.

Regional Development Agency A state or quasi-state institution designed to encourage and coordinate inward investment in a region.

Transnational Corporation (TNC) A commercial enterprise with operations in multiple national jurisdictions.

Travel-to-Work Area (TTWA) A functional space defined by the relationship between a central employment site and the labor market of its hinterlands.

Unitary System of Government A political system where state power is predominantly centralized in one geographic location.

Untraded Dependencies The economic gains associated with reciprocity between firms occupying a common regional location for example, the sharing of information or equipment.

Introduction

In most Western societies, regions represent an important site within which the activities of various policy elites are conducted. As an intermediate territorial unit, occupying a position somewhere between the national and the local, regions provide a logical platform for both the delivery of spatially targeted policies emerging from above, such as the national and supranational scales, and a forum within which local actors may seek to advance their agendas and ultimately influence local fortunes. As a result, the range of agencies that might be thought of as being regionally active is very wide and, taken together, constitutes a complex system of policy formulation, delivery, and reformulation. Despite the ongoing controversy regarding how regional spaces are defined, it is this characteristic of regions as a territorial meeting point for disparate agencies operating at spatial scales above and below which continues to ensure that regions are a key site of investigation for many human geographers.

In considering the scope of what might be meant by the term ‘regional actors’, it is first necessary to establish the political standing of the region itself. What is actually meant by a region differs from case to case and, in administrative terms, is functionally dependent upon the prevailing political system. For instance, in federal contexts, such as Germany, the region is a formally recognized component of the state hierarchy through the Länder system of regional legislatures. Elsewhere, as in unitary states such as the United Kingdom, regions are less clearly defined and have not historically been a spatial scale around which formal governing activities have been organized. These systemic differences in the extent and exercise of political power anchored at the regional scale inevitably color how regional actors are conceptualized and serve to expose important contextual differences from nation to nation. Nevertheless, particularly since the late 1990s, even in states where a formal tier of regional governance is not in existence, bespoke institutional forms designed to operate at this scale have become increasingly evident. It is around such a regionally orientated state apparatus that policy elites and actors may seek to organize themselves and, in jurisdictions that do not have formal regional government, choreograph aspects of the governance arrangements that prevail in nations that do.

Moreover, irrespective of what regional state apparatus can be found to exist and whatever its national standing may be, other interests may, regardless, seek to organize themselves at the regional level. This potentially includes groups defined by regional identity, such as the many separatist movements active around the globe, as well as trade and labor organizations and interest groups with a particular regional concentration. In short, regardless of national political context, the existence of regionally significant actors is evident across much of the Western world and so continues to represent a fundamental area of enquiry for observers and commentators.

Regional Actors in Context

Academic interest in the activities of actors at a regional scale stretches back at least as far as the advent of regional science as a discipline at the confluence of economics and human geography during the 1950s. Emerging from the view that the region, expansively defined to include functional as well as administrative subnational spaces, represented a logical site at which to situate spatial ana lyses, the first regional scientists were mainly preoccupied with economic phenomena. Fundamentally, this academic tradition was spawned to a large extent by the work of the US economist Walter Isard and focused initially on pro viding quantitative frameworks within which to analyze the locational choices of one of the most significant re gional actors, firms, as a function of the factors of pro duction (land, labor, and capital). Over the intervening half century, while the methodological outlook of regional science has remained in the positivist tradition, the ana lytical techniques have grown ever more sophisticated and the suite of regional actors that have come under its microscope has broadened to include not just traditional areas of interest such as labor and housing markets or the influence of infrastructure providers (for instance, trans port) but also the social and economic impact of region ally significant institutions such as major employers in both the public and private sectors.

Very much part of the post war quantitative turn in human geography, the subdiscipline of regional science subsequently waned in tandem with the decline of posi tivism and the ascent of critical and Marxist human geography in the 1960s and 1970s. During this time, however, while the methodological approach changed, regions and the activities of policy elites within them re mained a keen focus for academic activity. This trend has continued with analyses of action taken by indige nous peoples to protect culturally significant regional environments as well as insurgent politics based on re gional identity: for example, the activities of separatist movements such as in the Basque area of northern Spain/ southern France. Elsewhere in the mainstream of Marxist and neo Marxist human geography, others have focused on regions as a key strategic site in the spatial unraveling of the capitalist system. On this reading, regions are con ceptualized as a geographical container of contrasting forces, a meeting point at which centrally disseminated policy abuts the self organization of subnational interests. In the United Kingdom, key examples of this are found in the human geography of regional flash points such as the National Union of Mineworkers strikes in 1984–85 which involved a catalog of regionally significant actors – re gionally mobile strikers, or ‘flying pickets’; the police and judicial system; and national trades’ unions with particular regional strongholds.

The New Regionalism

Ultimately, while both positivist regional science and qualitative critical geography remain strong traditions in the broader discipline of human geography and have performed valuable work in narrating and analyzing the work of regional actors, the most recent mutation in academic thinking in this area has come through the emergence in the 1990s of the new regionalism move ment. Sharing a common worldview with much of the regional science literature, the new regionalism also draws more overtly on recent developments in political and administrative science and economics, including the new institutional economics. Building on work that draws these academic strands together, the new regionalism seeks to give a spatial account of work such as that of, for instance, Etzkowitz who has identified a ‘triple helix’ of parties said to underpin successful economies, namely, the state, business, and universities. For their part, the new regionalists have sought to identify the interaction of such agencies at a regional scale and, in so doing, es tablish the role played by these actors in both successful and comparatively less successful regional economies.

What should motivate the identification of the region as the most appropriate scale at which to view the ac tivities of these actors, however, is complicated. As was established earlier in many national contexts, statutory functions are not discharged at the regional scale. Fur thermore, most universities have at least a national, and in many instances an international, reach as do many of the most significant employers: frequently transnational corporations (TNCs) with globally diffuse shareholders. At one level, it is precisely this implied phenomenon of globalization that is cited by the new regionalists as the principal force pushing regions into sharper focus as the most appropriate scale of analysis for human geog raphers. Globalization, it is argued, erodes the capacity of central states to formulate a coherent and uniform policy for national territories that encompass disparately per forming regions. On this account, globalization has led to the fragmentation in the nation state and an accom panying settlement in which regional policy, and cor respondingly the actions of regional actors, are seen as increasingly important. This extends, for example, to the manner in which aspects of the regional environment can be shaped by and through the regulatory framework structured, to some extent, by the prevailing institutional configuration – economic development agencies, regional planning bodies, transport infrastructure, universities, and so on. From this perspective, the operation of the regional economy is determined by the relative efficacy of these regional actors in creating the conditions for regional development by, for example, engineering characteristics of the regional labor market and incu bating untraded dependencies between firms.

From this starting point, the new regionalists have identified exemplars where strong and well integrated regional actors, it is argued, underpin stellar economic performance. Such learning regions are said to include the Southeast of England, the Greater Boston region in Massachusetts, Baden Württemberg in Germany, and the Greater San Francisco–Berkeley region (popularly re ferred to as Silicon Valley) in northern California. In all such cases, it is argued that an economic performance well above the national average has been predicated upon the presence and interaction of significant regional actors – typically one or more globally important higher edu cation institution, innovative businesses, and an engaged regional state apparatus. A corpus of literature advocating regional development strategies designed to replicate the kinds of success enjoyed in such regions by copying their institutional arrangements has burgeoned since the 1990s, frequently with a focus on one of this triumvirate of general areas: for example, Porter's oft cited work on the regional clustering of firms or Florida's research on the development of the higher education sector to support the creation of a learning region. For some critical aca demics, for example, Lovering, the identification of such exemplars has not been established as anything more than temporally contingent – the redescription of a set of actor networks which happen to exist in those regions which are most keenly attuned to the current high growth sectors of the global economy. It has also been suggested by the same author that academic research in the new regionalist mold and policy advocacy have grown so close as to be practically indistinguishable, underpinning the proliferation of many new regional agencies in varying national contexts.

Understanding New Regional Actors – The UK, the EU, and the US

The rationale for this view can be found in the creation, particularly since the later 1990s, of policy vehicles situated at the regional scale, the justification for which has frequently drawn upon new regionalist thinking. In the United Kingdom, this has included devolution to the Celtic nations through the establishment of directly elected national assemblies in 1999, partly on the grounds of separate national identity and also as a mechanism to update and subsequently assimilate a long established institutional architecture, including, for example, the Welsh Development Agency, launched in 1976, and Scottish Enterprise which grew out of the Scottish De velopment Agency, active since 1975. In England, a more straightforward new regionalist account of the require ment for the creation of a regional bureaucracy was used shortly after the election victory of the first Blair gov ernment, to establish proposals for nine regional as semblies in 1998 and the same number of accompanying regional development agencies (RDAs), eight of which came into effect in April 1999, the ninth, the London Development Agency, delayed until April 2000 to coincide with a suite of other new institutions, including a mayor, for the capital. These RDAs augmented existing regional arrangements which previously centered upon the nine government offices for the regions created by the Conservative government of John Major in 1994 to co ordinate the activities of the plethora of subregional bodies which had sprung up over previous decades. The most recent attempt to add to the regional platform of governance has come through an abortive attempt to transform the appointed regional assemblies into directly elected chambers. The well documented rejection on 4 November 2004 of the proposal by referendum in the Northeast, then thought to have the strongest regional identity within England, has ensured that the policy is unlikely to be resurrected.

Elsewhere, similar processes of institution building at the regional scale have taken place in other European nations: to either reinforce in federal or create in unitary contexts a regional state or quasi state platform on which policy actors may seek to gather. This has been witnessed as taking place more or less simultaneously, both in older West European states including, among others, Austria, Belgium, Denmark, France, Germany, Italy, Portugal, Spain, and the Netherlands and in the accessor states in the former communist East, such as Hungary and Poland. This apparent trend toward establishing significant re gional actors across the European territory has sparked a great deal of interest both in cataloging and in comparing the varying types of agency, their roles, scope, and remit. Academic research in this area suggests that there are significant contextual differences between RDAs which are not only a function of the prevailing political system but also the sociology of the existing institutional setting. In short, among other factors, the effectiveness of new agencies at the regional scale is to some extent dependent on how their activities are integrated with those of other regional actors. Such issues may relate to questions of legitimacy and status; official and unofficial recognition; active or benign agencies; personnel, institutional culture, and organizational learning; and the degree to which agencies are permanent or *ad hoc*. Looking across these categories, some commentators have identified individual RDAs as particularly successful in supporting strong re gional economic performance. A notable example singled out for recognition in this respect has been ERVET (*Ente Regionale per la Valorizzazione Economica del Territorio*), the RDA for the affluent Italian region of Emilia Romagna. Otherwise, attention has focused primarily on identifying in similarly successful regions any contribution made to that prosperity by their institutional apparatus; examples here may include the network of RDAs operating in Baden Württemburg and the Paris and Ile de France Regional Development Agency (PRDA).

In all such cases, another factor shaping the emer gence of regional policy has been the influence of the

European Union – itself a significant regional actor. Important developments in this respect have included the creation in March 1994 (although subsequently disbanded in the late 1990s) of the European Committee for the Regions and ongoing funding disseminated through the four objective strands of the European Regional Development Funds (ERDF). As might be anticipated, the process whereby individual regions, sometimes enjoying formal statutory recognition within their nation and sometimes not, have engaged with the EU has been complex and multifaceted. At one level, the relationship between the EU and its constituent regions could be thought of in purely programmatic terms, similar to that which exists between the nation and its regions, whereby projects are delivered through spatially targeted regional development aid (such as ERDF). Beyond this, however, is a more political dimension, whereby the goal specified by the then EU Commission President, Jack Delors, in 1985 of a 'Europe of the Regions' and significantly not of the nations, is understood to be part of a wider process of eroding national identities and central state power. Instrumental in this has been the European Spatial Development Perspective (ESDP), active since 1988, with its commitment to territorial cohesion based on functional regional geographies to support European integration. One example of this can potentially be found in the articulation by the ESDP of a number of cross border regions defined by reference to functional actor based concepts primarily related to labor markets, which in some instances are named to reflect historical, prenational geographies, such as the New Hanseatic Interregio and the Baltyk Eurregio. In the other direction, the complex relationship between the regions and the EU can be found to exist where regional actors may seek to establish a connection with the EU and, in so doing, bypass the national state. An example of this has been said to exist in the United Kingdom since the introduction in 2000 of West Midlands in Europe, essentially a Brussels based office to support a direct dialog between the English region and the EU.

Finally, particularly in federal states, local interests perhaps from the city or city regional scale may well seek to congregate around the regional tier of government where local fortunes may be most immediately influenced. The most obvious example of this is the United States where the growth machine thesis and urban regime theory have sought to conceptualize the manner in which business interests attempt to pursue their agenda through a variety of subnational political channels. This approach seems to have most purchase in the United States due to the well established links between public and private sectors and the federal system of government. This said, a methodological approach drawing upon the regime theory toolkit has been employed in many other contexts to tease out a link between indigenous business interests (or 'parochial capital' in the growth machine terminology) and other subnational state and quasi state agencies. This includes not solely individual business interests but also representative agencies, such as chambers of commerce, which are a common feature of most Western polities, frequently organized at local, national, and, in many instances, regional scales of governance. The resultant framework is one anchored at the regional scale (expansively construed) which attempts to catalog, describe, and analyze the interaction of multiple actors: businesses, both indigenous and the *avatars* of TNCs – a variety of state and quasi statutory agencies with a determinable regional presence and various other communities of interest.

Conclusion

Fundamentally, what previous and current interest in regionally significant agencies and elites shares is the notion that the region represents an important analytical category within and upon which the activities of various policy actors are brought to bear. From this viewpoint, the region is a coming together point, a key strategic site at which myriad interests arrange themselves, and which consequently makes it a logical scale at which to pitch analysis. It is largely for these reasons that the activities of regional actors have been the subject of such intense interest over a sustained period of time.

At base, however, the principal controversy continuing to plague activity in this area is an equally perennial question in human geography: what is actually meant in the term 'regional actors' by the expression 'regional'? As has been noted throughout, there is little consensus on how regions can be unambiguously understood. The new regionalism, for instance, has been applied without distinction to, for example, geographically small nations which constitute an element of a wider nation state (e.g., Wales) as well as territorially expansive metropolitan regions (e.g., Greater Los Angeles). Even in contexts where administrative regions exist which occupy a formally acknowledged and well defined role in a national system, such administrative territories do not necessarily correspond to functional regions as defined by actor related concepts such as travel to work areas (TTWAs) or housing markets. This confused and confusing landscape is further obfuscated by the fact that even if a region could at a moment in time be uncontentiously identified, many of the actors operating within it may not, in themselves, be specifically tied to that region but may instead reach up to it from below or down from above.

It is from this perspective that analyses such as those in keeping with the regime theory approach which attempt to provide an account of the influence of regional actors in shaping regional fortunes have been criticized on the grounds that they do not adequately take account of the influence of nonregional actors in this regard. Building on this critique, radical human geographers have insisted that an understanding of regions, and similarly cities and city regions, and the agencies active at these scales be situated within the wider political and economic system prevailing at national and supranational levels. That is, regions are understood to be a construct of the central state and a function of its mediation of free market capitalism.

Despite such criticisms, what remains is an intuition that, even without rigorous academic definition, in most contexts, there are sensibly distinguishable subnational plateaux at and around which various interests organize themselves. This may be made manifest through the politics of regional identity, central conceptual categories in human geography relating to labor and the location of firms, or the analysis of spatially targeted policies and state agency exerted from a variety of different spatial scales. As a result, the natural impulse to seek an understanding of the impact of such actors is unlikely to wane. Furthermore, if the new regionalists are correct in suggesting that economic globalization is creating the conditions under which greater regionalization takes place, interest in the activities of regional actors perhaps has the potential to grow.

See also: Learning Regions; New Regionalism; Regional Science; Urban Regimes.

Further Reading

Allen, J., Massey, D. and Cochrane, A. (1998). *Rethinking the Region*. London: Routledge.

Amin, A. and Thrift, N. (eds.) (1995). *Globalization, Institutions and Regional Development Agencies in Europe*. Oxford: Oxford University Press.

Etkowitz, H. (2003). Innovation in innovation: The triple helix of university industry government relations. *Social Science Information* 42, 293 337.

Florida, R. (1995). Toward the learning region. *Futures* 27, 527 536.

Goodwin, M., Jones, M. and Jones, R. (2005). Devolution, constitutional change and economic development: Explaining and understating the new institutional geographies of the British state. *Regional Studies* 39, 421 436.

Hadjimichalis, C. (1987). *Uneven Development and Regionalism: State, Territory and Class in Southern Europe*. London: Routledge.

Hudson, R. (2003). Global production systems and European integration: De regionalizing, re regionalizing and re scaling production systems in Europe. In Peck, J. & Yeung H. W. C. (eds.) *Remaking the Global Economy*, pp 216 230. London: Sage.

Isard, W. (1956). *Location and Space Economy*. New York: John Wiley.

Jones, M. and MacLeod, G. (2004). Regional spaces, spaces of regionalism: Territory, insurgent politics and the English question. *Transactions of the Institute of British Geographers* 29, 433 452.

Keating, M. (1998). *The New Regionalism in Western Europe*. Cheltenham: Edward Elgar.

Lovering, J. (2001). The coming regional crisis (and how to avoid it). *Regional Studies* 35, 349 354.

Nuttall, M. (2000). Indigenous peoples, self determination and the arctic environment. In Nuttall, M. & Callaghan, T. V. (eds.) *The Arctic: Environment, People and Policy*, pp 621 638. London: Taylor and Francis.

Porter, M. E. (1998). Clusters and the new economics of competition. *Harvard Business Review* Nov/Dec, 77 90.

Scott, A. J. (1998). *Regions and the World Economy*. Oxford: Oxford University Press.

Stone, C. N. (1989). *Regime Politics: Governing Atlanta, 1946 1988*. Lawrence, KS: University Press of Kansas.

Storper, M. (1997). *The Regional World: Territorial Development in a Global Economy*. New York: Guilford Press.

Tomaney, J. (2002). The evolution of regionalism in England. *Regional Studies* 36, 721 732.

Regional Competition, Regional Dumping

I. M. M. Pires, Universidade Nova de Lisboa, Lisbon, Portugal

Glossary

Clusters These are geographic concentrations of interconnected companies, specialized suppliers, service providers, firms in related industries, and associated institutions (e.g., universities, standards agencies, and trade associations) in a limited territory, developing synergies from holding together through cooperation and competition relationships. There are specialized clusters (electronics, textiles, etc.) or territorial clusters; some are classified as industrial districts.

Competition It is a process by which regions, like people or firms, are using limited resources, trying to win something, or be more successful than the others. Regional competition is primary understood as a 'fight' for development and wealth.

Competitiveness It is the result of a successful competition process at the end of which a state, a region, a firm, or an institution is in a better position than its opponents to attract people and investments that will create employment and wealth.

Dumping It is defined as the act of selling goods at lower than market prices, a practice that is prohibited by World Trade Organization (WTO).

Globalization It is defined as the growing economic interdependence of countries worldwide through increasing volume and variety of cross-border transactions in goods and services, free international capital flows, and more rapid and widespread diffusion of technology. Globalization has economic, political, cultural, and technological aspects that may be closely intertwined.

Industrial Districts These are socioeconomic entities characterized by the active presence of a population of firms and a social community, localized in close proximity in a specific geographic region.

Knowledge-Based Society and Economy A society in which information and ICTs are in general use, and the most valuable asset is investment in intangible, human, and social capital. There are two main perspectives of the knowledge economy: one that only considers technology and innovation and the second that defines it by the wide and more efficient use of codified and tacit knowledge by organizations and people in order to enhance growth and development. In both approaches, knowledge is regarded as a driver of growth.

Social Network It can be defined as a set of nodes or actors (persons or organizations) linked by social relationships or ties of a specified type.

Do Regions Compete?

Competitive environments stimulate innovation. Creativity and the permanent need to do things better or differently has become a key driver of economic prosperity. In order to gain competitive advantage, regions (i.e., their institutions) adopt a diversified range of strategies with a view to engage in a worldwide competition for large scale investments, research and develpoment (R&D) activities, the location decisions of private art collections, museums, and large international festivals, and as destinations of choice for tourists. While some regions with various features (in terms of history, stage of development, labor market, institutional structures, economic mix, etc.) have proven successful in facing up to this challenge and remaining competitive in a changing environment, others have not. As a result of these diametrically different outcomes, competition can bring about both positive and negative impacts.

The current debate on regional competition has to do with how places, whether regions or cities, are met with either success or failure when it comes to being able to take advantage of their local endowments (in terms of financial, human, natural, and social capital) in order to attract foreign investments and promote both endogenous and exogenous growth.

Despite the increasingly global character of competition and the fact that the reduction in transport and communications costs are turning the world into a relevant arena in which to compete, location continues to play a crucial role. In the context of globalization, regions constitute location references or the (physical) local anchors of (virtual) global processes. Yet, are regions able to engage in competitive processes in the same way as firms?

These issues are not new among geographers and economists, but the current debate on regional competition has emerged both from the idea of regions as economic arenas and from the transposition of the concept of competition from firms to regions. Alongside other new schools of thought, both critical geography (CR) and the new economic geography (NEG) have opened up new angles of analysis within the debate of regional competition, introducing in this discussion aspects such as the relevance of the agglomeration economies and of noneconomic factors such as net working, dense social networks, proximity, trust, co operation, learning, and local knowledge accumulation and diffusion. These are some of the key intangible fac tors of the knowledge based economy.

Up until recently, regions were regarded as mere spaces which physically supported human activity. Within geography, however, the contributions by the Californian School, the Italian School, and the GREMI research on *milieus innovateurs*, from the 1970s and 1980s onward, have put regions at the core of economic de velopment processes (based on flexible systems of pro duction) and began to regard them as the engines of their respective national economies and as the prime focus of economic policy. Coming from economics, Porter also played a relevant role in extending the concept of com petitiveness from firms to regions and countries, re garding regions as relevant sources of competitiveness and prosperity.

Regions compete by mobilizing a different mix of local resources, depending on the environmental context. Competition may be shaped by the available resources and prevalent conditions in the region, by how efficiently those resources and conditions are used and produced (productivity), and by innovation and creativity.

Classical and neoclassical economic theories explain success based on the relative abundance of the various regions' factor endowments (such as land, raw materials, capital, and labor), which confer an absolute and/or comparative advantage vis à vis the regions where those factors are less abundant or are less efficiently used (production efficiency), under conditions of perfect competition. However, these theories fail to explain the increasing spatial unevenness in terms of economic de velopment, as this can hardly be explained simply by differences in terms of local endowments. Besides, in a shrinking world, the aforementioned factors can be easily brought in from outside the region, at even better con ditions. The attempt to understand why some regions have succeeded while others have failed has led econo mists to rediscover their interest in spatial issues and to new avenues of research within economic geography. The emergence of a wide range of new growth models within new trade theory, new growth theory, and, in the 1990s, NEG has been a consequence of that debate.

Two of the key issues that NEG sought to address consisted of unraveling the interaction between growth and location, and understanding the cumulative process of agglomeration. Agglomeration economies help explain why industries cluster within particular regions due to the numerous sources of productivity gains that stem from geographic proximity: lower transaction costs; eas ier access to a large and diversified pool of labor, in cluding specialized, highly skilled labor, to cater to the specific needs of the firms; shared costs of common infrastructures and technologies; and the presence of business services and complementary support activities. Agglomerations make it possible for firms to form and maintain linkages with each other within various kinds of networks, through both formal interactions and 'untraded interdependencies' (which include labor markets, public institutions, habits, and understandings). Agglomeration also facilitates the exchange of information and the local accumulation and diffusion of knowledge, thereby giving rise to incremental innovations.

Italian type industrial districts, clusters, innovative regions, and learning regions all constitute examples of regions that have based their economic success on the aforementioned factors. These regions are regarded as complex entities, made up of interdependent economic and social units (such as firms, universities, public and private institutions, entrepreneurs) that share a homo geneous value system as well as common objectives and codes of behavior, in which firms engage in a specific type of competition (coopetition), meaning that although they compete with each other they also have to cooperate in order to pursue and achieve certain common objectives and interests. Local institutions play a fundamental role in leading the process of competition, and institutional density is a decisive factor. Alongside the technological, social, infrastructural, and institutional assets at the dis posal of the firms located in a given region, these are the attributes that turn regions into 'sticky places in a slip pery space'.

However, geographic agglomeration alone does not ensure the emergence of regional networks, as shown by a comparison of the examples of Silicon Valley and Route 128, which exhibited different abilities to respond to quickly changing environments. The success of the for mer was due to the prevalence of dense social networks, an open labor market, and the porous functional boundaries between firms. In contrast, Route 128 was characterized by the existence of largely autarkic cor porations and a regional culture that favored stability.

Proximity and agglomeration also have positive im pacts upon creativity. The knowledge based approach emphasizes the critical role played by knowledge, R&D activities, innovation, and networks in the strategies of the most competitive regions and cities. This approach is based on the acknowledgment that sustainable regional competitive advantage will arise whenever those local assets are valuable, rare, not subject to substitution, and imperfectly imitated, thus enabling firms to produce new goods and services and become more competitive than others.

Creativity has explicitly become a driving force of economic growth. As argued by Richard Florida, cre ativity and talent are the regions' main engines of growth, and those regions in which tolerance levels are high will be more efficient in attracting key assets with which to compete. Open environments not only stimulate entrepreneurship, but also foster greater adaptability to change than is the case in more conservative regions.

Therefore, a region's competitive advantage is not simply a consequence of the availability of resources; rather, it depends on how those resources are used and on the productivity with which labor and capital are employed. Higher productivity arises when firms or regions compete in more sophisticated ways, through innovative and unique products and processes. Porter argues that the only meaningful concept of competitiveness is productivity, which constitutes the ultimate determinant of a region's standard of living in the long run. The 'location paradox' referred to by this author consists of how geographic, cultural, and institutional proximity still remain enduring sources of competitive advantage in an era of global competition, whereas everything that can be efficiently brought in from afar has become virtually irrelevant. Thus, Porter regards competition as a dynamic process based on innovation and strategic differentiation, which is sustained through a highly localized process. He then goes on to discuss the crucial role played by clusters, rather than isolated firms, in competitive processes, explaining how location affects competitive advantage through a 'diamond' of four interrelated attributes – factor conditions; demand conditions; firm strategy, structure, and rivalry; and related and supporting industries. The effect of one attribute often depends on the others and the diamond operates as a self reinforcing system.

With the deepening of the globalization process and the emergence of the knowledge based economy, the regions' sources of competitive advantage have been redefined, with the emphasis shifting from 'hard' factors to 'soft' one such as innovation, entrepreneurship, knowledge, creativity, or culture. However, fulfilling but a few of these conditions is not enough – and meeting all of them requires public–private cooperation. This means that in order for a region to be at the cutting edge of competition, it needs to be able to mobilize a complex, interdependent, and diversified set of factors. Another implication is that it has become harder for those regions that lag behind or which do not meet any of the aforementioned conditions to catch up with the winners, thereby increasing the risk of divergence, as shown by several empirical analyses of regional competitiveness in the European Union (EU).

Criticisms of Regional Competition

The idea that regions compete in the same way as firms is far from consensual, and a large group of both geographers and economists have dismissed the use of the concept of firm competition at the regional scale of analysis.

Some of the problems with the concept of regional competition have to do with the limitations of the region itself, as new forms of production organization and cooperation between producers, suppliers, and customers (e.g., industrial *filières*, supply chains, and Global Value Chains) and new forms of corporate organization (networked enterprise) are not contained within administrative boundaries. The networked territory thus seems to provide a more adequate alternative to the existing sociopolitical entities as a framework for the networked enterprise. Moreover, transposing concepts such as social relations over social groups and other forms of social behavior to social relations over specific bounded places is also beset by numerous problems.

Criticism stressing the problems with regard to regions as coherent and homogeneous entities with the ability to act as collective players may be linked to the debate that has been taking place in geography since the 1980s on the issues of scalar construction, the relevance of scale, and the need to take multiple scales of analysis into account in order to understand the global economy. Up until recently, scale was taken for granted as a concept; geographic scales such as the local, regional, national, and global were regarded as unproblematic and fixed. However, in the context of globalization, the boundaries between these scales are not easily identifiable. Thus, instead of drawing on strictly defined geographic scales (such as the region in its more traditional understanding), recent approaches regard scale as a social construction and seek to foster a paradigm shift from fixed to relational scales, by establishing and differentiating multiple spatial units in order to unravel the complexity and interdependence of social and economic processes.

In turn, other critics argue that, at the meso level, the concept is poorly defined and that there are some major differences between regions and firms: for instance, regions do not go out of business like firms do when they adopt unsuccessful competitive strategies and regions are able to benefit from public support in order to soften the impact of economic decline. Nevertheless, if a region is unable to attract and hold on to investments and creative workers and keeps losing them to other regions, a situation of 'bankruptcy or insolvency' akin to the ones experienced by firms may arise in the long run, as it will be unable to compete and prosper in the future without those critical assets.

The main criticisms point out the difficulty of relating the competitiveness of a firm located in a given region to the overall competitiveness of that region, since the two often diverge. Besides, it is usually not clear whether or not the region is a key element in the process, that is, whether the region also competes or only firms within the region do. Therefore, Porter's idea of extending and applying his model of competitive advantage from firms to regions seems problematic, as do the ideas by all the other authors who assume that whatever applies to firms equally applies to these other entities called regions.

Other sources of criticism include the idea that em phasizing the role of the region in enhancing firm com petitiveness makes it more difficult to assess the role of the firms in generating regional competitiveness. The existence of competitive firms does not necessarily imply an increasing standard of living or benefits to the region, for instance, if the industrial fabric is based on low value added products and unskilled, low paid labor. That is the case whenever regions are only capable of attracting the low value added segments of the value chain. Com petitiveness at the micro level does not necessarily en hance competitiveness at the macro level. Besides, in a borderless world, corporations, such as multinational enterprises (MNEs), will often tend to organize their production chains at the global level in ways that ignore solidarities based on proximity and the interests of the region in which they locate. Thus, the role played by the regions may be undermined by the strategies and actions of multinational companies as global players.

Because of the aforementioned criticisms and the general lack of adequate empirical studies providing in sights both on how regions as collective entities engage in competitive strategies and on the roles played by the firms and the soft factors in the competition process, the concept of regional competition has been dismissed by some authors as a fuzzy, meaningless, and dangerous metaphor that should be avoided.

Even though issues such as whether or not regions compete and whether or not they constitute the most adequate scale of analysis for understanding complex and interdependent phenomena like competition remain pertinent, there is widespread consensus on the fact that, at the very least, the characteristics of the regions and their level of 'institutional density' influence the per formance of the firms located in those regions and con tribute to improving the conditions under which local firms compete.

Regional 'Dumping'

Dumping may refer to any kind of predatory pricing that will result in price discrimination between the domestic and foreign markets.

The term 'dumping' usually has a negative conno tation, although there is no consensus as regards its im pact upon the economy. Some, especially those advocating free trade, argue that the harm to the local producers in the country receiving the imported, dumped products may be offset by the benefits to its consumers that result from lower prices.

For some firms, competition is turning into an ag gressive way of ensuring survival and ceaselessly acquiring market hegemony. The prevailing idea is that in order to survive in such competitive environments, economic and social agents become involved in an eco nomic rivalry, whereby only 'the fittest' come out on top as there is not enough demand to enable the survival of all. When competition is based on costs, a situation of predatory pricing by foreign firms requires the domestic firms to take one of two possible options, either close or relocate the firm or bring down the cost of production, namely, by cutting wages or reducing labor privileges. Both have negative consequences for the standard of living in the respective region.

To prevent this competition from surpassing certain limits, regulatory frameworks have been developed at both the global (e.g., World Trade Organization (WTO) regulation on trade on goods and services) and macro regional (e.g., EU Competition Policy) levels. Nevertheless, competition based unfair or less than eth ical strategies, such as squeezing costs by putting pressure on wages and natural resources or using government subsidies to lower production costs, and trade barriers to avoid competition still persist in both developed and developing countries, although they are perhaps most frequent in the case of the emerging and less developed economies that are most eager for international investments.

Dumping is an issue that is commonly addressed by economic trade theories, although it is less frequently used when applied to regions. However, competition based on 'race to the bottom' strategies that compromise ethical, labor, and environmental standards in order to gain advantage over other locations may be considered as a sort of regional dumping.

The unprecedented growth of foreign direct invest ment (FDI) in the last few decades and the fierce com petition between countries and regions to capture a share of those investments have led governments and regions around the world to set up specialized agencies, such as investment promotion agencies (IPAs) with a national, regional, or metropolitan scope, with the aim of pro moting the country or region and attracting new investments.

At the regional level, the various regions' degree of success in attracting large scale investments is ultimately determined by the regional institutions' capacity to engage in dialog with MNEs, their degree of autonomy vis à vis the central state and their ability to adequately identify strategic sectors and implement proactive policies.

Nevertheless, as competition increases and the strat egies to attract FDI become more aggressive, instances of regional dumping may emerge.

Regional policies or regional support to industry ac tions aimed primarily at protecting mature and low value added industries from international competition or at conquering high market shares through wage reductions, the use of immigrant labor, or pressures upon the levels of social and environment protection constitute examples of what may be considered as regional dump ing. One must regard these strategies with criticism. On the one hand, because they delay the necessary mod ernization of the regional industrial base and on the other hand, because they give rise to a growth pattern akin to Kaplinsky's 'immiserising growth', that is, a situation in which there is increasing economic activity but failing economic returns as a consequence of incapacity by the firms, groups of firms, and national economies to ad equately penetrate the global markets.

Another source of concern is the possibility of local authorities relaxing their environmental protection standards due to the economic pressure to attract FDI. Although there is no empirical evidence to support either the generalized 'race to the bottom' or pollution haven hypotheses, the actual results are mixed and depend on the characteristics of the industrial fabric and the level of development of each country or region. Such phenomena are more frequently found in the case of particular in dustries, typically resource or pollution intensive ones (such as resource extraction, chemicals, chlorine, and pesticides), for which environmental costs can be quite high and which are heavily regulated in developed countries. Also in the case of regions specialized in low quality products which find themselves trapped between better quality competitors in Europe and lower cost producers in Asia, governments and local authorities have showed reluctance to increase the level of this type of regulation, partly out of fear that other destinations will become more attractive and that they will lose their competitive edge.

The fierce interregional competition to attract scarce mobile investments is also becoming increasingly per verse in developed countries, leading to the adoption of low road strategies. Industries under pressure from glo bal competition, for which the cost of production is still of paramount importance despite the increasing rele vance of innovation, give rise to situations in which re gions, and sometimes even different facilities of the same corporation, compete against each other in offering the best conditions to attract new investments. The recent delocalization of a general motors (GM) facility from Azambuja (near Lisbon) to Saragoza, in Spain (after un successful negotiations with the government aimed at benefiting from additional incentives, based on the threat to relocate), and the direct competition among several Volkswagen (VW) facilities in Europe aimed at winning the right to produce a new model constitute paradigmatic examples. This upsurge of 'race to the bottom' com petition strategies is due to the fact that the scarcity of large investment projects and the increasing number of competitors improves the mobile investments' ability to ensure the best possible national and local incentives. It also compromises the acquired labor privileges in the EU, as the only alternative to the threat of relocation typically consists of substantial concessions in terms of employ ment protection and significant wage reductions. The recent academic debate on the impact of EU's eastern enlargement, which has increased the competitive pres sure upon the 'old' member states insofar as the 'new' ones have sought to attract FDI based on low labor costs, labor market deregulation, and other incentives con sidered as unfair in EU context, provides yet another relevant angle of analysis of these issues.

Competition can give rise both to positive impacts, as a consequence of the stimulus to improve efficiency and innovate, and to negative impacts such as economic and social costs, whenever regions afflicted by competitive disadvantage, vulnerability, and weak local authorities engage in competition strategies based on low road strategies and those strategies are replicated elsewhere in a vicious circle.

As correctly reminded by the Group of Lisbon, there are limits to competition, and too much of it can bring about negative consequences.

Policy Debate

Why some regions grow faster than others and show an efficient capacity to remain at the cutting edge, while others do not, remains an unsolved key question. How ever, regional competition constitutes an important target for governmental, regional, and local policies.

Even though high levels of wealth have in some cases been achieved, this has often occurred at the expense of some regions lagging behind, while other regions con stantly improve their ability to compete in the national and global arenas. The EU's ambition to become the most dynamic and competitive knowledge driven economy in the world (also known as the Lisbon Agenda) is being pursued against the background of a process of increasing regional divergence, in which a competitive group of re gions (including the most dynamic urban areas of some of the new member states) is in stark contrast to a large group of regions that seem unable to successfully adapt to the new paradigm. By intensifying territorial competition, the deepening of the process of European integration not only has had a positive impact upon the EU economy as a whole but also seems to have contributed to the accu mulation of wealth in the richest regions. Regional income disparities have widened as economic concentration in Europe's 'pentagon' increased.

The policy debate in the EU has been influenced by the wider debate on regional competition and on how to help regions become more competitive. The debate on policy actions to improve regional competitiveness tends to be polarized between those who regard them as wasteful, ineffective, expensive, and noninclusive (because they only benefit a few, if any, since they may even have negative impacts in the long run) and those who believe that they have a positive impact upon economic welfare.

The process of European integration has been giving rise to increasing territorial competition, here interpreted in a broad sense as the promotion of local economic development through competition with other territories in order to improve the regional standard of living. Policy actions are motivated by the desire to improve competitiveness, and policymakers believe that at the EU, national and local levels efforts should be made to help regions become more competitive. The European Regional Development Fund (ERDF), created in 1975, is a key instrument of the regional policy of the EU which allocates resources with the aim of helping lagging regions to catch up so as to strengthen the economic, social, and territorial cohesion of the EU.

On the other hand, competition policy sets limits to the national governments' ability to provide financial incentives to private investments. Most of the development policies of the EU that seek to enhance growth and competitiveness are designed within the framework of its Regional Policy and Structural Funds, exhibit an increasingly selective character, and their efficiency is subject to evaluation. They are typically used to finance a variety of projects and interventions, from infrastructures (such as transport, telecommunications, energy, and environment) to R&D activities and the qualification of human resources. However, there are many ways in which local and national agencies can circumvent these limitations, for instance, when negotiating relevant investments, orchestrating marketing programs, and allocating huge amounts of capital, skills, creativity, and talent to developing a positive image with which to compete (regional branding). More often than not, this sort of strategy is pursued by those regions that are already at the top of the competition league, as those that lag behind and are less developed are typically less able to invest in them.

Often, regions engage in promotion programs without having previously evaluated their efficiency, whether by copying or following the strategies adopted by other regions or by opting for 'traditional' programs based on traditional location factors such as capital, land, infrastructure, labor, or subsidizing lower production costs. These policies are often inefficient in the long run due to the ubiquity of those factors, the high mobility of capital, and the lack of local embeddedness in the local institutional framework on the part of those investments.

Regions end up losing more than they gain and, in some cases, the subsidies provided by the local and national governments actually exceed the expected benefits.

Assessing the Regions' Competitive Performance

Because of the increasing attention paid to the debate on regional competition, it has become fashionable to compile and create indices that measure and compare the performance and attractiveness of the various regions. Despite the methodological insufficiencies and the inconsistency and lack of precision of some of these indices, comparing the competitiveness and performance of the various countries and regions has become commonplace.

A wide range of indices currently exist that seek to monitor the competitive performance of regions, cities, and nations by drawing on national and regional databases and the use of appropriate indicators.

Up until recently, most of those indices focused on competitiveness among countries. Since 1979, the World Economic Forum has published its global competitiveness report (GCR), whose global competitiveness index (GCI) is composed of three indices (technology, public institutions, and macroeconomic environment) computed for a sample of 117 advanced and developing countries, using a combination of hard data and data drawn from the World Economic Forum's Executive Opinion Survey. In turn, the World Competitiveness Yearbook (WCY), published by the Institute for Management and Development (IMD) since 1989, focuses on the competitiveness of 61 national and regional economies based on 312 different criteria grouped into four competitiveness factors: economic performance, government efficiency, business efficiency, and infrastructure. The countries are then ranked according to a synthetic index that combines each country's score in each input group, and which is interpreted as the capacity of a given nation's environment to create and sustain firm competitiveness.

Other indices and assessment methodologies include the European Innovation Scoreboard (EIS), an annual assessment of innovation performance in the individual member states of the EU set up by the European Commission as part of the Lisbon Strategy, and the Global Innovation Scoreboard report (GIS), which compares the innovation performance of the EU25 to that of the other developed and emerging economies in the world. This latter scoreboard is drawn up using 20 indicators that measure human resources, the creation of new knowledge, the diffusion and application of knowledge, and innovation finance. A composite indicator provides an overview of national performances.

Differing considerably from the previous indices, Richard Florida's creativity index draws on other types of variables in order to measure the intangible advantages of nations. In a study sponsored by DEMO (an European think tank) and the Alfred P. Sloan Foundation, which followed an approach similar to the one used by this author in the context of the USA, Florida has put forth an

Euro Creativity index based on the 3T's (i.e., technology, talent, and tolerance) that can be used to compare the various EU countries in regard to their creative com petitiveness and their openness to difference and cultural creativity.

Competitiveness indices at the regional level are less well known, but have been receiving increasing attention. Robert Huggins Associates, a research based economic consultancy and think tank, publishes a competitiveness index that is applied at the global, European, and the UK levels, and which examines and measures the com petitiveness of Europe's regions and nations: the ECI is computed from a large number of variables grouped into five factors (i.e, creativity, the knowledge economy, sec toral productivity performance, economic performance, and infrastructure and accessibility) (see **Box 1**). Since 1997, the UK Department of Trade and Industry (DTI) has published a biannual compendium of regional com petitiveness indicators, which was reformulated in 2004 so as to incorporate several new indicators. Other indices have been developed at the regional level with a view to assess those regions' characteristics as compared to those of their potential competitors. For example, the Bureau of Business Research of the University of Texas has sought to assess how Austin rates in comparison with three other major high tech regions in the United States, by drawing on historical data on variables that reflect the welfare of the people, resource endowment, regional competitiveness, and quality of life.

However, the use of these indices is liable to criticism due to several limitations. On the one hand, economic variables tend to be overrepresented, whereas other key features of competitiveness referring to social and en vironmental aspects are usually underrepresented, be cause the latter are not relevant for the managers and business people for whom these indices were designed: when they convey their perceptions, they tend be overly sensitive to short run business cycle fluctuations, which explains the significant changes in the ranking from 1 year to the next. On the other hand, because they are easy to 'consume' and diffuse through the media, there is a risk that these indices be used uncritically to compare the performance of regions and cities, even though they can conceal diverse realities. Moreover, they can be re sponsible for stigmatizing the regions at the bottom of the list, which will then be compelled to contract consultants who have become suddenly active in fashioning local strategies and helping them to copy the lessons from those at the top, regardless of whether the contexts and characteristics are completely different.

Box 1 NEWS RELEASE
Sheffield, UK, 13th December 2005
SILICON VALLEY'S SAN JOSE IS THE WORLD'S MOST KNOWLEDGE COMPETITIVE ECONOMY

Nordic Regions Continue to Thrive, while Shanghai is Most Improved Performer

San Jose, the heart of California's Silicon Valley, heads the WKCI 2005, followed by the US regions of Boston, San Francisco, and Hartford. Seattle takes fifth place with Grand-Rapids in sixth position. The highest ranked non-US region is Stockholm in Sweden, which climbs seven places to eighth (…)

Alongside Stockholm, the Nordic regions have almost all improved their rankings for the second year in a row. Both West Sweden (37th) and South Sweden (46th) move up, and are now competitive with the majority of US regions, while Denmark (up eleven places) and Norway (up two places) also perform strongly. Uusimaa (Helsinki) consolidated its phenomenal growth in recent years and is 21st ranked. These rankings are based on very high levels of knowledge-based employment and high levels of research and innovation.

Elsewhere in Europe the picture is more mixed, with the UK regions of London (56th), South East England (55th), and Eastern England (62nd) dropping down the rankings. With the exception of the UK, most Western European regions have seen an improvement in their position. Ireland improved 17 places to 74th, indicating that the country's economic growth has been correlated with a similar expansion of its knowledge base.

Majority of US regions have seen their performance either remain static or fall. Rochester ranked as high as fourth in 2003, slips again this year, and now occupies ninth position. Austin has slipped further still: it began in second place back in 2003, but falls another ten places this year to 19th. All six ranked Canadian regions have seen their position fall since 2004.

The region that has shown the best growth in creating a thriving knowledge-intensive economy is Shanghai, which more than doubled its performance since 2004, and is now ranked second only to Tokyo in terms of patent registrations per capita. Also, Mumbai and Bangalore in India have seen improvements in the performance, although they remain towards the bottom of the index. (…)

Source: Robert Huggins Associates, NEWS RELEASE

http://www.hugginsassociates.com/
news archive wkci 2005.php

See also: Agglomeration; Competitiveness; Environmental Regulation; Industrial Districts; Local Economic Development, Politics of; Local Economic Development; Sustainable Development; Uneven Development.

Further Reading

A A. V V (2004). *Special Issue on Regional Competitiveness. Journal of the Regional Studies Association* 38(9).

Amin, A. and Thrift, N. (1993). Globalization, institutional thickness and local prospects. *Reveu d'Economie Régionale et Urbaine* 3, 405 430.

Amin, A. and Thrift, N. (1994). *Globalization, Institutions and Regional Development in Europe*. Oxford: Oxford University Press.

Batey, P. and Friedrich, P. (eds.) (2000). *Regional Competition*. Berlin: Springer Verlag.

Becattini, G. (ed.) (1987). *Mercato e forze locali: Il distretto industriale*. Bologna: Il Mulino.

Bohle, D. and Bela, G. (2006). Capitalism without compromising: Strong business and weak labor in Eastern Europe's new transnational industries. *Studies in Comparative International Development* 41(1), 3 25.

Brenner, N. (2001). The limits to scale? Methodological reflections on scalar structuration. *Progress in Human Geography* 25(4), 591 614.

Bristow, G. (2005). Every one is a 'winner': Problematising the discourse of regional competitiveness. *Journal of Economic Geography* 5, 285 304.

Budd, L. and Hirmis, A. (2004). Conceptual framework for regional competitiveness. *Regional Studies* 38(9), 1015 1028.

Cabus, P. (2001). The meaning of local in a global economy: The 'region's advocacy of local interests' as a necessary component of current global/local theories. *European Planning Studies* 9(8), 1011 1029.

Camagni, R. (2002). On the concept of territorial competitiveness: Sound or misleading? *Urban Studies* 39(13), 2395 2411.

Cambridge Econometrics (2003). Study on the factors of regional competitiveness (a draft final report for The European Commission Directorate General Regional Policy).

Cheshire, P. and Gordon, I. (1998). Territorial competition: Some lessons for policy. *Annals of Regional Science* 32, 321 346.

Coe, N., Yeung, H., Dicken, P. and Henderson, J. (2004). Globalizing' regional development: A global production networks perspective. *Transaction Institute British Geographers* 29, 468 484.

Department of Trade and Industry (2005). The 2005 Productivity and Competitiveness Indicators. Department of Trade and Industry, UK. http://www.autoindustry.co.uk/docs/p ind2005.pdf (accessed Jul. 2008).

Dicken, P., Kelly, P., Olds, K. and Yeung, H. (2001). Chains and networks, territories and scales: Towards a relational framework for analysing the global economy. *Global Networks* 1(2), 89 112.

ESPON (2006). ESPON atlas, mapping the structure of the European territory, ESPON Project 3.1. Bonn: *Bundesamt fur Bauwesen und Raumordnung.*

European Commission (2001, 2004, and 2007). *Reports on Economic and Social Cohesion*. Luxembourg: Office for Official Publications of the European Communities.

Florida, R. (2002). *The Rise of the Creative Class: And How is Transforming Work, Leisure, Community and Everyday Life*. New York: Basic Books.

Florida, R. and Tinagli, I. (2004). Europe in the creative age. Alfred P. Sloan Foundation and Demos. http://www.demos.co.uk/files/EuropeintheCreativeAge2004.pdf (accessed Jul. 2008).

Fujita, M., Krugman, P. and Venables, A. (1999). *The Spatial Economy, Cities, Regions and International Trade*. Cambridge, MA: The MIT Press.

Gardiner, B., Martin, R. and Peter, T. (2004). Competitiveness, productivity and economic growth across the European regions. *Regional Studies Association* 38(9), 1045 1068.

Group of Lisbon (1995). *Limits to Competition*. Cambridge, MA: MIT Press.

Hadjimichalis, C. (2006). Non economic factors in economic geography and in new regionalism: A sympathetic critique. *International Journal of Urban and Regional Research* 30(3), 690 704.

Hallin, G. and Malmberg, A. (1996). Attraction, competition and regional development in Europe. *European Urban and Regional Studies* 3(4), 323 337.

Kaplinsky, R. (2000). Spreading the gains from globalization: What can be learned from value chains analysis? *IDS Working paper* 110. Brighton: Institute of Development Studies.

Kitson, M., Martin, R. and Tyler, P. (2004). Regional competitiveness: An elusive yet key concept? *Regional Studies* 38(9), 991 999.

Kohno, H., Nijkanp, P. and Poot, J. (eds.) (2000). *Regional Cohesion and Competition in the Age of Globalization*. Cheltenham: Edward Elgar Publishing.

Kong, Y. (2003). Persistent dumping, competition and welfare. *Journal of International Trade and Economic Development* 12(1), 19 37.

Krugman, P. (1991). *Geography and Trade*. Leuven: Leuven University Press.

Krugman, P. (1994). Competitiveness: A dangerous obsession. *Foreign Affairs* 73, 28 44.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy* 99, 483 499.

Krugman, P. (1996). Making sense of the competitiveness debate. *Oxford Review of Economic Policy* 12(3), 17 25.

Malecki, E. (2002). Hard and soft networks for urban competitiveness. *Urban Studies* 39(5 6), 929 945.

Malecki, E. (2004). Jockeying for position: What it means and why it matters to regional development policy when places compete. *Regional Studies* 38(9), 1101 1120.

Markusen, A. (1997). Sticky places in a slippery space: A typology of industrial districts. *Economic Geography* 72(3), 293 313.

Maskel, P. and Malmberg, A. (1999). Localised learning and industrial competitiveness. *Cambridge Journal of Economics* 23, 67 185.

Morgan, K. (1997). The learning region: Institutions, innovation and regional renewal. *Regional Studies* 31(5), 491 503.

Moulaert, F. and Sekia, F. (2003). Territorial innovation models: A critical survey. *Regional Studies* 37(3), 289 302.

OECD (2002). *Environmental Issues in Policy Based Competition for Investment: A Literature Review*. Paris: Environmental Policy Committee, OECD.

Paasi, A. (2004). Place and region: Looking through the prism of scale. *Progress in Human Geography* 28(4), 536 546.

Peck, J. (2005). Struggling with the creative class. *International Journal of Urban and Regional Research* 29(4), 740 770.

Porter, M. (1990). *The Competitive Advantage of Nations*. New York: Free Press.

Porter, M. (1998). *On Competition*. Boston, MA: The Harvard Business Review Book Series.

Porter, M. (2000). Location, competition, and economic development: Local clusters in a global economy. *Economic Development Quarterly* 14(1), 15 34.

Raines, P. (2000). Regions in competition: Inward investment and regional variation in the use of incentives. *Regional Studies* 34, 291 296.

Saxenian, A. (2001). Inside out: Regional networks and industrial adaptation in Silicon Valley and Route 128. In Granovetter, M. & Swedberg, R. (eds.) *The Sociology of Economic Life*, pp 357 374. Cambridge: Westview Press.

Scott, A. (1998). *Regions and The World Economy: The Coming Shape of Global Production, Competition, and Political Order*. Oxford: Oxford University Press.

Scott, A. and Storper, M. (2003). Regions, globalization, development. *Regional Studies* 37(6 7), 579 593.

Storper, M. (1997). *The Regional World: Territorial Development in a Global Economy*. New York: Guilford Press.

Storper, M. (2006). Entrepreneurship, innovation and industrial development: Geography and the creative field revisited. *Small Business Economics* 26, 1 24.

Storper, M. and Scott, A. (1995). The wealth of regions: Market forces and policy imperatives in local and global context. *Futures* 25(5), 505 526.

Swyngedouw, E. (1997). Neither global nor global: 'Glocalization' and the politics of scale. In Cox, K. R. (ed.) *Spaces of Globalization: Reasserting the Power of the Local*, pp 137 166. New York: Guilford Press.

Thomas, K. (2003). Geographic scales and the competition for economic growth. *American Behavioral Scientist* 46, 987 1001.

Turok, I. (2004). Cities, regions and competitiveness. *Regional Studies* 38(9), 1069 1084.

Relevant Websites

http://ec.europa.eu
Analysis of Competitiveness, European Performance in Competitiveness, Enterprise Policy, European Commission Enterprise and Industry.

http://www.oecd.org
Competition, Organisation for Economic Co operation and Development (OECD).
http://www.cfo1ric.org/index.php
Centre for International Competeveness.
http://www.proinno europe.eu
PRO INNO Europe (Directorate General Enterprise and Industry), European Innovation Scoreboard, Comparative Analysis of Innovation Performance.
http://www.weforum.org
The Global Competitiveness Report (GCR), World Bank (World Economic Forum).
http://www.imd.ch
The World Competitiveness Yearbook (WCY), IMD World Competitiveness Center.
http://www.weforum.org
World Economic Forum.
http://www.hugginsassociates.com
World Knowledge Competitiveness Index (WKCI); The European Competitiveness Index at Robert Huggins Associates.

Regional Connectivity

M. Sokol, Queen Mary, University of London, London, UK

Glossary

Accessibility The measure of the ease with which one place can be accessed or reached from another, while taking into account the attributes of destination; usually measured through economic potential.

Beta Index A measure of the connectivity of a network, calculated as a ratio of links to nodes.

Connectivity Sometimes also referred to as connectedness: connectivity is a measure of the degree to which a network as a whole, or its parts, is internally connected.

Economic Mass One of the two functions in the measurement of accessibility; economic mass represents the attractiveness of destination in terms of the size of market, population size, income, gross domestic product, spending power, volume of activities, or opportunities to be reached.

Economic Potential Also referred to as market potential: a measure of accessibility which increases with the economic mass (i.e., population, market size etc.) and declines with impedance (i.e., distance, travel time, or cost).

Impedance One of the two functions in the measurement of accessibility; impedance represents the effort, time, distance, or cost needed to reach a destination.

Regional Connectivity–Accessibility The ease with which people, materials, and information can be moved from one location to another, within and between regions.

Self-Potential The economic mass of a zone or region divided by its estimated intrazonal distance.

Introduction

The aim of this article is to provide a definition of regional connectivity–accessibility, to highlight the differences between the concepts of connectivity and accessibility, to point at various ways in which connectivity and accessibility can be measured, to give examples of how regional connectivity–accessibility can be (geo)graphically represented, and to discuss the implications of connectivity–accessibility for regional development and regional policy. This article argues that the importance, the nature, and the exact direction of the causal link between regional development and regional connectivity–accessibility are open to debate.

Defining and Measuring Regional Connectivity–Accessibility

Connectivity–accessibility in general can be simply defined as the ease with which people, materials, and information can be moved from one location to another location. In the regional development context, however, a distinction needs to be made between intraregional connectivity–accessibility (i.e., connectivity–accessibility within a given region) and interregional connectivity–accessibility (i.e., connectivity–accessibility between regions). The former can be defined as the ease with which people, materials, and information can be moved from one location to another location within a given region, while the latter refers to the ease of movement between regions.

In contemporary regional economies these movements are usually supported by dedicated transport infrastructure (e.g., rail, road, air, and water) and telecommunication networks (e.g., telephone, telegraph, and Internet). Therefore, regional connectivity–accessibility could also be seen as a measure of the ability of these infrastructures and networks to reduce the friction of distance within and between regions.

Beyond these simple and rather general definitions, one finds that measuring regional connectivity and accessibility is a complicated task. It is important to realize that while these two terms (connectivity and accessibility) are often used interchangeably, there are in fact subtle differences between the concept of connectivity and the concept of accessibility. In addition to this, there are numerous ways in which both connectivity and accessibility can be measured, some of which are described below.

Connectivity

Connectivity (sometimes also referred to as connectedness) usually relates to the features of a given transport or communications network and can be measured in two basic ways. The first one is to measure the level of connectivity of an entire network, that is, the degree to which the network as a whole is internally connected. Numerous indices have been devised to measure such connectivity. One of the most frequently used is the beta index, which is calculated as the ratio of links (connections) to nodes. The higher the number of links to nodes, the higher the degree of connectivity of a given network. A beta index higher than 1 usually indicates a complex network with a high level of connectivity. It is often assumed that regional economies go through various stages of development and that the more a region develops, the denser its transport and communications networks are. The improving level of connectivity (beta index) is therefore often seen as an indicator of economic development. However, as it will become clear later in the article, the link between connectivity and regional economic development cannot be taken for granted.

The second option is to measure the level of con nectivity of individual places (nodes) within a network. In other words, node connectivity measures the degree to which individual nodes are connected with other nodes within the network. Again, this can be done in a number of ways. The simplest measure of node connectivity is a number of direct links that a particular node has with other nodes. The more direct links a particular node has, the higher its degree of connectivity. Such a measure of connectivity may be useful in some cases (e.g., identifying airports that offer direct flights to the largest pool of destinations) but less so in others. More sophisticated measurements of connectivity would, for instance, take into account the frequency of transport connections (e.g., number of flights to each destination) and/or the ca pacity of links (e.g., the capacity of each flight). Further to this, one can also measure the actual flow of people, materials, or information between the nodes (e.g., the actual number of passengers flying between different airports per day or per year).

Accessibility

Unlike the concept of connectivity, the concept of ac cessibility is based on the explicit recognition of the importance of both characteristics of nodes (attributes of destinations) and geographical distance between the nodes. Within the concept of accessibility, geographical distance (i.e., the length of links between the nodes) is considered as an impedance for the movement of people, materials, and information. While transport and tele communication infrastructures may reduce the friction of distance, it still takes a certain amount of time, money, and/or effort to move people, materials, or information around. Travel time and travel cost is often a direct function of the distance covered, although this does not always hold true. One way or another, the level of im pedance is clearly important to economic agents. To use the air travel analogy again, people want to know not only how many destinations they can fly to from their local airport (connectivity), but also how far these des tinations are and, therefore, how long the flight would take and how much it would cost. In addition to this, people want to know what these destinations have to offer. The attributes of destinations (i.e., the level of at tractiveness of nodes) represent the second dimension of the concept of accessibility.

From an economic point of view, the level of at tractiveness of destinations is frequently associated with the size of their markets, which can be expressed as an economic mass or a volume of economic activity either in terms of population size, income, spending power, gross domestic product, opportunities to be reached, or similar characteristics. This is based on an (contested) assump tion that the bigger the market a firm has access to, the better the prospects for the firm's profitability and growth. This way, the level of accessibility (to markets) is directly associated with the expected economic per formance and competitiveness of firms.

This assumption is encapsulated in the market po tential or economic potential model, which represents the most frequent way of measuring regional accessibility. In this model, the economic potential of a region is a positive function of the size of markets (economic mass) that can be accessed from the region, and a negative function of impedance (e.g., geographical distance) that needs to be overcome when reaching those markets. In the simplest measure of regional accessibility, the po tential of any region can be calculated by summing the population of all other regions in the system and dividing these by some measure of their intervening distance. The potential of a region, i, thus can be calculated as

$$P_i = (j = 1 \rightarrow j = n) \sum M_j / D_{ij}$$

where P_i is the potential of region i, M_j is a measure of economic mass (e.g., population) in region j, and D_{ij} is a measure of impedance (distance) from region i to region j, and n is the total number of regions in the system. The regions located closest to the biggest markets (and therefore presumably having the cheapest access to them) would be seen as having the highest economic potential. The exact same formula can be used to measure intraregional accessibility, that is, the economic potential of individual zones within a given region.

There are, however, two basic problems with the above measure of accessibility. First, it is not clear what weight should be given to the self potential of a given region in the calculation of its overall potential. The application of different indexes may lead to variegated measurements of accessibility.

Second, as indicated above, distance is not always the best indicator of an impedance. The example of air travel can again serve as a good example. Indeed, in an era of relatively low cost air travel, readers would be familiar with the ironic situation where their journey (by train, bus, or taxi) to their local airport costs more than the actual flight they are taking. Further to this, it is perfectly possible that, especially in big cities, the journey to the nearest airport at the outskirts may actually take more time than a short haul flight to a neighboring country. Therefore, alternative measures of impedance may be necessary, and transport costs and/or travel times are frequently used instead of travel distances to calculate accessibility more accurately.

Beyond these two points, there are several other difficulties associated with the above measure of accessibility. It is perhaps too optimistic to expect that two variables – economic mass and impedance – alone can capture the complexity of the movement of people, materials, and information within and between regions, characterized by different economic and social structures, different transport needs, different modes of transport, etc. Neither is the above formula for calculating economic potential sensitive to the actual users of transport and telecommunication infrastructures, barriers they may face in terms of age, gender, race, class, culture, or income, or their subjective perceptions of being connected or disconnected. Finally, the potential model is, for obvious reasons, inadequate when it comes to the measurement of accessibility of information through telecommunication means or electronic channels.

Regional Connectivity–Accessibility: Centrality and Peripherality

Despite the above limitations, various versions of the economic potential model are being used to measure both intra and interregional accessibility, which in turn is often used as an indication of the centrality or peripherality of regions or their parts. This can be illustrated by taking an example of an imaginary island region (**Figure 1**) in which the difference between the measures of connectivity and accessibility can be demonstrated. Let us imagine that our island region is of a rectangular shape (approximately 200 km wide and 400 km long) and its population lives in 15 compact urban centers spaced 100 km from each other, forming a perfectly regular grid. Twelve of these urban centers are located along the coast (A–L), three are located inland (X, Y, Z), but all are connected by a rectangular system of roads as shown in **Figure 1a**. The car is the only mode of transport on the island. If we imagine that these roads are narrow country roads with a maximum speed limit of 50 km h^{-1}, then it would take 2 h to drive from one urban center to another and 4 h for a two way return journey. It would, in all, take 12 h to travel from center A to center G located at the opposite side of the island.

What would be the degree of connectivity and accessibility on the island? Using the beta index, we can see that the overall network connectivity would be 1.47 indicating a high level of connectivity (**Figure 1a**). The level of connectivity of individual nodes within the network, however, clearly varies (**Figure 1b**).

Does this picture of connectivity differ from that of accessibility? If all the urban centers on the island were of equal size (e.g., towns with population 10 000 each) then the picture of accessibility measured through the economic potential model (with population as a measure of economic mass) would show a rather similar pattern (**Figure 1c**). Urban centers located in the geographical center of the island would display the highest level of accessibility while those at the geographical periphery would be considered as the least accessible.

If, for whatever reason, the populations of these urban centers start to vary, the picture of accessibility can change considerably due to changing market potential. If, for instance, urban center G were to become a city with a population of 500 000, the picture of accessibility of this regional economy would be dramatically transformed (**Figure 1d**). Indeed, city G with its population of half a million people would exceed, by several times, the population (market size) of the rest of the island. Therefore, due to the weight of its own self potential (however calculated), city G (formerly one of the least accessible places) would become the most accessible place on the island, all this without changing the levels of connectivity. Thanks to their proximity to city G, urban centers F and H would overnight become the second best accessible places on the island, perhaps attracting some population growth, thus further strengthening their own economic potential, while having a similar knock on effect on town Z (**Figure 1d**).

If this process continues, the all too familiar picture of the core–periphery differentiation of the regional economy is likely to happen (**Figure 1e**). In such cases, the economic core of the region continues to gain an increasing share of population and economic activity, while the periphery faces depopulation and economic decline. This uneven development, in turn, is likely to have an impact on the provision of transport and telecommunication infrastructure within the region. Indeed, the decline in the periphery may lead to a deteriorating road network there. In contrast, the increasing demand for transport in the core areas may be met by the building of new and better roads with greater capacity and higher driving speeds (**Figure 1e**). This way, the differences in the accessibility (and also connectivity) of these areas would be exacerbated, following the core–center–periphery pattern. City G in the core would probably become both the most accessible and the most connected place. In contrast, town A on the periphery would be one of the least connected and one of the least accessible places on the island (**Figure 1e**).

Significant debate exists among economists, economic geographers, transport geographers, and town planners as to how the regional economic and settlement structure will continue to develop under the conditions of constantly improving transport and telecommunications. Some argue that further polarization may occur through the continuing growth of metropolitan areas, others point at the evidence of counterurbanization and dispersal. Some predict that future spatial forms will take the shape

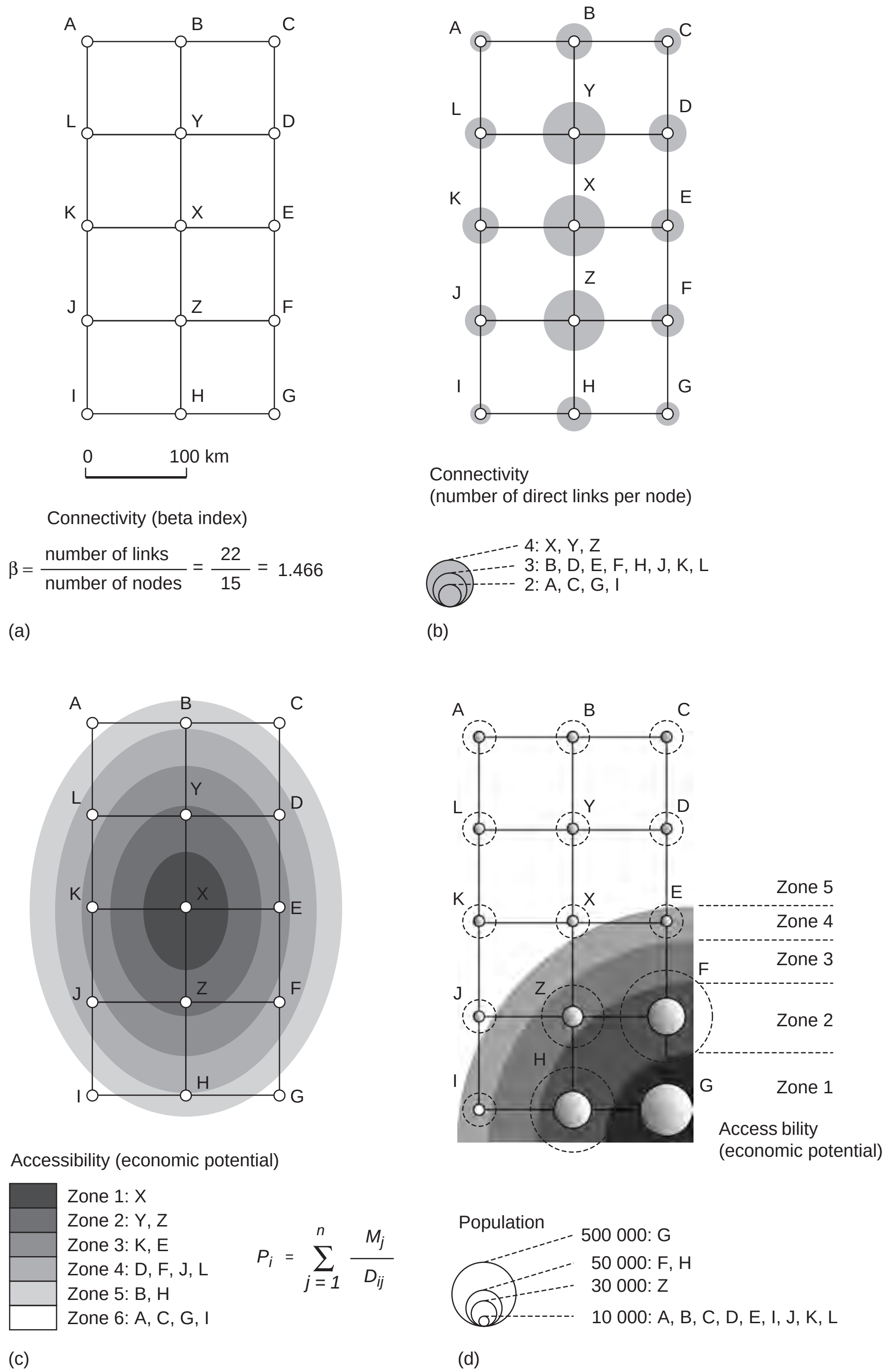

Figure 1 (a) Regional connectivity (beta index). (b) Regional connectivity (Direct links per node). (c) Regional accessibility (Economic potential). (d) Regional accessibility (Economic potential). (e) Regional connectivity accessibility (core-periphery). (f) Polycentric Mega-City Region. (a) Author: Martin Sokol © 2007. Cartographic support: Ed Oliver, Queen Mary, University of London. (b) Author: Martin Sokol © 2007. Cartographic support: Ed Oliver, Queen Mary, University of London. (c) Author: Martin Sokol © 2007. Cartographic support: Ed Oliver, Queen Mary, University of London. (d) Author: Martin Sokol © 2007. Cartographic support: Ed Oliver, Queen Mary, University of London. (e) Author: Martin Sokol © 2007. Cartographic support: Ed Oliver, Queen Mary, University of London. (f) Author: Martin Sokol © 2007. Cartographic support: Ed Oliver, Queen Mary, University of London.

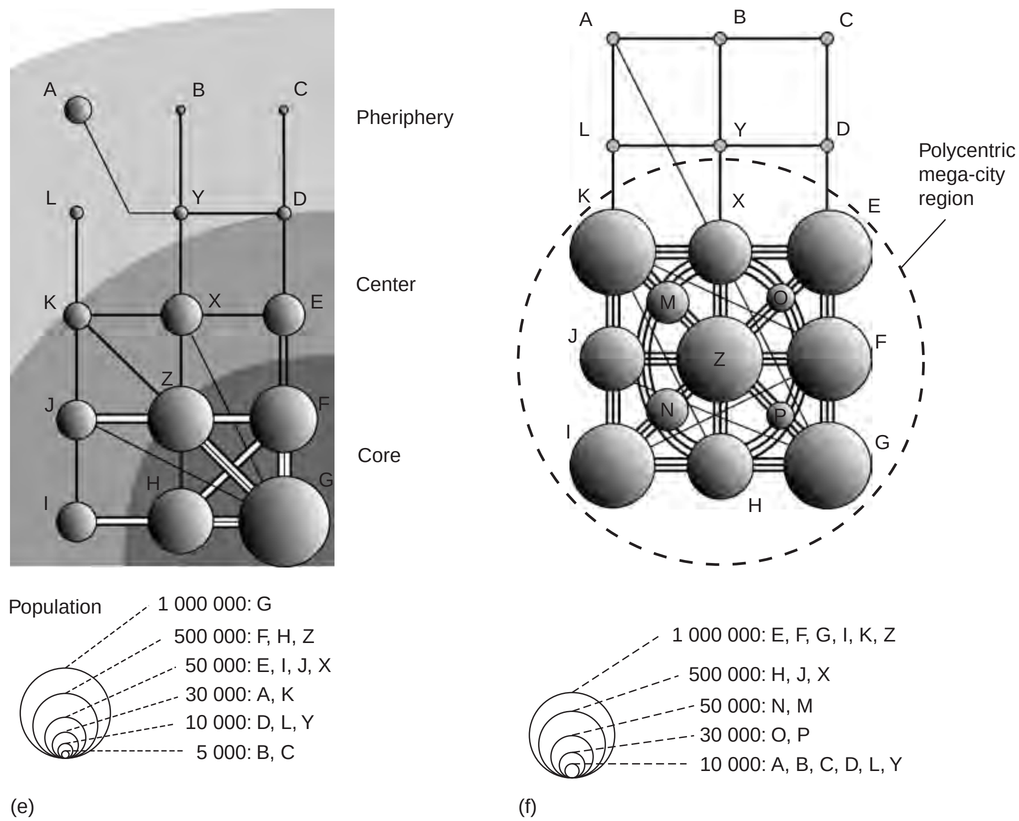

Figure 1 (Continued).

of densely urbanized zones along major transport corridors; others foresee the emergence of highly integrated polycentric urban regions or polycentric mega city regions (**Figure 1f**). It remains to be seen which of these trends will become predominant in the future.

In the meantime, policy makers are most likely to be confronted with regional differences of the core–periphery type. Their strategies to overcome peripherality through connectivity–accessibility usually follow one of the following basic policy options (**Figure 2**): (1) improving connectivity–accessibility within the periphery, for example, through a program of road building between the urban centers of the periphery (**Figure 2a**); (2) improving connectivity–accessibility between the periphery and the core zone, for instance, through the provision of a high speed train line (**Figure 2b**); and (3) the combination of these two approaches (**Figure 2c**). More recently, policies, which aim to improve electronic accessibility in peripheral areas, are seen as a way forward, although their outcomes are uncertain.

In addition to this, policy makers may aim to improve external (interregional) connectivity–accessibility (**Figure 3**). Opening a port in town A, for instance, would give it access to economic centers on neighboring islands, thus providing the impetus for its possible economic recovery (**Figure 3a**). However, if the island region as a whole is part of the periphery of a much wider economic system, then improved access to core regions (and their markets) may be a priority (**Figure 3b**). These model situations are not dissimilar from those policy makers face in real life situations. The exact impact of policy measures on connectivity–accessibility and their implications for regional development are, however, not always clear. In order to make informed decisions about the numerous policy options, and in order to understand, analyze, and visualize the impact of changing regional connectivity–accessibility, various (geo)graphical representations have been developed, some of which are presented below.

(Geo)graphical Representations of Regional Connectivity–Accessibility

There are numerous ways in which both regional connectivity and accessibility can be mapped or otherwise (geo)graphically represented. One of the simplest representations of accessibility are contour maps, such as those produced in the early 1980s in a pioneering study on regional accessibility and economic potential in Europe (**Figure 4**). Europe is a particularly good example, since the contours of geographical peripherality broadly correspond to the contours of economic peripherality. The economic potential of European regions has been calculated using a simple potential model and

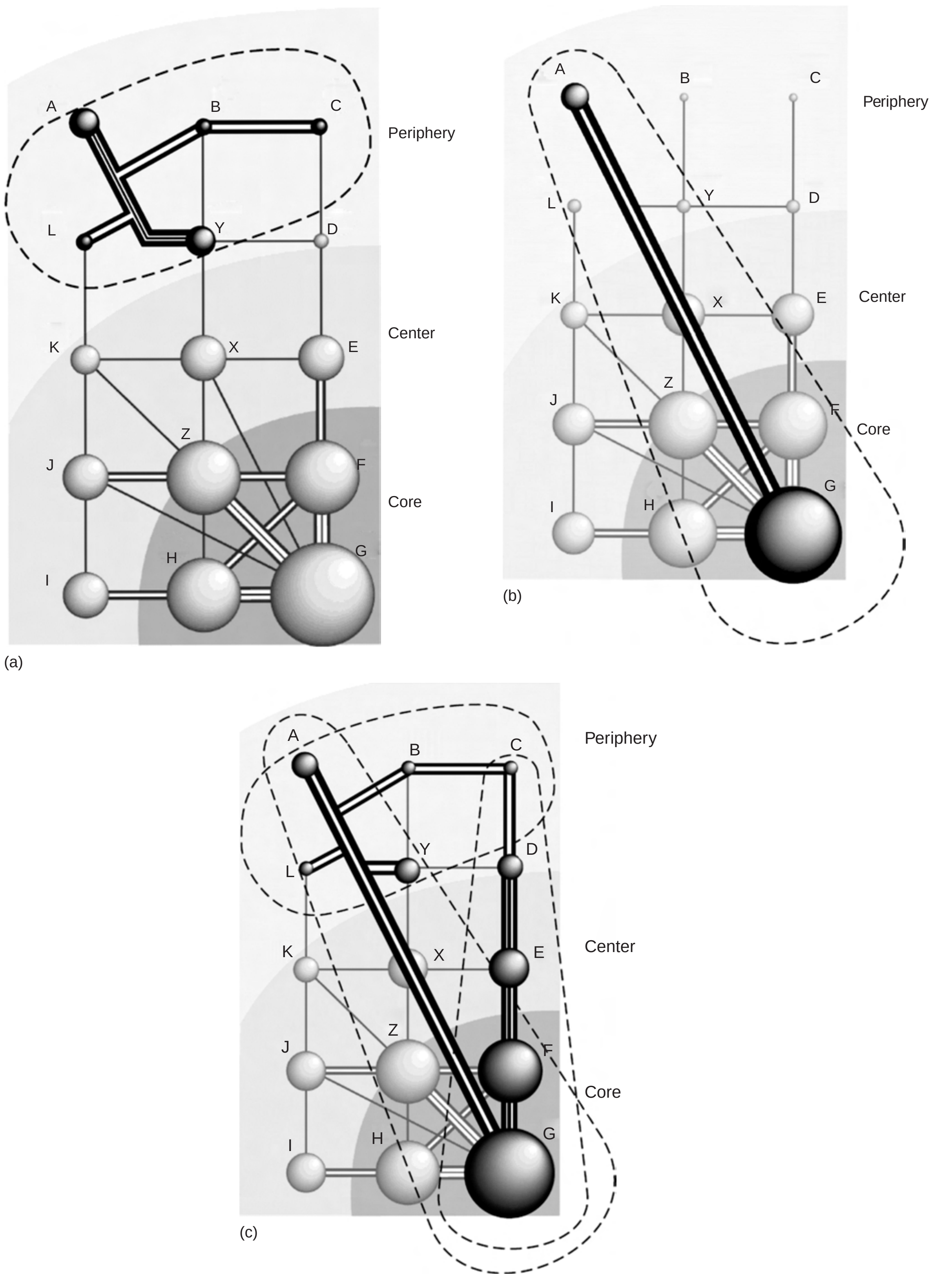

Figure 2 (a) Regional connectivity accessibility: policy option A. (b) Regional connectivity accessibility: policy option B. (c) Regional connectivity accessibility: policy option C. (a) Author: Martin Sokol © 2007. Cartographic support: Ed Oliver, Queen Mary, University of London. (b) Author: Martin Sokol © 2007. Cartographic support: Ed Oliver, Queen Mary, University of London. (c) Author: Martin Sokol © 2007. Cartographic support: Ed Oliver, Queen Mary, University of London.

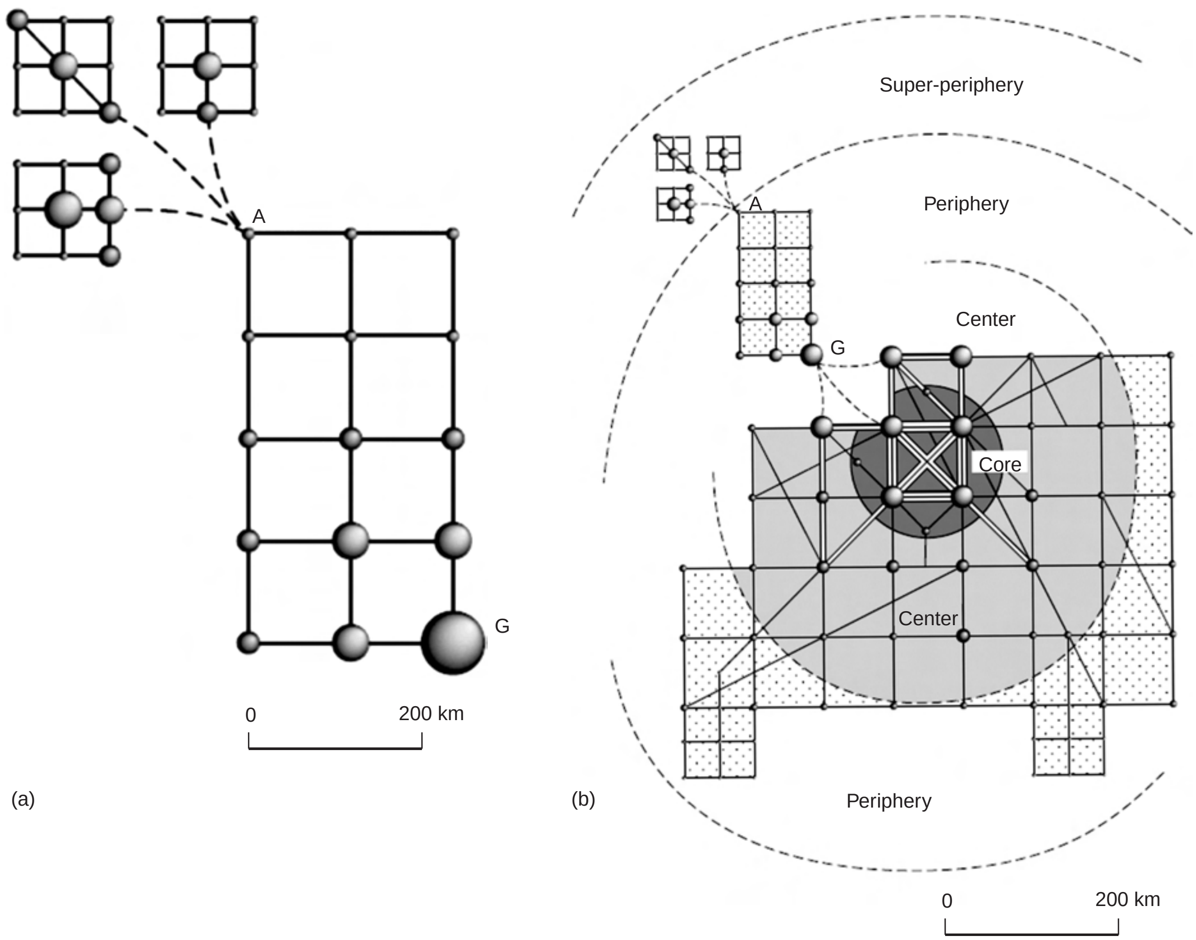

Figure 3 (a) Interregional connectivity accessibility. (b) Interregional connectivity accessibility. (a) Author: Martin Sokol © 2007. Cartographic support: Ed Oliver, Queen Mary, University of London. (b) Author: Martin Sokol © 2007. Cartographic support: Ed Oliver, Queen Mary, University of London.

mapped as potential contours. Each region's value is expressed as a percentage of the highest value found in Europe, with contours interpolated at successive 10% intervals up to the maximum. Based on 1977 figures, this exercise showed dramatic differences in the levels of accessibility in Europe. The regions with the highest potential were the centrally located regions in West Germany, Belgium, and the Netherlands, followed closely by Paris and London. The least accessible regions were to be found in the geographical periphery of Eur ope, that is, in Greece, Portugal, Spain, Southern Italy, and Ireland (see **Figure 4**).

A further elaboration of this approach is the use of accessibility surface maps (**Figure 5**) which have been used to evaluate the expected impact of tran European networks (TENs). TENs include high speed rail, roads, combined transport, inland waterways, and conventional rail network improvements which the European author ities have been putting in place to overcome the very accessibility gap identified above. Focusing on high speed rail, the study calculated the existing and expected levels of accessibility in Europe should the key high speed rail links be completed. The study concluded that while economic potential (measured as population potential) increases all over Europe, the general pattern of acces sibility would not change much. The highest accessibility levels would still be found in the major urban centers of central Europe (**Figure 5**).

An alternative to both contour maps and accessibility surface maps are simple accessibility maps where varie gated levels of accessibility are represented by variegated intensity of shading or color scale. When such maps are produced with the support of geographic information systems (GIS) a great degree of precision and geo graphical detail can be achieved, while allowing a spatial analysis to be performed. An example in **Figure 6** shows road accessibility in Europe in 2006, expressed as an index value in which the European average is set to 100.

Another way of representing the levels of accessibility is through time–space maps (**Figures 7** and **8**). **Figure 7** demonstrates the effects on accessibility of building a high speed train line between A and G on our imaginary island region (**Figure 7a**). Given that such a train can achieve speeds of up to 300 km h ^{1}, the travel time

Figure 4 Econom c potent a of Europe n 1977 (Contour map). Source: Keeb e, D., Owens, P. L. and Thompson, C. (1982). Reg ona access b ty and econom c potent a n the European Commun ty. *Regional Studies* 16(6), 419–432. [F g. 7, page 429]. Repr nted by perm ss on of Tay or and Franc s Ltd (http:// nformawor d.com).

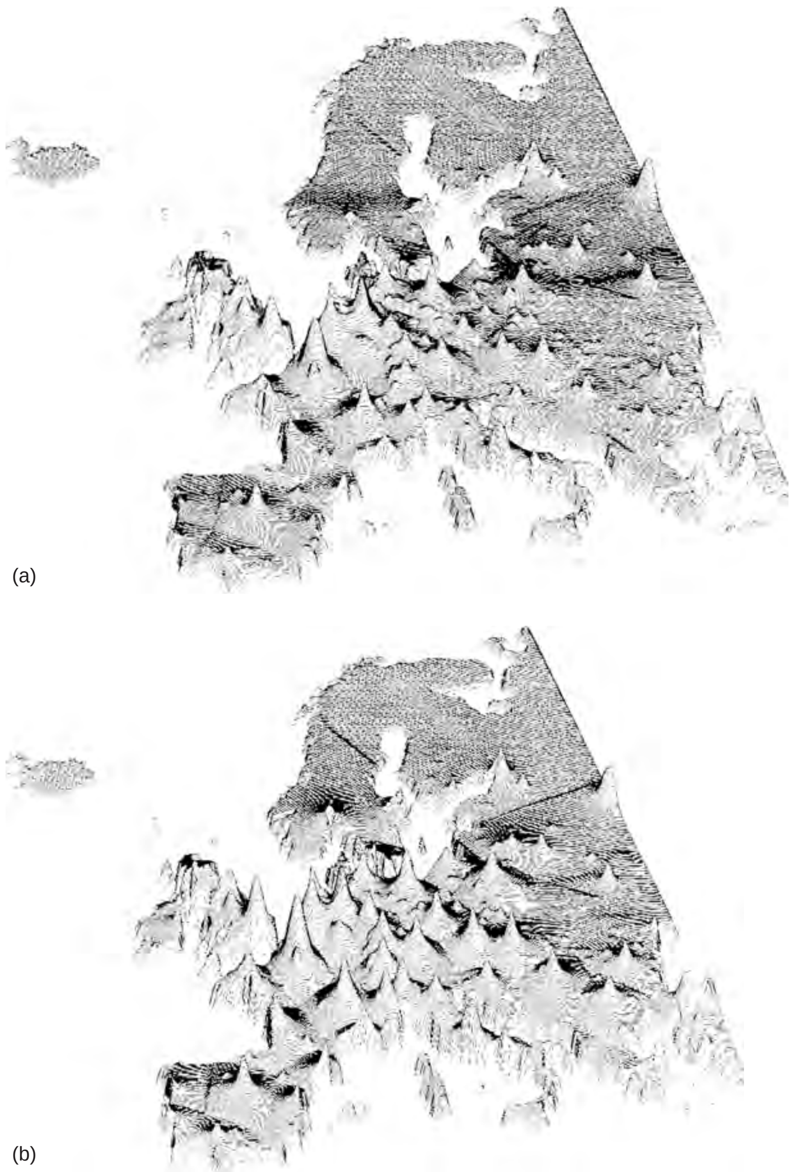

(a)

(b)

Figure 5 Accessibility surface of Europe by rail (accessibility surface maps): (a) 1993; (b) 2010. Source: Vickerman, R., Spiekermann, K. and Wegener, M. (1999). Accessibility and economic development in Europe. *Regional Studies* 33(1), 1–15. [Fig. 4, page 10]. Reprinted by permission of Taylor and Francis Ltd (http://informaworld.com).

between A and G could be easily reduced to less than 2 h. Meanwhile, town X located in the middle suffers from the so called tunnel effect, since the high speed train passes through it, but does not stop there. The same can be said about town Y, located nearby. Therefore, despite their spatial proximity to each other and to the high speed rail (but not being directly served by it), travel time between X and Y via train remains substantial (see **Figure 7b**). **Figure 8** shows time–space maps for the whole of Europe. The first map (**Figure 8a**) is the base map, which assumes a travel speed of $60\ \text{km h}^{-1}$ on all links of the network (time and distance are represented equally). The second map (**Figure 8b**) is adjusted for speeds achieved on the rail network in Europe in 1993. The third map (**Figure 8c**) shows how the European continent will have shrunk in time–space with the full completion of the high speed network in 2010. While such a time–space map is visually impressive, it is never accurate. Indeed, places that are close, but have no access, to major high speed train corridors appear well accessible on the map, but in reality suffer from negative tunnel effects. In other words, the map does not capture the full extent of the uneven nature of time–space shrinkage, in which, paradoxically, some places may become less accessible than before, in relative or absolute terms.

The following examples show some of the ways in which connectivity can be represented. **Figure 9** depicts estimated commuting flows between urban centers in the

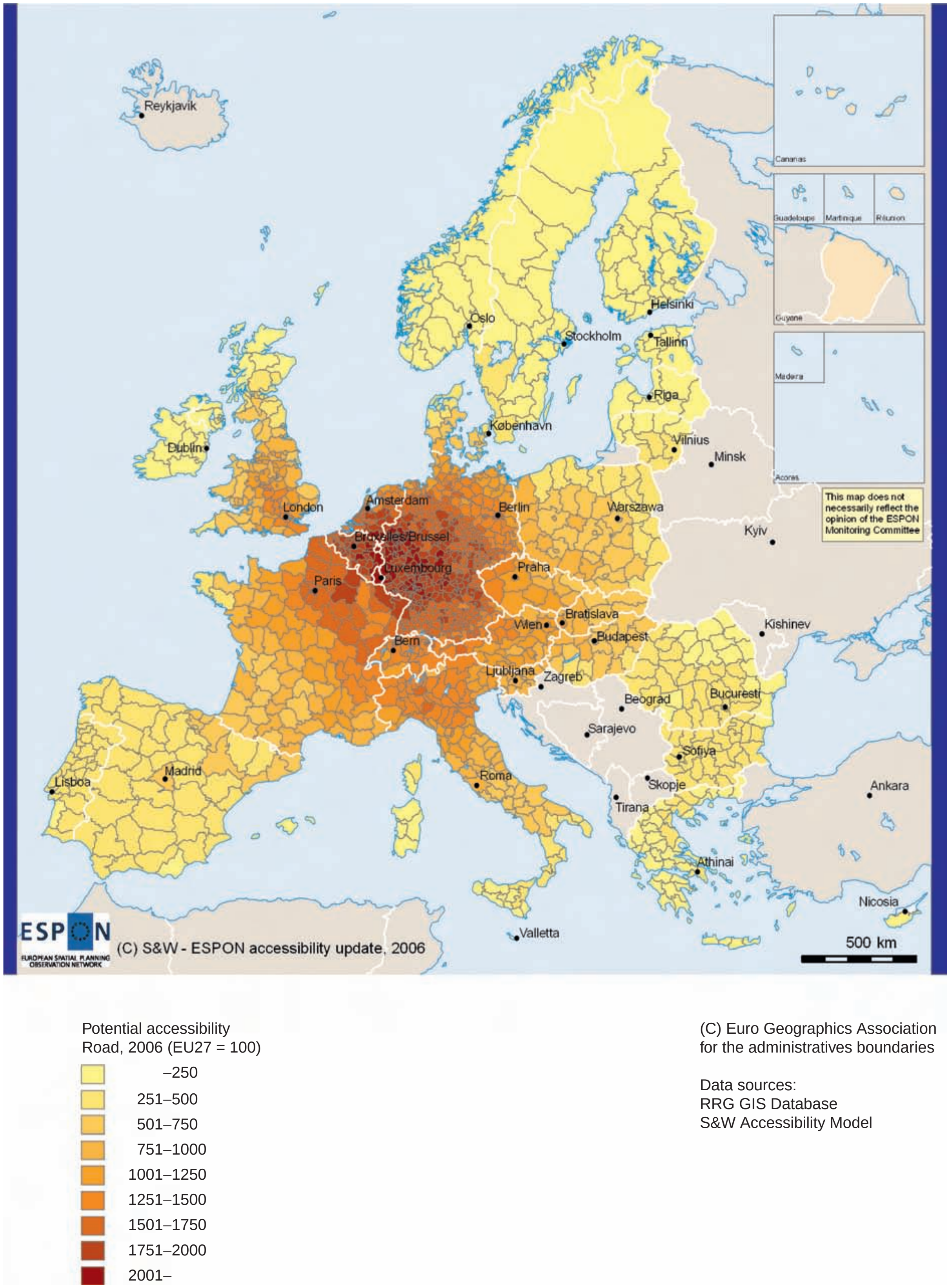

Figure 6 Potential accessibility by road in Europe in 2006. Source: ESPON and S&W (2007). *Update of Selected Potential Accessibility Indicators (Final Report).* Luxembourg: ESPON. http://www.espon.eu/mmp/online/website/content/projects/947/1297/file 2724/espon accessibility update 2006 fr 070207.pdf (accessed on January 2008). [p 16].

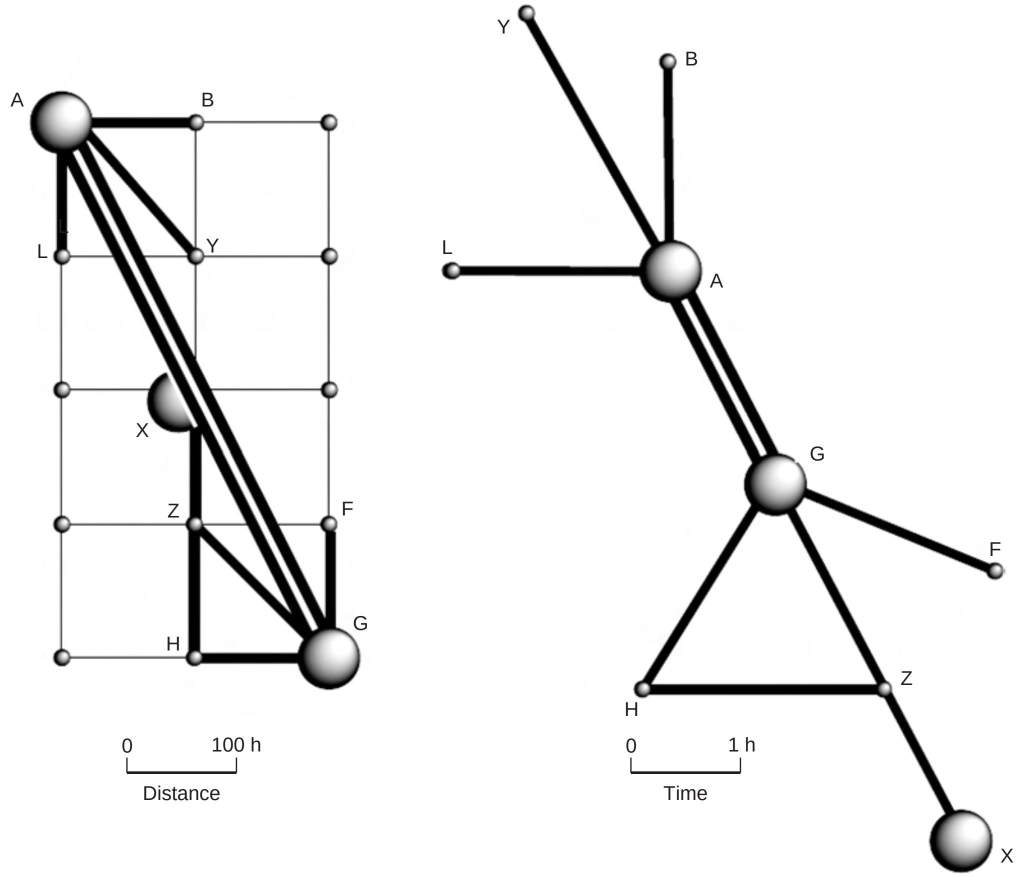

Figure 7 Regional connectivity accessibility: time space map. Author: Martin Sokol © 2007. Cartographic support: Ed Oliver, Queen Mary, University of London.

southeast of England, one of the biggest regional economies of Europe, showing the growing intraregional integration of its labor market in what can be seen as an emerging polycentric mega city region. **Figure 10**, on the other hand, shows the interregional connectivity of London with other major city regions in the world in terms of air traffic. The figure depicts the connections between the selected cities in terms of total passenger flows from January to August 2001. The size of nodes varies with the total number of incoming or outgoing passengers, while the size of links varies with the number of passengers flying between the two cities. The dominance of New York, London, Paris, and Los Angeles in these flows is immediately apparent.

This dominance broadly reflects the pattern of what Peter Taylor and his colleagues at the Globalization and World Cities Study Group and Network (GaWC) at the University of Loughborough (www.lboro.ac.uk/gawc) call a world city network. Various innovative techniques have been used to map information flows within the world city network. **Figure 11** shows levels of connectivity between 123 major cities of the world city network in terms of advanced producer service flows. New York (NY) and London (LN) come out on top, followed by other major city regions, such as Los Angeles (LA), Chicago (CH), Paris (PA), Milan (ML), Tokyo (TK), Hong Kong (HK), and Singapore (SG). The remainder of the cities included in the analysis display varying levels of connectivity, indicating a complex and uneven geography of information flows in the era of globalization.

The final example is from the work of the Centre for Urban and Regional Development Studies (CURDS) at the University of Newcastle upon Tyne (http://www.ncl.ac.uk/curds/) concerned with telecommunications networks and services. CURDS researchers have developed a number of indicators to measure ICT connectivity, both in terms of ICT supply and their uptake by businesses and households. **Figure 12** shows estimated levels of business telecommunications access and uptake, revealing major differences between European regions. The implications of regional connectivity–accessibility for regional development will be considered in turn.

Regional Connectivity–Accessibility and Regional Development

The relationships between regional connectivity–accessibility and regional development are numerous and

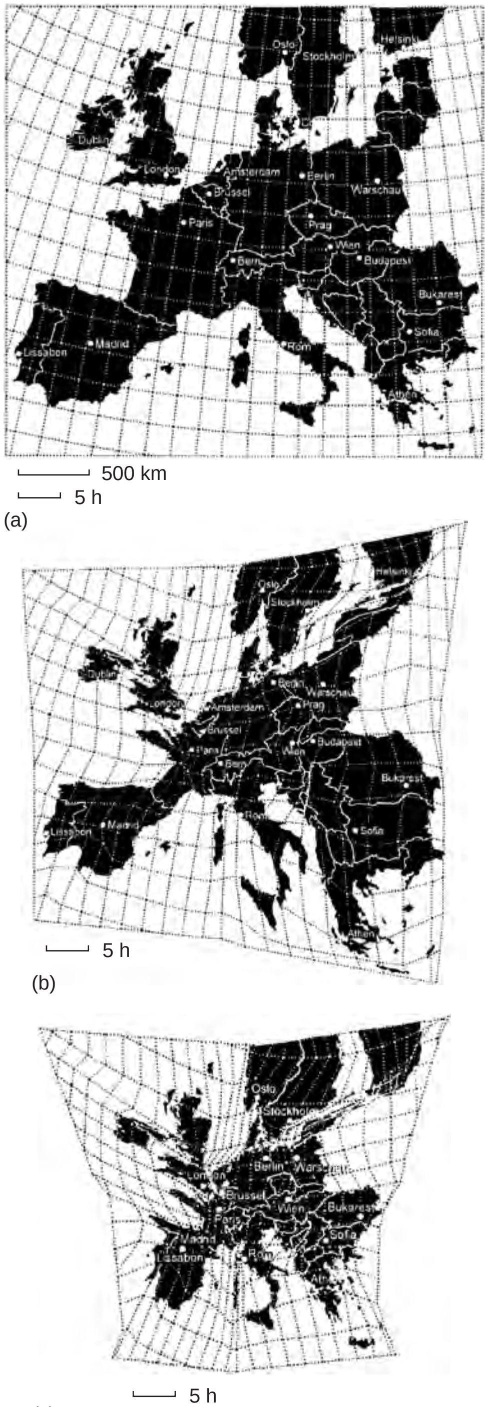

Figure 8 Time space maps of European rail network accessibility: (a) base map; (b) accessibility in 1993; (c) accessibility in 2010. Source: Vickerman, R., Spiekermann, K. and Wegener, M. (1999). Accessibility and economic development in Europe. *Regional Studies* 33(1), 1 15. [Fig. 3, page 9]. Reprinted by permission of Taylor and Francis Ltd (http://informaworld.com).

complex, but can be approached at three basic levels. The first approach is to analyze the economic contribution of activities directly associated with the movement of people, materials, and information within and between regions. The construction, maintenance, and operation of transport and telecommunications networks, provided by the state or private firms (e.g., road construction, rail maintenance, airline operators, telecom providers), with their contribution to employment and regional gross domestic product (GDP), for instance, would fall under this heading. The second approach would be to consider the production of transport and telecommunications means (e.g., shipbuilding, car production, and ICT production), which in some regions forms an important part of the regional economy.

The third approach is to analyze the impacts of regional connectivity–accessibility on regions in terms of their economic performance. This third approach is probably both the most intriguing and the most important one, although, arguably, also the most disputed. The key problem is that the impact of improved connectivity and/or accessibility on economic performance is ultimately much harder to quantify. Yet, the implications are crucial for the fundamental debate on economic efficiency (competitiveness) and regional equality (cohesion).

The conventional wisdom of neoclassical economics suggests that the improved provision of transport within a region will reduce overall production costs by reducing the transport cost that firms need to pay when delivering their products to the markets. In other words, better transport provision within the region increases the productivity of its firms. The second assumption of neoclassical economics relates to interregional connectivity–accessibility. According to it, improved transport links between regions will encourage trade between them, in turn promoting increased specialization (division of labor) between regions, with resulting efficiency gains shared by all involved regions. Third, the neoclassical equilibrium model assumes that if transport cost is reduced to zero or close to zero, distance between producers and their markets will cease to matter and firms will be free to locate anywhere, leading to a more even distribution of economic activities across space.

However, the real world is more complex than neoclassical economic models suggest. In fact, some empirical observations run against the neoclassical assumptions. Many economic activities continue to agglomerate in cities, often leading to highly polarized spatial structures. Meanwhile, the fruits of interregional trade are seldom evenly shared.

Finally, the assumption that increased levels of transport provision improve regional economic performance is problematic and cannot be taken for granted. Certainly, some evidence may suggest a degree of correlation between the two phenomena. Economic growth in many regions is often accompanied by improving transport infrastructures. However, the existence of the actual causal link between regional development and regional

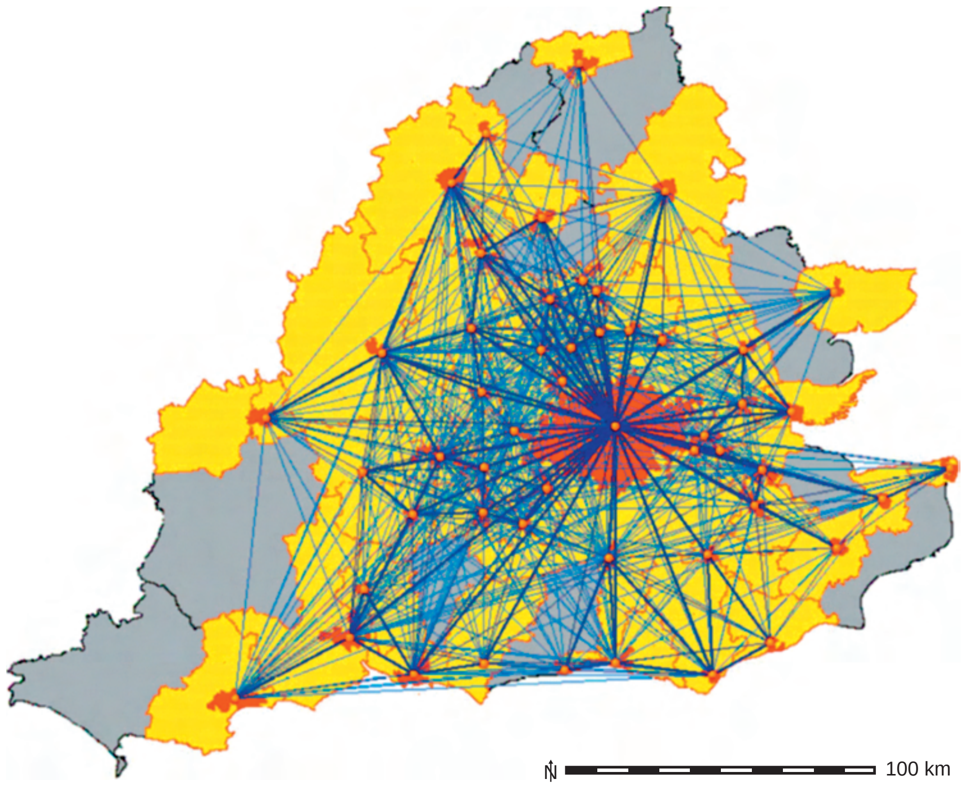

Figure 9 Southeast England mega-city region: Commuting flows, 2001. Reproduced with permission of Earthscan publications from Hall, P., Pain, K. and Green, N. (2006). Anatomy of the polycentric metropolis: Eight mega-city regions in overview. In Hall, P. & Pain, K. (eds.) *The Polycentric Metropolis: Learning from Mega-City Regions in Europe*, pp 19 52. London: Earthscan. [Fig. 2.10a, page 38].

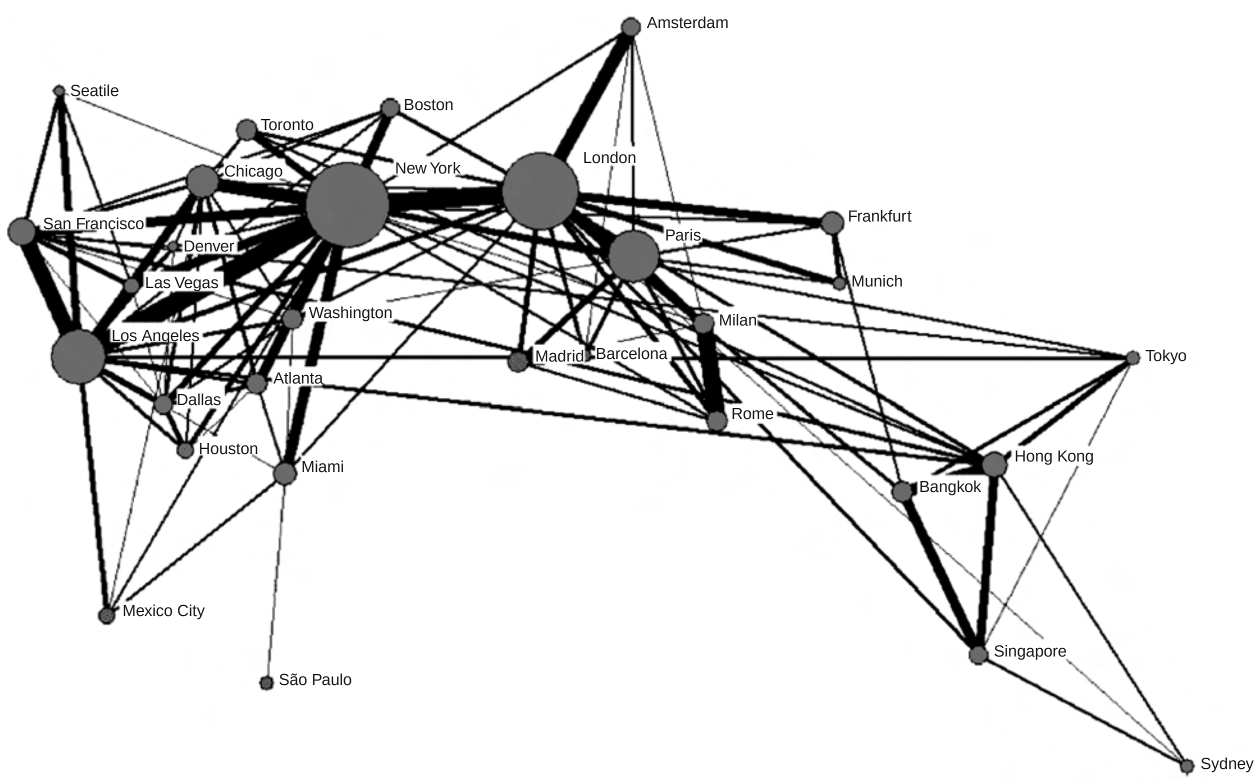

Figure 10 Air connectivity of the most important cities in the world city network (total number of passengers, January August 2001). Source: Derudder, B. and Witlox, F. (2005). An appraisal of the use of airline data in assessing the world city network: A research note on data. *Urban Studies* 42(13), 2371 2388. [Fig. 1, page 2384]. Reprinted by permission of Taylor and Francis Ltd (http://informaworld.com).

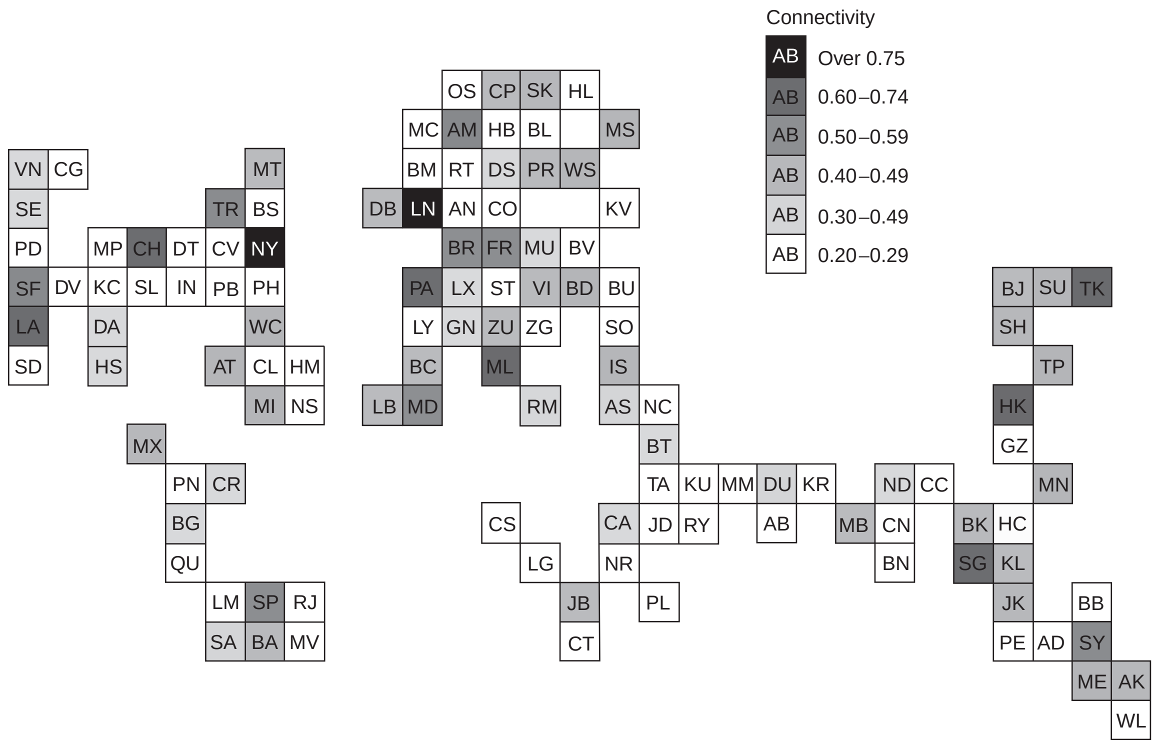

Figure 11 Global network connectivity: the most connected world cities, advance producer service connections. The world city archipelago defined by network connectivity. This cartogram places cities in their approximate relative geographical positions. The codes for cities are: AB Abu Dubai; AD Adelaide; AK Auckland; AM Amsterdam; AS Athens; AT Atlanta; AN Antwerp; BA Buenos Aires; BB Brisbane; BC Barcelona; BD Budapest; BG Bogota; BJ Beijing; BK Bangkok; BL Berlin; BM Birmingham; BN Bangalore; BR Brussels; BS Boston; BT Beirut; BU Bucharest; BV Bratislava; CA Cairo; CC Calcutta; CG Calgary; CH Chicago; CL Charlotte; CN Chennai; CO Cologne; CP Copenhagen; CR Caracas; CS Casablanca; CT Cape Town; CV Cleveland; DA Dallas; DB Dublin; DS Dusseldorf; DT Detroit; DU Dubai; DV Denver; FR Frankfurt; GN Geneva; GZ Guangzhou; HB Hamburg; HC Ho Chi Minh City; HK Hong Kong; HL Helsinki; HM Hamilton (Bermuda); HS Houston; IN Indianapolis; IS Istanbul; JB Johannesburg; JD Jeddah; JK Jakarta; KC Kansas City; KL Kuala Lumpur; KR Karachi; KU Kuwait; KV Kiev; LA Los Angeles; LB Lisbon; LG Lagos; LM Lima; LN London; LX Luxembourg; LY Lyons; MB Mumbai; MC Manchester; MD Madrid; ME Melbourne; MI Miami; ML Milan; MM Manama; MN Manila; MP Minneapolis; MS Moscow; MT Montreal; MU Munich; MV Montevideo; MX Mexico City; NC Nicosia; ND New Delhi; NR Nairobi; NS Nassau; NY New York; OS Oslo; PA Paris; PB Pittsburgh; PD Portland; PE Perth; PH Philadelphia; PL Port Louis; PN Panama City; PR Prague; QU Quito; RJ Rio de Janeiro; RM Rome; RT Rotterdam; RY Riyadh; SA Santiago; SD San Diego; SE Seattle; SF San Francisco; SG Singapore; SH Shanghai; SK Stockholm; SL St Louis; SO Sofia; SP Sao Paulo; ST Stuttgart; SU Seoul; SY Sydney; TA Tel Aviv; TP Taipei; TR Toronto; VI Vienna; VN Vancouver; WC Washington DC; WL Wellington; WS Warsaw; ZG Zagreb; ZU Zurich
Source: Courtesy of Peter Taylor. See also: Taylor, P., Walker, D., Catalano, G. and Hoyler, M. (2002). Diversity and power in the world city network. *Cities* 19(4), 231 241. [Fig. 2, page 235]; Taylor, P. J. (2004). *World City Network: A Global Urban Analysis*. London: Routledge. [Fig. 4.2, page 73]. Courtesy of Peter Taylor.

connectivity–accessibility is hard to prove. The importance, the nature, and the exact direction of the causal link between the two phenomena therefore remains disputed. Indeed, as we have seen in the case of our imaginary island region, good transport infrastructure may be as much the cause as it may be the consequence of economic development, perhaps being part of a wider circular and cumulative causation process of uneven development.

Attempts to overcome the problems of the periphery through its greater integration with the core area may not always yield the expected results. Various studies have shown that the improvement in regional connectivity–accessibility might be a necessary, but not a sufficient condition for economic development. On the other hand, there are examples of regions that have experienced dynamic economic growth despite relatively low levels of connectivity–accessibility (e.g., Ireland). In other cases, improved connectivity–accessibility (and thus increased integration with more powerful economic centers) may in fact result in the further erosion of the economic structure of peripheral regions (e.g., due to increased competition).

With the advent of the so called information revolution, much has been written about the ability of information and communication technology (ICT) to transform the landscape of regional economies. Many authors believe that with the emergence of the so called postindustrial information society, or knowledge based economy, the problems of uneven development created under industrial capitalism can be overcome. This

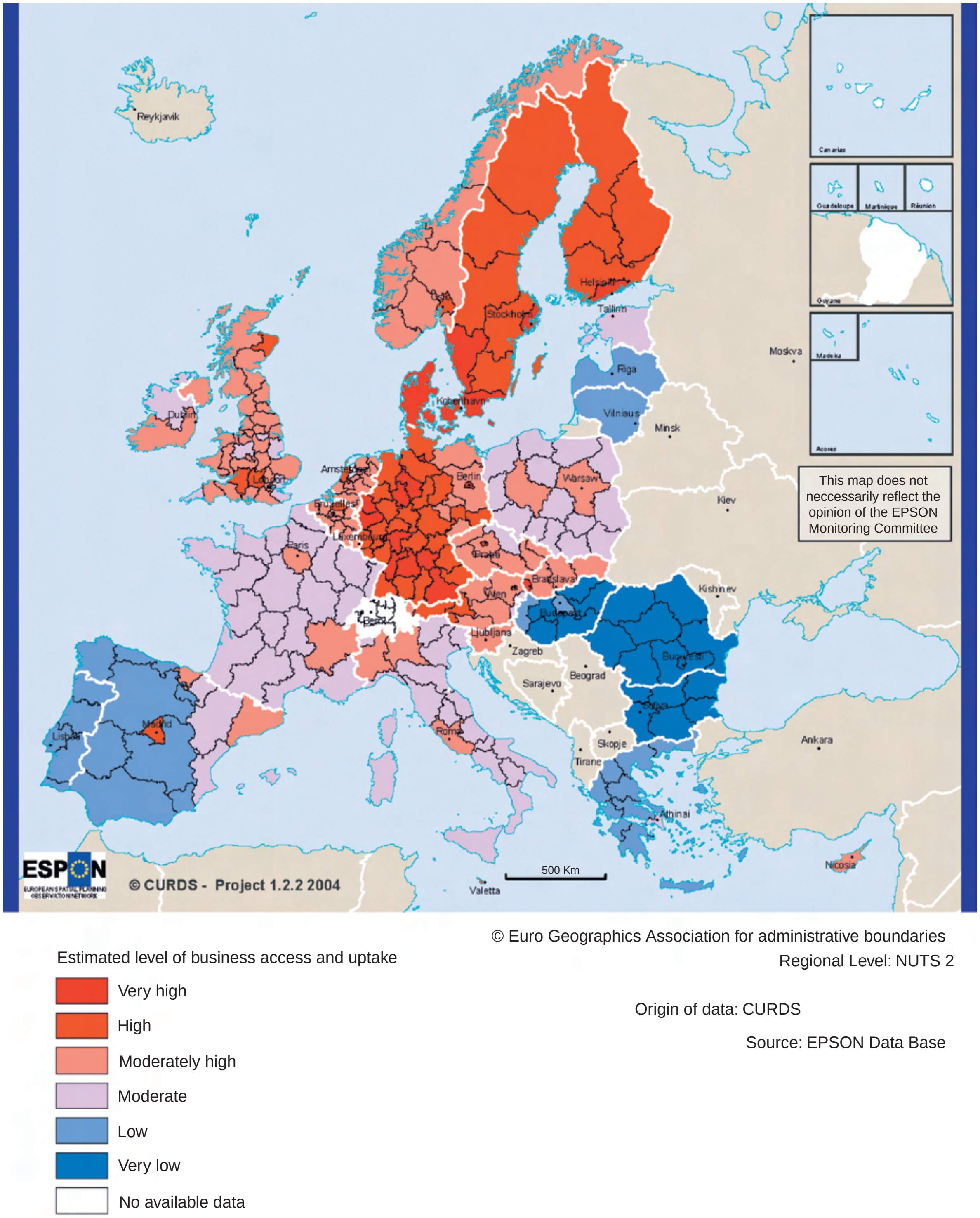

Figure 12 A typology of estimated levels of business telecommunications access and uptake in Europe. Source: ESPON and CURDS (2004). *ESPON Project 1.2.2 Telecommunication Services and Networks: Territorial Trends and Basic Supply of Infrastructure for Territorial Cohesion*. Luxembourg: ESPON. http://www.espon.eu/mmp/online/website/content/projects/259/653/file 1218/2.ir 1.2.2.pdf (accessed on January 2008). [Fig p 8].

expectation is based on the assumption that (1) in the knowledge based economy, information and knowledge replace tangible goods and materials as a dynamo for economic growth and (2) information, knowledge, and financial capital can be moved around the globe in stantaneously without constraints, courtesy of ICT. Op timistic voices would therefore argue that even the most disadvantaged regions (by their geographical location, poor transport infrastructure, or preexisting economic structure) have a fair chance to succeed in the global knowledge economy.

However, more cautious (and perhaps more realistic) approaches recognize that geographical location along side the preexisting regional economic structures still matters and that tangible goods and physical movements continue to play an important role in contemporary economies. It appears that ICTS facilitated (virtual) movements do not replace physical movements and the associated need for transport, but often complement and sometimes even encourage such a need.

Further to this, it could be argued that despite their potential to overcome the friction of distance, ICT in frastructures themselves are very unevenly distributed in space. Global cities, economic hot spots, and prosperous regions are usually served first, attracting the most ad vanced telecommunications networks and services. Meanwhile, many geographically and economically pe ripheral areas that would benefit the most from the po tential advantages of ICTs are often left behind. This process of geographical differentiation is likely to ac celerate further with the continuing privatization and liberalization of telecommunications and the retreat from the universal service provision previously imposed by nation states. Simply put, private telecom operators are likely to cherry pick the most profitable market segments (both geographically and socially) while bypassing less profitable social groups, localities, and whole regions. In other words, ICT infrastructure is being overlaid on the already very uneven social and spatial landscapes.

What is important to recognize, however, is that ad vances in transport and telecommunications are also par taking in the emergence of new, complex patterns of interregional and international divisions of labor in both manufacturing and services, underpinned by global com modity chains, expanding circuits of capital and value networks. Thus, instead of seeing the 'end of geography', we are perhaps witnessing a process of uneven time–space compression, which is likely to produce new economic geographies with complex landscapes of regional winners and losers and differentiated impacts on people and places.

See also: Core-Periphery Models; Digital Divide; Economic Geography; Information Technology; Internet, Economic Geography; Knowledge Economy; Mobility; Polycentricity; Spatial Division of Labor; Telecommunications; Transport and Accessibility; Transport Geography; Uneven Development.

Further Reading

Brotchie, J., Batty, M., Hall, P. and Newton, P. (eds.) (1991). *Cities of the 21st Century: New Technologies and Spatial Systems*. Melbourne: Longman Cheshire.

Castells, M. (1989). *The Informational City: Information Technology, Economic Restructuring and the Urban Regional Process*. Oxford: Blackwell.

Cornford, J., Gillespie, A. and Richardson, R. (2000). Regional development in the information society. In Ducatel, K., Webster, J. & Herrmann, W. (eds.) *The Information Society in Europe: Work and Life in an Age of Globalisation*, pp 21 44. New York: Rowman & Littlefield Publishers.

Derudder, B. and Witlox, F. (2005). An appraisal of the use of airline data in assessing the world city network: A research note on data. *Urban Studies* 42(13), 2371 2388.

Dicken, P. (2003). *Global Shift: Reshaping the Global Economic Map in the 21st Century* (4th edn.). London: Sage.

ESPON and CURDS (2004). *ESPON Project 1.2.2 Telecommunication Services and Networks: Territorial Trends and Basic Supply of Infrastructure for Territorial Cohesion*. Luxembourg: ESPON. http://www.espon.eu/mmp/online/website/content/projects/259/653/file 1218/2.ir 1.2.2.pdf (accessed on January 2008).

ESPON and S&W (2007). *Update of Selected Potential Accessibility Indicators (final report)*. Luxembourg: ESPON. http://www.espon.eu/mmp/online/website/content/projects/947/1297/file 2724/espon accessibility update 2006 fr 070207.pdf (ac cessed on January 2008).

Frost, M. E. and Spence, N. A. (1995). The rediscovery of accessibility and economic potential: The critical issue of self potential. *Environment and Planning A* 27, 1833 1848.

Gillespie, A. (1991). Advanced communications networks, territorial integration and local development. In Camagni, R. (ed.) *Innovation Networks: Spatial Perspective*, pp 214 229. London: Belhaven Press.

Gillespie, A., Richardson, R. and Cornford, J. (2001). Regional development and the new economy. *European Investment Bank (EIB) Papers* 6(1), 109 132.

Graham, S. (1999). Global grids of glass: On global cities, telecommunications and planetary urban networks. *Urban Studies* 36(5 6), 929 949.

Hall, P. and Pain, K. (eds.) (2006). *The Polycentric Metropolis: Learning from Mega City Regions in Europe*. London: Earthscan.

Hall, P., Pain, K. and Green, N. (2006). Anatomy of the polycentric metropolis: Eight mega city regions in overview. In Hall, P. & Pain, K. (eds.) *The Polycentric Metropolis: Learning from Mega City Regions in Europe*, pp 19 52. London: Earthscan.

Keeble, D., Owens, P. L. and Thompson, C. (1982). Regional accessibility and economic potential in the European Community. *Regional Studies* 16(6), 419 432.

Quinet, E. and Vickermann, R. (2004). *Principles of Transport Economics*. Cheltenham: Edward Elgar.

Rodrigue, J. P., Comtois, C. and Slack, B. (2006). *The Geography of Transport Systems*. London: Routledge.

Taylor, P. J. (2004). *World City Network: A Global Urban Analysis*. London: Routledge.

Taylor, P., Walker, D., Catalano, G. and Hoyler, M. (2002). Diversity and power in the world city network. *Cities* 19(4), 231 241.

Vickerman, R., Spiekermann, K. and Wegener, M. (1999). Accessibility and economic development in Europe. *Regional Studies* 33(1), 1 15.

Relevant Websites

http://www.ncl.ac.uk
Centre for Urban and Regional Development Studies (CURDS), University of Newcastle upon Tyne.

http://www.espon.eu
European Spatial Planning Observation Network (ESPON).

http://www.lboro.ac.uk
Globalization and World Cities Study Group and Network (GaWC), University of Loughborough.

Regional Development and Noneconomic Factors

J. O. Bærenholdt, Roskilde University, Roskilde, Denmark

What Is Noneconomic in Regional Development?

Regional development is the label of the efforts to develop certain areas of a country, with development usually understood in the socioeconomic sense. Regional development is thus not only measured in incomes, the number of jobs, and demographic trends in a certain area, but it can also point to the more general dynamics such as innovation and creativity in the region in focus. The definition of the region in question has not always been explicit, but when it becomes explicit, it becomes clear how regional development is connected to the making of territorial states. In the Keynesian welfare state, regional development was associated with regional policies aiming to overcome uneven development between different regions of the territorial state. This means that regional development was, and in some contexts still is, seen as a marker for the effort to overcome imbalances and unevenness within the territory of a state. In this more precise sense, regions are defined by being part of whole in relation to the state, and regional development is the agenda of regional policies in order to deal with a 'regional problem'. In the more Schumpeterian regimes, regional development has increasingly been seen as a process of innovation and entrepreneurship, aiming at growth in any region more than the explicit contribution to the solution of the particular 'regional problem'. Here, regional development is often associated with the growth and innovation of private business in a particular region, which in turn is rather loosely defined as the 'container' of socioeconomic activity.

Regional development can be associated with earlier traditions of regional geography, where regions were seen as the founding objects of geography and the main effort was to describe (i.e., to construct) regions as entities, which included noneconomic factors. Second, radical geography from the 1970s associated regional development with uneven development, where the features of region were often seen as economic and highly structured if not even materially determined. This approach was followed in the 1980s by the more discrete understandings of regional development as processes of restructuring and changes in the spatial divisions of labor. Third, economic geography, with its main focus on the performance of firms, increasingly took over the field in the 1980s and especially the 1990s, also taking noneconomic factors into considerations in explaining regional development. The progress in the sense of more specialized studies of firm development and firm networks, however, went along with decreased interest in regional development in the broader sense of the life and working conditions of the inhabitants in an area. Meanwhile, economic geography's increasing focus on noneconomic factors opens new possibilities for understanding the broader complexities of the field of regional development.

However, a general problem with much of the use of noneconomic factors as explanation to regional economic development is that these noneconomic factors are not analyzed in more detail. The most pertinent problem in this regard is the tendency to automatically make any noneconomic factor the attribute and generalized feature of the region in question. Another problem is the tendency to approach noneconomic factors as completely external, sort of outside resources, to the dynamics of regional development under study. There are, however, also human geographers that have contributed to the understanding of the making of regions in social, political, and cultural senses, where regional development becomes a field of place and regional identity, regionalism and regionalization, and regional administration and governance, though these contributions do not always take regional development in the socioeconomic sense into consideration.

At the same time, even economic approaches end up in pointing to the noneconomic factors that get into more or less black boxes, beyond further explanation or analysis. This raises the question whether or not the distinction between economic and noneconomic can be sustained. In recent studies in cultural economies, a new settlement is suggested, one that goes beyond this distinction. It is also a question to what extent these are 'factors' external to the processes of regional development. However, with these reservations in mind, the following sections introduce and summarize a selection of the concepts with which human geographers have tended to explain and understand regional development.

Common to these noneconomic factors are that they focus on the cultural, social, and material forms of organization of society and business, which are seen as historically produced (or constructed) through the practices of people. These are, in other words, examples of how contemporary social sciences understand the making of societies, which is an interdisciplinary effort, even beyond the many facets of human geography. Space, spatiality, and spatial organization is an integrated dimension in these processes, and therefore the geographies of regional development are much more than the external environments and 'stages', which geography all to often is limited to deal with. Regional development itself is seen as in an integrated aspect in the making and transformations of production, social relations, and soci eties. In many countries, including most of the European Union (EU) system, this understanding is increasingly embedded in social and political discourses. Regional development is often translated into the term of spatial development or spatial change of society, and as such connected with terms such as 'social inclusion' that in dicate the intrinsic role of regional (and urban) devel opment in social reproduction. This way, noneconomic does not mean 'not at all economic', but it does mean hybrids that are 'much more than only economic'.

Culture and Modes of Life in Regional Development

In traditional regional geography, dominating human geography in the first half of the twentieth century, re gions were often interpreted as entities of nature and certain modes of life, most famously in the works by Paul Vidal de la Blache on *genre de vie*. Regions were attributed and associated with specific cultures, as whole ways of life, most often defined by their, then, rather dominant de pendence on primary sector industries, including, for example, the various agricultural systems, as adapted to environmental possibilities. While some of this descriptive and regionalist, holistic tradition survived in educational practices for decades, it was not until the 1980s that human geography saw new approaches to culture and modes of life in relation to regional development (while there was also a cultural tradition in landscape geography in between). This new cultural turn took various routes in different contexts. One route was a new interest in con sumption, not only as part of social reproduction and urban development, but also as a symbolic practice, with new studies in economies of signs, lifestyles, and con sumption. Much of this interest was especially pre dominant in the Anglo American discourse, alongside the influence from French postmodernism. This route, con centrating on culture as product in various ways, has contributed significantly to contemporary studies in cul tural industries and the experience economy, within an increasingly interdisciplinary field across human geog raphy, cultural studies, business economics and marketing, communication, and performance studies. Quite another route was the increasing interest in modes of life and the social practices of everyday life, also nourished with human geography's long acquaintance with anthropology. This route became important in, among others, the Scandinavian countries, where the 'cultural turn' was as much a 'social turn', leading to analysis of different modes of life, often influenced by everyday life and critical theories from the European continent.

A parallel route was taken in Gramscian traditions in Southern Europe, not to mention many so called 'inter national development studies', often with analytical em phasis on the informal sector and household economies. Though there were affiliations with humanistic geog raphy, these approaches were as much interdisciplinary, among others with ethnology, sociology, anthropology, and cultural studies, in analyzing modes of life (or life forms) and social practices. They contributed to regional development research by studies ranging from place, locality development, migration, and urban life to in dustrial change, settlement structure, and resource management. While such approaches cannot contribute much to sweeping theories of regional development *per se*, their quality lies in their contextual studies of how people live and contribute to regional development by their everyday practices and their involvement in pro cesses of economic restructuring and demographic change. It remains then often to be a problem that such coping practices are not ascribed explanatory power to an extent that it allows for theoretical development in the field of regional development. Instead, we see the reading and use of certain fashionable concepts, such as social capital or the creative class that can be too generalizing and stereotyping attempts of theoretical interpretation.

Social Capital and Embeddedness in Regional Development

During the 1990s, research in regional development in creasingly became part of cross disciplinary debates across economic geography, economic sociology, eco nomic anthropology, political economy, political science, and others. These are debates that can be taken as indi cations of a more general 'spatial turn' in the social sci ences that helped to introduce more advanced discussions on the relation between social relations and space. However, in the field of regional development, with its penchants to economics, social relations and spatiality were still often kept at a distance, since the 'social' was assigned the role of an external explanation of regional development, as a spatial pattern. In the fol lowing section, four different scholastic groupings are discussed.

First, economic geographers, with the help of neigh boring fields, since the 1980s had formed an expanding discourse on post Fordism, flexible specialization, local ized learning, and innovation systems, with some affili ation with neo Schumpeterian innovation theory and institutional economics. Part of this tradition departed from neoclassical economics and focused on concepts such as transaction costs and clusters. But the trend was generally to question the explanatory power of economic factors. From studies of what used to be Marshallian industrial districts, and was now found in areas such as Silicon Valley and the Third Italy, concepts like untraded interdependencies and localized learning emerged. There was something more in the air than 'pure' economic interaction and trade. Rather than the direct economic interchanges between firms, where co location and agglomeration may be an advantage, it was the social networks, social capital, and embeddedness among employees that were assigned explanatory power to successes in regional development. Little ethnographic evidence of these relations was presented. Factors in focus were, for example, the importance of the circulation of tacit knowledge among skilled workers in the manufacturing of quality furniture or the common identification among people with their community or region. This way, the region could be seen as a nexus of localized features, responsible for regional development.

Alongside economic geography approaches, a second group of broader discussions of the concept of social capital emerged, initially with studies in Italy (Putnam and Italian associates), and of embeddedness with a background in the traditions of economic sociology (Granovetter) and economic anthropology (Karl Polanyi). In contrast to economic geography's focus on firms and businesses, this discourse had a wider approach to civic relations crucial to socioeconomic growth. It was shown how economic activities prospered if they were territorially embedded in areas with strong traditions of community building and voluntary (and political and religious) work. Trust, norms, and networks became essential noneconomic factors explaining success in regional development, and of regional government. While much of the surveys presented concentrated on the forms of networks associated with the identity or culture of communities, discussions increasingly addressed the issues of the looser forms of social capital and embeddedness, called bridging social capital and weak ties. Only some of these discussions also took into consideration the more analytical and critical approach to social capital suggested by Pierre Bourdieu. All of this pointed to the problems in assigning too much normative value to bonding social capital, equivalent to strong ties, that might well exclude others and hide the hierarchical or vertical power relations involved.

This is where the third grouping of approaches took their point of departure in the first decade of the twenty first century, when pointing to the problematic assumptions implicit in the social capital and embeddedness tradition. In consequence, more emphasis was given to autonomy and to relations performed at a distance, while highlighting the lock in problems in 'too much' social capital and embeddedness. It was now especially human geographers, trained with a modern approach to spatiality, the social construction of space, and spatial practice who were able to deconstruct the implicit spatial assumptions in the previous work. This led to new trends in economic geography with an interest in scrutinizing the more detailed spatial relations involved in regional development (see the last section). Thereby, it became clear that regional development is spatial not only in the form of regularized patterns and systems of settlement, location, transport, and communication, but also through the spatiality in (inside) the relations and practices, producing these patterns and systems. These new strands of regional development research were also nourished by the fact that they worked across earlier distinctions and included fields such as urban studies and the cultural economy.

Fourth, the inclination of regional development research to become economic geography or regional economics, focusing on the spatial practices and relations of firms, propelled the interest in broader approaches, looking for social innovation and new forms of community development. While the first and third group, mentioned above, tended to concentrate on business, the social capital tradition and the broader embeddedness tradition had the advantage of seeing regional development as a part of the formation of societies with class, gender, and ethnicity. This implies an interest in the more political questions of social justice, inclusion, redistribution, people's coping, empowerment, and governance, and when this interest is combined with the critics of the tacit spatial assumptions of earlier social capital studies, a new field of regional development studies can contribute to our understanding of the multiple processes of society production, which are always inherently spatial.

Networks in Regional Development

The concept of networks has become fashionable to include in every study, while the concept at the same time is used with many very different meanings. For example, the reciprocity involved in networks among humans belonged to the central defining features of social capital. But networks are also associated with the hardware of logistic systems in communication and transport. While some approaches celebrate the symmetries of 'soft' and 'hard' networks, the use of the concept has now become so chaotic (in parallel to, e.g., culture) that some clarification is needed.

The original meaning of the concept of networks was that of social networks built on reciprocity. The implication of this was that people through transactions with others build forms of interdependencies, which can become a resource for them, at another time. Especially in Bourdieu's approach to social capital, this understanding of networks, as connections through which other capitals can be mobilized, become clear. Therefore, social net works involve the power relations that are always part of people's coping practices. Here, Bourdieu's way of thinking allows us to take into consideration the dynamic struggles between people, also involved with other forms of capital. This makes apparent the problems in Putnam's assumption of causal links between civic participation, economic development, and well working political in stitutions. We cannot take for granted that the social capital 'of the region' is an independent variable. The spatiality of practices involved in regional development is much more complex.

This is not the least case because of the material networks, studied as commodity chains that include routes and flows of raw materials, products, goods, and information across the world. Meanwhile, there is also the whole field of governance networks, where networks though social and personalized, are associated with cer tain organizational bodies, just like networks between firms in economic geography. Actor network theory (ANT) then makes the point of symmetrical relevance of all of these networks, where it is the networks themselves that become actors. Likewise, relational approaches are celebrated across many fields and disciplines, but few of the uses of network in human geography relate to the earlier traditions of social networks analysis. An exemp tion to this generalization may be the use and study of networks in urban studies, across disciplines.

One of the key assumptions about networks is that they are enduring, in other words, a temporal assumption. This is, for example, the key feature attributed to net works as made of reciprocity, as studied in the anthro pological studies of gift exchange. Gifts and other interchanges are made to produce reciprocity over time following the formula, 'if I do something for you, I can expect that you will help one day, when I need your help'. It is this form that is then generalized in the Putnamian concept of social capital as the generalized reciprocity of civicness: "if I obey the rules and norms, I can expect that others will do the same." In line with Bourdieu, one can however argue that Putnam's understanding is based on a too rationalist 'economic man' like understanding of human practices. Notwithstanding discrepancies in ap proaches, many would agree that significant forms of stabilizations through institutionalization of social re lations are often crucial to regional development.

The emphasis on lasting relations can easily come into conflict with innovation and entrepreneurship, often nourished by more short term projects and team work, that is, exactly getting its creativity from knowing that the cooperation will not (have to) last. But bypass re lations and risky intersections, crucial to produce venues for regional development, may easily undermine social cohesion and solidarity, thus emphasizing the traditional distinction between individual and collective good (or freedom). Regional development therefore cannot escape the crucial questions of social justice and solidarity with who, how many, and how far away? Regional devolution attributing powers to new governance bodies at the re gional level, for example, may not help much if regional development is addressed without taking into consider ation distributional aspects.

In conclusion, the concept of networks has not only produced insights into the dynamics of regional devel opment, but has also contributed to confusion because of lack of concern of the differential aspects of the concept of networks: material networks channeling flows of re sources, traffic, and communication are very different from genuine social networks among people, reciprocally depending on each other – often at a distance – and then again from the governance networks involved in both distributional and innovation policies.

Power, Mobility, and Territoriality in Regional Development

Across the various, so called, noneconomic factors as sessed above, there is a cross cutting discussion of whe ther relations in proximity are more dynamic contributors to regional development than mobile or fluid relations, performed on the move and at a distance. The dynamic aspects not only are about more than entrepreneurship in the classical business sense, but also involve the roles of consumption, migration, and money flows within families and communities at a distance. Human geographers with a keen interest in spatiality and spatial practices contribute increasingly to advance understandings of the complex spatial relations involved in regional development. Regions become unbound, and regional development is readdressed in the more gener alized term of spatial development, for example, in the EU. A central distinction in regional policies takes place between those oriented to growth before equity in con trast to growth for equity.

With the acknowledgment of the constitutive role of mobility, places and regions can no longer be taken for granted. If territoriality is not the only basis for politics, this has profound consequences for our understanding of regional development and of regional policies. The agenda emerging is then one of fragile combinations of 'politics of propinquity' with 'politics of connectivity'. Much of these new understandings have been facilitated by, what has been termed, a cultural turn in economic geography, but there seems to miss a 'political turn' in studies of regional development. In one way, so called relational approaches seems to address power relations, without being explicit about this. With a point of de parture in Foucault's notion of power relations, which are productive and fluid, mobility can also be defined as a social practice of influence and control, not through boundaries, but through the movement of human beings, things, and information. This is not to say that practices and power relations of mobility are more relational than those of territoriality. Furthermore, mobility and territoriality also intersect in the complex spatiality of power relations involved in regional development.

International migration is one of the major issues, very important to regional development, which could be more aptly integrated in regional development studies, with an approach to power, politics, mobility, and territoriality. With the increasing role of labor migration and remittances, and also diverse forms of government, governance, and governmentality involved in attempts of regulation of migration, it becomes very clear how the economic, the cultural, and the material intersect in the practices that make regional development. Meanwhile, the diverse forms of networks, mobility, and territoriality involved in migration points to the problems in how the 'normal' understanding of associate territoriality with bonding networks and mobility with bridging networks.

Regional development is increasingly shaped by 'mobile bonding' practices of 'communities' (whether formed by ethnicity, religious or other beliefs, and political projects or the like) that are not based on territoriality but are mobile associations, dynamic because of their mobility. Their relevance to regional development is increasing, but their lack of territorial commitment to a specific group of inhabitants also raises the question of whose regional development they contribute to. While societies were traditionally assumed to hold together via a common, local, national, or regional identification 'territorial bridging' is meant to overcome and potentially level out conflict and injustice among people living in the same continent, country, region, or municipality, but then only due to the fact that people inhabit the same territory. Thus, territorial organization can be a – and is so the by far most effective – regime of solidarity and redistribution among citizens, that is, a society. The normative markers involved do not need to be more than 'universal' ideals of civil and human rights, democracy, and the like. Physical proximity is no condition for territorial bridging practices; it can work in various scales. The continued relevance in thinking about regional development resides in the fact that people still construct societies through citizenship of territorial states and other territorial units that define their mutual interdependence as inhabitants of the same territory, thus defined by the fact that people use territorial relations to establish and regulate their regimes of redistribution. Regional development, and the question 'for who?', reminds us that the regional question is still a question of how societies are made.

Noneconomic factors in regional development cannot be reduced to only cultural, social, or material factors that can be added to the analysis, when needed. Such an understanding rests on the assumption that regional development 'in the last instance' rests on pure economic factors. But is there anything that is purely economic? How economic relations are also cultural, social, and material especially becomes apparent when regional development is understood as the result of practices, politics, and power relations. The implication is not to confuse networks of economic transactions, social reciprocity, material flows, and governance with one another, but to acknowledge their interdependence and their common association with practices and power.

See also: Actor-Network Theory/Network Geographies; Cultural Turn; Culture; Development I; Development II; Economic Geography; Governance; Region; Regional Geography I; Regional Science; Regionalism.

Further Reading

Amin, A. (2004). Regions unbound: Towards a new politics of place. *Geografiska Annaler* 86B, 33 44.

Amin, A. and Thrift, N. (eds.) (2004). *The Blackwell Cultural Economy Reader*. Oxford: Blackwell.

Bærenholdt, J. O. and Aarsæther, N. (2002). Coping strategies, social capital and space. *European Urban and Regional Studies* 9, 151 165.

Bærenholdt, J. O. and Haldrup, M. (2003). Economy culture relations and the geographies of regional development. In Ohman, J. & Simonsen, K. (eds.) *Voices from the North, New Trends in Nordic Human Geography*, pp 49 66. Aldershot: Ashgate.

Bærenholdt, J. O. and Haldrup, M. (2006). Mobile networks and place making in cultural tourism, staging Viking ships and rock music in Roskilde. *European Urban and Regional Studies* 13, 209 224.

Bathelt, H. (2006). Geographies of production: Growth regimes in spatial perspective 3 toward a relational view of economic action and policy. *Progress in Human Geography* 30, 223 236.

Boschma, R. A. (2005). Proximity and innovation: A critical assessment. *Regional Studies* 39, 61 74.

Dicken, P., Kelly, P. F., Olds, K. and Yeung, H. W. C. (2001). Chains and networks, territory and scale: Towards a relational framework for analysing the global economy. *Global Networks* 1, 89 112.

Florida, R. (2002). *The Rise of the Creative Class: And How It's Transforming Work, Leisure and Everyday Life*. New York: Basic Books.

Grabher, G. (2006). Trading routes, bypasses, and risky intersections: Mapping the travels of 'networks' between economic sociology and economic geography. *Progress in Human Geography* 30, 163 189.

Hadjimichalis, C. (2006). Non economic factors in economic geography and in 'New Regionalism': A sympathetic critique. *International Journal of Urban and Regional Research* 30, 690 704.

Hudson, R. (2006). Regional devolution and regional economic success: Myths and illusions about power. *Geografiska Annaler* 88B, 159 171.

Illeris, S. (1992). The Herning Ikast textile industry: An industrial district in West Jutland. *Entrepreneurship and Regional Development* 4, 73 84.

Malmberg, A. and Maskell, P. (2002). The elusive concept of localization economies: Towards a knowledge based theory of spatial clustering. *Environment and Planning A* 34, 429 449.

Massey, D. (1979). In what sense a regional problem? *Regional Studies* 13, 233 243.

Moulaert, F. and Nussbaumer, J. (2005). The social region, beyond the territorial dynamics of the learning economy. *European Urban and Regional Studies* 12, 45 64.

Oinas, P. and Lagendijk, A. (eds.) (2005). *Proximity, Distance and Diversity: Issues on Economic Interaction and Local Development*. Aldershot: Ashgate.

Putnam, R. D. (1993). *Making Democracy Work, Civic Tradition in Modern Italy*. Princeton, NJ: Princeton University Press.

Storper, M. (1995). The resurgence of regional economies, ten years later: The region as a nexus of untraded dependencies. *European Urban and Regional Studies* 2, 191 221.

Taylor, M. and Leonard, S. (eds.) (2002). *Embedded Enterprise and Social Capital, International Perspectives*. Aldershot: Ashgate.

Woolcock, M. (1998). Social capital and economic development: Toward a theoretical synthesis and policy framework. *Theory and Society* 27, 151 208.

Regional Development and Technology

J. R. L. Howells, University of Manchester, Manchester, UK

Glossary

Agglomeration Economies These are associated with the benefits, savings, and cost reductions associated with the clustering of activities through economic scale and scope effects. The concept of agglomeration economies therefore relates to savings or benefits derived from the clustering of activities external to the firm and are consequently part of external economies.

Cumulative Causation A self-reinforcing process during which impulses (perturbations) activate positive or negative feedback leading to further growth (or decline), or other kinds of change within the same direction as the original impulse.

Innovation It is the generation, development, and introduction of a new idea in the form of a new product or service or an improvement in organization or process. Thus, innovation is associated with the process of converting knowledge and ideas into better ways of doing business or into new or improved products and services that are valued by society. The innovation process incorporates research and development, commercialization, and technology diffusion.

Innovation Clusters Geographically proximate concentrations of interconnected firms, including specialized suppliers and related service providers and innovation intermediaries, as well as other key institutions, such as universities and trade associations, which cooperate (and compete) in a particular set of technologies linked by a set of commonalities and complementarities. The geographic scale of these clusters ranges from a city, subregion, region, or federal state through to interregional or international scales of collaboration and networking.

Learning Regions A concept associated with regional innovation strategies that seek to develop models and policy frameworks for partnership-based learning strategies within regions.

Multiplier Effects These effects are associated with a numerical coefficient which relates the change of a component of aggregate demand to a consequent change in employment or income.

Regional Development It involves complex economic, technical, and social changes, often associated with economic growth, but also includes improvements in wider administrative institutional, organizational, and political frameworks.

Regional System of Innovation It is a localized network of actors and institutions in the public and private sectors whose activities and interactions generate, import, modify, and diffuse new technologies.

Technology It refers to the body of know-how (in both tacit and codified forms of knowledge) relating to the means and methods of producing goods and services. Modern technology is increasingly science based, but also includes methods of organization and practice as well as physical technique.

Introduction: Technology and Regional Economic Growth

Technology is core to regional development and change. It refers to the body of know how (in both tacit and codified forms of knowledge) relating to the means and methods of producing goods and services. Technology was originally defined as being innovation associated with tangible artefacts relating to physical products and processes. However, over time, the meaning of the word 'technology' has come to include less tangible innovation processes relating to improved methods of organization and practice, that is, described as disembodied technological change. Technology is the principal source of change and growth within national, regional, and local economies. This link between technology and economic growth has long been acknowledged. From Marshall through to Solow and Kuznets, there has been recognition that, directly and indirectly, technology changes economic activity and economic activity changes technology in constant rounds of change. However, how knowledge and technological change is seen to effect such change has shifted over time. Thus, early neoclassical approaches viewed knowledge and technology as being completely exogenous to the system and that the same technological opportunities were equally available to individuals and firms in all places. In turn, this was linked to viewing technology as being a public good implying that in the long run the rate of technological progress would be the same everywhere. In turn, growth paths of different countries or regions would also converge over the long term.

By contrast, newer growth theoretic models have emerged based on endogenous and neo Schumpeterian interpretations of economic growth. Thus, endogenous growth models have sought to build into their models the endogenous component of technological progress as an integral part of the theory of economic growth; namely, technological progress is seen as arising out of directed actions and investments by people through the allocation of key resources linked to human capital, and, more particularly, the amount of resources allocated to research. Neo Schumpeterian models have, in turn, introduced notions of monopolistic competition and the existence of intellectual property rights over new technology. Associated with both these two growth theoretic perspectives was that technology should be considered more like a private good rather than a public good and the acceptance that there could be increasing returns to knowledge generation. Under this neo Schumpeterian perspective in particular, not all countries or regions will be equally placed to generate and benefit from innovations and there will be strong tendencies for cumulativeness. Innovation will therefore be a strong disequilibrating factor in the processes of economic growth, giving rise to the pervasive differential growth rates between geographical areas. This neo Schumpeterian approach has echoes of earlier Marxist approaches toward technological change which stressed the disruptive nature of technology and its impact on social shaping and the shifting balance of power within new forms and structures of working.

What may be seen as hybrids of both neoclassical and new growth models have also emerged. However, although such models have acknowledged the private nature of technology in terms of patents and monopoly power, it still treated technology as being exogenous. Similarly, linear endogenous models exist which acknowledge that technology arises from the actions and investments of people, but treats technology as a public, nonrival good that is liable to geographical spillovers. However, there have been further synthetic, moderated developments of such new growth models, in particular in two spheres. First, there has been recognition that innovation and technological advance should include both endogenous and exogenous (to the system) elements. Second, there has been the realization that monopoly rents over technology remain incomplete and temporary. As such, technology has characteristics that become more like a public good over time and that knowledge and technology spillovers will occur (see later).

There is a clear and continuing link between technology and economic performance at a regional level. Studies using various proxies and indicators for technology and innovation have found a clear relationship between regional innovation, levels of prosperity, and economic growth. However, the correlation between innovative inputs and innovative outputs does not necessarily remain constant. Neither is research and development (R&D) nor patent activity necessarily linked to other economic performance measures, such as changes in productivity levels.

Significant differences in technological capacity still remain between regions in both developed and underdeveloped economies. As such, variations at a regional and subregional level of technological activity remain significant. Innovation, far from taking a uniform, rootless, and spaceless nature in this more globalized era, is influenced by its context and will remain highly uneven, distinctive, and differentiated in form. Innovations are therefore not evenly distributed over the whole economic system at random, but tend to concentrate in certain sectors and their surroundings.

Technology and Regions: Basic Phenomena, Processes, and Components

There are a number of basic phenomena, processes, and components that are crucial in our understanding of how technology and innovation influences, and is in turn shaped by, regional development. Thus, it has been revealed that areas close to existing successful innovative areas stand a much better chance of success in terms of performance and growth. By contrast, innovation poor regions surrounded by other innovation poor regions have been shown to do less well in terms of economic performance. What are some of the key elements (and concepts) in shaping this process. They can be grouped under three main dimensions, namely:

1. those associated with time;
2. those associated with space; and
3. those associated with flows.

Each of dimensions will be explored below.

Technology and Time: Cumulative Causation, Multipliers, and Evolutionary Change

Many of the effects of technology in regions in terms of economic growth and development are cumulative in nature, through the process of cumulative causation. Cumulative causation is a self reinforcing process during which impulses activate positive feedback leading to further growth, decline, or other kinds of change with the same direction as the original impulse. In turn, cumulative causation is closely associated with consequent multiplier effects where an initial change in spending or employment causes through demand mechanisms a further round of change in income or jobs. Thus, for example, new workers recruited by a firm due to growth from innovation and technological change create, through their spending in the local economy, additional jobs supported by these expenditures. The process of cumulative causation and economic multiplier effects is a key element in shaping regional development and competitiveness. In short, success breeds success; that is, 'virtuous circles' of expansion.

Equally, though, negative impulses can initiate 'vicious circles' of decline, which can be sustained and amplified by reverse patterns of cumulative causation and negative multiplier effects. Either way, it is unwise to suggest that regional development patterns get set for all time. Indeed, past patterns of technological success may make it difficult for firms and industries in a region to adapt to new demands and technologies and consequently they can suffer from problems of technological stasis and 'lock in'.

Technology, amplified by cumulative and multiplier effects, has a strong evolutionary change effect on regions that is far removed from the static equilibrium models of neoclassical economics. Technology is important in that it creates constant rounds of change and flux which are associated with path dependent processes (i.e., past events help shape events in the future), but also crucially help in creating variety and differential growth patterns. In Schumpeterian terms, innovation is central in creating these change processes and is part of the patterns of 'creative destruction', an acknowledgment that change involves decline as well as new growth and that such new growth is important in leading to decline and exit in other parts of the economy.

Technology and Space: Concentration, Agglomerations, Districts, and Clusters

The phenomenon of agglomerations and clusters is also central to the issue of how technology is shaped by, and impacts on, the development of regions. The benefits of agglomerations and clusters in terms of technological change and economic performance have long been recognized; although it was also accepted early on that there could also be significant disadvantages by locating in such districts. However, the debate over the nature and role of agglomeration economies and clusters is very much ongoing. Spatial concentration of firms, industries, and support infrastructures can help generate agglomeration economies which in turn can help foster and sustain innovation and technology in those regional and subregional centers. Such agglomerations may also be associated with cluster formation, which suggest closer network and linkage patterns and which may involve more specialist horizontal (same technology or industry) or vertical (combining backward and forward linkages through supplier and consumer networks) relationships. Whereas the longer established concept of agglomeration has tended to center more strictly on economic and industrial effects and benefits, the notion of clusters has tended to include wider social and institutional processes linked to trust, the role of mutual learning and informal networks. No consensus, however, has emerged on the size and nature of these latter, more informal set of capabilities. The question of "when do simple spatial concentrations of firms and technologies become more significant agglomerations or clusters?" still remains very open.

However, what is not contested is that such spatial phenomena do have a measurable impact on the size, performance, and rate of technological change within those areas. Firms and their associated technologies by being geographically concentrated and clustered do seem to benefit in significant and measurable ways. This is reflected, in turn, in the innovative and growth performance of such firms.

Technology and Spatial Flows: Knowledge Transfers, Diffusion, and Spillovers

Technology and innovation is also very much influenced by spatial flows of knowledge and specialized technology inputs and outputs. These flows can be related to a variety of mechanisms and processes. First, they involve the (formal) transfer of knowledge and technology between individuals within firms, universities, and other organizations involved in association with innovative activity. This is important in a regional and geographical context in that firms and organizations are becoming increasingly involved in distributed and networked forms of innovative and technological activities. No longer can firms solely rely on their own indigenous R&D and technical expertise and are entering in more collaborative arrangements in relation to research and technology. Second, and related to the above, flows of technology can be associated with diffusion. These diffusion processes may be related to knowledge which is key in the generation and developments of innovation. However, once developed, these new technologies, whether there are new products, services, processes, or organizational practices, are also diffused across space. Spatial diffusion of new technology is a critical determinant of regional economic competitiveness. The demand conditions of regions are important, influencing the more rapid spread and take up of new ideas and this is linked to a whole set of complex issues to do with openness, creativity, and risk sharing. The absorptive capacity of firms and organizations (i.e., the ability of firms to absorb new ideas, technologies, and practices from outside the firm) in a region is also central in influencing the spread and rate of diffusion of innovations. Above all, it has come to be acknowledged that the diffusion of new, disembodied organizational routines and practices associated with the innovation process are also critical in determining firm and regional performance.

Lastly, there are a set of largely unintended flows of knowledge and technology that are associated with geographical spillover effects. Geographical spillovers are an important element in influencing (through distance decay and other effects) the transmission of localized knowledge and technology which in turn influence the growth prospects of regions. The process whereby knowledge spills over and is appropriated by other firms has been largely assumed in traditional economic literature as an almost costless and frictionless process. Thus, knowledge was seen as a public good because it was seen as being impossible for its creator to prevent it being used by economic agents who do not pay anything in exchange for it.

There have been three ways in which researchers have sought to map these knowledge spillover processes and effects. Knowledge becomes generally available through various means, including publications, patents, and informal sharing, and this non excludability of knowledge is associated with such knowledge spillover effects. An early focus of such spillover studies was on examining the linkages associated with codified forms of knowledge and these studies confirmed the localization of knowledge transfers. Another approach has been to map the movement of people on the basis that knowledge moves with them. Such studies have similarly confirmed that the transfer of ideas is associated with mobility patterns between firms and that this is a further reflection of knowledge spillovers being geographically confined. A final stream of analysis has been based on the assumption that knowledge can be embodied in goods and that knowledge 'flows' can therefore be mapped, although there remain severe conceptual and methodological limitations of using this approach in relation to knowledge. However, regardless of what methodological approaches are used most studies have revealed that knowledge spillover effects can play an important and significant role in improving innovative performance (measured in terms of firm R&D and patent productivity levels or in output growth). As such, the transfer, absorption, and take up of the new technologies are crucial in shaping regional and subregional growth and development patterns.

Technology and the Regions: Regions as Learning Environments and Systems

The focus so far in the discussion has been largely on how technology influences regional development, but equally (as reflected in the earlier discussion concerning the role of agglomerations and clusters) regions also help shape and influence innovation and technological development. This has been associated with a conceptual shift away from seeing regions as being simply productive arenas toward the concept of regions representing localities where things are learned and knowledge shared. In particular, this work builds upon notion of the learning process within firms and their dependence on close interaction with their local environment. This concept can be traced back to some of the early writings on the benefits of agglomeration economies; however, instead it focuses on technology and innovation as a supply oriented, embodied processes in relation to production, instead it focused around issues associated with more demand led problem solving and learning activities. In this context innovation is undertaken to resolve specific technical problems experienced in producing goods and services and these advances in technology occur as a response to particular problems arising in existing centers of concentration.

These studies have stressed the highly localized nature of many forms of learning experience both in terms of what may be called 'psychological distance' and 'technological proximity', but geographical distance as well. 'Learning by doing' and 'learning by using' have been increasingly acknowledged as key components in successful innovation and are linked to direct and continuous physical proximity to the production process. The rise of the 'learning economy' and the 'learning region' where know how sharing (rather than knowing how to do things in isolation) has become more important in the innovation process, therefore heightening the importance of spatial proximity and the local relationships in technological and industrial performance.

Regions have therefore been seen as learning and problem solving environments, but also in a systemic way in relation to technology through the regional systems of innovation approach. Thus, regions are associated with a set of core elements (firms, their relationships, public sector organizations, and their institutional arrangements) which not only form a backdrop and framework for innovative and technological activities, but more specifically shape their whole set of relationships and their chances of success (performance). Regional systems of innovation are therefore associated with the development of these technological infrastructures based on collectively supplied, specific, industry relevant capabilities. As such, a regional system of innovation can be defined as a localized network of actors and institutions in the public and private sectors whose activities and interactions generate, import, modify, and diffuse new technologies. This is important in that the performance of regional economies cannot wholly be explained in terms of strategies and performances of firms, but they also crucially involve other types of economic actors and institutional arrangements that play a vital role in the generation and diffusion of knowledge and the development of new technologies.

Regions as systems of innovation can also be seen in an evolutionary sense as selection environments which can provide favorable conditions to innovation and technical change. In this context, new growth opportunities are shaped and constrained by the regional path dependency associated with the inheritance of regional structural characteristics from past knowledge accumulation and learning. Thus, on the one hand, good innovation environments make it more likely that innovations will be generated and succeeded, leading to a strong process of cumulative causation and positive feedback processes. On the other hand, though, certain system attributes (associated with the emergence and growth of institutional rigidities) may stagnate and ossify, leading to negative 'lock in' effects and thus negative feedback processes within the regional system.

The regional systems of innovative approach have been successful in highlighting the 'negatives'; namely, the lack of certain agents, resources, and competencies and the identification of system failures and obstacles within the innovation system. This is in comparison to agglom eration and cluster models which have tended to focus on highly successful, but fairly unique high technology dis tricts and regions (which in turn have been difficult, if not impossible, to copy and replicate elsewhere). In part, this respect of the 'negative', or what is lacking in a system, may be because there has been a focus on what elements there should be for a system to exist and be successful. It has therefore been used diagnostically to identify system failures and barriers which are hindering the overall in novative performances of regional economies.

However, problems remain with the concepts of learning regions and regional systems of innovation. Key issues remain; for example, "what attributes does a suc cessful local or regional innovation system or cluster possess?" Evidence, at least for successful regional in novation systems, suggests that they actually may be less prevalent in their full form than generally perceived. Arguably, therefore, fully functioning regional systems of innovation are relatively rare events.

Technology and Regions: Changing Perspectives

Over time the analysis of technology and its impact on regional development has changed. Thus, the concept of 'technology' associated with physical, embodied process of technological change have given way to the notion of 'innovation' which includes a much wider set of non technological activities, including organizational changes and more intangible forms of innovation, associated with, for example, design. However, changes in disembodied, nontechnological innovative processes, organizational arrangements, and markets still need to be given a greater weight in the analysis of regional development.

Equally, in the past, studies at the regional level have focused mainly on supply driven and producer defined models of technological change and innovation, and it has only been more recently that there has been growing attention toward more demand led notions of technolo gical change and the role of users in this whole process. The absorption, adaptation, and use of new technologies has long been a neglected theme within regional devel opment studies, even though it is likely to be a key factor in terms of long term growth and development.

Next, for an approach which espouses a dynamic and evolutionary perspective, the spatial study of innovation remains fundamentally rooted in the study of nodes or points and rather than in flows. This conceptual re striction means that in a geographical sense, it is much harder to map innovation processes (although changes in node patterns over time can provide a partial perspective on innovation dynamics). In summary, there is a need to further shift away from the focus on technology inputs, producers, and nodes in regional studies toward tech nology outputs, flows, and use. However to do this we need to develop better metrics and indicators at a re gional level to measure these processes and activities.

Lastly, and paradoxically, studies of regional develop ment and technology have tended toward a static rather than a dynamic analytical framework. Research has viewed regional development looking backward, moving from at best one time period to another but with little notion how one got from time period 'a' to 'b'; only that change has occurred. Similarly, innovation and technological pro cesses have been studied in a largely static way. There has strangely been very little work in trying to 'forward look' in the sense of developing concepts and models that try and predict future changes. This may be because they are seen as being too applied, although in truth higher level models and theories are exactly those concepts that have a strong (and successful) predictive element to them. This remains an important omission, as a much more thorough understanding of the relationships, communities, and networks in regional innovation systems need to be undertaken, if the process of innovation and technology is to be fully understood within such systems.

See also: Innovation; Knowledge Economy; Technology and Regional Development.

Further Reading

Boschma, R. A. and Frenken, K. (2006). Why is economic geography not an evolutionary science? Towards an evolutionary economic geography. *Journal of Economic Geography* 6, 273 302.

Cheshire, P. C. and Malecki, E. J. (2004). Growth, development, and innovation: A look backward and forward. *Papers Regulation in Science* 83, 249 267.

Gordon, N. R. and McCann, P. (2005). Innovation, agglomeration, and regional development. *Journal of Economic Geography* 5, 523 543.

Howells, J. (2005). Innovation and regional economic development: A matter of perspective? *Research Policy* 34, 1220 1234.

Malecki, E. J. (1997). *Technology and Economic Development: The Dynamics of Local, Regional and National Competitiveness*. Harlow: Addison Wesley Longman.

Martin, R. and Sunley, P. (1998). Slow convergence? The new endogenous growth theory and regional development. *Economic Geography* 74, 201 227.

Morgan, K. (1997). The learning region: Institutions, innovation and regional renewal. *Regulation Studies* 31, 491 503.

Ullman, E. L. (1958). Regional development and the geography of concentration. *Regional Science Association, Chapters and Proceedings* 4, 179 198.

Regional Development Models

M. Dunford, University of Sussex, Brighton, UK

Glossary

Circular and Cumulative Causation According to models of circular and cumulative causation, development and underdevelopment are self-reinforcing, while comparative development depends on the relative weight of unequalizing centripetal forces of attraction and suction and equalizing centrifugal forces of diffusion.

Development Development is often measured simply by calculating how much new wealth (gross domestic product) is created in a particular area in a particular year or how much income accrues to an area's inhabitants (gross national income). The creation of wealth is a means to an end (reduced poverty, inequality, and unemployment, and an increase in human capabilities defined as the freedom to achieve valuable beings and doings (themselves called functionings) which include good health, adequate nourishment, access to educational services, happiness, self-respect, and security).

Region The word region is most often used to refer to a delimited part of the surface of the Earth. In geography, regions are identified either on the grounds that the area as a whole shares certain common characteristics, or is in certain respects functionally interdependent or is an area over which economic, political, or administrative dominion or authority is exercised. Often considered as referring to a part of a nation, the word can also denote supranational territorial entities.

The Flying Geese Model A model developed to represent the role of a dominant growth center (Japan) at the head of a hierarchical group of East Asian economies that experienced rapid catch-up in the years after 1950. The idea was that the emerging economies in East Asia pursued one after another a sequence of industrial evolutions. The first involved a product cycle sequence (the import of modern manufactures, domestic production, export, and finally re-imports). The second involved an industrial sequence involving a movement from lower to higher value-added activities and a succession of industries (textiles, chemicals, iron and steel, motor vehicles, and electronic products). The third sequence was inter- rather than intra-national and involved the transfer of products and industries from countries that were more advanced to countries occupying lower positions in the hierarchy.

Unequal Exchange Models of unequal exchange involve the claim that, as wages are relatively low in economically under-developed countries, and rates of profit do not differ, the merchandise produced in under-developed areas is sold at relatively low prices: a worker in a rich area can therefore purchase for relatively few hours of his/her work the product of an entire day's work of a worker in an under-developed area.

Introduction

The word region is most often used to refer to a delimited part of the surface of the Earth. In geography, regions are identified either on the grounds that the area as a whole shares certain common characteristics, or is in certain respects functionally interdependent or is an area over which economic, political, or administrative dominion or authority is exercised. Often considered as referring to a part of a nation, the word can also denote supranational territorial entities. As for the word 'model', it can refer either to a particular historical or geographical pattern of development or to an abstract or simplified conceptual representation of the characteristics and determinants of real world situations. In the second case, the aim is to replicate or explain these workings and to identify underlying causal mechanisms. In the former, the aim is to identify the variety of forms that development can assume.

Generally speaking, the word development refers to the act of making the area more useful or more productive of useful things (development of or in an area) or to development of or for the people who reside in an area (human development) and is often associated with the idea that places and their inhabitants can pass from lower to higher stages of organization. Development is therefore a process in which people are at one and the same time the actors and the beneficiaries.

Development is often measured simply by calculating how much new wealth (gross domestic product (GDP)) is created in a particular area in a particular year or how much income accrues to an area's inhabitants (gross national income (GNI)). Most models of regional development are concerned either with explaining such increases in output and income or with explaining the actions and improvements that bringing about these outcomes. Although the aim of this article is to outline these models, it is important to note that wealth creation is essentially a means to an end, rather than an end in itself. According to Seers, the purpose of development is to reduce poverty, inequality, and unemployment. More recently, Sen has argued that development involves expanding the real freedoms that people enjoy: development should increase capabilities defined as the freedom to achieve valuable beings and doings (themselves called functionings) which include good health, adequate nourishment, access to educational services, happiness, self respect, and security. An important implication is that models of regional development should address questions of distribution as well as of growth. As we shall see, few do.

Figure 1 summarizes the existing literature. It indicates that an area's development, wealth, and income depend on the interaction of four sets of factors. The first comprises natural ('first nature') and created ('second nature') resource endowments (natural resources, infrastructures, population, skills, degree of development of technological capabilities and the forces of production, etc.) which influence the activities in which its inhabitants specialize and their role in wider divisions of labor. The second concern is the degree to which and the effectiveness with which resources and an area's human potential are mobilized. The extent and effectiveness of resource mobilization depends not just on the technologies and skills at an area's disposition but also on institutional and social relations, institutional capacities and performance, and the extent to which the wider national and international context is conducive to development: in a zero sum game, only the most successful prosper; in a positive sum game the opportunities for all are greater. The third is the extent to which its inhabitants can command and control the resources of others which depends on power relations, the degree of unequal exchange, and the extent of transfers of wealth. The fourth involves recognition of the fact that resources are created, conserved, and reproduced in and through development. As a result, resource endowments are a consequence as well as a cause of development, and development

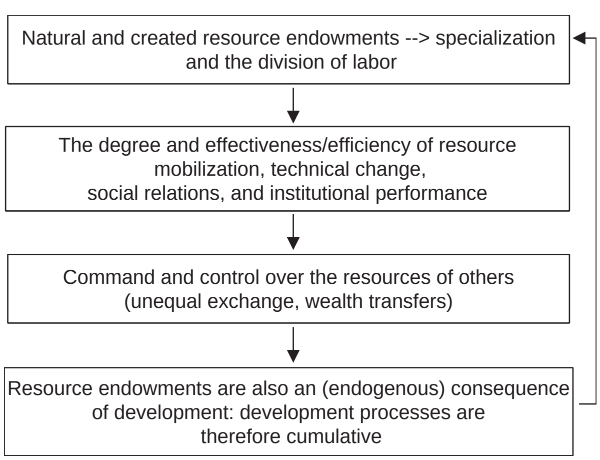

Figure 1 Development, resources, technologies, social relations, and cumulative causation.

processes are cumulative: development may promote further development through a progressive reinforcement of progress, or development may create obstacles that result in lock in and limit further progress. One implication of this fourth point is that development is endogenous and path dependent. Another is that it is cumulative and can involve self reinforcing virtual and vicious circles.

Models of regional development deal also with the evolution of regional disparities/inequalities and identify the mechanisms and actions that have shaped the past trajectories of regional economies, that explain their current evolutions and that are thought to permit them to arrive at a preferred future state. Existing models can be distinguished according to whether regional differences, GDP or GNI, are expected to get smaller (convergence or catch up) or wider (divergence) over time.

Neoclassical Convergence Models

Neoclassical growth, regional development, and trade models rest on the idea that regional development depends on resource endowments. Trade models treat resource endowments as exogenous. Growth models treat population growth and technical progress as exogenous, while capital resources depend on the rate of saving and investment/accumulation of physical capital, although the impact of additional physical capital on growth is itself assumed to dwindle to zero. Neoclassical regional development models explicitly allow for the mobility of capital and labor and their impact on regional resource endowments (e.g., increasing human resources in areas of net in migration). All three groups of models pay no attention to social relations and institutional performance. Instead, all assume that market mechanisms will automatically ensure that resources are employed fully and efficiently. To this end, all these models rest on (unrealistic) assumptions of constant returns to scale, implying that costs of production do not change as the volume of output changes; the related assumption of perfect competition, with its implication that the economy is made up of small enterprises none of which can have any impact on market prices; and an assumed instantaneous re employment of resources, implying that no resources remain unemployed. Externalities, market failure, information asymmetries, and coordination failures (if one enterprise invests, it is profitable for another to do so) which in practice play an important role in shaping regional development trajectories are assumed not to exist and not to matter.

The first of these groups of models includes the Solow growth model, and the growth accounting exercises to which it gave rise. At that stage, the dominant national and regional growth model was the Keynesian

Harrod Domar type model. This particular model attributed growth to the savings rate (on which capital investment depended) and the productivity of capital. In addition, it suggested that actual growth could fall short of the rate required to ensure full employment and identified instability as a possible feature of capitalist economies. The Solow model was radically different in its implications. The model rested on the idea that regional economic performance depends on (1) growth in the volume of capital and labor inputs, of which an area's resource endowment is made up and (2) increases in productivity, which depend on technical change. The Solow model predicts that the growth path of a country or region will converge on a steady state. As long as labor is mobile (or regional rates of population growth are identical) and technological knowhow is public, this steady state growth rate is the same for every country or region, while in this steady state per capita variables such as per capita output do not vary. According to these models, the existence of similar conditions in the economies under consideration would result in a convergence in per capita income. The reason for this is that economies that are less developed are expected to grow/approach the steady state faster than economies that are more developed, while, in the steady state, per capita variables do not change.

Two factors underpin the expectation of faster growth of less developed economies. The first is the existence of diminishing returns to capital and constant returns to scale. At any given level of technological development, these conditions imply that: (1) increases in per capita output get smaller as the amount of capital per worker increases; (2) increases in productivity cease once a steady state/equilibrium amount of capital per worker is reached; (3) increases in per capita output associated with incremental investments are greatest in areas/enterprises that are the least mechanized; and (4) countries that differ only in initial levels of per capita GDP will converge on the same level of GDP per capita. The second is the view that technology is a public good available at no cost to everyone, and that the diffusion of technology and of knowledge from advanced to less developed enterprises and areas will close technology and productivity gaps.

The second set of models are neoclassical models of regional economic development. These models suggest that the flow of capital to low income areas and of labor to high income areas will result in an equalization of income per head. In this case, the idea is that the movement of labor from low to high wage areas will raise wages in the former and reduce them in the latter. Capital, conversely, will move in the opposite direction as the rate of profit is expected to be higher in less developed areas (**Figure 2a**). Movement will continue, it is argued, until there are no further differences in rates of return in different areas. A further implication of this model is that the smaller the out movement of workers and the smaller the rise in local wages, the greater will be the capital in migration. A number of recent cases place a question mark over this conclusion: the fact, for example, that members of the Indian diaspora communities in Silicon Valley play an important role in channeling investment to Bangalore, indicates the possibility of an interdependence between exporting workers and importing capital.

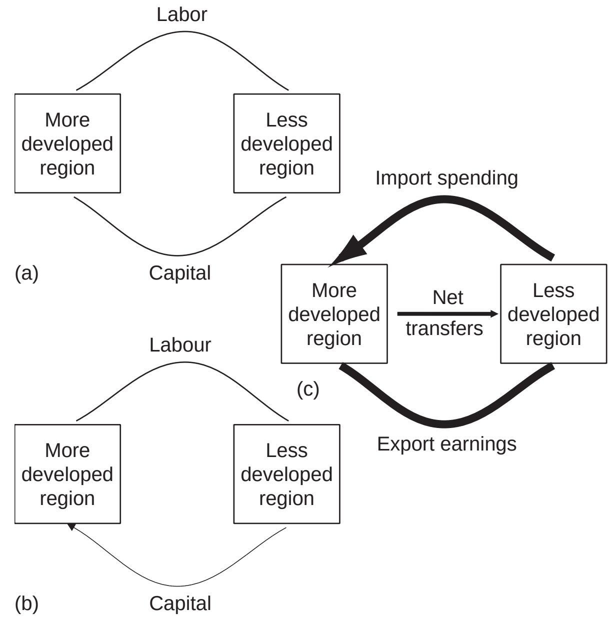

Figure 2 Factor and income flows and regional dynamics.

The third set of models are neoclassical factor proportions explanations of specialization and trade. These models predict an equalization of factor prices independently of factor mobility. According to these models, areas specialize in activities that make intensive use of resources that are relatively abundant: areas with abundant supplies of skilled workers specialize, for example, in skill intensive activities. Increases in the demand for relatively abundant resources and reduced demand for relatively scarce resources will raise the incomes of the former and reduce the incomes of the latter.

Circular and Cumulative Causation

Models of circular and cumulative causation (CCC) differ significantly from the neoclassical tradition. One reason for this is that cumulative causation models identify positive/negative feedback mechanisms and emphasize the endogeneity of factors such as resource endowments considered exogenous in the neoclassical models considered earlier. According to these models, comparative development depends on the relative weight of unequalizing centripetal forces of attraction and suction and equalizing centrifugal forces of diffusion. Myrdal, for example, called the former, backwash effects. An example is the way in which net flows of capital and labor are directed toward more rather than less developed areas (**Figure 2b**). As his aim was to explain increasing economic inequality between developed and underdeveloped areas, these forces received particular attention. Myrdal recognized, however, that there were also spread (trickle down) effects that worked so as to equalize development. In addition, he emphasized the importance of institutional factors in shaping development trends and argued for active policy intervention in order to promote greater equality which he thought would also contribute to greater economic growth. An example of such public action is provided by net flows of public expenditure put in place specifically to counteract the unequalizing effects of market mechanisms (**Figure 2c**).

Another strand of circular and cumulative causation models is rooted in the work of Kaldor who himself drew on the earlier work of Young. At the center of this approach was the idea that economies of scale were a fundamental characteristic of economic life (yet were assumed not to exist in the neoclassical tradition). Essentially, Kaldor argued that national and regional growth was export led. The growth of exports depends on the efficiency wage (the ratio of real wages and productivity). Increased exports of manufactured goods imply increased output. The consequent increase in output implies, as a result of economies of scale, external economies, or spill over effects and complementarities, increased productivity. Next, increased productivity improves competitiveness contributing to a virtuous spiral.

The emphasis on export demand is indicative of a second major difference. Neoclassical growth models are supply led. These models concentrate on the expansion of potential output on the supply side of the economy, and assume that aggregate demand will equal potential output. Kaldorian models are Keynesian models of demand led growth in which aggregate supply is adjusted to accommodate demand led changes in actual output through changes in capacity utilization and/or induced changes in accumulation, migration, and technical change.

Theories of Modernization, Dependency, and Unequal Exchange

Modernization theories have also informed analyses of regional development. These theories examine the emergence of the modes of economic and social life and economic and social organization that first appeared in Europe and were subsequently extended to other parts of the world. In extreme cases all areas are considered to occupy different positions on the same development path. A post World War II example is afforded by the stages of (neo American) economic growth identified by Rostow: traditional society; the preconditions for take off; take off; the drive to maturity; and the age of high mass consumption. Contemporary examples include neoliberal models of globalization and development. All of these accounts suggest that less developed areas simply find themselves at an earlier stage along a single development path already charted by the most developed capitalist economies in the world.

An influential East Asian variant of modernization theory was the flying geese paradigm (**Figure 3**). This model is centered on the idea that East Asian catch up involved the emergence of a dominant growth center (Japan) which subsequently acted as the leader of a hierarchical group of followers that included in the second tier, the Asian Tiger Economies and in the third tier Malaysia, Thailand, Indonesia, and other members of the Association of Southeast Asian Nations (ASEAN), and finally China and Vietnam (which is also an ASEAN Member State). Drawing on dynamic versions of trade theory, the idea was that the emerging economies in East Asia pursued one after another, a sequence of industrial evolutions. The first involved a product cycle sequence involving successively the import of modern manufactures, domestic production, export, and finally re imports once production was moved offshore. The second involved an industrial sequence involving a movement from lower to higher value added activities and a succession of industries (textiles, chemicals, iron and steel,

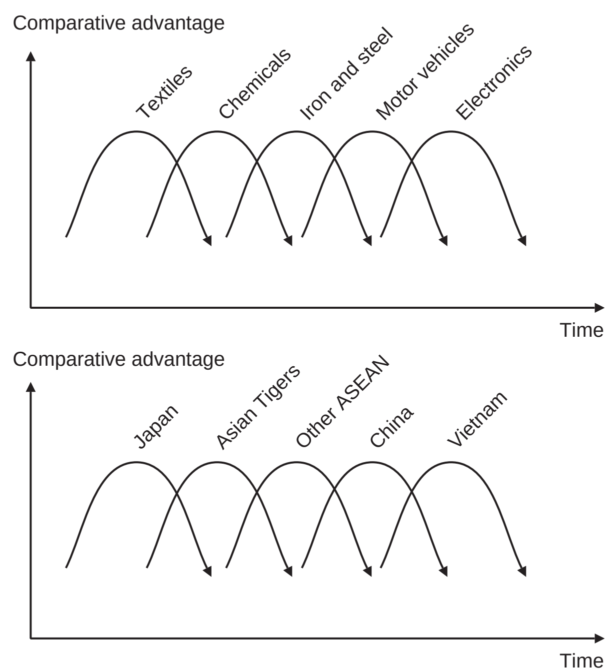

Figure 3 The flying geese model.

motor vehicles, and electronic products). The third se quence was inter national rather than intra national and involved the transfer of products and industries from countries that were more advanced to countries occu pying lower positions in the hierarchy.

Theories of dualism developed especially in studies of economically under developed areas are a variant of modernization theory. These theories emphasized the existence of sharp differences in the organization, degree of development, and goals of actors in the modern and traditional sectors of under developed areas and envis aged a progressive expansion of the former at the expense of the latter. A classic example is the Lewis model which rested on the idea that there was disguised unemploy ment in the traditional agricultural sector/rural areas so that the transfer of surplus labor from the agricultural sector/areas to the modern sector/more developed areas would leave agricultural output unchanged, yet increase the output of the modern sector. As with many orthodox economic models, the outcome was not the anticipated one: agricultural output declined; rural–urban migration, urban unemployment, and regional disparities increased and rural incomes were not large enough to provide a market for manufactured goods.

A number of the other models applied to under developed countries identified vicious circles of under development. An example is the case where insufficient savings leads to capital shortages, capital shortages lead to low productivity, low productivity to low real incomes, and low real incomes to insufficient savings completing the circle.

A wide range of strategies were designed aimed at breaking this and other vicious circles and at ensuring a progressive expansion of the modern sectors of under developed economies. Most of these models targeted investment. Models of balanced growth argue for a series of complementary investments capable of providing market demand for each other perhaps with the added impulse provided by a big push and active development planning. Models of unbalanced growth propose a con centration of investment in a few small projects whose backward and forward linkages sequentially create shortages, bottlenecks, and opportunities for further profitable investment. Growth pole models were a variant of an unbalanced growth strategy adapted to the devel opment plan approach that Hirschman opposed. For Perroux, a growth pole is an enterprise or industry that has propulsive effects on related activities through the creation of external and agglomeration economies and forward and backward linkages, while for Boudeville, the concept could also apply to geographical areas of con centrated development (growth centers and perhaps to industrial districts).

In the 1950s–1980s a more critical set of perspectives emerged in the shape of renewed interest in (internal) colonialism and imperialism on the one hand and the emergence of theories of dependency and world systems on the other. All of these models contested the idea that all areas were following similar development paths. At the center of these approaches was the idea that the causes of relative under development were to be found not so much in the internal conditions in less developed areas as in the impacts of their dependence on economically more advanced/industrialized areas. Although economically under developed areas did have distinctive characteristics, these characteristics were a result of relationships of de pendence that created structural deformations, center–periphery relationships, and class divisions between ex ternally oriented elites and marginalized people within economically under developed areas.

The original thesis was associated with Singer and Prebisch. At the center of this approach was the view that the application of the principle of comparative advantage in the global economic system condemned some under developed countries to a situation in which they spe cialized in and exported raw materials and agricultural products and imported from more developed countries, manufactured goods. Contrary to the expectation that productivity growth in manufacturing would reduce the relative price of manufactured goods, these writers ar gued that the prices of rich area manufactures increased relative to the prices of the under developed area exports. As a result, this division of labor created a self perpetuating structure of development and under development in which poor countries imported high value added goods and exported low value added goods, so that export earnings were never sufficient to pay for imports. A further possibility was that of immiserizing growth if increased exports from less developed areas are more than offset by deteriorating terms of trade.

More radical interpretations of the drivers of the terms of trade included theories of unequal exchange. According to these theories, the fundamental problem was that wages are relatively low in economically under developed countries, and that rates of profit do not differ (with capital actually flowing to economically developed areas) so that the merchandise produced in under de veloped areas is sold at relatively low prices: a worker in a rich area can therefore purchase for a few hours of his/her work the product of an entire day's work of a worker in an under developed area. This proposition was ex tremely controversial for it suggested that instead of capitalists exploiting workers it is workers in rich areas that exploit workers in less developed areas: the former make the latter pay for their high wages through unequal exchange.

The dependency thesis itself was also given a more radical interpretation with the rise of radical dependency and world systems approaches. These models explained the underdevelopment of dependent/peripheral/satellite and of semiperipheral countries by their political and economic subordination to the interests of dominant/center/metropolitan economies and to their domestic allies. Underdevelopment was seen as radically different from no development/development not having taken place (i.e., underdevelopment is radically different from no development/development not having taken place). The reason for this is that peripheral resources are actively used. The way in which they are used is to benefit metropolitan economies and involves the creation of internal rural–metropolitan imbalances as under developed economies are organized in order to satisfy the needs of their export sectors, the enclaves in which they are located, and their rich externally oriented elites.

Models of Endogenous Development I

Theories of dependency led to an emphasis on infant industry protectionism, import substitution, and auto centric development. These ideas were reflected in the emergence of a more general emphasis on the relative merits of self centered and locally controlled develop ment. Two factors underpinned this change of emphasis. The first was the impact of the economic crises of the 1970s on exogenous or externally controlled investments in many economically under developed areas. The sec ond was the wave of productive decentralization that these crises precipitated, the subsequent identification of industrial districts (especially in southern Europe) as drivers of regional economic development, and the yet later observation of the dynamics of new technology industries.

In many cases, these areas of industrial growth were dense concentrations of interdependent small and me dium sized firms (SMEs) in a single sector and in aux iliary industries and services. These districts were interpreted as the result of two sets of forces. The first was a set of economic forces that included: (1) scale economies that result from a high degree of special ization and division of labor; (2) external economies that arise from the existence of shared infrastructures, ser vices, and information; and (3) the availability of special skills and the pooling of the workforce. The second were the interactions between the economic and social system that generated a social atmosphere and communities of firms and people conducive to industrial development, whose consideration opened the door to models dealing with the social, cultural, political, and institutional foundations of the district model, including analyses of social norms and values, political subcultures, associa tionalism, good governance, institutional density and performance, conventions, trust, social capital, and entrepreneurship.

These theoretical developments were associated with some important shifts in the models that dominated re gional development strategy thinking. On the one hand, strategies of self reliance, autocentric growth, and de velopment from below were advocated as an alternative to development from above. On the other hand new engines of regional economic growth were identified: instead of the consumer and intermediate goods sectors that had dominated growth, attention switched in de veloped countries to investments in human capital and information and communications technologies (ICT) and the commodification of knowledge to produce infor mational goods. More generally, emphasis was placed on innovation and learning, investment in research and de velopment, the diffusion of knowledge and its social, cultural, and institutional determinants.

Models of Endogenous Growth II

The next step was the development of new neoclassical models of endogenous growth. Although these models were supply led, some of the conclusions were similar to those of demand led Keynesian cumulative causation models. Associated principally with Rohmer and Lucas, endogenous growth models assume nondiminishing or in the limit nonzero marginal returns to the accumulation of physical and human capital and to the accumulation of knowledge. Nondiminishing returns arise as a result of the association of factor accumulation with endogenous technical progress: the development of successive generations of equipment and of goods and services are associated with learning and with cumulative improve ments in human knowledge and technology and cumu lative increases in human skill and knowhow. At first, these improvements were considered to amount to spill overs that led to increasing returns to scale at the level of the economy but constant returns at the level of the firm. As a result, economists could continue to assume that firms operated as if they inhabited a world of perfect competition. At a later stage, allowance was also made for economies of scale at the level of the firm and for im perfect competition. This change reflected a final rec ognition of the fact that firms that carry out organized research and development can recoup the costs incurred by securing a temporary monopoly and restricting tem porarily, the diffusion of the innovations that result.

Growth can therefore stem from societal spill overs associated with the general advance of knowledge, and from the results of partially appropriable research and development. An important feature of this approach is the reconceptualization of the nature of knowledge. First, it is recognized that knowledge is a nonrival good that can be used simultaneously by different people, and over and over again by the same people at close to zero marginal cost. Second, it is recognized that knowledge is a partially excludable good. Third, investments that increase human knowledge yield increasing returns at the level of the firm. There are two reasons for this. First, insofar as enterprises can limit the diffusion of new ideas, they can spread the fixed costs of developing new products and technologies over a large volume of sales, yielding strong decreases in average costs as output increases. Second, investments that increase knowledge nonetheless generate strong externalities. The reason for this is that knowledge cannot be perfectly patented or kept secret (and is therefore only a partially excludable good): once it is known that something can be done, other firms can seek to replicate it, so that new knowledge has a positive effect on the production possibilities of other enterprises.

Applied to the study of comparative development, this movement away from constant returns to scale and perfect competition to a world of increasing returns and imperfect competition can imply cumulative causation and divergence rather than convergence, as new investments in places and enterprises that are already advanced and associated improvements in knowledge will lead to a continuation of investment and growth and create new development gaps, helping to explain why economically less developed areas fail to catch up.

A particularly striking application of these ideas is found in the work of Quah. Quah predicts the emergence of twin peaks in the distribution of regional income/household income in what he calls weightless (superstar) economies in which value is embodied in immaterial things (**Figure 4** which portrays the transition from an economy in which there are a lot of people with middle incomes and few rich and few poor people to a twin peak model). One cause of the twin peak model lies in the supply side effects identified earlier. For Quah, in weightless economies, ideas are themselves commodities. The immaterial goods that result are distinct in that they do not have to be transferred from one person to another but are simply replicated or copied at a marginal cost that is close to zero (and where, if equipment is required to use these goods, the upfront costs are small). Alongside this supply side factor, there are two demand side factors. First, the market for many immaterial goods is extremely large and often global in character (lifting the constraints on the division of labor posed by the extent of the market). Second, consumers often prefer to purchase the products/services of the famous to those of people who are less famous but whose talents, skills, and abilities may differ only marginally from those of the most famous. The combination of a very high level of demand for the goods/services offered by superstars and their very low costs of reproduction creates very high incomes in a world of winners and losers and wide inequalities. These mechanisms clearly have an impact on territorial development as immaterial goods and services are produced somewhere just as the people whose incomes are derived from them must live and work somewhere. What is more, Quah also suggests that the twin peak model may be applicable to regional economies.

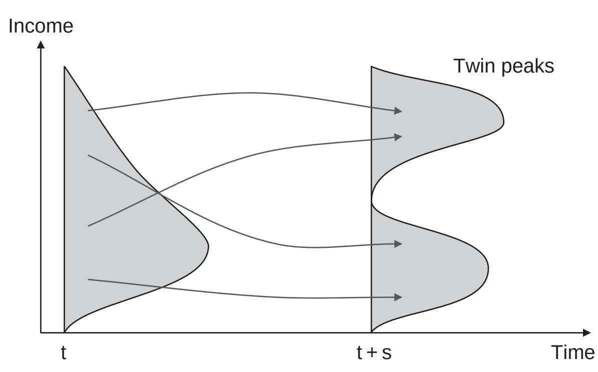

Figure 4 Twin peaks. From Quah, (1996).

All of these models deal with determinate dynamic systems which move toward a final state determined by exogenous data. Evolutionary models deal with indeterminate dynamic systems in which outcomes are path dependent. Analyses of indeterminate dynamic systems lead to models of regional development in which current growth is influenced by its own past record. Although the models examined earlier point to generic causal mechanisms, regional growth is historically contingent requiring analysis also in terms of concepts such as cumulative causation, lock in, hysteresis, and evolutionary change.

Economic Geography Models

Explaining regional development implies explaining the territorial division of labor (who does what, where, and when, what rewards they receive and in what relationships they stand to other people and economic activities in other places) and the constantly evolving resource endowments on which it depends.

In the past, geographical analyses of the territorial division of labor tended to concentrate on abstract models of economic landscapes. Examples include the work of von Thünen, Weber, Hotelling, Christaller, and Lösch. In the 1970s, the emphasis was placed by Hymer, Lipietz, and Massey on conceptually informed classifications of the functional roles of regional economies in wider national and international divisions of labor. In the 1980s and 1990s, a diverse array of conceptual models were advanced to characterize geographical systems of production, explain their structure, and account for their relative economic performance. Spatial categories examined included industrial districts; areas of specialized production, local productive systems, and system areas; localized industrial systems; clusters; new industrial spaces; technopoles and science parks that comprise innovative enterprises, research centers, and universities; worlds of production; milieux innovateurs; regional innovation systems; and learning regions. To explain these spatial structures, and to explain their reciprocal impact on development, three main groups of causal mechanisms are examined: (1) resource endowments; (2) industrial organization, industrial strategies, externalities, and transaction costs; and (3) innovation, systems of innovation, knowledge, individual and collective learning, and creativity. As a result, however, of interpretations of ideas associated with theories of regulation, institutional economics, or evolutionary sociology, these mechanisms are seen as operating in a series of specific historical contexts and of comprising a range of historical development models.

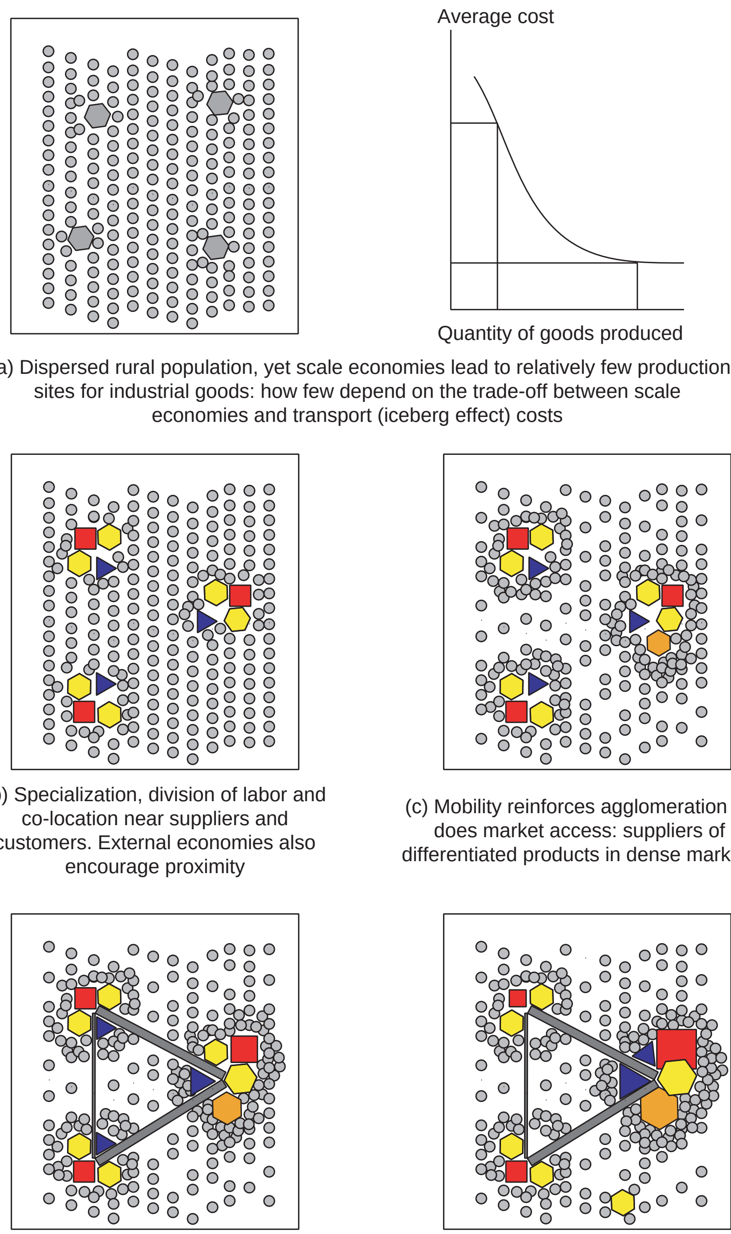

(a) Dispersed rural population, yet scale economies lead to relatively few production sites for industrial goods: how few depend on the trade-off between scale economies and transport (iceberg effect) costs

(b) Specialization, division of labor and co-location near suppliers and customers. External economies also encourage proximity

(c) Mobility reinforces agglomeration as does market access: suppliers of differentiated products in dense markets

(d) Differential transport costs lead to relative growth of increasing returns activities and population in transport hubs

(e) Endogenous growth/accumulation and competition differentially increase/decrease size of economic activities. Dispersion due to diseconomies/reduced transport/factor costs

Figure 5 Explaining economic landscapes. Reproduced from Dunford, M. and Greco, L. (2006). *After the Three Italies. Wealth, Inequality and Industrial Change*. Oxford: Blackwell Publishing, with permission from Blackwell.

Another important recent development is the emer gence of the new geographical economics which seeks to explain the emergence of economic landscapes (charac terized by spatial agglomeration, regional specialization, and core–periphery structures) and the regional resource endowments that were taken as the starting point for many earlier models. To this end, models are developed to examine the location of profit maximizing firms often in a world of increasing returns and imperfect com petition, drawing on analyses of the impacts of increasing returns, forward and backward linkages, external econ omies, endogenous growth, product differentiation, ex ternal diseconomies, etc., to generate imaginary economic landscapes (**Figure 5**).

Conclusions

In this article most attention has been paid to economic models of regional development, although some refer ence has also been made to political and cultural mechanisms, social relations, institutional capacities and performance and the wider national and international economic and geo political context. One reason why is that political and cultural mechanisms are not often considered in modelling approaches. Models are de signed to identify a small number of important factors seens as the main drivers of regional development. Economic mechanisms do play a particularly important role. It is nonetheless the case that many of these models pay insufficient attention in particular to the political foundations of market and capitalist societies. In the Marxist tradition associated for example with the work of Harvey particularly strong emphasis is placed upon the co existence of a logic of territory (the political, military and diplomatic control of territories and their resources) and a logic of capital as in the words of Arendt "a never ending accumulation of property must be based on a never ending accumulation of power". Even in main stream approaches this limitation is reflected in a recent increase in attention paid to institutional factors.

These economic models do nonetheless afford im portant insights. In particular, these competing models lead to the identification of two sets of mechanisms. On the one hand, there is a set of centrifugal forces, of which one of the most important is the transfer of technological and organisational knowledge to less developed areas or less advanced enterprises. Another is the possibility of stagnation and relative decline in relatively developed areas. These mechanisms lead to a tendency towards an equalization of the conditions of production and ex change and of levels of development. On the other hand, there is a set of centripetal forces of which some arise as enterprises and areas that are developed create new sources of competitive advantage thorough, in particular, further investments in knowledge and skills. These mechanisms lead in the direction of a constant differ entiation of the conditions of production and exchange and of the characteristics of regional economies. The value of the models summarised in this entry is that they identify these and other equalising and unequalising mechanism.

At any point in time the relative performance of dif ferent regional economies depends on the relative weight of these two sets of forces. At any moment in time existing resource endowments matter a great deal. An important implication of many of the ideas considered in this article is however that resource endowments are themselves a consequence of earlier stages in what are historical processes of development: what happens at each step depends in part on what has happened in the past, and while each step can be explained in retrospect it is usually not to predict what will happen next as it depends on choices that cannot be anticipated with any certainty. As a result regional development is endogen ous, cumulative and path dependent with distinct possi bilities of the emergence of virtual and vicious circles. The recent growth of interest in evolutionary models capable of dealing with indeterminate dynamic systems in which current growth is influenced by its own past record and in which outcomes are path dependent is therefore a significant step forward.

As the relative weight of equalizing and unequalizing forces can change over the course of time it also means that the evolutions of regional economies are associated with varying degrees of instability. Of the models con sidered it was the Harrod Domar and Keynesian type demand side models that were seen as predicting in stability, whereas orthodox neoclassical models consider market systems as self equilibrating. In Marxist and regulation approaches however capitalist development is also seen as involving phases of stable growth punctuated by periods of crisis. These crises are often important turning points in historical processes of regional devel opment and in the development of ideas. At the root of stable growth is the emergence of a sequence of new development models often centred on fundamental transformations of the preceding economic and social order. These new development models took shape in phases of crisis when older socioeconomic orders failed on the economic front and were rejected on the political and social fronts. As development models succeeded one another so did ways of thinking about regional devel opment. An example referred to in this chapter was the shift from the state led developmentalism of the 1950s and 1960s to market led approaches in the 1980s and 1990s, although at this moment in time one can surmise that contemporary financial crises and shifts in the centre of economic gravity in the world will open the way to further changes in direction.

See also: Dependency; Economic Geography; Growth Poles, Growth Centers.

Further Reading

Carter, I. (1974). The highlands of Scotland as an underdeveloped region. In de Kadt, E. & Williams, G. (eds.) *Sociology and Development*, pp 279 311. London: Tavistock Publications.
Dunford, M. (1990). Theories of regulation. *Environment and Planning D: Society and Space* 8, 297 321.
Dunford, M. and Greco, L. (2006). *After the Three Italies. Wealth, Inequality and Industrial Change*. Oxford: Blackwell Publishing.
Frank, A. G. (1971). *Capitalism and Underdevelopment in Latin America. Historical Studies of Chile and Brazil*. Harmondsworth: Penguin Books.
Garofoli, G. (1991). The Italian model of spatial development in the 1970s and 1980s. In Benko, G. & Dunford, M. (eds.) *Industrial Change and Regional Development. The Transformation of New Industrial Spaces*, pp 85 101. London: Belhaven.
Grabher, G. and Stark, D. (1997). Organizing diversity: Evolutionary theory, network analysis and postsocialism. *Regional Studies* 31(5), 533 544.
Hadjimichalis, C. (2006). Non economic factors in economic geography and in new regionalism: A sympathetic critique. *International Journal of Urban and Regional Research* 30, 690 704.
Harvey, D. (2003). *The New Imperialism*. Oxford: Oxford University Press.
Hirschman, A. O. (1958). *The Strategy of Economic Development*. New Haven, CT: Yale University Press.
Kaldor, N. (1970). The case for regional policies. *Scottish Journal of Political Economy* 17, 337 348.
Kaldor, N. (1972). The irrelevance of equilibrium economics. *Economic Journal* 82, 1237 1255.
Krugman, P., Venables, A. and Fujita, M. (2001). *The Spatial Economy. Cities, Regions and International Trade*. Cambridge, MA: The MIT Press.
Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics* 22, 3 42.
Martin, R. and Sunley, P. (2006). Path dependence and regional economic evolution. *Journal of Economic Geography* 6(4), 395 437.
Myrdal, G. (1957). *Economic Theory and Underdeveloped Regions*. New York: Harper and Row.
Perrons, D. (2004). *Globalization and Social Change   People and Places in a Divided World*. London: Routledge.
Prebisch, R. (1950). *The Economic Development of Latin America and Its Principal Problems*. Washington, DC: United Nations ECLA.
Roberts, M. and Setterfield, M. (2005). *What is Endogenous Growth Theory?* New York: Mimeo.
Romer, P. M. (1986). Increasing returns and long run growth. *Journal of Political Economy* 94, 1002 1037.
Rostow, W. W. (1960). *The Stages of Economic Growth: A Non Communist Manifesto*. Cambridge: Cambridge University Press.
Seers, D. (1969). The meaning of development. *International Development Review* 11(4), 3 4.
Sen, A. (1999). *Development as Freedom*. New York: Alfred A. Knopf.
Singer, H. (1950). The distribution of gains between investing and borrowing countries. *American Economic Review, Papers and Proceedings* 40, 473 485.
Solow, R. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics* 70(February), 65 94.
Stohr, W. B. and Fraser Taylor, D. R. (eds.) (1981). *Development from above or below?* Chichester: Wiley.
Wallerstein, I. W. (1974). *The Modern World System: Capitalist Agriculture and the Origins of the European World Economy in the Sixteenth Century*. New York: Academic Press.
Young, A. (1928). Increasing returns and economic progress. *The Economic Journal* 38(152), 527 542.

Relevant Websites

http://www.rri.wvu.edu/loveridgeintroregsci.htm
http://web.mit.edu/krugman/www/fkvintro.html

Regional Development Theory

P. Nijkamp, Free University, Amsterdam, The Netherlands
M. Abreu, University of Cambridge, Cambridge, UK

Glossary

Endogenous Growth Growth of output in the economy that is driven by long-run improvement of productivity of the production factors, with this improvement being sensitive to various aspects of the economy.

New Economic Geography Axiomatic approach to formal mathematical modeling of economic behavior of households and firms across space, allowing for imperfect competition and resources required for spatial interaction.

Regional Development: What and Why?

Regional development is about the geography of welfare and its evolution. It has played a central role in disciplines such as economic geography, regional economics, regional science, and economic growth theory. The concept is not static in nature, but refers to complex space–time dynamics of regions (or an interdependent set of regions). Changing regional welfare positions are often hard to measure, and in practice we often use gross domestic product (GDP) per capita (or growth therein) as a statistical approximation. Sometimes alternative or complementary measures are also used, such as per capita consumption, poverty rates, unemployment rates, labor force participation rates, or access to public services. These indicators are more social in nature and are often used in United Nations welfare comparisons. An example of a rather popular index in this framework is the Human Development Index which represents the welfare position of regions or nations on a 0–1 scale using quantifiable standardized social data (such as employment, life expectancy, or adult literacy).

The motives to measure regional development are manifold. But a prominent argument all over the years is that welfare positions of regions or nations may exhibit great disparities which are often rather persistent in nature. These in turn translate into large disparities in living standards. For example, in 1960, the world's richest country had a per capita income that was 39 times greater than that of the world's poorest country (after correcting for purchasing power), while by the year 2000, this gap had increased to 91. Areas in our world do not only have significant differences in welfare positions, but it also takes sometimes decades or more to eliminate them. As an illustration we take here Tanzania (the world's poorest country in 2000), which experienced on an average a modest growth rate of 0.6% per annum over the period 1960–2000. In order to reach the world's average per capita income of US$8820 per annum at its current rate of growth, it would need another 485 years. Even if the annual growth rate were to increase to 1.8% (i.e., the world's current average), it would need 161 years to close the gap. In addition, if it were to grow at the rate of South Korea (the fastest grower over the period concerned), it could close the gap in just 49 years. Persistent spatial welfare disparities are a source of frustration for both economists and policymakers.

This article is about regional economic development, in contrast to national economic development. The difference between regions and countries is not always very clear (for instance, several US states are larger than many countries in Europe), but the major distinction in most cases is the fact that regions are open spatial entities (in contrast to countries), while the competence of a region may normally be superseded by the nation. Regions display a spatial subdivision of a country and are characterized by a distinct degree of spatial diversity.

Regional development is clearly a multidimensional concept with a great socioeconomic variety that is determined by a multiplicity of factors such as natural resource endowments, quality and quantity of labor, capital availability and access, productive and overhead investments, entrepreneurial culture and attitude, physical infrastructures, sectoral structure, technological infrastructure and progress, open mind, public support systems, and so forth.

The literature on regional development has usually centered around two dominant issues: how is regional welfare created and how can we cope with undesirable interregional welfare discrepancies? The first question is normally referred to as 'allocative efficiency' and addresses the economic issue of an optimal spatial economic use of scarce resources (i.e., inputs such as capital, labor, physical resources, knowledge, etc.) so as to generate a maximum value of output. The second question is more of a sociopolitical equity nature and addresses the mechanisms and conditions (economic, policy interventions) that may help alleviate undesirable development disparities in the space economy. Normally, efficiently operating regions tend to grow faster than regions with less favorable development conditions, so that there is an inbuilt tension between efficiency and equity among a system of regions, at least in the short run. It goes without saying that the efficiency equity dilemma is one of the most intriguing issues in regional development policy which has extensively been discussed in the literature.

The present contribution aims to shed light on the complexity of regional development. It will start off from the heart of regional economics, viz. location and allo cation theory, and will include an exposition on neo classical factor endowment and infrastructure theory. Next, a more contemporaneous contribution will be offered on the modern drivers of regional development, viz. knowledge and entrepreneurship, while also paying attention to recent advances in endogenous growth and the new economic geography. In a subsequent section we will address more explicitly the so called convergence debate and the role of governments in regional devel opment policy. We will conclude with some retrospective and prospective remarks.

Location of Human Activity and Regional Development

The history of mankind has exhibited an interesting geographic pattern, where accessibility (e.g., river banks, coastal areas) and favorable physical geographical con ditions (e.g., climate) were decisive factors for settlement. These areas created the foundations for large agglom erations (such as New York, London, Tokyo, or Venice). Regional development appeared to be contingent upon the existence of large economic attraction poles. Thus, location of economic activity created the foundations for regional welfare. Even nowadays, persistent discrepancies in regional welfare have historical roots in locational conditions of these high potential areas. The present figures of our world are still striking: *c.* 1 billion people live on less than a dollar per day, while more than 2 billion people have no access to adequate sanitation. The gap between poor and rich is formidable and even increasing. For example, the top 20% of the world's population consumes about 85% of the world's income, while the bottom 20% lives on *c.* 1.5% of the world's income. Things get worse: a generation ago, people in the top 20% were 30 times as rich as those in the bottom 20%; nowadays, they are more than 70 times as rich!

The location patterns of people and economic activity in our world show apparently a great variation. Hence, location theory has played a central role in explaining not only the dispersion of economic activity, but also the dispersion of welfare among regions. Consequently, re gional development theory is deeply rooted in location theory.

Location theory has already a long history in regional economics and economic geography. Starting off from pathbreaking ideas set forth by Von Thünen, Christaller, Lösch, Isard, Hoover and many others, modern location theory has moved into a strong analytical framework for regional economics and economic geography. Cost minimization and profit maximization principles are in tegrated in a solid economic setting, in which both partial and general spatial equilibrium studies on the space economy can be found that highlight the geographical patterns of industrial and residential behavior. Further more, the theory is also able to encapsulate the impact of public actors (e.g., regional development policy).

Thus, the fundamentals of classical location theory are made up of a blend of physical geography (determining the accessibility of a location and the availability of re sources) and smart economic behavior (through a clever combination of production factors and market potentials in space).

However, location patterns are never static, but have an endogenous impact on newcomers. Thus, incumbent firms may attract others through scale, localization, and urbanization advantages (e.g., in the form of spatial economic externalities in a Marshallian district). Con sequently, agglomerations tend to become self reinforcing spatial magnets impacting on the entire space economy. Such concentrations of economic activity create welfare spinoffs for a broader regional system and thus determine the geographic patterns of welfare and regional devel opment. In this context, we may also observe a blend between location theory and urban economics or urban geography.

In recent years, we have witnessed the emergence of the digital economy through which actors could be net worked worldwide. As a consequence, the interaction between industrial networks and location as well as the access to telecommunication networks has gained much interest. Locations that offer the best available network services are the proper candidates for many firms in the information and communication technology (ICT), high tech and high services sectors, and are able to generate a high value added to regional development.

The availability of and access to infrastructure is an other critical success factor for regional development. In addition to the presence of labor as capital on traditional factor inputs, we observe an increasing interest in measuring the impact of infrastructure on regional de velopment. Especially in a world with shrinking dis tances, space–time accessibility of regions becomes a critical determinant of relative regional economic pos itions. Transport economics and transport geography have offered an abundance of theoretical and empirical evidence on the importance of physical infrastructure for regional growth. The uneven provision of infrastructure has also been identified as a key determinant of regional income disparities in less developed countries, as is witnessed in various World Bank studies.

Entrepreneurship, Innovation and the Knowledge Economy

Since Marshall, Schumpeter, and Kirzner, we know that innovation and entrepreneurship are the driving factors behind economic growth. There is an avalanche of literature on the importance of entrepreneurship for enhancing the innovative capacity and growth potential of regions.

Entrepreneurship is a complex and multifaceted phenomenon that finds its roots in risk taking behavior of profit seeking individuals in a competitive economy. But its determinants have also clear correlations with gender, age, education, financial support systems, administrative regulations, risk tolerance, and market structures. Entrepreneurship lies at the heart of innovation as the art of doing creative things for the sake of competitive advantage. The debate on entrepreneurship and innovation has – from a geographical perspective – prompted the emergence of new concepts such as innovative regions, innovative milieus learning regions, or knowledge based regions. Innovation has become the critical survival factor in a competitive space economy and determines the direction and pace of regional development. A key aspect of innovation in a modern space economy is the use of and access to the ICT sector. Consequently, ICT infrastructure is increasingly seen as a necessary resource endowment for regional development.

It goes almost without saying that ICT is a necessary ingredient of a modern knowledge based economy. That also holds for regions. Clearly, knowledge is a composite good with many facets, but from an economic perspective knowledge serves to enhance productivity and to induce innovations. There is indeed an ongoing debate on the unidirectional or circular relationship between knowledge and development, and this forms one of the central issues in endogenous growth theory.

Endogenous growth theory has played a central role in the growth debate since the 1990s. The main idea of these new contributions is that technological progress is not exogenously given, but an endogenous response of economic actors in a competitive business environment. Consequently, in contrast to earlier macroeconomic explanatory frameworks, the emphasis is much more as individual economic behavior of firms. In this way, it can be demonstrated that regional growth is not the result of exogenous productivity enhancing factors, but rather the outcome of deliberate choices of individual actors (firms and policymakers).

The importance of knowledge for innovation and entrepreneurship is thus increasingly recognized. The spatial distribution of knowledge and its spillovers are considered as an important success factor for regional development in an open competitive economic system. Thus, the geographical patterns of knowledge diffusion as well as the barriers for access to knowledge are decisive for regional development in a modern global and open space economy. Consequently, knowledge policy – often instigated by ICT advances – is a critical success factor for regional welfare creation.

Regional development policy appears to move increasingly toward knowledge and innovation policy. In recent years, it is also convincingly demonstrated that leadership and institutional qualities have a great impact on regional welfare, in particular, when the role of leadership is linked with innovation and knowledge creation. To the same extent that innovative entrepreneurship is critical for long term regional welfare growth, governance and leadership are essential for a balanced regional development. Leadership presupposes proactive behavior, visions for future development, awareness of institutional and behavioral processes, responses and bottlenecks, as well as acceptance by the population. The awareness of the importance of leadership and entrepreneurship lies in with the recognition of creative actions and learning actors. Studies on regional leadership are rare, but this is certainly a promising and important new field of research.

In recent years, we have also witnessed the emergence of a new strand of literature, coined the ‘new economic geography’, in the vein of endogenous growth theory. Although the term ‘new economic geography’ is arguably not appropriate (most concepts can already be found in the regional economics and regional science literature since the 1950s), this seemingly new approach has attracted quite some attention within the neoclassical economics literature. It marries the increasing returns monopolistic competition model with the micro foundations of spatial economic behavior, including interregional trade. This recent approach emphasizes the importance of agglomeration externalities (caused by increasing returns to scale) for regional welfare creation, in the context of global competitive forces where trade (between regions or countries) plays a critical role. Thus, regions are then part of a global competitive network system. Recent contributions within this literature have found that agglomeration can be a welfare improving outcome for workers in both core and periphery regions, for instance, if agglomeration raises the innovation rate. This result provides theoretical support for regional development policies destined to support and enhance existing clusters of specialization.

The Convergence Debate

Regions and nations in our world show complex development patterns. Textbook economics would teach us that under conditions of free competition, homogeneity of preference and technology parameters, and mobility of production factors all regions in the space economy would tend to converge to the same per capita income growth rate. In neoclassical economic growth models, convergence between regions takes place through capital accumulation. Regions that are further away from their state states grow faster in the short run, but in the long run diminishing returns to capital set in and the growth rate drops to the exogenous growth rate of technological progress. This tends toward a situation where the growth rate of GDP per capita falls and becomes constant (i.e., it becomes equal to the exogenously determined techno logical growth rate). The neoclassical growth models therefore predict that in the long run countries and re gions will converge in terms of per capita income levels, if one controls for the effects of differences in initial conditions.

A basic problem in the above neoclassical explanation of the world is that technological progress is not exogenous 'manna from heaven'. It is part of the complex architecture of a regional economy and is determined by both internal and external research and development (R&D) invest ments, on the job training, learning by doing, and spill overs from university research. Spillovers resulting from R&D expenditures and other activities generate increasing returns to scale for reproducible production factors, the existence of which implies the possibility of long run di vergence in per capita income levels.

The conflicting predictions of the neoclassical and endogenous growth models have generated intense scrutiny and a plethora of empirical studies, known col lectively as the 'convergence debate'. The literature has generally found that while per capita income levels be tween the poorest countries (of sub Saharan Africa) and the richest countries (Europe and the US) have diverged over the past few decades, there is convergence among countries that are similar in terms of initial conditions and policies, for instance, among the countries of the European Union or the fast growing East Asian econ omies (a phenomenon known as 'conditional con vergence'). The evidence also suggests that per capita income levels among regions within countries have di verged markedly in recent years, particularly in large, diverse countries such as India and China. An increase in regional disparities in fast growing regions such as India and China is not necessarily bad news, however. Im provements in living standards in vast countries such as these imply that global inequality as a whole may be decreasing (in tandem with improvement in living standards in these countries), and economic theory sug gests that an increase in agglomeration may lead to fur ther improvements in the long run, as knowledge spills over into other regions and sectors of the economy. The findings of the convergence literature therefore highlight the key role of regional development policies in pro moting economic growth and human development.

Regional Development Policy in Perspective

The policy response to regional inequality or spatial disparity is characterized by a great variety all over the years. Several strands of literature can be distinguished, in particular:

- supply side policy of a Keynesian nature with a pro nounced interest in public spending in less privileged regions;
- growth pole strategies, with a clear emphasis on a concentrated growth impulse in a few designated places or areas;
- infrastructure policy with the aim to create the ne cessary physical conditions (e.g., improvement of ac cessibility) in order to enhance the competitive capabilities of regions;
- self organizing policy where regions are encouraged to get their acts together on the basis of indigenous strength with a limited role of governments; and
- suprastructure policy in which regions are provided with favorable R&D conditions, educational facilities, knowledge centers, and the like in order to create the conditions for a self sustained development.

Regional policy has played an important role in shaping the European Union, as the vast differences in regional development among European regions would weaken social cohesion in Europe. The structural funds ad ministered by the European Union, and in particular, the regional development fund, have been strategic vehicles to cope with spatial disparities in Europe. It is note worthy that the various above mentioned explanatory frameworks for regional development differences have often been incorporated in the policy responses on spatial disparities in Europe, including the current interest in entrepreneurship and technological innovation.

An important contributor to regional development is technological progress, an extensively studied topic in the recent economic growth literature. From a geo graphic (regional, urban, or local) perspective, much at tention has been paid to the spatial conditions that induce technological progress (e.g., entrepreneurial cli mate, availability of venture capital, incubator facilities, etc.). Furthermore, also the spatial diffusion of technol ogy has obtained much attention, in particular in the geography literature. A particular case of knowledge and technology diffusion can be found in foreign direct in vestment (FDI). Several studies have demonstrated that FDI offers access to foreign production processes, so interregional or multinational technology spillovers may occur. These studies demonstrate clearly that the region is a dynamic player in an intricate web of spatial–economic interactions. Regional development is not a static phenomenon, but exhibits a dynamic pattern based on the interplay of various stakeholders (the business sector, households/workers, and governmental bodies), who have to face the challenge put forward by an open space economy.

See also: Agglomeration; Economic Geography; Location Theory; Networks; Regional Connectivity.

Further Reading

Abreu, M. (2005). *Spatial Determinants of Economic Growth and Technology Diffusion*. Amsterdam: Thela Thesis.

Acs, Z., Carlsson, B. and Karlsson, C. (eds.) (1999). *Entrepreneurship, Small and Medium Sized Enterprises and the Macroeconomy*. Cambridge: Cambridge University Press.

Acs, Z., de Groot, H. and Nijkamp, P. (eds.) (2002). *The Emergence of the Knowledge Economy*. Berlin: Springer Verlag.

Aghion, P. and Howitt, P. (1998). *Endogenous Growth Theory*. Cambridge, MA: MIT Press.

Asheim, B. (1996). Industrial districts as 'learning regions'. *European Planning Studies* 4(4), 379 400.

Asheim, B. and Gertler, M. (2005). The geography of innovation: Regional innovation systems. In Fagerberg, J., Mowery, D. C. & Nelson, R. (eds.) *The Oxford Handbook of Innovation*, pp 291 317. Oxford: Oxford University Press.

Audretsch, D., Thurik, A. R., Verheul, I. and Wennekers, S. (eds.) (2002). *Entrepreneurship: Determinants and Policy in a European US Comparison*. Dordrecht: Kluwer.

Baldwin, R. (1999). Agglomeration and endogenous capital. *European Economic Review* 43, 253 280.

Baldwin, R. E., Forslid, R., Martin, P., Ottaviano, G. I. P. and Robert Nicoud, F. (2003). *Economic Geography and Public Policy*. Princeton, NJ: Princeton University Press.

Barro, R. and Sala i Martin, X. (1997). Technology diffusion, convergence and growth. *Journal of Economic Growth* 2, 1 27.

Benhabib, J. and Spiegel, M. M. (1994). The role of human capital in economic development. Evidence from aggregate cross country data. *Journal of Monetary Economics* 34(2), 143 173.

Blakely, E. J. (1994). *Planning Local Economic Development*. Thousand Oaks, CA: Sage Publications.

Brakman, S., Garretsen, H. and van Marrewijk, C. (2001). *An Introduction to Geographical Economics*. Cambridge: Cambridge University Press.

Cameron, R. (2005). Spatial economic analysis. *Journal of Development Perspective* 1(1), 146 163.

Capello, R. (2006). *Regional Economics*. London: Routledge.

Carr, D., Markusen, J. and Maskus, K. (2001). Estimating the knowledge capital model of the multinational enterprise. *American Economic Review* 91(3), 693 708.

Coe, D. T. and Helpman, E. (1995). International R&D spillovers. *European Economic Review* 39(5), 859 887.

Crozet, M. (2004). Do migrants follow market potentials? An estimation of a new economic geography model. *Journal of Economic Geography* 4, 439 458.

Davis, D. and Stiglitz, J. (1977). Monopolistic competition and optimal product diversity. *American Economic Review* 67, 297 308.

Davis, D. and Thisse, J. F. (2002). *The Economics of Agglomeration*. Cambridge: Cambridge University Press.

Davis, D. and Weinstein, D. (1999). Economic geography and regional production structure: An empirical investigation. *European Economic Review* 43, 379 407.

De la Fuente, A. and Domenech, R. (2006). Human capital in growth regressions. *Journal of the European Economic Association* 4(1), 1 36.

Doring, T. and Schnellenback, J. (2006). What do we know about geographical knowledge spillover and economic growth? *Regional Studies* 40(3), 375 395.

Durlauf, S. N. and Quah, D. (1999). The new empirics of economic growth. In Taylor, J. B. & Woodford, M. (eds.) *Handbook of Macroeconomics*, vol. 1A, pp 235 308. Amsterdam: North Holland.

Englmann, F. C. and Walz, U. (1995). Industrial centers and regional growth in the presence of local inputs. *Journal of Regional Science* 35, 3 27.

Evans, P. (1996). Using cross country variances to evaluate growth theories. *Journal of Economic Dynamics and Control* 20, 1027 1049.

Findlay, R. (1973). *International Trade and Development Theory*. New York: Columbia University Press.

Fingleton, B. (ed.) (2003). *European Regional Growth*. Berlin: Springer Verlag.

Florida, R. (1995). Toward the learning region. *Futures* 27(5), 527 536.

Fujita, M., Krugman, P. and Venables, A. (1999). *The Spatial Economy*. Cambridge, MA: MIT Press.

Fujita, M. and Thisse, J. (2003). Does geographical agglomeration foster economic growth? And who gains and looses from it? *Japanese Economic Review* 54, 121 145.

Grossman, G. M. and Helpman, E. (1990). Comparative advantage and long run growth. *American Economic Review* 80, 796 815.

Hanson, G. (1996). Agglomeration, dispersion and the pioneer firm. *Journal of Urban Economics* 39(3), 255 281.

Heenan, D. and Bennis, W. G. (1999). *Co Leaders: The Power of Great Partnerships*. New York: Wiley.

Hofstede, G. (ed.) (1997). *Cultures and Organizations*. New York: McGraw Hill.

Islam, N. (2003). What have we learnt from the convergence debate? *Journal of Economic Surveys* 17(3), 309 362.

Judd, D. and Parkinson, M. (eds.) (1990). *Leadership and Urban Regeneration*. Newbury Park: Sage.

Keeble, D. and Wilkinson, F. (1999). Collective learning and knowledge development in the evolution of regional clusters of high technology SMEs in Europe. *Regional Studies* 33, 295 303.

Kirchhoff, B. A. (1994). *Entrepreneurship and Dynamic Capitalism*. Westport, CT: Praeger.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy* 99, 483 499.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics* 22(1), 3 42.

Malecki, E. (2000). Creating and sustaining competitiveness. In Bryson, J. R., Daniels, P. W., Henry, N. & Pollard, J. (eds.) *Knowledge, Space, Economy*, pp 103 119. London: Routledge.

Markusen, A. (1985). *Profit Cycles, Oligopoly, and Regional Development*. Cambridge, MA: MIT Press.

Markusen, J. (2002). *Multinational Firms and the Theory of International Trade*. Cambridge, MA: MIT Press.

Martin, R. (1999). Public policies, regional inequalities and growth. *Journal of Public Economics* 73, 85 105.

Martin, R. and Ottaviano, G. I. P. (2001). Growth and agglomeration. *International Economic Review* 42(4), 1003 1026.

Naudé, W. (2005). Geographical economics and Africa. *Journal of Development Perspectives* 1(1), 1 4.

Neary, P. (2001). The new economic geography. *Journal of International Literature* 39, 536 561.

Nijkamp, P. (2003). Entrepreneurship in a modern network economy. *Regional Studies* 37, 395 405.

Njikamp, P. and Stough, R. (eds.) (2004). *Entrepreneurship and Regional Economic Development*. Cheltenham: Edward Elgar.

Puga, D. (1999). The rise and fall of regional inequalities. *European Economic Review* 43, 303 334.

Redding, S. and Venables, A. J. (2004). Economic geography and international inequality. *Journal of International Economics* 62, 53 82.

Rivera Batiz, F. (1988). Increasing returns, monopolistic competition, and agglomeration economies in consumption and production. *Regional Science and Urban Economics* 18, 125 153.

Romer, P. M. (1986). Increasing returns and long run growth. *Journal of Political Economy* 94(5), 1002 1037.

Romer, P. (1990). Endogenous technological change. *Journal of Political Economy* 98(2), 71 102.

Serageldin, I. (2006). *Science: The Culture of Living Change*. Alexandria: Bibliotheca Alexandrina.

Simmie, J. (ed.) (1997). *Innovation, Networks and Learning Regions*. London: Jessica Kingsley.

Stimson, R., Stough, R. and Roberts, H. (2006). *Regional Economic Development*. Berlin: Springer Verlag.

Stimson, R., Stough, R. and Salazar, M. (2005). Leadership and institutional factors in endogenous regional economic development. *Investigaciones Regionales* 7, 23 52.

Storey, D. (1994). *Understanding the Small Business Sector*. London: Routledge.

Temple, J. (1999). The new growth evidence. *Journal of Economic Literature* 37(1), 112 156.

World Bank (2006). *World Development Report 2006: Equity and Development*. New York: Oxford University Press.

Xu, B. (2000). Multinational enterprises, technology diffusion, and host country productivity growth. *Journal of Development Economics* 62(2), 477 493.

Regional Development, Endogenous

F. Tödtling, Vienna University of Economics and Business Administration, Vienna, Austria

Glossary

Bottom-Up Development Represents a concept arguing that (regional) development is initiated and carried mostly by local actors and firms. It is oriented to the needs and objectives of the local population and it draws to a large extent, but not exclusively, from local factors and resources.

Industrial Districts Local concentrations of firms in specific industries, often specialized in traditional sectors such as textiles, clothing, leather, furniture, and machinery. They are characterized by competing and cooperating small and medium-sized firms, and a high level of entrepreneurship. The firms produce differentiated products using flexible machinery and skilled workers.

Less Developed Regions Regions or areas of countries, often in peripheral location, whose inhabitants have less wealth and economic and social opportunities than those in more advanced or central regions.

Regional Development Enhancement of wealth and living conditions in regions understood as subnational territories. It includes the growth of production, employment, and income as well as an improvement of social, political, and environmental conditions in these territories.

Regional Innovation System Consists of firms and institutions engaged in knowledge generation, diffusion, and exploitation in a region. This leads to innovations in the business sector, that is, the introduction of new products or processes.

Regional Learning Builds on regional competencies and skills, and includes knowledge exchange and the sharing of know-how and best practices between regional firms and organizations.

Top-Down Development Represents a concept arguing that (regional) development is designed and implemented by central government or external agencies. It emphasizes external factors such as interregional trade (exports, imports), or the mobility of capital (firms), labor, and technology between regions and countries.

Genesis and Core Concept

Endogenous regional development is a concept and a strategy for regional economic development which was introduced in the late 1970s and has become quite prominent since then. It has been applied in advanced economies as well as in developing countries. Its basic idea is that regional development should be driven in a bottom up manner by endogenous forces and factors. These are factors from the respective region such as land, natural resources, labor, and entrepreneurship. Included here are also factors which are intentionally created or upgraded by policymakers and related institutions, such as schools, training organizations, universities, and research organizations. Created factors, thus, refer to an educated work force, qualified labor, or to knowledge and technologies developed in the region, which might lead to new products, processes, or other new solutions. Endogenous forces also include social and political factors such as the engagement of social agents and civil society that trigger processes of self help, local initiatives, and social movements aiming at the improvement of living conditions in a particular region. Due to the strong role of local forces and factors such a development strategy has often also been called 'bottom up approach'. This refers to the idea that regional development is initiated and carried by local and regional actors and agents instead of the central government or external agencies, and that it is oriented to the needs and objectives of the regional population.

Endogenous regional development has evolved as a counterthesis to previous regional development approaches for less developed areas emphasizing external factors such as interregional trade (exports, imports) or the mobility of capital (firms), labor, and technology between regions and countries. More specifically, endogenous regional development strategies were formulated as a response to the often unsuccessful policies of the 1960s and 1970s which were strongly based on factors such as external demand (export base theory, trade theory), the attraction of leading international firms and technology to growth centers (growth pole theory), and the mobility of capital and labor between economically strong and weak regions (neoclassical growth theory). This previous external development paradigm has also been called 'top down regional development approach' in the literature because it was often designed and implemented by the central government or external agencies. A number of critical reviews and studies have pointed out that those external strategies had not been very effective for improving the economic situation of less developed and peripheral regions and countries. In particular, the gap to the economically leading regions, that is, the core areas, could often not be reduced. The critical points raised were the following:

- Only a small share of regional factors had been mo bilized and used due to the strong orientation on ex ternal demand and on specific comparative advantages of respective regions. In particular specific natural resources, tourist sites, and low cost labor have been exploited in less developed regions and countries, whereas other factors and potentials such as qualified labor, specific skills, and competences were not suf ficiently taken into account or further developed.
- The attraction of external firms and of branch plants has benefited only a few central locations or growth centers in these regions and countries. There were only a few beneficial effects or spillovers from these centers to more peripheral or less developed areas. These were mainly in the form of labor commuting to the centers and the dispersal of branch plants to the hinterland, less through input–output linkages and technology diffusion to peripheral firms.
- The branch plants have lacked higher level functions such as office and managerial activities, Research and Development (R&D), and innovation. They have offered only low quality jobs with few prospects of upward mobility for workers and employees.
- The potential of mobile plants for less developed regions has been reduced due to an increasing glob alization of the economy. New locations for multi national companies were coming up in newly industrializing countries in South East Asia and Latin America as well as in countries of Eastern Europe, offering abundant cheap labor as well as prospects for a high market growth. As a consequence, firms have been setting up new subsidiaries or branch plants in these new locations, and less so in backward regions of advanced countries.
- Due to the dominance of branch plants the external strategies did not raise the entrepreneurial potential and the innovation capability of less developed regions to a notable extent. As a consequence, there were few startups and a lack of innovation, in particular as re gards new products.
- Finally, the external strategies have often deteriorated the environmental situation in the respective regions, because ecological and related aspects had not been taken sufficiently into account.

As a reaction to these problems scholars and policy ob servers were demanding a new orientation of regional development concepts since the 1980s and policies in many advanced and developing countries subsequently were shifting more towards endogenous approaches of regional development. The new strategy was not based on a consistent new theory, but was defined rather as a counter model to the external approach. It comprised the following elements:

- The mobilization of endogenous regional factors and potentials, assuming that they existed and that they could be mobilized. Such potentials were seen to exist in particular in natural resources, landscapes and tourist sites, qualified labor and specific skills, or competencies of the respective regions.
- The inclusion of sectors other than manufacturing in development strategies such as agriculture, craft based industries, tourism, and services. The intention was to formulate integrated concepts trying to interrelate different but complementary sectors. This has in cluded, for example, the strengthening of input–out put linkages and other relationships between those various sectors. Examples are complementarities and interrelations between agriculture, tourism, and craft based activities in rural and peripheral regions.
- There was more concern for development problems and growth potentials of small firms as well as a stronger focus on entrepreneurship (formation of new firms). This was due to the fact that less developed regions often had high shares of small and medium sized firms in traditional sectors which did not benefit from the top down strategies. The formation of new firms, on the other hand, was seen as a key factor for generating growth and renewal in those regions.
- Innovation has received a prominent role. 'Innovation' was defined broadly, and included technological, business, and social innovations. The intention was to escape competition, in terms of cost, from low wage countries through a high quality of products and processes. Innovation, however, was also aimed at the solution of broader social and other problems and not confined only to technology and companies.
- Regional specificities in culture, local demand, and economic structure were regarded as important de velopment factors to be taken into account. Such pe culiar local and regional characteristics cannot easily be copied by competitors, and are therefore seen as a source of unique competitive advantages for regional firms. In this context a certain level of 'regional identity' was regarded as a favoring factor for regional development. In business this took the form of branding of regional products, and of using new ways of marketing such as direct selling.
- Beyond narrow economic factors, social and political forces were regarded as relevant. This included activ ities of social agents and the engagement of civil so ciety triggering and supporting processes of self help, local initiatives, and social movements aiming at the improvement of living conditions in particular regions.
- Decentralized decision making and policy com petences were seen as key factors and goals of regional

development. The decentralization of political and administrative competences towards the local and regional levels were regarded as favorable for economic development. It was assumed that there was a better understanding of problems, barriers, and potentials for regional development at those lower levels and since it allowed a better adjustment of development strategies to the needs and goals of the regional population.

- Finally, a key idea was that regional economic development should take a long term perspective and harmonize economic, social, and environmental goals. Economic growth, thus, should not deteriorate the environmental and ecological situation of the region, an idea which more recently has been introduced as 'sustainable regional development'.

Endogenous regional development is often equated with bottom up development. The central idea here is that regional development is to a high degree a bottom up process, initiated and carried out by local and regional actors, influenced and shaped by regional institutions and policy. There is a specific view of space underlying bottom up development differing from top down strategies. Space is not just seen as simple distance between places giving rise to transport and transaction cost, but is conceived as 'territory', representing the clustering of social relations, and having specific cultural and other features. Furthermore, public and other regional institutions play a key role in regulating and steering social and economic development. Regional specificities and the crucial role of regional institutions and of policy prevent the emergence of a uniform development model, leading to specific patterns of regional development. Endogenous regional development, thus, is a highly differentiated process in geographical space.

Key mechanisms for generating endogenous or bottom up development are development initiatives at the local and regional levels. These are started and implemented by actors from the region, using mainly regional resources and aiming at regional benefits. They respond to needs of the regional population and are created and controlled by individuals and groups of the respective community. Central objectives are often the creation of viable and worthwhile employment or the improvement of regional living conditions. An underlying hypothesis of this development approach is that the basic prerequisites of development – initiatives and entrepreneurship – are available or latent in most regions. Actions initiated at the regional level are seen to have several advantages compared to central or top down approaches such as direct problem perception, high intensity of interaction, regional synergies, regional strategy formulation, and collective learning within regional networks.

Endogenous regional development furthermore is seen as a process going through different stages. It starts with the emergence of entrepreneurship at the local and regional levels, and the start up and growth of regional firms. This is followed by an expansion of these firms beyond the region through exports and direct investments in other regions. A final state is an economic structure with control functions and business services under regional control. However, we have to take into account that regional development is always the result of both endogenous 'and' exogenous factors, and their interaction. This leads to a plurality of development paths and of development models, partly also due to the effects of different economic policies. Ideal cases or prefixed stages of regional development, therefore, do not exist. Furthermore, individual regions have a different capacity for endogenous development and therefore a differentiated need for central or external development inputs and efforts. It is a kind of paradox that particular peripheral and disadvantaged regions, which often were the target of endogenous regional development strategies, have a rather low potential for such strategies and require more external resources and support for triggering dynamic regional development.

As already mentioned, endogenous regional development is not a uniform concept. It has appeared and since the 1980s was further developed in different variants such as the bottom up or the local development approach sharing most of the stated principles and elements. Furthermore, the concept was inspired by and was further developed in a close relationship with other concepts and theories such as industrial districts, regional entrepreneurship and learning, and regional innovation systems. These can be regarded as related concepts sharing some of the basic ideas and elements of endogenous regional development but deviating in others.

Related Concepts and New Routes

Industrial Districts

The endogenous regional development approach was strongly stimulated by the literature on industrial districts, investigating regions of 'Third Italy' (e.g., Emilia Romana, Toscana, Veneto) and other regions of Europe such as Baden Württemberg, Jutland, or the South of France. For these areas a new pattern of industrialization has been observed characterized by territorial diffusion, leading to a change of core–periphery relationships in Europe since the 1980s. Often, these districts were specialized in traditional sectors such as textiles and clothing, leather and shoes, furniture, and machinery, and they were showing a novel form of development characterized by small firms, entrepreneurship, and flexible specialization. The firms involved had been producing differentiated products in small batches, with general purpose machinery and skilled workers, and, although they had often been competing fiercely, they had also maintained cooperative links and subcontracting relations. It has been demonstrated that firms in industrial districts were strongly embedded into the respective regional economy through input–output links, knowledge exchange and collective learning, and various kinds of social relationships such as family ties and relationships to local unions and other interest groups. Furthermore, there were elements of collective action, institutions, and of policy support, such as the provision of services in the fields of R&D, product development, technology upgrading, sales and distribution, and marketing.

Industrial districts were often regarded as role models for endogenous regional development, since they combined several relevant elements: local entrepreneurship, dynamic small firms in traditional sectors competing successfully on the world market through a high quality of products, incremental product innovation and continuous upgrading of technology, cooperation with customers and suppliers (vertical) and competitors (horizontal), as well as collective action and policy support at the local level. As a consequence, they were copied by policy makers in many regions across the world, even if the respective preconditions did not exist. Such conditions were a local concentration of SMEs in particular industries and a certain entrepreneurial potential, social capital and trust as preconditions for cooperation, and the existence of supporting institutions and services in the fields of education, qualification, and R&D.

Industrial districts cannot be regarded as a general model of endogenous development, however. There are convincing arguments in the literature that the model has worked only under highly specific social, economic, and institutional conditions which are not present in many regions of advanced and less developed countries. Then, industrial districts have been strongly challenged and transformed by the globalization of their industries since the 1990s. Emerging economies in Asia, Latin America, and in Eastern Europe have been catching up rapidly in some of those sectors. Industrial districts have reacted in various ways to this challenge: some of them have applied a globalization strategy by actively transferring production activities abroad to low cost countries; others have experienced an increasing penetration by large firms through a process of mergers and take overs by incoming foreign firms. A third group has successfully been upgrading by specializing in high end product segments or activities such as design, R&D, and marketing.

The literature on industrial districts, thus, has demonstrated some key problems and issues of endogenous regional development. First, only few regions have really good preconditions for an endogenous development route and strategy. Second, even those regions showing good conditions for and features of endogenous development are continuously challenged by globalization and changing external conditions such as new competitors, and the change of markets and technologies. And third, in the course of these changes, external factors and strategies of development are increasingly used and combined with endogenous elements. In this process, endogenous and external factors become closely interrelated leading to a stronger local–global interchange.

Since the 1990s we have been able to observe new routes of the endogenous regional development approach. These build on some of the core ideas and develop these further, but they do not accept all of the propositions. In general, these routes are inspired by evolutionary theory, innovation theory, and by studies on entrepreneurship. There have also been a number of policy programs, activities, and studies on local and regional development initiatives, supported, for example, by international organisations such as the Organization for the Economic Cooperation and Development (OECD), International Labor Organization (ILO), and the European Commission which have centered on the concept of endogenous regional development. Examples were the LEADER and RIS/RITTS programs for regions of the EU, or the LEED program by the OECD. The concept has been promoted also by the Economic Commission for Latin America (ECLA) in South America and has been applied in countries such as Guatemala, Costa Rica, Venezuela, and Brazil. In Asia it was taken up by countries such as India and Pakistan. In most of these newer approaches and related literature there is more emphasis on entrepreneurship, innovation, and regional learning compared to earlier versions of endogenous development. Furthermore, these new policy models take more account of local–global interactions and linkages, and depart from ideas like self reliance or regional autonomy.

Entrepreneurship and Regional Development

Entrepreneurship has been at the core of the endogenous development approach since its beginnings. In recent years, however, more research and policy studies have been undertaken in this field. These cover local conditions and regional differences of entrepreneurship and new firm formation, effects on local and regional development in terms of employment and economic growth, barriers for the setting up of new firms, and related policies and their results. As we have pointed out above, entrepreneurship is a key element in endogenous regional development because new firms often originate from the region, use local talent and labor, have more intraregional input–output and knowledge links, and they have their management and other key functions within the region. They are often inserted in local social and economic networks and strongly embedded in the region. The research on entrepreneurship, however, also shows the limits of this approach and of endogenous regional development. In particular, less favored regions usually have a limited entrepreneurial potential and there are many adverse conditions and limiting barriers in those regions. Furthermore, effects on local and regional de velopment in terms of employment generation and eco nomic growth remain often small, due to the small size and slow growth of many such companies.

Regional Learning

The central idea is that globalization increasingly leads to rapid shifts of markets, production, and technologies. Regional economies are challenged to adjust rapidly to accommodate with such changes. The argument is, fur thermore, that in a world of global knowledge flows via modern information and communication technologies, the sources of competitive advantages increasingly result from local unique competencies, skills, and tacit know ledge. Such competencies and knowledge are generated and exchanged at the local and regional level often through informal relationships and 'untraded inter dependencies'. Local learning, however, has some pre conditions such as a common understanding of problems and issues, or the existence of trust for engaging in net works and knowledge exchange. As in the case of entrepreneurship, such conditions are not to be found in every region.

Regional learning has a strong connection to en dogenous regional development since it builds on local competencies and skills, local knowledge adoption and exchange, and the sharing of knowledge and of best practices. The regional economy aims at a collective enhancement of know how, and an upgrading of practices and technologies. Such practices and mechanisms of collective local learning have been demonstrated and studied in particular in industrial districts, clusters, and in innovative milieux. As key mechanisms of collective learning have been identified the mobility of qualified labor, knowledge exchange through co operations and informal networks, and the setting up of spin off com panies. A key limitation and problem of local learning as a regional development strategy is that it often leads to the adoption of technologies or practices only, or to in cremental innovation, less often to more radical innov ations. This keeps firms on their existing technology paths and runs the risk of leading to 'lock in'.

Regional Innovation Systems

This approach sees innovation as a key driver of regional development. Such innovations may be new products, processes, or organizational practices and are at least partly based on knowledge or competencies from the region. Key endogenous actors in regional innovation systems are R&D performing and innovating firms, uni versities, research organizations, and education insti tutions. Also organizations for knowledge transfer, innovation finance, and support have an important role to play. Like in industrial districts and in regional learning an intensive interaction between firms (knowledge ex ploitation) and research organizations (knowledge gen eration) is essential to a successful regional innovation system. However, it should be pointed out that regional innovation systems are not only relying on endogenous actors and institutions. They are highly open and often multilevel systems, strongly related to the respective national innovation system as well as to international firms and institutions.

Like entrepreneurship and local learning, regional innovation system strategies are not appropriate or can not be the same for every region. In particular, core re gions and large cities often have better conditions for innovation, whereas less favored regions frequently have severe constraints and regional innovation system deficits. Peripheral regions are often lacking industry clusters and organizations of knowledge generation and education (universities and research organizations). As a con sequence, many firms are competing not on the basis of innovation but on other advantages. Old industrial areas on the other hand have industrial clusters as well as universities, schools, and R&D organizations. However, these organizations and institutions are oriented on old clusters and technologies, and often reinforce problems of lock in. This implies that 'endogenous' regional in novation system strategies cannot be applied to all kinds of regions, they require at least some innovation poten tial. Furthermore, they have to be well adjusted to the specific type of region, depending on the respective in novation problems and barriers.

Conclusions

Endogenous regional development is a concept and strategy for the economic development of regions in advanced and less developed countries. Originally it has been formulated as a counter thesis to the external re gional development approach. Core ideas and elements are the understanding of development as a bottom up process, the key role of local and regional actors and initiatives, including social agents and civil society, a high importance of decision making functions, as well as of policy competencies, and of regional institutions. It is conceived as an integrated approach taking account of sectoral and other socioeconomic interdependencies, and with an emphasis on ecological aspects and on a sus tainable use of natural and other resources.

Endogenous regional development was inspired by related concepts such as industrial districts, regional learning, and innovation systems. Partly, these concepts have distinct features and elements which were pointed out above. However, they also share common elements which relate them to endogenous regional development. Such common elements are the search for local and regional specificities, uniqueness, and identity also as a source of competitive advantages for firms, the key role of local and regional institutions, social capital and networks, the key role of entrepreneurship and innovation, and the view that learning and innovation are, despite globalization and modern information and communication technologies, to a high degree local processes, based on localized tacit knowledge and its exchange.

From a historical perspective, thus, endogenous regional development was a useful, stimulating, and influential concept evolving since the 1980s. It has responded to the problems and limits of a one sided top down or external development approach. It has focused on endogenous forces and factors, including social and political processes, and it has emphasized the key role of local initiatives, entrepreneurship, and innovation. It has prepared, thus, the ground for more recent concepts such as the fostering of entrepreneurship, regional learning, and innovation. However, we should not overlook the weaknesses and limits of the concept. It was formulated largely as a 'counter concept' against the then prevailing top down model and lacked coherence as a theory of its own. It has overemphasized endogenous factors and their control, neglecting the fact that successful regional development is usually the result of both endogenous forces and external factors such as mobile capital, technologies, talent, and knowledge. This problem gets reinforced through the ongoing processes of European integration and globalization. Regions and countries are nowadays highly open systems with ever increasing external flows of goods, services, finance and capital, people, and knowledge. Furthermore, institutions and policies for economic development and innovation are increasingly characterized by multilevel governance where regional actors and institutions are inserted in and strongly related to national and international (e.g., EU) levels and institutions. This was shown (e.g., for Europe) in recent studies on the role of Structural Funds and on regional innovation systems. Under such conditions endogenous regional development cannot be applied in its original conception. Its further development towards new variants and routes such as regional entrepreneurship and learning, and regional innovation system reflects this to some extent. We have to be aware, however, that these newer variants are not applicable to all kinds of situations either. Like the original model of endogenous regional development, they tend to work better in already successful regions and countries, thus further increasing the gap to those at the bottom league.

See also: Growth Poles, Growth Centers; Industrial Districts; Regional Development and Technology; Regional Innovation Systems; Uneven Regional Development.

Further Reading

Amin, A. (2003). Industrial districts. In Sheppard, E. & Barnes, T. J. (eds.) *A companion to economic geography*, pp 149 167. Oxford: Blackwell.

Asheim, B. T. (2000). Industrial districts: The contribution of Marshall and beyond. In Clark, G. L., Feldman, M. P. & Gertler, M. (eds.) *The Oxford handbook of economic geography*, pp 413 431. Oxford: Oxford University Press.

Asheim, B. and Gertler, M. (2005). Regional innovation systems and the geographical foundations of innovation. In Fagerberg, J., Mowery, D. & Nelson, R. (eds.) *The Oxford handbook of innovation*, pp 291 317. Oxford: Oxford University Press.

Aydalot, Ph. and Keeble, D. (eds.) (1988). *High technology industry and innovative environments: The European experience*. London: Rutledge.

Cooke, P., Boekholt, P. and Todtling, F. (2000). *The governance of innovation in Europe*. London: Pinter.

Garofoli, G. (ed.) (1992). *Endogenous development in southern Europe*. Aldershot: Avebury.

Hadjimichalis, C. and Papamichos, N. (1991). 'Local' development in southern Europe: Myths and realities. In Bergman, E., Maier, G. & Todtling, F. (eds.) *Regions reconsidered Economic networks, innovation, and local development in industrialised countries*, pp 141 164. London: Mansell.

Keeble, D. and Wilkinson, F. (eds.) (2000). *High technology clusters, networking and collective learning in Europe*. Aldershot: Ashgate.

Malmberg, A. and Maskell, P. (1999). The competitiveness of firms and regions: Ubifiquation and the importance of localized learning. *European Urban and Regional Studies* 6(1), 9 26.

Morgan, K. (1997). The learning region: Institutions, innovation and regional renewal. *Regional Studies* 31, 491 503.

OECD (2003). *Entrepreneurship and local economic development Programme and policy recommendations*. Paris: OECD Publications.

Stohr, W. and Taylor, D. R. F. (eds.) (1981). *Development from above or below? The dialectics of regional planning in developing countries*. Chichester: Wiley.

Stohr, W. B. (ed.) (1990). *Global challenge and local response Initiatives for economic regeneration in contemporary Europe*. London: Mansell.

Storper, M. (1995). The resurgence of regional economies, ten years later: The region as a nexus of untraded interdependencies. *European Urban and Regional Studies* 2, 191 221.

Todtling, F. and Trippl, M. (2005). One size fits all? Towards a differentiated regional innovation policy approach. *Research Policy* 34, 1023 1209.

Vazquez Barquero, A. (2006). *Endogenous development Networking, innovation, institutions and cities*. London: Routledge.

Regional Geography I

A. Paasi, University of Oulu, Oulu, Finland

Glossary

Chorology The study of the causal relations between different geographical phenomena occurring within a particular region. According to Hettner and Hartshorne, the chorological perspective aims at studying the character of regions through comprehending the existence together and interrelations between different realms of reality and their varied manifestations.

Exceptionalism An idea promoted by chorological thinkers that geography and history differ from other scientific fields because the former do not have a specific research object but rather look at the world from a specific viewpoint. Accordingly, geography studies the world in relation to space and history in relation to time. This idea was based on the philosopher Kant's classification of sciences.

Geographical Region The 'geographical region' was a significant idea in traditional regional geography. It is a regional unit that the researcher identifies as a result of the research process. The representatives of traditional regional geography struggled to identify regions as bounded units that are totally separate from each other. This idea has been challenged by new regional geographers and relational thinkers.

New Regional Geography A subfield of human geography whose representatives are interested in the social and cultural construction of regions, in their meanings for individual and social identities, and in the power relations that are involved in region-building processes.

Region The traditional key category of geographical thinking: a spatial unit that is somehow distinguishable from the surrounding areas. It has been understood both as a mental category that can be used in classification and as a 'real-world unit'. For new regional geographers, the region is an entity based on social practice and discourse.

Regional Geography According to a traditional definition, a subfield of geography that studies the diversity and the organization of nature and human aspects in an integrative regional framework.

Regionalization In traditional regional geography, 'regionalization' referred to the process of dividing space into regions. Regionalization also points to the political management of state space and, for instance, the decentralization of activities. In the global context, it may refer to the proliferation of regional blocks at a supranational scale.

Region Building The process through which regions and their meanings are socially constructed.

Systematic Geography It consists of systematic subfields of physical geography such as geomorphology, climatology, or biogeography; and of human geography such as population geography, urban geography, or political geography.

Introduction

"Regional geography – we don't teach that any more." This comment was posed by a senior colleague to Hugh Clout in the late 1980s when Clout mentioned that he was preparing a trend report on regional geography in the UK. Paradoxically, this comment, recalling many earlier but equally premature declarations on the death of regional geography, was offered in the middle of a new wave of interest in regional geography in the UK, something that was soon to become internationally recognized as 'new regional geography'. Furthermore, human geographers at the same time were also beginning to pay increasing attention to new possible ways of thinking regional geography, not only in the UK, but also in many other contexts such as the US, Germany, the Netherlands, and Finland.

This article scrutinizes both the long tradition of regional geography and the rise and key features of new regional geography. It aims to show how both the practice of making regional geography and regional geography as a theoretical enterprise have been deeply contested, and historically and geographically contingent. This disputability evinces the fact that both the practice of making and the discourses about regional geography have resulted from entangled societal and academic power relations. The likewise contested understanding of what the 'region' is has been part and parcel of these struggles. Due to these facts, the prime focus of this article is on methodological issues and, relatedly, in tracing how contested concepts of the region have been unfolded and utilized in geography. The article first looks at the contested features and definitions of regional geography, and then traces the roots and shaping of traditional regional geography, as well as examining the forms of critique that emerged after World War II. It then maps the background and key ideas of the so called new regional geography and the challenges that it has posed to geographical thinking. Finally, the challenge of globalization to regional geographic thinking and to the idea of region is reappraised.

Regional Geography: A Contested but Necessary Enterprise

Regional geography has been a highly significant source for both theoretical debate and concrete research practice by geographers over the course of many years. But, in spite of this creditable past, the phrase 'regional geography' often raises mixed feelings among contemporary geographers who have effectively enforced an increasing specialization in the field, and have witnessed the philosophical, methodological, and theoretical fragmentation of the discipline that has taken place since World War II. While most academic geographers doubtless appreciate the informational value of regional geography, including its role in the history of the field, and while many of them teach regional geography courses and some occasionally put forward pleas on the need to develop this field further, very few are ready to label themselves as 'regional geographers' today. And this is also the case when a regional emphasis may crucially inform their research, whether they are working with the region as a theoretical category or with regional issues related to such themes as urbanism (e.g., city regions), political geography (e.g., devolution or regionalism), or cultural geography (e.g., regional identity).

This inbuilt ambivalence, however, is not a new feature. In the later 1960s, Roger Minshull scrutinized the methodological problems that regional geographers addressed in a bewildering variety of different ways: that is, the methods of finding, defining, mapping, and describing the region. This vagueness goes even further back in disciplinary (and pre disciplinary) history. Some scholars had already suggested in the mid 1930s that there were nearly as many concepts of the region as there were regional geographers. Indeed, it is obvious that the ongoing debate on the idea of region has itself provided, in many countries, one basis to legitimate the existence of geography as a discipline in the academy and, consequently, the idea of region has become a major symbol for the identity of the field, a symbol that is in practice often taken for granted. For many scholars, regional geography has simply been an academic territory that serves as a safeguard against the absorption of geography as part of other academic fields.

The focus of the methodological problems recognized by Minshull has changed a lot since the 1960s. New interpretations of the region have perpetually emerged, often claiming the status of cutting edge knowledge, just as soon to fold under the next flow of dominant ideas. However, the meanings of the region, the importance of mapping it, and the power relations hidden in regional representations have all become, and for some time now remained, crucial features of regional geographic studies. Also, while the idea of a specific 'geographical region' that was for a long time regarded as the key product of a regional geographic research process *did* lose its importance in academic geography after World War II, the conceptualization of the region has continually motivated (regional) geographers and still does so today.

The geographical regions were originally identified in two different ways: by proceeding from small units to larger ones, and vice versa. The idea of the geographical region was not as narrow as was the idea of a 'natural region' that was in use in many countries before World War II. Most prominently, this idea arose in A. J. Herbertson's (1865–1915) efforts to distinguish large scale natural regions of the world by using primarily climatic indicators, an idea based on the supposed existence of 'natural boundaries'. This exploitation of a physical geographic basis in the delimitation of regions also manifested itself in political geography and boundary studies, where scholars discussed so called 'natural boundaries', and this idea was even apparent in political practice when state leaders and military authorities sought to negotiate, fix, and defend such boundaries through diplomacy and war. Regions and their boundaries were regarded as being dictated by nature, and in the extreme case regions were regarded as organisms that expand in the same way as biological organisms. While the notions of natural boundaries have been abandoned, current world regional geography textbooks – important indicators of the current, institutionalized understandings of regional geography in academia – still remind us of the tendency to divide the Earth surface into such large scale components, and their authors often lean on some macro scale regional frameworks simply to manage the complexity of multiple materials. This regionalization is also almost self evident in area studies: think of such expressions as Global South, Latin America, South Asia, Sub Saharan Africa, or Europe.

The aim of regional geographers to distinguish regional spaces is related to such continuing important social practices as regionalization and regionalism. Regionalization as a tendency to distinguish regions by dividing or integrating space is consistently important in both regional planning and the governance of the spatiality of social life. Think about, for example, the current tendency toward devolution in many states or the efforts to create suprastate institutional regional arrangements such as the European Union (EU) or North American Free Trade Agreement (NAFTA). Also regionalism, a cultural or political movement based upon regional features and interests, often draws on regional knowledge and at times on customary ideas of a certain regional unity. Regionalism has been, and still is, a significant basis for science, art (painting, literature), and political movements around the world. It has also been an important source of inspiration for many traditional regional geographers, such as E. W. Gilbert. Contrary to much traditional inwards looking regionalism, the so called 'new regionalism' that has emerged since the late 1970s has had a more open character, often being a vehicle for regional development strategies in a world where governance and power relations are rescaling. Both regionalism and regionalization are social practices that bring together political geographic and regional geographic aspects.

The most important and enduring site for the application of regional geographic knowledge, and at the same time one perpetuating the interest of geographers in questions about regional geography, has nevertheless been education and the production and reproduction of geographical world views in academic, educational, and popular circles. Generations of students in schools and university geography departments around the world have studied regional geography; and they still do so, whether under the label regional geography, area studies, world regional geography, or specially named courses. These courses exist in the syllabus at all levels of education, and hence have a strong niche in the academic and school geography textbook publishing market. The historical background here is the fact that it was not only the inquisitiveness of teachers or professionals, but also the passions of state and military officials, political leaders, and nation builders, that turned the world's attention to regional knowledge as a medium in both social reproduction and the construction of national identities. When geography entered universities during the late nineteenth century, the key ideological backgrounds for this step were nationalism and colonialism. Geographic education has hence been exploited not only to mediate 'geographical facts' but also to construct, and to reproduce, regional and national stereotypes (and at times even 'enemy' images). This politics of representation accentuates the connection between regional geography and power.

Regional geography thus has a potentially close relation to questions of racism, peace, and violence, themes studied currently by cultural, political, and postcolonial geographers. The need for such research was already suggested by United Nations Educational, Scientific, and Cultural Organization (UNESCO) after World War II. This theme is closely engaged with cartography, and perhaps in no other field of geography has the power of maps been as obvious as in regional geography. Regional depictions and narratives – always based on certain choices and simplifications from the complexity of the world – apparently 'objectify' the regional worlds that they are representing, and this effect is especially powerful when these worlds are inscribed on maps that 'fix' them as apparently timeless (and stop the historical process that a region actually *is*). Regional geographic education has effectively produced and reproduced, often by leaning on cartography, the image of a homology between cultures and bounded spaces, thereby forging a sense of 'us' ('here') and the 'Other' (or 'them, there'). Maps have thereby been a key vehicle to naturalize the boundaries between cultural and political realms. This practice of subjugating and homogenizing complex cultural relations in given spatial grids has been severely criticized, not only by geographers but also by anthropologists, during the last few decades. In spite of this critique, the link between regional geography and education has been a surprisingly neglected theme in critical human geographic research.

Defining Regional Geography

What then is 'regional geography' by definition? This has been and still is a deeply contested question. R. W. Dickinson defined regional geography at the turn of the 1970s as the treatment of a variety of spatially distributed phenomena in a particular area, whether it is a local area, a country, or a continent – thus implying that regional geographers operate at all spatial scales. The challenge for the regional geographer is, he argued, to discover integrating processes that give some measure of identity and uniqueness to an area. For him, regional consciousness and the names of regions, for instance, are important in 'mapping' such an identity. Regional consciousness and regional identity have been significant topics, particularly among continental European geographers, up to recent times.

The aforementioned E. W. Gilbert regarded geography in the early 1960s as the art of recognizing, describing, and interpreting the 'personalities' of regions. Regions were not static for him: rather, the 'characters' of regions are constantly changing and developing. For J. F. Hart (1982, p. 2 and p. 22), 20 years later, the task of regional geography was to produce "evocative descriptions that facilitate an understanding and appreciation of places, areas and regions;" and regions, for their part, "are subjective artistic devices, and they must be shaped to fit the hand of the individual user."

Whereas regional geography ceased to lie at the core of methodological debates in the discipline after the 1950s, the character of regional geography has often been discussed in nostalgic ways ever since, notably in presidential or memorial addresses. Some of these contributions have become widely recognized, like the address by Hart cited above or two addresses from the early 1960s: H. C. Darby's talk to the Institute of British Geographers or E. W. Gilbert's Herbertson Memorial Lecture. The implication is that regional geography has been regarded as valuable, something that perpetually reminds us of the time when geography was supposedly characterized by some sort of unity. Indeed, the idea of region, while always a contested category, has brought along a list of terms that have been related, or appealed, to this idea of unity. Expressions such as synthesis, uniqueness, total composition or complexity, compage, holism, whole, *Zusammenhang*, *Ganzheit*, individual, totality, organism, or personality, for example, have all been exploited by geographers over time to convince the community of geographers, and perhaps also others, of the power of the region as a unity or an organizing principle that brings nature and culture together in a regional context. For many scholars, the region apparently carries a certain primordialism, hence recalling the idea of 'community' in sociology that has generated similar massive debates. This ideal of unity can be clearly seen from the comments of S. W. Wooldridge (1956, p. 53) when he suggested that the aim of regional geography is "to gather up the disparate strands of the systematic studies, the geographical aspects of other disciplines, into a coherent and focused unity, to see nature and nurture, physique and personality as closely related and interdependent elements in specific regions."

Figure 1 The cover of the *Geographia Generalis* by Berhard Varenius. The division between general (systematic) and special (regional) geography was introduced in this book. From the Library of the University of Oulu, photo by Anssi Paasi.

The Roots of Regional Geography

The roots of regional geographic thinking are often traced back to the classical period, and Ancient Greek scholars such as Strabo (63/64 BC–*c.* AD 24) are recognized as forefathers of *chorographic* thinking. Chorography comes from *khŏros*, 'place', and *graphein*, 'writing', thus meaning literally writing about place. Much later, the German scholars Alexander von Humboldt (1760–1859) and Carl Ritter (1779–1859) were both developing comparative methods for geography, a field that was based on empirical work. Whereas Humboldt developed his research perspectives in the tradition of natural science by collecting data and making generalizations, Ritter was more committed to a historical approach and tried to identify separate regions, to find a unity in diversity. Humboldt's *Kosmos* and Ritter's massive series on comparative geography, *Erdkunde*, are often recognized as the founding works of modern geography. Whereas Humboldt is seen as the father of systematic geography, Ritter was the key person to develop regional geography. This division between 'general' or 'systematic geography' and 'special' or 'regional geography' has long roots that many scholars are ready to trace back to the classic work *Geographia Generalis* (**Figure 1**) written by the German Bernhard Varenius in the seventeenth century. These two subcomponents of geography have not been regarded as separate fields of inquiry, but instead have constituted a unity for many scholars, two sides of the same coin, and both have been regarded as significant for geographic research.

Regional geography has been a crucial notion in geography since the discipline as a whole became an institutionalized academic subject at the end of the nineteenth century. It has had somewhat different meanings and trajectories in various national contexts, but what was common in many contexts was that the first 'geographers' were recruited from other fields, like geology or biology. The ideas of region and regional geography seemed to provide a basis for a disciplinary identity, and for many geographers working in British, French, German, or North American universities it was for a long time the core or 'crown' of the discipline, a synthetic approach and principle that intended to bring together the formally separate subareas or systematic branches of geography, such as geomorphology, climatology, biogeography, urban geography, economic geography, and social geography. It may be summarized that, after the institutionalization of geography, regional geography dominated over systematic approaches until the 1940s–1950s, whereas since then systematic approaches have gradually taken a dominant position, albeit not leading to the total ousting of regional geography.

Contrary to the current tendency where Anglo-American theories and approaches are argued to dominate international human geography, the origin of regional geography in Britain (UK) and North America (especially the USA) largely mirrored developments in Germany and France, two academic contexts where many geographers are prepared to locate the roots of modern geography. Clout points out that in the UK, for example, most lecturers in university departments during the 1930s–1950s derived their inspiration for regional geography from the French classics. Paul Vidal de la Blache (1845–1918), the founder of modern French geography, argued emphatically for regional geography as the core of the discipline. In France, the ideal context for the then popular regional monographs were the *pays*, small scale regional units that corresponded to a distinctive landscape assemblage based on the close connection between the resident human beings and their physical environment, and on the ways of life (*genres de vie*) that emerged from this relation. It was supposed that this unity fostered the strong regional identity and identification felt by the inhabitants of their respective *pays*. The study of such unique regions was seen as the key task of the geographer, and French geography was characterized by a relatively high degree of unity before World War II, a fact based on the predominating influence of Vidal.

Vidal's approach was inductive and historical, but in some other contexts the key arguments behind regional geography were more theoretically grounded. Certain key representatives of chorological thought, such as Alfred Hettner (1859–1941) in Germany, Richard Hartshorne (1899–1992) in the USA, and many other scholars in Germany, Russia, Japan, Belgium, and Switzerland, based their arguments on the principle of there being a rational and logically defensible division of labor between scientific fields. Both Hettner in his *Die Geographie, ihre Geschichte, ihr Wesen und Ihre Methoden*, and Hartshorne, following Hettner, in his *The Nature of Geography* (see **Figure 2**) were drawing on the philosopher Immanuel Kant's (1724–1804) classification of sciences whereby geography was distinguished from other fields because, like history but contrary to all other major academic fields, it did not have any particular research object that would characterize its own field. Rather, it was their approach that distinguished geography and history from

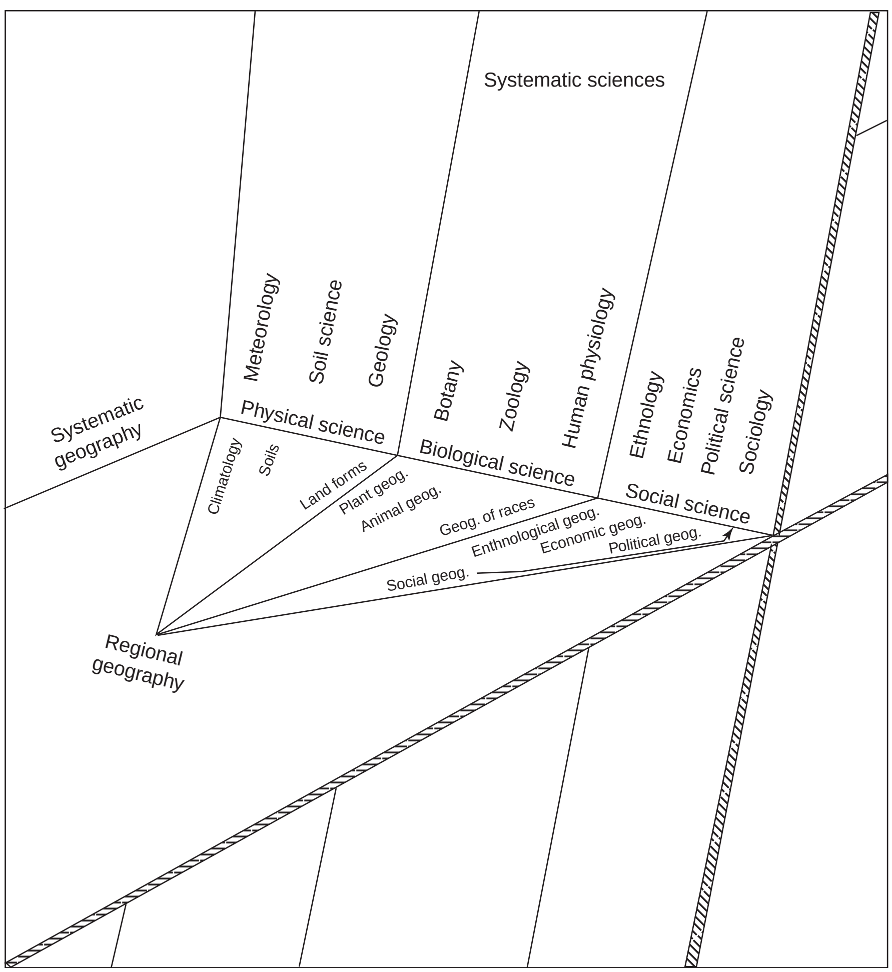

Figure 2 Richard Hartshorne's view on the relations between geography and systematic science. The figure was included in his *The Nature of Geography*.

other fields. According to this thinking, often labeled as 'exceptionalist', history was looking at the world in relation to time and geography in relation to space.

For chorologists, regional geography was a way to enact *the* geographical method: that is, to study the natural and human worlds according to their areal differentiation, stressing the 'areal' differences between regions as 'spatial sections of reality', and to map out (describe and explain) the causal relationships obtaining within the distinctive assemblages of phenomena within (and in effect constituting) different regions. Accordingly, the region was regarded as a crucial element for the identity of the discipline as a synthetic field, but also instrumental for the analysis of the differentiation of the Earth surface. Hettner's theoretical contemplations suggested a close relation between regional and systematic geography, and his reasoning on geography as a chorological science of regions paved the way for research, with a number of scholars following his ideas with great interest. Contrary to the situation in France, German geographers were not primarily studying Germany and its areas, but were often carrying out regional research overseas.

Much time has been devoted by 'traditional' geographers to reflect upon whether regional geography should be regarded as a part of the proper 'science' of geography or whether this should be rather seen as a form of 'art' or applied geography. This debate particularly characterized the pre quantitative tradition of geography, summarized by Hartshorne in his *Perspectives on the Nature of Geography* in the late 1950s. He reminded readers that we should not confuse either the regional concept or the regional method with what we commonly call regional geography, since the former tends to be used in every facet or at every level of geographic study, along the continuum from that of the studying the most elementary integrations (the extreme 'topical' or 'systematic' approach) to that of studying maximum integration (the extreme 'regional' approach). The confusion results from the fact, as Hartshorne noted, that we employ the word 'region' for a number of different concepts, each of which bears a different relation to the contrast between topical and the regional approaches.

In several countries, regional, historical, and landscape geographers have devoted much time to pondering over what a 'good' regional geography would entail. In some contexts, like in France, the classical regional monograph became highly significant, and practicing regional geography was regarded as a very exacting task where scientific knowledge, cartographic skills, and literary art were brought together. Some scholars, like the German Alfred Kirchhoff, tried to create a rather mechanistic scheme (*Das Länderkundliche Schema*) for carrying out regional geography, while another German, Hans Spethman, rejected this schema and struggled to create a dynamic regional geography concentrating on forces that were shaping the regions of the Earth. Regional geography has also been a crucial notion in relation to another key category, *landscape*. Whereas in the English language usage, landscape is above all a visual category, in the German and Scandinavian context, for example, region and landscape have been brought together in such categories as *Landschaft* or *Landskap*. This idea has not merely been developed in this context, but was also adopted by some historical geographers elsewhere. The most prominent representative of such thinking in the USA was Carl Ortwin Sauer (1889–1975), who suggested that the task of geographers should be to trace the transformation of natural landscapes into cultural landscapes and the areal connections of this process. In this approach, he developed the ideas suggested earlier by the German geographer Otto Schlüter.

The relation between region and landscape again brings in the question of art, both in the sense of skills and artistic or aesthetic representation. These problems were discussed by many authors in, for instance, France, England, Germany, and Finland. This theme was significant, particularly in German geography, perhaps reflecting the dominant continental hermeneutic and interpretative philosophical tradition as well as the importance of neo Kantianism. At the extreme were some German geographers like Ewald Banse (1883–1953) who claimed that a creative geography should be artistic, and he even wrote a small book on geography and expressionism in 1920. Hettner, who also discussed the relations between science, art, and esthetics in many publications, labeled Banse the *enfant terrible* of geography who believed that he was geography's 'Schopenhauer'. An interesting detail in the history of the discipline is also that such a towering German figure in natural science as Humboldt recognized the importance of artistic description. Similarly, the father of political geography, anthropogeographer Friedrich Ratzel (1844–1904), discussed at length the role of art in understanding nature in his posthumously published book *Über Naturschilderung*. Gilbert (1960, p. 159) suggested in the UK that "regional geography is an art comparable with other arts; both poetry and painting can and do portray landscape… if they [geographers: A.P.] are to make their regions vivid and alive, they should possess something of the expressive and sensuous imagination of poets and painters…." Hartshorne (1959, pp 171–172), for his part, was much more categorical in his opinions and claimed that "Effective description in geography … involves no small degree of art, not in the sense of subjective impressions, but in the objective sense of discernment and insight based on knowledge of those relationships that can be known." The issue of the artistic character of regional geography was hence raised by a number of scholars in various countries, and it has also been significant in so called 'literary geography' where novels and poems are crucial materials in the efforts of geographers to under stand and to represent regional diversity.

The Decline of Traditional Regional Geography

The fate of regional geography was different in various countries after World War II. In the US and UK, sys tematic approaches drawing on new theoretical, law seeking approaches, quantitative methods, and statistical data soon took a leading position in the analysis of the rapidly industrializing and urbanizing capitalist states, and this led to a decline in, and the replacement of, regional geography as an esteemed research field. Human geography gained new applied, instrumental forms. The gradual decline of traditional regional geography as an academic subject was based on several factors. Along with the rise of systematic areas in geography, the develop ment of quantitative methods and the adoption of new positivist philosophies, the representatives of this 'new geography' also formulated an explicit critique of re gional geography. Old concerns about the subjective judgments hidden in the methods of regional geography, and on the 'incestuous', inwards looking exceptionalist approach, were taken to a new level when they were combined with the philosophical claims of positivist conceptions regarding the correct routes to scientific explanation. In effect, the traditional geographical per spective, concerned with the real minutiae of world re gions, was now challenged by a spatial science separated from 'concrete ground'. Instead, then, it was abstract spatial relations, networks, and forms of interaction that would be studied in order to recognize general patterns, supposedly holding everywhere, not the unique and seemingly static material features of bounded regions.

In his posthumously published text of 1953, Fred K. Schaefer, a refugee from Nazi Germany to the US, challenged the descriptive regional geographic approach by suggesting that "it seems to me that as long as geog raphers cultivate its systematic aspects, geography's prospects as a discipline of its own are good indeed. … I am not so optimistic in case geography should reject the search for laws, exalt its regional aspects for its own sake and thus limit itself more and more to mere description" (Schaefer, 1953: p. 249). In the same spirit, in 1971 Ronald Abler and his co authors wrote in their much used textbook *Spatial Organization: A Geographer's View of the World* that "For a long time, geographers tended to be lieve that internally homogeneous regions, based on one or a number of place characteristics, existed independ ently of the principles which defined them, and that regions themselves were intrinsically worthy of delimi tation and study" (Abler *et al.*, 1971: p. 72). And they then added that, by the time of World War II, "it was about that time that geographers realized that further progress in geography depended upon generalization and analysis at levels more advanced than classification." Some rep resentatives of the new quantitative geography had a very negative view of the past: indeed, Peter Gould moaned in the late 1970s that, "with the exception of one or two works of scholarship in historical geography, it was practically impossible to find a book that could be put into the hands of a scholar in another discipline without feeling ashamed" (Gould, 1979: pp 140–141).

This new spirit rapidly spread to geography depart ments outside of the US and UK: what was now im portant were models, relative spaces, and distances. This shift was interpreted in highly progressive terms and rhetoric as the most fundamental change in the history of geography, since, as Abler and co authors suggested, it opened up an almost infinite number of new worlds to explore and to map. In locational analysis (or spatial science) regions still existed, but they now became in effect data bound cells in the spatial matrix that could be classified and analyzed by statistical methods, and re gionalization increasingly became classification and the making of taxonomies. It is important to recognize in this context the early, seminal role of such authors as Walter Christaller (1893–1969) in Germany and Walter Isard (1919–) in the USA, as well as the works of Torsten Hägerstrand (1916–2004) in Sweden, who were all pi oneering quantitative and theoretical geography and re gional science. In spite of his important role in this context, Hägerstrand was also deeply interested in the value of regional geographic perspectives, in which capacity his ideas reappeared in versions of a new re gional geography (as will be explained below).

In France, the regional geographic style and the re gional monograph as the crown of scholarship, both created very much in the Vidalian spirit, were still strong after World War II, and it took longer to replace them by new approaches. This is an interesting delay, since, as noted by Nigel Thrift, during the heyday of Vidalian geography the locally distinctive ways of life of the French peasantry were already disappearing under the onslaught of both capitalism and the activities of French state. The 1960s heralded an active debate in France on the purpose and methods of regional geography, and on the substance of human geography, and the emphasis was gradually turning toward questions of interaction and spatial distribution patterns. But, as been shown by Paul Claval, the region and regional geography still had a role to play in this national context.

The rise of new theoretical and quantitative ap proaches did not mean the end of regional geography in the 1950s–1960s. This persistence was based on the many roles of regional geography. Regional knowledge was, and still is, absolutely crucial in schools and higher education. The ideas of the region and the potential future roles of regional geography have been discussed frequently in geography journals, at times in connection with geography teaching, at times in connection with discussions about the methodologies of geography. Pleas for a balance between regional geography and a new technical and quantitative approach were already being set forth in the late 1960s when regional geography was declared 'dead' by the representatives of spatial science. It is also to be noted that, in spite of the major upheavals in geographical research practice, a strong regional undercurrent remained in North American cultural geography, where certain forms of regional geography lingered, and even continued to flourish, in the studies of vernacular regions and regional identities.

In Germany, some representatives of the strong tradition of *Sozial Geographie* continued to develop regional geography, but quantitative approaches did also diffuse here. The early 1970s witnessed a lively debate on the roles of regional geography in Germany, partly reflecting the generational gap, and many authors claimed that regional geography was still valid and needed. Since the late 1980s the region and regional approaches have therefore remained important in German geography, and geographers have studied subject matters such as the hermeneutic roots of regional geography and the regionalization of everyday life. British and Dutch geographers organized a binational symposium at the turn of the 1990s to develop new approaches in regional geography.

The Emergence of 'New Regional Geography'

One of the fundamental ontological and epistemological questions in traditional regional geography – and one central to understanding how such questions differ from the theoretical discussions raised later in new regional geography – has been whether regions do 'really exist': that is, are they visible and demonstrable realities or are they merely ideas or 'subjective artistic devices', mental entities which geographers develop and use in the classification of the variety of the world? This topic had been discussed extensively by Hartshorne and many later authors. For some scholars like George Kimble, the idea of the region was vague because such units simply do not straightforwardly exist in reality, an argument that was also used by Hartshorne when logically deconstructing the nature of geography. For Hartshorne, regions were neither objects nor phenomena of the material world, but rather creations of the student's mind and thoughts. Here, the region is cast as both an absolute space and an abstract field of experience where things and processes exist. There have been two forms for such constructions: the 'generic region', corresponding to the positivist mode of classification, and the 'specific' or 'unique region', forming the core of the exceptionalist thesis (and which has arguably been the more philosophically and methodologically testing).

Indeed, such questions on the nature of regions are perhaps more problematic than, superficially taken, they seem to be; but they also help us to understand some features of the so called new regional geography. A logical step further is to question what kind of philosophical and methodological commitments do we possess or reveal when answering such questions. For instance, does the ontological aspect deeply involved in questioning the nature of regions merely refer to the problem of empirical observation, so that the existence of regions would be based solely on the ability of researchers to 'find' regions in the world? The implications of such thought have long prevented the development of the geographer's concepts of region in other respects too. Central problems have been met in contrasts drawn between the understanding of regions as nonhistorical frames or containers, in which the various phenomena of nature and culture really are arranged in varying substantive fashions, and the conceptual reduction of regions to mental categories which the researcher creates on the basis of his or her research motives.

A theoretically oriented new regional geographer does not limit the question on the nature of regions merely to problems of classification or definition as such. A far more challenging issue is that taking seriously the precise ontology of the region forces scholars to trace abstractions by which they can reveal the crucial aspects of the 'regionality' phenomenon and, ultimately, 'draw' out, as 'objects' for concrete research, the regions which have been produced, reproduced, and maybe ultimately destroyed within various social and cultural practices and discourses. For a new regional geographer, both the questions and answers regarding the existence and manifestation of regions are inevitably based on social *practice*. Consequently, then, the representatives of new regional geography have been interested in the power relations, practices, and discourses through which people, social communities, and classes produce and reproduce 'regions' and localities in their daily life through various institutionalized practices, such as politics, administration, economy, education, media, communication, and so on. This complexity shows that new regional geographers often adopt an emancipatory interest. Hence, there is a fundamental step from – to use the concepts of Jürgen Habermas – a 'technical' interest in knowledge to a 'practical' (hermeneutic) and an 'emancipatory' one; a shift from collecting and presenting information on the facts of the 'objective regional world' to exploit as a technical device in the control of nature and society, to understanding the social and cultural practices and power relations inherent in region building processes, and ultimately to determining the social and psychological fetters which confine society, and in this way achieving mastery over society, so as to release people from their control.

Relatedly, geographers have been forced to rethink the question of the 'objectivity' of regions and to conceive them as processes performed, limited, symbolized, and institutionalized through numerous practices, discourses, and power relations that are not inevitably bound up with some specific spatial scale but may indeed 'stretch' in both time and space. A typical feature in new regional geography has been that it does not prefer any specific spatial scale: the understanding of regions requires recognizing and analyzing the processes which take place in and between different scales. This is a useful perspective for understanding both the complexities of current spatialities and the inherent relations of power. The emphasis on historical approaches, moreover, has not meant merely 'to take history seriously' or to place emphasis on past processes, but instead to insist upon the problematization of all assumptions concerning society, human beings, and social change. The works of many new regional geographers show that there is no reason to distinguish between historical and other geographies: regions and the making of them are always ongoing, never completed processes.

Whereas traditional regional geographers studied regional worlds consisting of a supposedly rich mosaic of regional entities, the challenge for the revival of regional geography is to develop methodological approaches for studying broader social and spatial transformations taking place in a world characterized by the increasing dynamics of the globalizing capitalism. Many scholars now pay attention not only to how regions and region building processes are materially embedded and constituted, as well as 'stretching' across supposed regional borders, but they also closely examine the regionalities and regionalizations of social and everyday life. Some geographers begin to consider carefully such questions as the sense of belonging, structures of feeling, loyalties, or mobilization of memory in a regional context. Regions and regionality thus start to appear as constructed in the dialectics of materiality and individual and social imagination. Since all of these elements are in flux in a globalizing world, this realization has forced scholars to develop new perspectives for regional research and also to look seriously at relevant contributions from social and cultural theory.

This transformation and reorientation initially took place during the 1970s–1980s. In the middle of exchanges between the still dominant positivist approaches and the emerging humanistic and Marxist critiques, Derek Gregory set forth – in his 1978 text *Ideology, Science and Human Geography* – the claim that, "Ever since regional geography was declared to be dead – most fervently by those who had never been much good at it anyway geographers, to their credit, have been trying to revive it in one form or other. … This is a vital task" (Gregory, 1978: p. 171). He suggested, in the spirit of an emerging Marxist vocabulary, that the challenge for the geographer is to study the constitution of specific, regionally embedded social formations, articulations, and transformations. Following such pleas to study regions in theoretically informed ways, the 'new regional geography' became an important category during the 1980s. This label, proposed by Thrift in 1983, became popular, along with Anne Gilbert's review in *Progress in Human Geography* in 1988. Gilbert brought together various theoretical perspectives on the region from both English and French speaking geography, including Marxist and humanistic conceptions, alongside approaches based on theories of practice. The latter particularly directed the theoretical attention of human geographers toward the works of certain sociologists. Especially significant sources of inspiration were the 'structuration theory' developed by the British sociologist Anthony Giddens and the theories of practice developed by the French sociologist Pierre Bourdieu, both of whom are discussed in the article on 'structuration theory'.

While some geographers saw new regional geography simply as a project 'coming from the left', Gilbert made a distinction between three different approaches in what she too called a new regional geography. The first approach may be labeled as 'materialist', looking at the spatial organization of social processes associated with a specific mode of production, and here it concentrates on the political economic basis of regions. It places stress on the role played by the logic of capital circulation within these social processes. This interpretation was typical for geographers who adopted a Marxist viewpoint, and perhaps the most illustrative examples were some of Doreen Massey's works from the late 1970s and early 1980s. For her, the region was a consequence of uneven economic development and successive, overlaying rounds of capital accumulation expressed spatially. She wrote in 1978 how:

> The approach which is suggested here begins from the process of accumulation and analyses the production of spatially uneven development without any pre specified regionalization of that space. From analysis on accumulation, it produces concepts of geographical organization in terms of the spatial division of labour. … In general terms, there is probably an increasing degree of agreement that analysis should start from accumulation rather than regions. (Massey, 1978: 114, 116)

The second approach identified by Gilbert begins from the region as a setting for social interaction, and typically argues that this setting or medium plays a fundamental role in the production and reproduction of social relations. As a result, space (as well as time), embracing its symbolic and ideological dimensions as well as its material basis (nature, economy), is understood as a social and cultural construct. Space and spatial patterns are not independent of social, cultural, and natural (or ecological) processes. Space is not seen as a causal power which would as such determine social processes, however, and rather social (and cultural) and spatial are taken as constituents and outcomes of each other.

The representatives of the second approach have been typically inspired by the ideas of structuration theorists, mainly Giddens, or have developed their own ideas on this basis. In general, these approaches accentuate the need to recognize how human agency and societal institutional structures are produced and reproduced in a dialectical process, and how diverging power relations are involved in this process. The first example of the use of these approaches is Allan Pred's theoretical effort to figure out place or region – he did not make any distinction between the two! – as historically contingent processes. Pred combines a structuration theoretic viewpoint with Hägerstrand's 'time geography', and also draws upon Vidalian thoughts about the importance of local *milieu* as a context shaping people's ways of life. His empirical examples are based on the historical geography of Sweden and the daily life patterns of local peasants. Another example is Anssi Paasi's work on the institutionalization of regions. For him, the region is a social process with a beginning and an end. From this process, he conceptualizes four abstractions: territorial, symbolic, and institutional shaping, and the establishment of the region in the regional system and social consciousness. In social practices, such as economy, politics, governance, media, or education, regions acquire their boundaries, which are important as social institutions and symbols rather than just as physical lines. Similarly, the often contested symbolism of regions is created, together with the institutional practices that are crucial in the maintenance of territorial and symbolic shapes. Regions, their boundaries, symbols, and institutions are hence not the result of autonomous and evolutionary processes, but instead expressions of a perpetual struggle over the meanings associated with space, representation, democracy, and welfare. Paasi's studies analyze the construction and meaning making of regionality in the case of Finnish provinces and, in later work, the Finnish national state.

The third approach identified by Gilbert stresses culture as the prime point of departure, concentrating on problems such as regional identification and regional identities. Here the region is understood primarily as a set of cultural relations between a specific human group and a particular place; it is a people bound category, conceived less with regard to individuals but rather more with social communities. Human geographers sometimes interpret these kinds of themes nowadays by referring to the idea of place rather than region. Along with the rise of humanistic geography, place became a significant category, but the usage of this category has expanded since the 1990s so that the element of experience accentuated by humanists like Yi Fu Tuan or Edward Relph ceases to be the key dimension. It is instead the socially constructed character of place, and the 'power geometries', in Massey's term, that are involved in this construction that draw the attention of scholars. Another important idea that is associated with new regional geography is that of 'regional identity', which does not refer to any search for separate, unique regions. What is important in this respect is the process through which such identity narratives are created and how they are used by individuals and institutional actors in the networks of power manifesting themselves in social and cultural life.

Overall, this new regional geography has been a diffuse phenomenon, something that was already obvious in the late 1980s, when the first reviews of this area of concern appeared. It is still a somewhat ambivalent brand: while many authors have noted the need to reconceptualize region and place, very few have been that explicit in proposing a new regional geography. It may thus be argued that new regional geography has not become a coherent approach. Rather, it is an umbrella term for research and for reflecting how region and/or place are constituted by and constitutive of social life, social relations, and identity. As we have seen above, one important and dynamic feature of regional geographic thinking since its infancy has been methodological discussions regarding the concept of the region and the processes of regionalization. The arguments of new regional geographers are merely one fold of this seemingly never ending discourse. It is hence no wonder that some authors, like Ron Johnston, have been ready to claim that, rather than a 'new regional geography', we actually need a more developed study of 'regions in geography'.

Regional Geography and the Concept of the Region

Sociologists like Bourdieu suggest that language plays a key role in scientific practice. Scientific facts, arguments, and theories are mostly constructed, communicated, and evaluated in the form of written statements, which simply means that scientific work is largely a literary and interpretative activity. The understanding of what the key categories are in scientific discourse is historically and spatially contingent, not merely because these categories are embedded in different languages but also because of different institutional contexts where scientific work is carried out, and because of existing societal and academic power relations. Even if theories may travel from one context to another, in social sciences they are at least partly context dependent. What German geographers, for instance, have meant by the region, and how regional geography has then been recognized and practiced in this context, has differed between the British, French, Finnish, and other contexts.

Preston James once wrote that words, whether they are technical ones or not, are valuable only as symbols standing for ideas or concepts. In academic geography, region, place, space, and some other related categories, such as location, distance, and direction, have been identified as such key categories. As we have seen above, the region has probably been the most powerful of concepts in geography, since it has been the central intellectual reference point for both the concrete research process, including the presentation of results (maps, pictures, text), and the philosophy and methodology of the discipline, meaning theoretical arguments for research frames and the concepts employed. One important feature of new regional geography has been the fact that, along with its rise, the terminology of regional geography has become broader, although not necessarily less ambivalent, than in the past. It has not merely been the region that has been scrutinized, but also the concepts of place and space. All too little serious research has been done on the possibilities for conceptualizing region and place together to reveal the complex dynamics of social world. Even for many new regional geographers, the meanings of region and place are more or less similar or overlapping.

Human geographers simultaneously use several different concepts of the region, which may be roughly categorized into three categories. The 'prescientific' concept of the region means that researchers do not problematize the meanings of regions. This is typical in applied research which is tailor made, for example, for planning bodies or other regional authorities that simply define in advance the 'regions' (e.g., statistical units) for researchers to inspect. Discipline centered regions are defined by researchers when using, for instance, uniformity of characteristics, homogeneity of content, or functional coherence as their criteria. In critical interpretations, regions are recognized as manifestations of often conflicting, social, political, and cultural practices and discourses. Even if it is still the critical geographer who abstracts and 'creates' the idea of the major dimensions of regions and their formation, these abstractions are explicitly created to reveal the mechanisms, processes, and power relations of the social, political, economic, and cultural worlds.

The discipline centered perspective regards regions as objects (e.g., traditional *Landschaft* geography) or, as results of the research process, often formal classifications of diverging empirical elements. These perspectives are sometimes used to legitimize a specific 'geographical perspective', and hence the theoretical debates on whether regions are 'real' units or imagined, mental categories. These debates are not only a historical curiosity but fitting illustrations of the struggle over legitimate conceptualizations, and the resulting 'regions' are also examples of academic socialization and power/knowledge relations. These regions show the power of geography: once 'invented', they can be powerful in shaping the spatial imaginations and actions of researchers.

Critical approaches to the region emerge from social practice, relations, and discourse. They aim at conceptualizing and scrutinizing current spatialities as part of wider networks of cultural, political, and economic processes, and also social divisions of labor. Most studies labeled as 'new regional geography' belong to this category: for example, those mapping individual and social identities and those regarding regions as manifestations of capital accumulation or settings for interaction. While these approaches have normally been pursued separately, a critical regional geography should ideally combine the political economic focus with questions of subjectification and identity formation. Furthermore, the view on regions as processes stresses both the importance of an historical perspective in understanding them as part of broader regional transformations, demanding the conceptualization of the relevant time scales of history in each case. The way that we understand and conceptualize the region may hence have a profound effect on how we shape and experience regional development, globalization, and the world.

Challenging Bounded Regions

A stable and somehow fixed image of traditional regional geography seems to have become a caricature that is overtaken by the representatives of newer approaches. Thus, new approaches often accentuate the missing historical reflexivity in their earlier counterparts, complete with the positing of the bounded character of regions and the fixity and inwards looking orientations in the traditional perspective. It is nevertheless important to recognize that a relational emphasis can be traced earlier in the history of regional geographic thought, even if the strive to develop regional geography often manifested itself in the search for regional unity and coherence, duly 'arresting' the spatial complexity into a specific form. Kimble, for instance, noted after World War II that regional geographers may perhaps be trying to put boundaries that do not exist around areas that do not matter! But think also of the following citations, taken from Richard Hartshorne in 1939 and Jan Broek in 1965, which display that these thinkers were not merely looking at hermetically sealed, bounded spaces. For

Hartshorne, this was a logical result of his 'Kantian' understanding of the region as a mental category rather than a 'really existing' unit.

> The problem of establishing the boundaries of a geo graphic region … presents a problem for which we have no reason to even hope for an objective solution… The most that we can say is that any particular unit of land has significant relations with all the neighboring units and that in certain respects it may be more closely re lated with a particular group of units than with others, but not necessarily in all respects. (Hartshorne, 1939: 275)

> Whatever the kind of place, we must always keep in mind its position in time as well as its location in space. A place cannot be understood by merely observing the inter action of present day forces. Knowing the legacy of the past and sensing the presence of change are essential qualities of the geographic mind. As to location, no place exists in isolation. We must, therefore, look beyond the internal nature of the are itself (the site) to perceive its external relations with other areas (the situation). (Broek, 1965: 6)

The relatively open character of regions was clear also to the representatives of spatial approaches. In the mid 1960s Peter Haggett, for example, regarded functional, nodal regions as open systems. He suggested that an analysis of the build up of such systems would require at first an analysis of movements (the channels along which movements occur), networks, nodes, and their organiza tion in a hierarchy, acknowledging that this complex system ultimately forms surfaces that we can identify. Currently important social scientific terms like networks and movement/mobility were thus identified as import ant categories in understanding spatial patterns. While scholars like Fred Lukermann recognized in the 1960s that the character of the places of the Earth is always emergent and becoming, the relational complexities of practices and power involved in region building pro cesses were not yet accentuated.

What is then the role of regions and regional geog raphy in a globalizing world characterized by new re gional and scalar divisions of space in economy and culture, processes of regional integration, reorganization, and rescaling of regional governance, regionalization processes, and place marketing and image building? Some scholars have suggested that, if geographers wish to contribute to globalization debates, regional geography has to be revived. As a result, some of them have sought to reshape thinking about the global regional geographies of the world system, some others the regional worlds of economic production, often in the context of city re gions. Still others have taken more political steps, and such representatives of relational thinking as Massey suggest that regions should be seen as nonbounded in the mobile world, and that the opening of borders is a challenge for a progressive (social) science and politics. They offer what is thus as much a political message as a scientific one. It is based on the challenges posed by the fact that – compared with the traditional efforts to create distinctive 'geographical regions' – most (con)temporary regions 'stretch' in space so that their social contents and relations are networked across borders, and appreciates that this networking indeed modifies and reconstitutes regions: regional boundaries and identities do *not* need to be exclusive and permanent. Contrary to the attempt of traditional regional geographers to find a certain unity and coherence for bounded regions, new regional geog raphers want to challenge any taken for granted essence of regions.

This sort of new, more open regional geography is certainly a challenge for both geographic education and research. The problem is that, while in some cases boundaries are quite insignificant, in other cases they are more persistent and cannot be erased very easily. Such cases were likely the prompt for the idea of the boundary that many traditional regional geographers had in mind when they tried to distinguish and to classify 'unique' regional spaces. A certain boundedness is obvious in the fact that many regions are *territories* deployed within processes of governance, and hence are made socially meaningful entities. This means, by definition, that these boundaries have a constitutive role in the management and control of social action. The importance of regions and their boundaries as catalysts for regionalist move ments and for planning strategies is also obvious in some contexts, evident in the case of many ethno territorial groups, but identity narratives produced and reproduced by regional activists, media, and governmental bodies are also typical examples of this politics of distinction. Such social movements hardly regard their own activities as regressive, although others may do so.

All of this is very obvious, but what is needed next is the rethinking of precisely what are 'boundaries'. Many relational thinkers apparently lean on a normal political geographic understanding of boundaries as dividing lines between social entities. It was suggested above that re gions acquire their boundaries in institutionalized prac tices, and that these boundaries are important as social institutions and symbols, not merely as physical lines. This has important implications for both political geo graphic and regional geographic understandings of boundaries and regions, making it easier to understand their persistence. This finding also implies that bound aries do not locate solely in border areas, but may be detected everywhere in a region, in diverging social practices, ideologies, and discourses. This idea becomes obvious in such national and regional examples as 'flag days', regional media, festivals, museums and archives, education, regional novels, electoral districts and voting, or sports events: these are all practices that are constitutive of the region, regional identity building and, indeed, boundaries! Regional boundaries are therefore not bounding the practice in some abstract way, but rather it is the practices that produce and reproduce such boundaries. This simply means that a crucial task for regional geographers is to study in each case how existing ideas of 'difference' are produced and reproduced in material and discursive practices. Positing of difference as itself problematic, and searching for how this is produced and reproduced, may be more fruitful than simply ignoring the role of 'boundedness' in social life.

Conclusion

Regional geography has been a variegated and historically contingent subarea of academic geography, both as a concrete research field and as an object of theorization. It was important in many countries until the 1950s–1960s, and since then has been declining in academic research while maintaining its role in both school geography and higher education. The so called new regional geography that emerged during the 1980s has brought regional geography and the region back on to the agenda in theoretical debates, however, albeit this time in a different way compared with the traditional perspective. New regional geographers have theorized both the ideas of the region and the role of regional knowledge in the frameworks provided by social and cultural theory, taking seriously historically contingent societal conditions, and this new regional geography is therefore *not* merely the collection of 'spatial knowledge' nor an internal methodological problem for the discipline of geography. The representatives of this approach have started to consider, during the last 10–15 years or so, how best to perform a new regional geography. Most scholars suggest that it occurs most effectively when exceeding the traditional boundaries between social sciences and humanities, and when scholars become interested in the power relations associated with region building processes, regional identity narratives, or regional development. Along with the rise of the new regional geography, some other traditional themes have also been developed further, above all the question of scale and the role of historical approaches.

Regions are nowadays regarded by new regional geographers as complicated institutional structures, indeed 'institutional facts', because they are dependent on both human agreement and such social institutions as political organization, governance, economy, media, or education systems. Most of these institutions operate across scales, which challenges traditional ideas about regions as separate, bounded units. This boundedness is not of course an either/or question. In the current mobile world, regional boundaries may be open in some social practices, in some others relatively closed. Actors, institutions, and organizations involved in the territorialization and meaning making of space may act both inside and outside of regions. Making regions always includes normative components because institutional structures are sedimentations of rules, power, and trust, in which boundaries, symbols, and institutions merge through material practice. Once created, they are also social facts in the sense that they can generate and are generated by action as long as people believe in their existence and they have a role in spaces of 'publicity' (e.g., in the media) or in processes of governance. This action, for its part, may be reproductive, resistant, or transformative at the same time.

Though regional geography lost its academic status during the 1950s–1960s, and even if new regional geography represents a new theoretically informed perspective on regional geography, the traditional perspective still often manifests itself in textbooks on regional geography where regions, created and classified by the researchers, are often described without contextual interpretation. One perpetual challenge for regional geographers is therefore to deconstruct the 'geographical' assumptions and inclusions/exclusions that such regions imply at various spatial scales. Textbooks often reproduce the historical meaning of geography: *ge* (Earth) and *grapho* (to write or describe). Of course we should not deny the cultural role of this information: whatever the critique says, regional geography is also, in this traditional, rather technical sense, important for our world views. But a critical understanding of the problems of this approach is important, and the key is to recognize the versatile and complex chiasm of power and regional geography. This simply means that a critical regional geography is also political geography.

See also: Hartshorne, R.; Hettner, A.; Place; Region; Regionalism; Regionalization/Zoning Systems; Space II; Structuration Theory; Structurationist Geography.

Further Reading

Allen, J., Cochrane, A. and Massey, D. (1998). *Rethinking the Region*. London: Routledge.

Bourdieu, P. (2004). *Science of Science and Reflexivity*. Cambridge: Polity.

Broek, J. O. M. (1965). *Geography: Its Scope and Spirit*. Columbus, OH: Charles E. Merrill.

Claval, P. (1998). *An Introduction to Regional Geography*. Oxford: Blackwell.

Clout, H. (1989). Regional geography in the United Kingdom: A trend report. In Paul, L. J. (ed.) *Post War Development of Regional Geography. Nederlandsche Geografische Studies* 86: pp 25 41. Utrecht.

Entrikin, J. N. (1981). Philosophical issues in the scientific study of regions: In Herbert, D. T. & Johnston, R. J. (eds.) *Geography and the*

Urban Environment. Progress in Research and Applications, vol. VI, pp 1 27. Chichester: John Wiley & Sons.
Gilbert, A. (1988). The new regional geography in the English and French speaking countries. *Progress in Human Geography* 12, 208 228.
Gould, P. (1979). Geography 1957 77: The Augean period. *Annals of the Association of American Geographers* 69, 139 151.
Gregory, D. (1978). *Ideology, Science and Human Geography*. London: Hutchinson.
Grigg, D. (1967). Regions, models and classes. In Chorley, R. & Haggett, P. (eds.) *Models in Geography*, pp 461 509. London: Methuen.
Hart, J. F. (1982). The highest form of geographer's art. *Annals of the Association of American Geographers* 72, 1 29.
Hartshorne, R. (1939). *The Nature of Geography*. Lancaster, PE: Association of American Geographers.
Hartshorne, R. (1959). *Perspective on the Nature of Geography*. Chicago, IL: RandMcNally.
Johnston, R. J. (eds.) *Geography and the Urban Environment. Progress in Research and Applications*, vol. VI. Chichester: John Wiley.
Johnston, R. J. (1991). *A Question of Place*. Oxford: Blackwell.
Johnston, R. J., Hauer, J. and Hoekveld, G. A. (eds.) (1990). *Regional Geography. Current Development and Future Prospects*. London: Routledge.
Jones, M. and MacLeod, G. (2004). Regional spaces, spaces of regionalism: Territory, insurgent politics and the English question. *Transactions of the Institute of British Geographers* 29, 433 452.
Kimble, G. H. T. (1951). The inadequacy of the regional concept. In Stamp, D. L. & Woolridge, S. W. (eds.) *London Essays in Geography*, pp 151 174. London: Longman.
Massey, D. (1978). Regionalism: Some current issues. *Capital and Class* 6, 106 126.
Minshull, R. (1967). *Regional Geography: Theory and Practice*. London: Hutchinson.
Paasi, A. (1986). The institutionalization of regions: A theoretical framework for understanding the emergence of regions and the constitution of regional identity. *Fennia* 164, 105 146.
Paasi, A. (2002). Region and place: Regional worlds and words. *Progress in Human Geography* 26, 802 811.
Pred, A. (1984). Place as historically contingent process: Structuration and the time geography of becoming places. *Annals of the Association of American Geographers* 74, 279 297.
Schaefer, F. K. (1953). Exceptionalism in geography: A methodological examination. *Annals of the Association of American Geographers* 43, 226 249.
Thrift, N. (1994). Taking aim at the heart of the region. In Gregory, D., Martin, R. & Smith, G. (eds.) *Human Geography*, pp 200 231. London: Macmillan.
Woolridge, S. W. (1965). *The Geographer as Scientist*. London: Thomas Nelson.

Relevant Websites

http://www.uwsp.edu
Internet resources for World Regional Geography, University of Wisconsin Stevens Point.

http://www.ifl leipzig.com
Leibniz Institute for Regional Geography (Leibniz Institut fur Landerkunde).

http://members.aol.com
World Regional Geography Web Exploration, AOL Hometown.

Regional Geography II

D. MacKinnon, University of Aberdeen, Aberdeen, UK

Glossary

Localities Research A major research initiative conducted by British geographers in the late 1980s and early 1990s. It involved a number of studies which examined the impact of global economic restructuring on different local areas. Locality studies also became a touchstone for methodological and conceptual debate at this time.

New Regional Geography This was a product of 1980s debates in human geography, attracting considerable interest. In contrast to its predecessor, the 'new regional geography' would adopt a more theoretical approach and focus on regions in the context of wider processes of change.

New Regionalism A convenient umbrella term for a number of related strands of research on regional economic development in the 1990s. These share an underlying emphasis on the increased prominence of 'the region' as a unit of economic organization and political action under late capitalism.

Old Regional Geography Dominant within geography from the 1920s to the 1950s, this was based on 'areal differentiation', requiring the description and interpretation of the variable character of the Earth's surface. In practice, this involved regional classification and description.

Post-Structuralism A set of theories derived from philosophy and literary and cultural studies which stress the fractured identities of individual subjects, seeing these as products of broader social categories and discourses.

Regulation Approach Originally developed by French economists in the 1970s, this approach became influential in human geography in the late 1980s and early 1990s. It highlights the role of particular institutional supports in stabilizing capitalism, allowing periods of sustained growth to occur.

Relational Approaches to Space Informed by post-structuralism and actor-network theory, these seek to overturn conventional conceptions of space as closed, static, and bounded; emphasizing instead, openness, change, and connectivity. From this perspective, places and regions have no pre-given essence or identity, but are constituted out of wider relations and networks.

Introduction

Regions have long been a key concept in human geography, becoming almost synonymous, at certain times, with the discipline itself. The term region has generally been taken to refer to a bounded portion of the Earth's surface which is held to be distinguishable from other such areas on the basis of its particular characteristics or features. The issue of scale in particular has complicated the study of regions which have been identified at a range of scales, from large 'world regions' spanning several countries and continents to small local areas or neighborhoods within cities. Regions have also been defined by a variety of criteria, ranging from features of the physical environment to language and culture. As this indicates, the concept carries a certain elasticity, allowing geographers and others to define regions in line with particular criteria and agendas. The regionalist approach is associated with an emphasis on the unique features of place, as opposed to the general principles underpinning the organization of space. It has closely informed the public image of geography which typically associates the discipline with the compilation of detailed factual information about particular regions and countries.

At the same time, geographers have periodically rebelled against their regionalist heritage, seeking to abandon or substantially revise the concept in the light of changing intellectual fashions and real world concerns. The focus on identifying and documenting distinctive regions has, from time to time, appeared somewhat backward and anachronistic in the face of modern communications and technology. The regionalist tradition has, moreover, threatened to consolidate geography's rather subordinate position in the academic division of labor as a descriptive, atheoretical backwater, eclipsed by the prestige of natural sciences like physics and chemistry and, to lesser extent, social sciences such as economics and sociology. One of the most fundamental underlying issues is the concern that human geography's regionalist tradition is at odds with the increasingly integrated nature of modern society with the latest transport and communication technologies enabling people to communicate and interact over distance, resulting in an effective 'shrinkage of space'. In a world in which satellite media broadcasts, e mail, and the Internet facilitate the instantaneous transmission of distant events while cheap air transport supports a growing level of international travel, the identification of unique and bounded local areas can appear futile and misguided. While globalization has brought this issue to the fore, it is far from new.

Despite such concerns, the concept of the region has proved surprisingly resilient. No sooner have geographers declared the regions to be dead, than they have been reborn in a new guise. The region, it seems, is an idea that just won't go away. Thus, while regionalist approaches were rejected by the cutting edge of the new scientific geography in the late 1950s and 1960s, the late 1970s and 1980s witnessed calls for the development of a 'new regional geography' which could bring together the study of general processes and particular social settings. Moreover, the so called 'new regionalism' of the 1990s conveys a sense in which regions, somewhat paradoxically, have become more prominent under globalization. This 'new regionalism' has, however, attracted criticism from advocates of relational thinking for apparently remaining underpinned by traditional assumptions of territorial fixity and boundedness. As such, the region can be viewed as something of a magnetic concept in human geography, serving to periodically attract and repel attention.

The 'Old Regional Geography'

The region became the organizing focus for academic geography in Western Europe and North America between the 1920s and the late 1950s. This drew on varied traditions of scholarship in different European countries, with the work of Vidal de la Blache in France and Hettner in Germany proving particularly influential in establishing the region as the key material expression of human–environmental relations. The American geographer, Richard Hartshorne, built on these foundations through a massive work of systemization and prescription to define geography as the science of 'areal differentiation'. This meant that its main purpose was to describe and interpret the variable character of the Earth's surface, accounting for those features that made one place distinct from others. This project usually began with aspects of the physical environment before proceeding to cover minerals, transport, settlement, and culture. Regions were defined at three main scales: world regions, defined either in terms of natural regions or culture regions such as Southeast Asia; subnational regions like North East England or California; and more localized areas (e.g., Upper Deeside or Southern Kincardineshire in North East Scotland).

The 'old regional geography' viewed the economy of a region as closely related to its natural resource base, reflecting the naturalistic tenor of this approach, defining the physical environment as a platform which facilitated human activity (earlier approaches had viewed this relationship more deterministically). In particular, minerals, soils, and climates were seen as providing the basis of regional economies as reflected in the profusion of industrial and agricultural regions in Western Europe and North America in the 1920s, 1930s, and 1940s. Such 'functional regions' can be distinguished from regions defined on the basis of their unique characteristics. While the former involved the development of regional typologies, the latter emphasized regional 'personalities' based largely on cultural criteria. As the economic depression of the 1930s took hold, regional geography became linked to the identification of problem regions by government agencies as the economies of specialized manufacturing regions collapsed in the face of reduced international demand and trade. In the early 1930s, for instance, the British government designated a number of 'special areas' (North East England, West Cumberland, South Wales, and Clydeside North Lanarkshire in Scotland) affected by high unemployment as eligible for special assistance.

The old regional geography, however, appeared increasingly anachronistic and backward in the post war era. Its descriptive approach seemed inadequate in the face of enthusiasm for science and the need for explanation and prediction of social and economic trends. At the same time, the concern with identifying unique regions was being undermined by increasing integration between regions, not least through the growth of modern transport and communication networks. This was memorably expressed by the American geographer, George Kimble, who observed, following an air trip over the Midwest, that it was the links in the landscape that stood out rather than the breaks, expressing his concern that "regional geographers may be trying to put boundaries that do not exist around areas that do not matter."

In response, a new scientific geography was advanced in the 1950s, utilizing quantitative methods derived from statistics and mathematics in order to develop models and theories of spatial organization. This new approach spread from its initial epicenters in the United States and United Kingdom to become widely adopted by the mid 1960s. As a result, the old regionalist approach was rapidly eclipsed, particularly in the 'cutting edge' fields of economic, social, and urban geography, although regions retained some residual presence within the new scientific geography through the development of quantitatively based typologies and classification schemes (functional regions).

The 'New Regional Geography', Localities and the Regulation Approach

By the late 1970s and 1980s, however, prominent human geographers such as Derek Gregory and Nigel Thrift were calling for the establishment of a 'new regional geography'. This was to be a rather different enterprise from its predecessor, adopting a more theoretical approach and focusing on regions in the context of wider processes of change. This reflected growing dissatisfaction, in certain quarters, with the spatial science of the 1960s and 1970s and the more structuralist versions of Marxism that had been embraced by some prominent researchers in the 1970s. Both these approaches were abstract in nature, emphasizing the general features of space and society over the particularities of place and the practices and beliefs of individuals. As such, the notion of a 'new' regional geography can be seen as part of an effort to create a more genuinely human geography in the 1980s which sought to examine the lives of individuals and groups based in particular regions, but in the context of the broader forces shaping society and armed with a sophisticated theoretical understanding. This often involved challenging what the political geographer, John Agnew, termed the 'territorial trap' of taken for granted political structures, particularly nation states and also substate jurisdictions such as local government areas, in order to define regions according to the particular process being studied.

A key theoretical inspiration was structuration theory, developed by the sociologist, Anthony Giddens. This emphasized the role of regions or 'locales' as settings for social interaction where people are socialized into society through various institutional processes, particularly education. Giddens's concept of locales remained highly elastic as these were envisaged as existing at a range of scales in relation to specific purposes or activities. For example, a classroom could be regarded as a locale in an educational context whereas a nation might be viewed as the crucial unit in the context of a telecommunications system. In addition, the rediscovery of regions in the 1980s was also the product of a Marxist informed understanding of regions as local responses to wider capitalist processes and a more humanist view of regions or places as the focus for cultural identification, blending into the 'new' cultural geography of the 1990s. Such rediscovery, however, remained largely theoretical in the sense that elaborate outlines of what the new regional geography would look like were not really matched by detailed empirical accounts of the construction and transformation of particular regions. Perhaps the major exception to this is the work of the Finnish geographer, Anssi Paasi, which developed a general theory of regionalization based on the identification of four distinct stages, exemplified by a thorough study of the construction of the Finnish–Russian border.

The Marxist rediscovery of regions as local responses to broader capitalist processes in the 1980s is best represented by the work of the British geographer, Doreen Massey. Her theory of spatial divisions of labor demonstrated how production was becoming 'stretched' across different regions as large corporations sought to capitalize on preexisting regional differences, particularly those related to the availability of labor. Increasingly, higher order functions such as research and development were being located in metropolitan regions like the South East of England where large reserves of qualified, professional labor could be found while skilled manufacturing was being concentrated in industrial regions such as the Midlands of England which contained skilled engineering labor (**Table 1**). At the same time, less skilled work related to the assembly of components was being relocated in peripheral regions of the UK and less developed countries where labor costs were much lower. As such, different types of regions were specializing in different parts of the production process. This approach successfully brought together a concern with general processes of capitalist development, relating the locational decisions of corporations to the dynamics of accumulation or investment, and an emphasis on local uniqueness in the form of individual regions.

Massey's work fed into the localities research program of the late 1980s in the UK which aimed to examine the impact of global economic restructuring on different local areas. This research initiative became a touchstone for methodological and conceptual debate among critical human geographers in the late 1980s and early 1990s. Three main criticisms were articulated. First, critics such as the Marxist geographer, Neil Smith, argued that the initiative risked licensing a new empiricism in human geography, favoring the detailed documentation of local outcomes over more theoretical issues related to the restructuring of the capitalist space economy. Second, the research was criticized for embracing a kind of economic reductionism which failed to really incorporate the role of culture and politics in shaping regional responses to broader processes of change. Third, there was a tendency for localities researchers to link the study of particular facets of reality to specific geographical scales such that 'the local' was defined in terms of responses and

Table 1 The spatial division of labor in manufacturing

Functions	*Characteristics of the division of labor*	*Location*
Research and development	Conceptualization/ mental labor; high level of job control	South East England
Complex manufacturing and engineering	Mixed; some control over own labor process	Established manufacturing regions like West Midlands
Assembly	Execution, repetition, manual labor no job control	Peripheral regions such as Cornwall or North East England

Adapted from Massey, D. (1995). *Spatial Divisions of Labour* (2nd edn.), p 76. London: MacMillan.

outcomes with 'the global' characterized as the scale at which causal processes and mechanisms operated.

In the late 1980s and early 1990s, the regulation approach became highly influential in human geography, particularly economic geography. Originally developed by French economists in the 1970s, the regulation approach focuses attention on the role of particular institutional 'supports' in stabilizing capitalism, allowing periods of sustained growth to occur. Regulationists conceive of these institutional 'supports' in terms of the mode of regulation, referring to the social norms and conventions which shape the process of capitalist development. Key aspects of capitalism upon which regulation is focused include labor and the wage relation, forms of competition and business organization, the monetary system, the state, and the international regime. When these different objects of regulation are linked together in a coherent fashion – achieving a balanced relationship between production and consumption – a period of stable growth, known as a regime of accumulation, ensues. In the 1980s, the regulation approach provided theoretical support for the idea that post war Fordism – based on mass production and mass consumption, linked through rising wages for workers and increased productivity in the workplace – had given way to a new regime of flexible accumulation, defined by the growth of more individualized consumption patterns and the deployment of new information and communication technologies in production.

In addition to providing a framework for understanding historic shifts in the organization of capitalism, the regulation approach was also deployed at the regional scale by economic geographers and others. For instance, Alain Lipietz, one of the original French regulationists, developed the concept of regional armatures to highlight the role of regional elites in regulating conflictual social relations through the construction of a distinctive regional interest and identity. The activities of this group are instrumental, Lipietz argues, in the conversion of a 'space in itself', defined by the prevailing social relations of production in a region, to a 'space for itself' endowed with the institutional capacity to intervene in processes of economic and social change. At the same time, Jamie Peck and Adam Tickell drew upon earlier research on industrial restructuring and uneven development to develop the concept of local modes of social regulation. According to them, these are distinguished by their "unique position within wider (national and international) structures of accumulation and regulation." They argued that the 'regulatory deficit' of South East England, reflecting the contradictions of uneven development under Thatcherism, meant that it was unable to reproduce the conditions for sustained growth into the 1990s. By contrast, the political geographers, Joe Painter and Mark Goodwin, stressed the uneven flow of regulatory processes over time and space, criticizing the over reliance on the concept of mode of regulation. From this perspective, regulation comprises a range of diverse practices grounded in particular local and regional contexts, with local state agencies representing important institutional channels through which national regulatory priorities are "interpreted and ultimately delivered," according to Goodwin and Painter.

The 'New Regionalism'

Discussions of the 'new regional geography' were paralleled and ultimately superseded by the growth of a 'new regionalism' in economic geography and regional development studies in the 1990s. The term 'new regionalism' is derived from the argument that regions have become more prominent as units of economic organization and political action under late capitalism. While, as shown below, the label covers a number of distinct strands of research, they are all underpinned by this common claim. As the global economy has became increasingly integrated in the 1970s and 1980s, post war structures of national economic regulation were undermined, resulting in increasingly direct competition between regions for inward investment and global markets. At the same time, the 'hollowing out' of national states in the face of globalization has involved a certain 'denationalization' of power as the subnational and supranational scales of government have become more prominent, evident in the increased role of regional authorities and the growth of institutions such as the European Union (EU). In general, the 'new regionalism' is characterized by an emphasis on the role of social and institutional factors within regions in shaping economic development. This is linked to the abandonment of the traditional regional policy of the post war period – whereby central government provided financial incentives for companies to locate plants and factories in 'special areas' characterized by high unemployment (**Figure 1**) – in favor of a new regional policy which strives to promote growth and innovation in all regions.

One strand of research in the late 1980s and early 1990s focused on the geography of flexible accumulation, highlighting the spatial agglomeration or concentration of certain forms of production in what were termed 'new industrial spaces'. This is sometimes termed the 'Californian school' since several of its leading protagonists, notably Allen Scott and Michael Storper, are based there, along with several of the agglomerations they studied, notably Silicon Valley and the Hollywood film production complex (**Figure 2**). According to their analyses, the logic of flexible accumulation was leading firms to subcontract certain functions to other firms in the face of shifting patterns of market demand. The

Figure 1 Areas eligible for regional policy assistance, 1945. Source: Mohan, (1999). *A United Kingdom: Economic, Social and Political Geographies,* p 182. London: Arnold.

Figure 2 The location of Silicon Valley. Source: Castells, M. and Hall, P. (1994). *Technopoles of the World*, p 13. London: Routledge.

associated pressure to minimize transaction costs – these are related to the problems inherent in coordinating activities between different firms such as manufacturers and their suppliers – encouraged spatial proximity between firms, enabling them to communicate and interact more effectively. As such, the 'Californian school' explained the rise of 'new industrial spaces' according to conventional economic factors such as changing organizational structures and interfirm relations (traded interdependencies). A later contribution by Storper, however, emphasized the role of what he called 'untraded interdependencies'. These refer to the informal linkages and relationships that tie firms together in particular locations, comprising an intangible set of skills, attitudes, habits, and understandings. In contrast to material linkages based on the exchange of specific goods and services, such untraded linkages are built and sustained through routine forms of communication and interaction.

A second strand of 'new regionalist' research adopted a more sociological approach to the understanding of industrial districts, relating this to the overarching shift in the organization of capitalism from mass production to flexible specialization. Derived from the writings of the English economist, Alfred Marshall, in the early twentieth century, the term industrial district refers to highly specialized local areas based on networks of small firms. While such districts had been eclipsed by the growth of mass production and large, integrated corporations over the course of the twentieth century, a number of sociologists and geographers were pointing to their renaissance in the 1980s in advanced economies such as Italy, France, Japan, Denmark, and Spain. The districts of Central and Northeastern Italy attracted particular attention, becoming known as the 'Third Italy' (in contrast to the rich, industrial North and the poor, agrarian South). A key theme concerned the role of wider institutional and sociocultural factors in supporting networks of small firms and entrepreneurs. According to the flexible specialization thesis, a dense network of local authorities, labor unions, industry associations, and chambers of commerce developed a sophisticated reservoir of knowledge, skills, and resources and provided business premises and a range of services to small firms.

A third strand of the 'new regionalism' has focused on the restructuring of states, emphasizing the 'downward' transfer of power from national states to regional agencies and an 'upward' shift to supranational bodies like the EU. Devolution is an important process here, involving a transfer of certain functions and responsibilities from central government to subordinate regional governments. It has been introduced in many different countries in response to the pressures on states from 'below' in the form of demands for greater regional autonomy and from 'above' through globalization and supranational integration. In more concrete terms, this research has examined the role of regional institutions in promoting economic development. Such institutions operate as part of a broader system of regional governance which incorporates private sector interests, specialist public agencies, and representatives of the voluntary sector alongside traditional local government, highlighting the importance of networks and partnerships in coordinating the activities of different organizations. In practice, however, this research has sometimes lost sight of relationship with the reorganization of the state at the national and supranational scales. As a result, the 'new regionalism' has tended to misinterpret the increased role of regional institutions as a product of regional level factors rather than as reflecting broader processes of state restructuring.

From the mid 1990s, earlier work on new industrial spaces and industrial districts gave way to a growing focus on regional learning and innovation processes. This reflects the significance of the meta or master narrative of competitiveness and the knowledge economy in structuring economic policy in recent years, reinforcing the shift away from traditional regional policy to a new approach based on growth and innovation. The latter adopts a more 'bottom up' outlook which focuses on the need to develop skills and to promote enterprise and learning within regions. As three leading British geographers, Ash Amin, Doreen Massey, and Nigel Thrift put it, "locally orchestrated regional development has replaced nationally orchestrated regional policy." Perhaps the key concept developed by economic geographers and others in this context is the learning region. In classic 'new regionalist' fashion, this emphasizes the renewed role of the region as a nexus of learning and innovation processes within global capitalism. The key idea is that the transmission of specialized, tacit knowledge, in which regional competitive advantage is held to be ultimately rooted, depends upon spatial proximity between firms, requiring regular face to face communication, in contrast to codified knowledge which can be exchanged globally through advanced communications networks. This is often linked to Storper's concept of untraded interdependencies, highlighting the informal conventions and routines that tie firms together in particular locations.

The Relational Region

In recent years, however, the 'new regionalism' has been subject to criticism from advocates of a relational approach to the understanding of space for assuming that regions are both internally coherent and externally bounded. In this sense, it would seem, the specter of geography's past maintains a lingering presence, with elements of traditional regional geography continuing to implicitly inform the 'new regionalism'. By contrast, informed by post structuralism and actor network theory, relational geography seeks to overturn conventional conceptions of space as closed, static, and bounded; emphasizing instead, openness, change, and connectivity. One important source of inspiration has been Doreen Massey's sustained efforts to rethink place in terms of the wider social relations that constitute it. Rather than representing some pre given identity, she argues that places are actually constituted out of wider relations and networks. Accordingly, places have no singular essence or identity, but are meeting points where different social groups and relations come together, emphasizing diversity and heterogeneity. This generates a distinctive 'politics of place' as different social groups and actors assert their claims to place, striving to shape its meaning and character according to their interests and orientations.

This conception of space has fostered a new concern with 'the relational region' which reimagines regions as open and discontinuous spaces, defined by the wider social relations in which they are situated. They are created for particular purposes by, for instance, policy makers, social movements, or academic analysts, meaning that they have no essential character or identity outside of these acts of creation and becoming. Instead of being internally coherent and unified, regions are discontinuous and divided. John Allen, Massey, and Allan Cochrane, based at the Open University, illustrate these arguments through an analysis of South East England, the emblematic growth region of neoliberal Britain in the 1980s and 1990s, defined by its close connections to the global economy through the City of London in particular. Within the South East, areas of economic decline and deprivation exist, complicating and confounding the overarching image of growth and prosperity, while the boundaries of the region can be seen as open and porous (**Figure 3**). Another strand of work has examined the development of cross border and transnational regions in the context of globalization and supranational integration. Such regions can be primarily economic, such as the Indonesia–Malaysia–Singapore growth triangle in Southeast Asia, or political in nature, with many of the cross border regions within the EU (Euroregions) falling into this latter category.

The advancement of this relational theory of the region has been accompanied by a critique of regionalist

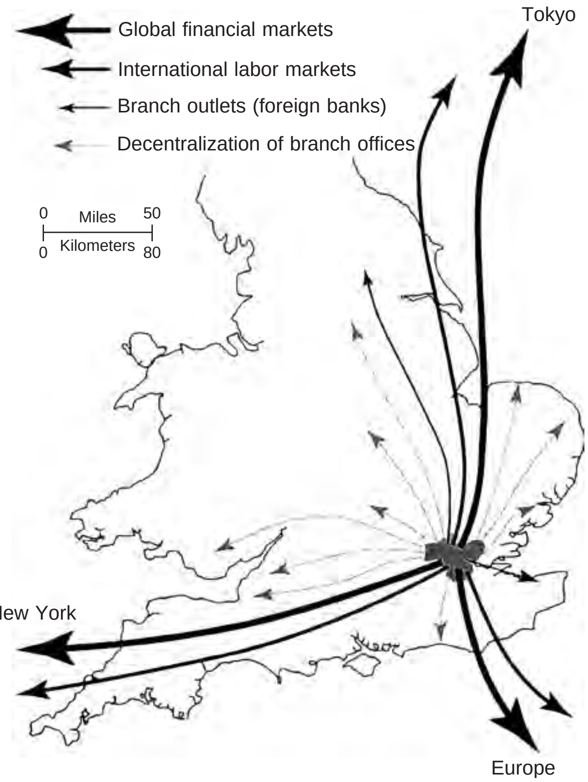

Figure 3 International financial links of South East England. Source: Allen, J., Massey, D. and Cochrane, A. (1998). *Rethinking the Region*, p 49. London: Routledge.

thinking. While the 'new regionalism' focuses on internal linkages between firms and institutions, theoretical assertion has often run ahead of empirical investigation, meaning that there is only limited evidence of the existence of industrial districts and learning regions outside of the classic 'hot spots' such as Silicon Valley, the City of London, and Baden Württemberg. This leads Ash Amin to dismiss economic localization as a myth. By contrast, relational geographers argue that globalization means that the connections between regions and the wider world have multiplied and proliferated. As such, regions can no longer be viewed as bounded and coherent entities, existing only as sites within wider corporate, governmental, or social networks. There is no reason to assume that the spaces of economic regionalism identified by the likes of the Californian school will coincide with the political regions created by devolution, still less with spaces of local identification and belonging. Regions, then, are internally heterogeneous and fractured, having been created and defined by different groups for particular purposes.

In particular, the central 'new regionalist' claim that geographical proximity facilitates knowledge exchange and learning through face to face communication and collaboration has attracted criticism. As a number of authors have argued, other forms of proximity can also be identified – namely cognitive, social, and institutional proximity – where closeness is defined in terms of mental frameworks, interpersonal relations, and shared habits and routines rather than geographical space *per se*. This theoretical observation has been reinforced by the empirical weaknesses of the regionalist thesis: despite a number of studies of dynamic growth regions, research has failed to really specify how learning processes operate geographically. The notion that localized relations between firms support close collaboration finds only limited support in the literature with several studies pointing to a lack of direct interaction, particularly in terms of 'traded interdependencies' which can be measured more easily.

By contrast, recent contributions have argued that learning can also occur within relational and organizational networks, based on the establishment of a shared cognitive frame of reference between actors that are not geographically co located. For instance, John Allen suggests that tacit knowledge can be transmitted over long distances, indicating that 'thick relationships' and 'shared understandings' depend on the quality of the interpersonal and organizational relations involved, not their geography. This argument relies on the notion of 'communities of practice' – derived from organizational studies and often opposed to geographical communities – which refers to the establishment of close relations between people working together within the same organizational or professional context.

These relationalist arguments provide a welcome and perhaps necessary critique of the 'new regionalist' literature. In many respects, however, they offer merely a mirror image or flip side, privileging relational over territorial understandings rather than vice versa. The suggestion that the local does not matter in the face of organizational and professional networks that operate over long distances hardly seems to represent progress over the opposing claim. By setting them up as opposites, it also results in a neglect of the potential overlap between the role of regions and networks in structuring learning and innovation. Furthermore, the relationalist argument that networks are more important than regions in supporting knowledge generation and learning under globalization, rests on equally weak, indeed perhaps even weaker, foundations than the 'new regionalist' claim about the geographically 'sticky' nature of tacit knowledge, lacking much detailed research to support it. Accordingly, neither approach has provided a clear understanding of how learning and innovation processes actually operate across space. Rather than viewing regions and networks as separate entities, research needs to examine the relationships between them, as indicated by a number of empirical studies which indicate that firms engage in both local and extralocal relationships to support innovation.

Summary

As this article has shown, regions have been a central focus of interest in academic geography since the 1920s. Indeed, they represented the central organizing focus of the dis cipline until the 1950s as the 'old regional geography' held sway, based on the identification and characterization of unique regions at a variety of scales. In the wake of geo graphy's Quantitative Revolution in the late 1950s and 1960s, however, regionalism was largely abandoned as geographers turned to the construction of abstract models and theories of spatial organization. Yet, by the 1980s, the wheel had turned full circle and regions were rediscovered by a new generation of researchers searching for a middle ground between abstract theory and local detail. This fostered talk of a 'new regional geography' which variously viewed regions as settings for social interaction, responses to capitalist restructuring processes, or foci of cultural identification. While the latter was largely subsumed into a revived cultural geography, the other strands were super seded by the 'new regionalism' of the 1990s in economic geography and regional development studies, based upon the underlying claim that regions were attaining a new prominence under late capitalism.

Since around 2000, however, this 'new regionalism' has itself been the subject of criticism from advocates of a relational approach to space for its adherence to tradi tional assumptions of internal coherence and external boundedness. These have become increasingly untenable due to the profusion of interregional networks and flows under globalization. While such criticisms are generally persuasive, it is important to avoid the wholesale aban donment of the regional concept. To echo Ron Johnston's distinction from the early 1990s, although efforts to construct academic edifices such as the 'new regional geography' or the 'new regionalism' can be viewed as misguided in retrospect, the study of regions in geog raphy remains an important task. In particular, research needs to explore the interfaces between regions and networks, reimagining and reinventing the region in an increasingly globalized world rather than rejecting it as anachronistic and regressive. Indeed, because of its geo graphically selective and uneven nature, globalization surely only heightens the need to understand processes of regional differentiation and transformation.

See also: Europe of Regions; New Regionalism; Regional Integration; Regional Planning and Development Theories; Regional Production Networks; Relational Economic Geography.

Further Reading

Allen, J. (2000). Power/economic knowledge: Symbolic and spatial formations. In Bryson, J., Daniels, P., Henry, N. & Pollard, J. (eds.) *Knowledge, Space, Economy*, pp 15 33. London: Routledge.

Allen, J., Massey, D. and Cochrane, A. (1998). *Rethinking the Region*. London: Routledge.

Amin, A. (2004). Regions unbound: Towards a new politics of place. *Geografiska Annaler* 86B, 33 44.

Amin, A., Massey, D. and Thrift, N. (2003). *Decentring the Nation. A Radical Approach to Regional Inequality*. London: Catalyst.

Gilbert, A. (1988). The new regional geography in English and French speaking countries. *Progress in Human Geography* 12, 208 228.

Goodwin, M. and Painter, J. (1996). Local governance, the crises of Fordism and the changing geographies of regulation. *Transactions of the Institute of British Geographers* 21, 635 648.

Johnston, R. J. (1990). The challenge for regional geography: Some proposals for research frontiers. In Johnston, R. J., Hauer, J. & Hoekveld, G. A. (eds.) *Regional Geography: Current Developments and Future Prospects*, pp 22 39. London: Routledge.

Jonas, A. (1988). A new regional geography of localities. *Area* 10, 101 110.

Kimble, G. T. H. (1951). The inadequacy of the regional concept. In Stamp, D. L. & Wooldridge, S. W. (eds.) *London Essay on Geography: Rowell Jones Memorial Volume*, pp 151 174. London: Longmans; reproduced in Agnew, J., Livingstone, D. N. and Rogers, A. (eds.) (1996). *Human Geography: An Essential Anthology*, pp 192 512. Oxford: Blackwell.

Lipietz, A. (1994). The national and the regional: Their autonomy vis à vis the capitalist world crisis. In Plan, R. & Gills, B. (eds.) *Transcending the State Global Divide: A Neo Structuralist Agenda in International Relations*, pp 23 43. London: Lynne Reimer.

Massey, D. (1994). A global sense of place in Massey, D. In *Place, Space and Gender*, pp 146 156. Cambridge: Polity.

Massey, D. (1995). *Spatial Divisions of Labour* (2nd edn.), p 76. London: MacMillan.

Morgan, K. (1997). The learning region: Institutions, innovation and regional renewal. *Regional Studies* 31, 491 504.

Painter, J. (2008). Cartographic anxiety and the search for regionality. *Environment and Planning A* 40, 342 361.

Peck, J. and Tickell, A. (1995). The social regulation of uneven development: 'Regulatory deficit', England's South east and the collapse of Thatcherism. *Environment and Planning A* 27, 15 40.

Sparke, M., Sidaway, J. D., Bunnell, T. and Grunday Warr, C. W. (2004). Triangulating the borderless world: Geographies of power in the Indonesia Malaysia Singapore growth triangle. *Transactions of the Institute of British Geographers NS* 29, 485 498.

Storper, M. (1995). The resurgence of regional economies, ten years later: the region as a nexus of untraded interdependencies. *European Urban and Regional Studies* 2, 191 222.

Relevant Websites

http://europa.eu/pol/reg/index en.htm
European Union Regional Policy.

http://www.michiganadvantage.org/
Michigan Economic Development Corporation.

http://www.englandsrdas.com/
Regional Development Agencies.

Regional Inequalities

M. Dunford, University of Sussex, Brighton, UK

Glossary

Convergence Hypothesis The convergence hypothesis is the idea that economies characterized by similar conditions would converge on a similar steady-state rate of growths and similar *per capita* incomes, with economies that are less developed growing faster than economies that are more developed. Empirically this claim is associated with the β and the θ convergence hypotheses. According to the first hypothesis the slope parameter, $(1 \quad e^{\beta t})/t$, of regressions of per capita GDP growth in t years, $\hat{y}_t$, on the logarithm of initial gross domestic product (GDP) per head, y_0, will be negative. According to the latter, the dispersion of regional incomes will diminish.

Gross Domestic Product (GDP) GDP is a measure of aggregate value added. As new wealth created in production is distributed to those who participate in economic life, GDP is also a measure of primary incomes (profits, interest, rents, and wages) accruing to those who contribute to a region's economic activities. The primary distribution of income is subsequently modified by state-administered redistribution.

The result is the secondary distribution of income. These secondary incomes are saved and/or spent, and thereby permit the establishment of claims over the goods and services produced. Measurement of these expenditures (conventionally divided into consumer and government expenditure, investment/savings, and net exports) offers a third way of measuring GDP.

Purchasing Power Standards (PPS) PPS are quantitative indicators of the allowance that must be made for differences in the prices of goods and services in different areas. The quantity of goods and services that a given amount of an international currency will exchange for is greater in a low-cost than in a high-cost region. The PPS measure allows for these differences in purchasing power and is, therefore, an index of the volume of goods and services that different local economies produce.

Region Regions are defined either as homogeneous regions that share certain characteristics, or as analytical or functional regions that are made up of places that are interdependent, or as administrative regions established to manage certain activities.

Regional Inequality Regional inequality is used to refer to differences in the quality of life, wealth, and living standards of people living and/working in different places. The emphasis is on differences in human welfare or well-being, and on related aspects of regional life.

Theil Index The Theil index has its roots in the concept of entropy developed in the field of information theory to measure the amount of information contained in a random event. The central idea was that the information content of an event was an increasing function of the degree of uncertainty associated with its occurrence. The more unexpected an event, the greater its information content, and *vice versa*: telling someone that a ball drawn from a box containing only red balls is red has, for example, zero information content as the outcome is certain. More specifically the information content of an event is inversely proportional to the probability of its occurrence: if there are n events, of which one is certain to occur, if the probability of a each event is x_i and if $\Sigma_{i \quad 1}^{n} x_i \quad 1$, the expected information content or entropy of an event is given by

$$H(x) = \sum_{i \quad 1}^{n} x_i log \frac{1}{x_i}$$

To adapt this index to the measurement of inequality, Theil replaced probabilities with income shares, identified a positive relationship between the degree of entropy and the degree of equality, and arrived at a measure of the degree of inequality by subtracting the resultant measure of equality from its maximum value (which occurs when all individuals get the same income share or all groups get a share that is proportional to the size of the group as a share of the total population).

Suppose that p_i denotes the population of the ith region ($i \quad 1,2, \ldots, n$), and that q_i denotes its GDP. The Theil coefficient (TC) is given by the equation:

$$TC = \sum_{i \quad 1}^{n} \frac{q_i}{\Sigma_{i \quad 1}^{n} q_i} log_e \left(\frac{\frac{q_i}{\Sigma_{i \quad 1}^{n} q_i}}{\frac{p_i}{\Sigma_{i \quad 1}^{n} p_i}} \right)$$

This index can be interpreted "as the expected information content of the indirect message which transforms the population shares as prior probabilities into the income shares as posterior probabilities", and as a measure of the gap between the share of the population that each individual or group accounts for and the share of income it receives.

Introduction

All regions are unique, as indeed are all of the natural and social phenomena about which generalizations are nonetheless made and theories are developed. Generalizations about regions are possible as all regions also have things in common with others: the types of soil, its fertility, and the types of settlement are all characteristics that may be found in other regions. The identification of characteristics that are, in significant respects, common and/or similar requires the use of a process of abstraction. In social science approaches to geography these abstractions are the starting point for the making of generalizations and the development of theories. Once developed, these theories are the starting point for a process of thought which moves in the opposite direction from abstract theories to concrete realties in which all of the details set aside are reintegrated such that particular regions emerge as a synthesis of many determinations. In this social science approach to geography, inequality is not used to refer in a general manner to differences or an absence of sameness. Instead inequality is used to refer to a narrower set of differences in a subset of the characteristics of different places. This subset includes the quality of life, wealth, and living standards of people living and/working in different places. The emphasis is on differences in human welfare or well being, and on related aspects of regional life.

The literature on regional inequalities (which derives from a number of disciplines that include geography, economics, sociology, political science, and regional science) deals with a large range of indicators representing the wealth, income, employment, life expectancy, health, and education of the inhabitants of different places. Most of this literature focuses on income inequalities, although attempts to explain these inequalities involve consideration of possible economic, demographic, political, and cultural differences that may cause income inequalities. In this article most attention will be paid to income inequalities. Attention will also be paid to the identification of regional entities and the implications of regional classifications, indicators of inequality, and the analysis of trends in regional inequality. Some attention will be paid to factors that underpin differences in regional incomes. Much of the research that seeks to explain regional inequalities is considered in the article on regional development models.

Which Territorial Units?

Essentially there are three ways in which regions can be defined. Homogeneous regions are regions that are made up of areas that share certain characteristics. Examples include climatic regions that share a common climatic regime. Other examples include areas classified according to their altitude, their soil and vegetation conditions, or certain population characteristics (such as ethnicity or religious affiliation). Analytical or functional regions are regions that are made up of places that are interdependent. A travel to work area comprising employment centers and the places where the people working in these centers reside, or shopping centers and the places from which people travel to use these retail outlets, are examples. Administrative regions are regions established to manage certain activities. Examples include nation states and subnational government areas, or areas for the provision of health or education services. In this case, regions are defined according, for example, to the size of the population considered necessary to carry out these tasks efficiently, perhaps respecting historical legacies or cultural distinctions. The principle underpinning the establishment of the French *départements* was, for example, the distance that an official could travel in 1 day on horseback. From an analytical point of view, the most sensible areas are functional, as they combine centers of wealth creation, for example, with the residences of the people who create this wealth, although functional areas do have the disadvantage that their limits change relatively rapidly. In most countries however regional statistics are collected for administrative areas whose existence may/may not make any functional sense.

An example of a regional classification used in regional analysis is the European Union's (EU's) Nomenclature of Units for Territorial Statistics (NUTS, although the acronym derives from the French name). The NUTS system is, at present, a four level hierarchical classification: each member state (NUTS 0) is divided into a whole number of NUTS 1 regions; each NUTS 1 region is, in turn, subdivided into a whole number of NUTS 2 regions; and so on. Generally speaking, the first three NUTS levels correspond to the main regional divisions in each member state: *Länder* and *Kreise* in Germany; *régions* and *départements* in France; *Comunidades autonomas* and *provincias* in Spain; and *regioni* and *provincie* in Italy. To secure a degree of international comparability and, more specifically, to try to ensure that the areas lie within certain size limits (3–7 million people in NUTS 1 areas, 800 000–3 million in NUTS 2 areas, and 150 000–800 000 in NUTS 3 areas) a third administrative level is identified for each member state: NUTS 1 for France, Italy, Greece, and Spain; NUTS 2 for Germany; and NUTS 3 for Belgium. At a more detailed level there are districts, municipalities, and communes, although these levels are not yet subject to the formal NUTS regulation.

The difficulty is that the national administrative and statistical regional divisions that are the foundation stones of the NUTS classification are not derived from a common set of criteria and, in some cases, make no functional sense whatever (e.g., Inner London is a NUTS

2 area). Although an attempt is made to identify areas that are comparable in economic size, the size limits are guidelines rather than strict constraints. Consequently, the NUTS classification does not provide a harmonized set of regional units. The 2003 gross domestic product (GDP) at purchasing power standards (PPS) of Level 2 regions varied from Euro 423 398 million in Ile de France and 882.2 million in Åland in Finland. The population of Åland was just 26 302. Another eight NUTS 2 areas had fewer than 250 000 inhabitants. At the other end of the scale, Ile de France had 11.234 million inhabitants, Lombardia 9.177 million, and Andalucia 7.502 million.

In spite of these difficulties, the system is used for three important purposes. The first is the collection, development, and harmonization of regional statistics. The second is the analysis of regional economies and societies. The third is the framing of EU regional policies: eligibility for Objective 1/Convergence Objective Structural Fund assistance is established at the NUTS 2 level, while eligibility under other priority Objectives is mainly established at the NUTS 3 level.

Which Indicators?

GDP *per* inhabitant is the indicator most often used to measure regional inequalities. A community's GDP is a measure of aggregate value added. As the new wealth created in production is distributed to those who participate in economic life, GDP is also a measure of primary incomes (profits, interest, rents, and wages) accruing to those who contribute to a region's economic activities, although the income that results does not necessarily accrue to a region's inhabitants: where the inhabitants of other regions have property rights in a region, there is an outflow of income, as there is if the human capital of other regions is used locally. GDP estimates do not include the value of those goods and services that people produce for their own use. This omission is significant as self provisioning, which depends on the resources a region's inhabitants control and can use, varies significantly from one place to another. The activities of the informal and hidden economies are conversely included at least insofar as government statisticians are able to estimate their magnitude. The primary distribution of income is subsequently modified by state administered redistribution. The result is the secondary distribution of income. These secondary incomes are saved and/or spent, and thereby permit the establishment of claims over the goods and services produced. Measurement of these expenditures (conventionally divided into consumer and government expenditure and investment/savings and net exports) offers a third way of measuring GDP. (The existence of inconsistencies between these three estimates helps in the identification of undeclared incomes.)

Inequalities in regional GDP *per* head can be measured in two ways. Measurements in a common international currency indicate the international value of the output of regional economies. The money value of regional output shows (1) what the output of the exposed sector can be sold for and can command on international markets, and (2) what the output of the protected sector can command directly on regional and indirectly on external markets. If a region's external trade is in equilibrium, this indicator reflects the 'quality' of its goods and services and its competitive strength. Measurements in PPS make allowance for differences in the prices of goods and services in different areas: the quantity of goods and services that a given amount of an international currency will exchange for is greater in a low cost than in a high cost region. The PPS measure allows for these differences in purchasing power and is, therefore, an index of the volume of goods and services that different local economies produce and the volume of output that their economic activities can command or exchange for in their own area. This second measure is, in other words, an indicator of differences in living standards. A related distinction can be made between qualitative and quantitative growth: increases in the value high quality goods and services can command on international markets and quantitative increases in the volume of goods and services that command low prices on world markets.

GDP *per* head and its growth can be split into productivity and employment rate elements, as the following identities indicate:

$$
\begin{aligned}
&\frac{\text{Gross Domestic Product}}{\text{Resident Population}}\\
&\quad\equiv \frac{\text{Gross Domestic Product}}{\text{Employed Population}} \times \frac{\text{Employed Population}}{\text{Resident Population}}\\
&\quad\equiv \left(\frac{\text{Gross Domestic Product}}{\text{Annual Hours Worked}} \times \frac{\text{Annual Hours Worked}}{\text{Employed Population}}\right)\\
&\qquad \times \frac{\text{Employed Population}}{\text{Resident Population}}
\end{aligned}
$$

and

$$
\begin{aligned}
&G\left(\frac{\text{Gross Domestic Product}}{\text{Population}}\right)\\
&\quad = G(\text{Gross Domestic Product}) \quad G(\text{Population})\\
&\quad = G\left(\frac{\text{Gross Domestic Product}}{\text{Annual Hours Worked}}\right)\\
&\qquad + G\left(\frac{\text{Annual Hours Worked}}{\text{Employed Population}}\right)\\
&\qquad + G\left(\frac{\text{Employed Population}}{\text{Resident Population}}\right)
\end{aligned}
$$

This partitioning of GDP *per* head into elements that reflect the productive performance of regional economies and some of the features of their labor markets is widely used as an initial step in the identification of the determinants of regional inequalities.

Which Measures of Inequality?

The simplest measures of inequality are the standard deviation, mean absolute deviation, Gini coefficient (GC), and Theil index. In the case of the first three measures, unweighted and weighted variants are used. An unweighted indicator gives equal weight to the average *per capita* income of each regional economy. An example is a simple standard deviation or coefficient of variation (CV). A weighted indicator makes allowance for the fact that the population of regional economies varies. In this case, the average income of each regional economy is weighted by its number of inhabitants. (Measures of social, as opposed to regional, inequality deal with the incomes of individual households or individuals, each of which is weighted equally instead of regional averages.)

The definitions of the most important of these indicators are set out in the following equations. If p_i denotes the population of the ith region ($i = 1,2, \ldots, n$), y_i denotes its *per capita* GDP, and q_i denotes its GDP, the population weighted standard deviation (WSD) is given by the equation

$$WSD = \sqrt{\sum_{i\ \ 1}^{n}(y_i \quad y)^2 \frac{p_i}{\Sigma_{i\ \ 1}^{n} p_i}} \qquad [1]$$

and the weighted absolute deviation (WAD) is given by the equation

$$WAD = \sum_{i\ \ 1}^{n} |y_i \quad y| \frac{p_i}{\Sigma_{i\ \ 1}^{n} p_i} \qquad [2]$$

To facilitate comparisons the weighted standard deviation and the WADs are expressed as percentages of the mean to give a CV.

The GC is given by the equation

$$GC = \frac{1}{2n\,\Sigma_{i\ \ 1}^{n} y_i} \sum_{i\ \ 1}^{n} \sum_{j\ \ 1}^{n} |y_i \quad y_j| \qquad [3]$$

As with the indicators based on the mean, the deviations from which the GC is calculated can be weighted by the product of the shares of the total population in each pair of regions.

Finally, the Theil coefficient (TC) is given by the equation

$$TC = \sum_{i\ \ 1}^{n} \frac{q_i}{\Sigma_{i\ \ 1}^{n} q_i} log_e \left(\frac{\frac{q_i}{\Sigma_{i\ \ 1}^{n} q_i}}{\frac{p_i}{\Sigma_{i\ \ 1}^{n} p_i}} \right) \qquad [4]$$

Territorial Units and the Measurement of Inequality

In analyzing regional inequality it is important to recognize that measured inequality increases with the level of spatial disaggregation. Suppose a country is divided into 16 areas (A1, A2, … , D4) with similar populations but different levels of GDP *per* head, and that these areas are grouped first into four and then into two regions (see **Figure 1a**). The standard deviation expressed as a percentage of the mean decreases from 38.5% (16 areas) to 10.6% (four areas: A1..B2, A3..B4, C1..D2, and C3..D4) and 3.22% (two areas: A1..B4 and C1..D4). It is important to note, however, that the choice of regional boundaries can affect the result. If, in **Figure 1a**, the 16 areas are divided horizontally rather than vertically into two groups (A1..D2 and A3..D4) the indicator falls to 9.67 instead of 3.22. Alternatively, if four areas are identified in the manner indicated in **Figure 1b** the CV will equal 24.7. Measured regional disparities depend, therefore, not just on the degree of spatial concentration of economic activities, but also on the regional division of the country: the number of areas and the choice of boundaries affect the measure of disparity, just as the delimitation of electoral districts shapes the outcome of elections. Clearly the ideal solution is to use functional economic areas which combine places of work with corresponding places of residence, though disparities between politically identified areas are significant as determinants of the resources over which different communities can exercise political leverage.

National and Regional Inequalities in the EU

Figure 2 records *per capita* GDP at PPS in 2003 for the NUTS 2 regions in the EU25, for NUTS 2 regions in Bulgaria, Romania, and Croatia, and the national average for Turkey. Countries are ordered according to their GDP *per* head from Luxembourg (233.9%) and Ireland (134.1%) to Turkey (26.5%), with national GDP *per* head represented by diamonds. The extreme values differ by a factor of more than 8.8. The figure for Ireland should however be treated with caution. Irish GDP far exceeds

	A	B	C	D
1	5	10	10	5
2	10	10	10	10
3	10	15	20	10
4	5	10	10	5

(a)

	A	B	C	D
10	5	10	10	5
11	10	10	10	10
12	10	15	20	10
13	5	10	10	5

(b)

Figure 1 Measured inequality and regional division.

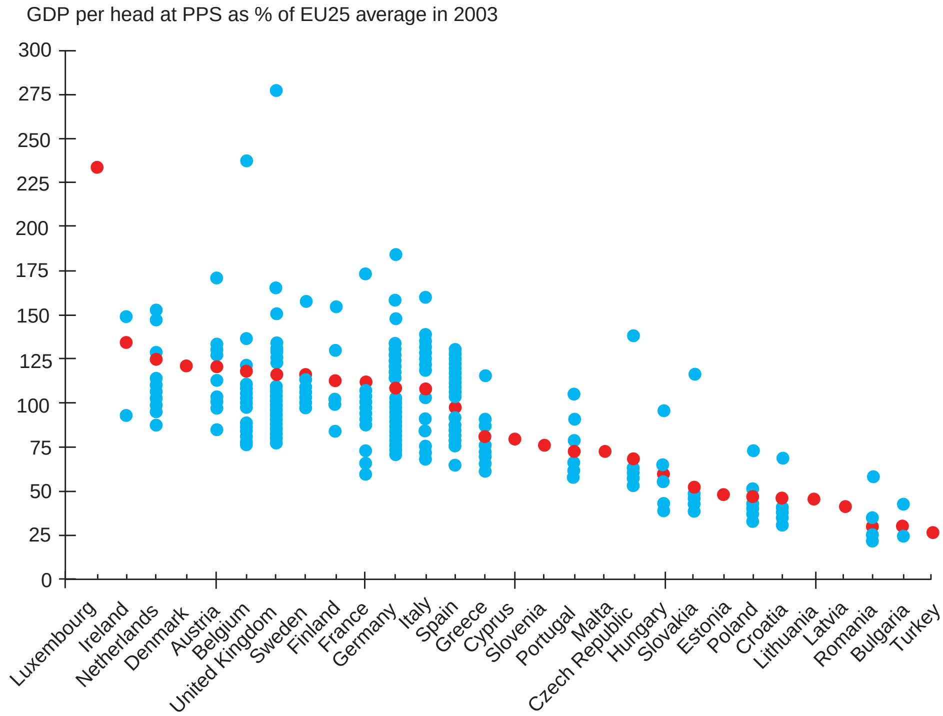

Figure 2 National and regional inequalities in the EU25 plus Bulgaria, Croatia, Romania, and Turkey. Source: Elaborated from data from Eurostat.

the Irish gross national product (GNP): the incomes that accrue to people who live in Ireland is about 17% less than the income created in Ireland. The reason for this discrepancy is that much of the wealth created in Ireland is produced by multinational companies that withdraw profits and engage in transfer pricing arrangements. Transfer pricing results in an overstatement of the wealth created in Ireland. For each country other than Turkey regional GDP *per* head figures are recorded in the col umns as circles. In most countries, and in this group of countries as a whole, there are wide regional disparities. At one extreme lie Inner London (277.6%), Brussels (237.6%), Luxembourg (233.9%), lle de France (173.3%), and Vienna (170.9%), although the figures for the first two areas must be treated with extreme caution as these areas are areas of substantial net inward com muting. At the other end of the spectrum were Nord Est (21.7%) and Sud (24.4%) in Romania, and Severen Tsentralen (24.1%), Yuzhen Tsentralen (24.6%), and Yugoiztochen (24.9%) in Bulgaria.

The Dynamics of Regional Inequalities: Convergence or Divergence?

In 1955, Kuznets suggested that there is an inverted U shaped relationship between inequality and income *per* head: as an economy starts to develop and people move from agricultural to industrial jobs and from rural to urban areas, inequality increases; as income *per capita* increases further, this relationship is reversed and in equality declines. (Other research has pointed to the existence of a nonlinear relationship between urban ization and development.) For Kuznets the technical reason for the increase in inequality was that the tran sition to an industrial society was unskilled labor saving. In 1965, Williamson extended this argument to regional inequalities and argued that regional inequalities will at first increase and subsequently decrease as the level of economic development increases. Williamson argued that the evolution of regional inequalities was a result of a set of spillovers associated with different stages of economic development including migration, capital flows, govern ment policies, and interregional trade. As evidence, Williamson relied on time series data for the first half of the twentieth century which showed that regional in equalities increase in less developed countries and de crease in the more developed countries and on cross sectional data which showed that regional disparities are greater in less developed and smaller in more developed countries. Williamson concluded that regional income inequalities are a natural consequence of economic de velopment and that any attempts to reduce them might inhibit development.

More recent work on the dynamics of regional in equalities involved investigation of the convergence hy pothesis. The initial idea derived from the expectation that economies characterized by similar conditions would converge on a similar steady state rates of growth and similar *per capita* incomes, with economies that are less developed growing faster than economies that are more developed. Empirically this claim is associated with the β convergence hypothesis. According to this hypothesis the slope parameter, $(1 - e^{\beta t})/t$, of regressions of *per capita* GDP growth in t years, $\hat{y}_t$, on the logarithm of initial GDP *per* head, y_0, will be negative (see **Figure 3a**). (It is important to note that this exercise is statistically ques tionable. What is involved is the estimation of an equa tion in which initial income appears on the two sides, once with a positive sign and once with a negative sign. In these circumstances a negative regression coefficient is, at least in part, a result of correlating a variable with itself.)

Two factors underpin the expectation of faster growth of less developed economies. The first is the existence of diminishing returns to capital and constant returns to scale. At any given level of technological development, these conditions imply that increases in *per capita* output get smaller as the amount of capital *per* worker increases (areas with low K/L grow faster than areas with high K/L), that increases in productivity cease once a steady state/equilibrium amount of capital *per* worker is reached, that increases in *per capita* output associated with incremental investments are greatest in areas/enterprises that are the least mechanized and that countries that differ only in initial levels of *per capita* GDP will converge on the same level of GDP *per capita*. The second is the view that technology is a public good available at no cost to everyone and that the diffusion of technology and of knowledge from advanced to less developed enterprises and areas would close technology and productivity gaps.

Empirical convergence hypothesis research showed that there were differences in regional steady state growth rates and permanent differences in income *per* head. As a result, emphasis was placed on theories of conditional and club convergence that predict cross sectional heterogeneity as a result of differences in

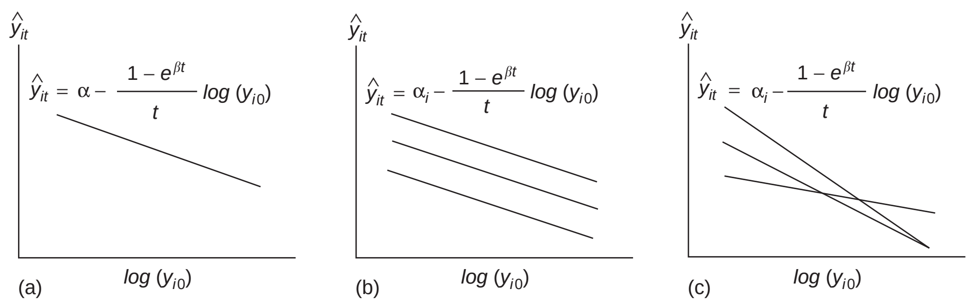

Figure 3 Convergence in reduced form neoclassical growth models: (a) neoclassical convergence; (b) club convergence; and (c) conditional convergence.

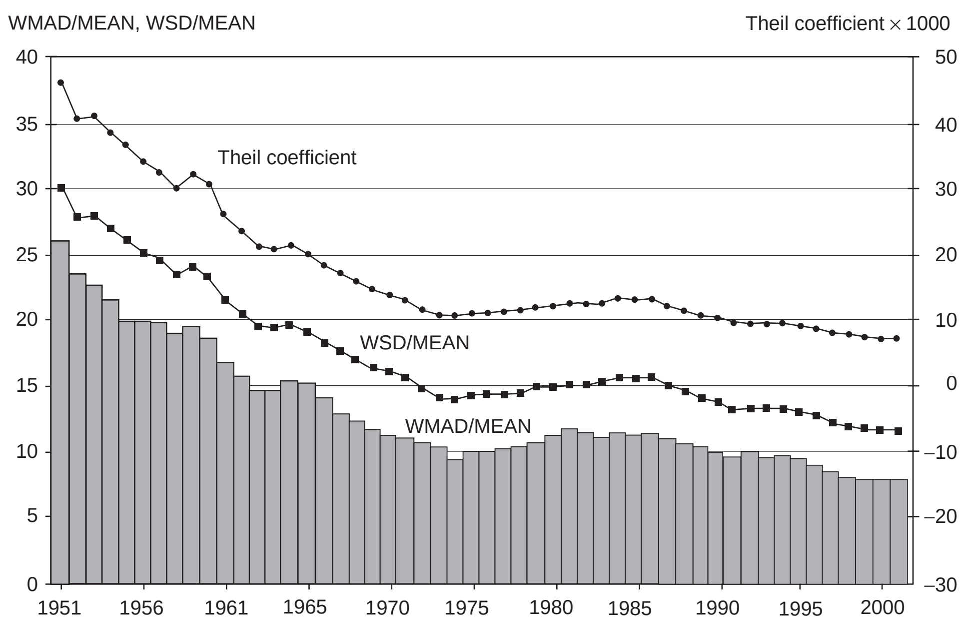

Figure 4 Trends in inequality in GDP per inhabitant in the EU15, 1951 2002. Source: Elaborated from Eurostat.

economic conditions (see **Figure 3b** and **3c**). Theories of club convergence attribute differences in performance to the fact that economies fall into different groups/clubs because of differing initial conditions. Theories of conditional convergence attribute differences to the cross area heterogeneity of control variables, such as rates of accumulation or the quality of modes of governance. In this much more restricted context the convergence that does occur is finally very slow: economies converge at only about 2% *per* year on their own steady state as determined by structural characteristics and starting conditions.

A second strand of empirical research concentrates more on the evolution of the dispersion of regional incomes (sometimes called θ convergence). In this case, the unweighted and weighted indicators identified earlier are used.

A third strand of empirical research is the distribution dynamics approach. This approach involves the application of Markov Chain techniques (using either discrete transition probability matrices or continuous stochastic kernels) to examine the evolution of the entire income distribution. Existing research drawing on these methods has indicated complex changes in the dynamics of relative income *per* head including divergence, the persistence of inequalities, movements within the distribution, and polarization. Modeling the evolution of the distribution of regional and individual income with these techniques, Quah found, for example, a tendency for the emergence of twin peak distributions with high frequencies for relatively high and relatively low incomes and relatively small frequencies for incomes in the middle of the distribution.

Trends in Regional Inequalities

Empirical evidence derived from research employing these indicators places a question mark over the Williamson hypothesis. The reason why is that in many developed economies regional inequalities, which did indeed diminish in the 1960s and 1970s, started to increase again in the 1980s and 1990s. What is more in these countries the rates of economic growth were faster in the years of diminishing regional inequality than in the subsequent years of widening regional inequality.

Figure 4 plots the evolution of disparities between the Member States of the EU15 from 1951 until 2001, while **Figure 5** plots the evolution of regional disparities in Italy. The first chart indicates that there was strong convergence amongst the EU15 Member States from 1951 until the mid 1970s. In this period (often referred to as a postwar Golden Age) European economies experienced unprecedentedly fast economic growth: growth was associated with increases in territorial (and social)

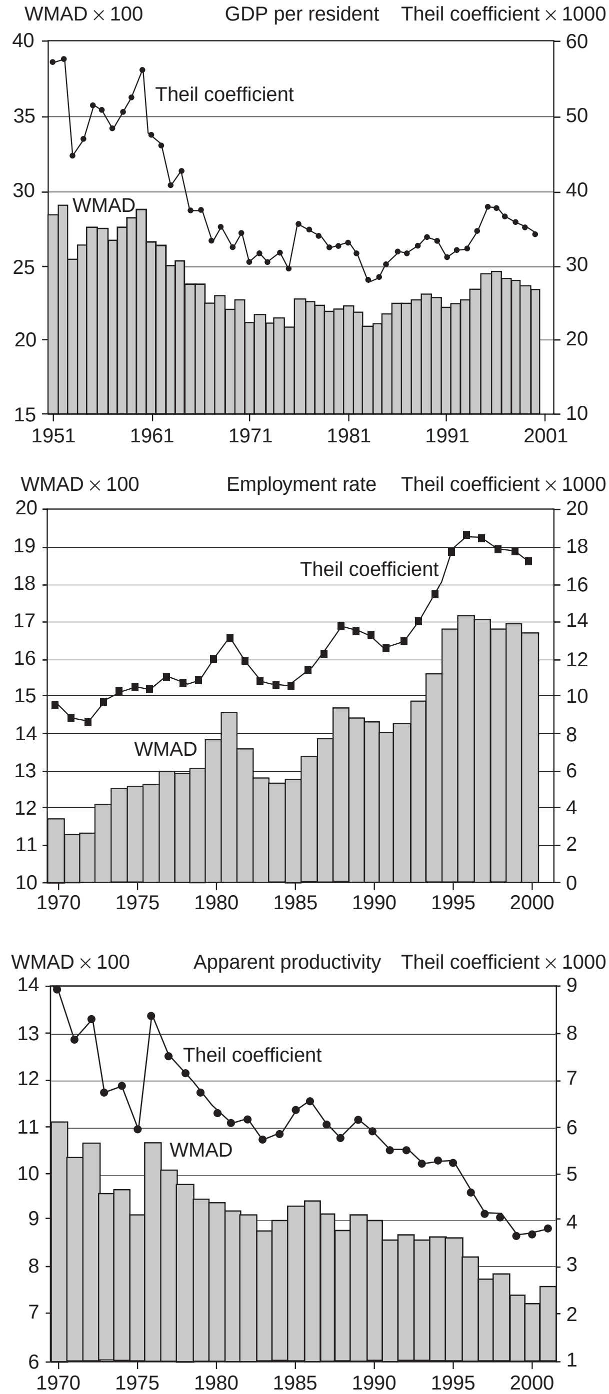

Figure 5 Trends in Italian regional inequality: Disparities in regional GDP per resident, apparent productivity, and employement rates, 1959 2000. Source: Dunford, M. and Greco, L. (2006). *RGS-IBG Research Series After the Three Italies: Wealth, Inequality and Industrial Change*. Oxford: Blackwell.

equality, while the movement in the direction of greater equality made positive contributions to European growth through, for example, their effects on the expansion of markets. Until the mid 1980s, however, divergence prevailed. At that point in time a renewed process of convergence was set in motion as a result of the relatively fast growth of the economically less developed Cohesion Countries and the relatively slow growth of more de veloped EU economies.

In the case of Italy's regional economies there was a sharp decline in disparities in 1951–53 after which terri torial inequalities increased in the 1950s. In 1960, the degree of inequality was close to its 1951 level. In 1960–75 the situation changed dramatically. For 15 years there was strong catchup as the less developed parts of Italy closed the gap on the more developed. (The late 1950s and early 1960s were the years of Italy's economic miracle.) After 1975 there was a reversal in the trend. Overall, there was a clear increase in inequality, with the TC reaching 37.8 and WMAD 24.6 in 1996. A not insignificant part of the relative improvement that had occurred after 1960 was reversed in 1975 and after 1983. In the last few years, conversely, there was a renewed phase of convergence, with the TC declining to 34.2 in 2000.

Also striking is the fact that the recent evolution of Italian regional inequalities is a result of opposite trends in regional productivity and employment rates. As **Figure 5** shows, productivity differentials have tended to decline. In 1972–96 employment rate differentials gen erally moved in the opposite direction. Clearly, it is not productivity differentials that account for the recent overall increase in Italian regional inequality, as pro ductivity differentials have tended to diminish, though productivity differentials remain a significant determinant of 'differences' in the levels of regional de velopment in contemporary Italy. At the root of the in equality increases up to 1996 were very sharp increases in employment rate variations.

As these two examples show, there are relatively sustained periods in which unequalizing tendencies outweigh equalizing tendencies and *vice versa.* (These empirical outcomes also reflect, of course, the impact of government policies designed to reduce development differences of which EU Structural and Cohesion Pol icies are an important example.)

A second important feature of empirical trends in relative development is the persistence of territorial

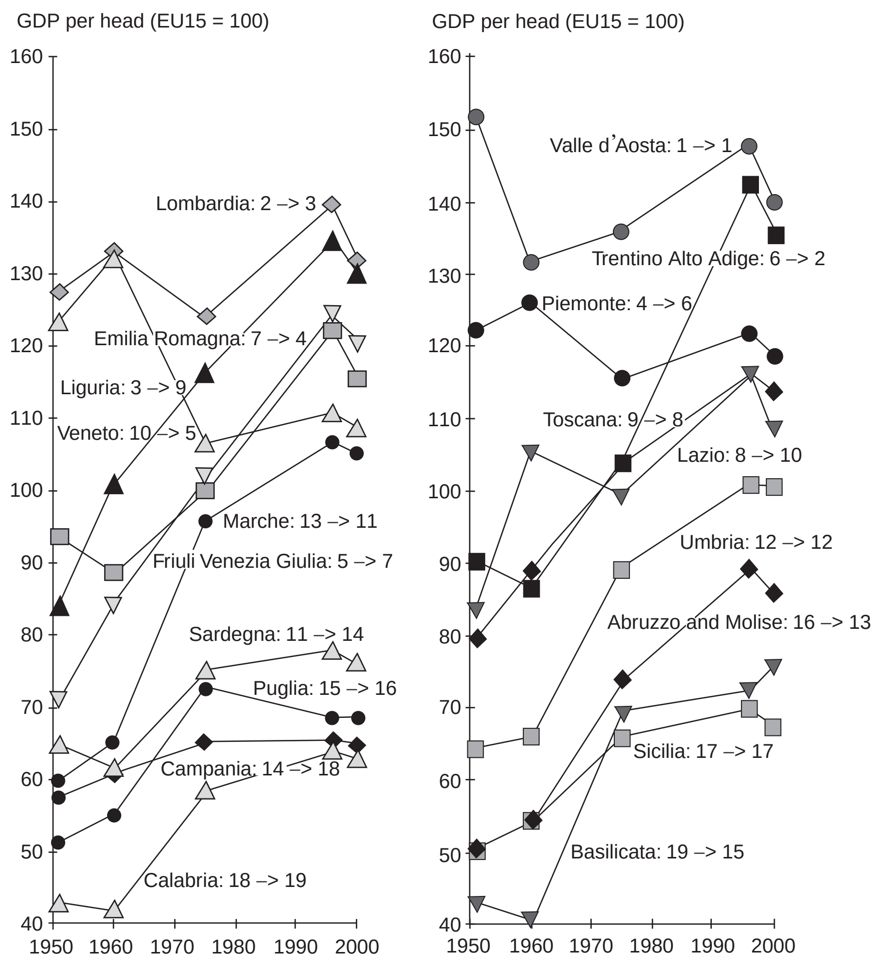

Figure 6 Growth of Italy's regional economies relative to the EU15 average, 1951 2000. Source: Dunford, M. and Greco, L. (2006). *After the Three Italies: Wealth, Inequality and Industrial Change*. Oxford: Blackwell.

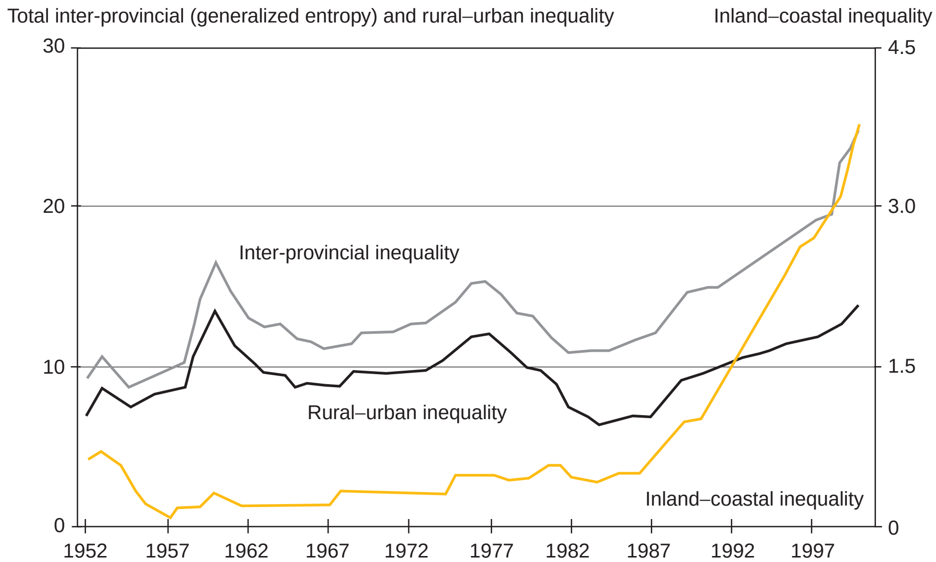

Figure 7 Spatial inequality in China, 1952 2000. Source: Elaborated from Kanbur, R. and Zhang, X. (2001). Fifty years of regional inequality in China: A journey through revolution, reform and openness. *CEPR Discussion Paper No. 2887*. http://ssrn.com/abstract 278564 (accessed April 2008).

inequalities. **Figure 6** records the evolution of Italy's regional economies relative to the EU15 average. Amongst other things, it reveals a high degree of inertia at the top and bottom end of the distribution: the top regional economies in 1951 remained for the most part at the top end of the distribution in 2000, just as the regional economies at the bottom end tended to remain at the bottom, although it also indicates some striking changes in the relative position of regional economies in the middle of the distribution.

Evidence of phases of increasing and diminishing inequalities is not confined to the EU. Kanbur and Zhang's analysis of inequalities in real *per capita* consumption in 28 of China's 30 provinces (Tibet and Hainan were excluded) also identified phases of convergence and phases of divergence. This research also pointed to a close relationship between trends in inequality between provinces, urban and rural areas, and coastal and inland areas (**Figure 7**) and phases in Chinese economic and political development. The early years of Communist rule, marked by land reform, were associated with a low and steady degree of inequality. In 1957 China embarked on the Great Leap Forward (sharply raising investment in heavy industries) which itself gave way to the Great Famine. In these years inequalities increased, peaking in 1960. From 1960 until 1967 inequalities diminished. 1966 saw the start of the Cultural Revolution which opened a phase of increasing inequality, with a peak in 1976. 1978 marked a major turning point. An initial phase of rural reform saw inequalities continue to decline. After 1984 in a period of political decentralization, increased foreign trade, and foreign direct investment, provincial, urban–rural, and inland–coastal inequalities all increased sharply. Overall inequality was relatively small when the development model was favorable to agriculture and the countryside.

Conclusions

In this article most attention has been paid to regional variations in wealth creation. The evolving geography of wealth creation is underpinned by the evolving geography of economic activities and the dynamics of labor markets and population distribution. These drivers are themselves shaped not just by economic mechanisms but also institutional factors and political choices (e.g., including specific policies designed to create a more equal distribution of people and work). Each of these drivers is associated with equalizing and unequalizing tendencies. Whether regional inequalities increase or decrease depends on the relative weight of these two sets of forces and, more generally, on the underlying model of development. Inequalities are, however, not just a result of the geography of wealth creation. Income redistribution and transfers toward economically disadvantaged regions are often substantial so that variations in income and consumption are often smaller than variations in wealth creation. Health, education, and collective services are often entitlements and as such are distributed far more equally than other economic activities.

See also: Europe; Regional Geography I; Regional Geography II; Uneven Development; Uneven Regional Development.

Further Reading

Armstrong, H. and Taylor, J. (2000). *Regional Economics and Policy*. Oxford: Blackwell Publishers.
Armstrong, H. W. and Vickerman, R. W. (eds.) (1995). *Convergence and Divergence among European Regions*. London: Pion.
Barro, R. and Sala i Martin, X. (1992). Convergence. *Journal of Political Economy* 100(2), 223 251.
Bishop, P. and Gripaios, P. (2005). Patterns of persistence and mobility in GDP per head across GB counties. *Tijdschrift voor Economische en Sociale Geografie* 96(5), 529 540.
Dunford, M. (1993). Regional disparities in the European Community: Evidence from the REGIO databank. *Regional Studies* 27(8), 727 743.
Dunford, M. and Greco, L. (2006). *After the Three Italies: Wealth, Inequality and Industrial Change*. Oxford: Blackwell.
Fan, C. C. and Casetti, E. (1994). The spatial and temporal dynamics of US regional income inequality, 1950 1989. *Annals of Regional Science* 28(2), 177 196.
Kanbur, R. and Venables, A. J. (eds.) (2005). *Spatial Inequality and Development*. Oxford: Oxford University Press.
Kanbur, R. and Zhang, X. (2001). Fifty years of regional inequality in China: A journey through revolution, reform and openness. *CEPR Discussion Paper No. 2887*. http://ssrn.com/abstract=278564 (accessed April 2008).
Martin, R. and Sunley, P. (1998). Slow convergence? The new endogenous growth theory and regional development. *Economic Geography* 74(3), 201 227.
O'Leary, E. (2001). Convergence of living standards among Irish regions: The role of productivity, profit outflows and demography, 1960 1996. *Regional Studies* 35(3), 197 205.
Quah, D. T. (1996). Twin peaks: Growth and convergence in models of distribution dynamics. *Economic Journal* 106, 1045 1055.
Williamson, J. G. (1965). Regional inequality and the process of national development: A description of the patterns. *Economic Development and Cultural Change* 13, 3 83.

Relevant Websites

http://ucatlas.ucsc.edu
The UC Atlas of Global Inequality.

Regional Innovation Systems

J. Revilla Diez and M. Kiese, Leibniz University of Hannover, Hannover, Germany

Background

Since the mid 1980s, research about innovation processes has been increasingly directed to the analysis of national innovation systems (NISs). The NIS concept interprets innovation as a systemic and complex process, which is not seen as a linear sequence, as in the neoclassical view, and not only dependent on one single entrepreneur's decisions. Instead, it is accepted that innovation is a result of intensive interactions between different actors within and outside a firm and therefore depends heavily on relationships between the firm and its external environ ment. The business environment influences a firm's in novation activities in two ways. On the one hand, innovative firms use cooperation with other actors to access information, knowledge, and other resources, to exchange experiences, and to jointly find adequate solu tions. Potential cooperation partners are other firms, for example, customers, suppliers, competitors, service pro viders and/or research institutes, financial institutions, and political decision makers. On the other hand, the business environment, which can change in the course of time, determines the firm's behavior. The social and cultural embeddedness, the institutional and regulatory framework, and the existing infrastructure influence entrepreneurial decisions. In this broad sense, innovation is the outcome of interplay between different actors, whose actions are determined by their environment.

A comprehensive analysis of innovation processes has to include numerous factors. According to Edquist, the concept of innovation systems is the only framework in which innovation processes can be examined: "If we want to describe, understand, explain – and perhaps influence – processes of innovation, we must take all important factors shaping and influencing innovations into account. The system of innovation approach – in its various forms – is designed to do this. Attempts to understand the structure and dynamics of such systems are at the core of modern thinking about innovation processes" (Edquist, 1997).

In their concept of an NIS, Nelson and Rosenberg comprise all institutions whose interactions determine the innovative performance of a nation's companies. This is by and large comparable to Freeman's definition of an NIS as a "network of institutions in the public and private sectors whose activities initiate, import, modify and dif fuse new technologies" (Freeman, 1987). Lundvall's ver sion of NIS goes beyond this rather narrow perspective. The much broader 'Aalborg version' of the NIS concept emphasizes the accumulation of tacit knowledge through routine processes of learning by doing, learning by using, and learning by interacting. Consequently, an NIS comprises "all parts and all aspects of the economic structure and the institutional set up affecting learning as well as searching and exploring." Hence, the NIS concept has successively evolved from a system of knowledge production through research and development (R&D) into a system of innovation and learning in which the generation of technological knowledge is linked to the economically relevant aspects of knowledge diffusion, transfer, and application.

Elements of Innovation Systems

Generally speaking, a system of innovation comprises all actors, organizations, and institutions whose actions and interactions influence innovation processes or set the framework for the intensity and direction of technolo gical change. Like all systems, systems of innovation thus consist of elements and networks of interaction between them. Across all territorial and nonterritorial scales dis cussed below, innovation systems can be generalized to include the following organizations or 'players':

1. Empirical research on innovation traditionally centers upon the Organisation for Economic Co operation and Development's (OECD) Oslo Manual definition of technological product and process innovations as im plemented technologically new products and processes and significant technological improvements in prod ucts and processes. This technological paradigm places manufacturing firms at the core of an NIS as they generate new technological knowledge through formal R&D and informal processes of learning, which they subsequently introduce to the market. Furthermore, they are also seen as prime users and adaptors of new technology. More recently, service firms have also been discovered as major sources of innovation provided that a broader concept of innovation is applied.
2. Besides being a significant source of innovation themselves, knowledge intensive business services (KIBSs) assume more and more importance for the support of innovation in manufacturing industries. As they are increasingly involved in processes of inter active learning, they speed up the circulation of technical knowledge, thus forming a 'second tech nology transfer infrastructure' and easily out performing the old established bridging institutions such as technology transfer agencies.

3. By performing basic and applied research, organiza tions of knowledge production and diffusion such as universities, polytechnics, and publicly funded re search institutions are essential to the development and diffusion of technical knowledge. Their inte gration into international research networks allows them to absorb new scientific knowledge worldwide, which they can then feed into their home territorial knowledge base through collaboration with national, regional, or local partners.
4. Furthermore, universities and polytechnics promote the diffusion of technology through their graduate output, as does the rest of the education system (schools, vocational training, etc.). The supply of human capital is critical for a system's absorptive capacity and hence especially for newly industrialized economies (NIEs), a point to be elaborated later on.
5. Bridging institutions were chiefly established in the 1980s by universities, chambers of commerce, and the like to match the needs of science and technology. They include not only transfer agencies, science parks, and business incubators, but also financial institutions such as banks and venture capitalists. Especially sci ence parks received a fair amount of criticism for essentially reflecting the outdated linear model of innovation; as the surge of KIBS shows, the transfer of knowledge is increasingly organized by the market.
6. Further organizations are responsible for the formu lation and implementation of policies affecting in novation, such as parliaments, ministries, other public agencies or councils for science and technology. Regulatory bodies setting standards and industrial norms or awarding certifications fulfill similar func tions, as do professional organizations, trade unions, academies, and so forth.

If these are the 'players', then their actions and inter actions are subject to a set of constraints or 'rules of the game', such as constitutions, laws, incentives, technical standards, cultural norms, values, conventions. The per formance of an innovation system depends crucially on the extent to which the different actors network with each other and aim at joint solutions. This requires an inclination for cooperation, which is in turn dependent on mutual trust. The sociocultural environment can produce institutional arrangements which can lead to routine like behavioral patterns and to a reduction of transaction costs in the face of technological as well as market uncertainties.

Systems of innovation can be delineated according to territorial or technological criteria. Studies on the geography of innovation naturally focus on territorial systems at various scales, that is, global, supranational, or continental, but most frequently at the national, regional, or local level. However, we are going to show that due to the limitations of territorial innovation models, non territorial systems and spaces of innovation should also be taken into account.

Regional Systems of Innovation

Since the mid 1990s, the increasing attention paid to the regional and local level led to a fairly high number of studies about regional innovation systems (RISs). Again, their point of departure is the chain linked innovation model which features feedback loops and interactions between different actors during innovation processes. The authors basically scale the systemic elements of an NIS down to the regional level. In short, they argue that region specific conditions and cooperation relationships between different actors influence the regional innov ation potential either positively or negatively. It is as sumed that regions which possess the full set of systems elements outlined above within an institutional milieu of systemic linkages and interactive communications among the innovation actors exhibit a better innovative per formance than other regions.

However, an RIS is commonly seen as more than a scaled down version of an NIS since the regional scenery can deviate from the national setting in a number of ways, such as a region's industrial structure, educational in frastructure, scientific and technological capabilities, or its capacity to absorb knowledge from outside the region. Both organizations and institutions can be limited in their territorial outreach. For instance, territorial innovation models such as industrial districts, innovative milieus, or learning regions stress the importance of a common cultural basis for building trust, thus reducing transaction costs and uncertainty while allowing for collective learning. Especially if a high level of knowledge tacitness is involved, it is argued that localized learning through frequent and often informal face to face interaction be comes a critical precondition for innovation, thus allowing for a faster regional circulation of knowledge and consequently innovative performance that cannot easily be achieved at a higher spatial scale.

Typology of Regional Systems of Innovation

According to a major comparative survey conducted by Cooke, Heidenreich, and Braczyk, RIS can be measured and identified along two dimensions. First, the govern ance dimension comprises public policy, institutions, and knowledge infrastructure as an integrated part of regional innovation processes. The key question is whether the organizational setting, for example, regional adminis trative bodies, animates or facilitates networking among actors of the innovation system. Second, the business innovation dimension is often linked to the industrial base characterized in terms of productive culture and systemic innovation. Of special interest are the level of the firm's absorptive capacity and the role of upgrading technological capabilities, the amount of R&D spending and the degree of linkages and network relationships.

Based on the business innovation dimension, the survey identifies different types: the localist, the interactive and the globalized RIS. A 'localist' RIS has relatively few large indigenous firms and externally controlled affiliates with major research facilities. Major public R&D innovation or R&D resources are also scarce, but there may be smaller private ones. However, the lack of large establishments is made up for by a high degree of associationalism within the business community, as well as between firms and policymakers. Examples of localist RIS include Tuscany in Italy, Tampere in Finland, or Slovenia. An interactive RIS displays a reasonable balance between large and small firms providing both regional research capacities and access to foreign sources of knowledge relevant to innovation. There is also an even mix of public and private research facilities and a dense network of associational activities. The German state of Baden Württemberg is still seen as the archetypal interactive RIS, the Spanish region of Catalonia and Gyeonggi in South Korea are given as further examples. Finally, globalized RISs are dominated by large firms with access to international, but largely internalized knowledge networks. A public research infrastructure may be helpful in upgrading the technological capabilities of small and medium sized enterprises (SMEs) in local supply chains. Typical examples of such globalized RIS dominated by foreign multinationals include Wales or Singapore.

The governance dimension provides an alternative classification into three forms of RIS, namely grassroots, network, and dirigiste. A grassroots RIS is generated and organized locally, at town or district level. Financial support and research competences are diffused locally, with a very low amount of supralocal or national coordination. Local development agencies and local institutional actors play a predominant role. Examples for a grassroots RIS are Tuscany, Italy, or Catalonia, Spain. By contrast, a network RIS is more likely to occur when the institutional support encompasses local, regional, federal, and supranational levels, and funding is often guided by agreements among banks, government agencies, and firms. The research competence is likely to be mixed, with both pure and applied, blue sky (exploration) and near market (exploitation) activities geared to the needs of large and small firms. As for the interactive business innovation model, the literature uses Baden Württemberg as the example of a network RIS. Finally, a dirigiste RIS is animated mainly from outside and above the region itself. Innovation often occurs as a product of central government policies. Funding is centrally determined, with decentralized units located in the region and with research competences often linked to the needs of larger, state owned firms in or beyond the region. A fitting example for such a dirigiste RIS is the aerospace cluster at Toulouse in Aquitaine, France.

During the last few years, a large number of studies were conducted in order to define the nature, activity, performance, and development of RIS. The studies focused on experiences of cities, agglomerations like metropolitan regions, and regions with high tech activities. Especially metropolitan regions interpreted as metropolitan innovation systems tend to show better innovative performance as they provide a wider pool of knowledge and potential cooperation partners. Recent studies have been done on nonurban/metropolitan innovation systems as well. Nevertheless, the recent focus on case study research, which is driven by a strong desire to identify preconditions for success and best practice models for policymaking, means that the resulting classifications remain essentially simplistic and descriptive. Business innovation and governance are undoubtedly two dominant dimensions of RIS, but clearly insufficient in themselves to understand the structure of RIS elements and their complex interactions.

The RIS concept can be credited for highlighting the systemic nature of regional innovation activities and the relevance of the subnational scale for innovation processes organized on superordinate territorial scales. However, both the RIS concept and the entire family of 'territorial innovation models' tend to pay unduly little attention to system openness and integration in national and international networks, that is, the cross scalar nature of innovation. Eventually, no RIS can exist in isolation from sectoral and/or territorial innovation systems at higher territorial scales, for which they fulfill specific functions.

Nonterritorial Systems of Innovation

Nonterritorial systems of innovation include all attempts to delineate systems of innovation based on technological or industry criteria; their territorial dimension arises as a complementary effect only, if at all. A technological system can be defined as a network of vertically and horizontally linked actors and organizations interacting in a specific industry with a specific infrastructure to generate, diffuse, and apply technological knowledge. In this perspective, technological systems comprise three elements: (1) the economic competence of firms which is shaped by learning, bounded rationality, as well as the national and regional environment; (2) strong reciprocal externalities resulting from the networking of producers, customers, and competitors; (3) organizations and institutions, including the educational system, bridging and financial institutions, associations, and the degree of internationalization of the industry in question.

Breschi and Malerba choose a different approach by defining a sectoral innovation system (SIS) as all firms participating in an industry's innovation activities by manufacturing products, or by developing or using the respective technology. These firms are systematically linked by interaction and cooperation in the generation of new technology, as well as by competition on the market. In contrast to the technological system, the SISs focus on firms as the key player and the importance of competitive dynamics and the market environment in the process of innovation. Quite differently from territorial systems of innovation discussed above, the territorial extent of their SIS is explained endogenously as a result of the technological regime. A regime is characterized by technological properties, such as the opportunities and conditions for appropriating technology, the degree of cumulativeness, the kinds of knowledge involved (e.g., tacit vs. codified), and the mode of transfer and communication. Since SISs are organizationally and institutionally embedded, in this view a territory does not define a system of innovation, but rather fulfills specific functions in various SISs.

From a spatial science perspective, it is quite unsatisfactory that SISs do not provide a proper means to delineate systems of innovation in territorial space. The concept treats the territorial embeddedness of technological trajectories only vaguely and largely neglects the institutional infrastructure that is inseparable from a specific territorial scale, that is, national or regional. Although concepts of nonterritorial systems of innovation usefully complement those of territorial systems of innovation, they alone are insufficient to replace them.

Concrete examples illustrate the overlap between territorial and nonterritorial innovation systems in practice, that is, innovation processes do not only cut across various territorial scales (local, regional, national), but also involve nonterritorial or sectoral systems of innovation at the same time. For instance, Singapore's semiconductor industry is integrated in the city state's NIS as well as in the industry's wider regional and global production networks. Oinas and Malecki (2002: 102–131) capture this multiscalar nature of innovation in their functional concept of spatial innovation systems which consist of "overlapping and interlinked national, regional and sectoral systems of innovation which all are manifested in different configurations in space." However, the relative importance of technological versus territorial factors in explaining technological capabilities largely remains uncharted territory, as do the interactions between technology and geographic space. To make matters worse, technology and territory form only two – albeit presumably not insignificant – dimensions of innovation space. Hence, multidimensional concepts of space are needed to capture collaborative innovation processes, of which territorial space is just one dimension.

Toward a Multidimensional Concept of Innovation Space

The distinction between territorial and nonterritorial systems of innovation eventually boils down to different concepts of space which can be employed to 'locate' innovation processes in a more differentiated manner.

Geographical proximity between innovation actors can foster their interaction, thus facilitating the exchange of tacit knowledge. However, the relevance of geographical proximity for interactive innovation processes tends to vary between industries and by the maturity of the innovation process itself, increasing with the ratio between tacit and codified knowledge employed. Industry differences are usually explained by technological parameters: if the technology is sufficiently stable and standardized, relevant knowledge can be transferred across any distance at low cost. As long as the technology is highly complex and still subject to frequent changes, geographical proximity can improve the innovative performance of the actors involved. Hence, geographical proximity is most important in times of radical change, or during shifts from one technological paradigm to another. During such a period of uncertainty, face to face contacts can assume paramount importance for the exchange of noncodified knowledge – a mechanism which explains the highly localized nature of new basic technologies, such as microelectronics in Silicon Valley's early days. While this argument is about technological 'systems', the relevance of geographical proximity tends to decline in the course of a 'product's' development process. During the early stages of the innovation cycle, tacit knowledge and learning by interacting demand frequent face to face contacts, before the successive codification of knowledge reduces the need for interaction, while at the same time allowing the use of other means of communication.

Consciously or not, scholars of the geography of innovation tend to overestimate the relevance of geographical space in collaborative innovation processes. Depending on the industry and region under study, empirical research into the relationship between geographical proximity and collaborative innovation yield very different results, altogether far from supporting a clear causal link. For instance, Boschma argues that unlike cognitive proximity, geographical proximity *per se* is neither a necessary nor a sufficient condition for interactive learning. If geographical proximity facilitates interactive learning, it mainly does so by strengthening cognitive, organizational, social, or institutional proximity. It therefore seems inevitable not just to investigate the cross scalar territorial nature of innovation, but also to take nonterritorial dimensions of space into account. Both territorial and nonterritorial spaces make up a multidimensional space in which interactive innovation processes can be mapped. Besides territorial space and its ensuing costs of overcoming physical distances, the following interdependent nonterritorial dimensions of space can be distinguished:

1. Techno economic space: To engage in collaborative innovation, actors must belong to the same field of technology, or to related fields at least. However, these fields of technology are not equivalent to product or industry classifications used in official statistics. To overcome this problem of concordance, they need to be matched using lists of corresponding products, industries and fields of technology. Input–output coefficients may yield another possibility for measuring the distance between actors in economic space. This dimension corresponds to the notion of cognitive proximity: in order to benefit from interactive learning, actors need to have a certain degree of overlap in their knowledge bases to enable mutual understanding. However, too much cognitive proximity limits the potential for learning as knowledge bases and interpretative schemes become too similar.
2. Organizational space: Organizations coordinate the action of their individual members by a set of specific rules, norms, and routines such as a common language or repository of knowledge. It is commonly argued that multinational corporations bridge their diverse geographic and cultural setups by building collective beliefs such as 'corporate cultures' that lead to a convergence of individual behavior, thus substituting territorial closeness by organizational proximity. Interactive learning between organizations thus requires actors first to develop common codes which help reduce the organizational distance between them.
3. The related concepts of 'relational', 'cultural', and 'social space' capture noneconomic relations between individuals based on similar attributes such as age, profession, language, mutual sympathy, and the like, values such as work ethics, behavioral norms, or shared opinions, a common cultural basis as well as mutual trust. These features determine the social cohesion between innovation actors, thus directly feeding into the innovative performance of organizations at the microlevel. At the macrolevel, the cultural space draws together the institutional environment that enables individuals and organizations alike to decode relevant information and employ it in their innovation processes. As evident from the diverse family of territorial innovation models, this dimension of space overlaps considerably with territorial space. Indeed, many mechanisms commonly attributed to geographical proximity are in fact based on a relational concept of space. However, the overlap is by no means complete, since modern communications technology allows relational proximity to be achieved at a distance.
4. The concept of 'institutional space' combines the organizational and relational dimensions by comprising both organizations or 'players' as formal and 'rules of the game' as informal institutions. A number of authors suggest that institutional proximity might be more important than territorial proximity in constituting the 'soft' architecture of learning.
5. Furthermore, 'temporal space' captures the notions that partners engaging in collaborative innovation projects need to have a roughly similar time horizon. This dimension can apply to intertemporal complementarities arising during the innovation process, as well as to a convergence of individual plans for the future, anticipations and behavioral patterns, which is frequently observed in cooperative projects. Temporal space, however, should not be confused with temporary territorial proximity as achieved on conferences or trade fairs within epistemic communities or communities of practice.

To conclude, it remains unclear to what extent the various dimensions of space can be substituted for each other. So far, the RIS concept still represents a territorial paradigm that fails to incorporate multidimensional concepts of innovation space. Taking RIS or territorial systems of innovation more generally forward thus requires integration with the readily available concept of SISs. Territorial and technological notions of space conflate in the popular concept of regional clusters of industry, which in turn needs to incorporate the systemic approach of the NIS and RIS literature more stringently. More conceptual and empirical work is clearly needed here. In the meantime, the complex interactions between different dimensions of space render it unrealistic to expect interactive learning to be induced by territorial proximity alone, or learning by being there.

See also: Embeddedness; Innovation; Networks.

Further Reading

Asheim, B. T. and Gertler, M. S. (2005). The geography of innovation: Regional innovation systems. In Fagerberg, J., Mowery, D. C. & Nelson, R. R. (eds.) *The Oxford Handbook of Innovation*, pp 291 317. Oxford: Oxford University Press.

Boschma, R. A. (2005). Proximity and innovation: A critical assessment. *Regional Studies* 39, 61 74.

Breschi, S. and Malerba, F. (1997). Sectoral innovation systems: Technological regimes, Schumpeterian dynamics, and spatial boundaries. In Edquist, C. (ed.) *Systems of Innovation: Technologies, Institutions and Organizations*, pp 130 155. London: Pinter.

Bunnell, T. G. and Coe, N. M. (2001). Spaces and scales of innovation. *Progress in Human Geography* 25, 569 589.

Cooke, P. (1992). Regional innovation systems: Competitive regulation in the new Europe. *Geoforum* 23, 365 382.

Cooke, P., Heidenreich, M. and Braczyk, H. J. (eds.) (2004). *Regional Innovation Systems: The Role of Governance in a Globalised World* (2nd edn.). London: Routledge.

Diez, J. R. and Kiese, M. (2006). Scaling innovation in Southeast Asia: Empirical evidence from Singapore, Penang (Malaysia) and Bangkok. *Regional Studies* 40, 1005 1023.
Doloreux, D. and Parto, S. (2005). Regional innovation systems: Current discourse and unresolved issues. *Technology in Society* 27, 133 154.
Edquist, C. (1997). *Systems of Innovation: Technologies, Institutions, and Organizations*. London: Pinter.
Fischer, M. M., Diez, J. R. and Snickars, F. (2001). *Metropolitan Innovation Systems: Theory and Evidence from Three Metropolitan Regions in Europe*. Berlin: Springer.
Freeman, C. (1987). *Technology Policy and Economic Performance: Lessons from Japan*. London: Pinter.
Lundvall, B. A. (1992). *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*. London: Pinter.
Moulaert, F. and Sekia, F. (2003). Territorial innovation models: A critical survey. *Regional Studies* 37, 289 302.
Nelson, R. R. and Rosenberg, N. (1993). Technical innovation and national systems. In Nelson, R. R. (ed.) *National Systems of Innovation: A Comparative Study*, pp 3 21. Oxford: Oxford University Press.
Oinas, P. and Malecki, E. J. (2002). The evolution of technologies in time and space: From national and regional to spatial innovation systems. *International Regional Science Review* 25, 102 131.

Relevant Websites

http://www.druid.dk/
DRUID.
http://www.globelics.org/
Globelics.

Regional Integration

A. Cumbers, University of Glasgow, Glasgow, UK

Glossary

Comparative Advantage A key principle of international trade theory dating back to the political economist, David Ricardo, in the early nineteenth century. Ricardo believed that in trading relations between two countries, each country should specialize in those goods that it can produce more cheaply than others and import those goods that are most expensive to produce locally. Through specialization both countries gain by producing those goods they have a relative advantage in and importing those that they have a relative disadvantage. The theory is still influential and has been adapted by neoliberal economists in the contemporary economy to argue for increased regional integration, particularly with regard to the Single European Market project and the North American Free Trade Agreement (NAFTA).

Global Triad A term that had developed as part of the debate about economic globalization to highlight the concentration of world trade within and between the three RECs or macro-regions of North America, Western Europe, and East Asia, at the expense of the rest of the world. The existence of a Triad raises questions about the real extent of globalization.

Preferential Trading Agreements Agreements between states to create larger economic spaces for increasing trade between the firms of the states concerned. At the same time, PTAs often exclude or disadvantage in some ways third countries who are not party to the agreement.

Regional Division of Labor A form of the spatial division of labor present in REBs where a core–periphery relationship develops between locations, with the higher value-added and more skilled jobs concentrating in core regions (e.g., London, New York) and the lower-paying and lower value-added jobs being competing for by more peripheral localities (e.g., Portugal, northern Mexico).

Regional Economic Blocs Blocs created at the macro-regional scale (e.g., Western Europe, North America, South and East Asia) to encourage greater trade and market integration, primarily through bringing down barriers to trade. They vary in the depth of the economic relations, from full political integration (e.g., EU) to looser federations that have more limited trading agreements.

Regional Integration A term referring to processes of geographical integration at the supranational level and increasingly prominent in all parts of the globe. Predominantly driven by processes of market integration and the geographical expansion of capitalism across the borders of nation-states, regional integration is primarily driven by economic processes although the need to regulate economic development is also leading to debates about greater political integration. So far though, only the European Union has witnessed significant development of legal and regulatory structures at the political level.

Introduction

A growing geographical trend since the end of World War II has been the tendency for groups of countries to create regional economic blocs (REBs) to stimulate processes of market integration and trade. The most advanced form of regional integration is the European Union (EU), where the attempt to create a continental economic trading bloc has been accompanied by forms of political and social integration. It has also seen suc cessive waves on enlargement, so that by 2007 it had grown from its original 6 members to 27 member states. However, every continent now has some form of REB, although the depth and extent of integration are highly uneven in practice. The other most significant devel opment has been the North American Free Trade Agreement (NAFTA), creating a single, tariff free market between Canada, the United States, and Mexico. At the other end of the scale, less substantive free trade zones or 'customs unions' have been created in East Asia (ASEAN Free Trade Association or AFTA) and in Africa (Central African Customs and Economic Union).

Although there has been much talk over the past two decades about a globally integrated economy, processes of regional integration have, if anything, been equally significant in transforming the world economy. Three world macro regions – North America, Western Europe, and East and Southeast Asia – sometimes referred to as the 'global triad' dominate the global economy, ac counting for 86% of world gross domestic product (GDP) and merchandise exports. Moreover, world trade has become highly 'regionalized' in recent years in the sense that, for the majority of countries, the most im portant trading partners are neighboring states. For the world's most significant trading region, Western Europe, over two thirds of trade takes place internally while for

Asia, as the second largest trading region, the figure is around a half, with the figure for North American internal trade being around 40% of the total.

Geographers have been at the forefront in developing our understanding of processes of regional economic integration. The simplistic view of a Triad in particular has recently been challenged by those who argue that a more complex set of geographical flows and connections can be discerned as part of globalization. The political geographer, John Agnew, has recently suggested that the notion of coherent regional trading blocs is often more to do with political aspirations than with economic realities. This view has received some support by the work of the economic geographer Jessie Poon who has carried out some detailed disaggregated analysis of foreign direct investment and trade flows between countries. Using analysis that compared flows of capital between 1985 and 1995, Poon and colleagues found a complicated picture whereby trade flows seemed to be becoming more regionalized but FDI flows were becoming more globalized. Rather than contiguous blocs emerging in terms of concentrated investment: "the realization of investment blocs is more likely to be articulated in the form of 'network' regions" (Poon *et al.*, 2000: 440) dominated by particular countries (e.g., US, Germany, Japan) and their global span of investments.

Politically, processes of regional integration have provoked considerable controversy; although the creation of larger market spaces has been viewed as an opportunity for many firms, the encroachment of competitors into hitherto protected or sheltered local and national markets has been a threat to the livelihoods of others. An important criticism has been that market integration favors larger firms and more advanced economies over smaller firms and less developed local and national economies. Countering this criticism, market enthusiasts contend that the creation of larger market spaces will produce more efficient economies and increase competition, benefiting consumers and enhancing economic growth in the longer term. Further criticisms have been voiced by nationalist politicians concerned that economic integration might lead to greater political integration and weaken the sovereignty of nation states. In reality, political and economic interests tend to be intertwined in processes of regional integration.

Defining Regional Integration

Regional economic integration is characterized first and foremost by the setting up of preferential trading agreements (PTAs) between independent states. In practice, such PTAs can take diverse forms with, as we have seen, a spectrum ranging from the highly regulated and developed EU (see below) to looser trading arrangements such as AFTA. An important point however that characterizes all PTAs is that they are state driven, in the sense that they involve agreements between governments to secure potential economic advantage for their own interests. Thus, though essentially economic entities, they are created through political action. There are also usually two aspects to them: creating an enlarged market space for increasing trade between domestic firms and those of other states in the PTA, but at the same time creating new barriers to exclude previous trading partners.

Dicken identifies four types of PTA, depending on the level of regional integration (**Table 1**). At one end of the spectrum is the 'free trade area' which involves the removal of restrictions on trade between countries but does not affect trade with other countries. The 'customs union' operates at a slightly higher level with a common external trade policy being introduced alongside the internal free trade area. The third type of PTA is the 'common market' which begins to resemble the economy of a single state with the free movement of the factors of production (land, labor, capital) being added to the features of a customs union. Finally, an 'economic union' represents the highest level of regional integration short of political union, with economic policy (e.g., monetary union and setting of interest rates) also being harmonized at the higher scale.

Although regional integration has been most pronounced in the period since the end of World War II, it is far from a new phenomenon. An early form of REB was created in Germany in 1834 when the then 38 independent states of the German Federation established a customs union as a spur to economic growth during the industrial revolution. It was also a protectionist measure, however, with an external tariff on products from other states. Primarily an economic union, like more recent forms of regional integration, there was a political dimension. The powerful northern state of Prussia was the major initiator of the Zollverein, which was seen as critical in developing an independent regional trading bloc to the Austrian empire which was a key rival for geopolitical supremacy in Central Europe. The Zollverein in this way helped pave the way for eventual German political unification in 1870.

However, processes of regional integration have been particularly notable from the late 1980s onwards and, as such, have been closely linked to broader processes of economic globalization. The increasingly interdependent nature of global economic activity has spurred an attempt to 'scale up' the process of governance, constructing new rule for international trade while at the same time removing barriers to the functioning of business and markets at the level of the nation state. The most developed form of REB has been that of the EU which we consider here in a little more depth.

Table 1 Types of regional integration with examples

Type of integration	*Examples*	*Countries involved*	*Key date(s)*
Economic Union	European Union	Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, France, Finland, Germany, Greece, Hungary, Italy, Ireland, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, United Kingdom	1957 European Common Market established; 1992 European Union and Single Market created
Common Market	Mercosur (Southern Cone Common Market)	Argentina, Brazil, Paraguay, Uruguay, Venezuela	1991
Customs Union	ANCOM (Andean Common Market)	Bolivia, Colombia, Ecuador, Peru, Venezuela	1969 (revived 1990)
Free Trade Area	AFTA (Asean Free Trade Agreement)	Brunei Darussalam, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand, Vietnam	1967 (ASEAN); 1992 (AFTA)
	NAFTA (North American Free Trade Agreement)	Canada, Mexico, United States	1994

Source: Derived from Dicken, P. (2007). *Global Shift: Mapping the Changing Contours of the World Economy*, ch. 6, pp 187 204. London: Sage.

The Case of European Integration

The EU – or European Economic Community as it was first known – was formed by six nation states (Belgium, France, Germany, Holland, Italy, and Luxembourg) through the Treaty of Rome in 1957. The integration process was originally motivated more by political con cerns than economic ones. In particular, the two world wars of the twentieth century, fought primarily on the European mainland, had the effect of producing a new generation of politicians (after 1945) determined to forge closer and more cooperative ties between their countries. A key element was forging stronger economic ties through creating cooperation initially in the coal and steel industries but later by extending this collaboration to create a 'common market'. The European Community developed initially therefore as a customs union, with internal barriers to trade and a common external tariff against imports from other countries established by the late 1960s. The latter was what distinguished the EC from a simple free trade area. Common policies on agriculture and fisheries were also introduced in the 1960s. Successive enlargements in 1973, 1981 and 1986, 1994, 2004 and 2007 have expanded the EU to 27 members, incorporating former communist countries in Central and Eastern Europe.

By the 1970s pressures for further integration were driven more by economic concerns and in particular the need for European firms to remain competitive in relation to the US and Japan. Competitiveness would be enhanced through economic integration based on the removal of barriers to trade between members, in cluding tariffs and various regulations and restrictions. After a period known as Euro sclerosis in the late 1970s and early 1980s economic integration gained mo mentum again in the 1980s when, in 1985, EU states agreed upon a program to establish a Single European Market (SEM) based on the free movement of goods, services, labor, and capital by January 1993. In this sense, the Single Market was the most advanced attempt to implement neoliberal free market ideology. In total over 300 non tariff barriers to trade were eliminated by 1992. The next major step in integration was the Treaty on European Union, agreed in Maastricht in December 1991, when member states agreed to create a single currency to replace national currencies. After a period of conversion, when participating members were to meet strict 'convergence criteria' such as low inflation and the reduction of budget deficits, the Euro was introduced in January 2002. However, it remains con troversial and only 12 of the 15 EU countries (of the time) signed up to it, with the UK, Denmark, and Sweden retaining their own national currencies. The advocates of European integration and the Single Market argued that it would benefit all regions of the European Union, by promoting greater economic efficiencies through increased competition, which in turn would raise productivity and reduce costs and prices, thereby fueling economic growth. However, the reality, as demonstrated later, has been somewhat different.

The EU is unique among REBs for also having political and social institutions attached to it with its own parliament, bureaucracy (the European Commission), and executive council (Council of Ministers). The Council of Ministers is the main decision making body, representing the member states. The heads of government and foreign ministers meet at least twice a year in the European Council, a high profile summit that attracts widespread media coverage. The main roles of the Commission (25 members appointed by the council) are to initiate legislation and proposals, to implement European legislation, budgets, and programs and to represent the Union on the international stage. The Parliament, consisting of directly elected Members of European Parliament (MEPs), has powers over legislation and the budget which are shared with the Council. It also exercises democratic supervision over other EU institutions, particularly the Commission. The EU operates on the basis of agreement and negotiation between independent states, although the member states have delegated some of their functions to the central institutions. As such, states have certainly ceded some powers to the EU as a supranational organization, but this has been a bottom up rather than a top down process. Rather than the superstate that some critics have projected, member states continue to shape how the powers of the EU are exercised and implemented, often seeking to further their national interests within the supranational space of the EU. This is evident even in areas like trade where the EU is the decision making body. While the EU is represented as one body within the current Doha round of the World Trade Organization (WTO) trade negotiations, different member states have adopted different stances. Britain's liberal free trade views, for example, are countered by the more cautious position of countries like France.

Such political realities reflect the difficult tensions thrown up by a process of regional integration that now involves 27 different states as opposed to the original six. A recent attempt to deal with the problems associated with an enlarged EU involved the establishment of a new EU Constitution to formalize the rules for economic governance and agree a system of voting that would avoid growing stasis, although this received a setback in 2005 when referendums held in the Netherlands and France rejected it (see below). In France, the 'no' vote represented a broad alliance of farmers, environmentalists, and radical trade unionists concerned that the new Constitution would provide businesses with more freedom to operate across borders without accompanying social, labor, and environmental legislation.

Spatial Tensions Underpinning Processes of Regional Integration

At the root of contemporary forms of regional integration are the forces associated with intensifying global economic competition. In the first place, large corporate and business interests have been important drivers for greater economic integration, arguing for the reduction of trade barriers and restrictions on the mobility of business at both the global and the regional scale. The ability to escape the limits of national economies, to exploit international markets and tap into cheaper and more differentiated sources of labor are key drivers. In this sense, and unlike previous processes of regional integration, contemporary processes have been accompanied by attempts made by states to forge a global system of trade regulation through the setting up of the WTO. The WTO was set up in 1995 to replace the General Agreement on Trade and Tariffs (GATT) of the post World War II era, and like its predecessor is primarily an attempt to agree upon a set of rules for international trade. However, due to the difficulties of finding consensus between the competing of interests of the multiplicity of nation states (especially those of the global north and south), the WTO has thus far been unable to agree upon a common framework. In this context, the forging of regional trading blocks is arguably a more effective way of pursuing market integration, particularly for European and American multinational corporations (through the EU and NAFTA).

There are two contrasting perspectives on regional integration, its benefits and disadvantages. One view, emanating from mainstream economics, is that the reduction of trade barriers will facilitate a more efficient use of resources, allowing countries to specialize in producing the goods or services that they have a relative cost advantage in. Weaker competition will be eliminated and scarce economic resources redeployed to more effective uses. Overall productivity will increase, prices and costs will fall, stimulating greater growth and wealth overall. This theory of comparative advantage was first associated with the economist David Ricardo and his book *The Wealth of Nations*, first published in 1817. This 'classical' theory of international trade was subsequently revived as part of a neoliberal economics in the 1970s, extolling the virtues of free trade, open markets, and the reduction of government intervention in the economy. Neoliberalism has become the dominant influence – though with some variation in its extent – on economic policymaking in all of the world's most advanced nation states. It was particularly influential in underpinning the Single European Market project.

An alternative perspective associated with those working in Marxist and critical political economy traditions is that increased economic integration without social and political constraints is likely to lead to more uneven and differentiated development between coun tries and localities. At the level of the firm, integration will strengthen larger and multi plant organizations capable of taking advantage of economies of scale at the expense of smaller and more localized firms. Similarly, in geographical terms, those states and localities that are more advanced in terms of skilled labor, capital, and technology will reap greater benefits than less developed ones. The mainstream response up until recently was that less developed regions would benefit in the longer term from lower costs of production, encouraging the shift away from high cost regions. However, some mainstream economists, notably the US academic Paul Krugman, have recently acknowledged the tendency for geo graphical equalities to persist due to monopolistic com petition and economies of scale.

Marxist geographers have pointed to the importance of spatial divisions of labor in shaping the economic land scape. This refers to the role played by large multinational corporations in producing new spatial patterns with the basic concept being a geographical separation of the headquarters and higher level administrative and research functions of the corporation from lower level (branch plant) activities. With an expanding market space, there is the potential for firms to site these activities wherever the production costs are lowest. Higher level functions are less mobile, requiring specialized and highly qualified labor and management, as well as the need to be close to the key centers of economic and political power. Within regional trading blocs such as NAFTA and the EU, decentralization and relocation of basic production activities have been apparent in recent years that appear to bare out this an alysis, giving rise to a variant of the spatial division of labor: the regional division of labor. Through NAFTA, there has been a wave of plant closures and job losses in existing industrial regions of Canada and the US and a massive shift to the border region of Mexico where it is estimated that over 600 000 jobs were created between 1994 and 1998.

In Europe, there has been a relocation of activities from west to east in industries as diverse as automobiles, clothing, and electronics. The impact of these changes has hit hardest in particular places because of existing spatial divisions of labor. In the UK for example, the impact of branch plant closures has been felt most in more peripheral regions such as south Wales and the north of England. The other side of the coin is that the new investments in Eastern Europe are highly concen trated in particular areas that have the appropriate labor force and infrastructure for Multinational corporations (MNCs) to be able to operate. In particular, metropolitan regions in the Czech Republic and Hungary have fared particularly well. The result of these changes is a com plex but divergent pattern of economic growth across

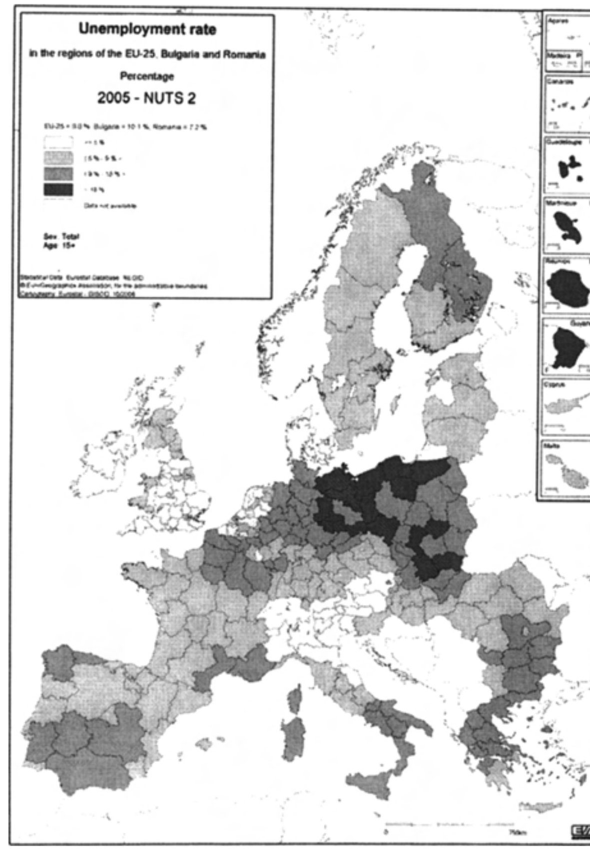

Figure 1 Regional unemployemnt in the EU 2005. Source: Eurostat, LFS.

Europe with little sign of the convergence expected by mainstream economics (**Figure 1**).

Resistance to Regional Integration

Although processes of regional integration have largely been driven by state and business actors, they have not gone uncontested. Trade unions, NGOs, environmental campaigners, farmers and peasants movements have, to varying degrees, protested against forms of market inte gration that are pro business without consideration of the broader social and environmental consequences. Protests against NAFTA and EU integration processes in recent years have been important elements of what is sometimes mistakenly referred to as the anti globalization move ment. This is a misnomer in two ways. (1) Because some of the most effective campaigns have been directed at REBs, notwithstanding the successful mobilization against the WTO meeting in Seattle in 1998. (2) Because the groups that make up the movement have a range of different positions with regard to globalization, with many proposing an alternative form of globalization, built around social justice and environmental regulation, ra ther than one forged according to free trade rules. REBs have become targets because they are perceived as threatening the livelihoods of local farmers, businesses and workers through exposure to unfair predatory competition, particularly from the actions of powerful MNCs.

Apart from the 'no' vote against the EU Constitution, the most celebrated example or resistance to an REB is the Zapatista guerrilla movement – or Ejercito Zapatista de Liberacion Nacional (EZLN) – which emerged from the jungle in the southern Chiapas region of Mexico on New Year's Day, 1994, on the day of implementation of the NAFTA. The specific focus of the Zapatista protest was the protection of the historic rights of indigenous communities in Mexico to communal land, which was threatened by the regime of private property rights and unrestricted trade at the heart of the NAFTA agenda. Indeed, the agreement has accentuated the growth of intensive agricultural production for US markets, displacing more localized farming systems – a process already underway as a result of the Mexican government's own market driven reforms. These developments are leading to the emergence of a small group of wealthy farmers and the dispossession and displacement of many indigenous Indian groups in particular.

The Zapatistas proclaimed an 'autonomous zone' to control their own land and resources. Following an armed insurrection and an uneasy peace with the Mexican government, the Zapatistas had been successful in achieving their own self governing communities, effectively creating a 'state with a state'. One of the reasons behind their success has been their ability to mobilize broader support for their campaign, using the Internet to publicize their struggle to a global audience. They have also been effective in linking their struggle with those of other indigenous and peasant movements across the globe resisting similar processes of neoliberalism and market integration. Their successful ability to scale up their protest suggests the power of opposition groups to develop their own geographical tactics against the spatial integration strategies of states and business interests.

The Future of Regional Integration

As we have seen, there is nothing inevitable about processes of regional integration. Although there are strong pressures to create ever larger and deeper geographical connections through REBs that can facilitate international trade and global economic integration, strong countervailing pressures continue to resist and frustrate such processes. The ambitious attempt to create a 'Free Trade Area of the Americas' effectively extending the NAFTA to the whole of the American continent has, for the present, been defeated by opposition from the developing countries of Latin America who resent continued US protectionism, particularly of its farm sector, and the US inspired neoliberal policies that were blamed for increasing poverty in the 1990s.

In developed countries, national political identities and protectionist concerns about job losses and damage to the livelihoods of their citizens mean that nation states will continue to confront scale dilemmas. Elected politicians cannot ignore the interests of their domestic voters in prosecuting a market integration agenda for business. In the first decade of the twenty first century protectionist currents are evident in the United States, Canada and Europe, countering to some extent the globalization and free trade rhetoric of the 1990s. Concerns are mounting about competition both from within their own REBs and further afield, particularly in relation to the growing economic powers of China and India. One Canadian politician was quoted as viewing NAFTA as a "nightmare of US continentalists come true: Canada's resources, Mexico's labor, and US capital" (cited in Dicken 2007: 199). In addition to such concerns, recent geopolitical instability and worries about security against international terrorism are putting new obstacles in the path of transnational economic integration. Growing conflicts over scarce natural resources might also lead to states retreating from expansionist policies to focus more upon narrower national interests.

Outside North America and Europe the trends toward regional integration vary. In Asia, negotiations are underway for China, Korea, and Japan to enter the ASEAN region. In Africa, trends towards regional integration remain weak, while in Latin America the Mercosur agreement has been effective at increasing internal trade though it lacks the institutional infrastructure of the EU. One interesting development in the 2000s involves the attempts by the radical left government of Venezuelan President Hugo Chavez to create an alternative anti capitalist bloc of countries pursing closer economic and social ties. Benefiting from considerable anti US feeling in Latin America, and with the benefit of his country's vast resources, Chavez has been promoting reciprocal agreements for the exchange of economic resources (e.g., oil for doctors with Cuba) with like minded leaders elsewhere. The election of left wing leaders in the neighboring states of Peru and Ecuador has increased this potential for this type of regional cooperation but it remains unclear as to how sustainable these developments will prove.

For economic geographers, there remains a considerable research agenda around questions of regional integration. Indeed, pressures for regional integration (particularly political and environmental) might increasingly come into conflict with capital's desire for more global dexterity. One set of issues revolves around how changing geographies of production – and in particular more complicated and extensive global networks of production – disrupt current processes of trade and market regulation centerd upon particular macro re gions. Another (countervailing) set of pressures around resource scarcity might actually strengthen regional in tegration as issues such as food and energy security rise up the political agenda, in response to climate change and oil depletion. Additionally, as fuel prices rise, more traditional geographical location factors such as transport costs may result in more regionalized and even localized forms of commodity production.

See also: Labor Markets, Regional; Region; Regional Integration.

Further Reading

Agnew, J. (1997). Regions on the mind do not equal regions of the mind. *Progress in Human Geography* 23, 91 96.

Coe, N., Hess, M., Yeung, H. W., Dicken, P. and Henderson, J. (2004). Globalising regional development: A global production networks perspective. *Transactions of the Institute of British Geographers* 29, 464 484.

Dicken, P. (2007). *Global Shift: Mapping the Changing Contours of the World Economy*, ch. 6, pp 187 204. London: Sage.

Dieter, H. and Higgott, R. (2003). Exploring alternative theories of economic regionalism: From trade to finance in Asian cooperation. *Review of International Political Economy* 10, 430 454.

Dunford, M. and Smith, A. (2000). Catching up or falling behind? Economic performance and regional trajectories in the 'New Europe'. *Economic Geography* 76, 169 195.

Harvey, D. (2005). *A Brief History of Neoliberalism*. Oxford: Oxford University Press.

MacKinnon, D. and Cumbers, A. (2007). *An Introduction to Economic Geography: Globalization, Uneven Development and Place*, chs. 5 7. London: Prentice Hall.

Massey, D. (1995). *Spatial Divisions of Labour* (2nd edn.). Basingstoke: Macmillan.

Ohmae, K. (1985). *Triad Power: The Coming Shape of Global Competition*. New York: Free Press.

Poon, J. P. H., Thompson, E. R. and Kelly, P. F. (2000). Myth of the triad? The geography of trade and investment 'blocs'. *Transactions of the Institute of British Geographers NS* 25, 427 444.

Schiff, M. and Winters, L. (2003). *Regional Integration and Development*. Washington, DC: World Bank.

Ziltener, P. (2004). The economic effects of the European Single Market Project: Projections, simulations And the reality. *Review of International Political Economy* 11, 953 979.

Relevant Websites

http://www.attac.org
ATTAC social movement campaigning for fair trade and social justice within and outside European Union.

http://europa.eu
European Union.

Regional Planning and Development Theories

E. W. Soja, University of California – Los Angeles, Los Angeles, CA, USA, and London School of Economics and Political Science, London, UK

Glossary

Entrepreneurial Regionalism Neoliberal approach to regional planning dominated by intense competition to attract global investments and tourism.

Localization Economies Cost savings for firms due to proximity of input sources, consumer markets, and labor supplies.

Neoliberalism Economic philosophy based on free markets, deregulation, and privatization.

New Regionalism The recent resurgence of regionalism as a political, economic, and cultural force and as the basis for a re-theorization of regions and regional development.

Regionalism Regional advocacy, an actively practiced belief that regions are useful for a wide variety of objectives, including efficient and equitable development.

Regionality The general condition of having regional qualities.

Regionalization The process of forming or identifying regions.

Underdevelopment Theory A critique that divides the development process into a dominant form for core capitalist countries and a dependent form for peripheral developing countries, also called dependency theory.

Urbanization Economies Advantages for firms that arise from the innovations and creativity associated with dense and heterogeneous urban agglomerations.

Welfare Regionalism Regional planning aimed at reducing regional inequalities within countries.

For at least the past 50 years, the theory and practice of regional planning has had an especially close and productive relationship with human geography and geographers. More so than other branches of planning, regional planning and development theory has been a primary focus for applied geographical concepts and perspectives. Part of this special connection reflects the traditional role of regional geography in all forms of geographical analysis. Regions provide a means of bringing together the multifaceted features of human geography into a coherent whole and have always been an important part of the geographer's conceptual toolkit.

The connections between human geography and regional planning, however, go well beyond sharing a regional perspective. Regions and regional thinking have played a central role in the development of geographic thought for more than a century. Regional perspectives were promoted widely in early German, Anglophone, Spanish, and Scandinavian geography, and French geographers such as Vidal de la Blache and the anarchist Elysée Reclus were major influences on the early twentieth century development of regional planning concepts in North America. The first entry for regional planning in the *Encyclopedia Britannica* in 1929, written by Benton MacKaye and Lewis Mumford, speaks of the 'great school of regional geographers' as one of the precursors in the formation of a regional planning profession. Geographers mentioned in this article include Vidal, Reclus, Ratzel, Herbertson, Fenneman, Smith, and Shaler.

Mid century geography, led by such influential figures as Carl Sauer and Richard Hartshorne, paid special attention to regions and regionalization, the method of constructing regions through geographical interpretation and synthesis. There was little direct influence of these ideas in regional planning, but much closer links were established after World War II. The rise of positivist or scientific geography in the 1950s was directly tied to the emergence of regional science as a new hybrid field aimed specifically at providing theories and concepts of use in urban and regional planning. Regional science sought to integrate all the social sciences around a regional spatial perspective, but its primary academic linkages were most closely tied and occasionally coincident with the subfield of economic geography.

By 1990, the heyday of both positivist geography and regional science was over, as a new critical human geography took shape from a mixture of neo Marxist, feminist, postmodern, and phenomenological perspectives. At first, regional approaches were relatively neglected, but the further development of critical human geography would later spur a resurgence of interest in regions and regionalism, especially as a means of integrating macro and micro geographical analysis. A subfield of regional political economy emerged, drawing together established but rarely connected approaches to analyzing the political economy of urbanization on the one hand, and the political economy of international development on the other. More recently, what is called a new regionalism has emerged focused on an integrative meso geographical perspective that ties together the global and the local, the macro and micro, exogenous and endogenous development processes, at a multiplicity of regional scales.

Regional planning, after 20 years of decline under the entrepreneurial policies of the neoliberal state, is currently experiencing a significant reconfiguration and renewal based on an infusion of innovative ideas about cities and regions as generative forces shaping all aspects of societal development. Critical human geographers and geopolitical economists, as well as a revived regional science, are playing a leading role in advancing this new regionalism. As has been true throughout the history of regional development theory and planning, these recent advances in regional thinking are attracting both positive attention and intensive critical evaluation. Before moving to a more detailed look at the evolution of this special relationship between human geography, regional planning, and regional development theories, some definitions are needed.

Defining Regions and Regionalism

Regions can be seen as empirically existing phenomena as well as products of a regionalization process imposed on the empirical world by observers, analysts, and administrators. In their simplest form, regions are definable areas with certain distinct characteristics. Three types of regions have been identified in the geographical literature. Formal or uniform regions are internally homogeneous with respect to a particular attribute. The Corn Belt, the shrinking Antarctic ice cap, or any governmental unit from the municipality to the nation state, are formal regions. Every point within their boundaries shares the same attributes, according to the defining criteria.

Functional regions are defined by their organizational structure. They are also called nodal regions because they usually revolve around a focal center, such as the downtown core of an urban region or the local service area surrounding a library or post office or shopping mall. Functional regions are behaviorally based activity areas and although they too can be defined by outside observers, they are also active parts of everyday life in the empirical world. Indeed, we all can be said to live within a multiscalar world of nodal regions ranging from the mobile human body, with its surrounding bubbles of personal space, to relatively fixed spaces of the household and home, the neighborhood and community, the city and metropolitan area, the district or province or state, the national state, and various supranational regions. At each of these scales, the nodal region has a distinctive core–periphery structure based largely on the spatial advantages of centrality or nodality.

The distinction between uniform and nodal regions as empirically existing phenomena is quite blurry, especially with respect to questions of governance and identity. France, for example, is both a legally defined uniform region within its set boundaries and a nodal region or set of regions based on its internal (and external) networks and nodal flows of people, goods, information, financial assets, etc. Is French national identity associated with the specific territory that we call France or with the active behavioral space associated with the French people? Political geographers have long argued that the ideal form of government at every scale should have the formally bounded region and the functionally organized region coincide, but rarely is this the case. International and intercultural conflicts often arise from this lack of regional congruence.

Programming or planning regions are a third type, defined explicitly as created for a particular purpose, literally *ad hoc*. The multistate area created for the Tennessee Valley Authority (TVA) in the 1930s is an example of a designated planning region. Planning regions can also be described according to their formal and functional characteristics, and though artificially imposed can develop more lasting effects on regional governance and identity.

All regions to some extent are the result of an imposed regionalization, whether created by outside observers or socially constructed from within. In either case, regions of many different scales, scope, and sensibility are part of the very nature of human life. The term regionality can be used to describe this general quality much like modernity refers to the general quality of being modern. Regionalization is the process of forming regions and regionalism the active promotion of regional perspectives, similar to modernization and modernism. Regional planners, regional geographers, and regional scientists are to varying degrees regionalists.

Regionalism in this broad sense is a form of advocacy, an actively practiced belief that regions are useful concepts for achieving a wide range of objectives. These objectives can include achieving greater theoretical insight and understanding, inducing more rapid and equitable economic development, improving administrative efficiency, enhancing political democracy and representation, fostering cultural identity, preserving the natural environment, and stimulating creativity and innovation. As a form of advocacy and collective action, regionalism is intrinsically political. It typically asserts the usefulness of regional ideas, organizations, and identities against established modes of thought and practice. Regionalism is thus directly tied to questions of governance, and especially to the territorial or spatial dimension of government, administration, regulation, social control, and the shaping of the built and natural environment.

Emphasizing this link to territorial governance is the Latin root *regere*, meaning to rule or govern over some defined domain. Hence, such associated terms as regime, regal, regent, and regulate. The actual politics of regionalism can be defined as primarily cultural or economic, intensely active or relatively weak, objectively real or subjectively imagined. As a form of advocacy and collective action, regionalism can occur at many different scales, from the global to the local, although regional planning throughout its history has tended to focus pri marily on subnational regions, such as Appalachia, Catalonia, and Quebec.

Regional planning and policymaking is a particular form of regionalism, a distinctive practice that uses re gions and regional thinking to achieve specified goals. Like all forms of regionalism, regional planning has a particular history and geography, that is, it develops to different degrees in different places at different times. Regional planning as an institutionalized activity and profession has been practiced for less than a century, but has evolved through series of distinct phases and forms that have earlier precedents. These phases in the evo lution of regional planning will be discussed here with particular attention given to co evolving theories of un even regional development, theories that attempt to ex plain why some regions develop faster than others.

Regional Planning as Resource Development: 1920–50

Regionalism emerged as a significant political force in Europe and North America in the second half of the nineteenth century, primarily in conjunction with the consolidation of national markets and the spread of na tionalist ideologies aimed at sustaining rapid industrial ization. In the US, regional divisions between North and South exploded in a bloody Civil War, while less violent regional uprisings multiplied throughout Europe as new nation states arose in Germany and Italy and older states such as France, Great Britain, and Spain attempted to integrate diverse regional cultures into unified national territories. Many of the cultural regionalisms that were politically asserted during this period, including those of Catalonia, Scotland, the Basque country, and the US South, have continued to the present as important pol itical, economic, and cultural forces.

Regional planning as a distinct form of public inter vention, however, first consolidated in the US in the 1920s during a period known as the Progressive Era, when many new experiments in urban government, public administration, and planning were initiated. This first phase drew upon European traditions of utopian socialism, anarchism, regional geography and sociology, and the regional political movements that arose in response to the homogenization of national markets and the deepening urban poverty that characterized late nineteenth century Europe. Particularly influential were such anarchist thinkers as Pierre Joseph Proudhon and Peter Kropotkin, and two leading planning innovators, the polymath Scotsman Patrick Geddes and Ebenezer Howard, founder of the garden city movement.

This first regional planning doctrine or paradigm viewed the large industrial capitalist city, with its teeming densities and unhealthy slum housing, as the primary cause of major social and environmental problems. It emphasized urban decentralization, typically into quasi socialist new towns and garden cities that would combine the advantages of both the city and the countryside, while hopefully ameli orating the problems of each through some form of com mon or public ownership. Although the European critique of urban industrial capitalism and the particular form it was taking in the industrial capitalist city significantly influ enced this first phase of regional planning, its more radical European regionalism was watered down in its American ization. What remained, however, were a strong anti urban bias, an idealized vision of the countryside and the en vironment, and a specifically spatial strategy of reform.

The appropriate planning region tended to be defined in almost organic or ecological terms, building on the social cohesion that was presumed to be coincident with the physical environment. The work of regional physical geographers was of particular importance here. The most innovative and exemplary expression of this first round of regional planning was the TVA, a semiautonomous multistate unit of government responsible for resource development and environmental preservation in the large Tennessee River drainage basin.

The TVA model of regional resource development, which would later be replicated in India, Mexico, the Soviet Union, and many other countries, linked together the two main branches of early regionalist thought in the US, one centered in New York in the Regional Planning Association of America, with Lewis Mumford its best known figure, and the other representing the southern states of the US, giving voice to problems of regional underdevelopment and, to a lesser extent, racial dis crimination. Metropolitan regional planning was also emerging at the same time in the US, but mostly as an adjunct to urban planning and often in antagonism with the resource based and larger scale ecological approach of the northern and southern regionalists.

This first phase, especially through the TVA, would have a significant influence on US national policy in the 1920s and early 1930s, but would lose nearly all its in fluence as the Great Depression deepened and national priorities shifted in response to the threat and reality of World War II. Once a grand utopian experiment, the TVA became little more than a munitions factory and a generator of cheap electricity for northern industries. By the time World War II ended, regional planning had al most disappeared in the US, although many other countries continued to create specialized agencies and initiate river basin hydroelectric and resource develop ment schemes similar to the TVA for decades afterward.

There was very little theorizing about uneven regional development during this first phase. There was some concern for urban poverty and a widespread critique of the overconcentration of both wealth and poverty in the booming industrial capitalist cities, but the regional discourse rarely addressed directly the rootedness of these problems in the capitalist development process. During this period, the discipline of geography for the most part was theoretically moribund and increasingly isolated from developments in the social sciences as well as in professional planning practice. After the swift expansion of regional planning and regionalist thinking in the first quarter of the twentieth century, peaking in the formation of the TVA, an equally rapid decline took place in the second quarter century.

Regional Planning 1950–80: Welfare Regionalism

A very different form of regional planning began to develop in the 1950s, in conjunction with the enormous growth in all the industrialized countries of national economic planning, the welfare state, and Cold War competition. Stimulated by the development of regional science, literally a science of regions that was absent in the first paradigm, and the related rise of a more decidedly scientific human geography, this second phase defined itself around the bold agenda of being able to maintain if not enhance administrative efficiency and national economic growth while at the same time dealing effectively with problems of uneven development and economic inequality through what was broadly defined as spatial systems planning.

This new mode of welfare regionalism operated primarily at the level of the national state and was bolstered by the expansion of national accounting systems and data collection on income levels as measured, for example, by such economic indicators as per capita gross national product (GNP). Ambitious but rarely well funded programs were initiated in almost every country in the world to deal with uneven regional development, either through special funding for development in 'lagging' regions or through comprehensive spatial planning for the entire national settlement system, from the smallest to the largest cities. In most cases, this involved some form of growth pole or growth center planning, described as a strategy of concentrated deconcentration of economic development aimed at a spreading of wealth to all regions of the national state.

It was argued by the new regional scientists and planners, including many in the socialist countries, that such welfare oriented regional policies were necessary to make all forms of post war economic development more equitable and democratic. Regional inequalities of income would not simply disappear over time, as neoclassical economic theory predicted. Some regional economic thinkers, such as Nobel Prize winner Gunnar Myrdal, referring both to the gap between First World and Third World countries as well as to pronounced regional inequalities within countries, such as between the north and south in Italy and in the US, argued that the regional problem would intensify over time without government intervention. Giving this new mode of regional planning an additional urgency was the Cold War competition between capitalist and communist approaches to development, a competition which also fostered the growth of the welfare state in Europe and, in a more diluted form, in North America.

New relations also developed between top down state centered regional welfare planning and metropolitan and local government, as regional planning expanded its scope to address the entire national space economy and its hierarchical network of urban settlements, ranging from the largest metropolis to the smallest hamlet. In many industrialized countries, strong metropolitan regional governments increased in number, aimed at dealing with the social and economic problems of large urban agglomerations and the sprawling and politically fragmented suburban areas surrounding them. In more peripheral countries, new institutions arose to promote the equitable spread of modernization throughout the entire settlement system, usually as part of a nation building process.

Emerging out of this period were the first broad outlines of a theory of uneven regional development, an attempt to explain the production and reproduction of regional, or more broadly spatial, inequalities within and between nation states. Such a theory was essential to buttress the extraordinary claim of regionalists and regional planners that they could achieve both efficiency and equity, something which economists and national economic planners generally thought impossible, at least under capitalism. In the period 1950–80, an extensive literature was produced to sustain this ambitious claim and promote the necessity for regional planning in every national economy. Although there was an unusual clutter of repetitious naming and re elaboration, and not a little confusion, a series of basic ideas were formulated about the causes of uneven regional development and the appropriate methods of regional planning that continue to be relevant today.

The starting point for a geographical theory of uneven development was simple enough, although at the time it was almost entirely unrecognized: development never takes place uniformly over space. Like all social processes, development, however it is defined, is always geographically uneven. Although rarely described as such, geographically uneven development (GUD) became the focal point for regional analysis, theory, and planning. The next step was to identify the major forces shaping GUD. Here the work of Francois Perroux was most influential. He argued from extensive empirical and historical analysis that uneven development arises primarily from geographical concentration, from dynamic urban agglomerations that he called *pôles de croissance*, variously translated as development poles, growth poles, or growth centers. Spatial agglomeration or polarization was therefore the key to GUD.

Theorizing further, Perroux and other key figures, such as Gunnar Myrdal and Albert Hirschmann, argued that development diffuses over space from these poles of growth in positive as well as negative waves. Myrdal called the positive effects spread and the negative backwash. Hirschmann used the terms trickling down and polarization, and others would add their own names for essentially the same processes. These concepts were developed in the 1950s and early 1960s and although they generated a huge ancillary literature there was very little advance made in deepening our understanding of the underlying dynamics of polarized development until more recent breakthroughs in geographical economics and the study of urbanization and agglomeration economies, a vital part of the currently developing new regionalism.

The basic argument seems clear in retrospect. GUD and the problem of regional inequalities are produced through positive and negative impulses coming from urban industrial centers. Myrdal would add to this his concept of circular and cumulative causation, arguing that once a development pole is spontaneously established, its initial locational and other advantages tend to accumulate over time. Similarly, if backwash effects are greater than spread effects, or the other way around, development would tend to move further in the same direction, leading either to the reduction of regional inequalities (convergence) or their intensification (divergence).

Myrdal as well as other pioneers in the new regional development theory did not just refer to uneven regional development within nation states but also to internationally uneven development, that is, to the differences between First, Second, and Third worlds. In both the national and international arenas, there was another dramatic implication. If major inequalities and deepening poverty were ignored, there was a strong chance that massive social unrest and rebellion would ensue. Although often ignored in practice, this darker, more threatening side of uneven regional development added greater urgency and necessity to regional planning intervention.

The logic of regional planning was straightforward. To promote efficient and equitable development, regional planners must facilitate positive spread effects and control negative backwash or polarization effects. How best to do this was more difficult to agree upon. Some control over the generative effects of existing urban agglomerations was obviously needed, especially where there was a heavy concentration of development in expansive primate cities. But controlling the negative effects of urban industrialization might also curb the most powerful positive developmental forces. Rather than actively intervening to limit the continued growth of the largest metropolitan areas, the alternative most often chosen was to identify the poorest regions and to try to stimulate new development there through a wide variety of methods, usually involving the designation of existing or proposed urban settlements as growth poles and growth centers. This was done in two related and often overlapping approaches.

Territorially based regional welfare planning specifically targeted the poorer regions of a country. Many different policies were formulated to attract and keep investment, to reduce poverty and unemployment, at times for the region as a whole but most often focused on selected growth centers. In the UK, the early garden city new towns, part of a strategy to siphon off population from overcrowded London, became in Mark II new towns part of a regional development strategy in such problem regions as northeast England, south Wales, and lowland Scotland. The Mezzogiorno, the poor south of Italy, became the target for a wide variety of development policies led by one of the earliest national agencies specifically charged with fostering regional development, the *Cassa per il Mezzogiorno*.

In the US, where regional planning of any kind was not usually welcomed by the federal government, an exception was made during the Kennedy presidency to create the Appalachian Regional Commission, like the TVA a multistate organization that covered mainly the South but also edged into the North as well. The Appalachian Commission never had the powers that were initially invested in the TVA, and the series of regional commissions that were created after Appalachia, as every region of the country began to demand equal attention, were even weaker and were soon closed down. Regional welfare planning probably had less effect in the US than it had in almost every other developed country.

A more ambitious form of regional planning addressed the entire national space economy and settlement system. It arose from the idea that reducing regional inequalities and spreading development more evenly throughout the country required a form of master planning that controlled or at least guided all locational decision making, public and private. Spatial systems planning typically involved designating a multilayered hierarchy of growth centers that ranged from the largest to the smallest urban settlements. This effort drew heavily upon the location theories developing in scientific human geography, especially the central place theory associated with

Christaller and Losch. The proposed hierarchy of growth centers, including established urban agglomerations as well as newly created ones, was to receive special attention in private and public investment. The aim was to streamline and integrate the whole national space into an efficient and cohesive system that would diffuse positive developmental impulses and lead to increased interregional equity.

Spatial systems planning was applied most often in developing countries, often in conjunction with nation building strategies, although something quite similar was tried in France under the direction of the *Delegation á l'Aménagement du Territoire et l'Action Regionale* (DATAR), the delegation or institute for territorial management and regional action, one of the most influential regional planning bodies in the world and still in existence today. French regional planners, building on Perroux's ideas, were particularly innovative in producing sophisticated methods to select the most propulsive industries, those with the greatest forward and backward linkages in the national economy, to locate in designated development poles. In most countries, however, especially in the Third World, industrial strategies involving national guidance over private capital investment were rarely applied. Even attempts to shape the location of public facilities investment, such as for post offices and schools, were difficult to implement successfully.

Given its need for powerful centralized control over locational decision making, spatial systems planning was never very successful in capitalist countries, whether in the First or Third worlds. It had more success in socialist command economies, although even here national security and military priorities often distorted attempts at equitable regional planning. Perhaps the greatest success story during this period was in Castro's Cuba, where urban and rural development strategies were combined to both control (often harshly) the growth of Havana and spread development, especially with regard to health and education, throughout the regions. The Cuban success story would influence politics and planning to this day in Latin America and to a lesser extent in Africa.

Regional Planning as Entrepreneurial Regionalism 1980–present

Triggered initially by the urban uprisings that took place all over the world in the 1960s and the worldwide economic recession of the early 1970s, the second model of regional planning began to decline significantly in importance as national attention shifted to other, more urgent, priorities. As occurred with the first model, regional planning flourished during a period of economic expansion, but once the boom ended there was a significant retrenchment as improvement in regional welfare (equity) became subordinated to the necessity for restoring expansive national economic growth (efficiency).

There were other factors leading to the decline of welfare regionalism and regional planning more generally. It had become increasingly clear that regional planning was having relatively little effect on reducing regional inequalities. Growth pole dynamics were very poorly understood and their use often led to a worsening of regional conditions. Even the most ardent regional planning advocates began to realize that spatial systems approaches, at least in capitalist countries, demanded a degree of control over decision making in both the private and the public sectors that was highly unlikely ever to be attained. The rather desperate argument that polarization reversal, and hence the success of regional planning programs, necessarily takes a long period of dedicated implementation was not easily accepted at a time of economic recession and political instability.

In addition, powerful radical critiques emerged that undermined the broader conceptual and ideological foundations of regional development theories and regional planning practice. The most important of these radical critiques, promoted especially forcefully by Latin American scholars, argued that an ideological and theoretical bias infiltrated the modernization theories and idealistic developmentalism that underpinned all efforts to promote capitalist development everywhere in the world. In what would be described as underdevelopment and/or dependency theory, the development process was no longer seen as a singular and linear process moving through sequential stages of growth but rather as a two sided dynamic that structurally divided the world into a developed and dominant capitalist core and an underdeveloped and dependent periphery, with each experiencing different development processes.

Underdevelopment was seen not just as a descriptive attribute of regions and countries but as an active process, as in the notion that dominant regions and states actively underdevelop other regions and states to promote their own economic expansion. This introduced a very different interpretation to one of the most influential models shaping regional planning policies, the inverted U shaped curve that linked the intensity of regional inequalities to levels of development. The model argued that inequalities were low at the earliest stages of development, grew to a maximum in the mid range (often exemplified by Brazil), and then declined as higher levels of capitalist development were reached. Instead of there being a single development process followed through sequential stages by all countries and regions, the radical theorists argued there were two different worlds of development, one for the advanced core countries tending at least somewhat toward convergence, another for the periphery, where inequalities intensified over time. In this alternative model, Brazil was much more underdeveloped than Nepal.

Theories of underdevelopment and dependency, un equal exchange in international trade, world systems, and other core–periphery approaches to understanding inter national as well as intranational economic development and modernization, made it increasingly difficult to ac cept the premises and promises not just of regional planning but of all capitalist development theory. A feeling of having been misled spread widely among committed development planners, policymakers, and scholars, especially in the Third World, as another reason was found for why regional development policies fail, even with the best of intentions and the most sophisti cated planning practices. The relative success of the Cuban model became more cogent than ever before.

At the same time as welfare regionalism came under attack from the left and liberal center, the very foun dations of the welfare state and its development policies came under challenge from more conservative forces, especially in the US and Great Britain. Regional welfare planning was being hit from all sides. It was seen either as a supportive tool to maintain the control and domination of advanced core regions over the underdeveloped and dependent peripheries (the radical critique), or as a wasteful luxury that created more problems than it re solved, especially at a time of reduced national economic growth and rising social unrest (the conservative cri tique). Where regional planning programs survived, in First, Second, or Third worlds, their aims tended to shift almost entirely from reducing inequalities to fostering more efficient economic growth.

By the late 1980s, a more entrepreneurial regionalism had consolidated in response to economic recession and the impact of the increasing globalization of capital in vestment. The national state, feeling some threats to its long standing economic power and political sovereignty, focused on crisis management and state restructuring in an attempt to adapt to changing world conditions. Led by the US and the UK under Reagan and Thatcher, many countries adopted neoliberal policies of deregulation, privatization, and strategic decentralization aimed at re empowering market forces at the expense of national welfare planning. With national funding virtually cut off, local initiatives to attract the attention of the private sector and global tourism came to dominate the activities of urban and regional planners.

A new context for regional planning emerged, shaped by complex processes of state restructuring that involved both decentralization (devolution) and recentralization of territorial authority and governing power. This had the result of transforming integrated, comprehensive, and nationally organized regional welfare planning into a more intensely competitive and geographically frag mented entrepreneurial regionalism in which local and regional governments competed fiercely with one an other for scarce resources and corporate investments.

At the national level, neoliberal trade policies and increasing privatization and deregulation were promoted to free the market from its Keynesian welfare restraints and Fordist inflexibility. This had the effect of weakening most existing subnational and metropolitan regional au thorities and planning institutions as they were func tionally reduced to serving the new national economic and political strategies. In contrast, city and municipal governments have often expanded their role in pro moting local economic development, especially through place marketing strategies, image oriented city booster ism, and the creation of deregulated enterprise or free trade zones to attract capital investment. Backed by various public private partnerships and special public subsidies, these localized strategies became the most important policy tools shaping regional and metropolitan growth.

It can be argued that more progressive forms of re gional planning and governance have been co opted by neoliberal market ideologues to serve the emerging new economy of flexible, post Fordist, and global capitalism. In this process, the fundamental character of regionalism as an assertion of the importance and power of regions and regional thinking changed significantly. Indeed, entrepreneurial regionalism is not so much a form of regionalism as it is an aggressive, competitive, and pa rochial localism. Regions become little more than a col lection of isolated territorial containers fighting each other for access and positioning within the growing global economy.

By the early 1990s, in most countries, earlier forms of liberal and progressive regionalism had either dis appeared or been replaced by this neoliberal and entre preneurial adaptation. Urban and regional planning became deeply absorbed into what some called territorial wars for jobs, investment, tourism, and global image making. As regional and other forms of inequality and injustice soared to near record levels in many parts of the world, welfare policies were cruelly weakened, as plan ners turned to seeking the next Olympic Games or global trade expos or iconic architectural extravaganzas such as the Frank Gehry designed Guggenheim Museum in Bilbao.

The Rise of a New Regionalism

As welfare regionalism declined and neoliberal entre preneurial planning strategies rose to prominence, re gionalism experienced an unexpected resurgence both in the material world of geopolitics and in the realm of regional development theory. This new regionalism, as it has come to be called, has only begun to affect regional planning practices and the theoretical developments are still in their infancy. There can be little doubt, however, that regions, regionalism, and regional thinking have reentered the political, economic, cultural, and academic arenas with unusual intensity nearly everywhere over the past 10 years. Entering 'new regionalism' on your Internet search engine today will result in an astonishing number and diversity of hits, spanning every geographical scale from the global to the local.

The renewal and resurgence of geopolitical regionalism can be linked to the same crisis generated restructuring processes that led to the decline of the welfare state and the rise of neoliberalism. These processes include the formation of a more flexible, global, and information intensive capitalist mode of development, now often described as a new economy; the accelerated and far reaching globalization of capital, labor, and consumer driven urban industrial culture; and, sustaining both economic restructuring and neoliberal globalization, a revolution in information and communications technologies.

Reminiscent of the many insurgent regionalisms that accompanied national integration movements in late nineteenth century Europe and North America, the past three decades have been marked by widespread irruptions of cultural and political regionalism. Examples include the regional disintegration of Yugoslavia, the Soviet Union, and Czechoslovakia, innovative forms of regional devolution in Spain (Catalonia and the Basque country), the United Kingdom (Scotland, Wales, the Greater London Authority), and Canada (Quebec), and secessionist movements of varying intensity and success in Eritrea, southern Sudan, northern Italy, Sri Lanka, Tibet, Kashmir, Chiapas, Kurdistan, Chechnya, and Aceh.

Rather than being triggered by the integration of national markets and the spreading imprint of nationalist ideologies, as happened in the nineteenth century, this latest round of regional uprisings and assertive regional political movements has been stimulated by the increasing integration of the global economy, the demands of a more flexible and expansive capitalism, the technology driven diffusion of a global consumerist culture, and the associated spread of neoliberal policies and ideologies, all of which have contributed to an intensification of economic inequalities and a widening of the gap between the rich and the poor. Stimulating regionalism further have been significant changes in the territorial authority and sovereignty of the nation state, what some have called state restructuring.

The result has been a major reconfiguration of uneven regional development at every geographical scale, from the global to the local. At the global scale, some scholars and business advisors have begun to speak of an increasingly borderless world in which once rigid national boundaries have become more open to transnational flows of migrants, trade, investment, information, and cultural influences. These intensified flows have led to the weakening of many older territorial attachments and nationalist identities, and contributed to a reshuffling of the international or global division of labor. What were once clearly identifiable as the First, Second, and Third Worlds are now much more difficult to define. Not only has the socialist Second World become radically deconstructed, the composition of the First and Third Worlds has changed significantly, especially with the emergence of a growing number of newly industrialized countries or NICs. A global North–South divide still persists, but it is very different from what it used to be.

The global restructuring of uneven development has been additionally shaped by three intertwining trends: an increasing cultural and economic homogenization (often described as Americanization) making the nations and regions of the world more alike in many ways; a countervailing process of increasing differentiation creating growing dissimilarities, divisions, and diversity; and a complex reconfiguration of global hierarchies and the uneven distribution of economic wealth and political power. There has probably been a greater change in the global geography of industrialization and economic development over the past 20 years than in the previous two centuries. Whether this has led to an overall increase or decrease in global inequalities is difficult to determine, but it is clear that these changes are challenging almost all existing theories of GUD, especially those based on relatively stable core–periphery structures.

Between the global and national scales has been another arena of intensified regional economic and political activity, exemplified best by the supranational regionalism associated with the formation of the European Union (EU). The EU, described by some as a Europe of regions, has been among the most important forums for preservation and hoped for expansion of progressive forms of regional planning. The EU Regional Fund has been shaping regional and national development for many years. More recently, several additional policies and programs have been adopted, such as INTERREG promoting transborder regional development and identity, and the evocatively named European Spatial Development Perspective (ESDP), aimed at reducing interregional socioeconomic inequalities and promoting related forms of sustainable urban and regional development. The ESDP has not yet had much tangible effect, but it provides an important channel for reopening and rethinking the debates on welfare regionalism, social exclusion, spatial planning, and theories of GUD.

Supranational regionalism has also manifested itself in the multiplication of multinational trading blocs, such as North American Free Trade Agreement (NAFTA) in North America, MERCOSUR in the southern cone of South America, and Asia–Pacific Economic Cooperation

(APEC) in eastern Asia. Although focused functionally on facilitating cross border trade, some regional blocs are adding new activities such as environmental planning and improving working conditions. The world map is be coming increasingly filled with these regional trading blocs or, as they are sometimes called, economic com munities, after the model of the earlier European Eco nomic Community that became the EU. Most continue to be dominated by neoliberal forces promoting an ideology of free trade and equal exchange, but there is some hope that more ambitious and progressive global coalitions may develop in the future to regulate the negative effects of neoliberal globalization.

In addition to state restructuring at the national scale, with its mixture of decentralization and recentralization, there have been dramatic regional changes taking place within nation states. In many advanced industrial coun tries, the regions that were the strongest engines behind earlier periods of economic growth, such as the American Manufacturing Belt and northeast England, experienced a significant and unexpected de industrialization and de cline after 1960, seemingly in contradiction with simplified theories of cumulative causation. It was this early period of de industrialization that led many scholars to embrace the notion of an emerging postindustrial society. De indus trialization, however, was occurring at the same time as an expansive re industrialization, the onset of a new and different period of urban industrialism that would enable some older industrial regions to recover and expand while also opening up new subnational industrial regions (NIRs?) to match the NICs at the global scale.

Three regional shifts have attracted particular atten tion. One was the pronounced power shift between the so called Frostbelt (or Rustbelt) and the Sunbelt in the US. De industrialization left behind devastated com munities and cities in Ohio, Pennsylvania, and Michigan while both a New South and a burgeoning Pacific Rim led the way to a new national geography of industrial ization. Economic recovery in many of the cities of the northeast as well as economic expansion in the Sunbelt states was spurred by another unexpected and para doxical development, the urbanization and industrial ization of parts of suburbia. The most iconic place here was Silicon Valley, perhaps the best known and studied example of a generative regional economy emerging in a formerly suburban landscape and sustained by innovative forms of regional planning. The Silicon Valley model, however, has been very difficult to replicate elsewhere, despite many attempts to do so.

Also receiving exemplary attention was the Third Italy, nestled between the northern region focused on the ad vanced industrial triangle of Milan–Turin–Genoa and the classic case of the underdeveloped South, the Mezzo giorno. While decades of regional planning aimed (un successfully) at reducing the North–South development divide, the in between regions of Emilia Romagna, Tus cany, and the Veneto forged ahead in a new kind of in dustrialization process, concentrated not in large cities as much as in smaller, flexibly specialized, urban clusters in the verdant countryside. The Third Italy became one of the leading innovative regions of Europe, although its leadership has been reduced in recent years, especially with northern centers such as Milan growing rapidly again.

Receiving great attention today has been the extra ordinary regional industrialization of China. China has engaged in an innovative stepwise model of regional development planning that has resulted in the most rapid urbanization led industrial expansion ever experienced by any country. Attention was concentrated first on the coastal region and the capital city, the most attractive areas for global capital. While development exploded around the three major poles of Shanghai, Beijing, and Shenzhen, plans were put into effect to spread develop ment further into the interior, in second tier industrial centers, which have grown rapidly in recent years. Seemingly aware from the onset that urban industrial ization also leads to increasing inequalities, widening income gaps, and potential social unrest, Chinese plan ners now appear to be preparing to shift away from maximizing the speed of industrial expansion to at tempting to deal with social and economic welfare. The mixed socialist–capitalist Chinese model may be the most important and innovative ongoing experiment in planning informed by the new regionalism.

Moving down in scale, the new regionalism has also been vividly expressed in the restructuring of the modern metropolis. Through the three intertwined forces of globalization, economic restructuring, and technological innovation, metropolitan regions have become more economically and culturally heterogeneous than ever before, with growing numbers of immigrants and ex panding investments from almost every country on Earth. Some have expanded to previously unheard of popu lation sizes, a few surpassing 25 million, and have become more polycentric and regionally networked. At the same time, abundant evidence exists to argue that this re structuring of the modern metropolis has been associated with increasing social polarization and widening income gaps between the rich and the poor, especially where welfare systems have been most weakened.

In many large metropolitan areas there has been a regional urbanization process that is significantly trans forming traditional suburbia through the growth of dense outer cities and edge cities; while at the same time both emptying out and refilling the urban cores with different populations and activities. Where regional urbanization has been most pronounced, as in northern and southern California, the dividing line between urban and suburban areas and lifestyles is being rapidly erased. In such technopoles as Silicon Valley and Orange County, the number of jobs locally now surpasses the number of bedrooms for commuters.

Mass regional urbanization and metropolitan re structuring have focused attention on the expansion of a global network of what have been called world cities or global cities or, more pertinent to the present dis cussion, globalized city regions. The majority of the world's population now lives in urban agglomerations of various sizes, but most of this urban population is in creasingly concentrated in 450 or so sprawling poly centric city regions or regional cities of more than one million inhabitants. In the largest of these city regions, such as on Honshu Island in Japan and in the Pearl River Delta and around Shanghai in China, megalopolitan re gional populations reach beyond 50 million. There is also little doubt that this hierarchy of million plus city re gions, economically led by the financial powerhouses of London, New York, and Tokyo, now contains well more than half of the world's economic wealth and political power as well as the vast majority of the world's one billion slum dwellers.

At an even more local scale, regionalism has begun to reach the grassroots level of activism and broad based social movements. Coalitions of labor unions, com munity based organizations, nongovernmental organiza tions (NGOs), and religious and ethnic groups have sprung up in a few large city regions to demand their residential rights to fundamental urban and regional services such as good education, appropriate and efficient public transit, adequate healthcare, a living wage, and protection from locational discrimination in the sitting of noxious facilities and the deleterious effects of pollution. Such coordinated local efforts to achieve environmental, social, and spatial justice are beginning to be imbued with a new regional consciousness, recognizing that regional dynamics are often the source of local problems.

At all geographical scales, there has been a growing recognition of the generative forces that arise from urban agglomeration and cohesive regional economies. Cities and city regions are seen as stimulating innovation, creativity, and economic development, but also as sources of increasing social and spatial polarization, environ mental degradation, inequality, and injustice. Making theoretical and practical sense of these simultaneously positive and negative generative effects has become the focal point for a new round of synergy between regional planners, development theorists, and human geographers.

Retheorizing Regions and Regional Development

A significant reconceptualization of regions and regional development has begun to emerge in association with the resurgence of geopolitical regionalism, and many of its key ideas have been coming from geographer planners, scholars who work in the overlapping spheres of critical human geography and urban and regional planning. Michael Storper, one of these geographer planners, presents a particularly pertinent discussion of the very different conceptual framework that underpins and ad vances the theoretical agenda of the new regionalism.

Storper argues that earlier approaches to regionalism and regional development treated the region almost en tirely as an outcome of underlying social, economic, and political forces. The region was a domain in which things happened, a stage for the playing out of such processes as economic development and cultural nationalism. Today, the region is being conceptualized quite differently. Storper describes the region as a fundamental unit of social life comparable in importance to the family, the state, and the market as a way of organizing human so cieties. Moreover, he claims that regions are among the fundamental driving forces of societal development, as potent as technology, social stratification, and interest seeking behavior and competitiveness, the lynchpins of urban industrial capitalism.

Seen as active developmental forces as well as dis tinctive social formations, regions significantly affect our lives in ways that go well beyond physical–environmental influences, resource availability, or relative locational advantage. Regions and in particular regional economies are now seen as generating development and change and stimulating innovation and creativity. At the same time, regional dynamics can work in the opposite direction, enhancing forces of social polarization and interregional conflict, conservatism and stasis, and repression and de cline. How to balance these countervailing forces in ac tual regional worlds of production and reproduction, maximizing one while minimizing the other, has become the central intellectual and practical–political challenge for the new regionalism.

To some extent, this view of regions and regional development represents a revival of older ideas about growth poles, spread and backwash effects, cumulative causation, and the development of underdevelopment. But it is a revival with some significant differences. First, it comes at a time when spatial and regional thinking has diffused to a wider audience of scholars, politicians, business leaders, and policymakers than perhaps ever before. It has become much less easy to ignore the voices of regional planners and academics in an age of global ization, accelerated urbanization, multiplying political regionalisms, and expanding cyberspatial technologies.

In addition, the new regional theorists have begun to advance significantly beyond these earlier ideas. Growth pole theory, for example, was essentially superficial and descriptive, rarely probing very far into what makes po larization work. Without this knowledge and with an overly idealized notion of their generative effects, it was difficult to understand why some designated growth poles led to more uneven development than less, as positive impulses were siphoned off to the already more advanced core regions. In short, there was little understanding of what makes development 'stick' in a region, rooting itself in local conditions and being sustained for long periods of time. This stickiness of regional development has been a key focus for the new regional theorists.

Regional development is now seen to be a complex and geohistorically contingent chain of processes that occurs within regional worlds of production, a comprehensive concept that is centered on the regional economy but extends to social, cultural, political, and environmental factors as well. Exogenous influences such as those arising from globalization and national policies significantly shape subnational regional development, but more attention is being given to endogenous social and spatial forces and to the role of regionally specific assets, those particular local qualities that distinguish one region from another and help link the region into the global cultural and political economy. This localized context of human and natural resources and the competitive regional advantages that arise from it are not fixed features but are significantly shaped and reshaped over time by regional cultures, politics, and planning.

In this more reflexive and relational approach, successful regional development depends on the formation of regional environments that are conducive to creative and innovative thinking and learning. These conducive conditions emerge largely through local social conventions, various kinds of social relations and associations, now often bundled under such terms as social capital, that help to create and sustain atmospheres of mutual trust, solidarity, and understanding. Vital here are the relational networks facilitated by proximity and face to face contacts and such 'soft' economic factors as untraded interdependencies, activities and linkages that do not work directly through markets, such as those of kinship, friendship, and worker solidarity. Much of the theoretical, analytical, and practical literature on the new regionalism revolves around gaining a deeper understanding of these conventions and agglomeration effects and their roles in the formation of learning economies and regional innovation systems.

The retheorization of regions has led to the revival not just of some of the best examples of earlier theories of GUD, including aspects of growth pole theory, but also of two particular ideas about what might be called urban spatial causality, how the spatial organization of cities and regions through the stimulus of urban agglomeration and clustering shapes all aspects of our lives. The two ideas are Alfred Marshall's century old theory of external economies and the arguments of Jane Jacobs about the 'spark' of urban economic life. What they lead to is a realization that without cities we would all be poor. There would be no economic or cultural or political development without the stimulus of urban–regional agglomeration.

Marshall was the first to systematically define and describe the productive forces that arise from urban contexts and how these forces lead to the clustering of economic activities in cities and industrial districts. He distinguished between two externality effects (operating outside the individual firm and production unit). Localization economies arise primarily from proximity and the efficiencies of time, energy, and cost that are associated with being close by in space. These efficiencies spur economic productivity through reducing the costs of inputs, expanding accessibility to labor pools at all skill levels, and facilitating the movement of outputs to appropriate consumption markets. More dynamic and propulsive, and also more difficult to measure in hard econometric terms, are what Marshall called urbanization economies, something in the very nature or atmosphere of urban life that creates impulses toward creativity and innovation, and hence toward expanded and sustained productivity.

Jane Jacobs would elaborate on these urbanization economies, focusing on the stimulating effects of heterogeneous urban density. Cities cause economic development, she would argue, not because people are smarter in cities but rather because urban densities and proximities produce a concentration of need and increased incentives to think about problems in new ways. Her ideas were ignored for decades until an interactive group of economic geographers and geographical economists revived them to promote the highly provocative idea that urbanization and urban agglomeration may be the primary generative force for all forms of capitalist economic development. Some even suggested that Jacobs deserved the Nobel Prize in economics for her discovery and began to label her version of urbanization economies 'Jane Jacobs' externalities'.

Despite these remarkable breakthroughs in thinking about urban spatial causality, we still do not know a great deal about the dynamics behind the generative effect of cities. What makes one city or city region more generative or creative than another? Are larger agglomerations more generative than smaller ones? Are there decreasing returns or agglomeration diseconomies related to urban size? What is the role of face to face contacts in innovation and learning? What is the best way to promote creative cities, regional innovation systems, and learning economies?

Perhaps the biggest challenge for regional planning today is how to make regional economies and societies more innovative and creative, more generative of positive economic and cultural development, while at the same time controlling for the almost inevitable negative spillover effects, the closely associated tendencies toward increasing inequality, exclusion, and social polarization. To meet these challenges, there must be an effective response to the growing critiques of the new regionalism, especially its tendencies toward a narrow economism that overemphasizes success stories and the positive effects of generative regional economies; as well as its premature dismissal of underdevelopment theory, the continued importance of core periphery relations, and the early lessons learned about the interdependency of spread and backwash. Development must never be reduced to a singular progressive process.

A Brief Conclusion

The new regionalism is still in its early stages of development and application, and many disagreements still exist with regard to its definition, scope, and theoretical foundations. Rather than being clearly framed or entirely new, it is highly eclectic and draws selectively upon many earlier ideas and traditions, making it difficult at times to distinguish the old from the new approaches. Some critics claim that the new regionalism is not much different from older forms of regional thinking and practice, and is little more than a modified version of the growth pole theories and spatial systems planning associated with (abandoned) welfare regionalism. For other critics, it is merely a currently fashionable catchword to attract attention for place marketing and related neoliberal forms of competitive and essentially entrepreneurial localism. And as has been noted throughout the history of regionalism and regional planning, there is the tendency to exaggerate the positive effects of regional development and neglect the negative accompaniments.

All these criticisms contain more than a grain of truth and need to be kept in mind whenever regionalists become overly enthusiastic and blithely optimistic in their theoretical and practical arguments and claims. Two conclusions, however, are relatively unquestionable. The first is that the new regionalism has opened up several highly innovative directions for geographical research and planning practice that are likely to shape both geography and planning far into the future. The second is that never before have the concepts of region and regionalism been given such potency in human affairs and in the wider frameworks of social, economic, and political theory.

See also: Aménagement du Territoire: Territorial Development; Cassa per il Mezzogiorno; City-Region; Concentrated Deconcentration; Core-Periphery Models; Devolution; Neoliberal Economic Strategies; New Regionalism; Polycentricity; Region; Regional Inequalities; Regional Innovation Systems; Regionalism; Uneven Regional Development.

Further Reading

Brenner, N. (2000). *New State Spaces: Urban Governance and the Rescaling of Statehood*. Oxford: Oxford University Press.
Brenner, N. and Theodore, N. (2002). *Spaces of Neoliberalism: Urban Restructuring in North America and Western Europe*. Oxford: Blackwell Publishers.
Friedmann, J. (2005). *China's Urban Transition*. Minneapolis, MN: University of Minnesota Press.
Friedmann, J. and Weaver, C. (1979). *Territory and Function*. Berkeley, CA: University of California.
Harvey, D. (2006). *Global Capitalism: Towards a Theory of Uneven Capitalist Development*. London: Verso.
Jacobs, J. (1969). *The Economy of Cities*. New York: Random House.
McLeod, G. (2001). New regionalism reconsidered: Globalization and the remaking of political economic space. *International Journal of Urban and Region al Research* 25(4), 804 829.
Saxenian, A. (2005). *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Cambridge: Harvard University Press.
Scott, A. J. (1998). *Regions and the World Economy*. Oxford: Oxford University Press.
Scott, A. J. (ed.) (2001). *Global City Regions: Trends, Theory, Policy*. Oxford: Oxford University Press.
Soja, E. W. (1989). *Postmodern Geographies: The Reassertion of Space in Critical Social Theory*. London: Verso.
Soja, E. W. (2000). *Postmetropolis: Critical Studies of Cities and Regions*. Oxford: Blackwell Publishers.
Storper, M. (1997). *The Regional World: Territorial Development in a Global Economy*. New York: Guildford Press.
Swyngedouw, E. (1997). Neither global nor local: 'Glocalization' and the politics of scale. In Cox, K. R. (ed.) *Spaces of Globalization*, pp 137 166. New York: Guildford Press.
Weaver, C. (1984). *Regional Development and the Local Community: Planning, Politics and Social Context*. Chichester: John Wiley.

Regional Production Networks

D. W. Edgington, University of British Columbia, Vancouver, BC, Canada

Glossary

Agglomeration Economies The economic benefits that firms obtain when locating near each other.

Branch-Plant Economy The phenomenon of transnational corporation companies building factories (branch-plants) in other countries, primarily to sell products in the local market.

Distance Decay Curve Representation of the weakening or fading of a process or pattern with increasing distance.

Flexible Specialization A strategy at company level of permanent innovation, and accommodation to, rather than control of, ceaseless change. This strategy is based on innovation, multiuse equipment, and flexible, skilled workers.

Focal Factories Factories having the capacity to operate all major production functions, from research and development (R&D) to parts manufacture and assembly.

Fordist Mass Production A form of production characterized by an assembly line (conveyor-belt factory system) and standardized outputs linked with the stimulation of demand brought about by low prices, advertising, and credit and exemplified by the mass-production systems based on the principles used by the car maker, Henry Ford (1863–1947).

Input–Output Analysis A method used in economics to trace the connections between products and services (the output) and the resources needed to produce them (the input).

Marshallian Industrial District A specialized manufacturing locality characterized by a high density of small and medium-sized firms that cooperate along the production system/chain.

NAFTA (North American Free Trade Agreement) Area One of the three 'triad' supranational regions, the others being the European Union and Asia Pacific.

NIES (Newly Industrialized Economies) The four Asian economies of South Korea, Taiwan, Hong Kong, and Singapore that gained industrial status sometime around the 1960s.

OEM (Original Equipment Manufacturer) Refers to where one company sells the product of another company, which it then sells into the market under its own brand. It usually refers to the company that originally manufactured the product.

Post-Fordism A system of production characterized by flexibility of both labor and machinery, the vertical breakup of large corporations, small-batch production, better use of links between firms so that subcontracting is increasingly used, just-in-time production, and niche consumption. It is associated with agglomeration and interaction between linked forms of economic activity.

Production System/Production Chain The different production stages through which a good or service passes before being sold to the buyer.

Roundaboutness of Production Adoption of more specialized, differentiated intermediate goods (e.g., parts and components) in the production process.

Sunbelt Region Refers to states such as Texas, California, and Florida in the south of the USA that experienced rapid growth in population and jobs in the last 40 years.

Systems Approach The interdisciplinary branch of science, dealing with analysis of systems in fields such as industrial processes, management, decision-making processes, and environmental protection processes.

TNCs (Transnational Corporations) Large companies having production and service operations spread over a number of different countries.

Vertical Disintegration The separation of a production process into separate companies, each performing a limited subset of activities required to create a finished product, as opposed to integration, in which production occurs within a singular organization.

The tendency of manufacturing establishments to cluster in space has been fairly constant and comprises a popular focus of geographic inquiry. Regional production networks (RPNs) provide a useful heuristic in examining this phenomenon and can be described as the full range of activities within a region necessary to take a product from its conception and design to end use and beyond. Accordingly, research into RPNs involves examining the nature of the 'production system' and (in the first instance) how firms are connected through the physical input and output of raw materials, manufactured parts, components, subassemblies, and eventually final products. Indeed, RPNs can be conceived as a chain of production that comprises a linked sequence of functions in which each stage adds value to the overall process of production of goods. These systems comprise networks of firms that are related to one another through flows of goods – as well as services – in order to manufacture a particular product, for example, an automobile, a computer, an airplane, or even industrial products such as machinery and equipment. Some studies also apply the

RPN concept to the 'manufacture' of services, such as the complex financial service cluster in key international cities – Tokyo, New York, and London. Whatever the final product involved, these types of networks are often structured in regions, and where they are sufficiently large in economic terms, intricate in structure and inte grated then they merit being called 'regional production networks' in their own right.

Introduction

Historical and contemporary examples of networks of production firms organized within a specific region in clude the heavy industries of the Ruhr valley in Ger many, the New York garment district, the jewelry and the gun trade in Birmingham, and the clothing, furniture, and engineering industries in London. In Japan, this phe nomenon occurs also; perhaps the most familiar being the motor assembly industry which has clustered in Tokyo and Toyota City through the use of just in time delivery systems for parts and materials. Other contemporary examples include a number of industrial districts spe cializing in textiles, ceramic tiles, and machine tool networks in Northern and Central Italy, the electronics, multimedia, and cultural products agglomerations in Los Angeles, and the technology intensive industrial cities and districts in the Baden Wurttemberg region of Ger many. Certain cities contain significant concentrations of life science industry (biotechnology) firms, including Boston, the San Francisco Bay area, Los Angeles, San Diego, North Carolina's Research Triangle, and the Philadelphia–New Jersey Pharm Country.

In examining a regional production network (RPN) and its economic underpinnings the key focus is on the particular firms and establishments involved rather than the 'industrial sector' *per se*. For instance, the intrinsic advantages arising from spatial clustering in a particular region will only be detected by exploring the concrete behavior of firms and their establishments rather than by merely examining industry wide statistical data, which comprise the sum of all firms. Moreover, even the geo graphical clustering of firms within similar industries does not by itself necessarily prove any connection or relationship. In this regard, RPNs are a useful approach for studying the clustering behavior of industries as they force the researcher to address the concrete relations between establishments, including assembly firms and their suppliers. Of course, such scrutiny of the backward and forward linkages of individual manufacturing plants and allied operating units can be very time consuming. Typically, managers with decision making powers (usu ally purchase or sales managers) have to be asked specific questions about their business operations and linkages, usually to ascertain connections with supply firms and customers within the regional economy.

Ideal Types

A number of ideal types of input–output structures within RPNs can be suggested (see **Figure 1**): (a) The production of a simple commodity in one establishment along artisan style manufacturing lines. This type hardly exists today, but hypothetical examples include fashion shoes made from a local shop, or milk supplied from a local farm. (b) Processing of resources and the transfor mation of raw materials into final products, often in one establishment, through 'flow production', such as oil re fining, petrochemicals, timber processing, newsprint, and metal smelting. (c) Sequential production systems where many firms are involved in generating a final product in a number of discrete stages. If the products of each step are the unique inputs of the next, it is a 'linear' production system. (d) If there are many parts brought together to be compiled into a final product then it may resemble a 'pyramidal' system. Specialization between plants and

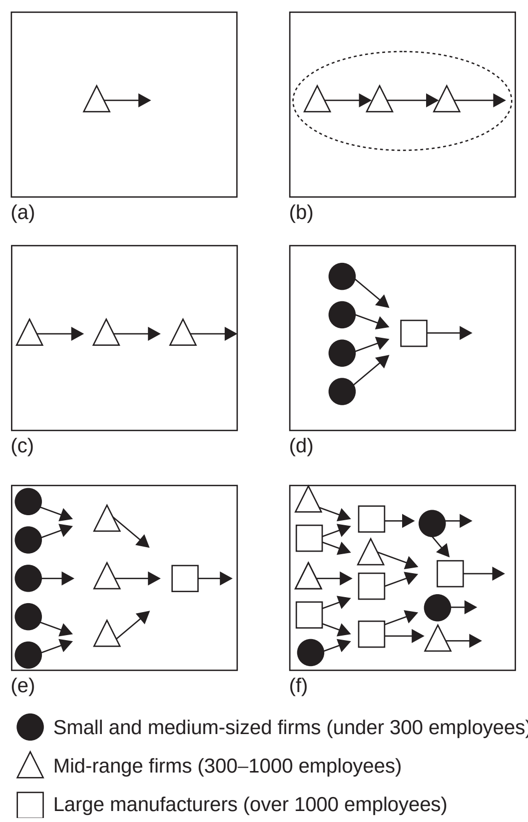

Figure 1 Six alternative representations of input output structures within regional production networks. (a) The production of a simple commodity in one establishment along artisan-style manufacturing lines; (b) the processing of resources and the transformation of raw materials into final products through flow production; (c) a linear production system sequential arrangement; (d) a pyramid production system; (e) nested systems of component-assembly production system; and (f) a more complex production system.

firms occurs along the various stages of production. (e) Nested systems of component assembly production occur where many parts and subassemblies are brought together to be compiled into a final product, such as an automobile or airplane. (f) More complex systems exist; in the vast majority of cases production systems overlap, and intermediate and final goods do not have unique downstream or upstream suppliers and customers. For instance, steel producers send their materials to a variety of industries, including car frames, construction, and miscellaneous machinery operations; semiconductors are likewise produced and then used in a number of allied electronic products.

An extension of this framework is to include activities other than production as part of the links that bind firms together in RPNs (see **Figure 2**). These include research and design, distribution, marketing, and services given to the final consumer, such as periodic repair and main tenance. Accordingly, if in practice production systems are discussed primarily in terms of linkages between manufacturing firms, then conceptually links with other types of firms – such as primary sector companies, re tailers, wholesalers, trading companies, business services, and financial organizations – can be readily incorporated. Indeed, many manufacturing firms themselves typically engage in 'nonmanufacturing' activities. Firms in RPNs are linked through financial systems and the use of credit and other service networks. A further flow connecting and coordinating establishments in an RPN involves the passage of market and technological information and decision making flowing back and forth between parts suppliers, assemblers, and final customers (see **Figure 2**).

An additional complication is how RPNs are organ ized in terms of corporations and labor markets. At one extreme, a 'single product' firm might achieve vertical integration among several plants and manufactures all the necessary intermediate materials, components, and parts in its own plants. In reality, most large manu facturing firms specialize on particular processes in the 'production chain', and purchase a large proportion of materials and parts from suppliers and subcontractors. Sometimes this is due to the workforce specializing in particular processes according to a complex division of labor. Whatever the particular arrangements, the decision of individual firms in the RPN as to whether to 'make' by themselves or 'buy in' from elsewhere is continually shifting as they make decisions about which functions to perform 'in house' and which to 'outsource' to other firms. In this regard, a number of ideal types of 'gov ernance structures' for RPNs have also been identified (see **Figure 3**). (a) RPNs are sometimes based on very dense localized transactions between small independent firms. (b) They may consist of a differentiated mixture of independent firms and of the branch plants and affiliates of multiplant firms, many of which are transnational corporations (TNCs). (c) RPNs can be organized in a 'hub and spoke' configuration with numerous firms linked together through strategic alliances or they may be part of business networks coordinated by key 'broker' firms, usually involving large assembly operations such as airplane or automobile production.

Yet another issue is that RPNs can be contained within a single geographic location – such as an indus trial district, metropolitan area, or industrial belt of several cities – or spread more widely across geographic space. Indeed, the term 'region' is somewhat problematic in that it has been applied to arrangements both at the subnational and supranational levels (e.g., the San Fran cisco Bay region on the one hand and the North American Free Trade Area (NAFTA) region on the

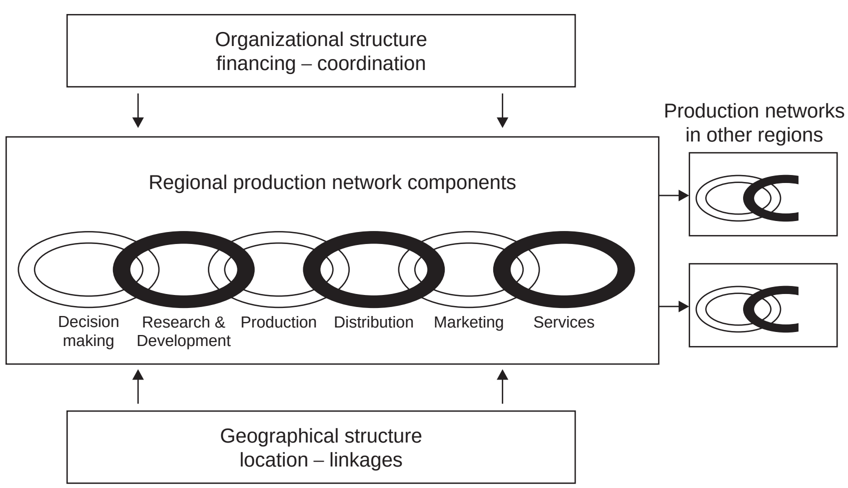

Figure 2 Regional production networks in geographical and organizational context. Adapted from Edgington, D. W. and Hayter, R. (1997). International trade, production chains and corporate strategies: Japan's timber trade with British Columbia. *Regional Studies* 31, 151 166.

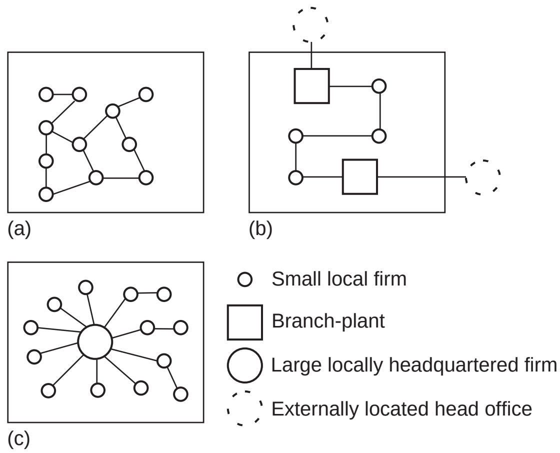

Figure 3 Three ideal governance structures for regional production networks. (a) Very dense transactions between small independent firms. (b) A differentiated mixture of small independent firms and branch-plants. (c) Hub-and-spoke configuration. Adapted from Markusen, A. R., Lee, Y.-S. and DiGiovanna, S. (eds.) (1999). *The Second Tier Cities: Rapid Growth beyond the Metropolis*. Minneapolis, MN: University of Minnesota Press.

other). As noted by Dicken, individual RPNs are themselves invariably enmeshed in broader production networks of inter and intrafirm relationships with intricate links – horizontal, vertical, and diagonal – forming multidimensional, multilayered lattices of economic activity (see **Figure 2**).

Approaches to Studying RPNs

Just as the structure of developed economies has changed over time, so has the composition, content, and focus of study of RPNs. Thus, prior to the 1970s the dominant research was a focus on the individual factory rather than systems of factories. However, changes in industrial organization during the twentieth century made traditional locational analysis lose much of its utility. These included adoption by entrepreneurs of mass production techniques resulting in an escalating scale of industrial plants and their evolution into multiplant corporations forging backward and forward linkages not only with other manufacturing activities but also with primary and tertiary activities. This necessitated the development in each firm of complex bureaucratic decision making organizations – often in office locations both separate from and integrated with factory functions. Because of these and other factors, research efforts in the 1970s diverged from the hitherto dominant mold of Weberian and Loschian analysis in two different ways. First, a fascination within social sciences on system theory and input–output analysis led to the first generation of research on RPNs. In a pioneering paper by economic geographers, Fredriksson and Lindmark studied the Swedish automobile firm Volvo and found that the core firm was dependent on suppliers and developed long term relationships with them to ensure the effective operation of its own production system. Second was a concern by human geographers with the geography of enterprise. The interest here was on the importance of corporate organization and behavior, especially with respect to the evolution of their spatial and linkage patterns and the relations between information flows, innovation diffusion on the impact on cities and regions. Early proponents of this approach combining industrial and corporate change with shifting industrial location emerged from the writings of Dicken, Krumme, and Watts, among others.

The Disintegration of RPNs

While RPNs have evolved largely in metropolitan regions their long term sustainability has depended to a substantial degree on regional closure and spatial monopoly, particularly in intermediate products. By the 1980s geographers argued that links between firms within RPNs were becoming weaker and that the growth of mass production factories dominated by large corporations led to many RPNs tending toward 'branch plant' and subsidiary economies. Thus, in many cases the growth sectors of RPNs were externally owned by firms having their 'focal factories' and headquarters outside of industrial regions and located in the major corporate centers of advanced economies. The implication was that a local factory's ability to expand may not depend entirely on its competitive stature and with very little R&D in branch plants, employment opportunities for skilled managerial talent was diminished. Consequently, the historical link between regional production and industrialization disappeared from view for many years in favor of a new nationwide (or even international) spatial division of labor based on the large corporation. Some economic geographers went so far as to argue that agglomeration economies were completely inoperative in the new corporate economy, on the grounds that material linkages, labor pools, and information were all internalized and commanded by the large firms. Hamilton and Linge in particular contended that RPNs in advanced industrial economies appeared to have lost some of their prior cohesion or had even disintegrated completely. Apart from the prevalence of nonlocal linkages, a second factor was the sharpened industrial competition at both a national and international level which made industrial systems and subsystems at the urban and regional level in developed countries vulnerable to shifts to low wage regions, either at home or offshore. Economic recession in the mid 1970s combined with the global shift of industries to developing countries in Asia Pacific and elsewhere brought about the displacement and substitution on world markets of a widening range of intermediate and finished goods. In other words, pro ducers in older industrial regions were being superseded by producers in emerging industrial areas, centers, and countries.

The impacts on many traditional RPNs were sub stantial contraction, if not virtual disappearance of major segments of intraregional, vertically integrated pro duction chains. Over the 1970s and 1980s this process occurred in regions involved in coal mining, ore mining, steel production, marine engineering, and shipbuilding in Western Europe and North America. Affected regions included the northeast USA, Lorraine in France, the Rhur in Germany, and the north of England, together with the industrial districts of many inner cities. The overall effect of competition with 'newly industrializing economies' (NIEs), such as South Korea, Taiwan, and Singapore, and new industrial districts in 'sunbelt re gions', for example, the south of the USA, led to rising unemployment and falling incomes in older industrial regions, leaving small subcontracting firms and remaining companies to restructure and increase their backward and forward linkages outside the region. Overseas, TNCs initiated and operated a significant proportion of industry in the regions of certain NIEs and developing countries; however, they seldom (at least initially) developed strong linkages with indigenous and small businesses. As a result of this and an increase in 'external control' of advanced regions by the headquarters of TNCs, it was argued that the regional specialization in the form of major segments of even entire production networks of a narrow range of industries was becoming displaced by specialization in limited segments of the production chain of a much wider range of industries.

The Rediscovery of RPNs

Still, almost at the same time, others maintained that in fact, industries and RPNs still prevailed over large firms and decentralized production networks as the building blocks of geographical industrialization. Indeed, at a time when many economic geographers were writing deeply pessimistic accounts about crises in older areas that had been dominated by Fordist mass production, an economic revival was becoming evident in certain regions that had developed apace on the basis of post Fordist industrial growth and an increasing 'roundaboutness' or a vertical disintegration in manufacturing into many specialized production firms and local labor markets. One of the new schools of thought may be designated as the Italian School, as represented primarily by Becattini and his co workers at the University of Florence, who revitalized the Marshallian idea of the industrial district to account for the spectacular growth of neoartisinal manufacturing in Northeast and Central Italy after the 1970s. The other is sometimes referred to as the Californian School of eco nomic geographers due to its focus on the industrial geography of Southern California and the San Francisco Bay Area. This group also emphasized the role of vertical disintegration, the rise of a 'round about' character of production, interindustrial transactional networks, and the importance of local labor markets as the main factors underlying a new spatial agglomeration displayed by many post Fordist forms of economic activity, often called 'flexible specialization'. Scott, Christopherson, and Storper in studying the printed circuit industry and the women's clothing industry in Los Angeles and of semi conductor production in Silicon Valley, and Hollywood's film and television industries, observed a strong process of vertical disintegration in response to market conditions.

In these and other industries, such as high technology products with high risk environments and rapid lead times to market, development cycles were short, proto typing was rapid, batches of production often small, the variety of products great, and costs spread across a wide range of goods. Consequently, the economies of spatial clustering remained extremely important, particularly for small firms in which there was a high degree of or ganizational interdependence between plants, a need to minimize transaction costs and maximize economies of scope rather than scale. Other scholars, such as Marku sen, examined high growth RPNs in a variety of lo cations yet concluded that it was much easier to identify 'hub and spoke' structures and 'satellite platforms' of branch plants rather than identify clean cases of Mar shallian industrial districts among fast growing RPNs. Thus, Silicon Valley in her study was represented as a conglomerate RPN with elements of hub and spoke, sat ellite platform as well as 'state centered districts' (where government supported operations were important) exist ing side by side with Marshallian relationships dominated by small scale stand alone firms.

Issues

RPNs and Technology Development

In recent years the study of RPNs has picked up two important themes. The first relates to issues of long term dynamics associated with technology development. As economic structures in advanced countries have changed from traditional to sophisticated production the role of technology and long term growth has been important to the study of both nations and regions. Along with such a shift it is possible to discern – broadly speaking – how early concerns with the material (input–output) dimen sion of production networks and transaction costs in the late 1980s was superseded in the 1990s by a focus on the sociocultural and institutional dynamics within regions.

As exemplified in the work of both Storper and Maskell, writing in this field is now replete with terms such as trust, norms, routines, common cognitive frameworks and decision rules, conventions, practices, and learning. There is a view that well integrated RPNs have the best chance of technological innovation through the role of these untraded interdependencies. While not diminishing the importance of conventional approaches to examining agglomeration economies, the recent stream of analysis suggests that the underlying dimension which confers competitive advantage on firms located in RPNs consists of ready access to a common knowledge base supported by close social interactions and by institutions building trust and encouraging informal relations among various actors, including skilled workers.

In particular, tacit knowledge, the deeply personalized knowledge possessed by individuals that is virtually im possible to make explicit and to communicate to others through formal mechanisms, has a very steep 'distance decay curve'. It requires direct experience and interaction and depends to a considerable extent – though not completely by any means – on geographical proximity. A similar interest in clusters as a source of competitive advantage has been taken by economists such as Krug man and management scholars such as Ohmae and Porter. Many 'technology regions' (or so called 'learning regions') have been identified as a source of national competitive advantage (e.g., Silicon Valley, Boston, Aus tin, Seattle, Boulder, Raleigh Durham, the M4 Corridor in England, Munich, Stuggart, Baden Wurtenberg, Paris Sud, Grenoble, Montpellier, Nice/Sophia Antipolis, Milan; and in Asia, Tokyo, Seoul Inchon, Tapei Hsinchu, and Kyoto).

RPNs and International Boundaries

The other novel theme revolves around the nature of RPNs that cross international boundaries. There is a widely accepted view that the geographical extensiveness of virtually all production networks has increased and now 'slice through' international boundaries in highly differentiated ways. Influenced in part by regulatory and nonregulatory barriers and local sociocultural conditions, TNCs and affiliated companies in different parts of the world create uneven, highly complex, and dynamic webs of production networks, economic spaces, and places connected together through threads of flows. Research into global value chains by Gereffi and others stresses both producer driven global networks, characteristic of those industries in which TNCs are other large inte grated industrial enterprise that play the central role in controlling the production systems (including its back ward and forward linkages), as well as buyer driven sys tems where control is exercised by the large retailers, brand name merchandisers, and trading companies. Global production networks have also been studied by Dicken and colleagues at the University of Manchester.

Two Case Studies – The Production of Watches and Toyota Automobiles

An interesting example of international competition and the need for RPNs to address technological change concerns a study by Glasmeier of the 250 year struggle for watch industry supremacy among a number of in dustrial regions. Starting in the mid eighteenth century watch manufacturing was conducted by 'putting out' watch parts among small farms and cottages scattered across the southwest Lancashire plain in England. While watches produced in this way retained their status of a product for the well to do, such a system operating on cottage industry lines could not resist competition from lower cost areas elsewhere, specifically the more ma chine intensive Jura region of Switzerland and the even more mechanized production system utilized in the northeastern states of the USA. At the beginning of the 1880s the so called 'American system' of mass production was on a par with Swiss production in integrated and semi integrated manufacturing firms. In the twentieth century Japan also gradually learned the secrets of watch production and Tokyo emerged to overtake the Swiss watch industry (by the 1970s), with a high degree of vertical integration and was among the first to move decisively into producing digital watches. At the same time, Hong Kong grew as the world's most formidable manufacturer of low priced watches alongside the com mercialization of quartz watch technology, which re placed traditional mechanical technologies. The dawn of the electronics age for watch making occurred with wage inflation in both the United States and Japan, forcing the search for cheap labor locations. Accordingly, Hong Kong and later the adjoining Pearl River delta region in southern China took over the low price end of the in dustry, based on labor intensive production and an elaborate subcontracting system. Despite a dramatic growth in global production capacity, Hong Kong pro ducers today continue to be successful in organizing the assembly of watches for original equipment manu facturing (OEM) distribution to retail outlets around the world, as well as developing their own watch brands. However, critical watch components such as watch movements are bought in from foreign suppliers in Switzerland, the USA, and Japan. Glasmeier's case study shows clearly that decline and growth in the older watch making regions in Europe, North America, and Japan did not occur in isolation but was the result of the com petition between RPNs for the leadership position as the dominant technology changed. For instance, the Hong Kong RPN in watches started due to its role as a parts distributor and service provider for Asian markets and grew into the established leader in the assembly at end of the production chain. While the transition in industrial dominance across the world took many years to emerge this study shows that different regions with quite dif ferent governance systems predominated over the life of the world watch industry.

Another study by Edgington focused on changes in the Japanese automobile industry during the 1980s and 1990s. Toyota Motor Company (TMC) with its network of suppliers in Toyota City (and adjoining Chukyo region in central Japan) has often been cited as a classic 'hub and spoke' type of RPN, one that has a single controlling firm linked through structured production networks to satellite rings of supplier firms. Indeed, the implicit hy pothesis of this and many other studies is that 'Toyota ism' (and the associated concept of 'lean production') has fostered the spatial clustering of activities, in contrast to the principle of the spatial division of labor as charac terized by 'Fordism': Henry Ford, of course, was the recognized inventor of the branch plant that enabled a spatial fragmentation of production processes along with the decentralization of car assembly plants to location near the major markets. However, in the late 1980s, TMC had to react to a new set of production problems that suggested that there were practical limits to the gains achieved in manufacturing efficiency by continued geo graphical clustering. Specifically, in the context of Japan's overheated 'bubble economy' of that time the over concentration of automobile production and just in time delivery systems led to traffic congestion on Chukyo's crowded roads. As a result, the just in time system (called *kanban* in Japanese) could no longer deliver parts from suppliers to assembly line operations within the severe 15–20 min turnaround time required. An even more critical concern was the impeding shortage of young blue collar workers who could not be found in, or attracted to, Chukyo to work in TMC's busy and de manding assembly factories. Faced with these new chal lenges, Toyota was forced to consider a radical change and a dispersal strategy seeking cheaper and more available land and labor within Japan. This led to its very first major assembly plant outside its original 'corporate fortress' to open in the rural prefecture of Fukuoka, Northern Kyushu region, some 700 km south from Toyota City – at Toyota Motor Kyushu (TMK). Al though affected by spatial distance from Toyota City, the new company used 'almost just in time' methods in order to adapt to the local situation and the longer supply lines for certain parts delivered from Chukyo. In fact, TMK used an innovative mix of nearby suppliers, in house production of parts, and reliance on more sophisticated materials and components from 'home' factories and suppliers in Chukyo. Nonetheless, the study illustrated that distance itself provided a surprising bonus. The new plant in Kyushu gave Toyota engineers the 'space' to rethink and redesign the approach to making a car. Many innovations were tried out in TMK and then in turn introduced into assembly factories back in Toyota City. This assisted production development overall by

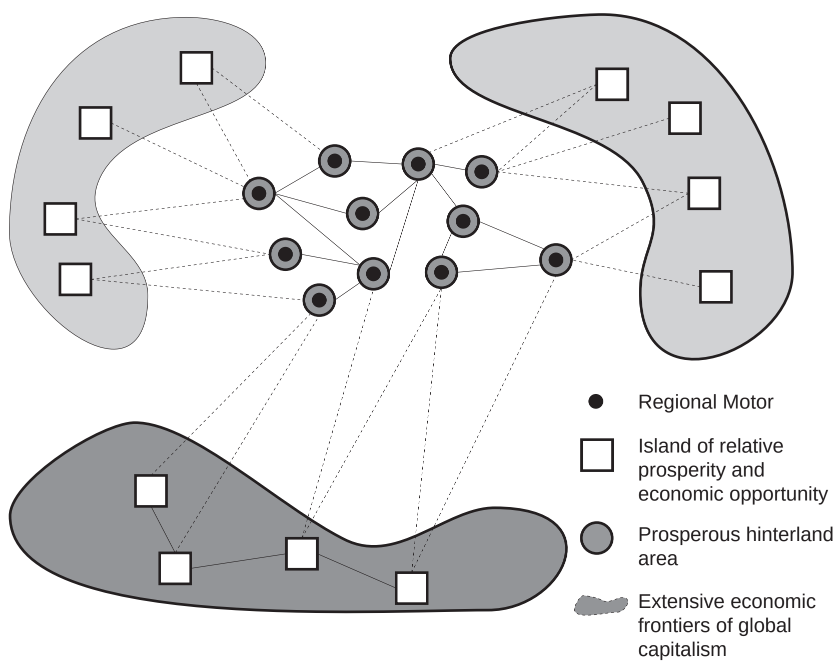

Figure 4 A world of regions. From Scott, A. J. (1998). *Regions and the World Economy: The Coming Shape of Global Production, Competition and Political Order*. Oxford: Oxford University Press.

facilitating access to ideas and techniques that were new to the Chukyo production region.

A World of Regions

The two broad streams of research identified earlier – technological change at the local regional level and links between RPNs across regional and even international boundaries – have begun to converge together as an in creasingly global world is now seen as anchored in a worldwide archipelago of industrial regions (see **Figure 4**). Conversely, it is realized that even advanced technology regions cannot be sustained and developed further entirely through indigenous networks. As illus trated in the Toyota case study, a key additional process involves the connections between some of the actors in a particular RPN with outsiders, for instance, firms in other regions who are suppliers, customers, or sources of specific information and knowledge. In other words, in addition to 'local buzz' of a dynamic RPN there also have to be 'pipelines', channels of communication to other actors in other places. This new perspective has sparked off research by geographers and others into the growth and development of RPNs in China, and other indus trializing countries of East Asia and elsewhere. Today, there are many 'new' industrial regions, yet on the other hand, many parts of the developing world have clearly been left behind in terms of manufacturing growth. As noted by Scott, it is hoped that geographical knowledge of RPNs and regional growth mechanisms might in some way offer alternative strategic approaches to development in economically backward areas.

See also: Agglomeration; East Asian Miracle; Economies, Branch Plant; Firms; Foreign Direct Investment; Global Production Networks; Industrial Location; Industrial Restructuring; Knowledge Communities; Regional Integration; Relational Economic Geography; Technology and Regional Development; Transnational Corporations in Developing Countries.

Further Reading

Bathelt, H., Malmberg, A. and Maskell, P. (2004). Clusters and knowledge: Local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography* 28, 31 56.

Becattini, G. (1990). The Marshallian industrial district as a socio economic notion. In Pyke, F., Becattini, G. & Sengenberger, W. (eds.) *Industrial Districts and Interfirm Co operation in Italy*, pp 37 51. Geneva: International Institute for Labour Studies.

Dicken, P. (2007). *Global Shift: Mapping the Changing Contours of the World Economy* (5th edn.). New York: The Guildford Press.

Dicken, P. and Lloyd, P. (1990). *Location in Space: Theoretical Perspectives in Economic Geography* (3rd edn.). New York: Harper and Row.

Edgington, D. W. (2002). Lean and its limits: The Toyota production system in Chukyo and Northern Kyushu, Japan. In McNaughton, R. B. & Green, M. B. (eds.) *Global Competition and Local Networks*, pp 216 246. Aldershot: Ashgate.

Edgington, D. W. and Hayter, R. (1997). International trade, production chains and corporate strategies: Japan's timber trade with British Columbia. *Regional Studies* 31, 151 166.

Fredriksson, C. G. and Lindmark, L. G. (1979). From firms to systems of firms: A study of interregional dependence in a dynamic society. In Hamilton, F. E. I. & Linge, G. J. R. (eds.) *Spatial Analysis, Industry and the Industrial Environment, Vol. 1 Industrial Systems*, pp 155 186. Chichester: John Wiley and Sons.

Gereffi, G. and Korzeniewicz, M. (eds.) *Commodity Chains and Global Capitalism*. Westport, CT: Praeger.

Glasmeier, A. K. (2000). *Manufacturing Time: Global Competition in the Watch Industry, 1795 2000*. New York: The Guildford Press.

Krumme, G. (1969). Towards a geography of enterprise. *Economic Geography* 45, 30 40.

Krugman, P. (1991). *Geography and Trade*. Leuven: Leuven University Press.

Markusen, A. R., Lee, Y. S. and DiGiovanna, S. (eds.) (1999). *The Second Tier Cities: Rapid Growth beyond the Metropolis*. Minneapolis, MN: University of Minnesota Press.

Ohmae, K. (1995). *The End of the Nation State: The Rise of Regional Economies*. New York: Simon and Schuster Inc.

Porter, M. E. (2001). Regions and the new economics of competition. In Scott, A. J. (ed.) *Global City Regions: Trends, Theory, Policy*, pp 139 157. Oxford: Oxford University Press.

Scott, A. J. (1988). *Metropolis: From the Division of Labor to Urban Form*. Berkeley: University of California Press.

Scott, A. J. (1998). *Regions and the World Economy: The Coming Shape of Global Production, Competition and Political Order.* Oxford: Oxford University Press.

Storper, M. and Christopherson, S. (1988). Flexible specialization and regional industrial agglomerations: The case of the US motion picture industry. *Annals of the Association of American Geographers* 77, 104 117.

Watts, H. D. (1981). *The Branch Plant Economy: A Study of External Control*. Harlow: Longman.

Relevant Websites

http://www.sed.manchester.ac.uk
A collaborative research enterprise at the School of Environment and Development, University of Manchester, "Making the Connections: Global Production Networks in Europe and East Asia."

http://www.globalvaluechains.org
'Global Value Chain Initiative', a collaborative website hosted by the Institute of Development Studies (IDS), University of Sussex, United Kingdom. The conveners of this initiative are Gary Gereffi (Duke University), John Humphrey (IDS), and Tim Sturgeon (MIT).

Regional Science

H. G. Overman, London School of Economics, London, UK

Glossary

Discrete Choice Estimation A type of regression analysis (see linear regression) where the dependent variable involves a choice between two or more discrete outcomes.

General Equilibrium Theory An economic theory that explains production, consumption, and prices of all goods in the economy. In contrast, 'partial equilibrium theory' is concerned with one particular good.

Imperfect Competition Any market structure where there is some competition, but firms still have some market power (i.e., the ability to exercise some control over prices prevailing in the market for its product).

Increasing Returns to Scale A situation in which the average cost of producing a unit of output decreases as more units are produced. It is sometimes also called economies of scale.

Linear Regression A method of modeling the expected value of one (endogenous or dependent) variable as a function of other (exogenous or independent) variables.

Microeconomic Models Theoretical models that start with assumptions on the behavior of economic agents and then try to explain the functioning of the economy from the 'bottom-up'.

Panel Data Data with repeated observations (e.g., over time) for specific individuals, firms, locations, etc.

Production Inputs and Outputs Inputs are resources used in a production process and outputs are the results of that production process.

Spatial Autocorrelation Roughly, a situation where there is a degree of dependence across observations in different locations. It is also referred to as 'spatial dependence' or 'interdependence'.

Spatial Heterogeneity It is used in spatial statistics/econometrics to refer to a situation where there are differences across space in terms of the process being studied. In human geography, similar concerns are reflected in the emphasis on context, contingency, and situation.

What Is Regional Science?

Papers in Regional Science (the journal of the Regional Science Association International) provides a definition of regional science as the "study of social, economic, political and behavioural phenomena that have a spatial dimension." This is clearly a rather broad definition and does little to separate regional science from the study of human geography more generally (apart, perhaps, from the latter's traditional emphasis on the human use of the physical environment). A second way to delineate the subject is to focus on the institutional history, and talk about the work of Walter Isard and his followers in founding and developing the organizations and journals specifically concerned with regional science. This is a rather well trodden path and, while interesting, is not always particularly enlightening for those that did not take part in that development. The third possibility is to take a narrower view of regional science as a methodological approach that is primarily concerned with the rigorous mathematical or quantitative analysis of the causes and consequences of spatial socioeconomic disparities. This is the approach taken here. Such a narrow delineation of the subject area is not without consequences and would certainly be contested by a number of regional scientists. It is, however, representative of the view of regional science held by the vast majority of social scientists. As this article is particularly concerned with the relationship between the social sciences and regional science, and specifically with the role of theory, it thus seems to be most appropriate to adopt here.

Theory

Taking this approach, initial developments, during the 1950s and 1960s, can be viewed in one of two ways: either as introducing geography into hitherto aspatial economic theory or else using economic theory to provide a deductive framework for a hitherto atheoretical, essentially inductive, economic geography. A considerable amount of energy has been expended discussing the appropriate characterization of this process. However, the outcome does not appear to be in dispute. The resulting spatial theory was, and continues to be, strongly associated with formal mathematical models. This approach to the treatment of spatial economies traces its roots back to von Thunen's model of the isolated state. von Thunen was concerned with the role that yield and transport costs played in determining patterns of agricultural land use around cities. His approach to modeling determinants of this land use was to start with a number of simplifying assumptions. A city in an isolated state surrounded by wilderness removes the need to think about links to other places. A flat featureless plain means that farmers involved in the same activity would be equally productive wherever they are located, removing any role for first nature geography (i.e., differences in the physical environment). Farmers ship their produce straight to market (literally – no transport network here) but incur transport costs to do so. Finally, farmers locate so as to maximize their profits. This simple and highly stylized model leads to the well known von Thunen rings with different activities occupying clearly defined rings around the city. The exact detail of what is produced, where, is less interesting than the fact that this ordered pattern is consistent with observed patterns of land use. From a theoretical point of view, the key point is that this ordered pattern emerges from the operation of competitive land markets without any coordination on the part of farmers. More than 100 years later, in the 1960s, the Alonso–Mills–Muth model (re)formalized these ideas to explain why, within a city, land values and population densities decline as one moves away from a central business district.

Of course, as with any stylized model, these theories of land use can be subject to criticism. Three issues in particular prove key in understanding further developments. First is the problematic assumption of the existence of a single center. Second is the fact that accessibility to this center is the only factor that explains urban structure, leaving no role for spatial interactions between individuals and firms. Third is the empirical observation that the theories may work quite well as models of land use, but they do not function so well as explanations of changes in land use. The reaction to these criticisms by different groups from the 1970s onward is informative. Geographers turned increasingly toward other social sciences and away from the quantitative geography that had hitherto overlapped extensively with regional science. The reaction from regional scientists was also to look to other disciplines, but in the sciences and not the social sciences. Drawing on social physics, the interaction between individuals was modeled mathematically using behavioral assumptions constructed from analogies with the laws governing the behavior of particles in physics. Alternatively, individual interactions were modeled using theories of mathematical ecology with animals and plants replacing particles as the basis for human behavior. A third group of urban and regional economists took the middle road, continuing to develop formal mathematical models which incorporated interaction between individuals but drew on economic and social theories of individual behavior. A similar story played out in another key area: that of location theory.

Alfred Weber (formalizing and publicizing ideas from Laundhart) provided an early model of industrial location. In the spirit of von Thunen, a number of simplifying assumptions make the problem tractable. In the simplest case, firms combine two inputs in fixed proportions to produce a final good. The two inputs are available in locations A and B, while the final good must be sold in the market at C. Transporting inputs or outputs between locations is costly. The assumption that firms face perfect competition (so prices do not depend on firm decisions) combined with the other assumptions on fixed inputs and transport costs reduces the firm's profit maximization problem to finding the location where total transport costs are minimized. The model can be complicated to allow for transhipment costs, new source of supply, etc., and (depending on the degree of complexity) algebra or algorithms solve for the firm's location decision allowing an understanding of the influence of factor prices, changes to transport costs, etc., on location. Extensions, by Moses, drop the assumption of fixed proportion technologies allowing firms to substitute away from expensive inputs. This approach to the industrial location decision has had some success as an empirical model for explaining the location decision of particular firms, but it is less satisfactory as a broad theory of location for two key reasons. First, the simplifying assumption of perfect competition ignores questions of increasing returns to scale and of imperfect competition which give firms spatial monopoly power. Second, the theory is only partial. It can explain the location of particular industries given the location of the market, but, as with land use theory, it cannot explain the location and existence of that market. Parallel developments in models of spatial competition, notably the work of Hotelling, addressed the first criticism. The second was addressed by attempts to provide a general equilibrium theory of location. This search for a general equilibrium theory has been a key objective of regional science from its foundation.

Early models, centered around the idea of central places, attempted to explain the spatial distribution of economic activities within an urban hierarchy. Christaller provided the first formulation by asking how a hierarchy of activities (in terms of the size of their market area) could be distributed to ensure that an evenly distributed rural population had access to all goods. The answer involved a series of hierarchical cities with nested market areas such that a location where a particular good is available also supplies all goods with smaller market areas. Whereas Christaller's central place theory drew heavily on empirical observation and inductive reasoning, Losch provided microeconomic foundations with market areas derived from consideration of the behavior of individual firms. Losch recognized that space meant that these firms would exercise some monopoly power (contrast Weber) and that this monopoly power would be limited by entry of other competing firms. However, lacking the necessary mathematical model of imperfect competition, closing the model in the general equilibrium case required *ad hoc* assumptions on the distribution of population and the assumption that all firms chose to locate at a common metropolis. One problem is that the resulting spatial distribution generates many inconsistencies with the underlying behavioral assumptions and so is not a true general equilibrium of the model. More importantly, Losch derives the efficient pattern, not the pattern that emerges from a decentralized decision making process. Isard, in his seminal book, introduced the work of these German scholars to an English language audience as well as formalizing the microeconomic foundations of these approaches to location.

That book, and subsequent volumes, one on the search for a general theory and the other on research methods essentially created the field of regional science. Fairly quickly, however, this new field reached the crossroads already described above for the theory of land use. Economic geographers reacted against abstraction with increasing emphasis on context, contingency, and situation. Economists focused on issues of spatial competition and the importance of spatial interaction in determining location patterns, but restricted themselves to the problems that were amenable to analysis based on behavioral assumptions drawn from mainstream economics (eventually, via Krugman's new economic geography – NEG – this would lead to the first general equilibrium economic theory of location). Regional scientists either turned to increasingly complicated partial equilibrium theories of firm location or worked on developing a mixture of abstract, mathematical, and general equilibrium models. The latter drew increasingly on analogies from the hard sciences. This story, of initial engagement with other social scientists (particular geographers and economists) followed by a period of increasing abstraction and progressive disengagement is repeated across other theoretical areas such as transport modeling or migration analysis. Some might argue that a nascent reengagement with economics (on location theory) and geography (on geographic information systems – GISs – and spatial data analysis) is underway. Before considering this, attention is turned to the other key plank of work in regional science – the development of a methodological toolbox for regional and urban analysis.

Empirical Methods

Although regional science draws on a wide range of methodologies, for many it is most closely associated with a smaller number of quantitative methods. Some would argue that the approach is also positivist, although as the role of theory is often ambiguous (discussed further below) this is a more contentious characterization. Any listing of the core methods will be subjective, but for many would include: econometrics and statistics (and the spatial variants thereof), regional impact models (particularly economic base models, input–output analysis, and more recently computable general equilibrium (CGE) models), spatial and social interaction models (including the gravity model) and, perhaps, GISs and associated spatial data analysis techniques. The focus here is on the three areas most closely associated with regional science: spatial econometrics, regional impact models, and spatial interaction models.

Spatial Econometrics

Econometrics is used in regional science, as in economics and other social sciences, to give empirical content to theory and to test hypotheses derived from that theory. To take a simple example, many location models predict that trade between locations is decreasing with respect to the distance between those locations. Econometrics can be used to test whether trade does indeed decline with distance (i.e., test a hypothesis derived from theory) and, if so, to provide an estimate of the degree to which trade declines as distance increases (i.e., to provide empirical content to the theory). While general econometric methods have been broadly applied in regional science, it is particularly associated with the development and application of spatial econometrics. Spatial econometrics traces its origins to the early 1970s when attempts were made to begin to deal with the methodological issues that arise in multiregion models when there is some form of statistical dependence between outcomes in different regions. Of course, aspatial econometrics also worries about such issues, but what sets spatial econometrics apart is its concern with spatial dependence. That is, with the notion that geographical space, broadly defined, would help shape the nature of any dependence. Spatial econometrics is also concerned with spatial structure or heterogeneity. Again, the feature that distinguishes spatial from aspatial econometrics is the concern with understanding and allowing for the role of heterogeneity across geographical space.

There are three main reasons for considering spatial effects, including spatial dependence and heterogeneity. First, the validity of a number of commonly used econometric techniques is based on underlying assumptions that will be violated in the presence of these spatial effects. Thus, correcting for these spatial effects is important if one is to reach valid conclusions about the nature of the relationships of interest. This 'space as nuisance' view of spatial effects has been a key concern of the spatial econometrics literature. Second, correctly modeling spatial effects can help extract information from data and improve predictions of spatially determined variables, even in situations where we do not understand why such spatial effects occur. This 'space as a source of information' view of spatial effects has long been a concern of the spatial statistics literature and has been of considerable interest in some areas of physical geography (e.g., kriging). In contrast to these nuisance and information views, the third reason to consider spatial effects is because 'space matters'. That is, the interest is in developing techniques that allow one to explain how space affects the relationship of interest. While clearly not mutually exclusive concerns, these three contrasting views, and the need to balance research efforts to address them, represent a source of ongoing tension in terms of spatial econometrics relationship both with regional science and with the wider social science community.

Initial interest in spatial econometrics came from researchers interested in multiregion models. Space clearly matters here, but this was not necessarily reflected in early developments which focused on detecting and correcting for residual spatial autocorrelation or on improving predictions in the presence of such autocorrelation. To take a stylized example, imagine a researcher interested in whether crime rate in a neighborhood was determined by socioeconomic characteristics of individuals living in the neighborhood. After collecting appropriate neighborhood data, the researcher runs a linear regression of crime rate on selected neighborhood characteristics. Using the estimated model the researcher is able to predict neighborhood crime rates on the basis of the available socioeconomic data. These predicted neighborhood crime rates can be compared to the actual and an unexplained 'residual' calculated as the difference between the two. Those residuals should be random thus displaying no systematic pattern. One possible departure from randomness, and a key issue of interest of spatial econometrics, concerns the spatial pattern of these residuals. For example, when plotted on a map the residual for a given neighborhood should be unrelated to those of other neighborhoods nearby. If, in contrast, positive residuals in one neighborhood tend to be associated with positive residuals in nearby neighborhoods (and similarly for negative), then the residuals display spatial autocorrelation. At best, this has implications for the statistical significance of the researcher's findings; at worst it means that the strength or even the direction of estimated relationships may be wrong. In addition, if interest is in predicting crime rates *per se*, then using the information on the nature of this spatial autocorrelation may help improve those predictions even if we do not understand the socioeconomic processes that actually drive that autocorrelation.

Clearly it would be useful if these kinds of errors could be detected and the spatial econometrics literature (often using insights from spatial statistics) has developed tests to do just that. The two most common are Moran's *I* and Geary's *C*, although other measures are available. Clearly, if spatial autocorrelation is detected the regression model should be respecified. Exactly how it should be respecified, however, depends on the source of the spatial autocorrelation. There are three possibilities, best illustrated through the continued use of the example on the relationship between neighborhood crime and socioeconomic characteristics. The first possibility is that the crime rate in a neighborhood increases and this, in turn, directly increases the crime rate in nearby neighborhoods. For example, a rise in crimes in a neighborhood encourages copycat crimes in nearby neighborhoods. This can be captured in the regression model through the inclusion of information on crime rates in nearby neighborhoods. The second possibility is that the socioeconomic characteristics of a neighborhood change in a way that increases crime in that neighborhood and also directly increases crime in nearby neighborhoods. For example, the number of young people in a neighborhood increases and they commit crimes in both that neighborhood and nearby neighborhoods. This can be captured in the regression model through the inclusion of information on the socioeconomic characteristics of nearby neighborhoods. The third possibility is that unexpectedly high crime rates in one neighborhood tend to be associated with unexpectedly high crime rates in nearby neighborhoods but that this effect does not work directly (through, e.g., copycat crime) or indirectly (through socioeconomic characteristics). This happens when there are factors that cause crime that are unobserved (at least to the researcher) and correlated across neighborhoods. This can be captured by assuming that there is spatial autocorrelation between the residuals of neighborhoods. That is, one solution to the problem of spatial autocorrelation of the residuals is specifically to allow for the spatial autocorrelation of the residuals in a revised specification! This feels somewhat circular and, in terms of understanding the underlying socioeconomic processes, is only appropriate if one can rule out the other two mechanisms through which spatial autocorrelation arises.

This discussion may well give the impression that it is hard to distinguish between these three different possibilities. The more formal treatment available in standard spatial econometrics texts confirms that this is indeed the case. It would be fair to say that these identification problems have received little attention in the spatial econometrics literature. Attention has, instead, focused on the specification and estimation of linear spatial regression models (including debates around the determination of appropriate 'spatial weight matrices') and the formal properties of the resulting estimators and associated test statistics. Efforts also went in to expanding the spatial approach to include panel data and discrete choice estimation. Increasingly, this emphasis and a growing interest in spatial dependence have moved spatial econometrics in to the mainstream econometrics literature.

While admirable, this progress in dealing with space as nuisance and as a source of information for prediction has not, however, been matched by comparable advances in the applied spatial econometric literature in increasing our understanding of situations in which space matters. There are two main problems here. First, the focus of too many applied spatial econometrics papers is on the implementation of spatial econometrics with the result that far too little attention is paid to constructing analysis that are informative about theory. The burgeoning growth convergence 'industry' is a good example of this. When attention is focused more directly on theory, the problem is that the proposed tests of many theoretical propositions regarding spatial behavior do not properly identify the precise mechanism through which interdependence occurs. Of course, in the spatial setting, this sort of identification is extremely difficult. In the crime example above, it is almost impossible to determine whether spatial interdependence in crime rates works through the direct or indirect mechanism. To separate out these two mechanisms one would need a way to exogenously change crime rates in one neighborhood and see what effect this had on nearby neighborhoods. In reality, the only way through which this might happen is by changing the socioeconomic characteristics of a neighborhood, but then both mechanisms will be in operation and there is no way to separate them out. In some situations, it may be possible to directly change the dependent variable but even then any change needs to be independent of changes in the other explanatory variables. For example, when considering tax competition between jurisdictions it may be possible to identify the interaction between tax rates, providing that changes do not reflect other changes in the neighborhoods. More attention to deriving clear predictions from theory and the associated search for identification should be central to the application of spatial econometrics by regional scientists trying to test spatial theories. It is not, and as a result, while spatial econometric theory is moving in to the mainstream econometrics literature, much applied spatial econometrics is ignored by mainstream economics. Of course, acceptance by mainstream economics is not the objective of many regional scientists. But the crucial issue here is the reason for that rejection not the rejection *per se*. A similar story, also involving the link between theory and empirics, plays out with respect to regional impact models, which represent another set of key methodological tools in regional science.

Regional Impact Models

Regional impact models are used to predict how changes in the economic environment, broadly defined, may affect a region's economy. Economic base models identify a set of export sectors (e.g., airplane manufacturing) that are independent of the local market and that form the base of the regional economy. The remaining sectors are dependent on the local market (e.g., supermarkets) and ultimately the size of the local market depends on the size of the economic base. If employment in the economic base increases, local market size increases, this in turn increases employment in the nonbase industries, which further increases market size and employment in base industries. Eventually, this circular process dampens down and overall regional employment has increased by some multiple of the original increase in employment. Input–output models work in an essentially similar manner, although they are considerably richer in their modeling of the interdependency between different industries (acting as one another's suppliers and customers) and the links between industry, consumers, government, and external economies. Again, though, the basic idea is that changes in one area of the regional economy impacts on other areas, which in turn have yet further impact on other areas. As each subsequent impact is smaller than the one that generates it, the effect eventually dampens down with the original impact having now been multiplied in the rest of the economy. Alternative assumptions about 'leakages' from the system (spending outside the region) and the exact nature of the interdependencies will change the magnitude of multipliers.

During the early days of regional science considerable emphasis was placed on the development of regional impact models both as a method of increasing understanding of regional economies and as an applied tool for advising on the impact of policy changes. The traditional approaches, however, have struggled to deal with two particular criticisms. First, the models ignore supply constraints and focus only on demand. Clearly this means that they will overstate the impact of changes in the economic environment if supply constraints are binding. The degree to which this is a problem is an empirical question, but some research suggests that this is more than just a theoretical concern. This focus on demand also means that the models cannot be used to predict the impact of supply side policies (in addition to a range of other policies that cannot be well specified within the model). The second problem is, arguably, more fundamental. The models rely on empirical estimates of the strength of the interdependency between sectors. Unfortunately, the strength of these interdependencies may change in response to changes in policy. This is the Lucas critique that reduced form parameters (e.g., the input–output coefficients) depend on the policy regime. Defenders of this kind of regional modeling, if they have addressed the criticism at all, have tried to argue that many policy changes are incremental and thus unlikely to lead to changes in the underlying parameters. Again, this may be true, but theoretical modeling tells us that these reduced form parameters may change in response to changes in the economic environment more generally (e.g., in prices and wages). Empirical evidence suggests that these changes do indeed occur. In the absence of a satisfactory response to this criticism, traditional regional impact models have become increasingly marginalized as tools for understanding spatial relationships, although they may remain useful for some types of policy analysis.

One interesting avenue which attempts to address both criticisms is the move toward CGE models. These models are developed from explicit assumptions on utility and production functions of consumers and firms. Even if these functions are not completely stable, they are much more independent of policy changes and the general economic environment than the reduced form parameters that are the basis for more traditional regional impact models. Unfortunately, data requirements for CGE modeling can be high and at the moment key parameters are often taken from international trade models and applied, almost certainly inappropriately, to subnational settings. Despite this problem, the number of applied regional models is certainly growing. Interestingly, the CGE approach is also in line with the numerical simulation approach used by NEG.

Spatial Interaction Models

Spatial interaction models seek to explain flows between locations. The two most common flows considered are those of people (e.g., migration) and goods (e.g., trade) and the most common way of modeling these is to use some form of gravity model where flows are proportional to the mass at destination and location and inversely proportional to the degree of friction between those locations. For example, trade flows between A and B are modeled as proportional to the multiple of gross domestic product (GDP) at the two locations and inversely proportional to the distance between the two locations.

As with regional impact models, gravity and other interaction models have shown themselves to be very useful for fitting data and thus provide a simple way to make predictions about the impact of certain policies. Researchers have also developed a large number of rigorous theoretical models that have an appropriately adjusted gravity model as the reduced form explaining flows as a function of mass and frictions. If one knew which of these theoretical models was appropriate, one could estimate the underlying structural parameters and then, in contrast to regional impact models, provide predictions that are not subject to the Lucas critique. Of course, the problem is that, if we only observe flows, location mass, and frictions, it is impossible to identify which theoretical model is appropriate because they all make the same prediction about the relationships between these variables. Thus, while in principle, the availability of a variety of rigorous theoretical underpinnings distinguishes spatial interaction modeling from regional impact modeling; in practice both are essentially reduced form of approaches to understanding the spatial economy.

Interestingly, the regional science versus social science split that was discussed with respect to location theory also plays out for spatial interaction modeling. The gravity model was originally based on an analogy with Newton's law of gravitation. In providing theoretical foundations, regional science has tended to focus on drawing on the hard sciences (e.g., physics) while other researchers in quantitative geography and economics have continued to emphasize behavioral assumptions drawn from the social sciences.

Regional Science and the Social Sciences

Regional planners and transportation departments (among others) need to make decisions that have spatial dimensions, and for this group the methodological toolbox developed by regional scientists has proved extremely useful. Academically, however, the fortunes of regional science have proved much more mixed. Beginning with Isard's 1954 call for an 'association of regional scientists' interest grew quickly. *Papers and Proceedings*, The Regional Science Association, was first published in 1955 and by 1958 regional science had another journal (the *Journal of Regional Science*) and its first department (the Regional Science Department at the University of Pennsylvania). By 1960, roughly 960 members were listed in the Regional Science Association directory. The field continued to grow in terms of both size and geographical coverage (to Europe, Japan, India, Latin America, and beyond). By the 1990s its World Congress could truly claim to be a global meeting of regional scientists. But by the late 1990s, the field's founding department at Penn had closed and, despite the existence of programs elsewhere, the capacity of the field to propagate itself remains unclear. In part, these more recent problems arise from the unresolved intellectual issues that have been the central focus of this article. Regional science certainly has not evolved into a well defined discipline. Instead, it continues to define itself as a multidisciplinary approach to problems having a spatial dimension. But a multidisciplinary approach must engage with other disciplines and, as discussed above, for a large part of the 1970s and 1980s the scientific turn has seen regional science drift away from other social sciences. This is particularly true of both mainstream geography and economics (the two obvious disciplines with which regional science might be expected to engage).

There are signs that a degree of reengagement may be underway. The interaction between spatial econometrics and main stream econometrics has already been discussed above. For quantitative geographers, GISs and spatial analysis provide the area of increasing overlap. For economics, Krugman's NEG and the continuing urban economics tradition provide obvious areas of synergy. In this context, two issues appear to be key for regional science's interactions with the wider social sciences. The first is the relationship between GIS and the wider human geography community. The second, and the main focus of this article, is the willingness to respond to the challenges raised by quantitative social science (e.g., economics) regarding regional science's approach to location theory, applied spatial econometrics, regional impact modeling, and spatial interaction. At this point in time, this reengagement is not necessarily viewed positively by researchers on all sides. Economists, in as much as they pay any attention to regional science, struggle with the behavioral assumptions (drawn increasingly from hard science) that underlie much of the theorizing in the field. Regional scientists, in turn, often fail to recognize the contribution of economists in constructing spatial theories based on microeconomic behavioral foundations. At best, this is because of a disagreement over the type of assumptions that are acceptable when micro founding quantitative models. At worst, it reflects disagreement over the need to micro found such models. It should be said, however, that the latter distinction more often provides the dividing line between economic geographers, on the one side, and regional scientists and economists on the other. The reactions to Krugman's NEG provide a good case study. Regional scientists felt that earlier work in the field had been ignored. Economists simply did not recognize earlier work as providing appropriate micro foundations. These arguments are being repeated with other areas or research. At this point, it is unclear whether the final outcome of these arguments will be renewed interaction between regional science and other social sciences or whether, once again, the various approaches to issues of spatial inequalities are destined to drift apart.

See also: Economic Geography; Economic Geography, Quantitative; Location Theory; Spatial Science.

Further Reading

Alonso, W. (1964). *Location and Land Use*. Cambridge, MA: Harvard University Press.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.

Beckmann, M. J. (1968). *Location Theory*. New York: Random House.

Cliff, A. D. and Ord, J. K. (1981). *Spatial Processes: Models and Applications*. London: Pion.

Florax, R. J. G. M. and Plane, D. A. (2004). Special Issue: The Brightest of Dawns. *Papers in Regional Science* 83, 3 385.

Fujita, M. (1989). *Urban Economic Theory: Land Use and City Size*. Cambridge: Cambridge University Press.

Fujita, M., Krugman, P. and Venables, A. J. (1999). *The Spatial Economy: Cities, Regions and International Trade*. Cambridge, MA: MIT Press.

Fujita, M. and Thisse, J. F. (2002). *Economics of Agglomeration: Cities, Industrial Location and Regional Growth*. Cambridge: Cambridge University Press.

Isard, W. (1956). *Location and Space Economy: A General Theory Relating to Industrial Location, Market Areas, Land Use, Trade, and Urban Structure*. Cambridge, MA: MIT Press.

Isard, W., Bramhall, D. F., Carrothers, G. A. P. *et al*. (1960). *Methods of Regional Science: An Introduction to Regional Science*. Cambridge, MA: MIT Press.

Loveridge, S. and Jackson, R. (eds.) (1999). *The Web Book of Regional Science* http://www.rri.wvu.edu/regscweb.htm (Accessed in May 2008). Morgantown, WV: Regional Research Institute.

Miernyk, W. H. (1965). *Elements of Input Output Analysis*. New York: Random House.

Mills, E. S. (1972). *Studies in the Structure of the Urban Economy*. Baltimore, MD: The Johns Hopkins Press.

Muth, R. F. (1969). *Cities and Housing: The Spatial Pattern of Urban Residential Land Use*. Chicago, IL: University of Chicago Press.

Richardson, H. W. (1969). *Regional Economics: Location Theory, Urban Structure and Regional Change*. New York: Praeger.

Wilson, A. G. (1970). *Entropy in Urban and Regional Modelling*. London: Pion.

Relevant Websites

http://www.elsevier.com
Regional Science and Urban Economics exists to facilitate and encourage high quality scholarship on important theoretical and empirical issues in urban and regional research and is published by Elsevier. *The Journal of Urban Economics* (also published by Elsevier) is a highly ranked journal that carries regional science content.

http://joeg.oxfordjournals.org
The Journal of Economic Geography published by Oxford Journals is also a highly ranked journal that carries regional science content.

http://www.blackwellpublishing.com
The *Journal of Regional Science* publishes original analytical research at the intersection of economics and quantitative geography and is published by Blackwell Publishing.

http://www.regionalscience.org
The Regional Science Association International (RSAI): RSAI is an international community of scholars interested in the regional impacts of national or global processes of economic and social change. *Papers in Regional Science* is the official journal of RSAI and is published by Blackwell Publishing.

Regionalisations, Everyday

B. Werlen, Friedrich-Schiller University, Jena, Germany

Glossary

Action Theory The theory grounded mainly in the work of Max Weber, emphasizing the core significance of subjective action for the understanding of society and its key role as a transformative power in the course of human history and shifting socioeconomic and sociocultural realities.

Constructionism It starts from the premise that all meanings of elements of the taken-for-real worlds are constituted on the basis of (inter-)subjective constituting processes. Therefore, social, economic, or cultural realities as meaningful realities can never be seen as objective realities, in the sense that, they exist independently from their construction within the course of human/social practices, but not independently from their (meaningful) constituting processes.

Locale A spatial context or setting for action comprised of material elements as well as of sets of social norms and culturally shared values, to be understood as a material, socioeconomic, and sociocultural constellation of action with inter-subjective shared meaning contents. Therefore, it is an action-related concept that cannot be turned into an objective fact or generalized as a social category, having the same meaning for members of a society (in a certain region).

Region One of the core concepts of geography, it is the outcome of academic or everyday practices of regionalization. Traditionally, it is understood as a form of spatial delimitation of natural and/or sociocultural units. According to the spatial approach, it is the outcome of an academic procedure of formal classification, referring to spatial as well as social, economic, cultural, or political categories, producing meaningful units for different types of spatial politics or geographical analysis. In an action-centered perspective, the region signifies the result of everyday regionalizations as one of the key forms of everyday geography-making.

Regionalization Traditionally understood as an academic practice of spatial delimitation of natural, social, economical, cultural, or political spheres of reality. According to an action-centered perspective, it is understood as an everyday practice, not for delimitating the 'world' spatially, but rather for using spatial references for the economical, political, cultural, etc., structuring of social realities. Therefore, the key issue is what could be called the production of different forms of 'world-binding' for the performance of human or, more precisely, social agency.

Regional Geography It represents social and economic, especially cultural and political, relations in spatial categories as spatial realities. Traditionally, regional geography referred to its object of study as an objectively pregiven entity, sometimes even as naturally pregiven in the sense of geodetermined reality. In more recent theoretical conceptualizations, the goal of regional geography is seen in the contextual analysis of social and economic processes or as the descriptive precondition for the analysis of the processes of everyday regionalizations.

Introduction

This article explains the concept of 'everyday regionalizations' elaborated in the context of action centered (or *praxis* centered) structurationist geography. In comparison to older forms of 'regional geography', this concept emphasizes that space is a social construct, ultimately a category of thought through which human beings make sense of, and then act within and upon, the world around them. Consequently, it is argued that what should be centralized in the study of regions is less about the physical constituents that the researcher might identify as comprising a given region, a delimitable portion of space on the Earth's surface, although this act of identification may continue to be a dimension of the research project. Rather, it is requested that we should take serious note of what is called 'geography making': the many ways in which human beings mark out – through their ideas and practices – what they evidently regard and act upon as meaningful places or 'regional' entities, or nation states perhaps, in the course of which they lend a constructed reality to such 'regions' within the conduct of their everyday lives (hence 'everyday regionalizations'). More philosophically, objection is made to theoretical positions that, however apparently sophisticated, lapse into treating regions as if they are preexisting entities, as prior 'containers' encompassing human actions, instead of acknowledging from the outset that the only reality possessed by these regions is as thought and practiced constructs of people for whom, for whatever exact reason, they have come to matter. Importantly, then, this approach highlights a crucial distinction between quite a bit of continental European geography, which insists that we must not dispense with the so called 'constructionist' understandings, and certain

Anglophone geography traditions, old and new, which too readily slip into assuming the physical 'realness' of regions (elevating them to ontological status or, in effect, lending *too much* importance to the category of space in deciding what, for the academic, is primary in the world).

What is at issue in this article, then, is once again the status and content of regional geography (as laid out in two other articles in the encyclopedia expressly tackling regional geography). Regional geography is still one of the main occupations of geographers, especially in teaching and in the field of geographical representations of the world, and over the last two or more decades, the most outstanding geographers have made several prop ositions for the conceptual revision of regional geog raphy. The focus of these propositions is clearly linked to the ambition of giving regional geography a more solid framework based on both the latest achievements in so cial theory and the most recent theoretical developments in human geography. In this debate, it becomes obvious that the attempt is no easy task, and that the reorientation of regional–geographical research is asking for a very serious, sustained effort. Indeed, nothing less than the revision of the overall relationship between (human) geography and social science, and we might add certain schools of philosophy, is at stake. Arguably, moreover, the propositions made so far are not going far enough to grasp in an appropriate way the current geographical conditions of modern, globalizing society, and are trap ped with the notion of the region – as this real, bounded entity – that is largely outdated in the face of collapsing physical (and political) borders. This article argues that, for scientific research – and Benno Werlen and others in the continental European heritage who are reluctant to drop claims to providing 'scientific' accounts of the human world – what is required is a change in per spective from regional geography, old and new, to a sci entific exploration of the processes integral to the everyday regionalizations occurring in the economic, political, and cultural registers.

This approach owes much to structuration theory and structurationist geography, and at root is an exercise in revised 'methodological individualism', perhaps more precisely called 'methodological subjectivism', accepting that ultimately it is in individual human beings as acting subjects – in their apprehension of the world, making sense of it through mental processes, attributing meaning to certain of its contents, and then acting upon this sense and meaning making – that the dynamic for all social life resides. To be sure, collectives of human beings may make inter subjective sense, attribute meaning, and en gage in actions that share much in common, out of which arise patterns that effectively sediment into social structures that subsequently serve – in classic struc turationist thinking – as inputs to, even constraints upon, subsequent rounds of sense and meaning making. It is only in the actions of the people involved, in what they do, how they do it, and why, that the structures are constituted, including those that might be said to have a spatial dimension (such as the location of capitalist en terprises). As such, the approach of everyday regionali zations can be cast as an 'action centered social geography', where it is human actions as *social* actions that routinely and repetitively produce the forms of so cial life – actions that must be themselves anchored in people's sense and meaning making. In this connection, it is also claimed that in many instances, particularly in the rapidly changing post or late modern world, there are thoughts and actions proliferating which serve con stantly to reembed older regionalizations, notably ethno regionalist or nationalist ones, in the teeth of globalizing pressures. Further, it is emphasized that people cannot but always be embodied or situated in time and space, as in the formulations of Torsten Hägerstrand's time geography, and that through proximate face to face interactions, as well as now through mediated inter actions over longer distances, they build up pictures of the geographies that they are inhabiting, suffusing them with meanings, often shared meanings, and therefore create a symbolically charged grounding – an everyday regionalization – that is central to all manner of ensuing economic, political, and cultural activities, events and processes.

The purpose of this article is to demonstrate the content and significance of these claims, and to bring into the encyclopedia a flavor of some distinctive, and in some ways controversial if in other eyes unfashionable, ideas less well known in Anglophone geography than they could be. The article proceeds in a somewhat unorthodox manner, however, by mounting a sustained critique of relatively recent theoretical maneuvers made in the study of regions by leading Anglophone human geographers, notably Nigel Thrift, seeking to show how they are cri tiqued *en route* to an alternative perspective concerned with everyday regionalizations.

Nigel Thrift's Contextual Regional Geography

Nigel Thrift, with his contextual reconceptualizations of regional geography from the early 1980s, notably in his seminal 1983 paper published in *Environment and Planning D: Society and Space*, through into a sequence of early 1990s papers in *Progress in Human Geography*, has delivered one of the most important propositions to the debates introduced above. His approach aims to take postmodern geographical conditions conceptually into account. The objective is to replace both the traditional approaches to regional geog raphy as typified by Alfred Hettner, Paul Vidal de la Blache, and Richard Hartshorne – 'compositional' approaches in the sense of creating a composition of disparate (physical and human) elements – and the spatial approach of spatial science, wherein regions largely cease to matter, by focusing on the progressive spatiotemporal disembedding and fragmentation of social worlds. Thrift's suggestion here focuses upon human agency, on what human beings think and do, and he regards the regional as a 'contextual' aspect of social practices, in that within regional contexts what matters is the manner in which social practices (of many different kinds) come together to lend distinctive character to the region in question. His can also be taken as an attempt to take into account changing ontologies, social as well as spatial, as tied up with a supposed 'speeding up' of social life in the contemporary era, banishing all older certainties about what is real (i.e., the traditional region as written about by Hettner *et al.*). The aim is hence to adjust regional–geographical research categorically and conceptually to contemporary sociospatial ontology.

It can nonetheless be argued that this attempt is not going far enough. A research program that is capable of overcoming traditional regional geography has to meet three requirements. First, it has to necessarily design a new framework based on a social theory that is compatible with the aforementioned changes to social and spatial ontologies. Second, based on a social–theoretical background, both traditional and Thrift's new contextual regional geography should be reorientated in an appropriate way. Third, this new action centered (or *praxis* centered) approach ought to offer new horizons for empirical geographical research. These are the arguments to be examined in some detail over the next few pages.

Compositional Regional Geography

According to Thrift, 'compositional regional geography' – as what traditional regional geography can be called – is designed as a three level program. The first level aims to illustrate geographical determinations like topography, geology, hydrology, and climate, and is conformable with what has sometimes been termed environmental determinism. The starting point of this geography is therefore based on an orientation that concentrates on physical rather than cultural or social aspects, and is not selectively focused on social or technical problems. Hence, the physical basis is not put into human perspective but remains the foundation of everything else. The second level is centered around economic aspects. Here, the study of the organization of production in a region, including product relations, labor forces and labor conditions, could eventually lead to an analysis and representation of the class structure of a region, as well as an examination of the gender , ethnic , race , and religious specific division of labor. The third level stresses the political side of a region, where the primary aim may be to investigate the locally realized form of the nation state.

Regarding the transition from 'compositional' to 'contextual regional geography' – the most important conception of regional geography, according to Thrift – the 'work process' must become central. 'Work' has to be understood as a vital medium between humankind and nature, and it can be regarded, particularly if not exclusively through Marxist lenses, as a mode of transforming nature into societal values. This understanding then leads to the question of how the work process can be defined contextually, a question that particularly arises whenever more than one mode of work is found in a region. If a region is characterized by only one type of work (and labor) process – for example, specific types of agricultural or industrial production – it can be represented easily and adequately by the compositional account. That is certainly one of the reasons why the best and most successful regional studies generally focus on regional life worlds where one work process, like wheat growing, ship building, or tin mining activities, is dominant. The problem of compositional regional geography emerges, though, if a region is characterized by a great variety of issues, such as the work process, that are to be researched; or, to put it differently, this kind of regional research is closely bound to assumptions about a certain level of sociocultural homogeneity within a thematically defined category (such as work). If this relative homogeneity is not present, problems over how adequately to represent and describe the situation become apparent.

Contextual Theory of Social Action in Space and Time

In his attempt to outline a new conception of regional–geographical research, one that might overcome deficiencies of the older approach outlined above, Thrift critically refers to the theory of structuration developed by Anthony Giddens, Roy Bhaskar, Pierre Bourdieu, and others. Thrift's approach is justified by the assumption that a transfer of social theory into human geography is impossible without revising the definition of the relationship between human geography and social science. According to Thrift, the first step of redefining the relationship lies in becoming aware that developing a contextual theory of human action in time and space represents a remarkable challenge, not only to human geography in particular, but also to the social sciences in general. Within this kind of dialectical understanding – stressing both the primordial importance of human geography for social theory and the undeniable necessity of social theory for human geography – the elaboration of social theory inevitably implies a fundamental problem of translation. The main difficulty here lies, according to Thrift, in the fact that social–theoretical statements about societal phenomena and events are characterized by high levels of generality and abstraction, whereas the geographical approach has tended to focus on unique places and regions. These divergent cognitive interests – generalized/abstract, on the one hand, versus idiographic/concrete, on the other – are not only typical for the relationship between human geography and social science, but also for that between social history and social science.

Redefining this relationship implies both a change in research perspective and the constitution of a new research object. This new orientation is asking for a shift of focus from the circumstances surrounding humanity in general to considering particular human actors embedded in material, social, cultural, and economic circumstance, and who cannot but be thinking in and acting out of specific sociocultural worlds. Thus, instead of formulating general and abstract ideas about historical circumstances, geographical research should rather refer to theoretical categories permitting the analysis of human action in concrete social situations. General theoretical categories are then to be used for structuring the empirical field so that it becomes possible to investigate the 'real' world with 'real' humans as they 'really' are.

Within this roughly outlined reorientation toward the projection of a reconstructed regional geography, a new research agenda has to be elaborated, one based on an action centered perspective. Such a theory of action should, first of all, enable a 'reconstructed regional geography': a regional geography that centers its interest on human beings acting within a frame of specific, imposed or chosen spatial and temporal conditions. This change of focus leads – taken radically – to a change of research perspective from the regional representation of sociocultural, economic, and political realities to geographical research into the constitutive processes of (everyday) regionalization and their different spatiotemporal reaches. Before this program can be developed in some detail, and its implications discussed, it is helpful to reconstruct at greater length the steps of argument within Thrift's proposition for a contextual regional geography.

First step: Redefining the regional

Thrift's reconstructed regional geography denotes a conception of regional research that is built upon the strengths of traditional geography (attention to spatial context, attention to local population, etc.) while at the same time readjusting this basis so as to focus on new theoretical objectives and emancipatory demands. With that move, the main line of argument already becomes obvious: the elemental perspective of traditional geography should not be eliminated, only reorientated. The primary access to the social world should be carried out in a spatial–contextual sense. Elaborating on this point, action related differentiations and emancipatory standards are to be introduced into this physically defined space; that is to say, the meaning of the regional with regard to action is here of lesser scientific interest than the physically defined space – in short, the region – as spatial context.

In this respect, the region is certainly regarded as a physically defined space within which social action takes place, but not solely in the sense of being an empty, meaningless environment. On the contrary, the region here is given a constitutive meaning – constitutive for human action. Thus, the status of the region can be qualified as a kind of mobilizing, meeting place of social structure and human action. Allocating this constitutive status should not, however, be interpreted as a possibilistic (after Vidal de la Blache) or causal (after Hettner) means of interpreting humankind's struggle with nature. The constitutive power of the region in this guise rather contributes to the consolidation of social structures without excluding the creativity of human action. The region, then, is in effect treated as a passive entity, just as it remains viewed as composed primarily of material circumstances, yet it is seen as indirectly active (i.e., constitutive) because 'structure', in the terms of structuration theory, has a structuring 'effect' on action.

Second step: From a compositional to a contextual regional geography

According to Giddens, and as followed by Thrift, contextual regional research would distinguish two components: 'locale' and 'social action'. Locale denotes a specific section of space that already displays a certain pattern of material artifacts as well as persons. Thus, locale is a matter of material context, the given constellation of action and interaction, and this constellation can also be termed a 'setting' (a setting for interaction). In order to scientifically frame a region as a spatial structure of interaction, an in depth understanding of interaction is therefore a fundamental prerequisite. Interaction can be understood as an activity linking social and spatial structure *qua* institutional interrelations. In this rather narrow sense, the region is no longer predetermined as a definite natural place or pregiven entity 'out there' in the world awaiting academic discovery. Instead, its delimitation should be the objective of scientific research: it is to find out which settings of interaction can be combined in order justifiably to constitute what might be designated a region. Consequently, the region in this sense has to be further qualified as constituted by a set of interconnected settings for interaction.

In the elaboration of Giddens's theory, categories of time geography besides those of structuration theory are included. Every region, now being regarded as predetermined, albeit not solely as a physical assemblage as in the older, compositional version of regional geography but now as what researchers 'determine' to be the interconnected settings for interaction, is seen as both enabling and restricting human actions. Both options clearly influence the life paths of the individuals who live in the region. If human actors are situated regionally, the life paths of individuals have the potential to interact simply because of their nearness to one another in time and space. Whether they really interact or not depends on the pattern of production/consumption and related connections, and not, on the locales relative to the different human actors present. The life paths of the regionally situated actors can be regarded as an outcome from the allocation of time between the aforementioned locales.

Consequently, the pattern – we could say structure – inhering within the arrangement of locales is considered to be an expression and component of the local social structure. In this regard, the varying importance of different settings becomes significant, since it falls on the organization of production to shape which locale is dominant. Dominant locales are characterized by the fact that time must be allocated to them on the basis of economic and/or state imperatives, for example; and we can understand the most dominant locales as manifesting a direct link between the interaction structure of the regionally situated actors and local social structure. A whole range of ontological imperatives are thereby established which characterize this new approach to regional geography. Defining the ontological status of locale is of vital importance for the project of a contextual regional geography, as the locale represents one of the two central planks of the overall conception. For this reason, these issues need to be commented upon further.

The first crucial specification results from a problematic assumption: the premise that spatial structures are expressions of social structures has led to the (at least implicit) conclusion that spatial structures and locales or settings *are* the social. Thrift explicitly claims that the organization of production and associated institutional arrangements *determine* actors' allocation of time to certain actions in certain locales. Having this linking function, locales – interpreted as the main sites of social production and reproduction – become constitutive for the social. It is important to note that in this understanding, it is not actions with their reference to the organization of production or institutional arrangements that become constitutive, but rather the locales or settings themselves. Resultantly, an analysis of locales is indeed equated with an analysis of the social; regional research becomes the study of society, of social structure, and not of the conditions for social action. According to this reasoning, locales have five main effects:

1. First, locales structure people's life paths in space and time inasmuch as they provide the main nodes through which life must flow.
2. Second, locales can have effects on other people's life paths.
3. Third, locales provide the arena within which interaction with other people takes place; thus locales are the sites where class conflicts take place and the main context in which experience about the world is gathered and common awareness engendered.
4. Fourth, locales provide the activity structure of everyday routines.
5. Fifth, locales are the major sites for the processes of socialization, from birth to death; they represent places in which collective modes of behavior are learned.

Summarizing, locales can be characterized as particular time–space patterns or structures having a determining impact on life paths. One of the major tasks of this new regional geography as a discipline of regional societal research is hence to unpack the impacts of institutional locales on peoples' life paths: that is to say, the objects of research are, above all, spatially defined. In spatially positioning society, to put it most bluntly, locales are to be considered as spatial–social entities.

Nevertheless, Thrift noticed that contemporary societies are increasingly characterized by interaction spanning large spatial distances ('interaction at a distance'), and that uniform social groups are more and more differentiated and fragmented by increasing mobility within (and beyond) locales. These societal processes are not, even so, considered to be reasons disqualifying a regional orientation to geographical research. Accordingly, the spatiotemporal embedding that remains under late modern conditions, despite being skewered by the interrelated mechanisms of disembedding and globalization, is still sufficient to maintain the logic of traditional (i.e., compositional regional) geography, even in the first part of the program of a contextual regional geography.

The second part of the program – paraphrased with 'social action' – identifies four aspects:

personality and socialization;
penetration and the availability of knowledge;
sociability and community; as well as
conflict and capacity.

Research in the field of 'personality and socialization' is to be understood as a historical geography of life path development. This includes investigations that deal with the spatial components of the constitution of personality (self, identity, and individuality). Analogous to social history, the regional *mentalité* – a subject's starting point of action – is at stake, but this time from a human–geographical point of view. Spatial components denote the bodily bound dimensions to the constitution of personality that are, in a way, always regionally differentiated due to the physical constitution of the body (i.e., it has to be 'placed' somewhere). Therefore personality cannot be universal, but rather has to be understood as a continuously negotiated expression of social relations, including economic relations in the work process, that can vary from region to region and from locale to locale. Thrift's analysis of the specifically regional aspects of socialization is based upon these considerations: since the constitution of personality is not only the result of individual activity but also of social interaction, the study of personality additionally implies the inclusion of specifically regional aspects to socialization. It is to be assumed, as Wolfgang Hartke has pointed out, that regional aspects have a similar diversifying effect on 'class', 'race', or 'gender'. Into what kind of population cohort an individual is born is, for everybody, of the highest importance; but every population cohort cannot but be characterized by collective experiences and socialization biographies that are distinctively regional. This first part of an action oriented contextual regional geography therefore tries to reconstruct facets of regional personality formation: that is, to explain specifically regional 'traits of character', such as 'Welsh', 'Scot', or 'Corsican'.

The second part of contextual regional geography focuses on the 'penetration and the availability of knowledge'. In doing so, Thrift wants to pave the way for a conceptual recognition of the conditions for knowledge based communication prevailing within a region, and he aims to reveal the connections that constitute a specifically regional social knowledge. It is important to note that Thrift emphasizes how the availability of knowledge varies very much, and that it depends first of all upon the particular setting of a region and second, being mediated by class or other regional group memberships. Research on this count should be completed by an investigation into the meaning of language for regional *praxis*. Since (regional) language is the central medium of action and meaning transfer, studying regional knowledge is essential for a scientific understanding of the conditions for regional communication, as well as of regional meanings (or myths) attached to the social and physical world. This regionality or limitation of language is explicated by the interrelation of practice and language; due to the fact that (physically mediated) practice is always confronted with specific regional conditions, language too shows regional differentiations.

The third part of contextual regional geography refers to sociability and community. This field of research concentrates on the relationship between 'sense of community' and 'sense of place'. The spatial pattern of the social arguably finds its expression in the rural, urban, or regional community. It is the symbolic reference to place here, to spatial context complete with physical attributes, which reinforces the sense of community and determines on which of those three levels – rural, urban, or regional – the sense of community finds its strongest expression. The fourth and final part of contextual regional geography comprises the arenas of conflict and capacity, two very important aspects of social action for certain groups 'living' in certain regions. Putting special emphasis upon class conflicts and class capacity, Thrift asks what specific form conflict and capacity take in different regions at different times, as well as what form they may take today. In this respect, capacity has to be understood as the ability or the potential of a social group to transform the personality formation of individuals, forms of knowledge, and forms of sociability in an interconnected fashion. **Figure 1** illustrates the interdependence of these elements: none of them can be treated separately from the general discourse of/about the region through which they are lived as social practices. Discrepancies between the elements will result in conflict, often in the form of spontaneous protest.

As **Figure 1** implies, though, Thrift constantly assumes that the region *as* spatial context is something within and through which people live. They live their region *qua* social practice, so to say. Despite all attempts to deepen the differentiating and theoretical ambitions involved, Thrift still assumes that the region is, for all social issues, always ultimately identifiable as a homogenous entity and, even more importantly, that – although not fixed for all times – it is a starting point for an analysis of society that is at the same time necessarily a regional–geographical research. The existence of a region is not at stake in this scientific discussion, after all, but rather is accepted as something given, as a given spatial context of social action. What is important is the theory orientated exploration of the region. Regional research and social research become identical because – as developed in more detail by Allen Pred – the spatial is considered to be the social and vice versa. The problematic consequences of this equation can be seen with politicians who hold on to their traditional worldviews, complete with homogenized (and likely exclusive and

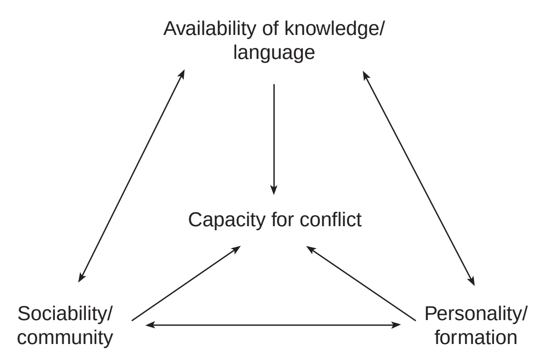

Figure 1 Elements of conflict in context from J. Thrift, N. (1983). On the determination of social action in space and time. *Environment and Planning D: Society and Space* 1(1), 23 57.

exclusionary) concepts of regions, even under globalized conditions.

Critique: From the Region to Analyzing Processes of (Everyday) Regionalization

Thrift's suggestions about how to develop a contextual regional geography entail a spatialization of the basic elements underpinning social theory. The reason for this outcome is that Thrift's perspective – just like Derek Gregory's, in his 1978 and 1982 contributions, and Allan Pred, in his 1981, 1984, and 1986 contributions – ultimately remains rooted in a space centered frame. He is thereby led to describe what he regards as spatially preconstituted social reality by means of social–theoretical terminology. It becomes clear that it is insufficient to refer solely to social theory to improve geographical research, however, even if the aforementioned problems of translation are solved. Rather, a conception of space is needed that is compatible with the basic premises of applied social theory: if this prerequisite is missing, then everyday conceptions of space – as held by real people in real circumstances – come into conflict, as they do here, as the consequence of applying structuration theory.

This tension becomes evident in the repeated calls for the integration of the subject, meaning the human actor or agent, into regional–geographical research. This demand is arguably opposed by an absolute or even concrete (material) conception of space, since a dominantly space centered perspective disables the full integration of the subject into regional–geographical research. A space centered perspective – one that takes spatial categories as primary categories for academic analysis – cannot be systematically linked with the integration of acting subjects; therefore, the ideal of a subject centered regional geography becomes contradictory. The best illustration of this highly problematic ordering of categories – space as primary category and subjects and actions as secondary categories – is given by Friedrich Ratzel's adoption of Ernst Haekel's notion of *Lebensraum* as the conditioning of all life forms and as the decisive entity in the selection of the (fittest of the) species. If space is taken as primary category of the constitution of the social, the biggest problem lies in the biological fallacy: the equation of the social with life, the implicit homogenization of the sociocultural realm, and the supposed determination of life forms by biological conditions.

If the call for a proper subject centered geographical research is to be realized, the approach must be based on a different order of categories (from agency to the social and to the spatial) and another conception of space different from those that have dominated geographical thinking and theorizing on the society–space nexus so far. We need a conception of space that is compatible with both the ontology of the social world and the agency capacity of human beings as social actors. On this basis, the constitutive power of the social world is – within a sociocultural context – to be primarily attributed to the subject, *not* to space, place, or region. In positioning the subject at the center of research, regional geography can no longer be one of the most prominent enterprises of geographical research. Instead, it needs to be replaced by a human (or, more specifically, social) geography that scientifically analyzes processes of everyday regionalization. In this move, the geographical production of regional knowledge (i.e., academic knowledge *about* regions) must no longer be the aim, but can only attain the status of a preliminary condition for investigating processes of regionalization, itself the starting point for the deconstruction of (the becoming of) action related regions.

Thrift, in his call for a contextual or 'new' regional geography, does not share this change of perspective, but instead identifies every sociocultural issue, as well as all subjects, as being *in* space. Action takes place in a region; in effect, for Thrift, even context is situated in a region. Even if he is arguing that the 'region' is lived *through*, not *in*, the basic line of argument opposes this view. Space and region are considered to be concrete and stable: it is then the geographer's task to observe what happens inside a region and report upon it in a theoretically informed manner. Speaking about social relations *in* regions, however, becomes increasingly difficult under contemporary socio ontological conditions. The perfect illustration of the conundrum is given by Thrift himself in his 1983 paper when referring to social and cultural processes 'in Bali', rather than characterizing them as regionalizing processes in the course of which something called 'Bali' is named, made meaningful, and acted upon or toward. In effect, thereby, Thrift claims that his statements are true for all people living on the island, and so the socially homogenizing effect of talking about social issues in spatial categories becomes obvious. The effect becomes still more problematic with the increasing importance of contemporary processes of disembedding, the increasing scope of individual decisions, and the increasing choices of individual life forms. If these processes are taken into account, life forms are no longer spatially separated in a socially homogenous way and, as a result, the appropriateness of describing sociocultural issues in spatial categories dramatically decreases. Hence, one of the core problems of Thrift's contextual regional geography is the explicit recourse to the strengths of possibilistic or Vidalian traditional regional geography, with its prioritizing of life ways lived *in* given regional units. Of course, this emphasis is inherent in the general run of the new regional geography against the spatial approach, which was criticized for its nomothetic orientation insensitive to regional variability, but the emphasis upon spatial context – upon being *in* regions – within Vidalian geography has gained special importance for Thrift's critique (and in several papers he explicitly re turns to his Vidalian influences, albeit refracted through a reading of Marx).

Nevertheless, this criticism does not go far enough: instead of revealing space laws, the aim is now to discover spatial uniqueness, an aim that is already inconsistent with a contextual regional geography insofar as the latter does not simply wish to revert to its Vidalian and equivalent forebears. More importantly, though, and as Werlen ar gued in 1997, a space centered geography of context *cannot* be a contextual geography. A contextual geography that is worth the name requires a scientific analysis of everyday 'geography making' as a mix of processes constitutive of what become lived, by subjects, as their routine socio cultural geographical realities. This analysis is impossible with a regional geography that is bound to the Vidalian roots. Despite the call for a concentration on the subject, moreover, Thrift fails to resist the tendency to spatialize the nonspatial. In the end, a contextual geography with a space centered research perspective is and stays a con tradiction because, logically, there cannot be a spatially defined context (a preexistent, predeterminant, and pre determining region) prior to human *praxis*. 'Context' is itself a relational concept, and as such can only be prop erly specified with regard to action and agency. Depending on the type of acting, different aspects of context become relevant, and that is the reason why a 'true' contextual geography always requires an action referenced, subject centered perspective.

According to Thrift, spatial structures are expressions of social structures mediated by action. For the very reason of spatial structures being 'expressions', however, they cannot be equated with social structures: which means that spatial structures are not social structures *per se*. They are indeed expressions, and may contribute to a persistence of social structures; but, as structuration theory shows, it is power relations (which must be con ceived first as social relations) that determine how spatial structures can be used. Instead of spatializing the social, therefore, it is more promising to analyze how spatially bound or place referenced symbols and markings are socially used. Then, however, it is not history or places that constitute persons but rather, and crucially, the knowledge about moments or points in time and space. This knowledge can only be obtained and used by other subjects, not by 'history' or 'place' themselves: it is the mutual knowledge of peoples together, a population co hort, that gives shape to a sense of history or place, call it a region if you will, that can subsequently become con stitutive of social action. In accepting this argument, re gional analysis as a means of investigating the relation between 'place' and 'subject' has to be given up; the an alysis should rather focus on the acting of subjects. This implies a change from a 'theory of context' to a 'theory of geographical practice'; or in other words, we need a change from 'regional geography' to the geographical exploration of 'everyday regionalization' processes.

See also: Anthropogeography (After Ratzel); Determinism/Environmental Determinism; Regional Geography II; Structuration Theory; Structurationist Geography; Time Geography.

Further Reading

Gregory, D. (1978). *Ideology, Science and Human Geography*. London: Hutchinson.

Gregory, D. (1982). *Regional Transformation and Industrial Revolution: A Geography of the Yorkshire Woollen Industry*. London: Macmillan.

Hartke, W. (1956). Die 'Sozialbrache' als Phanomen der geographischen Differenzierung der Landschaft. *Erdkunde* 10(4), 257 269.

Helms, G. (2001). Glasgow The friendly city, the safe city: An agency orientated enquiry into the practices of place marketing, safety and social inclusion. *Praxis Kultur Und SozialGeographie* 23, Universitat Potsdam.

Pred, A. (1981). Social reproduction and the time geography of everyday life. *Geografiska Annaler* 63B, 5 22.

Pred, A. (1984). Structuration, biography formation and knowledge: Observations on port growth during the late mercantile period. *Environment and Planning D. Society and Space* 2(3), 252 276.

Pred, A. (1986). *Place, Practice and Structure*. Cambridge: Polity Press.

Thrift, N. (1983). On the determination of social action in space and time. *Environment and Planning D. Society and Space* 1, 23 56.

Thrift, N. (1990). For a new regional geography 1. *Progress in Human Geography* 14(2), 272 279.

Thrift, N. (1991). For a new regional geography 2. *Progress in Human Geography* 15(4), 456 465.

Thrift, N. (1993a). For a new regional geography 3. *Progress in Human Geography* 17(1), 92 100.

Thrift, N. (1996). *Spatial Formations*. London: Sage.

Werlen, B. (1988). *Gesellschaft, Handlung and Raum*. Stuttgart: Franz Steiner.

Werlen, B. (1993). *Society, Action and Space: An Alternative Human Geography*. London: Routledge. (English translation of Werlen, B., 1988, above.)

Werlen, B. (2007). *Globalisierung, Region und Regionaliserung Sozialgeographie alltäglicher Regionaliserungen Vol. 2.* (2nd edn.). Stuttgart: Franz Steiner.

Regionalism

J. Tomaney, Newcastle University, Newcastle upon Tyne, UK, and Monash University, Melbourne, VIC, Australia

Glossary

Topophilia Affective bonds between a people and place or setting.

Untraded interdependencies Language, norms, customs and practices shared by firms and other actors that underpin economic action.

Introduction

In human geography the term regionalism has generally been used to refer to tendencies or practices based on the identification and/or promotion of subnational socioeconomic, cultural, or political territories. A concern with regionalism has not been confined to human geog raphy. Political scientists, sociologists, historians, and lit erary theorists – among others – have contributed to the debate on regionalism. The study of regionalism lends itself to multidisciplinary approaches. By contrast, in economics, and to some extent political science, the term has been used to refer to developments in international political economy, notably the formation of trading blocs (such as NAFTA or Mercosur) or customs and political unions (the EU). Recent debates in human geography have been concerned with the emergence of a 'new re gionalism'. Initially associated with debates about the EU, evidence of the new regionalism has been offered from virtually every continent. It speaks of the contemporary value attached to the region as a terrain of socioeconomic, cultural, and, perhaps above all, political action.

Regionalism is used to refer to efforts to organize (or reorganize) government into distinct administrative regions (as in the 'new metropolitan regionalism' in the United States, regionalization processes in contemporary Europe, or the debate about *doshu sei* in Japan), or it can signify the advocacy of such principles. Regionalism can describe an attachment to a particular region. Or, in cultural terms, it can indicate an emphasis on the regional characteristics of a setting, often using dialect, as in the 'regional novel', which is a subgenre of many national literatures.

If regions are social constructions, regionalism is the means by which they are constructed. Regionalism is a form of social, cultural, and political identification. It is associated with the production or protection of an identity and a consciousness attached to a particular territory within the nation state. Intellectuals, cultural producers, and political leaders contribute to its production, but re gionalism is reproduced in the practices of everyday life. Regionalism is contingent and is constantly being formed and reformed, negotiated and manipulated, promoted, and contested. Regionalism, then, is a historical phenomenon.

Regionalism presupposes a regional identity. Regional identity represents one form of collective identity, which is best understood in diachronic rather synchronic terms. The formation of regional identities is complex, contested, and contingent. Thus, regional identities are not pre given cultural phenomena, but are constituted through cultural and political institutions, social movements, and processes of governance. Regionalism may take the form of cultural activities by which a region tries to oppose the stand ardizing effects of the capital. It is a means of binding people and place and managing the relationships between the two. Interpretation of the past is a key element in the cultural representation of a region, as is the search for a 'usable past' which can guide regionalist thinking and action. Thus, regionalism is frequently associated with the revival of regional histories which involve challenges to dominant national interpretations of the past.

Regionalism, Culture, Politics

Regionalism is a 'performative discourse'. It seeks to achieve legitimacy for definitions of boundaries and to obtain approval for this definition in cultural and polit ical, and popular and official understandings. Thus, re gionalism is concerned with giving meaning to bounded material and symbolic worlds in an effort to create inter subjective meanings. Regionalism concerns process and story, more than artifact. It is concerned with em phasizing the moral particularity of a population boun ded by territory and with valorizing regional culture as a means to create regional identity. Regionalism is fre quently linked to 'topophilia', that is to the existence of affective bonds between a people and place or setting. Regionalism as a cultural phenomenon is frequently ex pressed in art, literature, music, and poetry. The *post bellum* American south has produced a vast literature which reflects and produces cultural life there most ob viously in the work of Mark Twain, Tennessee Williams, and William Faulkner. Women disproportionately con tributed to regionalist literature in the United States in the nineteenth century as evidenced by the work of Sarah Orne Jewett, Mary E. Wilkins Freeman, Alice Cary, Sui Sin Far, Zitkala Sa, Grace King, and Mary Austin. In contemporary regionalist novels such as those of the Kentucky writer Wendell Berry, the place – in this case the fictional Port William – becomes a character.

European regionalism typically has strong cultural foundations, but the political implications are sometimes ambivalent. In Ireland, the poetry of Patrick Kavanagh evokes Monaghan and an Ulster identity, in the self-conscious pursuit of 'parochialism'. But Kavanagh eschewed just about every type of politics. In England, regionalism appears in literature and poetry. The novels of Thomas Hardy bring alive an imagined Wessex, which today is used to market "Hardy Country" in Dorset. In poetry the proliferation of 'local epics' as represented in the work of Ted Hughes (*Remains of Elmet*), Geoffrey Hill (*Mercian Hymns*), or Basil Bunting (*Briggflatts*) evokes historical identities associated with distinctive regions. The architecture of Antoni Gaudí i Cornet was central to emergence of *Catalanisme* as a cultural and, ultimately, political force in the twentieth century. Gaudí contributed to the intellectual milieu which included Jacint Verdaguer i Santaló whose poem *L'Atlàntida* (1877/8) was written in Catalan and Josep Trueta who wrote *The Spirit of Catalonia* (1946) to promote its identity to the English speaking world.

Regionalist literature and art frequently show a concern with dialect and landscape. Landscape and dialect are intimately related because of the relationship of the latter to topographical nomenclature. Thus, we might say that regionalism is about the relationship between geography and biology, history, and myth. Contemporary regionalism is often closely related to a rising ecological awareness, in which attention is paid to the implications of human interaction with landscape and nature. From this viewpoint, regionalism concerns the defense of cultural distinction alongside the preservation of biodiversity. Ecological concerns and cultural concerns are frequently interlinked and can provide the foundations for political regionalism, although the links between cultural regionalism and political regionalism can be tenuous. A shared regional identity, in both its cognitive and affective dimensions, does not necessarily have political consequences, depending on whether this identity is used as a framework for perception and judgement of political issues.

Anti-Regionalism

The defense and promotion of territorial identity, especially regional identity, has been heavily criticized. Regionalists are criticized for emphasizing a topographical view of culture and politics, rather than recognizing the topological nature of postmodernity. In this view, contemporary life does not take place in territorially bounded regions, but in networks stretched through and beyond regions, although it is a viewpoint which tends to overlook the extent to which social practice, discourse, and power continues to focus on the making (and unmaking) of regional geographies, boundaries, and scales. This is the well spring of regionalism. Often implicit in views which emphasize topology is the association of regionalism with inherent conservatism and exclusionary social practices, with regionalists cast in the words of TS Eliot as "a small group of local malcontents." Regionalism is seen as being founded on primordial or ethnocultural understandings. The 'ontology of groupness' is contrasted with the progress of individualist (post)modernity. Critics maintain that regionalist practices are constructed with reference to an 'Other' in ways which are inherently exclusionary. Regionalism is criticized for offering a common geographical origin as a reason for solidarity and for neglecting the internal diversity, which can be found in any region. Regionalism is presented as legitimating regional power structures, rather than aiding regional emancipation. Escaping the constraints of the particularistic claims of territorial identities is a theme of radical commentators who advocate international solidarities and welfare politics.

Contemporary critiques of regionalism echo earlier ones. During the early period of nation building, the use of regionalism as a noun in English typically was pejorative. The *Manchester Guardian* referred to the "unfortunate 'regionalism' of Italy" (4 February 1881), while the *Edinburgh Review* referred to "The spirit of local individualism – in somewhat inharmoniously dubbed 'regionalism," (January 1887, p 107). In the twentieth century, it was suggested that innovations in communications and transport technologies would, in the words of Louis Wirth, 'undo regions'. The perceived cultural 'isolation' which gave regions their distinctiveness was neither feasible nor desirable according to this view, although in practice there is more evidence that transport improvements have enhanced comparative advantage in ways which have increased regional specialization and inequality. Moreover, normative accounts tend to overlook the way that despite the fragmentation and stretching of identities, regionalist discourses of belonging – as elements of 'scalar ontologies' – remain important. If regionalism concerns the defense and/or promotion of distinction, the question may be better posed as to how to maintain distinctive local characteristics that are palpable to both insiders and outsiders, without the damaging effects of isolation, introversion, and withdrawal. In 1948, TS Eliot asked whether "we can maintain the conception of a national culture which will draw its vitality from the cultures of its several areas, within each of which again there will be smaller units of culture having their own peculiarities." The appeal of regionalism lies in its offer of moral particularity as a means to negotiate the world of global networks. Regionalism may provide a means of being "at home in the world," a sense of being grounded in a particular society while being open to others. Regionalism represents the search for rootedness and belonging and the valorization of local distinctiveness in the face of the universalizing claims of globalization.

Regionalism Old and New

Despite the condescension toward regionalism and the power of nation building coalitions and their ideologies, states frequently accommodated regions in their political structures, albeit often reluctantly. Regional actors influenced patterns of state formation and nation building projects. The process of territorial integration was long, difficult, and incomplete. During the industrialization process, distinctive class practices acted as regional markers. The new regionalism acknowledges the reconfiguration of the nation state in the face of global economic integration, on the one hand, and the assertion of regional identities, on the other. Regionalism emerged after World War II as a reaction against processes of centralization and nationalization in economic and cultural life. The forms of regionalism which emerged in the 1960s were often based around language cleavages (Wales, Brittany, Flanders). Indeed, some such 'regions' made national and secessionist claims. The line between regionalism and nationalism is by no means always clear. In Europe, Quebec, Catalunya, and Scotland possess national identities, although they remain parts of larger states.

The 'new regionalism' was identified in Europe in the 1990s and reflected the growing cultural and political assertiveness of regional actors toward their respective national capitals. Regionalism became more widespread in the face of economic restructuring in the 1970s and 1980s. Modernization strategies of states were concerned with eliminating regional inequalities and integrating peripheral regions into national economies. Centrally directed regional policies and the regionalization of state structures were used to achieve these ends. But defining economic management in territorial terms fueled the growth of regionalism as actors mobilized around these institutions to make economic, cultural, and political demands. The new regionalism has become firmly founded on claims that economic restructuring is best managed by effective regional institutions. These arguments, though, have often been vigorously made by regions with strong regional identities. Indeed, it is claimed that such identities facilitate economic development through processes of 'untraded interdependencies' at the regional scale, although the evidence for this is less than conclusive. But the growth of regional institutions to promote economic development has encouraged claims for democratic regionalism – making regional institutions accountable to the populations over which they have jurisdiction.

New regionalism emerges when territorial relationships are difficult to manage through existing forms of political accommodation and exchange. The power and authority of the state is compromised by globalization, the advance of market forms of exchange, and changing cultural attitudes. Regionalist demands are both a response and a contribution to these trends. The ability of states to manage economies, ensure social solidarity, and promote national identities is made more difficult under these new conditions. New regionalism describes the propensity of subnational elites to respond to these developments by promoting their regions in international competition and developing their external links beyond national boundaries and seeking popular support for these strategies. Economic, cultural, and institutional developments interact dialectically. Protecting economic advantages can be the basis for political demands and cultural claims, as in the case of the Lega Nord in Italy. Elsewhere, economic disadvantage and cultural claims based on language underpin the case for political autonomy as in Galicia in Spain. Patterns of regionalism remain highly diverse in Europe, reflecting different national histories of state formation and nation building. At the same time, the existence of the European Union as a protective carapace provides a context in which regionalism can flourish. Indeed, the doctrine of 'subsidiarity' can be included in the repertoire of regionalist rhetoric and, to some extent, the EU has been stimulus to regionalist claims. This is evidenced by the growth of regionalism in the new member states where the creation of regional governments and the reassertion of regional identities has accompanied accession. In Poland, the rebirth of Upper Silesian regionalism provides a good example.

However, these developments are by no means restricted to the members of the EU. The complex mix of economic, cultural, and political factors shaping patterns of regionalism is visible in every continent. In Asia, regionalism is a notable force in Japan, South Korea, Indonesia, India, and China. The existence of regionalist tendencies in these countries cannot simply be reduced to ethnic differences, although these play a part in some cases. In Japan, regionalist claims are prominently associated with Hokkaido and Okinawa, where both regions are marked by strong and distinctive identities, including dialects and historical consciousness derived from a position on the Japanese periphery. The assertion of these identities is linked to efforts at economic development and the decentralization of political authority to the regions in ways which are not matched in other parts of Japan. The claims of Hokkaido and Okinawa are framed in the language of the new regionalism.

Regionalism has been an important force throughout modern Korean history despite the absence of linguistic and ethnic differences. Both Jeolla and Gyeongsang have marked identities and the tension between these regions has been a source of political conflict during both dictatorship and democracy. The growth of regional disparities as the Korean economy has internationalized – leading to the growing economic dominance of the Seoul city region – has both accentuated regionalist politics and required the Korean government to respond to them by decentralizing political authority to regional institutions. The language of new regionalism is prominent here too. While it is common to associate regionalism with ethnic identities in China and India, the picture in practice is more complicated. Regionalism has been a potent force in India since 1947. India has experienced a proliferation of regionalist parties. Parties which on the face of it have broader ideological ambitions, in practice, have close identification with particular regions such as the Communists in West Bengal. Some states have strong identities based on ethnic or, especially, language factors. The designation of Indian states and the agreement of the federal constitution in 1947 acknowledged language claims, but states such as Rajasthan, Andhra Pradesh, Tamil Nadu, and Maharashtra are not ethnically based, yet have strong identities, and in the case of Maharashtra have become more distinctive due to rapid economic growth. The Chinese republic saw the emergence of multiple movements for provincial autonomy in Hunan and Guandong, which were supported, among others, by the young Mao Zedong who hailed them as 'little Chinas'. Later, however, communist ideology condemned such movements as 'splitist'.

Conclusion

The new regionalism has dominated contemporary debates about the salience of the region as a focus for economic, cultural, and political action and has been a fruitful site for the investigation of contemporary social change. It provides a powerful and resonant account of regional change since the end of the 1970s and one which has appealed to policymakers, as well as academic researchers. For this reason the new regionalism can be seen as a new orthodoxy. Critics contend that the new regionalism canonizes the region as a functional space for economic planning and political governance. The region is presented as having a special efficacy in the governance of the flexible, post Fordist economy. But, according to the critics, such a viewpoint requires an overemphasis on the dynamism of an unrepresentative group of high profile regional agglomerations, where competitive advantage is embedded in civil and political networks characterized by 'institutional thickness'. Such examples do not represent the 'leading edge' of global change, but unique and irreproducible cases. Moreover, a focus on the internal characteristics of these places leads to a neglect of their position in wider economic divisions of labor and political systems. As such, the new regionalism is little more than the elaboration of a set of stylized 'success stories', which do not help us to understand broader patterns of regional change. Some critics suggest that the intellectual influence of the new regionalism owes less to its explanatory purchase and more to its instrumental appeal to regional political elites searching for narratives to justify their actions in an era of enhanced territorial competition. The simpler versions of new regionalism certainly underplay the complexities of the contemporary regionalization of business and governance, and the changing role of the state in relation to these. By fetishizing the region, we can overlook the contradictory nature of the reterritorialization of the state and the extent to which states and regions are entwined in multilevel systems of governance.

The power of regionalist narratives, however, cannot be ignored. The political and cultural construction of the contemporary regional world does include a cultural and political reawakening of regional consciousness. Regionalism, therefore, remains a key concept in human geography. Practically, it continues to shape culture and politics. Despite the power of the global language of new regionalism, regionalism remains a diverse and uneven process deeply influenced by varied patterns of state formation and nation building. As long as difference and inequality take a territorial form, regionalist claims are likely to be found somewhere in the world.

See also: New Regionalism; Region; Regional Actors; Regional Geography II.

Further Reading

Agnew, J. (2001). Regions in revolt. *Progress in Human Geography* 25(1), 103 110.
Allen, M. (2001). *Identity and Resistance in Okinawa*. Langham, MD: Rowman, Littlefield.
Amin, A. (2004). Regions unbound: Towards a new politics of place. *Geografiska Annaler* 86B(1), 33 44.
Darby, H. C. (1948). The regional geography of Thomas Hardy's Wessex. *Geographical Review* 38(3), 426 443.
Entrikin, N. (1991). *The Betweenness of Place*. Baltimore, MD: Johns Hopkins University Press.
Entrikin, N. (1999). Political community, identity and cosmopolitan place. *International Sociology* 14(3), 269 282.
Gilbert, E. W. (1957). Geography and regionalism In Taylor, G. (ed.) *Geography in the Twentieth Century*, 3rd edn. pp 345 371. London: Methuen.
Jones, M. and McLeod, G. (2004). Regional spaces, spaces of regionalism. *Transactions of the Institute of British Geographers* 29(4), 433 452.
Keating, M. (1998). *The New Regionalism in Western Europe*. Cheltenham: Edward Elgar.
MacLeod, G. (2001). New regionalism reconsidered: Globalization and the remaking of political economic space. *International Journal of Urban and Regional Research* 25(4), 804 829.
Morgan, F. W. (1939). Three aspects of regional consciousness. *Sociological Review* 31(1), 68 88.
Paasi, A. (1991). Deconstructing regions: Notes on the scales of spatial life. *Environment and Planning A* 23(2), 239 256.
Tomaney, J. (2007). Keeping a beat in the dark: Narratives of regional identity in Basil Bunting's *Briggflatts*. *Environment and Planning D: Society and Space* 25(2), 355 375.

Relevant Websites

http://www.cor.europa.eu
Committee of the Regions.
http://docsouth.unc.edu/southlit/regionalism.html
Library of the American South.

Regionalization/Zoning Systems

N. K. Shortt, University of Edinburgh, Edinburgh, UK

Glossary

Building Block Zones (BBZs) Basic input spatial units used in regionalization. The choice of BBZs is entirely up to the researcher.

European Regionalization Algorithm (ERA) A generalized version of the travel-to-work area (TTWA) algorithm. The ERA is a rule-based computerized method with the ability to produce contiguous macro-regions without the need to set a contiguity constraint within the model.

Modifiable Areal Unit Problem (MAUP) When data are aggregated to reporting zones, the results of the analysis will be affected by such a zoning system.

Synthetic Data Matrix (SDM) A synthesis of boundary sets allowing the researcher to analyze multiple input boundaries and test the strength of connectedness between the basic input building block zones.

Travel-to-Work Areas (TTWAs) Travel-to-work areas (TTWAs), used by the Department of Employment (in Great Britain) to publish unemployment statistics, arose from the need for consistently defined areas for which to compare 'like with like'. The advantages of such areas lie in the consistency of their definition and the thresholds of population and self-containment placed upon them.

Introduction

Rarely are lines drawn arbitrarily upon a map, but such lines, or boundaries, are often artificial. This artificial nature of administrative regionalization is inherently linked to space, with events and patterns of population flows being unevenly spatially distributed. Researchers within the 'regionalization/zone design' field have attempted to design regions which are not restricted by the boundaries imposed through such formal administrative division. Regionalization can thus be described as the process of demarcating boundaries for any purpose, including administration, education, healthcare, and electoral division.

Territorial division of space within any regionalization attempt is based on a desire to maximize the criteria set by the researcher in order to optimize the commonality of features within regions. This field of regional definition (also known as 'zone definition') came to the fore during the quantitative revolution in the 1960s and since then there has been a dependence upon a hierarchy of administrative units in analyzing spatial data. Renewed interest in zone definition has acknowledged that the problem of using contrasting administrative units defined for other purposes, and at different scales to report geographical phenomenon lies in the fact that these areas are not comparable and comparison between them should only be carried out for areas whose boundaries have been consistently defined. This problem exists because of the lack of other spatially defined units reflecting the 'real world', for the purpose of the analysis being carried out and the availability of data (predominantly census data) at the administrative level. There is a serious danger of procuring misleading and inaccurate results while using units which have not been designed for the purpose at hand.

Methods

Regionalization, or Zone design, can be seen as the aggregation of M small regions into N larger regions, where M is always greater than N. When defining regions the researcher is concerned with minimizing between region differences and defining regions that have more connection within than with other surrounding regions. It is measuring this interaction within and between regions, and defining regions on this basis, that interests geographers working within the field of regionalization. At the outset the researcher must decide what units will be used as building block zones (BBZs). These BBZs will then be amalgamated through a selected procedure in order to create final regions that satisfy criteria that may be specific to that particular zonation/regionalization problem. These criteria vary but they typically involve one, or all, of the following – homogeneity, equality (e.g., population size), and contiguity. A homogeneity constraint would require the researcher to define regions that are as similar as possible based on a selected population characteristic, for example, socioeconomic status. Related to this would be a measure of equality, most often population size, which would require areas to be as similar as possible to one another in this respect. The contiguity constraint, while not applied in all methods, is very common and, when used, rules that only neighboring BBZs should be merged to form regions. No single procedure may satisfy all criteria, so a degree of compromise is reached with the researcher often setting both minimum and target thresholds for each criterion.

The problem of regionalization can be summarized in **Figure 1**. Here we see eight spatial units, or BBZs, which

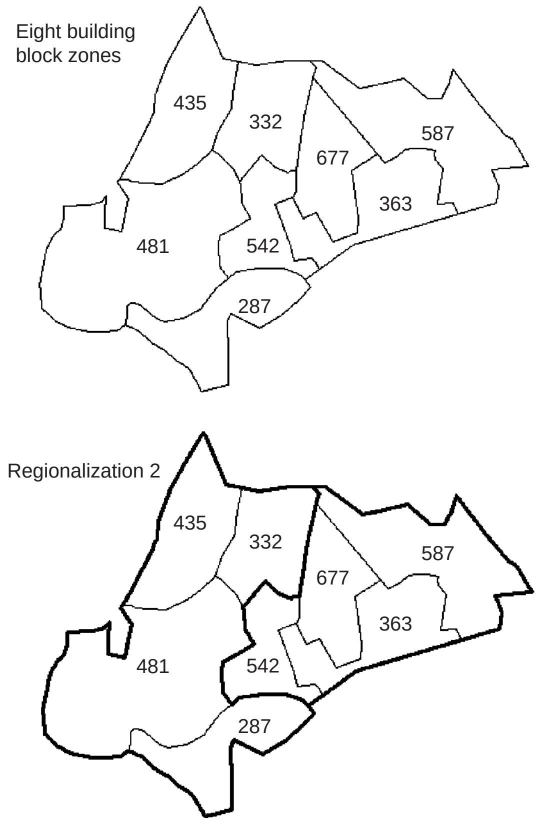

Figure 1 The regionalization problem.

we would like to combine to create just two regions. Within each of these we can also see the population of each BBZ, with an overall population of 3704. Our criteria could be to maximize population homogeneity, create two final output regions, and ensure spatial contiguity. Any solution should therefore aim to achieve two regions with similar sized populations, in this case with populations of approximately 1800 each. In this diagram we can see three possible iterations. In regionalization 1, our populations are 1745 and 1959; in regionalization 2, they are 1535 and 2169; and in regionalization 3, 1925 and 1779. Further regionalizations could be found by swapping the BBZs and removing the contiguity constraint.

Various methodologies have been applied to region alization with no ‘ultimate’ regionalization technique available. Using a ‘clustering procedure’, the researcher applies a methodology that groups individual building blocks into regions in a single process. Groups of BBZs are clustered around central units with constraints placed upon such factors as population size and connectivity. Readily available clustering methodologies are generally used (such as k means clustering). The automated zoning procedure (AZP) developed by Openshaw in 1977 is a typical example of this approach. Originally designed to examine the effects of the modifiable area unit problem (MAUP), the AZP has since developed to form the basis of Openshaw and Rao’s zone design (ZDES) system and the foundation for the design of the 2001 UK census output areas. These output areas are the base unit for census data release in the UK with a minimum size cri terion of 40 households but a target of 125. The starting point behind the AZP algorithm lay in the simple equation that $M > N$ in that M (a number of small zones) would be allocated to N (an output region). The key aims being to allocate all BBZs (M) to one and only one region (N), but all BBZs must be adjoining to pass the criteria of a contiguity constraint and be allocated to that region. Further criteria include a compactness of zone shape and that the researcher determines the number of regions required at the outset. This algorithm was enhanced successfully by Dave Martin in developing the 2001 census output areas for England and Wales. Thiessen polygons were constructed around the ordnance survey’s individual address point data, these were then dissolved to form unit postcode polygon building blocks and con strained to fall within wards and parishes (both ad ministrative subdivisions in England, in the analysis communities replaced parishes in Wales) while re specting features such as major roads and waterways. Constraints placed upon the algorithm included bound ary length, internal homogeneity (based on household tenure and dwelling type), and population size.

A further group of procedures involve hierarchical methods in which regions are built upon previous groupings of BBZs until certain criterion are satisfied. In the majority a contiguity constraint is applied so BBZs adjacent to one another are initially grouped. These regions are then tested based on the identified criteria and if they satisfy the regionalization is complete. Although results arising from these procedures are easily interpreted, limitations are recognized stemming from sample size, aggregation methods, and the initial spatial units considered for input. Researchers have highlighted that the strict inclusion of the contiguity constraint causes problems when attempting to satisfy further criteria, such as homogeneity, as the resulting regions can be hindered leading to inadequacy in the results. The initial groupings of the data are obviously critical.

Perhaps the most popular group of methods are those that employ a 'rule based approach' with the number of regions left unspecified throughout the analysis and BBZs allocated to core areas with those remaining unallocated treated as regions in their own right. At the outset, each area being used as a BBZ has an equal chance of becoming a region. Such a rule based method was used toward the definition of metropolitan areas in the United States as well as both the 'daily urban systems' and the 'travel to work areas' in the UK (TTWAs). The TTWA methodology is seen as a multistage rule based method that has the advantage of being able to produce contiguous macro regions without having to define an explicit contiguity constraint. The algorithm relies on the distance decay function and the strength of the interaction between areas rather than a direct constraint. The strength of this can be seen in the results of the TTWA analysis with only 0.5% of all BBZs not allocated to contiguous TTWAs. Multiple stages in the algorithm allocate zones to core areas and after each stage of the process the matrix of origin destination flows is updated to reflect the linkages in the previous stage. In doing this if regions i and j are linked in stage one, then all flows between ij are examined in stage two.

While all these methods have created regions that are fit for a particular purpose, a further problem arises when we are trying to reflect the multidimensional nature of regions. As all 'regions' are socially constructed based on a multitude of daily activities, the use of just one dataset, or boundary group, is limiting and a methodology allowing a simultaneous analysis of numerous boundaries is required. Various methodologies have been explored in the literature and one, developed by Mike Coombes, involves the creation of a 'synthetic data matrix', which is then analyzed using the European regionalization algorithm (ERA) (a variant of the travel to work areas algorithm used extensively in the UK). This procedure is comprised of four phases beginning with the collection of the data, in either flow or boundary form. In phase two, the flow data are transformed into a set of boundaries through the application of individual regionalization methods. Binary matrices of each boundary set are created that identify each pair of BBZs and whether these were grouped in each region or not. This ability of the ERA to analyze one dataset using numerous methodologies highlights the benefits of this approach. The third phase of the process involves the creation of the synthetic data matrix (SDM) through a summation of the individual matrices created in phase two. Finally, the ERA is applied to the SDM and a set of proto regions are created. Like the TTWA algorithm, the ERA is a rule based computerized method with the ability to produce contiguous macro regions without the need to set a contiguity constraint within the model. The method is also restricted by criteria set by the researcher that concentrate on self containment and population thresholds.

This process of regionalization (zone design) relates strongly to the MAUP. The choice of zoning system inevitably affects the results of both the resulting region boundaries and the ability to infer spatial relationships within and between such regions. As spatial units used in any analysis are generally already an aggregation of data (such as point data to polygon areas), any form of reaggregation is greatly affected by the MAUP. The MAUP has two effects, firstly, a scale effect that is related to the level of aggregation and secondly, the zoning effect that is related to the definition or partition of space. Related to this is the issue of the ecological fallacy, and in the case of regionalization when we analyze features of grouped areas (regions), we are subject to a degree of spatial representation error about the assumptions we can make regarding the individuals in our defined areas. While criteria such as homogeneity can help to reduce the level, all spatial units will be subject to this error. Although these fundamental problems are recognized, and have been examined extensively in prior research, automated regionalization procedures attempt to reduce these effects.

An Emerging Field?

Perhaps the downside of all the regionalization techniques mentioned is the level of technical expertise required to apply the algorithms and set the criteria. From an infinitely more user friendly perspective, many GIS systems now have built in redistricting wizards. ESRI's districting extension (available for both ArcGIS and ArcView) states that it will allow the user to redraw political boundaries, define schools and fire districts, and set up sales territories. When using the extension users must interactively assign BBZs to regions. Mapinfo also have a redistricting feature that allows users to run 'what if' scenarios and manually move BBZs between regions to maximize specific criteria. This availability of tech nology, while simplifying the regionalization process, also removes the ‘expert’ and as such we need to consider the possible problems. Maps, or in this case boundary sets, reflect the needs of any user and as such are value laden entities. We could term the emergence and use of these new technologies as a form of ‘everyday gerrymandering’ that neglects the importance of the researcher’s judgment and expertise, and the need for a degree of manual ma nipulation of the end product based on local knowledge. On the other hand, combined with the emergence of public participation GIS (PPGIS), the redistricting ex tensions available will enable communities to embrace spatial technology that can lead to boundary redefinition and indeed increased opposition to politically defined boundaries.

The use of detailed algorithms, computer programs, and automated procedures is not however value free. A plethora of research in the regionalization field has at tempted to define ‘communities’ or ‘localities’, but few have asked whether these output regions have any real meaning for their residents. Perhaps this is related to the scale at which we work and the subjective selection of the initial datasets with the application of the methodology being clearly dependent upon the input data. The choice and retrieval of primary datasets is crucial and must be made relevant to the application at hand, for example, what type of data should be used when creating school catchments? Further to this, the use of administrative units neglects the reality of flows and networks that are representative of everyday life. In an effort to ensure that automated regions reflect, to some degree, peoples’ levels of social contact and sense of place, alternative region alization methods could combine both a quantitative modeling approach and a more in depth exploratory qualitative approach. Such a qualitative dimension would attempt to capture local knowledge and perceptions of daily interactions and spaces while involving interactive participation within the research. The emergence of participatory GIS techniques could be used in novel approaches to regionalization and zone design. Further more, developments in GIS technology and the geor eferencing of address based data enable creation of zones based on detailed point data rather than arbitrarily de fined building blocks. Future research in this area should be compelled to recognize the enormous potential of the data rich resources now available. In any method of re gionalization we should concern ourselves with the an alysis of the most up to date and relevant interaction, boundary and qualitative data – that allows areas to set themselves aside as easily identifiable functional regions.

See also: Modifiable Areal Unit Problem; Neighborhood Effects; Thiessen Polygon.

Further Reading

Cockings, S. and Martin, D. (2005). Zone design for environment and health studies using pre aggregated data. *Social Science and Medicine* 60, 2729 2742.
Coombes, M. and Openshaw, S. (2001). Contrasting approaches to identifying ‘localities’ for research and public administration. In Frank, A., Raper, J. & Cheylan, J. P. (eds.) *Life and Motion of Socio Economic Units*, pp 301 315. London: Taylor and Francis.
Duque, J. C., Ramos, R. and Surinach, J. (2007). Supervised regionalisation methods: A survey. *International Regional Science Review* 30, 195 220.
Haynes, R., Daras, K., Reading, R. and Jones, A. (2007). Modifiable neighbourhood units, zone design and residents’ perceptions. *Health and Place* 13, 812 825.
Horn, M. E. T. (1995). Solution techniques for large regional partitioning problems. *Geographical Analysis* 27, 230 248.
Martin, D. (1998). Optimizing census geography: The separation of collection and output geographies. *International Journal of Geographical Information Science* 12, 673 685.
Martin, D., Nolan, A. and Tranmer, M. (2001). The application of zone design methodology to the 2001 UK Census. *Environment and Planning A* 33, 1949 1962.
Openshaw, S. (1984). *The Modifiable Areal Unit Problem: Concepts and Techniques in Modern Geography, Geo Abstracts*. Norwich: Geo Abstracts.
Openshaw, S. and Rao, L. (1995). Algorithms for reengineering 1991 census geography. *Environment and Planning, A* 27, 425 446.
Shortt, N. K., Moore, A. J., Coombes, M. and Wymer, C. (2005). Defining regions for locality health care planning: A multidimensional approach. *Social Science and Medicine* 60, 2715 2727.

Relevant Websites

http://www.esri.com/software/arcview/extensions/districting/index.html
Districting for ArcGIS.
http://www.ppgis.net/
Open Forum on Participatory Geographic Information Systems and Technology.
http://www.geog.leeds.ac.uk/software/zdes/
The Zone Design System.

Regression, Linear and Nonlinear

J. Pearce, University of Canterbury, Christchurch, New Zealand

Glossary

Dependent Variable The characteristic or event of interest (also known as *y*-variable, outcome variable, and response variable).
Independent Variable The factors that may influence the size and occurrence of the dependent variable (also known as *x*-variable and explanatory variable).
Intercept The value of the dependent variable at that point where the line crosses the *y*-axis (i.e., where x 0).
Least Squares Method Technique that finds the equation of the line that best fits the points on a scatter plot.
Linear Regression A technique that finds a line that fits the scatter of data points to provide for any given value of *x* the best estimate of the corresponding value of *y*.
Multiple Regression An extension of simple linear regression to simultaneously include more than one independent variable.
Regression Equation The mechanism for estimating a dependent value (*y*) from the respective explanatory value(s) (*x*).
Residuals The difference between the observed and the predicted value.
R-Squared Value The proportion of the variation in the dependent variable (*y*) that can be explained by the variation in the independent variable(s) (*x*).
Slope The change in *y* per unit change in *x*.

Introduction

Regression is a statistical technique that was adopted in human geography during the quantification of the subject in the mid 1950s and early 1960s. Regression analysis is used in identifying the relationship between a dependent variable (the *y* variable) and one or more independent (or explanatory) variables (the *x* variable). The technique allows geographers to understand and characterize the nature of the relationship between two or more variables of interest. One of the earliest studies in human geography to adopt regression methods was the work of Harold McCarty which provided a geographical analysis of Senator Joseph McCarthy's Wisconsin vote. Throughout the 1960s, quantitative human geography (including methods such as regression analysis) became integral to the discipline especially in North America and Britain, transforming geographical practice.

The technique can be thought of as a conceptual extension of correlation. Correlation describes the strength of linear association between two variables and is completely symmetrical (the correlation between *A* and *B* is the same as the correlation between *B* and *A*). While correlation provides the strength of a relationship between two variables, sometimes we also need to know exactly how one variable varies with respect to another and answer the question 'given *x*, how well can we predict *y*?'. Therefore, regression analysis is often more useful than correlation because it allows us to predict *y* from *x* and provides us with an insightful summary of the relationship between the two variables.

Regression analysis is a useful tool for a researcher who wishes to examine whether a particular outcome (the dependent variable) is in some way linked to variations in another phenomenon (the independent or explanatory variable). For example, a researcher interested in the geography of education may wish to know the extent to which the exam performances of the schools within a district are explained by the socioeconomic circumstances of the school catchment areas. In the researcher's regression analysis the dependent variable could be the average exam score of the pupils in each school and the explanatory variable a measure of social deprivation of the school catchment areas. Similarly, a health geographer may be interested in the relationship between neighborhood air pollution and lung cancer rates in the local community. In this example, the lung cancer rate of each neighborhood could be the dependent variable and a measure of local ambient pollution the explanatory variable. By using regression analysis we would be able to characterize the relationship between our variables of interest and identify the extent to which school performance and lung cancer rates altered with social deprivation and air pollution respectively.

Undertaking Regression Analysis

In the simplest example (one explanatory variable), we want to identify how one variable varies with respect to another. A useful starting point for understanding the relationship between two variables is to draw a scatter plot between them (see **Figure 1**). Unless there is a perfect fit between the two variables (which is very unusual), there will be a degree of scatter. Regression analysis constructs the best line of fit between the two

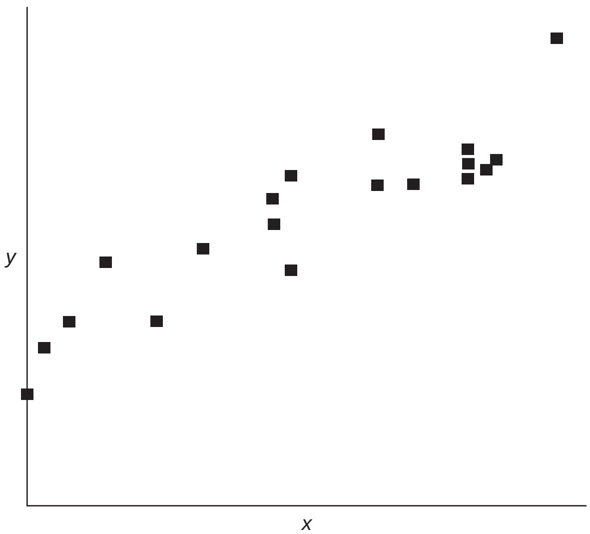

Figure 1 Example of scatter plot.

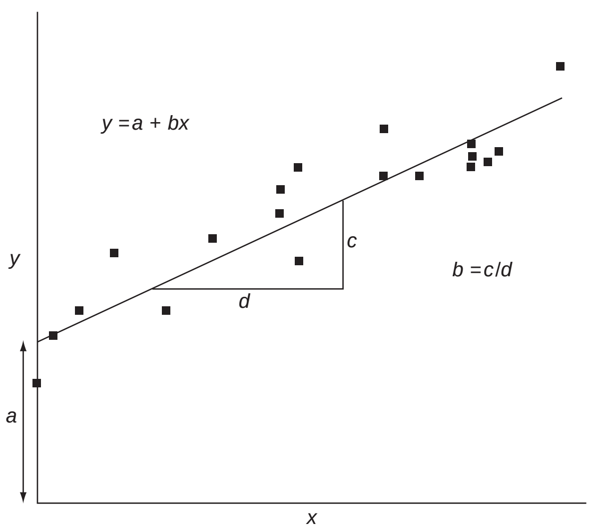

Figure 2 The intercept and slope of the regression line of the regression equation $y = a + bx$. The intercept (a) is the point where the line crosses the y axis (i.e., value of y where $x = 0$). The slope (b) is the increase in y corresponding to a unit increase in x.

variables and is the line through the scatter of points that fit the points as best as possible (**Figure 2**). The line of best fit allows us to quantify the association between the two variables. The direction in which the line slopes depends on whether the correlation is positive or negative. When the two variables increase or decrease together (positive), the line slopes upwards from left to right (as shown in **Figure 2**); when one variable decreases as the other increases (negative), the line slopes downwards from left to right.

The relationships between the two variables shown diagrammatically in **Figure 2** can also be usefully

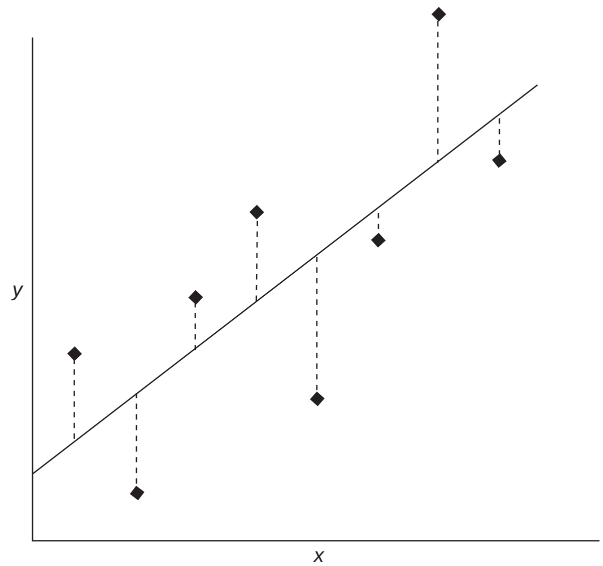

Figure 3 Linear regression line fitted by least squares.

represented in

$$y = a + bx$$

where y is the dependent variable, a is the intercept (where the regression line crosses the vertical axis), b is the slope of the line (often called the regression coefficient), and x is the independent (or explanatory) variable.

These values tell us much about the relationship between the two variables and help us to draw conclusions from our analysis. In particular, the intercept (a) tells us what value of y would be obtained based on the regression line if x were 0. The b coefficient provides us with a measure of how much change we observe in the dependent variable (y) with a unit change in the independent variable (x). In this way it represents the degree to which the line slopes upwards or downwards.

The line of best fit (the regression line) is most commonly calculated using least squares linear regression which minimizes the squared deviation between the data points and the regression line along the y axis (**Figure 3**). The line of best fit (b) can be calculated using an equation where the numerator is the covariation of x and y, and the denominator is the variation of x:

$$b = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2}$$

where $\bar{x}$ and $\bar{y}$ refer to the mean of x and y, respectively. Once we have worked out the value of b, it is easy to calculate the intercept (a):

$$a = \bar{y} - b\bar{x}$$

Residuals

From the regression equation it is possible to work out a prediction (or estimate) of y for any value of x. In practice, it is unlikely that the predicted value of the dependent variable will be exactly the same as the observed value. The difference between the observed and the predicted value is known as the residual. Therefore, the residual value is the vertical distance along the y axis from an individual point to the regression line $(y_i - \hat{y}_i)$. If an observation lies on the regression line then its residual value is 0. The residuals represent unexplained or residual variation after the fitting of a regression model. The residual values can be positive (observed is greater than expected) or negative (observed is less than expected) and, obviously, the residual values will tend to be small if the regression fits well and large if it does not.

Residual values can be calculated in two ways. First the absolute residual value is the difference between the observed Y value and the predicted value. Although this measurement is useful for identifying areas that do not fit the regression trend, they are often difficult to interpret. Residuals are expressed in measurement units of Y, so a large residual value does not necessarily indicate an important observation. To overcome this problem, one can calculate the second type of residuals – relative or standardized residuals. This approach has the advantage of indicating the magnitude of a particular residual in relation to the standard deviation of residuals. Standardization puts the residuals on a common scale of measurement, in units of standard deviations. Therefore, in tandem, the two approaches provide a useful way to interpret the spatial variation of the residuals.

An examination of the residual values is useful for two reasons. First, the residual values can be used as a basis to provide a goodness of fit measure of a regression model (e.g., R^2 or standard error of the estimate). Second, the residuals allow the researcher to test whether some key assumptions of regression analysis have been met. These two issues are considered in turn.

Goodness of Fit

It is often useful to measure the level of explanation of the dependent variable by the independent variable(s) that our regression model provides. The goodness of fit of our regression models is normally expressed using the 'R squared statistic' (R^2). This value is calculated by comparing the ratio of the sum of squared deviations explained by the regression (RSS) to the total sum of square deviations (TSS). **Figure 4** shows how the total sum of squares can be partitioned into its explained and unexplained components. The resulting R^2 value is known as the coefficient of determination. The R^2 statistic is summarized as

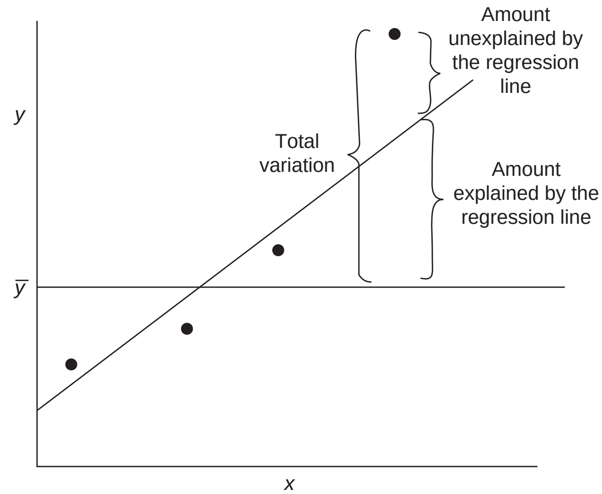

Figure 4 Partitioning the variability in the dependent variable (y) into the explained and unexplained components.

$$R^2 = \frac{\text{total variance} - \text{residual variance}}{\text{total variance}}$$

or more specifically,

$$R^2 = \frac{\sum (y_i - \bar{y})^2 - \sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

where $\hat{y}_i$ = the corresponding values of y_i shown on the regression line for any chosen value of x_i.

When our regression model is a good description of the relationship between the independent and dependent variables, the two terms will be similar in size and the R^2 will be high (close to 1). If our model is a poor description of the data then the RSS will be small and hence the R^2 will be low (close to 0). For example, an R^2 value of 0.055 indicates that only 5.5% of the variance is explained by our regression model. Sometimes the R^2 value is multiplied by 100 to provide a percentage figure instead.

Assumptions

Regression analysis, like other types of general linear models, relies upon six assumptions. These assumptions are:

1. Linearity: there is a linear relationship between the explanatory variable and the response variable (which can be transformed if necessary).
2. Normality: the residuals (i.e., the values of $(y_i - \hat{y}_i)$ for every value of x) are normally distributed. This assumption is required for significance testing of regression coefficients. The simplest way to check normality is to plot a histogram of the residual values

and compare it with a normal curve. There are also statistical tests for normality. If this assumption is not met then a transformation of the data may be necessary.

3. Mean of residuals equals zero: for every value of x the mean of the y residuals must be 0 (the mean of $(y_i - \hat{y}_i) = 0$). This can be checked by plotting the residuals against the predicted value of y. The points should be equally scattered about the line such that the value of the residual is approximately 0 for all predicted values of y.
4. Homoscedasticity: this assumption means that the variance of the residuals is the same for all values of the independent variable (x). This can be checked visually by plotting the residuals against the predicted values of y, and there are several statistical tests for homoscedasticity.
5. Autocorrelation: the residuals should be independent of each other. Therefore, one residual value should not be correlated with another. If the residuals are not independent then they are said to be autocorrelated. There are two forms of autocorrelation: temporal and spatial.
   (a) Temporal autocorrelation is when the residual at one point in time is correlated with that at an earlier (or later) time, and it exists in most time series analyses. For example, current retail prices strongly influence future retail prices as well as being influenced by earlier prices.
   (b) Spatial autocorrelation is particularly important in geographical studies and occurs where residuals at one place are correlated with those in nearby places. The residuals show a spatial pattern. Geographers have devised statistical tests for spatial autocorrelation in regression residuals, and if autocorrelation is detected, various spatial regression techniques can be used to incorporate spatial effects.
6. Lack of measurement error: all of the errors should be measured without error otherwise the regression coefficients may be biased.

When we are undertaking regression analysis it is important to consider each of these assumptions because if they are not met for the dataset being analyzed, then the model may be misspecified and the coefficients are likely to be inefficient and/or biased. In such cases, inappropriate conclusions may be drawn from the results.

Tests of Significance

The R^2 value provides us with details about the overall model fit. However, often we want to look closely at the individual variables in the regression model to see whether they are statistically significant or not. In other words, we test whether the true population value of the coefficient is significantly different from 0 or some other target value. We are, of course, usually interested in the prospect of rejecting the null hypothesis, thereby gathering evidence that the independent variable (x) is important in understanding the dependent variable (y). If we accept the null hypothesis then we are suggesting that x is not a good predictor of y, since the value of x has no association with the value of y.

There are two common and interrelated approaches to testing the significance of an independent variable in a regression model. First, t tests can be undertaken to derive p values which represent the probability of obtaining a given regression coefficient (b) if the true value is 0 (i.e., a value of 0 indicates no relationship between the independent and dependent variables). If this probability is sufficiently low, then the difference between the parameter and the statistic is often then described as statistically significant. Although a rather arbitrary choice, levels of 0.05 and 0.01 are most commonly used to reject the null hypothesis.

The second approach to testing the significance of an independent variable is to depict confidence intervals which occur on either side of the regression coefficient. The confidence intervals provide an indication of the reliability of the estimates of the dependent variable using the regression equation. The confidence intervals are calculated from the standard error of the residuals of each explanatory variable. Commonly, 95% or 99% intervals are used but there is no reason why other intervals should not be used instead. For 95% confidence intervals, for example, 95% of the sample regression coefficients can be said to lie within two standard errors of the observed value. An inspection of the confidence intervals provides an indication as to whether the independent variable is statistically significant or not. If the confidence intervals include zero then we can accept the null hypothesis and conclude that the x variable is not significant and unlikely to be associated with the phenomenon under investigation.

Multiple Regression

Regression analysis is not limited to examining just one independent variable. We can simultaneously consider a number of independent variables to inspect how they influence our dependant variable. This is an approach known as multiple regression (or multivariate regression). A multiple regression equation has the general form:

$$y_i = a + b_1 x_{1i} + b_2 x_{2i} + \cdots + e_i$$

In the above equation each value of b_j (this time called the partial regression coefficient) indicates the change in the value of y with a unit change in the value of the corresponding explanatory variable (x_j), assuming no change in the values of the other explanatory variables (i.e., they are held constant). In **Figure 2**, we saw how the data points between one independent and one dependent variable could be viewed. In multiple regression with two independent variables and one dependent variable, the data points must now be viewed in a three dimensional space with two dimensions for the two independent variable and one dimension for the dependent variable.

In multiple regression, it is assumed that the corre lation between the independent (x) variables is not high. If there is correlation between the independent variables then this effect is known as multicollinearity. If multi collinearity is high then the standard errors of the re gression coefficients become inflated, and the coefficients become very sensitive to the addition or deletion of a small number of variables and/or observations. In dependent variables that are not significant may appear to be statistically significant, and vice versa, and it is difficult for the researcher to isolate the true effects of the independent variables on the dependent variable. The general solution to the problem of multicollinearity is to drop one of the two correlated variables from the model or to use a method like factor analysis to create a summary index based on the correlated independent variables.

What is the utility of multiple regression? Sometimes we want to understand the effects of one variable in dependent of other factors. This can be particularly useful when we want to adjust or control for the effects of a confounding variable. For example, medical geog raphers have examined the effect of residential exposure to radon gas in explaining the incidence of lung cancer. It has been suggested that there are specific areas that have a particular type of underlying geology that can result in the build up in radon gas in people's homes which is a risk factor in contracting lung cancer. If we were to examine the relationship between radon and lung cancer in a bivariate regression model then this would be un satisfactory because we hadn't taken into account (or controlled for) all of the other known causes of the dis ease (particularly age and smoking behavior). This ex ample highlights a key advantage of multiple regression in that we are able to control for the effects of other related variables or what is known as confounding. Our regression analysis could be represented by the following equation:

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + e$$

where y is the rate of lung cancer in each area, a is the intercept, b_j is the regression coefficient (slope) for vari able x_j, x_1 is the proportion of the population aged over 60 in each area, x_2 is the proportion of the population who are smokers in each area, and x_3 is the mean resi dential radon level in each area.

By simultaneously including more than one variable into our regression model, we can control for the effects of confounders. In the example of radon and lung cancer, we may include the lung cancer rate of each area as our dependent variable and then add as independent variable, measures like the proportion of people aged over 60, smoking behavior, and finally radon levels. This approach enables us to examine the association between radon and lung cancer having controlled for the other important confounders. The values b_1, b_2, and b_3 in the above equation provide estimates of the effects of age, smoking, and radon while each of the other two independent variables are controlled.

Transforming the General Linear Model

So far we have considered simple regression models where the dependent variable is not constrained to a specific range of values (i.e., the dependent variable (y) is a numerical value). However, in some geographical ap plications it may be appropriate to constrain the dependent variable. For example, if the dependent vari able is a binary yes/no response, such as the occurrence (or not) of a disease (e.g., measles), then we would want to assign values of 0 or 1. Similarly if our dependent vari able is a count (e.g., the number of migrants or number of deaths in a neighborhood) then we must confine our dependent variable to non negative integers.

When we want to constrain the dependent variable in some way (such as binary or count outcomes), we model a 'transformation' of the dependent variable, which allows us to confine the regression line. These regression models are related to the outcome via a 'link function' such as the logit function or logarithmic function. In the case of a binary response variable (e.g., measles/no measles), the most common transformation used is the log of the odds of the outcome. This type of regression model is referred to as logistic regression. Logistic regression is useful for situations in which one wants to be able to predict the presence or absence of a characteristic or outcome based on the values of a set of independent variables. It is similar to linear regression but is suited to models where the dependent variable is dichotomous. Logistic re gression coefficients can be used to estimate odds ratio for each of the independent variables in the model. The odds ratio is defined as the ratio of the odds of an event occurring in one group to the odds of it occurring in another group and provides a way of comparing whether the probability of a certain event is the same for the two groups. An odds ratio of one implies that the event is equally likely in both groups. An odds ratio greater than one implies that the event is more likely in the first group

Table 1 Summary of the main regression models

Type of outcome variable (y)	*Type of regression model*	*Measure of effect*	*Further information*
Continuous	Linear (simple or multiple)	Difference in means	Rogerson (2001), Rothman (2002)
Binary	Logistic	Odds ratio	Wrigley (1985)
Count	Poisson	Rate ratio	Lovett and Flowerdew (1989), O'Brien (1992)

and an odds ratio less than one implies that the event is less likely in the first group.

Similarly, when modeling count data (such as the number of migrants from one city to another or crime incidence in an area) we use an alternative form of re gression analysis called Poisson regression. In a Poisson regression model the predicted value of the dependent variable (y) for case i is the maximum likelihood estimate of the mean of a Poisson distributed variable. The natural logarithm of the estimate is equal to a linear combination of the corresponding values of the independent variables. Poisson regression is particularly appropriate when the count of cases is small (e.g., there are a large number of zero values such as for the count of cases of a rare dis ease) and hence the mean is close to 0 because the dis tribution will be strongly skewed toward the vertical axis and cannot be approximated by the normal distribution. Although a comprehensive discussion of logistic re gression, Poisson regression, and other regression models is beyond the scope of this article, the various ways in which the data can be transformed, and suggested sources of further information, are outlined in **Table 1**.

It is often the case that we have a choice of which regression model to use, depending on how we decide to express the outcome data. For example, employment status may be expressed as a numerical, binary, or cat egorical variable which means we can use linear, logis tical, or ordinal logistic regression respectively. In making our choice as to which regression model to use, it is desirable to choose the model that uses as much of the information as possible. However, it is also often sensible to use simpler models before proceeding to more com plex ones and to assess whether the gain in precision of the estimates justifies the extra complexity and potential loss of ease in interpretation.

Nonlinear Regression

One of the basic assumptions of linear regression is that there is a linear relationship between the dependent and independent variables. However, sometimes exploratory analysis prior to the calculation of regression equations demonstrates that linearity is not always a sensible as sumption. Instead, relationships can be curvilinear or have some other form of relationship. One approach often taken by geographers is to 'restore' linearity by mathematically transforming the data (logs, exponential, etc.). The researcher can consider transforming either the x or y variables, or both. The effect of transformation can be evaluated by viewing a scatterplot to identify whether the dependent and independent variables now approxi mate a linear relationship, and/or examining whether using transformed variables results in a change in the R^2 value.

However, if there is a strong theoretical reason for doing so (i.e., the process under investigation is clearly nonlinear), it may be appropriate to consider using nonlinear regression models. Although a full discussion of nonlinear regression is beyond the scope of this report, in brief, nonlinear models may better describe some processes that are inherently nonlinear such as one that increases or decreases at a declining rate. An example might be participation in sport by age which may in crease from young children to adulthood and then de cline again among older age groups. To undertake nonlinear regression, a researcher must specify a function that accurately describes the relationship between the dependent and independent variables.

Concluding Remarks

Geographers have frequently found it profitable to adopt regression methods in their research. The approach is advantageous because it tells us important information about the relationship between the dependent variable and independent variables(s) of interest and helps us to develop our theoretical understanding of a range of geographical concerns. However, it is important to re member that regression analysis can only predict how one variable varies with respect to another and cannot necessarily prove a causal relationship. A strong and significant relationship between two variables of interest may often be confounded by the effects of a third vari able. Nonetheless, if there is a cause–effect relationship that one wishes to test, regression may well be the best method to use to investigate it.

See also: Ecological Fallacy; Geographically Weighted Regression; Spatial Autocorrelation; Spatial Expansion Method; Spatial Interaction Models; Statistics, Inferential;

Statistics, Overview; Structural Equations Models; Trend Surface Models.

Further Reading

Barnes, T. J. (1998). A history of regression: Actors, networks, machines, and numbers. *Environment and Planning A* 30, 203 223.

Dale, A., Fieldhouse, E. and Holdsworth, C. (2000). *Analyzing Census Microdata*. London: Arnold.

Johnston, R. J. (1986). *Multivariate Statistical Analysis in Geography: A Primer on the General Linear Model* (1st edn.). Harlow: Longman Scientific and Technical.

Kitchin, R. and Tate, N. J. (2000). *Conducting Research in Human Geography: Theory, Methodology and Practice*. Harlow: Prentice Hall.

Lovett, A. and Flowerdew, R. (1989). Analysis of count data using Poisson regression. *Professional Geographer* 41(2), 190 198.

O'Brien, L. (1992). *Introducing Quantitative Geography: Measurement, Methods, and Generalised Linear Models*. London: Routledge.

Rogerson, P. (2001). *Statistical Methods for Geography.* London: Sage Publications.

Rothman, K. J. (2002). *Epidemiology: An Introduction*. New York: Oxford University Press.

Sirkin, R. M. (2006). *Statistics for the Social Sciences* (3rd edn.). Thousand Oaks, CA: Sage Publications.

Walford, N. (1995). *Geographical Data Analysis*. Chichester: Wiley.

Wrigley, N. (1985). *Categorical Data Analysis for Geographers and Environmental Scientists*. New York: Longman.

Regulation

G. MacLeod and A. Holden, University of Durham, Durham, UK

Glossary

Fordism It is generally viewed to refer to the set of industrial, organizational, and technological innovations introduced by Henry Ford in the original Ford factory in Detroit, in particular the introductory of a moving assembly line and the scientific management of FW Taylor introducing strict demarcations between production and workplace tasks. The term is also stretched to consider the growing articulation between mass production (enabled not least by the workplace changes described above) and mass consumption that prevailed throughout much of the twentieth century in countries of the global 'north'.

Hegemony The ability of a dominant group or state administration to exercise control through processes of consent rather than brutal force or repressive tactics. Deriving in large part from the original writings of Antonio Gramsci – an Italian Marxist and political theorist incarcerated by Mussolini's Fascist regime between 1928 and 1935 – it is used to refer to the intellectual, moral and cultural institutions, and practices through which political parties and agencies achieve and maintain a legitimate and consensual form of government.

Introduction

For some time now, the theme of regulation has been a significant object of analysis for students of political geography. In orthodox terms, regulation refers to the power and institutional capabilities of governments to regulate certain functions in the economy and civil society, particularly relating to questions of law, and more generally vis à vis the idea of a regulatory form of the state. In more recent years, the concept of regulation has become closely associated with regulation theory: a large body of research and writings, originating in Paris, but now diffused across the world. In its early formulations, research by geographers drawing on regulation theory was largely about the geography of Fordism, post Fordism, and new industrial spaces, but more recently the regulation approach has become influential within urban politics and political geography. This article aims to provide a brief review of these debates as they relate to political geography. It takes the form of three sections: the opening section provides a general introduction to regulation theory and in particular how it relates to geography; the next section examines a number of ways in which regulation theory has been deployed and reworked to analyze a range of political geographical phenomena; and the final section provides some cautionary words about deploying regulation theory to geography.

From Regulation to Mode of *Régulation*

In conventional analysis, the study of regulation in political geography tends to refer to the measurable extent to which the state at various levels – local, regional, and national – is involved in actively governing various functions relating to the economy and society. Governmental regulation in this sense relates to a number of juridical and administrative functions, such as the rules regulating corporate law vis à vis antitrust regulations, and rules pertaining to competition or employment law in relation to a worker's holiday entitlement, and so on. In such discourse, the concept of regulation can also come to imply competing normative claims about the proper role of the state in relation to the market and civil society. In particular, it can relate to the extent to which the state could or should be intervening in and regulating certain functions, such as the labor market, things like industrial sectors, and indeed people and communities, or, alternatively, that the state should be deregulated, and that such functions should be set free from tight regulation by government and the state. Such thinking became particularly significant in the 1970s and 1980s when referring to the political strategies of Reaganism and Thatcherism, both of which advocated, at least discursively, minimal forms of regulation and a general withdrawal of the regulatory state that had built up over the period since 1945.

Over the last couple of decades, however, much discussion of regulation within human geographical inquiry has become more closely associated with the principles of a body of research known as regulation theory. Regulation theory emerged in Paris in the 1970s, and was initially pioneered by three economists, Michel Aglietta, Alain Lipietz, and Robert Boyer, all based at an organization called CEPREMAP. The early work of these researchers was concerned to develop an explanation of how the general capitalist system, replete with all sorts of contradictions, not only continues to survive, but during certain periods in history also appears to assume relative economic growth allied to social stability. The pioneering work of Aglietta sought to explain the rise, success, and incipient crisis of post war American Fordism: notably the dissolution of its social foundation and the break down of the synchronic articulation between mass pro duction and mass consumption. It was not long before Aglietta and his colleagues also turned their attention to the escalating crisis of capitalism in France. CEPRE MAP's proximity to the French system of state planning also afforded these key pioneers of regulation theory a brief role as advisors to the 1981 Socialist government.

Drawing intellectual inspiration from a blend of sources – Marx, Keynesian, and Kaleckian economics, econometrics, Gramsci's political theory, and Althusser – regulation theorists looked to analyze the historical and contextual *régulation* of capitalism in a way that disavows the orthodox Marxist understanding of inevitable crisis, while simultaneously transcending the structural ist problematic of societal reproduction. In doing so, they also looked to offer a rigorous critique of an emergent and influential neoclassical economic theory, problem atizing its assumption of a delimited sphere of econo mic relations marked by a tendency toward general equilibrium.

At the very outset, it needs to be emphasized that the regulationists' particular deployment of *régulation* should not be confused with the narrower sense of juridico political and administrative rule making discussed above, which itself translates more readily into French as *regle mentation*. Regulation analysts would never wish to imply that the economy is regulated only through the state and law, and their own use of *régulation* entails a much broader analysis of both the regularities of exchange relations alongside the complementary nonmarket institutions, collective identities, value systems, and social networks that help to secure capitalism's continued survival and expanded reproduction, in spite of its conflict ridden and often crisis prone character. In short, then, and in line with some forms of institutional and evolutionary eco nomic theory, regulation theory stresses the instituted form of the economy and the deeply contested, socially embedded nature of economic activity.

In elaborating a meso level research strategy, regu lation theorists have built up a distinctive ensemble of intermediate concepts with which to examine the di versity of practices and institutions which ameliorate capitalism's economic and social contradictions. The most rudimentary of these, the 'institutional form', refers to the codification of one or several fundamental social relations. It is worth emphasizing here that regulationists would insist that such institutional forms represent the outcomes of social struggles and political interventions: in this regard, they would deny any charge of func tionalism. Within the repertoire of regulation theory, five key institutional forms emerge: the wage relation, forms of competition, monetary and financial regulation, the state and governance, and the international regime. When aggregated these assume the form of a 'mode of regulation'.

For regulation theorists, the very institutionalization of socioeconomic and political life emerges through the contingent coupling between a distinctive mode of regulation and a particular 'regime of accumulation', which represents a macroeconomically coherent and re producible relationship between production, distribution, and consumption. It is the articulation of an accumu lation regime, and its complementary mode of regulation that helps to ameliorate crisis tendencies and which may secure for a period of time, the appropriate conditions to foster a long wave of capitalist development. However, regulation theory does not perceive of such a configur ation as a linear standardized inevitable pattern, struc turally written into the landscape of capitalism. Rather, the contingently generated patterns and processes of regulation become stabilized through coevolution to the extent that accumulation–regulation couplings are chance discoveries emerging out of a plethora of social and political struggles and strategies, which in turn oc casion a specific strategic terrain upon and through which subsequent strategies will take place.

While aiming to explain its capacity for survival, regulation theory still takes very seriously the crisis tendencies of capitalism. This can be illustrated in sub stantive discussions on Fordism, which identify at least three levels of crisis: (1) microcrises, which affect those specific sectors or fractions of capital that fail to adjust to changes in patterns of technology or consumption; (2) conjunctural crises, which reflect a period when sub stantial adjustments are required to counter cyclical downturns in the economy, often manifesting as a re cession; and (3) structural crises, where – through a process of institutional exhaustion – the correspondence between an accumulation regime and regulatory mode breaks down irrevocably. Relatedly, and amid the un relenting pace and socioeconomic complexity of global ization, regulation scholars have been forced to contemplate more deeply the regulation of space and emerging spaces of regulation. As part of this, certain members of the Parisian school have examined the international, global, and subnational scales.

In turn, the influence of regulation theory has stret ched well beyond the perimeter of Paris. Commentators have identified over a dozen schools of regulation, in cluding Grenoble, Amsterdam, Scandinavia, Canada, and the United Kingdom. Furthermore, throughout the last two decades, the translation and circulation of its core ideas has seen regulation theory make a profound impact on key debates in political science, international political economy, political sociology, social policy, legal theory, development studies, economic and political geography, and urban and regional planning. These research com munities have each helped to produce subsequent generations of regulationist research with their own distinctive objects of inquiry and levels of analysis. Given the variety and wealth of regulationist analyses, we thus concur with those to advocate its status as an ongoing research program or approach, as opposed to some strictly delimited theory.

Within geography, considerable emphasis has been placed on the capacity of the regulation approach (RA) to explain how economic and social dynamics vary over space and time. Yet, while the process of regulation is indeed constituted geographically and inherently uneven, most regulationist research has, in analytical terms, tended to prioritize time over space: the theory being largely one about the periodization of capitalist development. Nonetheless, the RA does emphasize that accumulation regimes and the crisis phases inherent to them take on different forms in different countries through unique processes of social and institutional struggle. Even allowing for this – and the Parisian studies mentioned above conducted on the international scale, international comparison, and subnational space – there continue to remain something of a theoretical lacuna with regard to understanding issues of uneven development and the spatiality of regulation. This was to change during the late 1980s and 1990s.

Regulation Theory and Political Geography

Within the early formulations of the RA, the role of politics was widely perceived to be something of an afterthought. A tendency to objectivize the regulation of accumulation at the expense of explaining the *régulation* of capitalist society more generally was viewed to have propelled many approaches to prioritize economic function over political form, or to subsume political struggle, and also the character of the state within the structural forms through which the regulation of accumulation is achieved *ex post*. The 1990s witnessed something of a reaction to this productionist debate, perhaps most notably evident within certain branches of economic and political geography in the UK and Western Europe.

Much of this work was channeled toward an examination of the emerging *regulation* of post Fordist economies and political societies. The relative dissolution of the distinctively national forms of economic space associated with Fordism alongside the rise of 'governance' as opposed to 'government' have each combined to generate a growing awareness of the interarticulation of subnational, national, and supranational levels of governance, and their uneven, inconstant geographies of regulation. Such shifts have forced researchers to develop new analytical tools in order to improve our understanding of the geography of regulation. To some extent, this has involved an explicit attempt to loosen the RA's grip on the nation state as both subject and object of analysis, and to spatialize the RA theory through methodological refinement, and analyses of spatial restructuring.

In an effort to comprehend this, RA scholars within radical geography have explored the conditions that might lead to particular 'local' or 'regional modes of regulation'. However, it is important to underline that a local mode of regulation is not to be understood as a mesocosm of a national mode, thereby merely relaying a nationalist agenda. Nor is the local mode intended to indicate the domain of exclusively local regulatory practices. Rather, a local mode of regulation is a regulatory system distinctive more for its unique position within wider national and international structures of accumulation and regulation.

In light of this, we can examine how various functions of the regulatory mode, say, perhaps the training of apprentice workers, may in fact be operated locally, the overall process through which it is actually regulated will be in and through a variety of superimposing and overlapping spatial scales. All of which implies that, amid the contemporary recomposition of political economic space, the RA would profit from a multiscalar perspective. Here, the theoretical and political priority is not afforded to any specific geographical scale, such as the national or local, but rather in the processes through which particular scales are constituted and contested. In selecting these diverse objects of analysis (state restructuring, city boosterism, and regional development), but in a way that is sensitive to the spatiality of regulation, some scholars have helped to offer more judicious guidance on what the building blocks of a post Fordist social formation might look like.

The sphere of urban politics has also seen a growing interest in the RA, particularly in understanding the changing nature of local government amid the shift from one mode of regulation to another. Three common trends in the transformation of urban politics may be identified. First, amid globalization, the local political arena has become increasingly significant as a site for and generator of proactive development strategies. Second, and in line with the departure from a tightly regulatory or managerialist mode of urban government associated with the Fordist era, city authorities have increasingly endeavored to construct what is known as an entrepreneurial city, featuring an emphasis on economic growth at the expense of collective services and welfare for local citizens. Third, as part of this reinvention, a range of new actors have been coaxed into the sphere of urban politics helping to form a whole range of private and semipublic agencies typically governed through a public–private partnership. It is these latter changes that have also encouraged some to explore a potential synthesis between the RA and urban regime and growth coalition approaches. More recently, this intensifying interest among regulationist scholars on the theme of urban governance has led them to explore the geopolitical character of neoliberalism. These are, in effect, the new 'urban modes of regulation'.

Others have sought to integrate the RA with a more explicitly state theoretical approach to political inquiry. A decisive contribution has been made by the British political sociologist, Bob Jessop, whose theoretical project has become articulated with the RA through a methodological critique of Marxist theories of the state. Jessop retains a basic concern with economic determination and institutional analysis but locates this within a much more expansive framework, looking to explain the relative and provisional unity of the state–economy relation with regard to its social basis in struggles for 'hegemony'. Through his development of certain meso level concepts that derive from the RA and neo Gramscian state theory – the accumulation strategy, the state form, and the hegemonic project – Jessop examines how the cohesive unity of the state is as underdetermined at the level of the state form as accumulation is at the level of the value form. Thus, just as accumulation strategies are required to give a particular substantive unity and direction to the circuit of capital, state strategies and state projects are needed to provide (if not necessarily guarantee) a given state some measure of internal unity and strategic guidance.

While undoubtedly complex in its formulation and exposition, Jessop's theoretical project has exerted considerable influence on those interested in theorizing the political geographies of the state in the purported transition from Fordism to post Fordism or after Fordism. This is evident in work on the changing modalities of the local state and the transformation of urban politics, the purported hollowing out and changing scalar configurations of the welfare state, and debates on the nature of urban and regional governance. Some have also drawn selectively on Jessop's regulationist informed state theory with the aim of reasserting 'the political' in political economy, particularly in relation to contemporary reconfigurations of citizenship vis à vis globalization and welfare state restructuring.

Finally, the work of Canadian political sociologist, Jane Jenson, offers an important conceptual framework for understanding how political spaces are actually created and how the process itself can challenge the existing configuration of political arrangements. For Jenson, crisis conditions thereby represent moments of efflorescence in the universe of political discourse, as newly visible actors, bearing a variety of collective identities, compete to present alternative models of regulatory conduct and social organization and to engage in political conflict over the terms of a new compromise. Further, if the crisis matures, a new social paradigm is likely to accompany it and to contain some contribution from actors who have gained new representational strength in the universe of political discourse. For example, the seedbed of the UK project of Thatcherism during the mid 1970s was in large part informed by new or rehabilitated discourses of political and economic life (such as liberalism), and narrated by newly influential political actors and organizations, such as the Social Market Foundation. Jenson's own work demonstrates how, with production based politics and identities no longer mobilizing at their pre Fordist and Fordist levels, new spaces of representation have been created around the family, the city, the environment, and the community, courtesy of a variety of social movements. The emphasis on cultural and representational media also adds to our understanding of the structuration of political regimes and helps to animate the whole regulatory process.

Extending Regulation

In recent years, the RA has proved seductive to many researchers seeking innovative tools for social scientific inquiry. This concern with extending regulation has led to some notable tendencies in the nature of RA inquiry, each of which has nontrivial implications for how the regulation approach is deployed in the study of political geography. First, a growing range of scholars from a range of disciplines (though often adopting a postdisciplinary approach) have supplemented the original Parisian concern with the economy to consider a range of phenomena relating to urban governance and politics, the national and local welfare state, international geopolitics, and the politics of identity. These new objects of analysis have led to some exciting research findings and interesting research experiments. At the same time, though, it is important to keep sight of what constituted the ontological vision of the original Parisian regulationists, and to consider how far the new research concerns can be analyzed within the original framework of the RA. In this sense, it is worth highlighting that the RA is not – and indeed could never be – a total theoretical solution to any conceivable problem in political economy.

This leads us onto the second tendency, which is the fact that a growing number of scholars have looked to combine, blend, or supplement the RA with alternative perspectives and theories, including urban regime theory, critical theories of the state, discourse analysis, and theories of urban and regional restructuring. Within the context of their own particular objects of analysis, they have tried to refine the methodology of the RA by variously stressing agency and cultural identity, space and scale, and the state and politics. Suffice to say, however, that while these studies have helped to address some 'missing links' in the regulationist repertoire, they have also stretched well beyond the compass of the original RA. This, moreover, forces us to explore the practical and theoretical commensurability between any given articulation between the RA and supplementary perspectives. In turn, the extent to which the RA can still be described as a distinctive 'approach' is of course a matter for debate.

The third and final tendency emerging from the extension and elaboration of the RA relates to the question of method. It is generally acknowledged that – unlike much of Marxian orthodoxy – the regulation approach moves down several levels of abstraction to focus on the material historical, geographical, and institutional specificities of capitalist development during a particular period of history. However, the examples outlined above have highlighted this further by pushing questions of space and spatiality to the heart of a regulation approach to capitalist relations. As a meso level 'theory' of capitalist regulation, then, the RA would seem to have an elective affinity with critical realist methodological strategies and its concern to approach capitalism's causal mechanisms in more concrete and complex terms than hitherto permitted in orthodox Marxism. Understood in these terms, the RA can indeed broadly be viewed as consistent with a range of nonessentialist perspectives on social relations (e.g., discourse analysis, neo Gramsican political theory, and feminist state theory).

In conclusion, we want to open up some questions and considerations about what a geographically sensitive regulation theory would entail? At this juncture, we need to face up to the fact that if researchers want to take seriously distinct issues of geographical process and to think spatially or geographically within a regulationist framework, then they probably do need to look beyond the contours of the RA. For while the processes of regulation are indeed constituted geographically and are inherently uneven, regulationist research has tended to privilege time over space both in its search for an explanation and its representation of the periodization of capitalist development. For better or for worse, then, the regulation approach is rooted in a historicist mode of theorization: one that is very much susceptible to what has been proven to be a powerful post structuralist spatial critique of modernist (and Marxist) modes of theorizing. In acknowledging these limitations, though, we would argue that, far from abandoning the RA, it demands us to rework it in relation to alternative perspectives, including different styles of theoretical analysis (such as those from post structuralist perspectives). There is the lingering question of what in *a fortiori* theoretical terms might best serve a geographically imaginative approach to regulation? In concluding the discussion with questions of spatiality and the geographical imagination – alongside its fundamental challenge to modernist–Marxist modes of reasoning and representing change – perhaps we are nudging toward the view that the regulation approach is Marxist in its theoretical focus but post Marxist in its explanatory strategies.

See also: Capitalism; Fordism; Governance; Neoliberalism; State Theory.

Further Reading

Brenner, N. (2004). *New State Spaces: Urban Governance and the Rescaling of Statehood*. Oxford: Oxford University Press.

Clark, G. (1992). Real' regulation: The administrative state. *Environment and Planning A* 24, 615 627.

Cox, K. (2002). 'Globalization', the 'regulation approach', and the politics of scale. In Herod, A. & Wright, M. (eds.) *Geographies of Power: Placing Scale*, pp 85 115. Malden, MA: Blackwell.

Goodwin, M. and Painter, J. (1996). Local governance, the crises of Fordism and the changing geographies of regulation. *Transactions of the Institute of British Geographers* 21, 635 648.

Jenson, J. (1990). Representations in crisis: The roots of Canada's permeable Fordism. *Canadian Journal of Political Science* 23, 653 683.

Jessop, B. (1995). The regulation approach, governance and post Fordism: Alternative perspectives on economic and political change. *Economy and Society* 24, 307 333.

Jessop, B. (ed.) (2001). *Regulation Theory and the Crisis of Capitalism*, 5 vols. London: Edward Elgar.

Jones, M. (1997). Spatial selectivity of the state? The regulationist enigma and local struggles over economic governance. *Environment and Planning A* 29, 831 864.

Lauria, M. (ed.) (1997). *Reconstructing Regime Theory: Regulating Urban Politics in a Global Economy*. London: Sage.

Le Galès, P. (1998). Regulations and governance in European cities. *International Journal of Urban and Regional Research* 22, 482 506.

MacLeod, G. (1997). Globalizing Parisian thought waves: Recent advances in the study of social regulation, politics, discourse and space. *Progress in Human Geography* 21, 530 553.

Painter, J. (1995). Regulation theory, post Fordism, and urban politics. In Judge, D., Stoker, G. & Wolman, H. (eds.) *Theories of Urban Politics*, pp 276 295. London: Sage.

Peck, J. (2000). Doing regulation. In Clark, G., Gertler, M. & Feldman, M. (eds.) *A Handbook of Economic Geography*, pp 61 80. Oxford: Oxford University Press.

Purcell, M. (2002). The state, regulation, and global restructuring: Reasserting the political in political economy. *Review of International Political Economy* 9(2), 284 318.

Tickell, A. and Peck, J. (1992). Accumulation, regulation, and the geographies of post Fordism: Missing links in regulationist research. *Progress in Human Geography* 16, 190 218.

Relational Economic Geography

N. M. Rantisi, Concordia University, Montreal, QC, Canada
J. S. Boggs, Brock University, St. Catharines, ON, Canada

Glossary

Actor-Network Theory A social theory concerned with how networks form, how they solidify, and why they fall apart. There are two attributes of this approach that distinguish it from other theories about networks. First, agency is accorded to nonhuman actors. All actors (human and nonhuman alike) are viewed as symmetrical elements of a network, because their respective identities, role, and positions within a network can only be defined in terms of their relation to the elements, that is, once they are actually part of the network.

Heterarchy A system of relations characterized by a governance structure in which constituent actors occupy the same horizontal position or status.

Innovation The process by which ideas are transformed into new products or methods that have commercial application.

Institutions Sets of informal and formal rules, conventions and routines, which govern the interactions of individuals or groups and have a self-reinforcing nature.

Late Capitalism The contemporary phase of capitalism, whose origins date back to the post-war years. This phase is often associated with the globalization of financial, production, and commercial capital, or what Frederic Jameson calls "multinational capitalism," and is characterized by an accelerated turnover of technological innovations.

Post-Structuralism A group of philosophical approaches that reject universal truths and the existence of a fixed subject and that maintain that identities and meanings are arbitrary and shifting.

Tacit Knowledge Implicit knowledge that is highly personal and acquired through experience and training. Tacit knowledge is difficult to articulate with formal language, particularly as individuals are not always conscious of it, and can only be understood by other people who share the conventions, norms, and routines that arise from a common experience.

Teleological Approach An approach to the study of social or natural phenomena that views them as purposeful and oriented to a final end state. Consequently, the focus of such approaches is on outcomes rather than the processes that generate outcomes.

Situating a Relational Economic Geography Approach

Geographer Allen Scott has suggested that theoretical preoccupations in economic geography in the last half century have been shaped by empirical trends as much as prevailing academic debates. In this regard, relational economic geography is no exception. While the relational turn does not signify a school of thought, *per se*, it does mark a broad search for alternative explanations to the sociospatial organization of production. The rise of a relational approach can be viewed as a response to new organizational trends in late capitalism as well as to the limits of neoclassical and Marxist political–economic approaches within economic geography.

The Rise of New Forms of Socioeconomic Coordination

The 1980s and 1990s witnessed a period of political and economic upheaval in advanced capitalist economies, with the de regulation of markets, the loosening of trade barriers, and innovations in transport and communication technologies. These developments, coupled with shifts in consumer tastes, have altered the basis of competition for firms and regions, whereby comparative advantage is defined less in terms of resource endowment or traditional factors of production (i.e., the levels of labor and capital) and more in terms of how, and for what purpose, those factors are used. The creative use of inputs and the knowledge that underlies this creativity is viewed as a new source of advantage.

A shift in the basis of competition has, in turn, translated into a shift in the organization of economic activities. For much of the post war period, the vertical integration of interrelated activities within firm boundaries was viewed as a means for controlling and coordinating functions that were deemed essential for supplying mass (and fairly stable) markets. In a context of market uncertainty, however, this model has come to be viewed as too rigid and firms have increasingly focused on core competencies and externalized peripheral activities to other firms. In this way, firms could socialize the costs and risks of production by concentrating their resources on what they do best and forging linkages with other specialist firms for the balance of their needs. In some cases, these linkages go beyond formal, arms length (or contractual) relations and cost minimization, and contribute to the creation of value added by promoting the often informal sharing of knowledge regarding product or process innovations. The most effective relations are said to be those based on frequent interaction and the development of a ‘common cognitive ground’, that is, shared norms and conventions on how business is done.

The Limits of Neoclassical and Marxist Political–Economic Approaches

While recent empirical trends suggest that the nature of interfirm linkages are significant determinants of economic growth and that divergent forms of relations often lead to divergent levels of performance, past theoretical models have not been able to accommodate an analysis of the micro level dynamics that underpin these relations. In both neoclassical and Marxist oriented approaches, little agency is accorded to economic actors (e.g., the firm), as economic practices and behavior are subsumed under generalized laws or structures.

In neoclassically inspired location theory, the practices of economic actors are subject to the laws of utility maximization in the case of consumers, and profit maximization in the case of producers. Since actors are viewed as rational and fully informed, their choices are viewed as optimal and as a consequence, the market, which is a constellation of these individual choices, is viewed as self regulating. An extension of this perspective to the location decisions of firms suggests that firms will seek to minimize the costs of accessing inputs and markets, such that if costs are rising in one region, they just relocate to another. By assuming that actors operate on an isotropic plane, the model posits a tendency toward economic convergence within and across regions. However, such an approach does not address the persistent underdevelopment of some regions or the persistent growth of others.

Marxian political–economic approaches are a critique of the limits of neoclassical models to explain persistent underdevelopment in a capitalist (i.e., ‘free market’) context. A key assumption underlying this school of thought is that power and politics play an important role in shaping relations between economic actors. Actors are viewed as strategic. They do not operate according to rational choices based on the aggregate supply and demand (or ‘price’) of a commodity; rather, they operate according to class interests, which set owners of the means of production (i.e., labor and capital) against workers (whose labor power is extracted). Applying these basic tenets on a larger scale, Marxist theorists such as Andre Gunder Frank explain uneven development between regions and countries by dividing the world economy into distinct economic zones, with capitalist classes in core zones (typically, the north) exploiting working classes in peripheral zones (typically, the south) by extracting labor power and raw materials to produce goods. The limits to this perspective however is that it cannot account for how some economic zones (e.g., the east Asian economies) have been able to move out of an ‘underdeveloped’ state, since the practices of economic actors are reduced to their relative positioning in broader class structures.

Thus, as with neoclassical approaches, outcomes in the Marxist approaches are deterministic and agency is accorded a secondary, if any, role in analyses of economic behavior. Both are teleological approaches that help to identify how markets are reproduced and stabilized, but offer limited insight into the dynamic processes by which markets are formed or transformed. Economic systems are viewed as closed or self referential, as the under socialized and unidirectional nature of these approaches do not allow for causation to run from agents to broader structures or laws and thus preclude the possibility of divergent forms.

An exception to these approaches within the field of economic geography are locality studies, which came into prominence in the 1980s as researchers sought to examine the consequences of economic restructuring by conducting detailed, empirical studies of different locales and comparing these studies to isolate structural outcomes from the contingent ones. While inspired by orthodox political–economic approaches, the premise of this body of work is that localities are not merely acted upon or narrowly defined by structural forces, but that place specific social and political organizations could play a role in mediating the effects of such changes. Accordingly, these studies seek to counter overly deterministic accounts, by considering the interrelation between structure and agency. These studies, however, have been criticized in turn for privileging the empirical at the expense of the theoretical. Moreover, the focus of these studies has tended to be on the experiences of decline. Yet, with the resurgence of regional economies in the Third Italy and the rise of Silicon Valley as a high technology hub, by the 1990s interest was sparked in analyzing factors relating to growth as well as decline and relational economic geography arrived as a response, providing a new lens for analyzing the sociospatial manifestations of late capitalism.

The Localized Institutions Strand within Relational Economic Geography

A dominant and earlier wave of studies within relational economic geography can be grouped under a general heading of ‘localized institutions’ approaches. In a similar vein to the locality studies approach, this strand is concerned with identifying the local (or endogenous) factors that shape the processes of local and regional economic development. However, this strand also endeavors to go beyond descriptions of success cases (although the focus of these studies have tended to be on such cases) and to theorize the relation between the economic competitiveness of firms and the localized set of institutions in which they operate.

The point of departure for most of these studies is evolutionary economics, which in contrast to the neoclassical branch, views firms as satisficing rather than optimizing actors. Information regarding new technology, new processes, and new markets is complex and constantly in flux; therefore, firms are not fully informed. Further, firms tend to institutionalize successful ideas or practices and to develop routines that govern their future practices; thus, firms are not viewed as 'rational', as success may be attributed to chance (i.e., having the right strategy at the right time). Scholars that fall into the 'localized institutions' camp take these tenets to another level by positing that the broader institutional setting (conventions, norms) in which firms are situated will condition the selection of practices and routines. For example, a key question that these scholars are concerned with is whether firms are situated in an institutional context that promotes an open exchange of practices and ideas, and thus enlarge a firm's selection environment.

When considering the context that is likely to shape firm strategies, most of the proponents of this strand draw attention to the territorial based nature of such a context. The view here is that much of the learning that occurs in the initial stages of product or process innovations involves tacit knowledge, a form of knowledge that has not yet been articulated or codified. As economists such as Ikujiro Nonaka and Hiro Takeuchi have suggested, the process of translating this tacit knowledge into a codified form is a social one. This process entails frequent, face to face interactions. Over time, such interactions can promote trust and the development of collective understandings or norms, which are the institutions that underpin an open exchange of information among firms. Physical proximity is said to contribute to the development of these institutions, since proximity not only facilitates the exchange of material resources and services (i.e., 'traded dependencies') but also facilitates planned or random encounters (e.g., meetings in bars or on the street) that can foster informal relations between firms in a particular industrial field (i.e., 'untraded interdependencies'). Once a culture of information exchange and circulation is inscribed within a locale, then industry firms that reside there can benefit without much investment. Or, to quote Alfred Marshall, the secrets of industry are then 'in the air'.

Economic geographers have employed a number of interrelated concepts to capture the spatial dimension of knowledge creation, such as 'learning region', 'industrial district', and 'regional innovation system'. All these concepts can be said to fall within the localized institutions strand of relational economic geography in that they encapsulate three key attributes that define this strand: (1) an emphasis on the social foundations of economic learning with a view that close relations governed by shared norms and conventions can facilitate the innovation process; (2) a privileging of the 'local' as the spatial scale most suitable to promote such relations; and (3) a focus on the firm as a principal agent in the innovation process. Here, the firm is viewed as a single, coherent actor and as the key entity through which linkages in a cluster are forged. However, as elaborated above, the firm does not operate as a free agent, since the institutional setting constrains and enables a firm's ability to create and exchange knowledge.

With respect to methodology, an implication of these attributes is that scholars within this strand of relational economic geography have tended to focus on the properties of localized settings in which industries are concentrated, or what Henry Yeung has termed 'relational assets'. This entails an examination of the local governance structure, with studies seeking to establish whether a locale is characterized by openness versus secrecy, cooperation versus competition, and heterarchy versus hierarchy among firms and between firms and other local or regional organizations (e.g., research centers, trade organizations, and government). The studies also seek to establish the extent to which there exist intermediary organizations that can facilitate informal exchanges and relations. The assumption here is that the prevailing 'culture' of a locale, in terms of its shared agreements and unwritten conventions, plays a significant role in shaping economic practice and thus performance. A representative study of this strand is Annalee Saxenian's comparison of Silicon Valley with Route 128 in Massachusetts, where she attributes Silicon Valley's economic success to a culture of openness and cooperation.

Limits to the Localized Institutions Strand in Relational Economic Geography

In recent years, studies falling within the localized institutions strand have been criticized on a number of grounds, and above all, for their tendency to fetishize and reify the 'local'. For example, critics suggest that a preoccupation with the 'local' risks the neglect of other spatial scales at which economic learning occurs and at which relations underpinning economic learning are governed. This is particularly significant in a context of globalization, where economic actors are increasingly operating, forging networks, and thus providing new opportunities for learning on a supralocal basis.

Moreover, critics point to the fact that equating a set of practices with a particular place implies that local institutions (or 'cultures') are fixed, rather than dynamic.

It assumes that Silicon Valley is, and will be, an open setting, and that Route 128, conversely, remains a closed setting. Taken to its logical conclusion, however, the corollary view to this is that once the culture of a place is defined, then actors within that place are expected to act accordingly. This has the effect of removing agency from actors. Although they are not 'fully informed' or 'rational', their specific actions and choices can still be read in a mechanical way off local institutions. In this way, approaches within a localized institutions strand risk falling back into a structuralist trap, leaving little room for variety or incongruity within a given context or even a given firm.

The Agency-Oriented Institutions Strand within Relational Economic Geography

The strand of relational economic geography reviewed above has recently been challenged by several lines of thinking that could be grouped under the broad heading of 'agency oriented' approaches. Like localized institutions approaches, these approaches are concerned with the undersocialized nature of past theories and the role that actors and their interrelations play in mediating economic activities. However, studies within this strand problematize the reification of the firm by opening up the black box of innovation. The firm is not viewed as a single, coherent actor, but rather as a unit embodying multiple, potentially conflicting, interests. The focus of these studies is therefore on individuals within firms, their respective networks, and the implication of those networks for firm practices. Furthermore, networks are defined broadly as either formal or informal and as local or nonlocal. Thus, another significant attribute of this strand is the de linking of social embeddedness with spatial proximity.

There are two bodies of work that have been particularly influential within this strand of relational economic geography. The first is the 'communities of practice' (or CoPs) literature. CoPs represent informal groups of individuals who share a common practice or domain of interest and who interact in order to learn from one another and build their competence in that domain. Individuals in these groups are practitioners, so their learning is a situated form of learning that can involve both tacit and explicit knowledge. The tacit forms can be translated into an explicit and collective pool through the development of a shared repertoire of resources: experiences, stories and routines, habits, or procedures. An important attribute of this concept is that the communities are not constrained by formal structures, since they may entail connections among people across organizational and geographic boundaries. Thus, this literature allows for a conception of the firm as a constellation of communities that can provide distinct but complementary knowledge bases. This literature also allows for a reconsideration of the spatial organization of learning, as it does not privilege a single scale, *a priori*. Rather, it suggests that organizational, or what proponents term 'relational', proximity is more significant than spatial proximity for the knowledge creation process. In cases where this proximity is strong enough, that is, where individuals share a repertoire of resources, learning can occur over long distances.

The influence of the CoPs literature in economic geography can be seen in a range of studies on how learning occurs in global networks, where spatially distant actors are linked up by their respective involvement in a common activity or a shared occupation. This learning can occur through regular interactions or through temporary encounters, such as trade fairs. The influence of the CoPs literature is also evident in studies that examine flexible forms of economic organization, such as projects. According to Gernot Grabher, projects are temporary constellations of goal directed social relations. Although these communities are short lived, however, they nonetheless form the basis for new constellations in later projects, as individual members gain trust and reputation through their project while building a stock of competencies. Such communities also provide a dynamic field of markers of success, which can enhance the practices of more enduring structures such as firms.

Informed by actor network theory and post structuralism, the CoPs approach posits networks as fluid and actors as interdependent subjects whose identities and resource capabilities are co constituted through relations with other actors and with artifacts, for example, a collective set of rules or codebook for practice. These perspectives are also shared by a second approach within the agency oriented strand, the 'hybridity' approach, which emphasizes the need for a decentered conceptualization of economic actors (be they firms or individuals) as the fundamental categories of organizing socioeconomic life. Like the CoPs approach, the hybridity approach emphasizes the significance of relations that may cut across formal organizational structures or geographic boundaries; however, this line of thinking does not restrict itself to an analysis of the economic networks of individuals. Individuals are viewed as political, social, and cultural as well as economic beings and their actions as being shaped by a host of economic and noneconomic networks, each with its own logic. Nancy Ettlinger, for example, illustrates in her studies on workplace practices how such an approach can allow for an analysis of the cases where noneconomic networks may be more effective than economic networks in enhancing coordination and performance.

In terms of methodology, agency oriented approaches such as the CoPs and hybridity approach, in contrast to the localized institutions strand, eschew the development of a typology of assets or key attributes from which one could identify successful versus unsuccessful cases. Instead, the agency oriented strand emphasizes the contingency of outcomes for each empirical case, and the need for detailed and nuanced analyses of the relations implicated in a particular outcome. Thus, this strand favors an intensive case study approach, which enables a researcher to get close to subjects and to trace relevant linkages, be they local or nonlocal, formal or informal.

Limits to the Agency-Oriented Strand in Relational Economic Geography

By moving away from a teleological approach and highlighting the indeterminate range of context specific network configurations and practices, the agency oriented strand challenges some key assumptions of the localized institutions strand regarding the stable, invariant nature of a system of relations and constituent actors. It shifts attention from identifying patterns of system governance to identifying the particular dynamics that distinguish a given system, positing systems as open rather than closed and self reinforcing. However, in emphasizing the 'particular', this strand has been criticized for privileging descriptive, as opposed to explanatory, accounts, and thus, offering little in the way of building a theory for understanding new organizational forms of economic adaptation.

Part of the limitation of this strand can be attributed to assumptions embedded in the social theories upon which it draws, such as actor network theory. In this theory, all actors are assumed equal, with power only exercised through relations and intentionality only accorded to the network. This relativist notion of power negates any kind of influence existing outside of a network. In particular, this perspective overlooks the respective resource capabilities or social capital that individual actors may bring to a network or that may enable such actors to access the network in the first place. Nonlocalized networks (or CoPs) may bring together actors with common understandings and experiences but those individual understandings are first developed somewhere, and the qualities and features of the place can condition that development process and in turn mediate how, when, and where relations are formed. According to some geographers, an aspatial and ahistorical conceptualization of networks fails to account for how structure and agency interact in shaping network dynamics and their consequent success or failure. Rather than overcoming the structuralist agency dualism characteristic of traditional economic geography approaches, this conceptualization merely reverses it by giving primacy to agency in the analysis.

The Future of Relational Economic Geography?

While approaches falling within the school of relational thinking have moved away from analyses that either reduce behavior to the dictates of macro scale structures or treat economic actors as atomistic entities, there are still conceptual and methodological challenges that remain for relational economic geographers. Most specifically, geographers such as Henry Yeung have pointed to the need to overcome the binary thinking that continues to pervade relational economic studies. Drawing on the recent works of John Allen, Yeung argues that an account of the role of power can be a means for overcoming the structure agency divide in sociospatial analyses of innovation. Adopting a broad definition of power as both resources (i.e., power in its potential form) and the effects of resource use within practice (i.e., the manifestation of power), Yeung allows for a relational conceptualization, suggesting that there are myriad forms of power that can characterize the relations between actors. Consequently, collective and individual economic performance depends on the realization of these different forms in specific empirical contexts. In other words, outcomes cannot be determined *a priori* by taking stock of the available resources, assets, or other structural features of a given locale, a given actor, or a given network of actors; rather, the causality of such features are only effectual through practice.

The value of a relational view of power, as advanced by Yeung and more recently by Harald Bathelt and Johannes Glucker, is that it accords structures (e.g., place based institutions or market competition) an influence without reifying them, acknowledging that structures are in turn shaped, and given meaning, by agency. When applied to analyses of networks, this relational view implies that a researcher examines not just the nature of relations within a given network, but also the position of that network relative to existing structures and competing networks. Similarly, when analyzing economic actors (be they firms or individuals), a researcher must consider both the structural position as well as the practice of the actor. In terms of methodology, this kind of analysis may require an ethnographic approach, such as the participant observation method more commonly employed in the field of anthropology, which can allow for an examination of the qualitative nature of relations among actors as well as an assessment of the spatial, temporal, and institutional contexts in which these relations are embedded. However, such a method also demands that the researchers themselves assume a relational position, acknowledging their own subjectivity relative to the object (or subject) of study and how this subjectivity informs the research process and findings.

Accordingly, the recent call by Yeung for a new conceptualization of the interconnection between structure and agency through an analysis of power suggests that relational economic geography represents an ongoing work in progress rather than a completed project. In particular, key questions remain as to how studies can move beyond the introduction of new themes for analysis (e.g., institutions or networks) toward the development of new theorizations, or of how heterogeneous configurations of power can be researched, in practice. As scholars continue to adopt more socialized accounts of the process of economic learning, these issues are likely to inform future studies and debates within the evolving school of relational thinking.

See also: Economic Geography; Economics and Human Geography; Industry, Historical Geographies of; Relational Economic Geography.

Further Reading

Allen, J. (2003). *Lost Geographies of Power*. Oxford: Blackwell Publishing.

Amin, A. and Cohendet, P. (1999). Learning and adaptation in decentralized business networks. *Environment and Planning D: Society and Space* 17, 87 104.

Barnes, T. J. (2001). Retheorizing economic geography from the quantitative revolution to the 'cultural turn'. *Annals of the Association of Economic Geographers* 91, 546 565.

Bathelt, H. and Gluckler, J. (2003). Towards a relational economic geography. *Journal of Economic Geography* 3, 117 144.

Bathelt, H. and Gluckler, J. (2005). Resources in economic geography: From substantive concepts towards a relational perspective. *Environment and Planning A* 37, 1545 1563.

Boggs, J. S. and Rantisi, N. M. (2003). The 'relational turn' in economic geography. *Journal of Economic Geography* 3, 109 116.

Dicken, P. and Malmberg, A. (2001). Firms in territories: A relational perspective. *Economic Geography* 77, 345 363.

Ettlinger, N. (2003). Cultural economic geography and a relational and microspace approach to trusts, rationalities, networks and change in collaborative workplaces. *Journal of Economic Geography* 3, 145 171.

Grabher, G. (2002). Cool projects, boring institutions: Temporary collaboration in social context. *Regional Studies* 36, 205 214.

Saxenian, A. (1994). *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.

Scott, A. J. (2000). Economic geography: The great half century. *Cambridge Journal of Economics* 24, 483 504.

Storper, M. (1995). The resurgence of regional economies, ten years later: The region as a nexus of untraded interdependencies. *European Urban and Regional Studies* 2, 191 221.

Thrift, N. and Olds, K. (1996). Refiguring the economic in economic geography. *Progress in Human Geography* 20, 311 337.

Yeung, H. W. (2005). Rethinking relational economic geography. *Transcactions of the Institute of British Geographers* 30, 37 51.

Reliability and Validity

O. Ahlqvist, The Ohio State University, Columbus, OH, USA

Glossary

Analysis of Variance (ANOVA) A statistical technique that compares the variation in separate data series in order to determine if there is a significant difference between them.

Pearson and Spearman Rank Correlation Metrics that indicate the strength of a relationship between two variables. Spearman's rank correlation does not require the relationship between the variables to be linear, nor does it require the variables to be measured on interval scales.

Positivism A doctrine stating that the only basis for knowledge can be provided by affirmative, scientific observations, grounded in an empirically accessible world.

Root Mean Square Error A metric of the average departure of measurements from the overall mean.

Introduction

No matter what scientific tradition or methodology a geographic inquiry follows, a primary concern is to make sure that it can deliver credible results. Scientific credibility is one of the more important components that separate science from mere investigative journalism. Part of making sure that results are credible is therefore a critical process in any academic research. Assume that an instrument record shows that the average temperature has increased by 0.4 °C. Does this mean that the climate has turned warmer? If two surveys, one before and one after the events of 11 September 2001, show that more people mention terrorism as a concern for their safety. Does that mean that more people are feeling threatened by terrorist attacks now than before 11 September? If the number of people with an annual household income below the federal poverty level is getting larger, does that mean that poverty is on the rise in a region? These questions regarding the interpretation of results are one of the most critical steps a scientist needs to perform as they address whether the results are significant, and if our interpretation of results is valid and worth attention. In a strict sense, validity is the degree to which empirical data truthfully measure a construct of interest. Sometimes a construct is very tangible and fairly straightforward to measure, such as the average temperature over a period of time. Other times a construct is of a more theoretical or conceptual nature, such as the amount of feeling threatened by terrorist attacks. In either situation we may come up with a method to measure or otherwise find ways to evaluate the construct and end up with a result that points us to an answer, but how valid is that answer? Did we really succeed in measuring temperature change in a truthful way? Does our instrument capture people's perception about terrorism threat? Is household income a valid way to measure poverty to begin with?

The question of 'validity' is related to the result of a study, and in an empirical setting it refers to the degree to which gathered information actually measures the phenomena of interest. Achieving validity obviously comes down to the skills of the investigator to come up with a proper research design and draw reasonable conclusions. Research design is a whole topic in itself but is in short made up of a number of components, including theory construction, data collection procedures, and analytical techniques. Each component is vital to the final validity of findings. One of the more important components of an empirical study is to identify proper, well known, or plausibly good working instruments to capture the information we want. A critical component of any such instrument, whether it is a thermometer or an interview protocol, is that it captures information in a consistent manner so that we get reliable data. 'Reliability' thus refers to the stability and consistency of the research instrument over repeated applications.

Although our instrument, for example, a community census, could be determined to generate reliable estimates of annual household income, we could still argue about its validity as a good metric for poverty. There are most likely other aspects of poverty that get lost in that formalization and the validity of any study that uses only income as a metric of poverty will not be entirely valid. On the other hand, if we were to agree that annual income is a valid metric of poverty, we could still not use it if our instrument was unable to give us reliable information on income. In this way the validity of empirical findings is dependent on the reliability of the empirical data used, but not necessarily the other way around.

A critical evaluation of reliability and validity can be done in many different ways: from a face value acceptance of results or procedures simply because they seem reasonable, to detailed and strict procedures of test retest, instrument calibration, use of reference data, and peer/member checking. However, the specifics of what evaluation method to follow and how to perform the evaluation differ depending on, for example, the research purpose, assumptions, or its methods.

The Scope and Meaning of Reliability and Validity in Human Geography

The different traditions and methodologies in geography typically put varying emphasis on the evaluation of validity and reliability. The quantitative, natural science research tradition has mostly shown a concern for achieving repeatability of experiments and the idea that scientific investigations can be made in an objective way. The social science tradition on the other hand needs to include more emphasis on qualitative information and puts emphasis on learning about the world first hand with a flexibility to capture rich, context dependent situations. It may be tempting to assume that a concern for reliability is only relevant in the quantitative research tradition where a positivist assumption assumes that there is an underlying universe where measurements can be replicated. Indeed, many types of investigations involving in depth interviewing and participant observation have paid little attention to evaluating the plausibility and validity of findings. Nevertheless, there is a wide recognition within social as well as natural sciences that evaluation is critical in achieving credible findings. But, because of the wide variety of methods for doing human geographic research, data reliability has significantly different meaning depending on the methodology employed.

Many accounts that deal with quantitative measures talk of reliability and validity in terms of uncertainty as the general term that includes many different components. Reliability and validity can be compared with two commonly mentioned and often interacting accuracy components – precision and bias – and they can be illustrated using a simple temperature measurement example. Consider a thermometer taking repeated measurements in a bucket of water where temperature is kept constant during the experiment. If the resulting recordings look like the strip plot in **Figure 1a**, we would think of the thermometer as very precise and unbiased and consequently reliable. If on the other hand the recordings look like strip plot (**Figure 1b**), we would regard this as a less precise but still unbiased instrument. If we were to produce a mean of the measurements we would end up with a very accurate result since the variation would even out. If on the other hand there is a systematic deviation from the true value such as the temperature readings illustrated by **Figure 1c**, we would think of this as a precise but biased instrument. If we can somehow compensate for the bias we could still end up with a very reliable instrument. The fourth situation, **Figure 1d**, illustrates the most unreliable instrument where there is a low precision and an element of bias as well.

Accuracy thus refers to the more general notion of validity that addresses how close measurements are to the 'true' value, and of the two separate components of accuracy, precision would most closely relate to the term reliability. This example is obviously based on the positivist notion of objectivity where ground truth at least theoretically could be measured. Unfortunately there are always problems of knowing the 'true' value. Accuracy then is also, even under a quantitative science interpretation, in practice a relative measure; the difference between a measurement and another measurement known to be of higher accuracy or one conforming to the general notion of the true value. Although this definition is inherently circular, it generally works under the pragmatic assumption that we can agree on what constitutes the most accurate measurement, typically one that has gained general acceptance or have used a larger effort to achieve accurate information.

Reliability and validity are still somewhat different in that they need to emphasize other aspects of the data than their quantitative siblings generally do. Since much qualitative information gathering strive to understand a phenomenon from inside using the words and actions of the participants as a foundation, strict protocols and standardized measurement are often not an option. Analysis and interpretation of the data are also not intended to be subject to standardization or calibration. This could make any validation of qualitative research findings entirely contingent on a face value trust of the author's ability to adequately summarize the data. However, there are other aspects of the research procedures that, given its careful consideration, would make assessment of reliability and validity achievable.

In many observational studies audio and video recordings can be retained or turned into detailed transcriptions allowing other people than the researcher direct access to the words and actions of research subjects. In cases where transcription is made of observations it is possible to maintain this direct access in the final transcript. Transcription can be done 'blind' making sure that persons doing the transcriptions are not aware of or influenced by the researcher's expectations. This is a data collection procedure common in the quantitative tradition to reduce the risk of introducing bias in the data collection stage. Personal integrity and confidentiality are obviously important issues but should not be an excuse for not maintaining some first hand account of the data used in an analysis.

In much text analysis, including analysis of observational and interview transcripts, it is common to use coding to distil the important information. To achieve reliability in this coding we want to achieve consistency in the coded transcripts so that a section coded as 'migration' would be coded the same way by another coder or if coding is needed to be redone. A way to achieve this is to ask coders to repeat coding of the same piece of text after

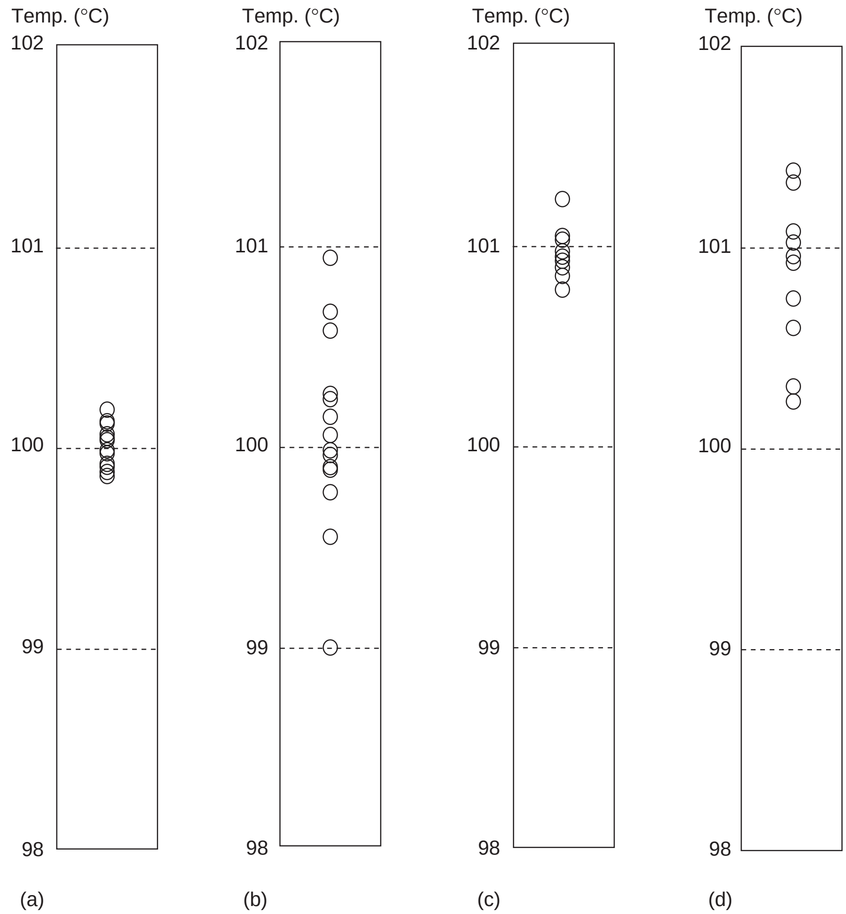

Figure 1 Strip plots of hypothetical measurement series where a constant temperature is repeatedly measured with four instruments (a d) of different accuracy; (a) high precision, no bias; (b) low precision, no bias; (c) high precision, with bias; and (d) low precision, with bias.

some time, or if several coders are working on a larger project, have them code the same piece of text, and then compare results to work out discrepancies. In a similar fashion a test–retest process can gather data on whether an instrument produces equal responses when repeated.

Achieving reliability of qualitative data focuses more on research procedures than the data itself so that a re liability check of a study based on interviews would ask questions such as: what information was collected and how? How were participants selected/sampled? How was the analysis performed? These are similar to what any critical reader of the results would ask for and require in enough detail to understand the full context of a study, context meaning not just the social and environmental setting of a study but also by what notions and study methods the research(er) makes observations. An upfront and transparent account of these procedures is necessary for evaluation of reliability.

Validity follows a similar line of reasoning in that it is increasingly thought of as a concept made up of different lines of validity evidence where a positivist interpretation of truth is not possible. Validity is evaluated with reference to a construct that is a scientifically informed idea developed to explain some phenomenon such as 'globalization', 'segregation' or 'poverty'. The process of evaluating validity then calls for accumulation and evaluation of evidence for a sound scientifically based interpretation where no single existing measure can be used as a criterion. Thus, validity assessment is con cerned with evaluating the interpretation of data or the degree to which accumulated evidence support a specific interpretation. Again, upfront and transparent reporting of engagement of hypotheses, available evidence, and interpretive arguments is the key to good validation.

Measuring Reliability and Validity

Quantitative measures are typically evaluated on their reliability using standard statistical approaches on the data. For example, the root mean square error, Spearman rank, or Pearson correlation are commonly used as in dices of the overall accuracy of an instrument; standard deviation and analysis of variance (ANOVA) methods can approximate the consistency of an instrument; and standard hypothesis testing is typically used to gauge the validity of measured variables. Despite these fairly straightforward procedures, the final validation is usually more complex. Questions surrounding temperature measurements may, for example, include if there is an overall warming trend, or if it can be attributed to human activities. These deeper questions are typically not answered by the statistical techniques but rather subject to interpretation, thus the need for fully transparent procedures and reliable data.

As mentioned previously, the reliability of qualitative measures is generally evaluated based on the procedures of data collection by providing a full account of the research activity. However, there are similar metrics as for quantitative data that can be used to measure qualitative data reliability. A survey instrument could, for example, use test–retest data to calculate the correlation between ratings in the first test and ratings in the second test. High correlation would suggest that the instrument generates consistent answers from respondents. When classification is the goal of a study or instrument question, for example, where multiple choice questions generate nominal data, a ratio of same answers to the total number of answers can be calculated. If the instrument is reliable, users would tend to pick the same answer when the instrument is repeated and that would generate a ratio close to 1. If users tend to fluctuate between alternatives in their answers, there may be confusion and that would generate a lower reliability ratio.

A variety of other approaches are possible too that address these and other facets of validity and reliability. Generalizability theory was developed specifically to recognize and separate different aspects of reliability; occasion variation emerges in test–retest situations, item variation occurs when alternative forms of the same instrument are used, and rater variation is simply the respondent's variability to the same stimuli. Obviously individuals exhibit some variation in their behavior in that they may try harder, making luckier guesses, being more alert, being in a good/bad mood, and have better/worse health than usual. Different generalizability coefficients, defined as the ratio of 'true' variance to observed variance, are generated through extensions of ANOVA methods to tease out these different aspects of reliability. However, these coefficients are typically most useful for comparing different test methods or procedures.

Concluding Comments

It has been argued that much qualitative work has not been rigorous enough in terms of its transparency about the purpose of a study, how information was collected, how data analysis was done, and so forth. That debate reflects an ongoing adjustment from a period where qualitative work went through a separation from the positivist, quantitative scientific tradition, to the current situation when mixed method approaches become more and more common. Quantitative researchers are also increasingly aware of the situated nature and context dependence of traditional numerical, quantitative approaches.

Because much qualitative research in human geography use methods similar to those in psychology and psychometrics, many approaches to address reliability and validity can be found in that literature. For example, the American Psychological Association has published standards for educational and psychological testing which provide detailed account of definitions and suggested approaches to validity, reliability, and errors of measurement.

Geographic observations normally describe phenomena with spatial, temporal, and thematic dimensions. There has been a deliberate focus on reliability and validity of qualities, that is the thematic value or category used to describe some spatiotemporal phenomenon of interest. It is important to realize that the idea of reliability and validity also applies to observations of spatial and temporal information, and that space, time, and theme are often dependent on each other. For example, the ability to locate a poor neighborhood depends on the same process of measuring poverty as the identification of poverty in the first place. Unreliable poverty metrics will result in unreliable spatial boundaries. Measuring spatial and temporal uncertainties has traditionally used quantitative techniques, but recent interest in qualitative spatial reasoning will ultimately end up needing similar types of validation and reliability assessment that have been described here.

See also: Ecological Fallacy; Error (Propagation and Modeling); Hypothesis Testing; Qualitative Geographic Information Systems; Selection Bias; Uncertainty.

Further Reading

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

Baxter, J. and Eyles, J. (1997). Evaluating qualitative research in social geography: Establishing 'rigour' in interview analysis. *Transactions of the Institute of British Geographers* 22, 505 525.

Cook, D. A. and Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: Theory and application. *The American Journal of Medicine* 119(2), 166.e7 166.e16.

Eyles, E. and Smith, D. M. (1988). *Qualitative Methods in Human Geography*. Totowa, NJ: Barnes and Noble Books.

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement* 38(4), 319 342.

Kirk, J. and Miller, M. (1986). *Reliability and Validity in Qualitative Research*. Beverly Hills, CA: Sage.

Silverman, D. (1993). *Interpreting Qualitative Data*. London: Sage.

Summers, J. O. and MacKay, D. B. (1976). On the validity and reliability of direct similarity judgements. *Journal of Marketing Research* 13(3), 289 295.

Religion/Spirituality/Faith

L. Kong, National University of Singapore, Singapore

Glossary

Numinous Rudolf Otto's (1917) concept describing the sense of awe in the presence of the sacred
Pooja Hindu term for prayer.
Sacralization The process of making sacred.
Ummah Pan-Islamic community.
Wicca A sub-branch of Paganism associated with witchcraft.

Introduction

Geographical research on religion, spirituality, and faith has grown over the last two decades, particularly beginning in the early 1990s, after a period of relative inactivity from the 1960s to 1980s. In part, that relative silence coincided with the Quantitative Revolution in geography, which did not embrace the study of personal, individual, inchoate experiences such as that associated with religion. In part, it coincided with a period of growing secularization in the world. Since then, religious revivalism, fundamentalism, the emergence of new religions, and the association between religion and terrorism, have reignited interest within geography in the study of religion.

Research Traditions

Right up till the early 1990s, geographical research on religion focused on a range of areas. Interest in mapping the spatial advance of religions (especially Christianity) in the world began as early as the sixteenth and seventeenth centuries but continues today in the form of mapping religious distribution in the world. This is a form of ecclesiastical geography, with its roots in Christian missionaries attempting to identify which parts of the world were 'saved', and which needed their transformative assistance. Another evidence of the Christian Church's influence is a form of biblical geography which attempts to identify places and names in the Bible and to determine their locations.

In the eighteenth and nineteenth centuries, scholars began to view landscapes, and in particular, nature, as a divinely created order for the well being of all life. For example, it was believed that the distribution of climates, plants, animals, and landforms was divinely planned for the well being of life. This close relationship between nature and religious worldviews became evident too in a school of thought termed environmental determinism in which the physical environment was sought as an explanation for human and social phenomena, including religion.

While environmental determinism offers less explanatory power now than it did in the past, what has gained ascendancy is an interest in knowing how religion in turn shapes the environment. Such a perspective was taken up particularly in the work of Carl Sauer and the Berkeley school of cultural geography. Landscapes became the main subject of analysis, and sacred structures were described and documented as evidence of how religion impacted the landscape. Relatedly, the role of religion in environmental degradation has also been studied, and such 'religious ecology' or 'environmental theology' has emphasized the ways in which religion has influenced attitudes toward plant and animal ecology, and environmental degradation.

Politics of Sacred Spaces

While some of the above approaches continue to underpin current research, more recent efforts increasingly recognize the integral interweaving of religion, spirituality, and faith with secular life. Sacred space is contested space, just as the sacred is a contested category. Particularly, sacred space is both the expression and outcome of power relations of domination, subordination, and negotiation, and is hence implicated in various politics. Such power relations are evident in the ways in which sacred spaces are produced, managed, and maintained, as well as the ways in which they are consumed and lived.

The most obviously sacred spaces are places of worship such as temples, churches, mosques, and synagogues. The politics surrounding their production are often the politics of secular–religious relations and majority–minority relations. In establishing places of worship, for example, power relations determine location. The secular power of the state, represented by 'rational' urban planning ideologies, capitalistic principles of land values, and principles of multiculturalism can run up against religious principles and divine guidance in the choice of location for a place of worship. Similarly, state discourse and practice surrounding burial and crematorial space often hinge on secular utilitarian views of planning, adopting principles of efficient land use and taking on board concerns about sanitation. On the other hand, local communities emphasize symbolic and religious meanings of graves as focal points of identity, expressions of relationships with the land, and crucial to the practice of religious beliefs and rituals. Thus, power relations also underpin the siting of cemeteries and columbaria.

The power of the majority group can also prevent a minority group from establishing its place of worship, an exclusion of the marginal 'other'. This can have the effect of forcing adherents to worship secretly, such as in properties meant for other purposes (residences, commercial premises). It is possible that there could be a confluence of the religious majority with the state, in which case dominance may be total. Religious adherents have to find ways of coming to terms with the primacy of the secular order, negotiating their conceptions of the sacred.

Besides the politics surrounding the production of religious place, the relations of domination and subordination also surround their management and maintenance. For example, tensions and negotiations may arise between indigenous people who revere sacred sites, and modern forces which want these sites for pragmatic, commercial, or even alternative religious purposes. Another example is the differential management of facilities associated with pilgrimage routes and sites arising from divergent meanings invested in these routes and sites. For example, the sacred meanings associated with sacrifice and privation along pilgrimage routes and at pilgrimage destinations lend themselves to particular ways of management and maintenance. On the other hand, these same pilgrimage routes and sites may be implicated in capitalist objectives of tourism promotion, and the corresponding promotion of comfort and enjoyment.

Poetics of Sacred Place

Even while sacred space is contested space, it is also important to recognize that for the religious person, the relationship with sacred space is more personal and the experience is more akin to a 'poetic' communion with the immanent and transcendent. Space becomes place. Perhaps because of the inchoate nature of this experience, few geographers have examined empirically this experience, particularly as mediated through place. The most influential literature in this regard still remains early works of scholars of religion: Mircea Eliade (1959), who conceives of the sacred as erupting in certain places as revelations (hierophanies), causing them to become "powerful centers of meaningful worlds," set apart from ordinary, homogeneous space; and Otto who describes the experience as a "numinous" one in which a variety of emotions not unlike ordinary happiness, anger, fear, and so forth are directed to the religious. Lane (1988) has also conceptually crystallized certain "axioms" of such sacred place: it chooses, it is not chosen; it is ordinary place, ritually made extraordinary; it is intimately linked to states of consciousness, such that it is possible to go by a place numerous times without recognizing it as sacred; it is both local and universal and can not only drive one to a quest for a particular center of divine encounter but also drive one out from that center with an awareness that God is never confined to a single locale. The last perspective parallels two general spatial orientations in the study of religion, the locative and utopian: the former is fixed, bounded, and requires the maintenance of one's place and that of others in a larger scheme of things; the latter is unbounded and unfixed to any particular location, breaking out of a prevailing social order.

Although some scholars believe that sacred place is not chosen, that is, not made by humans, others believe that places can be sacralized and choose to examine the manner in which place develops its sacred meaning. Anthropologists and geographers share a common interest in the ritualization that turns secular place to ritual and hence sacred place. Two examples will illustrate. First, drawing an example from new religions, Wiccans believe that sacred place can be set up anywhere (Wicca is a subbranch of Paganism associated with witchcraft). The processes of sacralization involve moments of quiet meditation prior to casting a circle, setting up the altar, laying out the witch's tools, ringing a bell to signal the commencement of the rite, and so forth. Second, drawing an example from an established world religion, Hindus can also sacralize the dwelling place by purifying the outside (e.g., through consecration of the land and planting of ritually significant plants), and sacralizing the inside (e.g., through lighting the sacred fire, anointing participants with ashes from the fire, and walking a cow through the rooms). Sacred place can thus be made and need not be singularly hierophanic.

In examining the poetics of sacred places, attention has refocused on some of the earlier contributions in humanistic geographies. Discussions about how religious places offer a sense of rootedness and identity find resonances with earlier ideas propounded by humanist geographers such as Tuan and Buttimer in nonreligious contexts. Kong has explored, for example, the personal and familial histories of religious adherents in Singapore and how they are tied up with churches and Hindu temples, contributing to the development of personal attachments and senses of place. Mazumdar and Mazumdar have focused on Hindu sacred place, arguing that the domestic *pooja* (prayer) area is viewed as a family heirloom and evokes a sense of rootedness. While representing continuities with the work of humanistic geographers, and illustrating the applicability of existing concepts, geographical research on religion has not contributed substantially to a reconceptualization of our understanding of place attachments. Attachment to religious place seems little different from attachment to secular place.

Poetics of Community

Studying the nature of place in mediating the experience of the sacred for the believer would be incomplete if not coupled with an understanding of the sense of identity, belonging, and community that accompanies member ship of a religious group.

The nature of experience of being part of a religious group – the poetics of religious communities – is enabled by the fact that religious places are often also social centers which facilitate community building. Religious places such as mosques and temples are social centers where adherents gather, not only to pray, but to engage in social activities as well. The quality of interaction and the mutual sense of commitment and sharing is part of the experience of being 'religious' for some religions (such as Christianity and Islam). However, it is important to remember that praying in the same place and engaging in activities together do not automatically mean that members 'belong' together as a 'community'. In as much as there are processes of sacralization of place, the making of community also relies on the experience of shared ritual. Thus, Turner's earlier anthropological work on pilgrimages introduced the ideas of "communitas" in which a sense of community emerges among those en gaged in the shared ritual of pilgrimage.

Whereas it is obvious that interaction arising from shared location can lead to the development of place based communities, spatially dispersed nonlocalized communities are increasingly enabled by technology. Communities have common needs and goals, a sense of the common good, shared lives, cultures and views of the world, and collective action. These rely on interaction and communication between community members, which are much more likely when there is unmediated face to face contact between people, which, in turn, means locatedness in a place. It is therefore understandable when religious communities that are grounded in particular places, such as parishes, develop. However, communities may also be spatially disper sed ('place free', 'stretched out'), and communicative media such as the telephone and the Internet allow for the construction of communities without territorial base. As in other domains of life, religious activities are not immune to the influence of technology, such as re ligious broadcasting and computer mediated communi cation (e mail, discussion lists, websites). They have the potential of creating communities that are not locked in place.

Politics of Identity and Community

To assume that 'belonging' to a parish or praying in the same place leads to a feeling of integration and community with other worshippers would be a pre sumptuous and somewhat uncritical treatment of re ligious places as social centers. At the same time, while 'traditional communities' are commonly conceived as harmonious entities with shared needs, goals, values, and activities, they are really a form of idealization. Often, the reality is characterized by tensions within the social group, with various forms of oppression, 'protecting the prevailing value system including its moral code', and displaying an 'intolerance of difference'. 'Community' is thus a process of becoming rather than a conferred status. The mediations and negotiations between groups and among members of a group con stitute the process of community making, just as the sacralization of place entails ritual performance. Such mediations are necessarily about relationships of power, of domination, resistance, and negotiation. Understand ing the construction of religious community thus entails engaging the politics of religious identity and community.

The Islamic 'community' is an excellent case for examining the politics of religious identity and com munity. The notion of the '*ummah*', a global community of Muslims, overlays local conceptions of place based communities of Islamic adherents, and tensions and conflicts. Also, reinforcements across the global and local are real. At the same time, within local place based communities themselves, the complications of com munity construction are already complex.

At one level, the local place based community already demonstrates a complexity of relationships that defy easy characterization as a harmonious entity with shared goals, values, and activities. For example, Dwyer's work on young Muslim women in a small town near London examines the ways in which community is not only a source of security and strength but also of constraint and oppression for young Muslim women. For these women, the sense of a local 'Asian community' was real, signaling a sense of security and acceptance (hence no racism) in the town. While this was positive, it came at a cost: living in an Asian community meant all sorts of surveillance by other members of the community about one's actions and behavior. The sense of contradiction of community thus confronted these young British Muslim women in real everyday ways.

At another level, the interweaving of a global com munity with local conditions adds another layer of complication and politics to an adherent's sense of identity and community. For example, reinforcing the sense of belonging to a transnational Muslim community among Singapore Muslims is achieved, *inter alia*, via ac cess to Islamic broadcasting from Malaysia, in which the universal teachings of Islam are purveyed. On the other hand, such religious broadcasting also contributes to the fracturing of the sense of global community. This challenge to transnational religious identity and community arises because Malaysian religious radio and television programs often address specifically local issues such as social problems particular to Malaysia, framed within a religious discourse. These may or may not be pertinent in Singapore, and may in fact be subject to divergent interpretations in a different context. Thus, whether the focus is on the global or local nature of the broadcast messages, the effect is an ongoing construction and contestation of multiple Islamic identities. Even while the constructed Islamic identity is a global one in that listeners and viewers are reminded of shared beliefs and a common identity, the Islamic identity is also a local one, tied up with specific social and political experiences in particular contexts. It underscores the view that the sense of religious identity and community is relevant to and requires understanding specifically within the context in which it is lived and practiced. In other words, there should be no *a priori* category called 'Islam' or a preexisting global community of Muslims. Rather, all 'Islams' and 'Muslims' are to be understood within the specificity of social, economic, and political conditions, and all Muslim communities are negotiated within such specific contexts.

No discussion of the politics of religious identity and community is complete without addressing the role of religious places in constructing and maintaining the boundaries between and within religious groups. Control over religious places can play an important role in this regard. For example, members of the Caribbean Hindu Society in London use the temple in a completely congregational manner, opening only for collective worship and remaining closed to individual and family based worship on weekdays; organizing communal activities where food of the Caribbean Indian variant is served; and reciting prayers congregationally, with the equivalent of church prayer books. The temple therefore becomes a significant means of consolidating and reproducing the Caribbean Indian Hindu community. Vertovec argues that this use of the temple has emerged because in Trinidad and Guyana, Hindus were at the bottom of the social structure and congregational worship provided a sense of mutual support and the maintenance of self esteem, demonstrating and reinforcing their ethnic identity. Those who migrated to Britain were still in an ethnic quandary, with the white British population thinking of them derogatorily as 'Paki' (subcontinental Indian); their official 'West Indian' status; and their harsh treatment by South Asians who saw them as a pariah group. For them, the temple helps to maintain and reinforce their sense of identity and community, emphasizing their legitimacy and presence. By contrast, where the need to consolidate and reinforce identity is not as marked, the temple does not play the important role of a 'community center'.

'New' Geographies of Religion

The article thus far illustrates how the literature on geographies of religion is a rich and varied one, reflecting the myriad issues confronting adherents in their religious lives, groups in their identity and community building, and states in their management of religions. Yet, there are many areas still that deserve more research attention.

One significant dimension is a willingness to reach beyond 'officially sacred' sites to understand the ways in which religious meaning, identity, and community can be underscored or challenged via sites other than places of worship, pilgrimage sites, and shrines. Ordinarily secular places can also be transformed into places for religious use, such as public streets for religious processions, private homes for domestic worship, schools for religious instruction, museums for religious display, and broadcasting or cyberspace for religious discussion or even ritual. These need to be more widely studied, in order that religion is not inadvertently categorized as a practice sustained in defined religious spaces only.

Quite as important is a recognition that the experience of religion – the numinous – is multifaceted, and the need to examine the role of the senses beyond the visual. As long as the focus has been on landscapes, the visual has been privileged. In religion, the aural is obviously an important part of the overall experience. So too, the olfactory. These may contribute toward a total religious experience, creating an emotional response currently little understood. Besides the different sensuous sacred geographies therefore, the role of emotion in religious experience also deserves further scrutiny.

The same religions are invariably interpreted and practiced in different historical and place specific contexts, and must be subject to context specific analysis. Analytic categories must not be mistaken for substantive categories and religion must not be taken as *a priori* theory. With different analyses in varied empirical contexts must come more nuanced theorizations.

Further, more analysis that connects various scales is important, given the interconnectedness across scales: global, national, regional, local, and indeed, that of the body. The reach of transnational religious groups set against the mediations of local forces demands attention, as does the question of how pan religious identities and communities (e.g., the *umma*) conflict with local and national affiliations. At the other end of the spectrum, in analyzing the body as a site of religious expression and contestation, dominant ideologies and representations endorsed 'universally' and 'globally' by a religious group (e.g., Islam) must be understood as subject to re interpretation and contestation in specific local contexts.

Thus far, there has not been sufficient attention paid to the different ways in which religion is experienced and shaped by different population constituents. For example, the differential situations with women and men, children and the elderly, and the different social classes cannot be conflated.

Finally, while morality and social justice may exist apart from religion, often, religion is the basis of morality and the impetus for social justice, as well as of intolerance and injustice. The ways in which different religions may inform the constructions of different moral geographies and the ways in which these constructed moral geog raphies contradict or are negotiated or reinforced by other secular agents of morality (e.g., the state) are im portant areas of study.

Moving ahead, the cultural and social geography of religion must therefore take into account a wide array of spaces, peoples, experiences, and practices. In so doing, geographers can contribute to refining theoretical understandings of the nature of the sacred (place, iden tity, and community). By enhancing our understanding of religion, geographers can play a vital role in a world where some of the most intractable and violent conflicts are related to religion.

See also: Diaspora; Identity Politics; Landscape.

Further Reading

Carmichael, D. L., Hubert, J., Reeves, B. and Schanche, A. (eds.) (1994). *Sacred Sites, Sacred Places*. London: Routledge.

Chidester, D. and Linenthal, E. T. (1995). Introduction. In Chidester, D. & Linenthal, E. T. (eds.) *American Sacred Space*, pp 1 42. Bloomington: Indiana University Press.

Davies, W. K. D. and Herbert, D. (1993). *Communities within Cities: An Urban Social Geography*. London: Belhaven Press.

Dunn, K. M. (2001). Representations of Islam in the politics of mosque development in Sydney. *Tijdschrift voor Economische en Sociale Geografie* 92(3), 291 308.

Dwyer, C. (1998). Contested identities: Challenging dominant representations of young British Muslim women. In Skelton, T. & Valentine, G. (eds.) *Cool Places: Geographies of Youth Cultures*, pp 50 65. London: Routledge.

Dwyer, C. (1999). Contradictions of community: Questions of identity for young British Muslim women. *Environment and Planning A* 31, 53 68.

Eliade, M. (1959). *The Sacred and the Profane: The Nature of Religion*. Translated from French by W. R. Trask. San Diego: Harcourt Brace Jovanovich.

Graham, B. and Murray, M. (1997). The spiritual and the profane: The pilgrimage to Santiago de Compostela. *Ecumene* 4(4), 389 409.

Holloway, J. (2006). Enchanted spaces: The séance, affect and geographies of religion. *Annals of the Association of American Geographers* 96(1), 182 187.

Hume, L. (1998). Creating sacred space: Outer expressions of inner worlds in modern Wicca. *Journal of Contemporary Religion* 13(3), 309 319.

James, W. (1902). *The Varieties of Religious Experience: A Study in Human Nature*. New York: University Books.

Johnston, R. J., Gregory, D. and Smith, D. M. (eds.) (1994). *The Dictionary of Human Geography* (3rd edn.). Oxford: Blackwell.

Knox, P. (1995). *Urban Social Geography: An Introduction* (3rd edn.). Harlow: Longman.

Kong, L. (1990). Geography and religion: Trends and prospects. *Progress in Human Geography* 14(3), 355 371.

Kong, L. (1993a). Ideological hegemony and the political symbolism of religious buildings in Singapore. *Environment and Planning D: Society and Space* 11(1), 23 45.

Kong, L. (1993b). Negotiating conceptions of sacred space: A case study of religious buildings in Singapore. *Transactions, Institute of British Geographers, New Series* 18(3), 342 358.

Kong, L. (2001a). Mapping new geographies of religion: Politics and poetics in modernity. *Progress in Human Geography* 25(2), 211 233.

Kong, L. (2001b). Religion and technology: Refiguring place, space, identity, and community. *Area* 33(4), 404 413.

Kong, L. (2002). In search of permanent homes: Singapore's house churches and the politics of space. *Urban Studies* 39(9), 1573 1586.

Kong, L. (2005). Religious schools: For spirit, (f)or nation. *Environment and Planning D: Society and Space* 23, 615 631.

Kong, L. (2005). Re presenting the religious: Nation, community and identity in museums. *Social and Cultural Geography* 6(4), 495 513.

Kong, L. (2006). Religion and spaces of technology: Constructing and contesting nation, transnation and place. *Special Issue on Geographies and Politics of Transnationalism: Environment and Planning A* 38, 903 918.

Lane, B. C. (1988). *Landscapes of the Sacred: Geography and Narrative in American Spirituality*. New York: Paulist Press.

Ley, D. (1974). The city and good and evil: Reflections on Christian and Marxist interpretations. *Antipode* 6, 66 74.

Ling, T. O. (1987). *Buddhism, Confucianism and the Secular State in Singapore*. Singapore: Department of Sociology, National University of Singapore, Working Paper No.79

Mazumdar, S. and Mazumdar, S. (1993). Sacred space and place attachment. *Journal of Environmental Psychology* 13, 231 242.

Naylor, S. K. and Ryan, J. R. (1998). Ethnicity and cultural landscapes: Mosques, gurdwaras and mandirs in England and Wales, Paper presented at the Religion and Locality Conference, University of Leeds, 8 10 September.

Otto, R. (1917). *The Idea of the Holy.* Translated by Harvey, J. W. (2nd edn.). London: Oxford University Press.

Pacione, M. (1999). The relevance of religion for a relevant human geography. *Scottish Geographical Journal* 115(2), 117 131.

Philp, J. and Mercer, D. (1999). Commodification of Buddhism in contemporary Burma. *Annals of Tourism Research* 26(1), 21 54.

Rath, J., Groenendijk, K. and Penninx, R. (1991). The recognition and institutionalisation of Islam in Belgium, Great Britain and the Netherlands. *New Community* 18(1), 101 114.

Silk, J. (1999). The dynamics of community, place, and identity. *Environment and Planning A* 31, 5 17.

Smith, D. M. (1999). Geography, community, and morality. *Environment and Planning A* 31, 19 35.

Smith, J. Z. (1978). The wobbling pivot. In Smith, J. Z. (ed.) *Map Is Not Territory: Studies in the History of Religions*, pp 88 103. Leiden: E. J. Brill.

Teather, E. K. (1998). Themes from complex landscapes: Chinese cemeteries and columbaria in urban Hong Kong. *Australian Geographical Studies* 36(1), 21 36.

Turner, V. (1974). *Dramas, Fields, and Metaphors: Symbolic Action in Human Society*. Ithaca: Cornell University Press.

Vertovec, S. (1992). Community and congregation in London Hindu temples: Divergent trends. *New Community* 18(2), 251 264.

Williams, R. (1977). *Marxism and Literature*. Oxford: Oxford University Press.

Young, I. M. (1990a). The ideal of community and the politics of difference. In Nicholson, L. J. (ed.) *Feminism/Postmodernism*, pp 300 323. London: Routledge.

Young, I. M. (1990b). *Justice and the Politics of Difference*. Princeton: Princeton University Press.

Remittances

B. Page, University College London, London, UK

Glossary

Diaspora A group of people who have been dispersed by migration from their original common homeland but who continue to share an identity because of their commitment to, and memories of, the homeland.

Group Remittances The movement of money or services from a group of migrants in the diaspora to their home country.

Hometown Associations Clubs of migrants who share a loyalty to a particular hometown and who meet regularly to support each other and sometimes to assist with the development of the hometown.

Repatriated Deposits Financial term for migrant workers' remittances to their country of origin that remain in the migrant's ownership after they are remitted.

Social Remittances The movement of ideas (rather than material goods or money) from migrants to their place of origin.

Unrequited Transfers Financial term for migrant workers' remittances to their country of origin that do not remain in the migrant's ownership after they are remitted. Also includes prizes, donations, dowries, and inheritances.

Introduction

Remittances are capital that is sent by the individual worker who has earned it from the place where it was earned to their place of origin. For example, a woman originally from Mexico who has moved to live and work in the US may send a proportion of her earnings back to her family in Mexico in order to help with their daily expenditures. She is then said to have remitted that money and the money she sends is described as a remittance. Remittances do not necessarily take the form of money – they could be in the form of easy to transport high value goods, such as jewelry or they could be in the form of consumer goods more generally. Remittances might not even be goods, they might be ideas, practices, or values – in which case they are referred to as social remittances. Remittances are not limited to capital, goods, or ideas that are sent internationally. For example if a school leaver moves from the hometown where they grew up to the capital of the country where they live to find work and they then send money back to their family in their hometown, then that money is also described as a remittance. However, most attention at the moment is paid to international remittances of money.

The practice of sending remittances is not new, but there is currently particular interest in the remittances sent back by workers who have migrated from the poorer countries of the world to the richer countries of the world and who are sending money back to the countries from which they originally came. These remittances are of interest because they constitute a large, growing, and stable flow of capital into the poorer countries of the world, and as such remittances are an absolutely crucial element in understanding the relationship between international migration and development.

The language of remittance studies is more complex than it initially appears and merits reflection. Most research is built around the central idea of a 'migrant' who moves from one country to another. The term 'migrant' implies an autonomous individual whose behavior is driven by a set of explicable determining factors, yet research has shown not only that it is misleading to separate the individuals who migrate from their wider family networks, but also that character, emotion, and caprice can be central to whether or not a particular person will send remittances. Furthermore, in an effort to keep it clear which of the two countries is being referred to authors often describe the place migrants have left as the 'home' country and the country they have moved to in search of work as the 'host' country. The terms 'home' and 'host' may appear to be a simple device used to clarify the story of particular groups, but they reveal a lot of the assumptions that underpin remittance analysis. These metaphors work to reinforce a set of politically contentious ideas. For example, they suggest that this migration is temporary and partial because they imply that the migrant's only real home is where they began their journey. They assume that if migrants are still sufficiently emotionally attached to the country from which they originally came to send remittances back, then they cannot be fully assimilated into the society of the country which they have joined. Such assumptions are contested in the literature on transnationalism, which considers those individuals with simultaneous and ongoing affinities to more than one nation. Equally the metaphor 'host' implies that these migrants are guests in the country to which they have moved, which carries certain implications about the way they should behave (their rights and their responsibilities). It is implied that they do not properly belong in the country to which they have moved. While it is necessary to have an agreed terminology to use when talking about remittances it is also necessary to reflect critically on the ideological work done by the terms that are being used.

The History of Remittances

The English term remittance first came into use in the eighteenth century. In the nineteenth century it was used to describe the regular allowance that was sent to individual colonists by their families back in Britain. Very often these colonists were the younger sons of aristocrats and wealthy families who were either being subsidized from the metropole in their colonial endeavors or were being paid to stay away and not return. Such individuals were described in derogatory terms as 'remittance men'. However, though the specific term may be relatively new the practice of sending money over long distances is very old. For example, in Tang Dynasty China (618–907 CE) provincial merchants travelled without carrying metal coinage but instead used *fei ch'ien* (literally 'flying currency'), in order to avoid the risks of robbery on long journeys. They brought goods into the capital from the provinces, which they sold. They then deposited their profits with a representative of the provincial government based in the capital (who used this money to pay tax obligations to central government). The merchant was issued with a paper certificate stating the value of the deposit. They then returned to the regions without the hazard of carrying money and were able to exchange the paper certificate for cash in a provincial government treasury. This practice was subsequently privatized and (in the nineteenth century) internationalized as Chinese families became established in two different places and issued the certificates to remit money first within the family and subsequently as a commercial venture.

Probably the better known historic remittance system is *hawala* (meaning 'transform' or 'change' in Arabic) or *hundi* (meaning 'collect' in Sanskrit). This system originally operated along the trade routes of the Indian Ocean, the Silk Road, the Eastern Mediterranean, and North Africa. The earliest evidence of *hawala* comes from an eleventh century Islamic legal treatise (*Al mabsut*) written by the Egyptian scholar Sarakhshi, but this does not exclude the possibility that the system was already operating much earlier. Again the system evolved as a means of minimizing the risk of carrying money during long distance travel. This remittance system remains active and popular today across the globe as a means of transferring money without incurring either the high charges of commercial bank transfers or the taxes of governments.

Hawala works in the following way. The person who wishes to remit money deposits it with a *hawaladar* (broker) who then contacts another *hawaladar* in the place where the person remitting the money would like it to be collected. An agreement is struck between the two brokers who also place a small charge on the transaction (usually between 0.25% and 1.25% of the total transaction) and the person for whom the money is intended can subsequently withdraw money in a different place. In the contemporary situation the discussion between the two *hawaladars* may take place via phone, fax, or e mail and some kind of 'collection code' may be given to the clients. Historically some kind of written note was used and then physically transported. The money does not move at all and so over time it is necessary for the two *hawaladars* to balance up their own accounts.

The *hawala* system works because of trust within the networks of *hawaladars*, which were (and are) very often kin based. Deceit by any one *hawaladar* was prevented through the potential social sanctions that would follow. Because of this emphasis on trust much is made in many descriptions of *hawala* of the absence of written documentation; however, the reality is that because of the varied and complex exchanges it is necessary to keep a written, if very private, record of the balance of accounts between different points within the network. Despite its marginalization in the nineteenth century; *hawala* remains popular because it is cheap, flexible, and quick. It has, however, been cast under suspicion in some quarters because of accusations that it is used for money laundering and funding terrorism and in recent years rather unsuccessful attempts have been made to regulate *hawala* in the US and UK.

Contemporary Remittance Channels

There are four other ways, apart from *hawala*, in which money is currently generally remitted internationally. First, it can be carried by hand by a migrant (or trusted friend or family member) who is going to the country where the migrant would like the money to be sent. This has the advantage of incurring no taxes or transaction charges (if one discounts the costs of human transport), but, as throughout history, it carries the risk of theft. In addition, it may be illegal if there are restrictions on either importing or exporting currency. It used to be assumed that hand carries were limited only to the very poor; research has shown this is not the case. Second, cash can be sent through the mail. This is a very high risk means of transporting money because of theft within the postal system, but empirical work with Mexicans in the US suggests that it is surprisingly popular because it is simple and cheap. Third, there are electronic transfers between bank accounts or via post offices and credit unions. These have the advantage of security and, theoretically, speed. However, the charges can be quite substantial, and in the case of banks and credit unions it is often required that senders have bank accounts. Studies suggest that quite a significant proportion of new migrants do not have bank accounts and are intimidated by banks, which makes this form of transfer mechanism less popular. Finally, there are the specialist commercial money transfer companies, such as Western Union. These are dedicated companies whose primary function is international money transfer. This is generally the most expensive way to transfer money, but has the advantage among commercial electronic mechanisms of not requiring a bank account or extensive identification from the sender. These companies also have very wide global networks of agencies where money can be collected, which means they are very popular despite the cost.

Government policies have attempted to steer remittances through particular channels. If the individuals remitting money from the richer countries of the world switched to cheaper services, it could redirect several billion dollars to families in poor countries, money that is currently taken as profits by money transfer companies and banks. From about 2000 onward governments in the richer countries of the world have tried to foster competition in the commercial money transfer market, often by increasing the flow of information about the different costs of sending money to migrants. Legislation (e.g., the 2001 Wire Transfer and Disclosure Act in the US) has tried to increase transparency in this market and commercial banks have also experimented with financial education programs. However, another potential area for savings that has been advocated by diaspora lobby groups themselves would be to offer tax relief to individuals who remit money on a regular and transparent basis. Here Western governments have shown less inclination to take action. Governments in those poorer countries that are receiving remittances would also prefer to see money passing through visible channels and offer incentives (such as access to foreign currency accounts) to encourage migrants to send their money through banks.

Data Collection and Reliability

The key central source for contemporary data on measuring and mapping international remittances at national and global scales is the *Balance of payments statistics yearbook* of the International Monetary Fund (IMF). While the IMF aspires to standardize the way this data is collected and compiled by individual governments around the world the reality is that data coverage at the present time is inconsistent and incomplete and the figures for remittances that are widely circulated by agencies, academics, and journalists have to be treated with some caution. Furthermore, while this data enables analysts to make reasonable estimates of the volume of remittances coming into countries, it is much less useful when trying to calculate remittances going out of countries; so there is a much better idea of where international remittances are going to than where they are coming from.

Most calculations of remittances into a country are based on the sum of three components from the balance of payments statistics: (1) the gross earnings of workers who are registered as living abroad for less than 12 months, (2) the money transfers sent by workers residing abroad for more than 12 months, and (3) the net wealth of migrants who move from one country of employment to another. This sum attempts to capture both unrequited transfers and repatriated deposits and treats both as remittances. This method of calculating total remittances may overestimate the result because not all the earnings of those overseas for less than 12 months will be remitted since some will have to be spent on subsistence. However, the general assumption is that this method of calculating remittances is much more likely to underestimate the total not only because it fails to include informal or unregistered international capital transfers (such as hand carries of cash and *hawala* transfers) but also because some remittances (such as jewelry) are recorded in the statistics as imports. Efforts to improve the quality of the data are being driven by the increasing interest in international remittances; however, the complexity of the task and the cost of data collection remain significant obstacles.

The Contemporary Significance, Scale, and Geography of Remittances

Geographers and other academics have become increasingly interested in remittances since the 1980s because of a growing awareness among those studying the poorer countries of the world that the remittance gain into a country is closely related to the loss of labor that results from international migration (known generically as 'brain drain'). Observations suggest that remittances from overseas are a vital source of hard currency for many families across the world. It is also generally assumed that receiving remittances benefits national economic growth, and for these reasons remittances are thought to be significant. Since attention has turned to quantifying and mapping these flows, the volume of remittances into developing countries has continued to grow steadily, while foreign direct investment and borrowing on international capital markets have been more volatile. At a global scale recorded remittances are now significantly larger than overseas aid flows (**Figure 1**) and add up to an annual flow of somewhere between 150 and 200 billion US dollars into the developing world.

The gross level of remittances has increased in recent years for a number of reasons. First, following heightened security alerts, particularly in the US, analysts have started to look for remittances more carefully, and as a result they have become more visible. Second, the total number of international migrants has increased. In 1970 there were around 80 million international migrants, in 2005 there were around 200 million international migrants, around half of whom are women. However, this increase in total numbers does not mean that the number of international migrants as a proportion of world population has changed significantly – indeed around 3% of the world's population today are international mi grants, not much more than a century ago. Third, high oil prices in recent years have boosted remittances from the Gulf States, Russia, and Kazakhstan. Remittances are important because they are a major, growing, and stable flow of capital into the developing world.

The geography of this global story, which is generally analyzed at the scale of nation states, requires some careful unpicking. The first point to make is that while remittances may be of particular economic significance in poorer countries it is wrong to assume that the flow of remittances is entirely from North to South. In gross terms India, China, and Mexico are the world's largest recipients of remittances, but France, Spain, Belgium, Germany, and the UK are also in the top ten remittance receiving countries. Even though the bulk of these European countries' remittances are from individuals employed outside their own country, but working within the European Union (EU), it is a salutary reminder that remittances are not only a development issue. More importantly remittances between countries in the global South can be very significant. Remittances from the Gulf States to the Middle East, Asia, and Africa would be an obvious example.

The second point to make is that the presumed benefits that flow from remittances are unequally dis tributed around the world between regions (**Table 1**). Despite the fact that approaching half the world's re mittances go into poorer Asian countries, the region has the lowest level of remittances per capita. Eastern Europe and the Middle East on the other hand, which benefit from proximity and, in some cases, easy access to Euro pean labor markets have disproportionately high levels of remittances. Similarly, Latin America and the Caribbean benefit from access to North American labor markets. Within any of these regions it is also necessary to look carefully at differences between nation states because the pattern of remittances is highly uneven. For example, Mexico, which is close to the US, is in a much better position to benefit from remittances than Paraguay, which is much further away. Similarly Poland, which is a member of the EU is in a much better position to benefit from remittances than the Ukraine, which is not.

The third point to make is that when examining re mittances in per capita terms or in relation to GDP it becomes clear that they are a crucial source of income for

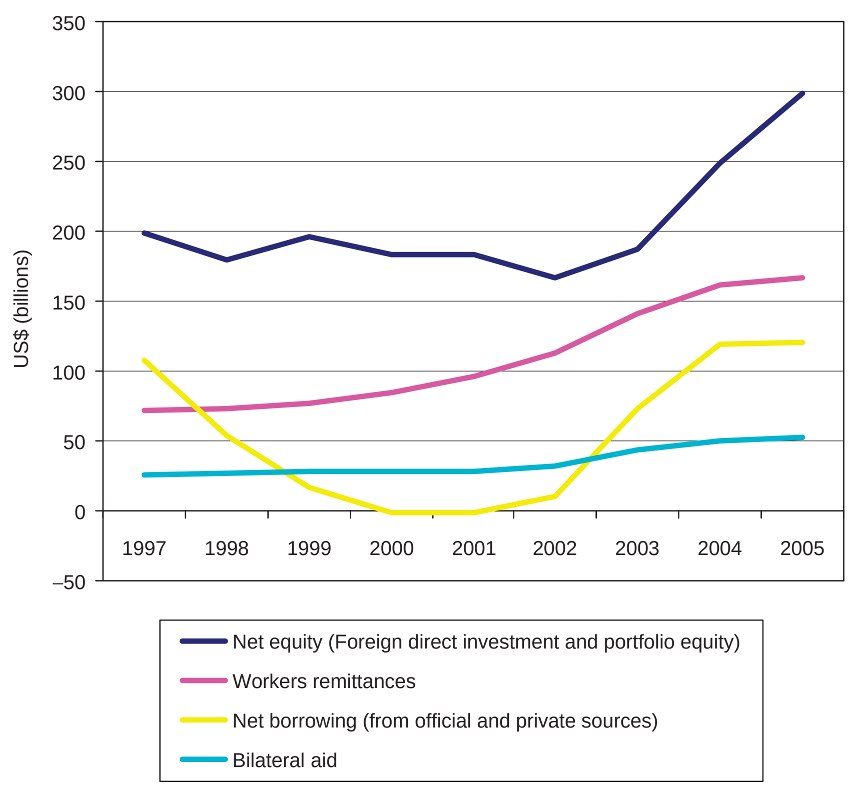

Figure 1 Capital flows into developing countries. From World Bank (2006). *Global development finance 2006: The development potential of surging capital flows.* Washington: World Bank.

Table 1 2002 share of global remittances by region in relation to population

Region	*Share of remittances going into developing world (%)*	*Share of world population (%)*
Africa	10	11
Asia	44	60
Eastern Europe	15	5
Latin America and Caribbean	21	8
Middle East	10	5

Source: International Monetary Fund Statistics Department (2004). *Balance of Payments Statistics Yearbook*, vol. 55. Washington, DC: International Monetary Fund.
United Nations Polulation Division Department of Economic and Social Affairs (2004). '*World Population Prospects: The 2004 Revision.*' New York: United Nations.

small and island nation states in the developing world, where they are often the single most important contri bution to the national economy. For example, in Tonga remittances represent over 40% of GDP. They also rep resent a significant proportion of GDP in the West Bank and Gaza, Lesotho, Jordan, Cape Verde, and Moldova. Fourth, in general remittances are of less significance to the low income countries than to middle income coun tries. The 2002 mean figure for all developing countries was US$ 28.53 remitted/capita of the population. Africa and Asia were below this average. Eastern Europe, the Middle East, and Latin America/Caribbean were above it. This is because there is more labor migration from middle income countries than from the world's poorest countries. Fifth, at a subnational scale the social geography of remittance distribution is also highly uneven because international migration is a selective process. International migrants tend not to come from the poorest families within a country because significant amounts of capital are needed to send a family member overseas. The result is that (at least initially) remittances are often assumed to exaggerate existing inequality between households and regions within a country, though the results from empirical studies trying to test this assumption are ambiguous.

The Developmental Impact of Remittances in Receiving Countries

Remittances are widely assumed to be good for the economy. Remittances reduce foreign exchange shortages in poorer countries and offset balance of payments def icits without incurring interest liabilities or necessarily increasing imports of foreign goods and services. Fur thermore, for those households that have access to them, there seems to be little doubt that remittances are an important source of income with welfare benefits. In the context of countries whose limited welfare services have been further stretched by neoliberal macroeconomic policies, such as the removal of subsidies and the intro duction of user fees for health and education services, remittances can be a crucial component of family bud gets. Household level studies have shown that money sent as remittances is spent in four basic ways by the recipients: consumption (including health and education costs), housing, land purchase, and productive invest ment. Productive investment is held to be unproblem atically good for national economic growth; however, most remittances are spent on consumption and real estate, the main benefits of which are indirect, for ex ample, multiplier effects that flow from injecting cur rency into local building trades. Nevertheless, putting money into local economies in this way stimulates de mand for goods and services. In addition, there is some evidence that remittances are being sent by diaspora groups, such as hometown associations, specifically for developmental projects in their home countries, for ex ample, water supplies, health clinics, and classrooms. However, on the negative side remittances can also have an inflationary impact, which can be significant in situ ations such as scares about potential food shortages, as occurred in Khartoum, Sudan in the late 1980s. While much research in this area remains to be done, the con sensus is that the net economic effects of remittances on recipient households and recipient countries are positive.

Why Do Migrants Remit?

Initial empirical observations suggest that there are a number of patterns that can be seen in the behavior of migrants who remit money, all of which invite explan ation. For example, it appears that migrant women remit a larger percentage of their earnings than men, it also appears that migrants from some countries or regions choose to remit more than migrants from neighboring countries or regions and it appears that migrants alter their remittance behavior in relation to political, en vironmental, and economic conditions in the country they are sending money to. Explaining such patterns is one of the main intellectual projects of remittance studies. Before outlining existing ideas it is necessary to introduce two caveats. First, while this is an interesting and valuable task it is also necessary to acknowledge that migrants are individuals whose choice of actions should not be reduced to a limited set of determining factors. It is also important to incorporate their own unique per sonal experiences and attitudes into the analysis. Much of the economic analysis of remittances casts explanation in very generalized, rationalized, and simplistic terms that allow little space for such influences as character or whim. Second, the focus of debate tends to be on the motivation of migrants to remit, but their ability to remit is equally crucial. Migrants with an income that only covers their subsistence needs will not be able to remit money, even if they want to.

For some time remittance behavior has been explained using the economic categories of altruism, self interest, and family loans. Migrants who are motivated to remit by altruism derive satisfaction from increasing the welfare of their relatives and family. Migrants who are motivated to remit by self interest focus either on keeping their family at home happy in order to secure their own future inheritance of assets or on sending money home in order to accumulate assets by investing in their home area ready for their own return in the future. The latter investments could be in real estate, financial assets, or social capital (to enhance prestige and political influence in the local community). A more nuanced conceptualization of migrants' remittance behavior puts it within the context of an extended family framework, in which remittances are (at least initially) seen as loan repayments to the family network which raised the capital sum necessary to enable migration (not only the costs of travel, but the costs of getting into a position of earning money – such as education). Remittances are a means to redistribute the benefits of migration across the family that initially invested in the project of sending a family member overseas. The family at home have leverage over the migrant and can, therefore, continue to expect remittances, as long as the migrant wishes either to return to his/her home country one day or the migrant needs the security from keeping in touch with his/her family. From the perspective of many international migrants to the richer countries this arrangement is experienced in doubly negative terms; not only are they often employed in badly paid, uninteresting jobs, but their relatives at home are continuously demanding remittances based on their false images of the migrant's new lifestyle.

While these economic models provide a basic framework from which to theorize migrants' remittance behavior there is considerable scope to add a much richer historical and geographical dimension to such work in order to capture the interplay of experiences and ideas that shape the behavior of any one individual migrant. Explaining the observation that women are more likely to remit a larger proportion of their income, for example, requires an understanding of the history of gender relations in particular places, and how these have changed. Interviews with recent African migrants in the UK suggest that there is a crisis of masculinity among this diaspora group because of increasing dependence within households on women's wages and the relative failure of men to attain lucrative or secure employment. Equally, understanding the differences in remittance behavior between African migrants and Eastern European migrants, for example, will also require investigating the different histories of the different places from which migrants come and the impact of those histories on migrants' values and aspirations. It is also important to get a better understanding of the different associational lives of different diaspora groups, the different relations between specific diasporas and the existing residents in the places where they now are, and the effect that these have on remittance behavior. Much of this project of understanding why migrants send remittances requires detailed ethnographic studies with which to supplement the existing basic economic models.

See also: Brain Drain; Development I; Development II; Diaspora; Foreign Direct Investment; Migration; Social Capital; Social Capital, Place and Health.

Further Reading

Blue, S. A. (2004). State policy, economic crisis, gender, and family ties: Determinants of family remittances to Cuba. *Economic Geography* 80(1), 63 82.

Buencamino, L. and Gorbunov, S. (2002). *Informal money transfer systems: Opportunities and challenges for development finance.* United Nations, Department of Economic and Social Affairs Discussion Paper Number 26. New York: UNDESA.

Conway, D. and Cohen, J. H. (1998). Consequences of migration and remittances for Mexican transnational communities. *Economic Geography* 74(1), 26 44.

Crisp, J. and Koser, K. (2005). *Migration in an Interconnected World: New Directions for Action*. Report of the Global commission on international migration. Geneva: GCIM.

Department for International Development (2006). *Moving out of poverty Making migration work better for poor people*. London: DFID.

Farrant, M., MacDonald, A. and Sriskandarajah, D. (2006). *IOM Migration Research Series No 22: Migration and development: Opportunities and challenges for policymakers*. Geneva: International Organization for Migration.

Ghosh, B. (2006). *Migrants' remittances and development: Myths, rhetoric and realities*. Geneva: International Organization for Migration.

De Haas, H. (2005). International migration, remittances and development: Myths and facts. *Third World Quarterly* 26(8), 1269 1284.

Jones, R. C. (1998). Remittances and inequality: A question of migration stage and geographic scale. *Economic Geography* 74(1), 8 25.

Lucas, R. and Stark, O. (1985). Motivations to remit: Evidence from Botswana. *Journal of Political Economy* 93, 901 918.

Poirine, B. (1997). A theory of remittances as an implicit family loan arrangement. *World Development* 25(4), 589 611.

Straubhaar, T. and Vadean, F. P. (2006). International migrant remittances and their role in development. In OECD (ed.) *International migration outlook*. pp 139 161. Paris: OECD.

World Bank, Global development finance 2006: The development potential of surging capital flows, 2006, World Bank, Washington.

Relevant Websites

http://www.livelihoods.org
Remittances and Migration, Livelihoods Connect, USAID/DFID.

http://www.gcim.org
The Global Commission on International Migration.

http://www.iom.int
The International Organization for Migration.

Remote Sensing

J. M. Read and M. Torrado, Maxwell School of Syracuse University, Syracuse, NY, USA

Glossary

Band An image layer containing reflectance data recorded for a specified range of wavelengths.

Electromagnetic Spectrum The range of wavelengths of electromagnetic radiation.

Georeferencing Relating locations on an image or map to real-world coordinates.

Multispectral An image with more than one layer (band) of reflectance data.

Pixel (picture element) The smallest unit of a remotely sensed image.

Raster Spatial data structure based on dividing an area into a grid of pixels.

Sun-Synchronous Orbit Orbit that follows the Sun such that it crosses the equator at the same time each day.

Introduction

Remote sensing provides information about objects at or near the surface of the Earth and atmosphere based on radiation reflected or emitted from those objects. The information is usually captured at a distance from above in the form of image data. Such data allow us to determine the composition and nature of the Earth's surface and atmosphere from local to global scales, and assess changes by analyzing images captured at different points in time. In this sense, remote sensing is useful in providing spatial information that is otherwise difficult or impossible to obtain. In the social sciences remote sensing is useful for visualizing (providing alternative and synoptic views) and classifying human environments. Social science researchers commonly integrate remotely sensed data or its derivatives with other socioeconomic data sets within geographic information systems (GIS) to conduct spatial analyses.

History of Remote Sensing

Modern day remote sensing can be traced back to 1858 when a French photographer and balloonist, Gasper Félix Tournachon (nicknamed Nadar), took the first aerial photograph from a balloon over Paris. It was not until 50 years later that Wilbur Wright was the first to take an aerial photograph from an airplane, beginning the highly successful era of aerial photography. Since then, remote sensing has evolved from aircraft based, analog, visible wavelength photographic systems capturing small areas of Earth's surface and accessible to a few specialists; to space based, digital, multi and hyper spectral systems with coverage of the entire Earth surface and available to most people with a computer. Aerial photography, however, still remains an important data source in much of the world's remote sensing applications.

During World War I, aerial photography was used for reconnaissance and surveillance missions. Subsequently, during the 1920s and 1930s, it was adopted primarily, but not exclusively, by government agencies for mapping applications, such as topographic mapping, soil and geologic mapping, and forest and agricultural inventories. Personnel were trained, and the science of photogrammetry (making accurate measurements from photographs) really came into its own. By 1934, the American Society for Photogrammetry (later renamed the American Society for Photogrammetry and Remote Sensing) was formed. With World War II came some of the first applications of photography using infrared and microwave radiation, after which, applications of aerial photography expanded beyond mapping to assessment of conditions of different land cover types, such as crop health, improved characterization of vegetation, assessment and prediction of crop yields and nutrient uptake for precision farming, soil conditions, and environmental hazards mapping.

Space based remote sensing began in 1960, when TIROS 1 (Television and Infrared Observation Satellite), the first experimental meteorological and weather satellite, was put into orbit by a group of US government agencies. The only other space satellite programs at that time were classified military programs, such as the United States' Corona system, which was operational until 1972 and declassified in 1995. In the absence of other data sources, manned space craft provided important photographs for civilian purposes, although none were suitable for scientific analysis. In the 1960s, the United States' National Aeronautics and Space Administration (NASA) began a remote sensing research program named Earth Resources Technology Satellites (ERTS), later renamed Landsat (land satellite). Landsat 1 was the first space based satellite to capture digital multispectral data of much of the Earth's surface, with predictable and repetitive coverage available for civilian uses. With the availability of digital data, digital image processing and quantitative analysis methods and algorithms could be developed; perhaps the most useful of these are automated classification algorithms, indices, and statistics.

Today, many raster based GIS analyses incorporate digital remotely sensed data. In addition to national and international space satellite programs, such as those of the United States, Canada, India, Brazil, Japan, Russia, the European Space Agency, and others, several commercial systems are operational, with more planned and under development. Hyperspectral systems were first developed in the 1980s, allowing for capture of fine detail about Earth's surface through spectroscopy. In the 1990s, satellites that capture data from the entire surface of the Earth were put into orbit, and in 1999 the first commercial satellite sensor was launched, providing much finer detail than was previously available. Today, radio detection and ranging (RADAR) and light detection and ranging (LIDAR) systems are increasingly being used for mapping terrain and other landscape features. There are many social science applications of Earth resources satellites, including urban mapping, public health, rural and urban planning, precision agriculture, land use analyses, resource use, transportation, disaster management, demographic and economic studies, and studies of human–environment interactions. Nowadays, increased data availability and Internet based applications, such as Google Earth and Microsoft's Virtual Earth, provide lay users with access to remotely sensed data sets, some with near real time availability. The United States Geological Survey (USGS) National Map and other data clearinghouse sites also provide access to remotely sensed data sets; and online visualization and mapping tools, such as the USGS Global Visualization Viewer (GloVis), allow users to find and browse imagery.

Relationship of Remote Sensing to Other Geographic Information Technologies

While many remote sensing operations involve visual interpretation of imagery, digital image processing is commonly employed for classification and other analyses. In addition, global positioning system (GPS) technologies are frequently used to assist with georeferencing remotely sensed imagery to allow points on an image to be associated with locations on the ground or with other georeferenced data sets. In addition, because remote sensing represents an important source of environmental data, it is often imported to a GIS for analysis and integration with other data sets. Thus, remote sensing specialists will commonly use not only remote sensing, but also GPS and GIS technologies in their research.

Electromagnetic Radiation

All real objects above absolute zero temperature reflect and emit radiation. The measurement and interpretation of emitted and reflected electromagnetic radiation from Earth's surface represents the foundation of remote sensing. In remote sensing, radiation is typically measured and categorized based on wavelength using a logarithmic scale. The range of wavelengths is known as the electromagnetic spectrum (**Figure 1**). Remote sensing uses physical principles of radiation to determine characteristics about an object emitting or radiating at a specific wavelength. Most remote sensing employs passive systems that capture the Sun's reflected or emitted light. When remote sensing scientists talk about visible, infrared (IR), and microwave radiation, we are referring to arbitrarily designated regions of the electromagnetic spectrum (**Figure 1**). In remote sensing, we measure reflected solar radiation in the ultraviolet, visible (i.e., visible to the human eye), and near IR (NIR) wavelengths; reflected and emitted solar radiation in mid IR; and emitted radiation only in the far IR (in the form of thermal energy) and microwave wavelengths. Active systems, as opposed to passive, such as RADAR and LIDAR, generate and measure returns of their own electromagnetic radiation.

Energy interacts with both the atmosphere, as it travels through, and with the surface of the Earth that is being sensed. Electromagnetic radiation is scattered, absorbed, and refracted as it passes through the atmosphere. For satellite sensors, light must travel large distances through the atmosphere, and in the case of reflected light, it must travel through the atmosphere twice (from the Sun to Earth's surface, and from Earth's surface to the sensor) (**Figure 2**). Some important atmospheric effects resulting from scattering and absorption of light are wavelength dependent, and determine how much energy of a specific wavelength is able to pass through the atmosphere and arrive at the sensor. Shorter wavelengths, for instance, tend to be scattered more by the atmosphere than longer wavelengths, and as a result are more affected by haze and clouds.

The degree to which energy is reflected, absorbed, and/or transmitted when it interacts with Earth's surface features depends not only on the feature itself but also on its wavelength. In the case of reflected energy, the amount of reflected light, known as spectral reflectance, will depend on surface roughness as it relates to wavelength, and angle of incidence. Earth's surface features have distinct spectral reflectances as a result of the differing amounts of energy that they reflect, absorb, or transmit at different wavelengths. Due to atmospheric and other (e.g., temporal) effects on spectral reflectances, it is not possible to define a precise spectral reflectance value for an object under normal remote sensing conditions. As a result, in remote sensing we think in terms of spectral response patterns. It is through these spectral response patterns that different objects and conditions can be distinguished (**Figure 3**). **Figure 4** illustrates how spectral responses of different features vary with wavelength. Water absorbs

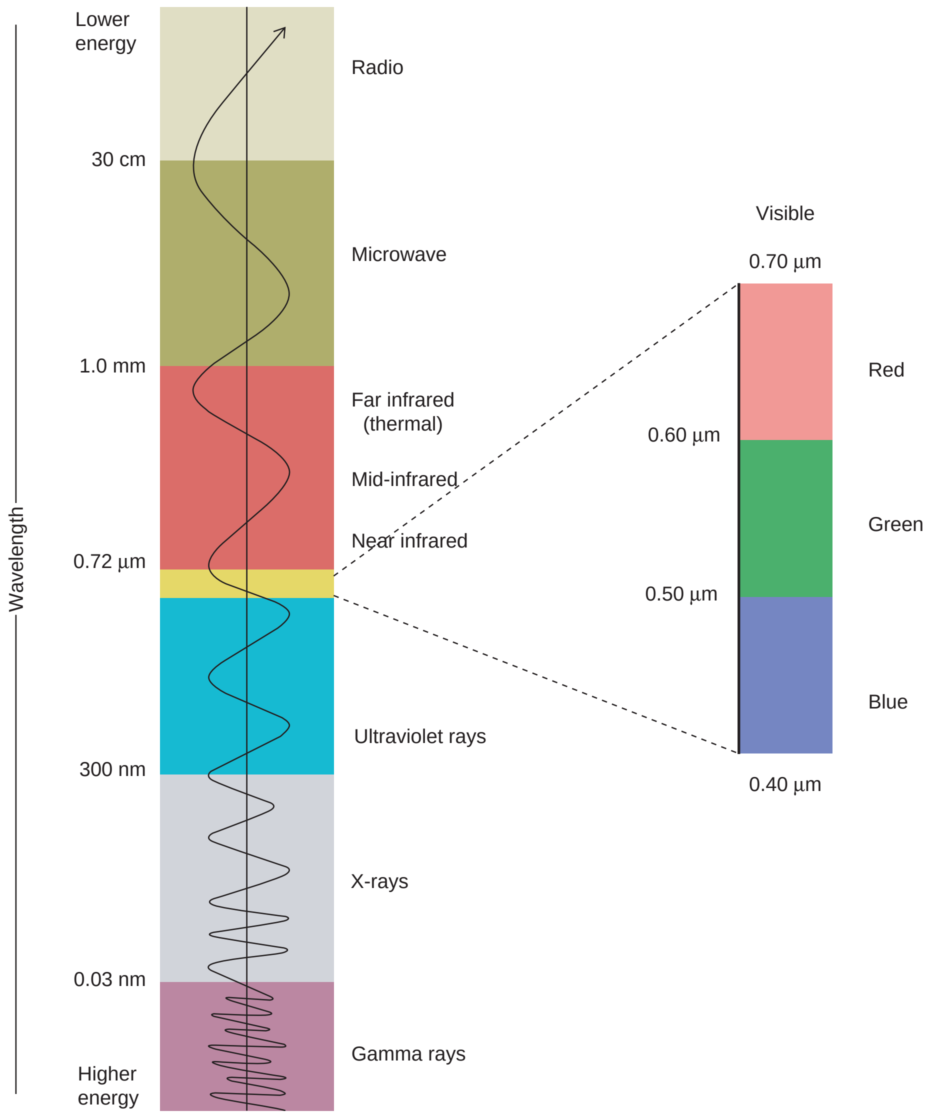

Figure 1 The electromagnetic spectrum, showing arbitrary divisions use in remote sensing.

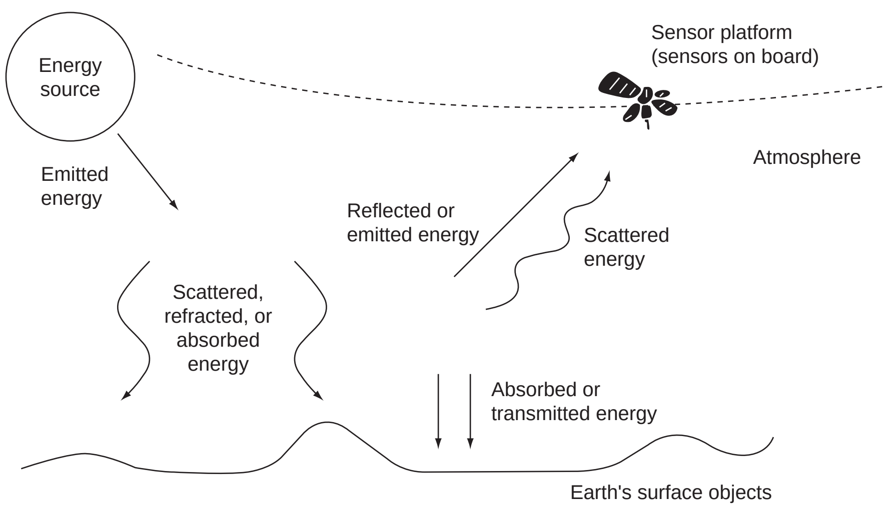

Figure 2 Energy pathways and components of a typical remote-sensing system.

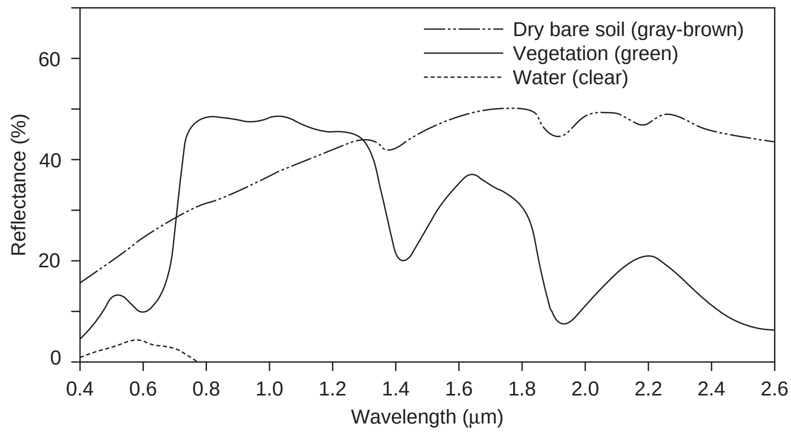

Figure 3 Typical spectral reflectance curves for vegetation, soil, and water. Reproduced from Lillesand, T., Kiefer, R. and Chipman, J. (2004). *Remote Sensing and Image Interpretation* (5th edn.). New York: Wiley, with permission from John Wiley and Sons.

incident energy in NIR and longer wavelengths, as evidenced by the difference in the response patterns of the deep water of Onondaga Lake (A) and shallower water of the Inner Harbor (B). Healthy vegetation, on the other hand (as demonstrated by the spectral response of the park at C), absorbs in the visible (and particularly the green portion of the spectrum) and reflects strongly in the NIR wavelengths. Note also that there are differences in the responses at different wavelengths of the mall buildings and its surrounding parking lot at D.

Characteristics of Remotely Sensed Data and Data Collection

There are three basic components for a remote sensing system: a light source (usually the Sun's energy), a sensor flown on a platform to capture and measure the reflected or emitted light, and a recording medium, which may be analog or digital (**Figure 2**).

Traditional aerial photography is flown in an airplane equipped with a camera with a lens designed to refract and focus the light onto photographic film. By flying overlapping flight lines, a mosaic of stereoscopic aerial photographs can be recorded for an area. Photographic film may be either panchromatic or color. The altitude of the airplane determines the spatial resolution (see below) of the image.

Digital sensors record photons for a specific range of wavelengths, and store digital numbers (also referred to as brightness values or digital counts) representing brightness as an array of pixels (**Figure 5**). The instantaneous field of view (IFOV), which is determined by the height of the sensor, relates to the area of ground being sensed (ground resolution) and defines the finest level of detail (**Figure 6**). With multispectral data, different regions of the electromagnetic spectrum are recorded simultaneously to generate separate bands (layers) to make up an image (**Figure 6**).

Resolution

Resolution refers to the amount of detail, or ability to distinguish different objects, in a remotely sensed image. Resolution is determined by the characteristics of the remote sensing system as a whole, including the characteristics of the sensor(s) and the orbital parameters. There are four types of resolution (spatial, spectral, radiometric, and temporal) that determine the detail captured and displayed.

Spatial resolution determines the smallest object that can be detected, and is defined by the IFOV, or ground resolution (**Figure 6**). This is most often referred to in terms of pixel size. For instance, advanced very high resolution radiometer (AVHRR) with 1.1 km × 1.1 km pixels may be considered a coarse resolution sensor; Landsat's Thematic Mapper (TM) with 30 m × 30 m pixels, medium resolution; and the IKONOS sensor with 4 m × 4 m pixels, fine spatial resolution (**Table 1**). **Figure 7** illustrates how the increased spatial resolution of IKONOS data allows smaller features, such as the parking garage and university buildings to be resolved, compared with Landsat TM data where only larger features, such as the sports dome, can be identified. Resolution is relative, however; some researchers working with 2 cm × 2 cm pixels might not consider a 4 m × 4 m pixel to be so fine. Social scientists interested in detecting individual buildings in an urban setting would need to select a sensor with spatial resolution similar to IKONOS or QuickBird data; those wanting to determine urban versus rural land cover would likely need a coarser sensor, such as Landsat TM.

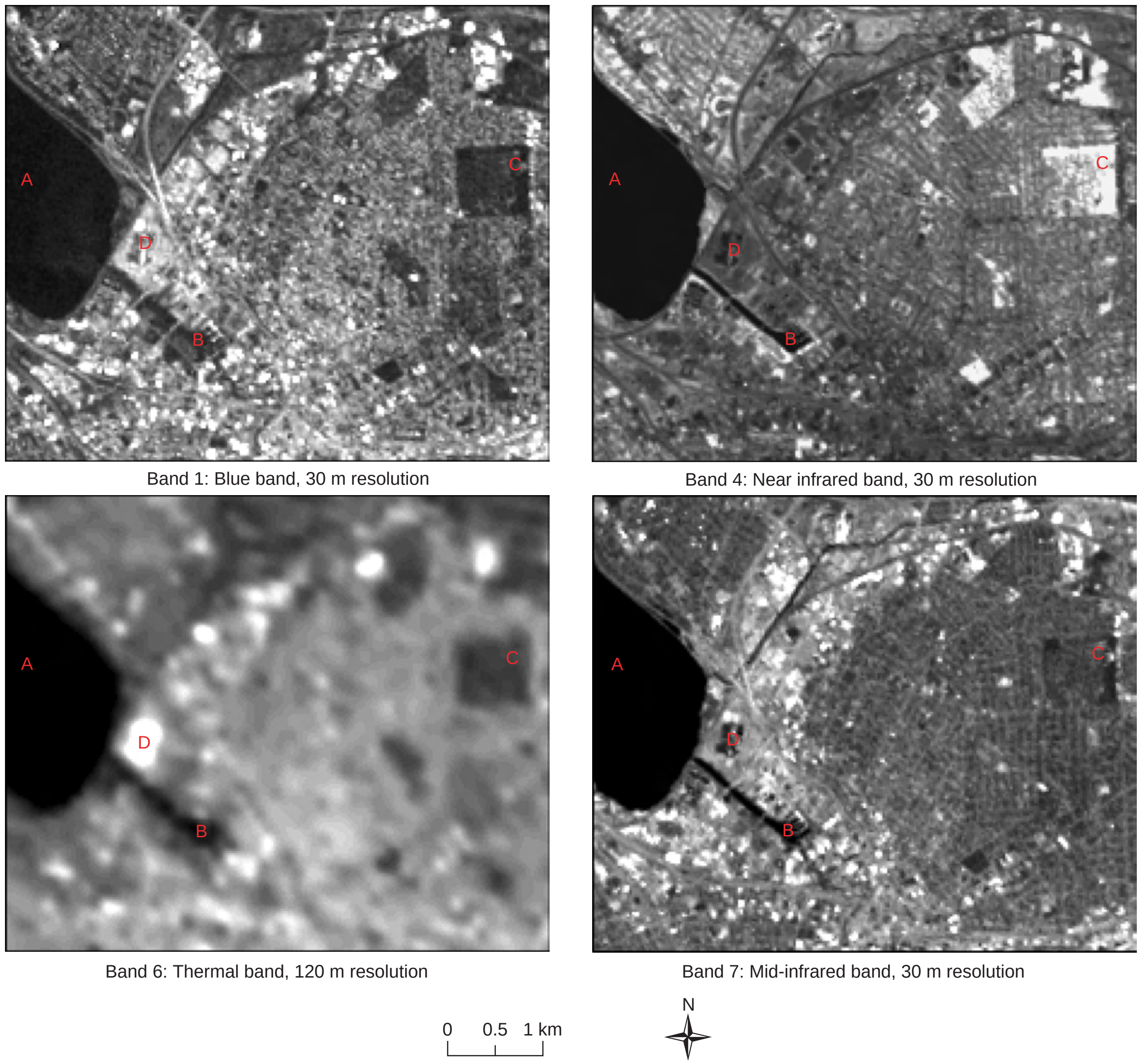

Figure 4 Selected bands of a Landsat Thematic Mapper image of an area of the City of Syracuse, New York (2000), demonstrating the different spectral responses of land-cover types: A deep water at the south end of Onondaga Lake; B shallower water in the Inner Harbor; C vegetation; D mall buildings and surrounding parking lot. Dark shades represent low reflectivity; light shades represent high reflectivity.

Spectral resolution is defined by the number and width of spectral bands recorded, in addition to the sensitivity of the sensor to record the number of photons across the entire bandwidth. More narrow bands provide finer spectral resolution. Landsat TM has seven bands; IKONOS has four bands; hyperspectral sensors typically may have more than 200 very narrow bands (cf. Landsat's seven bands).

Radiometric resolution refers to the number of po tential brightness values that can be stored for a pixel. Finer radiometric resolutions store and display more subtle differences in brightness values than lower reso lutions, allowing finer detail of features to be recorded. Landsat TM scenes are stored as 8 bit data, which provide 256 possible brightness values per pixel; whereas IKONOS and QuickBird data have finer resolution 16 bit data providing 65 536 possible values per pixel.

Temporal resolution relates to a sensor's repeat cycle, which determines how frequently the sensor images a location. This is important for applications that seek to uncover change through time, especially in cases of natural disasters, for example, where frequent updates are needed. The Moderate Resolution Imaging Spectro radiometer (MODIS) images the entire Earth every 1 or 2 days, whereas Landsat repeats once every 16 days

90	91	90	86	82	78	78	82
89	100	94	85	79	78	82	86
91	102	100	89	80	75	81	84
97	89	81	79	83	82	77	78
92	81	78	86	98	101	80	78
83	77	87	94	114	117	79	76
80	82	91	84	86	100	99	90
77	85	81	66	66	88	104	105

(a)

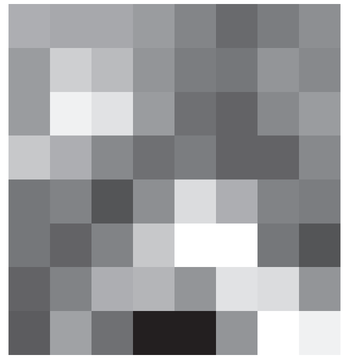

(b)

Figure 5 (a) An array of brightness values stored as individual pixels, and (b) a corresponding image shown in grayscale (low to high values range from dark to light shades).

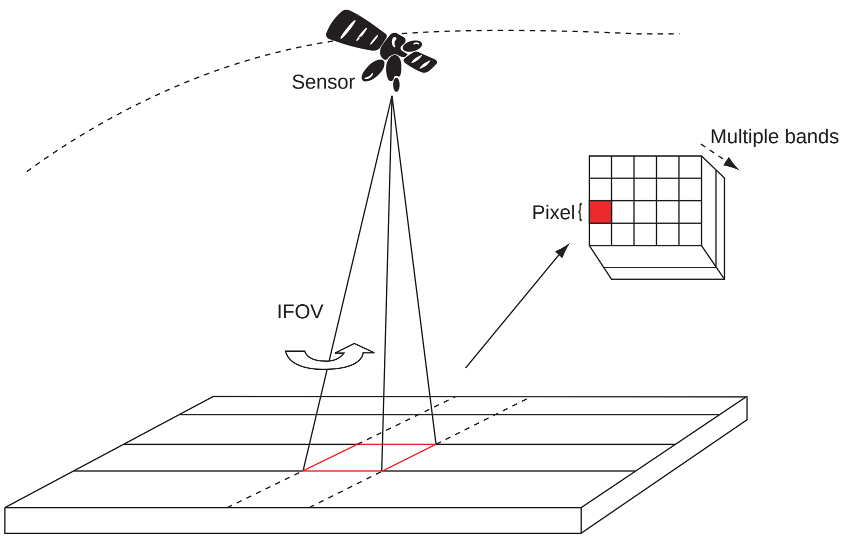

Figure 6 Schematic showing the relationship between the instantaneous field of view (IFOV) of the sensor and the pixels and different bands of a multispectral image.

(**Table 1**). Some sensors have pointing capabilities which means that the frequency can increase.

A remote sensing system has its own unique set of resolutions that determine image coverage (size and frequency), radiometric, spectral, and spatial resolutions. There tend to be technological tradeoffs between the different types of resolution.

The amount of detail in an image will also depend on the characteristics, configuration, and context of the objects in the scene. For instance, two land cover classes with similar spectral reflectance patterns may not be distinguishable regardless of the configuration of the sensing system. Or mixed pixels (pixels that represent more than one land cover type or object) may confound an analysis. Increased spatial resolution may lead to fewer mixed pixels, but finer resolution may also resolve more features, thereby leading to more mixed pixels. Thus, the amount of detail captured in a scene depends on a number of different factors relating to the remote sensing system and the scene being imaged.

Digital Image Processing and Analysis

While visual display of remotely sensed data can provide a powerful tool for many applications by aiding analysis of analog or digital data, the speed at which operations can be performed confers distinct advantages to digital image processing. Many different operations can be performed on digital imagery. Essentially, image values are transformed by passing each pixel through a process that produces a new image with new pixel values. The operations may range from very simple, to combinations of more complex operations.

Digital image processing often involves several steps, including: (1) preprocessing operations (such as

Table 1 Character st cs of se ected remote sensors

Sensor system	*Sensors*	*No. of bands*	*Nominal pixel size*	*Orbit height (km)*	*Repeat cycle (days)*	*Nominl swath width (km)*
High-resolution sensors						
~Qu ckB rd (2001–)	Panchromat c	1	0.61 m	450	1–3.5	272–435
	Mu t spectra	4	2.40 m			
~ KONOS (1999–)	Panchromat c	1	1 m	681	<11	11
	Mu t spectra	4	4 m			
~SPOT 1 (1986–90) ~SPOT 2 (1990–) ~SPOT 3 (1993–)	2 × h gh-reso ut on v s b e (HRV)	4	1 × 10 m (pan); 3 × 20 m (× S)	822	26	60–80
~SPOT 4 (1998–)	2 × h gh-reso ut on v s b e and nfrared (HRV R)	4	1 × 10 m (pan) and 3 × 20 m (× S)	822	26	60–80
	Vegetat on 1	4	1000 m		1–2	2250
~SPOT 5 (2002–)	2 × h gh-reso ut on geometr c (HRG)	6	2 × 5 m (pan); 3 × 10 m (× S); 1 × 20 m (N R)	822	26	60–80
	1 × h gh-reso ut on stereoscop c (HRS)	1	5–10 m (pan)			60–120
	Vegetat on 2	4	1000 m		1–2	2250
~NASA/JPL (1989–)	A rborne v s b e/ nfrared mag ng spectrometer (AV R S)	210	4–20 m depend ng on a t tude			11
~(1989–)	Compact a rborne spectrograph c mager (CAS)	288	Var es depend ng on a t tude			
~NASA/JPL	Therma nfrared mu t spectra scanner (T MS)	6	7.6 m			
~Japanese Earth Resources Sate te (JERS-1) (1992–98)	Opt ca sensor (OPS)	7	18 m	570	44	75
	Synthet c aperture radar (SAR)					
Moderate-resolution sensors						
~Landsat 1 (1972–1978)	Return beam v d con (RBV)	3	80 m	900	18	185
~Landsat 2 (1975–1982)	Mu t spectra scanner (MSS)	4	56 m × 79 m			
~Landsat 3 (1978–1983)	Return beam v d con (RBV)	1	30 m	900	18	185
	Mu t spectra scanner (MSS)	5	56 m × 79 m			
~Landsat 4 (1982–1993)	Mu t spectra scanner (MSS)	4	56 m × 79 m	705	16	185
~Landsat 5 (1984–)	Themat c mapper (TM)	7	Bands 1–5 and 7: 30 m Band 6 (therma): 120 m			
~Landsat 7 (1999–)	Enhanced themat c mapper (ETM+)	8	Bands 1–5 and 7: 30 m Band 8 (pan): 15 m Band 6 (therma): 60 m	705	16	185

(*Continued*)

Table 1 Cont nued

Sensor system	Sensors	No. of bands	Nominal pixel size	Orbit height (km)	Repeat cycle (days)	Nominl swath width (km)
~ nd an Remote Sens ng (RS)-1A (1988–) and RS-1B (1991–)	L SS- L SS-	4 4	72.5 m 36.25 m	904	22	148 146
~ nd an Remote Sens ng (RS)-1C (1995–) and RS-1D (1997–)	L SS- Panchromat c W de fie d sensor (W FS)	4 1 1	3 × 23 m and 1 × 70 m 5.8 m 188 m	817	24	142–148 70 774
~RADARSAT-1 (1995–)	Synthet c aperture radar (SAR)	4	8–100 m	798	24	45–510
~EO-1 (2000–)	Hyper on	220	30 m	705	16	7.5
Coarse-resolution sensors						
~EOS Terra (1999–) and Aqua (2002–)	Moderate reso ut on mag ng spectro-rad ometer (MOD S)	36	250–1000 m	705	2	2330
~NOAA m ss ons 6–17 (1979–2002)	Advanced very h gh reso ut on rad ometer (AVHRR)	5	1.1 km	833–870	4–9	2400
~Sea-v ew ng W de F e d-of-v ew Sensor (SeaW FS) (1997)	Loca area coverage (LAC) G oba area coverage (GAC)	8	1.13 km Subsamp ed every 4th ne and 4th p xe	705	var es 2	~2800

pan = panchromatic × S = multispectral NOAA = National Oceanographic and Atmospheric Administration NASA = National Aeronautics and Space Administration JPL = Jet Propulsion Laboratory

Sources

Campbell J (1996) *Introduction to Remote Sensing* (2nd edn) New York Guilford Press
Erdas (1999) *Erdas Field Guide* (5th edn) Georgia ERDAS nc
Digital Globe Homepage *Standard Imagery* Retrieved March 2007 from http //www digitalglobe com/product/standard_imagery shtml
Geo mage Website Retrieved October 2006 from http //www geoimage com au/
Geospatial Data and nformation System *Spot Satellite Imagery* Retrieved October 2006 from http //www geovar com/spot htm
Lillesand T Kiefer R and Chipman J (2004) *Remote Sensing and Image Interpretation* (5th edn) New York Wiley
Lindborg C Update April 20 2000 *IRS (Indian Remote Sensing)* Retrieved October 2006 from http //www fas org/spp/guide/india/earth/irs htm
Modis Web *About Modis Components* Curator Maccherone Brandon Retrieved October 2006 from 〈http //ladsweb nascom nasa gov/〉
NASA Earth Observing-1 Homepage Retrieved March 2007 from http //eo1 gsfc nasa gov
NOAA Coastal Services Center Update June 30 2006 *About LIDAR Data* Retrieved March 2007 from http //www csc noaa gov/products/sccoasts/html/tutlid htm
NOAA Satellite and nformation Service Update March 27 2007 *Advanced Very High Resolution Radiometer – AVHRR* Retrieved October 2006 from http //noaasis noaa gov/NOAAS S/ml/avhrr html
SeaWiFS Project Homepage *An Overview of SeaWIFS and the SEASTAR Spacecraf; Spacecraft Description* Retrieved October 2006 from http //oceancolor gsfc nasa gov/SeaWiFS/SEASTAR/SPACECRAFT html
SPOT mage webpage Retrieved March 2007 from http //www spotimage fr
USGS Update April 6 2005 *Thermal Infrared Multispectral Scanner (TIMS) Sensor/Instrument* Retrieved March 2007 from http //eosims cr usgs gov 5725/sensor_documents/tims_sensor html#1

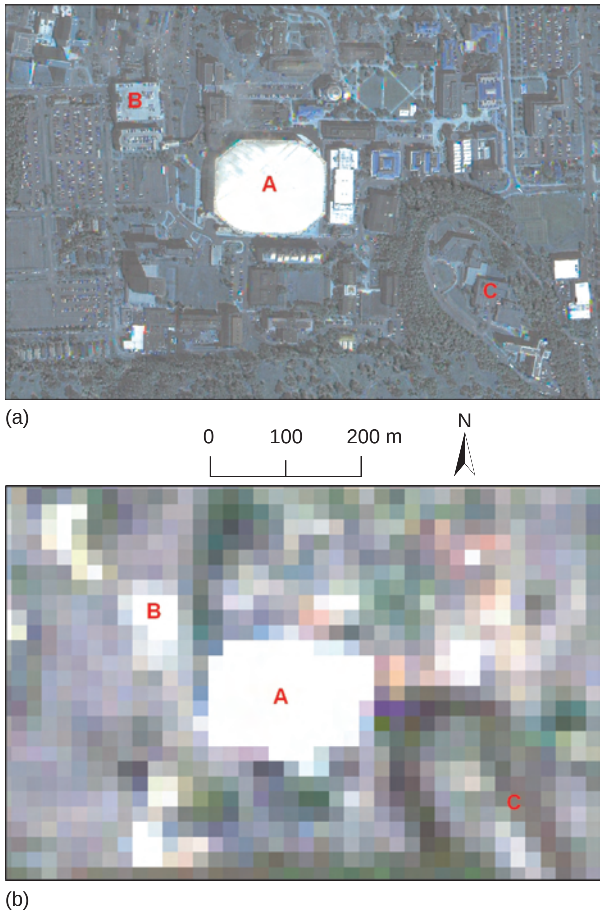

Figure 7 (a) 1 m pan-sharpened IKONOS data of the City of Syracuse, New York, USA (2001), (b) 30 m Landsat Thematic Mapper data of the same area (2000): A sports dome, B parking garage, C Syracuse University buildings.

georeferencing, correction of atmospheric, sun angle, and sensor effects, or conversion from brightness values to radiances); (2) feature extraction (methods to highlight and isolate specific features in an image); (3) data fusion (merging different types of data to take advantage of the characteristics of both data types); and (4) other digital analyses, such as classification.

Image Display

Images are displayed based on their digital numbers; generally, the higher the brightness value, the brighter the display color. In gray scale those colors range from black through shades of gray, to white. A computer monitor has three color guns (red, green, and blue), each of which traditionally displays up to 256 brightness levels. It is possible to display up to three bands at any one time by assigning each band to a different color gun (**Figure 8**). By selecting different band combinations, it is possible to identify characteristics of objects being sensed through their different spectral pattern responses. Image enhancement functions in image processing software can improve visual interpretation by altering the digital display values to make subtle differences in brightness values more obvious.

Visual Image Interpretation

Visual image interpretation is extremely important in remote sensing and, in some cases, may represent the only use of the data, or it may be incorporated into the beginning or intermediate stages of a more complex set of analysis operations. An image, whether aerial photograph or other digital sensor image, through patterns of brightness values across the scene, has tone (lightness/darkness), texture, shadows, patterns, and contextual attributes. By taking into account all of these attributes, often subconsciously, an interpreter can identify, classify, and quantify objects or regions, assess the condition of those objects or regions, and make measurements.

Classification

Remotely sensed data are often used in classification analyses, whereby individual pixel values are classified into meaningful categories. Classification may be based on spectral, spatial (texture, proximity, etc.), or temporal (changes through time) information in an image or images. Probably the two most commonly used automated classification methods are supervised and unsupervised classifications. These two methods are inherently different. On the one hand, supervised classification leverages the operator's *a priori* knowledge of the study area to drive the classification process – this assumes that the knowledge of the operator is complete and that the pixel values can be assigned to those classes that the operator is defining. On the other hand, unsupervised classification uses the statistical distribution of pixel values to assign pixels to statistical classes, which are subsequently interpreted by the operator into meaningful classes. This ensures that the full range of variability of pixel spectral values is considered, but can result in meaningless classes. A third classification method, known as hybrid classification, uses a mix of both methods. The end result of a classification is a new image with classified values, where each value represents a unique land cover category. Accuracy is assessed through comparing the resulting classification with reference data; a classification error matrix (**Figure 9**) is commonly reported, along with Kappa statistics which assess the result against the possibility of it being generated randomly.

In the social sciences, classification is important in characterizing and mapping environments such as urban, suburban, and agricultural land uses. Other forms of analyses which social scientists might apply to their studies include various automated change detection methods, which compare raw, transformed, or classified imagery from one time to another. Change detection

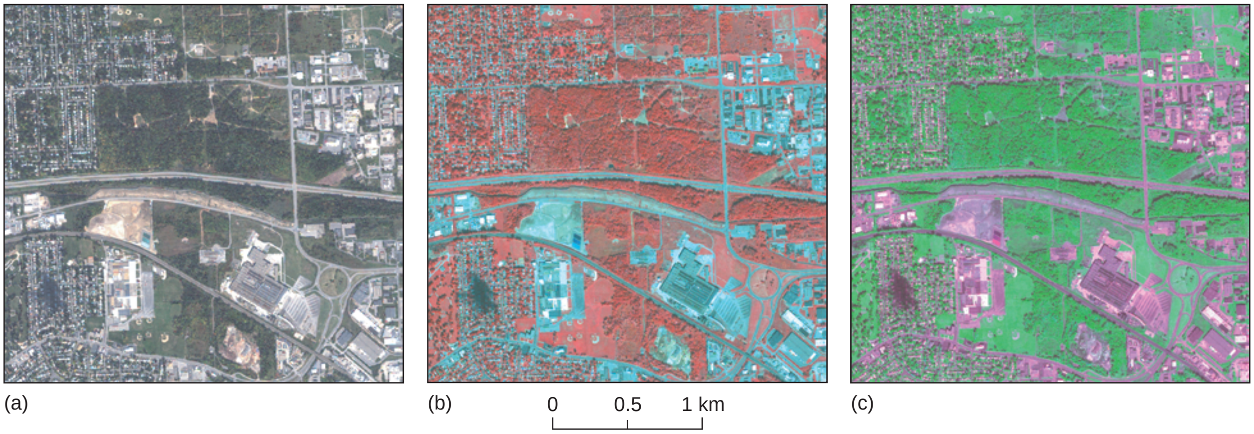

Figure 8 IKONOS data of the City of Syracuse, New York, USA (2001): (a) bands 3, 2, and 1 displayed on the red (R), green (G), and blue (B) color guns, respectively; (b) bands 4, 3, and 2 displayed on RGB, respectively; (c) bands 1, 4, and 2 displayed as RGB, respectively, where band 1 blue, band 2 green, band 3 red, band 4 near-IR.

Classification accuracy assessment report

	Reference data				
Classified data	Vegetation	Impervious	Water	Row totals	User's accuracy
Vegetation	112	24	0	136	112/136 = 82.35%
Impervious	24	100	40	164	100/164 = 60.97%
Water	0	0	40	40	40/40 = 100%
Column total	136	124	80	340	
Producer's accuracy	112/136 = 82.35%	100/124 = 80.65%	40/80 = 50.00%		Overall accuracy 252/340 = 74.12%

Kappa = 0.7359

Figure 9 Typical accuracy assessment error matrix.

applications can be on a long timescale to assess historical to present day changes, or rapid or near time change, such as emergency response and disaster monitoring and management. Other useful analysis functions include calculation of indices based on pixel values from one or more bands. A commonly used index is the normalized difference vegetation index (NDVI), which uses the red and NIR band values to allow identification of healthy vegetation. NDVI can be useful in helping to characterize changing crop health, for example, which could in turn assist in predicting famine and human migration patterns in a particular place.

Remote-Sensing Data Types

Aerial Photographs

Aerial photography, generally flown from an airplane, is still widely used in the creation of topographic maps worldwide, and represents a relatively cheap and accessible data source. Photography can provide black and white, color, or color IR data in either film or digital form. All photographs will have some geometric distortion inherent. An orthophoto is a photograph that has been corrected for any geometric distortions, thereby allowing for distortion free measurements. Present day aerial photography can be commissioned from one of the many commercial firms specializing in such data. The USGS makes available digital orthophoto quadrangles that correspond to USGS 7.5 min topographic sheets. The USGS is also a good source for US archival photography (and other remotely sensed data sets). Because aerial photography dates back before satellite imagery, it represents a valuable source of historic landscape data. Land use studies often rely on panchromatic aerial photographs for compiling land use histories. Another advantage to aerial photography is that oftentimes it is captured at relatively low altitudes, thereby capturing fine detail, such as buildings, stands of trees, roads, bodies of water, etc., which renders it highly useful for visual interpretation.

Multispectral Earth Observation Sensors

Most multispectral Earth observation sensors are carried in satellites that orbit at approximately 400–900 km above the Earth (refer to **Table 1** for details of select systems). Geostationary satellites, which carry communications and meteorological sensors, remain in a fixed position above the Earth at much higher altitudes, whereas Sun synchronous satellites are placed in orbits that follow the Sun. Satellite sensors confer the advantages of sensing large areas in a single pass, repetitive and predictable passes, and coverage of all or most of the Earth. Here we describe some of the more commonly used land observation sensors.

The US government's Landsat program (**Table 1**) far surpassed initial expectations. When the first Landsat satellite was launched in 1972, it was not known whether the system would work. Landsats 1–3 carried the return beam vidicon (RBV) and multispectral scanner (MSS); the MSS, which had been put on board as an afterthought, was the more successful sensor of the two. Landsats 4 and 5 carried MSS and the TM sensor; Landsat 6 carried the Enhanced Thematic Mapper (ETM), but failed on launch, and the latest, Landsat 7 carries the ETM+. At the beginning of 2007, data from Landsat 5 were only being collected over North America, and Landsat 7 was functioning with a scan line corrector error that results in missing data. The USGS has recognized the importance of continuing the Landsat program, which has provided a continuous stream of land data for over 34 years, with 2 million scenes archived as of December 2006. With Landsats 5 and 7 due to complete their useful lives by 2010 or sooner, it is crucial that the plans for a Landsat 8 mission are given high priority.

In 1978 the National Oceanographic and Atmospheric Administration's (NOAA) AVHRR sensor was launched on a polar orbiting satellite (**Table 1**). The sensor was designed to collect meteorological data; however, it has been highly successful and the data are also used for land sensing. Indeed, the highly used NDVI was originally developed for AVHRR to provide global coverage of Earth's vegetation coverage.

Other medium resolution multispectral systems include the French earth observation system, SPOT, which began operations in 1985 and is now receiving data from the SPOT 5 satellite (**Table 1**). This latest system is capable of providing earth data at 5, 10, or 20 m pixel widths, depending on the wavelength. The India Remote Sensing (IRS) instrument is another moderate resolution sensor that has had a high degree of success (**Table 1**).

In 1999 the first commercial high resolution sensor, IKONOS (with four 4 m multispectral bands and a 1 m panchromatic band), was launched, and 2 years later QuickBird followed suit. Several other high resolution sensors have since been operationalized, and others are in development. Applications of medium and high resolution sensors are similar, including forestry, agriculture, resource use and management, oceanography, hydrology, rural and urban planning, to mention but a few. However, whereas the medium resolution sensors are useful at a scale of 1–2 ha or more, the high resolution sensors allow much more specific studies at fine detail. For example, while medium resolution sensors can be useful in detecting urban–rural boundaries, high resolution sensors can detect specific urban changes at the scale of individual buildings and other fine scale phenomena.

Other Classes of Sensor

Other sensor classes include hyperspectral sensors, which collect almost a continuum of very narrow bands of the electromagnetic spectrum to provide very detailed spectral information about features. A typical hyperspectral sensor may have 200+ bands (**Table 1**). Hyperspectral data are useful in studies of atmospheric composition, and detailed agricultural applications, such as determining leaf composition, soil moisture content, and salinity.

RADAR data are acquired by the sensor sending out known wavelengths of energy and recording the time it takes for the energy to return. RADAR has the advantage that it can penetrate ('see') through clouds. RADAR data are good for mapping topography, drainage, and other clearly defined landforms. In the social sciences, the most common use of this type of data is for elevation mapping as well as analyses of land cover change. LIDAR is another active sensor that works similarly to RADAR, except that it uses beams of light. Very detailed profiles of three dimensional features can be built up using LIDAR data, and has been used for urban mapping, highway mapping applications, assessment of forests and protected areas, monitoring of contaminants, such as air pollution, and also changes in land cover.

Finally, thermal sensors, sensitive in the mid and far IR and microwave regions of the spectrum, sense surface temperatures and thermal properties of features. It can be useful for identifying and mapping objects, and generally is better as a qualitative interpretation tool, rather than for precise temperature measurements. In the social sciences, thermal remote sensing data have been used to study urban heat island effects, as well as monitoring of forest fires.

Summary

In the last 5–10 years, remote sensing has developed rapidly into a technology that is now available to a wide audience, and is providing valuable spatial environmental information to support a wide range of social science applications. Limitations in using remotely sensed data in the social sciences derive from the nature of remotely sensed data itself, as well as logistical issues of data availability, accessibility, cost, and ease of use. While remotely sensed data cannot provide direct information about social conditions of individuals, households, com munities, or global systems, they can provide information about the nature and spatial configuration of the physical environment in which those entities operate. Remote sensing can also be useful in providing indicators of, and predictions for, social processes that manifest themselves in the physical environment, such as urban revitalization projects, or processes that develop as a result of changes in the physical environment, such as famine and out migration as a result of drought. As more remote sensing systems are developed, and data availability and access to user friendly software continue to improve, applications of remotely sensed data in the social sciences are likely to continue to expand.

See also: Georeferencing, Geocoding; Global Positioning/GPS; Scale; Photogrammetry/Aerial Photography.

Further Reading

Campbell, J. (1996). *Introduction to Remote Sensing* (2nd edn.). New York: Guilford Press.

Congalton, R. G. and Green, K. (1999). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Boca Raton, FL: CRC Press.

ERDAS (1999). *Erdas Field Guide* (5th edn.). Atlanta, GA: ERDAS.

Fox, J., Rindfuss, R., Walsh, S. and Mishra, V. (2003). *People and the Environment; Approaches for Linking Household and Community Surveys to Remote Sensing and GIS*. Norwell, MA: Kluwer Academic Publisher.

Lillesand, T., Kiefer, R. and Chipman, J. (2004). *Remote Sensing and Image Interpretation* (5th edn.). New York: Wiley.

Lindborg, C. (1997). IRS (Indian Remote Sensing). http://www.fas.org/spp/guide/india/earth/irs.htm (accessed Mar. 2008).

Liverman, D., Moran, E. F., Rindfuss, R. R. and Stern, P. C. (eds.) (1998). *People and Pixels: Linking Remote Sensing and Social Science*. Washington, DC: National Academy Press.

McCoy, R. M. (2004). *Field Methods in Remote Sensing*. New York: Guilford Press.

USGS (2005). Thermal Infrared Multispectral Scanner (TIMS) Sensor/Instrument. http://eosims.cr.usgs.gov:5725/sensor documents/tims sensor.html#1 (accessed Mar. 2008).

Williams, D. L., Goward, S. and Arvidson, T. (2006). Landsat: Yesterday, today, and tomorrow. *Photogrammetric Engineering and Remote Sensing* 72, 1171 1178.

Relevant Websites

http://www.csc.noaa.gov
About LIDAR Data, NOAA Coastal Services Center.

http://noaasis.noaa.gov
Advanced Very High Resolution Radiometer AVHRR, NOAA Satellite and Information Service.

http://www.asprs.org
American Society for Photogrammetry and Remote Sensing (ASPRS).

http://oceancolor.gsfc.nasa.gov
An Overview of SeaWiFS and the SeaStar Spacecraft, SeaWiFS Project, Ocean Color.

http://eo1.gsfc.nasa.gov
Earth Observing Mission 1, Goddard Space Flight Center.

http://www.esa.int
European Space Agency (ESA), information on European remote sensing programs and activities.

http://www.geoimage.com.au
GeoImage.

http://glcfapp.umiacs.umd.edu
Global Land Cover Facility Earth Science Data Interface, access to remotely sensed and other data.

http://www.isprs.org
International Society for Photogrammetry and Remote Sensing (ISPRS).

http://ladsweb.nascom.nasa.gov
MODIS Level 1 and Atmosphere Data Products, LAADS Web.

http://gcmd.gsfc.nasa.gov
National Aeronautics and Space Administration (NASA) Global Change Master Directory, information on satellite and earth science data.

http://www.spotimage.fr
SPOT Image.

http://www.geovar.com
Spot Satellite Imagery, Geospatial Data and Information System.

http://www.digitalglobe.com
Standard Satellite Imagery, Digital Globe.

http://edc.usgs.gov
United States Geological Survey (USGS) Earth Resources Observation and Science (EROS) data center, information and access to aerial and satellite data of the United States (Go here for links to USGS Global Visualization Viewer (GloVis) and Earth Explorer to search EROS data center archives).

http://landsat.u0sgs.gov
USGS Landsat, comprehensive information on the US Landsat program.

Representation and Re-presentation

A. Kobayashi, Queen's University, Kingston, ON, Canada

Glossary

Discourse The power-laden communicative relationship between human beings. All human relations are discursive.

Nonrepresentational Theory Attempts to go beyond representation by focussing on the material acts, or performances, of social life than on interpretations of the visible.

Power/Knowledge The term used by Michel Foucault to describe the quality of discourse.

Representation The practice of constructing meaning through language, a discursive or signifying act.

Introduction

Representation is the practice of constructing meaning through language. Although historically philosophers, particularly those of the eighteenth century, make rep resentation synonymous with human thought as the capacity to apprehend the external world, recent post structuralist approaches view representation as the social practice of signifying acts through which products such as texts, pictures and other images, and landscapes are formed and given meaning. During the 1990s, geog raphers and other social scientists referred to the 'inter pretive turn' as the basis for understanding acts of representation that give and are given meaning. Cultural geography in particular has been concerned with study ing the representations that comprise the human land scape. Recent critical geographers are also concerned with the representation, and re presentations that com prise the discipline of geography.

The issue of representation is perhaps the most dif ficult and persistent concept that has been wrestled with by Western philosophers. In the eighteenth century classical philosophy from Descartes onward, represen tation was synonymous with thought, or the capacity for reason, as the basic attribute that permits knowledge of the world and distinguishes human beings from other beings. Most notable is Kant's work, which posits a transcendental realm in which human ability to com prehend the world is structured; inherent characteristics of human thought and the ability to reason allow rep resentation to take place as the connection of an idea and an object in the world. Understanding the transcendental realm was for Kant the basis for understanding the possibility and the validity of knowledge. He placed great emphasis on the importance of defining the conditions under which knowledge could thrive. Subsequent chal lenges to Kantian idealism have asserted either that representations of the world can arise through internal creativity, without the constraints of a transcendental epistemology, or that such individual agency is mean ingless in the face of external structures that condition thought and meaning.

The question of representation in the discipline of geography has been most deeply influenced by the writings of Michel Foucault who, among other post structuralist thinkers, saw problems with a definition of human consciousness as either externally determined, as occurs in both positivist and Marxist interpretations, or purely interior, as occurs in phenomenological forms of existentialism. Foucault's concept of the 'analytic of finitude' connects the condition of subjectivity and the historical experience of objectification. Knowledge is neither created *sui generis* by the mind nor imposed de terministically upon the mind, but rather produced through the bodily act of discourse; that is, in the pro duction of language. All human relations are discursive. Foucault's depiction of discourse is always laced with power, including the power of the body to represent itself in particular historical conditions as well as the power of historical institutions to regulate the body. He writes thus of power/knowledge as the nexus through which the world is organized. Human beings are subjected to power/knowledge through the organization of the world (which, for Foucault is only a material world, including its discourses), and people are produced through dis course. There is no absolute truth but only the creation of 'regimes of truth' that act as the normative conditions for a particular time and place. Representation is thus the result of constructing particular regimes of truth.

What LaClau and Mouffe call the 'revolution of our time' is the realization that the act of discourse is the human way of engaging, organizing, and classifying the world that surrounds us. There is no 'nature' therefore except a nature that is subject to human classification. Contemporary epistemology rests on understanding that no discursive representation, whether of nature or human artifact, can be read transparently or unproblematically. Nonessentialism thus stated might also be seen as the revolution of geography. No intellectual development has had a more profound effect upon geographical method ology during the past four decades as geographers have sought to question the assumptions that underlie the process of representation.

Foucault wrote of the various forms of discourse, and of the various landscapes, in which people are constituted through discourse, in a range of institutions, including prisons and psychiatric institutions. Foucault's work also relies heavily on the power of visualizing the world by organizing power/knowledge in particular ways. He seizes upon Bentham's panopticon, for example, as one of the most powerful means of enacting a discourse of control. Although geographers such as Chris Philo have read Foucault critically, and point out that he employs an abstract and geometrical notion of space and an insuffi ciently grounded notion of place, they have also been keen to take up the Foucauldian notion of power, which makes sense of the ways in which representations are produced within specific landscapes. Landscapes as rep resentations in this sense are the medium through which discourse is enacted. This recognition has met very fertile ground in the discipline of geography, which has always been a highly visual discipline, but which now questions the act of visualizing, the ways, for example, that the map is a discursive representation that organizes human beings in particular ways, and that our concepts of landscape likewise impose normative limits upon human action.

Linking representation to power has also led geog raphers to take up Gayatri Spivak's distinction between two types of representation: representations that speak of and representations that speak for. Speaking 'of' others raises the ethical question of the ways in which the lives of Others are constructed, particularly as they are usually constructed according to historically specific norms of race, class, and gender. Following Edward Said, 'Orien talism' is a way of depicting the colonial Other in ways that justify colonial power and subjugation. Such repre sentations of the body of the Other become the basis for the practice of racialization. Similarly, Gillian Rose points to the powerful ways in which gender is repre sented, or re presented, in a variety of visual and power charged discourses. The act of positioning people within a landscape powerfully represents historically gendered norms of man's place and women's place.

Although cultural geographers commonly recognize that speaking of others is itself a discursive act – a rep resentation of others' representations – most continue to believe that it is possible to apply a practiced geo graphical eye to a critical interpretation of landscapes as repositories of meaning. Such perspectives 'interpret' the design of the world in terms of buildings that represent capitalist power systems, parks that contain tributes to nature or monuments to national heroes, or landscapes of war that represent the inhumanity of violence. Land scapes also signify belonging and exclusion through icons of gender, class, and race that are deployed to regulate who is welcome. Signs may be expressed literally in words, figuratively through a wide range of graphic im ages, or more subtly through the arrangement of objects that signify taste, or cultural or religious values. Land scape signification can beckon, like the golden arches of Macdonald's now recognized by children throughout the developed world; or it can repel, like the sinister, barbed wire topped walls of the prison or military compound.

Geographers have also been very attentive to repre sentations that depict place through the arrangement of images in photographs, motion pictures, or the graphic arts. Photographic practices change over time to indicate the meaning of different places such as home or work, and the value that people attach to memories of places in which significant forms of human relationship are rep resented. Landscape paintings indicate the normative construction of nature or the importance of property as measure of social standing. There are many ways of representing places that are distant, foreign, and exotic, whether such places are located across the world from the originator of the image, or in the midst of a dominant society. Advertising, political campaigns, and religious messages are examples of signifying practices that rep resent bodies in normative ways that are meant to in fluence the larger population by invoking subtle cultural cues.

The ethical dilemma of speaking of others has also led to considerable discussion of the need for auto reflexivity in human geography research. Reflexivity uncovers the privilege of the researcher's positionality and her/his power to represent the lives of Others and thus to in fluence the ways in which they are constructed as well as the ways in which they construct their own identities. Critical reflexivity can avert practices such as Oriental ism that represent the distant Other in ways that justify practices of colonialism and imperialism. Some would argue, however, that no amount of reflexivity can change the fact that in speaking of Others, the researcher ap propriates their voice and limits their possibilities for self representation.

Speaking 'for' others potentially provides a basis for a politics of representation through which social groups can mobilize in order to create the 'discursive space' for social change. New social movements acknowledge the historical ways in which power/knowledge has been constituted and believe that new meanings, discourses, and forms of social relations can be constructed through new forms of representation. 'Re' presentation can work either to replicate and strengthen established discourses or to create new forms of resistance. Geographers have been concerned with the ways in which the politics of representation work – actually or potentially – to create landscapes free of oppression based on racism, sexism, class, religious persecution, ablism, or environmental degradation. A politics of representation therefore redefines or resituates the scholar as an activist who not only interprets (speaking of) but also becomes an agent of change (speaking for). But it also creates the conditions under which people might re present their own subject positions within landscapes that are reconfigured to reduce the element of control that is so dominant in Foucault's institutions. Recently, geographers have ex perimented with innovative forms of community repre sentation and empowerment, using methodologies such as Photo Voice or Geographic Information Science in action research projects in which community members represent their experiences to policymakers and polit icians. Photo Voice, for example, represents a collabor ation among community members to share images of their lives both as a means of self understanding and in order to portray to others something of their experiences. It can have therapeutic effects that come from sharing images with others and thus enhancing interpersonal understanding. Community GIS represents a powerful interpretive tool for re presenting the world of com munity groups in ways that may be easier for policy makers to understand, and also as a means of problem collaborative problem solving.

Notwithstanding the advances of post structuralist thought in clarifying that representation is an active so cial act grounded in specific historical and geographical conditions, the geography of representation remains highly fraught. Over the past decade, an increasing number of geographers have pointed out that the inter pretive, or cultural, turn has resulted in a deadening of cultural geography, reducing culture to a series of un problematic representations that are divorced from the bodily experiences of human beings. Cultural geography as 'mere' representation has a tendency to become a kind of homily to the cultural landscape, in which meaning, even meaning that is the result of discursive action, is divorced from human experience. The cultural turn has resulted in university and college courses that send stu dents scurrying to mine the landscape as a repository of meaning, bringing back representations that may be in spirational for the geographical imagination but that are problematic for their normative assumptions. The recent trend toward 'more than representational' theories in cultural geography highlights a need to treat landscapes, or any form of representation, as always in flux, both in their creation and in their interpretation. The inten tionality of representation is appropriated at its inception by the discursive process, to become contested, resisted, re presented. As Foucault suggested, subjects may pro duce particular representative objects within a landscape, and endow those objects with particular intended meanings, but they do so only within an overarching episteme that positions both the subject and the inter preter within a culturally created context. Human action is thus a constant state of active representation and, through discourse, appropriation of the representative act; so for geographers to think of representations as fixed or static is not only an ontological error, but one that has ethical and methodological connotations. *Post hoc* repre sentational attributions capture neither the active process of landscape representation nor the possibilities for social change that occur through active engagement in that process.

Other geographers have gone further, drawing upon those theorists, such as Gilles Deleuze and Felix Gua terri, who are influenced by Spinoza's notion of affect as the pre conscious possibility of emotion as a historically constituted human experience. According to this per spective, paying attention to the embodied possibilities of affect will allow human beings to break free of the conditioned, and normative, emotional responses that are inherent in culturally structured representations, thus creating previously un thought of possibilities. Geog rapher Nigel Thrift points to forms of representation such as dance that are difficult to freeze and therefore depict the creative potential to imagine bodies in new ways. This emphasis highlights the importance of the body itself as a landscape of social transformation. Critics of this perspective, however, point out that no act is pre reflectively disconnected from its historical context; at best then, and paradoxically, the nonrepresentational act is a particular form of oppositional representation. Moreover, to go beyond representation is not necessarily to shed the importance of understanding the historical and geographical power of representation to construct the world.

Debates over the place of representation in human geography return us to earlier and still unsettled ques tions about the ontology of representation. While most would agree that the act of representation is ongoing and in constant flux, there is wide disagreement over the status of the representational product. Some opt for placing a focus upon those moments of representation that are most powerfully fixed, in the built environment or in a text or a painting, in an attempt to achieve understanding of the ways in which human beings have sought to imbue their surroundings with meaning and in the process have created systematic forms of represen tation that comprise culture. There is a growing belief, however, that such an approach lapses easily into essen tialism and cultural normativity. The 'representation of representations' reproduces dominant meanings, with implications for ongoing processes of racialization and other forms of power laden signification.

At the heart of this debate is the question of how we interpret what Jean Paul Sartre called the 'accumulation outflow', or the products of human actions. How can we understand the cumulative effects of human action as the continual and seamless flow of representation if those effects apparently become frozen in place? From a political perspective, which is more powerful, the normative representation that seeks to control populations through signifying practices that maintain a *status quo* and a conservative interpretation of social change, or the nonrepresentational act that seeks to disrupt meaning and transcend historically constituted affects? Is it mere utopianism to search for landscapes in which the power to represent Others is sufficiently subdued to overcome oppression, or is it possible that geographers can imagine such a landscape?

See also: Cultural Materialism; Cultural Turn; Landscape; Non-Representational Theory/Non-Representational Geographies; Orientalism; Postcolonialism/Postcolonial Geographies.

Further Reading

Barnes, T. and Duncan, J. (1992). *Writing Worlds*. London: Routledge.
Castree, N. and Macmillan, T. (2004). Old news: Representation and academic novelty. *Environment and Planning A* 36, 469 480.
Cosgrove, D. and Daniels, S. (eds.) (1988). *The Iconography of Landscape: Essays on the Symbolic Representation, Design and Use of Past Environments*. Cambridge: Cambridge University Press.
Duncan, N. and Sharp, J. (1993). Confronting representation(s). *Environment and Planning D: Society and Space* 11, 473 486.
Foucault, M. (1973). *The Order of Things*. New York: Vintage.
Hall, S. (ed.) (1997). *Representation: Cultural Representations and Signifying Practices*. London: Sage.
Lorimer, H. (2005). Cultural geography: The busyness of being 'more than representational'. *Progress in Human Geography* 29(1), 83 94.
Philo, C. (1992). Foucault's geography. *Environment and Planning D: Society and Space* 10(2), 137 161.
Rose, G. (2003). Family photographs and domestic spacings: A case study. *Transactions of the Institute of British Geographers NS* 28, 5 18.
Said, E. (1977). *Orientalism*. New York: Phaiden.
Spivak, G. (1988). Can the subaltern speak? In Nelson, C. & Grossberg, L. (eds.) *Marxism and the Interpretation of Culture*, pp 217 313. London: Macmillan.
Thrift, N. and Dewsbury, J. D. (2000). Dead geographies, and how to make them live again. *Environment and Planning D: Society and Space* 18, 411 432.

Representation, Politics of

H. V. Scott, Aberystwyth University, Aberystwyth, UK

Glossary

New Cultural Geography A strand of cultural geography that emerged in the 1980s, drawing above all on British cultural studies and on post-structuralist theory.

Representation A complex term which embraces several distinct but interconnected meanings, representation has become one of the most debated concepts in human geography. It is possible, however, to identify two principal ways in which the term is used and understood. First, representation refers to a portrayal, or equally to the act of portrayal of certain ideas, objects, places, or people. It may therefore be understood as an act or material object that stands for something else. Representations may take the form of a material product, such as a painting or a written text, but are also produced by intangible means of communication such as the spoken word. Second, the term refers to the active inclusion and consideration of the interests, needs, knowledges, and identities that define particular individuals or groups. While this can mean inclusion in a formal political system (for instance, through the granting and exercising of electoral rights), it also refers more broadly to inclusion in diverse social, cultural, intellectual, and economic spheres of human life, as well as in 'real' geographical spaces.

The Politics of Representation A phrase that attained common usage in human geography by the 1990s, the politics of representation conveys the notion that all forms of representation are biased and selective, favoring the interests and ideas of certain groups while marginalizing or excluding others, and hence are open to contestation. It also refers to social struggles that are played out over and through representation, in both senses of the term.

Introduction

Until the latter half of the twentieth century, work in human geography was founded upon a deep seated faith in the researcher's ability to produce unbiased and transparent representations of the world by means of rigorous and objective investigation. The emergence of positivist approaches in the 1950s signalled, on the one hand, a radical break with existing practices of representation in human geography, for the empiricist traditions of detailed written description were rejected in favor of abstracted visions of the world based on mathematical models. On the other hand, however, these new theory driven impulses, which were inspired by the 'hard' sciences, perpetuated what had gone before in terms of their inbuilt assumptions about the possibility of producing and conveying knowledge in a manner untainted by factors such as personal experience, social background, or cultural identity.

In the 1970s, a humanistic backlash against the abstractions of quantitative geography fostered scholarly recognition of the role of subjectivity, and hence of the researcher's presence, in the production of geographical knowledge. Yet, despite the efforts of humanistic geographers to bring human experience and subjectivity to the center of geographical inquiry, the politics that are now deemed to be ever present in representations of the world (whether produced by academics or other groups) and in the circumstances of their production remained largely unchallenged. Only with the embrace of post structuralist thought in the late 1980s did the politics of representation begin to achieve real prominence as a focus of analysis and debate.

Geography and the Politics of Representation: An Overview

The late twentieth century diffusion of post structuralism resulted in the profound destabilizing of old certainties that had once underpinned academic thought and practice across the humanities and social sciences. In human geography, as in other academic arenas, the emergence of doubts about the existence of a world filled with naturally given and enduring meanings that could be unlocked, as well as of new understandings of knowledge as inescapably partial and power laden, produced what is referred to as a 'crisis of representation'. Despite the negative connotations of this expression, the emergence of geographical anxieties about representation provided a stimulus for intellectual creativity that encouraged the development of new theoretical, methodological, and thematic directions in social and cultural geography. These new directions were guided by the argument that representations are never mirror images of reality, but instead are always the product of diverse and ever shifting contexts, and hence are never innocent, unbiased, or divorced from the realm of power and politics. Such concerns have been shared by academics in a range of other disciplines that include art history, literary criticism, cultural studies, and anthropology, and human geographers have both drawn on and contributed to work in these other disciplines. Despite the absence of clearly defined disciplinary boundaries, geographical approaches are nevertheless distinguished by a particularly pronounced interest in exploring the politics of representation through the prism of geographical concepts such as space, place, and landscape.

Since the late 1980s, critical geographical engagement with the politics of representation has involved, first, the deconstruction or 'reading' of diverse representations both textual and material, ranging from written texts and visual images (such as travel accounts, maps, and paintings) to physical landscapes. Focused above all on representations that are produced by powerful groups or individuals, critical analyzes show how representations work to promote particular ways of knowing and imagining the world and particular constellations of power and privilege within it. Demonstrations of the ways in which representational practices can have tangible effects in the world, including the marginalization of certain knowledges and peoples or social groups, have therefore been central to geographical interest in the politics of representation.

Crucially, however, cultural and social geographers have been concerned to show that representations are always open to multiple readings and interpretations and, consequently, to challenge and contestation. While scholarly energies initially dwelled on uncovering the politics embedded within particular representations, the emphasis increasingly shifted toward exploring the discursive, spatial, and embodied ways in which dominant representational practices may be challenged or subverted by the marginalized or oppressed. In attending to the ways in which representations are contested, social and cultural geographers have placed at the center of their inquiries a broad spectrum of groups and identities which, both within academia and in the wider world, have experienced varied forms of exclusion or marginalization on the grounds of categories such as race, class, or sexuality.

Inspiration for such projects has undoubtedly been drawn directly from the work of key post structuralist theorists such as Foucault and Derrida. However, currents of thought in feminism, postcolonialism, and cultural studies (all of which engage closely with post structuralism) have played a prominent role in developing geographers' concerns, not only for analyzing representations and their effects in the world, but also for addressing the profoundly political issue of whose voices, views, and ways of knowing are represented in disciplinary geography, either as subjects of intellectual inquiry or as examples of valid geographical knowledge.

The analysis of representations has necessarily been accompanied by self reflexive engagements with issues surrounding the politics of representation in contemporary academic geography. No longer a taken for granted aspect of 'doing geography', the production of representations is now understood as an undertaking that brings with it serious moral responsibilities for the researcher, precisely because representations advance situated, partial understandings of the world that can promote certain worldviews and practices beyond the academy and therefore have real effects on people's lives. The difficult questions surrounding contemporary practices of representation in geography are illustrated by past debates about whether contemporary researchers who seek to undermine discriminatory representations (such as those produced by agents of colonialism) by making them the object of critical analysis run the risk of inadvertently reinforcing the very ideas and ideologies that these representations convey. Since the 1990s, the impact of such concerns in social and cultural geography has become increasingly apparent in the careful contextualization of contentious representations that are reproduced as objects of critical study in research articles and books, as well as in a heightened awareness of authorial use of language and style.

At the same time, however, anxieties have emerged over the legitimacy with which geographers may claim to represent the views and experiences of other groups or individuals or, to put it another way, to provide them with a 'voice', especially when researchers and research subjects are separated by profound cultural, social, or educational differences. The very complexity of recent debates over representation in geography is reflected, therefore, in ongoing tensions between the desire to achieve greater inclusiveness in terms of who is represented in contemporary geographical research, and the recognition that any academic claims to speak with authority on behalf of others are severely circumscribed.

Landscape

Since the emergence of the new cultural geography in the late 1980s, landscape has constituted one of the most prominent arenas of inquiry in which the politics of representation have been explored by both cultural and social geographers. Within this context, an ongoing concern among academics is to contest the apparent innocence and taken for granted existence of landscapes by demonstrating that the interests and influence of particular groups in society are reflected in representations that are made of certain landscapes as well as in the material form that those landscapes take. While some landscapes are self evidently a focus of struggle (for instance, those marked by overt military and political conflict, such as the highly contested barrier separating Israel and the West Bank), others, such as affluent suburban landscapes of contemporary North American cities or rural European scenes captured in eighteenth century paintings, partially conceal their politics beneath a mundane or estheticized exterior. Departing from the premise that all landscapes may be imbued with multiple meanings that reflect distinct and often competing interests, identities, and beliefs, a great deal of research has examined the conflicts and negotiations over land scape that are played out by means of representational practices, whether in the context of 'exceptional' land scapes or those that appear to be everyday and mundane.

Work on landscape in social and cultural geography is too disparate and diverse to allow one to trace a clear progression in terms of how it has engaged with issues of representation in recent decades. Nevertheless, since the late 1980s, the focus has largely shifted away from the critical analysis of representations of landscape such as paintings and maps and toward an exploration of the processes by which landscapes are continually (re)shaped and struggled over by both material and representational means. However, whether their work is predominantly concerned with the analysis of individual representations or with tracing how landscape unfolds as a process, cul tural geographers have displayed an ongoing interest in examining which groups possess the power to shape and inhabit particular landscapes in ways that reflect and reinforce their own interests and beliefs, and which are marginalized within or excluded from them.

Politics in the Cartographic Landscape

One form of representation that attracted considerable attention from cultural geographers in the 1990s was cartography. Although the critical study of landscape paintings by scholars such as Cosgrove and Daniels provides a more frequently cited example, the seminal work that exposed the biases and exclusions embedded within cartographic representations has been equally influential, not only within human geography but also beyond. Initiated above all by J. B. Harley, who drew particular inspiration from Foucault's writings on power and knowledge, these critical explorations of cartography have presented a dramatic challenge to the notion that 'scientific' maps – that is, those produced within a modern, Western mapmaking tradition that emphasizes mathematical accuracy – are innocent, objective, and universally valid. Some maps, of course, are overtly political, such as world maps that were made to show the territorial extent of the former British Empire. High lighting in red Britain's colonies and protectorates, they were intended to convey unequivocal messages to British and colonial subjects alike about the global reach and magnitude of the Empire. However, even if the overtly political content is stripped away from such cartographic representations, leaving only bare territorial outlines, they are still neither innocent nor objective.

As Harley's work demonstrates, the historical rise to dominance of a Europe centered world map, accepted by most inhabitants of the contemporary West as an objective reflection of geographical reality, is inextricably con nected to power and politics, for it involved the im position of a culturally specific way of imagining the world that accompanied and propelled European imperial expansion. Critical studies of cartography, however, are not only concerned with what such representations con vey, but equally with what they leave out. By revealing how (to cite one of many possible examples) colonial maps of the Americas mask Amerindian ways of per ceiving and representing the continent's landscapes and geographies, the work of Harley and other scholars demonstrates that such taken for granted representations are in fact historically implicated in the marginalization of non Western knowledges and worldviews.

The Material Landscape as a Contested Site of Representation

Just as certain knowledges are omitted from cartographic representations, so too particular groups in society are excluded from or rendered marginal (whether in a physical, cultural, or political sense) in 'real' or material landscapes. Material landscapes may be regarded as sites of representation that favor the interests of select groups in society by giving their ideas and values tangible form. Although a variety of research themes could be men tioned, the processes of gentrification that transformed many urban landscapes of the industrialized West throughout the 1990s, and the conflicts and struggles that accompanied these processes, generated numerous studies by social and cultural geographers. A prevailing objective of this work was to reveal the conflicts that are frequently concealed beneath the estheticized and glossy exterior of revitalized city centers and urban neighbor hoods in cities such as Glasgow, Dublin, and New York.

As geographers have shown, the transformation of these landscapes is inseparable from a politics of repre sentation that revolves around exclusions that are both physical and symbolic in nature. While the visible fea tures of the gentrified landscape (such as luxury housing and boutiques) function as representations which com municate and reinforce the values of the privileged groups who create and use them, other populations, such as homeless people or those on low incomes, may be physically removed or displaced. In examining the phe nomenon of gentrification, then, social and cultural geographers such as Lees have asked questions about who possesses the power to shape the urban landscape and hence be represented through it, and how these inter ventions affect and are experienced by less powerful social actors. No less importantly, they have shown how physical and symbolic exclusions are contested, whether in the form of resistance to physical displacement, through the modification or use of the gentrified landscape in officially sanctioned ways, or by means of organized political protest that involves the production of contestatory representations.

This subsection and the preceding one identify two commonly encountered approaches to the politics of representation in landscape research, one involving the analysis of landscape representations such as paintings and maps and the other focusing on the material landscape as a contested site of representation. By no means, however, are these approaches mutually exclusive. The struggle to be represented – that is, to possess a recognized place in a particular landscape and a stake in shaping it – is shown by social and cultural geographers to be intimately connected to the production of textual representations. Indeed, a significant concern of many recent studies has been, precisely, to explore the connections between representations and material landscapes – to show how each is influenced by the other in an ongoing process of interaction that changes (or perpetuates) how landscapes are portrayed in diverse media and whose interests and worldviews are represented in, or excluded from, the material landscape.

Feminism

When feminist interventions began to make a tentative mark on human geography in the 1970s, representation was yet to attain prominence as a source of widespread intellectual anxiety and as a focus of critical analysis. Without doubt, these early interventions played a crucial role in the development of broader geographical concern for issues of representation and influenced the theoretical and methodological directions that critical studies of representation have taken. At a time when calls for a feminist agenda in human geography were being made, women were conspicuous in geography largely by their absence, whether as subjects of geographical inquiry or as academics. Initially, therefore, feminist geographers were principally engaged in a practical politics that strove simultaneously for the representation of women in geographical research and within the academy.

In order to challenge women's invisibility as subjects of inquiry, early feminist research in geography strove to bring their lives and experiences to the fore. In doing so, they not only demonstrated that women's mobility, their everyday use of space and experiences of place can differ markedly from those of men, they also drew attention to the injustices and inequalities that these differences represented. So, for example, it was shown that female mobility in the industrialized West was severely restricted in comparison to that of men, due to factors that included their spatial isolation as suburban housewives, women's lower economic status, and their avoidance of certain public spaces due to fear of crime or harassment. Whereas many male geographers conceptualized the home as a place of refuge and security, the emerging feminist scholarship argued that such assumptions sidelined the experiences of many women for whom the home is instead experienced as a space of violence, fear, and confinement.

Feminism and Representation since the Late 1980s

Reflecting broader trends in human geography, feminist geographers increasingly turned their attention to the analysis of visual and textual representations from the late 1980s onward, drawing on psychoanalysis as well as on post structuralist theory. In particular, their work has paid critical attention to the politics of representation that surround depictions of the female body in Western societies, whether in art, journalism, or advertising, and has highlighted the ubiquity with which women's bodies are represented as sexualized objects of heterosexual male desire and as sites of moral danger and irrationality that are in need of constant regulation and control. Feminist scholars have examined, for instance, the intimate connections between representations of female bodies and the imposition of discriminatory social expectations regarding women's bodily practices and their behavior in a variety of public and private spaces such as the workplace, the street, and the home. Research has also focused, however, on the body as a contestatory site of representation, and on how those female identities which present a challenge to gender based social expectations and norms are expressed by means of women's embodied performances.

Feminist engagements with the politics of representation have also involved a critique of the representational analyzes and practices employed by (predominantly) male academics. Cultural geographers' explorations of the European tradition of landscape painting and gardening became a particular target of criticism in the 1990s. Associated above all with the work of Gillian Rose, feminist discussions of landscape research argued that cultural geographers had not only ignored a crucial political dimension by sidelining the gendered nature of landscape portrayals, but inadvertently reproduced unequal gender relations by taking pleasure in the visual mastery of landscape and in the notion of landscape as 'a way of seeing'. Not only have women been historically subject to an exploitative male gaze that identifies them with and as part of a passive natural landscape, Rose argues in a frequently cited essay, that male gaze is in fact perpetuated in the landscape research of cultural geography.

Early feminist work in Anglo American geography tended to overlook the fact that the experiences of women in the industrialized West do not necessarily represent those of women in other parts of the world. However, the reception of feminist critiques from beyond the West, along with the embrace of postcolonial theory, brought about the development of increasingly cautious and self reflexive representational practices that recognize the cultural, educational, and socioeconomic differences that divide women as a group, and acknowledge the limitations inherent in any attempt by Western feminists to speak on behalf of other women. To a great extent, this 'crisis of representation' encouraged a pronounced focus in Anglo American research on Western women's experiences – a trend that is in fact widely reflected in post 1980s social and cultural geography. However, these anxieties have also ensured that feminist studies of representations produced by women have largely been critical and nuanced, rather than merely affirmative.

To give a brief example, feminists working within cultural geography have insistently challenged masculinist accounts of geography's history by highlighting the largely unacknowledged contributions that women have made to the construction of Western geographical knowledges since the nineteenth century, whether as travel writers, missionaries, or as the companions and assistants of male geographers. However, recent work on Victorian women's representations of colonial worlds beyond Europe not only emphasizes their distinctive geographical insights and spatial experiences, but also demonstrates how their diaries and descriptions reveal complicities with the Empire and the articulation of perceived race and class based differences that divided women as a group. Such studies provide a powerful reminder that representations are never innocent, even when produced by relatively disenfranchised individuals (in this case, nineteenth century European women), and demonstrate that they are always complex products of multiple, intersecting identities.

Since its emergence in social and cultural geography, feminist research has actively engaged with a politics of representation that endeavors to provide a place for women as academic contributors to geography or as subjects of geographical research, as well as for members of other groups who experience exclusion on the grounds of categories such as sexuality, race, or disability, whether female or male. Just as significantly, it has played a leading role in imbuing human geographers with a sensitivity toward the inescapably partial and situated nature of their own practices of representation and with a consequent awareness of the need for self reflexivity.

A New Crisis for Representation?

Recent years have brought growing discontent among social and cultural geographers with regard to the study of representation and increasing interest in the 'non representational' or the 'more than representational'. As Castree and Macmillan explain in a detailed analysis of these trends, many human geographers feel that there is no longer any need to provide further examples of how representations are power laden and partial, or of how certain spaces, places, and landscapes favor or exclude a bewildering spectrum of cultural groups and identities. Indeed, as the example of landscape studies illustrates, recent work in social and cultural geography has progressively moved away from the critical scrutiny of representations toward a concern for exploring embodied, precognitive practices and experiences. Although scholarly enthusiasm for the analysis of representations is clearly on the wane, a broader concern for representation in human geography is unlikely to disappear. Recent lively debates about the extent to which the voices and knowledges of non Anglophone scholars are marginalized in Anglophone human geography provides just one of many clear indications that the politics of representation will continue to be a prominent focus of discussion and inquiry.

See also: Cultural Geography; Feminism/Feminist Geography; Gentrification; Humanism/Humanistic Geography; Landscape; Non-Representational Theory/ Non-Representational Geographies; Positivism/Positivist Geography; Poststructuralism/Poststructuralist Geographies.

Further Reading

Cosgrove, D. and Daniels, S. (eds.) (1988). *The Iconography of Landscape*. Cambridge: Cambridge University Press.
Barnes, T. and Duncan, J. S. (eds.) (1992). *Writing Worlds: Discourse, Text and Metaphor in the Representation of Landscape*. London: Routledge.
Barnett, C. (1997). 'Sing along with the common people': Politics, postcolonialism, and other figures. *Environment and Planning D: Society and Space* 15, 137 154.
Castree, N. and Macmillan, T. (2004). Old news: Representation and academic novelty. *Environment and Planning A* 36, 469 480.
Desbiens, C. and Ruddick, S. (2006). Speaking of geography: Language, power, and the spaces of Anglo Saxon 'hegemony'. *Environment and Planning D: Society and Space* 24, 1 8.
Duncan, J. S. and Duncan, N. G. (2004). *Landscapes of Privilege: The Politics of the Aesthetic in an American Suburb*. New York: Routledge.
Harley, J. B. (2001). *The New Nature of Maps: Essays in the History of Cartography*. Baltimore, MD: John Hopkins University Press.
Lees, L., Slater, T. and Wyly, E. (2007). *Gentrification*. New York: Routledge.
Livingstone, D. (1998). Reproduction, representation and authenticity: A rereading. *Transactions of the Institute of British Geographers* 23, 13 19.

McDowell, L. and Court, G. (1994). Performing work: Bodily representations in merchant banks. *Environment and Planning D: Society and Space* 12, 727 750.
McDowell, L. and Sharp, J. P. (eds.) (1997). *Space, Gender, Knowledge: Feminist Readings*. London: Arnold.
Nash, C. (1996). Reclaiming vision: Looking at landscape and the body. *Gender, Place and Culture* 3, 149 169.
Nash, C. (2000). Performativity in practice: Some recent work in cultural geography. *Progress in Human Geography* 24, 653 664.
Rose, G. (1993). *Feminism and Geography: The Limits of Geographical Knowledge*. Cambridge: Polity Press.
Soderstrom, O. (2005). Representation. In Atkinson, D., Jackson, P., Sibley, D. & Washbourne, N. (eds.) *Cultural Geography: A Critical Dictionary of Key Concepts* (1st edn.), pp 11 15. London: I. B. Taurus.

Relevant Websites

http://www.sagepub.co.uk
Cultural Geographies (journal), Sage.
http://www.tandf.co.uk
Gender, Place and Culture: A Journal of Feminist Geography; Social and Cultural Geography (journal), Taylor and Francis.

Representation-Mapping

J. Mennis, Temple University, Philadelphia, PA, USA

Glossary

Cartographic Symbolization The use of graphic elements to stand for geographic phenomena on a map or other visualization.
Geographic Visualization The display of spatial data, typically using advanced computer technologies for the purpose of data exploration and knowledge discovery.
Semiotics A branch of philosophy dealing with how concepts are tangibly expressed.
Spatial Cognition How the human mind acquires, encodes, retrieves, and applies spatial information.
Spatial Data Model A computational strategy for encoding spatial data.

Representation – Mapping Defined

Representation with respect to mapping concerns how geographic phenomena are represented in spatial data bases and geographic visualization. In its most basic form, representation in the context of maps can be understood as the choice of cartographic symbols that stand for actual geographic phenomena. For example, a line on a road map may represent a highway, or an area shaded blue on a topographic map may represent a lake. This same principle may be extended to more sophisticated computational geographic visualization environments, such as three dimensional displays, where, for example, a display of a volumetric solid may be used to represent a geological structure.

Issues of representation in mapping, however, encompass far more than the simple choice of cartographic symbolization. Representation implies a model of reality – just as a model airplane represents a real airplane, a map represents the real landscape. As with all models, a map necessarily emphasizes certain aspects of reality while ignoring others. Representation with regard to mapping concerns the cognitive and computational processes of how certain aspects of geographic reality are transformed into a representation, as well as the structures that are used to encode that representation. The representations themselves may be visual or computational in nature, and the encoding structures used to support the representation may be graphic elements or computer database elements.

The representation of geographic phenomena can be considered a process whereby geographic information is initially conceptualized by a person, encoded in a storage medium (e.g., a computer), graphically displayed via a visual medium (e.g., on a paper map or computer screen), and consequently interpreted by an analyst or map reader. Because of the widespread use of computers for storing and managing the geographic data from which most maps are made and computational geographic visualizations generated, the nature of computational encoding for spatial data plays an important role in the process of cartographic representation. Likewise, the cognitive processes through which people learn about geographic information, and store it in memory, are also key to the computational representation of geographic phenomena.

The Representation Paradigm in Cartography and Geographic Information Science

In the period following World War II, the development and standardization of academic cartography programs in the US centered on establishing cartography as a scientific discipline. While a handful of academic cartographers emphasized mathematical and analytical applications, the 1970s saw the emergence of what has been called the communication paradigm in cartography, an adaptation of principles of communication theory to cartography. The communication paradigm took an engineering perspective on developing optimal design strategies for communicating geographic information through maps. Inherent in the communication paradigm was the idea that the purpose of maps is to communicate an objective and observable geographic reality to the map viewer, and that cartographic research should focus on developing design principles that maximize the efficiency of, and minimize potential error within, that communicative process. Thus, cartographic representation in this context concerned developing the most accurate cartographic symbolization strategies for effective communication of geographic information.

The communication paradigm eventually came under criticism for its rigid adherence to cartographic representation as seeking an ideal depiction of objective reality, with no explicit acknowledgement of the political and cultural contexts within which the representation was created or applied. In the 1990s geographic visualization emerged as an alternative to the communication paradigm. The emphasis here is on representation and analysis, with the explicit acknowledgment that a map does not simply communicate objective information but is a subjective expression of geographic knowledge. In addition, the geographic information ascribed to a map is, in part, produced by the map reader using *a priori* knowledge. It was argued that the process of map use can be understood using principles of cognitive psychology and semiotics, as well as the biophysical processes of vision, for explaining how geographic information is interpreted, stored, and applied to problem solving by a map reader. Instead of developing optimal map design strategies for efficient communication, research in geographic visualization emphasizes the development of visual techniques to facilitate geographic data exploration and hypothesis generation from maps and other geographic data displays, often using advanced computer technologies.

The cognitive psychology and philosophical frameworks that underlie much research in geographic visualization are related to similar developments in geographic information science (GISci) research, particularly in the context of spatial data models. Spatial data models are encoding strategies for storing spatial data in a computer. Because most maps and other graphic displays of geographic information are currently generated from spatial data stored in a computer, the nature of the spatial data model is an integral part of representation in mapping. Early research in spatial data modeling was largely driven by practical issues, such as the effective transfer of geographic information from paper map to computer database, efficiency of data retrieval, speed and convenience of plotting, and the maintenance of administrative records.

In the 1980s, researchers began to investigate the relationship between computational and cognitive strategies for storing and processing geographic information. This research agenda was catalyzed in part by the recognition that the conventional spatial data models, based primarily on simple geometric encoding strategies used to mimic the look of paper maps, were limited in their ability to capture geographic semantics, that is, to represent not only georeferenced measurements but also the complex entities and relationships that compose geographic domains. Thus, as in geographic visualization, researchers began to address the issue of representation of geographic phenomena not only in terms of efficient and effective storage and display strategies, but also in terms of the cognitive and computational processes of geographic information processing. This new research emphasis was applied to developing richer and more sophisticated computational representations of geographic domains.

Geographic Representation and Cognition

Categorization

Knowledge is cognitively organized in a categorical hierarchy, which supports maximum information retention with minimum effort. Categories are arranged vertically in the hierarchy by kinds (taxonomies). A category is a conceptual grouping of entities that are assumed to share a common set of properties. For example, consider the categorical hierarchy: ‘furniture’, ‘chair’, and ‘bar stool’, where ‘chair’ is a subcategory of ‘furniture’ and ‘bar stool’ is a subcategory of ‘chair’. Categories are often defined by their parts and by their use, or how a person interacts with objects within the category. For example, a chair typically has legs, a seat, and a back (though not all chairs have all of these specific parts) and is used for sitting. A higher level in the hierarchy implies a higher degree of abstraction and a greater inclusiveness of the category. Categories at the same level of the hierarchy have the same level of inclusion. For example, we may consider the category ‘table’ to be at the same level of ‘chair’ in the above hierarchy. The basic level category is that category in a hierarchy that is most useful and is learned first.

Cognitive categories do not behave as discrete mathematical sets; their boundaries are not rigid, they may be modified over time, and their member objects may have varying degrees of centrality of membership. Cognitive categories can be defined in terms of prototypes, those attributes most characteristic of the category, so that certain examples may be considered more central to a category than other examples. For example, a bar stool may be less central to the category ‘chair’ than the typical dining room chair. Prototypes may be used to identify new instances of that category in the environment.

These cognitive principles of categorization are also used to organize geographic information about places and events. For example, consider the categorical hierarchy: ‘settlement’, ‘city’, and ‘industrial city’, where ‘industrial city’ is a kind of ‘city’. In addition, characteristics of specific places may be organized within a categorical hierarchy based on nested spatial relations, for example, in the spatial hierarchy implied by a US state, a city within that state, and a neighborhood within that city. Here, one may cognitively associate the characteristics of, say, a particular city based on the state within which that city is contained.

Cognitive structuring of categories is informed by interaction with objects in that category, most importantly motor interactions at the scale of the human body. For example, as noted above, the key property of the category ‘chair’ is that one sits on it. These interactions that occur at the scale of the human body serve as cognitive metaphors with which geographic information is organized, even though human interaction with geographic entities at a scale larger than, say, an individual house, is typically not through direct physical interaction but is more abstract. For example, one cannot possibly experience an entire city or country, yet geographic information is cognitively encoded in a manner similar to that of human body scale objects.

Image schemata are the cognitive means by which metaphors generated from human scale interaction with the physical environment are used to structure knowledge in the mind. Since they are generated from motor interaction, image schemata are generally spatial in nature, for example, as with the image schemata 'container', 'center periphery', and 'path'. These image schemata are used to interpret more abstract information.

Perception and Object Recognition

Geographic knowledge may be acquired through direct sensory experience, most prominently vision. When one views a scene, the initial sensory input may be interpreted as a raw image of colors. Individual objects in that scene may be discerned through the recognition of patterns in the image that indicate basic forms and boundaries between those forms. The interpretation of the image as to its contents is informed by previous knowledge about the kinds of objects that may be found in the image. This *a priori* knowledge used in the interpretation of the raw image can take the form of image schemata, not only through the application of basic metaphors like 'container', but also in a manner that provides more specific propositional information. These propositional schemata may be considered precognized patterns associated with a particular category and used to recognize instances of that category.

A visual scene may be interpreted by matching patterns observed in the scene with schemata stored in the mind. A matched pattern may be identified as an object or event associated with a particular category, and is given meaning by the semantic information provided within that categorical hierarchy. Geographic learning is thus a process whereby sensory information is interpreted by *a priori* knowledge stored in the mind, and at the same time information gathered from sensory input may be used to modify that knowledge. Schemata serve as mediators between sensory input and stored knowledge.

Representation in Spatial Databases

Spatial Data Models

The representation of geographic phenomena in a computer is a process that involves the conceptualization of the phenomena to be represented, the use of a general geometric spatial data model for capturing that conceptualization, and the implementation of the data model in a specific data structure that can be physically encoded in computer hardware. Geographic phenomena are typically conceptualized as either a set of individual objects occurring within an empty space or as a quality that varies continuously throughout space. These conceptualizations are referred to as the object based and field based views, respectively. An example of the former would be the conceptualization of a number of buildings as a set of objects, each with a discrete boundary and with empty space occurring between them. The field based view is exemplified by elevation (above sea level), where elevation is considered to occur everywhere throughout space and can be measured at any location.

Different geometric spatial data models can be used to encode geographic phenomena that are associated with the object and field based views. Two general strategies can be distinguished: vector and raster. Development of the vector model was based on a cartographic encoding, in which the motivation was to mimic the traditional, pre computer means of handling spatial data – the paper map. Consequently, vector is composed of the point, line, and polygon geometric primitives. The development of the raster data model, on the other hand, was primarily hardware driven and was derived from the use of the grid for computer graphics and the use of the line printer for cartographic output. Thus, raster is based on an exhaustive tessellation of space into, typically square, grid cells.

For phenomena conceptualized using the object view, the vector approach is typically used, where each object is encoded as an individual point, line, or polygon. Alternatively, the raster approach may be used where each grid cell encodes the occurrence or nonoccurrence of the object at that grid cell location (**Figure 1**). For phenomena conceptualized using the field based view, vector approaches may be taken where space is exhaustively tessellated into regions. The modeled quality is encoded as varying continuously throughout each region according to a function that may be constant, linear, or of a higher order. Nominal land cover (e.g., residential and commercial), for example, is a field like phenomenon that can be represented by exhaustively partitioning the land surface into regions of uniform land cover type, where the

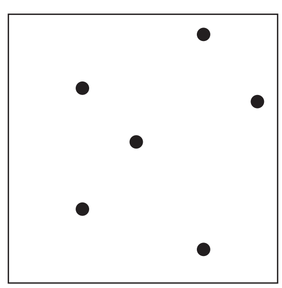

0	0	0	1	0
0	1	0	0	1
0	0	1	0	0
0	1	0	0	0
0	0	0	1	0

Figure 1 Vector and raster representations for geographic phenomena conceptualized using the object view, for example, a set of stores. The vector approach is shown on the left, where each store is represented as an individual point. The raster approach is shown on the right, where each grid cell encodes whether that cell contains, or does not contain, a store using the values of '1' and '0', respectively.

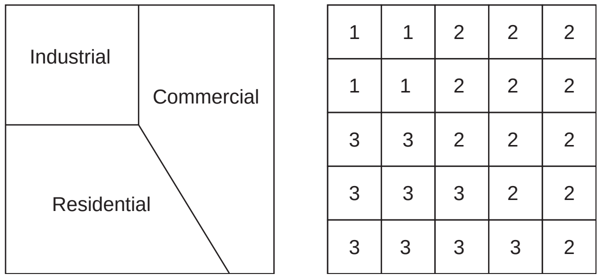

Figure 2 Vector and raster representations for geographic phenomena conceptualized using the field view, for example, nominal land cover. The vector approach is shown on the left, where the space is exhaustively partitioned into regions of uniform land cover (i.e., industrial, commercial, and residential). The raster approach is shown on the right, where each grid cell encodes the land cover at that location using values of '1' (industrial), '2' (commercial), and '3' (residential).

land cover value is considered to be constant within each region. Raster approaches for the field based view encode a value for each grid cell location (**Figure 2**). The vector and raster data models are implemented using a variety of data structures that describe the specific file organization, or database organization if implemented within a database, for geometric encoding as well as for the management of related nonspatial (i.e., attribute) data.

Extensions to the Conventional Spatial Data Models

Raster and vector data models have long been criticized for providing only weak representational ability because of their dependence on simple geometric primitives as the primary spatial unit of information to which other, nonspatial attributes are attached. For example, the vector model is relatively well equipped to represent geographic entities that are generally static (seldom change through time), spatially discrete (have crisp and easily discernable spatial boundaries), and simple (not composed of multiple parts that also must be represented within the database). However, there are clearly many entities in the world that are dynamic, have graded boundaries, and are complex (have multiple components), such as landslides and traffic jams. The representation of such geographic entities using vector and raster models is therefore problematic.

Researchers have utilized a variety of approaches for developing more robust geographic database representations. Many of these are computational analogs to the principles of geographic cognition described above. For example, criteria for membership within a category can be encoded as a set of rules (i.e., a rule base), relationships among objects can be encoded using a semantic network, and centrality of membership or uncertainty can be represented using fuzzy and rough sets. Arguably, the most prominent research direction in advancing geographic database representation has focused on object oriented approaches. Object oriented is term that has been applied to the design of programming languages, database design, and conceptual modeling strategies. It provides formal mechanisms for representing categorical hierarchies with taxonomic organizations, as well as part–whole relationships.

Object oriented techniques have been used to develop programming strategies for most efficiently storing geometric primitives and the relationships of composition that allow the construction of composite, or complex, geometric structures. In this sense, this approach may be considered an extension of the vector model with an object oriented wrapper that allows for a more sophisticated database representation than that allowed by the conventional spatial data models. The geometric approach has been criticized for limiting semantic database representation to that which can be represented geometrically. As an alternative approach, a number of authors have stressed the need to approach the issue of geographic database representation from a conceptual, as opposed to implementation, perspective. In other words, one should approach geographic database representation by first considering the nature of the domain to be represented – apart from the constraints of implementation or physical storage of the data.

Object oriented modeling has consequently been used to support the development and implementation of geographic ontology, the formal specification of a conceptualization of an application domain – how the salient features of an application domain may be symbolized in a formal manner. An ontology of a particular domain therefore specifies those entities and their properties that are deemed important for the purpose of representation in a computational environment. Research in ontology in the context of geographic databases addresses how database users' conceptual models of geographic domains may be elicited, formalized, and represented within a geographic database context.

Representation in Geographic Visualization

Cartographic Symbolization

Representation in geographic visualization can be introduced through cartographic symbolization. Cartographic symbols may be isomorphic to the geographic phenomena which they are intended to represent, or they may be more abstract in nature. The scale of the visual representation also plays a crucial role in cartographic symbolization, as the granularity of the visual representation is dependent in large measure on the scale. The term generalization refers to the degree of simplification and detail in the cartographic symbolization relative to the complexity of the geographic phenomenon the symbolization is intended to represent. Generally, the degree of generalization increases at smaller scales.

Common graphical symbol types are the geometric primitives point, line, and area symbols, as well as volu metric symbols. The placement and type of symbol typically indicates the location and basic categorical membership of a phenomenon. In addition, symbols have other attributes that are used to represent information about the phenomenon. These attributes are called the symbol dimensions, and include characteristics of shape, size, color (hue, value, and saturation), and pattern (orientation, arrangement, and texture). Other symbol dimensions have been proposed for dynamic displays, including display date, duration, order, rate of change, frequency, and synchronization.

Various combinations of the symbol types and attri butes may be matched to different levels of measurement, including nominal, ordinal, interval, and ratio data. Nominal data are categorical and indicate nonequivalency of membership, as with the division of land into agricultural and nonagricultural regions. Ordinal data imply a hierarchy of ranking of classes, without the ability to determine the metric measurement between pairs of classes. An example of the ordinal data type is the clas sification of cities into small, medium, and large. Interval data are numeric where the numeric unit remains constant throughout the range of the data. Ratio data are also nu meric and differ from interval data in that the origin of the numeric range is not arbitrary but meaningful. For ex ample, temperature in degrees fahrenheit is considered interval while temperature measured in degrees kelvin is ratio. An example of a map showing how symbol type and attribute may be related to level of measurement is pro vided in **Figure 3**, which shows a choropleth map of population density. Here, ratio level data are encoded in polygon symbols and visually represented by color.

Semiotics

The use of cartographic symbolization to represent geographic phenomena can be understood through the

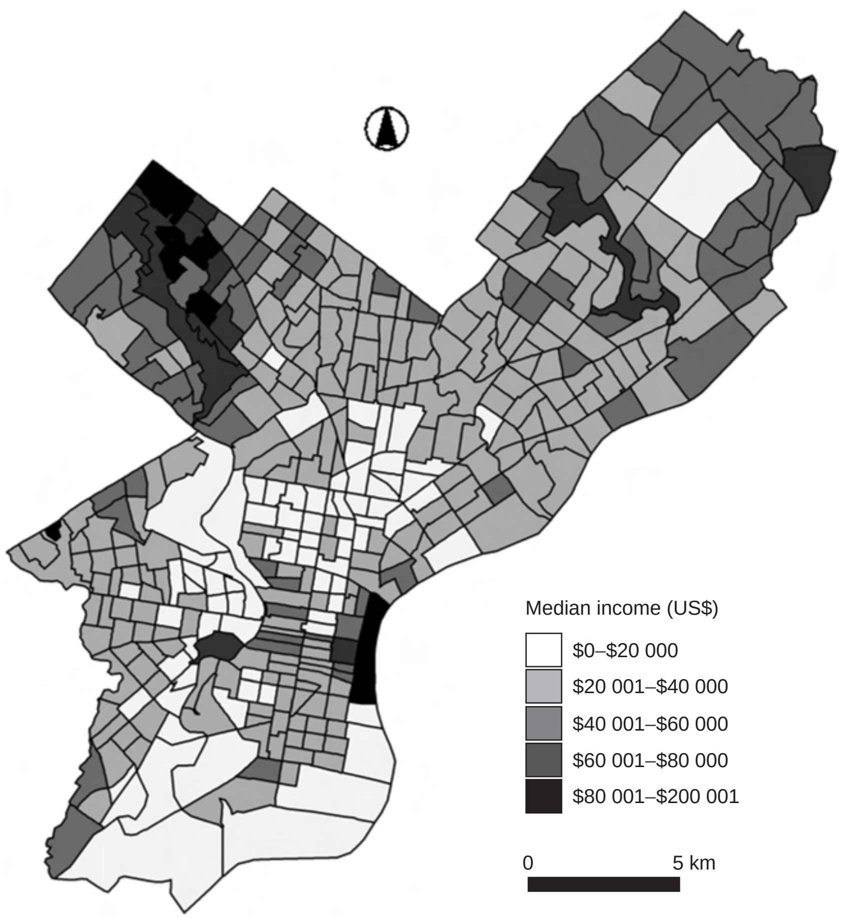

Figure 3 A choropleth map of 2000 median household income (US dollars) by US Bureau of the Census tract for Philadelphia, Pennsylvania, US. Each tract is represented as a polygon, and falls within one of five classes based on its income value. Each class is assigned a lightness value of gray based on its value range a darker shade of gray indicates higher income.

lens of semiotics, which focuses generally on how concepts are expressed and communicated, and is thus closely related to many of the principles of cognitive psychology described above. The semiotic triangle serves as a useful framework with which to understand cartographic representation, and describes the relationship between the sign, interpretant, and object, where each forms a corner of the triangle. The sign is a device that carries meaning, as with a symbol. The object refers to that phenomenon to which the sign refers. The interpretant refers to the conceptual meaning associated with the sign. The interpretant may be considered to mediate between the sign and the object, so that one understands which object a sign may refer to based upon a concept with which that object is affiliated.

In applying the semiotic triangle to representation in geographic visualization, the cartographic symbol may be considered a sign. Each cartographic symbol stands for a particular geographic feature, the object. The mediation of the sign with the object occurs through the interpretant, the concept that allows for a particular cartographic symbol to be matched to a particular geographic feature. It has been suggested that the relationship of object to interpretant is related to cognitive categorization, such that the object is an instantiation of the interpretant. Thus, a cartographic symbol is deemed appropriate for a particular geographic feature based on the feature's perceived membership within a given category of geographic phenomenon. Cartographic communication is dependent upon agreement among the map maker(s) and map reader(s) on the relationship between sign and object via the interpretant.

Cartographic symbolization may be organized by grouping map symbols according to how the sign–object relationship is established. Pictorial symbols seek to present a direct visual representation of the represented geographic phenomenon, while geometric symbols use geometric shapes that have no apparent relationship to the look of the phenomenon. Associative symbols present a visual form that suggests, or in some cases is a physical part of, a particular type of feature, such as the use of a cross symbol to stand for a church. Rules for map symbolization may be based broadly on the principle that similar objects should share similar signs.

Geographic Visualization

In geographic visualization, the emphasis is on data exploration, often using advanced computational technologies that extend the visual display of spatial information beyond the bounds of the traditional map. Visualization supports learning through cognitive processes whereby a display is seen and interpreted using stored knowledge about the visual patterns one might expect to find. These stored visual patterns may be considered schemata, which serve to identify instances of categories of geographic phenomena. The display itself is composed of signs. Thus, learning through visualization is the process of identifying patterns in signs and interpreting those patterns as an instantiation of a cognitive category. The process of learning is iterative, in which a viewer may identify sign patterns in the display, match those patterns to an existing schema, and revise the schema based on new information gathered from the visualization. This process may then initiate the viewer into interacting with the display to acquire new visual information, identify new sign patterns, and so on.

Critiques of Cartographic and Spatial Database Representation

The research emphasis on the technological aspects of representation in cartography and geographic information science has engendered a variety of critiques by those interested in the cultural and political contexts within which geographic representations are embedded. These critiques include criticisms of conventional approaches to cartographic and spatial database representation, as well as investigations into the underlying social and political processes that have produced those representations. Perhaps of foremost concern here is that cartographic or spatial database representations are socially constructed objects that not only reflect those aspects of the landscape that they are intended to represent, but also reflect the political, social, and cultural views of those creating the representation. Of particular note is the central role of spatial technologies in developing spatial database and cartographic representations, as the computer hardware and software used to develop a geographic representation necessarily constrains and influences the nature of that representation.

It has been argued, for example, that the reliance of conventional spatial data models on a Cartesian coordinate system within which to locate all represented geographic phenomena hinders the ability of maps and spatial databases to represent information about places and spatial relationships as conceptualized by many non-Western cultures. Additionally, there are many salient social characteristics and processes that are not well suited for representation by the simple geometric primitives of the conventional spatial data models. The emotional and spiritual characteristics that people ascribe to places, for example, do not easily lend themselves to representation using simple points, lines, and polygons.

A related issue concerns the ways in which geographic representations themselves may influence the very places and people they are intended to represent through their application in domains such as business and public policy. This concern stems from a confluence of ideas: (1) that spatial technologies necessarily constrain the types of geographic phenomena that can be represented; (2) that cartographic and spatial database representations are often thought of as representing an objective reality as empirically verified, and therefore tend to confer a ven eer of 'truthfulness' on the representation; and (3) that geographic representations and their application are often controlled by those with social and/or political power, and thus play a role in re enforcing *status quo* power relations among social groups. For example, maps of national boundaries may reinforce the political div isions they aim to depict, as well as influence the national identification of those nations' inhabitants.

See also: Spatial Analysis, Critical; Spatial Ontologies.

Further Reading

Bertin, J. (1983). *Semiology of Graphics*: *Diagrams*, *Networks*, *Maps*, Berg, W. J. (trans.). Madison, WI: University of Wisconsin Press.

Brewer, C. A., MacEachren, A. M., Pickle, L. W. and Herrmann, D. J. (1997). Mapping mortality: Evaluating color schemes for choropleth maps. *Annals of the Association of American Geographers* 87, 411 438.

Burrough, P. A. and Frank, A. U. (1996). Concepts and paradigms in spatial information: Are current geographical information systems truly generic? *International Journal of Geographical Information Science* 9, 101 116.

Egenhofer, M., Glasgow, J., Gunther, O., Herring, J. and Peuquet, D. (1999). Progress in computational methods for representing geographic concepts. *International Journal of Geographical Information Science* 13, 775 796.

Kosslyn, S. C., Chabris, C., Marsolek, C. and Koenig, O. (1992). Categorical versus coordinate spatial relations: Computational analyses and computer simulations. *Journal of Experimental Psychology: Human Perception and Performance* 18, 562 577.

MacEachren, A. M. (1995). *How Maps Work: Representation, Visualization, and Design*. New York: The Guilford Press.

Mark, D. M., Freksa, C., Hirtle, S. C., Lloyd, R. and Tversky, B. (1999). Cognitive models of geographical space. *International Journal of Geographical Information Science* 13, 747 774.

Mennis, J. L., Peuquet, D. J. and Qian, L. (2000). A conceptual framework for incorporating cognitive principles into geographical database representation. *International Journal of Geographical Information Science* 14, 501 520.

Peuquet, D. J. (2002). *Representations of Space and Time*. New York: The Guilford Press.

Pickles, J. L. (2004). *A History of Spaces: Cartographic Reason, Mapping and the Geocoded World*. London: Routledge.

Robinson, A. H. and Petchenik, B. B. (1976). *The Nature of Maps*. Chicago, IL: University of Chicago Press.

Tversky, B. and Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General* 113, 169 193.

Wood, D. (1992). *The Power of Maps*. New York: The Guilford Press.

Relevant Websites

http://www.geovista.psu.edu
GeoVista Center.

http://www.ncgia.ucsb.edu/giscc/units/u006/u006 f.html
Human Cognition of the Spatial World.

http://www.geoanalytics.net/ica/
International Cartographic Association Commission of GeoVisualization.

Research Funding Bodies

R. Johnston, University of Bristol, Bristol, UK

Origins of the Dual Funding Structure

University systems in many parts of the world – especially those where the English language predominates but also in Scandinavia and the Netherlands, for example – have evolved a dual structure for funding research activities. Alongside support for teaching and related activities, the universities' core funding streams sustain the institutions' basic research infrastructure, such as library and computing resources as well as staff research time; other sources have to be tapped from a second stream for grants/contracts to fund personnel, equipment, and other resources, plus travel costs (including attendance at conferences to present results) for specific projects and programs. Elsewhere – as in Germany and the former Soviet Union and to a lesser extent in France – universities are predominantly teaching institutions, with research funding concentrated on separate laboratories, not all of which are located within universities.

In the English speaking world, very little money was obtained by geographers through the second of these streams until after World War II. Research – as in many other (nonscientific) disciplines – was a small scale, individual activity, for which the main expenditure was on travel and subsistence while the researcher was in the field/archives collecting information. Research activity and productivity (through publications) was not a priority of most academics' work, and career prospects were not closely linked to research performance. Much of the work undertaken involved either original data collection in the field (including archival sources for historical studies) or the synthesis of material from a variety of secondary sources to provide accounts of the characteristics of a region or place. A great deal of it was done with virtually no financial support, undertaken in vacations almost as a sideline from the main task of teaching. As Pruitt describes the US situation:

> Before …1948, research was conducted by geographers mainly on a one person basis, financed by the individual with perhaps a small grant of a few hundred dollars from the university, the Social Science Research Council, or a foundation. It was mostly a summertime activity, or the occupation of a sabbatical. Occasionally graduate assistants had the opportunity to go into the field with the professor on these research undertakings if financial support could be found. The scale of the enterprise was generally modest and called for considerable sacrifice on the part of the geographer. Today, it is hard to believe that there was once a time when a person was expected to *pay* the costs of the studies he was conducting, and to finance his own travel, photography, drafting, and attendance at scientific meetings. [Although many geographers still contribute to such activities from their own incomes.] (Pruitt, 1979: 103)

Even relatively large projects undertaken at the time obtained little or no external support. The largest in the UK, for example, was the Land Utilisation Survey, launched in 1933 and directed by Dudley Stamp from the London School of Economics. Its goal was to map the use of every piece of land in Great Britain (a Northern Ireland survey was undertaken later). The mapping was undertaken by volunteers – mainly high school students – as was collation and interpretation of the data by academics. Small sums of money were obtained to cover administrative and other expenses (Stamp mortgaged his home at one stage to raise money) but most of the work was done for free – producing maps that proved extremely valuable for land use planning during and after World War II. Similarly, Estyn Evans relied on volunteers to do the fieldwork for his survey of prehistoric monuments in Northern Ireland and the large enterprise directed by Clifford Darby to map and analyze the 1086 *Domesday* returns for England involved no external funding, although Darby ensured that the university departments which he headed during the period employed cartographers to draft all of the diagrams.

Funding was available before 1939 to support certain types of research activities, especially travel, with the main goal being to allow scholars to interact with their peers in other countries and to develop an appreciation of geographical practices elsewhere. Thus, for example, in the 1930s, Robert Dickinson – a pioneer British urban geographer – had a Rockefeller Travelling Scholarship to spend the academic year 1931–32 in the USA familiarizing himself with the work of leading scholars; and in 1936–37 he had a further scholarship from the same source to spend a year in France and Germany for the same general purpose. But he was very much the exception: few geographers were able to raise the money to support their travels/subsistence and relieve them from their normal teaching and administrative responsibilities for such long periods. During World War II, a number of academics gained initial field experience overseas during their military service. A number of British academics also gained such experience through appointments or secondments to new university institutions in the British

Empire: Dudley Stamp was a professor of geology and geography at the University of Rangoon, 1923–26, for example; and Oskar Spate, who wrote a pioneering text on India and Pakistan, was at the same institution between 1937 and 1939; and Keith Buchanan held posts at both the University of Natal and the University of Ibadan before moving to the London School of Economics in 1953 and then to the Victoria University of Wellington.

This situation had been changing for other disciplines – especially those in science and engineering – which needed laboratories and equipment, plus dedicated staff to run them, in order to meet the needs of societies increasingly dependent on scientific (including medical) and technological research as the basis for economic and social advancement. Geographers were not entirely excluded from such sources – but as in general they did not undertake large research projects, they had little need for substantial resources.

Some of the first geographical research projects to obtain substantial external support – largely to pay for dedicated researchers to undertake large scale data collection, collation, and analysis – were funded through the US Office of Naval Research (ONR), established in 1946. Although most of its work focused on issues relevant to naval operations, a Geography Branch was established as a result of the efforts of Evelyn Pruitt (who obtained a geography PhD from UCLA in 1943 and then worked for the Coast and Geodetic Survey, Aeronautical Chart Branch). Under this branch, major research programs were established on the Arctic and in coastal studies generally (the latter including the Coastal Studies Institute based at Louisiana State University's department of geography), and there was funding for pioneering work on air photo interpretation and what later became known as remote sensing. In addition, the branch also supported various aspects of systematic and regional geography, including pioneering studies of urban central business districts and separate investigations led by William Garrison (at the University of Washington), Brian Berry (University of Chicago), and William Warntz (at the American Geographical Society). Much of this involved cutting edge quantitative studies, including the use of computers, in both human and physical geography. (On such early research, see Barnes, 2008.)

Winning Recognition

Although the ONR's Geography Branch pioneered funded research by geographers and facilitated a considerable number of PhD submissions, human geography in the US – as elsewhere – was on the margins of large scale research funding development. Particularly important for this was the National Science Foundation (NSF), the largest source of public funds for (nonmedical) research in US universities: its expenditure in 2006 was $6406 million. Geography – especially human geography – was not recognized as a science by this body and although some individuals were able to tap its resources, there was no established path for human geographers nor any moneys set aside specifically for the discipline. To alter that situation, geographers lobbied the NSF and the National Academy of Sciences (the country's preeminent scholarly society) for an opportunity to make their case for sustained support. A committee produced a book on *The Science of Geography*, which promoted the discipline as a 'useful science' through the examples of four topical areas: physical geography, cultural geography, political geography, and location theory. As a result, geography gained recognition within the NSF framework with the establishment of a geography and regional science program – directed by a series of academics seconded to NSF. This has provided substantial funding for a range of geographical work over the last 40 years, including the major initiative taken when Ron Abler was the program director to fund a national center for work in GIS.

For human geographers, there was the allied issue of gaining recognition for their discipline as a social science, to which there was considerable resistance. In the late 1960s, the US National Academy of Sciences and the Social Science Research Council (SSRC) jointly funded a Survey of the Behavioral and Social Sciences. Geography was included within this, though reluctantly: Gauthier reports that

> … geography initially had not been selected to participate in the survey, because the other panels in economics, sociology, psychology, anthropology and political science did not view the field as a viable social science. Not until the geography panel issued its report in 1969 was there any acknowledgement that geography was indeed a legitimate social science. The panel members changed people's minds by making a strong argument for geography as a social science that made contributions to society and the development of public policy. (Gauthier, 2002: 577)

The report stressed the location theory (or spatial science) approach to human geography even more than *The Science of Geography* did, emphasizing that human "geography is concerned with giving man an orderly description of his world" (Taaffe, 1970: 5). It presented the discipline as an integral component of the social science portfolio, increasingly sophisticated in its analytical tools, focusing on spatial organization, and offering particular skills in mapping and data acquisition.

These two reports, and the greater research activity by human geographers within their much altered discipline, facilitated increased access to research funds from *c.* 1970 onward and helped to gain the discipline a much higher profile nationally within the scientific and, especially, social scientific communities. (In the first 8 months of 2007, the NSF's Geography and Regional Science Pro gram allocated grants totaling some $30 million, some of which was to support doctoral dissertation projects.) Nevertheless, the position remained somewhat tenuous, and in the 1990s the Association of American Geog raphers successfully lobbied for a further opportunity to promote their discipline to the national scientific com munity – because of a "well documented perception (external to geography as a discipline) that geography is useful, perhaps even necessary, in meeting certain societal needs" (NAS–NRC, 1997: ix). The report on *Rediscovering Geography: New Relevance to Science and Society* once again stressed human geographers' technically based analytical skills, especially those involving spatial analysis and GIS.

In the UK, as in the USA, human geographers were required to establish their credentials within the social sciences in order to gain access to national sources of research funding. An SSRC was not established there until 1965 (a Medical Research Council, a Science and En gineering Research Council, and a Natural Environment Research Council already existed), and human geography was not included in the disciplines for which it initially offered both research funds and postgraduate student support. To counter this, a group of senior academics asked if they could present a case for their discipline's inclusion to the SSRC chairman, who agreed to receive it. Preparation of the case was largely undertaken by Michael Chisholm, against some opposition within the discipline from senior individuals who wished to maintain the *status quo* and continue to get studentships (though no access to research funds) through the established route of the De partment of Education. Their opposition was largely against their discipline being repackaged as a social sci ence, something that Chisholm then did in a number of books. With the clear support of the SSRC Chairman Lord Young, the campaign succeeded – though there was little apparent enthusiasm shown toward geography by representatives of most of the other social science disciplines: human geography joined the SSRC (later renamed the Economic and Social Research Council – ESRC) in 1967, though rather than have a separate committee for the discipline (as was the case for all of the founding disciplines) it was linked with planning for the allocation of research grants: there was a separate geog raphy subcommittee for allocating research studentships. Among the first recipients of research grants was Peter Haggett, and major SSRC funded projects directed by Alan Wilson were operating at Leeds during the 1970s. Separate disciplinary committees were abolished in the mid 1980s, and replaced by a Research Grants Board and a Training and Development Board which handled grants and studentships, respectively, and on which each major discipline – including human geography – is represented. There is also a Research Resources Board which invests in major, social science wide initiatives, such as the Data Archive, and in obtaining free access for academic users to all UK census data.

Funding Streams

Research councils and other funders operate in two main modes – responsive and directive. Under the re sponsive mode, an amount of money is set aside each year to fund projects proposed by individual researchers and research groups. Applications are peer reviewed, and the allocations go to those with the highest evalu ations in a competitive situation. Under the directive mode, the funders determine particular lines of research in which they want to invest, and invite applications against a set brief. Those lines might be particular re search themes of immediate public interest – such as devolution in the UK since 1998 and the major Can adian Metropolis project; they may involve sustaining research activity, such as funding widely used surveys (the American National Election Study, the British Election Study, and the British Household Panel Study, for example), making census data freely available to academic researchers, promoting major cross disciplin ary research initiatives (such as GIS, which was sup ported by the ESRC funded Regional Research Laboratories in the UK and the NSF funded National Center for Geographic Information and Analysis) or establishing a Data Archive where datasets are stored and disseminated for further analysis.

In the UK, the research councils also provide funding for postgraduate students with award schemes covering both tuition costs and a salary. The ESRC now operates what is termed a '1 + 4' system, whereby the first year of a studentship involves attendance at formal courses de signed to provide the breadth and depth necessary for a research career (for which a separate qualification may be earned – such as MRes – and the other three are deployed in undertaking the research for a PhD degree). Only a small number of departments are recognized to provide the basic training, and students compete for the awards available there; a slightly larger number can accept stu dents for the research component. All UK universities offer research degrees, however, and some offer financial support. In North America, most postgraduates are supported by their individual universities: some offer competitive scholarships; most employ postgraduates as part time teaching and/or research assistants. Support for fieldwork and other research expenses may also be avail able, but may have to be obtained through competitive grants from the NSF and other bodies – in some cases as part of a research grant awarded to students' supervisors.

Recognition by the main research funding bodies in the US and UK not only gave human geographers access to important – and prestigious – sources of money, but it also established their discipline's credibility within the social sciences, facilitating the search for support from a wider range of sources. These include major charities established to support research in the social sciences and other disciplines: in the UK, these include the Nuffield Foundation, the Leverhulme Trust, the Joseph Rowntree Foundation (which is linked to several separate trusts), the Carnegie UK Trust, and the Wellcome Trust. There is also a wide range of other sources for research funding which human geographers are able to tap, including the British Academy, the recently established Arts and Humanities Research Council, many government departments (national, regional, and local) and various arms of the European Union (EU) which has become a major source of research funding.

The EU operates through a series of Framework Programmes, each of which has a major theme: that for the Seventh Framework Programme (covering 2007–13) is competitiveness and innovation, which has four subprograms – cooperation (budget, €32 billion), ideas (€7.4 billion), people (€4.7 billion), and capacities (€4.2 billion). Grant allocations are made by a European Research Council following similar procedures (emphasizing peer review) to those operated by national bodies. While some parts of the program are quite specific (€2.3 billion on energy research, for example), others are more general – such as €1.8 billion for research infrastructures (data archives, for example) and €4.7 billion for career development grants: under the ideas component, the European Research Council has €7.4 billion to spend on "the most talented scientists, supporting risk taking and high impact research, and promoting world class scientific research in new, fast emerging fields". The opportunities for geographers – usually in collaborative, international teams – are very substantial (as with the European Spatial Planning Observation Network: ESPON); a list of some EU programs in which UK geographers were involved in the period 1996–2000 was made by Thrift and Walling, which also discusses non EU international collaborative programs in which UK geographers have participated.

US human geographers also have a wide range of potential sources, including the prestigious John Simon Guggenheim Memorial Foundation, whose fellowships have funded research by a number of scholars, largely by paying for them to be released from their teaching and administrative responsibilities. Of course, US geographers participate in major international collaborative programs organized by a variety of organizations – some under the umbrella of the International Council of Science and others through nongovernmental organizations (NGOs) such as the World Wildlife Foundation.

Research Contracts

Although much emphasis with regard to research funding goes to the winning of grants, in part because they are perceived as prestigious, being allocated on the basis of peer reviewed competition, considerable attention is also given – and increasingly so within human geography – to obtaining research contract income. Contracts differ from grants in that the latter generally involve what is sometimes termed 'blue skies' research – the pursuit of knowledge for its own sake with little or no concern about its application (although research councils and others are increasingly concerned – and pressed by their government funders – to have 'user inputs' and focus on 'useful knowledge production') – whereas contracts are normally undertaken for an end user (generally, though not always, in the private sector) as a form of applied research. Academics are encouraged only to obtain contracts that are fully funded, which have a substantial intellectual component (i.e., there is original research involved), and from which the results can be published (i.e., do not remain classified because they are commercially sensitive), but the availability of money for research that might otherwise not get funded is important.

Many geographers have undertaken funded research through such channels. Some major research centers have been established in this way, such as the G Map company established at the University of Leeds to undertake location allocation studies for commercial and public sector clients (later sold to commercial clients and operating as GMAP Consulting) and the Flood Hazard Research Centre at the University of Middlesex.

Research Funding and Research Evaluation

Obtaining research funding is now a central part of academic activity in the British and North American university systems: without the separate inputs from research councils, charities, government departments, and other sources, the dual funding structure would not operate. There are several reasons for this, the most important of which is that research is becoming increasingly expensive, because it is more labor intensive in the collection and analysis of data. Major research projects can no longer be operated as summertime activities: they are year round enterprises, many of which require the principal investigators to be released from their other increasingly onerous university tasks linked to teaching and administration, and a considerable number also require associated research assistance – from postdoctoral workers (gaining additional research experience and expertise as the first stage of an academic career), postgraduate students dedicated to parts of the project or program, and also from contracted workers, such as those engaged to conduct major surveys as part of a research initiative.

That funding sustains not only the research activity itself but also the infrastructure on which it is founded. Most of the national research councils provide an 'overhead' element additional to the direct costs of the project, for example, which is a contribution to the host institution's costs of managing and underpinning that activity – as well as preparing for further rounds of ap plications and investments. In the UK, for example, the overhead element was for some years set at *c.* 40% of the direct costs of a project but the principle of 'full eco nomic costing (FEC)' for research grants has now been accepted, and means that the overhead element may be 100% or more of the direct costs. Some research funders, such as a number of charities in the UK, decline to meet this overhead element – claiming, for example, that the infrastructure for research in universities should be funded out of general taxation, to which they are major contributors; the funding formulae for universities take this into account, providing a parallel stream of income for researchers/institutions which cannot get the FEC of their grant/contract income.

Research grant income is also important because it is increasingly used as an indicator of research excellence, on the argument that the more you earn to undertake research the better you must be at it. This is particu larly important in the UK where a series of six Research Assessment Exercises (RAEs) has been held since 1986 as part of the evaluation of research activity and ex cellence. Increasingly, the allocation of the infrastruc tural funding for research in universities (the so called quality related funding QR element) is determined by the results of this exercise: the departments with the highest grades (currently on a seven point scale) get the bulk of the money, with those in the lowest categories getting none.

Those RAEs have largely been based on evaluations of research outputs, notably publications, although depart ments/universities have been required to provide infor mation on the number and size of research grants/ contracts won as well as the number of postgraduate studentships. In 2006, however, the government launched an initiative to replace those evaluation processes – at least in the sciences – by a metrics based procedure, in which the QR allocations would be a function of certain indicators. Research funding was presented as the main candidate, although other metrics – such as citation counts – are also being considered. If (more probably when) this system is introduced, earning income will clearly become the core of most research activity there – although some activities will remain which the individual researcher can undertake with little more than a modi cum of support for, say, travel.

See also: Citation Geography; Cross-Cultural Research; Education; Knowledge Economy.

Further Reading

Barnes, T. J. (2008). Geography's Underworld: The Military industrial Complex, Mathematical Modelling and the Quantitative Revolution. *Geoforum* 39, 3 16.

Buchanan, R. H. (2006). Emyr Estyn Evans 1905 1989. In Armstrong, P. & Martin, G. J. (eds.) *Geographers: Biobibliographical Studies*, vol. 25, pp 13 33. London: Continuum Books.

Chisholm, M. (1971). *Research in Human Geography*. London: Heinemann.

Chisholm, M. (1975). *Human Geography: Evolution or Revolution*. London: Penguin Books.

Chisholm, M. (2001). Human geography joins the Social Science Research Council: Personal recollections. *Area* 33, 428 430.

Chisholm, M. and Rodgers, H. B. (eds.) (1973). *Studies in Human Geography*. London: Heinemann.

Darby, H. C. (1977). *Domesday England*. Cambridge: Cambridge University Press.

Fisher, C. A. (1979). *Three Times a Guest: Recollections of Japan and the Japanese, 1942 1969*. London: Cassell.

Gauthier, H. L. (2002). Edward 'Ned' Taaffe 1921 2001. *Annals of the Association of American Geographers* 92, 573 583.

Johnston, R. J. (2004). Institutions and disciplinary fortunes: Two moments in the history of UK geography in the 1960s II: Human geography and the Social Science Research Council. *Progress in Human Geography* 28, 204 226.

Johnston, R. J. (2006). Research quality assessment and geography in Australia: Can anything be learned from the UK experience? *Geographical Research* 44, 1 11.

Mead, W. R. (2007). *Adopting Finland*. Helsinki: Niilo Helander Foundation.

Mikesell, M. W. (ed.) (1973). *Geographers Abroad: Essays on the Problems of and Prospects of Research in Foreign Areas*. Chicago, IL: University of Chicago, Department of Geography. Research Paper 152.

Murphy, R. E. (1972). *The Central Business District*. Chicago, IL: Aldine Atherton.

NAS/NRC (1965). *The Science of Geography*. Washington, DC: The National Academy of Sciences/National Research Council.

NAS/NRC (1997). *Rediscovering Geography: New Relevance for Science and Society*. Washington, DC: National Research Council.

Pruitt, E. L. (1979). The Office of Naval Research and Geography. *Annals of the Association of American Geographers* 69, 103 108.

Sidaway, J. D. and Johnston, R. J. (2007). Geography in higher education in the UK. *Journal of Geography in Higher Education* 31, 57 80.

Spate, O. H. K. (1954). *India and Pakistan: A General and Regional Geography*. London: Methuen.

Spate, O. H. K. (1965). *Let Me Enjoy: Essays, Partly Geographical*. Canberra: Australian National University Press.

Stamp, L. D. (1936 1950). *The Land of Britain: The Report of the Land Utilisation Survey of Britain (92 parts)*. London: Geographical Publications Ltd.

Stamp, L. D. (1948). *The Land of Britain: Its Use and Misuse*. London: Longman.

Taaffe, E. J. (1970). *Geography*. Englewood Cliffs, NJ: Prentice Hall.

Thrift, N. J. and Walling, D. E. (2000). Geography in the United Kingdom 1996 2000. *The Geographical Journal* 166, 96 124.

Turner, B. L. II (2005). International research agendas and practice. *Progress in Human Geography* 29, 173 177.

Walker, H. J. (2006). Evelyn Lord Pruitt. *Annals of the Association of American Geographers* 96, 432 439.

Willatts, E. C. (1987). Geographers and their involvement in planning. In Steel, R. W. (ed.) *British Geography 1918 1945*, pp 100 116. Cambridge: Cambridge University Press.

Relevant Websites

http://canada.metropolis.net
About Metropolis Canada: Overview, Metropolis.
http://www.ahrc.ac.uk
Arts and Humanities Research Council.
http://www.arc.gov.au
Australian Research Council.
http://www.britac.ac.uk
British Academy.
http://cordis.europa.eu
CORDIS.
http://www.devolution.ac.uk
Devolution and Constitutional Change.
http://www.esrc.ac.uk
Economic and Social Research Council.
http://www.espon.eu
EPSON.
http://ec.europa.eu
European Commission Research, EUROPA.
http://www.fhrc.mdx.ac.uk
FHRC (Flood Hazard Research Centre).
http://www.gf.org
Foundation Program Areas, 2005, John Simon Guggenheim Memorial Foundation Home Page.
http://www.rsnz.org
Funding, Royal Society of New Zealand.
http://www.eurodirect.co.uk
GMAP Consulting: The Market Analysis & Retail Planning Experts, GMAP.
http://www.ihdp.uni bonn.de
IHDP (International Human Dimensions Programme on Global Environmental Change).
http://www.ipcc.ch
IPCC (Intergovernmental Panel on Climate Change).
http://www.unesco.org
ISSC (International Social Science Council), UNESCO.
http://www.icsu.org
ISU (International Council for Science).
http://www.gf.org
John Simon Guggenheim Memorial Foundation.
http://www.neh.gov
National Endowment for the Humanities.
http://www.nih.gov
National Institutes of Health.
http://www.noaa.gov
National Oceanographic and Atmospheric Administration.
http://www.nrc cnrc.gc.ca
National Research Council.
http://www.nationalacademies.org
National Research Council, The National Academies.
http://www.nsf.gov
National Science Foundation.
http://www.nerc.ac.uk
Natural Environment Research Council.
http://www.nsf.gov
Social Science Research Council: National Science Foundation.
http://www.fordfound.org
The Ford Foundation.
http://www.jrf.org.uk
The Joseph Rowntree Foundation.
http://www.leverhulme.ac.uk
The Leverhulme Trust.
http://www.nuffieldfoundation.org
The Nuffield Foundation.
http://www.fulbright.co.uk
The US UK Fulbright Commission.
http://www.wellcome.ac.uk
The Wellcome Trust.

Resistance

J. Cupples, University of Canterbury, Christchurch, New Zealand

Glossary

Discourse Poststructuralist concept used to refer to the circulation of meanings and ideas in a way which establishes relations of power.
Everyday Life Used to refer to activities or events which occur on a routine or everyday basis.
Globalization A series of economic, political, and cultural processes which increase the global circulation of goods, services, people, and ideas and promote more intense interconnections between places in the world.
Hegemony Concept elaborated by Antonio Gramsci to explain the ways in which the state and the institutions of civil society secure consent for the rule of dominant social groups so that it appears to be natural or legitimate.
Identity Politics The mobilization of particular social identities such as gender, class, or sexuality in order to achieve political goals.
Neoliberalism Economic model promoted by international financial institutions which emphasizes free trade, privatization, and a reduced role for the state.
Resistance Organized or spontaneous action which attempts to bring about or prevent political or social change.
Transnational Used to refer to a geographical process which operates beyond or outside of the borders of the nation-state.

Introduction

There is a wide range of geographic scholarship with engages with the concept of resistance, broadly understood as the political and cultural struggles carried out by social actors to oppose, challenge, or undermine dominant workings of power or the normative production of space. Human geographers in their endeavors to understand and address the uneven and unequal world in which we live have focused attention on the concept of resistance from a wide range of theoretical perspectives.

Resistance encompasses many multifaceted forms of response, struggle, and activity. Resistance can be organized, collective, and highly visible taking the form of armed struggle, terrorism, strikes, boycotts, riots, occupations, marches, and demonstrations. Resistance can also be more subtle, mundane, individual, and hidden and take place as part of the practices of everyday life. Everyday forms of resistance might include coping strategies, smoking, gossiping, time wasting, watching soap operas, consuming/not consuming particular products, or using spaces and places in ways that were not intended by planners or owners. Resistance is also commonly expressed in many cultural products such as music, painting, graffiti, blogs, poetry, films, and novels. While defining resistant practices might appear to be straightforward, many practices are far more contradictory or ambivalent and combine resistance with elements of compliance with dominant power relations. Resistance can also be reactionary, as well as progressive. It can be used to bring about political or social change but it can also be used to prevent change and resecure a dominant order.

Essentially geographic scholarship on resistance illuminates the possibilities which exist for people to effect political and social change in their lives and those of others and the geographies that are created through those practices.

Theories of Resistance

The geographic scholarship on resistance draws on a diverse range of theoretical perspectives including Marxism and Gramscian theory, feminism, poststructuralism, and postcolonialism. In conventional Marxist or liberal accounts, resistance was understood to oppose power or domination in a binary form of confrontation. Power was seen as located in particular institutions such as the state or the media or possessed by certain social classes to the detriment of others. According to Marxist thought, revolutionary political change would be brought about through class conflict and the inherent contradictions of capitalism. These conventional accounts understood resistance as a kind of once and for all struggle for victory and paid little account to the everyday capacities of ordinary people to resist and shape on a micro level the world in which they live and to the inherent instability and fragmentation of domination.

There is a broad consensus that conventional structural forms of analysis have proven inadequate to grasp the cultural dimensions of political struggles and the precariousness of dominating power. Scholarship in the new cultural geography which draws on post Marxist, poststructuralist, and postmodernist insights and pays attention to questions of difference and identity, has highlighted the inability of conventional structural analyses to get to grips with the cultural politics of re sistance. It has become clear that many political struggles are concerned as much with the transformation of meaning as they are with changing difficult material conditions or gaining access to resources.

Many scholars in the new cultural geography and cultural studies turned to the work of Italian theorist, Antonio Gramsci to escape the reductionism of con ventional Marxist accounts. Gramsci asserted that dom ination or hegemony is secured not through coercion but through common sense. Dominant groups are able to exercise hegemony by gaining the consent of subordin ated groups. Consent is secured and hegemony exercised in the cultural sphere where meanings are made, such as in the institutions of civil society. While hegemony might function as a consensual order, it is however always fragile and unstable and subject to de stabilization by counter hegemonic forces. Gramsci believed that capit alism was subverted by ‘organic intellectuals’ who ar ticulate counter hegemonic discourses which challenge the commonsense understandings of the dominant.

Geography’s understanding of how power works and the importance of the cultural sphere as a site of political engagement and change were similarly revolutionized by the work of Michel Foucault. Foucault saw power not as belonging to any particular group or class but as far more diffuse and de centered and as something which is productive and enabling, as well as repressive and con straining. For Foucault, power is exerted through tech niques rather than being something which is possessed like property. This understanding, which refuses a binary conception of power, means that everybody always has some power at their disposal and that domination and resistance are instead mutually constituted, entangled, and thoroughly hybrid. This line of thinking on resist ance was taken up in an edited collection by Joanne Sharp *et al.* This text explores the spatiality of en tanglements of domination and resistance and em phasizes that while domination and resistance cannot exist independently of one another, they are not re ducible to one another either. Consequently, as geog raphers we can focus on how (fragmented) power at times becomes sedimented in particular spaces or institutions producing injustice and discrimination.

Foucauldian insights on power and resistance have been extended by a number of theorists including Michel de Certeau and James Scott. These theorists emphasize that even ordinary people who are marginalized by existing forms of domination have some power at their disposal to remake the landscapes in which they live. De Certeau conceptualizes the everyday practices of resist ance as tactics, by which he means the playing with and manipulating of space in ways which escape or subvert authority. Tactics are distinct from the strategies deployed by dominant groups such as the state or the military who have the power to construct and control place. Similarly, James Scott’s work on peasant resistance in Malaysia also focuses on the minor and more hidden subversions such as foot dragging, petty sabotage, and feigned ignorance. Scott calls these everyday forms of resistance infrapolitics and he makes a distinction between what he calls public transcripts, or the discourses of the powerful, and hidden transcripts, the discourses or weapons of the weak. Re sistance can therefore take place without it being noticed by the dominant and over time such resistance can gen erate alternative ways of knowing and being.

Feminist geography has also made an important contribution to our understanding of both organized and everyday resistance. Feminist geographers are interested in the gendering of particular spaces and places as sites of oppression or resistance. A focus of much feminist geo graphic inquiry has been into how particular practices engender resistant subjectivities and how gender and other social identities and place meanings are renegoti ated through such practices. The political struggles which emerge from women’s everyday lives and practical difficulties and constraints (such as fighting for clean water or electricity or access to paid work or freedom from domestic violence) make decisive contributions to the transformation of dominant understandings of femi ninity, sexuality, or domesticity and often serve to blur and de stabilize dichotomized understandings of public and private.

Development and Resistance

In terms of development geography, a significant com ponent of resistance scholarship has documented resist ance to development in its multiple forms, including colonialism, neocolonialism, imperialism, Westernization, modernization, and neoliberalism.

Colonialism as an early exercise in development be fore World War II produced its own forms of both overt and covert resistances by the colonized. Geography as a discipline played a key role in colonial expansion through its cartographic production and through the development of environmentally determinist accounts of people in colonized parts of the world. Human geography has begun to engage with the legacies of its colonial origins through engagement with key debates in postcolonial studies and through empirical treatments of historical resistance to colonialism and slavery, as well as explor ations of the colonial present. These intellectual en deavors serve to question and expose geography’s past and present complicity with imperial relations of all kinds and constitute a crucial part of intellectual attempts to de colonize geography as a discipline.

Development geography has also provided insights into the resistance movements and new social movements which have developed more recently in response to the failures of development, particularly in its modernizing and Westernizing guises. In this respect, the new social movement (NSM) school has been particularly influen tial. NSMs are broadly understood as a new kind of political activism which emerged in the 1970s and 1980s in response to changes in late capitalism and in which activists moved away from formal spheres of politics such as trade unions and political parties and began to or ganize around diverse axes of identity such as gender, sexuality, ethnicity, and so on. NSMs include women's movements, indigenous movements, environmental groups, neighborhood lobbying organizations, human rights groups, and lesbian and gay movements. They are often perceived to be more spontaneous and partici patory and less hierarchical than conventional political movements. The NSM approach has broadened our understanding of what constitutes resistance and how it is to be theorized.

Many recent resistance movements in both the First and the Third Worlds have developed in response to neoliberalism and the Washington Consensus, the set of economic policies promoted and designed by the World Trade Organization (WTO), the International Monetary Fund (IMF), and the World Bank. As debt burdened nation states have embraced (or been forced by the international financial institutions to embrace) forms of economic globalization such as trade liberalization, pri vatization, and cuts in public spending, they have become subject to political challenges from within. Such policies have had devastating social and environmental costs in many places, bringing inequality, exclusion, poverty, and environmental degradation and exacerbating the already precarious situation of small farmers, women, and in digenous people. Resistance to free market capitalism and the predatory forms of global governance to which Third World nations are often subject takes many col lective and individual forms including increased migra tion and crime but often also includes spectacular high profile political action which attracts significant academic and media attention. Examples of this nature include the Zapatista uprising in Chiapas in Southern Mexico to protest the North American Free Trade Agreement (NAFTA) and other dimensions of neoliberalism, the mobilizations in Brazil by the Movimento Sem Terra (MST) to challenge the overwhelming concentration of agricultural land in the hands of a minority, the protests against water privatization in Cochabamba, Bolivia and movements to prevent the construction of World Bank sponsored and environmentally destructive dams along the Narmada River in India.

Resistance to neoliberalism also includes everyday forms of mobilization. Women across Latin America, resistant to the idea that they should be the shock ab sorbers under neoliberal economic policies, have developed a multiplicity of strategies to cope with falling real wages and declining social services. These strategies include running communal kitchens in order to feed their families, creating handicrafts or saving cooperatives, lobbying government agencies or nongovernmental or ganizations (NGOs) to secure drinking water or elec tricity in their communities, ending relationships with unproductive or violent partners, and pulling children out of school. Such place based and community oriented approaches to struggle and resistance have been extended by the work of J. K. Gibson Graham, whose work has shown how the sense of inevitability promoted by pro ponents of neoliberalism is fragmented by the everyday anticapitalist practices which are adopted by people negatively affected by neoliberalism. There are hundreds of examples in which ordinary and marginalized people have mobilized to defend lands, livelihoods, and en vironments under threat. All of these movements raise important questions about nationhood, citizenship, and belonging and often serve to undermine many of the Western and masculinist biases of the development process.

Up until recently, resistance to neoliberalism involved mainly extra institutional forms of struggle, but is now increasingly involving governments of the South. The political landscape of Latin America has shifted con siderably in the last few years as a result of the election of a number of Latin American governments who cam paigned on anti neoliberal tickets. The governments of Venezuela, Bolivia, Ecuador, and Nicaragua are all seeking to find alternatives to economic dependency and modify the forces of globalization through the national ization of key natural resources such as oil and natural gas and through the creation of alternative trading blocs to the US sponsored Free Trade Area of the Americas (FTAA) such as the Bolivarian Alternative for the Americas (ALBA). Venezuelan president, Hugo Chávez and Bolivia's first indigenous president, Evo Morales have become important reference points and symbols of re sistance for activists worldwide.

There is also significant and growing resistance to free market capitalism in the First World which is fre quently developed in solidarity with the Third World or as a protest about the effects of neoliberalism in the Third World. Meetings held by the international financial institutions are now almost always accompanied by pro tests and large amounts of money are spent on security to combat what some media commentators unhelpfully refer to as 'antiglobalization' activists. In recent years, there have been high profile protests at a whole range of meetings, including the WTO meeting in Seattle in 1999, the IMF and World Bank meeting in Prague in 2000, the G20 meeting over the FTAA in Quebec City in 2001, the G8 summit in Genoa in 2001, the WTO meeting in Cancún in 2003, and the G8 summit in Scotland in 2005.

Resistance to the neoliberal economic model in the First World is also driven by alternative consumer practices through the boycott of goods which involve exploitative labor relations or environmentally destructive practices or through the consumption of ethically and sustainably produced products. The growing consumption of fair trade coffee, tea, chocolate, and other products creates a form of mutually beneficial solidarity between Third World producer and First World consumer and produces tangible benefits in the lives of producers and their families.

The Globalization of Resistance

While many resistance movements are place based, they are also increasingly global or transnational in their scope and generate connections beyond the countries and regions in which they take place. The Zapatista rebellion, which involved masked guerrillas taking over a number of towns and cities in Chiapas on New Year's Day, 1994, the day that NAFTA came into force, was well timed to take advantage of the global media attention focused on Mexico because of NAFTA. An innovative use of Internet and other forms of global media created massive transnational solidarity with the Zapatistas, who have been referred to as cyber Indians. In the 1990s hundreds of websites dedicated to the Zapatista cause appeared and spread an anti NAFTA, anti neoliberal, and pro indigenous message across the world. The information on the Internet was not however distributed by the Zapatistas themselves from a central command site but developed in many different sites around the world. Their messages and communiqués combined counter hegemonic discourses which had originated elsewhere and were already in global circulation with indigenous discourses of reciprocity and communal land ownership. Other indigenous resistance movements across the Americas have also gained transnational support and strengthened their claims for land rights and autonomy by engaging with First World environmentalist discourses, as well as indigenous understandings. The Bolivians who successfully evicted corporate water giant Bechtel from Cochabamba in 2002 deployed Andean cosmological discourses to spread an anti privatization message and stress the importance of water as a fundamental human right.

Globalization and the development of information and communication technologies (ICTs) such as the Internet, e mail, and satellite television have brought resistance movements into closer dialog and promoted more unified forms of struggle. It is now possible to share ideas and information and coordinate activities in ways which enable activists and the discourses and ideas they work with to 'jump scale' or to operate across various scales simultaneously. ICTs have therefore facilitated the creation of spatially stretched networks of social movements and NGOs who are united in their opposition to create alternative futures for people and places harmed by corporate power and free trade. One such powerful network is the World Social Forum (WSF) which brings together activists from all over the world under the slogan "Another World Is Possible." Since its first modest meeting in Porto Alegre in Brazil in 2001, the WSF has blossomed into an international network of activists which enact a kind of globalization from below. In 2007, the Forum met in Africa for the first time with some 50 000 organizations present in Nairobi.

Geography and Resistance

An important component of the study of resistance from a geographic perspective is the analysis of the multiple geographies at play which forces us to think simultaneously about the geographies which make resistance possible or necessary and about the geographies which resistance creates. While human geographers are not the only scholars in the social sciences and the humanities to study resistance, one of geography's contributions to this body of literature has been to enhance the theoretical attention to space, place, and scale. Given that the operation of power is inherently spatial, human geographers and others who take spatiality seriously have a unique contribution to make to the theorization of resistance. Indeed, geography's key contribution has been to highlight how political and politicized spaces are created, occupied, and maintained, often beyond the realm of institutional politics, as well as within it. The scalar dimensions of resistance are also particularly important to consider. The transnational character of many resistance movements has forced a rethinking of how the local shapes the global and vice versa and how the local and the global are not natural spheres but are constituted through situated practices and made effective in specific places and spaces. Recent work has however questioned the notion that there is such a thing as a global civil society united in its rejection of global neoliberalism given ongoing place based and identity centered differences in positions and objectives and competing interests among different social movements. At any rate, it seems pertinent to conceptualize the politics of resistance as cultural politics. As spaces and places are reshaped through acts of resistance, meanings and knowledge are simultaneously rearticulated in ways which might be more progressive or liberatory.

Despite these theoretical advances, resistance defies easy theorization and there are a number of ongoing and unresolved debates. One problem with the concept of resistance is that resistance is often considered to be everywhere. If resistance is indeed so ubiquitous, it raises questions about whether it can be theorized as a distinct and well defined practice. Some people might question whether the coping strategies deployed by people in order to survive in the face of severe hardship constitute acts of resistance. There is therefore no clear consensus about how to recognize resistance or how to distinguish a resistant act from a compliant one or an ambivalent one. One view is that resistance needs to contain some element of intentionality. While this approach is certainly valid to some extent for organized political movements, it does not capture the moments when dominating power is undermined accidentally or spontaneously. It also over looks the ways in which many organized political movements accept and conform to certain demands of the dominant system and contribute to their repro duction or how they are susceptible to co optation by dominant forces.

Whether struggles against neoliberalism are better fought on the streets of Prague, Seattle, or Cochabamba with the TV cameras rolling or in the mundane spaces of everyday life is an ongoing debate. Clearly, then, there is a tension between organized resistance and everyday resistance, and between universal and postmodern forms of struggle, although theorists such as de Certeau believe there is something at least systematic about the everyday practices of resistance. This debate is important because it involves a consideration of the variable effectiveness of different political strategies and resistance movements. It also establishes a clear link between academia and pol itical activism through analyses of how researchers might work to unsettle forces of oppression such as capitalism, racism, or sexism.

It is possible that the most spectacular forms of re sistance (while important in their own right) might not be the most effective in terms of chipping away at capitalism or bringing progressive social change. While high profile forms of resistance such as Zapatista incursions into towns in Chiapas or the protests to block water privat ization in Bolivia have undoubtedly produced victories, their challenges to neoliberalism remain limited. Most indigenous people in southern Mexico are still poor and marginalized and there are still huge difficulties with the provision and distribution of water in Cochabamba. Moreover, research has shown that it is not uncommon for organized resistance movements to create their own forms of internal domination and hierarchies and to suppress or silence forms of internal dissent. Foucault, de Certeau, and others have, for example, expressed skepticism toward organized oppositional politics which often develop totalitarian or authoritarian tendencies. On the other hand, it is possible that if resistance is too subtle and goes unnoticed, large scale power relations might not change very much.

An additional problem with theorizing resistance is the overemphasis given to dominating or hegemonic power. A key question about resistance is whether mar ginalized or less powerful people are only able to re spond to the terms or strategies defined by the dominant. Mitch Rose believes that framing political practice in terms of resistance tends to treat the systems of power that are being opposed as preestablished or as something already operative in the world. Consequently, Rose urges us to avoid overemphasizing the hegemonic system and being complicit in its reproduction by focusing on the performative, the textual, and the symbolic and on the inherent instability of systems or enacted coherences.

Conclusions

Recent research on resistance has documented the ways in which resistance takes space in response to oppressive power relations such as the forces of corporate neoli beralism. Geographers have focused on the spaces and places in which resistant actions occur and from which resistant subjectivities emerge and how through such practices alternative knowledge and ideological repre sentations come to the fore. While there is ongoing debate around whether a global resistance movement that is unified in its opposition to neoliberalism is emerging, there is no doubt that the sites in which re sistance politics can be practiced has proliferated to include global networks of activists, NGOs, and civil society organizations. Changing economic and political conditions have led to new and innovative ways of doing politics. In many countries, civil society has become more diverse and robust and developments in ICTs have facilitated transnational collaboration and the formation of alliances. Key ideological assumptions that were previously dominant such as the notion that free trade is good for everyone or that the environment should be subordinate to the economy are now more unsettled and contested than ever before. The globalization of resistance as a counterpoint to the globalization of trade and commerce is rearticulating the meanings of devel opment in ways which enable alternative futures to be envisaged.

See also: Civil Society; Fair Trade; Feminism/Feminist Geography; Foucauldianism; Postcolonialism/ Postcolonial Geographies; Scale; Transnationalism.

Further Reading

Alvarez, S. E., Dagnino, E. and Escobar, A. (eds.) (1998). *Cultures of politics, politics of cultures: Re visioning Latin American social movements*. Boulder: Westview Press.
Amoore, L. (2005). *The global resistance reader*. London: Routledge.

Cresswell, T. (1996). *In place/out of place: Geography, ideology and transgression*. Minneapolis: University of Minnesota Press.
de Certeau, M. (1984). *The practice of everyday life*. Berkeley: University of California Press.
Featherstone, D. (2003). Spatialities of transnational resistance to neo liberal globalization: The maps of grievance of the inter continental caravan. *Transactions of the Institute of British Geographers* 28, 404 421.
Gibson Graham, J. K. (2006). *A postcapitalist politics*. Minneapolis: University of Minnesota Press.
Pile, S. and Keith, M. (eds.) (1997). *Geographies of resistance*. London: Routledge.
Power, M. (2003). *Rethinking development geographies*. London: Routledge.
Rose, M. (2002). The seductions of resistance: Power, politics, and a performative style of systems. *Environment and Planning D: Society and Space* 20, 383 400.
Routledge, P. (2003). Convergence space: Process geographies of grassroots globalization networks. *Transactions of the Institute of British Geographers* 28, 339 349.
Scott, J. C. (1985). *Weapons of the weak: Everyday forms of peasant resistance*. New Haven: Yale University Press.
Sharp, J. P., Routledge, P., Philo, C. and Paddison, R. (eds.) (2000). *Entanglements of power: Geographies of domination/resistance*. London: Routledge.

Relavant Websites

http://www.mstbrazil.org
Brazil's Landless Workers Movement, Movimento Sem Terra.
http://www.resist.org.uk
Globalize Resistance.
http://www.makepovertyhistory.org
Make Poverty History.
http://www.agp.org
Peoples' Global Action.
http://www.forumsocialmundial.br
World Social Forum.
http://www.ezln.org.mz
Zapatista Front for National Liberation.

Resource and Environmental Economics

H. Neo, National University of Singapore, Singapore

Glossary

Contingent Valuation A method used to determine the willingness to pay for services or environmental amenities.

Ecological Economics An approach in studying the relations between economics and ecology where the economy is seen as a subsystem of, and not separate from, the Earth's ecosystem.

Externalities An externality is a cost or benefit to one party (e.g., firm, household, or individual) that results from the activities of others.

Hedonic Pricing A valuation method used to derive values for environmental resource or environmental services that directly affect market prices. It is most frequently used to value environmental amenities that affect the price of residential properties.

Use Value The actual utility derived from using a commodity for its intended purpose.

Definitions of Resource and Environment

From the perspective of economics, resources are considered factors of production. Resources can be mined, extracted, or harvested as raw materials that go into sustaining our economy. Such natural resources are distinguished from other forms of human resource (like labor or technology). The environment in general (e.g., a nature reserve) can also be considered a form of resource although it is not typically viewed as a direct input toward economic reproduction. The field of resource and environmental economics aims to ascertain the optimal price of resource, as well as to mitigate, prevent, and mediate the externalities that come with resource use and economic production.

Natural resources can be divided into two categories: renewable and nonrenewable. Examples of renewable resources include solar energy, fisheries, and forests. Most energy and all minerals resources, however, are nonrenewable. These include natural gas, coal, and copper. Nonrenewable resources will clearly be depleted at some point in time, regardless of how slowly or little we use them – as long as we do not stop using them. Renewable resources can also be depleted if they are overexploited. The concepts of stock and flow can help illustrate this point.

For example, a nonrenewable resource, like oil, is considered a stock resource. Each time we consume oil, we are depleting its stock. On the other hand, flow resources cannot be depleted because its base (or stock) can generate flows indefinitely. This is analogous to drawing on the interest payment generated out of a lump sum of money in perpetuity, without ever touching the principal. An example of a flow resource would be solar energy. Yet, in reality, many resources which we would consider as flow resources are running the risk of being depleted as we start drawing from and depleting their stock. Thus flow resources, like fisheries, collapse swiftly once we start drawing from their stock.

The environment serves four functions in economic terms. First, it provides public goods, such as air to breathe in or, simply, pretty sceneries. Second, the environment is a provider of natural resources, as detailed above. Third, the environment is also seen as a receptacle of wastes from human activities and finally, the environment provides space for industrial and residential purposes, for agricultural production, and other infrastructural uses (like the building of roads).

Relationship between Resource Extraction and Economic Development

There cannot be economic development without the use of the environment and natural resources. Yet, the relationship between resource extraction and economic development is not a straightforward one. Many resource rich countries still rank as the poorest in the world. Two concepts have been used to describe this apparent anomaly: the 'Dutch disease' and the 'resource curse'. The Dutch disease hypothesized that a booming natural resource sector can lead to a decline in the development of other tradable sectors (most frequently the manufacturing sector, but may include the agricultural sector as well). This happens when a disproportionate amount of labor is pulled into the natural resource sector and shifts production away from the manufacturing sector. There is thus an over reliance of one sector which, to make matters worse, is itself price sensitive and subject to a volatile global market.

Similarly, the resource curse postulates that being resource rich is not necessarily a blessing. Various reasons have been offered to explain this paradox. First, many resource rich countries, particularly in the less developed world, were previous colonies that had been exploited for their natural resources, with minimum regard to the overall well being of their economy. This historical legacy tends to persist even after independence from their colonizers and severely compromized the abilities of these countries to diversify their economies. For example, Angola's economy is almost entirely made up of oil and diamond exports. Second, the over reliance on a single resource often leads to internal strife and conflict as different factions within an already fractious country fight to take control of the exploitation of the resource. Sierra Leone and Angola, blessed with rich diamond deposits, are excellent examples of countries torn by the exploitation of a natural resource. Moreover, as with the Dutch disease, the failure to diversify one's economy not only leads to wild fluctuations in export prices, it also makes long term economic planning difficult.

Both concepts point to the problems of the misuse and overexploitation of natural resources in economic de velopment. Indeed, the consequences of overexploitation can be extensive and global in nature. For example, one of the contributing factors to global warming is un inhibited deforestation in the past few decades while the pollution of fresh water resource from economic activ ities has always been a chronic problem. Environmental economics employs a variety of mainstream neoclassical economic tools and policies to ensure that the environ ment and natural resources are used in a market efficient manner.

Key Ideas of Environmental Economics

Types of Values

Given that the concept of pricing lies at the heart of environmental economics, valuation of environmental costs and benefits is imperative. In environmental eco nomics, three different types of values are distinguished, that is, use value, option value, and nonuse value. Use value refers to the direct use of a given resource for its intended purposes. For example, drawing water out of wells to quench one's thirst. Option value is the value that people place to reserve the right to use the resource in future. It is hence not a direct use value but a potential use value. Option value can change due to, among other things, technological innovation. For example, the option value of oil would drastically drop if a cheaper substitute for it can be found. The final type of value, the nonuse value, describes the value that people place on certain resource even if they would never get the chance to utilize it. For example, the loss of a nation's largest nature reserve is palpable even for those citizens that might not have visited (nor, indeed, have any inclination to do so). Nonuse value represents an intrinsic value that is quite different from use and option values; both of which are derived from the direct or potential use of the resource. Nonuse value is, moreover, shaped to a large extent by social norms and economic status.

Direct Methods of Valuation: Market Price and Contingent Valuation

The price of a resource is derived from its overall value, which is in turn made up of these three different types of values. Knowing the costs of different environmental options is necessary to make an informed choice on which option to take. For example, consider the clearing of a plot of forest for timber that will result in the extinction of a species of primates. Does one harvest the timber or save the primate? Decisions like this are arrived at through cost benefit analysis. Cost benefit analysis entails comparing the relative economic benefit of different courses of actions. In our hypothetical ex ample, if the economic benefit of harvesting the timber outweighs the economic benefit of allowing the primate to live, then the forest would be cleared. In order to make that comparison possible, a common 'currency' needs to be used. Almost always, a monetary price is assigned. Thus, the use value of harvesting the timber can be translated into monetary terms by estimating how much the timber would fetch in the market (i.e., its market price). However, pricing the nonuse value of the primate is less direct and involves contingent valuation. Contin gent valuation is a way of deriving values which cannot be directly obtained. In this case, it might involve asking respondents how much they are willing to pay in order to preserve the primate.

The major problem with contingent valuation is that respondents might give biased answers on the amount of money that they would pay for a particular environ mental option. Such bias can arise, for example, if the respondent provides a higher figure in the hope of in fluencing the decision making process or if the re spondent has insufficient knowledge or understanding to make an informed response. However, many studies have shown that the source of bias is not as large as one might think.

Indirect Methods of Valuation: Hedonic Price Approaches

In cases where it is not possible to directly extract value and price out of an environmental resource, hedonic price approaches are used. The principle behind hedonic price approaches is to infer environmental values from other economic activities that might not be considered environmental themselves. For example, the price which a person is willing to pay for good, unspoiled scenery could be derived from property values. This method is known as hedonic property values. Travel costs could also be used to estimate the value of a given environ mental resource (e.g., a bird sanctuary). Travel cost methods gather information about how far visitors traveled and how much they spent to get to the site to enjoy the resource.

Wage values is the third common hedonic price approach. For example, the value of a clean coastline could be indirectly inferred from the wage levels of people paid to clean up oil spills. The main problem with hedonic wage values is that while it attempts to isolate that component of a worker's wage which serves to compensate him for exposing to environmental risks (e.g., coal miners); deriving the environmental benefit of a safe mining working condition from the wage levels of miners is complex. Not least, miners in two difference countries would have accepted drastically different wages. This suggests that not everyone can afford the same level of environmental security and more broadly, that the value of life can be priced. As will be elaborated later, both these implications are contentious from the standpoint of critical geography.

Externalities and the Use of Economic Policy Instruments

An externality occurs when the cost of or benefit to one party (e.g., firm, household, or individual) is the result of the activities of others. For example, a chemical factory that dumps its waste into the river is imposing an externality (in the form of fresh water pollution) on people living downstream who depend on the water for survival. Negative externalities (i.e., externalities that are a cost) often develop when there is a lack of well defined property right. In economics, both positive and negative externalities are signs of market inefficiency. Environmental economics seek to resolve such externalities through several means.

Property Rights Approach

In our example above, there is no reason for the chemical factory to increase its costs by using alternative, more expensive, ways of disposing its wastes because, technically, nobody owns the river. Property rights approach argues that ensuring that environmental assets are owned and managed by relevant stakeholders is a necessary step toward the elimination of externalities. Specifically, property right sets out owner's rights, privileges, and terms for the use of the resource. More importantly, a rigorous property rights regime needs to fulfill three criteria, namely, exclusivity, transferability, and enforceability. To elaborate, the owner of the resource should be exclusively responsible to all the costs and benefits associated with maintaining the resource; they should be able to transfer their property rights freely to another party should they wish to; the property rights of the owners must be enforced and protected from involuntary seizures or encroachment.

Property rights are less formal and legalistic in places which practice common property regimes. Resource rights in this case may sometimes be protected by a set of informal rules, derived from tradition and cultural norms. However, in many Third World countries, what were initially common property resources (e.g., grazing land) become open access resources as increasing pressures deplete the resource, forcing owner users to flout the implicit rules that have traditionally governed its usage. Forced privatization of common pool resources would similarly lead to rapid deterioration of the management of the resource. Open access resources are essentially resources with no ownership. Drawing from open access resources, in the face of dwindling resource, will invariably lead to a tragedy of the commons where users exploit the resource as much as they want and as fast as they can.

However, the presence of common property and open access resources that have historically been managed prudently begs the normative question as to whether exclusive property rights is morally correct and whether it can be successfully implemented in different socioeconomic contexts.

Taxation and Subsidies

Besides conferring and upholding ownership rights to environmental resources, economic instruments like taxes and subsidies have often been used to eliminate externalities. For example, pollution tax could be imposed on firms which pollute beyond a certain level. Subsidies could also be given to firms so that they could install newer, more efficient technology or to encourage them to engage in more environmental R&D. One major problem with the imposition of pollution tax is that it sometimes merely internalizes the externalities. In other words, imposing pollution tax on a firm might merely lead to an increase in the price of its products. In this case, not only do the consumers pay more for the goods, but the environment continues to be negatively impacted. Another problem associated with taxing emissions is the cost effectiveness of measuring and monitoring emissions. For large firms, regulatory agencies often rely on the formers' self reporting, backed by occasional checks. However, the costs of monitoring smaller firms or households might not be cost effective.

Regulatory Approaches

Regulatory approaches can be divided into regulating standards and regulating quotas. For the former, this would entail the governing agencies to impose a predetermined standard of effluent, ambient, or technology in firms and the products they produce. An example would be legislation that mandates cleaner car engines. Quotas have also been implemented to control pollution and reduce externalities. In this case, a regulatory agency decides on a cap on the amount that a pollutant can be emitted. Firms then receive credits that will allow them to emit pollutants up to, but not beyond, the cap. If they manage to keep emission levels below the cap, they can use the surplus credits in a variety of ways. They could sell these credits to other firms who have emitted more pollutants than the quota allows for. In some quota sys tems, these firms can 'retire' these credits. The effect of this act would be that remaining credits become more expensive for firms who need to buy them to increase their emission allowance. Thus, the incentive to reduce emissions becomes greater the less free credits remain in the quota system.

Clearly, the success of this system of tradable quotas depends crucially on the effectiveness of governance of the emissions trade. It is also important to actually come up with a reasonable cap in the first place, although caps could be lowered subsequently. Others have argued that the use of tradable permits and quotas is a disincentive for firms to innovate new technology toward better emission standards. The Acid Rain Program of the 1990 Clean Air Act in USA (which sets quotas for the emission of sulfur dioxide) and the European Union Emission Trading Scheme (which deals with greenhouse gases) are two of the largest and most well known emissions trading schemes. The formation of the International Emissions Trading Association (IETA) in 1999 is an indication of the popularity of this form of regulation to control pol lution emission. IETA is led by industry leaders to de velop effective systems for the trading of greenhouse gas emissions by businesses, in accordance with the United National Convention on Climate Change.

Limits of Environmental Economics

Touted as a solution and reaction against deteriorating environment and depleting natural resources, environ mental economics has been criticized at various levels. This is apart from the various methodological challenges, detailed earlier, that might plague the key concepts of pricing and valuation in environmental economics. In pointing out the limits of environmental economics, ecological economics argue that the market, which lies at the core of environmental economics, is but one possible mechanism for the allocation of resources.

The Critique from Ecological Economics

Ecological economics recognizes the importance of effi cient allocation of resources but seeks a much deeper understanding of the relationship between economic development and resource exploitation. It situates the fundamental problem of environment economy not in market failures, but in humans' inadequate understanding of their role, impacts, and responsibilities within the larger ecological system. Consequently, the goals of ecological economics are more far reaching than those of environmental economics. While, as with the latter, valuation of resources is still an important goal, eco logical economics is also interested in issues of sus tainability; ecological–economic system accounting; ecological–economic modeling at various spatial scales; and exploring innovative tools for environmental man agement. In other words, ecological economics has a much more transformative agenda (in terms of modifying our relationship with the environment) than environ mental economics. In fact, the latter is often seen as more concerned with maintaining the capitalist status quo. To illustrate one difference between the two, one of the specific goals of ecological economics is to maintain the total natural resource stock at, or above, the current level; however, there is no compulsion for environmental eco nomics to consider such ecological limits as long as ex ternalities are eliminated or minimized and the market allows for a fair exchange between buyers and sellers.

The core of the ecological economics critique is that it is wrong to treat the environment as merely a factor of production that can be analyzed in isolation to, and ex tracted out of, the whole ecological system.

The Critique from Marxist and Critical Geographies

Taking the ecological economic critique of treating re source in isolation further is the body of work on the commodification of nature, developed by Marxist geog raphers. The Marxist notions of 'privatization' and 'valuation' can be used to highlight the limits of en vironmental economics. Although the notion of privat ization is central to environmental economics' property rights approach for the resolving of externalities, con temporary Marxist analyses of privatization of erstwhile common pool and open access resources argue that pri vatized ownership of resources often leads to situations where resources are controlled and used for the benefit of few, to the detriment of the public. The privatization of freshwater resources in urban cities is representative of such works.

The Marxist critique of the notion of value is long standing. Among other things, it argues that by repre senting different types of values (and different types of commodities) as money, true value is obscured. It is precisely the reduction of environmental degradation as money that leads to situations where externalities are internalized and passed on as increased costs to con sumers. It follows that contingent valuation methods are opposed on ontological grounds by many Marxist and critical geographers and, by extension, one needs to more critically engage with the ethical bases of contingent valuation.

At a more general level, environmental economics is seen as merely addressing the symptoms of environmental pollution and degradation without looking into the root of the problem. In particular, critical geography, with its focus on social inequalities, justice, and emancipatory politics across scales and spaces, has critiqued the amoral underpinnings of environmental economics. Many critical studies on the geographies of natural resource have detailed how global firms are increasingly drawn to less explored and less developed regions of the world in search of more resources to exploit. Due to their low integration to the global capitalist systems, such remote places also have limited knowledge and the lack of power to demand and enforce adequate environmental standards to safeguard their environment and resources. The mining and oil industries are oft cited examples of such global predatory tactics.

Conclusion

The critique of environmental economics is not meant to dismiss its usefulness in helping one make the correct environmental decisions nor to discount its role in addressing the problems of wanton environmental exploitation and pollution. Rather, it is to point out its inadequacies especially in dealing with issues of justice and equity in the use and misuse of resources and the environment. Moreover, environmental economics in its present form and scope, framed in the logics of neoclassical economics, can offer little toward an alternative vision of a more environmentally enlightened world.

See also: Environmental Policy; Environmental Regulation; Natural Resources; Resource Industries; Sustainable Development.

Further Reading

Auty, R. M. (1993). *Sustaining Development in Mineral Economies: The Resource Curse Thesis*. London: Routledge.
Bakker, K. (2000). Privatizing water, producing scarcity. *Economic Geography* 76, 4 27.
Blomley, N. (2006). Uncritical critical geography? *Progress in Human Geography* 30, 87 94.
Castree, N. (2003). Commodifying what nature? *Progress in Human Geography* 27, 273 297.
Constanza, R. (ed.) (1991). *Ecological Economics: The Science and Management of Sustainability*. New York: Columbia University Press.
Daly, H. E. and Farley, J. (2004). *Ecological Economics: Principles and Applications*. London: Island Press.
Emel, J. and Bridge, G. (1995). The earth as input: Natural resources. In Johnston, R. J., Taylor, P. J. & Watts, M. J. (eds.) *Geographies of Global Change*. Oxford: Blackwell Publishers.
Markandya, A. and Richardson, J. (eds.) (1992). *The Earthscan Reader in Environmental Economics*. London: Earthscan.
Siebert, H. (1995). *Economics of the Environment* (4th edn.). Berlin: Springer Verlag.
Tietenberg, T. H. (2005). *Environmental and Natural Resource Economics* (7th edn.). Boston: Addison Wesley.
Tietenberg, T. H. (2006). *Environmental Economics and Policy* (5th edn.). Boston: Addison Wesley.
Van Wijnbergen, S. (1984). The 'Dutch Disease': A disease after all? *The Economic Journal* 94, 41 55.

Relevant Websites

http://www.ecosystemvaluation.org
Ecosystem Valuation.
http://www.ieta.org
International Emissions Trading Association.

Resource Industries

R. Hayter, Simon Fraser University, Byrnaby, BC, Canada

Glossary

Corporate Strategy The allocation of resources and investments to meet the long-run objectives of corporations, especially in relation to rivals and markets.

Cost Minimization The view that location of economic activities are primarily motivated to minimize costs.

Entrepreneurial Model of Development The organization of economic activities around individual, relatively small firms with little or no power over markets.

Flexible Specialization The organization of economic activities around interacting populations of entrepreneurial firms.

Geopolitics The influence of government policies, especially of powerful states, over the spatial organization of territories and of economic activities.

Green Paradigm The idea that economic activities must be collectively organized in environmentally sustainable ways.

Industrial Resource Values The benefits to individuals (producers and consumers) to be gained from exploitation of nature to produce commodities to be sold in markets so long as prices cover costs and provide a profit.

Informal Resource Exploitation The exploitation of resources by individuals who have not been granted formal rights to access the resource. These activities are often considered illegal by states even though they may be vital for local survival of local populations.

Nonindustrial Resource Values The social benefits provided by nature to sustain life in all forms and to meet social requirements related to biodiversity, spirituality, and esthetics.

Nonrenewable Resources Resources whose exploitation means depletion.

Plantation Model The organization of economic activities around large-scale, hierarchical organizations that seek to exploit internal economies of scale. While the term 'plantation' originated specifically in the agricultural sector, it is used as a generic name for the local sites of mass production by large organizations.

Primary Manufacturing Manufacturing activities whose principal material inputs are natural resources or minimally processed natural resources.

Primary Sector Economic activities engaged in the direct exploitation of natural resources.

Property Rights Legal claims to income streams that imply exclusivity, security, and transferability. The most important categories of property rights are private property, common property, government controlled property, and open access property.

Renewable Resources Resources whose exploitation need not mean depletion, even if in practice depletion occurs.

Resource Curse The idea that nations and regions commonly fail to translate resource abundance and exports into sustained economic growth.

Resource Cycles The idea that resource exploitation features a boom and bust cycle, the latter defined the end of resource cycles.

Resource Wars A metaphor used to represent the many deep-seated conflicts that occur over how resources are to be controlled and exploited.

This chapter examines the location dynamics of resource industries in terms of their spatial patterns and implications for local development. This mandate is more problematic than often portrayed in economic geography. The spatial evolution of resource industries is shaped, not just by cost minimizing imperatives as often supposed, but by the nature of resource rights, industrial organization, development aspirations, and geopolitics. Moreover, resource industries are distinguished by their embeddedness in nature and by patterns of exploitation (resource cycles) that deplete resources and threaten both industrial competitiveness and nature's 'nonindustrial values' in the form of environmental benefits, esthetics, spirituality, and ways of life. How societies value and use resources varies in space and time and the values of different interest groups are often hard to reconcile. Indeed, resources are vital to industrialization and civilization but their exploitation is often metaphorically associated with the 'resource curse' that seemingly traps resource rich countries and regions in poor economic performance and 'resource wars' that reflect intense conflicts of different environmental, cultural, political, and industrial interests over resource exports regarding their income streams and impacts on nonindustrial values.

This article unravels the location dynamics of resource industries from a political economy or institutional perspective that privileges how resource exploitation is controlled and organized, notably (but not exclusively) by powerful multinational corporations (MNCs), and influenced by interacting economic, environmental, cultural, and political processes. Resource industries are extraordinarily varied and conventionally comprise primary activities (e.g., agriculture, fishing, mining, and logging) and primary manufacturing (e.g., sawmills, oil refineries, and sugar refineries) that directly exploit 'natural' re sources, in raw or semiprocessed forms, to create com modities for society's use. The resources utilized by resource industries may be classified as renewable or nonrenewable, based on whether or not depletion is in herent to use, although, in practice, renewable resources are often 'mined' without giving thought to renewal. The article proceeds by first noting alternative forms of re source organization, their connections to resource rights, and implications for local development. Second, the lo cation dynamics of resource industries is summarized in terms of cost minimizing theory and then assessed ac cording to the implications of corporate strategy and geopolitics. Third, resource industries are interpreted ac cording to the resource cycle thesis to appreciate their boom and bust characteristics and the vulnerability of resource peripheries. Services, notably tourism and rec reation that directly rely on the provision of natural re sources, are excluded from this discussion.

Resource Rights and Organization

Resource exploitation is rooted in the assignment of property rights that, in general, are claims to benefits or income streams that imply exclusivity, security, and transferability. Exclusivity assigns sole rights of use to (resource) property, security is provided by law or social sanction, and transferability means that owners can sell or trade their property rights. Resource rights are held pri vately by individuals or corporations, in common by specified groups of people, and by governments; 'open access' resources lack property rights. To crudely gener alize, historically common property resources were widespread, different forms of aboriginal and peasant cultures were rooted in tradition, custom, and the as signment of rights by elites. Industrialization and 'The Great Transformation' toward market economies, how ever, privileged (legally enshrined) private resource property rights, often imposed at the expense of tradition and common property arrangements by, for example, the enclosure movement in Europe and settler colonies in the New World. The deforestation of the Amazon illustrates the continuation of this process as the government assigns resource rights to peasant farmers, corporate ranchers, and MNCs while often disregarding aboriginal rights.

Around the globe, the 'plantation model' and the 'entrepreneurial model' are two polar institutional types that summarize the organizational variation of resource industries that serve markets (**Figure 1**). Each type is associated with fundamentally different routines, con ventions, organizational structures, and strategies, and each can be further differentiated. Although con ventionally associated with the large scale crop mono culture, the plantation model in the present context represents 'all' forms of large scale (mineral and biotic) resource production that seek economies of scale and size and are controlled by hierarchical organizations. Thus plantations refer to the subsidiaries of MNCs, large do mestic firms, agencies of government, or joint ventures among these organizations. Entrepreneurial models em phasize the fragmentation of production among many locally based firms, family controlled enterprises, and related operations that collectively realize external economies of scale.

Private property rights are typically assigned to the resources themselves and utilization depends on the decision making discretion of owners, limited to some extent by government policies. There are significant ex amples – Canadian forests are an illustration – in which

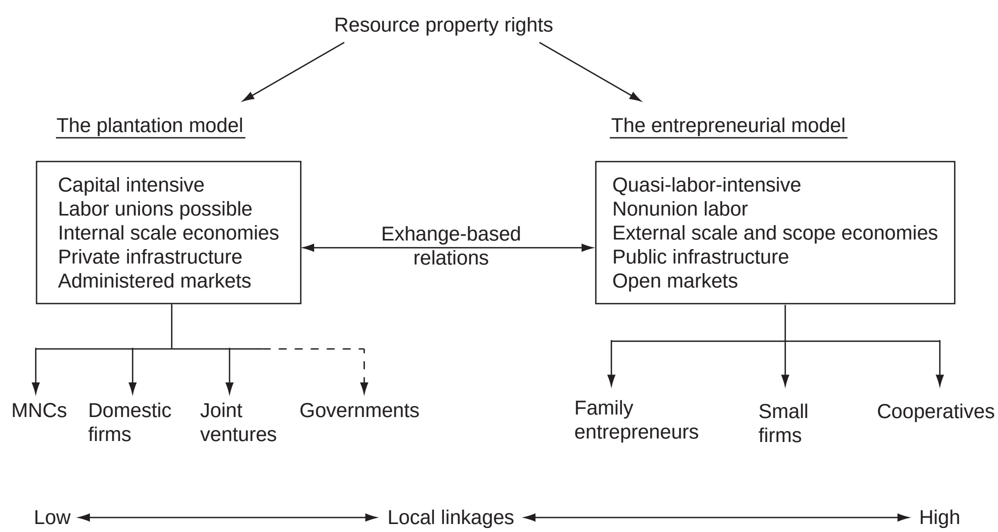

Figure 1 The organization of industrialized resource exploitation: Basic institutional types.

governments retain ownership of resources and privatize their utilization via 'resource contracts' that typically specify time horizons, renewal possibilities, rent (tax) payments and minimum or maximum levels of pro duction, and potentially other aspects as well. In some cases, especially in relation to water and energy, even among rich market economies, governments opt to fully control resource industries. In developing countries, re source exploitation also occurs in myriad forms without proper assignment of property rights, and although typically considered illegal by national authorities, such activity is not necessarily corrupt or 'antisocial'. Indeed, 'artisanal', 'semiindustrial', and 'informal' small scale mining is a widespread strategy among desperately poor people seeking to eke out a living. In the coal mining region of Raniganj in India, for example, Lahiri Dutt estimates that there are around 200 000 people involved in informal coal mining in which individuals collect coal from the tailings of wastes of the large mines and trans port their coal, by foot or bicycle, to markets such as households, shops, etc.

International Trade

Large scale resource production of all kinds is organized to meet domestic needs within highly localized pro duction chains. With the expansion of civilization and urbanization, however, resources always have been sig nificant components of interregional and international trade; motives for imperialism, colonialism, and geo politics; as well as for economic development. Resources remain a big component of international trade. Thus according to the World Trade Organization (WTO), total trade in 'visible' goods (all raw materials and manu factured goods) amounted to almost $6 trillion in 2001 (while 'nonvisible' (service) trade was worth $1.5 trillion). Agricultural and mining products, the WTO's definition of the resource sector, accounted for 22.3% of the value of visible trade, and the contributions of primary manufacturing such as forest products and iron and steel would further increase this share. Oil dominates resource exports; in 2002 its export values were four and six times higher than natural gas and aluminum, the second and third export categories.

Generally speaking, rich countries are the world's major resource consumers and importers, although re cently Asian industrialization has led to significant shifts in global trade in many resources notably toward China and increasingly India. Within these rapidly changing trade dynamics, poor countries are the most export dependent on resources. However, the extent to which poor countries can exploit the resource strengths is limited because rich countries produce and consume vast amounts of resources domestically and are often major resource exporters themselves. The US, for example, in 2002 was among the top three global exporters of wheat, rice, fruit and nuts, cotton, oil seeds, wood, fish and coal, and an important supplier in many others. In contrast, in poverty stricken Angola and Gambia, just three com modities provide almost 100% and 79% of national ex ports, and domestic markets are small. The resource exports of poor countries are extremely dependent on rich markets that enjoy much discretion regarding sour ces of supply. The market power of rich countries is further reflected in organizational arrangements (**Table 1**).

Implications for Local Development

The plantation and entrepreneurial models have differ ent implications for local development (**Figure 1**). Large scale resource exploitation by big organizations, espe cially MNCs, realizes internal economies of scale at both the level of individual operations (mines, wells, and agricultural plantations) and at the firm level, for ex ample, in the form of (internal or administered) access to global market networks. Such large scale operations typically pay relatively high wages and rely on dedicated,

Table 1 Share of three leading commodities in total exports of selected countries and country groups (in %)

Country or group	*1965*	*1980*	*1995*	*2000*
Developing countries	11.9	61.2	17.7	21.8
Brazil	56.7	27.6	14.1	12.5
China	5.7	28.8	6.8	5.2
India	23.9	14.5	10.4	6.0
Least developed countries	29.1	42.0	34.1	45.7
Angola (fuels, diamonds, fishery commodities*)	56.4	87.1	94.3	99.8
Burundi (coffee, tea and mate, hides and skins*)	97.5	94.3	87.8	73.6
Gambia (sugar, fishery commodities, groundnuts*)	75.2	89.2	70.0	79.1
Mauritania (iron ore and concentrates, live animals, fishery commodities*)	95.2	89.9	58.5	59.3
Somalia (live animals, ananas, fishery commodities*)	65.1	89.9	63.8	56.8
Developed countries	4.9	9.5	4.7	6.0

* Three leading commodities in 2000.

Source: UNCTAD (2005). UNCTAD Handbook of statistics 2005. New York: United Nations Publications.

often subsidized infrastructure, notably transportation links, housing, and community amenities. The plantation model provides local economies with major investment, established expertise, access to markets, export income, and relatively fast means of growth. However, plantation models increase external and foreign control of local economies, rendering them vulnerable on decisions made at distant head offices, while parts, equipment, services, and research and development (R&D) are often imported, including from affiliated sources, thus reducing the possibilities, for these activities to be developed locally. Dedicated infrastructure also implies exclusive use. Payments by plantations for head office services, R&D and parts, along with the profits repatriated to pay for the original investment and provide rates of return, are drawn out of the local economy and may or may not be reinvested locally. Moreover, a basic rationale for foreign investment in the resource sector is to extract resources for processing in home countries. For local economies, MNCs incorporate resource industries within the most remote locations within global (space) networks but in ways that can limit diversification possibilities.

In contrast, the entrepreneurial model is based on local ownership, and the collective impacts of many small scale operations are more conducive to the generation of external economies of scale, for example, through the stimulation of local services and machinery suppliers and creation of local marketing, R&D, and lobbying associations while social and economic infrastructure that is created in the form of roads, railways, and housing is typically a public good, accessible to all. That is, the entrepreneurial model is strongly place centered; local entrepreneurs are more likely to equate self and local interests and be more willing to add value, pursue economies of scope, and reinvest profits locally. However, entrepreneurial models cannot control markets and prices to the same extent as MNCs, they are vulnerable to competition and the policies of powerful and distant buyers (markets) and suppliers, and they are not necessarily viable alternatives to large scale resource production that exploits internal economies of scale. Traditional models of resource exploitation are autarkic and virtually all development is local.

In general, the plantation model has increasingly dominated resource industries. Some of the biggest MNCs in the world, such as Shell and Exxon, are resource based and many resource industries, including oil, are dominated by vertically integrated organizations that have considerable control over the extraction, primary processing, and marketing of resource 'production chains'. Many entrepreneurially dominated activities have experienced increased corporate control including in farming and fishing, although reversals of the plantation to entrepreneurial model have occurred. Controversy over the development implications of different forms of resource industry organization is widespread. Thus in the 1960s and 1970s several countries nationalized the resource activities of MNCs, changing the nature of the plantation model from foreign, private to local, public ownership. On the other hand, as in the case of water supply in the UK, resources long considered to be part of the public sector have been privatized. In these contexts, the case for public control emphasizes that natural resources are 'public goods', owned by all citizens who should have rights to the income streams generated by resources or equal access to their consumption. In contrast, privatization proponents emphasize greater efficiencies in production and innovation.

In practice, resource based production chains are organizationally and spatially varied and plantation and entrepreneurial models coexist with one another, as rivals and in complimentary supplier relationships. In the forest sector of the Pacific Northwest in Canada and the US, for example, big companies rely extensively on subcontracted logging, partly to access the greater efficiencies and lower (nonunion) wages of smaller operations and partly to devolve the risks and costs of maintaining highly seasonal (weather affected) operations.

Oil companies extract, transport, process, and market oil, connecting directly with final consumers in franchised gas stations. Japan's sogo shosha transport coal from large scale (plantation) mines from around the world to various plantation consumers in Japan, such as iron and steel mills and energy companies. The same sogo shosha seek out timber supplies from around the world, provided by both big and small companies, and delivered to complex distribution chains within Japan many of which terminate in purchases by individual carpenters building Japanese style homes. Coffee from Ethiopia is produced by small scale growers, sometimes sold cooperatively and sometimes by local merchants who deliver to local processing plants, warehouses, and local auctions from where the coffee is sent to the roasteries owned by giants, such as Nestlé, and smaller firms, such as Costa Coffee, where the beans are grounded or passed to the various retail outlets.

In these and other resource based production chains, location choices for individual activities reflect the interaction of cost minimizing principles, corporate strategies, and geopolitical considerations.

Resource Industries in Space

Natural endowments exert profound effects on the location of resource industries. In this regard, conventional neoclassical location and land use theories were conceived in relation to the entrepeneurial model, and with strong assumptions about competitively constrained economic rationality rooted in cost minimizing or profit maximizing behavior. Nevertheless, neoclassical location theories demonstrate that space matters to the viability of resource industries while, more generally, the neoclassical interpretation of resources as commodities has cast a giant shadow.

A Cost-Minimizing Starting Point

Neoclassical location theories distinguish between 'point' resources such as minerals that are found in particular places and 'diffuse' resources such as forests and agricultural land that are spread over large areas. In the case of a mineral ('point') resource, for example, mining activity occurs if the revenues received from consumers cover processing costs and transportation costs to markets, plus at least a 'normal' profit margin, defined as the minimum profit required to induce production. While costs of transportation tend to vary systematically around markets, shaped by the nature of transportation networks, the costs of exploitation are more irregular, dependent primarily upon the size and nature of the resource and its ease of working. In this model, spatial pattern of mining reflects underlying cost structures and, for example, remote mines can operate when nearby mines are closed by substituting lower per capita processing costs for higher transportation costs. Indeed, the dispersal of mining operations to the distant corners of the globe is closely associated with increasing operating size that realize economies of scale, as well as investments in specialized transportation systems that reduce freight rates. Alternatively put, the dispersal of mining and increasing scale of operations has been mutually stimulating. Individual categories of resources are also internally differentiated, associated with distinct market, technological, and location conditions.

For diffuse resources, economic viability similarly depends on revenues covering the costs of production and transportation costs to the market. Direct competition among different uses of land, however, adds a complicating dimension. Thus in von Thünen's theory of agricultural location, different farming systems and specific crop and farming types compete with each other to use land and serve markets. In this model, land use is controlled by location rent, that is, the difference between revenues generated at the market less transportation and production costs. The model predicts that land uses that generate the most location rent will be chosen and that in situations in which transportation costs are 'important', different types of land uses will vary systematically, or in zones with distance from the market. In particular, 'intensive' land uses, defined as land uses that apply relatively high levels of factor inputs (capital, machinery, labor, fertilizer, etc.), that generate high yields and revenues per unit of land will occur close to markets to reduce the transportation costs per unit of land (associated with higher yielding uses). The converse applies to more remote locations where extensive land uses prevail. Over time, in this model, land uses change if location rent changes. Thus, the model predicts that as demand increases (decreases) and transportation costs decrease (increase), each land use zone and the outer margins of production are spatially extended (contracted) as higher rent replace lower rent generating activities. However, as transportation costs have declined significantly, the implications of resource endowment for productivity have become a more important control over location rent, and spatial zoning a less evident part of the landscape.

As resources are processed, whether point or diffuse, the neoclassical location calculus takes into account the procurement (as well as distribution) costs of assembling resources at factory sites and variations in processing costs associated with supplies of labor, energy, agglomeration economies, and government policies. In this view, alternative locations offer different cost structures and factories should choose the one that minimizes overall costs. In this regard, the search for new, low cost locations for resource activity is constrained by the advantages enjoyed by established factories that utilize existing (and paid for) equipment, infrastructure, and labor pools. Even if no longer least cost, 'old' resource locations can survive so long as operating revenues exceed operating costs.

Corporate Strategies and Geopolitics

Costs (prices) are vital to understanding the location of resource industries. Yet, in the plantation model, costs are not simply 'given' (exogenous) datum to which firms passively respond. Rather, the prices corporations pay for resource rights, and associated conditions of resource exploitation, are subject to bargaining among various stakeholders, especially those with ownership rights. In resource bargains, large corporations have considerable power that is associated with their accumulated expertise, size, and (actual and potential) global networks, rooted in alternative production sites and marketing connections that integrate and transcend individual political jurisdictions. For MNCs, decisions regarding specific resource sites are shaped by the imperatives of long run corporate strategies and existing location structures. The strategies of MNCs are further profoundly shaped by geopolitics or government policies that seek to control and influence geographic space to serve national, economic, and political interests, including issues of national security.

In the context of resource bargaining, issues related to 'security and uncertainty' profoundly shape location dynamics. As Odell observed, with respect to the (plantation dominated) oil industry, the location evolution of resource exploitation is more diversified and extensive than anticipated by cost minimization principles alone. Thus MNCs compete for access to resources in 'over lapping location' strategies, driven by the fear of rivals controlling resources that are low cost and/or high quality. Resource consuming countries have sought to control and secure access to vital resource supplies around the world. Indeed, a major stimulus to European colonization in the nineteenth century was to gain con trol over 'their own' resources, a policy that resonates today, for example, with respect to Japan's 'multiple sourcing' strategies, organized globally by the sogo shosha for a wide range of resources. In the US, UK, Europe, and Japan, as domestic coal and iron ore re sources became depleted, since the 1960s new invest ments in iron and steel emphasized coastal locations that provided 'location flexibility' to alternative sources around the world, facilitated by cheap water transporta tion. Such flexibility reduces the risks of dependence on particular resource sites, maintains value added in rich countries, and provides close access to markets.

The diversification of supply sources is further re inforced by competition among governments in resource peripheries that wish to exploit resources as a source of economic growth and diversification. These desires are complicated by the contrary impulses to be competitive by offering resources at low cost and to extract value in the form of higher taxes and increased primary processing (and machinery and other input supplies). Moreover, the desire to negotiate resource bargains that add value places resource peripheries in tension with resource consuming countries who seek stable, low cost resources and value added in the form of profits and jobs. Indeed, the location of value added has been an insistent source of tension within resource based production chains.

With respect to primary manufacturing, government and corporate policies frequently counter the advice of neoclassical theory to locate close to resource supplies whenever processing reduces weight, bulk, and/or per ishability of resource inputs. Tariff policies have major relevance in this regard. Historically, the rich countries of North America, Europe, and Japan pursued policies of significantly reduced tariffs on imported raw materials that they required while maintaining high tariffs on manufactures that processed these materials. The transfer pricing policies of integrated MNCs have often re inforced such tariff structures. A MNC that processed bauxite (from Australia) in Iceland, for example, reported artificially low costs on alumina to reduce the taxes it paid and to minimize costs prior to final processing stages in Europe. In general, resource MNCs seek to extract resources and transfer them to home (and consuming) countries for further processing. Tariff policies have strongly reinforced this behavior as regards market ori ented activities, in which processing adds bulk, weight, perishability, and/or is design intensive, but have been more ambiguous with respect to input oriented resource based activities.

To create resource bargains in their favor, resource peripheries have raised taxation and royalty levels, re stricted the export of raw materials by export taxes or outright bans, formed cartels with other producers to restrict supply and increase prices and nationalized re source industries, in whole or in part, to stimulate do mestic corporate champions. For example, resource cartels such as the Organization of Petroleum Exporting Countries (OPEC) seek to control prices by controlling supply; Russian policy has recently shifted to reduce the role of foreign MNCs in its oil fields; and many forest exporting nations, including Canada, limit raw log ex ports. From a corporate perspective, the obsolescing bargain hypothesis proposes that after the initial invest ments in resource exploitation, bargaining advantages shift away from the MNC toward local governments. In this view, initially, MNCs have location options and monopolize resources and know how. Subsequently, however, the bargaining power of host governments is increased by the fixity of investments, increases in local resource exploiting know how, and policy options in cluding nationalization. In practice, the realization of the obsolescing bargain – from the perspective of resource peripheries a perceived improved bargain – is restrained by competition among resource producers, the global reach and expertise of MNCs, and the hard realities of the resource cycle.

Resource Cycles and Vulnerabilities

According to the resource cycle, exploitation follows a life cycle curve based on discovery, slow initial growth, rapid growth, and a plateau effect as the region matures, followed by depletion and industrial decline. During the course of resource cycles, 'normal' business cycles of growth and recession are ultimately represented by boom and bust, the latter encouraging resource exploitation to move on to the next region. For nonrenewable resources, continued depletion inevitably leads to collapse, even if cycles are modified by technological innovations that expand the size of the resource. Similarly, crises of de forestation, overfishing, and dustbowls indicate failures to regenerate renewable resources that intimate powerful self destructive tendencies.

Indeed, for all kinds of market driven resource ex ploitation there are deep seated tendencies to extract profits as fast as possible while minimizing costs, such as renewing stocks with long term, uncertain returns. Moreover, the best, most accessible resources are typically extracted or harvested first so that industry's costs of ex ploitation typically increase. In contrast, resource prices are often dampened by investments encouraged by regional and corporate rivalries that result in oversupply and increased competition, and technological innovation that creates substitution possibilities among resources (and from synthetics). As a result, in the context of growing fears and realities of depletion, resource industries are often caught in long term cost price squeeze. This squeeze is tempered by the advantages of established capital and human resources while the opening of new resource spaces costs requires major investments in pro ductive capital, infrastructure, and the hiring, training, and housing of workers. Resource firms in older resource re gions may also be able to produce more valuable products, think up ways to reduce costs, and/or lobby their gov ernments for subsidies or protection from cheaper im ports. In fact, government subsidies in rich countries in support of domestic resource industries are a significant, controversial feature of international trade conflicts.

In the case of agriculture, the US, the EU, and Japan have massive domestic industries that they protect from foreign competition, especially from poor countries, by substantial subsidies. In the US, a large domestic timber industry has effectively lobbied the government to restrict imports from Canada. Although leading advocates of free trade, these countries justify their protectionist behavior by linking free trade policies to interpretations of 'fair' trade. Thus agricultural exports from poor countries are deemed unfair because wages are low and Canadian lumber exports to the US are deemed unfair because taxation of lumber on public land is similarly deemed too low. However, such definitions of fairness are only possible because of the greater market power of consuming countries. The recent (2006) collapse of global free trade talks organized by the WTO stemmed from a failure of rich countries to grant free market access for the agri cultural produce of poor countries (who in turn refused to remove some barriers to foreign investment). Paradoxic ally, resource industries are variously on the leading edge of free trade and the center of major trade disputes.

Resource Peripheries

Resource peripheries are regions where resource industries are significant dimensions of economic activity, especially in relation to exports; the vulnerabilities of resource in dustries during short term business cycles and especially over longer resource cycles are shared and compounded in resource peripheries. Especially in the context of the plantation model, once closely overlapping developing as pirations between community (or region) and corporation are increasingly likely to diverge as resource cycles mature. Ghost towns that symbolize the end of resource cycles need not imply ghost corporations. Resource based MNCs have location options for investment and this location flexibility is itself a source of instability in peripheries, for example, with respect to labor bargaining or negotiations over taxes, the cost of local services and subsidies. Tech nological changes that reduce jobs (and costs) increase corporate profitability but have more problematic impacts on communities. Moreover, resource communities, dis persed and specialized according to where resources hap pen to be, are often (increasingly) remote from markets with limited opportunities to diversify. These restrictions themselves are built into resource cycles. During boom times, when wages, profits, employment, and optimism for resource longevity are high, there is no local incentive, nor any mandate by MNCs or unionized workforces, to con template diversification. During busts, remoteness, foreign control, specialized infrastructure and worker skills, and competition with other places restrict the ability to diversify.

In Australia and Canada, government resource plan ning policy has been reformulated by the apparent re lentless imperatives of resource cycles. Thus the opening up of remote resource regions during the 'Fordist' boom years from the 1950s to the 1970s was celebrated by the planned frontier (resource) town offering a full range of social amenities to workers and their families. This concept is now replaced by 'fly in' workforces, brought in to exploit the resource as needed on a planned temporary basis, with no need for fixed community investments nor worries for future obsolescence.

The logic of resource cycles resonates closely with Auty's thesis that resource rich developing countries ex perience a 'resource curse' and by the parallel notions that resource peripheries in Canada and the US are 'trapped' within or 'addicted' to resource cycles. No matter how powerful, resource cycle dynamics that culminate in de pletion and trapped, addicted, and cursed peripheries are nevertheless tendencies not laws that need to be placed in distinct commodity, historical, and geographical contexts and related contingencies. At any particular period and place, resource cycles and curses (or traps) need to be understood in relation to specific local–global dynamics that include the trade regulations by foreign governments and the assignment of resource rights by domestic governments, and how these regulations and rights are organized by MNCs, local entrepreneurs, and others within production chains linking resources to markets. Dauvergne, for example, interprets the deforestation of Southeast Asia in terms of the cumulative interdependent effect of MNCs with local, traditional 'patron client' pol itics, based on mutual favors and asymmetric power re lations, driven by Japanese demand for wood.

Remapping Resources Toward Nonindustrial Values

In recent decades, resource industries and peripheries have been embroiled in 'resource wars', sometimes literally violent, that are remarkably prevalent around the world.

A striking characteristic of many of these conflicts is op position to actual or planned resource exports by MNCs from environmental and indigenous groups who desire to remap resources according to their priorities that re spectively emphasize ecological values and cultural values rooted in local control and identity. This proliferation of resource conflicts is itself an expression of globalization rooted in the increased scale and intensity of resource cycles that by the 1960s had stimulated concerns for global environmental degradation, an increased public policy commitment to environmental sustainability and the birth of modern forms of environmental nongovernment or ganizations (ENGOs). As the sites of resource exploitation, resource industries have been the principal target of ENGO opposition in a plethora of 'antifishing', antimin ing', and 'antilogging campaigns'. ENGOs are themselves multinationals, possessing high levels of spatial mobility and able to orchestrate global campaigns against resource developments around the globe. Moreover, in opposing industry, ENGOs have frequently been associated with indigenous rights movements, overwhelmingly located in resource peripheries, although the remapping priorities of the latter are not necessarily the same as the former.

For resource industries and peripheries, the concerns for environmental and aboriginal justice are an increasingly important component of resource bargains and how re sources are to be exploited. In some cases, major projects, including ones sponsored by the World Bank as 'develop ment' priorities, have been stopped. The longer term im plications of environmentalism (and aboriginalism) for resource peripheries are to emphasize a broader inter pretation of the social benefits of resources, especially with respect to their nonindustrial values and to give more recognition to alternatives to the mass production option that offer more sustainable forms of development. The renaming of British Columbia's central coast as The Great Bear Rain Forest, based on a re allocation of forest re sources from timber leases that required large scale log ging, to a mix of conservation areas, land under aboriginal control and for other community uses, illustrates this remapping.

In general, the remapping of resource peripheries re mains highly contested and varied but there is an urgent effort to protect environmental values with varying con sequences for resource industries. Thus conservation parks, which exclude resource industries entirely or per mit them under highly regulated circumstances, are an increasingly important symbol of the remapping of re sources away from property rights regimes that assume commodification. The very creation of such parks reflects the belief that market forces cannot protect the environ ment and associated public benefits. A friendly amend ment to the pure conservation park is biotechnology prospecting that involves a search for new medicines from tree sources. By no means least, remapping initiatives have encouraged greater consideration of the entrepreneurial models of (localized) development. Thus proposals for community forests and farmers' markets illustrate highly localized forms of development that anticipate reliance on local suppliers and consumers. Lahiri Dutt's plea for the peasant 'scavenging' of coal tailings in India to be assigned legalized resource rights as part of a 'moral economy' is a related initiative. There are also illustrations of flexibly specialized developments in the resource sector, for ex ample, in the Canadian and Japanese forest industries, in which populations of locally based small firms exploit economies of scale and scope to serve local and distant markets. If these initiatives are problematic, in some cases experimental, they also probe the limits of the mass pro duction model.

From a long term perspective, the scale and intensity of resource conflicts around the globe raise the question of whether society is at a turning point in thinking about the balance between industrial versus the nonindustrial values of resources. Admittedly, global demands for re sources is encouraging the continued search for new resources in increasingly isolated locations, beneath the seas, at the polar extremes, in the remotest of places where 'fly in' workforces are necessary, and perhaps even other planets. On the other hand, the Kyoto Protocol is leading and symbolizing accumulating efforts by national, regional, and local governments to rethink resource use as a central part of strategies to achieve environmental sustainability and to resolve global, as well as local en vironmental degradation. In Freeman's terms, these ini tiatives suggest a 'green paradigm' in which economic development maintains the nonindustrial values of re sources but does not disadvantage poor and developing countries. The restructuring of resource industries is a truly vital challenge facing the green paradigm and economic geography has an important role in under standing and prescribing the global–local dynamics of positive sum processes and outcomes in relationships between environment and economy.

Acknowledgments

The author appreciates the support of SSHRC, specif ically grant # 410 2001 0071, for financial assistance in the research for this article.

See also: Resource and Environmental Economics; Resource Management, Rural.

Further Reading

Auty, R. M. (1995). *Patterns of development: Resources, policy and economic growth*. London: Arnold.
Auty, R. M. (2001). *Resource abundance and economic development*. Oxford: Oxford University Press.

Bakker, K. (2005). Neoliberalizing nature? Market environmentalism in water supply in England and Wales. *Annals of the Association of American Geographers* 95, 542 565.

Baldwin, R. E. (1956). Patterns of development in newly settled regions. *Manchester School of Economics and Social Studies* 24, 161 179.

Ballard, C. and Banks, G. (2003). Resource wars: The anthropology of mining. *Annual Review of Anthropology* 32, 287 313.

Bradbury, J. (1979). Toward an alternative theory of resource based town development in Canada. *Economic Geography* 55, 147 166.

Bradbury, J. and Sendbuehler, M. (1988). Restructuring asbestos mining in Western Canada. *The Canadian Geographer* 42, 129 144.

Bradshaw, B. (2003). Questioning the credibility and capacity of community based resource management. *The Canadian Geographer* 47, 137 150.

Bridge, G. (2002). Grounding globalization: The prospects and perils of linking economic processes of globalization to environmental outcomes. *Economic Geography* 78, 361 386.

Clapp, R. A. (1998). The resource cycle in forestry and fishing. *The Canadian Geographer* 42, 129 144.

Clapp, R. A. (2005). Wilderness ethics and political ecology; remapping the Great Bear rainforest. *Political Geography* 23, 839 862.

Crook, C. and Clapp, R. A. (1998). Is market oriented forest conservation a contradiction in terms? *Environmental Conservation* 25, 131 145.

Edgington, D. and Hayter, R. (1997). International trade, production chains and corporate strategies: Japan's timber trade with British Columbia. *Regional Studies* 31, 149 164.

Fleming, D. (1967). Costal steelworks in the Common Market. *Geographical Review* 576, 48 72.

Freeman, C. (1992). *The economics of hope*. London: Pinter.

Freundenburg, W. R. (1992). Addictive economies: Extractive industries and vulnerable localities in a changing world economy. *Rural Sociology* 57, 305 332.

Hanink, D. M. (2000). Resources. In Sheppard, E. & Barnes, T. J. (eds.) *A companion to economic geography*, pp 227 241. Oxford: Blackwell.

Hay, A. (1976). A simple location theory for mining. *Geography* 71, 65 76.

Hayter, R. (2003). "The war in the woods": Post fordist restructuring, globalization and the contested remapping of British Columbia's forest economy. *Annals of the Association of American Geographers* 93, 706 729.

Hayter, R., Barnes, T. and Bradshaw, M. (2003). Relocating resource peripheries to the core of economic geography's theorizing: Rationale and agenda. *Area* 35, 15 23.

Hayter, R. and Le Heron, R. B. (2002). Industrialization, techno economic paradigms and the environment. In Hayter, R. & Le Heron, R. (eds.) *Knowledge, industry and environment: Institutions and innovation in territorial perspective*, pp 11 30. London: Ashgate.

Hecht, S. B. and Cockburn, A. (1989). *The fate of the forest: Developers, destroyers and defenders of the amazon*. New York: Verso.

Lahiri Dutt, K. (2003). Informal coal mining in Eastern India: Evidence from the Raniganj Coalbelt. *Natural Resources Forum* 27, 68 77.

Le Billon, P. (2001). The political ecology of war: Natural resources and armed conflict. *Political Geography* 20, 561 584.

Leslie, D. and Reimer, S. (2006). Situating design in the Canadian household furniture industry. *The Canadian Geographer* 50, 319 341.

McCarthy, J. (2006). Neoliberalism and the politics of alternatives: Community forestry in British Columbia and the United States. *Annals of the Association of American Geographers* 96, 84 104.

Odell, P. R. (1963). *An economic geography of oil*. London: Bell.

Polanyi, K. (1944). *The great transformation*. New York: Rinehart.

Parker, P. (1995). Canada Japan coal trade: An alternative form of staple production model. *The Canadian Geographer* 41, 248 267.

Patchell, J. and Hayter, R. (1997). Japanese precious wood and the paradox of added value. *Geographical Review* 87, 375 395.

Peet, J. R. (1969). The spatial expansion of commercial agriculture in the nineteenth century; a Von Thunen interpretation. *Economic Geography* 45, 282 301.

Prudham, W. S. (2002). Downsizing nature: Managing risk and knowledge economies through production subcontracting in the Oregon logging sector. *Environment and Planning A* 34, 145 166.

Rees, K. G. and Hayter, R. (1996). Flexible specialization, uncertainty and the firm: Enterprise strategies and structures in the wood remanufacturing industry of the Vancouver metropolitan area, British Columbia. *The Canadian Geographer* 40, 203 219.

Reiffenstein, T. and Hayter, R. (2006). Domestic timber auctions and flexibly specialized forestry in Japan. *The Canadian Geographer* 50, 503 525.

Skúlason, J. B. and Hayter, R. (1998). Industrial location as a bargain: Iceland and the aluminum multinationals. *Geografisca Annaler* 80B, 29 48.

Storey, K. (2001). Fly in/fly out and fly over: Mining and regional development in Western Australia. *Australian Geographer* 32, 133 148.

UNCTAD (2005). *UNCTAD Handbook of statistics 2005*. New York: United Nations Publications.

Vernon, R. (1971). *Sovereignty at bay: The multinational spread of US enterprises*. New York: Basic Books.

Watkins, M. (1963). A staple theory of economic growth. *The Canadian Journal of Economics and Political Science* 29, 141 148.

Relevant Websites

http://permanent.access.gpo.gov/lps35389/2000/natural resources.html
CIA World Fact Book of natural Resources.

http://www.fao.org
Food and Agricultural Organization of the United Nations.

http://www.fao.org
Greenpeace International.

http://comtrade.un.org/db/
United Nations Commodity Trade Statistics.

Resource Management, Rural

L. M. B. Harrington, Kansas State University, Manhattan, KS, USA

Glossary

Agroecology The application of ecological concepts to agricultural management.

Billfish Large fish with a long swordlike bill or beak, including swordfish and marlin.

Biotic Resources Resources that are derived from living things, or are themselves living things, such as timber and wildlife.

Dustbowl A region of the Great Plains of North America that suffered very large dust storms and wind erosion in the 1930s as a result of agricultural expansion and removal of vegetation, followed by a period of severe drought. This led to farm abandonment and population exodus from the region at that time.

Ecosystem The collection of interacting biotic and abiotic components of a particular area, normally with reference to natural components and processes.

Ecosystem Services Processes and conditions provided by ecosystems that benefit humans and other living things. These include nutrient cycling, waste assimilation, and maintenance of biological diversity.

Erosion Loss of soil materials due to water or wind.

Genetically Modified Organism (GMO) A living thing whose genetic makeup has been altered from that found in nature by genetic engineering, commonly by addition of DNA from another organism in order to achieve some desirable characteristic.

Goods Materials or conditions which are seen as having value, and for which people generally are willing to pay.

Groundwater Water occupying spaces in layers of rock or sediments below the surface, generally below the soil.

Holistic Management Management of resources with concurrent consideration of environmental, economic, and social conditions, along with both short- and long-term goals.

Management Practices and decision making pertaining to the use and maintenance of resources.

Mill Towns Towns that generally grew or were built around factories, such as textile mills. In some cases the town or housing was owned by the company running the mill and employing town inhabitants.

Nutrients Chemical elements that support plant life and growth.

Resources Materials and conditions of use or value, generally taken to specifically apply to human use.

Shellfish Aquatic animals with exoskeletons or shells, including both mollusks (clams, oysters, scallops) and crustaceans (lobsters, shrimp, crabs).

Sustainability An ability for conditions or resources to be maintained or improved over the long term.

Introduction

Rural people and places historically have been intimately tied to the natural resources of the place. Use and management of resources have built thriving societies, from ancient Mesopotamia to major modern agricultural regions, and have led to regional decline when human related stresses have caused major reductions in resource quality or availability, from salinity triggered decline of Mesopotamia to the North American Dustbowl of the 1930s.

The resources most clearly associated with rural regions and rural prosperity are the natural resources upon which agriculture most relies – soils and water. However, rural places, including small to medium size towns that are closely linked to the activities of the countryside, can be affected by other natural resources, including climate, plant life, fisheries and wildlife, minerals, and scenery (**Table 1**). Human and cultural attributes should also be considered as significant rural resources, representing not only labor and brainpower, but also history and ethnicity. Like scenery, cultural resources serve some small towns as tourist draws and help to sustain the rural communities where these attributes have been recognized and given value.

Table 1 Rural livelihoods and key resources

Resource-based rural livelihoods	*Key resources*
Agriculture	Soil, water, climate; livestock breeds, crop varieties
Fishing	Finfish and shellfish stocks
Timber production	Forests; tree species
Alternative resource gathering	Forests; edible and pharmaceutical biota, ornamental plants
Mining	Fuel and nonfuel minerals
Recreation and tourism	Scenery, historical structures and districts, ethnic and cultural heritage

Resource Dependency

The concept of resource dependency has been applied to rural areas. In essence, many rural regions and small towns are so dependent on their particular resources that a particular resource or resource based activity represents by far the most important basis for the economic well being of the place (and often the general health and sustainability of the community). The US Department of Agriculture's Economic Research Service (ERS) has categorized rural counties as farming dependent, mining dependent, manufacturing dependent, government dependent, and nonspecialized. Other, potentially overlapping, ERS categories include retirement destination, federal lands, commuting, persistent poverty, and transfers dependent. The discussion below focuses on several resources on which communities may be dependent, including some not made explicit in the ERS labels.

Widely varied rural environments exist around the world, and within individual nations. Correspondingly, the main forms of rural livelihoods are connected to native natural resources, and in some situations the cultural and historical development of the area (**Table 1**). Among the natural resources, many are renewable if well managed. These include the biotic resources relevant to agriculture, fisheries, timber production, and some alternative resource gathering activities. In general, water resources are considered renewable, though use of some groundwater sources has brought them to nonrenewable status. If managed carefully, soil may be considered as a renewable resource – as long as erosion is not allowed to exceed soil formation rates in the long term. High occupational hazard levels accompany many resource oriented activities, including mining, ocean based fishing, logging, and even agriculture.

Agriculture

Agriculture is the activity that the general public most clearly identifies with rural places. It also is clearly dependent on a number of natural and modified resources; farming communities are in turn dependent on agricultural activities. Climate constitutes one of the natural resources on which agricultural activities are based; plants have particular temperature and moisture requirements and farmer choice of crops (and livestock) will be affected by climate. In some regions where rainfall would be insufficient to grow certain crops, there may be a high degree of reliance on water provided through irrigation systems. This adaptation can overcome some climate limitations where suitable supplemental water sources are available. In some areas, however, significant dependence on nonrenewable groundwater resources has developed over time, and future adjustments to declining water availability are likely. In regions where both rainfall and supplemental water sources are limited, agriculture may rely on crops that are better adapted to seasonally dry conditions, such as wheat, or take the form of ranching where the success of even drought tolerant crops is unlikely.

In addition to climate and water, agriculture is reliant on soil. Like other natural factors, soil varies from location to location. Soil characteristics, particularly texture and chemistry, affect plant access to moisture and nutrients essential for growth. Organic material, generally found in the upper layers of the soil, is important to soil texture and nutrient status. Overuse of soil can result in a loss of nutrients and fertility reduction.

Biotic resources – crop plants and domestic livestock – also constitute resources for agriculture. These often are far different than their wild progenitors, but are essential for farming. Their productivity is dependent on other resources of the farm, and their management can affect future productivity.

Forests and Timber

In some rural regions, the dependence is on timber production and other forest related resources. Timber based communities in North America often are in areas dominated by publicly owned (federal, state, or provincial) forest areas. However, in this and other world regions, large forest areas also may be owned by corporations as industrial forestlands, and many smallholders who may be involved in agriculture or other activities own woodlots which may be used for household purposes or occasional timber sales. In addition to the obvious timber resource of forests, nontimber forest products may be drawn on by individuals seeking income supplements or household use, including food sources (berries, wild meat, mushrooms), floral and decorative products (ferns, evergreen branches, Christmas trees), and medicinal products (ginseng, cascara (laxative), taxol (cancer treatment drug from yew trees)). Forest scenery and wildlife can contribute to recreation and tourism. Management of forested land also can be important for maintaining water quality and reliable flow. Forest resources generally are considered as renewable, or flow, resources – if managed properly, supplies of these resources should not diminish over time because they are naturally replenishable.

Fisheries

Though not recognized in the ERS county types, communities along oceans or seacoasts, or major lakeshores, may exist with a high level of dependence on fisheries. Numerous fishing villages exist in both the Northern and Southern Hemispheres, many with very long histories of connections to the sea. Although fisheries also constitute renewable resources, widespread overfishing has become a concern regarding the sustainability of the resources and of fishing communities. Overfishing in the North Atlantic has caused major declines in cod and billfish stocks. In Nova Scotia, for example, economic hardships have led many to seek employment elsewhere in Canada. In addition to finfish, shellfish capture or production can be important to some communities. This resource can include lobsters, crabs, shrimp, clams, or oysters, depending on the environment.

Mineral Resources

Unlike farming, forest dependent activities, and fisheries, mining and fossil fuel energy resource dependencies are based on nonrenewable resources. From the coalfields of Europe, which began to be tapped extensively beginning in the late 1700s with the Industrial Revolution, to the silver and gold boomtowns of the New World, communities have sprung up, expanded, and sometimes collapsed, depending on mineral markets and the status of local deposits. Mining areas are particularly sensitive to the economics of market changes and resource depletion. Fossil fuel resources (petroleum or oil, natural gas, and coal) are likewise nonrenewable. In some rural regions, these may constitute a significant portion of the income, particularly in coal producing areas. The Appalachian coal belt of eastern North America has long been known as a relatively low income rural region, with many communities closely tied to work in the mines. In other areas, including places in North America, South America, and Africa, mining for metallic minerals provides a high proportion of jobs.

Recreation, Tourism, and Retirement

In the modern era, recreation and tourism activities within society and to the economies of specific places have become important. Attractions include scenery (both natural and human, including more picturesque farming areas), environmental conditions suitable for particular outdoor pursuits (downhill skiing; mountain climbing; hiking, tramping, or hillwalking; bicycling; recreational fishing or angling; hunting; agritourism; and dude ranch stays), historic buildings, and cultural or ethnic heritage (with shopping opportunities oriented toward regional or community heritage). Some communities previously dependent on other resources, such as mining towns in mountain areas, have been able to make a switch to recreation based economies. Fishing villages in Ireland are now touted as scenic tourist destinations. In a few situations, a cultural identity has been constructed for a place in order to serve as a unique tourist resource. An example of this is Leavenworth, Washington State, USA, which has created and marketed a 'Bavarian village' identity for itself. Oftentimes the scenery and other amenities that are attractive to tourists also attract retirees to small towns and rural areas. The pensions or social security payments that go to retirees constitute additional income to the local area.

Human Resources

Although not normally considered in discussions of resources of rural areas, for some places a labor pool also serves as a resource, generally as a draw for industry. During the early twentieth century, the attraction of a combination of cotton availability and a cheap labor pool drew textile companies to relocate from New England to the southeastern states. The southern mill towns have since declined, due to internationalization and cheaper labor and raw materials overseas.

Resource Management: Considerations and Approaches

Individual and Private Enterprise Resource Decisions

Economics

Rural resource management usually is linked to economics, whether it be for a household, a community, or a large business entity. For most rural resource users, narrow profit margins and volatile markets create situations where short term economic benefits are emphasized. With short term economic survival or benefits as the focus, longer term issues, including resource sustainability, often are overlooked. This may, in turn, lead to environmental damage, resource decline, and decreasing economic benefits over time. In the United States, some forestlands are owned by large timber corporations. Over time, their harvesting activities have surpassed tree growth rates and have reduced availability of timber. This has led to increased demands on national forests, which are managed, in part, for timber production.

In farming, economic factors have tended to lead to greater concentration of livestock in confined feeding operations within more industrialized agricultural systems. These operations tend to be oriented toward production of a single type of animal – often a single breed or genetic type, in order to ensure consistent products and ease of processing. This follows on earlier trends toward monocultural crop farming in the 1960s and 1970s. Both trends toward greater specialization have led to biological simplification within agriculture. While economically efficient, specialization to this degree does increase concerns with vulnerability to disease outbreaks.

Economic goods or commodities based on rural resources in a particular place can shift as rural inhabitants continue to seek means by which to improve income levels, as knowledge of both options and environmental relations improve, and as urban consumers change in their tastes and desires. In this way, scenery and recre ational opportunities have been commodified, and now either supplement historic resource based activities, or may have become dominant in local rural economies. Tourism often feeds commodification of rural scenic, recreational, and historic amenity areas. In historically mining and timber based areas of the US Rocky Mountains, for example, (re)development of some small towns is oriented toward tourism, skiing, whitewater rafting, and bicycling. Housing developments are occur ring in rural areas that in the past had been fairly isol ated, but that have amenity values for which more people today have the desire and economic wherewithal (and perhaps the technology to work from home) to obtain. These types of developments are controversial, with those seeking growth and development pitted against those seeking to maintain traditional landscapes and livelihoods.

Those whose livelihoods are based on natural re sources seek to increase efficiency and output of products in demand. When societal demands change – developing, for example, a market for organically produced foods and 'exotic' vegetables – agriculturalists respond. New crops and products, and shifts in production techniques, have appeared with other changes. Acceptance of changes to products and methods of production can vary widely among individuals. Industrialization of agriculture, in volving more intensified production and emphasis on the use of modern technologies, has continued for the last several decades, in answer to desires for greater effi ciency. Highly specialized farming involving high reliance on synthetic inputs (potentially including gen etically modified organisms, GMOs) has come to be called 'factory farming', as resources and processes used now create production that resembles manufacturing more than traditional farming. 'Factory farms' can in clude, for example, modern swine operations with large numbers of nearly identical animals housed in mechan ically cooled and heated buildings, and fed supplements of antibiotics that can help promote weight gain, as well as health. Alternatively, more organic production, first by small acreage farmers and more recently by larger op erations and corporate producers, has been an answer to growing demand for alternatives to foods and production practices involving synthetic chemical inputs.

For some individuals and nonprofit resource owners, economic benefits are not the only desirable environ mental 'good'. Decisions regarding privately held lands and resources may be made based on other con cerns, including preservation of wildlife, ecosystems, and scenery. Management decisions regarding resources in governmental ownership likewise take environmental goods, services, and other values into consideration, along with commodity production goals. For example, because national forests also are valued for their maintenance of natural processes, ecosystem services, natural (or seminatural) communities, biodiversity, and scenery, demands on these forests for timber production can be controversial and new timber cuts often are met with opposition from other interest groups.

Scale

Scale is a concern with rural natural resource conditions and management. Individuals as property owners gen erally control relatively small acreages (to several hun dred acres, depending on history and environment). Some may control thousands of acres, mainly in places like the North American Great Plains and the interior of Australia, where environmental conditions require larger landholdings in order to produce enough for the eco nomic survival of the farm or ranch operation. Corporate control may reach across millions of acres: Weyerhaeuser, a forest products company, manages over 34 million acres (approaching 14 million ha) of forest in five countries. Decision making of many landowners based on a patchwork of smaller landholdings will have different bases, including varying economic pressures, and lead to a patchwork of outcomes; decision making where large land areas are involved tends to be adjusted to spatial variations in resource availability. These differences result in variations in resource use and connected environmental effects. Actions that may lead to reduced resource sustainability at one scale or in a specific area may have different results regarding sustainability at another scale.

Management options

Soil and water

A wide range of specific methods are utilized for man aging natural resources in rural areas. These have con tinued to develop as understanding of resources has improved. Regarding soil management, nutrient return is generally accomplished through fertilization, additions of organic matter, and crop rotations. Overuse can result in soil loss (erosion) exceeding the rate of soil formation. In recent decades, techniques to better manage soil re sources have been more widely adopted, including not only crop rotation, terracing, and plowing on the contour, but also no till or reduced tillage farming, grassed waterways, stream buffer zones, and site specific or pre cision agriculture. Many of these techniques constitute improved management of both soil and water resources – reductions in soil loss to streams both help to maintain soils on the farm or forest, and help to reduce sedi mentation and chemical contamination of water bodies.

Broad concerns with usage of agrochemicals, in cluding both pesticides and fertilizers, have led to ex pansion of alternative production methods, including low input agriculture (specifically reduced inputs of manufactured chemicals), integrated pest management (again, involving a reduction of conventional chemicals), and organic agriculture. Such reductions in chemical use can positively affect soils, biota and ecosystem structure, and water quality.

Water resource management regarding quantity available is most critical in areas with insufficient or marginal rainfall for crop growth. However, concerns are growing in regions that have higher rainfall but are stressed by growing populations and occasional drought. Efficiency of water use has been improved since the 1970s, with new irrigation delivery technologies. The newest of these is subsurface drip irrigation, which delivers water directly to the root zone of plants and greatly reduces evaporative loss to the atmosphere. Low water demanding crops can replace those with high moisture needs, though this also tends to reduce income from crop sales. Work is progressing on developing higher yielding crop varieties that have lower water needs.

Timber and forests

Forest management has numerous options, as does agricultural land management. In some cases similar actions, like stream buffer zones, are utilized for water protection. Indeed, stream buffers have been mandated in areas where past practices had included logging up to the edge of waterways. Timber management may also include chemical use (fertilization and pesticide applications), and often involves replanting of trees after harvest utilizing nursery grown seedlings. In some areas, timber production takes the form of tree plantations, with planting of a single species in rows leading to areas that greatly resemble rows of a crop on a farm.

In many cases, the replanting procedures lead to creation of simplified forests, as compared to more species diverse natural forests. Not only is there a tendency to replant a single species of tree, but the process of completely clearing an area of all timber at one time (through clear cutting), followed by replanting at one time, leads to even age stands of trees. (Even age stands are natural in some locales, where forest regrowth after a fire will tend to create stands of trees of the same age, and where fire adapted species are apt to dominate.) From a harvesting perspective, this creates a more efficient situation where all trees of the same type and age can be removed and processed at once. It does, however, diminish biodiversity in most cases. An alternative harvest method is selective cutting, where specific trees or groups of trees are identified for removal, leaving others behind. Selective cuts are less economically efficient than clear cuts, though they are more esthetically pleasing and may be much better for biodiversity (depending on the basis of the selection).

Fisheries

Regarding fisheries management, much of the damage to fish populations has come about due to advances in technology to the point very large vessels can fish large areas over large periods of time, processing and storing the catch for days to weeks. Bottom trawling has become more efficient, not only allowing for capture of more fish, but also causing widespread habitat destruction. The nature of fisheries makes control of these activities very difficult, and a high demand for seafood continues to drive much of this type of resource extraction. The large corporate fishing vessels stand in stark contrast to more traditional small fishing boats owned by individual fishers.

Other dramatic changes to fisheries include fish farming, where finfish are grown in ocean enclosures or freshwater ponds, or shellfish are grown in enclosures (in the case of shrimp) or in tidal mudflat areas (in the case of mollusks). These activities are more sustainable than those based on large open ocean vessels, but environmental concerns accompany them, as well, including loss of mangrove forests protecting tropical coastal zones and concerns about escape of farmed fish and their impacts on native species. Additionally, there are concerns about biocide use in some fish farming operations, as well as potential overfertilization of water bodies through the feeding process.

Minerals

Because minerals are a nonrenewable resource, there are few management options available to maintain production and the communities based on these resources. Production is driven by economics. Technological change can lead to extraction of minerals that were unavailable under prior technologies, and so can lengthen the economic life of mineral deposits and the rural communities dependent on them. Because many mining activities result in undesirable environmental impacts, it has fallen to government mandates for particular actions for cleanup and mitigation of mine related problems.

Alternative approaches

With greater recognition of the interconnectedness of processes in natural and human managed systems, some people have promoted resource management techniques that consider these connections to optimize productivity and sustainability while reducing negative side effects. At the scale of a farm or other private enterprise unit, agroecology and holistic management have been promoted as systems to better manage resources for long term goals. Holistic management is perhaps a broader concept, as it is not agriculture specific. The main idea of holistic management is that environmental, economic, and social conditions should be considered simultaneously, and consideration must be given to both short and long term goals. Holistic range management has been promoted as a regional scale approach to resource management. The central idea behind agroecosystem management is that a farm works like an ecosystem, al beit with a very strong component of human influence. Rather than managing only for a specific crop, the interactions of that crop with other biota and with the nonliving environment should be considered along the same lines as processes at work in any ecosystem. Managing for the processes, and for goals beyond pro duction of the crop by itself, should enhance sustain ability. Most small organic farms are approached, in essence, with an agroecological frame of thinking. Agro ecology also can be approached at a broader scale, rather than the single farm, and much of the focus is given to agricultural systems.

Government Policy

Driving forces: environmental and economic concerns

Governments have regulated certain resources for cen turies. The ownership and control of forest resources and wildlife by the government dates back at least to estab lishment of royal 'forest' areas designated as under the control of the Crown for management of timber and game resources (e.g., Sherwood Forest, of Robin Hood fame). Government ownership of specific land areas and control of resources continues in nations around the world, including national forests, national parks, reserves, and other areas.

Protection of wildlife continued with regulation of hunting early in colonial settlements, though this was not universal. Regulations regarding natural resources more apt to noticeably display the results of overuse, like wild game and forests, were addressed fairly early. Likewise, rules regarding water use were established in Europe, and transplanted to the New World, to reduce conflicts over this resource.

Lands in government ownership exist mainly because the resources were seen as either needing protection from rampant individual or corporate exploitation and accompanying environmental damage, or to provide conservation of resources seen as vital for national strength. For the United States, widespread deforestation and resulting erosion and flooding in the 1800s were driving forces in developing not only national forests, but forest policies to promote conservation. At that time, forests were highly important to national defense, as well as to settlement and the economy. Concern over damage to and potential loss of natural scenery, including Niag ara Falls views and Yellowstone thermal features, led to federal reservation of areas with dramatic scenery as exploration and settlers spread westward. Similar trends occurred in Canada and Australia as Western European settlers and European traditions became established. Later on, the US Dustbowl period in the 1930s brought about national awareness of soil management needs and led to the creation of an agency and policies addressing soil resources and conservation. Heightened awareness of general environmental conditions in the 1960s and 1970s led to numerous laws and expanded policies regarding soil, water, air, and forest resources. Over time, multiple agencies addressing rural resource issues have been es tablished in individual countries.

In the US, one of the most specific policies relevant to management of rural resources is the Conservation Re serve Program (CRP). Established by agricultural legisla tion in 1985, the CRP allows farmers to enroll land with highly erodible soils to receive payments for production foregone, while cover vegetation is planted and the land is maintained without cultivation. The CRP originally was meant to serve the dual purposes of promoting soil pro tection and helping to reduce agricultural production (and thus lead to increased prices through reduced supply). In recent years, it also has been recognized as a way to provide wildlife habitat and help sequester soil carbon.

New Zealand has perhaps attempted more integration of resource management around the concept of sustain ability than any other country. Through its Resource Management Act of 1991, resource and land management legislation were integrated into one law oriented toward sustainability. Decision making became more locally fo cused, though concerns at broader scales could override these. New Zealand, unlike other countries, also re oriented its local governments to watershed based regions – watersheds often are considered a more rea sonable approach to resource management than areas defined by more arbitrary boundaries due to the en vironmental linkages found within watersheds.

Governments develop economically oriented policies that clearly affect resource management, including trade. The United States and Canada have had a timber trade dispute for several years that has restricted imports of Canadian softwood lumber into the US; this policy dis pute has affected production in both countries. The European Union Common Agricultural Policy (CAP) addresses farm management for environmental quality in farm payments. CAP also addresses farm product sub sidies and international trade, which also can affect farmer decisions. The European Union also emphasizes sustainability in its rural policies, and countries work together to develop forestry and fisheries policies, as well as agricultural policies. In addition to addressing activ ities like farming, fishing, mining, and forestry, govern ment action also is taken to promote rural community development through tourism promotion, creation of facilities, and other activities.

For public land management, goals vary with creative legislation. In some cases the designated purposes for land management are very narrow. In others, multiple uses (potentially including timber production, wildlife habitat, grazing, water resource protection, recreation, support of local economies, and preservation of natural features) may be specified, requiring careful judgments and balancing of conflicting interests.

Scale issues and a governmental alternative approach

Governments are in a position to address scale issues far differently than those pursuing household or business goals. While a large number of landholders and their individual decisions lead to fragmented resource management, government policy and programs can potentially lead to greater integration of resource management decisions over larger areas and for longer time frames. Ecosystem management is an approach similar to that of holistic management, though it often is approached over areas so large that government bodies are likely to be involved. Like holistic management, the focus is on integrating environmental, economic, and social goals, and like agroecosystem approaches, consideration of ecosystem processes and interconnections is critical. A large project in the United States known as the Interior Columbia Basin Ecosystem Management Project covers portions of several states and involves several federal agencies. Much of the achievement of the project has been in terms of scientific data development and integration. For successful implementation of management over a large ecosystem with multiple landowners, broad communication and cooperation among multiple levels of government, businesses, and individuals is necessary. Other countries and international agencies also have adopted an ecosystem management framework for particular projects, including the World Conservation Union (International Union for Conservation of Nature, IUCN).

Nongovernment Organizations

A number of nongovernment organizations are active in pursuing resource conservation and preservation of rural landscapes and lifestyles. They may be fairly broad in focus (e.g., the Natural Resource Defense Council), though other organizations exist with more specific goals, such as the American Farmland Trust, American Forests, the Rare Breeds Survival Trust, and American Water Resources Association. Professional organizations also exist to support use and management activities related to the specific resource. Given the nature of demands on resources and land areas to provide multiple goods which may not always be compatibly produced, various resource interest organizations frequently are in conflict with each other.

Conclusions

Rural resources include a very broad set of natural and cultural resources. Historically, rural regions and small towns have been highly dependent on their local natural resources. In some areas, dependence has shifted to cultural resources as attractions for tourism, but the rural landscape still often serves as a component of recreation and tourism attraction.

Management of natural resources in rural regions has changed over time, with changes in technology, improved understanding of resources and environmental interconnections, and changes in human populations and demands for resource based goods. Indeed, rural landscapes and resources are constantly changing, although those changes often are not noticeable to the casual onlooker, as when crops and crop varieties are shifted from one year to the next. In some places, improvements to management have occurred, although economic conditions and demands on resources also have resulted in resource damage. Greater concerns with resource sustainability have become evident, with both individual landowners and government bodies pursuing these goals. The well being of rural regions relies to a large extent on recognition of and continued improvements to management of their resources.

See also: Agricultural Land Preservation; Agriculture, Sustainable; Agri-Environmentalism and Rural Change; Animal Geographies; Conservation and Ecology; Economic Development, Rural; Ecotourism; Environmental Policy; National Parks; Rural Communities; Rural Geography; Sustainability; Tourism, Rural; Water Management.

Further Reading

Altieri, M. A. (1995). *Agroecology: The Science of Sustainable Agriculture*. Boulder, CO: Westview Press.

Aplet, G. H., Johnson, N., Olson, J. T. and Sample, V. A. (eds.) (1993). *Defining Sustainable Forestry.* Washington, DC: Island Press.

Beesley, K. B., Millward, H., Ilbery, B. and Harrington, L. (eds.) (2003). *The New Countryside: Geographic Perspectives on Rural Change*. Halifax: Brandon University and St. Mary's University.

Bowler, I. R., Bryant, C. R. and Nellis, M. D. (eds.) (1992). *Contemporary Rural Systems in Transition, Vol. 1: Agriculture and Environment, and Vol 2: Economy and Society.* Wallingford: CAB International.

Cloke, P., Marsden, T. and Mooney, P. H. (2006). *Handbook of Rural Studies*. London: Sage.

Essex, S. J., Gilg, A. W., Yarwood, R. B., Smithers, J. and Wilson, R. (eds.) (2005). *Rural Change and Sustainability: Agriculture, the Environment and Communities*. Wallingford: CABI Publishing.

Furuseth, O. and Cocklin, C. (1995). An institutional framework for sustainable resource management: The New Zealand model. *Natural Resources Journal* 35, 243 273.

Ilbery, B., Chiotti, Q. and Rickard, T. (eds.) (1997). *Agricultural Restructuring and Sustainability: A Geographical Perspective*. Wallingford: CAB International.

Owen, L. and Unwin, T. (eds.) (1997). *Environmental Management: Readings and Case Studies*. Oxford: Blackwell.

Perkins, H. C. (2006). Commodification: Re resourcing rural areas. In Cloke, P., Marsden, T. & Mooney, P. H. (eds.) *Handbook of Rural Studies*, pp 243 257. London: Sage.

Pound, B., Snapp, S., McDougall, C. and Braun, A. (eds.) (2003). *Managing Natural Resources for Sustainable Livelihoods: Uniting Science and Participation*. London: Earthscan.
Savory, A. and Butterfield, J. (1999). *Holistic Management: A New Framework for Decision Making* (2nd edn.). Washington, DC: Island Press.
Wallach, B. (1991). *At Odds with Progress: Americans and Conservation*. Tucson, AZ: University of Arizona Press.
Woods, M. (2005). *Rural Geography*. London: Sage Publications.

Relevant Websites

http://www.europa.eu
Activities of the European Union, Agriculture Fisheries, Environment, and Other Issues at European Commission.
http://www.ec.europa.eu
Agriculture and Rural Development, European Commission.
http://www.farmland.org
American Farmland Trust.
http://www.dfo mpo.gc.ca
Atlantic Fisheries Policy Review, Fisheries and Oceans Canada.
http://www.planeta.com
Global Journal of Practical Ecotourism.
http://www.ers.usda.gov
Natural Resources and Environment, US Department of Agriculture Economic Research Service.
http://www.nrcs.usda.gov
Natural Resources Conservation Service, US Department of Agriculture.
http://www.affa.gov.au/brs
Natural Resource Management; Bureau of Rural Sciences at Australian Government Department of Agriculture, Fisheries and Forestry.
http://www.rural.gc.ca
Rural Tourism Pathfinder, Canadian Rural Information Service.
http://www.fs.fed.us
State and Private Forestry, USDA Forest Service.
http://www.maf.govt.nz
Sustainable Resource Use, Ministry of Agriculture and Forestry (New Zealand).
http://www.nal.usda.gov
US Department of Agriculture National Agricultural Library Rural Information Center, Promoting Tourism in Rural America.

Retail Geographies

N. Wrigley, University of Southampton, Southampton, UK

Glossary

Corporate Lock-In The inability of a firm's management to initiate changes in strategy and drive through essential corporate restructuring in the face of changing competitive conditions.
Corporate Restructuring A strategy of corporate change that materially alters the composition of a firm's asset portfolio and/or claims against those assets.
Digital Natives Demographic groups who have grown up in the Internet age.
E-Tailing Retailing via B2C (business to consumer) electronic commerce (e-commerce) – alternatively, and less frequently, via the 'infomediary' role in which no inventory risk is assumed and profitability of the e-tailer is dependent on making a commission on transactions from either buyer or seller.
Ethical Trading Responsible global sourcing involving codes of conduct for labor standards at sites of production.
Food Deserts Conventionally defined as areas of poor access to the provision of healthy affordable food where the population is characterized by deprivation and compound social exclusion.
New Retail Geography The name which emerged in the 1990s for a reconstructed subdiscipline of retail geography characterized by theoretical engagement and by an appreciation that the transformation of retail capital and of its production and consumption spaces offered some of the most challenging areas of study in human geography.
Retail FDI Foreign direct investment by retail transnational corporations (TNCs).
Urban Regeneration Partnership Stores Retail stores developed in underserved, deprived urban communities in the form of a partnership between retailers, city authorities, labor unions, and local community groups involving measures such as job guarantee schemes as a route into work for unemployed local residents.
Virtual Products Products such as those generated from textual, musical, and visual data which can be sought out, transacted, transported, and consumed all within electronic space.

Introduction – The New Retail Geography

Conventionally defined during the period spanning the 1960s to 1980s as the study of interrelations between the spatial patterns of retail location and organization on the one hand and the geography of consumer behavior on the other, 'orthodox' retail geography during the late 1980s became subject to increasing criticism of what Ducatel and Blomley described as its 'wholly inadequate analysis' of retail change lacking a 'systematic account of retail capital and its transformation'. In response, and driven by:

1. a growing realization among social scientists of the progressive concentration of capital in the retail industry, the very visible rise of the retail corporation (the 'mega chain' such as Wal Mart in the USA and Tesco in the UK), the increasing importance such corporations had begun to assume in several Western economies, and the emergence of retail corporations as lead firms in what Gereffi termed 'buyer driven' global commodity chains;
2. arguments that retailing and its emerging corporations had become a pivotal industrial sector in the shift to 'post Fordism'; and
3. the beginnings of a wider movement across the social sciences and humanities toward the study of consumption, consumer society, and consumer culture;

a reconstructed subdiscipline of retail geography emerged – often referred to Wrigley and Lowe as the 'new retail geography'. That reconstructed retail geography of the 1990s and early 2000s, responded to Blomley's challenge to take both its economic and cultural geographies seriously. It was characterized above all by theoretical engagement, and by an appreciation that the transformation of retail capital and of its production and consumption spaces offered some of the most fascinating and challenging areas of study in contemporary human geography.

This article outlines some of the key themes which have emerged within the 'new retail geography' over the past 15 years focusing, because of space constraints, on five issues relating to the economic geographies of retail capital. A broader treatment which also tackles the 'cultural logic of retailing', and which explores consumption spaces and places and how those spaces are 'actively produced, represented, and contested', is contained in the major summary of the 'new retail geography' provided by Wrigley and Lowe.

Geographies of Retail Concentration, Corporate Restructuring, and Regulation

By the early 2000s, when ranked in terms of global sales, Wal Mart had become the world's leading industrial corporation. By 2006–07 its annual sales of $345 billion were larger than the gross domestic product (GDP) of some sizeable industrial countries, and it had profound impacts on many aspects of the US economy from consumer price inflation, through labor productivity rates, to trade relations (e.g., accounting alone for over 10% of all US imports from China). Likewise in UK, the leading retailer Tesco (with 2006–07 global sales of approximately $90 billion) had emerged as the largest private sector employer of labor (280 000 workers in 2006–07) and held a similar pivotal position in the national economy. Both examples are illustrative of the way retail had been radically transformed in most Western economies during the final quarter of the twentieth century, by intense processes of consolidation and a strong trend toward the concentration of capital, into an industry increasingly dominated by 'big capital' in the form of large corporations. An industry, which in many countries had previously been dominated by small firms operating small stores, had been transformed into one increasingly characterized by large firms, frequently operating large stores, which in certain cases had become major corporations, some of which had developed a global presence.

The geographies of retail corporate restructuring – waves of mergers and acquisitions, periods of intense new store construction and frantic competition between retailers for the most attractive development sites (eras of 'store wars'), and the spatial consequences of the reconfiguration of the financial and organizational structures of retail firms – which characterized the radical transformation of the industry during the late twentieth century, have been documented in an increasingly rich and substantial literature. In turn, that literature has highlighted three related issues of conceptual and policy significance.

1. How are rapidly consolidating retail industries, subject to waves of horizontal merger and acquisition, appropriately regulated to maintain competitiveness both at the level of the industry and at the level of the local market – that is to say what is known about and what contributions have geographers made to the study of 'retail regulation' and the regulatory constrained nature of corporate restructuring? In this context, regulatory authorities such as the Federal Trade Commission in the USA or the Office of Fair Trading and Competition Commission in the UK have been concerned not only with the emergence of 'market dominant' retailers and their impacts – that is to say with potential 'scale' or 'complex' monopolies – but also with ensuring that local markets remain competitive, open to new entry, and that consumer welfare is protected. In particular, proposed horizontal acquisitions and mergers involving rival retail firms have been subject to increasingly stringent scrutiny in many countries (albeit with important national differences) – a trend which in some countries has been accompanied by and linked into tightening land use planning regulation. Geographers have made important contributions not only to understanding what Marsden *et al.* term the 'shaping of competitive spaces' within that regulatory environment but also, more broadly, to exploring the tensions which have developed within the complex private/public regulatory relationships which have emerged between governments and retailers.
2. What are the important issues in relation to retail corporate restructuring and its geographies which are raised by market contraction, competitive failure on the part of the individual firms, and periods of retrenchment rather than expansion? In this context, processes of divestment, asset redeployment, and market exit become the key issues within corporate restructuring, and those processes of strategic reassessment have produced a growing literature. In turn, those processes raise issues of how the executive/managers of retail firms respond to changing competitive conditions, and the way some firms suffer from 'corporate lock in' – the inability to initiate changes in strategy and drive through essential corporate restructuring in the face of rapidly changing market conditions. Here the work of Schoenberger shows that corporate lock in is intrinsically bound up with the social asset structures of managers and the issue of sunk costs, and suggests that the most severe obstacles a retail firm is likely to encounter in remaining competitive lies 'in the social being, position, and perceptions of the people who run it'.
3. What are the links between corporate restructuring and issues of corporate governance in the retail firm? In this context, Schoenberger's stress on the importance of the 'social asset structures' of managers in determining the strategy and prospects of the retail firm, raises the related and important issue of how suppliers of finance to the firm – the institutions/people who sink capital into the firm and essentially own it – seek to control its managers. To what extent are they able to influence strategies of corporate restructuring and the 'strategists' (the top management of the firm), and what are the implications? These issues of the 'internal' regulation of the retail firm (known as 'corporate governance') and its impacts, together with the links between the internal and external regulatory environments of retail firms are highlighted by Wrigley and Lowe and by Marsden *et al.* but, as yet, remain relatively underexplored within the 'new retail geography'.

Geographies of Retail Sourcing Networks and Supply Relations

As retail industries consolidated and major corporations emerged and became increasingly dominant during the final quarter of the twentieth century, there was an accompanying and fundamental shift in the balance of power away from manufacturers and toward retailers. Retailers were able to exercise power and increasingly shape and control supply relations because of four factors. First, because of their ability to deny, or credibly threaten to deny, manufacturers' access to the retail markets vital to the sale of their branded products. Second, because of their ability, as buyers of increasingly large volumes of any particular product, to demand specially negotiated and arguably 'discriminatory' price discounts related to their market dominance. Third, because they could substitute own (private) label products for the branded products of manufacturers and, by allocating high quality display space to those own label products, could place significant pressure on the lower ranked manufacturers' brands in any product field. (Additionally, they could use own label to fill gaps left in the range of products offered by manufacturers, taking more risks with product innovation and developing those own label products under tighter control.) Fourth, because of their increasing ability to take control of their own distribution systems – developing centralized, logistically efficient, quick response systems which allowed the reduction of retailer inventory holdings and the amount of capital tied up in those holdings, passing back to manufacturers more and more of those costs.

Within these reconfigured supply systems, geographers have concerned themselves with how, most appropriately, to conceptualize the nature of retailer–supplier relations, and whether different power dynamics characterize supply relations in the context of own label products compared to manufacturer branded products. Within the context of own label supply by small and medium size manufacturers, Doel draws attention to the shift toward 'associative' relationships – characterized by long term collaboration and cooperation and value adding information sharing. However, in the context of the hierarchical 'preferred supplier' relationships which emerged within clothing retail during the 1990s, Crewe and Davenport pose important questions about a reality of retailer dominance and supplier dependency within systems which appear to exhibit superficial aspects of long term collaboration and product design/information sharing. Such dominance/dependency relationships were brutally reinforced in the UK during the late 1990s as the leading clothing retailer rapidly shifted toward global sourcing networks in response to pressure to cut supply costs and maintain market share.

In the latter context, and following Gereffi *et al.*, geographers have also concerned themselves with, and made important contributions to, understanding the rise of retail corporations as lead firms within 'buyer driven' global sourcing networks and the governance of those networks. Empirically rich research on sectors ranging from horticulture, wine, and cut flowers, to clothing, furniture, and footwear, have provided insight into the rapidly expanding and closely managed global sourcing operations of major retailers and the 'buyer driven' forms of coordination and organization used by retailers within those sourcing strategies. By considering the flow of goods through these supply chains, the social and economic practices which shape that flow, and by tracing connections between different sites along the chains, geographers have also linked global networks of commodity production and retail sourcing to local sites and practices of consumption, and have traced ways in which the meanings of commodities are reworked along and across different sites and spaces in those chains.

Geographers have also contributed significantly to understanding the rise of ethical trading initiatives as a key means of governance – particularly of labor standards – in retailers' global sourcing systems since the mid 1990s. This research has focused on the implementation of codes of conduct at sites of production, on the multistakeholder bodies which have emerged to organize ethical trade and, more broadly, on the rise of corporate social responsibility (CSR) as a key issue within retailers' corporate strategies in the early twenty first century. These issues, not least, reflect the increasing significance to the investment community of the ethical performance of large corporations. In turn, ethical trading initiatives and the progressive imposition of 'private standards' of quality and safety by retailers on suppliers in global sourcing networks, provide examples of the displacement which took place in the 1980s and 1990s of regulatory functions from traditional governmental institutions within nation states to the corporate retailers. The rise of what Marsden refers to as 'private interest regulation'.

Inconstant Geographies and the Spatial Switching of Retail Capital

Consideration of the rise of corporate retail power, the reconfiguration of corporate structures in retailing, the character and shifting power balance of retailer–supplier relations, issues of retail regulation and governance, and so on provide the essential groundwork within the 'new retail geography' for a more sophisticated understanding of the 'reciprocal' nature of the relations between space and retail activity. In particular, there has been an important focus on what can be termed the 'inconstant geography' and spatial switching of retail capital, focused on the active creation and 're'creation of markets by retail capital.

In this context, geographers have shown how the retail landscape provides one of the clearest examples of what Harvey describes as the 'creatively destructive' nature of the spatial switching of capital in the built environment, with retail investment being shown to be constantly on the move creating, then abandoning, and then re discovering, spaces of profit extraction. The most well documented example of this involves the shuttling of retail investment between central cities/downtowns and the suburban periphery – in the process relocating downtowns, shifting retail investment to the suburban frontier, devalorizing and polarizing retail environments of the inner city, then rehabilitating and revitalizing some of the most degraded of these areas, perennially reversing the preexisting logic and location of profit extraction and the sources and flows of creativity in the retail landscape. Here the work of Berry which documented the de centralization of retail in Chicago during the 1950s and 1960s as the city experienced a 'rolling tide of residential segregation' provides a lasting achievement and a basis for understanding not only the subsequent periods of rehabilitation and revitalization of American downtowns, but also potential future shifts in the logic and location of profit extraction and flows of retail investment in American cities, and the reaction of European cities prepared to use land use planning regulation to protect the vitality and viability of central city retailing.

Out of this conceptual focus within the new retail geography on making sense of the retail landscape's 'structured coherence', and understanding its inconstant geographies and spatial switching of capital, two new areas of research activity began to emerge strongly by the early 2000s – a concern with geographies of retail ex clusion and underserved markets, and with geographies of transnational retail.

Geographies of Retail Exclusion/Inclusion

In both North America and Europe, the evolution of retail systems and their commercial landscapes in the latter part of the twentieth century left behind 'under served' urban areas and communities. In the USA, those were often racially segregated communities/areas of the inner city. In the UK, they were frequently highly de prived local authority housing estates developed around the periphery of the city.

Concerns regarding the consumption problems of the poor in those deprived and underserved urban areas have, in different periods, generated considerable debate across the social sciences. Following the urban riots in North American cities during the late 1960s, for example, de bates focused on the question of 'do the poor pay more?' for basic goods and services in such areas, and the extent that was a factor in urban unrest. By the late 1990s, however, it was the public health problems of these de prived underserved areas which were attracting attention. In the UK, deprived areas of cities with poor access to the provision of healthy affordable food become known as 'food deserts' and it was a metaphor of social exclusion which as Wrigley showed, captured the imagination of policymakers. Whether 'food deserts' exist in any simple sense, and can be identified in a systematic manner, is a debatable issue. However, there are clearly many signifi cantly 'unsupportive local food environments' in both North American and British cities in which food access has been (or remains) problematic, where diets of resi dents fail significantly to meet government recommended nutrition targets, and which are characterized by high levels of diet related ill health. During the late 1990s/ early 2000s, the public health consequences of under served communities/unsupportive local food environ ments became increasingly important debates to which geographers made important contributions.

The reverse side of 'underserved' urban communities, however, is that they offer potentially significantly underexploited markets and considerable development opportunities. Indeed, in the USA, Porter's 'Initiative for a Competitive Inner City' in the late 1990s argued exactly that. In the UK similarly, and also because of progressively tightening retail planning regulation, the major retailers drew the same conclusion. As a result, and for what they freely described as 'enlightened self interest' reasons, several of the major retailers adopted an urban regeneration agenda to gain planning permission for new large store development. Significant numbers of 'urban regeneration partnership stores' (involving city authorities, retailers, labor unions, and local community groups) now exist in deprived urban communities in the UK and similar initiatives are being developed in the USA. To what extent retail led urban regeneration can help alleviate social exclusion and public health problems and promote social inclusion and community cohesion in previously underserved communities, and how this re lates to the development of what geographers refer to as 'alternative food networks' is an important topic. Related issues are discussed in a significant 'food systems plan ning' literature which is developing in North America.

Geographies of Transnational Retail

Alongside the emergence of retail led (buyer driven) global sourcing networks during the 1990s, there was also a less noticed but equally significant step change in the intensity of retail foriegn direct investment (FDI) into the emerging markets of the global economy, together with the rapid rise of a small group of retail transnational corporations (TNCs) – see **Table 1**. By the mid 2000s

Table 1 Leading transnational retailers, ranked by sales outside home market, 2005

Rank	*Name of company*	*Country of origin*	*International sales (US$m)*	*International sales (% of total, 2005)*
1	Wal-Mart	United States	62 700	20
2	Carrefour	France	50 050	52
3	Ahold	Netherlands	45 352	82
4	Metro	Germany	38 502	54
5	Aldi	Germany	20 119	45
6	Lidl & Schwarz	Germany	19 832	43
7	Tesco	UK	19 640	24
8	Auchan	France	19 535	45
9	Delhaize	Belgium	18 893	79
10	IKEA	Sweden	18 868	92
11	Tengelmann	Germany	16 706	51
12	Rewe	Germany	15 207	31
13	Ito-Yokado	Japan	12 010	34
14	Casino	France	11 849	42
15	Pinault	France	11 775	46

Source: From Coe, N. and Wrigley, N. (2007). Host economy impacts of transnational retail: The research agenda. *Journal of Economics Geography* 7, 341 371.

those retail TNCs typically operated store networks across 10–30 countries – a level of internationalization comparable with many manufacturing industries – and were beginning to have profound impacts on the host economies and societies of many countries in East Asia, Latin America, and Central Eastern Europe.

The scale and importance of this spatial switching of Western retail capital into the emerging markets began to attract the attention of scholars across a wide range of disciplines. Researchers in development studies and agricultural economics, for example, conceptualized the development as 'the supermarket revolution in de veloping countries' and viewed retailers as critical inte grators of value chains in those countries. Geographers have made important contributions to these debates and have sought to redress what they perceived as the un deserved neglect of these developments within the wider literature on economic globalization. They have ad vanced a conceptualization of the retail TNC which builds on relational/network approaches in economic geography and which views retail TNCs as being ne cessarily embedded and essentially networked. Specific ally, the conceptualization draws attention to three features. First, to the necessarily high levels of 'territorial embeddedness' (in local cultures of consumption, real estate and land use planning systems, and supply net works), which retail TNCs must invest in to achieve organizational legitimacy in host markets. Second, to the fact that the network relationships (interfirm and extra firm) of the retail TNCs are simultaneously shaped by the institutional, regulatory, and cultural contexts of 'both' the home economies/societies from which the re tail TNCs have emerged and also the host economies/ societies those firms enter in the emerging markets. Third, to the 'mutual' transformation of host economies/ societies by the retail TNCs and, reciprocally of the retail TNCs themselves as they become more complex organizations as a result of operating across different national and institutional contexts.

Based on this conceptualization, Coe and Wrigley have focused on the key dimensions of the host economy developmental implications of retail FDI, and geog raphers more widely have begun to produce empirically rich studies of both the impacts of and strategic chal lenges faced by retail TNCs in emerging markets. In turn, that has influenced research in other disciplines. They have also described a process of spatial switching of retail investment into international markets in which the achievement of 'global reach' by a retail TNC is inevit ably a labored, uncertain, and above all, contested process. In that process, flows of knowledge and organ izational learning through intrafirm networks assume critical importance.

Virtual Geographies of Retail Capital

Toward the end of the 1990s, just as the major retail corporations were beginning to assume an increasingly significant and visible role in the global economy, they were suddenly challenged by the rise of B2C (business to consumer) e commerce which, for a brief period, ap peared to offer the opportunity for new business models which could alter the economics of the industry. E commerce was viewed as a potentially 'disruptive tech nology' with the ability to alter the existing basis of competitive advantage in the industry, lowering barriers to entry and circumventing the value chains of traditional retailing (a process termed 'disintermediation'). B2C 'e tailers' like Amazon.com emerged, and similar firms appeared to have the credible option of being able to reach national and global consumer markets without the need to establish a significant 'physical' presence and to incur the associated fixed and sunk costs. Gereffi's vision was of a 'new digital era of globalization' and, within that, the significant growth prospects assumed for e commerce were encapsulated during the 'dot.com' bubble at the end of the 1990s within the spectacular market valuations generated by some small and unproven startup e tailers.

In the early 2000s, the dot.com bubble came to a rapid and dramatic end, and the robustness of the 'revo lutionary' business models of the e tailers was suddenly tested against the reality of capital markets whose en thusiasm to subsidize the startup losses of e tailing had abruptly disappeared. During a period which has been referred to as 'dot bomb', the potential flaws in the e tailers business models – specifically the 'killer costs' of 'picking' and delivery involved in operating an infra structure to fulfill orders placed over the Internet, together with issues of tactility/sociality and security, were exposed and it became clear that e commerce re mains bound by geography to a far greater extent than is often suggested. Additionally, and significantly, pure play e tailers faced a growing challenge from traditional store based retailers who began to migrate part of their business online and to remodel themselves in a multi channel organizational form as 'bricks and clicks' re tailers. Relative to the pure play e tailers who struggled to achieve profitability under the burden of heavy start up costs, problems of fulfillment, and the scale require ments of the buying process and inventory control, the major store based retailers held considerable competitive advantages. They had the warehouse infrastructures, supply chain and inventory control systems, distribution systems, customer support centers, and product return networks already in place, and could seek to leverage those investments and their economies of scale (not least in sourcing) by adding e commerce operations to their existing businesses.

As pure play e tailers crashed into bankruptcy during the 'dot bomb' era, multichannel 'bricks and clicks' re tailing became the new commercial reality. Geographers have made important contributions to understanding what goes on 'behind the web store' within the organ izational structures of multichannel retailing. In par ticular, they have exposed the techniques used to manage fulfillment costs, and control what is referred to as the 'last mile' problem of e commerce. Additionally, virtual geographies of retail display – the way web stores can be engineered to entice and retain consumers, and to foster new forms of engagement with consumers – have been addressed.

Five years on from the dot bomb era, e commerce had recovered strongly and was beginning to grow rapidly in many countries – albeit penetrating different product sectors differentially due to issues of tactility, sociality, and ease of fulfillment. The growth was driven by increasing numbers of established store based retailers (trusted by consumers) developing multichannel oper ations, faster Internet and broadband technology be coming the norm, growing consumer trust in both the processes and security of online shopping, increasing consumer recognition of the price comparison strengths of online shopping, and an increasing range of 'virtual products' – products such as those generated from tex tual, musical, and visual data which could be sought out, transacted, transported, and consumed all within elec tronic space. E commerce had also begun to have in creasing impacts on the nature of physical shopping in traditional 'high streets' and shopping centers, with im portant substitution, complementarity, and modification impacts being observed by geographers. In addition, the emergence of a generation of 'digital natives' – demo graphic groups who had grown up in the Internet age – was progressively creating new types of consumption relationships between consumers, brands, and retailers. These new relationships can be illustrated by what Currah refers to as the 'digital storm' of online consumer piracy challenging copyright capitalism in the music, film, TV, and publishing industries and its implications for traditional retail channels, also by e commerce's ability to both expose and drive consumer demand down the 'long tail' of products which had previously struggled to find an audience in the physical world of retail dis tribution. Geographers are increasingly beginning to explore such shifts in consumption relations within a broad 'new retail geography' perspective.

Future Agendas

Fifteen years on from the initial development of 'the new retail geography' the subject continues to offer some of the most challenging but rewarding areas of study in human geography. In particular, the increasingly obvious rise of retail power in national and global markets and the consequences of that rise are prompting new agendas which are focusing the attention of both geographers and social scientists more widely. Three interrelated issues stand out.

1. The rise of retail has provoked increasingly wide spread attempts to 'rein in' the power and dominance of the leading corporate retailers. In Western econ omies, market dominant retailers have faced both nationally organized campaigns (e.g., the Tescopoly campaign in the UK) and, at the local level, increas ingly vocal and effective community resistance to their efforts to 'prise open' which Marsden *et al.* refer to as new territories and terrains of profit making. In the USA, in particular, a so called 'war on Wal Mart' has erupted in the wake of attempts by the corporate giant to enter the major urban markets, and that 'war'

has been bitterly fought out in Los Angeles, Chicago, and New York. Similarly in emerging economies which have experienced significant levels of retail FDI (Southeast Asia provides an example), attempts have been made to curb the power of the increasingly market dominant transnational retailers either by re regulating previously liberalized retail FDI regimes, or by tightening land use zoning, trading hours, and/or selling practice regulation.

2. In Western economies, in response to such pressures, the more innovative and agile of the corporate retailers have become increasingly sensitive to the fact that both their future customer mandate and corporate prospects rest on moving rapidly to embrace agendas of community responsiveness, sustainable development, urban re generation, responsible/ethical sourcing, etc., increas ingly demanded by their customers.
3. In particular, and interrelated to the above, the 'lead firm' status of the largest Western corporate retailers has placed them into a position where they have little commercial alternative other than to address the wider environmental challenges faced by the society/com munities from which they draw their mandate. As a result, it is the major retailers in several Western economies who are now at the forefront of promoting a 'sustainable consumption' agenda. However, what the implications and consequences are of retailers assuming such a role, and the extent to which that agenda might translate into and across the developing economy operations of the largest retail TNCs are as yet unexplored.

It is clear, however, that together these issues represent a fascinating nexus – the study of which will allow geog raphers to make an important contribution to both aca demic and public policy debate over the next decade. It also confirms what Wrigley and Lowe suggested at the time of the initial development of the 'new retail geography', namely, "the central importance, both to the broader pro gress of critical human geography and to the wider study of circulation and consumption activities in the social sci ences, [of having] a vibrant reconstructed retail geography positioned at the cutting edge of theoretical debate."

See also: e-Business and e-Commerce; Embeddedness; Firms; Food Networks; Food Networks, Alternative; Foreign Direct Investment; Global Production Networks; Globalization and Transnational Corporations; Regulation.

Further Reading

Berry, B. J. L. (1963). *Commercial Structure and Commercial Blight*, Univ. of Chicago, Dept. of Geography. Research Paper 85.

Blomley, N. K. (1996). 'I'd like to dress her all over': Masculinity, power and retail capital. In Wrigley, N. & Lowe, M. S. (eds.) *Retailing, Consumption and Capital: Towards the New Retail Geography*. London: Addison Wesley Longman.

Brunn, S. (ed.) *Wal Mart World: The World's Biggest Corporation in the Global Economy.* New York: Routledge.

Coe, N. and Hess, M. (2005). The internationalization of retailing: Implications for supply network restructuring in East Asia and Eastern Europe. *Journal of Economic Geography* 5, 449 473.

Coe, N. and Wrigley, N. (2007). Host economy impacts of transnational retail: The research agenda. *Journal of Economic Geography* 7, 341 371.

Crewe, L. and Davenport, E. (1992). The puppet show: Changing buyer supplier relationships within clothing retailing. *Transactions of the Institute of British Geographers* NS17, 183 197.

Cummins, S., Findlay, A., Petticrew, M. and Sparks, L. (2005). Healthy cities: The impact of food retail led regeneration on food access, choice and retail structure. *Built Environment* 31, 288 301.

Currah, A. D. (2003). The virtual geographies of retail display. *Journal of Consumer Culture* 3, 5 37.

Currah, A. D. (2006). Hollywood versus the Internet: The media and entertainment industries in a digital and networked economy. *Journal of Economic Geography* 6, 439 468.

Currah, A. D. and Wrigley, N. (2004). Networks of organizational learning and adaptation in retail TNCs. *Global Networks* 4, 1 23.

Doel, C. (1996). Market development and organizational change: The case of the food industry. In Wrigley, N. & Lowe, M. S. (eds.) *Retailing, Consumption and Capital: Towards the New Retail Geography*, pp 116 136. Harlow: Addison Wesley Longman.

Ducatel, K. and Blomley, N. K. (1990). Rethinking retail capital. *International Journal of Urban and Regional Research* 14, 207 227.

Friedberg, S. (2003). Cleaning up down south: Supermarkets, ethical trade and African horticulture. *Social and Cultural Geography* 4, 27 43.

Frieden, B. J. and Sagalyn, L. B. (1989). *Downtown Inc: How America Rebuilds Cities*. Cambridge, MA: MIT Press.

Gereffi, G. (1994). The organization of buyer driven global community chains. In Gereffi, G. & Korzeniewicz, M. (eds.) *Commodity Chains and Global Capitalism*. Westpoint, CT: Greenwood Press.

Gereffi, G. (2001). Beyond the producer driven/buyer driven dichotomy: The evolution of global value chains in the Internet era. *International Development Studies Bulletin* 32, 30 40.

Guy, C. M. (2006). *Planning for Retail Development*. London: Spon Press.

Harvey, D. (1985). *The Urbanization of Capital*. Baltimore: John Hopkins University Press.

Hughes, A. L. (2000). Retailers, knowledges and changing commodity networks: The case of the cut flower trade. *Geoforum* 31, 175 190.

Hughes, A. L. (2005). Corporate strategy and the management of ethical trade: The case of the UK food and clothing retailers. *Environment and Planning A* 37, 1145 1163.

Hughes, A. L. and Reimer, S. (eds.) (2004). *Geographies of Commodity Chains*. London: Routledge.

Humphrey, J. (2007). The supermarket revolution in developing countries: Tidal wave or tough competitive struggle. *Journal of Economic Geography* 7, 433 450.

Marsden, T. K., Harrison, M. and Flynn, A. (1998). Creating competitive space: Exploring the social and political maintenance of retail power. *Environment and Planning A* 30, 481 498.

Marsden, T. K. and Wrigley, N. (1995). Regulation, retailing and consumption. *Environment and Planning A* 27, 1899 1912.

Miller, D. (ed.) (1995). *Acknowledging Consumption: A Review of New Studies*. London: Routledge.

Murray, R. (1989). Fordism and post Fordism. In Hall, S. & Jaques, M. (eds.) *New Times: The Changing Face of Politics in the 1990s*, pp 38 53. London: Lawrence and Wishart.

Reardon, T., Henson, S. and Berdegué, J. (2007). Proactive fast tracking diffusion of supermarkets in developing countries: Implications for market institutions and trade. *Journal of Economic Geography* 7, 399 431.

Schoenberger, E. (1994). Corporate strategy and corporate strategists: Power, identity and knowledge within the firm. *Environment and Planning A* 26, 435 451.

Weltevreden, J. (2007). Substitution or complementarity: How the Internet changes city centre shopping. *Journal of Retailing and Consumer Services* 14, 192 207.

Wrigley, N. (1992). Antitrust regulation and the restructuring of grocery retailing in Britain and the USA. *Environment and Planning A* 32, 1977 1989.

Wrigley, N. (2002). Food deserts in British cities: Policy context and research priorities. *Urban Studies* 39, 2029 2040.

Wrigley, N., Coe, N. M. and Currah, A. D. (2005). Globalizing retail: Conceptualizing the distribution based transnational corporation (TNC). *Progress in Human Geography* 29, 4437 4457.

Wrigley, N., Guy, C. M. and Lowe, M. S. (2002). Urban regeneration, social inclusion and large store development: The Seacroft development in context. *Urban Studies* 39, 2101 2114.

Wrigley, N. and Lowe, M. S. (eds.) (1996). *Retailing, Consumption and Capital: Towards the New Retail Geography*. Harlow: Addison Wesley Longman.

Wrigley, N. and Lowe, M. S. (2002). *Reading Retail: A Geographical Perspective on Retailing and Consumption Spaces*. London: Arnold.

Wrigley, N., Lowe, M. S. and Currah, A. D. (2002). Retailing and e tailing. *Urban Geography* 23, 180 197.

Wrigley, N., Warm, D. L. and Margetts, B. M. (2003). Deprivation, diet and food retail access: Findings from the Leeds 'food deserts' study. *Environment and Planning A* 35, 151 188.

Rio Summit*

E. S. Norman, University of British Columbia, Vancouver, BC, Canada
D. Carr, University of California – Santa Barbara, Santa Barbara, CA, USA

Glossary

Civil Society The sphere or space between individuals and the state and/or market.

Civil Society Organizations Voluntary civic and social organizations as differentiated from state and commercial organizations and institutions.

Destatization A major outcome of globalization whereby civil society and regional authority develop as state power wanes.

Glocalization A process of rescaling in which a simultaneous shift occurs upward toward the supranational scale (e.g., North American Free Trade Agreement and European Union) and downward toward the local or global, products and services tailored to local places.

Governance The processes and systems through which organizations and society operate.

Sustainable Development "Development that meets the needs of the present without compromising the ability of future generations to meet their own needs" – World Commission on Environment and Development (1987).

Introduction

The United Nations Conference on Environment and Development (UNCED), widely known as the Rio Earth Summit, convened 3–14 June 1992 in Rio de Janeiro, Brazil. The Rio Summit focused on developing a global framework for addressing environmental degradation through sustainable development. Through the participation of both state and nonstate actors, the main themes and agendas of the Rio Summit were condensed into several documents and institutional mechanisms. The documents provided guidance for communities worldwide who desired to integrate sustainable development goals into their governance structure. The main documents produced at the Summit include Agenda 21, the Rio Declaration on Environment and Development, the Statement of Forest Principles, the United Nations Framework Convention on Climate Change, and the United Nations Convention on Biological Diversity. The Summit hosted a record 172 108 governmental officials, 2400 representatives of nongovernmental organizations, and 17 000 attendees at a parallel nongovernmental organization (NGO) forum. With an estimated 10 000 journalists on site, the widely reported Summit was heard by millions of people around the world.

Global Events on Environment and Development Pre-Rio

The 1972 Stockholm Conference on the Human Environment was the predecessor to the Rio Earth Summit and one of the first mega conferences to tackle questions of environment and development at a global scale. One of main outcomes of the Stockholm Conference was the Declaration of the United Nations Conference on the Human Environment, which outlines a series of principles and proclamations that aims to both "inspire and guide the peoples of the world in the preservation and enhancement of the human environment" (United Nations Environment Programme 2006). Although the declaration clearly recognizes the sovereignty of each nation, stated under principle 21, it also strongly encourages each country to cooperate at an international scale to help protect and enhance the human environment. Finding the balance between state sovereignty and transnational responsibility was crucial to the success of the Stockholm talks and continues to play a fundamental role in negotiating global environmental governance regimes.

Although the Stockholm Conference addressed myriad topics, one of the central debates that transcended the meetings centered on development and poverty. The question of whether development and economic growth are inherently destructive to the environment created sharp divisions between pro growth advocates and those concerned about the negative consequences of unchecked economic growth. These debates, in many cases, accentuated the polarization of governments from the Global North and South, and more generally, between pro growth governments and NGOs advocating for limits to growth.

More than a decade later, the Commission on Environment and Development was established to reflect on the escalating environmental crisis and devise a global agenda for change. The 1984 commission, also known as the Brundtland Commission after the chair, Norwegian

*Some material included in this chapter is reprinted from *Geoforum*, Vol 39(1), D.L. Carr and E.S. Norman "Global Civil Society? The Johannesburg World Summit on Sustainable Development," 358 371, 2008, with permission from Elsevier.

Prime Minister Gro Harlam Brundtland, produced the widely received and provocative document *Our Common Future*, which decried the world's failure to achieve sus tainable development and outlined several far reaching actions that needed to occur to mitigate anticipated en vironmental disaster. This document was instrumental for laying the foundation for the topics to be discussed at the Rio Summit and for shaking up the global com munity to convene another global summit.

On 22 December 1989, the United Nations General Assembly called for a global meeting that would address the issues raised in the Brundtland Report. The aim of the meeting was to formulate strategies to stop and re verse the effects of global environmental degradation "in the context of increased national and international efforts to promote sustainable and environmentally sound de velopment in all countries." This meeting would come to be known as the Rio Earth Summit.

The Rio Summit, Agenda 21, and the Millennium Development Goals

The Rio Summit was unlike other UN conferences, in terms of its large scale, public nature, and wide ranging topics. The Summit hosted state leaders from across the globe (representing 172 governments and 108 heads of state) as well as 2400 NGO representatives. The Summit attendees collaborated on innovative ways to address pressing environmental concerns and largely focused on the need for broad based, environmentally focused sus tainable development.

The Rio Summit set a precedent for greater civic engagement by allowing increased participation from NGOs and nonstate actors. The increased presence of civil society in both the pre conferences and the Summit itself was indicative of a widening in governance struc tures in the global arena, which continues today. The greater participation of NGOs also left an indelible mark on the proceedings and accords – the legacy of the Summit. The immense number of participants and journalists at the Summit further launched the concept of sustainable development into the public discourse and sparked the world's interest in environment and devel opment issues.

The Rio Summit also set the agenda for all subsequent UN conferences examining the interconnections between human rights, population, health, social and economic development, and environmental sustainability. Regional and international development policies were explicitly developed with consideration of environmental and economic impacts. These discussions led to critical examinations of the environmental and health impacts of common household agents such as gasoline and pesti cides. Similarly, the Summit focused awareness on the increasing scarcity of freshwater and diminishing fishing resources and explored ways to replace fossil fuels with alternative energies such as wind, solar, and water.

Attempting to find economic based solutions to en vironmental problems gave rise to other issues at the Summit, including global equity, governance, and North–South relationships. In particular, North–South rela tionships weighed heavily on discussions of poverty eradication, access to resources, and global aid. Some argue, for example, that the development policies in which the northern 'industrialized' countries are pos itioned as the grantors and providers of technical ex pertise and the southern 'less developed countries' are positioned as the grantees and recipients of information reinforce neocolonial relationships. These relationships, although often developed in an attempt to 'level the playing field' between the Global North and the Global South, arguably reify northern privilege and silence southern voices. Throughout the last 30 years, the UN and other development agencies have worked toward developing a more equal global representation in devel opment. However, despite these efforts, the Global South continues to be a minority in global forums (as discussed below in 'progress since Rio').

Agenda 21

The Rio Summit was successful at moving beyond mere discussions of development and the environment. As a direct result of the Summit, three major agreements (ratified by 108 governments) and two legally binding conventions substantiated the talks. The agreements in cluded Agenda 21, the Rio Declaration on Environment and Development, and the Statement of Forest Prin ciples. The conventions included the UN Framework Convention on Climate Change and the Convention on Biological Diversity. As a way to support the imple mentation of these wide ranging sustainable develop ment goals and commit to the discussions of the Summit, the UN created three institutional bodies the UN Commission on Sustainable Development, the Inter agency Committee on Sustainable Development, and the High level Advisory Board on Sustainable Development.

The most widely cited of the documents produced at the Summit is Agenda 21's Declaration on Sustainable Development, which builds on the 1972 Stockholm Declaration and reaffirms the commitment to environ ment through multiscalar cooperation (international, national, local, civil society). Agenda 21, adopted on the last day of the Rio Summit (14 June 1992), is a comprehensive program of action implemented by gov ernments, civil society, development agencies, UN organizations, and independent sector groups where human and economic activity affects the environ ment. The premise behind Agenda 21 is that wider socioeconomic issues need to be addressed to stave off environmental degradation. The document broadens the charge of ecosystem protection to wider issues including poverty, hunger, sickness, and illiteracy. Further, it promotes a coordinated and far reaching approach to environmental management, as stated in the preamble to Agenda 21: "No nation can achieve this on its own. Together we can – in a global partnership for sustainable development."

Although the wider agenda of the document is designed for international collaboration, the UN Commission on Sustainable Development (UNCSD) encourages each country to identify individual strategies to implement goals which will work within their nation state framework. These individual goals, known as the National Sustainable Development Strategies (NSDSs), clarify priorities and sustainable development efforts at a national level. Furthermore, within each country, the responsibilities of translating Agenda 21 to the local level has manifested into a program called Local Agenda 21 (LA21). LA21 recognizes that local communities are a major group that can implement a global planning process for sustainability through their own programs and initiatives. For example, Chapter 28 of Agenda 21 identifies the importance of local governments in fulfilling sustainable development objectives:

> Local authorities construct, operate and maintain economic, social and environmental infrastructure, oversee planning processes, establish local environmental policies and regulations, and assist in implementing national and sub national environmental policies. As the level of governance closest to the people, they play a vital role in educating, mobilising and responding to the public to promote sustainable development. (Agenda 21, Chapter 28.1)

Simply, LA21 attempts to embody the adage 'think globally, act locally'. This downshifting or devolution of responsibilities to the local level has been met with varied results, however. For example, active LA21 programs are found in countries such as Australia and Japan, but other countries such as China, have been reticent to adopt such plans. Although LA21 is a tool to help implement the goals and principles outlined at the Rio Summit, and then later reinforced at the Johannesburg Summit (as Action 21), it is by no means a quick fix to complicated international environmental governance issues.

Millennium Development Goals

Other mechanisms aimed at implementing the goals outlined in Agenda 21 include the Program for the Further Implementation of Agenda 21 (1997), the United Nations Millennium Development Goals (MDGs) (also known as the International Development Goals) (2000), and the Johannesburg Plan of Implementation (2002). The MDGs are a compiled version of agenda items identified by international conferences and summits held throughout the 1990s – including Rio's Agenda 21. The goals were set as part of the UN's Millennium Conference held in New York City, which aimed at galvanizing efforts to meet the needs of the world's poorest people. The MDGs provide a common framework for measuring development progress and encourage those participating within the UN system to work coherently together. The goals are organized into eight categories and include several quantitative indicators to be achieved by the target date of 2015. The goals include (1) eradicate extreme poverty and hunger; (2) achieve universal primary education; (3) promote gender equality and empower women; (4) reduce child mortality; (5) improve maternal health; combat HIV/AIDS, malaria, and other diseases; (7) ensure environmental sustainability; and (8) develop a global partnership for development.

Progress since Rio

There has been some progress of note since Rio in achieving the Agenda 21 plan of action and the MDGs. The percentage of the world's grinding poor fell from one quarter to one fifth of the world's population. Although there are more than 2 billion nutrition deficient people, the world faces minimal food scarcity. And the sometimes recalcitrant US administration recently announced a 50% (US$5 billion) increase in US foreign aid spending over 3 years.

However, there is plenty of bad news regarding progress since Rio in environmental and human indicators. Population grew from approximately 5 billion to 6 billion people as the number of people who earn less than US$2 per day increased by tens of millions. Carbon emissions have increased nearly 10% (nearly 20% in US). HIV has exploded in sub Saharan Africa (20–40% infection rate in southern African nations). And increasingly, uneven development is implicated in desertification, the plummeting of fish stocks, widespread drought and salinization, and wildlife species disappearing at unprecedented rates, much of these due to tropical deforestation claiming approximately 200 000 square kilometers every year. Yet the political commitment of the world's wealthiest nations, particularly the US, to reversing these processes remains far from matching the cost they exact on society. The US' proposed aid increase brings its foreign aid package to not even one fifth the 1992 pledge to earmark 0.7% of annual gross domestic product (GDP) to foreign assistance. This amounts to less than 1 billion dollars a month; a meager sum next to a US military budget in excess of 1 billion dollars per day. At the Rio Earth Summit 10 years prior, the world's richest nations committed to halve poverty by 2007, to eradicate hunger, reduce under five mortality by two thirds, and to enroll all children in school. The UN estimates the cost of achieving these goals at between 40–60 billion dollars over the amount that would be produced from the 0.7% commitment – merely several weeks worth of US military spending or one sixth of what the West spends to subsidize its farmers. In an attempt to rectify some of these problems, and lack of international commitment in solving them, the UN convened a summit 10 years after Rio, the 2002 Johannesburg World Summit on Sustainable Development (WSSD).

The WSSD, which convened from 26 August 2002 to 4 September 2002 in Johannesburg, South Africa, aimed at assessing developments since the Rio Earth Summit and reinforcing multilateral commitments to sustainable development. More than 20 000 participants, from governmental (representing 180 countries) and NGOs, the private sector, and the scientific community addressed increasing challenges in environmental degradation and sustainable development. The large number of unmet accords inherited from the 1992 Rio Summit – due to flaccid commitment by the wealthiest nations to mobilize around these concerns and in no small part to the US government's failure to participate in key global treaties (such as the Kyoto Accord) – called for significant structural changes at the 2002 Johannesburg Summit. One change was to revisit the accords some changes include halving the proportion of people that lack access to clean water or proper sanitation and restoring depleted fisheries by 2015. In reassessing progress in the MDGs and Agenda 21 since Rio, Summit organizers also decided to strengthen collaborations beyond government. Sectors explicitly included were business, women and children, trade unions, indigenous, local nongovernmental groups, farmers, and the scientific community. Most saliently, the organizers included a parallel event to strengthen civil society participation, the Stakeholders Forum Implementation Conference.

Global Environmental Governance

Civil Society Representation

The number of NGOs participating in UN sponsored events has increased drastically since Stockholm. Fewer than 300 NGOs participated in the 1972 Stockholm Conference on the Environment. Twenty years later at the Rio Summit, 2400 NGOs were officially registered and 17 000 representatives attended a parallel summit specifically for NGO participants. Estimates suggest that since Rio, a growing tens of thousands of NGOs participate in UN conferences through lobbying, policy formulation, and monitoring/implementing international accords.

Despite the rapid growth in the number of NGOs participating in UN conferences, a long road remains to attain a truly global civil society. Civil representation has remained unbalanced geographically, with most donor money consolidated among northern NGOs. At Rio, despite the ample participation, NGOs were considered ancillary collaborators to states and largely remained distant from the central politics shaping Agenda 21. Similarly, in Johannesburg, the main conference remained closed and the heightened security and hour's drive between the NGO and government sectors compromised the ability for civil society to fully participate in the UN Summit. Furthermore, for both NGOs and states, differences between developed and developing economies, and wealthy and poor remain primary divisions claimed by stakeholders in UN conferences. Myriad and fragmented interests continue to balkanize civil society organization (CSO) efforts as they operate at UN conferences. States still dominate agendas. And governments inconsistently welcome CSO participation. These inequities, though abated since Stockholm, are still significant factors in CSO participation, representation, and voice within global environmental governance.

Despite these persistent shortcomings, the recent swell in CSO involvement at global decision making forums, including UN conferences, is rich with implications for shifting dynamics in global political processes and representation. For example, many argue that the roles and goals of NGOs have significantly expanded. This increase in capacity makes them more significant players in the global governance arena. Furthermore, with large international institutions such as The World Bank and the UN promoting economic and human development through CSOs, important questions emerge regarding representation of local interests and implications for 'trickle up' decision making.

Civil Society at the WSSD: The Stakeholders Forum Implementation Conference

A significant example of increased capacity of civil society and 'trickle up' decision making is the UN sponsored Implementation Conference: Stakeholder Action for Our Common Future (IC), which immediately preceded the WSSD. The IC represented a new approach to civic engagement and was designed to mobilize stakeholder participation and facilitate the implementation of commitments established in Rio and embodied in Agenda 21. The organizers of the conference, the UK based Stakeholders Forum for Our Common Future, hoped that integrating civil society at an international level might reverse disappointing trends since Rio. By including stakeholders in the planning and implementations process, the organizers also hoped for a better record for achieving proposed international accords. The IC met in Johannes burg for three days subsequent to the WSSD, with 331 representatives from 50 countries, and had four main theme groups and 24 subgroups.

A central idea of the IC was the belief that partner ships and stakeholders can create solutions regardless of the level of involvement of individual state governments. The IC was designed as a forum for stakeholders to identify, prioritize, and commit to tangible action plans. The IC, as defined by the organizers, was about "action, not about lobbying governments." By providing a phys ical space for civilian organizations to dialog, the IC sought to foster concrete partnerships that transcended the Johannesburg meetings. Specifically, the IC aimed to (1) bring global attention to the burgeoning stakeholder movement and (2) influence the decision making at the WSSD and the language of WSSD documents.

A novel approach at the WSSD was the concept of 'Type II' agreements that brought public, private, and civil actors in partnership. These agreements were de veloped by delegates at the IC and resulted in formal partnerships between NGOs, intergovernmental organ izations, private companies, and scientific institutions. The Type II agreements aimed at enabling stakeholders to advance concrete contributions toward the official outcomes of the WSSD covenant. This new form of governance widens the scope of participation to include both private and public sectors, from multiple scales of governance, to align themselves with international en vironmental and developmental organizations. Whether this increasingly neoliberal governance structure is in dicative of an increased social responsibility in the business sector, or a co opting of nongovernmental and governmental organizations to conform to business elites, is open to debate.

Although this was a laudable first attempt at engaging civil society at a global scale and widening the dialog for stakeholders at mega international conferences, struc tural changes need to occur if the IC and other sub sequent fora are to meet their goals. In order for civil society representation to be truly strengthened, and the IC to achieve its stated goals, structural changes need to occur. For example, the participant representation at the conference failed to faithfully reflect the geographic di versity that was expected at a global civil society event. The IC drew a large number of participants from North America and Europe; however, representations from Asia, the Middle East, and South America were largely underrepresented (**Figure 1**). Large disparities also oc curred within Africa. Unsurprisingly, South Africa was very well represented; however, Western and Central Africa had sparse representation. Limited institutional capacity within these southern regions poses challenges for future conferences where equitable representation is desired (Carr and Norman, 2008).

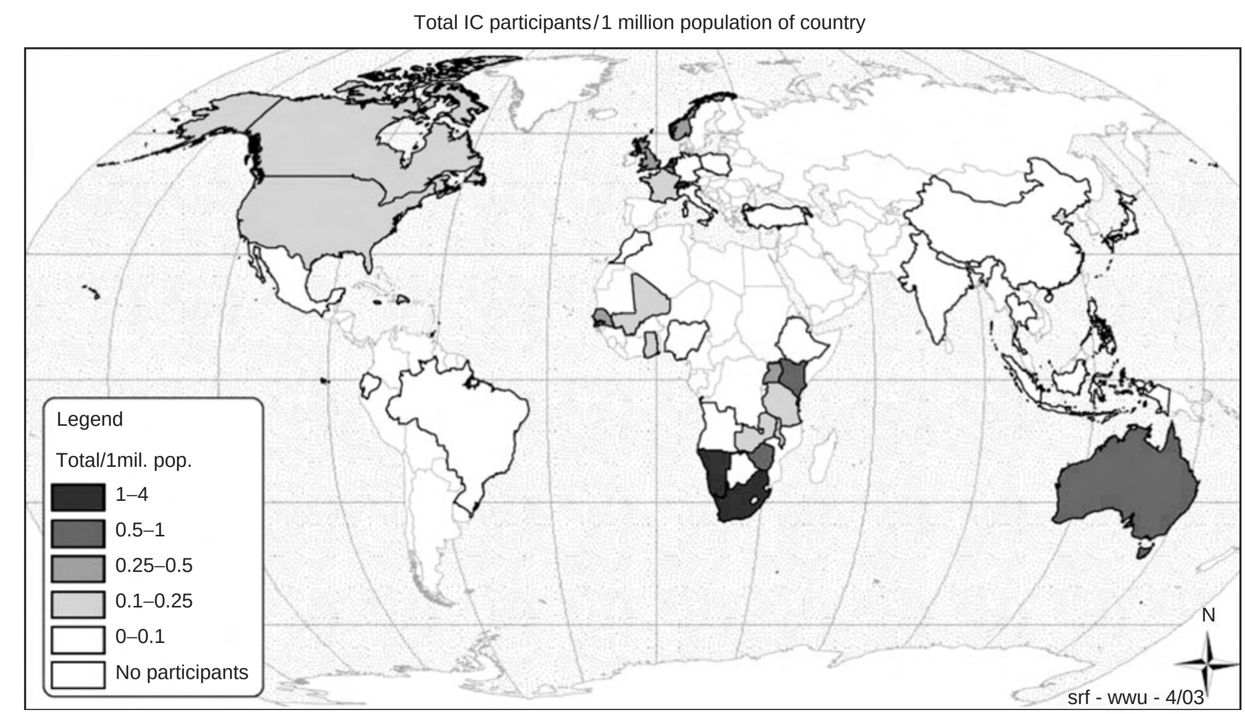

Figure 1 Implementation conference participation based on country population. Reproduced from Carr, D. and Norman, E. (2008). Global civil society? The Johannesburg World Summit on Sustainable Development. *Geoforum* 39 (1), 358 371, with permission from Elsevier.

The increase of nonstate actors and CSOs in the global environmental community points to significant changes within the world governance structure. Scholars have pointed to trends of 'hollowing out of the state' in which the functions of government are transferred to other parties through the rescaling of the state to non state institutions (e.g., private sector, NGOs, voluntary organizations). Similarly, these shifts in governance are sometimes referred to as 'glocalization', a simultaneous shift in power upward toward the supranational and downward toward to the local. However, an increase in governance does not necessitate the loss of government power as if it were a zero sum game. Despite the increase in nonstate actors in decision making and political in fluence, governance is still considered 'in the shadow of government' with the state as a key sphere of influence.

Thus, the changes in global environmental governance structures, as exemplified by the trends in UN environ mental summits to include nonstate actors, is not ne cessarily indicative of a weakening of state sovereignty as originally predicated by former secretary general of the UN Boutros Boutros Ghali when he called for a re structuring of the UN. Nor does this new structure lead to a borderless world, where globalized economies erase nation state delineations as famously described by Jap anese businessman Kenichi Ohmae in the *The End of a Nation State*. Rather, this change in global environmental structure should be seen as a 'widening' of the nation state to include greater participation of civil society, a trend that continues to play out in the global environ mental arena.

See also: Aid; Civil Society; Development I; Development II; Environment; Governance; International Organizations; Nongovernmental Organizations; Sustainable Development.

Further Reading

Carr, D. L. (2005). U.S. foreign policy and a geographically challenged U.S. citizenry. *Journal of Geography* 104(1), 41 42.

Conca, K. and Dabelko, G. (eds.) (1998). *Green Planet Blues: Environmental Politics from Stockholm to Kyoto*. Boulder: Westview Press.

Food and Agricultural Organization of the UN (FAO) (2001). State of the world's forests. *Report of FAO*. New York: Department of Forestry, United Nations Secretariat.

Haas, P. M., Keohane, R. O. and Levy, M. A. (1991). *Institutions for the Earth: Sources of Effective International Environmental Protection*. Cambridge, MA: The MIT Press.

Herz, S. and Ebrahim, A. (2005). A call for participatory decision making. Discussion Paper on World Bank Civil Society Engagement. Johannesburg: Civicus.

Morphet, S. (1996). NGOs and the environment. In Willetts, P. (ed.) *The Conscience of the World: The Influence of Non governmental Organisations in the U.N. System*, pp 116 146. Washington, DC: Brookings Institutions.

Ohmae, K. (1995). *The End of the Nation State: The Rise of Regional Economies*. New York: The Free Press.

Ohmae, K. (2002). *The Borderless World: Power and Strategy in the Global Marketplace*. London: Profile Books.

Pfeffer, J. and Salancik, G. (1978). *The External Control of Organizations: A Resource Dependence Perspective*. New York: Harper and Row.

Swyngedouw, E. (1997). Neither global nor local: 'Glocalization' and the politics of scale. In Cox, K. (ed.) *Spaces of Globalization: Reasserting the Power of the Local*, pp 115 136. New York: Guilford Press.

Swyngedouw, E. (2000). Authoritarian governance, power, and the politics of rescaling. *Environment and Planning D: Society and Space* 18, 63 76.

Tamiotti, L. and Finger, M. (2001). Environmental organizations: Changing roles and functions in global politics. *Global Environmental Politics* 1(1), 56 76.

Taylor, R. (ed.) (2004). *Creating a Better World: Interpreting Global Civil Society*. Bloomfield, CT: Kumarian Press.

Van Rooy, A. (1997). The frontiers of influence: NGO lobbying at the 1974 world food conference, the 1992 earth summit and beyond. *World Development* 25(1), 93 114.

Van Rooy, A. (2004). *The Global Legitimacy Game: Civil Society, Globalization and Protest*. London: Palgrave.

World Commission on Environment and Development (1987). *Our Common Future*. Oxford: Oxford University Press.

Relevant Websites

http://www.globalpolicy.org
Global Policy Forum: Social and Economic Policies.

http://www.un.org
United Nations Conference on Environment and Development.

http://www.unep.org
United Nations Environmental Program: Declaration of the United Nations Conference on the Human Environment.

http://www.un.org
United Nations Summit.

http://www.un.org
United Nations Rio Declaration on Environment and Development.

Ritter, C.

M. Doevenspeck, University of Bayreuth, Bayreuth, Germany

Glossary

Erdkunde It is the science of Earth, a literal German translation for the word geography.

Comparative Geography It pursues the goal of a geographical typology by a generalizing and comparative search for laws.

Compendium Geography The prescientific stage of geography lasting until the early 1800s, a mere description and summary of facts.

Geographical Providential Device Transfer of the idea of providence in geography, for example, interpretation of relief molding as God's plan.

Inductive Reasoning Moving from specific observations to broader generalizations and theories.

Ritter, Carl (1779–1859)

The German geographer Carl Ritter (**Figure 1**) was born in Quedlinburg (then Prussia). As one of six children whose father died when he was 2 years old, Ritter enjoyed an unorthodox education by his teachers who tried to practice new teaching methods based on the educational ideas of Jean Jacques Rousseau and inspired by the Enlightenment: being responsive to the nature of children instead of learning by rote and searching a God in nature instead of being paralyzed by church doctrines. Since his religious mother had a great influence on him, education and Christianity remained decisive for Ritter and his work.

On finishing school, with financial support from the banker Bethmann Hollweg, Ritter attended university in Halle where he studied under *inter alia* Matthias Christian Sprengel, a specialist on India and an early critic of colonialism and slavery. In return for this support Ritter worked as tutor in Bethmann Hollweg's house in Frankfurt until 1813 using his own teaching materials. These indicate his place as the leading teacher in geography of his time: they included thematic maps, relief models, and the interpretations of his data. During this time Ritter made several journeys to Switzerland and Italy where he conducted barometer, thermometer, and hygrometer measurements as well as completed panoramic drawings. He also met the Swiss educational reformer Johann Heinrich Pestalozzi several times, who was to have a great influence on his work. In 1807, the year he met von Humboldt for the first time, Ritter published the second volume of his *Geography of Europe*, a work which is not very well known today, but which went far beyond the then conventional compendium geography. In 1813 Ritter accompanied his students to Göttingen, where he lived for 6 years. He studied geography, history, philosophy, botany, chemistry, and mineralogy and worked on an introduction and the first two volumes of his major work *Die Erdkunde*, (literally, 'Earth science') which were published in 1817 and 1818. After several months as a grammar school teacher in Frankfurt, he became a professor of history at the Friedrich Wilhelms University, and then the first chair of geography in Germany at the military academy in Berlin in 1820. He stayed for nearly 40 years in Berlin teaching geography and establishing the Berlin Geographical Society as one of the leading geographical associations at that time. Through his students, such as Elisée Reclus, the important French geographer and professor of comparative geography in Brussels, Arnauld Guyot, who became professor of physical geography and geology at Princeton (then the College of New Jersey) and his

Figure 1 Carl Ritter, a painting by Carl Begas, 1844. Reprinted from Staatsbibliothek Preußischer Kulturbesitz (ed.) (1979). Carl Ritter in Seiner Zeit. *Ausstellungskataloge* 11. Berlin.

admirers, such as Paul Vidal de La Blache and Peter Kropotkin, he strongly influenced geography in France, Russia, and North America.

Since Ritter was highly religious; he may be described as a Christian geographer. For him, nature was divinely arranged by God and as part of this arrangement, the Earth was the setting of divine revelation, created by God as the home of man. Against the prevailing secularized *zeitgeist*, Ritter appears as one of the last to hold on to ideas of divine providence. Even if his religiosity influenced his scholarship, it did not bar anarchists like Réclus or Kropotkin from following him in his conception of geography. Ritter attached great importance to teaching a 'new scientific geography' that distances itself from the traditional summation of facts about states and cities. In contrast to the simple description and inventory of the objects on Earth's surface, he tried to understand the interdependencies between these objects and, even if he formulated very few laws himself, he promoted inductive reasoning. Like von Humboldt, he recognized the complexity of nature and the interconnections and interrelations in it, but differed from von Humboldt by looking at the same place in different historical episodes rather than comparing different places at the same time. The consideration of history was an important *leitmotiv* for Ritter. After his early classification of physical and economic geography in 1806, he came to the conviction that analyzing the Earth's inhabitants means to consider their particular history. Subsequently, the strong tie to economics no longer satisfied him and he favored the term historic in order to emphasize geography's link to man, who stood in the center of his studies. For Ritter, the Earth's surface was therefore best conceived of as a dynamic sequence of historical–geographical layers.

Ritter worked on many of the geographical subjects that still matter today, including, for example, migration models and research on large conurbations (e.g., Jerusalem) and river catchments, always accompanied by an impressive cartography. Ritter's regional approach was picked up in the early 1900s by the American geographer, Richard Hartshorne, who refined the method and named it 'areal differentiation'. Between 1824 and 1852, Ritter traveled all over Europe; but, unlike von Humboldt, he did not leave the continent, relying instead on reports from European discoverers, researchers, missionaries, and merchants. He also used the available Arabic, Asiatic, and Chinese literature.

The full title of his main work *Die Erdkunde* reads *The Science of the Earth in Relation to Nature and the History of Mankind; or, General Comparative Geography as the Solid Foundation of the Study of, and Instruction in, the Physical and Historical Sciences*. It again reveals the importance Ritter gave to history. To this day, Die Erdkunde is the largest work of a single author in geographical literature and even more comprehensive than Réclus' *Nouvelle Géographie Universelle*. But if one dares to evaluate this work especially in respect of its claim to follow a comparative approach, it is important to note that it remained fragmentary. Of the originally intended three parts that were outlined in a never published handbook, only the first, with the object of dealing with the continents ('the solid form or the continents') was published. Nonetheless, this ran to some 19 volumes with over 20 000 pages between 1817 and 1859, and even that part covers only Africa and Asia. Ritter also never finished the second part on hydrography, meteorology, and volcanology ('the liquid form or the elements'), and the third on mineralogy, botany, and zoology ('the bodies of the three empires of nature'). What is particularly evident here is that mankind was not to be treated especially but was regarded as a 'vital mirror of nature' – the guiding idea of all three parts.

Like his contemporary von Humboldt, Ritter is thus acknowledged as one of the discipline's founding fathers, with his comparative approach and attempt to create geography as an integrative science having legacies which still continue to unfold.

See also: Christian Geography; German-Language Geography; Kropotkin, P.; National Schools of Geography; Reclus, E.; Vidal de la Blache, P.; von Humboldt, A.

Further Reading

Beck, H. (1979). *Carl Ritter, Genius of Geography: On His Life and Work*. Bonn Bad Godesberg: Inter Nationes.

Beck, H. (1982). *Große Geographen. Pioniere Außenseiter Gelehrte*. Berlin: Dietrich Reimer.

Lenz, K. (ed.) (1979). Carl Ritter *Geltung und Deutung. Beiträge des Symposiums anlässlich der Wiederkehr des 200. Geburtstages von Carl Ritter*, November 1979 in Berlin (West). Berlin: Dietrich Reimer.

Martin, G. J. (2005). *All Possible Worlds: A History of Geographical Ideas*. Oxford: Oxford University Press.

Plewe, E. (1978). *Die Carl Ritter Bibliothek. Erdkundliches Wissen 50*. Wiesbaden: Steiner.

Richter, H. (ed.) (1983). *Carl Ritter. Werk und Wirkungen. Beiträge eines Symposiums im 200. Geburtsjahr des Gelehrten, Quedlinburg, DDR. Wissenschaftliche Abhandlungen der Geographischen Gesellschaft der Deutschen Demokratischen Republik 16. Gotha: VEB Hermann Haack*.

Ritter, C. (1822). *Die Erdkunde im Verhältnis zur Natur und zur Geschichte des Menschen oder allgemeine vergleichende Geographie als sichere Grundlage des Studiums und Unterrichts in physikalischen und historischen Wissenschaften. Erster Theil, Erstes Buch. Afrika*. Berlin: Reimer.

Ritter, C. (1883). *Geographical Studies*. Boston, MA: Gould and Lincoln.

Ritter, C. (2007). *Comparative Geography*. Montana: Kessinger Publishing.

Schmitthenner, H. (1951). *Studien Über Carl Ritter. Frankfurter Geographische Hefte 4*. Frankfurt: Dr. Waldemar Kramer.

Staatsbibliothek Preußischer Kulturbesitz (ed.) (1979). Carl Ritter in Seiner Zeit. *Ausstellungskataloge* 11. Berlin.

River Basin Development

C. Fox, Dartmouth College, Hanover, NH, USA

Glossary

Ecosystem The complex of a community of organisms and its environment functioning as an ecological unit.

Ecosystem Services The conditions and processes through which natural ecosystems, and the species that make them up, sustain and fulfill human life.

Geopolitics It describes the state's power to control space or territory and shape the foreign policy of individual states and international political relations.

Global North The world's most prosperous countries, as measured by indicators such as gross national product (GNP) per capita and the UN human development index (HDI). Although many of the countries are located in the Northern Hemisphere, the Global North is an economic rather than geographic term.

Global South The lower-income and less-industrialized countries of Asia, Latin America, and Africa. It is a preferred term for 'Third World' or 'developing' countries.

Governance It refers to more nonformalized means of regulating human activities, drawing on the knowledge and practices of many actors at multiple scale levels. Governance does not depend on governments, but instead draws on civil society – local communities, nongovernmental organizations, and epistemic communities – linked through various kinds of networks and organized in a nonhierarchical manner. Because governance is more flexible, responsive, and adaptable than more bureaucratic government structures, it may be better suited to address complex political–ecological problems.

Human Security It indicates more than just the absence of violent conflict. Human security encompasses issues such as access to a clean environment, freedom from fear, access to education, the provision of clean water, and the right to sustain one's livelihood.

In-Stream Values Values relating to the environment of the stream or watercourse channel.

Integrated River Basin Management The process of coordinating conservation, management, and development of water, land, and related resources across sectors within a given river basin, in order to maximize the economic and social benefits derived from water resources in an equitable manner while preserving and, where necessary, restoring freshwater ecosystems.

Participatory Development Development that strives to include all those affected by development projects and processes.

River Basin The area which contributes hydrologically to a first-order stream, which, in turn, is defined by its outlet to the ocean or to a closed lake or inland sea. River basin is synonymous with watershed and catchment.

Sustainable Development Development that meets the needs of the present without compromising the ability of future generations to meet their own needs.

Transboundary River Basin A river basin that transcends international boundaries.

Introduction

Most of the world's land surface can be divided by river basins. River basins are defined as topographic areas that collect precipitation and transport runoff and sediment to a common outlet. River basin ecosystems sustain terrestrial and aquatic species as well as human societies. While the well being of human societies is closely tied to a healthy natural environment, the history of river basin development is replete with examples of the prioritization of human needs at the expense of river ecosystems. As a result, the freshwater ecosystems at the heart of river basins are some of the most altered and degraded in the world. While there are hopeful trends toward more ecologically minded development and management, important ecosystem services continue to be lost in the name of river basin development around the world.

Efforts to develop and control rivers for human benefit go back thousands of years. Evidence points to the development of irrigation and flood control measures in Mesopotamia along the Tigris and Euphrates rivers as early as the sixth century BC, while the rivers of present day Jordan were dammed 5000 years ago. The river basin as a unit of development is, however, more recent. In the 1890s, Sir William Wilcocks was one of the first to suggest that the river basin (in this case the Nile river) be considered as a unit of development. In the late 1800s, John Wesley Powell proposed dividing the American West into 300 river basin units, which would serve as natural districts of social, technological, and political organization and development. In the 1860s in New Zealand, river management was organized along basin lines in the form of river boards. Attention to the basin scale was also evident in the US in the nineteenth century, when urban commissions were established to survey, acquire, and manage watersheds that protected drinking water sup plies. None of these early efforts can be compared, however, with the scale and impacts of contemporary river basin development, which began in earnest in the early twentieth century. At that time, flood control mea sures, navigation improvements, irrigation schemes, and hydropower development began to dramatically trans form river basins. Such transformations required ad vances in technical knowledge and engineering skills, as well as a particular understanding of human–environ ment relations, one that conceived of nature as a resource to be controlled and harnessed primarily for the benefit of human society. In the words of the US Interior Sec retary Harold Ickes at the dedication of the Hoover Dam in 1935, "Pridefully, man acclaims his conquest of nature."

Twentieth-Century River Basin Development – Controlling Nature to Benefit Society

River basin development in the early and mid twentieth century was underpinned by an unquestioned faith in the ability and right of humankind to control and alter nature for its own benefit. Development was fueled by techno logical optimism, with little recognition of the ecological, social, economic, and cultural benefits of allowing water to flow freely. Rather, water that flowed out to the sea was seen as wasted. Rivers were managed and manipulated in order to make their unpredictable and variable flow as regular as possible, because the natural variability of a river's flow was seen as an impediment to its rational use. River basin development was the domain of engineers, agriculturalists, and economists, with almost no partici pation by those members of society who benefited from or saw value in free flowing rivers. Moreover, power over a river basin ecosystem was associated with human pro gress and celebrated as a symbol of national develop ment. The first prime minister of India, Pandit Jawaharlal Nehru, revealed his country's pride in river development when he spoke of the Bhakra Dam, which he praised as a "stupendous and magnificent work…the symbol of a nation's will to march forward with strength, determin ation, and courage." In a similar vein, when Woody Guthrie sang "Roll on, Columbia, roll on, roll on, Col umbia, roll on/ Your power is turning our darkness to dawn/ Roll on, Columbia, roll on," he was celebrating the way in which Columbia river basin development contributed to national development through hydro power generation.

From India to the United States, the legacy of human attempts to control and manage nature has altered river basins dramatically. Dams, reservoirs, channelization projects, intra and interbasin water transfers, irrigation schemes, navigation improvements, and flood control measures all have contributed to this alteration. The di versity of projects in any given basin reflects the mul tiple use nature of most river basin development. Yet, the above uses also reveal the extent to which some other uses – notably ecological preservation and restoration – have been absent from development schemes. The con sequences of this oversight range from extirpated species and physically fragmented rivers to unusually severe floods and degraded landscapes. Also absent from most multiple use plans is consideration of the many ways in which people rely on intact river basin ecosystems to sustain their livelihoods. The disregard for local com munities that too often has accompanied development reveals the extent to which inequitable power relations have been bound up with the history of river basin de velopment. Yet, even while local communities and eco logical values have tended to suffer disproportionately from river basin alteration and development, the regions in which those basins are located have often benefited significantly in terms of economic development. Eco nomic development results from enhanced agricultural production, a more secure water source for urban areas, better transportation, energy generation, and increased employment opportunities. River basin development is therefore often seen as an important strategy for lifting up economically lagging regions, as was the case with the Tennessee River valley in the southeastern US, and as is the case with the ongoing Southeastern Anatolia Project in Turkey and the Narmada River projects in Gujarat, western India.

River basin development, by its very nature, has im pacts in and around the region where it occurs. The mixed consequences of regional impacts mean that, when people have the political freedom to do so, basin devel opment projects almost always face some opposition from those who feel they are bearing costs unfairly while not sharing in the benefits. The regional impacts of devel opment vary depending on a whole host of factors: the scale and types of projects, the physical geography of the basin, the political context, the goals of the project, the nature of scientific assessments, mitigation measures, and local perceptions and values, among other variables. What is common across the world is that water projects affect regional development through numerous pathways. Geographer Gilbert White captured the essence of the regional implications of river basin development when he noted: "Every change in landscape or resource use on a large scale disturbs a complex set of relationships." The disturbance and reconfiguration of human–environment relations is unavoidable in the case of modern river basin development.

River Basin Development in the Global North

Some of the earliest examples of large scale, multipurpose river basin development come from the US. In 1933, the Tennessee Valley Authority (TVA) embarked on a comprehensive river basin development plan focused on the Tennessee River basin. Charged with development of land and water, the TVA constructed nine major dams on the Tennessee River and 42 dams on tributaries. While power generation, navigation improvements, and flood and erosion control have been cited as positive outcomes of the TVA projects, the flooding of prime farmland, the uneven distribution of benefits, and the massive alteration of the river ecosystem have been more problematic. The significance of the TVA lies not only in the fact that it was the first attempt to introduce regional development via comprehensive river basin development, but also that it became a model for river basin development around the world.

Conceived of during the New Deal, the TVA was an expression of the changing role of the federal government in American society. It was, according to President Roosevelt, also an important antidote for the Great Depression. The TVA initiated government led regional planning through the use and development of natural resources to alleviate rural poverty. The project included not only river development but also policies and strategies for managing fields, forests, and urban areas in the flood prone areas of the basin. Importantly, the project aspired to a total reorganization of life in the Tennessee River valley – it was not simply about developing the river. To that end, the TVA was charged with planning for power generation and flood control on a watershed scale. In addition to hydropower generation, the TVA worked to educate the public about land conservation and management, provided recreation areas, engaged in town development, and constructed public buildings. The legacy of the TVA can be seen today in the practice of river basin analysis and development, in the multipurpose use of rivers and reservoirs, and in the increased federal involvement in environmental regulation and policymaking.

Other US river basins targeted for large scale development in the early twentieth century include the Columbia, Colorado, and Mississippi. The Columbia has been transformed by 14 major dams on the mainstream and hundreds of large and small dams and diversions along its tributaries. One of the most massive multipurpose dams on the river is the Grand Coulee, which was begun in 1933 to create jobs, generate hydropower, provide irrigation water, and facilitate navigation. The Grand Coulee and other dams have transformed the Columbia from a natural ecosystem into the economic engine for the Pacific Northwest, captured by idea that the river has 'died and been reborn as money.' Irrigation water from the river has changed the basin landscape, making possible a shift from marginal farming to large scale, commercial agriculture. While the benefits of regional economic development have been significant, they flow disproportionately to electric power users and farmers, with costs borne by Canadian First Nations, Native Americans, and the salmon and steelhead upon which their cultures depend.

The Colorado River basin has been the object of similar large scale multipurpose development. Water was first diverted from the river in 1901 to irrigate farmland in the Imperial Valley of California. After the construction of the Hoover Dam in 1935 came the Glen Canyon, Headgate Rock, and Parker dams, among others. Colorado basin development transformed the Southwest, turning desert to cropland and making possible the growth of cities such as Phoenix and Las Vegas. Yet, while the Colorado has been a key resource spurring regional development, many would argue that development that depends on far off water is increasingly unsustainable. Today, the Colorado River is overallocated between the six states that share the basin, with little water to spare for the river ecosystem itself, which has been seriously degraded by decades of development. A multiyear drought, population growth, the prospect of reduced precipitation as a result of climate change, and increased pressure from environmentalists to return water to the river will create unprecedented challenges for sustainable management of the Colorado in the twenty first century.

The Mississippi River provides yet another example of unfettered basin development in the early twentieth century. Beginning in the 1930s, a series of 29 locks and dams were built on the upper Mississippi, most of which were intended to maintain a channel for commercial barge traffic. The lower river's development came in the form of levees and dikes to control flooding and secure a navigation channel. The levees and channelization, along with substantial loss of bottomland forests to agriculture, have resulted in a loss of wildlife and fish habitat, decreased water quality, and an expansion of the hypoxic zone in the Gulf of Mexico. The massive 1993 flood in the Mississippi River basin, which was very destructive to human populations, demonstrates the fiction of complete human control of nature. Even a large array of levees, dikes, dams, and locks cannot really contain a river. Again, while Mississippi River basin development has clearly had regional and national economic benefits, human and environmental costs have been high. The disastrous impacts of hurricane Katrina resulted, in part, from decades of environmental alteration of the Mississippi River basin.

Outside of the US, river basin development was taking place on a similar scale in the former Soviet Union, where government planners saw electrification as the key to industrialization. Beginning in the 1930s and accelerating in the post war period, massive alteration of rivers such as the Dneiper, with six large dams, and the Volga, with nine large dams, took place in the name of national development. Also following the example of the US dam builders, Australia began constructing dams and weirs on rivers such as the Murray–Darling in the 1930s. Development of that basin, which drains one seventh of the Australian landmass, has turned a semiarid region into the most productive agricultural region in the country. While this has clearly benefited regional development, the ecological costs of infrastructure development (along with other land use changes such as the introduction of sheep grazing) have been high. The river has been so overallocated that it is estimated that only 20% of the water in the river basin at any one time actually reaches the ocean. As in the American West, drought and climate change projections will make sustaining people and ecosystems in the basin a tremendous challenge in the coming decades.

In France and elsewhere in Europe, rivers such as the Rhone, which boasts multiple hydroelectric dams, locks, and irrigation networks, underwent radical technological and environmental transformation in the name of reconstruction and modernization after World War II. Even water rich countries such as Canada have found it necessary to develop river basins to promote nation and industrial development. A central feature of river basin development in that country has been water transfers, because water is more often available in the northern part of the country, where the population is minimal. Examples include the massive La Grande project in Quebec, which has four dams to divert and double the flow of the La Grande River to generate hydroelectricity.

All of the above examples are from the Global North, where the majority of large rivers have been dammed, diverted, or channelized to support industrialization, commercialization of agriculture, and to make possible human settlement in flood plains. After decades of unabated river basin development from the 1930s through the 1970s, scientists began to accumulate evidence of impacts on sediment movement, water quality, vegetation, aquatic species, and social systems. Also during the 1970s, the environmental movement raised awareness about the consequences for the environment of modern, industrialized, and high consuming societies. Ecologists, citizens, scientists, and policymakers began to question whether the ecological and cultural costs of environmental alteration of river basins and other ecosystem might outweigh the benefits. In geography and other social sciences, researchers began to write critically about the fact that some people and ecosystems bore disproportionately heavy burdens and saw little benefit from development. Ironically, just as the Global North was beginning to reflect on lessons learned from decades of development, river basin development in the Global South was accelerating, or in some cases, just getting underway.

Exporting the Model – River Basin Development in the Global South

Some of the most active regions of river basin development today are in the Global South. This is primarily because comprehensive basin development came later to these areas, and because basin development is seen as an important strategy for lifting up these often economically disadvantaged regions. The North's model of large scale river basin development and the technology to implement it were introduced after World War II, when the US and Western Europe were seeking to improve the economic status of the world's former colonies. River basins were seen as important engines of economic growth, which would stimulate modernization and regional development through irrigation, flood control, and hydropower generation. While the experience of basin development was initially similar to that of the North, in the past few decades it has become more contentious and politicized than it was during the heyday of development in the US and Europe. While the critiques that drive this politicization are due largely to increased awareness of the negative social and ecological impacts of basin development schemes, they are also closely linked to more general critiques of development and the discourses that sustain it. These critiques of development, put forward by academics and activists, reveal that development has often failed to deliver on the promise of a better life for many people in the Global South. Rather, environmental degradation, marginalization, dependency, and inequitable distribution of costs and benefits are too often the unintended consequences. These problems are especially evident in river basin development projects, since the benefits tend to flow to cities, industry, and large irrigators, while those who depend on river basin resources to support their livelihoods tend to bear the immediate costs of development. In this regard, river basin development often creates human insecurity in specific populations.

An example of how the North's model was exported to the South comes from the Mekong River basin in Southeast Asia (**Figure 1**). In the mid 1950s, the US Bureau of Reclamation and the Wheeler Commission (headed by Lt. General Raymond Wheeler, former chief of the US Corps of Engineers) came to the region

Map No. 4112 Rev. 2 UNITED NATIONS
January 2004

Department of Peacekeeping Operations
Cartographic Section

Figure 1 Greater Mekong subregion.

offering ‘expert’ assistance in planning a multipurpose scheme modeled on the TVA, which would take the form of a cascade of 180 dams on the river and tributaries. Not only would the TVA serve as a model, but, according to President Johnson, the potential for hydro development was so great that the Mekong scheme would “dwarf the TVA.” Notably, this plan was not strictly about economic development, but also bound up with Cold War geopolitics. Multipurpose river basin development was intended to promote modernization, which, according to prevailing theories, would prevent the region from falling under the influence of communism. Due to the US–Vietnam war and the isolation of the region that followed, the original plans were put on hold for many decades. With peace and economic integration, the river is once again the focus of large scale development. Today, revived development plans are much more politicized and contentious than they would have been 50 years ago, because there is much greater awareness of the way in which river basin development around the world has negatively impacted livelihoods for the sake of national development goals.

While the nature and scope of development may be more contested today than in earlier decades, the idea of regional development through river basin development remains a very powerful idea in the Mekong basin. Geographically, the idea of regional development is expressed through the Greater Mekong Subregion (GMS), which comprises Cambodia, the People’s Republic of China, Lao People’s Democratic Republic, Myanmar, Thailand, and Vietnam. The GMS was conceived in 1992 with the assistance of the Asian Development Bank, and it provides a forum for the six basin countries to promote subregional economic cooperation and the free flow of goods and people through the development of infrastructure such as transportation corridors and shared energy grids. The GMS envisions the subregion as an international growth area linked by the Mekong River.

River basin development in the Global South sometimes occurred under colonial rule, as was the case with the Kariba dam on the Zambezi River, built by the British in 1959. Under conditions of such inequitable power relations, the beneficial regional development that resulted has to be weighed against far reaching negative impacts. The Kariba dam, for example, had significant environmental impacts and even more devastating social consequences. The dam forced the relocation of 57 000 Gwembe Tonga people, with disastrous consequences for their health, culture, and well being. This case is, unfortunately, not unusual. Upward of 50 million people worldwide have been displaced in the course of river basin development, generally when a large dam is involved. Today, in places where resettlement is necessary – such as the Lesotho Highlands Water Project, which is an ambitious replumbing and redirecting of the country’s water to support factories, mines, and homes in South Africa – authorities are trying to provide adequate compensation and to give affected communities more of a stake in the benefits of the project. Results are mixed, but most resettled communities experience significant insecurity. In some cases, there is little resistance to displacement due to lack of political freedom, as in China’s Three Gorges Dam project, which will displace 1.3 million people when fully completed. In other instances, such as development of the Narmada River basin in Gujarat state in India, resistance has been dramatic, with villagers refusing to leave their homes as the waters behind the Sardar Sarovar dam rose. In the Narmada basin, the majority of those being displaced are marginalized tribal peoples who depend on an intact river ecosystem to sustain their livelihoods and cultures. In democratic India, basin development raises questions about who owns the rivers and fish and who has a voice in the way in which those resources will be used. Critics suggest that river basin development has not lived up to its promise in that country, noting that India has 3600 large dams, but 200 million people do not have safe drinking water and 600 million people lack basic sanitation. River basin development has done little to address these important aspects of human insecurity. Still, others see basin development as the logical path to economic prosperity in water scarce regions such as Gujarat.

Latin America has been a particularly active site of river basin development during the last few decades. The region is a leader in another trend in river basin development in the Global South, that of privatization. An example of this is the Bio Bio River in Chile, which flows from the Cordillera of the Andes to the Pacific Ocean. Over 1 million people use the resources of the Bio Bio for drinking and irrigation water, recreation, and fisheries. ENDESA, the largest private company in Chile, is planning to construct six hydroelectric dams on the Bio Bio. The largest of the dams will displace 700 Pehuenche Indians and flood over 70 km of the river valley, inundating the richly diverse forest and its wildlife. Environmental and indigenous rights groups oppose the project, while the state of Chile sees it as essential for energy security of the country. Critics are concerned that private companies may be less willing to spend money to mitigate social and ecological impacts, given their need to maximize profits.

In Turkey, the Southeastern Antatolia Project (GAP) represents a particularly ambitious plan for regional development through river basin development. The GAP is a $32 billion, multisectoral, integrated, regional development project that seeks to enhance the social, institutional, and economic situation in this economically disadvantaged part of the country. River basin development is the focus of regional development, and the project, when completed, will include 22 dams, 19 hydropower projects, and extensive irrigation and drainage systems.

Rethinking River Basin Development in the Twenty-First Century

While ecologically and socially problematic river basin development continues around the world, it is also the case that rivers are increasingly valued as more than transportation routes and commodities. River basins are recognized as ecosystems and landscapes worthy of protection and restoration. In the Global North, where the legacy of three quarters of a century of river basin development is increasingly evident, millions of dollars are being spent to restore the physical integrity of river basin ecosystems and to support species that have been adversely affected by development. Sometimes this takes the form of actually 'un developing' a river through dam removal, as was the case with the Edwards dam on Maine's Kennebec River in 1999, when the river flowed freely for the first time in 162 years. Sometimes basin restoration involves giving water back to a river, as happened in 2006 when Los Angeles returned water to the Lower Owens River after 93 years of diverting it to serve urban users. Giving back water to in stream uses means taking it from other users, which can be very contentious. In Oregon's Klamath basin, during a drought in the summer of 2001, farmers were denied irrigation water in order to leave water in the river to protect en dangered salmon. Conflict between in stream needs, farmers, fishermen, and Native Americans remains un resolved in the basin today. River basin protection and restoration can also be inspired by recognition of the economic value of ecosystem services. This is the case in the Catskill/Delaware basin, which supplies 90% of New York City's water. The water is unfiltered as a result of watershed protection measures. Protecting the basin costs about $100 million a year, whereas a filtration plant large enough to support the Catskill/Delaware system would cost from 6 to 10 billion dollars, with an additional $110 million annually in operation and maintenance. Pro tecting the land and water that comprise the river basin saves money.

While it is debatable whether large scale river basin development can ever be truly sustainable, natural re source managers and policymakers are certainly making efforts to be more ecologically and socially responsible in both the development and management of river basins. In places where development is ongoing, scientists, advo cacy networks, nongovernmental organizations (NGOs), and local communities have come together to thwart deeply problematic projects, as was the case with South ern Okavango Integrated Water Development Project in Botswana in the early 1990s. The project would have caused irreparable harm to a unique ecosystem and local inhabitants. Others have gone forward, but with con sultation from ecologists and anthropologists, such as the case of the Nam Theun II project in Laos. The need to reform river basin development practices has been con firmed by the recent work of the World Commission on Dams, which conducted a global study of large dams. In 2000, the Commission, which was made up of antidam activists as well as representatives of the dam building industry, issued a report that, while recognizing the benefits of dams to society, was quite critical of the dis regard for people and ecosystems that has characterized much large dam building in the twentieth century.

As a result of changing attitudes and increased sci entific knowledge, most policymakers and natural re source managers aspire to implement a process called integrated river basin management (IRBM). IRBM rests on the notion that management of river basins must in clude maintaining ecosystem functioning as a paramount goal. The approach is holistic, linking socioeconomic development and environmental protection. IRBM is driven by an ecosystem approach, which recognizes river basins as dynamic over space and time. All activities within a river basin must be managed in a coherent and coordinated manner in order to maximize the economic and social benefits while preserving and, where necessary, restoring freshwater ecosystems. Additionally, IRBM seeks to be participatory, including all the major stake holders in decision making. The ultimate goal is sus tainable development of river basins. While IRBM sounds very promising, its focus on management, and its overreliance on technocratic solutions fails to confront the inequitable power relations that are so often at the root of unsustainable policies and practices.

Transboundary River Basin Development

All of the challenges associated with sustainable river basin development and management become immensely more complicated in a transboundary context. Put sim ply, river basin ecosystems do not match up with political boundaries. The mismatch creates opportunities for both cooperation and conflict between countries in the de velopment and management of rivers. Transboundary basins are surprisingly common – 40% of the world's population live in a transboundary river basin and 145 countries share a basin. Transboundary basins account for 60% of global river flows. Most of the world's great rivers contain numerous countries in their basins. The Danube River basin encompasses 19 countries, the Mekong six, the Nile 11, the Zambezi nine, the Congo 13, and the Amazon nine. Interestingly, the number of shared basins grew in the last decades of the twentieth century, mostly due to the breakup of the former Soviet Union. Whereas there were 214 transboundary basins in 1978, now there are 263. These basins bring development issues into the realm of geopolitics, raising a host of questions about who has the right to develop shared international resources and how to determine rights and responsibilities. Trans boundary basins, among other global environmental problems, call into question the logic of assigning re sponsibility for natural resource management to sover eign, territorial states. No single state can effectively protect and manage the resources of a transboundary river. Moreover, states may have different visions of the role of a river basin in regional development.

With regard to the potential for conflict over basin resources, some countries – generally upstream ones – are in more advantageous positions with respect to river basin resources. China in the Mekong basin and Turkey in the Tigris–Euphrates basin are two such examples of upper most riparians with ambitious river basin development plans. China has completed construction of two dams on the Mekong mainstream and has plans for as many as nine on the river and its tributaries. Downstream countries such as Cambodia and Vietnam, which rely heavily on basin resources for agriculture and fisheries, will be impacted by this development, but they have little power to influence China's dam building. Similarly, Turkey has plans to de velop 22 dams and irrigate 1.7 million hectares as a part of its Southeastern Anatolia Project. This will affect flow to downstream countries Syria and Iraq.

In such situations of hydrological interdependence, one might expect conflict – 'water wars' – to erupt be tween countries. In general, this has not been the case. States have more often resorted to cooperation as a way to deal with international tensions over water. Even in water scarce regions with high levels of political in stability such as the Middle East, water wars between sovereign states are almost nonexistent. Cooperation often takes the form of treaties – more than 3600 of which have been signed by organized political bodies since AD 805. States also rely on international watercourse law to adjudicate disputes and promote rational and equitable use of international water resources.

This is not to imply that conflict is absent from transboundary basins, but it usually exists at scales other than the international scale, and it is not always violent. A focus on water wars can be problematic to the extent that it diverts attention from conflicts at other scales, conflicts that both contribute to and reflect human insecurity. Moreover, with regard to an issue as complex as river basin development, the 'cooperation or conflict' dichot omy is both simplistic and somewhat deceptive, particu larly to the extent that cooperation is seen as an unmitigated good. It is very possible that cooperation between states to massively develop a basin would be undesirable from both an ecological and livelihood per spective. Similarly, conflict that is not violent can lead to positive outcomes if it forces people to rethink a prob lematic *status quo* or shifts power relations in positive ways.

The desire to cooperate around transboundary basins (and to use the basins as the focus for regional devel opment) has spawned a host of transboundary river basin organizations and, it could be argued, is leading to the emergence of new forms of global governance. Examples of river basin organizations include the Mekong River Commission, the Zambezi River Commission, the Dan ube Environmental Programme, the Nile Basin Initiative, Senegal River Development Organization, Limpopo River Basin Commission, and the Okavango River Basin Commission. While these formal commissions are doing important transboundary work, most of them remain very state centric. Some, as in the case of the Danube En vironmental Programme, represent what has been termed 'deep' institutional cooperation. This is because the Danube Programme is one of the few transboundary river basin organizations designed through a process of genuine public participation, including NGOs, journal ists, and local authorities, and academics. The inclusion of this array of actors in the decision making process represents a move toward more governance and less formal government in transboundary basin politics. Many geographers would argue that governance, which draws together nonstate actors across national boundaries, is a better option than more government when addressing complex political–ecological problems. This is because these networks of actors, operating at a variety of scales, tend to be more flexible, responsive, and adaptable than government bureaucracies. In a number of transboundary basins, such as in the Mekong and Zambezi basins, groups of nonstate actors have come together to form advocacy networks that work to influence or stop problematic projects. These advocacy networks tend to work outside of the official, state centric river basin commissions (the Danube differs in this regard, because governments have welcomed nonstate actors into the official river basin commission), and they are, some argue, actively gener ating transnational spheres of more participatory and democratic decision making. The evidence to support the emergence of these spheres is somewhat limited, but it is clear that in the twenty first century, river basins – whether domestic or transboundary – will be critical sites for working toward and establishing sustainable society–nature relations.

See also: Natural Resources; Sustainable Development; Water.

Further Reading

Adams, W. M. (1992). *Wasting the Rain: Rivers, People, and Planning in Africa*. Minneapolis, MN: University of Minnesota Press.

Biwas, A. K., Unver, O. and Tortajada, C. (eds.) (2004). *Water as a Focus for Regional Development*. Oxford: Oxford University Press.
Conca, K. (2006). *Governing Water: Contentious Transnational Politics and Global Institution Building*. Cambridge, MA: MIT Press.
Finger, M., Tamiotti, L. and Alloughe, J. (eds.) (2006). *The Multi Governance of Water: Four Case Studies*. Albany, NY: State University of New York Press.
Freidmann, J. and Weaver, C. (1979). *Territory and Function: The Evolution of Regional Planning*. Berkeley, CA: University of California Press.
Hambright, K. D., Ragep, F. J. and Ginat, J. (eds.) (2006). *Water in the Middle East: Cooperation and Technological Solutions in the Jordan Valley*. Norman, OK: University of Oklahoma Press.
McCully, P. (2001). *Silenced Rivers: The Ecology and Politics of Large Dams*. London: Zed Books.
Newson, M. D. (1997). *Land, Water, and Development: Sustainable Management of River Basin Systems*. London: Routledge.
Wescoat, J. L. and White, G. F. (2003). *Water for Life: Water Management and Environmental Policy*. Cambridge, MA: Cambridge University Press.

Relevant Websites

http://www.irn.org
International river network.
http://www.un.org
Mekong River basin (public domain).
http://www.tva.gov
Tennessee River basin (public domain).
http://www.thirdworldcentre.org
Third World centre for water management.

Rural Communities

S. Salamon, University of Texas at Dallas, Dallas, TX, USA
K. A. MacTavish, Oregon State University, Corvallis, OR, USA

Glossary

Community of Place Sum total of interactions among people who live in a particular locale.

Ideal Type A concept developed by Max Weber as a method for comparative study. It is a model constructed using a detailed empirical study of a phenomenon against which similarities to, and divergences from, can be compared. An ideal type does not represent the best image or what ought to be. It does, however, present a clear and unambiguous pure case of the phenomenon.

Sense of Community The experience of belonging and loyalty to a larger group of human relationships associated with a community of place.

Social Capital The investment of participation people make in their communities which creates social resources, such as trust and watchfulness, that have reciprocal effects on the community, its civic institutions, and cooperation that benefits them, and the greater good.

Sustainable Community Development An approach to community development that seeks to ensure environmental, social, and economic viability.

Voluntarism A tradition of unpaid activity performed beyond home and family is relied on to provide services in rural communities, due to a sense of obligation. It is a hallmark of rural life.

Definitions

A rural community is spatially, physically, and socially distinctive. Typically, a rural community is strongly identified with a particular geographic space. As a community of place, it occupies a small scale, compact settlement with clear boundaries. Spatially dispersed across the countryside, the low density rural communities are set apart by distances that reflect early travel of less than a day by foot, horse, or carriage. Historically, such places were often relatively isolated by being in remote areas lacking links to the wider society such as good roads, navigable rivers, or more recently, access to train, plane, bus, or boat transportation. Remoteness favored small, homogeneous, indigenous populations, having an identity derived from a shared culture and history, language and religion, residential stability, and face to face interaction. People in rural communities express greater satisfaction with their lives than do urban people. But for almost every indicator of well being, employment, income, socioeconomic status, and population growth, rural communities lag behind urban places. For many societies, their rural communities continue to represent stability and embody their unique national identity, culture, and values in the environment of rapid globalization and high international connectivity.

Rural Community as a Social and Theoretical Concept

Rural community as a concept is associated with the development of sociology as a discipline in the nineteenth century. The concept of rural community from the outset was shaped by social, artistic, and philosophical thought that emerged as a reaction to the rapid industrialization and urbanization of European societies undergoing transformation by the French and Industrial Revolutions. Before these two revolutions most people lived in rural areas where work and life took place in the contexts of the extended family and the peasant community. After the revolutions, (1) population mobility increased with mass migration to urban areas in search of work in newly emerging factories and (2) work became more specialized and focused outside the household. People no longer produced most of what they consumed. Money and labor became the primary modes of exchange (as opposed to barter or trade). Europe became dominated by urban places incorporating its remaining rural communities into large, centralized nation states. Elsewhere, in Asia, Africa, and the Americas, rural communities remained the prevailing residential form.

Nineteenth century sociological theorizing focused on comparing rural communities with those in newly emerging urban areas. Starting with Alexis de Tocqueville's early nineteenth century study of the new US, rural communities became associated with discussions about social capacity creation and its availability as a resource for community and human development processes. Tocqueville reported that US rural communities were characterized by high rates of voluntarism, which he saw as both the result of and crucial to the maintenance of a democratic society. Classical sociological theorists, beginning with Ferdinand Töennies, followed by Emile Durkheim, Max Weber, Karl Marx, and Georg Simmel, and the US anthropologist Robert Redfield, address how industrialization, modernization, and urbanization affect social organization at the community level. These theorists use community types in comparative typologies, following the Weberian tradition of conceptualizing essential characteristics into an 'ideal type', to explain the causes and consequences of the community change they observed. Typically, a continuum from ideal type rural to urban forms of community organization is posited to examine the social costs or benefits from rural communities being superseded by larger, complex urban communities. In particular, these theorists look at the spatial shift of people over time and whether that shift entails loss of community or creates new community forms. Fundamentally, these typologies reflect concern that, due to rapid social change, a sense of community would be lost or changed beyond recognition, and yield societal chaos.

Researchers no longer believe that any one community is representative of a nation or is a microcosm of society, as did the classical theorists. But the rural community as a concept still figures largely in discussions about whether a sense of community is threatened by twenty first century social change, spurred by globalization and communication technologies. Pre World War I is considered the golden age of US rural communities, in a nation still largely rural. After World War II, US rural communities were transformed as the nation's population grew. Suburbs sprouted (attempting to reproduce small town life) and rural communities became increasingly integrated in the larger society via the media; transportation systems; and urban centralization of legal, educational, financial, medical, and commercial institutions. According to the sociologist Roland L. Warren, this transformation of society meant that rural communities lost isolation and autonomy but retained other distinctive characteristics. Warren's ideas motivated the influential 1950s rural community study, *Small Town in Mass Society* by Arthur Vidich and Joseph Bensman. These sociologists argued that rural communities are not distinctive social units but rather their integration in the wider society undermines them as increasingly as they become controlled by urban forces. Later work, exemplified by Robert Putnam's *Bowling Alone*, argues that US rural communities remain distinctive due to their association with rich stores of social resources, or *social capital*, generated by the sustained voluntarism needed to make them work.

In response to the loss of autonomy connected with contemporary society, perhaps nostalgically, rural communities remain more comprehensible units for people and are thus thought to be a location for experiencing the salience of group life. Such associations with rural communities have motivated the vast migration of urban people into the exurban or rural hinterlands beyond the suburban rings around metropolitan areas (see **Figure 1** to see recent population changes in the rural US). People continue to associate rural communities as the locus for experiencing a sense of community in an increasingly urbanized world.

Theoretical and Ideological Debates about Rural Communities

Rural community as a concept is entwined with national myths and beliefs about a distinctive societal identity. Sociologists, geographers, and planners use the concept of rural community to debate issues of autonomy, modernization, distinctiveness, regional triage style rural planning, and causes for changing migration streams. Rural community as a concept is the basis for the continuing debate about whether quality of rural communities deteriorates when agriculture becomes more industrial in scale. The concept also underlies the debate about whether remote rural communities should be abandoned entirely, with the land and wildlife restored to presettlement conditions. Proposing a 'Buffalo Commons' for the US Great Plains is one such idea. A philosophical search for civility and belonging in industrialized nations, and the myth that rural communities are morally superior as 'the' good community, fuel arguments that losing them signals more generally a decline in a society.

Rural communities are simultaneously stereotyped by urbanites as places of ignorance, isolation, backwardness, provincialism, and outside the mainstream of thought and practice. This view is connected with rural areas generally lagging behind urban areas in the last century according to most indicators of well being. Dependence on natural resource extraction for employment in a context of globalization, outmigration of youth, aging populations, and lower levels of human capital are all factors associated with the rural/urban gap. Rural stereotypes, both positive and negative, are found in literature, television, advertising, and films.

Key Types/Characteristics

A wide range of social–organizational forms are incorporated by the category of rural community. Communities in more remote rural areas include: voluntarily isolated indigenous peoples, tribal, nomadic, and peasant settlements. In more developed nations, small towns and, more recently, vacation home developments or suburban like settlements fall under the umbrella term of rural community. Most rural community types are associated with a subsistence economy dependent on the land, or some combination of resource extraction from the land: agriculture (farming, animal husbandry, fishing, hunting, and gathering), forestry and extractive industries (mining or energy resources), as well as tourism and transportation. When natural resource dependent activities diminish in financial value, whether for ecological,

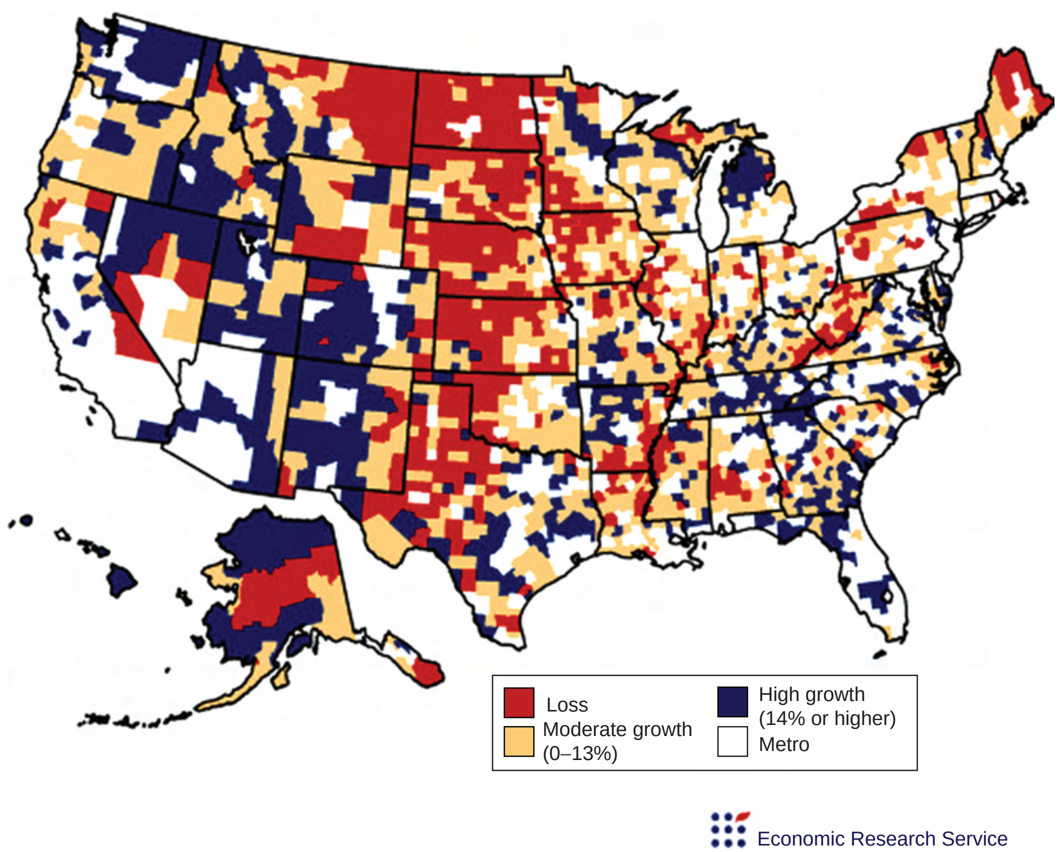

Figure 1 Recent population changes in the rural US (1990 2000). Reproduced from USDA.

local, or global reasons, rural communities everywhere exploit their hinterland location and their distinctive culture (life ways, architecture, food, music, dance, or crafts) for tourism or recreation activities.

Worldwide, formerly remote rural communities are being transformed whether by urban sprawl, or from natural resource exploitation by their own nation or multinational firms. In highly urbanized nations, rural areas have become attractive to urbanites seeking nature, solitude, and space. The historic identity of rural com munities is potentially transformed by exurban expansion into what were previously remote settlements. The productive reason for its economy is replaced by a resi dential or recreational purpose, as when transformation is into a bedroom community. When a change is made in the subsistence basis of a place, other aspects of the community culture are profoundly altered.

Rural community settlement patterns change over time as various uses and reasons for existence wax and wane or persist according to economic and technological developments in the wider society. Archetypal rural community forms include: railroad towns, resource dependent towns, cattle and ranching towns, market towns, factory towns, energy boom towns, tourist towns, exurban towns, ghost towns, utopian communities, mili tary base towns, and trailer parks. Other sorts of primary economic engines accounting for diversity in rural community include specialized employers such as aca demic institutions, religious institutions and shrines, medical facilities or institutions, government instal lations, prisons, national parks, wildlife refuges, or archeological excavations.

Distinctive Social and Economic Processes

Rural communities are socially distinctive due to their small size. Such places range from settlements called hamlets or villages of fewer than 500 people to small towns with populations of 5000 to 50 000 – the latter size constitutes a small metropolitan area, according to the US census. Each small population is integrated by a shared history and norms of social control based on a lack of privacy, watchfulness, gossip, and concerns for family reputations. High levels of trust derive from familiarity, face to face interactions, interdependence, and cooper ation to sustain the prized self sufficiency of a place. The spatial separation of rural communities fosters a value of egalitarianism (even when it is understood that differ ences exist), social networks that are overlapping and reinforcing (e.g., the same people are in networks of work and church), the belief that children belong to everyone, and dependence on the local informal economy (bartering and trade). Because assistance is far away, rural community folk are knowledgeable about how to live off the land, have skills to build or repair buildings and machinery, and know how to keep themselves safe and secure, unlike urbanites. Traditions of voluntarism exist in rural communities to provide the services urban people take for granted (like emergency medical care), and to make these small places function. Church and community overlap because of the relative homogeneity of the population and the likelihood that most residents share the same religion. Thus, social control is suffused with religious morality more so than in urban places. Rural communities are consistently associated with a greater satisfaction of life than are urban communities.

The distinctive social processes of rural communities are altered by the migration flow associated with exurbanization or suburbanization. In developed nations, as well as other parts of the world, a belief that life in rural communities is closer to nature makes their ambience attractive to urban people who seek solace in the supposed quiet and security of these spatially separate places. Whether for the second/vacation home in the countryside or retirement to a scenic place, urban people living in a rural community are often out of sync with the local culture and morality.

Ethnocultural Variation

Across rural communities, even in the same region, ethnocultural variation is high. Distinctive architectural and settlement patterns are the most visible features to typify regional and national variation. Settlement patterns often originate in the agrarian past of the present rural communities. Whether farmers lived in town and traveled to the fields as in much of Europe, had seasonal mobility as in mountainous regions like Switzerland, or lived scattered across the countryside as in much of North America explains the distinctive community form of housing density. Architectural style of homes – roof pitch and materials, number of stories, incorporation of animal housing, or utilization of barns, wood, or stone construction – derives from what materials were readily available for construction. Kinship ideology also makes a mark on agrarian settlements by embodying rights to land, inheritance of property, household form and size, and landownership patterns. Housing size or the spatial clustering of dwellings reflects whether extended kin groups live and work together. Whether churches anchor a community or are scattered across the countryside in small hamlets varies according to whether religious denominations are centralized in the state or locally organized. Church architecture similarly is localized or dictated by central authorities. A church is often the tallest structure in the rural countryside and serves as a landmark for travel or property surveying.

Regional cultural differences are the wellspring for distinctive food, music, language, and dance that combined account for distinctive rural cultures. Customary life course events like birth, marriage, or funerals are typically celebrated by the entire rural community, because many people have a kinship bond, and because everyone knows everyone else. Community ritual events are also used to mobilize cooperative labor for major endeavors like building a school, repairing a church, maintaining irrigation systems, celebrating the seasons, or honoring specific religious days. Such celebrations simultaneously enact ritual events and build a sense of community. As a nation urbanizes its distinctive rural regions and cultures become exploited for tourism purposes and as agriculture and other extractive industries wane in economic importance, ethnic and cultural variations tend to emerge as the major economic driver of tourism for rural communities.

Rural communities – home to minorities of caste, class, color, or ethnicity – may be characterized by persistent and pernicious poverty. Rural poverty is sometimes worsened through the control or exploitation by external or internal elites whose best interests lie in maintaining the status quo. The US examples include African American descendants of slaves in the rural South, Appalachian mining communities where the coal is controlled by outsiders who deal with locals through an entrenched elite, and remote rural American Indian reservations that lack resources that can provide jobs. Elsewhere, some African or Chinese rural, agrarian populations were and are exploited to sustain urban areas and/or colonial rule. The isolation of impoverished rural communities conveys a sort of invisibility to the wider society that shrouds their exploitation, inequality, and suffering. Rural community diversity in these contexts is equated with ignorance, and people deserving of their treatment. Similarly, the natural resources specific to a rural setting like water, coal, oil, gas, or minerals, such as diamonds or rubber, become inherent to processes of rural exploitation. The landscape itself and its indigenous rural communities are exploited for international tourism or recreation. Examples of this phenomenon are the grand ski stations of the French Alps, the Costa Rican rain forests, or the US Grand Canyon, and other national parks.

Rural Communities and the Life course

That 'It take a village to raise a child' embodies a universal belief in rural communities that childrearing requires a group effort. In some languages, older women share the kin terms of address of mother, auntie, or grandmother in reflection of their behavior toward all children. Rural communities are often considered good places to raise children not only because of the watch fulness and shared responsibility, but also because of being close to nature and perhaps safer than urban places. For children, the downside is that everyone is nosey, and if a child gets into trouble, parents quickly know about it. Watchfulness is typically by older people, allowing them to contribute to the rural community when no longer economically productive.

Watchfulness and support in rural communities, a product of familiarity and routinization of daily life, while important to childrearing also helps other de pendent groups, like the elderly. Elders can depend on support from kin and neighbors after they lose self suf ficiency. When an older person's routine is altered somehow, neighbors know to check on their situation. Thus, a rural community with its daily routine and its small territory make it a good place to grow old, among people who care. Rural community involvement in the shared tasks of raising its children or caring for the elderly is a source of the social resources that allow residents to mobilize for other enterprises that benefit the greater good. Sometimes called social capital, such re sources are needed by rural communities to sustain themselves or renew themselves.

Rural communities worldwide are places youth leave to seek a better life. This accounts for the migration flow from rural communities to urban areas worldwide of those with the energy and the drive for social mobility. Mexico, for example, loses young workers who migrate to the US seeking work. Eastern Europeans, such as Al banians, migrate to Italy and Africans migrate to Western Europe destinations where more jobs exist than in their native rural communities. Sending areas are typically home to the old and the young or other dependent in dividuals. In agrarian areas, women are left behind to work in the family fields because the men are in urban places. Those that stay behind expect the out migrants to send back financial support. Migrants may return when they grow old, to live out the remainder of their lives. Thus, rural communities frequently have a hollowed out demographic profile with more very young and elderly people than youth or young adults living there. Resi dential stability of older adults who never migrate, is termed 'aging in place' and is a process particularly as sociated with rural communities.

Rural Communities in the Information and Globalization Age

Because rural communities lack the communication in frastructure typical of high density areas, they may be left behind as work becomes more dependent on these facilities and associated skills. Rural communities, how ever, are leapfrogging over their lack of traditional tele communication infrastructure, by adopting cell phones. Thus technology can overcome the spatial distance in herent to rural life. Migrating family members utilize cell phones and other technology to keep in touch with those in the sending community. Similarly, long distance trucking transport and overnight delivery services link even isolated rural places to goods and services not available locally. Because remote communities are built along historic roads connecting places, freight trucks travel through small places on the only road available. Long distance trucking in many places has destroyed the quiet and solitude of remote rural communities.

Does any place really remain rural when connected to the world via the Internet, urban media, continental transportation networks, highways, television, movies, or communication technology? While rural communities eagerly adopt cutting edge technology and entertain ment, their people argue they can have the best of both worlds. They can be selective about what they take as their own, while preserving the intimacy and support that makes their community form distinctive in this rapidly changing world.

Issues of Viability and Community Development

Rural communities worldwide face an array of challenges to their economic and social viability. Declines in agri culture and the extractive industries and the movement of manufacturing jobs globally generate problems with retaining living wage jobs, managing the out migration of youth, mitigating loss of a distinctive cultural and eco nomic heritage, or losing local control over natural re sources. Grass root strategies innovated by rural places underlay the early community development movements and theory. Exploitation of cultural and natural amenities is a current development scheme for attracting tourism, recreation, sports, or shopping. Similarly, marketing local cuisine and music, museums showcasing historic settle ment patterns, a rural way of life, or battlefields are drawn on for economic development purposes. Conflicts emerge with community development strategies that involve exurban, upscale scattered site housing, second home owners, hobby farmers, or amenity migrants. Cultural clashes between longtime, permanent residents and newcomers are common. In addition, environmental costs in the forms of heightened fire danger , depleted water resources, and threatened wildlife habitat arise in the context of low density exurban residential development.

More recently, the approach of sustainable rural community development has gained momentum worldwide, even as definitions for sustainable community development remain unclear. In general, this approach seeks to balance environmental sustainability with eco nomic and social/societal aspects of community devel opment. To achieve sustainability, that is, development that ensures future viability; a regional perspective is adopted that recognizes the interconnectedness of eco logical, economic, and social systems. The environmental justice movement, global feminist movement, and com munity based economic development approaches have shaped efforts to attain sustainable rural communities. Such efforts are reflected in the development of worker cooperatives and micro loan funds targeting women, for example. An emphasis on greater local governance for natural resource management is also a common charac teristic of this approach to rural community develop ment. Innovation and entrepreneurship in developing new products for niche markets by enhancing existing assets (e.g., organic produce or range fed beef) are critical for successful sustainable development. In the US, such programs seek to build alliances between rural and urban population centers to counter the political power of suburbs and reinforce greater understanding of the vital contribution rural places make to national diversity.

Planning and the Rural Community Icon

Several planning and community development schemes have origins in, and continued connections to, the settlement patterns and social processes of an ideal type rural community. 'New urbanism' promotes the housing clusters, small street setbacks, front porches, and common spaces characteristic to small towns and rural com munities. Such design arrangements seek to encourage face to face interactions and the kind of watchfulness, trust, and a sense of community important to ideal small town life. The urban 'smart growth' movement similarly borrows from typical rural community settlement pat terns. Smart growth promotes urban growth boundaries that ensure clustered development rather than unfettered sprawl. These developments attempt to integrate work and residential life and thereby foster an interactive community. More recently, aspects of the smart growth movement are being adopted by rural communities as 'smart land use' – that is, rural land use practices that promote, in particular, the clustering of development and the retention of open spaces. Such practices are common for rural communities located in the urban periphery.

See also: Community; Migration; Neighborhoods and Community; Poverty, Rural; Social Capital; Urban-Rural Continuum.

Further Reading

Brown, D. L. and Swanson, L. E. (eds.) (2003). *Challenges for Rural America in the Twenty First Century*. University Park, PA: The Pennsylvania State University Press.
Childress, H. (2000). *Landscapes of Betrayal, Landscapes of Joy: Curtisville in the Lives of Its Teenagers*. Albany, NY: SUNY Press.
Duncan, C. M. (1999). *Worlds Apart: Why Poverty Persists in Rural America*. New Haven, CT: Yale University Press.
Falk, W. W. (2004). *Rooted in Place: Family and Belonging in a Southern Black Community*. New Brunswick, NJ: Rutgers University Press.
Green, G. P. and Haines, A. (2002). *Asset Building and Community Development*. Thousand Oaks, CA: Sage Publications.
Hoff, M. (1998). *Sustainable Community Development: Studies in Economic, Environmental, and Cultural Revitalization*. New York: Lewis Publishers.
Lapping, M. B., Daniels, T. L. and Keller, J. W. (1989). *Rural Planning and Development in the United States*. London: Guilford Press.
Netting, R. M.cC. (1993). *Smallholders, Householders: Farm Families and the Ecology of Intensive, Sustainable Agriculture*. Stanford, CA: Stanford University Press.
Popper, F. J. and Popper, D. E. (1994). Great Plains: Checkered past, hopeful future. *Forum for Applied Research and Public Policy* 9, 89 100.
Putnam, R. (2000). *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon and Schuster.
Salamon, S. (1998). Cultural dimensions of land tenure in the US. In Jacobs, H. M. (ed.) *Who Owns America? Social Conflict over Property Rights*, pp 159 181. Madison, WI: University of Wisconsin Press.
Salamon, S. (2003). *Newcomers to Old Towns: Suburbanization of the Heartland*. Chicago, IL: University of Chicago Press.
Tocqueville, A.de. (1990). *Democracy in America* (first publ. 1840). New York: Vintage Classic Books.
Tonnies, F. (1957). *Community and Society* (first publ. 1887; transl. Loomis, C. E.). New York: Harper Torchbook.
Vidich, A. J. and Bensman, J. (1958). *Small Town in Mass Society: Class, Power and Religion in a Rural Community*. Princeton, NJ: Princeton University Press.
Warren, R. L. (1987). *The Community in America* (3rd edn.). Lanham, MD: University Press of America.

Relevant Websites

http://www.carseyinstitute.unh.edu
Carsey Institute at University of New Hampshire.
http://www.cfra.org
Center for Rural Affairs.
http://www.esrs.hu
European Society for Rural Sociology.
http://www.irsa world.org
International Society for Rural Sociology.
http://www.rene net.org
Rural Extension Network.
http://www.rupri.org
Rural Policy Research Institute.
http://www.ers.usda.gov
US Department of Agriculture, Economic Research Service.
http://www.comm dev.org
USA Community Development Society.
http://www.ruraledu.org
USA Rural School and Community Trust.
http://www.ruralsociology.org
USA Rural Sociological Society.

Rural Geography

M. Woods, Aberystwyth University, Aberystwyth, UK

Glossary

Counterurbanization The movement of population from urban to rural areas, normally associated with urban to rural migration, but also produced by differential rates of population growth.

Cultural Turn A period in human geography in the late 1980s and 1990s marked by increasing interest in the study of cultural trends and phenomena; the adoption of concepts and methods from cultural studies; and the introduction of postmodern and poststructuralist theories into the discipline.

Political-Economy Approach An approach to analyzing and theorizing geographical observations of the world derived from neo-Marxist structuralist theory which emphasizes the connection of social and political phenomena to the capitalist economy and the imperative of the accumulation of capital.

Rural A term that describes nonurban geographical areas and the social and economic activities, lifestyles, cultures etc., associated with them. Rural areas have traditionally been identified with agriculture, but the precise definition and delimitation of 'rural' is heavily debated.

Rural Restructuring The reshaping of social and economic structures in rural areas during the late twentieth century and early twenty-first century produced by various, interconnected processes of change including the declining economic significance of agriculture, the rise of the service sector, urban to rural migration, and so on.

Subdiscipline A body of academic work that focuses on a particular specialist field of study within a broader discipline.

Introduction

Rural geography may be simply defined as the study of people, places, and landscapes in rural areas, and of the social and economic processes that shape these geographies. This simple definition, however, is problematic in that it describes rural geography as a field of study in relation to its territorial focus on 'rural areas' – a categorization of space which itself is often imprecise and contested. Indeed, the definition of 'rural' and the delimitation of 'rural areas' have been recurring preoccupations for rural geographers. This work has developed over time from mapping rural regions and constructing statistical indices of rurality to the examination of social and economic processes that might produce distinctively 'rural' phenomena, to the more recent recognition of 'rurality' as a social construct and analysis of the discourses through which rurality and rural areas are imagined, represented, and contested. As such, rural geography today might be better defined as the study of the processes through which rurality is produced, reproduced, and contested, and of the places and practices that are associated with 'rural' ways of being.

The scope of rural geography is hence extremely broad, encompassing research stretching from economic analyses of farm businesses and agricultural trade to cultural analyses of media representations of rural life. In common with all academic fields of study, rural geography is also a dynamic subdiscipline, evolving over time in terms of its key concerns and major theoretical and methodological approaches. There are, however, a number of 'traditional' concerns that can be identified as consistent areas of research by rural geographers over the past four decades. These include:

Agricultural geography. The study of farm landscapes, businesses, and households, and of agricultural trade and policy, has long been one of the primary concerns of rural geography. The focus of work on agriculture by rural geographers has however shifted, reflecting both changes in agriculture and agricultural policy and the influence of theoretical and methodological developments in rural geography. Early work on the geographical distribution of agricultural types was superseded by analysis of agriculture as a capitalist industry, including the rise of global agribusiness. Policy reforms during the 1980s and 1990s, notably in the European Union, prompted debates about the nature of a so called 'postproductivist transition' and adoption of 'multifunctionality' in agricultural policy, supported by empirical research on farm diversification and the implementation of agri environment schemes. Recent work has also introduced a more cultural perspective into agricultural geography, exploring, for example, gender roles in farmwork, human–livestock interactions, and the influence of consumer cultures in shaping markets for agricultural produce.

Rural resource management and conservation. Alongside agriculture, rural geographers have also traditionally studied other primary industries such as forestry and mining, as well as the management and conservation of natural resources in rural areas, including water resources. The nature of work in this subfield has also shifted over time, with an increasing emphasis on conservation from the 1980s onward, as well as on the repackaging and exploitation of traditional rural resource sites for tourism and recreation. Recent work continues to exhibit a strong interest in conservation, but has also explored both the neoliberal restructuring of rural resource industries and the development of alternative practices of resource use and management, such as community forestry.

Rural land use and planning. The study of geographical patterns of rural land use and questions about how rural land use should be planned and regulated have also been long standing concerns of geographers. This body of work has always had a strong applied focus, seeking to engage with and contribute to land use planning strategies and policies. In particular, rural geographers have focused on the control of settlement growth and development; the organization and regulation of industrial land uses; and the protection of the rural environment through planning regulations and the designation of protected areas such as national parks. From the 1980s, a more critical strand of research on this theme developed, concerned not with contributing to planning policy but with analyzing the political processes through which planning policy is constructed and implemented, including the role of the state.

Population, migration, and social structure. The fourth traditional area of work by rural geographers has followed the shifting population trends of rural areas, with an initial focus on rural de population moving to the new phenomenon of counterurbanization. At one level, rural geographers have enumerated and mapped migration patterns in rural areas, including comparative studies between different regions and countries. Research has, however, also sought to analyze and understand the processes shaping migration patterns, leading, in the early 1990s, to work exploring migrants' perceptions of rurality and the influence of these perceptions on decisions to move and on their integration in rural communities. Rural geographers have also studied the consequences of population change on the social structure of rural communities. In the 1980s and 1990s this work focused primarily on class recomposition, particularly the significance of the 'service class' in counterurbanization, and the accompanying gentrification of rural communities. More recently, issues of age and ethnicity in the composition of the rural population have also begun to be examined more fully. Additionally, a more critical perspective on counterurbanization has been advanced in recent research, highlighting the complexity of rural population trends and patterns.

Rural economic development. In addition to research focused on agriculture and other primary industries, rural geographers have also been concerned with the restructuring of the wider rural economy, and the rural development strategies adopted to assist rural areas in adjusting to economic change. In part, this work has charted and analyzed the fluctuating fortunes in rural regions of nonagricultural economic sectors such as manufacturing, telecommunications, and the service sector. However, as with work on planning, engagement with rural development has had a strong applied dimension with rural geographers involved in formulating rural development strategies, evaluating rural development programs and initiatives, and working with rural communities on rural development projects. Alongside this applied engagement, a more critical perspective has again been developed that has examined the political processes shaping rural development policies and strategies, connecting moves toward more 'bottom up' or endogenous strategies in rural development to shifts in the mode of governmentality and the rise of neoliberalism.

Rural settlement patterns and communities. The study of rural settlement patterns and their historical evolution is one of the oldest areas of geographical research on rural areas, but its significance in Anglophone rural geography dwindled in the 1970s and 1980s and is now at best marginal to the contemporary subdiscipline. It has retained greater significance in some countries, such as Germany where models of rural settlement patterns continue to be integrated into rural planning studies. In general, the interest of rural geographers has shifted from the physical form of rural settlements to the social dimensions of the rural community, particularly in the context of rural restructuring. The adoption of ethnographic techniques to explore these issues in some recent research has reconnected with an older tradition of rural community studies that was present in both geography and sociology in the immediate post war period, as discussed further below.

Rural infrastructure. Rural geographers have undertaken a number of studies of rural infrastructure provision, including transport networks, electricity and water supplies, public and commercial service provision, and, more recently, IT connections. Studies on this theme have varied from the mapping and modeling of infrastructure provision; to critical analysis of the influence of political trends, such as the de regulation and privatization of public services; to examination of the impact on rural communities and their residents of the closure of local services and facilities. The study of rural housing provision and quality, including the provision of public housing, has also been a long standing concern in rural geography.

Recreation and tourism. The final 'traditional' concern of rural geography is the study of recreation and tourism in rural areas, which has grown in significance as the contribution of agriculture to the rural economy has declined. Early studies tended to focus on the management aspects of rural recreation and tourism, including the environmental impacts, and while this strand of work has continued, it has been supplemented by more critical perspectives. Research in the 1990s, for example, drew on the Marxist concept of commodification to examine the repackaging of rural sites and rural experiences for consumption by tourists and visitors. More recent studies have focused on the experiences of tourists and recreational users of rural space, from walkers to adventure tourists, exploring their perceptions of rurality and their embodied interactions with the rural environment.

Alongside these eight 'traditional' concerns, a further four concerns can be identified as emerging during the development of rural geography from the 1970s onward, and remaining significant areas of study today. These are discussed below.

Poverty, deprivation, and social welfare in rural areas. Although there is a longer history of research on rural poverty in rural sociology, this did not really become a significant concern for rural geography until the 1980s. Indeed, early work in this field involved a critique of the role of both lay and academic discourses in obscuring the presence of poverty and disadvantage in rural communities. Subsequent research has sought to record and describe the conditions of poverty in rural areas and to examine the political and economic processes creating these conditions. As well as fiscal poverty, this work has also explored a range of forms of disadvantage and exclusion, including homelessness, transport poverty, and ill health. Later research has further examined the relations and dynamics of social exclusion within rural communities, and the experiences of those who are excluded and their perceptions of rural life.

Rural governance and politics. The development of a more critical approach in rural geography in the 1980s fueled a growing interest by rural geographers in the political processes shaping rural economies and society. This resulted in research on the operation and apparatuses of the state in rural areas, and analyses of rural policymaking processes, applying the concepts of corporatism and policy networks. More recently, research has examined the restructuring of governing processes and institutions in rural areas and the development of a 'new rural governance' with an emphasis on partnership working and community participation. Rural geographers have also traced and explored the emergence of social and environmental conflicts in rural areas, often produced by tensions created by social and economic restructuring, and studied the political mobilization of rural actors against perceived threats from policy reforms, trade agreements, and other external pressures.

Social difference and experiences of rural life. The third 'new' area of research in rural geography recognizes the multiplicity of social groups, and consequently of experiences, in the countryside. Research on gender roles and relations was developed in the late 1980s and moved from an initial focus on farm women and women in employment, to a broader range of studies including work on rural masculinities. The wider exploration of social difference in the countryside was however stimulated by a review article by Chris Philo published in the *Journal of Rural Studies* in 1992 – discussed further below – which critiqued the narrow focus of mainstream rural geography on middle aged, middle class, white, able bodied men. Subsequent studies of 'other' groups in rural society have examined a wide range of groups, including, among others, children and young people, the elderly, ethnic minorities, people with physical and mental disabilities, and sexual minorities, as well as marginalized 'lifestyle' groups such as new age travelers. Collectively, this research has helped to produce a more nuanced, multivocal representation of contemporary rural life.

Rural culture and media representations. The final key area of work in rural geography covers a fairly disparate body of studies exploring the articulation of rurality and rural identity through various cultural practices, and the representation of rurality through the media. These include work on media representations of rurality and rural life through newspapers, magazines, film, television, and literature, and their role in reproducing ideas of the rural idyll and establishing norms for rural society. Other research has examined the articulation and reproduction of rural identities through local festivals and events, participation in particular activities such as hunting, and the reiterative performance of everyday routines, such as milking cows, that constitute a distinctive rural 'lifescape'.

Cutting across the above areas of work in rural geography is the underlying theme of change. The changing countryside of the developed world, characterized by declining agricultural employment and agricultural restructuring, economic readjustment, migration and social recomposition, political reforms, the rationalization of local services, environmental pressures and changing public attitudes, has formed the context for the bulk of the work undertaken by rural geographers. Moreover, the empirical observation of processes of change and their effects in rural areas has been a key element driving the development of rural geography as a subdiscipline. At the same time, the development of rural geography has also been influenced by its engagement with other areas of human geography and with cognate disciplines including agricultural economics and rural sociology, particularly through the import of new conceptual and methodological approaches. This subdisciplinary history is discussed further in the remaining sections of this article.

The Birth of a Subdiscipline

In comparison with many other fields of geography, the consolidation of rural geography as a coherent, distinctive subdiscipline is a relatively recent development. The first textbook in English on rural geography, the appropriately titled *Rural Geography: An Introductory Survey,* by Hugh Clout, was not published until 1972, with arguably the first 'rural geography' textbook published in the United States, *The Look of the Land,* by John Fraser Hart, following in 1975. The Rural Geography Study Group (RGSG) of the Institute of British Geographers (IBG) came into being in 1974 with the renaming of the Agricultural Geography Study Group, but in the Association of American Geographers, a Rural Geography Speciality Group was not established until 2002 with the merger of the Contemporary Agriculture and Rural Land Use Speciality Group (CARLU), and the Rural Development Speciality Group.

Clout's book is regarded by many as the founding document of Anglophone rural geography because it presented the first systematic framework for the integrated geographical study of rural areas. With chapters covering rural de population; people in the countryside; the urbanization of the countryside; land use planning; structural changes in agriculture; forestry; landscape evaluation; settlement rationalization; manufacturing; passenger transport; and the integrated management of the countryside (which included economic development), the volume was wide ranging and brought together previously disparate areas of research in geography. Clout outlined a model for the components of 'rural geography' in which farmland, forest, water, and settlement were all linked as the land use components of the rural landscape, with the study of farmland feeding back to analysis of the farm population, market demands, equipment, land, tenure conditions, farm size and layout; and the study of rural settlements encompassing the range of services, transport, employment, and both the farm and nonfarm rural populations. Hart's book, though narrower in focus than Clout's and less clearly branded, nonetheless brought together elements identified as components in the moulding of the rural landscape: plant cover, land division, farm size and tenure, farm management and decision making, farm buildings, rural settlements, mining, forestry, and recreation.

The development through these publications and models of a coherent entity called 'rural geography' in the early 1970s was assisted by growing awareness at the time of the changing nature of the countryside in Britain and North America. Both Clout and Hart reflected public concerns about the urbanization of rural areas; the impact of increased mobility and leisure time for both rural and urban populations; the consequences of agricultural modernization and broader rural economic restructuring; and the rationalization of rural services. Moreover, analysis of population statistics in the United States would shortly reveal the first signs of counterurbanization.

Yet, Clout in particular was also influenced by more immediate factors, notably pressures for human geography in British universities to demonstrate its contemporary relevance, and Clout's own desire to consolidate his research interests on rural Britain and France into a distinctive niche. As a consequence, either intentionally or inadvertently, Clout set two parameters for Anglo American rural geography that have largely held to this day. First, that rural geography is focused on the contemporary countryside (or at least the post war countryside), omitting the historical study of rural landscapes, settlements, and people, which is left to historical geographers and rural historians. Second, that rural geography is focused on the rural areas of the developed world, with relatively weak ties to the study of rural areas in developing nations undertaken by development geographers. This latter point is less true for rural geography outside Britain and North America, with rural geographers in countries such as France, Germany, and New Zealand in particular tending to undertake research in both developed and developing world contexts.

The study of rural areas by geographers, however, long pre dated the emergence of a distinctive subdiscipline of rural geography. The regional approach, which dominated geography in the early to mid twentieth century, was largely a *de facto* study of the economies, land use, and settlement patterns of rural areas. The study and classification of rural land use, in particular, was a key focus for human geography at this time, associated most notably with the work of the British geographer Dudley Stamp and his 'land utilization survey' of Britain during the 1930s (which was to be a major influence on British agricultural planning during World War II and on the establishment of town and country planning system in the post war period), as well as with studies based on the concentric model of the nineteenth century German economist Johann Heinrich von Thünen.

As geography began to fragment into subdisciplines, the study of rural areas continued to be a major concern for specialists within several fields. Rural settlement patterns, for example, were seen as central to the work of both historical geographers – undertaken largely through numerous local studies of historical settlement patterns – and population geographers, with a population and rural settlement group being one of the first research committees established by the Association of American Geographers. The more systematic approaches to settlement analysis developed in the central place theories of Walter Christaller and August Lösch were influential in the development of rural geography in Germany and Scandinavia, but are generally regarded as mainly contributing to a new quantitative urban geography in the 1950s and 1960s. Similarly, ethnographic studies of rural communities were a distinctive aspect of mid–twentieth century social geography; rural land scapes were a frequent focus for the Berkeley School of cultural geography; while the study of agriculture was a major component of economic geography until the 1950s.

The development of systematic and quantitative ap proaches in geography, however, refocused the attention of geographers away from rural areas and on to urban systems. One response was the attempt to create a more systematic agricultural geography. As noted above, agri culture had been a focus of study by economic geog raphers since the early years of the twentieth century, including research on the global geographies of par ticular commodities, associated with British geography, and regional analyses of crop combinations and farm problems in particular areas, associated with continental Europe. The regional approach was extended by American geographers such as Derwent Whittlesey in the 1930s, with quantitative models developed to aid com parative analysis of different regions. These models became the basis for the assertion of a distinctive, sys tematic agricultural geography as economic geography moved away from agricultural concerns. A committee on 'types of farming' was the first study group to be estab lished by the IBG, in 1960, and had by 1968 been joined by a 'terminology of the agrarian landscape' study group. These were to combine into the 'agricultural geography study group' in 1969.

Yet, agriculture was by this time a rapidly declining presence in the rural economies of developed nations. The new 'rural geography' developed in the 1970s sought to integrate agricultural geography with studies of other dimensions of rural societies and economies, but the terms 'agricultural geography' and 'rural geography' were often used interchangeably into the 1980s. Furthermore, analysis of the register of rural research periodically compiled by the IBG RGSG shows that agriculture continued to be the single largest area of work in British rural geography throughout the 1970s. Not until the early 1980s did a significant shift toward studies of rural development and planning, and of the rural population, become discernable (**Table 1**).

The trajectory of rural studies in geography observed in Britain and the United States was paralleled to a greater or lesser degree in other countries, albeit shaped by national contexts. In France, the study of rural *'pays'* had been a fundamental feature of early French geog raphy. More thematic studies of rural landscapes, rural populations, and agriculture emerged in the 1930s, forming a nascent rural geography, although the term was more narrowly employed in 1963 for Pierre George's book *Précis de géographie rurale* to refer to the economic geography of agriculture. These studies had a strong influence on Hugh Clout, who worked extensively in France, though he was careful to distinguish his definition of 'rural geography' as going far beyond that of George. A

Table 1 Projects listed in the 'rural geography study group' (RGSG) register of research

Category	*Percentage of projects*			
	1973/4	*1978*	*1981*	*1987*
Agriculture	40.6	40.2	25.0	21.8
Forestry, fishing, and mineral extraction	1.6	4.0	7.8	3.8
Rural settlement	19.7	16.1	16.4	10.7
Rural population	12.0	10.3	15.2	18.0
Rural transport	3.6	4.5	5.3	3.8
Recreation and tourism	6.8	11.6	9.0	9.7
Development and planning	15.7	13.4	21.3	32.0
Total number of projects recorded	n/a	224	244	206

From: 1973/4 & 1978: Clark, G. (1979). Current research in rural geography. *Area* 11(1), 51 53; 1981: Clark, G. (1981). *Register of Research in Rural Geography*. Lancaster: RGSG; and 1987: Robinson, G. (1987). *Register of Research in Rural Geography*. Occasional Paper No. 6. University of Edinburgh, Department of Geography.

more synoptic approach was developed in the 1970s, marked by the publication of an integrative textbook, *Géographie rurale*, *methods et perspectives* by Bonnamour in 1973, the establishment of chairs in rural geography at several French universities, and a special series, *Etudes Rurales*, which provided a vehicle for the dissemination of work exploring the 'urbanization' of rural France and its integration into wider economic structures.

In Germany, by contrast, rural geography was con solidated with more restricted parameters, its primary focus being on rural settlements and spatial planning. This partly reflected the continuing influence of Von Thünen's and Christaller's models, but also the strongly applied nature of German rural geography and its active engagement in the spatial planning process.

The potential for rural geography to contribute to policymaking and implementation was also noted in Britain. A reflective paper delivered by the then chair of the RGSG, Alan Williams, to the IBG annual conference in 1977 argued for the RGSG to act as a platform for disseminating the expertise of rural geographers to pol icymakers, lamenting the failure to present a corporate submission in recent public consultation exercises on agricultural landscapes and transport policy. Williams suggested that rural geographers should seek to con tribute, both individually and corporately, to similar na tional consultations in future, as well as to local inquiries on structure plans, new rural motorways, and the like.

For other rural geographers, however, the future de velopment of the subdiscipline lay not in applied en gagement with policymakers, but in engagement with the new theoretical perspectives being advanced elsewhere in geography and the broader social sciences.

Critical Rural Studies and the Political-Economy Approach

The new 'rural geography' may have enjoyed a brief period of fashionability (the RGSG was the fourth largest study group in the IBG by 1983, with 280 members), but it remained peculiarly detached from the theoretical developments underway elsewhere in human geography, notably the rise of a Marxist political economy perspective stimulated by David Harvey's work in urban geography. Research in rural geography was resolutely empiricist and some rural geographers sought to explicitly present the subdiscipline as being 'technically neutral' and 'theory free'. In fact, rural geography was dominated by 'applied positivism', a conceptual approach that was for some satisfyingly clean from overt politics and ideology, but which was nonetheless most certainly not 'theory free', as critics such as Henry Buller, Keith Hoggart, and Richard Munton pointed out.

For many younger researchers drawn to rural geography in the 1970s, the lack of conceptual engagement limited the advancement of the subdiscipline. A review of the field by Paul Cloke published in *Progress in Human Geography* in 1980, for example, commented that rural geographers had been 'tardy' and 'uninspired' in the introduction of rigorous analytical methods, and argued that some form of conceptual framework was required on which to found strands of investigation and analysis.

In searching for a conceptual framework, critical rural geographers were influenced by Marxist geography, but also, more directly, by the development of a 'critical rural studies' through the work of British sociologist Howard Newby and colleagues. Newby's research on farmworkers in East Anglia may have been serendipitous (he was primarily interested in subservience and turned to farmworkers only when his initial proposed project on domestic servants fell through), but his class analysis of the relations between laborer and farmer (published in 1977 as *The Deferential Worker*), and his later extension of the work to study the rural power structure and its ties to property and labor relations (*Property, Paternalism and Power*, 1978), demonstrated the applicability of political economy approaches to the analysis of rural economies and societies.

Dialog between critical rural sociologists and sympathetic rural geographers in Britain was facilitated by the creation in 1979 of the 'rural economy and society study group' (RESSG), which had as a core aim the encouragement of theoretically informed investigation and analysis of rural issues. Through a series of conferences and publications, the RESSG formed a space for the propagation of a political–economic approach in British rural studies separate to the mainstream of rural geography which remained wedded to applied positivism. Membership of the new group rapidly exceeded 100.

A further vehicle for critical rural studies was established in 1985 with the founding of the *Journal of Rural Studies*. Constituted as an interdisciplinary journal, the *Journal of Rural Studies* nonetheless became in effect the flagship periodical for rural geography, helping to position rural geography at the center of critical rural studies, in Britain at least. Similarly, although not explicitly aligned with any theoretical perspective, the opening editorial by Paul Cloke noted the conceptual conundrums confronting rural studies and solicited further contributions to these debates.

The political economy approach presented rural geographers with new perspectives on the exploration of traditional areas of study, and four key concerns soon emerged. First, rural geographers contributed to work on the political economy of agriculture. In Britain, this was exemplified by the program of research undertaken by Richard Munton, Terry Marsden, Sarah Whatmore, and others on the capitalism subsumption of family farming, recognized as making important contributions to explaining the geographical and temporal specificity of agricultural structures that moved the theoretical debate beyond deterministic structuralism.

Second, critical analyses of rural planning and economic development processes led to work on the political economy of the state, and its role in reproducing the conditions for capitalist exploitation in rural areas, including through the planning system, agricultural policy, and the rationalization of rural services. Third, rural geographers developed analyses of class relations and social inequality in rural areas, informed by structuralist theory. This included innovative research on rural poverty and disadvantage, but also research on the burgeoning rural middle classes and the processes of class recomposition in rural communities. One key aspect of this work was the development of a critique of Newby's thesis of property as the basis for rural class relations, and the assertion of a new occupation based class hierarchy framed around the 'service class'. Fourth, critical rural geographers emphasized the centrality of capital restructuring to processes of rural social and economic change. In particular, this work interrogated realignments in the spatial division of labor, as expressed through urban to rural shifts in manufacturing industry (and later service sector industries), and population movement under counterurbanization. Later work also drew on regulation theory to position and interpret rural restructuring in the context of the shift from Fordist to post Fordist modes of regulation in the global economy.

Collectively, these strands of research generated lively debate in British rural geography in the 1980s and 1990s about the nature, interpretation, and conceptualization of rural restructuring. However, the political economy approach was by no means the prevailing perspective within British rural geography during this period, and its percolation through rural geography outside Britain was generally limited to individual geographers and research groupings. In the United States and continental Europe in particular, structuralist critical rural studies found its home in rural sociology, not in rural geography.

One limiting factor in the spread of the political economy approach was the problem that it ultimately posed for the treatment of 'rurality' in rural geography. Structuralist analyses argued that social and economic outcomes in rural areas were the product of large scale overarching structures, moderated by particular locality effects. The intervening scale of the 'rural' was stripped of any explanatory power, and even questioned as legit imate or useful classification. Paul Cloke and Jo Little, for example, in a study of the planning system as an ex pression of state power concluded that there were no discernable 'rural' effects in the operation of the state, only local factors in their book, *The Rural State*. The implication was taken to its logical conclusion by Keith Hoggart in a paper in the *Journal of Rural Studies* in 1990, provocatively titled 'Let's do away with rural', in which he argued that 'rural' was an 'obfuscatory' and 'chaotic' concept that had no analytical purpose.

The Cultural Turn and the Renaissance of Rural Geography

Political economy or structuralist perspectives had been the most visible expression of critical rural studies, but it was not the only attempt to bring social theory into rural geography. Humanistic approaches were also employed, particularly in work on rural communities and social geographies, and positioned as a counter to structuralist explanations. Perspectives from feminist theory were introduced in the late 1980s as a critique of the absence of gender in political economy accounts of rural change, and applied in a new body of research on gender relations in both agriculture and rural communities. Similarly, other perceived limitations in political economy ac counts were addressed by borrowings from cultural studies that were part of the broader 'cultural turn' in human geography at the end of the 1980s and early 1990s. For example, work on the 'service class' began to explore their cultural practices and consumption preferences; while research on rural tourism sites combined the Marxist concept of commodification with cultural an alysis of promotional material. In a progress report on rural geography published in *Progress in Human Geography* in 1991, Sarah Whatmore anticipated a flourishing of conceptually rich research drawing on political economy perspectives, feminist scholarship, and post structuralist theory.

Of particular importance was the influence of social constructivism, imported from postmodern and post structuralist thinking. By recognizing the 'rural' as a so cial construct – that is, as an entity that exists only in terms of the meanings attributed to it in representation – rural geographers were able to reclaim the 'rural' as a legitimate object of analysis and opened up new areas of study. For instance, work by Keith Halfacree explored the importance of social constructions of rurality in coun terurbanization, producing a nuanced understanding of the motivations of migrants. Representations of rurality more broadly were explored in a number of contexts, drawing on work on representation in cultural geography that ultimately owed its lineage to hermeneutics.

The defining episode in the renaissance of rural geography, however, was an exchange between Chris Philo and Jonathan Murdoch and Andy Pratt in a series of articles in the *Journal of Rural Studies* in 1992, 1993, and 1994. The first of these, by Philo – an outsider to rural geography – was a review essay inspired by Colin Ward's book, *The Child in the Country*. Philo observed that Ward's account provided insights into the experiences and nar ratives of rural children that were absent from a rural geography which, he contended, was in danger of por traying rural people as 'Mr Averages' – men in employ ment, white, without sexuality, healthy and able in body, and devoid of religious or political identity. Noting the emerging sensitivity in social and cultural geographies to the multiple perspectives of 'other' human groupings, he called for rural geography to engage with the 'neglected rural geographies' of 'other' social groups from children to tramps.

Responding to Philo's article, Murdoch and Pratt sought to take his argument a step further by suggesting that rural geography needed to take postmodernism more seriously. Their paper explored the production of knowledge about rural areas, presenting an 'archaeology' of modernist social science research from early com munity studies to the political economy approach and locality studies, which had systematically neglected certain social groups, cultures, and identities. Rural geography, they argued, would not be able to satis factorily address Philo's concern for 'neglected rural others' until it escaped this modernist narrative. Hence, they proposed the adoption of a 'sociology of post modernism' with a reflexive perspective that rejected universal concepts such as 'rural' and which instead examined the power by which certain actors imposed their 'rurality' on others. They labeled this new approach the study of the 'postrural'.

The consequences of the Philo/Murdoch and Pratt exchange were threefold. First, Philo's call to arms was answered by a plethora of studies in the mid and late 1990s that explored the 'neglected rural geographies' of social groups including children and young people, the elderly, ethnic minorities, people with physical and mental disabilities, and sexual minorities, as well as marginalized 'lifestyle' groups such as new age travelers, among others. Second, Murdoch and Pratt's postmodern perspective, though less readily adopted, generated both analyses of the production and circulation of discourses of rurality, and the further exploration of 'rurality' as a networked or hybrid form. Third, the ongoing theoretical debate and the theoretically informed empirical studies prompted by the exchange briefly moved rural geography from a marginal position chasing the mainstream of contemporary geographical thought to a position nearer the center of intellectual development in geography.

The transformation was acknowledged by Paul Cloke in an editorial for the *Journal of Rural Studies* in 1997, in which he noted an increased number of graduate students working on rural topics, a heightened production of publications on rural themes, and the interest in rural questions of nonrural geographers, as evidence of a re surgence of rural geography in Britain. Cloke proceeded to outline four areas of 'current excitement' in rural geography, all of which reflected the influence of the cultural turn. These were, first, research on the inter connections between nature and society including studies of nonhuman agencies; second, investigation of dis courses of rural experience and imaginations, including work on 'othered' groups; third, the analysis of symbolic texts of rural cultures, including pastoral art, novels, country magazines, films, television dramas, country furnishings, advertisements, country music, and so on; and finally, emerging research on 'movements', including new approaches to studies of rural tourism.

The introduction of new concepts in rural geography during the 1990s was not, however, limited to post modern and post structuralist theory. Political economy perspectives also continued to be drawn upon in new ways. These included a significant body of work on the so called 'new rural governance' which engaged with the concept of 'governance' developed by the sociologist Bob Jessop and employed in urban geography; and new re search in agricultural geography which engaged with concepts such as commodity chains and food regimes, influenced by critical economic geography and regulation theory. Additionally, political economy and cultural perspectives were combined in areas such as a renewed rural social geography exploring issues of poverty and social exclusion, as exemplified by the work of Paul Cloke and Paul Milbourne on rural homelessness.

The breadth of rural geography at the end of the twentieth century is indicated by analysis of the articles published in the *Journal of Rural Studies*, which reveals significant clusters of papers on rural governance and politics, gender, 'other' rural social groups, food branding and consumption cultures, and the representation of rurality through the media and cultural practices (**Table 2**). Yet, the analysis also shows the persistence of traditional areas of study including agriculture, economic development, planning, and population changes. The precise details of subjects studied within these broad categorizations demonstrate progression – work in agricultural geography, for instance, was increasingly dominated by studies of farm diversification and agri environment schemes – but not necessarily heightened conceptual engagement. Like the political economy perspective in the 1980s, the cultural turn in the 1990s was limited in the extent of its adoption within rural geography, and particularly in terms of its dissemination outside Britain.

Table 2 Categorization of articles published in the *Journal of Rural Studies*, 1991 2005

Category	*Number of articles*		
	1991 1995	*1996 2000*	*2001 2005*
Agriculture	39	39	30
Governance and politics	10	16	16
Forestry, rural resource management, and conservation	15	13	13
Social difference and 'other' rural geographies	10	10	17
Economic and community development	13	16	8
Nonagricultural rural economy	10	11	9
Population change and migration	9	10	7
Food and consumption	4	2	17
Planning and land use	12	3	5
Rural communities	4	10	6
Culture and media representations	2	3	8
Transport and infrastructure	2	5	5
Definitions and discourses of rurality	4	3	5
Recreation and tourism	4	1	1
Poverty, deprivation, and social welfare	5	1	0
Other	7	8	15

Rural Geography and Cognate Disciplines in International Perspective

If the above account exhibits a distinct British bias, it is a reflection of the distinctive position and experience of British rural geography since the 1970s, which in turn is a product of the disciplinary structure of rural studies in Britain. Up to the 1980s, the institutionally strongest discipline in British rural studies was agricultural eco nomics. Agricultural economics departments had been established at a number of universities in the inter war period as part of a state sponsored drive for agricultural modernization. The Agricultural Economics Society was founded in 1926, and the *Journal of Agricultural Economics* established 2 years later in 1928. Through its application of economic theories and models to agricultural questions, agricultural economics became regarded as the conceptual engine house of British rural studies, and its institutional position helped it to lay claim to areas of research that might otherwise have been studied by sociologists or geographers. Indeed, one of the imperatives for the formation of a systematic agricultural geography was to develop links with agricultural economics, and agricultural economics continued to be influential on agricultural studies in British rural geography up to the 1980s – although its neoclassical economic theories were increasingly at odds with the neo Marxist political economy approach of critical rural studies. Moreover, by espousing neoclassical economics, agricultural economists ultimately undermined their own institutional position as role of the state in agricultural planning and advice was cut back in the 1980s. Many agricultural economists sought to integrate themselves into mainstream economics, and by the 1990s, most agricultural economics departments in British universities had been closed or subsumed within larger schools.

The strength of agricultural economics in Britain, however, militated against the development of rural sociology to any significant degree. Notably, although several key figures in modern British sociology, including Howard Newby, Ray Pahl, Peter Saunders, Colin Bell and John Urry, worked on rural topics and contributed to the emergence of critical rural studies, they tended not to define themselves as rural specialists and moved on to work in urban contexts. With the exception of Newby, none of these figures were particularly active in the interdisciplinary rural studies community, and none developed significant groups of rural sociology graduate students. The institutional weakness of rural sociology in Britain enabled rural geography to establish itself as the primary rural social science discipline as agricultural economics faded, but it also meant that rural geography itself became at least in part an interdisciplinary field. Individual scholars moved between geography, sociology, and planning departments, and it is notable that some of the most prominent centers of rural research in Britain today, including the Centre for Rural Economy at Newcastle and the School of City and Regional Planning at Cardiff are not geography departments, but many of their staff are rural geographers by training and disseminate their work through rural geography publications and conferences.

In the United States, in contrast, rural geography has tended to operate in the shadow of rural sociology. The first university course in rural sociology in North America was taught at the University of Chicago in 1894, and the discipline expanded rapidly in the inter war period. It soon became detached from mainstream sociology, with its own departments, own association, the Rural Sociology Society, and own journal, *Rural Sociology*, founded in 1936. The growth of rural sociology reflected in part, political concerns about the future of rural America under pressure from modernization and urbanization, and rural sociology always included a strong applied focus and a commitment to university 'extension' programs reaching out to farm populations. The key concerns of rural sociology were with the nature of rural society and the problem of 'urbanization'; social relations in rural communities and in agriculture; the farm household and farm life; social provision in rural communities; and rural health and education. Rural sociology in the United States hence monopolized work on the 'social' dimensions of the countryside, including themes that were picked up by rural geographers in Britain. American rural geography consequently differentiated itself by focusing on questions associated with rural 'land' – from the geographies of agricultural production and the exploitation of natural resources to analysis of the rural landscape and land use planning. Such work has often been approached through a regional lens and, reflecting the strength of connections between American universities and their localities, has frequently had an applied dimension involving working with rural agencies and communities, as evident in the careers of figures such as Owen Furuseth and Mark Lapping.

From its beginnings, American rural sociology had engaged with sociological theory, and thus proved more to be more open than rural geography to later theoretical innovations, including critical rural studies. In particular, American rural sociologists such as Fred Buttel and Philip McMichael have made significant contributions to the study of the political economy of agriculture. Yet, as with British rural geography, the extent of uptake of the political economy approach (and later cultural perspectives) was limited and much rural sociology continues to be empiricist and positivist. There are, increasingly, areas of overlap between rural geography and rural sociology in the United States, both in terms of focus and in terms of concepts, but there are also continuing clear distinctions between the disciplines and interaction is relatively limited.

Rural sociology has also tended to be stronger than rural geography in continental Europe, although there are differences between individual countries. As in the United States, rural sociology in Europe developed as a distinctive discipline in the inter war period and was strongly associated with both agricultural modernization and an anti urbanization movement. A pan European association, the European Society for Rural Sociology (ESRS) was established in 1957, with its own journal, *Sociologia Ruralis*. Institutionally, rural sociology in many European countries tends to be housed as the social studies arm of specialist agricultural universities or rural research institutes, coexisting alongside agricultural economics and agricultural sciences, whereas rural geographers tend to be employed in university geography departments. This positioning is reflected in the centrality of agriculture and the farm household to European rural sociology, and the greater emphasis in rural geography on more explicitly 'geographical' themes such as settlement patterns and spatial planning. For instance, Dutch rural geographer Joeke Veldman, writing in 1983, suggested that in the Netherlands, rural geography was primarily concerned with the interaction between the physical material structure and the sociospatial systems of rural areas; rural sociology with the social problems of agriculturally defined areas; and agricultural economics with the prosperity of agriculture.

In some cases, rural geographers and rural sociologists have competed for prominence, for example, in the cartographic representation of rural France; yet, French rural geography also has a strong tradition of interdisciplinary collaboration. Formal interactions between rural geography and cognate disciplines continue to be relatively limited in many other European countries; however, since the 1980s, European Union funded research programs have helped to create new networks that are both international and interdisciplinary. There is also a discernable broadening in the scope of European rural geography. In France, the Netherlands and Sweden, for example, rural geographers have engaged with many of the same new themes adopted in British rural geography, including alternative rural lifestyles, rural governance, the commodification of the countryside, and media representations of rurality. At the same time, the practice of rural geography continues to be informed by particular national contexts. In the countries of Central and Eastern Europe, for example, rural geographers are increasingly addressing issues of agricultural reform, economic restructuring, population movements, the rationalization of rural services, and social polarization, reflecting the impacts of the post socialist transition.

In Australia and New Zealand, rural research is more evenly balanced between geography and sociology and, to a lesser extent, anthropology. Although the disciplines have traditionally had distinctive fields of study, there is now significant overlap in terms of both subject and approach, and a degree of interdisciplinary collaboration. Rural geography in Australia and New Zealand has tended to be more in tune with British rural geography than has been the case in the United States or continental Europe, but work by geographers such as Neil Argent, Chris Cocklin, and John Holmes in Australia, and Richard Le Heron, Warren Moran, and Ruth Panelli in New Zealand, has also reflected the particular contexts of the two countries. Research on agricultural restructuring and the impacts on rural communities has responded to the radical policies of de regulation pursued in both countries during the 1980s and 1990s, while specific cultural and geographical factors have been reflected in work on service provision in remote rural communities; resource management, including dealing with drought; new forms of adventure tourism in rural localities; and the cultural and social position of indigenous communities.

Similar observations can be made about rural geography in Canada, which has also tended to mirror British rural geography more strongly than has rural geography in the United States, and which has also combined conceptually engaged research with concerns that reflect national context, including significant work on marginal communities, forestry, and mining. At the same time, Canadian rural geography has shared the strong regional and applied emphasis of US rural geography, with individuals such as Chris Bryant in Quebec and Greg Halseth in British Columbia actively engaging in the practice of rural development and governance in the province.

The uneven development of rural geography as a subdiscipline has militated against the creation of a strong and coherent international forum for rural geography equivalent to the International Rural Sociology Association. The International Geographical Union's Commission on the Sustainability of Rural Systems was launched in 1992 and now has over 400 members in nearly 50 countries, yet, as its title implies, its emphasis is on agriculture and the rural environment and hence does not represent the full breadth of contemporary rural geography. Similarly, the RGSG in Britain has for three decades organized bilateral meetings with rural geographers in other European countries and in North America, but until recently these meetings have tended to be dominated by traditional rural geography concerns and have not functioned as means of disseminating the more radical innovations in British rural geography. Critical rural geographers in Britain often preferred to build links to like minded rural sociologists, or to explore connections to human geographers outside rural geography engaged in similar work in other countries.

There are, however, signs of growing convergence in rural studies, both of subdisciplines and of national fields of study. Participation at international conferences is increasing and broadening, including a degree of crossing of disciplinary boundaries. Moreover, the internet has facilitated easy access to publications from around the world and across disciplines. Individual rural geographers are hence less constrained than previously by national disciplinary traditions. As such, the current developments and new directions outlined in the next section have the potential to percolate far more extensively through rural geography than has been the case with previous waves of innovation.

Current Developments and New Directions

The study of rural geography, like the contemporary countryside, is characterized by both continuity and change. The continuity comes from the persistence as significant components of rural geography of traditional areas of study, including agricultural geography, rural resource management, rural planning and rural development. The change comes from the development of new areas of study, the shifting focus of work within traditional concerns, and the adoption of new conceptual perspectives. All of these aspects are evident in the current developments in rural geography outlined below. As in previous eras, these developing areas of inquiry have been stimulated in part by new theoretical influences, and in part by the questions posed by the ongoing restructuring of rural economies, societies, and polities.

These two influences have been brought together in recent developments that have infused the study of agricultural restructuring and the study of rural resources with critical inflections. Analysis of the purported postproductivist transition in agriculture by rural geographers such as Brian Ilbery in Britain and Michael Troughton in Canada, and debates over the usefulness of the postproductivist transition as a concept, have led to recent work that has embraced the concept of 'multifunctionality', the idea that rural environments have a range of commodity and noncommodity use values, and that the economic and environmental sustainability of the countryside is best advanced through the balanced exploitation and management of these values. Rural geographers such as Clive Potter and Geoff Wilson have examined debates around the adoption of multifunctionality into the European Union's Common Agricultural Policy and by the World Trade Organization, as well as studying the local implementation of multifunctionality as a principle in various rural localities worldwide. A more critical strand of research is questioning the policy rhetoric of multifunctionality and its social and environmental implications.

A second development in research on agricultural restructuring has been to connect production to consumption through the examination of 'food geographies'. This has included work exploring consumer practices and cultures, and in particular, studies investigating consumer preferences for various forms of certified foods, such as organic produce and produce identified with certain geographical places of production. The latter includes the growing interest in the consumption of local food, and the rise of alternative sites of food retailing, such as farmers' markets. It also includes an expanding body of work on the global agro food industrial complex developed by geographers such as Margaret Fitzsimmons, Richard Le Heron, and Michael Watts, and connecting with similar work by rural sociologists and agricultural economists including Hugh Campbell, David Goodman, and Geoff Lawrence.

Similarly, the traditional field of rural resource geographies has also been revitalized by the introduction of a more critical edge. In particular, recent research has explored the management and exploitation of rural resources in the context of the 'neoliberalization of nature', through which complex nature entities are reduced to commodities through the assignment of a price or value. To some extent, this work has built on earlier studies of the commodification of the countryside, but proponents such as James McCarthy and Scott Prudham also critique the emergence of neoliberal forms of environmental governance, which hold that the only way to preserve nature is to give it a price and assign property rights. A related body of work has examined alternative economic strategies aimed at challenging the neoliberal hegemony by, for example, creating community forests in which local communities are given greater governance responsibility over forests and are to pursue a range of noncommodity goals. This, in turn, engages with recent work on social forestry that seeks to understand how individuals and communities relate socially and culturally to forests, and how they assign forests with noncommodity value.

The interaction of the rural environment with wider dimensions of rural economy and society have also been explored through two further recent developments in rural geography. First, the 'integrated management of the rural environment and its social and economic dimensions' has been the focus of both theoretically informed and applied research. In particular, this includes studies in Britain funded by the government sponsored 'rural economy and land use' (RELU) program, launched in 2003 in the wake of the 'foot and mouth' crisis in Britain and against the backdrop of growing public concern over the impact of climate change. Studies funded under RELU include work on the food chain, flooding, and pollution, as well as the application of sociology of science approaches to analyze the scientific construction of knowledge relating to natural hazards in rural areas. Second, one of the most notable developments in recent years has been introduction of 'a political ecology perspective' into the study of rural North America. This has been led by political ecologists such as Michael Watts and Peter Walker who have turned the techniques and concepts employed in work in developing nations, on to the United States. In part, this development reflects the observation that political ecologists have often explored similar questions about the rural experience in developing nations as rural geographers have explored in the developed world. However, political ecology has also introduced questions about environmental politics and conflict, rural land use and management, the environmental dimensions of rural economic re structuring and the regulation of the rural environment, complementing work exploring the environmental di mensions of rural sustainability by geographers such as Margaret Fitzsimmons and Lisa Harrington.

One stimulus for political ecologists to turn their attention to the developed world was the increasing interconnection of the rural economies, societies, and environments of the Global North and the Global South under globalization. 'Globalization and its rural effects' is itself a major developing area of concern for rural geographies. The broad range of studies encompassed by this overarching theme include work on global com modity chains for rural resources; the role and power of transnational corporations, including global seed companies and supermarkets; the liberalization of global trade in agriculture; global movements of people; and the global dissemination of media representations, and social and environmental values. Some recent research has begun to investigate the cumulative impact of different manifestations of globalization on specific rural localities and the responses of rural residents, asking questions about precisely how rural localities are being remade within what Michael Woods has described as the emer gent 'global countryside'. Additionally, rural geographers are also exploring attempts to construct alternatives to neoliberal globalization, from community supported agriculture projects, to the fair trade movement, to the mobilization of rural antiglobalization protesters.

Alongside studies of antiglobalization protests, the broader 'politics of rural restructuring' continues to be a significant focus for rural geographers. The political mobilization of rural residents in protest activity, and the incorporation of rural conflicts into mainstream national politics, have been observed and studied in a number of countries. Although few comparative studies have been conducted so far, attempts have been made to concept ualize the observed trends, with debates revolving around the coherence of rural protests as a possible social movement, as proposed by Michael Woods. Continuing work on the politics of the countryside also includes ongoing research on the regulation and governance of rural areas, and, in particular, on the development of neoliberal strategies of rural governmentality that pro mote rural self help and partnership working, and assert new roles and responsibilities for rural citizens.

The above developments all broadly build on the political economy perspective; however, the legacy of the cultural turn is also evident in contemporary rural geography. Most notably, this can be seen in interest in the 'performativity of rural life and rural experiences'. Studies within this field include, for example, work on neo peasant lifestyles in France, and on middle class lifestyles in rural Britain. Research has explored the significance of particular 'rural' activities and rituals such as hunting to rural identity and culture, as well as the materiality of rural performances indicated by the cen trality of certain objects as props for a rural lifestyle. There is also growing interest in the place of emotion in rural life, including emotional attachments to particular rural places and ways of being, and the construction of rural 'lifescapes' through the routine reiteration of ac tivities such as milking cattle described by Ian Convery and colleagues. The performances of visitors to the countryside have also been explored, investigating the embodied experiences of particular rural spaces and how these enable tourists to develop connections with rural places, including work by Paul Cloke and Harvey Perkins in New Zealand. Connections between the body and the performance of rural identities are also being explored in the context of gender and youth identities by geographers such as Jo Little.

Finally, the study of 'rural population change and migration' has also been infused with new critical perspectives, reappraising the concept of counter urbanization and exploring the nuances of contemporary patterns of migration. These include, for example, new critical analyses of rural gentrification by Martin Phillips and Darren Smith, and work by Keith Halfacree on nonconventional rural in migration; as well as renewed interest in North America on the concept of exurbani zation and amenity migration. International migration into and between rural areas has additionally emerged as a growing area of research, including the contribution of migrant workers to rural economies and the impact on rural communities. The increasing ethnic diversity of some rural regions as a consequence of migration, not ably in the southern and western United States, is also an expanding area of research, including the develop ment of what Jeff Popke and Rebecca Torres call 'rural cosmopolitanism'.

Cutting across these new directions for research are three developments that are reshaping the practice of rural geography and moving the parameters of the sub discipline. First, the confinement of Anglo American rural geography to the study of the developed world is weakening as more rural geographers have begun to trace the interconnections between the Global North and Global South – for example, in commodity chains; to explore the comparative experiences of rural localities in developed and developing countries; and as concepts and methods from developing world studies have been introduced into developed world rural geography. Sec ond, researchers have begun to explore the 'more than human' geographies of rural areas, engaging with the animals and plants that are part of the hybrid or net worked rural. Third, rural geographers have started to reexplore the boundaries of urban and rural – both the physical boundaries, as examined through research on city hinterlands, and the socially constructed boundaries of 'urban' and 'rural' as cultural forms. French geographers such as Claude Lacour, Sylvette Puissant, and Jean Didier Urbain, in particular, have taken up this challenge, analyzing not only the professed 'urbanization' of the rural, but also a suggested 'ruralization' of the urban.

See also: Animal Geographies; Cloke, P.; Counterurbanization; Dudley Stamp, L.; Gender and Rurality; Historical Geographies, Rural; Identity and Otherness, Rural; Nature-Culture; Political Ecology; Resource Management, Rural; Rural Communities; Rural Populations; Rurality and Post-Rurality; Services, Rural; Tourism, Rural.

Further Reading

Clark, G. (1979). Current research in rural geography. *Area* 11(1), 51 53.

Clark, G. (1981). *Register of Research in Rural Geography*. Lancaster: RGSG.

Cloke, P. (1985). Whither rural studies? *Journal of Rural Studies* 1, 1 10.

Cloke, P. (1989). Rural geography and political economy. In Peet, R. & Thrift, N. (eds.) *New Models in Geography*, vol. 1, pp 164 197. London: Unwin Hyman.

Cloke, P. (1997). Country backwater to virtual village? Rural studies and 'the cultural turn'. *Journal of Rural Studies* 13, 367 376.

Cloke, P., Marsden, T. and Mooney, P. H. (eds.) (2006). *Handbook of Rural Studies*. London: Sage.

Lowe, P. and Ward, N. (2007). British rural geography: A disciplinary enterprise in changing times. In Clout, H. (ed.) *Contemporary Rural Geographies: Land, Property and Resources in Britain: Essays in Honour of Richard Munton*, pp 1 20. London: Routledge.

Murdoch, J. and Pratt, A. (1993). Rural studies: Modernism, postmodernism and the 'post rural'. *Journal of Rural Studies* 9, 411 428.

Philo, C. (1992). Neglected rural geographies: A review. *Journal of Rural Studies* 8, 193 208.

Robinson, G. (1987). *Register of Research in Rural Geography*. Occasional Paper No. 6. University of Edinburgh, Department of Geography.

Sautter, G. and Kayser, B. (1990). French geography and the rural world. In Lowe, P. & Bodiguel, M. (eds.) *Rural Studies in Britain and France*, pp 97 116. London: Belhaven.

Walker, P. A. (2003). Reconsidering 'regional' political ecologies: Toward a political ecology of the rural American West. *Progress in Human Geography* 27, 7 24.

Woods, M. (2005). *Rural Geography: Processes, Responses and Experiences in Rural Restructuring*. London: Sage.

Relevant Websites

http://people2.moreheadstate.edu/orgs/rgsg/index.htm
AAG Rural Geography Speciality Group (US).

http://www.geog.umontreal.ca/igu rural systems/
IGU Commission on the Sustainability of Rural Systems.

http://www.elsevier.com/wps/find/journaldescription.cws home/348/description#description
Journal of Rural Studies.

http://www.geog.plymouth.ac.uk/ruralgeography/default.htm
RGS IBG Rural Geography Research Group (UK).

Rural Populations

K. H. Halfacree, Swansea University, Swansea, UK

Glossary

Demographic Equation The formula to calculate population change: Population growth/decline births – deaths ± net migration.

Gaze A structured/organized way of looking at and interpreting a specific perceived experience.

Other A person or entity who is different or opposite to the (commonplace) self.

Postmodernism The rejection of the grand theories or narratives that underpin modernist social theory, from positivism to humanism to Marxism.

Representation The presentational way humans imagine and understand the world around them.

Rural Idyll A highly positive representation of the countryside that typically emphasizes diverse aspects of both high physical and high social quality, such as scenic beauty and community spirit, respectively.

Weberian Class Analysis The analysis of a social phenomenon where class is seen as of paramount explanatory importance but is defined by the three dimensions of position in the workplace, status, and political power rather than in more narrow Marxist terms of workplace exploitation.

Introduction: Rural Populations Today

Are There Rural Populations?

Beginning on a cautious note, the salience of the idea of 'rural populations' hinges on the largely implicit separ ation of 'rural' as a category from some nonrural cat egory, usually 'urban' or 'the city', whose populations are seen to be different in some significant way. Furthermore, although compact cities originated at least 5000 years ago, it is only in the last few hundred years that their associated condition – the urban – has become wide spread enough to make this category more than some kind of rather unusual subset of a rural norm. For ex ample, consider England in 1801, when outside London, already a metropolis of 900 000 people, only six other towns had populations of over 50 000. In this context, it would not really make much sense to talk of the 'rural population' as compared just to the population. In other words, the idea of rural populations only really emerges in the wake of both extensive and intensive industrial ization and urbanization; 'rural populations' is therefore as modern a notion as 'urban populations'.

Furthermore, here in the twenty first century we re turn to the problematic of whether we (still) have what we can 'scientifically' term rural populations. Today, of course, the issue has become one of whether almost all the human population is now urban, with rural being a category that, as Keith Hoggart memorably put it, social scientists might 'do away with'. To appreciate this further, the reader is directed to the entry in this encyclopedia on 'rurality and post rurality' but throughout all of the discussion below the core question of the distinctiveness of the populations in question from those resident in neighboring large towns and cities should always be under consideration, even if remaining largely in the background.

What Are Rural Populations?

Curtailing discussion by noting that there are consider able conceptual questions and academic doubts over the very idea of rural populations today would, of course, be to overlook the ubiquity and frequency with which these populations are acknowledged in every country in the world, both officially by formal institutions and usually more implicitly by people in their everyday lives. How, then, are such populations recognized and defined?

A good place to start is with statistical breakdowns of the populations of different countries across the world. First, we can do this by looking at countries grouped into either continents or their location in more or less de veloped regions. This is presented in **Table 1**, where we see immediately that around the millennium the world changed from having most of its population rural to having most people resident in urban areas. However, there are still major differences in the relative distri bution of the population between (and within) the less or more developed regions. In the former, and especially in Asia and Africa, most people reside in rural areas. In particular, for Africa the population growth rate for the early twenty first century has been estimated at over 1% per year, leading to an absolute projected increase in rural populations from 0.5 to 0.7 billion between 2000 and 2030. In contrast, Asia's rural population is expected to remain at around 2.3 billion. In the more developed regions, although the legacy of urban focused develop ment is very clear, even by 2030 nearly a fifth of its population will still be rural. Nonetheless, the rural population of Europe is expected to decline from around 193 million in 2000 to just 131 million in 2030. Overall, though, all of these fairly crude figures and trends point

Table 1 Rural populations, % total by world area, 1975 2030

Rural population (%)	*1975*	*2000*	*2015*	*2030*
World	62	53	46	39
More developed regions	30	25	21	18
Less developed regions	73	60	51	43
Africa	75	63	55	61
Asia	75	63	54	46
Europe	33	27	24	20
Latin America and Caribbean	39	25	20	15
Northern America	26	23	19	13
Oceania	28	26	24	25

Source: Based on the data from United Nations (2004).

Table 2 Rural populations, % total by selection of countries, 2003

Western Africa	58
Benin	55
Burkina Faso	82
Cape Verde	44
Côte d'Ivoire	55
Gambia	74
South-eastern Asia	58
Cambodia	81
Indonesia	54
Lao People's Democratic Republic	79
Malaysia	37
Viet Nam	74
Western Europe	19
Austria	34
Belgium	3
France	24
Germany	12
Netherlands	34
South America	19
Argentina	10
Bolivia	37
Colombia	24
Guyana	62
Uruguay	7

Source: Based on the data from United Nations (2004).

to the sustained significance of any immediate idea of rural populations for the twenty first century.

Second, from this same statistical perspective, we can also consider the rural populations of individual coun tries. A small selection of these as estimated for 2003 is given in **Table 2**. This reveals clearly how even con tiguous countries can vary considerably in their pro portionate size of rural populations and cautions against overuse of the kind of regional generalizations given in **Table 1**. An important caveat to note here is that the definition of 'rural' varies considerably between countries (see **Table 5**), an inconsistency which clearly influences the impressions given by both **Tables 1** and **2**.

Moving away from official statistical definitions of rural populations, it is important to acknowledge how these populations are both imagined and talked about by the general public within everyday life. Again, this is likely to vary considerably geographically, reflecting both the 'reality' of who lives in rural areas and spatial vari ations in our cultural imaginations. We will see some thing of these latter cultural constructs in the sections of this article that follow a brief discussion of some of the key trends that shape rural populations across the globe.

Key Trends Shaping Rural Populations

When considering the populations captured by tabula tions such as those given above, a crucial issue to ac knowledge – also suggested by the variation in definition of rural that underpins them – is that these rural popu lations should never be seen as somehow existing in isolation from their urban counterparts. Crucially, we can only understand the characteristics of rural populations, especially today, by recognizing their connections with the urban and beyond. This is clear when we note a few of the widely researched trends that have shaped and continue to mold rural populations.

Out-Migration

Out migration of people from rural areas toward the cities is the main worldwide motor of urban growth both today and in the past. While most clearly represented in key historical episodes, such as the 1750–1850 period of the Industrial Revolution in Britain or in China since the mid 1980s, rural to urban migration is ubiquitous (if highly uneven), spatially and historically. It is under pinned most strongly by diverse economic factors – better wages, availability and diversity of job opportun ities, training and educational opportunities, availability and affordability of housing or food – but should also be seen as expressing cultural factors, from the attractions of the 'bright lights' of city life to the ability to live in a manner that would not be tolerated or feasible in the rural environment.

The consequences of rural to urban migration for rural places, and hence for rural populations, are nu merous. Most attention has probably been paid to the negatives. For example, out migration in developed countries is often given the evocative label 'drift from the land', emphasizing, on the one hand, the decline of agriculture which this migration typically expresses and, on the other hand, the more general rural decline this engenders, as expressed in **Figure 1**. In contrast, a de veloping world perspective is more likely to stress the positive consequences for rural populations. Out migra tion is often associated with moving for better paid work, either to the country's cities or abroad and rural popu lations benefit as this labor migration does not usually

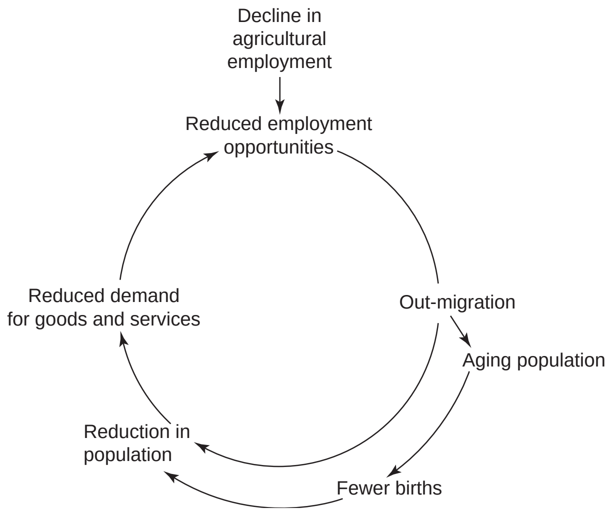

Figure 1 The cycle of rural decline. Source: Simplified from Gilg, A. (1983). Population and employment. In Pacione, M. (ed.) *Progress in Rural Geography*, p 95. Beckenham: Croom Helm.

signify a clean break for the migrant from family that remain behind. Instead, migrants send money back as remittances and may eventually return to settle in the rural location, again bringing with them various resources.

Counterurbanization

In the United States in the mid 1970s researchers began to notice that a major demographic shift affecting both rural and urban populations was gathering pace. In short, they noted a 'population turnaround', with a previously dominant trend of rural to urban migration being re placed by a dominant urban to rural migration trend. The latter, soon termed counter( )urbanization or counterurbanization, suggested a revival – some even claimed a renaissance – for rural populations. Indeed, counterurbanization was soon found across much of the developed world, its form expressed generally a nega tive linear correlation between population growth and settlement size.

As already suggested, counterurbanization was largely migration driven, with people 'voting with their feet' as they sought, in particular, better employment conditions and/or a perceived better quality of life. However, as with migration generally, these flows of people were selective, biased in favor of the categories listed in **Table 3**, thereby suggesting in turn a selective rural population renais sance. Moreover, we should also note the often con siderable geographical (even within one country) and historical variations displayed by counterurbanization, suggesting that its impact on any specific rural population always needs careful individual consideration.

Table 3 Some key overlapping counterurbaniser groups

- Higher social classes, including higher levels of education
- High levels of self-employment, at least after migration
- Middle aged and retired adults
- Nonreturn migrants (see below) but often from nearby locations
- Owners of houses rather than renters
- Households of two or more people, not single people

Other Migration Trends

At least two other forms of migration also have a po tentially important role in shaping rural populations. First, there is the circulation of people through rural areas, again often related to work. Second, as has already been noted, some out migrants return to rural areas later in life as what are called return migrants. Both of these flows can significantly shape any specific rural population through things such as the input of resources and novel cultural ideas that the migrants can bring with them.

Births and Deaths

A contemporary developed world perspective on the demographic factors shaping rural populations tends to focus on in and out migration and the selective character of those involved. However, we should not lose sight of the other two elements of the demogra phic equation: births and deaths. Thus, even in these countries there are rural populations exhibiting relatively high or low numbers of births or deaths. For example, many rural areas are very attractive to retired migrants, with the resulting demographic restructuring prompt ing lower birth rates and higher death rates than the rural norm.

In the less developed world, the role of births and deaths in molding rural populations is probably more significant. Specifically, there are considerable geo graphical variations in both rural birth rates – explained by interrelated factors such as culture, contraceptive use, life expectations – and rural death rates – where the spectre of HIV/Aids looms large. Some indications of these variations are shown in **Table 4**, where we see considerable diversity even for one country.

Studying Rural Populations

This section briefly notes how rural populations have been studied to date, drawing out four broad and com plementary families of work. From the well established legacy of anthropology, we also note research that draws more on sociological and geographical research conventions.

Table 4 Estimated rural and urban birth and death rates for the larger states of India, 2001

	Birth rate		*Death rate*	
	Rural	*Urban*	*Rural*	*Urban*
Andhra Pradesh	21.4	19.7	9.0	5.6
Assam	27.9	18.6	9.9	6.6
Bihar	32.3	23.4	8.5	6.3
Gujarat	26.7	21.5	8.9	5.6
Haryana	27.9	22.9	7.7	7.4
Karnataka	23.6	19.0	8.2	6.4
Kerala	17.4	16.7	6.8	6.2
Madhya Pradesh	32.9	23.1	10.8	7.2
Maharashtra	21.1	20.2	8.5	5.9
Orissa	24.0	19.7	10.8	6.8
Punjab	22.1	18.7	7.2	6.4
Rajasthan	32.4	24.8	8.3	6.2
Tamil Nadu	19.6	17.8	8.5	6.0
Uttar Pradesh	33.2	27.0	10.6	7.8
West Bengal	22.9	13.8	7.2	6.4

Source: Based on the data from http://www.indianngos.com/

Table 5 Criteria used by various countries to classify settlements as rural

Country	*Criteria*
Germany	Communes with fewer than 2000 inhabitants
Canada	Areas with fewer than 1000 inhabitants or a population density of less than 400 persons per square kilometer at the previous census
Peru	Areas that are neither populated centers with 100 dwellings or more grouped contiguously nor administrative centers of districts
Nigeria	Areas outside towns of 20 000 inhabitants or more whose occupations are mainly agricultural
Japan	Areas outside Densely Inhabited Districts (groups of contiguous basic unit blocks with a population of 5000 inhabitants or more and with a population density of at least 4000 inhabitants per square kilometer and/or with public, industrial, educational and recreational facilities)

Source: Based on the data from United Nations (2004).

The Anthropological Tradition

Traditional anthropology, as the study of humanity, has long concerned itself with default rural populations, using its predominant ethnographic approach to produce very detailed works based on long term and in depth immersion in 'other' cultures, both abroad and at home. For example, one cannot think of the history of the subject without acknowledging early twentieth century works such as Malinowski's studies of the Trobriand Islanders, Arensberg and Kimball's work on rural Ireland or Redfield's insights into the folk cultures of Mexico's Yucatan.

Within this anthropological tradition, we must note the decision by some anthropologists, notably in the UK and Ireland, to bring anthropology 'home' after World War II and seek to study the domestic 'other' – interestingly seen overwhelmingly as the rural populations. This body of work, mostly completed from the 1940s to the 1960s, became known as 'rural community studies'. It comprised highly empirical profiles of places such as Llanfihangel yng Ngwynfa and Glynceiriog in Wales, detailing their economic and social structures, religious, political, and cultural practices and, of course, kinship and family relations. Although, of great interest and value in their own right, these studies did not present a more general view of rural populations and, partly in consequence, researchers subsequently turned to more quantitative methods.

Statistical Work

With most countries in the world now producing some kind of fairly regular census of their populations as we have seen, it is unsurprising that social scientists have used these as a key source to profile rural populations, usually drawing comparisons with their urban counterparts. Data from a vast array of other detailed surveys (see below) can be used for similar purposes.

Statistics can provide a sense of rural populations in two ways. First, they give indirect appreciation through the ways they have been involved in defining rural. For example, the statistics involved in constructing the 'official' definitions of rural given in **Table 5** speak of the 'ideal' imagined character of the rural population: small numbers and low density, a high percentage working in agriculture, low 'political' importance, and a lack of services. Of course, we must take direct measures of the rural populations (e.g., **Table 4**) to be more certain of their actual as opposed to idealised characteristics. Hence, once we have agreed on our rural/urban spatial breakdown – by no means an easy task – statistics can be used to profile employment, housing conditions, family size and structure, etc.

Survey Work

A wide variety of types of social survey, from pre coded questionnaires through to more open ended interview schedules, have long been used to profile rural populations and continue to provide great insight. With the data collected, sometimes direct comparison is then made with urban populations but when neither the resources available nor the immediate interests of the researchers can stretch to such a task, comparisons that are subsequently made tend to be more implicit.

In the wake of community studies and the realization of rural change through counterurbanization, many social scientists sought to investigate the contemporary character of the rural populations of the more developed countries, subsequently enabling us to be able to move beyond stereotypes, as discussed in the next section. Pioneering was British sociologist Ray Pahl's PhD work in the late 1960s, which identified eight 'social groups in the countryside' based on a Weberian class analysis:

1. large property owners;
2. salaried immigrants with some capital;
3. 'spiralists';
4. those with limited income and little capital;
5. the retired;
6. council house tenants;
7. tied cottages and other tenants; and
8. local tradespeople and owners of small businesses.

An equally wide variety of survey work has also taken place in less developed countries, often stemming from work on health issues, much of it again helping to undermine often deeply embedded stereotypes of their rural populations (see below).

The Ethnographic Tradition Revisited

The last couple of decades have seen something of a return to more ethnographic work on rural populations. Here, the focus is on the excellent indicative example of US sociologist Michael Mayerfield Bell's 1994 study of 'Childerley', a pseudonym for a village in southern England that he lived in for 6 months in the early 1990s. Bell's particular interests were 'morality' and 'nature' but his study also very much expressed people's experiences of rural change, not least counterurbanization. Bell demonstrated that although most residents were white, politically Conservative and Protestant, a wide range of other people were also present. He saw differences largely in terms of class, which affected the cultural 'styles' that villagers adopted. While poorer villagers displayed a 'back door style' – informal, communal, locally embedded – wealthier residents displayed a 'front door style' – formal, individualistic, geographically dispersed. These styles were reflected throughout residents' lives, even influencing which pub was visited and subsequent behavior in it. The operation of power was also revealed through villagers seeking to deny the reality of class by an appeal to a unifying discourse of the 'country person', with the landscape embodying 'nature' and 'real' communities.

Beyond Stereotypes of Rural Populations

It was noted earlier the importance of acknowledging what ordinary people in their everyday lives understand by 'rural populations'. However, as with representations generally, and representations of groups of people in particular, academics have shown popular understandings to be partial, selective, and biased. Such stereotypes tend to fall into three families: negative, positive, and over simplistic.

Negative: 'Idiocy of Rural Life'

Ever since rural–urban divides could be readily identified, each side of this divide has tended to portray the other through some form of demeaning stereotype. In modern urban dominated societies, these have been especially strong culturally in denigrating the rural population. Reflecting Marx's and Engels's infamous description of the supposed 'idiocy of rural life', rural populations have been represented as backward, stupid, and reactionary. In Britain, the farming community has been particularly vilified by certain sections of popular culture, illustrated by the posting on the Chavworld internet forum reproduced verbatim in **Table 6**. Equally well established negative rural stereotypes in the USA, such as that of the people of the Appalachian region, have been shown to have knock on effects on the lives of these people, such as limiting their educational attainment. Similarly, in present day Hungary, the mass media in particular has promoted a negative representation of rural which is very unhelpful in promoting much needed rural development.

The less developed world has also manifested negative stereotyping of its rural populations, this time both by its own urban populations and also as part of a more, typically racist, developed world gaze. The latter is commonplace once again in the popular media where, for example, news stories about Africa position the rural population within a rhetoric of disaster that typically embraces a mix of drought, famine, warfare, over population, and epidemic disease. The notorious representation of Africa as the 'Dark Continent' looms large here.

Positive: 'Rural Idyll'

Not all stereotypes of rural populations are negative. Indeed, the noted trend of counterurbanization would fit poorly with the migrants involved having negative

Table 6 Post from weeble ii, 07/06/2005

farmer stereotype
wear old raggy clothes
has a flat hat
has straw in mouth chewing
goes 'ooohh arrr'
has a red tractor and a comb bine 'arrrvistar
has a wife who makes him a ploughmans salad with chunky sliced bread
sends sons to cut the logs
smokes
goes 'i've been farmin' for fortyyy year my son arr'

Source: Based on the data from http://www.chavworld.co.uk/forum/about1462.html.

Table 7 Logistic regression analysis of rural housing preferences in the Netherlands, by characteristics and valuation of the countryside

Rural characteristic	*B*	*Odds ratio*
Beautiful	0.372	1.450
Vivid	1.048***	2.853
Quiet	0.077	0.926
Safe	0.083	0.921
Clean	0.876**	2.402
Traditional	0.019	1.020
Attractive buildings	0.854***	2.348
Individualistic	0.178	1.195
Sociable	0.268	1.307
Well accessible	0.397	0.672
Positive valuation	1.289***	3.630
Constant	0.989	0.372
N	525	
2 log likelihood	321.872	
χ^2	75.194***	
Nagelkerke ρ^2	0.251	

*p < 0.10; **p < 0.05; ***p < 0.01.
Source: Table 4 in van Dam, F., Heins, S. and Elbersen, B. (2002). Lay discourses of the rural and stated and revealed preferences for rural living. Some evidence of the existence of a rural idyll in the Netherlands. *Journal of Rural Studies* 18, 461 476.

stereotypes of rural residents. Consequently, as shown by numerous studies, strongly positive representations of rural living are a key component in explaining the emergence and growth of counterurbanization. For ex ample, **Table 7** presents results from a Dutch study that uses quantitative methods to link a preference for rural housing with various positive images of the countryside. Moreover, as the table suggests, while these represen tations initially tend to emphasize physical qualities of the environment – space, beautiful scenery, quietness – they also have a strong social element. Many of the key themes in the latter – a slower pace of life, a greater sense of community, a more friendly and welcoming environ ment, less criminality – clearly construct a positive image of the rural population. Overall, although it does still contain within it some of the negative stereotyping noted above, the celebrated 'rural idyll' portrays a very positive picture of rural people, albeit one that will not appeal to everyone!

Oversimplified: Mr Average

While negative and positive stereotypes of rural popu lations clearly oversimplify, this can also take place without such explicit value judgment. Oversimplification can stem from lazy academic analysis or, more charitably, from an overemphasis on the most commonplace. How ever it arises, it is perhaps of heightened concern during these times of especially rapid rural restructuring when so much of what is taking place can all too easily be lost in the mix.

The classic statement on academic oversimplification of the rural population has still to be Chris Philo's 1992 journal article, that moved from a review of Colin Ward's account of children's lives in the countryside to consider rural studies more generally, and not least the subject's perspective on rural populations. Overall, whilst ac knowledging much fine work, and influenced by post modernism, Philo lamented a predominantly narrow emphasis on what he labeled 'Mr Average'; males who were middle class, employed, white, heterosexual, Eng lish, able bodied, and apolitical.

The legacy of Philo's intervention is well known. Al though there was noted work before his article appeared, it triggered a landslide of studies on the 'neglected rural geographies' he championed, now extending beyond the United Kingdom: for example, Philo's own work on the Shaker communities of North America or Barbara Pini's exploration of women's involvement in the Australian sugar industry. Indeed, work on 'rural others' has rapidly become a core element of contemporary rural studies. While some of this work has raised ethical concerns over issues such as the researcher's degree of commitment to and understanding of the people studied, here at least the stereotyping of rural populations has been challenged in a significant way.

Finally, there are clearly numerous negative con sequences that can stem from any stereotyping, some of which have already been suggested. A sadly topical one that has been associated with a Mr Average style sim plification concerns the perceptions and responses to rural homelessness in the developed world. Typically represented as an 'urban problem', homelessness is often as proportionately commonplace in rural areas as it is in the cities but has persistently failed to have been rec ognized as such, whether by the general public, policy makers or even academics. This stems from a number of factors, such as the relative absence of 'cardboard cities' in rural areas, but a crucial role is played by cultural factors. Specifically, where 'idyllic' cultural represen tations of rural populations predominate, they have little 'place' for homeless people; such people are not 'sup posed' to be found there. Moreover, when their presence can no longer be denied, this can be explained as a temporary anomaly, with the homeless also re repre sented and denigrated as out of place outsiders.

Conclusion: Recognizing Diversity

The jury may well still be out as to the ultimate social scientific salience of rural populations in an increasingly urbanized, globalized, and interconnected world, but it is clear that people living in rural areas across the globe remain both significant numerically and are of major interest to academics, students, policymakers, and the public at large. This article has sought to illustrate this point, arriving ultimately at the need to recognize the considerable variety that (still) exists within the rural population, anywhere in the world one chooses to look. Rural populations are not somehow isolated from their urban counterparts, and neither are they any less com plex or diverse.

See also: Census Geography; Migration; Rurality and Post-Rurality.

Further Reading

Bell, M. (1994). *Childerley*. London: University of Chicago Press.
Boyle, P., Halfacree, K. and Robinson, V. (1998). *Exploring Contemporary Migration*, ch. 6. Harlow: Addison, Wesley, Longman.
Cloke, P. (1997). Country backwater to virtual village? Rural studies and 'the cultural turn'. *Journal of Rural Studies* 13, 367 375.
Durand Drouhin, J. L., Szwengrub, L. M. and Mihailescu, I. (eds.) (1981). *Rural Community Studies in Europe*. Oxford: Pergamon.
Halfacree, K. (2003). Landscapes of rurality: Rural others/other rurals. In Roberston, I. & Richards, P. (eds.) *Studying Cultural Landscapes*, pp 141 169. London: Arnold.
Hoggart, K. (1990). Let's do away with rural. *Journal of Rural Studies* 6, 245 257.
Milbourne, P. and Cloke, P. (eds.) (2006). *International Perspectives on Rural Homelessness*. London: Routledge.
Philo, C. (1992). Neglected rural geographies: A review. *Journal of Rural Studies* 8, 193 207.
Woods, M. (2005). *Rural Geography,* chs. 1, 2, and 6. London: Sage.
Zlotnik, H. (2004). World urbanization: Trends and prospects. In Champion, T. & Hugo, G. (eds.) *New forms of Urbanization. Beyond the Urban Rural Dichotomy*, pp 43 64. Aldershot: Ashgate.

Relevant Websites

http://www.rural.gc.ca/discpaper e.phtml
Canadian Rural Partnership.
http://wordpress.com/tag/rural people/
Blogs about: Rural People.
http://www.ruralhome.org/pubs/hsganalysis/ts2000/index.htm
Building Rural Communities.
http://www.fao.org/sd/erp/ERPRuralYouth en.htm
Food and Agricultural Organization of the United Nations.

Rurality and Post-Rurality

K. H. Halfacree, Swansea University, Swansea, UK

Glossary

Actor-Network Theory (ANT) A sociological theory whose primary hub is the concept of the heterogeneous network. It is distinctive in that ANT's networks contain not merely people but other living things, inanimate objects, and organizations. All are collectively referred to as actors and are thus accorded a degree of agency.

Affordance Some quality of an object or environment that allows or facilitates a specific action.

Environmental Determinism The idea that the environment – usually the physical environment – determines behavior, culture, societal forms, and so on.

Nonrepresentational Theory A theory that argues that instead of studying and representing social relationships we should focus upon the everyday practices of human life, where 'doing' is typically much more than the act of 'contemplation'.

Political–Economy Perspective An approach that argues that the inextricably intertwined spheres of the political and the economic are the primary cause of everyday human experience.

Simulation An attempt to recreate as closely as possible a supposedly 'real' thing, condition, or process.

Introduction: The Problem with Rurality

To identify rurality is to identify the various things that make somewhere, someone, or something rural. This task is likely to prove much more challenging for social sci entists than for the general public who are mostly fairly, immediately able to speak of what they see to be these key rural characteristics. Thus, for example, **Figure 1** shows the answers that four cohorts of Swansea Uni versity students taking the Contemporary Rural Britain third year module gave to the question (asked in the first lecture), "Please list three words/terms/expressions that you associate with 'rural' (countryside)." For these stu dents, rurality is dominated by the practices and land scapes of farming, supported by a sense of quiet, remote environments amenable to leisure; to a lesser extent by a strong sense of community, rustic icons from mud to cider to tractors; and, for a few, by socially negative features such as inbreeding and cultural backwardness. Such a wide set of associations is not surprising in its British context, and has been reiterated frequently in a wide range of academic and journalistic studies.

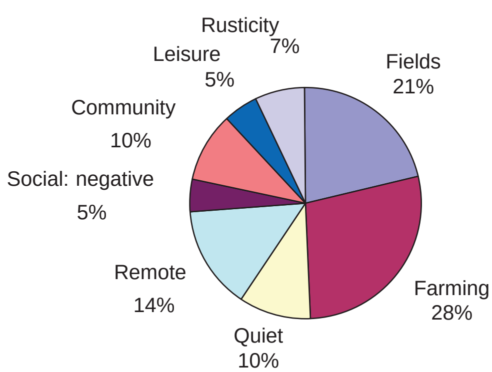

Figure 1 Characteristics of rurality, geography students 2003 2006.

For social scientists, however, just building up this popular picture of rurality is, of course, far from ad equate. Indeed, for some it is completely the wrong approach to take. Instead, the fundamental challenge is not merely to re present popular 'knowledge' but to find out whether or not one can identify any aspects of rurality that speak of more substantial and significant differences between the qualities of the rural and some other spatialized category or categories, the most obvi ous ones being the urban and the city. Reflecting this challenge, four perspectives on how one can position rurality today are now presented, instead of dwelling on its content, although the latter is unavoidably referred to at times.

Perspective 1: Rurality Effaced

At the very least, ever since Marx and Engels (in) famously depicted an 'idiocy of rural life' overturned through the backward countryside becoming subjected to the progressive rule of the towns, critical social science has been wary of accepting either the rural or rurality as valid social scientific concepts. This wariness, in par ticular, reflects the lasting influence of Marxian political–economy perspectives that stress the need to avoid 'chaotic conceptions' that lump together things that have little in common or, and most important here, divide that which should not be divided.

The general argument is that with all that was once seen as solid melting into air, the spatiality of con temporary capitalism has progressively done away with formerly significant geographical demarcations and bor ders, including that between rural and urban. The scales of operations within an increasingly speeded up – and, of course, globalized – capitalist society are constantly being rewritten, the net result being that the overall significance of any 'rural' scale has been incessantly and plurally undermined. From local, national, and global perspectives, therefore, an anachronistic rural has become 'a minor player in the line up of localities'. Sometimes it may reflect spatially significant difference but usually what takes place in rural areas is little different from what is taking place at regional, national, international, or even global scales.

This conceptual irrelevance today of erstwhile rural space was famously outlined by James Copp in a Presidential address to the US Rural Sociological Society in 1972 and has been reiterated periodically since. All of these declarations advocated doing away with 'rural' for theoretical progress. Indeed, continued belief in any town versus country divide may even be seen as ideological, both denying and confusing human understanding of the spatiality of contemporary capitalism. The only way to exculpate rurality would be if one could show 'significant' social processes in operation delineated at a local spatial scale, with their resulting spatial inscriptions enabling 'rural' to be distinguished from one or more 'nonrural' environments. In summary, one needs to pinpoint spaces distinguished enough by their own causal forces that they can be labeled 'rural'.

The battle over the theoretical salience of rurality can be illustrated with respect to agriculture. As already noted, agriculture and the landscapes of farming are usually seen as bedrock elements of rurality. However, building an understanding of rurality that foregrounds agriculture tends to single out this industry's supposed relative uniqueness when compared to other industries. The validity of this distinction is debatable. Indeed, it was in many ways the desire to challenge the 'exceptionalism' of understanding agriculture in a capitalist age that spurred on the political–economy approach in rural studies in the 1980s. In short, agriculture can be seen in many ways to resemble other industries in a capitalist economy, through the roles played by commodities, capitalist labor relations, class, profit, and so on, even if at the same time it does still retain many distinctive features, such as its continued heavy reliance on family labor and natural growth rhythms and seasons.

Of course, as the example of agriculture illustrates, the general argument for doing away with rurality as a key social scientific concept is not the same as saying that every place is becoming identical. Rural places and what goes on within them, so deeply inscribed in some cultures, although of course not in all, do still have their own characteristics and shadings. For example, **Table 1** shows the chapter titles from a recent textbook on the rural geography of the Global North. As the titles imply, with their reiteration of the words 'rural' and 'countryside', the author considers rural places to have some degree of distinctiveness but, via his topics, from the changing economy to poverty and social exclusion, it is

Table 1 Chapter titles in a recent rural geography textbook (selection)

• Globalization, modernity and the rural world	• Protecting the countryside
• Agricultural change	• Rural conflicts
• The changing rural economy	• Changing rural lifestyles
• Social and demographic change	• Living in the countryside: Housing, health and crime
• Changing communities: Restructuring rural services	• Growing up and growing old in the countryside
• Environmental change and rural areas	• Working in the countryside
• Rural policy and the responses to restructuring	• Hidden rural lifestyles: Poverty and social exclusion
• Rural development and regeneration	• Rurality, national identity and ethnicity
• Rural governance	• Alternative rural lifestyles
• Selling the countryside	

From: Woods, M. (2005). *Rural Geography.* London: Sage.

also very clear that rurality is often subordinate in any causal sense to more general structural features of twenty first century capitalism, as filtered through culture, politics, etc. Moreover, there is often as much or even more variation in these topics within countries or even within regions as there is between them as a group and the (equally diverse) urban world. In short, then, as a significant explanatory concept, rurality can be said to have been substantially effaced and one might therefore argue that scholarship should reflect this diminished status.

Another key issue that modulates the distinctiveness of the prospective rural locality is the scaling of the population geography of the area concerned. While acknowledging the dangers of environmental determinism, what are predominantly seen as rural areas in often (but not always) very large countries, such as the USA, Canada, Russia, or Australia, are likely to be more economically, politically, socially, and/or culturally distinctive than rural areas in the relatively small and densely populated countries of Western Europe, such as the Netherlands, Germany, or England (compare **Figures 2** and **3**). Again, of course, intranational variation must also be recognized; for example, compare the remote north of Sweden with the densely populated south.

Perspective 2: Rurality's Popular Resilience

In spite of the commonplace rejection of rurality as an important underpinning for academic research, it remains, of course, the framework for a vibrant area of scholarship, even for many of the aforementioned critics. This reflects in large part the last section's idiomatic elephant in the room, namely the strong cultural, popular presence of rurality within everyday life, which leads to its widespread understanding and usage. In short, and again it should be noted that in some cultures more than others, rurality comprises a powerful 'lay narrative' or

Figure 2 The Great Plains in South Dakota, USA: Extreme rurality? Source: Holly Barcus.

Figure 3 The heart of a rural idyll? Thorverton, Devon, England, 2007.

'conversational reality' regardless of the different spatialities that may be favored by twenty first century capitalism. Even if such narratives can be regarded as ideological, this does not negate their study.

Besides researching lay narratives of rurality in their own right and/or in order to critique them, a central reason for acknowledging the popular resilience of rurality is that it has very real material, geographical, and sociopolitical consequences. In short, people act on or through their understandings of rurality in their everyday lives and the rural world is partly produced thus. A good way to illustrate this is with reference to migration, although other practices such as tourism could also be used.

Net migration of people to more rural areas, or counterurbanization, is now recognized as a central, if geographically, historically, and socially uneven, demographic feature of most countries in the Global North, having been first identified as a mass phenomenon in the US by researchers in the mid 1970s. Much academic effort has subsequently been spent seeking the causes of this ‘demographic turnaround’ (from previous urbanization). On the one hand, as illustrated by political–economy inspired work, there are strong economic and class based explanations for counterurbanization, with the phenomenon ultimately rooted in the uneven spatial dynamics of the capitalist production process. This explanation has room for ‘rurality’ to feature only in a secondary capacity, with primary attention given to the capitalist restructuring that, as seen in the last section, has little respect for the rural/urban divide. On the other hand, more agency centered explanations of counterurbanization emphasize what the previous explanation downplays, namely how little of the precise location of the migrant’s new home is really explained by these economic factors. Especially given the prevalence of often quite long distance commuting, counterurbanizers – especially if middle class – usually have considerable leeway in deciding exactly where to live. Specifically, popular ideas or ‘representations’ of rurality typically feature strongly in guiding these moves. Such representations are even more significant for retired migrants and others unencumbered by job considerations.

From work on in migration to three northern Dutch rural areas in 1996, **Table 2** summarizes the responses given by at least 5% from either migrant category from a total of 374 survey respondents asked to give the two main reasons for choosing their new place of residence. All had settled in the research area within the 5 years prior to the survey. The importance of environmental perception is clear, with, more specifically, ‘beautiful/landscape/remote’ being positively associated with migrants for urban areas and ‘rural/countryside’ being positively associated with rural origin migrants at the 95% confidence level. Even though one should be cautious with both the exact figures and any implications of causality, **Table 2** suggests a clear role for rurality in human behavior – a representation that acts as a resource for human actions.

What are the representations of rurality that prove to be such significant resources in formulating behavior? There is not enough space to pursue this question in detail but of central significance are various forms of the ‘rural idyll’. These representations, which vary geographically, culturally, and socially, position the countryside as somewhere that is more relaxed and relaxing, scenic yet human scale, ‘organic’, and somehow ‘external’ to or otherwise distanced from (the negative features of) modern society. Some elements of a rural idyll have already been suggested in **Figure 1** but **Table 3** lists more, and **Figure 3** portrays a prospective modest, English, rural, idyllic village landscape. The English example is highlighted here in part due to the strength of this representation within dominant English culture(s) but, as suggested by **Table 2**, various idyllic representations of rurality are to be found across the Global North.

Table 2 The importance of the rural character of the destination to migrants to rural Netherlands

	Urban origin (n 158)	*Rural origin (n 216)*
Job related	16	22
Physical environment	44	43
Beautiful/landscape/ remote	(16)	(8)
Tranquility	(8)	(11)
Nature/wildlife	(11)	(14)
Rural/countryside	(1)	(6)
House price	9	5
House availability	6	2
Back to roots	9	13
Social relations (family/ friends)	3	7
Total	100	100

From: van Dam, F., Heins, S. and Elbersen, B. (2002). Lay discourses of the rural and stated and revealed preferences for rural living. Some evidence of the existence of a rural idyll in the Netherlands. *Journal of Rural Studies* 18, 461 476.

Table 3 Some ‘physical’ and ‘social’ elements of the English rural idyll

Physical elements	*Social elements*
Small villages	Timelessness
Small farms and fields	Stability and social harmony
Narrow lanes	Community and closely knit
Tranquility	Knowing one’s neighbors
Scenic beauty	Strong sense of tradition
Attractive vernacular housing	Relaxing environment
Village core (village green, pond, benches, etc.)	Little crime or trouble
Pub(s), shop(s), school	Mutual support

From: Halfacree, K. H. (author).

Perspective 3: Rurality beyond the Rural

The differing emphases of the last two sections indicate that social scientists have something of a dilemma with respect to their professional ‘attitude’ toward rurality. On the one hand, as reflecting a key spatiality within contemporary capitalist society, it seems somewhat minor and therefore merits little engagement. On the other hand, its fecundity within everyday life in cultures with a strong sense of the rural seems undiminished. This suggests that it merits much more attention from researchers, although one must nonetheless avoid reinforcing any rural/urban ideology. Indeed, this need for critical engagement can be taken further when one notes representations of rurality being deployed to shape existing rural spaces, or even to corral assemblages of rurality ‘beyond’ the rural. Such maneuvers increasingly reflect Jean Baudrillard's age of simulations, where the map no longer follows on from the territory, seeking to represent it, but instead 'precedes' and 'engenders' it – rurality, in other words, before rural space.

One can again look at the consequences of counter urbanization when identifying, first, how rurality is being deployed in shaping rural spaces. Excellent examples come from a research project that examined the re structuring of Aylesbury Vale district in Buckingham shire, England in 1989–1992. In particular, there was a private developer's construction of a new settlement of 800 houses on a largely self contained site just outside Aylesbury town. This new settlement, Watermead, al though to comprise substantial amounts of prestigious executive housing and to feature two lakes designed for water sports and an artificial ski slope, was also meant to represent a 'traditional village'. Thus, its publicity and marketing material spoke of accessing such things as 'the warmth and charm of a traditional Edwardian village' and the 'distinctive village square, with its pink and cream painted pub, restaurant and shopping mall, set around an attractive piazza'. Overall, a particular quality of life infused heavily with particular representations of rurality was constructed in this corner of the village in England.

At a second level, and more subtly, existing rurality can be 'reconstituted' to resemble more closely what rurality is 'supposed' to look like. This is very much driven by the changing and enhanced consumption de mands of rural areas, and is expressed in Aylesbury Vale by the 'rustification' of village centers to cater to in migrant expectations. The growing presence and de mands of rural tourism are also often implicated strongly in this rural reconstitution, complicating it still further. For example, the settlement of Cavendish on Prince Edward Island in Canada was the setting ('Avonlea') of L.M. Montgomery's popular 1908 novel *Anne of Green Gables*. Today, the Green Gables site is very much managed for heritage tourism, including being re constructed in the style of the imagined 1908 idyll, which includes unambiguously fictive elements becoming ma terial created so that the 'real' place resembles the lit erary place more closely.

At a third level, the Watermead example also points to how rurality is deployed to assemble many elements of the rural even beyond what most people would ac knowledge as rural space. Again, housing development is immediately instructive of rurality's apparent pseudo revanchist intent, with clearly urban developments often badged and marketed using associations typically seen as hallmarks of rurality. More generally, one can see asso ciational elements of the rural increasingly being brought into the city, noted examples being the introduction of 'nature' into otherwise sterile, super urban retail en vironments, such as Canada's celebrated West Edmonton Mall, and planners striving to make rural 'values' such as community, local place, and identity integral to urban (re)development. **Figure 4** shows such an example, from Pamplona in Spain. The idea behind this development of Chantrea Vive (Chantrea Lives) is to build a neighbor hood that is 'alive', so as to make the town more 'human'; for that, the developers include pseudo rural trees and little gardens.

The loosening and even severing of rurality from its material spatial referent has led to calls to acknowledge a 'post rural', whereby the rural should now no longer be seen as fixed in any way but as inherently open: to be thought of, molded and changed, experienced, and lived. From this understanding, instead of searching for any essence of post rurality, one should instead focus on the ways in which post rurality is constructed and deployed in practice (here one sees the influence of actor network theory). Moreover, such inherent mutability, mobility, and flexibility also help explain its present fecundity.

Perspective 4: Rurality beyond Representation

It is very clear that rurality speaks strongly and deeply to many people in many societies today and clearly has not been effaced culturally, even though its more abstract explanatory legitimacy is quite problematic. Indeed, as just shown, it has so much belligerent vitality in repre sentational terms that it is even deployed beyond the rural space to which it (arguably) normatively connects. However, examination of this vitality may be taken in a different direction to suggest that rurality is not just of importance at the level of representations and their material expression but also has more affective significance.

A core element within what has come to be known as nonrepresentational theory is that of affect, or the feel ings, emotions, and even actions brought about through engagement with the materiality of the world. Thinking affectively about rurality takes one away from a discursive idealism that concentrates on the rural's expression through representation to refocus on what it is like to be 'in' the rural, to take the rural's own forces seriously – to go from viewing the rural to being in the rural. This is an area that is yet to be developed significantly, in part reflecting the predominance of representation issues within rural studies' cultural turn to date, an anti materialist, pro social bias, and also the fear of espousing some form of environmental determinism. With regard to the latter, it is important not to see any one way re lation between the (rural) environment and action but to see the environment more as a set of affordances or opportunities.

Considering a small patch of Welsh rurality, for ex ample, such as that shown in **Figure 5**, what is its rural affect? Imagine what it would be like to be there (and

Figure 4 The development of Chantrea Vive, Pamplona, Spain. Courtesy of María Jesús Rivera Escribano.

Figure 5 Dinas Bran Castle, Llangollen, Wales, 2006.

compare it with the Great Plains in **Figure 2**). How would you 'feel' if you were walking in this environment? What is it that would 'strike' you most if you were among those trees? What would you get 'existentially' from such a place? And, since affect should not be seen as inherently subjective, what do such places 'mean' in their own terms? Answers to all of these questions will almost inevitably pull in the seemingly indelible association that rurality has with 'nature'. In other words, it is the 'nature of rurality' that holds the key to its affective power. While there is not enough space in this article to explore the many thorny issues around the definition of nature, perhaps the key aspect of 'nature' here is the sensual manifestation of the physical world, notably its sights, smells, and feelings, but also a sense of the mystical and unexplained. Thus are soil, rocks, water, animals, plants, insects, weather, temperature, and even supernatural forces brought (back) into human understanding of rurality; one goes back to the land. From this, the key to nature/rural affect is exposure to and reception of this earthy living sensuality, bodily and biologically in an immediate sense, as well as, of course, through cultural representational filters.

Pursuing the argument a little further, to explain the significance of this nature/rural affect refers ultimately to the importance of nature for human being and becoming. Again, examining what this entails cannot be developed here but a number of recent studies have begun to suggest the values of nature in terms of its linkages to things such as renewing the human ecological bond with the natural world and the land; accessing (an authentic (*sic.*) grounded sense of) dwelling and (ecological) community; its therapeutic potential for both physical and psychological ailments; its capacity for inspiration, long appreciated by artists; its ability to surprise and enchant; and even its expression of resistance, notably to dominant aspects of the social world. Nature, it seems, is still far from transcended and rurality is a key access point to it.

Conclusion: Integrating the Four Perspectives

This article has introduced four interpretations of rurality today. While some of these are largely complementary, others, notably the idea of rurality as effaced, are more difficult to reconcile with one another. Nonetheless, in this conclusion, such reconciliation is posited through the model shown in **Figure 6**. This model adopts and adapts Henri Lefebvre's well known threefold sense of spatiality to engage with debates about how rural space is defined. It argues that rural space is a hybrid of three intertwined aspects:

- rural localities – inscribed through distinctive spatial practices;
- representations of the rural – formal expressions; and
- everyday lives of the rural – what it is like be(com)ing in the rural.

Rurality, as that which makes somewhere, someone, or something rural, can be mapped onto this threefold model. First, the idea of rurality as effaced suggests the weakness and even failure of the rural locality at first sight, with most practices occurring in rural space not being especially distinct to that space. Therefore, rural space is now more defined in terms of strong representations – reflected in migration to desired 'rural' destinations or marketing places as 'rural' – and in terms of distinctively experienced everyday lives. Within the latter, an intermingling of engaging rurality through popular representation with rurality's affective impact creates a powerful sense of the rural as distinctive. This distinctiveness, tagged on the ground, for example, through the low density of people, in turn acts back on the spatial practices acting through rural space to create a degree of distinctiveness within rural localities after all. Moreover, less tied to predetermined space on the ground than is the 'traditional' spatiality of the rural locality, the representational rural is also able to express itself 'away' from accepted rural places as post rurality.

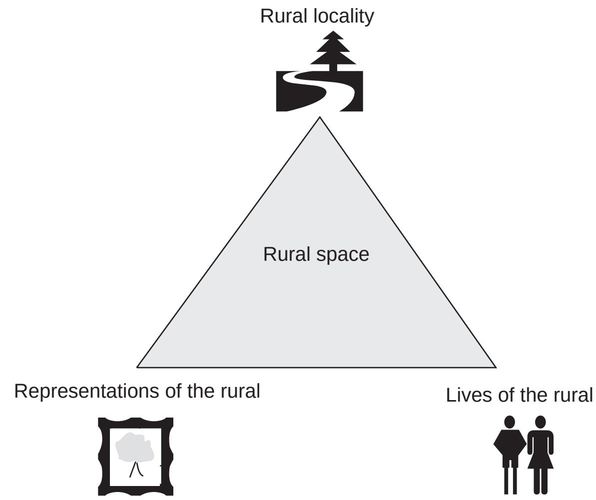

Figure 6 Modeling the hybrid totality of rural space. From Halfacree, K. (2006). Rural space: Constructing a three-fold architecture. In Cloke, P., Marsden, T. & Mooney, P. (eds.) *Handbook of Rural Studies,* pp 44 62. London: Sage.

However, there remains a tension here as the very rootedness of the affectual dimension of lived rurality still suggests that particular places – the countryside and remote areas sparsely populated by humans, generally – remain and retain the ontological basis of a rurality that is not so easily either done away with 'or' transformed into a simulation. Thus, in conclusion, rurality 'is' clearly strongly social, but it is also more than this.

See also: Counterurbanization; Nature; Rural Geography; Rural Populations; Urbanism.

Further Reading

Cloke, P. (1989). Rural geography and political economy. In Peet, R. & Thrift, N. (eds.) *New Models in Geography*, vol. 1, pp 164 197. London: Unwin Hyman.
Cloke, P. (ed.) (2003). *Country Visions*. Harlow: Pearson.
Cloke, P. (2006). Conceptualizing rurality. In Cloke, P., Marsden, T. & Mooney, P. (eds.) *Handbook of Rural Studies*, pp 18 28. London: Sage.
Copp, J. (1972). Rural sociology and rural development. *Rural Sociology* 37, 515 533.
Halfacree, K. (1993). Locality and social representation: Space, discourse and alternative definitions of the rural. *Journal of Rural Studies* 9, 23 37.
Halfacree, K. (2006). Rural space: Constructing a three fold architecture. In Cloke, P., Marsden, T. & Mooney, P. (eds.) *Handbook of Rural Studies*, pp 44 62. London: Sage.
Hoggart, K. (1990). Let's do away with rural. *Journal of Rural Studies* 6, 245 257.
Murdoch, J. and Marsden, T. (1994). *Reconstituting Rurality*. London: UCL Press.
Murdoch, J. and Pratt, A. (1993). Rural studies: Modernism, postmodernism and the 'post rural'. *Journal of Rural Studies* 9, 411 427.
Squire, S. (1992). Ways of seeing, ways of being: Literature, place, and tourism in L.M. Montgomery's Prince Edward Island. In Simpson Housley, P. & Norcliffe, G. (eds.) *A Few Acres of Snow. Literary and Artistic Images of Canada*, pp 137 147. Toronto: Dundurn Press.

van Dam, F., Heins, S. and Elbersen, B. (2002). Lay discourses of the rural and stated and revealed preferences for rural living. Some evidence of the existence of a rural idyll in the Netherlands. *Journal of Rural Studies* 18, 461 476.

Woods, M. (2005). *Rural Geography*. ch. 1. London: Sage.

Relevant Websites

http://rurality danh.blogspot.com/
Linking Global and Local in Rural Living.

http://www.nrdsbd.org/
Noakhali Rural Development Society.

http://www.rurality.org/
Rurality.

http://rurality.blogspot.com/
Rurality.

http://www.rural.co.uk
UK Villages Online.

Russian-Language Geography

V. Kolossov, Russian Academy of Sciences, Moscow, Russia
A. Treivish, Russian Academy of Sciences, Moscow, Russia

Introduction

Human geography of the world's largest state is very specific in style, progress, and troubles. Its major peculiarities related to the key features of Russia as a state or to the Russian scientific tradition can be reviewed as follows:

1. a relatively late but rapid surfacing in the eighteenth century and a catching up development;
2. an academic origin because of lacking medieval university and weak monastic science;
3. a tendency toward large scale and nationwide approaches and a certain 'statism';
4. a focus on economic, mostly resource based development;
5. a forced isolation from the world human geography during the Soviet period;
6. an elaboration of some original theories, such as territorial production complex (TPC) and energy production cycle, economic geographical position, basic framework (*opornyy karkas*), and territorial structure of settlement or economic systems, etc.;
7. human geographers were a minority in both Soviet and post Soviet geography traditionally dominated by specialists in physical geography; and
8. a falling authority (coupled with that of Russian language) in the post Soviet space.

These trends, both positive and negative, grant to Russian human geography its own face and can be better understood in the historical contexts of three major development stages: the pre Soviet, Soviet, and post Soviet.

Pre-Soviet Stage

The Kievan Russia's chronics mentioned peoples, towns, and waterways in Europe, North Africa, and Middle East – a typical horizon of the then Europeans. The Mongol yoke shook the country, turned it eastward to Asia, and, finally, granted to it the space of the horde proper. Between 1581 and 1647, Russian trapper explorers, though often illiterate, advanced from the Urals to the Pacific Ocean in search of fur.

There had been no professional science until 1724 when Peter I created the Academy in St. Petersburg. Then Ivan Kirilov, Vasily Tatishchev, Mikhail Lomonosov, Petr Pallas, and other universal scholars introduced both geography and political (by Tatishchev) or economic geography (the term by Lomonosov, the head of the department of geography in the Russian Academy since 1758). Their expeditions applied Western methods to the exploration of the Eurasian landmass which shaped Russia's geographical traditions, with its accent on economic zoning, colonization, and an analysis of industrial location. Kirilov and Tatishchev selected the points of localization for the Ural iron foundries in future cities of Orsk and Yekaterinburg.

In the nineteenth century, economic and population geography, along with statistics, became truly academic disciplines. In 1818, Constantine Arsenyev, the contemporary of Humboldt, Ritter, and Thünen, divided Russia into ten natural and economic, mostly agricultural, 'spaces'. After Arsenyev, many Russian geographers believed in the objective existence of economic regions to be revealed. Arsenyev criticized serfdom and was nearly blamed for the 1825 Decembrists' riot. Even so, he became an academician and a founder of Russian Geographic Society in 1845. Petr Köppen, the chairman of the society's statistical department, calculated population density by province and published the ethnographic map of European Russia in 1851.

The next phase is associated with Semenov Tyan Shansky. Petr Semenov went back from his travel to Central Asia in the late 1850s, participated in the abolition of serfdom, and became a brilliant social and physical geographer. He headed the Central Statistical Committee (1864–97) and Geographic Society (1873–1914), arranged many works, including the first all Russian census of population in 1897, and wrote or edited numerous books, such as *Geographical Statistical Vocabulary* and the popular *Russia: Full Geographical Description of Our Motherland* based on his network of 12 European economic regions. In 1880, he published a sample study of peasant strata in his native Murayevenskaya Volost (between Ryazan and Lipetsk).

The time created *de facto* human geographers, or 'anthropogeographs' after Ratzel, among them the climatologist, demogeographer, and environmentalist Alexander Voeykov; philologist, ethno , and historical geographer Vladimir Lamansky; and anthropologist Dmitry Anuchin, who founded the first cathedra of geography in Moscow University in 1888 (political emigrants Leon Mechnikov and Peter Kropotkin wrote mostly in French or in English).

Dmitry Mendeleev, the great chemist and advocate of industries, studied Russian coal and oil deposits. His book *On Cognition of Russia* placed the 1897 census data into his own 19 regions and divided households in three groups: "getters who take raw material from nature, manu facturers who process this material, and those who offer somewhat quite different from daily bread," or just pri mary, secondary, and tertiary sectors (introduced by A. Fisher and C. Clark near 1940). Mendeleev defined the then centroids of Russia's space and population and predicted the drift of the latter "towards the plentiful south and the land abundant east." Boris Veinberg cor rected his formulas in 1912 and calculated the long time series of both centroids in 1915. The recipe was adopted by Centrographic Laboratory (until the opponents of its leader Eugene Svyatlovsky ruined his center of spatial analysis).

Veniamin Semenov Tyan Shansky, the son of Petr, symbolized the continuity of Russian human geography. Born in 1870, he first studied geology in 1889–93 and died in 1942 siege of Leningrad. Veniamin joined his father's statistical projects in 1900 and became original scholar around 1910, with editing the *Trade and Industry of European Russia by Region* in 12 volumes and wrote *The Town and the Country in European Russia* – the pioneer book in urban geography. He identified 1237 'true' cities (by population and trade industrial turnover) instead of 761 official, of which 227 were not 'true', revealed the ring shaped structures around both Russian capitals and five macro clusters of urban life that survived through the twentieth century.

World War I stirred geopolitical ideas by H. Mack inder and R. Kjellén and others. The B. Semenov's 1915 article 'On powerful territorial possession as applied to Russia (essays on political geography)' added three forms of empire: 'circular' (Roman, etc.), 'patchy' overseas (British), and solid 'from sea to sea' (the USA and Russia). Semenov called for the growth of mid Russian Eurasia between the Volga and the Yenisey in order to move its cultural center nearer to the spatial one. The eastward shift was already Russian *idée fixe* and a part of 'Eur asianism', the next geopolitical concept after 'Panslavism' which, in turn, replaced that of Westernization in the mid 1800s. Eurasianism is normally regarded as a thought of the 1920s' emigrants (Nicholas Trubetskoy and Peter Savitsky, etc.). In fact, they removed abroad a domestic concept born together with the Trans Siberian and other big Russian railways.

After 1917, Veniamin Semenov Tyan Shansky fo cused on methodology of regional studies, mapping, and teaching. In his *Region and Country*, he suggested a system of geographical knowledge, of its general concepts (man, space, movement, etc.), and laws and predicted the Baikal Amur and the second Central Asian railways. Actually, the 'last universal geographer' generated Russian urban, theoretical, and political geography. Their close links with the landscape science (*landshaftovedenie* taken first from Germany and developed by Leo Berg and many Russian physical geographers) supported the unity of future Soviet geography.

Soviet Stage

Deep social transformations after October 1917 gave rise to sharp ideological discussions. During first 15 years since the arrival of Bolsheviks to power, human geog raphy in Russia/Soviet Union represented a changing mosaic of different schools. Some geographers still re mained faithful to Western approaches, such as German *Anthropogeographie*, stressing the impact of environment on human activities (A. Kruber) or French traditions (L. Sinitsky), others developed their original theories proposed before 1917, such as E. Svyatlovsky or V. Semenov Tian Shansky, still others believed that economic geography was a branch of economic sciences describing the territorial distribution of production with statistical methods (V. Den). Some radicals called upon an 'even' location of economy over the territory of the Soviet Union not depending on resources and distances. But the core of the future, original Soviet school of economic geography, which had a considerable influence on the geographical thinking in East Central Europe and in developing countries, gradually emerged from the fa mous State Plan of Electrification of Russia (GOELRO) adopted in 1920 and the works on further economic plans.

The founders of Soviet (Marxist) geography believed that the territorial division of labor cannot be indifferent; it is always constructed in the interests of a social class or group. Under capitalism it is never equal: some countries and regions are winners, other territories are losers. The territorial division of labor has a dual nature: on the one hand, it reflects social relations, but on the other hand, it depends on the level of technologies and development of productive forces. It explained a few similar territorial patterns in capitalist and socialist countries, as well as the existence of old patterns inherited by socialist patterns from the past. But the socialist territorial organization of production and of society as a whole is quite different, because it is being realized in the interests of all people.

For the so called regional school of new Soviet eco nomic geography, the central problem was economic re gionalization. This school was founded by Ivan Alexandrov and especially Nikolay Baransky (1881–1963) who wrote pioneering works and textbooks, initiated the creation of a number of geographical institutions (departments, journals, and publishers), and has for a long time been the leader of human geographers at Moscow State University. The main idea of the regional school was that in socialist conditions economic geography should play a constructivist role: it was necessary to overcome its traditionally descriptive nature. Economic geography should ground a new efficient territorial division of labor between different parts of the country. The task was to identify rational relations between the use of natural resources, material values, and skills of population accumulated in the past, and new technologies, and to get the best possible territorial combination of economic branches within an area (economic region) whose boundaries should be identified in course of a geographical study. At the same time, each economic region should represent a large, integral productive complex. But a relative self sufficiency of an economic region did not mean its isolation; on the contrary, the objective was to define an optimal specialization of each region at the scale of all national economy and to deepen the territorial division of labor between economic regions, which should allow us to rationalize the whole system of production and, in particular, to diminish transportation expenses. Thus, the economic integrity of a region means that its features are not a simple sum of characteristics of each branch of economy or settlements within its boundaries. Their combination offers a new quality. The system of economic regions should serve a basis for economic and territorial planning. This system should be hierarchically organized. At each territorial level, the boundaries of an economic region of the respective scale should match the boundaries of an administrative unit (with the exception of national areas). It should increase the efficiency of planning and facilitate the governance.

These ideas have been for the first time realized in the construction of the Dnepr hydroelectric power station in the Ukrainian city of Zaporozhye, using its energy and a common productive and social infrastructure. They were also applied to a number of large projects in Urals, Siberia, and other parts of the Soviet Union. In the 1920s, the central government began to integrate old and new provinces and recently created national autonomies within very large regions (*krai*), which looked like economic regions. But in the early 1930s, these regions were split again. The centralized governance through branch economic ministries (called first 'people's commissariats') permitted to strengthen political and economic control of the regions and to easily mobilize the means needed to realize big industrial constructions works, and particularly to maintain the military–industrial complex more or less competitive at the international scale. But in 1957, during the famous Nikita Khrushchev 'thaw', the Soviet leadership made the second and the last attempt to balance the territorial and the branch governance in creating the so called 'councils of national economy' (*sovnarkhozy*) – first within the limits of regrouped administrative regions, then within the regions themselves. This attempt was definitively abandoned in 1964 for the same reasons: the central government needed to concentrate financial and material means for funding the development of space and other mostly military technologies.

But at the level of separate, large projects, the ideas of the regional school and its followers have been certainly used in territorial and physical planning (called in Russia as 'regional planning' – *raionnaya planirovka*). Physical planning developed in 1930s and especially after World War II in a powerful interdisciplinary applied science, where geographers played an important role. Numerous institutes specialized in physical and urban planning were created in Moscow, Leningrad, and other cities. Physical planning was understood by E. Pertsik and other leading authors in this field as the elaboration of real projects of a combined localization of various economic objects (from large industrial plans to sanatoriums), cities, and other settlements in a concrete region (locality). The objectives also were the protection of environment, the rational use of natural resources, and a strict local functional zoning. But in reality, branch economic ministries were interested only in the increase of production, often did not follow the projects, and did not fulfill their obligations to protect the environment or to develop the social infrastructure needed for a normal life in a new city (department stores, kindergartens, services, etc.).

Since the early 1950s, they were developed and completed by the concept of the TPC. Usually, it was supposed that each large economic region was based on a single TPC, though some authors believed that it could have several complexes. TPC was understood as a territorial combination of industrial plants integrated by a certain leading (main) branch. This branch determined the role of a region in the national and/or international territorial division of labor. Other branches were divided into several types: those branches which supplied the whole complex with raw materials and energy; the branches processing by products of the leading branches and/or completing them in using the same raw materials; complementary branches of specialization; the branches satisfying internal needs of the region; and infrastructure. The accent was made, on the one hand, on the study of interactions and combinations of various elements of TPCs and the optimization of the use of labor and natural resources and the minimization of costs. On the other hand, economic geographers focused their studies on external relations of TPCs: the objective was to identify the branches of their natural (efficient) specialization. In 1954, the leading planner and professor of economic geography at Moscow State University Nikolay Kolosovsky introduced a related concept of the energy productive cycles (EPCs), which was widely used in the studies of TPCs. Kolosovsky's main idea was to create theoretical models of the most efficient *in situ* interaction between different branches of economy – from the extraction of raw materials up to the production of ready goods, including side and complementary industries, services, etc. The best example of such chain (cycle, in terms of Kolosovsky) was the combination of metallurgy with iron ore, coal, and chemical industry, electric power station, the production of construction materials, etc. Each TPC included one or several EPCs. They proposed economic methods of calculations of TPCs' efficiency. Graphic and mathematical models of TPCs served as a basis for physical planning.

The ideas of the regional school and the theories of TPCs and EPCs were developed by several generations of geographers from all parts of the Soviet Union. This is not by chance that the discipline was called economic, and not human geography. The primary purpose was to develop production, while social needs of a territory remained relatively secondary. Few years after the end of World War II, Nikolay Baransky launched his famous reproach to Soviet geographers: "They forgot the man!!!" But with time the situation changed: labor and, in particular, skilled labor was becoming a more and more valuable resource. It was not possible anymore to ignore environmental consequences of human activity. Geographers to a growing extent included in their studies and TPC models: education, research, housing, personal services, tourism, and other leisure activities. Since the 1960s the differentiation of old 'economic' geography was progressing very fast. Social, cultural, and even political geography (using mostly the data on foreign countries) became more popular among young scholars than traditional fields such as industrial or agricultural geography. For example, in the early 1970s, geography of personal services emerged as a separate and a quickly growing discipline, which included four main elements: (1) studies of conditions and factor of services' development, (2) the relationship between the real and the normative level of their consumption established by the state planning committee or other institutions, (3) the territorial organization, and (4) the behavior of consumers. Economic geography was even formally transformed into socioeconomic (or human) geography.

Settlement systems were considered as important elements of the territorial organization of society and spatial planning. Urban geography was a flourishing branch of economic geography, especially since the 1950s. Perhaps, the most original and practically oriented theory in this field was the concept of the supportive framework (literally –'skeleton') of the settlement system (*opornyi karkas rasselenia*) worked out by Georgy Lappo, Viktor Vladimirov, Pavel Polian, and other scholars. They considered the network of the largest towns and urban agglomerations together with the pattern of the main ways of communications integrating them in a single system. The 'hubs' or the foci of this network are the centers of regional and local settlement systems. These centers organize the surrounding territories and provide them with different services. It allows us to analyze the national network of towns and other settlements as a single system, to plan, and to coordinate its development. It was shown that due to the creation of linear structures of major cities along the main ways of communications, the average distance between them is significantly lower than that between large cities as a whole. It reduces transportation costs and partly compensates the low demographic and economic density in many regions of the country.

Urban studies stimulated the studies of the economic geographical location. This category was worked out by several generations of Soviet geographers: N. Baransky and I. Mayergoyz, V. Pokshishevsky and V. Varlamov in Moscow, A. Topchiev in Odessa (Ukraine), and many others. It was understood as a combination of spatial relations of countries, areas, regions, cities, and other settlements enterprises, etc., toward external objects affecting their functioning and economy. It was stressed that the economic geographical location was not predetermined by physical space and other realities forever, but was a historical category depending on changing technologies, the social organization of production, and other factors, modifying, on their turn, the territorial division of labor. The economic geographical location can be evaluated as favorable or not favorable. But a favorable location is only an opportunity; its realization depends on historical and socioeconomic circumstances. It was proposed, first, to analyze the economic geographical location at different levels – from the global to a local: the 'integral' is defined by a specific combination of its features at various scales. For instance, the shortcomings of the location at the national level can be to a certain extent compensated by its advantages at the regional and local levels. Second, they distinguished between different, particular kinds of location – political (toward political units, centers, processes, and phenomena), transportation (toward transportation axes and hubs, transit communications, and important ports), etc. Again, the evaluation of the integral location requires an analysis of the combination of particular locations. Since the early 1970s, a number of Soviet human geographers have worked on quantitative methods, helping to evaluate the economic geographical location.

Economic (later human) geographers have always been a minority in most Soviet geographical institutions usually headed by physical geographers (universities' departments, and central and regional academic institutes of geography). This situation had many problems but, at the same time, certainly had some positive sides: Soviet human geography paid much attention to the use of natural resources and the protection of environment; a number of ambitious projects (for instance, atlases) were realized in close cooperation between physical and human geographers. However, in the 1960s the recognition of the unity of geography as a science studying the interactions between different geographical envelopes of the Earth, including the social envelope, required a long and sometimes dramatic discussion, be cause some of its participants accused their opponents in ideological mistakes and the attacks against Marxism. Vsevolod Anuchin, an ardent champion of the unity of geography, played an outstanding role in this discussion.

Post-Soviet Stage

During the 1990s and early 2000s, both decay and re newal affected Russian science. The number of re searchers has halved to 0.4 million while that of university teachers has grown from 0.2 million up to 0.3 million. The shares of geographers (0.5%) and of human geographers (below 0.1%) did not change in the face of new challenges. The articles identified as human stably make some 30% in four geographical journals but shift from economic to social, cultural, and political themes (see **Table 1**). This is partly balanced by the fact that economy, demography, social, and political science are more 'regional' than before. In terms of space, human geography shrinks itself from USSR–CMEA–CIS to Russia proper and, inside it, from Asia to Europe (or, at least, is now better proportionate to their relative power).

The reforms and crisis of the 1990s attracted critic ism and alarm. Many authors, such as E. Animitsa, M. Bandman, S. Lavrov, and N. Petrov *et al.*, argued that the way to democracy and the market would shar pen regional problems. The term *regionalizatsiya* (by V. Kagansky and V. Leksin) reflected new regionalism often accompanied with the claims for a new spatial policy in order to soften growing disparities and political risks. The center–periphery gaps generated by previous 'big jumps' and 'spatial mobilization' were described and compared with the Western and Central European trends (O. Gritsay, A. Treivish, T. Nefedova, and N. Zubarevich). Regional identities and vernacular regions have been studied by M. Krylov and S. Pavlyuk.

Yu. Nikulnikov, N. Petrov, B. Rodoman, V. Kagansky, V. Lyubovny, and B. Lagutenko criticized the Soviet heritage, especially the concepts of TPC and economic region, in the very early 1990s. The last listed authors in 1994 called the TPC a schism and a source of personal well being for former generations of economic geog raphers and regional economists. Yet the pro TPC views were strong in Syktyvkar, Novosibirsk, and Irkutsk. These disputes did not sound as loud as, say, those be tween the old Soviet and the new liberal Russian economists. In the post crisis 2000s, they took the form of discussion about TPC versus 'cluster' model by M. Porter and M. Enright, analyzed in Russia by I. Pilipenko in 2005. Some other studies pretend to modernize the Christaller–Lösch heritage (V. Shuper and A. Vazhenin) or generated original ideas, such as the geo evolutional concept of transportation networks (by S. Tarkhov).

Human geography is now pluralistic, if not eclectic, both theoretically and empirically. Quite opposite views are often published under one cover. Never before have so many new topics been introduced simultaneously, many of them unusually 'negative', such as unemploy ment, industrial decline, poverty, ethnic conflicts, sep aratism, boundary, and other geopolitical problems (as Yu. Gladky, P. Baklanov, V. Baburin, and V. Shuvalov stated). Hence, a 'russianization' of those who were prepared for this by studying in foreign countries (S. Artobolevsky, V. Kolosov, V. Pulyarkin, G. Sdasyuk, L. Smirnyagin, R. Turovsky, L. Vardomsky, B. Zimin, etc.).

In the 2000s, economic recovery supports interest to outer world and the studies by A. Gorkin, A. Fetisov, N. Kaledin, Yu. Lipets, N. Mironenko, L. Sintserov, and other scholars whose main field remains distant or global. The contacts with the Western colleagues have increased due to joint projects and meetings, while the ties with the CIS human geographies have been maintained but weakening, partly due to their switching from Russian to

Table 1 Topical restructuring of human geography by publications in four journals[a]

Number of publications and their major subjects	*1980 82*	*1990 92*	*1995 97*	*2000 02*
Articles, annual average	139	129	124	117
Same as percent of their total number in journals	30.0	27.9	30.6	30.1
Human geography by topic, in %:	100	100	100	100
General: theory, history, and regional synthesis, etc.	22.0	25.9	25.2	24.9
Economic	25.7	20.0	17.2	16.6
Industrial and agricultural	20.8	15.6	11.1	10.4
Tertiary	4.8	4.4	6.1	6.2
Social, cultural, and political	24.0	25.9	29.0	30.2
Population settlement	16.0	10.9	9.5	10.5
Social (narrow sense), medical, and perceptual, etc.	6.9	7.8	10.8	10.7
Cultural, ethnic, and political	1.1	7.2	8.7	9.0
Resource environmental	28.3	28.2	28.6	28.3

[a] *Izvestiya Academii Nauk* (Geography series, Moscow); *Vestnik Moskovskogo Universiteta* (Geographical series); *Izvestiya Russkogo Geo graficheskogo Obshchestva* (St. Petersburg); and *Geografiya i Prirodnye Resursy* (Irkutsk).

their native or to European languages. During the late 1980s and the 1990s, dozens of geographers left Russia and did not necessarily join new expert communities. Some other human geographers (such as L. Smirnyagin, I. Shabdurasulov, and T. Regent) occupied relatively high official positions but have been replaced by professional politicians, as it has happened to many intellectuals.

Teaching is also progressing. Textbooks written by university teams or individually by A. Alekseev, V. Dro nov, E. Faibusovich, G. Lappo, V. Kholina, V. Kolosov, E. Kovalev, V. Maksakovsky, N. Mironenko, A. Naumov, I. Rodionova, A. Skopin, and other geographers form a new educational base. *Geografiya* schoolteachers' weekly (O. Korotova, S. Rogachev, and D. Zayats, *et al.*) is a helpful organ even for professional researchers. The monographs and readers by D. Zamyatin in cognitive geography (perception and images) also help education and popularization of human geography. Despite this, its class time has been reduced in secondary and in some higher schools.

So the discipline did survive, although it did not achieve much success. Its future remains unclear, espe cially in rapidly ageing academic institutions.

See also: Asia; Europe; Language.

Further Reading

Alayev, E. B. (1983). *Sotsial'no Ekonomicheskaya Geografiya: Ponyatiyno Terminologicheskiy Slovar (Social Economic Geography: Conceptual Terminologic Dictionary)*. Moscow: Mysl'.

Anuchin, V. A. (1972). *Teoreticheskiye Osnovy Geografii (Theoretical Foundations of Geography)*. Moscow: Mysl'.

Baranskiy, N. N. (1980). *Izbrannye Trudy. Stanovleniye sovetskoy ekonomicheskoy geografii. Nauchnye Printsypy Geografii* (Selected works, in 2 vols., *Formation of the Soviet Economic Geography. Scientific Principles of Geography*). Moscow: Mysl'.

Gladkiy, Yu. N. (2006). *Rossiya v Labirintah Geograficheskoy Sud'by (Russia in Labyrinths of the Geographical Fate)*. St. Petersburg: Izd. R. Arslanov Yuridicheskiy Tsentr Press.

Kolosovsky, N. N. (2006). *Izbrannye Trudy (Selected Works)*. Smolensk: Oykumena.

Lappo, G. M. (1978). *Razviiye Gorodskih Aglomeratsiy v SSSR (Development of Urban Agglomerations in the USSR)*. Moscow: Nauka.

Leyzerovich, E. E. (2004). *Ekonomicheskiye Mikrorayomy Rossii (Setka i Tipologiya) (Economic Mycro Regions of Russia (Network and Typology))*. Moscow: Trilobit, Rialteks.

Mayergoyz, I. M. (1986). *Territorial'naya Struktura Khozyaystva (Territorial Structure of Economy)*. Novosibirsk: Nauka.

Mendeleev, D. I. K. (1905 1907). *Poznaniyu Rossii (On Cognition of Russia)*, 6 edns. St. Petersburg.

Nefedova, T., Polyan, P. and Treivish, A. (eds.) (2001). *Gorod i Derevnya v Evropeyskoy Rossii: Sto Let Peremen (The Town and the Country in European Russia: One Hundred Years of Change)*. Moscow: OGI (United Humanitarian Publishers).

Saushkin, Yu. G. (2001). *Izbrannye Trudy (Selected works)*. Smolensk: Universum.

Semenov Tyan Shansky, P. P., Semenov Tyan Shansky, V. P., Lamanski, V. I. *et al*. (eds.) (1899 1914). *Rossiya: Polnoye Geograficheskoye Opisaniye Nashego Otechestva… (Russia: Full Geographical Description of our Motherland…)*, 19 vols. St. Petersburg: Izd. A.F. Devriena.

Semenov Tyan Shansky, V. P. (1910). *Gorod i Derevnya v Evropeyskoy Rossii: Ocherk po Ekonomicheskoy Geografii (The Town and the Country in European Russia: An Essay on Economic Geography)* (Zapiski Imperatorskogo Russkogo geograficheskogo obshchestva po otdeleniyu statistiki. Tom 10. Vypusk 2. P. 1 212). St. Petersburg.

Semenov Tyan Shansky, V. P. (1915). *O Mogushchestvennom Territorial'nom Vladenii Primenitel'no k Rossii (ocherki po Politicheskoy Geografii) (On Powerful Territorial Possession as Applied to Russia (Essays on Political Geography))*. Petrograd: Tipografiya V.V. Stasyulevicha.

Semenov Tyan Shansky, V. P. (1928). *Rayon i Strana: Posobiye Dlya Vysshey Shkoly (Region and Country: Teaching Aid for Universities)*. Moscow: Leningrad.

S

Sampling

L. Lo, York University, Toronto, ON, Canada

Glossary

Bias It refers to how far the average statistic lies from the parameter it is estimating. It is the error due to chance which arises when estimating a quantity. Errors from bias will not cancel out each other.

Precision It is a measure of how close an estimator is expected to be to the true value of a parameter. It is usually expressed in terms of imprecision and related to the standard error of the estimator. Less precision is reflected by a larger SE.

Random Number Table It is a table of numbers in which there is no discernable pattern of occurrences.

Sampling Fraction It is the ratio between sample size and population size. It represents the proportion of the population that is sampled.

Sampling Interval It is determined by dividing the population size by a desired sample size. It is the standard distance between sampling units selected from the sampling frame in a systematic sampling framework.

Unit of Analysis It is the major entity that is being analyzed in the study. It is the 'what' or 'whom' that is being studied. In human geography research, it can be individuals, groups, organizations, artifacts, or geographical units. This should not be confused with unit of observation, the unit on which data are collected.

Introduction

Sampling, normally associated with surveys, is a means of providing statistical data on a wide range of subjects for both research and administrative bases. Central governments make considerable use of surveys to inform them of conditions of their people in terms of employment and unemployment; income and expenditure; housing, education, nutrition and health conditions; migration, immigration, and newcomer integration. They also conduct surveys of organizations such as banks, retail outlets, farms, schools, and hospitals to check on financial and human health. Local and regional governments use travel and transportation surveys for local planning purposes. Market researchers carry out surveys to identify markets for products and services, to discover how they are used and perform, and to determine consumer reactions. Opinion polls keep track of the popularity of political leaders and their parties. Many university students and researchers in human geography as well as other disciplines use surveys (questionnaires and interviews to collect both quantitative and qualitative information) to address their research questions of diverse nature, context, and geography. As an example, a scan of two recent issues in *Environment and Planning A* (vol. 39, issues 3 and 8) finds (1) primary survey data were used to examine Chinese immigrant grocery shopping behavior in Toronto, Canada; (2) secondary survey data were used to examine the relationship between ethnic segregation and educational performance in Bradford and Leicester, England, the utilization of external business advice by small and medium size firms in four English towns, and space and gender relationship in post Soviet patterns of penalty; and (3) interview data were used to trace the development of the information technology industry in Suzhou, China, and the rise of boutiques in London's corporate finance industry.

While the subject of interest varies, there are common questions about the gathering of data. What information should we collect? From whom do we get the information? Where do we get the information? And how will the information be used? These are interrelated questions, answers to which determine the design of a survey. Before a survey can be conducted, the researcher has to decide on a method to select participants or respondents, and the size of the sample in addition to developing a list of possible respondents/participants for sampling. This article focuses on sample design as an integral part of the overall survey design. Two major principles underlie all sample design. The first is the desire to avoid as much bias as possible in the selection procedure; and the second is to achieve the maximum precision for a given outlay of resources.

The article begins with a brief outline of the history and logic of sampling. Then various sampling methods are explained, describing their strengths and weaknesses. Finally, the statistical aspects of sampling are explored.

History of Sampling

Although the idea of (random) sampling is an old one, mentioned a few times in the Bible, sample survey has a relatively short history, occurring mostly since the 1930s. First, it is the theory of small sample statistics (*t* distribution) developed by William Sealy Gossett, a chemist and statistician in England, in the early 1900s that put the subject on a more rigorous basis. Then in the US, an influential general interest magazine, the *Literary Digest*, published from 1890 to 1938 and then merged with *Time Magazine*, put it prominently on our radar map, when it successfully predicted the US presidential elections in 1920, 1924, 1928, 1932, but not in 1936. In 1936, George Gallup introduced quota sampling, and used this method to correctly predict the elections in 1936, 1944, but not in 1948. Considerable advances in survey methodology have since taken place.

It is safe to say that there have been two modern phases of development in sample survey. The first phase, started with the *Literary Digest*, dealt with the development of basic probability sampling theory, and the acceptance of probability sampling as the method of choice for obtaining representative samples. Personal interviews and household surveys were widely applied to a broad range of topic areas. The second phase took place especially vigorously in the late 1960s and early 1970s. This phase, tied in to the quantitative revolution in social sciences, witnessed an even broader utilization of survey sampling methodology and updated many techniques and procedures in line with broader societal and technological changes.

The Logic of Sampling

Sampling is the process by which a sample is chosen. A 'sample' is some part of a larger body specifically selected to represent the whole. Whereas information can theoretically be gathered about all members of a population, practicality (considerations of time, money and other resources, and member availability) determines that in the majority of cases, it is impossible to observe and sample all members. This leaves sampling a subset as the only alternative.

Even when enumerating the whole population is possible, sampling has its own advantages. It costs less, it can be processed more speedily, and it allows more timely reporting. Moreover, concentrating resources on a sample may lead to better data quality and hence more accurate results than a complete enumeration.

Population

In statistics, the term 'population' differs from our general understanding of the word which can mean all inhabitants of the earth, all individuals living in a country, or all residents of a city. In statistical terms, population refers to the collection of all 'elements' or 'units of analysis' under study; the element or unit of analysis can be an individual, a family, an ethnic group, a firm, a nongovernmental organization (NGO), a cultural organization, a school, a building, a street festival, or any event that is of interest to us.

It is important to define exactly what we mean by population since findings from a representative sample are often generalized to the population under study. The definition of population needs to be precisely and carefully specified according to the objectives of the survey. For example, if we are interested in New Yorkers' expenditure patterns, the population will include all households in New York because expenditure on housing, transportation, education, healthcare, food, clothing, and other consumer goods and services are often household based; households can be single- or multiple-person. On the other hand, if the objective is to discover if there is support for the introduction of a new transit system in Los Angeles, we will have to consider if the survey should be confined to persons living within the municipal boundaries of Los Angeles and if those persons should be above or below 15 years of age. Alternatively, if we are interested in the commuting patterns of Los Angeles' immigrant workers, we will have to decide whether we just focus on weekday work trips or include those in the weekends, and also whether it is just rush hours that matters. This says that in defining a population, it is important to consider both the spatial and temporal factors.

The ideal population that is required to meet survey objectives is known as the 'target population'. Sometimes, the target population has to be modified to exclude certain people for practical reasons. For instance, in national surveys, people living in remote communities are often excluded because they are too difficult or costly to enumerate. Exclusions can also include those living in prisons and mental health institutions as they are either inaccessible or mentally incapable of providing meaningful answers. Target population with explicit exclusions due to these practical constraints is known as the 'survey population'. Each member of a survey (or target) population constitutes an element or a unit of analysis.

In the remainder of this article, the terms 'population', 'target population', and 'survey population' may be used interchangeably.

Sampling Frame

A sampling frame is an actual list of elements or units corresponding to the target population. It can be a list of eligible voters, a list of bank locations, or a list of all households living in government subsidized housing units. It can also be a list of geographical areas such as neighborhoods, towns and cities, regions, or countries. A sampling frame provides the means of accessing the target population. A good sampling frame should be up to date, complete, accessible, and should contain sufficient information to identify and locate each unit.

Sampling frames are not always readily available. To develop one and ensure it is not biased, it is necessary to make sure that it comes from a variety of sources that represents the diversity in the population. For example, to construct a sampling frame of senior citizens living in one's community, the sources one can use to identify seniors include the phone book, the voters list, municipal departments that deal with seniors, membership lists of senior clubs and associations, senior housing developments, lists of seniors using home support services, and property assessment/tax rolls. One may include or exclude those who live in nursing homes, home for the aged, or extended care facilities, depending on one's research objective.

After constructing the overall list, one should check and eliminate duplicates, and then make sure that members in the list should be randomized so that there is no particular order in how they would appear.

All members in the frame are elements considered for selection in the sampling stage. Each is considered a 'sampling unit'. Information collected about these units provides the basis of analysis. If information is collected from each and every unit of the sampling frame, one has a 'census'. If collection of information is based on part of the population, one has a 'sample'. There is a set of rules and procedures that specifies how a sample is to be selected.

Sampling Methods

Sampling methods are about choosing participants for the survey. They can be probability or nonprobability based. The choice between the two and the choice within each group depend on the objectives and scope of the survey, including the budget of the survey, the availability of time, the method of data collection, the kind of respondents needed, reliability of the data, and the desire for precision. Each has its own advantages and disadvantages. Probability sampling aims for objectivity, conforms to statistical theories, and allows statistical inferences to be made to the population. It is however a costly approach. Nonprobability sampling generally lacks theoretical support. The sample may not be representative; hence population estimation is often not possible. Unlike data from probability sampling, the quality of the data from nonprobability sampling cannot be measured. Convenience and economy are its main attractions.

Probability Sampling

In probability sampling, each unit in the population has a known, nonzero chance of being included in the sample. Inferences can be made about the population which the sample represents. Methods range from a simple random selection to more complex selection strategies. **Table 1** outlines a general rule of thumb.

Simple random sampling

Simple random sampling is a scheme in which each unit has the same chance of being selected, and a combination of units in the frame has the same chance of making up the sample. Samples with this quality are often known as equal probability of selection method (EPSEM). A random sample can be selected by the lottery method (e.g., picking numbers out of a drum), by means of a table of random numbers available in many statistics and research design or method books, or using an appropriate function in any statistical software such as Microsoft Excel, SPSS, or SAS. **Table 2** shows part of a random number table.

Suppose we have a sampling frame of 200 members and we are to draw, for illustrative purpose, a random sample of five (which is too small by any standard) using a random number table, we can name the units 1 to 200. Since 200 is a three digit number, we use the first three digits of the numbers in the table. To start the selection process, we can randomly choose one number, say, by closing our eyes and pointing to a spot on the table. Suppose the number on row 1 and column 2 was selected. 187 then becomes the 'random start' or the first sampled element. The next number is 781, but since we do not have a member with that number, we work down the column until we find four more numbers with the first three digits lying between 001 and 200. The five numbers

Table 1 Choice of probability sampling methods

Population	*Probability sampling method*
One big group with relatively homogeneous units	Random or systematic sampling
Many groups; each internally homogeneous in the survey variables	Stratified sampling
Many groups; each internally heterogeneous in the survey variables	Cluster sampling

Table 2 An example to generate a random sample

	Col. 1	*Col. 2*	*Col. 3*	*Col. 4*	*Col. 5*
Row 1	3009	**187**9	**044**0	7916	6643
Row 2	9865	7813	7468	8493	3293
Row 3	1196	**125**1	2368	**126**2	5769
Row 4	1067	3716	8897	1970	8799
Row 5	5160	5563	6527	7861	3477
Row 6	4560	**009**4	8284	7604	1667
Row 7	7697	2151	4860	0739	4370

Adapted from Johnson, R. A. and Tsui, K. W. (1989). Statistical Reasoning Methods. Appendix B, Table 1, p. 553.

are highlighted in **Table 2**, and the corresponding numbers on the list/frame will become our sample.

Simple random sampling is the easiest method of sampling and it is the most commonly used. Advantages of this technique are that it has a sound mathematical and theoretical basis. On the other hand, this technique does not make use of auxiliary information present on the frame (such as number of employees in each business) that could make the design of the sample more efficient. While easy to apply to small populations, simple random sampling can be expensive and infeasible for large populations, especially if personal interviewers are required over geographically spread populations. It is also not recommended for heterogeneous population.

Systematic sampling

Systematic sampling is a modification of random sampling. In this method, every kth unit in the population is selected; k is the 'sampling interval' (N/n) best determined by dividing the population size (N) by a desired sample size (n). Using the example above where population is 200 and sample size is 5, the sampling interval $(k) = 40$. As with simple random sampling, a random start, for example, 7, is selected. The sampling interval (or standard distance between elements selected from the sampling frame) prescribes that every 40th member in the sampling frame will be selected. The sample then consists of 7, 47, 87, 127, and 167. This procedure produces an EPSEM sample. The method is simple to apply, and the sample spreads over the entire target population. This method is only problematic if characteristics of the prearranged list coincide with the sampling fraction/interval.

Stratified sampling

Simple random sampling and systematic sampling are good to use when the population is relatively homogeneous. However, in many sampling problems, the population can be composed of a set of groups with different characteristics. Stratified sampling is used when we require different estimates for different groups or when we are concerned with the precision of the overall population estimates. In this case, the population is divided into relatively homogeneous groups or strata based on income, age, gender, race/ethnicity, geographical regions, home ownership, possession of an automobile, and so on. As regards to number of strata, as many as practically possible should be used, and each stratum could be made as homogeneous as possible.

Once strata have been established, separate samples are selected from each stratum. Strata sample size can be proportionate or disproportionate to strata population size. Sampling units within each stratum can be selected by the simple random or systematic method. The example in **Figure 1** illustrates a proportional allocation scheme and a simple random selection method. The population of 117 items is divided into four strata (a, b, c, and d) with 21, 34, 35, and 27 elements, respectively. Suppose the budget for sampling is \$1000, and the cost of collecting information from one element is \$34. This allows a maximum of 29 ($=$\$1000/\$34) elements to be surveyed. This gives an overall 'sampling fraction' of 1/4 ($=29/117$). Proportional allocation yields five elements or observations in stratum a, eight in stratum b, nine in stratum c, and seven in stratum d. The observations in each stratum are randomly selected.

Stratified sampling increases sample efficiency or the precision of the overall population estimates. As well, estimates can be done at the stratum level. This method ensures sufficient sample with certain key groups, and allows the use of different sample designs for different portions of the population. It is also administratively convenient.

Cluster sampling and multistage sampling

Where a population consists of a set of groups and the groups are internally heterogeneous in the variables of interest, each group can be considered a cluster as opposed to a stratum consisting of relatively homogeneous elements. A random sample of the clusters can be selected to represent those not selected in the survey. If all units in the selected clusters are included in the sample, the method is known as cluster sampling or single stage sampling (see **Figure 2**). If a hierarchy of clusters exists, the process of selecting a final sample may involve two or more successive stages of cluster selection.

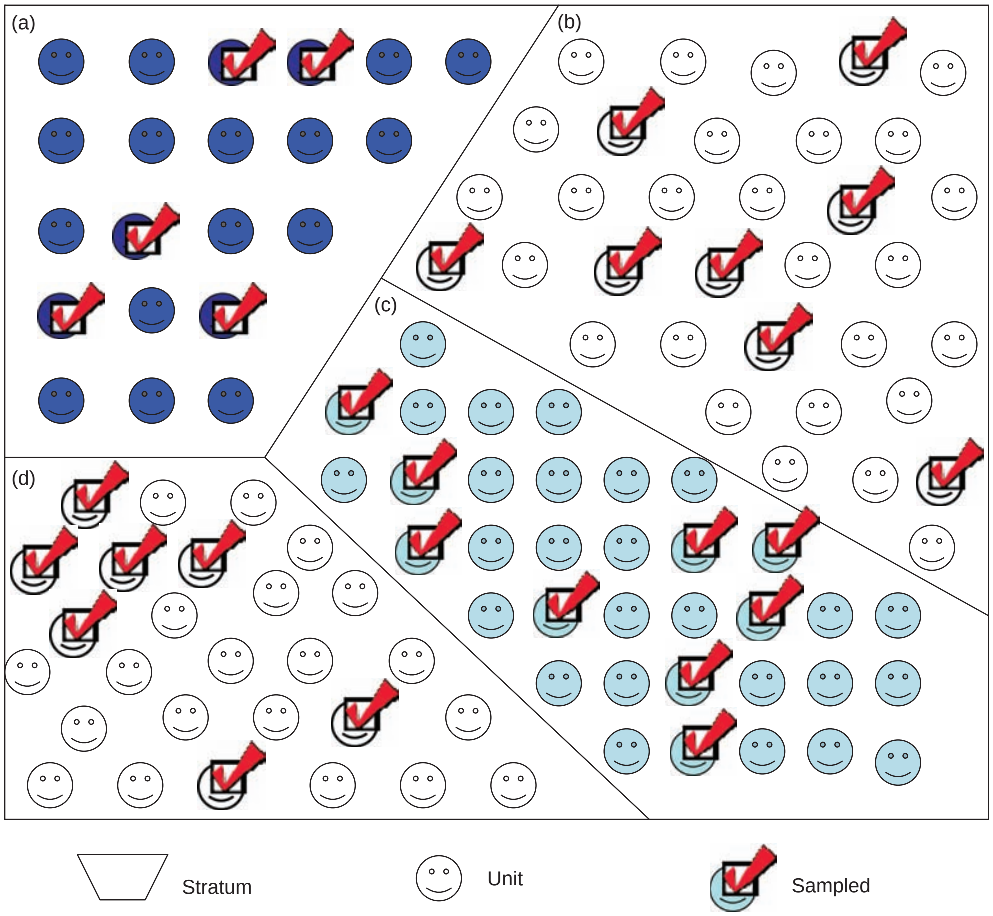

Figure 1 Stratified sampling. Adapted from Statistics Canada. (2001). Surveys: Start To Finish. Workshop presented by Statistical Consulting Group. Toronto. November 27 29. p 40.

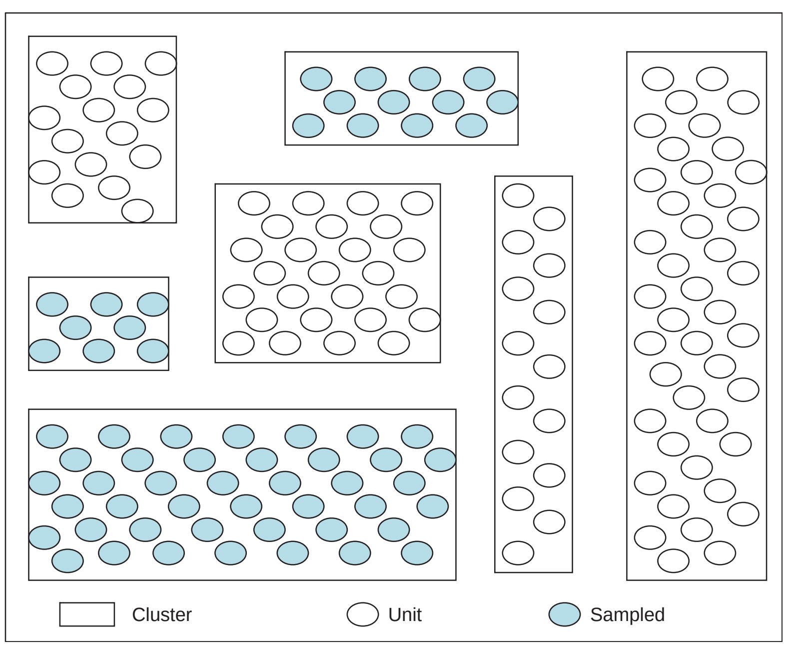

Figure 2 Cluster sampling. Adapted from Statistics Canada. (2001). Surveys: Start To Finish. Workshop presented by Statistical Consulting Group. Toronto. November 27 29. p. 42.

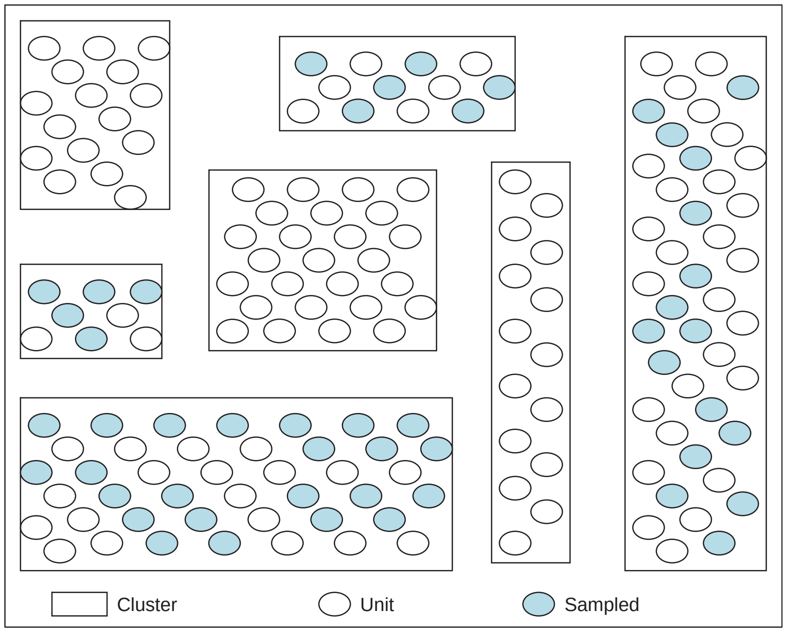

Figure 3 Two-stage sampling. Adapted from Statistics Canada. (2001). Surveys: Start To Finish. Workshop presented by Statistical Consulting Group. Toronto. November 27 29. p. 44.

This gives rise to multistage sampling. For example, in a national study of student performance in the public school system, the clusters selected in various stages can be cities, neighborhoods, schools, grade levels, and selected classes. **Figure 3** illustrates a two stage sampling.

Cluster sampling is used when a satisfactory sampling frame does not exist. It is less precise than simple random sampling because neighboring units tend to be similar. It is however justified by the economy it creates for sampling and data collection.

Area sampling

Area sampling is a common form of cluster sampling in which areas serve as the primary sampling units. In this approach, the population or total area under investigation is divided into mutually exclusive and exhaustive subareas using maps, and a random sample of subareas is selected. If all the units (e.g., households and buildings) in the selected areas are used in the study, it is one stage area sampling. If the subareas themselves are further divided and subsampled, the procedure is multistage area sampling.

Thus in area sampling, maps, rather than lists, serve as the sampling frame. This is the main method of sampling where population is geographically dispersed and where adequate population lists are rare. A case in point is developing countries. A grid, like the one in **Figure 4**, is drawn and superimposed on a map of the area of concern. Sampling points are selected on the basis of numbers drawn at random that equate to the numbered columns and rows of the grid. **Figure 4** shows a pattern of sampled areas that emerges from this process. If the areas are large, they can be further subdivided into subareas with another grid superimposed on them.

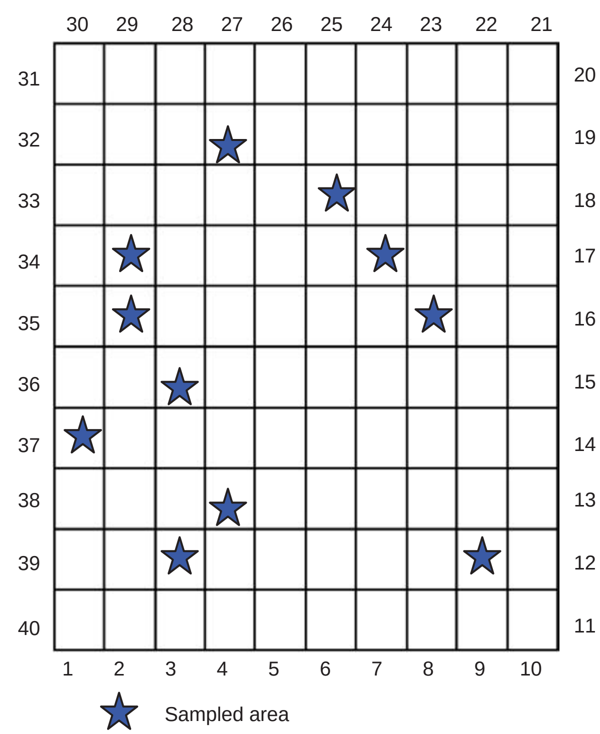

Figure 4 Area sampling. Adapted from Crawford, I. M. (1997). Marketing Research and Information Systems. Rome: Food and Agriculture Organization of the United Nations. Figure 7.3. http://www/fao.org/docrep/W3241E00.htm, accessed 7 May 2007.

Area sampling is more economical than simple random sampling. It incurs fewer travel expenses and administrative costs, and sampling frame is only needed for the chosen areas. On the negative side, sampling errors are larger and are difficult to measure.

Nonprobability Sampling

Unlike probability sampling when results can be gener alized to the whole population, nonprobability sampling is marked by subjectivity which precludes the develop ment of a theoretical framework and evaluation by assumption free statistical methods. In nonprobability sampling, the sample is not representative of the popu lation, so no population estimate is possible, and one cannot measure the quality of the data.

Nonetheless, nonprobability sampling is widely prac ticed, mainly for reasons of cost, timeliness, and con venience. It is useful when our objective is exploratory, when it is impossible to obtain a large enough sample, when qualitative information is required, and when stat istical reliability is unimportant. Nonprobability methods are common in pretest, pilot studies, and focus groups.

Convenience/haphazard/accidental sampling

As the name suggests, units in a convenience or acci dental sample are selected on the basis of convenience and expediency. Examples include conducting interviews on a street corner, surveying people who happen to be at a shopping mall or who frequent a seniors' drop in centre, using volunteers in medical/clinical/psycho logical experiments, and pooling responses from people answering a television request for opinions or a ques tionnaire on a website. Convenience sampling is useful if the population is homogenous. Otherwise, large biases are likely introduced as the volunteer population is often very different from the general population.

Purposeful/judgment sampling

A purposeful or judgment sample is chosen on the basis of knowledge of the research problem or expert judgment concerning the makeup of a population. Expertise allows the selection of typical units for inclusion in the sample. For instance, an educational researcher chooses a selec tion of schools in a city to give a cross section of school types, and a social planner recommends a group of social service providers for examining the utilization of settle ment service by immigrants. An advantage of this method is reduced cost and time. Yet sample selection can be severely biased.

Quota sampling

Quota sampling consists of two or more stages. The early stages are often carried out by probability methods, in the same way as for a probability sample. These stages create subgroups (similar to strata in stratified sampling) based on age, gender, employment status, and/or geographical area. On the basis of available data on the representativity of desired groups in the population, quotas are calculated and assigned to each subgroup. The quota controls can be independent (e.g., 100 men, 200 women, 150 people under 35 years of age, and 150 above 35 years of age), or connected (such as 300 men who are older than 65 and are still working, and 400 immigrants from China who moved to Canada within the last 5 years). The aim of assigning quotas is to avoid the selection biases that would occur in subsequent sampling.

For each quota group, it is unlikely that a list exists. So in the final stage, an existing list such as a phone book is often used to select possible respondents. The interviewer goes through the list and picks those who fit into a preestablished group by means of screening.

Quota sampling differs from stratified sampling at the final stage of selecting sampling units. This method is also known as modified probability sampling. It is useful for comparing subgroups, and is widely used in opinion and market research surveys. Quota sampling is less costly than random sampling, and can be administered easily. It is therefore good to obtain immediate public reaction to some event. Like other nonprobability methods, it faces the issue of precision.

Snowball sampling and respondent-driven sampling

In snowball sampling, the first few potential respondents are often initiated by word of mouth. They are contacted and asked whether they know of anybody with the characteristics that we are looking for in our research. This is based on the premise that they may well have knowledge of others with similar characteristics or in similar situations (e.g., drug users and cancer survivors). This method of using existing subjects to recruit future subjects will result in a sample that appears to grow like a rolling snowball.

Snowball sampling is a good method for populations that are hard to reach, and small relative to the general population. Examples of such not well delimited and often hidden populations include the homeless, sex workers, gangsters, and people with hidden disability. It may be extremely difficult or cost prohibitive to locate respondents with these characteristics. The snowball technique can dramatically lower search costs, Yet there is no exact idea if the sample represents a good cross section of the population. The argument is that people who have many friends are more likely to be recruited into the sample.

The bias due to snowballing can be tremendous. The recent advent of respondent driven sampling is a step up. Respondent driven sampling combines snowball sam pling with a mathematical model that weighs the sample to compensate for the fact that the sample was not collected in a random way. In essence, respondents recruit their peers, as in network based samples, and researchers keep track of who recruit whom and their numbers of social contacts. A mathematical model of the recruitment process then weighs the sample. This method allows theoretical development and enables researchers to provide both unbiased population esti mates and measures of the precision of those estimates.

Other Aspects of Sampling

Sample Size

Sample size, margin of error, and level of confidence

A question often asked in sample design is how many observations, subjects, or members of the population one needs to pick for the study. As mentioned earlier, ideally we sample the whole population. At the least, we need a sample that is representative of the target population.

In determining a sample size, it is important to con sider the survey objectives and possible implications of the results. If the objective is to generalize results for the target population, and use the results to make major decisions concerning the target population, one should strive for the highest level of accuracy. If the purpose is for general information about the target population, to identify issues facing them, or to explore their views, one needs a minimum level of accuracy and a smaller sample will do.

The sample size must be reasonable in terms of the budget available, timeliness of the data; and any operational constraints imposed by the availability (or ease of construction) of a sampling frame or list, the availability and ease of recruiting interviewing, data coding and editing staff, and processing facilities. Usually a positive correlation exists between resource availability and sample size. These constraints also determine the sampling method and data collection method.

As a rule of thumb, the larger the population, and the greater the variability in the population, the larger the sample needed when one is concerned with data reliability and expected accuracy of results. How well the sample represents the population is gauged by two statistics: margin of error and confidence interval.

'Margin of error' refers to the degree of fuzziness one can live with. A census (or surveying the whole population) gives the true response to a question. A sample survey provides an estimate of the true response. How precise is the estimate is expressed by the margin of error. Generally, the larger the margin, the less precise is the estimate. As sample size increases, margin of error decreases.

'Confidence interval' (α) or its counterpart, 'level of confidence' ($1-\alpha$), indicates the reliability of the findings. A 5% confidence interval or a 95% level of confidence means that if the survey were conducted 100 times with different samples, in 95 times, the results obtained would be similar to the results obtained with another sample. Alternatively, it also means that if one could have talked to the whole population, there is a 5% chance that the findings would vary from the survey results.

Putting them together, if a survey has a margin of error of plus or minus 3% at a 95% level of confidence, it means that if the survey were to be repeated, 19 times out of 20 (or 95 times out of 100), the results obtained would be within three percentage points above or below the results provided by the survey sample. A 95% level of confidence also means that 5% of the surveys will be off the wall with numbers that do not make much sense.

The relationship between sample size, margin of error, and confidence interval is unambiguous. **Table 3** shows that at any confidence interval, margin of error decreases as sample size increases. However, beyond a certain point, diminishing returns click in. For example, a sample of 1000 has a margin of error of 3% at 95% level of con fidence. By doubling the sample to 2000, the margin of error only decreases from plus or minus 3% to plus or minus 2%.

A 95% level of confidence is an industry standard although sometimes 99% or 90% is used. For the same sample size, a lower level of confidence yields a smaller margin of error. Similarly, a smaller sample can obtain the same margin of error at a lower level of confidence.

Strategies for determining sample size

For reasons of reliability, no samples should be smaller than 25 although this is common and acceptable in practice, particularly with qualitative studies. In statistics, 30 observations or less constitutes a small sample. The following are common approaches to determining sample size:

1. Using a census for small populations: although cost considerations make census impossible for large

Table 3 Sample size, margin of error, and level of confidence

Survey sample size	*Margin of error (%)*[a]	
	Level of confidence	
	95%	90%
50	±13.8	±11.6
100	±9.8	±8.2
500	±4.3	±3.6
1000	±3.0	±2.5
1500	±2.4	±2.0
2000	±2.1	±1.7

[a] Assume normal distribution.

populations, a census is attractive for small populations. It provides data on all the individuals in the population and eliminates sampling error.
2. Using a sample size of a similar study: a review of the literature can provide guidance about 'typical' sample sizes which are used. Care however should be exercised as to not repeat errors another study made in determining the sample size.
3. Using published tables: a third way is to rely on published tables such as one represented by **Table 4**. These tables provide the sample size for a given set of criteria regarding precision, confidence levels, and population variability. The sample sizes are the number of 'obtained' responses, and not necessarily the number of surveys mailed or interviews planned (as nonresponses are common).
4. Applying formulas: although tables can provide a useful guide for determining the sample size, one may need to calculate the necessary sample size for a different combination of levels of precision, confidence, and variability. One of several formulas available to determine sample size is

$$n = (Z\alpha/2\ \sigma/E)^2$$

where Z is the critical value associated with the confidence interval α, σ is the population standard deviation, and E is the desired margin of error. This formula applies if the variable of interest is a normally distributed continuous variable. It assumes that σ is known. Otherwise, a small scale preliminary sampling is necessary to obtain an estimate of σ to be used in the formula. Other formulas can be obtained easily from statistical textbooks.

Error and Bias in Sampling

Error and bias in sampling can be very costly. A classic example deals with the failure of *Literary Digest* (now part of *Time Magazine*) to predict a Republican win in the US presidential election of 1936. The Digest obtained a sample size of 1 million through magazine subscription lists and telephone directories, and did not realize that these lists were heavily biased towards Republicans. The resulting sample, though very large, was deeply flawed. The bias in it cost the Republican Party and its candidates a great deal in monetary, time, energy, and emotional terms. So it is pertinent to look at the various sources of error and bias and exercise tight control in both sample design and sample surveying stages.

Sampling error

A sample is expected to mirror the population from which it comes, however, there is no guarantee that any sample will be precisely representative of its population. Why does this happen? A common explanation is the presence of sampling error.

Sampling error is the difference between the results of a sample and a census, due to the particular units that happen to have been selected. There are two basic causes of sampling error. One is chance, which is measurable (margin of error), and can be compensated by a larger sample size. The other is sampling bias which is usually the result of a poor sampling plan. A poorly designed sampling plan, such as administering a survey during holiday seasons, can lead to a high level of nonresponse. The bias is due to the fact that nonrespondents usually have characteristics different from the respondents.

Nonsampling error

Like sampling error, nonsampling error may be an innocent byproduct of the sampling procedure (random error) or produced by participants in the survey (bias). Nonsampling error results primarily from the manner in which sample elements are selected and measured. The effects of random errors approximately cancel out if fairly large samples are used. Biases tend to create errors in the same direction and cumulate over the entire sample. The sources of nonsampling bias can be classified into coverage, processing, and nonresponse errors.

Coverage error occurs if the sampling frame does not adequately cover the target population. A less than adequate coverage can be the result of omission (as in unlisted phones), duplication (such as using both a voter list and a ratepayer list), or wrongful inclusion of units (such as an obsolescent employee list dated several years ago) in the sampling frame. It is also caused by errors in field procedures that result in missed households or persons.

Processing errors may occur when the data are requested, provided, received, recorded, edited, and imputed. No two interviewers are alike. Personal prejudices of either the designer of the study or the data collector may induce bias. The same person may provide different answers to different interviewers. Respondents may

Table 4 Sample size table

Population	*Sample size*
50	4
100	80
500	222
1000	286
5000	370
10 000	385
100 000	398
1 000 000	400
10 000 000	400

Source: Statistics Canada. (2001). Surveys: Start To Finish. Workshop presented by Statistical Consulting Group. Toronto. November 27 29. p. 56.

give incorrect answers to impress the interviewer, and knowing why a study is being conducted for who may create a different and possibly incorrect response, for example, to an income or even age question. Data can be wrongly coded or entered. Errors may not be identified at the edit phase, and compounded by poor imputation procedures.

The presence of nonrespondents and incomplete/imperfect information provided by the respondents produces nonresponse error. Nonrespondents are particularly problematic as they may possess characteristics different from those of the respondents. Nonresponse rate is negatively affected by a lack of publicity of the survey, a lack of enthusiasm from potential respondents, and a wrong timing when respondents are approached.

These potential biases are present in both census and sample surveys. They are the principal causes for concern about the quality of data. They are often difficult, if not impossible, to measure.

Control of errors

Sampling errors have to do with the sampling process. Apart from the need to develop an efficient sampling plan and make proper use of available information in developing the sample design, sampling errors can be effectively controlled by increasing the sample size since size of sampling error decreases with increase in sample size.

Nonsampling errors are more various. Errors due to coverage issues can be reduced by obtaining the most up to date and accurate sampling frame. To reduce response and processing errors, pretesting the questionnaire, including in the questionnaire specific instructions to respondents, providing good training program for survey personnel, instituting quality control programs (such as getting a different coder to recode the sample or verifying data entry), and deploying good edit rules and decision tables are effective controls.

To attain a good response rate, one starts with a good questionnaire design that ensures professionalism, elicits interests from potential respondents, and is manageable. One also needs to publicize the survey, and promote its usefulness with advance notices, reminder notices, and rewards if necessary. Apart from encouraging participation, one should avoid survey during holiday seasons or summers or repeat calling not at home potential respondents at different times of the day and different days of the week.

See also: Hypothesis Testing; Quantitative Methodologies; Questionnaire Survey; Selection Bias; Statistics, Inferential.

Further Reading

Fowler, F. J. (2001). *Survey Research Methods. Applied Social Research Methods Series 1*. Newbury Park, CA: Sage.
Frankel, M. R. and Frankel, L. R. (1987). Fifty years of survey sampling in the United States. *The Public Opinion Quarterly* 51, S127 S138.
Heckathorn, D. D. (1997). Respondent driven sampling: A new approach to the study of hidden populations. *Social Problem* 44, 174 199.
Henry, G. T. (1990). *Practical Sampling. Applied Social Research Methods Series 21*. Newbury Park, CA: Sage.
Johnson, R. A. and Tsui, K. W. (1989). *Statistical Reasoning Methods*. New York: John Wiley.
Kalton, G. (1983). *Introduction to Survey Sampling*. Newbury Park, CA: Sage.
Kish, L. (1995). *Survey Sampling*. New York: John Wiley.
Lohr, S. L. (1999). *Sampling: Design and Analysis*. Pacific Grove, CA: Duxbury.
Montello, D. R. and Sutton, P. (2006). *An Introduction to Scientific Research Methods in Geography*. Thousand Oaks: Sage.
Nesbary, D. (2000). *Survey Research and the World Wide Web*. Boston, MA: Allyn and Bacon.
Salganik, M. J. and Heckathorn, D. D. (2004). Sampling and estimation in hidden populations using respondent driven sampling. *Sociological Methodology* 34, 193 239.
Satin, A. (1993). *Survey Sampling: A Non Mathematical Guide*. Ottawa, ON: Statistics Canada.
Statistics Canada. (2001). Surveys: Start To Finish. Workshop presented by Statistical Consulting Group. Toronto. November 27 29. p 56.

Relevant Websites

http://www.raosoft.com
Sample size calculator.
http://www.surveysystem.com
Sample size calculator.
http://www.statcan.ca
Statistics Canada: data and learning resources.
http://www.abs.gov.au
The Australian Bureau of Statistics: data and learning resources.
http://qb.soc.surrey.ac.uk
The Question Bank is an information resource, in the field of social research, with a particular emphasis on quantitative survey methods.
http://www.statistics.gov.uk
UK's Office for National Statistics: data and learning resources.
http://www.census.gov
US Census Bureau: data and learning resources.
http://www.socialresearchmethods.net
Web center for social research methods. Based in Cornell University, this website consists of a web based textbook and tutorial web projects that address all of the topics in a typical introductory undergraduate or graduate course in social research methods.

Santos, M.

M. A. de Souza, Universidade de São Paulo, Campinas, Brazil

Glossary

Devil's Whole Poor people's lack of access to disposable new global technologies.
Psychosphere This is the result of beliefs, desires, passions, and habits which inspire philosophical and practical behavior and interpersonal relations of communication with the universe.
Rugosity Psychoterritorial inheritance which appears as a form, constructed space, or landscape along with what is left from the past.
Sociospatial Formation Defined as the territorial specificities of the projection of world history on national territories. This process leads to the formation of social dynamics and territorial configuration in space.
Technosphere This is the result of the increasing artificialization of space, by which nature is replaced more and more by a technical sphere, whether in a rural or urban environment.
Territorial Configuration Arrangement of natural and artificial elements present in a territory.
Tropical Flexibility Copying merchandise process that increases the popular and informal market (electronics, CDs, shoes, etc.).

Santos, Milton (1926–2001)

Milton Santos (**Figure 1**) founded a new school of Geography in Brazil based on his reinterpretation of traditional concepts of geography and the indissolubility of human and nonhuman, of objects and actions, and of time and space; he used these insights to develop a criticism of globalization which emphasized the situ ation of the poor. His reinterpretation of traditional geographical concepts led him to base his work on the foundational concepts of 'socio material techniques' (conceived of as a single, universal system of social and material methods used by man to conduct his life and relate to the physical and social world) and 'geographical space' (the indissolubility of ever more technical objects and actions projecting onto and producing space). These concepts led him to emphasize the world as a whole, intimately connected to the situation of a given place, which permeated much of his later theorizing.

Figure 1 Photo of Santos, Milton.

Milton Almeida dos Santos was born on 3 May 1926, in Brotas de Macaúbas, in the interior of the state of Bahia, in Brazil, and passed away in São Paulo on 24 June 2001. His grandparents were slaves, but his parents were teachers, and he learned algebra, French, and good manners at home. He attended secondary school in Salvador (BA), where he fell in love with geography in the writings of Josué de Castro, although he later studied law because it was considered more prestigious. In 1948 he received a degree in law from the Federal University of Bahia, and from 1949 to1964 he worked as a journalist and political advisor for the government, both state and national, with 4 years spent in France completing his doctorate degree in geography. Upon his return, he founded the Applied Geomorphology Laboratory in the Institute of Geosciences of the Federal University of Bahia.

After being imprisoned during the military dictator ship in 1964, he was eventually released and exiled. He first worked at various universities in France, but later worked in Canada, Venezuela, Tanzania, and the United States.

After returning to Brazil in 1972, he spent a full year in Bahia without finding work; he then went to São Paulo, where he initially worked as a consultant for the state government, later teaching urban geography at the University of São Paulo and, up to 1982, at the Federal University in Rio de Janeiro. In the last 20 years of his life, he returned to the University of São Paulo, this time teaching in the Department of Geography. Overall, Santos wrote 50 books, was leadership and organized 14 edited collections, contributed some 76 chapters to edited books, and wrote and published 260 scientific articles. In addition, he contributed columns to many newspapers and magazines, and gave 62 interviews to scientific magazines and newspapers in the last 5 years of his life alone. Milton was 'Doctor Honoris Causa' of 12 universities worldwide.

Santos was a free thinking intellectual, devoted to critical thought and especially concerned with the situation of the poor. His philosophy was based on the existence of socio material techniques and their role in the creation of geographical space. The consideration of the world as an indivisible whole led him to reflect on the way in which events around the world are reported electronically as they occur. The resulting planet makes global action, including economic transactions, possible (convergence of times), and led to his proposal of a sin gle, driving force, that of surplus capital (in Marxian terms). This convergence of space–time through elec tronic means, however, leads to the social exclusion of those without access to information technology. He felt, however, that these socially excluded 'poor and slow' people would some day be responsible for the future of the world.

For Santos, geographical space was regarded as a dynamic whole in constant movement, with the objects involved being the product of the social elaboration resulting from labor and the actions consisting of intentional human behavior, changing things, but also suffering change.

Santos' highly politicized conception of geographical space as 'territory in use' – a common space of all humans and institutions – hence facilitated a scientific discussion of sociospatial inequality. For Santos, the highly 'techni fied' geographical space of today incorporates human labor, with the traditional relations between people and nature having been replaced by a new relationship between man and a second nature, the product of human labor. To deal with the socioeconomic specificities of capital and production of countries, he created the con cept of sociospatial formation, based on Marx's economic and social formation.

Places, regions, and landscapes therefore took on new interpretations in Santos′s geography. Places are defined globally as spaces for social relations converging on common interests, whether or not the participants in volved are continguously located. Thus, the directors of a multinational company need not be physically present in the same space as their factories, but they can still dominate the relations and activities in these places. Regions are no longer a local construction, but rather a functional subdivision of the total space, whether national or local, and larger cities can be subdivided into regions or participate in various international regions. Landscapes, which involve finished work in space, are the material revelation of human labor, visible for as far as can be seen, simultaneously including past and present action and objects of action.

These new formulations or concepts led to some uniquely derived concepts which help understand the global world, such as technosphere and psychosphere. The former is the result of the increasing artificialization of the geographic environment, with the growing intro duction of technology in both urban and rural areas. It is accompanied by the psychosphere, consisting of all the human beliefs, desires, passions, and habits which influ ence interpersonal relations and communication with the world. These two spheres encompass interdependent processes essential for the functioning of the modern world.

Santos felt that globalization could only be under stood if the horizontal and vertical organization of regional spaces were taken into consideration. He con sidered territory as the arena of opposition of the market (which individualizes) and the techniques of production, organization, and civil society (which are generalized to include all individuals). The territory of the market supports the verticalized networks of 'egoistic' and 'utilitarian' norms of interest to hegemonic agents, whereas local and regional norms involving the con tiguous construction of lives constitutes the horizontal aspect. Vertical and horizontal aspects are seen to inter fere with each other, with the global influencing the local the local, and vice versa.

Santos accordingly borrowed concepts from eco nomics and philosophy in his construction of a new geography. His philosophical studies led him to propose the concept of the 'devil′s whole', based on three main theses. The first of these involves 'capital' as the unifying agent, with a coherent strategic orientation and homo geneous behavior. The second concerns the decisive effects of technology on social functioning, even though these may be invisible, and the third involves the active role of geographic space in social dynamics. Santos reflects a strong Anglo Saxon tendency in his interpre tation of the world on the basis of Marxism and suggests that capitalist innovations in an underdeveloped country expose the country to external influences and accentuate its dependence on the dominant means of production. Such a formation involves a unitary whole; the structure prevailing when development is controlled from outside without the participation of those living inside is not that of the country itself, but rather the global structure of the capitalist system, with the forms introduced serving the needs of the formation of the country. Santos considered this international whole to be a perverse whole belonging to 'the devil'.

From economics, Santos developed a critical reflection of globalization, denominated 'tropical flexibility'. In large cities, especially in the Third World, he suggested that the majority of the population exists under precar ious conditions, although they produce the fundamental necessities for the consumption of the more well to do classes. This division of labor, once installed, reproduces itself. Agents are mobile and exercise various roles, depending on circumstances. This metamorphosis of the work of the poor creates the flexibility found in tropical or southern countries. There are an infinite number of jobs, with an immense capacity for adaptation of social subjects, which are sustained by a hybrid of materiality and social relations in their local geographical medium. The divisions of labor are unstable, but plastic, and adapt to internal and external stimuli. Localized social relations are created *in situ*, while the solidarity imposed by hegemonic interests is dictated from outside. These provide the means of escape for the 'poor and slow', who can form relations outside the totalitarianism of ratio nality, a privilege otherwise reserved for the upper classes, and thus find a path to the future.

Santos' criticism of globalization was preceded by a profound criticism of urbanization in poor countries, with the division of urban space into two circuits: the superior and the inferior. In underdeveloped countries, cities do not function as a whole because the urban system itself is dependent on other higher level systems. The superior circuit consists mainly of banks and modern urban commerce and industries, while the inferior circuit consists basically of the noncapital intensive means of production such as small, traditional businesses and services.

Santos' view of the poor in relation to globalization is based on three considerations. The first is the idea that globalization is a fable consisting of oft repeated fantasies providing an apparently solid basis for inter pretation. These fantasies include the idea of the global village, the myth of reduced distances, and the idea that the whole world is within the reach of all. The ideology of globalization requires these fantasies. The second point is the perversity of globalization. For much of humanity, globalization is a factory of wrongs: chronic unemployment, increasing poverty, decreasing quality of life for the middle classes, hunger, homelessness, in creasing illnesses, the lack of a good quality education, and the deepening of moral crises such as corruption. These problems are all laid at the door of globalization. The third point is that there will be another globalization which will result from the tremendous miscegenation of races, cultures, and tastes around the world. This miscegenation includes the ideas of nonrational, non European philosophers, while the increase in the agglomeration of people in ever smaller areas leads to greater dynamicity of mixing. The result is socio diversity, historically even more important than biodiversity. In theory, it is thus possible to imagine a new discourse, a new metanarrative. For the first time in the history of humanity, there exists an empirical uni versality which is not the product of the mind of phil osophers, but rather of the ordinary experience of each individual.

Shortly before his death in 2001, Santos presented a manifesto synthesizing his ideas and his theory of a 'new' geography. Presented at the XII National Meeting of the Association of Brazilian Geographers, it was an affirmation of what was necessary to inhibit the fragmentation of Geography as a distinct discipline. The ten methodological precepts presented outlined a path for geographers to follow so that they could act in favor of the majority of the population. These items included the idea that geography should participate in the trans formation of society; that geographical space should be considered as territory in use, the result of historical processes and the material and social basis for new human actions; and that geography should also consider the society as a whole in movement, since a view of only portions leads to isolation and fragmentation on a theoretical plane. For Santos, the consideration of geo graphic space, not as a whole, but rather as political, economic, anthropological, touristic, and urban spaces, also leads to the fragmentation of the discipline, making it impossible to explain the whole. Fragmentation and specialization, however, have been chosen by the market and politics, although they would lead to the establishment of an infinite number of independent geographies, which would be of limited value on an op erational level. His major premise is that geographical thought must emphasize the whole, despite pressures toward fragmentation and domination by the market. A holistic view can help reestablish the unity of the world as a whole.

Rather than construct a simple, unified view of Geography and the world, Santos contended it would be preferable to develop a coherent set of integrated ideas leading to an integrated and dynamic view of the world. Geography must thus be based on a system of concepts that make it possible to comprehend the indissoluble system of objects and actions.

The commitment of Milton Santos to the improve ment of the theory and methodology of geography was linked to his central concern with constructing a coherent system of ideas capable of assuming its role of providing explanations for the world as a whole, a task which remains essential for the understanding con temporary human geographies.

See also: Globalization, Economic; Landscape; Place; Region.

Further Reading

Levy, J. (2007). *Milton Santos/Philosophe du mondial, citoyen du local*. Lausanne: Presses polytechniques et universitaires romandes.

Santos, M. (1971). *Le métier du géographe em pays sous développé*. Paris: Ed. Ophrys.

Santos, M. (1978). *Por uma Geografia Nova*. Sao Paulo: HUCITEC EDUSP.

Santos, M. (1979). *Economia Espacial: Críticas e Alternativas*. Sao Paulo: HUCITEC.

Santos, M. (1979). *Espaço e Sociedade*. Petrópolis: Vozes.

Santos, M. (1979). *The Shared Space: The Two Circuits of the Urban Economy and Spatial Repercussions*. London: Methuen.

Santos, M. (1982). *Pensando o espaço do homem*. Sao Paulo: HUCITEC.
Santos, M. (1987). *O espaço do cidadão*. Sao Paulo: Nobel.
Santos, M. (1988). *Metamorfoses do espaço habitado*. Sao Paulo: HUCITEC.
Santos, M. (1994). *Espaço, Tempo e Técnica*. Sao Paulo: HUCITEC.
Santos, M. (1996). *A Natureza Do Espaço. Técnica e Tempo Razão e Emoção*. Sao Paulo: HUCITEC.
Santos, M. (2000). *Por uma outra globalização. Do pensamento único a consciência universal*. Rio de Janeiro: Record.
Souza, M. A. (1996). *O mundo do cidadão, um cidadão do mundo*. Sao Paulo: HUCITEC.

Sauer, C.

M. Williams, Oxford University, Oxford, UK

Sauer, Carl (1889–1975)

Carl Sauer was a major figure in human geography and a number of cognate disciplines (**Figure 1**). The term 'polymath' has been applied to him because of his wide knowledge in the humanities and natural sciences, especially where these impinged on his research interests in Ibero America, plant domestication, and early human activity during the Pleistocene. He displayed an ability to link diverse evidence together and then take bold, intuitive leaps and speculative sweeps. It is a measure of his importance and appeal that there are nearly 400 works written by others about him and his work, and more keep on appearing.

Because of the broad sweep of his work it is difficult to define his precise contribution to human geography. Nevertheless, one can say that he bequeathed a number of ideas to the subject that have had an enduring legacy. He had a firm belief in the importance of fieldwork in the empathetic understanding of the quality of life of 'ordinary folk' (especially rural dwellers). He also had a keen interest in regional study, of cultural history and historical and ecological methods in geographical analysis, and a concern about environmental and cultural degradation, paralleled by a belief in cultural, ecological, and methodological diversity. All of this was informed by a deep distrust of the theoretical and mechanistic solutions associated with the social scenes from the 1940s onward, as well as a dissatisfaction with the cultural homogeneity of Western (specifically American) imperi alism. It has been fashionable of late to consider him as a conservative, who created the new orthodoxy of cultural geography (which became known as the Berkeley School). But it would probably be more accurate to typify him as a maverick who was simply 'out of step' with his contemporaries, and a revolutionary.

Figure 1 Carl Sauer.

Sauer was born in Warrenton, in the district known as 'the Missouri Rhineland' west of St. Louis, on 24 December 1889. At the age of nine his German born father took the family to Germany so that Carl would get a thorough education, especially in languages and the classics, and he spent 3 years at the Gymnasium in Calw (about 35 miles west of Stuttgart). Consequently, he was well read in German culture and literature, which was reinforced by 6 years back at the Central Wesleyan College, Warrenton, Missouri, a bilingual institution (now defunct) granting first degrees. His father, William, was professor of music and French and was also a keen student of geology and botany. He was a scholarly, religious, and kindly man whom Carl admired and respected deeply. As the College's star pupil Carl then moved to Northwestern in 1908, and completed his PhD in the first full fledged department of geography at Chicago University in 1915, under the varied and different influences of Ellen Churchill Semple, Harlan Barrows, and the geologist and charismatic head of department, Rollin Salisbury, all of who had a lasting influence on him.

During these years he was commissioned by the State Geological department to do a survey of the geography of the Upper Illinois Valley, and also completed the fieldwork for his PhD on the Ozark uplands of Missouri and adjacent parts of Arkansas. Although Carl admired Ellen Semple's charm and lecturing ability, he com pletely rejected what he saw as the sterile environmental determinist she espoused that emphasized humans as response mechanisms to physical factors, which had become the dominant paradigm in American human geography at the time. His study of the Ozarks gave an intimate insight into how the people of the uplands lived by knowing how that living was fashioned. It was his first eloquent evocation of the importance of people and place, and people in a place – and acclaimed as a model of regional cultural geography. It was the beginning of his distinctive form of a new, humane, cultural geography – informed by anthropological approaches – which seemed to offer a means of dealing with the diversity of humankind and its cultural landscapes through time. This was reinforced by his work on the Economic Land Survey of the problems of settling the denuded timber lands of northern Michigan (the Cutovers), which showed that humans radically transformed the Earth, often for the worse, and in the process created 'cultural landscapes'.

Sauer taught at the University of Michigan until 1923. Here he keenly felt the prejudice that arose with the anti German hysteria of World War I and which culminated in an attempt by some staff members to remove him. He never recovered fully from his disillusionment with those sections of American society which regarded some of its members as 'less than one 100% American', and subsequently his political sympathies during the 1920s and 1930s were very much for the 'underdog' in the form of non Anglo immigrant communities. He moved to Berkeley in 1923 where he taught for 34 years (32 years as Chair) and established one of the most distinctive graduate schools of American geography that would always be associated with 'cultural geography'.

On arriving in Berkeley, however, he made a few unsuccessful forays into the geomorphology of southern California, and soon found natural soul mates in the anthropologists Alfred Kroeber and Robert Lowie. The concept of 'culture' subsequently pervaded all his teaching and writing. In *Morphology of Landscape* he distilled an almost wholly German geographical literature, established the primacy of human agency in the formation of cultural landscapes "fashioned from a natural landscape by a culture group" and the importance of a time based approach. In addition, he placed great importance on observation and contemplation in the field – *Verstehen* – an empathetic understanding (phenomenology) and intuitive insight into behavior or object in order to achieve "a quality of reasoning at a higher plane" than the tangible facts alone.

Sauer wrote *Morphology of Landscape* in order to 'emancipate' himself from environmental determinist thinking, and realized that he would have to 'disagree pretty thoroughly with almost all the geographers in this country'. Once completed he put the epistemological game behind him and started substantive fieldwork on early settlement and society in Lower California, Arizona, and Mexico. During the 1930s he produced many works (including some on early Mexican historical geography) which were published in a new monograph series *Ibero Americana* that he founded in the same year with Alfred Kroeber and the historian H. E. Bolton. These investigations drew him into the controversy about New World plant domestication and plant origins, and into collaboration with botanists, archeologists, and ethnologists, who (from the evidence of his correspondence) he found to be more congenial intellectual company than most geographers.

With maturity Sauer's horizons got wider, his ideas more speculative, and his ethical values more refined. He played a leading role in the establishment of the United States Soil Conservation Service, was consultant to the President's Advisory Science Board, and a leading member of the Guggenheim Memorial Fund Committee for 25 years. In 1938 he wrote 'Destructive exploitation in modern colonial expansion', a brief, but seminal paper which was probably the most sustained and biting critique of the destructive social and environmental impact of European imperialism and colonialism since the writing of Karl Marx. He drew inspiration from the work of George Perkins Marsh on the human transformation of the Earth and Ernst Friedrich's concept of *Raubwirtschaft*, or destructive exploitation. His fieldwork and knowledge of Latin and Central America and their history suggested that the Spanish conquest had led to a devastating and permanent impoverishment of the land and of its cultures and societies. In his opinion, disease, warfare, and enslavement had disrupted traditional value systems and caused catastrophic population decline, being one of the first people to publicize the idea of demographic collapse. Also, he pioneered the idea that humanly induced fire had made major alterations to many ecosystems. Hence, he suggested the diffusion of technologically superior societies could affect cultures and human life just as much as it destroyed physical resources.

His world reputation was firmly established during the 1950s with works that speculated about and reviewed the place of humans on Earth. First came 'Agricultural origins and dispersals' in 1952, followed by a string of publications into the human uses of the organic world, and early humans in the Americas from the Ice Age onward. Unfortunately radiocarbon dating came too late to inform Sauer's writing; although he may not have provided the answers, he brilliantly defined the questions. Second, in 1956, he masterminded the symposium on *Man's Role in Changing the Face of the Earth*, the theme of which thereafter became his overriding interest. All his learning culminated in this volume, and particularly his chapter on 'The agency of man on Earth'. The focus here was on the capacity of humans to alter the natural environment ('deformation of the pristine'), the cult of progress, the waste that stemmed from mass production ('commodity fetishism'), and the alien intrusion of humans into world ecology. This was an early and influential statement about the degradation of the en vironment. It also had a global dimension as he returned to the theme of destructive exploitation. His suggestion was that Western "imperialism of production" was as bad as the old, colonial imperialism, and might ultimately be no better than Marxist totalitarianism; mass culture was eliminating not only biological diversity but also the cultural diversity of older and less robust societies. Mass production promoted the homogenization of society, not only in America but also globally as America became the superpower after the 1950s, with what he saw as a disdain for the 'lesser breeds.' Somehow, humans had to rise above this mindless, short term exploitative mode and try to conserve 'a good Earth.'

Although he retired in 1955 he enjoyed 20 remarkably productive years (before his death in 1975) that saw the publication of four books and a score of influential papers, all distinguished by big and speculative ideas, which were the fruits of a lifetime's reflection and unhurried reading. His reputation soared so that an "aura of sage, philosopher king, and even oracle surrounded him." Sauer's simply articulated ideas and pithy language had a resonance with many intellectuals, including Californian and the Black Mountain School of avant garde poets and literati, who extolled his work as an example of cultural and ecological sensitivity and respect, tinged with deep historical insight, and scholarship. It is clear that he also tapped a deep spring of feeling during the 1960s and 1970s at the time of Vietnam and student unrest with their concerns on the limits of the Earth and abuse of technological/political power. Yet, Sauer was a complex mix. He was congenitally nonconformist but deeply conservative, and although profoundly concerned with conservation was never formally an 'environ mentalist'; indeed, he thought the movement was little more than an 'ecological binge'.

Although Sauer distained methodological and epi stemological discussion, his writing reveals that he was, as Entrikin (1984: 387) has said, "a philosopher in spite of himself." Cultural history became his model of social science, and it was drawn from the natural sciences not the social sciences. He was basically a pragmatist who was influenced by the writings of the German cultural geographers Friedrich Ratzel and Eduard Hahn, and by the methods of geology and anthropology and their emphasis on the provisional character of working hy potheses which were no more than a means to an end. This was reinforced by his later association in Berkeley with people who worked on 'tangible things', like plant ecologists, agricultural scientists, botanists, and population geneticists. He argued for a theoretical and methodological pluralism that stemmed, in his view, naturally from the inherent diversity of nature and culture. The natural science idea of a dynamic balance and diversity that arose from organic evolution was deeply embedded in his thought and he felt that the modern world had disrupted these processes so that it was out of balance – hence his deep distrust of American capitalism and all bureaucratic systems which would destroy diversity and local community. He felt liberal social scientists, such as economists, educationalists, sociologists, and political scientists, who designed, plan ned, and directed community life were more likely to destroy than enhance it by imposing universalizing concepts of social organization, and by ignoring the in herent pluralism and diversity of nature and culture that stimulated the naive curiosity about the world, which was the essence of geography. Theoretical and normative social science grated with him in terms of its exaggerated confidence in the statistical and the inductive, and its "dialectic atmosphere." Their focus on the present pre cluded a better insight into the origins and evolution of any topic, and gave "an exaggerated accent on contemporaneity."

Sauer's social science was firmly based on history and geography – this was cultural history. Consequently the Sons of Daedalus (Daedalus was the cunning and arro gant craftsman of Greek mythology who thought he could invent anything) were the target of his criticism because their well funded procedures and programs emasculated the independence of the impressionable younger scholars. His 'Folkways of the social sciences' was an impassioned plea to social scientists to "give back the search for truth and beauty to the individual scholar to grow in grace as best he can" and to reinstate time and place into the study of America. Sauer was aware that he was 'out of step' with his colleagues and intellectuals, and spoke of himself as an 'unofficial outsider' and even as a 'peasant'. That was partially an outcome of his upbringing and rural youth during a simpler age. But he anticipated society's disenchantment and fears with progress, science, the elimination of diversity, and the degradation of the environment. Perhaps the ultimate relevance of Sauer's work was in his groping toward a rapprochement be tween the social sciences, the humanities, and the bio logical life sciences. By being behind he was far ahead.

See also: Berkeley School; Cultural Geography; Landscape; Place; Region.

Further Reading

Callahan, R. (ed.) (1981). *Selected Essays 1963 1975: Carl O. Sauer.* Berkeley, CA: Turtle Island Foundation.

Entrikin, J. N. (1984). Carl O. Sauer, philosopher in spite of himself. *Geographical Review* 74, 387 408.

Hooson, D. (1981). Carl O. Sauer. In Blouet, B. W. (ed.) *The Origins of Academic Geography in the United States*, pp 165 174. Hamden, CT: Archon Books.

Kenzer, M. S. (ed.) (1987). *Carl O. Sauer: A Tribute*. Corvallis, OR: Oregon State University Press.

Kenzer, M. S. and Mathewson, K. (eds.) (2003). *Culture, Land, and Legacy: Perspectives on Carl Sauer and Berkeley School Geography*. Baton Rouge, LA: Louisiana State University.

Leighly, J. (ed.) (1963). *Land and Life: A Selection from the Writings of Carl Ortwin Sauer*. Berkeley, CA: University of California Press.

Leighly, J. (1978). Carl Ortwin Sauer, 1889 1975. In Freeman, T. W. (ed.) *Geographers' Biobibliographical Studies*, vol. 2, pp 99 108. London: Mansell.

Sauer, C. O. (1925). Morphology of landscape. *Publications in Geography* 2(9), 271 302. (Reprinted in Leighly, 1963: 315 350.)

Sauer, C. O. (1938). Destructive exploitation in modern colonial expansion. *Comptes Rendus du Congrés International de Géographie* 2(Sect. 3c), 494 499. (Reprinted in Callahan, 1981: 330 363.)

Sauer, C. O. (1952). *Agricultural Origins and Dispersals*. Bowman Memorial Lecture, Series 2. New York: American Geographical Society.

Sauer, C. O. (ed.) (1952). *Folkways of Social Science. In the Social Sciences at Mid century: Papers Delivered at the Dedication of Ford Hall*. Minneapolis, MN: University of Minnesota Press. (Reprinted in Leighly, 1963: 380 388.)

Sauer, C. O. (1956). The agency of man on earth. In Thomas, W. L. (ed.) *Man's Role in Changing the Face of the Earth*, pp 49 69. Chicago, IL: University Press of Chicago.

Williams, M. (1983). The apple of my eye: Carl Sauer and historical geography. *Journal of Historical Geography* 9, 1 28.

Scale

N. F. Sayre and A. V. Di Vittorio, University of California – Berkeley, Berkeley, CA, USA

Glossary

Construction of Scale The process of creating observational scale.

Extent The range of measurement of a scale.

Observational Scale Epistemological moment of scale as defined by extent and resolution; the reference by which attributes and processes are measured.

Operational Scale Ontological moment of scale; the spatiotemporal attributes of a particular process.

Production of Scale The process of creating operational scale.

Resolution (or Grain) The smallest measurable unit of a scale.

Scale as Level A qualitative category defined by position on a quantitative scale.

Scale as Relation A relative framework for comparing attributes or processes through observation and relational category definition.

Scale as Size A quantitative measure of an attribute or process.

Introduction

Nearly everyone acknowledges the importance of scale in human geography, but a common understanding of what it means and how to use it remains stubbornly elusive. Articles on the subject have proliferated in the past 15 years, reflecting an effective consensus that scale belongs among such fundamental geographical categories as space, place, environment, and location. Indeed, interest has burgeoned throughout geography as a whole and in many natural science disciplines as well. Ecologists, for example, have declared scale a fundamental concept for all sciences. Yet at the same time, a plethora of definitions, uses, and theories of scale has generated considerable confusion. Common definitions of scale do not exist within disciplines, authors use scale differently within articles, and debates rage over whether scale is an epistemological tool or a material attribute of the world.

This article reviews the uses and meanings of scale in recent human geography to demonstrate the variability, complexity, and power of scale. The focus here is on noncartographic concepts of scale, yet the topic requires a brief introduction to cartographic scale for completeness. Cartographic scale refers to the spatial relationship between a map and the space it represents: it is the mathematical ratio of units on a map to units in the world. A map scale of 1: 62 500 depicts 1 mile in the world as 1 inch on the map. A small scale map shows more area than a large scale map, thus cartographers use the terms large and small scale in an opposite way to scholars in other fields. Cartographic scale is not prominent in the following discussion, but it shares a common basis with all other types of scale: relationality.

Scale has multiple meanings that are distinct yet internally related to one another. On the one hand, they are genuinely different meanings for the same term – one cannot declare one of them right and the others wrong – and in some ways they appear incompatible. On the other hand, they are conceptually linked, and the fact that they share a single term is not a coincidence. This conceptual and terminological complexity has led to frequent conflation and persistent confusion.

Fortunately, scholars can alleviate confusion if they understand scale in terms of relations and processes. The significance of scale lies in the processes that continually reconstitute it as theoretically, materially, and politically important; rather than in a particular geographical scale, such as the local or the global. These processes range from academic analysis to political border definition to road building, and each process and its outcomes change with changes in scale. This is because a change in scale represents a shift from one set of relations among processes to another. Then what is scale and how does it change? First of all, the structure of scale comprises three interrelated aspects – size, level, and relation. Size and level are relatively clear on their own, but they can obscure the fundamental aspect of scale as relation.

Scale as Size, Level, and Relation: An Illustration

To identify some salient features of scale without entangling current debates in human geography, it is useful to begin by considering another kind of scale altogether: the kind used to weigh things.

Imagine that you wish to weigh the following six items: a gold earring, a newborn baby, yourself, a truckload of waste, a trainload of grain, and the carbon emitted into the atmosphere last year. For present purposes, we ignore the distinction between mass and weight. In each case you seek to know the same quality – namely, weight – and thus all six could be expressed in one unit of measurement, such as grams. In theory, a single scale is sufficient because the weights differ only in quantity; you are abstracting from all six items in the same way.

This illustrates the first aspect of scale: as size. Whether one is measuring weight, distance, area, volume, velocity, duration, temperature, or some other quality, one uses a scale to do so. Scale as size is absolute insofar as it relies on the division of abstract weight into uniform segments for measuring particular weight. Scale in this sense is a tool, arbitrary in its units, that an observer employs to derive knowledge about the world; by abstracting from qualitatively different things in a standardized way, a scale renders them comparable in quantitative terms.

But of course you would not actually use the same scale in all six cases. The apparatuses and methods for measurement would be quite different – for example, we measure a portion of carbon emissions and estimate the total with models – and the units would also be different: grams for the earring, for example, and gigatons for the carbon. Why not use a single scale? The simple answer is technical: no single apparatus has been devised that could weigh both an earring and a train; a scale strong enough to withstand a train could not detect the weight of a single earring. One cannot put atmospheric carbon on a scale at all. We weigh different things with different scales, understood as both apparatuses and units of measurement: gold with a jeweler's scale, in grams (or ounces); humans with a bathroom scale, in kilograms (or pounds); grain with a railroad scale, in metric tons; and global carbon emissions with optical and chemical detectors and models, in gigatons. In purely quantitative terms, these different scales are separated only by orders of magnitude – all could be reduced, without distortion, to a single scale expressed in scientific notation – and each scale also represents a common order of things measurable with it. And in practice, each scale helps to constitute an order of things: gold and cocaine are very different substances, but they are weighed in the same kind of scale and expressed in the same units, and in this sense they are similar.

This illustrates the second aspect of scale: as level. Things measured in grams and things measured in tons occupy different levels of weightiness, so to speak. Note that the terms scale and level are interchangeable in this sense of scale. A scale – such as a bathroom scale – implies the existence of a class of items that can be measured on that scale, and (generally) a unit of measurement that goes with that class. Insofar as different classes can be arranged conceptually according to another, second order scale, the first order scales (jeweler's, bathroom, truck, railroad) become distinct levels in the second order scale. Each level represents one of a series of regions ordered along a conceptual measurement scale (**Figure 1**). Note further that these levels are distinguished qualitatively. When we observe that the weight of the earring is on a different scale from that of a train, we do not deny that quantitative reduction is possible, but rather recognize that the quantitative difference is so great as to represent, in some significant sense, a difference in kind. Things measured in a jeweler's scale will not register on a railroad scale, and things measured on a railroad scale will crush a jeweler's scale to smithereens. Scale as level presupposes, but abstracts from, scale as size.

Technical limitations are not the only reason for using different scales in different situations. Even if a railroad scale capable of weighing a gold earring could be engineered, there would be no reason to use it for that purpose. Similarly, there is no need to know the weight of a trainload of grain down to the nearest gram. You could easily weigh yourself in grams or ounces – the way you would a newborn – but why bother? For the newborn, such precision is warranted because it can signify important differences in health and vitality, but by childhood this is no longer the case. In adults, such precision would merely generate noise, as your weight would rise and fall noticeably over the course of a day. Scales are internally related to the practical purposes for which they are devised. If gold were as cheap as grain, we would measure it in kilos or tons rather than grams. Conversely, the techniques of aggregation and measurement – including the apparatuses known as scales – can have profound effects on the practices in which they are employed. Take, for example, nineteenth century Midwestern US grain production: the dramatic increase in corn distribution and trade due to railroad and silo technologies rapidly transformed farming.

Here we see the third aspect of scale: as relation. It concerns, for example, the relation of birth weight to infant health (why newborns are weighed in grams), or the relation of weight to value (why a gold earring is worth measuring in grams whereas a plastic one is not). When you drive a truckload of waste to the dump, it is weighed in units suited to the interaction of the truck, the dump, and the cost of disposing of the waste. If any one of these were to change significantly – if your waste were radioactive, say – the scale would change as well. Or change might occur the other way around: if dump space became extremely scarce, or if truck scales capable of much greater precision became sufficiently widespread, then dump charges might soon be calculated per kilo or even per gram. This would in turn redound on your behavior in deciding how much waste to generate or haul to the dump.

Scale as relation abstracts from scale as level to ask why processes have certain spatiotemporal dimensions, or – put another way – what difference those dimensions make in the outcomes of processes. If scale as size is quantitative, and scale as level is qualitative, scale as relation concerns where quantitative change becomes qualitative change. A relative scale defines and orders levels based on observed relationships between processes, or entities, rather than forcing observations into

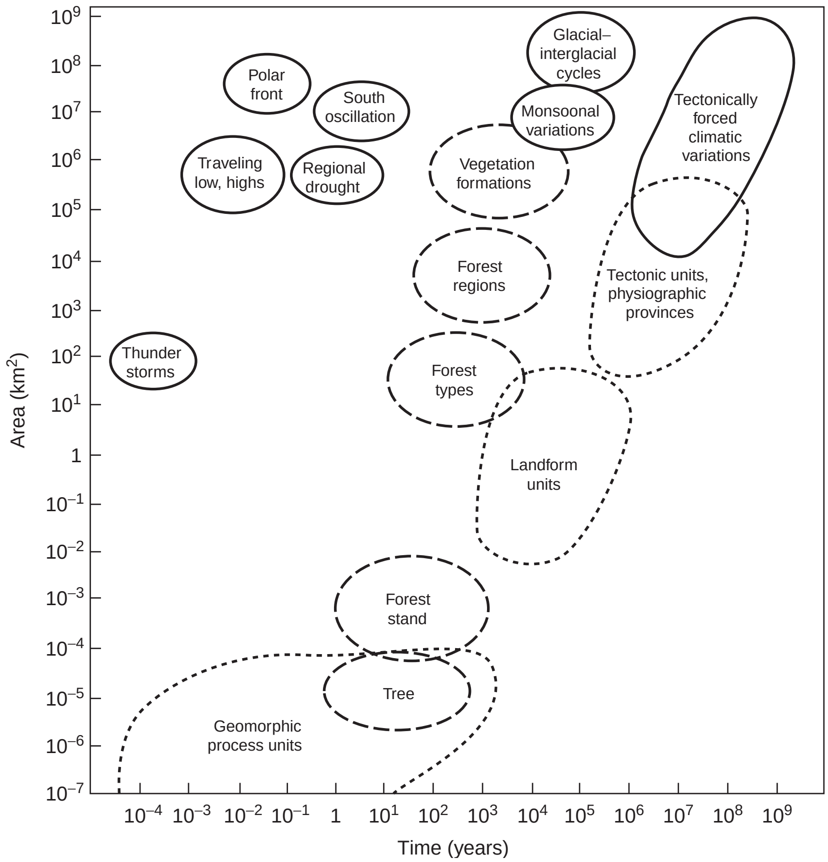

Figure 1 Delineation of conceptual levels on a quantitative scale. Solid lines denote climatic levels, dashed lines denote forest ecology levels, and dotted lines denote geomorphic levels. From Meyer, W. B., Gregory, D., Turner, B. L. II. and McDowell, P. F. (1992). The local-global continuum. In Abler, R. F., Marcus, M. G. & Olson, J. M. (eds.) *Geography's Inner Worlds*, pp 255 279. New Brunswick, NJ: Rutgers University Press.

predefined levels. The emphasis is on how entities are positioned in relation to each other. Geographers use hierarchy, network, and complex system theories to organize observations into relative scales for analysis (see below). Musical scales provide an analogy for demon strating how relative scales represent significant changes in quality with quantity. The relations among a whole ensemble of notes define a musical scale, and a quanti tative shift in scale represents a rearrangement of relations between the notes. The new scale will produce a qualitatively different set of sounds than the old. Simi larly, a different arrangement of geographical levels in a relative scale will represent a qualitatively different group of processes.

Current Scale Debates in Human Geography

How does the preceding illustration help to understand scale in human geography? **Table 1** summarizes the various meanings of scale discussed above. The uses of scale by human geographers in recent decades all fall into one or more of the above meanings, but the distinctions among them have frequently gone unrecognized. Often, scholars have blurred two meanings together, either in formulating their own arguments or in juxtaposing their arguments to those of other scholars. Such conflation is understandable insofar as scale does, indeed, refer to: (1) measuring phenomena – scale as size; (2) organizing phenomena into classes according to such measurements – scale as level; and (3) exploring the interactions among (processes at different) levels – scale as relation. The result, however, has been persistent confusion and occa sional, unnecessary acrimony. The following overview of scale debates will highlight similarities among the various uses of scale.

Scale's Metaphysical Status

First, there is debate over whether scale is an epi stemological or an ontological category. On the one hand, scholars argue that scale is simply a methodological issue. One's choice of scale may have material effects, but it does not exist independently of human observation. Many commentators have discussed the metaphors used

Table 1 The multiple aspects of scale

Scale as	*Also known as*	*Metaphysical status*	*Expressed*	*Consists of*	*Concerned to measure or understand*
Size	Observational scale; absolute scale	Epistemological	Quantitatively	Grain and extent	Weight, size, area, distance, duration, speed, etc.
Level	Conceptual scale. May be observational or operational; ongoing effort to reduce disparity between the two	Either epistemological or ontological	Qualitatively	Multiple scales-as-size arranged functionally and/or hierarchically	Different orders within one such metric
Relation	Operational scale; relative scale	Ontological	Both: where change in quantity becomes change in quality	Processes interacting across scales-as-levels; relations between scales (e.g., how to 'scale up' or 'scale down')	Scaling effects; thresholds or nonlinearities produced by cross-scale interactions; scale mismatches

to express or illustrate scale, such as pyramids, scaffolds, nested hierarchies, or Russian dolls, suggesting that scale is a figurative device or conceptual framework, not something whose reality is independent of human cognition. The focus therefore falls on the metaphors, measurement tools, and metrics employed to generate questions and knowledge about the world. For human geographers, such scales conventionally include the human body, household, neighborhood, city, metropolitan area, province/state, nation state, continent, and globe. A research project might employ one or more of these scales; arrayed as a series, they represent qualitatively distinct levels for organizing and conducting research. For example, studies of globalization focus on economic and political processes across the local, national, and global levels; studies of childcare address personal, economic, and political processes across the individual, household, community, municipal, provincial, and national levels; and studies of the empowerment of American women examine how economic, political, and cultural practices at the household scale influence related processes across geographic levels of analysis.

On the other hand, much of the recent interest in geographical scale stems from the insight that it is not only a methodological issue. After all, geographical scales such as the household, the urban, and the national are not only measurement tools – they also refer to material and social realities. Some geographers obscure the material and political effects of spatial scale by focusing on the methodological construction of scale. What is lost is the influence of scale categories on politics, practical activity, and the material production of space. Recognition of this influence reflects a broader shift in geography from absolute to relative conceptions of space. For example, capitalism produces a historically unique material geography that scholars can readily differentiate in terms of space, economic production, reproduction, labor and industrial relations, and politics. The resulting categories, however, are relative to capitalism and would not necessarily be applicable to noncapitalist societies.

Clearly, there is truth in both positions; the two sides are implicitly referring to different aspects of scale and, consequently, arguing past each other. In terms of the introductory illustration, grams, kilograms, tons, and gigatons are human constructs that allow us to know and compare the weight of items in the world, but they are not themselves ontological categories. The observer uses a scale to measure things in the world, and this choice is principally a methodological consideration. Indeed, it is an unavoidable issue for any science: science seeks to identify and explain patterns, and this requires an epistemological reference – an observational scale.

Human geographers rarely discuss observational scale, but biophysical geographers, landscape ecologists, and other natural scientists define observational scale by extent and resolution. Resolution, also known as grain, is the smallest unit of measurement employed in a given method; extent is the overall size, magnitude, or duration over which measurements are taken. The extent and resolution of observational scale determine the patterns an observer can resolve and also influence particular observations (**Figures 2** and **3**). Finer grain studies typically have smaller extents, whereas large extents typically require coarser grains. In terms of the introductory illustration, grain refers to the precision (or smallest unit) of a scale; extent to the heaviest amount that a scale can weigh. The two tend to covary: the jeweler's scale, with a grain less than a gram, typically

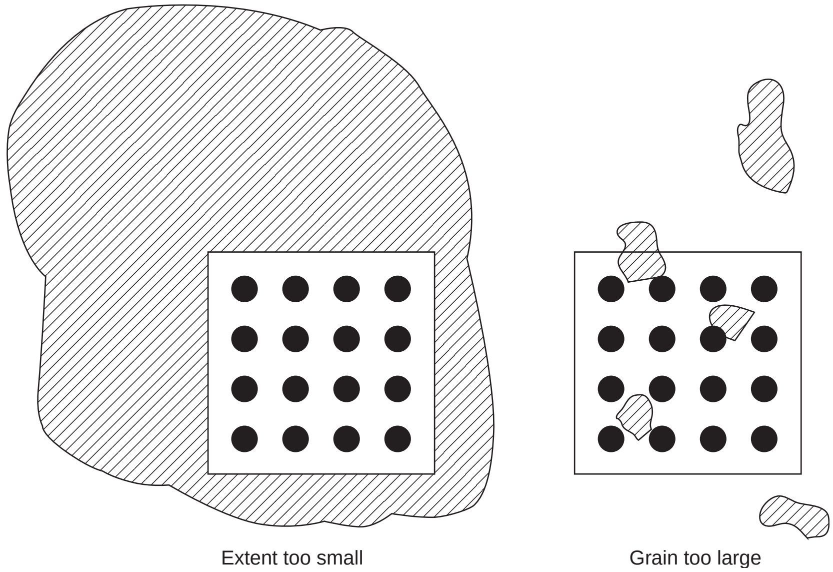

Figure 2 Grain and extent of observational scale. If the extent is too small, then the observer cannot measure the entire object of study. If the grain is too large, then the observer will not obtain data appropriate to the object of study. In both cases the observed pattern does not accurately represent the object of study. From Goodwin, B. J. and Fahrig, L. (1998). Spatial scaling and animal population dynamics. In Peterson, D. L. & Parker, V. T. (eds.) *Ecological Scale*, pp 193 206. New York: Columbia University Press.

Figure 3 Choice of scale affects the observed orientation of the Maine coast. From Giampietro, M. (1994). Sustainability and technological development in agriculture: A critical appraisal of genetic engineering. *BioScience* 44(10), 677 698.

cannot weigh items of tens of kilograms or more; the truck scale can weigh something of many tons, but not with precision down to grams.

Scholars on the other side of this debate are tacitly referring to operational scale – that is, scale as an attribute or characteristic of the processes being studied, not merely as a lens or metric of observation. For processes such as governance, economic production, or social reproduction, it is clear that spatial scale can and does have profound material effects. These processes operate within bounded (if not definite) realms and have determinate (but not static) spheres of influence. For example, changes in fishery governance altered ownership, employment, and production practices in Alaska; need for childcare support created different financial responsibilities for the Canadian state, its provinces, and cities that resulted in different types of administration and support in different cities; and local economics and politics can affect global capital flows.

Operational and observational scale are distinct but related analytics: if the patterns one sees depend on the observational scale one employs, this suggests that different processes operate at different spatial (and temporal) scales. One cannot posit *a priori* the correct observational scale for a given process; rather, one must ascertain this empirically, bearing in mind that patterns may be artifacts of one's observational scale. In principle, any discrepancy between observational and operational scale should diminish over time as our knowledge and understanding of the process improve. In summary, scale cannot be understood as either ontological or epistemological: these are not mutually exclusive, nor can they be collapsed into one another. Rather, scale has both ontological and epistemological moments.

For example, Peter Taylor's foundational definition of geographical scale incorporates both observational and operational scales. His framework is an observational scale that correlates with distinctly urban, national, and world processes. Many geographers have since challenged Taylor's framework on the grounds that the scale and the processes are not fixed in relation to each other; the levels of observation and the operational extent of the processes are historically contingent. These critiques have implicitly addressed both the construction of observational scale and the production of operational scale while asserting priority of one over the other.

Scale and Hierarchy

A second point of debate concerns whether or in what sense scale is hierarchical. Taylor's framework is inherently hierarchical. The urban, national, and world levels are vertically nested and complementary to horizontal differentiation in space. Hierarchy theory does indeed provide a framework for constituting relative, second order scales. A hierarchy of levels expresses relationships between and within distinct orders of a particular quality. The spatial hierarchy in **Figure 4** represents a four level conceptual scale arranged by size. Each level represents a set of processes with stronger intralevel than interlevel connections. Moving from a smaller to a larger level represents a decrease in the strength of intralevel connections, and the level arrangement restricts the influence of processes at one level on those of another.

Some geographers, however, contend that relations between levels are not necessarily hierarchical (**Figure 5**). They argue that hierarchical paths do not capture the

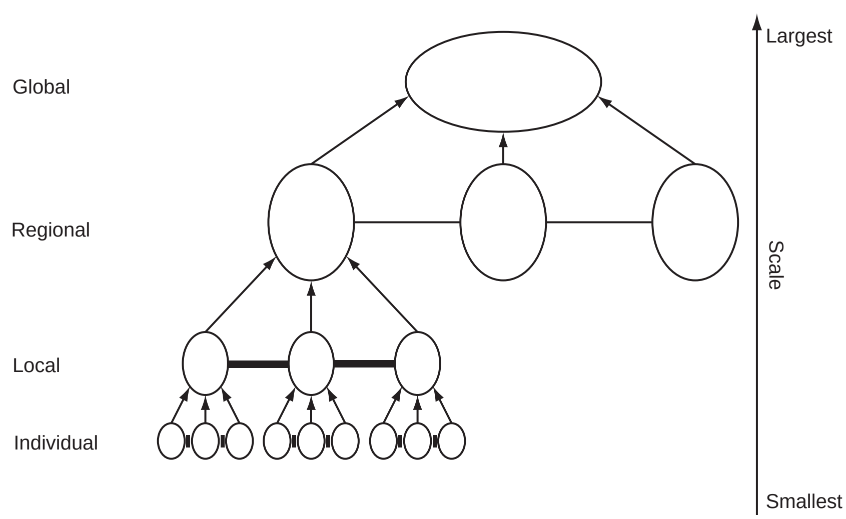

Figure 4 A scalar hierarchy. Levels higher in the hierarchy are at larger spatial scales. The thickness of the interconnecting lines indicates the degree of interaction within levels; thicker lines indicate a greater probability of interaction. From Goodwin, B. J. and Fahrig, L. (1998). Spatial scaling and animal population dynamics. In Peterson, D. L. & Parker, V. T. (eds.) *Ecological Scale*, pp 193 206. New York: Columbia University Press.

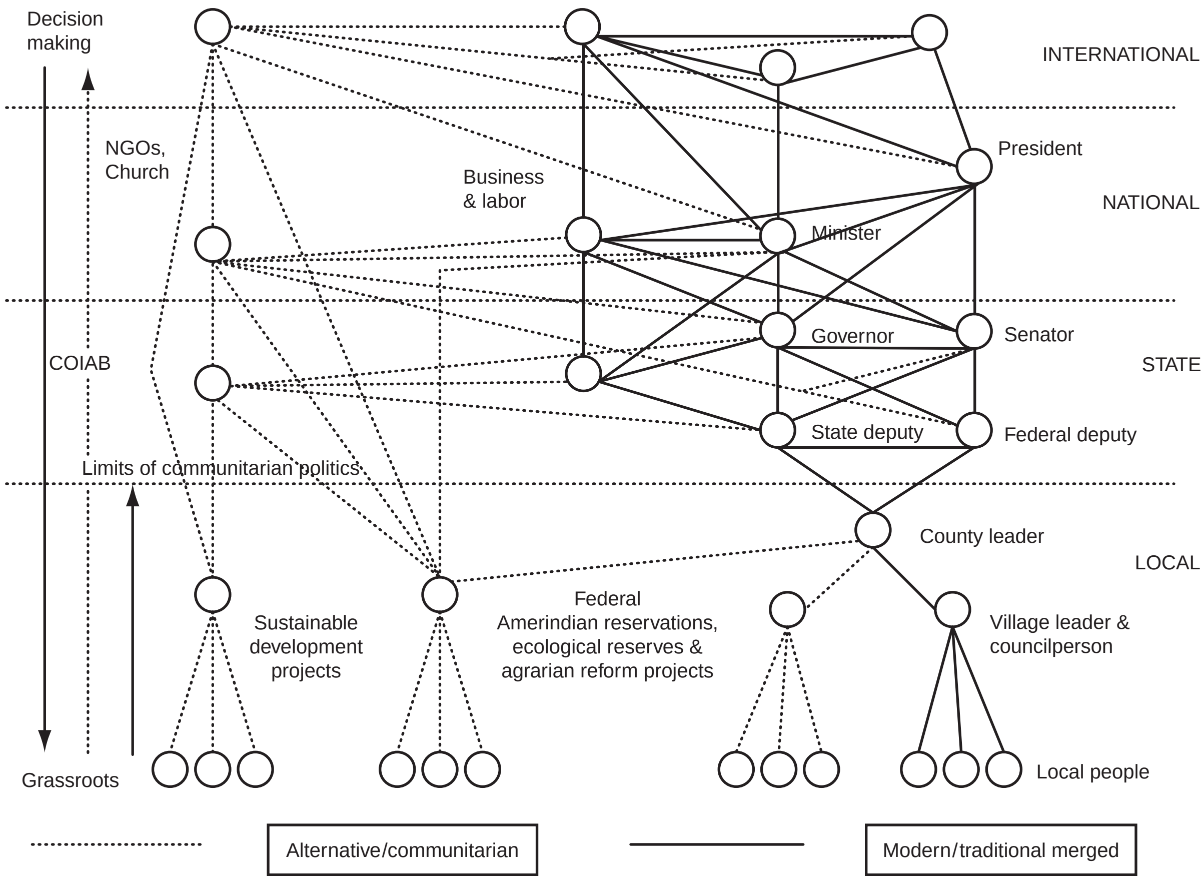

Figure 5 Hierarchical vs. nonhierarchical politics. From Hoefle, S. W. (2006). Eliminating scale and killing the goose that laid the golden egg? *Transactions of the Institute of British Geographers* 31, 238 243.

complexity and dynamics of actual interactions. Alter native models include scaled networks that span space to provide links between individual agents; and self organizing complex systems with emergent properties that operate at particular scales. One group of scholars insists that scale's hierarchical baggage renders it in effective and proposes to eliminate scale from the geo graphical lexicon.

In spite of the debate, scale is remarkably useful in all of its manifestations. Nested levels provide insights to post Fordist patterns of globalization. It is argued that globally oriented capital accumulation has surpassed nationally organized capital accumulation in magnitude and importance, while local economics and politics have gained prominence in a global rather than national market. National policies, however, still mediate inter actions between local and global processes, as demon strated by US fishery policies that boosted local production by encouraging foreign investment. Geog raphers often refer to a change in dominant level of a process, such as national to global capital accumu lation, as a jump in scale rather than a jump in level because such change alters the extent and outcomes of a process.

Scholars have also characterized globalization as a pattern of interdependent economic, political, and regulatory processes that continually (re)produce their own operational scales. Similarly, European transnational networks have shifted political power from the national to the municipal and regional levels. Furthermore, indi vidual purchasing and employment behaviors, which are often gendered and personal, emerge nonhierarchically in national and global business, marketing, and pro duction practices.

The Social Construction of Scale

The notion that scale is socially constructed is widely affirmed; indeed, it is one of the few points about scale on which human geographers are in agreement. Yet the implications of this are variously understood. Generally, scholars invoke social construction to emphasize the political significance and real world ramifications of scale – otherwise known as the politics of scale. The point is that geographical scale is not fixed: ongoing interactions among humans in their social and physical environment create local, regional, global, etc., realms of action. The fact that people can recognize these as levels in the world indicates that socially constructed scales have material effects and ontological status. Often, however, scholars contrast the social construction of scale to conceptions of scale as ontologically preordained – in the sense that levels of analysis are given by an immutable set of processes, independent of human cognition. This is clearly intended to compel critical scrutiny of the scales one employs in research – rather like pointing out that weighing a newborn on a railroad scale won't produce knowledge – and therefore resides in scale's epistemological moment. The confusion is further compounded by the fact that relations, unlike things, are not physical objects that one can directly observe. People constitute the national level through practice and analysis, but it can only be measured indirectly, by observing processes that operate at the national level.

To avoid this confusion, it is useful to distinguish between the construction and the production of scale. Scales are constructed in the epistemological moment, as observational scales for measurement (such as grams and tons for weight, or local and national for space). These should not be taken as ontologically given because they are not ontological at all; nor should they be taken for granted (or otherwise used uncritically) because they have important epistemological and methodological consequences. It need hardly be added that such construction is social. Scales are produced, on the other hand, in the ontological moment, as the operational scales of processes that occur independent of any act of observation. To say they are produced is to emphasize their historical specificity and contingency while allowing for both social and other determinations: insofar as they are ontologically real, these scales interact with material processes (e.g., biological, geological, and chemical) that may exceed or stand apart from the social realm.

Much of the confusion attending these debates can be traced to a failure to theorize level. This middle aspect of scale occupies a curious, polyvalent position: a level may be observational or operational, epistemological or ontological. Ecologists are more explicit in their use of the term; their levels of individual, population, community, ecosystem, landscape, biome, and biosphere are not scales at all. Rather, ecologists delineate observed processes into levels along a spatiotemporal scale based upon size and duration. Thus, they use observational scales to measure operational scales of processes, and then categorize the processes into epistemological levels. These levels are usually defined relative to their position on a quantitative scale: a biome is larger than an ecosystem, which is larger than an individual.

Geographers, however, recognize that each level, once postulated, can also define the extent and resolution of an observational scale. National level processes operate over particular areas and times that limit the observations of a national study. People construct the national level through ideology, and produce it through practice. Thus, a level can simultaneously be epistemological through analysis and planning, and ontological through praxis. Geographers observe processes at particular levels along one observational scale, and then define an observational scale for each level based upon the operational limits of its processes. The result is that national becomes both a level and a scale: the former is a level on a scale comprising local, global, etc., levels; and the latter is a scale used to analyze relations among processes operating within national limits. Studies of globalization often utilize a three level scale – local, national, and global – to demonstrate a shift in capital accumulation from the national to the global along with a shift in political influence from the national to the local. But examination of national and local scale economic and political processes associated with Alaskan fisheries reveals the strengthening of national political power at the expense of national capital with the effect of increasing local production through foreign investment.

In short, a level can be postulated as a category for classifying and thinking about empirical phenomena, or it can be a characteristic or attribute of the phenomena themselves. Human geographers are correct to point out, moreover, that the two kinds of levels may interact and affect each other – how people think and talk about the world can affect how the world and its people are organized – into nation states, for example.

Studies of fisheries provide good examples of the reciprocal relationship between the construction of observational scales and the production of operational scales. Geographers who study human–environment interactions increasingly characterize environmental problems as the result of mismatches between ecological and geographical scales. Processes operating at individual, population, community, ecosystem, seascape, and global levels govern fish populations; and processes operating at individual, family, community, municipal, regional, national, and global levels govern human activity. These two observational scales have similar extent, but between them the operational scales of biological, physical, and social processes often do not match (**Figure 6**). Likewise, the ecological and geographical levels do not match because ecologists and geographers measure operational scales of different processes to determine respective levels of analysis. According to this model, economic and biological collapse of human and fish populations occurs when the observational scale of research and decision making contributes to human practices that operate at different scales than those of nonhuman processes. The mismatch between operational scales of fishing based upon economics, politics, and culture; and fish reproduction based upon biology, climate, and physical environment undermines the viability of both systems. It follows that the social construction of

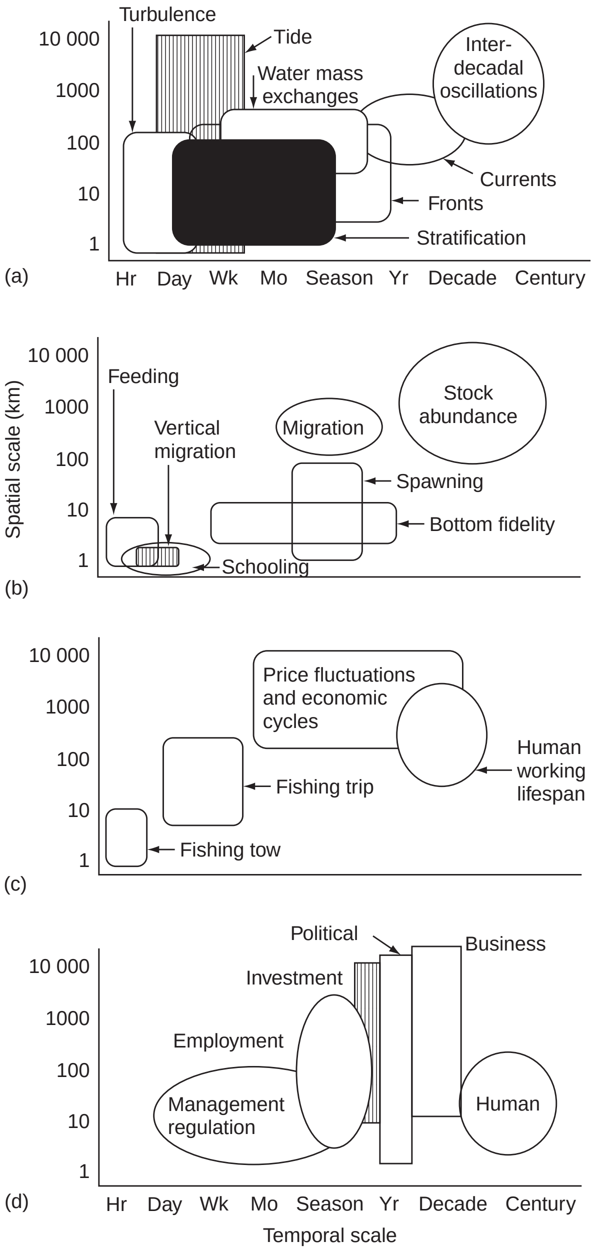

Figure 6 Space/timescale diagram of characteristic processes from the natural sciences: (a) physical; (b) biological; and from the social sciences: (c) fishing; (d) fishing communities. From Perry, R. I. and Ommer, R. E. (2003). Scale issues in marine ecosystems and human interactions. *Fisheries Oceanography* 12(4/5), 513 522.

observational scale needs to integrate human and non human processes to match the operational scales at which coupled human natural systems operate.

Conclusion: Processes and Relations

The many uses and meanings of scale can be reconciled if they are understood in terms of relations and processes. Ultimately, even scale as size is relational: measurement relates one thing to another to produce information or knowledge about the thing being measured. However, the scientific importance of scale lies not in measure ments of size, but in exploring relations among phe nomena so measured. These relations are produced by processes – hence the theoretical priority of process over scale *per se* – and both the processes and their relations are ontologically real regardless of whether they are anthropogenic or not. Processes are simultaneously spa tial and temporal; there is no single correct scale for science, although any given process may have an appropriate or best scale for research. Scale has both an ontological moment, insofar as it is an immanent characteristic of material processes, and an epistemo logical moment, insofar as one's scale of observation may determine the patterns (or lack thereof) that one observes. Understanding these as moments of a dialectic underscores the manifold relationality embed ded in scale: between subjects and objects, observers and observed, theory and practice, and knowledge and the world.

See also: Capital and Space; Complexity Theory, Nonlinear Dynamic Spatial Systems; Critical Geography; Dialectical Reasoning and Dialectical Materialism; Local-Global; Nature, Social; Networks; Relational Economic Geography; Space-Time; State Theory.

Further Reading

Brenner, N. (1998). Between fixity and motion: Accumulation, territorial organization and the historical geography of spatial scales. *Environment and Planning D: Society and Space* 16, 459 481.

Brenner, N. (2004). *New State Spaces: Urban Governance and the Rescaling of Statehood*. New York: Oxford University Press.

Giampietro, M. (1994). Sustainability and technological development in agriculture: A critical appraisal of genetic engineering. *BioScience* 44(10), 677 698.

Gibson, C. C., Ostrom, E. and Ahn, T. K. (2000). The concept of scale and the human dimensions of global change: A survey. *Ecological Economics* 32, 217 239.

Goodwin, B. J. and Fahrig, L. (1998). Spatial scaling and animal population dynamics. In Peterson, D. L. & Parker, V. T. (eds.) *Ecological Scale*, pp 193 206. New York: Columbia University Press.

Gunderson, L. H. and Holling, C. S. (2002). *Panarchy: Understanding Transformations in Human and Natural Systems*. Washington, DC: Island Press.

Hoefle, S. W. (2006). Eliminating scale and killing the goose that laid the golden egg? *Transactions of the Institute of British Geographers* 31, 238 243.

Howitt, R. (1998). Scale as relation: Musical metaphors of geographical scale. *Area* 30, 49 58.

Howitt, R. (2003). Scale. In Agnew, J., Mitchell, K. & O'Tuathail, G. (eds.) *A Companion to Political Geography*, pp 138 157. Oxford: Blackwell.

Jessop, B. (2000). The crisis of the national spatio temporal fix and the tendential ecological dominance of globalizing capitalism. *International Journal of Urban and Regional Research* 24, 323 360.

Levin, S. A. (1992). The problem of pattern and scale in ecology. *Ecology* 73, 1943 1967.

Marston, S. A. (2000). The social construction of scale. *Progress in Human Geography* 24, 219–242.
Marston, S. A., Jones, J. P. III. and Woodward, K. (2005). Human geography without scale. *Transactions of the Institute of British Geographers* 30, 416–432.
Meyer, W. B., Gregory, D., Turner, B. L. II. and McDowell, P. F. (1992). The local–global continuum. In Abler, R. F., Marcus, M. G. & Olson, J. M. (eds.) *Geography's Inner Worlds*, pp 255–279. New Brunswick, NJ: Rutgers University Press.
Perry, R. I. and Ommer, R. E. (2003). Scale issues in marine ecosystems and human interactions. *Fisheries Oceanography* 12(4/5), 513–522.
Peterson, D. L. and Parker, V. T. (eds.) (1998). *Ecological Scale: Theory and Application*, pp 499–522. New York: Columbia University Press.
Sayre, N. F. (2005). Ecological and geographical scale: Parallels and potential for integration. *Progress in Human Geography* 29, 276–290.
Sheppard, E. and McMaster, R. B. (eds.) (2004). *Scale and Geographic Inquiry: Nature, Society, and Method*. Malden, MA: Blackwell Publishing.
Smith, N. (1984). *Uneven Development: Nature, Capital, and the Production of Space* (2nd edn. 1990). Oxford: Basil Blackwell.
Swyngedouw, E. (1997). Neither global nor local: 'Glocalization' and the politics of scale. In Cox, K. R. (ed.) *Spaces of Globalization: Reasserting the Power of the Local*, pp 137–166. New York: Guilford Press.
Taylor, P. J. (1982). A materialist framework for political geography. *Transactions Institute of British Geographers, New Series* 7, 15–34.

Scale Analytical

T. W. Crawford, East Carolina University, Greenville, NC, USA

Glossary

Centroid The geometric center point of a two-dimensional object.

Critical Geography Scholarship that emphasizes the social production of space and the roles of gender, race, ethnicity, materialism, and political economy.

Degree The number of links connecting a given node in social, spatial, or sociospatial networks.

Granularity Measurement scale of data observations that ranges along a continuum of coarse to fine resolution.

Hierarchy Theory Ecological thought arguing that systems are best understood by examining at least three scale levels to include a focal scale of interest and scales one step above and below the focal scale.

Modifiable Area Units Artificially imposed boundaries for two-dimensional areal objects that can be rezoned or resized leading to aggregation and scale effects in analytical results.

Scale Dependence A condition where patterns and/or process relationships change according to scales of observation.

Scale Invariance Geometric objects, pattern/process relationships, or statistical frequency distributions that appear identical or self-similar upon scale reduction or magnification.

Semivariance The sum of squared differences between coupled observations at defined distance intervals divided by 2 times the number of observations per interval, used to identify scales where spatial dependence is operative.

Spatial Analytic Tradition A paradigm that emphasizes locational analysis typically using quantitative methods.

Introduction

Scale is a fundamental concept that permeates virtually all topics in human geography. Prior to the discipline's quantitative revolution and subsequent engagements with post positivist paradigms, scale was mainly con ceptualized in terms of map scale (i.e., cartographic scale) and its impacts on cartographic representation. During the latter half of the twentieth century and continuing to the present, human geographers have developed multiple ways of thinking about, representing, and analyzing scale. Substantial efforts have been devoted to scale explicit analytical investigations. Among the social sciences, human geography is noted for its attention to spatiality, and its major modes of viewing the world include: (1) integration in place, (2) interdependencies between places, and (3) interdependencies among scales. Geog raphers often strive to link pattern, process, and scale – a perspective that privileges no single scale and instead recognizes that geographtic patterns and generative processes may vary across scales. Ongoing scale related developments within the spatial analytic tradition and debates among critical geographers demonstrate that human geography is deeply concerned with issues of scale.

Meanings of Scale

Despite the avowed importance of scale, semantic confusion frequently abounds with detrimental effects for communication both within and outside the discipline. Geographers have developed a framework of selected scale concepts and terminology to ameliorate this unfortunate situation (**Figure 1**).

1. Cartographic scale (i.e., map scale) refers to the ratio between distance measurements on a map and actual measurements on the ground. It may be stated as a representative fraction (1:24 000), verbally (1 cen timeter = 100 kilometers), or through the use of a graphic scale bar. Large scale maps cover smaller areas and convey more detailed information than small scale maps which cover larger areas with less detail.

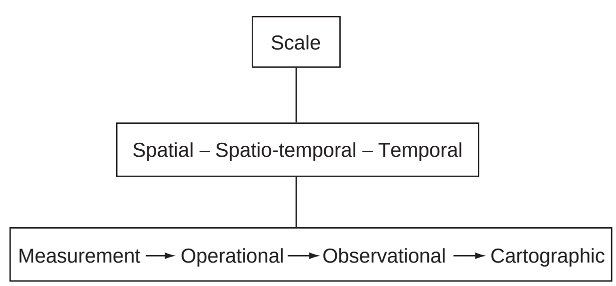

Figure 1 Meanings of scale (after Lam, N., Catts, D., Quattrochi, D., Brown, D. and McMaster, R. (2005). Scale. In McMaster, R. B. & Usery, E. L. (eds.) *A research agenda for geographic information science*, pp 93 128. Boca Raton, FL. CRC Press).

2. Observational scale refers to the spatial extent being investigated; or, the area of coverage. Large scale studies investigate larger regions compared to small scale studies that encompass smaller regions. With this usage, a study of national voting patterns across an entire nation is a large scale study compared to a study of voting patterns across a single state (or district) within a nation state.
3. Measurement scale (i.e., resolution) refers to the smallest observable (or observed) parts of a spatial entity. A synonymous term is granularity (or grain) and the adjectives fine and coarse may be used to communicate differences in measurement scale (**Figure 2**). For example, observations may range from the individual person to households to local neighborhoods to defined cities/town or to states located within a nation state. Analyses relying on individual observations are considered fine scale and those relying on individual observations aggregated to states are considered coarse grain. One may employ a large observational scale (i.e., entire nation state) in concert with a fine measurement scale (i.e., individual person). It is also possible to employ a smaller observation scale (i.e., California) with a coarser measurement scale (i.e., census tracts within California).
4. Operational scale refers to the scale at which a geographical process occurs. For example, spatial mismatch theory posits that American inner city populations have lower access to jobs due to the process of employers moving from inner cities to suburbs. This process exemplifies a metropolitan operational scale. The loss of US manufacturing employment is due to international economic processes operating at a global scale. Alternatively, the process of gentrification is localized to smaller pockets within the inner city which reflects an intracity scale. An important caveat is that multiple observational scales may be relevant to various processes. For example, intracity gentrification may occur via local real estate processes in concert with global cultural and economic processes impacting gentrifying household locational behavior. Geographers are often interested in the scale at which variance is maximized since maximal variation suggests a scale 'where the action is' which may in turn suggest an appropriate operational scale.

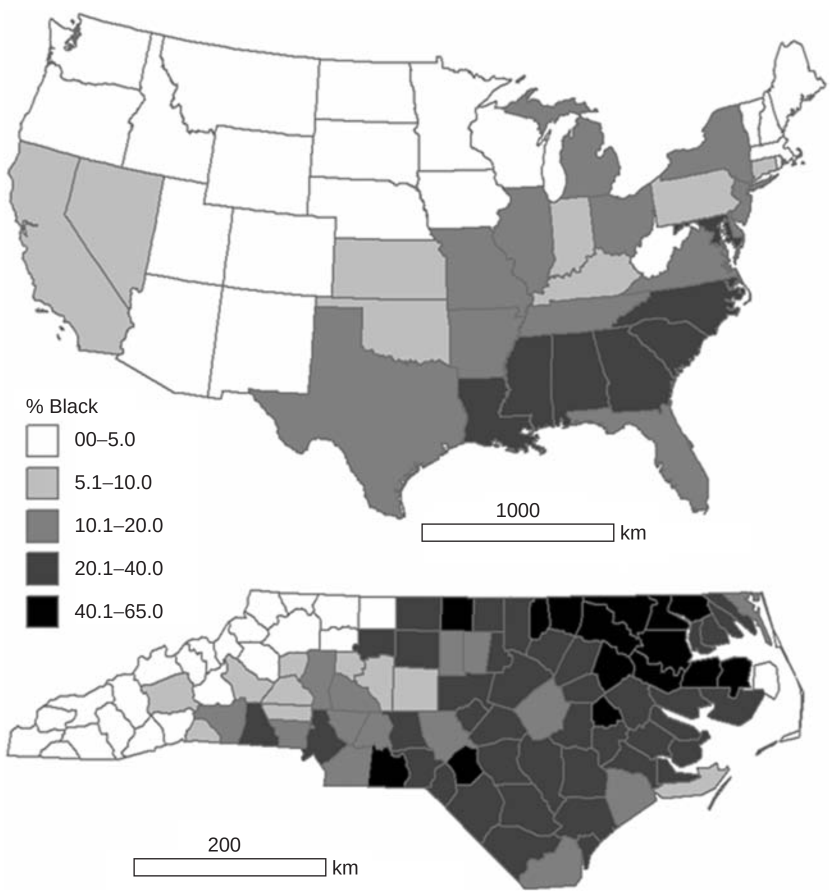

Figure 2 Population maps demonstrating a larger observation scale (United States) with coarser measurement scale (states) relative to a smaller observation scale (state of North Carolina, US) with finer measurement scale (counties in North Carolina).

The four meanings of scale as presented above fall largely within the domain of spatial scale, and spatial scale has attracted much interest among geographers. Similar scale definitions and examples pertaining to both temporal and spatio temporal scales are also employed. Explicit integration of space and time and their associ ated scales has attracted attention from geographers particularly within the Geographic Information Science community.

Another meaning of scale refers to levels of meas urement applied to the entities that we observe and represent. Whereas spatial and temporal scales deal with issues of 'where' and 'when', levels of measurement are scales that deal with the 'what' of entities located in space and time. Nominal level representation uses distinct, nonquantifiable categories such as place names (e.g., London, New York) or land uses (e.g., commercial, residential). Ordinal level representation is also cat egorical, but the categories have a ranked position implying quantifiable differentiation – for example, countries with high, medium, or low poverty rates. Interval and ratio representations are noncategorical, numeric levels of measurement. Interval scales are those lacking an inherent meaning for zero such as temperature (e.g., 0 °C and 0 °F). Ratio scales are numeric but have an inherent zero. For example, a region with zero people really does mean zero people. Additionally, 100 people is twice as many people as 50 people, whereas 100 °C is not twice as hot as 50 °C.

Metaphors of Scale

In recent decades human geographers have questioned how scale is talked about and represented. A dominant discourse centers on a scale continuum ranging from the local to the global. The global scale is implicated in the idea of globalization which denotes the process of increasing interconnectedness of many parts of the world through common processes of economic, environmental, political, and cultural change. Diffusion of market economies, increasing rates of natural resource con sumption, the spread of democratic government, and adoption of Western style lifestyles are hallmarks of globalization that, in theory, may promote a homogen ized global human geography. However, globalization is manifested at distinctly local scales that can act to modify and adapt global diffusion processes to local geographical contexts. Witness a popular US comic book hero (Spiderman) that has been successfully exported and marketed to Indian (Asian) youth in a culturally modified form that bends to local naming conventions and religious understandings. Adaptation of the McDonalds restaurant chain to local consumer preferences is another more prominent example of interactions between global and local scales. This highlights a process of glocaliza tion that can be described as a local consumption of a globalizing world. These examples suggest that neatly ordered meanings of scale (**Figure 1**) are complemented by more complex and nuanced understandings, or metaphors, of scale.

One metaphor of scale is the hierarchical ladder (**Figure 3a**). Here the various scales are viewed as rungs on a ladder with a rigid progression between the various scale levels. The global is conceptualized as above the local and all other scales, and each scale is treated as distinct from every other scale. This metaphor promotes a vertical way of thinking about scale relationships and has affinities with hierarchy theory in ecological sciences. Hierarchy theory suggests that understanding of geo graphical systems requires understanding of the scale of interest (level 0) in addition to at least two other scales – a scale above (level $+1$) and a scale below (level -1). The higher level scale (level $+1$) acts to impose top down external constraints and opportunities and is gen erally greater in spatial extent than level 0 and slower moving (temporally). The lower level scale (level -1) is typically faster moving (temporally) and smaller in spa tial extent than level 0. Level -1 processes provide the mechanisms that explain level 0 behavior. The diffusion and cultural adoption (adaptation) of a US comic book hero within India (level 0) can be viewed as a scalar process involving the interplay of a higher scale process (level $+1$ – global US capitalist and cultural expansion) and lower scale processes (level -1 – burgeoning Indian youth population, increased youth literacy rates, local religio mythic contexts).

Another popular metaphor conceptualizes scale and scalar relationships as a progression of increasing larger concentric circles (**Figure 3b**). Here a flat horizontal perspective is conveyed with distinct scale levels still recognized as separate entities; however, each scale is encompassed relationally within successively higher scales. This suggests analytical possibilities whereby human geographic entities might interact across all scale levels so that actors are not limited to vertical scale relations at the sub (local) or super (national) levels as implied by a vertical perspective. Instead, scalar rela tionships may shift and occur among scales. For example, regional political action groups concerned with public lands management policy may simultaneously employ strategies that shift scales (i.e., jumping scale) by appealing to local civic associations, lobbying through state or na tional political institutions, or garnering regional, national, or global support through media exposure campaigns. Individuals in such groups, while located geographically in bodily activity space, have the possibility of being located and interacting within and across multiple scales.

A nested hierarchy scale approach is commonly applied in analyses using administrative spatial units

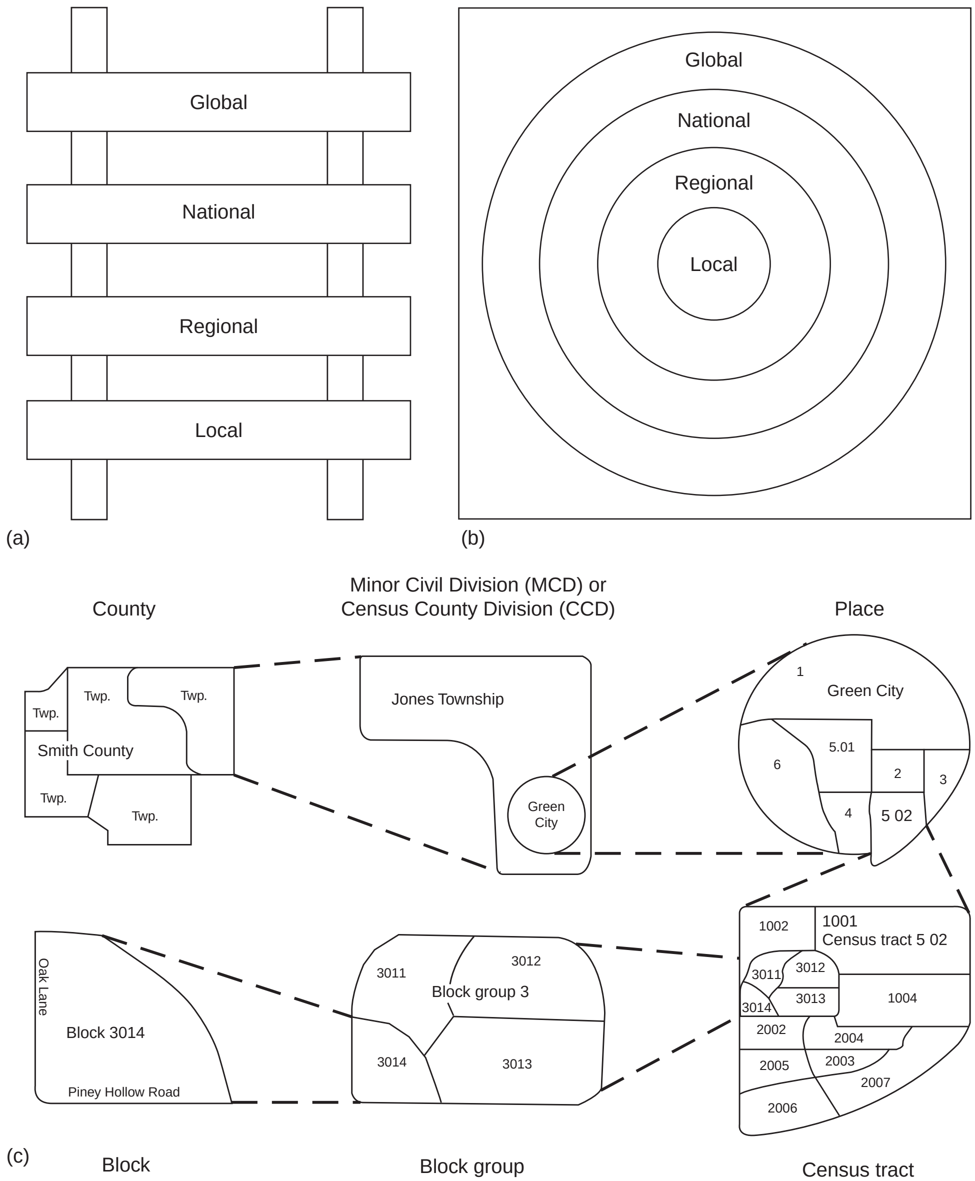

Figure 3 Metaphors of scale: (a) hierarchical ladder (b) concentric rings (a and b after Herod, A. (2003). Scale: The local and the global. In Holloway, S., Rice, S., and Valentine, G. (eds.) *Key concepts in geography*, pp 229 247. London: Sage Publications), and (c) nested hierarchy (source: US Census, http://www.census.gov).

such as census enumeration districts. Taking US census geography as an example (**Figure 3c**), socioeconomic and housing data are aggregated from individual or household level observations and reported at successively higher and nested spatial scales (e.g., blocks, townships, places, counties). Geographers commonly use data aggregated within nested scale hierarchies in statistical analyses (e.g., correlation analysis, multivariate regression) to describe, explain, and predict spatially differentiated human geographic patterns and processes. Critique centers on the fact that, unlike some environmental phenomena such as watersheds which naturally occur within nested scales, politico administrative units of aggregation are not naturally occurring and are imposed by political processes and concerns of expedient data collection and enumeration.

A fourth metaphor is the linear network (**Figure 4**). This approach departs from those that conceive of space as being organized at various levels of bounded two dimensional space (i.e., areas). The networked conception of scale focuses on sites of human agency (i.e., individuals, households, firms, institutions) corresponding to local, regional, national, and global scales that connect to and are intertwined with one or more other sites in a complex fashion forming a space of flows. Cross scale interactions afforded by modern information

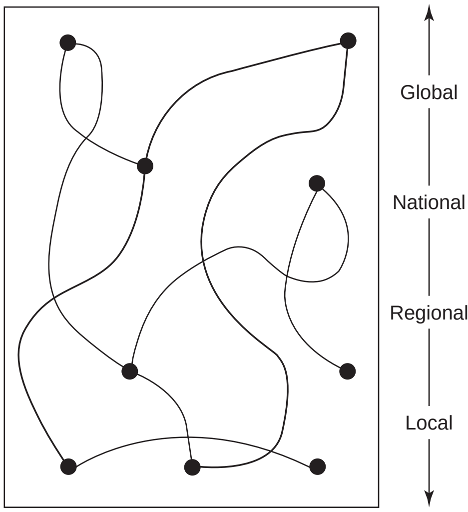

Figure 4 Scale as complex linear networks (after Herod, A. (2003). Scale: The local and the global. In Holloway, S., Rice, S. and Valentine, G. (eds.) *Key concepts in geography*, pp 229 247. London: Sage Publications).

technologies and command and control structures of multinational corporate and financial institutions and their effects on human geographic patterns and processes exemplify this linear based conception of scale and scalar relationships.

Modifiable Areal Unit Problem

Inferential statistical analyses relying on hierarchically nested scales are subject to the modifiable area unit problem (MAUP). This problem is especially relevant for studies that use census or census like data that involve aggregation from lower to higher scale levels. It can be decomposed into two effects. The scale effect refers to the fact that, on average, the larger the unit of aggregation the larger the correlation between two vari ables. The aggregation effect refers to the very large number of ways that entities may be aggregated using different zoning schemes. For example, a city of 100 000 people might be comprised of 500 census enumeration districts of varying areal sizes and shapes. Upon rezoning these districts using different zonal boundaries, aggre gated observations will change in turn yielding different inferential results. Monte Carlo simulation of large numbers of zonal schemes and associated aggregated data observations has shown that it is possible to obtain correlation coefficients spanning the full range of -1.00 to $+1.00$.

Related to MAUP is the problem of ecological fallacy which occurs when inferences based on a higher scale level are used to make interpretations at a lower level. For example, a high correlation at the state level between race and literacy level taken to imply a similar correlation at the individual (person) level is erroneous. Ecological fallacy can be avoided by collecting data and making inferences at appropriate operational scales; however, this often entails primary data collection such as sampled surveys which implies *a priori* knowledge regarding which scale to observe and measure.

Regression and Scale Dependence

The scale effect, while a component of the MAUP, has been used as a mechanism to investigate scale depend ence. Scale dependence is a condition where multivariate relationships and/or process mechanisms differ de pending on levels of aggregation. For example, human–environment studies of land use and cover change have employed multivariate regression models using different social (households, villages, districts) and landscape (varying raster pixel sizes) aggregation levels and found varying model fits and parameter estimates that suggest scale dependent patterns and processes. Such an ap proach may help reveal different scales of pattern–process relationships. Alternatively, it can be critiqued as a manifestation of the MAUP.

Multilevel Modeling

Multilevel modeling is a statistical method that examines the nature of relationships at several spatial scales simultaneously as opposed to conventional regression analysis that assesses relationships at a single scale. It is particularly relevant to geography where national, re gional, and local social and/or environmental configur ations provide contextual effects posited to influence individuals. Modeling spatial behavior at the individual level misses the spatial context in which individual be havior occurs. Modeling behavior at the aggregate level is prone to the ecological fallacy. Multilevel modeling attempts to avoid these problems by combining dis aggregated, microlevel individual behavior with macro level, contextual measures that may explain or predict variations in behavior.

Spatial Dependence and Scale

Spatial dependence refers to the degree of spatial auto correlation between independently measured values observed in geographical space. Global measures of spatial autocorrelation (e.g., Moran's I, Getis G) provide a singular measure of spatial dependence. Local indi cators of spatial association (LISA) perform a similar function but yield multiple location specific measures of spatial dependence allowing researchers to assess nonstationarity; or, how spatial dependence varies throughout space. Spatial dependence prevails in many geographical datasets and exemplifies Tobler's First Law of Geography – that is, things closer together are more similar than things further apart. For example, housing values tend not to be randomly distributed in space. Instead, higher or lower priced homes and neighborhoods tend to be spatially clustered due to multiple sociospatial processes.

Scale is related to spatial dependence due to the fact that spatial autocorrelation is frequently present within certain distance thresholds beyond which observations are statistically independent. Semivariance analysis is a method that assesses scales (distances) within which spatial dependence occurs. Semivariance is defined by:

$$\gamma(h) = \frac{1}{2n}\sum_{i=1}^{n}[z(x_i) - z(x_{i+h})]^2$$

Where z is an observed value at locations x, h is the lag distance between coupled observations at x_i and x_{i+h}, and n is the total number of observations. The sum of squared difference is divided by $2n$ to account for the coupling of observations at each distance lag so that this magnitude is not double counted. A plot of semivariances versus lag distances is known as the semivariogram (**Figure 5**). In practice, semivariances are binned and averaged over defined lag increments for semivariogram plots. The lag distance where semivariance levels off is known as the range and indicates the scale at which observations are no longer spatially dependent (i.e., they become spatially independent).

Semivariance analysis is often employed in physical geography and remote sensing analyses where observations are sampled using regular distance increments. It also serves as a fundamental method in geostatistical applications such as of kriging interpolation. Human geographers can apply semivariance analysis to sampled point locations or centroids of areal units like census enumeration districts, although caution is warranted due to varying sizes and shape of such areal units. Spatial autoregressive and geographically weighted regression are statistical modeling techniques that account for spatial dependence so that parameter estimates are unbiased in the presence of spatial autocorrelation. Spatial weights matrices are calculated to incorporate spatial dependence in both modeling approaches and include a distance threshold as input that can be determined via semivariance analysis to represent scales at which spatial dependence is operative.

Fractals

A fractal is an irregular, self similar geometric object that can be subdivided into parts each of which is a reduced size copy of the whole. Fractals are therefore scale invariant since they appear identical following scale reduction or magnification. Fractal concepts were originally developed during the 1960s by mathematician Benoit Mandelbrot. Canonical examples of pure mathematical fractals include the Koch snowflake, Sierpinksi triangle, and other fractal space filling curves. Fractal analysis is prevalent in physical geography where fractals exist widely in nature (e.g., coastlines, plants, stream morphology). Within human geography, a tradition of fractal research exists primarily in studies of urban morphology and transportation networks beginning in the 1980s.

Geometrically, a straight line and flat plane have Euclidean dimensions of 1 and 2, respectively. Taking a straight line with Euclidean dimension 1 and reducing its

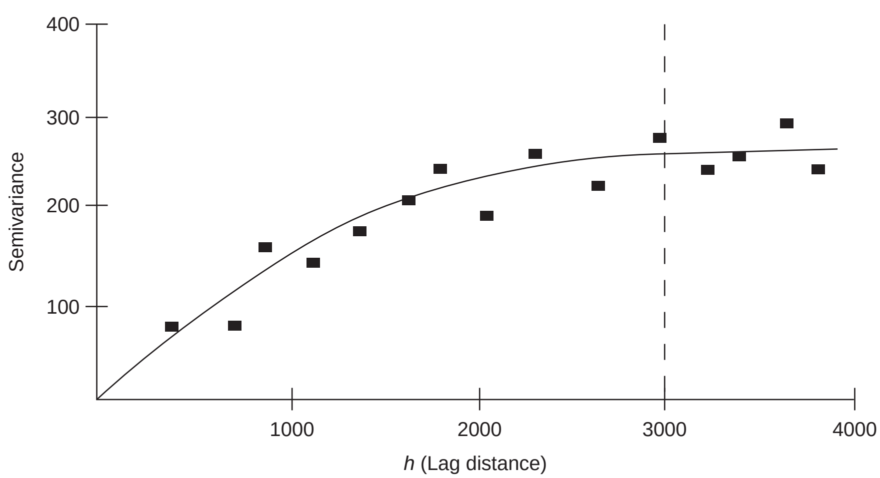

Figure 5 An example of semivariogram plot of lag distance versus semivariance; spatial dependence is operational up to a range of 3000 distance units.

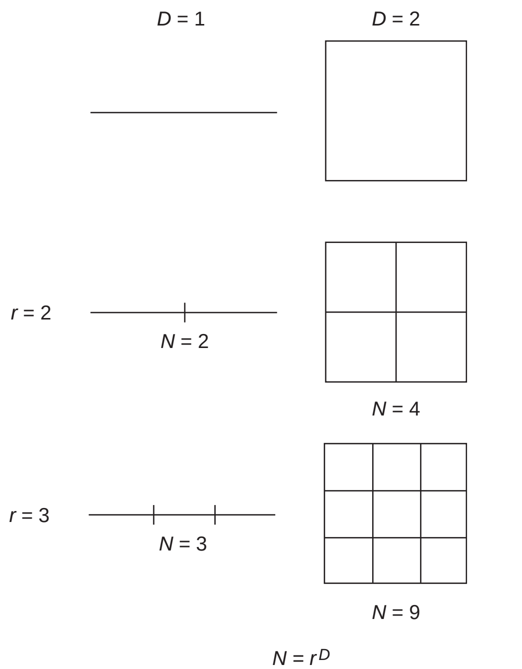

Figure 6 Haufsdorff dimension to characterize Euclidean and fractal dimensions.

linear measurement increment (i.e., the 'yardstick' used to measure) by $1/r$, its length would increase to $N = r^D$ times the original (**Figure 6**). Applying a log transfor mation we get $\log(N) = D\log(r)$ and, with simple algebra, $D = \log(N)/\log(r)$. Taking $r = 2$, we get $D = \log(2)/\log(2) = 1$. Taking a square with Euclidean dimension 2 and reducing its areal measurement increment by $1/r$, its area would increase to $N = r^D$ times the original. Again taking $r = 2$, $D = \log(4)/\log(2) = 2$. This method for cal culating dimension yields what is known as the Hausdorff dimension. Objects with integer values of D have dimensions identical to their Euclidean dimensions and therefore are not fractals as with a straight line or square. Many complex real world geographical objects have noninteger (i.e., fractional) values of D which measures their fractal dimension. Intuitively, complex linear objects such as urban boundaries have a fractal dimension between 1 and 2. The same underlying concept can be extended to three dimensional objects. In practice, empirical studies have used a variety of methods, often involving geographical information science (GIS), to calculate fractal dimension including but not limited to: the caliper method, box counting method, pixel dilation method, and mass radius method.

Through fractal characterization of urban boundaries for many different cities, geographers have identified fractal dimensions ranging from slightly larger than 1.0 to approximately 1.8. Studies that trace the temporal evo lution of urban fractal dimensions suggest that urban morphology becomes more complex (D increases) and space filling as cities progress from industrial to post industrial economies. In selected studies, links between historical politicoeconomic events and trajectories of fractal dimensions have been made. A critique of such approaches is their emphasis on geometric pattern de scription and relative neglect of dynamic sociospatial processes. More recent research deriving from complex systems paradigms (i.e., complexity theory) is attempting to include process mechanisms and explanations using dynamic systems modeling such as cellular automata and agent based modeling.

Power Laws

Power laws describe scalar relationships between two quantities that can be written as $y = ax^k$, where a is a constant and k is the exponent of the power law.

Relationships obeying power laws appear as straight lines on log–log plots since the prior equation becomes $\log(y) = k\ \log(x) + \log(a)$ upon logarithmic transforma tion which is in the same form as the equation for a line. Unlike fractals which deal with self similar geo metrical objects, power laws characterize scale invariance of multivariate relationships or frequency distribution for descriptive attributes of geographic phenomenon. The rank size rule of city size distributions is a well known empirical example of a power law defined as: $P_k = P_1 k^{-b}$, where P_k is the size of the kth largest city, P_1 is the size of the largest city and negative b is the exponent describing the power law. Power law relationships and the rank size rule are identical in principle to Zipf's law – a rank size relationship for word frequency in texts – and the Pareto distribution, a power law distri bution of allocation of wealth. Stochastic growth models involving positive feedback (i.e., 'the rich get richer') and self organized criticality have been posited to explain power law distributions; however process mechanisms continue to be debated and investigated.

With a power law distribution, there is no character istic scale (size) for the phenomena under investigation – it is 'scale free'. This contrasts with the normal, bell shaped distribution for which the mean can be interpreted as the characteristic scale. A resurgence of interest in power laws has occurred during the past decade due to research that has identified scale free social and spatial networks. For example, studies of Web page links have found that some nodes act as highly connected hubs with many links, that most nodes have a small number of links, and that the distribution of numbers of links per Web page follows a power law. For spatial networks, power laws have implications for failure tolerance. If node failures occur randomly, then the network will not lose its connectedness due to the fact that the vast majority of nodes have low degree (i.e., connections) and the likelihood of failure for the highest degree nodes is almost negligible. Alternatively, a failure scenario affecting a few major nodes causes a power law networked system to fall apart and become disconnected. Issues related to scale free networks are relevant to several human geography topics including but not limited to: security studies, transportation and utilities analysis, geographies of the Internet, economic production chains, and social network analysis.

Critiques of Scale

In the early 1980s, critical geographers working largely in economic and political geography began to question mainstream understandings of scale that emanated from geography's spatial analytic tradition. Issues such as economic globalization, supranational institutions, local ized activist political movements, and scale bending media and information technologies did not fit into rigid, areally and hierarchical based conceptions of scale. Arguing that space is socially produced, geographers have critiqued the ontological status of scale. Scale is not given but is produced through processes of social, economic, and political struggle for given historically and technologically contingent contexts. Scaled units of aggregations can be viewed in this light as artificial epistemological constructs that enable methods of data observation, description, and analysis. More extreme critique proposes to eliminate scale altogether as a con cept in human geography. Critical approaches have worked to complicate notions of analytical scale, engaged more visibly with critical social theory, and employed combinations of discursive, qualitative, and quantitative methods.

Nevertheless, analytical scale persists as a vibrant topic of human geography. The various meanings and metaphors of scale, methods of scale characterization and analysis, and critiques of scale developed during the past several decades demonstrate significant intellectual advancement in the treatment of scale and its linkages to human geographic patterns and processes.

See also: Ecological Fallacy; Kriging and Variogram Models; Modifiable Areal Unit Problem; Network Analysis; Spatial Autocorrelation.

Further Reading

Allen, T. F. H. and Starr, T. B. (1982). *Hierarchy theory: Perspective for ecological complexity.* Chicago: University of Chicago Press.

Atkinson, P. M. and Tate, N. J. (2000). Spatial scale problems and geostatistical solutions: A review. *Professional Geographer* 52, 607 623.

Batty, M. (2005). *Cities and complexity.* Cambridge: MIT Press.

Herod, A. (2003). Scale: The local and the global. In Holloway, S., Rice, S. & Valentine, G. (eds.) *Key concepts in geography*, pp 229 247. London: Sage Publications.

Hudson, J. (1992). Scale in space and time. In Abler, R. F., Marcus, M. G. & Olson, J. M. (eds.) *Geography's inner worlds: Pervasive themes in contemporary American geography*, pp 280 297. New Brunswick, NJ: Rutgers University Press.

Jones, K. (1991). *Multi level models for geographical research*. Norwich: Environmental Publications.

Lam, N., Catts, D., Quattrochi, D., Brown, D. and McMaster, R. (2005). Scale. In McMaster, R. B. & Usery, E. L. (eds.) *A research agenda for geographic information science*, pp 93 128. Boca Raton, FL: CRC Press.

Lam, N. and De Cola, L. (1993). *Fractals in geography*. Englewood Cliffs, NJ: Prentice Hall.

Marston, S. A., Jones, J. P., III. and Woodward, K. (2005). Human geography without scale. *Transactions of the Institute of British Geographers* NS 30, 416 432.

Openshaw, S. (1984). *The modifiable a real unit problem*. Norwich: Geobooks.

Peuquet, D. J. (2002). *Representations of space and time*. New York: The Guilford Press.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review* 15, 351 357.

Sheppard, E. and McMaster, B. (eds.) (2004). *Scale & geographic inquiry: Nature, society, and method*. Oxford: Blackwell Publishing.

Tate, N. J. and Atkinson, P. M. (eds.) (2001). *Modelling scale in geographical information science*. New York: Wiley.

Taylor, P. (1982). A materialist framework for political geography. *Transactions of the Institute of British Geographers* NS 7, 15 34.

Relevant Websites

http://classes.yale.edu/fractals/
a comprehensive page on mathematical fractals.

http://www.csiss.org
a premier site for spatial analysis in the social sciences.

http://www.ncgia.ucsb.edu/~jgotts/murray/murray.html
an example of MAUP from the ncgia web site.

http://www.colorado.edu/geography/gcraft/notes/cartocom/cartocom f.html
geographer's craft lecture notes on cartographic communication which covers map scale and levels of measurement.

http://www.globalization101.org/
globalization.

http://www.glocalforum.org/
glocal.

http://www.census.gov
U.S. Census Bureau.

Science and Scientism, Cartography

D. Turnbull, ACSIS, Melbourne University, Melbourne, VIC, Australia

Glossary

Commensurability/Incommensurability Having or not having an established common measure.

Contrapuntal Cartography Incommensurable modes of mapping held in dialogical tension.

Distributed Mapping Dispersed, internet-based, multi-user mapping.

Emergent Mapping The unpredictable forms of creating links and boundaries that come out of distributed contrapuntal mapping.

Knowledge Spaces The spaces within which particular modes of knowledge become possible as a result of linking people, practices, and places.

Ontological/Spatiality/Temporality, Narratives An ontology is a worldview within which certain assumptions are made about what exists, what space and time are.

Performativity The assumption that knowledge is not just representational or propositional, but is also tacit and based in practice. Along with the claims of co-production and social construction; it is the claim that knowledge of the world is shaped by the structures of the society that produces it, and that society is conjointly shaped by the ways we have developed for knowing the world.

Reflexivity The inherent self-referential character of knowing about forms of knowing.

Representationalism The assumption that there is or can be a non-problematic, correspondence between a representation and a reality.

> off the map: out of existence; into (or in) oblivion or an insignificant position; of no account; obsolete; [...] on the map: in an important or prominent position; of some account or importance; in existence; [...] OED

In 1815 William Smith published the first geological map of an entire country. This remarkable map metamorphosed geology into a science through the introduction of history. Using fossils as markers of temporal ordering, it introduced chronological sequencing as the organizing taxonomic principle for what had previously been a chaotically diverse range of geological phenomena. But what is equally notable is that this birthing of geological science through temporal ordering was effected through a map – a spatial display of underlying relationships. There was, in effect, a double moment of scientism and cartography expressed in Smith's map. "It changed the world" as Winchester puts it. A map laid the foundations for a science, but just as importantly maps and mapping became dominant tropes for knowledge and its creation.

It was in the period that Smith carried out his laborious surveys of England (1780–1820) that cartography itself became established as an empirical science and maps became the *sine qua non* for knowledge. The new national states were all busy conducting topographic surveys; while Diderot and d'Alembert described their *Encyclopédie* (1751–65) as 'a kind of world map' and Linneus (1735) called his taxonomic system a *mappa naturae*. But, the term 'scientist' was first coined by William Whewell in 1840, and the first usage of 'cartography' was in 1859 and 'scientism' in 1877. It appears that the issue of scientism in cartography is closely connected to the territorial expansion of science in the Enlightenment project and the emerging professionalization of scientists and cartographers in the mid nineteenth century.

On this account, scientism in cartography would seem a fairly simple matter. Scientism is a mildly pejorative rhetorical term used by those who would argue that there are or ought to be limits to scientific cartography and/or that the discipline of cartography has non scientific components such as esthetics. So debates about scientism in cartography reflect negotiations and controversies over the boundary conditions of professional cartography.

The apparent simplicity of such a claim is just that. A moment's careful reflection shows it is far from simple beneath the surface appearance. If you ask what kind of claim is 'debates about scientism in cartography reflect negotiations and controversies over the boundary conditions of professional cartography', what is the evidence for it, and then ask what are boundary conditions, how are they established, what is cartography, and indeed what is science; it is immediately obvious that there are profound issues concerning the nature of disciplinary practice, epistemology, ontology, the nature of theory and explanation, along with questions about the wider social and historical context of both science and cartography, and the questions of how they are to be understood socially and historically. Indeed these are the very questions that also inspired much of the early work in the sociology of scientific knowledge (SSK) which was concerned to understand what was taken then as a given of modernity – the authority of science.

The question of science and scientism in cartography is problematic; on the one hand because of its reflexive character, and on the other because of the complex relationship between thinking or analysis of any kind, and the kind of space that is assumed in the equation of maps with knowledge. As Doel puts it "One cannot think without *spacing*, nor can one space without *thinking*." The very assertion of there being boundaries is itself a matter of spatial relations and therefore a subject of cartography. If cartography is the science of spatial representation whose boundaries, like those of all disciplines, are in continual historical process driven by intersecting dynamics of a struggle for authority and technological development within the wider complex political and economic developments at the state and global levels, then there can be no simple answers, only controversial, contested, complex, and contrapuntal tensions between alternatives. The intensely reflexive character of the issue is inherent in the reality that the analytic endeavor to reveal such social dynamics is itself a spatial exercise in representing the production, assemblage, transmission, and legitimation of knowledge.

The International Cartographic Association defines cartography as 'the discipline dealing with the conception, production, dissemination, and study of maps'. It goes on to say 'A map is a symbolised image of geographic reality, representing selected features or characteristics resulting from the creative efforts of cartographers, and is designed for use when spatial relationships are of special relevance'.

By contrast Harley and Woodward give a more catholic definition; 'maps are graphic representations that facilitate a spatial understanding of things, concepts, conditions, processes, or events in the human world'.

For Cameron and Edge:

> Scientism is present where people draw on widely held images and notions about the scientific community and its beliefs and practices in order to add weight to arguments they are advancing, or practices they are promoting, or to values and policies they are advocating… The concept of scientism implies an *attitude to science*: those who use scientistic language acknowledge and respect *the authority of the scientific community*, and wish to capitalise on that authority, in order to make their discourse more persuasive. In doing so they reinforce and consolidate that authority.

While Thomas Gieryn points out that the authority and credibility of science is a cartographic accomplishment, it results from boundary work separating out the legitimate and authorative from the illegitimate and incredible.

The ICA definition of maps as images of 'geographic reality' is arguably an instance of scientism in action, and equally there is boundary work going on. Harley and Woodward's definition of maps is much more inclusive. However, both definitions are also exclusive. They distinguish cartography from closely related practices such as surveying, geodesy, communication theory, GIS/Global Positioning System (GPS), and information and communication technology (ICT). They distinguish cartography from the wide range of disciplines and sociotechnical infrastructures that make cartography possible. But it is also possible to define cartography to be maximally inclusive: as the study of the ways in which beings create, perform, and represent spatial understandings of the world including their presence in it. This would expand their representationalist definitions to include the performative; it would undo the divide between mapping and mapmaking.

In addition to these divides over definition, location, and orientation, some of the most contested are over unity and plurality; is cartography, and for that matter is science, unified or disunified? Is cartography a science or an art? Where should the boundary be drawn between Western, scientific maps, and indigenous maps? While others are more concerned with the outer limits of cartography, is it to be immured against power and ideology? Do maps colonize? Is it just a technology or a tool? Is it a primary mode of knowledge assemblage and discovery with ever expanding boundaries? All of these debates have their parallels in science; unsurprisingly, since science and cartography are intertwined, and the trickiest forms of boundary work are enacted at their intersections in the constant juggling of the benefits of identity and differentiation. This article canvasses map the variety of issues wrapped up in determining the boundary conditions of cartography. What are the spatial dimensions of spatial science? What are the most salient features in the landscape of cartographic scientism, how can they be ordered, classified, related, and mapped, according to what kind of science or analytic discipline?

The major dimensions are epistemological, ontological, historiographical, technological, and sociological, but they bear no essential relationship to one another nor do they have a natural hierarchy. How they relate and interact depends on the interplay of the divisions and boundaries they are used to articulate or oblate. Clearly these are not easy questions; they are reflexively inseparable, as is the conjoint reality that cartography is also inseparable from its history and its disciplinary self conception.

Cartography can be purified, isolated, distilled into its essences and held to be nothing other than the technique of mathematical display of spatial relations. As does for example Cordell Yee, historian of Chinese cartography, who recently rose to the defense of scientism in cartography. He argues that the inclusion of the symbolic, the linguistic, the cultural, and the political, in the history of cartography is unwarranted, resulting from a mistaken reaction against scientism, and that it represents a rejection of the Enlightenment project. He insists there is 'a universal method valid across cultures and for multiple points of view' and that 'universal knowledge is possible through mathematics and the object of this knowledge is the nature of the physical world'. His ideal type is the modern scientific map on which things are measurable because 'the measurable or quantifiable is privileged'. [His] credo is 'What held for Descartes still holds: the most certain things are those that can be measured. Modern cartography has expressed a Cartesian urge in focusing on the most certain'. Counter to Harley, Edney, Wood, Ingold, and others, there is no 'cartographic illusion' or fallacy of 'cartographic mimesis':

> The goal for the physical sciences is a transparent representation of the way things are. With maps part of the transparency rests on a one to one correspondence between represented and representation. Maps are spaces representing spaces. Locations on maps represent locations in space. Thus there is some ground for the authority granted to maps based on measurement.

For Yee the issues are clear, unproblematic, and ultimately a question of cartography's authority. This resurgent intervention in defense of scientism is not, however, a quaint piece of nineteenth century antiquarianism, nor is it just a return to the simplistic history of fifty years ago when it was possible for Crone to claim 'the history of cartography is largely that of the increase in the accuracy with which the elements of distance and direction are determined and in the comprehensiveness of map content'. Yee is explicitly responding to the History of Cartography project initiated by Brian Harley and David Woodward which has had a profound impact on cartography and its self understanding, which in adopting an inclusionary definition of maps mentioned earlier, has resulted in an entire volume being devoted to that which was previously excluded – indigenous cartography. That inclusion changed the ontological boundaries of cartography. Where once indigenous knowledge and cartography had been excluded, because their narratives of spatiality and temporality were place based and history free, now their inclusion implies there are forms of space and time other than those assumed by Western scientific rationality. This undoing of one of 'the great divides' as Latour called them, between Western science and indigenous knowledge is itself packed within the making and unmaking of another set of divides or boundaries between unity and plurality, science and cartography.

This complex co presence of knowledge and mapping, science and cartography, is apparent in that ways in which the defining representational process of the modern age has often been said to be the conquest of the world as map. While the cartographic process of knowledge assemblage provides the dominant metaphor for science as a unified body of knowledge. As the postcolonial critic Sumathi Ramaswamy puts it, it is 'the staging of the world on a map before one's eyes as an enframed whole that can be ordered, secured, rendered knowable, and ultimately masterable, [which] holds the key… to the European domination of the globe that is such a diagnostic feature of the modern age'. But, for Ramaswamy, the hegemony of the map is not complete; cartographic resistance is possible through the mapping of difference. Whereas, for Walter Ong maps are markers of a difference between planned exploration, and haphazard traveling, that irredeemably separates scientific from indigenous knowledge traditions.

> Only after print and the extensive experience with maps that print implemented would human beings, when they thought about the cosmos or universe or 'world', think primarily of something laid out before their eyes, as in a modern printed atlas, a vast surface or assemblage of surfaces… ready to be 'explored'. The ancient oral world knew few 'explorers', though it did know many itinerants, travellers, voyagers, adventurers and pilgrims.

The portrayal of the history of cartography as one of unified and unifying scientific progress is a central scientistic trope whose articulation has been partly undone by the History of cartography project and by Edney's analysis. The image of cartography as a continuously expanding, spatial database built on the assumption of scale independent, commensurable, geographical facts that can be geolocated to any required degree of accuracy has to be set against Edney's revised history. His key point is that cartography is not just one single practice, it has multiple modes with differing histories and practices, consequently there are different types of cartographic information, including thematic, systematic, chorographic, charting, and topographic. Each cartographic mode has different social, cultural, and technological relations, and differing spatialities. Each mode is socially constructed in the same manner as art, science, technology… as such, the boundary between science, technology, art, cartography, etc., "is a matter for social negotiation and represents no underlying distinction".

However, there is a parallel making and unmaking of the same divide between unity and disunity in science; between claims there is a growing body of coherent and consistent facts, and claims there are multiple disciplines and approaches each with their own styles, methods, and ontologies.

A key process in these territorial dynamics, in both cartography and science and in their interactions, is the achievement of standardization and commensurability. In a world of difference, multiplicity, and messiness, how are stability, unity, and coherence accomplished, how is the world made one, and how is knowledge made assemblable? One way of understanding this process, promulgated by Yee, Crone, and others that has already been mentioned, is that it is achieved through increasingly sophisticated mathematical techniques and ways of measuring, now including satellites and GPS. By contrast other historians have argued that considerable social and technical work has had to go into creating knowledge spaces by linking people, practices, and places, and that standardization and commensurability are products of social and historical as well as technical and scientific processes which have themselves been extremely varied both in time and place.

Science, cartography, and the nation state were effectively co produced in the seventeenth and eighteenth centuries in Western Europe, through the myriad of interlinking processes required to build national observatories, determine the distance of the Sun and the shape of the Earth, establish national surveys and link them, agree on an international date line, and a common system of measurement. The apparent seamlessness of the socially constructed knowledge space that was co produced with science, cartography, and the modern state has now been revealed as even more fragmented and diverse, as historians of cartography have dug deeper and have also moved their focus to countries outside Western Europe.

In Mexico for example, the state, despite strenuous efforts, has not succeeded in a fundamental cartographic task – 'that of regulating property and controlling the space of the state' through a comprehensive property register. The failure was due in part, on Craib's account, to the complex historical transformations of *ejidos* – community owned lands, which has led to an inability to achieve a correspondence between the map and geographic reality.

Similarly the work of Winichakul in Thailand, Yonemoto in Japan, and Smail in Marseilles show that space is not an abstract, neutral, featureless void within which actors move, nor does the modern state ineluctably impose its own forms of identity and spatiality upon the citizens.

Rather, as Craib puts it 'people make their own geography as well as their own history', but 'the ambiguities of (and struggles in) history are reconciled and suppressed through spatial order as the open ended yields to the inevitable. The complexity, contingency, messiness, and irony, that 'is' human history… all are flattened and neutralized in the teleological quest for legitimacy, foundational coherence, and the naturalisation of the social world'.

However, the processes of naturalization are also discerned by Corrigan and Sayer for whom the where 'ontological and epistemological premises of a particular and historical form of social order' are the legitimated result of 'defining, mapping, [and] naming reality', of being 'on the map'. But they are never complete, the processes are always in a state of open ended dialogical tension. Mathew Sparke, for example, points out that maps like the National Atlas of Canada have counter hegemonic effects as well as hegemonic ones. In an instance of 'contrapuntal cartography' the land claims of two indigenous tribal groups of Canada, the Gitxsan and the Wet'suwete'en, were successfully mediated by the state's own maps against the state's claims.

This process of both rigidifying and destabilizing the boundaries between the abstractions of the cartographic discipline and the engaged, embodied, performances of mapping, and being mapped reveals scientism as being, in part, a legitimation strategy. It is a rhetorical move in the struggle for authority, which is deployed by all parties to the continuous restructuring of the field as it is reordered both by technology and developments in socio historical research and by the actors. Those who would favor a progressivist and techno mensurational account seek to display science as revealing and displaying a natural order of things while concealing or denying the agency of the actors. Those who would favor a constructivist account are similarly claiming to reveal the 'real' social ordering while concealing their own assumptions about the agency of technology, artifacts and infrastructure, and forms of spatiality. There can be no escaping this reflexive dilemma.

The agency of the users and of the technology reveals a similar open ended dynamic within and without cartography, especially in the past twenty years or so with the advent of the new technologies of GIS, ICT, GPS, remote sensing and satellite imaging, and the Internet, which together have transformed the boundary conditions of cartography, its location, and its orientation. The introduction of GIS in cartography and geography sparked a flurry of debates about its role. These debates had two principal foci; where is it to be located? Is it a tool or is it a science? What are the social implications of its use? Both of these foci raise the question of scientism; where are the boundaries of cartography to be located in the face of the introduction of a new technology? During the 1990s in cartography's equivalent of the 'science wars', opposing sides either tried to resist what they saw as the expansion of technocratic positivism or tried to welcome the advent of the new technology which would give cartography a new scientific legitimacy. A compromise has now been reached in which the boundaries of cartography have been radically expanded to include critical cartography on the one hand and the establishment of GIS as a science on the other. But what has produced an even more radical transformation is the new mobility of cartography. It has become unbounded. Protagonists on both sides of the earlier debates seem agreed that cartography is no longer in the hands of the professionals but in the hands of the users. It is claimed that anyone with a home computer and an Internet connection can utilize 'open source' collaborative tools, mobile mapping applications, and geo tagging to make their own maps. Arguably this is a form of scientism in that the various geo coding technologies have been black boxed in standardized, unexamined, and largely unexaminable packages, and equally the techniques cannot be used without at least a minimum of expertise. While of course a new divide opens up between the computer haves and the have nots.

However, the scientistic efflorescence of unbounded, mobilized, cartography into every and any area the users want to take it, has been accompanied by a less remarked but nonetheless substantial transformation in the nature of knowledge and the ways it is assembled. The ultimate panoptic dream of Enlightenments cartography was the assemblage of all knowledge in one place and time. A dream that was premised on the creation of equiva lences through the geo coding of all data. Knowledge was to be standardized by the deployment of common systems of commensuration; through abstraction from place, context, and practice; and through the attribution of unique locational identifiers. But, at the very moment when totalizing cartography has become a reality rather than a dream, the antithesis of unified universal know ledge is also being realized. The criticisms that were leveled against GIS and the scientization of cartography are now being incorporated in new forms of spatial representation. In earlier debates about scientism in cartography, there was a concern over the boundary between science and esthetics. That boundary has now been largely dissolved; not only does cartography acknowledge the necessity of esthetics in its own prac tices, as became apparent in the controversy over the Peters projection, but cartography has also become an important mode of artistic expression. The main thrust of the social critiques of GIS were leveled at its hom ogenizing effects on knowledge. The processes of com mensuration turned knowledge into information and information into data; the performative was displaced by the representational. The particularities of place and context were eliminated along with differences in prac tices and ontologies. Now the recognition that there are multiple and distributed knowledges based in com munities, practices, and places with differing narratives of spatiality and temporality, has moved from the confines sociology of scientific knowledge and cartographic theory into the wider arenas of business organization, knowledge management, cyberspace, and indigenous and local communities. In these arenas 'distributed' and 'emergent mapping' are now being vigorously pursued. The new unbounded technologically enhanced cartography is no longer insistent on commensuration, it can and does work 'with' incommensurability. Indigenous and local communities can now map their own territories and traditions. But, this does not stop the giddying gyre of cartography's mapping of mapping in cyberspace, nor does it stop the marking of new boundaries and divides in the constant interplay between knowledge, space, science, and authority.

What lies behind the emergence of distributed map ping is a switch in the models of knowledge and space that inform cartography, from physics and information sciences to ecology and the social sciences. The switch is really one to thinking of knowledge as relational, based in communities of practice, and in movement. Nonetheless it is an example of scientism, of the expansion of an area of science into a region where it previously held no sway. This suggests that scientism in cartography is not an inherently bad thing; it is only problematic where the contrapuntal dynamic is foreclosed, where differing narratives of spatiality and temporality are silenced and unperformed. One possible resolution to the reflexive dilemma of the spatiality of knowledge is to work with the fecundity of the tension between differing and incommensurable ontologies at every level from the lived reality of the actors to the analyst's abstractions.

See also: Art and Cartography; Counter-Mapping; Critical Cartography; Indigenous Mapping; Mapping, Non-Western; Performative and Embodied Mapping; Qualitative Spatial Reasoning; Spatial Ontologies.

Further Reading

Cameron, I. and Edge, E. (1979). *Scientific Images and Their Social Uses: An Introduction to the Concept of Scientism*. London: Butterworth.

Corrigan, P. and Sayer, D. (1985). *The Great Arch: English State Formation as Cultural Revolution*. Oxford: Oxford University Press.

Craib, R. (2004). *Cartographic Mexico: A History of State Fixations and Fugitive Landscapes*. Durham, NC: Duke University Press.

Crampton, J. W. and Krygier, J. (2006). An introduction to critical cartography. *ACME: An International E Journal for Critical Geographies* 4, 11 33.

Crone, G. (1953). *Maps and Their Makers*. Folkstone: Dawson.

Doel, M. (1996). A hundred thousand lines of flight: A machinic introduction to the nomad thought and scrumpled space of Gilles Deleuze and Felix Guattari. *Environment and Planning D: Society and Space* 14, 421 439.

Edney, M. H. (1993). Cartography without 'progress': Reinterpreting the nature and historical development of mapmaking. *Cartographica* 30, 54 68.

Edney, M. (1994). Mathematical cosmography and the social ideology of British cartography 1780 1820. *Imago Mundi* 46, 101 116.

Harley, J. B. and Woodward, D. (eds.) (1987). *The History of Cartography: Cartography in Prehistoric, Ancient and Medieval Europe and the Mediterranean*, vol. 1. Chicago, IL: University of Chicago Press.

Olson, D. (1994). *The World on Paper: The Conceptual and Cognitive Implications of Writing and Reading*. Cambridge: Cambridge University Press.

Pickles, J. (ed.) (1995). *Ground Truth: The Social Implications of Geographic Information Systems*. New York: Guilford Press.

Ramaswamy, S. (1999). Catastrophic cartographies: Mapping the lost continent of Lemuria. *Representations* 67, 92 129.

Smail, D. L. (1999). *Imaginary Cartographies: Possession and Identity in Late Medieval Marseille*. Ithaca, NY: Cornell University Press.

Sparke, M. (2005). *In the Space of Theory: Postfoundational Geographies of the Nation State*. Minneapolis, MN: University of Minneapolis Press.

Turnbull, D. (2003). *Masons, Tricksters and Cartographers: Comparative Studies in the Sociology of Scientific and Indigenous Knowledge* (2nd edn.). London: Routledge.

Winichakul, T. (1994). *Siam Mapped: A History of the Geo Body of a Nation*. Honolulu, HI: University of Hawaii Press.

Woodward, D. and Lewis, G. M. (eds.) (1998). *The History of Cartography, Vol 2, Book 3: Cartography in the Traditional African, American, Arctic, Australian, and Pacific Societies.* Chicago, IL: University of Chicago Press.

Wright, D., Goodchild, M. and Proctor, J. (1997). GIS: Tool or science? Demystifying the persistent ambiguity of GIS as 'tool' versus 'science'. *The Annals of the American Association of Geographers* 87, 346 363.

Yee, C. D. K. (2001). A reaction to the reaction against scientism: On the power and limits of the textual analogy for maps. In Woodward, D., Delano Smith, C. & Yee, C. (eds.) *Approaches and Challenges in a Worldwide History of Cartography*, pp 203 222. Barcelona: Institut Cartografic de Catalunya.

Yonemoto, M. (2003). *Mapping Early Modern Japan: Space, Place and Culture in the Tokugawa Period (1603 1868)*. Berkeley, CA: University of California Press.

Relevant Websites

http://www.nacis.org

Scientific Method

R. Flowerdew, University of St Andrews, St Andrews, UK

Is Human Geography a Science?

Human geography for some decades has been unsure as to whether it is an art or a science. Many human geographers would reject this dichotomy, claiming that it is a social science, but the question remains as to whether human geography should try to proceed in the manner of a science or not. During the quantitative revolution, human geographers (and physical geographers) tried to redefine themselves as scientists. Among other things, this involved the adoption of positivist ontology and epis temology – more accurately, Popper's critical rationalism, with its emphasis on falsification. The way science was supposed to work was then, and still is today, described as 'the scientific method', despite recent studies in the sociology of science which show that scientists seldom, if ever, follow the scientific method in practice.

The scientific method is a label given to a set of prescriptions for conducting research, associated with the positivist tradition. These prescriptions are also known as the hypothetico deductive approach. The starting point is a consistent set of statements about the world which constitute a theory. A theory is based on a set of axioms, statements which are assumed to be true, and primitive terms, which are taken as basic entities, in relation to which all other terms must be defined. Other terms may then be introduced by defining them, and further state ments may be made, which together constitute a theory. Statements may be logically deduced from the axioms and definitions – these may be called 'theorems'.

A theory needs to have no practical relevance to the real world. Many mathematical theories are developed by changing the axioms and seeing what happens. For ex ample, Riemannian geometry, in which the angles of a triangle can be more than 180°, was originally thought to have no practical relevance, but subsequently appli cations have been found in Einstein's relativity theory. A theory is only regarded as scientific if it can be applied to the real world. This means that there must be entities in the real world which can be regarded as equivalent to primitives and defined terms in the theory. Then the scientist can investigate whether the axioms and theo rems are true in reality.

This requires the existence of testable hypotheses – in other words, statements which can be evaluated using observational methods. The scientific method is in tolerant of entities which are not clearly defined and of relationships which cannot be observed, which are labeled as 'metaphysics' and rejected. Thus, most religious dogma is rejected, as is the Marxist concept of social class. They are not regarded as acceptable within the scientific method because there is no way to disprove their existence.

The scientific method thus is based on applying a theory which can generate testable hypotheses – in other words, it must be possible to make observations that would lead to rejecting the hypothesis. If observations do lead to a rejection of the hypothesis, the scientist should return to the theory and modify it. The modified theory must then be used to generate a new set of testable hypotheses – the results of testing either according with the theory, in which case the theory can be developed further, or conflicting with the theory, in which case it must be further modified. The scientific method does not however tell us how the theory should be modified or developed, and there is scope for individual scientists either to recognize the unsuccessfully tested hypothesis as an exception to a generally true theory, or to make more fundamental changes to the theory. Kuhn, in his 1962 study of scientific paradigms and how they may be overthrown, discusses the conditions under which each of these alternatives may occur.

The result of testing a hypothesis empirically may be in line with theoretical predictions or it may not. If it is, it is tempting to regard the result as verification of the theory, but this may be a mistake. It may be that we can think of many possible theories that would predict the result obtained; the next step in that case is to find a testable hypothesis whose result will allow a decision among these alternative theories. This point is made by Popper who argued that falsification rather than verifi cation was the key process in scientific work. The most robust scientific theories are those which have survived the most powerful attempts to falsify them.

The Scientific Method in Human Geography

The scientific method became important in human geography as a result of the 1960s quantitative revolution, and those writers most concerned to construct human geography as a science urged their readers to adopt the scientific method (notably textbooks by Abler, Adams and Gould, and by Morrill). The scientific method was described and discussed in some detail by Harvey, though he adopted a much broader definition of scientific method as "the setting up and observing of decent intellectual standards for rational argument.” However, a few years later he had largely abandoned the scientific method in favor of Marxist dialectics. Nevertheless, others continued to advocate the use of the scientific method, notably Amedeo and Golledge. A short introduction to the topic taking Popper’s view on falsification is provided by Marshall.

One of the best known examples of theory in human geography is central place theory. This makes some simple assumptions about the world and derives an elaborate spatial pattern of settlements. Central place theory was first formulated by Christaller and Losch but was developed into a formal theory by Dacey. It is a good example of a theory which is internally consistent but does not fully accord with real world characteristics. In particular, the central places are assumed to be located on an isotropic plane, with transport equally easy in any direction. It is also assumed that central places can adjust themselves to the best set of locations, without any consideration of historical development or settlement growth. There is nowhere in the real world which fits the assumptions of central place theory. Settlement planners in a few places which were being developed for the first time in a flat area have tried to apply central place theory to real world conditions, for example, in reclaimed land in the Zuyder Zee (IJsselmeer) and in the southern Israeli desert. These experiments have not been very successful, partly because of better transport links and greater economies of scale. Changes in these factors cannot lead to adjustments in settlement spacing, and lower order settlements have not grown as the theory suggests that they should.

This is not to say that central place theory is a total waste of time, just that it is unrealistic to expect real settlement patterns to fit its conclusions when they clearly do not reflect its assumptions. It can be argued that one has a better grasp of a complex system if one has constructed a complete model of part of the system, even though one knows that important aspects of the system are ignored, as opposed to a set of unconnected generalizations whose relationships have not been fully thought through. Thus central place theory may give more insight into the size and spacing of settlements than a set of unconnected generalizations about the sites and situations enjoyed by individual settlements. Against this, Harvey in 1989 has argued that theories in human geography tend to cover only the everyday and trivial aspects of reality, and never predict really important new events.

The real world is a complicated place, and theories as human productions cannot hope to cover more than a few aspects of reality. Theories will be based on simplifying assumptions. Thus, it should not be expected that a theory will reflect the real world exactly; it is bound to be a simplified version. What matters is whether the simplifications lead to a better understanding of the aspects of reality they are applied to, or whether the factors omitted are so important that the factors covered by the theory are trivial or inapplicable.

Critique

It can be argued that the scientific method is often (perhaps always) inapplicable in human geography because human geography is not a science, or at least not a physical or natural science, and hence scientific method is not appropriate. Graham discusses the difference between naturalism and antinaturalism in human geography, where antinaturalists argue that social sciences, unlike natural sciences, have to take account of human actions and activities, and hence use methods (e.g., like interviewing) that are not available to natural scientists. Naturalists, in contrast, admit that additional methods are available, but argue that the same process of theory development, hypothesis generation, and empirical testing is just as relevant to the study of people as to the study of rocks or rivers, and can easily incorporate the particular concerns of social behavior and social science methods. Indeed, mainstream psychology and economics appear to use theories and methods that look little different from those used in the natural sciences. However, others would argue that this is a problem not an advantage for those disciplines.

By the mid 1970s, positivism and the scientific method were under severe criticism from several angles. Structuralists rejected the idea that science should be confined to the observable. Realists argued that conjunction of processes was not sufficient to establish causality. Humanists argued that the scientific method could not give an insight into meanings and values, the things which are most important to people.

In 1978, Derek Gregory published a devastating critique of the scientific method in human geography. He goes back to the ideas of Auguste Comte, the founder of positivism, showing how the scientific method is dependent on:

1. direct experience of reality;
2. the principle of replicability – the requirement that any scientist testing the same hypothesis under the same conditions should get the same result;
3. the formal construction of theories with testable hypotheses;
4. technical utilizability, concerned with ‘means’ not ‘ends’, hence nonideological in the sense that results are independent of ethical positions and value judgments; and
5. a relative and unfinished account of scientific knowledge, which grows slowly as more hypotheses are tested successfully.

All of these are open to criticism and have indeed been widely criticized. The principal points of dispute include:

- the claim that experience can be a sound basis for scientific knowledge;
- that science should deal only with observable phe nomena and not with abstract or hypothetical entities;
- that it is possible to distinguish between an a theoretical observation language and a theoretical language;
- that theoretical concepts have a 1:1 correspondence with 'reality' as it is observed;
- that scientific laws are based on constant conjunctions between events in the world; and
- that 'facts' and 'values' can be separated.

Strictly speaking, these points are presented in relation to positivism, but they are applicable to the positivist concept of the scientific method.

Hay in 1985 attempted to defend the applicability of the scientific method. Although aware of the problems inherent in the scientific approach, Hay believes that scientific methods will be retained in geography, in part because 'knowledge of a scientific type is required by society for its purposes of managing social and natural systems'. More specifically, he sees a continuing need for the research hypothesis, the hypothesis test and predic tion as a testing device. However, the idea that one set of observations can conclusively validate or falsify a theory should not be retained. Instead, the accumulation of results will lead to the advance or decline of rival research programs.

So far, 'scientific method' has been used in accordance with positivist and critical rationalist approaches. Ad herents of other epistemologies may wish to maintain the stance that the methods they use are just as scientific (if not more so) as the positivist hypothetico deductive ap proach. Thus, Sayer argues that realist methods "have every right to claim to be theoretical, explanatory and, if the word means anything, 'scientific'." Even Marxist scholars claim their methods to be 'scientific'. There can then be different scientific methods dependent on one's philosophical position. However, the term 'scientific method' is usually reserved for the hypothetico deductive method outlined above.

Most telling, perhaps, as a criticism of the scientific method are observations from social scientific studies of scientists at work. Here, social science research by Latour and others has established that few scientists actually use the scientific method in their work. Scientists' empirical work seldom involves the development of testable hypotheses from theory, and is likely to be more explanatory, looking at informal hypotheses with the aim of investigating empirical phenomena rather than trying to test existing theories. It can also be observed that scientists may persist with theories even after testing has shown that a hypothesis is wrong, or may abandon a theory on the basis of a more attractive one becoming known, rather than on the basis of any tests that the theory has failed.

In this context, Kaplan has developed a distinction between reconstructed logic and 'logic in use'. The latter describes the reasoning actually used by scientists on a day to day basis which is far more intuitive than the scientific method, whereas the former, which does apply the scientific method, is used when reporting the work, often creatively adapting the chaotic process of much scientific research to fit in with the supposed norms of the research process, represented by scientific method.

Overall, it is probably fair to say that the popularization of the scientific method in human geography in the 1960s and 1970s served a useful purpose in getting human geographers to think more systematically (and more sci entifically) about their work. More recently, however, its use has declined because it seems inappropriate for many aspects of human geography research, and because it is no longer seen as central by scientists themselves. Some human geographers, who might be referred to as positivists or critical rationalists, still attempt to construct and test hypotheses, but do so in a more informal way than dictated by the scientific method in its strictest sense.

See also: Central Place Theory; Hypothesis Testing.

Further Reading

Abler, R., Adams, J. S. and Gould, P. R. (1971). *Spatial Organization: The Geographer's View of the World*. Englewood Cliffs, NJ: Prentice Hall.
Amedeo, D. and Golledge, R. G. (1975). *An Introduction to Scientific Reasoning in Geography*. New York: Wiley.
Blaikie, N. (1993). *Approaches to Social Enquiry*. Cambridge: Polity Press.
Graham, E. (1997). Philosophies underlying human geography research. In Flowerdew, R. & Martin, D. (eds.) *Methods in Human Geography*, pp 6 30. London: Longman.
Gregory, D. (1978). *Ideology, Science and Human Geography*. London: Hutchinson.
Harvey, D. (1969). *Explanation in Geography*. London: Arnold.
Harvey, D. (1989). From models to Marx: Notes on the project to 'remodel' contemporary geography. In Macmillan, W. (ed.) *Remodelling Geography*, pp 211 216. Oxford: Blackwell.
Hay, A. M. (1985). Scientific method in geography. In Johnston R. J. (ed.) *The Future of Geography*, pp 129 142. London: Methuen.
Kaplan, A. (1964). *The Conduct of Inquiry*. San Francisco: Chandler.
Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
Latour, B. and Woolgar, S. (1979). *Laboratory Life: The Construction of Scientific Facts*. London: Sage.
Marshall, J. U. (1985). Geography as a scientific enterprise. In Johnston, R. J. (ed.) *The Future of Geography*, pp 113 128. London: Methuen.
Morrill, R. L. (1970). *The Spatial Organization of Society*. Belmont, CA: Wadsworth.
Popper, K. R. (1959). *The Logic of Scientific Discovery*. New York: Basic Books.
Sayer, R. A. (1985). Realism and geography. In Johnston, R. J. (ed.) *The Future of Geography*, pp 159 173. London: Metheun.

Scott, A.

C. Gibson, University of Wollongong, Wollongong, NSW, Australia

Glossary

Capitalist Mode of Production System of production and consumption underpinning capitalist society, in which the methods of production, distribution, and exchange for commodities and services are privately owned, and in which inputs and outputs are available through commercial markets.

Cultural Economy Alternatively understood as the intertwining of human economic and cultural activities and trends; the application of cultural theory and epistemologies to economic subjects and topics; or a new form of capitalist production and consumption overtly commodifying culture (including popular culture, design, and the arts).

Flexible Accumulation A particular configuration of capitalism thought to have risen to prominence since the 1980s, in which labor is increasingly 'flexible', capital more mobile, and production vertically disintegrated. Workers become multiskilled, are less unionized, and are often employed in small teams or firms, or in subcontracting relationships with other firms. Fordist production lines and social contracts between capital, the state, and labor unions are replaced by offshore subcontracting of manufacturing, and regional specializations of tasks such as design, finance, logistics, and assembly.

Social Divisions of Labor Specialization of workers and tasks in the capitalist labor market through mechanisms of social control and power. Differentiated from technical divisions of labor, where workforce specialization is driven by technical necessity.

Spatial Agglomeration Concentrations of human activities in geographical space. Often refers to the clustering of capitalist firms in cities and regions within close proximity to each other.

Vertical Disintegration New production processes for commodities that break up old production line methods and replace them with separate, specialized tasks and use subcontracting relationships. Typically occurs in industries with high levels of risk, fashion cycles, and structural volatility, such as film and music.

Allen J Scott is a leading economic geographer whose ideas have circulated over four decades. Born in Britain in 1938 and educated in geography at both Oxford and Northwestern Universities, Scott became a pivotal thinker on industrial organization and location, and systematic economic change. Although he is most recently known for his work on creative industries and the cultural economy, he has contributed in substantial ways to at least two other definitive phases in economic geography, with advances in knowledge on regional economic development, industrialization, and the spatial and institutional organization of production. In 1999 he was elected Corresponding Fellow of the British Academy and in 2003 was awarded the prestigious Lauréat Prix International de Géographie Vautrin Lud – geography's Nobel Prize equivalent. In 2000 he was named Director of the new Center for Globalization and Policy Research at the University of California, Los Angeles. At the time of writing he is Distinguished Professor at the Department of Geography, University of California, Los Angeles (UCLA).

Throughout economic geography's different phases – and notwithstanding their respective and divergent emphases – Scott has pursued research around a set of core questions: where is economic activity located, and why does it locate there? How does the location of economic activity shape (and in turn how is it shaped by) tendencies in modern capitalism? How does economic activity change the relationships between places and actors, across geographical scales? In answering these questions, Scott has examined a raft of industries – from medical devices, electronics assembly, and aircraft parts to music, fashion, furniture, and film – and has maintained a strong interest in linking research to policy making. Scott's work has impacted on thinking and practice well beyond geography; he is widely read and cited in microeconomics, planning, public policy, media, and cultural studies. As well as his work in English, which has often been translated into other languages, he has also written original research findings regularly in French. In this respect he works across linguistic boundaries that have otherwise unhelpfully segregated the production of global geographical knowledges.

After his PhD at Northwestern, Scott was appointed as Assistant Professor in Regional Science at the University of Pennsylvania (1965–66), and subsequently at University College, London (1969–73) and the University of Toronto (1973–81). Scott was an important contributor to the so called 'quantitative revolution' transforming geography at that time – in Scott's own words

> a period of great intellectual and professional struggle in geography between traditionalists and reformers, with the latter seeking to push geography out of its perceived idiographic torpor and on the basis of quantitative methodologies and formal modelling into a more forthright engagement with theoretical ideas. (Scott, 2004: 481)

Scott published his early ideas about network analysis, central cities, and locational equilibrium at the height of the quantitative turn in economic geography. The emphasis in this early work was on experimental application of mathematical models to understand and predict human spatial activities, seeking to improve on previous models toward universal explanations – especially of the internal organization of cities.

The 1970s saw economic geography transformed as recognition of capitalism's vulnerabilities amplified and as scholars sought different ways to theorize apparent economic crisis. Some geographers adopted a revolutionary stance on prior quantitative approaches, throwing out old models and approaches in favor of a more politically charged perspective. Although Scott's work would be influenced by this so called 'radical' shift, he would graft new thinking to his core questions, rather than discredit previous approaches. Scott continued to work on quantitative modeling but increasingly sought to engage with critical questions about the capitalist system.

In the early 1980s, Scott moved to UCLA, and by then he had become convinced of the efficacy of understanding the spatial pattern of economic activities based on critical conceptions of "the workings of capitalist commodity production, the accumulation process, and the internal evolution of the capitalist firm" (1982: 111), being "firmly situated within an emerging urban problematic that seeks to root the city as a social phenomenon within the structure of the capitalist mode of production at large" (1982: 112). The city became more fully the site and scale of analysis, while labor markets and the firm were brought into focus in this "rectified urban science" (1982: 112). It was in this period that Scott's interest in spatial agglomeration really developed. It also signaled a subtle, but important shift in epistemology from his earlier modeling approaches – with observations and empirical inquiry brought to the foreground. With more confidence Scott began to distinguish between associative and causative trends: land use and land price might appear to be related in some universal manner (e.g., firms decentralizing to gain access to cheap land for single storey factories), but extending this observation in some predictive manner would be doomed to failure without requisite understanding of the 'structure and evolution of the capitalist firm' and its internal dynamics which lead it to make locational decisions.

It was at UCLA that Scott's central geographical insights matured into a long run research argument – that capitalist economies tend to spatial agglomeration; and that regions are the motors of economic activities, rather than merely recipients of them. For Scott, "the structure and dynamics of modern production systems give rise to massive agglomerations of capital and labor forming the basic building blocks of the large metropolis" (2005: xi). The region is accordingly "a nexus of production relationships and associated social infrastructures" (2004: 486). Expanding on this thesis were empirical projects examining specific industries such as printed circuits, garments, and film animation, allowing extension into related questions of the interrelationships between regions, trade flows, competition, and globalization.

Outputs of Scott's long run examination of regional agglomeration have contributed to important debates across several decades, but particularly in the 1980s, when Scott was a key thinker in debates about flexible accumulation, and the links between new industrial regions, forms of production, and social divisions of labor. In *New Industrial Spaces* Scott was one of the first to analyze the rise of the 'Third Italy' complex (so called because it was different from the conventional 'north–south split' in the Italian economy), where industrial growth was accompanied by high rates of new firm formation, decreases in average firm size, and nonmonopolistic, artisanal production in labor intensive, vertically disintegrating industries such as pottery, furniture design, shoes, and clothing. That book also introduced Scott's increasingly important work on high technology regions (in the Ile de France and Silicon Valley). It was in many respects a manifesto – albeit understated – for moving beyond both the arcane elements of neoclassical modeling (which up until that time had focused rather abstractly on individual, rational, maximizing behavior) and the binary tendencies of neo Marxist economic geography, with its focus on labor capital relations in mass production, and *de rigueur* case study locations in the de industralizing cities of Europe and North America. Instead, Scott for the first time systematically documented another trend in the world economy – known as the shift to flexible accumu lation or post Fordism – which was utterly transforming places at that time. New industrial regions grew on the back of nascent industries, and were characterized by clusters of workers in design and information based jobs, low rates of unionization, multiple levels of subcontracting (including indirect exploitation of low wage labor), and extensive use of marketing. Such regions were not merely bit players in a new global production regime: they had become powerhouses, driving new forms of development and shifting the world geography of economic power. Accordingly, in this period Scott was awarded both a Guggenheim Fellowship, and the Honors Award of the Association of American Geographers for his work.

Scott's keen instinct to look beyond mass production industries and traditional centers of manufacturing production subsequently led him to be among the first economic geographers to take seriously new cultural industries such as music and film, and examine growing sunbelt cities such as Los Angeles. In Scott's words,

> so much of the literature on urban geography generally was posited on the model of the northeastern US metropolis… [this] prompted us to attempt to formulate an alternative view based on the very different forms of industrialization and urbanization that were becoming increasingly apparent in Southern California and the rest of the sunbelt. I myself had the extraordinary experi ence… of being informed by an anonymous referee of a research proposal on the manufacturing economy of Los Angeles that there was no economic activity of this type of any interest in Southern California, and that if you wanted to study urban manufacturing you had better focus on cities like Chicago, Detroit or Pittsburgh. (Scott, 1999: 34)

In concentrating on post Fordist style industries (such as film, furniture, electronics assembly, medical devices, clothing, and music) in Los Angeles in the 1980s and 1990s, Scott's work would prove cutting edge. It documented the shift to image and symbolic based production. Because of this, Scott was often considered part of the 'LA School' of geography, meaning a group of researchers based in Los Angeles who proposed that this city was the paradigmatic postmodern metropolis, although his commitment to systematic examination of the dynamics of capitalism located Scott firmly in a political–economic, rather than postmodern tradition. The extent to which Los Angeles and flexible accumulation were paradigmatic concepts for explaining a new global economic geography became the subject of much debate at this time.

As well as respond to these debates, Scott's interests with cultural products industries such as music and film led to theorization on another ground breaking theme: the cultural economy. Scott emphasized the dual notions of the 'culturalization of the economic' (where economic activity is increasingly governed by cultural logics) and 'economization of culture' (where cultural life is in creasingly commodified). In the process – and perhaps more than any other single author – Scott firmly pos itioned geography as a key disciplinary site of knowledge production on the cultural and new media and com munications industries. He is owed much credit for the current popularity of themes such as globalization, clustering, interfirm power relations, and place sensitive empirical work in subsequent cultural and policy analysis. Notwithstanding this, Scott would remain cautious about the cultural turn in economic geography, critical of its "peculiar obsession with evacuating the economic con tent from economic geography" (2004: 480). But at the same time, Scott has recently acknowledged the manner in which social and cultural geography can provide analysis sensitive to the local conditions often under pinning inter regional variations in economic activity (2004). Scott was led to conclude that the discipline demands "methodological variety and openness" (2004: 494) including a respect for much maligned quantitative methods. Moreover

> a particular point of focus… is provided by the con tinuing commitment by significant numbers of economic geographers to *critical* analysis and to the search for progressive social change. The pursuit of some sort of social democratic agenda and the fight against global neoliberalism, it seems to me, must stand high in any set of priorities in this regard at the present time, and all the more so as our world remains an arena in which tremendous variations in living standards, economic opportunity, and possibilities for cultural self realization persist tenaciously from place to place and country to country. (Scott, 2004: 494)

With this in mind, and in addition to working on old staples (e.g., Scott's 2006 book, *On Hollywood*, which won the AAG's Meridian Book prize), Scott has increasingly linked his ideas about regions as motors of the economy to theories of development and underdevelopment, well beyond the West.

Perhaps Scott's most pervasive (if underacknowledged) influence has been to shape a certain approach to the writing of economic knowledges. His work is well the orized, but not a slave to theory; it builds on observation, but moves beyond mere description through elegant synthesis of ideas and interpretations. Scott's writing emphasizes the situatedness of economic actors and activities in places, and draws on quantifiable evidence, and is also informed by ethnographic accounts of the internal workings of industries. Above all, Scott is keenly committed to the value of rigorous empirical work con ducted at the local–regional scale and in response to real problems and patterns of change.

See also: Cultural Economy; Edge Cities; Industrial Location; Planning, Urban; Political Economy, Geographical.

Further Reading

Amin, A. and Robins, K. (1990). The re emergence of regional economies? The mythical geography of flexible accumulation. *Environment and Planning D* 8, 7 34.

Fan, C. C. and Scott, A. J. (2003). Industrial agglomeration and development: A survey of spatial economic issues in East Asia and a statistical analysis of Chinese regions. *Economic Geography* 79, 295 319.

Gibson, C. and Kong, L. (2005). Cultural economy: A critical review. *Progress in Human Geography* 29, 541 561.

Hudson, R. (1989). Labour market changes and new forms of work in old industrial regions: Maybe flexibility for some but not flexible accumulation. *Environment and Planning D: Society and Space* 7, 5 30.

Power, D. and Scott, A. J. (eds.) (2005). *Cultural industries and the production of culture*. London: Routledge.

Scott, A. J. (1971). *Combinatorial programming, spatial analysis and planning*. London: Methuen.

Scott, A. J. (1976). Land use and commodity production. *Regional Science and Urban Economics* 6, 147 160.

Scott, A. J. (1982). Locational patterns and dynamics of industrial activity in the modern metropolis. *Urban Studies* 19, 111 141.

Scott, A. J. (1983). Industrial organization and the logic of intra metropolitan location 2: A case study of the printed circuits industry in the greater Los Angeles region. *Economic Geography* 59, 343 367.

Scott, A. J. (1984). Territorial reproduction and transformation in a local labor market: The animated film workers of Los Angeles. *Environment and Planning D: Society and Space* 2, 277 307.

Scott, A. J. (1984). Industrial organization and the logic of intra metropolitan location 3: A case study of the women's dress industry in the greater Los Angeles region. *Economic Geography* 60, 3 27.

Scott, A. J. (1986). Industrial organisation and location: Division of labor, the firm, and spatial process. *Economic Geography* 62, 215 231.

Scott, A. J. (1988). Flexible production systems and regional development: The rise of new industrial spaces in North America and Western Europe. *International Journal of Urban and Regional Research* 12, 171 186.

Scott, A. J. (1988). *Metropolis: From the division of labor to urban form*. Berkelely: University of California Press.

Scott, A. J. (1988). *New industrial spaces*. London: Pion.

Scott, A. J. (1991). Theorizing post Fordism: Why contingency matters Rejoinder. *International Journal of Urban and Regional Research* 15, 302 302.

Scott, A. J. (1996). The craft, fashion, and cultural products industries of Los Angeles: Competitive dynamics and policy dilemmas in a multisectoral image producing complex. *Annals, Association of American Geographers* 86, 306 323.

Scott, A. J. (1997). The cultural economy of cities. *International Journal of Urban and Regional Research* 21, 323 339.

Scott, A. J. (1999). The cultural economy: Geography and the creative field. *Media, Culture and Society* 21, 807 817.

Scott, A. J. (2000). *The cultural economy of cities*. London: Sage.

Scott, A. J. (2002). Regional push: Towards a geography of development and growth in low and middle income countries. *Third World Quarterly* 23, 137 161.

Scott, A. J. (2004). A perspective of economic geography. *Journal of Economic Geography* 4, 479 499.

Scott, A. J. (2006). The changing global geography of low technology, labor intensive industry: Clothing, footwear, and furniture. *World Development* 34, 1517 1536.

Scott, A. J. (2006). *On Hollywood: The place, the industry*. Princeton: Princeton University Press.

Scott, A. J. and Soja, E. W. (eds.) (1998). *The city: Los Angeles and urban theory at the end of the twentieth century*. Berkeley: University of California Press.

Relevant Websites

http://www.geog.ucla.edu
Allen J Scott, UCLA Department of Geography.

http://www.spa.ucla.edu
Allen J Scott, UCLA School of Public Affairs.

Second Homes

M. Haldrup, Roskilde University, Roskilde, Denmark

Glossary

Authenticity It Relates to a person's effort to be (or become) faithful to his/her own self and the immanent qualities of nature and culture. Most often contrasted with the alleged 'inauthenticity' of modernity. Tourism and leisure research has often emphasized that consumption practices within theses fields to wide extend reflects a quest for authentic experiences compensating for inauthentic social relations and practices of modern work and everyday life.

Deindustrialisation A process in which especially manufacturing jobs are removed or reduced within particular regions and nations. From the 1970es onwards deindustrialisation has been experienced by many regions and nations in the 'First World' especially as a consequence of outsourcing of industrial production to (former) Third World countries.

Space–Time Compression Concept coined by geographer David Harvey capturing how innovations in communication, infrastructure and transport technologies 'compresses' distance by reducing time. The process of space–time compression is analogue to what sociologist A. Giddens has reffered to as 'time–space distanciation'.

Introduction: A Brief Overview

Second homes, second home owners, and renters count for a significant proportion of tourist consumption in many parts of the world, yet the concept is often under theorized in academic geography. This relative absence from academic discourse partly has to do with the in tangibility of what second home tourism actually is. The term 'second home' covers a whole range of very dif ferent forms of ownership, from the literally privately owned 'second' home to holiday houses, cabins, and commercially driven rentals of houses and holiday homes. Furthermore, the term also covers caravans, tents, mobile homes, and so on. The concept of second home tourism is equally vague; it often seems to include very diverse practices that link up with 'nontourist' forms of leisure and mobility. Many second home owners and renters either commute to work or undertake distance work. This makes the role and use of second homes in contemporary society an interesting example of how patterns of mobility change social relations and practices; this is all closely related to technological, social, eco nomic, and cultural transformations. Finally, the apparent lack of a coherent definition of ownership forms and leisure practices means that second homes do not imply the 'usual' characteristics of tourist consumption. Thus, the notion of the 'second home' is an intangible and fuzzy label that needs to be contextually defined when applied.

The Historical Roots of Second Homes

The geography of second homes has historically been subject to change relating to broader social transforma tions. Thus, the intangible and fuzzy features of the notion of the second home can be explained partly by the varied roles and functions of second homes in different historical contexts. Thus, contemporary second home cultures build on the historical trends of the leisure practices of earlier European elite cultures. For example, the patricians of the Roman Empire would often own several estates between which households could move during the year and similar practices were common among European feudal aristocrats. The 'roots' of contemporary second house tourism are however to be found in bour geois culture which developed in Europe during the nineteenth century. In particular, influences from romantic culture were important in forming a leisure culture based on the 'discovery' of Europe's mountainous and coastal areas as leisure spaces. Until then, these areas had not been highly valued, because of their lack of significance for the upper classes of the then primarily rural based economy and society, but under the influence of the work of writers, poets, and painters depicting the grandeur of the forces of nature in coastal, mountainous, and woodland settings, these areas became popular leisure spaces.

Often, the families of the new upper classes would go to such areas for their month long summer holidays in resorts which quickly developed within these new leisure spaces. Here, families with households could stay at beach and mountain hotels. Moreover, there was also a rapid increase in the purchase of 'second homes', especially in the rural hinterlands of the metropolitan areas as they became the favorite settings for the new urban upper classes. Thus, several European cities and towns developed densely built up areas along, for example, coast lines in rural areas located within convenient traveling distance from the urban centers. As part of a general in crease in living standards, the ownership or rental of a second home gradually became popular preconditions for holiday spending also for middle class and working class

Figure 1 The development of second homes in the twentieth century was closely related to the rise of the nuclear family, relative affluence of the middle classes, and the family car. These allowed working- and middle-class families to relocate parts of their lives to the spectacular settings in the countryside as on this picture of 'My second home Quebec, Canada 1965'. Photograph by Peter Ellis. Reproduced from "*My Second Home*, *1965*" *http://www.flickr.com/photos/pcellis/4166763/*, *with permission.*

families, and around the turn of the century we find middle class, second home cultures developing across North America, Europe, and the oversees territories such New Zealand. While there are of course major differences both at regional and national levels of these second home cultures, we find common elements in terms of the same combination of second home ownership (or rental), access to cars, and the ideals of the nuclear family and its leisure demands emerging throughout these cases (see **Figure 1**). Thus, the geography of second homes that we see today on the fringe of urban centers or in amenity rich land scapes in coastal or mountainous areas is closely related to the rise of the postwar Fordist economy in the Western part of the world.

The Changing Geography of Second Homes

The emergence of the contemporary second home landscape in many countries has been brought about by a specific sociotechnological network made up of societal institutions, technological innovations, and everyday routines and practices. Therefore, the character of con temporary second home ownership and use has also gone through significant changes following social and technological transformations. Above all, changes in the infrastructure and transport technology have been significant in two ways. First, new means of trans portation have substantially changed the geography of second homes over time, as the horse carriage was re placed by the train, the train by the family car, and so on. The rise of second home especially during the twentieth century was also closely linked to and indeed reinforced by the social policies of the Fordist welfare state. Regu lations concerning minimum incomes and working hours meant that workers should have at least one yearly holiday, meant that during the twentieth century, it also became possible for working class families to own and maintain second homes or at least save up to rent a 'second home' in the holiday season. In some countries, for example, pre 1989 East Central Europe, and also more broadly, second homes were often passed on by employers or trade union organizations.

First, whereas the pattern of second home ownership and rental have been regionalized to a large extent, this has taken place in various ways within or across national borders, but still in driving distance from the larger metropolitan areas where owners and renters live. The development of new forms of transport, especially that of cheap air traffic, has enabled people to travel longer distances to their second homes. This has meant that a North–South flow (which is also known from other types of contemporary mass tourist travel) in second home ownership and rental has begun to emerge.

Thus, since the late twentieth century, there has been an increase in the number of retirement resorts and purchases of second homes or apartments in the Mediterranean countries by northern Europeans. Countries such as Spain, Bulgaria, Turkey, and Egypt have all experienced a significant rise in second homes owned by Northern Europeans. In the Americas, countries such as Mexico have experienced similar developments with regard to North American second home ownership and retiring migrants. This is a development that primarily has been driven by cheap property prices in these countries compared to, for example, Northern Europe and North America, as well as the increasing availability of cheap air travel. Thus, the geography of second home tourism has to some extent been reshaped, changing from being a largely domestic phenomenon to a regional one in general and a North–South in particular.

Second, and related to this, changes in communication technologies and labor relations have also significantly altered the character of second home ownership and renting. Whereas the second home holiday of the nuclear family of the Fordist welfare state was a retreat from the spaces of work and settlement, new flexible forms of work and innovations in communication technologies, such as mobile technology, laptop, computers, and Internet access to work, have enabled distance work, so that the second home may also be a place of work. Thus, there are some indications that this could lead to a more flexible use of the second home as a home not only reserved for leisure purposes but also simply as a place to relocate work and everyday life routines when convenient.

Both types of transformations suggest that the second home landscape is likely to change as technological changes in transport and communications enable and reinforce further space–time compression of social relations. An understanding of how changing patterns of mobility shape everyday practices and social institutions is therefore essential for explaining changes in the role of second homes in contemporary society. Issues of migration, transnationalism, and flexible uses of the second home are likely to become a central concern for geographic research into second homes in the future.

Tourism, Leisure, and Second Homes

In human geography, the subject of second homes is not a particularly well researched field. The role of second homes has mainly been approached indirectly in relation to, for example, their influence on the environmental degradation of landscapes or regional planning. This has partly to do with the undertheorized status of 'the second home' which again is linked to the fuzzy character of second home tourism.

One central reason for this is that it is very difficult to define precisely what a second home is. Second homes may be purposely built as weekend or vacation homes, such as a hunting lodge or a time share apartment in Southern Europe. In addition to this, it may be difficult to distinguish clearly between primary and secondary homes. In many countries, the ownership of second homes is related to family tradition and place attachment. The 'second home' has often been in the possession of the family for generations and is handed down from one generation to the next without being sold. In some periods it may be used as a primary home, in other periods as a second home, it may be rented out more or less permanently, and it may eventually be sold as a house to become either a primary or a second home.

Furthermore, second home tourism seems in significant ways to deviate from the forms of leisure consumption normally assumed to be linked to tourism. Second home owners and renters seem not to be preoccupied with 'consuming' difference. So, it has been argued that second home tourism in important ways departs from the model of the *flâneur* engaging in gazing at and visually decoding landscapes, places, and people different from his/her own. While the escape from city life and the discovery of spectacular landscapes and people certainly were an important feature in the early development of the geography of second homes; more recent qualitative research seems to indicate that second home tourism seems to deviate from what is normally assumed to be defining characteristics of tourist consumption in significant ways. Theories of tourism have most often emphasized the drive for consuming the difference of other people, places, and cultures as the main characteristic of tourist consumption. Furthermore, it is assumed that tourists tend to seek 'authenticity' rich experiences of, for example, spectacular landscapes, heritage, or local cultures. Cultural studies of second home tourists, however, seem to suggest a much more introvert form of tourism. The authenticity preferred by many second home tourists is apparently not an objective authenticity of other people, places, and cultures, but rather a social or existential authenticity of having the possibility of retreating to a 'home away from home'. The second home is valued because it offers such a space of retreat, a space to dwell rather than a space of exciting experiences, a space where, for example, social relations within the family as well as own capabilities of craft skills and so on can be experienced in a setting where the constraints of normal everyday life are suspended. Thus, the second home affords a context for social experimentation however much leisure activity in relation to second homes is significantly preoccupied with introvert practices rather than consuming difference. It should however be noted that there are, of course, huge variations in this and that it is difficult to generalize the social and cultural character of second home because of the enormous variation covered by the very term. Thus, it is important contextually to examine the complex and flexible use of second homes in contemporary society, and especially the role that new patterns of mobility and migration plays for the

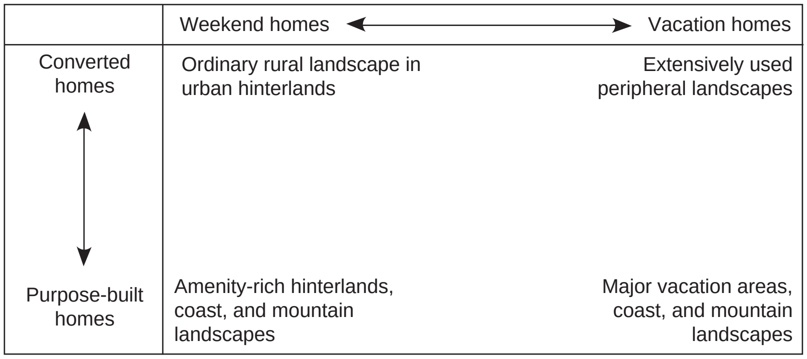

Figure 2 'Second home' is a term that comprises types of particular uses that may vary to a great extent. Thus, weekend homes in the urban hinterlands and purpose-built vacation homes in amenity-rich landscapes may in fact play very different roles. Therefore, it is clearly important to define the notion of 'second home' contextually. Source: Hall, C. M. and Muller, D. (eds.) (2004). *Tourism Mobility and Second Homes. Between Elite Landscape and Common Ground*. Clevedon: Channel View Publications.

geography of second homes. Vacation homes in the wilderness and weekend homes close to the urban centers have historically, and also today, played very different roles with regard to their actual use (see **Figure 2**).

These difficulties in any precise definition of second homes and the role of second homes in relation to other types of tourism have frequently meant that the role of second homes both in relation to the social and cultural aspects of tourism as well as in relation to regional development have tended to be underexamined in human geography, and that issues related to this have been partly neglected both in tourism geography as well as in geographies of regional development.

Second Homes and Regional Development

The role of second homes in relation to development has been a highly debated issue. While the development of second home tourism has been a central issue of much regional planning and strategizing for rural regions, there have been heated debates over whether second homes are a 'curse or blessing' for the development of rural regions. While much second home ownership is (and historically has been) domestic, second homes are also an important component of international tourism in many countries, especially in Central Europe and Scandinavia. In many cases, second homes account for the highest number of tourists' overnight stays. Therefore, second home development has also been central for regional strategies and policies, especially in relation to economically declining and/or rural regions. The exact impacts of second home development on local economies have been debated. This is partly because it is very difficult to clearly identify the sources of impacts on the local economy, and partly because second home tourists tend to spend less than other types of tourists while on vacation.

The main reasons for regional and local planners to encourage second home development in rural regions have most often been to rejuvenate areas after periods of economic decline often associated with crises in the agricultural and fishery sector. Thus the anticipation of increased local spending in shops as well as increased revenues from local taxes has been a central motivation. There are, however, numerous local concerns that have been voiced in relation to this. Of particular importance is the link between second home development and social inequalities, and the transformation of the countryside from productive to leisure space. Thus, there are important economic, social, and cultural potentials for local conflict and disagreement over second home development.

First, we have to distinguish between the various forms of second homes. Second home developments have taken many forms in different countries, and the very notion of the 'second home' covers a very broad spectrum ranging from the primitive shaft or cabin (such as the 'bach' in New Zealand or the mountain cabin in Scandinavia) to planned and regulated second home villages, for example, in marine resorts and to ordinary homes that are now being kept purely (or primarily) in family possession for leisure purposes. In some cases, tents, caravans, and mobile homes are registered as 'second homes'. Following from this very heterogeneous character of second homes, there have often been heated debates on the regulation of second home ownership which has often been recognized by regional planners as unwelcome intruders into the built environment. Huge efforts have therefore been made, especially in the densely populated countries in Northern Europe, to avoid spontaneous second home development, especially in coastal areas.

Second, it is often argued that the development of second homes leads to increased social and economic inequalities. The expectations of second homes' positive impacts on the local economy rest on the central assumption that second home owners do not replace people living in the area designated for second home development. One of the concerns that have often been voiced in debates about local and regional planning is that second home development tends to increase property prices, hence squeezing locals out of the property market. While this has been voiced as a central concern in several cases, it is also oversimplified. Contemporary transformations of the countryside are not primarily a result of the growth of second home ownership, but rather related to more general and global processes. The decline of labor in agriculture and industry faced by many rural regions within 'the First World' is intimately linked to processes of globalization and deindustrialization. As new relations between work and leisure are made possible by more flexible relations in the labor market as well as new transport and communication technologies, the borders of the urban hinterlands move further away, and it becomes possible to commute to the urban core from larger distances. Thus the reason for settling in what is perceived to be 'rural areas' may increasingly be based on lifestyle preferences and not on relations to, for example, local employment. In addition to this, the lack of diversity of housing adequate for all social groups, as well as what is in fact often an abundance of empty houses caused by depopulation following economic decline in such areas is an important aspect of this problem too.

Third, the fact that much second home development happens in areas where the local population seems to be less wealthy than second home owners and renters may cause conflicts. The fear that second home owners should emerge as a class of 'absentee landlords' over the local population is clearly a source of potential tension, for example, in relation to second home investment from the North in the South, such as the widespread constructions of second homes on the Bulgarian Black Sea Coast or along the Mexican Gulf. However, similar concerns have also been raised in relation to second home development in Northern Europe. Here, the discussion has often been related to arguments concerning cultural and/or national difference. This concern has, for example, been persistently voiced in relation to discussions of 'English' second home ownership not only in Wales, but also across national borders this has been an important element. For example, it is not possible for other European Union citizens to own second homes in Denmark, an exception made explicitly to avoid Germans taking over the second home property market along the North Sea coast. While these discussions have often made reference to the impacts of local property markets by 'foreign' ownership and so on, questions of national identity ownership have certainly been important in such discussions too.

Fourth, a central concern has been conflicts over identity and esthetic preferences. Apart from conflicts based on national identity, other issues of potential cultural conflict relate to the more general transformations of the countryside from production landscapes to amenity landscapes. Thus, it has been argued that there is a danger that the countryside is turned into a theatrical spectacle for the growing leisure industries, as second homes and commuters move into such areas. While this may not only be a consequence of second home development, but is also more broadly linked to the changes described earlier, conflicts arising from different sets of lifestyle based esthetic preferences are clearly an unresolved issue in many cases. Thus, different views related to, for example, how to prioritize local amenities valued by second home owners and the interests of local communities in social issues, such as infrastructure, schools, and employment, is clearly an important concern in relation to second home development.

The Future Landscape of Second Homes?

The landscape of second homes is not a static field. It contains huge variations in both historical and contemporary society. Second homes are a vital part of the tourism sector in several countries, with regard to both domestic and international tourism. In addition, second home ownership, rental, and use are also important for the way in which mobility configures contemporary society. This implies that the geography of second homes is likely to change as patterns of mobility change. Above all, the impacts of three tendencies in the transformations of contemporary mobility patterns are interesting in relation to the future landscape of second homes.

First, changes in transport technology and prices have enabled a space–time compression in relation to second home ownership. The emerging North–South orientation of retirement migration and second home ownership may expand the distances between owners and second homes, pointing to complex global relations of second home ownership and use. Second, changing relations on the labor market point both to more flexible labor conditions as well as to more mobile types of work. Mobility for work (and the use of the second homes as temporary habitation located in convenient commuting distance from (temporary) places of work), as well as the possibility of relocating the place of work temporarily to the second home, may lead to the situation where the purchase and ownership of second homes will have increasing importance in relation to the organization of work. Third, migration and transnationalist networks are important features of the future of the second home landscape. Many transnationalist migrants keep (or purchase) ownership of houses and property in origin countries while immigrating. The role of these types of second home ownership both with regard to their impact in countries and regions of origin (e.g., on local property markets and communities) as well as on the (re)production of cultural identities of diaspora communities are interesting and important elements of future relations of second home ownership and use. All these three types of changing global mobility pattern point to a space–time com pression of the second home landscape.

See also: Mobility; Tourism, Rural.

Further Reading

Bærenholdt, J. O., Haldrup, M., Larsen, J. and Urry, J. (2004). *Performing Tourist Places*. Aldershot: Ashgate.

Coppack, J. T. (1977). *Second Homes: Curse or Blessing?* Oxford: Pergamon Press.

Hall, C. M. and Muller, D. (eds.) (2004). *Tourism Mobility and Second Homes. Between Elite Landscape and Common Ground*. Clevedon: Channel View Publications.

Jaakson, R. (1986). Second home domestic tourism. *Annals of Tourism Research* 13, 367 391.

Lofgren, O. (1999). *On Holiday. A History of Vacationing*. Berkeley, CA: University of California Press.

Second World

J. Oldfield, University of Glasgow, Glasgow, UK

Glossary

Marshall Plan The Marshall Plan (1947) was named after George Marshall, US Secretary of State during President Truman's term of office. The Marshall Plan represented a substantial investment of US capital for the countries of Europe following the end of World War II. Its stated purpose was to facilitate economic revival and economic integration within the region and to encourage closer economic ties between Europe and the USA. At the same time, it is clear that the US administration was wary of a spread of communism from the East during the early post-war years. The Soviet response to the plan was understandably negative and their subsequent rejection of assistance ensured that the majority of aid went to the UK, France, Western Germany, and Italy.

NATO (North Atlantic Treaty Organization) NATO was formed in 1949 as a military alliance between the USA, Canada, and ten Western European states. There have been five rounds of enlargement since this time and membership currently stands at 26 countries (2007). The most recent expansion occurred in March 2004 when seven Central and Eastern European states joined the alliance. These included Bulgaria, Estonia, Latvia, Lithuania, Romania, Slovakia, and Slovenia.

Post-World War II

The tripartite division of the globe into 'First', 'Second', and 'Third' world regions has its roots in the Cold War period. The ideological face off between the capitalist West (spearheaded by the USA) and the socialist USSR created the geopolitical context within which the 'Third World' was formed and shaped during the early 1950s as both sides attempted to extend their influence post World War II. While the Third World concept was linked to the emerging discourse of development, the First and Second World categories were propagated by the West and rooted firmly in ideological difference. From the West's perspective, the Second World came to represent those countries characterized by a socialist development model and the pursuit of communism. The two main players were the USSR and the People's Republic of China (PRC). The implied uniformity of the Second World region disguised a marked degree of internal heterogeneity along sociopolitical and economic lines both between the USSR and China, as well as within their respective spheres of influence. At the same time, it should be noted that both Soviet and Chinese conceptualizations of the global geopolitical situation differed from that promulgated by the West.

After the end of Word War II, the USSR was intent upon securing a protective belt along its western border. The USSR's subsequent efforts to bring Central and Eastern Europe (CEE) within its sphere of influence proceeded unevenly across the region and yet by the end of the 1940s a large number of CEE countries were considered to be under Soviet control (i.e., Poland, Romania, Bulgaria, Hungary, Czechoslovakia, German Democratic Republic (GDR), and Albania (until 1961)). In contrast, the communist leadership of Yugoslavia resisted Soviet influence and maintained a degree of independence from 1948 onward. The USSR's hold over what for many in the West became known simply as Eastern Europe, in recognition of the location of the region to the east of the 'Iron Curtain' and the presumed uniformity imposed by Soviet rule, was exercised through various means. Military force was used at times in order to bring countries back in to line, most notoriously in the cases of Hungary (1956) and Czechoslovakia (1968). With respect to economic influence, the USSR initiated the creation of the Council of Mutual Economic Assistance (CMEA) in 1949, which incorporated the satellite states outlined above. This was essentially a trading organization which grew in influence during the mid to late 1960s with the general intention of creating an economically interdependent socialist bloc. Together with the 1955 Warsaw Pact (dealing with security and defense concerns), such initiatives were a direct response to the strengthening of Atlantic alliance between Western Europe and the USA as evidenced by the Marshall Plan for European economic reconstruction and the establishment of the North Atlantic Treaty Organization (NATO). The vision of the CMEA was never achieved due largely to the overwhelming dominance of the USSR and the associated resistance from member countries to its activities. Nevertheless, it did result in a marked geography of trading activity characterized broadly by the USSR supplying CEE with cheap raw materials in return for manufactured goods (see **Figure 1**).

The PRC was formed in 1949. Writing in 1964, the British geographer, John Cole, captured the importance of the recent developments in China by suggesting that

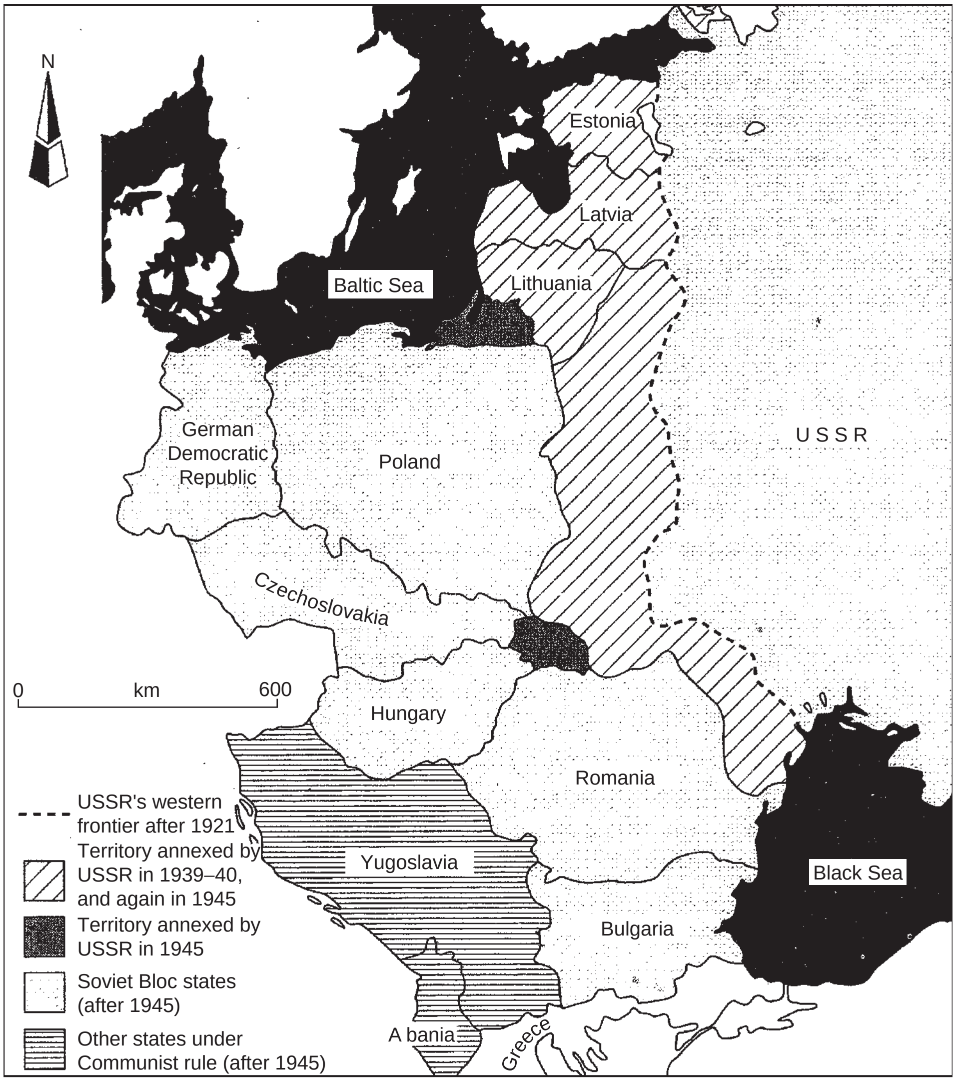

Figure 1 Central and Eastern Europe. Source: Shaw, D. J. B. (1998). 'The chickens of Versailles': The new Central and Eastern Europe. In Graham, B. (ed.) *Modern Europe: Place, Culture and Identity*, p. 123. London: Arnold. Reproduced by permission of Edward Arnold (Publishers) Ltd.

"The emergence…of modern Communist controlled China as a major world power is perhaps one of the most significant events in world affairs in the present century." Superficially, there would seem to have been consider able scope for interaction between the communist re gimes of the USSR and the PRC. However, Sino Soviet relations, which reached their apex in the 1950s before deteriorating, were characterized by a range of problems reducible to differing domestic and foreign policies, as well as ideological differences. Geopolitically, while the USSR looked to exercise control over the afore mentioned states of CEE, as well as making forays into sub Saharan Africa and elsewhere (Cuba also became an ally following Castro's success in 1959), the PRC occu pied an uneasy position as a significant regional actor in the Asia Pacific region.

Beneath the Cold War Facade

The 40 year face off between the USSR and the USA was characterized by alternating periods of heightened tensions and détente. Furthermore, with the Cold War manifest in large parts of the globe, periodic conflicts such as those in the Middle East, Southeast Asia, and the African subcontinent did much to undermine the strength of both superpowers, particularly toward the end of the Cold War period.

Within the context of the Cold War, crude charac terizations and mutual misunderstanding abounded. The backlash against alleged communist sympathizers in the USA during the 1950s driven by Senator Joe McCarthy is illustrative of this. Generalizations and negative stereo types were often seized upon in order to advance the respective geopolitical ambitions of the two power blocs. In a similar vein, Jim Glassman notes the way in which the generalizing work of certain Western geographers during the Cold War period helped to propagate a unified conceptualization of Southeast Asia despite the region's internal heterogeneity. Such an approach was particularly useful in positioning the region in opposition to neighboring communist China.

More broadly, the emergence and consolidation of a more or less homogenous 'Second World' within the geopolitical imagination of the West remained influential throughout the Cold War period. In reality, while there were of course marked similarities between the political, economic, and social structures of the USSR, China, and their satellite regions, these existed together with notable differences in ethnic and cultural characteristics, as well as in the type of socialist model practiced. Mention has already been made of the ideological differences that emerged between China and the USSR with Mao Zedong developing his own interpretation of Marxist thought. With respect to CEE, the propensity for differing mani festations of the socialist development model was given additional impetus by the relatively greater freedoms accorded to the region following the death of Stalin in 1953, although the region was still prone to brutal treatment by Moscow when deviations from the Soviet model were considered too great. At the same time, the general failure of the USSR to keep up with the economic achievements of the West gave credence to the search for alternative versions of the socialist development model and associated experimentation such as the intro duction of Hungary's New Economic Mechanism during the late 1960s.

The differences evident between the various socialist states were at least partially reflected in the academic engagement with the socialist region. For example, British geographers tended to focus on either the USSR or else one or more country in CEE in recognition of the range of languages spoken, the existence of distinct cultural characteristics, and the different manifestations of socialism apparent throughout the region. There were also more prosaic reasons linked to the relative ease of accessing certain countries in contrast to others.

End of the Cold War

The end of the Cold War, heralded by the symbolic destruction of the Berlin Wall (1989), the collapse of socialist regimes across CEE, and the disintegration of the USSR (1991), undermined the perceived unity of the Second World and allowed the noted heterogeneity of the region to be debated and expressed more openly. The region has been subject to a range of renderings since this time as different factions within the West (academic, political, economic) struggle to readjust to, and make sense of, the new reality.

CEE and the Former Soviet Union

The fall of communist regimes across CEE and the former Soviet Union (FSU) was largely unexpected, at least with respect to the speed and finality of the changes that took place between 1989 and 1991. After 40 years of ideological confrontation, it was inevitable that the failure of the USSR's socialist experiment would generate triumphalist rhetoric in the West. Nevertheless, such sentiments were joined by both concern and uncertainty over the geopolitical consequences of socialism's failure in the region. George Demko and Matthew Sagers, writing in the newly renamed journal *Post Soviet Geog raphy* (formerly *Soviet Geography*), captured something of the dynamism of the period with particular reference to geographical engagement with the region:

> Changes in the geographical study of the area once known as the Soviet Union as well as in geography as practiced and taught in the states and regions that were once a part of that country have been occurring at an accelerating pace since 1985, and have been absolutely dizzying in their impact upon the world in recent months. Now, the formal end of the USSR as a state in late December [1991] and the rise of a series of successor states signal an even more dramatic shift in develop ments. After more than 70 years, Communism as a state ideology is largely gone, and with it much of the "Soviet" system, including the centrally planned economy. These are the key factors that our new, post Soviet geography must come to grips with. (1992: 1)

The 15 former republics of the USSR were character ized by considerable variation in terms of economic struc ture, as well as social and ethnic composition. At a general level, at least four more or less distinct groups were dis cernible: a relatively industrialized Slavic core centered on Russia and Ukraine; a Western oriented Baltic region; a predominantly agriculturally based Central Asian per iphery; and the Caucasus region incorporating Azerbaijan, Armenia, and Georgia. This phase of nationalistic uprising gave rise to a number of conflicts most notably in parts of the Caucasus and Central Asia. Furthermore, attempts by the predominantly Muslim Russian republic of Chechnya to break away from the Russian Federation emerged as one of the bloodiest and most disheartening conflicts of the period and remains unresolved.

There was also uncertainty over the changing geo political situation within CEE. The nationalist separatism that characterized much of this region post 1989 resulted in the appearance of several new countries. Notwith standing the emergence of former Soviet republics this period of (re)birth included the 'velvet divorce' of

Czechoslovakia and the disintegration of Yugoslavia. These nationalist assertions were generally welcomed at the time as a legitimate expression of the region's com plex ethnic composition. Nevertheless, as the historical geographer, Michael Heffernan notes, and with particular reference to the Balkan wars during the 1990s, there was also tragedy wrapped up with the shifting geo political situation and the re emergence of the territorial imperative:

> Despite a half century of European debate in which sovereignty and citizenship seemed to be increasingly uncoupled from land and territory, the Yugoslavian crisis demonstrates the remarkable persistence of older forms of geopolitical reasoning both amongst the participants in the war and amongst those who sought to arbitrate. Herein lies the 'sadness of geography'. (1998: 230)

Building Market-Based Democracies in CEE and the FSU

The early to mid 1990s witnessed a remarkable effort by the West to encourage the development of market economies and democratically based governments in the former socialist countries of CEE and the FSU. The experience of international financial institutions in parts of the developing world during the 1980s was used as a basis for devising regional financial and economic reform programs suffused with notions of 'shock therapy' and 'short term pain for long term gain'. Such programs were influential in supporting and propagating the notion of societies in 'transition' from socialism to liberal democracy. The World Bank, the European Bank for Reconstruction and Development (EBRD), and other major financial organizations, as well as political organizations such as the EU have been active in attempting to reduce societal change within the region to a series of planned economic, social, and political development paths. Importantly, these generalized agen das have been criticized for internalizing a relatively simplistic and undifferentiated conceptualization of the post socialist region grounded on linear models of development.

Despite such critical attention, idealized versions of a market economy and democratic forms of govern ment came to represent the yardstick by which the performance of post socialist countries was measured. The initial result of the aforementioned policy recom mendations was indeed traumatic for the countries involved. While CEE generally began to recover, at least with respect to macroeconomic indicators, during the mid 1990s, countries of the FSU (with the partial exception of the Baltic States) continued to suffer marked levels of economic and social difficulty (see **Table 1**). A similar East–West split was also noted with respect to the efficacy of developing systems of political representation with the countries of CEE characterized in general by more robust democratic systems of governance.

Efforts to direct societal change within the post socialist region were accompanied by attempts to (re) locate the countries of CEE/FSU within the context of the global economy. In particular, recent analysis has associated the countries of CEE, the FSU, and China with the concept of 'emerging markets'. The use of this concept gained momentum during the 1990s and is intended to highlight those countries across the globe located outside of the core advanced economies and

Table 1 General socioeconomic data for selected countries

Country	*Life expectancy at birth (years), 1981 90*	*Life expectancy at birth (years), 2002*	*Avg. annual population growth (%), 1990 2003*	*GDP[a], avg. annual % growth, 1990 2003*	*GDP[a] per capita % growth, 2002 2003*	*Carbon dioxide emissions, Millions of tons, 2000*
Bulgaria	71.4	72	– 0.8	– 0.2	4.9	42.3
Czech Republic	71.1	75	– 0.1	1.4	2.9	118.8
Hungary	69.6	72	– 0.2	2.4	0.7	54.2
Poland	70.9	74	0.0	4.7	4.9	301.3
Azerbaijan	69.6	65	1.1	2.4	10.5	29.0
Estonia	70.0	71	– 1.2	1.5	5.3	16.0
Kazakhstan	68.2	62	– 0.7	– 0.6	8.7	121.3
Russia	68.8	66	– 0.3	– 1.8	7.8	1435.1
Ukraine	70.0	68	– 0.5	– 5.3	10.2	342.8
Uzbekistan	68.1	67	1.7	1.2	3.0	118.6
China	68.3	71	1.0	9.5	8.4	2790.5

[a] Gross domestic product.

Sources: World Bank (1996). *From Plan to Market: World Development Report 1996*, p 172. New York: Oxford University Press; World Bank (2004). *A Better Investment Climate for Everyone: World Development Report 2005*, pp 256 257, 260 261. New York: Oxford University Press.

yet conforming more or less to market conventions and considered to possess sound investment potential. The emergence of the concept is indicative of the changing global geopolitical situation. As James Sidaway and Michael Pryke suggest:

> [T]he general notion of places as 'emerging markets' is symptomatic of the diminishing credibility of a passé geopolitical division of the world into 'Third', 'Second' and 'First'. Beyond the Cold War, in a 'New World Order' the old poles of North and South and East and West are less credible. (2000: 189)

More focused reworkings of global economic space have also emerged with Russia and China grouped together with Brazil and India (so called 'BRIC' economies) as countries characterized by both large populations and the potential for marked economic growth in the medium term. While such categories can be useful in challenging old regional geographies, they can also gloss over underlying heterogeneity (see **Figure 2**).

China

In contrast to the dramatic economic and political changes that characterized CEE and the FSU from the late 1980s, China's societal transformation has been far more gradual with its origins in the late 1970s. China remains a communist country and the willingness of its leaders to allow gradual economic change has arguably been important in helping the country to avoid the dramatic changes that characterized the final years of the USSR. Where the countries of CEE and FSU have relied heavily on market blueprints imported from the West in order to negotiate their path away from a centrally planned economy, China's more pragmatic approach has enabled it to record impressive levels of economic development according to macroeconomic indicators during the course of the last three decades. China is emerging rapidly as a key global economic power, which in tandem with its potential for future growth, has increased markedly its presence on the international stage. China has long borders with both Russia and a number of Central Asian states, and political relations between the different countries have become increasingly fluid since the end of the Soviet period. Energy is a key factor underpinning these relations with China interested in acquiring a secure supply of hydrocarbon fuels from the southeastern part of Siberia, as well as the Caspian basin region. In spite of its relative economic success, China's communist regime continues to rule in a dictatorial style and this has drawn criticism from many quarters in the West particularly with respect to human rights issues.

The Enduring East–West Division

While the events of 1989/91 are often taken to symbolize the end of East–West rivalry and division within Europe, they did not prevent new forms of division emerging which, according to Merje Kuus, represent an active process of 'reinscription of otherness'. Indeed, Europe's East–West split has a long history rooted in deep seated religious and ethnic divisions and traceable to the eighteenth century Enlightenment period and beyond. In short, the western parts of Europe are typically posited as both culturally refined and progressive in contrast to a relatively backward and unsophisticated East. Importantly, this East–West division is highly mobile, and the border between the two regions has moved markedly during the last century to reflect the changing geopolitical situation. Since 1989/91, this division has been further challenged by the collapse of socialist regimes and the progressive enlargement of the EU. The borders of the West are typically and uncritically associated with those of the EU.

A related aspect of the reworked geographies of CEE/FSU concerns the pervasive notion of 'a return to Europe' among many countries of the region. This has obvious connections with the noted stereotyping of the East and West whereby certain former Soviet republics such as the Baltic States, as well as former satellite countries perceive considerable political, social, economic, and cultural advantages associated with their realignment with Western Europe. In addition, the purposeful move to the West is less ambiguously related to a collective desire amongst many of these countries to distance themselves politically from the Russian Federation. Evidence of this repositioning is found in the reformulation of maps and other nationalist symbols. At the same time, the chief expressions of this desire to return to Europe are evidenced by the expansion of NATO membership, as well as the ongoing enlargements of the EU. Eight former socialist countries were incorporated into the EU in 2004 (together with Cyprus and Malta). Both Romania and Bulgaria acceded in January 2007 and a further seven countries, including significant parts of the Balkans, as well as Turkey, are considered either 'candidate countries' or else 'potential candidate countries'.

In response to the enlargement process and changing geopolitical situation, the EU has reworked its relationships with adjacent regions. In particular, since 2004, a large number of bordering countries, stretching from North Africa and the Mediterranean Basin in the south to Belarus in the northeast, are now termed 'neighbours' in line with the EU's European Neighbourhood Policy. The stated aim of this policy "is to share the benefits of the EU's 2004 enlargement with neighbouring countries in strengthening stability, security and well being for all

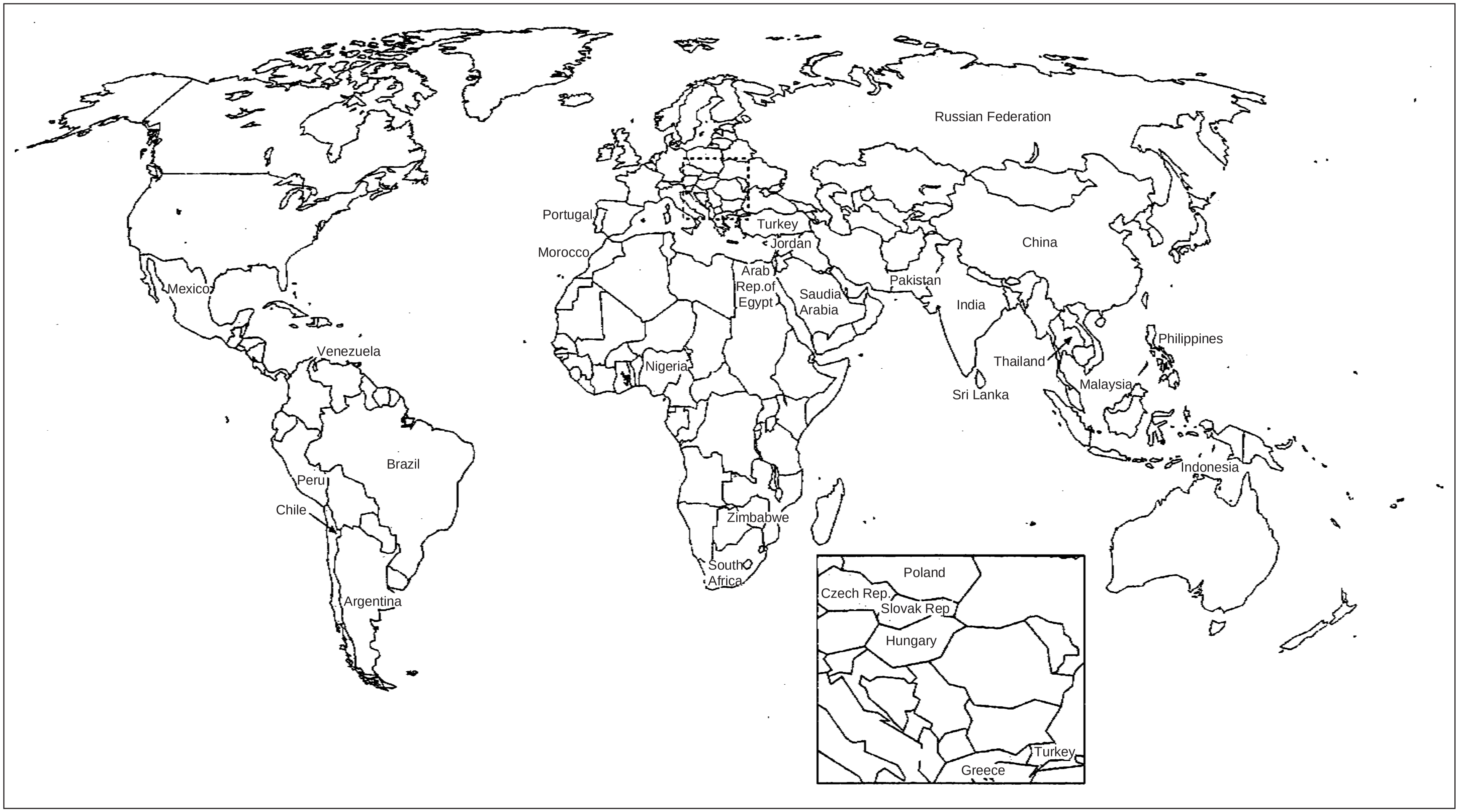

Figure 2 Emerg ng markets 1998. nternat ona F nance Corporat on (FC) emerg ng markets da y ndex n 1998. Source: S daway, J. D. and Pryke, M. (2000). The strange geograph es of emerg ng markets'. *Transactions of the Institute of British Geographers NS* 25, p 192.

concerned." Importantly, Russia has resisted such rhetoric and relations between Russia and the EU have developed along different lines with the establishment of a 'common space' dialogue.

Conclusion

The Second World category discussed here has its origins in Western geopolitical imagination during the aftermath of World War II. As such, it differs from contemporaneous Soviet and Chinese conceptualizations of the global geopolitical situation. The Second World category served a purpose during the Cold War period by drawing attention to the ideological division between the capitalist countries of the West and the communist regimes centered on the USSR and the PRC. At the same time, it allowed for the clear demarcation of those parts of the world which fell outside these two dominant blocs and were deemed 'underdeveloped' in line with the emerging developmental rhetoric of the period, a rhetoric driven by the USA. The efficacy of the Second World category discussed here was seriously challenged by the revolutions of 1989/91 in CEE and the USSR and it has little, if any, contemporary relevance. While the shared experience of socialism remains a common feature uniting former socialist countries, it is important to acknowledge the variety of socialisms that actually existed during the Cold War period. Furthermore, the range of societal changes that have taken place across CEE and the FSU since the late 1980s coupled with the particularities of the Chinese model of development ensures that any generalizations concerning shared characteristics must be treated with care.

See also: First World; Geopolitics; Third World.

Further Reading

Bradshaw, M. J. B. and Stenning, A. (eds.) (2004). *East Central Europe and the Former Soviet Union: The Post Socialist States*. Harlow: Pearson.

Cole, J. (1964). *Geography of World Affairs* (3rd edn.). Harmondsworth: Penguin.

Cooper, J. (2006). Of BRICs and brains: Comparing Russia with China, India, and other populous emerging economies. *Eurasian Geography and Economics* 47, 255 284.

Demko, G. J. and Sagers, M. J. (1992). Post Soviet geography. *Post Soviet Geography* 33, 1 3.

Glassman, J. (2005). On the borders of Southeast Asia: Cold war geography and the construction of the other. *Political Geography* 24, 784 807.

Grasso, J., Corrin, J. and Kort, M. (2004). *Modernization and Revolution in China: From the Opium Wars to World Power* (3rd edn.). Armonk, NY: M.E. Sharpe.

Heffernan, M. (1998). *The Meaning of Europe: Geography and Geopolitics*. London: Arnold.

Hobsbawm, E. (1995). *Age of Extremes: The Short Twentieth Century 1914 1991*. London: Abacus.

Horschelmann, K. (2002). History after the end: Post socialist difference in a (post)modern world. *Transactions of the Institute of British Geographers NS* 27, 52 66.

Kochan, L. and Keep, J. (1997). *The Making of Modern Russia: From Kiev Rus' to the Collapse of the Soviet Union* (3rd edn.). London: Penguin.

Kuus, M. (2004). Europe's eastern expansion and the reinscription of otherness in East Central Europe. *Progress in Human Geography* 28, 472 489.

Nijman, J. (1992). The limits of superpower: The United States and the Soviet Union since World War II. *Annals of the Association of American Geographers* 82, 681 695.

Nolan, P. (1995). *China's Rise, Russia's Fall: Politics, Economics and Planning in the Transition from Stalinism*. London: Macmillan Press.

Nove, A. (1992). *An Economic History of the USSR: 1917 1991* (3rd edn.). London: Penguin.

Shaw, D. J. B. (1998). 'The chickens of Versailles': The new Central and Eastern Europe. In Graham, B. (ed.) *Modern Europe: Place, Culture and Identity*, pp 121 142. London: Arnold.

Sidaway, J. D. and Pryke, M. (2000). The strange geographies of 'emerging markets'. *Transactions of the Institute of British Geographers NS* 25, 187 201.

Smith, A. (2002). Imagining geographies of the 'new Europe': Geo economic power and the new European architecture of integration. *Political Geography* 21, 647 670.

Smith, G. (1999). *The Post Soviet States: Mapping the Politics of Transition*. London: Arnold.

Taylor, P. J. and Flint, C. (2000). *Political Geography: World Economy, Nation State and Locality* (4th edn.). Harlow: Prentice Hall.

Westad, O. A. (ed.) (1998). *Brothers in Arms: The Rise and Fall of the Sino Soviet Alliance, 1945 1963*. Washington, DC: Woodrow Wilson Center Press.

World Bank (1996). *From Plan to Market: World Development Report 1996*, p 172. New York: Oxford University Press.

World Bank (2004). *A Better Investment Climate for Everyone: World Development Report 2005*, pp 256 257, 260 261. New York: Oxford University Press.

Relevant Websites

http://www.nato.int
North Atlantic Treaty Organization (NATO).

Segregation

M. F. Poulsen, Macquarie University, Sydney, NSW, Australia

Glossary

Assimilation The policy of transforming all racial/ethnic groups to the cultural norms of the majority group.

Integration Equal attainment by all racial/ethnic communities in education, the labor market, and the housing market.

Multiculturalism The policy of transforming the cultural norms of society to that of a mix of racial/ethnic cultures.

Segregation The residential separation of people within urban areas on the basis of race/ethnicity and income.

Urban Governance The system by which cities are governed and resources allocated.

Residentially segregated areas within Western cities are primarily ethnic enclaves of unskilled workers, and as such are low socioeconomic areas. By contrast, there has been under globalization, increasingly the development of ethnoburbs, which are ethnic enclaves – usually suburbs – with skilled workers at higher socioeconomic levels. Particularly striking has been the development of segregated residential populations in Western European nations, most of which were previously sources of immigrants, not recipients. The major political issues that face these multicultural societies are the result of the conflict between a focus in public policy on developing and maintaining ethnic communities – multiculturalism – over that favoring spatial and cultural assimilation. While accepting the need for immigrants to support steady population growth essential for economic growth, few governments would support policies favoring cultural assimilation. The expectation is that cultural assimilation will occur across the generations through the educational system. Such issues are today viewed within the context of debates over integration and social cohesion. These are difficult issues and are at variance with the experience of immigration to the New World in the late nineteenth and early twentieth centuries. In that period, the expected process of integration was that of assimilation. That is, individuals from different white ethnic communities gave up their cultural identity and values and became part of their new society. Key to that assimilation was their development of language skills. In an age when manual labor was central to the labor market, both integration and social cohesion were rapidly achieved. Individuals from different European origins blended into the new nation, often within a generation only distinguishable by their surnames.

Residential segregation in geography is a major topic because segregated areas form major geographic features within our cities. It is difficult to view major Western cities without considering the impact and likely outcomes of residential segregation. The role of geographers in their study of residential segregation has also changed. In the information age in which we live the geographer is no longer the impartial observer but increasingly the provider of accurate empirical and qualitative information to public policymakers. To support this endeavor the geographic literature cover major themes such as: how we measure segregation; what are the temporal changes in segregation; the comparative nature of segregation across countries; determinants of segregation; individual life changes; and individual experiences.

Introduction

Within Western cities residential segregation is primarily a feature of large cities with substantial race/ethnic populations, where home ownership forms the dominant form of housing tenure. Residential segregation is defined here as locations where 70% or more of the population are of nonwhite backgrounds: in some cases this population is a visible minority. These measurements are usually conducted on data for small census areas where the total population is less than 500 persons, in that the boundaries of ethnic enclaves and ethnoburbs tend to cut across larger spatial units. Within Western cities residential segregation for the unskilled ethnic population occurs primarily within some of the cheapest dwellings – usually rented property – in the city. This property in some cities is centrally located and in others its location is suburban. In general, residential segregation is seen to be a normal outcome for unskilled immigrants during their initial period of settlement and as such is seen as a positive outcome in that these communities are very supportive of each other. These are places where cultural and religious differences are maintained, while the population over time becomes increasingly integrated into the general labor market through educational attainment.

Residential segregation usually only becomes a major public policy issue when it persists across the generations or if a major incident occurs. Such communities would probably not be problematic except when members of the ethnic community are trapped at the bottom of the socioeconomic ladder by living in those areas and are likely to remain so. As a result, there is a fear of the development of parallel societies and second, concern over residential segregation limiting the life changes of the children within those communities.

Five Periods of Research

Geographers in the 1950s became involved in the study of residential segregation because these areas are major geographical features within cities. Since that initial research this field of study has gone through a number of phases. While it is difficult to be exact about the dating of each phase, in the first period covering the 1950s and 1960s, the aim was to acquire information on the nature of segregation. The focus was on segregation of the African Americans in the United States because that was where the extremes of segregation existed – and continue to exist. The approach was primarily observational. To do so the index of dissimilarity was used as the standard measure. What this index measures is for each individual racial/ethnic group in a city the percentage of that population that would have to move to be distributed in the same manner as the total population. As a result of those measurements, it was established that in many American cities more than 90% of the African American population would have to move. Today, primarily due to the increase in the Latino populations and the resulting competition for residences, and educational attainment that percentage has dropped by up to 10% in many urban areas.

The complexity of the links between racial/ethnic and income residential segregation was initially estab lished in geographic research in the late 1960s and early 1970s (the second so called period of segregation research). The mapping of the three dimensions from social area analyses (socioeconomic status, familism, and ethnicity) established that racial/ethnic segregation only overlapped with low income segregation in a limited number of census areas in the city. Furthermore, outside of America, racial/ethnic residential segregation was often found to be associated with lower middle class areas. These findings suggested a special relationship existed between racial/ethnic residential segregation and income residential segregation. Why this was so and its importance became a major focus of research over the next three decades, with the research focus shifting to linking the research field across the social sciences.

Throughout the 1970s and early 1980s there were the development and adoption of an increasingly diverse array of racial/ethnic segregation measures. In this, the third period of residential segregation research, the most important measure adopted was the index of isolation. Again a single segregation index for the whole of a metropolitan area, the index of isolation measures the probability that a person within a metropolitan area will meet another co ethnic within that city. This pro liferation of measures of segregation during this third period of research by the mid 1980s led to the estab lishment of five dimensions of segregation: evenness, exposure, concentration, centralization, and clustering. The first two are aspatial measures, the last three are spatial measures. The purpose in identifying these five dimensions was to refocus segregation research. Whether there are five dimensions of segregation or just two (aspatial and spatial) is disputed. In general, the aspatial measures are considered the primary measures of resi dential segregation with universal applicability, while the spatial measures are contextual measures. The level of segregation, measured by evenness and exposure indices, are seen to be explained by three generic variables: (1) the size of the city, (2) the proportion that the minority population constituted of the metropolitan population, and (3) the national context.

By the early 1990s the relationship between income segregation and racial/ethnic segregation, the double aspects of segregation, had been accepted. In this, the fourth period of research, geographers studied a wide variety of topics. In particular, the study of the dynamics of the process of segregation; the relationship between racial/ethnic residential segregation and con centrated poverty; the roles of religion, educational attainment, access to the labor market, access to the housing market, geographic context, geographic scale; and the comparative nature of segregation between cities in different countries. Geographers have also focused on developing spatial measures of segregation based on area and shape measurements, although this work has had limited impact because of the unique geography of each city.

More recently the focus has changed yet again. Since 2000, in what we might call the fifth period, at least five themes have emerged. The first theme is the role that urban governance plays in residential segregation. One issue here has been whether different systems of urban governance promote residential segregation.

The second theme is the continued development of research into the concept of residential segregation, in particular, research into delimiting the individual com ponents of segregation. Increasingly today, geographers are called upon to provide sound empirical evidence to policymakers so that policies can be developed and evaluated. Many of these policies are geographically specific in that they focus on assistance to communities at the local area scale. Invariably, this involves the devel opment and use of more meaningful measures of segre gation than the single general summary indices for a city developed since the 1950s and the use of qualitative methods. Hence the third theme is the development of better information on residential segregation. The fourth theme is the development and use of better data sets. In particular, the use of longitudinal surveys that record the histories of individual household members over subsequent waves of interviews. The fifth theme is the development of qualitative research aimed at capturing the experiences and attitudes of individuals.

The Role of Urban Governance

Turning to some of these recent themes and issues, how cities are structured in terms of urban governance is seen to be fundamental to understanding residential segre gation. In America, the role of local government by way of the establishment of a multitude of incorporated places within a metropolitan area is fundamental to the continued residential segregation of the African American population. Relying on the ownership housing market to develop income segregation, the white population, by incorporating places of privilege, by de veloping exclusionary zoning, by basing the educational districts on those places as well as the tax base, has en sured that racial/ethnic segregation persists. In the poor inner cities where the African American population live, and increasingly in some cities where large migrant Latino populations reside, educational attainment levels are well below those in the white privileged suburban incorporated places. As a result, access to the labor market for many in the ethnic groups is restricted, and, in turn, as is access to the housing market. It is the reliance on the ownership housing market, as the basis of individual capitalism, the structures of urban govern ance and the electoral system that make it impossible for many individuals from racial/ethnic communities in American cities to have the same life chances, in particular, the African American population and the Latino poor. Without the replacement of the in corporated places by a single metropolitan government for each city, racial/ethnic residential segregation is not going to be overcome. White privilege will continue.

The alternative to the ownership housing market is rental social housing for all incomes. Where social housing exists as the dominant market for all income groups – as it does in a number of Western European nations – low levels of income segregation and racial/ ethnic segregation occur. Unfortunately, neoliberalism under globalization has changed many of these housing markets into ownership markets. This has occurred at the same time as increased levels of immigration into these countries. As a result, increasing levels of both income segregation and racial/ethnic segregation have de veloped. In between these two extremes are the nations where social housing applies only to a residual of the population: the poor. Invariably, in the European context, this means housing on estates, where a significant proportion of that poor population are members of minority ethnic groups we have double segregation. The result is essentially the same as the American experience – for the African American and some Latino populations – with poor levels of educational attainment, resultantly poor access to the labor market, and in turn poor access to the housing market.

Ghettos, Enclaves, and Ethnoburbs

During the last decade there have often been debates over what constitutes a ghetto and if ghettos exist in a particular country. The word 'ghetto' is derived from the Italian word for the process of 'casting out' in the pro duction of iron. Therefore, if most of the population in a group is forced to reside in an area of a city because they have been cast out of living in other residential areas, then the area they live in can be classified as a ghetto. Such instances are very rare, and usually relate to the African American experience in the early twentieth century and the earlier Jewish experience in Europe. Where residential segregation occurs because of low in come alone then it is not a ghetto but rather an ethnic enclave. The concept of the 'ethnoburb' introduces an other dimension to the classification of ethnic com munities – the socioeconomic dimension. An ethnoburb is an ethnic enclave that has developed or risen up above the base of the housing or income ladder. The classic example of an ethnoburb is that of the residential seg regation of the Jewish community, as displayed in Lon don. Many of these ethnoburbs are located in the suburbs, as in some of the Chinese communities in Auckland, Vancouver, and Sydney.

Self-Segregation

The concept of 'self segregation' implies that individuals in an ethnic group out of choice decide to segregate out from the larger community in terms of their resi dential location. A good example is some members of a religious faith, such as the Jews, who like to or are re quired to live within close proximity of their place of worship; or alternatively, groups that form a community so as to support their co ethnics. Self segregation for all but a few ethnic groups is viewed by many geographers as a contested concept when it occurs beyond the first and second generations. Most research suggests that second and subsequent generations would prefer to live in a mixed community and share in the economic benefits of society, rather than live in their own communities. How ever, in many cities that choice is not available to them.

Better Measures of Residential Segregation

The way residential segregation is measured within cities has been a major theme of geographical research into segregation across the decades, and it is likely to continue to be so, particularly with the increasing focus in geography on the provision of information for public policymakers. The three most privileged measures of segregation are the indexes of dissimilarity, isolation, and diversity; these are very sound measures when applied to extreme levels of residential segregation. But when applied to high or moderate levels of residential segregation in a multicultural society they are problematic. The reason is that these single measures are averaging measures. For example, take the situation where the number of nonwhite persons living in segregated areas (defined as areas within a city with 70% or more nonwhite) increases over time, while the number of nonwhite persons living in mixed areas also increases (defined as areas within a city with 20–70% nonwhite). If the change in the mixed areas exceeds that of the segregated areas, then all of those indices will measure a decline in segregation; when, in fact, the number living in the segregated areas has increased. Hence, geographers have developed a better threshold methodology to separate those differences in the data (see **Table 1** and **Figure 1**). Since the early 1990s geographers have also made a major contribution to the spatial measurement of residential segregation over time at a submetropolitan scale. These measures usually based on census data at sequential censuses have focused on shape, clustering, and the locational changes in the segregation. The problem with spatial measures of segregation is their comparability between cities, in that each city has a unique topography underlying it. However, these measures do reveal important information about changes in residential segregation within an individual city, especially when those measures are conducted across a set of major divisions within that city.

Life Chances

Geographers in Western nations assessing segregated areas and the potential for spatial assimilation today focus on a basic process model containing three components: (1) educational attainment, (2) access to the labor market, and (3) access to the housing market. Each component is interlinked. Educational attainment is seen to determine access to the labor market, and access to the labor market determines access to the housing market. In terms of integration, the geographic literature displays that it is in the field of educational attainment that the initial problems occur for some residentially

Table 1 Conflicting change between the general indexes of segregation: differences in these indices for London, 1991 and 2001

	Index of dissimilarity	*Index of isolation*
White	0.028	– 0.047
Black Caribbean	– 0.057	0.000
Black African	– 0.059	0.059
Indian	– 0.031	0.006
Pakistani	– 0.042	0.015
Bangladeshi	– 0.042	0.022

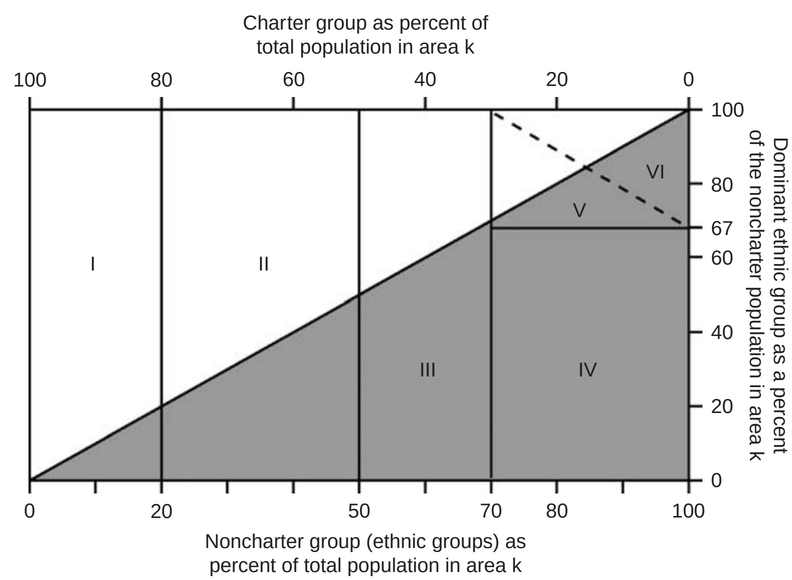

Figure 1 Segregation typology.

segregated groups. Because of their small catchment areas primary schools tend to be more highly segregated than the residential area in which they are situated. Within this environment some students within some ethnic communities fail to achieve the same levels of educational attainment compared with that achieved by the white population living in higher socioeconomic areas, although at the same socioeconomic level some Asian ethnic groups achieve higher levels of attainment than the white population. Similarly, although some nonwhite populations have higher proportions attending university, they do not have higher proportions covering the broader range of further education. It is in exposing these issues through the assessment of life chances that geographers assist with the development of public policy (**Figure 2**).

Qualitative Research

One of the major issues in segregation research has been the relationship between information that can be acquired from empirical measures derived from census and longitudinal data and the experiences of individuals at the local level. In the last decade very valuable information on individual experiences has been derived from qualitative research. In general, numerous vignettes conveying the attitudes and experiences of both white and nonwhite populations in local settings have brought to life the experiences and differences between individuals in different places. Geographers however, increasingly need to combine both quantitative and qualitative methods.

Issues

The span of geographic research on topics covering racial/ethnic residential segregation and income residential segregation has expanded in recent years. Trying to integrate a profusion of findings from a wide variety of geographies and topics has become increasingly a daunting, if not impossible, task. The traditional topics such as measurement of the levels of segregation in a set of metropolitan areas still appear in the literature, as do studies presenting new methods of measuring segregation (**Figure 3**). In almost all cases, geographers aim to supply information that will help those who are disadvantaged to achieve a better life outcome. The imperative of social justice remains dominant, even though at times it might seem that some research yields information that could be misused. Qualitative studies focusing on the experiences of individuals are currently yielding very valuable information. A focus on the community impact of migrant groups within different geographical settings, including the experiences of asylum seekers, and race/ethnic histories forms an important body of contemporary geographical research. Segregation, politics, and public policy are also important geographical topics.

Moving forward, we need to establish a framework within which the key research goals can be placed. One such framework divides the geographic research field into four major subfields: (1) the measurement of segregation, (2) the understanding of segregation, (3) the experiences of segregation, and (4) projections of segregation. Within this framework we can position research into the role of urban governance structures, social housing, educational attainment, access to the labor

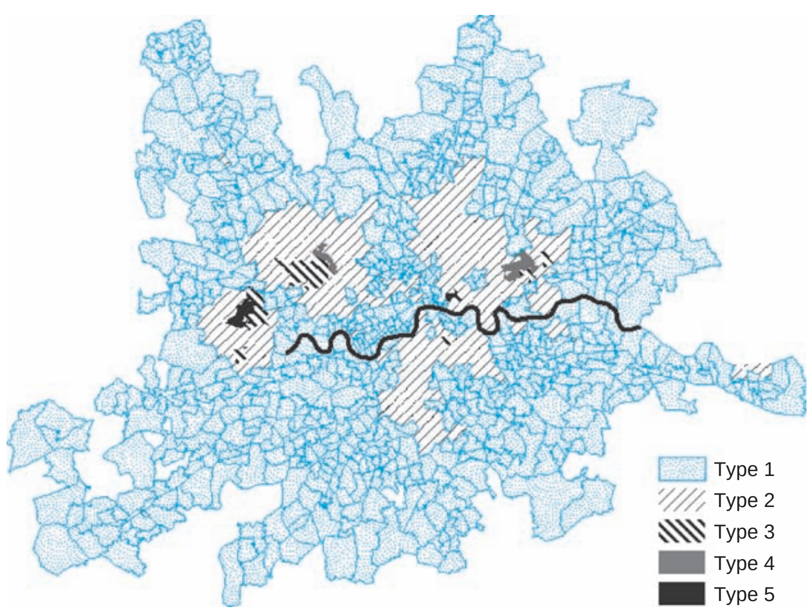

Figure 2 Character and extent of residential segregation in London (1991).

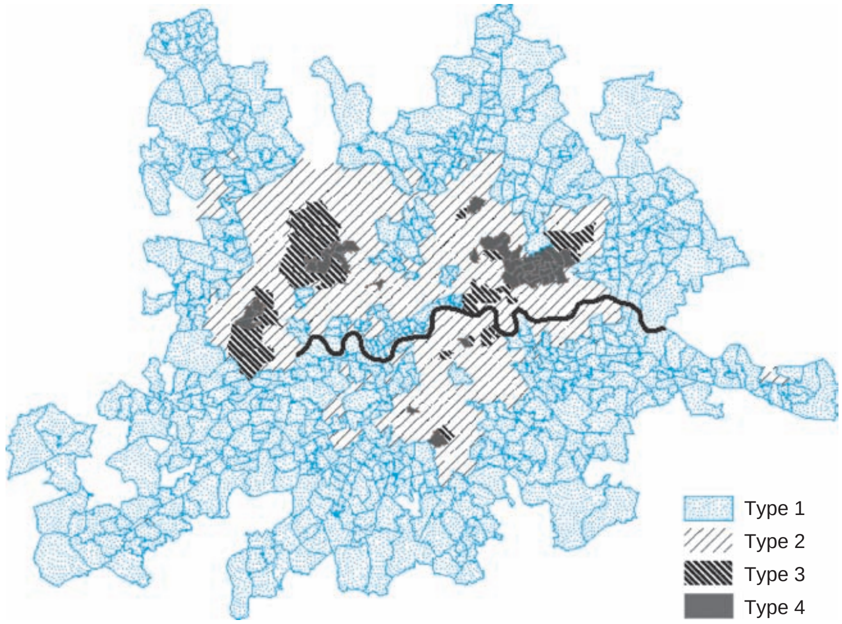

Figure 3 Character and extent of residential segregation in London (2001).

market, access to the housing market, integration, and individual experiences.

Issues of residential segregation are regularly published in the media. Invariably they come to the fore when a major event occurs, such as the Rodney King incident in Los Angeles, the riots in the northern towns of England, the riots in France, the bombings in London, and the riots in Sydney, Australia. Crime linked to low income levels also forms the basis for the media to focus on residential segregation. In each case, the media focuses on the negative aspects of residential segregation, and rarely comments on the positive aspects of the process. Invariably, the media likes to build up a false vision of a divided society, a nation in trouble, in which ethnic communities live in ghettos, and multiculturalism as a failed policy. They do so because such stories sell newspapers. Geographers therefore need to contest many of these perceptions promoted by the media through well developed and accurate research.

Geographic research into residential segregation needs to be developed within the context of each country, while comparative research is also important, particularly as different countries have different data sets and levels of residential segregation. For example, in America we need to establish why different racial/ethnic groups achieve different levels of educational attainment? What impact does Latino immigration have on the relationships between the white population and the African American population? In Australia, the issues tend to be on how do we speed up the process of integration so that it does not take generations for all individuals in an ethnic group to have equal life chances as members of the white population? How do we redistribute ethnic groups to cities other than Sydney and Melbourne? Whereas in New Zealand how do we address the problem that of some children among the Maori and Pacific Islander ethnic groups do not achieve their potential position within society? And what advice do we provide public policymakers in Western European nations where social housing for all income groups is declining and as a result income and ethnic segregation is increasing? Or how do we deal with, as in the United Kingdom, ethnic residential segregation in the northern English towns where access to the labor market is limited and the white population is declining? Or how do we overcome the containment of many ethnic communities within the lowest socioeconomic stratum? Unfortunately, there are no simple answers in any country, nor global answers that span all countries.

In reality, most racial/ethnic groups in Western countries do not live in ghettos. Instead, they live and work in a society where they are needed by the majority white population who are increasingly reliant on these ethnic communities to undertake the lower skilled occupations. While many in the population are uncertain over where their societies are heading when they are made up of a set of separate communities, most see huge benefits in a multicultural society. To assist in evaluating likely futures geographers are currently developing modeling, which allows us to make projections on levels of residential segregation in 10 and 20 years hence. To illustrate one such model based on spatial assimilation starts with an all white urban area, where the initial inflow of unskilled immigrants establishes the ethnic communities that in time form residentially segregated areas. Individuals stay within these areas for the first two generations, with some in the second and third gener ations moving out to areas of mix in surrounding areas within the same socioeconomic stratum. A limited number achieve higher levels of educational attainment and move to areas of higher socioeconomic status, and hence achieve spatial assimilation. Over time, those moving into mixed areas at the same socioeconomic stratum will see their areas transform back into segre gated spaces due to the further settlement of members of the ethnic groups in those areas. Hence, what starts off as both increasing populations in areas of segregation and areas of mix, with mixed areas dominant, will under the worst case scenario time result in areas of increasing segregation unless movement up through the socio economic stratum is achieved by individuals. Hence the importance of spatial assimilation, educational attain ment, and the development of shared values in the development of a new society. Essentially, only through educational attainment can individuals move to higher paid employment within the labor market, and to better quality housing in the housing market. And only through shared values will cities with a set of ethnic based com munities achieve integration and social cohesion.

See also: Governance, Urban; Segregation.

Further Reading

Allan, J. and Turner, E. (2005). Ethnic residential concentrations in United States metropolitan areas. *The Geographical Review* 95, 267 285.

Boal, F. W. (1999). From undivided cities undivided cities: Assimilation to ethnic cleansing. *Housing Studies* 14, 585 600.

Clark, W. A. V. (2003). *Immigrants and the American Dream: Remaking the Middle Class*. New York: Guilford Press.

Ellis, M., Wright, R. and Parks, V. (2004). Work together, live apart? Geographies of racial and ethnic segregation at home and at work. *Annals of the Association of American Geographers* 94, 620 637.

Hiebert, D. and Ley, D. (2006). Introduction: The political economy of immigration. *Tijdschrift voor Economishe en Sociale Geographie* 97, 2 6.

Johnston, R. J., Burgess, S., Wilson, D. and Harris, R. (2006). School and residential segregation: Analysis of variations across England's Local Educational Authorities. *Regional Studies* 40, 973 990.

Massey, D. and Denton, N. (1993). *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, MA: Harvard University Press.

Peach, C. (2006). Muslims in the 2001 Census of England and Wales: Gender and economic disadvantage. *Ethnic and Racial Studies* 29, 629 655.

Poulsen, M. F., Johnston, R. J. and Forrest, J. (2004). Is Sydney a divided city ethnically? *Australian Geographical Studies* 42, 356 377.

Skop, E. and Li, W. (2005). Asians in America's suburbs: Pattern and consequences of settlement. *Geographical Review* 95, 167 188.

Wong, D. W. (2005). Formulating a general spatial segregation measure. *The Professional Geographer* 57, 285 294.

Segregation Indices

D. W. Wong, George Mason University, Fairfax, VA, USA

Glossary

Checkerboard Pattern Spatial distribution with alternate white and black cells like the checkerboard.

Composite Population Count Sum of population in the reference unit and populations in the neighboring units.

Global Measures Summary statistics or indices that describe the overall characteristics of a region.

Local Measures Statistics or indices that describe the characteristics of subunits with a region and can be assigned to each local unit.

MAUP The inconsistencies of analytical results due to changes in the sizes or scale of areal units (scale effect) and the zonal boundary structure (zoning effect).

Spatial Autocorrelation How the values of observations are related in space, clustering, or dispersion.

Introduction

Relationships among population groups have always been a concern in several social science disciplines, including sociology, economics, geography, political science, and, to some extent, psychology. The nature of the relationship is often determined by how the population is divided into groups according to demographic and socioeconomic characteristics. Among these characteristics, race or ethnicity is frequently used. The general concern is to what extent different population groups interact or are separated. From a geographical perspective, it is important to evaluate how different population groups are separated in space – a topic often known as segregation.

As segregation can be a highly charged political and social topic, researchers have been attempting to develop methods for measuring the magnitude of segregation. These measures are relevant for policy formulation when places are compared according to their segregation levels, and partly to better describe phenomena that are somewhat related to subjective feelings, such as being discriminated against. Sociologists led the efforts in the 1950s to develop measures of segregation when modern census data became widely available, and later economists attempted to explain and model individual level behavior leading to the emergence of segregated patterns. Segregation measurements developed by sociologists decades ago are still widely used in contemporary studies. Geographers joined the effort to develop segregation measures, starting in the 1970s with the focus to develop measures capturing the spatial dimensions of segregation, followed by another surge in the early 1990s in searching for better spatial measures. Measuring segregation is also of interest to ecologists and biologists in studying the separation of animal and plant species.

This article first discusses the nature of segregation and different dimensions of segregation in the next section. Then it reviews several popular measures of segregation. Because these popular indices are not spatial in nature, several approaches in developing spatial segregation measures will be discussed. Most measures are global measures that summarize segregation for the entire region; however, local versions of these measures have been developed recently depicting the levels of segregation for areal units within a region. These local measures are reviewed, if they exist. Finally, software for computing segregation indices are discussed.

Nature and Conceptualizations of Segregation

Measurements of segregation are dependent upon the conceptualizations of segregation. Most people equate segregation with separation. When one population group is separated from another one, it is generally acknowledged that segregation exists. Depending on where the separation occurs, we can label different types of segregation. If the segregation is manifested in the residential space, we label it as residential segregation. If it happens among schools, we call it school segregation. Therefore, we can label segregation according to where it occurs.

However, segregation also involves categorizing population into groups or classes. There are many variables we can use to divide the population into groups. Race or ethnicity is the most common attribute used in segregation studies. One can also use income, socioeconomic status, occupation, demographic (e.g., age, gender), and cultural (religion and language) variables.

Sociologists have suggested five dimensions of segregation, reflecting how segregation is conceptualized and how segregation can be measured. A very common way to characterize segregation is based on the evenness of a group's distribution. Two extremes are possible when the evenness dimension is concerned. If all population groups are distributed across all units in the same manner, then perfect evenness or no segregation is achieved. Note that this is not the same as having the equal number of people of the group in all units, but rather the group composition or population mix is uniform across all units. This also means that the population mix in each unit is the same as the population mix of the entire study area. On the other hand, if each areal unit is occupied exclusively by one population group, then a perfectly uneven situation or perfect segregation arises. Evenness is a very important conceptualization of segregation. It is best represented by the index of dissimilarity *D* or the segregation index.

Another segregation dimension is exposure, which refers to the opportunity or magnitude of one group's exposure to another group. The opposite of exposure is isolation. In the geographical literature, exposure is related to the concept of interaction. If one group is highly exposed to another group, the two groups can potentially interact and therefore, the segregation level should be relatively low, and vice versa. Note that the exposure concept is often not symmetrical. When the minority has high exposure to the majority in an area, the reverse is likely not true. However, when the concept of interaction is used, the relationship is often symmetrical.

While the exposure dimension focuses on the relationship between groups, the concentration dimension is primarily for one group. It is based upon the concept of density. High density of one group may result in dominance. When one group dominates an area, it is likely that this group is segregated from other groups. Many measures have been used to measure concentration, besides the simple density measure. Location quotient is frequently used in human geography, and can be used to analyze population concentration.

Historically, minority populations were concentrated close to the central cities in many North American metropolitan areas. Because of the disadvantaged socioeconomic condition, minority populations seemed to be trapped in the inner cities, while populations of higher socioeconomic status were suburbanized. Therefore, the fourth dimension of segregation, centralization, refers to the concentration of minority population in the city center. Measures reflecting this dimension are often based upon simple proportions of minority residing within the central city out of the entire metro area. This dimension is both country and time dependent. Many Latin American cities have an urban structure inside out to North American cities: minorities or poorer people reside along the fringe of the cities while the higher socioeconomic classes occupy central locations. Even in some North American cities, a recent trend is that minority groups are moving to suburban locations. Therefore, the general applicability of the centralization dimension is context dependent.

As separation of population is the essence of segregation, one manifestation of group separation is clustering. When population groups exhibit high degrees of clustering, different groups occupy their own territories and therefore are segregated. Some researchers claim that this conceptualization of segregation is the most spatial of all. Geographers have been deriving measures to reflect levels of clustering. When similar values cluster, positive spatial autocorrelation exists. Many statistics have been suggested to evaluate the level of spatial autocorrelation, including Moran's *I* and Geary Ratio *C*. These measures are appropriate to measure the spatial clustering of population groups.

The five dimensions (evenness, exposure, concentration, centralization, and clustering) cover many facets of segregation, but some geographers strongly feel that measures associated with these dimensions fail to effectively capture the spatial separation of population – the most important aspect of segregation. Recent resurgence in segregation research in geography is partly in response to this methodological deficiency. New spatial measures have been proposed to include the modeling of spatial relationship of population groups either implicitly or explicitly using some versions of distance, which include adjacency.

Another direction of segregation research that has emerged recently the development of local measures to facilitate neighborhood level analysis. Most common segregation measures can be regarded as global measures because one value is used to summarize the overall segregation level of the entire region, which may be a state or a metro area. But segregation often varies quite significantly within the study region. Therefore, relying on one summary value fails to recognize different experiences the residents have within the region. Some local measures are simply the spatial decompositions of the regional or global summary measures, an approach consistent with the development of local spatial autocorrelation statistics.

Traditional Measures

Several common types of segregation measures will be discussed below. These measures are applicable only to nominal or categorical scale population variables. If ordinal or interval/ratio scale variables are used to classify population, other types of measures will be needed or some modifications of measures discussed below will be required.

Dissimilarity Measures

Among the five dimensions discussed above, the evenness dimension has been regarded as the most important dimension, and the index of dissimilarity *D* or segregation index is regarded as the most effective measure for this dimension.

The dissimilarity index D was developed in North America when African American was the dominant minority group. Therefore, the two groups used in the formulation of the index were white and black. To generalize the discussion, we use group A and B. The D index is formally defined as

$$D = \frac{1}{2}\sum_{i=1}^{n}\left|\frac{a_i}{A} - \frac{b_i}{B}\right| \qquad [1]$$

where a_i and b_i are the population counts of the two groups in areal unit i, A and B are the total population counts of the two groups in the entire study region, and n is the number of areal units in the study region. The $|\cdot|$ is the absolute value of the difference of the two ratios, which are the proportions of group in areal unit i over the group populations in the region. If the first group is being allocated proportionally more to one unit than the second group to that unit, then the first group will be more dominant. Thus the proportion difference indicates how the two groups are different in their shares in that unit. Bigger differences indicate the dominance of one group. The D index does not indicate which group is more dominant, as the absolute values are taken on their differences. The index is just an overall indicator of dominance or segregation. The absolute proportional differences of all areal units are summed and half of that is the D index value.

The D index has a theoretical range from 0 to 1, where 0 reflects no segregation and 1 indicates perfect segregation. When the proportions of the two groups are the same for all areal units, or

$$\frac{a_i}{A} = \frac{b_i}{B} \;\forall i$$

then $D = 0$, indicating no segregation. When each areal unit is exclusively occupied by one group, then $D = 1$, indicating perfect segregation. Sometimes, the D value is converted into percentage to show the percent of population needed to be moved in order to achieve no segregation. This interpretation of D is one of the reasons why the index is frequently adopted.

Using Washington, DC 2000 Census data at the tract level as an example, **Figure 1a** shows the proportion of black based on the white–black total at the tract level (the two ovals will be discussed later). The residential segregation of the two groups is clearly visible with blacks

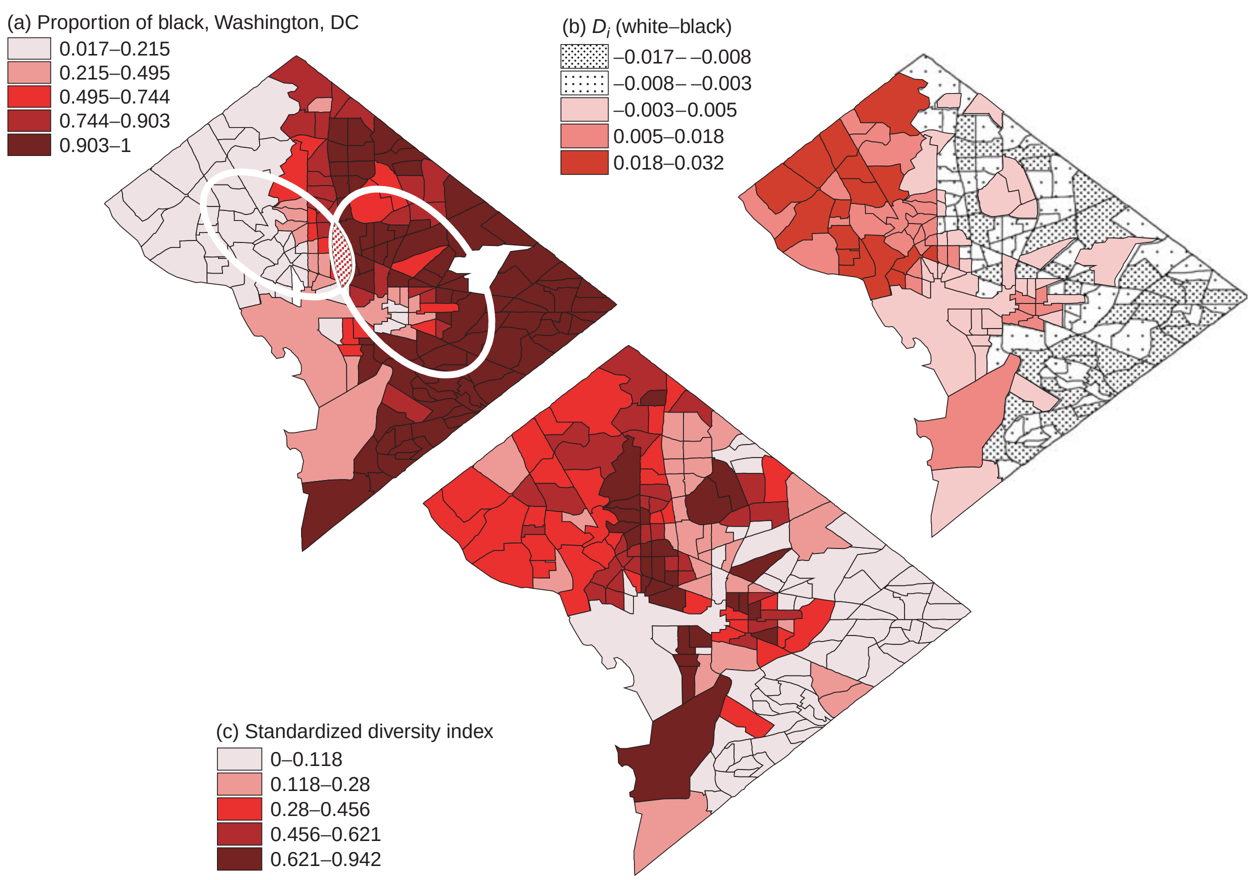

Figure 1 (a) Proportion of black, Washington, DC, 2000 Census tracts, using white black totals as the base; (b) D_i index; and (c) Standardized diversity index.

concentrated on the east side and whites concentrated on the west. As a result, the distribution offers a D value of 0.7758.

A criticism of D is that it can handle only two groups, but multiethnicity is the norm rather than the exception. To overcome this limitation, pair wise comparisons of all possible pairs of population groups can be performed using D. An alternative is to use a multigroup version of D, which is defined as

$$D(m) = \frac{1}{2}\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}|N_{ij} - E_{ij}|}{\sum_{j=1}^{m} N P_j(1 - P_j)} \qquad [2]$$

and

$$E_{ij} = \frac{N_{i\cdot} * N_{\cdot j}}{N}$$

where N_{ij} is the count of population group j in areal unit i. There are n units in the region and m population groups. Then, $N_{i\cdot}$ is the total population in unit i across all population groups and $N_{\cdot j}$ is the total population of group j across all units in the region. N is the total population of the entire region and P_j is the proportion of population belonging to group j in the region. This formulation is similar to the χ^2 (chi squared) statistic. $D(m)$ has the same numerical properties as D, such as bounded by 0 and 1. In the Washington, DC example using white, black, and Asian, the $D(m)$ value is 0.7430, very close to the D in the two group case.

Both D and $D(m)$ are global measures because they offer one index value for the entire study region, regardless of how large the region may be. Therefore, they cannot reflect the internal variation of segregation level within the region. Local measures may be used to depict segregation at the neighborhood scale and one approach to develop local measures is to spatially decompose global measures. For each areal unit i, the local D_i is

$$D_i = \frac{a_i}{A} - \frac{b_i}{B} \qquad [3]$$

D_i is essentially the difference in the group proportions for each areal unit. D_i can be positive or negative. Positive values indicate that group A is dominant in i, and vice versa. An advantage of using local measures such as D_i is that index values can be mapped to show the variation geographically and thus this approach is appealing for local policy formulation. **Figure 1b** shows the D_i values of Washington, DC. The local index effectively depicts the dominance of the two groups in different parts of the city.

Entropy-Based Diversity Measures

Another popular measure under the evenness dimension is the entropy based diversity index H. Different from D, the H index can handle multiple groups. If p_k is the proportion of population group k in unit i, or $p_k=$ population of group k in i/total population in i, then, H_i, the diversity for unit i, the local version of the diversity index, is

$$H_i = -\sum_{k=1}^{m} p_k * \ln(p_k) \qquad [4]$$

where $\ln(\cdot)$ is the natural logarithm and m is the number of population groups. Note that the logarithm of other bases can also be used. Though the diversity index has a minimum value of 0 indicating no diversity, its maximum value is dependent upon the number of groups involved. Therefore, to facilitate comparison, the index is often standardized by dividing by its maximum value, $\ln(m)$. Because this index measures diversity, higher values in H_i indicate less segregation, while lower values reflect more segregation. The standardized diversity value for each census tract in Washington, DC is mapped in **Figure 1c**, including white, black, and Asian in the comparison.

To derive a global version of the index for the entire study region, there are several ways to 'scale it up'. One method is to compute an average over all units within the region, but is weighted by the population sizes of units. Formally,

$$H = \frac{\sum_{i=1}^{n} H_i P_i}{\sum_{i=1}^{n} P_i} \qquad [5]$$

where n is the number of units in the region and P_i is the population count in unit i.

Exposure and Other Measures

Recent studies have increasingly utilized the exposure concept to evaluate the level of segregation. The exposure index, ${}_aP_b^*$, describes the chance that a person in group A will encounter another individual in group B. Formally,

$${}_aP_b^* = \sum_{i=1}^{n}\left(\frac{a_i}{A}\right) * \left(\frac{b_i}{N_i}\right) \qquad [6]$$

where N_i is the total population in areal unit i and all other terms are defined as before. The exposure of group A to group B is not the same as group B to group A, but ${}_bP_a^*$ can be derived in a similar manner. These are global aspatial exposure or isolation measures. Spatial versions of these measures and spatial local exposure measures have been developed.

Measures for the concentration dimension are mostly density based measures, including the location quotient. For centralization measures, most of them are simple proportions of population between the inner cities and the entire region. Therefore, they will not be elaborated here. For the clustering dimension, some measures are spatial autocorrelation measures, including Moran's *I* and Geary Ratio *C*. Some measures are specific to measuring segregation but spatial in nature. They will be reviewed in the next section.

Spatial Measures of Segregation

Most popular segregation measures have a common deficiency: they cannot distinguish different population distribution patterns effectively. The so called 'checkerboard' pattern is the best example to illustrate this weakness. According to most measures, the checkerboard pattern has perfect segregation, as is true for other patterns when each unit is exclusively occupied by one group. It is because most measures assume that interaction across unit boundaries is not permitted, even though those boundaries are often artificially demarcated. If all black cells of the checkerboard are moved to one side and white cells to the other to form a two subregion pattern, the level of segregation should increase, but traditional measures will indicate that all these patterns have the same level of segregation (perfectly segregated). As a result, many spatial measures have been proposed to overcome this weakness. In order to effectively distinguish different spatial distribution patterns, these measures incorporate information about the spatial relationships among areal units.

One way to measure the spatial relationship between areal units is distance. Therefore, early spatial segregation measures, such as the distance based segregation index (DBI), incorporate a distance component. As *D* reflects the proportion of population needed to be moved in order to achieve no segregation, the distance that people are required to move can also be taken into account. Therefore, DBI is the minimum distance weighted by the populations that are required to move to achieve no segregation. To derive the index, optimization methods are required and the details are beyond the scope of this article.

Also using distance as an operator to capture spatial relationship between people or areal units, the spatial proximity index was proposed. The index can be defined as

$$P = \frac{AP_{aa} + BP_{bb}}{NP_{nn}} \qquad [7]$$

where *A* and *B* are the total population counts of the two groups in the region and *N* is the combined total of the two groups. P_{nn} is the average proximity between all persons in the city regardless of the group membership. Therefore, P_{aa} is the average proximity of population in group *A* and P_{bb} is the average proximity of population in group *B*. For the average proximity of a group, let us say *A*, the computation follows:

$$P_{aa} = \frac{1}{A^2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j d_{ij} \qquad [8]$$

where d_{ij} is a general distance measurement between units *i* and *j*. The distance measurement can accommodate various distance specifications (e.g., Euclidean, network, or Manhattan, etc.) and distance decay functions.

Besides distance, contiguity can also be used in analyzing spatial relationships between population groups. A family of spatial segregation measures modifies the *D* index by accounting for the interaction across neighboring units. The traditional *D* measure overstates segregation level by ignoring interunit interaction. The spatial version of *D* suggested by Morrill gives a segregation value of 0.6444, compared with 0.7758 of the aspatial *D* for Washington, DC. The spatial version proposed by Wong offers a segregation value of 0.6588, very close to the one by Morrill.

Another approach to accommodate interunit interaction is to utilize the concept of composite population counts. To implicitly allow interaction across unit boundaries, the composite population count of a unit includes populations of the reference unit and its neighbors so that populations in the reference and neighboring units will be treated as if they are within one unit. The composite population count of a unit *i* is defined as

$$ca_i = \sum_{j=1}^{n} f_{ij}(\cdot)a_j \qquad [9]$$

where a_j is the population count in unit *j*, and $f_{ij}(\cdot)$ is a neighborhood function, which may carry a binary form of 0 and 1 with 1 indicating that *j* is a neighbor of *i*, and 0 otherwise. Then using the composite population counts instead of the raw counts to compute segregation measures, population mix of neighboring units will be considered implicitly. Several spatial measures, including a generalized *D* measure, the spatial multigroup *D* index, the spatial version of the entropy based diversity indices (both local and global, eqns [4] and [5]), and the local spatial exposure indices adopt the composite population count concept in measuring segregation. The spatial multigroup *D* index for Washington, DC is 0.6553, compared to 0.7430 of the aspatial multigroup *D*.

Several other directions in developing spatial segregation indices are worth mentioning. The *D* index is conceptually linked to the Gini coefficient or Lorenz curve for measuring inequality. Like many indices used in human geography, spatial relationship of areal units is not taken into consideration in Gini coefficient and a spatial version of the Gini coefficient has been proposed to measure spatial segregation.

In another direction, because spatial distribution of a population may be graphically captured by a standard deviational ellipse, an ellipse can be created for each population group and ellipses for different groups can be overlaid. The overlapping area of ellipses reflects the spatial correspondence of these population groups. Therefore, an index formed by 1 – (area of ellipse intersection/area of ellipse union) can serve as a possible spatial measure of segregation. **Figure 1a** shows the el lipses for whites (left) and blacks (right). The intersected area was also highlighted. The two groups did not overlap much spatially. As a result, the segregation value is 0.9757.

Related Issues on Measuring Segregation

Most global traditional measures are quite straight forward to compute, but for spatial measures, special software, such as geographic information system (GIS), is needed especially to capture the spatial relationship. Currently, several software packages are available for students and researchers to compute both traditional and spatial segregation measures. A package based upon ArcView 3.X GIS is available for download. Another stand alone program has implemented some 40 segre gation measures and is available.

In addition to measuring segregation, researchers have also explored how and why segregation emerges. One approach to explaining the emergence of segregation is through the modeling of individual behavior. Decades ago, Thomas Schelling, the Nobel Laureate in eco nomics, proposed a micro level simulation model to explain the emergence of segregation patterns. Some researchers have implemented Schelling's simulation model (e.g., SimSeg originally developed by Mark Fossett and Chris Cook). On the other hand, research shows that segregation is not just a result of individual behaviors but of social and institutional forces in different social spaces, including residential, work, and school spaces. Housing market discrimination and redlining are well documented processes which create and reinforce segregation.

Most measures of segregation rely on ecological data – data gathered for aggregated spatial units. When ecological data are involved, the so called modifiable areal unit problem (MAUP) exists. The MAUP states that results vary depending on the sizes and boundaries of areal units used in measuring segregation. While the literature has addressed how segregation measures are affected by the MAUP, a simple example is used to illustrate the essence of the problem. If Washington, DC (**Figure 1a**), is divided into two units only, we can choose the division from north to south along the middle. Then each unit will be very much dominated by one group and the pattern will result in a highly segregated situation. If we divide the region horizontally, both the northern and southern units will have relatively mixed populations and the region overall will have a much lower level of segregation. Therefore, segregation level is affected by how the boundaries of areal units are drawn. The MAUP effects on measuring segregation have been analyzed and, to some extent, are understood. However, it is advisable to take into consideration the MAUP impacts on meas uring segregation in research and practice. An important and challenging topic in ongoing research is to handle and model the MAUP impacts effectively on measuring segregation.

See also: Modifiable Areal Unit Problem; Spatial Autocorrelation.

Further Reading

Coulter, P. B. (1989). *Measuring Inequality A Methodological Handbook*. Boulder, CO: Westview Press.

Dawkins, C. J. (2004). Measuring the spatial pattern of residential segregation. *Urban Studies* 41, 833 851.

Duncan, O. D. and Duncan, B. (1955). A methodological analysis of segregation indexes. *American Sociological Review* 20, 210 217.

Jakubs, J. F. (1981). A distance based segregation index. *Journal of Socio Economic Planning Sciences* 15, 129 141.

Lieberson, S. (1981). An asymmetrical approach to segregation. In Peach, C., Robinson, V. & Smith, S. (eds.) *Ethnic Segregation in Cities*, pp 61 82. London: Croom Helm.

Massey, D. S. and Denton, N. A. (1988). The dimensions of residential segregation. *Social Forces* 67, 281 315.

Morrill, R. L. (1991). On the measure of geographical segregation. *Geography Research Forum* 11, 25 36.

Reardon, S. F. and Firebaugh, G. (2002). Measures of multigroup segregation. *Sociological Methodology* 32, 33 67.

Schelling, T. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology* 1, 143 186.

Cook, C. The Schelling Segregation Model. http://www.econ.iastate.edu/tesfatsi/demos/schelling/schellhp.htm (accessed on 23 August 2007).

White, M. J. (1983). The measurement of spatial segregation. *American Journal of Sociology* 88, 1008 1018.

Wong, D. W. Spatial Segregation Measurement Research. http://geog.gmu.edu/seg (accessed on 23 August 2007).

Wong, D. W. S. (1993). Spatial indices of segregation. *Urban Studies* 30, 559 572.

Wong, D. W. S. (1997). Spatial dependency of segregation indices. *The Canadian Geographer* 41, 128 136.

Wong, D. W. S. (1998). Measuring multi ethnic spatial segregation. *Urban Geography* 19, 77 87.

Wong, D. W. S. (1999). Geostatistics as measures of spatial segregation. *Urban Geography* 20, 635 647.

Wong, D. W. S. (2002). Modeling local segregation: A spatial interaction approach. *Geographical and Environmental Modelling* 6, 81 97.

Relevant Websites

http://www.simseg.com
Advanced Simulation for the Social Sciences.

http://www.inrs ucs.uquebec.ca
Institut National de la Recherche Scientifique. Urbanisation, Culture et Société.

Segregation, Urban

R. Atkinson, University of Tasmania, Hobart, TAS, Australia
J. McGarrigle, Universidade de Lisboa, Lisbon, Portugal

Glossary

Area Effects The net contribution to the life-chances of an individual based on living in a particular neghborhood or area

Gated Communities Residential development that has gates and walls or fences and which may be patrolled by guards or have other surveillance mechanisms. Gated communities are also distinctive in combining these physical features with socio-legal and contractual obligations that govern the type of housing and conduct of residents and visitors

Lifeworld Taken from sociology, the world around us that is shared and taken as a given for daily social experience.

Socially Tectonic Relationships Used to describe the close spatial proximity of social groups, in a neighborhood, but whose social interaction is very limited. Initially used by Tim Butler and Gary Robson in their study of gentrification in London to describe the general absence of social contact between gentrifiers and existing local, working class populations.

Introduction

Many commentators have acknowledged that social relationships tend to be conducted with people who are in some way like ourselves. This notion of homophilia, the desire to be with people like oneself, suggests that those who hold similar values and similar status positions, including ethnic identities, will tend to band together. When these tendencies in social behavior are expressed spatially, in towns and cities, this creates patterns of social similarity within neighborhoods. Where choice charac terizes residential decisions, we often find that social areas develop which may show a prevalence of particular social characteristics. Tim Butler, in the context of middle class migration and gentrification in London, talks about these tendencies in terms of a desire to be with 'people like us', the sense that people do not necessarily express exclusive ideas about identity but value a local community that reflects their own identities and aspirations.

It is important to remember that these social forces are played out within the context of private housing markets so that choice is never fully unconstrained. In housing markets, price operates as an indicator of the value of locational attributes which thereby excludes those who cannot afford properties or rents in such areas. This means that a sorting process of social identity comes about through the interplay of these market forces combined with social forces and produce patterns of relative segregation on the basis of household income. Nevertheless, provision of affordable and public housing has often been planned in ways, particularly in regions like Europe, to maintain social diversity at the neigh borhood level and thus avoid areas of concentrated poverty.

In cities like London, patterns of segregation and social mix occur unevenly. Areas like Islington, for example, have not only seen extensive gentrification but also retained public housing, whereby rich and poor live, to some extent, side by side. As Butler has shown, the worlds of these poorer renters and new affluent owners are often not just different but separate from each other. Knowledge of each other is mutually exclusive, in what are characterized as 'socially tectonic' relationships – wherein the affluent float separably above a poorer and minority ethnic grouping within the same neighborhood space. This raises a further issue which relates to the importance of scale and how we interpret the prevalence and degree of segregation. At what point can we actually say that segregation exists to one degree or another? We might reasonably set an arbitrary neighborhood size at 2000 households. If this area contained entirely Black residents we might suggest that it was segregated from other areas surrounding it that were entirely White. However, if our imagined neighborhood and the sur rounding areas had an equal number of Black and White residents we would find it difficult to express the extent of segregation. A segregation index would suggest that there was an even distribution and therefore no segre gation. Yet, what if the Black and White residents in many neighborhoods were divided within these spaces?

This discussion raises two key issues. First, the measurement of segregation is contentious but remains important because it informs broader public debates about how cities and housing should be managed to achieve socially equitable outcomes. If we believe that segregation expresses prejudice and social division then understandings of its scale are clearly important. As Richard Sennett has argued, the fact that areas like downtown Manhattan are ethnically diverse does not mean that those groups interact with each other.

However, he argues that local social difference influences our understanding of those who are different from our selves. If we do not see this difference in daily life then empathy is made more difficult, with all of the impli cations this has for debates about citizenship and diver sity in civic life. As shown later, the existence of *de jure*, or legally sanctioned, segregation in countries such as South Africa and the US has shown that contemporary patterns of ethnic clustering are not only merely due to chance or the operation of the housing system but also reflect the historical attempt to minimize contact with their Black populations. The social inequity and injustice that has stemmed from this has become a key feature of contemporary urban and ethnic relations in these countries.

Second, where social groups tend to live apart neighborhoods may operate as social containers for the residents of poorer areas that are separated from the wider social and economic life of the city. This issue has been given significant attention by researchers who have focused on what are known as area effects. This concept refers to the net contribution to a person or a household's life chances that stem from living in an area of ghettoized poverty. If one is on a low income and lives in an area with many other people on low incomes, it has been suggested that such neighborhoods tend to further the reproduction of poverty. These processes operate in various subtle ways. For example, low income areas tend to be stigmatized and unpopular with some prospective employers who see residents as lazy or welfare depend ant. This can, in turn, lead to self fulfilling prophecies that cast low income groups as feckless or incapable. Other problems may stem from the additional burden on essential services, such as schooling, which may be pressured by high concentrations of poorer households. Equally, service providers in such neighborhoods may not see residents as worthy of help.

The work of William Julius Wilson has been import ant in highlighting how Black communities have often been stripped of their working and middle classes; a process which has, in turn, had important effects on the socialization of young people, notably by the loss of important local role models. There is clearly some degree of paternalism in such debates and a fear, in European countries, that such arguments run the risk of casting poor, minority ethnic households as an underclass responsible for its own plight. Contrary to this, Jargowsky has shown that, as broader economic opportunities rose across the US in the 1990s, work participation rates also rose in poorer, ghettoized Black communities, sug gesting that segregated poverty is not disconnected from potential economic opportunities.

Segregation may not simply be good or bad and may vary in its functions depending on the social and ethnic context of such patterns. This is particularly the case in the European context where the extent of segregation and ghettoization is less advanced than in North American cities. Peach argues that segregation is distinct from ghettoization. For example, the relative segregation of groups such as the Irish, Poles, and others have all tended to change over time as assimilation has occurred and economic well being has improved. Here, spatial clus tering in cities promoted what Peach describes as sup portive 'ethnic villages', or sanctuaries, which served protective and supportive roles for their residents. In other words, patterns of segregation may change over time as the relationships between ethnic groups also change. However, Peach also sees that segregation can have negative con notations, where it may represent an attempt by dominant groups to keep underprivileged ethnic populations spa tially and socially excluded.

Global Views on Variations and Explanations of Segregation

Much of the literature on segregation has focused on North American cities. Notable contributors such as Massey and Denton have described the growth of seg regation and its underlying causes. The problem has tended to be seen as emerging most strongly following the late 1960s, though ghettoization and concentrated poverty were much in evidence, particularly from the 1930s. Even then we might look further, back to earlier phases of urbanization where such sociospatial divisions were in evidence and a key feature of the work of the Chicago School. By 1970, the majority of the US popu lation lived in the suburbs of its cities, a result, in part, of technological advances, cheap fuel, and 'White flight' from the urban cores where unrest and rioting arising from concentrated poverty and an absence of civil rights became pronounced. As a result, patterns of ethnic differentiation became stark in many US cities such as Detroit which moved from being a city with around 16% Black and foreign born in 1950 to 81.6% in 2000, with the overall population dropping from 1.8 million to around 900 000 in 2004. The cost of this segregation and poverty concentration was the perpetuation of social and racial disadvantage, often locked out of suburbanizing work opportunities in what some saw as a spatial mismatch between transport, poor Black households, and the opportunities in new suburban areas.

Another key ingredient in the kind of segregation seen in the US was its mandating by law since the late nineteenth century. To a significant extent the features of social life in American cities have continued this legacy and have generally been unable to overcome the abo lition and integrationist concerns that led to desegre gation in schooling and social life from the 1950s onward, and which culminated in the Civil Rights Act of 1964.

To some extent this enshrining of ethnic separation in law stemmed from that used in apartheid South Africa where similar concerns about racial mixing and the potential for racial dilution through marriage were prohibited by law. In Nazi Germany, prohibitions against marriage between Jews and German 'Aryans' from the 1930s led to the ultimate spatial separation and collective isolation of the ghettos of Warsaw, among others in Poland. The term ghetto, while widely used today to refer to places of concentrated poverty, originally stemmed from the fourteenth century concentration and, later, containment of Jews in the foundry area of Venice (ghetto being a word derived from the Italian *ghetta* which was used to refer to the cooling waste material from iron production). While the word ghetto is often associated with areas of extensive Black poverty, the terms barrio and slums are used to refer to areas of Hispanic and White poverty today.

Religious segregation continues to be a feature of many cities today. In its most extreme form, in cities such as Belfast, its Catholic and Protestant communities continue to be highly segregated. Middle class movement out of the city has historically served to further isolate and polarize those left behind in the inner city. Cities such as Glasgow, with its history of Irish migration, have retained a less sharply divided set of sectarian boundaries, yet continue to display ritualized religious and ethnic difference through social life and sport.

In the European context, urban life has been characterized by divergent levels of migration and minority ethnic concentration and generally less extensive sociospatial segregation than is found in US cities, yet concern with the specific problems of such concentration goes back at least to the 1950s, as new waves of migration from the Commonwealth to the UK followed the wake of the postwar boom. In recent years, concern has switched to the integration, opportunities, and relative concentration of new migrants in European cities who have often fled wars and human rights abuses in countries such as the former Yugoslavia and Iraq, and who have also been attracted by the economic opportunities opened up by European integration for countries in the former Soviet and Eastern European states.

All of this raises further issues when segregation is considered in relation to recent housing and urban policies. For example, the desire to build new housing developments with some degree of social mix has become a prominent concern in the UK and in European countries. The use of planning arrangements to encourage stronger neighborhood levels of social diversity also says something about deeper assumptions that relate to the desirability of social diversity as a means of combating segregation and reducing the costs of public service provision in deprived areas. Such assumptions continue to assert a value to social diversity though evidence that greater mix might facilitate stronger futures for poorer residents has remained contentious.

In recent years, the emerging picture of segregation has not been positive. Writers such as Kaufman have discussed a new urban poverty in terms of a growing spatial separation based on race in cities such as Chicago. This kind of hypersegregation has been produced as a result of broader shifts in the economies of cities which have seen an erosion of middle tier occupations with a shift to advanced service sector employment in the US and professionalization in Europe, where neoliberal welfare reform has also become more advanced. These processes have also become further entrenched in countries like Brazil, among other emerging Third World economies, where a combination of neoliberal market reforms imposed by institutions like the International Monetary Fund (IMF) and social disintegration from economic problems have conspired to produce not only social polarization between rich and poor but also the 'fortification' of gated affluent areas marked off as clear territories from the poorer favelas. This has led commentators to suggest that these forms of segregation and social separation may be emulated in other Western cities in the future.

Measuring Segregation – Debates and Issues

There are longstanding debates and considerable controversy over the measurement of segregation. Measures of segregation are often linked conceptually to five different dimensions – unevenness, exposure, centralization, clustering, and concentration (defined in more detail below). Some of these dimensions of segregation, however, are less applicable to European cities, which have tended not to follow the same model of sociospatial differentiation as American cities. Although there is an ongoing debate regarding which scale might provide the best measurement of spatial segregation, the two most commonly used indices in both the US and Europe are the index of isolation (P^*), a measure of exposure, and the index of dissimilarity (ID), a measure of evenness.

Measures of evenness measure the distribution of minority and majority groups across subareas of a city (e.g., if all of the subareas of a city were found to be 90% and 10% Black, the Black population would be considered to be evenly distributed). The most widely used measure of evenness is the ID. It measures the percentage of a minority population that would have to move to create an even distribution. It computes the sum total in a larger area of the differences in the relative populations in subareas. A figure of zero would denote no pattern of segregation, since the distribution is exactly the same or equal between the two groups being compared. A figure of 1 would indicate total segregation, where the residential distribution between the two populations being compared is totally uneven, or lumped together spatially. In between these two values, an ID less than 0.4 is considered low, 0.4–0.59 as moderately high, and 0.6–0.69 as high and above 0.7 is considered very high:

$$\mathrm{ID} = \frac{1}{2} \cdot \sum_{i=1}^{k} |x_i - y_i|$$

where k = the total number of subareas in the city, x_i = the proportion of the city total of the x population in area i, and y_i = the proportion of the city total of the y population in area i.

The ID is measured as half the sum of the difference between the proportion of the two populations living in each individual k areas of the city. There are key weaknesses that have called the reliability of this measure into question. First, it only measures two groups at a time, so when there are more than two groups it is a less reliable indicator. In this sense it may be more useful for cities where there are only two ethnic groups, though clearly most cities today are multiethnic. Second, it is the problem of scale since the index is aspatial and measures only the degree of segregation, but not its spatial patterns. For instance, if the measured ID is zero, which denotes no segregation, then it could mean that half of the area is Black and the other half White; however, we are unable to tell if the Black and White residents are mixed or if they are each occupying a separate half of a particular area. This is also true of mapping techniques (the second technical approach used to measure segregation), since unless a very small spatial scale is employed we are unable to tell how segregated or evenly distributed the minority group is within an area. Furthermore, erroneous comparisons may be made between studies using different spatial units.

Indices of exposure or isolation measure the degree of potential contact between minority and majority group members within the area of the city where they live, so it is dependent upon the extent to which groups live in the same subarea of the city. This differs from measures of evenness as it attempts to measure the 'experience of segregation' rather than being based on an assumption of ideal, or even, distribution. There are two measures of exposure: the first, and less commonly used, is the isolation index, which measures the degree to which members of minority groups are exposed only to one another. The second, more commonly used measure is P^*, an interaction index. This measures the potential exposure of the minority group to the majority group. P^* is also used to ascertain the group 'from whom and to whom' the exposure is directed; it differs from measures of evenness as it takes into consideration the size of the groups being compared, as there will be greater probability that a member of a smaller group will come into contact with a member of the larger group:

$$P^* = \sum \{(x_1/X)(y_1/t_1)\}$$

where X = the total number of group X in the city, x_1 = the total number of group X in a given subarea, y_1 = the total number of group Y in a given subarea, and t_1 = the total population in a given subarea.

Indices of isolation and interaction are related and can be used together when there are only two groups (adjustments must be made for more than two groups); the isolation and interaction indices sum to 1.0, so lower values of interaction and higher values of isolation both indicate higher levels of segregation.

Clustering measures the degree to which spatial units occupied by minority groups are joined to one another. While the other measures consider the distribution of minority areas relative to a set point (i.e., the center of the city) or the distribution of minority and majority group members (the linked measures of exposure and evenness), this measure looks at the distribution of minority areas in relation to each other. If we compare two cities that have exactly the same scores when the other measures are computed, but in the first city the minority areas are scattered throughout the urban area, and in the second they are adjoining, it would be suggested that the latter is more segregated.

Measures of concentration, which are not widely used and are fewer in number, measure the actual amount of physical space occupied by a minority group, relative to the total land area of the city. This measure (normally termed the Delta index) is based on the assumption that through processes of residential discrimination, minority groups become marginalized and restricted into relatively few neighborhoods that overall occupy a small proportion of the city. There are clear weaknesses with the assumption that high density denotes either social marginalization or discrimination. For example, minority ethnic groups of Asian origin have displayed preferences to live in higher densities in extended family households. Furthermore, lower income populations in several modern British cities have been 'deconcentrated' from the inner city to peripheral housing estates with more space without greater social mix or amenity.

Mobile and 'Static' Forms of Segregation

Much of what we have discussed so far relates to what might be seen as static representations of social life, contained within residential neighborhoods, often as though such areas operated as the fundamental daily lifeworld. As we have seen from the area effects debate, these spaces are often more important for low income groups, yet movement beyond those boundaries may be common. In this respect, theories of segregation have, while including exposure as a measure, often tended to downplay the spatial movement and social integration of social groups in daily life. With the rise of gated communities and other 'master planned' communities, where price and lifestyle operate in excluding ways, we see a growth in nodal geographies in which work, play, socialization, education, and consumption appear as more or less restricted and privatized functions. This kind of segregation can be expressed as a time–space trajectory, wherein patterns of residential sociability, particularly those of the affluent, are extended across daily patterns of spatial movement. The perception of risky urban spaces and engagement with dangerous social 'others' has led to segmented spatial trajectories in which groups associated primarily with people like themselves as a defence against such fears. As writers like Graham and Marvin have noted, the splintering of urban areas into 'premium network spaces' (such as 'pay to use' highways) has forced a new subtle set of excluding moments in urban life. This kind of urbanism means that ability to pay, combined with social identity, has produced forms of segregation which operate across neighborhood bound aries and through fractured corridors of social contact and disconnection.

If segregation is a problem because it expresses deeper social fears and prejudices, while consigning lower in come and excluded social groups to poorer residential conditions, then its dynamic forms serve only to broaden our understanding of these social problems. Rising real incomes and mobile technologies mean that, for affluent groups, privacy and personal space is produced as an exclusive space from which other poorer or socially different groups may be excluded.

See also: Ecological Fallacy; Gated Communities/Privatopias; Gentrification; Segregation; Segregation Indices; Suburbanization.

Further Reading

Atkinson, R. and Flint, J. (2004). Fortress UK? Gated communities, the spatial revolt of the elites and time space trajectories of segregation. *Housing Studies* 19(6), 875 892.

Butler, T. (2003). *London Calling: The Middle Classes and the Re Making of Inner London*. London: Berg.

Caldeira, T. (2000). *City of Walls: Crime, Segregation, and Citizenship in Sao Paulo*. Berkeley, CA: University of California Press.

Dorling, D. and Rees, P. (2003). A nation still dividing: The British census and social polarisation 1971 2001. *Environment and Planning A* 35, 1287 1313.

Gorard, S. and Taylor, C. (2002). What is segregation? A comparison of measures in terms of strong and weak compositional invariance. *Sociology* 36(4), 875 895.

Graham, S. and Marvin, G. (2001). *Splintering Urbanism: Networked Infrastructures, Technological Mobilities and the Urban Condition*. London: Routledge.

Jargowsky, P. (1997). *Poverty and Place: Ghettos, Barrios, and the American City*. New York: Russell Sage Foundation.

Massey, D. S. and Denton, N. A. (1993). *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, MA: Harvard University Press.

Musterd, S. and Ostendorf, W. (eds.) (1998). *Urban Segregation and the Welfare State: Inequality and Exclusion in Western Cities*. London: Routledge.

Peach, C. (1996). Good segregation, bad segregation. *Planning Perspectives* 11, 379 398.

Wilson, W. J. (1996). *When Work Disappears: The World of the Urban Poor*. New York: Knopf.

Relevant Websites

http://enceladus.isr.umich.edu/race/calculate.html
Calculation Formula for Segregation Measures.

http://www.censusscope.org/segregation.html
Racial Segregation Statistics for US Cities and Metropolitan Areas.

Selection Bias

T. J. Cooke, University of Connecticut, Storrs, CT, USA

Glossary

Heckman Two-Step Model A statistical procedure used to address problems of selection bias in regression analysis.

Sample-Selection Bias A type of selection bias that occurs when sample data are gathered in a nonrandom manner.

Selection Bias A problem of statistical inference when estimates are made from a nonrandomly selected sample of the population.

Self-Selection Bias A type of selection bias that occurs when behavioral processes sort individuals into nonrandom samples.

Treatment Effects Model An extension of the Heckman two-step model which provides information on the effects of unchosen alternatives on measured outcomes.

Introduction

Selection bias is an important concern for quantitative human geography. There are two main types of selection bias: self selection bias and sample selection bias. Sample selection bias may occur when sample data are gathered in a nonrandom manner. Self selection bias may occur when behavioral processes sort individuals into nonrandom samples. The essential problem of selection bias is that statistical inferences made from nonrandomly generated samples are biased estimates of the population. In human geography, the most important source of selection bias is from residential mobility and migration. Both processes are highly dependent upon self sorting and thus any comparison of populations across space is likely affected by selection bias. Thus, selection bias is a ubiquitous problem in human geography but one that is both frequently misunderstood and underappreciated.

Selection Bias in Regression Modeling

Selection bias occurs in many types of geographic data analysis but is most frequently thought of in terms of the regression analysis of individual level data. Selection bias is formally thought of in the following manner. Define a sample regression equation as

$$y_i = b_0 + b_1 x_i + u_i \qquad [1]$$

In the presence of both sample and self selection, the estimates of eqn [1] are biased.

Assume that selection into the sample is determined by a variable, z_i, which is a function of an observed variable, w_i, and an error term e_i (i.e., an unobserved variable):

$$z_i = b_0 + b_1 w_i + e_i \qquad [2]$$

In normal circumstances, z_i is not observed but rather an individual is selected into the sample if z_i exceeds some threshold (c). Therefore, the selection process is defined as follows:

$$Z_i = 1 \quad \text{if } z_i > c \qquad [3]$$

$$Z_i = 0 \quad \text{if } z_i \leq c$$

In the presence of selection the correct sample regression function is

$$y_i = b_0 + b_1 x_i + b_2 \lambda_i + \hat{u}_i \qquad [4]$$

where λ_i is the inverse mills ratio (defined below in eqn [10]) and reflects, as shown below, the probability that an observation is included in the sample as a function of w_i. Viewed from this perspective, selection bias is an example of missing variable bias because in eqn [1] λ_i is missing.

Examples of Selection Bias in Regression Models

Two examples demonstrate the impact of selection bias on regression analysis results. Both examples rely upon a randomly generated sample of 10 000 observations: e_i and u_i are randomly generated from a bivariate standard normal distribution with a correlation of 0.45, and v_i, w_i, and x_i are randomly generated from a standard normal distribution. Following the previous logic, z_i, and y_i, are defined as follows:

$$z_i = w_i + e_i \qquad [5]$$

$$y_i = x_i + u_i \qquad [6]$$

Selection is determined by the value of z_i, and OLS regression is used to estimate the parameter associated with x_i under various levels of selection which are displayed in **Figures 1** and **2**.

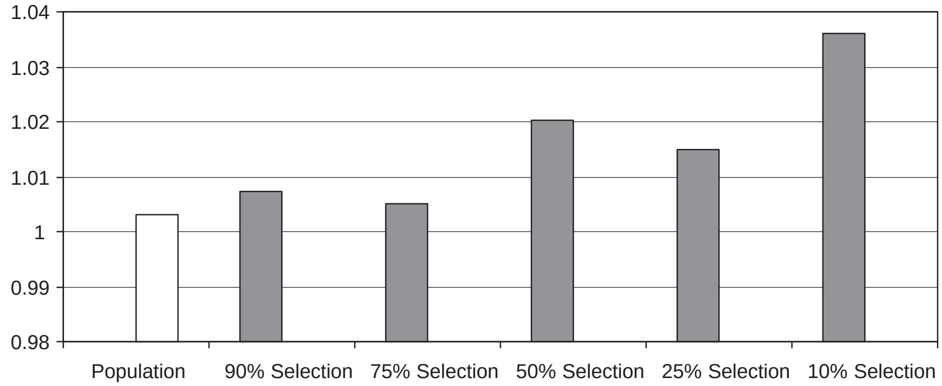

Figure 1 Selection bias in bivariate regression.

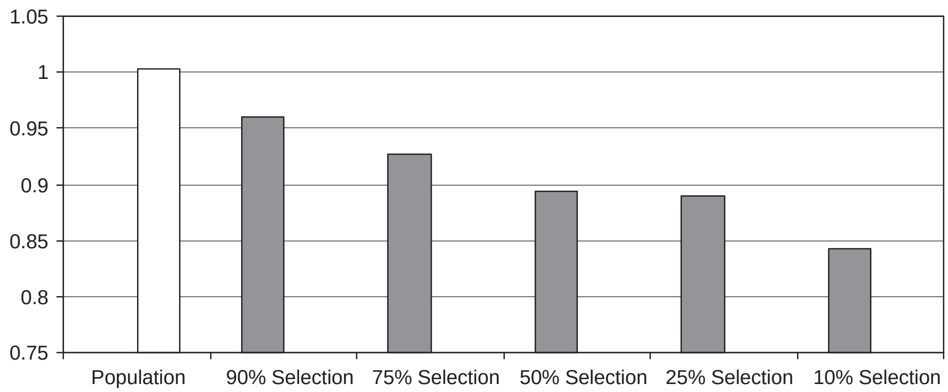

Figure 2 Selection bias, example 2.

The first estimate in **Figure 1** is for the entire sample. This serves as the reference for selection bias since it is, in effect, the population value. The second estimate in the figure is calculated by estimating eqn [6] but only for observations where z_i exceeds the 10th percentile, the third estimate in the figure is then calculated by estimating eqn [6] but only for observations where z_i exceeds the 25th percentile, and so on. **Figure 1** demonstrates that as the degree of selection moves from very low levels of selection to very high levels of selection, the degree of bias increases. However, in this example the degree of selection bias is limited. The population value is 1.003 and at the most extreme level of selection, when only 10% of the population is selected into the sample, the parameter estimate increases only to 1.036.

The second example introduces a bit more complexity by first including x_i in the selection process and second by adding another random component (v_i) to y_i:

$$z_i = w_i + x_i + e_i \quad [7]$$

$$y_i = x_i + v_i + u_i \quad [8]$$

Figure 2 follows the logic of **Figure 1**. Here, however, the degree of bias becomes much more meaningful with an increase in the level of selection. In this example, the population parameter of the effect of x_i on y_i is 1.003. Even when only 10% of the population is eliminated from the sample, the estimate of 0.960 displays meaningful bias. When only 10% of the sample is selected, the parameter estimate is reduced to 0.843.

These two examples point out the most misunderstood issues of selection bias. In both cases, the source of the bias is the correlation between the unobserved values of the selection process (eqn [7] in the second example) and the unobserved values of the regression model (eqn [8] in the second example). Thus, selection bias cannot merely be solved by including additional independent variables in the regression model and the bias affects all of the parameters of the regression model, not just those associated with the selection process.

Addressing Selection Bias in Regression Models

The standard solution to selection bias in this context is the Heckman two step model. The procedure follows the above discussion very closely. In the first step a probit model of the probability of being selected into the sample is estimated. Using the second example (eqns [7] and [8]), this would mean estimating the following probit model:

$$\text{Prob}(Z_i = 1) = b_0 + b_1 w_i + b_2 x_i + e_i \qquad [9]$$

and using the predicted values from this model to calculate

$$\lambda_i = \varphi(\gamma w_i)/\Phi(\gamma w_i) \qquad [10]$$

where $\varphi(\gamma w_i)$ is the probability density function and $\Phi(\gamma w_i)$ is the cumulative probability function of being selected into the sample as estimated by the probit model. This is the inverse mills ratio and is effectively a function of the probability of being selected. The second step of the Heckman two step estimation procedure is to include λ in the selected sample regression (e.g., eqn [4]).

Figure 3 shows the results of estimating the Heckman two stage model with the example used in **Figure 2**. In this example, the Heckman two step estimation procedure provides estimates which are much closer to the population parameter than the original models.

The Heckman two step estimation procedure is now commonly included in statistical software packages. Two crucial aspects to estimation are identification and the estimated correlation between e_i and u_i. First, there is considerable controversy in the social science literature as to the degree to which the probit model must include a variable which is exogenous to the regression model and the reader is urged to study this literature carefully. Second, the Heckman procedure assumes that e_i and u_i are correlated. This correlation (ρ), and its statistical significance, is usually calculated and reported by software packages. A low level of correlation or statistical insignificance could indicate that there is no need for the correction procedure itself or that the models are not properly specified. In any event, familiarity with the statistical literature and the issues that this approach presents is necessary because there are many cases where use of the Heckman approach exacerbates rather than relieves selection bias. Finally, there are alternative approaches to addressing selection bias when longitudinal or geographic data allow for the use of fixed effects models.

Selection Bias and Geographic Research

Selection bias has been presented as a problem but when the process is well understood, it can provide some very sophisticated insights. The most important use of selection models is in program evaluation because the Heckman two step approach effectively controls for selection bias by including information in the model about alternatives choices. For example, in geography it has been applied to migration research. Migration is thought to contribute to success in the labor market because it enables migrants to select a destination which is the best match with their abilities and skills. Migration is also thought of as a means for accumulating experience and skills. Therefore, migrants are frequently found to have higher incomes and higher status occupations than nonmigrants. However, migrants are also self selected. Migrants differ from nonmigrants in any number of both observable (e.g., age, education, and occupation) and, importantly, unobservable ways (innate ability, delayed gratification, etc.). The problem of self selection bias therefore becomes a question of whether differences in income, for example, between migrants and nonmigrants are due to the benefits of migration itself or due to differences between migrants and nonmigrants. Even more importantly, because self selection occurs with respect to unobserved factors, attempts to take into account differences in age, education, or occupation between migrants and nonmigrants are not sufficient to correct for self selection bias. In this situation, one extension of the Heckman two step model that has been used is to estimate a probit model to calculate λ_i but rather than just estimating the second model among migrants it includes a dummy variable indicating whether or not the individual was a migrant or not (i.e., a treatment effects

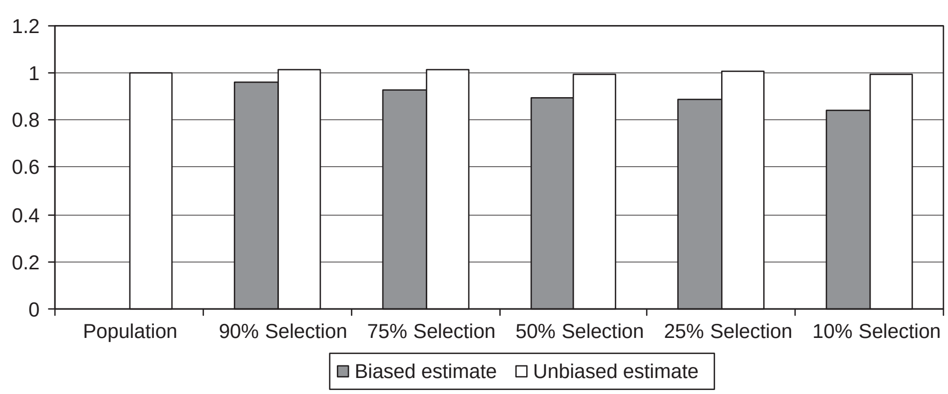

Figure 3 Heckman two-step procedure.

model). The resulting estimates allow for the accurate estimation of the actual effect of moving by comparing how well the same individual would have done had they moved or not. This provides the most accurate estimate of the actual effect of moving. This example demon strates that by fully understanding the sources of selec tion bias and how selection bias can be addressed, geographers can not only estimate more accurate models but also answer questions that were not previously pos sible to answer.

See also: Migration; Statistics, Inferential; Statistics, Overview; Quantitative Data; Quantitative Methodologies.

Further Reading

Bailey, A. J. and Cooke, T. J. (1998). Family migration and employment: The importance of migration history and gender. *International Regional Science Review* 21(2), 99 118.

Cooke, T. J. and Bailey, A. J. (1996). Family migration and the employment of married women and men. *Economic Geography* 72(1), 38 48.

Cooke, T. J. and Ross, S. (1999). Sample selection bias in models of commuting time. *Urban Studies* 36(9), 1597 1611.

Greene, W. H. (1993). *Econometric Analysis*. Englewood Cliffs, NJ: Prentice Hall.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica* 47(1), 153 161.

Stolzenberg, R. M. and Relles, D. A. (1997). Tools for intuition about sample selection bias and its correction. *American Sociological Review* 62(3), 494 507.

Self-Other

D. Sibley, University of Leeds, Leeds, UK

Glossary

Liminal Spaces Spaces which are in between, undecided, resisting categorization and often contested.

Orientalism The term generally associated with the work of the Palestinian-American writer, Edward Said, who used it to describe the misrepresentation and stereotyping of the Middle East by western, colonial powers.

Psychogeography The term is given various interpretations but used here to indicate imagined geographies, fantasies which have a real bearing on their experience of place.

Introduction

Self–other is a dualism that has been used widely in the humanities and social sciences to convey the idea of domination, exploitation, denigration, and dehumanization. If the self is the subject, the other is the nonperson, the nonagent. Self and other may not be so starkly opposed as this in discourses of othering but the construction of an individual or group as other generally reflects a power asymmetry and involves negative stereotyping. In representing someone or some group as other, there is an implicit assumption on the part of those doing the labeling that they are whole and 'normal'. The other is in some respect imperfect and deviant. In this article, ways in which the self–other dualism has been incorporated in geographical analyses of group relations, how the unconscious process of 'othering' might be related to the construction of social and cultural spaces and the material world at different spatial scales, and the implications of the argument for the future of group relations, are considered.

Studies of Otherness in the Social Sciences

Although self–other is concerned with interpersonal relations, it is aspects of relations between groups that have been the principal concern of social science research and where the concept of the other has been used to signify domination, marginalization, or exclusion. The list of othered groups includes racialized minorities, women, gays, and lesbians, people stigmatized because of mental illness or learning disabilities, and so on. At the largest scale, we can see the problem affecting global geopolitics as, for example, in the characterization of Russians during the Cold War as Asiatic, implying something less than civilized, and the stereotyping associated with orientalism which still infects western perceptions of the Middle East and contributes to current conflicts in this region.

The process of othering has been recognized and challenged in feminist and postcolonial writing in critical human geography, as in other social sciences, although it could be argued that a discipline like social anthropology, where much early research involved colonial encounters, demonstrated an awareness of the problem of othering rather earlier than geography where intensive engagement with other cultures has been less common. However, using different vocabularies, the analysis of anxieties and conflicts associated with the othering of social groups is part of a fairly long tradition in the social sciences, particularly since the 1960s. Thus, Erving Goffman's work on stigma in relation to mental illness and Stanley Cohen's analyses of moral panics triggered by supposedly deviant youth cultures were centrally concerned with self–other relations. Cohen's writing was particularly important because he was able to demonstrate how dominant societies were able to maintain their moral boundaries by amplifying the perceived deviance of groups of young people – such as Mods and Rockers in England in the early 1960s and exaggerating the threat posed by adolescent bust ups in the weakly regulated liminal spaces of English coastal resorts. He identified the popular press as a particularly potent agent in maintaining the boundaries of the normal and stereotyping the other. The recurrence of the 'moral panic', with the othered group, or 'folk devil', changing from one period to another but with the contours of the panic being remarkably consistent, suggested that there were fundamental social and mental processes generating tensions and conflicts. What Cohen's analysis points to is the transgressive quality of the other, its actual or potential for boundary crossing which threatens the integrity of the self, and this can be a problem at any spatial scale.

The perspective developed by Cohen and other British sociologists was used in analyses of sociospatial relations in human geography somewhat later, probably beginning with David Sibley's studies of relationships between Gypsies and the larger society in the late 1970s and early 1980s. More recently, the othering of individuals and groups has engaged geographers in a wide range of contexts. These include Gill Valentine's writing on food and gender, exploring the implications of 'the ideal body' for sociospatial relations, and work by Paul Cloke and others on 'the rural other'. The latter provides useful critiques of popular conceptions of the rural community and helps our understanding of the treatment of minority groups in rural areas in Britain. Most of this writing, like Stanley Cohen's, has been concerned with representations of groups as other – in literature, media reports, conversation, and so on. These studies of discursive representations are akin to Edward Said's writing on orientalism – the stereotypical representations of the east in western discourses. However, we could also look at the process of othering as a problem of the psyche. Some psychoanalysts, such as Robert Young, have argued that anxieties about others are a necessary consequence of the way in which the self develops and discursive representations are a reflection of psychic processes.

Self–Other and the Unconscious

The self–other dyad is fundamentally concerned with mental states. Thus, we should explore further some psychoanalytical arguments in order to understand why individuals and groups have a need for an other. The writing of Melanie Klein is particularly suggestive. Although Klein's theorizing was based on observations of behavior in early infancy, she came to the conclusion that the primitive emotions of infancy were retained in adulthood. She argued that we can never escape the primitive. Essentially, Klein saw libidinal and aggressive impulses as fundamental to experiencing oneself as 'good' (both loved and loving) and 'bad' (both hated and destructive). In early infancy, these polarized feelings are derived from the sense of the 'good breast' and the 'bad breast', the initial sources of pleasure and pain (when the breast is not available) and, in Klein's thesis, they provide the enduring basis of our feelings of anxiety.

The first state of anxiety she labeled the 'paranoid–schizoid', where 'paranoid' refers to a persecutory anxiety, a fear of some invasive malevolence coming from the outside, and 'schizoid' describes the defense against persecutory anxiety, entailing the sharp separation of the good ('the good breast') from the bad ('the bad breast'). In adulthood, this defense against anxiety, often described by modern Kleinians as 'psychotic', is similarly manifest in the clear, unambiguous separation of the good and the bad. Thus, the primitive emotions of early infancy provide a patterning of adult emotions in circumstances where a person senses a threat to the integrity of the self emanating from the outside. Psychotic anxiety becomes a part of social experience through the process of 'projection' whereby elements of 'the good self' and 'the bad self' are projected onto others. Projection was a term initially used by Freud to describe the phantasized (unconscious) or fantasized (conscious) location of unwanted aspects of the self in others. Klein took this further in suggesting that a 'bad self' or 'good self' could be located in an other. The individual thus keeps a connection with the other through a process of what Klein called 'projective identification'. For example, a person who has deep anxieties about their own sexual feelings may defend against their psychotic anxiety through hostility to what they perceive as threatening sexualities. Collective, social and cultural, constructions of the bad and the good can be traced back to primitive anxieties rooted in the unconscious but these anxieties are maintained and reinforced through the 'introjection' of good and bad stereotypes, that is, these stereotypes are continuously feeding in to the unconscious.

Klein also recognized a developing awareness in infants about the pain of the mother and then a growing range of others which provided the basis of a sense of guilt (for causing pain or damage to the child's own loved objects through her own destructiveness) and anxiety about others' suffering, maturing into a desire for reparation and making connections with others. This she described as 'depressive anxiety'. Kleinians would argue that with socialization, primitive anxieties become less pronounced mental states. They do not govern behavior but, still, we oscillate between psychotic and depressive positions, reverting to the primitive defense against psychotic anxiety when we feel threatened by a transgressive other, some people exhibiting this tendency more than others. Some psychoanalysts, following John Steiner, have argued that people with serious personality disorders may defend against deep anxiety by phantasizing about a 'psychic retreat'. This could be a literal space – a house, a cave, a desert – or it could be some kind of collective which provides a strong sense of security, like a religious cult. Steiner's construction of this defensive geography of the unconscious leads nicely to a consideration of the ways in which psychotic anxieties, in particular, may contribute to the creation of geographies of otherness and, conversely, how social and material geographies can exacerbate or lessen anxieties about transgressive otherness.

Some Geographies of Self–Other

Fragmentation and separation characterize residential space in many societies. This is an expression of power to secure territory, the capacity to move widely across space, and uneasiness about the real or imagined mobility of those others who are placed elsewhere, deemed 'not to belong' and who might disturb the order and harmony of privileged enclaves. This problem has been associated primarily with cities in the most developed societies. In Chicago in the late nineteenth century, for example, the othered groups who were causing anxiety for the affluent suburbanites were primarily recent Italian immigrants, often stereotyped as anarchists, and current anxieties in urban America are still largely associated with racialized minorities and, occasionally, the mentally ill. The same pattern has developed in Latin American cities, however, where the material defenses against transgressive others – razor wire, security guards, guard dogs, and so on – are increasingly evident. What this suggests is, first, that individual projections of bad objects onto stereotyped deviant others are also collective; second, that the spatial structuring of cities, including a range of collective defenses, like gated communities, and family or individual defenses, like panic rooms, are confirming and deepening self–other differences. In other words, the defense against anxiety becomes a part of the problem. We could suggest that, metaphorically, the problem of social conflict and distanciation expressed in the self–other dualism reflects a decline of touch and a shift toward an increasingly capsular society. For the more privileged in developed and some developing societies, life is experienced within a number of capsules which insulate them from others – homes, retail spaces, workplaces, and cars. The car, in particular, is a capsule that allows extensive movement without the intimate experience of others, as well as connecting a number of other well secured capsular spaces. The prospects for better social relations are clearly not good. This is not just because there is a lack of engagement with othered groups, but also because the attempts by the privileged to defend against psychotic anxieties through securing their material spaces are bound to fail. Sigmund Freud, in his reflections on 'the uncanny', recognized that the desire to acquire a sense of secure homeliness (the *heimlich*) only encouraged the return of repressed emotions, unsettling feelings (the *unheimlich*). Similarly, the French psychoanalyst, Julia Kristeva, identified abject sensations associated with bodily residues as a source of anxiety, or through elision, connected with others who constituted bad objects, 'imperfect people' who, like body hair and sweat, are always hovering and will return. Thus, we can conclude that attempts to insulate spaces against the imagined incursions of others actually ensure that anxiety about sociospatial relations is maintained.

Self–Other Ambiguity

In other contexts, the self–other question appears more complex. An enduring problem in some societies has been the conflict between sedentary populations and nomads. In the most developed societies, sedentarism is 'normal', confirmed by the legal system, and nomads are characteristically identified as discrepant and transgressive. However, in European societies, Gypsies and Roma, have been romanticized as cultures at one with nature and with a remarkable capacity for survival. In this idealization, in psychoanalytical terms, Gypsies are projected as a good object, or a benign other. They represent something that settled society has lost and, as a good object, they express a desire on the part of sedentary selves to reconnect with nature. The consequence of this, however, is that the nomads who disturb the order of settled society, who are constructed as discrepant, as 'dirt', are compared unfavorably with their romanticized selves. In racist discourse, the fantasized romantic Gypsy is 'real', and the Gypsies actually encountered are confirmed as bad objects and deviant others through comparison with this mythical benign other. It could be argued more generally that this demonstrates the ambivalence of the other, that both good and bad are projected onto others. They do not have singular identities, although in the mix of repulsion and desire, it is the former that is dominant in the shaping of sociospatial relations.

The Others of Others

In a critical assessment of the use of the self–other binary to represent processes of stereotyping, distanciation, and boundary formation, it is important to recognize those who are othered in relation to dominant groups in society not as passive recipients of the label, other, but as people whose own social relationships and use of space are affected by anxieties triggered by an array of bad objects. Thus, for the writer bell hooks, as a black girl growing up in the American South, the white people and the white part of town constituted a terror. For the children of recent immigrants living on a run down council estate in North London, there may be a whole range of people and spaces beyond the confines of the flat or the tower block onto whom their anxieties are projected. This psychogeography can seriously limit their use of the larger spaces of the city.

Understanding self–other relations presents a challenge for human geography. If an appreciation of the ways in which people relate to each other spatially, as individuals, groups, cultures, and nations, is affected by unconscious and conscious anxieties and desires, it is essential for the researcher to engage with others – to listen, to converse, and to spend time with people. This is not easy. It requires commitment and this kind of involvement raises ethical questions about privacy, intrusion, and producing knowledge that may not necessarily serve the interests of the group. However, if othering is a process of dehumanization, to understand the distortions and misrepresentations, we have to get involved.

See also: Orientalism; Psychoanalysis.

Further Reading

Cohen, S. (1972). *Folk Devils and Moral Panics*. London: MacGibbon and Kee.
Lucey, H. and Reay, D. (2000). Social class and the psyche. *Soundings* 15, 139 154.
Nast, H. (2000). Mapping the unconscious: Racism and the oedipal family. *Annals of the Association of American Geographers* 90, 215 255.
Segal, J. (1996). *Melanie Klein*. London: Sage Publications.
Segal, H. (1995). Hiroshima, the Gulf War, and after. In Elliot, A. & Frosh, S. (eds.) *Psychoanalysis in Contexts: Paths between Theory and Modern Culture*, pp 191 204. London: Routledge.
Sibley, D. (1995). *Geographies of Exclusion*. London: Routledge.
Wilton, R. (1998). The constitution of difference: Space and psyche in landscapes of exclusion. *Geoforum* 29, 73 85.
Young, R. (1994). *Mental Space*. London: Process Press.

Relevant Websites

http://www.learnerautonomy.org/kumar2008.pdf
B. Kumaravadivelu, 'The other side of othering'.
http://www.mythosandlogos.com/Klein.html
Robert Young, 'Primitive space: psychotic anxieties'.
http://www.rae.com.pt/Freud1.pdf
Sigmund Freud's 1919 essay, 'The Uncanny'.
http://www.onlineclassroom.tv/sociology/catalogue/classic collection/stan cohen folk devils and moral panics
This includes an interview with Stan Cohen.

Semiotics

K. E. Foote, University of Colorado at Boulder, Boulder, CO, USA
M. Azaryahu, University of Haifa, Haifa, Israel

Glossary

Code An interpretive system in the form of rules and conventions that allows us to encode and decode the messages signs communicate.

Connotation The implied, associated, or figurative meanings of a sign often using tropes or stemming from metaphorical, metonymical, or synecdochical relationships among signs.

Denotation The explicit, literal, commonsense meaning of a sign.

Myth The implied connotative message (cultural values, social norms, ideology) that is associated with the denotative message of the sign. This association leads to naturalization: the connoted system of values appears as a self-evident fact or denotation when, in fact, the meaning is only implied or figurative.

Polysemy The condition when a sign has many meanings as well as the likelihood that different people associate the signifier with a different signified.

Sign The basic unit of semiotics, the sign is composed of a signifier (the carrier of meaning, the form or expression, the vehicle of signification) and the signified (the idea or mental concept it represents, the content). The icon (or emblem), index, and symbol are three common types of sign.

Signification The meaning of the sign as established by the relationship between the signifier and the signified.

Text A system of signs. Originally used to distinguish written or spoken sign systems, the term text is now used widely to refer to written and visual sign systems such as architecture, material culture, elements of popular culture, the visual arts, and other realms of symbolism and communication.

Building upon the work of the linguist Saussure and the logician Peirce, semiotics (semiology) is the study of signs and sign systems as modes of communication. Semiotics is a branch of the study of culture that explores the communication of meanings and how messages are encoded and decoded. Semiotic theory has been employed in geography, though not as often as in nearby fields such as social and cultural anthropology, ethnology, sociology, sociolinguistics, social and critical theory, tourism studies, art and architectural theory, and literary and cultural studies. The influence on geography has, however, been relatively indirect. Saussure's ideas about the fluid and dynamic character of human communicational systems are foundational to structuralism, as well as textual, postmodern, literary, and critical theory. Geographers working in these theoretical areas are heirs to Saussure's theory, as can be seen in Duncan's analysis of the 'city as text'; Olsson's writings on language, knowledge, and power; recent methodological writings by Laurier and Doel on the limits of text and transcriptions in conversation analysis; and Richardson's approaches to material culture and performance. Peirce's framing of semiotics has been less influential in geography than Saussure's, but is fundamental to pragmatic philosophy, logic, and mathematics. The indirect borrowing of semiotic concepts does, however, have its limitations as noted by Sitwell, Foote, and others, in that geographers are not always aware of some of the nuances and ambiguities of semiotic theory. This article focuses on the semiotic concepts and their points of contact most relevant to human geography.

Key Concepts

The basic unit of a system of signification and hence of semiotic analysis is the sign (symbol, index, icon). The sign is composed of a signifier (the carrier of meaning, the form or expression of the sign) and the signified (the mental concept or content evoked by the sign). Yet a sign is more than a correspondence between a signifier and a signified – it is an expression recognized by society as correlated with content. This correlation is established by a code (rule, custom, or convention) that, as a shared interpretive framework, is used to encode and decode messages. Social, moral, political, architectural, scientific, and esthetic codes invest signs with meaning and facilitate the understanding of their messages.

From a semiotic perspective, the meaning of a sign consists of both denotation and connotation, and a distinction is made between a denotative and connotative signified. Perhaps more attention has been focused on issues of connotative meanings because they express, in Geertz's terms, complex 'webs of significance' linking cultural values, social norms, political ideologies, and patterns of language use and communication. Explorations of these webs of significance are fundamental to the 'cultural turn' in human geography over the past 20 years. In his *Mythologies*, Barthes also drew attention to the fact that connoted meanings which may seem self evident, even obvious, are often subtle constructions expressing social conventions and power relations. Bahktin's notion of dialogic thinking has also been influential in drawing attention to the way the meaning of words and things are dynamic and relational and are constantly being shaped by how they are used in the past, present, and future; that language is a process of endless re descriptions of the world. All of these ideas extend Saussure's argument that the relation between signifier and signified is always arbitrary and somewhat unstable. Any system of signification (text, language, gesture, etc.) involves interpretation and negotiation, is almost always in flux and transformation, and is very much influenced by the social and material context in which the signifier is used.

The prominent use of terms like text and reading is evidence of the linguistic roots of much semiotic theory. Beyond literary texts, films and works of art that are products of high culture, semiotic inquiry extended to include the reading of texts of popular culture and importantly, such that belong to the sphere of lived experience, where meanings are often embedded in social practices, cultural norms, and specific modes of behavior.

For critics, the bias in favor of language as a prototype reflects the tendency to privilege language as the natural vehicle through which meaning is communicated. This reflects the linguistic fallacy, according to which it follows that since all languages are systems of signs, all systems of signs are languages. In fact, the *a priori*, nonempirical nature of much semiotics theory has also been questioned extensively by linguists, sociolinguists, cognitive scientists, and researchers in nonverbal communication. Substantial differences exist among oral, written, visual, nonverbal, performative, and symbolic communicational systems so that casting them all as text or viewing them as language like, fails to account for their own intrinsic dynamics. The bias in favor of representational systems like language also obscures the fact that semiotic analysis can be effectively applied to performative, ritual, and nonverbal systems which are enacted rather than produced, a point which links semiotics to nonrepresentational theory as developed by Thrift and others.

However, taken at a more abstract level, semiotic inquiry offers insight into meanings and meaning making, although different theoretical orientations can be discerned. In structural theory, semiotics provides a very useful way of conceptualizing how individuals and collectivities communicate and interact in both denotative and connotative terms. Clear relationships can be seen between semiotics and Blumer's symbolic interactionism, Turner's comparative symbology, and Goffman's frame analysis as well as to comparable themes in the works of Cassirer, Douglas, Durkheim, Firth, Greimas, Langer, Morris, Rossi Landi, Sebeok, Simmel, Singer, and Warner. For structural semioticians, much attention focuses on cross cultural commonalities of communicational behavior, for example, the use in many cultures of binary oppositions such as inside/outside, sacred/profane, culture/nature, female/male. Structural semiotics tends to be synchronic and static, with little interest in change through time and the social and historical dynamic of cultural systems.

Semiotics has received perhaps more attention within postmodernist theory because it provides tools for deconstructing communicational behavior to reveal ex plicit and implicit power relationships and genealogies of authority and social control. Much work in semiotics in recent decades has called attention to the fact that 'sig nifieds' are transient and signification is not about de finitive meanings. Of primary importance in this connection is the notion of polysemy: signs and symbols have different meanings for different users and con sumers, and these meanings may be incongruent and even conflict with each other. Moreover, new signifieds are constantly produced by groups and individuals in the context of social action and everyday life and that control of communicational systems offers control of social relationships. This postmodernist dimension of semiotics has been employed effectively to draw critical attention to ways in which gender, sexuality, ethnicity, and race are embedded in communicational systems.

The notion of polysemy also underlies social semi otics, a subfield which focuses the multiplicity of inter pretations implicit in communicational behavior and how they relate to historical and social context. Social semiotics is less concerned with the formal relationships between signs. Its main focus is on semiotic procedures that underlie cultural signification in specific social contexts and in relation to particular communities: sig nification is not only a semiotic procedure in the formal sense but, importantly, also a social process. The emphasis is on meaning making in the context of social formation. It employs textual and discourse analysis and draws on social and critical theory. Based on the premise that the political economy of signs and social formation are interrelated, social semiotics explores ideology as a cultural form and myths as focal points of culture (ideology). It investigates the sociopolitical dimension of signs employed in meaning making with a special interest on patterns and representations of everyday life.

Since it offers important insights into, and analytical tools for the study of, communication, symbolism, and ideology, semiotics and social semiotics are especially relevant for cultural geography. Not only do cultural geographers have a long standing interest in these issues, but they can make contributions to semiotic theory by drawing attention to the spatial, geographical, and material dimensions of geographical texts and objects. These extend beyond the familiar literary (fiction and documentary) and cinematic texts that are intentionally produced and contain a fixed set of signifiers to also include landscapes and landscape features, the built environment and architecture, toponymy, and represen tational (symbolic) texts such as maps and works of art.

Architecture and the Semiotics of Objects

Architectural objects are constitutive elements of the cultural landscape and have received more attention in the semiotics literature than most other phenomena of interest to geographers. The analysis of architectural styles has been a long established field of study where the distinction between form and function has figured prominently. A semiotic approach to architecture and to objects of use in general emphasizes the symbolic aspects of architectural features and how they encode and communicate meanings. In his treatment of the semiotics of architecture, Eco maintains that architectural objects denote a certain function and connote a certain ideology about this func tion. He distinguishes between two levels of signification: denotations of utilities and symbolic connotations that refer to all other modes of communication associated with an architectural object. The distinction is between primary functions, such that are denoted (utilitarian functions) and a complex set of secondary functions, which are connoted (symbolic functions). Sidney's Opera House is not only a place where operas are performed, but also an archi tectural icon of Australia. The popular uses of the building extended its intended function. Importantly, the dis tinction between primary and secondary functions does not imply hierarchical positioning but rather the second ary functions are supported by utilitarian functions.

This semiotic model takes into account history in that both primary (denoted) and secondary (connoted) functions may vary in time. The cultural history of architectural objects is an account of the dynamic interplay between different primary and secondary functions. This history reveals the extent to which the symbolic functions of architectural objects are anchored in social contexts and embedded into cultural (ideo logical) codes. The utilitarian function of water towers is to serve the water supply system but, in Israel, the structures have assumed different symbolic and nostalgic meanings, once abandoned.

The City as Symbol and Text

The semiotics of architecture refers to buildings as objects of use. A different issue is urban semantics in general: the city as a system of signification, as an ensemble of object messages to be deciphered, and, ultimately, as a text or sometimes as cosmographical representations of religious or cultural values as raised in Wheatley's work on the traditional Chinese city. Another possibility based on recent theories of telecommunica tions and cyberspace is a communicational or message city: a city that functions in order to communicate rather than a city that communicates in order to function. The notion of city as a symbol or as text draws attention to the built environment and material objects in particular as a complex system of signification. It suggests the city as writing, while the user of the city, who moves about the city, is a reader engaged in decoding the messages inscribed unto the built environment. Reading the city is embedded into urban experience; the readings actualized by the users of the city are fragmentary and partial, each a permutation of parts of the text, conducted in the space–time coordinates of everyday life.

Such ambitious approaches – as in the work of Duncan in *The City as Text* – have to take into account the fact that the text has different authors and, unlike literary texts, reading is not necessarily intentional. Concurrently, it is questionable whether reading and use of the city are necessarily synonymous. The notion of the city as a text ignores aspects of urban experience that cannot be reduced to signs. Further, a reading of the city has to take into account that the city – apart from some instances where urban centers grew up around ceremonial complexes – does not have a single and def inite meaning but comprises a vast number of signifiers that communicate different meanings to different audi ences, depending on personal experiences and filtered through social, ethnic, generational, and gender prisms.

Reading Landscapes: Ordinary, Elite, and Symbolic

The notion of the landscape as a system of signification suggests that landscapes are texts that are subjected to readings. Another issue is whether landscape is written, namely, if the meanings read are intentionally inscribed. This topic has gained the most attention in studies of ordinary, common, or folk landscapes – those designed or constructed without the help of professional architects, designers, or builders. These studies follow in the humanistic tradition of Jackson, Stilgoe, Meinig, Tuan, Jordan, Jakle, and Ford, where the focus is on what commonplace elements of everyday life – houses, barns, lawns, graves, parking lots, or fences – signify about those who built them.

The notion of reading suggests that beyond the reference to the same physical objects, reading of the landscape is about interpretations and the production of meaning. Meanings are not simply 'out there': different readings of the landscape represent the cultural per spectives and concerns, and ideological dispositions and points of view of the readers. Thus, landscape can be read as nature, pristine and pure; as history, where physical landscape is but the external expression of the drama of history; as a human artifact that is imprinted with the signature of human interference, with humans as creators who redesign the physical environment for the benefit of society; as a place that has a particular character and in this sense it is unique.

On the level of denotative signification, landscape is a configuration of physical features and objects. On the level of connotative signification, landscape represents and concretizes cultural values, social relations, and structures of power. Since landscape is not only familiar but also self evident, ostensibly an aspect of the natural order of things in the world, the social relations and ideologies inscribed in the landscape are being natural ized. As experienced in the context of everyday life, the ideological dimension of the landscape is taken for granted. As an act of interpretation, a semiotic analysis of the landscape and the notion of a critical reading of the landscape as a sociopolitical text also involve making explicit that which is implicit by decoding the ideological messages of the landscape.

In contrast to the attention focused on reading ordinary landscapes, geographers have applied semiotic concepts to interpreting elite or privileged landscapes and those that are created – often explicitly – and pre served to serve as emblems of power and authority. This work has been pursued on many fronts in widely varied landscapes. The work of James and Nancy Duncan has focused on elites and the politics of the esthetics in the American suburb, Domosh has considered the connota tive meanings of skyscrapers in the nineteenth and twentieth century US, and some of Cosgrove's early research concentrated on the sixteenth century land scapes of Venice and the Veneto. Cosgrove's icono graphical approach – a methodology of the fine and visual arts – underscores his approach to elite landscapes. These and many other works affirm the important role that elite and powerful groups often play in shaping and controlling meaning within the semiotic process. For this reason, elite landscapes continue to attract considerable attention, including the planning of inter national and national expositions, major urban buildings, monuments, public art and civic plans, and debate over historical preservation.

The Semiotics of Popular Culture

The semiotic perspective has been employed often to analyze elements of popular culture – the beach, the shopping mall, the tourist site, and the theme park such as Disneyland. A semiotic reading entails decoding the meanings of signs and an exploration of how signs align with specific myths, concepts, and themes.

These sites, such as beaches, have gained such attention because they are often positioned at interfaces between public and private, commercial and non commercial, and nature and culture. Reading the beach, for example, involves an analysis of the cultural codes and social norms that govern modes of behavior as well as patterns of meaning that are inscribed into the local architecture. It also involves an exploration of how uses of the beach signify myths of popular culture and the conflicts that arise when different ideologies compete to gain control over beach uses.

Shopping malls on the other hand are focal points of the ideology of consumerism. Their primary function as retail spaces is concealed by design forms that sustain and support the appeal of the goods offered for sale. The stark distinction between exterior and interior signifies that though being a quasi public space, the shopping mall is a protected and regulated space that serves as a destination which upon entering becomes a haven of consumption. The spatial layout of the mall is directed to achieve an optimal efficiency. The iconography of shop ping malls abounds with signifiers including corporate emblems and logos as well as those intended to connote popular traditions and notions of community.

Tourism has also attracted semiotic attention because a main concern of the tourism industry is the transfor mation of sites into sights. From a semiotic perspective, an attraction is a sign that involves the relationship between markers (signifiers) and a sight (signified). The marker includes the information about the sight: its name, outline map, interpretive brochures and guide books, advertisements, postcards, and plaques. When recognized as a tourist attraction, the sight becomes indistinguishable from its markers to the extent that markers are transformed into sights: the information about a site and the site itself become entwined as a full fledged attraction (sign). Possibly a sight is a marker of another sight: The Eiffel Tower is a popular sight that serves as an iconic representation (marker) of Paris.

Of all the tourist sites which have been studied, theme parks have perhaps received the most attention. As contrived environments, theme parks are of particular interest because of the way they are consciously con structed as a set of signifiers grouped around a unifying and explicit theme. Explicitly stated, the theme belongs to the plane of denotation. On another level are socially shared ideologies, values, and myths that, as connotative codes, provide for the popular appeal of the park as a tourist attraction.

Maps and Meanings

From a semiotic perspective, a map is a cartographic text that comprises words and graphic signs. The iconic aspect is essential to understanding the function of the map as a source of information. Maps as selective representations of reality impose certain interpretations of the world which are underlain by the cultural and ideological presuppositions of the cartographer, points reinforced in the writings of Harley, Cosgrove, Monmonier, Crampton, and Wood. A map translates certain needs and interests into a visual composition of points, lines, and areas differentiated by orientation, value, hue, placement, size, pattern (texture), and shape and annotated with words – points made explicit in Bertin's *Semiologie Graphique*. The visual organization of the information encoded in a map is an aspect of the cartographic text: graphic design combines the level of denotative signification (geography) with the level of connoted signification (ideology, values).

On the level of myth, maps are instrumental in rendering objective and hence natural, certain concepts (statehood, cultural identity) by means of a set of signifiers that are aligned with the particular concept. For their users, reading a map is about obtaining geographical information about locations, orientations, distances. Reading a map semiotically is about decoding the ideology and values that belong to the level of connotative signification. It is about exposing the mythic layers of the map and the structures of power implicit in the cartographic text.

Toponymy

Real places are located in physical space, but significantly they are also constantly created in the symbolic realm of cultural signification. Names and the nomenclatures they belong to occupy a central place in any cultural system. The shaping or reshaping of sets of toponyms is a profound example of the connection between political and cultural processes. Place names belong to the cultural and historical geographies of landscapes and territories and to the political geographies of power and dissent. From a semiotic perspective, toponyms signify social and political ideologies and myths. A semiotic approach to the study of toponyms highlights their formation in specific political and social contexts. It sheds light on the ideological dimensions of (re)naming procedures and on the belonging of place names to discourses of identity.

Place names are, for example, laden with the reputation(s) of the places they denote. The name Las Vegas both denotes a city in Nevada and conjures up its reputation as 'Sin City'. The name New York City conjures up its reputation as a vibrant metropolis that never sleeps. In some cases, place names associate events, such as Waterloo (a battle, 1814) or Yalta (a conference, 1943). The symbolic meanings associated with place names are historically dynamic; place names may acquire new connotations, while older ones may lose their power of evocation.

Place names also belong to the language of modern nationalism. In a nationalist context, the reference to places by certain names not only acknowledges these places as constituents of the physical world, but also evinces nationalist assumptions and arguments. This is especially significant when different nomenclatures compete for recognition as legitimate and definitive, as is the case with contested areas. The use of place names and the rejection of others is a feature of a conflict. For instance, designating a place as Yarzum (Spanish) or Oiartzun (Basque) is a political statement in the context of the Basque struggle for independence.

Naming and renaming places is a taxonomic measure that entails the notion of appropriation. It is a by product of exploration as well as (re)settlement activities. It is also an aspect of nation building. In Greece after 1830, Turkish, Slavic, and Italian place names were Hellenized; after 1867 it was also practiced in Hungary, and after 1918, also in the new states founded in Central and Eastern Europe. Reshaping toponyms is also an aspect of postcolonial situations, when changes of place names proclaim the rejection of colonial history by a return to precolonial names. Thus, Leopoldsville, the capital of Belgian Congo became Kinshasa, the capital of Zaire. Recently, Bombay was renamed Mumbai.

Public Memory and Commemoration

The final area in which semiotic analysis crosscuts geography is in the area of public memory and commemoration. This research focuses the ways in which discourse of the past is constructed socially and expressed materially in landscape, public memorials, and heritage sites in some form of contemporary experience – a tourist's visit to a battle site, a community's celebration of its centennial, the dedication of a new memorial, historical pageants and reenactments, and many other types of events. Memorials have gained much attention because they are instrumental in naturalizing the past they commemorate: the signified (the past) being naturalized as the signifier (the memorial) seems to be a permanent feature of the landscape. Memorials are laden with symbolic meanings and their visual prominence attracts attention, but the work of DeLyser and Hoelscher also call attention to other ways in which history is constructed and staged.

Commemorative place names have also gained attention in this area. Though the primary function of place names is to denote a location, they often also have a commemorative function which is symbolic and politically significant. The denotative signification of commemorative toponyms refers to both their utilitarian (location) and commemorative (history) functions. In respect to their commemorative function, toponyms connote a certain ideology about what should be com memorated and where. Commemorative toponyms are constantly evoked in contexts that are commonly de tached from the sphere of ideology and the realm of politics. In this symbolic capacity they naturalize the commemorated history: since they appear to be obvious, commemorative toponyms effectively implicate politics and ideology into the practices of everyday life.

The ostensibly natural and hence obvious relationship between a memorial (signifier) and history (signified) is disrupted when memorials become controversial. Sus ceptible to different and possibly contradictory interpret ations, the meanings of geographical commemorations are dynamic: their meanings are constantly reevaluated as the result of changes in cultural codes, social values, ideo logical premises, and political power relations. The com memorative history of memorials is possibly punctuated by discontinuities that evince the ideological priorities of successive regimes.

When a commemoration is identified with the ideol ogy of a former regime, the signifier is vulnerable to obliteration, as is the case when, following a regime change, monuments are pulled down and streets are renamed. Another possibility is redesigning the memorial to attune its message with the ideology of the new regime. Berlin's Neue Wache underwent three successive conversions in the twentieth century: from a memorial of the Weimar Republic taken over by the Nazi regime it was later converted into a memorial of the communist East Germany. After German re unification it was re dedicated as a national shrine of the Federal Republic of Germany.

Conclusion

Despite these overlapping themes, the contacts between semiotics and geography are likely to remain modest, for two key reasons. First, although it would not be entirely fair to say that semiotics is aspatial, the major theoretical writings of semiotics have had relatively little to say about space, location, and landscape. For semioticians, geography is just one of many disciplines which provide examples of sign systems for analysis. Although semio ticians sometimes write about phenomena of geographic interest like architecture and popular culture, the ana lyses often pay scant attention to the nuances of space and place which are of most interest to geographers. Second, a major thrust of contemporary geography has been to assert the importance of space, place, and environment in economic, political, and social theory. Semiotic theory, which is weak in its conceptualization of space, environment, and geographic patterns and processes, does not offer the sorts of theoretical and methodological tools of most interest to geographers. The result will be, as outlined in this article, close contact between semiotics and geography in some areas of cul tural and social geography, but not widespread theore tical interaction between the two fields.

See also: Cultural Geography; Dialogism (After Bakhtin); Landscape; Material Culture; Memorials and Monuments; Memory; Place Names; Street Names and Iconography; Structuralism/Structuralist Geography; Symbolic Interactionism; Symbolism, Iconography; Text and Textuality.

Further Reading

Azaryahu, M. (1996). The power of commemorative street names. *Environment and Planning D: Society and Space* 14, 311 330.
Azaryahu, M. (2001). Water towers. A study in the cultural geography of Zionist mythology. *Ecumene* 8, 317 339.
Barthes, R. (1972). *Mythologies* (trans. Lavers, A.). New York: Hill and Wang.
Cosgrove, D. E. (1984). *Social Formation and Symbolic Landscape*. London: Croom Helm.
Duncan, J. S. (1990). *The City as Text: The Politics of Landscape Interpretation in the Kandyan Kingdom*. Cambridge: Cambridge University Press.
Duncan, J. S. and Duncan, N. (1988). Re reading the landscape. *Environment and Planning D: Society and Space* 6, 117 126.
Eco, U. (1976). *A Theory of Semiotics*. Bloomington, IN: Indiana University Press.
Eco, U. (1986). Function and sign: Semiotics of architecture. In Gottdiener, M. & Lagopoulos, A. Ph. (eds.) *The City and the Sign. An Introduction to Urban Semiotics*, pp 55 86. New York: Columbia University Press.
Fiske, J., Hodge, B. and Turner, G. (1987). *Myths of Oz. Reading Australian Popular Culture*. Boston, MA: Allen and Unwin.
Foote, K. E. (1988). Object as memory: The material foundations of human semiosis. *Semiotica* 69, 243 268.
Foote, K. E. (2003). Semiotic aspects of geography. In Posner, R., Robering, K. & Sebeok, T. A. (eds.) *Semiotics: A Handbook on the Sign Theoretic Foundations of Nature and Culture*, vol. 3, pp 2636 2642. Berlin: Walter de Gruyter.
Gottdiener, M. (1995). *Postmodern Semiotics: Material Culture and the Forms of Postmodern Life*. Oxford: Blackwell.
Gottdiener, M. and Lagopoulos, A. Ph. (eds.) (1986). *The City and the Sign. An Introduction to Urban Semiotics*, pp 55 86. New York: Columbia University Press.
Hopkins, J. (1998). Signs of the post rural: Marketing myths of a symbolic countryside. *Geografiska Annaler, B* 80, 65 81.
Laurier, E. (1999). Geographies of talk: Max left a message for you. *Area* 31(1), 36 45.
MacCannell, D. (1976). *The Tourist. A New Theory of the Leisure Class*. New York: Schocken Books.
Meinig, D. W. (1979). *The Interpretation of Ordinary Landscapes: Geographical Essays*. New York: Oxford University Press.
Meltzer, E. (2002). Performing place: A hyperbolic drugstore in Wall, South Dakota. In Coleman, S. & Crang, M. (eds.) *Tourism: Between Place and Performance*, pp 160 175. Oxford: Berghahn Books.
Sitwell, O. F. G. (1981). Elements of the cultural landscape as figures of speech. *Canadian Geographer* 25(2), 167 180.
Sitwell, O. F. G. (1990). Reading urban landscapes: Review of Duncan's *'The City as Text'*, *The Semiotic Review of Books* 2(3). http://www.chass.utoronto.ca/epc/srb/srb/urban.html (accessed Jul. 2008).

Wood, D. (1992). *The Power of Maps*. New York: The Guilford Press.

Relevant Websites

http://www.chass.utoronto.ca/french/as sa/index.html
Applied Semiotics.

http://www.cspeirce.com/
Arisbe: The Peirce Gateway.

http://www.semioticon.com/
Open Semiotics Resource Center.

http://www.semioticon.com/seo/index.html
Semiotics Encyclopedia Online.

Sense of Place

K. E. Foote, University of Colorado at Boulder, Boulder, CO, USA
M. Azaryahu, University of Haifa, Haifa, Israel

Glossary

Sense of Place Emotive bonds and attachments, both positive and negative, that people develop or experience in particular locations and environments. Also used to describe the distinctiveness or unique character of particular localities and regions.

Sense of place refers to the emotive bonds and attach ments people develop or experience in particular en vironments, from the national, regional, or urban levels all the way to the personal scale of the neighborhood and home. Sense of place is also used to describe the distinctiveness or unique character of particular localities and regions which emerge from their particular histories or environmental settings. In most cases, research has concentrated on exploring the positive bonds of comfort, safety, and well being engendered by place, home, and dwelling. Affective attachments are, however, neither always strong nor always positive. Edward Relph coined the term 'placelessness' to characterize the weakened bonds of attachment to community and home produced by the forces of modernism and postmodernism and Tuan and others have explored senses of fear and dys phoria induced by some places and situations.

The concept of sense of place has served an important role in debate in human geography over the past 30 years. Although the concept had its strongest roots in Anglo American geography, it has gradually had a broader international and interdisciplinary impact. When first introduced, the concept drew attention to the often subjective nature of human environmental experiences, as well as to the perceptual and cognitive dimensions of these experiences. By focusing on these issues, researchers questioned assumptions underlying pre vailing, objective models of human behavior and agency. The result was that, over the past generation, sense of place has played an important role in reorienting some subfields of human geography by asserting the import ance of considering the subjective and perceptual dimensions of human action, as well as the value of considering geographical phenomena from humanistic perspectives. Sense of place continues to serve as a conceptual bridge among a number of subfields such as social, cultural, and behavioral geography. As a link between widely contrasting methodologies, the term is as likely to appear in humanistic studies of place, dwelling, and the existential experience of the lived world as in positivistic investigations of spatial perception and behavior.

Sense of Place as Critique

In some respects, the rise of the concept of sense of place can be viewed as a critique of the dominant trends in human geography of the 1960s and 1970s. Under the influence of the quantitative revolution in the 1950s and 1960s, human geographers started to focus largely (although not exclusively) on the theoretical analysis of functional relationships and spatial systems employing spatial analytic and positivistic epistemologies. This research valued the universal over the particular and emphasized objective, quantifiable data and abstract, geometric space. It tended to side step issues of psy chological, social, and cultural meaning and the osten sibly subjective human perceptions and experiences of spaces of everyday life. This approach emphasized space and location as objective categories. The indi viduality of places was no longer considered a key con cern of academic geography, though the distinctiveness of place was often the subject of popular culture and addressed by writers such as Barry Lopez, Jan Morris, Peter Matthiessen, John McPhee, and Patricia Solnit, journalists, and filmmakers whose concern for the qual ities of the particular and the experience of place have always found a large audience, as seen in bestsellers like *A Year in Provence* (1989) by Peter Mayle and *Under the Tuscan Sun* (2005) by Frances Mayes.

Calls for research into sense of place questioned in two key ways the assumptions of the increasingly dom inant spatial analytic approaches to human geography. First, sense of place drew attention to the perceptual worlds of individuals and to the importance of under standing the perceptual and cognitive dimensions of human decision making. The work of Gilbert White, his students, and collaborators had begun to stress the importance of these factors in understanding human response to natural disasters, and other researchers, like Kenneth Boulding and Kevin Lynch, were having a similar invigorating impact on geography with their writings on the value of considering perception, images, and imagability in the social sciences. The first wave of environmental perception research in the 1960s and 1970s affirmed in numerous studies the strong sense of attachment people often feel and express for their home environment, city, state, region, or nation. This first wave of research was so influential that the methods and insights of the field of environmental perception spread rapidly through human geography. Many geographers readily incorporated perceptual and cognitive models in their research which, in turn, led to important theoretical and applied advances across human geography, especially in the subfields of urban, social, and economic geography and in cultural and political ecology. The influence of these perceptual models continues today in many subfields of geography, though usually involving more sophisticated cognitive models than those considered in the first sense of place studies. In this way, sense of place pushed geographers to expand the scope of the positivistic methodologies they applied to questions of spatial perception and behavior.

Second, at the same time, some sense of place studies questioned this prevailing positivistic orthodoxy altogether. Radical and Marxist geographers were already questioning mainstream orthodoxies of the time, but sense of place served as a reminder of geography's long standing humanistic tradition – a concern for human interests and values; a stress on the dignity of the individual; and a focus on morality and reason. These themes had been implicit in the work of a number of first and second generation academic geographers as seen in the writings of Vidal de la Blache on *genre de vie*; J. K. Wright on *terrae incognitae* and the role of imagination in geography; as well as in the link established between Carl Sauer and the Black Mountain poets and writers influenced by Charles Olson. However, it was only later with the rise of humanistic geography and interest in sense of place that these ideas were developed more explicitly from the point of view of a broad range of philosophical approaches to human values, morals, and rationality. The term sense of place was well chosen to spearhead this twofold critique of human geography.

The term place, by itself, has a long pedigree in geography, both ancient and modern, and was used frequently in the spatial analytic literature to mean a particular region, settlement, location, or position on the surface of the Earth. By joining the objective concept of 'place' with the subjective notion of 'sense', writers such as Yi Fu Tuan, Anne Buttimer, and David Seamon were asserting the need to balance objective and subjective approaches to understanding human geographical phenomena. In his book, *The Betweenness of Place*, J. Nicholas Entrikin discusses this issue in detail and argues that the objective/subjective dichotomy is fundamental to the academic exploration of place – that geographic epistemology rests 'between' the objective and subjective approaches to knowledge. Rising interest in the 1970s and 1980s in sense of place can therefore be interpreted as attempts to bring into balance a relationship between objective and subjective philosophies and methodologies that had swung too far toward the objective extreme during the 1960s.

This counter swing was set in motion in the 1970s by explorations of phenomenology, existentialism, constructivism, transactional analysis, and other humanistic, cognitive, and hermeneutic methodologies which had not previously been used in geography. Geographers saw great potential in works such as *La Poétique de l'Espace* by Gaston Bachelard; *Body, Memory and Architecture* by Charles Moore and Kent Bloomer; *Genius Loci: Towards a Phenomenology of Architecture* by Christian Norberg Schulz; and *Interpreting Environments: Tradition, Deconstruction, Hermeneutics* by Robert Mugerauer and sought to emulate the work of these philosophers and architectural theorists. Quite quickly, place, sense of place, and placemaking became popular terms that highlighted particularities and uniqueness in the human experience of place. Interest in sense of place also resonated with a new wave of postmodern critical theory devolving from the writings of Foucault, Derrida, Baudrillard, de Certeau, Guatarri, and others which sought to explore power structures embedded in everyday life.

Applications and Insights

Writings by Buttimer, Tuan, and others were particularly influential in establishing sense of place as a term fundamental to new approaches to human geography interested in the psychological, social, and cultural aspects of the engagement of humans with the environment. In 1974, Tuan coined the term *topophilia* (meaning 'love of place') to emphasize the intimate relationship between human beings – either individuals or communities – and places to which they are attached. Tuan later suggested the perspective of experience, thereby stressing the uniqueness of places as a function of human experience rather than an inherent, objective quality of the place. Space becomes place as it acquires experiential depth.

Within contemporary cultural geography, place is about what it means to people. These meanings are expressed in terms of values, obligations, intentions, and commitments, as well as emotional and social involvements. Since people define themselves through attachment to particular places, place is important to identity. Place specific shared experiences and common notions about the past and the future of place are important to a sense of community. Recent literature in this general area by Mike Crang, Steve Pile, Tim Edensor, Veronica Della Dora, and others includes explorations of consumption, material culture, nostalgia, memory, and autobiography in everyday life. Work outside geography, particularly in anthropology and architecture, also remains important in this area including, for example, the writings of Tim

Ingold on perceptions of environment and everyday placemaking.

Sense of place is a common allusion to what makes a place distinct, special, or unique. Notably, what sense actually means is rather fuzzy. Sense is the faculty of perception of stimuli such as the sense of sight or a mental capacity such as sense of humor. It denotes a discerning awareness, as in sense of community, or rationality, as in to make sense. It signifies a feeling. Significantly it also denotes meaning and suggests signification. On the other hand there is certain confusion as to what a sense of place exactly refers to – a feeling, perception, reasoned awareness, or all of them simultaneously. This ambiguity is not necessarily a disadvantage since it opens up possibilities as to the different ways in which humans address the special and particular about a place as the result of their engagement with space and landscape. However, it also means that usage of the term is shrouded with vagueness, suggestive as it may be.

As it is commonly used, sense of place means the logic and perception of place in connection with the qualities and attributes that distinguish a place from others, give it a sense of authenticity, and induce feelings of attachment and belonging. Sense of place emphasizes the individuality of place in terms of unique personality or distinct character. It suggests a particular feel or a specific character that makes the place stand out among other places, as is the theme of Jacquelin Burgess's writings on Britain's Fens or of Margaret and John Gold's studies of Scottish national identity. These same emphases can be seen in Pierre Nora's writings on the role of place in the representation of the French national past. Sense of place is thus tied to the process of placemaking through which social, cultural, ethnic, or national groups shape the environment and landscape to leave distinct imprints of their ways of life. This involves considering how people shape the environment to reflect their social identities and tastes symbolically in the built environment of cities and suburbs and the landscape design of gardens, parks, yards, and homes. David Lowenthal, Hugh Prince, Nancy Duncan, and James Duncan have all been instrumental in calling attention to symbols of identity and landscape tastes such as these.

For some, sense of place refers to specific characteristics that are an intrinsic quality of the place, an objective trait that is 'out there' and only has to be experienced in order for it to be perceived for what it is. In this sense of the term, sense of place is closely related to the 'spirit of place', or *genius loci*, or even to practices of geomancy and divination like *ilm al raml* (Arabic) and *feng shui* (Chinese). In Roman lore, *genius loci* was the protective spirit of a place, and hence belonged to the place itself irrespective of human beings. As a modern metaphor, *genius loci* does not refer to any supernatural entity, but to the essence of the place as evinced in its particular qualities and distinct character.

For others, sense of place consists of feelings and perceptions. Accordingly, sense of place is a function of human sensitivity and cognition rather than an inherent quality of the place itself: it is a relationship between humans and space and locations that is determined by experiences and evinces forms of knowledge on both individual and communal levels. In this meaning it is a foundational term of human geography that is concerned with the psychological and cultural aspects of the engagement of humans with the environment.

The concept of a sense of place emphasizes the particularity and uniqueness of place. In this it is the opposite of placelessness. As coined by Relph, placelessness is the result of the erosion or fracturing of the coherence of place since the mid nineteenth century as the result of homogenization and standardization, consumption and commodification, and last but not least: the influence of international patterns of planning and architectural styles which gradually effaced regional and local differences in style. Placelessness means the loss of individuality, distinctiveness, and authenticity. Moreover, the erosion of sense of place also entails a decline of a sense of community, since place based communal attachments become geographically fragmented to the effect that the basic association of place and community is severed.

Although Tuan and others have also noted that dysphoria, even fear, can be elements of place experience, it was Relph's idea of placelessness that captured the attention of many researchers. If sense of place was built upon the direct and full engagement of all senses, placelessness seemed very much tied to the rise of mediated environmental experiences, such as those mediated through literature, film, television, music, photography and, mostly recently, the Internet, and other network technologies. Joshua Meyerowitz's *No Sense of Place*, for example, argued that television had altered the balance between public and private and, in so doing, blurred traditional age, gender, role distinctions, and social boundaries resulting in tremendous social change. This happens because television undermines the clear spatial boundaries among distinct social and personal places by splitting them apart and reassembling them almost at random in narrative combinations which would be impossible without using the editing and splicing capabilities of electronic media.

These ideas have been taken up by geographers who have, in part, argued that media can certainly disrupt a sense of place, as well as easily create new types of experiences. Paul Adam's notion of television as a gathering place, Stuart Aitken's idea that the cinematic image event can challenge views to question everyday expectations and myths, or Jacquelin Burgess's ideas about capturing regional identity in film all point to ways in which sense of place can actually be enhanced through mediated experiences. Indeed, one of the areas in which sense of place has had an important impact is the geographical study of literature, film, and television where scholars like Douglas Pocock and J. Douglas Porteus have drawn attention to the rich senses of place created by writers and artists.

Conclusion

Sense of place continues to be an important concept in contemporary human geography. Because the concept weaves together the varied concerns of perception, cognition, individual behavior, cultural value, symbolism, and meaning, it provides a point of intradisciplinary contact, even unity within today's pluralistic discipline. For example, in the recent *Geography in America at the Dawn of the Twenty first Century* edited by Gary Gaile and Cort Willmott, sense of place or human attachment to place was discussed in chapters reviewing developments in eight subfields of US geography: aging and the aged; American ethnic geography; American Indian geography; cultural geography; environmental perception and behavioral geography; historical geography; population geography; and recreation, tourism, and sport. Of equal importance is the role sense of place has played in sustaining humanistic traditions within geography. This has involved adding to the philosophical and methodological richness of the discipline through the introduction of phenomenology, post phenomenology, existentialism, and even post humanistic epistemologies, as well as expanding the range of topics geographers have sought to explore in the arts, humanities, and social sciences.

Since the 1990s, however, the interest in place as a subject matter of geographical and anthropological inquiry has shifted from sense of place as an inherent quality of place to placemaking procedures, processes, and rituals. Placemaking is about measures undertaken by interested parties to invest place with specific cultural characteristics. The shift from sense of place to placemaking reflects the growing academic interest in issues pertaining to social construction and cultural formation. Yet placemaking and sense of place are different perspectives rather than different issues altogether: placemaking is about production, whereas sense of place is the end result.

Placemaking underlies campaigns to promote cities, where the manipulation of urban images in the context of public relations campaigns serves promotional purposes. Intended to market the city, such campaigns employ branding strategies to associate the city with positive images.

Underlying the transformation of places into destinations, placemaking is also a concern of the tourism industry. Since destinations compete with each other for the attention of potential tourists, they must be differentiated, which also entails focusing on the special and even unique. From the perspective of the tourism industry, placemaking assumes the form of place promotion by means of advertisements and guidebooks, travel reports in newspapers, and TV documentaries. By specifying the distinct character of a destination and the unique experiences it offers, placemaking associates places with a particular sense of place. In this respect, sense of place, albeit contrived and manipulated, is an asset that can be converted into profit.

Sense of place is also relevant to the discourse and analysis of globalization, in particular since the disappearance of vernacular landscapes and a sense of placelessness are an inevitable by product of homogenization and standardization of cultural landscapes. The erosion of sense of place and the possible obliteration of traditional features and meanings of place as the result of globalization may lead to conflicts where concerns about traditional senses of place are set against the pressures exerted by globalization.

See also: Behavioral Geography; Existentialism/Existential Geography; Humanism/Humanistic Geography; Landscape; Landscape Perception; Phenomenology/Phenomenological Geography; Place; Positivism/Positivist Geography; Posthumanism/Posthumanistic Geographies; Post-Phenomenology/Post-Phenomenological Geographies; Sensorium; Space I.

Further Reading

Buttimer, A. and Seamon, D. (eds.) (1980). *The Human Experience of Space and Place*. London: Croom Helm.

Casey, E. S. (1997). *The Fate of Place: A Philosophical History*. Berkeley, CA: University of California Press.

Cresswell, T. (2004). *Place: A Short Introduction*. Malden, MA: Blackwell.

DeLyser, D. (2005). *Ramona Memories: Tourism and the Shaping of Southern California*. Minneapolis, MN: University of Minnesota Press.

Entrikin, J. N. (1991). *The Betweenness of Place: Towards a Geography of Modernity*. Baltimore, MD: Johns Hopkins University Press.

Meinig, D. W. (ed.) (1979). *The Interpretation of Ordinary Landscapes: Geographical Essays*. New York: Oxford University Press.

Meyrowitz, J. (1985). *No Sense of Place: The Impact of Electronic Media on Social Behavior*. New York: Oxford University Press.

Olwig, K. (1984). *Nature's Ideological Landscape: A Literary and Geographic Perspective on Its Development and Preservation on Denmark's Jutland Heath*. Boston, MA: Allen & Unwin.

Pocock, D. C. D. (ed.) (1981). *Humanistic Geography and Literature: Essays on Experience of Place*. Totowa, NJ: Barnes and Noble.

Porteus, J. D. (1990). *Landscapes of the Mind: Worlds of Sense and Metaphor*. Toronto, ON: University of Toronto Press.

Relph, E. (1976). *Place and Placelessness*. London: Pion.

Seamon, D. and Mugerauer, R. (eds.) (1985). *Dwelling, Place, and Environment: Towards a Phenomenology of Person and World*. Dordrecht: Nijhoff.

Stein, H. F. (1987). *Developmental Time, Cultural Space: Studies in Psychogeography*. Norman, OK: University of Oklahoma Press.

Tuan, Y. (1974). *Topophilia: A Study of Environmental Perception, Attitudes, and Values*. Englewood Cliffs, NJ: Prentice Hall.

Sensorium

D. McCormack, Oxford University Centre for the Environment, Oxford, UK

Glossary

Haptic Of, or relating to, the sense of touch.
Ocularcentrism The privileging of vision as a mode of sensory engagement.
Olfactory Of, or relating, to the sense of smell.
Sensorium Originally referring to the physiological seat of sensation in the brain, but now often used to designate a historically and culturally specific configuration of the senses.

Introduction

The relation between sense and space can be understood in at least two ways. To begin, and most obviously, the senses refer to those familiar perceptual registers of everyday experience such as vision, smell, hearing, touch, and taste through which space and place are experienced. So in this respect the senses are integral to the practice and production of geographical knowledge, particularly if we think of such knowledge as an activity involving some kind of empirical encounter with the world, perhaps as part of fieldwork. At the same time, and second, sense can also be taken as the outcome of these encounters and the disciplinary practices and technologies with which they are enabled. Indeed, whether it is everyday and unformalized (think of using a map to orientate yourself around an unfamiliar city) or academic and formalized, the practice and production of geographical knowledge is often understood and described as a process of making sense.

If sense can be both the raw 'data' presented by perceptual faculties and the product of knowledge practices, defining the term is complicated further by the fact that while it might seem to be dependent solely on the physiological capacities of the body, the experience of spatial sensing and sense making is not independent of the social and cultural context within which it is enacted. Consider, for instance, the degree to which touch is or is not socially acceptable either with strangers, or even with loved ones if in public. The kind of implicit spatial regulation of such sensory intense conduct is reflective of the fact that this conduct is caught up in particular sets of culturally inflected dispositions to act which may vary, in subtle ways, across different historical and cultural contexts. If the spaces and practices of touch vary, so do the ways in which we learn to see the world, a point that can be illustrated through the example of the artistic representation of place, particularly landscape painting. Such representations are not just objective depictions of material space that reveal landscape 'as it really is'. They also rely upon certain techniques and technologies of visual perception, the most significant of which is perspective. Yet, while perspective seems to have naturalized a particular way of seeing, the culturally embedded conventions upon have been challenged by a whole range of artistic which it is based developments, each of which has sought to infuse representations of landscape with the multisensory nature of place.

Any discussion of the relation between human geography and the senses must therefore consider not only the different sensory registers through which space and place are experienced and made meaningful. It must also attend to wider debates about the role of the senses in geographical knowledge production, and the socio cultural–technological contexts within which such debates are articulated. Recent decades have seen human geographers engage with the complexity of both, prompted by developments within the discipline and by a wider expansion of interest in the senses across the social sciences and humanities. On one level, this engagement has involved efforts to map out the spatialities of familiar sensory registers such as smell, sound, touch, and, to a lesser extent, taste. Such efforts have also been extended to the consideration of how spatial experience is articulated through kinesthesia, or the sense of movement. But geographers have also begun to inquire with much greater theoretical rigor into the very conditions that allow sense making to take place, paying particular attention to the body as a site of spatiosensory involvement within the world, albeit one caught up in the techniques and technologies of wider sensoria.

Situating Vision

Any consideration of the question of the senses in human geographical knowledge must address the traditional primacy placed within the discipline upon vision. Indeed, one of the most important reasons for the growth in interest in questions of sense has been an increased critical attention to the central role played by vision in geographical knowledge. The ocularcentrism of the discipline is most obviously manifest in the simple fact that geographers have and continue to employ a range of visually based representational devices – maps, diagrams, sketches, photographs, and more recently, PowerPoint slides. This reliance upon the visual registers of sensory perception is of course not disciplinary specific, but reflects a wider privileging of vision within Western traditions and practices of knowledge production. The reasons for the wider dominance of vision are complex, but can be distilled into at least three factors. In the first place, vision seems to support and facilitate a kind of higher level thinking in which cognition is conceived primarily in terms of the transmission and manipulation of image like representations. At the same time, and second, vision, more than the other sensory registers, seems to offer the most transparent and durable mode of conveying information about the world – after all, in common parlance we tend to equate seeing with believing. Finally, of all the senses, vision appears most detached from the influence of the body and the unwieldy energies and forces it contains, particularly passions, emotions, and feelings.

Taken at face value then, vision presents itself to us as the sensory register most closely aligned with the production of truthful, objective, and disembodied knowledge. However, recent scholarship across the social sciences, including human geography, has sought to explain the primacy of vision within Western knowledge practices in terms of emergence and reproduction of historically and geographically specific arrays of sense – or what we can call sensoria. Part of this work has involved exploring how understandings of the relation between different sensory registers have shifted over time, of which the relation between vision and touch is perhaps most notable. Throughout the seventeenth and eighteenth centuries, touch actually served as a model for visual perception in the work of philosophers such as Berkeley, Descartes, and Diderot. In turn, this alignment of touch and vision was also predicated upon a specific understanding of space as a kind of geometrical container whose contents could be mapped with a degree of stability and predictability. However, this alignment broke down with the emergence, in the nineteenth century, of a much more fluid and mobile visual economy, particularly one caught up in the accelerating tendencies of capitalism. In the process, and as part of a more rigid separation of the senses in the nineteenth century, vision became detached from touch, and assumed an abstract status divorced from the sensuous embodiment of the observer.

Such critical contextualizing of vision has proved of particular value for geographers. Rather than offering a space of transparent truth, vision has been recast as part of discursive practices that define the limits of what kind of knowledge and geographies matter. This critical reworking of vision has also been accompanied by a greater appreciation of how the kinds of knowledge underpinned by vision are implicated in the production and reproduction of various forms of power – economic, colonial, and geopolitical. At the same time, geographers have sought to explore how the tendency to place primacy upon vision and to render subordinate the other senses is implicated in the mundane practices – such as fieldwork – through which geographical knowledge is produced and reproduced. This tendency is nevertheless loaded with a certain tension, in which the primacy of vision is layered with the pleasures of more than visual sensory encounters with the field, the prospect of which was famously captured by John Kirkland Wright in a 1946 presidential address to the Association of American Geographers: "In the course of field work or on a summer holiday we have all climbed a mountain and gazed over uninhabited and unfamiliar country. In the contemplative mood that the mountaintops induce, we have brooded over the view, speculated on the lay of the land, experienced a pleasurable sense of the mysterious – perhaps even felt a touch of the sinister. We have heard the Sirens' voices."

There is much in this brief passage. Most obviously, it places vision at center of the fieldwork – and indeed geographical experience. At the same time, it seems to suggest that the sensory elements of encounters with landscape are generative of a distinctive, if also a somewhat disturbing sense of pleasure – one that owes much to the kind of aesthetic sensibility associated with romanticism. Crucially, the way in which Wright signals the disquieting associations of this pleasure is by invoking other sensory registers – touch and sound, the latter of which is given a distinctively gendered inflection. The underlying suggestion here is that while it may well be pleasurable, the sound of the Siren's voices threatens to lead geographical knowledge astray, to cloud its clarity and transparency. Wright's configuration of the relation between the senses and geographical knowledge is hardly surprising. It associates the nonvisual with a realm of sensory experience not easily accommodated within the epistemological framework underpinning geography's claims to be a natural and social science. This framework appeals to a form of rational, abstract thinking impervious to the influences and vagaries of extraneous sense experience. As Wright suggests, in the terms of such a framework, senses other than vision are at best part of the fringe benefits of geography, and at worst, serve to adulterate the logical, abstract rationality of geographical thought.

Humanistic Geographies and the Place of Sense

A second important influence on the renewed interest in the senses has emerged from a critique of precisely these abstract tendencies of human geography. The idea and ideal of geography as a kind of rational and objective form of knowledge was perhaps best exemplified in the emergence and disciplinary promise of spatial science in the 1960s. One of the most significant critiques of spatial science concerned the fact that it conjured up an abstract world populated by equally abstract individuals. This was a conception of space in which little if any account was taken of the fact that everyday life involved embodied sensory engagements and encounters. This conception of space became the focus of extensive critique in the writing of humanistic geographers throughout the 1970s and early 1980s. For these geographers, everyday life could not be reduced to such abstract terms – rather it was better apprehended in terms of the dynamic experience of 'sense of place', given consistency by multisensory modes of attachment, inhabitation, and involvement through which places became meaningful. In this respect the senses had a crucial role to play. Rather than beginning with an abstract observer, geographical knowledge began with the experiential, sensing body. Also, rather than the passive container for a rational mind, the body became refigured as the corporeal vehicle for human encounters with the environment.

While humanistic geography opened up a space for the cultivation of geography through the proximate spatial relations of what James Porteous called 'intimate sensing', its contribution needs some qualification. In the first place, much of this work tended to avoid direct engagement with the relations and structures of power that shape sensory experiences of place. It is not enough to reassert and reaffirm the multisensory experience of place: attention also needs to be paid to how the privileging of different sensory registers is implicated in the politics of knowledge. Sense and sense of place are differentiated through discourses that affirm certain experiences (explicitly or implicitly) as masculine or feminine, rational or nonrational, white or non white, etc. Furthermore, and second, as its name suggested, underwriting the experiential phenomenology of humanistic geography was the possibility of developing a fully 'human' geography. While this aim may have provided an important corrective to many of the problems of spatial science, it also reaffirmed a view of the human subject as coherent and centered. As such, it can be critiqued on the basis that it fails to recognize adequately the fact that the geographies of sense and sense making are the product of historically and spatially contingent processes involving a range of more than human agencies.

These qualifications notwithstanding, the significant influence of humanistic geography can be discerned in one of the key works on the relation between human geography and the senses – Paul Rodaway's *Sensuous geographies: Body, sense and place*. In the context of developing an argument for cultivating geographies attuned to the multiple senses of the world, Rodaway pays particular attention to touch, smell, sound, and vision. As Rodaway observes, of these sensory registers it is perhaps smell that seems most tangential to the concerns of geographical knowledge. Unlike vision, smell proves difficult to map through the attachment of the suffix 'scape', because olfactory experiences do not present themselves in the form of views. Nor indeed are they something experienced at a distance. They are diffuse, sensate atmospheres within which the moving sensing body is situated. At the same time, smell is a crucial way in which the affective resonance of particular places are generated – chemically, imaginatively, and emotionally. Like taste, to which it is intimately related, smell can often interrupt the experience of one place with the virtual memory of another. To map spells, therefore, perhaps as part of what Rodaway calls 'olfactory geography', is not just to grapple with a spatially diffuse sensory phenomenon. It is also to engage with the ways in which temporality is folded and refolded through the repetition and difference of particular sensory experiences.

Nonrepresentational Senses of Space–Time

Precipitated by a critique of vision, and by a reaffirmation of the experiential element of space and place, the renewed interest by geographers in the senses has more recently been amplified by nonrepresentational theory. Mobilizing a diverse range of theoretical resources, this work has contributed to the rehabilitation of the senses in important ways. To begin, and most basically, nonrepresentational theory (like some of humanistic geography) reaffirms the centrality of the sensing body in theories and practices of space and place. In this work the body is conceived as a site of mutual sensorial intertwining between self and world. At the same time, nonrepresentational styles of work pay particular attention to how the activity of making sense – both in its 'everyday' and more narrowly academic inflections – involves more than the operation of cognitive and representational processes. So, to sense something – through taste, touch, etc. – is not necessarily a process akin to forming an image in the mind, recognizing this image as a distinct unit of sense data, and then responding with appropriate behavior. Sensing often happens far too rapidly for such contemplative thinking to occur. As such it exists as a processual, prereflective background to so much of what we do, a background that often only becomes visible when it breaks down.

Furthermore, while foregrounding the role played by the senses in everyday experience, nonrepresentational theories also seek to displace the kind of subject centered accounts of space that underpinned the writing of much humanistic geography and certain strands of phenomenologically influenced work. In the terms of nonrepresentational styles of work, making sense is refigured as an emergent outcome of an aggregate of corporeal, perceptual, and affective processes, many of which take place without much in the way of deliberation or contemplation. Such work also actively distributes the geographies of sense and sense making across a multiplicity of agencies and artefacts operating at a variety of scales, from the molecular to the sociotechnical. In doing so it opens up the possibility that the space of sensing is not solely the purview of the human. It is also something implicated in a range of more than human geographies, involving texts, technologies, and other forms of animate life.

Mapping the Multiple Senses of Space

Influenced by the developments outlined above, but also by recent work beyond the discipline, contemporary geographical engagements with the senses can be defined along a number of interrelated trajectories. The first is a sustained attempt to flesh out the spatialities of different sensory registers. Of these sound has arguably garnered the most attention. This is evident for instance in the reasonably well established disciplinary interest in geographies of music and in work engaging with the distinctive nature of acoustic space. Such geographies cannot simply be incorporated within the representational logics of much of human geographical research. The attempt to think through the properties and processes of acoustic space must avoid the move that would reduce it to the kind of knowing upon which human geography has traditionally been based. The challenges of thinking through the senses are as pertinent to touch, taste, and smell, as they are to sound (and indeed vision). Such challenges are also crystallized in attempts to engage with the kinesthetic element of sensory involvement within the world. On one level the relation between movement – and our sense of it – seems intrinsic to the corporeal experience of spatiality, an experience which can be apprehended through the kinesthetic energies of somatic practices such as dance which produce particular forms of affective experience through heightening the sensory capacities of the moving body. Yet things become complicated as soon as we try to think through this relation as a question of becoming rather than a matter of being. The moving body is no longer a stable sensory anchor but a site of ongoing spatial transformation. The importance of this complication is that it serves as a reminder that foregrounding the sensory movement of the body in space does not necessarily provide a shortcut to authentic, phenomenological experience. Instead, it opens up the question of sensory experience to a much wider, and more complicated set of questions about the relation between space, sensation, and movement.

A second trajectory in contemporary human geographical research on the senses revolves around the question of how different sensory registers are caught up in wider assemblages of technique and technology – what was referred to above as sensoria. Kinesthetic experience provides a useful way of tracing this, particularly insofar as it is thoroughly implicated in some of the key spatial transformations of modern life. For instance the emergence of new transport technologies such as the railway and the automobile was generative of distinctive forms of sensory experience associated with the pleasures of moving at speed through landscapes whether by car, automobile, or plane. Similarly, while not premised upon the physical movement of the body through space, the emergence of cinema at the end of the nineteenth century also worked upon both the visual and kinesthetic senses, providing a new logic and experience of movement space. More recently Nigel Thrift has argued that we are witnessing the emergence of a "new sensorium based on qualculation – which assumes a world of movement." One of the ways in which the emergence of this new sensorium can be traced is through the elaboration of forms of hapticity that extend the geographies of touch over a range of distances and degrees of intimacy.

A third, if perhaps underdeveloped trajectory along which research into the senses is moving concerns the opportunities for expanding the terrain of the ethical and political afforded by a renewed interest in this topic. The senses have already been politicized through the critique of the role of vision in human geography and its complicity with the production and reproduction of discursive power. However, any revalorization of the senses has the potential to contribute to a rethinking of the space of power in ways that supplement and challenge the primacy of ideology and discourse as political registers. As philosopher Jacques Rancière has argued, politics is intimately bound up in a certain "distribution of the sensible," where the sensible refers to that which can be apprehended by the senses. For human geographers this is of more than theoretical relevance. Indeed it provides a reminder that the production and practice of human geography is implicated thoroughly in a particular kind of distribution of the sensible – in which the visual is bound up with the denigration and partial exclusion of other sensory registers from the serious business of knowledge production. Furthermore, intervening in the politics and ethics of knowledge production, including geographical knowledge production, might also involve imagining and inventing new distributions of sense and sensibility, where sensibilities are understood as corporeal–cultural dispositions to act.

Conclusion

Despite the discussion above it would be a mistake to exaggerate the extent to which geographers have expanded the theoretical and empirical logics of thinking space to embrace and affirm the many modes of sense and sense making. Human geography remains a discipline in which vision is dominant, and it remains remarkably difficult for knowledge produced through other sensory registers to be taken seriously. Furthermore, important questions remain: about the precise conceptualization of the senses and their relation to the production of geographical knowledge; about the kind of methodo logical techniques through which the multiple logics of sense can be apprehended; and about the ethics and politics of apprehending space–time through mul tiple sensory registers. In addressing these questions, human geographers have much to gain from a greater engagement with emerging work on the senses across the social sciences, the best indication of which is perhaps the recent establishment of a journal dedicated to the area – *The Senses and Society*. This does not mean however that human geographers need to disengage from more traditional spatial concepts and problematics. Indeed, one of the most valuable aspects of a renewed interest in the senses is the opportunity it provides for geographers to revisit and revise familiar geo graphical questions such that they become part of the expanded conceptual and empirical sensorium of human geography.

See also: Affect; Becoming; Emotional Geographies; Humanism/Humanistic Geography; Soundscapes.

Further Reading

Buttimer, A. and Seaman, D. (eds.) (1980). *The human experience of space and place*. London: Croom Helm.

Crary, J. (1990). *Techniques of the observer: On vision and modernity in the 19th century*. Cambridge, MA: MIT Press.

Harrison, P. (2000). Making sense: The sensibilities of embodiment. *Environment and Planning D: Society and Space* 18(4), 497 519.

Howes, D. (ed.) (2004). *Empire of the senses: The sensual cultures reader*. London: Berg.

Paterson, M. (2006). Feel the presence: Technologies of touch and distance. *Environment and Planning D: Society and Space* 24, 691 708.

Paterson, M. (2006). Seeing with the hands, touching with the eyes: Vision, touch and the enlightenment spatial imaginary. *The Senses and Society* 1(2), 224 242.

Pocock, D. (1993). The senses in focus. *Area* 25(1), 11 16.

Porteous, J. D. (1985). Smellscape. *Progress in Human Geography* 9(3), 356 378.

Porteous, J. D. (1986). Intimate sensing. *Area* 18, 250 251.

Rodaway, P. (1994). *Sensuous geographies: Body, sense and place*. London: Routledge.

Seamon, D. (1979). *A geography of the lifeworld: Movement, rest, and encounter*. London: Croom Helm.

Smith, S. (2000). Performing the (sound)world. *Environment and Planning D: Society and Space* 18(5), 615 637.

Spinney, J. (2006). A place of sense: A kinaesthetic ethnography of cyclists on Mount Ventoux. *Environment and Planning D: Society and Space* 24, 709 732.

Thrift, N. (2004). Movement space: The changing domain of thinking resulting from the development of new kinds of spatial awareness. *Economy and Society* 33, 582 604.

Thrift, N. (2004). Summoning life. In Cloke, P., Crang, P. & Goodwin, M. (eds.) *Envisioning Human Geographies*, pp 81 103. London: Arnold.

Services, Professional

A. James, Queen Mary, University of London, London, UK

Glossary

Absorptive Capacity A firm's ability to value, assimilate, and apply new knowledge to commercial ends.

Emotional Labor Workers' deliberate manipulation of their emotions and personalities to create a pleasant experience for the customer/consumer, invariably involving the public display of an emotion that they may not necessarily feel privately.

Expatriation The relocation of workers out of the corporate core to overseas countries in order to staff branch offices as firms export their services internationally.

Externalities Sources of competitive advantage which lie outside of firms' individual boundaries, but which accrue to firms as a function of their collocation in space.

Introduction: Professional Services and the 'New Economy'

It is now widely accepted that fundamental changes within advanced capitalist economies over the last three decades are representative of a transition to a new era of capitalist development. Beginning in the early 1980s, we witnessed a major reorientation in the pattern, form, and sources of capitalist economic growth, involving – on the one hand – a sustained decline in the contribution of heavy industrial manufacturing to national economic output, employment, and trade ('deindustrialization'), combined with a dramatic growth of employment and output in service activities ('tertiarization') on the other. In large part, these changes have been ushered in by technological innovations in information and communication technologies (ICTs), developments which have also underpinned an ever increasing functional integration of economic activities at the international scale. This major reorientation of the productive structures and dominant sources of employment within advanced capitalist economies has been conceptualized in a number of different ways, in terms of a transition to: a 'post Fordist regime of accumulation'; the 'fifth kondratiev wave'; 'knowledge capitalism'; 'friction free capitalism'; 'soft capitalism'; or most commonly espoused within policy and media circles, the 'new economy'. Whatever the moniker used, all are united in their according central importance to innovation, learning, and creativity to economic success; that is, in asserting that the dominant source of wealth creation lies in the production of ideas, knowledge, and know how. In other words, physical capital and human capital are argued to have been usurped by 'knowledge capital' and 'creative capital' as the primary drivers of wealth creation and prosperity; that the future depends more on 'brains rather than brawn'; and that we are now (according to some commentators at least) 'living on thin air'.

Despite many of these claims surrounding the 'new economy' having been somewhat exaggerated, nowhere are they perhaps more valid than within the knowledge intensive 'professional services' sector, a sector which has exhibited remarkable growth over the last three decades. In professional service firms (PSFs), it is the embodied knowledge, skills, and expertise of highly educated, trained, and accredited employees that comprise firms' most valuable assets, and which are used to deliver customized solutions to clients' specific problems. Professional services are characterized, therefore, by intangible products delivered through intangible personal relationships with clients buttressed by intangible reputations, experience, and trust. In the subsequent sections of this article, the causal drivers of recent growth in professional services are sketched out, along with a discussion of the distinctive (emotional) labor process that lies at the heart of wealth creation within this sector. Two distinctive features of the emergent geographies of professional services are then analyzed in terms of their implications for knowledge production and management: namely, localized agglomeration in functionally integrated 'clusters' versus the internationalization of professional service firms. The article concludes with some possible directions for future geographical research on professional services and on employees' experiences of working in this sector.

Professional Services: How Do We Know Them When We See Them?

When geographers talk about 'professional services' they are referring both to the long established, formally regulated professions of law, accountancy, architecture, and real estate, as well as emergent 'new economy' growth sectors of advertising, business and management consultancy, market research, public relations, financial services, data processing, and computing services. These various sectors are recognized as exhibiting a number of distinctive commonalities by which 'professional service firms' might usefully be characterized as a coherent entity, centered on the nature of their workforce, the labor process, corporate ‘product’, and firm–client relationships. These unifying characteristics, then, are (1) a well paid, highly educated workforce whose entry into the labor market is controlled and sanctioned by a formal institutional body which authorizes and accredits to particular forms of professional competence achieved through prolonged training; (2) a corporate dependence on the embodied skills, idiosyncratic knowledge, cre ativity, trustworthy reputations, and social connections of staff as firms' most valuable assets; (3) wealth creation through largely autonomous employee engagement in task based, symbolic analytic problem identification, problem solving, and strategic brokering activities on behalf of clients; (4) customized ‘product’ delivery at the point of demand through interpersonal relationships with clients in which expert workers exercise a high degree of personal judgement; and (5) professional practice is governed by a code of ethics.

Scholars continue to disagree on which of these five characteristics of professional services should lie at the heart of a single overarching definition of this sector (with consequent implications for which sectors are included under this banner heading in the first place!). Thus for some scholars, it is the degree of contact and interaction with the client that ultimately defines professional services; for other scholars it is the degree of product customization; and for yet others, it is the intensity and duration of professional training that matters. At the same time, scholars also distinguish between professional service ‘firms’ (the corporate enterprises that comprise this in dustrial sector); professional service ‘occupations’ (forms of work involved in the production of professional service products); and professional service ‘functions’ (the uses that clients obtain by consuming products from professional service firms)! Despite this definitional and classificatory complexity, however, it remains a truism that ‘all’ professional services are fundamentally based on highly intangible knowledge inputs to, and outputs from, the production process. At the same time, the clients of professional service firms are typically other firms, organizations, or government departments, such that the outputs of PSFs invariably constitute intermediate inputs to the production processes of client organizations. In this way, professional services differ significantly from other (personal) service industries whose outputs are used for final consumption by private individual consumers.

On the Historical Growth and Development of Professional Services

Professional services have attracted considerable atten tion from economic geographers in recent years as they have grown in economic significance and scale of oper ations. Indeed, professional services are among the fastest growing sectors within advanced capitalist economies over the last three decades, exhibiting an average 10% growth per annum in revenues since the early 1980s, with current revenues of over US$1000 billion globally. Employment in this sector has exhibited a similar rate of growth over the same time period and, according to OECD data, professional services now account for around 20% of all employment in the US and Western Europe. This growth dynamic can be explained in a number of ways. One explanation concerns post Fordist strategies of externalization or outsourcing of many service functions previously carried out in house by large vertically inte grated manufacturing firms in the pursuit of enhanced flexibility, efficiency and competitiveness, and shared market risk. However, practices of subcontracting and externalization are not exclusive to manufacturing firms; recent years have also witnessed increased subcontracting of in house service functions by large service firms.

A second set of drivers of professional services growth concerns the ever growing demand for such expertise in relation to a new macroeconomic environment. Specif ically, as widespread state practices of de regulation have increased the uncertainty faced by many firms, this has prompted them to develop more elaborate administrative bureaucracies and engage the assistance of financial, ad vertising, sales, legal, marketing, and management experts in order to make sense of, and respond appropriately to, an increasingly complex decision making environment. Third, globalization processes facilitated by new devel opments in information and communication technologies continue to intensify the degree of international com petition between firms, further reinforcing corporate concerns to (re)assess the ways in which they produce goods, and to extend their incorporation of external professional service expertise during the pre production, during production, and post production phases.

In turn, the result of these various drivers is a pro fessional services sector that is increasingly polarized into a dual economy comprised of a small number of giant firms operating in conjunction with a mass of smaller companies. The decline of medium sized firms is popu larly explained in terms of their not possessing the economies of scale, international office network and range of expertise offered by large firms, nor the flexi bility, speed, and personal attention to customers as provided by their smaller competitors.

Understanding the Professional Services Labor Process: Embodied Work and the (Gendered) Performance of Emotional Labor

The nature of the labor process within professional service work has provided economic geographers with a very fertile research agenda, not least because it has helped force a fundamental rethinking of previous 'culture'/'economy' binaries within the discipline. In contrast to manufacturing where the production and consumption of 'hard' physical goods occurs at different points in time and space, in professional services, the spheres and production are merged. That is, production and consumption of the service occur simultaneously, such that the cultural identity, personality, appearance, embodied personality, and social characteristics of the worker become 'part of' the product consumed. Consequently, a central task within professional service occupations involves workers using their emotions and personalities (as well as their physical and intellectual capacities) to create a pleasant experience for the client. Arlie Russell Hochschild has described this phenomenon, in which workers actively and consciously manage their feelings and emotions to create a publicly observable facial and bodily display as 'emotion management'. The term 'emotional labor' refers to these activities when they are done for a wage in the public sphere. Thus, for example, the daily work activities of lawyers, solicitors, management consultants, stock brokers all involve personal interaction and face to face contact with clients.

Business success is crucially dependent on employees' presentation of self, negotiating skills, and relationship management in ongoing personal encounters, on the basis of which workers develop and maintain client portfolios. Moreover, professional service workers invariably must display publicly an emotion that they may not necessarily privately feel. As such, management of feeling shifts from being a strictly private act performed almost unconsciously, to a public act performed consciously according to others' guidelines. A number of geographical studies in the professional services sector have demonstrated how the performance of emotional labor cannot be regarded as a gender neutral phenomenon; rather, that the gendered construction of social identities fundamentally controls access to, and success within, many professional service occupations. Most notably, Linda McDowell's work on professional service work within the merchant banking sector in the city of London has demonstrated how women continue to be concentrated within secretarial and clerical grades with little opportunity for career movement into professional occupations (save for a select group of 'honorary male' female traders). This is explained in terms of a set of accepted workplace norms and everyday behaviors within this sector which emphasize a particular version of masculinity based on 'competitive spirit', 'self confidence', and 'toughness' as required attributes for workers' success. As such, dominant notions of the ideal worker are socially constructed as inherently masculine; in contrast, women are cast as 'other'.

On PSF Clustering: From Transaction Cost Reductions to Interfirm Knowledge Spillovers, Collective Learning, and Innovation

One of the most distinctive features of the geography of professional services is a tendency toward spatial agglomeration in functionally integrated 'clusters', characterized by exceptional levels of external networking and interfirm linkages. On one level, scholars have explained PSF clustering tendencies in terms of beneficial transaction cost reductions that accrue to firms as a function of their spatial proximity. This thesis is based on a threefold argument. First, instability of markets and an accelerated pace of technological change are met by a disintegration of the production process, and hence a deepening in the social division of labor between firms. This then allows firms to maximize the benefits of specialization and to minimize exposure to risks of overcapacity, labor force hoarding, and dangers of technological lock in, and hence to become more flexible. Second, as interfirm transactions become more important, so the costs of transport, communication, information exchange, search, scanning, and so forth, become more significant. The increased costs associated with increased external transactions are argued to create a 'spatial pull'; that firms will seek to agglomerate to minimize the costs of those transaction costs through external economies of scale. Third, spatial agglomeration is also argued to create specialized institutions which further lower transaction costs and hence increase economic efficiency, as labor and other factors of production are pooled among large numbers of specialist producers, and which firms would otherwise not be able to afford individually.

However, while this explanation goes some of the way to explaining why PSFs cluster, it is but part of a bigger picture. For professional service firms, competitiveness is sustained by becoming a moving target through continuous technological learning and the rapid development and commercialization of new ideas: those firms that can learn and innovate faster become more competitive because their knowledge is scarce and therefore cannot be immediately imitated or transferred to new entrants. Accordingly, scholars have been keen to examine how spatial clustering also enhances processes of knowledge creation, information dissemination, learning, and innovation on which the economic competitiveness of professional service firms is founded. A number of concrete mechanisms of knowledge transfer and spillover between professional service firms within clusters have been identified, based on the fundamental premise that innovation is an 'interactive' learning process in which complementary forms of information and knowledge are combined, to create new forms of knowledge greater than the sum of their constituent parts.

First, in addition to formal cooperation agreements between professional service firms, personal contact networks between individual employees have been widely identified as important conduits of interfirm information exchange, reinforcing more formal types of interaction. While this may mean losses of proprietary knowledge for individual firms, allowing other firms to appropriate returns on their investment, such personal contact networks also allow firms to learn new technologies, opportunities, and market challenges more quickly because of higher densities of interaction within these networks. Such informal social interactions are found both within work settings (so called 'cafeteria effects') and nonwork settings within PSF clusters, the latter being premised on a porous division between work, social and leisure activities. Employees frequently meet at industry seminars, talks, and other social activities organized by local business organizations, as well as in more informal venues such as bars, clubs, pubs, cafes, and coffee shops. In these social contexts, personal contacts are established and maintained, providing workers with access to a wider diversity of ideas and bases for comparison with ideas and practices from other firms, often leading to new, often unexpected, ideas or synergies. Indeed, intricate social networks based on personal contacts offer potentially 'the' most efficient organizational arrangement for sourcing information given that information is difficult to price in a market and difficult to communicate through a hierarchy.

Second, as communications technologies have improved the transfer of codified (formal) knowledge between professional service firms, so economic competitiveness is increasingly dependent upon their ability to access sources of new 'tacit knowledge', which is highly personal, context specific, and difficult to formalize (or as Karl Polanyi summarized it, that "we know more than we can tell"). However, tacit knowledge is difficult to transfer between firms in the absence of labor mobility, given its embodiment in individuals. It is therefore significant that scholars have documented higher rates of skilled labor mobility within professional services clusters. These interactions are advantageous at two levels: (1) acting to transfer knowledge once and for all by workers taking expertise and knowledge with them to competitor firms; or (2) establishing an 'ongoing' link between the firms via the personal relations maintained between the various employees. Either way, as new employees interact and engage with new colleagues with partially overlapping knowledges, comparisons of evolving ideas are made with other practices in the firm that are not internally generated. Thus, there is an increased potential for new unexpected ideas, interpretations, and synergies to develop; that is, for increased learning and innovation. Indeed, interfirm spillovers of embodied expertise through labor mobility are particularly significant in the professional services sector as firms seek to poach workers in order to access their personal client portfolios and established reputations, as discussed in the previous section.

In sum, continuous technological learning and innovation within professional service firms are widely recognized as highly dependent on firms' abilities to 'access' external sources of information and knowledge, abilities which are enhanced through spatial agglomeration. Fundamentally however, successful innovation and learning within professional service firms is also dependent on firms' abilities to assimilate, reconfigure, transform, and apply new information to commercial ends once it enters the firm, that is, on their 'absorptive capacities'. Different absorption rates are not random but depend upon both the social and cultural structures within firms, because the ability to absorb new knowledge within a firm will always depend on sociocultural constructions of what is acceptable and desirable. The innovation literature has thus consistently highlighted a set of cultural norms that, if widely shared by the members of a firm, actively promote the generation of new ideas and help in the implementation of new approaches. These norms include a climate of openness in which debate is encouraged; a willingness to listen to other people's ideas; creative dissent or the right of employees at all levels to challenge the status quo; and multiple advocacy, that learning requires more than one 'champion', if it is to succeed. Successful innovation within professional service firms therefore presumes a necessary relationship between learning and active employee involvement at all levels: that all employees can act as independent agents, take responsibility, experiment, and make mistakes as they learn. At the same time, the open exchange of ideas among workers also depends on all employees being able to make themselves heard. As such, the gender inequalities outlined in the previous section potentially constrain professional service firms' absorptive capacities and restrict firms' abilities to capitalize fully on the skillsets of female employees.

Internationalization of Professional Services

A second distinctive feature of the emerging geography of professional service firms is the development of spatially dispersed, interdependent organizational networks at the international scale as PSFs have 'gone global'. This globalization process needs to be understood in the context of a number of causal drivers: to follow client firms overseas; to reduce risk by operating in more than one geographical marketplace; to follow other professional service firms in order to reduce their competitive advantage through imitation; and to expand operations through merger and acquisition activities. This process of PSF internationalization through international branch plant development is particularly extensive within the accountancy and management consultancy sectors, well known corporate examples including Deloitte, PricewaterhouseCoopers, Ernst and Young, and KPMG. Indeed, PriceWaterhouseCoopers is now the largest accountancy firm globally, with 130 000 employees in almost 150 countries worldwide.

However, in stark contrast to many media accounts which portray this globalization phenomenon as a relatively straightforward process, professional service firms are faced with a number of significant knowledge management challenges as a function of their capital accumulation strategies ultimately being dependent on the embodied knowledge, intricate personal networks, skills, expertise, and trustworthiness of their professionally accredited employees who engage clients in face to face interactions in order to generate fee income. Above all, how to 'stretch' locally rooted tacit knowledge and corporate memory to clients overseas?

One organizational response by British legal firms has been to expatriate partners and lawyers to export 'English common law' to their international overseas offices, while simultaneously employing the services of local staff to supply jurisdictional law. In this way, British legal firms are able to develop, manage, and diffuse idiosyncratic knowledge from the corporate core to overseas subsidiaries, as well as meeting locally specific client expectations of quality of service and working practices. Accordingly, such strategies are necessarily neither uniform nor homogenous across different professional service firms, nor indeed across all overseas branch office locations within a single professional service firm. As such, the globalization activities of professional services provide further significant evidence to critique the 'end of geography' thesis.

Future Research Directions

Research on the socioeconomic geographies of professional services continues to develop a distinctive contribution to our understanding of emerging global service networks and new international divisions of labor, as well as the new forms of work and employment which characterize knowledge intensive firms in the new economy more generally. There are a number of directions in which future research on this sector might usefully develop, not least because professional services are widely seen as a 'flagship' new economy sector, where work and employment practices are developed first before diffusing out to other industrial sectors.

First, in light of persistent structures of inequality within professional service firms which militate against female and ethnic minority workers progressing as quickly through employment hierarchies as their white male colleagues, future research should continue to dereify so called 'glass ceilings' of constraint by grounding them in everyday concrete workplace interactions, social processes, and their associated causal agents – an attribution of responsibility upon which policy actors might then act. Moreover, by demonstrating empirically the ways in which these same worker inequalities based on gender and ethnicity themselves constrain corporate innovative performance, future studies might also develop arguments for workforce diversity couched in the language of economic competitiveness that firms (and governments) particularly appreciate and understand.

Second, building on previous case studies in the legal services and advertising sectors, there exists considerable scope for exploring the role of local informal and formal institutional context in shaping the internationalization strategies of other types of professional service firms. In so doing, geographers might offer increased empirical and conceptual insights into (overcoming) the spatial limits to knowledge and learning which are likely to underpin divergent development trajectories in the emerging 'new economy'.

Finally, future work might also usefully explore the emergent operations of professional service firms outside of the USA, Western Europe and South East Asia, particularly in non English speaking countries – or in other words, outside of economic geography's domain of empirical inquiry as traditionally demarcated vis à vis development studies. In this way, new work might contribute to the nascent 'postcolonial economic geographies' research agenda which, while conceptually exciting, nevertheless continues to suffer from a rather narrow empirical evidence base.

See also: Business Services; Consumption; Corporate Responsibilities; Economies, Borderland; Economy, Informal; Ethnic Economies; Feminism and Work; Foreign Direct Investment; Governance, Corporate; Industrialization; Neighborhoods and Community; Trade, International; Tropical Geography.

Further Reading

Amin, A. and Thrift, N. J. (1992). Neo Marshallian nodes in global networks. *International Journal of Urban and Regional Research* 16, 571 587.

Ascher, B. (1993). Business and professional services: Competing in a more mobile world. In Aharoni, A. (ed.) *Coalitions and Competition: The Globalization of Professional Business Services*, pp 20 31. London: Routledge.

Beaverstock, J. V. (2004). 'Managing across borders': Knowledge management and expatriation in professional service legal firms. *Journal of Economic Geography* 4, 157–179.

Beaverstock, J. V., Smith, R. G. and Taylor, P. J. (1999). The long arm of the law: London law firms in a globalising world. *Environment and Planning A* 31, 1857–1876.

Bryson, J. R., Daniels, P. and Warf, B. (2004). *Service Worlds: People, Organizations, Technologies*. London: Routledge.

Burton Jones, A. (1999). *Knowledge Capitalism: Business, Work and Learning in the New Economy*. Oxford: Oxford University Press.

Coe, N. M. and Townsend, A. R. (1998). Debunking the myth of localized agglomerations: The development of a regionalized service economy in South East England. *Transactions of the Institute of British Geographers* 23, 385–404.

Faulconbridge, J. R. (2006). Stretching tacit knowledge beyond a local fix? Global spaces of learning in advertising professional service firms. *Journal of Economic Geography* 6, 517–540.

Hochschild, A. (1983). *The Managed Heart: Commercialisation of Human Feeling*. Berkeley, CA: University of California Press.

Kroes, N. (2005). Better Regulation of Professional Services UK Presidency Seminar Brussels, 21st November 2005. http://www.europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/05/711&format=HTML&aged=0&language=EN&guiLanguage=en (accessed Jul. 2008).

Keeble, D. and Nachum, L. (2002). Why do business service firms cluster? Small consultancies, clustering and decentralization in London and southern England. *Transactions of the Institute of British Geographers* 27, 67–90.

Leadbeater, C. (1997). *Living on Thin Air: The New Economy*. London: Viking.

Lowendhal, B. (1997). *Strategic Management of Professional Service Firms*. Copenhagen: Copenhagen Business School Press.

Lowendhal, B., Revang, O. and Fosstenlokken, S. N. (2001). Knowledge and value creation in professional service firms. *Human Relations* 54, 911–931.

McDowell, L. M. and Court, G. (1994). Performing work: Bodily representations in merchant banks. *Environment and Planning D* 12, 253–278.

Nachum, L. (1999). Measurement of productivity of professional services: An illustration on Swedish management consultancy firms. *International Journal of Operations and Production Management* 19(9), 922–949.

Warf, B. (2001). Global dimensions of US legal firms. *Professional Geographer* 53, 398–406.

Relevant Websites

http://www.sbs.ox.ac.uk
Clifford Chance Centre for the Management of Professional Service Firms, University of Oxford.

http://www.oecd.org
OECD indicators of regulatory conditions in the professional services sector, Regulatory Reform: Economic Issues, Indicators of regulatory conditions in the professional services, Economics Department.

Services, Rural

A. C. Vias, University of Connecticut, Storrs, CT, USA

Glossary

Back-Office Services Producer services traditionally found in urban centers (e.g., credit card processing), now often found in rural areas as communication technologies improve.

Big Box Stores Large-scale stores owned by national and transnational corporations that take advantage of economies of scale and scope to provide lower prices.

Distributory Services Activities oriented to moving products of the extractive (e.g., farming) and transformative (e.g., manufacturing) sectors to their markets.

High Fliers and Lone Eagles New producer services (often information services) that take advantage of telecommunications technologies to locate in rural high amenity areas.

Personal Services Activities providing services directly to households and individual consumers.

Producer Services Activities oriented to providing services to businesses and government.

Public Administration Activities of government often grouped with the service sector.

Social Services Activities supplied to businesses and households, often nonprofit in nature.

Introduction

The extraordinary growth of the service sector in general, and in rural areas specifically, is a result of a number of broad changes in the global economy. These changes include an increased demand for services as discretionary income rises in the developed nation and a lack of productivity gains often found in the primary and secondary sectors that keeps labor demand high. Furthermore, changes in the manufacturing sector have led to the outsourcing of service employment and the creation of new independent service firms, especially in the producer services. Finally, new complex manufacturing processes now frequently have a need for new specialized services such as high speed information and computer systems, needs that did not exist a decade or two ago.

The above characteristics show the multifaceted processes that drive growth and change in the service sector as a whole, and with respect to its constituent parts. This makes analyses of this part of the economy difficult because aggregate trends hide significant variations: within the sector in terms of growth; in the spatial distribution of service activities; and with respect to the impacts of services. For example, the processes driving the development of producer services are different from those in retail, and rural areas are experiencing service sector changes much different from those in urban areas. Also, producer services offer great potential for employment and income growth, but many researchers argue that only a small percentage of rural areas can take advantage of these opportunities. This means that any description of rural services needs to tease out this large section of the economy to determine the rationale for locating various types of service establishments, and to find out which parts of the service sector are winners or losers with respect to their ability to promote economic growth in rural areas.

Nature of Rural Services

Definition and Classification

In recent decades, scholars have devoted considerable effort to untangling the nature and impact of restructuring in rural services. However, services in general remain a difficult part of the economy to analyze. Depending on the focus of the research, a researcher may use different definitions and terminology to describe all or portions of the service sector. At the most rudimentary level, services are a residual that form the tertiary sector, or any employment that is not part of the primary or secondary sectors. This type of definition is overly broad, and really hides the heterogeneity that exists within the tertiary or service sector. In this approach, services can be said to include the following broadly defined sectors: transportation, communications, and public utilities; wholesale trade; retail trade; finance, insurance, and real estate; services; and public administration. A complicating factor is that the terminology used is often quite inconsistent and/or poorly defined. For example, services is the term often used to describe the entire tertiary sector, while others use the term services to describe a subset of the tertiary sector. Another problem, at least when analyzing US data, is that the classification systems have changed over time. Newer classification systems are superior in many ways because they take into account services in a more meaningful way. For example, information services are a distinct category. However, any changes in a major employment classification system make any longitudinal analyses very challenging.

Other definitions of services reallocate employment using criteria such as the market served by firms, trad ability of the good or service over space, occupational type, along with hybrid definitions that combine these criteria. For explaining prevailing trends in services, especially their role in rural areas, the hybrid scheme by Singelmann that focuses on markets served is the most useful one. In this system, four service categories are used: producer services (finance and insurance; real estate, rental and leasing; management of companies; administration); personal services (arts, entertainment, and recreation; accommodation and food services); social services (healthcare and social assistance; education; and sometime public administration); and distributory services (wholesale and retail trade; transportation and warehousing; and utilities). Additionally, the equivalents of the primary and secondary parts of the economy are called the extractive and transformative sectors.

Another issue unrelated to the various classifications has to do with defining services in rural versus urban areas. In most cases, and for understandable reasons, services have been defined in the context of urban areas, creating descriptions that often may not be appropriate for the rural economy. This issue plus the traditional problems centered around defining rural versus urban using politically defined geographic areas, such as met ropolitan areas, make any empirical analysis problematic. What's more, the low densities and numbers found in rural areas often mean that data for these areas are suppressed by government agencies worried about con fidentiality issues. Overall, the result of all these prob lems is that all analyses of rural services using data currently available require the use of multiple trade offs, sometimes leading to contradictory findings, or results that are not as reliable as scholars or policymakers would like. That said, many of the most important trends and issues surrounding rural services can still be discerned using current data sources.

Some Aggregate Trends in Rural Services

The increasing importance of the service sector in the economy of the US is easily shown with data from the Regional Economic Information System (REIS) de veloped by the Bureau of Economic Analysis. Although these are US data, many of these trends are quite applicable to other developed countries as well. **Table 1** shows the distribution of employment in 2001 and 2005 utilizing the broad sector categories described above: extractive sectors (including farming), transformative sectors (including manufacturing and construction), and services. Public administration, often included as part of services, is also shown. The data show that services, broadly defined, now comprise well over three quarters of all employment in the US. There are some issues with

Table 1 Sectoral employment in metropolitan and nonmetropolitan areas in the US as a percentage of total employment, 2001 2005

Region	*Sector*	*% Employment*	
		2001	*2005*
US	Extractive	2.9	2.7
	Transformative	16.1	14.8
	services	67.1	68.8
	Public administration	13.9	13.7
	Total	100.0	100.0
US Metro	Extractive	1.8	1.7
	Transformative	15.4	14.1
	services	69.4	70.0
	Public administration	13.4	13.3
	Total	100.0	100.0
US Nonmetro	Extractive	9.3	8.8
	Transformative	19.9	18.5
	services	54.5	56.7
	Public administration	16.3	16.0
	Total	100.0	100.0

Source: REIS 2006. (Sectors based on NAICS, 2003).

the data as presented, including the use of full and part time employment rather than full time equivalents, and the utilization of metropolitan and nonmetropolitan areas (using US Office of Management and Budget des ignations of 2003) as proxies for rural and urban areas. Nevertheless, there is little doubt about the dominance of services, and the fact that the role of services continues to grow. Perhaps more interesting is that the next two parts of **Table 1** also show why theories based on services in urban areas may provide flawed insights on services in rural areas. In nonmetropolitan areas, the extractive and transformative sectors are more important, but employ ment in the service sector is at least ten percentage points lower than that found in metropolitan areas. These structural differences have always existed, although the percentage of people employed in services in metro versus nonmetro areas is slowly converging. That said, there is likely to be significant differences in service employment between metropolitan and nonmetropolitan areas in the foreseeable future.

Besides nonmetro areas having a smaller percentage of people employed in the service than metro areas, there are significant differences where people are employed within the service sector as well. **Table 2** shows the percentage of total employment in the four primary types of services described by Singelmann: producer, distribu tive, consumer and social services, along with employ ment in public administration. Clearly, the distinguishing characteristic of services in metropolitan versus non metropolitan areas is employment in producer services, with metropolitan areas in general having twice as many people employed in the producer services. This means nonmetropolitan areas are vastly underrepresented in

Table 2 Service employment in metropolitan and nonmetropolitan areas in the US as a percentage of total employment, 2001–2005

Region	*Sector*	*% Employment*	
		2001	*2005*
US	Distributive services	18.5	18.0
	Producer services	23.6	24.5
	Personal services	13.8	14.3
	Social services	11.2	11.9
	Public administration	13.9	13.7
US Metro	Distributive services	18.6	18.1
	Producer services	25.5	26.3
	Personal services	13.0	14.4
	Social services	11.4	12.2
	Public administration	13.4	13.3
US Nonmetro	Distributive services	17.9	17.6
	Producer services	13.1	14.4
	Personal services	13.6	14.0
	Social services	9.9	10.5
	Public administration	16.3	16.0

Source: REIS 2006. (Sectors based on NAICS, 2003).

areas of the economy like finance, insurance, management, and information services. Only in public administration was employment in nonmetropolitan areas overrepresented compared to metropolitan areas.

The impact of these differences in service employment are significant for two reasons. First, over the past few decades, the producer services have been a major area of employment growth compared to other types of services. Second, and perhaps more importantly, producer services generally produce higher salaries for each worker than other types of services. In many instances, these differences in salary are very large, with salaries in many types of producer services like information and management services over twice as high as other types of services such as retail or personal services. Sadly, while the percentages of employment in services in metro versus nonmetro areas are converging, salaries between different parts of the service sector are generally diverging.

On the whole, the data show the increasing significance of the service sector in both rural and urban areas of the US. These trends are generally found in other developing nations as well. Unfortunately, there are some key differences between the service sector in rural and urban areas, with the data indicating that service sector jobs in rural areas are generally less attractive and lower paying than service jobs in urban areas.

The Geographic Rationale of Rural Services

Services and Economic Growth

For decades, services were treated as a residual, and never the central, focus of rural economic research because scholars did not believe services were a part of the export base of a community. Traditionally, it was the primary and secondary sectors of agriculture, mining, and manufacturing that formed the export base and was capable of promoting economic growth, especially in rural areas. In recent decades, the role of services in urban and rural areas has been reassessed. There is strong evidence that now suggests that several components of the service sector can act as part of the export base. That is to say, many types of service activities do not derive all their income from local sources. Additionally, new ideas on the role of services in the local economy have emerged, and geographers have increased their study of the location criteria of service establishments. Despite this reevaluation of the service sector, the research on services is often too geographically focused on urban areas. In answer to this dearth of work, there has been an increasing amount of research over the last 10–20 years that has examined services in the context of rural areas. An examination of the components of rural services displays the diversity of the sector, and why rural areas need to be examined separately from urban areas.

Underlying these structural changes in services are the changing patterns of consumption of society in general. Researchers have long analyzed consumption and its significant role in the capitalist economy. One of the most significant shifts in consumption patterns was noted by Veblen at the start of the twentieth century, in his writings on the development of a strata of society enjoying 'conspicuous consumption'. While Veblen spoke of a small fragment of the richest in society, in recent years higher incomes in developed countries have made mass consumption a way of life for a majority of the population. Rising incomes in developing countries have also increased the demand for a much greater variety of goods and services. Additionally, in recent decades, globalization processes have made consumers aware of and capable of obtaining goods from around the world. For this new large segment of society, rural areas have an allure that was rarely found in people living in large cities and the suburbs, especially in terms of environmental and cultural amenities. This applies to people simply visiting rural areas as well as those taking up permanent residence. These new patterns of consumption have led to a new demand for goods and services previously not found in rural communities. The impacts of these social changes and their role in rural services are outlined in more depth below, especially in the area of personal services.

Distributory and Personal Services

In many ways, distributory activities, as defined by Singelmann, are the hardest segment of the service sector to analyze because of the diverse jobs within the category.

Complicating issues is that most scholars actually do not view transportation, utilities, and wholesale trade as part of the service sector, and hence, little has been written about those activities, at least as it relates to services. Conversely, retail trade has been the focus of many studies, and many aspects of retail make its study as part of the service sector a valid approach. Given the orientation of retail toward consumers, analyzing retail within the context of personal services makes more sense.

Over the decades, personal services have been viewed as serving the local community, and with little export potential. Central place theory supports this idea, stating that personal services are driven by local demand and the size of the settlements, all located in a hierarchy of places. While rural areas have never had the variety of personal services found in urban areas, personal services have diversified in recent decades as rural consumers began to demand the same types of goods previously available only in the cities. Despite overall growth in personal services in rural areas, the spatial distribution of these activities has fluctuated tremendously because of the link with local demand, and hence, population. Many of the most remote rural areas, especially in less dense nations like the US and Australia (but also in parts of Europe as well), have experienced population losses along with the decline of personal service establishments and small town central business districts. This is in contrast to certain types of high amenity rural areas experiencing rapid population growth through in migration with an associated expansion and diversification of the types of personal services available to the new residents. These newcomers often have very different tastes in goods and services, and display consumption patterns more often found in large urban areas. However, as discussed later in this article, this type of growth has limited potential for most rural areas.

Underlying this spatial reconfiguration of personal services is a broad process of economic restructuring, especially in the retail sector. For example, the emergence of 'big box' stores and giant merchandisers like Walmart and K Mart, and very large international grocery chains like Tesco and Woolworths, all relying on huge economies of scale and scope, has created a difficult competitive environment for locally owned 'mom and pop' stores. For example, in many rural areas of the US, stores like Walmart locate in one service center serving surrounding rural communities and hinterlands. The net result of these new megastores is that retailers of all sorts in surrounding rural areas (any many in the same service town with the new Walmart) are not competitive with the low prices offered by the big box chains. While this decimation of small stores has been widely discussed in the popular media, it is worth noting that many locally owned stores serving distinctive market niches can actually complement big box stores in these regional service centers, provided they do not directly compete with the large chain stores. Finally, an added problem for personal services in the most rural towns, but not directly related to retail restructuring, is the loss of smaller, locally owned stores as customers drive farther and farther (as a result of better transportation infrastructure) to shop and take advantage of the greater variety of stores in larger market center malls along with the convenience of one stop shopping. This structural change has been encouraged by increasing car ownership in rural areas as rural transit services decline, especially in places like the UK, as opposed to the US where this type of long distance shopping has been common for a longer period of time.

As a result of all these processes, the geography of personal services in rural areas is highly uneven, in constant flux, and very different from that found in larger metropolitan areas. Over time, these changes at multiple spatial scales may lead to a more stable configuration of stores in rural areas, but the process is far from over.

Producer Services

Compared to personal services, producer services generally represent a growth area of the economy. Many rural areas have benefited from the movement of manufacturing firms out of urban centers in search of lower costs, bringing with them many of the producer service activities upon which manufacturers rely to produce goods. Reinforcing this trend has been another structural change forced on manufacturers trying to stay competitive, outsourcing, along with the shift of many producer service operations (e.g., many back office operations) to rural areas as a result of the new telecommunications technologies. A classic example of this phenomenon is the well studied move of Citbank's credit card processing unit to the rural Great Plains in the 1980s. Finally, the transfer to rural areas of small and individual proprietor service companies, often called 'high fliers' and 'lone eagles', also represents another example of the impact of technological change. This has been especially advantageous for rural areas with the desirable amenities often attractive to these new entrepreneurs.

Besides creating new types of establishments in rural areas, producer services are an important indirect ingredient of the rural economy by serving traditional local industries like farming. However, the dynamic aspects and growth potential of producer services, especially as described in much of the literature, may only benefit a small number of rural communities. Most often, producer services that move out of big cities end up locating on the periphery of metropolitan areas, while the most remote areas remain untouched by this growth. Perhaps more worrying is the fact that in the long run, these firms may be so geographically mobile that they end up moving completely offshore where labor costs are even lower than in rural areas, following manufacturing firms that have been offshoring their operations for decades.

Social Services

Finally, social services, which include employment in education and health services, along with public administration, are an important part of the rural economy, even more important than in urban areas. Throughout rural areas, social services are among the more stable parts of the economy in terms of employment. Unlike personal services, their stability is partly a result of the nature of many of these jobs, which are most often supported through funds from local, state, and federal sources, not the open market. In some cases, employment is actually increasing in the social services as the federal and state governments push for increased educational opportunities in rural areas, and as devolution processes push services to more geographically dispersed locations under the control of local communities. Also, as the population ages, there is an increased need for healthcare for the elderly, although it is unclear whether the funds will be available to provide these services in the most distressed rural areas. However, in the long run, neutralizing the potential for the growth and development of new social services in rural areas is their link with the size of the local population. Hence, for the many rural areas experiencing outmigration, employment losses for these types of activities are still a real possibility, if not a certainty.

There are other issues related to social services in rural communities that may be more difficult to overcome than the annual battles with budget cutting at the national and regional levels. In recent years, social services – more so than other types of services dominated by the private sector – have been affected by broader political and economic changes. These changes have taken place in the US, but seem to be even stronger in Europe and Australia. A broad neoliberal agenda has promoted less government regulation and intrusion into the economy, and the devolution of more and more functions to regional and local governments. Foucalt's ideas on govermentality help explain these new types of relationships between policy making at the national level, and local public and private institutions (like volunteer groups) that govern through 'community'. For functions such as education, healthcare, and postal services, the ideal is for decentralized government bodies to more fully 'own' and control their local social services and economic development policies, rather than rely on traditional top down approaches.

Unfortunately, the end result of these changes has often been the elimination of many types of cornerstone or iconic services in rural governments unable to support these services financially. For example, in the UK, postal services are being rationalized resulting in the loss of thousands of post offices, leading to a decline in an institution that is more than just another service outlet to handle mail and parcels. In the UK, postal outlets often serve more as a local hub for community activities and as a lifeline for many rural residents to the goods and services not available locally. While rural residents may be saddened by the loss of many types of privately owned rural services (e.g., personal services), the loss of iconic services like local schools, post offices, and medical facilities, all part of the social fabric of rural communities, has led to broad protests in the UK. In the US, these broad changes in government have led to the consolidation of school districts and rural hospitals in remote rural areas (like the Great Plains in the US) to take advantage of economies of scale, a process hastened by declining populations and improved accessibility to urban centers for many (but not all – e.g., the elderly) rural residents.

Impact of Services on Rural Areas

Underlying this entire debate on employment and growth in the service sector is the nature of the jobs available. A major issue for many types of service sector jobs is that they are often part time with lower wages and few benefits. For rural areas, this problem is even greater because low wage labor is more likely to be found in the types of services oriented to local consumer demand in rural areas (e.g., personal services, retail, and social services), as opposed to the higher paying producer service jobs most areas would prefer. Thus, even though services as a whole are growing in rural areas, the impact of this employment growth may not be very substantial. Despite this problem, many argue that any vital, vibrant, and sustainable rural community must have a range of services – including producer and personal services not reliant on external demand – that is attractive not only to current residents and businesses, but also to potential businesses and migrants who may want to move into the area. It is important to note that although many producer services generally require greater skills and generate higher paying jobs, there are segments in this area of services with low paying jobs. Unfortunately, these low paying jobs are often the types of producer services likely to locate in rural areas while higher paying producer services remain in urban areas.

Future Growth (and Decline) of Rural Services

As the service sector continues to grow significantly in the economy of developed nations, there is strong evidence that rural areas are not benefiting from this growth as much as urban areas, especially with respect to service sector jobs likely to promote independent economic growth through external/export demand (e.g., producer services). In most cases, the odds on attracting the most dynamic types of producer services to rural areas remain low. That said, there are small niches of the service economy that can take advantage of new telecommunications technologies like the internet. Additionally, services of all types still play an important support role in all rural communities, even for areas currently reliant on the extractive and transformative industries.

A good example where rural areas can develop vibrant service sector growth is for communities endowed with an abundance of environmental and cultural amenities attractive for tourism and recreation. The demand for these new types of goods and services fits quite well with the changing patterns of consumption, and the desire for increasing the numbers of people who can experience the rural environment. These types of services bring external income into rural areas, but unfortunately, are often low paying and seasonal in nature. There is much more to be said about this newly emerging area of rural service growth, but this topic is covered in more depth in another article of this encyclopedia. For other high amenity areas experiencing population growth as a result of in migration, the impact on rural services may be more beneficial. For example, certain types of producer service workers who can take advantage of improved communications technologies (e.g., lone eagles and high fliers taking advantage of the internet) may be a source of long term growth in some rural communities, although this group of employers is relatively small. There is also potential for new rural service business opportunities for residents self employed in small establishments that sell more tradi tional goods. These small entrepreneurs can often take advantage of new technologies like the internet to cater to small niche areas of the market, allowing them to sell goods previously available only locally (e.g., foods, clothing, and other commodities emblematic of the local culture). However, this also remains a small portion of the rural economy in goods and services.

While rural communities with attractive environmental and cultural amenities offer some real opportunities for growth in the service sectors, in the end there are only a limited number of rural areas with the amenities that can lead to significant long term economic growth. Addition ally, there are many rural areas in the developed nations experiencing some growth as a result of retiree migration. This is now possible because this large cohort is able to take advantage of mobile, nonemployment income sources, and can now find communities that offer the services needed by an elderly population. Unfortunately, there is also an older cohort of people found in remote rural areas with limited amenities, and where out migration is leading to a population aging in place. In this type of situation, services are often on the decline.

Finally, and perhaps the most hopeful scenario for rural services in the near future, is to focus on retaining services already in the community. These services are vital to the overall health of the community, and to other businesses. This geography of rural services is more likely to prove successful and stable in the long run, as opposed to efforts aimed at attracting high paying pro ducer services jobs (sometimes called the 'new buffalo hunt'). These sought after service establishments and jobs – although desirable in many ways – are just as likely to move on from rural areas as their spatial flexibility increases with globalization processes. In the meantime though, many of the most rural areas must contend with depopulation and out migration, a devastating situation for local demand of personal services. Furthermore, economic restructuring in service sectors like retail presents a host of other problems, even for areas main taining their population. This process often decimates locally owned businesses and creates jobs in new mega stores like Walmart, a corporation often disparaged (fairly or unfairly) for the low quality of its jobs. Clearly, socioeconomic forces at the local, national, and inter national scale are transforming the geography of rural services in a process that is far from over.

See also: Business Services; High-Tech Industry; Hinterland Development; Local Economic Development; Retail Geographies; Rural Populations; Services, Professional; Tourism, Rural.

Further Reading

Adamchak, D., Bloomquist, L., Bausman, K. and Qureshi, R. (1999). Consequences of population change for retail/wholesale sector employment in the nonmetropolitan Great Plains: 1950 1996. *Rural Sociology* 64, 99 112.

Beyers, W. and Lindahl, D. (1996). Lone eagles and high fliers in rural producer services. *Rural Development Perspective* 11, 2 10.

Bocock, R. (1999). Consumption and lifestyles. In Bryson, J., Henry, N., Keeble, D. & Martin, R. (eds.) *The Economic Geography Reader*, pp 279 290. Chichester, UK: Wiley.

Bryson, J., Daniels, P. and Warf, B. (2004). *Service Worlds: People, Organizations and Technologies*. London: Routledge.

Glasmeier, A. and Howland, M. (1995). *From Combines to Computers: Rural Services and Development in the Age of Information Technology.* Albany, NY: State University of New York Press.

Goe, W. R. (2002). Factors associated with the development of nonmetropolitan growth nodes in producers service industries, 1980 1990. *Rural Sociology* 67(3), 416 441.

Gray, D., Farrington, J., Shaw, J., Martin, S. and Roberts, D. (2001). Car dependence in rural Scotland: Transport policy, devolution and the impact of the fuel duty escalator. *Journal of Rural Studies* 17, 113 125.

Herbert Cheshire, L. (2000). Contemporary strategies for rural community development in Australia: A governmentality perspective. *Journal of Rural Studies* 16(2), 203 215.

Higgs, G. (ed.) (2003). *Rural Services and Social Exclusion*. London: Pion.

Hornbeck, J. F. (1994). *The Discount Retail Industry and Its Effects on Small Towns and Rural Communities*. Washington DC: Library of Congress, Congressional Research Services.

Kassab, C. and Luloff, A. (1993). The new buffalo hunt: Chasing the service sector. *Journal of the Community Development Society* 24(2), 175 195.

Pamiagua, A. (2002). Urban rural migration, tourism entrepreneurs and rural restructuring in Spain. *Tourism Geographies* 4(4), 349 371.

Poe, N. and Shaw, T. (2004). Exploring the current and future role of market towns in servicing their hinterlands: A case study of Alnwick in the North East of England. *Journal of Rural Studies* 20, 405 418.

Singelmann, J. (1978). *From Agriculture to Services*. Beverly Hills, CA: Sage Press.

Vias, A. C. (2004). More stores, bigger stores or no stores: Paths of retail restructuring in rural America: 1988 1999. *Journal of Rural Studies* 20(3), 303 318.

White, S., Guy, C. and Higgs, G. (1997). Changes in service provision in rural areas. Part 2: Changes in Post Office provision in Mid Wales: A GIS based Evaluation. *Journal of Rural Studies* 13, 451 465.

Sexuality

P. Howell, University of Cambridge, Cambridge, UK

Glossary

Closet Space The spatiality of the closet, which not only operates to render gay life invisible to a homophobic society, but also exists as a strategy for gay sociability and survival.

Heteronormative/Heteronormativity The institutionalized assumption that heterosexuality is normal and natural.

Performativity The queer theoretical argument that not only gender but also sexual identities are artificial, practiced, and repeatedly performed; the norm of heterosexuality is thus no more than an authorized fiction.

Phallocracy Feminist term, literally 'rule by the phallus', referring to the ruling power of men simply by virtue of their sex.

Queer Theory/Queer Studies An interdisciplinary academic/activist movement, opposed to heterosexual norms, considering all sexual identities to be socially constructed and sanctioned, and inherently unstable.

Sex/Gender System The argument that sexuality is actually an effect of gender, rather than the other way round; sex and gender are analytically distinct, but particular social formations institutionalize the ideology that gender identities are a matter of sexual orientation as well as sexual difference.

Sexology The pioneering scientific or pseudoscientific analysis of sex and sexuality, largely and uncritically based on the methods of natural science.

Sexuality A term popularized by nineteenth- and twentieth-century sexual science, referring to the forms of sexual behavior and practice, and with a particular stress on sexual types and identities.

Sexual Citizenship The relationship between individuals, as sexual beings, and the political communities of which they are a part; traditionally exclusive and based on compulsory heterosexual norms, more inclusive forms of sexual citizenship may be proposed and advanced.

Sex Work Contemporary term grouping together the variety of modern commercial sex activities, including but not confined to prostitution.

Introduction

The poet Philip Larkin famously averred that sexual intercourse began in 1963, ruefully adding "(Which was rather late for me)." From the perspective of academia, however, it came all too prematurely, as many disciplines had to wait until the late 1970s and 1980s for serious and systematic consideration of sexual themes; none more so than geography, a subject whose initial, hesitant explorations of sexuality made Larkin appear precocious. Since the 1980s, however, the analysis of sexuality has come to occupy a prominent position in academic research, certainly no longer relegated to the grubby margins. In geography, as in history, the humanities and the human sciences in general, there has been a tremendous upsurge of interest and research. Sexuality has become, in the words of the historian Jeffrey Weeks, "a continent of knowledge, with its own rules of exploration and its own expert geographers." This metaphor is a distinctly troubling one; however, the notion of taking possession of virgin epistemological land recapitulates the problems of the late nineteenth- and early twentieth-century sexologists who were the first to train their spyglasses on this new continent. We owe to this tradition the whole notion of sexuality as a distinct domain, separate from the other fields of human behavior and social order, powerful, and perhaps even determining. This is a contentious formulation, and, at the very least, all studies of 'sexuality' must be self-consciously critical of their epistemological history.

Geographies of Sexuality

Sexuality, as it emerged as a domain of knowledge, was always localized. To put it crudely, this was firstly in an inner human nature that was both biologically universal and irreducibly individual; but secondly within an equally problematic social field wherein sexuality was properly installed in the private sphere yet threatened to spill out into the promiscuous public world. The implications of the first of these spatializations would be explored by the various psychological sciences, and by Freudian psychoanalysis, in particular. Though reliant on geographical tropes (Freud famously developed the notion of female sexuality as a 'dark continent' for instance), much of this work serves to efface any material and social geography. But the second spatialization equally obviously proposes a socially situated moral geography in which sexuality is recognized as a powerful and fundamental force that requires acknowledgment, channeling, discipline, and perhaps, above all, careful spatial regulation. Sexuality here is normatively private, but is always already threateningly public in its pathological forms, in 'public' sex (like prostitution) and various other abhorred practices (such as homosexuality). As a concept, then, sexuality's geography was from the start thoroughly ambiguous, at the same time nowhere and everywhere. It is this problem that geographical studies have attempted – belatedly – to resolve.

These studies might be considered, initially at least, in terms of three related but distinct fields of enquiry: what might be called 'psychosocial', 'feminist', and 'socioeconomic' geographies of sexuality. Geographers have had relatively little to say on the psychology and psychoanalytics of sexuality, to take the first of these strands, but it is clear that such an engagement offers rather more than is indicated by the crude characterization offered above. Rather than simply a matter of discursive localization in the psyche, commentaries on Freud and perhaps more fruitfully on Jacques Lacan and his critics suggest that a more materially grounded, historically and geographically specific, psychosocial analysis of sexuality is possible. There has been some important research, for instance, drawing on psychoanalysis and developmental psychology, focusing on geographies of affect and emotion, and above all of embodiment. In such geographies the role of space and place in the production and consolidation of sexual subjectivity is considered in some detail. Some of this work still adheres to a dated and dubious universalism, for sure, but more promising approaches exist, positing for instance the negotiation of sexual identities in spaces of encounter between 'self' and 'other', and out of a dialectic of personal experience and social expectation. One criticism that might be leveled at some studies of sexuality is that there has been little enough concentration on sex, on bodily difference and sexual experience; seeing 'sex' as an inflection of wider sociospatial sexual relationships is therefore a real step forward in our understanding. At the very least these pioneering psychosocial geographies of sexuality do not prematurely foreclose the claims of either nature or nurture, and they may allow for a rapprochement between psychoanalytical and social constructionist approaches to sexuality.

The field within the discipline that grappled most insistently and most promptly with these issues, however, is feminist geography. Though initially concerned with gender at the expense of sexuality, the redirection of feminist research toward a conjoined analysis has eliminated much uncritical naturalism, and pointed to the spatial practices upon which the sex/gender system depends. Perhaps, the key move here has been the denaturalization of the so called public and private spheres, which have for so long misrepresented sexuality as purely and properly a private matter; and the marginalization of 'feminine' private space reflects the epistemological and actual exclusion of sexuality from the 'masculine' public sphere. At the same time, feminists have been suspicious of attempts to reduce sexual difference to the question of the social construction of gender alone; and without returning to biological essentialism, the importance of embodied difference for social power remains a powerful argument. Analysis of the role of sexual violence, or at least the fear of it, in underwriting men's privileged access to public as well as private space is one important example of this recognition. And even if the fiercest critiques of phallocracy and patriarchy are uncongenial, it can hardly be denied that the privileges (and injuries) of gender are intimately and politically related to the privileges and injuries – hidden or not so hidden – of sex and sexuality. The relationship between sex, gender, sexuality, and space is particularly complex, but the still largely unanswered challenge of feminist theory and practice is of central importance for understanding the geographies of sexuality. Most of the questions over the constitution of sexual identity, and the role of sexual ideology and sexual experience, or of nature versus nurture, have been rehearsed in geographical work that has had the question of sex and gender at its heart.

There is a third strand of thinking about sexuality and space that can be suggested, though. The complex relationship between capitalism and the modern order of sexuality has been raised and analyzed within social and economic analysis. At a rather abstract level, the production of sexuality can be seen as necessarily, not merely contingently, tied to modern capitalism's production of space. For one thing, though all societies need to regulate fertility and ensure the reproduction of the labor force, capitalism has been seen as having particularly insistent requirements for the control of sexuality in the service of both biological reproduction and the reproduction of workers' energies. Sexuality, in the form of the reproductive matrix, is inseparable from the historical geography of the family in the modern, capitalist West. In this analysis the normative heterosexual system, with its clearly demarcated geography of separate spheres – public and private, men's work and women's work, the male breadwinner norm, and the ideal of domesticity – is fundamental to the very survival of capitalist society. The ability of capitalism to harness and exploit sexuality in the service of profit has clear implications, not only for high theory, but also in concrete geographical studies of, *inter alia*, the sexualized workplace, sex work and sex tourism, spaces of commodified leisure and pleasure, place marketing and promotion, the creation of urban villages, and the nature of city life. It may be noted here that the last category encompasses the well recognized phenomenon of gay men's gentrification of urban space, first properly investigated in Geography by Manuel Castells. This is an indication for many not just of defensive mobilization but also of capitalist society's ability to co opt even potentially sexually subversive identities and lifestyles. The very notion of distinct sexual identities is suspect for its understanding of sexuality as a matter essentially of individual sexual consumer preference: a norm of sexual object choice increasingly authorized and perhaps even required by the dynamics of contemporary capitalism.

Space, Discourse, and Sexual Identity

These three strands of research – crudely labeled as they are – remain productive and indeed essential for the geographical analysis of sexuality. Of course, they overlap in many significant ways; they are separated out here only to gesture at the subdisciplinary emphases, out of which a geographical analysis of sexuality has emerged. These three are complemented and challenged, however, by the various projects conducted under the aegis of the 'cultural turn' in social theory. This is an imprecise phrase, but it does encompass the emphasis placed on discourse, social construction, and anti essentialist social theory. In this context, the analysis of geographical discourses of sexuality is prominent, and the influence of Foucault, as in the other social sciences, has been especially profound. Perhaps most provocatively, Foucault argued in the first volume of his *History of Sexuality* that sexuality is the product of an explosion of discursive activity in the modern West, the result of a great process of transforming sex into discourse. One major implication is that the typological classifications so characteristic of the sexological tradition are fundamentally misconceived. Instead of natural sexual identities being revealed by science, Foucault suggests how these identities are produced within discourse and within wider relations of power. Now in many ways this analysis is pre empted by the feminist critique of the sex/gender system, but the best known illustration of this argument is Foucault's discussion of male homosexuality. Foucault decisively asserted that male same sex sexual relations were transformed, at or around the end of the nineteenth century, from an understanding in which sexual acts are dominant (with the gendered roles of active/masculine and passive/feminine determining the nature of individuals' sexual deviance) to one in which sexuality is invested in types of persons (both partners, whatever their role, now become equally 'homosexual').

This remains a contentious analysis, and all too easily misconstrued (one long running and largely unproductive debate is whether we have to say that there was no 'sexuality' before the modern era), but the power of the general argument for the construction of sexuality is largely accepted. What this means for geographers, of course, is that sexuality is not something that preexists the social order and its discursive and material spatialization. Rather, sexuality is dependent upon the geopolitical work of culture and discourse. It draws, on the one hand, upon geographical tropes that define the norms and nature of sexuality; and on the other, it redraws the material and imagined geography of the social world according to its dictates. As an example of geographical discourses of sexuality, we can point to the role that geographical imaginations of sexuality at the colonial margins played in the critique and constitution of Western sexuality. It can be noted here that Foucault's own contrast of a 'Western' sexual science and an 'Eastern' *ars erotica* is thoroughly indicative of the importance of the geographical imagination for the transformation of sex into discourse, at the same time as being seriously misconceived in its reliance upon orientalist fantasy. Postcolonial critics, inside and outside Geography, are increasingly elaborating on the argument that metropolitan sexual norms were, and are, constructed out of the discursive contrast with the geographical imagination of colonial and non Western sexualities.

This attention to discourse can be expanded into a wider spatial history of sexuality, moreover, particularly insofar as it is engaged in the particular and specific manifestation of sexual geographies. A second thematic here points to the importance of spatial practices in the authorized production of sexual subjectivities. The regulation of sexuality, a project or set of projects grounded in space and spatial principles, is certainly one aspect of this theme. Foucault's own work directs us to various 'sites of sexuality' in which eroticism and sensualism are contained and channeled, but which also invite, through the perceived threat of deviant sexuality, surveillance and discipline by authority. The activities of nineteenth and twentieth century moral reform agencies are thus replete with concerns over urban sites in which transgressive sexualities are seen to flourish; these sites of sexual danger invited intervention by the authorities, which may of course be the point of this discursive activity. But the notion of the regulation of sexuality must also allow for the unsuccessful, contradictory, and self consciously restrained nature of societal intervention, the careful selection of socially constructed targets, and a suspicion of private morality agencies, all within self professedly liberal regimes in which subjects are granted and even required to have certain freedoms (sexual and otherwise). The regulation of heterosexual space in modern liberal societies was thus far more complex than a matter of straightforward discipline and repression. The nature of this heterosexual space is still under theorized and researched, but it is clear that it is intimately connected to the more glaring social and spatial manifestations of heterosexual deviance, and to nonheterosexual forms that provide the counterpoint to such heterosexual norms.

While we can readily agree on the need to recognize the importance of geography for the discursive construction of sexuality, we must take this further. The concern with discourse must consider the policing and regulation of sexuality as the result of representational strategies that are themselves the products of specific spatial, procedural, and cultural circumstances. Words, discourses, do not simply produce geographies of sexu ality; they are also interventions in very specifically situated geographies of sexual encounter, experience, and self understanding. In this regard, it is clear that sexual identities are also the product of spatial practices that are not strictly supervised or authorized from above. They are associated in fact with the world of the street and even, to continue the spatial metaphor, with the under ground or sexual underworld. The most well known such practice is undoubtedly that of the nineteenth century urban wanderer and voyeuristic spectator, the *flâneur*, but recent studies that have extended and critiqued this lit erature include work on homosexual cruising and on the sexual implications of cross class slumming. Such work confirms the importance of social space for the emer gence of sexual subjectivities, but is also careful to specify these experiences of modernity. It is particularly im portant here not to valorize such activity uncritically as forms of 'resistance' to power and to the projects of sexual 'repression' that come from above. These practices reveal in fact the uneven privileges (of gender, class, race, and so on) that characterize sexual citizenship. Sexual sub jectivities are formed in and through socially differen tiated landscapes of power, and contribute to a complex spatial politics of inclusion and abjection. So Foucault's suspicion of simple narratives of sexual 'repression' and 'liberation' is of vital importance. Historical and con temporary geographies of sexuality have consistently qualified the dyad of power and resistance. Though there remain difficulties inherent in an opposition between spaces of regulation and spaces of experience, there is no going back to a cosy and comforting narrative of repression and liberation, the spatial story of Victorian public denial and private hypocrisy consigned to history by our contemporary openness about sex.

Queer Times, Queer Frontiers

Nowhere perhaps is this more important than in the discussion of homosexuality and other transgressions of the heterosexual norm. It is of course perfectly right to value the resistance of gay men and women to the norms of a heterosexist and homophobic society, and many accounts do this. It is no part of this article to downplay the important work on the political mobilization of sexual minorities, subcultures, dissidents, and radicals in claiming space and thereby rights as sexual citizens; it remains the case that all over the world people are harassed, discriminated against, threatened, and killed for their sexual orientation, sexual desires, and perceived sexual and gender transgression. Making space, aca demically and politically, for gay men and women, bisexuals, transgenders, and transsexuals, among others, remains of the utmost importance. At the same time, however, there has been a reaction against the kind of social movement theory and identity politics that invests uncritically in such typologies. The expanding canon of sexualities resembles in its ever increasing complexity the classifications of the early sexologists (*urnings*, *ura nians*, *dionings*, and so on), and is subject to the criticism that it naturalizes and essentializes sexual difference. The dichotomization of 'gay' and 'straight' obviously relies upon the late nineteenth century, and Western, discovery not just of 'homosexuality' but also of 'heterosexuality'. The problem is that 'gay' becomes universalized just as much as 'straight' does; and the suspicion is that this not only does little to subvert what has been called 'hetero normativity' but that it may actually subsidize it. The creation of gay men's urban enclaves, for instance, is for many an historic achievement, but from a different angle it is yet another kind of 'tolerated zone' which assumes that there is indeed something to 'tolerate'. It may be that such neighborhoods become an unchallenging kind of heterotopia – an alibi, literally – that underwrites the heteronormativity of the sociospatial order. Sexual 'de viants' become associated with particular 'marginal' spaces in a process that effectively legitimates the norms of heterosexual spatiality.

The various interventions collected under the name 'queer theory', which set out to destabilize this hetero sexual norm, are based on demonstrating the socially and spatially constructed nature of all sexual identities. Not only is this necessary in terms of the supposed solidity, stability, and security of heterosexuality, it is an inevit able challenge to the notion of a unitary, transcultural, and transhistorical gay identity and subculture. The historian George Chauncey's landmark study of *Gay New York* demonstrates for instance the visibility, vitality, and viability of an early twentieth century 'gay male world' – a world whose sexual topography was inseparable from the creation of alternative sexual practices and identities. Framing this as a challenge to the narrative of repression and liberation contained in the metaphor of 'the closet', Chauncey, along with others who have focused on the historical geography of queer urban culture, shows how sexual difference is constituted in dynamic geographies of public sex and sociability. This is emphatically not about uncovering some pre established but hidden gay culture; the task is rather to show how all forms of sexuality are produced in specific sites and locales, through specific spatial practices, subject to regulative regimes, but often enough irrepressibly and uncontain ably creative. These studies show how geographies of sexual difference were constituted long before the modern, post closet, era of supposed sexual openness and authenticity. It is important also to note that this historiography challenges Foucault's wholly inadequate narrative of the late nineteenth century emergence of the homosexual as a singular type of person, an argumentative *aporias* that serves only to obscure the sociogeographical construction of sexuality. Several commentators have shown that the supposed deviance of the male homosexual – his 'queerness' – remained tied to his bodily and spatial disposition long after this supposed watershed. Sexual transgression, then and now, was a matter of having an appearance, behavior, and attitude deemed 'out of place' in particular environments.

Whether these ideas and interventions go far enough is something of a moot point, however. Geographers have been both appreciative of Chauncey's description of the spatiality of a pre closeted gay sexuality and critical of his revisionism. The recovery of gay or queer space(s), marvelous in itself, hardly exhausts the project of a queer geography. Not only does 'queer' refer to an enormous diversity of practices, its restless disruptiveness militates against any attempt to pin down and specify sexual identities. It will not do, therefore, simply to counterpose 'queer spaces' such as the bathhouse or the cottage, to heteronormative space. The point is rather to show the artifice and ideology behind all forms of sexual spatialization. Here the influence of the theorists Eve Kosofsky Sedgwick and Judith Butler has been particularly important: both have provided thoroughgoing critiques of essentialist positions on sex/gender, and pointed to alternative understandings. And though neither has been particularly keen to specify a materially geographical analysis, their work has been influential in this direction. Butler's stress on the 'performativity' of sex, sexuality, and sexual identity has been extended in geographers' emphasis on the spaces which allow and authorize such performances, or which are transgressed by them. Sedgwick's analysis of the necessary contradictions of a culture epistemologically distorted by the reification of heterosexual/homosexual difference also feeds directly into a more straightforwardly material analysis of what she herself refers to as the contingencies and geographies of the closet. Michael Brown's recent study of the spatiality of the contemporary closet, demonstrating the diversity and scalar complexity of what he calls 'closet space', is a direct response to this question, and a signal contribution to the field. It is a study that, along with other cutting edge work, points the way toward a political geography of sexual citizenship.

While Butler's work appears to take a strong constructivist position on sexuality, in Sedgwick's work we also have an attempt to bypass the potentially sterile opposition between essentialism and constructivism, nature and nurture. Sedgwick famously borrows the terms 'minoritizing' and 'universalizing' to point to the epistemological work of discourses that, on the one hand, propose the homosexual/heterosexual definition of interest only to a fixed 'homosexual' minority, and, on the other, the view that sees this as an issue for all of us, across the spectrum of sexualities. It is thus not a matter of whether essentialists or constructivists are right, but rather of understanding how different discourses of sexuality – themselves contradictory – are deployed in different forms at different times, sometimes trading on the identification of minority sexual identities, and sometimes asserting the instability, indeterminacy, and plasticity of sexuality. While Sedgwick uses these terms to address the problematic modern discourse of homosexuality, it is clear that many or perhaps all sexual discourses trade on minoritizing and universalizing ideas. In the nineteenth century, for instance, the idea of the 'prostitute' as a woman set apart is ranged alongside the notion of the 'fallen woman' that identifies the potential of social and sexual transgression in all women. We might even suggest that sexuality itself is a discourse that is both minoritizing and universalizing: that is, proposed firstly as universal and innate, but at the same time specifically and problematically attached to particular, 'minority' communities and identities. The implications for geographies of sexuality are that sexuality is conceptually all over the map, as Sedgwick puts it: inseparable from social space in its entirety and complexity, and yet particularized in specific sites attached to particular sexual identities. This is of course to return to the contradictory conception of sexuality noted at the beginning of this article, but not this time simply as a matter of frustration; while we must certainly enjoin skepticism toward the science and conventions of sexuality, we can proceed by analyzing and critiquing the landscape that is the result of sexuality's paradoxical nature. Queer theory directs us to interrogate the compromised and contradictory cultural work of sexual categorizations, to repudiate the oppositional categories upon which these categorizations depend, and to work toward a fairer political geography rather than a more accurate map of sexuality. Queer's notably nomadic spatial valence, its outlaw work, has certainly highlighted the political and academic work that is left to be done: such as resisting spatial fixes of sexual identity and affirming spaces of sexual citizenship in an age of globalization and in the time of AIDS.

It may well be that queer has not been fully able to transcend the epistemological categories of sexology – and a post queer (though certainly not anti queer) understanding of sexuality is increasingly advocated. Geographers of sexuality have, however, been directed by the career of queer theory to a strikingly expansive and engaging new topography of sexuality. We recognize that all space is sexed, in that it encodes and endorses societal norms of sexual behavior and identity, but also that, because of the unstable and contradictory nature of these sexual categories, all space is actually or potentially queer. To revisit the metaphor introduced at the begin ning of this article, we might be tempted to substitute a proposed 'queer frontier' – a liminal, lawless space in which traditional roles and rules are suspended or freely transgressed. This still has troubling colonial overtones, for sure, but it is in this territory that the emergent geographies of sexuality are likely to be found.

See also: Body, The; Cultural Studies and Human Geography; Cultural Turn; Culture/Natures; Feminism and Work; Feminism/Feminist Geography; Feminist Geography, Prehistory of; Feminist Methodologies; Flãneur, The; Foucauldianism; Gay Geographies; Gender and Health; Gender, Historical Geographies of; Gender in the City; Heteronormativity; HIV/AIDS in Developed Countries; Identity Politics; Lesbian Geographies; Masculinities; Mental Health; Moral Landscapes; Nature-Culture; Natures, Gendered; Other/Otherness; Parenting/Motherhood/Fatherhood; Performativity; Pregnancy and Childbirth; Psychoanalysis; Queer Theory/Queer Geographies; Social Geography; Social Justice, Urban; Urban Village.

Further Reading

Bell, D. and Valentine, G. (eds.) (1995). *Mapping Desire: Geographies of Sexualities*. London: Routledge.
Binnie, J. (2004). *The Globalization of Sexuality*. London: Sage.
Bristow, J. (1997). *Sexuality*. London: Routledge.
Brown, M. (2000). *Closet Space: Geographies of Metaphor from the Body to the Globe*. London: Routledge.
Butler, J. (1990). *Gender Trouble: Feminism and the Subversion of Identity*. London: Routledge.
Chauncey, G. (1995). *Gay New York: The Making of the Gay Male World, 1890 1940*. London: Flamingo.
Duncan, N. (ed.) (1996). *BodySpace: Destabilizing Geographies of Gender and Sexuality*. London: Routledge.
Foucault, M. (1998). *The History of Sexuality: The Will to Knowledge*, vol. 1. London: Penguin.
Hemmings, C. (2002). *Bisexual Spaces: A Geography of Sexuality and Gender*. London: Routledge.
Hubbard, P. (2002). Sexing the self: Geographies of engagement and encounter. *Social and Cultural Geography* 3, 359 381.
Ingram, G. B., Bouthillette, A. M. and Retter, Y. (eds.) (1997). *Queers in Space: Communities, Public Places, Sites of Resistance*. Seattle, WA: Bay Press.
Knopp, L. (1992). Sexuality and the spatial dynamics of capitalism. *Environment and Planning D: Society and Space* 10, 651 669.
Mort, F. (2000). *Dangerous Sexualities: Medico Moral Politics in England Since 1830* (2nd edn.). London: Routledge.
Namaste, K. (1996). Genderbashing: Sexuality, gender and the regulation of public space. *Environment and Planning D: Society and Space* 14, 221 240.
Sedgwick, E. K. (1990). *Epistemology of the Closet*. London: Prentice Hall.
Spargo, T. (1999). *Foucault and Queer Theory*. Cambridge: Icon Books.
Valentine, G. (2001). *From Nowhere to Everywhere: Lesbian Geographies*. San Francisco, CA: Harrington Park Press.
Weeks, J. (1989). *Sex, Politics and Society: The Regulation of Sexuality Since 1800* (2nd edn.). London: Longman.

Relevant Websites

http://virtual.yosemite.cc.ca.us/bhutchison/sexuality websites.html
A guide to websites on human sexuality.
http://pegasus.cc.ucf.edu
Geography on gender and sexuality.
http://www.queertheory.com
Queer theory.
http://www.utexas.edu/utpress/journals/jhs.html
The history of sexuality.
http://www.glbtq.com

Shift-Share Analysis

R. W. Jackson, West Virginia University, Morgantown, WV, USA
K. E. Haynes, George Mason University, Fairfax, VA, USA

Glossary

Growth Decomposition Technique A method of analysis in which growth is broken down into components to better understand the nature of overall change.

Mix Regional change that would be expected if its parts grew at the same rate as the parts at the national level.

National Share Regional change that would be expected if its parts (e.g., industries and sectors) changed at the same rate as the nation as a whole.

Productivity The amount of output per unit of an input such as labor.

Regional Shift Change that is not explained by either the national share or mix components, and is therefore specific to the region.

Introduction

Change, the essence of behavior, forms the focus of a dominant portion of social science research. While de scription must often precede explanation, explanations almost always focus on the nature of changes within sys tems or in components of systems. One approach to a better understanding of behavioral processes is the de composition of change into various components of be havior that can be understood by reference to other changes that are well studied. Within this approach, it is often possible to decompose change into components that are expected or attributable to known and well understood forces and the balance of change for which an explanation has yet to be developed. The shift share method in its most fundamental form fits this description.

As it is applied in geography, particularly within economic geography, shift share has become a popular tool most commonly used to decompose regional growth into component parts to enhance the understanding of regional economic growth and development. Its popu larity can be attributed in part to the fact that it is relatively simple and straightforward to implement, it relies upon easily accessible data, and it generates a set of results that has clear and intuitive interpretation. In this context, shift share can be defined as a method of dis aggregating regional change to identify and understand better the components of that change. It is derived from a set of accounting relationships that we develop and present below.

In developing the method on the following pages, we focus on regional employment change by industry since that is one of the most common application contexts of this method of regional analysis. However, other values of interest (such as income or value added) could form the focus. Indeed, the basic shift share method has been ex tended and refined in a variety of ways. First, however, we turn to the presentation of the method in its most fun damental form.

The Classic Shift-Share Model

Change requires the passage of time. The measurement of change requires values for the beginning and the end of the period of interest. A variety of sources for employment data are available, each of which reports employment levels for geographic regions in total and for the industries that make up the regional economies. In most cases, the employment data will be pegged to a specific date or to an average level for a calendar period such as a month, a quarter, or a year. To make the pre sentation of the method as general as possible, we assume that the analyst has an appropriate set of data, and that the period of interest can be represented by one unit of time. Specifically, we use time $t-1$ as a reference to the relevant date for the beginning period, and time t for the relevant date for the ending period. In this way, we can examine the change in employment (or income, value added, etc.) from time $t-1$ to time t.

We now define e_t^r as the total employment in region r at time t, $e_{i,t}^r$ as the employment in region r industry i at time t, e_t^n as the total national employment at time t, and $e_{i,t}^n$ as the national employment in industry i at time t.

Then for beginning and ending periods, we calculate the following three growth rates:

$$r_i = e_{i,t}^r / e_{i,t-1}^r, \quad \text{growth rate for regional industry } i \quad [1]$$

$$n_i = e_{i,t}^n / e_{i,t-1}^n, \quad \text{growth rate for national industry } i \quad [2]$$

$$n = e_t^n / e_{t-1}^n, \quad \text{average national total growth rate} \quad [3]$$

By definition

$$e_{i,t}^r = e_{i,t-1}^r r_i = e_{i,t-1}^r \left(e_{i,t}^r / e_{i,t-1}^r \right) \quad [4]$$

If regional industry i grew at the average national rate, we could write

$$e^r_{i,t} = e^r_{i,t-1} n \qquad [5]$$

and we would attribute the change in industry i in region r simply to overall system change. If industry i is a particularly fast growing one nationally, however, there will be a difference between n and n_i. Further, if there were no regional variation in industry i's growth rate, we could account for the deviation of specific industry performance from the average by

$$e^r_{i,t} = e^r_{i,t-1}[n + (n_i \quad n)] \qquad [6]$$

The order of the rates in the new term in parentheses is not arbitrary, because if the industry is fast growing (nationally), $e^r_{i,t}$ must be greater than the value obtained from eqn [2].

If there is variation in industry i's growth rate across geographical regions, then the difference between r_i and n_i will be nonzero. To reflect this variation in our formal representation, we would then transform eqn [6] into

$$e^r_{i,t} = e^r_{i,t-1}[n + (n_i \quad n) + (r_i \quad n_i)] \qquad [7]$$

The bracketed term in eqn [7] clearly reduces to $r_i = (e^r_{i,t}/e^r_{i,t \quad 1})$, duplicating and confirming the identity in eqn [4].

If we are interested in examining directly the change in employment, we can subtract $e^r_{i,t \quad 1}$ from both sides, yielding

$$e^r_{i,t} \quad e^r_{i,t-1} = e^r_{i,t-1}[n + (n_i \quad n) + (r_i \quad n_i)] \quad e^r_{i,t-1} \qquad [8]$$

or

$$\Delta e^r_i = e^r_{i,t-1}[n + (n_i \quad n) + (r_i \quad n_i) \quad 1] \qquad [9]$$

and

$$\Delta e^r_i = e^r_{i,t-1}[(n \quad 1) + (n_i \quad n) + (r_i \quad n_i)] \qquad [10]$$

where Δe^r_i is the change in regional industry i employment over the study period.

We can separate the components on the right hand side of eqn [10], however, yielding the following values and interpretations:

$$N_i = e^r_{i,t-1}(n \quad 1)$$

the national share component, which is the growth in the regional industry that can be attributed to the national average rate of growth,

$$M_i = e^r_{i,t-1}(n_i \quad n)$$

the industry mix component, or the growth in the regional industry that can be attributed to the difference between the specific industry's national growth rate and the national average rate of growth, and

$$R_i = e^r_{i,t-1}(r_i \quad n_i)$$

the regional shift, that is, growth in the regional industry that cannot be attributed to industry or national effects: often assumed to reflect comparative advantage.

Given these definitions,

$$\Delta e^r_i = N_i + M_i + R_i \qquad [11]$$

Finally, we can sum both sides over all industries and show that

$$\sum \Delta e^r_i = \sum (N_i + M_i + R_i) \qquad [12]$$

$$\Delta e^r = N + M + R \qquad [13]$$

A Shift-Share Example

As an empirical example, we present in tabular form below a classical shift share analysis for the West Virginia for the 2001–05 period, using Bureau of Economic Analysis Employment Data (**Tables 1–3**).

From **Table 3**, we can see that 16 of 24 industries in West Virginia performed more poorly than in the US. Eight of the 24 industries outperformed their US counterparts, but only three of these eight were strong performers nationally. Professional and technical services, arts entertainment and recreation, and other services, in particular, would certainly merit further, in depth attention by analysts concerned with West Virginia economic development.

Early Development

Early applications of shift share, sometimes also called mix and share, concentrated on the study of overall regional economic performance rather than on the relative growth or decline of individual industries. The focus lay in the regional shift summary measure – the sum of the regional shift over all industries – called the 'differential shift'. This part of regional growth or decline is region specific and cannot be attributed to either national or industry wide trends. It is a short conceptual step from this interpretation to the assumption that a positive shift can be associated with the comparative locational advantage of the region for an industry, and vice versa.

Subsequent applications focused more directly on shifts in individual industries rather than the region as a whole. Implicit within such applications is the normative

Table 1 Employment for West Virginia and the US, 2001

	Er(t 1)	*Er(t)*	*En(t 1)*	*En(t)*	*Total*
Proprietors employment	149 524	164 940	28 183 700	33 282 600	15 416
Farm employment	22 861	22 640	3 056 000	2 914 000	221
Forestry, fishing, related activities	4254	3950	1 022 500	1 012 200	304
Mining	28 210	30 971	811 400	820 000	2761
Utilities	7003	6381	618 800	594 100	622
Construction	51 149	54 376	9 846 700	10 845 700	3227
Manufacturing	74 820	65 081	16 994 600	14 860 900	9739
Wholesale trade	26 363	26 517	6 273 400	6 401 300	154
Retail trade	111 678	112 986	18 528 800	18 941 100	1308
Transportation and warehousing	25 678	25 545	5 474 000	5 510 100	133
Information	15 256	13 268	4 053 800	3 577 100	1988
Finance and insurance	27 988	26 773	7 839 600	8 186 600	1215
Real estate and rental and leasing	19 941	23 339	5 551 400	6 934 300	3398
Professional and technical services	35 870	39 442	10 575 800	11 488 700	3572
Mgmt of companies and enterprises	2968	4134	1 779 300	1 857 000	1166
Administrative and waste services	39 463	40 490	9 621 000	10 645 100	1027
Educational services	11 882	12 106	3 058 300	3 552 900	224
Health care and social assistance	106 641	115 005	15 611 400	17 267 000	8364
Arts, entertainment, and recreation	13 428	17 191	3 243 100	3 517 300	3763
Accommodation and food services	58 422	62 393	10 825 200	11 728 300	3971
Other services, except pub admin	50 608	55 617	9 049 600	9 758 900	5009
Government and gov't enterprises					
Federal, civilian	21 235	21 993	2 728 000	2 790 000	758
Military	10 140	9951	2 099 000	2 027 000	189
State and local	117 270	119 679	18 353 000	19 020 000	2409
Total	1 032 652	1 074 768	195 198 400	207 532 200	42 116

Table 2 Shift-share components

	r_i	n_i	N_i	M_i	R_i
Proprietors employment	1.10	1.18	9448	17 604	11 635
Farm employment	0.99	0.95	1444	2507	841
Forestry, fishing, related activities	0.93	0.99	269	312	261
Mining	1.10	1.01	1782	1483	2462
Utilities	0.91	0.96	442	722	342
Construction	1.06	1.10	3232	1957	1962
Manufacturing	0.87	0.87	4728	14 121	345
Wholesale trade	1.01	1.02	1666	1128	383
Retail trade	1.01	1.02	7056	4571	1177
Transportation and warehousing	0.99	1.01	1622	1453	302
Information	0.87	0.88	964	2758	194
Finance and insurance	0.96	1.04	1768	530	2454
Real estate and rental and leasing	1.17	1.25	1260	3707	1569
Professional and technical services	1.10	1.09	2266	830	476
Mgmt of companies and enterprises	1.39	1.04	188	58	1036
Administrative and waste services	1.03	1.11	2494	1707	3174
Educational services	1.02	1.16	751	1171	1698
Health care and social assistance	1.08	1.11	6738	4571	2945
Arts, entertainment, and recreation	1.28	1.08	848	287	2628
Accommodation and food services	1.07	1.08	3691	1182	903
Other services, except pub admin	1.10	1.08	3198	769	1042
Government and gov't enterprises					
Federal, civilian	1.04	1.02	1342	859	275
Military	0.98	0.97	641	989	159
State and local	1.02	1.04	7410	3148	1853
Total	1.04	1.06	65 249	854	22 279

Table 3 Shift-share summaries

	N_i	M_i	R_i	M	R
Proprietors employment	9448	17 604	11 635	+	
Farm employment	1444	2507	841		+
Forestry, fishing, related activities	269	312	261		
Mining	1782	1483	2462		+
Utilities	442	722	342		
Construction	3232	1957	1962	+	
Manufacturing	4728	14 121	345		
Wholesale trade	1666	1128	383		
Retail trade	7056	4571	1177		
Transportation and warehousing	1622	1453	302		
Information	964	2758	194		
Finance and insurance	1768	530	2454		
Real estate and rental and leasing	1260	3707	1569	+	
Professional and technical services	2266	830	476	+	+
Mgmt of companies and enterprises	188	58	1036		+
Administrative and waste services	2494	1707	3174	+	
Educational services	751	1171	1698	+	
Health care and social assistance	6738	4571	2945	+	
Arts, entertainment, and recreation	848	287	2628	+	+
Accommodation and food services	3691	1182	903	+	
Other services, except pub admin	3198	769	1042	+	+
Government and gov't enterprises					
Federal, civilian	1342	859	275		+
Military	641	989	159		+
State and local	7410	3148	1853		
Total	65 249	854	22 279		

Interpretations:

	indicates that the industry was a poor performer nationally, and even poorer in the region than in the nation (9 WV industries)
+	indicates that the industry was a poor performer nationally, but better in the region than in the nation (5 WV industries)
+	indicates that the industry was a good performer nationally, but poorer in the region than in the nation (7 WV industries)
+ +	indicates that the industry was a good performer nationally, and even better in the region than in the nation (3 WV industries)

assumption that an industry in a region should grow at the same rate as for the nation as a whole, an assumption which is the foundation of one of the earliest criticisms of method. Later applications explored the use of the shift share method in regional economic forecasting. If the overall national and national industry growth rates could be projected (perhaps econometrically), all that would be needed would be a projection of the regional shift to project total regional employment. In a forecasting con text, the 'constant shift' assumption is that the future regional shift is equal to the past.

Scholars debated whether the regional shift com ponent would be stable enough over time to support its use as a forecasting tool, but no consensus developed. Results of forecasting models based on shift share modeling varied by: size of place, level of disaggregation, time frame used, and certain mathematical assumptions (e.g., homotheticity). Although shift share forecasting effectively failed to show adequate potential, the debate uncovered a number of other conceptual issues relevant to even the descriptive, *ex post* applications of the method.

Analytical Issues

Among the issues identified in the study of shift share forecasting potential was the choice of appropriate reference region. Shift share is a method that clearly can only be used when comparing a study area of functional interest to a specific reference region. For example, one might choose to study a US state relative to the Census region within which it is embedded, or a county within a US state, or simply a state/province relative to a country (i.e., a state within the US). The choice of region will influence the results and interpretations.

A second issue concerns the choice of an appropriate time period for analysis. The effects of a measurement date relative to business cycle fluctuations can lead to very different results from analyses with start and end dates shifted only a few years. For this reason, analysts typically try to select start and end dates that lie in similar positions on the business cycle (e.g., trough to trough or peak to peak), or apply the shift share analysis in a series of successive estimations over short time periods – a method known as 'dynamic' shift share.

With annual data for an extended period, for example, one could conduct the shift share analysis on a 1 or 2 year interval basis, and assess the results both graph ically and with the use of summary statistics. Dynamic shift share provides the opportunity to smooth changes by averaging.

A third issue concerns the question of whether the regional shift component truly is reflective of some comparative advantage, and if so, whether this advantage is likely to persist. Positive shifts could, for example, bid up the prices of capital, labor, and other factors of pro duction in the region needed as inputs for the industry, which would ultimately reduce the initial comparative advantage. Alternatively, positive shifts might generate growth that creates agglomeration economies. Lastly, the shift share method ignores inter industry linkages. Are specific combinations of industries especially likely to achieve rapid growth rates? A voluminous literature on industry complexes and clusters exists, supporting the importance of inter industry relationships.

Despite shortcomings, the simplicity of the basic shift share model and the ease with which it can be applied has provided a pragmatic justification for its widespread use in regional analysis and has spawned extensions and adjustments in attempts to respond to various criticisms. Since its introduction in 1960, shift share has been adapted and applied to a wide variety of regional issues including intra urban dynamics, analysis of public decision making, manufacturing labor productivity, occupational gender composition, impact of transporta tion employment on regional growth and port competition, regional economic forecasting, regional employment change, regional demographic change, migration turnaround, rural occupational adjustment, analysis of disease death rates, ethnic/racial division of labor, differential employment of citizens and noncitizens in science and engineering, regional convergence in Europe, evaluation of regulationist history, spatial structure and employment change, and telecommunica tions access and efficiency. The classic shift share model has been reformulated as a statistical (two way ANOVA) model and has been shown to have a linear econometric equivalent. However, due to some of the criticism noted earlier and others, there have been several major exten sions and new formulations of the classic model described above.

Shift-Share Extensions

During the late 1980s and 1990s, new formulations and extensions, and a wide variety of applications of shift share analysis replaced the widely criticized original model of the 1970s. Yet despite the increasing compu tational capacity of researchers and the availability of more advanced regional models, it has not been totally displaced. This is because in a relatively short time and with little cost, shift share provides reasonable, easily interpreted results. Its simple logic, analytic clarity, and relatively easily accessible data requirements make the model popular both among researchers and practitioners.

Output and Productivity Interaction

Productivity growth is one of the most important sources of economic expansion and hence it is a key source of the rise in output and income. There is another important aspect of productivity: its impact on employment change. While productivity growth plays a crucial role in eco nomic growth in the long run by attracting new busi nesses and hence creating new jobs, it may have a negative impact on sector specific employment change in the short run because it will be possible to produce the same amount of output with less input, including labor. In that sense, understanding, measuring, and explaining prod uctivity growth and its impact on the regional economy is very important and is a major concern to elected policymakers. This is also important in the promotion of sound economic development policies and programs.

To investigate the interaction of productivity and output change on regional economies, several extensions of classic shift share analysis have been developed. In this context, the shift share equations have been expressed in the following forms:

$$TS_L \equiv NS_L + IM_L + RS_L \qquad [14]$$

$$\begin{aligned} NS_L &\equiv NS(a_L) + NS(b_L) \\ &= \sum E_{ir}(a_{nL} + b_{nL}) \end{aligned} \qquad [15]$$

$$\begin{aligned} IM_L &\equiv IM(a_L) + IM(b_L) \\ &= \sum E_{ir}[(a_{inL} \quad a_{nL}) + (b_{inL} \quad b_{nL})] \end{aligned} \qquad [16]$$

$$\begin{aligned} RS_L &\equiv RS(a_L) + RS(b_L) \\ &= \sum E_{ir}[(a_{irL} \quad a_{inL}) + (b_{irL} \quad b_{inL})] \end{aligned} \qquad [17]$$

where the *a*'s represent the rate of employment change in the region and the reference area that result from vari ations in output over time with productivity constant, and the *b*'s represent the rate of employment change that result from variations in productivity over time with output constant.

The employment change resulting from nonlabor factors' contribution can be calculated as the difference between actual employment change and employment change that results from output and productivity vari ation. This can be shown as

$$\Delta EP = \Delta E \quad \Delta EL \qquad [18]$$

where ΔE is the actual employment change over time in the region, ΔEL is the employment change in the region or state resulting from the change in labor productivity and output. ΔEP is the employment change resulting from the other production factors' contribution to total factor productivity. In fact, nonlabor factors, productivity and output change, and their impact on employment can be integrated into shift share models. From eqns [8] and [9] for sector i in region r,

$$TS \equiv NS + IM + RS \qquad [19]$$

$$\begin{aligned} TS \equiv{} & [NS(a) + NS(b)] + [IM(a) + IM(b)] \\ & + [RS(a) + RS(b)] \end{aligned} \qquad [20]$$

We can rewrite eqn [19] in terms of employment growth rates as

$$\begin{aligned} \Delta\ln E_{ir} \equiv{} & \Delta\ln E + (\Delta\ln E_i \quad \Delta\ln E) \\ & + (\Delta\ln E_{ir} \quad \Delta\ln E_i) \end{aligned} \qquad [21]$$

This equation corresponds to the traditional shift share model and $\Delta\ln E$ represents the national share, $(\Delta\ln E_i - \Delta\ln E)$ represents the industry mix, and the regional share is represented by $(\Delta\ln E_{ir} - \Delta\ln E_i)$.

The set of further extensions and variations of the reformulated model is rich and varied. For example, equations can be developed to take into account output change and productivity change resulting from improve ment of capital and labor and a residual that reflects dimensions such as faster than average regional technical progress, scale and external economies, and infrastructure.

International Trade and Shift-Share Analysis

The traditional shift share model assumes that there is no interregional trade and the region under investigation serves only the regional economy. Further extensions implicitly incorporate interregional or international trade flows by including output change in the model, but these are only marginally linked to the international dimen sion. Other shift share extensions that incorporate international dimensions pay special attention to inter national economic forces essential for the integration of regional analysis into global dynamics. These extensions of shift share that encompass trade (imports and exports) as well as the domestic economy are critical additions in modernizing the method to make it useful in the global economy of the twenty first century.

Reference Area–Region Interaction

Despite various new formulations and extensions and a wide variety of applications of the shift share method, one important issue remained unaddressed in earlier applications: reference area–region interaction. In the traditional model and all of its extensions, the reference area contains total employment (income, value added, etc.) including the region's (state and county) share. In some cases, particularly at the county level of analyses, a large region or sector could dominate the reference area employment having a large share of total employ ment. Hence, below (above) average regional employ ment, growth (decline) could affect substantially the employment growth of the reference area. In such cases, the results of the shift share analysis may be misleading by over or underestimating the components of the shift share model used. This issue poses a serious problem for practitioners because they generally use a given shift share model to investigate the sectoral growth patterns of a given region, state, or county, rather than comparison of regions, states, or counties. Neither earlier formulations and extensions nor the above models have explicitly dealt with a classic reference area component problem.

This issue can be addressed by redefining the refer ence area component as the rest of the nation (region and state), excluding the region of interest. In this way, it is possible to isolate the influence of a larger region or sector on the reference area. Since we are investigating the region under consideration relative to the reference area, this redefinition approach provides more realistic information about the region, and avoids over or underestimation problems. This approach can be applied to all extensions of the shift share analysis. Some may argue that this redefinition removes one of the shift share models' most attractive properties, regional comparison, but in cases where regional analysis is more important than comparison it can be very useful. Note that even in the simplest form of shift share model, the region–reference area interaction is important in its own right and needs to be addressed. However, as the model gets more refined and as the level of sectoral and regional disaggregation increases, this issue becomes increasingly important because the various ratios used become more sensitive to issues of scale.

Conclusions

Shift share analysis is a convenient and effective tool when used judiciously. Its convenience of calculation and logical, straightforward interpretations make it a valuable addition to the regional analysis tool kit. Judicious use requires a clear understanding of the method's strengths and limitations. Shift share is a descriptive tool. It is designed to reveal change regional relative to a reference region benchmarks, but it does not explain differences in growth patterns.

A negative shift alone does not indicate that the economy is or will continue to be in decline. One or a set of indicators can signal areas that warrant further scrutiny, but in and of themselves are not necessarily cause for alarm or policy intervention. As with any *ex post* analysis, shift share components refer to the past, and may no longer be applicable due to structural or other changes underway. Shift share is best used with other regional analysis methods, such as input–output for industrial structure or data envelopment analysis for as sessment of efficiency and the analytic hierarchy process.

Shift share has proven its value over nearly a half century of use. Its flexibility and intuitive appeal will no doubt ensure its continued popularity in decades to come.

See also: Input-Output Analysis; Time Geographic Analysis; Time Series Analysis.

Further Reading

Barff, R. A. and Knight, P. L. III (1988). Dynamic shift share analysis. *Growth and Change* 15, 1 10.
Bekele, G. and Jackson, R. W. (2005 06). Theoretical Perspectives on Industry Clusters, Working Paper 2006 5, Regional Research Institute, West Virginia University (http://www.rri.wvu.edu).
Dinc, M., Haynes, K. E. and Qiangsheng, L. (1998). A comparative evaluation of shift share models and their extensions. *Australasian Journal of Regional Studies* 4(2), 275 302.
Dunn, E. S. (1960). A statistical and analytical technique for regional analysis. *Papers of the Regional Science Association* 6, 97 112.
Haynes, K. E. and Dinc, M. (1997). Productivity change in manufacturing regions: A multifactor/shift share approach. *Growth and Change* 28(2 Spring), 201 221.
Haynes, K. E. and Machunda, Z. B. (1987). Considerations in extending shift share analysis: A note. *Growth and Change* 18(2), 69 78.
Hellman, D. A. (1976). Shift share models as predictive tools. *Growth and Change* 7(3), 3 18.
Knudsen, D. C. (2000). Shift share analysis: Further examination of models for the description of economic change. *Socio Economic Planning Sciences* 34(3), 177 198.
Loveridge, S. and Selting, A. C. (1998). A review and comparison of shift share identities. *International Regional Science Review* 21(1), 37 58.
Mulligan, G. F. and Molin, A. (2004). Estimating population change with a two category shift share model. *Annals of Regional Science* 38(1), 113 130.
Nazara, S. and Hewings, G. J. D. (2004). Spatial structure and taxonomy of decomposition in shift share analysis. *Growth and Change* 35(4), 476 490.
Perloff, H. S., Dunn, E. S. Jr., Lampard, E. E. and Muth, R. F. (1960). *Regions, Resources and Economic Growth*. Lincoln, Nebraska: University of Nebraska Press.
Plane, D. A. (1989). Population migration and economic restructuring in the United States. *International Regional Science Review* 12(3), 263 280.
Sihag, B. S. and McDonough, C. C. (1989). Shift share analysis: The international dimension. *Growth and Change* 20(3), 80 88.
Smith, S. A. (1991). Shift share analysis of change in occupational sex composition. *Social Science Research* 20(4), 437 453.

Relevant Websites

http://www.bea.gov
Bureau of Economic Analysis Employment Data.
http://www.georgiastats.uga.edu
Northeast Regional Project 1011, Shift Share Analysis Website.

Simulation

S. M. Manson, University of Minnesota, Minneapolis, MN, USA

Glossary

Difference Equation A system of mathematical equations in which each equation relies on the results of other equations, calculated over discrete time steps.

Emergence Process by which complex systems gain qualities that result from synergistic interactions among system components.

Expert Elicitation The processes of extracting and encoding expert personal knowledge in a way that makes it available to others.

Hierarchical Decomposition Solving a problem or understanding a system by reducing it into smaller and more easily solved or understood components.

Microsimulation Computer model of a large number of entities whose behaviors are well understood in aggregate, such as households, and their characteristics such as the number of children per family.

Object-Oriented Programming A computer programming approach emphasizing modularity through use of discrete software objects.

Scenario A plausible story of possible futures or pasts that vary with changes in model inputs or structure.

Stochasticity Random or probabilistic changes in model inputs that vary according to known statistical distributions or empirical estimates.

System Set of real or hypothetical elements interconnected in a manner that gives the system an overall identity and behavior.

Overview

Simulation is a form of computer modeling. Although the term simulation also applies to physical models such as flight simulators or role playing games, the most common definition of simulation refers to the process of translating a conceptual model of some system into a computer model of the system. As with modeling in general, a conceptual model can range from a simple mathematic equation scrawled on the back of an enve lope to a complex natural language description spanning several volumes. A simulation model represents key features of a conceptual model through a software pro gram. This computer model takes data as input, executes its underlying software programming, and produces output data for analysis.

A simulation model represents a system. A system is a set of real or hypothetical elements interconnected in a manner that gives the system an overall identity and behavior. Systems can range from the atoms in a mol ecule to people in a city to galaxies in the universe. In creating a model of any system, it is imperative for the modeler to describe germane characteristics of system elements and capture important relationships among these elements. The modeler must also describe system boundaries because a system almost always exists within a larger environment. Regional economies or ecosystems are self contained in some senses, for example, but they have connections to larger economies and environments, respectively.

Simulation offers both advantages and challenges to modelers, discussed below, but in general it is a useful way to extend knowledge of real or hypothetical systems. Simulation allows the modeler go beyond collecting and analyzing data about a system by using these data to build models of the system in a computer. The use of simulation is growing given the increasing amount of digital data, the ready availability of computers, improvements in programming languages and simulation methods, and the increasing numbers of supporting applications such as geographic information systems.

Uses of Simulation

Simulation models are used in a variety of interrelated ways across research, policy, and educational contexts. At their simplest, simulations can explain a system and examine how the system will look in the future or times past, or in many possible futures or pasts. They also support human decision making, especially in systems that are too complicated for any one person to easily understand. The process of simulation can also help elucidate new information by making clear what the modeler does and does not know. Finally, simulation is increasingly used in education and entertainment.

Explanation: Simulation is one way to understand and explain the underlying processes of a given pattern in a system or the potential patterns associated with processes in a system. While some critics of simulation see it merely as a way of encoding existing knowledge, the ability of a simulation to link together data and theory about multiple system components actually allows it to generate better explanations. Simulations of the global climate system or transportation networks, for example, combine the research and data from multiple scientific and policy domains to create a better understanding of how these systems work.

Prediction: Simulations are used to predict aspects of the future state of a system when the simulations can adequately capture some of its underlying processes. Simulation models are regularly used to project the complex properties of human systems such as populations, cities, or economies. In an urban setting, for example, modelers can use socioeconomic predictions to answer questions like how many schools to build in a city or where to route transportation and infrastructure to minimize future ecological harm.

Simulation is also a means to recreate past events in a manner analogous to predicting future ones. This use of simulation is especially useful for understanding how past cultures or civilizations changed, or in cases such as the Classic Maya in modern day Central America or Anasazi peoples of the present day southwest United States, collapsed or disappeared.

Prediction and retrodiction often imply that simulations are designed to identify a single past or future outcome, but simulation is often used to construct different scenarios, or plausible stories, of possible futures or pasts that vary with changes in model inputs or structure. Scenarios have been used to assess the effects of different irrigation schemes on a dam project, for example, and to examine the impacts of variations in future climate change on global sea levels and agriculture.

Simulation can supplement or replace expert knowledge in decision making. This use of simulation is widespread in industry and commerce, for example, where it is used to help plan product rollouts or guide the choice of materials in manufacturing. There is also a growing number of simulations designed to allow various publics and policy makers come to better informed decisions about situations like responding to disease pandemics or siting of nuclear power plants.

In order to build a simulation, a modeler must understand much about the system being simulated. At the same time, the process of creating a simulation model also helps the modeler order and integrate current knowledge. Simulation can also identify gaps in current knowledge and thereby help prioritize research needs. Finally, simulation also provides, and relies on, approaches such as qualitative interviews or expert elicitation to encode personal knowledge in a way that makes it available to others.

Simulation has long been used for educational and entertainment purposes, and the growth of simulation in general has increased these uses as well. Simulation has long been used in the military, industry, and medicine to train people in situations where actual activities are difficult, dangerous, and expensive (such as flying airplanes, engaging in combat, and conducting surgery, respectively). Simulation is also used in schools and universities to provide a form of 'virtual laboratory' to model urban planning processes or changes in the climate.

Simulation Steps

While there is a broad array of simulation approaches, most simulation modeling employs a similar set of steps that include establishing the problem, designing the system, creating the model, conducting model runs, and analyzing the outputs.

First, it is critical to establish which problem the model will address. As all modeling involves simplifying features of a real or hypothetical system, it is important to ensure that clear questions are posed of the model. Instead of asking "What will this city look like in the future?" the modeler can ask "How will increases in population change the balance between residential and commercial land use in thirty years?" or "What blend of agricultural and residential land use will best protect the urban songbird population?"

Second, system definition follows from problem definition. In order to project future population and land use, the modeler must specify the key demographic and economic conditions that influence land use, and from these, important entities (such as firms and households) and their interrelationships (like travel to work or demand for land). The modeler also defines the system boundaries and determines influences, such as interest rates or immigration, that are external but still important to the system.

Third, the modeler must create the model, including gathering input data, choosing a method or combination of methods, and establishing desired outcomes, such as maps and figures. The modeler must also evaluate the model to ensure it works correctly; this process is detailed below.

Fourth, the modeler runs the simulation. This process is similar to experimentation in a laboratory, where inputs are modified to produce a range of outcomes. While simulation can be prescriptive, or used to determine optimal system configurations, it is more often used for scenario generation. Scenarios ask a range of 'what if' questions by repeatedly changing one or more inputs and seeing how the simulated system subsequently changes. Sometimes these changes are stochastic, which requires introducing random changes into the inputs according to well understood statistical distributions or directly reflecting the variability found in reality.

Finally, the modeler analyzes model outputs in light of the original problem specification – did the model answer the question asked of it? Analysis also involves examining simulation outcomes in aggregate to understand how the system as a whole changes in response to varying inputs. In some cases, the model will create a suite of largely similar outcomes, while in others there may be two or more sets of drastically different outcomes. At this point, it is important to seek expert opinion to establish whether these divergent outcomes are possible or whether the model needs improvement. The modeler will often employ other methods, such as statistical modeling or graphical analysis, to analyze simulation model output.

It is important to note that this five step process is highly idealized. In most cases, the modeler must shuttle between the various steps as the problem changes, new questions arise, certain types of data cannot be easily collected, or new methods become available. Many of the advantages of simulation modeling arise not from running the model to create outputs, but from the process of creating the model itself. Establishing a reasonable problem and defining the system, for example, are two activities that are often very worthwhile in and of themselves.

History and Approaches

Computer simulation in the natural sciences and engineering extends almost as far back as the invention of the digital computer in the 1940s. The earliest significant use of simulation was in government sponsored weapons development, meteorology, and hydrodynamics. It was not until advances in computer hardware and software became widespread in the 1960s that researchers began to use simulation methods on a large scale, particularly in the natural sciences and the social sciences following soon thereafter. The private sector was also quick to adopt simulation, particularly in heavy industries and manufacturing such as aerospace and defense. It was also during this time that models began to be used, or at least encouraged for use, by policy makers for urban, transportation, and social services planning.

Three approaches came into their own in this period: mathematical methods, statistical models, and microsimulation. First, mathematic equations have long been used across the sciences to model systems assumed to be static or existing in equilibrium. 'Spatial interaction models', for example, are used to estimate population flows among cities. When the inputs are varied probabilistically, a mathematical model becomes a simulation model that can create scenarios. Second, simulation models also rely on statistical methods that derive mathematical equations from empirical observations. These equations may then be applied to a variety of new inputs to simulate a range of outcomes. 'Space–time modeling with statistical approaches' is an example of the use of statistical methods. Third, since the 1950s, microsimulation approaches have been used, which are essentially collections of formulas that predict changes over time in large collections of statistical data. They portray systems with a large number of entities whose behaviors are well understood in aggregate, such as households, and their characteristics such as the number of children per family or life expectancy of the head of household. Changes in these characteristics, such as a changing birth rate or increased longevity, can be traced in the artificial population over a series of simulated years. Microsimulation was, and still is, used in planning large scale systems. This approach is particularly popular for national scale governmental issues such as social service provision or public pensions.

In the 1970s and 1980s, simulation methods improved but there was a good deal of debate in the academic and policy communities about the use of simulation, particularly about whether it was advisable or even possible to simulate complicated systems such as cities or economies. While simulations on the whole became more sophisticated and varied, there were some high profile disappointments in planning and predicting urban land use, economic change, and population growth from local to global scales. In many respects, however, these failures were less a problem with the simulations than with the way they were oversold as catchall solutions to a host of problems. Any model, let alone a simulation, is only as good as its underlying data, methods, and theory. Few real world systems of interest to scientists or planners, such as economies or cities, have a single, well established body of data or theory that can undergird the model. In these situations, the model is perhaps best used as a way of extending theory, establishing new methods, identifying possible futures, and establishing research and data priorities.

System modeling became popular during this time of foment and growth for simulation models. This approach remains one of the most common methods, to the point where system modeling is synonymous with simulation for many practitioners. System models are based on a series of difference equations, or those in which each equation relies on the results of other equations, calculated in discrete time steps, that in turn are often tied together with intermediary functions and data. System models focus on the stocks and flows of information, material, or energy in a system. They are applied to a broad range of problems in human geography and beyond, such as urban or population modeling, and have been an important approach to modeling human–environment and natural systems since the 1950s.

From the 1990s onward, greater amounts of data, better modeling tools, and increasingly inexpensive computers ensured that simulation became a mainstay in many disciplines and industries. Geography has long led the way in spatially explicit simulation, which is increasingly common with access to geospatially referenced data from sources such as 'remote sensing' or 'census mapping'. Associated technologies include 'graphical computer programming', 'space–time model ing with statistical approaches', 'spatial data mining', and 'geographic information systems'.

This last period saw rapid growth in the number and variety of simulation methods, including 'artificial intelligence', 'expert systems', 'neural networks', 'evo lutionary programs', 'cellular modeling', and 'agent based modeling'. While most of these approaches have been available for some time (often in more rudimentary forms), it is only in the last decade or so that they have seen broad use across the natural and social sciences. Much of this work is linked to an attendant growth in the use of 'chaos and complexity theory'. Artificial intelli gence and expert models are designed to have computers act in a way that mimics human intelligence. These approaches use logical rules combined with data to allow the models to simulate novel situations. Conversely, neural networks and evolutionary programs are pat terned after biological phenomena. Artificial neural net works are computational analogs of neural structures in animals (including humans) that evolve over time to recognize relationships between inputs to the simulation and desired outcomes. The related evolutionary pro gramming approach creates computational models of Darwinian evolution to similarly link inputs to outputs. These approaches are used, for example, to predict land use as a function of social and environment factors or translate remotely sensed imagery into land cover. Cellular models examine interaction among cells in a lattice that represents phenomena ranging from vege tation to house occupancy. Cells evolve over time according to their current state and that of surrounding cells. The related agent based modeling approach uses collections of agents, or software programs, that represent adaptive autonomous entities that use information from their surroundings to inform perception, planning, and learning. Key to cellular modeling and agent based modeling is their focus on how local interactions among cells or agents can create complex systems.

The future of simulation is necessarily difficult to predict, but it is likely we will see a continued trend of creating simulations that are fully integrated with the system being modeled. We increasingly rely on computer models to guide system behavior in real time, such as routing package delivery, controlling highway traffic, or monitoring communication networks such as phone systems or the Internet. Many cities use traffic simu lations, for example, to integrate real time inputs such as weather data with traffic flow information from sensors embedded in roads. These simulations create short term projections of traffic flow and provide guidance on the timing of stop lights and lane reversals.

In sum, given the complexity and range of systems simulations can model, simulation approaches run the gamut from the relatively simple to the extraordinary complex. While all approaches rise and fall in popularity over time, each offers distinct strengths and challenges to use. In addition to modelers continually improving in dividual methods, there is also a movement toward better integration and hybridization of different approaches. Models of urbanization and land use, for example, in creasingly combine traditional approaches such as stat istics and system models with more recent approaches such as agent based modeling or cellular modeling.

Critical Evaluation and Challenges

Simulation models offer a number of advantages to work in research, policy, and education while also posing significant challenges. Simulation is first and foremost a means to develop new theories, abstract features of reality, and create knowledge. Much of its power derives from offering a tractable means of incorporating sto chasticity and conditionality in an efficient and modular way. Simulation is not without challenges. Evaluating simulation models is a difficult process, often due in part to their great data needs. Simulation must also account for sensitivity to changing conditions and the issues posed by modeling systems across spatial, temporal, and organizational scales. Finally, one of the greatest chal lenges to simulation is adequately conveying exactly what the model does and why.

Advantages

Simulation is considered a 'third way' of doing science because it falls between induction (generalizing from empirical observations to theory) and deduction (testing theory against empirical observations). The classic mode of science holds that a theory will lead to testable hypotheses that we evaluate against observations of reality. This model of science holds true for many situ ations in the natural sciences because theories can be expressed as mathematical formulas and tested against empirical observations. This approach is less effective in social science or for complex natural systems, however, in which case simulation models act as a bridge between reality and theories that are too complex to express as mathematical hypotheses.

A key advantage of simulation is the ability to model possible outcomes for situations that do not exist in reality, exist in the future, or for situations that are dangerous or expensive to perform, such as testing nuclear weapons. Simulation is therefore increasingly forwarded as a way to complement, and perhaps even replace, some kinds of experimentation that are impractical. Mathematics or natural language may also be used in this manner, of course, but simulations do it better. Large urban or climate simulations, for example, offer insights into the future of these systems that are not available from other approaches.

Like other scientific methods, simulation can be used to create knowledge. As noted above, in order to create a simulation, the modeler must identify the necessary data, which in turn leads to consideration of how to meet data needs. When a simulation encapsulates the state of knowledge in a field and yet does not replicate reality (i.e., is wrong), it can help define new hypotheses and lines of research. Simulation similarly points out the limits of current knowledge and data.

Simulation allows the modeler to integrate a large amount of data and information on relationships between elements in a system. At the same time, simulation can address problems that are analytically intractable when expressed as mathematic equations. While less expressive than natural language, simulation modeling is more flexible than most mathematical or statistical methods. For this reason, simulation is sometimes termed the third symbol system of science because it complements the first symbol system, natural language, and the second, mathematics. In many respects, simulation also acts as a bridge between natural language descriptions of systems and their mathematical expression.

Simulations can incorporate random noise or a range of values for individual parameters in a way unavailable to mathematics. Unlike natural language, simulation allows the modeler to consider the effects of a range of subtly different inputs on complex systems. This allows the modeler to examine the effects of changing inputs on outcomes and, importantly, the amount of possible vari ability in these outcomes. Importantly, some simulation models are also able to capture how minor changes in inputs can create disproportionately large changes in outcomes. This is considered below in greater detail.

Simulation is an ideal way to examine conditional, partial, and parallel processes because multiple simulated processes can occur almost simultaneously and interact accordingly. The rise of object oriented programming and attendant approaches such as agent based modeling has allowed simulation models to represent the inter actions among many entities. Consider the modeling of foot traffic in a city by simulating thousands of autono mous pedestrians or highway commuting by simulating hundreds of thousands of vehicles. These models can capture how a single local interaction between indi viduals, such as pedestrians bumping into one another or two automobiles having an accident, will scale up to create system wide gridlock.

As noted above, simulations are increasingly inte grated with, and serve to control, the systems they model. More broadly, simulations can be a cost effective way of designing real world systems by offering a means to create virtual prototypes. Industry in particular uses simulation to help design products in the computer before creating actual physical prototypes. This approach is particularly important in civil, automotive, and aero nautic engineering because these industries build very complex objects, such as vehicles or planes or bridges, that must safely meet a variety of needs.

Simulation models are often modular, or comprise multiple interconnected pieces. Modularity supports what Herbert Simon termed hierarchical decomposition, or taking a divide and conquer strategy to problems. The model can be changed or updated by changing one component without having to change all other com ponents. With methods such as agent based modeling, this modularity extends to individual model elements such as people or vehicles.

Challenges

Evaluation is a broad term that encompasses simulation calibration, verification, and validation. These equate to, respectively: fitting the model to data and theory; determining whether the model runs well; and com paring its structure and outputs to data not used during calibration. Mathematic models are generally axio matically correct by definition, while simulations can rely on computer code susceptible to errors and poor pro gramming. And unlike most mathematical models, simulation models can create a large number of outputs, especially when multiple runs or scenarios are generated. Validating these outputs against empirical evidence or other models is difficult and time consuming.

Simulations can require a large amount of data, which follows naturally from a methodology that models many facets of a complex system. Simulations often rely on parameters (e.g., population growth rates, taxes) that can often be derived from actual data, but sometimes are the result of educated guesses. An example of the potential problems with this approach is seen in The Limits to Growth, a report written in 1972 that used a simulation model to project the future of global population and economic systems, sometimes incorrectly. While Limits to Growth has been often unfairly critiqued, one of the drawbacks noted by the authors themselves was the use of expert opinion to estimate critical parameters.

Complex systems in reality and in simulations can exhibit sensitivity, or large changes in system behavior that occur in response to relatively small shifts in inputs. This attribute of real and modeled systems complicates the creation and use of simulation models because many methods assume that incremental inputs changes will lead to similar changes in outputs. Complicated situ ations therefore require more sophisticated testing and evaluation that accommodate sensitivity.

Simulation gives a variety of ways to examine social, temporal, and organizational 'scale'. Key to this examina tion is the idea of emergence, or the process by which complex systems gain qualities that result from syn ergistic interactions among system components. To continue an example from above, there is nothing intrinsic to vehicles in highway traffic that creates grid lock, but the interactions among these vehicles can readily lead to gridlock. In part derived from chaos and complexity theory, such systems are termed self organizing because they can change their structure over time in response to either internal or external changes. Self organization and scale also complicate the notion of causality in real and model systems because neither system characteristics nor their causes are known before hand. The ability to model self organization makes simulation a valuable approach to modeling complex systems, particularly with methods such as agent based modeling, cellular modeling, evolutionary programming, and artificial neural networks. The key challenges of scale for any method lie in matching data collected across scales (e.g., household information vs. census information vs. regional data) and linking entities across scales (e.g., how do households interact with social systems?).

One of the most challenging aspects of simulation is the need for clarity in model structure and the way in which this structure is communicated by the modeler. One key advantage shared by both natural language and mathematics is that most people are conversant in one or both of these symbol systems. In contrast, a large range of specialized programming languages and systems are used to create simulations. To this source of opacity is added the fact that simulations are often used to model many features of very complex systems, to the point where even a moderately sophisticated simulation can have hundreds or thousands of individual model com ponents. In these complicated models, it is very difficult for the modeler to convey basic features of model com ponents, let alone their interconnections. With the growing importance of simulation there is a movement toward increasing model transparency through common user interfaces and reporting requirements. A significant obstacle to these changes is the wide variety of situations in which they are used, which range from academic researchers to the private companies to public agencies.

Conclusion

Simulation via computer modeling has been, and will remain, an important approach to understanding a wide variety of human, human–environment, and natural systems. It offers a way of converting conceptual models of real and abstract phenomena into software programs, allowing us to combine existing data and theory to build virtual worlds. Simulation will likely continue to grow in popularity and utility because it can help researchers and policy makers take advantage of the growing amount of digital data, readily available computers, improved simulation methods, and varied cognate approaches such as geographic information science and artificial intelli gence. While simulation will continue to offer a number of key advantages, ranging from theory building to ease of modeling, simulation modelers will continue to address an array of challenges that are shared by human geog raphers in general, such as scale or data needs.

See also: Artificial Intelligence and Expert Systems; Census Mapping; Chaos and Complexity; Spatial Data Models; Neural Networks; Remote Sensing; Scale; Spatial Data Mining, Cluster and Pattern Recognition; Spatial Data Mining, Geovisualization; Spatial Interaction Models.

Further Reading

Batty, M. (1994). A chronicle of scientific planning. The Anglo American modeling experience. *Journal of the American Planning Association* 60(1), 7 17.

Gilbert, N. and Doran, J. (1994). *Simulating societies: The computer simulation of social phenomena*. London: UCL Press.

Gilbert, N. and Troitzsch, K. G. (1999). *Simulation for the social scientist*. Buckingham: Open University Press.

Goodchild, M. F., Anselin, L., Appelbaum, R. P. and Harthorn, B. H. (2000). Toward spatially integrated social science. *International Regional Science Review* 23(2), 139 159.

Hanneman, R. (1988). *Computer assisted theory building: Modeling dynamic social systems*. Newbury Park, NJ: Sage Publications.

Hannon, B. and Ruth, M. (1994). *Dynamic modeling*. New York: Springer Verlag.

Jackson, L., Trebitz, A. and Cottingham, K. (2000). An introduction to the practice of ecological modeling. *Bioscience* 50(8), 694 706.

Landis, J. (1994). The California urban futures model: A new generation of metropolitan simulation models. *Environment and Planning B: Planning and Design* 21(4), 399 420.

Lee, D. B. (1973). A requiem for large scale models. *Journal of the American Institute of Planners* 39, 163 178.

Meadows, D. H., Meadows, D. L., Randers, J. and Behrens, W. W. (1972). *The limits to growth*. New York: University Books.

Parker, D. C., Manson, S. M., Janssen, M., Hoffmann, M. J. and Deadman, P. J. (2003). Multi agent systems for the simulation of land use and land cover change: A review. *Annals of the Association of American Geographers* 93(2), 316 340.

Starfield, A. M., Smith, K. A. and Segal, J. K. (1990). *How to model it: Problem solving for the computer age*. New York: McGraw Hill.

Relevant Websites

http://www.scs.org/
The Society for Modelling and Simulation International.

Situated Knowledge, Reflexivity

A. Kobayashi, Queen's University, Kingston, ON, Canada

Glossary

Epistemic Subject The knowing subject, the subject that constitutes herself or himself through knowledge, as opposed to the use of knowledge in the construction of an object, especially other people, as an object of knowledge.

Gaze A term used especially by postcolonial scholars to indicate the practice through which the colonial power constructs the 'other' by envisioning the other in a subaltern position.

Partial Knowledge Knowledge that is not complete or universal. Postcultural theorists recognize that all knowledge is partial because it comes from a particular vantage point, that such vantage points are constantly shifting, and that the partiality of knowledge is reinforced because it is shared with others.

Reflexivity The practice on the part of a researcher of reflecting upon his or her own identity or positionality in relation to those studied, the recursive practice of self-examination and self-transformation in the exchange of knowledge between the researcher and those studied.

Introduction

The concept of situated knowledge was adopted initially by feminist geographers during the 1990s, in response to historian of science Donna Haraway's definition of 'situated knowledges' as epistemic positions that are partial, embodied, and localized. Geographers have used the concept to deepen their understanding of the ways in which geographical knowledge is produced, emphasizing both the specific social location of, and the spatial interactions among, people.

Since the introduction of the term by Haraway, it has been used in a variety of ways and in many disciplines. Journals of business administration feature articles on how situated knowledge can be used as a marketing strategy by placing people in the field in such a way that their 'grounded' knowledge allows them to develop a better understanding of consumer needs. In contrast, nursing journals emphasize the importance of understanding the role of the caregiver as context specific. Education specialists examine the relationship between social situation and learning. So, situation affects not only the production of knowledge, but also the discursive process through which information is received.

For geographers, the concept of situated knowledge has been used primarily in two ways. The first draws upon the work of Donna Haraway to encourage auto reflexivity in geographical research, acknowledging the subject position of the scholar in relation to 'others'. A second approach draws also from antiracist scholarship, and in particular from the concept of 'Orientalism' introduced by Edward Said, to emphasize the complexity of social landscapes in which 'others' are defined by the process of geographical imagination.

Science as Situated Knowledge

The concept of situated knowledge emerged in intense discussions by feminist theorists through the 1980s and early 1990s over questions of feminist epistemology: Who holds the power to know? What do they know? What is the role of gender in informing epistemology? Concern was over not only who has access to the power of the academy – including access to publication outlets, research grants, and the ear of the public – but also over how the 'situation' of the researcher affects the production of knowledge and the lives of those we study, particularly women and minority groups.

These questions led feminist scholars such as Sandra Harding, Nancy Hartsock, and Dorothy Smith to develop 'standpoint theory', which argues that feminist scholars occupy a privileged epistemic position because of their own social situation as women. Social and epistemic positioning are interrelated, so that 'who' knows is just as important as what they know. This approach is both post Kuhnian in its emphasis on the dynamic nature of knowledge creation, and Marxian, based especially on Georg Lukacs' contention that privilege (of the capitalist class) impedes the ability to understand exploitation, while the underprivileged (workers) possess a dual vision of both their own lives and those of their oppressors; thus social position is inversely related to epistemic position.

The original standpoint theorists have been criticized for a number of reasons, including their reliance on the coherency of gendered perspectives as opposed to other perspectives, and their tendency toward essentialism by privileging gendered knowledge. Postmodern feminists claimed that standpoint theory fails to take account of the 'multiple' positionalities, of women and others, that affect the process of epistemic privileging. Nonetheless, feminist geographers and others have been strongly influenced by the ability of standpoint theory to differentiate one knower from another, and by the recognition that all knowledge is context specific, or situated. These points provide an important conceptual basis for under standing that all knowledge has a geography. It occurs somewhere, grounded and placed.

Recognizing scientific knowledge as situated is more about context than location. Donna Haraway's path breaking contribution was to emphasize that all know ledge is a product of the context in which it is developed. She coined the term 'situated knowledges' in a 1988 essay (and a subsequent book) that claims that all knowledge is local, partial, and embodied. Masculinist and racist science performs the 'God trick' of establishing itself outside the lives of its subjects, upon whom it visits a neutral, impartial, disembodied, and unmarked gaze. A critical epistemology, in contrast, situates the knower as an embodied subject in relation to other knowers. The act of situating knowledge is often paradoxical, however, because partial knowledge means that situation is never complete: to resort to a belief in the fixity of situation is to revert to the God trick. Rather than (like the original standpoint theorists) positing an alternative position from which to construct yet another form of privileged epistemology, therefore, Haraway calls for a recognition of partial knowledge as a basis for redefining 'objectivity' as shared, mediated knowledge founded on political alliances, and as a means of developing critical under standing of the significance of social location and its contingencies.

All Views Are Views to/from Somewhere

All views are views 'to somewhere'. Landscapes, the stuff of geographical knowledge, are both constituted by geo graphical vision and the constituent matter that human beings organize. Landscapes are places placed. Although this point has strong antecedents in the discipline, particularly in the Berkeley School tradition of cultural geography, it is the situation of the researcher within the landscape that differentiates recent critical approaches and their imperative to become more reflexive. For geographers, the concept of situated knowledge has been adopted along with the principle of social construction to emphasize that geographical knowledge is always socially produced. Geographical vision–knowledge is partial, shifting, contingent, and powerful – and has profoundly affected the ways in which geographers see the world and convey their understanding to others. The situation of geographical vision–knowledge provides a vantage point from which to observe the world, actuating a gaze that both positions others and establishes the observer within a discursive field. That gaze is always partial, shifting, and always embodied. Geographical vision is also a view from somewhere.

The analytical power of the concept of situated knowledge in geography is that it elucidates the inter secting ways in which both researcher and subject, knower and known, are embodied, positioned, and empowered within specific landscapes. The point of understanding situated knowledge as double vision, to and from a particular vantage point, is of course that there is a recursive relationship between knower and known, and that detachment of one from the other is impossible, however much we may wish to step outside of the embodied relationship. Nonetheless, the genealogy of situated knowledge within the discipline of geography is such that there has been a tendency to emphasize either one or the other, the epistemic production of the self, through reflexivity, or of the other, by understanding the historic construction of difference.

Situated Knowledge as Reflexivity

To be reflexive is to reflect upon the self in a self conscious examination of the relationship between the researcher and the researched. Self reflexivity is a deliberate attempt to situate one's own knowledge, to emphasize that such knowledge is not detached from the world, that the researcher embodies a range of affects, social meanings, and power. It asks not only the epi stemological question of how one's own situation affects the production of knowledge, but also the moral question of what right one has to speak about, to, or on behalf of 'others', especially when those others occupy marginal social locations that render them less visible or less likely to be heard.

Situating the body of the research through self reflection poses some difficult moral and theoretical challenges. There is a growing number of geographical texts that begin with an apologetic situation of the author(s) as white, male, able bodied, and privileged. While such statements indeed call attention to the fact that geography is a social construction dominated by white males who have had a secure hold on the pro duction of knowledge at least since the Enlightenment, such attention does not necessarily 'change' the fact that the bodies, if not the ideas, of geographers are still white, male, able bodied, and privileged. Such texts still reinforce the authority of dominant voices.

Equally problematic has been the tendency to use reflexivity to shift attention from subjects in the field to the researcher himself or herself. The geographical literature in recent years contains no small number of such articles, in which privileged academics reflect upon intellectual affect, often in beautifully written, evocative prose about what it is like, how it feels, to conduct research, and what kinds of moral and political dilemmas are encountered. Such musings can play an important role in questioning, and defining, the kinds of research in which geographers engage, and the kinds of subject they choose to study. Situated knowledge does not just happen, in other words, but involves active situation of the researcher, making transparent his or her subjectivity. To situate is to act.

But an ironic aspect of such self reflexivity is that it tends to 're place' the focus on the researcher himself or herself, thus privileging the researcher even more by centering knowledge production, albeit with the 'acknowledgment' of the power that the researcher embodies. Reflexivity runs the risk, therefore, of becoming a kind of academic self indulgence used to justify the fact that change within the discipline occurs only slowly. This conundrum has led some geographers to try to develop methodologies for 'turning the gaze' of others back upon the researcher, making the body of the researcher the site of social contestation. Some, however, have claimed that even such attempts run the risk of recentering the privileged body, thus emphasizing the otherness of the other.

Such reflexivity, therefore, is an extremely weak political act. Indeed, to continue to occupy the position of power, however reflexively marking the unmarked body, is to maintain the power of difference. Reflexivity is therefore, for most critical geographers, only a point of departure for political action, as self reflexive scholars are above all concerned about the potential for recreating or reinforcing the forms of social exclusion that they study, starting with the differences of class, race, sexuality, or ability that separate them from their subjects, focusing attention on how to overcome the distance and on the ethical questions of speaking for/with/as marginalized others. Reflexivity thus demands political, moral, and personal engagement with the lives of others and the conviction to make a difference.

A more powerful way to resituate the body of the researcher is through direct political activism. Growing numbers of geographers now engage in critical social action research, in which they insert themselves into community projects as only one among many working for a common end. Social action research involves going beyond self reflexivity to engage in mutual reflexivity as the researched become researchers. Such attempts have mixed results, and the lessons learned are often intense. The projects often involve considerable investments of time and emotion, and can become risky as the researcher gives up almost completely on the notion of controlled detachment over an agenda set by others. Some have found that despite the explicit, Haraway esque recognition that reflexive knowledge is partial and shifting, the distance from some subjects is too great, and the points of common identity too fleeting, for sustained engagement. For action, research requires more than a commitment to political activism; it also requires social acceptance, credibility, and some degree of common experience, and mutual reflexivity is nothing if not an exploration of those commonalities, and a critical analysis of how they provide a basis for successful political action.

Still others, however, have found that social action research can at times achieve positive political ends, either because it is emancipatory and empowering for those involved or because it attains some specific goal, such as changing the direction of an urban renewal project, or bettering conditions for workers, immigrants, racialized groups, or others situated in marginal social positions. Successful projects also involve risks, of course, as researchers may find that the contingency and spontaneity required of successful political actions results in a remaking of the self; the geographer is thus resituated through the process of the research in ways that may be unsettling, morally provocative, and certainly epistemologically challenging, often involving anxious negotiations that tend to unsettle any pretense to complete knowledge. The resituation of the self always involves the ongoing reformulation – reflexion – of knowledge.

It is not only previously unmarked bodies that are called into question through reflexivity. As the academy has slowly become more diverse, an increasing number of geographers situate themselves and their analyses from the perspective of marginalized groups, writing as women of color, as disabled women, and the like. Their positioning further blurs the boundaries between inside and outside the academy, and offers significant analytical benefits for research that is firmly grounded in the community and based on deeply experienced participant observation. It also sets a challenge, however, not to repeat the error of standpoint theorists by positing new forms of epistemic privilege. To adopt an essentialized epistemic position based on identity is to deny the very process through which identity is socially constructed, multiply positioned, and contingent. To carry such logic to its furthest extreme, moreover, is to embark on the slippery slope of solipsism: if representational authenticity depends only upon common identity, then the separation of knowers within any discursive field, their alterity, invalidates the ability of any one knower to speak or to know on behalf of anyone but her or himself. To fall into that trap is to reinvoke the God trick from the other side, distancing oneself from one's subjects not by objectivity but by subjectivity.

Recognizing situated knowledge through reflexivity is thus deeply ironic and paradoxical. Any attempt to refashion the relationship between knower and known by acknowledging the power of the knower inevitably results in a repositioning of the knower, and emphasizes the contingency of that which is known. To succumb to the seductive appeal of reflexivity as a means of understanding the self, therefore, reinforces the extent to which all knowledge – including that of the self – is partial.

Situated Knowledge as Geographical Imagination

A second approach to situated knowledge emphasizes the construction of the other as a historical project through the production of imagined geographies. This approach entered geography not through feminism directly but through postcolonial and antiracist studies, influenced especially by the work of the literary theorist Edward Said. Rather than focus (at least not primarily) on the role of the researcher as a reflexive subject, that is, the construction of the 'self', it addresses the powerful discursive forms through that support the construction of the 'other'. The most in depth work by a geographer from this broad perspective is Derek Gregory's *Geographical Imaginations*.

Said's concept of orientalism has been engaged by geographers to show that colonialism is not only a particular set of political economic practices but also a system of power–knowledge, 'situated' in such a way as to deploy a set of profoundly powerful discursive formations, or metanarratives, through which the 'Orient' has emerged as a place. That place lies beyond the everyday knowledge of the West, yet is broadly implicated in the ways in which the West imagines both itself and the other, the oriental. Orientalism as power–knowledge has been carried forward to the present, resituated in present day relationships between the West and the other, as expressed in an immense and complex array of cultural practices, socioeconomic relations, and political realities. Most importantly, the present day construction of the 'Middle East', as a product of Orientalism and a site of violence, can be mapped through a cartography of difference produced under an orientalist gaze from colonial times until the present.

This concept is particularly compelling for geographers because it contains the important notion that orientalism, like all metanarratives, is also a spatial formation. Geographical writing inspired by Said's insights has consistently emphasized that spatial formations need to be understood not only in a metaphorical sense (recognizing of course that the power of metanarratives is often metaphorically fueled). Said himself acknowledged the spatiality of the colonial relationship, and indeed the spatiality of all discourse. He wrote little, however, about either the experiences of people embodied by colonial knowledge, or the specific material ways in which spatiality is lived out in particular landscapes. That project has been taken up, often passionately, by geographers.

Situated knowledge is about the positioning of the heterogeneous knowing subject. Geographers insist that the heterogeneity of subject positions is simultaneously epistemic, powerful, and 'spatial'. Heterogeneity, or alterity, in itself provides the condition for multiple forms of social and spatial outcomes. Subject constitution – or its obverse, object construction – is therefore a contingent process. We are not concerned, however, with contingency as some kind of meaningless, myriad swirl of outcomes but, rather, with the specificity of place and relative situation, or context, through which human relations and meanings are negotiated, resisted, and mediated in multiple ways. The gaps and disjunctions of knowledge, its transformations, its partiality, and its fragmentation, do matter as much as its consistency, hegemony, and continuity across and between real places in which knowledge is situated. People are bounded, transported, connected, and made visible, as situated bodies that are simultaneously knower and known. Their knowledge, however, to repeat, is always situated within a landscape, as a vantage point to and from somewhere. People are positioned within a discursive field, in other words, that influences what they say and do, the power that they wield, with whom they speak, the extent of their inclusion or marginality. It matters, moreover, 'where' they are positioned, in buildings, streets, or cities, in what part of the world, and with whom they share proximity, either through direct bodily contact or through the myriad ways in which knowledge is transmitted spatially, shared partially, and embodied meaningfully.

The 'discursive turn' of the 1990s gave geographers the opportunity to locate knowledge within the production of texts, broadly conceived as stretching from words to landscapes, or any form of representation that carries meaning, and can be 'read'. The post structuralist recognition of knowledge as socially constructed has been extended in explicit discussions of the ways in which knowledge, and the human relations and forms of social organization through which knowledge is mediated, are situated. Geographers focus on landscapes, as complex material constellations in which meaning and form are inseparable, in order to ground, or spatialize, the discursive formations that Michel Foucault has shown to be so important to the invention of modern institutions and cultural practices. Despite important differences in their thinking, both Said and Foucault contributed to our understanding that the construction of the other, through practices of surveillance, organization, naming, subjugation, and other means of knowledge–power, is at once an imaginative engagement and the imposition of a normative gaze upon/within a specific landscape. Imaginative geographies thus constituted are simultaneously limitations upon human knowledge, and engagements with the possible.

Geographers drew similar inspiration from Benedict Anderson's notion of 'imagined community' (often erroneously interpreted as the 'invention' of the nation as though a nation could emerge through some sort of idealized act of will) as a set of epistemic bonds through which the nation is made both sovereign and bounded, a place united not (or at least not necessarily) by direct human relations, but by common understanding of what national belonging entails. Imaginative geographies in this sense carry broad implications for who resides inside and outside places of privilege, and for understanding the spatial strategies according to which situated knowledge is engaged to produce social difference.

Like the practice of reflexivity, the practice of geographical imagination is a contingent process that can have many outcomes. The geographical literature is replete with examples of how geographical imaginations have been mobilized historically in constructing the excluded other, as well as the ways in which present day practices of envisioning difference respatialize colonial relations in many ways. In particular, in the times since 11 September 2001 in which practices of imagining national identities across global distances have become particularly intense, geographers have responded with a flurry of studies that emphasize some specific ways in which knowledge is being mobilized to embody war, to envision the normal, and to spatialize difference. Such studies are paradoxical: while chronicling the effects, often horrific, of geographical imaginations, thus situating knowledge within a contemporary context, they are at the same time engaged in reformulating geographical imaginations to make a 'different difference'.

To return to Gregory, geographical imaginations comprise the history of a discipline that situates knowledge through specific cartographic acts: visualizing the world, embodying its inhabitants, in conjunction with particular forms of spatial organization such as colonialism, or under the rubric of metanarratives such as modernism or postmodernism. The challenge to geographers in the present is to advance both the epistemic project of understanding the relationship between knowledge–power and spatiality *and* the moral project of making the act of corporealizing vision, one that elucidates the lived results of past geographical imaginations and projects new ways of seeing, in other words, resituating knowledge.

Situated Knowledge and Moral Geography

The development of the concept of situated knowledge in geography has involved a moral engagement that goes back for several decades. During the 1970s, geographers engaged in what later became known as 'relevance debates' over the significance of geographical knowledge for human being in the world. Those debates emerged simultaneously out of a general uneasiness with the abstraction of geography as a science of 'space', based on the epistemological arrogance of objectivity and detachment. They emerged also from the social conditions of the 1970s: the anti war movement, the civil rights movement, the upsurge of social concern on the part of students the world over, and from realization of the horrors of life frozen in constant Cold War. Since that time, human geographers have become more and more epistemologically critical, influenced both by the discursive turn of the 1990s and by more recent calls to emphasize the material conditions in which discursive events take place. Recent geographical writing has become much more explicit about the connections between the situation of knowledge and the moral agenda of the academy.

The production of geographical knowledge is thus simultaneously a moral and a political project. The discipline of geography has always been in this sense a moral project, even at its most abstract moments of embracing the science of space. But critical epistemologies, especially, but not by any means exclusively among those engaged in feminist and antiracist scholarship, have shown that situated knowledge is more than simply the 'location' of knowledge – it involves the act of situating, and thus positioning, embodying, and engaging the knowing self and the known other. The two strands of geographical thought outlined above, emphasizing on the one hand the need for reflexivity in situating the self, and on the other hand the implications of engaging geographical imagination to position the other within a discursive field, are not of course exclusive. They are part of a common project of revisioning the world, resituating knowledge, and reimagining geography as a moral project.

See also: Content Analysis; Culture/Natures; Monte Carlo Simulation; Neoliberalism and Development; Tourism, Rural.

Further Reading

Code, L. (1981). Is the sex of the knower epistemologically significant? *Metaphilosophy* 12, 267 276.

Code, L. (2006). *Ecological Thinking: The Politics of Epistemic Location*. Oxford: Oxford University Press.

Gregory, D. (1994). *Geographical Imaginations*. Oxford: Blackwell.

Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies* 14, 575 599.

Haraway, D. (1991). *Simians, Cyborgs, and Women: The Reinvention of Nature*. New York: Routledge.

Harding, S. (1983). Why has the sex/gender system become visible only now? In Harding, S. & Hintikka, M. (eds.) *Discovering Reality: Feminist Perspectives on Epistemology, Metaphysics, Methodology, and the Philosophy of Science*, pp 311 324. Dordrecht: D. Reidel.

Harding, S. (1991). *Whose Science? Whose Knowledge? Thinking From Women's Lives*. Ithaca, NY: Cornell University Press.

Harvey, D. (1996). *Justice, Nature and the Geography of Difference*. Oxford: Blackwell.
Kitchin, R. M. and Hubbard, P. J. (1999). Research, action, and 'critical geographies'. *Antipode* 31, 195 198.
Kobayashi, A. (1994). Coloring the field: Gender, 'race', and the politics of fieldwork. *Professional Geographer* 46, 73 80.
Merrifield, A. (1995). Situated knowledge through exploration: Reflections on Bunge's 'geographical expeditions'. *Antipode* 27, 49 70.
Rose, G. (1997). Situating knowledge: Postitionality, reflexivities and other tactics. *Progress in Human Geography* 21, 305 320.

Situationism/Situationist Geography

D. Pinder, Queen Mary, University of London, London, UK

Glossary

Dérive Literally 'drift', a form of moving through cities that is usually conducted on foot and in small groups, embodying a critical and ludic attitude toward spaces and their possibilities.

Détournement A practice akin to montage that involves diverting, appropriating, or hijacking elements from cultural forms and discourses, and juxtaposing them to create new and potentially subversive effects.

Everyday Life A field outside of specialized activities that is the common ground of human activities and thought; for the situationists, it is both increasingly 'colonized' by the commodity and the state and also a terrain of resistance and struggle, such that they called for its revolutionary critique and for the recreation of its totality.

Psychogeography The study of urban ambiances and the effects of environments upon emotions and behavior, not simply as a recording or tracing but as an active intervention.

Spectacle A stage in commodity production and the accumulation of capital that assigns particular importance to consumption and the image, and that is marked by the 'colonization' of everyday life and by the alienation of people so that they become more like passive spectators than active subjects.

Unitary Urbanism Presented as a critique of urbanism rather than a doctrine or plan, it envisions the collective creation of the urban environment through the integration of artistic and technological means, and the production of emancipatory spaces that would be dynamic, transcend current divisions and separations, and be lived by their creators.

Introduction

The Situationist International (SI) was a revolutionary organization founded in July 1957 by members of several small European avant garde groups. Until its disband ment in 1972 it developed an incisive critique of the economic, social, political, and cultural conditions of modern societies, in opposition to both capitalism and bureaucratic state socialism. The situationists sought the radical transformation of social relations and spaces, believing that revolution must entail changing not only forms of political rule but also the very fabric and texture of everyday life. It should be a process by which people overcome the alienation, fragmentation, and division characteristic of life dominated by the commodity form, and discover their true selves and desires.

The situationists insisted that to change everyday life it was necessary to change everyday spaces. Radical change would only come about as people transformed their environments and produced spaces appropriate for new ways of living. The central theorist of the group, Guy Debord, therefore described proletarian revolution in his book *The Society of the Spectacle*, published in France in 1967, as a "*critique of human geography*" (1994: thesis 178). Geographical concerns were indeed central to the situationist project and the group's ideas and practices have recently become increasingly well known and influential within human geography, even though they were never meant to court respectability and were forged at the time not as contributions to academic discussion or theory but as weapons with which to harm powerful interests.

The SI is sometimes dubbed the 'last avant garde'. This refers to the militant model of a small group of artists positioned in advance or at the cutting edge of social developments, typically working collectively and internationally, defining its position through manifestoes and radical demands, and seeking to transgress the boundaries between art and everyday life so as to change not merely artistic practice through developing innova tive techniques and modes of representation, but also art institutions and the social world beyond. Whether or not the viability of that model ended with the situationists, it was not a term they usually favored for they rejected the notion of acting as a vanguard or artistic movement. Many of the early situationists were nevertheless artists and their critique of society took up positions within the culture and art of the time, even if they saw their role as undermining existing artistic edifices as they helped to theorize and fuel a new revolutionary contestation.

The SI's roots lay in the historical avant gardes of the early twentieth century, in particular dada and surreal ism. Believing surrealism to have become a largely spent and conservative force by the 1950s, the situationists argued that in its original form it still contained vital constructive possibilities and they took up the challenge of going beyond its efforts to exceed the terrain of art and to change life. While the means of fulfilling this task were the subject of considerable debate, the group came to present it dialectically as the supersession and real ization of art, involving its transcendence as a separate specialist category through the revolution of everyday life. Marxism was crucial in the development of their theory and practice, although their engagement with it was critical as they operated outside of Communist Party strictures and were vehemently opposed to overt or residual Stalinism in the contemporary Left, as well as currents of Maoism that attracted many political radicals during the 1960s.

Since the SI's formal demise, interest in its ideas and practices has continued in radical political and anarchist circles. Elements of its iconography, language, and slogans have also seeped into popular culture, in Britain and North America especially through punk rock. After years of marginalization and neglect the situationists have further become increasingly common reference points in art practice, art and cultural history, and critical theory, as recognition of their significance has grown with a spate of exhibitions and publications since 1989 positioning them alongside other more established twentieth century thinkers and avant gardes. The process of historicization has not been uncontested, though, with some commentators concerned about its selectivity and the loss of political edge. The situationists often combated the silencing, diluting, or misrepresenting of their positions in their own time and repeatedly rejected the term 'situationism', viewing it as a label concocted by those wanting to freeze their ideas into a doctrine that could be more easily incorporated into the reigning cultural discourse. That term's increasing use today, including in the title chosen for this encyclopedia article, could be seen as an extension of the impulse to package them in a respectable format alongside other 'isms' in a way that strips them of their subversive force.

Yet, situationist ideas and activities are not only of historical interest for they also speak powerfully to a range of current issues and political struggles, hence the continuing importance of returning to and reassessing them in the light of contemporary circumstances. Their significance for those concerned with geographical questions centers around three themes: critiques of the society of the spectacle; exploration and contestation of urban spaces, including through forms of psychogeography; and utopianism and attempts to imagine and construct emancipatory urban futures. After brief historical contextualization, these three themes will be discussed in turn.

Critiques of the Spectacle

The SI brought together selected members of two main movements. The first was the Letterist International (LI), which was based in Paris and included Guy Debord and Michèle Bernstein. The second was the International Movement for an Imaginist Bauhaus (IMIB), which had established an experimental laboratory in Alba, Italy, and included the artists Asger Jorn, Pinot Gallizio, and Constant. Also participating in the founding was Ralph Rumney of the London Psychogeographical Committee, an 'organization' of which he was the only member and whose title he had invented after encounters and discussions with Debord and other letterists in Paris. The SI had a total of 72 members from 16 different countries. They employed a variety of means to develop and disseminate their positions, central among them being writing in the form of pamphlets, books, and their journal *Internationale situationniste*, which was published in 12 issues between 1958 and 1969 (since reissued in a complete facsimile edition). They also produced exhibitions, films, paintings, maps, models, graphics, collages, reworked comics, and a range of interventions, provocations, and political practices.

The group was intentionally small and tightly organized as it strove to counter hierarchies and avoid ideological stasis, and to find new modes of political organization. It rejected the idea of 'disciples', wanting instead to encourage self determination through actions by autonomous revolutionary organizations. Internal disagreements included those over the role of artistic practice. Most of the practicing artists were expelled or had resigned by 1962, with a number regrouping as a Second Situationist International around Jørgen Nash and others, who had founded a Movement for a Scandinavian Bauhaus Situationist at the Drakabygget farmhouse, in Sweden. Histories of the SI typically present this schism as a move from an artistic and urban phase to a political and theoretical one, centered increasingly on Debord in Paris. Yet, although there were changes in personnel and approach, such characterizations obscure continuities as the group remained committed to changing both art and politics, and to struggling for the transformation of both society and space, retaining a critical concern with urban geographies. Such historical accounts also sideline the different perspectives on these issues among the various situationist groupings, especially those based in Scandinavia and Germany.

The most influential concept associated with the SI is that of the spectacle. The situationists argued that the accumulation of capital had entered a new stage during the twentieth century, a stage that was increasingly evident after World War II and that was associated with the commodification of ever more realms of social life. Many thinkers and activists of the time noted the significance ascribed to consumerism, leisure, the mass media, and the power of images within the restructuring of capitalism. The growth of television, marketing, branding, and the like indicated an era in which appearances and image management took on new importance. The concept of the spectacle addressed such concerns but in a totalizing way and from a basis in total opposition. It was developed within the situationist journal and found its most corus cating articulation in Debord's *The Society of the Spectacle* in 1967.

In that book Debord targeted the spectacle as a world brought under the sway of the commodity econ omy, a pseudo world apart that is the object of contem plation rather than direct intervention. As the picture of the ruling economic order, it cultivates spectators rather than active agents and induces passivity. Under a barrage of images, life is experienced at one remove, its meaning and prestige derived from appearances. The distinctive qualities of places evaporate along with authentic historical consciousness, to be replaced by the management of a perpetual present. Debord emphasized that the spectacle cannot be reduced to any particular technology such as television, advertising, or propaganda, which are only so many of its manifestations. More fundamentally it is "a social relationship between people that is mediated by images'. It is 'both the outcome and the goal of the dominant mode of pro duction', and 'the very heart of society's real unreality" (1994: theses 4 and 6).

Much academic interest in situationist writings has centered on this concept of the spectacle and compared it with post structuralist or postmodernist theorizations of contemporary social conditions. The term spectacle has thus become something of a buzzword within media and cultural studies. It has also been employed in geography and urban studies to address the emphasis on visual consumption within capitalist urban development and the construction of spectacular landscapes, both histor ically in terms of nineteenth century Paris and London, for example, and in relation to more recent attempts to reconstruct and reimage cities and urban areas, especially since the 1970s. However, references to Debord's work often dilute its critical force by neglecting its roots in Marxism and specifically Marx's writings on alienation and commodity fetishism. They also tend to sideline its basis in revolutionary politics and the ways in which it was forged alongside analyses of contemporary issues that included urban uprisings in the United States, and the political situations in Algeria, the Middle East, and China, as well as attempts to establish a basis for radical alternative political organization embodying direct democracy through workers' councils.

Debord's reading of Marx was Hegelian and had sig nificant connections with work in the Western Marxist tradition by Georg Lukács and Henri Lefebvre. Debord had a close and mutually influential association with Lefebvre for several years and, along with fellow situationist, Raoul Vaneigem, shared his interest in developing a critique of everyday life that emphasized the importance of changing it. In a significant move the situationists wrote of the 'colonization of everyday life', a term that was also taken up by Lefebvre in the second volume of his *Critique of Everyday Life* in 1961, and that was meant to indict the occupation of more and more of social life by the commodity and state power. They coined the phrase at a time when a rapid state led modernization drive in France was accompanied by processes of de colonization and struggles for in dependence in the country's external colonies, and when there was a concomitant turn 'inward' by capital and administrative techniques to focus with new intent on objects and spaces that included leisure, private life, and the built environment.

In using the term colonization, the situationists em phasized the geographical nature of the processes involved, as well as the need to resist them and to reap propriate spaces through processes of de colonization. In this way they understood cities and urban spaces, which also became major concerns for Lefebvre in his series of influential books on the subject through the late 1960s and early 1970s, as key battlegrounds that were sites both of social control and alienation and of potential freedom. Their critique of urbanism built on previous explorations of the subject by Debord and the LI from the early 1950s when they attacked urban developments that were driving the working classes and ethnic minority popu lations from the city center to the suburbs, and destroying vibrant streets and neighborhoods. In their view these developments were part of a process by which the urban as a meaningful site of encounter, sociality, passionate interaction, and historical consciousness was being eradicated.

Critiques of modernist and functionalist architecture were also developed during the 1940s and 1950s by the Danish artist, Asger Jorn, a prolific painter and theorist who had been a prominent member of the Cobra art group, and during the 1950s by the Dutch artist, Con stant, who had been a fellow member of Cobra and was moving away from painting to embrace forms of spatial construction. When they joined forces in the SI, they placed urban questions at the center of the group's pro gram. The situationists' criticism of modernist urbanism was shared by a number of architects, critics, and groups at the time that were closer to the modern movement of architecture but were forging paths beyond its ortho doxies. Unlike architectural and design groups such as the Independent Group and Team 10, however, the situationists took a revolutionary position against con temporary urbanism as a significant component of the society of the spectacle, viewing its creation of oppressive 'environments of abstraction' as a means of reproducing class power through ensuring the isolation of individuals massed together, and hence in need of complete trans formation through a struggle against the totality of the spectacle itself.

Psychogeography and Reclaiming Cities

The situationists disputed the view held by many intel lectuals during the 1950s and 1960s that the working class was declining in importance in a supposedly post industrial and consumerist age. On the contrary they argued that it now included all those who had no direct control over their own lives and the spatial–temporal organization of society. The name situationist was derived from a concern with constructing situations, based on an appreciation of the significance of particular situations for the constitution and conduct of social life, and of the need to intervene in their production. "Since human beings are moulded by the situations they go through," the group wrote, "they need the power to create situations worthy of their desires" (SI, 2006: 178). This meant moving from a detached and interpretative position of the kind they associated with most philo sophical and academic studies, such as the existentialism of Jean Paul Sartre, to an active stance that sought to intervene in everyday life and to transform situations.

The situationists scoured the political horizon for cases of refusal and resistance, seeing one of their main roles as theorizing and helping to articulate emanci patory struggles through which people reclaimed aspects of their lives and environments. They addressed anti imperialist struggles around the world, as well as revolts such as those that shook the Watts district of Los Angeles in 1965 and inner city Detroit in 1967, and wildcat strikes and spontaneous actions by those who were op pressed, alienated, or seemingly bored or nihilistic. In the process they believed they were charting the outlines of a new proletarian assault against class society that was outside the machinations of traditional political parties and trade unions. Such critical activities were one part of the situationist project but others involved direct actions and experiments as the group explored the operation of power and the play of possibilities within cities with the intent of changing them for the better.

During its early years the SI devoted considerable energy to psychogeography. Previously developed by Debord and other members of the LI during the early 1950s, psychogeography was the study of urban environments in terms of their effects on emotions and behavior. It involved investigating ambiences, obscure resonances, and zones of attraction and repulsion so as to inform understandings of the interplay between sub jectivity, emotion, behavior, and urban space. As such it shared certain preoccupations with academic traditions of cognitive mapping, as well as later explorations within humanistic geography, although for the letterists and situationists the accent was on activism and political intervention more than detached inquiry. The study of urban ambiences was combined with resisting or pro posing urban changes in an effort to defend favored sites and neighborhoods from forces of capitalist development and planning, and to consider how they might be trans formed in accordance with different values.

Psychogeography also had a playful and oppositional side especially through its principal mode, the dérive or drift. The latter technique was a mode of urban passage mainly on foot, typically in groups of 2 or 3 people, and lasting anything from several hours to weeks or even months. It entailed letting go of the usual practical rea sons for moving related to work or leisure, and allowing one to be drawn by the terrain and what was found there. Besides providing the basis for psychogeographical in vestigation, it embodied an experimental, adventurous, and anti authoritarian way of urban living, in particular a refusal of work and leisure as structured within capitalist society, and a transgression of their attendant space–time discipline. As such it was a lived critique of everyday life. Connections may be drawn between situationist dérives and surrealist strolls in Paris during the 1920s, not least in their questing nature, in their search for possibilities in the here and now, and in their shared interest in ways of defamiliarizing cities and in unfettering movement and activity. The dérive was less reliant on chance and the unconscious, however, involving a more conscious and collective process of investigation and contestation. Its politicized and collective stance also differentiates the practice from much recent loose use of the term, including in contemporary art practice and criticism where it is frequently invoked.

Urban explorations led to the remapping of cities, including the production of psychogeographical maps of Paris by Debord in 1956–57. These maps, which have since become influential in debates about subverting cartographies, were based on findings from dérives and were composed through a process termed détournement. The latter was a key tactic of the letterists and situ ationists that entailed diverting, reworking, or hijacking existing cultural materials and recomposing them to create different propagandizing effects. Recognizing that language and cultural products are realms of class conflict and struggle rather than transparent means of communication, the groups used détournement for a range of interventions in different media that included writings, films, paintings, cartoons, and architectural compositions. In the case of cartography, Debord cut existing maps of Paris into segments, each indicating a 'unity of ambiance' discovered on dérives, and re positioned them with arrows suggesting movement toward or away from the sections. The images challenged standard cartographies to depict what Debord called the city's 'psychogeographical relief', along with the search for alternative passages through its spaces. Traces of dérives further surfaced in letterist and situationist publications and films made by Debord, where their combative, impassioned, and frequently drunken spirit is apparent. However, psychogeography and its techniques of mapping were also meant to look to the future and to the potential construction of new cities. As such they informed utopian visions of cities and struggles for different futures.

Utopianism and Sociospatial Transformation

The notoriety of the situationists was established in particular through the revolts in Paris in May 1968, which they helped to instigate and fuel. Events in that city were part of worldwide revolutionary political struggles during the course of 1968. In France occupations of universities, institutions, and streets in the capital soon led to around 10 million people in the country taking strike action. Daily routines were halted and time and space temporarily transformed. The situationists regarded it as an emancipatory festival in which hierarchies melted, and in which communication and community found a new reality and terrain. The influence of the situationists was through not only their participation in the occupations movement in Paris but also their prior activities and writings, which anticipated many aspects of the May struggles.

The situationists previously developed the idea of revolution as a festival and as involving the transformation of everyday space and time in their analysis of the working class reclamation of Paris during the Commune of 1871, which they characterized as a form of revolutionary urbanism. Questions about revolutionizing urban spaces were indeed long standing concerns as their commitment to creating situations and criticizing existing urbanism attest. They framed this early on in terms of 'unitary urbanism', a concept that referred to the collective, conscious, and continuous creation of the urban environment by people as they construct a different world. Unitary urbanism was utopian in the ways in which it anticipated the transcendence of current social and spatial divisions through the creation of a dynamic space that would be mobile, playful, and lived by its creators. However, the situationists insisted that it was not a doctrine or a plan for some distant future but was rather a critique of urbanism that found expression in immediate experimentation and political struggle.

The fullest attempt to outline unitary urbanism came with Constant's work on an urban future that he entitled New Babylon, and that he developed as a member of the SI before his resignation in 1960 and then independently over the following decade. This utopian project challenged contemporary urban conditions and ideologies through promoting 'an other city for an other life' in models, plans, maps, drawings, lithographs, paintings, photographs, films, and extensive writings. Constant's proposals stemmed from the opposition he shared with the other situationists to the oppressive results of capitalist processes of urbanization and planning, and from their belief that other worlds were possible. They also came out of his desire to find a means of creative revolutionary activity while rejecting the dominant culture of his time along with its traditional individualized modes of artistic practice.

Constant envisaged an urban environment for a postrevolutionary and postscarcity society in which land and the means of production were socialized, work was automated, and life was devoted to play. He presented New Babylon's structures as largely devoted to social space and entirely flexible, intended as the ever evolving products and medium for the life play of the inhabitants, whose creativity would be unleashed and whose lives would become nomadic as they were freed from the constraints of utilitarian society. Constant insisted that New Babylon was not primarily a planning or design project but was instead an attempt to provoke different ways of imagining and acting in cities, and to encourage, as well as to outline spaces for a new way of urban living. He further denied his vision was utopian in the common usage of that term to denote an impossible but desired place; rather, he claimed that it was realizable so long as the potentialities of current productive forces and technologies were seized by people and put to different ends.

Constant's focus on revolutionizing social space brought him into a close dialog with Debord and later Lefebvre, who wrote favorably about his experiments. However, once Constant resigned from the SI after tensions concerning the influence of other artists came to breaking point, the group publicly dismissed his work as its focus on urbanism soon moved away from visionary proposals and toward political critique and action. Debord and others still insisted that a revolutionary struggle must transform the totality of existing conditions to be effective. They did not approach this through the lens of urbanism or esthetics as such, however, with Debord referring more generally to the need "to reconstruct the entire environment in accordance with the needs of the power of established workers, councils" (1994: thesis 179). Constant for his part continued work on New Babylon until the early 1970s, after which he returned to painting.

Long neglected by architectural historians, New Babylon was 'rediscovered' and its significance for that field asserted in the late 1990s especially by Simon Sadler and Mark Wigley, who traced connections and influences with contemporaneous designers such as Team 10 and Archigram, as well as later practitioners. The situationists more generally have also received increasing attention in geographical and urban debate as Alastair Bonnett, David Pinder, and others have explored and contextualized their writings and activities, and deployed them for critical ends. In returning to situationist engagements with geographical issues, they have revealed the significant ways in which situationist perspectives anticipate, connect with, or challenge other radical geographical positions, in particular concerning the production and contestation of space within Marxist traditions, and the possibilities of countering hierarchical power and developing autonomous spaces associated with anarchist traditions. Andy Merrifield has similarly asserted the importance of Debord's vibrant yet undervalued contribution to the rich lineage of Marxist writings on cities.

Situationist ideas may have recently moved from the margins to become strangely respectable, as Erik Swyngedouw remarks, but they have not been completely sanitized. Not only are their critiques of the spectacle remarkably prescient in current image saturated times, as battles rage over control of the image realm, but they also continue to offer provocative insights and critical impetus for exploring cities and the politics of urban space, and for challenging the processes of the capitalist production of space with a view to how other worlds might be constructed. At a time when dominant powers insist there is no alternative to capitalism and liberal democracy, Pinder asserts that the utopianism of the situationists is of particular significance, retaining the capacity to scandalize and inspire through its critique of the spectacle and its desire to construct radically different worlds. In addition to considering situationist critiques of urbanism and their experiments with psychogeography, he addresses Constant's New Babylon, situating it in relation to other currents of utopian critical thought and practice so that its often neglected political consequences move to the fore, along with the challenges it continues to pose to conceptions of spatiality and processes of urbanization.

Much recent literature on the SI has focused almost exclusively on Debord and, especially in France but also increasingly in Britain and North America, treated him as a great moralist and prose stylist, detaching his writings from the wider situationist terrain, as well as their Marxist roots and political engagements. As a result the situationist legacy becomes construed as primarily literary, artistic, and even bound up with individual heroism, a pure stance to be admired from afar. As suggested above, however, in contrast to this overly narrow view of Debord and the SI, others are positing and activating a more open, entangled, and collective understanding of the group's historical geographies in the plural and hence of their potential legacies. In this context it is worth stressing that those drawing inspiration from situationist activities are found but especially among political activists, artists, and revolutionaries.

One prominent legacy of the SI's geographical concerns lies in the flowering of interest in psychogeography since the 1990s in Europe and North America that typically traces its lineage back to the experiments of Debord and fellow explorers of Paris. This has multiple activist, literary, artistic, and dilettantish manifestations, some quite far from the original source. Among its expressions have been the establishment of a number of psychogeographical groups, associations, and festivals, as well as a more diffuse influence on ways of exploring, imagining, and contesting cities. These have included the revival of a London Psychogeographical Association, which took its name from Rumney's original formulation and was active during the 1990s, as well as other groups in New York, Manchester, Nottingham, Toronto, Milwaukee, and elsewhere. Their activities have ranged from activist struggles over urban change to explorations, mappings, and imaginatively engagements with places through various techniques, both low tech and new media. The term has also been popularized by the British writers, Iain Sinclair and Will Self.

While recent psychogeographical activities have done much to bring into mainstream discussion certain geographical ideas and practices associated with the situationists, the latter's insistence that these be connected to revolutionary praxis, committed to bringing about sociospatial transformation, has rescinded from view. Indeed, in some hands the political engagement at the heart of letterist and situationist psychogeography is lost altogether. Yet, another important situationist legacy is apparent in the positions, tactics, and language adopted by a number of radical and revolutionary groups, as they struggle against capitalism and oppressive state power, and as they seek to envisage and build alternatives with an awareness of how geography matters for both the constitution and contestation of power. Far from being a dead terrain, then, different understandings of 'situationist geography' are continuing to be drawn upon and activated in a variety of ways, both intellectual and political. Their role in future emancipatory struggles may not be negligible.

See also: Anarchism/Anarchist Geography; Avant-Garde/Avant-Garde Geographies; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Psychoanalytic Theory/Psychoanalytic Geographies; Radical Geography; Surrealism/Surrealist Geographies.

Further Reading

Bonnett, A. (1989). Situationism, geography, poststructuralism. *Environment and Planning D: Society and Space* 7, 131 146.
Bonnett, A. (1992). Art, ideology, and everyday space: Subversive tendencies from Dada to postmodernism. *Environment and Planning D: Society and Space* 10, 69 86.
Bonnett, A. (2006). The nostalgias of situationist subversion. *Theory, Culture and Society* 23(5), 23 48.
Debord, G. (1994). *The Society of the Spectacle*, Nicholson Smith, D. (trans.). New York: Zone Books (Originally published in 1967).

Internationale situationniste 1958 1969 (1997), collected facsimile edition. Paris: Arthème Fayard.
Jappe, A. (1997). *Guy Debord*, Nicholson Smith, D. (trans.). Berkeley, CA: University of California Press.
Knabb, K. (ed. and trans.) (2006). *Situationist International Anthology*. Berkeley, CA: Bureau of Public Secrets, revised edition.
McDonough, T. (ed.) (2002). *Guy Debord and the Situationist International: Texts and Documents*. Cambridge, MA: MIT Press.
Merrifield, A. (2002). *Metromarxism: A Marxist Tale of the City*. New York and London: Routledge.
Merrifield, A. (2005). *Guy Debord*. London: Reaktion Books.
Pinder, D. (1996). Subverting cartography: The situationists and maps of the city. *Environment and Planning A* 28, 405 427.
Pinder, D. (2005). *Visions of the City: Utopianism, Power and Politics in Twentieth Century Urbanism*. Edinburgh: Edinburgh University Press.
Sadler, S. (1998). *The Situationist City*. Cambridge, MA: MIT Press.
SI (2006) Questionnaire. In K. Knabb (ed. and trans.), *Situationist International Anthology*. Berkeley, CA: Bureau of Public Secrets, pp 178 83 (originally published in 1964).
Swyngedouw, E. (2002). The strange respectability of the situationist city in the society of the spectacle. *International Journal of Urban and Regional Research* 26(1), 153 165.
Vaneigem, R. (1994). *The Revolution of Everyday Life*, Nicholson Smith, D. (trans.). Left Bank Book and Rebel Press.
Wigley, M. (ed.) (1998). *Constant's New Babylon: The Hyper Architecture of Desire*. Rotterdam: Witte de With/010 Publishers.

Relevant Websites

http://www.bopsecrets.org
Bureau of Public Secrets.
http://www.notbored.org
Not Bored Situationist International Archive.
http://www.cddc.vt.edu
Situationist International Online, Center for Digital Discourse and Culture.

Situationist City

Q. Stevens, University College London, London, UK

Glossary

Dérive An aimless, playful, drifting walk through the city.

Détournement The practice of pirating, reorganizing, and parodying bourgeois art and text in ways which communicate new messages.

Psychogeography The study of the laws and effects of the geographical environment on the perceptions, emotions, and behavior of individuals.

Spectacle Propaganda, advertising, commodities, and built environments which present images that help to justify a system of social relations, to disguise those social relations, and to produce them, especially by placing the viewer in a position of passive receptivity.

Unitary Urbanism A critique of existing urbanism through action, by appropriating and exploiting existing spaces and establishing various new ambiences. Also a spatial milieu conducive to dissolving existing social and functional separations; a complexified terrain for free social play which can further people's self-realization.

Introduction

The Situationist International (SI) (1957–72) was a small avant garde and revolutionary group of writers, artists, and social and political theorists based in Paris. Key contributors included Guy Debord, Raoul Vaneigem, and artists Asger Jorn and Constant (Nieuwenhuys). Their critique of cities is chiefly known through the translation and republication of numerous short articles from their journal *Internationale Situationniste* and from Potlatch, the journal of the preceding Lettrist International (1952–57).

The Situationists' artistic roots lay with the nihilism of Dada and the subversion of surrealism, both of which illustrated the potential for everyday experience and familiar objects to be charged with intensity, a diversity of desires, and creativity. They were also interested in the phenomenology of urban space. Politically and philosophically, their thinking was informed by Hegel's dialecticism, Marx, Lukacs, and Marcuse, and by the existentialism of Sartre, who was the first to use the term 'situation' to describe the individual's need to constantly negotiate the complexities of everyday life.

The Situationists were critical of rapid technological and cultural changes in the 1950s: the functionalism of industrialized production; the growing power of national governments; conservative bourgeois social attitudes; modernist town planning; and consumerism in Western Europe, with its commodity fetishism, advertising, and passivizing forms of leisure, and in the East, the cult of personality. The SI referred to these as the diffuse and concentrated forms of spectacular society, respectively, and rejected the French Left's acceptance of the bureaucratic oppression of state socialism. The SI viewed these issues in relation to a broader existential problem: alienation, and the suppression of individual experience, expression, and creativity. Everyday life was characterized by a homogeneous mediocrity, boredom, predictability, a lack of control, or any sense of adventure. While the methods and objectives of the SI were diverse and changeable, their core ambition was the development of new, unalienated ways of living focused around creative play. They sought to progress this notion dialectically, chiefly by critiquing the separateness of art and integrating it into everyday living, and through general critiques of modern society which attempted to recuperate and redeploy what they saw as the positive elements of its cultural and technological development, such as mass media, communications, and high speed transport.

Through their rhetorical, experimental, strategic, and often explicitly self critical approach, the SI took care to emphasize that their writings, public actions, films, and architectural models served principally as situated critiques of existing cities and societies, rather than presenting an enduring movement based on universal objective truths and tested methods. The latter would have been incompatible with their dialectical understanding of social praxis and their emphasis on direct participation. Their main role was as agents provocateur, hoping to incite citizens to overcome alienation, reclaim urban space, socialize playfully, and invent new ways of life, all on their own terms without mediation or control by experts. Politically engaged, their ideas played a critical role in the popular uprising in Paris in May 1968.

The term 'Situationist City' was coined by Sadler, who first drew academic attention to the Situationists' critique of the urban environment, their principles for urban planning and urban living, and their design proposals for urban forms. From the SI's own perspective, such a construct as 'the city' is in itself a reification of built forms, and images of them, as spectacular objects of consumption, a reification of the dominant social order, and a masking of the history of power relations and social processes which produced it. The SI viewed existing urbanism as one of the key causes of alienation and as the enemy of revolutionary social practice. The existing city was of core interest to the Situationists because it was the spatialization of the most advanced form of consumer capitalist society, presenting both the extreme of instrumental alienation of everyday life and the greatest potential for emancipation and revolution.

This article explores in turn the nature of Situationist urbanism as an analytical concept, as an object of analysis, and as a critical proposition.

The Situationist Critique of Urbanism

The Situationists criticized the banality of contemporary life, ruled by technological conveniences, architecture, and urban planning which had all liberated man from poverty and the vagaries of the natural environment, but had themselves become fetishized and led to the ossification of everyday life. They argued that the key objective behind Hausmann's urban renewal plans had been the circulation of troops, and that contemporary urbanism's key objective was the circulation of motor vehicles. They suggested future urbanism might instead apply itself to the functional enhancement of psychogeographical possibilities, through the creation of varied psychic and sensory atmospheres, and to the free circulation of people.

Their detailed examination of modern urbanism critiqued its spectacular ideological character, which encouraged passivity, and which obscured the complicity of planning with the control of urban space and the conditioning, surveillance, and policing of social behavior in the interests of capitalism and bureaucracy; the reduction of density and the segregation of functions, including leisure, in the name of utility, which fixed people in space and led to social disintegration; capitalism's emphasis on divisive contests for space; and the promotion of traffic circulation which furthered alienation and restricted social encounters. While the SI praised functionalism's embrace of designing around function and comfort, its rejection of ornament and its technological innovation, they criticized its unconscious alignment with conservative social and ethical principles (such as health, relaxation, property, and the family) and its role as a static doctrine. The SI noted that the historic texture of the city connected people to the past, although the catalyzing power of its symbolism was diminished and in need of rejuvenation by bringing new meanings to its landmarks and monuments. They called for a spiritual transformation of urban life based on the rediscovery of forgotten desires and the creation of new ones.

Many elements of the SI's critique of urbanism were particular to the context of post war urban reconstruction in France, and specifically Paris, and the class and ethnic tensions underpinning it. They disparaged the museumization of inner cities, their presentation as spectacles for tourist consumption, and their gentrification by the growing managerial elites, as well as the Grand Ensembles, a massive modernization program of new office towers, department stores, cultural institutions, and freeways, which all furthered the circulation of people and commodities and the dominance of commerce and the bureaucratic state. These changes were exemplified by the destruction of Paris's historic wine market and its replacement with the expanded modern university; the destruction of the market at Les Halles to construct an underground commuter rail interchange and shopping center; and the destruction of the adjacent poor neighborhood, Beaubourg, to build a showcase national center of art and culture, Centre Pompidou, where a retrospective exhibition of the SI was held in 1989.

The SI was neither nostalgic about old cities nor opposed to modernization. They saw outmoded spaces and ways of life as both evidence of and model for resistance to the state's authoritarian decision making. In a letter to *The Times* regarding plans to demolish London's Chinatown, they noted:

> We protest against such moral ideas in town planning, ideas which must obviously make England more boring than it has in recent years already become… if modernization appears to you, as it does to us, to be historically necessary, we would counsel you to carry your enthusiasm into areas more urgently in need of it, that is to say, your political and moral institutions. (Bernstein *et al.* in Andreotti and Costa, 1996: 54)

The Situationists criticized the banishment of Paris's urban poor and immigrants to new Banlieue (suburbs) on the periphery, and the use of similar methods in France's colonial domination of Algeria. SI member Abdelhafid Khatib was twice arrested for breaking a police curfew on Algerians.

The SI criticized the construction of industrial new towns in France along modernist lines, where the labor and social relations of people's employment were reproduced in the segregation of housing types and locations for single people, married couples, families, and high salary employees, with managers living in other, older cities. They often compared modernist housing blocks in these suburbs and new towns to military barracks, concentration camps, or indigenous reservations and compared single family homes to fall out shelters, completely isolated.

The SI were also critical of new towns elsewhere in Europe, and of the new modernist colonial cities Chandigarh and Brasilia, which illustrated and effectuated a centralized bureaucratic worldview rather than progressive, modern social ideals. Potlatch had also boycotted a 1956 festival celebrating Le Corbusier's Radiant City plan. They described Corbusian tower blocks as slum construction, vertical ghettos, and prisons, denouncing Corbusier's elimination of the street with its prospects for encounters.

Methods of Analysis and Representation

The SI developed numerous experimental techniques for studying urban spaces and communicating their perceptions of them. These methods were closely tied to their theoretical perspective, reflecting their interests in intervention, praxis, and dialectics. Their analytical methods sought to directly put their theories into practice, to transform everyday life and achieve the more actively engaged, multidimensional, specific, and authentic experience which modern society denied, and to aid in the further development of their ideas, by understanding the possibilities and the internal contradictions within social life as they experienced it.

The central aim of their investigations was to further the field of psychogeography, their most unique and significant contribution to urban geography. Their key field technique was the *dérive*:

> One or more persons abandon, for an undefined period of time, their usual motives for action and movement, their relations, their labor and leisure activities, and let themselves be drawn by the attractions of the terrain and the encounters they find there… the *dérive* includes both this letting go and its necessary contradiction: the domination of psychogeographical variations by the knowledge and calculation of their possibilities. (Debord in Andreotti and Costa, 1996: 22)

The SI argued that large cities were particularly conducive to this kind of distraction. The conscious effort to recognize and pursue particular physical décor, ambience, and social context distinguished *dérive* from its models, *flânerie*, and surrealist chance excursions, as did the emphasis on suspending one's class consciousness, and active, tactical intervention to expand urban possibilities. The Situationists often used military metaphors to describe both the existing city and their own activities and objectives; they saw urban space as their theater of operations. *Dérive* was political, critical, and combative activity because it emphasized and helped produce the use value and emotional qualities of space, contesting the routinized and spectacular nature of urban life and its domination by exchange value and constrained notions of functionality.

Some Situationist writings took the form of reports on *dérives*: an investigation of the ambience of Les Halles; the discovery of a psychogeographic hub surrounding a ruined rotunda by Ledoux in the 20th arrondissement; troubling encounters with immigrants in a variety of bars. Debord's now unavailable film *On the Passage of a Few Persons Through a Rather Brief Period of Time* employs *dérive*, providing a wandering documentation of Paris's Left Bank and Les Halles, critiquing their architecture and depicting the everyday life of marginal people spending their time freely. Keillor's 1993 film *London* illustrates a fictional series of *dérives* purporting to grasp that city's moods, power relations, and untold histories. A slow sequence of still shots of marginal and mundane locations is accompanied by the ambient sounds of everyday events which crisscross the scene. An unknown narrator makes free associations to the feelings and memories of the fictional *dériver*, and mentions boredom, disappointments, detours, and unexpected events during the journeys. The narration is often disjunctive with the images. The film is primarily about the situated circumstantiality of urban experience; it fails to explain the city.

The SI applied a different kind of creative reappropriation to the images which consumer society produced – *détournement* – which exposed, contradicted, and destabilized the dominant narratives of power propagated through high culture, propaganda, and advertising. Unlike surrealist collage and postmodern pastiche, *détournement* was seen as a collective, anonymous, and political activity. Although *détourned* graphics and public graffiti could illustrate the foibles of consumer society and inspire action, the main link between *détournement* and urban space was through the reception, reuse, and production of maps.

Debord and Jorn's *Mémoires*, a *détourned* book of poetry, was composed from scattered pieces of maps of Paris, text, and images. One page features map fragments corresponding to Paris locations visited during a 1953 *dérive* by members of the Lettrist International. On another page, a historical battle scene is mirrored against a map of Contrescarpe; the Lettrists frequented this area of Paris, admired its playful and forgotten atmosphere and opposed urban renewal projects there. Collaged text here and throughout the book reinforces the combative theme. Another page is crowded with exemplars of picturesque and primitive architecture from various exotic civilizations, depicting the city as a provocative mixture of vernacular and monumental forms.

Another similar map, *Discours sur les passions de l'amour*, and its claimed antecedent, a 1653 map of the Land of Love, emphasize the metaphorical link between journeys through the city, its erotic stimulations, and the emotional trajectory of social relationships. The fragments collaged from a perspectival rendering of Paris included several indigenous working class neighborhoods which Debord and Jorn felt should be studied and preserved.

The 'Naked City' is the Situationists' clearest example of a *détourned*, psychogeographical map. This was a collage of fragments of a map of Paris linked together by sets of large red arrows. The collage technique undermines the instantaneous unity, totality, order, and detachment that a map imposes. The Naked City emphasizes instead that cities are experienced successively in time and in motion by actual, situated subjects, whose movements are guided by the attractions and repulsions of particular atmospheres, and shaped around key psychogeographic loci, to form habitual patterns. The fragments of the map were carefully trimmed to include all streets which contribute to a particular local ambience, and rotated and separated to emphasize the smooth drifting journeys between them along the psychogeographical slopes represented by the arrows. A noted antecedent was maps of Paris published by the sociologist Chombard de Lauwe in 1952: those of residential quarters, discontiguous in shape because they were determined by the behavior of the inhabitants, and with arrows showing inward flows; and a plot of the narrow scope of all the journeys taken over 1 year by a bourgeois student.

These maps are narrative descriptions of events and feelings rather than objective tools of universal knowledge. They present a critique of town planning and of the instrumental nature of space and behavior; they report on the Situationists' own experiences with *dérive*, and also provide a stimulus and a guide to escaping conventional ways of seeing the city and moving through it.

Proposals, Constructions, and Actions

The Situationists also made concrete proposals for changes to existing cities, to their planning and management, and to social practice. These proposals, like their critiques, are characterized by a freely creative, playful, humorous attitude which itself provides a counterpoint to the pragmatic, conservative, and dull metaphysics of urban planning.

The Lettrists' plan for rational improvements to the city of Paris proposes numerous *détournements*. They advocate opening the underground at night after the trains had ended, as well as public parks – although both were to remain unlit or dimly lit for psychogeographical effect, and streetlights given switches so the public could adjust them as desired. They recommended opening buildings' rooftops to the public and linking them together, and opening prisons for free access. Cemeteries, museums, monuments, and statues were to be eliminated, and the artworks redistributed to bars and *cafés*. Train stations' destination boards would be removed or scrambled to promote *dérive*.

Situationist theses on traffic parodied the International Congress of Modern Architecture's Athens Charter, criticizing planning for the automobile and commuting (which increased alienation and reduced free time) and the zoning of work and residence, and proposing instead unitary urbanism: not another new doctrine, but a dialectical, lived critique of existing urbanism which would expose its deceptions, manipulations, and limits and turn its messages and mechanisms back on themselves. Unitary urbanism aspired to an idea of functionalism beyond the immediate and obvious usefulness of objects, and beyond habits, social conventions and taboos. It promoted the appropriation, exploitation, and reintegration of existing urban spaces, buildings, and *décor*, including the pedestrianization of city centers. It required the unification of social space to enable social participation and free exchange, but also the establishment of new and varied ambiences to stimulate a spectrum of feelings. They appropriated the term 'quarter' to highlight this ongoing enhancement of differentiation. Unitary urbanism also called for the permanent creative transformation of the city, in opposition to fixed patterns, museumization, and monuments. The *dérive* was an example of unitary urbanism because it was simultaneously a means for studying the urban milieu and a social game played within it. It aided the discovery of fragments of urbanism bypassed or forgotten by spectacular urbanism, as highlighted on Situationist maps.

Unitary urbanism could be achieved through the *détournement* of existing architecture, urbanism, poetry, and cinema or the creation of new forms. The SI produced several avant garde architectural projects intended to *détourn* the separation of art, urban space, and intense personal experience and to promote playful socialization. The Cavern of Anti Matter was based on 'industrial painting': a long roll of cloth painted with abstract strokes by a rudimentary machine was spread around walls, ceiling, and floor. Mirrors, lights, perfumes, and music added to a dramatic atmosphere intended to stir visitors into action. A model dressed in the fabric demonstrated its use as clothing, depicting the total merger of the user with their environment. Cheap, mass produced, and sold by the meter, industrial painting demonstrated the SI's general aspiration that industrial means could help ordinary citizens transform their environment, come closer to art, and experience a new, liberating social atmosphere.

A subsequent unrealized installation for Amsterdam's Stedelijk museum was a labyrinth that merged interior and outdoor environments, featuring artificial rain, fog and wind, variations in light, temperature, and sound, and other provocations. Situationists would *dérive* together through this labyrinth, remaining in walkie talkie contact with other groups *deriving* the streets and canals of central Amsterdam. Constant would direct these *dérives*, map their progress, give instructions, and arrange impromptu experiments.

A far larger, more ambitious, and more utopian environment designed to promote situations was Constant's New Babylon. Through several recent books and exhibitions of its models, plans, and architectural drawings, this scheme has become a synecdoche for the Situationist City. Disagreement on the appropriateness of pursuing such concrete and spectacular spatial visions was the grounds for Constant leaving the SI, the most significant of numerous splits and resignations which marked their history.

New Babylon originated with a permanent camp Constant designed for Gypsies: a lightweight, transportable structure, with movable partitions. New Babylon was an endless megastructure, an irregular labyrinth of buildings elevated on large pylons. The ground surface was left open for vehicles, agriculture, nature, and monuments. Above, the building had no functional zoning and few private areas. The interior was a multilevel open plan social space with moveable and demountable partitions, floors, roofs, staircases, and bridges which could be constantly reorganized by New Babylon's nomadic inhabitants. Light, color, acoustics, ventilation, odor, texture, temperature, and moisture could all be mechanically modified by the users to create different ambiences to suit their changing needs and desires, or to stimulate new emotional states. The intention was partly to disorient and to confuse conventional spatial ordering; routine, customary behavior, and inhabitation would be impossible. New Babylon was typically presented in models and drawings as 20–30 ha sectors – the term indicating Constant's aversion to the restrictive bourgeois social relations and possessive domesticity of 'neighborhood.' Constant frequently described the scheme's many means of movement: wandering on foot, taking cars or underground trains between sectors, or by plane from rooftop airstrips. He later suggested airports provide a harbinger of the future city of nomadism, enabling social contact with difference and long distance adventure for the masses.

Later presentations showed New Babylon's network at different scales superimposed over maps of various European cities and regions. In these views, it bears comparison to the arrows on Debord's Naked City map: it is an armature which provides new connections between previously unrelated parts of the historic urban landscape.

Constant's utopian environment was premised on a technologically advanced society of automation which would allow people to devote their time and energy to adventure, pleasure, play, free and varied socialization, and thereby the development of their creative capacities. Free time and mobility would enable people to engage in a permanent worldwide *dérive* through this landscape which they were constantly involved in producing: continual rearrangement and reinvention of the space itself was to be part of the game.

Critics disagree on interpretations of Constant's proposals. Some emphasize New Babylon's architectural and ideological similarities to other contemporary technocratic megastructural fantasies, by Utopie (France), the Metabolists (Japan), and Archigram (UK), although these varied in their social program and political objectives, and were not as driven by political critique. Some highlight its technical shortcomings, impracticalities, and paradoxes. Others praise New Babylon's metaphoric potency as a critique of the existing historical city, and as a provocation, illustrating how social space could be transformed alongside an unalienated social life. The SI themselves saw Constant's spatial constructions as preliminary experiments in form which helped to clarify the problem of unitary urbanism.

It was fundamental to the SI's dialectical, materialist, and phenomenological approach that ideas had to be integrated as much as possible with actions in lived spaces. They aimed to construct situations, and to promote people's free and active participation in such construction, as a means to help citizens recognize and better understand how different, playful combinations of people, environments, and events could produce certain ambiences and feelings. They also believed that through practices like the *dérive*, everyday action could provide insight into new possibilities for experience and for space. Their ultimate aim was to stimulate the raising of everyday consciousness and the ongoing transformation of everyday life, of politics, and of the environment itself. To the SI, this was not just an academic methodology, but a core aspiration for everyday life in general.

The Situationist City Today

Urban geographers before the 1990s did not engage with Situationist thought in great depth, tending to theorize spectacle primarily in terms of architectural imagery and failing to examine its pervasive role in structuring social relations and practices, although French sociologists posited that spectator observers could, consciously or unconsciously, resist classification and control by spectacular, textual urban landscapes.

The 1990s saw a dramatic revival in interest in the SI. This interest is splintered, with different disciplinary focuses on their means of analysis, their critique of urban areas, and their design proposals. New Babylon has been widely exhibited, becoming a spectacularized version of the Situationist City, a symbol of the recuperation of Situationism by the very forces they most despised: the art market, the intellectual elite, and their passive bourgeois audiences. Much recent secondary literature on the Situationists has been criticized for treating their visions of urban life and the city, their writings, and their personalities as commodities, intellectualizing, estheticizing, and depoliticizing them. This interest ignores Situationism's eclectic and shifting membership, aspirations, and strategies, as well as ignoring the practical realities of everyday, situated action in real places. In contrast to this paradoxical reification of Situationist concepts and objects, other reappropriations of urban space have been pointed to: Los Angeles' 1964 Watts riots and recent situations such as Reclaim the Streets and worldwide antiglobalization protests.

The primary influence of the Situationists on con temporary urban geography is methodological, show cased by the short lived journals *Transgressions* and *The Journal of Psychogeography*, and evidenced through the *dé rived* writings of Iain Sinclair and photography of Martha Rossler. There have also been broader impacts on critical theory: the attempt to imagine or uncover a post struc turalist Marxist dialectical utopian urbanism, viewed as a social project which challenges present conditions and where both built form and urbanization processes are open to possibility, social difference, and the develop ment of new norms.

See also: Dialectical Reasoning and Dialectical Materialism; Leisure; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Phenomenology/ Phenomenological Geography.

Further Reading

Andreotti, L. and Costa, X. (eds.) (1996). *Theory of the Dérive and Other Situationist Writings on the City*. Barcelona: Museu d'Art Contemporani de Barcelona ACTAR.

Bonnett, A. (1989). Situationism, geography, and poststructuralism. *Environment and Planning D: Society and Space* 7, 131 146.

Borden, I. and McCreery, S. (eds.) (2001). *New Babylonians*. *Architectural Design* 71(3), 67 73.

Debord, G. (1994). *The Society of the Spectacle*. New York: Zone Books.

de Certeau, M. (ed.) (1984). *Walking in the City. The Practice of Everyday Life* pp 151 160. Berkeley: CA: University of California Press.

de Zegher, C. and Wigley, M. (eds.) (2001). *The Activist Drawing: Retracing Situationist Architectures from Constant's New Babylon to Beyond*. New York: The Drawing Center.

Kaplan, A. and Ross, K. (eds.) (1987). Introduction. *Yale French Studies* 73, 1 4.

McDonough, T. (ed.) (1997). Guy Debord and the Internationale Situationniste. *October*, Vol. 79 (Winter), republished and expanded as McDonough, T. (ed.) (2002). *Guy Debord and the Situationist International*. *Texts and Documents*. Cambridge, MA: MIT Press.

Pile, S. (2002). The problem of London, or, how to explore the moods of the city. In Leach, N. (ed.) *The Hieroglyphics of Space*. London: Routledge.

Pinder, D. (2000). Old Paris is no more: Geographies of spectacle and anti spectacle. *Antipode* 32(4), 357 386.

Pinder, D. (2001). Ghostly footsteps: Voices, memories and walks in the city. *Ecumene* 8(1), 1 19.

Pinder, D. (2005). Arts of urban exploration. *Cultural Geographies* 12, 383 411.

Pinder, D. (2005). *Visions of the City: Utopianism, Power and Politics in Twentieth Century Urbanism*. Edinburgh: Edinburgh University Press; New York: Routledge.

Plant, S. (1992). *The Most Radical Gesture: The Situationist International in a Postmodern Age*. London: Routledge.

Sadler, S. (1998). *The Situationist City*. Cambridge, MA: MIT Press.

Sinclair, I. (1997). *Lights Out for the Territory*. London: Granta.

Swyngedouw, E. (2002). The strange respectability of the Situationist City in the society of the spectacle. *International Journal of Urban and Regional Research* 26(1), 153 165.

Vaneigem, R. (1983). *The Revolution of Everyday Life*. London: Rebel Press.

Wigley, M. (2000). *New Babylon The Hyperarchitecure of Desire*. New York: Princeton University Press.

Relevant Websites

http://www.notbored.org
NOT BORED! is an autonomous, situationist inspired, low budget, irregularly published, photocopied journal.

http://www.nothingness.org
Situationist International.

Slums

R. Harris, Hamilton, ON, Canada

Slums

A slum is a residential area with substandard housing that is poorly serviced and/or overcrowded, and therefore unhealthy, unsafe, and socially undesirable. A single dwelling can fit this description, but the term is usually reserved for larger areas, rural or more commonly urban. Slums, like the poor who inhabit them, have always existed, but the term was not coined until the early nineteenth century, when it was first used in the largest city of the day, London, England. They became extensive with the rise of the industrial city and are now ubiquitous in the developing world, where they commonly take the form of squatter settlements. The term is culturally defined and pejorative, with social as well as physical connotations. It is usually applied by outsiders, some of whom have used it to justify their interventions in neighborhood affairs, including slum clearance, and has often been resented and resisted by local residents. The discourse of slums has always been contested.

Definition: Physical and Social Aspects of the Slum

Slums are culturally defined, both in terms of their physical characteristics and also in their social correlates. Certain characteristics of dwellings, for example, their materials of construction and quality of electrical wiring (if any), directly affect health and safety, notably fire risk. In general, however, there are no absolute standards of physical adequacy, crowding, or of the residential services that are necessary, even for public health. Most obviously, requirements vary with the physical environment. In warm, dry climates people spend more time outdoors and require less interior space; then, too, the main use of interiors is for sleeping, which requires only limited space. In areas with low residential densities (dwellings per acre), quite high levels of crowding (persons per room) may be more tolerable, while piped water and sewer systems may be desirable but not absolutely necessary. How people use their housing is just as important. Certain cultures have long emphasized healthy practices such as breast feeding and strict domestic hygiene. This is why, for example, rates of infant mortality were relatively low among North American Jewish households in nineteenth and early twentieth century cities, even though many of these lived in rundown neighborhoods.

Just as bad physical conditions do not always guarantee poor health, better conditions may not be adequate. Rural migrants bring domestic animals, and habits of waste disposal, that make adequate urban housing unhealthy. Those unaccustomed to sanitary facilities may not use them properly. In England, some middle class reformers claimed that ignorant working class tenants in new council (public) housing would keep coal in their baths. Patronizing and possibly apocryphal, this story dramatized the usefulness of education programs for public health. Such programs became common in early twentieth century Europe and North America, as medical knowledge and epidemiology improved.

The influence of living conditions on social pathology, although widely assumed, is even less predictable. There has usually been an association between slum conditions and social problems, notably crime. To a large extent this association reflects the underlying influence of poverty. Most people who live in slums do so because that is all they can afford and their low incomes make them vulnerable to crime. Low income households are not mobile; they shop locally, buy in small quantities, and pay more than average for items of food and clothing. Their incomes are not only low but also uncertain, which undermines their ability to save, whether for their children's education, periods of illness, or old age. They may rationally conclude that it makes more sense to live in the present rather than try to plan for an uncertain and uninviting future. They are a vulnerable target for those engaged in illegal activity, while the young, especially, may be drawn into crime themselves by the absence of attractive alternatives. This becomes more likely because the stress of living hand to mouth often pulls families apart, so that children have to grow up early, without proper adult supervision and role models.

If slum conditions reflect the poverty of their occupants they also exacerbate that poverty and its associated social problems. This influence is of particular interest to social liberals. Conservatives often argue that poverty and slum conditions reflect low moral standards and the poor choices that some individuals have made. They leave little room for environmental influences. In contrast, rather than blaming the victim, liberals have argued that, for most slum dwellers, the choices are limited. From this point of view, among the important constraints that slum dwellers face is the geography of the built environment, not just of the slums themselves but of the larger metropolitan area. It is generally agreed that social problems are compounded where the segregation of the poor occurs on a large scale. An extensive slum may be the only world that children and many adult inhabitants ever see, or know. In school and in the 'hood, a child's peers may all come from poor households, with no role models or social networks to make social mobility possible. The recent African–American experience is a good example. Since the 1960s, a substantial black middle class has emerged and its opportunity to move into still segregated, but prosperous, neighborhoods has increased. For the most part it has vacated the inner city, leaving behind a concentrated underclass whose problems are compounded by their physical isolation from middle class peers and employment. In the same period, jobs have moved to the suburbs where they can only be reached by car, and the economic prospects for the residents of inner city neighborhoods have deteriorated. In some cities this has been counterbalanced by inner city revitalization and residential gentrification. In others, however, the divorce of jobs and low income households has produced a spatial mismatch.

The presence of immigrants and ethnic minorities is one of the most important ways in which the social composition of slums can vary because it typically affects the degree and pattern of social interaction. Slums vary in their social cohesion. Because they are likely to contain a high proportion of people with low levels of education and other forms of social capital, and of those who have physical or mental disabilities, slums may lack social networks and organization. This fact has often been overstated by middle class observers, either because they were unfamiliar with life in slum neighborhoods or because they had preconceived ideas as to what should count as an appropriate form of social interaction. In varying degrees, every slum contains its own social order. This is especially true where many local residents share a minority culture, since the creation and maintenance of local social institutions may be viewed as necessary to group survival. Immigrants, above all, are likely to create formal and informal associations. Slums that contain a high proportion of immigrants may boast dense social networks, richer than those of more affluent areas where private and individualistic modes of life are viable.

Sometimes, discrimination may compel all members of a minority to occupy a ghetto. Ghetto may describe a concentration of people of any kind (e.g., 'student ghetto'), but the term usually refers to people of a particular ethnicity. It is not necessarily a slum: in recent years, many Chinese immigrants in North America have chosen to settle in affluent ethnic enclaves. In contrast, involuntary ghettoes that harbor minorities who face discrimination often contain a mixture of incomes. Here, even well to do members of a community may have to occupy what becomes a slum, in part because crowding is forced upon them by the absence of residential alternatives. Jewish ghettoes in some European cities during the 1930s were made into slums in this fashion as were some African–American neighborhoods in the early post war decades. Since 2001, rising tensions between Muslims and host societies in Western Europe, North America, and Australia will probably help ghettoize this community, producing similar – although hopefully less serious and only temporary – effects. The enforced concentration of educated, middle class people in slum areas can be a radicalizing force.

Changing Patterns and Types of Slums

The character of slums has changed, and now varies greatly. Slums, like the poor themselves, have always been with us. Concentrated districts of crowded and insanitary housing had always been a feature of ports. They became larger and more common as European powers, first Portugal and Spain, and then the Netherlands, France, and Britain, developed colonial empires from the sixteenth century onward. They grew in colonial cities such as Batavia and Mombasa and in metropolitan centers, notably London. But it was the industrial revolution which gathered momentum from the late seventeenth century that put slums on the cultural map. Some writers have suggested that this was because, for the first time, urban conditions for a growing middle class had improved sufficiently that it was now possible to treat as pathological the sorts of conditions that had previously been viewed as the norm. But most have assumed that the new concern with slums reflected the unprecedented growth of cities, coupled with the emerging political ambitions of the working class. The industrial revolution not only created a new type of urban center, the industrial city, but also enabled for the first time a sustained process of urbanization that, by 1850, made Great Britain the first urban nation. The term slum dates from the 1820s, being first applied to rooms and then to poorer districts in London and soon the burgeoning industrial cities of the United Kingdom, northwestern Europe, and then North America. Among the most graphic contemporary accounts are those provided by Charles Dickens, of London, by Friedrich Engels, of Manchester, and Emile Zola, of a fictional mining town in northern France. By the 1880s the term 'slumming' had been coined to describe how the middle classes visited and patronized slum districts, whether for amusement or charity. Soon, studies of poverty and slums by Henry Mayhew and Charles Booth (London), by Seebohm Rowntree (York), and by the Pittsburgh survey in the United States, helped pioneer the methods of social science.

In these cities, 'slum' was used in various ways. Sometimes it referred only to the very worst concentrations of dilapidation, disease, and vice. Often, however, it was used more broadly to refer to extensive districts of the working class. It became part of a polar ized discourse: Benjamin Disraeli spoke of *The Two Na tions* and by the end of the nineteenth century the social counterpoint became couched as slum versus suburb. It was in this era that the moral dimensions of slum dis course were most fully developed. Reformers sought to distinguish between the deserving low income worker and the undeserving poor and to separate them physic ally so that the former would not become infected by the undesirable traits of the latter. The decentralization of jobs and the suburbanization of the working class came to be seen as the ideal solution to the problem of the slums.

The twentieth century has seen a reorganization of the global geography of slums. In Western and then Eastern Europe, as in North America, Australia, Japan, and in a few of the newly industrialized nations such as Singapore and South Korea, slums have become less common. This trend has been hindered as well as helped by the state. Probably the most important, but least heralded, state initiative has been the introduction and enforcement of building regulations, and then mainten ance by laws. By the nineteenth century, all major European and North American cities had regulations to limit fire risk. By the early twentieth century, new regulations sought to improve ventilation and sanitation. By the 1930s cities commonly enforced by laws that re quired owners to maintain their properties to minimum standards. In theory, and in countries like Sweden in practice, controls almost eliminated substandard housing. In some countries, regulations have been complemented by large scale social housing schemes. These have been a mixed blessing. Low rise, garden city style housing in urban settings often worked well. But some projects were placed in isolated settings so that residents found it dif ficult to find, and reach, work. High rise blocks proved unsuitable for families, especially when they were poorly maintained. Too often, as in the 'grands ensembles' of suburban Paris, poor location and design combined to exacerbate and even create major social problems, as was dramatized by riots in November 2005.

Other state initiatives have done more harm than good. In the nineteenth century cities undertook slum clearance schemes that made crowding worse because they displaced people without providing an alternative. In England these were encouraged by Parliamentary Acts after 1868. Many schemes were used for purposes of civic beautification and pride, even in a city like Joseph Chamberlain's Birmingham, which was in other ways an exemplar of municipal socialism. Other schemes enabled railways to build lines to centrally located stations (e.g., London), or to make it easier for the state to exercise control over the unpredictable urban masses (e.g., Haussman's Paris). In the twentieth century, clearance was given a more altruistic tinge. In theory, for example under Britain's Greenwood Act (1930), it combined demolition with the construction of social housing. In the colonial world some of the earliest and largest examples are provided by India. There, beginning in Bombay in 1896, Improvement Trusts widened streets and built chawls (tenements). In practice, however, more people were displaced than were re housed; there was always a lag of several years; most of those who occupied the new units were more affluent than those displaced. It was a similar story with the various types of urban renewal schemes that most developed nations fashioned after 1945. Some decent housing was destroyed; many house holds were displaced but not re housed, and with little compensation; the new housing was physically adequate but often inappropriate and disliked. If slums have waned in the developed world, especially since 1945, it is largely because suburbanization has reduced the population pressure on inner city districts, and above all because rising incomes have made it possible for the majority of households to occupy at least minimally adequate hous ing. These trends, combined with the enforcement of building and housing regulations, have ensured that in western nations the housing problem no longer takes the form of the slum but of the homeless, for whom none of the available housing is affordable.

Since 1945 it is the developing world that has seen the most rapid rates of urbanization and, with it, the flour ishing of slums. These typically take the form of squatter settlements at the urban fringe and are different from the slums of western cities in the industrial era in a number of ways, mostly bad. First, the new slum settlements have not always depended upon the growth of urban industry, or indeed of any type of employment. There are striking exceptions, beginning with Hong Kong, and then in cluding Taiwan, South Korea, and Mainland China. But since 1945, in the developing world many people have moved to cities because of poor conditions in rural areas and not because city jobs are abundant. Their urban prospects, then, and those of the slums they occupy, are poor. Second, many of the new slum dwellers are squatters with no legal right to the property they occupy. They have little or no recourse when, as happened in the early 2000s in Zimbabwe, the state evicts them. Then, too, without title, owners cannot use property as col lateral, and have little incentive to make improvements. Third, the scale of modern slums is unprecedented. Some are huge, amounting to hundreds of thousands of people and constitute cities in their own right. Even those lo cated in smaller urban areas loom large in relative terms. In many cities in the developing world the majority of residents live in districts that may fairly be labeled slums. Even in the time of Dickens, London was never quite like that. This changed proportion undermines the capacity of municipalities to handle the problem. The closest analogy might be with those low income suburbs in North America which have had to provide basic services from a weak tax base. Examples in the Chicago area include Robbins, an African–American suburb of the 1920s which never prospered and Stone Park, a small white suburb that lacked the resources to prevent Al Capone making it a crime base in the 1930s. Financial problems compelled all of Toronto's low income suburbs to go bankrupt during the Depression, but at least these could be bailed out by the Provincial government. This is not usually an option in the developing world today.

A fourth difference is in the location of many of these settlements at the urban fringe. This point can be overstated. On the one hand, suburban slums were always more common than observers chose to emphasize. In the early twentieth century, the suburbs of many European cities, including Paris, Rome, and Athens contained thousands of low income workers, many of whom had built their own homes in poorly serviced districts. The same was true in North America, around cities like Los Angeles or Toronto, where contemporaries spoke of the suburban shacklands. These attracted less attention than the pockets of poverty in the inner city: they were less visible to the social elite and perceived to be less threatening. But they were common enough. On the other hand, cities in the developing world contain many inner city districts which can be described as slums. For example, the social geography of the poor in Cairo or Rio de Janeiro includes a mixture of city and suburban slums. But the overall balance is different. The usual nineteenth century slum was found in the inner city; the typical modern slum lies at the squatter fringe. This can have certain advantages: fringe land is cheaper and more abundant, and so settlements there are lower in density and tolerate poorer services. It can also have disadvantages, especially isolation from jobs, above all those in the formal sector that provide higher and steadier incomes.

There is one aspect in which the modern slum dweller is better off than her nineteenth century counterpart. Rates of infant mortality, the single best indicator of population health, sometimes exceed 0.100 (one in ten) in the poorer squatter settlements, more than ten times as high as in the cities of developed nations. This is an appalling toll. But rates were higher still in many nineteenth century cities, commonly exceeding 0.200. Physical conditions in modern slums are as bad as those of an earlier era, but medical knowledge is greatly improved, underlining the fact that culture, in the broad sense, matters.

Underlying these distinctions between slums in the developed and the developing world, but sometimes cutting across them, is the varied manner in which they have developed. Those areas that grow through a formal process of development take time to become slums. The concept of the formal sector refers to those parts of an economy that are visible to, and regulated by, the state. In real estate, it encompasses land sales that are legally registered, builders and tradesmen who are certified, and companies that pay relevant taxes. In societies, or areas, where the formal sector is responsible for land development, very little (if any) housing will be built by the private sector for the poor. It is rarely profitable. Without state intervention, the only way low income households can obtain accommodation is through a process of filtering down: dwellings built for the middle, or upper middle classes slowly deteriorate or become obsolescent. As they decline in value relative to new housing, they become occupied by lower income households. Larger dwellings may be subdivided, which enables still lower income households to move in, while hastening physical decay. Decline may become more rapid where a high proportion of the housing stock is owned by absentee landlords, who calculate maintenance on the basis of profitability. It is hastened when financial institutions red line an area, refusing to make loans because the risks are perceived to be too great. As housing deteriorates and is subdivided, families double up, and respectable areas decline into a slum. This process is not inevitable, as the recent gentrification of many inner city areas has shown. But it has been common enough to be one of the more common ways in which slums have formed.

Alternatively, many areas were slums from the beginning. There are three ways this can happen. First, they can arise in squatter settlements where owner builders lack the incentive, knowledge, and/or resources to build decent housing and so erect shacks where municipal regulations that govern construction are nonexistent or ineffectively enforced and where no municipal government is able or willing to provide basic services. Second, less excusable, are the public housing estates that became instant slums because they were allowed to become overcrowded, because construction and then maintenance was skimped, or because the design of the housing was inappropriate to their occupants' culture and needs. Overcrowding has often been allowed by public housing agencies because evictions force people onto the street; skimping has reflected political corruption as well as the usual shortage of funds; poor designs have often reflected the use of styles developed in the west as symbols of modernity. Third, occasionally entrepreneurs see profits in the building of cheap tenements that are poorly built and designed, as well as overcrowded.

Slum Discourse and Public Policy

Not all of the districts labeled as slums have required public action to solve their problems. Like other labels, 'slum' simplifies a complex reality and, when made the basis for public action, can acquire a regrettable power. Outsiders, including middle class reformers, have often used the term inappropriately to promote a physical solution that was unnecessary. Sometimes they have failed to understand that a physically inadequate area could be socially viable, and be able to address most if not all of its own problems. Sometimes they have de ployed the label cynically to justify policies, for example, slum clearance, that were intended to accomplish un related goals. From the late nineteenth century until the 1960s, the discourse of slums was a powerful tool that was used by social reformers and urban planners, not always wisely or justly. Beginning, especially, during the 1960s its continued misuse helped inspire a backlash among writers and academics. They raised questions about the meaning and usefulness of the term, to the point that it is now rarely used in the developed world even in situations where it could be justified. They also raised questions about previous attempts to solve the slum problem.

The evolution in slum discourse has been apparent among geographers, who have used the term extensively because it meshes with persistent themes in the discip line. Marked by high densities and physical deterioration, slums are distinctive elements in the urban landscape. With concentrations of the poor, and other disadvantaged minorities, they dramatize the way social distinctions and exclusions are inscribed in space. Then, too, the associ ation of crowding and poor sanitation of slum areas with ill health and social deviance suggests how the built environment can affect social life and public health. Geographers have a particular stake in the concept.

Influenced by Charles Stokes, many writers now distinguish between slums of hope as opposed to des pair. The former, which are likely to be new and oc cupied by owner builders and recent migrants, are those where residents are in a position to make improvements, as long as the state allows them to do so. They may or may not require public assistance, in the form of mu nicipal services. Slums of despair are typically older and deteriorating, occupied by tenants and, in general, by those who, unaided, can do little to improve their lives. It is these districts that require the most vigorous public intervention. In reality, there is a continuum, while the prospects of even the slums of hope depend critically upon local job growth which in turn is shaped un predictably by the global economy. Hope, like slum, is a relative term.

Public intervention can take various forms. Until the 1960s, planners assumed that neighborhood decline was inevitable and that once an area had become a slum the only solution was to clear it. Influenced by Jane Jacobs, during the 1960s this view was successfully challenged in at least two ways. First, writers and planners now appreciate that physical planning may relocate, and perhaps improve, but cannot eliminate slums unless it is accompanied by social and economic programs that generate and/or redistribute wealth. Second, they now accept that physical upgrading is usually both cheaper and more effective than clearance, that the best methods will vary with local conditions (including patterns of property ownership), and that to be effective they must involve local residents rather than relying on the pater nalist decisions of local officials, however well inten tioned. Accordingly, starting in the 1970s, in the developed world, governments have encouraged the renovation of existing houses and neighborhoods. In the developing world, international agencies encouraged governments to support owner building by providing sites and services to upgrade slums (often by installing piped water and sewers) and to improve the governance of urban areas. Although these programs are better than those of the past, they have made little headway against the tide of urbanization. The United Nations estimates that one third of the world's urban population now lives in slums, mostly in the developing world. The challenge that they pose is partly one of urban planning and gov ernance, but even more it is one of poverty, inequality, and global development.

See also: Gentrification; Ghettos; Homelessness; Housing; Housing, Neighborhoods and Health; Industrial City; Squatter Settlements; Transport and Social Exclusion; Urban Policy.

Further Reading

Byington, M. (1910). *Homestead. The Households of a Mill Town*. New York: Russell Sage.

Clinard, M. (1966). *Slums and Community Development. Experiments in Self Help*. New York: The Free Press.

Davis, M. (2006). *Planet of Slums*. London: Verso.

Dyos, H. J. and Reeder, D. A. (1973). Slums and suburbs. In Dyos, H. J. & Wolff, M. (eds.) *The Victorian City. Images and Realities. Shapes on the Ground/A Change of Accent*, vol. 2, pp 359 386. London: Routledge and Kegan Paul.

Engels, F. (1952). *The Condition of the Working Class in England in 1844*. London: Allen and Unwin.

Faure, A. (2006). Local life in working class Paris at the end of the nineteenth century. *Journal of Urban History* 32, 761 772.

Fried, M. (1966). Grieving for a lost home. Psychological costs of relocation. In Wilson, J. Q. (ed.) *Urban Renewal. The Record and the Controversy*, pp 359 379. Cambridge, MA: MIT Press.

Harris, R. (1996). *Unplanned Suburbs. Toronto's American Tragedy, 1900 1950*. Baltimore: Johns Hopkins University Press.

Jacobs, J. (1961). *The Death and Life of Great American Cities. The Failure of Town Planning*. New York: Random House.

Mayne, A. (1993). *The Imagined Slum. Newspaper Representation in Three Cities, 1870 1914*. Leicester: Leicester University Press.

Riis, J. (1890). *How the Other Half Lives. Studies among the Tenements of New York*. New York: Scribner's. http://www.yale.edu/amstud/inforev/riis/chap6.html (last accessed 21 May 2007).

Shapiro, A. L. (1985). *Housing the Poor of Paris, 1850 1902*. Madison, WI: University of Wisconsin Press.

Stedman Jones, G. (1971). *Outcast London. A Study in the Relationship between Classes in Victorian Society.* Oxford: Clarendon.
Stokes, C. J. (1962). A theory of slums. *Land Economics* 38, 187 197.
Suttles, G. D. (1968). *The Social Order of the Slum. Ethnicity and Territory in the Inner City*. Chicago: Chicago University Press.
United Nations Human Settlements Programme (2003). *The Challenge of Slums. Global Report on Human Settlements 2003*. London: Earthscan.
Ward, D. (1989). *Poverty, Ethnicity, and the American City, 1840 1925. Changing Conceptions of the Slum and the Ghetto*. Cambridge: Cambridge University Press.
Yelling, J. A. (1992). *Slums and Redevelopment. Policy and Practice in England, 1918 45, with Particular Reference to London*. London: University College London Press.

Relevant Websites

http://www.laguardiawagnerarchive.lagcc.cuny.edu
La Guardia and Wagner Archives.
http://memory.loc.gov
Library of Congress. American Memory.
http://www.old8thward.com
Michael Barton and Stephanie P. Gilbert, Harrisburg's Old 8th Ward.
http://www.tenement.org
New York. Lower East Side Tenement Museum.
http://digitalgallery.nypl.org
New York Public Library Digital Gallery.
http://www.portcities.org.uk
Port cities UK.
http://www.archive.org
Prelinger Archive.

Smith, N.

A. L. Hansen, Lund University, Lund, Sweden

Glossary

Imperialism The production and protection of unequal economic, cultural, and territorial relationships (often, but not merely, between nation-states).

Lebensraum 'Living space'. The term comes from the German geographer Friedrich Ratzel's *Politische Geography* from 1897.

The International Critical Geography group (ICG) An international association established in the late 1990s to provide an alternative to the increasingly institutionalized and corporate culture of universities. The group is comprised of geographers and nongeographers committed to developing the theory and practice aimed at dismantling prevalent systems of capitalist exploitation; oppression on the basis of gender, race, and sexual preference; and imperialism, neoliberalism, national aggression, and environmental destruction.

Smith, Neil (1954 –)

Public intellectual and pioneer in the exploration of connections between space, nature, social theory, and history, Neil Smith was born in Leith, Scotland, in 1954 (**Figure 1**). He earned his BSc from the University of St. Andrews, where Joe Doherty was an important mentor in the early stage of Smith's formal education. He left Scotland for Baltimore in 1977 to initiate his PhD research on gentrification – a phenomenon he first encountered in Rose Street in Edinburgh's New Town in 1972, again as an undergraduate in Philadelphia's Society Hill, and later in New York City's Lower East Side (and many places around the globe since then). Early on, Smith related the process to broader forces of uneven development. With David Harvey as his supervisor at Johns Hopkins University, Smith became deeply engaged with the work of Karl Marx, and he fully embraced the transition from the so called spatial science, taught to him as an undergraduate, toward a Marxist approach that equipped him to analyze modern capitalist society, first from an urban point of view, later from multiscalar perspectives.

After graduating from Johns Hopkins in 1982, Smith taught at Columbia University and at Rutgers University, where he was Chair of the Geography Department (1991–94) and a Senior Fellow of the Center for the Critical Analysis of Contemporary Culture. From 2000 he was distinguished Professor of Anthropology and Geography at the Graduate Center of the City University of New York (CUNY) and he directed the Center for Place, Culture and Politics. Through the center, Smith has created a vigorous interdisciplinary intellectual environment where graduate students, faculty members, and the public are engaging in the exploration of crucial contemporary issues. The center's themes have included: the new internationalism, crossing borders, imperialism, war patriotism and resistance, transformative cities, democracy shrugged, catastrophe and discontent, and geopolitics and insecurity. He has received many awards and honors for his work. For example, his 2003 book, *American Empire* won the Los Angeles Times Book Prize for biography, the AAG Globe Book Award, and the Henry Adams Prize.

Neil Smith has produced groundbreaking intellectual contributions in the field of human geography for three decades. His work spans the identification of theoretical concepts such as the rent gap, as well as extensive archival studies that underpin studies such as the prize winning *American Empire.* In 1979 Smith published an article where he identified the 'rent gap' (the difference between actual and potential grond rent) as a key force behind urban change, emphasizing the importance of patterns of investment and disinvestment in the built environment. In a later book on *The New Urban Frontier* he wrote more broadly on New York City, characterizing it as a revanchist city, in which the logic of the market, the state, and police force produce unjust conditions for the urban poor and other socioculturally weak groups while the upper classes 'reclaim' the city. The book became a landmark text in

Figure 1 Neil Smith.

gentrification research, commonly contrasted with David Ley's book *The New Middle Class*.

Smith's book *Uneven Development* (1984–90–2008) concerns political economy and the production of uneven development, illustrating how space is constructed and reconstructed as expressions of (primarily capitalist) social relations. From this perspective, uneven development is the hallmark of the capitalist mode of production. In the same vein as David Harvey's *Limits to Capital* and Henri Lefebvre's *The Production of Space*, the book shows how and why space matters to capitalism – an element left almost untouched in Marx's work. Human societies under capitalism, the book argues, socially produce nature. A 'first nature', external to human societies, has been replaced by a socially produced 'second nature' through uneven means of the logic of market, private property, and free trade. The book provides a lens to critically analyze contemporary power struggles over the production of nature: e.g., global warming and the Kyoto Protocol, genetically modified organisms, biotechnology, and working and health conditions. Another key interrelated idea that derives from *Uneven Development* is the importance of the understanding of the dynamics of the production of geographical scale and, in later work, Smith followed this train of thought and developed a research agenda that has inspired a whole generation of social scientists to explore the question of 'scale'.

American Empire, published in 2003 to some acclaim, demonstrates how US global ambitions in the 1940s amounted to an attempt to create American global economic *Lebensraum*. According to Smith, the US led wars subsequently are to be viewed not only as wars on terrorism. Nor did the US just want to control the important oil resources. Rather the wars were designed to complete a US centered globalization that has characterized the third moment of US ambition since the 1980s. 'The endgame of globalization', as Smith calls it in a later book, is about gaining global control through geoeconomic means via the long term installation of neoliberal institutional frameworks.

Smith's work collectively identifies a series of shifts that crystallize a new stage in uneven development. In this new stage, Smith argues, the geographic dynamic of uneven development has shifted from national and regional economies to the nexus between global and local. From this perspective, the geographical scales of human activity are the product of changing social relationships, and are to be understood as something different from the traditional distinction between fixed or given urban, regional, national, and global scales. The transformation of the global political economy since the 1970s has involved a restructuring of scales at which different kinds of political, economic, and cultural activities are organized, leading to a new form of politics of scale.

Apart from major contributions to the understanding of critical issues of contemporary global societies, what makes Smith's work especially appealing is his unflagging energy to search for more just alternative futures, whether through his engagement with alternative survival strategies for homeless people (e.g., the homeless vehicle), his commitment to the discussions of a democratized production of nature, or in terms of his midwifery of, and ongoing passion for, the International Critical Geography group (ICG).

See also: Critical Geography; Gentrification; Harvey, D.; Imperialism, Cultural; Ley, D.; Nature; Neoliberalism; Public Space; Scale; Space I.

Further Reading

Castree, N. (2004). Neil Smith. In Hubbard, P., Kitchin, R. & Valentine, G. (eds.) *Key Thinkers on Space and Place*, pp 264 268. London: Sage.

Smith, N. (1979). Towards a theory of gentrification: A back to the city movement of capital not people. *Journal of the American Planning Association* 45, 538 548.

Smith, N. (1987). Dangers of the empirical turn: Some comments on the CURS initiative. *Antipode* 19, 59 68.

Smith, N. (1992). Contours of a spatialized politics: Homeless vehicles and the production of geographical scale. *Social Text* 33, 54 81.

Smith, N. (1996). *The New Urban Frontier: Gentrification and the Revanchist City*. London: Routledge.

Smith, N. (2002). Scales of terror and the resort to geography: September 11, October 7. *Society and Space* 19, 631 637.

Smith, N. (2003). *American Empire: Roosevelt's Geographer and the Prelude to Globalization*. Berkeley, CA: University of California Press.

Smith, N. (2004). Scale bending and the fate of the national. In Sheppard, E. & McMaster, R. (eds.) *Scale and Geographic Inquiry. Nature, Society, and Method*, pp 192 212. Oxford: Blackwell.

Smith, N. (2005). *The Endgame of Globalization*. London: Routledge.

Smith N. (2008) *Uneven Development: Nature, Capital and the Production of Space* (3rd edn.). Athens, GA: University of Georgia Press.

Relevant Websites

http://web.gc.cuny.edu
Center for Place, Culture and Politics.

http://econgeog.misc.hit u.ac.jp
ICG (International Critical Geography group).

Social Capital

A. Bebbington, University of Manchester, Manchester, UK

Glossary

Heuristics Devices or methods that – often in commonsense, easy-to-grasp ways – help solve problems in a manner that, though not perfect, is reasonably close to an adequate solution. These devices aid in thinking about problems and phenomena that, though they may lack conceptual precision, bring intelligibility.

Rent-Seeking A term used in economics and other social sciences to refer to efforts to secure financial gain not through production, trade, or other 'legitimate' economic activity but rather through manipulation of legal, regulatory, fiscal, or other such rules governing the economy, or through monopoly control of certain assets. It can refer to means that are both legal and illegal.

Washington Consensus (and Post-Washington Consensus) The Washington Consensus was a term initially coined by John Williamson to refer to ten economic policy recommendations that, he argued, constituted a more or less agreed-upon package for resolving macroeconomic crises and eliciting stabilization and growth. It has also been used in a derogatory way – by critics of such reform and adjustment-related policies – to refer to policy prescriptions deemed neoliberal.

Introduction

Since 1993, the term 'social capital' has rapidly risen to prominence in social science and policy debates. In development studies it has become a buzzword and the author whose work triggered its rapid popularization, political scientist Robert Putnam, was by some accounts the most cited social scientist of the 1990s. This remarkable explosion of interest in and use of the concept also generated a sort of 'counter literature', dissecting the limitations of the concept and the many ways in which it is used, and teasing out the (deliberately or not) hidden political project that lay behind the rise and rise of social capital.

The concept has become important in a range of intellectual and policy domains, among them health, community development, urban renewal, and local governance. Another such domain was international development, the focus of this article. In what follows, meanings that have been given to the concept are explored, tracing both their points of overlap and difference. The next section maps how the concept traveled into development debates, while the subsequent section discusses certain ways in which the concept has been used in development organizations and policy as well as in development studies. The penultimate section outlines some of the criticisms that have been advanced about the concept, and the article wraps up with a reflection on what may have been lost possibilities in these otherwise quite sharp and fierce debates on social capital and development.

Meanings of Social Capital

One criticism of the way in which social capital is used in these various policy and research domains has hinged around the multiple meanings given to the term (see later). This diversity of meanings should come as no great surprise, given the equally divergent understandings offered in the foundational works on the concept. Some authors suggest three such foundational authors – Pierre Bourdieu, James Coleman, and Robert Putnam – and we add a fourth here, Mancur Olson. Use of the term 'social capital' predates each of these authors, but it is in their work that it received some of its most systematic and widely accessible treatment.

For Bourdieu social capital is "the aggregate of the actual or potential resources which are linked to possession of a durable network of more or less institutionalized relationships of mutual acquaintance and recognition – or in other words, to membership in a group – which provides each of its members with the backing of the collectivity owned capital, a 'credential' which entitles them to credit, in the various senses of the word" (1986: 248–249). It constitutes one of several 'forms of capital' variously referred to as economic, cultural, symbolic, and social. Bourdieu insists that these forms of capital – though not reducible one to another – must be understood in relation to each other. That is, the distribution of social capital is related to that of economic capital, and together with cultural capital is used (consciously and unconsciously) to reproduce power relations and forms of access to resources that tend to reproduce existing distributions of power and capitals. Bourdieu's is a concept of social capital suffused with notions of power and political economy.

Coleman comes from a quite distinct intellectual lineage, though in important ways his approach to social capital overlaps with Bourdieu's. For him too, social capital is a resource that inheres in relationships – a feature of social structure – but can be appropriable by individuals. This social capital is understood (*inter alia*) as the norms, sanctions, obligations, and expectations in hering in these relationships. This capital can be mobil ized by individuals in their forms of action, and as a means of accessing resources. However, while Bourdieu's approach is one in which structure and institutionalized behavior are paramount, Coleman's analysis lays much more emphasis on individuals' calculations in the de ployment of social capital.

In their different ways, Bourdieu and Coleman em phasize the individually appropriable and contextual nature of social capital – its embeddedness in particular social structures is what gives it its value. In the work of Robert Putnam, while Coleman is invoked, social capital is understood less contextually and more collectively. Putnam emphasizes the norms of trust and reciprocity that emerge from participation in civic associations, and argues that these norms underpin a series of collective benefits: the creation of responsive government (through promoting civicness and accountability), dynamic econ omies (through easing transaction costs and reducing rent seeking behavior), and more generally, vital, en gaged communities (his book *Bowling alone* applies the argument to the US). Putnam also appears more inter ested in the properties of trust and civility that emerge from participation in groups rather than in the dynamics and groups themselves. This leads him toward a less contextualized understanding of social capital, and he certainly shows a greater willingness to work with ag gregate, de contextualized indicators and data.

This emphasis on trust and aggregate qualities of so cial life that enhance collective gain also underlay Mancur Olson's approach to social capital. Olson, how ever, was far less sanguine about the values of partici pation in organizations (particularly large ones), convinced as he was that many sociopolitical organiza tions rapidly lead to rigidities in the economy and be come vehicles for rent seeking behavior. He was, though, convinced of the value of generalized trust in society and for him – following Douglass North approach to insti tutions and economic performance – social capital in hered in those macro institutions producing trust.

Across these four traditions we have quite differing approaches to social capital: social capital can rest pri marily in micro sociological or macro institutional foundations; has to be understood in context, or does not; is appropriable by individuals or collectivities; and has to be understood in relationship to political economy, or does not. As one way of ordering this diversity, Uphoff distinguishes between approaches that emphasize cog nitive forms of social capital, and those that emphasize structural forms of social capital. The former hinge around values and perceptions in which norms of trust and civicness are paramount: they are primarily inter ested in the role of trust in influencing other societal (and development) outcomes and dynamics. The latter focus on social structures (networks, organizations, and rela tionships) and their role in facilitating access to other resources and institutions. These structural forms of so cial capital can in turn be broken down into different types. One schema distinguishes between so called bonding social capital (close ties between geographically and socially proximate actors), bridging social capital (connecting groups and organizations that, though simi lar, are distant in other regards, especially spatially), and linking social capital (those relationships between or ganizations and actors who are otherwise quite different, socially and in terms of their relative power). These distinctions help map out ways in which the concept was taken up in development studies and policy.

Travels to Development: The Take-up of Social Capital

There is no doubt that it was the publication of Putnam's *Making democracy work* in 1993 that launched the concept of social capital into wider public debate, and specifically into the world of international development. Various channels were at work here, but two are of particular importance: the World Bank, and a project of the American Academy of Arts and Sciences (AAAS), coordinated by sociologist Peter Evans. These two channels worked with the concept in somewhat differing ways that would each be apparent in later work on social capital and development.

Putnam's work flowed into the World Bank through two distinct routes. The first of these was via the Development Economics Vice Presidency. This is where the World Bank's economics research is located. Here the concept triggered two broad types of study. The first understood social capital in its cognitive sense. This research worked with attitudinal datasets, exploring the role of trust in macroeconomic and macro societal dynamics. The publi cations of Knack and Keefer are particularly important here and raised the significance of macro institutional arrange ments in helping foster such trust. Mancur Olson's con ceptualizations were as influential as Putnam's in this research. The second line of research sought to combine quantitative and qualitative work as a means of exploring the extent to which measures of social capital might give more robustness to more traditional econometric ap proaches to poverty. As Lant Pritchett, one of these authors would later note, the discussion on social capital helped focus the attention of 'socially aware' economists on the 'social something' that they knew helped explain forms of collective action, information sharing, and the like, but which they had not yet been able to account for or model adequately.

The second channel passed through the Vice Presi dency for Environmentally Sustainable Development, whose chief, Ismail Serageldin, began talking of social capital as part of the wealth of nations and as one of four assets for sustainable development. Serageldin suggested that a more complete assessment of the national wealth should include – and indeed account separately for – stocks of human, natural, and social capital as well as the more common measures of financial and fixed capital. The argument was that GNP/GDP and income per capita measures alone were inadequate measures of wealth. It was also argued that sustainability could best be thought of in terms of the mixes and trade offs among produced capital, natural capital, human capital, and social capital, and that sustainability could be con ceptualized in terms of the rules governing the substi tution among these different stocks of capital.

Here, as in Bourdieu and Coleman, social capital was being understood as one of several forms of capital. Also, again as in Bourdieu and Coleman, it was increasingly seen as a structural rather than a cognitive phenomenon. By the time of the high profile and polemic World Development Report of 2000/2001, *Attacking Poverty*, social capital is dis cussed primarily in terms of social relationships and organ izations, and in its bonding, bridging, and linking forms. This, however, was as far as the similarity with Bourdieu went. His attention to the structural relationships between the distri bution of social and economic capital is largely absent in these approaches, as is any real politicized understanding of power and political economy (and some readers will recall that one of the reasons why *Attacking Poverty* was polemic was precisely because of an effort to excise some such sensibilities from the report). In these broader senses, in the final, ap proved text social capital was conceived more as in Putnam's work – an asset whose dynamics and qualities could be considered apart from political economy.

A structural understanding of the role of social capital in development also emerged from Evans' AAAS project, though in this case the understanding was also far more contextualized and politically sensitive. This project – enquiring into the place of social capital in development – concluded that one of its most important roles was in fostering synergies: between state and civil society, be tween government and community, and among different civil society actors. Notwithstanding its importance, Evans and colleagues argued, such synergy was generally over looked in debates typically pitched in terms of state or market or civil society. Yet it could play important roles in fostering the emergence of pro poor public services, the creation and maintenance of collective goods, and the consolidation of pro poor and more democratized forms of government. Underlying such synergy were social rela tionships bridging institutional and political boundaries that served to foster both complementarity and account ability among actors. These relationships – understood as social capital – could be built both through long standing interactions embedded in deeper social structures, as well as through policy interventions 'and' in some cases, pol itical conflict itself. The possibility of their emergence also depended very much on prior political economies of de velopment and distribution.

Social Capital and Development Studies

As development studies' use of the concept of social capital has unfolded since these early interventions, structural (rather than cognitive) notions of social capital have become progressively more prominent, and the need for greater political economic sensibility has become ever more apparent. Social capital – understood as the re sources inhering in social relations, networks, and organ izations – has been understood in two main ways: (1) as an asset among several forms of capital and one that facilitates access to other capitals and institutional spheres; and (2) as a source of synergy among distinct development actors.

The asset based understanding of social capital has been most visible in approaches to livelihood, poverty, and welfare that understand these as a function of household access to different types of capital. The quantitative variant of this approach seeks to model the relative effects of social capital (as compared to other assets) on some de fined measure of welfare. As something of a derivative of this work, quantitative and economic approaches to social risk management have also explored the importance of social capital as an asset that helps people reduce risk and cope with crisis periods when they occur. Such quantita tive approaches have, at certain moments, played an im portant role in making the importance of social relations more visible in environments otherwise dominated by rational choice approaches in economics.

The qualitative variant of asset based conceptual izations has been perhaps most apparent in the large body of writing on livelihoods and vulnerability. Moser's work on the asset vulnerability framework was formative here, shortly followed by the sustainable livelihoods framework. Each afforded an important role to social capital, under stood as facilitating access to other assets and to insti tutions such as schools, healthcare centers, or financial service organizations that provide those assets. Social capital was deemed important for such institutional access through providing the necessary contacts and/or by guaranteeing the person's reliability and accountability. This argument has been particularly used in micro fi nancial services literature and practice, in which social capital (in the form of group membership) is taken as a guarantee that loans will be repaid. Another strand in this writing has viewed social capital as an important safety net, a means of reducing vulnerability. Here social rela tionships (formal or informal) are valued for the role that they can play in helping people recover from or cope with crisis, violence, or other sources of risk and perturbation.

The sustainable livelihoods framework was adopted by the UK Department for International Development (DfID) just as DfID itself was becoming a far more as sertive (and better financed) actor on the international stage. As such, DfID actively promoted the framework among its various national and international, govern mental and nongovernmental partners, and there can be little doubt that this helped further disseminate the no tion of social capital as an asset of the poor. Subsequently, enthusiasm for asset based livelihood frameworks has declined somewhat and – as with sustainability itself – the sense seems to be that the frameworks were more helpful as heuristics than as research tools. Indeed, if they were to play an important role in structuring research they still had to be combined with broader theories of development and of the dynamics driving asset distri bution. Too often, perhaps, use of the framework re placed any serious engagement with theory.

Social capital as an asset fostering institutional syn ergy has received somewhat less attention, a negligence that is perhaps ironic if one were to consider the extent to which people operating within development organiza tions (and development studies institutes) depend on social capital assets as they build the institutional ar rangements that allow them to do their own work. That said, there have been interesting attempts to understand the workings of national poverty reduction programs in these terms. Among academic researchers, Elinor Ostrom's important corpus of work on institutional ar rangements for delivering development (in its varying dimensions) continues to demonstrate the importance of social capital as a source of synergy. Jonathan Fox's in depth research in Mexico does the same, while adding a political economy analysis to understanding why certain forms of social capital and synergy emerge (or are absent) in particular contexts.

Criticisms

The concept of social capital has, in development studies at least, been criticized with as much gusto as it has at other times been celebrated. Some of these criticisms hinge around what are seen to be the analytical short comings of the concept and its use; others are more concerned about the larger political project that the concept is deemed to serve and that those using it are deemed to be promoting (deliberately or not). In this section, we consider each type of criticism.

Analytically, probably the greatest complaint about social capital has been that it has been used to refer to so many dimensions of social life that it becomes vacuous. At different times, social capital refers to trust, values of reciprocity, and norms of civicness; at other times, it is used to refer to the social relationships that produce these values. At different times, social capital is used to refer to organizations and associations, while at other times, it refers instead to the qualities inherent within the everyday social relationships that can make possible the building of organizations; and so on. The concept is sometimes used to refer to dimensions of social relations that deliver particular results (e.g., trust), at other times it is used to refer to that trust, and at still others it is used to refer to both the relationships and their effects. Put an other way, much social capital writing suffers from con fusion over what constitutes the dependent and the independent variables at play.

A closely related problem in social capital research has been to find adequate indicators. Depending on the study in question, studies have used as indicators everything from answers to attitudinal survey based questions regarding how far respondents trust other people through to the number of organizations people belong to and the frequency with which they participate in their activities. Some efforts have been made to de velop standardized tool kits, with different indicators for different types of research and development problems, but it remains the case that social capital research is plagued both with lengthy explanations to justify the selection of particular indicators, and then with critiques which hinge as much on the limitations of the chosen indicators as on more analytical and substantive issues.

Another set of criticisms relates to the explanatory power of the concept. Here the argument is that, even if the structure of a person's social relationships might help explain elements of their livelihood, far more variance in both economic and political practice can be explained by variables other than social capital. Authors drawing on certain Marxian traditions argue that class – defined, broadly, in terms of relationships to the means of pro duction – is far more powerful an explainer than is social capital. Often this argument is made in broad theoretical terms though it has also been made empirically. Das for instance uses quantitative and qualitative data to argue that poverty levels in rural India are far more evidently an effect of class position than of wage laborers' social capital. He also argues that for the very poorest in India, social relationships are not much of an asset because they are generally with equally poor and asset deprived per sons and so open up very little by way of access to other resources. Cleaver has made similar criticisms of the concept on the basis of research in Tanzania. For her, social capital is of little use either to explain the origins of chronic poverty or to identify ways out of such pov erty. This is so because of:

> three factors, which severely constrain the ability of the poorest people to actively construct or benefit from social capital through association and public representation. The dependence of very poor people on their own

> able bodiedness and their inability to sustain this is the first constraining factor. Secondly, they experience little room for maneuver in their kin and wider social relationships, due to small fragile families, unstable marital arrangements, and wider derogatory perceptions of the poor. Finally, the poorest people are constrained by their inability to ar ticulate successfully in public fora and even where the voices of the poor are heard, they are given little weight and exert negligible influence (Cleaver, 2005: 895 6).

In short, people are too subject to illness, disablement, lack of education, and general structural constraint to be able to participate in and gain advantage from social organizations.

Observations, such as Cleaver's, suggest a further problem with the concept, or at least the way it is gen erally used. That is, it encourages a focus on social re lationships, networks, and values in a way that draws attention away from political economy. Emphasizing the potential role of social capital in escaping poverty or building democratic institutions has the effect – intended or not – of de emphasizing the influence of local, na tional, and global political economy factors in producing poverty, weakening democracy, and constraining liveli hood options. Whether the use of social capital 'needs' to be so blind to political economy is, however, another question – as we have noted, Bourdieu's version of the concept is explicitly embedded in a notion of structure and the relationships between economic capital and so cial capital. Still, it is the case that development writing has been far more influenced by Putnam's political economy weak version of the concept than by Bourdieu's political economy strong approach.

If social capital is so limited in its explanatory cap acity, critics ask 'why has it proven so popular and so resilient?' The answer often given opens up a different terrain of critique. This critique questions the political motivations – conscious or unconscious – behind the use of social capital. By ignoring relationships of power and difference, and the relationships between social struc tures, economic arrangements, and the dynamics of ac cumulation, the concept lends itself to work that takes as given the basic arrangements governing contemporary social and economic life. As such, it is – the critique goes – a neoliberal (and neoliberalizing) concept which is therefore deemed to serve the larger projects of both the Washington and post Washington Consensus. In some hands it might become a reactionary concept, accepting the *status quo* and delivering interpretations that for many critics have the whiff of 'culture of poverty' and 'blame the victim' arguments that see people as poor not be cause of massive resource transfers out of their countries, but because they have not made the effort to invest sufficiently in their own social organizations and networks.

Variants of this position seem broadly shared in the literature in critical development studies and critical development geography, and seem to explain the energy with which some authors attack, and dedicate so much effort in trying to destroy, the concept (indeed, for some this has been the explicitly stated goal). It raises the question, however, as to how far the concept 'necessarily' serves such a project. The article ends with a brief re flection on this.

Final Words

The concept of social capital has proven to be something of a lightening rod, in social science in general, and de velopment studies/development geography in particular. The debates that it has elicited have served to mark clear lines between scholars and between political projects. These lines mark differences not just between radical and conservative projects, but also between radical and re formist projects within the broad socialist/social demo cratic left. To some extent, the debate has been characterized by more or less relentless insistence on the part of radical critics that social capital can only be de politicizing, reactionary, and obfuscating; and a more or less equally relentless insistence from those who have worked with the concept that, on the contrary, it can be useful for understanding how institutional and political boundaries are transgressed and how reformist progress is made toward more inclusive and democratic societies.

Potentially bridging these different arguments is the question as to whether it is possible to build what Rankin calls a 'Marxian approach to social capital' one which would be centered on "a structural analysis of social capital … [that] … draws attention to the ideological as well as material dimensions of social change." Such a project would seem to have been the intent (though clearly not the major one) in Bourdieu's conceptual ization of social capital as one of several forms of capital, a project that did seek to keep hold of some notion of class and political economic structure while also seeking the sources of variance in the practices that derive from such structures.

There is some evidence of development scholars slowly moving away from the concept of social capital as a research tool (if not as a heuristic). This seems to have happened before any serious and sustained attempt has been made to respond to Rankin's challenge. This seems unfortunate, for it is also the case that class based and structural analyses suffer many explanatory limitations, especially when it comes to understanding the nitty gritty of building social movements and political alliances for ushering in and sustaining progressive social change. A Marxian approach to social capital may have had a lot to offer here, and it is not casual that a social movement theorist such as Crossley suggests that one of the most fruitful ways forward for theorizing social movements is to work with Bourdieu's theory of practice and forms of capital. One wonders though, whether within development studies the intensity and fierceness of the debate on social capital has had the effect of crowding out the possibility of elaborating this reflection – a reflection that would have been as strategically and politically valuable as it would have been analytically progressive.

See also: Civil Society; Governance; Livelihoods; Sustainable Development; Vulnerability.

Further Reading

Bebbington, A., Woolcock, M., Guggenheim, S. and Olson, E. (eds.) (2006). *The Search for Empowerment. Social Capital as Idea and Practice at the World Bank*. Bloomfield: Kumarian Press.

Bourdieu, P. (1986). The forms of capital. In Richardson, J. G. (ed.) *Handbook of Theory and Research for the Sociology of Education*. New York: Greenwood Press.

Cleaver, F. (2005). The inequality of social capital and the reproduction of chronic poverty. *World Development* 33(6), 893 906.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology* 94, S95 S120.

Crossley, N. (2002). *Making Sense of Social Movements*. Buckingham: Open University Press.

Das, R. (2004). Social capital and poverty of the wage labour class: Problems with the social capital theory. *Transactions of the Institute of British Geographers* NS29, 27 45.

Evans, P. (ed.) (1996). *State Society Synergy: Government and Social Capital in Development*. Berkeley, CA: Institute for International Studies.

Fine, B. (2001). *Social Capital versus Social Theory. Political Economy and Social Science at the Turn of the Millennium*. London: Routledge.

Knack, S. and Keefer, P. (1997). Does social capital have an economic payoff? A cross country investigation. *Quarterly Journal of Economics* 12(4), 1251 1288.

Moser, C. (1998). The asset vulnerability framework: Reassessing urban poverty reduction strategies. *World Development* 26(1), 1 19.

North, D. (1990). *Institutions, Institutional Change, and Economic Performance*. New York: Cambridge University Press.

Olson, M. (1982). *The Rise and Decline of Nations: Economic Growth, Stagflation, and Social Rigidities*. New Haven: Yale University Press.

Putnam, R. (1993). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, NJ: Princeton University Press.

Rankin, K. (2001). Social capital, microfinance, and the politics of development. *Feminist Economics* 8(1), 1 24.

Scoones, I. (1998). *Sustainable rural Livelihoods: A Framework for Analysis*. Working Paper 72. Brighton: Institute for Development Studies.

Uphoff, N. (1999). Understanding social capital: Learning from the analysis and experience of participation. In Dasgupta, P. & Serageldin, I. (eds.) *Social Capital: A Multifaceted Perspective*. Washington: World Bank.

Woolcock, M. (1998). Social capital and economic development: Towards a theoretical synthesis and policy framework. *Theory and Society* 27(2), 151 208.

Relevant Websites

http://www.socialcapital foundation.org
A Belgium based NGO dedicated to social capital issues.

http://www.livelihoods.org/
A repository of thinking about livelihoods in which concepts of social capital feature. Closely related to the UK Department for International Development.

http://www.hks.harvard.edu/saguaro/
The Saguaro Seminar is an initiative of Robert Putnam to discuss social capital in the US. Many useful on links to Putnamesque thinking on social capital and community.

http://go.worldbank.org/VEN7OUW280
This links to World Bank resources on social capital.

Social Capital, Place and Health

L. Twigg and J. Mohan, University of Portsmouth, Portsmouth, UK, and
University of Southampton, Southampton, UK

Glossary

Collective Efficacy The undertaking of coordinated, group, or collective activities on behalf of the common good.

Ecological Fallacy The fallacy of making inferences about individuals from aggregated data.

Ecological Studies Research using data for areas as opposed to data for individuals.

Epidemiology The study of the causes, distribution, and control of diseases in the population.

Multilevel Models The technique of simultaneously modelling data collected at a number of geographical levels, whereby lower level units (e.g., individuals) nest in upper level units (e.g., neighborhoods, counties or regions).

Prospective Study A study designed to follow participants forward through time.

Reciprocity The act of undertaking a favor for another person, or group of people, without necessarily expecting something in return immediately. Good will prevails and the favor may or may not be called in at a later date.

Social Epidemiology The study of the links between the social environment and the causes and distribution of disease within populations.

Introduction

Interest in geographical variations in health status and health outcomes can be traced to the writings of the fifth century physician, Hippocrates, whose well known work, *Of airs, waters and places*, discussed the natural environ ment on health. Today the dominant focus is concerned with the influence of the social environment in ex plaining persistent health inequalities across individuals and populations. This focus, up until the late 1990s, centered on material circumstance and lifestyle. How ever, even after allowing for these factors, there remains unexplained variation, and so analytical attention has moved to the notion of social capital in determining health outcomes.

Concepts, Definitions, and Measurement

Broadly speaking, social capital refers to the social or ganization of neighborhoods and communities and in particular to the formal and informal patterns of associ ational life and community interaction through which relationships of trust are (or are not) established. In areas where social capital is strongly present, it is suggested that economic transactions, trust in governmental insti tutions, and psychosocial well being are all facilitated, to the benefit of all. The term 'social capital' has been present in various guises throughout the twentieth cen tury across many subject areas. It is claimed that social capital has beneficial effects on both individuals (e.g., promoting better health, facilitating child development) and communities (e.g., stimulating economic sustain ability and promoting organizational development). Due to this apparent ubiquity, social capital has been taken up by governments and NGOs in both the developed and developing world. The World Health Organization, for example, introduced a 'Civil Society Initiative' to help promote a wider understanding of the concept and the World Bank has introduced a Social Capital Initiative to improve understanding and promote dissemination of information regarding policy interventions and best practice to aid in poverty reduction.

For a geographical audience the work of Robert Putnam, on social capital as a property of communities, has attracted much attention, but the reason Putnam has reached a much wider audience is because of his ex tensive writings about the demise of social capital in the US. In Putnam's formulation, emphasis is placed on the amount and quality of social interactions among indi viduals, through participation in relatively non hierarchical associational activities which cut across social divides (Putnam's celebrated examples include choral societies and bird watching clubs); this is said to inculcate norms of trust and reciprocity. Communities with high social capital are made up of individuals who tend to work for the benefit of the group. They have high levels of respect for fellow community members and treat others as they would expect to be treated themselves. It is through such community interplay that levels of col lective efficacy are maximized.

Distinctions can also be made between bridging, bonding, and linking social capital. The first form refers to the relatively weak links that exist between different sets of connections or different social groups. This form of social capital is seen as inclusive and tends to help people get ahead through networking. The example commonly described is the use of such connections when someone is trying to find a job or change their career direction. Second, bonding social capital represents much stronger attachments within existing groups that are so cially similar, for example, based on ethnicity, university fraternity, or professional grouping. These tend to be exclusive, inward looking, and help people get along by strengthening the bond between group members. How ever, individuality may be suppressed and outsiders may have difficulty gaining access and negative outcomes are always a possibility (criminal gangs can be regarded as exemplars of bonding social capital). Third, linking social capital refers to vertical linkages across hierarchies of power or status. An example would be the interplay be tween a community based tenants' group and the local agency or authority responsible for social housing in the area.

How can social capital be measured? Analysts have distinguished between the structural and cognitive components of social capital. Structural elements refer to the quantity and quality of associational links or activity and cognitive elements relate to perceptions of support, reciprocity, and trust (some authors have argued that social surveys should routinely include questions asking people what proportion of their fellow citizens can be trusted). The former are arguably more readily open to operationalization and subsequent measurement (e.g., memberships of community organisations), whereas the latter are sometimes more problematic to consistently assess, particularly in comparative studies due to the context sensitive nature of responses (e.g., regarding ob ligations to or perceptions of one's neighbors: inter national comparisons show that the interpretation of these issues varies across nation states). In other words, peoples' perceptions of neighborhood are influenced by the locality in which they reside and variations in per ceptions of trust, for example, may not necessarily be an index of social capital but rather an outcome of it.

Most of the indicators of social capital described thus far are not measures that are truly ecological in nature. Rather, they are based on the composition of areas and comprise aggregated counts of individual activities (e.g., volunteering, social club membership, church attendance, voter turnout) and/or attitudes (e.g., levels of trust and perceived levels of community spirit). Genuine eco logical measures are much more difficult to capture al though some North American research has attempted to do this through covert monitoring of neighborhood social interaction through the use of hidden videocameras. Other indicators have included the speed with which householders clear snow from the public pathway in front of their property after a snowstorm, though clearly this is context specific!

How do these concepts differ from traditional ideas of social support and levels of human capital? Most defin itions argue that it is the ecological nature of social capital that distinguishes it from the former and geographical analyses now regularly uses indicators of social capital for geographical areas of varying size, such as levels of civic participation (e.g., voter turnout), rates of volunteering, density and levels of club membership, perceptions of trust, and the levels and nature of inter action between immediate neighbors.

Geography and Social Capital

It has already been noted that social capital is usually thought of as a property of the places in which people live, and spatial variations in social capital have been associated with numerous outcomes including health. Social capital is conceptualized as a contextual or eco logical (i.e., areal level) explanatory concept, but attach ing a measure of social capital to a particular place involves a margin of error, because it is unlikely that aspects of social capital (e.g., the extent to which people feel they can trust their fellow citizens) will map neatly onto the spatial units for which appropriate data are available. Moreover, indicators of social capital such as civic participation and membership of voluntary associ ations will vary across different ages, genders, ethnicities, and socioeconomic groups. There is, of course, a geog raphy to these compositional variables and so the geog raphy of social capital will reflect spatial variations in such characteristics. Other features of places may be conducive to the formation of social capital, such as more affordable and/or accessible spaces to conduct activities. Furthermore, spatial variations in social capital may also reflect uneven area targeting policies associated with social capital activities. Charitable or voluntary organ izations, for example, may target requests for help in more affluent areas. This challenges the assumption that high rates of social capital precede economic develop ment, as opposed to following it.

A useful illustration of these issues is the geography of voluntary activity in England. There are regional vari ations in the proportion of the population who are active volunteers, which prompted the influential Wolfenden Committee on voluntary action (1978) to remark that "the soil for voluntary activity appeared more fertile in some locations than others." On the face of it such a quote would support a view that there are significant regional variations in social capital. However, when al lowance is made for compositional factors such as gender, socioeconomic status, and age, geographical variations in voluntary activity reduce substantially. There are still variations, which reflect unobserved heterogeneity (e.g., variations in the propensity to volunteer might re flect variations in the distribution of organizations for which volunteering takes place). The variations that exist are less substantial than appears to be the case at first sight.

Social Capital and Health

Historically in the social epidemiology literature, ideas relating to social capital, such as social support and isol ation, have regularly been used to explain variations in health outcomes. One of the most widely cited examples is the Alameda County prospective study which illustrates that strong ties with close relatives and friends are signifi cant predictors of lower 9 year mortality risk for persons aged less than 70 years at baseline. However, since the mid to late 1990s, there has been a dramatic rise in interest in the impact of social capital upon health. If the search phrase 'social capital and health' is used in the Science and Social Science Citation Index, 16 hits are returned for the period 1970–98; in contrast, for the time period 1999 to June 2006, 369 publications are listed. There are a number of ex planations for this sudden growth in interest.

The beginning of this article indicated that debate concerning the link between social capital and health builds on previous work that attempts to assess the relative impact of the social environment on health. Two ex planatory problems here were the apparent inability of material circumstances (both individual and areal) and lifestyle differences to satisfactorily explain spatial in equalities in health, and the persistence of health gradi ents, even among the more affluent in society. These problems led the epidemiologist Richard Wilkinson to hypothesize that the existence of social hierarchies *per se* impact on health outcome. He argued that, for developed countries, it is the most egalitarian, rather than richest nations that have the best health outcomes. One of the major pieces of evidence in support of this idea is illus trated in **Figure 1** where those countries that exhibit largest income inequalities tend to have the poorest life expectancies. In contrast, **Figure 2** indicates that the link between life expectancy and GDP per capita is not so strong. According to Wilkinson and his supporters, large income differentials in any one nation lead to a breakdown in social cohesion, resulting in isolation, anxiety, and in security among individuals which in turn leads to chronic stress. Such psychosocial processes lead to poorer health 'inside' (i.e., through biological processes) and 'outside' of the body (i.e., through social processes). Negative bio logical processes may result in adverse health outcomes directly via the psycho neuro endocrine mechanism or indirectly as a result of stress induced unhealthy lifestyle behaviors (e.g., increased alcohol consumption or tobacco consumption). Due to the existence of gradients in society (which are evident within organizations which employ large numbers of well educated individuals, such as the British civil service – there is a strong mortality gradient between different grades of the service), social cohesion

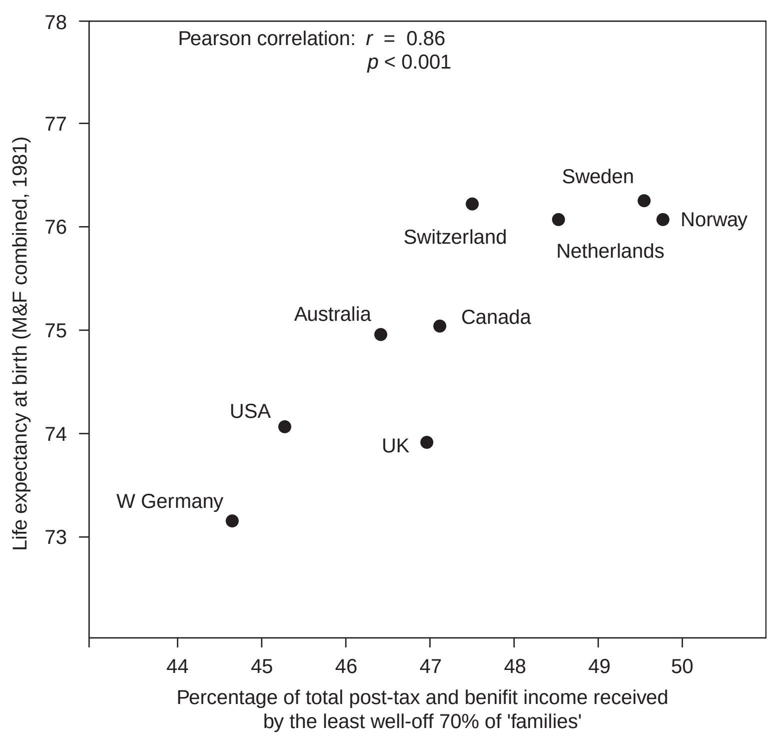

Figure 1 A scatterplot showing the relationship between male and female life expectancy at birth and income inequality. Income inequality decreases to the right-hand side of the *x*-axis where life expectancy tends to be higher. Reprinted from Wilkinson, R. G. (1996). *Unhealthy societies*: *The afflictions of inequality.* London: Routledge. Source: Data from Bishop, J. A., Formby, J. P. and Smith, W. J. (1989). International comparisons of income inequality: Luxembourg Income Study, Working Paper 26.

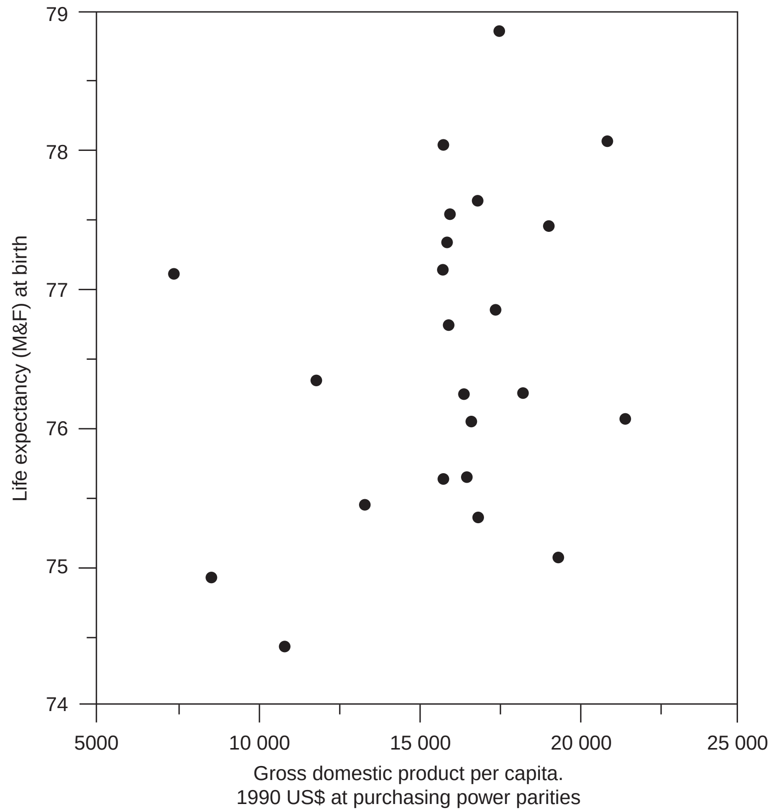

Figure 2 GDP per capita and life expectancy in OECD countries in 1990 (classes are split by standard deviation). Reprinted from Wilkinson, R. G. (1996). *Unhealthy societies*: *The afflictions of inequality*. London: Routledge.

breaks down as invidious comparisons across the hier archies are made, creating the potential for antisocial be havior and an erosion of social capital. Therefore in the Wilkinson hypothesis, social capital is seen as the inter vening variable connecting income inequality and health. Putnam draws on this research, and on numerous quan titative and qualitative studies, to argue that social con nectedness is one of the most influential determinants of our health and well being. He is not averse to making strong claims, such as the argument that as a 'rough rule of thumb' an increase in social engagement for an individual of belonging to just one group (as opposed to none) re duces the risk of dying over the next year by 50%.

The inequality theory has been tested across many developed countries and much of the early evidence involved large scale studies using data for the US states. For example, **Table 1** shows the correlation results from a study investigating the associations between income inequality, low levels of trust, group membership, and firearm violent crime. At the state level, large positive correlations are derived between rates of homicide, fire arm crime, and indicators of mistrust. Negative correl ations with levels of group membership are also illustrated. More importantly, a path analysis reveals that the relationship between income inequality and age adjusted firearm homicide is, in part, mediated by social capital.

Consistent associations have also been found in other nations such as Russia and Hungary but conflicting re sults have been found for other OECD countries (e.g., Canada). Furthermore, there is a wealth of literature and evidence that challenges the income inequality and health link on a number of points. First, the results are not always replicated in other developed nations. Similar research, such as our own investigation of the effect of social capital on mortality variations across the 15 standard regions of England, finds no substantive evi dence for the direct impact either of social capital (as measured through notions of community and friendliness of neighbours) or of income inequality on all cause mortality, after controlling for median income. Moreover there is no evidence to suggest that social capital operates as a mediator in the relationship between income in equality and health. Second, some strong claims were made by Putnam and others for the causal efficacy of social capital on the basis of relatively simple statistical analyses based on aggregate data for large regions. Apart from the difficulty of conceiving how US State or re gional measures of social capital can be meaningful when it is assumed that social capital is a community or

Table 1 Correlations among indicators of social capital, income inequality, and firearm violent crime (39 US states)

		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
1	Age-adjusted homicide									
2	Age-adjusted firearm homicide	0.99*								
3	Firearm assault	0.61*								
4	Firearm robbery	0.57*	0.56*	0.53*						
5	Firearm suicide	0.41*	0.48*	0.50*	0.45*					
6	Income inequality	0.73*	0.76*	0.48*	0.31*	0.43*				
7	Group membership	−0.51*	−0.49*	−0.34*	−0.33*	−0.05*	−0.40*			
8	Trust 1[a]	0.82*	0.83*	0.55*	0.52*	0.49*	0.73*	−0.54*		
9	Trust 2[b]	0.72*	0.73*	0.34*	0.46*	0.44*	0.71*	0.65*	0.79*	
10	Helpfulness[c]	0.72*	0.75*	0.60*	0.54*	0.51*	0.70*	−0.54*	0.81*	0.78*

[a] Percent responding: 'most people would try to take advantage of you if they got the chance'.
[b] Percent responding: 'you can't be too careful in dealing with people'.
[c] Percent responding: 'people…mostly look out for themselves'.
* $p < 0.05$.
Source: Reprinted from Kennedy, B. P., Kawachi, I., Prothrow Stith, D., Lochner, K. and Gupta, V. (1998). Social capital, income inequality, and firearm violent crime. *Social Science and Medicine* 47(1), 7 17. With permission from Elsevier.

neighborhood concept operating at local scales, these investigations are problematic because of the ecological or aggregative fallacy. The results are based on averages for very large areas and while some of these studies attempt to control for material circumstances by including average or median state income, very few of these early studies include individual level measures of material circumstance, poverty or wealth. Hence, research cannot distinguish contextual effects (those due to the nature of a place) from compositional effects (effects of what is in a place – in this case, the characteristics of the people who live there).

Following on from and responding to these technical criticisms, the statistical approaches have become more elaborate and robust by using multilevel methodologies to model individual health outcomes against area levels of social capital, while controlling for individual and contextual covariates. Some support has been provided for the effects of social capital, but research findings are not always consistent. For example, complex effects of community level social capital on the probability of reporting self rated poor health were found in a study of 40 communities in the US. Higher levels of community social trust were associated with a lower probability of reporting poor health. Controlling for individual level perceptions of trust, however, rendered the main effect of community level social trust statistically insignificant. However, there was a complex interaction effect, in that the health promoting effects of community level social trust were apparently greater for high trust individuals. This suggests that if there are health benefits to be gained from social capital, they are not equally available to all. Furthermore, once allowance was made for the individual compositional effects of socioeconomic status, there was little impact of communities on poor self rated health. Such quantitative, multilevel studies that are able to use community or neighborhood level data across large geographical areas are relatively rare. While social capital questions are becoming a common occurrence in many social surveys, such sources are usually unable to generate reliable results below the region or state level and through necessity, most quantitative studies employ measures of social capital summarised for these relatively large geographical levels. The scale of these units – American states, British standard regions, etc. – renders it difficult to identify and interpret causal processes which might link social capital and health across these large areas.

To overcome this problem data on social capital are needed for small areas, but the kind of systematic social observation discussed above is not feasible for large numbers of areas. An alternative on which we have drawn is the growing body of work on synthetic estimation via multilevel modeling. **Figure 3**, for example, shows estimates of volunteering across the electoral wards of England and Wales. These were generated by applying small area census data to multilevel model equations which summarize the probability of an individual undertaking voluntary work for more than 2 weeks over an annual period (hence the term 'core' volunteering in the title of the map). The multilevel equations incorporate the importance of individual characteristics on the likelihood of volunteering (e.g., age, gender, and marital status) as well as including estimates for the impact of contextual or small area variables (e.g., levels of car ownership, deprivation scores, social class, and ethnicity).

These small area estimates can be validated against *ad hoc* surveys and the technique has been successfully applied to a number of social capital indicators including political, social and altruistic activity, perceived importance of local friends, perceptions of neighborhood,

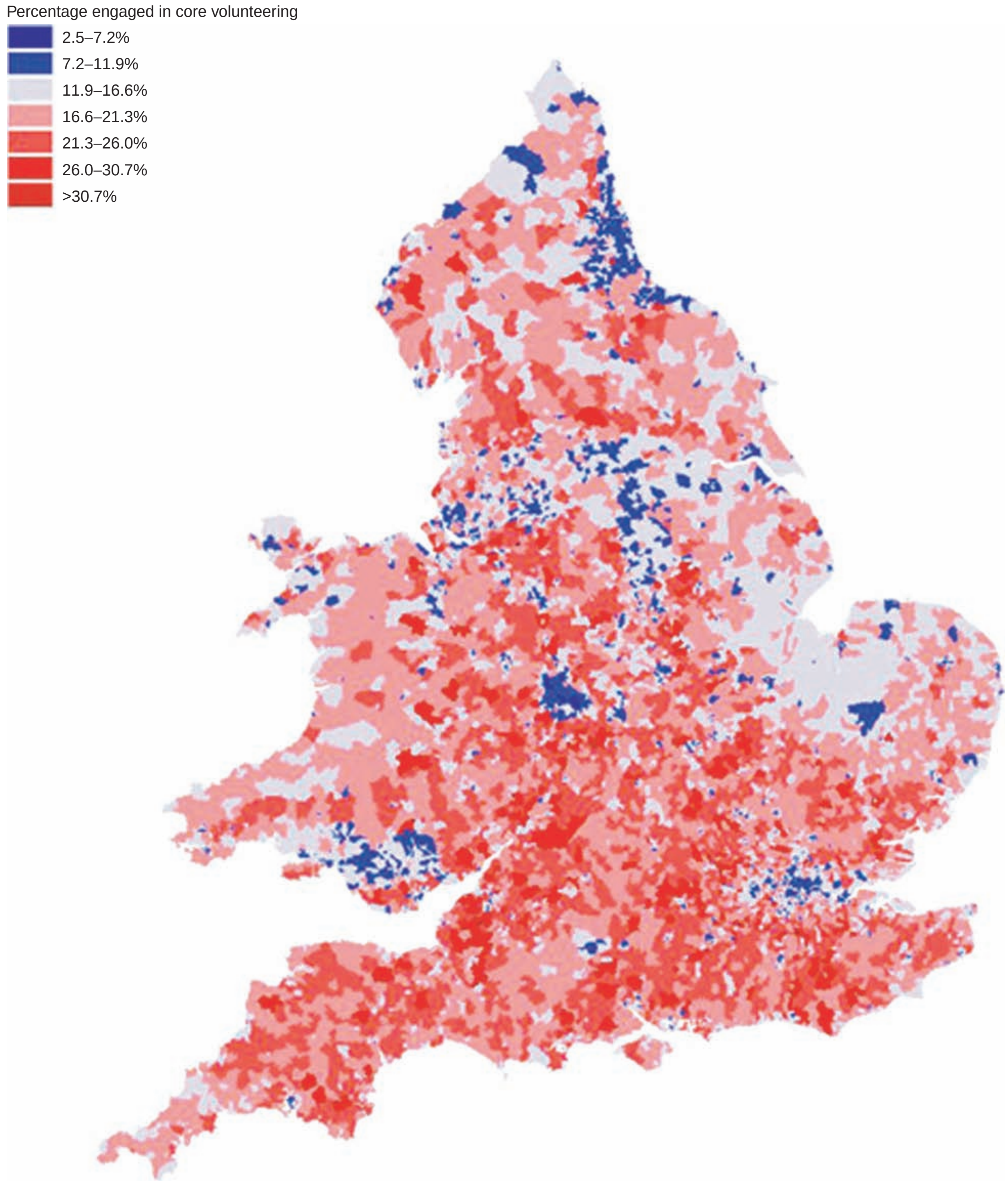

Figure 3 Synthetic estimates of 'core' volunteering. These have been derived via multilevel model equations which summarize the impact of individual and contextual influences on the likelihood of volunteering. Reprinted from Mohan, J., Barnard, S., Jones, K. and Twigg, L. (2004). Social capital, geography and health: Developing and applying small-area indicators of social capital in the explanation of health inequalities. In Morgan, A. & Swann, C. (eds.) *Social capital for health*: *Issues of definition*, *measurement and links to health*, pp. 83 110. London: Health Development Agency.

willingness to work with others to improve the neigh borhood, voter turnout, and perceptions of community spirit. Furthermore, these have since been used in an analysis of a UK health and lifestyle survey of 9003 in dividuals that was initially undertaken in the mid 1980s. Since this preliminary survey sweep, mortality records have been linked to the relevant individual records, thus facilitating a multilevel analysis of the impact of small area measures of social capital upon mortality whilst controlling for other individual and areal covariates. The findings suggest that there may be an ecological effect for social capital. However, social capital and area level de privation are highly collinear and separating out their effects is problematic. While there is some evidence to suggest that health related behaviors and individual material circumstances mediate the relationship between social capital and health, individual material circum stance far outweighs the importance of social capital on health outcome. Such findings suggest that caution should be exercised in drawing conclusions about the effects of area level social capital on health outcomes.

Qualitative Approaches

Many of the extensive, quantitative studies are heavily criticised for being under theorized. Causal processes cannot be fully explored and vague associations are noted as causal without any real justification of plausible mechanisms. Alternative qualitative approaches avoid causal generalizations and focus on identifying aspects of social capital that facilitate health promotion or health enhancing activities. Much qualitative research originates from the health promotion arena and attempts to seek out aspects of community social capital that increase the ability of certain individuals to access health related in formation and health related resources or to engage in more healthy lifestyles. Importantly, these studies are usually context specific and highlight subtle differences that exist in patterns of networks and civic engagement in otherwise similar neighborhoods. The well known ex ample of the Pennsylvanian town of Roseto can be called upon here to illustrate how higher levels of social capital and social cohesion (in comparison to similar neigh boring towns) during the 1950s were used to explain the more favorable rates of coronary heart disease within this community. Again, as with the quantitative approaches, overall findings from the qualitative studies tend to be inconclusive. A number of neighborhood based studies in the UK have suggested a weak relationship between so cial capital and health. However, in contrast, a recent investigation which attempted to determine the extent to which high rates of common mental disorder (CMD) could be linked to deficits in social capital in Gospel Oak in London found no evidence to suggest such a connection. Here, social capital was assessed using the phenomenological experience of residents derived via interviews, focus groups, and participant observation. In general the residents found the area to be rich in social capital and it was argued that compositional character istics were more important in explaining the higher rates of CMD within this area.

Qualitative approaches have also been useful in recognizing the ‘down’ side of social capital, sometimes termed negative social capital. We have already hinted that the ties that bind might be too tight in close knit communities to allow individual to make choices that optimize their future health and well being. The as sumption that all aspects of social capital are good for your health in all types of contextual setting and at every lifestage is obviously problematic and may explain why null or negative findings are sometimes derived.

Social Capital, Public Health, and Policy Implications

Theoretical debates concerning the efficacy of social capital within the public health arena suggest three dif fering perspectives. The first views the health benefits of social capital as arising from mechanisms of broad social support. This perspective acknowledges definitions based on the collective properties of places (e.g., social co hesion, civic participation, and social integration) as well as recognizing individual aspects of social capital (i.e., the ability of individuals to make use of networks to improve social support before, during, and after illness). The re search detailing the benefits of social support has a long history within and outside of the social capital literature, *sensu stricto* and covers a vast array of health outcomes (e.g., mental health, heart disease, overall mortality, and health related behaviors).

Second, as we have already outlined, social capital is said to be a part of the psychosocial pathway linking income and social inequalities to health outcomes. Those societies that are dominated by liberal market economies have resulted in relatively large disparities between the rich and the poor which leads to a breakdown in social cohesion, increased levels of stress and anxiety and a breakdown in social support mechanisms. In this argu ment it is the existence of hierarchies and inequalities in income *per se* that create the problem, irrespective of absolute income levels. To the extent that social capital mitigates the effects of these hierarchies and gradients, it can be beneficial for health.

In contrast to this ‘inequality’ thesis, the third per spective, often termed a ‘neo material’ or ‘political economy’ interpretation, challenges the simple accept ance of social capital as a universal remedy. In this view, not only is the concept of social capital fraught with conceptual problems and difficult to measure, but it is simply a way of redescribing material inequalities. Its introduction therefore is a distraction from the traditional, but still important, links between health and absolute income and material disadvantage.

There are also two very different views regarding the policy implications of social capital. The first sees social capital as a positive and important element in the transition from state based welfare systems toward 'third way' alternatives involving greater community control and participation. Here it is argued that understanding the concept and processes surrounding community social capital can be used positively to empower local populations to tap into local networks, to 'use' local stocks of social capital for mutual benefit. The voluntary sector plays a pivotal role in such policies as the state transfers traditional responsibilities to nongovernmental organizations. The role of the state is to educate its citizens, to promote sharing, trust, and communication, and to use social capital to enhance health by informing public health policy on the best ways to reach 'hard to reach' target groups. In contrast, the alternative stance regards policies that embrace social capital as unproductive since they will ultimately lead to 'community victim blaming'. The focus on social capital as a major health determinant detracts attention from the more severe problems associated with structural inequalities and the direct effects of material disadvantage. Furthermore, reliance on social capital may contribute to a deficit model, in which it is the responsibility of neighborhoods and communities to build up and enhance stocks of social capital. The community can therefore be blamed if it fails to do this and this is politically convenient for governments who are absolved from responsibility for solving social problems.

Conclusions

Social capital remains a popular concept in the health inequalities arena. What were initially somewhat simplistic associative analyses have been complemented by more sophisticated work, based on multilevel techniques, which simultaneously analyze the relative role of individual and area characteristics. There has also been a determined effort to develop a qualitative understanding of the causal processes through which social capital might influence health. Finally, there is an emerging body of work on the utility of social capital as a means of community empowerment to help in the delivery and dissemination of neighborhood based health promotion campaigns. Nevertheless, the precise significance of social capital is still the subject of considerable dispute. Substantial disparities in health experience continue to exist, and these are primarily due to material disadvantages, but social capital may have some influence on remaining variations in health status. As an overall conclusion, it is fair to say that the jury is still out as to whether social capital has an influence on health independent of the effect of material circumstances.

See also: Civil Society; Health and Development; Health Geography; Health Inequalities; Housing, Neighborhoods and Health; Participation; Voluntary Sector.

Further Reading

Bishop, J. A., Formby, J. P. and Smith, W. J. (1989). International comparisons of income inequality: Luxembourg Income Study, Working Paper 26.

Kawachi, I., Kennedy, B. P., Lcohner, K. and Prothrow Stith, D. (1997). Social capital, income inequality and mortality. *American Journal of Public Health* 87, 1491 1498.

Kawachi, I., Kim, D., Coutts, A. and Subramanian, S. V. (2004). Commentary: Reconciling the three accounts of social capital. *International Journal of Epidemiology* 33, 682 690.

Kennedy, B. P., Kawachi, I., Prothrow Stith, D., Lochner, K. and Gupta, V. (1998). Social capital, income inequality, and firearm violent crime. *Social Science and Medicine* 47(1), 7 17.

Lochner, K., Kawachi, I. and Kennedy, B. P. (1999). Social capital: A guide to its measurement. *Health and Place* 5, 259 270.

Mackinko, J. and Starfield, B. (2001). The utility of social capital in research on health determinants. *The Millbank Quarterly* 79, 387 427.

Mohan, J., Barnard, S., Jones, K. and Twigg, L. (2004). Social capital, geography and health: Developing and applying small area indicators of social capital in the explanation of health inequalities. In Morgan, A. & Swann, C. (eds.) *Social capital for health: Issues of definition, measurement and links to health*, pp 83 110. London: Health Development Agency.

Mohan, J. and Mohan, G. (2002). Placing social capital. *Progress in Human Geography* 26(2), 191 210.

Mohan, J., Twigg, L., Barnard, S. and Jones, K. (2005). Social capital, geography and health: A small area analysis for England. *Social Science and Medicine* 60, 1267 1283.

Portes, A. and Landholt, P. (1996). The downside of social capital. *The American Prospect* 26(May June), 18 21.

Putnam, R. (2000). *Bowling alone: The collapse of and revival of American Community*. New York: Simon and Schuster.

Shortt, S. E. D. (2004). Making sense of social capital, health and policy. *Health Policy* 70, 11 22.

Subramanian, S. V., Kim, D. and Kawachi, I. (2002). Social trust and self rated health in US communities: A multilevel analysis. *Journal of Urban Health* 79, 521 534.

Szreter, S. and Woolcock, M. (2004). Health by association? Social capital, social theory and the political economy of public health. *International Journal of Epidemiology* 33, 650 667.

Whitley, R. and Prince, M. (2005). Is there a link between rates of common mental disorder and deficits in social capital in Gospel Oak, London? Results from a qualitative study. *Health and Place* 11, 237 248.

Wilkinson, R. G. (1996). *Unhealthy societies: The afflictions of inequality*. London: Routledge.

Wolfenden Committee (1978). *The future of voluntary organisation*. London: Croom Helm.

Relevant Websites

http://www.socialcapitalgateway.org
Social Capital Gateway Resources for the study of social capital edited by Fabio Sabatini.

http://www1.worldbank.org
Social Capital (Topics in Development: Poverty), World Bank Group, The World Bank.

Social Class

N. Duncan, University of Cambridge, Cambridge, UK

Glossary

Aestheticization Everyday cultural products and practices are appreciated for their aesthetic qualities. Aestheticization is a kind of mystification occurring when such appreciation obscures the social relations (e.g., class relations that have gone into their production and reproduction). For example, there are hidden stories of unjust class relations behind beautiful landscapes.

Class-for-Itself Class interests as determined by objective criteria of which those who share these interests may or may not be aware. Objectivity is of course a matter of degree.

Class-in-Itself The individual and collective consciousness of class relations (such as exploitation) and political action based on recognition of common class interests.

Contradictory Consciousness Related to the Gramsian concept of hegemony. It is seen as less simplistic and totalizing than the Marxian term false consciousness. Most people have mixed or partial understanding of their class interests. However, it is extraordinarily difficult to trace all the connections and relations that make up many individual's complex embeddedness in class processes. This may be most pronounced when an individual occupies a mixed labor-capitalist class location.

Labor Theory of Value The value of a commodity reflects the wages (costs of reproducing laborer power) plus a profit (average yeild on capital). In Marxist theory, capitalism and the exploitation it enables is based on this logic (commodities produce more value than they cost in labor power). Labor time, wages, prices, etc., vary especially between distant regions; thus, this value is not the same as price. Many Marxists and others today believe this theory is not necessary to explain exploitation.

Introduction

Stratification based on structured inequalities in the distribution of wealth, status, and power has existed in most, if not all, societies. Inequalities in life chances and power have been structured in more or less rigid ways largely depending upon the rank, order, caste, estate, or class into which an individual happens to be born. The types of structured inequalities and their bases have varied throughout history and across societies. After reviewing some of the classical Marxist and Weberian definitions of social class and revisions to these conceptualizations, some recent social and cultural geographical approaches to class in modern capitalist societies will be discussed.

Despite obligatory reference to class among what Butler refers to as proliferating etceteras (including gender, race, sexuality, and age), class has received significantly less attention in recent geographical literature than these other bases of social differentiation, identification, and inequality. This is due to widespread acceptance of post structural critiques of essentializing, reductionist, totalizing, and determinist models of social structure.

Among the more favored approaches to social research in recent decades are those that pay attention to ambiguity, fragmentation, and hybridity and which focus on the fluidity, richness, instability, and general messiness of social life. These emphases have resulted in the avoidance of bipolar structural categories such as labor and capital or oppressor and oppressed. However, it would appear that rather than reformulating definitions of class, many geographers simply avoid using class as an analytic category. It is important to note that there are some prominent exceptions among contemporary geographers. Neil Smith and David Harvey are among those who have been most vehement in decrying the fact that class relations are largely left off current research agendas in geography. Sayer and Walker have suggested that despite dramatically restructured class relations and an upsurge of global class organizing during the last three decades, class has been a neglected concept. Smith says that it would be distressingly ironic if class remains an undeveloped area of geographical research given the unprecedented industrial expansion since the 1970s that has resulted in new global class formations.

In response Smith calls for the replacement of what he describes as thuddingly inflexible conceptualizations of class inherited from 1970s and 1980s Marxism. He suggests that more nuanced understandings of class would be compatible with attention to other cross cutting facets of social identity, agency, and structure such as gender, race, ethnicity, nationality, disability, age, and sexuality. For example, racialization of labor segmentation has become especially relevant with increasing global labor migration. Neo liberated global capital has created competition between regions, increasingly bringing labor into competition with labor from distant regions, dividing labor geographically, undermining labor organization, and reinforcing and deepening wage and employment inequalities. These developments call for increased at tention to the experience, consciousness, and, most im portantly, the active agency of workers in the production of capitalist space or what Herod calls labor's spatial fix in response to capital's spatial fix.

Jackson and many others have called for a general re materialization of social and cultural geography which would bring together advances in cultural analyses of the fluidity, richness, and fragmentation of experience, per formance and identity, grounding these in material cul ture, concreteness of nature, and uneven structures of capitalism. Peet suggests that such a project would link economic with cultural analysis supported by an rich array of intellectual resources which conventional eco nomic thinking currently ignores.

When geographers do use the term class, it is usually in a loose, popular sense referring to identity or social status of individuals based on education, occupation, cultural capital, income, and/or social capital. The idea of classes as collective actors and class struggle as a world historical metanarrative have largely disappeared. In fact, it has become increasingly rare to use the concept of class in any way that relates it explicitly to the logic of capitalism. Instead, studies of nonclass bases of identity and experi ence have predominated. Income groups with differing lifestyles and consumption patterns are described, but not usually analyzed in terms of exploitative class relationships or the structuring of inequalities. Furthermore, the post structural crisis of representation has undermined the confidence that academics once displayed in making judgments about misperceptions or unrecognized interests.

Classical Definitions of Social Class: Marx and Weber

The term class came into common usage in Europe during the late eighteenth and early nineteenth centuries in Europe. Most formal definitions today owe their ori gins to nineteenth century social and political thought especially the work of Marx and/or Weber. Although Marx's sociological descriptions refer to an intricate system of class divisions and ranks, his more abstract political economic analyses of capitalism posit an es sentially dichotomous class system based on members' relationship to the means of production. It is this class relation that constitutes capitalist societies as capitalist. However, during the nineteenth century concepts such as race and class were used in loose and interchangeable ways to refer to all sorts of divisions such as ethnic groups, nationalities, criminals, and breeds of animal and Marx's writing are no exception.

Nevertheless, in Marx's formal analysis of capitalism, he predicted that over time members of all the various groupings within a society would eventually coalesce into two divisions, a broad class of capitalists who own the means of production such as factories and land and a class of laborers who control only their own labor power which they can sell to the capitalists who pay them enough in wages to maintain themselves. These two broad classes are internally related (mutually dependent for their existence as classes). Capitalists as a class depend on laborers as a class to sell them enough of their labor power to generate surplus value (value beyond the wages paid) which the capitalists then use to increase their capital and hence the means of future production and the reproduction of their power over the laborers. Marx's notion of surplus value is important in explaining the source of exploitation hidden within the logic of capitalism.

Marx predicted that those capitalists with access to only small amounts of capital would tend to sink into the class of proletarians as they failed to compete with more powerful capitalists. As factory work became more and more technologically advanced, it would have an in creasingly de skilling effect that would further this process of proletarianization. Laborers largely worked under conditions set by their employers and over time these conditions were more and more alienated from them. For example, one of the incentives for capitalists to create factories was to further increase control over labor. Through unionization and governmental legislation the then newly enfranchised workers managed to recover some degree of control over labor conditions. However, as Marx was unable to forsee, during the late twentieth century the vastly increased geographical mobility of capital due to diminishing transport and communication costs in relation to the relative immobility of labor undermined such efforts on the part of labor. Conditions of capitalist exploitation are thus deepening and broad ening with the segmentation of labor markets inter nationally and the spread of neo liberalism globally. The class consciousness that Marx predicted would develop has not done so due to many factors such as the spatial frag mentation of workers internationally and the cross cutting and particularizing of class with other bases of identifi cation and inequality such as race, gender, and age. Un fortunately, Marx's bipolar framework does not easily lend itself to nuanced accounts of these intersections.

There are some more recent attempts to extend Marxist class theory by taking their cue from Weberian and popular discourse on class fragmentation while retaining the basic structural logic of exploitation. Nevertheless, there still remain many unanswered questions as to the relation be tween what Marx referred to as class in itself (objective class interests) and class for itself (collective and individual consciousness of class relations such as exploitation and the mobilization of common class interests).

It should be clear that the above is Marx's bare bones definition of class, one that he and many later Marxist thinkers have elaborated in many ways. What may be most important to retain from this basic perspective, however, is the relationality of the concept of class and the idea that there is a sustained hidden logic to the (albeit fluid) structuring and restructuring of class relations over time.

Weber's concept of class drew on Marx, but in the spirit of critique. Weber agreed with the basic logic of class exploitation in capitalism. However, he rejected the teleology of class struggle which Marx foresaw resulting in a 'dictatorship of the proletariat' and ultimately in a classless society with no privately owned property. It should be added that this teleology is often also recognized in more recent versions of Marxist class theory as either naïve or as having been purposely simplified by Marx for polemical purposes. The teleology is generally recognized to have been thwarted over time. Weber stressed the complexity of stratification based on status, the experiential dimension of capitalist class systems, and differential bases of power. Weber focused much of his analysis on the bureaucratic and highly rationalized nature of capitalism. For Weber the ownership or nonownership of the means of production is but one factor affecting life chances.

Weber's more pluralist view saw the distribution of power in society as based not only on class membership, but also on status based on honor, prestige, and lifestyle as well as membership in organized political parties and movements. The latter two he saw as more closely associated with feelings of attachment and conscious identification. Classes, therefore, are not cohesive communities, but highly variable objective situations based on market determined life chances that may or may not provide a basis for mobilizing communal action. He saw social stratification as having multiple and overlapping dimensions involving complex sets of social relationships. Not all power is economically conditioned; power may be achieved through political organization or complexly related to status, either inherited or acquired. Sometimes power is mobilized and sometimes not, depending upon general cultural conditions and individual life histories and dispositions. Marx saw the ever present underlying source of exploitation and oppression lying hidden within the basic, inescapable logic of capitalism. For him, experience that does not recognize this results from false or contradictory consciousness. In contrast, Weber introduces a complex social psychological dimension to class that is seen as not necessarily, but often reflective of actual class situations.

Class as Stratification

In the 1940s American anthropologist Warner published in five volumes an empirically comprehensive study of class as status. It was a study of class attitudes, occupational prestige, and social mobility based on nearly a decade of detailed ethnographic research in a small town in the eastern US by a team of researchers. His use of class is descriptive rather than explanatory and is typical of much twentieth century academic and nonacademic writing on class. The study divided the town into classes: the upper-upper class was composed of what he referred to as "old money" or people with inherited wealth, the lower upper class composed of "new money" or people who have made large sums of money, the upper middle class including highly paid professionals and corporate executives, the lower middle class being less well paid white collar workers, the upper lower or working class including blue collar and manual workers, and the lower lower class which consisted of the working poor, the homeless and the permanently unemployed.

The dominant sociological paradigm in the US during the mid twentieth century was structural functionalism that posited various similar gradational class schemas based on occupation. These were thought to sort the population into class positions based on capabilities, qualifications, and status attainment. The stratification of occupations was seen as important to the smooth functioning of society. Such stratification models more closely matched popular conceptions of class such as those described by Warner than the abstract relational or conflictual models. Many similar systems of status stratification have been proposed for the US. Some of these are based on functional reasoning and others merely describe and measure social status and occupational prestige based on differences in skills, incomes, knowledge, autonomy, and authority or status (Parsons, Coleman and Rainwater). Such American models tend to highlight harmony and naturalized differences in ability and attainment in contrast to British models which have tended more toward conflict and structured inequality models.

Revisions to Classical Theories: Neo-Marxians and Neo-Weberians

Sociologists such as Goldthorpe and Wright use computer technology to analyze large scale social survey data and to empirically test Marxian and Weberian theses. Goldthorpe's theories attempt to explain class relations rather than simply describe prestige or income hierarchies. His main goal has been to understand social mobility through cross national research.

Wright has developed a nuanced, multidimensional concept of exploitation by studying class across various types of capitalist societies. He is interested in conceptually clarifying structurally ambiguous or contradictory locations within class relations. He shows that the middle class shares in the power of the capitalist class. This includes control over the material means of production, authority over the labor conditions of others, and/or control over money capital or investment. This fragmentation and dispersal of power results in unclear class boundaries. In opposition to dichotomous class models, Wright argues that many people occupy inter stitial positions between the bourgeoisie and proletariat. For example, some people control the conditions of the labor of others, but do not own the means of production. Wright says that managers and supervisors occupy a class location between the bourgeoisie and the proletariat. He has undertaken large scale empirical research which supports his views. For example, he has shown that in contemporary North America, members of the middle class receive significantly higher income as a return on education than do other workers due to their elevated status based on their positions of authority. Small em ployers also occupy the middle ground between the bourgeoisie and the petty bourgeoisie.

Wright makes it clear, however, that classes are not aggregations of occupations, but can be defined only in relation to other classes. He says of the three possible bases for class formation (material interests, lived experience, and capacity for collective action), material interests are primary. This view goes against the grain of much con temporary theorizing which sees speculation about the objective material interests of others as presumptuous (see Laclau and Mouffe). On the other hand, one could argue that given the global reach of capitalist relations and the complexity of intersecting bases of inequality, it is extraordinarily difficult for social researchers to trace the complex causal connections that constitute class relations and produce differing interests. It would be even more difficult for individuals to trace the various multiple causal connections and relations in order to understand how their own interests are structured.

The lived experiences of people who are positioned economically and socially in similar ways can vary con siderably and thus they may not recognize their com monalities or form unified organizations. In fact, there are usually good explanations (including the obvious geo graphical factors such as living in different parts of the world) why they will not do so. Thus, one can see why Wright chose material interests as the most useful basis for determining class, the other two factors being far more tenuous. Wright's empirical findings appear to contradict some Marxist propositions such as the growing proletarianization of societies. He discovered instead an increasing number of supervisors and managers in the many capitalist countries. He argues, however, that one should not restrict ones analysis to national economies. If one looks globally, it becomes evident that proletarian ization is indeed happening, but it is centered in Third World countries. Thus, his work points to the changing geography of capitalism.

Other European sociologists also focus attention on structured inequalities and class conflict. Poulantzas, for example, draws on Althusserian Marxism, but places more emphasis on class. He argues that classes are distributed by the relations of production and are constituted only in class struggle. Miliband analyzes class relations in terms of domination, but leaves little room for working class agency, hence is less sensitive to class conflict. Others less sure than Poulantzas of the inevitability of class con sciousness pose the development of such consciousness as an empirical question. For example, Lockwood proposes that the determinism in Marxist theory be rejected leaving the relationship between class position and class con sciousness as an open empirical question. Lockwood's re search shows that the status differences between 'clean' white collar and 'dirty' factory work produce different class attitudes and mutual antagonisms. As many low status jobs including low paid domestic work have been done by women, there has always been a cross cutting of class and gender which complicates the attribution of class especially within families. Answers to the question of which sorts of low paying white collar jobs might be considered working class or proletarian have varied. Control over conditions of work is often seen as a com ponent; however, attitudes and self ascription are also often thought to play a constitutive role.

Empirical research on class psychology and the lived experience of class relations such as the studies done in the 1970s by Sennett and Cobb (in the US) or Willis (for Britain) or the more recent studies by Argyle on the psychology of class or Charlesworth on the phenomen ology of working class experience are relatively rare and tend to be national in focus. These types of study cannot therefore adequately deal with the growing transnatio nalization of class processes. Such approaches, however, do help to explain the hegemony of capitalism and the lack of consciousness of conflicting class interests.

Geographical Approaches to the Analysis of Class Relations

Geographical analysis of class processes emerged in the field of geography with the rise of Marxist approaches in the early 1970s. Cultural geographers had generally avoided the study of social processes and social geog raphy was a young subfield at the time. In recent decades, geographers have looked at issues such as the spatial division of labor, the expanding global migration of labor, and the geographical restructuring and transnationaliza tion of class formations. They have researched the ways class relations are transformed unevenly across space due to the differential mobility of labor vis à vis capital. However, this literature has tended to theorize class ab stractly and has not yet sufficiently tackled the empirical issues of how individuals are related to class processes.

Self identification, consciousness, and multiple and contradictory locations of individuals within a class system have not been subjected to in depth social–psychological or cultural analysis. The relation between culture and class has been neglected, while interest in cultural identities based on race, ethnicity, and sexual orientation has flour ished. When class is mentioned by cultural geographers, it is usually in an uncritical, gradational sense of status or distinction based on occupation and income or on cultural capital as defined by Bourdieu. Often the term class is used in imprecise, descriptive ways referring to naturalized popular categories such as middle class consumption practices or working class attitudes.

Cultural and social geographers rarely use class in rigorous Marxian or Weberian relational senses of the term for fear of economic determinism or essentialism. On the other hand, some Marxist and other political economic discussions of class tend to avoid explanatory uses of culture for fear of blaming the culture of those who are oppressed, rather than the political–economic structures that facilitate exploitation. Although there is a danger in some contemporary work in social and cultural geography of over emphasizing agency and thus in advertently blaming the oppressed for their own op pression, there are some good reasons to look at class categories as they are constructed in the popular im agination. Sayer, writing on the moral significance of class, for example, argues that class inequalities are based not just on differences in wealth and income, but in ac cess to valued ways of life. People tend to include or exclude, or be included or excluded, on the basis of in tangible attributions of social class, some of which are included in Bourdieu's notions of cultural capital and distinction. Sayer points to the fact that there are in numerable minor, subtle, unintended acts of symbolic domination and resistance to these attributions of class in everyday life. These acts of domination, hegemony, and resistance are thought of by people as class practices and should be studied as such, even if these are only tan gentially related to the logics of capitalism. If class matters to people and effects their interactions with others, even if – or perhaps especially if – they find it difficult to articulate, then it is an important topic for social and cultural geographers to study.

Soja uses the phrase 'space hides the consequences' to help understand how people may not be able to see the relations between their lives and large scale political–economic processes. Cultural geographers such as Cos grove, Daniels, Mitchell, and Duncan and Duncan look at class conflict as embedded in landscapes (again this is confined to the local or national scale). This work depends upon a critical analysis of naturalized class interests. As such it adopts a more distanced or objective approach to class consciousness than many post structural and phe nomenologically oriented cultural geographers favor. Mitchell, and Duncan and Duncan have filled out Har vey's notion of the aestheticization of class processes empirically in their studies of cultural landscapes. They show how through complex interrelationships individuals are complicit in producing unintended negative outcomes for others. As it is not in such peoples' collective interest to trace these linkages, the consequences of their collective actions remain either unnoticed or consciously allowed to remain unremarked.

In the 1980s, Thrift and Williams looked at the question of how individuals are recruited to class political move ments. However, they and others who have adopted similar approaches have found it difficult to relate the conscious ness of individuals within communities to processes at other geographical scales. In the 1990s, Kearns and Withers looked at the practices of class identity especially the cultural and experiential aspects of inequalities, but again such work proves difficult to employ in studies of class transformation at wider geographical scales.

Pred and Watts raise the question of how to rethink class in nonessentialist terms. Massey argues that classes are not empirical groups of men and women and that there is no necessary political unity or consciousness of common interests. Gibson Graham define classes not as social groupings, but as processes and experiences of exploitation (the producing, appropriating, and distrib uting of surplus labor). Their work goes perhaps the farthest in arguing that the hegemony of capitalism is fragile, fragmented, and fluid. In their view the world is less dominantly capitalist than most people assume. They claim that noncapitalist processes are far more prevalent than is usually acknowledged. Class is over determined meaning that it is constituted not just through economic relationships, but through many intersecting social and cultural processes some of which are capitalist and others not. Gibson Grahams' arguments are intended to call attention to the urgency of empirically based social psychological and cultural approaches to the under standing of class processes and to imagine beyond the all encompassing totality of capitalism in order to under mine capitalist exploitation. The question is whether they place too much emphasis on the potential performative power of thinking outside capitalocentric discourse. This question is best answered through empirical research and community activism as they suggest.

Sayer's work on the morality of class is the most sustained analysis by a geographer of the relation be tween class and consciousness. He argues that not only are entrenched inequalities and the hidden injuries of class difficult to explain, the moral dimension of un deserved privilege and suffering based on natal class lo cation is so uncomfortable and embarrassing that it is often simply ignored by academics and nonacademics alike. As subjective definitions of class are only tangen tially related to capitalist economic processes, class re mains a complex, contested, perhaps even a chaotic concept. Individuals collectively engage in class practices;

these practices and performances are both conscious and practical (unarticulated). These practices may or may not be consciously organized; usually they are not. Indi viduals may have multiple, overlapping, or contradictory class based interests which may or may not be under stood as such. Clearly, to understand consciousness, practices, and performance of class more cultural and social geographic research is required.

See also: Capital and Space; Capitalism; Capitalism and Division of Labor; Cultural Capital; Labor Geography; Neoliberalism; Structural Marxism; Transnationalism and Labor Geography; Underclass.

Further Reading

Bradley, H. (1996). *Fractured Identities: Changing Patterns of Inequality*. Cambridge: Polity Press.

Daniels, S. (1989). Marxism, culture, and the duplicity of landscape. In Peet, R. & Thrift, N. (eds.) *New Models in Geography* vol. 2, pp 196 220. London: Unwin Hyman.

Duncan, J. and Duncan, N. (2004). *Landscapes of Privilege: The Politics of the Aesthetic in an American suburb*. London: Routledge.

Duncan, N. and Legg, S. (2004). Social class. In Duncan, J., Johnson, N. & Schein, R. (eds.) *A Companion to Cultural Geography*, pp 265 278. Oxford: Blackwell.

Gibson Graham, J. K. (1996). *The End of Capitalism (as we knew it): A Feminist Critique of Political Economy*. Oxford: Blackwell.

Harvey, D. (2001). *Spaces of Capital: Towards a Critical Geography*. Chicago, IL: University of Chicago Press.

Massey, D. (1984). *Spatial Divisions of Labour*. London: Macmillian.

Mitchell, D. (1996). *The Lie of the Land: Migrant Workers and the California Landscape*. Minneapolis, MN: University of Minnesota Press.

Sadler, D. (2000). Concepts of class in contemporary economic geography. In Sheppard, E. & Barnes, T. (eds.) *A Companion to Economic Geography*, pp 325 340. Oxford: Blackwell.

Sayer, A. (1999). Valuing culture and economy. In Ray, L. & Sayer, A. (eds.) *Culture and the Economy after the Cultural Turn*, pp 53 75. London: Sage.

Sayer, A. (2005). *The Moral Significance of Class*. Cambridge: Cambridge University Press.

Sennett, R. and Cobb, J. (1972). *The Hidden Injuries of Class*. New York: Vintage.

Smith, N. (2000). What happened to class? *Environment and Planning A* 32, 1011 1032.

Willis, P. (1977). *Learning to Labor: How Working Class Kids Get Working Class Jobs*. New York: Columbia University Press.

Wright, E. (ed.) (2005). *Approaches to Class Analysis*. Cambridge: Cambridge University Press.

Relevant Websites

http://www.centerforworkingclassstudies.org
Working class and Labor Geography Bibliography.

Social Geography

R. Panelli, University College London, London, UK

Background

Social geography has different cultural roots and mean ings. Within selected European traditions, it has been understood as an overarching form of geography that integrates numerous other subdisciplines focusing on populations, land uses, and cultural practices so that a composite socio spatial understanding can be derived for social life in a particular place or landscape. This inte grating, 'big picture' notion of social geography contrasts with the more intense and detailed, specialist studies of socio spatial life undertaken in most Anglo American traditions.

Anglo-American Social Geography

In Anglo American settings, social geography has been widely understood as a subdiscipline that grew in rele vance and popularity in the last quarter of the twentieth century. Driven in part by modernizing tendencies and a strong belief in the benefits of scientific approaches for solving human and environmental challenges, social geography emerged from a wider geographic attention to postwar problems. Especially characteristic of this period was a belief that problems could be overcome with the correct application of scientific (including social scien tific) thinking and the adoption of 'universal' truths sur rounding progress, development, and prosperity. Via extensive mapping and modeling exercises, anxieties surrounding slums, ghettos, urban vices, and depressed rural conditions were laid in stark contrast to some of these supposed truths and ideals. As a consequence, considerable focus was given to the study and planning associated with slum clearance, ghetto management, crime prevention, and widespread service provision.

Cumulatively, social geography was recognized in some quarters as an applied and policy relevant sub discipline that would contribute to the monitoring and achievement of postwar prosperity: the 'building of a better world'. But social geography was more potent than this. Drawing on diverse heritage, including nineteenth century anarchic geographies and the emerging energies of the 'civil rights', 'women's liberation', and diverse 'left and socialist' movements, some sectors of social geog raphy also took up more radical agendas and actions. These resulted in a range of critiques about the uneven and unjust nature of societies in space. Some of the verve and clarity of these theses would be profoundly challenged in later years, but the critical energies of the late 1960s and 1970s laid a foundation of questions and commitments that ensured the growing popularity of social geography as a taught subject and field of research endeavor in Anglo American related universities (in cluding those of Australia and New Zealand). But to fully appreciate the diverse, progressive, and critical strands of the subdiscipline, together with the more recent chal lenges and developments, an awareness of the contrasting philosophical and theoretical approaches is first required.

Contrasting Approaches

The breadth of social geography has resulted from the different philosophical and theoretical traditions that have punctuated human geography more generally. These traditions shaped contrasting engagements with social issues and therefore resulted in the production of different types of social geography (summarized in **Table 1**).

Some of the European forms of social geography re flected the attention to description and synthesis achieved in the French regional geography tradition. However, the dominant Anglo American shape of social geography, emerging in the latter part of the twentieth century, initially reflected an interest in positivism and the adoption of scientific method and quantitative ap proaches. The belief in 'geography as a spatial science' involved the adoption of statistical techniques and mathematical modeling in consort with increasingly so phisticated computer aided cartography. This produced a hegemony of explanation and prediction that was ap pealing and potent to many of the dominant political and social interests of the time. Residential patterns, migra tion processes, service needs, and problematic or ideal ized forms of social, economic, and political interaction could all be monitored, analyzed, and commented on. Especially popular was the belief that the manifestation of problems could be located and predictions could be posed, based on the continuation or adjustment of nu merous purportedly causal (or at least related) factors.

While the power of quantitative and neopositivist scientific analysis of the social world would retain an application and relevance well into the contemporary era, the 1970s saw a growth in two alternative approaches, drawing on contrasting philosophical traditions. First, an interest in hermeneutic and phenomenological phil osophies, combined with ethnographic methods, enabled

Table 1 A summary of contrasting traditions and results in social geography

Tradition	*Foci on social issues*	*Types of social geography produced*
Regional	Social life and relations are interconnected with other (physical and human) dimensions of the region under consideration.	• Integrated descriptions of specific areas or regions demonstrating how local society was interwoven as a tapestry of interrelated environmental, demographic, economic + cultural dimensions
Neo-positivist Spatial Science	Social life can be understood as a function of universal processes and laws that link society and space.	• Statistical descriptions of existing patterns of society operating over space (e.g., migration, poverty and affluence) • Mathematical models, assisting scholars in hypothesizing, testing + predicting relations between social groups + space (e.g., ethnic concentrations, housing markets, and social service needs)
Humanist	Social life is seen as individual and contextual (rather than universal) and understood via experiences and meanings that individuals and groups will assemble about their lifeworlds.	• Text-rich accounts of individuals' + groups' experiences and the values + meanings they ascribe to everyday life and the places to which they are attached
Marxist	Social life is recognized as organized around the material conditions and relations that sustain a society. Critique of social conditions and processes sustaining capitalism was initially common.	• Theoretical critiques of how society is organized + regulated to enable capitalism to be reproduced • Case studies exposing unequal, social power-relations + material conditions faced by contrasting classes in capitalism • Geographies of social conflict at both inter-class + intra-class levels
Feminist	Social life is recognized as organized around the gender divisions and relations that shape and reproduce a society. Critique of patriarchy and gender inequalities was initially common.	• Theoretical critiques of the unequal gender relations reproducing patriarchy • Case studies of women's unequal life experiences (e.g., access to housing, employment, and struggles over identity) • Later consideration of hegemonies + diversities of masculinity that help shape women's, and some men's, lives
Postmodern + Post-structural	Society and space are seen as diverse, multiple, dynamic and constantly reproduced/contested via discursive meanings and practices. Deconstruction of core social terms and multiple readings of place and space were initially common.	• Readings of diverse + dynamic constructions of social groups + spaces • Deconstructions and critical accounts of contested social + spatial relations • Exploration of possible radicalized geographic praxis combining past + new approaches to social concerns/action

humanist approaches to challenge the homogenizing and universalist claims of spatial science approaches. They could provide a starkly different way of creating social geography. Specifically, humanist geography paid attention to the complexity of social life and individual experiences of space that defy measurement. They developed conceptually and empirically detailed accounts of everyday people and their experience of place.

Second, a belief in radical critique and action enabled both critical philosophies based on Marxism and feminism to provide alternative perspectives on the social world. Following the social protests and political critiques emerging through the 1960s and 1970s, these approaches offered contrasting forms of practice to those of quantitative and humanist traditions. Theoretical frameworks conceptualizing social reality – and explaining the source and perpetuation of inequalities – were established. Consequently, documentation and critique of notions such as 'capitalism' and 'patriarchy' were detailed along with recommendations on how social relations could be changed.

Cumulatively, humanist, Marxist, and feminist approaches provided theoretical foundations that enlivened social geography with detail and highlighted the dynamism of society–space relations. Geographies of everyday life and critiques of disadvantage and difference consequently flourished through the latter part of the twentieth century. But a further philosophical current was to challenge these newer perspectives. With the rise of postmodern and post structural thought, questions developed concerning the surety and essential or uni versal qualities of concepts such as human life, capital ism, class, patriarchy, and so forth. The social world, and the frameworks we use to understand it, was redefined as a set of cultural constructions. Social life and relations were now understood to be products of particular dis courses and articulations of power; products that would be contingent, relative, multiple, and contested. Con sequently, the effect of postmodern and post structural thought on social geography opened up whole new spectrums of diversity and produced challenging new insights regarding the dynamics of social life.

Some sense of the impact of each of these approaches is outlined in the following three sections. These dis cussions highlight both the contrasting outcomes re sulting from these different approaches, as well as acknowledging that each social geography is not only constructed from within particular broad traditions, but also produced by specific people with their own geo graphic biographies and contexts.

Geographies of Social Patterns and Everyday Life

While other facets of human geography have specialized in studies of human populations (e.g., demography, mi gration) or experiences (e.g., of labor, health or urban/ rural settings), the broader scope of social geography has resulted in scholars emphasizing the multiple and intersecting dimensions of everyday life. Two significant forms of this focus have involved either quantitative, spatial science accounts of social patterns and problems, or the more humanist inspired investigations of locally specific meanings and experiences of life, articulated by individuals and groups.

Mapping Social Patterns and Problems

Quantitative approaches to social geography have a long and valued history focussed on the identification of pat terns and processes that are broad in scale and helpful to formal policy interests and activities. During post World War II interest in modernization and social progress, countries and governments such as those in Australia, the United Kingdom, and the United States of America sought ways to identify existing social patterns and predict and plan for future social needs. Geographers' numeric, cartographic, and modeling skills were in high demand. Studies in the 1960s included projects, such as Richard Morrill's, that responded to the then common imperative of quantification and prediction. Morrill mapped and modeled the dynamics of ethnic minority concentrations in urban space, producing results that were significant for both academic geography and wider political debates. Morrill's location and simulation of the expanding 'Negro Ghetto' in Seattle was a classic approach to a social problem of the day, and it stimulated similar work in other Anglo American contexts through the 1970s and 1980s. It also emphasized the detailed, locational, and predictive knowledge that could be generated while also avoiding the more difficult questions of racial discrimination and resi dential segregation.

Similar types of quantitative geography were gener ated to identify problems and progress in rural social settings. The 'rurality index' first developed by Paul Cloke in the 1970s and later replicated by others in the 1980s and 1990s, enabled a composite measure of living conditions in England and Wales based on a range of indicators of more modern and/or urban life (e.g., household amenities, occupations, and commuting pat terns). These measures were indicative of social hardship or improving quality of life that varied spatially at the time. Importantly, the purported 'objectivity' of index ation techniques and the policy relevant demands of the time underpinned Cloke's choices. It also enabled this work to gain popularity, both within geography and the wider policy and planning environment as England and Wales sought to change and improve living conditions for rural dwellers.

In summary, as shown in **Figure 1**, numeracy, car tography, and modeling were core resources for social geographers. They enabled researchers to assist wider authorities in understanding and managing social prob lems concerning postwar Anglo phone societies and pursuing then assumed, universal goals of progress and development.

The potency and applicability of these types of geo graphic studies have faced critique, but retain a clear currency because of their capacity to shed light on the profiles of changing societies and their needs. Most re cently, studies of increasingly complex migration and ethnicity patterns have been important in a range of Anglo American settings (e.g., the heterogeneity of non white British employment and residential patterns; the social and political consequences of defining black/white and Latino categories in the USA; and the in/visibility of Muslim populations in Australian society).

Tracing Meanings of Everyday Lives and Places

In contrast to the macro level, and numeracy based char acter of quantitative studies, humanist social geographers moved away from what they perceived as the problematic objectivity and homogenizing qualities of such works. In contrast, humanist approaches offered theoretical and methodological bases from which to build geographic knowledge as contextually focused, even unique. Attention was given to the everyday, mundane, life of individual people and groups of people at personal and local levels.

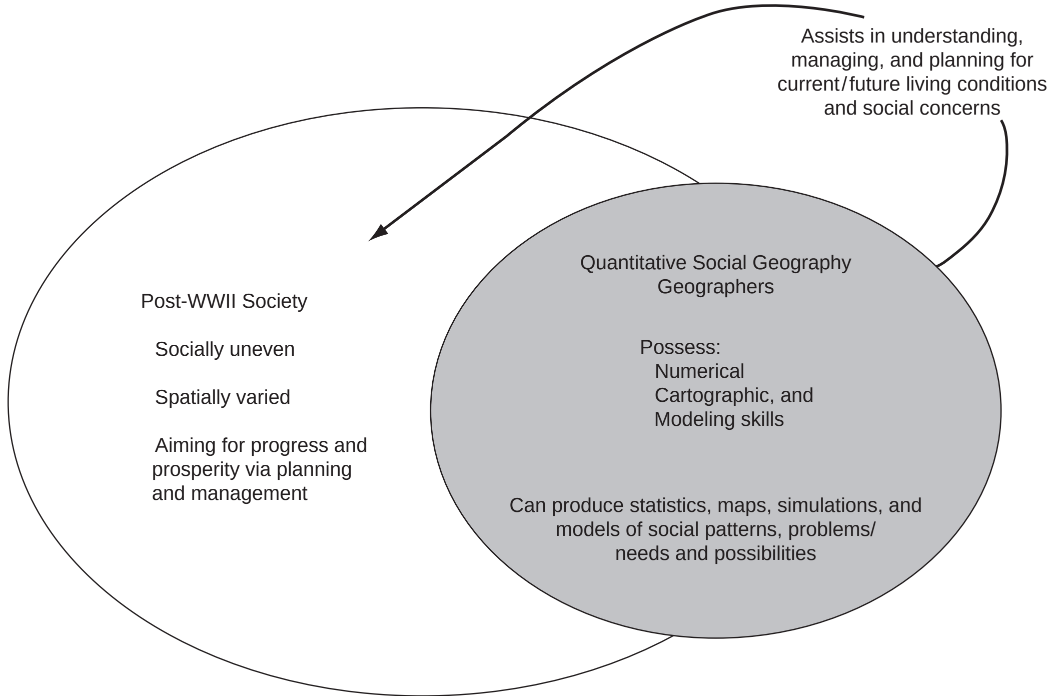

Figure 1 Quantitative social geography's relationship with wider society.

In particular, a focus was made on the 'dynamism of the lifeworld' or the 'world of everyday life', both via theoretical and empirical endeavors. For instance, theoretical considerations of phenomenology in geography were outlined by Anne Buttimer and Edward Relph, working to establish the philosophical credentials of humanist geography as a valid alternative to the 'quantitative revolution' occurring in human geography in the early 1970s. In a complementary fashion, case studies, such as those of Roman Cybriwsky and David Ley, showed the nuanced interpretations that could be generated through a detailed study of the social structure, understandings, and daily life of contrasting groups within an individual neighborhood. Patterns of home making, neighborliness, racial tensions, and intergenerational interaction in a Philadelphia suburb could all be captured through lengthy ethnographic study, and the spatial variation of these social dynamics could be represented via photographs (e.g., of graffiti or street scenes), quotations from local residents, and mapping of different activity sites.

Across humanist social geography, considerable emphasis was given to agency and human–place relations, in recognition of the significance accorded to notions of human essence, place based experience, and meanings. In some cases, like the work of David Seamon, these were registered at the level of individual bodies and 'place ballets'. In other cases, attention was given to cumulative and collective meanings attributed to places that were assembled by people over varying lengths of time, via both daily life in specific residential settings and imaginary literary constructions of place and landscape.

Considering these contrasting quantitative and humanists projects and interests, it is possible to see how both continue to resonate with more contemporary projects in social geography. While preferences in theoretical approach have changed over time, the continued merit and relevance of studying the geographies of social patterns and the nuances of everyday life has retained currency across social geography. Moreover, new applications of, and questions about, quantitative and humanistic geography are informing contemporary dimensions of social geography, as will be shown later in this entry.

Critiquing Disadvantage and Difference

While mapping social patterns and investigating detailed cases of social life were indisputably popular in the second half of the twentieth, the social and political unrest of the 1960s combined with the development of radical social theory produced the basis of a further form of social geography. Broad social and political challenges presented via the 'civil rights' and 'women's liberation movements', together with French students' and wider Socialist protests enabled interested scholars to review the purpose, practice, and potential of many disciplines, including human geography. Across the various subdisciplines of geography, ideals of modernization, progress, development, and objectivity were criticized. Social geography was especially affected by these trends, for the living conditions and social relations that were so central to the emerging subdiscipline (and its public policy interface) were now core foci of the new critical devel opments. In particular, questions of welfare, social justice, inequality, and exploitation would take center stage.

Key writings by David Smith and David Harvey reflect both their own experience of geography and wider social concerns in the 1970s and 1980s. Each in their own (wel fare and Marxist) geographies would challenge other social geographers to consider not only the nature of uneven spatial patterns but also of living conditions and social opportunities. They would not only identify where social problems were located, but they would also ask the 'why' questions. Why are social resources unevenly distributed? Why do some people face poverty while others enjoy substantial affluence? Proposals for a greater commitment to comprehensive welfare for all sectors of society, and for the exposure of social disadvantage, injustice, and socio economic exploitation would grow from this work.

A core effort to emerge at this time involved the es tablishment of Marxist geographies that complemented developments in economic and political geography. In social geography, Marxist approaches enable scholars to explain the uneven access and control of space (e.g., for housing, leisure, and transport) in terms of the unequal dynamics of a capitalist mode of production. The concept of class would form a key concept in social geographies of the late 1970s and 1980s and research would variously expose the unequal class relations occurring in housing and labor markets in many urban, rural, and regional settings. For instance, Michael Dunford, Paul Cloke, and Nigel Thrift would each be influenced by Marxist de bates and give increasing attention to uneven and un equal conditions, and show the social disruption and class restructuring that occurred during different phases of capitalist development in France and England. Similarly, harnessing the critical impulses of this period, Mike Berry would document the housing inequalities that sustained capital accumulation in the USA and Australia. Furthermore, the role of the state would also be impli cated, as geographers such as David Harvey and Ruth Fincher, wishing to extend the benefit of Marxist cri tiques in North American geography, would demonstrate how state apparatus and government practices helped to sustain and reproduce the necessary social contexts in which capitalism could flourish in periods of growth, and reassemble itself in periods of 'crisis' or 'restructuring'.

Beyond problems of housing and employment, other issues that had previously been studied as 'social prob lems', to be mapped and analyzed, were equally re considered in the light of Marxist approaches. In particular, geographies of sexuality would highlight the economies of prostitution (such as those documented in England and the USA by Richard Symanski at varying levels of State, metropolitan, street, and bar activity). The heterosexual reproduction of capitalist industries and space would also be exposed. In the latter instance, the impact of feminist geography (discussed below) would intersect with Marxist interests as scholars like Kim England, Larry Knopp, and Ruth Fincher would illus trate how differing forms of urban space, housing stock, and social services sustained hegemonic inequalities that privileged capitalist class relations, and patriarchal and heterosexual practices.

Together, Marxist and feminist approaches to social geography would elevate the importance of critiquing social differences (such as class, gender, and sexuality) with the aid of radical theories and detailed case studies. That which had previously been mapped as 'social variation' would now be critically named and analyzed as 'social difference'. Implicit and explicit links between difference and unequal power relations would become transparent and widely studied in critical social geographies.

In relation to feminist social geography, two broad developments emerged from the end of the 1970s. First, women were identified as a missing or under represented group within human geography research. While this would influence all aspects of human geography, social geography was a particularly fertile subdiscipline for the growth of new, gender sensitive studies. Commencing with a new focus on women (before later interrogating masculinity in the 1990s), scholars made radical and socialist feminist analyses of the unequal experiences and opportunities women encountered vis à vis men across all forms of daily life, for example, housing, employment, unpaid domestic, care and community work, access to transport, and social services.

Socialist feminist approaches were especially popular in British, North American, and Australian studies of the social and economic inequalities women faced in paid and unpaid work. These investigations would show how disadvan tageous class, gender, and sexual relations entwined to en sure the social and material reproduction of capitalism. They would also show how these relations would vary spatially, where different configurations of gendered space and place specific manifestations of gender would occur (e.g., in English, family farm units, compared to urban, Canadian, working class households).

Radical feminist critique of the heteronormative focus of much of this work was also produced. Studies, such as that by Louise Johnson in the mid 1980s, supported both the growth in lesbian and gay geographies and later postmodern and post structural geographies of gender, sexuality, and the body as a site of contested identities and performances (see below). Other differentiation in feminist social geography resulted from the growing nuances being documented between gender and eth nicity. Studies such as those by Susan Hanson and Ger aldine Pratt, and Melissa Gilbert would begin to expose the complex contrasts in gendered experiences when analysis of ethnicity and racism were included in feminist social critiques of housing, employment, service access, and social networks.

The second development in feminist social geography reflected the wider trends in feminist geography, namely, the recognition of the academic discipline as a gendered social world in its own right. In this respect, numerous women and geography groupings would critique the male domination of departmental and research cultures. They would also establish wide ranging critiques of research practice, reviewing the traditional power relations so often established between researcher and researched, and calling for greater recognition and respect for women's own voices and agendas in research design and reporting. These innovations would also intersect with broader interests in activist and participatory research practices developing more widely in and beyond geography where emancipation, affirmation, or a reversal of past power hierarchies guided social research (e.g., development of youth, disabilities, and postcolonial geographies).

Diverse Deconstructions of Social Life: Postmodern and Post-Structural Geographies

Postmodern and post structural approaches to social geography have focused on the fluid, heterogeneity of everyday life in the era characterized by postmodernity and post industrialism. These approaches document how social life, and previously taken for granted assumptions (e.g., about work, space, identity, and social interaction), have been unraveled. In their place diverse new configurations of postmodern cities and postproductivist or multifunctional countrysides have been noted. No single, ideal, coherent, or unified notions of 'the city' or 'the rural community' are possible. Instead, geographers have concentrated on questioning hegemonic or normalizing assumptions about societies and environments, showing how multiple, shifting, and contestable these meanings and practices are in individual cases.

Geographers in the last decade of the twentieth century mobilized the skepticism and multiplicity characterizing postmodern thought. They recognized that postmodern social geographies might be created to reflect either the configuration of a particular postmodern era or object (e.g., social relations in a postmodern city), or the style and methods that have come to be associated with postmodernity. Specifically, these include a skepticism and critique of 'modern' values and generic 'truths' such as 'scientific objectivity', 'progress', 'development', 'equality', 'the individual', and 'society'. They also included an interest in pastiche, and the mixing and collage of different genres as well as an explicit attention to the multiplicity of meanings and experiences that might be studied and represented in geographic research.

Drawing on cultural and literary theory social geographers have also used post structural strategies to deconstruct the many previously dominant understandings of society and self, production and consumption, belonging and marginalization, social power relations or resistance. They have also reassessed the notions of class, gender, race/ethnicity, and sexuality that have been so central to social geography. By showing these notions are discursively constructed and culturally contingent, post structural social geographies have created critical readings of social norms and showed how everyday life could be read as a more complex negotiation of multiple subject positions. For instance, the work of Katherine Gibson and Julie Graham would move from classic interests in Marxist critiques of capitalism to explore the possibility of post structural and antiessentialist readings of society and economy. They also used postmodern styles, unsettling the previously coherent and unitary notion of 'author' (to create a new collaborative, plural writing persona), and reconstructing processes of theoretical conceptualization and book structure – inviting multiple and shifting readings of their work. In presenting a post structural feminist critique of Marxist political economy, they showed how essentializing definitions of class could be deconstructed and shown to be overdetermined by (or mutually constituted in connection with) a range of other social dimensions: gender, ethnicity, sexuality, and so forth. Similarly, gender was shown to be diversely constructed and experienced by differing groups of women and men.

Gender identity as a key individual and social construction was also recognized by many scholars as providing at least some women and some men with a range of discourses and subject positions by which to make meaning and choose tactics associated with their lives and material or political interests. In addition, race and ethnicity were interrogated for the ways in which hegemonic 'white' interests could construct racialized meanings. Diverse work ranging from readings of historic geography text books, through to contemporary census instruments or official place names, have illustrated how dominant (historic and contemporary) notions of whiteness can control and marginalize minorities in specific places or across whole institutions and systems of social thought. Similarly, nuanced interrogations of dis/ability and youth/age also began to emerge at this time.

Subsequently, diverse sets of identity politics were investigated, illustrating how disadvantaged groups could mobilize strategic meanings and identities from a range of available subjectivities in order to challenge entrenched norms and claim material spaces and symbolic places for these more marginalized interests and activities. Readings of the mobilization of gay pride events in the United States, Australia, and New Zealand, and the construction of strategically particular neighborhoods (based on ethnicity and religious observance) exemplify these types of works.

The complexity of social relations and spatial struggles highlighted in these investigations of identity politics stimulated two developments. First, social geographers came to recognize the need to consider the heterogeneity within analytic categories they had previously tended to hold as unities. For instance, past advocacy for the geographies of commonly disadvantaged groups (e.g., working classes, women, ethnic minorities, and homosexuals) was now accepted as a far more diverse and complex project than previously acknowledged. These issues had begun to appear early in geographies of class and gender, as radical geographers recorded the range of contrasting experiences faced by different groups of women and men, class fractions and intra class conflicts, and groups falling outside any form of income or capitalist/production based classes. Through post structural social geographies, the range of class and gender positions became more visible, and critical readings and deconstructions of these categories were carried out. Similar trends occurred in geographies of ethnicity and sexuality, complemented by postcolonial and queer theories that highlighted the hegemonies of Western and Northern norms that had tended to fix binaries and hierarchies around race and sexuality. In contrast, these new geographies disturbed the homogeneity and presumed stability of categories. Moreover, they highlighted the distinct spatial and sociocultural differences across 'non white' populations and the instability of definitions and shifting political possibilities across a continuum of sexual identities and practices that refuse categorization in a simple hetero:homo sexual binary.

Second, these developments also emphasized the need to build more nuanced critiques of dominant social categories. In this way, the complexity of elite classes, masculinity, whiteness, and heterosexuality has been increasingly exposed. Scholars have shown both the powerful normalizing tendencies associated with some discourses and power relations associated with class, whiteness, heterosexuality, and so forth, together with the consequent spatial and social advantages accorded to such groupings. Alistair Bonnett's and Wendy Shaw's interrogations of privileged forms of whiteness (in contrast to 'color', 'Blackness', and 'indigeneity') in the United Kingdom and Australia exemplify these efforts. These types of work have also exposed the degree to which social life varies for individuals who are positioned at apparently advantageous points along such axes of difference. In this way, work such as Lucy Jarosz and Victoria Lawson's analysis of 'redneck' geographies in rural USA records the simultaneous 'advantage' of white, male positions as well as the obsolete and derogatory character that can be associated with these groupings. Similarly, the variation within heterosexual social relations has been increasingly exposed by scholars such as Jo Little, Phil Hubbard, and Mary Thomas in detailed studies of 'where and how' sexual relations are negotiated by different ages, in different spaces, and associated with different sociocultural expectations.

While social geography has continued to highlight the significance of established categories of social difference, the growth of postmodern and post structural approaches in human geography changed the character of this differentiation. Attention to diversity has meant there has been an increasing interest in differences beyond the established axes of class, gender, race/ethnicity, and sexuality. Consequently, readings of difference associated with issues as diverse as age/youth, dis/ability, mental health, rural poverty and homelessness, and forms of itinerant and gypsy life have increased voluminously. By concentrating on the discursive constructions of these experiences and the spaces where they are commonly relegated, post structural accounts of marginalization and exclusion have been assembled, showing how tactics of ostracism and constructions of abjection could be focused upon these minorities. Deconstruction of ablest discourses, adultist and ageist prejudice, and readings of madness and itinerancy as degenerate, even evil, illustrates just some directions this burgeoning work has taken.

The opportunities stemming from the deconstruction of dominant meanings and the attention given to the diversity of social groups and practices have provided a further direction for social geography. Specifically, those interested in the critical possibilities and the alternatives for social power and action have harnessed some of the analytic and practical potential of post structuralism. Geographers such as Fiona Mackenzie and Paul Routledge have demonstrated how groups can mobilize dominant and reverse discourses to political advantage in Canada and Scotland. Similarly, J.K. Gibson Graham and Jenny Cameron have recorded the diverse social and economic possibilities of participatory and action research that explicitly seeks to unsettle – and work beyond – hegemonic assumptions about capitalism and development.

Rematerializing and Feeling Social Geography

While postmodern and post structural approaches were enormously popular and influential in social geography around the turn of the century, several other perspectives have begun to challenge or complement these developments. In response to the perceived overemphasis on social constructions and representations as a way to analyze social life, some scholars have sought to moderate the 'cultural turn' that expanded with the use of postmodern and post structural perspectives. They have called for a refocusing on to the material conditions of life. Social geography has been invited to 'rematerialize' research foci and the geographic knowledge being produced.

Results of this 'material re turn' involve a whole range of diverse materially attentive studies that highlight social and political implications of quite mundane aspects of life. This includes studies of arrangements of bodies; the use and reuse of clothing; diverse processes of home making; and negotiation of workplaces and public sites of formal or informal significance. For instance, drawing on recent feminist traditions in New Zealand geography, Victoria Guyatt has demonstrated how gendered social relations are embodied and contested through specific struggles over clothing and behavior in contrasting sites within the work spaces of a local hotel. In contrast, engaging the more varied debates surrounding migration, ethnicity, and post/colonialism in British settings, Divya Tolia Kelly has developed detailed accounts of how specific and contradictory identities surrounding Englishness and Asian ness are articulated via artifacts and the organization of diverse matter within particular homespaces.

Benefits of such work are reported to lie in both: (1) the re engagement with the conditions characterizing daily life (including the continuing prevalence of social–spatial inequalities); and (2) the production of socially relevant and policy informing research. Studies showing the strikingly uneven socioeconomic indicators across national populations or the deprived living conditions of immigrant populations in large cities provide one example of this type of rematerialized geography. Other forms of study interrogate the way social identities are materially assembled and experienced through the everyday engagement with bodies, homes, work places, or the encounters with specific education, leisure, and health choices. These types of geography enable an understanding of how lives are socially negotiated through varying material worlds. But in some cases they also provide resources for informing the diversity of policy environments and processes, stretching across government and NGO interests in immigration, employment, health, education, and social development. Rematerialized social geography has the potential, therefore, to re engage connections between social geography and policy spheres that partly lapsed as post structural perspectives appeared less accessible and relevant for many policy stake holders.

A second way that scholars have sought to qualify discursive and representational approaches to social geography centers on studies of affective experience and relations. At times this overlaps with rematerialized geographic endeavors, but primarily the focus of this second development rests on the contention by Kay Anderson and Susan Smith that emotions constitute an underrecognized set of relations shaping the social world. Appreciation of this position has some of its roots in humanist geography interests concerning people's 'sense of place'. But the focus on emotional relations especially developed from the work of a range of feminist scholars who have produced detailed and critical reflections on their experiences (including emotions) 'in the field'. Acknowledging that geographic research involves a plethora of uneven, socially navigated, and place based, emotions explicitly situated geographers themselves within the varied terrain of emotions. Extending from these reflexive considerations, emotional geographies have begun to be explored across diverse topics in social geography. Commencing with the immediately personal, some scholars have shown the multiple emotions and uneven consequences that surround experiences of fear, memory, and/or affection as they are played out in contrasting family, leisure or public sites. Gillian Rose's work on emotions infusing experiences and practices surrounding family, motherhood, memory, and photographs illustrates this genre at household levels in Britain. Others have taken interests and concerns, focusing on individuals' and groups' navigation of emotional experiences and 'emotion work', encountered through health issues, music, dance and sexuality. Still others (such as Paul Cloke, Harvey Perkins, Owain Jones, and Jo Little) have drawn on the twin understandings of discursive and biophysical dimensions in rural geography, to argue that emotions ranging from fear to adventure and desire are implicated in social relations with specific British and Antipodean places, landscapes, and constructions of nature.

In future, the local and everyday experiences of emotion are likely to further intersect with wider critical and political geography interests. This is especially possible in emerging critiques of the geopolitics of fear as articulated in globalized, national, and local forms. Past social geographies of crime and fear will engage with established attention to gender, ethnicity, and age to expose the varied emotional topography of contrasting populations as they seek to make meaning of current and future life chances and world events.

Further Social Geographies: Chasms and Connections

At the time of writing this entry, social geography appears ever more crucial to understanding social problems, and imagining, or agitating for social change. While it is evident that advances in opportunities (if not equality) have been achieved in certain places for some groups, it is also true that in the past three decades, Western or Northern societies have become economically more differentiated and socially and politically more polarized, in part replicating the severity of divisions found in other parts of the world. Differences based on ethnicity and identity politics form arguably deeper chasms than ever as geographers have shown recently in critiques of protests against asylum centers in England and mosque developments in Australia.

Critique of Anglo-American, Anglophone, and White Geographies

Equally, ongoing inequalities within geography have become increasingly obvious following recent documentation of white, Anglo American, and Western hegemonies in geography. European scholars, such as Judit Timar (highlighting an East–West experience), have noted the complexity of these dominant ways of producing and circulating geographic knowledge. In addition, from the particular historical and political circumstances of geography in Greece, Dina Vaiou has called radical geographers to debate Anglophone hegemony and disturb existing patterns of definition, control, and regulation. Similarly, in considering women of color in geography, Minelle Mahtani has pointed to the need to critique the social profile and practices of geography itself. Mahtani and others highlight continuing privilege of whiteness within the academy and the need for changing practices and support for both faculty and students, and their research. In terms of the textual reproduction of the discipline, some positive responses have been made, including the initiation of country reports in the *Journal of Social and Cultural Geography* and a range of special issues in diverse journals addressing these concerns.

Indigenous Perspectives

There has also been an increasing concentration on indigenous geographies, and the formation of special working groups to address indigenous issues in some geographical associations (e.g., Association of American Geographers, Institute of Australian Geographers). This work promotes strikingly diverse geographies. Acknowledging the specifically different geographic worlds of social groups decimated via invasion and colonization within current Australian, Canadian, New Zealand, and the USA nation states, scholars are highlighting how indigenous perspectives can challenge and invigorate social (and other) geographies. For instance, using different, and more integrating perspectives, insights into previous human–environment or society–nature binaries have been recorded as Lisa Palmer, Nicole Gombay, and Kathleen Wilson have shown the intimate interrelations connecting people with their environments, history, traditional cultural practices, and contemporary shifting relations, in Australian and Canadian settings. These emerging indigenous sensitivities also have significance for the way geographic studies are conducted. For example, in 'maori' or 'tauiwi', dialog surrounding the practice of geographic research in New Zealand, Robert McClean, Lawrence Berg, and Mike Roche have argued that critical reflection of 'boundary crossings' and 'aroha' are required in cross cultural investigations. Similarly, Paul Hodge and John Lester, working from an indigenous research center, record ways geographers can de colonize their own field through reflexive cross cultural innovations.

Ethics and Morals

More broadly, the current tensions, anxieties, and discrimination evident in many Western or Northern settings have prompted social geographers to become more reflective on matters of ethics and morals. In part, this has involved a renaissance of moral concerns previously articulated through welfare geographies and questions of social justice. David Smith, long time advocate for socially just and morally reflective geographies, has written extensively on the value of moral concerns. He sees these not as fundamentalist, or repressive codes, but as frames of reflection to deter excesses of relativism. In a complementary fashion, other social geographers – working from quite contrasting traditions and inspirations – have increased the attention given to morally self conscious research design and action, as well as a more ethically reflective and morally nuanced approach to the teaching of geography. Cumulatively, these diverse scholars present the possibilities of imagining fairer, more inclusive, and careful societies in future.

Posthuman and More-Than-Human Social Geographies?

Finally, and in part stimulated by disparate environmental concerns, and debates surrounding ethical, embodied, and material experiences of life, social geography is being influenced by debates about hybridity and posthumanism. Bruce Braun identifies posthumanism as a questioning of "the figure of the human, its fixing and bounding" and an emphasis on the connections between 'human' and other forms and relations. Moving beyond anthropocentrism, Sarah Whatmore's *Hybrid Geographies* highlights the possibilities of ethically interrogating 'human' life as complexly connected with other, nonhuman forms and processes. This resonates as a new Anglophone 'discovery' – in concert with long held indigenous knowledge systems. The implications for social geography, for considering more than human beings and relations, are only just beginning to be considered as some scholars (in and beyond definitions of social geography) commence re reading our material, embodied, and environmentally related, everyday lives via more intimate entanglements. These include: our experience of walking outdoors; human–plant relations and ethics in garden settings; indigenous, organic, and slow relations with food; or our passage through illnesses and experi ence of death, including the interface between 'human' and other forms of matter, meaning, and memorial.

Social geography will continue to be a hugely diverse subdiscipline of human geography. Its practitioners will craft knowledge of social life, problems, and possibilities from widely contrasting traditions and personally pos itioned contexts. Social geographies will always be re cords written by multiply identified scholars constructed from diverse institutional, philosophical, and social con texts. But social geography also now exists as one of numerous human geographies that can face – and benefit from – the challenge of moving beyond historic divisions between human and physical geography. Considering indigenous and posthumanist perspectives in new, more than human interconnections between people and their surrounding worlds will invigorate new material, emo tional, and political accounts of social life; both at the mundane and personal level as well as the wider geo political levels where division and distance most threa tens the possibility of social connection and respect.

See also: Ageism and Age; Body, The; Capitalism; Colonialism I; Difference/Politics of Difference; Emotional Geographies; Ethnicity; Feminism/Feminist Geography; Heteronormativity; Homelessness; Humanism/Humanistic Geography; Hybridity; Identity Politics; Indigeneity; Inequality; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Masculinities; Participatory Action Research; Patriarchy; Positivism/Positivist Geography; Posthumanism/Posthumanistic Geographies; Postmodernism/Postmodern Geography; Poststructuralism/Poststructuralist Geographies; Race; Radical Geography; Sense of Place; Sexuality; Welfare Geography; Whiteness; Youth/Youth Cultures.

Further Reading

Anderson, K. and Smith, S. (2001). Editorial: Emotional geographies. *Transactions of the Institute of British Geographers* 26, 7 10.

Bonnett, A. (1997). Geography, race and whiteness: Invisible traditions and current challenges. *Area* 29(3), 193 199.

Castree, N., Nash, C., Badminton N., *et al*. (2004) Mapping posthumanism: An exchange *Environment and Planning* A 36 2004 1341 1363

Cybriwsky, R. A. (1978). Social aspects of neighbourhood change. *Annals of the Association of American Geographers* 68, 17 33.

Dunn, K. (2004). Islam in Sydney: Contesting the discourse of absence. *Australian Geographer* 35, 333 353.

Fincher, R. (1981). Analysis of the local level capitalist state. *Antipode* 13, 25 31.

Gibson Graham, J. K. (1996). *The End of Capitalism (as We Knew It): A Feminist Critique of Political Economy*. Cambridge, MA: Blackwell.

Gombay, N. (2005). Shifting identities in a shifting world: Food, place, community and the politics of scale in an Inuit settlement. *Environment and Planning D: Society and Space* 23, 415 433.

Guyatt, V. (2005). Gender performances in a service orientated workplace in Aotearoa/New Zealand. *New Zealand Geographer* 61, 203 212.

Harvey, D. (1976). The Marxist theory of the state. *Antipode* 8, 80 89.

Hubbard, P. (2000). Desire/disgust: Mapping the moral contours of heterosexuality. *Progress in Human Geography* 24(2), 191 217.

Hubbard, P. (2005). Accommodating otherness: Anti asylum centre protest and the maintenance of white privilege. *Transactions of the Institute of British Geographers NS* 30, 52 65.

Jarosz, L. and Lawson, V. (2002). 'Sophisticated people versus rednecks': Economic restructuring and class difference in American's west. *Antipode* 34(1), 8 27.

Jones, O. (2003). "Endlessly revisited and forever gone": On memory and emotional imaginations in doing children's geographies. An 'addendum' to "To go back up the side hill": Memories, imaginations and reveries of childhood' by Chris Philo. *Children's Geographies* 1, 25 36.

Knopp, L. (1992). Sexuality and the spatial dynamics of capitalism. *Environment and Planning D: Society and Space* 10, 651 669.

Little, J. (2003). 'Riding the rural love train': Heterosexuality and the rural community. *Sociologia Ruralis* 43(4), 401.

Mahtani, M. (2006). Challenging the Ivory Tower: Proposing anti racist geographies within the academy. *Gender, Place and Culture* 13, 21 25.

McClean, R., Berg, L. D. and Roche, M. M. (1997). Responsible geographies: Co creating knowledge in Aotearoa. *New Zealand Geographer* 53, 9 15.

Morrill, R. L. (1965). The Negro ghetto: Problems and alternatives. *The Geographical Review* 55, 339 361.

Pain, R. (2001). *Introducing Social Geographies*. London: Arnold.

Pain, R. (2006). Policy research in social geography: Seven deadly myths. *Progress in Human Geography* 30, 50 59.

Panelli, R. (2004). *Social Geographies: From Difference to Action*. London: Sage.

Pratt, G. and Hanson, S. (1988). Spatial dimensions of the gender division of labour in a local labour market. *Urban Geography* 64, 299 321.

Shaw, W. (2000). Ways of whiteness: Harlemising Sydney's aboriginal Redfern. *Australian Geographical Studies* 38, 291 305.

Smith, D. M. (1994). *Geography and Social Justice*. Oxford: Blackwell.

Thomas, M. (2004). Pleasure and propriety: Teen girls and the practice of straight space. *Environment and Planning D: Society and Space* 22, 773 789.

Timar, J. (2004). More than 'Anglo American', it is 'Western': Hegemony in geography from a Hungarian perspective. *Geoforum* 35, 533 538.

Valentine, G. (1993). (Hetero) sexing space: Lesbian perceptions and experiences of everyday spaces. *Environment and Planning D: Society and Space* 11, 395 413.

Valentine, G. (2001). *Social Geographies: Space and Society*. Harlow: Prentice Hall.

Whatmore, S. (2002). *Hybrid Geographies: Natures Cultures Spaces*. London: Sage.

Women and Geography Study Group (1984). *Geography and Gender*. London: Hutchinson with Women and Geography Study Group of the IBG.

Social Justice, Urban

K. Newman, Rutgers University, New Brunswick, NJ, USA

Definition

Urban social justice is a term that suggests that inequalities are socially produced rather than bound in universal truths. In contrast with theories of social justice that are based on a theoretical construct, the term urban social justice suggests that social justice is embedded in social processes that are directly related to the mode of production as it is expressed spatially. Cities are spaces of oppression and inequality but they are also spaces of political liberation. For Marxist political economists urban social justice is created and produced through a dialectical relationship between the oppression and inequality created by the factors of production in cities and by the spaces of freedom also located in cities.

Much like traditional thinking about justice and social justice, urban social justice is about equitable processes and outcomes and it assumes a role for the state in their production. Questions focus on what is being distributed, how it is distributed, and with what outcome. Will just processes produce just outcomes or are inequalities better addressed through redistribution? Can the mode of production remain the same and produce just outcomes? At what scale is social justice produced? Is economic inequality the only source of oppression and injustice? If it is not, what alternative processes are necessary to produce social justice? Is it possible and/or desirable to create a universal idea of justice in multicultural societies? If justice is socially produced and is different for different groups, do intra and inter group cleavages become problems?

Evolution and Main Concepts

Social justice theories consider how social processes produce and reproduce space and how the resulting spaces affect social justice. The term urban social justice diverged from earlier conceptions of social justice in that it has a decidedly spatial component. Early conceptions of urban social justice recognized that cities play critical functions within capitalist economies as centers for production and exchange. Cities are both the places in which goods are produced and the place to invest surplus capital. In the United States, efforts to increase capital accumulation produced metropolitan uneven development as capital expanded outwards, from the urban core to the suburban periphery and beyond, leaving economically devastated inner cities in its wake. Those who remained in the inner city were left to lament: "Search for justice, and what do you find? Just us on the unemployment line, just us sweating from dawn to dusk. There's *no* justice, there's – *Huh!* – just us" (Berman, 1997: 164).

Even as capital fled outward, agglomerations of fixed capital sought to draw capital back to the urban core. These fixed capital actors partnered with the state to create what Harvey Molotch called a growth machine to further the potential for capital accumulation by enhancing the competitive position of the city relative to other cities. Cities seek to retain existing and attract new mobile capital and labor by minimizing the costs to capital, such as providing tax breaks, regulating land use, providing cultural amenities and infrastructure, reducing the cost of labor, reducing regulations, and increasing the appeal of residential neighborhoods. This entire process enhances opportunities for capital accumulation, produces uneven development, and creates a tension within cities between the city as a place for capital accumulation and the city as a lived place for individuals and communities. As Merrifield has observed, the entire city is used to maximize exchange value. Globalization has accelerated these processes and pushed them out into the world creating uneven development at a global scale.

Right to the City

Cities may be places in which capital seeks to further its interests, but cities are also places in which people live, have families, and build community. Lefebvre argues that those who use the city should have a right to the city. He does not prescribe how inhabitants use that right in practice but it certainly happens outside of the existing formal mechanisms for political participation. Lefebvre points out that existing electoral structures are too narrow to influence the space of the city which is produced and reproduced through social processes. A right to the use of the city challenges the liberal right of property ownership giving those who use the city a right in its production along with those who value it for exchange. Social justice, then, according to Lefebvre is produced as a result of the processes of inhabitants living their daily lives.

Don Mitchell builds on Lefebvre's right to the city and considers how public spaces are used, how that use is governed, what rules shape the city and who creates them, and what that means for who has rights in the city. Campaigns to 'clean up' public spaces have removed the homeless which Mitchell argues has removed their right to the city because it eliminates 'the right to inhabit.' The right to the city means participating in the discussions about producing and reproducing the city. If the home less are removed physically, if they do not have the right to inhabit, they do not have the right to the city. Mitchell argues that a right to property should not be a require ment to live in the city and therefore participate in urban governance. Mitchell sees protest and legal action as ways to engage in the discussion about the future of the city that can produce a more equal society.

A right to the city seems to suggest that using that right would take place at the local scale. But what is local and does local participation create a right to the city? Purcell argues that while Lefebvre likely did not anticipate that all action would be local, many who have used the idea of the right to the city have translated it into local action with the expectation that it would produce democratic processes and outcomes. Purcell questions the assumption that local action is inherently democratic suggesting that it is en tirely possible that local action is undemocratic. If the right to the city is translated into a process by which local inhabitants control decision making, Purcell questions whether this process necessarily produces socially just outcomes. He asks how the needs of local interests are interrogated against the needs of broader publics. A pro cess which favors local interests might produce democratic and socially just outcomes, or it might not, but the process, Purcell argues, is not democratic because it favors decision making dominated by the few over the interests of the broader public. The right to the city in Lefebvre's initial conceptualization was about making the city for those who use it instead of solely for those who seek to use it to increase exchange value. This could happen at many dif ferent scales. In the case that Purcell reviews, community groups and residents in a section of Seattle that is slated for rezoning to encourage the development of bio technology firms could participate in the processes locally but he argues that participation at that scale could actually produce undemocratic outcomes. The most likely way to make decisions that would increase use value would have to take place at a broader scale such as that of the city.

Processes and Outcomes

Social justice theories vary on preferred processes for achieving socially just outcomes. Lefebvre and Mitchell focus on the production and reproduction of space. It is by producing and reproducing space that residents can claim the city as their own. By claiming their right to the city they can transform the city as a place merely for exchange value to one in which residents can maximize use value. But participating in the production and re production of the city is only one strategy for creating urban social justice. Other theories of social justice focus on distribution of the social surplus. The challenge is to develop an equitable system for equitable redistribution. Embedded in that challenge are two additional challenges. What sort of system is necessary and what does one use as a metric for distribution. Who has need and how is that determined?

Marxists challenge whether it is possible to arrive at a socially just end without changing the production system. They ask whether distribution in an inequitable system can ever produce socially just outcomes. Few Marxist political economists argue for whole scale economic change but many argue for some modification in the mode of pro duction to ensure equitable outcomes. Post structuralist theories follow a similar logic; only they question whether the sources of oppression are solely related to the pro duction system. These theorists, like feminists, point out that even if economic injustice was eliminated, many groups would still be oppressed. They argue that linking social justice to a universal conception of justice, even to one on the left, will not produce social justice in multi cultural societies as groups experience injustice differently. Post structural theorists offer what is sometimes referred to as a cultural politics in which difference and individual and group identity are valued as strengths. Instead of merging groups to create one homogenous population with one theoretical universal idea of social justice, retaining dif ference and identity, in their view, strengthens democracy.

Iris Marion Young, for example, takes Marxist con ceptions of urban social justice as a starting point. She defines social justice within the context of social processes but she does not limit the source of oppression to the economic system. In much the way that Marxists might argue that social justice cannot be achieved without transforming the mode of production, Young argues that social justice remedies that are limited to distribution will not produce social justice if there are additional sources of oppression embedded in social processes. No amount of distribution will produce equitable outcomes if oppression and domination are embedded structurally because the inequalities will be reproduced through distribution. Moreover, distributing things like wealth and income does not address alternative sources of oppression.

To better understand alternative sources of injustice, Young identifies five faces of oppression: exploitation, marginalization, powerlessness, cultural imperialism, and violence. The first three are related to the way in which people participate in the economic system and their ability for self determination and self development. Ex ploitation can be seen in unequal labor relationships, such as women's unpaid labor within the household which benefits men or how certain groups, such as ethnic and racial minorities, work in jobs that are less likely to allow for self determination and self development. Marginalized people such as those who receive welfare benefits are left out of the labor system entirely. In her view, because marginalized people are unable to participate in the labor force themselves, they are also unable to practice full citizenship rights in the economic system. Powerlessness is the last face of oppression that is related to the economic system. Those who are powerless have jobs in which they have little leeway and must work for others such as pro fessionals compared with menial laborers. Cultural im perialism imposes dominant group ideas and stereotypes and violence is the remaining face of oppression.

Acknowledging that society is not homogenous, Young views social justice as something that is produced through a process of deliberative democracy in which groups and individuals maintain their identity and strengthen themselves through discussion and sharing with the goal of facilitating self development and self determination. Young argues that deliberative process deepens democracy and increases accountability by providing a space in which groups articulate alternative viewpoints and make justice claims. This alters group dynamics by making each group accountable to the others. Young argues that this process has the potential to produce social justice under an ideal set of conditions that include "inclusion, equality, rea sonableness, and publicity" (Young, 2000: 23).

Not everyone is convinced that cultural politics will produce socially just outcomes. Many critics argue that re taining individual and group identities and using that to socially produce justice will lead to relativism and a weak ening of the concept of social justice. Or they fear that competing group claims will make achieving social justice an impossible feat. Fainstein doubts that post structuralism offers a way forward suggesting instead that it leads to new forms of conflict and group oppression. In her critique she suggests that these oppressed groups can themselves oppress other groups and can even oppress subgroups within their own group, all the while under the cover of post structur alism. Fainstein points out the challenge, if not impossibility, of creating processes that produce equitable outcomes.

Further critique suggests that strengthening group identity may weaken community identification which is closely tied to social justice. Low and Gleeson argue that by focusing on individual and group claims social justice claims become fragmented. They offer an alternative approach that involves thinking about communities as things that overlap and interrelate. Social justice, they argued, is created through the idea of community. Instead of narrowing the idea of community to individual groups, the idea of community can be broadened beyond neighborhood, ethnic, race, gender, sexual orientation, and the like and even the nation state expanding com munity and claims for social justice. The goal is to extend the boundaries of community to build connections among people. Low and Gleeson make these ideas real by considering global relationships. For example, people in one country such as the United States may desire access to many inexpensive goods and services. How these goods are produced, who produces them, and how that production affects their environment may not come into consideration. But if the idea of community is expanded outward so that the consumers in the United States are part of the same community as the producers in another country, then the lives and communities of both become closely intertwined. Justice can be produced as the ties of community draw these communities together. Viewing community as something broader than individuals, neighborhood, or nation state may shape desires and political action. Low and Gleeson suggest the limitations of geographical difference and favor an appeal to human rights and a global discourse through international in stitutions to pursue universal rights and needs.

Challenging Neoliberalism

As capitalism evolves these contradictions are increasing and they are doing so on a global stage. Harvey observes how capitalism, as it has evolved in the early part of the twenty first century, is producing dramatic uneven devel opment making some places very wealthy and leaving some exceptionally poor. Uneven development has "in tensified" as a result of what Harvey calls "free market utopianism" which produces "spiraling income inequalities, uneven geographical developments, externality effects on the environment, and the like" (Harvey, 2000: 178). The neoliberal regime argues for free market solutions but it does this even as it is shaped by state policy. The state has not faded away in the era of neoliberalism; it has not been hollowed out. Far from it! The state has been reshaped to further the advances of capitalism. When it hits the ground with little regulation it produces dramatic inequities.

The idea of challenging capital's structural dominance seems daunting but Castells argues that even though structure is important, it does not eliminate opportunities for activism. Perhaps counter intuitively, capital's global expansion presents new opportunities for activism. Har vey questions whether geographic difference is at odds with a global discussion of human rights. He counters the traditional Marxist claim that rights are the product of the bourgeoisie and instead suggests that globalization offers a new scale for organizing. Workers are no longer limited by state borders giving them more power to build bigger broader organizations. In fact, the globalization of capital has made the contradictions ever more clear as uneven development is spread globally with far greater extremes between the haves and the have nots. While Harvey suggests the potential for organizing at the global scale, efforts to reshape the global can occur at other scales. Globalism could be altered by local action. Wil liam Sites suggests that global capital has to actually take place somewhere and how states respond to global actors locally shapes globalization. If states pursue neoliberal agendas then a neoliberal agenda gets reproduced at the global scale. But, if states shape global actors locally, then they can also shape globalization. If the state continues to act in response to neoliberalism through a hands off policy of facilitating capital's expansion, then it merely reproduces that at the global.

But perhaps it is within the city that social justice is produced and reproduced through struggle. Andrew Merrifield questions whether social justice must be created from a universal idea or ideal type or whether social justice is an activist concept. His question emerges from challenges not to the traditional liberal thinkers but rather to those on the left. He questions whether Marxist approaches that require a change in the mode of production to achieve social justice are necessary if not unnecessarily utopian. He wonders instead what happens if one tries to make these ideas – of urban social justice – real? Instead he suggested adopting a 'practical dystopian politics' which he views as an approach that links theory with reality in contrast to the Marxist 'utopian alternative'. This more realistic approach involves a dialectical struggle between urbanization and urbanism. Merrifield defines urbanization as the process and urbanism is the place as we experience it. Merrifield does not see this as something that will be ultimately resolved. Instead it is a process – a dialectic. Social justice, then, is produced through these conflicts, through the effort to make social justice real in the city. Eliminating inequality through a change in the production system is unlikely. It is only through group conflict in the city that social justice is created. These battles in turn create and recreate the city so the struggles take place in space but they are really over the space – what it is, who it is for, and what it will be. These struggles are witnessed in any city on a daily basis. In New York City, for example, the battles between labor and employers takes on a larger than life character as union workers regularly protest developers who hire low wage nonunion laborers by erecting enormous inflatable rats in front of job sites. In response, developers erected their own inflatable cat. What started out as a narrow effort by some unions has expanded from construction workers to many different types of laborers with some unions even renting their rats out for use by other groups.

The battle to create the city, to negotiate exchange versus use value, and to gain the right to the city is evident on the streets. Monthly bike rallies challenge whether the streets are solely to increase the flow of goods and labor and the accumulation of capital or is their other value in their use? In many cities once a month these streets become a place of public dialog and discussion as bike riders take to the streets in an effort known as critical mass. This mass bike ride started in San Francisco in 1992 and later spread to cities across the country. As bike riders take over the streets of San Francisco and New York and other cities across the country, the streets become contested ground, the contested ground that helps to continually reshape the city. The streets are places of liberation as people seek to influence how public space is used and challenge who the city is for. But the same streets are also spaces of oppression in New York City as the police department videotapes critical mass rides to collect evidence for future use in court. The battle for public space continues as the discussion over the use of public space winds its way through the streets and the courts. These processes, these struggles for space, are what Merrifield calls a dialectical urbanism. He sees the city as both a space for emancipation and a space for oppression. The city is the place in which injustices occur and also in which justice is sought.

See also: Neoliberalism, Urban; Uneven Development; Urban Growth Machine.

Further Reading

Castells, M. (1985). *The City and the Grassroots*. Berkeley, CA: University of California Press.
Fraser, N. (2003). *Redistribution or Recognition: A Political Philosophical Exchange*. New York: Verso.
Harvey, D. (1973). *Social Justice and the City*. Baltimore, MD: The Johns Hopkins University Press.
Harvey, D. (1989). *The Urban Experience*. Baltimore, MD: The Johns Hopkins University Press.
Harvey, D. (1997). *Justice, Nature and the Geography of Difference*. New York: Blackwell.
Harvey, D. (2000). *Spaces of Hope*. Berkeley, CA: University of California Press.
Lefebvre, H. (2003). *The Urban Revolution*. Minneapolis, MN: University of Minnesota Press.
Leitner, H., Peck, J. and Sheppard, E. (eds.) (2007). *Contesting Neoliberalism: Urban Frontiers*. New York: The Guilford Press.
Logan, J. and Molotch, H. (1987). *Urban Fortunes: The Political Economy of Place*. Berkeley, CA: University of California Press.
Low, N. and Gleeson, B. (1998). *Justice, Society and Nature: An Exploration of Political Ecology*. New York: Routledge.
Merrifield, A. (2002). *Dialectical Urbanism: Social Struggles in the Capitalist City*. New York: Monthly Review Press. 2002.
Merrifield, A. and Swyngedouw, E. (1997). *The Urbanization of Injustice*. New York: New York University Press.
Mitchell, D. (2003). *The Right to the City*. New York: The Guilford Press.
Peck, J. and Tickell, A. (2002). The urbanization of neoliberalism: Theoretical debates neolibearlizing space. *Antipode* 34(3), 380 404.
Purcell, M. (2006). Urban democracy and the local trap. *Urban Studies* 43(11), 1921 1941.
Sites, W. (2003). *Remaking New York*. Minneapolis, MN: University of Minnesota Press.
Smith, N. (1990). *Uneven Development*. Oxford: Blackwell.
Smith, N. (1996). *The New Urban Frontier*. New York: Routledge.
Young, I. (1990). *Justice and the Politics of Difference*. Princeton, NJ: Princeton University Press.
Young, I. (2000). *Inclusion and Democracy*. New York: Oxford.

Relevant Websites

http://chicagocriticalmass.org/
Chicago Critical Mass.

http://gothamist.com/2005/10/27/cat vs rat at radio city music hall.php
Gothamist: Cat vs. Rat at Radio City Music Hall. October 27, 2005.

http://www.righttothecity.org/
Right to the City Alliance.

Social Movements

M. Ramutsindela, University of Cape Town, Rondebosch, South Africa

Glossary

Bioregionalism Ideas and practices that relate to connecting peoples' cultures to the environment in which they live at a small scale.

Environmental Movement A group of people who are mobilized around the health of the planet and the impact of human behavior on it.

Green Parties Political parties that are formed to promote particular environmental policies and actions through an electoral process.

Social Mobilization The process by which specific concerns become a rallying point for mass action.

Social Movement An organized and sustained collective group of people, who seek to challenge power holders and institutions entrenching inequalities of all sorts.

Social Movement Theory A set of ideas and perspectives that are used in the process of explaining social movements.

Transnational Public Sphere An environment that enables people to organize themselves across national boundaries.

Universalist Propositions Statements that are applied everywhere with no regard for geographical differences.

Introduction

The study of social movements within geography and the relevance of geography to their study have recently attracted the attention of geographers. With regard to their place in geography, there is a view that the discipline has no well developed tradition of studying social movements. Moreover, geographers have been highly selective in engaging with the literature on social groups. This undermines geography's contribution to the broad interdisciplinary debates on social movements. Equally, research in disciplines such as sociology and political science has neglected the relevance of geography in the development of social movement theory. The importation of geography into the study of social groups and vice versa takes place by inserting the importance of space and scale in the analysis of these groups. There is a realization that political mobilization, which is at the heart of social movements, has a spatial context that cannot be ignored if we are to fully understand them. Geographers have also argued that there are geographical differences and variations in the development of social groups and identities; the distribution of resources that are mobilized by social groups; the scales at which social movements operate; and the circumstances and influences of particular places. Against this backdrop, the interdisciplinary nature of social groups and insights from the analyses of scale and place in geography provide opportunities for deepening our understanding of human action within the broad field of human geography. The question that could be asked is how best can human geography harness those opportunities? This short article uses political geography as one of the platforms on which human geography can contribute to, and benefit from, social movement research. This demands understanding social groups as a political force, the strategies they adopt, and the influence they have on the political landscape.

Social Movements and Research Trends

Social movements bring groups of people who collectively seek to change systems that govern or affect their lives. They are concerned about the social and political power and the presumed legitimacy of existing institutions and practices; the gaps between freedoms and entitlements, and the ways in which benefits are extracted and shared; imposition of lifestyle choices; oppression and emancipation, and so forth. In this way, social movements are different from other formations/groupings in that they represent a conscious and sustained effort to change society and have a longer lifespan than, say, a single protest. They are composed by ordinary people coming from diverse backgrounds who aspire to change society for the betterment of everybody's life. The pertinent questions that have occupied most social movement research have been why and how such movements are formed, and what strategies do they use with what effect? Answers to these questions have generated various intellectual responses, which set the following general theoretical foci:

1. Before the 1960s, scholars studied social movement through the lens of an angry mob, meaning that they were frightened by the mob, which they regarded as dangerous. Theories of the origin of social movements focused either on the character of participants or on conditions in the broader environment which the movement faces, the intention being to find out how best to prevent social movements. Theorists

collaborated with law and order enforcement agencies in identifying social forces and to 'know the enemy', as was the case with Project Camelot. The Project was funded by the US Department of the Army and operated as a joint venture between US and Latin American scholars with the aim of identifying social forces committed to change in Latin America and to develop counterinsurgency. Generally, social movements were seen as a problem and as outsider actors from 'normal' politics. It should be noted that scholars were generally less sympathetic to participants in social groups at the time, because they regarded those participants as irrational and as people who blindly and stupidly followed their leaders.

2. The 1960s saw a dramatic change in attitudes toward social movements. This was particularly occasioned by the emergence of the Civil Rights Movement, which not only attracted people of different social strata but also made the demand for freedoms and improvement in the well being of those at the bottom of society, a justifiable course. In other words, what changed the perception of social movements as irrational groups was in part the change in the rank and file members of the movements, which included students, the middle class, professionals, and so forth. Indeed, social movements in the 1960s and 1970s were regarded as 'new' social groups partly because they were seen as an outcome of a rational choice. The feminist movement and liberation struggles added another layer to the rationality and acceptability of social movements in the 1960s and 1970s.
3. In the 1980s, emphasis was placed on the cultural side of social groups, particularly the processes by which organizers frame their issues in order to recruit members and the methods by which a social group develops its collective identity. Unsurprisingly, a strong body of work focused on identities and nationalisms in different socioeconomic and political settings.

The consequence of the shift in the perceptions of social movements was that research not only sought to offer various explanations for the rise of these movements, but also sought, and still seeks, to understand the strategies and resources at their disposal. It has been suggested that in the developed capitalist societies, social movements emerged from the industrial working class struggle and the political ideology, which justified the contrast between conservative right and reformist left. However, this does not answer the question of the geographical spread of these movements and the differences between them. To be precise, why do the same conditions (e.g., socioeconomic) not lead to similar patterns of social mobilization? Notably, social movements result from the ways in which difference and inequalities in society are framed as a political significance in order to mobilize social action. The process of mobilization in turn depends on the capabilities of social groups, the resources available, and the environment in which that process takes place. Social movements have also emerged as a result of globalization, which prompted people to mobilize to protect their lands, identities, and their autonomy. It should be noted that the attempts to explain the emergence and character of social groups are central to theories of social movements. For example, the theory of movement origin focuses either on the characteristics of participants or on conditions that the movement faces. The resource mobilization school emphasizes variation between social movements as a result of the availability of resources. It suggests that the more resources are available to social movements the more such movements are formed. For its part, the political process approach pays attention to the opportunities that changes in the state give to social mobilization. These theories have led to some conclusions regarding the interconnections between the strategies of social groups and those of the state:

1. The encounters between social forces and the state are underscored by the dynamics on both sides of the equation. Social movements therefore continuously monitor their tactics both to evaluate their effectiveness and to respond to the state apparatus. Equally, protest and police tactics adapt to each other through a process of reciprocal influence, involving innovation and adaptation.
2. Collective action is channelled around policy specific opportunities or group specific environment. That is to say social movements are opposed to one or more elements of the existing social and political order.
3. More repressive, diffuse, and hard techniques of policing tend to discourage the mass and peaceful protest on one hand, while fuelling the more radical fringe on the other hand. In other words, the nature of the state's response to social movement plays an important role in shaping the character of social movements. For example, the banning of liberation movements and the use of the army in crashing down opposition to the state created conditions for armed struggles in most of the former colonies around the world.
4. Apart from overt and repressive measures, the state can also respond to protest by social movements through a deliberate inaction. This is meant to create conditions for counterviolence, especially where there are major divisions between sectors of society. For example, the apartheid state fuelled and sponsored violence between ethnic and racial groups in South Africa in the 1980s in order to undermine the liberation struggle.
5. The growth of international agreements among governments has been accompanied by a proliferation of transnational civil society associations. For example, the creation and practices of the World Trade Organization provided a stimulus for the rise of fair trade movements while the international recognition of

indigenous collective rights opened up opportunities for the engagement of the state by indigenous social movements.

6. Transnational activist networks cast the goals of movements in universal terms with local contexts. The creation of a transnational public sphere to advance shared agendas expands the strategic choices available to social movements. This is possible, because a transnational public space not only allows movements to draw their resources from a wide range of sources, but also helps them to articulate local struggles and to entwine them with global issues of all sorts. The environmental movement best illustrates the creation of a transnational public sphere and the main strategies that movements can adopt in their engagement with the state.

Lessons from the Environmental Movement

Unlike other social movements (e.g., feminism and black consciousness) that are associated with political identities, the environmental movement is mobilized around a shared sense of humanity and offers universalist propositions. Notwithstanding the caution that universalist propositions have legitimized the subjection of other peoples and regions, the environmental movement garners support from various sectors of society across the world, because it confronts challenges to the protection and management of the environment and adopts a multipronged approach and a multilayered discourse. For the purpose of our discussion in this article, attention is paid to the environmental movement's challenge to, and intimate relationships with, the state.

The environmental movement challenges the state on several grounds. These include the ways in which the state participates in the destruction of the environment; its incapacity to respond to environmental challenges; and the desirability of the state as an appropriate social organization. These challenges have shaped the environmental movement's engagement with the state, and also resulted in the mushrooming of different environmental interest groups. These are those that cooperate with the state, those opposed to the state, and those that are unconcerned with the state. While there are no fixed boundaries between these categories, each of these categories on their own deepens our insights into the state–social movement interactions.

Cooperating Environmental Movements

The world's largest environmental nongovernmental movements, the World Wide Fund for Nature (WWF), the World Conservation Union or International Union for the Conservation of Nature and Natural Resources (IUCN), The Nature Conservancy (TNC), and Conservation International (CI) all consider conserving biodiversity as their global mandate, which can successfully be achieved in cooperation with the state. They cooperate with the state so as to avoid confrontation, to influence state policies, and more importantly, to access bureaucratic and informal opportunities offered by government institutions and agencies. In order to achieve these goals, they consider the state an important partner. For example, as of January 2007, the IUCN had 1020 members consisting of national nongovernmental organizations (NGOs) (717), government agencies (109), state members (82), international NGOs (82), and affiliate members (30). These organizations adopt conventional lobbying strategies and use donor funding, scientific information, and international platforms to influence government policies and to push for international agreements. The cooperation between these NGOs and the state has also been elevated to the global level in order to establish an international environmental governance structure. Effectively, that cooperation has cemented the relationship between environmental movements and the state, thereby blurring the boundaries between formal and informal politics and allowing government officials to become environmentalists and vice versa. By way of example, the former South African Minister of Environmental Affairs and Tourism, Valli Moosa, became the President of the IUCN after successfully hosting the World Parks Congress in Durban (South Africa) in 2003.

Environmental Movements in Opposition to the State

There are environmental groups that challenge the state by directly engaging the state as a political opposition. They participate in the electoral system as is the case with most Green parties. They do this in order to gain formal political channels that would provide opportunities for combining civil protests and legal–administrative strategies. Their common point of entry into the electoral system is an ideological critique of the state's preoccupation with materialist concerns and the exclusion or marginalization of environmental issues in state policies. Unsurprisingly, the majority of Green parties originated in industrialized countries, where industrialization and nuclear energy and weapons stimulated the mobilization of social groups. Practically, the participation of Green parties in elections through the 'green issue' ticket has narrowed the political space for most of these parties. This is so, because most states have incorporated environmental issues in their policies as required by international laws and protocols, and politicians are increasingly using 'green issues' in their political campaigns. The point here is that, by participating in national and regional elections, Green parties such as the German Green (national) and the European Federation of Green Parties (regional) have collapsed the boundaries between formal and informal politics at both national and regional levels. Notably, not all oppositional environmental movements regard participating in political elections as a strategy for engaging the state. Instead, they could form a pressure group to influence state environmental policies and actions, as is the case with the environmental justice movement. The principal goal of the environmental justice movement is to challenge environmental policies and programs that undermine the rights and values of inhabitants.

Environmental Movements Independent of the State

Though bioregionalism shares the same concerns with other environmental movements, it does not take the state as its point of departure in as far as the future of the planet and humans is concerned. Instead, it propagates a complete reorganization of the world into bioregions in order to integrate human cultures into their life support systems. It dismisses the modern state on the grounds that the state imposes monoculture on people of diverse backgrounds; it has failed to address the crisis of ecological peril and the political systems that have caused ecological disintegration; and that it is at odds with natural laws, which govern biotic communities without coercion, organized force or recognized authority. Its radical stance against the capitalist economy, industrialization, exploitations of humans and nonhumans, cultural imperialism, and so on, makes it attractive to other social movements which are not necessarily concerned with environmental issues.

The groups of environmental movements discussed above illustrate the artificial separation between social movements and the state as representative of informal and formal politics, respectively. The complex relationships between the state and social movements, together with the broad literature on social groups, offer prospects for expanding and enriching the content and scope of political geography, as is shown below.

Social Movements as a Subject Matter of Political Geography

The fact that the most significant changes in society in the twentieth century were brought about by social movements makes these movements a 'natural content' of political geography. Indeed, political geographers have made a contribution to the study of imperialism and decolonization, nationalism and identity formation, all of which emphasize the need to understand the spatial contexts and the role of power relations as the bases for social mobilization. There has been a sustained interest in the study of feminism as a social movement and on the contribution that feminist perspectives have made to political geography and to social science as a whole. For their part, electoral studies have highlighted the interactions between citizens and the state. The studies contributed to our understanding of the impact of circumstances and factors on the outcomes of elections. These include electoral systems, the nature and financing of campaigns, party support, voting behavior, and the demarcation of voting zones. Notwithstanding these important contributions, the study of social movements forms a negligible part of political geography. With few exceptions, most political geography textbooks rarely refer to the concept of social movements in their entries. This is more than an omission, but reflects the weak conceptualization of social movements as part of 'normal politics' and a subject matter of political geography.

It would be fair to suggest that political geography has not yet maximized its contribution to the study of social movements despite its traditional focus on the state and politics, because of the following reasons.

First, political geography has been preoccupied with formal politics (e.g., elections, nation state, identity, borders, citizenship, etc.) for much of its history. In this way, the artificial dichotomy between formal and informal politics led to the neglect of social movements, which are generally seen as informal and noninstitutional. As the brief review of social movements above has shown, the dichotomy between formal and informal is less useful if we are to understand political behavior. The intellectual space that social movements offer to political geography is that they blur the distinctions between the two (i.e., formal and informal politics). This happens through the ability of social movements to push issues from the informal arena into the domain of the formal political agenda, thereby involving ordinary people to participate directly in the political process. Social movements are often the first to articulate new political issues and ideas. The space for inserting such issues and agendas by the state itself is severely limited, because bureaucracies in modern societies are preoccupied with economic and political control in order to ensure stability while preventing the unexpected.

Second, a combination of the dichotomy referred to above and the state centric view that dominated much of the research in political geography imposed limitations on our understanding of the impact of the state beyond the narrow confines of formal politics. Social movements provide the space for appreciating conditions that the state creates or imposes upon the masses and how the masses respond to those conditions and why. The nation state is by and large a mobilizing context for social movements. The state's legislative and judicial system, electoral rules, alliances and conflicts among elites, and capacities for state action, provide conditions for the rise of social movements. There are two dominant views on how the character of the state impacts on the emergence and nature of social movements, the first being that democratic states are conducive for social movements mainly because they are tolerant. The second view is that repressive regimes lead to the development of violent social movements than regimes that are tolerant.

Third, though political geography has contributed to our understanding of the role and function of the state in a capitalist world economy , there is still a need to further grapple with questions of the extent and nature of the impact that the state's participation in the capitalist world economy has on social movements. It has been suggested that movements making ethnic demands are largely the product of the failure to build a national state in the per iphery of the capitalist system. The relationship of the state to the capitalist economy could shape collective action against the state, not least because the process of capital accumulation (e.g., privatization) and its effects (e.g., widening inequality) create conditions for protests. For example, the dismantling of rural development programs through neoliberal reforms in the 1980s and 1990s provided the incentive for indigenous social movements to protest.

Fourth, there has been an enormous outpouring of work on theories of the state that emphasize the lineal evolution of the modern state. In that linear evolution, the intricate relationships between political parties, states, and social movements is given less attention. Political parties and social movements are mutually dependent in shaping politics as the example of the en vironmental movement above has shown. This is so be cause there is a fuzzy and permeable boundary between institutionalized and noninstitutionalized politics. Social movements do not only shape elections, legislatures, and courts, but also give rise to political parties. Participants in social movements can adopt a dual membership by being activists in social movements while, at the same time, presenting themselves as political candidates. In reality, social movements can also become political par ties and political parties do appeal to particular interest groups. Labour parties are one such common example in that they naturally draw their strengths from labour unions. It is commonly held that the ascendance of social movements into the levers of the state has generally weakened the transformative agenda of most move ments. In other words the consequences of blurring the boundaries between social movements and formal insti tutions of power not only change the power dynamics between the two, but also result in compromising desir able social changes. In Africa most liberation movements that became post independence governments have failed to pursue the agendas of the liberation struggles for all sorts of reasons.

Conclusion

The emergence, character, and activities of social groups are shaped by a combination of factors, which in turn differentiates one group from another. Notwithstanding their differences, social movements are intimately con nected to the activities of the state and therefore blur the boundaries between formal and informal politics. This blurring of boundaries provides fertile ground for in corporating social groups into the analysis of the political in political geography. The fact that social movements have contributed to changing the world' political map means that they are a political force that cannot be ig nored by political geographers.

See also: Activism; Colonialism I; Environmental Studies and Human Geography; Nationalism; Political Boundaries; Sovereignty.

Further Reading

Aberley, D. (1999). Interpreting bioregionalism: A story from many voices. In McGinnis, M. V. (ed.) *Bioregionalism*, pp 13 42. London: Routledge.
Della Porta, D. (1988). Recruitment process in clandestine political organisations: Italian left wing terrorism. In Klandermans, B., Kries, H. & Tarrow, S. (eds.) *International social movement research* vol. 1, pp 155 169. London: Jai.
Dryzek, J. S., Downes, D., Hunold, C., Schlosberg, D. and Hernes, H. (2003). *Green states and social movements: Environmentalism in the United States, United Kingdom, Germany and Norway*. Oxford: Oxford University Press.
Goldstone, J. A. (ed.) (2003). Bridging institutionalised and noninstitutionalised politics. In *States, parties and social movements*, pp 1 24. Cambridge: Cambridge University Press.
Goodwin, J. and Jasper, J. M. (2003). Introduction. In Goodwin, J. & Jasper, J. M. (eds.) *The social movements reader: Cases and concepts*, pp 3 7. Oxford: Blackwell.
Laurie, N., Andolina, R. and Radcliffe, S. (2005). Ethnodevelopment: Social movements, creating experts and professionalising indigenous knowledge in Ecuador. *Antipode* 37(3), 470 496.
Miller, B. (2000). *Geography and social movements*. Minneapolis: University of Minnesota.
Nash, J. (2005). *Social movements: An anthropological reader*. Oxford: Blackwell.
Painter, J. (1995). *Politics, geography and 'political geography': A critical perspective*. London: Painter.
Ramutsindela, M. (2007). *Transfrontier conservation in Africa: At the confluence of capital, politics and nature*. Nosworthy Way: CABI.
Staeheli, L. A., Kofman, E. and Peake, L. J. (eds.) (2004). *Mapping women, making politics: Feminist perspectives on political geography*. London: Routledge.
Tarrow, S. (1996). States and opportunities: The political structuring of social movements. In McAdam, D., McCarthy, J. D. & Zald, M. N. (eds.) *Comparative perspectives on social movements*. Cambridge: Cambridge University Press.
Wignaraja, P. (ed.) (1993). *New social movements in the South: Empowering the people*, pp 76 100. London: Zed Books.

Social Studies of Scientific Knowledge

B. Greenhough, Queen Mary, University of London, London, UK

Glossary

Actor-Network Theory An approach to Science Studies developed by Michel Callon, Bruno Latour, and John Law, which asserts that scientific knowledge is co-constructed by human and nonhuman actors.

Enlightenment A philosophical movement, usually dated to late seventeenth and eighteenth centuries of European history, which believed that the real nature of the world could be revealed through the application of careful and rigorous scientific methods.

Posthumanism A philosophical position that challenges the notion of a singular universal human subject by drawing attention to the ways in which human identities are formed through relationships with human and nonhuman others.

Reflexivity A methodological approach which involves considering and/or reflecting on the researcher's own social background and standpoint.

Relativism A view that all knowledge and experience is understood differently dependent on the standpoint or position from which one encounters it.

Science and Technology Studies (STS) An approach to Science Studies which, unlike SSK, stresses the role of nonhuman agencies (animals, machines, institutions) as well as human societies in shaping scientific knowledge.

Social Constructivism A hermeneutic approach to social science which asserts that all forms of knowledge are shaped by the social and material cultures within which they are produced and consumed.

Social Studies of Science or Science Studies An interdisciplinary research area that draws on social science methods, and in particular social constructivism, to explore the production and consumption of scientific knowledge.

Sociology of Scientific Knowledge (SSK) An approach to Science Studies which stresses that scientific knowledge is shaped by the social conditions of its production and therefore can never produce absolute or universal truths about the world.

Strong Program An approach to SSK, advanced by the Edinburgh and Bath schools of Science Studies, which argued that all kinds of scientific claims – both successful and unsuccessful – were equally susceptible to social influences (often contrasted with the so-called weak program of Thomas Kuhn who used social factors only to explain the failure or decline of scientific theories).

Introduction

Social Studies of Scientific Knowledge, or Science Studies as they are more commonly known, are concerned with what counts as scientific knowledge and how it comes to count as scientific knowledge. Traditionally, what distinguishes scientific knowledge from other forms of claims about the world (such as those made by religions or cultures) is the belief that scientific methods offer a definitive view of the world as it really is, rather than a subjective interpretation of it. Science is supposed to be about discovery. Scientists uncover 'facts' about the world through a process of dispassionate and objective observation. As a result, the sciences appear to offer a disembodied 'view from nowhere', producing knowledge that is universally true or applicable. Conversely, Science Studies scholars argue that scientific knowledges are 'views from somewhere', the result of particular studies conducted at particular times and in particular places, and that as a result they are socially constructed or shaped by the social environment within which they are produced. In order to support this claim, Science Studies research draws attention to the local conditions and material resources necessary in order to generate a particular scientific observation or discovery. However, the approach taken to this research varies widely, ranging from historical and archival studies exploring the social, political, and intellectual environments that surrounded particular scientific discoveries, to in depth anthropological investigations of science in action in the field or the laboratory.

While Science Studies scholars share a conviction that scientific knowledge is socially constructed, there are considerable differences of opinion regarding the extent to which science is socially constructed and as to how we should study science and its knowledge making practices. The origins of critical Science Studies are usually traced back to 1962, when Thomas Kuhn published his *Structure of Scientific Revolutions*. Kuhn argued that the failure of some forms of scientific theory was the result not of scientific error, but of changes within the culture of scientific institutions (which he called paradigm shifts). This work formed the basis for the development of a critical Sociology of Scientific Knowledge (SSK) by David Bloor and his colleagues at the Edinburgh School of Science Studies, who contrasted Kuhn's focus on scientific failures with their own 'strong program' of research which argued that all kinds of scientific claims – both successful and unsuccessful – were equally susceptible to social influ ences and were the product of particular locations, times, and places. Scholars of the Bath School of Science Studies added a further empirical dimension to this program, producing in depth analyses of scientific con troversies to highlight the uncertainty and doubt that surrounds many scientific discoveries. Unsurprisingly, SSK met with considerable opposition from within the scientific community. Social constructivism, by ques tioning whether or not scientific techniques could access the 'truth' about the world, represented a direct threat to scientific authority. This intellectual conflict, which continued throughout the 1990s, became known as the science wars.

In 1979, sociologists Steven Woolgar and Bruno Latour published a book called *Laboratory Life*, which was to in spire a new approach to Science Studies known as Science and Technology Studies (STS). *Laboratory Life* was an ac count of an ethnographic study of a neuroendocrinology laboratory in La Jolla, California. In this account, Latour and Woolgar seek to demonstrate how scientific facts are socially constructed by mapping out the geographically specific combinations of laboratory environment, person nel, materials, machines, scientific peer review procedures, and publication processes necessary to produce a scientific fact. Latour and Woolgar's ethnographic observations of science in practice mark a turning point in Science Studies. They seem to find a middle ground between the two extremes of absolute scientific objectivity and those versions of social constructivism that argued scientific knowledge is entirely the product of social processes. Furthermore, they move from a focus on the texts and arguments produced by science in the form of books, papers, and publications, to focus on the materials of sci ence through which that knowledge is produced and verified, including machines, instruments, and even the subjects and objects of scientific analysis themselves. What Latour and Woolgar describe is a kind of sociomaterial constructivism. They still assert that scientific knowledge is produced rather than discovered, but highlight the role of nonhuman (technologies, materials, organisms) as well as human factors in that process of production. One of the key outcomes of STS has been the actor network theory (ANT). Developed by Latour, along with Michel Callon and John Law, ANT is an approach that looks at the different actors (scientists, samples, machines, animals and institutions, among others) involved in the production of a particular scientific discovery or fact, and considers the diverse ways in which they are connected to or relate to each other (networks). ANT provides a means of investi gating the relationships and interconnections between human and nonhuman factors in an analysis, and therefore it holds considerable appeal for geographers working in fields that combine scientific and social interests such as agriculture, environment, and medicine.

Alongside these developments, a strong feminist ap proach to Science Studies has also emerged which both highlights and questions the very male centered environ ment of scientific research. Feminist scholars such as Donna Haraway, Sarah Franklin, and Evelyn Fox Keller have ar gued in various ways for a more situated understanding of scientific knowledge which takes into account the social background of the scientist. They use the insights of social constructivism to question the kinds of knowledges pro duced by scientific research and the social, political, and economic motives of the scientists and institutions that produce them. For example, Haraway has explored the links between the development of the Internet and the Cold War; Franklin, the links between patriarchy and reproductive technologies; and Fox Keller, the different approach to sci entific investigation undertaken by female scientists like the biologist Barbara McClintlock. Methodologically, this work places a great deal of emphasis on reflexivity, the need for the researchers to take their own social position into account when conducting and reporting research encounters.

At the same time, Haraway (among others, including Katherine Hayles) also sees within the hybrid, messy, and unnatural products of scientific innovation, such as cy bernetic organisms or genetically engineered mice, a metaphor for a new kind of perspective on the world sometimes referred to as posthumanism (although Haraway herself strongly resists this label). These per spectives stress how innovations in biotechnology – like genetic engineering – are making it hard to establish where humanity ends and the rest of the world begins, making a singular human perspective increasingly dif ficult, if not impossible, to establish. Such work demon strates how scientific knowledge does not stand apart from the world, but actively intervenes in it, reordering the relationships between human subjects and the world. Rather than seeing scientific knowledge as a form of privileged, god like anthropocentric perspective, which promises control and domination over the natural world, geographers have drawn on posthumanism to begin to explore how scientific knowledge is reconfiguring our relationships with nonhuman others ranging from do mestic and wild animals to drugs and machines. This work also links to Latour's more politically orientated works including *We Have Never Been Modern*, which place emphasis on bringing the nonhuman into political de bates. Latour argues that modern society has created a false division between the natural and social worlds, in which we accept that the only way to access the natural world is through the facts of scientific representation. As a result, scientific controversy and nonhuman agency have been excluded from the political realm. Latour's work suggests we rethink the place of scientific knowledge (and those nonhuman agents it claims to speak for) within political debates. Instead of seeing scientific knowledge as a source of externalized authority, we should perhaps see it as a form of what Donna Haraway calls situated knowledge, one which offers new ways of understanding and intervening in the world. We should also recognise that these new perspectives are partial and open to uncertainty and contestation like any other form of political representation.

To date, the influence of Science Studies on mainstream geography has been limited, but SSK, ANT, and posthumanism have been drawn upon by geographers seeking to question both the geographically contingent nature of scientific practice and, historically, the use of scientific approaches within geography.

History of Scientific and Geographical Knowledge

One way in which geographers have engaged with the social construction of scientific knowledge is through work in historical geography. The historian of science, Steven Shapin, has argued that the social study of science poses particular challenges for geographers as scientific knowledges, by definition, are supposed to be universal and independent of the times, places, and cultures within which they are produced. Critical historical geographers respond to the claim that science offers a universal viewpoint by 'placing the view from nowhere', relocating scientific ideas within their social and cultural contexts. Our images of science tend to associate science with stereotyped sites such as the laboratory or the scientific conference. This view is challenged by the work of historical geographers like David Livingstone and Charles Withers, who explore how scientific knowledge is produced, manipulated, circulated, and consumed in a wide range of sites and locations. These sites range from conventional sites such as laboratories and museums, to ships, coffee houses, pubs, and pigeon breeding clubs. Furthermore, they show that in each of these locations, the kinds of scientific knowledges that are produced and consumed change (or are manipulated) in response to the situation. The kind of science presented as a museum exhibit or discussed in pigeon breeding clubs is very different from the kind presented to peers at a scientific conference. The key point is that space matters in the production of scientific knowledge.

A second dimension to the historical geography of science arises from the recognition that geography itself is in many ways a scientific discipline. Geographers share one of the key dilemmas faced by sociologists of science, the recognition that they need not only to engage critically with the methods and claims of those they are studying (scientists), but also to reflect critically on their own quasi scientific methods and approaches as well. This need for reflexivity in geographical approaches to science is in some senses anticipated by wider changes in the discipline. Since the 1970s and the cultural turn in human geography, the discipline of geography has seen a growing divide between those scholars who adopted a scientific approach to geographical research (including many physical geographers, modelers, and quantitative geographers) and those who, drawing on postmodernist literatures and philosophies, sought to question the claim that those 'scientific' approaches offer an objective view of the 'real' world. Historical and cultural geographers became increasingly interested in how geographical knowledge was in itself socially constructed, and how it reflected and was influenced by the dominant intellectual traditions, beliefs, methods, and approaches of its time. In particular, these scholars sought to tease out the relationships between the origins of geographical science and the Enlightenment period (late seventeenth and eighteenth centuries) of European history, which is usually associated with the triumph of scientific method and reason over more subjective, mostly religious explanations of the world. For example, in 2001, the historical geographer Felix Driver drew on a combination of Science Studies and history of science approaches in his book, *Geography Militant*, which paints a picture of the geographical enterprises of discovery and observation as a set of cultural practices that involve mobilizing people and resources across vast distances, and translating observations and samples collected in one context into a form of knowledge that would make sense back home. He describes in great detail how the practices of discovery were not standardized and objective procedures, but highly contested and diffuse approaches where competing scholars sought to proclaim the superiority of their own methods and discredit the validity of their rivals.

Furthermore, *Geography Militant* highlights how geographers, through the application of scientific methods to processes of surveying, mapping, collecting, and bioprospecting, were complicit in the Enlightenment's colonization of the globe and the promotion of Western science as the one, true objective worldview. Acknowledging the influence of social factors in the production of scientific knowledge was an important breakthrough for studying geography as science, because it highlighted considerable historical inequalities between those who had the knowledge, resources, and status to be involved in the production of scientific facts, and those who did not. For example, during the Enlightenment period, women were actively excluded from many of the gentlemen's clubs and saloons where scientific debate took place and were rarely noted as contributors to scientific research, though some were considerable scholars in their own right. In accounts of the exploration of the areas which later became European colonies, such as Africa and the Americas, the contribution of native peoples to scientific expeditions as bearers, guides, or field assistants was rarely acknowledged. Scientific objectivity became a way in which Western society separated itself out from those worlds and societies it colonized, serving to legitimize the beliefs and practices of a colonial government. As an Enlightenment science, geography played a key role in legitimating and upholding colonial authority within the British and other European colonies, often at the expense of other supposedly nonscientific worldviews. Science Studies are now drawn on by contemporary geographers in order to critically interrogate colonial science. Bruce Braun draws on a combination of Science Studies and post structuralist approaches to highlight how scientific understandings of nature became incorporated into political projects of colonization. Braun shows that by representing the forest landscapes of British Columbia as a natural resource ‘empty’ of society, both colonial scientists and contemporary environmentalisms effectively write out the presence of First Nations peoples and their claims to land and resources.

More recently, Science Studies has become a resource for economic and quantitative geographers seeking to reflect more critically on their own quasi scientific methods. In particular, the economic geographer Trevor Barnes has drawn on the work of Haraway and ANT in order to interrogate the shift from quantitative to more hermeneutic approaches to the study of economic processes. More generally, Doreen Massey, Felix Driver, and Beth Greenhough have all drawn on the work of Latour, STS, and ANT to draw critical parallels between the way in which geographers in the past have imagined their field sites as a process of scientific discovery, and the ways in this same geographical imagination shapes contemporary fieldwork in geography.

Rethinking the Place of Science in Nature–Society Relations

Science Studies has not only allowed the development of a critical and reflexive history of science and geography, it has also informed more contemporary critiques of the discipline. Drawing on the methods and findings pioneered by Science Studies scholars, geographers have begun to rethink how scientific knowledge shapes the way we approach the study of natural and social worlds. These geographers critically engage with the assumption (inherited from Enlightenment science) that the natural world is distinct and separable from the social world. But in contrast to other approaches which seek to explore the relationship between nature and society either through cycles of economic production and consumption (Marxism) or through practices of representation (cultural geography, postmodernism), geographers drawing on social studies of science argue that nature and society are inseparably mixed. Therefore, rather than seeking to explain phenomena as either natural or social, this work takes an actor network approach, exploring the relationships between the different human and nonhuman actors that co produce a particular phenomenon. Rather than focusing on science explicitly (as in the historical projects discussed above), this work instead considers how scientific knowledge is enrolled in wider debates about, for example, agricultural production or medical research. The result is an understanding of a phenomena or problem which does not prioritize either human perspectives and claims to objectivity (as does science) or human issues and concerns (as does social science), but which instead, considers the role of scientific interventions alongside the agency and interests of a diverse range of actors. There are three broad concerns that have arisen through this research.

First, removing humans as the central focus of research generates new ideas about exactly who should be considered when thinking through the social, political, and ethical implications of a particular scientific practice. This approach calls for an expanded political community, one in which the interests of nonhumans are taken into account in social and political decision making processes. For example, in her essay on relational ethics, Sarah Whatmore maps out the different networks and relationships which link human food producers and consumers with plants, animals, and environments in a range of different locations (**Figure 1**). What she seeks to argue is that when debating the ethics of a particular form of scientific intervention, such as the use of hormonal treatments in beef production, we need to consider not just human interests, but also the rights and interests of other constituents within the network as part of an expanded ‘ethical community’. Science is studied not just as a source of knowledge about the world, but as a particular kind of intervention in the world, which affects the world through its relationships with (in this case) other elements in the food chain.

Science Studies has led geographers to rethink the relationship between scientific knowledge and its objects, interrogating how the subjects and objects of science effectively answer back or object to the ways in which science seeks to understand and represent them. For example, Steve Hinchliffe and Nick Bingham have demonstrated how entities as varied as BSE disease prions and machines (respectively) resist their scientific characterizations and in so doing articulate their own form of nonhuman agency.

Second, STS has opened up the possibility of rethinking the way we look at environmental issues such as global warming. Following on from earlier historical

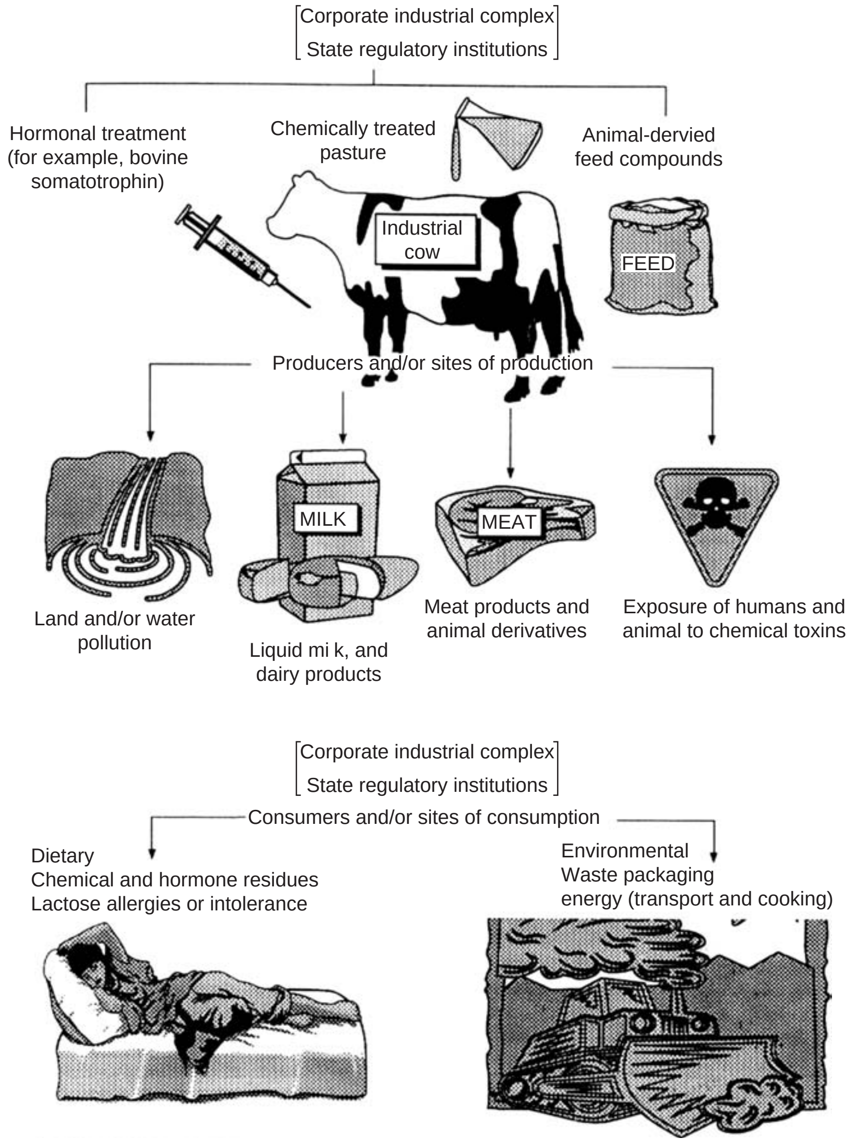

Figure 1 Diagram of agro-food network from Whatmore (1997) showing the complex relationships between producers and consumers of milk.

critiques of geographical practices, David Demeritt has used STS to look at the social construction of global warming and forestry science. His work points to how the interplay between the social and political aims and ob jectives of particular interest groups (governments, en vironmentalists, industry) can both influence and be influenced by the kinds of scientific research undertaken and the kinds of scientific knowledge produced. For ex ample, different scientific understandings of forestry can be used to argue for either the conservation of ancient woodland with no commercial exploitation or the bene fits of sustainable commercial forestry. Politically, such moves are important as they create links between seemingly abstract processes of scientific experimentation and their social, cultural, and material interventions in the world. Noel Castree and Bruce Braun, through both their own work, and through two edited collections in 1998 and 2001, *Remaking Reality* and *Social Nature*, have sought to bring together writers whose work shares an interest in how different scientific and nonscientific representations of the natural world are drawn upon to justify political and social positions. In short, this work poses the question that if nature is socially constructed who (or what) is it constructed for?

Third, geographers have combined Science Studies with an interest in political economic geographies to think through how scientific innovation and bio technology are transforming processes of production and consumption. Genetic engineering has resulted in chan ges in the way we produce and consume plant and animal resources by creating new forms of genetically modified agricultural crops and animals. Economic geographers have shown how new scientific techniques of genetic extraction and manipulation have allowed plant and animal resources to be commodified in new ways as genetic resources for the development of crops, farm animals, drugs, and pharmaceuticals. Authors such as Bronwyn Parry have highlighted how such commodifi cations often involve complex and unequal transactions between so called less developed countries that are rich in genetic resources and the American and European pharmaceutical firms that have the technologies and re sources to exploit them. The perspective of STS helps these geographers to trace the pathways through which nature is commodified and ask again, 'who benefits from this commodification of nature'?

Remapping the Human

A final strand of work has drawn on feminist Science Studies and posthumanist philosophies in order to re think the human in human geography. Traditionally, human geography has arguably seen the human as a discrete and autonomous human subject. Posthumanism challenges this by drawing on Science Studies, and, in particular, its attention to the transformative potential of scientific knowledge and perspectives, to suggest that there is no one human, no essential version of what it is to be human. Instead, they argue it is our relations in and with the world that define us, including the myriad of complex ways we are linked to our biological and ma terial environment. These more conceptually orientated studies have provided useful metaphors for thinking through exactly what it means to be human in an era of genetic engineering and nanotechnology. In a famous essay written in 1980, Haraway draws on the cyborg figure as metaphor for rethinking our political status (as humans) by recognizing that our bodies as well as our minds are engaged in and affected by political decisions and scientific innovations:

> I am making an argument for the cyborg as a fiction mapping our social and bodily reality and as an im aginative resource suggesting some very fruitful coup lings. By the late twentieth century, our time, a mythic time, we are all chimeras, theorized and fabricated hy brids of machine and organism; in short we are cyborgs. The cyborg is our ontology; it gives us our politics.

This cyborg metaphor provides a potent device for repositioning our gaze or perspective on the world and emphasizing our connections with nonhuman others, a process Donna Haraway calls figuration (see link to the Open University website for a webcast of a lecture given by Haraway on 'When Species Meet'). If we accept that the conceptual world that we inhabit as society and the material world into which science intervenes are in separably linked, we are faced with a whole new series of challenges to come to terms with. These challenges range from acknowledging our own role in setting scientific agendas to thinking through how our own bodies become sites of scientific experimentation when we ingest gen etically modified foodstuffs or try out a new drug. We make decisions not as discrete human individuals, but as people who consume genetically modified foods, donate and receive blood, organs, and DNA, co habit with pets and other animals, and enhance our bodies with visual, auditory, and other forms of technologies.

The cyborg and other metaphors have provided in spiration for empirical research within geography look ing at how science is reshaping the relationships between humans and the world. Bronwyn Parry and Beth Greenhough have studied the ways in which our bodies are increasingly redistributed and traded across space as bodily commodities – blood, tissue, DNA, and personal medical data. Gail Davies has marked how, at the same time, our bodies become the site for nonhuman inter ventions through the processes of organ donation and xenotransplantation. Catherine Nash has charted how genetic technologies and the possibility of tracing our genetic ancestry can create new kinds of genealogical identities and histories, and new ties between people and places across the globe. In different ways, these authors suggest that the human is not a finished project, but that we are being constantly made and remade in ways that tie us to diverse geographical sites and locations, or as Sarah Whatmore would suggest, how we are always becoming more than human. On a larger scale, science fiction has also provided fertile ground for trying to make sense of a society transformed by scientific innovation. Urban geographers in particular are attracted by the ability of the cyborg metaphor to capture the complex inter meshing of natural, social, and technical elements that makes up the contemporary city.

Critiques

While geography's engagement with Science Studies is fairly recent in origin, there are perhaps three key areas of conflict. First, there is a concern that a focus on actors and relations, rather than underlying social economic structures (like capitalism) or issues (like social inequal ity), means that ANT and Science Studies are apolitical or relativist. In particular, writers from a Marxist per spective are less sure of the ability of ANT to provide theoretical basis for engaging with what political econ omy sees as underlying structural inequalities and social injustices within society. This leads to a second set of concerns, namely the way in which, so far, a focus on

'Western' forms of scientific knowledge has led to a marginalization of other non Western forms and ways of knowing. The root of Science Studies in the critique of Enlightenment science tends to assume that western knowledges will always dominate/obscure other forms of knowledge. However, other accounts, such as Michael Bravo's study of navigation published as part of an edited collection on *Geography and Enlightenment* suggests scien tific knowledge was both exchanged and negotiated, even within colonial encounters. Third, posthumanist and social constructivist approaches have also received cri tique from the more scientifically orientated areas of human geography, including quantitative geographers, who, not unlike scientists in other areas, are concerned that Science Studies undermines their claims to object ivity and thereby their contribution to geographical knowledge. In addition to these, it is interesting to note that much of the critical engagement with geography as a science has focused on geographers from the past, with little work being done at present to critically engage with the work of contemporary geographical scientists. There are some notable exceptions to this rule, however, in particular a series of conversations organized by Doreen Massey, and colleagues in physical geography including Stephan Harrison and Keith Richards.

New Directions

Medical geography is one area within which the insights from Science Studies promise to produce geographies that are both theoretically innovative and politically sig nificant. For example, Edward Hall's studies of the role of genetics in heart disease treatment provides inspiration for the development of more critical health geographies by combining a Science Studies approach to studying biomedical innovation with an empirical interest in health and care practices. Second, scholars are combining the ethnographic approach employed by Science Studies with theoretical ideas on materiality and embodiment in order to interrogate how scientific insights and innovations are being incorporated into our everyday life. For example, Jon Murdoch and Emma Roe have combined an interest in ANT with work on esthetics to explore how science is changing the way we think about and react to so called natural products such as food, changing our perceptions of what is edible and inedible, healthy and unhealthy. Third, emerging work in geographies of biosecurity has begun to explore how genetic and information technol ogies are reproducing and challenging geographical boundaries and identities. By shifting the attention from human to nonhuman entities and by acknowledging nonhuman agency, geographers can explore how our so cial conventions and political boundaries are transgressed by a whole host of nonhuman entities. This work is exploring how viruses, pets, and feral species challenge the ways in which we seek to order the world, creating fear, disruption, and panic.

Overall, the relationship between geography and Science Studies looks to remain both provocative and productive, offering a means for geographers to engage with the implications of scientific innovation while re flecting critically on their own quasi scientific inter ventions in the world.

See also: Animal Geographies; Cyberspace/Cyberculture; Human-Nonhuman; Nature, Historical Geographies of; Nature, Social; Nature-Culture; Posthumanism/ Posthumanistic Geographies; Urban Habitats/Nature.

Further Reading

Barnes, T. (2001). Retheorizing economic geography: From the quantitative revolution to the cultural turn. *Annals of the Association of American Geographers* 91, 546 565.
Biagioli, M. (1999). *The Science Studies Reader*. New York: Routledge.
Bingham, N. (1996). Objections: From technological determinism towards geographies of relations. *Environment and Planning D* 14, 635 657.
Braun, B. and Castree, N. (eds.) (1998). *Remaking Reality: Nature at the Millenium*. London: Routledge.
Castree, N. and Braun, B. (2001). *Social Nature: Theory, Practice, and Politics*. Oxford: Blackwell.
Demeritt, D. (1996). Social theory and the reconstruction of science and geography. *Transactions of the Institute of British Geographers* 21, 484 503.
Driver, F. (2001). *Geography Militant: Cultures of Exploration and Empire*. Oxford: Blackwell.
Franklin, S. (1997). *Embodied Progress: A Cultural Account of Assisted Conception*. London: Routledge.
Greenhough, B. (2006). Tales of an island laboratory: Defining the field in geography and Science Studies. *Transactions of the Institute of British Geographers* 31, 224 237.
Greenhough, B. and Roe, E. (eds.) (2006). *Theme Issue: Towards a Geography of Bodily Biotechnologies*. *Environment and Planning A* 38, 416 498.
Hall, E. (2004). Spaces and networks of genetic knowledge making: The 'geneticisation' of heart disease. *Health and Place* 10, 311 318.
Haraway, D. (1991). *Simians, Cyborgs and Women*. London: Free Association Books.
Harrison, S., Massey, D., Richards, K. *et al*. (2004). Thinking across the divide: Perspectives on the conversations between physical and human geography. *Area* 36, 435 442.
Hayles, K. (1999). *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature and Informatics*. Chicago, IL: University of Chicago Press.
Keller, E. F. (1983). *A Feeling for the Organism: The Life and Work of Barbara McClintock*. San Francisco, CA: W. H. Freeman and Co.
Latour, B. (1993). *We Have Never Been Modern*. Harvard: Harvard University Press.
Latour, B. (1999). *Pandora's Hope*. London: Harvard University Press.
Latour, B. and Woolgar, S. (1986). *Laboratory Life: The Construction of Scientific Facts*. Princeton, NJ: Princeton University Press.
Livingstone, D. N. and Withers, C. W. J. (1999). *Geography and Enlightenment*. London: University of Chicago Press.
Massey, D. (2003). Imagining the field. In Pryke, R. & Whatmore, S. (eds.) *Using Social Theory: Thinking through Research*, pp 71 88. London: Sage.
Nash, C. (2005). Geographies of relatedness. *Transactions of the Institute of British Geographers* 30, 449 462.
Parry, B. (2004). *Trading the Genome*. New York: Columbia.

Shapin, S. (1998). Placing the view from nowhere: Historical and sociological problems in the location of science. *Transactions of the Institute of British Geographers* 23, 5 12.

Whatmore, S. (2002). *Hybrid Geographies: Natures, Cultures, Spaces*. London: Sage.

Relevant Websites

http://www.ensmp.fr
Bruno Latour has a personal website which includes copies of many of his articles in both French and English, as well as a biography of Latour and details of some of his many projects.

http://stadium.open.ac.uk
The following website contains a recording of a lecture given by Donna Haraway at the Pavis Centre, Open University on the Wednesday 11th October based on her forthcoming book, *When Species Meet*. In the talk she outlines the process of figuration, and how it helps her re think nature society relations.

Socialism

S. Engel-Di Mauro, SUNY New Paltz, NY, USA

Origins

By the time it circulated in print in the 1820s in Britain and France, the term socialism already described several, sometimes mutually antagonistic political currents in different parts of Europe. In part, this diversity can be attributed to the relatively deep roots of socialist political movements and philosophies. The peasant dominated revolts, alternative communities, and neo Platonic and millenarian philosophical schools beginning some 500 years ago all contributed to the shaping of critical per spectives that eventually evolved into what is now called socialism (some argue for even earlier origins, but the social contexts and philosophical content of earlier re bellions, such as the one led by Spartacus, are much too different from and hardly linked to specifically socialist struggles).

Responding to late feudal social contradictions and tensions, these rebellions, alternative political practices, and new conceptions of society were inspired by com munalistic ideals, which were often supported through religious principles and argumentation and posited in contrast to an emerging social order that threatened or destroyed peasant livelihoods and many forms of socially cooperative practices. Alternative societal ideals were also influenced by the first sustained and ultimately de structive contacts between Native Americans and Afri cans and largely Western European colonists. The spread of knowledge in Europe about societies radically differ ent from those then known (among European elites) in spired much speculation about alternative social realities.

Utopianism

Combined with major philosophical and religious shifts, such speculation resulted in the first writings engaging with the social tensions associated with an emergent social order (e.g., Thomas Moore's *Utopia* and Tommaso Campanella's *Civitas solis*). Such works envisioned the development of morally just societies based on the cul tivation of social harmony through the imposition of strict behavioral codes and asceticism and on the sup pression of money, profit, and/or private property. In his *History of political economy* of 1839, Jerome Blanqui would refer to these precursors as 'utopian socialist', a term adopted and popularized through Marx and Engels' *Communist manifesto* of 1848, to demarcate these earlier prescriptive imaginings from what was viewed as a more progressive scientific approach. These precursors to socialist thought were largely isolated from the mostly peasant movements and millenarian philosophies, which arose in response to the economic duress imposed by aristocratic elites and hardly sought the formation of an entirely new social order.

Though a direct link between utopians and peasant uprisings did not exist, utopian critiques had gained widespread influence among both social reformers and leaders of mass movements through the circulation of translations in debating circles among both clerical and secular elites and scholars (e.g., the revolutionary theol ogy of Thomas Müntzer, Huldreich Zwingli's social program in 1520s Zürich). Such influence on movement leaders can be discerned in the insistence on specifically communal access or even control of land, besides the demands for the reduction of levies, which already questioned the legitimacy of elite privileges. Through land struggles, peasant movements challenged the dis tribution of wealth on the basis of private property. This can partly explain the especially savage and vindictive manner through which the feudal monarchies suppressed millenarian movements. However, peasant revolts were not the only a manifestation of late feudal social tensions. There were sporadic, urban centered artisan movements struggling for higher wages, better working conditions, and greater control over the means to livelihood. The first such struggle occurred with the Ciompi Rebellion of 1388, which met a fate similar to peasant revolts. At times, the interests of peasants and artisans converged so that the two groupings merged into single mass movements.

The extreme reprisals and moderate concessions of the sixteenth century contributed to a relative hiatus in the demands for greater political inclusion and more economic equality. The exception was in Britain with the Levelers and Diggers (whose ideas were largely repre sented in the writings of Winstanley) in the 1640s and 1650s. Communities sometimes inspired by mixtures of millenarian and utopian socialist principles were also founded in the American colonies, at times to seek refuge from persecution (e.g., Bruderhof Hutterite communes). Otherwise, peasant and urban insurrections continued, with hundreds if not thousands of peasants murdered by better organized armies, but their content was largely limited to reactions to overwhelming taxation systems (e.g., France under Louis XIV) or foreign rule (e.g., against the Spanish Crown in the Kingdom of Naples). Radical intellectual circles were similarly decimated and it was not until the mid to late eighteenth century that utopian socialist writings re emerged in trenchant critical form, most famously with those of Rousseau but also through Morelly and later Mably, who advocated reforms toward the development of a communist society. These critiques were in reaction to philosophical apologists of capitalism, such as Quesnay and Voltaire, with whom they engaged in lengthy debates.

During the seventeenth and eighteenth centuries, with the twin developments of peasant dispossession and proletarianization and industrialization, mainly in Britain and its North American colonies, waged worker discontent added considerably to preexisting peasant unrest. Though early worker revolts manifested themselves through general strikes (e.g., in 1740 and 1786 in France), radical intellectual circles were hardly responsive to or representative of workers' and peasants' conditions or demands, possibly because most intellectuals' interests lay in the reform of monarchical and emerging capitalist arrangements, rather than their overthrow. The exception was again Britain and its North American colonies, where the increasing proletarian population was more effectively integrated through greater political concessions thanks to the war and succeeding constitutional reforms of 1688 (workers already associated into craft unions by the 1720s).

In other parts of Europe, regimes were much more dictatorial and capitalist elites were embryonic, so that conditions for the emergence of specifically socialist alternatives did not yet exist. In Germany, for example, it took political economic unification, state directed industrialization, and a repressive monarchy, and the defeat of Lassalle's brand of socialism, among other processes, to produce viable socialist and trade union movements (which enabled the formation of the most influential socialist force in Europe in the Social Democratic Party by the 1890s), and this did not transpire until the nineteenth century. However, in mid to late eighteenth century France, famine, excessive taxation burdens, a general fiscal crisis, and political repression, among other factors, led to a simultaneous revolt of both a rising bourgeoisie and the peasant and proletarian poor. The latter swelled the ranks of a highly variegated, grassroots movement (the *sans culottes*) that was strongly influenced by Rousseau's ideas and therefore excluded demands for the abolition of private property and/or money, in contrast to earlier political struggles described above.

Constitutional changes following the toppling of the monarchy in 1789 therefore addressed only formal political structures. In any case, various forces conspired violently in the 1790s to ensure that only a privileged male capitalist minority would mostly benefit from the revolution (as reflected in the guillotining of playwright and women's rights advocate Olympe de Gouges in 1793). Despite the nonsocialist nature of the revolution, events in France became a fulcrum for the development of socialist thought and movements. This is partly due to a minority of radicals that took the egalitarian promises of the revolution seriously and thereby endeavored to develop systematic socialist programs. A conceptualization of politics in terms of class struggle emerged later, along with a delineation of a specifically communist and more systematic political platform. At first, it was very much the view of a small minority, especially associated with the Conspiracy of the Equals (a secret society led by Baboeuf), most of whose membership became another victim of postrevolutionary repression.

In the aftermath of violence and political repression in France and with the spread of large scale industrial methods of production, solutions to the 'social question', as tensions connected to capitalist development came to be understood, were increasingly formulated in terms of the collective management of the economy. Self described socialist groups began to form by the 1820s and envisioned a progressive improvement of workers' conditions through formal political bargaining and/or an extension of private property to economically disadvantaged groups, whose suffrage would be thereby secured. However, each group developed different strategies for a socialist transformation of society.

The followers of Fourier, Owen, and the cooperative movement built or promulgated writings on exemplary, self contained communities or producers' societies in Europe, as well as the Americas. The followers of Saint Simon pressed the state for the establishment of a technocratic order based on the large scale scientific management of the economy. Most syndicalists, especially in Britain, organized workers to improve social conditions and largely accepted the legitimacy of the existing social order, at least in economically prosperous years. Finally, Jacobins, a prominent grouping during the French revolution, strove to turn state institutions into organs for the gradual extension of political rights for all. Shared by all the major socialist movements was the belief that everyone involved in production could eventually cooperate to improve society for mutual benefit, which assumed a harmonious relationship between, for instance, peasants, industrialists, and workers.

Scientific Socialism, Internationalism, and State Power

It was not until the 1830s and certainly following 1848, with the suppression of mass revolts and the onset of even more repressive regimes, that most socialist political movements and philosophies increasingly converged toward a more Baboeuvist tendency, though not necessarily in the form of secret societies. This tendency intermingled with existing evolutionary theories that saw social change unfolding through contradictory tensions within society itself, as formulated by Cabet, Spence, and Saint Simon, for example. Such strands of thought, as well as those of Hegel and the Ricardians later on, are what enabled Marx and Engels' analysis of capitalism as a system of dispossession and exploitation and their development of dialectical historical materialism. Their ideas eventually impacted socialist movements inter nationally with the publication of the *Communist manifesto* and through the debates in the International Working men's Association (the First International). They became highly influential from the 1890s onward as part of the political philosophies propounded by the German Social Democratic Party. Eventually, a few of the Marxist scions evolved into statist and vanguardist approaches that formed the ideological backbone of state socialist dicta torial regimes (these approaches are reminiscent of the communist current represented by the Conspiracy of the Equals in late eighteenth century France and, in the succeeding century, that of Blanqui). Other theories also developed out of similar critical perspectives, such as Proudhon's federalism and Godwin's social individual ism. Their ideas prepared the ground for the reception of theories from Bakunin, Kropotkin, and the Owenite Benbouw, among others, who stressed the inseparability of state and capitalism and therefore the necessity to overcome both simultaneously. Such ideas influenced the formation of revolutionary syndicalist and anarcho communist movements, especially in southern Europe, and revolutions in places like the Ukraine, Mexico, and Spain, in the 1900s.

In other words, by the time an internationalist socialist organization took shape in the form of the First Inter national in 1864 (which fell apart less than a decade later), socialism was already broadly divided into syn dicalist, anarchist, communist, and statist factions. This diversity deeply affected any attempts at international coordination among socialist movement. From the very beginning, the major contentions among and within the different camps were over the level and type of cen tralization and coordination (both internal organizational structure and a future and/or intermediate society), the type of general strategies (national and international, seizing the state and stage based progression, gradualism through electioneering and reforms, immediate singular or succeeding revolutions, etc.), methods of political engagement (propagandist, insurrectional, parliament arian, and/or interventionist, etc.), and the relationship to and role of state institutions. The lack of consensus on these issues often degenerated into bitter struggles for political influence and/or institutional supremacy. The effects were such salient self defeating events as the lack of support from British and some French socialists for the Paris Commune in 1871 and the folding of both the First and Second International (the Third International was essentially reduced to a foreign policy organ of the USSR), as well as the subsequent proliferation of inter national socialist organizations that ignored each other, multiplied similar efforts, or actively sabotaged one another.

Some socialist groupings eventually succeeded in gaining political power through statist strategies (e.g., in 1848 France, 1890s Germany, in 1917 Russia) and evolved into social democratic or labor parties and/or repressive rulers of socialist states or states rising from anticolonial national liberation struggles (if one were to confine the analysis to mere labels, socialist political practice could even include neoliberal capitalism, judging by the policies of socialist parties in Western Europe represented in the Socialist International). As part of institutionalized political formations, socialism is therefore often difficult to distinguish from other mainstream political practices centered on modernization and capitalist economic ar rangements (e.g., wage based employment).

Other socialist camps remained largely on the fringes of the political mainstream and at times suffered mas sacres at the hands of both liberal democratic and Soviet or Maoist state socialist forces. However, they were influential in the global revolts of 1968, which was an outpouring of liberation movements long excluded within most socialist institutions, and have arguably been important in current antisystemic movements, through the World Social Forum, among other international groupings.

In sum, since the 1800s, socialism developed into di verse theories and movements that gradually gained prominence in Europe and in parts of the world colon ized by European empires. Partially as a result of global capitalist expansionism and the collapse of one exploit ative statist model (the USSR), socialist ideas and politics continue to proliferate worldwide at the level of both grassroots organizing and state and international insti tutions. The persistence of this influence has prompted reactions, sometimes violent, on the part of elites op posed to any concessions to socialist or socialist inspired demands.

Socialisms and Geographers

Even though geography had been institutionalized by the 1820s (first in Germany), socialist ideas did not become influential to any appreciable degree until the late 1960s, when institutional socialism, reduced either to Soviet or Maoist dogma or parliamentary reformism, was already under pressure from alternative grassroots movements. In fact, from their very beginnings, all socialist variants were challenged and shaped by actual social practices and historical events, such as the Haitian Revolution of 1804 and early feminist movements of the nineteenth century, that is, prior to the emergence of human geography as part of academic institutions. However, the questioning of colonialism, racism, and masculinism implied in such events and movements rarely took center stage in the debates and policies of the various socialist camps, at least not until the 1960s. The overwhelming majority of academic geographers genuinely believed in the neutrality of their work or shied well clear of politically risky involvement. The tendency, regardless, was to reinforce, if not willfully contribute to prevailing ideologies, state apparatuses, and imperialistic foreign policies. The nineteenth century writings and political engagement of the anarchist geographers Peter Kropotkin and Elise Reclus represented very rare instances of social critique.

By the late 1960s, heterodox variants of Marxism and, to a lesser extent, collectivist anarchism gained more academic support more than a century later as a result of changes in the social composition of academic geography, changes that fostered greater sensitization to contemporary political upheavals. Within the rising radical geography wing of human geography, it was Marxist theories that gained prominence, due to their greater theoretical systematicity, larger body of available literature, and relatively more politically prominent history and clout. Arguably, it was in select Western European and North American geography departments where a major shift occurred and where deep dissatisfaction with existing academic orientations compelled a minority to shape new areas of study of greater relevance to wider society. The dissatisfaction originated from a dissonance between the theoretical potentials discovered through political activism and the prevailing geographical epistemologies and methodologies (largely closed and descriptive quantitative perspectives implicitly imbued with capitalist ideologies) that constrained the development of social theory. Richard Peet's founding of the journal *Antipode* in 1969 provided a pivotal academic outlet for radical geographers, whose early work primarily engaged with issues of social justice (e.g., discrimination against women, racism, worker exploitation, uneven development, environmental degradation). Internal theoretical debates and the establishment of the Union of Socialist Geographers in 1974 eventually steered radical geography toward more socialist and especially Marxist perspectives. Scholarly works of wide intellectual influence subsequently propelled such theoretical approaches into the mainstream of Anglophone geography by the 1980s (e.g., David Harvey's *Social justice and the city*, published in 1973; Neil Smith's *Uneven development*, published in 1984; or the 1980s recuperation of Henri Lefebvre's *The production of space*, published in 1974).

These socialist strands were soon challenged from within the same radical intellectual circles. This was due to a lack of involvement with other coeval social struggles and theory building that addressed inequalities and repressive practices related to cultural constructions of race, gender, sexuality, ability, among other processes exploitative and suppressive of corporeal diversity that were largely ignored by most Marxists. Feminist scholars, for instance, questioned the treatment of working class perspectives in universalist terms, since they often reflected the realities of only some men in some social contexts. Such empirically erroneous universalizations on the part of many Marxists bear unfortunate resemblances to historically tunnel visioned political practices, such as the indifference of trade unions or socialist organizations and parties toward the outcomes of the strikes of the 1830s by women workers in Lowell textile factories in the United States or the women's suffrage movements in different parts of the world.

Since the late 1980s, there have been many debates about the status of class relative to other social processes that have problematized the masculine, white European, heterosexual subject implied in much of socialist thought and political practice. In spite of the splintering of various groups and sometimes acrimonious nature of argumentation, the debates have resulted in constructive shifts in different socialist strands, within and outside human geography, which finally take seriously and integrate analytically the many axes of social processes, including patriarchy, heterosexism, racialization, ablism, and agism. The influences of many social movements, including feminist, antiracist, postcolonial, and queer, on socialist groups and on human geography have greatly improved the epistemological and ontological premises of socialist practices and theories in academic settings, leading to integrative approaches such as Marxist and anarchist feminisms. At the same time, greater concern for ecological and nonhuman realities have opened up different forms of socialism to more holistic understandings that are now emerging in the forms of such movements as ecosocialism, ecoMarxism, material ecofeminism, and Marxist feminist environmentalism. In spite of these convergent tendencies, there remain many obstacles to integration of different strands of socialism within human geography (as in the world generally), while some existing socialist movements and human geographers of socialist persuasion continue to ignore the above mentioned critiques, especially within institutionalized political formations that still claim affinity to socialist philosophies.

In the Soviet dominated or influenced contexts, geographers mostly followed (or were forced to follow) a rigid catechism that was a parody of Marx's (and even Lenin's) writings. Academic geography was often subordinated to the functioning of the administrative organs of the state. No socialist geography actually developed that was independent of ruling party elites, but there were major shifts in the late 1950s (e.g., expressed in Anuchin's work) and the late 1970s and 1980s, with the appearance of social or humanistic geography. This was also part of a general reaction against the reductionism of quantitative approaches that was facilitated by wider political apertures, a reaction not too dissimilar to its radical counterpart in Western Europe and North America and that would merit a comparative analysis.

An account of alternative academic geographies en compassing other parts of the world remains to be writ ten. However, one might expect substantively different intellectual histories as well as continuities. Though largely Eurocentric socialist theories (especially of the Soviet and Maoist varieties) have often been super imposed in a facile manner in (post)colonial situations for geopolitical expedience (e.g., in national liberation struggles), most non European intellectuals have also been trained through Eurocentric canons. Heterodox geographical traditions may therefore be ensconced within non European socialist theories, not necessarily within academic geography (as in the case of Bogdanov in Russia or Foucault and Lefebvre in France). The emergence of these traditions may be enabled by the USSR's demise and the intensification of capitalist ex pansionism, which has facilitated renewed interest in socialist theories and politics and a blossoming of re search on socialism in practice.

See also: Anarchism/Anarchist Geography; Capitalism; Colonialism I; Colonialism II; Imperialism, Cultural; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Patriarchy; Racism and Antiracism; Social Class; Structural Marxism.

Further Reading

Arrighi, G., Hopkins, T. K. and Wallerstein, I. (1989). *Antisystemic movements*. London: Verso.

Blunt, A. and Wills, J. (2001). *Dissident geographies: An introduction to radical ideas and practices*. New York: Prentice Hall.

Braunthal, J. (1967/1961 3). *History of the international* (2 vols.). 1967/1961 3, New York: Praeger.

Cole, G. D. H. (1953 1960). *History of socialist thought* (5 vols.). London: MacMillan.

Guérin, D. (1970). *Anarchism: From theory to practice*. New York: Monthly Review.

Guillaume, D. (1993). *L'Internationale Socialiste. Histoire et sociologie du socialisme international (1945 1990)*. Paris: Presse de la Fondation Nationale des Sciences Politiques.

Hofmann, W. (1971). *Ideengeschichte der sozialen Bewegung des 19. und 20. Jahrhunderts*. Berlin: Walter de Gruyter & Co.

Marshall, P. (1993). *Demanding the impossible: History of anarchism*. London: Fontana Press.

Mazurkiewicz, L. (1992). *Human geography in Eastern Europe and the former Soviet Union*. London: Belhaven Press.

Monk, J. and Hanson, S. (1982). On not excluding half of the human in human geography. *The Professional Geographer* 34, 11 23.

Peet, R. (1998). *Modern geographical thought*. Oxford: Blackwell.

Peet, R. and Thrift, N. (eds.) (1989). *New models in geography*. London: Unwin and Hyman.

Renton, D. (2004). *Dissident Marxism: Past voices for present times*. London: Zed Books.

Robinson, C. J. (2001). *An anthropology of Marxism*. Aldershot: Ashgate.

Salleh, A. (1997). *Ecofeminism as politics: Nature, Marx, and the postmodern*. London: Zed Books.

Salmon, S. and Herod, A. (2003). Socialist geography. In Gaile, G. L. & Willmott, C. J. (eds.) *Geography in America at the dawn of the 21st century*, pp 209 220. Oxford: Oxford University Press.

Smith, G. E. (1989). *Planned development in the socialist world*. Cambridge: Cambridge University Press.

Spini, G. (1992). *Le. origini del socialismo. Da Utopia alla bandiera rossa*. Torino: Giulio Einaudi Editore.

Wallerstein, I. (1991). *Unthinking social science: The limits of nineteenth century paradigms*. Cambridge: Polity Press.

Relevant Websites

http://dwardmac.pitzer.edu
Anarchist Archives.

http://www.marxists.org
Marxist Internet Archive.

Society–Space

S. M. Ruddick, University of Toronto, Toronto, ON, Canada

Glossary

Fordism A system of mass production and consumption that originated with Henry Ford, and led to stabilization of wage relations, guarding against overproduction or underconsumption through state entitlements, regulation of trade agreements, and a culture of mass consumption.

Keynesianism A system of social entitlements that included unemployment insurance, healthcare, mother's allowance, and welfare, originating with John Maynard Keynes and instituted after the Great Depression of the 1930s in many industrialized nations.

Neoliberalism A strategy of governance that attempts as much as possible to reduce a state's responsibility for social entitlements and to download responsibilities for care onto cities, communities, families, and individuals.

Post-Fordism A globalized system of production and consumption characterized by the increasing use of robotics in production systems, externalizing uncertainties of the market to small-time suppliers who produce on demand (just in time), and the marketing of specialized goods to globalized 'niches' of consumers.

Post-Structuralism A movement in French continental philosophy centered on the de-stabilization of claims to truth and variously connected to the antihumanist rejection of a unified rational subject; a rejection of binary oppositions; and a view which sees knowledge as fluid, unstable, and discursively produced rather than ideological.

Introduction

When figures as divergent as conservative British Prime Minister, Margaret Thatcher, and left leaning political philosopher, Ernesto Laclau, declare society is an "impossible object" we might ask – why investigate the relationship between society and space? The topic appears so vast that when properly charted – like Borges' famous map drawn at a 1:1 scale – so much ground is covered as to become practically and analytically useless. Moreover, the relationship would seem a truism: societies create spaces that best express their needs; spaces in turn constrain or enable societal developments. But how we conceive of this relationship subtends any attempt to understand, critique, or transform, the social world. Our understanding of society–space, as a totality, determines the vantage point from which we understand the social world. It locates these delimited processes, both literally and figuratively, in relation to a social whole.

The first approach views society–space in terms of exploitative relations organized in a 'structured coherence' – a social formation, whose territorial boundaries loosely approximate politicojuridical boundaries. Rooted in a Marxist vision of social inequity, whereby societal relations are defined by exploitation – alienated labor – its debates have focused on the logical primacy of particular unequal relations (classed, gendered, and racialized) and their attendant spatialities (workplace, home, underdeveloped neighborhoods, regions, and nations). The second approach sees society–space as a 'strategic field' – constituted through productive relations of governance – an occupied zone in which struggles resemble the engagements of a war. Underpinned by philosophical approaches rooted in the work of Foucault, Nietzsche, Heidegger, and Clausewitz, in this view, societies are organized through a constellation of discursive logics, enacted in everyday spaces, and through these practices the abnormal are contained, controlled, excluded, or policed. The third sees society–space as 'performative field', beginning with the premise that society is fundamentally unstable, its subjects perpetually hybridizing, caught in structured instabilities – in a social field that is conflictual and paradoxical. In the fourth approach, society–space is an 'immanent field', an expression of continuous differentiation, its subjects are unstable because they continuously transform themselves, engaging in acts of conjunction, connection, or collaboration with their milieu.

These four approaches share a recognition of persistent structural social inequities, a conviction that space is complicit in the production and possible transformation of these relationships, and the belief that one should not merely analyze societies but attempt to change them – to create a more equitable world.

Looking Backward: Overcoming Newton and Kant

The contemporary view is that the relationship between the spatial and social is an iterative one. This is by now so commonplace as to seem self evident. However, this insight has only recently returned to geography after a long conceptual slumber in the seductive embrace of mathematical description. Mathematical approaches enabled a complex description of either social or spatial processes, but little theorization, and as David Harvey noted, no strategy at the interface between the spatial and the social. Geography's significant role in articulating of a concept of society and space, and the discipline's part in a larger spatial turn in social theory can best be appreciated through a brief historical detour.

The mapping presented here attempts to be as comprehensive as possible, but does not exhaust the field. Other trajectories include systems based, phenomenological, psychoanalytic, ethnomethodological approaches, which are part of contemporary geographical investigation. There is, as well, a more complete archeology of influences on geography than addressed here, including the influences of Vidal de la Blache on geographical thought. The emphasis here, moreover, is on the changing status of the 'spatial' in the society–space couplet. A shifting understanding of the social through the course of twentieth century Continental and English speaking geography is charted elsewhere by Philo and Söderstorm.

Early social theorists – the sociologist, Georg Simmel, and geographer, Friedrich Ratzel, for example, posited an iterative relation between society and space. Inspired by the rapidity of urban industrialization in the late 1800s, they counterposed rural existence to the distinctive qualities of urban life. Urban life, they theorized, became blended and compressed, a locus where anonymity encouraged nonconformity in thought. Apart from their theoretical insights, geography was for the most part, ruled by idiographic tendencies – detailed descriptions of specific places, which were useful for exploration, colonization, and conquest, but resistant to theoretical production.

By the 1940s normative and methodological impulses prompted a move away from even these early theorizations. Ratzel's concept of *Lebensraum* was discredited through its appropriation by the Nazi party: a development that, perhaps, propelled German geography at the end of World War II back to the safe terrain of idiography and description. The rising methodological hegemony of quantitative approaches encouraged the revival of a Newtonian view of space as a container of social processes, and a search for pure pattern in the mathematical and geometrical expression of social processes in their spatial forms.

Social area analysis, which reached its zenith in post World War II quantitative geography, viewed human societies as an aggregate of myriad individual decisions, which became routinized and achieved equilibrium in distinct spatial forms. Undergirding this approach was a neoclassical economic model of rational choice – perhaps suited to the economic and political stability following World War II, but ill equipped for what was to come afterward. By the late 1960s social unrest began to trouble the implied equilibrium of this framework. For Anglo American geographers, including David Harvey, the 'problem of the ghetto' remained resistant to quantitative explanations. Harvey's *Social Justice and the City* was the first of a collection of transitional texts in geography, charting a course from the limitations of quantitative approaches to the possibilities of Marxism, setting a radical reorientation of human geography that has persisted to this day.

Society–Space as Structured Coherence

This is the first approach. The turn to Marxism in the late 1960s and early 1970s compensated for quantitative geography's weaknesses: class analysis provided the tools to explain social inequity; the concept of dialectically induced crisis, propelled by the conflict between capital and labor, offered a mechanism of societal transformation. The deductive power of the dialectic, moreover, compensated for both the idiographic tendencies in descriptive accounts that characterized early work in geography and the agnostic politics of quantitative approaches that came after.

However, for geographers, Marxism came with additional challenges. Marx himself had tended to ignore space entirely or reduce it, as Soja noted, to Hegel's notion of space as a state idea, remaining within the 'simplifying assumptions of a closed national economy'. With Marxism this attitude shifted from mere indifference to active hostility. Beginning with the Second International and writings of György Lukács (considered the founder of Western Marxism), attention to space was considered a fetish and diversion from class struggle. Under Stalinism, it was relegated along with culture and politics to the realm of superstructure, an irrelevant epiphenomenon.

Geographers countered this legacy with a frontal attack, arguing it was precisely the undertheorization of geography and space from Marx forward that had impoverished radical political theory and practice. Early debates in Marxist geography revolved around how far one could push the significance of space without risking spatial fetishism or subordinating class struggle – a social struggle – to a spatial one. Hyphenated Marxisms (feminist, antiracist, etc.) within geography were intent on challenging the explicit hierarchy of laboring forms within the Marxist framework. The latter privileged particular subjects (the white male Fordist worker) and their attendant spatialities (the factory in industrialized Western countries) as the linchpin of societal transformation.

From Spatial Fix to Socio-Spatial Dialectic

One line of enquiry reconceptualized the relationship between society–space: the '–', if you will. Here concepts of spatial fix, time–space compression, social space, and the sociospatial dialectic, each charted a different course between the Scylla of ignoring space entirely, and the Charybdis of fetishizing it altogether.

Working perhaps most closely with Marx's texts, David Harvey interrogated the role of space in the circuit of capital. For Harvey, space constituted a limit to the process of circulation – Harvey's concept of spatial fix. However, technological developments in the labor process could also erode distances between places – time–space compression. Harvey's concept of spatial fix unmasked the geographical and territorial dimensions of the concept of dead labor – labor invested in particular infrastructures and technologies that weighed 'heavily on the living'; time–space compression on the other hand expressed the manner in which capital transformed technologies (and the division of labor) to overcome this predicament, at the same time, annihilating fixed space with accelerated time. These two extremes of fixity and flight delimited the terrain of societal reform, and the extent to which surplus might be diverted from capital that was bound to a region, before it could leave.

It was one thing to argue that space could act as a break or an accelerant to the circulation of capital; it was quite another to propose a historical inversion in the relationship between capital and space. Lefebvre's concept of social space – a counterpart to Marx's concept of social labor – did just that. Space left the realm of covert strategy and was elevated to the status of necessity. Space was not merely a productive component of contemporary capitalism; capitalism was now produced, itself, by space. Working along multiple trajectories Lefebvre charted the geohistory of spaces. This he traced from the world of early human history organized primarily through natural spaces (in which natural features were discovered and exploited), to contemporary abstract spaces, wherein distinctions between places in the globe were erased and nature was subordinated to a grid of functions. In this vision, (inspired by Lefebvre's own lifetime experience of the French state's relentless modernization of its countryside), modernity and modernization were not emancipatory conditions, but rather an expression of a commodification tending toward barbarism. Modernization was a means by which capitalism ensured, by stealth, its survival. Social space raised questions about the centrality of class as the *sine qua non* of social transformation. The city was elevated from its status as a micro expression of larger socioeconomic processes to a central point of intervention. Social space reconceptualized the terrain of struggle: spilling out of the factory and on to the streets.

For some geographers, such as Harvey, this position bordered dangerously on a fetishizing of space; for others – Edward Soja in particular, it marked the return of space to its proper place in critical social theory after a century of its subordination.

Soja's sociospatial dialectic took Lefebvre's dialectical materialism, a dialectic no longer tied to temporality, as a point of departure. Drawing on Lefebvre's argument that space was in some ways homological to class structure, expressing major contradictions and acting as a vehicle for societal transformation, he reassessed the latent spatiality of the work of Marxist theorists within the discipline and beyond.

The sociospatial dialectic neatly side stepped the question of logical primacy in the society–space couplet. More provocative than the dialectic was Soja's argument that the intersection of simultaneous synchronic socio spatial processes, emanating from multiple locales, took precedent over the diachronic, which viewed social change as a sequence of local developments unfolding in place. This viewpoint amplified a similar argument made by scholars of English literature (such as John Berger), that the linear narrative form of the novel had been superseded by a multiplicity of intersecting viewpoints. Although space and the social were iterative, in this view space now trumped time; geography trumped history.

Socio-Spatial Formations

Although these contributions addressed the role of space in the society–space couplet, they still begged the question as to how one bounded this concept. Where did one society begin and another end? The political implications were clear: one had to understand what 'larger' processes must be considered in any study of social inequity in order to appraise the political field that might over determine it.

Marxists sought to rescue politics and culture from their status as epiphenomena and counter Stalin's base superstructure model by drawing on and extending Lenin's and Marx's concepts of the socioeconomic formation. The resulting conception of the social formation was understood as the complex articulation of social, economic, and political practices that together expressed a coherent society – a structured coherence, or totality of instances. Its boundaries were roughly congruent with national economies, but this scale reflected the hegemonic importance of national economies in the 1970s – today the European Union might be considered a social formation.

The sociospatial formation (introduced by South American radical theorist, Milton Santos, in collaboration with geographer, Dick Peet), marked the first tentative vision of active space, countering the Newtonian view that the nation state was merely the container of relevant social, political, and economic processes. Space figured actively in both production and social reproduction. Santos' inspiration for the concept came from a Hegelian understanding of things (as endowed with content and purpose), and practical observations about the role of modern planning infrastructures, which, in nonaligned South American economies at that time, acted as a means to attack and transform noncapitalist economies. Here the reorganization of space had preceded economic and political transformation as a mechanism of accelerated capitalist modernization.

The implicitly national agenda that underpinned early conceptions of the sociospatial formation reflected an unstated understanding that the national arena was central to any serious political agenda for societal transformation, and the vehicle from which to revive an international politics. The structured coherence of societies coalesced, logically, at the national level – a view reinforced by the centrality of the Keynesian welfare state during the Fordist period. The implication was that the nation state itself was the target of any necessary change.

The demise of the Keynesian bargain, the differentiation of the economic fates of regions and cities, the rise of frontier zones, and the intensification of globalization were to erode this view – echoed in evolving debates about regional and urban specificity, the interpenetration of global and local, and finally, the salience of scale itself as a concept. At the same time theorists were beginning to grapple with the complexities of global–local relationships and their articulation in specific places. Doreen Massey's concept of power geometries captured the differentiated nature of globalizing societal processes, which propelled jet setting elites around the world with increasing ease, while confining the less fortunate to increasing immobility or distended travel times – long waits in bus lines to gain access to shopping centers, green grocers, and basic amenities to cover the bare necessities of life.

The Paradigmatic City

In the mid 1970s the national context so predominated as to provoke a debate around the specificity of the urban: not simply a microcosm of larger societal processes (within the structured coherence of a sociospatial formation), nor entirely its own separate reality. By the 1980s however, the city not only asserted its own specificity: it was considered paradigmatic of a new social–spatial form. For Soja, Los Angeles undermined the underlying precepts of the structured coherence of the sociospatial formation and offered, in its place, the city as a global node in an urban hierarchy, defined by its relative position in a global network of cities.

This polycentric, post Fordist agglomeration became the platform from which to theorize three interrelated restructurings of the society–space relationship: post historicism, that is the ascendancy of geographical over historical thought; post Fordism, which was the shift from national economies of mass production and consumption to global economies that organized geographically dispersed labor pools, an international division of labor, and marketed to niches; and postmodernism, which admitted a cacophony of social subjects and processes into the arena of societal change. Some theorists considered this singular focus on Los Angeles a somewhat paradoxical maneuver – to declare a city as paradigmatic at the same time as it demonstrated spatial heterogeneity raised questions about the significance of other places.

Nevertheless, *Postmodern Geographies* was perhaps the most forceful elaboration of the thesis on the centrality of space in the society–space couplet and another key text marking a turning point in geographic thought. For feminist geographers however, Soja's recognition of cacophonous, differentiated voices had not gone far enough: a geography on the postmodern, to be sure, but written from a modern perspective. Although the book acknowledged the significance of a multiplicity of different voices, questions about the specificity of these differences awaited further investigation.

The view of society–space as a social–spatial formation has been the subject of two principle critiques. The first – against diffusionism – challenged the implied hierarchy between the First and Third World or developed and underdeveloped nations; the second – privileging social reproduction – challenged the primacy of subjects and spaces of production.

Against Diffusionism

Theorists working on underdevelopment, and later postcolonial studies, raised questions about the unstated Eurocentrism and diffusionism of many Marxist approaches, which presumed that societies, (nationally configured) were expected to follow the European model of modernization and transition to capitalism. C. L. R. James and the Caribbean school challenged the implied teleology of European development, arguing that slavery in Haiti provided the economic basis for the French Revolution. Geographers deepened this analysis. James Blaut's *The Colonizer's Model of the World* offered an extensive critique of the many manifestations of Eurocentrism and myths of diffusionism – which suggested that European models of society were the most highly developed forms, to be adopted over time by more backward societies.

Clyde Woods' *Development Arrested* countered similar mythologies about the United States. The typical view was that societal development in the United States was led by a modernizing industrializing north, contrasted by an anachronistic form of economic and social development in the American south. Woods' research documents the persistent underdevelopment of the Mississippi Delta, whose peoples and resources have been continuously plundered by plantation economies from the late 1600s on. The political and economic forms in the south moreover, presage contemporary neoliberal strategies of dispossession.

Geographers have offered ways to conceptualize the resonances between processes of underdevelopment common to the First and Third World. Jim Blaut's conception of the American ghetto as an internal neocolony has contemporary relevance to understanding the wanton neglect of infrastructures leading to the destruction of

New Orleans. Katz's concept of disintegrating develop ments links parallel processes of impoverishment in places as far flung as New York and Howa, recently documented in *Growing up Global*.

The full significance of these works, and the extent to which they demonstrate the complicity of slavery and nonindustrial forms of exploitation in the rise of capit alism and modernity – not just as precursors, but active laboratories, has yet to be fully appreciated. These works challenge the implied hierarchy of sociospatial develop ment, symptomatic of a persistent colonial mindset, which suggests that only when countries have passed through specific stages of capitalist development might they be ripe for any kind of transformation.

Spaces of Social Reproduction

Feminist geographers opened a battle on two fronts – first to insist on a Marxism that did not consider gender issues and the production of labor power as derivative, but rather as central to social struggle, and second, to demonstrate the specificity of gendered spaces in this dynamic.

They were inspired by early work of radical feminists, and Torsten Haagerstrand's insights on time geographies, which detailed the particular spatial patterns emerging from people's daily journeys from home to work, shop ping, school, and the like. The challenge initially, was to document the difference that gender made. Feminists charted the gendered characteristics of capitalist space – from suburb to gentrified central city, the former badly serving women who had joined a secondary labor force by the mid 1970s, and the latter reducing distance be tween home and work. The doubled roles of women as homemakers and as part of a paid labor force introduced its own spatial dynamics into the development of capit alism. For example, industry's flight to the suburbs in the 1970s was as much to exploit a female labor force tied close to home as it was to flee organized unions in urban areas. The difficulties in advancing a feminist perspective in the early years should be underscored. Feminist geographies were sometimes castigated by established academics in the field, as being 'not the stuff of geog raphy', but rather more like urban sociology.

Declining state responsibilities for services associated with social reproduction, the aggressive downloading of these responsibilities to cities, communities, and families that characterized the 1990s, all conspired to rework the scalar nesting of home, national, and international spaces in complex ways. Peripheralization at the core was one manifestation of this reworking, exemplified in the re organization of home space with the transnational mi gration of childcare workers. Feminist geographers also began to unpack the ways that apparently local sites of social reproduction, from school classrooms to hospital birthing stations, supported national agendas.

Feminist geographers reworked the spatial imaginary along two axes. First, they argued for a shift from understanding social functions as operating in bounded quasi territorial containers (the city, the nation state) to a focus on nodes and flows of activities and practices (diasporas, transnational migrations). Second, they ex plored the disruption of the comfortable scalar nesting of practices: local, regional, and global were no longer thought of in terms of a kind of Russian doll set of containers; local practices might produce national or international effects; and inversions of all kinds were possible.

Society–Space as a Strategic Field

The second approach envisions society–space as a stra tegic field. The limits of the field are not political boundaries; the dispersal of particular institutional sites and technologies of rule that are constituted through them, map out the extensive space of the discursive re gime – for example, the spread of asylums, prisons, or hospitals. Subjects are differentiated in terms of their relationship to societal determinations of normal and abnormal (e.g., the criminal, the insane). Foucault's work is central to this approach, but it includes others inter ested in the societal production and regulation of trans gressive subjects: how different societies define disability, sexuality, or aging, for example, becomes a means through which to understand how societal norms are constituted.

In this approach subjects (prisoners, pupils, soldiers, teachers, doctors, the ill, the insane, the priest, the penitent) are not defined by their alienated labor, they are produced through discursive logics about what is true and what is false, what is normal and what is abnormal – the rules of conduct governing conduct itself. These find expression in a variety of institutional mechanisms and practices to produce the norm and to curtail, contain, and exclude the abnormal (the school, the military, prison, the hospital).

Over the past decade geographers have been ex ploring the changes in discursive regimes arising in response to shifts in governance. With the rise of neoli beralism, states are increasingly interested in promoting right behaviors to limit collective costs (e.g., exercise and healthy eating to reduce healthcare costs). New cat egories of social exclusion – for example, the enemy noncombatant, or the illegal alien – place subjects outside of the range of entitlements afforded ordinary citizens and juridical protections decreed through international law. Space is being deployed in new ways as a technology of control, for example, through the banning of convicted drug users from certain parts of the city, and the demarcation of limited 'zones of protest' as a space of free speech for demonstrators. Biotechnologies, such as the iris scanner, face recognition technologies, thermal registers (deployed at airports during the severe acute respiratory syndrome (SARS) epidemic) form part of a range of new technologies of surveillance. Additionally – biopower – the politics of 'life itself' has become a subject of governance as interest in the productive capacities of living matter (DNA, genetically modified organism (GMO)) raises regulatory questions and becomes part of the lexicon of governance of human populations. Campaigns around bioethical citizenship forge a link between people's ability to act responsibly, to pursue the right lifestyle, and their ability to gain access to social services.

Dispositif

Within the strategic field, particular kinds of spaces (e.g., the school or the prison) become organizing and orienting mechanisms for a larger society – a *dispositif*, or underlying diagram of power relations. Theorists have argued for the iconic status of different spaces in different historical periods: for instance Foucault suggested the school/prison/barracks expressed the logic of disciplinary societies and template for subject production in the modern era, whereas Agamben has argued that the concentration camp is a better signpost. The camp, for Agamben, marks the division between those included within the productive technologies of the state and those over which the state has absolute power, but who are reduced to bare life – states of exception. Agamben's approach has become increasingly popular as we witness a growing range and variety of excluded individuals not subject to the protection of conventional laws (e.g., refugee populations, illegal migrant workers, unrecognized states, enemy combatants).

Within this approach, changes in the *dispositif* emerge genealogically – through accidents, adaptations, and improvisations, rather than through an internally driven logic (the unfolding of the dialectic characteristic of structured coherence). The discursive technologies that exclude or contain transgressive subjects can become the organizing principle for the entire social field. These iconic subjects and their corresponding spaces become a vantage point from which the underlying discursive logic of an entire society can be revealed, although in any given period, technologies may coexist, and strategies overlap, fuse, or evolve separately.

The concept of power in this approach draws on the philosophical heritage of Nietzsche and Clausewitz: it is nondialectical – opponents confront one another on a terrain much like a battlefield. Society is a strategic field of engagement, in which each side defines the terms of conflict differently. Power focuses on relationships of governance rather than economic relations. Shifts in strategies of governance relate only opportunistically to economic changes: political and economic forces are not bound by a shared logic. Power is not defined through class relationships and then refracted through other forms. Power is productive: organized neither through ideology nor repression, (as Deleuze was to remark). Although power is dispersed – it is 'everywhere' – it also takes specific forms at each site, something critics of Foucauldian analysis often neglect to acknowledge. It is expressed as a series of specific technologies of rule; passing ubiquitously through individuals as a relay – each one subordinated to and exercised by power in specific ways.

Historically, shifts in power relations express changes in state (and other institutions') strategies of government: from the patrolling of boundaries of nation states to guaranteeing internal security within the population. In class based analyses the concept of population is often viewed as an obfuscation: here it becomes an object of rule. Governance of the daily conduct of subjects mitigates societal processes such as unemployment, migration, birth and death rates. Interests of governing elites are only tangentially linked to economic concerns. Although relations of power are negative to the extent they channel behaviors, they are also productive, rather than strictly repressive or ideological: subjects are not duped nor coerced, they are invested in their identities. Feminists have critiqued this viewpoint as reproducing the colonizer's view of the colonized, unable to acknowledge the extent to which subjects are cognizant and critical of their subject positions, aware of their paradoxical position even as they participate in their production.

This approach emphasizes the optic and technical qualities of space: the space of the gaze, disciplinary spaces, spaces of surveillance, are organized in such a way as to enact particular discursive technologies. The uniform rows of desks in the classroom, for instance, allow for maximum observation by the teacher, and emphasize the uniform status of the pupil; the organization of the hospital according to specific illnesses allows for enclosure, partitioning, and creation of functional sites and the ranking of their uses.

Geographers have explored the limits and possibilities of this approach through investigations of its philosophical heritage. On the one hand, they have critiqued the unidimensional tendencies of this view of space, which merely reflects the discursive program that has organized them, inattentive to the multiple possibilities for redefining, challenging, and inverting relationships within them. Harvey and Thrift have each lamented this treatment of space as rigid and inanimate, rather than 'lively' – the former arguing for a return to the dialectic, and the latter suggesting an appreciation of space that would be haptic, as well as optic. On the other hand, geographers have clarified philosophical influences within the approach. In *Mapping the Present*, Stuart Elden explores connections to Heideggerian and Cartesian views of space: as techne and extension. As techne, space is a technology expressing and enacting relations of power. As extension, institutions, and corresponding discursive arrangements, disperse across a social–spatial field, not limited by political boundaries – for example, schools, hospitals, and prisons have proliferated across Western nations. Power relations are enacted in a dispersal of institutional arrangements through space.

Civil society is thus thought of as an occupied zone, a battlefield. Localized practices are not understood as resonating within a larger societal field, a structured coherence of larger economic, political, and cultural processes. Instead societal discourses themselves, emerge through micro geographies, the localized space of the school, the church, the barracks, the prison cell, the confessional. The interconnection, fusion, or coordination of their logics comes to express different discursive constellations whose variation marks a historical shift from one system of technologies of rule to another. This is a bottom up rather than a top down approach, resonant with Nietzsche's frog's eye view of power relations. Different historical periods are distinguished by distinct *dispositifs* – diagrammatic relations of power, connections between discursive logics at different sites, which subtend the organization of society as a whole, and reflect how societies think about and organize the relationship between what is considered normal and abnormal.

Genealogy's Geography

Geographers have also explored the relationship between genealogical transformations and their context, guarding against a totalizing view that sees new spaces and discursive technologies as simply steamrolled across an indifferent or accommodating landscape. In this account, variations in location, topography, and social environment are not simply relegated to the status of accident or contingency. Environment (geography, territory, already built form) figures in the genealogy of discourses as an active element in this process. One excellent example is Philo's exploration of the geographies of madness. This rich account details the variation in therapeutic regimes in England and Wales – not simply imposed upon an indifferent landscape, but rather wrought from the Earth, an 'original chaotic space', a geographically differentiated brew of highway wanderings, sacred healing waters, and secluded forests and abbeys.

This reanimation of the geography of discourse has extended to an analysis of places of scientific knowledge production – the locales and networks, which were the sites of development of quantitative techniques for the analysis of population. Geographers have explored the spaces of exception that correspond to states of exception. Guantanamo Bay, the nonstate of Palestine, and the refugee camp have been figured as extraterritorial extrastate spaces. However, there are internalized states and spaces of exception as well, such as the workspace of transnational domestics who migrate to work in foreign lands without the benefit of local regulatory protections.

Critiques of the 'strategic field' approach have focused on its tendency to view resistance as inseparable from the power structures that constitute their subjects; an indifference to the multiple subject positions that might confound an individual's relationship to a particular discursive structure; and a resultant reluctance to explore the multiplicity of contested positions that might coexist in relation to a given discursive regime.

Society–Space as Performative

The third approach takes as its starting point the instability of multiple subject positions as a productive uncertainty – a performance or a becoming. Society is inherently unstable, improvisational, continuously differentiating. Its stability must be explained. The emphasis on marginality is neither simply exclusion – the Marxian view of the lumpen proletariat – nor subject production, the Foucauldian view. The spaces the marginal occupy are seen as performative, paradoxical, or cramped, expressing the ways they must knowingly negotiate their own subordinated positions within discursive constructs. These spaces become the vantage point from which to uncover blind spots in hegemonic discourse, or as de Lauretis put it "chinks and cracks in the power knowledge apparati."

This perspective is subtended by an emphasis on radical instability and difference: there is no singular, privileged vantage point from which to understand society–space as a totality – rather this perspective reveals a range of heterogeneous spaces embodying different modalities of oppression and possibilities for resistance.

Subject positions are fundamentally unstable because subjects must negotiate between competing discursive frameworks. Subjects are not preexisting, autonomous entities that relate to one another across clearly defined boundaries. Subject identity is in a state of flux – emerging in the in between space negotiated with others. Central to this project is the idea that the position of marginality is not simply one of exclusion: marginality offers a vantage point from which to critically appraise hegemonic relations of power. The political impetus for this approach has been to understand resistance to societal norms that might emanate from outside of hegemonic power relations, rather than simply reacting against them – something that Foucault and his adherents gesture toward but do not fully explore.

Paradoxical Spaces

For theorists working in this vein, inspiration has come from the conscious acts of negotiation or resistance by transgressive or marginalized subjects (e.g., the act of passing for white when black, or straight when gay or transgendered) and from social struggles that have set their own terms of reference and political objectives outside of the 'normal' frame of negotiation. Exemplary in this regard is Gillian Rose's *Feminism and Geography*, and her concept of paradoxical space. Other geographers have extended her perspective to investigate phenom enon of passing. Griselda Pollock, a scholar of the arts, has undertaken a similar reassessment of spaces of modernity through analysis of impressionist paintings. For example, Morisot's *On the Balcony*, is not viewed by Pollock as a masculinist vision of modernity, but rather a record of a white bourgeois woman's experience of spa tialized, gendered difference, marked by the distance between balcony, and street – semiprivate and public space. While the margin marks a site of exclusion, it is thus understood as a potential site of recuperation and resistance. Hooper and Soja deepen this line of argument, drawing upon bell hooks' insights on the meaning of home for black women. They conceptualize the space that difference makes as a productive, counter hege monic project.

As Caroline Desbiens notes, the question remains as to whether this space of resistance exists in a beyond (implying an as yet unrealized utopia) or an elsewhere, as sites of critical engagement that are part of the here and now, akin to the space off in film (denoting the space off screen). The idea that these spaces operate in the here and now has been forcefully demonstrated in McKit trick's mobilization of paradoxical space in *Demonic Grounds*. Here she explores the range of micro geog raphies that depict American slavery as intimately bound up in the constitution of modernity. For example, the auction block displayed a perverse rationality as a site of measurement of human worth. However, she also argues these spaces can afford possibilities for resistance. Thus a cramped attic space, which functioned for 7 years as a space of extreme confinement for an escaped slave was also a transit point to freedom. This approach invites a deeper exploration of politics through a range of spa tialized concepts – third spaces, passing, transgression, border zones, a politics of location, 'between ness', cen ter margin all come under scrutiny.

Society–Space as Immanent Field

The fourth and final approach shares much with the third, in that it sees instability as an endemic, persistent societal condition. This approach can be linked most strongly to post structuralist feminist and Deleuzian in fluences. Where it differs from the third approach is that the instability of subject positions does not arise through a structured paradox. In this view society flees. The 'lines of flight' or 'becomings' of the subjects are the paths wrought by their perpetual improvisation – subjects, bodies are not bounded entities, containers, but rather 'organizational arrangements' in a continual state of re action to their milieu. These improvisations can become channeled into stable recurring relationships, or blocked, diverted, or even overthrown. Matter is, in this view, wayward – kinetic, dynamic – and systems arise through patterns of exchange, the coding of flows of matter (people, goods, sewage, water, microorganisms) within a milieu. Through this coding stable assemblages emerge.

Theorists working in this vein eschew Cartesian boundaries to think differently about configurations of space and subjects. Their work includes alternative conceptualizations of the flows, conjunctions, and con vergences of processes in a world rendered increasingly complex through globalization, and a rethinking of con figurations across traditionally conceived boundaries of human/nature, human/animal, human/machine, sub jects/spaces. Longhurst and other feminist geographers, for example, have focused on the affective and leaky qualities of bodies and other spatialized concepts of human subjectivity that belie the neat and bounded categorizations of Cartesian thought, suggesting that an approach open to the porosity and flexibility of subject production might also offer new kinds of alliances.

Others have focused on complexities of landscapes. Bonta's study of Honduras, excerpted in *Deleuze and Geophilosophy*, explores a variety of practices that consti tute its landscapes. He suggests that the point is not to disaggregate complex spaces into their constituent parts – 'peasant and rancher, human from forest' – but rather to understand how spaces and subjects are increasingly 'entangled'. These places do not constitute discretely bounded territories: they jostle one another at the edges, 'beans taking over forest; forest taking over ranch land'. This process of becoming, the nomadism of space and subjects is central to this approach.

In this view of society–space as an immanent field, space is not strictly a techne, nor a juridicopolitical boundary, but the manifestation of process itself, a kind of becoming. Space is thus deterritorialized and reterritor ialized; smoothed – in a process where relationships within a milieu are relatively free form, determined only by one's tactical position in relation to others; or striated – in a process whereby subjects' capacities are invested through particular roles that they carry regardless of their location. This is a Leibnizian rather than Cartesian space: relational, enfolded, understood in terms of interfaces, interconnections, conjunctions, and interpellations. Even bodies themselves are not viewed as discrete bounded objects but rather a coding of the flows of heterogeneous materials, which confounds conventional notions of 'inside' and 'outside'. Indeed one might argue that in this view, the boundary between 'space' and 'subject' has dissolved. All of matter is viewed in terms of its active or expressive capacities, rather than reduced to its repre sentational forms.

In social terms, assemblages channel desire and code activities into particular flows and pathways. These forms – the band, the state form, and the capitalist axiomatic – do not represent discrete historically distinct phases, but rather modus operandi that coexist and overlap. Subjects do not resist societal norms and inequities simply by reacting against structures: they draw on their own inspiration and creativity, ask their own questions, and formulate their own problems to conceive a more equitable world. Not to do so, risks remaining within the reference frame, channels, and codes, of the very struc tures one seeks to transform.

This view of resistance draws upon a concept of power that explicitly rejects Hegel's (and Marx's) resort to a negative dialectic – a dialectic whose movement is achieved exclusively through a reaction against intoler able conditions. It draws instead on Spinoza and Leibniz to rework concepts of power and space. Power is viewed as manifesting in two distinct forms: potestas – an alienating form of power which separates its subjects from their own capacities, and tends to be associated with feelings of melancholia and sadness; and potentia – or an emancipatory power which enhances subject's capacities to act in their best interests, and tends to be associated with feelings of joy. It is this second form of power, which provides the impetus for societal transformation based on affirmation, on the positive (rather than the dialectical movement of the negative). One example of such is the wildcat strike of autoworkers in Turin, Italy in the 1970s. They refused a simple negotiation of better working conditions, rallying around the slogan "*Volle Tutte*" – We Want Everything! Another is the creative, inventive, and constantly evolving political agendas and practices of social movements around the globe in the late 1960s.

In this view, individuals must decide themselves whether a relationship expresses potentia or potestas – no one else can determine this for them. Space and subjects are not thought of solely in terms of their optic qualities – how subjects are organized in space, but their haptic qualities – the space of sensation, emotion, affect, which in turn help to constitute a normative, ethical basis for a politics.

The approach offers a different vantage point from which to understand societal transformations, or social movements (both negative and positive). Subjects can just as equally desire fascism as liberation. For example, in the rise of Nazi Germany the masses wanted to be led, they wanted fascism, a view expressed most strongly by Deleuze and Guattari. The point is to understand the basis of that desire, not simply assert that subjects have somehow been duped. Desiring production is a mech anism through which societies are constituted. The manner in which desire becomes channeled, blocked, or confounded with other objectives is central to this understanding of society–space.

Geographically inspired variants of this approach have concentrated at its antipodes. At one pole they have explored the constitution of the social in the interface between bodies and sensations. Post structuralist feminist works and nonrepresentational geographies, the latter pioneered by Nigel Thrift, have investigated the affective and expressive, rather than representative dimensions of society and space. They begin with the body as a vantage point, and suggest new ways of approaching the political and the transformative, emphasizing the haptic – the sensory – dimensions of everyday life as a starting point from which social change begins. At the other pole, we find a Deleuzian reworking of complexity theory in the geo coding of flows of matter. The most provocative of these, perhaps, is Michael DeLanda's *A Thousand Years of Non Linear History*. Although he is not a geographer, his work provides a provocative intersection of both geo logical and human activities, reinterpreting the last mil lennia through flows of geological, biological, and cultural–linguistic matter.

Arun Saldahna suggests one oversight of these appli cations (works of Bonta and others aside) is that the political questions of exploitation and the restrictive constitution of subjects are often overlooked – politics has been eviscerated, leaving us with a simple celebration of creativity, reformulated narrowly as cultural com modity production, or wondering about the complexity of existence. One might argue, however, these explor ations have attempted to extend the terrain of the pol itical, and indeed there is much to suggest that research on bodywork and other engagements of affect might prove a potent avenue of enquiry. But clearly much work remains to be done on the ways in which flows and codings help to constitute institutions that enshrine state power, advance imperialism, and channel desire into frameworks that justify and perpetuate, rather than transform, social inequities. An investigation of this di mension of Deleuzo–Guattarian space remains firmly on the agenda.

Issues

Before the radical turn in geography some 30 years ago, David Harvey lamented the lack of strategy at the interface in quantitative geography – expressed in the inability of quantitative methods to reconcile social and spatial processes. In working through the four broad approaches outlined about, geographers have been adept at demonstrating the salience of space to social theory and restored its dynamism in the society–space couplet. In each of these approaches, geographers have offered analytical insights and new conceptual orientations toward space as a site of both research and praxis. At times they have built on or sharpened spatial concepts elaborated by theorists working outside the discipline and at times created their own, particularly where the sali ence of space had been ignored.

Space is indeed 'back on the agenda'. New chasms, however, have appeared in our thinking of social–spatial relations. The most pressing concern in contemporary human geography is how one might think of the interface between these competing visions of the subject and its attendant spatialities. It would be difficult to suggest that one of these approaches is more relevant than the others. In a world where the gap between rich and poor is, if anything, accelerating with attendant decline of cities and regions; where new and troubling forms of sur veillance (with the introduction of biotechnologies) are coming into being; where new hybridized subjects are surfacing; and where desire is being mobilized to sustain affective regimes and intensified exploitative produc tion, each of these approaches offers insights into con temporary societal transformations and social inequities.

These varied approaches are often characterized as ir retrievably antagonistic, even in the face of evidence to the contrary in the original writings that they draw upon. Al ternately they are combined, sometimes indiscriminately. Work needs to be done to explore their commonalities and differences more fully – to specify more carefully the complex terrain in which these diverse subjects are pro duced, and the particular configurations of society and space that sustain them. A new strategy at the interface is required, in order to be able to imagine different possi bilities for life and work in a more equitable world.

See also: Feminism/Feminist Geography; Fordism; Fordism, Post-Fordism and Flexible Specialization; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Non-Representational Theory/ Non-Representational Geographies; Poststructuralism/ Poststructuralist Geographies.

Further Reading

Blaut, J. M. (1993). *The Colonizer's Model of the World*. New York: The Guilford Press.
Bonta, M. and Protevi, J. (2004). *Deleuze and Geophilosophy*. Edinburgh: Edinburgh University Press.
Crampton, J. W. and Elden, S. (2007). *Space, Knowledge, Power. Foucault and Geography*. Hampshire: Ashgate.
Crang, M. and Thrift, N. (2000). *Thinking Space*. London: Routledge.
Elden, S. (2001). *Mapping the Present*. London: Continuum.
Fincher, R. and Jacobs, J. (1998). *Cities of Difference*. New York: Guilford Press.
Gregory, D. (1994). *Geographical Imaginations*. Cambridge, MA: Blackwell.
Harvey, D. (1973). *Social Justice and the City*. Baltimore, MD: Johns Hopkins University Press.
Harvey, D. (1996). *Justice, Nature and the Geography of Difference*. Cambridge, MA: Blackwell.
Katz, C. (2006). *Growing up Global Economic Restructuring and Children's Everyday Lives*. Minneapolis, MN: University of Minnesota.
Keith, M. and Pile, S. (1993). *Place and the Politics of Identity*. London: Routledge.
McKittrick, K. (2006). *Demonic Grounds. Black Women and Cartographies of struggle*. Minneapolis, MN: University of Minnesota.
Philo, C. (2004). *A Geographical History of Institutional Provision for the Insane from Medieval Times to the 1860s in England and Wales: The Space Reserved for Insanity*. Wales: Edwin Mellen Press.
Philo, C. and Soderstrom, O. (2004). Social geography: Looking for society in its spaces. In Benko, G. & Strohmayer, U. (eds.) *Human Geography. A History for the Twenty First Century*, pp 105 138. London: Edward Arnold.
Pile, S. and Keith, M. (1997). *Geographies of Resistance*. London: Routledge.
Rose, G. (1993). *Feminism and Geography. The Limits of Geographical Knowledge*. Minneapolis, MN: University of Minnesota.
Soja, E. (1989). *Postmodern Geographies*. London: Verso.
Thrift, N. (1996). *Spatial Formations*. London: Sage.
Woods, C. (1998). *Development Arrested*. London: Verso.

Soja, E.

A. Latham, University College London, London, UK

Glossary

Exopolis A generally low density, sprawling, suburban-like form of urban settlement, often without clear structures of public governance, that has become an increasingly distinctive element of much contemporary urbanization. California's Orange County is a paradigmatic example of the exopolis.
Postmetropolis The contemporary American metropolis. It refers to how the present postmodern epoch has transformed the dynamics of contemporary urbanization. In essence, the term is simply a contraction of the term postmodern metropolis.
Simcities Cities or parts of cities, which are built around the simulation of other, often fantasy, times, and/or places.
Sociospatial Dialectic The dialectical relationship between space, society, and history.
Synekism 'The stimulus of urban agglomeration' according to Soja (2003). Synekism refers to the generative properties of urban density; the way in which the presence of large, diverse populations creates a whole range of dynamic and often unexpected relationships not possible in smaller, less diverse populations.
Thirdspace Refers simultaneously to a style of analysis that places a spatialized trialectics at its analytical center, and a conceptual space that captures the full material and imaginative texture of people's everyday experience of the world.
Trialectics An extension of the established Marxian dialectic to include space as well as time. Trialectics is essentially a reworking of Edward Soja's earlier term sociospatial dialectic.

Soja, Ed (1941–)

Urban geographer, social theorist, and key member of the so called 'Los Angeles School' of urban studies, Ed Soja has been a Professor at the UCLA department of urban planning since 1972, and has been a Visiting Centennial Professor at the London School of Economics since 1999. Soja became internationally renowned for developing a postmodern human geography that offered both a radical critique of mainstream social science, and a thoroughgoing consideration of the foundations needed for a genuinely spatialized critical social theory. His central argument has been that critical social theory, like mainstream social science, has failed to take space seriously and this has profoundly undermined its ability to understand the dynamics of the contemporary capitalist system. To begin to take space and spatiality seriously, he argued, required the rebuilding of critical social theory from the ground up. Thus, in a trilogy of books, *Postmodern Geographies: The Reassertion of Space in Critical Social Theory* (1989), *Thirdspace: Journeys to Los Angeles and Other Real and Imagined Places* (1996), and *Postmetropolis: Critical Studies of Cities and Regions* (2000), Soja undertook that very task. In these books, and a parallel series of articles, Soja presents an intellectual archeology of the origins of critical social theory's repudiation of the spatial, highlights a range of alternatives – and spatially attuned traditions – of critical social theorizing that offer ways of radically spatializing critical social theory's conceptual imagination, and lastly, provides a series of exemplary case studies focusing on contemporary Los Angeles to show what a properly spatialized – postmodern – critical social theory might look like. (**Figure 1**)

Born and raised in the Bronx, New York City, Soja began his academic career in the 1960s as a specialist on African development. In 1972, after a period of teaching at Northwestern University following the completion of his doctoral studies at Syracuse University, Soja moved to The Graduate School of Architecture and Urban Planning at the University of California, Los Angeles. This

Figure 1 Soja, E.

was to mark a major shift in the intellectual and geographical focus of Soja's work. While his early publications on Africa had been broadly sympathetic to spatial science and the then dominant paradigm of modernization theory, through the 1970s and early 1980s, Soja increasingly distanced himself from these traditions. At the same time, the focus of his work shifted from Africa to Los Angeles. In a series of essays that culminated in the publication of his most highly regarded and cited book *Postmodern Geographies: The Reassertion of Space in Critical Social Theory*, Soja undertook an intellectual archeology of critical social theory's neglect of the spatial. How was it, Soja asked, that within this otherwise compelling intellectual tradition that space had come to be "treated as the dead, the fixed, the undialectical, the immobile [...while] time on the contrary was richness, fecundity, life, dialectic?" (Michel Foucault, cited in Soja, 1989: 10). Simultaneous with this archeology, Soja also sought to bring to attention an emergent countercurrent within the critical social theoretic tradition that repudiated this 'historism' and instead reached toward a profoundly spatialized mode of thinking based on a triple dialectic of 'space, time, and social being'.

With their direct engagement with social theory, and their attempt to relocate human geography at the center of contemporary social theory, these essays, and indeed *Postmodern Geographies* itself, were closely aligned to the work of Marxist human geographers like David Harvey, Richard Peet, Neil Smith, and Doreen Massey. But Soja's aim was not simply to bring human geography to the center of the western Marxist tradition: it was to reshape critical social theory. Thus, besides drawing on Marxist geographers, Soja assembled a diverse collection of writers – some Marxist, some not – that ran from Henri Lefebvre, Ernst Mandel, Jean Paul Sartre, and Louis Althusser through to Anthony Giddens, Michel Foucault, John Berger, Marshall Berman, Fredric Jameson, and Martin Heidegger all of whom (he argued) had in some way sought to place spatiality at the center of their intellectual projects.

Of these figures the most influential on Soja was Henri Lefebvre. A Marxist French philosopher, Lefebvre had since the 1960s been arguing that the production of spatiality rather than history had become the central armature of capitalist development and contradiction. Lefebvre's thesis became the central logical armature to *Postmodern Geographies*, as it is to Soja's whole postmodern project. Following on from Lefebvre, Soja argues that both the theoretical project of postmodernism and the concept of postmodernity understood as a distinct historical epoch are defined by the centrality that space and spatiality play within them. The spatially oriented definition of the postmodern and postmodernism set out by Soja in *Postmodern Geographies* was a distinctive, if not to say, highly idiosyncratic reading of the postmodern. Indeed, while most postmodern theory was defined by what the philosopher Jean François Lyotard famously termed an 'incredulity towards meta narratives' – in particular, those of Marxism and historical materialism – Soja tied his postmodernism tightly to both of these traditions. For him, his postmodern geographies were truly postmodern because they recognized the salience of Lefebvre's claim that it is 'space more than time that hides things from us'. If the modern era and modernism are defined by their fixation with time and the temporal, postmodernity and postmodernism are defined by the importance of space and the spatial. Hence, Soja's account retained at its center a notion of a coherent, structured, historical progression. Thus, following the Belgian Marxist economic historian Mandel, Soja suggested that the current postmodern epoch should be read as but the latest – if highly distinctive – manifestation of a series of long waves of capitalist development. Simultaneously, the aim of a postmodern critical social theory is to make sense of the emergence of the distinctive capitalist restructurings that have brought this postmodern epoch into being. In short, Soja provided a 'very' modern version of postmodernism.

Postmodern Geographies was a hugely influential book. It presented one of the first and most compelling responses by a human geographer to the challenge of postmodernism. It made an impressive case for the need to put space at the center of critical social theory. But if the breadth of Soja's scholarship was widely admired within human geography and, significantly, the broader social science community, the vision of a postmodern geography set out in *Postmodern Geographies* also met with widespread and frequently vociferous criticism within the discipline. Some critics argued that the basic premise of *Postmodern Geographies* was wrong. The world might be changing in rapid and unexpected ways but this, they argued, was no reason to ditch established modes of analysis in favor of some kind of unproven postmodernism. Others, more sympathetic to postmodernism, wondered about Soja's style of postmodernism. A key attraction of postmodernism was its democratic, eclectic, and inclusive ethos. The best postmodern writing sought to break down established barriers between popular and high culture, open up spaces within which previously marginalized groups might be heard, and possessed a healthy irreverence toward established schools of knowledge. The postmodernism of *Postmodern Geographies* appeared to be strangely suspicious of this pluralistic spirit. Still other critics wondered about the way *Postmodern Geographies*, despite its claims to be profoundly concerned with everyday spatialities, constantly reduced the everyday to little more than an epiphenomenon of underlying economic processes. This is most obvious in the final two chapters in *Postmodern Geographies* that focused primarily on Los Angeles. As Derek Gregory complained in an in depth review of *Postmodern Geographies*, while stating his intention to show how the experience of space and time in the metropolis is being transformed, Soja actually offered the reader an account of the city 'rarely disturbed by human forms'. That said, related accusations were made of David Harvey's *Condition of Postmodernity*: together, both served as key reference points in debates surrounding geography's future trajectories as the discip line approached the millennium.

Thirdspace: Journeys to Los Angeles and Other Real and Imagined Places (1996) and *Postmetropolis: Critical Studies of Cities and Regions* (2000), the two books that Soja authored after the publication of *Postmodern Geographies,* offer both a response to these critiques and at the same time further elaborate Soja's vision of a radically spatialized, post modern analysis. In both books, Soja introduced a number of new concepts meant to help clarify and crystallize the key coordinates of his project. For example, the 'socio spatial dialectic' first introduced in an *Annals of the Asso ciation of American Geographers* paper in 1980, and developed in Chapter 3 of *Postmodern Geographies,* was refined into the notion of a 'trialectics of being' – the insight that the ontology of being can only be interpreted through the interlocked terms 'spatiality, historicality, and sociality'. In a similar way, he used the term 'thirdspace' to refer to a style of analysis that placed a spatialized trialectics at its center as well as a particular texture of everyday lived space that exceeds the compartmentalized knowledges of the conventional social sciences. Furthermore, Soja was keen to demonstrate the openness of his postmodern/ thirdspace project. So while in both *Thirdspace* and *Post metropolis* Lefebvre retained his privileged position, a new range of postcolonial and feminist writers like bell hooks, Homi Bhabha, Edward Said, Gayatri Spivak, Gloria Anzaldúa, and Gillian Rose were added to – and were absorbed into – Soja's intellectual project.

Yet, for all the opening of Soja's theoretical horizons to postcolonial and feminist writers, huge swathes of the social sciences were simply dismissed for their narrow focus on what are dubbed firstspace ('the 'real' material world') or secondspace ('imagined') representations of spatiality perspectives. For Soja, thirdspace is the privil eged space of analysis, despite the fact for all the hun dreds of pages of *Thirdspace* and *Postmetropolis,* it remains a uniquely slippery and flexible term. In Soja's own words:

> Everything comes together in thirdspace: subjectivity and objectivity, the abstract and the concrete, the real and the imagined, the knowable and the unimaginable, the re petitive and the differential, structure and agency, mind and body, consciousness and unconsciousness, the dis ciplined and the transdisciplinary, everyday life and unending history (original emphasis). (Soja, 1996: 56 7)

Thirdspace is accordingly claimed to encompass everything there is to say about anything. Hence, in his attempt to continue fashioning a radically new style of thinking, and his keenness to stress just how new his 'radical postmodernism' was, Soja often de anchored himself from any established intellectual tradition. This compelled him to work at a level of abstraction that, in some eyes, undermined the most productive and inter esting elements of his argument.

Indeed, while much of the critical reception of both *Thirdspace* and *Postmetropolis* focussed on Soja's theoretical arguments, the most compelling sections of both books are where Soja turns to describing the contemporary metropolitan condition. In these sections, the majority of which focus on *Los Angeles*, Soja presents a series of routes through what he calls the contemporary postmetropolis: the new spaces of industrialization and economic ex ploitation ('the postfordist industrial metropolis'); the city turned inside out ('the exopolis'); the city reshaped by globalization ('the cosmopolis'); the city of multiple and divided sociality ('the fractal city'); the privatized and sanitized city ('the carceral archipelago'); and spaces of urban simulation ('simcities'). Taken together, and com bined with Soja's attempt in the opening chapters of *Postmetropolis* to outline a geohistorical account of the importance of urban agglomeration (what he calls 'synekism') to all social development, these accounts offer some of the most comprehensive and daring at tempts to summarize and interpret what is happening to the western cityscape over the past two decades.

See also: Gregory, D.; Harvey, D.; Historical-Geographical Materialism; Lusophone Geography; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Massey, D.; Postmodernism/Postmodern Geography; Scott, A.

Further Reading

Scott, A. and Soja, E. (eds.) (1996). *The City: Los Angeles and Urban Theory at the End of the Twentieth Century*. Berkeley, CA: University of California.

Soja, E. W. (1980). The socio spatial dialectic. *Annals of the Association of American Geographers* 70, 207 225.

Soja, E. W. (1985). The spatiality of social life: Towards a transformative retheorization. In Gregory, D. & Urry, J. (eds.) *Social Relations and Spatial Structures*, pp 90 127. London: Macmillan.

Soja, E. W. (1986). Taking Los Angeles apart: Some fragments of a critical human geography. *Environment and Planning D: Society and Space* 4, 255 272.

Soja, E. W. (1987). The postmodernization of human geography: A review essay. *Annals of the Association of American Geographers* 77, 289 296.

Soja, E. W. (1989). *Postmodern Geographies: The Reassertion of Space in Critical Social Theory*. London: Verso.

Soja, E. W. (1996). *Thirdspace: Journeys to Los Angeles and Other Real and Imagined Places*. Oxford: Blackwell.

Soja, E. W. (2000). *Postmetropolis: Critical Studies of Cities and Regions*. Oxford: Blackwell.

Soja, E. W (2003). Writing the city spatially. *City* 7(3), 269 280.

Sound and Music

M. Duffy, University of Melbourne, Carlton, VIC, Australia

Glossary

Embodiment This refers to the bodily aspects of human subjectivity and is the central theme of European phenomenology. Recent cultural geographic research has drawn on developments in social theory, in particular that of psychoanalysis and the work of Henri Lefebvre and Michel Foucault.

Music Although the definition of music has been challenged and expanded by composers, generally it is defined as an art form constructed from certain patterns of sounds and silences, and associated elements such as melody, rhythm, dynamics, articulation, and texture.

Noise Any undesirable or unwanted sound is defined as noise. How sound is experienced as noise physiologically is dependent both on the intensity and the frequency of the tones. However, defining what is undesirable or unwanted also arises out of particular cultural contexts and power relations. Moreover, contemporary composers and sound artists have used ideas of noise to rethink ideas about the definition of what constitutes music.

Rhythm At a very simple level this means regular repetition. It can mean regularity, lawfulness, and measure, but it is also an expressive or compelling motion, gesture, or shape. Rhythm is an important means of shaping the processes of communication. The interaction we have with each other is rhythmic and interpreted as expressive and meaningful because it occurs within a background pulse as well as recognized cultural codes.

Sound It is a disturbance of mechanical energy that propagates through matter as a wave. It is further characterized by the generic properties of waves, that is, frequency, wavelength, period, amplitude, speed, and direction. Sound is perceived through hearing as well as through vibrations felt by the body.

The Difficult Nature of Sound (and Music)

Although physical properties of sound, such as pitch, rhythm, timbre, and tempo, can be measured in precise ways, such measurements tell us little about the ways in which sound is embedded within and articulates ideas circulating through society. Sound and music are not simply physiological outcomes of hearing sonic qualities arising within the environment. As John Cage's composition, *4′33″*, demonstrates, what often goes undetected in everyday life are the ways in which certain acoustic properties are accorded particular and ideologically loaded meanings. For example, sound becomes music when placed within an appropriate context, be that the physical space of a concert hall or the marking out of a temporal frame as performance. Further, as Jacques Attali argues in his seminal text, *Noise: The Political Economy of Music* of 1992, the ways in which sound is defined as music or noise is an expression of power, one where the appropriation and control of sound reflects and affirms the sociopolitical structures of a community.

The effects and impact of sound can be difficult to capture. Music in particular is generally described – or represented – as something that has happened in the past, although there are exceptions to this in research undertaken outside of geography, such as in the work of sociologist Tia De Nora and ethnographer Stacy Holman Jones. Sound and music unfold in time and place, and methods need to take into account this immediacy as well as sound's fleeting nature. Nor do the ears alone hear sound; our bodies are in fact sounding boards that pick up the vibrations so that we are enveloped by sound. This immersion has important implications in geographical study. While sight distances us from the world, sound makes us aware of our connectivity. These qualities mean that sound, and music in particular, appears to speak directly, to communicate ideas that nonetheless remain ambiguous and open to multiple interpretations.

Geographers need, then, to consider methods appropriate to sound's multisensory qualities, the numerous processes in which sound constitutes and is constituted by the everyday (and the extraordinary), as well as strategies for (re)presenting such research in meaningful ways.

Further to this focus on the acoustic and sonic qualities of place, sound and music can be very useful methods in themselves in thinking about research frameworks, as well as in the teaching and learning of geographical concepts and practices. This is exemplified in recent work in human geography which has deployed the notions of 'rhythm' and 'scale' as a means to develop conceptual frameworks that contextualize various social interactions operating at spatial and scalar relationships.

Sound Methods

The majority of geography's earliest studies of sound and music draw on the discipline's more traditional methods of fieldwork analysis. In particular, music is conceptualized as an object – a score, a performance, and a sonic environment – and hence a text that can then be read and deciphered. Although less engaged with the specific qualities and unique aspects of the sonic in the world, this methodology does uncover significant detail about geo graphic and social relations.

Cartographies: Diffusion

Music is linked to place in very specific ways. Music can be characterized as a discrete object, with definable fea tures that mark it as representative of a clearly demar cated and spatialized community or group. These features then are used as representative of complex entities such as nation, or as groups defined by ethnicity, or cultural and subcultural practices, gender, and sexuality.

The mapping of sound and musical styles or genres can therefore be used to demonstrate the migration of groups across space, or the movement of cultural practices and knowledge through a mapping of the music itself, dis seminated via technology or performance tours. These mapping methods also uncover the ways in which sound and music are contained and controlled. This is important as these mappings of boundaries – how sound is defined and located, who defines sound as ‘noise’, ‘music’, or ‘si lence’ – are politically motivated expressions of power.

Much early research, particularly that of George O. Carney in North America since the 1970s, has focused on the diffusion of certain musical forms across geographical locations. This work itself drew on early studies in an thropology and ethnomusicology, and was based on some what essentialist assumptions of identity. In this approach, certain sound characteristics are defined as representative of specific group identities. For example, musical styles or genres such as folk music, gospel, and blues are associated with particular racial, national, or class groups. These early geographic studies, typified by Alan Lomax’s cantiometrics project of 1976, mapped the movement of music genres and styles as they traveled out from a designated origin. Such a methodological framework and analysis produces detailed local studies with an emphasis on place bound theories around identity and culture.

However, this fixing of identity through certain essentialized characteristics is problematic, as it dis regards the fluid and hybrid natures of the concepts of identity, place, and the cultural product of music itself. A further important critique of this focus is that it fails to recognize the social and cultural contexts in which music is produced and consumed. Even within a defined group, there will be different expressions of culture and identity, and such differences will be subject to hegemonic pro cesses of control. Finally, this framework constructs space and place as simply sites in which sound/music occurs or through which it diffuses. Sound is an active agent in the formation and creation of place such that space and time are combined in ‘becoming’ and bring into being that space as it unfolds.

Cartographies: Cultural ‘Distinctiveness’ and Place

Many researchers have noted the centrality of geography in determining the relationship between sound and place. This cartographic emphasis is particularly strong in studies of popular music, where music styles are de scribed in ways that intimately link musical activities to specific local and regional places, as well as in the con struction of local musical scenes. Studies of late 1990s youth cultures in the UK, such as the relationship be tween Manchester and rave culture, or Blackburn’s warehouse parties, examine the ways in which naming of specific places such as buildings and streets creates these imagined geographies of music culture. As such work demonstrates, the material elements associated with these activities are significant to creating localized feelings and experiences, further reinforcing the connections between music scene and local place.

Strategies of economic and social regeneration also draw on the distinctiveness of place as created through particular forms of music and music practices. For ex ample, Chris Gibson and John Connell in their recent publication, *Music and Tourism*, point out that the rise of niche tourism within the broader tourism industry has led to the development of markets with a specific music orientation. Music festivals and other music related events are used to promote designated places with the aim of increasing social and economic benefits. Methods investigating sound and music in this context require an unpacking of the complex processes involved – issues of culture, such as notions of authenticity, heritage and es sentialism; economic considerations of investment, em ployment, and the stimulus of local, regional, and even international economies; as well as the significance of recreation, enjoyment, and nostalgia to individual and community well being.

Imagery: Lyrics, Narratives, and Language

The consumption of music, that is, how listeners imagine place, has been particularly important in popular music research, and the methodologies of linguistics, semiotics, and musicology have influenced and shaped similar re search in geographic inquiry. This type of study primarily examines how listeners imagine place, and such an ap proach generates very detailed and rich material on the evocation of place through text. In this way, it is an obvious means of expressing localized, geographic connections.

However, song lyrics are not isolated from social practices, but embedded within a sound structure with its own set of semiotic codings. A closer reading of song is demonstrated in the work of Lily Kong, and her study of popular music in Singapore. Rather than simply searching song lyrics for geographical references, she demonstrates how the place of Singapore is conveyed also through the use of the rhythms and instrumentation of the various different ethnic and cultural groups resident in Singapore, as well as through the juxtaposition of their languages (Malay, English, and Chinese) and the distinct form of English used in Singapore, 'Singlish'. These aural sounds of place accompany a lyric content that tells of a distinctly Singaporean society, such as the city's cuisine and its multiracial character.

Kong's studies demonstrate that it is equally important to examine extramusical elements, those activities and processes in which the music and the music scene are embedded. Visual culture, for example, album covers and promotional material, is important in helping mark out the spaces of these scenes. The materiality of the places in which these scenes occur locate and anchor the scene to specific locations. In this way, performance venues are an integral part of the music scene's geographic imaginary, while the existence of local record labels, independent record stores, an active local press, and local radio promote and reinscribe the connections between the music scene and place. Development of technology, particularly the increase in internet use and the employment of web based sources such as YouTube, has significant implications for the geographic study of sound. The music–place relationship is not simply one of text and context. As research into song shows, sound's intertextuality is an important means for understanding its significance to place making. This has important implications for understanding the political power of sound/music.

Cartographies: Dissemination

Much of the entanglement between the sounds of music scenes and their material constitution can be approached using existing ethnographic methods (see also below). However, the meanings assigned to music scenes and their components are significantly part of larger networks – of youth cultures, of the music industry. Local and regional sounds therefore may be captured, marketed, and disseminated well beyond that of the originating geographic location, but in ways that retain, and even emphasize, the scene's geographic origins. This, in turn, may impact on other parts of the network, with new cultural alliances forming, but which themselves are embedded within localized forms of cultural production. For example, studies of rap, a style originating in the Bronx and Los Angeles, have demonstrated the importance of this music as a means for repressed groups to challenge dominant ideologies of contemporary society. Strongly associated with its origins in inner city 'ghettoes' and black struggles for recognition of equality, rap and hip hop have become important forms with which to express resistance to marginalization in many Western societies. Yet each geographic location gives rise to variations that make this genre 'local', so that we can now discern French rap, say, from New Zealand or Italian hip hop.

Recent research has drawn on cartographic methods, such as geographic information system (GIS), as a means to capture the nodes and networks of cultural processes, such as the production, reception, and consumption of music cultures. This approach is now used in studies of the creative industries as a means to map out the connections between people, place, industry, and leisure so as to understand and better guide urban development.

Ethnographies

Ethnographic studies are an important means of teasing out the ways in which sound and music are embedded within a wide ranging set of relations – social, sensual, symbolic, and economic. This methodology draws on anthropological methods of observation and participant observation in order to examine how sounds, images, events, and artifacts circulate within communities and are given meaning within and through social practice. In this framework, sound and particularly music are conceptualized as active ingredients in social formation and therefore play a significant role in establishing, maintaining, and transforming social relations.

Social relationships arising within sound/music practices enable (or prevent) interactions between people. For example, relations of kinship, friendship, and fandom can bind people to place physically, conceptually, and emotionally, through their collective involvement in communication, exchange, reciprocity, and investment in activities and associated places of music making. These various interactions themselves operate within a range of spatial relations: local power inequalities, local and regional relations, as well as those interactions at local, national, and international levels.

The detailed material generated out of ethnographic methods (anthropologist Clifford Geertz's 'thick description') demonstrates that there is no simple correspondence between sound/music, place, and identity, although, as anthropologist Sara Cohen points out, different places do make different noises. Rather, these studies reveal the contested nature of place making. Sound and music are cultural materials that can be used to create or reinforce place specificity. Music notably has the ability to affect and articulate mood, atmosphere, and emotion, and to trigger the imagination. Sound, music, and music making not only reflect but also produce notions of place, hence influencing people's experience of place. The music–place relationship is therefore a powerful source of belonging or alienation.

Ethnographic methods can, however, be somewhat limiting. They usually involve small scale research populations and face to face interaction, so generating very detailed and profuse amounts of material that then requires appropriate management. The size of the research population raises issues around representative sampling techniques – can the analysis derived from this group be applied to others? Presentation of this very descriptive data also presents its own challenges. The detailed material collected needs to be incorporated in ways that enrich the study and avoids banality. Too much concern with local details can also be detrimental, as we may ignore the ways in which this particular sense of the local is part of wider networks, activities, and places.

Interviews and Surveys

In many ways the interview process is not separate from ethnographic methodology. However, interviews are an important means of accessing more in depth material; in particular they may provide different perspectives on sound in its social context. The type of interview conducted – in depth, semistructured, and formal, or informal – is also an important consideration, as is the choice of participants targeted and the number of interviews undertaken. For example, do you want more qualitative or quantitative data? Is the project examining audience responses to music performance, or is the focus on the ways in which a music scene is shaped by local council bylaws?

The timing of interviews may also be crucial. More generally, when conducting research into music related activities, interviews with concert organizers, promoters, and audience members, carried out at the performance space before, during, and after musical events, can provide information on not just about how performances are set up and/or staged in particular ways to appeal to certain audiences, but they can also tell us something about the manipulation, mutation, and management of material space within a musical event. For some projects, it is important to collect a more raw response, for example, by carrying out short, 'on the spot' interviews conducted during or immediately after the performance. While perhaps less articulate, using this method may give the researcher access to the affective responses of those interviewed, in comparison to the more considered responses of formal interviews.

Quite different data are generated through survey methods, which nonetheless yield very detailed information for a range of parameters, but this does require a significant size of population for meaningful analysis.

Economies of Sound

Although sound and music can be framed as having very specific geographical links, analysis of the ways sound is produced and consumed raises different and often much more complex sets of spatial relations. Moreover, in projects involved in the study of the music industry, researchers need to keep in mind the tension between music conceptualized as an ephemeral even nomadic activity, with industry's attempts to construct a product, something that is static, fixed, immobile, and available for purchase. A focus on music as part of the cultural economy explores the ways music is not only bound up in industrial and commercial enterprises and shifting of capital, but can also help us better understand the networks and connections between the 'economic' and the 'social'.

The marketing of place through sound has become an important strategy for economic regeneration. Many places have used the music festival structure as a means to brand a local community in the hopes of generating income through ticket sales and associated activities. The recent focus on the creative classes generated by Richard Florida's *The Rise of the Creative Classes* in 2002 has also demonstrated the significance of music in creating and developing social capital within specific spatialized ways that then provide economic benefits for local communities. Understanding the importance of how certain spatial relations operate in these contexts requires examining such things as the pattern of employment in the music industry, as well as the regulatory and institutional settings within which music is created, developed, and promoted.

Further to this, one needs to be mindful of the socially constructed linkages between people and industries. These networks of trade are also networks of influence and knowledge transfer, which are not themselves static but negotiated, renewed, severed, or generated in new ways. Sound, and particularly music, is significant in these economic/social networks; its ability to rapidly transform place is because of its spatial mobility. Therefore, these networks are open to negative influences such as exploitation, the erosion of local differences as a result of globalization, cultural imperialism, and the loss of intellectual property. On the other hand, examination of the economies of sound can also demonstrate the ways in which trade practices are opened up, or how interaction between culturally and spatially diverse groups may give rise to new, hybridized form of cultural practice.

Experiencing Sound

A considerable amount of cultural geographic focus on sound has been on music and its significance to constructions of place and the associated concepts of identity and belonging. However, sound is also understood as an important signifier of nature and particular types of environment, such as that of the urban or rural. Soundscape studies are then, too, an examination of power relations, as certain sounds are defined in terms of nature, the unnatural, the built environment, and well being.

One aspect of this type of research is the analysis of the environment for 'appropriate' sounds, and their regulation and control. The measurement of sound levels is the chief methodology used here. Sound levels are measured in units of decibels (dB), with normal conversation measured at around 65 dB, while at 125 dB humans feel pain. Although not necessarily related to such measurements, noise is defined as any undesirable sound. However, the disturbing effects of noise do depend on both the intensity and the frequency of its sound. Definitions of sound as 'noise' or 'music', and associated ideas of such sound as 'unhealthy' or 'beneficial', have important implications in the construction and development of place through zoning laws and public health acts, as well as building design. Analyses of the debates that define sounds as acceptable or not – which sounds can or should be present in the public realm? Is nature quiet? What sounds are of benefit to community well being? – point to contested values within the community. As David Matless suggests, these evaluations are exercises in mapping a moral geography.

While the use of sound and music to construct identity is a common strategy for community representation, this obscures the ways in which music shapes and influences our actions in, and experiences of, place. Identities are not simply constructed through musical forms, rather subjectivities arise and are constituted within the time–space of the event. The study of sound and music in geographic inquiry, therefore, is a significant factor in understanding spatial relations that arise out of sensual, affective, and embodied modes of being.

The methodologies used for this type of research draw our attention to the acts of sound making, and the ways in which spaces and people are brought into being through engagement with sound. Susan J. Smith has done significant work in this area. This approach emphasizes immediacy and presence in the relationship of music, place, and sociality. Place and social relations are constituted through sound in the very acts of people engaging with the event of the sound's performance. This framework moves away from assumptions about musical communication as unidirectional and unmediated, and conceptualizes music as a dialogic process in which various subject positions arise within the sound space.

Recent, exciting work in cultural geography has experimented with a range of techniques that draws on nonrepresentational practices of place making, such as performance as method. This concern with the experiential can uncover unexpected and hidden, but nonetheless important, processes at work in the everyday.

The Body and Space

An experiential approach asks us to think less about the body's cultural meaning and to focus more on its physiological responses in place. In this context, the body's materiality – its movements in space, its relation to other bodies and objects, and its soundings – is about the becoming of spaces of musical possibility, what Nigel Thrift has called 'as if' spaces.

The body's response to music in the form of dance has provided a creative means to consider embodiment and affect. Dance, as Thrift argues, is an example of the expressive and playful elements of embodiment, yet these very characteristics have made it a difficult form to address in the social sciences. The dancing body can be conceptualized as an important strategy of eluding power, and demonstrates different experiential frameworks that are beyond that of the spoken or written word. Through studying the corporeality of the experience of sound – the impact of rhythm, timbre, and melody on the body – we can examine how the physical practice of making sound shapes esthetic forms and their social spaces. Capturing and reflecting on these processes have, in turn, led to inventive analysis practices, such as McCormack's depiction of the body's responses and actions in (sonic) space through nonrepresentational diagrammatic practices.

Listening

A further important strategy is the practice of listening, which highlights our involvement with sound as a participatory activity. Listening invites response and the possibility of dialog. This focus can reveal how emotional engagement with sound constitutes spaces of sociality, how the individual connects to the broader community, as well as the potential this has to giving rise to an intensification of relations in place. Therefore, listening, as opposed to seeing, focuses on the ways we are connected rather than how we differ.

Future Directions in Research

Very creative methodological approaches in cultural geography have come out of art practices, particularly the work of sound artists, along with increased availability of recording technology to a mass market. Sound walks – in which participants move through particular spaces guided by portable soundtracks that tell of stories, memories, and other forms of narrative about that place – offer an experiential mode of engaging with place, a means of placing the listener almost literally in someone else's shoes.

These methods also emphasize that we listen carefully to place. As Michael Bull and Toby Butler demonstrate in their explorations of urban spaces through such portable sound technology, this refocusing of attention enables an embodied, sensory, and more dynamic experience of place, and offers an important means of accessing someone else's relationship to place.

Methodologies such as these open up discussion on a range of issues germane to the heart of geographic research – our relationship to place – and the complex manifestations of this through concepts of physical movement, subjectivity, representation, experience, mem ory, performativity, intercultural relations, and spatial unconsciousness.

See also: Affect; Anthropology and Human Geography; Belonging; Creativity; Cross-Cultural Research; Cultural Economy; Cultural Geography; Cultural Studies and Human Geography; Diaries (Video, Audio or Written); Embodied Knowing; Emotional Geographies; Ethnography; Festival and Spectacle; Fieldwork; Heritage and Culture; Identity Politics; Interviews: In-Depth, Semi-Structured; Non-Representational Theory/Non-Representational Geographies; Participant Observation; Performance, Research as; Performative and Embodied Mapping; Performativity; Popular Culture; Post-Phenomenology/Post-Phenomenological Geographies; Poststructuralism/Poststructuralist Geographies; Representation, Politics of; Sense of Place; Sensorium; Social Movements; Soundscapes; Subjectivity; Time Geography; Tourism; Tourism, Urban; Urban Design; Wellbeing; Youth/Youth Cultures.

Further Reading

Bull, M. (2000). *Sounding Out the City: Personal Stereos and the Management of Everyday Life*. Berg: Oxford.
Carney, G. (1997). *The Sounds of People and Places: Readings in Geography of American Folk and Popular Music*. Lanham, MD: University Press of America.
Connell, J. and Gibson, C. (2003). *Sound Tracks: Popular Music, Identity and Place*. London: Routledge.
DeNora, T. (2000). *Music in Everyday Life*. Cambridge: Cambridge University Press.
Duffy, M., Waitt., G. and Gibson, C. (2007) Get into the groove: The role of sound in generating a sense of belonging at street parades. *Altitude* online at www.altitude21c.com (last accessed 15 May 2008).
Gibson, C. and Connell, J. (2005). *Music and Tourism: On the Road Again*. Clevedon: Channel View Publications.
Kong, L. (1996). Popular music in Singapore: Exploring local cultures, global resources and regional identities. *Environment and Planning D: Society and Space* 14, 273 292.
Leyshon, A., Matless, D. and Revill, G. (1998). *The Place of Music*. New York: The Guilford Press.
Matless, D. (2005). Sonic geography in a nature region. *Social and Cultural Geography* 6(5), 745 766.
McCormack, D. (2002). A paper with an interest in rhythm. *Geoforum* 33, 469 485.
Revill, G. (2004). Authenticity and nonrepresentational theory cultural geographies in practice: Performing French folk music: Dance. *Cultural Geographies* 11, 199 209.
Rodaway, P. (1994). *Sensuous Geographies: Body, Sense and Place*. London: Routledge.
Smith, S. J. (2000). Performing the (sound) world. *Environment and Planning D: Society and Space* 18, 615 637.
Smith, S. J. (1994). Soundscape. *Area* 26, 232 240.
Wood, N., Duffy, M. and Smith, S. J. (2007). The art of doing (geographies of) music. *Environment and Planning D: Society and Space* 25(5), 867 889.

Relevant Websites

www.sounddesign.unimelb.edu.au
Australian Sound Design Project.
www.naisca.ca
New Adventures in Sound Art Sound Travels.
www.ascc.gov.au
Noise control.
www.reverberant.com
Noise control.
http://interact.uoregon.edu
World Forum of Acoustic Ecology.
www.sfu.ca
World Soundscape Project.

Soundscapes

A. Saldanha, University of Minnesota, Minneapolis, MN, USA

Glossary

Crowd Concentration of bodies in space, especially in cities, and its affective impacts.
Diffusion Transformation of cultural traits through their spreading in space.
Embodiment The given that people maintain sensuous and particular relationships to their environment and each other through physiology and movements.
Music Exceptionally and intentionally organized sounds with emotional significance for a particular group.
Phenomenology Philosophical analysis of the meaning to a person of things, other people, time, and space.
Representation Meanings arising through culturally regulated reference.
Soundscape A space–time cohered by particular sounds; the representation of such a space.

'Soundscape' is a crucial term to conceive how sound gives meaning to spaces and places. By filling not only time but space, sounds such as music help orientate the lives of all humans, as much as sight does. However, to be sensed, sound has to be embodied. Because bodies differ and because bodies cohere in crowds, embodiment introduces social differences into soundscapes. Given geography's emphasis on vision, the geographic study of local soundscapes is embyronic. Mostly, geographers have studied the representation and international diffusion of music. Nonetheless, the embodiment of soundscapes invites fruitful exchange with cognate disciplines.

Sight, Sound, and Music

The European tradition of geography has always placed central importance on the faculty of sight, which made the planet's surface legible through navigation instruments, maps, satellite imaging, and remote sensing. Visual depictions of exotic species and landscapes have also contributed to geography's quest to visualize the Earth, as have photography and cinema. Writing itself, whether administrative or literary, is mostly a visual medium. All these media interlock to form a global visibility that has been crucial not only for academics, but also for governments, artists, armies, planners, companies, consumers, and nongovernmental organizations (NGOs). It is impossible to deny, therefore, that Western culture's emphasis on vision has been immensely productive and useful. The question is whether visual technologies do not actively block out the other kinds of information that surroundings contain. The marginal status of the geography of sound, smell, touch, and the other senses would strongly suggest so.

Sound is an obviously spatial as well as temporal phenomenon. It necessarily occurs within a place, and gives meaning to that place according to the recognizable cultural associations it provokes. Almost all sounds in our lives are what in semiotics is called 'indexical': their meanings are quickly understood by most humans because they are largely given by their obvious physical origin. The bark or smell of a dog signifies 'dog', a deep rumble signifies 'airplane'. However, some sounds are intentionally organized in extremely particular ways, such that their meaning becomes much more than indexical and highly culturally specific. These sounds are called 'music'. Musical sounds are per definition capable of producing feelings that are central to a sense of identity. Music has been studied by academics for a couple of centuries, of course, but almost exclusively as a temporal, formal, and disembodied process. Only since the mid 1990s have a handful of scholars started paying consistent attention to the spatial aspects of music and other sounds. Though the work remains scattered, an interdisciplinary framework for soundscape studies is emerging in which geography plays an integral role.

There is a larger epistemological stake involved. Although the production and circulation of knowledge are almost inconceivable without writing, images, and mapping, paying more geographic attention to nonvisual media and faculties enriches our understanding of what it means to be human and to interact with the environment. Still marginal to the discipline, the geography of sound and music can be expected to grow simply because its object is ubiquitous. As fleshy beings, and however able bodied, humans continually engage in a myriad of senses. More critically, the study of these other senses reveals which social groups might benefit from the conventional overemphasis on sight. Over the 1990s, cultural and feminist geographers have criticized the discipline's disproportionate dependence on vision. It would seem that some kind of emancipation of the senses of hearing, touch, smell, taste, balance, temperature, etc. was then inevitable. This has profound implications for the way the production of knowledge is understood.

In geography and cognate disciplines, such as history, cultural studies, urban studies, and anthropology, roughly three relationships between space and sound have been researched. First, sound is heard (or listened to) and thus embodied within particular spaces or soundscapes. Second, sounds and music spread over space, to particular cities or over particular channels. Third, musical sounds signify places to the people who know and enjoy these sounds. Research into these three geographical aspects of soundscapes requires methodological experimentation and a phenomenological sort of conception of everyday life.

The Embodiments of Sound

Silence and the Body

In the mid 1950s, the American composer John Cage pronounced that there is no such thing as silence. In a soundproof chamber or floating in outer space, there are still sounds that cannot be eliminated: the sounds of the body, like breathing, blood circulation, digestion, and even brain activity. As organisms living in air we cannot avoid inhabiting soundscapes. Sound is after all merely air molecules vibrating. A first set of questions in the study of soundscapes revolve around classifying environments in terms of how their sounds affect the body. A subset of questions ask where and why these sounds are called music, and by whom. As avant garde music since the 1950s has shown, the boundaries between silence, music, and noise are contentious. Some environments could be called 'silent' when there is a relative minimum of ambient sounds, with the definition of this minimum depending on the situation.

Soundscapes such as a village dance, funeral, or sitar concert become 'musical' if they are designed to facilitate certain special feelings and social bonding. 'Noise' is unwanted sound, again depending almost entirely on how the situation is defined and by whom. What 'ambient music' attempts to do is erase the distinction between music and background noise. Silence, music, and noise are only comprehensible in relation to each other. There has been more work in esthetics, music therapy, and psychoacoustics addressing these questions, than in geography.

Soundscapes

With the conceptions of silence and sound challenged in their foundations, Cage's influence on sound art and its academic study were decisive. Henceforth music was no longer something artificially isolated from its surroundings, but both performatively and conceptually situated within the ever changing sonic environments of concert halls, art galleries, and the city. The concept of soundscape was comprehensively developed by the Canadian composer and pedagog R. Murray Schafer in the late 1960s. He demanded of musical composition not only a physico acoustic understanding of sound and the environment, but an esthetic appreciation of everyday sounds as music. The latter includes a historical and anthropological knowledge of the importance of sound to human culture. Music is only a very unusual, and therefore deeply meaningful, soundscape. Industrial modernity has greatly increased the amount of noise in our midst, for city dwellers threatening to obliterate the sounds of the nonhuman environment, such as birds, insects, frogs, rain, rustling leaves, and streams.

A sound, if it is heard by a human, necessarily envelops his or her body. Ears cannot be shut, and their evolutionary function is precisely to be able to register information from places the eyes cannot see. As the deep bass in dance clubs testifies, airwaves of sufficient magnitude are registered throughout the body. 'Volume' is not only a linear measure of amplitude, but also a three dimensional quality of sound. It is precisely the 'scape' of the soundscape. The 'higher' the volume, the greater its extension, the more impact it has on present bodies. The location of a body within the volume, say the distance from the speakers or the stage, affects the way that body inhabits the soundscape. Volumes above 140 dB can be painful; a body 5 m from a drill experiences about 100 dB. From teenage bedrooms to flight paths above residential zones, volume is a political problem worth investigating by human geographers.

Sonic Sense of Place

Human geography has since the 1970s been heavily influenced by phenomenological philosophy and became interested in studying the meanings places have for people within it. The sense of place of a location is the collection of experiences and associations the location provokes in its users. As phenomenology comprehends well, in order to experience a place fully, a range of bodily senses come into play, among them hearing. In fact, given their biological importance for orientation and balance, the ears might be as important to sense of place as the eyes are. The current question for human geography is how rapidly changing soundscapes (e.g., as a result of cellular phones and the iPod) alter modern people's senses of place.

A sharp increase of research into sonic sense of place and auditory culture has occurred in art criticism and popular music studies, which experienced a turn toward space and the body over the 1990s. In ethnomusicology, there has longer been a trickle of studies of sound cultures outside the West. Unfortunately, research on contemporary urban soundscapes tends to be concentrated in urban Europe and the United States. Judging from its long tradition of engaging with both phenomenology and cross cultural fieldwork, human geography seems bound to follow suit, and widen understandings of the diverse effects of sound on place and identity.

Historical Soundscapes

Cultural history, long interested in visual landscapes, has since the 1990s gathered evidence of what soundscapes in the past must have been like. Though most historical documents are visual, they nonetheless contain crucial clues as to what soundscapes were like in the past. Indeed, soundscapes are seen, painted, and written down, as much as they are heard. Much of the pleasure of a concert, for example, derives from being able to observe musicians playing their instruments. Richard Leppert's work has spurred scholarship exploring the reciprocal relationships between sight and hearing in history.

Another flourishing line of investigation is the interplay of modernity and recording and transmission technologies. One historian examines the working together of electroacoustics, architecture, and the organization of social order in inter war New York. Another historian has inquired into what cohesive function church bells had for French villages. Bells did not only mark important times of the day and people's lives (weddings, funerals, wars, etc.), but both civil and church authorities knew how to use them in order to structure social life on the countryside. One could think of the use of radio for totalitarian propaganda.

In popular music studies and cultural geography, the unusual history of music of economist Jacques Attali has also been influential. Turning the Marxist base superstructure metaphor on its head, he argues that music heralds socioeconomic change. If listened to carefully, soundscapes can be shown to point to certain deep tendencies of the social formation they are part of. The philosophical implications of conceiving sound, music, and noise as more fundamental to an era than writing or science have yet to be drawn out.

Crowds

Anthropological and evolutionary evidence suggests that music is crucial for keeping human formations together. This is most clearly appreciated in village dances, armies, and sports games. In these soundscapes, many bodies move in unison to emotionally charged sounds, and thereby attain a strong sense of purpose and identity. In cities the world over, crowds of bodies cohere during fairs, festivals, parades, protests, and markets, not only through physical proximity, but through sharing sounds. Few authors have approached these collective events as soundscapes. A soundscape approach would emphasize the many bodies that are required to inhabit and produce the coherence of the soundscape. According to one accomplished historian, the bodily togetherness, the shared heat and pleasures, and the mimicked rhythmic movements of such phenomena as dance and marching have accompanied much of world history.

Sound and Power

An important qualification feminist geography adds to the phenomenological soundscape approach is the fact that different people have, or rather are, different bodies. If a soundscape is necessarily embodied, the question becomes what kind of bodies we are talking about, what they are allowed to do, and in which circumstances. Treating crowds as homogeneous might benefit counting or policing them, but is not helpful to understanding why they should dissipate or converge in the first place. A riot or carnival is evidently experienced differently by women and men, children and adults, immigrants and natives, and tourists and terrorists. Some bodies will shout more, while others will keep to themselves.

Fragments of scholarship exist across the human sciences that show how soundscapes are fully enmeshed in relations of power. If the turn toward the body in human geography leads to the study of soundscapes, it will have to be demonstrated that collective embodiment of sound and music can result in both identity (the coherence of bodies) and differentiation (the segregating of groups of bodies). Shouting political slogans coheres some bodies, but threatens others. The everyday example of the annoyances of cellular phone conversations on the bus attests that a soundscape can contain more than one meaning, as different bodies are involved in it. Differential embodiment brings the critical study of soundscapes immediately to considering the politics of place.

The historical soundscape literature shows that some soundscapes are clearly manifestations of the spatial strategies of the nation state, urban planning, capitalism, and public order. Sound is controlled for a wide range of purposes; in fact, without the control of sound social life cannot run smoothly. The tranquility of places of worship, the strict patrolling of speech in hospitals and libraries, the Muzak played in hotel lobbies or subway stations, and the Christmas carols in shopping malls demonstrate the almost unconscious infusion of politics in sonic space. The sounds of 'nature' or 'yesteryear' that technological change excludes (of the rooster, the blacksmith, or the steam engine) thereby attain nostalgic, potentially reactionary ideological meanings. On a variety of scales, therefore, soundscapes matter deeply to political geography.

The Diffusion of Soundscapes

Music Geography

A small subfield of American cultural geography, music geography has concerned itself mainly with the origins and spread over space of genres, musical instruments, and performance styles. Questions asked include where European folk dances survive, where blues was taken by African American migration, and why hip hop exhibits regional variations. This music geography shares with ethnomusicology and American studies an interest in explaining the existence of music through its function for local culture, but geographers, in addition, map musical similarities over multiple cities, regions, and countries. Economic geography meanwhile explains the commercial and legal nature of music's international distributions. It can be inferred from the study of diffusion/distribution that there are some similarities in the soundscapes within places where a given musical phenomenon exists. The diffusion of electronic dance music, for example, largely coincides with the diffusion of clubbing culture.

In the wake of British cultural studies over the 1980s and 1990s, a number of problems with both economic geography and American cultural geography were pointed out. First, the supposed origin or hearth of a musical trait only leads to further questions about which routes went into constituting it. Second, diffusion is also transformation, as is easily observed in musical styles and instruments. There cannot be certainty about the similarity of soundscapes from merely knowing a musical trait is present in a location. If diffusion attempts at maintaining a pure version of cultural traits, as is the case with European classical music, critical music geography would ask what class, gender, and ethnic distinctions this underscores. Third, diffusion is always uneven, reflecting the unevenness in wealth, population density, access to media technologies, and cultural tastes in the world. While economic geography to some extent maps the conditions for the creation of markets for a certain genre, the concrete dynamics of bodies and sounds within the actual soundscapes have to be included to obtain a sense of the collective pleasures that that genre fulfills.

A final, fourth way of problematizing the diffusion of music derives from the fact that the meanings and pleasures of musical soundscapes are not automatically apparent to everyone who hears it. Work in the sociology of music shows that the diffusion and enjoyment of music intersects with social relations of race, class, gender, and sexuality. Simply mapping the presence of jazz or reggae in certain places does not tell us which groups within those places attain a sense of identity through the music, and which groups resist being exposed to it. Thus, whether or not a city's soundscape includes reggae depends on much more than reggae's musical characteristics. A critical geography of the globalization of music would suggest that power relations in society and between societies greatly affect the spatial reach of a musical genre, dance step, or pop star.

The Global Soundscape

Except music, there has been close to no academic work on the diffusion of everyday soundscapes. With more humans living in cities than in the countryside, the planetary soundscape is transforming rapidly. Not only are there many documented cases of biotopes being disrupted by the onslaught of sonic modernity, but there are also a few examples of species adapting to modern noise (pigeons and squirrels). While the spread of cinema houses, televisions, cars, telephones, cassettes, jet fighters, computers, fridges, vacuum cleaners, and highways has been hotly debated from the 1970s onward, the same cannot be said for their accompanying sounds. Noise pollution has of course been an established social and legal concern since the nineteenth century, but systematic geographical scholarship is still wanting. Given the growing theoretical interest in sound and the ecological effects of changing soundscapes across the globe, however, it can be expected that it is a matter of time before geographers attend to the acoustics of industrial modernity.

Sounds Representing Place

The most indirect and enigmatic relationship between space and sound occurs when sounds make absent places psychologically present. Whereas embodiment and diffusion deal with sounds inhabiting and thereby giving meaning to tangible places, in this third case, space is presented to minds in a virtual or imagined way. Such sonic representation can only take place if the listener is familiar with the ways that the sounds are supposed to mediate. For example, to someone unfamiliar with Chinese opera, that singing will not evoke China. The lyrics of a Bollywood movie hit enable making imaginative connections to India easier for a migrant who knows Hindi. Soundscapes are therefore capable of becoming more than merely volume circumscribing a physical place, and more than a dot within a global distribution. Through the working of the imagination, soundscapes can also transport its participants to faraway locations and landscapes. This representational function of soundscapes is of course most appreciated in music. Though indexically music can be said to refer to the instrument and the player's body, the musical representation of places can in nonindexical ways go much further in space.

Representation is a notoriously elusive concept, perhaps especially so when it pertains to music. It is generally used to conceive how language is related to the physical world of things and situations. For example, the word 'cow' represents the milk giving farm animal. However, there need to be rules learnt quite consciously to be able to understand that the three letters c o w are supposed to represent this animal. Without knowing the rules the letters mean nothing. This is why learning a new language takes a lot of time. Most scholars believe the meanings of music come about very differently than the meanings of words. This is because musical sounds vary so much and so subtly, and are embodied in such variable ways (mostly unconsciously) that they cannot be easily formalized into rules. By enveloping the body and never stopping, music seems to affect people in a more immediate and inevitable manner than words. Though scholars agree it is somehow meaningful, there is no agreement on whether music 'represents' anything at all. Still, insofar as there are meanings in music that need to be learned by its users, and insofar as music makes things present that can also be described in words, many would agree music is capable of representing places, persons, or concepts, even if it is often enjoyed without any identi fiable representation.

Music geography and musicology have both elucidated this evocation of place in music, which can happen in one of three ways. First, there is a properly linguistic com ponent of song, namely, the lyrics. Songs are in a language identified with a particular region or country. They speak about circumstances, people, sentiments, mythologies, and landscapes relevant to people from a specific origin. Sec ond, the sounds of instruments from a particular place can invoke that place to those who know it, for example, the fiddle for Ireland or the didgeridoo for Australia. Melodies, harmonies, rhythms, and timbres can all represent cultures and regions. In much popular music this is used efficiently to add a hint of the exotic, as with the sitar and India. With sampling technology, the entire soundscapes of places are reproduced accurately, and thus we have albums with soundscapes of the rain forest or dolphin games.

Third, within the soundscape, there might be media other than sound or lyrics that represent places. Photo graphs of Thailand might adorn a restaurant with Thai music in Vietnam. The embodied reception of music, es pecially in collective dance, can itself come to stand for a place or country, as when immigrants maintain a tradition of dancing on special occasions. Thus, in order to probe into a soundscape's representational work, its entire space–time including tastes, smells, rituals, and architecture, not just its sound, has to be taken into account. This holistic and more 'fleshy' or ethnographic approach to represen tation is not yet an explicit research agenda item, but is found in some work, especially on migrant place making. It balances an overemphasis on lyrical content and formal musical analysis that much work on the imaginary geog raphies of music has tended to.

Future Directions

Given geography's long standing marriage to visual methods, the growth of soundscape studies within the discipline has been slow compared to art criticism and history. Yet, cognate disciplines continue learning much from geography's long standing theorizations of space, place, scale, unevenness, and flow, even if these were primarily studied through sight. As geographers became intensely interdisciplinary over the 1980s and 1990s, scattered exchanges on music and sound started occur ring. These are the conditions in which a geography of the embodied, diffuse, and representational nature of soundscapes can take hold and grow.

See also: Affect; Body, The; Difference/Politics of Difference; Diffusion; Embodied Knowing; Media; Phenomenology/Phenomenological Geography; Place, Politics of; Sound and Music; Visuality.

Further Reading

Attali, J. (1985). *Noise: The Political Economy of Music*. Minneapolis, MN: University of Minnesota Press.
Bull, M. and Back, L. (eds.) (2004). *The Auditory Culture Reader*. Oxford: Berg.
Cage, J. (1961). *Silence*. Cambridge, MA: MIT Press.
Carney, G. S. (ed.) (2003). *The Sounds of People and Places: A Geography of American Music from Country to Classical and Blues to Bop* (4th edn.). Lanham, MD: Rowman & Littlefield.
Connell, J. and Gibson, J. (2003). *Sound Tracks: Popular Music, Identity and Place*. London: Routledge.
Corbin, A. (1998). *Village Bells: Sound and Meaning in the Nineteenth Century French Countryside*. New York: Columbia University Press.
Feld, S. (1990). *Sound and Sentiment: Birds, Weeping, Poetics, and Song in Kaluli Expression* (2nd edn.). Philadelphia, PA: University of Pennsylvania
Press.
Leppert, R. (1993). *The Sight of Sound: Music, Representation, and the History of the Body*. Berkeley, CA: University of California Press.
Leyshon, A., Matless, D. and Revill, G. (eds.) (1998). *The Place of Music*. New York: Guilford.
Matless, D. (2005). Sonic geography in a nature region. *Social and Cultural Geography* 6(5), 745 766.
McNeill, W. (1997). *Keeping Together in Time: Dance and Drill in Human History*. Cambridge, MA: Harvard University Press.
Saldanha, A. (2002). Music tourism and factions of bodies in Goa. *Tourist Studies* 2(1), 43 62.
Schafer, R. M. (1994). *The Soundscape: Our Sonic Environment and the Tuning of the World*. Rochester, VT: Destiny.
Smith, S. J. (1997). Beyond geography's visible worlds. *Progress in Human Geography* 21(4), 502 529.
Thompson, E. (2000). *The Soundscape of Modernity: Architectural Acoustics and the Culture of Listening in America, 1900 1933*. Cambridge, MA: MIT Press.

Relevant Websites

http://www.acousticecology.org
Acoustic Ecology Institute.
http://www.6villages.tpu.fi
Acoustic Environments in Change Project.
http://www.globalmusicproject.org
Global Music Project.
http://www.ubu.com
UbuWeb, online avant garde sound poetry database.
http://interact.uoregon.edu
World Forum for Acoustic Ecology.

South Asia

G. P. Chapman, Centre for Advanced Study, Oslo, Norway

Glossary

Caste A social group maintained by a restricted endogenous marriage group, whose status is ascribed and differentiated within a hierarchy of castes. Caste is word of Portuguese origin. In India the terms 'varna' (colour) and 'jati' (the marriage group) are more commonly used.

Diaspora Emigrants who in their country of adoption nevertheless retain distinctive identitive links with their country of origin. Initially the term was mostly used to refer to the Jewish diaspora scattered from Palestine. Nowadays it is frequently used with reference to South Asians living outside South Asia.

Westphalian Replacing the religious-imperial overlordship of the Pope and the Holy Roman Empire, the term refers to the system of independent sovereign nation states which was the outcome of the Treaty of Westphalia (1648).

Introduction

Use the term 'South Asia' in any context, be it as the title of an academic course, as part of a newspaper headline, or as the setting for a nature film, and then ask the listener/reader/viewer where the feature is set, and he/she will most often respond 'Southeast Asia'. The term South Asia has not registered in the minds of the public of many beyond (and as we will see, to some extent, within) the region, whereas 'Southeast Asia' definitely has. It is not clear when either term was first used, but it is almost certain that they both date from during or soon after World War II and the start of the Cold War. The United States and the Western Allies contrived to contain the Soviet Union with a series of encircling alliances. In the west was the North Atlantic Treaty Organization (NATO), which included Turkey; further east came the Central Treaty Organization (CENTO), which included Turkey, the then monarchies of Iraq and Iran, and the military regime in Pakistan, plus Western powers; and the Southeast Asian Treaty Organization (SEATO), which was intended as a bulwark against communism in Indochina and the Philippines. SEATO also partly overlapped with the Association of Southeast Asian Nations (ASEAN), a later economic (and geopolitically rooted) group growing in the same region. Both were in part the result of the Vietnam Wars, and perhaps they were responsible for raising the profile of Southeast Asia in Western minds.

There never was a SATO – much though the US would have liked to have had one – bringing independent India into the grand coalition of the mischievously named 'Free World'. This was because what had once been known and seen as a single world region – India, or the Indian Empire (the 'jewel within the crown' of the British Empire) – had broken up at independence in 1947, into the larger part, the Republic of India, and the smaller part, Pakistan, which were almost immediately at war with each other over disputed territory and whose creation involved massive population transfers and violence that would change the social ('communal') geography of the region. Ultimately both sides looked for allies, and as Pakistan aligned itself with the US (and later sought links with China), India sought aid and arms from the USSR. Hence the divisions of the Cold War became entrenched in South Asia, and the wholeness of the region was further riven. There are, however, regional destinies. The Cold War fracturing of Europe between the west and east has largely been mended. SATO may not exist, but the South Asian Association for Regional Cooperation (SAARC, founded in 1985) does, which to some extent has taken ASEAN as its model, and which reflects the realization of what an integrated South Asia could achieve. However, the regional destiny will not be completed until there are defense alliances (or at least a culture of security and cooperation) between the current great protagonists, when divisions within South Asia are mended and the states collectively face the outside world. It is almost inconceivable that the term 'India' will be used for this greater entity as once it was – but, not impossible.

The Distinctiveness of South Asia

The wholeness of South Asia that suggests this regional destiny is not imaginary; it has framed much of the history just discussed. The wholeness starts with plate tectonics and geological history, and the directly consequent geopolitical history. There is a Eurasian tectonic plate, spanning the Old World from the Atlantic to the Pacific. The Indian plate is now attached to the south of the Eurasian plate, but that arrived from elsewhere. Until about 200 million years ago, what is now the Deccan Block was part of what geologists call Gondwana(land), the southern supercontinent that included Africa, Madagascar, India, Australia, and Antarctica. The name Gondwana comes from the Gond(i) tribes of the central Deccan, in whose lands the first recognition of the geological signature took place. The Indian plate broke off, became an island (like Madagascar or Australia), and drifted across the Indian Ocean, beginning its collision with Eurasia about 60 million years ago. It lifted the Tibet plateau, pushed up the Himalayas, the Hindu Kush, and the mountains on the Assam–Burma border. The collision is ongoing, and the mountains continue to be pushed up in what is one of the world's most seismically active zones. Looking at a physical map of the world, it is easy to see this huge dent which 'India' has made in Eurasia.

The high mountains block the seasonal migration of the jet stream, and hence they impact on the onset of the monsoon; they are the source of three of the world's greatest rivers – the Indus, the Ganges, and the Brahmaputra; they are the source of the sediments which have built the world's greatest riverine plains and the world's greatest delta; and they block easy migration and invasion from all sides, of all forms of life, from flora and fauna to human beings. The South Asia thus framed has so many extremes – the highest mountains, the flattest plains, the wettest place on Earth (in Meghalaya), and one of the hottest and driest (the Thar Desert). But, despite this diversity, it always remains true that movement within South Asia is easier than movement by land into or out of South Asia.

It is easy to see how this large region, attached to only just a part of Eurasia, has been labeled a subcontinent. Indeed, the idea that it is 'the subcontinent' lives on in that inheritance from the British Raj, cricket. West Indian and English cricket teams go on a tour of 'the subcontinent', in the process usually playing matches against India, Pakistan, Bangladesh, and Sri Lanka. In geopolitical designations too, the idea of a subcontinent has proven enduring.

At the beginning of the twentieth century, the British geographer Sir Halford Mackinder considered that the Columbian epoch of European expansion was over, and that the world now constituted a global system. (This was about globalization before the word was coined for the current phase of it.) In wondering how the physical structure of the land masses and oceans impacted on political history, he surmised that the world was composed of two major parts – a heartland, or, the Eurasian land mass (which excludes maritime Europe), and the rest – the maritime dependent outer arc. His thinking was about the possibilities of movement. The heartland either has internal drainage or drains to the Arctic Ocean, and is therefore not open to the naval powers. It is an arena of land movement, in premodern times, by horse. At the time he wrote, long distance economic integration on land was possible for the first time because of the advent of the railways. The maritime arc constituted those territories amenable to integration by ships. To begin with, this lumping together may come as a surprise. But when the distribution of population in this arc is considered, it becomes more persuasive. The populations of both North and South America are on the east and west coasts, not the interior. The populations of Africa are mostly by the coasts, and so are those of Australia; even most of China's great cities (and today, its special economic zones) are on or near the coast.

Mackinder's visions influenced much world geopolitical thinking through both world wars, were significant in the Cold War, and are arguably applicable (and much debated with respect to) to the new rise of Russia. One of the recent interpreters of his ideas, Saul B. Cohen, divides the world first into its two major geostrategic regions, and then subdivides those into geopolitical regions, areas of similar culture and proximity and movement, wherein integration is possible. So the Caribbean is a geopolitical region within the maritime world, and so is the Maghreb of Africa.

In his scheme, the subcontinent is unique: it is the only geopolitical region which is independent of a geostrategic region. It has incubated independent land empires, but it is skirted by sea. It is attached to Asia, but protected from it. This independence and ambiguity sums up the subcontinent's history.

The Cultural Origins

Early human settlement was almost certainly a result of the migration of hunter gatherers by land. It seems likely that a significant group among these settlers were the forebears of the Dravidian people – a category whose meaning is disputed, but which refers today to the peoples of South India who speak Dravidian languages (Tamil, Malayalam, Kannada, Telugu, and others). By 2600 BC they had founded one of the world's first great hydraulic civilizations, by the Indus in Sind, where the capital city of the Harappan civilization was Mohenjo Daro. Starting from about 2000 or 1500 BC, waves of Aryan invaders/settlers arrived, through the passes of the northwest frontier. These people were lighter skinned, and spoke a language (Sanskrit) which is an offshoot of the Indo European stem. It is disputed how much of Hinduism stems from the Aryans, from the Dravidians, or from an assimilation of both traditions, but there are some facts which cannot be disputed. The linguistic invasion of India is incomplete: the northern languages from Baluchistan to Bengal, from Himachal Pradesh to Maharashtra and Orissa – thus spanning Pakistan, India, and Bangladesh – are descended from Sanskrit, in much the way the Romance languages in Europe are descended from Latin. But the southern states remain Dravidian speaking. A second fact is that there are complex 'racial' gradients in South Asia. In general (and to simplify) in any locality, the higher castes are lighter skinned, and there is a geographical gradient from the lighter north west to the darker south. While these discriminations were bolstered by the categorizations made under colo nial rule, they cannot be simply reduced to that, and have proven persistent.

Hinduism is not a 'revealed religion' with an im mutable source of authority from a specific date – as Islam and Christianity have, for example. But it does have sacred texts, the Vedas, starting from oral traditions be fore 2000 BC. It is commonly said that it is not a religion at all (or must be understood on its own terms, not through comparison with Islam or Christianity), but a way of life, following rich, complex, and sometimes contradictory philosophies. Most people who would consider themselves Hindu have the following in com mon (though to varying degrees and few of these things are exclusive to Hinduism): a belief in reincarnation; veneration of cows, and an inclination to vegetarianism; divinity in all creation, including rivers and mountains and animals; and acceptance of arranged marriages within their jati (local caste formation). Many and com plex are the ways that castes are ranked in any locality, but most can be mapped approximately into Brahmin (priestly), Kshatriya (warrior and ruler), Vaisya (mer chants and trading), and Sudra (menials and laborers). Beneath them come the outcastes, the untouchables, whose tasks are the most polluting ones, of cleaning cesspits and cleaning carcasses. Gandhi made a point of staying with the poorest, and called the outcastes 'Hari jans' (the children of God). In modern India, the term Dalit is used. There have been swathes of legislation promoting positive discrimination in favor of poor and backward castes, but of course, enforcement means the recognition of caste and caste membership. In general, modern India may have provided better education and many new job opportunities, but this many millennia old social system is never far from the surface.

Well before the rise of the West, the Ganges valley was, in comparative world terms, a densely settled agricultural area supporting a large population. Early Hindu empires were based around the modern day city of Patna in Bihar – ironically Bihar is now seen as the most 'backward', cor rupt, and violent state. It was also in Bihar that Gautama, a Kshatriya Hindu prince, found enlightenment and became the Buddha and for the next few centuries it seemed that Buddhism rather than Hinduism might become India's dominant belief system. (The two share much, but Bud dhism rejects caste.) It was from this base that the first emperors such as Asoka extended their sway into the Deccan and as far as Afghanistan (where the great Bud dhas of Bamiyan stood until their ideologically motivated destruction by the Taliban in 2001). The pattern estab lishes some general truths about South Asia.

Invasions normally occurred through the northwest frontier; power became based on the agricultural wealth of the Ganges valley; India provided sufficient resources and geographical range that empires became self suf ficient; and closed off the northwest frontier, they ex tended into the Deccan as a land based power, using large armies, building roads, and using elephants and pack animals. In the heart of the empire, movement on the Ganges and its tributaries was important to sustain the imperial centers. The scientific advances of early Indian empires were spectacular, in astronomy, in me tallurgy, in mathematics, in philosophy, and a lot else, and much of what Europe thought it obtained from the Arabs, including their decimal numerical system, actually has it origins in India.

The cycle of the empire is a topic which exercises many a historian. Their rise and fall may be partly a result of internal dynamics. It is suggested that over generations, the ruling elite became progressively effete, and there was increasing and increasingly costly bur eaucracy. It is not proposed to get into these details here, but suffice it to say that successive empires in South Asia have declined, often accompanied by new invasions from the northwest, to be followed usually by a constellation of pettier kingdoms and chieftainships. But the cultural continuity has remained. This India, imperial or divided, was and is one of the world's great cultural heartlands.

After the death of the Prophet Mohammed, Islam was spread rapidly by Arab armies in Arabia and North Af rica. In India it was almost certainly spread into the Indus valley quite early by Sufis, who were of a missionary mystic nature. It was probably introduced by sea, to southern India and to Bengal. It made its first large scale impact by land with Afghan invasions from the northwest frontier, starting in the twelfth century. By the middle of the thirteenth century, the first of the Delhi Sultanates had been established, recognized by the Khalif of Bagh dad. This was an empire that geographically matched that of Asoka, more than a 1000 years earlier, except that the most significant cities, like Lahore and Delhi, were further west than Bihar. But in cultural terms it was new: the ruling elite shared a religion with a revealed source of truth, the Koran. Some conversion of the masses was attempted, but the cultural distinction between rulers and the ruled had its uses, bonding the former together. 'The Sultanate' was a period which embraced several dynasties, internal strife, and assault from without. It ended when a new Islamic force, that of the Mughals (so named because they originated from Mongolia), invaded in 1525, and defeated the army of Ibrahim Lodhi at Panipat, northwest of Delhi. This was the beginning of the Mughal Empire that became famed in Europe for its power and wealth: to this day, in England the term 'mogul' is used to indicate the tycoons who head the media empires or the oil industry. Ironically, the empire grew as Europeans were establishing themselves on the coast for the first time, starting with the Portuguese in

Goa in 1510, followed by the Danes, the Dutch, the British, and the French.

India (like China) was one of the world's leading manufacturers. The Europeans wanted above all, not just cotton and silk cloth, but finished articles of clothing, in addition to spices, jewels, and saltpeter. Mostly what the Mughal Empire wanted from them was silver bullion, in which the rupee was denominated, which was unavailable in India. The British East India Company had independent centers on the west coast (Bombay, now Mumbai), the south (Madras, now Chennai), and Bengal (Calcutta, now Kolkata), often in close proximity to other European factories.

After two centuries, the Mughal Empire went into decline at the center, and in 1739, Nadir Shah of Persia sacked Delhi, and withdrew with the Peacock Throne. A little later, the practically independent Nawab of Bengal tried to evict the British, but instead was himself overthrown and the Company took charge of Bengal. So for the first time, power had arrived in South Asia by sea, from the maritime arc. Drawn in by the instability and power vacuum following the decline of the Mughals, the British had dominion over almost the whole of India by the early nineteenth century. Since not much of it was of commercial interest to them, in many areas they ruled indirectly, by accepting the internal autonomy of princely states. Many of these were in the central and northwestern Deccan, some in the mountains, the largest of which was Kashmir. The nature of trade changed radically: the industrial revolution in Britain took off, raw materials like cotton were needed, supplanting handspun cotton and finished goods; and Britain used India as a market for its cutlery and milled cotton. Later, the railways entrenched the new coastal ports, and suddenly India's geography had been turned inside out, and its economy de industrialized. The great land cities of Lahore, Delhi, Agra, Hyderabad, and Mysore were eclipsed by Bombay, Madras, and Calcutta. Ironically, in 1911 the maritime power proclaimed its inheritance by shifting its capital to the land base of Delhi.

The Imperial Intrusion of Modernity

Britain also implanted in much of India (and what became Pakistan and Bangladesh), begrudgingly and starting from small steps at the lowest levels of municipal government, the seeds of democracy, new forms of state apparatus, and university education in English, a true *lingua franca* for the whole of India's intelligentsia, for the first time. Within the British Empire there also circulated a Westphalian 'national goal' – as colonies became self governing states (achieving 'home rule'). The question that was posed as the British Empire crumbled was therefore: "Is it possible for a democratic Westphalian nation to emerge as inheritor of the whole of the British Indian Empire?" The guiding force of the nationalists, Mahatma Gandhi fervently hoped that the answer would be yes. But even in the nineteenth century, enlightened and concerned British commentators such as John Bright MP thought that a more 'European' Westphalian outcome was likely – with five or six large, successor states. The founders of the Muslim League that challenged Gandhi's vision believed, to begin with, that there would be separate Muslim states in the Indus valley in the west, the central Deccan, and in Bengal, and that Rajasthan and Sikh Punjab should also be independent. The Muslim leadership feared Hindu overlordship, and did not believe that a secular India was either desirable or likely to last. The outcome of the struggle was a composite, with a very large India embracing its myriad languages and cultures, and a smaller Pakistan, in two halves. The human loss of life as this partition occurred was horrific – more than a million people died as Punjab and Bengal were split between the successor states, and more than 12 million people were (to use a phrase that has been applied to other such forced transfers) 'ethnically cleansed' (though in practices, ethnicity and religion did not coincide in any straightforward way). Then, 24 years later, another one million died as Bangladesh hived itself off from Pakistan.

The partition has not achieved the separation of all Muslims and Hindus. India is the world's second largest Muslim state, because 15% of 1.1 billion is a lot of people. This is problematic for the Organisation of the Islamic Conference, on which India has observer status. India has remained a democratic state with a secular constitution, but in recent decades Hindu nationalism has made major political advances, and some political leaders have either tolerated or instigated attacks on some Muslim communities in India. For all of this, and the continuing dispute over Kashmir, there is a linguistic and cultural continuity between Pakistani and Indian Punjab, and between Indian and Bangladeshi Bengal. Indeed, most Pakistani Punjabis might feel more affinity for their Indian kin than their Pathan compatriots in the northwest frontier.

Westphalian South Asia

After independence, India rejected the colonial economy of the maritime trading worlds, and sought relatively autarchic development with a major role for central planning and state industry. In the first four decades of independence, it grew slowly – the term 'the Hindu rate of growth' was coined as a rather derogatory epithet. But a measure of social justice was achieved, and the sense of nationhood built. This was the inheritor of the land empire model. In the modern world the trajectory failed, as technology is now traded internationally, and no state 'does it all' by itself. After a foreign exchange crisis, change was forced on India. Liberalization since the 1990s has brought pains to many, particularly poor farmers, but also a flood of inward investment, much of it from rich Indians in the diasporas in Europe and North America, and the economy is now growing at a speed not far short of China's. Growth is regionally unequal. The lower Ganges areas of Uttar Pradesh, Bihar, and West Bengal (and Bangladesh) contain the world's greatest number of poor people – far more than in Africa. It is in the south, west, and northwest that the cities are fueling India's economic spurt (notably Mumbai, Bangalore, and other centers that have become plugged into a new wave of globalized capitalism). There are tensions here that could threaten the union in the future.

Pakistan has one of the world's fastest growing populations, and has had a very poor record in education. Its public services have been stunted by defense ex penditure to balance the perceived threat, to its in dependence, by India. Whether or not the public support this has hardly been tested: the military are more often than not the government, setting their own budget, and when not formally the government, still tell it what to do. For many observers, three 'A's rule Pakistan: Allah, the Army, and America. For 24 years, it exploited East Pakistan (now Bangladesh) and when it was forced out, that state was impoverished, described by a senior economist as 'Malthusia'. It is still remarkably poor, but both Bangladesh and Pakistan have in recent years begun to grow fast economically, both of them in an era of greater economic liberalization. Bangladesh now exports finished garments, as it used to when the British first arrived. It is also exploiting its saline/brackish coastal areas from shrimp production. Cotton textiles are im portant for Pakistan too, which has bought up much of the machinery being decommissioned in Europe. In sum, in the new globalizing world, it seems that South Asia is becoming a more significant part of the maritime arc.

Modern South Asia is probably best defined by the membership of SAARC. The seven founding countries were the Maldives, Pakistan, India, Nepal, Bangladesh, Bhutan, and Sri Lanka. They have been joined since 2007 by Afghanistan, although whether this re inclusion of this ancient Indian realm will endure is yet to be seen. The United Nations includes Iran in its definition of South Asia, perhaps not only because of its location, but because Iranians speak Farsi, an Indo European language. Farsi (Persian) was the court language of the Delhi Sultanate and the Mughal Empire and in their early period was used by the British administrators as well. Hindustani which has embraced Persian words and which is written in the Per sian/Arabic script is known as Urdu, and is the official language of Pakistan and one of the official languages of India. Hindustani which is written in Devanagari script and which uses a vocabulary more derived from Sanskrit is now known as Hindi, and is the most important language of South Asia.

India, the regional hegemon, has fought several wars with Pakistan, in one of them acting as midwife to the birth of Bangladesh; reversed a coup in the Maldives; imposed a peace keeping force in the predominantly Tamil part of northern Sri Lanka; for a while blocked trade with Nepal; and formally dominates Bhutan's for eign policy. A free and powerful SAARC would soon see the peripheral small nations forming alliances against Indian power. Because of this, India insists that SAARC discusses only general issues of interest to all, not bi lateral ones. So, for example, Bangladesh cannot raise specific cross border water issues with India inside SAARC. Because of this, and its very limited budget, SAARC is widely seen as toothless.

South Asia and the Diaspora

In the fifteenth century, Europe knew of the fabled eastern lands of silk and spices, the Indies, but was blocked from direct overland trade by Arabic and Turkish Empires. The 400 year Columbian epoch began when Christopher Columbus set sail westward to reach the Indies from the other side. When he struck the New World he assumed at first that he had completed the circumnavigation and had reached the Indies. The word stuck; the inhabitants of the New World became known as Indians (only later to be distinguished as Red Indians), and the islands of the Caribbean became the West Indies. So, an early export of South Asia in the first phases of globalization was its own identity. After the establishment of European empires in Asia and Africa, and after the abolition of the slave trade, labor was still often in short supply in many of the colonies, but abundant in India. Thus one of the great migrations of the nineteenth century and early twentieth century was indentured labor from India to South Africa, East Africa, the Caribbean, Burma, and Malaya, and even the Pacific (Fiji). These migrations were often summed up as tro pical–tropical. After the end of World War II and the end of Empire, there have been several more waves of out migration, but of rather different types. From Pakistan, India, and Bangladesh the first migrants into the UK were mostly in relatively low skilled factory jobs. They were followed by entrepreneurs setting up small businesses – one sector being the 'Indian' restaurant. Since the ma jority of Indian restaurants in Britain were owned by Bangladeshis (from one district – Sylhet) the adjective Indian is inaccurate but ineradicable. India, Pakistan, and Bangladesh have a new currency in the UK which pro vokes debates about integration and assimilation, com munalism, and secularism. Educated professionals have made their mark in the health services and in banking and IT. Indian multinational companies are now purchasing companies in Europe. The migration of educated scientists and engineers (particularly in IT) to the US and Canada has been pronounced – and has necessitated the introduction in America of the term Asian Indian (or sometimes East Indian), since in North America, the terms Asian and Indian have not usually been associated with folk of South Asian origins or heritage. From both the UK and North America, there is return migration and investment to South Asia. With these migration streams and the constant exchange of visits flow the cultural exchanges. From Britain to India, the export of chicken tikka masala curry (a dish forged by Bengalis to satisfy English tastes), and from India to Britain and the USA, the flamboyant movies and music of Bollywood which have in recent years become an influence on popular culture in the West, have provoked debate about hybrid cultural identities and their geographies.

A Common Future?

Five times in history, most of South Asia has been incorporated into an imperial state. In between, it has been split into smaller kingdoms and tribal territories. In the last three empires (the Sultanate of Delhi, the Mughal, and the British) the ruling elite had a cultural background that distinguished them from the masses, and that distinction was part of what maintained the integration of the area. The peasantry remained impoverished, hardly served by communications, and served not at all by education. Has South Asia's history been not just written by the elite, but also exclusively made by them? In recent decades, the authors behind a project known as Subaltern Studies have tried to write a new history of local empowerment and of resistance to the elite, which provides a new understanding of the true roots of 'India'. Different events in history are reexamined and given a new interpretation. A classic example is the term 'First War of Independence' used in India to describe what the British call the Indian Mutiny of 1857, when British forces very nearly lost control of India. The events of the mutiny were uncoordinated and spasmodic, more a kind of reaction to and collective punishment of the British 'intruders' than an attempt to rescue a homeland. Delhi was recaptured with the help of Sikh and other forces who had no conception that they had something in common with Bengali sepoys. Whatever degree of agency was granted to the masses, they were not about to achieve independence of their own accord.

How then, and by whom, was an independence movement led? The answer, by whom, can be given very easily: by the relatively very tiny, educated class of Indians that had emerged under British rule. But they were small enough to be easily watched, and put in and out of prison as and when necessary. The how therefore required what in modern jargon would be called force multiplication – use the masses. It was part of Gandhi's genius that he could and did connect with the masses. But the masses had little or no idea of an Indian identity – they were Muslim or Hindu Bengalis, Muslim or Hindu Gujaratis, Muslim, Hindu, or Sikh Punjabis. The leaders of the independence struggle may have been nationalist (and communalistic in some cases) but the masses were frequently communalistic. When they were roused, intercommunal strife often rose. Nehru, leader of the All India National Congress, had imbibed from the West, the idea of a secular state: but other Congress leaders (particularly at provincial level) and the Muslim League had not. The failure of Congress to persuade Jinnah that Muslim interests would be protected in a future federal and secular state, was a major cause of the partition. This is why the struggle over Kashmir is so intense. It (the Indian part of Kashmir) is the only Muslim majority state in India, and 'proof' that a separate homeland for the Muslims was not necessary. For the Pakistanis, the insurgency in Kashmir is proof that Muslims need a homeland of their own. A military or political victory in favor of either state would enable them to rewrite history.

Although Nehru passionately believed in the identity of India, and had been persuaded by his own propaganda that perhaps the peasantry did too, he was well aware of the fragility of the new state. The autarchic development plans were part and parcel of the huge effort to persuade the people of the largest democratic state ever created, to retain their unity. The simultaneous adherence to democracy and the achievement of national identity was one of the most stupendous political feats of the late twentieth century. Westphalian India seems a huge success. These days, Indian newspaper editors talk of the 'post national' class of readers. These are not people who have gone beyond nationalism: indeed, they are fiercely nationalistic. They are post nationalist because these are the business leaders and the elite in Chennai or Mumbai or Kolkata, who accept that India is now a reality, accept it as a working proposition, and who can look outward with confidence. But the common man's identity is still mostly within the local community. Pakistan by contrast is only a state because its borders (most of them) are recognized and accepted by the international community. It failed to develop a national identity between its original constituent wings (West and East Pakistan), and is now failing to develop a common identity between Punjabi, Sindhi, Pathan, and Baluchi, let alone accommodate the fundamentalist and moderate wings of Islam.

The word Hindu means the people who live by and beyond the Indus (known in Sanskrit as Sindhu). The Pakistanis and others frequently refer to India as Hindustan – as the British used to recognize the north of

India as Hindustan and its language as Hindustani. The word India has the same root, but is not used within vernacular languages in India – in which it is known as Bharat, a word meaning emperor and empire from the first Hindu empires. The expression of less secular Hindu India has been marked by the rise of the Bharatiya Janata (people's) Party – the BJP – espousers of Hindutva. The divisions within South Asia are those of class and col lective identity – and these are often parochial in soci eties which are only half literate, where poverty is still extensive, and languages distinct. Gandhi thought of the two world wars as Europe's civil wars. It has taken Europe those tragedies to get from there to its current state of mutual respect and acceptance, where what is common is valued, and what is different is largely restricted to economic, cultural, and sporting rivalry.

Perhaps we can accept that South Asia will take its time too, to find its regional destiny, to realize that what unites is more significant than what divides. The biggest leap of all would be when the people of the Indus and the Ganges and the Deccan are happy again with that embracing and secular term, India. Then perhaps 'South Asia' will have been forgotten, before the term is lodged in the minds of the outside world.

See also: Asia; Diaspora; Empire; Ethnicity; Imperialistic Geographies; Religion/Spirituality/Faith; South Asia.

Further Reading

Allchin, B. and Allchin, R. (1982). *The Rise of Civilization in India and Pakistan*. Cambridge: Cambridge University Press.
Basham, A. L. (1954). *The Wonder That Was India: A Survey of the History and Culture of the Indian Sub continent before the Coming of the Muslims*. London: Sidgwick Jackson.
Chapman, G. P. (2000, 2003, 2008/9). *The Geopolitics of South Asia: From Early Empires to the Nuclear Age*. Ashgate: Aldershot.
Chapman, G. P. and Thompson, M. (eds.) (1996). *Water and the Quest for Sustainable Development in the Ganges Valley*. London: Mansell.
Cohn, B. S. (1971). *India: The Sociology of a Civilization*. Princeton, NJ: Englewood Cliffs.
Corbridge, S. and Harriss, J. (2000). *Reinventing India: Liberalization, Hindu Nationalism and Popular Democracy.* Cambridge: Polity Press.
Coupland, R. (1943). *Report on the Constitutional Problem in India, London, Pt I The Indian Problem 1833 1935, Pt II Indian Politics 1936 1942, Pt III The Future of India*. Oxford: Oxford University Press.
Dumont, L. (1970). *Homo Hierarchicus: The Caste System and Its Implications, Translated from the French by Mark Sainsbury.* London: Weidenfeld & Nicolson.
Khilnani, S. (1997). *The Idea of India*. New York: Farrar Straus, and Giroux.
Kumar, D. and Desai, M. (eds.) (1983). *The Cambridge Economic History of India*, Vol. 2, pp c1757 c1970. Cambridge: Cambridge University Press.
Michel, A. A. (1967). *The Indus Rivers: A Study in The Effects of Partition*, 595p. New Haven and London: Yale University Press.
Nehru, J. (1961). *The Discovery of India*. New Delhi: Asia Publishing House.
Philips, C. H. and Wainwright, M. D. (1970). *The Partition of India: Policies and Perspectives 1935 1947*. London: MacMillan.
Rahula, W. (1967). *What the Buddha Taught*. London: Gordon Fraser.
Raychaudhuri, T. and Habib, I. (eds.) (1982). *The Cambridge Economic History of India*, Vol. 1, pp c1200 c1750. Cambridge: Cambridge University Press.
Robinson, F. (ed.) (1989). *Cambridge Encyclopaedia of India, Pakistan, Bangladesh and Sri Lanka*. Cambridge: Cambridge University Press.
Schwartzberg, J. E. (1992). *An Historical Atlas of South Asia*. New York: Chicago University Press. Oxford University Press.
Talbot, I. (1998). *Pakistan: A Modern History*. London: Hurst and Co.
Wallace, P. (ed.) (1985). *Region and Nation In India*. New Delhi: Oxford and IBH.

Relevant Websites

http://www.londonsocialisthistorians.org
London Socialist Historians Group.
http://www.saarc sec.org
South Asian Association for Regional Cooperation.
http://www.tufts.edu
Tufts University whose links provide a very useful summary and lots of useful references.

Southeast Asia

Eric C. Thompson, National University of Singapore, Singapore

Glossary

ASEAN Association of Southeast Asian Nations.
Association of Southeast Asia A regional cooperative grouping of Malaya, Thailand, and the Philippines, which predated the formation of ASEAN.
Entrepôt A transshipment port rather than final destination along a trade route.
Eurocentric Ideology and other processes that take Europe as their primary point of reference.
Geo-Body Territorial dimensions of a polity produced through cultural processes and technologies, particularly mapping.
Maphilindo A proposed regional confederation of Malaysia, the Philippines, and Indonesia, which predated the formation of ASEAN.
Plural Society A society composed of two or more culturally differentiated groups.
Polity A system of social–political organization, such as a city-state, nation-state, or kingdom.
Sanskrit Language of politics and scholarship through much of Southeast Asia in the first and early second-century CE; it originated in India.
SEATO Southeast Asia Treaty Organization.
Swidden A system of shifting cultivation in which land is prepared by cutting away and burning existing vegetation, commonly referred to as slash and burn.

Southeast Asia is a regional geographic construct. It refers to an area south of contemporary China and east of the subcontinent of contemporary India. The region of Southeast Asia is currently organized politically under the ten member Association of Southeast Asian Nations (ASEAN), which includes Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar (Burma), the Philippines, Singapore, Thailand, and Vietnam. These constructs and concepts of Southeast Asia as a region and ASEAN as a regional political forum are founded on a complex and contested spatial, physical geographic, social, and cultural processes.

Southeast Asia is currently more contested (less taken for granted) than some other regional concepts such as Africa, Asia, Europe, or Latin America. In many accounts, both past and present, Southeast Asia is presented as an artificial construct driven by exogenous geopolitics and cultural forces. Yet during the second half of the twentieth century, Southeast Asia has emerged as a locally relevant regional frame of reference, in ways that some other well known concepts, especially the Orient or Far East, have not.

Realism and Constructivism

In 1984, Donald Emmerson produced a key article on Southeast Asia, in which he contemplated whether the region should be considered more of a rose or a unicorn. The former, invoking Shakespeare's famous lines from *Romeo and Juliet*, implies that Southeast Asia is a name referring to something with an underlying reality not contingent on the process of naming or conceptualization itself. The unicorn, on the other hand, implies that Southeast Asia is primarily a product of creative imagination; only through the human cultural capacity to create order out of chaos does Southeast Asia come into being. In this latter case, the search for any underlying reality to Southeast Asia, apart from its cultural invocation in area studies writing or geopolitical maneuvering, is a chimera. Emmerson traces a process through which the imagined unicorn like Southeast Asia was gradually emerging as a rose like reality, a trend which has, if anything, consolidated since the 1980s.

Emmerson highlights the dominant and competing realist and constructivist approaches to Southeast Asia. The simplicity of this dichotomy covers not only a wide range of approaches, but also two very different aspects of realism and constructivism. On the one hand, each of these is a scholarly, theoretical standpoint. In this regard, realists favor descriptions and understandings of the region that focus on processes such as trade patterns, politics, geography, and the like that may or may not demonstrate Southeast Asian regionalism. Constructivists, from the standpoint of scholarship, focus on various more or less intentional or conscious processes through which Southeast Asia is brought into being – in other words, as a social and cultural construction. But constructivism and realism are also used in debates over Southeast Asia as an object itself, not in reference to the theoretical inclinations of particular scholars. Realism in this sense takes the region to contain an underlying, if evasive, unity while constructivism takes the region to be an area of relatively discontinuous cultural, economic, geographic, political, and social patterns woven together in various narratives of scholars and politicians more out of convenience than for any other reason. The distinction is important, since at least some scholars who are constructivists – emphasizing the social and cultural construction of reality – nevertheless take objects thus constructed, such as Southeast Asia, to be no less real than products of other processes, such as geological systems or trade networks. A fair amount of scholarship on Southeast Asia is also marked by strident rejectionism or enthusiasm, of those who either vociferously reject the notion of regional legitimacy or authenticity and others who with equal vigor defend the notion of Southeast Asia as both legitimate and authentic.

The realist–constructivist debate cuts across many disciplinary fields – from history to political science to geography. The disciplines most actively involved in debates over the status of Southeast Asia as a region have been history and political science (specifically international relations specialists), with some inter ventions from geographers, anthropologists, and others. Most scholars working in these latter disciplines have been more concerned with examining spatial and social processes within the region rather than the status of Southeast Asia as a region *per se*. Nevertheless, the area studies focus, especially dominant in American uni versities, has been important in framing research as Southeast Asian studies.

Physical Geography

Southeast Asia is conventionally divided into mainland and maritime subregions. Maritime Southeast Asia en compasses the Indonesian and Philippine archipelagos and the Malay Peninsula, as well as the coastal regions of contemporary Vietnam on the South China Sea, the Bay of Siam, and the Bay of Bengal. This coastal, maritime, archipelagic region has been a major thoroughfare over millennia for trade connections stretching from the East Asian regions of China, Japan, and Korea through to India, the Arab peninsula, Africa, and the remote regions of Europe. Mainland Southeast Asia is defined by a series of river basins bordered on the north and west by mountain ranges which created a natural barrier to human interactions. These features created a unique propensity toward land based isolation and sea based interconnections both within the region and between the region and the rest of the world. They also account for a second dimension along which the human and physical geographies of the region intersect; that of highlands and lowlands. Highlands have historically been home to populations with subsistence economies based on swid den cultivation considered outside of or marginal to the lowland state societies organized around irrigated agri culture and trade. The relationship between lowland and highland populations continues to be a major area of scholarship, especially in anthropology (see **Figure 1**).

Since at least the middle of the first millennium CE, a series of mainland and maritime based states have risen and fallen. The mainland states have been situated primarily in the fertile river valleys of the Irrawaddy, Chao Phraya, Mekong, and Red Rivers. The latter two rivers in particular link the areas of contemporary Southern China, especially Yunnan province, to main land Southeast Asia. In recent years, a new regional concept has gained some currency – the Greater Mekong Subregion, encompassing the five nation states of main land Southeast Asia and Yunnan province.

Physical geography did not determine the location and extent of state formation in a simple and straight forward fashion. But features of the mainland clearly contributed to the location and extent of various states, including current nation states. Moving across mainland Southeast Asia from west to east, it is relatively easy to discern the natural features that have facilitated and hampered political and cultural integration within the region. The Irrawaddy and Chao Phraya river basins have been fertile soil for a succession of polities and currently form the center of the contemporary states of Myanmar (Burma) and Thailand. Forerunners of Myanmar and Thailand have played out a long standing rivalry over many centuries, including numerous wars and border skirmishes, but neither has exerted sustained domination over the other. To the east of the Chao Phraya, the Khorat Plateau, bordered on the north and east by the middle Mekong, and the basin defined by the confluence great lake Tonlé Sap and lower Mekong, form two further natural lowland regions conducive to early state consolidation. The latter was the central site of the Angkorean empire, mainland Southeast Asia's most ex tensive, sustained and successful early polity, and now the territory of Cambodia. Contemporary Laos is situated on the northeastern portion of the Khorat Plateau, while the bulk of the plateau has been incorporated into Thailand. Historically, this middle region of the mainland has been the site of shifting, expanding, and contracting polities. Finally, much of the coastal strip of the mainland that is now the site of Vietnam was at an earlier point in history claimed by Champa, a constellation of more or less integrated ethno linguistically Malay polities. Further north, the Red River Valley was home to the ethno linguistically Kinh or Viet polity known as Dai Viet. For roughly the first millennium CE, the Red River basin was a direct vassal state of Imperial China. After achieving independence in CE 939, especially from the thirteenth to fifteenth centuries, the Dai Viet expanded down the length of the coast at the expense of the Cham states.

The mainland geography of mountains, highlands, river basins, and valleys insulated the region to some degree from overwhelming Sinic or Brahmanic influence, while the maritime geography connected the region to those neighboring civilizations. An important geographic feature of maritime Southeast Asia is the narrow Strait of Malacca, between the Malay Peninsula and the island of

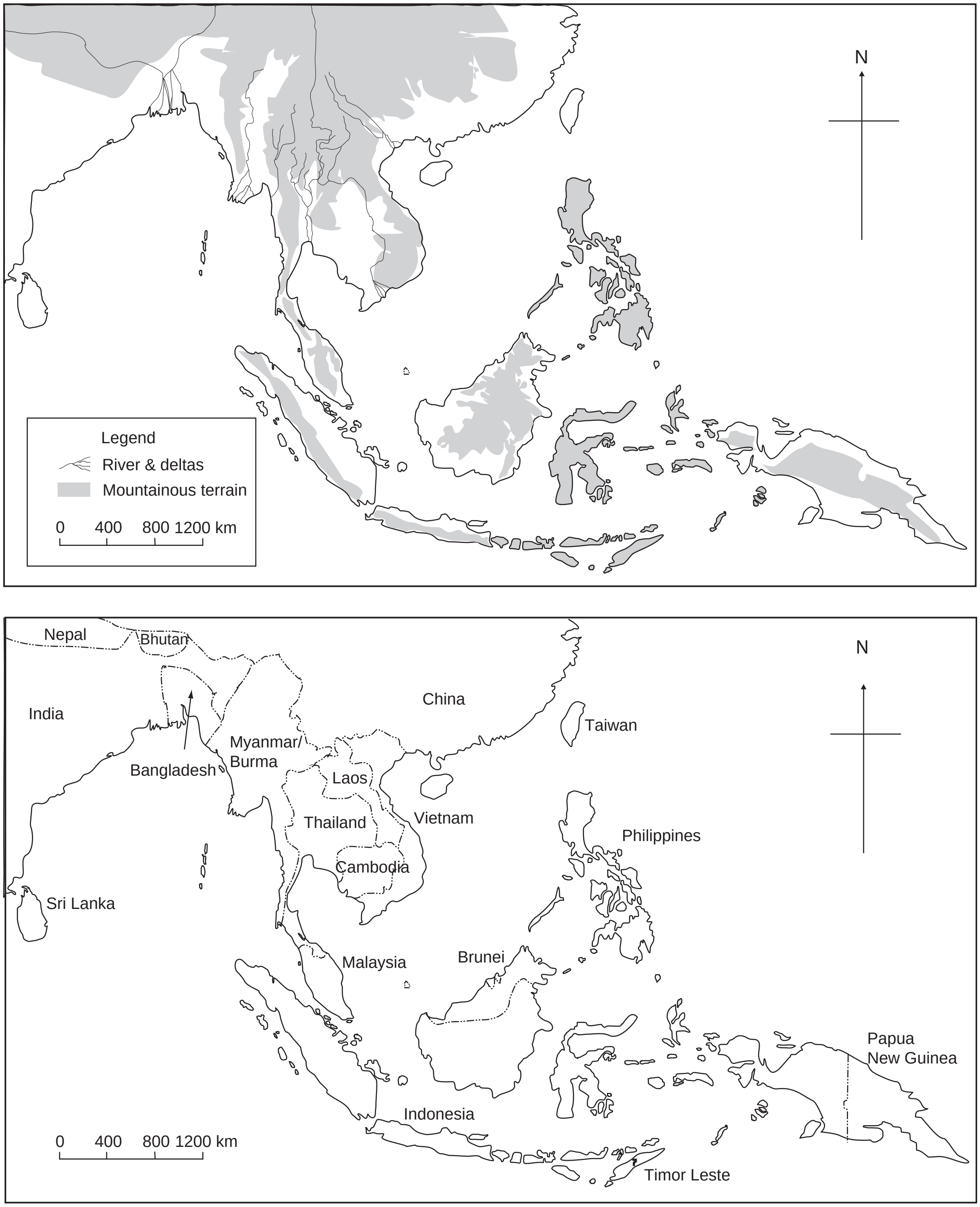

Figure 1 Two views of Southeastern Asia. From Kratoska, P. H., Raben, R. and Nordholt, H. S. (2005). *Locating Southeast Asia: Geographies of Knowledge and Politics of Space*. Singapore: Singapore University Press.

Sumatra. A succession of port cities have thrived on es tablishing themselves as the preeminent *entrepôt* for trade flowing through this choke point of Asian maritime commerce. The most sustained power being the Srivi jayan Empire centered on Jambi/Melayu on the East coast of Sumatra from the eighth through twelfth cen turies. Srivijaya was succeeded by Pasai, Malacca, and other lesser ports and later by competition between British and Dutch colonial powers contesting dominance over the Straits from either side of the waters. The current economic preeminence of Singapore in the re gion is a testament in part to its inheritance of the mantle of foremost Malaccan Straits *entrepôt.* The Malacca Strait ports were supplemented and interconnected with other important *entrepôts* between the Malay Peninsula and southern China as well as throughout the Indonesian archipelago.

This mainland maritime division of Southeast Asia was never a simple divide. Not only did trade form an ongoing basis of circulation of things, people, and ideas between mainland and islands, the archipelago was home to important inland polities, such as Mataram on the island of Java, and some mainland polities, such as the Ayutthaya located on the Chao Phraya river, were also maritime powers. But even with the regional systems and subsystems, conditioned at least in part by these features of the physical geography, Southeast Asia is still con strued mainly as a residual category, as the lands south China and east of India. The region is defined as much by connectivity, a maritime crossroad of the Asian continent, as it is by demarcating boundaries.

Historical Antecedents

There is an active and ongoing debate among historians regarding the coherence of the region over many cen turies or even millennia preceding its formalization in the past hundred or so years. Anthony Reid has argued that important intraregional connections can be identi fied especially in what he periodizes as the "early mod ern" period, or "age of commerce" from the fifteenth through seventeenth centuries. Reid argues that this was a historical era when Southeast Asia was operating as a regional sociocultural system, subsequently disrupted by the disintegrative effects of European colonialism under which colonial powers carved up the region and inhibited interactions among the indigenous populations. Reid's thesis has been challenged by Victor Lieberman, who argues that Reid's thesis may work for maritime or coastal Southeast Asia but is less applicable to the mainland. In Lieberman's view, Reid also privileges intra Southeast Asian connections at the expense of a more complex set of broader global and Eurasian processes. More generally, the work of Reid and others has been criticized for projecting a twentieth century perception of Southeast Asia backward into the sixteenth century.

A useful historical antecedent for thinking about contemporary Southeast Asia is *Suvarnabhumi*, a Sanskrit term, meaning literally the golden land. References to *Suvarnabhumi* throughout Southeast Asia date to the first and early second millennia CE, described by Sheldon Pollock as an era of Sanskrit cosmopolitanism. Polities of Southeast Asia, such as Angkor, Pagan, Ayutthaya, Sri vijaya, as well as a multitude of smaller entities were culturally interrelated through the Sanskrit lingua franca which carried with it a loosely integrated set of political and cosmological beliefs related to Brahmanic traditions later to become known as Hinduism and Buddhism. A limited amount of historical (written) documentation and a greater wealth of archeological evidence demonstrate that a substantial circulation of goods, people, and ideas bound together (at least loosely) a region stretching from the Cham principalities of what is now south and central Vietnam to the Indian subcontinent and through the Indonesian archipelago and mainland Tai, Khmer, and Bamar polities (where Thailand, Laos, Cambodia, and Myanmar are now located). In this era, the Red River Delta was under direct Chinese political control and cultural influence and the Philippine archipelago was largely outside of the maritime flows of commerce and culture.

Suvarnabhumi remains a poetic inheritance in many places in contemporary Southeast Asia, reappearing in forms such as Charles Keyes' *The Golden Peninsula*, the name of the new international airport in Bangkok, and references in the National Museum in Kuala Lumpur, where it is translated somewhat parochially as *tanah Melayu* (the Malay lands). But any sense of *Suvarnabhumi* as a substantial geopolitical region did not survive the disintegrative effects of European colonialism. With Islamization of maritime Malay polities from about the eleventh century onward, any regional sense of *Suvarnabhumi* was probably already very much on the wane well before the colonial era and in any event was never even remotely established to the extent that the concept of Southeast Asia became in the late twentieth century.

Nevertheless, *Suvarnabhumi* remains an instructive comparison to contemporary Southeast Asia, as its co gency derived from many analogous processes (insofar as it did become a relatively widespread signifier of a region stretching from contemporary Vietnam to the Indonesian archipelago and through the mainland Tai and Bamar polities – where Thailand and Myanmar are now lo cated). The diffusion of the *Suvarnabhumi* concept rested on the diffusion of Sanskrit as a privileged language of religious and political discourse in the region, much as English is the preeminent *lingua franca* today. This was also the basis for intraregional political relations between polities and individuals who not only could communicate between mutually intelligible languages, but implicitly comprehended a political grammar and syntax of polities organized on principles described by Stanley Tambiah as "galactic polities" and O.W. Wolters as "mandala states."

As useful as *Suvarnabhumi* may be for thinking about some of the region forming processes, such as language, politics, cosmology, commerce, and the like, Southeast Asia as now understood is undoubtedly the cultural product of the past century.

Convenient Constructs and Contingent Devices

The use of Southeast Asia as a term to designate the region south of China and east of India appears only in idiosyncratic references prior to the twentieth century. From the sixteenth century onward, European colonial powers laid claim to their territories with terms such as the Netherlands East Indies and French Indochina. Only well into the twentieth century do we find Southeast Asia and variants used with regularity in different languages, including for instance *Asia Tenggara* in Malay and Indonesian, *Asia Akhane* in Thai and Lao, and *Đong Nam Á* in Vietnamese.

The Allied South East Asia Command established in 1943 during World War II is often cited as the earliest widely influential use of Southeast Asia. However, Shimizu Hajime has shown that by 1919, the term *Tonan Ajiya* (literally Southeast Asia) had entered Japanese geography curriculum in modern state schools. Hajime argues that the displacement of earlier terms such as *nanyo* (South Seas, derived from the Chinese *nanyang*) played an important conceptual role in guiding Japan's military colonial expansion into the region in the subsequent decades. Hajime and others suggest that the Allied concept of Southeast Asia, therefore, was largely a reaction to Japanese imperial ideas and actions.

Following World War II, the idea of Southeast Asia received further support with the formation of the Southeast Asia Treaty Organization (SEATO), established as an American initiative to counter Chinese and Soviet influence in the region in the context of the Cold War. But from within the region, only the Philippines and Thailand were members of SEATO (others being Australia, France, New Zealand, Pakistan, the United Kingdom, and the United States). The organization clearly had more to do with American Cold War activities than the ideas or interests of governments or people living in the region.

A very strong, if not dominant, stream of scholarship on Southeast Asia since the inception of the term in English language discourse has focused largely on the SEATO like constructions of the region. That is, the forces and ideas acting largely on the region from without, based on extra regional interests and motives with only marginal interest or input from individuals, institutions, organizations, or states within the region. Likewise, much attention has been given to the establishment and role of Southeast Asian Studies in the United States particularly in the context of the Cold War, a central focus of Emmerson's important article.

In a 2005 article, Jim Glassman provides a useful review of this sort of external, Cold War driven conceptualization of Southeast Asia, focusing especially on English language scholarship in the discipline of geography. Glassman compares the environmental determinist approach of Charles Fisher to the pragmatic constructivist approach of Donald Fryer, a dichotomy played out in various ways across English language social science and humanities scholarship. Glassman's approach also typifies the stream of scholarship that treats Southeast Asia as tinged with inauthenticity and largely ignores the role of governments and people within the region – who are acted upon, but are not themselves agents of the ideological and social processes constructing Southeast Asia.

Given the prevailing dominance of English medium cosmopolitanism over the past century, English language scholarship on Southeast Asia has been particularly powerful in shaping recent ideas about the region. One of the more significant effects is that English cosmopolitanism is plagued by a concern for authenticity of the other; that is to say authenticity of the native, conceived of as standing outside of the space and time of most English language scholarship. Implicit standards of native purity lurk in the subtext of this scholarship; with the effect that the 'polluting' effects of foreign ideas – in particular the effects of English cosmopolitanism itself – are taken as signs of inauthentic regionalism. When such signs can be discerned, Southeast Asia is cast as a somehow less than worthy cultural construct.

Critiques of regionalism – of scholarship on the subject and of active diplomacy and interactions – are at their most vigorous carried off with a combination of cynical realism and rational choice reductionism. Somewhat inexplicably, these arguments particularly in the realm of international relations scholarship take nation states (which did not exist a century ago) to be real actors with real interests while casting Southeast Asian regionalism as an elite Panglossian fantasy. In a similar vein, many of the same authors as well as others insist that forces external to Southeast Asia – such as Cold War superpower actors and American based scholars – must be understood as substantially instrumental in the formation of Southeast Asian regionalism, while intraregional initiatives are to be understood as insubstantial, fleeting, and illusory.

Realizing Southeast Asia

Given the Anglo American centered, English medium biases of scholarship over the past several decades, the most underappreciated and discursively dismissed conditions of Southeast Asian regionalism are the initiatives undertaken from within the region. Amitav Acharya's succinct analysis of Southeast Asian regionalism published in 2000 made an important contribution to addressing this lacuna by developing the story of Southeast Asian regionalism with a focus on intraregional actors, but with less overt advocacy than found in much of the writing of ASEAN enthusiasts. Acharya's thesis is that Southeast Asia has in many respects been realized – brought into being – through a combination of broad political, economic, social, and cultural conditions of the late twentieth century and the active role of Southeast Asian politicians and diplomats. He traces the cultivation of Emmerson's rose, but with much greater attention to the local gardeners.

In the mid twentieth century, the fraught and con tested process of decolonization entrenched nationalism as the basis for new polities in the region. At the same time, European colonial powers were supplanted by Cold War superpowers of the United States, the Soviet Union, and China, each with their own designs on the region. While not dismissing the substantial effects of Cold War geopolitics on the region, the significance of Acharya's thesis is to undermine the overbearing, Washington, Moscow, or Beijing reading of the region as a theater in which their own passion play was performed, with local petty lords as bit players. Acharya reads the field instead from the point of view of the motives, desires, and ac tivities of local actors, with the Cold War geopolitics, the retreat of Europe, and the rise of America, China, and the Soviet Union a backdrop. From within Southeast Asia, the master narrative is one of decolonization and the rise and normalization of nation states, not the global ma neuverings of Cold War superpowers.

From the post World War II era through the 1990s a range of initiatives by actors within Southeast Asia aimed at establishing linkages transcending Cold War divisive ness. In 1946, for instance, Acharya notes that Ho Chi Minh was writing to the Prime Minister of Indonesia in an initiative to develop a "Federation of Free Peoples of Southern Asia" to include Vietnam, Indonesia, India, Burma, and Malaya. But in the early post war period, when much of the region was involved in violent strug gles over decolonization, an inchoate sense of regional distinctiveness was weak at best. Rather, Ho Chi Minh, Sukarno, and others had grander visions of broad Third World alliances. These initiatives undoubtedly reached their zenith in the 1955 Afro Asian Conference in Bandung. Such attempts at broad alliances waned in subsequent years, as their shear scope made them dip lomatically, strategically, and politically unworkable.

In the 1960s and 1970s, the first substantial initiatives toward a more modest and Southeast Asia focused internationalism were launched. These included the Association of Southeast Asia spearheaded by Malaysia's Tunku Abdul Rahman and Malphilindo (Malaysia, Philippines, Indonesia) initiated by Philippine President Diasdado Macapagal. While each of these ideas was short lived, they set the stage for the August 1967 establish ment of the ASEAN.

ASEAN was composed of the pro capitalist and anticommunist governments of Indonesia, Malaysia, the Philippines, Singapore, and Thailand. In the first decades of its existence, it was commonly looked upon as a pro capitalist block at odds with the Vietnamese dominated pro communist region of Indochina (including Laos and Cambodia) and with socialist Burma. Here again there is cause to see some dissonance between American and other external Cold War readings of the region and the primary concerns of actors within ASEAN and Southeast Asia. The anticommunist concerns of the ASEAN gov ernments had far more to do with internal struggles for power in the context of making new nations than with ideological and geopolitical concerns of distant Cold War superpowers. The ASEAN governments were concerned with the countries of Indochina (and perhaps more so China) as communist states, but primarily to the extent that Vietnam or China might give support to the internal communist movements in each ASEAN country. From its inception, ASEAN had the aim to incorporate all of the countries considered geopolitically within South east Asia and its chief driving ideology was non interference between territorially defined nation states, not anticommunism.

From its founding through the 1990s, ASEAN did much to solidify the concept of Southeast Asia. The nation state came to define both the limits and foun dations of Southeast Asian regionalism. ASEAN de veloped as a mapped geo body, in much the same way that its constituent members developed as territorially recognized polities, by cultural and political processes described in great detail by Thongchai Winichakul for Thailand. Various ambiguities were worked out, such as the status of Ceylon (Sri Lanka) – considered for inclu sion in ASEAN at an early date, but eventually denied membership. And by the end of the 1990s, with the rapid dissolution of Cold War conditions and just as import antly the dissolution of serious internal communist threats with member states, ASEAN came to include its current ten member states, incorporating Indochina and Myanmar (Burma). Of course, such historical closure is always subject to ongoing changes; the ambiguous status of newly independent Timor Leste (East Timor) being the most obvious example. Even at its most complete, the contemporary ASEAN based notion of Southeast Asia remains contested within and beyond the region.

One of the most common bases of discursive dismissal (in frequent and almost reflexive discounting of local initiatives) is the Eurocentric concept of political re gionalism as nation state writ large. This is not just vaguely Eurocentric, but the very specific tendency to take Europe as the model for all political regionalisms. Europe and the political entity of the European Union is premised on an ethno racial identity that is more than analogous to the idealized ethno racial identities which (often problematically) underwrite the political legitim acy of nation states. Europe and the European Union are taken as prototypical of what a region should be and any entities that lack the characteristics of this prototype are then found wanting. This is grossly evident throughout assessments of Southeast Asia and ASEAN in particular.

The ASEAN community does not replicate a nation state writ large so much as a plural society writ large. The strengths and weaknesses, possibilities, and limits of ASEAN reiterate those of plural societies more so than those of nation states. It is worth recalling that the pre eminent theorist of the plural society, J. S. Furnivall, was himself casting it as an unworkable construct of colonial machinations, in which the ability of natives to forge the common will of a European type national society was thwarted by cultural divisions fostered by and for the benefit of European colonialists. At the same time, the conditions identified by Furnivall became political real ities for postcolonial states and fostered multiculturalist political ideologies, such as Indonesia's 'Unity in Diversity'.

In Acharya's work and that of others of similar in clination, there is ongoing ambivalence toward the origins and legacy of Southeast Asian regionalism. It is possible – as Reid and others have done – to trace historical con tinuities into the present from the millennia old patterns of commerce and cultural diffusion, not to mention the physical geography and environment. Yet at the same time, it is easy to see how such sense of region is always contingent; a reality made and unmade through systemic political, economic, and social forces as much as through intentional human agency and imaginative practices. Southeast Asia has been an imaginative geographic construct for Westerners, Japanese, Chinese, and Soviets. Embodied at present in the form of ASEAN, Southeast Asia has also been substantively realized through in digenous imaginative practices. But realization at present is not the end of history. Moreover, Southeast Asia cum ASEAN remains as much an aspirational construct as a realized entity. In parallel to the concept of 'nations of intent', developed by Shamsul A. B. to describe Malay sian nationalism, we could likewise consider ASEAN as a 'region of intent'. Southeast Asia is as much a process as a product or an object.

See also: Asia; Ethnicity; Geopolitics; Physical Geography and Human Geography; Place, Politics of; Political Geography; Regionalism; South Asia.

Further Reading

Acharya, A. (2000). *The Quest for Identity: International Relations of Southeast Asia*. Oxford: Oxford University Press.

Anderson, B. (1998). *The Spectre of Comparison: Nationalism, Southeast Asia and the World*. London: Verso.

Emmerson, D. K. (1984). Southeast Asia: What's in a name? *Journal of Southeast Asian Studies* 15(1), 1 21.

Glassman, J. (2005). On the borders of Southeast Asia: Cold War geography and the construction of the Other. *Political Geography* 24(5), 784 807.

Kratoska, P. H., Raben, R. and Nordholt, H. S. (2005). *Locating Southeast Asia: Geographies of Knowledge and Politics of Space*. Singapore: Singapore University Press.

Lieberman, V. (2003). *Strange Parallels: Southeast Asia in Global Context, c. 800 1830. Vol. 1: Integration on the Mainland*. Cambridge: Cambridge University Press.

Pollock, S. (1998). The cosmopolitan vernacular. *Journal of Asian Studies* 52(2), 261 297.

Reid, A. (1988). *Southeast Asia in the Age of Commerce 1450 1680, Volumes 1 and 2*. New Haven, CT: Yale University Press. [Reprinted 1993, New Haven, CT: Yale University Press.]

Rigg, J. (1997). *Southeast Asia: The Human Landscape of Modernization and Development*. London: Routledge.

Shamsul, A. B. (1996). Nations of intent in Malaysia. In Tonnesson, S. & Antlov, H. (eds.) *Asian Forms of the Nation*, pp 323 347. London: Curzon Press.

Winichakul, T. (1994). *Siam Mapped: A History of the Geo Body of a Nation*. Honolulu: University of Hawaii Press.

Wolters, O. W. (1999). *History, Culture, and Region in Southeast Asian Perspectives (Revised Edition)*. Ithaca: Southeast Asia Program Publications, Cornell University.

Sovereignty

M. Coleman, The Ohio State University, Columbus, OH, USA

Glossary

Critical Geopolitics Study in political geography of geographical assumptions embedded in discourses about world politics, and how these enable particular stories to be told about the shape and substance of world politics.
Deterritorialization Undoing and/or remaking of particular territorialities.
External Territoriality Use of borders and boundaries to control movement and restrict access to a territory.
Internal Territoriality Creation of regulations to manage how resources are used or how people act within a territory.
International Relations Subdiscipline of political science dedicated to the study and understanding of world politics.
Power Entangled and sometimes incoherent social relationships of authority, domination, coercion, consensus, compulsion, confession, etc., may be proximate or distant.
Realism Perspective within international relations which treats states and their national security interests as the most important constituents of world politics, major focus is on the state as a rational and unitary entity.
Spatiality The organization of power and social relations through and across space, place, and territory.
State Government over some territory which claims a monopoly on the exercise of force, as well as overriding political and legal authority, typically thought to be centralized.
State Territoriality Assumption that state accords to a discrete and carefully demarcated territory.
Territoriality Strategy of classifying, communicating, and enforcing control over space used by individuals, groups, and organizations.
Territory General term to describe a chunk of space occupied by a person, group, or state.

Introduction: Final and Absolute Legal–Political Authority

That sovereignty is a commonplace yet systematically under investigated concept in political geography and international relations (IR) has led many contemporary critical scholars concerned with the spatiality of power to defer the search for a decisive definition of the term. Their point is important, and bears careful consideration: that sovereignty does not submit neatly to a logic of representation, in which analytical concepts are simple and surefire referents of things in the world. Rather, sovereignty discloses a quite different set of intellectual practices concerning the stabilization and inheritance of knowledge, which some scholars refer to as 'disciplinarity'. From this perspective sovereignty is literally about the setting of limits to social practice – whether in the form of statecraft or scholarly inquiry about statecraft – via epistemological appeals about how best to know and act in the world. Among other things this suggests that one must not necessarily look to shore up a definition of sovereignty, as if an affirmation of something 'out there'. Rather, the task is to understand how, for what ends, and in what contexts the term is used and expounded upon, as well as attend to the work done by it.

This said, we can still attempt a definition, or at least review how the concept has been used by others. Sovereignty is typically held as the exclusive, independent, and transcendent right of an entity (i.e., deity), person (i.e., monarch, citizen, dictator), or political body (i.e., parliament, legislature, executive, judiciary, the nation, community) to exercise ultimate legal–political authority over an indivisible geographic area and its attendant population, conceived as an amalgamated body politic. This is the definition favored in E. H. Hinsley's classic on the topic, *Sovereignty*. Indeed, Hinsley describes sovereignty as the final and absolute political authority in a community for which there is no other source of final and absolute authority. Community and sovereignty are, in this reading, expressly statist terms. For example, Hinsley attributes the development of sovereignty as an intellectual construct to 'primitive' and 'stateless' peoples' abandonment of homogeneous 'tribal groups' and their gathering into state collectivities. His claim is that this new constellation of power demands an answer to the question of the location and nature of political authority, which sovereignty ultimately provides. (This account is corroborated by Claude Lefort who argues similarly that sovereignty is an attempt to fix states' political authority in a certain transcendent location. His point of departure is that this attempt is in the end doomed and that the result is a totalitarianism of one sort or another.) From this standpoint, then, states – of all stripes and colors, democratic, monarchical, theocratic, aristocratic, authoritarian, etc. – can be said to be sovereign if they enjoy unrestricted and effective authority over all peoples and places within a discrete and unified territory, absent interference from external interests or forces.

Classical liberalism speaks slightly differently of sovereignty. The emphasis here is on 'individuals' as free, equal, independent, knowledgeable, self interested, indivisible, primarily private, minimally obliged, self sufficient, and socially unencumbered political subjects. From this perspective, sovereignty is about a person's intrinsic self determination, as well as their self possession of inalienable rights, with respect to other similarly self governing subjects. However, this reading of sovereignty is not totally distinct from the long and complicated history of scholarship of which Hinsley's book is a part. Indeed, there are important connections between classical liberalism's definition and sovereignty as is perhaps more familiarly encountered in Anglo American scholarship on power as a defining attribute of the modern interstate system, and specifically of the states that together comprise this system. For example, sovereignty in mainstream research on states tends to reproduce classical liberalism's atomistic individualism in their discussion of what constitutes the essence of states. Sovereignty in much political geography and IR scholarship on states is frequently analogized to individual sovereignty, as if the former was a scaled up version of the latter. Indeed, neorealist scholarship in IR makes exactly this link, and proceeds as if states are coherent, individual, and unitary actors. This accounts for why sovereignty – as in state political authority – is frequently described as neither episodic nor ruptured, but as emanating from a center point and extending in a continuous and smooth fashion to the edges of the state, where it abruptly ends.

Sovereignty as Supreme Command, Sovereignty as Political Authority

Sovereignty raises some conceptual question marks in part because of the uneasy ways it conjoins power with authority. On the one hand, for example, sovereignty has been elaborated as a sort of transcendent command or injunction to obey. Jean Bodin (1530–96), in his *Les Six Livres de la République*, which was written in the aftermath of Europe's sixteenth century religious conflicts and which in essence substantiated the power of the French king against his many challengers, put forth what is generally considered to be the inaugural definition of sovereignty along these lines. For Bodin, sovereignty was the unfettered power to make laws and exact obedience to them from a population, regardless of consent. Sovereignty here translated as a matter of compelled and centralized authority in the head of state – a power bestowed by office but no less absolute.

However, if Bodin understood *souveraineté* to be in the end limited by customary, natural, and divine law, it was Thomas Hobbes (1588–1679), writing during the turmoil of the English civil war, who proposed a theory of unlimited sovereignty. Treated as a key building block in sovereignty's intellectual edifice by realist IR scholars in particular, Hobbes articulated a pessimistic and competitive view of human beings. He described, in the *Leviathan*, a hypothetical 'state of nature' which was, as he famously put it, "solitary, poor, nasty, brutish and short." Hobbes argued that individuals, however primarily self interested, might exit this presocietal condition of conflict and turmoil by entering into a social contract or covenant governed by some overseeing authority. Hobbes called this authority the Leviathan – an 'artificial person' – to whom individuals could cede their rights in return for a peaceful existence. The Leviathan (supreme ruler) was to maintain order and peace through a monopoly on coercive authority or what has been referred to as a 'right of the sword', clearly displayed on the frontispiece to *Leviathan* (see **Figure 1**). Importantly, Hobbes posed absolute command as the only alternative to anarchy.

This tradition of sovereignty as supreme command has influenced many thinkers, including Weimar Republic era political theorist and jurist, Carl Schmitt. As

Figure 1 Frontispiece to Thomas Hobbes' *Leviathan* (1651).

he wrote in *The Leviathan and the State Theory of Thomas Hobbes*, post World War I liberalism's emphasis on human rights and constitutional reform sapped the state of its sovereign authority and turned it into a mere association among myriad other private groupings. Against this 'le galization' of what he understood as sovereign legitimacy and order, Schmitt articulated a strong definition of sovereignty as the power to decide on the exception, or to act if need be extrajudicially and without consultation through emergency powers to stall civil conflict or otherwise safeguard the erosion of the public sphere.

On the other hand, sovereignty has been explained differently as political authority. John Locke (1632–1704) objected that a sovereign's authority could be considered absolute. In *Two Treatises of Government*, Locke put forth that legitimate government was a revocable authority bestowed by the ruled on the rulers in order to facilitate and safeguard an individual's rights of liberty, health, property, and life. In comparison to Hobbes this repre sented a sort of decentralization of sovereignty, as well as an argument against the wholesale transfer of political authority from individuals to a princely figure standing above and apart from them. Jean Jacques Rousseau's (1712–78) objection to sovereignty as supreme command was stronger still. In *The Social Contract* Rousseau wrote with Hobbes that political authority was something conferred to the sovereign by the people. However, against Hobbes Rousseau concluded that sovereignty – 'popular' sovereignty, that is – was derived from the people's 'general will' and as such could not be given up to the sovereign in the manner sketched out in the *Le viathan.* For Rousseau this meant that political authority could never be forcibly estranged from the people. As a result he claimed that people should deliberate together on matters of government, participate directly in the production of law, and moreover should not be obligated to laws which are deemed to contravene the common good.

Opposition to sovereignty as unrestrained and limit less command finds a more recent home in arguments made by Hans Kelsen, twentieth century Austrian legal theorist and legal advisor to the United Nations War Crimes Commission. Kelsen, Schmitt's principal intellectual antagonist during the 1930s, championed the curtailment of a narrowly conceived state sovereignty – which he understood as the conditions of possibility for imperialism – through the institutionalization of parlia mentarism and constitutionalism on the domestic front, as well as through a pacifist global legal order to which state governments would ultimately be accountable. It was exactly Kelsen's reconfiguration of sovereignty in relative terms – checked 'at home' in the form of parliamentary democracy and in account 'abroad' to some overarching juridical structure, such as the United Nations – that Schmitt understood as itself imperialistic, designed to shift authority from a transcendent sovereign seat above the social world with which a territorial col lectivity could affiliate, to multiple, disconnected private seats defined by uninterrupted social contest and material labor.

The point here – well summarized in the clash be tween Schmitt and Kelsen – is that sovereignty is defined in the canon as both supreme power and political au thority, and that these competing translations are not without their points of conflict. However, in general we can say that today sovereignty is referenced in main stream political geography and IR as a condition of 'political authority and legitimacy' rather than as an ex ercise of unconditional and supreme power. (This does not include a resurgent critical literature on sovereignty's states of exception, which we will get to below.) Political scientist David Held, for example, summarizes sover eignty in his *Democracy and the Global Order* as an "im personal and privileged legal or constitutional order, delimiting a common structure of authority, which spe cifies the nature and form of control and administration over a given community."

Sovereignty and State Territoriality

It should be obvious from the comments above that sovereignty in political geography and IR, defined in either way as supreme command or political authority (or indeed as some combination of both), is regularly defined in explicitly statist terms and as such is used – often with limited reflection, or at least on the presumption that the concept is stable across time and space – to describe the location, scope, and efficacy of political order 'in and by states'. The result is ultimately a state territorial template for mapping political power in the global political economy. In this context it makes sense to speak ex plicitly of sovereignty qua state sovereignty. Indeed, sovereignty as conventionally defined is held as syn onymous with 'sovereign statehood', the 'geography of state power', the 'state', the 'nation', the 'nation state', as well as 'centralized state authority', which as a result are terms too commonly collapsed into one another. That sovereignty – that is, political authority – might not be an exclusive property or characteristic of the territorial state or a community of territorial states is too seldom con sidered, as is the possibility of a political authority or power exercised nonterritorially or through a scattered network like structure of nodes and flows.

Typically, the advent of sovereignty qua state sover eignty is dated to the 1648 Peace of Westphalia, which put an end to the Thirty Years' War in Europe. The idea and practice of an overarching territorial authority was not novel to 1648, however. The prior English and French experiments with absolutism, as well as the free cities of late medieval Europe all disclosed a territorial ordering of political authority. However, as Steven Krasner argues in his *Sovereignty – Organized Hypocrisy*, the conventionally held opinion in IR (and to this we can add political geography) is that the princely authority over religion (*cuius regio, eius religio*) occasioned by the Westphalia treaty marked a watershed transition to a new geography of noninterventionist and absolute political power – often dubbed 'Westphalian sovereignty' or 'the Westphalia model'. The result was none other than the undoing of a feudal landscape of poorly delineated, nonmonopolistic, overlapping, fragmented, noncontiguous, multiply constituted and nonexclusive power networks (i.e., through parishes, manors, guilds, etc.) and their replacement with a modern structure of political authority rooted in the rule and apparatus of the centralized territorial state. As such, the Westphalian model amounted to an exclusive agglomeration of political authority in delimited state spaces. Indeed, the often told story is that the treaty's emphasis on political and religious self determination, nonintervention, and formal legal equality spelled out, for the first time, sovereignty as specifically a state's twofold ability to: on the one hand, 'containerize' or rule over a territorially captive population through an effective 'inward looking' internal political organization of state space ('internal sovereignty', 'intraterritoriality'); and on the other hand, to exclude or limit the influence of other states on this newly circumscribed population through an effective 'outward looking' external bounding of sovereign state space based on reciprocality, or the mutual recognition of state authority by other states ('external sovereignty', 'interterritoriality').

Of course, in practice Westphalia by no means instantaneously constituted a fully formed and unchanging sovereign state 'chessboard' of domestic and world politics. As many political sociologists have argued, the territorialization of political authority in states in the aftermath of the seventeenth century proceeded in fits and starts via state elites' coercive and consensual organization, communication, and enforcement of control over socioeconomic and political space. The three centuries after 1648, for example, saw the gradual territorial organization or bundling of political authority in states, through conscription, the waging of war, taxation, and the organization of markets, among many other things. This period also witnessed the piecemeal making and breaking of sovereignty as a principle of interstate relations, with war a frequent violation of the principle of noninterference. The upshot, as Alec Murphy has usefully mapped out (**Figures 2** and **3**), is that if state sovereignty as a solution to the where and what of political authority 'within' states was more or less steadily incremental in the 300 plus years after Westphalia, then certainly sovereignty as a principle governing relations 'between' states waxed and waned over the same period. These historical trajectories – which pose sovereignty as a social practice, as well as an ideal typical spatialization of political authority – mean a quite significant shift of emphasis from sovereignty as a straightforward referent of something given in the world to a sociogeographical analysis of the contingent and context specific techniques of government underpinning the measured territorialization and reterritorialization of sovereign political authority and power in states after 1648.

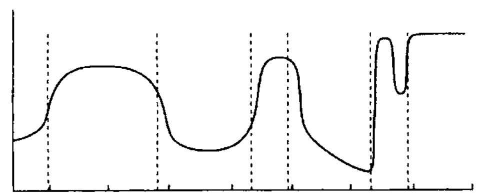

Figure 2 Dominant views of the nature of sovereignty as a principle governing relations between states, as seen from the core of the sovereign state system. Reproduced from Murphy, A. (1996). The sovereign state territorial system as political territorial ideal: Historical and contemporary considerations. In Biersteker, T. J. and Weber, C. (eds.) *State Sovereignty as Social Construct*, pp. 81 120. Cambridge: University of Cambridge Press, p 88. Reprinted with the permission of Cambridge University Press.

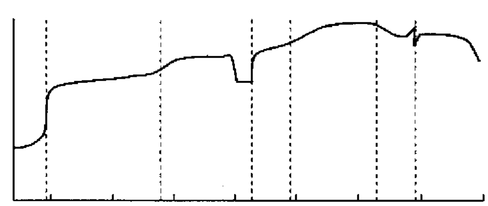

Figure 3 Dominant views of the appropriate political territorial structure for international society, as seen from the core of the sovereign state system. Reproduced from Murphy, A. (1996). The sovereign state territorial system as political territorial ideal: Historical and contemporary considerations. In Biersteker, T. J. and Weber, C. (eds.) *State sovereignty as social construct*, p. 81 120. Cambridge: University of Cambridge Press, pp 89. Reprinted with the permission of Cambridge University Press.

Sovereignty, Security, and the Limits of Political Community

Sovereignty qua state sovereignty is an inherently geographical problematic, if by geography we in part mean attention to spatial division and delimitation. This is so because sovereignty qua state sovereignty describes the conditions of possibility for a plurality of formally distinct and nonoverlapping territorial blocs of space containing distinct peoples and places. In other words, sovereignty qua state sovereignty makes at least theoretically possible the partitioning or bracketing of global space into singular, fixed, and differentiated parcels of political authority.

As pointed out by critical IR scholars in the late 1980s and as reiterated by critical geopolitics researchers in the early 1990s, this spatialization of political authority in the territorial state places starkly bordered limitations on the scope and substance of political community. Due to its emphasis on the divisibility and exclusivity of political rule and authority across space, the concept of sovereignty by definition circumscribes the spatial and temporal horizons of political community and belonging to life within state borders. As a result, politics (democratic or otherwise) is made conditional on the physical integrity of state borders, as well as on mechanisms of state surveillance and policing over territorially enclosed populations. (This is precisely the logic that allows contemporary theorists of world politics to divvy the world up into so called 'tame' and 'wild' zones comprising, respectively, orderly democratic states, as well as their 'collapsed' or 'failed' counterparts, who are unable to control their populations, as well as what crosses their borders.) From this standpoint, as R. B. J. Walker points out in his *Inside/Outside – International* Relations *as Political Theory*, sovereignty in standard political geography and IR research is best approached as a philosophical or political theoretic articulation which, in setting apart a 'thick' domestic sphere from a 'thin' international sphere, not so subtly consecrates a very historically and geographically specific account of the possible 'wheres' and 'whats' of political identity and community.

This commitment to sovereignty qua state sovereignty is not, however, unique to political geography or IR research. Rather, the assumption that sovereignty qua state sovereignty is the grounds for political community that constitutes much of the contemporary scholarship on democracy in political theory. Many liberals, for example, presuppose that rights bearing political subjects, while pursuing comprehensive moral projects in the private sphere, adhere to constitutional rules and procedures brought to bear throughout a sovereign statist public institutional–political sphere. Communitarians too, although typically counterposed to liberals in terms of their emphasis on goods and not rights, adhere to a similar modeling of sovereignty qua state sovereignty. Communitarians' conceptualization of membership, for instance, is typically imagined in terms of a conventional state territorial geography of political authority and legitimacy, such that belonging and entitlement are often imagined as dependent on a prior condition of geographical closure, as well as on the nonrecognition of political and ethical claims by subjects beyond state borders. It is perhaps only a relatively recent school of radical cosmopolitan democrats who have queried sovereignty qua state sovereignty as the geographical basis for politics, and who have insisted instead on a less grounded and more networked modeling of political community and democratic engagement.

The work done by this pervasive domestic/foreign ordering of places – that is, by sovereignty's mapping of an indivisible civic culture on the inside (deserving of the moniker 'politics') and mere *raison d'état* on the outside – accounts, among other things, for the moral differentiation between violence against 'citizens' (us), deemed illegitimate, and violence against 'others' outside the state (them), deemed unfortunate but necessary, in the name of national security. The result is most often a strict territorialization of empathy to the 'inside' and danger and incomprehension to the 'outside' – a common geopolitical theme in the US after September 2001. The 'war on terrorism', for example, has in large measure been justified by a sovereign language of internal indivisibility/comprehension and external dangerousness/incomprehension.

That sovereignty in this way 'enables' conflict has lead political geographers to roundly criticize the prevailing assumption that sovereignty provides for security. Indeed, sovereign state territoriality is frequently held in security studies and elsewhere as a 'prerequisite' for politics – a protective bulwark which holds an anarchic and lawless 'outside' at bay such that politics can proceed in an orderly fashion on the 'inside'. Feminist geopolitics scholars in particular have argued compellingly that state security and human security (i.e., access to food and shelter, public safety, meaningful employment, and education, etc.) are not analogous, and specifically that the latter is not always – if ever – provided for by the former. Their point is that the provision of security for select interests within the state, for example, via war, by necessity implies the subjugation of people, livelihoods, and communities elsewhere, and thus that the practice of national security is as such directly a condition of possibility for human insecurity. Other scholars have also pointed out, in a corroborative mood, that the prevalent sovereign state territorial reading of security 'inside' and danger 'outside' conveniently forgets about geopolitical violence and injustice meted out within the state, for example, against indigenous populations.

Spatialities of Power, Plural

Although sovereignty qua state sovereignty continues to inform much research in political geography and IR, it is now also the case that scholars of world politics have embarked on a substantial rethinking of the spatiality of political authority beyond what John Agnew has called the statist 'territorial trap'. The empirical reasons for this are legion, and can only be but hinted at here: the abrogation of Bretton Woods, and the attendant explosion of foreign direct investment and global finance; the growth of transnational production and transnational firms, as well as free trade regimes governing states' control over production and investment; the scaling up of political authority to entities such as the European Union; the emergence of global protest movements; neoliberalism and the gutting of the welfare state in the global North; the debt crisis in the global South; the phenomenon of un authorized global migration; the end of the Cold War and the emergence of networked forms of nonstate geopolitical practice; and among things, what might be referred to in the shorthand as US imperial ambition on the world stage in the new millennium. Indeed, due to these and other events, which lead in common to a fundamental ques tioning of the sovereign state as a final and absolute locus of political power and authority, there has been a signifi cant rethinking of the role of states and state authority in the modern world economy.

This has in turn led to the development of mappings of power and authority exercised through diffuse net works rather than over blocs of segmented space. Indeed, Agnew suggests (**Figure 4**) that contemporary transfor mations in the geography of political authority necessi tate a more careful historical interrogation of political power as tending over time from simple 'ensemble of worlds' and 'field of forces' models (perhaps closest to the sovereignty qua state sovereignty ideal) to more complex 'hierarchical' and 'world network' models in which pol itical power and authority function in ways that funda mentally challenge the Cartesian coordinates of the modern interstate system. Agnew adds that authority and power in the contemporary global political economy are

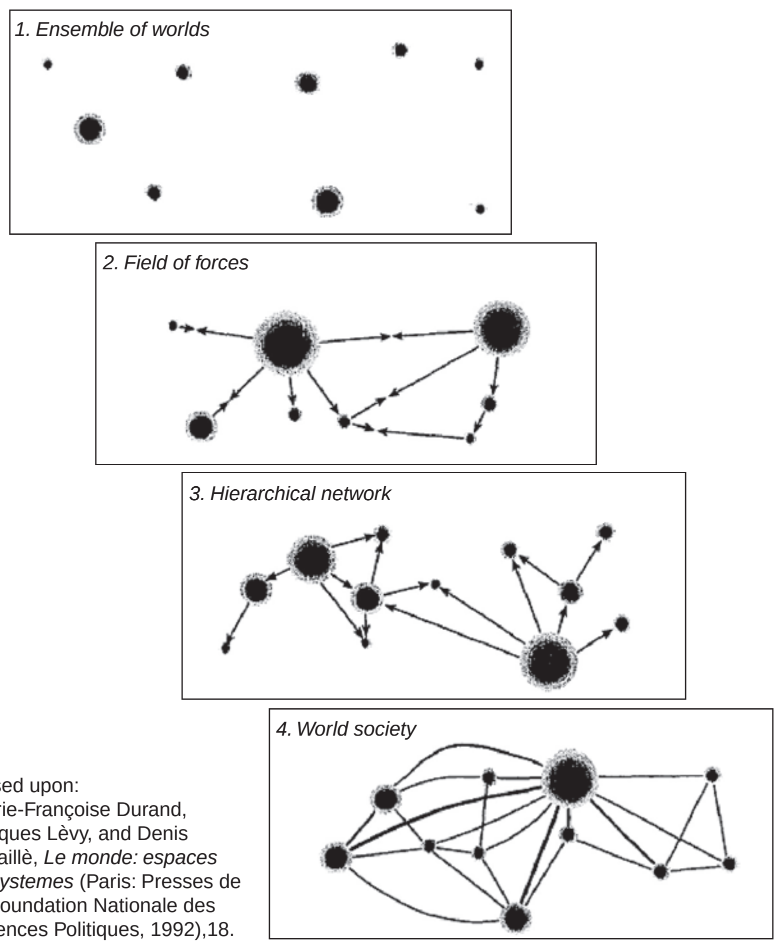

Based upon:
Marie-Françoise Durand, Jacques Lèvy, and Denis Relaillè, *Le monde: espaces et systemes* (Paris: Presses de la Foundation Nationale des Sciences Politiques, 1992),18.

Figure 4 Spatialities of power and political authority. Reproduced from Agnew, J. A. (1999). Mapping political power beyond state boundaries: Territory, identity, and movement in world politics. *Millennium* 28, 499 521.

no longer neatly ‘pooled up’ in territorial states; rather, they are to be found in the ‘effective co presence’ of the four models of power outlined in **Figure 4**.

Global Sovereignty, States of Exception

There has been a recent return to sovereignty, at least nominally, to explain the shape and substance of current world politics. We can think of the two most prominent models as sovereignty as biopolitics and sovereignty as abandonment.

On the one hand, there has been considerable interest in the emergence of a ‘global’ sovereignty. Michael Hardt and Antonio Negri, for example, have written of the emergence of a new form of ‘placeless’ political authority – anathema to state power – predicated on the globalization of the American constitutional experiment of ‘divided powers’. Indeed, in their much read and celebrated *Empire*, Hardt and Negri suggest that political authority in world politics, now best described as disorderly and multiply sited, accords to a loosely structured three tier pyramid. At the height stands the US as a military superpower acting in collaboration with other states through the United Nations. In this top layer are also private credit and debt rescheduling organizations such as the Paris and London Clubs, as well as the G7 states which control key global monetary instruments and regulatory organizations. In the second tier, transnational corporations (TNCs) weave vast production networks throughout the globe, and bargain with a general set of increasingly hollowed out states. The third tier concerns structures of global civil society that percolate upward through states, media conglomerates, religious institutions, and networks of nongovernmental organizations NGOs. All three tiers feed parasitically off the creative and productive energies of the ‘multitude’, a mass of mostly disenfranchised subjects whose labor constitutes the material basis of the global economy (hence sovereignty as biopolitics). Overall, the point is that political authority has shifted from a sovereign state territorial template to a hybrid, flexible, and modulating apparatus of control which Hardt and Negri typify as a more or less ‘smooth space of uncoded and deterritorialized flows’.

On the other hand, due in large part to post 9/11 US geopolitical strategy in the ‘war on terrorism’ – which has brought to light such practices as ‘extraordinary rendition’, as well as a slew of extrajurisdictional prison camps (i.e., Abu Ghraib, Bagram, Guantánamo, etc.) – scholars have become increasingly interested in sovereign power as a ‘state of exception’. Perhaps the most important thinker here is Giorgio Agamben. Unlike much contemporary discussion of authority and power in terms of opaque webs of relations, Agamben zeroes in on the sovereign’s unique ability to abandon political beings to a state of exception. He calls this abandoned form of ‘creaturely life’ *homo sacer*, a form of existence made possible by one’s exclusion from the law, as well as by the physical possibility of death or subjection to extreme violence allowed by this exclusion. In other words, sovereignty as abandonment operates as a suspension of the law and a surrendering of bodies to a zone of indistinction where they can literally be left to die with impunity. Although in some ways this model of sovereignty can be read as a return to older theories of political authority, specifically by virtue of its interest in sovereignty qua state sovereignty, it may also be considered a quite radical refashioning of the canon. For example, among other things, it is the sovereign’s double placement ‘before’ the law and ‘as’ the law – its ability to literally embody and straddle both the ‘state of nature’ and the ‘state of law’ – that gives it what amounts to the ability to forcibly surrender bodies to spaces of neglect and decay.

See also: Empire; Geopolitics; State; Superpower; Territory and Territoriality.

Further Reading

Agamben, G. (1998). *Homo Sacer: Sovereign Power and Bare Life*. Stanford: Stanford University Press.

Agnew, J. A. (1999). Mapping political power beyond state boundaries: Territory, identity, and movement in world politics. *Millennium* 28, 499 521.

Anderson, J. (1996). The shifting stage of politics: New medieval postmodern territorialities? *Environment and Planning D: Society and Space* 14, 133 153.

Bartelson, J. (1995). *A Genealogy of Sovereignty*. Cambridge: University of Cambridge Press.

Hardt, M. and Negri, A. (2000). *Empire*. Cambridge: Harvard University Press.

Held, D. (1995). *Democracy and the Global Order From the Modern State to Cosmopolitan Governance*. Stanford: Stanford University Press.

Hinsley, E. H. (1966). *Sovereignty*. London: C. A. Watts and Co.

Hyndman, J. (2001). Towards a feminist geopolitics. *Canadian Geographer* 45, 210 222.

Krasner, S. (1999). *Sovereignty Organized Hypocrisy*. Princeton: Princeton University Press.

Mann, M. (1986). *The Sources of Social Power* vol. 1. Cambridge: Cambridge University Press.

Murphy, A. (1996). The sovereign state territorial system as political territorial ideal: Historical and contemporary considerations. In Biersteker, T. J. & Weber, C. (eds.) *State Sovereignty as Social Construct*, pp 81 120. Cambridge: University of Cambridge Press.

Sack, R. (1986). *Human Territoriality: Its Theory and History*. Cambridge: Cambridge University Press.

Shapiro, M. J. (1996). *Violent Cartographies*. Minneapolis: University of Minnesota Press.

Walker, R. B. J. (1993). *Inside/Outside: International Relations as Political Theory*. Cambridge: University of Cambridge Press.

Weber, C. (1995). *Simulating Sovereignty Intervention, the State and Symbolic Exchange*. Cambridge: University of Cambridge Press.

Relevant Websites

http://www.lboro.ac.uk/gawc/
Globalization and World Cities website.

Space I

S. Elden, Durham University, Durham, UK

Introduction

Space is, in many ways, the key topic of geography. It is probably the most often cited term that gives geography its identity and unifying focus. While biology studies life, and sociology society, geography lacks this specificity, as the etymology of the term would suppose a focus on the world (*geo*) rather than space. Yet, attempts to redefine geography as spatial science privileged this term, and though criticisms of this position offered alternative ap proaches, they did not fundamentally shift the emphasis from space. Thus both human and physical geographers use the word regularly, but it is frustrating to define and to many appears so diverse as to lack any useful specificity.

One of the reasons that space is so difficult to define is that it has many uses and a complicated history. Indeed, it seems incontestable that space is a historical concept, determined by advances in science and debates in philosophy. Many of the initial disciplinary histories of geography tended to obscure differences in how space was thought in order to lend a degree of coherence to the imagined trajectory of the subject. Where space is defined, or perhaps better, assumed, it tends to be un problematically done so as a container for things to exist in, a place where events happen. Others are content to leave these big questions to one side as they pursue more narrowly defined questions within their own specialism. Yet some awareness of these debates is in valuable to a geographer, however concrete the focus of their work.

Key philosophical questions that geographers have engaged with include:

- Do we experience in space or experience space?
- What is the distinction between space and place?
- What is the relation between space and time?
- Is the world in space or is space in the world?

Space in the Philosophical Tradition

The word ‘space’ is derived from *spatium*, which in clas sical Latin meant a distance or a stretch. It could also mean a fixed distance such as the course in a race – a *spatium* or a *curriculum* was a lap in a chariot race. In the late Middle Ages some writers started to use the term as more akin to a container, as a synonym of a particular sense of *locus*, or place.

For later writers such as Descartes, what defined space, *spatium*, was its extension – or stretch – in three dimen sions: length, breadth, and depth. For Descartes this gave rise to a particular determination of the material world as what he called *res extensa* – an extended thing – and because of this fundamental determination the world becomes accessible to us through the science that best allows us to comprehend it. This, for Descartes, is geometry, which he reconfigures in such a way to enable each point in space to be mathematically determined, and therefore effectively the same as any other. The notion of Cartesian coordinates (x, y, and z) derives from this understanding. Space in these terms becomes more understandable, and therefore controllable, through being understood in such a way.

Philosophers since have written about space in a number of ways, though it is interesting that it is much less the focus in itself but rather as one of a pair of terms. This is either because space is paired with ‘time’, or opposed to ‘place’.

Space and Time

It is tempting to suggest that space is like time in at least one respect. This is St. Augustine’s famous remark in his *Confessions*, Chapter XIV, that if nobody asks us what it is, we know; yet if we are asked, we do not know. In the philosophical tradition, time has generally been accorded a priority over questions of space. Whether this is because of a belief that while all events happen in time some happen solely at the level of thought and are therefore unplaced, or whether the priority is accorded to temporality and thinking spatially is derived from it depends on the par ticular thinkers. Some suggest that space measurements are ultimately reducible to ones of time, and that therefore time is prior, logically, to space. Often time is looked at as progress, change, and development; space as static, fixed, and merely a container. Kant argued that the senses of space and time were effectively hard wired into our brains that we could not perceive except in space and time. Kant however thought that time was prior to space. For some writers, ideas of the three dimensions of space – x, y, and z – are partnered by a similarly calculated dimension of time, t. Temporal change has long been known to change the dynamics of space – Heraclitus famously declared that you cannot step into the same river twice, since other waters are continually flowing in. In Einstein’s work the relation between time and space is not merely coexistent, but co dependent.

Space and Place

Many writers see a distinction between space and place, in that place is the more experienced, lived form of our encounter with our environment, and space is the more mathematical, abstract imposed view of this. Place in these terms is understood in large part through the body, and our perception of what it moves through and interacts within. Yet it is worth remembering that many of our units of measure are derived from our body – foot, inch, fathom, and cubit, for instance, and an acre was the amount of land one man and an ox team could plow in a day. Others see that place is to be understood as a more localized way, whereas space is larger and less personal. On that basis place is 'in here' and space 'out there'. Other approaches see that place is more open and dispersed, whereas space is bounded and exclusive. While two things can be in the same place, this is not the case with space, as things take up space, they occupy it. This view, which was certainly held by Descartes, is founded on the claim that the calculative measurement of space is the same both for the volume of objects and the 'container' they are within. As the remainder of this article notes, many of the ways in which space is understood within contemporary human geography are closer to some of the traditional understandings of place, and some closer to the more mathematical. Other ways of seeing this are attempts to mediate between the two positions.

Spatial Science

The quantitative revolution in geography was often characterized as seeking to reconceive geography as a 'spatial science'. This was part of a wider development within the social sciences, but saw space as a fundamental variable which conditioned both the social context and the actions of individuals within it. Space is thus the focus, and scientific method the means to explore and analyze it. Within this approach to space, issues of distance and arrangement take on a particular priority, alongside direction and connection. The goal of accurate general rules and prediction led to an increase in the objects of study and the development of tools to make this possible.

The pursuit of the spatial order that conditioned spatial relations thus becomes important, which requires more than a straightforward view of space as a static box. Objects and events take on an importance in terms of their spatial proximity and relationships. While this was often unquestioning of what space actually was, the use of mathematical methods and quantification tends to lend support to the idea of space as a calculative, measured construct. In these terms, this is strongly influenced by positivism as a scientific approach. Advances in geography more recently, including geographic information systems, work on a similar understanding of space. Critiques tended to suggest that this either saw space as a separate domain or it had causal properties – separatism or fetishism. In these terms, geographers suggested that other disciplines from within the humanities were more useful in grounding geography rather than the social sciences, or at least the latter with the stress more on the social than the scientific.

Structuralist thought approached space from a related perspective. Many structuralist writers, including the anthropologist Claude Lévi Strauss and the Marxist theoretician Louis Althusser, made great play of spatial language and metaphors in their work. These included terms such as field, terrain, site, and position. Yet they also stressed the question of space, often at the expense of time. Structuralists sometimes claimed that their work was interested in the freezing of time in order to look at things in terms of their relations and structures. This was known as concentrating on the synchronic rather than the diachronic, which examined change through time. The argument was that this was a way of containing temporal change in order to examine the deeper structures of human thought and society. Spatial rather than temporal relations became important in the framing of these structures. It was a way of avoiding the idea that different societies in the world were simply at different stages of development, where they were situated as further along a simple temporal succession. The problem, as many of the structuralists themselves came to recognize, was that this tended to occlude the possibilities of change through time, even within these societies. Later thinkers recognized that though structures or frameworks exist, these are dependent on knowledge, power, and truth, all things that are historically constructed. In this way, they began the process of thinking questions of temporality and spatiality together.

Ways of Thinking Space

Contemporary human geography tends to use the word space in a wide range of ways. While their methodological approaches differ in numerous ways from spatial science, the object of inquiry remains. Indeed, the advent of social/spatial (or sociospatial) theory has tended to emphasize the interrelated nature of social and spatial phenomena, suggesting that to think one without the other is to miss an indispensable part of the picture. This stresses the way in which space functions as a constituent element in social processes – not in a purely determinist way, but rather as continually intertwined, made, remade, and making. This is not the spatial fetishism that some geographers have been accused of in the past, where landscape makes people, and spatial variation is unproblematically assumed. Social relations are not merely spatial relations, but neither is there any way to separate them. These claims suggest that for too long social theory has neglected the spatial aspects of human interaction, while it has generally recognized the historical elements. Many geographers have seized upon Michel Foucault's suggestion that while the obsession of the nineteenth century was history, the present epoch will be the one of space. Yet, as with Foucault's work, this development has not been simply to reintroduce an unproblematic sense of space, but rather to recognize the plurality and fecundity of this term.

This gives rise to the relation between space and property, with an emphasis placed on the difference between privately owned land and public space. Politic ally, much of the debate about space is related to the notion of territory, from that which belongs to individual states to smaller scale territories of groups, gangs, or individuals. In some of this work there is a link between human and nonhuman geographies. Territory in its modern sense of the land controlled by a state is closely related to the development of techniques of measuring land and controlling terrain. In this respect, territory can be understood as a particularly political way of rendering space, and territoriality as a strategy of humans to control and dominate it.

In these terms the question is not so much what space 'is', but how ways in which space have been understood have conditioned particular types of human interaction and understanding. These issues are not confined to geography, of course, since space clearly exceeds that disciplinary focus. Physics, philosophy, and art all have sense of space which is important in understanding the changing senses of space. The space of Renaissance art – of linear perspective and representation – for example, was shattered by the fragmented geometries of Cubism or abstract expressionism. Art in this sense was a challenge to form as well as content. Categorizing these different understandings of space can be done in many ways, yet three ways have proved to be particularly influential. These are absolute versus relative space, and the later introduction of a notion of relational space; Henri Lefebvre's distinction between kinds of space; and the work of Deleuze and Guattari on smooth and striated space.

Absolute, Relative, and Relational Space

When space emerged as an explicit category of Western thought, it tended to be thought in a way that has come to be characterized as 'absolute'. This is space as fixed, as a container, as something in which all things happen. This is the space of classical physics, in the writings of Newton, Galileo, and Descartes. Space can therefore be represented as a two or three dimensional grid, imposed over a landscape in surveying or cartography. Each point is effectively the same in a homogenous expanse, amenable to scientific procedures based on coordinate geometry. Space, as understood in this, is directly understandable through measurement and calculation. This is the view of space that underpins modern science generally, with an emphasis on the arrangement of things or phenomena within space and less on the things or phenomena them selves. Thus, space is built up from points to lines, to planes to volumes.

Space, in this absolute sense, fits with the axioms of Euclid's *Elements*, giving rise to the term 'Euclidean space'. This is somewhat misleading, since Euclid's Greek does not contain a word that is the equivalent of 'space', but the term can be understood to mean a view of space that accords with his axioms and postulates.

In opposition to this view of space, yet sharing many of its essential elements, is 'relative' space. One way to understand this is to recognize that space is not an empty container, but filled with objects and relations. Space is, on this argument, in part dependent on that relation to objects, as it is a positional quality of a world of material things. Relative space can also be based on challenging the fixed geometries of absolute space. This would include the development of multiple geometries that rejects some of the key postulates of Euclid, while re taining a coherency and consistency, and recognizing that the perspective of the observer plays a key role. Spherical geometry, for instance, shows that two parallel lines will meet if laid on a spherical surface, such as the curved space of the globe. Two parallel lines laid north to south do meet at the poles. This has important implications for triangulation as a means of measuring the Earth's surface and helps to explain the reason for the distortion of two dimensional maps, such as the Mercator projection, which sacrifice shape and area for the ability to plot straight line courses. Notions of latitude and longitude depend both on this, and also on a clear sense of how absolute space allows division and abstract models.

When looking at speeds close to the speed of light, Einstein realized that space was related to time, and that space–time is the most appropriate way to see this. Other writers have focused on how space appears dependent on other things. On this view, space and our experience of it is relative to other things – to time, to cost or other economic factors, to social interaction, or the way in which we cognize or perceive it. In these terms, prox imity may not be the same as distance – I am closer, for instance, to the person I am speaking to on the telephone than to the neighbor next door; I have more in common with the other participants on an e mail discussion list, dispersed across the world, than the people in the next town. This can equally be understood as the difference between pattern and process.

Leibniz's dispute with Clarke is a famous instance of a debate between absolute and relative views of space.

Clarke defended the absolute view of space of Newton against criticisms leveled by Leibniz. Leibniz had many points to make, but crucial to them was that space was not an absolute reality, devoid from any substance, or it would be more real than any substance. Kant agreed with this criticism, and Clarke is generally acknowledged to have had little response to this. Leibniz also argued that motion was not something that could be measured only through space itself, but solely understood through relation to other objects. Clarke replied that even objects moving in a void were effected by forces, which could not be related to anything else. Leibniz also suggested that direction in space is relative to the things within it – up and down, left and right, and compass directions are all dependent rather than self evident. Space for Leibniz is thus relative, as it cannot be understood solely in relation to the single point, but only in relation to others around it. This challenges some of the assumptions of positivism or empiricism about how we can measure things. Kant attempted to reconcile these positions by suggesting that space, as part of our internal perceptive apparatus, was real enough to be amenable to scientific examination, yet not an attribute of things aside from through their perception by us.

More recently, some of these ideas have been developed further, but as relational space. In this sense, space is relational because objects exist only as a system of relationships to other objects. Space is thus seen as the product of interrelations, as constituted by them. Space is a multiplicity, heterogeneous rather than homogenous, plural rather than singular. Space is, therefore, always in the process of making, never finished or closed. Doreen Massey, in particular, has suggested that we should understand space in this way. If the differences between this and relative space are perhaps overdrawn, this is a way to understand space that exceeds the mathematical more generally, that is the space of social relations, metaphorical spaces, spaces of poetry, art and emotion, and dream and aspiration. Relative space is, for some contemporary geographers, too close to the understandings of spatial science, and the stress on the relational allows them to escape from more narrowly conceived ideas of the relative. It is through practical analyses that spatial issues are seen to be embedded in social relations. Geographies of emotions and the recent emphasis on affect link to these ideas of space in a less confining way.

Lefebvre's Understandings of Space

Henri Lefebvre was a French Marxist philosopher and sociologist, who wrote on a wide range of topics. Although not a geographer, his work has been widely read by geographers, especially since the 1991 translation of *The Production of Space*. This was a book he came to write only in his mid 70s, after several decades of work on urban and rural sociology, studies of everyday life, and Marxist theory. Lefebvre argued that space was poorly understood and some specificity needed to be introduced. Yet, this was not to propose a fixed, unitary definition. Like other ways of looking at space, and in partial relation to the work of David Harvey, Lefebvre proposed a dialectical way of thinking about space. Harvey and other writers – notably Edward Soja and Mark Gottdiener – had previously introduced Lefebvre's ideas into debates in English language geography. Although Soja's appropriation of Lefebvre was criticized his stress on the essential spatial nature of social life – the argument of the sociospatial dialectic – and the argument that space needed to be reasserted in social theory had a major impact in geographical work and in related fields such as urban studies.

Although Lefebvre took many of his ideas about space from German philosophy, including Ernst Cassirer and Martin Heidegger, he read them through a deep attachment to Marxist politics and the spaces of twentieth century France. Lefebvre argued that there was undoubtedly a way of thinking of space as physical and material – the spaces of architecture, towns, and nature. This was space as we perceived it in the world around us. He called this 'spatial practice', and some have suggested it can be understood as real space. This is the organic, physical space of the world. Lefebvre also argued that there were 'representations of space', space as we conceive of it. This was a mental construct, and would include the spaces of maps, of plans, space as geometry, and space as imagined. This tends to dominate ways in which space is thought. Lefebvre therefore suggests that the traditional ways of thinking of space have tended to be concrete or abstract.

Lefebvre's dialectical point is that these two ways of thinking about space are, in themselves, understandable, but that in our everyday lives we encounter space in a way that, while it has elements of both, exceeds their determinations. These he calls 'spaces of representation' (the English translation renders them as 'representational spaces'). This is the social space of lived interaction, the spaces of symbolism and meaning and, in Edward Soja's term, space as simultaneously 'real and imagined'. Lefebvre therefore challenges both a crudely materialist analysis and a politically detached idealist one. He suggests that this lived space is dominated by the tactics of abstraction, and overlays the physical space of the world, imbuing it with meaning.

As a Marxist, Lefebvre was particularly interested in introducing this analysis into his historical materialism. Some have suggested that this spatial element makes for a historical and geographical materialism, but this perhaps runs the risk of too much emphasis on the additionality, and not enough on the disruptive introduction. To inject this kind of spatial perspective is simultaneously to challenge how we think about time and history, with change through space and time rather than simply historical progress being one of the results. Lefebvre also claimed that time should also be understood in a lived sense rather than a mechanized or measured way.

Lefebvre particularly emphasized the way in which spaces were produced – that is constructed, coded, recoded, and used through a range of practices. For Lefebvre, social space is a social product. These practices were not only confined to material building projects or abstract schemata of town planners, or regional developers but also the everyday interactions of people and groups. His analysis therefore not only owes much to political economy, in that space is an economic and social product, with analysis of rent, exchange and use value, and profit, but also exceeds this to develop a more general politics of space. Space and time thus are not neutral, preexisting frames for human events, or the way in which we experience other things. Space and time are produced and shaped through actions and reactions, experienced in themselves.

Lefebvre's notion of social space can also be understood as lived, embodied space, and so has links to some of the literature on 'place'. This is somewhat complicated by Lefebvre's use of the term *espace*, which, in French, arguably carries a wider range of resonances than the English 'space'. Many of those meanings would not straightforwardly be understood as 'space', and are perhaps closer to the English use of 'place'. Lefebvre rarely used the term space in isolation from some kind of qualifying adjective. Indeed *The Production of Space* is filled with a number of other spaces – concrete space, abstract space, contradictory space, and differential space. This is an important and neglected point, in that Lefebvre proposes the three part schema only to complicate and historicize it in later chapters of *The Production of Space*. For Lefebvre the mode of production is important in terms of determining how space will be seen, and he is particularly interested in how capitalism produces space.

Smooth versus Striated

One final way to think about space is found in the work of the French thinkers Gilles Deleuze and Félix Guattari. Deleuze was a philosopher by training while Guattari was a Lacanian trained psychoanalyst. They wrote extensively individually – especially Deleuze – and a number of books together, of which *A Thousand Plateaus* is probably the most influential in geography.

Deleuze and Guattari suggest a distinction between smooth and striated space. Smooth space is crisscrossed with lines, overlapping and jumbled points in a web of connections. This is space as immense and sprawling, space defined through movement, lines, and trajectories. Striated space is organized, hierarchical and largely static, channeled, and divided into shapes of various sizes. Politically these can be represented as the globalized world of the third millennium, of far flung points connected in a network society, nodes in a system; as opposed to the nation state map of ordered Europe, which progressively eradicated the overlapping sovereignties of medievalism, redistributing land to remove almost all the enclaves and exclaves.

However, Deleuze and Guattari convincingly argue that smooth space only appears to be free of striation, and that while striated and smooth space can be theoretically disassociated, in practice they continually entwined. This has been useful for some geographers in trying to understand the way in which the smooth space of global networks is imposed over the striated spaces of modernity, with nation states still attempting to cling to their sovereignty and territorial integrity in an age of globalizing markets and culture and emergent global modes of governance. To use two of their other terms, this is, for Deleuze and Guattari deterritorialization partnered with reterritorialization, the continual making and remaking of spatial relations.

Conclusion

In his recent work, Harvey has suggested that in many of the ways geographers operate, we think of space as absolute, relative, and relational all at the same time. While perspective is important, this does not mean that we can abandon some of the ways in which absolute understandings of space condition our daily lives. Harvey's point is that contemporary geography may be concentrating too much on the relational and forgetting the concrete and relative; just as the positivism of a previous generation was too obsessed with the absolute and relative. For Harvey, this fits with his views of space as dialectical.

Harvey has also attempted to reconcile the absolute/relative/relational schema with Lefebvre's ideas, suggesting that these can be turned into a matrix of meanings for understanding space. If this perhaps is too neatly dialectical in attempting to reconcile all differences within a confining system, it does point to the ways in which these different ways of understanding space are not mutually exclusive, and it is possible to recognize geographers working with a number of these ideas and frames of understanding today.

Many of these ways of conceiving, perceiving, and living space illustrate the unequal dynamics of human society. Space can be understood economically as a form of property, particularly through the notion of land; as a military strategic issue of terrain; and as the political notion of territory. Certain spatial frames are imposed, made to seem obvious and unproblematic, and yet have profound impacts on people's lives, in a variety of ways. If we think about power as creative, and not just repressive, and therefore about the making and remaking of social relations, it is clear that thinking about power requires a thinking of questions of space. Similarly, thinking about space requires an understanding of questions of power, since it is about interaction, determination, and control. In that sense 'space' really is at the heart of what geog raphers – both within and beyond the discipline of 'geography' – do.

See also: Mapping, Philosophy; Marxism/Marxist Geography II; Military Geographies; Phenomenology/ Phenomenological Geography; Philosophy and Human Geography; Place; Positivism/Positivist Geography; Quantitative Revolution; Scale; Territory and Territoriality.

Further Reading

Benko, G. and Strohmeyer, U. (eds.) (1997). *Space and Social Theory: Interpreting Modernity and Postmodernity*. Oxford: Blackwell.
Bonta, M. and Protevi, J. (2004). *Deleuze and Geophilosophy: A Guide and Glossary*. Edinburgh: Edinburgh University Press.
Casey, E. S. (1997). *The Fate of Place: A Philosophical History*. Berkeley: University of California Press.
Cloke, P. and Johnston, R. (eds.) (2005). *Spaces of Geographical Thought*. London: Sage.
Crang, M. and Thrift, N. (eds.) (2000). *Thinking Space*. London: Routledge.
Deleuze, G. and Guattari, F. (1987). *A Thousand Plateaus: Capitalism and Schizophrenia*, vol. II. Minneapolis, MN: University of Minnesota Press.
Elden, S. (2004). *Understanding Henri Lefebvre: Theory and the Possible*. London: Continuum.
Gottdiener, M. (1985). *The Social Production of Urban Space*. Austin, TX: University of Texas Press.
Gregory, D. and Urry, J. (eds.) (1985). *Social Relations and Spatial Structures*. London: Macmillan.
Harvey, D. (2006). Space as a keyword. In Gregory, D. & Castree, N. (eds.) *David Harvey: A Critical Reader*, pp 270 293. Malden: Blackwell.
Jammer, M. (1958). *Concepts of Space: The History of Theories of Space in Physics*. New York: Dover.
Kern, S. (1983). *The Culture of Time and Space 1880 1918*. Cambridge, MA: Harvard University Press.
Lefebvre, H. (1991). *The Production of Space* (translated by D. Nicolson Smith). Oxford: Blackwell.
Lefebvre, H. (2009). State, space, world. In Brenner, N. & Elden, S. (eds.) *Terror and the State of Territory*. Minneapolis: University of Minnesota Press.
Massey, D. (2005). *For Space*. London: Sage.
Ray, C. (1991). *Time, Space and Philosophy*. London: Routledge.
Sack, R. (1980). *Conceptions of Space in Social Thought*. London: Macmillan.
Soja, E. W. (1989). *Postmodern Geographies*. London: Verso.
Soja, E. W. (1996). *Thirdspace*. Oxford: Blackwell.
Thrift, N. (2003). Space the fundamental stuff of human geography. In Holloway, S. L., Rice, S. P. & Valentine, G. (eds.) *Key Concepts in Geography*, pp 95 107. London: Sage.
Tuan, Y. F. (1977). *Space and Place: The Perspective of Experience*. Minneapolis, MN: University of Minnesota Press.
van Fraassen, B. C. (1970). *An Introduction to the Philosophy of Time and Space*. New York: Random House, 1970.

Space II

R. Kitchin, NUI Maynooth, Maynooth, Republic of Ireland

Glossary

Absolute Space Space is understood as a geometrical system of organization (usually Euclidean geometry with *x*, *y*, and *z* dimensions) within which people and objects are located and move through. Here, space is understood as natural, given, essential, and measurable.

Cognitive Space Space as mentally constructed. While space *per se* is absolute in nature, to be able to operate in the world and undertake complex spatial choices and decisions people rely on their ability to remember and think about spatial relations. We might occupy absolute space, but we live in cognitive space.

Idiographic Science A form of science that focuses on the individual or the unique, and is illustrated in geography by a case-study approach to people and places.

Masculinist Space Feminists argue that both absolute and relative conceptions of space are highly masculinist in character, underpinned by a masculinist rationality and reason that seeks to be autonomous, context free, and objective. Space in these terms is masculinist in conception, something that can be rationally and scientifically understood and mastered.

Metaphorical Space A kind of relational space wherein the production of space is articulated through metaphor.

Nomothetic Science A form of science that involves the search for abstract universal principles, in the case of geography, spatial axioms, and laws about the world.

Ontogenesis A form of ontological thinking that focuses not on what something is, but rather how something becomes. In so doing, it rejects the notion that objects or concepts are ontologically secure – fixable, definable, and knowable – instead arguing that their ontological status is contingent, relational, and unfolding.

Ontology The branch of philosophy that studies the nature and operation of reality or being. It concerns the set of specific assumptions about the nature of existence underlying a theory or system of ideas, beliefs about what exists and can be observed, and therefore known.

Paradoxical Space The notion of paradoxical space recognizes that much about the production of space is unrepresentable and unknowable, given that it is diversely produced by multiple actors and actants who are often working in contradictory ways.

Relational Space In contrast to absolute space, relational understandings of space conceive space to be contingent and active, as something that is produced or constructed by people through social relations and practices. Space is not an absolute geometric container in which social and economic life takes place, rather it is constitutive of such relations.

Spatialization A process whereby attributes with no spatial qualities are given spatial form. For example, attributes of size or number might be transformed into attributes of distance, or proximity or territory, producing a metaphorical space that can be visualized and analyzed spatially.

TimeSpace The dyadic conjoining of time and space that recognizes that they are mutually constituted, and it therefore makes little sense to conceive of them separately. However, time–space extends beyond the idea of four-dimensional space–time (a four dimensional version of absolute space) to recognize a multiplicity of space–times that are relational, contingent, dynamic, and paradoxical.

Virtual Space The space within the realm of technologies, such as the Internet, that are entirely immaterial, consisting entirely of ones and zeros.

Introduction

Space is a key geographical concept. Along with other core concepts such as place, landscape, scale, mobility, nature, and environment, it helps define the discipline as one that is explicitly spatial in its focus and thinking. Unsurprisingly then, geography is often described as a spatial science. As such, while human geographers are interested in social, political, cultural, economic, and environmental issues and undertake historical analyses, they do so cognizant of the role of space in shaping the world around us and using theories and methods that illustrate why space and spatial processes matter. This article details how thinking about space has evolved significantly since the 1950s, focusing in particular on how theorists have conceptualized the ontology of space. To illustrate the differences between the various ways of thinking about space, an example of how cities are understood within different ontological frameworks is used.

Absolute Conceptions of Space

Implicitly Absolute Space

Interestingly, given the centrality of space to geographical work, prior to 1950s, it is fair to say that beyond the works of theoretical physics (such as works of Newton, Liebniz, and Einstein) and branches of philosophy little conceptual work had been undertaken concerning the ontology of space. Geographers, who might have been expected to be most interested in such philosophical thinking, were more concerned about spatial processes across and within space, rather than the nature of space itself. Space was simply understood as a container which things happened; space was implicitly absolute in nature having fixed dimensions across which things could be mapped. As such, while not formally recognized by those working at the time, conceptually space was understood as natural, given, and essential, and spatial processes were teleological and measureable. There was no need to seriously think through the ontology of space, as it was beyond question – it was simply the spatial dimensions in which life took place. As a consequence, geography as a discipline was highly descriptive detailing spatial patterns and processes and the interrelationships between places. Underpinning this kind of geographical research was an exceptionalist belief that geography and its method were unique, being an ideographic science (fact gathering), as forwarded by influential geographers such as Richard Hartshorne, rather than a nomothetic (law producing) science. Epistemologically empiricist, research was then largely analytically naïve, consisting of the accumulation of facts as evidence for generalist theories. From this perspective, cities were understood as unique, but related entities, and analyzed by mapping patterns of different variables, such as commerce, transport, and types of residence, and by charting the functional relationships between cities and regions.

Absolute Space

From the early 1950s, a number of geographers started to challenge the *status quo* in geographical research by arguing that the discipline needed to become more scientific in its thinking and method. For example, Frederick Schaefer, in a paper often cited as the key catalyst for the adoption of scientific method in human geography, argued that "geography has to be conceived as the science concerned with the formulation of the laws governing the spatial distribution of certain features on the surface of the earth" (Schaefer, 1953: 226–249). For Schaefer and others, this meant explicitly recasting the implicit ontology of space using the language of science. Here, space was defined and understood in absolute terms as a geometrical system of organization (usually Euclidean geometry with x, y, and z dimensions) within which people and objects are located and move through. Spatial processes operating within this space could be measured objectively and scientifically, then analyzed using quantitative techniques and spatial statistics. Highly essentialist in formulation, space is effectively reduced to the essence of geometry, its properties natural and given. For converts to this new way of thinking, geography became the science of spatial laws wherein spatial relations could be explained through functional equations and could be modeled and simulated. Although few of these converts referred to the philosophy of positivism in their work, it is clear that many of spatial science's central tenets are drawn loosely from this school of thought. This mode of thinking became very popular during the 1960s with the development of spatial laws and models with respect to many phenomena and remains the cornerstone of much spatial science and geographical information science (GISc) research. From this perspective, cities have an absolute system of geometry that directly shapes spatial processes and behavior and are complex systems of interlocking spatial relationships and laws which can be calculated, modeled, and predicted.

Cognitive Space

During the 1970s, the concept of absolute space was complemented with that of cognitive space. This perspective argued that while the space in which people live is absolute in nature, it is not perceived or cognized as such. Rather, to be able to operate in the world, to undertake complex spatial choices and decisions, people rely on spatial understandings of places, their ability to remember, and think about spatial relations. Consequently, human spatial behavior and, therefore, most spatial processes of note are based on cognitive space – space as mentally constructed. As such, behavioral geographers and environmental psychologists argued that while we might occupy absolute space, we live in cognitive space, and therefore we should study the interrelationship between the two. For them, cognitive space is ontologically abstract, representational, and intangible; it is a product of the mind. Epistemologically, there are two dominant schools – analytic and phenomenological. The first takes a classic psychology approach seeking to scientifically measure and analyze the properties of cognitive space – its components and its geometries – and to model how it approximates to, and is mapped onto, the absolute space of the world. The latter is more concerned with the sense of place and peoples' beliefs, values, understandings, and attachments to particular spaces. For behavioral geographers, cities may well consist of absolute spaces, but spatial behavior and many fundamental spatial processes are founded on cognitive space and how people think spatially about the city.

Relational Conceptions of Space

Relational Space

Also developing from the 1970s onward, as a more explicit counter to the ontology of absolute space, were the more relational ontologies of space. These ontologies differed markedly to the concept of relational space as defined in physics and were first explicitly articulated within human geography by radical geographers (e.g., Marxist and feminist scholars) who were challenging the ideas and ideology underpinning spatial science. These theorists argued that spatial science was highly reductionist and that absolute notions of space emptied space of its meaning and purpose and failed to recognize the diverse ways in which space is produced. Space, it was argued, was not a given, neutral, and passive geometry, and essentialist and teleological in nature. Instead, space was conceived as relational, contingent, and active, as something that is produced or constructed by people through social re lations and practices. Space is not an absolute geometric container in which social and economic life takes place, rather it is constitutive of such relations.

In such thinking, it is recognized that the spaces we inhabit – the built environment, transport systems, and the countryside – do not simply exist, preformed and awaiting meaning. Rather they, and the spatial relations they engender, are produced – made, shaped, managed, and given meaning by people; they are the products of diverse material and discursive practices that in turn actively shape social relations. Conceived of in these terms, an everyday space like a football stadium can be seen to be both a physical form constructed by certain agents and institutions for particular ends as well as a space given meaning through myth, language, and ritual: its use and occupation is shaped both by its material form and the immaterial meanings that coalesce around it. Cities are thus composed of relative spaces, produced in contingent and relational ways by people. Epistemo logically, what this relational conception of space de manded was a shift from seeking spatial laws to focus on how space is produced and managed to create certain sociospatial relations.

This relational understanding of space is perhaps most fully developed by Henri Lefebvre in his book *The Pro duction of Space*. As Merrifield notes, Lefebvre sought to develop a 'unitary theory of space' that would provide a "rapprochement between 'physical' space (nature), 'mental' space (formal abstractions about space), and 'social' space (the space of human action and conflict)" (Merrifield, 2000: 167–182). He suggested that these seemingly different types of space are actually of the same substance and force, each produced through the entwining of three elements, which he determined were key in the making space: spatial practices, representations of space, and spaces of representation (or represen tational space). Spatial practices refer to the processes, flows, movements, and behaviors of people and things that can be perceived in the world. Representations of space refer to the discursive media (e.g., images, books, films, maps, plans, and so on) which serve to represent the world spatially in order to make sense of it and to think through what is and might be possible. These represen tations work ideologically to legitimate or contest par ticular spatial practices and as such do not simply represent space but rather produce space. Spaces of representation are the spaces that are produced by the body in everyday practice; the spaces lived and felt by people as they weave their way through everyday life. The relations between these three elements are complex, and analytical priority cannot automatically be given to one element over the others, but in combination they bind together Lefebrve's three forms of space (physical, mental, and social) to produce space. Crucially, the re lationship between the three elements varies over time and with context, as Lefebrve demonstrated by trans forming Marx's periodization of capitalism into a history of spatial production, thus showing how different con figurations produced different spatial relations across time and place. In so doing, Lefebvre's work implies that the main struggle in society is one of spatial conflict, of contesting the production of space.

Masculinist and Paradoxical Space

From the early 1980s, feminist geographers have been making important interventions into both disciplinary practices and the foci, theorization, and praxis of human geography. Geographers such as Doreen Massey, Susan Hanson, Jan Monk, Linda McDowell, and Gillian Rose and others have highlighted that what counts as geo graphical knowledge and who produces such knowledge is skewed by the fact that the discipline has been highly dominated by men. As such, they argue, geographical theory and praxis is highly masculinist in character, largely underpinned by a masculinist rationality and reason that seeks to be autonomous, context free, and objective. As a consequence, they suggest that our understanding of space is similarly masculinist – space is something that can be rationally and scientifically understood and mastered.

In contrast, Gillian Rose in *Feminism and Geography* argues that space is never fully knowable as we can never achieve an all encompassing, exhaustive God's eye view of the world, only views from particular positions that are differentially shaped. She forwards the notion of para doxical space – of an understanding of space that reflects the researchers' viewpoint, while at the same time is sensitive to the myriad ways in which space is experi enced and produced by others, and acknowledges that there is much about the production of space that is unrepresentable. For her, this involves producing relational geographies that recognize the complex, entangled, and often contradictory power geometries that produce space, rather than constructing grand, all encompassing theories that seek to 'tell it like it is' and, in so doing, silence different viewpoints (as with, for example, much Marxist analysis). As Rose notes: "space is multidimensional, shifting and contingent. It is also paradoxical, by which I mean that spaces that would be mutually exclusive if charted on a two dimensional map – centre and margin, inside and outside – are occupied simultaneously" (Rose, 1993: 140). Space from this per spective is simultaneously knowable and unknowable, representable and unrepresentable, produced by complex and often contradictory forces to produce a "precarious conceptual geometry of the non Euclidean type" (Rose, 1993: 141). In such a view, city spaces are diversely produced and understood and their analysis requires the careful uncovering of its paradoxical nature.

Metaphorical Space

Metaphorical space is a particular kind of relational space, most often articulated in disciplines beyond geography, especially literary and cultural studies. It became a popular way to conceive of space from the 1990s onward. Here, space and its production are seen to take metaphorical qualities that are far removed from absolute conceptions of space. So, for example, the pro duction of city space is seen akin to a text that is written and read (city as a text), as organic and living entity (city as a body), as a massively complex assemblage of nuts and bolts (city as a machine), as a network of flows and fluxes (city as a network), and so on. Such metaphorical understandings of space seek to detail the diverse nature of space with respect to its constitution, meaning, func tion, and complexity that is not always easily captured or expressed in other articulations. In other words, meta phor is used to try and articulate the paradoxical nature of space – to provide a shape to explain complex geographies.

Virtual Space

More recently, with the development of Internet tech nologies and the advent of cyberspace, spatial theorists have sought to think through the nature of virtual space. Often considered as aspatial (spaceless and placeless) by many commentators, as many geographers have demon strated, cyberspace is inherently spatial and possesses diverse spatialities. Cyberspace consists of a diverse collection of interlinked domains. Some of these domains are explicitly spatial in nature with direct geographic referents (e.g., virtual models of a geographic location), some are explicitly spatial in nature without a geographic referent (e.g., virtual worlds and multiuser dimensions (MUDs)), some have real world referents but no explicit spatial form/attributes (e.g., a list of names, a Web page), and some have no or little geographic referents or spatial form/attributes (e.g., computer file allocation tables). While the latter two lack formal spatial qualities, they are often given spatial form through a process of spatialization, that is, a spatial structure is created through the conversion of defined attributes into spatial attributes (e.g., size to distance/proximity or territory) producing a metaphorical space that can be visualized. All forms of virtual space, however, are very different in nature to geographic space being entirely immaterial, consisting entirely of ones and zeros, produced through code. Such spaces are free to possess geometries and forms impossible to recreate in geographic space. What this means is that virtual spaces can possess very chaotic geometries that lack Cartesian logic being multidimensional and noncontinuous, where 'travel' between domains is nonlinear and rhizomic; every location being each others' next door neighbor. Spaces can be both territory and map, with space itself also the means of navigation such as with hypertext in web pages. From this perspective, cities are complemented with virtual spaces of information, interactions, and transactions that can be mapped onto geographical space in diverse ways.

TimeSpace

So far, the discussion has focused exclusively on space. Yet, everyday life occurs in both time and space. For some geographers, it therefore makes little sense to think of time and space as two separate categories as the two are inherently mutually constitutive. They sug gest we should think not about time or space but of TimeSpace. The interlinking of time and space in con temporary geography has its roots in the work of Torsten Hägerstrand and his notion of time geography, developed initially as part of his doctoral research, completed in the 1950s, and continued throughout his lifetime. Hägerstrand was interested in patterns of settlement, migration, and innovation diffusion over time, often at an individual level, and how they could be understood, modeled, and simulated. For him, it was important to place spatial changes into a temporal context in order to map how patterns had evolved. This was not an act of simply producing historical geographies, but of pro ducing explicit time space paths of movement of people, objects, and ideas – literally mapping in four dimensions (x, y, z, and time). This contrasted with much other quantitative research at the time which tended to focus purely on analyzing spatial processes and patterns without placing them fully in a temporal context.

Just as thinking about space has evolved, so has thought concerning the interlinking of space and time. For Hägerstrand, these were two interlinked but separate variables (hence the hyphen often used to join them – time space or space time). For some contemporary thinkers, they are two sides of the same coin (hence the lack of a hyphen to denote they are dyadic – mutually constituted). In their book *TimeSpace*, Jon May and Nigel Thrift argue that thinking about time and space as separate categories tends to lead to one being prioritized over the other. This has negative consequences because a prioritization of time produces a 'debilitating historicism' that reduces space to a neutral backdrop and a prioritization of space leads to debilitating 'spatial imperialism' that overemphasizes space at the expense of time. Yet, time and space work in conjunction with each other – everything happens at some time in some place – and so time and space need to be conceptualized as conjoined: TimeSpace.

For them, this is not simply thinking of things as four dimensional space time (a kind of four dimensional absolute space), but of recognizing a multiplicity of space times that are relational, contingent, dynamic, and paradoxical; that our experiences and understandings of TimeSpace are shaped by timetables and rhythms, disciplining regimes, instruments and devices, and various discourses. What these produce are a heterogeneous set of TimeSpaces that are complex, sometimes contradictory, and which need to be carefully mapped, as they demonstrate through a discussion of time space compression. Most analyses of time space compression argue for a significant speeding up and shrinking of the world post Industrial Revolution due to advances in transport and communications. In contrast, they highlight how time space compression was uneven (across the globe, class, race, etc), contradictory (just as some things speeded up, others slowed down), worked at different rhythms, and how other technologies had significant impacts on TimeSpace, such as power, light, cinema, and so on, to restructure in multiple ways, rather than accelerate TimeSpace. From this perspective, geographers need to examine the timespaces of cities, tracing out the various heterogeneous rhythms of daily life.

Ontogenetic Conceptions of Space

All of the conceptions of space outlined above are unified through the examination of the ontology of space, wherein they explore 'what space is'. More recently, a group of scholars have begun to challenge such an ontological position and to develop and forward ontogenetic conceptions of space. In so doing, they change the central question of enquiry from 'what space is' to 'how space becomes'. Space (and everything else in the world), they argue, is not ontologically secure – a fixable, definable, knowable, predetermined entity. Rather, space is always in the process of becoming; it is always in the process of taking place. Space, in these terms, is a practice, a doing, an event, a becoming – a material and social reality forever (re)created in the moment. Here, space gains its form, function, and meaning through 'practice'. Space 'emerges' as a process of ontogenesis. As Marcus Doel has pointed out, from this perspective, space can be seen as a verb rather than a noun, with him suggesting that the term 'space' might better be replaced by 'spacing' to better capture its ceaseless production. These ideas have been extended to other core concepts underpinning spatial thought, such as scale, place, nature, and landscape, recasting each within ontogenetic terms (challenging the ontological security of the concept itself, and rethinking each as emergent in nature).

Michel de Certeau in *The Practice of Everyday Life* sought to move beyond theories centered on representation and behavior to consider 'practices' that are constitutive of both. In particular, de Certeau (drawing on Foucault) was interested in how people live within, negotiate, and challenge circuits of power and the 'proper' order of space as reproduced by dominant elites such as the state and corporations. Here, space is an outcome of the complex interplay between discursive and material strategies that seek to reproduce "places in conformity with abstract models" (de Certeau, 1981: 29) of scientific rationality and political economy, through persuasion, seduction, coercion, domination, intimidation, violence, and so on, and resistive tactics that seek to undermine such citational practices by "manipulat[ing] events in order to turn them into opportunities" (de Certeau, 1981: xix); for example, avoidance, organizing protest, transgressing social norms, and so on. de Certeau understood tactics as performative, as emerging unconsciously within a context, so that as individuals "move about, their trajectories form unforeseeable sentences, partly unreadable paths across a space" (de Certeau, 1981: xviii), where a trajectory "*comes into being*, the product of a process of deviation from rule governed … practices" (de Certeau, 1981: 5; original emphasis). Individuals actualize spatial possibilities, making space exist as well as emerge; they invent and transform space; they create a "mobile organicity, a sequence of phatic *topoi*" (de Certeau, 1981: 99; original emphasis). This actualization is citational, "making credible the simulacra produced in a particular place" (de Certeau, 1981: 189). de Certeau explains:

> Space occurs as the effect produced by the operations that orient it, situate it, temporalize it, and make it function in a polyvalent unity of conflictual programs or contractual proximities. On this view, in relation to place, space is like

> the word when it is spoken, that is when it is caught in the ambiguity of an actualisation, transformed into a term dependent upon many conventions, situated as the act of a present (or of a time), and modified by the transformations caused by successive contexts. … In short, *space is a practiced place*. Thus the street geometrically defined by urban planning is transformed into space by walkers. (de Certeau, 1981: 117; original emphasis)

In other words, for de Certeau space emerged in context through practices.

From a related perspective, Gillian Rose draws on Judith Butler's theory of performativity to argue that: "space is a doing, that does not pre exist its doing, and that its doing is the articulation of relational perform ances … space then is not an anterior actant to be filled or spanned or constructed … [i]nstead, space is prac tised, a matrix of play, dynamic and iterative, its forms and shapes produced through the citational performance of self other relations" (Rose, 1999: 248). To Rose, space itself, and thus its production, is brought into being through performativity – through the unfolding actions of people. She thus argues that this produces a 'radically unstable notion of spatiality' that allows for a critical analysis of space as "extraordinarily convoluted, multiply overlaid, paradoxical, pleated, folded, broken and, per haps, sometimes absent" (Rose, 1999: 247). In other words, she suggests that a performative understanding of space allows for a nuanced analysis that appreciates in dividual differences across place, time, and context, and the paradoxical, contradictory, and complex nature of sociospatial relations as lived and expressed by people.

Drawing on the ideas of Butler, Latour, and Deleuze, among others, Nigel Thrift has developed the notion of nonrepresentational theory. Thrift suggests the world emerges through spatial practices that are often unre flexive and habitual, and that are not easily represented and captured because they are unconscious and in stinctive; they are performed without cognitive and rational thought. These human practices are com plemented by other actants – animals, objects, machines, circuits, networks – that do diverse work in the world. In particular, Thrift is interested in how new sentient technologies automatically produce space, that is, bring space into being without human interference.

Dodge and Kitchin have extended this work by con sidering how software does work in the world – pro cessing information, making decisions, controlling technologies that shape peoples lives in a myriad of ways (how software augments, mediates, controls domestic appliances, transport systems, communications technol ogies, healthcare, work environments, utilities, financial networks, and so on). Software, they suggest does not simply help produce space, it transduces it – transforms it from one state to another. They argue that space is constantly brought into being as an incomplete solution to ongoing relational problems – enabling people to meet or work or play, getting from A to B, etc. For them, this means that software is helping to transduce different formations of space into being. Two such spaces are what they term code/space and coded space. Code/spaces are spaces dependent on code to transduce them into being, wherein the relationship between code and space is 'dyadic' – that is, without code the space would not come into being as intended. For example, checking in areas at airports are now dependent on software systems to function – the old manual way of checking in has been discontinued for security reasons. If the software crashes, there is no other way of checking passengers in and the space is produced not as a checking in area but as a waiting room. Coded space, on the other hand, is where a transduction is mediated by code, but whose relationship is not dyadic – software mediates the solution to a problem, but it is not the only solution. For example, a PowerPoint presentation transduces the space of a lecture theater, but if the computer crashes the space continues to be produced as a lecture theater, although the talk might not be as effective as it might have otherwise been.

The ceaseless production of space articulated by these theorists can be illustrated in many ways. With respect to spatial form, it is clear that the world around us is not static and fixed. Instead, spatial forms are constantly being altered, updated, and constructed through the interplay of complex sociospatial relations in ways that alter, in often subtle and banal ways, the spaces we live in. At a macro scale, there are new local, regional, and national development schemes that are enacted daily to transform and regenerate built environments, transport infrastructures, and 'natural' landscapes. For example, modifications in road layout, new buildings and infra structure, additional and reorganized public transport provision, new zones of business and housing, land management schemes such as drainage or irrigation, and so on, that alter the physical landscape and time–space relations of places. At a more micro scale, infrastructure is modified, repaired, redesigned, and so on so that streets and rooms are always in a process of being refashioned and remodeled and spatial layouts rejigged. For example, streets are dug for cabling, shop fronts updated, shop interiors redesigned and maintained, trees are planted, buildings painted, grass mowed, litter dropped and cleaned up, and so on. In other words, the material fabric of space is constantly (re) created through spatial prac tices that vary in their pacing, so some changes are more immediately noticeable than others. As processes of erosion and entropy at abandoned buildings demonstrate, however, all places are in the course of change, slowly mutating to another state.

Similarly, the function of spaces are not static but alter with time (e.g., seasonally – tourist destinations;

daily – day and night time economy) and the use of space is negotiated and contested between individuals and groups. Spaces thus have multiple functions, and through the daily flux of interactions, transactions and mobilities are always in the process of being made differently. For example, Trafalgar Square in London functions as somewhere to meet, to have lunch, to chat, to visit museums, to gather for protests, to party, to take tourist photos, to travel across, to feed pigeons, to catch a bus or tube, to sunbathe, to people watch, and so on. It is a space in flux, constantly being created in the moment as a collective manufacture composed of hundreds of recursive, interconnected relationships between people and place. Trafalgar Square does not simply exist, fully formed – a still landscape. It is endlessly remade, ceaselessly reterritorialized.

Likewise, the meanings associated with spaces shift, ever changing with mood, action, memory, events, and so on. Again in relation to Trafalgar Square, the meanings inscribed on that location vary as a function of how the space is used (as tourist, or Londoner), how the viewer interprets Nelson's Column and the surrounding buildings (as visually stimulating scenery or imperialist celebration), the social background and attitudes of a person, that person's memories and understandings of the square, and so on. Similarly, meanings attached to home, work, buildings, and routes mutate over time. How space is related to, and the spatiality that engenders, is therefore never static, but emerges, varying over time and across people and context.

Drawing the examples above together, it should be clear that Trafalgar Square (and indeed the notion of what Trafalgar Square is) is always in the process of taking place – its form, function, and meaning is ever shifting across people, time, and context. Its reproduction as Trafalgar Square appears to be relatively stable, because it is maintained as such through a diverse set of discursive and material practices, including street cleaning, stone repair, policing, traffic management, history lessons, reading guide books, viewing postcards, sitting on steps, splashing in fountains, and so on. In other words, Trafalgar Square is constantly remade through repeated, iterative practices enacted by people and things. These practices are citational, in that they endlessly, but imperfectly, cite the previous moment and thus give the appearance of coherence and continuity. Trafalgar Square then is something that happens rather than something that is. Cities emerge ceaselessly through multiple, overlapping spatial practices.

Conclusion

Space, far from being simply the unquestionable backdrop to everyday life, is open to various different ontological conceptualizations. Since the 1950s, geographers along with others have developed a number of different ways to think about space with relative and ontogenetic understandings seeking to replace absolute conceptions. This article has outlined a number of these ways, but it should be noted that beyond geography, theorists of cosmology and physics, and of theology and philosophy, have also continued to produce experimental, observational, and theoretical reflections on space and time.

It is fair to say that the conceptions of space outlined, from implicit to ontogenetic, are today all in use by geographers around the world. For example, absolute conceptions of space still predominate in spatial science and GISc and relative conceptions of space are popular with radical and feminist geographers. While new conceptual thinking is evolving all the time, as some geographers seek ever more sophisticated ways to think about and analyze the world, older ideas persist rather than simply being replaced. What this means is that space remains a fertile concept for contemporary philosophical thinking and debate.

See also: Human Geography; Place; Scale.

Further Reading

Amin, A. and Thrift, N. (2002). *Cities*. Boston, MA: Polity Press.
Bunge, W. (1962). *Theoretical Geography*. Sweden: Lund Studies in Geography, Royal University of Lund.
Curry, M. (1995). On space and spatial practice in contemporary geography. In Earle, C., Mathewson, K. & Kenzer, M. (eds.) *Concepts in Human Geography*, pp 3 32. Lanham, MD: Rowman and Littlefield Publishers.
de Certeau, M. (1981). *The Practice of Everyday Life*. Berkeley, CA: University of California Press.
Dodge, M. and Kitchin, R. (2000). *Mapping Cyberspace*. London: Routledge.
Dodge, M. and Kitchin, R. (2005). Code and the transduction of space. *Annals of the Association of American Geographers* 95(1), 162 180.
Doel, M. (1999). *Postructuralist Geographies: The Diabolical Art of Spatial Science*. Edinburgh: Edinburgh University Press.
Gatrell, A. (1983). *Distance and Space: A Geographic Perspective*. Oxford: Oxford University Press.
Golledge, R. G. and Stimson, R. J. (1997). *Spatial Behavior: A Geographic Perspective*. New York: Guilford Press.
Hagerstrand, T. (1967). *Innovation Diffusion as a Spatial Process*. Chicago, IL: University of Chicago Press.
Haggett, P. (1965). *Locational Analysis in Human Geography*. London: Arnold.
Hartshorne, R. (1939). *The Nature of Geography*. Lancaster: Association of American Geographers.
Kitchin, R. and Blades, M. (2001). *The Cognition of Geographic Space*. London: IB Taurus.
Lefebvre, H. (1972/1991). *The Production of Space*. Oxford: Blackwell.
Massey, D. (2005). *For Space*. London: Sage.
May, J. and Thrift, N. (2001). *Timespace: Geographies of Temporality*. London: Routledge.
Merrifield, A. (2000). Henri Lefebvre: A socialist in space. In Crang, M. & Thrift, N. (eds.) *Thinking Space*, pp 167 182. London: Routledge.

Rose, G. (1993). *Feminism and Geography*. Cambridge: Cambridge University Press.
Rose, G. (1999). Performing space. In Massey, D., Allen, J. & Sarre, P. (eds.) *Human Geography Today*, pp 247 259. Cambridge: Polity.
Schaefer, F. K. (1953). Exceptionalism in geography: A methodological examination. *Annals of the Association of American Geographers* 43, 226 249.
Thrift, N. and French, S. (2002). The automatic production of space. *Transactions of Institute of British Geographers* 27, 309 335.

Space-Time

J. D. Gatrell, Indiana State University, Terre Haute, IN, USA
E. W. LaFary, The University of Auckland, Auckland, NZ

Glossary

Biography Formation The concept of biography refers to the interactions between structure and human agency and how those interactions shape the lives of individuals and determine the collective geography of places.

Chorology The study of the spatial distribution of people, phenomena, or other characteristics and how the distribution can be used to create a 'synthesis' of a place or region.

Space–Time Compression The phenomena of diminishing the relative distance between people, places, or infrastructures that is the result of changing technologies and/or social change.

Structuration A theory of social organization that recognizes agents (or people) interacting across space and over time with other agents within a given collection of existing social structures. In turn, agents simultaneously modify or influence social structures within a given space–time configuration.

Structure–Agency Debate During the 1980s, a great deal of intellectual energy was focused on the politics of micro- and macro-level explanations – in essence geographers were attempting to situate social theory at either the local (agency) or more global (structure) scale. The structure–agency debate contributed to the development of a new middle ground known as structuration. In the process, geographers began to address not only issues of scale – but (re)define the relationship between time and space.

Introduction

Time complicates geography in a fashion similar to the way in which geography, or more accurately space, complicates the natural and social sciences. Ironically, time is the primary social construct that unites the sometimes disparate worlds of physical and human geography insofar as geographers investigate the changing landscape and the ways in which humans interact with or alter the natural environment and/or order the political, economic, and social worlds. Consequently, the challenge for geographers is to (re)conceptualize the inherent relationship between chorology and chronology. By leveraging the interaction between time and space, geographers create synoptic accounts of the world, as well as the interdependencies of its people, regions, places, and landscapes, in both real and imagined spaces. In this article, the historical development of key concepts associated with space–time will be examined including diffusion, time geography, place construction, and technology. Additionally, the article will examine the potential for rearticulating the space–time relationship as a unifying theme within the discipline – particularly as space–time issues impact geographic information science (GIScience). Finally, the article will examine contemporary debates surrounding space–time that increasingly challenge the discipline's definition of space as a physical attribute and the resulting space–time geography.

Diffusion

Geographers interested in the cultural landscape and the diffusion of new ideas, species, and other physical/social phenomena recognized the relationship between time and movement across space. This work was pivotal to the development of a 'time' informed discipline as it bound early regional geographies with similarly scaled histories to construct new syntheses of the natural and physical world. As the following paragraphs illustrate, the manner in which the landscape, regional and quantitative traditions incorporate time and space–time varies and continues to inform contemporary conceptualizations of space–time within the discipline.

Cultural Landscape: A Human–Environment Interface

Geography is primarily concerned with the society–environment interface. While the definition of the interface and the myriad environments (social, physical, and even virtual) of geography investigations are increasingly complex and at times disjointed, the relationship between human beings and their natural environment has been a primary concern of human geography. Indeed, much of contemporary human geography finds its genesis in investigating this essential dynamic and can be traced to the early works of the Berkeley School. The Berkeley School – which was dominated by the influential work of Carl O. Sauer, as well as his students – emphasized the impact of human agency upon the natural environment. Specifically, this collection of researchers explored the changing dynamics of the landscape over time and the role of humans as agents of change. In the process, geographers began to sort out geographies of diffusion and established a new culture of geography that was inherently bound in both space and time known as the cultural landscape. By investigating and dating artifacts in the cultural landscape, these researchers began to tease out the relationship between space and time within the context of the spatial diffusion of innovation insofar as innovation was understood to be cultural practice.

Regional Geography

Within geography, areal differentiation – more commonly referred to as the 'regional concept' as it came to be embodied in the work of regional geographers – served as the empirical foundation for much of the discipline's academic legitimacy throughout the twentieth century. Fundamentally, regional geography treated space as a timeless commodity that could be observed, cataloged, and described. In the process of cataloging space and drawing regional boundaries, the simple chorology of early regionalists acknowledged time but did so in a fashion that was decidedly historical. While the regional paradigm recognized certain people migrated to a region over time or an industry emerged, the primary concern was (and is for those few individuals still performing research in this paradigm) who or what was there at the time of observation. By that, we mean to suggest time – as a variable – was not conceptualized. In fact, Richard Hartshorne – recognized by most geographers as the most prominent regionalist in the United States – was mindful (even insistent) that regional geography was not history. For this reason, the region was a static feature with items, people, and phenomena to be observed. Put succinctly, regional geographies offered well defined 'slices' of space–time – and thus failed to interrogate change as a 'process' situated in and contextualized by both space and time. By neglecting time, the relationship between society and physical systems, regional geographies obscured the human–environment interface and observed conditions were frozen in place and time. As a result, contemporary readers of regional geographies published in the 1940s or the 1970s would be left with the misguided impression that regions exist today as they did then. Conceptually, change over time and across space were in some respects the enemy of the regional concept as they underscored the primary weakness of traditional regional geography in that regions might become more similar or more different in the future. As a result, time challenged the wisdom of simple empiricism of the region – and insofar as the cultural landscape approach began to ask questions that sought explanations – and would eventually give way to a new explanatory approach that would come to dominate the discipline for nearly two decades.

Modeling Diffusion: A Quantitative Approach

While the notion of the cultural landscape would emerge as the dominant theme of the Berkeley School, the dynamic relationship between social and physical systems and how these relationships change over time and across space would serve as the foundation for a great variety of geographic research in future years. More importantly, within the context of this article, the work of these early human geographers coupled human and physical geographies in space and time under the rubric of innovation and in the process established a new human geography of innovation.

By focusing on innovation and how ideas, concepts, and practices move across and change the landscape over time, the earlier geography of the 1920s would emerge as an explanatory science in the 1950s. Like all sciences, the premium on explanation led to the development of sophisticated (to be read complex) and elegant (to be read simple) statistical models in the 1950s and 1960s. As in earlier years, diffusion of ideas, peoples, and phenomena were dominant themes of early quantitative geography. Early on geographers recognized the simple explanatory power of distance and space relative to the diffusion of phenomena from an initial point of origin. Building on the works of other social scientists, the primordial research focused on the diffusion of ideas from urban centers to the hinterland and matured to include examinations within and between urban hierarchies. Likewise, geographers were using diffusion models to understand migration patterns of people, language, and characteristics. Yet, the emphasis on modeling, methods, and abstractions of space as embodied in the simple geometries of distance tended to alienate the temporal dynamics of sociospatial processes. That is, time while certainly an interesting and informative facet of innovation was little more than the 'x axis' extending to infinity upon which spatial diffusion occurred and hence an external factor that could not be fully engaged within the context of a theoretical geography and the then emerging field of regional science. By the late 1960s though, quantitative geography's overemphasis on simple metrics began to obscure the very interactions and dynamics at the core of geography's disciplinary identity – the human–environment interaction. Indeed, a small collection of geographers began to openly question and engage the interaction between time and space and began to assert that 'time geographies' should be investigated.

Time Geography

The reunification of space and time within the discipline, referred to as 'time geography', began with the collective work of Torsten Haagerstand. Influenced by the field of human ecology, time geography evolved in the 1960s from earlier quantitative efforts at modeling diffusion and the recognition that both innovation and diffusion are contextualized processes that are simultaneously bound in space and time resulting in spatially and historically distinct diffusion patterns. In many respects, time geography was an initial step toward repeopling quantitative geography. It was this conceptual reunification of people as agents within human geography that forms the bedrock of time geography. By repopulating human geography, time geographers provided the impetus for a more detailed construction of time and space within and across the social sciences – and these principles serve as the foundation for the emergence of new social theories and conceptualizations of historical geography.

Hägerstrand's Ideas

Hägerstrand, a proponent of geography's quantitative revolution, began to recognize the inherent complexity of spatial modeling and proposed the existence of the so called 'space–time prism' through and within which all actors and structures exist. The space–time prism facilitated the visualization of what Hägerstrand termed space–time paths. Implicit to the model are restrictions placed upon individuals as they interact with and (re) shape the places/spaces/persons within their domain.

According to Hägerstrand, 'prisms' account for and define the activity space of individuals. Within prisms, the activities of individuals are limited by three factors: authority, capability, and coupling. Authority refers to social constraints that limit the movement of individuals or groups. For example, the social constructs of citizenship or class may serve to limit movement across national borders or urban spaces, respectively. Yet, time geography recognized that interactions would occur beyond the structured boundaries, if permissible. In contrast, capability refers to the physical limitations of everyday life. Indeed, the spatiality of everyday life may be limited by commute times, household responsibilities, or other factors that limit the spatial extent of activities. Finally, coupling is the temporality of grouped interaction (e.g., the number of hours one works or the time spent in queue for a coffee) within and across individuals or groups.

Based on these three limiting factors, the model charts the prisms, paths, and stations of everyday life. In concert, prisms, paths, and stations form an interconnected assemblage of paths at stations within the prism which form 'webs' of social interaction producing both real and imagined landscapes. Ultimately the model suggests that time and space are limited; a postulation that is at the heart of contemporary debates by geographers, as well as others on the nature of time and space or space–time within the social sciences (see **Figure 1**). Nonetheless, the works of Hägerstrand and others advanced theoretical conceptualizations of the daily interactions of people, places, and societal structures.

Biographies and Geographies: Structuration

The explosion of research across the social sciences into the role of human agency in shaping social systems gave rise to a vast literature on 'structuration'. The structure–agency debate as it came to be called resulted in a new accounting of how space and time interact to produce unique social structures and ultimately new geographies. This section will also examine space–time as it relates to place construction.

Within geography, time geography served as the initial salvo in an aggressive attempt at repopulating human geography with people as agents of social change within a quasi quantitative framework. One mechanism for understanding human agency in a systematic way was to creatively borrow from the critiques of logical positivism, particularly structuralism and humanism, to understand how people acting within localized (or even global) structures create 'places'. For the humanists, place is the product of human agency and for the structuralists place represents the concretization of social structures at the local scale. Through time the two divergent understandings of the lived experience in place would be reconceptualized under the new theory of structuration.

Structuration – as defined by sociologist Anthony Giddens and later translated by geographers – recognized that 'place' is a dynamic scene which cannot be adequately frozen in time and is constantly (re)created by human agency and whereby human agency is constantly (re)created by social structure. In this sense, Giddens built on the earlier work of Hägerstrand to construct a theory of structuration that asserts that people influence the cultural, social, and physical landscape and that these landscapes in concert shape the geographies of individuals. In other words, structuration was the confluence of actions and processes that produce and reproduce individual belief systems, institutions, and structures and in the process construct an agency–structure dialectic from which everyday geographies are derived or more accurately unfolded. Furthermore, structuration includes the idea that rules and norms of society are both creative and inhibitive in regards to human agency. The processes of structuration are not static. They occur at both multiple spatial scales and various points in time and therefore act on one another through the regular interaction of individuals, as well as institutions. And it was structuration's interest in the changing geographies that result from the interaction between people and

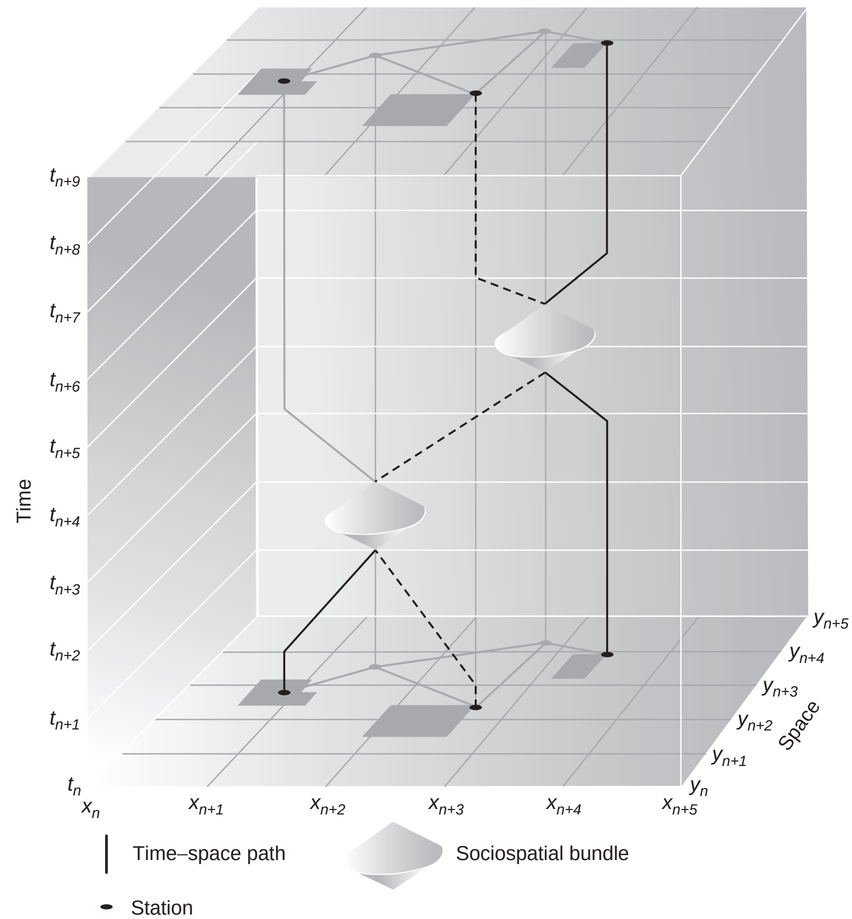

Figure 1 The time space model.

institutions that produced local, regional, and global change over time that contributed a reconceptualized regional geography. As a result, a great deal of geographic research examined how local communities and local decision makers interacted with more global structures – such as economic restructuring. Still other geographers were interested in how local politics create a distinct sense of community and place distinct from neighboring cities or even nearby neighborhoods.

While structuration was a popular approach throughout the 1980s and appeared to capture the sociological complexities of the structure–agency dialectic – two prominent critiques were eventually levied by geographers. First, structuration could not be explicitly linked to the personal scale of individuals. As such, structuration does little to truly afford people agency *per se* and/or related individuals to local structures. Second, structuration was ahistorical and in the here and now – or at the very most could only be used to understand contemporary change. More importantly, Gidden's innovative theory was insufficient to capture the importance of history on placemaking – specifically, the theory was unable to capture uninterrupted space–time flows.

Place, Biography, and History: New Regional Geographies

To address the deficiencies of structuration, geographers interested in place construction and the historical development of place began to situate people and their biographies (personal narratives) within known social structures to understand the role of agency in producing and reproducing local (and even global) social structures. For these historical time geographers, charting individual paths and institutional projects were essential to understanding places. As such, the everyday biographies of people (and how these biographies are shaped by contemporary institutions) could be used to unlock the hidden dynamics of place over time.

The basics of biography formation – as it would be known and articulated by Allen Pred – borrowed from Haagerstand's imagery of the path asserted personal everyday tasks have both spatial and temporal characteristics and the collection of these activities are necessary to complete institutional or societal goals. In addressing the role that biography plays, biography is defined as uninterrupted individual paths through time and space which is at all times and across all spaces (or in space–time) constantly subject to multiple constraints. The formation of individual biographies and its simul taneity with the already constructed and continuously modified physical environment is inextricable. That is to say that as one travels through the physical environment, their thoughts are inherently linked to action. Con sequently, individual thoughts, feelings, and actions are derived from the memories acquired through a lifetime of their daily paths intersecting with other paths and institutional projects. In short, individuals make decisions based upon the influences of intersecting paths on a daily basis as defined by already existing power relations and institutional projects or much more succinctly, by means of chance. It is the intersection of a collection of individual paths that construct a shared sense of place and history.

The interactions of human agency and the 'natural' landscape – and the landscape's constituent elements – are so deeply embedded in the process of becoming that extrication is not possible. These elements are the product of all previous interactions among locations, humans, laws, norms, mores, and things within the physical environment. The result being that alteration of both natures (nature and outer nature) are directly linked to localized struc turation and a myriad of relevant power structures. Those power relations, which are both historically contingent and place specific combine, dialectically, ensuring that citizens must accept the power of an institution or project over themselves. As place becomes, individual and collective biographies are formed as normality overtakes the exotic granting the authority to no longer question one's inter actions along daily and life paths. Biographies are shaped by imprinting their experiences upon the material and physical landscapes. In addition to localized actors upon the development of biography, as well as the sense of becoming place are extralocal actors. A key determinant of whether or not a place will be altered by noninhabitant entities or individuals is the degree to which they are or are not embedded within that place. Specifically, 'who is making the decisions at/for the local level?' is a question which should be explored. Obviously, if decisions are made by local actors then the biography will retain its current trajectory of becoming.

Throughout the late 1980s and through the 1990s, structuration emerged as a dominant theme in new cultural and regional geographies. In contrast to the descriptive approach of earlier researchers, 'new' geographers were building on the work of Pred and others to understand (and more importantly explain) the persistence of culture in place and over time. For example, the historical geographies of Gillian Rose actively explored how a shared history and contemporary politics can be used to construct and pre serve a sense of identity and self sufficiency within East London's Poplar district. Put simply, shared histories and the sustained politics of local places that are derived from the compounded actions of human agents within an existing structure simultaneously reproduce sameness and differ ence and preserve 'place'. As such, the geographies of Rose, as well as other new cultural and regional geographers moves beyond simple geographies of the way places 'used' to be and use space–time to explain observed social conditions and the persistence of place. In this sense, place, as Alan Pred asserts, is a historical contingent concept and process.

Planes, Trains, and WiFi? Technology and Space

Technology plays a crucial role in altering the landscape that geographers investigate. Indeed, technology func tionally and literally reconfigures how people interact with each other, the firm (e.g., workplace), their com munity, and larger social structures. As the popular author and New York Times columnist, Thomas Fried man, often notes, technology has functionally removed barriers between people, countries, and continents. While proponents of his 'world is flat' paradigm envision tech nology as a great egalitarian equalizer, the reality is that it – like all capitalist structures – is inherently contra dictory and produces an abundance of uneven social relations. While it is debatable whether technology (particularly information technology) will make global structures more equitable, it is undeniable that rapid technological change continues to alter the landscape and redefine the everyday lives of people in local places around the globe.

Technology–Space–Time Compression

To understand the transformative role technology plays in geography, it is essential to refer to quantitative geo graphy's emphasis on distance and geometric primitives. When considered within the context of distance and the relationship between distance and time, technology con tributes to space–time compression. That is, technology has the capacity to make space (e.g., distance) 'virtually' irrelevant insofar as the distance between two places as measured in convenience (e.g., time) is decreased.

The wider debate surrounding the role of technology in producing space–time compression focuses on not only compression (e.g., reducing the functional signifi cance of space) but also convergence and distanciation. A (re)conceptualization of space–time is necessary as technological innovation continues at a tremendous pace and these innovations are transforming how people, firms, and institutions interact in meaningful ways at increasing distances instantaneously and with limited disruptions. By facilitating time–space compression and compressing space–time (which are related but distinct), technology reconfigures the geography of everyday life. To begin the process of understanding these changes, recent conceptual modifications of space–time and related applications are presented in the following paragraphs within the context of situated narratives, networks, and place. This section will begin with advances in transportation and will close by focusing on the space–time implications of new social spaces – specifically the Internet – and information technologies inclusive of GIScience.

Transportation

The distance that personal space could incorporate with the addition of the automobile radically altered the landscape of towns, cities, regions, and states. Resulting in the requisite of (re)conceptualizing one's personal space, the extent of one's daily paths, professional space, but most profoundly – urban morphology. Cities were no longer developed with the pedestrian as the center of concern. As the automobile became an ubiquitous aspect of contemporary society, the needs of combustion trumped those of the sustainable commuter. No longer did citizens have the option of ignoring technology; instead, a dialectical relationship, which continually unfolds, was borne between automobiles and humans, impacting the urban morphology, biography, and more generally becoming. As people and their cars spend increased time together, a codependence develops as the car becomes an extension of its operator. Informatics, ergonomics, and artificial intelligence, each influence the ability for technolgies to be extensions of the user. Things interacting and in instances anticipating the actions of their human operator have transformed things distinct from humanity into expressions of human agency. It is this extensibility afforded by the seemingly perpetual evolution of technology that allows for cyberspace to become a new social space rather than a technology limiting social intercourse.

The Ubiquity of Space, Time, and Power in the Age of the Blogosphere: Human Agency in Action

Geographies of time and space and the role technologies play in creating new sociospatial structures – such as the Internet – are rapidly changing the ways in which people interact with each other, engage in economic enterprise, maintain social structures, and reconfigure social (and even physical) space. More importantly, the proliferation of new complex and user friendly technologies have afforded individuals seamless access to new worlds, economic opportunities, and sociospatial relationships, as well as empowered and legitimized individual critiques of dominant social structures. From the geographer's perspective, as Paul Adam asserts, technology is reducing the friction within, between, and across both physical and virtual space and the result is to broaden the scale, scope, and impact of individuals on social structures. Insofar as technology is increasingly ubiquitous in the developed world and new notions of extensibility (the capacity for individuals to overcome the friction of physical, virtual, and social distance), the role of the individual in promoting social change has and will continue to evolve as new technologies converge and access to these technologies expands.

While the utopian values (e.g., more consumer choice, access to information, telecommuting, and so on) are often attributed to the Internet, the promise is only now being realized. However, the Internet's most dramatic development has been in the area of politics and current events. Specifically, technology has expanded the space for 'other' voices from other 'places' and in doing so afforded new legitimacy to those alternative voices vis à vis increased access, technology integration, and media convergence (the integration of telecommunications, print, video, and electronic resources). Conceptually, the Internet has centered traditionally marginalized voices, rescaled narratives, and transformed local politics into global forces. Functionally, the World Wide Web has proven to be an effective 'real time' mechanism for mobilizing spatially disparate groups vis à vis complex technical and social networks to affect real social and political events at a variety of scales (e.g., scale jumping).

One of the most visible mechanisms for social change on the Internet is the 'blog'. Initially conceptualized as either recreational or reflective activities akin to daily journaling, blogs have become highly specialized spaces that focus on everything from current events to athletics. Since the mid 1990s, blogs have evolved from simple online diaries to hybridized chronologies of everyday life that include political analysis, edgy commentaries, and interactive editorials that allow authors to communicate directly with their readers. In the process of specialization, blogs became a recognized vehicle for news delivery – particularly stories (and gossip) situated at the margins of the traditional print and broadcast media. By the turn of the century, blogs had become a powerful force in the public discourse that enabled writers, journalists, and would be journalists to forward controversial positions that were often difficult to place in more commercial outlets through the space of new media. Simultaneously, the blogosphere is able to forward new unheard stories from writers and individuals who authored daily diary entries from around the world, in combat zones and transitioning regions. While these stories do not enjoy the visibility of celebrity bloggers like 'Wonkette' Ana Marie Cox or Andrew Sullivan, these personal web diaries provide new texture to the complex geographies of people in local places embedded in very real global conflicts, facing the economic uncertainty of globalization, or struggling to promote social change in their local communities.

Beyond the blog, converging technologies from file sharing to instant messaging and activities (e.g., online equities trading, shopping, etc.) are transforming global and local flows of data, restructuring capitalism, and redefining normative values. Yet, as the geography of cyberspace continues to evolve it is becoming apparent that new hierarchies and sociospatial relationships are emerging which both challenge and reinforce existing social structures. For example, the blogosphere has become a highly articulated space comprising conservative, progressive, and radical networks. In concert, these networks have developed sophisticated strategies for moving news, opinions, or YouTube downloads from the realm of new media to the mainstream media to forward their ideological agenda. In this sense, a new – albeit alternative – collection of gatekeepers (which has simultaneously re visioned and reproduced existing social relations) has evolved and these new social relationships are shaping and defining the structure and flow of information at all scales.

Beyond the blog, geographic research on the space–time implications of the Internet has already demonstrated that information technologies have altered the everyday time geographies of individuals in key ways. For example, the Internet relaxes space–time constraints on individuals in urban areas. Specifically, the web provides new intervening opportunities to shop online, communicate, and obtain information thereby offering individuals new opportunities and flexibility to access urban resources at different times and locations. Of course, the ability of cyberspace to relax space–time constraints is uneven and determined by a variety of access issues (to be discussed later). In addition to providing new opportunities in cyberspace that relax real world constraints, geographers are also investigating ways in which cyberspace is informed by and informs contemporary conceptualizations of space and place, redefines cartography, and affords ready access to a myriad of opportunities to visualize a great number of space–time conceptualizations ranging from Internet maps of real time instant messaging activity to the dizzying world of Second Life.

In many respects, cyberspace has defined new spaces and places that are responsible for redefining institutions at multiple scales. Specifically, the convergence of technology facilitates the globalization of local events and the ability of local agents to challenge traditional social structures. Going forward, it is the challenge of geographers to unlock the complexities of new media convergence (and technological convergence more broadly); chart the new paths of virtual space, and identify the constraints that structure these new geographies. Moreover, the capacity for the Internet (and no doubt a successor technology) to facilitate new and exciting space–time interactions will continue to expand and redefine the landscape of time geography.

Time as Unifying Force: GIScience, Human and Physical Geographies

This section examines emerging discourses on space–time, specifically as it relates to GIScience and the potential for time to be recast as a dominant theme that unites an increasingly fragmented discipline. While one of geography's greatest strengths is its ability to spatialize virtually any phenomenon, human geography's aggressive and expansive spatiality has created a cleavage between physical and human geography and more recently aided in the development of GIScience (which examines both human and physical geographies). Since human geography's spatial term has not been paralleled in physical geography, it can be argued that many potentially rich areas of enquiry – particularly human–environment interactions, public policy, or applied geography – have not been fully realized. Given this reality, recent 'space–time' advancements in the field of GIScience may have the potential to facilitate new linkages between human and physical geography – and in doing so strengthen the relationship between GIScience and the discipline. In short, space–time has the potential to reintroduce the physical world (broadly conceptualized) to human geography and repopulate the world of physical geography.

Time in GIScience

The role of time and history in contemporary geography is often a difficult and complicating factor for any researcher. In the case of GIScience, time proves especially problematic as it – unlike space – is a difficult variable to model in a digital landscape. While certainly geographers can chart change over time in the form of simple land use and land cover change models, the exact role of time – as a factor or force – is unknown. Instead of trying to capture the relative impact of time as a process, time is reduced to a simple variable that can be frozen, dissected, and examined as a snapshot independent from events on either side of its chronological ordering. This response is not unexpected as it mirrors the earlier work of quantitative geographers who embraced the time series. Yet, the Swedish school's time geography demonstrated that time was indeed an important factor to account for when performing geographic research. By rejecting that phenomena or activities can be independent events, time geographers asserted that phenomena (particularly social action) represented an accumulation of earlier interactions with people, spaces, and institutions over both time and space. While geography was radically altered by the Swedish concept of time, data collection proved problematic (at multiple scales, particularly the individual) and limited in depth enquiry.

Temporal GIS and Spatiotemporal Modeling

Geographic information system (GIS) and the field that would become GIScience has long understood the importance of accounting for and modeling space–time elements. Like earlier quantitative geographers before them, the reality of developing a temporal GIS proved especially difficult and collecting data at the individual scale was frustrated by software and hardware limitations. Due in part to these limitations, initial efforts at building a temporal GIS adopted the 'snapshot' approved that assessed change over time. Yet, this approach – as mentioned earlier – fails to capture the complexities of space–time and the degree to which both space and time structure human interactions in the social and physical world. By the late 1990s though, hardware improvements, new data structures – specifically object oriented databases, enhanced commercial software, and emerging concepts from computer science were beginning to lay the foundation for more dynamic spatiotemporal modeling. In concert, these advancements were enabling GIScientists to visualize and model human and physical activity in both space and time. More recently, wireless technologies coupled with location aware technologies such as Global Positioning System (GPS) are enabling real time data collection at the level of the individual.

Today, GIScience and the related field of visualization are well positioned to realize the ambitious goals of Hägerstrand and the Swedish School. For example, technology now permits the visualization and analysis of daily paths through space–time. The ability to model space–time now allows researchers to ask (and answer) new questions and opens up entirely new research possibilities. As Mei Po Kwan's research illustrates, GIScience models that interrogate space–time and associated constraints have the capacity to yield new insights into the everyday lives of individuals and cities. In the process, Kwan's analysis of gendered urban opportunities problematizes traditional 'locational' models that tend to focus on distance and the less time sensitive variable of commute times. Still other researchers have proposed using a range of novel techniques such as cellular automata to situate people in space–time. Likewise, GIScientists are interrogating the intersection between virtual and physical space and how the interaction of these two produce new space–time prisms that simultaneously situate individuals locally in physical space and globally in virtual space.

As these examples underscore, space–time research in GIScience is pushing the boundaries of both human and physical geography. While the above examples are primarily human oriented, advancements in visualization and space–time are also transforming geomorphology, climatology, and nearly all subareas of physical geography. However, GIScience has the potential to unite human and physical geography insofar as spatiotemporal GIS applications can be used to investigate human–environment interactions. For example, GIScience has the capacity to model human responses to natural hazards such as flooding or hurricanes and in the process contribute to the formation of more efficient and effective emergency planning. In any event, GIS technologies will no doubt empower geographers to (re)conceptualize space–time as it relates to the morphology of the landscape, as well as the effects of human agency and their related biographies on the environment.

The Problem(s) with Space–Time

Time geography and space–time are contested concepts within and outside the discipline. While many geographers acknowledged the importance of time and space–time and still others have diligently worked to develop methodologies for unlocking the everyday geographies of peoples and places at multiple scales and around the world; time geography in general has been widely critiqued as unimportant, reductionist, imprecise, and even ethnocentric. Over time though, these same critiques have led to important methodological advances including new methodologies for charting space–time and the integration of new technologies to more accurately measure movement across space and over time. Likewise, critiques have led to conceptual advances – like structuration and more nuanced understandings of place. Indeed, the critiques have prompted geographers to reconsider the meaning of space and the historical context that informs our working definition of it. And in the process, contemporary human geography has expanded our understanding of space–time in unique and creative ways and challenged the historical underpinnings of space and space–time.

Perhaps the most vexing facet of the space–time debate is the degree to which time or temporality is (or should be) embedded within the geographer's notion of spatiality. That is, how essential is time – as a concept – to the investigation of human or physical geography. Accordingly, Doreen Massey asserts time is an essential construct and process for geographers to investigate as conceptualizations of space and place change over time and are inherently linked to the specific historical geographies from which meanings are attributed in both the physical and human 'worlds'. For this reason, if time is a unifying theme for the discipline, then interrogating the relationship between time and space is an essential task. However, the exact interaction between space and time and the role of space–time in geographic research is unclear insofar as contingents within the discipline of geography – aided by the expansion of GIScience – embrace natural science methodologies that have the potential to minimize the significance of space–time and the historical perspective.

In recent years, this debate within geography has been increasingly centered on the evolving relationship between geography, other disciplines, and interdisciplinary subfields. While time geography has strengthened the connection between geography and the social sciences, particularly sociology, its impact on the natural sciences has been less distinct. For this reason, critical discussions of the themes of space–time and change over time and space in fields such as geology, geomorphology, and geoarchaeology have been more or less absent from the literature. For example, how might space–time be used to more fully understand – not necessarily explain – geologic processes? Likewise, can space–time be used to forge new linkages between disciplines and illuminate the unique contributions of geography to the subfields of geomorphology and geoarchaeology? That is to say, the interconnectedness of spatial and temporal characteristics in geomorphology, geoarchaeology, and even biogeography serve to underscore the geography in each. Or alternately, to what degree have allied subfields developed increasing aspatial research agendas rooted in the scientific rigor of 'hard' sciences that serve to marginalize both the spatiality and temporality of each. Having said this, the potential exists for subfields in allied disciplines to demonstrate how space–time can effectively be used to understand fluvial systems, landscape ecology, and human settlement – and in the process unite physical and human geography. It is essential for both human and physical geographies to constantly reconceptualize space–time and develop an appreciation for historiography.

The Differential Velocity Issue, Technology Gaps, and Other Complications

The efficacy of time geography – as conceptualized by the Hägerstrand 'Swedish School' – has been questioned from a variety of perspectives. First, modeling space–time is problematic as it focuses almost exclusively on 'physical' interactions. As a result, time geography assumes shared access to social and physical infrastructures that facilitate known and discrete interactions. Unfortunately, as critics note, the concepts presented in the prism model are exceedingly diffuse. Indeed, what exactly constitutes a coupling/bundling, intersection, or path? Do these basic concepts truly represent the inherent complexity of paths through space–time or is the notion of a single 'path' itself too simplistic? Similarly, are the accumulated influences of intersections and bundling in the real world equivalent to virtual intersections? What types of interactions produce change and which interactions serve to reproduce existing structures? As these questions suggest, the reductionist and physical nature of time geography poses a range of methodological complications relative to accuracy and emerging technologies. More importantly though, the meaning (or significance) of time geography hinges on the answers to these very real concerns.

A second and salient critique of time geography and emerging geographies of time (particularly cyberspace) are rooted in differential velocity (i.e., the absence of uniform velocity). That is, access to key technologies and infrastructures is highly uneven and how (and at what rate) people across the street or around the globe proceed through physical and virtual space differs greatly. For this and related reasons, the most elegant models and collection of potential path trees (e.g., a map of potential interactions) are unable to account for the reality that billions of people do not and will not reasonably have access to key space compressing technologies such as the Internet, airplanes, or even cars. Moreover, many conceptualizations of space–time are decidedly Western in orientation and subject to charges of ethnocentrism. Consequently, the ability of blogs to capture everyday geographies in places around the world is highly suspect. Further, space–time (and its inherent linkage to technology and related infrastructures) is situated in a historical and spatial context and may or may not be relevant to the experiences of individuals in developing contexts. Therefore, space–time is an inherently contextual concept informed by an individual researcher's own positionality. Within the context of geography as a 'science', the situational and subjective nature of space–time limits its adoption and legitimacy across the discipline.

However, differential velocity and access alone are not merely infrastructure issues. Time geographies are also derived from more global processes, flows, and associated power relations. The result of these more global processes has been the creation of what Doreen Massey calls a 'power geometry' that is inherently uneven as a result of time–space compression and rapidly revised notions of space. In short, the instrumental (or technological) facets of globalization are creating new and ever shifting collections of social relations, institutional contexts, and social divisions of space from which an increasing collection of people are (or may be) absent, disengaged, or in Massey's more dramatic terms 'imprisoned'. In a very practical sense, the power geometry of globalization and its spatial and temporal dynamics has contributed to not only uneven geographies at the global scale but also at the local scale. Hence, the geometries of decision makers, followers, and the disenfranchised are produced, reproduced, and deepened in place as a result of the interaction between time and space. In this sense, space–time and time geography reside in a highly contested landscape facilitated by technology (e.g., time–space compression) and characterized by a new social division of space.

As these critiques suggest, scientific geography's emphasis on uniformity, specificity, and measurement continue to place constraints on the ability of time to inform the geographer's understanding of space. Yet, the strength of time geography, the elegance of the prism model, the efficacy of social networks, or the utility of the temporal may not be their collective ability to replicate the real world experience – but their challenge to social scientists (and natural scientists, too) to imagine the world in space–time and the diverse collection of social, physical, and virtual forces that influence human decision making and alter the landscape.

Conclusion

Nearly a century after the Berkeley School conceptual ized the interconnected nature of human agency and the physical environment – arguably widening the gap be tween geography's physical and human realms – GIScience is allowing the reunification of geography's two primary research areas by incorporating facets of human ecology and time geography to (re)conceptualize space–time. Likewise, continued efforts to interrogate the meaning of physical space, imagined space, and the spatiality of post structural thought have made time geography a fertile ground from which geographers can reconceptualize the relationship between chorology and chronology. Theoretical debates such as post modernism, post structuralism, structuration, and actor network theory have enabled the researcher to study and consequently more fully understand the dialectical relationships among human agency, biographies, societal projects, the physical environment, and the roles of lan guage and text. Technology is redefining space–time as space–time concomitantly redefines the tools of technology and the manner in which geographers materialize the world. These modifications in the land scape of geographical thought ensure the debate is far from over as to how 'best' to define the spatiotemporal characteristics central to geography. Thus, both space and time and the resulting space–time need to be regularly (re)conceptualized to ensure appropriate research methodologies, as well as for the development of theo retical concepts for both the social and physical sciences. A new way forward, as it turns out, is not altogether new.

See also: Critical GIS; Diffusion; GIS and Society; GIS, Mobile and Locational Based Services; Positivism/ Positivist Geography; Space-Time Modeling; Spatial Science; Structuration Theory; Time and Historical Geography.

Further Reading

Adams, P. (1995). A reconsideration of personal boundaries in space time. *Annals of the American Association of Geographers* 85, 267 285.

Adams, P. (2005). *The Boundless Self: Communication in Physical and Virtual Spaces*. Syracuse, NY: Syracuse University Press.

Brabyn, L. and Skelly, C. (2006). Modeling population access to New Zealand hospitals. *International Journal of Health Geographics* 1, 3.

Giddens, A. (1986). *The Constitution of Society*. Berkeley, CA: University of California Press.

Hagerstrand, T. (1967). *Innovation Diffusion as a Spatial Process*. Chicago, IL: University of Chicago Press.

Kwan, M. (1999). Gender and individual access to urban opportunities: A study using space time measures. *The Professional Geographer* 51, 211 227.

Kwan, M. (2002). Time, information technologies, and the geographies of everyday life. *Urban Geography* 23, 471 482.

Kwan, M., Janelle, D. and Goodchild, M. (eds.) (2003). *Special Issue: Accessibility in Time and Space*. *Journal of Geographical Systems* 5(1).

Langran, G. (1992). *Time in Geographic Information Systems*. London: Taylor & Francis.

Lenntorp, B. (1999). Time geography: At the end of its beginning. *GeoJournal* 48, 155 158.

Massey, D. (1999). Space time, 'science' and the relationship between physical geography and human geography. *Transactions of the Institute of British Geographers* 24(3), 261 276.

Massey, D. (2005). *For Space*. London: Sage.

Miller, H. (2007). Place based versus people based geographic information science. *Geography Compass* 1, 503 535.

Pred, A. (1977). The choreography of existence: Comments on hagerstrand's time geography and its usefulness. *Economic Geography* 53(2), 207 221.

Pred, A. (1984). Place as historically contingent process: Structuration and the time geography of becoming places. *Annals of the Association of American Geographers* 74, 279 297.

Sykora, L. (1994). Local urban restructuring as a mirror of globalization processes: Prague in the 1990s. *Urban Studies* 31, 1149 1196.

Space-Time Modeling

M. Yuan, The University of Oklahoma, Norman, OK, USA

Glossary

Conceptualization A mental process to abstract and generalize reality or a problem domain.

Data Models or Database Models A formalization of objects, attributes, and relationships in information systems.

Georelational Database Model A way to organize spatial data by linking sets of spatial objects (i.e., points, lines, and polygons) and attribute tables to characterize geography.

Geospatial Lifelines The space–time path of an individual's continuous movement with a sequence of time-stamped locations in geographic space.

Object-Oriented Database Model A way to organize spatial data in which a data object encapsulates both spatial and nonspatial properties.

Ontology A specification of a conceptualization for knowledge sharing and reuse.

Representation A surrogate for reality in physical, graphical, or digital forms.

Introduction

Space and time are two fundamental dimensions in geography. As early as 1961, geographer James M. Blaut promoted the idea of space as the basic organization of geographic concepts and "every empirical concept of space must be reducible by a chain of definitions to a concept of process" (Blaut, 1961). His argument repre sents a classic space centered view of geography and a process based approach to unite space and time. Indeed, uniting space and time is essential to understanding geographic worlds as space–time interactions drive change and movement in geography.

Space–time modeling, in the context of information science, refers to spatiotemporal information constructs of reality and organizations of these information con structs. Specifically, geographic information science (GIScience) approaches to space–time modeling aim to capture essences of space, time, and geographic things manifesting themselves in space and time in support of information analysis and mathematical modeling. To be explicit, space–time information modeling may be used to distinguish GIScience approaches from mathematical approaches. While there is only one reality, information constructs and their organizations can be perceived in many distinctive ways from different perspectives.

GIScience approaches to space–time modeling in volve three main issues: conceptualization, represen tation, and data modeling of geographic space over time. Conceptualization is a cognitive process that abstracts and generalizes meaningful properties of space and time as well as identifies geographic entities or features that are significant to problems on hand. At the highest level of distinction, conceptualization of space and time can be absolute or relative. The Newtonian view considers space as a container populated with objects, whereas Einstein rejected the idea of empty space. In Einstein's *Relativity*, space–time exists not on its own but as a structural quality of a field.

The absolute versus relative conceptualizations of space and time correspond to object and field based representations of reality. Geography, on the one hand, exhibits discreteness as many geographic entities can be conceptualized as individual wholes, such as countries, lakes, wells, and roads. On the other hand, continuous properties are also apparent in many geographic fields, such as elevation, soil moisture, population density, and census tracts. In some cases, individual wholes, such as lakes, can be transformed conceptually into a continuous field, such as lakeness. Temporal extensions to object and field based representations have helped enhance geographic information system (GIS) abilities to repre sent complex geographic features and their relationships beyond what can be recognized in static object or field only perspectives.

Conceptualization processes determine geographic entities and relationships significant to particular space–time concerns. Subsequently, representation processes address how space and time should be framed in a GIS and how geographic entities and relationships should be coded digitally. Representation involves creating a surrogate for reality, and in the context of GIS, a digital surrogate. Representation of space and time is a complex, multidimensional issue, and existing spatiotemporal databases encompass only limited properties of space and time. How geographic entities are represented in a GIS determines the type of queries that can be supported and the accuracy of GIS solutions. Any entities that cannot be represented are considered nonexistent in the GIS. Results from GIS analysis are consequently limited to what can be represented in the system. Simply put, if a GIS is incapable of representing three dimensional (3 D) features, the GIS cannot support analysis of groundwater flows or movements of people in a high rise building. In a nontemporal GIS, temporal objects and relationships are considered nonexistent.

Following representation, space–time modeling involves the development of formal models that specify coding schemes for entities and relationships. In addition to considerations of objects, fields, entities, and relationships, formal models of space and time detail specific data structures, attribute sets, data types, and data linkages. Georelational database models that link tables of geometries and attributes to represent geographic entities have been popular in GIS applications. Later object oriented data models have attained much attention to address the need to encapsulate geometry and attribute in structuring data objects as well as hierarchical relationships among data objects. Modern GIS data models tend to be object relational, taking advantages of both relational and object oriented approaches in computation and integration. Specifications of data models are expressed in a data modeling language, such as entity–relation diagram or more recently Universal Modeling Language (UML) with computer aided software engineering (CASE) tools. In addition to space, space–time data models are further categorized depending upon the way and type of time incorporated in the space–time model.

GIScience approaches to space–time modeling address issues on conceptualization, representation, and data modeling of space and time. The following discussions highlight the major developments since the late 1980s, except for algorithmic development.

Incorporating Time into Database and GIS

Time can be categorized as world time (also referred to as 'valid time' when the fact of interest is true in the modeled reality), transaction time (also referred to as 'database time' when a database fact is current in a database), and user defined time (i.e., an attribute defined by the user). A database without transaction time or valid time encoded is a snapshot database. If transaction time is encoded, transaction relations support rollback databases in which every transaction relation can be mapped to exactly one system supported transaction time. On the other hand, valid time relations support historical databases because a valid time relation maps to exactly one system supported valid time. When a database has both valid time (i.e., the intrinsic time aspect of a fact) and transaction time (i.e., the extrinsic time aspect of a fact), we have a bitemporal database. Bitemporal relations are considered 'fully temporal relations' because when combining historical relations and transaction relations, the database enables incorporation of both intrinsic and extrinsic time aspects of facts and therefore provides the full story.

Temporal database systems and temporal GIS share common thoughts in research developments. Specifically, both communities apply various time stamped approaches to handle temporal data. Time stamps may correspond to world time, transaction time, or user defined time. Temporal database systems mainly apply time stamping techniques to incorporate time into relational or object oriented databases by time stamping tables, attributes, or values. In parallel, temporal GIS adopts the snapshot model by time stamping individual data layers, the space–time composite model by time stamping individual attributes associated with a spatial object, and the space–time object model by time stamping individual values with attributes of a spatial object (**Figure 1**).

The space–time composite model and space–time object model face similar challenges in spatiotemporal data management. Because both models only allow spatial objects with uniform properties, a given attribute can only have one value at the given time for a specified spatial object. However, temporal properties may change to only a portion of the spatial object. Both models require reconfiguring a new set of objects (or atoms in the case of space–time object models) to account for the partial changes to a former spatial object. Consequently, the space–time composite model must assign spatial identifiers to the new set of spatial objects. Keeping temporal lineages of spatial objects before and after a spatial reconfiguration is not a trivial task, especially when the study area becomes more and more fragmented over time. The space–time object model is able to overcome the issues related to tracking temporal lineage of spatial objects and is useful for recording cadastral histories, for example, but it cannot represent movement, dispersion, and complex geographic processes.

From Space–Time to Geographic Dynamics

Representation of change, movement, dispersion, and other geographic processes requires thinking geographic dynamics in space–time modeling.

Change and movement are observables of geographic dynamics. Without change or movement, geographic dynamics arising in space and time will go on without being noticed. Brian Berry's geographic matrices of places and characteristics are effective means of spatial representation to support place to place comparison of geographic variables, and when arranged in a temporal sequence for a set of places, geographic matrices can facilitate study of changes in spatial association and areal differentiation.

In analyzing possible dimensions that geographic dynamics can manifest, Goodchild *et al.* proposed a

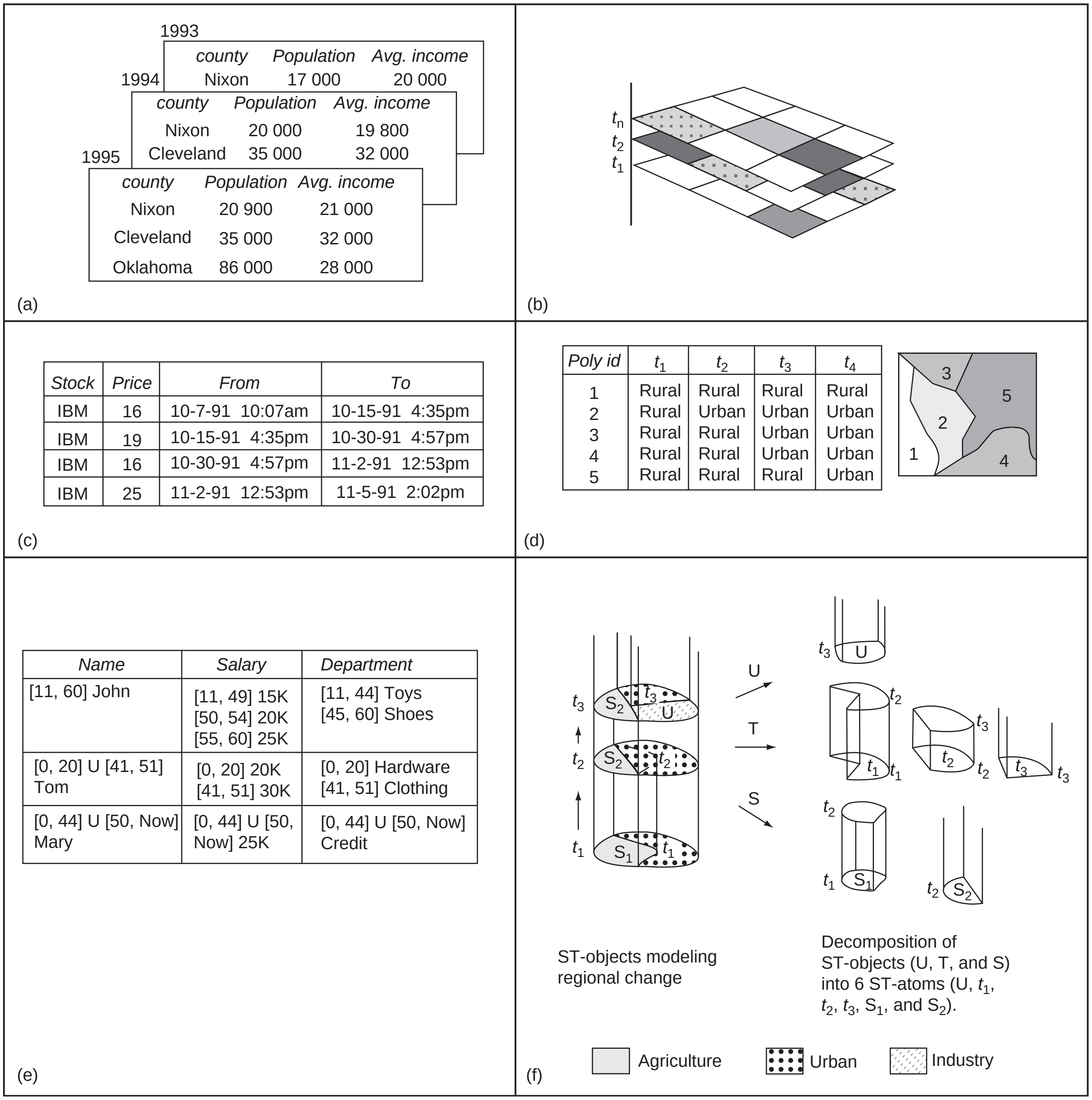

Figure 1 "Similar developments in incorporating time into database systems and GIS. Panel (a) time-stamped tables; (b) time-stamped layers; (c) time-stamped tuples (rows): an ungrouped relation; (d) time-stamped attributes (columns): space time composites; (e) timestamp values (cells): a group relation [11, 60] represents a period starting at t11 and ending at t60; and (f) time-stamped space time objects: the spatiotemporal object model. Part (c) from Snodgrass, R. and Ahn, I. (1986). Temporal databases. *IEEE Computer* **19**(9), 35 42; part (d) from Langran, G. and Chrisman, N. R. (1988). A framework for temporal geographic information. *Cartographica* **25**(3), 1 14; part (e) from Gadia, S. K. and Yeung, C. S. (1988). A generalized model for a relational temporal database. *Proceedings of ACM SIGMOD International Conference on Management of Data*, pp. 251 259. New York: ACM; and part (f) from Worboys, M. F. (1992). A model for spatio-temporal information. *Proceedings: The 5th International Symposium on Spatial Data Handling*, pp. 602 611. Bristol, PA: Taylor and Francis. Whole figure modified from Yuan, M. (1999). Use of a three-domain repesentation to enhance GIS support for complex spatiotemporal queries. *Transactions in GIS* **3**(2), 137 159.

conceptual cube to identify eight basic types of possible transformations of geospatial phenomena (**Figure 2**). The cube provides a basis for GIScience researchers to conceptualize different ways of space–time modeling to capture a full range of geographic dynamics. Currently, GIScience research addresses well attribute changes and location changes (i.e., movement). Significant research efforts are needed to address changes to geometry and changes to internal structure. With attempts to sum marize developments in this research area, the following

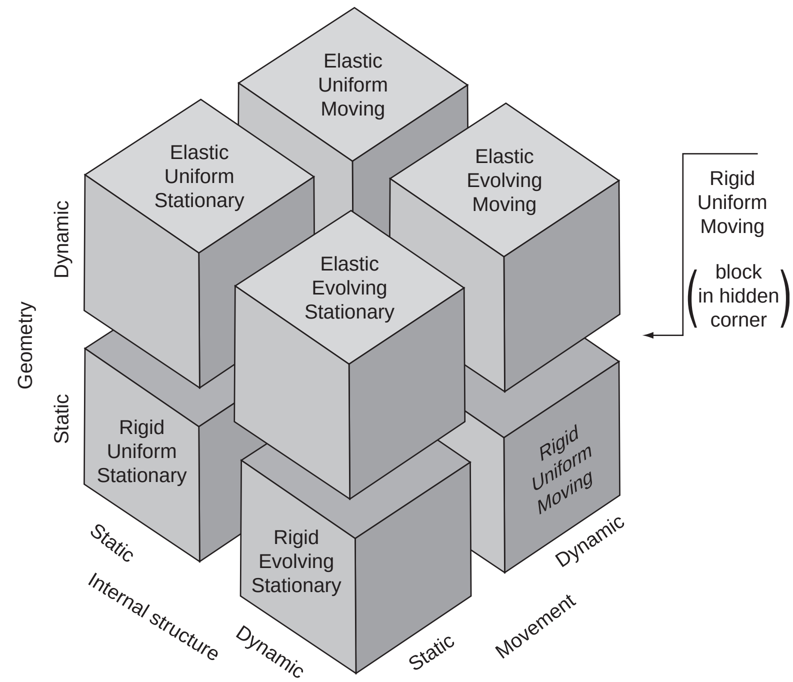

Figure 2 Dimensions of geographic dynamics. From Goodchild, M. F., *et al.* (2007). Towards a general theory of geographic representation in GIS. *International Journal of Geographic Information Science* 21(3), 239 260.

sections discuss change and movement, in which change is mostly related to change in attributes. Discussions on changes to internal structure and geometry are limited due to the infancy of the research emphases.

Change in the Context of Geographic Dynamics

Change is often inferred by comparing properties and forms over time, and inferring processes from forms to understand geographic dynamics. Sauer's school of cul tural–historical geography stressed the importance of research to record, illustrate, and synthesize geographic patterns and explore geographic processes. In doing so, historical and cultural geographers build a 'retrospective science' of cultural and historical processes in geography that enables us to look forward. Geographic dynamics pervades space and time, and their pervasiveness em powers the past as the key to understanding the present and predicting the future.

The time stamped approaches in space–time data modeling are well suited for change detection and change analysis. Change detection can be performed by spatial overlays of consecutive snapshot layers. Intuitively, one would attribute the discrepancies between two snapshot layers as change from t_1 to t_2. However, there are many potential causes for the discrepancies, such as data quality, scale issues, and ontological consistency. To as sure change, one must relate the discrepancies to "actual processes, not just the geometry of space–time" (Chris man, 1998). The space–time composite model addresses this issue by representing change explicitly in GIS databases and builds histories at locations with temporal transitions of uniform properties over space. The space–time object model uses changes to identify space–time atoms and assemble them to space–time objects. Change is also central to the event based spatiotemporal data model (ESTDM) in which an event is recognized by change to attribute values in raster cells, and the model organizes locations of change in a time series to facilitate temporal data retrieval and analysis (**Figure 3**).

In order to determine 'change' based on comparing attributes at two different points in time, we assume that the observed location or object remains the same identity from t_1 to t_2. Persistent identities (i.e., the identifier of a location or an object remains constant in the database after updates and other data manipulations) allow the observer to recognize and trace the same objects for change analysis. Hornsby and Egenhofer proposed an identity based change model with a set of primitives and operations. The set of primitives include 'identity states of objects' and 'transitions'. There are two basic identity states of objects: existence and nonexistence. Transitions express the progression of an object from one state to another, that is, temporal sequences of identity states of an object. Moreover, the model attaches two qualifiers: 'with history' and 'without history', to existence and nonexistence and therefore results in four possible identity states. Operations then can be used to determine

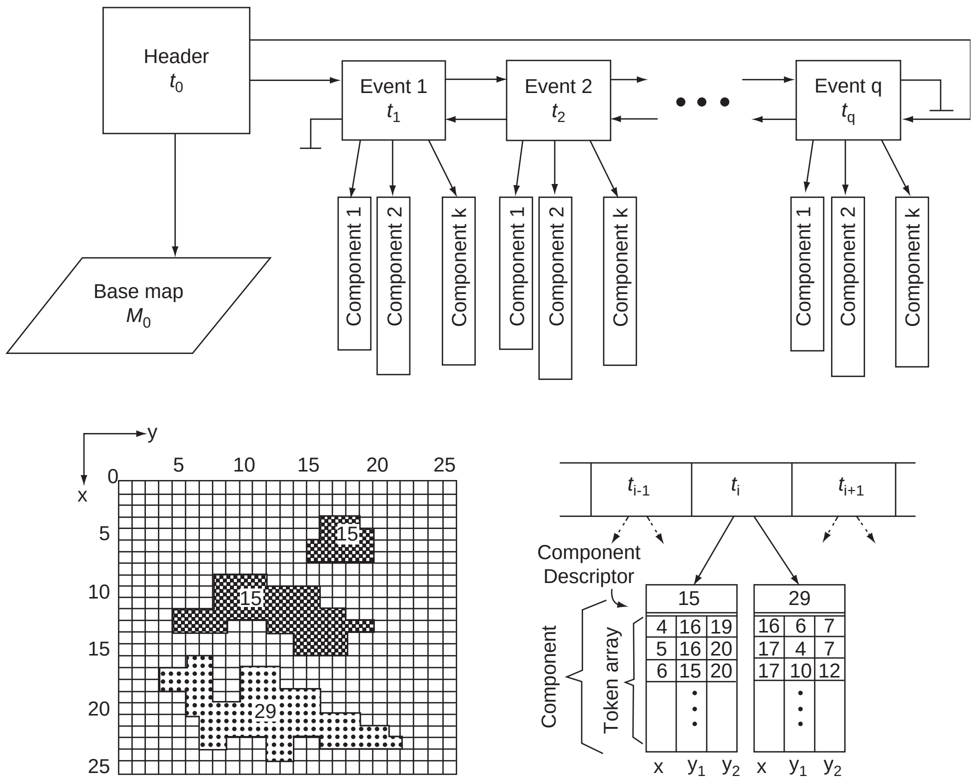

Figure 3 The event-based spatiotemporal data model. From Peuquet, D. J. and Duan, N. (1995). An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data. *International Journal of Geographical Information Systems* 9(1), 7.

the legality of transitions from one of the four possible identity states to another.

When changes occur to shape (or geometry) and internal structures, multiple spatial representations are needed to record the varying spatial properties. The spatiotemporal object model applies spatiotemporal atoms to capture geometries at different times and forms spatiotemporal objects by aggregating spatiotemporal atoms (**Figure 1**). As long as spatiotemporal objects keep persistent identities, spatiotemporal objects represent well the development of a defined property (such as a land use type) over time. One constraint to the spatiotemporal object model is that the defined property needs to be homogeneous (i.e., no spatial structure or variance) over space and time.

Changes to the internal structure of geographic objects have only been recently recognized in GIScience literature, particularly for some geographic objects exhibiting spatially heterogeneous properties. For example, storms move as individuals, but a storm's properties, such as precipitation intensity and temperature, vary spatially within the storm. Another example relates to cities. When a city grows, its internal structure, such as demographic distributions and population density, changes accordingly. Both the internal structure of geographic objects and changes to the internal structure are resulted from dynamics operating in and carried by the geographic objects (**Figure 4**). Complex, hierarchical spatial objects are needed to trace changes in the internal structure in which large scale objects constitute spatial aggregates of component objects at a smaller scale. One solution is the use of 'field objects' to emphasize internally varying spatial properties. A hierarchical field object data model was proposed to capture the entities as individual and their embedded field properties (**Figure 5**).

Movement in the Context of Geographic Dynamics

Hägerstrand's *Time Geography* developed in 1970s provides the conceptual basis for movement analysis in GIScience. Time geography is based on a simple idea that an individual's movement can be considered as a space–time path in a 3 D (2 D space and 1 D time) cube. Depending upon the individual's constraints on time and means of transportation, a space–time prism can be constructed to determine the potential space where an individual can reach at any given point on the individual's space–time path. The potential spatial extent that one can reach provides an estimate of opportunities for activities, such as shopping and running errands.

The combination of time geography and GIS analytical and visualization functions has promoted research on

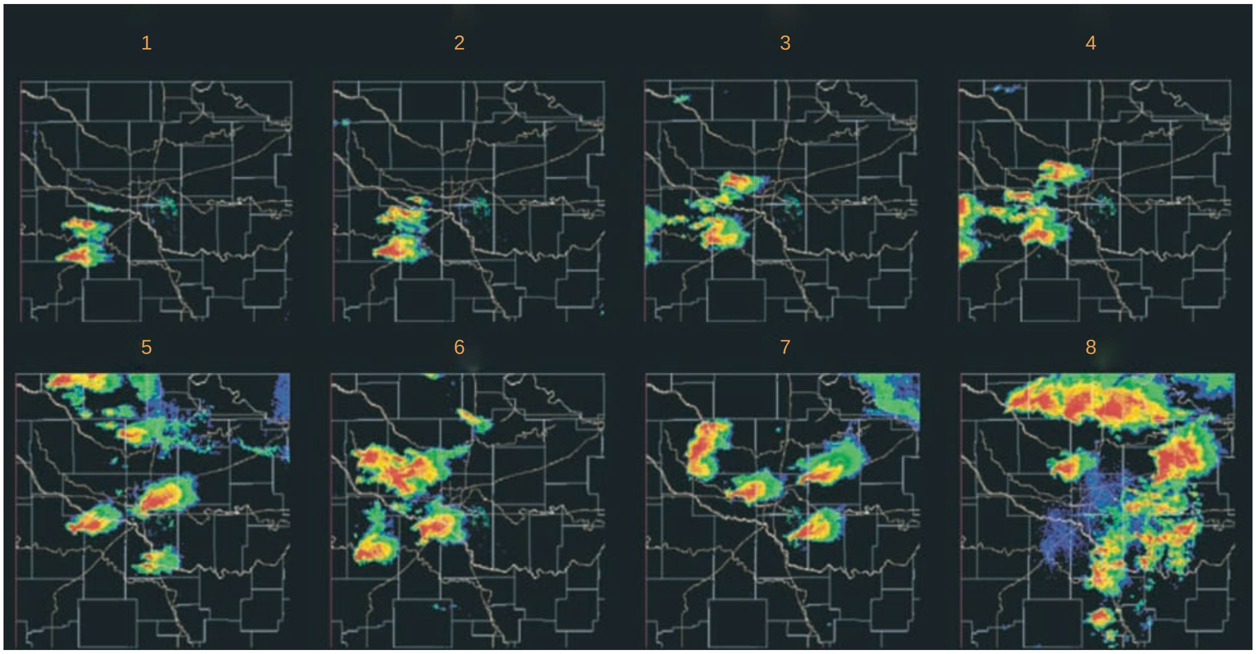

Figure 4 Examples of geographic objects with spatially variant properties. Precipitation intensity varies spatially within each storm cells. The panels are National Weather Services NexRAD images in a time sequence during the 3 May 1999 Oklahoma City tornado outbreaks. From Yuan , M. and K. Stewart. (2008). *Computation and Visualization for the Undertanding of Dynamics in Geographic Domains: A Research Agenda*. CRC/Taylor and Francis.

human activity patterns and use of space by different populations, such as the space–time paths (**Figure 6**) of women with different ethnicities and demographic char acteristics. Hybrid activities (i.e., activities with multiple purposes) and embedded meanings became apparent when spatially contextualizing these activity paths. For example, Kwan's study of women's movements in Portland, Oregon, suggests that African American women appear more geographically confined when compared with Asian and Caucasian women. Kwan's research on the daily activities and feelings of Muslim women after the 11 September event reveals 'a geography of fear' by generating space–time trajectories from her surveys to recognize where Muslim women feel most threatened in Columbus, Ohio.

With space–time prisms of multiple individuals, an alysis can be made to the degree of spatiotemporal 'co incidence' in which these individuals' space–time paths coincide – for example, a group of people meeting at a coffee shop. Employing GIS overlay concepts, inter sections among space–time prisms can be resolved to determine the potential interactions among individuals. In addition, spatial and temporal constraints in com munication, and how the constraints are affected by new communications technologies have been analyzed via space–time modeling in GIS.

The idea of space–time paths can be applied to daily activities or long term geospatial lifelines, such as one's history of residency. Geospatial lifelines are temporally ordered discrete points sampled at regular or irregular time intervals, and therefore, geospatial lifelines are essentially stored as multiple points with estimates of intermittent paths among these points in a GIS database. When contextualizing environmental factors in GIS, geospatial lifelines can reveal one's environmental ex posures, other life experiences, and proximity of resi dence over a period of time. They can be also used for fine tuning a lifeline's temporal granularity or general izing the lifeline's pattern at a coarse scale. Speed and direction provide the reference for movement patterning and parameters for interpolation and generalization.

While geospatial lifelines represent individuals as points, spatiotemporal helixes (STH) (see **Figure 7**) consider 2 D objects and track the objects' outline de formation over time. An STH consists of a spine and prongs. Spines are the trajectory of the center of a given object, and prongs represent the expansion or collapse of the object at a given time. When identifying geographic features from motion imagery, the STH model first applies self organizing map technique to identify spines and then uses the deformable contour models (a.k.a. snakes) to elicit prongs.

Space–Time Ontologies and Beyond

GIScience approaches to space–time modeling can be addressed in the frame of space–time ontology. Rooted in philosophy and linguistics, ontology is "the metaphysical

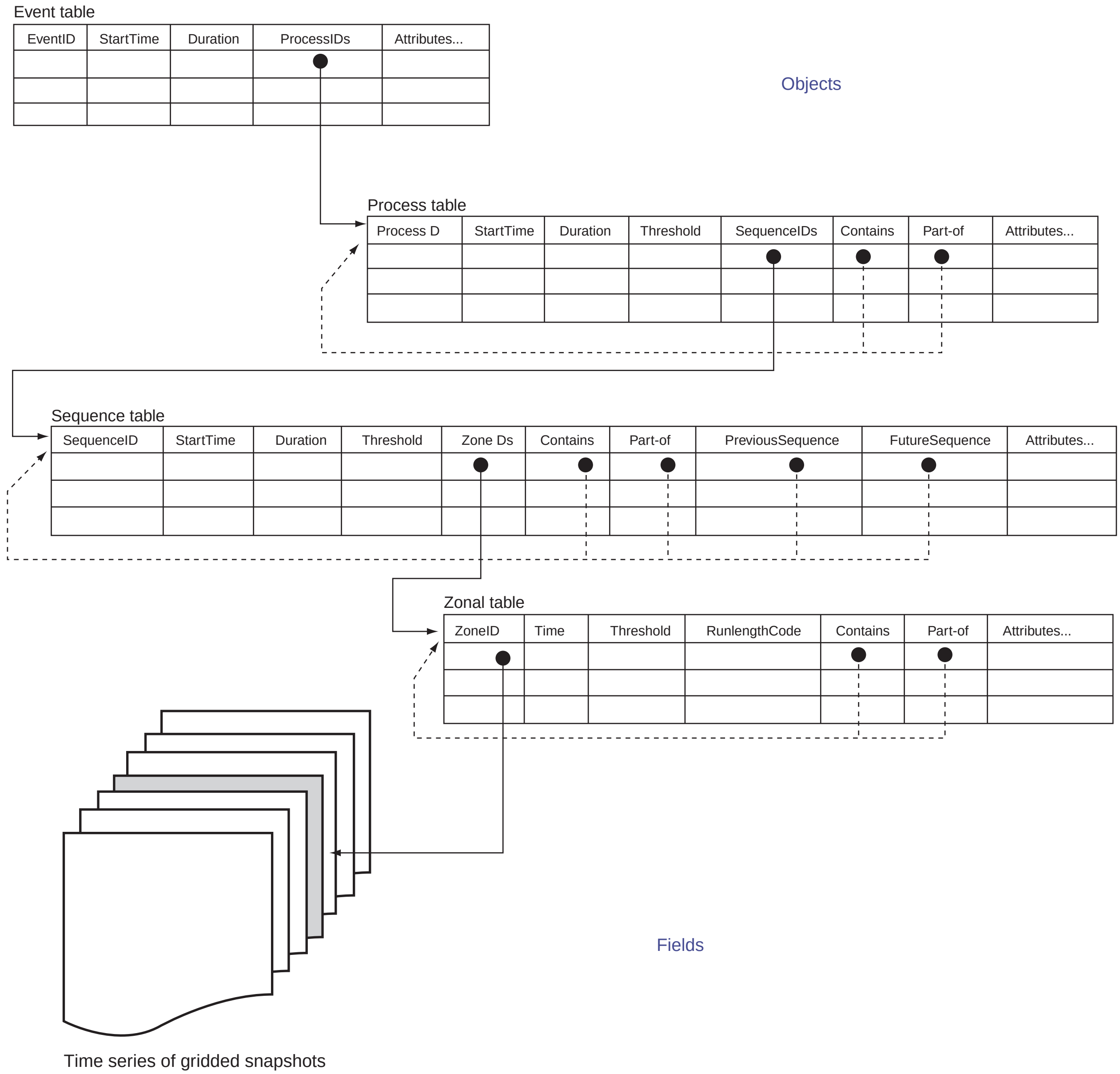

Figure 5 A hierarchy of grids and vector attribute tables to capture internal structures of rainstorms. From McIntosh, J. and Yuan, M. (2005). Assessing similarity of geographic processes and events. *Transactions in GIS* 9(2), 223 245.

study of the nature of Being and Existence" (Fellbaum, 1998). When applied to information science, ontological study examines fundamental elements, concrete or abstract, in our world. GIScience approaches to space–time modeling need to consider spatial ontologies, temporal ontologies, and their integration.

SNAP and SPAN present upper level ontologies to distill concepts about space–time occurrents and space–time continuants. Both ontologies have been addressed in GIScience approaches to space–time modeling, and strong connections to the intended domain of applications are important. There are two alternative ontologies of geographic dynamics: SNAP considers successions of instantaneous snapshots of the world and SPAN emphasizes a unified view of space and time. The SNAP ontology centers on existence at a defined instant, and consequently, SNAP entities have no temporal parts; they are spatiotemporal occurrents. SPAN entities, on the other hand, are continuants; they may have existed and will continue to exist for some time.

The snapshot model and geospatial event model follows the SNAP ontology. This is because a GIS considers objects on different data layers as different objects, even if they represent the same geographic entities at different times. Changes can be identified by comparing two snapshots or states before or after an event through

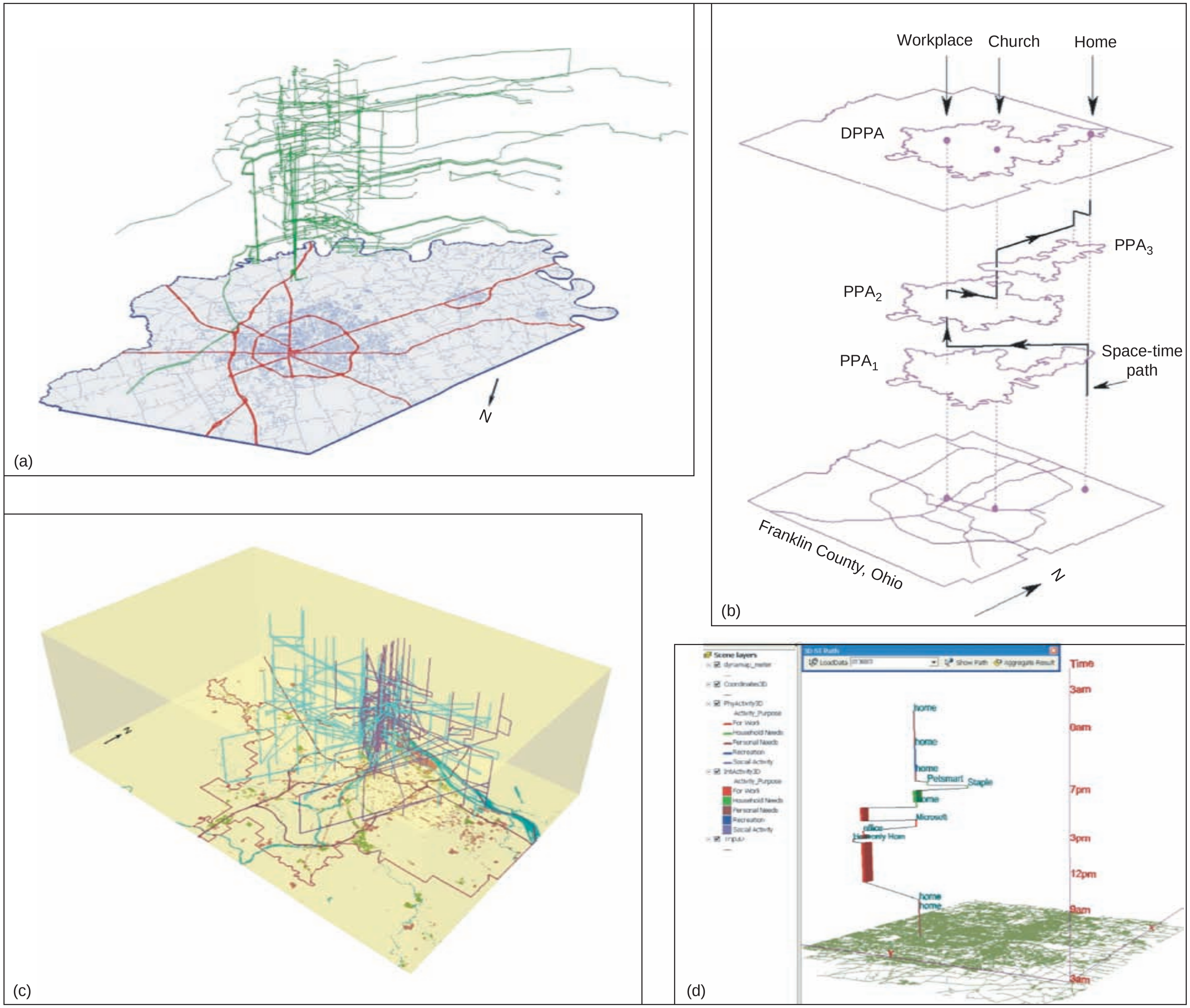

Figure 6 Space–t me paths of nd v dua s' da y act v t es. (a) Space–t me paths of act v t es from a g ven popu at on; (b) space–t me paths annotated w th centers of act v t es and the r nkages; (c) an aggregat on of space–t me paths from mu t p e popu at ons; and (d) the v sua zat on and ana ys s of human act v t es. From Kwan, M.-P. (2004). G S methods n t me-geograph c research: Geocomputat on and geov sua zat on of human act v ty patterns. *Geografiska Annaler*, *Series B: Human Geography* 86(4), 267–280.

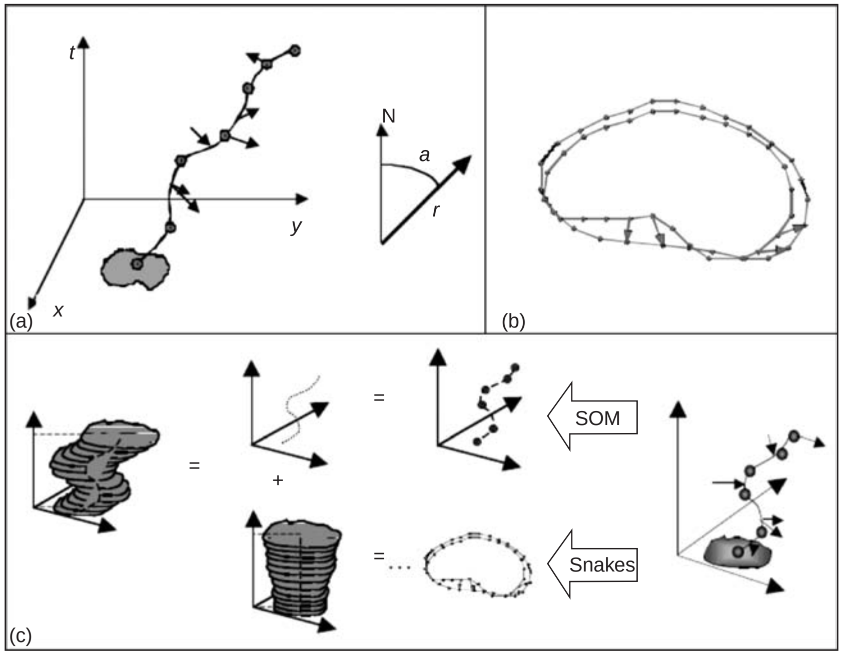

Figure 7 The spatiotemporal helix model with a spine and prongs (a). The model is based on extracting moving objects from motion imagery. Spines are identified using self-organizing map techniques, and prongs are derived from differential snakes method (b and c). Spines represent objects' movement trajectories, and snakes keep tracks of outline deformations. From Stefanidis, A., Eickhorst, K. et al. (2003). *Modeling and Comparing Change Using Spatiotemporal Helixes*. The eleventh ACM international symposium on advances in geographic information systems. New Orleans, LA: ACM Press.

spatial overlays in GIS. A SPAN entity, on the other hand, consists of temporal parts which constitute its history or lifeline. The space–time composite model, spatiotemporal objects model, event based spatiotemporal data model, and process based field object model all address the SPAN ontology. Because the SPAN ontology considers the continuum of an entity over space and time, SPAN based space–time models support identification, infor mation, and analysis of deformation and movement.

Research efforts are needed to expand SNAP, SPAN, and other ontologies relevant to spatiotemporal databases to include geographic dynamics, processes, events, and activities for a comprehensive coverage of the eight possible geographic dynamics in **Figure 2**. In addition to unifying space–time knowledge, ontologies have been shown to be an effective means for data integration. Ontological understanding of and ontological commit ments to geographic dynamics can promote conceptual integration and exchanges of space–time modeling in GIS databases to strengthen GIS facility to spatio temporal analysis and geocomputation models. Similar to the development of diverse GIS functions to address object and field based geospatial data, comprehensive treatments of SNAP, SPAN, and other space–time ontologies shall result in rich sets of GIS functions for spatiotemporal analysis and geocomputation methods to address the full spectrum of dynamics in geographic worlds.

Summary

GIScience researchers have made theoretical and tech nological advances in space–time modeling since late 1980s. From the perspective of information systems, space–time modeling address issues about conceptual ization, representation, and data modeling of spatio temporal properties in geography. Diverse views of space and time promote distinctive conceptual models of space and time as continuous, discrete, absolute, or relative. These conceptual models are captured by a wide spectrum of space–time representational schemes and accordingly implemented in numerous spatio temporal data models. Ultimately, the goal of GIScience approaches to space–time modeling is to capture geo graphic dynamics in GIS databases and computation. Change and movement, as the two primary observables of geographic dynamics, have been the foundation in the development of space–time data modeling and analysis. The above discussion highlights some of the space–time data models proposed in the GIScience literature and the analysis of their representational power in capturing spatiotemporal properties. The emerging ontological study of geographic dynamics promises to promote knowledge sharing of space, time, and dynamics in geography. A robust ontology of geographic dynamics can serve as a foundation to further integrative and innovative GIScience in space–time modeling for a wide range of physical and human research.

See also: Simulation.

Further Reading

Blaut, J. M. (1961). Space and process. *The Professional Geographer: The Journal of the Association of American Geographers* 13(4), 1 7.

Chrisman, N. R. (1998). Beyond the snapshot: Changing the approach to change, error, and process. In Egenhofer, M. J. & Golledge, R. G. (eds.) *Spatial and Temporal Reasoning in Geographic Information Systems*, pp 85 93. New York: Oxford University Press.

Fellbaum, C. (ed.) (1998). *WorldNet: An Electronic Lexical Database. Language, Speech, and Communication*. Cambridge, MA: The MIT Press.

Galton, A. (2004). Fields and objects in space, time, and space time. *Spatial Cognition and Computation* 4(1), 39 68.

Goodchild, M. F., Yuan, M. *et al*. (2007). Towards a general theory of geographic representation in GIS. *International Journal of Geographic Information Science* 21(3), 239 260.

Grenon, P. and Smith, B. (2004). SNAP and SPAN: Towards dynamic spatial ontology. *Spatial Cognition and Computation* 4(1), 69 104.

Hornsby, K. and Egenhofer, M. (2000). Identity based change: A foundation for spatio temporal knowledge representation. *International Journal of Geographical Information Science* 14(3), 207 224.

Kwan, M. P. (2004). GIS methods in time geographic research: Geocomputation and geovisualization of human activity patterns. *Geografiska Annaler, Series B: Human Geography* 86(4), 267 280.

Longley, P. A., Goodchild, M. F. *et al*. (2005). *Geographic Information Systems and Science*. Hoboken, NJ: John Wiley.

McIntosh, J. and Yuan, M. (2005). Assessing similarity of geographic processes and events. *Transactions in GIS* 9(2), 223 245.

Miller, H. (1991). Modeling accessibility using space time prism concepts within a GIS. *International Journal of Geographical Information Systems* 5(3), 287 301.

Peuquet, D. J. and Duan, N. (1995). An event based spatiotemporal data model (ESTDM) for temporal analysis of geographical data. *International Journal of Geographical Information Systems* 9(1), 7.

Sinha, G. and Mark, D. M. (2005). Measuring similarity between geospatial lifelines in studies of environmental health. *Journal of Geographical Systems* 7(1), 115 136.

Stefanidis, A., Eickhorst, K. *et al*. (2003). *Modeling and Comparing Change Using Spatiotemporal Helixes*. The eleventh ACM international symposium on advances in geographic information systems. New Orleans, LA: ACM Press.

Yuan, M. (1999). Use of a three domain representation to enhance GIS support for complex spatiotemporal queries. *Transactions in GIS* 3(2), 137 159.

Yuan, M. (2001). Representing complex geographic phenomena with both object and field like properties. *Cartography and Geographic Information Science* 28(2), 83 96.

Yuan, M., Mark, D. *et al*. (2004). Extensions to geographic representations. In McMaster, R. & Usery, L. (eds.) *Research Challenges in Geographic Information Science*, pp 129 156. Boca Raton, FL: Taylor and Francis.

Relevant Websites

http://www.cs.auc.dk
Aalborg University, Department of Computer Science: consensus glossary of temporal database.

Spanish Language Geography

A. Albet, Autonomous University of Barcelona, Bellaterra, Spain
P. Zusman, University of Buenos Aires, Buenos Aires, Argentina

Glossary

Association of Spanish Geographers Association created in the 1975 gathering of both academic and professional geographers in Spain.

Latin-American Geographers' Meetings Congress of Latin American geographies and geographers. Held every 2 years, it is the most successful and massive meeting of the region.

Modern Geography Project Discipline proposal inspired by European geography dating from the beginning of the twentieth century and intended to disrupt the Encyclopedic Geography project in place in Latin America.

Pan-American Institute of Geography and History Created in 1928 under the auspices of the Pan-American USA policy, to promote knowledge in areas such as cartography, geography, history, and geophysics.

Spanish Language Geography Geography developed on countries where Spanish is the main official language (Spain and Latin America).

Geography in Spain

Slow Beginnings

In Spain, the Public Instruction Act (known as 'Ley Moyano') that was passed in 1857 rationalized the numerous and chaotic syllabi and reforms of public instruction that had appeared since the beginning of the nineteenth century. This law became the basis for all the legislation related to education until the early decades of the twentieth century and firmly established the teaching of geography in primary and secondary schools with basically a descriptive bias and closely related to history.

One could expect, therefore, that many chairs of geography should have been created in Spanish universities and that the institutionalization of geography would have occurred at an early date. But, in fact, the few chairs of geography that were created were subordinated to other curricula and were taken by professors trained in other fields, as was the case with geologists, botanists, and mathematicians in the Faculty of Sciences and historians in the Faculty of Arts. In the latter, historical geography was initially taught, but at the end of the century it was changed to political and descriptive geography. At the university level, geography was also included, for several decades, in the Faculty of Sciences under titles such as cosmography and Earth physics, geodesy, phytography, and botanical geography. In spite of the steps taken by the Geographical Society of Madrid (created in 1876, the first one in Spain), in response to this situation, there was no significant change until the 1940s. Geographers were unable to influence the decisions of the Spanish government as long as their scientific projects were in competition with other projects.

Geography without University Geographers: The Contribution of Catalan Geographers

Academic geography ("la géographie des professeurs" in Yves Lacoste's words) was almost nonexistent until the 1940s. But things were different with the utilization of geographical knowledge and with the professional practice of geography outside the university. Geographical knowledge was highly valued in the nonacademic circles (like cultural associations, as well as political and ideological lobbies). According to the cultural initiatives, from its beginning, there was a close connection with the hiking movement, especially in those regions of Spain where this movement had a strong nationalistic flavor as was the case of Catalonia. On the political–ideological contributions, the geopolitical approach brought by geographers (in the form of books, articles, and open lectures) should not be overlooked. They mainly refer to Spain's colonial problems or to the geostrategic position of the Iberian Peninsula in the Mediterranean and the European context. It has to be pointed out that this strongly ideological and geopolitical discourse persisted during the first years of Franco's dictatorship (especially between 1940 and 1945) due to its close connections with the German National Socialist régime.

During the 1930s, professional practice of geography in Catalonia showed a vitality, quality, and modernity that clearly contrasted with its low level of institutionalization in the Spanish universities. In 1932, the Generalitat (the Catalan autonomous government) assigned the geographer Pau Vila to prepare a project of territorial division. Although Vila was considered to be a follower of the Vidalian tradition of regional geography (he published the first regional monography in Spain following Vidal de la Blache's proposals), together with a group of professionals, he planned the new administrative map of Catalonia that was really very innovative for his time. Vila (who had never taught at the university but was a recognized teacher of primary school and an educator in training schools) and a small group of geographers also designed, for the first time in Spain, a syllabus to obtain a specific university degree in geog raphy at the University of Barcelona: this syllabus had to be implemented in the 1936–37 academic year but the Civil War destroyed this project, among many other things.

Times of Growth

Geography as an academic discipline, both at school and at university level, was held in high esteem by the Franco régime (1939–75). Geography (as well as history) was conceived as a tool to transmit the national–fascist ideology, to reinforce the sense of unity of the country, and to enhance the Spanish patriotic spirit. For this reason, teaching geography in primary and secondary schools was expanded but its presence at the university level was more conspicuous in the Faculty of Arts than in the Faculty of Sciences, where it was represented by only one chair of physical geography in Spain. Nevertheless, topics and approaches uncongenial with the dominant ideology were taken out of the official syllabi. Lacking a critical methodology, and the most prominent geog raphers being in exile, geography easily became a de scriptive and apolitical discipline.

Between 1940 and 1965, the scientific community of geographers increased slowly in parallel with the rela tively low number of pupils at secondary schools and universities: in 1959, there were only eight full time professors of geography in the Faculty of Arts and four in the Faculty of Sciences. Between 1965 and 1980, the number of university students enrolled in the Faculty of Arts increased from 17 000 to 100 000, and geography gained advantage by this process. This spectacular growth was a key factor in the institutionalization of academic geography, now almost exclusively concentrated in the Faculty of Arts, where specific units of geography ap peared during the 1970s. Growth was impressive in every aspect that could be considered: the emergence of new units, the number of students, the number of doctoral dissertations, the publication of new journals, etc.

For instance, before the 1970s, geography was present in the Spanish universities only through a dozen subjects, including those taught in the Faculty of Sciences by geologists. As a matter of fact, in the Faculty of Arts only three subjects were taught, that is, general geography, geography of Spain, and world's descriptive geography. But at the end of the 1970s, about 150 different titles were taught in Spanish departments of geography. It is important to point out that in 1974, a decree was passed that allowed the creation of departments wherever spe cific geography courses were taught, and this eventually led to the creation of specific curricula in geography in some universities. This formed the basis for the impressive increase in the number of students majoring in geography during the 1970s and 1980s.

Again, the pioneer role that Catalan geography played at the end of the 1960s and the beginning of the 1970s has to be taken into account. Some young lecturers spent periods at geography departments in British and North American universities and became deeply interested in and receptive to new ideas. The atmosphere was one of methodological crisis and there was an urgent need to break away from the past (mostly the Vidalian tradition) and find more means of analysis and interpretation of geographical facts. Unlike the rest of the Spanish uni versities, in 1968–70 a flexible study curriculum was es tablished in the Faculty of Arts at the University of Barcelona and the Autonomous University of Barcelona, providing a strong basis for the development of modern geographical studies, introducing quantitative, Marxist, and humanistic geographies for the first time in Spain. As a matter of fact, for the first time in any Spanish uni versity, a degree in geography was created in 1968 at the University of Barcelona and it proved to be an important factor in giving geography a firm foothold in the Catalan university. In the recently founded Autonomous Uni versity of Barcelona, geography courses were imparted within the Department of Social Sciences, a vinculum which encouraged marked interdisciplinary overtones in geography and a definite tendency toward the develop ment of a very innovative human geography.

Introducing New Academic Systems

The University Reform Act of 1983 introduced a major set of regulations that, among many other things, allowed the majority of Spanish universities to offer, for the first time, the right to design their own study plans in geog raphy and, also for the first time, a full first degree in geography: in the past, universities had to follow plans laid down by the Ministry for Education, and geography was mainly considered within or as a part of a degree in history or in arts. For a minority of Spanish universities, in particular the Catalan ones, the Act proved to be a step backward because it established many compulsory com mon subjects that constrained the decisions on their syllabus and promoted a more traditional type of geography.

Since 1983, many new public and private universities have been created in Spain and a degree in geography is offered only in public universities. It is also remarkable that a new degree in environmental sciences created in 1995 is offered in 26 universities (four of them private) and often syllabi are partially provided from geography departments that do not take great advantage from the social and professional development of this new field. This degree has great social success and good marketing promotion, and geography students see it as one of their main sources of professional rivalry. Since 2001, the Autonomous University of Barcelona has been offering, for the first time in Spain, a full degree in geography through the World Wide Web.

Since 2005, the 'Bologna system' introduced a new degree similar to that in the rest of the European universities; in fact, the new system has been an occasion to seriously question the degree in geography, looking for impossible private market profitabilities in public universities. Anyway, restructuring of most geography departments and readjustment of faculty staff are probably unavoidable steps for recasting the discipline within the Spanish university system.

Apart from the first degree studies ('licenciatura'), some departments offer doctoral studies to obtain the PhD degree and many others participate in postgraduate programs usually with a very applied orientation. In many Spanish universities, geography finds inclusion in the syllabi of other related degrees like humanities, environmental sciences, social and cultural anthropology, tourism, sociology, political sciences, etc.

Publications and Foreign Influences

Despite the synthetic nature of many of the subjects taught, there are as yet few publications in Spanish language within the scope of students. A substantial part of the bibliography is made up of foreign books, only few being translated from British or North American text; before 1980, authors and works from the French school dominated. Although considerable efforts are being made by certain Spanish teachers and publishers, very few books by Spanish writers have been published. Paradoxically there are a large number of geographical journals, published by different universities.

Difficulties in fluently reading and writing in English have slowed not only publishing articles in foreign journals but even the normal use of basic texts. This is especially problematic when the Spanish university system demands its staff to publish in peer reviewed British–American journals. For many Spanish academic geographers, the Erasmus program of the European Union has brought, for the first time, the possibility of creating contacts and some research links with foreign universities. Due to the historical evolution of Spanish geography (and also budget shortages), the relation with Latin American geography is not as fluid and systematic as it should be and is supposed to be. Although at some time Spain has provided new texts to Latin American geography, the reverse has been very rare; thus Argentinian and Mexican publications are only occasionally present in Spanish geography.

Although not firmly recognized, the Vidalian tradition is still strong in many Spanish universities: classical regional geographies are common, and a very descriptive geography is still taught, although the introduction of some quantitative methodologies seem to give it a touch of modernity. In many circumstances, geography still balances between humanities and social sciences, but nowadays geography is almost definitively removed from the faculties of science.

The Role of the Spanish Academic and Professional Geographical Associations

The development of geography led to the creation of the Association of Spanish Geographers (AGE) in 1975 (with around 1200 members, mostly academic geographers). This association promotes geographical research as well as teaching, by encouraging academic meetings and general conferences every 2 years. It also tries to increase relations with other national and international geographic associations. In the mid 1980s four working groups were established: rural geography, quantitative methodology, physical geography, and teaching geography. Lately, more groups are established, that is, climatology, Latin America, population geography, regional studies, services geography, economic geography, tourism and leisure geography, urban geography, information technology, history of geographical thought, and local development. Since 1984, AGE has been publishing the biyearly 'Boletín de la Asociación de Geógrafos Españoles,' a well known reference journal in Spain.

In recent years, geography has established itself as an emerging profession in Spain, mainly in four different but related fields: city and regional planning, environment, cartography, and geographic information systems (GISs). Since the arrival of democracy in 1977, the creation of regional autonomous governments and the need for planners and policy managers at municipal level has brought an excellent opportunity for geography graduates and their skills, but concurrence with other professionals is strong. In this sense, academic curricula pay too much attention to market demands and consequently reinforce those technical and applied aspects in which the expertise of geographers can be most valuable. Still, a solid ground of theoretical concepts and intellectual resources should be secured as well as a clear ability for interdisciplinary work focusing especially on the interpretation and intervention on the territory with a global and integrated perspective.

However, the prospects for geography as a professional field appears to be vulnerable because of a great dependence on the public sector. For instance, about 70% of the members of the Catalan Professional Geographers Association (created in 1988) are employed in the regional and local administration. The rest of the members are employed in the private sector, mainly self employed, and a large proportion of them have the public sector as their main, final customer. In order to overcome this dependency, imaginative efforts should be directed to promote the presence of geographers in the private sector, mainly in service firms not related to the ad ministration. A significant step toward this end is the Official Spanish Professional Geographers Association (College, created in 1999), which is a useful tool to safeguard their professional rights and promote a wider social comprehension of what geography and geog raphers are able to do.

Because in Spain there is no direct relation between the academic degree in geography and its professional field, the popular image of geography is mainly linked with a descriptive, cartographic, and memoristic tradi tional field. Only very recently landscape, water, sus tainability, meteorology, and GIS are, in some way, newly identified with geography. Anyway, there is a possible drift by which geography becomes just a qualified pro fessional study with a low theoretical background, easily displaced by other fields with a more substantial social presence.

Geography in Latin America

Geographical Associations and Military Geographical Institutes

Between the end of the nineteenth and the beginning of the twentieth centuries, the term 'geography' in the Spanish speaking countries of Latin America was asso ciated with the practices of the modern nation states defining their territories (exploration, reconnaissance, and appropriation of areas under indigenous control; definition of international borders; and territorial dis putes). Geographical societies (as the Mexican Geo graphical and Statistical Society, 1833; the Argentinian Geographical Institute, 1879; the Argentinian Geo graphical Society, 1881; the Lima Geographical Society, 1888; the La Paz Geographical Society, 1889; the Sucre Geographical Society, 1900; the Geographical Society of Colombia, 1903) played an important role in the pro motion of territorial practices. Imitating their European counterparts, they had an active role in the production of information (descriptive, cartographic, and pictographic) about the land. Geography was seen as a useful know ledge that guides the training of geographic engineers in countries like Mexico, Colombia, and Argentina.

One of the first instances of the institutionalization of geography was through education. Knowledge of the physical and demographic characteristics was considered relevant to educating the inhabitants of the countries (often, ethnically differentiated) in the national values defined by the ruling elite. The participation of geog raphy in the creation of the imagined community of every Latin American country made the training of geography teachers a requirement in many countries at the beginning of the twentieth century. While the first curriculum and textbook proposals were devised on the basis of an encyclopedic geography, at the beginning of the 1920s this perspective was substituted by the modern geography project influenced by the French, German, North American (in the case of Cuba), and British proposals.

It should be stressed that geography had a strong presence in the military sphere. Thus, for example, the Military Geographical Institute in Argentina (1904), Chile (1922), Colombia (1935), or Bolivia (1936), were in charge, for a long time, of handling all issues related on geodesy and cartography in these countries. In some areas, these institutes ended up being the official geo graphical organization in each country.

From Applied Geography to Critical

Around the 1950s, in many Latin American countries, geography was separated from history and institutional ized as a degree within the Faculty of Arts or within the Faculty of Sciences (Argentina, 1948; Perú, 1948; Cuba, 1962; Venezuela, 1964; Chile, 1966). This institutional ization was accompanied by a deeper professionalism in the degree. In fact, the preexisting model, with clear references on a humanistic, erudite geography and pri marily devoted to the training of teachers, was brought into question to introduce a new model more directed toward the professionalization of the degree both in terms of research and as an outlet for new jobs outside the academia. It was perhaps in Cuba where this model of professional geography was first developed: there, from the very creation of the degree itself, research had an applied character to be useful for the economy, the productive processes, or the social organization, together with exchanges between Cuba and countries of Eastern Europe.

Of course, this level of professionalism did not mean that the traditional perspectives, now inspired by the regional model of Hartshorne, disappeared. In fact, re gional geography continued being strongly used in the sphere of teaching, and geography meetings organ ized by certain associations of geographers assured the continuity of this model (Association of Geographical Studies in Argentina, 1922; Chilean Association of Geographers, 1955; Colombian Association of Geog raphers, 1967).

Geography in Latin America was not oblivious to the political and social changes that were experienced on a world and regional level in the 1960s. It was the period in which a proposal for a critical geography was designed to prioritize the analysis of social issues from a Marxist perspective. At the same time, the idea of 'social space' appeared as a key concept of the analysis. In this context, two meetings on new geography took place in Latin

America: in Salto (Uruguay, 1973) and Neuquén (Argentina, 1974) where socially committed geographers, such as the Uruguayan Germán Wettstein and the Brazilian Milton Santos, as well as many students, participated.

With the dictatorships that were established in the Southern Cone, social geography lost all possibility of spreading. Many of the mentors had to go into exile, confining themselves to private centers or professional practice. Hartshorne's geography and the theoretical–quantitative geographies supported and legitimized the military plans while, at the same time, they removed social geography from the universities. The model of social geography, mostly recognized as critical geography, was taken up again around the mid 1980s when democracy returned to the region and when displaced geographers came back to the universities in the 1970s. The impact of critical geography had been different in each country. It was evident in Argentina, Uruguay, Venezuela, and Colombia. Milton Santos was one of the most influential figures in this area in all of the Spanish speaking countries of Latin America. His thoughts on space, globalization, and on the spatial dynamics of the economy, had an impact on almost all of the countries in the region.

The Recent Renaissance of Geographic Studies

Between the end of the 1980s and 1990s, many of the geography courses in the region had organized postgraduate studies in geography (Pedagogic and Technological University of Colombia, 1984; University of Panamá, 1993; University of San Marcos, Peru, 1995) or in related subjects (Masters in Land and Environmental Policy, Buenos Aires, 1995). Simultaneously many geographers in these countries went to study in Europe or North America. In certain national contexts, an attempt was made to eliminate geography as a subject and integrate it into the social sciences. Nevertheless, the interest raised by the environmental issues and by the GIS, have led to a renaissance in geographic studies. This can be seen in the fact that, in the last few years, new geography degrees have been organized in certain countries (Autonomous Metropolitan University, Mexico, 2002; University of Concepción, Chile, 2003; National University of Córdoba, Argentina, 2004).

Most of the geography graduates in Latin America become school teachers. Currently, however, a growing number of geographers can be found working in ministries and services related to infrastructure, transportation, irrigation, town planning, social programs, poverty programs and policies, education, housing, health, and state modernization and decentralization. Others work in city governments, in positions such as administrators, executives, or consultants in community development works, planning departments, education health councils, and environmental offices. The demand has also increased in the service sector and consulting firms on environmental issues, productive investments, and project location, as well as on consulting firms owned by geographers in fields related to regional planning, natural resources, environment, and cartography. Some of these geographers work as academicians in higher educational facilities in the teaching, research, and expansion of geographical knowledge. Finally, some have been integrated into international nongovernmental organizations of both regional and world scale.

The first precedents for organizing congresses oriented toward the whole region can be found in those periodically organized by the Pan American Institute of Geography and History. This institution was created in 1928 under the influence of the USA Pan American policy, and was oriented toward the promotion of knowledge in the areas of cartography, geography, history, and geophysics. Currently, the most successful regional events are the Latin American geographers' meetings, where all kind of themes, models, and positions are presented. To date, 11 such meetings have taken place: the first one in 1987 in Aguas de São Pedro (Brazil) and the last one in Bogotá (Colombia, 2007). The São Paulo (Brazil) sessions in 2005 brought together 2000 specialists from the region.

Fields of Research and Professional Interest

Spanish speaking Latin American geographers work primarily on environmental issues and regional planning and management. Social vulnerability, such as facing natural threats like hurricanes or floods, is a major issue for Central American countries. The application of GIS also attracts the attention of many of these specialists. Another theme that sparks the interest of geographers is the impact of globalization and its relation with investment in rural productive activities and in the process of valuation in urban areas. Latin American urban geography is devoted to study new functional uses of urban industrial land, new paths of urban consumption, new wealthy (gated communities) and poor districts, and the processes of gentrification; in general, these analyses are associated with the study of social exclusion (polarization, urban segregation). City, governance, and local government are also subjects of increasing interest for Latin American geographers. Rural geography seems to be interested in studying the way in which agrarian agents participate in the new productive procedures, in life strategies of small scale farmers, and in rural social movements, as well as the effects of neoliberal policies on industrial farming concentration and on regional restructuring. Lately, a new field of work lies on the re lation between rural areas, environment, and local development.

The strong relation between geography and education explains the relevance of geography in all kinds of edu cational plans all over the region. Epistemological and cultural approaches are tackled only by a small number of geographers. Geography of gender is mainly devoted to analyzing women's participation in the labor system in rural and urban areas, sometimes migrant, often through illegal jobs. A subject that has started to grow since the 1990s and that is attracting increasing attention is the relationship between tourism, land values, and local development.

Geography in the Spanish speaking countries of Latin America receives influences both from French and from British–North American geographies. But some authors think of a 'third geographical tradition,' following the theoretically hybrid character of present day geography, by which foreign influences are adapted through the lenses of other social sciences and through the local problems and perspectives.

See also: Anglo-American/Anglophone Hegemony; Francophone Geography; Geography, History of; Human Geography; National Schools of Geography; Possibilism; Santos, M.

Further Reading

Abalos, J. A. (2004). Resena de: La geografía profesional en Chile, pasado y presente, de Oscar Liendo. *Polis. Revista de la Universidad Bolivariana* 3(8), 1 9.

Berdoulay, V. and Mendoza, H. (eds.) (2003). *Unidad y Diversidad del pensamiento geográfico en el mundo. Retos y perspectivas*. México: Instituto de Geografía, UNAM.

Capel, H. (1976). La geografía espanola tras la Guerra Civil. *Geocrìtica* 1, 1 35.

Ccente Pineda, E. A. and LaTorre Ruiz, F. M. (2003). *El devenir de la geografía en el Perú*. Lima: Universidad Nacional de San Marcos. http://cybertesis.unmsm.edu.pe/sisbib/2003/ccente pe/html/index frames.html

Conferencia de Geógrafos Latinoamericanos (2006). El estado actual de la geografía en los países hispanoamericanos. In *Reunión de la Asociación de Geógrafos Americanos*, pp 1 49. Chicago. http://sites.maxwell.syr.edu/clag/GeoLA.pdf.

Dory, D. and Douzant Rosenfeld, D. (1995). Geografía y geógrafos en Bolivia y en Cuba: ensayo de sociología histórica comparativa. *Documents d'Anàlisi Geogràfica* 27, 57 73.

Garcia Ramon, M. D. (2005). Enfoques críticos y práctica de la geografía en Espana. Balance de tres décadas (1974 2004). *Documents d'Anàlisi Geogràfica* 45, 139 148.

Garcia Ramon, M. D., Albet, A. and Zusman, P. (2003). Recent developments in social and cultural geography in Spain. *Social and Cultural Geography* 4(3), 419 431.

Garcia Ramon, M. D. and Nogué, J. (1994). Nationalism and Geography in Catalonia. In Hooson, D. (ed.) *Geography and National Identity*, pp 197 211. Oxford: Blackwell.

Gómez, F. J. (1990). Los estudios e investigaciones geográficas en Euskal Herria. *Lurralde* 13, 9 18.

Gómez Mendoza, J. (2001). La geografía espanola: final y principio de capítulo. In Asociación de Geógrafos Espanoles (ed.) *Actas del XVII Congreso de Geógrafos Españoles*, pp 19 27. Oviedo: Asociación de Geógrafos Espanoles.

Hiernaux, D. and Lindón, A. (eds.) (2006). *Tratado de geografía humana*. Barcelona: Anthropos.

Mendizàbal, E. and Albet, A. (2005). Una aproximació a la geografia als Paısos Catalans (1985 2005). *Afers. Fulls de recerca i pensament* 50, 153 175.

Trinca, D. (2006). Los encuentros de geógrafos de América Latina. *Revista Geográfica Venezolana* 47(1), 5 7.

Valenzuela, M., Mollá, M. and de Lázaro, M. L. (2004). Geography in Spain. *Belgeo,* 2004(1), 145 158.

XXX IGU Congress, Spanish Contribution (2004). Spanish geography on the challenges of contemporary society. *Report of the XXX IGU Congress, Spanish Contribution*. Glasgow: International Geographical Union, Spanish Committee. (http://www.realsociedadgeografica.com/es/pdf/geografia ingles.pdf.)

Zusman, P., Castro, H. and Soto, M. (2007). Cultural and Social Geography in Argentina Precedents and Recent Trends. *Social and Cultrual Geography* 8(5), 775 798.

Relevant Websites

http://age.ieg.csic.es/
Asociación de Geógrafos Espanoles / Spanish Geographers Association.

http://www.geografos.org
Colegio Oficial de Geógrafos / Professional Geographers Association (Spain).

http://www.ingeba.euskalnet.net/
Euskal Geografi Elkargoa / Basque Geographical Institute.

http://www.ub.es/geocrit/menu.htm
Geocrítica.

http://www.geoinstitutos.com/
Geoinstitutos. Plataforma de los Institutos Geográficos Latinoamericanos / Geoinstitutos. Platform for Latin American Geographical Institutes.

http://www.ieg.csic.es/ieg.do
Instituto de Economía y Geografía CSIC / Economy and Geography Institute CSIC (Spain).

http://www.ipgh.org
Instituto Panamericano de Geografía e Historia / Pan American Institute of Geography and History.

http://www.realsociedadgeografica.com
Real Sociedad Geográfica / Royal Geographical Society (Spain).

http://www.sgxe.com/
Sociedade Galega de Xeografia / Galician Geographical Society.

http://scg.iec.cat/
Societat Catalana de Geografia / Catalan Geographical Society.

Spatial Analysis, Critical

T. Schwanen, Utrecht University, Utrecht, The Netherlands

Glossary

Complexity Theory A diverse set of theories and notions for understanding the development and functioning of complex systems. More recent versions contend that complex systems are defined more by the relationships between constituent parts than by those parts themselves.

Epistemology Theories about the nature of knowledge and what counts as (scientific) knowledge.

Essentialism The tendency to attribute natural or inherent qualities to certain categories and states. Essentialist thinking has been discredited in critical geography because it fails to appreciate the contingent nature of the associations between states and qualities.

Ontology While the term has multiple meanings in philosophy, geographers generally use it with reference to theories and beliefs of what exists in the world.

Positionality The notion central to (feminist) post-structuralist epistemology that researchers always see, interpret, and understand the world from specific embodied locations, which thus necessarily mediate in the knowledge they produce.

Reflexivity A strategy of self-critical introspection that researchers can adopt to better understand the ways their own positionality mediates the research process and the outcomes produced. This strategy has been advanced in particular by feminist geographers.

A *Contradictio* in *Terminis*?

Every new theoretical tradition within human geography since the late 1960s has heavily criticized spatial analysis or quantitative geography. Over the years, numerous radical, humanist, feminist, post structuralist, and most recently nonrepresentational geographers have dismissed and denounced the use of quantitative methods (including geographic information system (GIS)), considering them inappropriate for the type of analysis they advocate. The long list of criticisms on the practices and philosophical underpinnings of spatial analysis makes it tempting to conclude that critical geography and quantitative inquiry are indeed at odds with each other and that the differences between them are insuperable.

There is, nonetheless, a long standing tradition within quantitative geography that seeks to contribute to progressive social change but does so on the basis of largely positivist understandings of scientific practice and knowledge. Though sometimes disparaged by other critical geographers, this strand of literature is indicative of a mismatch between the actual practices of quantitative geographers and the representations about spatial science circulating through certain sectors of academia. Additionally, most critiques on mainstream spatial analysis take the link between quantitative methods and positivism for granted. However, as more and more critical geographers point out, the association between quantitative methods and positivism was actively created during the heydays of the quantitative revolution in Anglophone geography in the late 1950s and 1960s. Accordingly, these (critical) geographers argue that utilizing quantitative methods is not identical to accepting the philosophical position of positivism, and that adherents of post positivist geographies should be careful not to throw away the baby with the bathwater. Quantitative methods are thus considered one set of methods on the repertoire of the critical geographer that can be used to address at least a subset of the research questions in which s/he is interested. This is not to suggest, however, that reconciling quantitative methods and critical inquiry is always easy and straightforward. Nor is it true that critical spatial analysis constitutes an internally homogeneous and coherent body of work. Researchers have elaborated or are elaborating their own versions, which, as a whole, are sometimes contradictory.

This article provides an overview of critical spatial analysis. It is organized around the most familiar criticisms of mainstream spatial analysis and the ways critical geographers deploying quantitative methods have alleviated and circumvented these. To this end, the remainder of this text starts with those critiques on quantitative geography and then proceeds to discuss reactions and responses to them. It should be noted that the terms spatial analysis and quantitative geography are used interchangeably hereafter. Additionally, the term 'quantitative methods' is used to denote simple enumeration (counting), statistical analysis, mathematical modeling, and GIS. Despite recent contributions to the literature that challenge the usual alignment of GIS with more traditional tools for quantitative analysis, these methods are considered together here, because critical spatial scientists try to overcome the quantitative/qualitative dualism and because some of the most exciting developments within critical spatial analysis pertain to GIS, particularly in the USA. Finally, the term 'critical geography' is used to denote those currents in human geography that challenge different forms of oppression and exploitation, promote tolerance and difference, and actively bridge the gaps between scientific research and progressive social movements.

Criticizing Spatial Analysis and Beyond

Criticisms leveled at quantitative geography provide a good starting point to appreciate what critical spatial analysis is about. It is beyond this article to review all criticisms (partly because some are not relevant anymore and others have never been very accurate); the focus will be on those issues that critical analysts have sought to address. A first major group of issues are ‘political’; many geographers have rejected quantitative methods – and continue to do so – because of the links between spatial analysis and forms of social conservatism, such as neoliberalism, imperialism, and patriarchy. Though originally put forward by radical geographers in the 1970s, the argument that quantitative approaches perpetuate the social *status quo* continues to be used today. Even though quantitative methods are used widely in applied and policy oriented research with the best intentions of effectuating progressive social change, many argue that statistics and GIS are used widely to support non progressive government and corporate interests.

The second major group of criticisms of mainstream spatial science pertains to the ‘positivist worldview’ (ontology and epistemology) with which quantitative geography became associated in the 1950s and 1960s. Perhaps positivist understandings of objectivity have been criticized most extensively. While radical geographers in the 1970s and 1980s already challenged the alleged value neutrality of scientific research, feminist geographers in particular have expanded these notions and also deplored the detached and disinterested observer stance and associated view from ‘nowhere’ (a neutral, transcendental location). These epistemological premises have been denounced because they imply that the positionality of the researcher is hidden and his/her connectedness to the people or phenomena under study lost. Feminists have also questioned the hierarchical and asymmetrical power relations with an active and controlling observer, on the one hand, and a passive subject, on the other, that often characterize positivist research processes.

Another common criticism of mainstream spatial analysis concerns its tendency to concentrate (only) on similarities, regularities, and universalities, to be grasped through transcendence of and abstraction from concrete contexts. Therefore, the resulting knowledge tends to lack attention for difference and heterogeneity, and can be characterized as de contextualized and disembodied. The focus on similarities cannot be seen as separate from most spatial analysts’ commitment to reductionism – the assumption that the world is an aggregate of distinct objects and phenomena behaving in regular ways. This reductionism is convenient insofar as it allows analysts to grasp the whole through studying the actions and motivations of individual decision makers and to resort to analytically tractable models. It should be noted, however, that some spatial analysts have addressed both similarities and heterogeneity through the development of context dependent or local statistics. Examples include the expansion method, geographically weighted regression, and multilevel regression modeling.

Also aligned to the focus on similarity is the positivist belief in the prevalence of order and stability. Much spatial science has moved away from its early focus on geometry and ordered patterns to focus more extensively on spatial processes and context dependence; however, this work generally continues to be premised on a structured ontology based on fixity and categorization. While practices of fixing and classifying are inherent to almost all research, spatial scientists often unreflectively adopt rigid, dualistic, and/or essentialist taxonomies of the events and phenomena under scrutiny. Consequently, they tend to use taken for granted, hard and fast theoretical categories of, for instance, gender and race/ethnicity. These categories are not only unable to capture the complexity and richness of everyday practices and experiences; they also conceal the relational processes through which they have been constituted. That quantitative geographers tend to start from taken for granted understandings is also evident in dominant representations of space time. Space and time are often seen as containers, as neutral and external Cartesian grids, in which structures, behaviors, and changes evolve and take place. The notion that space time is produced by and internal to those structures and behaviors in often highly uneven ways remains counterintuitive to many spatial scientists.

Critical geographers using quantitative methods subscribe to many or all of these criticisms. They have adopted various strategies to accommodate or circumvent these issues, which can be grouped into two main categories: those centering on research questions and those evolving around the separation of method and philosophical basis. Where the former are concerned with the articulation of marginalized views and perspectives, the latter interrogate the possibilities of using quantitative methods given a realist, feminist, post structuralist understanding of scientific practice and knowledge. As argued before, the latter strategies follow from the recognition of the historically contingent connections between positivism and social conservatism with spatial analysis. Most of these efforts argue that quantitative methods can be reappropriated for critical purposes and used to destabilize hegemonic representations of reality in the language of dominant discourses, which often draw support from numerical information. Critiques of those discourses using the same frame of reference cannot be dismissed as irrelevant easily, it is hoped.

More Progressive Research Questions

A first type of studies discussed here consist of quanti tative investigations of issues of social distribution that question the current state of affairs in subtle ways. The subtlety lies in the adoption of largely conventional notions of epistemology and ontology that also dominate mainstream spatial analysis. These studies pragmatically draw on mainstream methodology understandings – for instance, with respect to notions of objectivity – to enhance the visibility of underexposed social groups or interests in dominant intellectual and policy discourses. While examples can also be found in the radical geog raphies of the 1970s and early 1980s, this version of critical spatial analysis is perhaps best exemplified through the body of work known as feminist empiricism.

Feminist empiricism constitutes one of the oldest forms of inquiry in feminist geography; though still important, its peak arguably fell in the 1980s and early 1990s. Consistent with traditional epistemology, it seeks to produce more accurate representations of the world by showing that disregarding women and gender issues biases research outcomes. Early work aimed to articulate the spatial dimensions of women's inequality. For in stance, Hägerstrand's constraint oriented time geography provided a useful conceptual framework for largely quantitative assessments of the implications of women's prime responsibility for domestic and caring responsi bilities for everyday activities and commuting patterns. Other studies have shown how gender, income, and race/ethnicity interact in shaping women's access to employ ment and services and their everyday life in general.

Work along such lines continues to be very helpful in documenting women's lives and in generating the numbers that inform the formulation of public policies attempting to mitigate inequalities. While feminist empiricist research may not be so prevalent in geography as it used to be in the 1980s, similar work is now undertaken more frequently in neighboring disciplines, such as urban and transportation studies. This research is also informed by conventional epistemologies but does not focus on gender issues alone; it also concentrates on race/ethnicity, age, and other axes of social differentiation.

Challenging Positivist Worldviews

During the 1980s and 1990s, methods became more and more essentialized, which intensified the association between qualitative research and critical geography and resulted in more suspicion toward using quantitative analysis for critical purposes. This made critical geog raphers with more favorable opinions about quantitative methods rethink the usefulness of quantitative analysis for critical geography and identify various contributions.

First, it has been argued that quantitative methods can be employed to give 'partial' answers to 'some' of the research questions of interest to critical geographers. Some feminist geographers, for example, have contended that especially questions focused on the broad contours of dif ference on the basis of such axes of social differentiation as gender, class, or race/ethnicity can be addressed effectively with quantitative methods. While there is some overlap with the research in the previous section, these studies distance themselves more explicitly from positivist epistemologies, instead of exploring the possibilities of combining quanti tative methods with notions from critical realism and post structuralism. Feminist geographers have been attracted in particular to the notions by Donna Haraway about situated knowledges. In brief, her argument holds that all knowledge is socially constructed and that the researcher's embedd edness in power relations always mediates the ways s/he studies phenomena and collects and interprets evidence, implying that knowledge is necessarily partial and political. She contends, therefore, that only acknowledging this par tiality and exploring its implications promises a useful version of objectivity and a view from somewhere. Har away's work has allowed critical geographers to reclaim GIS and other quantitative methods in ways that challenge traditional yardsticks of scientific rigor like generalizability and detached objectivity. Reflexivity is often employed as an analytical strategy to arrive at quantitative narratives in which researchers try to identify and explain how their own positionality, feelings, beliefs, and understandings of phenomena have mediated the research process. That is, reflexivity is used to produce quantitative geographies from somewhere rather than from nowhere.

Critical geographers who favor statistics and math ematics but may not identify themselves so directly with feminist geography have also suggested various epistemological strategies. One example is the use of Bayesian statistical theory, which allows researchers to incorporate their prior beliefs into *ex post* evaluations of modeling outcomes. Another is (very) cautious deploy ment of inferential statistics; critical geographers tend to be far more reluctant to generalize conclusions from analyzed data to larger populations than most spatial analysts. Still another strategy is to focus statistical and mathematical research on difference rather than on similarities using the advanced modeling approaches mentioned above.

Second, critical geographers have argued extensively that quantitative inquiry might be used to contextualize more in depth case study research with qualitative meth ods. More specifically, the outcomes of enumeration, statistical, or mathematical analysis can be used to sub stantiate the selection of particular case studies or to support statements about the transferability of the out comes of in depth case study research to other geo graphical contexts. This implies that quantitative analysis becomes integrated into mixed method research in which different methods are employed to answer 'different' research questions. This way of utilizing quantitative analysis also constitutes a departure from traditional notions of epistemology; it is actually better compatible with the (philosophical) pragmatism by Richard Rorty and others than with positivism. At any rate, this way of mixing methods is often defended on the grounds that it allows researchers to capitalize on the strengths of different methods. This mode of combining methods in a single research project has become very popular in human geography in recent years.

Third, quantitative methods can also be accorded a larger role in mixed method research than contextual izing other modes of inquiry. Some critical geographers have argued that outcomes of quantitative methods can be triangulated against those obtained with qualita tive methods. This implies that different methods are employed to answer ('more or less') 'the same' research question(s). Method triangulation is usually adopted to detect consistencies and corroborations across methods; the production of similarities is then considered a form of validating findings and data. However, though much less common in the literature, it is also possible to use triangulation to explore inconsistencies and the silences between data produced by different methodologies. In this way the situated and partial nature of different forms of scientific knowledge can be explored; this form of triangulation thus draws directly on the writings of Donna Haraway. A rare but very good example of this approach is the work by Andrea Nightingale who trian gulated aerial photo interpretation and in depth inter views in a study of forest change in Nepal to show the insufficiency of hegemonic representations produced by positivist, statistical research that dominate policy discourses.

Fourth, critical researchers have been keen to address the often hierarchical and asymmetrical power relations in research processes, especially with regard to GIS in the USA. Participatory GIS is a process oriented approach to GIS for altering the relations between researcher and the researched so as to empower and enhance the participation of the latter in planning pro cesses and decision making. One strategy to achieve this is the joint development of GIS applications with mar ginalized groups and/or resource poor organizations that enable the inclusion of their views, perspectives, and needs. In this way, a view from somewhere is produced in line with the work of Haraway and others. Of course, similar forms of participatory action research in which critical academic researchers collaborate closely with grassroots movements and resource poor organizations also exist for other analysis techniques, but these have received less attention in the academic literature in recent years than participatory GIS.

Fifth, some of the ontology related criticisms of positivist spatial science have been addressed using mathematical models and GIS, although the number of studies engaging with such issues is rather limited to date. In the modeling tradition it is increasingly argued that models drawing on complexity theory can to some extent bridge the differences between mainstream spatial science and social and cultural geographies. This is pri marily because complexity theory is holistic rather than reductionist due to the central role it accords to the concept of emergence, which implies that system effects are more than the sum of the effects of components and also stem from interactions between constituent parts.

For example, work by Paul Plummer and Eric Shep pard suggests that mathematical models informed by complexity theory can mediate between mainstream spatial science and critical geography. These authors have specified models for regional competitiveness in capitalist economies that seriously question positivist approaches prioritizing stability and order. Their model allows for the ontological assumption that structure and agency are mutually constitutive. This enables them to trade the assumptions of rational choice theories for those favored by critical economic geographers. Not only may knowledge be imperfect (also accommodated in some rational choice theories), preferences can also be shaped by the social and cultural contexts in which agents are situated and actors can engage in collective action. Moreover, their specified model also enables alternative representations of space that are no longer Newtonian (space as a container) but reflect endogenous and uneven spatial interdependencies. It goes without saying that models of this kind produce outcomes that clearly differ from those cast in the mould of mainstream spatial science. However, the additional level of com plexity comes at the price of substantial difficulty when it comes to model estimation. The development of dedi cated computational approaches – not only algorithms but also guidelines for how to deploy these in ways consistent with critical methodologies – is still in its in fancy, at least in geography.

Critical GIS is another burgeoning field where critical researchers are broadening the ontological and repre sentational spectra of spatial science. Recent, feminist inspired work has sought various ways to include con textual complexity into knowledges and representations produced with GIS. These include, but are not limited to, the incorporation of difficult to quantify phenomena, such as place bound emotions and webs of economic practices. For instance, in her work on place specific feelings of safety and anxiety among Muslim women after 9/11, Mei Po Kwan employed time–space diaries, in depth interviews, and ethnography to construct what she terms collaborative three dimensional (3 D) GIS videography from the vantage point of the research participant. Her research thus also addresses the claim that GIS – and spatial science in general – produces disembodied knowledges. However, as before, the levels of intensity and complexity of such GIS approaches pose significant technical challenges. While technological developments are rapidly expanding the possibilities of 'qualitative' GIS, the scaling up of produced represen tations to the levels of aggregation needed for convincing interventions in policy and decision making processes remains an important concern for the near future.

Finally, the above endeavors also suggest that critical quantitative geographers attempt to add nuances to the hard and fast theoretical categories common to main stream spatial science. This remains a difficult challenge, however, because making events and phenomena dis cernible and commensurable is required for quantification to become possible. The use of fuzzy logic, which accepts that category membership is gradual (expressed through a [0,1] range) rather than a binary condition, may mitigate some of the problems identified but still assumes that different ontological states exist. Other strategies that have been outlined and employed to circumvent rigid classifications are less technically intensive. These include reflexively accounting for the adoption of certain cat egorizations; sensitivity analyses to assess how outcomes are affected by different categorizations; and the use of segmentation and interacting variables in statistical analysis and mathematical modeling. Work by Geraldine Pratt in the early 1980s already showed that segmentation and interacting variables help researchers avoid overly universalizing claims. She used a segmentation approach to statistical analysis to interrogate Marxian statements about the links between housing tenure and political ideology, which showed that this connection was mediated by occupational status and unionization.

Conclusion

Critical spatial science is a variegated and heterogeneous body of work that takes issue with the criticisms leveled at positivist analysis and that develops novel ways to connect critical/qualitative and analytical/quantitative geog raphies. Though earlier versions in particular started from positivist worldviews, most recent work tends to be premised on the notions that using a set of techniques does not necessarily entail a specific philosophical position. This implies that not only progressive research questions are addressed, but also that alternative under standings of scientific objectivity and ontology are explored with quantitative methods. Despite its internal heterogeneity, recent critical spatial analysis also exhibits various commonalities. These include the recognition that quantitative methods may not be appropriate for all questions interesting critical geographers; a focus on difference, contextuality, and contingency instead of, or in addition to, regularity and similarity; and a process orientation to doing research. Moreover, much research transcends the qualitative/quantitative binary, primarily by mixing different research methods and by outlining simi larities in the practices associated with different methods.

Despite its vibrancy, the field of critical spatial science also faces various challenges. For instance, one important trend is the use of ever more complex tools and appli cations demanding high levels of expertise from the user. While this tendency is very welcome, it is important that efforts to advance new tools are not only restricted to technical aspects. Equal attention needs to be devoted to the development of guidelines for using those applications as a researcher and for communicating research to others in and especially beyond academia in ways compatible with radical, feminist, and post structuralist understandings of science. Additionally, given that utilizing such tools and applications is often time and resource consuming, issues of unequal access to these developments are likely to remain relevant. Notwithstanding these and other challenges, it is essential that the initiatives to integrate critical theory and spatial analysis are continued and expanded. This is because quantitative information pervades all aspects of contemporary societies and appears to become even more important, implying that the significance of numbers as tools of empowerment is also likely to increase. The attempts described in this article allow for constructive critiques of dominant scientific practices and represen tations of the social world that leave the voices and per spectives of certain groups unarticulated and thus form a necessary ingredient of a critical discipline.

See also: Chaos and Complexity; Complexity Theory, Nonlinear Dynamic Spatial Systems; Critical GIS; Cultural Turn; Economic Geography, Quantitative; Feminism/Feminist Geography; Feminist Methodologies; Mixed and Multiple Methods; Positivism/Positivist Geography; Poststructuralism/Poststructuralist Geographies; Pragmatism/Pragmatist Geographies; Quantitative Revolution; Scientific Method; Triangulation; Visualization, Feminist.

Further Reading

Barnes, T. J. and Hannah, M. (2001). Theme issue: The place of numbers: Histories, geographies, and theories of quantification, part 2. *Environment and Planning D: Society and Space* 19, 505 572.

Elwood, S. (2006). Critical issues in participatory GIS: Deconstructions, reconstructions and new research directions. *Transactions in GIS* 10, 693 708.
Fotheringham, A. S. (2000). Context dependent spatial analysis: A role for GIS? *Journal of Geographical Systems* 2, 71 76.
Hanson, S. and Pratt, G. (1995). *Gender, Work, and Space*. London: Routledge.
Haraway, D. J. (1991). *Simians, Cyborgs, and Women: The Reinvention of Nature*. New York: Routledge.
Jones, J. P. III. and Hanham, R. Q. (1995). Contingency, realism, and the expansion method. *Geographical Analysis* 27, 185 207.
Kwan, M. P. (2002). Quantitative methods and feminist geographic research. In Moss, P. (ed.) *Feminist Geography in Practice: Research and Methods*, pp 160 173. Oxford: Blackwell Publishers.
Kwan, M. P. (ed.) (2002). Special Issue: Feminist Geography and GIS. *Gender, Place and Culture* 9, 261 303.
Kwan, M. P. and Knigge, L. (eds.) (2006). Theme issue: Qualitative research and GIS. *Environment and Planning A* 38, 1999 2074.
Mattingly, D. J. and Falconer Al Hindi, K. (eds.) (1995). *Focus: Should women count? Professional Geographer* 47, 426 466.
Nightingale, A. (2003). A feminist in the forest: Situated knowledges and mixing methods in natural resource management. *ACME: An International E Journal for Critical Geographies* 2, 77 90.
Philo, C. (ed.) (1998). Theme issue: Reconsidering quantitative geography. *Environment and Planning A* 30, 191 332.
Plummer, P. and Sheppard, E. (2006). Geography matters: Agency, structures and dynamics at the intersection of economics and geography. *Journal of Economic Geography* 6, 619 637.
Pratt, G. (1989). Quantitative techniques and humanistic historical materialist perspectives. In Kobayashi, A. & Mackenzie, S. (eds.) *Remaking Human Geography*, pp 101 115. Boston, MA: Unwin Hyman.

Spatial Autocorrelation

D. A. Griffith, University of Texas at Dallas, Richardson, TX, USA

Glossary

Auto- A prefix literally meaning self; spatial autocorrelation means self-correlation, or values within a given variable are correlated, resulting in the variable being correlated with itself.

Correlation A description of the nature and degree of a relationship between a pair of quantitative variables.

Geary Ratio An index of spatial autocorrelation, involving the computation of squared differences of values that are geographic neighbors (i.e., paired comparisons), that ranges from 0 to 1 for negative, and 1 to approximately 2 for positive, spatial autocorrelation, with an expected value of 1 for zero spatial autocorrelation.

Geographic Connectivity/Weights Matrix An n-by-n matrix with the same sequence of row and column location labels, whose entries indicate which pairs of locations are neighbors.

Map Pattern The systematic organization of values for some variable across a map resulting in visually conspicuous texture that consists of global, regional, and local trends, gradients, swaths, or mosaics.

Moran Coefficient An index of spatial autocorrelation, involving the computation of cross-products of mean-adjusted values that are geographic neighbors (i.e., covariations), that ranges from roughly (–1, –0.5) to nearly 0 for negative, and nearly 0 to approximately 1 for positive, spatial autocorrelation, with an expected value of $-1/(n-1)$ for zero spatial autocorrelation, where n denotes the number of areal units.

Moran Scatterplot A scatterplot of standardized versus summed nearby standardized values whose associated bivariate regression slope coefficient is the unstandardized Moran coefficient.

Negative Spatial Autocorrelation For the geographic distribution of some variable across a map, high values tend to be geographic neighbors of low values, intermediate values tend to be geographic neighbors of intermediate values, and low values tend to be geographic neighbors of high values.

Positive Spatial Autocorrelation For the geographic distribution of some variable across a map, high values tend to be geographic neighbors of high values, intermediate values tend to be geographic neighbors of intermediate values, and low values tend to be geographic neighbors of low values.

Redundant Information Information in data that is duplicated, and hence unneeded for spatial statistical analyses; for georeferenced data, this duplication arises from locational closeness that results in mutually shared information, allowing an awareness of nearby values once the value for a given location is known.

Introduction

'Spatial autocorrelation' is the correlation among values of a single variable strictly attributable to their relatively close locational positions on a two dimensional (2 D) surface, introducing a deviation from the independent observations assumption of classical statistics. Spatial autocorrelation exists because real world phenomena are typified by orderliness, (map) pattern, and systematic concentration, rather than randomness. Tobler's first law of geography encapsulates this situation: "everything is related to everything else, but near things are more related than distant things." To this maximum should be added the qualifier: "but not necessarily through the same mechanisms." In other words, spatial autocorrelation means a dependency exists between values of a variable in neighboring or proximal locations, or a systematic pattern in values of a variable across the locations on a map due to underlying common factors.

Selected physical models portray the existence of spatial autocorrelation. First, and foremost, pictures (analogous to map pattern) would not be discernible on a television/computer screen without spatial autocorrelation. Nor could sliding tile or jigsaw puzzles be solved. Magnetic sculptures constructed from separate pieces of metal piled upon a magnetized base more completely illustrate spatial autocorrelation: when the pieces of metal are placed on their magnetic base, they can be sculpted into a 3 D figure; when the pieces of metal are moved far from their magnetic base (as well as any other magnetic source), they simply are independent pieces of metal that only can be brushed into a pile. The spatial autocorrelation mechanism here is the magnetic field that is transferred from piece to piece by metal pieces touching. Meanwhile, a number of real world examples also portray the existence of spatial autocorrelation. Most mineral deposits cluster at relatively few locations on the Earth's surface; they are not ubiquitous. House prices and house value assessments are established in the real estate market by comparisons between a house and similar nearby houses. Zoning ordinances force similar land use types to group together in coterminous locations. Diseases, such as West Nile virus, creep across a landscape through contagion. Finally, unabated water (wind) pol luters generate negative consequences for those down stream (downwind) of their locations; a similar but more fully 2 D example is furnished by defunct smelter superfund sites.

Conceptual Meanings of Spatial Autocorrelation

Spatial autocorrelation has many interpretations. The most dismissive is as a 'nuisance parameter'. Spatial autocorrelation is captured by a model specification because its presence is necessary for a good description, but it is not truly of interest and only interferes with the estimation of other model parameters that are of true interest. This interference tends to be for parameters such as variances, rather than means. Scientists increas ingly are deciding that spatial autocorrelation should not be treated as a nuisance parameter.

Interpreting spatial autocorrelation as 'self correlation' is literal. Correlation arises from the geographic context within which attribute values occur, and as such can be expressed in terms of the Pearson product–moment cor relation coefficient (r) formula, but with neighboring values of variable y replacing those of variable x. Here, the correlation being ascertained is an average of that between location specific time series of a variable for all possible pairs of locations. But these time series are not observable, and hence the assumption invoked is exchangeability (i.e., the set of time series can be permuted without affecting results – the order in which a time series mechanism generates values across a map is irrelevant).

Interpreting spatial autocorrelation as 'map pattern' emphasizes conspicuous trends, gradients, swaths, or mosaics across a map. Consider a constant, which is the degenerate case (i.e., a constant has no variance) of perfect positive spatial autocorrelation: once the value of a constant is known at a single location, it is known at all locations. Next, consider a variable that portrays a north–south (or east–west) linear trend across a map. If this variable has a mean of 0, then it is geometrically independent of any constant. These north–south and east–west oriented linear trend variables also are in dependent. A variable with mean 0 whose values' mag nitudes form a 3 D symmetric hill (or valley) in the center of a map constitutes yet another mutually in dependent map pattern. These three variables display maximum levels of positive spatial autocorrelation when geographic variance is present, and may be described as global geographic patterns. Alternating sequences of moderately large hills and valleys with either an east–west or a north–south orientation portray moderate positive spatial autocorrelation, and constitute regional map patterns. Alternating sequences of small hills and valleys with either an east–west or a north–south orien tation portray weak positive spatial autocorrelation, and constitute local map patterns. This fragmentation con tinues through randomness (zero spatial autocorrelation) to arrangements of increasingly alternating values (i.e., single value hills and valleys), which portray increasing negative spatial autocorrelation. Most substantive vari ables have geographic distributions that can be described by combinations of some subset of these mutually in dependent varying hill–valley cluster size map patterns.

As a 'diagnostic tool', spatial autocorrelation plays a crucial role in model based inference, whose foundation is a set of valid assumptions rather than a scientific (i.e., random) sampling design. Detected spatial auto correlation can signify model misspecification, including treating nonlinear relationships with a linear specifi cation, such as an exponential relationship between a predictor variable, x, and a response variable, y, that is described with a straight trend line. It also can signify 'missing variables' for a regression equation, and as such serves as a 'surrogate' for variation otherwise un accounted for because these variables are missing. This surrogate role occurs when autocorrelation map patterns displayed by predictor variables align with autocorrela tion map patterns displayed by y. Moreover, the same set of distinct map patterns is common to both a set of missing variables and y. Accounting for spatial auto correlation in this context can do a surprisingly good job of representing missing variables in an equation specifi cation. In contrast, spatial autocorrelation that remains unaccounted for tends to distort classical correlation coefficient interpretations.

The term correlation alludes to the notion of 'redundant information'. If x and y are perfectly correl ated, then knowing x means exactly knowing y. In other words, the information content of y is perfectly dupli cated in x; this degree of duplication decreases as the correlation coefficient moves toward 0. Spatial auto correlation extends this notion of redundant information to georeferenced data. The value at a given location can be predicted with some degree of accuracy from the values at nearby locations; this spatial data feature con stitutes the foundation of cartographic interpolation. Recalling the example of housing prices, such redundant information results from 'spatial spillovers': house values at one location spill over to impact upon house values at nearby locations. Building an expensive house near an inexpensive house tends to reduce the value of the expensive house while increasing the value of the inexpensive house, all other things being equal. This mechanism is similar to a river that floods its banks into its flood plain. Another example, at the micro level of geographic resolution, is second hand smoke generated by smokers. And yet another example is the noxious smell generated by such facilities as sewage treatment or rendering plants that permeates their surrounding neighborhoods.

Spatial autocorrelation can materialize from some course of action operating over a geographic landscape, such as contagious diffusion of a disease, resulting in it being interpreted as a 'spatial process mechanism'. The diffusion of West Nile virus across the coterminous United States (US) illustrates this interpretation. This disease quickly became a serious problem in the US, extremely rapidly diffusing from Long Island in 1999 throughout the remainder of the country following northeast to west and south paths. A weather front moving across a geographic landscape can be viewed in a similar way because it results in highly spatially autocorrelated local weather conditions.

Finally, interpreting spatial autocorrelation as an 'outcome of areal unit demarcation' relates it to the modifiable areal unit problem (MAUP), whereby results from statistical analyses of georeferenced data can be varied at will simply by changing the surface partitioning used to demarcate areal units. For example, a standard eight by eight checkerboard pattern exhibits negative spatial autocorrelation between the red and black colors of its squares. But if these squares are aggregated into compact clusters of four (i.e., two by two groupings), and the red and black colors averaged (resulting in a constant dark red color across the checkerboard), then the spatial autocorrelation becomes maximally positive. Political redistricting involving gerrymandering (i.e., electoral district or constituency boundaries are manipulated in order to achieve a prespecified geographic aggregation result, such as some political advantage) exemplifies this general situation in practice.

Illustrations of Spatial Autocorrelation

Most empirical spatial autocorrelation cases involve moderate, positive relationship tendencies between nearby values on a map. Remotely sensed satellite images are one exception, almost always displaying a very strong positive relationship. Most socioeconomic/demographic data display a moderate positive relationship. And, negative spatial autocorrelation rarely is encountered in practice.

Strong Positive Spatial Autocorrelation

Remotely sensed images are one exception to the georeferenced data norm of moderate positive spatial autocorrelation, frequently displaying very strong positive spatial autocorrelation. This feature partly results from light reflectance scattering, rather than being neatly contained in pixel boundaries, which are imaginary, hence spilling over into nearby pixels measured by a satellite's sensors.

A 1000 by 1000 pixel subset was extracted from a satellite image of the Florida Everglades for illustrative purposes (**Figure 1a**). The amount of green vegetation in this region can be quantified with a normalized difference vegetation index (NDVI), which is calculated with the spectral reflectance values for the near infrared (B5) and visible red (B3) bands as follows:

$$\frac{\mathrm{B5} \quad \mathrm{B3}}{\mathrm{B5} + \mathrm{B3}}$$

Positive values beyond about 0.3 tend to represent green vegetation, whereas negative values tend to represent swamp areas; positive values closer to 0 tend to represent soil. The 1 000 000 selected pixels yield only 3033 different NDVI values. This measurement was modified mathematically (i.e., transformed) in order to better align it with a bell shaped curve (**Figure 1b**).

The image appearing in **Figure 1a** displays very strong positive spatial autocorrelation. The outline of the Everglades is apparent, as is the southeast coast of Florida, south of Miami. The absence of autocorrelation would result in this image appearing as a shuffling of the set of colors, similar to the picture distorting momentary white specks that appear on a television screen when, for example, atmospheric static is dominant when, say, a cable connection is lost.

Moderate Positive Spatial Autocorrelation

Maps of population density tend to display moderate positive spatial autocorrelation, in part due to urbanization at a regional or national scale, and zoning at a local scale. Data from the 1993 census of Peru furnish population counts by districts across the Cusco department; an ArcGIS shapefile furnishes area measures for these 108 districts. Population density tends to be skewed, with a natural lower bound of 0, and few areal units with relatively sizeable concentrations.

Population density by district in the Cusco department ranges from 0.8 to 11 579.2 per unit area. Its inverse square root, after adding 11 to each density, better mimics a bell shaped curve (see **Figure 2**). This transformed population density displays moderate positive spatial autocorrelation, forming an elongated mound map pattern with a single peak. The highest density is in the city of Cusco, with the next highest densities stretching along an economic corridor formed by the Vilcanota River valley; the lowest densities are in the most rural areas of this department.

Moderate Negative Spatial Autocorrelation

Few empirical examples of negative spatial autocorrelation are reported in the literature. Usually this phenomenon is discussed conceptually in terms of geographic

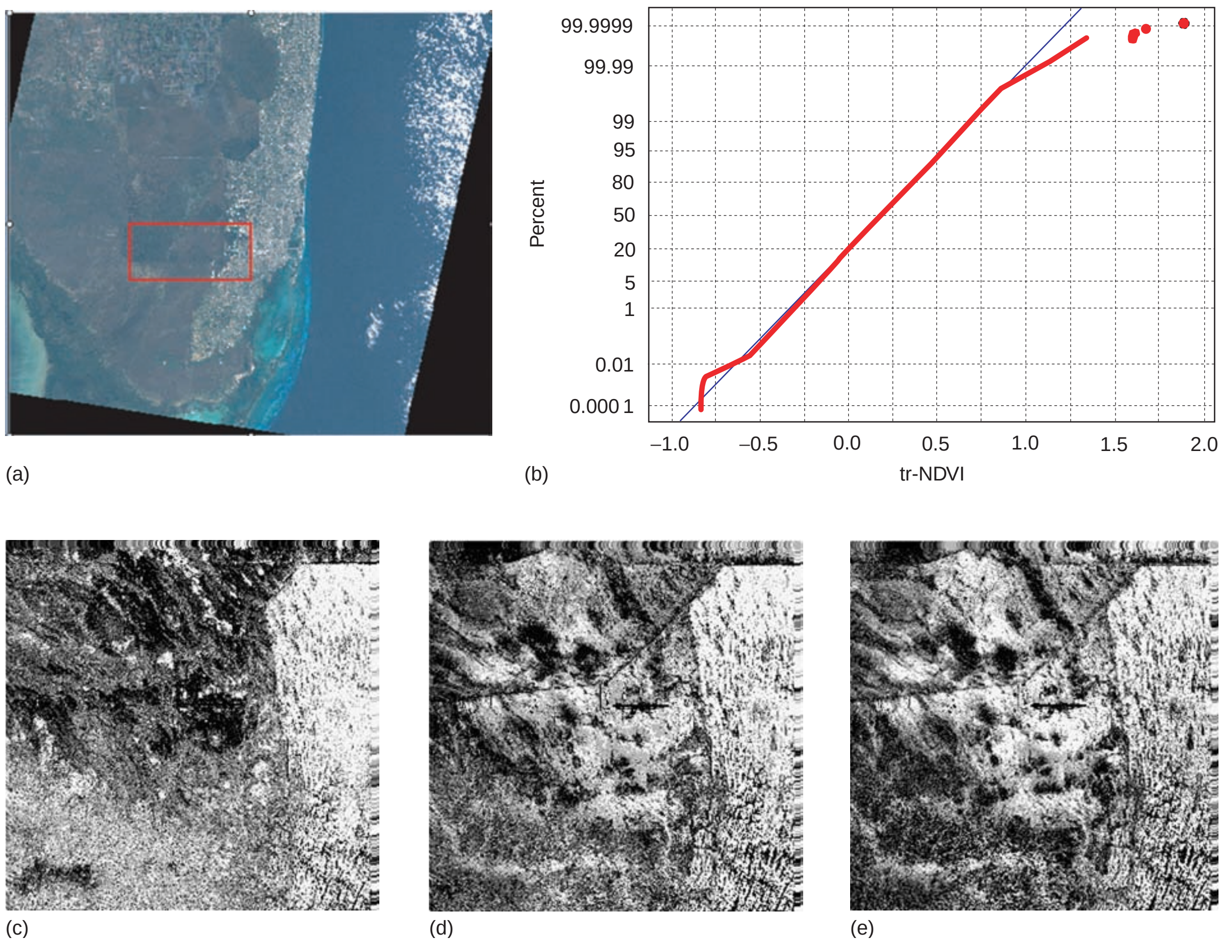

Figure 1 A Landsat 7 Enhanced Thematic Mapper Plus (ETM+) image of the Florida Everglades for 1 January 2002. (a) A composite image constructed using all spectral bands, with the subregion of analysis outlined with a red rectangle and the nonimage areas blacked out; (b) a quantile plot of the 1 000 000 transformed NDVI values for the subregion; (c) visible red spectral band (wavelength 0.63–0.69 microns (i.e., nanometers/1000)) for the subregion, with data values in the range 13–175; (d) mid-infrared spectral band (wavelength 1.55–1.75 microns) for the subregion, with data values in the range 1–173; and (e) transformed NDVI, with values in the range −0.84 to 1.89. Note: the gray scale for the maps is directly proportional to the pixel values.

competition. In other words, if a finite amount of land is available, gains in land size of one territory can occur only through the loss of land size in nearby territories. The World Wars fought on the European continent illustrate this situation.

Continental Europe principally in terms of the European Union is partitioned into 22 countries. Treating the capital of each country as its seat of power, and hence its focal point, this portion of the continent also can be partitioned into Thiessen polygons (**Figure 3b**). The ratio of each country's actual to corresponding Thiessen polygon land size gives an index of local competition, whose normal quantile plot (**Figure 3a**) implies a frequency distribution that conforms well to a bell-shaped curve. This areas ratio displays negative spatial autocorrelation: Switzerland, Luxembourg, Slovenia, and the Czech Republic are islands of very low ratio values that are completely surrounded by countries that have the highest ratio values (**Figure 3c**).

Estimators of Spatial Autocorrelation

Historically, once the spatial autocorrelation concept was established and widely recognized, spatial scientists became interested in quantifying it, and then testing hypotheses about it, with an ultimate goal of incorporating it into models. The two most commonly used quantitative indices are the Moran coefficient (MC) and the Geary ratio (GR).

From *r* to Moran Coefficient

Exploiting the interpretation of self-correlation, spatial autocorrelation can be expressed in terms of the formula for r, but with neighboring values of variable y replacing those of x

$$\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})/n}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2/n}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2/n}}$$

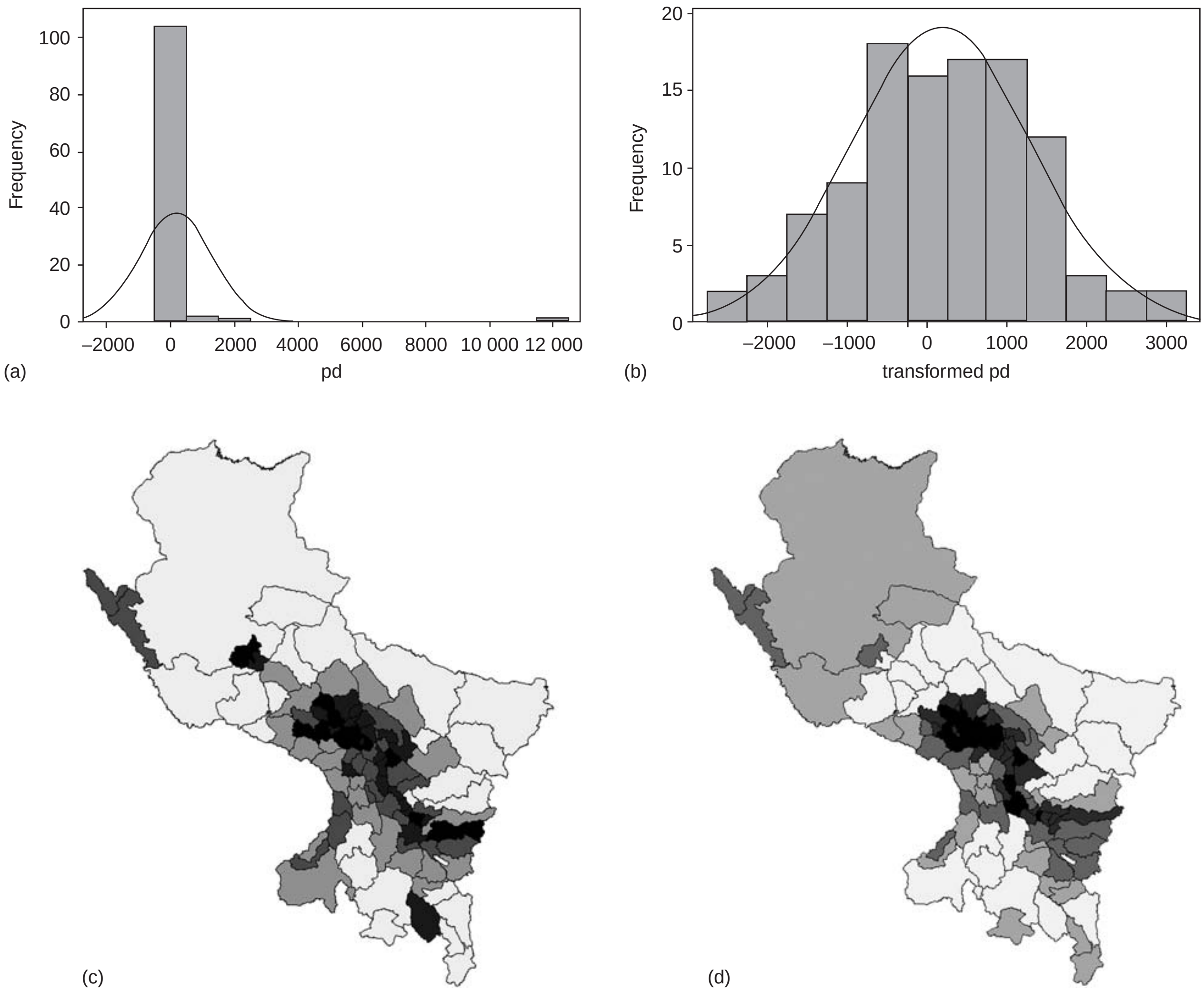

Figure 2 The 1993 population density across the department of Cusco, Peru. (a) A histogram constructed with the 108 raw population densities; (b) a histogram constructed with the 108 transformed population densities; (c) a quintile map of the geographic distribution of transformed population density values; and (d) a quintile map of the composite spatial autocorrelation map pattern latent in the geographic distribution of transformed population density values. Note: the gray scale for the maps is directly proportional to population density.

becomes

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}(y_i - \bar{y})(y_j - \bar{y}) / \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2/n}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2/n}}$$

The left hand expression converts to the right hand one by replacing y's with x's in the right hand side, by computing the numerator term only when areal units i and j are nearby (c_{ij} is an indicator variable – often called a spatial weight – whose value is 1 for neighbors, and 0 otherwise), and by averaging the numerator cross-product terms over the total number of pairs denoted as being nearby. The denominator of this revised expression is the sample variance of Y, s_Y^2. But unlike the values ± 1 for r, the extreme values of the MC are determined by sophisticated mathematical quantities called eigenvalues that are computed when the set of n indicator variables is organized into a table, much like a Microsoft Excel spreadsheet, called a matrix. These values rarely are ± 1; more often, the lower bound is between –1 and –0.5, and the upper bound is between 1 and 1.1. Nevertheless, like the relative values for r, MC values greater than $-1/(n - 1)$ indicate positive spatial autocorrelation (e.g., using the ratio MC/MC_{max}, where MC_{max} denotes the maximum possible MC value, 0.25 to 0.50 denotes a weak, 0.50 to 0.70 denotes a moderate, 0.70 to 0.90 denotes a strong, and 0.90 to 1.00 denotes a marked degree), whereas values less than $-1/(n - 1)$ indicate negative spatial autocorrelation.

The 'MC' is a covariation (i.e., pairwise products of deviations from the mean) index, and is defined by the

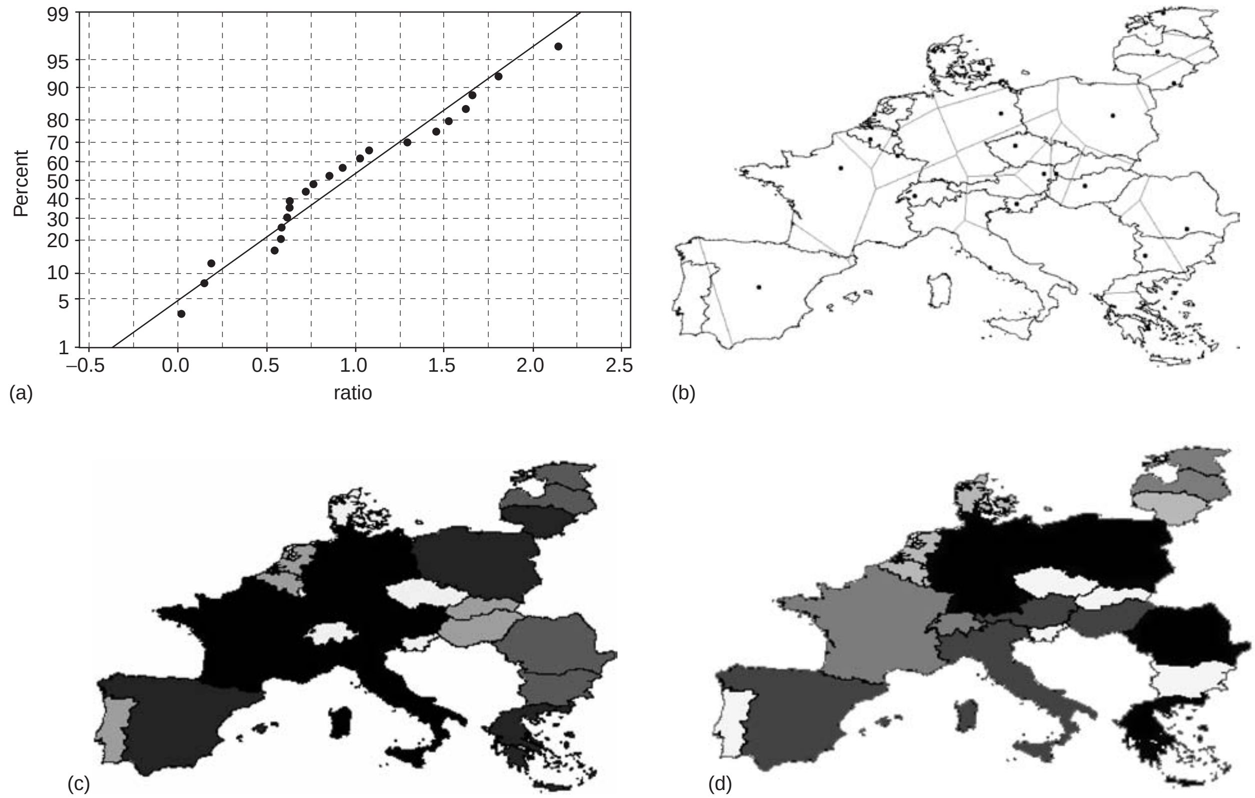

Figure 3 The geographic distribution of hypothetical land size competition across continental Europe. (a) A quantile plot of the ratio between actual and Thiessen polygon areas, by country; (b) the Thiessen polygon partitioning of the continent (gray lines), with country capitals as focal points; (c) a quintile map of the geographic distribution of the areas ratio values; and (d) a quintile map of the composite spatial autocorrelation map pattern latent in the geographic distribution of the areas ratio values. Note: the gray scale for the maps is directly proportional to the size of the values mapped.

preceding right hand expression. Its sampling distribution can be constructed in one of two ways: randomly sampling from a hypothetical probability distribution or taking the set of observed values as given and constructing all possible permutations of their values across a given surface partitioning. In the former case, present distributional results assume that a normal probability model underlies variable y. In either case, the expected value of the sampling distribution is $-1/(n-1)$, the boundary separating positive and negative spatial autocorrelation.

The standard error formulae for the MC are rather complicated, but can be well approximated by

$$\sqrt{\frac{2}{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}}}$$

when n is at least 20; the sum $\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}$ counts the number of 1's in the n indicator variables, which equals twice the number of neighbors. This standard error result is not surprising, recalling that the standard error of a sample correlation coefficient relates to n, and that all neighbors are double counted (both i to j and j to i).

The MC values for the preceding empirical examples are 0.78 for the Everglades transformed NDVI, 0.51 for the Cusco transformed population density, and –0.33 for the European areas ratio. Their respective standard deviations are: 0.058 78, 0.000 71, and 0.171 18.

The 'GR', the other popular spatial autocorrelation index, is based upon paired comparisons (i.e., pairwise squared differences), and may be defined as follows:

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}(y_i - y_j)^2 / \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}}{2\sum_{i=1}^{n}(y_i - \bar{y})^2/(n-1)}$$

This expression, which involves an average squared differences term, also involves the unbiased variance estimate for variable y. A '2' appears in the denominator because of double counting

$$(y_i - y_j)^2 = (y_i - y_j + \bar{y} - \bar{y})^2 = (y_i - \bar{y})^2 + (y_j - \bar{y})^2 - 2(y_i - \bar{y})(y_j - \bar{y})$$

The expected value of GR is 1. It ranges between roughly 0 and 2, with 0–1 signifying positive spatial autocorrelation (i.e., $(y_i - y_j)^2$ goes to 0), and with 1–2 signifying negative spatial autocorrelation (i.e., $(y_i - y_j)^2$ is increasing in magnitude). Again, the actual extreme values are a function of eigenvalues affiliated with the matrix constructed with their spatial weight indicator variables. In addition, 0 can never be reached, because it is associated with the degenerate case of a constant across a geographic landscape, and hence yields a division by 0.

The GR values for the preceding empirical examples are 0.22 for the Everglades transformed NDVI, 0.41 for the Cusco transformed population density, and 1.67 for the European areas ratio.

Relationships between the Moran Coefficient and the Geary Ratio

The MC and GR values tend to give consistent implications when y is a normal random variable. In this case, a useful rule of thumb is that the two coefficients should sum to approximately 1. Dramatic deviations from 1 by this sum suggest that y may be non normal. Meanwhile, most – but not all – empirical cases yield roughly the same MC and GR values for raw data as well as data transformed to better mimic a bell shaped curve. And, the MC is the statistically most powerful of the two indices: it does the better job, on average, of differentiating between null and alternative hypotheses. The algebraic relationship between the MC and the GR highlights this feature

$$\mathrm{GR} = \frac{n-1}{n}\left[\frac{\left[\sum_{i=1}^{n}\left(\sum_{j=1}^{n} c_{ij}\right)(y_i - \bar{y})^2\right] / \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}}{\sum_{i=1}^{n}(y_i - \bar{y})^2/n} - \mathrm{MC}\right]$$

This equation reveals that the GR incorporates locational information in addition to that included in the MC. This additional information is the ratio of squared deviations times their number of neighbors, divided by the sample variance. If outliers are present, then this numerator can become excessively large; if an areal unit has a large number of neighbors, then this numerator can be markedly influenced by the corresponding deviation. If variable y conforms to a bell shaped curve, then this additional information term is approximately 1; as n becomes increasingly large, the ratio $(n-1)/n$ converges on 1. Therefore, adding MC to both sides of this equation would result in $(\mathrm{MC}+\mathrm{GR}) \approx 1$.

The sign for the MC term in this preceding equation is negative, emphasizing the negative relationship between the MC and the GR: MC values approaching 1 correspond to GR values approaching 0 (positive spatial autocorrelation), and MC values approaching −1 correspond to GR values approaching 2 (negative spatial autocorrelation). Zero spatial autocorrelation corresponds to an expected value for the MC of $-1/(n-1)$, which asymptotically converges on 0, resulting in the corresponding GR value converging on 1, its expected value for zero spatial autocorrelation. This negative relationship can be seen simply by constructing a scatterplot of MC and GR values for the three preceding empirical examples.

Graphical Portrayals of Spatial Autocorrelation

Spatial autocorrelation is a concept that lends itself to visualization, including scatterplots and mappings. Various graphical portrayals of this concept can be constructed, including a Moran scatterplot, a semivariogram plot (based upon the GR and interareal unit distances), and a spatial correlogram (based upon MC and GR values for 1st, then 2nd, then 3rd, and so forth, nearest neighbors). The first of these, a Moran scatterplot, can be constructed by: converting georeferenced data values to z scores (i.e., subtract the mean and then divide by the standard deviation), summing the surrounding z scores for each areal unit (i.e., $\sum_{j=1}^{n} c_{ij} z_j$), and then plotting these pairs of z scores (horizontal axis) versus sums of surrounding z scores (vertical axis). The slope of the resulting trend line is proportional to the MC (i.e., it needs to be divided by $\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}$). Moran scatterplots for each of the three preceding empirical examples appear in **Figure 4**.

The Moran scatterplot portrays $\sum_{j=1}^{n} c_{ij} z_j$ versus z_i, whose trend line highlights the global trend across a given geographic landscape. These sums of neighboring values' quantities also can be visualized with a map. Doing so produces local indices of spatial autocorrelation (LISA) statistics, which enable clusterings on a map to become more conspicuous. Again, hills (i.e., clusters of surrounding values above a mean) and valleys (i.e., clusters of surrounding values below a mean) constitute the patterns of interest. LISA quantities highlight local trends across a given geographic landscape, emphasizing any clusterings in the deviations from the global trend line. These individual contributions to an MC reveal whether spatial autocorrelation essentially is the same in all parts, or differs from one part to another part, of a geographic landscape. Local indices can be constructed with a global GR, too, as well as with statistics computed from geostatistical quantities (e.g., the pair of Getis–Ord statistics).

Inspection of a map often suggests the nature and degree of the spatial autocorrelation it contains. This composite map pattern can be viewed in terms of global, regional, and local map pattern components. For example, the composite map pattern associated with the geographic distribution of population density across the Cusco department appears in **Figure 2d**, and comprises 14 individual distinct map patterns, of which 11 are global and regional (strong to marked positive spatial autocorrelation), and three are local (weak to moderate

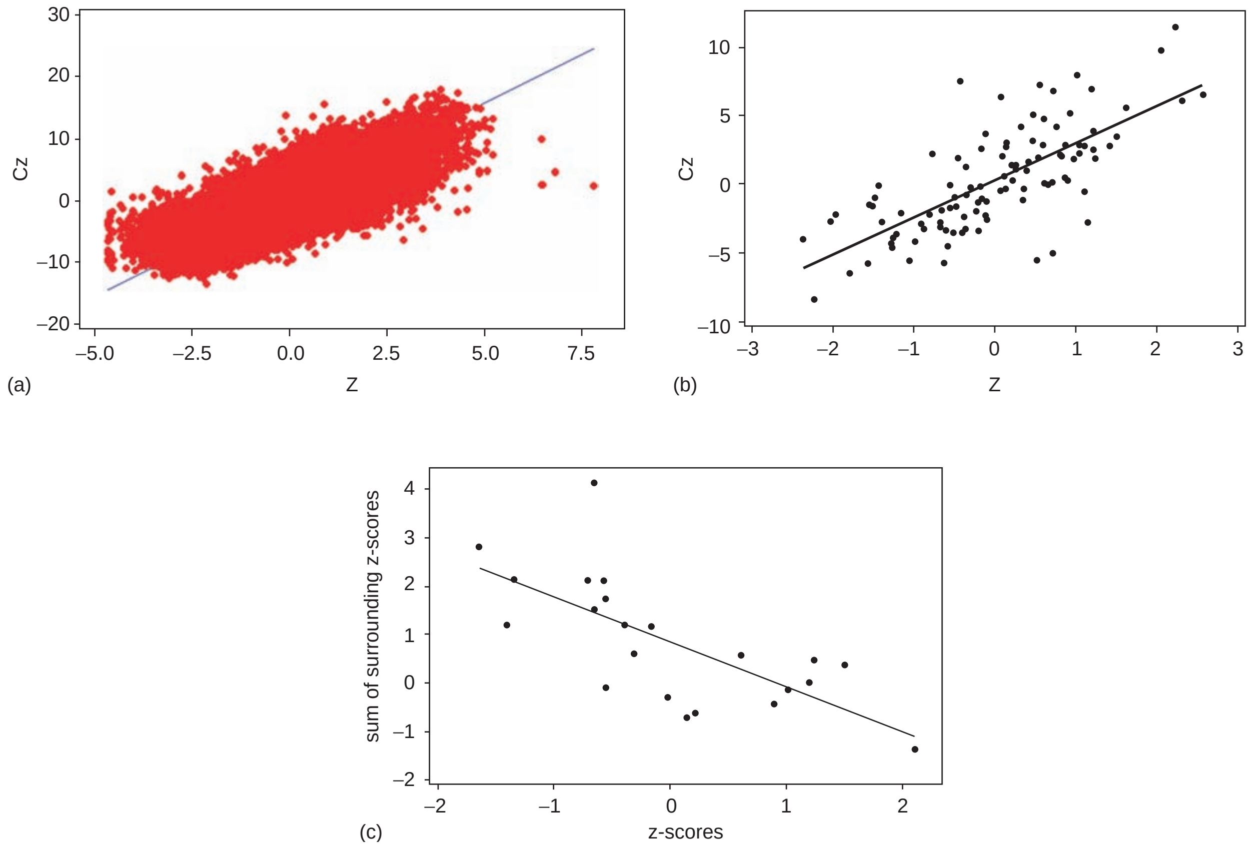

Figure 4 Moran scatterplots for the three empirical examples. (a) Everglades transformed NDVI subregion; (b) Cusco transformed population density; and (c) European areas ratio.

positive spatial autocorrelation). The most prominent among this selected set of map patterns is a northeast–southwest trend, followed by the hill shaped map pattern; the most fragmented of these map patterns (i.e., a local pattern) includes nine visually detectable clusters scattered across the department, many of which are partly merged. This composite map pattern accounts for roughly 62% of the geographic variation (i.e., redundant information), and for all but a trace amount of the positive spatial autocorrelation in population density across the department. Similar descriptions can be furnished for the Everglades NDVI and the European areas ratio maps. These local map patterns relate to LISAs. The totality of these map patterns relates to geographically weighted regression (GWR).

Theoretical Statistical Properties of Spatial Autocorrelation

Classical statistics furnishes criteria to distinguish useful from useless spatial autocorrelation measures, including unbiasedness, efficiency, sufficiency, and consistency.

The arithmetic mean of the sampling distribution for an 'unbiased estimator' equals its corresponding population parameter. This property has been used extensively to evaluate the case of zero spatial autocorrelation measures, and the impact of nonzero spatial autocorrelation on conventional sample statistics. For many, but not all, conventional statistical models, sample means, and regression coefficients tend to be unbiased, whereas sample variances and correlation coefficients tend to be biased, by nonzero spatial autocorrelation. Thus, the principal impact of spatial autocorrelation is on standard error calculations. Because a correlation coefficient is the ratio of a covariance and two standard deviations, spatial autocorrelation impacts in both of these calculations largely cancel each other when the ratio is calculated.

An 'efficiency estimator' is both unbiased and has the smallest possible standard error. Because spatial autocorrelation mostly impacts upon variance calculations, which in turn are used to compute standard errors, this statistical property is the one most affected by nonzero spatial autocorrelation. In the presence of positive spatial autocorrelation, a variance tends to be inflated. The net result is that the corresponding sampling distribution is flatter than traditional statistical theory indicates. Redundant information introduced by spatial autocorrelation results in, for example, sample sizes being misleadingly large. Consequently, more geographic sam ples are needed to acquire a given margin of error.

A 'sufficient estimator' utilizes all information con tained in a sample that is relevant to a particular par ameter. Conventional statistics applied to georeferenced data overlook locational information contained in nearby values. Recognition of this information is required for sufficiency to be preserved. For example, an arithmetic mean not only needs $\sum_{i=1}^{n} y_i$, but also the cross product term $\sum_{i=1}^{n} y_i \sum_{j=1}^{n} c_{ij} y_j$, which is used to measure spatial autocorrelation.

Finally, a 'consistent estimator's' sampling distribution concentrates at the corresponding parameter value as n increases. The efficiency criterion implies that this will occur for unbiased statistics, such as the arithmetic mean, but at a slower rate when spatial autocorrelation is present than is suggested by conventional statistics. Asymptotic concentration will occur even with variance calculations, although not necessarily at the correct value. If geographic sampling intensifies in a given region (infill sampling), then sample points increasingly become closer as n increases, resulting in spatial autocorrelation increasing, and hence concentration increasingly slow ing down. If geographic sampling involves expanding a region while maintaining the same average spacing between sample point (increasing domain sampling), then the rate of convergence will be constant rather than decreasing, and thus consistency relies only on the finiteness of all or part of the globe.

Summary and Contemporary Issues

Spatial autocorrelation has many faces, with its most common interpretations being expressed in terms of self correlation, map pattern, and redundant information. Its preferred measure is the MC, and one of its two most popular graphical portrayals is the associated Moran scatterplot. Few empirical examples of negative autocorrelation have been found, with most empirical examples involving moderate positive spatial auto correlation; remotely sensed images tend to display strong positive spatial autocorrelation. Impacts of spatial autocorrelation on conventional statistics can be assessed with standard mathematical statistics criteria, such as unbiasedness, efficiency, sufficiency, and consistency.

Today, spatial scientists routinely compute measures of spatial autocorrelation, and rather than test hypotheses about its presence, automatically include it in their model specifications. Doing so often costs only 1 degree of freedom. Spatial autocorrelation with linear models is well understood, and has yielded spatial autoregressive tools used in spatial statistics and spatial econometrics. Spatial autocorrelation with generalized linear (mixed) models is not well understood, with only a few cum bersome tools available to handle it. Spatial filtering, which is in its infancy and exploits the map pattern in terpretation, offers an approach that spans both linear and nonlinear statistical models with tools that account for the presence of spatial autocorrelation. This meth odology decomposes an underlying composite map pat tern into global, regional, and local components of spatial autocorrelation.

The frontiers of spatial autocorrelation research entail fuller development of contemporary techniques such as spatial filtering, and efficient extensions of existing techniques to massively large datasets. For example, the Everglades remotely sensed image contains 41 611 007 land coverage pixels, whereas the simple preceding analysis was challenged by dealing with only 1 000 000 of these pixels. Furthermore, a need still exists for de velopment of quality spatial autocorrelation measures, especially robust ones, for non normal data.

See also: Regression, Linear and Nonlinear; Segregation Indices; Spatial Clustering, Detection and Analysis of; Spatially Autoregressive Models; Statistics, Spatial.

Further Reading

Anselin, L. (1995). Local indicators of spatial association LISA. *Geographical Analysis* 27, 93 115.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society B* 36, 192 225.

Getis, A. and Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24, 189 206.

Griffith, D. (1987). *Spatial Autocorrelation: A Primer*. Washington, DC: Association of American Geographers Resource Publication.

Griffith, D. (1992). What is spatial autocorrelation? Reflections on the past 25 years of spatial statistics. *l'Espace Géographique* 21, 265 280.

Griffith, D. (1996). Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying geo referenced data. *The Canadian Geographer* 40, 351 367.

Mardia, K. and Marshall, R. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* 71, 135 146.

Richardson, S. and Hémon, D. (1981). On the variance of the sample correlation between two independent lattice processes. *Journal of Applied Probability* 18, 943 948.

Tiefelsdorf, M. and Boots, B. (1995). The exact distribution of Moran's I. *Environment and Planning A* 27, 985 999.

Relevant Websites

http://www.geoda.uiuc.edu
Geo Da Centre for Geospatial Analysis and Computation.

http://ncg.nuim.ie/ncg/GWR/
National Centre for Geocomputation.

http://www.ecoevol.ufg.br/sam/
Spatial Analysis in Macroecology.

http://spatialfiltering.com/
Spatial Filtering.

Spatial Clustering, Detection and Analysis of

Y. Lu, Texas State University – San Marcos, San Marcos, TX, USA

Glossary

Complete Spatial Randomness (CSR) A process in which any event has an equal probability of occurring at any location in the study area and the location of any event is independent of the location of any other event.
Correspondence of Spatial Clustering The spatial autocorrelation in paired point locations. It evaluates if a subset of points cluster when the subset of their corresponding (paired) points cluster.
Hot Spot A spatial cluster of point events or of high values. A spatial cluster of low values is called cold spot.
Multivariate Spatial Analysis Analyze the spatial correlation patterns between different variables or different types of events.
Space–Time Cluster The cluster of events through space and dependent on time.
Spatial Autocorrelation The correlation of a variable with itself through space. Positive spatial autocorrelation refers to the patterns where nearby or neighboring values are more alike; while negative spatial autocorrelation refers to the patterns where nearby or neighboring values are unlike.
Spatial Cluster Positive spatial autocorrelation.
1st-Order Analysis Global scale analysis, aiming at assessing if the pattern under investigation shows spatial autocorrelation.
2nd-Order Analysis Local scale analysis, aiming at detecting the location and spatial extent of the cluster.

Introduction

Clustering detection and analysis examines the spatial autocorrelation patterns imbedded in the distribution of geographical phenomena. A clustering pattern exists if the nearby or neighboring measurements tend to be similar. The identification of spatial clusters often facilitates the establishment of hypothesis related to the patterns of the geographical phenomenon and the relationship between the phenomenon and the underlying process. The testing for such a hypothesis often leads to better understanding of the phenomenon and the underlying process.

Cluster detection fulfills two tasks: to examine if the distribution is clustered and to reveal the location and extent of clusters. These two aspects are investigated through 1st order and the 2nd order analysis, that is, global and local scale analysis. For both types of analyses, the observed measurements are compared with the expected measurements, which are often derived from random distribution. A spatial distribution of geographical phenomena might display three different types of patterns: regular pattern, random pattern, and clustered pattern. **Figure 1** illustrates the three different patterns using point events as examples. A set of events show complete spatial randomness (CSR) if: (1) the number of events in a region follows a Poisson distribution, and (2) given n events in the region, the location of the ith event is independent of that of every other event in the distribution. In other words, CSR refers to a pattern in which any event has an equal probability of locating at any position within the study area, and the location of any event is independent of that of any other event. The detection of clusters is normally conducted through evaluating the divergence of the patterns under investigation from a random pattern. If nearby values are more similar to each other than in a random pattern, or if there are more observations within a neighborhood than would be for a random distribution, the distribution under investigation is claimed to be clustered. The neighborhood that contains more similar values or more observations defines a cluster.

Fundamental Considerations for Cluster Detection

There are three fundamental issues for cluster detection and analysis: (1) the measurement of spatial separation, (2) the scope of analysis, and (3) the searching strategy. A cluster detection method is normally developed based on a certain approach for each of these issues. Since no clustering analysis method can take care of all the three issues from all aspects, each cluster detection technique has its strength and weakness.

To assess if spatial autocorrelation exists in a set of geographical events is to examine if the nearby measurements are more similar than expected or the events are closer to each other than expected. In both cases, the measurement of spatial separation is the key. Cluster analysis techniques generally adopt two types of approach to measure spatial separation between events: continuous distance function and a grid of spatial cells (regular or irregular) draped over the study area. These two approaches are commonly called as 'distance view approach' versus 'neighborhood view approach' or 'distance method' versus 'quadrat method'. The distance

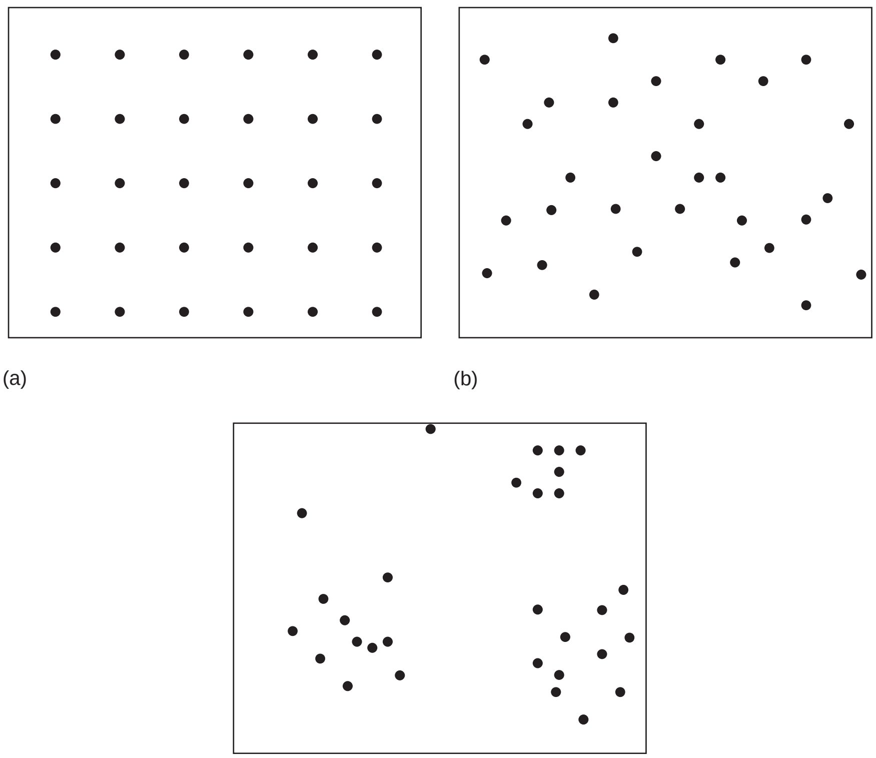

(a) (b) (c)

Figure 1 The three different types of point patterns (with 30 point events in each set): (a) a regular pattern, (b) a random pattern, and (c) a clustered pattern.

method has the flexibility of evaluating the clustering patterns on any preferred spatial scale; the cells of a quadrat method can be designed to accommodate the spatial structure of the underlying process. However, it can be a challenge for a distance method to account for the spatial variation of risk population. For example, one should not expect to observe the same number of disease cases within equally sized subregions everywhere as population distribution is uneven. Another potential problem with the distance methods is the tendency for overcomputation. The same subset of events might be evaluated for multiple, overlapping subregions; this can cause repetition of clustering evaluation. Similarly, there are challenges for the quadrat methods. First, the definition of the quadrat impacts the output of cluster detection. A point events distribution might show a clustering pattern when investigated with one set of quadrats, but no clustering with a different quadrat set. This is a typical modifiable areal unit problem (MAUP). Second, when the pattern is analyzed through a series of scales, the discrete nature of the cell size might result in significant clustering patterns that exist at a certain spatial scale not being identified.

Cluster detection techniques assess spatial clustering at different scopes. Some are based on an exclusive examination of the entire data set (global methods) aiming at assessing if clustering patterns exist in the whole distribution. Other techniques examine local variation in clustering (local methods) by evaluating the observations at a neighborhood scale to identify the locations of clusters. Put another way, global methods evaluate clustering within the entire study area, whereas local methods assess clustering within distinct subregions of the study area. Logically, a global view test should be conducted before any local view investigation. In practice, given the availability of great computing power nowadays, some local cluster detection methods are directly applied to the whole data set provided that the whole study area is examined thoroughly. Examples of such include methods based on moving window approaches or a spatial scanning process. **Table 1** provides a quick summary of some of the commonly used clustering analysis methods.

The third aspect that differentiates cluster detection methods is related to the searching strategy for clustering detection. There are general tests and focused tests. General tests conduct an exclusive search through the whole study area to identify all the clusters in the area. Focused tests evaluate the existence of clusters around certain, predefined locations. For example, a focused test can be conducted to see if there are more drunk and drive cases near taverns. When using a focused search

Table 1 Most frequently used techniques for spatial cluster analysis

		Scale of observation	
		Global view	*Local view*
Perception of separation	Continuous distance	Ripley's *K* function, Nearest neighbor index	Getis' and Ord's G*, Kernel density
	Discrete zoning	Moran's *I*, Geary's *C*	Anselin's LISA, Local Moran, Kulldorff SaTScan

strategy, a focus location is first identified and then the events surrounding the location are evaluated for clustering. When there is no obvious hypothesis about the spatial relationship between a facility and the events, a general cluster analysis of the whole event set is usually conducted. The results of cluster analysis, when plotted together with the distribution of facilities, might aid in establishing hypotheses about the relationship between the events under investigation and the locations of facilities. John Snow's mapping and analysis of the locations of cholera cases and water pumps in 1854 in Broad Street, London, is a well known example.

It is important to note that, although frequently seen for point data analysis, spatial cluster detection and analysis is very valuable for areal and linear/flow data as well. Examples include investigating the spatial clustering of socioeconomically disadvantaged neighborhoods and the positive autocorrelation (clustering) of journey to work trips. The former is commonly analyzed to reveal urban segregation or to investigate social or environment injustice issues. The latter pattern can provide valuable recommendations related to urban land use and road construction planning.

Spatial Clustering Analysis for Point Data

Many geographical phenomena can be conceptualized as point events, for which the spatial information is recorded as point locations and attributes may be associated with each point. Examples of point events include the distribution of a species of trees, the patterns of diseases or crime activities, and the spatial arrangement of stores within a study area. The techniques of spatial cluster detection for point data were first developed and applied by botanists and ecologists in the early years of the twentieth century. It was not until the mid decades of the last century when spatial point pattern analysis became widely accepted and applied by mainstream geographers.

Point clustering detection methods are different in multiple aspects. One group of techniques examine the points pairwise to see if they are closer to each other than expected or if similar measurements (variables measured at the point locations) are closer to each other than expected. Nearest neighbor analysis and related methods are such examples. Another group of techniques compare the observed number of points within a subregion or the measurements within a subregion with the expectance for a same subregion to decide if the observations are clustered. A subregion can be defined by a set of quadrats or a zoning system over the study area. It can also be defined as within certain distance from a focal location or any location in the study area. **Figure 2** illustrates how these two approaches define a subregion for clustering analysis. Not intended to exclusively discuss all cluster detection techniques, the following section summarizes the basic ones for point pattern cluster analysis.

Moran's *I*

Moran's I statistic is a quadrat method and assesses the clustering pattern of point distribution on a global scale. It applies a zoning system (consisting of regular or irregular cells) to the study area and uses a contiguity matrix to define neighboring zones. The index is calculated according to the following formula:

$$I = \frac{n \sum_i \sum_{j \neq i} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\left(\sum_i \sum_{j \neq i} w_{ij}\right) \sum_i (x_i - \bar{x})^2} \qquad [1a]$$

where n is the total number of zones, x_i is the intensity of points in zone i, $\bar{x}$ is the mean of x_i in the study area, and w_{ij} is weight measure to indicate the contiguity of zone i and zone j. The expected value of Moran's I is 0. A positive value indicates a spatial cluster.

A variation of Moran's I is the Local Indicator of Spatial Association (LISA) indicator. The local Moran's I is defined as

$$I_i = Z_i \sum_{j \neq i} w_{ij} Z_{j_{ij}} \qquad [1b]$$

Where Z_i represents the point intensity in zone i in a standard form. A pseudo significance level test of I_i can be obtained by either a conditional randomization or a permutation approach. A small p value indicates the presence of a clustering pattern in the neighborhood surrounding zone i.

Geary's *C*

Geary's C is similar to Moran's I. Instead of checking the variation of the observations from the mean, Geary's C is based on the weighted sum of the square difference

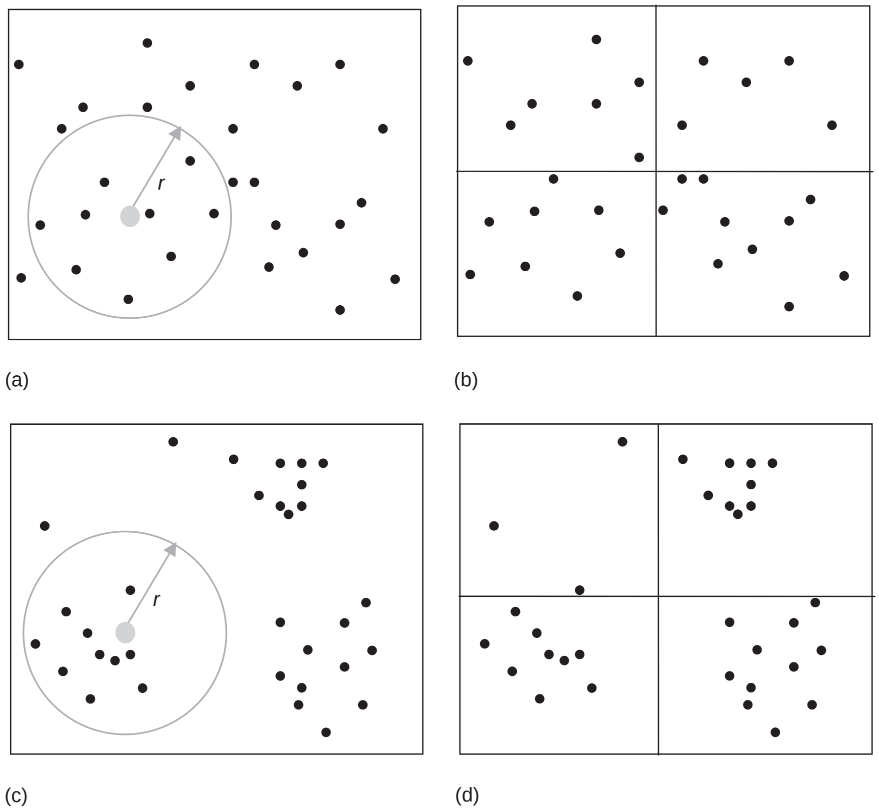

Figure 2 Different ways of measuring spatial separation by different cluster detection methods: (a) and (c) illustrate the definition of a subregion using distance methods for a random pattern and a clustered pattern, respectively; (b) and (d) illustrate the definition of subregions using quadrat method for a random pattern and a clustered pattern, respectively.

between observations. It is defined as

$$C = \frac{(n-1)\sum_i \sum_{j \neq i} w_{ij}\left(x_i - x_j\right)^2}{2\left(\sum_i \sum_{j \neq i} w_{ij}\right)\sum_i \left(x_i - \bar{x}\right)^2} \qquad [2a]$$

While the denotation in the above equation is the same as eqn [1a], the expected mean for Geary's C is 1. A value of less than 1 typically indicates the presence of clusters. Compared with Moran's I, Geary's C is more sensitive to the variation between immediate neighbors.

Local Geary's C_i also provides a local scale measurement for cluster analysis. It is defined as

$$C_i = \sum_{j \neq i} w_{ij}\left(Z_i - Z_j\right)^2 \qquad [2b]$$

A pseudo significance value can be obtained. A small p value indicates the similarity of the observation at i with those surrounding it.

Getis' *G*

Getis' G is a global statistic. It is defined as

$$G = \frac{\sum_i \sum_j w_{ij} x_i x_j}{\sum_i \sum_j x_i x_j} \qquad [3a]$$

An approximately normalized G can be calculated as $Z(G)$. A high positive $Z(G)$ indicates that the spatial pattern is dominated by clusters of high values, while a strong negative $Z(G)$ indicates that the spatial pattern is dominated by clusters of low values. One limitation of the G statistic is that it can only be applied to positive data value with a natural origin.

For local scale cluster analysis, $G_i(d)$ was developed to assess the concentration of observations in all zones that are within distance d to the focal zone, i. $G_i(d)$ is defined as

$$G_i(d) = \frac{\sum_j w_{ij}\left(x_j - \bar{x}_i\right)}{\sqrt{\frac{w_i(n-1-w_i)\sum_j \left(x_j - \bar{x}_i\right)}{(n-1)(n-2)}}} \; j \neq i \qquad [3b]$$

where $\bar{x}_i = \frac{1}{n-1}\sum_{j \neq i} x_j$, $w_i = \sum_{j \neq i} w_{ij}(d)$, and $\{w_{ij}\}$ is a symmetric binary spatial weight matrix. A significant and positive $G_i(d)$ indicates a cluster of large values surrounding the location i whereas a significant and negative $G_i(d)$ indicates a cluster of small values surrounding i. This statistic cannot identify negative spatial association (i.e., high value with surrounding low values and vice versa.).

Nearest Neighbor Index

Nearest neighbor index (*NNI*) is a distance method. It compares the average distance between the nearest neighbor points with that in an expected, random point distribution. *NNI* is the ratio of the observed nearest neighbor distance, d_{nn}, and the expected nearest neighbor distance, d_{ran}:

$$NNI = \frac{d_{nn}}{d_{ran}} \qquad [4a]$$

If *NNI* is smaller than 1, the point events are closer to each other than would be expected on the basis of chance. If *NNI* equals 1, then the observed point pattern is a random distribution.

It is worth pointing out that *NNI* only tests for the clustering of the nearest neighbor points. While it is expected that most clusters consist of more than two points, *NNI* can be extended to high order neighbor points. The *K*th order *NNI*, NNI_k, is calculated as

$$NNI_k = \frac{d_{nn,k}}{d_{ran,k}} \qquad [4b]$$

where $d_{nn,k}$ is the average distance between a point and its *K*th order nearest neighbor point, and $d_{ran,k}$ is the mean random distance between a point and its *K*th order nearest neighbor point.

K-function

K function is also called Ripley's *K* function. It is a dis tance method and is commonly used to detect the spatial extent of clustering in a point distribution. *K* function provides a continuous evaluation of clustering over a range of distances, and hence is claimed to be free from MAUP. Given a set of points, *K* function compares the observed *K* value at distance *d*, $\hat{K}(d)$, with the expected *K* value, *K*(*d*). By increasing *d* from small to large, typically with 50–100 small increments to cover the extent of the point set, a series of $\hat{K}(d)$ can be obtained and plotted against *d*. By comparing the curve of $\hat{K}(d)$ with that of *K*(*d*), one can tell if the observation set is clustered, randomly distributed, or dispersed at certain distances.

For significance testing, Monte Carlo simulations are usually conducted to generate a series of random distri butions of *n* points in the same study area. For each simulated distribution, a *K*(*d*) value is calculated for each distance band. It is a common practice to generate 100 or 1000 simulated distributions. The observed $\hat{K}(d)$ is then ranked with the set of *K*(*d*)s from the simulated distri butions. If $\hat{K}(d)$ is ranked to be significantly larger than the simulated *K*(*d*)s (e.g., larger than 99% of the *K*(*d*)s), the points are said to be clustered at distance *d*. In **Figure 3**, the dotted lines represent the envelope defined by the maximum and minimum of the simulated *K*(*d*) values across all distance bands, and the solid line rep resents the observed $\hat{K}(d)$s. The graph indicates that the point data are clustered up to 4.51mi; the data show a random pattern between 4.51 and 4.78 mi, and a dis persed pattern beyond 4.78 mi.

Kulldorff's Scan and Other Spatial Scan Techniques

Kulldorff developed a spatial scan statistic which uses circles as scan windows to test for spatial clustering.

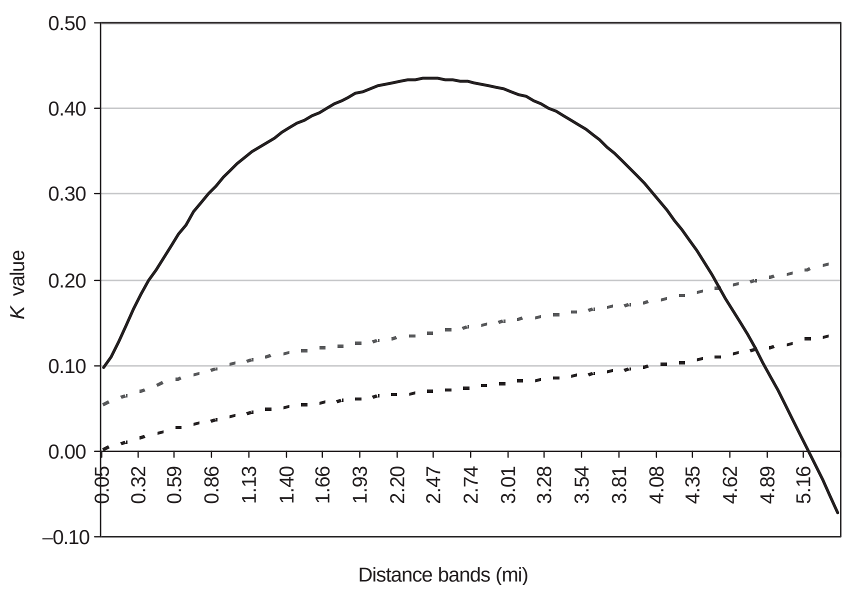

Figure 3 An example of the output from *K*-function. The dotted lines indicate the envelope defined by the maximum and minimum *K* values for all the distance bands from the 100 simulations, and the solid line represents the *K* value for all distance bands for the observed point distribution.

The spatial scan statistic defines a circular window that moves across the study area. The window centroid can be placed at the centers of polygons that cover the study area (e.g., census units) or at arbitrarily defined regular grid points. The radius of the window varies continu ously from zero to a predefined upper limit. To deter mine if clustering exists within a window, a likelihood ratio is derived by comparing the observed number of points within the window with the expected number of points according to a particular point distribution model (Poisson or Bernoulli). Based on the likelihood ratios, the method provides a list of the scan windows that most likely have clusters in them. Kulldorff's scan technique is implemented in the downloadable software, SaTScan.

There are other scan based techniques to detect cluster patterns. Geographical Analysis Machine (GAM) is one of these. GAM was progressively developed by Openshow during the 1980s and was modified in 1990s. Similar to Kulldorff's method, GAM uses an increasing sized, moving window to scan the study area for spatial clusters. Based on the number of points within the moving windows, GAM produces a surface map to illustrate the clustering patterns.

Most scan techniques allow for a series of increasing sized scan windows to search through the study area. The point count or density of the observed patterns in the scan window is compared with that of the expected patterns. A common challenge in this group of techniques is to model the expected point distribution when the point process is unknown or hard to describe. Monte Carlo simulation methods are commonly used to gen erate a large number of expected point distributions, against which the observed pattern can be compared. Another challenge for the scan based cluster detection methods is the efficiency of the search strategy. When the whole study area is seamlessly scanned, significant redundancy exists in the searching process: multiple search windows of the same size include the same subset of points. This can result in statistical bias associated with multiple comparisons.

Clustering Detection for Linear and Areal Data

In addition to points, geographical phenomena are fre quently conceptualized as lines and polygons. Migration trajectories, telephone communications, and monetary flows are examples of linear features. Flood zone, election zone, and urban neighborhood are normally represented as area objects. Cluster analyses for point data, linear/flow data, and polygon data are related. They all share the same fundamental concerns relating to the measurement of spatial separation, the scope of analysis, and the searching strategy. The techniques for point cluster detection are relatively more developed and form the foundation for many linear and areal clustering detection techniques. The clustering analysis of linear or areal data is closely linked to that of point data.

Linear feature clustering detection and analysis can be conducted either as vector autocorrelation analysis or the analysis of cluster correspondence between the cor responding end points. Vector clustering refers to the spatial pattern of a group of vectors that are located close to each other and have similar direction and length measurements. There is a general lack of techniques and applications for vector autocorrelation analysis in geography. Berglund and Karlström's investigation of vector association collapses back to the G^* statistic through assigning the vector outflow and inflow zone information to the end points of vectors. The cluster correspondence analysis developed by Lu and Thill proposes to 'decompose' a vector into a 'from point', an 'end point', and the distance and direction from the former to the latter. Therefore, the autocorrelation pat terns of flow data can be revealed by examining the clustering of the 'from points' and 'end points', respect ively, and then linking them together. A 'high–high' cluster correspondence between two sets of paired points indicates the existence of vector clustering.

Clustering analyses of areal data are also closely related to those of point data. The measurements within a polygon can be aggregated to the polygon centroid or the attribute of a polygon can be assigned to its centroid. Either way, the techniques of point clustering detection can be applied to the analysis of the patterns of polygon centroids. The spatial separation among areal features is commonly measured by a contiguity matrix. Thus, the traditional clustering detection methods, such as Moran's *I*, Geary's *C*, Getis' *G*, and SaTScan, are all applicable to areal data.

Other Issues in Spatial Cluster Detection and Analysis

Scale Issue and Cross-Scale Analysis

Spatial clustering detection is very sensitive to the def inition of the study area. A set of events with a regular pattern may appear clustered if the boundary of the study area is expanded. It is very important that the study area be defined to fit the spatial extent of the geographical distribution and of the area of interest.

Furthermore, the same set of events may show different patterns when analyzed at different scales. For quadrat approaches, different definitions of the quadrat grid will result in different estimates of point intensity in the zones, which may lead to different conclusions about clustering patterns. Changing from one geographical scale to another can generate different clustering results. For the same set of crime data, the clustering pattern detected at the census block group level might be different from that at the census tract level. Similarly, for the distance approaches, clusters identified at one distance might become insignificant at another distance. This is true for both the *K* function analysis and the scan based techniques.

Cross scale analysis of clustering pattern is frequently conducted when the dominant scale of the underlying process is not clear or hard to detect. Cross scale cluster analysis involves applying the same cluster detection technique at a series of spatial scales to reveal the dynamics of clustering patterns in the same distribution. The purpose of cross scale cluster analysis is more as a means to obtain a comprehensive understanding of the clustering patterns across scales rather than to identify the best scale for the analysis. For example, an environmental justice study on the relationship between the locations of landfills with hazardous materials and the residences of socioeconomically disadvantaged populations can be conducted at block group, census tract, county, and even state level. The major contribution of a cross scale study on this topic is not to reach a consistent conclusion regarding environment injustice; rather, it provides a full picture of the environment injustice problem at different spatial scales, with implications for policymaking and public practice at those scales.

Edge Effects

Cluster detection methods are influenced by edge effects, which are biases that occur when evaluating clustering close to the boundary of the study area. Statistics calculated along the boundary may be inaccurate because of the reduced window size and the 'missed' geographical events in the 'chopped off' area. To address this, one can build a buffer zone along the boundary of the study area. This is done either by defining the study area as smaller than the spatial extent of the observations or by obtaining more samples beyond a fixed study area. A second type of approach is to simulate the distribution of geographical events beyond the boundary. This is commonly done by mirroring the distribution within the boundary to adjacent areas outside the boundary.

Multivariate Clustering Analysis

Cluster detection methods have been extended to analyze spatial clustering of multiple types of variables. Developed in botany and ecology, these methods evaluate if events of one type tend to be clustered or dispersed around events of another type. This group of techniques can be used to examine if certain types of stores tend to be located close to another type of stores or if one type of crime is spatially very close to another type of crime. Both the quadrat approach and the distance methods for univariate clustering detection can be modified and extended to handle the cluster detection of multivariate geographical patterns.

Clustering of Multilocation Events: Correspondence of Clusters and Other Issues

Certain geographical phenomena are associated with more than one location. 'Journey to work' and patient's migration trajectory are such examples. These phenomena are multilocation events. Each event involves multiple locations; the specific connection between locations of different types is preserved for the clustering analysis. For 'journey to work', each trip is specifically defined by the connection between a residence location and a workplace. Any change in either the residence or the workplace results in a different journey to work event. Correspondence of cluster analysis investigates the spatial clustering of multilocation events. When a set of type A locations cluster, cluster correspondence exists if the corresponding set of type B locations are also clustered. Cluster correspondence analysis tests for the clustering patterns of multilocation events by examining the coexistence of clustering patterns among the corresponding locations that define the multilocation events.

Some multilocation events consist of events that occur through time. In such cases, the associated time stamp for each location is critical. 'Geospatial lifelines' is a term coined to describe geographical events each consisting of a sequence of locations at specific time periods. Clustering analysis of such events needs to consider both space and time simultaneously. Techniques for space–time clustering analysis have been developed to search for clusters of events in space and time.

Summary

Spatial clustering methods analyze the positive spatial autocorrelation patterns embedded in the geographical events under investigation. These events can be represented as points, lines/flows, or areas. Clustering analysis evaluates the distribution of the observed events and compares it with that of the expected. The fundamental considerations for analyzing clustering of point, linear, or areal features are the same. Clustering methods are well developed for point features, and cluster detection for linear and areal features often collapses back to point pattern analysis one way or another. That being said, quite a few point pattern analysis techniques were developed mainly for the analysis of areal data. Such examples include Moran's *I* and Getis' *G* statistics. Nevertheless, clustering detection techniques for vector data remain an important challenge. Another challenge is to develop methods for space–time clustering detection of multilocation events, or geospatial lifelines. Further development of clustering detection methods also calls for approaches to account for the dynamics of clusters in the same data set through time, that is, cluster sur veillance. The evaluation of local scale clustering when large scale spatial autocorrelation exists is yet another challenge for future research.

See also: Edge Effects; Exploratory Spatial Data Analysis; Modifiable Areal Unit Problem; Point Pattern Analysis; Spatial Autocorrelation; Statistics, Spatial.

Further Reading

Bailey, T. C. and Gatrell, A. C. (1995). *Interactive Spatial Data Analysis*. Harlow, UK: Longman Scientific & Technical.

Berglund, S. and Karlstrom, A. (1999). Identifying local spatial association in flow data. *Journal of Geographical systems* 1, 219 236.

Boots, B. N. and Getis, A. (1988). *Point Pattern Analysis*. Newbury Park, CA: Sage Publications.

Diggles, P. J. (2003). *Statistical Analysis of Spatial Point Patterns* (2nd edn.). London: Arnold.

Everitt, B. (1974). *Cluster Analysis*. London: Heinemann Education Books.

Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics Theory and Methods* 26, 1481 1496.

Lu, Y. and Thill, J. C. (2003). Assessing the cluster correspondence between paired point locations. *Geographical Analysis* 35, 290 309.

Mark, D. M. and Egenhofer, M. J. (1998). Geospatial lifelines. In Gunther, O., Sellis, T. & Theodoulidis, B. (eds.) *Dagstuhl Seminar Report No. 228: Integrating Spatial and Temporal Database*. Wadern, Germany: Dagstuhl.

Odland, J. (1988). *Spatial Autocorrelation*. Newbury Park, CA: Sage Publications.

Ord, J. K. and Getis, A. (1995). Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis* 27, 286 305.

Rogerson, P. A. (1997). Surveillance systems for monitoring the development of spatial patterns. *Statistics in Medicine* 16, 2081 2093.

Sinha, G. and Mark, D. M. (1995). Measuring similarity between geospatial lifelines in studies of environmental health. *Journal of Geographical Systems* 7, 115 136.

Tobler, W. (1979). Cellular geography. In Gale, S. & Olson, G. (eds.) *Philosophy in Geography*, pp 379 386. Dordrecht: Reidel.

Upton, G. J. G. and Fingleton, B. (1985). *Spatial Data Analysis by Example, Vol. I: Point Pattern and Quantitative Data*. New York: Wiley.

Relevant Websites

https://www.geoda.uiuc.edu
GeoDa An Introduction to Spatial Data Analysis.

http://www.satscan.org
SaTScan.

http://www.terraseer.com/products clusterseer features.php
TERRASEER.

Spatial Data Mining, Cluster and Pattern Recognition

T. A. Arentze, Eindhoven University of Technology, Eindhoven, The Netherlands

Glossary

Clustering The task of grouping objects in a database into meaningful classes (clusters) based on similarity or dissimilarity of objects given their attributes.

Data Mining The process of discovering interesting and useful patterns in large databases through automated methods.

Discrimination Classification The task of assigning objects to one of a set of predefined and mutually exclusive classes based on their values on one or more attribute variables.

Feature Selection The selection of attributes to be used for training a classifier from a database with the aim to improve the accuracy or interpretability of the classifier or both.

Inductive Learning The process of discovering relationships in data of a finite set of cases in a domain that are generalizable to other, unseen cases in the same domain.

Knowledge Discovery The process of extracting rules from data or from patterns extracted from data generally by using an inductive learning technique and sometimes by direct mapping techniques.

Spatial Data Mining Data mining of 'spatial' databases, that is, databases where one or more attributes used to describe units of observation are dependent on the specific locations of the units in geographic space.

Supervised Learning Inductive learning of rules for classifying objects based on examples of objects of which the target class is known.

Unsupervised Learning Inductive learning of rules for classifying objects based on examples of objects of which the target class is unknown.

Validation of a Classifier Estimating the ability of a classifier to predict classes of unseen cases accurately usually by testing goodness of fit of predictions on a holdout set (i.e., cases that originate from the same dataset as those used for training the classifier).

Introduction

Data mining is the process of identifying interesting and useful patterns in large databases by means of automated methods. The databases of interest in (human) geography include spatially referenced observations of some kind that are described in terms of spatial as well as nonspatial attributes. The detection of clusters and patterns in spatial databases is a relevant issue in many areas of geographic research such as the interpretation of remote sensing data of the Earth's surface, the analysis of spatial dispersion of crime, diseases, and other phenomena, the study of space–time behavior of individuals, households or firms, and so on. As for terminology, the terms data mining and knowledge discovery in data (KDD) are generally used as synonyms. However, extracting rules from patterns revealed by a data mining algorithm is sometimes implemented as a post processing step to aid interpretation activities. Then, the terms data mining and knowledge discovery are also used to refer to the two components of a single KDD process.

Data mining is a relatively young field of research which emerged in the 1990s in response to a combination of mainly two developments: the increasing availability of very large datasets through automated data recording methods (e.g., transaction data in marketing) and the increasing availability and power of machine learning methods emerging from work in the fields of machine learning and artificial intelligence. The field brings together various disciplines, including statistics, pattern recognition, artificial intelligence, machine learning, and information systems. Unlike established statistical methods of data analysis, machine learning and pattern recognition methods implement local search algorithms which can handle very large feature spaces efficiently. Large databases are common in geographic analysis, due to a general aim to study spatial phenomena on a high spatial (and temporal) resolution, the wide application of automated data collection methods such as remote sensing and Global Positioning System (GPS) tracking methods, and availability of powerful geographic infor mation system (GIS) software to manage and process the data. Furthermore, data mining techniques do not rely on specific assumptions about the form of data distri butions and, hence, are suited to reveal nonlinear, irregular, as well as linear relationships in data.

Clustering and pattern recognition is only one of the standard tasks considered in data mining. Other data mining tasks include regression, summarization, dependency modeling and change, and deviation detection. Clustering and pattern recognition are forms of classification problems that differ on an important characteristic. In clustering, the classes to which objects are to be assigned are not predefined, but rather to be identified based on distributions of objects on (spatial and/or nonspatial) attributes. The aim is to identify groupings of objects such that objects of the same group are as similar as possible to each other and objects of different groups are as dissimilar as possible. On the other hand, in pattern recognition, also known as dis crimination classification, classes are predefined and the task considered is to assign objects to classes based on their attributes.

Learning based data mining techniques to develop classifiers of these types are known as unsupervised learning (clustering) and supervised learning (dis crimination). In both forms of learning, rules are to be learned by induction, that is, by identifying generalizable characteristics in data from a limited set of cases provided to the classifier. In supervised learning, the so called training cases are preclassified, whereas in unsupervised learning the training cases have no class label. Well known examples of inductive learning techniques are decision trees and neural networks. Apart from algorithms based on principles of inductive learning, numerical algorithms stemming from longer traditions in statistics and pattern recognition fields play a role.

In the remainder of this article, we discuss issues and developments that play a role in these areas of (spatial) data mining. The discussion is organized in several sec tions. The first two sections look closer into existing methods of clustering and (discrimination) classification. The sections that follow then focus on issues and methodology that are relevant in the application of both types of classifiers. These pertain to validation testing, feature selection (FS), knowledge discovery, and treat ment of spatial data.

Classification by Clustering

In this area, numerical algorithms stemming from statistics and pattern recognition play a more important role than inductive learning algorithms. Three groups of numerical approaches can be distinguished: partition based algo rithms, density based algorithms, and hierarchical clus tering algorithms. Partition based algorithms assume that the number of clusters, k, to be identified is given and try to find a way of assigning objects (i.e., data points) to clusters that maximizes the homogeneity of clusters or distance between clusters in attribute space. A well known example of this group is the k means algorithm. This algorithm tries to improve an initial solution of cluster centers by an iterative procedure of moving cluster centers and re allocating data points in each solution to the nearest center. An error score of the cluster solution is calculated based on measuring the distance in attribute space between each data point and cluster center. Density based algorithms too assume the number of clusters, k, as given. However, rather than an all or nothing classification they use cluster membership probabilities and try to find the cluster solution that maximizes a likelihood function. The expectation–maximization (EM) algorithm is often used to estimate the parameters of a cluster solution that maximizes this likelihood function. Hierarchical clustering algorithms, in contrast, do not assume a predefined num ber of clusters, but rather develop a hierarchy of possible cluster solutions. Based on a matrix of distances between data points, these methods produce a hierarchy usually by a process of successive merging of rows and columns into clusters. The hierarchical cluster structure has a useful visualization, known as a dendrogram, which supports the analyst in determining the optimal degree of clustering considering cluster number and cluster distance criteria.

In the study of spatial dispersion of phenomena such as crime, diseases, etc., which is an important topic area of geographic study, the data points are located in geographic space. Established algorithms reviewed above are not readily applicable to this problem. Because the population that produces the phenomena considered is generally not equally spread across space, the relevant question is not ‘where are spatial clusters of the data points?’, but ‘where are clusters with higher occurrences of these ‘events’ than expected given the background population?’. Furthermore, given the large size and high spatial resolution of study areas typically involved in these studies, methods relying on the construction of a distance matrix are often not feasible as they would require too long computation times. Finally, the Gaussian form of density distributions, which are assumed by parametric methods, such as EM, often do not hold in the geographic domain where distributions are commonly skewed. The Geographical Analysis Machine (GAM) developed by Openshaw and colleagues is an example of an influential spatial cluster detection technique that intends to overcome all these problems. In this method, a cluster is defined as a circular region with greater than average point density (e.g., of a disease) and an exhaustive search for clusters in all possible circles at all possible locations (in a grid) is implemented. Improving com putational efficiency and reducing the number of false positive detections are problems that have been addressed in alternative approaches or extensions of the GAM. Genetic algorithms and graph based algorithms are alternative approaches that have been developed.

In the study of space–time behavior of individuals, the problem of identifying clusters in activity travel data has attracted attention. A special characteristic of this cluster problem is that the units of analysis are not points but rather sequences of points in space and time, namely activity sequences conducted by individuals in some time window (usually a day or a trip chain). String alignment methods have been developed to measure distances (i.e., dissimilarities) between activity patterns as an alternative to traditional Euclidean based methods, to take sequence information into account. String alignment methods measure the distance between two strings of characters as the amount of effort required to make the strings identical through deletion, insertion, and substitution operations. They originate from microbiology where they are used to measure (dis)similarity between deoxyribo nucleic acid (DNA) strings and were introduced to the domain of activity analysis by Wilson in the late 1990s. Applications in activity analysis have focused on finding clusters in activity pattern data either through aligning all strings simultaneously (multiple string alignment) or by constructing a pairwise distance matrix first and then using an existing hierarchical cluster algorithm to iden tify clusters. An extension to the case of multidimen sional strings has been developed by Joh and colleagues to better deal with the multidimensional nature of activity sequences. Methods to detect common elements in sequences and to deal properly with location and time dimensions of activity strings are topics of current research.

Classification and Supervised Learning

In this type of task, the classes objects can belong to are known *a priori* and the objective is to discover clas sification rules based on a set of examples. This task is relevant in several areas of geographic research. The area where it is investigated most extensively is, without doubt, information extraction and pattern recognition in remote sensing data (e.g., to detect or predict land cover changes). However, it also finds application for the dis covery of patterns in datasets from various sources such as census data, count data, administrative record data, survey data, etc. Although inductive learning algorithms are more important in this type of task compared to clustering, they are not the only class of techniques used. Overall, we can distinguish the following approaches: decision tree induction (or, decision tree discovery), decision rule induction, Bayesian classification, nearest neighbor approaches (e.g., case based systems), max imum likelihood classification (e.g., logistic regression), and neural network approaches. Neural networks and decision trees are the most widely used techniques and because they are illustrative for a wide range of approaches we will consider them here in more detail.

Neural networks are used for more data mining tasks than just classification. When used for classification, the feedforward network or, more specifically, the multilayer perceptron (MLP), with a backpropagation learning al gorithm, is usually the most suitable type of network. **Figure 1** shows the architecture of an MLP with one hidden layer. The number of hidden layers and number of nodes within layers should be chosen depending on the characteristics of the problem at hand. Nodes in the input layer represent the attribute variables used for

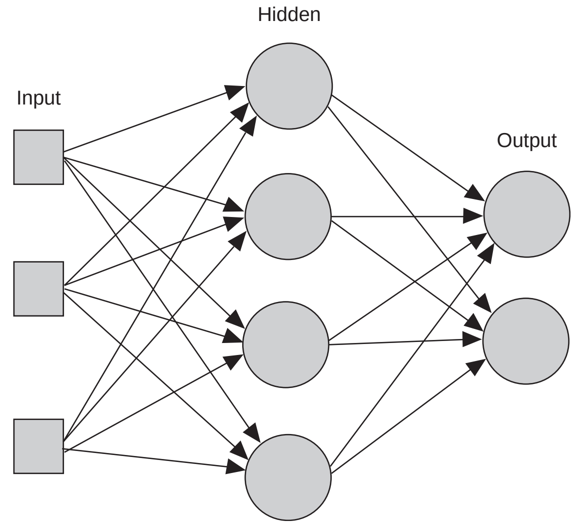

Figure 1 Architecture of an MLP with one hidden layer.

classification and the nodes in the output layer stand for the class labels. Input nodes simply transmit their input values to the next layer. Nodes in hidden and output layers use a function to aggregate received inputs (usually, weighted summation), a function to determine the activation level (usually, a logistic function), and a function to calculate the output (usually, the identity function). The output pattern represents the classification result. Usually, the assigned class is identified as the class represented by the node having the highest activation level. During the network training, the weights of links are adjusted through backpropagation of classification error at the output nodes. The link weights after training represent the knowledge the network has extracted from the data.

A decision tree has also a graph representation, but here the nodes and links constitute decision rules. **Figure 2** shows an arbitrary example of a decision tree. There are two types of nodes in a decision tree: test nodes and leaf nodes. A test node defines a test to be carried out on the value of an attribute variable. The branches at the node represent the possible outcomes of the test. Leafs are the terminal nodes of the tree. They specify the class to which an unlabeled case is assigned by the tree. A decision tree is learned from examples by an inductive learning algorithm. Well known induction systems, such as C4.5 and CART, use a top down approach. In this approach, a tree is developed starting from the root node by recursively splitting the training set on attribute variables until cases within partitions are homogeneous in terms of the class label. The way of splitting the sample at each node of the (growing) tree is the result of selecting, in a greedy manner, the candidate test that maximizes a splitting criterion. For example, C4.5 uses the information gain (or, to be precise, the gain ratio) and CART the Gini index as splitting criterion.

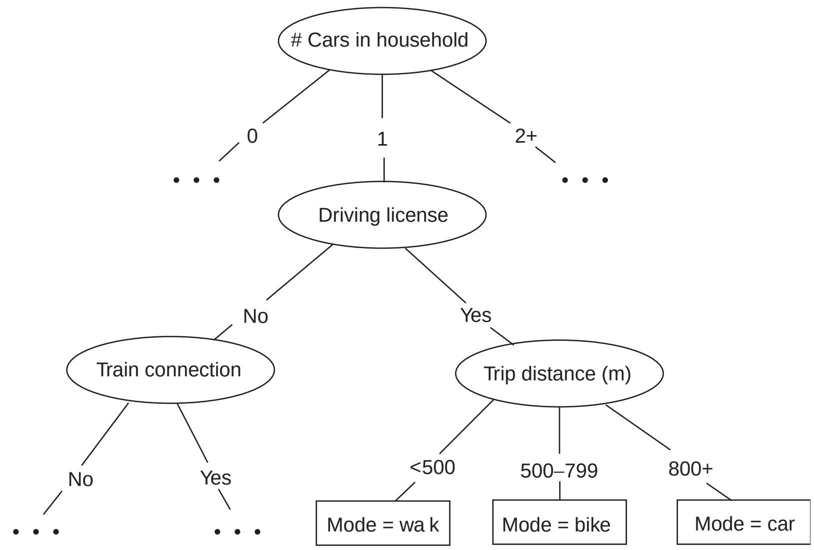

Figure 2 Example of a decision tree.

Systems either use a stopping rule to avoid the formation of insignificant branches or let the tree grow fully and prune insignificant branches afterwards. The latter approach reduces the risk of not identifying branches that only at deeper levels of the tree turn out to be significant. After having developed the tree, an action is assigned to each leaf node based on the classes observed in the corresponding partition of the training set. Usually, the modal class is selected. To obtain a probabilistic classifier, alternatively, class membership probabilities can be taken.

A common characteristic of neural networks, decision trees, and other types of discrimination classifiers is that they define a partitioning of the attribute space to try and separate objects of different classes. A fundamental property of a classification problem in this regard is linear separability. **Figure 3a** visualizes this concept graphically for a two class, two attribute classification problem. In the case shown, it is possible to separate the squares and triangles by means of two different linear combinations of the attributes. In two dimensional (2 D) space, the edges are lines, in a 3 D space they appear as planes, and in a general *n* D space they are referred to as hyperplanes. In case of an MLP, every neuron in the hidden layer usually defines a hyperplane in attribute space. An increase in number of hidden neurons, thus, means an increase in the power of the model to separate classes. In case of a de cision tree, hyperplanes are defined by test nodes and run orthogonal to the attribute axes. **Figure 3b** illustrates how a set of decision rules partitions attribute space.

The MLP is an important type of neural network for constructing classifiers, but not the only one. The radial basis function (RBF) network is another neural network model with specific strengths for classification. These networks have an input layer, an output layer, and a single hidden layer. Unlike MLPs, they use RBFs as activation function for the neurons in the hidden layer. These functions define a closed region around a center point rather than a hyperplane in attribute space. **Figure 3c** illustrates how an RBF network partitions attribute space by fixing a collection of such regions. It shows that RBF networks are potentially powerful for problems where classes are not linearly separable but tend to cluster around prototypical points in attribute space. By that property, RBF networks resemble another type of neural network called the self organizing map (SOM). Similarly, an SOM subdivides an attribute space by defining ellipsoidal or Voronoi like regions, depending on the function used. As these networks are trained using unlabeled instances, they belong to the class of unsupervised learning techniques.

In the area of decision trees, alternative forms of the representation model are used as well. So called oblique decision trees allow tests at nodes to be linear combi nations of attributes. As a result, hyperplanes defined by test nodes are no longer restricted to run parallel to at tribute axes. This increases the power of the trees to learn effective classification rules. A disadvantage, however, is loss of comprehension. Fuzzification of decision trees is an approach that has been proposed to introduce vagueness in assigning cases to outcomes at test nodes. The parametric action decision tree is yet another development that re places a static action assignment rule at each leaf node by a discrete choice model to predict class membership prob abilities as a function of one or more numeric attributes.

Accuracy and Validation

The accuracy of a classifier is generally assessed by using some measure of goodness of fit of predictions on

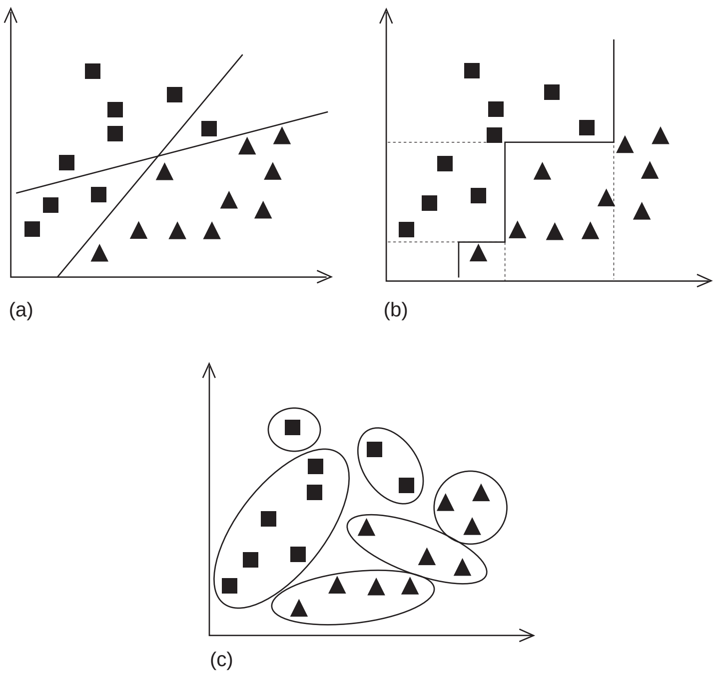

Figure 3 Classification in a two-class, two-attribute classification problem by an MLP neural network (a), decision tree (b), and RBF neural network (c). Modified from Gahegan, M. (2000). On the application of inductive machine learning tools to geographical analysis. *Geographical Analysis* 32, 113 139.

observations. Examples of such measures are percentage of correctly predicted cases, likelihood of the model, distance based measures (e.g., aggregate pairwise dis tances between data points and cluster centers), and entropy based measures. These measures can be calcu lated on the same data as used for training, but then no information is obtained on the generalizability of learned rules to unseen cases. Overfitting the training data occurs if the rules learned incorporate specific characteristics of the training data. Then, accuracy drops when the classifier is applied to new cases. In case of neural network classifiers, the choice of number of hidden neurons and hidden layers is critical for expected accuracy. If this number is set too low, the learning capability of the network is affected and if set too high, nongeneralizable features of the training data will be extracted. In case of decision tree learning, threshold parameters of the stopping rule or pruning rule, such as minimum number of cases at leaf nodes and minimum level of significance of branches, provide the means to control the risk of overfitting. In partition based cluster methods, the number of clusters, *k*, can be varied to reveal the relationship between number of clusters and accuracy in cluster solutions.

Using a holdout set for testing is a commonly used method to assess the extent to which overfitting occurred in training a classifier. A disadvantage of this method is that not using all data for training may reduce the ability to discover rules or may increase costs of data collection. This problem does not occur when data is available in abundant quantity or can be collected against low costs. However, generally, economic use of data is an issue and research effort has focused on methods of validation testing that minimize loss of data for training. One such method which is widely used is *k* fold cross validation. In this method, the dataset is split into *k* mutually exclusive subsets of nearly equal size and the classifier is trained *k* times on *k* 1 subsets of the data, each time using a different subset for validation. As a result, all cases are used for training and every case is used once for testing. The average performance across *k* runs on the holdout cases provides a reliable estimate of accuracy. A typical setting in applications of this method is $k = 10$. Another well known method is bootstrapping. In this method, bootstrap samples are created by random sampling of cases with replacement from a dataset for additional validation testing.

Usually, validation methods, such as holdout sets and cross validation, are used to test a classifier after training. The methods may also play a role, however, in con figuring a classifier preceding training. For example, *k* fold cross validation may be used to determine the number of clusters in partition and density based cluster methods. Holdout sets are also used as part of automated methods to optimize the number of hidden layers and hidden neurons in an application of a neural network. Finally, a holdout set plays a role in some training procedures. If a gradient descent learning algorithm is used to train a neural network, the moment of stopping the training process is an important decision to avoid overfitting. It is a very common practice to use a holdout set to test performance after each cycle of training and stop the process the moment performance on the holdout set starts to decline.

Feature Selection

Much of the power of data mining methods (for classi fication and clustering) arises from the ability of these methods to deal efficiently with high dimensionality data, that is, where cases are described in terms of many attributes. Nevertheless, there are sometimes good rea sons to make a selection of the most relevant attributes for a classification or cluster task in a data preprocessing phase. This is known as FS. There are several reasons for considering FS. First, learning methods are based on local search algorithms and as a consequence may become trapped in local optima. The presence of irrelevant attributes increases this risk and, hence, FS may lead to a better learning result. Second, decision trees and rule based classifiers in general, when they are induced from data, may become very large and complex. Eliminating less relevant attributes in a preprocessing stage may help, just as pruning afterwards, to reduce the size and improve comprehension of a tree or rule set against minimum costs of loss in accuracy.

A variety of FS methods have been developed for application in machine learning. The methods commonly consist of a component to generate subsets of attributes by forward (adding attributes to an empty set) or back ward selection (eliminating attributes from an exhaustive set) and a component to evaluate subsets generated. Two major approaches can be distinguished: wrapping and filter techniques. Wrapping techniques are, as it were, built around the classifier they intend to support. The classifier is used to measure the suitability of a subset of features (i.e., attributes) by training the classifier based on this subset and evaluating its performance after training. Thus, in wrapping methods, training of the classifier is embedded in a search process for the best FS solution. In contrast, filter methods are implemented prior to training and use an independent measure to evaluate an FS solution. An argument in favor of the wrapping approach is that the classifier provides a better measure than an independent measure. A disadvantage of the approach is that wrapping methods increase computation time by a huge factor.

Knowledge Discovery

The patterns revealed through automated (learning) methods may be difficult to interpret either because the representation model used does not support this (e.g., neural networks) or because the extensiveness and complexity of the patterns hamper this (e.g., rule based models). Knowledge discovery, in the more limited sense of the term as used in this section, refers to a post processing of patterns generated by a data mining algorithm to support interpretation activities. This has received attention particularly in the area of (feed forward) neural networks. Two approaches can be dis tinguished. In one approach, rules are extracted on the level of individual neurons by mapping the input–output functions the neurons in hidden and output layers com pute into if–then rules. In another approach, the neural network is treated as a black box and rules are discovered using some rule induction method from predictions of the network for a (synthesized) sample of cases. In so called neuro fuzzy systems, fuzzy rules are used to represent the description or the behavior of a trained network.

Another aspect that has received attention is that rule based models, such as decision trees, do not provide quantitative information on the relevance and signifi cance of attribute variables for class membership pre diction. In contrast, estimates of coefficients obtained in parametric methods indicate sizes and directions of im pacts of explanatory variables on the dependent variable. A question of interest is whether such information can also be extracted from a rule based model. In the CART system, the reduction of classification error attributed to an attribute variable is provided as a measure of signifi cance. In another approach, the impact of each attribute variable is measured as the sensitivity of predictions for variation over the value range of that attribute variable in training or holdout data.

Treatment of Spatial Data

Databases involved in geographic analysis include spatial as well as nonspatial attributes. To a certain extent, spatial attributes do not require special treatment and can be supplied to algorithms for learning classifiers just as nonspatial attributes. However, this is true only to some extent. Attributes of objects that are close to each other in geographic space tend to be correlated; a phenomenon which is known as spatial auto correlation. To make use of this information, it is important that geographic structure is represented in data used to construct a classifier. The concept of a neighborhood graph is used to represent spatial relationships effectively for developing classifiers. In this approach, attributes of neighboring objects are used for classification as well as those of the actual object. Alternatively, smoothing and generalization methods have been proposed to represent spatial struc ture in data. Spatial relationships are also explicitly used in the phase of FS. A central concept in spatial FS is that the relevance of an attribute increases if neighboring objects of the same class share that attribute or if neighboring objects of a different class do not share that attribute. Finally, effective ways of encoding geographic location of objects for classification and clustering have attracted attention. Using x and y coordinates as separate attributes may complicate the task of a classifier as they need to be processed simultaneously to pinpoint an object in space. Schemes such as Morton coding provide methods to map pairs of x and y coordinates into a 1 D key and thus may offer more effective ways of spatial coding.

Conclusion

The field of data mining, although it is a relatively young field, is characterized by a large variety of approaches. This reflects the multitude of research disciplines involved and offers the analyst a wide choice range of techniques. The choice of technique will depend on the nature of the task (clustering or classification), characteristics of the database, and relative importance assigned to different performance aspects of methods. Regarding the latter, a strong learning capability and ready interpretability of output of a data mining algorithm sometimes stand in a conflicting relationship to each other. This article discussed numerical cluster methods, the neural network, and decision tree in more detail as they are the most important techniques and together representative for a wide range of approaches.

See also: Neural Networks; Spatial Autocorrelation; Spatial Clustering, Detection and Analysis of.

Further Reading

Arentze, T., Hofman, F., van Mourik, H., Timmermans, H. and Wets, G. (2000). Using decision tree induction systems for modeling space time behavior. *Geographical Analysis* 32, 330 350.

Conley, J., Gahegan, M. and Macgill, J. (2005). A genetic approach to detecting clusters in point data sets. *Geographical Analysis* 37, 286 314.

Fayyad, U. M., Piatetsky Shapiro, G., Smyth, P. and Uthurusamy, R. (eds.) (1996). *Advances in Knowledge Discovery and Data Mining*. Menlo Park, CA: The MIT Press.

Frank, E., Wang, Y., Inglis, S., Holmes, G. and Witten, I. H. (1998). Using model trees for classification. *Machine Learning* 32, 63 71.

Gahegan, M. (2000). On the application of inductive machine learning tools to geographical analysis. *Geographical Analysis* 32, 113 139.

Joh, C. H., Arentze, T., Hofman, F. and Timmermans, H. (2002). Activity pattern similarity: A multidimensional alignment method. *Transportation Research B* 36, 385 403.

Klosgen, W. and Zytkow, J. M. (eds.) (2002). *Handbook of Data Mining and Knowledge Discovery*. Oxford: Oxford University Press.

Luo, J. C., Leung, Y., Zheng, J. and Ma, J. H. (2004). An elliptical basis function network for classification of remote sensing images. *Journal of Geographical Systems* 6, 219 236.

Murray, A. T. (1999). Spatial analysis using clustering methods: Evaluating central point and median approaches. *Journal of Geographical Systems* 1, 367 383.

Pal, S. K. and Mitra, P. (2004). *Pattern Recognition Algorithms for Data Mining*. London: Chapman and Hall.

Quinlan, J. R. (1993). *C4.5 Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann Publishers.

Swingler, K. (1996). *Applying Neural Networks: A Practical Guide*. London: Academic Press.

Witten, I. H. and Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. Amsterdam: Morgan Kaufman.

Spatial Data Mining, Geovisualization

C. Kim, University of Cincinnati, Cincinnati, OH, USA

Glossary

Co-Location Pattern It is a group of spatial features that are frequently co-located in the same region. The goal of spatial co-location pattern mining is to find subsets of spatial features frequently located together in spatial proximity. Examples of co-location patterns include services requested frequently and located together from mobile devices (e.g., personal digital assistants (PDAs) and cellular phones). For co-location pattern mining, previous studies often emphasize the equal participation of every spatial feature. As a result, interesting patterns involving events with substantially different frequency cannot be captured.

Exploratory Spatial Data Analysis (ESDA) It is usually the very first step in any spatial analysis. It detects spatial patterns in data, formulates hypotheses based on the geography of the data, and assesses spatial models. Linking numerical and graphical procedures with the map is most essential. Spatial data mining provides very specialized ESDA tools for outlier analysis, normality test, and distribution analysis.

Hot Spot Analysis It is the process of finding unusually dense event clusters across time and space. Many criminal justice agencies are exploring the benefit provided by computer technologies to identify crime hot spots in order to take preventive strategies such as deploying saturation patrols in hot spot areas.

KDD (Knowledge Discovery in Databases) It is the field that is evolving to provide automated analysis solutions. Knowledge discovery is defined as the nontrivial extraction of implicit, unknown, and potentially useful information from data. The knowledge discovery process takes the raw results from data mining and carefully and accurately transforms them into useful and understandable information.

Markov Random Fields (MRFs) These provide a convenient and consistent way to model context-dependent entities such as image pixels and correlated features. The spatial property can be modeled through different aspects such as the contextual constraint.

Spatial Autoregression (SAR) Model It is a knowledge discovery technique used for mining massive geospatial data in many application domains. It is an extension of the classical regression model for incorporating spatial dependence. It is popular for prediction and classification of spatial data in regional economics, natural resources, and ecological studies. Estimation of the parameters of the exact SAR model using maximum likelihood (ML) theory is computationally very expensive because of the need to compute the logarithm of the determinant of a large matrix in the log-likelihood function.

Variogram Cloud This may reveal extreme outlying points that are dominating the estimate of the sample variogram.

Geovisualization is a research led field that integrates approaches from visualization in scientific computing, cartography, image analysis, information visualization, exploratory data analysis, as well as geographic information system (GIS). It uses visual representations to facilitate thinking, understanding, and knowledge construction about human and physic environments, at geographic scales of measurement. Geovisualization is versatile in that it can perform both cartographic display and limited analysis. It augments human visual ability in perceiving high complex structures, and detecting, exploring, and exploiting significant patterns. It integrates scientific visualization with traditional cartography, and can be utilized at data pre processing, data mining, and knowledge construction. The main purpose of geovisualization, however, is on insight rather than maps. The insight learned from discovery and explanation leads to decision making.

Exploration and communication make geovisualization an attractive way for the collaboration of specialized users and the public. In terms of data exploration and analysis, geovisualization has been referred to as visual data mining. Maps have historically been a database of information. With the increased availability of geospatial data, however, maps are no longer the physical database but rather they are visual projections of the databases that they represent. As a result, maps have become a tool that users can interact with to visualize the outcomes of a multitude of inputs. Maps are used to stimulate visual thinking about geospatial patterns, relationships, and trends. The use of visual geospatial display explores data, generates hypothesis through exploration, develops problem solutions, and constructs knowledge. The most successful geovisualization method is the one that takes the fullest advantage of human sensory and cognitive systems developed for interacting with the real world. At the same time, the best way of spatial data mining is to explore data using maps.

International Cartographic Association (ICA) Commission on Visualization and Virtual Environments proposed four main research agendas in geovisualization:

cognitive and usability issue in geovisualization, repre sentation and its relationship with cartographic visual ization, the integration of geographic visualization with knowledge discovery in databases and geocomputation, and user interface issues for spatial information visual ization. First, cognitive and usability issues are con sidered in the context of six major research themes: geospatial virtual environments (GeoVEs), dynamic representations, including animated and interactive maps, metaphors and schemata in user interface design, individual and group differences, collaborative geo visualization, and evaluating the effectiveness of geo visualization methods. Second, representation and its relationship with cartographic visualization issues are stressed with five themes: the nature of what kind of data and phenomena are to be represented, the form of representation chosen, the purpose for which represen tation is undertaken and used, the users for whom representation is undertaken, and the methods and the technologies that enable representation to be accom plished. Third, research agenda is synthesized in the form of three sets of research questions addressing: visual approaches to knowledge discovery, visual support for knowledge construction and geocomputation, and databases and data models that must be satisfied to make visually led knowledge construction a reality in the geographic sphere of interest. User interface issues for spatial information visualization are addressed with four themes: the general concepts of interfaces and their use for interacting with representations of geography, the design of metaphors and access routines, manipulation tools and techniques as well as navigation, the import ance of universal access with international Internet standards efforts, and the practical implementation of interfaces using new technologies.

Geospatial Data

There are several important characteristics of geospatial data. First, observations of spatial data are dependent on its location. Second, the size of spatial data is enormous. Third, spatial relationships are imbedded in spatial data. Fourth, spatial attributes are defined in spatial data as well as nonspatial attributes. Efficient geovisualization tools for extracting information from geospatial data are crucial to organizations which make decisions based on large spatial datasets, including NASA Earth Observing System (EOS), the National Imagery and Mapping Agency (NIMA), the National Cancer Institute (NCI), the National Institute of Justice (NIJ), the Census Bur eau, National Institute of Health (NIH), and the United States Department of Transportation (USDOT). These organizations are spread across many application do mains, including ecology and environment management, public safety, transportation, Earth science, epidemi ology, and climatology.

Spatial data are different from nonspatial data due to the complexity of spatial data (e.g., multidimension and autocorrelation) and spatial relationships (e.g., adjacency, connectivity, direction, and inclusion). There are several major knowledge constructions for spatial data mining: spatial clustering, spatial prediction, spatial character ization, spatial association, spatial trend detection, and spatial classification. Each has its own algorithms and prototypical applications.

Spatial clustering is the task of grouping objects of a database into meaningful subclasses. It is a process of grouping a set of spatial objects into clusters so that objects within a cluster have high similarity in comparison to one another, but are dissimilar to objects in other clusters. The members of a cluster are as similar as possible whereas the members of different clusters differ as much as possible from each other. Spatial clus tering algorithms utilize spatial relationships among data objects in determining clusters of input data. For ex ample, spatial clustering is used to determine the hot spots in crime analysis and disease tracking. Spatial prediction incorporates spatial dependence to predict events at particular geographic locations.

The spatial autoregressive (SAR) model and Markov random fields (MRFs) are spatial data mining techniques of spatial prediction. For example, habitat of endangered species or water pollution sites can be identified by location prediction models. Spatial characterization is to find a compact description for selected subset of the database. To obtain a spatial characterization, not only the properties of the target objects are considered, but also the properties of their neighbors up to a given maximum number of edges in the neighborhood graph. The task of spatial characterization is to discover the set of all properties for which the relative frequency in a set targets, extended by its neighbors, is significantly differ ent from the relative frequency. A geographic database may be used by economic geographers to discover spatial characterizations of the economic power or other inter esting properties of communities. Some nonspatial at tribute such as the unemployment rate is chosen as a relevant indicator. Spatial association is based on detailed descriptions of their formal properties and background knowledge. A co location pattern is a type of spatial as sociation where a group of spatial features or events are frequently co located in the same region. The underlying space of a co location pattern is based on continuous space rather than discrete sets. A typical example is that human cases of West Nile virus often occur in regions with poor mosquito control and the presence of birds.

Spatial trend detection identifies a regular change of one or more nonspatial attributes when moving away from a start object. It can be used for local and global trends. To detect and explain such spatial trends is an important issue because the identification of spatial outliers or discontinuities can lead to the discovery of unexpected knowledge. Creating a variogram by plotting attribute difference and distance for each pair of points is a technique for spatial outliers. Spatial classification assigns an object to a class from a given set of classes based on the attribute values of the object. It mainly considers the distance, direction, or connectivity relationships among spatial objects.

Geovisualization Software

Since geovisualization in spatial data mining is such a new emerging field, many research prototypes and commercial software are being developed. GeoVISTA Studio supports independent development of data analysis, including exploratory statistical and visualization components that can be combined in a highly coordinated manner using a flexible visual interface. GeoDa is highly interactive and extends the principle of dynamically linked windows (observations highlighted in one graph or map being simultaneously highlighted in all) to incorporate the notions of global and local spatial and space–time autocorrelation from spatial statistics. This allows the investigation of clusters and outliers in geospatial datasets, a powerful collection of methods with applications ranging from criminology, epidemiology, to public health and environmental analysis, with an important outreach component. The software includes these techniques in addition to standard exploratory data analysis tools, such as scatter plots, parallel coordinate plots, and three dimensional (3 D) scatter plots.

There are several spatial data mining software available providing geovisualization functionalities. S Plus SpatialStats is the first comprehensive, object oriented software package for the analysis of spatial data. It was created specifically for the exploration and modeling of spatially correlated data. Its major functionality includes visualization/exploratory data analysis, analyzing geostatistical data (e.g., kriging and variogram), analyzing lattice data (e.g., spatial regression models), and analyzing spatial point patterns (e.g., *K* function). There are many modules of spatial analysis available for R. The modules of spatial, splancs, and spatstat can assist spatial prediction (e.g., kriging) or spatial co location pattern mining (e.g., cross k function). Spatial statistics toolbox for Matlab/Fortran and spatial econometrics library for Matlab compute simultaneous and conditional spatial autoregressions and mixed regressive spatially autoregressive models. The ClusterSeer/BoundarySeeer/SpaceStat by TeraSeer software support spatial clustering, spatial autocorrelation analysis, *k* function, and classification. Neighborhood expectation–maximization (NEM) clusters a given dataset, using the NEM algorithm. GeoMiner is a prototype of a spatial data mining system, including a spatial database server. SPIN! represents the web based integration of spatial data mining and geovisualization for such applications as public health, environmental protection, seismology, or marketing. It includes live Oracle based queries and data visualization. The GIS software such as ArcGIS, Intergraph GeoMedia, and MapInfo Professional contains a full set of exploratory data analysis tools for spatial data, and these tools could be used to enhance spatial data mining.

Spatial Uncertainty in Geovisualization

The notion of uncertainty has been associated with geospatial data. The spatial uncertainty typically came from errors in data collection or display. Little research has been focused on eliminating the spatial uncertainty. It is important to note that the uncertainty is not in data quality but in the qualitative nature of some spatial data and descriptions. The uncertainty of qualitative descriptions can be caused by either one of two problems or both: multiplicity and vagueness. Multiplicity implies that more than one target location fits the qualitative description in question. Results of spatial data mining can return multiple suitable target sites. Vagueness implies that there are unclear boundaries to delineate an appropriate target site.

The challenge to geovisualization is to deal with these problems of spatial uncertainty and researchers have found a variety of ways to do just that. For the problem of multiplicity, geovisualizations can employ proportional symbol mapping to visually depict the suitability of each target site to the initial query. Suitability can be indicated by symbol size or color. For the problem of vagueness a methodology to create a discrete target area that can be identified by name as to whether or not they fall within the region in the initial query is necessary. If a query suffers from both multiplicity and vagueness the user can create a fuzzy 3 D surface where multiplicity is addressed by suitability values being represented by height (z value) and vagueness can be addressed by an overlaid fog map.

Geovisualization for Spatial Data Mining

Geovisualization as a way of spatial data mining is the process of discovering interesting and previously unknown but potentially useful patterns from large spatial datasets. These patterns are valid, novel, useful, and understandable. However, finding knowledge constructive patterns from spatial datasets is not simple due to the complexity of spatial data types, spatial relationships, and spatial autocorrelation. The power of spatial data mining is that it truly facilitates the discovery of new knowledge from new or existing data. The spatial relationships detected through geovisualization are rarely known prior to analysis and are also not always obvious.

Geovisualization focuses primarily on mapping spatial attributes and employing the art of cartography. It is increasingly dealing with the exploration of patterns and relationships in such large datasets for understanding underlying geographical processes. At the same time, spatial data mining focuses on charting and graphing nonspatial attributes using statistical methods. It also expands the number of attributes for further analysis by deriving new attributes. Geovisualization is able to dis play spatial patterns in the distribution of large datasets while spatial data mining is able to detect spatial asso ciations in data in the form of spatial clusters, spatial relationships, and regularities in data. While geovisuali zation can be used as a basic visual analytical tool, it is not suited for in depth quantitative analysis that spatial data mining can provide. The benefit of geovisualization as one of the means for spatial data mining in knowledge discovery is that it can be used to find spatial relationship patterns in data that would otherwise go unnoticed. While geovisualization is intended for unspecialized end users, spatial data mining is more challenging and com putationally intensive often using complex algorithms to determine the relationships between individual data fields in a larger dataset.

The process of knowledge discovery in spatial data relies on geovisualization as a means for spatial data mining. Geovisualization is required to display spatial patterns in large datasets. Many of these patterns can be made more obvious by spatial data mining. Where spatial data mining is strong at achieving detailed statistical analysis of data, geovisualization is weak at displaying and communicating these relationships. Thus, geovisua lization and spatial data mining complement each other to discover hidden knowledge in large spatial datasets. With the increase in spatial data sources anticipated over the next decade it is clear that data managing method ologies will play an increasingly important role in spatial data interpretation. They will continue to grow as geo spatial data expand and new techniques are developed.

Future Research

Research needs in geovisualization are extensive. Geo visulation in spatiotemporal databases is a promising area for future research. The explosive growth of spatial data and extensive use of spatial databases emphasize the need for the automated discovery of spatial knowledge. Other noticeable areas include geovisualization in remote sensing data and spatial object oriented databases, and parallel algorithms and distributed infrastructures for spatial data mining.

Effective geovisualizations of spatial relationships are also important areas. Geovisualization is useful to iden tify interesting spatial patterns. The data inputs of spatial data mining have both spatial and nonspatial features. The complexity of spatial data and intrinsic spatial relationships limits the usefulness of conventional data mining techniques for extracting spatial patterns. To facilitate the geovisualization of spatial relationships, research is needed on ways to represent both spatial and nonspatial features. For example, many visual represen tations have been proposed for spatial outliers. However, a way to emphasize spatial outliers within geovisualiza tions of spatial relationships is not fully studied yet. For instance, the spatial relationship between a single spatial outlier and its neighbors is not obvious in vario gram cloud and scatterplot visualizations. It is essential to convey the information back to the original map in geographic space to ensure neighbor relationships. Another important area to be improved is efficient computation. Mining spatial patterns is often compu tationally expensive. For example, the estimation of the parameters for the spatial autoregression model is more complicated than that for the linear regression in classical data mining. Research is needed to reduce the compu tational costs of spatial data mining algorithms.

The areas of geovisualization in spatial data mining are very wide ranging. Its technology can be applied practically to any domain. Although geovisualization and spatial data mining have been used as separate technol ogies, there are numerous areas where they can be integrated. Spatial data can be passed easily if geovisua lization and spatial data mining operate under the same system. One of the attempts in geovisualization research has been to use artificial neural networks as a technology especially useful in situations where the numbers are vast and the relationships are often unclear or even hidden. Geocomputation is another area on the geo subjects trying to find new or better solutions to existing problems via a computational route.

See also: Exploratory Spatial Data Analysis; Geovisualization; Spatial Autocorrelation; Spatial Data Mining, Cluster and Pattern Recognition.

Further Reading

Foody, G. M. (2003). Uncertainty, knowledge discovery and data mining in GIS. *Progress in Physical Geography* 27(1), 113 121.

Gahegan, M., Wachowicz, M., Harrower, M. and Rhyne, T. M. (2001). The integration of geographic visualization with knowledge discovery in databases and geocomputation. *Cartography and Geographic Information Systems* 28, 29 44.

Huang, Y., Pei, J. and Xiong, H. (2006). Mining co location patterns with rare events from spatial data sets. *GeoInformatica* 10(3), 239 260.

Jiang, B. and Li, Z. (2005). Geovisualization: Design, enhanced visual tools and applications. *The Cartographic Journal* 42(1), 3 4.

Keim, D. A. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics* 7, 100 107.

Kraak, M. J. and MacEachren, A. (1999). Visualization for exploration of spatial data. *International Journal of Geographical Information Science* 13(4), 285 287.

Kwan, M. P. (2000). Interactive geovisualization of activity travel patterns using three dimensional geographical information systems: A methodological exploration with a large data set. *Transportation Research C* 8, 185 203.

Kwan, M. P. and Lee, J. (2004). Geovisualization of human activity patterns using 3D GIS: A time geographic approach. In Goodchild, M. F. & Janelle, D. G. (eds.) *Spatially Integrated Social Science*, pp 48 66. New York: Oxford University Press.

MacEachren, A. M., Gahegan, M., Pike, W. *et al*. (2004). Geovisualization for knowledge construction and decision support. *Computer Graphics and Applications* 24(1), 13 17.

Miller, H. J. and Han, J. (2001). *Geographic Data Mining and Knowledge Discovery*. London: Taylor and Francis.

Shekhar, S., Zhang, P., Huang, Y. and Vatsavai, R. (2004). Trends in spatial data mining. In Kargupta, H. & Joshi, A. (eds.) *Data Mining: Next Generation Challenges and Future Directions*, pp 4 192. Cambridge, MA: AAAI/MIT Press.

Wang, W., Yang, J. and Muntz, R. (2000). An approach to active spatial data mining based on statistical information. *IEEE Transactions on Knowledge and Data Engineering* 12(5), 715 728.

Yao, X. and Jiang, B. (2005). Visualization of qualitative locations in geographic information systems. *Cartography and Geographic Information Science* 32(4), 219 229.

Relevant Websites

http://www.buzztracker.org
Buzztracker World News, Mapped.

http://kartoweb.itc.nl
Cartography Netherlands.

http://www.geoda.uiuc.edu
GeoDa an introduction to spatial data analysis.

http://www.geovista.psu.edu
GeoVISTA Center, Penn State University.

Spatial Data Models

L. Bian, University at Buffalo, State University of New York, Amherst, NY, USA

Glossary

Bona Fide Boundary Genuine discontinuities in the world, such as the boundary between a lake and land.

Fiat Boundary Boundary that does not exist physically, but can be precisely placed in space, such as an administrative boundary.

Field Spatial phenomenon that is spatially extended, and its attributes and processes continuously vary across space.

Object-Orientation A computing paradigm that represents phenomena as objects. Objects have identification, properties, and behavior, and they can be organized into a hierarchy of classes.

Spatial Data Model A model to discretize a space into discrete geometric units, such as points, lines, polygons, or grid cells.

Spatial Object Spatial phenomenon that has discrete boundaries, properties, and processes, and can be mobile or movable.

Spatial Region Spatial phenomenon that has not only the characteristics of a spatial object but is also a part of a continuous field.

Introduction

Spatial data models are a formal representation of space. Because the computing environment is finite and discrete, a geographic space must be partitioned into a finite number of discrete pieces to accommodate the computing environment. Spatial data models describe the design of the discretization and the relationship between the discretized pieces. Spatial data models are one of the most fundamental concepts in geographic information systems (GIS). They appear in every GIS textbook, GIS software packages, academic research, and professional applications.

Three types of information are associated with a spatial data model, location, attributes, and topology. A location is represented by a pair of coordinates in east and north directions. Attributes describe properties of partitioned pieces. Topology, on the other hand, records the spatial relationship between these pieces. Two GIS data models – the vector data model and the raster data model – are commonly used. These are two fundamentally different data models, representing two different ways of partitioning the space.

Vector models partition the space according to the form and function of spatial features, and represent them as geometric forms such as points, lines, and polygons. The implementation of a vector model explicitly records the coordinates of points, turning points of lines, and outlines of polygons. Each point, line, or polygon can have multiple attributes. Because of a clear identification of the spatial features, the vector data model can explicitly represent topology between them, such as which polygonal features are on either side of a line feature or which point features are at either end of a line. Vector data models are advantageous for phenomena that require a clear identification of spatial features, the spatial relationships between them, and a rich set of attributes associated with each. This is especially true for those spatial features that are perceived to be homogeneous and stable.

Raster models partition the space into a matrix of regular cells. Because of the regular geometry (shape and size) of raster models, the coordinates of a cell are implicitly referenced by its row and column positions. Attribute data are explicitly represented as a single value for each cell. The adjacency between cells in the four cardinal and the four diagonal directions is clear. Topology between spatial features, which is usually formed by aggregates of cells, is not recorded because single cells, instead of spatial features, are the basic unit of the raster model. Raster models are advantageous for phenomena that are heterogeneous and dynamic. The explicit adjacency between cells is particularly effective for supporting localized simulations that often query information about areas adjacent to a local cell.

Spatial Data Models for Human Geography

Human geography is concerned with the spatial differentiation and organization of human activities and with human use of the physical environment. Modeling and analysis in human geography attempt to represent and analyze these activities and the environments. In a generic sense, a model is a representation of certain aspect of reality. Spatial data models represent the spatial aspect of the real world in an abstract way. Unlike those models that directly describe human activities and the human use of environment, spatial data models focus rather on the spatial representation of these activities, such as geometry and topology. Spatial data models are often implemented in spatial databases that provide tools to manipulate the data. These tools are mostly for basic spatial operations, but not for the actual modeling and analysis.

The spatial databases are built as generic tools. There is no information in them to describe what phenomena that these data models should represent. The separation of the spatial data models from the phenomena that they represent allows the spatial databases to be flexible in order to support a wide range of applications. This separation, in the mean time, makes it critical for users to choose data models appropriate for intended modeling and analysis. There is always more than one spatial data model that can be used to represent the same phenomenon. The choice of the model can have significant impact on the modeling and analysis and the subsequent outcome. An appropriate choice depends on many factors, but the foremost is how a phenomenon is perceived and how it subsequently should be represented.

Spatial Representation

Levels of Representation

Three levels of abstraction are involved in the representation of space, conceptual models, spatial data models, and spatial data structures. Conceptual models describe the perceived space. Their purpose is to identify essential spatial components and the relationships between them. For example, a population can be conceptualized as either individuals at various locations, subpopulations within administrative regions, or population density varying continuously in space. With the components identified, the relationships refer to those between locations, between regions, and within a continuous space. The second level of the abstraction, the spatial data model, formalizes the perceived spatial components and relationships in terms of basic geometry (e.g., points, lines, polygons, and grids) and topology. In turn, the spatial data models guide the implementation of these components and relationships at the third level, spatial data structures. This level is concerned with how to implement various data models in a computer.

Similar multiple levels of abstraction have been proposed for spatial representation. These include the conceptualization, function, and implementation levels, and the scientific, logical, and physical models. The three levels of abstraction represent three distinct viewpoints, that of an observer of the world, a software specifier, and a software implementer.

Object and Field Views

In perceiving the space (the conceptual model), the object–field dichotomy has offered a profound framework for the conceptual representation of spatial phenomena. With the object view, the reality is perceived as independent objects littered in space (hereafter referred to as 'spatial objects'). In contrast, the field view perceives the reality as a continuous plane with variations at an infinite number of locations. The identification of spatial objects in contrast with the identification of fields has been discussed as two opposing conceptualizations of space.

In spatial investigations, certain phenomena are easily perceived as discrete spatial objects and others as continuous fields. A continuous phenomenon can be divided indefinitely without changing its essential nature, while a discrete phenomenon most likely cannot be divided without altering its identity. For example, the water in a bucket may be continually halved and yet remain water, but half a bucket is no longer a bucket. The object and field views are not exclusive. A phenomenon can be conceptualized as either a spatial object or a field depending on the spatial scale of the observation, the purpose of an investigation, and convention. The two conceptualizations can coexist as well.

In addition to easily recognizable spatial objects and fields, spatial regions have received substantial attention in the literature. This is because spatial regions have dual qualities. They appear to be discrete as spatial objects, but are not quite as independent because they are always parts of a continuous field. The characteristics of spatial objects, spatial regions, and fields have been actively discussed in the geographic information science literature.

The two spatial data models, vector and raster, are not directly paired with spatial objects and fields, respectively, or spatial regions. It is most important to first understand the characteristics of spatial phenomena and, based on this understanding, select the most appropriate data models for the intended modeling and analysis. The discussion below summarizes the characteristics of spatial objects, spatial regions, and fields.

Identifying Spatial Objects, Regions, and Fields

Spatial Objects and Regions

Many characteristics have been used as criteria to identify spatial objects and object like spatial regions. These can be synthesized into five categories, namely, spatial scale, boundary, attributes, process, and mobility.

The spatial scale criterion uses human body as the reference scale and human experience as the basis to categorize spatial phenomena into four types. The first type of spatial phenomena includes everyday tabletop objects that are smaller than the human body and can be moved or manipulated. The second type is an extension of the small objects into large scale space. This category includes phenomena that are larger than the human body, but perceived as objects, such as buildings. The third type refers to spatial phenomena in large scale space, such as an urban landscape, that cannot be experienced com pletely all at once. These phenomena are perceived as fields. The fourth type is an extension of fields and refers to those large scale phenomena that are beyond the range of direct human experience, such as space beyond the earth. The spatial scale criterion can be readily used to identify those spatial phenomena that are easily per ceived as spatial objects, either small or large.

Boundary is another criterion that has been used to identify spatial objects and regions. From an onto logical perspective, boundaries are as essential as the internal content to the ontological make up of a spatial object. Both the aforementioned small space and large space objects tend to have well defined boundaries. The boundary criterion has also been used to identify spatial regions extracted from continuous space. With bound aries, these spatial regions tend to be perceived as objects, as if they were small scale objects. Boundaries for spatial objects or regions have been categorized into three types according to their origin. The first type, bona fide boundaries, includes those corresponding to genuine discontinuities in the world, such as the boundary between a lake and land. The second type is called fiat boundaries. It includes those that do not exist physically, but can be precisely placed in space, often as a reflection of human intention. The most typical example of this type is administrative boundaries. The third type refers to those that do not exist physically, yet can be placed in space, but not precisely. For example, the boundary between 'Midwest' and 'Northeast' is such an imprecise boundary. With boundaries, spatial objects can be readily identified and spatial regions be perceived as objects.

Attributes are a third criterion. From the ontological viewpoint, things are known to the world and dis tinguished from one another by their unique properties. Spatial objects can be formally described as $<i, a_1, a_2, \ldots, a_m>$, where i is an object and a_1 through a_m are attributes of the object. Because spatial objects are often identified based on their boundaries, the importance of attributes is most evident when identifying spatial regions. These spatial objects and spatial regions are identified by the magnitude and spatial distribution of attribute values. These include, but are not limited to, the magnitude or threshold of attribute values (e.g., pollution zones), the homogeneity or dominance of attribute values (e.g., land use and land cover zones), or the spatial association of attribute prototypes (e.g., cities composed of residential, commercial, and industrial areas).

Process, the fourth criterion, is interpreted inter changeably with concepts such as activity, operation, and function. Process may change the state of a phenomenon, and all things change. The information about process is essential in the identification of spatial objects and regions. This criterion is particularly important for human geography because, by definition, human geog raphy is concerned with human activities. Similar to attribute, the importance of process is more evident in the delineation of spatial regions than that of spatial objects, because the latter can be readily identified by their boundaries. There are plenty of examples of spatial regions that are delineated by various social, economic, legal, political, and cultural processes, such as adminis trative regions, electoral districts, and traffic generation zones. The magnitude, dominance, and direction of processes are often used to identify spatial objects and spatial regions. Because process can be represented as attribute change, the attribute and process criteria are often paired together.

Mobility has often been used as a criterion to identify spatial objects in small space or large space. Because mobility implies the independence of objects from locations, it is a criterion in its own right, although it can be considered as a special case of the process criterion or a spatially explicit attribute criterion. While moving, objects maintain their identity, properties, and behavior. Many spatial objects in human geography studies can be mobile or can be moved. When an object, such as a vehicle, moves, all parts of the object move together, along with its structure and function. Spatial regions can also be mobile, but it is its attributes and process that move, not its parts. A close analogy is the movement of 'wave' in a sport stadium. By a sequence of vertical motion of audiences (standing up and sitting down), a wave moves around the stadium, while its parts (audi ences) do not change location. This 'pseudo' movement of spatial regions has many implications in selecting spatial data models.

In summary, spatial objects and spatial regions can be identified through any number or combination of the aforementioned criteria. The genuine spatial objects exist in both small space and large space, have discrete boundaries, properties, and processes, and are mobile (or movable). Spatial regions are extracted out of continuous space, mostly in large space. They have definable, but nonexistent boundaries, and can be mobile. Properties and process are most important for the identification of spatial regions. Both spatial objects and spatial regions can be conceptualized as objects, although spatial regions carry dual qualifications.

Fields

As opposed to objects, fields are spatially continuous. By this definition, fields are viewed as a mapping between attributes and an indefinite number of spatial locations. The formal model of fields can be expressed as $<x, y, z_1, z_2, \ldots, z_m>$, where x, y are continuous locations, and the attributes at the location x, y are represented by the set z_1 through z_m. Fields can be categorized in several types, including scalar, vector, and tensor fields. A scalar field presents a scalar value of an attribute for each location (e.g., population count or density). A vector field identifies the direction and magnitude of a phenomenon at a location (e.g., population migration). A tensor field represents strains at multiple directions through a matrix at every location (e.g., direction of population migration).

Many researchers have sought to identify the characteristics of fields. Although fundamentally different from objects, fields can also be described using the criteria discussed earlier, namely, spatial scale, boundary, attributes, process, and mobility. Fields are most often in large scale space. They are spatially extended without a boundary or the boundary is not a concern. Each location in a field is associated with attributes and processes, and these attributes and processes vary continuously across the field. The motion of parts or the entire field is represented through the movement of attributes and process. These characteristics, that is, the large scale, no boundaries, the continuous change of attributes and processes, and the pseudo mobility, fundamentally distinguish fields from spatial objects.

Spatial regions, as parts of a field and often conceptualized as objects, do not comply fully with the continuous definition of fields or the discrete definition of objects. The dual qualification of spatial regions presents challenges, as well as flexibilities in their representation.

Networks are a special, one dimensional field embedded in a two dimensional space. It is considered a field because it is continuous, though in a linear form. The variation of its attributes occurs only on the network, not in the intervening spaces. The formal model of fields, $<x, y, z_1, z_2, \ldots, z_m>$ is applicable to networks, where x, y identify an infinite number of locations constrained to this linear field. The attributes represented by z_1, z_2, …, z_m can be any of the types for two dimensional fields. Linear segments extracted from a network are equivalent to those spatial regions in a two dimensional field.

Spatial Objects, Regions, and Fields in Human Geography

Human geography deals with a variety of spatial phenomena. Some of these can be easily thought as spatial objects, some as typical fields, and others as spatial regions extracted out of continuous fields. Using the five criteria discussed above, namely, scale, boundary, attributes, process, and mobility, these spatial phenomena can be categorized into several types of spatial objects, spatial regions, and fields. The identification of these categories helps the appropriate selection of spatial data models in human geography studies.

1. Individuals: These individuals exist in both small and large scale space, have clear boundaries, attributes, and processes. These individuals can be mobile or sedentary. Mobile individuals are independent of spatial locations, while the sedentary individuals are bound to their locations. The most typical examples of mobile individuals are individual humans or vehicles, while the typical examples of sedentary individuals can include individual buildings in a city. The unique identity, attributes, and behavior (including the mobility) of individuals in this category are of great interest in modeling and analysis. These individuals, the mobile and the sedentary, are normally conceptualized as objects. Both types are fundamental subjects of many human geography studies.
2. Masses of individuals: This category is an extension of the category of individuals. As the constituents of the mass, these individuals are identifiable, but are small in size or large in quantity, or both. These constituents can be mobile or sedentary. The most common examples of the masses of mobile individuals include a human population composed of mobile individuals, or a traffic flow composed of individual vehicles. Examples for the masses of sedentary individuals include urban landscape composed of individual buildings. The collective attributes and behavior, the continuous distribution, and the dynamic or sedentary nature of the masses are important in modeling and analysis. Whether a mass is dynamic or sedentary is relative to the time step and duration of a study.
3. Regions of individuals: This category is an extension of the masses of individuals or another extension of the individuals. These are spatial regions extracted from the continua in the masses of individuals. A region can be mobile or sedentary, dependent upon whether the location of the region as a whole is mobile or not, and independent of whether the constituents of the regions are mobile. Electoral districts extracted from the spatial distribution of population and residential areas extracted from an urban landscape are two examples of sedentary regions. Examples of mobile regions include the spread of disease and urban sprawl. The unique identity, attributes, and behavior (or function) of the regions are of great importance in modeling and analysis. Regions are often conceptualized as objects. Whether a region is dynamic or sedentary depends on the temporal scale of a study.
4. Continuous masses: Phenomena of this type are spatially extended, continuous, and spatially varying. The attributes and function of this type of phenomena vary across space, and the mass may be dynamic or sedentary. Land and roads are examples of sedentary masses in two and one dimensional space, respectively. Population density and air pollution can be examples of the dynamic masses (note that the population density refers to a continuous surface expressed as a rate, while the population example

used in the masses of individuals refers to a collection of discrete individuals). The continuous form and the dynamic or sedentary nature of the attributes and function of this type are important to modeling and analysis. Both continuous sedentary mass and continuous dynamic mass are typically conceptualized as fields.

5. Regions in mass: This category is an extension of the continuous masses as the regions are extracted out of the continua. Similar to regions of individuals, the constituents of the regions in mass can be mobile or sedentary, but the mobility of these regions depends on the location of the region as a whole. Land use and land cover areas extracted from the continuous land and high accident sections of a road are examples of sedentary regions in mass. The population density of a certain workforce and plumes of air pollution are examples of dynamic regions in mass. Both sedentary and mobile regions are often conceptualized as ob jects. The unique attributes and behavior (or func tions) of these regions are important for modeling and analysis.

The five categories discussed above represent a spectrum of spatial continuity with the most typical objects on one end and the most typical fields on the other, converging to spatial regions in the middle. The first three types begin with phenomena that are perceived as spatial objects and then aggregated into spatial regions and fields. The other two types begin with continuous fields and are then discretized into spatial regions. Spatial phenomena in human geography can be placed anywhere along this spectrum of spatial continuity, not necessarily in distinct categories. A same phenomenon may qualify in a number of categories depending on the objective and scale of a study, as well as the convention in a dis cipline. For example, a population can be perceived as either mobile or sedentary individuals, a mass of mobile or sedentary individuals, regions of individuals, a con tinuous mass if presented as population density, either dynamic or sedentary, or regions in the mass of popu lation density.

The examples used in aforementioned categories focus on ‘things’, such as humans, vehicles, buildings, land, roads, etc. For human geography that deals with human activities, it is often the activities that are per ceived as objects, regions, or fields. Examples of the activities may include crime incidents, cases of disease, the flow of information, the transportation of goods, the migration of populations, and the dispersion of pollutants. In these situations, the activities are perceived as the spatial features, and spatial data models are used to register the spatial distribution of these activities. Regardless of things or activities, the same criteria apply in the identification of types of objects, regions, and fields.

The identification of these spatial objects, spatial regions, and fields helps select the appropriate spatial data models for the intended investigation. The section below examines spatial data models implemented in GIS and their compatibility with the representation of spatial objects, regions, and fields.

GIS Data Models for Spatial Objects, Regions, and Fields

Object-Oriented Implementation of Data Models

Since the 1990s, object orientation has become the computing paradigm and software industry standard. GIS data models have been implemented in the object orientation environment since the later part of the 1990s.

The object orientation approach consists of a number of basic principles. Of these, the encapsulation and composition principles tend to be considered conceptual, while others are implementation concerns. The en capsulation concept states that each object has an iden tity, properties, which are represented as attributes, and behavior, which is represented as methods. Attribute values describe the states of the object, whereas methods can change the state. The identity, properties, and behavior are encapsulated within the object. The com position principle states that objects can be organized into a class hierarchy. Objects in a subclass inherit properties and behavior from its superclass, in addition to their own unique properties and behavior. Objects can also be organized in aggregations and associations. An object is an integral part of the aggregated whole or a member of the associated set.

Extensive research has been conducted in the design of object oriented GIS databases. The object oriented version of spatial data models has adopted the same space partition as in the traditional data models. Geometric primitives, such as points, lines, polygons, their deriva tives in the vector data models, and grids in the raster data models are implemented as software objects. These software objects are encapsulated with attributes and functions and are organized into various hierarchies, associations, and aggregates.

Despite the identical geometry, the object oriented vector models differ from their traditional equivalents. In traditional vector models, data are organized based on location, expressed as coordinates of spatial features. Attributes and topology are attached to the location. Any change in the location requires that the coordinates be updated and the attributes and topology be reestablished. This design makes the representation of movement difficult, thus limiting the usefulness of the vector data model, especially for mobile objects and mobile regions.

With the object oriented implementation, spatial features, represented as software objects, are built on their identifiers. Location is no longer the basis for organizing spatial data. Instead, it is treated as an attri bute, such as the 'shape' attribute in ArcGIS. With this treatment, changes in location can be easily implemented by updating the location attribute, while the identity, other spatial and aspatial attributes, and functions of the object remain intact. This change is one of the major improvements of object oriented vector models over the traditional ones and has significantly increased the use fulness of GIS in many applications.

The differences between the traditional version of raster data models and their object oriented version are less prominent than those for vector data models. Object oriented implementation of the raster data model not only keeps the same spatial partition as in the traditional one, but also maintains the basic form of array of cells. The cells may be aggregated into different levels of blocks according to the needed generality. Each block can be treated as a software object. Note that it is the entire array of cells or a block of cells, rather than an individual cell, which is implemented as a software object. Occa sionally, individual researchers may treat cells as objects in their own in house object oriented GIS applications. While technically feasible, the conceptual advantage of this practice (individual cell objects) is, however, subject to debate.

Both vector and raster data models can support the representation of the five types of spatial objects, spatial regions, and fields, though to various degrees. The appropriate choice of spatial data models depends on the compatibility between a data model and the conceptual model of phenomena, that is, spatial objects, regions, and fields.

Spatial Data Models for Spatial Objects

Spatial objects include the category of individuals dis cussed earlier. The vector representation for spatial objects is rather straightforward. The discrete form and the use of spatial features (points, lines, polygons, or their derivatives) as the basic unit (as opposed to grid cells in rater data models) make vector data models inherently compatible with spatial objects. This allows vector data models to readily support the representation of scale, boundary, attributes, processes, and mobility of spatial objects.

Attributes of spatial objects can be internal (geo metric, socioeconomic, etc.), background environment, and spatiotemporal. Vector data models can readily support them. The representation of the scale and boundary of spatial objects can be accommodated by internal attributes, such as shape. The mobility of an object can be supported by updating the spatial and temporal attributes, as discussed earlier. The processes of spatial objects may include the actions of individuals, the interactions between them, and their interactions with the background environment. Because spatial databases mostly support spatially oriented processes, additional tools may be required to support the representation of some of the other spatial and nonspatial processes.

For raster data models, a single cell or an assemblage of connected cells may be used to represent a spatial object, depending on its scale and geometry. Because the basic unit of raster data models is a cell, not a spatial feature, the representation of boundary is implicit. For attributes and processes, their representation is, in gen eral, straightforward.

Because a raster database does not have the built in ability to identify a spatial object, for processes that require spatial objects as the basis, additional program ming tools are normally needed. It is seldom a challenge for raster data models to support these tools. The mobility of spatial objects, one of the typical processes, can be represented as a sequential change of attribute for a cell or an assemblage of cells along the path of movement.

Raster data models are the simplest way to partition a continuous and complex space into a finite number of discrete pieces. Strictly speaking, raster data models are not entirely compatible with spatial objects conceptually, and a square cell does not resemble most spatial features. It is the simple form and regular arrangement of cells that provide raster data models with the flexibility to support the representation of spatial objects.

Spatial Data Models for Spatial Regions

Spatial regions include two categories, the regions of individuals and regions in mass, as discussed earlier. Spatial regions are most likely to be represented as lines or polygons in vector data models. Because a spatial region can be perceived as a spatial object, as opposed to a field, the vector data models are conceptually com patible with its representation. The vector representation can support the scale, boundary, attributes, and processes associated with spatial regions, as would they for spatial objects. Unlike spatial objects that are independent of their background environment, spatial regions are bound to the field from which they are extracted. To represent their mobility, it is necessary not only to update the location of a mobile region, but also the rest of field needs to be updated in order to fill the 'hole' left by the region.

Raster representation of spatial regions is similar to that for spatial objects. An assemblage of cells, rather than a single cell, is most likely to be used to spatially approximate a region. Because of the inherent advantages in representing heterogeneous and dynamic phenomena, it is easier for raster data models to represent the mobility of both spatial objects and regions than vector models.

Because of the dual quality of spatial regions, that is, they are perceived as discrete spatial objects, but are always parts of a continuous field, a number of repre sentation strategies have been proposed to accommo date this dual quality. These strategies use the raster data model for the internal variation of a region, and integrate a vector data model to support the identity of the region.

Spatial Data Models for Fields

Fields include the categories of masses of individuals and continuous masses. Both vector and raster data models can be conceptually compatible with a field and, thus, able to support the representation of their characteristics. These characteristics include the spatial continuity, the spatial variation of attributes and processes, and the dynamics of these attributes and process. Six data models have been discussed specifically for the representation of fields. These are polygons, triangulated irregular networks (TINs), contours, cell grids, point grids, and irregular points. Of these, polygons, TINs, contours, and irregular points are vector data models, while cell grids and point grids are raster models.

Of the four vector models, the polygon and TIN models partition space into contiguous regions with explicit boundaries. They support the representation of those fields that consist of discrete spatial regions. Each individual polygon may be perceived as a spatial object, but collectively, they are parts of a continuous field. The polygon model can readily support the representation of the spatial continuity of a field, and the spatial variation of attributes, processes, and their dynamics, as it would for spatial regions. For fields with a linear form, such as networks consisting of discrete linear spatial regions, the same principles apply. The TIN model can be considered a special kind of polygon model. The delineation of tri angles in a TIN is based on the locations of prominent attribute values, such as value peaks, ridges, passes, and valleys. The resultant triangles contain homogeneous values in slope and aspect and linearly varying values in elevation. TINs, however, are not used frequently in human geography.

Contours and irregular points, though in a vector form, are intended to represent a continuously varying space, rather than points, linear shaped spatial objects, and regions. Strictly speaking, the contour lines and irregular points do not partition the space. They present attribute values of a field at sampled locations, either along the contour lines or at certain point locations. If needed, attribute values for the rest of the field may be estimated based on those at the sampled locations. These two models can support the representation of spatial continuity, spatial variation of attributes and processes, and their dynamics through a sampled form. The dynamics can be supported by updating the attri butes of the contours and points.

The representation of field can take full advantage of raster data models, including cell grids and point grids. Their simple form, the explicit representation of attri butes, the easy representation of spatial variation, and the support for spatial operations allow the grid models to effectively represent the spatial continuity, spatial variation of attributes and processes, and their dynamics. Due to their inability to identify discrete spatial objects and regions, grid models are most appropriate for con tinuous fields lacking explicit regions. In these fields, individual grid cells are the result of a simple form of spatial partition. Each cell is a part of a continuous field, not any meaningful spatial object or region.

Composite Structures

Both raster and vector data models have been imple mented in a layered structure. Raster data models are inherently constrained to a single layer. For vector data models, an individual layer is determined by both theme (e.g., land ownership, delivery areas, or electoral districts) and geometry (e.g., points, lines, or polygons). For example, an urban landscape may consist of residential, commercial, industrial areas (polygons), streets (lines), and individual humans (points). In many GIS software packages, this landscape must be prepared in three layers, a polygon, a line, and a point layer, respectively, although all belong to a single theme. It is theoretically and technically feasible to develop a 'composite' structure to support the representation of different geometries or themes in a single layer. For the example above, all three types of geometric features can be presented in a single layer to support modeling and analysis. Computing methods, such as object oriented and component based modeling, can support the implementation of the com posite structure. The theoretical design of such com posite structure, on the other hand, has been discussed in the literature only recently.

A multigeometry and multitheme 'object fields' has been developed for environmental management recently. Vector data of various geometries and themes are inte grated with a continuous raster field of a different theme for comprehensive modeling. Furthermore, the nine intersection system is a notable theoretical development to support the complex topology between points, lines, and polygons. The implementation of a composite structure requires sophisticated vector databases. Con tinued theoretical development in the research com munity and the technical development from the GIS software industry are critical for advancing this com posite structure or any other more advanced spatial data models.

See also: First Law of Geography.

Further Reading

Bian, L. (2007). Object oriented representation of environmental phenomena is everything best represented as an object? *Annals of the Association of American Geographers* 97, 266 280.

Burrough, P. A. and Frank, A. U. (eds.) (1996). *Geographic Objects with Indeterminate Boundaries*. London: Taylor & Francis.

Couclelis, H. (1992). People manipulate objects (but cultivate fields): Beyond the raster vector debate in GIS. In Frank, A. U., Campari, I. & Formentini, U. (eds.) *Theories and Methods of Spatio Temporal Reasoning in Geographic Space*, pp 65 77. Berlin: Springer Verlag.

Cova, T. J. and Goodchild, M. F. (2002). Extending geographical representation to include fields of spatial objects. *International Journal of Geographical Information Science* 16, 509 532.

Egenhofer, M. J. and Franzosa, R. (1991). Point set topological spatial relation. *International Journal of Geographical Information Systems* 5, 161 174.

Goodchild, M. F. (1992). Geographical data modeling. *Computers and Geosciences* 18, 401 408.

Kemp, K. K. (1997). Fields as a framework for integrating GIS and environmental process models. Part 1: Representing spatial continuity. *Transactions in GIS* 1, 219 234.

Martin, D. J. (1999). Spatial representation: The social scientist's perspective. In Longley, P. A., Goodchild, M. F., Maguire, D. J. & Rhind, D. W. (eds.) *Geographical Information Systems*, pp 71 80. New York: Wiley.

Mcintosh, J. and Yuan, M. (2005). A framework to enhance semantic flexibility for analysis of distributed phenomena. *International Journal of Geographical Information Science* 19, 999 1018.

Montello, D. R. (2003). Regions in geography: Process and content. In Duckham, M., Goodchild, M. F. & Worboys, M. F. (eds.) *Foundations of Geographic Information Science*, pp 173 189. London: Taylor & Francis.

Peuquet, D. J. (1994). It's about time: A conceptual framework for the representation of temporal dynamics in geographic information systems. *Annals of the Association of American Geographers* 84, 441 461.

Peuquet, D. J., Smith, B. and Brogaad, B. (1999). *The Ontology of Fields, Report of Specialist Meeting Held under the Auspices of the Varenius Project*. Santa Barbara, CA: National Center for Geographic Information and Analysis.

Smith, B. and Mark, D. M. (1998). Ontology and geographic kinds. In Poiker, T. K. & Chrisman, N. (eds.) *Proceedings of International Symposium on Spatial Data Handling (SDH'98)*, pp 308 320. Vancouver, BC: International Geographical Union.

Worboys, M. F. (1994). Object oriented approaches to geo referenced information. *International Journal of Geographical Information Systems* 8, 385 399.

Zubin, D. (1989). Untitled. In Mark, D. M., Frank, A. U., Egenhofer, M. J., Freundschuh, S., McGranaghan, M. & White, R. M. (eds.) *Languages of Spatial Relations: Report on the Specialist Meeting for NCGIA Research Initiative 2, Technical Report 89 2*, pp 13 17. Santa Barbara, CA: National Center for Geographic Information and Analysis.

Spatial Databases

A. C. Winstanley, National University of Ireland, Maynooth, Republic of Ireland

Glossary

Database A collection of interrelated information stored by a computer system.
Database Management System (DBMS) A computer system designed to store, organize, retrieve, and combine information.
Domain A classification of data according to their characteristics, for example, integer numbers and Boolean values.
Relational Model A method of structuring data as a set of tuples or records in relations or tables so that the interdependence of data is modeled.
Spatial Data Data that record the location as well as other attributes of objects.

Introduction

Databases are computer systems designed to store information in a systematic way so that their contents can be easily accessed, managed, changed, and augmented. Spatial databases are such systems designed specifically to include data with spatial attributes, such as geographical location, distance, and extent. The software used to manage and query a database is known as a database management system (DBMS). The most prevalent type of database is the relational database, a tabular scheme in which data are defined in terms of simple data types and operations so that it can be reorganized and accessed in a number of different ways. Most spatial DBMSs extend the relational model to include more complex spatial data types and operations. All geographic information systems (GISs) use a spatial database with, in addition, functions for presentation, visualization, and analysis of the spatially referenced data.

The Relational Data Model

The relational data model was first defined by E. F. Codd in 1970 and is based round the notion of data being stored in a series of interrelated relations of data items. A relation is a mathematical notion of groups of data items whose order is not significant. They can be represented as a regular table, with each group represented by one tuple or row, the individual data items occupying cells in columns. In a database, each relation/table holds data on one kind of object, each row representing one object of this kind and each cell one attribute. Each cell holds an atomic datum, which cannot be meaningfully subdivided into smaller items. A relational database is a collection of these tables, along with queries, forms, reports, and macros, all stored in a computer program and all of which are interrelated.

As an example, **Table 1** shows a relation representing commercial businesses. Each row represents one business, each column an attribute to describe a business, and each cell contains an indivisible data item. All values in each column are of the same simple domain or type, for example, integer number, string, or Boolean. According to the relational model, the order of rows and columns has no significance and relations with similar structures can be combined using set operations (e.g., union and intersection). Each relation has a key, one or more attributes whose value uniquely identifies each row.

A database will typically consist of several relations, each representing a different kind of entity. For example, the business database might also include an owner table containing details of the people who own businesses (**Table 2**). The key attribute here could be that holding the name of each person. Each row also contains a value from the key attribute of the businesses table, a foreign key, which records the business owned.

The strength of the relational data model is its rigorous foundation in the mathematical notion of relations which has properties proven to make it efficient and flexible. This includes the ability to normalize a set of relations to minimize data redundancy and optimize

Table 1 A database relation representing businesses

Id	*Name*	*Phone*	*City*
1	Conservation Products Ltd.	472846435	London
2	Timecare	547623810	Birmingham
3	Design Ltd.	657347384	Belfast
4	Archive Books	657839210	Dublin
5	Tom Meek Artist	648390184	New York
6	Northern Music	475627439	Liverpool
7	Technologia	484756282	Manchester
8	Gen Supplies	968567595	London
9	Arcare	947465484	London

Table 2 A database relation representing owners of businesses

Id	*Surname*	*First name*	*Busid*
101	Smith	John	9
102	Jones	Alan	2
103	Murphy	Patrick	4

them for querying. Notations can be devised to specify queries as a set of combinations, manipulations, and ex tractions of values from relations. Mathematical methods, such as relational calculus and relational algebra, form the basis of computer based notations of which the most common is Structured Query Language (SQL). Queries can be reduced to a series of primitive operations on relations: selection of specific rows of a relation based on particular values of attributes or projection of specific columns from a relation. The most powerful operations consist of joining two relations together based on com mon values in key fields – the join. For example, to find the name of whoever owns a particular business we can join together the relations in **Tables 1** and **2** using values in the key field, select the row containing the business name, and project out the column containing the owner name. This is expressed in SQL as:

```
SELECT owner.name
FROM business, owner
WHERE business.id = owner.busid and business.name
= Archive Books
```

Queries are the most common way to interrogate a database. They create new, usually temporary, relations or views of a database. Most DBMSs include tools to analyze data and generate reports also. Languages such as SQL also contain operations to alter or add to the database. Typically, selection involved extracting rows based on an attribute being equal to a value or in a range between two values using the standard Boolean operators.

```
SELECT owner.name
FROM business, owner
WHERE business.id = owner.busid and business.turnover
> 100000;
```

Often to speed up querying, the system compiles indexes, sorted according to natural ordering, to access the relevant rows more efficiently. A typical DBMS will also include facilities for multi user access, versioning and transaction control to ensure the database is updated in a coherent way. The well founded methodologies for de signing and implementing well structured relational data bases, most notably a technique based on entity relationship modeling, together with its sound mathematical basis, has resulted in the almost universal dominance of relational databases for commercial and scientific applications.

Spatial Relational Databases

A spatial DBMS typically follows the same pattern as relational databases but includes extra data domains as standard which can model spatial attributes such as location, shape, and extent. Also included are spatial operations to allow these attributes to be manipulated in a similar way to the standard domains. What makes spatial domains different is that they can consist of complex data structures which may vary considerably depending on the application. For many applications the most appro priate is the vector model representing spatial attributes of features explicitly as their geometry of points, lines, and areas. A point would be defined using a suitable coordinate system; a line would be defined as an ordered collection of points; an areal feature would consist of an ordered col lection of lines that form a closed polygon. Spatial data base systems using the vector model have built in domains that allow users to model spatial features easily. However, many applications need to represent attributes that vary continuously over space, such as temperature, that are more naturally modeled as raster data grids.

Whatever the underlying model, spatial data types usually have no natural linear ordering on which the usual search, selection, and indexing algorithms can work. Also the topological relationships between objects are generally as important as spatial relationships. Typical queries required by applications may include measuring distances and calculating the areas of features. They also include spatial functions such as the creation of new features through the combination of existing ones (e.g., the intersecting area of two polygons or the buffer zone around a feature) and spatial predicates (e.g., 'is there a bus stop within 500 m of this building?').

A spatial DBMS provides facilities that enable these operations that would otherwise be difficult to efficiently program in a conventional database. Typically, these in clude tools to edit and create geometrical objects and link them to relational tables. In addition, spatial querying is provided which considers the spatial relationship between objects. Typical queries might involve determining the dimensions or centroid of a single object. Others deter mine whether objects touch, intersect, overlap, or contain one another – a group of operations classified as spatial joins analogous to the natural join using simple data types.

These spatial query terms are sometimes provided through extensions to SQL or by a dedicated spatial query language. Core domains and operations for spatial data have been defined as international standards by the Open Geospatial Consortium (OGC) and have been implemented in query languages such as Spatial SQL and MsSqlSpatial. So, for example, a query to find all the businesses within 10km of 'Arcare' might take the following form:

```
SELECT b1.name
FROM business b1, business b2
WHERE b2.name = 'Arcare' and
        distance(b1.location, b2.location) < 10;
```

Other Boolean spatial predicates include ones to detect adjacency, intersection, and containment between objects.

Spatial queries are efficiently implemented using spatial indexing. Unlike standard data domains that can be indexed alphabetically or numerically, spatial domains require a more complicated scheme that takes into account their two or three dimensional nature. Typical schemes include using a superimposed grid system, R trees based on the bounding rectangles of objects and quad trees that use recursive divisions of space to pro duce indices based on object density.

Other Spatial Database Models

Although the relational model is widespread, it is par ticularly suited for spatial data that uses the vector model of points, lines, and areas. However, not all tasks are amenable to, and not all spatial data come in, this form. Therefore, some spatial databases are based on alter native schemes such as the network or hierarchical models. Such problems include those dealing with topology and those concerning networks of features, such as roads in transport applications. In addition a large amount of data, particularly that used in environmental modeling or captured using remote sensing, is in the form of raster images or stored as data grids, and the applications that use them lend themselves better to flat file organization. Many commercial, spatial database systems, such as Oracle Spatial and ESRI ArcGIS, pro vide facilities for several types of data model to coexist and interact.

One drawback with the relational model is that the data describing an object tends to be distributed over several relations. The object oriented model combats this through the use of software engineering concepts, such as encapsulation and inheritance. Indeed, these concepts are very good at modeling object geometry and nonspatial attributes. However, this model inhibits issues, such as indexing and query optimization – factors very important with the large quantity of data usual in spatial databases.

See also: Qualitative Spatial Reasoning; Spatial Ontologies.

Further Reading

Burrough, P. A. and McDonnell, R. A. (1998). *Principles of Geographical Information Systems*. Oxford: Oxford University Press.
Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM* 13, 377 387.
Date, C. (2003). *An Introduction to Database Systems*. Boston, MA: Addison Wesley.
Date, C. (2005). *Database in Depth: Relational Theory for Practitioners*. Sebastopol, CA: O'Reilly.
DeMers, M. N. (2000). *Fundamentals of Geographic Information Systems*. New York: Wiley.
Guting, R. H. (1994). An introduction to spatial database systems. *International Journal on Very Large Data Bases* 3, 357 399.
Heywood, I., Cornelius, S. and Carver, S. (1998). *An Introduction to Geographical Information Systems*. London: Longman.
Longley, P. A., Goodchild, M. F., Maguire, D. J. and Rhind, D. W. (2001). *Geographic Information Systems and Science*. New York: Wiley.
Papadias, D., Zhang, D. and Kollios, G. (eds.) (2007). *10th International Symposium, SSTD 2007: Advances in Spatial and Temporal Databases*. Boston, MA: Springer.
Rigaux, P., Scholl, M. and Voisard, A. (2001). *Spatial Databases With Application to GIS*. Francisco, CA: Morgan Kaufmann.
Ritchie, C. (2002). *Relational Database Principles*. Florence, KY: Int. Cengage Business Press.
Shekhar, S. and Chawla, S. (2003). *Spatial Databases: A Tour*. Saddle River, NJ: Prentice Hall.
Worboys, M. and Duckham, M. (2004). *GIS a Computing Perspective*. Boca, FL: CRC Press.

Relevant Websites

http://www.opengeospatial.org
Open Geospatial Consortium.

Spatial Division of Labor

A. Smith, Queen Mary, University of London, London, UK

Glossary

Bloody Taylorism Despotic employment conditions in labor-intensive export industries often located in export-processing zones, usually in less-developed economies.
Division of Labor The separation of tasks and their allocation to different groups of workers as part of the relations of production.
Fordism A system of mass production involving the alignment of worker productivity with higher wage levels and the development of mass consumption.
Outward Processing A system of production organization involving the relocation of certain tasks in the labor process, often of a labor-intensive nature, to other production sites for assembly and reimportation into the original market.
Upgrading A process by which firms attempt to enhance their role in supply chains by adopting higher-level and higher value-added tasks and responsibilities.

In late 2006 the fate of a factory employing about 300 workers in South Wales made the international stage when Hollywood actor and son of the Rhondda Valley, Ioan Gruffudd, joined a campaign against the closure of the plant and the proposed relocation of its production of Burberry polo shirts to China or South America. This was the latest in a whole series of clothing factory closures in the UK and elsewhere in the 'old' European Union (EU) over the last 20 years which, in the UK, has seen employment in the industry drop, according to Annual Business Inquiry data, by 88 000 jobs or 64% in the 5 years since 1998 alone. In the wider EU, between 1985 and 1995, employment in the clothing industry is estimated to have declined by 40%. The Burberry factory in Ynyswen was no longer deemed competitive following a supply chain review by the global brand which is famous for its high end fashion garments and distinctive British sensibility and, more recently, for finding favor among sections of working class youth culture. The local economic impact is likely to be devastating in a region that has had its industrial heart torn out over the last 20 years. As one long standing employee of 40 years, 56 year old Alan Williams described:

> "Who's going to employ me at my age?". …He is bitter about Burberry's decision to close the plant, which has been in the Rhondda since 1938 and once employed 2,500 people. "We have worked to get Burberry where they are today, and they're just dumping us on the scrapheap. They're going to China to earn 2% more than they earn now." The GMB estimates Burberry will make an extra £4 m by switching production to Asia or South America; operating profits in the six months to September 30 were £84m; full year profits are on course to be double that. (Moss, 2006)

It is unusual for a day to go by without such media discussion of changing spatial divisions of labor, such as that in the Ynyswen case. While the language may be different, issues involving the changing location and organization of supply chains, patterns and shifts of employment, and the dynamism of organizational structures in the global economy attest to the very rapid transformations in economic geographical life we are currently witnessing. Indeed, the Ynyswen case highlights the central concept of changing spatial divisions of labor. From its headquarters in London, the design and decision making of Burberry is geographically removed from the sites of assembly production which operate in a quite fluid and increasingly liberalized global landscape of economic costs and social relations of production. Increasing competitive pressures in the global economy lead key decision makers in capitalist economies to take decisions which impact upon people and places from South Wales to China. From the call centers of India to the emergence of China as the putative 'workshop of the world', new geographies of economic activity and new spatial divisions of labor are continually being recreated and the geographical organization of the economy is thereby reshaped.

At such a time of profound reconfiguration of the geographies of the global economy, it is therefore more than appropriate to consider what spatial divisions of labor mean for understanding contemporary processes involving the de localization of jobs, economic upgrading, and national and regional repositioning in the global economy. Central to the current recreation of spatial divisions of labor is the internationalization of labor markets. Following the collapse of state socialism in East Central Europe and the thorough going, yet uneven, liberalization of the Chinese and Indian economies during the 1990s, there has been an internationalization of global labor markets. During the 1990s, this internationalization effectively doubled the size of the global labor force integrated into global capitalist commodity production. Access to large pools of lower cost labor resulted in a dramatic relocation of employment, especially in labor intensive manufacturing and service sectors from clothing assembly to call centers and business processing operations, away from traditional centers of economic and manufacturing activity. This has taken the form of the de localization of certain types of jobs away from the core economies to lower cost locations and increasing national and regional special ization, along with uncertainty and unemployment for some in the labor markets of core economies. In some treatments, these new spatial divisions of labor are rendered as a race to the bottom; a relocation of jobs of global proportions attendant upon the further removal of decent conditions of employment, living wage levels, and resulting in a further immiserization of the global working class. For others, they demonstrate the benefits of glob alization as industrial and service sector jobs are spread around the global economy, leading to increases in in comes, and the core economies are left to focus on higher value added activity, thereby generating greater wealth than a focus on more labor intensive activity allows.

Global and regional outsourcing has therefore dra matically reconfigured the map of employment and global spatial divisions of labor. From clothing production to call centers and business processing operations, jobs have been relocated and reconfigured in a complex and changing spatial division of labor. This has been accompanied, in the face of mass unemployment and de industrialization, by attempts to upgrade spatial divisions of labor in core parts of the global economy into knowledge based in dustries. For example, the EU's strategy for competitive ness has stressed the importance of knowledge based economies and concentration on highest value added activity to enable Europe to secure its place in an in creasingly competitive and globalized world.

What, then, is a spatial division of labor? Conventional approaches to divisions of labor seek to distinguish be tween the differentiation tasks and the differentiation of occupations. These approaches highlight how, over time, economies become more complex as tasks – and increasingly occupations – are split into many com ponents allowing specialization of activity. More recently, others such as the geographer Doreen Massey have focused on divisions of labor across space relating to how the economic geographies of control in the economy structure relations between places. In the following sec tions the origins of the concept of the division of labor are traced. This is then followed by a discussion of spatial divisions of labor and other more recent and related attempts to understand the changing geographies of divisions of labor in the global economy.

Origins: From the Ancient Greeks to Adam Smith and Karl Marx

Discussions of divisions of labor can be traced as far back as ancient Greek philosophy and debates about the country and the city and the organization of economic life. The most widely acknowledged conceptualization of a division of labor has its roots, however, in the Scottish enlightenment of the eighteenth century. The Scottish economist, Adam Smith, famously acknowledged the importance of the division of labor in establishing eco nomic efficiency. Using the example of the pin factory, Adam Smith argued that the division of labor allows for: (1) increasing 'dexterity' in production processes as workers concentrate on specific aspects of production; (2) time savings made because of the eradication of movement for the worker from one set of tasks to another; and (3) increased efficiencies through the use of machinery adapted to the increasingly specialized tasks of production. Through a deepening division of labor, economies thus become more efficient and the increasing segmentation and division of tasks and occu pations is, in this view, to be welcomed.

The analysis of the division of labor was taken much further and more critically by Karl Marx in his nine teenth century analysis of capitalism. For Marx, the development of capitalist manufacturing away from craft forms of production became central to the capitalist mode of production and entailed relations of production that resulted in the de skilling of workers as they con centrated on routine and repetitive activity in an in creasingly mechanized economy. The range of skills that any worker may possess became narrowed and resulted in the alienation of the worker from creative and satisfying work, which compounded the exploitative working con ditions and social relations embodied in the pursuit of profit and surplus value. At the same time, occupational categories developed with different positions in the production process leading to a labor aristocracy that – while still exploited by the capitalist class – retained some power, in positions such as overseers, foremen, or management, over the more mundane work of working class manual laborers. The result, for Marx, was the need to abolish the division of labor and to allow all workers to materially create conditions of existence that allow them to creatively express themselves as human beings.

Of course, this lineage of the origin of the concept of a division of labor can be traced to times when economies were in many ways much less globally extensive. For both Adam Smith and Karl Marx the focus was very much on divisions of labor and of tasks within the factory and within capitalist societies more generally. Later, Alfred Marshall acknowledged the territorial basis for divisions of labor, situated in local and regional pro duction systems, which became known as industrial dis tricts. Indeed, Marshall and later work on industrial districts in the Third Italy and beyond highlighted how such districts create locally competitive positions through the agglomeration of activity that combine tasks in the production process across many smaller specialized firms.

Neither approach, however, developed a strong geo graphical imagination to understanding how divisions of labor were established through space. It was not until the publication of Doreen Massey's *Spatial divisions of labour* in 1984 that research began to take seriously the geo graphical basis of divisions of labor.

Spatial Divisions of Labor

The most important geographical intervention in debates concerning spatial divisions of labor has been and con tinues to be Massey's conceptualization of the spatial division of labor. For Massey, the critical issue – as other economic geographers have continually argued – was that economies cannot be understood as operating on the head of a pin. Rather, she argued that the combined and uneven development of capitalism required the production of a patterned landscape, involving the geo graphical arrangement of particular functions in the division of labor. For Massey, it is precisely the organi zation of the capitalist economy over space that creates the social relations through which inequality and uneven development become produced:

> But new spatial divisions of labour are also more than just new patterns, a kind of geographical re shuffling of the same old pack of cards. They represent whole new sets of relations between activities in different places, new spatial patterns of social organisation, new dimen sions of inequality and new relations of dominance and dependence. Each new spatial division of labour repre sents a real, and thorough, spatial restructuring (Massey, 1984: 8).

For Massey this entailed an analysis and under standing of the uneven development of the UK economy and the increasing concentration of economic and pol itical power in London and the southeast of England, concomitant with the relative marginalization of the branch plant economies of the North:

> The fact that one region has jobs only in direct pro duction while another claims all the headquarters, or that areas differ in their dominant industries, or that in one area the jobs on offer are overwhelmingly manual while in another there is a sizeable slice of white collar and well paid employment in research, all these differences reflect different forms of geographical organisation of the relations of production. What are often called 'inter regional relations' reflect the same thing. Structures of dominance and subordination between economic activ ities in different places mirror the way in which … the relations of production are organised over space. 'Inter regional' relations are the relations of production over space (Massey, 1984: 67 8).

Indeed, in the United Kingdom during the 1970s and 1980s, 'Relations of production are increasingly central ized in the south and east of the country. There is a growing geographical, as well as social, division between conception and execution. …This is the new spatial division of labor, and it brings with it a new north–south divide.'

At the same time, it is important to distinguish between sectorally based spatial divisions of labor, in which individual economic sectors are concentrated in particular places, and hierarchically based spatial div isions of labor, in which different stages of production in one sector are located in different regions. As Morgan and Sayer argue, this distinction reflects the geographies of social and technical divisions of labor, respectively.

One illustrative example of a hierarchically based spatial division of labor in the UK is provided by the clothing and footwear industry that had been facing, by the time Massey was writing, a slow and painful economic decline and loss of contracts and production to other parts of the global economy. Massey highlighted the way in which small and large capital in the clothing industry were able to pursue different spatial strategies in this wider context of economic crisis. Small capital sought out cheaper labor reserves locally largely through outsourcing production to homeworkers and lower cost, often illegal, workers in the existing sites of production, such as London. Large capital, by contrast, was able to pursue a strategy involving the relocation of production, often to lower cost regions such as the northeast of England and Northern Ireland. Consequently, "in their choice of new locations[,] geographical dispersal and monopoly over local labor markets [in those new locations] enabled these large firms to reinforce the vulnerability of their new employees". The increasing hierarchical organization of the industry meant that the command and control functions – investment, design, marketing, etc. – were consolidated in London and other large urban centers while production and assembly activity was distributed to often feminized labor markets in peripheral and rural regions.

Spatial divisions of labor are thus reflective of wider technical, organizational, class, and employment divisions of labor in the capitalist economy, and matter critically for the politics of regional policy and working class organization. In particular, for Massey, the politics of spatial divisions of labor revolve around the fact that the "issue of external control [of regional economies] is not simply a matter of the geography of ownership … Any real solution would have to tackle also the question of the *social* location of control and the deeper organ isation of production itself". Interregional relations and the regional problem, as it was known, of uneven development are therefore the result of the stretching out of the relations of capitalist production across space.

For Massey, then, the very nature of the class organization of British capitalism was at stake:

> Nor is it only "capitalist relations of production" in some general sense which must be challenged. It is also the specific class from which they take, for which they are the basis in the United Kingdom, just as much as facing up to the problems of the national economy, means attacking the dominance political, economic and spatial of international capital. It also means challenging the social power of certain strata in British society, in particular the upper echelons of white collar hierarchies, managerial, professional and technical, and those who occupy and maintain the archaic structure of the British establishment in its broadest sense (Massey, 1984: 305).

Massey's conceptualization of spatial divisions of labor provided a much needed impetus to the rethinking of economic–geographical scholarship and continues to resonate in many debates in the subdiscipline. However, the increasing globalization of national economies meant that other research was beginning to focus more on the particular forms that new international divisions of labor were taking. It is to these more geographically extensive spatial divisions of labor that we now turn.

The New International Division of Labor

By the 1970s, the increasing internationalization of manufacturing activity – through the role of multinational manufacturing operations and through increasing outsourcing of production and global trade – led to recognition that conceptions of the bounded national economy and its spatial divisions of labor could no longer be sustained. Processes that Massey had identified operating in the UK national economy, relating to relocation and the reworking of spatial divisions of labor in manufacturing, were becoming increasingly related to international processes of production relocation. In Europe, there was increasing recognition that manufacturing activities were being relocated to other parts of the world economy creating new spatial divisions of labor. For Fröbel *et al.* in their detailed study of German industry, this involved what they termed a New International Division of Labor (NIDL). A major dynamic behind the NIDL involved the outsourcing of production from West European countries to lower cost production sites in Southern Europe, Asia, and – at that time – state socialist East Central Europe. Production costs were lower primarily because of the lower costs of labor in these regions – differences that continue today (**Table 1**). The major impetus for this relocation was the attempt to deal with the decreasing competitiveness – resulting from the saturation of markets and increasing costs – in the core economies through the search for low cost labor.

Table 1 Labor costs in the clothing industry

Country	*Cost per minute (US $)*
Chinese Mainland CMT	0.05
Belarus	0.06
Ukraine	0.06
Moldova	0.07
Bulgaria	0.08
Romania	0.09
Slovakia	0.14
Poland	0.16
Czech Republic	0.18
Hungary	0.18
Hong Kong	0.32
Greece	0.38
Portugal	0.42
Spain	0.48
Italy	0.60
Germany	0.62

Source: SECO Sector Consulting survey (2002) reported at: http://www.tdctrade.com/imn/03010902/clothing075.htm and http://www.just style.com/article.aspx?id=93035&lk=s

In Asia, the NIDL thus underpinned the emergence from the 1970s of the newly industrialized economies – the five Asian tigers – while in Southern Europe it involved the uneven integration into the wider European economy following the end of the dictatorships in Spain and Portugal. For French regulation theorist, Alain Lipietz, this new international division of labor involved the establishment of what he called bloody Taylorism, involving despotic employment conditions in labor intensive export industries located in export processing zones (EPZs), alongside the development of peripheral Fordism involving elements of factory work relations characteristic of Fordist production systems in Western Europe and the USA (the compartmentalization and segmentation of industrial work into strict categories in return for uneven union recognition and relatively stable jobs in some sectors). This occurred alongside the development of the state institutions of Fordism, notably a partial welfare state for the industrial proletariat. However, Lipietz insisted that this new spatial division of labor involved only a peripheral form of Fordism due to limited domestic markets and few skilled jobs being relocated from the core economies.

In East Central Europe, the spatial divisions of labor involved in this process of internationalization of subcontract production provided a basis for export led growth in the latter years of communist power and during the period of more general economic decline after the collapse of state socialism in 1989. In a range of labor intensive industries, but notably in clothing and footwear, systems of outward processing trade (OPT) were developed to enable the relocation of labor intensive aspects of the division of labor to lower cost production sites in East Central Europe and beyond. For enterprises in East Central Europe this provided much needed access to new export markets and new machinery, which was often shipped directly from factories being closed down in Western Europe. This system of outward processing trade which helped to establish a new spatial division of labor in Europe was largely driven, however, by the requirements of an industry in Western Europe facing increasing competitive pressure in home markets and the need to reduce costs. The outward processing trade system enabled the duty free reimport of textiles which had been exported to enterprises in East Central Europe for sewing into clothing. This customs system enabled textile firms in Western Europe and designer manufacturers to retain activity while outsourcing the more labor intensive parts of the clothing supply chain. It also enabled the development of quick response strategies allowing relatively fast delivery times because of geographical proximity and responses to rapidly changing fashion markets in core EU markets. In this latter sense, new international spatial divisions of labor are not only the result of differential labor costs. Issues of proximity to main markets, indigenous production capacities and skill levels in new production sites, and the changing requirements of markets in core economies have all worked to place East Central Europe as an important player on the EU clothing industry scene since the 1980s.

This process paralleled that in North America which was stimulated by a variety of regional trade arrangements culminating in the North American Free Trade Agreement (NAFTA), enabling production sharing and outsourcing from the USA to Mexico and, more recently, elsewhere in Central America. A variety of related regional trade agreements have extended these outsourcing arrangements beyond Mexico to Central America and the Caribbean Basin and, through the African Growth and Opportunity Act (AGOA), to Africa. Like the OPT arrangements in Europe, such regional trade agreements are designed to ensure a regional spatial division of labor partly protecting core country production of textiles and yarn, for example, from the US, with labor intensive sewing taking place in lower cost economies.

More recently, these regionalized divisions of labor have been challenged in part by the emergence and later consolidation of low cost, labor intensive production in China. Intensified by the 2005 completion of the phase out of quota constrained trade in textiles and clothing, China is witnessing dramatic increases in export production to European and US markets, raising implications for the long term sustainability of industries in these core areas. The growth of Chinese export production, however, has its roots in earlier post 1978 reforms allowing for the emergence of export oriented industrialization in regions such as the Pearl River Delta (PRD), neighboring Hong Kong, and Macau. Producers in Hong Kong and Macau were experiencing increasing operating costs which coincided with the planned, state led liberalization of regions such as the PRD. Foreign investment began to flow out of Hong Kong and Macau, drawing upon long standing ethnic ties, and into the PRD fuelling the rapid industrialization of what was previously a largely rural backwater. The PRD saw rapid economic growth based on the region's integration into a system of 'triangular trade' coordinated through Hong Kong and Macau for the main export markets in the US.

New Spatial Divisions of Labor: Value Chains, Commodity Chains, and Global Production Networks

These conceptualizations of dynamic and changing spatial divisions of labor – through the new international division of labor – have been developed further over the last 10 years or so, specifically with regard to debates over global commodity chains, global production networks, and networked forms of value production. The further development of spatial divisions of labor into new conceptual frameworks involving commodity chains and production networks has been very profitable. In particular, these new frameworks have focused on identifying the organizational forces leading to particular changes in spatial divisions of labor. This has enabled a focus on what has been called the governance of global production. Gereffi's focus on global commodity chains, for example, has stressed the importance of understanding the role of lead agents in the global organization of production. In particular, Gereffi emphasizes the distinctive roles of buyer driven and producer driven commodity chains (**Table 2**). Producer driven chains involve capital intensive and technology intensive industries, usually coordinated by large multinational enterprises and are characteristic of sectors such as automobile, aircraft, and computer manufacturing. Buyer driven chains, which have received much more attention in the research literature, are organized in a less hierarchical way, in which large retailers and buying organizations establish geographically distributed contracting networks with a large number of manufacturing operations, sometimes mediated by middlemen organizations, such as trade companies. Such chains are typically found in sectors such as clothing and footwear, where labor intensive operations are more important. These lead firms design and market the product but usually do not manufacture goods. Unlike producer driven chains, where profits come from economies of scale, high volume production, and technological innovation, buyer driven chains combine higher value design, sales, marketing, and financial services in the activities of retailers allowing them to act as strategic

Table 2 Characteristics of producer- and buyer-driven commodity chains

	Form of economic governance	
	Producer-driven	*Buyer-driven*
Controlling type of capital	Commercial	Industrial
Capital/ technology intensity	High	Low
Labor characteristics in production	Skilled/high wage	Unskilled/low wage
Controlling firm	Manufacturer	Retailer
Production integration	Vertical/ bureaucratic	Horizontal/ networked
Control	Internalized/ hierarchical	Externalized/ market
Contracting/ outsourcing	Moderate and increasing	High
Suppliers provide	Intermediate goods	Finished goods
Examples	Automobiles, computers, aircraft, electrical machinery	Clothing, footwear, toys, consumer electronics

Adapted from Kessler and Appelbaum (1998).

brokers in geographically distributed networks of over seas factories and traders. These factories are linked to the main markets for consumption by the buyer driven agents in the chain, who are thereby able to accumulate profits from the control of distributional and retail net works, brands and their image, and higher level design and control functions in these global spatial divisions of labor.

More recently, this framework has been extended in several directions to focus on value chains rather than commodity chains (through Gereffi's collaborative ini tiatives with John Humphreys at the Institute of Devel opment Studies and Tim Sturgeon at MIT), on global production networks (through the University of Man chester based Global Production Networks project), and on networks of value and new spatial divisions of labor (through collaborative research on regional development in the enlarged European Union). Each of these approaches has sought to build upon the original global commodity chains framework but to extend it in two main ways to understand new spatial divisions of labor. First, these new frameworks have sought to move away from a focus on the commodity to the social relations and organization of production and value creation, distri bution and appropriation in spatial divisions of labor. This has the potential to return more fully to Massey's original framework emphasizing that all forms of pro duction are embedded in social relations and that cap italist commodity production processes and the divisions of labor they create are about the (uneven) distribution and development of control of surplus value and profit in the economy. Second, a shift has occurred away from a more linear conceptualization of chains, in which com modities pass through different links in a unidirectional chain, to one emphasizing the complex organization of networks across space. These production or value net works are the lines through which value is created and divisions of labor established in the global economy.

More recently, emphasis has been placed on under standing how firms in particular places in an international spatial division of labor can attempt to reposition them selves and to upgrade activity to capture higher levels of value added and profit. This process of upgrading has involved the capturing of certain types of activity and some of the command and control functions in the global economy from global buyers. An illustrative example is again provided from the clothing industry in which some assembly firms in lower cost production locations such as Mexico, Hong Kong, Turkey, and East Central Europe have been taking on the role of what is called full package production. Buyers and retailers in core markets have increasingly relinquished these activities and have been seeking suppliers who can purchase fabric and trim inputs (moving away from outward processing type arrange ments), organize production networks to schedules de termined by buyers, and provide packed, hung, bar coded and labeled clothing direct to warehouses and stores in core markets. This process of upgrading enables some firms to reposition themselves in global spatial divisions of labor, but recreates new divisions through potentially outsourcing production to new firms in new locations. Questions have also been raised about the extent to which firm level upgrading translates into improvement in the conditions of work for employees and into wider regional level processes of economic upgrading and value capture.

Conclusion

Spatial divisions of labor continue to undergo dramatic change in a wider context of globalization and present real challenges to public policy makers, governments, businesses, and workers. The changing evolution of spatial divisions of labor has been accompanied by a very profitable development of conceptual frameworks over the last 20 years. These frameworks have focused on a shift from understanding regional divides in countries such as the UK to those understanding the changing organization of the world economy. A shift has also occurred toward understanding the organiza tional dynamics of firm and economic networks in the creation of spatial divisions of labor. However, although not explicitly making connections, these more recent approaches develop in different ways the original con ceptions of Massey and the spatial division of labor. In particular, they seek to understand contemporary processes of global economic transformation – often implicitly – in relation to how social and technical div isions of labor translate across space and constitute relations between places. This relational approach to such networks, and the spatial divisions of labor they give rise to, is most explicitly rooted in geography in the work on global production networks and on networks of value and regional development. However, such work extends into new interdisciplinary endeavors across economic sociology, international political economy, and global business studies, underlining the importance of cross disciplinary approaches to understanding the geographies of contemporary economic change.

See also: Commodity Chains; Export Processing Zones; Global Commodity Chains; Global Production Networks; Globalization and Transnational Corporations; Industrial Restructuring; Uneven Development.

Further Reading

Bair, J. and Gereffi, G. (2003). Upgrading, uneven development, and jobs in the North American apparel industry. *Global Networks* 3, 143 169.

Begg, R., Pickles, J. and Smith, A. (2003). Cutting it: European integration, trade regimes and the reconfiguration of East Central European apparel production. *Environment and Planning A* 35, 2191 2207.

Coe, N. M., Hess, M., Yeung, H. W. C., Dicken, P. and Henderson, J. (2004). 'Globalizing' regional development: A global production networks perspective. *Transactions of the Institute of British Geographers* 29, 468 484.

Frobel, F., Heinrichs, J. and Kreye, O. (1980). *The new international division of labour*. Cambridge: Cambridge University Press.

Gereffi, G., Humphrey, J. and Sturgeon, T. (2005). The governance of global value chains. *Review of International Political Economy* 12, 78 104.

Gereffi, G., Spener, D. and Bair, J. (eds.) (2002). *Free trade and uneven development: The North American apparel industry after NAFTA*. Philadelphia, PA: Temple University Press.

Henderson, J., Dicken, P., Hess, M., Coe, N. and Yeung, H. (2002). Global production networks and the analysis of economic development. *Review of International Political Economy* 9, 436 464.

Kessler, J. and Appelbaum, R. (1998). The growing power of retailers in producer driven commodity chains: A 'retail revolution' in the US automobile industry? Unpublished manuscript, available from Department of Sociology, University of California at Santa Barbara, USA.

Lipietz, A. (1987). *Mirages and miracles: The crises of global fordism*. London: Verso.

Lu, L. and Wei, Y. D. (2007). Domesticating globalisation, new economic spaces and regional polarisation in Guangdong Province, China. *Tijdschrift voor Economische en Sociale Geografie* 98, 225 244.

Massey, D. (1984). *Spatial divisions of labour: Social structures and the geography of production*. London: Macmillan.

Morgan, K. and Sayer, A. (1988). *Microcircuits of capital: 'Sunrise' industry and uneven development*. London: Polity.

Moss, S. (2006). 'They've got no reason to take it away from us', *The Guardian Unlimited*, 12 December (http://www.guardian.co.uk/g2/story/0,1969948,00.html).

Pellegrin, J. (2001). *The political economy of competitiveness in an enlarged Europe*. New York: Palgrave.

Pickles, J., Smith, A., Buček, M., Begg, R. and Roukova, P. (2006). Upgrading, changing competitive pressures and diverse practices in the East and Central European apparel industry. *Environment and Planning A* 38, 2305 2324.

Rantisi, A. (1982). *Marx and the division of labour*. London: Macmillan.

Sayer, A. and Walker, R. (1992). *The new social economy: Reworking the division of labour*. Oxford: Blackwell.

Smith, A., Rainnie, A., Dunford, M., Hardy, J., Hudson, R. and Sadler, D. (2002). Networks of value, commodities and regions: Reworking divisions of labour in macro regional economies. *Progress in Human Geography* 26, 41 63.

Relevant Websites

http://www.globalvaluechains.org/
Global Value Chains Initiative.

http://www.guardian.co.uk
Guardian Unlimited, Guardian News and Media Limited.

Spatial Expansion Method

E. Casetti, Ohio State University, Columbus, OH, USA

Glossary

Dual Expansion Method A technique for creating a linear terminal model from two linear initial models that have the same dependent variable. The technique involves redefining the parameters of any one of the two initial models into linear functions of the independent variables of the other.

Expansion Equations Equations that redefine the parameters in an initial model as functions of variables defining the initial model's context.

Expansion Method A technique for constructing a terminal model by redefining the parameters appearing in an initial model into functions of variables that define the initial model's context.

Models All the models referred to in this glossary are assumed to relate a dependent variable to one or more independent variables.

Parametric Drift It occurs when the parameters of a model have different values at different points of the model's context.

Introduction

This article is divided into five sections. In this section the expansion method and the spatial expansion method are defined and discussed in generalities. A demonstration consisting in the step by step construction of a model of polarized growth follows in the second section. In the third section, the dual expansions extension of the method is described and illustrated. Selected classes of applications of the spatial expansion method are outlined in section four. The research philosophy implications of the expansion methodology are the object of the concluding section of the article.

The expansion method is a technique for constructing mathematical models that involves widening the scope of an initial model by expansion equations that redefine some or all of its parameters into functions of expansion variables. The expansion variables index a substantively relevant context of the initial model. The terminal model thus produced includes both the initial model and a specification of its contextual variation. The mathematical models can be deterministic, stochastic, or mixed, and can be intended for solving, maximizing, or estimating. The expansion modeling can be applied to all these types of models.

The relevance of the expansion method to the spatial disciplines becomes readily apparent considering that models born nonspatial can be expanded to encompass spatial contexts defined by spatial variables, and that models born spatial can be expanded to encompass nonspatial contexts defined by nonspatial variables.

A model is spatial if at least some of the variables in it are spatial. Examples of spatial variables are spatial dummies, distances, directions, spatial coordinates, location quotients, measures of spatial inequality, and spatially referenced measures of phenomena that vary over space or index geographic environments. Spatial models are found not only in geography and regional science, but also in the spatial peripheries of virtually all social science disciplines. The spatial expansion method involves constructing spatial models either from a spatial initial model and nonspatial expansions, or from a nonspatial initial model and spatial expansions, or from a spatial initial model and spatial expansions.

Let us sketch a rough typology of spatial expansions constructed from classes of initial models and classes of expansion variables. The initial models can originate from spatial theory, such as spatial interaction models, models from central place theory (for instance, nearest neighbor distance of centers as function of centers' sizes), models from mathematical land use theory (for instance, population density models), transportation models, and so on. Alternatively, the initial models can originate within nonspatial social science disciplines.

The parameters of any of these models can be redefined into functions of nonspatial and spatial variables. The spatial expansion variables more frequently used include the distances from substantively significant geographical reference points, the variables that appear in spatial polynomials such as trend surfaces or two-dimensional Fourier polynomials, indexes of regional characteristics or factors extracted from these by factor analyses, and spatial systemic measures such as indexes of accessibility, of metropolitan dominance, or of inequality.

The traditional regional studies that emphasize spatial differentiation can be effectively integrated with nonspatial models (or vice versa) using the expansion methodology. For instance, this integration can be accomplished by redefining the parameters of a nonspatial model as trend surfaces. These spatial polynomial expansions can identify and map the statistically significant spatial variation of the model that can be then evaluated and commented upon in terms of the complex geographic circumstances the regional geographers are experts in. Alternatively, spatial expansions can be based on indices of economic, demographic, or social spatial differentiation, or on factors extracted from these indices.

An Example

In this section the spatial expansion method is demon strated by constructing a model intended for estimation and testing. Consider an analyst who wishes to investigate the spread of modern economic growth from a pole to the pole's hinterland. During the early stage of modern economic growth the product per capita tends to grow exponentially, and consequently the analyst chooses a simple initial model relating the natural logarithm of product per capita, y, to a linear function of time, t,

$$y = \alpha_0 + \alpha_1 t + \varepsilon \quad [1]$$

where ε is a vector of error terms, and the parameters α_1 and α_0 denote respectively the growth rate of y and the value of y at time zero. This model does not involve space. To extend it to encompass space, first its parameters are redefined by expansion equations into functions of distance from the pole, s, and then the expanded par ameters are replaced into the initial model to produce a terminal model that encompasses both time and space.

Specifically, the analyst chooses to expand all the parameters of the initial model into functions of distance from the pole, s, by the following expansion equations:

$$\alpha_0 = \gamma_{00} + \gamma_{01} s \quad [2]$$

$$\alpha_1 = \gamma_{10} + \gamma_{11} s \quad [3]$$

By substituting the right hand sides of eqns [2] and [3], for the parameters α_0 and α_1 in eqn [1], the terminal model [4] is obtained:

$$y = \gamma_{00} + \gamma_{01} s + \gamma_{10} t + \gamma_{11} ts + \varepsilon \quad [4]$$

Let us assume that we have data for the pole and the areal units in its hinterland at a number of unequally spaced points in time. Then our dataset consists of the variables y, t, and s, where y is the natural logarithm of product per capita, t stands for time, and s is the distance between the centroid of each areal unit and the centroid of the pole. The percentage rate of growth of product per capita over time, y', obtained by taking the time deriva tive of both sides of [4] is the following function of dis tance from the pole

$$y' = \gamma_{10} + \gamma_{11} s \quad [5]$$

Assume that the terminal model [4] has been estimated, and that the results of all appropriate post estimation tests do not warrant the respecification of the model or the use of a different estimation approach. Then the estimated model can be used to test for the possible occurrence and modalities of a number of types of growth, including polarized growth. Let us consider se lected tests relevant to the growth of product per capita:

1. If the hypothesis that $\gamma_{10} = \gamma_{11} = 0$ cannot be rejected, we have no evidence of growth of product per capita throughout the region.
2. If the hypothesis that $\gamma_{11} = 0$ cannot be rejected but the hypothesis that $\gamma_{10} = 0$ can, and if the estimated γ_{10} is greater than zero, growth of product per capita occurs at the same rate throughout the region.
3. If the hypothesis that $\gamma_{10} = \gamma_{11} = 0$ can be rejected, and the estimated γ_{01} and γ_{11} are respectively greater and smaller than zero, the notion that the region ex periences polarized growth over the time interval considered is validated.

Dual Expansions

The dual expansion method is an extension of the expansion method and builds upon the proposition that any given linear initial model and linear expansion equations imply a second initial linear model and linear expansion equations that yield the same terminal model. The two sets of initial models and expansion equations are referred to as primal and dual expansion formu lations, respectively. They imply each other much in the same way as primal mathematical programming formu lations imply dual ones, and vice versa.

Let us illustrate the primal and dual expansions by extending the spatial temporal model of polarized growth discussed in the previous example. The example for malizes by eqns [1]–[3] an initial model relating product per capita, y, to time, t, and expansion equations relating this initial model's parameters to distance from the pole, s. Suppose that the analyst defines a new initial model

$$y = \beta_0 + \beta_1 s + \varepsilon \quad [6]$$

relating y to distance from the pole, and expands its parameters β_0 and β_1 with respect to time:

$$\beta_0 = \gamma_{00} + \gamma_{10} t \quad [7]$$

$$\beta_1 = \gamma_{01} + \gamma_{11} t \quad [8]$$

By substituting the right hand sides of eqns [7] and [8], for the parameters β_0 and β_1 in eqn [6], we obtain the same terminal model [4] arrived at in the example.

The expansion formulation specified by eqns [6]–[8] is the dual of the one specified by eqns [1]–[3]. Actually, which one of the two formulations is primal is arbitrary, but when an initial model and expansion equations are specified, a dual formulation is also specified, by implication.

Clearly, the independent variable in the primal initial model is the same that appears in the dual expansion equations, and vice versa. Consequently, also the primal and dual initial models define the terminal model. They do so because the terminal model can be generated by expanding the parameters in one of the initial models into the independent variables appearing in the other initial model. Thus, we can regard the terminal model as a supermodel that combines and/or interacts the primal and dual initial models.

The dual expansions method shows that the model context perspective is not the only foundation of the expansions approach to constructing models. A second general foundation rests on the notion of combining and/or interacting alternative or competing models. These models could originate from different theories, possibly in different disciplines or in different subfields of the same discipline, and/or could reflect distinct empirical regularities.

In terms of a model context perspective, terminal model [4] is arrived at by expanding the parameters of initial models with respect to contextual variables. The initial model [1] is temporal and its context is spatial. The dual initial model [6] is spatial and its context is temporal. In terms of the models combining perspective, the terminal model [4] combines or interacts models [1] and [6] by expanding the parameters of one into the independent variable of the other.

Applications

The applications briefly described in this section are grouped into five clusters referred to as distance cluster, trend surface cluster, spatial context cluster, rank size cluster, and nonlinear dynamics cluster.

The distance cluster encompasses applications in which functions of geographical distance appear either in the initial model or in the expansion equation. For instance, spatial diffusion, change in urban population densities, polarized growth and polarization reversal, and temporal change in spatial interactions can be modeled by expanding into distance polynomial(s), the parameters of a time function that appears in an initial model. For example, spatial diffusion models have been generated from an initial model relating the percent of adopters of an innovation to a logistic function of time, and by redefining the logistic parameters as polynomial func tions of distance from a diffusion pole. In order to investigate the effects of proximity to large and to small cities on fertility, the parameters of an initial model relating fertility levels to the distances to the nearest primary and to the nearest secondary urban centers were expanded into quadratic polynomials in time. The spatial variation of international trade has been modeled by expanding gravity models' parameters into variables suggested by theory and previous empirics.

Initial models or expansion equations in which trend surfaces appear, characterize the trend surface cluster. In an intriguing paper by Eldridge and J. P. Jones, the friction of distance parameter in a gravity model was expanded into a quadratic trend surface. The estimated terminal model obtained was then used to produce dis torted maps that depict the spatial warping implied by the variation of the estimated friction of distance parameter. In other applications the spatial temporal dynamic of the acquired immunodeficiency syndrome (AIDS) epidemic was estimated using terminal models generated by expanding into trend surfaces initial models relating the number of AIDS cases to time. Methodo logically, similar applications have addressed the spatial temporal dynamic of population, of fertility decline, and of economic development.

Initial models and/or expansion equations in which geographically referenced variables appear, characterize the spatial context cluster. In a number of applications falling in this cluster, initial models relating housing prices to housing attributes (such as size, or number of bathrooms) have been expanded into neighborhood quality variables. These spatially expanded hedonic equations produce estimates of the relation between the value of each housing attribute considered and neigh borhood quality. This line of work was pioneered by Ayse Can and has spawned a growing literature. In other applications falling within this cluster the households' propensity to move has been related to households' attributes, and the resulting initial models were then expanded using indices descriptive of geographical con texts. Other examples of expansions with respect to spatial context dealt with manufacturing productivity, with the determinants of poverty, and with women's economic progress.

In the rank size applications of the expansion method the parameters of a rank size model have been expanded into functions of time. In the classical rank size models the natural logarithm of a city's population is linearly related to the natural logarithm of its rank. The time expansions of the parameters of this type of models have been used to investigate the differential growth of urban centers and the dynamic of regional income inequalities.

The nonlinear dynamics cluster includes applications of the spatial expansion method to nonlinear dynamics modeling. The expansion method is well suited to modeling complex spatial dynamics because it allows the linking of geographical entities such as countries or regions to dynamics that vary across space and time. Thus, geographical entities can be associated with stable or unstable equilibria. In one such application the expansion method is applied to the catastrophe modeling of modern economic growth. Another application centers on the spatial spread of modern economic growth in Europe, and yields estimates of the time when each of 16 European countries transitioned from premodern quasi stagnation to modern economic growth. A third application investigates the favorable and perverse dynamic equilibria that historically materialized in distinct clusters of countries, and that are associated respectively with strong manufacturing growth and with strong growth of the service sector.

Concluding Remarks

These remarks are concerned with the research phil osophy implications of the expansion method. Let us briefly sketch a line of reasoning suggesting why this is the case. Both conventional and expansion modeling link substantive relations among variables to mathematical structures such as equations, probability distributions, and stochastic processes. In the expansion modeling, however, the analyst starts from a preexisting initial model and redefines some or all of its parameters as functions of contextual variables by expansion equations.

Redefining parameters into variables is tantamount to assuming that the initial model drifts, namely, holds with different parameter values across a context. The con textual variables index aspects of the real world that are external to the initial model and with respect to which it can be presumed on substantive grounds that the initial model drifts. Thus, a familiarity with the expansion method is likely to engender a readiness to question whether any model's parameters should be assumed to be stable. Concurrently, it is also likely to motivate search ing across which contexts these parameters might drift and why.

This switch from the default presupposition that parameters are stable to the default presupposition that they are likely to drift across contexts has a substantial impact on how research is carried out and leads to a change in the way models are perceived. Models cease to be candidates for law status, and become tools to ask questions about the manner in which they drift across real contexts and the reasons why they do so. In the limit, models are no longer expected to be true. Instead, they become devices for interrogating shifting realities, with the objective of explaining which shifts occur and why.

See also: Complexity Theory, Nonlinear Dynamic Spatial Systems; Hypothesis Testing; Regression, Linear and Nonlinear; Spatial Autocorrelation; Spatially Autoregressive Models; Trend Surface Models.

Further Reading

Can, A. (1990). The measurement of neighborhood dynamics in urban house prices. *Economic Geography* 66, 254 271.

Can, A., Anselin, L. and Casetti, E. (1989). Spatial variation in the propensity to move of households. *Sistemi Urbani* 1, 107 127.

Casetti, E. (1972). Generating models by the expansion method: Applications to geographical research. *Geographical Analysis* 4, 81 91.

Casetti, E. (1986). The dual expansion method: An application for evaluating the effects of population growth on development. *IEEE Transactions on Systems, Man, and Cybernetics SMC* 16, 29 39.

Casetti, E. (1989). The onset and spread of modern economic growth in Europe: An empirical test of a catastrophe model. *Environment and Planning A* 21, 1473 1489.

Casetti, E. (1992). Bayesian regression and the expansion method. *Geographical Analysis* 24, 58 74.

Casetti, E. (1997). Catastrophe models and the expansion method: A review of issues and an application to the econometric modeling of economic growth. *Discrete Dynamics in Nature and Society* 1, 185 202.

Casetti, E. (1997). The Expansion method, mathematical modeling, and spatial econometrics. *International Regional Science Review* 20, 9 33.

Casetti, E. and Pandit, K. (1987). The non linear dynamics of sectoral shifts. *Economic Geography* 63, 241 258.

Eldridge, D. J. and Jones, J. P. (1991). Warped space: A geography of distance decay. *Professional Geographer* 43, 500 511.

Fan, C. C. (1988). The temporal and spatial dynamics of city size distribution in China. *Population Research and Policy Review* 7, 123 157.

Jones, K. and Bullen, N. (1994). Contextual models of urban house prices: A comparison of fixed and random coefficients models developed by expansion. *Economic Geography* 70, 252 272.

Jones, J. P. and Casetti, E. (eds.) (1992). *Applications of the Expansion Method*. London: Routledge.

Kristensen, G. (1997). Women's economic progress and the demand for housing: Theory, and empirical analyses based on Danish data. *Urban Studies* 34, 403 418.

Malecky, E. J. (1980). Growth and change in the analysis of rank size distributions: Empirical findings. *Environment and Planning A* 12, 41 52.

Zdorkowski, T. R. and Hanham, R. Q. (1983). Two views of the city as a source of space time trends in economic development and the decline of human fertility. *Urban Geography* 4, 54 62.

Zhang, J. and Kristensen, G. (1995). A gravity model with variable coefficients: The EEC trade with third countries. *Geographical Analysis* 27, 307 320.

Spatial Filtering/Kernel Density Estimation

G. Rushton and C. Tiwari, The University of Iowa, Iowa City, IA, USA

Glossary

Bandwidth The radius of a circular kernel.

Choropleth Map A map that shows data values for discrete areas that are grouped into data classes.

Kernel A defined area over which observations will be summed, weighted, and averaged to estimate the density of the observations that will be attributed to the center of the kernel.

Privacy Protection A method for ensuring that individually identifiable information cannot be construed from a map.

Window Width Same as bandwidth.

Density Estimation

Much of geography is the study of patterns of points. This article focuses on methods for studying the spatial pattern of densities of points. This includes densities of observations or densities of the ratio of one distribution to another – such as the ratio of infant deaths and infant births over small areas to make a map of infant mortality. The focus is not on the absolute locations of things but on their relative location. Two questions are raised by this interest: over what area should the density be measured and how should each point be weighed when computing the density at any location. A density can never be defined literally at a point because by its nature it requires an area to be defined. But clearly the size of this area can be variously defined. As shown below, the larger the area over which a density is defined, the smoother the density will be. The area over which density is measured is known as 'the kernel' or 'the spatial filter', and the spatial weight function is known as the probability density function over which probabilities sum to one. In the simplest case, all occurrences within the kernel are weighted equally and the kernel density is simply a two dimensional moving average. In other cases, however, the weight assigned to an observation within the kernel declines with some function of distance from the point for which the density is estimated. Increasingly, with access to geographic information system (GIS) functions as well as specialized software, geographers are producing 'kernel density maps' or 'spatially filtered maps' as an important part of the process of investigating spatial aspects of many problems.

Examples of application areas that use density estimation techniques include disease rates (number of diseased cases/number of people at risk); crime rates (number of crimes/number of people); segregation rates; or poverty rates. In other applications, the densities may be spatial rates of change – as when the number of children in grade $x+1$ in time $t+1$, divided by the number of children in grade x at time t is a density that describes the change in the number of children in the cohort over a 1 year period in the area of the filter. This is called the grade progression rate. Density surfaces are sometimes computed for the purpose of testing hypoth eses. More commonly, they are used as an exploratory data analysis tool to suggest patterns in data, to 'smooth' the data, or to separate 'signals' from 'noise'.

When the spatial filter is circular and symmetric, it is usually called a 'kernel' and the size of the circle is called the 'bandwidth' or the 'window width'. The character istics of density maps are that values are continuously distributed across the maps rather than in discrete values that change at borders of spatial collection units as in choropleth maps. In this respect, spatial density functions are solutions to the 'modifiable area unit problem' where patterns are dependent on the area units over which data are gathered and displayed.

Relationships between Choropleth Maps and Density Maps

Although choropleth maps do not show continuously distributed values, they often portray densities. **Figure 1a** shows point based data – in this case the approximate locations of 41 238 births and 396 deaths of children within 1 year of birth in the greater Des Moines, Iowa, metropolitan area between 1989 and 1996. The resi dences of mothers of these children were geo coded to their street address. **Figure 1b** shows one method of illustrating the density of these point based distributions. It shows, for US Census Block Groups, a map of the infant mortality rate, which is the ratio of deaths to births (×1000).

In the language of kernel density maps, such chor opleth maps show densities for nonoverlapping kernels of different shapes and sizes. Viewed in this way, it is becoming common to question the appropriateness of displaying densities on choropleth maps unless avail ability of data absolutely prevents the making of kernel density maps. It is difficult to make the case that a spatial density map constructed for nonoverlapping, different sized and shaped areas, is the most appropriate way to describe spatial variations of densities. Yet, many

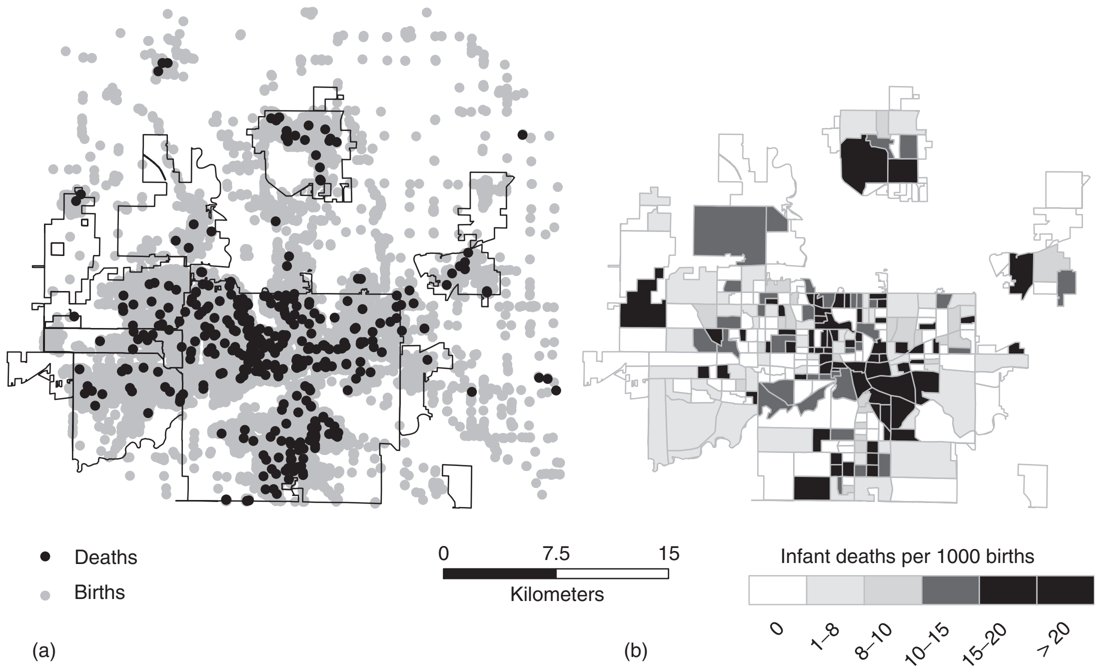

Figure 1 Example of point-based data for making a kernel density map of infant mortality in the Des Moines (Iowa) metropolitan area. Births and deaths, 1989 96. Note that the locations of the births and deaths shown on the map have been spatially masked to preserve the privacy of the individuals shown on the map. (b) A choropleth map of infant mortality for the same area. Data is aggregated for US Census Block Groups.

choropleth type maps continue to be constructed in this way as illustrated in **Figure 1b**.

Kernel Density Methods

In preparing a density map, a kernel is placed suc cessively on the crossing points of a grid and at each point a density is measured. The kernels overlap so that densities at successive points on the grid are similar because they share a considerable proportion of the data values in the kernels. By design, therefore, the density map is usually smooth.

The densities are often interpreted as probabilities of events occurring – the probability of a crime being committed in a given location; the probability of a person having a disease, etc. Kernel density estimation is a nonparametric statistical method for estimating the probability of a defined event; when gathered across all grid points a probability distribution of densities is pro duced. The density estimate is produced directly from the data. After the densities have been computed for all grid points, the density map is computed by inter polating the surface from the pattern of density estimates. A spatial interpolation function in a GIS is typically used to do this. The use of an interpolation function to create a smooth surface can be avoided by computing the density for every pixel on the map (known as a raster layer). Since the kernel density estimates obtained at grid points are already smoothed, a raster layer generated from them can be used to create the density surface.

Typical Input Data for Density Maps

Density maps can be created from point data (individual entities) or from aggregates of data for small areas. For each grid intersection a count is made of the observations within the kernel or, with spatially aggregated data, a count is made of the small area aggregates for all areas whose centroids lie within the window. In making an infant mortality map for a city, for example, the grid points might be 0.5 km apart, the window might be 3 km in radius, and the data may be aggregated from small census areas. Such a map is made as a density defined as the ratio of the aggregates of births and infant deaths, respectively, in each kernel.

Mathematical Discussion

A mathematical introduction to kernel density estimation methods is found in Silverman. We distinguish here between weighted and unweighted kernel methods. In unweighted methods, each data point contributes equally to the measure of density. In weighted methods, each data point contributes to the measure of density according to a function that is inversely related to the distance of the observation point from the location for which density is to be estimated. Data points not located within the kernel window have zero weight. There are several inverse distance weighting functions in the literature. Silverman provides more details.

Given data values $x_1, \ldots, x_n$;

1. An unweighted kernel density for a fixed window size h is defined at location (loc) as:

$$p(\text{loc}) = (1/m)\ \Sigma x_i \quad (i = 1, \ldots, m)$$

   $x_1, \ldots, x_m$ define the subset of values of x in order of their distance from loc for the kernel or window centered at loc of size h. For any location, m will vary in size with the density of the data observations. In this case h is a fixed distance size and m is variable (see **Figure 2a**).

2. An unweighted spatially 'adaptive' kernel density is defined as

$$p(\text{loc}) = (1/M)\ \Sigma x_i \quad (i = 1, \ldots, M)$$

   $x_1, \ldots, x_M$ define the subset of values of x in order of their distance from loc for the kernel or window cen tered at loc of whatever size h is sufficient to find M data points. For any location, M is set at a fixed value and the kernel size h is whatever distance is needed to capture m observations. For this adaptive kernel case, h is of variable size and M is fixed (see **Figure 2b**).

3. A weighted kernel for a fixed window size h is defined as

$$p(\text{loc}) = (1/m)\Sigma w_i x_i \quad (i = 1, \ldots, m)$$

   where $w_i = w_i/\Sigma w_i \quad (i = 1, \ldots, m)$, where $\Sigma w_i = 1$. Silverman describes several functional forms for defining this weights function.

Effects of Varying the Grid Size

Densities are measured from points on a grid. The resulting density map should not depend on the locations of the grid points. This is achieved if the resolution of the grid is much smaller than the window size – a good rule of thumb is that the window size should be at least four times the distance apart of the grid points. Such a ratio ensures that a large proportion of data points will overlap between neighboring window counts and the density values in neighboring grid points will be very similar.

Effects of Varying the Window Width

The size of the kernel is important in kernel density estimation. The window width is often referred to as the 'spatial filter', the 'bandwidth', or the 'smoothing parameter' since all data within the window is averaged for the point to which it is referenced. The window corresponds with the spatial scale at which the density is examined. Small kernel sizes give local detail and hence, in most cases, large spatial variability in the density surface. Large sizes give regional detail and smooth the local detail to reveal regional patterns of density change. It follows that increasing the size of the window, increases the smoothing of the map and decreases the variability in the map. There is no universally accepted approach for selecting the window width; instead, there are accepted subjective and automatic methods for selecting it.

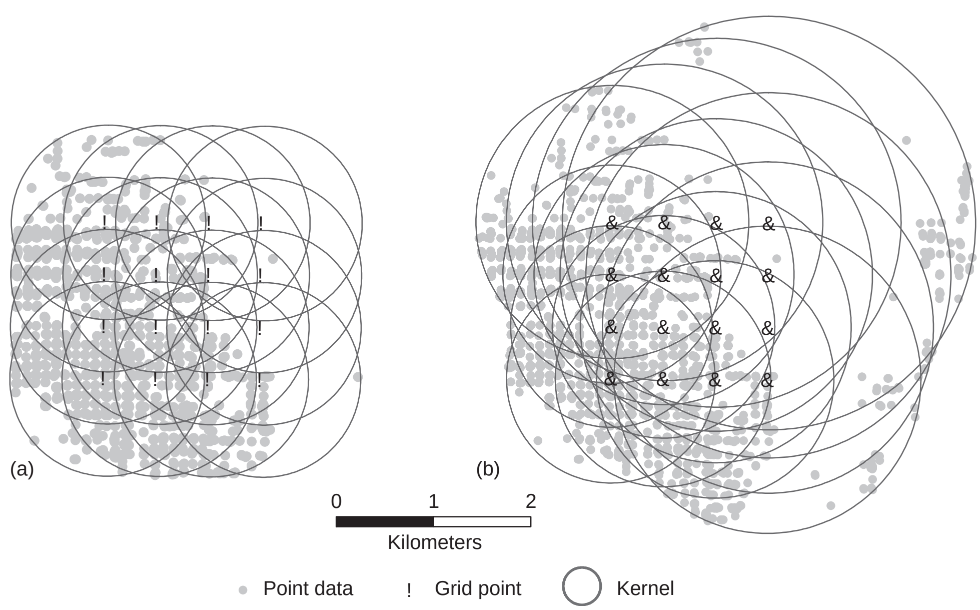

Figure 2 A comparison of fixed and adaptive kernels in relation to the spatial pattern of grid points for which densities are estimated.

Common guidance for making a subjective choice is to make several maps using different window widths and explore what is revealed as the window size becomes smaller. Geographers refer to this practice as 'exploring scale dependent variation' since it is well known that the processes that generate geographic distributions are frequently scale dependent. For many problems, choosing a value of h somewhat larger than the value at which the densities begin to break up spatially is recommended as the optimal bandwidth. Silverman describes this approach as the test graph method. Studies of rates of asthma in children have been shown to decrease with distance from congested urban highways by varying the spatial window as a linear feature oriented to the centerline of the highway. Likewise, the practice of varying the window width from small to large is often called 'upscaling the data'.

The Adaptive Kernel Method

In the adaptive kernel estimate, areas with low densities of observations are smoothed more than areas with high densities. The reasoning here is that the more data contributes to a given density rate, the more reliable that rate is. One rule to use is to select a size for a kernel adaptively based on the data around each point on a grid for which it is measured such that an equal number of observations is used for each point. It follows, that the kernel will be larger in areas of low density of observations. The geographic scale of the resulting map is variable across the map. **Figure 2b** illustrates the kernel size for a case where observations are shown. For the spatially adaptive kernel where m is fixed and h is variable, it is the value of m that needs to be determined and that too can be selected by the test graph method in which the value of m is varied systematically. Spatially adaptive kernel density maps are especially useful when the density measured is interpreted as an estimate of risk which is based on numbers of events that occur with small probabilities. In such cases, the reliability of measures of risk is directly related to the size of the sample from which they are estimated. An adaptive kernel will ensure that estimates are based on the same size samples of data and therefore will be of approximately equal reliability.

Illustration of Density Estimation Methods

Using the data shown in **Figure 1a**, we illustrate in **Figure 3** the two kernel density estimation methods shown in **Figure 2**. As in **Figure 1b**, the density is defined as the ratio of deaths and births within the kernel area. By convention, this ratio is multiplied by 1000 so that the interpretation is 'deaths per 1000 births'. Instead, however, as in **Figure 1b**, showing this density for discrete areas of different sizes and shapes, using kernel density estimation methods, we now control the sizes of the kernels. In this sequence of maps, map (a) uses the fixed filter of 1.8 miles and map (b)–(d) uses spatially adaptive kernels of 750, 1250, and 1750 births, respectively. The median size of the adaptive kernels in map (c) was 1.8 miles (approximately 3 km). Note that in this sequence, the larger the kernel size, the more smooth is the rate and fewer high and low spots are found.

Protecting the Privacy of Point Data

In many applications of density estimation methods, the actual locations of events are considered private information that cannot be publicly disclosed. Armstrong *et al.* discuss a variety of methods, known as geographic masks, that can be used to protect privacy. In the case of the births and deaths in **Figure 1a**, a spatial randomization mask has been used to place the point randomly within a selected distance (200 feet) of its true location. By agreement with the institution that provided this data, this masking rule was used for any public display of the data. The locations of actual residences of the mothers of these children were used in the analyses shown in **Figure 3**.

Issues of Statistical Significance

In spatial epidemiology, kernel density estimates are often made and tested for statistical significance. Boscoe *et al.* and Kulldorff make maps that show those kernels where densities exceed a given level of statistical significance. Openshaw *et al.* compute maps that show the likelihood that the observed rates are higher than expected rates based on a stated hypothesis of the probability that each person may be positive for the event examined. These are known as 'probability maps'.

Some Limitations

There are other spatial analytic techniques that can be compared with kernel density techniques when the interest is in density surfaces. These include trend surface analysis, splines, and kriging. The main difference of these approaches is that they attempt to model the surface either as a function of the location coordinates (trend surface analysis) or as spatial interpolation from known values at sample locations (kriging), rather than to construct the surface descriptively from local data (kernel density estimation). These other approaches are superior when the problem is that of uncovering spatial structures

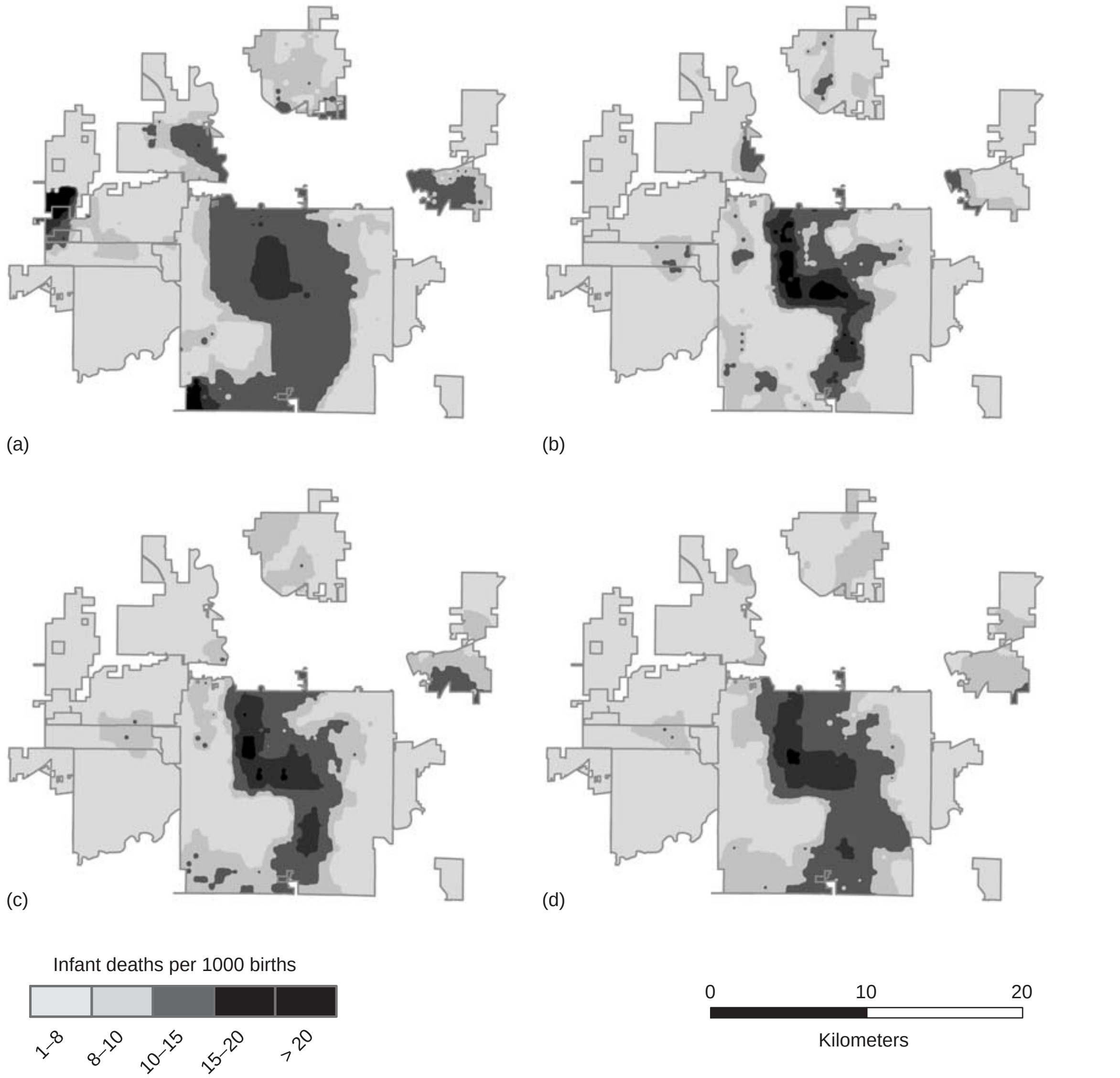

Figure 3 Four kernel density maps of infant mortality rates for the Des Moines, Iowa, area, 1989 96. (a) Fixed kernels, 1.8 miles radius (~3.0 km); (b) adaptive kernels, 750 births; (c) 1250 births; and (d) 1750 births.

in the data. An important question is determining the number of observations required to make valid density estimates. Adequate sample size varies with the spatial structure of the data. When the true (unknown) density is unimodal and not complex, a smaller sample size will suffice to estimate the density surface. However, when the density surface is complex, a larger sample will be required to make accurate density estimates. Spatially adaptive kernels have been developed to ensure that each density estimate will have the required spatial basis of support. Gatrell *et al.* discuss several possible extensions to kernel estimation as solutions to this problem.

Densities computed along the edges or corners of study regions suffer from the loss of data when the kernels centered within the study area have large areas without data. Density estimates from processes that occur according to defined probabilities are therefore less reliable in these areas unless spatially adaptive kernels are used to compensate for the loss of data along edges by increasing the window width to gain more strength for the points in these areas. A limitation of the spatially adaptive kernel method is that the user of such maps generally does not know the precise area that has contributed to any given value on the map. Maintaining accurate ‘readouts’ is an important property of maps.

Software for Kernel Density Applications

Users of several commercial GIS software offer kernel density estimation methods as options in their application areas. A common approach is to use the buffer function in a GIS centered on the grid values. The National Institute of Justice has distributed software – CrimeStat – for application of the method to crime data. The National Cancer Institute has distributed software – SaTScan – for identifying hot spots of disease that do not show the entire kernel density surface but, instead, show those kernels where densities exceed defined levels of statistical significance after correcting for multiple testing. The University of Iowa distributes DMAP which is software for making kernel density estimates for a grid.

Conclusions

The use of kernel density estimation methods in human geography is increasing rapidly as data for small areas as well as geo coded data become available. Also contrib uting to their use is the availability of software and GIS that implement these methods. In addition to these technical forces, is the increasing philosophical interest in measuring and studying local differences, theorizing why such differences exist or, as in many practical imple mentations of these methods, making decisions to employ resources to change the patterns. Kernel densities are proving to be useful for these purposes.

See also: Edge Effects; Kriging and Variogram Models; Modifiable Areal Unit Problem; Point Pattern Analysis; Spatial Clustering, Detection and Analysis of.

Further Reading

Armstrong, M. P., Rushton, G. and Zimmerman, D. L. (1999). Geographically masking health data to preserve confidentiality. *Statistics in Medicine* 18, 497 525.

Bailey, T. C. and Gatrell, A. C. (1995). *Interactive Spatial Data Analysis*. New York: John Wiley.

Boscoe, F. P., McLaughlin, C., Schymura, M. J. and Kielb, C. L. (2003). Visualization of the spatial scan statistic using nested circles. *Health and Place* 9, 273 277.

Diggle, P. J. (1985). A kernel method for smoothing point process data. *Journal Royal Statistical Society C* 34, 138 147.

Gatrell, A. C., Bailey, T. C., Diggle, P. J. and Rowlingson, B. S. (1996). Spatial point pattern analysis and its application in geographical epidemiology. *Transactions, Institute of British Geographers NS* 21, 256 274.

Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics: Theory and Methods* 26, 1481 1496.

Openshaw, S., Charlton, M., Wymer, C. *et al*. (1987). A Mark I Geographical Analysis Machine for the automated analysis of point data sets. *International Journal Geographical Information Systems* 1, 335 358.

Rushton, G. and Lolonis, P. (1996). Exploratory spatial analysis of birth defect rates in an urban population. *Statistics in Medicine* 15, 717 726.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Waller, L. A. and Gotway, C. A. (2004). *Applied Spatial Statistics for Public Health Data*. Hoboken, NJ: John Wiley.

Relevant Websites

http://www.icpsr.umich.edu
About CrimeStat, Inter University Consortium for Political and Social Research.

http://www.satscan.org
Software for the spatial, temporal, and space time scan statistics.

http://healthgis 01.iowa.uiowa.edu
Spatial Analysis for Environmental Health Surveillance.

Spatial Interaction Models

M. E. O'Kelly, The Ohio State University, Columbus, OH, USA

Glossary

Beta A parameter used to control the response of interaction to distance.
Destination A place to which interaction is attracted.
Origin A place from which interaction emanates.
Spatial Interaction Flow or movement over space which results from a decision process.

Introduction

Gravity models evolved from observed empirical regu larities in phenomena such as migration. Early work by Ravenstein identified the role of 'push and pull' factors and derived law like statements on the role of distance on diminishing regular interactions between places. The ideas in this field are often collected under the term 'gravity model', based on a fairly direct analogy with the Newtonian concept of interaction between masses at a distance. Despite these earlier crude generalizations, the model has in fact spurred many sound behavioral hypotheses and prompted much careful thinking in geography and regional science on the modeling of interaction phenomena. The later versions of these ideas are often collected under the title of spatial interaction (SI) modeling.

The earliest formulations had a striking similarity to the Newtonian model familiar from physics, with driving forces like 'masses' (V_i and W_j) which in this case are assumed to be proportional to measures of the popu lations or flows originating from place i (O_i) and destined for place j (D_j) representing the key push and pull effects respectively. These areas interact in a positively reinforcing way that is a multiplicative function of their masses, and at the same time their interactions are diminished by the role of intervening distance. Further, the model requires a 'k' factor to ensure the concordance of the units to some observed total flow. A typical notation for the model is

$$M_{ij} = kV_iW_j/d_{ij}^{\beta}$$

where M_{ij} is the expected interaction between origin i and destination j; V_i and W_j are representative of the masses of the origin and destination respectively; and d_{ij} raised to an exponent β ('beta') is the distance effect acting as an impedance on the interactions. Note that the distance term appears in the denominator of the equation. The model can also be written as

$$M_{ij} = kV_iW_jd_{ij}^{-\beta}$$

Note the conventional entry of β with a negative sign reflecting the fact that it enters the denominator of the original equation. Thus, a value of $\beta = 2$ would imply that interaction diminishes with the square of distance. The indices 'i' and 'j' could denote a city, node, or regional area from which flows come (origins) and to which flows go (destinations). Contributing to the early use of the model was its relatively easy estimation using a log linear transformation, essentially making this model amenable to estimation by multivariate regression. Such simplistic efforts at estimation have since been eclipsed by more rigorous derivations using a wide range of statistical methodology (Poisson regression, generalized linear models, maximum likelihood, genetic algorithms, hybrid numerical methods such as neural networks, and so on). Even as geographers have downplayed the role of 'analogies' and law like statements as a basis for such models, the empirical usefulness of this type of model in areas such as air passenger forecasting has continued to keep it in the forefront of the spatial analyst's toolkit.

The model in its modern guise is known as an SI model and is widely studied, accepted, and is one of the principal contributions of spatial analysis to the fields of demography, regional science, and traffic estimation. The model was useful as an early basis for trade, air passenger traffic, and migration, but perhaps its strongest contribution was to stimulate the derivation of a family of models from a much more rigorous basis: that is, the model moved away from analog reasoning toward der ivation using information theory, choice theory, and behavioral processes. The fact that a similar functional form can be obtained from such diverse perspectives is regarded by many as a strength of the underlying fun damental conceptual framework.

Simple Gravity Model

The gravity model posits that, other things being equal, interaction increases between larger places, and between places relatively close to each other. The notion of *ceteris paribus* or literally, holding 'all else equal' is important to SI theory, as it is necessary to derive and, if possible, separate the independent effects of a plethora of ex planatory variables. The pull and push factors that drive interaction would be technically referred to as attraction (positive impact on flow) and repulsion (negative impact on flow), and typically these effects are measured indirectly through some proxy variables. Perhaps some of clearest examples would be the idea that migration should be positively attracted to areas with thriving economies, good wages, and plentiful affordable housing. In common practice however, these effects are rarely aligned exclusively in favor of a single dominant region, and in the case of migration, places with high in migration may very well also have high out migration and return flows, due perhaps to rising house prices, failed expectations on the part of migrants, pressure of high competition in a congested metropolis, and so on.

Interaction forces therefore change over time in response to changes in attraction and repulsion factors. Perhaps the greatest challenge facing interaction mod elers is not simply the notion of movement as a function of explanatory independent variables, which seems to be reasonably well established, but the need for accurate forecasts of the many driving forces and complexities that underlie the effects. In part this is made more difficult by the fact that some aspects of flows and attraction are intertwined: in a sense, attraction of a destination for migrants depends on the pool of migrants already in the community. Such endogenous effects give rise to a non trivial estimation task. The relevant econometric litera ture in migration has of course encompassed these general theories of movements (attracting attention from Muth, Alonso, Anselin, and Getis among others). As these systems are usually 'open' and inherently subject to un predictable changes, the art of interaction forecasting becomes more difficult.

Spatial Interaction Models

Spatial interaction models, in general terms, deal with flows and movement between places, based on (1) their spatial separation; (2) their complementarity; and (3) whatever other intervening opportunities or spatial structural elements serve to augment or diminish the expected flow. Spatial separation in turn is measured by cost or time involved in the interaction, and is often represented by a distance decay function that shows diminishing interaction as distance increases.

Complementarity is an indicator of the suitability of the particular pair of places for the movement in ques tion, so for example, a place with surplus crop production would be in a sense complementary to a place that has demand for food products. Whether trade flows actually take place between such complementary places depends on the distance, prices, and institutional barriers. Recognizing that interactions may be diverted to inter vening opportunities begins to raise the complex ques tion of the relative location of the interacting agents vis à vis other places. These fundamental driving forces (often attributed to Ullman) provide a general set of broad explanatory factors that help to measure, model, explain, and in some cases predict spatial interaction.

Family of Constrained SI Models

By and large, the literature on measuring SI is structured around the notion of constrained origin–destination (OD) or trip distribution tables that are 'pinned down' on the marginal totals over either the rows, or columns, or both. When a table with row (or origin) total is con strained, it is required to match the data as well as pos sible with the known trip productivity from each origin. When a table with column (or destination) total is con strained, one is attempting to match the table as well as possible with the known trip pull or attraction to each destination. There is thought to be a trade off between better fitted tables with more constraints (i.e., a higher quality fit) versus a larger number of predictive effects in the unconstrained case. Each kind of model has its par ticular use and application.

Production constrained models predict the share of a fixed amount of trip production from an origin to a given set of destinations:

$$T_{ij} = A_i O_i W_j f(c_{ij})$$

$$A_i = \frac{1}{\Sigma_j} W_j f(c_{ij})$$

where T_{ij} is predicted flow, A_i are balancing factors that ensure that the model reproduces the observed origin trip rates (O_i), and $f(c_{ij})$ is a distance impedance factor that limits trips to match an observed empirical trip length. Such models have the property of predicting an unconstrained estimate of the flows arriving at the jth destination, while reproducing the expected flow from the ith origin (O_i). They are especially well suited to retail trade area estimation.

Attraction constrained models predict the share of a fixed amount of trips arriving at a destination that come from a particular origin:

$$T_{ij} = V_i B_j D_j f(c_{ij})$$

$$B_j = \frac{1}{\Sigma_i} V_i f(c_{ij})$$

where T_{ij} is predicted flow, B_j are balancing factors that ensure that the model reproduces the observed destin ation trip rates (D_j) and $f(c_{ij})$ is a distance impedance factor that limits trips to match an observed empirical trip length. Such models have the property of predicting an unconstrained estimate of the flows departing from the ith origin, while reproducing the expected flows for the jth destination (D_j). They are especially well suited to residential land use allocation around fixed work places.

Doubly Constrained Models

An intensive use of SI models is in the area of trip dis tribution, an important step in urban travel demand models; it is also called a 'doubly constrained' model. These interaction equations are basically of the form devised by Wilson:

$$T_{ij} = A_i O_i B_j D_j f(c_{ij})$$

where T_{ij} is predicted flow, A_i and B_j are balancing factors that ensure that the model reproduces the observed origin and destination trip rates (O_i and D_j, respectively), and $f(c_{ij})$ is a distance impedance factor that limits trips to match an observed empirical trip length. In this special case where both row and column totals need to be bal anced, an iterative process akin to biproportional matrix adjustment is used to solve for the flows as a function of the main variables (O_i D_j and $f(c_{ij})$).

Recent Extensions and Elaborations

Access connotes proximity to other places and activity centers. Whether access then has a 'diminishing' effect on flow (due to competition) or an 'augmenting' effect due to complementarity is an empirical question. In the first case, openness to competition comes from access and relative location near to other places; for example: an 'open economy' is prone to suffer whatever turns of fortune should befall their larger neighbor(s). So accessibility can mean vulnerability to competition. On the other hand, access can promote larger markets and spillover effects from adjacency to opportunity. The balance of these forces is an important question; the answer can determine whether a place benefits from proximity to other complementary destinations or whether it is downgraded because of loss to close com peting destinations. (This is a theory that has been extensively elaborated by Fotheringham.)

The spatial structure of destinations and the relative location of opportunities on this surface are important facts. Not only does one need to know what is close to a particular location, but before drawing conclusions about the interaction with any one specific place, one ought to reckon which (if any) alternatives are interspersed between the origin and destination. Having made that determination, one would then gauge the impact of 'intervening opportunities'. For example, if we consider consumer choice of branch banks (places to transact retail over the counter banking), we are aware that the consumer faces an array of spatial alternatives: clearly distance is but one attribute of their spatial choice, and empirical observations confirm that people often pat ronize a facility other than the nearest. This of course is impacted by the convenience orientation of the product; traditionally, low order retail goods (central place the ory) are highly convenience sensitive, and so the pre ponderance of trips might be expected to be handled by the nearest retail service point (with resultant Voronoi diagram for trade areas). This is even the case for gro ceries though one has observed some tendency to bypass the nearest alternative on some occasions. Such decisions might arise from other contingencies in the trip maker's travel agenda or may result from the desire to accom modate other destinations and activities as part of a multipurpose tour. As one goes to higher order retail goods and services, the hold of distance and convenience becomes even more tenuous as the sole factor, and such important processes as 'variety seeking' behavior, or complex trip chaining begin to trump the basic idea of going to the nearest source.

In all these instances, the SI model permits a degree of dispersion even from the same origin to a variety of destinations, in a way that allows for realistic represen tation of trade areas and their resultant geographic expression. In other words, a person's activity space depends on where the destinations are relative to their travel area. Further, the new spatial choices could come and go both as a result of dynamics on the supply side (the stores open and close and rearrange their positions) and as a result of the changing residential and travel landscape in the city. In fact, some of the remarkable transformation of the supply of retailing in the city has been capture mobile patrons at accessible nodes, but also to create nodes of accessibility to draw patrons toward new destinations.

Without added consumer behavioral data, however, these spatial analyses are devoid of some of the more important social processes that are critical to the oper ation of an urban area. It remains to the analysts to enliven these descriptions with reasons 'why' we need to consider them and explanations (if possible) for their importance. Thus, we are not surprised to hear the contrast between 'core' and 'periphery' meaning the difference between isolated locations, and the most central accessible parts. Recognize too that one region's periphery may be very much more 'central' when we adjust the scale. (Growth poles are supposed to be places where the advantages of agglomerated activities promote further impetus to grow.)

So as d_{ij} increases, we take it as axiomatic that the impact of the activity or entity at j, on i, decreases. At exactly what rate this decline occurs, and whether it is smooth, or whether is suddenly drops off to zero after a certain radius is what allows for a good range of flexibility in representing realistic geographical processes. To take an extreme example, one would not wish to include the population of Santa Barbara, California in the growth and population potential measures for Buffalo, New York. But what is a correct distance weighting? That question has fascinated geographers interested in spatial interaction models for several decades.

Applications

Given this broad scope, it is not surprising that appli cations of SI concepts include migration, commuting, communications, retail consumer behavior, and inter regional trade. SI largely focuses on the beneficial (i.e., desired or positive) effects of connecting separated activities. Economists and demographers interested in migration, and transport geographers interested in intra urban mobility, think of spatial interaction as the outcome of 'where a place is relative to what is going on around it', and the resultant flows then become a target of inference and explanation for analysts. Spatial interaction is also a source of complex phenomena such as spatial interdependence and spatial externalities; these later effects require careful adaptation of standard econo metric methods. Note too that interactions are not always 'positive goods', as exemplified, for example, by the form of negative spatial externality displayed in congestion on highways, or pollution from moving sources.

Modelers are also interested in the potential for interaction between places based on their relative lo cation. Such measures of traffic generation are clearly of great importance, especially as a factor in economic de velopment. Accessibility to a national market, or a claim that a particular location has a large population within easy access range are commonly used to report a city's merits as say an industrial location, or a theme park, or an expansion franchise for major league sport. Potential market access, in terms of people within fixed proximity to an attraction, is often a preliminary screen used to gauge site suitability. Access in this sense can become a self reinforcing magnet for further migrants, when one considers processes such as family and kinship networks reinforcing the information that prospective migrants have about a destination. This aspect of population growth altering the accessibility to markets is a crucial dynamic element of spatial interaction in changing urban systems. A long standing concern of demographers has been to document the impact of population redistribution on aggregate measures of segregation, population het erogeneity, or the general mixing of people with various degrees of mobility (movers and stayers). Finally, a powerful application of the concept of spatial interaction leads to a measure of market size for the optimal location of a proposed retail site: as potential interaction reinforces the suitability of a site to accommodate a new business, or for expansion of an existing business.

Limitations

Despite the rich inclusion of spatial interaction factors there are possibly missing effects that are influential on interactions, and these topics are of considerable ongoing research. Among these would be impacts of the relative location of surrounding nodes. Intrazonal flows also pose a recurring puzzle in interaction modeling, as the zonal arrangements impact the internal interactions. This can be quite consequential for traffic models in cities. Once a zoning scheme is set, there are inevitably intrazonal flows; to the extent that it is possible to have the bulk of the interactions being interzonal, the model is thought to be better specified. In the context of what is referred to as the modifiable areal unit problem, the size of self contained flows can be enormously consequential for model accuracy. The possibility of designing the optimal zonal scheme for modeling interactions is currently a topic of research.

Inadequate treatment of activity patterns and multi purpose trips in aggregate interaction models have di minished their relevance for travel demand modeling, and instead the spatial interaction approach has been eclipsed somewhat by behavioral approaches. To the extent that the SI model is part of a larger system, the simplifying assumptions may lead to cascading errors, reinforcing the likelihood that interactions and travel impedance are in need of simultaneous estimation, re quiring a dynamic, stochastic, equilibrium network flow.

See also: Location Analysis; Network Analysis; Space II.

Further Reading

Alonso, W. (1978). A theory of movement. In Hansen, N. M. (ed.) *Human Settlement Systems*, pp 197 211. Cambridge, MA: Ballinger.
Fotheringham, A. S. (1983). A new set of spatial interaction models: The theory of competing destinations. *Environment and Planning A* 15, 15 36.
Fotheringham, A. S. and O'Kelly, M. E. (1989). *Spatial Interaction Models: Formulations and Applications*. Dordrecht: Kluwer Academic.
Greenwood, M. and Hunt, G. L. (2003). The early history of migration research. *International Regional Science Review* 26, 3 37.
Isard, W. (1960). *Methods of Regional Analysis*. Cambridge, MA: MIT Press.
Nijkamp, P. and Poot, J. (1987). Dynamics of generalized spatial interaction models. *Regional Science and Urban Economics* 17(3), 367 390.
Pooler, J. (1994). An extended family of spatial interaction models. *Progress in Human Geography* 18(1), 17 39.
Roy, J. R. (2004). *Spatial Interaction Modelling: A Regional Science Context*. Berlin: Springer Verlag.
Taaffe, E. J., Gauthier, H. L. and O'Kelly, M. E. (1996). *Geography of Transportation*. Englewood Cliffs, NJ: Prentice Hall.
Wilson, A. G. (1971). A family of spatial interaction models, and associated developments. *Environment and Planning* 3, 1 32.
Wilson, A. G. (1974). *Urban and Regional Models in Geography and Planning*. New York: Wiley.

Spatial Interpolation

N. S. Lam, Louisiana State University, Baton Rouge, LA, USA

Glossary

Areal Interpolation A mathematical procedure used to interpolate data values defined in one set of aggregated boundaries to another.

Kriging A spatial interpolation method named after D. G. Krige who introduced the method in the early 1970s.

Point Interpolation A mathematical procedure used to estimate values from a set of sample points with known values to locations with unknown values within a study area.

Pycnophylactic Interpolation An interpolation method introduced by W. R. Tobler in late 1970s for volume-preserving isopleth mapping.

Spatial Interpolation A mathematical procedure used to estimate values from locations with known values to locations with unknown values within a study area.

Variogram A diagram depicting the relationship between variance and distance for a set of sample points.

Introduction

Spatial interpolation is used to estimate values at locations whose values are unknown based on a sample of locations with known values. Before the 1980s, the spatial interpolation problem was confined mostly to point interpolation, where the sample data are in the form of point locations. Goodchild and Lam expanded the traditional notion of spatial interpolation by incorporating areal interpolation, where data defined in one set of areal boundaries (polygons) are used to estimate values of areas defined using another set of boundaries. Together, point and areal interpolations have since become a major tool for many scientific investigations. Hence, the spatial interpolation problem can be broadly defined as follows: given a set of spatial data either in the form of discrete points or aggregated subareas, the spatial interpolation problem is to find the function that best represents the whole surface so that prediction of values at other points or subareas within the study area whose values are unknown can be obtained.

Spatial interpolation is a crucial technique in many disciplines and has been utilized in a wide range of applications. Perhaps the most common and earliest use is in cartography, where point interpolation is employed to create a continuous topographic surface from sampled elevation points. The interpolated topographic surface is usually mapped in the form of contour lines or three dimensional block diagrams. In climatology, spatial interpolation is an essential routine for estimating temperature and other climate parameters for the entire globe from a network of weather stations. Spatial interpolation techniques have often been used to estimate air pollution parameters such as SO_2, radon gas, and particulate matter (PM) concentrations using samples from monitor stations. In physical geography and geological applications, spatial interpolation is the key method for estimating subsurface geological strata and mining potentials. Similarly in human and economic geography, spatial interpolation is frequently applied to estimate population density, market potential, income potential, health parameters, voting behavior, and crime patterns.

With the development of geographical information systems (GIS) technology, the interpolated surface data can be stored in digital form called digital elevation models (DEM) or digital terrain models (DTM) (even though the variable of interest is not necessarily elevation). These digital data can further be used to calculate slopes, aspects, and volumes, and/or integrated with other data for more comprehensive analysis such as landscape planning, visibility assessment, route planning, and cut and fill engineering projects. This infusion of GIS and digital technology has opened up lots of application potential and makes the interpolation process even more indispensable and popular than before.

As in many other spatial estimation methods, spatial interpolation is subject to error, the magnitude of which depends on a number of factors. First and foremost, the mathematical function used in every spatial interpolation model to estimate unknown values is only an assumption; it may not be the best in approximating the underlying surface, thus resulting in error. Many interpolation methods have been developed, as well as refined and extended, in an attempt to find the best fit function that represents the underlying surface so as to increase interpolation accuracy. The second major source of error arises from the fact that spatial interpolation deals with spatial data that have x and y dimensions; the nature of the problem is thus different from nonspatial interpolation such as interpolation of time series data. The spatial interpolation process will need to consider spatial proximity or neighborhood effects. The added dimension makes the interpolation process more complicated and subject to ambiguity, and factors other than the mathematical function itself, such as the grid size used during the interpolation process and the spatial distribution of samples, will have great impacts on the resultant interpolation error.

With the availability of spatial interpolation routines in GIS and other mathematical software, spatial interpolation nowadays is simple to use and has been frequently applied. However, many people perform spatial interpolation without realizing that there could be significant errors involved in the estimates. Moreover, some methods may yield more accurate results than others for different types of data surfaces. An understanding of the major methods in spatial interpolation and their error sources should help in improving the accuracy of the interpolated estimates, as well as raising the awareness of the errors involved in subsequent analysis that is based on the interpolated estimates.

In the following, a classification of spatial interpolation methods is described, which should help in putting the various methods in perspective. Then, major methods in each class are briefly explained. Factors affecting the interpolation accuracy are outlined, followed by some suggestions toward future research.

A Classification of Spatial Interpolation Methods

There are many methods for spatial interpolation, and a classification of these various methods according to some schemes is necessary to better understand the method itself and its relationship with the data. Using Lam's classification framework developed in 1980, spatial interpolation includes both point and areal interpolation (**Figure 1**). Point interpolation deals with data collectable at a point, such as temperature readings or elevation, whereas areal interpolation deals with data aggregated according to some boundary delineation, such as population counts by census tracts.

Point interpolation is further divided into 'exact' and 'approximate' methods according to whether they preserve the original sample point values. Exact methods are employed when there is high confidence of the accuracy of the original sample data, whereas approximate methods are preferred when the original sample data are thought to have some noise. Other classification schemes for point interpolation have been used in the literature. For example, some classify the methods according to the spatial extent into either 'global' or 'local' methods. Global methods utilize all sample points in determining value at a new point, whereas only nearby points are used in local methods. Another classification scheme distinguishes point interpolation methods according to whether it is a statistical technique or not. As such, Kriging is in the group of statistical technique, and the remainder of point interpolation methods is in another group. This article adopts the classification that categorizes methods into 'exact' or 'approximate', because the property of preserving the original sample point values on the estimated surface seems fundamental in analyzing accuracy and in examining the nature of interpolation methods.

Common examples of the methods of the 'exact' type include interpolating polynomials, most distance weighting methods, Kriging, spline interpolation, and finite difference methods. Examples of the group of 'approximate' methods include power series trend surface analysis, Fourier models, distance weighted least squares, and least squares

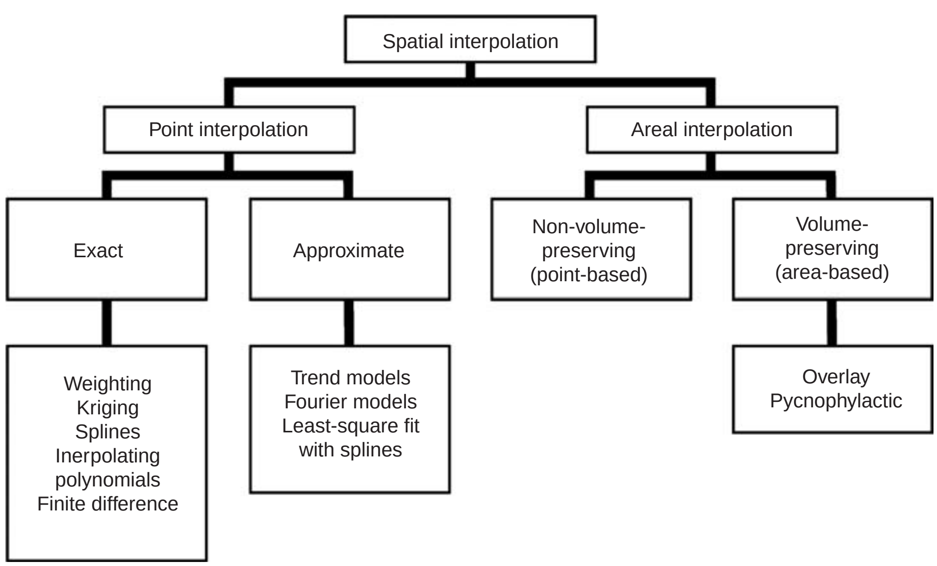

Figure 1 A classification of spatial interpolation methods. Adapted from Lam, N. S.-N. (1980). *Methods and Problems of Areal Interpolation*. PhD Dissertation, University of Western Ontario.

fitting with splines (**Figure 1**). These examples reflect the most generic version of the method itself. It is noted that there are many variations and extensions of the generic versions, with some methods being a combination of different basic methods. Moreover, other criteria can be used simultaneously to help understand a method. For example, an 'exact' method can also be global or local, statistical or nonstatistical.

For areal interpolation, two different approaches can be used. Following the terminology by Goodchild and Lam, the geographic areas for which data are available are called source zones and those for which data are needed are called target zones. The non volume preserving approach, which is an older approach, does not preserve the original source zone data values, whereas the newer approach, the volume preserving approach, does. Again, similar to point interpolation, the preservation of original source zone values is considered an important property for accuracy assessment. It has been shown in subsequent research that those methods that can preserve the original source zone values are far more accurate than those that cannot. Examples of non volume preserving methods include a number of point based interpolation methods, whereas examples of volume preserving methods include polygon overlay and pycnophylactic interpolation. Again, these are basic versions of the methods, and variations and extensions of these two methods have been made to compensate the shortcomings existed in each method.

Point Interpolation

Grid versus TIN

Since sample points are most likely irregularly spaced, prior to point interpolation, a decision is needed to make on whether the interpolated values will be in the form of regular grid cells or a triangulated irregular network (TIN) (**Figure 2**). For the 'grid' approach, a mesh of grids will first superimpose on the study area. A point interpolation method is then applied to interpolate from irregularly spaced sample points to regularly spaced grid centers or intersections. The size of the grid is often arbitrarily determined, which has great impact on the accuracy of resultant grid estimates, as discussed below.

The TIN approach links the irregularly spaced sample points into a set of triangles. Interpolation will be directly made on any point along the edges of the triangles. This approach was frequently used to produce contour maps and shaded relief maps, as it is more direct and generally faster in computation. However, inconsistencies arise because there are many ways of making the triangulation of the same set of points. Furthermore, for different sets of points, different TINs will result. This makes the TIN approach less desirable

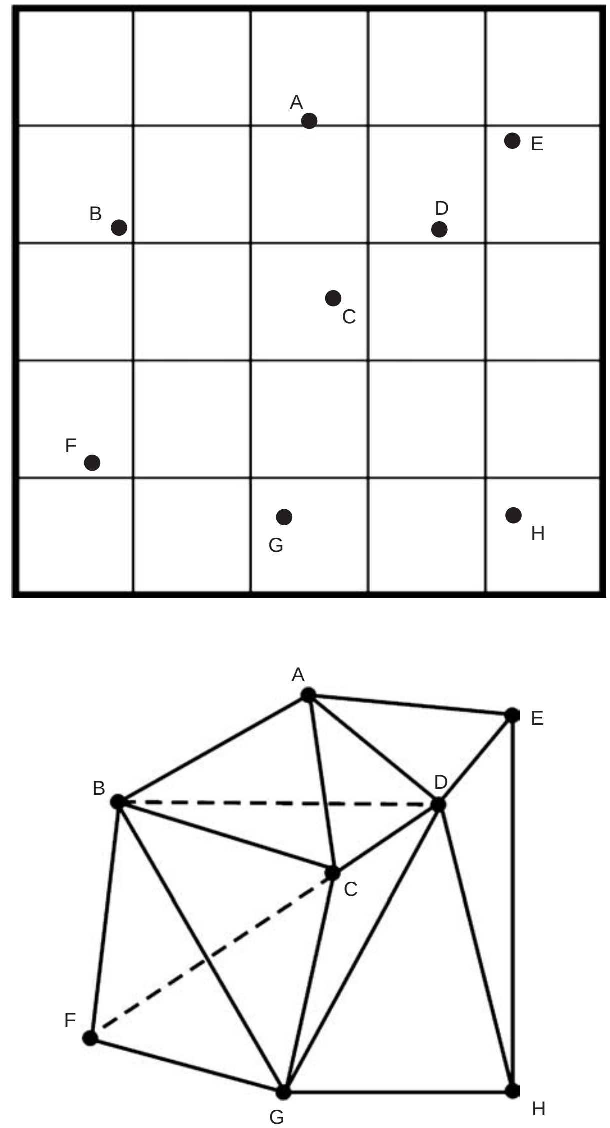

Figure 2 Grid vs. TIN configurations for eight sample points. Note that TIN configuration is not unique; alternative triangles can be connected between two points (e.g., dotted line $\overline{FC}$ instead of $\overline{BG}$). Adapted from Lam, N. S.-N. (1980). *Methods and Problems of Areal Interpolation.* PhD Dissertation, University of Western Ontario.

because of the difficulty to compare from one surface to another. Integrating the interpolated surface data with other types of data in a GIS setting using TIN configuration would be impossible. The grid approach enables comparison of different surfaces by using the same grid size; hence, it is often preferred for point interpolation.

Exact Methods – Distance-Weighting

Distance weighting method is probably the most commonly used point interpolation method, as it is flexible, easy to understand, and easy to use. Given a set of N data points, the distance weighting method is to assign more weight to nearby points than to distance points. This follows the logic of spatial autocorrelation, spatial de pendency, and the Tobler's law of geography, where things close by are more related than things located farther apart. The general mathematical expression for computing the value at a grid center (x, y) is

$$f(x, y) = \frac{\left[\sum_{i\ 1}^{N} w(d_i) z_i\right]}{\left[\sum_{i\ 1}^{N} w(d_i)\right]} \qquad [1]$$

where $w(d)$ is the weighting function, z_i is the data value at point i, and d_i is the distance from point i to (x, y).

The main variation for this method is the weighting function. Unfortunately, the choice of a weighting function is often arbitrarily made, and it may or may not represent the underlying surface well. Some weighting functions will place far more weights for the points nearby and very little weights for the points farther, resulting in a more varying surface, such as the family of inverse distance functions (e.g., $w = d\ ^{2}$). Other weighting functions may place a less drastically reduced weight for distant points, such as a linear weighting function (e.g., $w = a - bd$), thus resulting in a smoother surface. It is important to note that the weighting method can be 'exact' or 'approximate', though most of the functions used in geographical applications lead to 'exact' results. For those weighting functions where $w(0) = \infty$, such as $w = d\ ^{1}$, the weighting method will give the exact value of the original sample points. On the other hand, for a negative exponential weighting function, such as $w = 2e\ ^{d^2}$, the method will only approximate the original values. Moreover, the distance weighting method can be made either a 'local' method, where only nearby data points are used to estimate the value at the grid center, or a 'global' method, where all data points in the study area are utilized to estimate a grid value.

The simplicity and flexibility of the distance weight ing method make it a very popular method for point interpolation. However, in addition to the problem of selecting the appropriate weighting function, the dis tance weighting method is subject to two additional disadvantages. First, the method is easily affected by uneven distributions of data points, since an equal weight will be assigned to each of the points even if it is in a cluster (**Figure 3**). This problem has long been recog nized and can be handled by additional algorithmic ma nipulation, such as averaging all the points or selecting a single point to represent the cluster. Second, as a weighted average method, it has a smoothing effect, and the resultant interpolated values will be bounded by the maximum and minimum values of the samples. This means that if the original sample points do not capture the underlying true maximum and minimum surface values (e.g., mountain top and valley bottom), then the interpolated surface will never reach these maxima or minima.

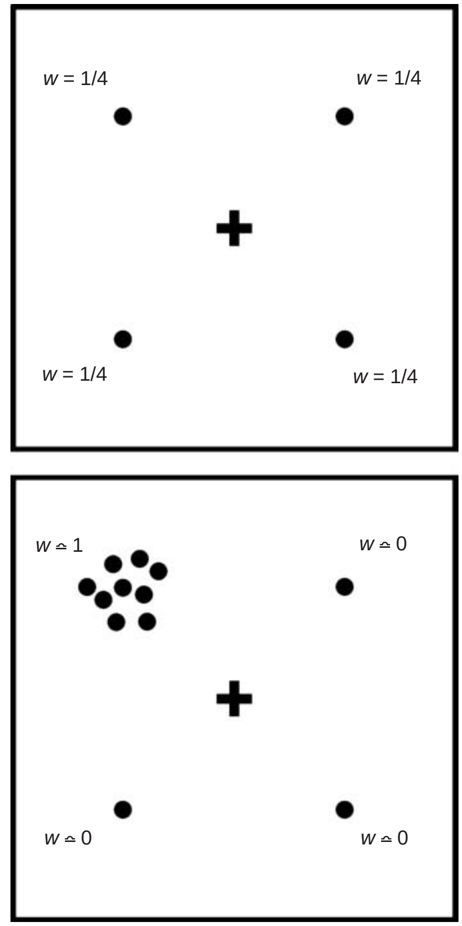

Figure 3 Effects of clustered data points on distance-weighting methods. Adapted from Lam, N. S.-N. (1980). *Methods and Problems of Areal Interpolation*. PhD Dissertation, University of Western Ontario.

Exact Methods – Kriging

The term Kriging is named after D. G. Krige, who introduced the use of the technique for mining appli cations in the early 1970s. Since then, Kriging has been extended from simple Kriging to other variations such as universal Kriging, block Kriging, co Kriging, and others. Applications of Kriging have been increasing, especially now that the technique is available in popular GIS and statistical software packages. Kriging is a statistical technique, which is different from the rest of the point interpolation methods. It is the main method in the field of geostatistics.

In its simplest form, Kriging assumes that the surface to be interpolated is a statistical surface that has a certain degree of stationarity. The structure of the surface, also termed regionalized variable in Kriging literature, can be represented by its variogram, which is a function

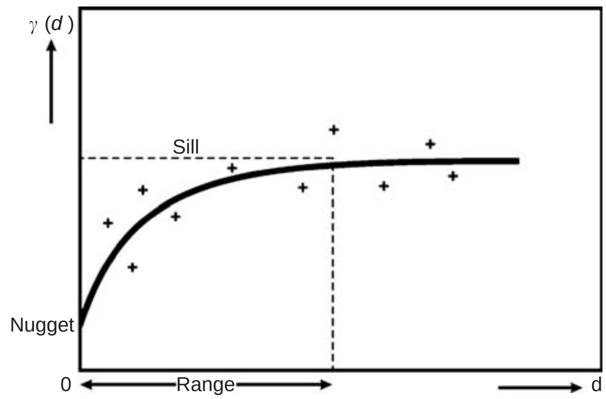

Figure 4 A hypothetical variogram function estimated from observed data points. Adapted from Lam, N. S.-N. (1980). *Methods and Problems of Areal Interpolation*. PhD Dissertation, University of Western Ontario.

depicting the relationship between the variance of the sample points and their distance (**Figure 4**). Mathematically, the variogram (2γ) or semivariogram (γ) is defined by

$$\gamma = \frac{1}{2}N\sum_{i=1}^{N}\left[z_{(x_i+d)} - z_{(x_i)}\right]^2 \qquad [2]$$

where d is the distance between two sample points. This function (variance) is expected to increase with increasing distance between two points. But just like many geographical phenomena, the increase is expected to taper off at a distance 'range', reaching a level called 'sill'. Sometimes, there is a small nugget effect, where samples taken from the same location (zero distance) may not have exactly the same value.

Based on the empirical estimated variogram function, points that have unknown values can be estimated as a weighted average of the sample point values according to their distances to these points. The system of linear equations to solve for the weights λ_j for an unknown point A is

$$\sum_{j=1}^{n}\lambda_j\gamma(d_{ij}) + u = r(d_{\mathrm{iA}}) \ \text{ for all } i = 1, n \qquad [3]$$

$$\sum_{i}^{n}\lambda_i = 1 \qquad [4]$$

where $\gamma(d_{ij})$ is the semivariogram function value for the distance between sample point i and j, $\gamma(d_{\mathrm{iA}})$ is the semivariogram function value between point A and sample point i, and μ is the Lagrange multiplier. Once the λ's are found, the Kriging estimate of an unknown point A is

$$z^*_{\mathrm{A}} = \sum\lambda_i z_{(x_i)} \qquad [5]$$

This is an estimate that is derived using two criteria: (1) that it is an unbiased estimate; and (2) that it minimizes the expected variance. Hence, the Kriging estimate obtained is also called the best linear unbiased estimate (BLUE). The corresponding error provided for each estimate is the Kriging error. For points that belong to the set of samples, Kriging returns the original data values; hence, it is an 'exact' interpolation method.

Although Kriging is a weighted average method, it has several major advantages over the distance weighting method. First, unlike the distance weighting method where the weighting function is arbitrarily determined, Kriging uses the variogram estimated from the empirical data. The more the sample points, the more accurate the empirically derived variogram function, hence the more accurate the estimates. Second, as a statistical technique, Kriging provides an error estimate and confidence interval for each unknown point. This error information reflects the density and spatial distribution of sample points and the degree of spatial autocorrelation within the surface, and therefore is very useful in analyzing the reliability of the Kriged estimates. This proves to be very useful for future improvement of the estimates by showing where more samples are needed. Finally, the Kriged estimates are not bounded by the sample data minimum and maximum.

On the other hand, if the assumption of stationarity is not met and if the variogram function is inaccurate, then the advantages of Kriging will diminish. Subsequent empirical studies have shown that Kriging will perform the best with a large number of sample points. Moreover, even with a small number of data points, Kriging will still return with more accurate estimates than distance weighting methods. Hence, Kriging is considered to be the preferred point interpolation method over the remainder, and it should be used whenever possible.

Approximate Method – Trend Models

This method is to find the best fit polynomial function that assumes values at the data points approximately but not generally equal to the observed values. Thus, there will be an 'error' or residual at every data point. The general equation is

$$f(x, y) = \sum_{i,j=0}^{m} a_{ij}x^i y^j + e \qquad [6]$$

where a_{ij} are the coefficients to be estimated from observed data point values, e is the residual, and m is the degree of the polynomial, which is less than the total number of points, N. These methods are called trend surface models since they are often used to simplify the surface into a major trend with associated residuals. Once the coefficients are estimated, they can then be used to determine values at other points. This method is highly affected by the extreme values and uneven distribution of data points. Moreover, the goodness of fit of the poly nomial (i.e., the r^2 value) is not a true estimation of the error, but rather an underestimation (**Figure 5**).

Factors Affecting Point Interpolation Accuracy

To summarize, there are four major factors that affect the accuracy of point interpolation estimates using the grid approach, and these four factors are also interrelated. First, the size of the grid cell affects the resultant esti mates. Unfortunately, there remains no general rule on the optimal size of the grid and how it relates to the data density. A small grid size may lead to a surface that is artificially complicated, whereas a large grid size may not be sufficient to portray the real details. Second, the number of data points affects the estimates. Generally, the larger the number of points, the more accurate the estimates. However, this is not always true, as some points are more 'informative' (e.g., peaks, pits) than others. This is further complicated by the third factor, which is the spatial distribution of data points. Clustered sample points may not add much information, compared with sample points that are more evenly spread out. In fact, clustered sample points may bias the estimate, as shown in the distance weighting method. Finally, the inter polation method itself, such as the weighting functions used, will seriously affect the accuracy of the estimates.

Areal Interpolation

Areal interpolation, coined by Goodchild and Lam in 1980, was regarded as a variant of the general spatial interpolation problem. As GIS applications increase, the need to integrate various types of data increases. Areal interpolation has been used to transform data defined in one set of boundaries to another, such as census data to political boundaries. It has also been used to transform time series data where boundaries change over time to enable analysis. Lately, remote sensing images as well as other data layers have also been utilized to improve the accuracy of the interpolated estimates. The following describes the two most generic approaches using the classification scheme, non volume preserving and vol ume preserving.

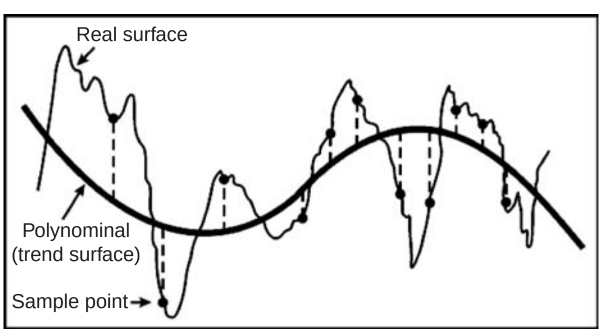

Figure 5 An example of underestimation of true variance at unknown point. Adapted from Lam, N. S.-N. (1980). *Methods and Problems of Areal Interpolation*. PhD Dissertation, University of Western Ontario.

Non-Volume-Preserving Methods

Consider the areas for which data are available as source zones, and the areas for which data are needed as target zones (**Figure 6**), the non volume preserving approach generally proceeds by overlaying a grid mesh on the study area and assigning a control point to each source zone. A point interpolation is applied to interpolate the values at each grid. Then, the estimates of the grid points

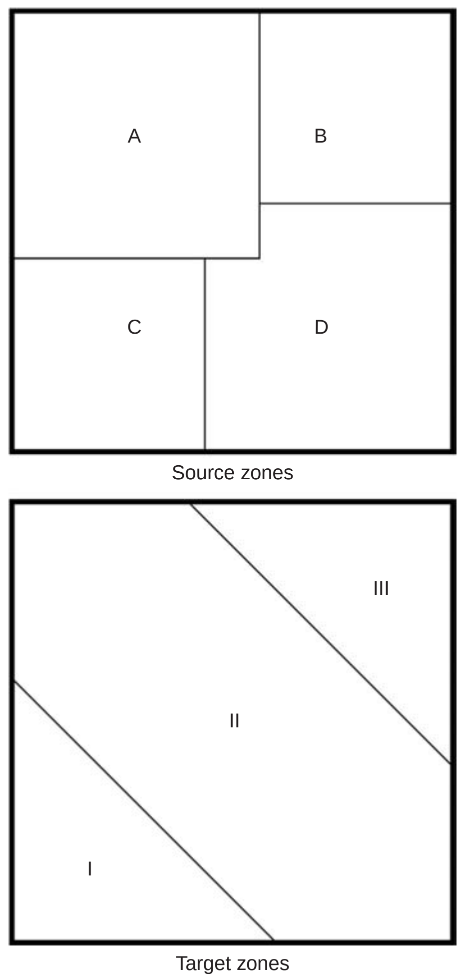

Figure 6 Hypothetical source and target zones. Adapted from Lam, N. S.-N. (1980). *Methods and Problems of Areal Interpolation*. PhD Dissertation, University of Western Ontario.

are tabulated and averaged according to the new target zone boundaries, yielding the final target zone estimates.

In this approach, the major variations between the numerous methods are the different point interpolation models used in assigning values to grid points. This ap proach suffers the same sources of error as point inter polation methods. It does not guarantee that the original source zone values are preserved; hence, it is called non volume preserving or point based approach.

Volume-Preserving Methods

In this approach, the original source zone values are preserved. Also, the zone itself is now used as the unit of operation instead of the arbitrarily assigned control point, and no point interpolation is involved. Two basic meth ods utilizing this approach can be distinguished.

The 'overlay' method, also called area weighting method, superimposes the target zones on the source zones. The values of the target zones are then estimated as a weighted sum or average based on the size of overlapping areas. For example, if target zone II has 70% overlapping area with source zone A, then 70% of the population in source zone A will be assigned to target zone II. The equation is

$$V_t = \sum_s U_s a_{ts} / \sigma_s \qquad [7]$$

where V_t is the value for target zone t, U_s is the source zone value, a_{ts} is the area of overlaps, and σ_s is the total area of source zone s. There will be minor variations of the formula, depending on whether the variable repre sents counts or density.

The 'pycnophylactic' method was originally de veloped by Tobler in 1979 for isopleth mapping. Lam later expanded its uses in areal interpolation. The method starts by superimposing a grid mesh on the polygons and assigning the mean source zone value to each grid within the zone. Then, a number of iterations will be carried out to modify the grid values according to two criteria: smoothing and volume preserving. An ex ample of the smoothing condition is to change each grid cell value into the average of its four neighbors. The volume preserving condition is enforced by summing all the grid values in each source zone and then decreasing/increasing all grid values within a zone if the total is above/below the observed source zone value. The final step is to reaggregate the grids according to the target zone boundaries to obtain the target zone estimates.

Factors Affecting Areal Interpolation Accuracy

The quality of areal interpolation estimates depends largely on how the source and target zones are defined and aggregated and the characteristics of the partitioned sur face. Judging from the theoretical and subsequent empir ical studies, the volume preserving approach is far more accurate than the non volume preserving approach; hence, the former should be used whenever possible. Within the volume preserving approach, the overlay method assumes spatial homogeneity within each source zone, the more homogeneous the source zones, the more accurate the estimates. On the other hand, the pycnophylactic method assumes a varying smooth surface. Accuracy of the pycnophylactic estimates hence depends on whether this assumption of the underlying surface is true or not.

Subsequent improvements and extensions of these two basic versions of areal interpolation hence focus on im proving the knowledge of the source zones by using add itional layers of information such as remote sensing images. These images can be used to identify different land cover areas within source zones so that the area weighted estimates are more accurate, a procedure that is similar to an old cartographic practice called dasymetric mapping.

Future Directions

Spatial interpolation encompasses a class of methods that are crucial to producing estimates for the analysis of spatial phenomena. Many controversial studies have based on spatial interpolation estimates to derive conclusions without realizing the interpolation error involved in these estimates. A case in point would be the current debate of global warming. While there are many solid scientific studies that document gradual increase in global tem perature, the exact magnitude of increase through time and space will need to be carefully interpolated and calibrated using a better interpolation methods such as Kriging or its extension. Another case in point would be for estimating the air pollution PM concentration level through interpolation; the resultant estimates are often used to link with human health, and the conclusions of such studies are often controversial and subject to scrutiny.

Although there have already been numerous studies on interpolation error, it is suggested that future research should focus on investigating the causes and con sequences of interpolation error. The emphasis of future studies should be more on developing realistic strategy to manage interpolation error, with an understanding that spatial interpolation error will exist no matter what. The task we are facing is to find the best strategy to mitigate and minimize such error, so that decisions and conclu sions basing on the estimates can still be made with a minimal degree of uncertainty.

See also: Kriging and Variogram Models; Overlay (in GIS); Spatial Data Models.

Further Reading

Abraham, J. S. and Comrie, A. C. (2004). Real time ozone mapping using a regression interpolation hybrid approach, applied to Tucson, Arizona. *Journal of the Air and Waste Management Association* 54(8), 914 925.

Fisher, P. F. and Langford, M. (1996). Modeling sensitivity to accuracy in classified imagery: A study of areal interpolation by dasymetric mapping. *The Professional Geographer* 48(3), 299 309.

Fisher, P. F. and Tate, N. J. (2006). Causes and consequences of error in digital elevation models. *Progress in Physical Geography* 30(4), 467 489.

Flowerdew, R., Green, M. and Kehris, E. (1991). Using areal interpolation methods in geographic information systems. *Papers in Regional Science* 70, 303 335.

Goodchild, M. F. and Lam, N. S. N. (1980). Areal interpolation: A variant of the traditional spatial problem. *Geo Processing* 1, 297 312.

Goodchild, M. F., Anselin, L. and Deichmann, U. (1993). A framework for the areal interpolation of socioeconomic data. *Environment and Planning A* 25, 383 397.

Kennedy, E. and Kennedy, L. W. (2004). Using dasymetric mapping for spatially aggregated crime data. *Journal of Quantitative Criminology* 20(3), 243 262.

Lam, N. S. N. (1980). *Methods and Problems of Areal Interpolation*. PhD Dissertation, University of Western Ontario.

Lam, N. S. N. (1981). The reliability of spatial interpolation models. *Modeling and Simulation* 12(3), 869 876.

Lam, N. S. N. (1983). Spatial interpolation methods: A review. *The American Cartographer* 10(2), 129 149.

Liao, D., Peuquet, D. J., Duan, Y. *et al*. (2006). GIS approaches for the estimation of residential level ambient PM concentrations. *Environmental Health Perspectives* 114(9), 1374 1380.

Tobler, W. R. (1979). Smooth pycnophylactic interpolation for geographic regions. *Journal of the American Statistical Association* 74(367), 519 536.

Spatial Ontologies

N. Schuurman, Simon Fraser University, Burnaby, BC, Canada

Glossary

Formalization It is the process through which ideas, text, and spatial entities are translated into digital objects so that they can be manipulated in a computer. It is the process of rendering the nondigital world into code so that it can be computed.

Interoperability It is the process of exchanging data and code between diverse computing systems and databases. It frequently involves conflation of near but nonidentical categories or computing protocols.

Ontology It means something very different in philosophy than it does in information systems. In the former, it means the question of being. In computing and information sciences, an ontology is a formal universe in which each entity is precisely defined and its relationship with every other entity in the specific categorical or computing realm is also predetermined. Ontologies in this context are the range of what is possible. They can be thought of as simply a classification system or a data dictionary.

Definition: Spatial Ontologies

The term 'spatial ontologies' is an amalgamation of two words that point to very different concepts. Spatial can be distinguished from geography as the former refers to space while the latter refers specifically to the Earth's surface. Space is a concept that has many manifestations at scales from the interior of cells to entire galaxy and ultimately the universe. Space often refers to two dimensions but can be extended to three and even four dimensions. Space can be mathematical (e.g., a set) or philosophical (e.g., epistemological implications of a conceptual framework). Space and the adjective spatial are in effect complex descriptions of a concept that embodies area at a range of scales and epistemologies.

Adding the noun ontology to the adjective spatial is a complicating strategy. Ontology is multifaceted term as it is understood very differently across disciplines. Human geographers and social scientists have interpreted ontology in an Aristotelian manner. In this guise, ontology signifies a foundational reality and is often referred to as the essence of an object or phenomena. Alternatively, ontologies answer the question: what must the world be like – or what conditions must exist – for us to perceive and know something. Computing scientists and systems theorists, on the other hand, use an engineering interpretation of the word ontology. In this view, an ontology is a complete and internally logical system (e.g., a classification system). In cartography and geographic information systems (GISs), ontology is interpreted in the information science tradition. Each ontology in this guise is a unique statement of logic, a way of describing spatial entities from one perspective or knowledge system.

Formal ontologies comprise a logical universe, and an ontology in a formal environment is equivalent to a logical theory – but is often expressed as a classification system in databases. Formal or systems ontologies have three levels of convention: representational, communication protocol, and content specification. The realm of existence is limited to that which can be represented (often in a database). Domain knowledge must be formally declared and this becomes its own world of discourse. Ontologies thus described appear circumscribed. They lack the mystery of philosophical ontologies yet this requirement of exact specification is a means of communicating different epistemologies within an information system.

Each spatial ontology is a product of an epistemology – or a way of knowing the world. There is a tradition of essentialism in information sciences which argues that the world is divided into natural classes that reflect an essential reality and these classes can be discerned. This epistemological view relies on the power of human observation to divide artifacts into natural classes. There is increasing recognition, however, that scientific classification is a mixture of metaphysics and epistemology and that the classification systems we use to represent the world are dependent on perspective, measurement techniques, and agenda. In geographic information science (GIScience), many scholars now acknowledge that context shapes the formation and selection of categories. Changes in context – or point of view – lead to a shift in perceived and/or enumerated categories. Each representation of categories leads potentially to an alternative ontology that could be implemented in a computational environment. Though there are GIScientists who would argue that categories are based on naturally occurring, stable categories, most people agree that GIS must be flexible enough to represent more than one system of classification. Multiple ontologies permit multiple stakeholders with divergent agenda to be represented in a single GIS.

In a world with multiple spatial ontologies, scientific classification is just as vulnerable to error as lay classification. Increased recognition of the role of epistemology in creating ontologies is a radical departure from objectivism which is premised on one 'real world' independent of the human mind and social context. When we talk of multiple spatial ontologies, we are incorporating recognition that each of those ontologies represents an epistemology and a point of view.

How Do Spatial Ontologies Look Like?

A spatial ontology then is a formal list of objects or terms that are permitted within a specified range set by a group. Medical data registries, for instance, contain fields (columns in the database) that are strictly controlled. Using the example of a Trauma Registry, we find a number of fields including injury date, ICD 10 codes, postal code of patient's usual residence, injury severity score (ISS), external cause code, etc. In combination, these database fields comprise a classification system. Such a classification system, like many in common use, implicitly describes the developers' epistemology – or ways of knowing and state of understanding – in that discipline. The ontology can be considered spatial because even though it describes nonspatial attributes (such as patient's date of birth), the injury location is a critical element. A Trauma Registry from a different country might include quite different fields or more commonly the interpretation or scale of the data in individual fields may vary widely. In one registry, location might refer to postal code while in another it may be entered using Global Positioning System (GPS) coordinates. Likewise, injury description may be described using a different coding system (e.g., ICD 9). One of the challenges of working with multiple ontologies is that comparing and integrating them may be difficult because each is developed based on differing assumptions about the world. And as the state of understanding in a field changes, ontologies change in subtle and often imperceptible ways.

A reasonable question at this point might revolve around the difference between a simple classification system and an ontology. In a computational interpretation of ontology, the only discernable product is the database. But before an ontology is fully developed, the stakeholders might consider the skeletal classification system the ontology. Moreover, geologists consider a map legend a fully fledged ontology while a GIScientist might consider a range of spatial entities (e.g., mountain, volcano, valley, and foothills) an ontology. And a database analyst may refer to a data dictionary as an ontology. The truth is that there is a range of acceptable interpretation of spatial ontology in GIScience as illustrated in **Figure 1**. But every ontology is linked by its relationship to a particular epistemology and its role in determining the range of acceptable categories in a classification.

Moreover, ontologies in information science are ultimately translated into formal terms that can be implemented computationally. Formalization is the process through which ideas and objects are described in terms that can be encoded digitally so that the objects can be manipulated in a computer. It is this process of formalization that separates organized but unstructured thoughts and conceptual frameworks from computational data structures populated with data. A necessary stage of formalization is the creation of succinct coherent ontologies.

Formalization and Ontologies

GIScientists frequently refer to 'formalization' as the process of rendering concepts into a form that can be ultimately represented in a digital environment. This is more easily said than done. There is a great difference between the abstract concepts that we routinely use to describe the world and the formal language – and computing code – that is required to implement technology. GIS, like all information sciences, is based on a hierarchical relationship between conceptualization and formalization. If we think of conceptualization as a step toward understanding spatial processes and relationships, there remains a pressing need to express these relationships in a mathematical or formal notation as a precursor to coding them. The trade off, however, is information loss associated with formalization. There is a danger that we suppress detail in favor of abstract and formal relationships. On the other hand, abstraction – based on information loss – is a powerful way of understanding structure.

The process of abstraction assists us in making sense of the physical and human world and, in GIS, we use abstractions to model reality. The process of formalization is based on abstraction of concepts into a symbolic form.

A mechanism for moving between cognitive impressions of reality and database representations of it is required, however, to achieve this level of abstraction. In a seminal paper by Jean Brodeur and others, the researchers describe five ontological stages associated with geospatial data and analysis. At the top level is reality (R), followed by cognitive models of reality (R^1). The third stage consists of conceptual representations and is referred to as R^2. The fourth ontological stage (R^3) converts concepts to database representations. Finally, the fifth ontological level (R^4) is the set of spatial concepts that can be produced from the database. **Figure 2** illustrates the conceptual model of ontologies.

To human geographers, such a taxonomy might seem mechanistic, but it is essential in GIS to be able to express cognitive models in a formal language as a

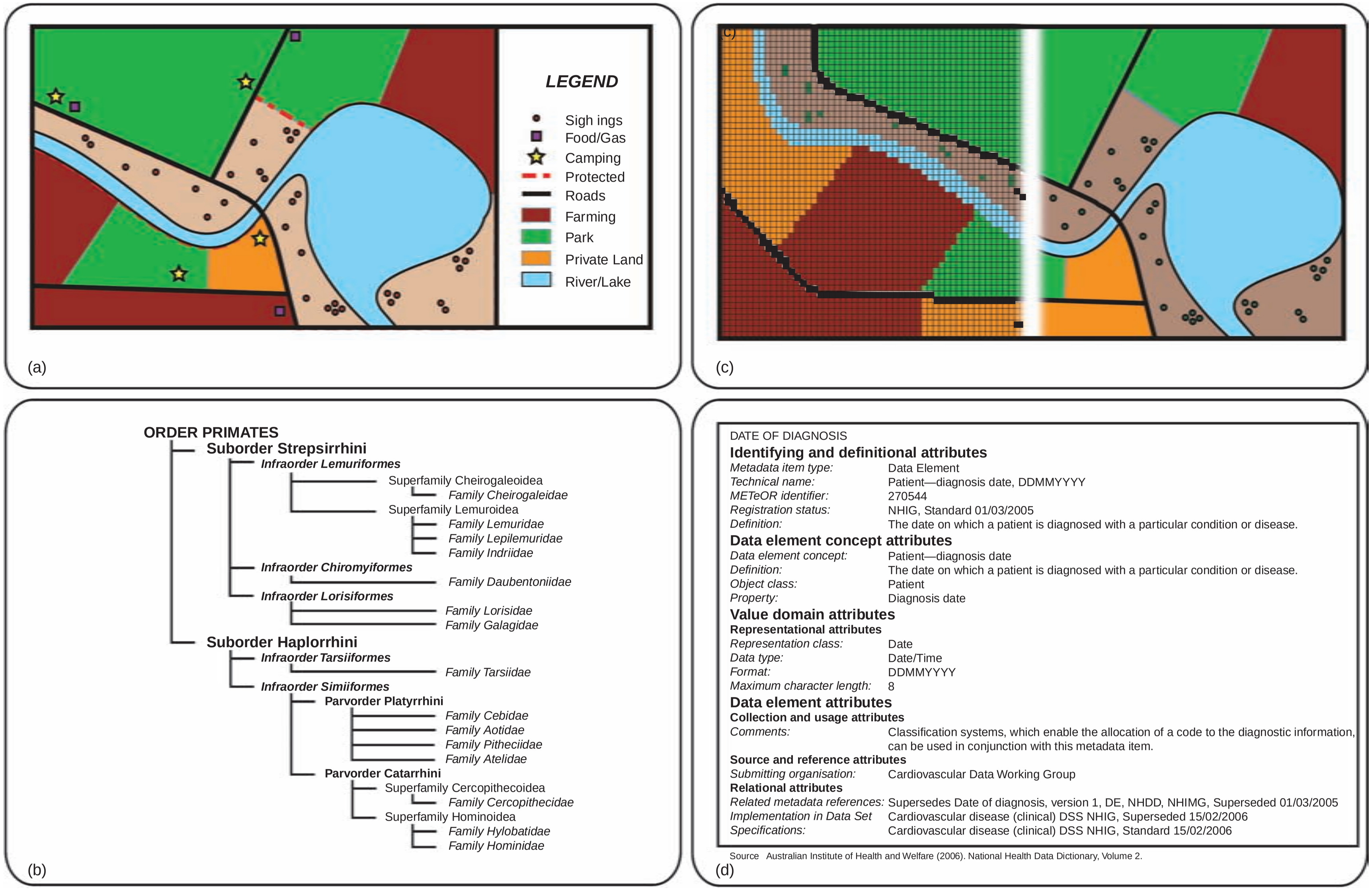

Figure 1 Mu t p e nterpretat ons of onto ogy. (a) ustrates a map egend. For many geo og sts and cartograhers, the egend represents the ent re forma un verse of an onto ogy. n (b), data mode s represent d fferent onto og es. (c) s a c ass ficat on systems or taxonomy – another nterpretat on of an onto ogy. (d) ustrates a data d ct onary that defines exact y what terms can be represented spat a y and how they are defined. Each of these are a ternat ve terat ons of onto ogy.

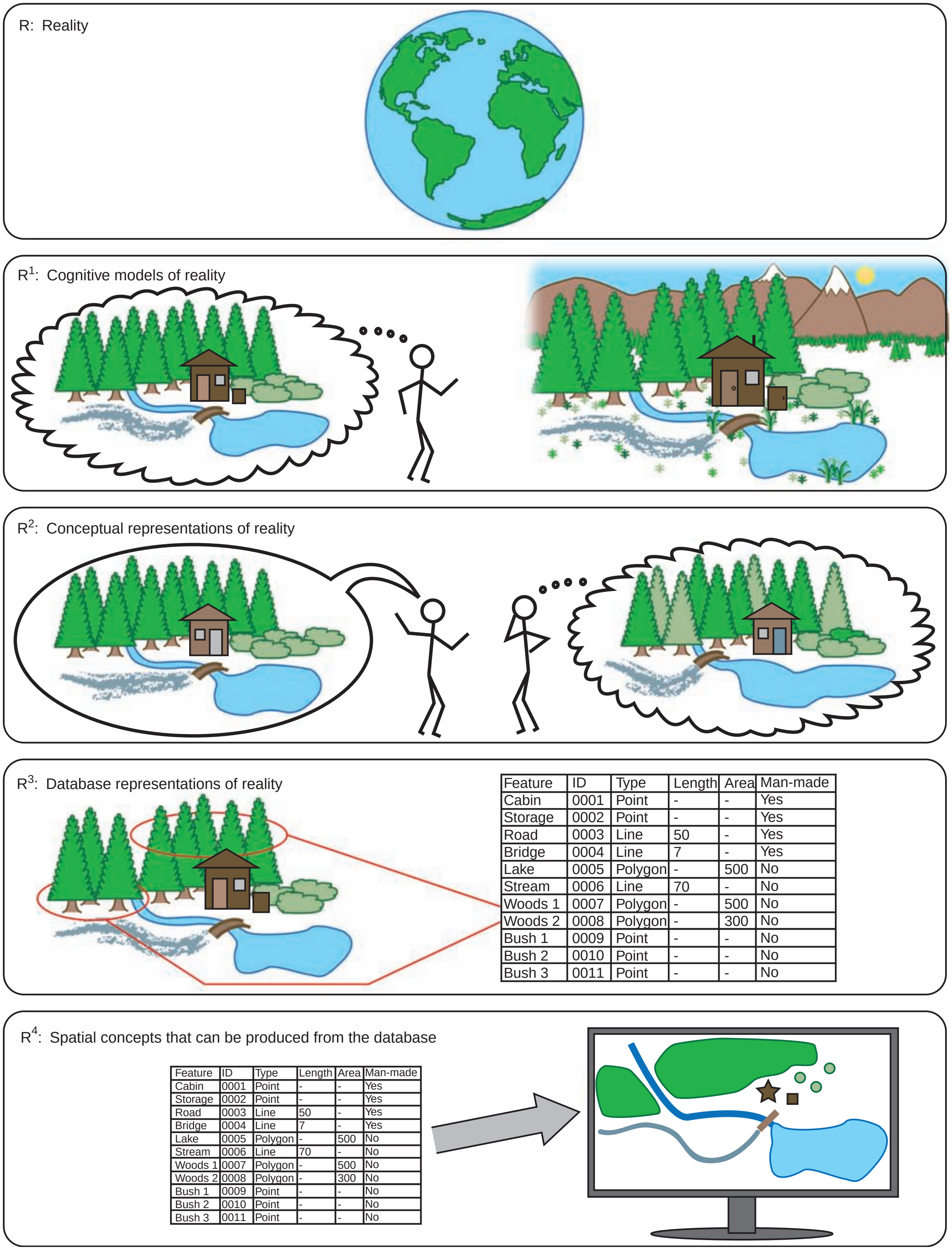

Feature	ID	Type	Length	Area	Man-made
Cabin	0001	Point	-	-	Yes
Storage	0002	Point	-	-	Yes
Road	0003	Line	50	-	Yes
Bridge	0004	Line	7	-	Yes
Lake	0005	Polygon	-	500	No
Stream	0006	Line	70	-	No
Woods 1	0007	Polygon	-	500	No
Woods 2	0008	Polygon	-	300	No
Bush 1	0009	Point	-	-	No
Bush 2	0010	Point	-	-	No
Bush 3	0011	Point	-	-	No

Figure 2 Five levels of abstraction associated with computational ontologies. The highest level of reality is the world as it is a traditional philosophical interpretation of ontology. R^1 represents the way that humans perceive the world transformed through cognition. When we conceptualize reality, we alter it again (R^2). Database representations of reality (R^3) introduce another profound transformation and the spatial concepts that can be produced from a database (R^4) are further restricted. In this illustration, we see how at each step, representations of the world are altered (and limited) as we move through different levels of abstraction. For example, the first viewer conceives the scene of the forest, house, pond, and bridge differently than she is able to convey it to the interlocutor. And representations of reality are again transformed when articulated in databases and re-articulated as spatial concepts that can be computationally displayed.

precursor to coding it. R^2 or the set of conceptual rep resentations marks the fuzzy dividing line between the conceptual and implementational realms. A structured vocabulary must be used to describe objects and object classes. Moreover, database representation requires a method of embedding context. For example, we need to know whether the word 'crown' refers to the degree to which a tree blocks light in a forest or a metal adornment intended to rest on a queen's head. In everyday (non digital) life, such meaning is extracted from verbal or environmental context. In a digital environment, it must be supplied. Context is necessary for understanding im plied use, background, and common meaning of spatial entities. Ontologies are a way of encoding context.

Ontologies can exist, be described and understood, but this does not mean that they have been formalized. Formalization is the critical step referred to by Brodeur's R^3 and R^4. Formalization allows us to take descriptive concepts and render them digital; it makes them real in computational terms. GIScientists are currently de veloping methods of formalizing and encoding diverse ways of looking at the spatial world – a critical com ponent of ontology research.

Ontology Research in GIScience

Ontology research in GIScience began in the mid 1990s when a number of scientific journal papers first explored the issue. Predictably, there were many variations in in terpretation of ontology but they were linked by the assumption that (1) we need to represent multiple epis temologies; and (2) formalization is a constraint in rep resentation at the computational level. There is a common understanding that concepts must be filtered through a formalization process. The process of formal ization is the designated arena for fixing meaning – for creating a fixed universe of discourse.

Representing multiple ways of seeing and under standing the world in a computer has been tackled by a number of researchers in different ways. One way is to use semantic networks that employ node–arc structures from graph theory to achieve this goal. To achieve this, nodes are equated with spatial entities or concepts and arcs with relationships between the entities. Close con cepts are physically close on the graph structure. The goal of the semantic proximity approach is to identify close or near concepts and label database elements (e.g., mountains or valleys) so that proximate concepts and entities can be linked. Other researchers have developed complex sets of inclusion rules that must be incorporated during classification of entities. Whether or not a spatial entity is included in a future classification can be theo retically determined by the rules. Each of these methods is linked by emphasis on including context in the development of databases – as a means of ensuring that the terms of classifications can be recognized in different settings.

The goal of maintaining ontological context is con sistent with one of the chief objectives of present day GIScience: interoperability. There is a trend toward distributed digital knowledge with different institutions housing different databases that are occasionally inte grated for comparative or global analyses. Such insti tutions, including businesses, government agencies, and university researchers, need ways to integrate spatial information even when it is described from different epistemologies at different levels of abstraction using multiple schematics (e.g., different database organiza tion). Sharing spatial data and attributes requires stability of language across culture and geography – assumptions that are seldom true. Researchers are forced to develop methods of delivering intended meaning along with field names in databases in order to allow 'intelligent' semantic sharing. Semantic interoperability implies translating concepts into code and back again. The work of semantic interoperability is enhanced by thinking of each unique set of spatial objects as part of an ontology – with its own context.

Spatial Ontologies of Land Use: An Example

Land use categories are used by municipal planners to organize and categorize public and private lands under their purview. Land use refers to the activities occurring on a particular piece of property or area of public land. Land use classification is intended to include both the intent and the reality of how a given land area within a metropolitan area may be altered by habitation and human use. Land use categories, however, vary widely among communities and most have different taxonomies (e.g., ontologies). Because it is complicated by semantic heterogeneity, integration of land use taxonomies is complicated. Each municipality collects its land use data using different assumptions about categories. These epistemological and ontological differences become en coded in the data – often with no record. Though the distinctions are often subtle (e.g., skateboard parks might be included as 'green space' in one community but ex cluded in another because they are concrete), they affect the reliability of the data. Data always contain gray zones of meaning. While they are being used in the context in which they are collected, this is usually not a problem. The meaning of green space in the dataset is implicit, for example, in the offices of individual municipal govern ments. But if that database is moved to a different municipality, workers there might interpret it differently – most likely in the same way that their community presently understands the term. When data are integrated in broader datasets, however, or sold as secondary datasets, their contextual meaning can be lost resulting in the propagation of uncertainty in the attendant analysis.

Once subtle distinctions are encapsulated with the data, users make assumptions about their meaning. For example, neighboring municipalities might integrate their land use categories to develop a regional classification. Superficial integration of the datasets based on the assumption that categories such as 'green space' or 'recreation centre' mean the same thing in every jurisdiction embeds error into the resulting merged data. The problem, in effect, is one of different ontologies. Comparing urban land use across municipal lines in the Greater Vancouver metropolitan area of British Columbia (BC), Canada, is a telling example of the problems associated with different spatial ontologies that affect everyday GIS and cartographic practices.

The Greater Vancouver Regional District (GVRD), located within the province of BC, houses over 2 million people residing within 21 municipalities. It constitutes a single contiguous metropolitan area. Land use classification is a municipal government responsibility in the GVRD. There is no regional land use classification standard, and resulting taxonomies of land use are disparate and thereby incommensurable. For example, one community may identify 'single family' as a land use designation; this designation may include two or three different types of single family types in a competing classification (see **Table 1**).

Municipalities characterize land use by employing codes to describe differences. **Table 2** illustrates the superclasses used by the neighboring cities of Vancouver and Burnaby. The superclasses differ somewhat as would be expected – but the differences are more profound than might first appear. The city of Burnaby uses detailed codes to identify land use at the scale of individual properties. They give each property outline a very specific code. For example, the numerical code range 000–099 constitutes the 'residential series' and describes various housing types and their associated forms of tenure. The city of Vancouver, by contrast, operates with many fewer subclasses – and their database is much more general as a result.

The general categories may be similar, but the problem is that there is no way of digging deeper. Fundamental ontological distinctions are a product of epistemological differences in data collection as well as varying institutional cultures in the individual municipalities. When these ontologies' differences are accounted for, one realizes that a trivial integration of the superclasses contains uncertainty. While the mapping presented in **Table 2** is schematically legitimate, it fails to convey the true (lack of) symmetry between the land use classes. The city of Vancouver bases its land use coding on municipal 'zoning', a legal classification encoded into, and mandated by, city policy (bylaws).

Zones are determined for large areas at a time, and they encompass many individual properties. Each property is assigned a land use 'code' based on which zone their property lies in. Land use information for Vancouver is based on the 'what should be there' principle.

By contrast, the city of Burnaby determines based on the activities associated with individual cadastral parcels – as determined by the taxation authorities. The city of Burnaby also determines its own municipal zoning. Unlike Vancouver, it does not consider zoning and land use as equivalent. These epistemological differences in how

Table 1 Number of different zoning designations for various categories including of industrial, low-rise residential, high-rise residential and single-family residential land use

	Industrial	*Res low*	*Res high*	*Single family*
City of Coquitlam	9	2	2	5
City of North Vancouver	7	1	1	3
City of Port Coquitlam	5	2	2	1
City of Richmond	7	1	1	1
City of Surrey	5	2	2	3

The table illustrates that the five sample communities have a varying range of subclassifications. Each land use categories and its subclassifications constitute an ontology. Schuurman, N., Leszczynski, A., Fiedler, R., Grund, D. and Bell, N. (2006). Building an integrated cadastral fabric for higher resolution socioeconomic spatial data analysis. In Riedl, W., Kainz, A., & Elmes, G. (eds.) *Progress in Spatial Data Handling, 12th International Symposium on Spatial Data Handling*. Berlin: Springer.

Table 2 Cadastral land-use classification for the cities of Vancouver and Burnaby

Vancouver 'District schedules'	*Burnaby Land-use 'series'*
One-family dwelling Two-family dwelling Multiple dwelling	Residential series
Limited agricultural	Farm series
Industrial	Industrial series
Commercial	Commercial series
- *no corresponding class* -	Utility series *includes railways, pipelines, etc.*
Cultural and recreation (*a subclass*)	Institutional series
Historic area - *no corresponding class* -	government buildings, etc.
Comprehensive development	- *no corresponding class* -

Note that the superclasses are not commensurate but can be manually synced. From Schuurman, N. and Leszczynski, A. (2006). Ontology based metadata. *Tranactions in Geographic Information Science* 10(5), 709 726.

land use is assigned in two different communities reveal ontological differences. The differences can be summarized as the distinction between 'what should be there' versus 'what is actually there'. As a result, it is difficult to directly compare land use categories between Burnaby and Vancouver. The two classification schemes are irreconcilable at a very basic level because they emerge from two different institutional contexts – in other words, they represent two separate, though overlapping, ontologies.

The example of land use reveals the impact that different ontologies have on spatial data – such as land use – that may otherwise seem unproblematic. If data users are to understand their data from a qualitative perspective, they need to think about ontologies. Are the two ontologies comparable based on their classification systems; attribute characteristics; terminology; naming conventions; rationale for collection; scale; and measurement systems? Collectively, such parameters constitute data 'context'. Context is likewise the basis for understanding how different spatial ontologies shape data.

Spatial Ontologies: Present and Future

In the past decade, GIScience researchers have become increasingly concerned with optimal methods for representing multiple epistemological perspectives about the same spatial events and entities. At the same time, there is a pressing need to attach contextual information to database elements in order to identify different ontologies in the interest of enhancing interoperability of data across institutions and geography.

Spatial ontologies enable multiple stakeholders to represent different scenarios, agenda, and interpretations of the geographic world. Formal ontologies are a means of implementing these diverse views in a structured digital environment. Acknowledging and understanding alternate perspectives on the same geographical phenomena or spatial framework is a means of preserving heterogeneity of perspective with implications for democratic policy and governance. Spatial ontologies are a means of achieving these goals in a computational environment.

See also: Critical GIS; Critical Realism/Critical Realist Geographies; Feminism, Maps and GIS; GIS and Cartography; Maps.

Further Reading

Agarwal, P. (2005). Ontological considerations in GIScience. *International Journal of Geographical Information Science* 19(5), 501 536.
Bowker, G. (1996). The history of information infrastructures: The case of the international classification of diseases. *Information Processing and Management* 32(1), 49 61.
Bowker, G. C. (2000). Biodiversity datadiversity. *Social Studies of Science* 30(5), 643 683.
Gregory, D. (2000). Ontology. In Johnston, R. J., Gregory, D., Pratt, G. & Watts, M. (eds.) *The Dictionary of Human Geography*. Oxford: Blackwell.
Harvey, F., Werner, K., Hardy, P. and Yaser, B. (1999). Semantic interoperability: A central issue for sharing geographic information. *Annals of Regional Science* 33, 213 232.
Kuhn, W. (2001). Ontologies in support of activities in geographical space. *International Journal of Geographical Information Science* 15(7), 613 631.
Schuurman, N. (2006). Formalization matters: Critical GIS and ontology research. *Annals of the Association of American Geographers* 96(4), 726 739.
Schuurman, N. and Leszczynski, A. (2006). Ontology based metadata. *Tranactions in Geographic Information Science* 10(5), 709 726.
Schuurman, N., Leszczynski, A., Fiedler, R., Grund, D. and Bell, N. (2006). Building an integrated cadastral fabric for higher resolution socioeconomic spatial data analysis. In Riedl, W., Kainz, A. & Elmes, G. (eds.) *Progress in Spatial Data Handling, 12th International Symposium on Spatial Data Handling*. Berlin: Springer.
Smith, B. and Mark, D. M. (1998). Ontology and geographic kinds. Paper read at *The 8th International Symposium on Spatial Data Handling*. Vancouver, BC, Canada.
Winter, S. (2001). Ontology: Buzzword or paradigm shift in GI science? *International Journal of Geographical Information Science* 15(7), 587 590.

Spatial Science

Ron Johnston, University of Bristol, Bristol, UK

Spatial science is an approach to human geography which came to prominence in much of the Anglophone world and a number of other countries – notably Sweden and the Netherlands – in the 1960s and 1970s, through what Burton heralded as the discipline's 'quantitative and theoretical revolution'; it has remained a significant component of the disciplinary portfolio ever since. It deals with aggregate patterns in space and their generating processes, and is characterized by two major features: the description, explanation, and (in some cases) prediction – even control – of general patterns of spatial behavior and their reflection in the landscape; and the application of rigorous mathematical and statistical procedures in that task. The approaches taken within this broad framework have changed considerably over the last half century, however, as have the methods applied.

This approach within geography is closely associated with parallel work in other social and natural science disciplines. In geology, for example, spatial science is a major component of the enterprise, and some developments within human geography have been stimulated by geologists' work. There is also considerable interdisciplinary interaction with workers in other disciplines, such as economics, sociology, political science, and archaeology, where the spatial approach – as in human geography – is one component of a broader disciplinary framework. And there are applied disciplines – notably city and regional planning – which build on those foundations.

Origins

With most substantial changes in a discipline's portfolio, harbingers of those changes which nevertheless had little or no impact on its practices can be found in its research literature; they are contributions to knowledge which were not taken up by others and so very largely disregarded. This was certainly the case with spatial science. Innovative pieces of work can be identified from the 1930s, and even earlier, which either went almost entirely unnoticed or attracted no followers among the small number of professional geographers, who chose to avoid such challenges to their established ways of working. Major change in disciplinary orientation usually only occurs when a number of individuals develop an alternative agenda to that currently dominant and promote their ways of working through a variety of aggressive political strategies. They mobilize support for their ideas and seek to change disciplinary practices – not only how research is undertaken and to what ends but also how the discipline is taught to undergraduates and presented in graduate training programs.

Such academic social movements emerged in both the USA and the UK in the late 1950s and early 1960s, taking the discipline by storm and rapidly ensuring a major shift in its foundations. In the USA, the origins of this shift have been traced by Barnes and Farish to the experience of a large number of human geographers seconded to the Office of Strategic Services (OSS) during World War II to provide information that would underpin military campaigns and their aftermath. Unlike the parallel situation in the UK where geographers worked separately, geographers in the OSS were required to work in multidisciplinary teams, in which some became aware of deficiencies in their disciplinary practices that constrained their ability to contribute to various forms of policy development. They were convinced that their discipline had to change if it was to remain viable within academia, contributing applicable knowledge – in effect, it had to become more like economics in adopting the approaches considered typical of the natural sciences such as physics. Some – such as Ackerman and Schaefer – preached this need for change: others began to implement it.

One influential change agent was Edward Ullman, who produced two pioneering essays on theoretical approaches to urban geography – one, much cited, was coauthored with Chauncy Harris. Ullman returned to Harvard after the war but following the closure of the geography department there moved to the University of Washington at Seattle. In the mid 1950s, a number of graduate students were attracted to work with him on the new theoretically based approach he was promoting, focused on the role of transportation as a major influence on human spatial behavior, and the resultant spatial patterns of areal differentiation. Ullman was away for some time, however, and they gravitated to work with another professor – William Garrison – who was committed not only to that approach but also to the application of quantitative methods. Together, through a variety of strategies and practices, that remarkably talented and influential group (including Brian Berry, Ron Boyce, William Bunge, Michael Dacey, Art Getis, Duane Marble, Richard Morrill, John Nystuen, Forrest Pitts, and Waldo Tobler) stimulated others in the discipline to adopt the 'new geography'. Their subsequent appointments to teaching positions in leading geography departments, from which in turn their students moved to other centers, facilitated a rapid spread of the new ways.

In the UK, a parallel introduction of similar approaches focused on the University of Cambridge in the late 1950s to early 1960s, stimulated by two recently appointed lecturers, Richard Chorley (with an undergraduate degree in geography from Oxford and then trained as a geologist in the USA by one of the pioneers of a new, quantitative, approach to geomorphology – Arthur Strahler) and Peter Haggett, who first encountered the sort of material that later became central to the 'new geography' when teaching economic geography at University College London in the mid 1950s. (Haggett succeeded Brian Law as a lecturer there: Law introduced location theory to undergraduates – including Brian Berry. Similar material – notably the work of Hoover on industrial location – was also taught at the London School of Economics in the 1940s–1950s, but had little wider impact.) Chorley and Haggett introduced the teaching of quantitative methods at Cambridge in 1960 and Haggett later taught a course that became his pioneering text on *Locational Analysis in human geography*. Their students – like the Washington graduates a few years earlier – rapidly spread the 'revolution' to other centers, as did those from Bristol after Haggett moved there in 1966.

Chorley and Haggett also fostered the new approach in the UK's secondary schools through five mid 1960s summer schools that were the progenitors of two seminal collections they edited together. Changing approaches to the nature of geography in British schools – thus influencing the backgrounds of those who studied the subject as undergraduates – were also strongly influenced from another source. Stanley Gregory, a London trained climatologist, initiated the teaching of statistical methods at the University of Liverpool in the 1950s, published the first textbook on the subject in 1963, and inspired a further stream of graduate students who moved into university careers. He was also influential in changing the syllabuses for the public examinations taken at age 18 ('A' level) that determined potential for university entrance, which included work through the Geographical Association. Meanwhile, John Cole, at the University of Nottingham, pioneered in the development of teaching materials for students at primary (age 5–11) schools.

From these origins, interest in quantitative and theoretical work rapidly spread, not only through those two countries but also a number of others, mostly English speaking, such as Australia and New Zealand. By the early 1970s a number of pioneering textbooks had been produced conveying the approach – mainly to undergraduates in the UK but postgraduates in the US – reflecting the different positions of geography in the two countries' educational systems. Notable and highly influential among these were Haggett's *Locational Analysis*, plus Abler, Adams, and Gould's *Spatial Organization: The Geographer's View of the World*, and Morrill's *The Spatial Organization of Society*. Harvey's *Explanation in Geography* provided a detailed philosophical and methodological foundation for the practices the pioneers espoused. Spatial science soon occupied a major role within the discipline, with compulsory courses in quantitative methods normal in degree schemes and material on location theory central in – if not dominating – many substantive courses. This provided the foundation for continued substantive and technical developments over the next four decades, and also the basis for major critiques and subsequent campaigns to change the nature of geographical practices in different directions: the 'revolution' – if 'revolution' it was – was hardly over before counter revolutions were launched.

Location Theory, Hypothesis Testing, and Quantitative Analysis

Many of the stimuli for this fresh approach to human geography came from outwith the discipline; geographers found both substantive and methodological inspiration in a range of disciplinary literatures, notably economics in the first instance. There was no coherent body of work within those disciplines that focused on the approaches taken up by geographers, but the contributions of a number of workers – mostly considered to be on the fringe of their discipline by economists – provided the initial stimuli that geographers followed up. Many – such as Hoover, Hotelling, Lösch, Palander, Alfred Weber, and also von Thünen, whose work was published in the early nineteenth century – concentrated on the location of economic activities, especially manufacturing industries. They sought to account for observed distributions not by identifying the rationales for particular decisions through empirical case studies but rather by uncovering general principles underpinning all locational decisions through theoretical development followed by hypothesis testing: model became a key word in their vocabulary.

This impetus for a 'new' approach to human geography occurred in parallel with the inauguration of regional science, led by an economist – Walter Isard – wanting to introduce spatial perspectives to mainstream economics. Isard enrolled a number of geographers in his long running, extensive campaign to gain disciplinary status for regional science; many became (and some geographers remain) active in the Regional Science Association, which has branches throughout the world. But an autonomous separate discipline represented in major universities never eventuated, and although the Association remains a vibrant interdisciplinary organization – running conferences and publishing journals – it has never threatened geography's separate existence, as some originally feared.

For the pioneering geographers, the key principle in the theories and models they were attracted to related to transport costs – reflecting an assumed importance of distance (or space) in decision making. Of the various costs involved in manufacturing products, it was the one that could be promoted as clearly geographical – a case made by Ullman and adopted by his followers. Manu facturing industries, for example, have to bring raw materials and component parts for a product together – which involves costs (both time and money) – and then distribute them to their markets, involving further costs. The more that these costs can be held down, the more efficient the production and distribution of the product and hence the lower the price to the consumer – as suming all other costs are held constant. Thus, the core geographical problem was presented as identifying the least cost location at which transport costs were min imized. Knowledge of an industry's cost structures, in cluding those of transporting its various inputs as well as its outputs, could thus be deployed to suggest where factories should optimally be located, given information on where the raw materials and other inputs could be obtained and the geography of demand for its products.

This 'ideal' was a model of the 'real world', a pre sentation of what the distribution of an industry should be in a country or region if those involved in location decisions were making the most economically rational choices. Elements of that model could then be tested empirically: hypotheses – speculations about what the world should look like – could be generated whose val idity could be assessed by, in effect, comparing a hypo thetical map of an industrial distribution against the 'real thing'. Most geographers, following Harvey's approach in his *Explanation*, chose to test their hypotheses using the verification approach, favored by positivists – is my hy pothesis right? – rather than the falsificationist approach favored by Popper – is my hypothesis wrong?

The importance of transport costs was not constrained to studies of industrial location, particularly when broadened to the more general impacts of distance – closely linked to transport costs. As geographers dis covered through searching cognate discipline literatures, others were studying the impact of distance on a wide range of behaviors. With migration, for example, it was argued that it is easier to get information regarding a potential destination for a move about a nearby place than about one further away: there is a friction of distance to information flows. As a result, people are more likely to search nearby than distant places when planning to move. They are also more likely to select a nearby des tination, because longer distance moves are more costly. The result is a postulated distance decay effect – what one sociologist, Zipf, called a 'law of least effort' – in a wide range of spatial behaviors, not only permanent migrations but also, for example, journeys to work, to shop, to a doctor's surgery, to university, to church, and so on: the greater the distance between two places, the lesser the interaction between them.

The hypothesized existence of distance decay effects could be tested by analyses of movement patterns, many of which were set in the context of the so called gravity model, using an analogy from Newtonian physics (on which Olsson provided an early introduction). According to the basic formula:

$$I_{ij} = \alpha(M_i M_j)/d_{ij}^2 \qquad [1]$$

where M_i is the mass at point i; M_j is the mass at point j; d_{ij} is the distance between points i and j; α is the co efficient of proportionality; and I_{ij} is the amount of interaction between points i and j.

In modeling migration flows between places i and j (I_{ij}), the two terms in brackets M_i and M_j might refer to the populations of towns i and j, respectively: the more people in i the more who would move to j (and vice versa); the more people in j, the more attractive it would be for people in i. And the shorter the distance between i and j (which is squared in Newtonian physics), the more people who would move between them – *ceteris paribus*.

Alongside industrial location and migration, a third area of work dominating much early spatial science was central place theory, addressing issues relating to settlement patterns. A number of nineteenth and early twentieth century scholars had explored regularities in the distri bution of central places – where consumers purchase goods and services from shops and offices – but a German geographer (with a strong background in economics), Walther Christaller, first formalized this, although his 1993 publication was not taken up by English speaking geog raphers until the 1950s. Christaller's work was also based on the 'law of least effort', assuming that people visit the nearest center offering their needed goods and services. Arguing that some baskets of goods and services are nee ded daily, some weekly, and others less frequently, Chris taller hypothesized that service centers would be organized in hexagonal patterns across the landscape (assuming no major topographical barriers), with denser patterns of small centers providing the most frequently demanded services, nested within less dense patterns of larger, less frequently visited, centers, etc. This theory – and a more complex variant developed by Lösch, which does not assume bas kets of goods/services all bought with the same frequency – was the basis for much data collection and hypothesis testing, seeking regular patterns with hexagonal foundations.

Finally among the most influential sets of ideas – though taken up a little later than some of the others – was work on spatial diffusion developed by a Swedish geographer, Torsten Hägerstrand. This, too, was founded on the 'law of least effort' – reformulated by Tobler (1970: 234–240) as "the first law of geography: everything is related to everything else, but near things are more related than distant things." If flows of information are spatially structured, ideas should diffuse outwards from innovation centers in wave like formations, thus producing ordered geographies of change. According to Morrill, Hägerstrand's initial work on migration was later extended by him and others after his 1959 visit to Seattle – to address a wide range of phenomena which have spread over landscapes, such as diseases. He also pioneered use of Monte Carlo methods to simulate patterns of spread and diffusion, and was one of the first human geographers to promote use of computers.

Most of these approaches were summarized by Garrison into six basic spatial problems:

1. the transportation problem – what is the most cost efficient flow of goods through a network of supply and demand points?;
2. the spatial price equilibrium problem – what prices for goods and services will result from such flow patterns?;
3. the warehouse location problem – what is the best set of locations for a series of supply points, to serve a given geography of demand?;
4. the industrial location problem – what is the optimum location for an industry given the geography of supply for its inputs and that of demand for its outputs?;
5. the interdependencies problem – what is the best pattern of linked plants so as to maximize profits jointly?; and
6. the boundary drawing problem – what is the most efficient set of boundaries for the hinterlands of places (such as schools), which minimizes transport costs?

Others synthesized the spatial science approach in different ways. Haggett, for example, used a diagram to identify the six major components of spatial systems (**Figure 1**), comprising: movements; the channels along which they flowed; nodes on those networks of channels; hierarchies of nodes according to their size and importance; surfaces, representing internodal patterns; and diffusions – which could flow down the hierarchies, along the channels and across the surfaces. Later work included classification and regionalization – dividing the surfaces up into relatively homogeneous areas on selected criteria.

These idealizations of spatial structures and processes stimulated a large number of empirical studies seeking ordered patterns in the landscape – point patterns of settlements and hierarchies, for example; line patterns of transportation routes; smooth (trend) surfaces of internodal land use patterns and land values; diseases diffusing down through urban hierarchies and across the surfaces – as well as ordered patterns of behavior, such as distance decay patterns in various forms of interaction. The emphasis was very much on pattern. As Bunge, Harvey, and others argued, geometry was the language of this 'new geography'. To test the geometric hypotheses, the literature of statistical applications in a range of disciplines was explored for relevant procedures – such as nearest neighbor analysis for point patterns, graph theory for line patterns, the gravity model for migration flows, and trend surface analysis. Alongside theoretical

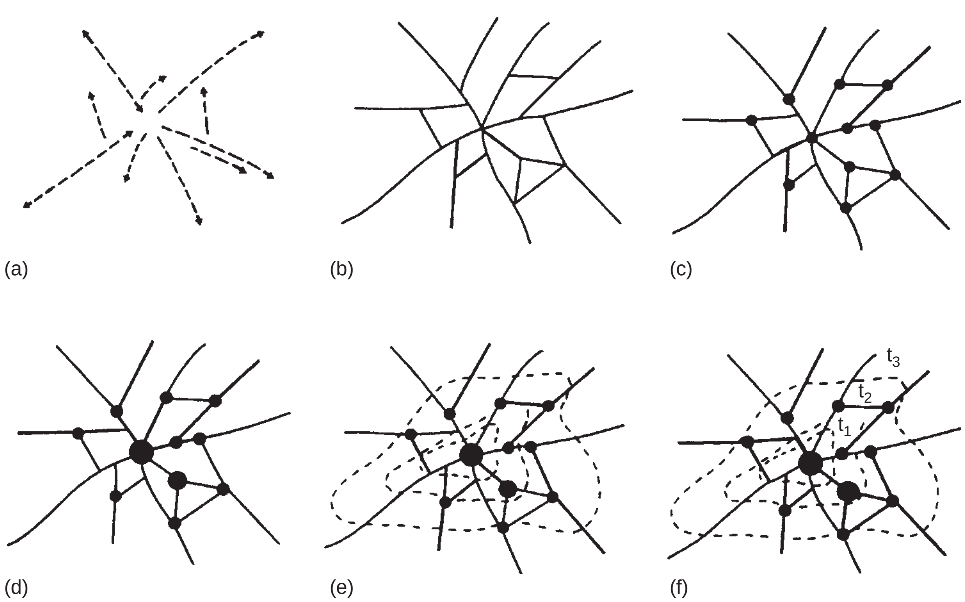

Figure 1 Haggett's representation of the six core components of a spatial system: (a) flows; (b) channels; (c) nodes; (d) hierarchies; (e) surfaces; and (f) diffusions.

exploration there was a great deal of methodological experiment in spatial science's early years.

One set of procedures came to dominate, however – the general linear model, developed in statistics for testing hypotheses regarding associations and (putatively causal) relationships, and widely applied in other social sciences. Within this still evolving family particular attention was given to regression and associated measures of correlation. Regression relates two or more variables – one of which is the hypothesized effect and the others the postulated cause(s) – to inquire how well, if at all, the causes (the independent variables) predict the effect (the dependent variable). The gravity model (eqn [1]), for example, could be expressed in a regression framework as

$$\log I_{ij} = a + b_1(\log M_i) + b_2(\log M_j) \quad b_3(\log d_{ij}) \pm \varepsilon \qquad [2]$$

In this, the regression coefficients b_1, b_2, and b_3 indicate the relative importance of the size of M_i, the size of M_j, and the distance between places i and j, respectively as influences on the amount of migration between those two places; the constant term a is equivalent to the coefficient of proportionality in Newton's formulation [1], and ε is the (normally distributed) error term. The positive signs for b_1 and b_2 indicate expectations that the volume of migration increases the larger each of the places, whereas the negative sign for b_3 indicates an expectation that the amount of migration between two places – holding constant their size – will decline, the greater the distance between them. (In Newton's model b_3 is equal to 2, but in social science applications no theoretical justification could be found for that; empirical models had to be calibrated to the datasets deployed to obtain domain specific estimates.) The associated correlation coefficient – r, or R in multiple regression with more than one independent variable – varies between 0 and 1.0: the associated coefficient of determination – r^2 or R^2 – indicates the proportion of the variation in the dependent variable that can be accounted for statistically by the independent variables.

When fitted to data, statistical models such as these provide technical answers to the question 'to what extent are variations in the dependent variable (in this case, migration) accounted for statistically by variations in one or more of the independent variables?'. (When applied geographically, they, in effect, ask questions about the correlations among two or more maps: how similar is the mapped distribution of x to those of y and z?) Such statements of association are frequently given a causal interpretation, within the context of the theory that stimulated the test: the volume of migration was generated/caused by the size of the places involved and the distances between them. Statistical relationships are thus taken as means of verifying theoretical expectations – or testing hypotheses. If the expected patterns are observed, then the postulated causative mechanisms can be assumed to operate: they are provisional statements of the processes governing spatial patterns, open – as always – to critical scrutiny and modification (perhaps even falsification) by further research.

Regression modeling – and associated techniques such as factor analysis – remain at the heart of much spatial science. But over the half century since they were introduced to human geography applications have become much more sophisticated, allowing a wider range of hypotheses to be tested and datasets of different types to be explored. Log linear modeling, logit analysis, multi level modeling, and a range of other procedures are now widely applied: all are based on the general linear model and allow similar interpretations. The testing is undertaken in one of two ways:

1. deductively, formal hypotheses regarding the size, strength, and direction of relationships among variables are stated and those expectations compared against empirical data; and
2. inductively, patterns are identified in datasets which suggest ordered behavior.

Both approaches use measures of statistical significance to assess the importance of their findings. How likely is it, that the pattern uncovered in one's dataset is a chance occurrence? Or, could it have occurred so rarely if chance were the only operating factor that one is justified in claiming there is a pattern with substantial meaning? The answers to such questions are only valid if the data meet (or at least substantially approximate) the major assumptions of the general linear model – the regression residuals should be independently distributed, for example.

Whereas a fundamental goal of the 'revolution' that brought spatial science to the foreground of human geography was to demonstrate that its approach provided not only better descriptions of the world but also reasonable explanatory accounts for what was described, this was not necessarily an end in itself. For many involved – including the Seattle group – the representations they produced were also to be deployed in changing the world. As Yeates noted when reviewing research stimulated by Brian Berry during his immediate post Seattle years at the University of Chicago, it was characterized by a strong utilitarian purpose with much of it undertaken for external sponsors, both public and private sector. Similarly, all of the modeling of urban and regional systems undertaken by Alan Wilson and his co workers at Leeds was aimed at more efficiently planned and operating cities and regions: a company he established – GMap – uses those models as the foundations for consultancy work seeking optimal locations for a range of commercial activities.

Wilson is one of the few examples of a geographer who took an alternative approach from most of the early spatial scientists. Formally trained as a mathematician, he did not test hypotheses statistically. Instead, he sought to model entire systems. Most researchers took the gravity model as the source for hypotheses about flow patterns, for example. Wilson's *Entropy in Urban and Regional Modelling* restructured it as a way of representing an entire system, concerned, for example, not just with the decisions underpinning travel behavior – as with commuting in a city – but also those regarding where homes and jobs are/might be located. His models, based on the thermodynamic concept of entropy, represent the system as a whole, and data were collected to illustrate that, rather than extract one element of the system for testing against empirical data. And then, by changing certain system elements, such as new home and job locations, he can estimate their impacts on other components – such as traffic flows.

The concept of a system attracted considerable attention among spatial scientists, largely as a rhetorical device rather than as a format for undertaking detailed research. Only Wilson and a relatively small number of others (such as Batty and Bennett) extended the interest in mathematical models, looking, for example, at catastrophe and chaos theories to explore system dynamics in space–time. This approach was marginal to most spatial scientists' concerns, however.

Moving on I – from Normative to Behavioral Theory

Many reactions to/critiques of early spatial science argued that its models were too simplistic, reducing decision making in a complex world to a single key influence – transport costs; distance was presented as a basic determinant of behavior (thereby sustaining geographers' claim for a separate place within the social sciences). The result was deterministic models of spatial patterns based on rather simplistic assumptions about spatial behavior. Although transport costs are undoubtedly important to much decision making, large and small (which shop shall I go to today for a newspaper?; where shall I locate the next supermarket in the chain I am developing?), many other factors influence spatial decisions. Furthermore, very few – if any – decision makers have all of the information needed to resolve a particular locational problem optimally, and/or have the capacity, when making major decisions that require large volumes of data, to make the optimal decision. In addition, the outcomes of many decisions are affected by those made by others – the profitability of a new supermarket may well depend on how close to it competitors put their new outlets (so once sunk costs are incurred optimal decisions may not be possible). Perfect decision making is rarely feasible; people make rational decisions using the information available to them, which will in many cases reflect not only who they are but also where they are because, as Hägerstrand has shown, information is not evenly distributed. A consequence is that the equilibrium patterns at the core of models such as central place theory are unattainable: systems are continually in flux, which requires models built into a dynamic location theory. The 'traditional' location theories with their relatively unsophisticated expected geographies based on the operation of a spatial single variable only – distance, or travel costs – are thus largely obsolete: nevertheless, as a small number of researchers, such as Fingleton, continue to demonstrate, it remains possible to derive testable hypotheses from economic theory – of variations in wages, for example.

The realization of the power of these arguments stimulated a major shift in the nature of much spatial science. The models and theories that dominated its early years – such as central place theory – became much less important (indeed, some almost disappeared from the disciplinary portfolio), and were replaced by approaches that initially went under the name of behavioral geography. These are now so predominate within the spatial science component of the discipline that the term is rarely used. They haven't entirely taken over spatial science, however. Some formal modeling has been undertaken using alternative approaches to economics from the neoclassical such as welfare economics by Chisholm, Marxian by Webber and Rigby, and Ricardian/Sraffian approaches by Sheppard and Barnes. Over the last decade a school of 'new economic geographers' has emerged from within economics. The *Journal of Economic Geography* was founded to bring them together with geographers who deploy formal modeling approaches based on neoclassical economics, seeing transport costs as a significant element underpinning the creation of spatial patterns – such as trade flows and agglomeration/urbanization.

The core argument from behavioral geography's promoters is not that people are irrational decision makers but rather that they are boundedly (as against perfectly) rational: the boundedness reflects the limits to their information, their ability to process it, and the utility functions they apply when evaluating options. This does not mean that everybody has to be treated as both unique (having characteristics that nobody else shared in total) and singular (sharing no characteristics with others), so that no general patterns can be found and theories developed. Rushton makes this point in contrasting studies of behavior in space – for example, mapping individual journey to work routes – with those identifying rules of spatial behavior, the general patterns within all those choices that can be uncovered from the individual pieces of data and which suggest common behavioral decisions. The reasons for such commonalities can then be explored by, for example, identifying shared information underpinning such decisions, as in Peter Gould's pioneering work on mental maps, the spatial depictions we use when evaluating the places we want to visit, move to, etc.

Most of this work is theory driven, based on generalized hypotheses of how people behave, but much of the analysis involves exploring data assembled to identify patterns rather than explicit hypothesis testing – most of those tested are general expectations rather than firm statements that can be rigorously evaluated. This very largely reflects the 'immaturity' of such studies; our knowledge of how people make spatially relevant decisions and behave is limited and so the best we can come up with at this stage is broad expectations. In electoral studies, for example, there are strong theoretical arguments that self interest will strongly influence how people vote, and that economic concerns will underpin that self interest. But it may be expressed in different ways: some may vote sociotropically, according to their views of the national, regional, and/or local economic situation (voting for the governing party(ies) if they are optimistic about the economic situation but against it if they are pessimistic); some may vote egocentrically, according to their perceived personal financial situation; and some vote altruistically, in the interests of their (geographical) neighbors even when these don't coincide with their own. Those theoretically based expectations generate testable hypotheses but – as with the gravity model – they cannot predict in advance the size of the relevant coefficients for each type of voting, because of the many other contingent factors (some of them spatial in nature) which influence electoral behavior.

The results of such analyses provide understanding of general trends and patterns in behavior and its outcomes, but the term 'law' – widely deployed in the early years of spatial science – now rarely appears. Indeed, much contemporary spatial science adopts a less formalistic approach to knowledge acquisition. Early efforts to identify laws of spatial behavior were criticized as not only oversimplistic approaches to appreciating complex realities but also for treating individual decision makers as lacking free will – models were perceived as deterministic despite early essays separating macro and micro approaches to cause and effect. Spatial scientists still make generalizations – identifying general patterns of behavior within a given context (dataset) – but no longer imply that these are widely applicable, invariant law like statements.

In searching for these generalizations, a great deal of spatial scientific work now involves alternative data acquisition strategies. Much locational theory – such as central place and industrial location theory – concerned patterns observable in the landscape; settlement distributions, for example, and shopping centers within cities. The data were there – on maps, or in the field. The shift to a behavioral approach requires new forms of – usually bespoke – data, recording not only what people do but why, what information they have, how they evaluate it, and on what criteria, etc. Such data have to be collected from individual subjects, in most cases using questionnaires, requiring geographers to develop new skills. Not all studies require bespoke data collection by the researcher, however. A variety of data sources – such as national censuses and surveys – can be used to address many geographical questions, and there is an increasing tendency for researchers to deposit their datasets in archives so that others can reanalyze them.

A wide range of studies, in most of human geography's subfields (urban, rural, political, social, electoral, economic, etc.), illustrates the great variety of subject matter studied within these rigorous approaches. Electoral geographers examine spatial variations in turnout, for example, arguing that the more intense the local campaigns by political parties the more of their supporters who go to the polls. At aggregate scales this has been tested by relating voting patterns to the amount spent on local campaigns: the more spent, as an index of campaign intensity, the better a party's performance (especially by challengers to the incumbent party). This is complemented by individual scale studies, showing that those who report having been contacted by a party during a campaign are more likely to vote than those who do not, but that in addition the more spent on the campaign there the more likely they are to vote. And a substantial program of work on a variety of diseases has similarly thrown light on the processes involved in their diffusion through environments where geographical contingency plays crucial roles in rates of exposure and contraction.

Moving on II – Methods

In the early years of spatial science, geographers adopted methods developed in other disciplines, such as the 'family' of methods later categorized as the general linear model (regression, correlation, principal components and factor analysis, discriminant analysis, and canonical correlation analysis): these continue to serve them very well. But problems with applications to geographical (spatial) data were identified, which require alternative – more sophisticated – procedures, some developed by geographers themselves, often working in collaboration with statisticians and others.

The first of these problems – spatial autocorrelation – reflects Tobler's first law of geography. The general linear model assumes that the residuals from a regression are independent, that the value for one sample point is not affected by the values for neighboring individuals: if it is, the estimated coefficients are imprecise. But Tobler's law says that they must be so affected, that spatial autocorrelation is the core of geography. Instead of deploying the general linear model, researchers – notably a group based in Bristol around Andrew Cliff – created procedures that provide accurate estimates of relationships in the presence of spatial autocorrelation. Others argue that the presence of spatial autocorrelation indicates (as yet unidentified) spatial patterning needing investigation: the model being tested is misspecified in at least one of a number of ways. Important spatial variables may have been omitted, for example: the message is that more research is needed.

The issue of spatial autocorrelation continues to inform much spatial science. As an example, O'Loughlin, Flint, and Anselin analyzed the geography of voting for the National Socialist German Workers' Party (NSDAP) (the Nazis) at the Weimar Republic's 1930 general election and found, using regression, that the main predictor of NSDAP success was the percentage of the Kreis (electoral district) population who were Protestants: the more Protestants in the area the more votes for the NSDAP. Residuals from that regression (which included other, less important independent variables) are spatially clustered – that is, autocorrelated. A range of measures collected under the generic term local statistics, developed as part of a larger program of work in spatial statistics/econometrics led by Luc Anselin, separated this geography from the generic pattern associated with Protestantism. Local statistics identify 'hot spots', clusters of values out of line with the general pattern – in this example clusters of areas with either unexpectedly high or low levels of Nazi support characterized by O'Loughlin as 'islands that are distinctive from surrounding regions'. These are where support for the NSDAP deviated significantly, either positively or negatively, from the general pattern which could only be accounted for by particular reasons applying to those places alone. In this way, statistical analysis is linked to a growing argument among geographical researchers that 'place matters': local statistics identify places where general relationships don't operate, suggesting the importance of spatial context (locale in the language of structuration theory) as an influence on behavior.

An alternative approach to spatial autocorrelation issues developed by geographers is geographically weighted regression (GWR), which explores spatial variations in a defined relationship across a set of areas, autocorrelation having been detected. There may be a positive relationship between electoral turnout and voters' average age in most parts of an area (suggesting that older people are more likely to vote), but a negative link in some (younger people are more likely to vote). A single regression equation fitted to all districts in the area would not identify this difference: GWR, which focuses on the relationship in various parts of the area by giving greater weight to nearby observations – the first law of geography again – can do so, and pose further questions for the analyst.

GWR is linked to a wider range of inductive, exploratory spatial analysis procedures developed by Openshaw. These were initiated with his Geographical Analysis Machine (GAM) which identified significant clusters of points – such as childhood leukemia cases – unlikely to have occurred randomly and which therefore raised questions regarding their cause: did local sources of radiation generate such outbreaks? This was extended into a Geographical Explanations Machine (GEM), which explored spatial variations in relationships, such as regression equations. As in many other areas of science and social science, such inductive rather than deductive approaches (data led rather than theory led) are becoming more important. We have a great deal of information about the world: exploring it is a very fertile source of ideas and knowledge, as a first stage to developing fuller understanding of spatial variations.

Also deployed for exploring variations across places in relationships, multilevel modeling was developed by educational statisticians to evaluate differences between different settings (e.g., schools, classrooms) in student achievement, holding constant other influential variables such as family social class background. The method explores whether there are differences among places (levels in the terminology) in relationships, suggesting that processes vary across space; it can incorporate a number of levels, such as districts within counties, within states. For example, Datta *et al.* show that the prevalence of smoking among black women in the US is greater, the lower an individual's educational attainment, and also greater among those 'living as married' than those who were married; it also varies by age group (younger and older people were less likely to smoke than the middle aged), and professionals are less likely to smoke than those in nonprofessional occupations. They also find higher smoking rates among those who live in neighborhoods with higher poverty rates, as well as in states with higher poverty rates. Individual characteristics are important determinants of behavior, but similar people living in different places apparently act differently when presented with the same stimuli – reflecting the context within which they make those decisions. Multilevel modeling identifies and evaluates those geographical contingencies – the importance that place makes – dealing with places as discrete, predefined entities. (GWR treats space as continuous, although it may have to conduct its analyses using predefined entities.)

A major issue for much spatial science is the modifiable areal unit problem. Much spatial analysis – across the social sciences – uses data for areal units, such as census data for counties. Analysts may explore whether there is a relationship between average house prices and the percentage of the workforce employed in managerial and professional occupations. There may be 50 counties in the study area, but this is only one way of amalgam ating a large number of smaller areas – census tracts, for example – into such areas. Will the relationship between the two variables be the same whichever amalgamation is used? Experiments by Openshaw and Taylor showed not, but rather that different configurations can produce dif ferent – sometimes markedly different – results. So 'which is the right one?'; even 'is there a right one?' – and if not, what are the implications for spatial analysis?

Openshaw and Taylor showed that the modifiable areal unit problem results from the interaction of two separate problems operating when creating larger spatial units from smaller, spatially identified, building blocks. The first is the aggregation problem: the same set of *x* building blocks can be aggregated into *y* larger units in a (possibly very large) number of ways, while meeting preset criteria – each *y* unit should comprise the same number of *x* building blocks, for example, and all *x* building blocks in a *y* unit should be contiguous. The second is the scale problem: different sizes for *y* (i.e., numbers of *x* per unit *y*) can be produced. In many cases, there is a clear relationship between the scale of the analysis and the size of the correlation coefficient. But the interaction of scale and aggregation effects can con siderably blur such a general tendency.

The modifiable areal unit problem is significant not only because of the analytical difficulties it raises – what reliance can one place on one set of results given that the same data aggregated in a different formation could generate different results? – but also its practical rele vance. The definition of electoral districts exemplifies this. There are many ways in which UK parliamentary constituencies and US Congressional districts can be produced as aggregations of smaller areas meeting size, shape, compactness, and contiguity criteria, and they may generate very different election results. Even where relatively neutral nonpartisan procedures are deployed it is possible to generate equivalents of both major potential abuses in electoral cartography – malapportionment (the scale problem) and gerrymandering (the aggregation problem). This is illustrated by the biased outcomes in UK general election results: the Labour party has re cently benefited very substantially in the translation of votes into seats because of the geography of its support relative to that of constituency boundaries. But Open shaw realized that the problem is also an opportunity, because with experimentation it should be feasible to identify the optimum zoning of an area to meet pre determined criteria. Such zone design principles were deployed to create homogeneous small areas as reporting units for 2001 UK census data.

Closely linked to the modifiable areal unit problem is that of ecological inference, which involves drawing conclusions about individuals from aggregate data. In the classic presentation, Robinson found a correlation of 0.946 between the percentage of the population black and the percentage who were illiterate, using 1930 aggregated data for American census divisions. Using data for the 48 states, the correlation was somewhat weaker, at 0.773, and using individual data it was much weaker, at 0.203. Relying on aggregated data at either scale, analysts might conclude that blacks were much more likely to be illit erate than whites, but the individual level data show a weaker relationship (16% of blacks were illiterate, ac cording to the 1930 census, as against 3% of whites, a fivefold difference: the aggregate data analyses suggested a 10–15 fold difference). Robinson could point up these differences because individual level data were available, but if they are not, ecological inference appears to offer the only – although potentially flawed – way of drawing conclusions about individual behavior from aggregate data. Such inferences may be very weak. The vote for the British National Party – BNP – is generally higher in areas with large ethnic minority populations than where there are few representatives of those groups. But this should not be taken as indicating that members of ethnic minority groups vote for the BNP. Very few do, and higher levels of voting for the party in areas where such individuals are relatively common reflect voting by white residents there expressing support for an anti immi gration party. This is, in part at least, a scale problem reflecting the size of the area used to analyze the voting patterns: few of the whites who vote BNP actually live in the same streets as nonwhites, so at a fine enough scale you would get a negative relationship between per cent of the population who are nonwhite and voting for the BNP.

It is possible, however, to use aggregate data to draw fairly robust inferences about individual relationships and behavior. One of the most widely used methods was developed by a political scientist – Gary King – but geographers have developed alternative procedures. Entropy maximizing methods, for example (first used by Alan Wilson in 1970 to model traffic flows and other interaction patterns more successfully than the gravity model), have been deployed to disaggregate voting pat terns, by, for example, Johnston and Pattie, and the as sociated method of microsimulation (developed in economics) produces similar types of estimate (as in Ballas *et al.*'s study of the impact of various social pol icies). Johnston, Pattie, and Allsopp's 1988 study, for ex ample, showed that between the 1983 and 1987 general elections in the UK, whereas the national percentage remaining loyal to the Conservative party was 70, across the country's 627 constituencies it varied from 37 to 84. They use these estimates to explore reasons for that spatial variation, finding that Conservative loyalty was lower than average in areas where support for the party was already relatively weak in 1983, but especially in those with high levels of unemployment, where voters were much more likely either to switch their support to another party or to abstain in 1987.

Spatial science analyses spatial patterns and processes using rigorous quantitative procedures on datasets (many large and complexly structured). Its development has been much assisted – in many cases only made possible – by improvements in the speed and power of computer hardware and associated software. In particular, much re cent development has been closely linked to that of geo graphic information systems (GISs) and GIScience. These combined hardware and software resources for the col lation, visualization, integration, and analysis of spatially referenced data not only facilitate the creation and an alysis of large datasets in spatial science but also provide platforms without which many analyses are virtually im possible. Spatial science is about the analysis of maps: GIS is about creating those maps and facilitating their analysis. The field of geodemographics, for example, builds on work classifying areas according to their population character istics using census data by adding other datasets and identifying common patterns of consumer behavior, on which marketing strategies are based.

Many of these approaches to spatial statistics can now be applied using bespoke software packages, some of them developed by geographers – such as Anselins' spatial econometrics package, GeoDa. Furthermore, the approach is increasingly being taken up for the analysis of spatial data in the other social sciences: sociology, for example, has "taken a significant spatial turn" (Downey, 2006, 567) and realized the value of GIS, as have economists and political scientists.

Contemporary Spatial Science

Spatial science is a substantial component of con temporary human geography. It is very closely associated with quantitative methods: without them and associated technologies for their application much (most?) spatial science would not be feasible. The methods are a means to an end, however, for answering the basic question posed by Fotheringham (2006, 237): "Does what I'm doing provide useful evidence towards the better understanding of spatial processes?"

Most answers to this question from spatial scientists use aggregate datasets to research broad patterns. Spatial science is about patterns and processes involving large numbers of decision makers, seeking order in even the most complex situations. Even when data are available on individual decision makers, the goal is to distil general patterns – as in Rushton's distinction between behavior in space and rules of spatial behavior. Unlike other ap proaches to human geography, therefore, spatial science is macro /meso in scale, rather than micro – though it may deploy micro scale data in its search for macro/ meso scale patterns. Its findings might pose questions that can only be addressed by studying individuals, but its core methodologies focus on wholes rather than parts.

In some situations this approach is the only one possible given the subject matter. Patterns of (changing) regional development and underdevelopment, for ex ample, can only be studied as statistical aggregates – as can many indices linked to this concept, such as inflation, unemployment rates, land values, house prices, prod uctivity, etc. Aspects of our worlds are only accessible in such formats, and these are sensibly analyzed using spatial science approaches. In other cases, whereas indi vidual level data are available (or can be obtained through bespoke surveys) – on illness, turnout at elec tions, etc. – much can be gained from aggregating them and studying general patterns, of morbidity and mortality rates, for example. Much work in, for example, popu lation, medical, social, and electoral geography is of this type – though such subdisciplines are not constrained to spatial scientific approaches.

A further justification for macro /meso approaches lies in the probabilistic nature of much of our understanding and representation of the world. For example, it is now generally accepted that smoking causes lung cancer – but not all people who smoke get lung cancer and not all people who get lung cancer have smoked at some time in their lives. It is a probabilistic relationship – albeit one with a high probability. Part of the reason why we cannot say that smoking causes lung cancer without qualification other variables can either accelerate or decelerate, even block, the processes involved; and part is because is be cause our knowledge is incomplete – the scientific work continues to explore all the processes involved, and is to some extent constrained because of the difficulties of conducting controlled experiments. Similar arguments apply to a whole range of behaviors studied by (physical as well as human) geographers: the processes involved are so complex and difficult to unravel – because we cannot conduct experiments (although a range of techniques is being developed in other disciplines that may be of value in such situations: see for instance the work of Sherman and colleagues). Hence work in spatial science is almost of necessity conditional; it represents the state of our current understanding and can only be phrased in probabilistic rather than deterministic terms. Take the example of tossing a coin. We invariably treat this as a random process: assuming there is no bias in either the coin or how it is tossed we assume – and can demonstrate experimentally – that the probability of heads being the outcome is 0.5. But if we could control every possible influence – where the coin is placed on the thumb; the distance between thumb and floor; how strongly it is flipped; the nature of the surface it lands on; etc. – then we could predict the out come much more precisely.

The meso /macro patterns in many aspects of contemporary society described and analyzed by social scientists are very relevant to people's daily lives and experiences – the 'meaningful nature of life' explored in other geographical practices. Levels of ethnic segregation in neighborhoods and schools in many cities provide the context within which not only lives are (partially) lived but also people's life chances and relationships within civil society are influenced. Geographical concentrations of poverty similarly structure many people's life chances while biased election results – reflecting the operation of the aggregation and scale issues underpinning the modifiable areal unit problem – strongly influence government formation, and policies directed at those concentrations, in many countries.

Alongside their intrinsic interest and importance to understanding spatial patterns and behavior, therefore, meso /macro scale work is relevant to public policies – either studies of their impact on geographies or analyses of geographic patterns that call for public intervention. Geographies of mortality, for example, may identify areas where further investment of medical resources is warranted. Much public policy has impacts – direct and indirect – on topics of interest to human geographers and, although directed at individuals, is delivered to areas, as with the location of healthcare clinics: if the world operates through spatial aggregates, then it should be analyzed accordingly (though not exclusively so).

Similarly, whatever the immense variety of human spatial behavior, public policy which intervenes in it in some way (and much private sector policy too) is almost invariably phrased in aggregate terms. New highways are planned to link places where demand either currently outstrips supply or modeling suggests that it soon will. New commercial establishments are located where potential (unfulfilled) demand is deemed greatest (and usually implies some distance decay pattern in usage, as represented by the gravity model): much of the work by GMap (see above) is based on this aspect of Tobler's law.

Many critiques of spatial science from the 1960s onward equated its approach with the philosophy of positivism, and its goal – as in the natural sciences – of establishing general laws. (Harvey's, *Explanation in Geography* set out the basic tenets of this approach – without ever referring to it as positivism.) This aspect of the positivist approach has long been abandoned by almost all spatial scientists: they adopt some of the precepts of the approach – particularly that the things they study can be observed and measured; that the statements which they derive can be tested for their veracity; and that their studies can be replicated. For them knowledge production involves careful observation, measurement, analysis, and interpretation, generating statements which identify synoptic patterns – the broad pictures which might then be decomposed to see how they are produced and what they mean for the people who live in them. They do not – as Fotheringham (2006, 239) puts it – ignore "all the emotions and thought processes that are behind what is sometimes … highly idiosyncratic behaviour"; they accept those as valid topics for study, calling for different approaches. For spatial scientists a whole range of subjects can be addressed at the aggregate level, from which valid generalizations can be drawn to illuminate aspects of the human condition.

Spatial science is not – and should not be considered – an approach to human geography having few points of contact with other disciplinary practices. At the core of much contemporary geographical scholarship is the concept of contingency, with particular reference to place as the locale within which that contingency is played out. If, according to this argument, places matter, then with certain types of behavior this should be demonstrable through approaches associated with spatial science. This was the goal, for example, of Openshaw's original work with his GAM: exploring whether places had significantly different cancer rates, as a preface to identifying why that was the case. (The answer, higher rates were found close to nuclear power plants, which suggested that the cause was radiation either from exposure to the reactors or from the waste.) GWR similarly explores large and complex datasets to identify places with particular characteristics, and work using bespoke neighborhood data enabled Johnston and Pattie to produce strong circumstantial evidence that people are influenced by talking to their neighbors when making electoral decisions. Spatial science's core purpose is to generate rigorous descriptions of complex realities, through clear pictures which suggest general patterns of behavior within those complexities. It is theory driven, with foundations based in beliefs about how people behave (individually and in aggregates), and its findings contribute to theory development – but contemporary spatial scientists are neither as ambitious nor as naïve as the pioneers; complex realities continue to take much effort to understand.

See also: Entropy-Maximising Models; Haggett, P.; Modifiable Areal Unit Problem; Positivism/Positivist Geography; Quantitative Methodologies; Region; Regression, Linear and Nonlinear; Space II; Spatial Autocorrelation.

Further Reading

Abler, R. F., Adams, J. S. and Gould, P. R. (1971). *Spatial Organization: The Geographer's View of the World*. Englewood Cliffs, NJ: Prentice Hall.

Ackerman, E. A. (1958). *Geography as a Fundamental Research Discipline*. Chicago, IL: University of Chicago, Department of Geography, Research Paper 53.

Anselin, L., Florax, R. J. G. M. and Rey, S. J. (eds.) (2004). *Advances in Spatial Econometrics: Methodology, Tools and Applications*. Berlin: Springer.

Anselin, L., Syabri, I. and Kho, Y. (2006). GeoDa: An introduction to spatial data analysis. *Geographical Analysis* 38, 5 22.

Ballas, D. M., Rossiter, D. J., Thomas, B., Clarke, G. and Dorling, D. (2005). *Geography Matters: Simulating the Local Impacts of National Social Policies*. York: Joseph Rowntree Foundation.

Barnes, T. J. and Farish, M. (2006). Between regions: Science, militarism and American geography from world war to cold war. *Annals of the Association of American Geographers* 96, 807 826.

Batty, M. (2006). *Cities and Complexity: Understanding Cities with Cellular Automata, Agent Based Models, and Fractals*. Cambridge, MA: The MIT Press.

Bennett, R. J., Haining, R. P. and Wilson, A. G. (1985). Spatial structure, spatial interaction, and their integration: A review of alternative models. *Environment and Planning A* 17, 625 645.

Bunge, W. (1966). *Theoretical Geography*. Lund: C. W. K. Gleerup.

Burton, I. (1963). The quantitative revolution and theoretical geography. *The Canadian Geographer* 7, 151 162.

Chisholm, M. (1966). *Geography and Economics*. London: G. Bell.

Chorley, R. J. and Haggett, P. (eds.) (1965). *Frontiers in Geographical Teaching*. London: Methuen.

Chorley, R. J. and Haggett, P. (eds.) (1965). *Models in Geography*. London: Methuen.

Cliff, A. D. and Ord, J. K. (1973). *Spatial Autocorrelation*. London: Pion.

Datta, G. D., Subramanian, S. V., Colditz, G. A. *et al*. (2006). Individual, neighbourhood and state level prediction of smoking among US Black women: A multilevel analysis. *Social Science and Medicine* 63, 1034 1044.

Fingleton, B. (2005). Beyond neoclassical orthodoxy: A view based on the new economic geography and UK regional wage data. *Papers in Regional Science* 84, 351 375.

Fotheringham, A. S. (2006). Quantification, evidence and positivism. In Aitken, S. & Valentine, G. (eds.) *Approaches to Human Geography*, pp 237 250. London: Sage.

Fotheringham, A. S., Brunsdon, C. and Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: John Wiley.

Garrison, W. L. (1959). Spatial structure of the economy II. *Annals of the Association of American Geographers* 49, 471 482.

Gould, P. R. and White, R. R. (1974). *Mental Maps*. London: Penguin Books.

Gregory, S. (1963). *Statistical Methods and the Geographer*. London: Longman.

Hagerstrand, T. (1968). *Innovation Diffusion as a Spatial Process*. Pred, A. (trans.). Chicago, IL: University of Chicago Press.

Haggett, P. (1965). *Locational Analysis in Human Geography*. London: Edward Arnold.

Harris, C. D. and Ullman, E. L. (1945). The nature of cities. *Annals of the American Academy of Political and Social Science* 242, 7 17.

Harvey, D. (1969). *Explanation in Geography*. London: Edward Arnold.

Isard, W. (2003). *History of Regional Science and the Regional Science Association International: The Beginnings and Early History*. Berlin: Springer.

Johnston, R. J. (1978). *Multivariate Statistical Analysis in Geography: A Primer on the General Linear Model*. London: Longman.

Johnston, R. J. and Pattie, C. J. (2006). *Putting Voters in Their Place: Geography and Elections in Great Britain*. Oxford: Oxford University Press.

Johnston, R. J., Pattie, C. J. and Allsopp, J. G. (1988). *A Nation Dividing? The Electoral Map of Great Britain 1979 1987*. London: Longman.

Johnston, R. J., Rossiter, D. J. and Pattie, C. J. (2006). Disproportionality and bias in the results of the 2005 general election in Great Britain: Evaluating the electoral system's impact. *Journal of Elections, Public Opinion and Parties* 2, 37 54.

King, G. (1997). *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton, NJ: Princeton University Press.

Martin, R. L. (1999). The 'new geographical turn' in economics: Critical reflections. *Cambridge Journal of Economics* 23, 65 92.

Morrill, R. L. (1970). *The Spatial Organization of Society*. Belmont, CA: Wadworth.

Morrill, R. L. (2005). Hagerstrand and the quantitative revolution: A personal appreciation. *Progress in Human Geography* 29, 333 336.

O'Loughlin, J. (2002). The electoral geography of Weimar Germany: Exploratory spatial data analyses (ESDA) of Protestant support for the Nazi party. *Political Analysis* 10, 217 243.

O'Loughlin, J., Flint, C. and Anselin, L. (1994). The geography of the Nazi vote: Context, confession and class in the Reichstag election of 1930. *Annals of the Association of American Geographers* 84, 351 380.

Olsson, G. (1965). *Distance and Human Interaction: A Review and Bibliography*. Philadelphia, PA: Regional Science Research Institute, Bibliography Series 2.

Openshaw, S. and Taylor, P. J. (1981). The modifiable areal unit problem. In Wrigley, N. & Bennett, R. J. (eds.) *Quantitative Geography: A British View*, pp 60 70. London: Routledge and Kegan Paul.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review* 15, 351 357.

Rushton, G. (1969). Analysis of spatial behavior by revealed space preferences. *Annals of the Association of American Geographers* 59, 391 400.

Schaefer, F. K. (1953). Exceptionalism in geography: A methodological examination. *Annals of the Association of American Geographers* 43, 226 249.

Sheppard, E. S. and Barnes, T. J. (1990). *The Capitalist Space Economy: Geographical Analysis After Ricardo, Marx and Sraffa*. London: Unwin Hyman.

Sherman, L. W. (2003). Misleading evidence and evidence led policy: Making social science more experimental. *Annals of the American Academy of Political and Social Science* 589, 6 19.

Sherman, L. W. and Strang, H. (2004). Experimental ethnography: The marriage of qualitative and quantitative research. *Annals of the American Academy of Political and Social Science* 595, 204 222.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46, 234 240.

Ullman, E. L. (1941). A theory of location for cities. *American Journal of Sociology* 46, 853 864.

Ullman, E. L. (1956). The role of transportation and the bases for interaction. In Thomas, W. L. (ed.) *Man's Role in Changing the Face of the Earth*, pp 862 880. Chicago, IL: University of Chicago Press.

Webber, M. J. and Rigby, D. L. (1996). *The Golden Age Illusion: Rethinking Postwar Capitalism*. New York: Guilford Press.

Wilson, A. G. (1970). *Entropy in Urban and Regional Modelling*. London: Pion.

Wilson, A. G. (2000). *Complex Spatial Systems: The Modelling Foundations of Urban and Regional Analysis*. Chichester: John Wiley.

Wilson, A. G., Rees, P. H. and Leigh, C. M. (1977). *Models of Cities and Regions*. Chichester: John Wiley.

Yeates, M. H. (2001). Yesterday as tomorrow's song: The contribution of the 1960s 'Chicago School' to urban geography. *Urban Geography* 22, 514 529.

Relevant Websites

http://joeg.oxfordjournals.org
Journal of Economic Geography.

http://www.regionalscience.org
The International Regional Science Association.

Spatially Autoregressive Models

D. A. Griffith, University of Texas – Dallas, Richardson, TX, USA

Glossary

Auto-Model A phrase coined by Besag that describes a probability model where attribute values for a single variable distributed across *n* locations covary with weighted sums of nearby, neighboring attribute values for the same variable, and whose conditional probabilities can be written in terms of these weighted sums.

Autoregression A regression model specification that includes memory or feedback, in terms of weighted sums of attribute values, such that the system of observations that it describes is interdependent and contains important internal structure, and the response variable appears on both sides of the equation.

Generalized Linear Model A model specification in which some nonlinear transformation of a response variable is expressed as the weighted sum of a set of covariates whose weights are coefficients computed from observed data, without including an error term (the random component is a function of the mean response and is generated by the underlying probability distribution), and whose variation may be assumed to be non-normally distributed (e.g., Poisson, binomial, negative binomial) according to the mathematical link between the transformation and the non-normality.

Georeferenced An ordering and then tagging to a two-dimensional surface (e.g., the Earth's surface) of observations with Cartesian coordinates, such as longitude and latitude or Universal Transverse Mercator (UTM) coordinates, as symbolic descriptors of place.

Linear Model A model specification in which a response variable is expressed as the weighted sum of a set of covariates whose weights are coefficients computed from observed data, to which a random error term is added whose variation is assumed to be normally distributed, where scatterplots for response variables and each covariate approximate a straight-line graph.

Maximum Likelihood A popular inferential statistical technique that yields parameter estimates for a selected probability distribution by maximizing the joint probability for *n* observed values in a given dataset – the product of *n* individual probabilities for independent observations – resulting in estimates that make the observed data most likely.

MCMC (Markov chain Monte Carlo) An iterative mechanism for drawing dependent samples in situations where regular sampling is difficult, if not completely impossible, in order to compute simultaneous likelihood results, that generates Markov chains with Monte Carlo simulation, exploits sufficient statistics, makes use of marginal probabilities, and frequently utilizes pseudo-likelihood results.

Normalizing Constant A constant by which all possible probabilities of a probability model must be multiplied to ensure that they sum to 1, for discrete values, and integrate to 1, for continuous values.

Spatial Autocorrelation The self-correlation of values of a given attribute arising from their nearby locations in geographic space, and the essence of the first law of geography.

Spatial Filter A model specification that decomposes a georeferenced response variable into the sum of two synthetic variates, one containing spatial autocorrelation and one free of spatial autocorrelation. These former synthetic variates can be a linear combination of the eigenvectors of a geographic weights matrix.

Background: The Literal Meaning of Autoregression

The prefix 'auto' in the term 'autoregressive' (AR) means 'self' and the root 'regressive' means regression. In other words, a variable containing n observed attribute values is being regressed on itself, which results in the response variable, Y, whose values often are denoted by the n by 1 vector $\mathbf{Y}$, where n is the number of observations, appearing on both sides of the equal sign. The selection of values of Y to be on the right hand side of an equation is by observation, which is locational in the spatial context, and is dictated by some rule in such a way that each value of Y is not being regressed on itself, but rather on some subset of the remaining $n-1$ values of Y. For geographic models, the adjective 'spatial' dictates that these subsets comprise neighboring values of each location. Because of their closeness in geographic space, neighboring values tend to contain duplicated information. This situation is similar to taking a sample and falsely doubling its size merely by duplicating each of its entries. Spatially redundant information needs to be accounted for by regressing Y on neighboring values of Y. Otherwise, results can be misleading, as would be the case for the artificially doubled sample. Differentiation of neighboring and nonneighboring values frequently is achieved with n binary $0-1$ indicator variables (0 if the locations of Y values are not neighboring, and 1 if the locations are), each containing n entries. These n indicator variables then are collected into an n by n table, commonly called a geo graphic weights matrix, and usually denoted by matrix $\mathbf{C}$. Consequently, in matrix form, vector $\mathbf{Y}$ is on the left hand side and vector $\mathbf{CY}$ is on the right hand side of an AR equation. Because vector $\mathbf{CY}$ is the sum of sur rounding neighboring observed values of variable Y, it frequently is replaced with its row standard version (i.e., each cell entry in matrix $\mathbf{C}$ is divided by its corres ponding row sum – the number of 1s in its row), denoted by $\mathbf{WY}$, which is the average of the surrounding values of Y. Individual ith observation row entries for these two versions are denoted, respectively, by $\mathbf{C}_i\mathbf{Y}$ and $\mathbf{W}_i\mathbf{Y}$.

Introduction: The Meaning of Spatial Correlation

Correlation refers to the relationship between two vari ables, X and Y, can be quantified with a Pearson product–moment coefficient, and often is portrayed with a scatter diagram (a two dimensional plotting of individual ob servation value pairs, (X_i, Y_i), where i denotes an obser vation's name). Positive linear correlation indexes a tendency for high values of X to be paired with high values of Y, medium values of X with medium values of Y, and low values of X with low values of Y. As this tendency becomes more pronounced, these pairings in creasingly align along a line – a straight line when a relationship is linear. Negative linear correlation indexes a tendency for high values of X to be paired with low values of Y, medium values of X with medium values of Y, and low values of X with high values of Y. Finally, no correlation (i.e., randomness) indexes a tendency for high values of X to be paired with high, medium, and low values of Y, medium values of X with high, medium, and low values of Y, and low values of X with high, medium, and low values of Y. Conventional linear regression models capture this correlation.

Spatial correlation – also known as spatial auto correlation – replaces each value of X with $\mathbf{C}_i\mathbf{Y}$ or $\mathbf{W}_i\mathbf{Y}$ in this preceding relationship scenario, can be quantified with a Moran coefficient or a Geary ratio, and often is portrayed with a Moran scatterplot or a semivariogram plot, which are customized scatterdiagrams. Accordingly, positive spatial correlation indexes a tendency for high values of, say, $\mathbf{W}_i\mathbf{Y}$ to be paired with high values of Y, medium values of $\mathbf{W}_i\mathbf{Y}$ with medium values of Y, and low values of $\mathbf{W}_i\mathbf{Y}$ with low values of Y. Negative spatial correlation indexes a tendency for high values of $\mathbf{W}_i\mathbf{Y}$ to be paired with low values of Y, medium values of $\mathbf{W}_i\mathbf{Y}$ with medium values of Y, and low values of $\mathbf{W}_i\mathbf{Y}$ with high values of Y. Finally, no spatial correlation indexes a tendency for high values of $\mathbf{W}_i\mathbf{Y}$ to be paired with high, medium, and low values of Y; medium values of $\mathbf{W}_i\mathbf{Y}$ with high, medium, and low values of Y; and low values of $\mathbf{W}_i\mathbf{Y}$ with high, medium, and low values of Y. Most spatial correlation is moderate and positive; one ex ception is spatial correlation for remotely sensed images, which almost always is very strong and positive. Spatial AR models capture this spatial correlation.

The Auto-Normal Probability Model

Some interval/ratio georeferenced datasets have fre quency distributions that mimic a bell shaped curve (e.g., geo coding errors). One principal, and perhaps the most conspicuous, impact of spatial autocorrelation in terms of this type (as well as other types) of variable is on vari ation. The overwhelming majority case of positive spatial autocorrelation results in variance inflation, a well known feature of georeferenced data, which for a normal frequency distribution means the central frequencies (i.e., numbers of values near the mean) decrease and the tail frequencies increase – peakedness (i.e., kurtosis) is altered because the curve is flattened. Less well known is that the intermediate frequencies can become dis proportionately reduced in order for increased concen tration very close to the mean (see **Figure 1**). Spatial auto normal AR models seek to furnish a description of these data.

Three spatial linear regression model specifications have become favored in the literature to describe geor eferenced data, namely the

conditional AR (CAR) model:

$$\mathbf{Y} = (\mathbf{I} \quad \mathbf{D})\mathbf{Y} + \mathbf{DX}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \mathbf{D}^{\mathrm{T}}\mathbf{D} = (\mathbf{I} \quad \rho\mathbf{C})\mathbf{M} \qquad [1]$$

AR response model:

$$\mathbf{Y} = \rho\mathbf{WY} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad [2]$$

and simultaneous AR (SAR) model:

$$\mathbf{Y} = \rho\mathbf{WY} + (\mathbf{I} \quad \rho\mathbf{W})\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad [3]$$

where $\mathbf{X}$ is an n by $(p+1)$ matrix of covariates, $\boldsymbol{\beta}$ is a $(p+1)$ by 1 vector of regression coefficients, ρ is the spatial AR parameter, $\mathbf{M}$ is an n by n diagonal matrix, often set equal to the n by n identity matrix $\mathbf{I}$, and $\boldsymbol{\varepsilon}$ is an n by 1 vector of independent and identically normal distribution random errors. The first of these specifi cations describes spatial autocorrelation in terms of a first order inverse covariance structure (i.e., $(\mathbf{I} - \rho\mathbf{C})$, which involves the first order term $\mathbf{C}$); the other two describe it in terms of a second order structure (i.e., $(\mathbf{I} - \rho\mathbf{W})'(\mathbf{I} - \rho\mathbf{W})$, where prime denotes the matrix transpose operation, which involves the second order term $\mathbf{W}'\mathbf{W}$). A first order model represents weaker spa tial autocorrelation over neighboring locations within a relatively short distance band; a second order model

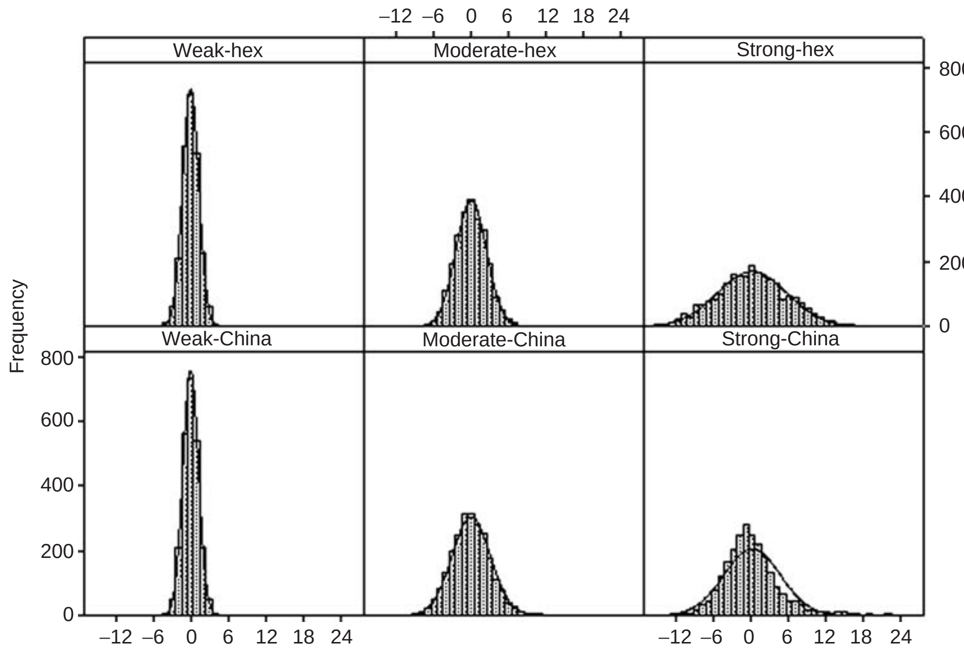

Figure 1 Histograms for simulated normal data containing varying levels of positive spatial autocorrelation: $\mu = 5$ and $\sigma = 1$. Top: a 45-by-54 regular hexagonal tessellation ($n = 2430$) and bottom: China's 1990 county irregular surface partitioning ($n = 2379$).

represents stronger spatial autocorrelation over neighboring locations in a greater distance band. The AR model, called the lag model in spatial econometrics, is a special case of the SAR model in which the spatial lag component coefficients for the covariates are restricted to zero rather than set to ρ. The AR specification conceptualizes spatial autocorrelation as a direct effect outcome; for example, high population density in a given location tends to promote high population density in nearby locations. In contrast, the CAR and SAR specifications conceptualize spatial autocorrelation as an indirect, error effect; for example, a rain storm that moved across a field of crops may well create spatial patterning that is not captured by a crop yield model specification. Spatial scientist practitioner preferences tend to reflect employment of the CAR model in remote sensing, the AR model in spatial econometrics, and the SAR model in spatial statistics. One deviation from this usage pattern is employment of the CAR model in hierarchical Bayesian modeling by spatial statisticians, and hence epidemiologists who engage in disease mapping.

Parameter estimation most often is achieved using maximum likelihood techniques, which optimize the observed data likelihood function. If the spatial covariation is denoted by matrix $\mathbf{V}$ – that is, $\mathbf{V}^{-1} = (\mathbf{I} - \rho\mathbf{C})$ for the CAR model with $\mathbf{M} = \mathbf{I}$, and $\mathbf{V}^{-1} = (\mathbf{I} - \rho\mathbf{W})'(\mathbf{I} - \rho\mathbf{W})$ for the SAR model – then maximum likelihood estimates are given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} \qquad [4]$$

$$\hat{\sigma}^2 = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})/n \qquad [5]$$

and $\hat{\rho}$ is yielded by MIN:

$$[(\mathrm{e}^{-\mathcal{J}(\rho)})^2(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})]^{-n/2}$$
$$\text{over the range } \left(\frac{1}{\lambda_{\min}}, \frac{1}{\lambda_{\max}}\right) \qquad [6]$$

where $\mathbf{V}^{-1}$ is a function of ρ, and $\mathrm{J}(\rho)$ refers to the calculus concept 'Jacobian of a transformation', which ensures that all probabilities integrate to 1. Meanwhile, for the AR model,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1/2}\mathbf{Y} \qquad [7]$$

$$\hat{\sigma}^2 = \mathbf{Y}'(\mathbf{V}^{-1/2})'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{V}^{-1/2}\mathbf{Y}]/n \qquad [8]$$

and $\hat{\rho}$ is yielded by MIN:

$$[(\mathrm{e}^{-\mathcal{J}(\rho)})^2(\mathbf{V}^{-1/2}\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{V}^{-1/2}\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})]^{-n/2} \qquad [9]$$

where $\mathbf{V}^{-1/2}$ commonly is $(\mathbf{I} - \rho\mathbf{W})$.

Equation [4] renders unbiased estimates whose expectations are equivalent to its ordinary least squares (OLS) counterparts (i.e., $\mathbf{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$); OLS first moments often are unbiased in the presence of spatial autocorrelation. Equation [7] renders biased OLS estimates (i.e., $\mathbf{E}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1/2}\mathbf{X}\boldsymbol{\beta}$). The standard bias correction adjustments for eqns [5] and [8] are to divide by $(n - p - 1)$, for p covariates, rather than just n; but this adjustment does not render unbiased OLS estimates in the presence of spatial autocorrelation without also adjusting for the autocorrelation – this is where spatial autocorrelation introduces variance inflation, and where statistical efficiency is impacted upon. In ideal situations these estimates contain sufficient statistics, with the set of complete sufficient statistics now including a spatial covariation term.

An alternative specification to eqns [1]–[3] exploits the geographic structure latent in matrix **C**. Pre and post multiplying this geographic weights matrix by the multivariate statistics projection matrix $(\mathbf{I}-\mathbf{11}'/n)$ coincides with a centering of the data and yields the Moran coefficient numerator term $(\mathbf{I}-\mathbf{11}'/n)\mathbf{C}(\mathbf{I}-\mathbf{11}'/n)$. Now the right hand side of a regression equation can be specified using the matrix decomposition $(\mathbf{I}-\mathbf{11}'/n)\mathbf{C}(\mathbf{I}-\mathbf{11}'/n)=\mathbf{E\Lambda E}'$, where **E** and **Λ,** respectively, are the matrix of eigenvectors and the diagonal matrix of eigenvalues of this modified matrix **C**. This projection matrix essentially replaces the principal eigenfunction with one whose eigenvalue is 0 and whose eigenvector is **1**. Judiciously selecting from among the n eigenvectors yields a subset of them, say $\mathbf{E}_\mathrm{K}$, which can be used to replace AR terms like $\rho\mathbf{C}$ and $\rho\mathbf{W}$ in eqns [1]–[3]. Accordingly, the spatial filter (SF) model may be written as

$$\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}_\mathrm{X}+\mathbf{E}_\mathrm{K}\boldsymbol{\beta}_\mathrm{E}+\boldsymbol{\varepsilon} \qquad [10a]$$

where the subset of eigenvectors $\mathbf{E}_\mathrm{K}$, which are mutually orthogonal and uncorrelated, can be selected with a stepwise linear regression procedure, $\boldsymbol{\beta}_\mathrm{X}$ and $\boldsymbol{\beta}_\mathrm{E}$, respectively, are regression coefficients for the p covariates and the K selected eigenvectors, and spatial autocorrelation detected in non SF OLS residuals is captured by $\mathbf{E}_\mathrm{K}\hat{\boldsymbol{\beta}}_\mathrm{E}$, rendering the resulting residual vector, **e**, void of spatial autocorrelation. An appealing feature of this specification is that distribution theory for the Moran coefficient is known for the resulting linear regression residual vector **e** (e.g., a sampling distribution mean of

$$\frac{n \quad \mathrm{TR}\{[(\mathbf{XE}_\mathrm{K})'(\mathbf{XE}_\mathrm{K})]^{-1}(\mathbf{XE}_\mathrm{K})'\mathbf{C}(\mathbf{XE}_\mathrm{K})\}}{\mathbf{1}^\mathrm{T}\mathbf{C1} \qquad n \quad 1 \quad p \quad K} \qquad [10b]$$

where TR denotes the matrix trace operation, and an approximate standard deviation of $\sqrt{2/\mathbf{1}^\mathrm{T}\mathbf{C1}}$.

A simple illustration of how this spatial statistical modeling may alter the understanding of empirical situations can be gleaned by revisiting Haining's 1991 Glasgow health data. The bivariate correlation between respiratory and cancer death rates across the 87 health districts is 0.63. Both variables contain weak to moderate positive spatial autocorrelation, respectively, exhibiting roughly 1/4 and 1/3 redundant information. Results from SAR models indicate that this correlation actually is 0.51, a significantly lower value. Results from SF models indicate a value of 0.45, and reveal that spatial autocorrelation in the two variables has two map patterns in common, which inflates the correlation beyond 0.63, coupled with two and five map patterns that are unique to, respectively, respiratory and cancer death rates, which deflate the correlation to 0.63.

The Auto-Poisson Probability Model

Many georeferenced datasets contain counts whose upper limit is unconstrained (e.g., lighting strikes in a given region), and hence have frequency distributions mimicking that for a Poisson random variable. Changing the probability model specification from a normal to a non normal, one does not necessarily alter the principal impact of spatial autocorrelation on variation. The most common case of positive spatial autocorrelation also results in variance inflation for a Poisson frequency distribution, which means the chances of smaller and of extremely large values increase, at the expense of intermediate values (see **Figure 2**). Excessive zeroes as well as outliers may materialize. Winsorized (i.e., truncating unlikely large extreme values) spatial auto Poisson AR models seek to furnish a description of these data when spatial autocorrelation is positive.

The common spatial Poisson regression model specification – which has Y on both sides of the equation – may be written as

$$P(Y=y_i)=c^{-1}\mathrm{e}^{-(\mathbf{X}_i\boldsymbol{\beta}+\rho\mathbf{C}_i\mathbf{Y})}(\mathrm{X}_i\boldsymbol{\beta}+\rho\mathbf{C}_i\mathbf{Y})^{Y_i}(y_i!) \qquad [11]$$

where y_i is the count for location i, ! denotes the factorial operation, $\mathbf{X}_i$ is the individual ith observation row entry in the covariate matrix **X**, which includes a 1 for the constant term, and $c \quad ^1$ is the normalizing constant that functions like the Jacobian term for an auto normal model (i.e., it ensures that all probabilities sum to 1). Unfortunately, $c \quad ^1$ is unknown, complicated, and intractable.

Fortunately, Markov chain Monte Carlo (MCMC) techniques enable the parameters of eqn [11] to be estimated. MCMC is used to simulate from a target joint auto Poisson probability distribution, which is known only up to the normalizing constant $c \quad ^1$. MCMC begins with the observed conditional (marginal) distributions, and outputs a sample of parameters drawn from their joint distribution. Starting with any initial conditions (e.g., independent and identically distributed Poisson random variables) and the transition matrix defined by eqn [11], the idea is to recursively and iteratively simulate each location's value using eqn [11]. Initial parameter estimates can be obtained from a conventional Poisson regression that includes the spatial lag term, $\mathbf{C}_i\mathbf{Y}$ (i.e., pseudo likelihood estimates). An iteration involves visiting each location on a map in turn, perhaps in a random sequence, using its updated local Poisson mean to draw a

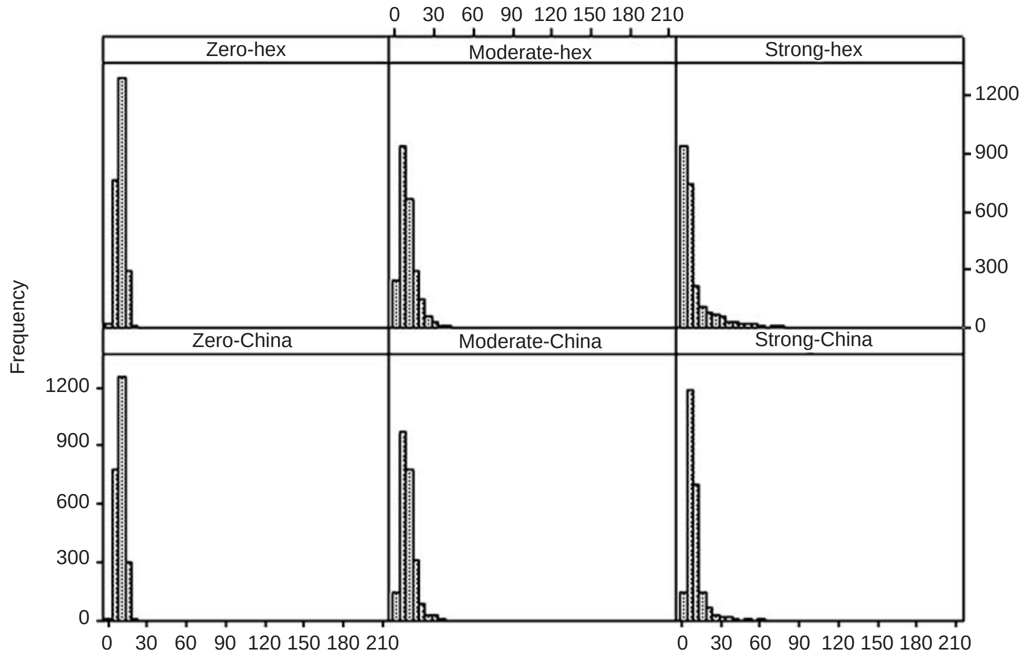

Figure 2 Histograms for simulated Poisson data containing varying levels of positive spatial autocorrelation: μ 5. Top: a 45-by-54 regular hexagonal tessellation (n 2430) and bottom: China's 1990 county irregular surface partitioning (n 2379).

random value from a Poisson distribution, and then immediately replacing the existing value with this new value. Winsorizing is performed when a simulated value exceeds some prespecified maximum value. Many tens of thousands of iterations must be performed. A burn in set of beginning iterations is discarded so that the arbitrary initial map of counts becomes irrelevant. Next, time series plots of sufficient statistics need to be inspected, as do their time series correlograms, to ensure an absence of patterns and serial correlation, and convergence of parameter estimates. Frequently conspicuous patterns can be removed by increasing the burn in period. Often serial correlation can be minimized by extracting every *m*th map (i.e., weeding). Convergence can be checked by generating and then evaluating several of these sequences (i.e., chains) with analysis of variance procedures. This MCMC scheme is known as Gibbs sampling, and is a method for generating a joint empirical distribution of multiple variables from a set of modeled conditional distributions for each variable when the structure of data is too complex to either implement mathematical formulas or directly simulate.

The resulting parameter estimates are obtained by constructing ratios of the converged upon distribution and the initial distribution. Constructing these ratios places the unknown normalizing constant into both the numerator and the denominator, allowing it to divide out of the problem, and hence remain unknown. Because the simple auto Poisson model can capture only negative spatial autocorrelation, Winsorizing tends to result in $\hat{\rho}$ values that are quite close to zero, and at best represent weak to moderate positive spatial autocorrelation. For corroboration purposes, these results can be approximated with an auto normal model specification by analyzing the square root of the observed counts (the variance stabilizing Box–Cox power transformation for a Poisson variable), or the logarithmic transformation LN(y_i $+\delta$), where δ is a translation coefficient that helps to better align these transformed values with a normal distribution.

The SF equation once again replaces $\rho\mathbf{C}$ with $\mathbf{E}_K\boldsymbol{\beta}_E$, yielding

$$P(Y = y_i) = c^{-1} e^{e-(\mathbf{X}_i\boldsymbol{\beta}X+\mathbf{E}_i K\boldsymbol{\beta}E)}\left[\mathbf{e}^{(X_i\boldsymbol{\beta}_X+\mathbf{E}_{iK}\boldsymbol{\beta}_E)}\right]^{Y_i}/(y_i!) \qquad [12]$$

where the subset of eigenvectors $\mathbf{E}_K$ can be selected with a stepwise Poisson regression procedure. Some criteria suggest that this specification may well be superior to its Winsorized auto Poisson counterpart. One appealing advantage eqn [12] exhibits over eqn [11] is that it can capture extremely high levels of positive spatial autocorrelation. Another is that it does not require MCMC techniques for parameter estimation because its normalizing constant is known and tractable; rather, it can utilize standard generalized linear modeling statistical theory and software.

The Auto-Binomial Probability Model

A number of georeferenced datasets comprise counts that have upper limits (e.g., votes by precincts for political

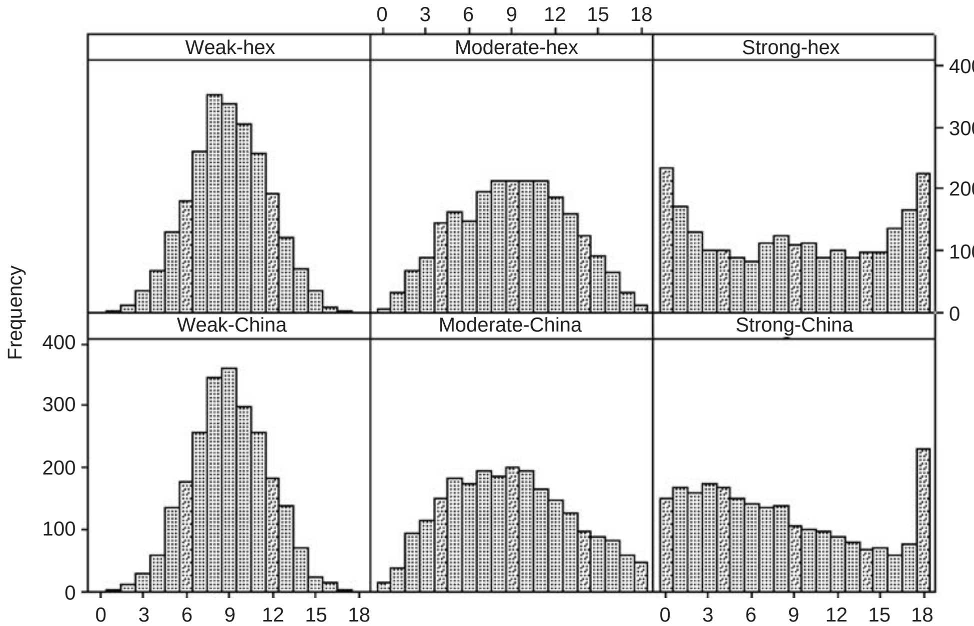

Figure 3 Histograms for simulated binomial data containing varying levels of positive spatial autocorrelation: N 18, p 0.5. Top: a 45-by-54 regular hexagonal tessellation (n 2430) and bottom: China's 1990 county irregular surface partitioning (n 2379).

candidates), and hence have a frequency distribution mirroring that of a binomial random variable. Positive spatial autocorrelation also results in variance inflation for a binomial frequency distribution, which means the chances of the extreme probabilities increase at the ex pense of more centrally located intermediate prob abilities (see **Figure 3**). This effect is not detectable in binary, Bernoulli (i.e., logistic regression type) data. Spatial auto binomial/logistic AR models seek to furnish a description of these data.

The common spatial binomial regression model specification – which has Y on both sides of the equation – may be written as

$$P(Y = y_i) = c^{-1}\binom{N_i}{y_i}\left[\frac{\mathrm{e}^{\mathbf{X}_i\boldsymbol{\beta}+\rho\mathbf{C}_i\mathbf{Y}}}{1+\mathrm{e}^{\mathbf{X}_i\boldsymbol{\beta}+\rho\mathbf{C}_i\mathbf{Y}}}\right]^{y_i} \times \left[1 \quad \frac{\mathrm{e}^{\mathbf{X}_i\boldsymbol{\beta}+\rho\mathbf{C}_i\mathbf{Y}}}{1+\mathrm{e}^{\mathbf{X}_i\boldsymbol{\beta}+\rho\mathbf{C}_i\mathbf{Y}}}\right]^{N_i-y_i} \qquad [13]$$

where N_i is the total base statistical population at location i. Again, the normalizing factor c ^{1} is unknown, complex and intractable, but as with the auto Poisson model, parameter estimates can be obtained here with MCMC techniques.

The SF equation once again replaces $\rho\mathbf{C}$ with $\mathbf{E}_\mathrm{K}\boldsymbol{\beta}_\mathrm{E}$, yielding

$$P(Y = y_i) = \binom{N_i}{y_i}\left[\frac{\mathrm{e}^{\mathbf{X}_i\boldsymbol{\beta}_\mathrm{X}+\mathbf{E}_{i\mathrm{K}}\boldsymbol{\beta}_\mathrm{E}}}{1+\mathrm{e}^{\mathbf{X}_i\boldsymbol{\beta}_\mathrm{X}+\mathbf{E}_{i\mathrm{K}}\boldsymbol{\beta}_\mathrm{E}}}\right]^{y_i} \times \left[1 \quad \frac{\mathrm{e}^{\mathbf{X}_i\boldsymbol{\beta}_\mathrm{X}+\mathbf{E}_{i\mathrm{K}}\boldsymbol{\beta}_\mathrm{E}}}{1+\mathrm{e}^{\mathbf{X}_i\boldsymbol{\beta}_\mathrm{X}+\mathbf{E}_{i\mathrm{K}}\boldsymbol{\beta}_\mathrm{E}}}\right]^{N_i-y_i} \qquad [14]$$

Now parameter estimates can be obtained with standard stepwise logistic regression or generalized linear model procedures.

Summary and Contemporary Issues

Spatial autocorrelation frequently has little impact on regression coefficient estimates for geographically dis tributed non distance based covariates, regardless of whether the statistical distribution of a variable is normal, Poisson, or binomial. In contrast, spatial autocorrelation impacts upon the variance of these variables, with posi tive spatial autocorrelation inducing variance inflation. This increased variance tends to flatten a bell shaped curve, and create overdispersion in Poisson and binomial variables. Consequently, frequency distributions are dis torted when spatial correlation is present. As positive spatial autocorrelation increases: (1) the center is squa shed and the tails are thickened in a normal curve; (2) the shape of a Poisson distribution increasingly resembles that of an exponential distribution, with the frequency of zero and/or near zero values increasing concomitantly with the appearance of marked outliers; and (3) the shape of a binomial distribution first approaches that of a uni form, and then that of a sinusoidal distribution. These metamorphoses are the only way variance can increase for these types of variables. In addition, spatial depend ency effects compromise the descriptive ability of con ventional versions of these probability models, as well as statistical inferences associated with them. Fortunately, technology now exists that enables the parameters of AR versions of these probability models to be estimated, furnishing much better descriptions of georeferenced data.

A number of unresolved issues are raised by the for going discussion. Foremost is the absence of formal tests for generalized linear model residual spatial auto correlation. Such tests also need to be developed for auto normal residual spatial autocorrelation. Second is the lack of standard software, such as GeoBUGS for Bayesian map analysis, for implementing the necessary MCMC techniques needed to estimate the parameters of auto Poisson and auto binomial models. SF model specifications help circumvent this latter problem.

Nevertheless, spatial AR models have brought new power to the analysis of georeferenced data, capturing effects of redundant locational information contained in these data. Their descriptive abilities better portray ob served data, and have become, and continue to be, in creasingly appealing to scientists working in the spatially integrated social sciences, as well as in spatial epidemi ology. They offer a more fully developed nomothetic approach that, in turn, will support an increasing em phasis on the geographically varying local statistics that fluctuate around their global parameter counterparts.

See also: Regression, Linear and Nonlinear; Spatial Autocorrelation.

Further Reading

Anselin, L. (1988). *Spatial Econometrics*. Dordrecht: Kluwer.
Arbia, G. (2006). *Spatial Econometrics*. New York: Springer Verlag.
Cliff, A. and Ord, J. (1980). *Spatial Processes*. London: Pion.
Cressie, N. (1992). *Statistics for Spatial Data*. New York: Wiley.
Griffith, D. (2000). A linear regression solution to the spatial autocorrelation problem. *Journal of Geographical Systems* 2, 141 156.
Griffith, D. (2003). *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization*. Berlin: Springer Verlag.
Griffith, D. (2004). A spatial filtering specification for the auto logistic model. *Environment and Planning A* 36, 1791 1811.
Griffith, D. (2004). Extreme eigenfunctions of adjacency matrices for planar graphs employed in spatial analyses. *Linear Algebra and Its Applications* 204, 201 219.
Griffith, D. (2006). Assessing spatial dependence in count data: Winsorized and spatial filter specification alternatives to the auto Poisson model. *Geographical Analysis* 38, 160 179.
Griffith, D. and Lagona, F. (1998). On the quality of likelihood based estimators in spatial autoregressive models when the data dependence structure is misspecified. *Journal of Statistical Planning and Inference* 69, 153 174.
Haining, R. (1990). *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge: Cambridge University Press.
Haining, R. (2003). *Spatial Data Analysis*. Cambridge: Cambridge University Press.
Huffer, F. and Wu, H. (1998). Markov chain Monte Carlo for autologistic regression models with application to the distribution of plant species. *Biometrics* 54, 509 524.
Kaiser, M. and Cressie, N. (1997). Modeling Poisson variables with positive spatial dependence. *Statistics and Probability Letters* 35, 423 432.
Upton, G. and Fingleton, B. (1985). *Spatial Data Analysis by Example: Point Pattern and Quantitative Data*, vol. I. New York: Wiley.

Relevant Websites

http://www.spatial econometrics.com
Econometrics Toolbox for Matlab by James LeSage, University of Toledo.
http://www.spatial statistics.com
Spatial statistics software for Matlab and Fortran by Kelly Pace, LSU.

Squatter Settlements

K. D. Willis, Royal Holloway, University of London, Egham, UK

Glossary

Burgess Model Developed by Ernest W. Burgess in 1925, this modeled the urban social structure of industrialized cities based on the case of Chicago. The model consisted of a series of concentric rings around the central business district, with working-class populations living near the center where industries were located, and middle-class households living further out from the center.

Gentrification Processes of urban social and economic change, which involve increased property prices and the influx of higher-income residents. May be associated with urban government infrastructure projects.

Micro-Finance Term given to schemes which aim to provide funding to very poor individuals or families who are unable to obtain loans from formal lending institutions.

Neoliberalism A political and economic ideology and set of practices which focus on reducing the role of the state and expanding the involvement of the private sector.

Introduction

As more and more of the world's population lives in urban areas, access to housing has become a growing concern. This is not because housing is necessarily difficult to find in urban areas, but because the rates of urbanization and the availability of appropriate housing do not match. The majority of people living in the cities of the Global South (where rates of urban growth are highest) are poor. Governments are unable or unwilling to provide sufficient housing, and private sector housing provision is too expensive for most households. Given this housing gap, poor urban households have developed their own solutions. These often include what have been termed 'squatter settlements' where households occupy land and build their own houses. While housing quality and access to services may be limited, such opportunities are key in providing millions of people shelter.

In this article, the main debates surrounding squatter settlements are outlined, focusing on definitional issues and their geographical distribution. The changing nature of policies toward these settlements is also covered.

Definitions

The term squatter settlement is often used as a general term to encompass low quality housing, occupied by the poor, usually on the periphery of cities in the Global South. Such use makes it synonymous with other terms such as 'shanty town' and 'informal settlement'. Despite the fact that they are often used interchangeably, these terms actually refer to different forms of shelter, with distinctions by land tenure, building type, construction method, and legal status. Formally, a squatter settlement is identified by land tenure, with residents occupying land illegally, that is, squatting. This may be through an organized land invasion, but is more often through the gradual accretion of the settlement as people move in over time. In many parts of the Global South, squatters occupy public or communal land as it is often harder to remove them from such locations, rather than privately held terrain.

In contrast, shanty towns are formally defined according to the quality of the buildings; shanties are usually flimsy constructions made out of materials such as cardboard, plastic, and corrugated iron. The buildings lack access to services such as water, electricity, and sanitation. Poor housing quality is also implied by the term 'slum', but this term also encompasses rental housing and overcrowded conditions in buildings constructed from more permanent materials.

The more general term 'informal settlement' is used to encompass residential areas which may have dubious land tenure, for example, when settlers buy plots of land from landowners who have subdivided and sold their property without legal permission. Informal settlements are also those where housing regulations have not been followed and the houses have been built by the occupants (hence the terms 'self built housing' or 'self help housing') or informally employed builders. Such settlements usually lack services early in their history due to their informal status, but this may change over time as discussed below. Due to its broad scope, the term informal settlement is often used in official documents and will be used in this article unless land tenure is the focus of the discussion. Local terms for such forms of housing are also found throughout the world (see **Table 1**).

Distribution and Growth of Informal Settlements

Informal settlements are most associated with the cities of the Global South, reflecting the mismatch between

Table 1 Alternative names for squatter settlements

Name	
Squatter settlement	Housing built on illegally occupied land.
Shanty town/shanty	Area of poor-quality housing, built from materials such as cardboard and corrugated iron. A shanty is an individual dwelling made of such materials.
Informal settlement	Area of housing which does not meet legal requirements of tenure, housing quality, and land use.
Self-help housing	Housing which has been built by the residents themselves.
Bustee/Favela/Pueblo joven/ Bidonville	Local names in India/Brazil/Peru/ Algeria.
Spontaneous settlement	Settlement which has developed without formal planning. May be due to a land invasion, or gradual growth over time.

Table 2 Slums by Global Region, 2001

	Urban population (%)	*Slum population as % of urban*
Northern Africa	52	28.2
Sub-Saharan Africa	34.6	71.9
Latin America and Caribbean	75.8	31.9
Eastern Asia	39.1	36.4
Eastern Asia (excluding China)	77.1	25.4
South-Central Asia	30	58
South-Eastern Asia	38.3	28
Western Asia	64.9	33.1
Oceania (excluding New Zealand and Australia)	26.7	24.1
World	47.7	31.6

Source: Adapted from UN Habitat (2003). *Slums of the World: The Face of Urban Poverty in the New Millennium?* Nairobi: UN Habitat, (Table 3, p. 27).

urban population growth and affordable housing pro vision. They have, however, been and continue to be found in the Global North. For example, in the United States during the Great Depression of the 1930s, thou sands of people who had lost their homes or were forced to migrate for work lived in what became known as 'Hoovervilles' after President Herbert Hoover. As in parts of the Global South today, shelters were built from available scrap material and located in areas of wasteland, railway yards, streets, and parks, including New York's Central Park. In today's Global North, such settlements persist, for example, those built by migrant agricultural workers in California, or asylum seekers in Paris. With the collapse of the Soviet Union, informal settlements have become increasingly common in parts of the in dependent republics. Regardless of where or when they are built, their presence reflects the exclusion of residents from better quality housing produced by either the public or private sectors.

While informal settlements can be found throughout cities, the common image is of large scale settlements being set up in the urban periphery. This has resulted in models of the 'Third World city', with a decreasing socioeconomic gradient from the city center to the pe riphery, so that the rich live centrally and the poor on the margins. This is contrasted with models of 'industrial cities' where low income populations reside in the city center and the rich in spacious accommodation in the suburbs, as in Burgess' concentric zone model, for ex ample. Such models fail to recognize the heterogeneous nature of urban space; higher income households are able to afford to live in more desirable areas, thus leaving more polluted or dangerous locations to the poorer populations. Poorly drained locations or steep slopes are more likely to be occupied by the poor as these locations present environmental hazards, while richer households will also seek to locate far from polluting industries. For example, the population most affected by the release of deadly poisonous gas from the Union Carbide chemical plant in Bhopal, India, in 1984 were the low income residents of informal settlements nearby.

Models of Third World and industrial cities also fail to consider the dynamism of urban structures. Gentrifi cation has occurred in the center of many cities, par ticularly in the Global North, leading to shifts in demographic profiles, while in parts of the Global South, upper class households are moving to gated com munities, often toward the city edge. Additionally, in many cities, long established informal settlements have been consolidated and upgraded, thus becoming formally recognized areas occupied by middle class households. This process is discussed in more detail below.

Globally, about one third of the population lived in slums in 2001, but the importance of informal settlements and slum housing varies greatly (see **Table 2**). Collecting accurate figures for such housing is difficult, given both definitional variations and the fluid and transitory nature of some of the settlements and their populations. The United Nations Human Settlements Programme (UN Habitat) talks generally about 'slums' to refer to poor quality housing. The importance of such housing for urban populations does not reflect the level of urban ization in a region. Sub Saharan Africa had an urban population of just over one third in 2001, but of these, nearly three quarters lived in what were classified as slums. In contrast, Southeastern Asia had a slightly higher urban population, but less than a third lived in slums. Such differences reflect, among other things, different urban growth rates, levels of economic development, land tenure systems, land availability, and political factors.

Within regions, there may also be very different patterns. For example, peripheral squatter settlements have been much more important in Mexico and Peru (most notably in Mexico City and Lima) than in Chile (after 1973) or Colombia. This reflects not only the availability of land for such settlements but also the way in which squatter settlements were viewed politically. As the section below demonstrates, government attitudes to informal settlements have varied greatly from wholesale clearance to tacit support as a form of patronage. In Mexico and Peru, in the past, politicians have turned a blind eye to informal settlements in return for electoral support.

Approaches to Informal Settlements

When examining approaches to informal settlements, it is vital to recognize the diversity of opinion from a range of actors. The main groups involved are the informal settlement residents themselves, other urban residents, and government. International agencies, such as the World Bank and the United Nations, also have changed their policies toward low income housing provision over time.

Clearance

In the 1950s and 1960s when many Latin American and Asian cities were growing rapidly, informal settlements were often considered by governments and other urban residents (particularly the middle classes) as great problems. Such perspectives were not only because these settlements were viewed as being unhealthy for the inhabitants, but also because they were seen as a threat to the city's population as a whole. The living conditions were unhygienic because of overcrowding and a lack of water and sanitation, resulting in high levels of disease. In addition, the settlements were viewed by outsiders as being havens for criminals and gangs.

Because of these perspectives, many municipal governments cleared informal settlements, sometimes by sending in bulldozers with little warning, and in other cases, developing relocation schemes involving housing that was viewed by planners as more suitable. This alternative housing was often in the form of high rise flats, located far from the original settlement, as in the case of the 'superblocks' of Caracas and the low income housing constructed as part of the Brazilian government's development of Brasília as the new national capital in the 1950s. Similar clearance and relocation schemes were adopted in Asia, with Singapore and Hong Kong being the most notable examples.

While the housing provided through relocation schemes was of better quality and had basic urban services, such schemes often failed in the aim of providing housing for the poorest urban dwellers, although the Singapore and Hong Kong examples were certainly much more successful than their Latin American counterparts. These failures were largely due to planners' limited understanding of the lives of the poor residents. The new flats had low rents relative to other formal housing in the city, but even these low rents were too high for most of the urban poor and the need for regular payments did not fit with the often sporadic and precarious earning capacity of workers largely employed in the informal sector. Additional costs for water and electricity also made the housing costs too high. The result was that many relocated families sold or sublet their flats to middle class households who could afford to live there and returned to informal settlement living.

The location of the new flats was also a severe problem. Residents in informal settlements were often able to operate businesses from their homes, or chose to live near their place of work. Being uprooted from economic and social networks to be relocated elsewhere often destroyed these economic opportunities just at the time when the cost of living was going up because of rent and service costs. Families might also have had a small amount of space to keep animals or grow crops in the informal settlement; activities which were very difficult to continue in the apartment blocks.

The general failure of such clearance and relocation schemes, combined with increasing pressures on state finances, meant that they became much less common in the 1970s and beyond. This was also because of changing attitudes toward informal settlements and their residents by governments and international agencies. Clearance of informal settlements still takes place, not least as part of urban regeneration projects associated with gentrification, the rise of middle class consumer culture, and the demand for large shopping malls, and high profile international events such as the Olympics or a regional heads of state meeting.

Upgrading and Self-Help

An ignorance of the lives of people in the informal settlements lay behind many of the problems identified with state housing schemes. In the late 1960s and early 1970s, there was some change in perspectives on this sector of the urban population due to research. The two key individuals involved were William Mangin, an anthropologist, and John Turner, an urban planner. They both conducted research in the squatter settlements of Lima, Peru. Rather than being havens for criminals, the only option for the urban poor and home to the large migrant population that had flocked to the city to seek a better life, Mangin and Turner's research concluded that residents of the *pueblos jovenes* or *barriadas*, as they were termed, had often lived in the city for many years, were hardworking, and were seeking to improve their housing conditions as a form of investment for their children.

Turner developed the concept of 'bridgeheaders' and 'consolidators' as a way of categorizing poor urban dwellers in Lima. This idea was based on the recognition that the urban poor, including migrants, made rational decisions about where and how to live. On first arriving in the city, the key issue for migrants was employment, so they tended to live in rental accommodation in over crowded tenements in the city center. These were the 'bridgeheaders'. After a few years when they had become more established in the city, Turner argued that the priority shifted to housing and investment for the future. Given the cost of land and housing in the city center, squatting on land on the urban margins and constructing a house as and when you could afford it was a rational decision. Over time, housing quality would improve when families had funds available and communities would work together to access services from the gov ernment. These people were termed 'consolidators' and the process of improvement over time was called 'up grading'. As a result of this research, the peripheral in formal settlements were often viewed as 'slums of hope' in contrast to the 'slums of despair' in the city center.

The research of Mangin and Turner resonated with what was being seen elsewhere and encouraged many governments to view informal settlements in a more positive light. Here was a solution to housing problems for the poor which did not involve high levels of gov ernment spending and allowed the urban poor to build the housing appropriate to their lifestyles and income levels. This approach was also taken up by international agencies such as the World Bank.

Rather than implementing state house building and settlement clearance, governments and inter national agencies increasingly sought to provide support to the self build process. In some cases, this involved the regularization of land tenure so that householders were given the legal title to their land and therefore the security to improve their houses without fear of eviction. In many cases, however, households improved their homes without the official paperwork as they felt secure enough due to political support or other indications of official approval, such as service provision. Some gov ernments also developed schemes to provide subsidized building materials for home improvements. In most cities, major infrastructure services were provided by the state as in the case of the high profile Orangi Pilot Project in Karachi, Pakistan.

Having supposedly learned from earlier mistakes, some governments began to develop site and services schemes. The World Bank also invested millions of dol lars in such schemes in over 50 countries during the period 1972–1990. Households were provided (usually at a low cost) with a plot of land which had connections to water, sanitation, and electricity services. They were then able to build the kind of house they wanted when they had available funds. The premise of the schemes was that it was easier and cheaper to put the services in before the development of the settlement and by giving households the freedom to build the house they wanted, the gov ernment was not trying to house everyone in uniform residences. While such schemes received large amounts of funding, issues around cost and location proved highly problematic once again. Households with low or irregular incomes were unable to participate, and this particularly hit women headed households who tended to have low incomes. As with the earlier relocation schemes, many site and service plots were sold or sublet to less poor families.

Self help housing of this sort often failed to house the very poorest. Its adoption by governments also received criticisms as it was viewed as a way in which governments could save money and ignore their responsibilities to the poorest people in the city. The fact that the ideas and policies were largely based on the Latin American ex perience, particularly that of Lima, also needs to be considered. Lima had vast areas of arid land surrounding the city where squatters could move to build their homes; not all cities have this space available. Additionally, compared to many residents in cities in sub Saharan Africa and South Asia, Latin American poor urban dwellers often had more economic opportunities and therefore chances to invest in their own homes. For many migrants to cities in sub Saharan Africa, the lack of housing in the city centers means that the marginal in formal settlements are the only options; again contrasting with the bridgeheader–consolidator model developed in the Lima case.

Processes of upgrading and self help have led to shifts in the urban structure of many cities. What were previ ously areas of poor quality housing occupied by very poor families have become middle class districts with all urban services. However, this is not always correlated with the improved lives of the original residents. As services are introduced and settlements are formally recognized, the cost of living there increases, so meaning that original residents may be forced to move elsewhere. **Figure 1** shows one of the main streets in Colonia Aurora on the northwestern edge of Oaxaca City, Southern Mexico. This settlement started in the 1970s as an in formal subdivision and was occupied by residents from elsewhere in the city and also new migrants from the rural areas. By 1990, some residents had been able to build brick dwellers and there was electricity supply, although not all houses were connected. There were no connections to the water or sanitation system. By 2002, water and sewerage systems had been installed, the road was paved, and housing quality had improved greatly (see **Figure 2**). Some of the families who had lived there in

Figure 1 Street in Colonia Aurora, Oaxaca City, Mexico, 1990.

Figure 2 Street in Colonia Aurora, Oaxaca City, Mexico, 2002.

1990 were still residents, but others had been unable to afford the increased cost of living and had moved away. Others, seeing the chance for income generation, rented out their homes or sold them. Thus, upgrading may not lead to improved housing for all.

The State as Enabler or Facilitator

Since the early 1980s, government housing policy in the Global South and approaches adopted by international agencies have increasingly moved toward seeing government as an enabler or facilitator, rather than as a provider of housing. Within this model, housing is provided by the residents themselves, carrying on with the self build idea, and services and support are provided by nongovernmental organizations (NGOs) or the private sector. These policies clearly resonate with the broader neoliberal agenda which has become so prevalent.

Rather than providing housing, the state will now provide an environment that will facilitate the development of housing markets. It is argued that this is more efficient and provides greater choice for all urban residents. One of the key enabling processes relates to land titles. It is argued by researchers such as Hernando de Soto that housing provision and economic development in the cities of the Global South are limited by the lack of legal tenure and the amount of time it takes for tenure legalization due to bureaucratic procedures. Providing home owners in informal settlements with legal titles to their land and homes will, it is argued, give them the security to improve their houses, and will also enable them to sell their property, should they wish to. In addition, legal home ownership provides collateral to borrow money for other investments, such as home improvements and business expansion.

Receiving legal land title is popular with residents, but it does not necessarily have the expected outcomes. As outlined earlier, while residents of informal settlements may not have formal land titles, they have often felt secure enough to improve their property, and an informal land and housing market has developed in most cities. Legal titles and tenure regularization may also push out the poorest and most vulnerable households from informal settlements as it is associated with increased costs.

Housing finance has become an increasingly important issue, especially as squatting as a route to access land has become scarcer around urban areas. Within an enabling framework, governments now attempt to promote private sector or NGO financing activities, often based on a micro finance model. This has achieved success in some cases, but again there are limits for households with very low or irregular incomes. Using housing as collateral for formal bank loans has also been effective for some, but this varies greatly; some banks still view low income applicants with suspicion, even if they do have formal title to their property.

As infrastructure services have been increasingly privatized in cities of the Global South, issues around cost recovery have become more pressing. Private sector water, sanitation, and electricity companies have been encouraged to provide services to low income residents, including those in informal settlements, by being allowed to charge for these services. It is argued that this is because governments lack the funding and capacity to provide such services themselves. Given the importance of such basic services, many governments have implemented policies to ensure that the very poorest are not excluded. Such policies include fixed tariffs and subsidized services. In the case of South Africa, the very poorest urban households are provided with a set amount of free water per month. Consumption above this level must be paid for.

By involving informal settlement dwellers in both the financing and the management of upgrading activities, governments, UN Habitat, and many NGOs are echoing Mangin and Turner's ideas from the 1960s regarding the agency and capacity of such people. Such policies can lead to great improvements in housing quality and the concomitant betterment of health and environmental conditions. However, it is vital that the poorest households are not excluded from such processes and that the participation of residents is meaningful, rather than purely for presentational purposes.

Conclusions

Squatter settlements or informal settlements have been a very important part of many cities in the Global South. Shifting government and international agency attitudes toward them since the 1960s have reflected a growing recognition of the capacity of the urban poor to adapt and sometimes to thrive in very difficult circumstances. As the world's urban population grows, there will be increasing pressure on both land and housing. The shift toward market mechanisms for both land and housing delivery has been beneficial in some cases, but without forms of support and protection, millions of poor households will be excluded and left to fend for themselves in the diminishing number of available spaces in the world's cities.

See also: Housing; Neoliberalism and Development; Postcolonial Cities; Poverty; Third World Cities.

Further Reading

Datta, K. and Jones, G. (2001). Housing finance in developing countries: Invisible issues on research and policy agendas. *Habitat International* 25, 333 357.

Davis, M. (2006). *Planet of Slums*. London: Verso.

De Soto, H. (2000). *The Mystery of Capitalism: Why Capitalism Triumphs in the West and Fails Everywhere Else*. London: Black Swan.

Drakakis Smith, D. (2000). *Third World Cities* (2nd edn.). London: Routledge.

Gilbert, A. (2002). On the mystery of capital and the myths of Hernando de Soto. *International Development Planning Review* 24, 1 20.

Hardoy, J. and Satterthwaite, D. (1989). *Squatter Citizen*. London: Earthscan.

International Development Planning Review, 2006. Special Issue on State Society Relations in Land Delivery Processes in African Cities. *International Development Planning Review* 28, 127 285.

Mangin, W. (1967). Latin American squatter settlements: A problem and a solution. *Latin American Research Review* 2, 65 98.

Moser, C. and Peake, L. (eds.) (1987). *Women, Human Settlement and Housing*. London: Routledge.

Potter, R. B. and Lloyd Evans, S. (1998). *The City in the Developing World*. Harlow: Longman.

Turner, J. F. C. (1968). Housing priorities, settlement patterns and urban development in modernizing countries. *Journal of the American Institute of Planners* 33, 167 181.

UN Habitat (2003). *Slums of the World: The Face of Urban Poverty in the New Millennium?* Nairobi: UN Habitat.

Verma, G. (2002). *Slumming India: A Chronicle of Slums and Their Saviours*. London: Penguin.

Ward, P. M. (1982). *Self Help Housing: A Critique*. London: Mansell Publishing Ltd.

Relevant Websites

http://www.bshf.org
Building and Social Housing Foundation. Independent research organization which focuses on the development of people centered and environmentally sustainable housing policy.

http://www.citiesalliance.org
Cities Alliance. A global coalition seeking to eradicate slums and urban poverty.

http://www.sdinet.org
Slum/Shack Dwellers International. Network of urban poor from the Global South who share experiences and approaches to housing improvement.

http://www.unhabitat.org
United Nations Habitat. United Nations agency responsible for human settlements.

http://www.makingcitieswork.org
USAID Urban Policy.

State

M. Jones, Aberystwyth University, Aberystwyth, UK

Introduction

Perhaps the most difficult task in this article is defining the state, and this has been a challenge for political and human geographers, as well as social and political scientists. As we shall see in this article, the state is a highly complex phenomenon and no single approach can ever fully capture this. States and interstate systems provide a moving target because of their complex developmental logics and because continuing attempts to structure and restructure states and the state system leave their own traces on forms, functions, and activities. History, geography, and social agency matter. The state is, in short, a series of multifarious processes in motion and to get a handle on this we need to freeze the moment and see what is there, and go back and do this several times over. Getting a grasp on the state is further complicated because of the embedded relationship between states, society, and space. There can be no adequate theory or position on the state without a corresponding theory or take on society and its spatial relations. The state and the political system are bound up with the broader geographical ensemble of social relations and one cannot adequately describe or explain the state, state apparatus, state projects, or anything else associated with the state, without referring to its relationship with such ensembles. All this calls for sensitivity to the historical and institutional specificity of the state's social and spatial development. This gives political geographers a unique opportunity to say something interesting and this article seeks to tackle this.

The article has three sections. First, it outlines the different dimensions of the state, somewhat of a cookery book approach, which helps to set the scene further. It then turns to look at how the state can be theorized from different popular standpoints. The discussion focuses on three distinct approaches: Marxism, Weberian historical sociology, and post structuralism. All these have been used, at different times, in political geography and human geography more broadly. It then seeks to put the state in its place by looking at developments across time and over space. Periodization matters and we consider historical and geographical specificity. This is done to get beyond two interrelated blocks – the 'territorial trap' and 'methodological nationalism' – that have hindered and can still hinder a political geography of the state, especially that pervading theory construction and empirical research. Such 'traps' comprise three assumptions. First, the state is said to possess sovereign and mutually exclusive control over its territorial borders. Second, binary position of 'domestic' and 'foreign' is adopted as a fixed feature of the interstate system. Third, the state is conceived as a static, timeless territorial 'container' that encloses economic and political processes. The approach in this article is more variegated and intimately geographical. Future issues for research and reflection conclude the discussion.

Tracing the State

The most basic but least asked question, by all scholars and not just geographers, in thinking about the state is defining the 'properties' of the state. The most general feature of the state (premodern as well as modern and precapitalist as well as capitalist) is that it comprises a set of institutions concerned with the territorialization of political power. This involves the intersection of politically organized, coercive, and symbolic power, a clearly demarcated core territory, and a fixable population on which political decisions may be made collectively binding (such as the exercise of law and order, and the collection of taxation and other sources of revenue). Thus, the key feature of the state is the historically variable ensemble of technologies and practices that produce, naturalize, and manage territorial space as a relatively bounded container within which political power is exercised to achieve various, more or less well integrated, and changing policy objectives. A system of formally sovereign, mutually recognizing, mutually legitimating national states exercising sovereign control over large and exclusive territorial areas – the contemporary scenario – is only a relatively recent institutional expression of state power. Other modes of territorializing political power have existed, some still coexist with the so called Westphalian system (initially established by the Treaties of Westphalia in 1648 but realized only stepwise, at times chaotically, and often subject to reversals, during the nineteenth and twentieth centuries), new expressions are emerging, and yet others can be imagined. In short, there remains much work to be done on the state by political geographers.

The state, then, is distinct and different from say a multinational corporation, by virtue of its territorial integrity and its political legitimacy. The state is also different in the various roles that it can play. As we shall see below, to know the state we need to also know the history and geography of capitalism. States can respond to the contradictions, dilemmas, and problems of capitalism by creating the general conditions for the production and social reproduction of the capital relation, that is, the environment for economic growth and development. The state does this in part by seeking to regulate even development and/or by responding to the effects of this, that is, uneven growth, change, and restructuring. The state though is omnipresent: due to its development and penetration into most spheres of life, it is everywhere and nowhere. This presents geographers with challenges of analysis and interpretation.

Adopting these general parameters enables us to note that political power is not always exercised in and through territory and to explore the different modalities of territorialization, deterritorialization, and reterritorialization. For territory is not a given but something that gets constructed and institutionalized in different ways in different times and places. Recognizing this from a political geography standpoint should exclude criticisms of 'spatial fetishism', that is, the belief that space as such had a linear causal efficacy in relation to the state. 'State/space', as discussed elsewhere, is a lot more contingent, as the terrain of the state is forged through the ongoing engagements between agents, institutions, and concrete political and policy circumstances. The engagements change across time and space, which makes things even more interesting. One way of holding all this down to grapple with the state is to first dissect the state's 'institutional relations' and second then to consider the ways in which the state interacts with its wider 'social environment', taking note of the points above. These two divisions can be subdivided further into analytical categories for conducting research, which we will now consider. In short, there are six dimensions to the state's organization form and sociopolitical bases. Political geography can help us look at these.

Starting first with the 'institutional relations' of state, it is possible to identify a 'representational regime'. This is concerned with delimiting patterns of representation and the state in its inclusive sense. It uncovers the territorial agents, political parties, state officials, community groups, parastate institutions, regimes, and coalitions that are incorporated into the state's everyday policymaking practices. These are the state's peopling and peopled properties. We can then identify the 'internal structures of the state'. This is the institutional embodiment of the above and it underscores the distribution of powers through different geographical divisions and departments of the state and its policy systems. This reading not only allows research to study the apparatus of government; but it also explores the ways in which political strategy helps to create spaces and scales of policy intervention and delivery. For political geographers, the relationship between the different politically and socially charged scales of intervention is interesting. Last, we can think about 'patterns of intervention' to analyze the different political and ideological rule systems that govern state intervention, such as frameworks of rights and responsibilities, the balance between the public and private, and the perceived roles of the social partners in the policy process. Concerns here can include the discourses of citizenship, social inclusion/exclusion, universal versus targeted and selective service provision, and equality versus allocation through competition.

Second, turning to the wider 'social environment' and social relations, and taking note of the point above that we cannot have a position on the state without a perspective on society and space, the 'social basis of the state and its power' draws attention to the consolidation of the representational regime through civil society, that is, those social forces outside the political system. We can explore here the different ways in which uneven development is mobilized into the political system through targeted state strategies. In turn, thinking about 'state strategies and state projects' brings some overall coherence to the activities of the state, its forms of intervention, and its policymaking priorities. The state is seen here as a political strategy and its various policy and power networks can privilege some coalition possibilities over others and some interest groups over others. Last, by using the idea of 'hegemonic projects', we can examine language and other semiotic codes, and seek to uncover ideological programs of action, that is, how forms of knowledge and discourses become codified and mobilized to advance particular interests.

Theorizing the State: Three Positions

These properties of the state are mostly accepted by those working on the state apparatus. The question becomes how do we use and interpret this framework and what drives the various categories of analysis, and how is the state ultimately linked to perspectives on society? The state cannot be reduced or restricted to studies of government and a political geography of the state has to proceed in tandem with more general theoretical and empirical research on the geography of socioeconomic relations. We will now look at this.

Marxist perspectives on the state have certainly been popular in political geography over the years. With its emphasis on accumulation, regulation to counteract uneven development, and the role of history and materialism in making capitalist space economies, Marxism was well placed to capture the changing capitalist world of the 1970s and 1980s. This perspective has the benefit of highlighting big issue 'why' questions and a neat and often closed explanatory power, for the main tenet of geographical Marxist criticism – that sociospatial re lations are constructed and constituted according to the political–economic circumstance in and through which they are produced – might well immediately be accepted as self evidently true. The difficulty and controversy lies entirely in deciding how close the influence is. The 'determinist' position argues that the state is formed by, and acts in accordance with, the forces of capital and class, which are intrinsically interlinked through the ongoing territorialization of political power. Those that own and control the means of production ultimately have privileged access to the state apparatus (representational regime and internal structures) and are able to direct and redirect resources (patterns of intervention) to serve their own class and social interests. All this is reinforced by the control of ideology (hegemonic projects), in what can often read like a conspiracy, when things in reality are more complex.

There are, of course, many different versions of the Marxist thesis, with recent years accepting the critiques of essentialism (the essential role of structural forms like the economy driving things) and exploring a 'liberal' line, where the forces of capital and class are much more distant, subtle, and can be traced through the distinctive nature of political strategy. Those advocating a neo Marxist regulationist approach, for instance, have ex plored this road and are interested in causal powers (seen through 'state strategies and state projects') that bring the forces of capital and class to bear on society and space. The leading edge of this is the strategic relational ap proach (SRA), formulated and deployed by Jessop. This draws on Marxist accounts of capital being a social re lation and applies this to the state as social relation so that the state is seen as neither a neutral instrument (equally accessible to all forces and useful for any pur pose) nor a rational calculating subject (with a pregiven unity and clear purposes). State power is strategically and spatially selective and results from a continuing interaction between structures, strategies, and balance of forces operating within and beyond the state. The state needs to be thus thought of as a 'medium and outcome' of processes that constitute its many interventions. The state is both a social relation and a producer of strategy and, as such, it does not have any power of its own, other than the forces acting in and through its apparatus.

Historical sociology, and especially drawing on the work of Weber, offers a different perspective on the state and state power. In contrast to more structural ap proaches above, they are interested in the polymorphous crystallization of state power, which assumes a territorial shape through the interplay between the states 'infra structural' or 'administrative' power in and across space. Here, the state is an institutional ensemble of social powers that can be organized under the dominance of various macro social logics, with four being highlighted: economic, politico military, representational, and cul tural ideological. This approach is partly a critique of Marxist essentialism, where the economy is determinant in the last instance, with other possibilities being im agined and detailed. Thus, a historical sociological anal ysis of state formation emphasizes the important role for long periods of cultural ideological, politico military, and representational principles. The balance between these can be highlighted historically by drawing on the various sources of social power and interesting accounts have been presented on this. Critics of this approach do not deny the influence of, say, the state's politico military dimensions – particularly in the current era – but cannot accept that this is either foundational or can possibly exhaust the complexities of the modern world.

Post structuralism sits at the other end of the spec trum to Marxism. Difficulties and controversies with Marxist approaches to the state are self evident and have centered on reductionism, functionalism, teleology, class bias, structural determinism, and among other things the writing of peopled experiences out of the script of cap italist geographies. This is one major entry point for post structuralism – a diverse field covering everything from studies of narrative, subjectivity and body, the new his toricism, and philosophies of deconstruction. All these probably have one thing in common. Ontology (forms of knowledge on the state) is always already an outcome of the epistemology (means to knowing the state) of our socially constructed ways of knowing. Post structuralists insist that all knowledge and subjectivity, for instance on the state, is not out there waiting to be found but is antiessentialist. Elements of ontology are always framed in thought in the first instance and meanings and interpretations become fixed and known through social practices, power relations, and geographical sites of performance. This does not make knowing the state impossible; we instead need to develop frameworks on the 'state in action'. Political geographers have used this in studies of critical geopolitics, whereby international statecraft is made through discursive practices to repre sent the world as characterized by particular types of people, places, and dramas. Visual performances of the social field are presently in vogue. Research on everyday, or prosaic, 'state effects' is also becoming popular. Here, we are drawn to studying the situated, performative, complex, chaotic, and problematic nature of existence and the multifarious struggles and articulations of pol itics. Things are messy and cannot be explained by a single causal force. Last, there has been an explosion of interest in the techniques and technologies of governing, using notions of 'governmentality' to examine the his torically evolving and shifting mentality of government. This is useful for looking at the creation of political objects, subjects, and spaces of government and govern ance. Critics of post structuralist approaches to the state normally take issue with their arguments on the every where nature of the state and state power (its supposed immanence) and urge accounts to focus on structural forces, causal properties, and how mechanisms trigger and situate those contingent everyday experiences.

Despite the nature of some approaches, arguments, languages, and writing styles even in contemporary geography, none of the three approaches is necessarily more right than the others. Claims to 'truth' on the state ultimately rest on the particular research problems being addressed, the time frame in which they are being explored, and the kinds of explanations ultimately being sought by geographers. Marxism, for instance, is not as helpful as post structuralism for investigating the everyday and mundane world of state practices, such as the regulation of social behavior and the changing nature of political expression. Conversely, post structuralism offers less of an insight than Marxism into technical change, geographical specificities of economic re structuring, the general changing spatial divisions of labor, and the scalar geometries of state spatial re structuring. Last, historical sociology perhaps has a unique handle on the state apparatus and its internal development, and is useful for conducting longitudinal studies of 'state/space' in this more functional manner. These are the theoretical horses for our research courses: we make a choice, run with it, and endeavor to hold on tight. Occasionally, it can be a bumpy ride. Nobody is necessarily right.

Historical Specificity: Periodization and Territory

One way of political geography making a difference to our understanding on the state is to explore the spatiality of the state, state restructuring, and state intervention through the creation of 'spatiotemporal fixes'. This sug gests that the state performs the role of securing the relative stabilization of society by endeavoring to manage the various economic and political contradictions within the state system. This is inherently spatial, and always temporary, as we shall see. The window into this per spective is normally a neo Marxist approach, building on the above, interested in exploring state intervention and action through constructions of spaces (scales, levels, horizons, etc.) of intervention, the fixing of borders, the stabilization of places, and in short, attempts made to produce and reproduce a territorially coherent and functioning socioeconomic landscape. Using the work of Brenner on 'new state spaces', the remainder of this section outlines this and is interested in the historical specificity of the state in mainly North America and Western Europe in three distinct phases: 'encagement' (1890s–1930s), 'entrenchment' (1950s–early 1970s), and 'de nationalization' (post 1970s onward). As noted in the 'Introduction', this challenges the national container like assumptions of state and political geography.

Until the late nineteenth century, there was a distinct polarization between industrializing city regions and predominately rural agricultural peripheries, which was reinforced at a global scale through colonialism and imperialist expansion. The global depression of the late 1890s and after, however, drove a search for a state managed and organized capitalism. Despite the con tinued exploitation of 'peripheral' zones in Asia, Africa, and Latin America, an emerging 'encagement' of socio economic relations was taking place due to the increasing spatial convergence between the interests of capital and the state, which fueled the development of global urban hierarchies and increasingly territorial states. Geoeco nomic interests began to dominate here and the con tinued expansion of capital growth and development drove the search for a spatial symmetry between a na tional economy, state bureaucracy, civil society, and na tional culture, all feeding each other. Infrastructure development across territorial space, industrial infra structures, national economic planning, and an expanding state apparatus helped to secure this. Over a 40 year period, the state thus acquired a major role in the con struction of numerous interlocking territorial infra structures – public transportation, education and housing facilities, communication networks, utility supplies, and other aspects of the expanding urban fabric.

Following the Great Depression of the 1930s, and certainly after World War II, this model of statehood was consolidated further through modernization, national ization, and rationalization to make space for a new wave of capitalist expansion, despite the onset of serious and lasting economic crises. Consequently, throughout the Fordist–Keynesian period, roughly from the 1950s to the early 1970s, the role of the territorial state as a geographical container of capital accumulation and social development intensified and solidified. Though propelled by the dynamism of large scale manufacturing regions in Northeastern USA, the English Midlands, Northern France, Belgium, and the Ruhr district, the Fordist–Keynesian model of economic development was arranged into distinctly national space economies within a world system organized under the geopolitical and geoeco nomic hegemony of the USA. Control of finance was key here, first through a stable international monetary regime and second an increasingly interventionist state regime capable of influencing the spatial location of capital and managing demand throughout the economy. Already during the 1930s states had begun to engage directly in attempts to devalue and revalue capital through subsi dies, grants, loans, tax advantages, public investments, and state ownership of production facilities. This practice progressed and regional and local state institutions became transmission belts for interventionist central state politicis. Their goal was simultaneously to maximize growth and to redistribute its effects as evenly as possible on a national scale. Between the 1950s and the 1970s, a range of regional industrial policies were introduced to promote industrialization within each state's under developed peripheries. This was supported by the ex pansion of the welfare state, increasing universal rights and social citizenship, and politics predominately related to achieving and raising social benefits. The state's role in territorializing capital and in securing a spatiotemporal fix for social relations, therefore, converged around the national scale. In this manner, the national scale came to operate as the most fundamental geographical framework for capitalist production and exchange, as the key insti tutional site of sociospatial polarization, and as the most central arena of sociopolitical contestation. This 'entrenchment' period can, therefore, be characterized as that of the Keynesian Welfare National State (KWNS).

'De nationalization' has had more profound impli cations on geography of world capitalism and the polit ical geography of the state than the two preceding waves of spatiotemporal fixing. On suprastate levels, since the 1970s, new macro geographies have been emerging and consolidating as Fordist–Keynesian national economies are being challenged and superseded by a configuration of the world economy dominated by the super regional blocs of Europe, North America, and East Asia. Shifts within the international division of labor, aided by technical change, have brought about intense economic restructuring, which has transformed some industrial heartlands into wastelands and provided new post industrial opportunities for others. Accordingly, on substate levels, interspatial competition has intensified among cities and regions competing with one another to attract capital investment and secure state subsidies. Meanwhile, a global mosaic of world cities and indus trialized urban regions has emerged as a crucial terri torial infrastructure for the activities of transnational corporations. Consequently, the circulation of capital can no longer be adequately conceived with reference to autocentric national economies or the image of a world economy packaged into distinctive national territorial spaces. A mosaic of postnational socioeconomic and sociospatial relations has been created.

In addition to these globally induced reconfigurations of the national scale, there has been a series of highly contested forms of reterritorialization involving the state and its interventions, restructuring impacts, and re sponses to this. Two overarching broad trends are worth mentioning and these have been defining research agendas in political geography and human geography more broadly in recent years.

First, scholars have highlighted a 'relativization of statehood'. In contrast to the era of Fordism–Keynesianism, described above, there is no privileged level around which the state can influence the unfolding of capitalism. The relativization of statehood thus im plies: (1) an increasingly tangled hierarchy of overlap ping, continually changing arrangements associated with multilevel interventions; and (2) the systematic lack of any dominant scale, or system of metagovernance, that encompasses or subsumes competing scales of political economic organization. To understand this, the crisis of Fordist capitalism and the subsequent burst of global ization should be understood as a de centering or 'hol lowing out' of nationally scaled regulatory arrangements, as subnational and supranational scales of political eco nomic life have acquired a renewed, and in some cases unprecedented, significance through a variety of trial and error, often *ad hoc*, political initiatives, in especially economic development and regeneration. This often in volves the shift from government to governance, as new arrangements for state management now involve an array of different socioeconomic partners and not just those in formalized state structures. The challenges of this are considered in the conclusion.

While the Keynesian welfare national states of the post war era were intent to harmonize the equalization of wealth, population, and infrastructure across national territories, contemporary post Keynesian state projects are promoting territorial competitiveness within certain strategic subnational sites such as cities, city regions, and industrial districts, which in turn are to be positioned within supranational and global circuits of economic development. While certain aspects of this entre preneurial reorientation of local and regional policy has occurred from below, as fiscally stressed localities and regional states have attempted proactively to attract new sources of investment through the actions of develop ment agencies and other institutions, the 'new localism' and the 'new regionalism' must also be construed, to a significant degree, as national state projects. Indeed, provoked by hegemonic discourses of globalization and business civilization alongside a political rhetoric of fiscal prudence, national states have actively sought to reduce commitments to integrated welfare entitlements and redistributive urban and regional policies in favor of supply side neoliberal initiatives intended to promote technological innovation, labor market flexibility, lean management, and endogenous growth.

Second, these state activities have been bringing with them 'authoritarian politics', that is, they can have deep consequences across cities and regions as inherited and hard won forms of popular control over welfare state institutions and capitalist firms are being gradually ero ded. Insofar as they have entailed a 'productivist' re ordering of social policy and the increasing privileging of economic competitiveness, shifts to entrepreneurial governance have also contributed to a reduced

Table 1 Properties of the KWNS and SWPR

	Economic policies	*Social policies*	*Scale*	*Means of compensating market failure*
KWNS	Full employment, demand management, and provision of infrastructure to support mass production and consumption	Collective bargaining and state help generalize norms of mass consumption; expansion of welfare rights	Relative primacy of national scale in economic and social policymaking with local as well as central delivery	Market and state form a 'mixed economy'. State is expected to compensate for market failures
	Keynesian	*Welfare*	*National*	*State*
SWPR	Focuses on innovation and competitiveness in open economies, with increasing stress on supply-side reform	Subordinates social policy to an expanded notion of economic policy; downward pressure on 'social wage' and attack on welfare rights	Relativization of scale at expense of national scale. Competition to establish a new primary scale but continued role of national state(s)	Increased role of self-organizing governance to correct both market and state failures, but state gains greater role in the exercise of metagovernance
	Schumpeterian	*Workfare*	*Postnational*	*Regime*

Based on Jessop, B. (2002). *The Future of the Capitalist State.* Cambridge: Polity.

commitment to the concerns of collective consumption for poorer sections of the working class. The reworking of state spatiality and territory in major urban regions has also frequently been linked to the construction of nonaccountable institutional arrangements, dominated by corporate elites and property owners who are largely insulated from democratic control and subordinated to narrow productivist goals, such as performance measures and target setting. Finally, the rescaling of political economic governance has also entailed significant trans formations of urban and regional space, as large scale urban development projects and metropolitan insti tutional reforms are imposed from above, without ex tensive democratic deliberation or consultation. New landscapes of power have been created in which new forms of exclusion, subordination, and also resistance are inscribed into urban and regional space. One notable ingredient in this political endeavor has been 'workfare', which aims to secure a new relationship between the state and its subjects by requiring work or active labor market activities in return for unemployment benefit and welfare assistance. Collectively, then, we are witnessing the arrival of Schumpeterian Workfare Post National Regimes (SWPRs) to challenge what is left of the KWNS. **Table 1** summarizes this transition.

Future Issues and Directions for Research

This article concludes by raising three future issues for consideration on political geography and the state. We are living in exciting times and some things to think about are:

How is the 'relativization of statehood' unraveling in different geographical contexts and is there an emerging spatiotemporal fix to replace the primacy of national territory, national economic space, national social for mation, and national networks? Are new forms of state configuration emerging that seem to be consolidating and ossifying to create a relatively stabilized landscape of regulation, or are multilevel arrangements set to continue with an ongoing search for different possibilities? Polit ical geographers are getting excited at the moment about notions of multilevel governance, that is, the proliferation of different levels or scales of the state. Multiple rounds of 'filling in' are occurring, which challenges assumptions that with globalization the state is 'hollowing out', de clining, and being eroded away. This raises many re search questions, with some being: who has access to the state apparatus? And who participates when and how? What is the relationship between different types of state and nonstate actors and between different levels? How are issues of democracy and legitimacy dealt with? And how are the conflicting demands of competitiveness, co hesion, and sustainability dealt with?

Building on this, interesting work can be undertaken on the 'geographies of regulatory failure'. Given that state actors walk a difficult line between ensuring the right environment for economic growth and development but at the same time guarding against a crisis of electoral support (issues of legitimation crisis), instances of regu latory failure across cities and regions around the world are becoming apparent, as state policymaking constantly switches economic problems in concerns of state ra tionality that can be more easily addressed through public policy. State actors appear to be continually re inventing policy initiatives, often in response to the problems and contradictions caused by previous rounds of state intervention, in a search to get things right. Re search can be undertaken on how state actors are dealing with these challenges, that is, how is failure being presented and interpreted, addressed, new solutions proposed, and are these becoming successful? The growing knowledge based economy (KBE) across most of the developed world is a fertile ground to explore this, with governments searching for a fix to hold down globalization and ensure that their citizens have the skills they need for sustainable employment throughout their working lives. Attempts made to coordinate are also particularly interesting for political geographers to re search through work on metagovernance experiments. The spreading 'city regions' project is also something to watch, as it promises joined up thinking and super co ordination, social cohesion, economic competitiveness, and prosperity for all in a stretched relational new economy. This is currently political in vogue and is being implemented with very little critical analysis and reflection.

Drawing on more 'everyday' accounts of the state, there are numerous opportunities to undertake research on what appears to be an increasingly paternalistic ad vanced capitalist state, that is, the embedded links be tween the state in/as political society and the state in/as civil society. The 'authoritarian politics' highlighted above is one road into this topic with the globalization of welfare to work and workfare schemes as part of a 'new paternalism'. Taking more of a cultural political geog raphy slant, there is much research to be undertaken on more banal and prosaic interactions, which appear to be becoming more commonplace as state power spreads throughout civil society. Research can look at the pro duction and reproduction of nationalisms, regionalisms, localisms, and other spaces of identity and expression in a postnational state constellation. Devolution and consti tutional change across Western Europe is opening up spaces for this kind of cultural research. Subnational political geographies are being reinforced, resisted, re visited, and in some cases invented through language, history, and the symbolic naming and shaping of territory. The increasingly ideological nature of state power is also important to follow. Witness, for instance, the fostering of state sponsored environmental and sustain able citizenship regimes and responsibility codes in schools, institutions of higher education, the workplace, and the home – despite state government reliance on a rather unsustainable mode of growth across much of the world!

See also: Governance; Governmentality; Hegemony; Nation; Nationalism; Space II; Territory and Territoriality.

Further Reading

Agnew, J. and Corbridge, S. (1995). *Mastering Space: Hegemony, Territory, and International Political Economy*. London: Rouledge.

Brenner, N. (1998). Between fixity and motion: Accumulation, territorial organization and historical geography of spatial scales. *Environment and Planning D: Society and Space* 16, 459 481.

Brenner, N. (2004). *New State Spaces: Urban Governance and the Rescaling of Statehood*. Oxford: Oxford University Press.

Brenner, N., Jessop, B., Jones, M. and MacLeod, G. (eds.) (2002). *State/Space: A Reader*. Oxford: Blackwell.

Clark, G. L. and Dear, M. (1984). *State Apparatus: Structures and Language of Legitimacy*. London: Allen & Unwin.

Collinge, C. (1999). Self organization of society by scale: A spatial reworking of regulation theory. *Environment and Planning D: Society and Space* 17, 557 574.

Corbridge, S., Williams, G., Srivastava, M. and Véron, R. (2005). *Seeing the State: Governance and Governability in India*. Cambridge, MA: Cambridge University Press.

Jessop, B. (1990). *State Theory: Putting Capitalist States in Their Place*. Cambridge: Polity.

Jessop, B. (2002). *The Future of the Capitalist State*. Cambridge: Polity.

Jessop, N. (2006). Spatial fixes, temporal fixes and spatio temporal fixes. In Castree, N. & Gregory, D. (eds.) *David Harvey: A Critical Reader*, pp 142 166. Oxford: Blackwell.

Jessop, B. (2007). *State Power*. Cambridge: Polity.

Jones, M. (1997). Spatial selectivity of the state? The regulationist enigma and local struggles over economic governance. *Environment and Planning A* 29, 831 864.

Jones, M. and Jones, R. (2004). Nation states, ideological power and globalisation: Can geographers catch the boat? *Geoforum* 35, 409 424.

Jones, R. (2007). *People/States/Territories*. Oxford: Blackwell.

Kirby, A. (1993). *Power/Resistance: Local Politics and the Chaotic State*. Bloomington, IN: Indiana University Press.

Mann, M. (1986). *The Source of Social Power*, vol. 1. Cambridge, MA: Cambridge University Press.

Mountz, A. (2007). Smoke and mirrors: An ethnography of the state. In Tickell, A., Sheppard, E., Peck, J. & Barnes, T. (eds.) *Politics and Practice in Economic Geography*, pp 38 48. London: Sage.

Painter, J. (2006). Prosaic geographies of stateness. *Political Geography* 25, 752 774.

Swyngedouw, E. (2000). Authoritarian governance, power and the politics of rescaling. *Environment and Planning D: Society and Space* 18, 63 76.

State Theory

B. Jessop, Lancaster University, Lancaster, UK

What Is the State?

Everyday language sometimes depicts the state as a subject – the state does, or must do, this or that; and sometimes as a thing – this economic class, social stratum, political party, or official caste uses the state to pursue its projects or interests. But how could the state act 'as if' it were a unified subject and what could constitute its unity as a 'thing'? Coherent answers are hard because the state's referents vary so much. When pressed, a common response is to list the institutions that comprise the state, usually with a core set of institutions with increasingly vague outer boundaries. From the executive, legislature, judiciary, army, police, and public administration, the list may extend to education, trade unions, mass media, religion, and even the family. Such lists typically fail to specify what lends these institutions the quality of statehood and who can legitimately be described as an agent of the state.

One alternative is to define the state in terms of means rather than ends or statal entities. This approach informs Weber's celebrated definition of the 'modern' state as the "human community that (successfully) claims the *monopoly of the legitimate use of physical force* within a given territory" (Weber, 1948: 78), as well as definitions that highlight its formal sovereignty vis à vis its own population and other states. This does not mean that modern states exercise power largely through direct and immediate coercion but that coercion is their last resort in enforcing decisions. For, when state power is regarded as legitimate, they can normally secure compliance without such recourse. Even then states reserve the right – or claim the need – to suspend the constitution or specific legal provisions in exceptional conditions and many rely heavily for their survival on force, fraud, and corruption and their subjects' inability to organize effective resistance.

Other theorists regard the essence of the state (premodern as well as modern) as the territorialization of political authority. This involves the intersection of politically organized coercive and symbolic power, a clearly demarcated core territory, and a fixed population on which political decisions are held to be collectively binding. Thus the key feature of the state is the historically variable ensemble of technologies and practices that produce, naturalize, and manage territorial space as a bounded container within which political power is then exercised to achieve various policy objectives. The so called Westphalian system of independent sovereign states exercising control over large and exclusive territorial areas dates back formally to treaties signed in 1648 but, as an effective institutional expression of state power, it has been more or less fully realized only in the last 60 years and was being challenged even as it was being completed. Other modes of territorializing political power have existed; some still coexist with the Westphalian system; new expressions are emerging; and others can be imagined. Thus one can ask whether the European Union (EU) is a new form of state power, a rescaled 'national' state, a revival of Medieval political patterns, a post sovereign form of authority, or a new type of empire. Similarly, are we moving toward global governance or even a world state?

The Origins of the State and State-Building

State formation is not a once and for all process nor did the state develop in just one place and then spread elsewhere. It has been invented many times, had its ups and downs, and seen recurrent cycles of political centralization and decentralization, territorialization and deterritorialization. Although its origins have been explained in various monocausal ways, none provides a convincing general explanation. Marxists focus on the emergence of economic surplus to support a specialized, economically unproductive political apparatus that seeks to secure cohesion in a (class)divided society; military historians focus on the role of military conquest in state building and/or the demands of territorial defense in stimulating state capacities to penetrate and organize society. Others emphasize the role of a specialized priesthood and organized religion (or other forms of ideological power) in giving symbolic unity to the population governed by the state. Feminist theorists have examined the role of patriarchy in state formation and the state's continuing role in reproducing gender divisions. And yet other scholars focus on the 'imagined political communities' around which nation states have been constructed (classically Anderson 1991).

The best approach is multicausal and recognizes that states change continually, may well break down, and must be rebuilt in new forms, with new capacities and functions, new scales of operation, and a predisposition to new types of failure. In this context, as Mann notes, the state is polymorphous – its organization and capacities can be primarily capitalist, military, theocratic, or democratic and its dominant 'crystallization' is liable to challenge as well as conjunctural variation. There is no guarantee that the modern state will always (or ever) be primarily capitalist in character and, even where this is the case, state managers typically take account of other functional demands and civil society in order to promote institutional integration and social cohesion within its territorial boundaries. Whether it succeeds is another matter.

Modern state formation has been analyzed from four main perspectives. First, there are historical studies of the path dependent origins and development of one or more parts of the modern state (such as standing armies, modern tax system, formal bureaucracy, parliament, universal suffrage, citizenship rights, and recognition by other states). Such studies assume that the history of the modern state is decomposable into the histories of its constituent parts and, as such, can be written for all states. Second, a necessarily more restricted set of studies focuses on how just some states have acquired certain distinctive 'formal' features of a modern state, such as its institutional separation from other spheres of society, its own political rationale and modes of operation, and distinctive constitutional legitimation, based on adherence to its own procedures rather than values such as divine right or natural law. Not all states display these features, of course, and those that do acquire them at different stages in their development may also experience reversals. This approach poses additional questions about other state forms – authoritarian or totalitarian regimes and/or kleptocratic, narco, and other failed states. Third, other scholars study the elaboration and pursuit of state projects that give a substantive (as opposed to merely formal) unity to state actions, enabling the state to act 'as if' it were relatively unified. They characterize states in terms of their dominant state projects, for example, mercantilist, liberal, welfare, fascist, developmental, or competition states. Fourth, 'configurational' analyses explore the distinctive nature of state–civil society relations and seek to locate state formation within wider historical developments. Rokkan on the intersection of economic, religious, political, and geographical cleavages in state formation in Europe over the last 400–500 years is exemplary.

Marxist Approaches to the State

There have been many attempts to complete the Marxist theory of the state based on selective interpretations of the disparate writings of Marx and Engels. Most attempts tend to explain the state's form and/or functions in economic terms and/or tend to see the state as a simple instrument that can be wielded by the temporarily dominant class forces for any purpose. They thereby ignore the distinctive 'political' logic of the state in capitalist societies as an institutional ensemble separate from the profit oriented, market mediated logic of capital accumulation as well as Marx's argument that the form of state that underwrites the economic exploitation and political oppression of the proletariat (and other subordinate classes) cannot be the instrument of their emancipation. Analyses affirming one or both of these points were revived in the 1960s and 1970s in response to the seeming ability of Keynesian welfare states to secure crisis free economic expansion and/or to the 'end of ideology' allegedly produced by boom conditions. It was argued that, notwithstanding economic growth and the extension of social welfare, states could not abolish capital's contradictions and crisis tendencies and still had a key role in maintaining class domination. This raised the question of 'relative autonomy', a topic much debated in the 1970s and 1980s, which essentially explored the relative freedom of states (or, better, state managers) to pursue policies against the 'immediate' interests of the dominant economic class(es) without becoming so autonomous that they undermined the latter's long term interests too.

The best work in this period formulated two key insights with a relevance that extends far beyond Marxist concerns. First, some scholars explored how the typical form of the capitalist state caused real problems for capital accumulation and political class domination. For the institutional separation of the state from the market economy, a separation regarded as a necessary and defining feature of capitalist societies leads to the dominance of different (and potentially contradictory) institutional logics and modes of calculation in each social field. Thus, political outcomes may not serve the overall needs of capital – even if states (and state managers) are operationally autonomous – because they face distinctive political pressures that point in other directions. This conclusion fueled work on the structural contradictions and strategic dilemmas of specific state forms as well as interest in how the complex interplay of social struggles and institutions influenced state formation. Second, Marxist theorists began to analyze state power as a complex social relation. This involved studies of the biases inscribed in different state forms and political regimes and the factors that shaped their strategic capacities. Attention focused on the variability of these capacities and their impact on the state's ability to project power into social realms well beyond its institutional boundaries. This led in turn to more complex studies of struggles, institutions, and political capacities.

State-Centered Theories

The flourishing of Marxist state theories in the 1970s prompted a counter movement in the 1980s to 'bring the state back in' as a critical explanatory variable in social analysis. 'State centered' theorists argued that state activities and their impact are easily explained in terms of its distinctive properties as an administrative or repressive organ and/or the equally distinctive properties of the broader political system encompassing the state. Societal factors, when not irrelevant, were certainly secondary; and their impact on state affairs was always filtered through the political system and the state itself. The main focus of state centered work was detailed case studies of state building, policymaking, and implementation. Six themes were emphasized: (1) the geopolitical position of different states in the interstate system and its implications for the logic of state action; (2) the dynamic of military organization and the impact of warfare in the overall development of the state – reflected in Tilly's claim that, not only do states make war, but wars make states; (3) the state's distinctive administrative powers – especially those rooted in its capacities to produce and enforce collectively binding decisions within a centrally organized, territorially bounded society – and its strategic reach in relation to all other social subsystems (including the economy), organizations (including capitalist enterprises), and forces (including classes) within its domain; (4) the state's role as a distinctive factor in shaping institutions, group formation, interest articulation, political capacities, ideas, and demands beyond the state; (5) the distinctive pathologies of government and the political system – such as bureaucratism, political corruption, government overload, or state failure; and (6) the distinctive interests and capacities of 'state managers' (career officials, elected politicians, and so on). The main conclusions were that there are distinctive political pressures and processes that shape the state's form and functions; give it a real and important autonomy when faced with pressures and forces emerging from the wider society; and thereby endow it with a unique and irreplaceable centrality both in national life and the international order. In short, the state is a force in its own right and does not just serve the economy or civil society.

'State centered' theorists argued that it is quite usual for the state to exercise autonomy in pursuit of its own, quite distinctive, interests. They emphasize: state managers' ability to exercise power independently of (and even in the face of resistance from) nonstate forces – especially where a pluralistic universe of social forces gives them room for maneuver; and the grounding of this ability in the state's distinctive political resources and its ability to use these to penetrate, control, supervise, police, and discipline modern societies. Neostatists also argue that state autonomy is not a fixed structural feature of each and every governmental system but differs across states, by policy area, and over time. This is due partly to external limits on the scope for autonomous state action and partly to variations in state managers' capacity and readiness to pursue a strategy independent of nonstate actors.

Neostatist empirical research has usefully countered one sided class and capital theoretical work but it has also been criticized on various grounds. The most serious charge is that it rests on a fundamental theoretical fallacy. It posits clear and unambiguous boundaries between the state apparatus and society, state managers and social forces, and state power and societal power; the state can therefore be studied in isolation from society. This renders absolute what are really emergent, partial, unstable, and variable distinctions. This excludes hybrid logics such as corporatism or policy networks; divisions among state managers due to ties between state organs and other social spheres; and many other overlaps between state and society. If this assumption is rejected, however, little is left to distinguish state and society centered approaches.

Foucauldian Approaches

Foucault aimed to challenge the centrality of the state, sovereignty, or law in the study of power relations by advancing three main criticisms of earlier work. First, state theory is essentialist because it tries to explain the state and state power in terms of their own inherent, pre given properties. Instead, it should explain state development and functioning as contingent outcomes of practices that are not necessarily (if at all) located within, or openly oriented to, the state. Second, state theory tends to explore the modern state in terms of 'Medieval' notions of monarchical authority and/or a unified sovereign power. This ignores the tremendous dispersion and variety of institutions and practices involved in exercising state power. Third, state theorists adopt a top down approach to state power. They are preoccupied with the summits of the state apparatus, the legitimation of political sovereignty, and the state's capacity to penetrate society from above. Foucault advocated a bottom up analysis of the many sites where power is actually exercised and may only subsequently be codified into a general pattern of domination. His 'micro physics of power' focused on actual practices of subjugation rather than macro political strategies. For state power involves the active mobilization of individuals and not just their passive targeting; and it is associated with multiple discourses, strategies, and institutions. In short, power is not concentrated in the state: it is ubiquitous, dispersed, and immanent in every social relation.

Nonetheless Foucault did not reject all concern with macro political questions. He came to see the state as the crucial site of statecraft and 'governmentality' (or governmental rationality). What interested him was the art of government, a skilled practice involving the reflexive monitoring of the population and, with all due prudence, attempts to align it behind specific state projects. *Raison d'état*, an autonomous political rationality, set apart from religion and morality, was the key to the rise of the modern state. It was linked to different modes of political calculation or state projects, such as the early modern 'police state' (where 'police' refers to policies oriented to the public good) or the late modern 'welfare state'. These governmental rationalities or state projects colonize local sites of power and link them into ever more general mechanisms of overall domination. Foucault also identified the close connections between power and knowledge, highlighting the state's role in surveillance, gathering knowledge about individuals and the overall population, shaping individuals as specific types of subject (e.g., pupil, patient, parent, prisoner, worker, citizen), and securing an appropriately balanced population. More recent studies inspired by Foucault have addressed the reorganization of the state and the reorientation of statecraft (especially in a neoliberal direction) and the conditions that enable states to engage in effective action across many social domains.

Feminist Approaches

While feminists have carefully explored the gendering of social relations and powerfully criticized 'malestream' political philosophy and political theory, they have been less interested in developing a feminist theory of the state. The best work shows that patriarchal and gender relations make a difference to the state but it also refuses to prejudge the form and effects of this difference. Thus, "acknowledging that gender inequality exists does not automatically imply that every capitalist state is involved in the reproduction of that inequality in the same ways or to the same extent" (Jenson, 1986). An extensive literature on the complex and variable intersections of class, gender, and ethnicity in particular state structures and policy areas has since revealed the limits of gender essentialism. This approach has been taken further by third wave feminists and 'queer theorists', who stress the social construction and instability of dominant views of sexual and gender identities and establish the variability of masculine as well as feminine identities and interests. This has prompted revival of explicit interest in gender and the state, including the construction of masculinity and femininity, the variability of gender identities, interests, roles, and bodily forms, and the issue of how some of these *relata* are privileged in state discourses, institutions, and material practices. This rules out analyses of the state as a simple expression of patriarchal domination and questions the patriarchy's utility as an analytical category. But feminists reveal basic flaws in malestream theorizing and indicate the need for research on gender biases in state structures and capacities as well as its fundamental role in reproducing specific patterns of gender relations.

Discourse Analysis and Stateless State Theory

Discourse analysis studies the production of intersubjective meaning in and through a wide range of oral, written, audio–visual, and digital 'texts' and their implications for identities and practices. Some recent work in this tradition suggests that the state does not exist but is, rather, an illusion – a product of political imaginaries. For belief in its existence depends on state discourses that provide a vocabulary of motives and a political grammar that orient (in)actions toward the 'state', leading social forces to act 'as if' it existed. Since there is no common discourse of the state (at most there is a dominant or hegemonic discourse) and different forces orient their action at different times to different ideas of the state, the state is basically a polyvalent phenomenon that changes shape and appearance with the forces acting toward it and the circumstances in which they act. There is also growing interest in specific narrative, rhetorical, or argumentative features of state power. Thus case studies of policymaking suggest that state policies do not objectively represent interests located in or beyond the state or objectively reflect 'real' problems confronting the political system. Instead policies are discursively mediated, if not wholly discursively constituted, products of struggles to define and narrate 'problems' which can be dealt with in and through state action. The impact of policymaking and implementation is therefore closely tied to their rhetorical and argumentative framing and this also holds, *a fortiori*, for the capacity of different parts of the state apparatus to work together in a relatively coherent manner. Such discourse theoretical work clearly differs from state centered theorizing and a Foucauldian micro physics of power. In contrast to neostatism, it rejects the reification of the state; and, in contrast to a bottom up approach to state power, it highlights the critical role of narrative and rhetorical practices in creating belief in the existence of the state. This role is variously defined as mystification, self motivation, pure narrativity, or self description but, regardless of standpoint, discourses about the state have a key role in shaping the state as a complex ensemble of political relations linked to society as a whole.

New Directions of Research

Despite declining interest in the more esoteric and abstract modes of state theorizing, substantive research on states and state power exploded from the 1990s onwards. Major themes include the variability of statehood (or stateness); the relative strength or weakness of states; the future of the national state in an era of globalization and regionalization; the state's changing forms and functions; scale, space, territoriality, and the state; and the rise of governance and its links to government.

First, interest in stateness arises from growing disquiet about the abstract nature of much state theory (especially its assumption of a ubiquitous, unified, sovereign state) and rising interest in the historical variability of actual states. Some theorists focus on the state as a conceptual variable and examine the varied presence of the idea of the state. Others examine the state's differential presence as a distinctive political form. Thus Badie and Birnbaum usefully distinguish between the 'political center' required in any complex social division of labor and the 'state' as one possible institutional locus of this center. For them, the state is defined by its structural differentiation, autonomy, universalism, and institutional solidity. France is the archetypal state in a centralized society; Britain has a political center but no state; Germany has a state but no center; and Switzerland has neither. Such approaches historicize the state idea and stress its institutional variety from the local to the international scale.

Second, there is growing interest in factors that make for state strength. Internally, this refers to a state's capacities to exercise authority over events and social forces in the wider society; externally, it refers to the state's power in the interstate system. This concern is especially marked in recent theoretical and empirical work on predatory and/or developmental states. The former are essentially parasitic upon their economy and civil society, exercise largely the despotic power of command, and may eventually undermine the economy, society, and the state itself. Developmental states also have infrastructural and network power and use it in allegedly market conforming ways. Unfortunately, the wide variety of interpretations of strength (and weakness) threatens coherent analysis. States have been described as strong because they have a large public sector, authoritarian rule, strong societal support, a weak and gelatinous civil society, cohesive bureaucracies, an interventionist policy, or the power to limit external interference. In addition, some studies risk tautology insofar as strength is defined purely in terms of outcomes. A possible theoretical solution is to investigate the scope for variability in state capacities by policy area, over time, and in specific conjunctures.

Third, recent work on globalization casts doubt on the future of national states in general and nation states in particular. This issue is also raised by scholars interested in the proliferation of scales on which significant state activities occur, from the local through the urban and regional to cross border and continental cooperation and a range of supranational entities. However, initial forecasts of the imminent demise of the national territorial state and/or nation state have proven erroneous. This reflects the adaptability of state managers and state apparatuses, the continued importance of national states in securing conditions for economic competitiveness, political legitimacy, social cohesion, and so on, and the role of national states in coordinating the state activities on other scales from the local to the triad to the international and global levels.

Fourth, following a temporary decline in Marxist theoretical work, interest has grown in the changing forms and functions of capitalist states. Key issues include the state's role in: (1) securing conditions for private profit – the field of economic policy; (2) reproducing the population, especially wage labor, on a daily, lifetime, and intergenerational basis – the broad field of social policy; (3) managing the scalar division of labor; and (4) compensating for market failure. On this basis Jessop characterizes the typical state form in the advanced capitalist economies during the post war expansion as a Keynesian welfare national state. Its distinctive features were an economic policy oriented to securing the conditions for full employment in a relatively closed economy, generalizing norms of mass consumption through the welfare state, the primacy of the national scale of policymaking, and the primacy of state intervention in compensating for market failure. He also describes the emerging state form from the 1980s onwards as a Schumpeterian workfare postnational regime. Its distinctive features are an economic policy oriented to innovation and competitiveness in relatively open economies, the subordination of social policy to economic demands, the relativization of scale with the movement of state powers downwards, upwards, and sideways, and the increased importance of governance mechanisms in compensating for market failure. Other types of state, including developmental states, have been discussed in the same terms.

Fifth, there is interest in the changing scales of politics. While some theorists tend to see the crisis of the national state as displacing the primary scale of political organization and action to the global or regional scale, others argue for a relativization of scale. For, whereas the Fordist period of post war European and North American expansion was associated with national economies, national states, and national societies, the after Fordist period has witnessed dispersion of economic, political, and sociocultural issues across many scales of organization, with none of them clearly primary. This in turn poses problems about securing the coherence of action across different scales. This has prompted interest in the novelty of the EU as a new state form, the reemergence of empire as an organizing principle, the prospects for a global state.

Sixth, with growing recognition of state as well as market failure, interest has grown in modes of 'governance' that rely neither on government command nor the anarchy of market forces but on different types of self organization. Governance operates on many scales (from the expansion of international and supranational regimes through national or regional public–private partnerships to more localized networks of power and decision making). Although this trend is often taken to imply a diminution in state capacities, it could well enhance the (national) state's power to secure its interests and, indeed, provide it with new (or expanded) role in the metagovernance (or overall coordination) of different governance regimes and mechanisms.

Finally, interest in governance is sometimes linked to the question of 'failed' and 'rogue' states. All states fail in certain respects and normal politics is an important mechanism for learning from, and adapting to, failure. In contrast, 'failed states' lack the capacity to reinvent or reorient their activities in the face of recurrent state failure in order to maintain 'normal political service' in domestic policies. The discourse of 'failed states' is often used to stigmatize some regimes as part of interstate as well as domestic politics. Similarly, the term 'rogue state' is used to denigrate states whose actions are considered by hegemonic or dominant states in the interstate system to threaten the prevailing international order. Moreover, according to some radical critics, the USA itself has been the worst rogue state for many years.

Final Remarks

The state presents researchers with a challenge: it is tangible and elusive, seemingly constant but changeable. While its analysis is central to political geography, especially because a defining feature of the state is its role in the territorialization of political power, states and state power also impact, directly or indirectly, all other aspects of social life. This entry raises some central issues posed by the complexity of the state and emphasizes the need for various approaches that can, together, address its polymorphous character. Moreover, because state power cannot be studied in isolation from the wider sets of social relations in which states are embedded, it also discusses broader approaches that seek to put the state in its place. Lastly, as the territorialization of political power has been undergoing important changes, emerging areas of research on states and state power have been identified.

See also: Governance; Scale; War.

Further Reading

Abrams, P. (1988). Notes on the difficulty of studying the state. *Journal of Historical Sociology* 1(1), 58 89.
Anderson, B. (1991). *Imagined Communities: Reflections on the Origin and Spread of Nationalism* (2nd edn.). London: Verso.
Badie, B. and Birnbaum, P. (2003). *The Sociology of the State*. Chicago, IL: University of Chicago Press.
Barrow, C. W. (1993). *Critical Theories of the State: Marxist, Neo Marxist, Post Marxist*. Madison, WI: University of Wisconsin Press.
Barry, A., Osborne, T. and Rose, N. (eds.) (1996). *Foucault and Political Reason*. London: UCL Press.
Beck, U. and Grande, E. (2007). *Cosmopolitan Europe*. Cambridge: Polity.
Chomsky, N. (2001). *Rogue States: The Rule of Force in World Affairs*. London: Pluto.
Draper, H. (1977). *Karl Marx's Theory of Revolution. Part I. State and Bureaucracy*. New York: Monthly Review Press.
Evans, P. B., Rueschemeyer, D. and Skocpol, T. (eds.) (1985). *Bringing the State Back In*. Cambridge: Cambridge University Press.
Fairclough, N. (1992). *Discourse and Social Change*. Cambridge: Polity.
Foucault, M. (1979). *Discipline and Punish: The Birth of the Prison*. Harmondsworth: Penguin.
Foucault, M. (1980). *Power/Knowledge: Selected Interviews and Other Writings 1972 1977*. New York: Pantheon.
Foucault, M. (2007). *Security, Territory, Population*. Basingstoke: Palgrave.
Hennessy, R. (2000). *Profit and Pleasure. Sexual Identities in Late Capitalism*. London: Routledge.
Jenson, J. (1986). Gender and reproduction: Or, babies and state. *Studies in Political Economy* 20, 9 46.
Jessop, B. (1982). *The Capitalist State*. Oxford: Martin Robertson.
Jessop, B. (2002). *The Future of the Capitalist State*. Cambridge: Polity.
Jessop, B. (2007). *State Power: A Strategic Relational Approach*. Cambridge: Polity.
Knutilla, M. and Kubik, W. (2001). *State Theories: Classical, Global and Feminist Perspectives*. London: Zed.
Lauridsen, L. S. (1991). The debate on the developmental state. In Martinussen, J. (ed.) *Development Theory and the Role of the State in Third World Countries,* paper no. 2. Roskilde: Roskilde University.
Mann, M. (1986). *The Sources of Social Power*, vol I. Cambridge: Cambridge University Press.
Melossi, D. (1990). *The State and Social Control*. Cambridge: Polity.
Rokkan, S. (1999). *State Formation, Nation Building and Mass Politics in Europe: The Theory of Stein Rokkan*. Oxford: Oxford University Press.
Rose, N. (1999). *Powers of Freedom*. Cambridge: Cambridge University Press.
Schram, S. F. and Neisser, P. T. (eds.) (1997). *Tales of the State*. Lanham, MD: Rowman & Littlefield.
Shaw, M. (2000). *Theory of the Global State*. Cambridge: Cambridge University Press.
Slaughter, A. M. (2004). *A New World Order*. Princeton, NJ: Princeton University Press.
Tilly, C. (1973). Reflections on the history of European state making. In Tilly, C. (ed.) *The Formation of National States in Western Europe*, 96pp. Princeton, NJ: Princeton University Press.
Weber, M. (1948). Politics as a vocation. In Gerth, H. H. & Mills, C. W. (eds.) *Essays from Max Weber*, pp 77 128. London: Routledge & Kegan Paul.
Zeitlin, J. and Trubek, D. M. (eds.) (2003). *Governing Work and Welfare in a New Economy*. Oxford: Oxford University Press.

Relevant Websites

http://www.cddc.vt.edu/feminism/
Feminist Theory Website, Virginia Tech University.
http://www.marxists.org/
Marxist Internet Archive.
http://www.psa.ac.uk/www/
Political Studies Association.
http://www.foucaultsociety.org/
The Foucault Society/Resources.
http://www.theglobalsite.ac.uk/index.html
The Global Site, Critical Gateway to World Politics, Society and Culture.

Statistics, Descriptive

J. Lee, Kent State University, Kent, OH, USA

Glossary

Kurtosis A statistic that measures the degree of peakedness in the distribution of values in a data series.
Maximum The largest numeric value in a data series.
Mean The arithmetic mean is commonly called the average. It is the result of dividing the sum of a set of numeric values by the number of the values.
Mean Center The location that is defined by the mean of the *x*-coordinates and the mean of the *y*-coordinates in a set of points.
Mean Deviation The average of absolute differences between individual values to the mean in a data series.
Median The value in the middle of an ordered series of numeric data.
Median Center The location as defined by the median of the *x*-coordinates and the median of the *y*-coordinates.
Minimum The smallest numeric value in a data series.
Mode The value that has the highest frequency in a data series.
Percentile It is a value on a scale of 100 that indicates the percent of a data series that is equal to or below it.
Range The numeric difference between the minimum and the maximum values in a data series.
Skewness A statistic that measures the degree of symmetry in the distribution of values in a data series.
Standard Deviation The square root of variance.
Standard Deviational Ellipse This is a closed curve formulated by modified standard distances with adjustments based on the orientation of data points in space.
Standard Distance This statistic measures the degree to which points are concentrated or dispersed around the mean center.
Variance The average of squared differences between individual values to the mean in a data series.

Descriptive and Spatial Statistics

When working with a set of numeric data, statistics are quantitative measures derived from the data to describe various aspects of the data. If classified by their functions, we have 'descriptive statistics' and 'inferential statistics'. Both can be used to analyze the data, to compare data with others, or to test the data against preformulated hypotheses. Descriptive statistics are calculated from a data series to describe how the values are distributed within the data series. For example, these include the maximum, minimum, range, percentile, mean, median, mode, standard deviation, variance, skewness, and kurtosis. Inferential statistics are calculated from sample data for the purpose of making an inference to a population or for making comparison between series of data values.

Depending on the area of application, 'classical statistics', or conventional statistics, are generally used in different application areas, such as sociology, political science, medicine, and engineering. These statistics have been modified or extended to accommodate specific application areas. For example, 'spatial statistics' are extensions of classical statistics that can be applied to deal with geographic data, or data with spatial references.

The fundamental assumption for any statistical analysis is that the data should be independent, or not influenced by each other. In human geography, however, this is often not the case because geographic objects or phenomena that we study could in some way be related to each other. This is often described as the 'first law of geography': "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970: 236). For this reason, many classical statistics have been extended to spatial statistics to account for the spatial association between geographic objects or phenomena being studied.

For example, we may want to do a study that compares the economic status of Portage County and Summit County, two adjacent counties in northeast Ohio. Recognizing that the two counties are of different sizes, we could start by randomly selecting 100 households in each county as the basis for comparison. First step in the analysis would be to calculate what the average household incomes are in these two counties. The average income figures allow us to make a quick comparison but they only offer a partial story to what the sampled household incomes provide. The average income figures may be affected by the minimum and maximum values in each set; so it would be good to know the range of data values in each set. Still, data values could have different degrees to which they cluster around the average figures or could have different proportion of income figures that are mostly in the higher end or in the lower end. The additional descriptive statistics of standard deviation, variance, skewness, kurtosis, and others can then be used to provide more details that better describe the distribution of sampled income figures from the two counties.

Remember that households have street addresses that can be converted to locations. Therefore, household in comes can also be considered to be spatial data. In that case, we could use spatial statistics to help us understand the spatial distribution of these income figures even better than using classical statistics alone. With mean center and median center, we can identify the center of a set of lo cations. With standard distance and standard deviational ellipse, we can assess how a set of locations may be clus tered or dispersed. Furthermore, it is even possible to give individual locations specific weights to reflect the im portance in that location with respect to others in the set.

In summary, descriptive statistics, classical or spatial, are measures that help us understand data series in terms of not only central tendency but also variation of values in the series. These statistics are simple to calculate but are powerful in providing snapshots of the data values and in providing the basis for further analysis.

Descriptive Statistics

For the purpose of illustration, let us use $x_1, x_2, x_3, \ldots, x_n$ to represent n numeric values in a data series, X. Fur thermore, let us simplify this representation to using x_i, where $i = 1, 2, \ldots, n$, to represent the ith numeric value in the series.

Minimum and Maximum

Given a set of 'numeric' values, the minimum and maximum can be easily identified as the smallest and the largest values in the set. If $x_1, x_2, x_3, \ldots, x_n$ are in as cending order, x_1 is the minimum and x_n is the maximum.

If data values in a distribution are categorical in nature and can be ordered, the identification of the minimum and maximum would depend on the ordering of the various categorical values. In general, the minimum value will be the value at the very beginning of the categories when they are ordered into an ascending order while the maximum would be the very last category. For example, areas of different altitudes may be classified into low altitude, medium altitude, and high altitude regions. Given a set of ten regions of these three categories, the minimum value would be the low altitude region while the maximum value would be the high altitude region. It is to be noted that categorical data normally are not suitable for any mathematical calculation such that none of the other descriptive statistics are applicable.

Range

For a set of 'numeric' values, the range is defined as the difference between the maximum and the minimum in the dataset. Simple arithmetic subtraction of the minimum from the maximum would yield the range. Using the same notation as defined above, the range of a distribution can be calculated as $x_n - x_1$. To provide a quick glance of the distribution of a dataset, a greater range implies a wider spread of data values in the dataset than a smaller range.

Percentile

Percentiles identify the positions of values in an ordered dataset as defined by grouping all values into 100 subsets or clusters. For example, the 50th percentile value in a dataset would have 50% of the data values above it and 50% of the data values below it. It is the middle value in an ordered dataset. Similarly, the 25th percentile in a dataset will have 25% of the data values before it and 75% of data values after it in the ordered dataset. Per centiles are commonly used for describing proportions of the data that should fall above or below a given value.

Mean

A very popular and useful representative of a set of 'numeric' values is the mean, or the average of all values in the dataset. This statistic is useful because it provides a quick snapshot of how the values in a dataset distribute. It is popular because it can be easily calculated. The mean of a set of data values can be calculated in two simple steps. First, add up all values in the dataset and then divide this sum by the number of data values in the set to yield the mean. Using the notation defined earlier, the mean, $\bar{X}$, is calculated as $(x_1 + x_2 + \cdots + x_n)/n$, for the n values in the dataset. In short, the formula for calculating the mean can be expressed as

$$X = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

or

$$X = \frac{\sum_{i\ \ 1}^{n} x_i}{n}$$

While easy to calculate, the mean of a data distri bution is sometimes affected by extreme values. Users of the mean should be sensitive to this possible problem when comparing means from different datasets. For ex ample, a set of values (1, 3, 5, 7, 9) have a mean of 5 because $5 = (1 + 3 + 5 + 7 + 9)/5$. This mean is not too different from the mean of 5.4 from another set of values: (3, 3, 5, 7, 9). If the first 3 in the second dataset is replaced by -100, such that the second set becomes (-100, 3, 5, 7, 9), the new mean from the second set will become -15.2, a significantly different value from 5 while four out of five values in each set are still identical. It can be easily seen that the value of the mean can be affected by only one extreme value in the dataset.

Median

One way to avoid the influence of extreme values in a dataset when measuring central tendency is to use the median value of the dataset. The median of a dataset is the data value in the middle of the ordered data values. Alternatively, the median can be thought of as the 50th percentile in a data series.

In x_i, $i = 1, n$, if n is an odd number, the median is $x_{(n+1/2)}$. If n is an even number, the median is $x_{(n/2)}$. Using the earlier example of the datasets (1, 3, 5, 7, 9) and (−100, 3, 5, 7, 9), the medians of these two datasets are both 5, not affected by the extreme value of −100.

The relative positions of the mean and median of a dataset depend on whether the data distribution is symmetric or asymmetric. In a symmetric distribution, the mean and the median are the same, indicating that both are at the center of the data distribution. Both the mean and the median mark the division of the data distribution into two equal parts: 50% of the data values on one hand and another 50% on the other. In an asymmetric distribution in which there are more data values in the series that are smaller than the mean than data values greater than the mean, the mean of the series will be to the right of the median because the mean will have a greater value than the median. Conversely, in a distribution that has more values greater than the mean than values less than the mean, the mean will have a smaller value than the median in the data series.

Mode

With categorical data, mean and median cannot be calculated. In those situations, mode is used. The mode of a set of data values is defined as the value that occurs most frequently in the dataset. In other words, the mode is the value that has the highest frequency. While used less often for numeric data, mode can also be identified in a set of numeric values. It provides another way to represent the distribution of values in a dataset. Since mode is defined as the value occurring most frequently in a distribution, it is located at the peak point of a frequency distribution curve. Together, the mean, the median, and the mode are three indicators of central tendency in a data distribution.

Depending on whether the data distribution is symmetric or asymmetric, their relative positions can be outlined in **Figure 1**. In **Figure 1**, vertical axes are frequency of data values while the horizontal axes are data values. The example being described (**Figure 1**) is a data series with more data values greater than the mean in the series than those smaller than the mean. As a result, the median is greater than the mean. This also highlights the fact that the mode does not need to be the mean and the median does not need to be the mode.

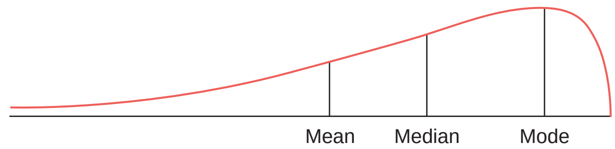

Figure 1 Relative positions of mean, median, and mode in an asymmetric frequency distribution.

Mean Deviation

Deviation is the difference between an individual value and the mean of that dataset. The mean deviation, in turn, is the average of all such deviations in a data series. It measures the average extent to which individual data values deviate from the mean. This statistic is often used to estimate how much individual values may deviate from their mean. Since some deviations are to be positive and others negative, absolute values are used in calculating the mean deviation. Specifically, the mean deviation, D, can be calculated as

$$D = \frac{\sum_{i=1}^{n} |(x_i - X)|}{n}$$

Variance

Mean, median, or mode can be used to provide a representative value of a set of data. While useful, none of these representative statistics inform us of how different values in the data series vary. To derive information on how values vary, the variance statistic can be calculated. Variance is the average of squared deviations between each individual data value and the mean of the dataset. Specifically, the variance, V, can be defined as

$$V = \frac{\sum_{i=1}^{n} (x_i - X)^2}{n}$$

Variance describes how individual values differ from the mean value of the dataset. It is similar to mean deviation except that the deviations are squared rather than being expressed in absolute values. Squaring eliminates the offsets between positive and negative deviations so that they can be summed and averaged.

Variance describes how wide or narrow the range of data values is. It provides a useful complement to the mean in describing the distribution of values in a dataset. In **Figure 2**, it can be seen that a distribution with a smaller variance has data values concentrated mostly around the mean while a distribution with a greater

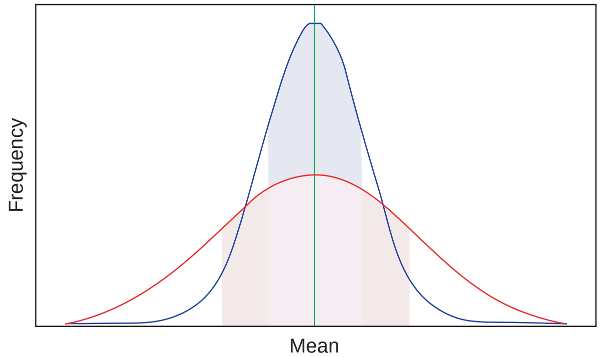

Figure 2 Two frequency distributions with different variances.

variance has data values concentrated less around the mean. The larger the variance, the greater the dispersion, or difference, of data values.

Standard Deviation

Standard deviation is defined as the square root of variance. Since variance is calculated as the average squared differences from individual values to their mean, taking the square root of variance will return this statistic to the same numeric magnitude as the data values. Therefore, standard deviation is at the same numeric scale as the data values. The standard deviation, S, can be calculated as

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - X)^2}{n}}$$

Because the standard deviation of a dataset is at the same numeric scale as that of the data values, it is used much more often than the variance in describing the degree of difference, or dispersion, among data values.

Skewness

Given the mean and the standard deviation, most distributions of data values can be described sufficiently. However, more descriptive statistics can be used to provide even more detail of the distributions. One of such descriptive statistics is skewness. This statistic provides information about whether a distribution is symmetric or skewed to either the end of lower values or the end of higher values. Skewness, SK, can be calculated as

$$SK = \frac{\sum_{i=1}^{n}(x_i - X)^3}{nS^3}$$

In **Figure 3**, the left diagram describes a normal curve that is symmetric and without skewness. The middle diagram describes a skewed distribution with more values smaller than the mean than values greater than the mean.

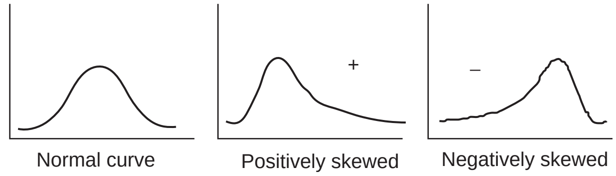

Figure 3 Skewness of frequency distributions.

The middle diagram is said to be positively skewed. The right diagram, in turn, has more values greater than the mean than values smaller than the mean and is negatively skewed.

Kurtosis

The last descriptive statistic is kurtosis, which provides information for the degree of peakedness of a data distribution. Peakedness in a data distribution is the degree to which data values are concentrated around the mean. Datasets with high kurtosis tend to have a distinct peak near the mean and tend to decline rapidly, and have heavy tails. Datasets with low kurtosis tend to have a flat top near the mean rather than a sharp peak.

Data distributions that have the same means, the same standard deviations, and the same degrees of skewness can have different kurtosis because they may have different degrees of the peakedness. For example, **Figure 4** describes two distributions with the same means and the same standard deviations but with obvious different degrees of peakedness around the means. The curve with solid line has greater kurtosis than that of the curve in dash line.

Kurtosis, K, can be calculated as

$$K = \frac{\sum_{i=1}^{n}(x_i - X)^4}{nS^4}$$

Greater values of K indicate higher degrees of peakedness whereas smaller K values indicate flatter distributions. In practice, however, we often subtract 3 from the calculated K values so that we would have positive K values or negative K values with respect to 3. This is because K value for a normal curve often approximates 3 and that comparisons are often made with respect to normal curves. In this case, a positive kurtosis indicates a distribution whose data values are more concentrated around the mean than those in a normal distribution. Similarly, a negative kurtosis, in this case, indicates a distribution that is more disperse than a normal distribution.

Grouped Data

When the objects of study are in large numbers or when it is not appropriate to reveal individual identities of the

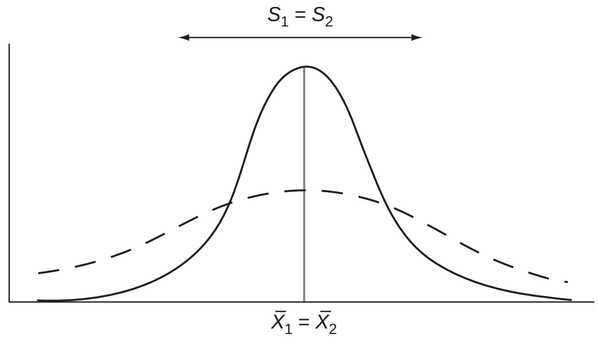

Figure 4 Two frequency distributions with the same mean and the same standard deviation but different kurtosis (peakedness).

study objects, researchers often group them into categories for further analysis. In such cases, a representative value, typically the midpoint value, from each category is used with the number of data values in the corresponding category for calculating descriptive statistics. For example, let p_i with $i = 1, n$ represent the midpoint value of the ith category and f_i as the number of data values (frequency) in that category, the calculation of the mean for grouped data, $\bar{X}_G$, can be summing the products of midpoints and frequencies and then dividing this sum with the total frequency

$$X_G = \frac{\sum_{i=1}^{n} f_i p_i}{\sum_{i=1}^{n} f_i}$$

Similarly, variance (V_G), standard deviation (S_G), skewness (SK_G), and kurtosis (K_G) for grouped data can be calculated by extending the corresponding formulas discussed above to the following:

$$V_G = \frac{\sum_{i=1}^{n} f_i (p_i - X_G)^2}{\sum_{i=1}^{n} f_i}$$

$$S_G = \sqrt{\frac{\sum_{i=1}^{n} f_i (p_i - X_G)^2}{\sum_{i=1}^{n} f_i}}$$

$$SK_G = \frac{\sum_{i=1}^{n} f_i (p_i - X_G)^3}{\left(\sum_{i=1}^{n} f_i\right) S_G^3}$$

and

$$K_G = \frac{\sum_{i=1}^{n} f_i (p_i - X_G)^4}{\left(\sum_{i=1}^{n} f_i\right) S_G^4}$$

With the proliferation of computers and computation technology, it seems that the descriptive statistics based on grouped data are less useful because we can always include every data value in our calculation. However, there are situations where such formulas should be used to better reflect the various characteristics of data distribution.

For example, for calculating the average family income of all 88 counties in the state of Ohio in the United States, we can use the 88 average family income figures in the state. Simply adding these 88 figures and then dividing their sum by 88 will yield an average that may not be very indicative to the true income trend. This is because the average family income figures from large counties are treated the same as those from small counties. A more desirable way to calculate this statistic will be to weight each average family income figure with the number of families in that county in the computation procedures. In essence, we treat each county as a 'group' and calculate the grouped mean of average family incomes by county.

Spatial Descriptive Statistics

In human geography, we often work with data that are associated with locations. We could be measuring areas of political units, we could be counting population of census tracts, we could be averaging family incomes among sampled neighborhoods, and so on. All these types of analyses call for additional information beyond what classical statistics can provide. For example, knowing the average family income of a set of sample neighborhoods may be achieved by calculating the mean of those family income figures for the neighborhoods in the study. However, we may want to know if these neighborhoods are very far apart from or clustered with each other. To assist analysts in describing such spatial patterns of the data and their locations, spatial descriptive statistics have been developed from extending the classical descriptive statistics.

Mean Center

Given a set of points, P, with n points defined in two dimensional space as $p_i = (x_i, y_i)$, where $i = 1, n$, the mean center, C, can be calculated as

$$C = (x_c, y_c) = \left(\frac{\sum_{i=1}^{n} x_i}{n}, \frac{\sum_{i=1}^{n} y_i}{n}\right)$$

The mean center gives an estimate of the spatial focal point of a set of points. It gives the central tendency of a set of points.

When points in the set have different relative importance, it is possible to calculate a weighted mean center by multiplying each coordinate with its predefined weight and then calculate the weighted mean from all of the weighted x coordinates and the weighted mean from all of the weighted y coordinates as

$$C_w = (x_{cw}, y_{cw}) = \left(\frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}, \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i} \right)$$

where w_i is the weight associated with point (x_i, y_i).

Median Center

Using the concept of the median, the median center is the point whose coordinates are defined as the median of all x coordinates and the median of all y coordinates of the geographic objects in the dataset. Expanding from the classic descriptive statistics, the mean center and the median center have similar pros and cons. The mean center may be more affected by any outliers than the median center. It should be noted that the median center cannot be precisely defined to a unique pair of x coordinate and y coordinate. This is because its x and y coordinates can be anywhere between the most centered geographic objects.

Standard Distance

Extending from the concept of standard deviation in classic statistics, the standard distance provides a similar measure in two dimensional space. Combining the standard deviation in x axis and that in y axis, the standard distance, SD, in a point set can be calculated as

$$SD = \sqrt{\frac{\sum_{i=1}^{n} (x_i - x_c)^2 + \sum_{i=1}^{n} (y_i - y_c)^2}{n}}$$

Similarly, if each point is associated with a predefined weight, then the weighted standard distance would be

$$SD = \sqrt{\frac{\sum_{i=1}^{n} w_i (x_i - x_{cw})^2 + \sum_{i=1}^{n} w_i (y_i - y_{cw})^2}{\sum_{i=1}^{n} w_i}}$$

When examining the spatial distribution of a set of points, analysts often draw a closed circular curve around the mean center with the standard distance as the radius. The size of this circle indicates how concentrated or how dispersed the set of points are around the mean center.

Standard Deviational Ellipse

The spatial distribution of a set of points is often affected by many factors such that the distribution may have a unique orientation, or directional bias. For example, housing settlements along a river valley are often dictated by the river course. Similarly, occurrences of algae on the surface of an elongated lake would tend to show an elongated pattern of point distribution. To account for this directional influence, the circle defined by the mean center and the standard distance is modified for that purpose to the standard deviational ellipse.

This ellipse can be calculated by first performing algebraic operations of transforming all coordinates to a new coordinate system that has the mean center as the origin and then apply an equation (given below) to calculate the directional angle that describes the directional bias. Finally, this angle is incorporated into a second set of equation to calculate the deviation along the x axis and the deviation along the y axis.

The transformation of point coordinates to a new coordinate system that has the mean center, (x_c, y_c), can be done by the following algebraic operations:

$$x_i' = x_i - x_c$$
$$y_i' = y_i - y_c$$

The calculation of the angle for rotating the point set to fit the directional bias is by this operation:

$$\tan\theta = \frac{\left(\sum_{i=1}^{n} (x_i')^2 - \sum_{i=1}^{n} (y_i')^2\right) + \sqrt{\left(\sum_{i=1}^{n} (x_i')^2 - \sum_{i=1}^{n} (y_i')^2\right)^2 + 4\left(\sum_{i=1}^{n} x_i' y_i'\right)^2}}{2\sum_{i=1}^{n} x_i' y_i'}$$

Concluding Remarks

Descriptive statistics provide summarizing information of the characteristics and distribution of values in one or more datasets. The classical descriptive statistics allow analysts to have a quick glance of the central tendency and the degree of dispersion of values in datasets. They are useful in understanding a data distribution and in comparing data distributions. For human geographers, it is often necessary to take into account the locational references of the data we work with. The spatial descriptive statistics allow analysts to assess the central tendency and variation of data in spatial context. The two types of descriptive statistics are complementary. Combining both statistics, analysts are able to study the geographic phenomena they work with.

See also: Distance.

Further Readings

Clark, W. A. V. and Hosking, P. L. (1986). *Statistical Methods for Geographers*. New York: Wiley.

Ebdon, D. (1988). *Statistics in Geography* (2nd edn.). Oxford: Blackwell.

Fotheringham, A. S., Brunsdon, C. and Charlton, M. (2000). *Quantitative Geography: Perspectives on Spatial Data Analysis*. London: Sage Publications.

Griffith, D. and Amrhein, C. G. (1991). *Statistical Analysis for Geographers*. Englewood Cliffs, NJ: Prentice Hall.

Lee, J. and Wong, D. W. S. (2001). *Statistical Analysis with ArcView GIS*. New York: Wiley.

Rogerson, P. (2001). *Statistical Methods for Geography*. London: Sage Publications.

Taylor, P. (1983). *Quantitative Methods in Geography: An Introduction to Spatial Analysis* (2nd edn.). Prospect Heights, IL: Waveland Press.

Tobler, W. R. (1970). A computer model simulation of urban growth in the Detroit region. *Economic Geography* 46(2), 234 240.

Wong, D. W. S. and Lee, J. (2005). *Statistical Analysis of Geographic Information with ArcView GIS and ArcGIS*. New York: Wiley.

Statistics, Inferential

D. M. Scott, McMaster University, Hamilton, ON, Canada

Glossary

Hypothesis An unproven statement concerning a problem under investigation.
Parameter A numerical descriptive measure of a population that is calculated from the values comprising the population.
Population A complete set of data (values or measurements) associated with some process or phenomenon.
Probability Sampling A form of sampling whereby every element contained within a sampling frame has a known probability of being included in the sample.
Random Variable A rule that assigns one numerical value to each simple event of an experiment.
Sample A subset of data selected from a population.
Sampling Frame A structure containing the entire set of elements from which a sample will be drawn.
Statistic A numerical descriptive measure of a sample that is calculated from the values comprising the sample.

Introduction

The goal of scientific research in human geography is to further our understanding of geographic processes and geographic phenomena. In statistics, the complete set of values or measurements pertaining to a particular process or phenomenon is called a population. The commuting times of all Canadian workers, for example, form a population. For a variety of practical and theoretical reasons, however, research typically proceeds based on a probability sample from a much larger (or even infinitely large) population. To be useful for scientific inquiry, the sample must be typical or representative of the population from which it is drawn – that is, it must accurately reflect the underlying characteristics of the population without bias. If this is indeed the case, then a human geographer can use inferential statistics to make generalizations about a population based on a sample, thus fulfilling the goal of scientific inquiry.

There are two basic types of statistical inference: estimation and hypothesis testing. In some instances, a human geographer might simply be interested in estimating a population characteristic, otherwise known as a parameter, from a sample statistic. For example, the mean commuting time of all Canadian workers could be estimated from a sample of commuting times. However, given that probability sampling involves some uncertainty, it is unlikely that the sample statistic will equal exactly the true population parameter. For this reason, a confidence interval is placed around the statistic. This is known as interval estimation.

Hypothesis testing is more involved than estimation and is often considered the core of statistical analysis due to the important role that it plays in scientific research. Simply put, in human geography, the formulation and testing of hypotheses is what furthers our understanding of geographic processes and geographic phenomena. For example, a hypothesis test could be used to determine if the commuting time for a sample of Canadian workers influences time spent participating in social activities (i.e., activities with co workers, friends, and family). If a significant relationship is found between the two variables, then it can be concluded that the relationship also holds for all such workers. In turn, this finding can help improve our understanding of how work travel impacts the quality of life of Canadian workers.

The key to understanding inferential statistics is to first understand what some may argue is the most important concept in statistics, the sampling distribution, which serves as the foundation for both interval estimation and hypothesis testing. Given their importance to inferential statistics, sampling distributions are discussed in detail in the next section. After this, estimation is addressed followed by a section detailing hypothesis testing. The material presented in the following three sections pertains to what is commonly known as classical inference. An alternative to this approach is Bayesian inference, which is mentioned briefly in the final section.

Sampling Distributions

Sample statistics such as the mean ($\bar{x}$), the standard deviation (s), and the proportion (p) are random variables meaning that their values can vary from one sample to another. As such, each statistic computed from a sample of size n drawn from a population is associated with a probability distribution known as the sampling distribution of the statistic. Simply put, a sampling distribution depicts the probability of occurrence (i.e., relative frequency) of all values a statistic can assume when computed for all possible independent, random samples of size n. The number of such samples, and therefore the number of statistics, is easily determined using combinatorial mathematics. Specifically, the number of ways of selecting a sample of size n from a population of size N

is computed as:

$$\binom{N}{n} = \frac{N!}{n!(N - n)!} \qquad [1]$$

where ! is the factorial symbol. Holding sample size constant, as population size increases and approaches infinity, the same is also true for the number of samples. On the other hand, holding population size constant and varying sample size results in a symmetrical distribution whereby the number of samples is maximized for a sample size that is half of the population size.

Sampling distributions are theoretical entities. While it is possible to draw all samples of a particular size from a finite population and compute a statistic for each sample, it is impractical to do so for a host of reasons, including the sheer number of samples that might have to be drawn (as suggested above), and more importantly, the simple fact that such an endeavor would require information for the entire population. Such information is rarely available (an exception would be a national, 100% census of the population), and if it is, then there is no need to use inferential statistics in the first place, which implies that there is no need to create a sampling distribution.

For the sake of making the concept of the sampling distribution, including its properties, more concrete, a series of three is constructed using data on commuting times for Canadian workers obtained from Cycle 19 of the General Social Survey of Canada. A total of 7309 respondents reported commuting times on their survey day. For illustrative purposes only, the commuting times of these workers are considered to be the population from which samples are drawn. The distribution of commuting times is shown in **Figure 1**. As can be seen, it is positively skewed (skewness has a value of 6.13), has a mean (μ) of 51.10 minutes, and a standard deviation (σ) of 54.43 minutes. Overall, the distribution is not normally distributed.

Although sampling distributions could be created for any one of a number of statistics, the mean ($\bar{x}$) is chosen for demonstration. Sampling distributions of the mean are derived for sample sizes of 250, 500, and 1000. In each instance, the sampling distribution is approximated by drawing 2500 random samples from the population using a computer program written explicitly for this purpose. The mean of each sample is then computed. Ideally, the sampling distributions should be created from all possible samples. However, in each case, the number, as calculated according to eqn [1], is astronomically large. The three sampling distributions, along with fitted normal curves, are shown in **Figure 2**.

Like any other distribution, the sampling distribution of a statistic has both a mean and a standard deviation, which is more commonly known as the standard error of the statistic. For each sampling distribution in **Figure 2**, the values of these statistics, respectively, μ_x and σ_x are shown. It is clear that for each sampling distribution, μ_x is extremely close to μ. Furthermore, it can be seen that as sample size increases, σ_x decreases. In turn, this implies that the mean ($\bar{x}$) of a larger sample has a tendency to be closer to μ than the mean of a smaller sample. Finally, despite the fact that the population is not normally distributed, the sampling distributions are.

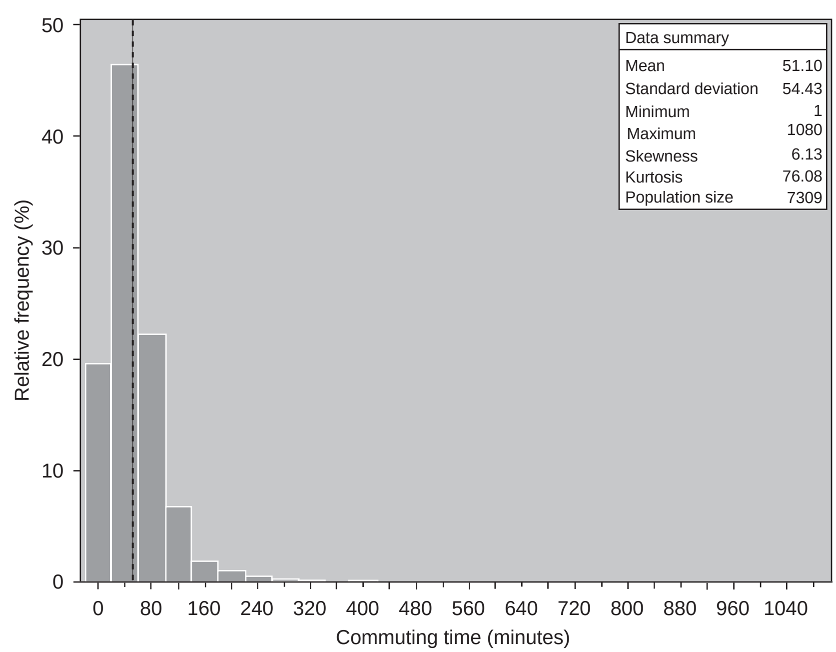

Figure 1 Distribution of commuting times for Canadian workers. From Cycle 19 of the General Social Survey of Canada.

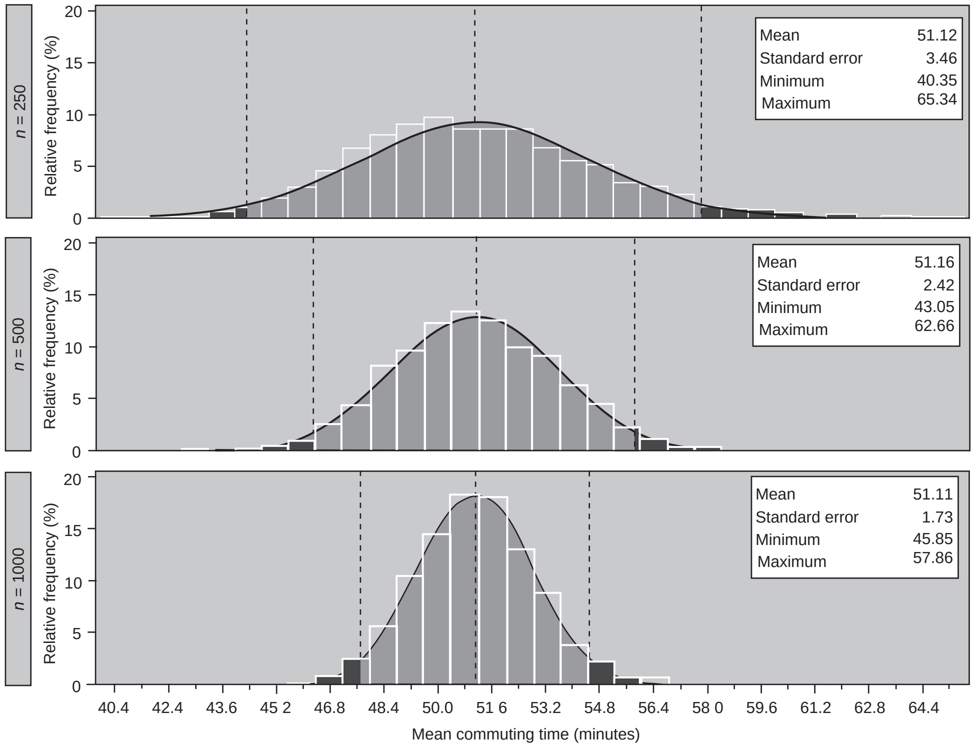

Figure 2 Sampling distributions of mean commuting times for Canadian workers for sample sizes of 250, 500, and 1000. Source: computed by author using data obtained from Cycle 19 of the General Social Survey of Canada.

For sample sizes of $n \geq 30$, the central limit theorem summarizes important properties of the sampling distribution of the mean, two of which are mentioned above. Specifically, the sampling distribution of the mean will have a mean (μ_x) equal to μ, have a standard deviation or standard error (σ_x) equal to $\sigma/\sqrt{n}$, and be normally distributed around the mean (μ_x).

Thus far, the discussion of sampling distributions has focused on those created by drawing all possible single samples from a population. However, sampling distributions can be created by drawing and computing statistics for all possible pairs of samples, triplets of samples, or other sets of samples. This is often the case for hypothesis testing. For example, the sampling distribution for the two sample difference of means z test is derived from the differences in sample means for all possible pairs of samples of sizes n_1 and n_2 drawn from a population where each sample size is ≥ 30. This sampling distribution is known as the sampling distribution of the difference between means.

It should be clear that many sampling distributions are used in inferential statistics. Each differs from the next according to sample size, whether single or multiple samples are drawn from a population, and the statistic being calculated. Fortunately, to reiterate, sampling distributions are theoretical entities, which means that we know what they look like without actually forming them. In practice, they are approximated by well known distributions such as the standard normal (z) distribution, the t distribution, the F distribution, and the chi square (χ^2) distribution, to name a few. The exact shape of virtually all such distributions is, however, determined by sample size, which is often measured in terms of degrees of freedom.

Estimation

Point Estimation

The simplest form of statistical inference is point estimation. Simply put, a statistic is calculated from a sample and used as an estimate of its corresponding population parameter. Such statistics are called point estimators. A point estimator is a rule or formula that is applied to a sample to calculate a single number or point that can be used as an estimate of a population parameter. For example, the sample mean ($\bar{x}$) is a point estimator of the population mean (μ). Likewise, the sample proportion (p) is a point estimator of the population proportion (ρ).

Often many different point estimators can be found to estimate the same population parameter. Information about each estimator is contained in its sampling distribution. Such information is crucial in determining the best point estimator or statistic for a particular parameter. For instance, a point estimator should be unbiased. An unbiased estimator is one whose sampling distribution has a mean that is equal to the population parameter that it intends to estimate. Obviously, this implies that if the mean of the sampling distribution is not equal to the population parameter, then the point estimator is biased – that is, it either underestimates or overestimates the pa rameter. The three sampling distributions of the mean shown in **Figure 2** provide irrefutable evidence that the sample mean ($\bar{x}$) is an unbiased estimator of the popu lation mean (μ). Specifically, the mean commuting time of each sampling distribution (μ_x) equals the mean com muting time of all workers (μ) – that is, 51.1 min.

Another feature of sampling distributions that can help determine the best point estimator for a particular parameter is the standard error of the statistic (i.e., the standard deviation of the sampling distribution). Quite simply, given a choice between two or more unbiased estimators, the one with the smallest standard error should be chosen. The reason for this is that standard error measures how closely values of a statistic cluster around an unknown population parameter. **Figure 2** il lustrates such a choice. The sampling distributions for three point estimators of the mean commuting time of all workers are shown. These estimators vary only accord ing to sample size (i.e., 250, 500, and 1000). Clearly, the smallest standard error, 1.73 min, is associated with the largest sample size. Given a choice between the three estimators, the sample mean ($\bar{x}$) derived from a sample size of 1000 is chosen.

In summary, when making an inference about a population parameter, the best point estimator should have an unbiased sampling distribution. Such estimators are a matter of record. Quite simply, the best (unbiased) point estimator of a population parameter is its corres ponding sample statistic. If a choice exists among two or more unbiased estimators, which is sometimes the case, then the best point estimator is the one whose sampling distribution has the smallest standard error.

Interval Estimation

We know from the discussion thus far that a point estimate of an unknown population parameter can vary from one sample of size n to another. Furthermore, the exact location of any particular statistic within its sampling distribution is unknown. Given these facts, it is necessary to assess the accuracy of any given estimate of a population parameter. Interval estimation is used for this task. Simply put, a range or interval of values that has a known probability of cap turing the population parameter being estimated is calcu lated from a sample. The formula that specifies how to use a sample to derive this interval is called an interval estimator. For example, the interval estimator for the population mean (μ) for large samples ($n \geq 30$) is

$$x \pm z_{\alpha/2}\sigma_{\bar{x}} \qquad [2]$$

where $\bar{x}$ is the sample mean, z is the z value (from the standard normal distribution) for a given significance level α, and σ_x is the standard error of the mean. For small samples ($n < 30$), z is replaced by a t value from the t dis tribution. Generally, interval estimators for other popu lation parameters are constructed in a fashion similar to that shown in eqn [2].

The range of values derived by an interval estimator, bounded by upper and lower values known as confidence limits, is called a confidence interval because there is a known probability that the interval surrounding a sample statistic captures the population parameter being esti mated. This probability and its expression as a percentage are respectively referred to as the confidence coefficient ($1 - \alpha$) and the confidence level for the interval estimate. Likewise, the probability that the population parameter falls outside the interval is called the significance level (α) of the estimate.

The concept of interval estimation is further ex plained by means of illustration by turning, once again, to **Figure 2**. For each sampling distribution, all of which are normally distributed, 95% of all mean commuting times fall within the lightly shaded region delimited by the outer dashed lines. This region is equivalent to 1.96 standard errors above and below the mean, which is identified by the dashed line at the center of each dis tribution. It follows that any sample mean computed from a random sample of size 250, 500, or 1000 has a probability of 0.95 of falling within the shaded region of its respective sampling distribution. If a sample mean does fall anywhere within this region, surrounding it on both sides by 1.96 standard errors (i.e., $\bar{x} \pm 1.96\sigma_x$) will identify a range of values that encompasses the mean commuting time of all workers (μ). The confidence limits provided by $\bar{x} \pm 1.96\sigma_x$ define the 95% confidence interval, which has a known probability (0.95) of cap turing the unknown population mean (μ).

As with point estimation, an interval estimate of a population parameter should be as precise as possible. This implies that narrow confidence intervals are pre ferred over wider ones. The width of a confidence interval, and therefore the precision of an estimate, is governed by three factors. All other things being equal, interval width increases as confidence level increases, as the amount of variability in the population increases, and as sample size decreases. Of these factors, we have little control over the second. However, to some extent, we can manipulate the first and third to produce narrower intervals for a price that involves trade offs between precision, accuracy, and cost/effort. In particular, when modifying confidence level, we must bear in mind that the trade off concerns precision and accuracy. Increased precision can only be achieved by reducing the con fidence level and therefore the accuracy of the estimate. Although we can produce interval estimates for virtually any confidence level, our choice is by and large tied to tradition – that is, the most commonly accepted confidence levels in practice are 90%, 95%, and 99%. When it comes to sample size, the trade off is between precision and cost/effort. Simply put, both precision and accuracy can be enhanced by purchasing a larger sample. However, the relationship between precision and cost/effort is subject to diminishing returns. In other words, the rate of reduction in interval width decreases with sample size, which means that a point is reached whereby it makes little sense to further increase sample size. Fortunately, interval estimators, such as that shown in eqn [2], can be manipulated to determine the minimum sample size needed to estimate a confidence interval with a desired width (precision) and confidence level (accuracy).

Hypothesis Testing

As mentioned earlier, hypothesis testing is more involved than estimation. To be more precise, it consists of six steps (**Table 1**). Together, these steps constitute a formal procedure whose goal is to draw inferences or conclusions about the values of one or more population parameters based on sample statistics estimating those parameters. As with interval estimation, hypothesis testing also involves some uncertainty – that is, it is possible to draw the wrong conclusions about values of population parameters based on sample information. This possibility is, however, controlled by the researcher who, as part of the formal hypothesis testing procedure, must assign a measurable probability of committing such an error to each conclusion.

Hypothesis testing begins by defining two complementary hypotheses. As such tests involve either a comparison of one or more population parameters or a relationship between two variables, the null hypothesis (H_0) usually states that no significant difference or relationship exists. Moreover, any observed difference or relationship is merely the result of sampling error (recall that values of a statistic vary from one sample to another). In contrast, the alternative hypothesis (H_A) states that the difference or relationship is indeed significant – that is, what is observed is too great to be attributed to sampling error. Furthermore, the alternative hypothesis takes one of two forms. It is either nondirectional or directional, which implies that it can be written using one of three inequalities: not equal ($\neq$), greater than ($>$), or less than ($<$). The choice between forms depends on whether *a priori* information exists to justify a statement of directionality. If such information does not exist, then the nondirectional alternative is specified.

Table 1 Hypothesis testing procedure

Step	*Task*
1	State null and alternative hypotheses
2	Select appropriate statistical test and list its assumptions
3	Select significance level (α)
4	Establish the decision rule for rejecting the null hypothesis
5	Calculate test statistic or its *p*-value
6	Make decision regarding the null hypothesis

The second step in hypothesis testing involves selecting an appropriate statistical test to address the research problem at hand. Oftentimes, this is perceived as the most difficult step in hypothesis testing given the wide variety of tests available to choose from. Fortunately, some simple rules of thumb can be used to guide this selection. First and foremost, the choice of statistical test depends on the type of hypothesis being tested – is it a comparison or is it a relationship? The answer to this question should be obvious from the research problem and its translation into null and alternative hypotheses. Furthermore, if the answer is comparison, then the number and type (i.e., dependent or independent) of samples being compared must also be considered. Third, data can be measured at various scales: nominal, ordinal, interval, and ratio. Likewise, statistical tests are designed to work with data at specific levels of measurement. Traditionally, these tests are organized into two groups: parametric and nonparametric. Simply put, parametric tests can only be applied to interval or ratio data, and nonparametric tests to nominal or ordinal data. A final consideration in test selection concerns the assumptions underlying parametric tests. For example, many such tests assume that the population from which a sample is drawn is normally distributed with a mean μ and standard deviation σ. For samples of size $n \leq 2000$, this assumption can be tested using a Shapiro–Wilk test (it is assumed that the sample is representative of the population from which it is drawn). If this or any other assumption is violated then an appropriate nonparametric alternative should be applied to the data. This is especially true for sample sizes of $n < 30$. In turn, this means that the data must first be downscaled to either nominal or ordinal form according to the requirements of the statistical test chosen. For instance, Spearman rank–order correlation (nonparametric), which uses ordinal data, could be used in place of Pearson product–moment correlation (parametric). **Table 2** lists some common statistical tests and identifies the criteria governing their selection.

The third step in hypothesis testing involves selecting an appropriate significance level (α), which is the probability of rejecting the null hypothesis when it is true (see **Table 3** for a summary of possible outcomes from hypothesis testing). This type of error, known as a type I error, is attributed to sampling error. This means that an observed difference or relationship can occur due to the fact that values of a test statistic vary from one sample to

Table 2 Some common statistical tests along with criteria governing their selection

Test	*Type of hypothesis*[a]	*Number of samples*	*Type of samples*[b]	*Type of data*[c]	*Normality assumption*[d]
Chi-square goodness of fit	C	1	I	N/O	N
Kolmogorov Smirnov goodness of fit	C	1	I	O	N
Wilcoxon matched-pairs signed ranks	C	1/2	D	O	N
Wilcoxon rank sum	C	2	I	O	N
Mann Whitney *U*	C	2	I	O	N
Kruskal Wallis *H*	C	3+	I	O	N
One-sample difference of means	C	1	I	I/R	Y
One-sample difference of proportions	C	1	I	I/R	N
Matched pairs *t*	C	1/2	D	I/R	Y
Two-sample difference of means	C	2	I	I/R	Y
Two-sample difference of proportions	C	2	I	I/R	N
Analysis of variance (ANOVA)	C	3+	I	I/R	Y
Contingency analysis	R	1	I	N/O	N
Spearman rank order correlation	R	1	I	O	N
Pearson product moment correlation	R	1	I	I/R	Y

[a] C = comparison, R = relationship.
[b] I = independent, D = dependent.
[c] N = nominal, O = ordinal, I = interval, R = ratio.
[d] N = no, Y = yes.

another. In other words, like an ordinary sample statistic (e.g., $\bar{x}$, s, or p), a test statistic has a corresponding sampling distribution, which means that extreme values, although highly unlikely, are not impossible to obtain. Thus, to minimize the likelihood of committing a type I error, low values of α, such as 0.05 and 0.01, are usually selected.

Once a significance level is selected, it is used to es tablish a decision rule whereby the null hypothesis is re jected. At this stage, the researcher must choose between one of two methods of hypothesis testing for this and subsequent steps. The first method, often referred to as the classical method, identifies the critical value of a test statistic that corresponds to α for a one tailed (directional) test and $\alpha/2$ for a two tailed (nondirectional) test. Based on this value, the sampling distribution of the test statistic is then divided into regions of rejection and nonrejection. If the statistical test is two tailed, the rejection region is distributed equally across both tails of the sampling dis tribution rather than being placed in one as is the case for a one tailed test. A decision rule, which identifies the re jection region, is then specified. Obviously, this implies that the rule will be written with reference to the critical value of the test statistic. In contrast, when following the p value method of hypothesis testing, the decision rule for rejecting the null hypothesis is written merely in terms of α for both one tailed and two tailed tests. More precisely, the probability value (p value) must be less than α (i.e., if p value $< \alpha$, reject H_0).

The fifth step in hypothesis testing involves calcu lating a test statistic using the formula specified for a particular statistical test. For example, the formula for calculating the test statistic for a one sample difference of means test involving a large sample ($n \geq 30$) is as follows:

$$z = \frac{x \quad \mu}{\sigma_{\bar{x}}} \qquad [3]$$

where z is the test statistic. If the p value method of hypothesis testing is being used, then it is also necessary to convert the test statistic into a probability. This is ac complished as follows. If the test is one tailed, the p value is equal to the area of the sampling distribution of the test statistic beyond the calculated value of the statistic in the same direction as the alternative hypothesis. Alter natively, if the test is two tailed, the p value is equal to twice the area of the sampling distribution beyond the calculated value of the test statistic, thus accounting for both tails. The p value calculated measures the exact probability of making a type I error.

Table 3 Possible outcomes from hypothesis testing

Decision	*Reality*	
	H_0 is true	H_0 is false
Do not reject H_0	Correct decision (prob. 1 α)	Type II error (prob. β)
Reject H_0	Type I error (prob. α)	Correct decision (prob. 1 β)

Based on sample evidence, hypothesis testing cul minates with a decision to either reject or not reject the null hypothesis. This decision is rendered by comparing the test statistic or *p* value calculated in step five to the decision rule established in step four. If either of these values meets the conditions of the decision rule, then the null hypothesis is rejected. Failing this, we simply state that the sample evidence is insufficient to reject the null hypothesis at the specified significance level. By wording our decision as above, we avoid risking a type II error, which is accepting the null hypothesis as true when it is actually false (see **Table 3**).

Concluding Remarks

To reiterate what was said at the beginning of this article, the goal of scientific research in human geography is to further our understanding of geographic processes and geographic phenomena. Central to this task is inferential statistics. For this reason, human geographers, like many other social scientists, usually receive formal training in the application of such statistics as part of their under graduate and, to a lesser extent, graduate studies. The subject matter covered in this article (i.e., sampling dis tributions, estimation, and hypothesis testing) forms the core of such training, which implies that human geog raphers are educated in a form of statistical inference known as classical inference. Sampling distributions, which describe what will happen in the long run should all samples of a given size *n* be drawn from a population, are at the heart of classical inference. The probabilities formed are empirical probabilities, which measure the likelihood of specific events. In turn, this implies that significance level, in both interval estimation and hy pothesis testing, measures the probability of rare events occurring.

Another approach to statistical inference is Bayesian inference. Unlike classical inference, Bayesian inference treats probabilities as personal measures of uncertainty based on available evidence. In other words, such prob abilities are subjective. This subjectivity is captured in the specification of prior probabilities, which reflect a researcher's intuition concerning the value of a popu lation parameter. These probabilities are modified by sample information to derive posterior probabilities for the parameter. Unfortunately, few human geographers receive formal training in Bayesian inference. As it is beyond the scope of this article to delve further into this form of inference, the interested reader is referred to the texts listed under further reading for an extensive treat ment of the subject.

See also: Monte Carlo Simulation; Sampling; Statistics, Descriptive; Statistics, Overview.

Further Reading

Béchard, M. and Marchand, I. (2006). *General Social Survey Cycle 19: Time Use (2005) User's Guide to the Public Use Microdata File*. Ottawa, ON: Statistics Canada.

Burt, J. E. and Barker, G. M. (2003). *Elementary Statistics for Geographers* (2nd edn.). New York: Guilford Press.

Casella, G. and Berger, R. L. (2002). *Statistical Inference* (2nd edn.). Pacific Grove, CA: Thomson Learning.

Diekhoff, G. (1992). *Statistics for the Social and Behavioral Sciences: Univariate, Bivariate, Multivariate*. Dubuque, IA: Wm. C. Brown Publishers.

Fotheringham, A. S., Brunsdon, C. and Charlton, M. (2000). *Quantitative Geography: Perspectives on Spatial Data Analysis*. London: Sage Publications.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2003). *Bayesian Data Analysis* (2nd edn.). Boca Raton, FL: Chapman & Hall/CRC.

Gibbons, J. D. and Chakraborti, S. (2003). *Nonparameteric Statistical Inference* (4th edn.). New York: Marcel Dekker.

Griffith, D. A. and Amrhein, C. G. (1997). *Multivariate Statistical Analysis for Geographers*. Upper Saddle River, NJ: Prentice Hall.

Iversen, G. R. (1984). *Bayesian Statistical Inference*. Beverly Hills, CA: Sage Publications.

Keller, D. K. (2006). *The Tao of Statistics: A Path to Understanding (with No Math)*. Thousand Oaks, CA: Sage Publications.

McClave, J. T. and Sincich, T. (2006). *Statistics* (10th edn.). Upper Saddle River, NJ: Prentice Hall.

McGrew, J. C. and Monroe, C. B. (2000). *An Introduction to Statistical Problem Solving in Geography* (2nd edn.). Boston, MA: McGraw Hill.

Rogerson, P. A. (2006). *Statistical Methods for Geography: A Student's Guide* (2nd edn.). Newbury Park, CA: Sage Publications.

Wong, D. W. S. and Lee, J. (2005). *Statistical Analysis of Geographic Information with ArcView GIS and ArcGIS*. Hoboken, NJ: John Wiley.

Statistics, Overview

G. M. Robinson, Kingston University, Kingston upon Thames, UK

Glossary

Algorithm A procedure or set of instructions for accomplishing a (statistical) task.

Chaos It describes the behavior of some nonlinear dynamic systems, which under certain conditions exhibit dynamics sensitive to initial conditions (popularly termed the butterfly effect). The behavior of chaotic systems appears to be random, even though they are deterministic in that their future dynamics depend on their initial conditions, and there are no random elements involved. This behavior is known as deterministic chaos.

Cluster Analysis The classification of objects into different groups by the partitioning of a dataset into clusters so that the data in each cluster share a common trait (defined usually by a correlation coefficient or distance measure).

Confidence Interval That interval within which a parameter (such as the mean) of a population is calculated to have a stated probability of lying.

Confidence Limits The lower and upper boundaries of a confidence interval.

Contingency Table Tables used to record and analyze the relationship between two or more variables (usually for categorical data).

Correlation The strength and direction of a linear relationship between two random variables.

Dendrogram A tree diagram frequently used to illustrate the arrangement of clusters in a dataset produced by a clustering algorithm.

Dummy Variable A qualitative variable (measured on a binary scale) that can be incorporated into a regression analysis.

Entropy A measure of the uncertainty associated with a random variable.

Factor Analysis A statistical data reduction technique used to explain variability among observed random variables in terms of fewer unobserved random variables called factors.

Fractal A rough or fragmented geometric shape that can be subdivided in parts, each of which approximates to a reduced-size copy of the whole. It is derived from the Latin *fractus* meaning 'broken' or 'fractured'.

Frequency Distribution A list of the values that a variable takes in a sample.

Hypothesis An empirical proposition testable by experience. It is a statement in the form of an unproven theory.

Inferential Statistics Inferences made about a population on the basis of a random sample drawn from it.

Markov Process A sequence of states for which the conditional probability distribution of a state in the future can be deduced using only the current state. The past states carry no information about future states.

Monte Carlo Methods Simulation methods that are nondeterministic, usually by incorporating a random element.

Neighborhood Effect The propensity of geographically proximate locations to be more similar than places geographically remote from one another.

Population The entire aggregate of individuals or items (sometimes called observations) from which samples are drawn.

Probability Distribution A function of the probabilities of a mutually exclusive and exhaustive set of events.

Regression The relation of a dependent or response variable to specified independent variables.

Sample A set of individuals or items (sometimes called observations) selected from a population so that properties or parameters of the population may be estimated.

Spatial Autocorrelation Spatially distributed variables being characterized by correlations between adjacent values.

Statistics The art and science of gathering, analyzing, and making inferences from numerical data.

Variable Any set of numerical values being investigated, such as values of prices, earnings, and distances. The term 'variable' denotes that it possesses a number of different values. To indicate these different values a special notation is used: x_i.

Introduction

The first genuine statistical inquiries were instituted in the United Kingdom (UK) and continental Europe in the seventeenth century as greater emphasis was placed on the exchange value of commodities, and therefore on the quantifiable aspects of people, events, and objects. There was a demand for more precise measurements and ac counting procedures. Subsequent advances in science led to statistical techniques being developed for analyzing experimental data, for creating effective designs in gathering data, and for controlling the quality of manufactured products. Over the course of the next two centuries, into the nineteenth century, these develop ments represented the evolution of statistics as a clearly recognizable subset of mathematics. However, this im pinged little upon the work of geographers prior to the 1940s. Even then, it was only relatively simple de scriptions of geographical data that were employed. Averages, percentages, and graphical depictions of data, often in maps, were geographers' main contact with statistics.

This changed in the 1950s and 1960s as part of geo graphy's quantitative revolution, pioneered in the United States and the UK. A wide range of techniques was ap plied to geographical data as increased emphasis was placed on utilizing quantitative analysis wherever pos sible. In some cases this was part of a deliberate pursuit of the generation of law like statements; in others it was a more speculative empiricism fueled by growing avail ability of large datasets and improvements in computer technology to handle such information. The initial wave of adoption of statistics by geographers emphasized techniques to measure the strength of relationships (correlation and regression), but there was also wide spread use of statistics to test hypotheses, and a gradual move to increased statistical sophistication as exemplified in the adoption of complex multivariate techniques and statistical models.

By the end of the 1960s, few parts of the discipline remained unaffected by the introduction of statistical techniques, though concerns were voiced about in appropriate applications and the implications of the philosophical underpinnings of the new methods. Nevertheless, there had been a series of important de velopments in the use of statistics by geographers:

1. developing the use of statistical hypothesis testing to incorporate the full range of tests found in inferential statistics;
2. extending the study of relationships from the mea surement of the strength of the relationship between two variables (bivariate measures) to multivariate situations;
3. generating specifically spatial statistics, and moving from very simple descriptive applications to more complex analyses of spatial structure;
4. linking the spatial and temporal realms through time series analysis, analyzing the spatial form of temporal change through forecasting and predictive methods, and examining spatial diffusion processes; and
5. using new techniques to both analyze and generate maps (e.g., trend surface analysis).

From 1970 onward, as human geographers developed alternatives to approaches reliant primarily on statistical methods, statistics were used both in more specialist applications and in new ways, notably in examining aspects of public policy. However, major advances have also occurred in the development of spatial statistics and the use of sophisticated statistical models. General trends have included:

1. greater concentration on process rather than form;
2. reappraisal of the nature and appropriateness of statistical inference;
3. the stimulus of applied forms of geographical re search, directing it into new areas, generating a need for different approaches and forms of analysis, in cluding the use of statistics within geographic infor mation systems (GISs) and statistical modeling; and
4. criticism of the use of quantitative methods as part of a widespread critique of their underlying philosophy, and greater reliance on qualitative methods with dif ferent philosophical bases of inquiry.

This overview outlines key aspects of statistical analysis undertaken by geographers. It includes sections on the collection of numerical data through use of sampling procedures, the application of inferential statistics, an alysis of the strength of relationships, multivariate stat istics, use of generalized linear models (GLMs) and categorical data analysis, statistical models, spatial stat istics, research involving both spatial and temporal dimensions, and the most recent advances in the appli cation of statistical analysis to geographical problems. It demonstrates the wide and multifaceted application of statistical techniques by geographers, ranging from sim ple descriptive work to complex modeling and analytical applications.

Sampling and Statistical Description

Sampling

One of the central pillars of statistical analysis is the use of sampling procedures and the use of sample data to represent the characteristics of larger statistical popu lations. Populations can comprise discrete objects (e.g., people sampled using a questionnaire) or they can be continuous (e.g., slope angles measured on a hillside).

There are six basic stages in a sampling process:

1. define the population;
2. define the sampling frame containing all the entities of the population, such as persons and places. Examples include telephone directories, electoral registers, and locations on a map;
3. specify the sampling unit (e.g., city streets, households);
4. determine the sampling method;
5. determine the size of sample; and
6. specify the operational procedures necessary for tak ing the sample.

Samples are used as representations of populations, and inferences about populations are made by analyzing samples (and hence the term 'inferential statistics').

When employing maps as their sampling frame, geographers have used three different ways to ensure that all parts of the map may be selected in the sample (**Figure 1**):

1. point samples (generally based on grid intersections or map coordinates);
2. line samples (transects); and
3. area samples (quadrats) as used in botanical sampling.

There are two basic types of sampling design: random and systematic sampling. In the former, an individual element in the population is as likely to be included in the sample as any other. This is generally achieved using either a random number table or a tombola system. Systematic sampling involves selection from the population at regular intervals, achieving a uniform coverage, whereas random sampling may overlook parts of a population and provide over representation of others. Systematic sampling can sometimes replicate underlying regularities in the population. This therefore generates bias, or the inclusion of unrepresentative fractions of the population, which sampling is seeking to avoid.

Stratification is the commonest variant to the basic designs: separate samples are drawn from each subset of the population, for example, using divisions based on gender, age, or socioeconomic status. Stratified systematic unaligned sampling combines elements of random, systematic, and stratified sampling when selecting point samples from maps.

More complex multistage designs are employed in national surveys, enabling very small samples to represent a national population, for example, in the UK, national political opinion polls sample just over 1000 voters. Pioneered in the US by George Gallup, this approach employs a series of stratifications of the voting population: by region, type of constituency, ward, block, and household.

Descriptive Statistics

Once data have been collected, the first stage of analysis is description, the basis of which is the four statistical 'moments': the arithmetic mean, the standard deviation, skewness, and kurtosis. The mean (or average) is a measure of central tendency (**Table 1**) and is sometimes accompanied by a related measure, the mode (the most frequently occurring value in a set of observations). Several measures of dispersion are used as statistical descriptors. Three in particular are commonly used:

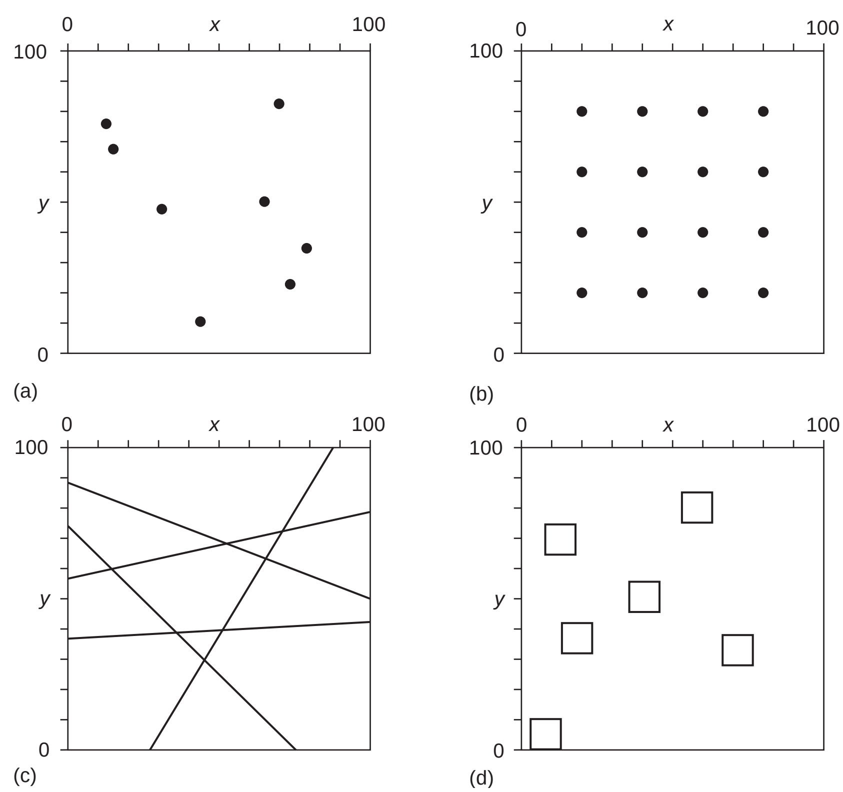

Figure 1 Types of sampling with a spatial sampling frame: (a) point sample: simple random sampling; (b) point sample: systematic sampling; (c) line sample: random coordinates; and (d) quadrant sample: random squares.

1. skewness: the extent to which observations occur on one particular side of the mean;
2. kurtosis: the degree of peakedness of a set of obser vations when plotted as a frequency distribution; and
3. standard deviation (the square root of the variance): measures the average deviation of a set of observations about the mean. When two sets of observations are compared the standard deviation is often replaced by a relative measure, the coefficient of variation, which expresses the standard deviation as a percentage of the mean.

More extensive use is now made of various descriptive graphical methods termed exploratory data analysis (EDA), first developed by statistician John Tukey in the 1970s. EDA emphasizes careful examination of data using simple numerical description in conjunction with im aginative graphical representation.

Table 1 Statistical symbols[a]

For a population		*For a sample*	
Number of observations	$= N$	Number of observations	$= n$
Mean	$= \mu$	Mean	$= x$
Variance	$= \sigma^2$	Variance	$= s^2$
Standard deviation	$= \sigma$	Standard deviation	$= s$

[a] $\sum =$ sum all scores or quantities that follow, e.g., the notation $\sum x$ means that all values of a variable x should be added.
$\sum_{i\ 1}^{n} x_i =$ sum all quantities from x_1 to x_n in the series, $x_1 + x_2 + x_3 + \cdots + x_n$.
$\sum_{i\ 5}^{15} x_i$ means, starting at x_5 sum the sequence to x_{15}.
$\text{Mean}(\bar{x}) = \frac{1}{n}\sum_{i\ 1}^{n} x_i$.
$\text{Standard deviation}(s) = \sqrt{\frac{1}{n}\sum_{i\ 1}^{n}(x_i \quad \bar{x})^2}$.

Sample Size

When sampling, the most desirable sample size may be determined using the central limit theorem. If a random sample is taken from a population and the mean value, $\bar{x}$, is calculated, it is unlikely that this sample mean will correspond exactly with the population mean (μ). If this process is repeated several times, using the same size of sample each time, a series of values of $\bar{x}$ will be gener ated, termed the sampling distribution of $\bar{x}$. Provided this is repeated at least 30 times, when plotted as a graph it gives a bell shaped curve, known as a normal or Gaussian distribution, centered on the population mean (μ) and symmetric about the point $\bar{x} = \mu$. Around one obser vation in three lies more than one standard deviation from the mean, and about one observation in 20 lies more than two standard deviations from the mean (**Figure 2**).

If values of a normally distributed variate, x, are ex pressed in terms of deviations from the mean, the newly created values are termed standardized ones. The units of expression are proportions of the standard deviation and are termed z values, where $z =$ the number of standard devi ations by which x departs from μ. From this, probabilities of particular events can be easily estimated using the distri bution, by employing pre prepared z tables (**Figure 3**).

Applying the central limit theorem, the sample mean can be treated as a normally distributed random variable, with $z = (\bar{x} - \mu)/\sigma_X$. The number of standard errors (the standard deviation of a sampling distribution) can be de noted as z values. The two z values utilized most fre quently are 1.06 and 2.58 as these correspond to the 95% and 99% confidence limits respectively (i.e., 95% and 99% of the area under the normal distribution lie within these values). Confidence limits are the lower and upper bands of z (e.g., ± 1.96). The distance between the limits is the confidence interval. So a desired minimum size of sample

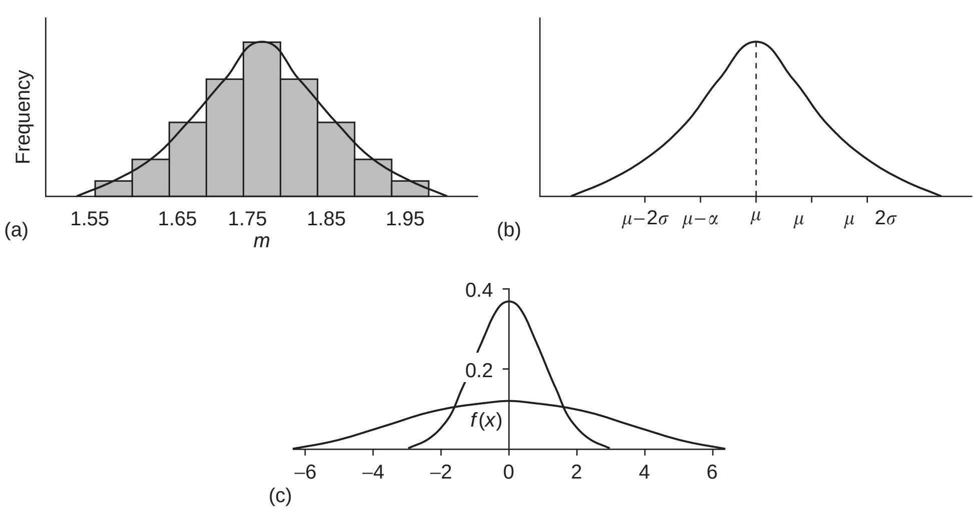

Figure 2 The normal distribution: (a) the distribution of a sample of men's heights; (b) the normal distribution centered on μ; and (c) comparison of two normal distributions.

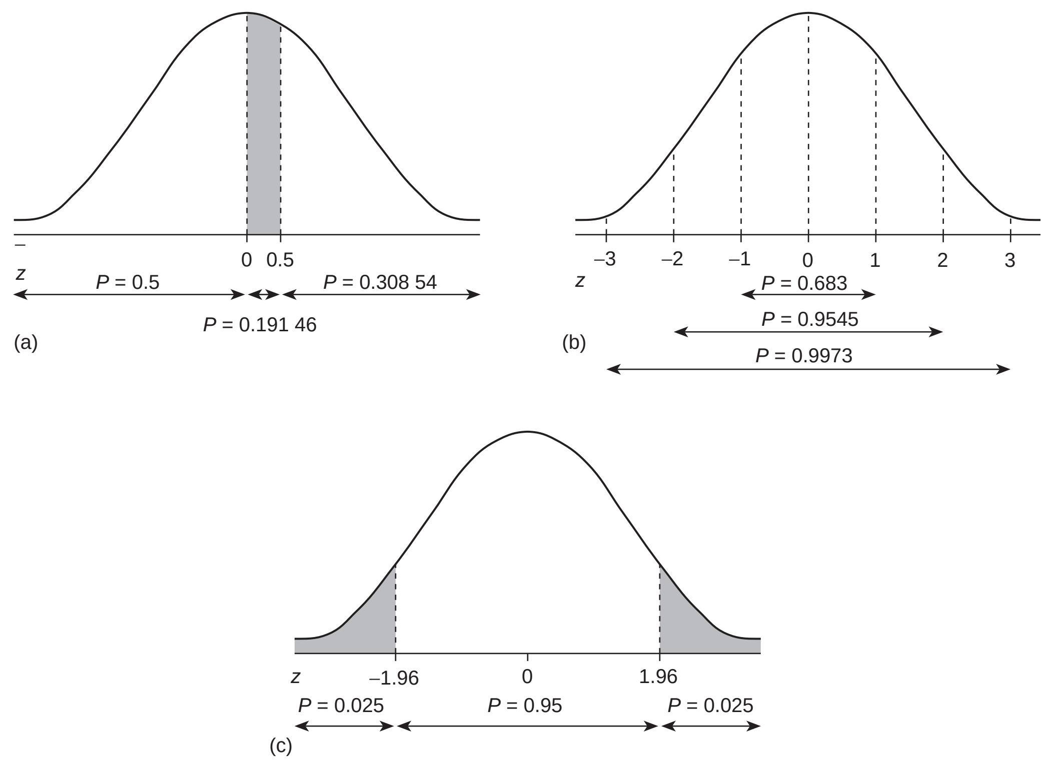

Figure 3 The normal distribution: (a) *P*(*z* 0 to 0.5); (b) probabilities and *z* values; and (c) regions of rejection (*P* 0.05).

may be sought to ensure that μ can be estimated within specified limits. The larger the sample size, the more accurate is the estimate of the parent mean.

As a general rule, for a sampling design to be more appropriate than a simple random sample, for the same sample size, it should yield estimates of population parameters with smaller standard errors. Alternatively, it should yield the same level of accuracy, but using a smaller sample size. Generally, the final sample size utilized in any research project is a compromise between cost, desired accuracy, time, and ensuring sufficient numbers for realistic analysis of subgroups in a population.

Testing Hypotheses

Hypotheses and the Language of Probability

The basis of much statistical analysis is inferential statistics, in which theory is employed to estimate particular characteristics of the population on the basis of sample data. Underpinning inferential statistics is probability theory, and the products of any inferential analysis are statements of a probabilistic nature. This is because an inference is made from a sample to a population. The variability of information in a sample does not allow exact estimates of population parameters to be made. Instead, estimates are generally given within a particular range of values to which a level of probability is attached. Hence, much statistical analysis draws on probability theory, utilizing knowledge of probability distributions to represent or replicate events. Common distributions are the binomial (for two mutually exclusive, independent events), the normal (producing a curved frequency distribution symmetrical about the population mean), and the Poisson (for rare events within long periods of time or space).

Two types of hypotheses may be identified:

1. an unrestricted hypothesis in which the actual hypothesis being tested represents a large number of similar situations (e.g., all birds have wings); and
2. a restricted hypothesis that is only relevant to the context within which it is being tested (known as contingently related conditions) (e.g., incomes in place A are higher than incomes in place B).

Most testing of hypotheses by geographers has been of the latter type, so that the notion of particular events being accounted for as examples of a general law has been avoided. It has been spatial irregularities, as opposed to the world of models and general laws, which have usually been the subject of hypothesis testing by geographers. Although this implies limitations within this approach, key advantages are replicability, the use of probabilistic conclusions, and avoidance of individual prejudices and dogmas through reliance on a repeatable and transferable method.

Statistical testing of hypotheses makes use of the language of probability developed in the 1950s by statisticians Jerzy Neyman and Egon Pearson. So the events under consideration, such as shopping behavior and migration, are conceptualized in terms of this lan guage. The hypothesis employed in statistical testing is termed the null hypothesis (H_0), which generally takes the form 'there is no difference between A and B', for example, productivity per business in group A and productivity per business in group B, that is, H_0: A = B. The object of the hypothesis test then becomes to accept or reject this null hypothesis for a specific level of stat istical significance.

Parametric and Nonparametric Tests

Once H_0 has been formulated various tests can be utilized, the selection of the most appropriate being partly determined by the type of sample data collected. Parametric tests require data on the interval or ratio scales (e.g., 23.7; -127.6), while nonparametric tests operate on ordinal scale (ranked) or nominal scale (categorical) data. Parametric tests possess three additional requirements:

1. the population from which the sample is drawn is normally distributed;
2. the population is heteroscedastic (samples should have equal variances); and
3. observations are independent of one another (it should not be possible to order individual observations in the population and have a better than random chance of predicting values of some when the values of others are known).

Despite these requirements, parametric tests have often been preferred as they are deemed 'more powerful' (where power refers to the ability of a test to evaluate H_0 correctly), and hence non normally distributed data have been transformed mathematically to convert them into normal form for use in parametric tests. However, over time, the qualities of nonparametric tests have been better appreciated because:

1. they can deal with very small samples;
2. they can utilize data on a variety of measurement scales;
3. they can analyze samples from several different populations;
4. they provide probability statements that are exact probabilities; and
5. they are usually easier to apply because they involve less complex calculations.

In a hypothesis test the significance level (α) is usually set at 5% (probability, $P = 0.05$) or 1% ($P = 0.01$), so there is less than five/one chances in 100 that A is different from B.

Table 2 shows the tests most commonly employed within geography. The nonparametric chi squared (χ^2) test is the most utilitarian. It operates on data in the form of frequencies and can be applied to one, two, or more independent samples, so that it can be used to see if one set of data corresponds to a particular type of distri bution, as well as comparing two or more samples. The latter is performed on a contingency table, an example of which is shown in **Table 3**. The two principal parametric tests, Student's t test and Snedecor's F test, are based on particular statistical parameters, the former on the mean and the latter on the variance, and hence it is commonly known as an analysis of variance (ANOVA). The calcu lations for many parametric tests and especially the more complex ANOVAs generally rely on computer programs.

Measuring Associations

Correlation

Initial work by geographers on measuring the strength of a linear relationship between two variables (x and y) re lied upon the Pearson product–moment correlation co efficient (r_{xy})

$$r_{xy} = \frac{\sum(x_i \quad x)(y_i \quad y)}{(n \quad 1)S_x S_y}$$

This has values lying in the range from $+1.0$ to -1.0. A value of $+1.00$ represents a perfect positive relationship

Table 3 A 2 × 2 contingency table for a χ^2 test

Origin	*Under 30s*	*Over 30s*	*Row total*
Urban	5	12	17
Rural	20	13	33
Column total	25	25	50

H_o = there is no difference in the origins of the two age groups.

Table 2 Statistical hypothesis tests commonly employed within geography

Number of samples	*Nonparametric*		*Parametric*
	Nominal data	*Ordinal data*	*Interval data*
1	χ^2	Wilcoxon signed-rank test	*z*-test
	Sign test		*t*-test
	Kolmogorov Smirnov test		
	Runs test		
2	χ^2	Mann Whitney *U*-test	*t*-test
	Kolmogorov Smirnov test	Wilcoxon rank-sum test	
2	χ^2	Kruskal Wallis *H*-test	Snedecor's *F*-test

between the two variables (as y increases so does x to an equal degree). A value of $r_{xy} = 0.0$ indicates the absence of any statistical relationship between x and y, while a value tending toward ± 1.00 indicates high positive or high negative correlation. The relationship's strength can be expressed further by the coefficient of determination ($100r^2_{xy}$), the proportion of the variance in the two variables 'explained' by their correlation. A correlation coefficient of $+0.90$ produces a coefficient of determination of 81%, leaving 19% of the variance between the variables that is not coincident. The value of the correlation coefficient may be reinforced by an assessment of the relationship's statistical significance, which may state that such a value could only occur by chance once in 100 samples ($P = 0.01$). Student's t test or calculation of z values are commonly used for this purpose.

Pearson's product–moment correlation coefficient analyzes interval and ratio scale data, whereas Spearman's rank coefficient (ρ_{xy}) uses ordinal or ranked data:

$$\rho_{xy} = 1 \quad \frac{6\sum D^2}{n(n^2 \quad 1)}$$

where D = differences in ranks of corresponding values of x and y; and n = numbers of pairs of x and y values.

Bivariate relationships are often indirect or casual, with correlations dependent on intermediate factors. Therefore, much geographical research examines multivariate situations rather than bivariate ones.

Four key problems can affect the results of correlation analysis of geographical data:

1. Modifiable areal units: for areal units (e.g., enumeration districts, counties), correlation may measure not only the variation of the variables but also the arbitrary spatial units utilized. The impact of this arbitrary 'spatial mesh' is difficult to predict and measure.
2. Ecological correlation: different results may be obtained if information is analyzed at different spatial scales. Correlations between individual observations are different to correlations for those individuals aggregated into different spatial units. This is termed ecological correlation and is related to the problem of modifiable areal units. It demonstrates that careful nuances of interpretation are required when applying correlation analysis to data for areal units.
3. Closed number sets: when using proportions (which sum to 1) or percentages (which sum to 100), correlation structures will be biased by recording the same thing more than once. This is termed closure or the ability to predetermine a given value for a variable once other values are known (e.g., if there are three land use categories, one representing 50% of an area and another 30%, then the third category must cover 20% of the area). Wherever possible, closed ratios such as this must not be correlated or recognition should be made of possible outcomes (if these can be determined).
4. Spatial autocorrelation: it is common for many variables to be characterized by correlations between adjacent values in space or time. This is termed auto or serial correlation. Its significance was not appreciated by geographers until work by Andrew Cliff and John Ord in the late 1960s. Essentially the presence of spatial autocorrelation violates the assumption of independence of observations. The implication is that applying correlation analysis to situations in which samples possess locational characteristics can conflict with essential assumptions of the statistical techniques. This has often been ignored in many geographical applications, though some geographers have designed spatial models incorporating autocorrelation and have developed tests for and measures of spatial autocorrelation.

Regression

In two particular circumstances geographers have extended the investigation of the strength of bivariate relationships by conducting regression analysis.

1. If one variable is deemed to be dependent on another, for example, if crop yield is considered to be dependent on labor input, in this case, the dependent variable is crop yield (y) and labor input is the independent variable (x). The correlation (r_{xy}) can be calculated, and a scattergraph of y versus x can be produced. The line of best fit through the scatter of points in the scattergraph is the linear regression line whose equation ($y = a + bx$) can be utilized in various ways (a = the intercept on the y axis and b is the slope of the line).
2. When there is a desire to predict the value of the dependent variable for given values of the independent variable, the coefficient of determination (C), (where $C = 100r^2_{xy}\%$) is the level of explanation of the dependent variable by the independent variable. $C = 100\%$ if all observations lie directly on the regression line. This is rarely the case in real life situations and so several individual observations may lie at some distance from the best fit regression line. The distance along the y axis from an individual point to the line is termed the residual (or the difference between the actual and predicted values of y). As the size of the residuals increases so the utility of the regression line decreases in terms of its predictive value (**Figure 4**).

Each residual can be expressed in a standardized form as standard error units from the regression line, where

$$\text{Standardized residual} = \frac{x_i \quad x}{\sigma_e}$$

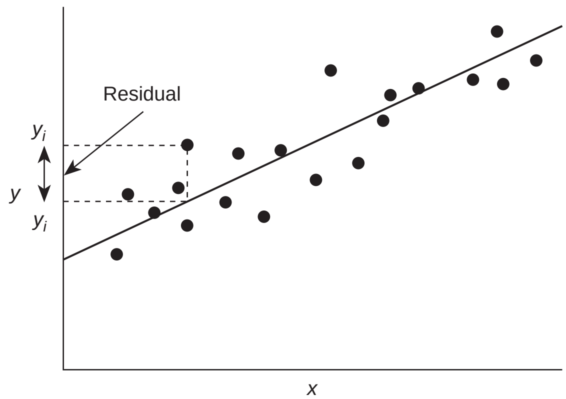

Figure 4 Regression line and scattergraph with residuals.

Mapping of residuals can be used to help form hy potheses, to examine regional boundaries or identify study areas for intensive fieldwork.

Multiple Regression

Simple linear regression [$y = f(x)$] may be extended to more complex multiple regression [$y = f(x_1, x_2, x_3, \ldots, x_n)$]. In multiple regression, the equation takes the form:

$$y = a + b_1 x_1 + b_2 x_2 \pm e$$

where a = the intercept on the y axis, b_1 and b_2 = partial regression coefficients, x_1 and x_2 are the independent variables, and e = an error term. The partial regression coefficients can only be compared with one another if they are transformed into beta weights (β_i), where $\beta_i = b_i(s_{xi}/s_{yi})$, b_i = slope coefficient of independent vari able x_i, and s_{xi} and s_{yi} = standard deviation of the dependent and independent variables respectively.

β_i weights show how much change in the dependent variable is produced by a standardized change in one of the independent variables, with the influence of the other independent variables controlled. This enables assess ment of the effects of the individual independent variables in the regression equation: the higher the β_i weight, the greater the rate at which the dependent variable increases with an increase in the particular in dependent variable.

In multiple regression the multiple correlation co efficient ($ry.x_1x_2x_3{\ldots}x_n$) measures the correlation between y and all the independent variables. If there is correlation between the independent variables then collinearity exists. This reduces the extent of the explanation of the dependent variable and indicates redundancy in the independent variables. This may be eliminated using stepwise entry of variables into the equation by adding variables to the equation in their order of importance in reducing the variance of y, starting with the most im portant first.

Multivariate Analysis

Analysis of datasets based on large numbers of variables was first performed by geographers and sociologists in the 1950s as a means of testing models of urban social dif ferentiation developed by human ecologists at the Uni versity of Chicago. Initially, this work was termed social area analysis. It used population census data to classify census tracts that could be mapped and compared with the urban social structure in the ecologists' models.

Factor analysis was the statistical technique employed for social area analysis, so its application was termed factorial ecology. During the 1960s factorial ecologies were performed for North American cities, and then for cities throughout the world and also for other multi variate datasets. Factor analysis reduces data for a num ber of intercorrelated variables (m) to a smaller set of factors (k) that represent amalgams of the variables. This involves a four stage sequence:

1. the geographical matrix (places vs. variables);
2. the correlation matrix (variables vs. variables);
3. factor loadings (variables vs. factors); and
4. factor scores (places vs. factors).

From stage 2, there are two basic applications: principal components analysis (PCA) and principal axis factor analysis (PAFA).

Principal Components Analysis

PCA produces as many components as there are vari ables. The first component always accounts for the greatest variance among the variables and the mth com ponent the mth highest. The variance among the variables accounted for by each component is equal to the sum of the squares of the loadings, known as the eigenvalue or latent root. It is usually expressed as a percentage of the maximum possible variance and is a measure of the im portance of each component.

The first four or five components extracted from a set of 40 or 50 intercorrelated variables should account for at least two thirds of the overall variance to be regarded as an effective replacement. These few components repre sent the basic differentiating characters of the dataset. To determine the probability of a particular loading having occurred by chance, a significance test can be employed. The first few components are termed common com ponents as they have high loadings on many of the variables.

The components can be mapped and represent a measure of the importance of each component in each place, expressed in standard deviation units from the average score. The map of component scores replaces a series of maps of the individual variables. The component score (c_{ab}) for place a on component b is:

$$c_{ab} = \sum_{i=1}^{m} (S_{ai})(l_{ib})$$

where m = the number of variables; s = the standard score for place a of variable i on component b, with $s_{ai} = (i - \hat{\imath})/\sigma_i$ and l_{ib} = the component loading of variable i on component b.

Principal Axis Factor Analysis

PAFA was developed by psychologists as a hypothesis testing device; m variables are replaced by a smaller number (k) of underlying factors. PAFA splits the variance into three parts:

1. common variance (or communality): the degree to which one variable varies in conjunction with other variables;
2. unique (or specific) variance: the individuality of the distribution of each variable. The ratio of common to unique variance becomes larger as successive factors are extracted; and
3. error variance (or unreliability): any imperfections (such as sampling bias) in the original data.

Variables with high communality are the ones of interest, and so more importance is attached to these variables when extracting factors. The different weightings applied to each variable modify the factor structure, in contrast to PCA, where each variable is treated as if it had the same communality.

PAFA can employ rotation of the factor axes until the factor loadings created are either near ± 1 or near 0. This aids interpretation and labeling. Rotation can be performed with factors at right angles (orthogonal rotation) or obliquely (oblique rotation).

Partly because of the similar nature of the variables employed, the many factorial ecologies of North American cities typically produced three factors accounting for the majority of variance: family status (varying concentrically with distance from the central business district), economic status (varying sectorally), and ethnic status (clustered, reflecting concentrations of different ethnic groups; **Figure 5**). More variation was found outside North America, but the value of such work was strongly questioned in geography's 'relevance debate' in the 1970s. The end result of maps of factor/component scores was seen as misguided at a time when social inequalities within cities were being publicized daily in the mass media.

Classification

Factor scores can be used to delimit particular regions as a form of classification. Classification can be based either

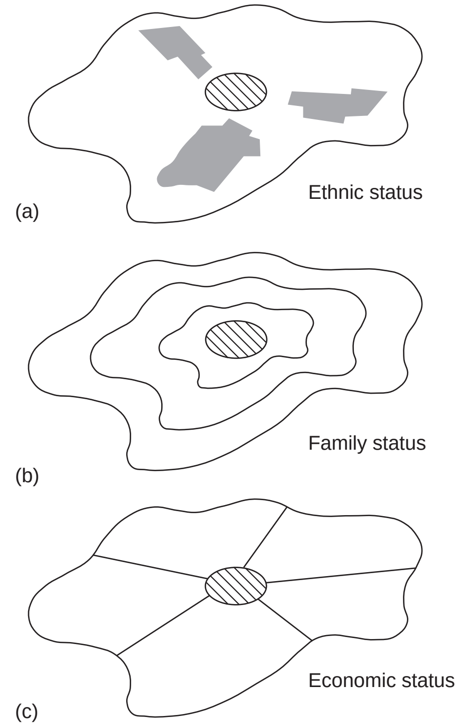

Figure 5 Schematic portrayal of the spatial distribution of the general components from PCAs of American cities. Based on Murdie, R. A. (1976). Spatial form in the residenmtial mosaivc. In Herbert, D. T. & Johnston, R. J. (eds.) *Social Areas in Cities: Volume 1. Spatial Processes and Form*. London: John Wiley.

on a correlation (r) or distance (d) matrix depicting relationships between observations. The distance coefficient (d_{ij}) is the squared distance between observations (places) i and j:

$$d_{ij} = \frac{\sqrt{\sum_{k=1}^{m} (x_{ik} - x_{jk})^2}}{m}$$

where x_{ik} = the kth variable measured on object (or place) i; x_{jk} = the kth variable measured on object (or place) j; m = the number of variables; and d_{ij} = the distance between object i and object j. A small value of d_{ij} indicates that two objects are similar or close together. It is not constrained within the range $+1$ to -1.

The classification process itself is usually based on the representation of the correlation/distance matrix in a dendrogram, so named because of its branching structure. The initial step in the dendrogram is to select the mutually highly correlated pairs of observations, which are then grouped with unattached observations until the grouping procedure has combined all the observations into one single group. Attaching a single observation to a group can be done in two basic ways:

1. single linkage clustering using the highest correlation between the object and any other object in the cluster and
2. weighted pair group clustering on the average of the correlation between the object and each individual in the cluster.

In a classification exercise, there is no *a priori* knowledge of existing clusters. In contrast, discriminant analysis usually begins with already recognized clusters or groups. It allocates individuals to these groups by es tablishing a linear combination of variables measured for two clearly defined groups that produces the maximum difference between the groups. If this linear combination (a discriminant function) produces a significant differ ence, it can then be used to allocate samples of unknown origin to the known groups.

A linear discriminant function transforms an original set of observations into a single discriminant score. The distance between the mean of two sets of observations along the discriminant function is the Mahalanobis distance, a measure of variance of the two groups. Dis criminant functions are used to test hypotheses, to evaluate a classification, and to assign new members to pre recognized groups.

Generalized Linear Models and Categorical Data Analysis

Generalized Linear Models

In the last two decades the range of statistical techniques applied to geographical data has been extended to in clude specialist tools to analyze categorical data. Stand ard regression techniques are often inappropriate for such data. Instead, it may be better to use a family of GLMs for estimating linear relationships with nominal data. GLMs are based on a suite of exponential prob ability distributions (**Table 4**), which can be used in the context of a simple equation whose form is similar to that for simple linear regression

$$y_i = \mu_i + \varepsilon_i$$

where y_i = an independent random variable; μ_i = the part of y_i whose value is predictable; and ε_i = the part of y_i whose value is a random error. Any model of the variability of y_i requires information on the character istics of other variables thought to influence the variation in y_i, which is expressed as a linear predictor (η_i). It also requires information on the linkage which relates μ_i to these other variables. μ_i is related to the linear predictor η_i by a link function (g). There are various link functions and hence various models, for example, logarithmic (log μ), logit ($\log(\mu/n-\mu)$), probit ($\Phi^{-1}(\mu/n)$), square root ($\sqrt{\mu}$), and reciprocal ($1/\mu$).

Investigating Contingency Tables

These various models are applied to contingency tables created by the cross classification of categorical variables. Such variables may be dichotomous (e.g., yes or no), unordered polychotomous (more than two categories without a definite order, e.g., manufacturing, service, primary), and ordered polychotomous (e.g., young, mid dle aged, elderly). These models enable an investigation of contingency tables to examine relationships besides independence, while 'traditional' statistical tests, such as χ^2, only focus on independence. Alternative models can be fitted to data in a contingency table and then evalu ated, utilizing every possible form of interdependence in the data.

One of the most popular forms of GLMs is log linear analysis, which attempts to obtain a model that is linear in the logarithms of the expected frequen cies of a contingency table and which describes or fits the associations and interactions that exist in the table. As with regression analysis, log linear analysis can involve the examination of residuals to see which individual values in the contingency table have the greatest re siduals. Log linear analysis can be extended from small 2×2 contingency tables to more complex 'multi way' designs that require more complicated modeling.

The five principal models that have been used to analyze data are summarized in **Table 5**. The top row, with cells a, b, and c, represents the traditional models of simple linear regression, dummy variable linear regression, and ANOVA. The bottom row refers to categorical models and can be addressed by logistic, logit, and probit models.

Table 4 The suite of exponential probability distributions

Distribution	*Use*
Normal	Regression analysis
Gamma, beta	Stochastic process modeling
Binomial, multinomial, Weibull	Analysis involving discrete choices and individual behavior
Poisson, binomial, multinomial	Parametric analysis of contingency tables

Table 5 Classification of statistical problems

Response variables (dependent)	*Explanatory variables (independent)*			
	Continuous	*Mixed*	*Categorical*	*None*
Continuous	a	b	c	
Categorical	d	e	f	g

Based on Wrigley, N. (1981). Categorical data analysis. In Wrigley, N. & Bennett, R. J. (eds.) *Quantitative Geography: A British View*, pp. 111 122. London: Routledghe and Kegan Paul.

Statistical Models

During geography's quantitative revolution, progress was made in the use of statistical and mathematical models to represent different facets of economic and social geography. These were developed from simple schemes for the internal structure of cities and the arrangement of central places. They incorporated theories, laws, hypotheses, structured ideas, roles, relations, equations, syntheses, maps, graphs, and computer hardware and software. From this catholic interpretation more sophisticated mathematical models have emerged, with both deterministic (normative) and probabilistic forms. Emphasis has been placed upon spatial interaction problems that can be addressed using models.

Gravity Models

One of the foci of statistical modeling has been the geometry of cities, especially the forecasting of flows, which is essential in making land use and transport investment decisions. Predictions can be used to reduce the costs of spatial interaction while maintaining diversity of choice. Spatial interaction models are used to predict spatial choices reflected in flows of people or goods between origins and destinations, expressing trade offs between the accessibility of alternative destinations and the attractiveness of these opportunities.

The gravity model is the basis of spatial interaction models. It provides a direct analogy with Sir Isaac Newton's law of gravitation

$$F = \frac{G(m_1 m_2)}{R^2}$$

where F = magnitude of the attractive force, G = a gravitational constant (the size of which depends on the system of units used, and which is a universal constant), m_1 and m_2 = masses of the two bodies, and R = the distance between them. In the late nineteenth century, E. G. Ravenstein's laws of migration were based on this Newtonian idea.

The basic gravity model has been used as a predictive tool, starting in 1931 with W. J. Reilly's law of retail gravitation, predicting the flow of retail trade. Important respecifications were made in the 1960s, constraining production or attraction or both. Of the many variants of gravity models, the most popular is the Garin–Lowry model, based originally on allocating population and services by linking the gravity model with submodels relating population and services via an economic base mechanism.

Alternatives to Gravity Models

One alternative to gravity models is intervening opportunity models, first developed in 1940 by S. A. Stouffer. These measure basic characteristics of the destination in terms of the number of opportunities available there. Instead of distance *per se* influencing the chance of a trip being made, the crucial factor is the number of intervening opportunities: those opportunities closer to the trip's origin than any particular destination being considered by a traveler. This model can be adjusted to allow for overlapping catchment areas, depending on how the rate of decline of demand over distance is calculated.

Another alternative is entropy maximizing models, which analyze the most likely state of a system by drawing upon concepts of entropy in thermodynamics and information theory. They were developed in the late 1960s by Alan Wilson who treated movements of people and objects in the manner in which particles in gases were analyzed in statistical mechanics. He modeled the journey to work in cities, dividing the city into zones of residence and employment, with interaction between zones measured through analysis of trip origins and destinations. This shares with gravity models the intention of predicting a trip matrix.

Entropy maximizing models have been used for forecasting different components of complex flows in urban and regional contexts, and they have formed the basis for estimating the structures of nonlinear models. These and other interaction and allocation models have been widely used by planners, especially through the addition of algorithms that relax some of the more simplistic aspects of the basic models. They have been especially useful in helping retailers decide on the location of new stores and in predicting traffic flows. However, they have been criticized for failing to make adequate representation of the behavioral processes that lead to individuals selecting a particular journey. The models are aggregate ones, averaging out user behavior. Nevertheless, by adding underlying elements of economic and social structure to the models they have become much more realistic and more widely applied, drawing upon the increased power of modern computers to deliver rapid solutions to complex locational problems.

Spatial Statistics

Geographers have used statistics and geometry to analyze spatial form, focusing on the locational pattern of objects and treating them as points, lines, and areas. The content of this analysis has ranged from use of simple descriptive indices to highly complex measures of spatial structure.

Descriptive Measures

The Lorenz curve

This compares the cumulative percentages of two distributions on a graph, with one distribution plotted as the

y axis and the other on the x axis. The Lorenz curve is the resulting curve: the bigger its variation from the diagonal, the bigger the difference between the cumulative proportions.

The index of dissimilarity, I_d

This is the percentage of a population group that would have to shift its residence in order to reproduce a spatial distribution identical with that of the group with which it is being compared. I_d is the maximum difference in the cumulative percentages of the two distributions shown in the Lorenz curve

$$I_d = \frac{1}{2}\sum(x_i \quad y_i)$$

where x_i and y_i are the individual percentages of each variable. Typically, the variables refer to occupational groups or ethnic status, for which I_d becomes an index of segregation (I_s), with a correction factor applied to eliminate the presence of y_i (proportion of ethnic group y) occurring in the total population, T

$$I_s = \frac{I_d}{1 \quad \frac{\sum y_i}{T}}$$

Indices of 80% and above for African Americans in city blocks in the US cities led to the label 'ghetto' being applied. However, the size of the administrative unit on which I_s is calculated can significantly affect its value, so it can only be used as a general indicator of the degree of segregation.

Location quotients

These measure the extent to which a particular distribution differs from the norm, typically comparing employment in a particular industry with the national average for all industries

$$\mathrm{LQ} = \frac{g_x}{k}$$

where g_x = % of activity x in an area, and k = the national %.

1. LQ >1 = a high degree of concentration of a particular activity in an area;
2. LQ <1 = low concentrations; and
3. LQ = 1 = the distribution is the same as the national average.

The Gini coefficient (I_G)

This compares the distribution of an attribute with a hypothetical equal distribution. It is related to the Lorenz curve and is an alternative measure of dissimilarity

$$I_G = \frac{1}{2}\sum_{i \quad 1}^{n}|P_i \quad I_i|$$

where P_i = % values of one attribute, for example, population numbers, and I_i = % values of another attribute, for example, income levels. I_G is constrained between 0 (exact correspondence) and 100 (maximum inequality).

The nearest-neighbor statistic (R)

This measures the dispersion of points in relation to properties of the distribution of distances between each point and its nearest neighboring point. Developed initially to analyze patterns in plant ecology, in human geography the points have often been locations of settlements. Using the Poisson probability distribution (which applies to random events), the average expected distances (r_e) between points $= 0.5\sqrt{(A/n)}$, where A = area of study region, and n = number of points.

$$R = r_o/r_e$$

where $\bar{r}_o$ = observed distances.

Values of R range from 0 (all points occupy the same location) to 2.149 (points are equidistant from six other points in a hexagonal pattern). A value of 1 suggests a random pattern. However, R is highly dependent on how A is defined, and the same value of R can be generated from different point patterns.

Spatial Structure

Detailed analysis of spatial structure has been performed by geographers since Cliff and Ord's pioneering work in the 1970s. This examined spatial autocorrelation or the association between a particular quality in an area and its presence in a neighboring area (also known as the neighborhood effect). This is a form of self correlation and, if present, violates the basic condition of independence among observations required by parametric statistical tests. Positive values of autocorrelation indicate a greater presence than expected of the quality concerned, and negative values indicate a lower presence than expected.

Spatial autocorrelation is created by linkages and flows between places that create spatial dependencies. Such linkages are common but difficult to measure, especially for irregular spatial grids (e.g., maps of administrative areas). Various measures have been formulated for regular grids, initially based on nominal data (presence or absence), though they are generally complex and have proved difficult to incorporate in standard statistical computing packages. Two frequently used ones are Moran's I and Geary's C statistics, which can be applied to data for both regular and irregular grids. These measures are not normally incorporated in standard statistical computing packages, but they are commonly available in GIS programs and in stand alone packages for spatial analysis (e.g., GEODA).

Related work by Stan Openshaw applied ideas of spatial autocorrelation to testing for clusters of disease, using computer programs to test repeatedly for clus tering at a range of spatial scales. This type of analysis has been significant in various epidemiological investigations, though usually employing more advanced methods, such as Bayes estimates, based on the joint probability of two events written as the product of one event and the con ditional probability of the second event, based on the first event.

Recent Advances

Recent advances in point pattern analysis have used three sets of properties of spatial stochastic processes:

1. First and second order properties: describing the way in which expected values of the process vary across space or are correlated at different spatial locations;
2. Kernel estimation: estimates of disease intensity at each grid point are made by counting events per unit area with a kernel or three dimensional (3 D) func tion centered on a number of locations arrayed in a fine grid superimposed on the study region. This can help to identify clusters of disease occurrence; and
3. The *K* function: this describes the extent to which there is a spatial dependence in a pattern of events, which therefore can be used to detect clustering and different types of clustering.

$\lambda K(d) = E($number of events $<$ distance d of an arbitrary event), where, $E(\ldots.) =$ expectation (as a probability); and $\lambda =$ intensity or mean number of events per unit area.

Further advances in the development of spatial stat istics are based on the assumption that variation in spatial data arises from three components that represent a probability model:

1. a deterministic structured element, usually given a functional representation that can be equated with the mean of a probability distribution;
2. a stochastic structured element, usually represented in terms of the second order properties of that prob ability distribution; and
3. a local random element or noise.

It is usually assumed that

$$E[Z(x)] = m \quad \text{for all } x$$

where $E =$ mathematical expectation, $Z(x) =$ the random variable at location $x = (x_1, x_2)$, and $m =$ a constant.

Space and Time

Spatial Diffusion Processes

Geographers have frequently applied statistical techniques to analyze changes in spatial patterns through time. Typical concerns have been the spread of diseases and the diffusion of innovations. Models of spatial diffusion processes are termed simulation models or Monte Carlo models as they utilize random numbers. There are three general types of diffusion processes:

1. Expansion diffusion: knowledge of an innovation is communicated throughout a population so that the number of people with knowledge of an innovation increases steadily over time. Such diffusion can apply to transmission of diseases through direct contact between people/animals. This process is sometimes termed contagious diffusion. Epidemic models divide the population into susceptibles, infectives, and removals.
2. Hierarchical diffusion: some innovations can leap over intervening people and places, following a path that concentrates, for example, initially on only large urban centers or significant individuals before filtering down to smaller places or beyond the immediate circle of community leaders.
3. Relocation diffusion: holders of information may migrate taking their knowledge with them, thereby transplanting new ideas into a different environment. Common examples of this relocation diffusion are large scale migrations of different cultural groups (**Figure 6**).

In reality, many diffusion processes represent a combin ation of expansion, hierarchical and relocation diffusion, and especially of the first two.

Simulating Spatial Diffusion

The first substantial attempt to model spatial diffusion processes was the work of the Swedish geographer Tor sten Hagerstrand, who likened the expansion diffusion process to a wave like pattern which loses its strength as it moves away from its source of origin. Working on innovations among farming communities in southern Sweden, Hagerstrand simulated the expansion diffusion process in which a distance decay function was operating. The likelihood of and frequency of contact among the farmers were determined by the distance between the individuals concerned (a neighborhood effect or con tagious growth). Therefore probabilities of adoption were highest in the vicinity of an earlier adoption and de creased with distance away from an adoption. The wave like pattern can be represented as a common pattern of adoption of an innovation in the form of a normal curve

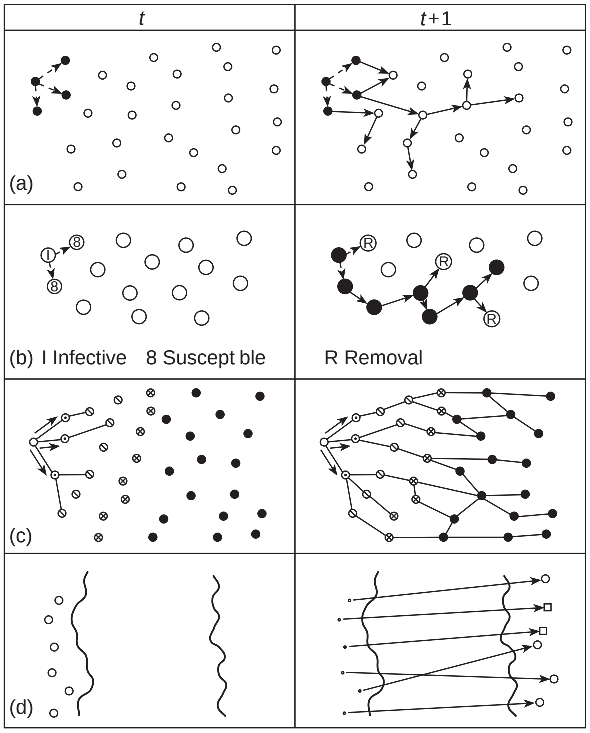

Figure 6 Diffusion patterns: (a) expansion diffusion; (b) the general epidemic model; (c) hierarchical diffusion; and (d) relocation diffusion.

or its cumulative form, the logistic curve, expressed as

$$P_t = (1 + e^{a-bt})^{-1}$$

where P_t = the proportion of adopters at time t, a = the intercept, and b = the slope coefficient. Similarly,

$$y_t = k(1 + e^{a-bt})^{-1}$$

where y_t = number of adopters, and k = the maximum possible number of adopters, that is, the saturation level.

$$r_t = \quad bp_t(1 \quad p_t)$$

where r_t = the rate of change in p_t (proportion of adopters) over time, $(1-p_t)$ = proportion of the popu lation who have not adopted at time t, $p_t(1-p_t)$ = the probability that a random meeting between two indi viduals is between an adopter and a potential adopter, and b = the rate at which meetings take place (the rate of mixing, which can be modified to reflect the neighbor hood effect: inhomogeneous mixing).

To simulate the pattern of spatial diffusion, a mean information field (MIF) is superimposed over the study area. This is a grid of probabilities representing the probability of someone in a particular cell receiving in formation of the innovation in question (i.e., direct con tact). These probabilities have a distance decay function from the point of origin of the innovation. Derivation of these probabilities is problematic: Hagerstrand initially based them on migration patterns and telephone traffic. These gave a sharp distance decay effect. However, less steep effects prevail in modern society with greater ease of mass communication and travel. The MIF is oper ationalized using a random number table so that change through time and space may be predicted, and predic tions compared with reality or used for forecasting. The highly empirical nature of these simulations is one of their limitations, but they have been widely used to examine characteristics of the early adopters of innova tions, the influence of different forms of information, and barriers to adoption.

Large scale simulations of the spread of disease have been developed, notably to analyze human immuno deficiency virus/acquired immunodeficiency syndrome (HIV AIDS), influenza, and measles. These are known as epidemic models, which use more sophisticated 'rules' for the MIF and can incorporate other statistical techniques, such as regression analysis, to allow for the effects of influential variables. For HIV AIDS this might include information on drug use, sexual activity, availability of hospitals, and characteristics of people's travel movements. More advanced methods use equations of stochastic pro cesses that formalize the various transitions from the three main stages experienced by an individual during an epi demic's progress: from susceptible to infective to removal. One variant of this is the expansion method which in vestigates the variation in the temporal growth of a disease in response to changes in population density.

A variant on the basic Monte Carlo model is the in jection of a more complex set of processes that can include adoption, rejection, and readoption of an in novation. This sequence of events can be incorporated in a Markov chain: if the probability of being in a state (or states) at time t is wholly dependent upon the state(s) at some preceding time(s), it is said to be a Markov process. In a Markov chain the MIF is replaced by probability values comprising a transition matrix (P_{ij}), where P_{ij} represents the probability of a change in state from a preexisting state i to a future state j. The transition matrix may be used to predict future outcomes by multiplying the matrix by a vector. The chain may be used to portray both spatial and aspatial sates, and a stochastic element may be injected to render outcomes probabilistic as op posed to deterministic (**Figure 7**).

New Frontiers

Recent developments in the use of statistical techniques by geographers have used advances in computing to analyze complex systems, notably those involving catastrophes or discontinuities and chaotic or irregular systems. The theory behind this work has only been developed from the

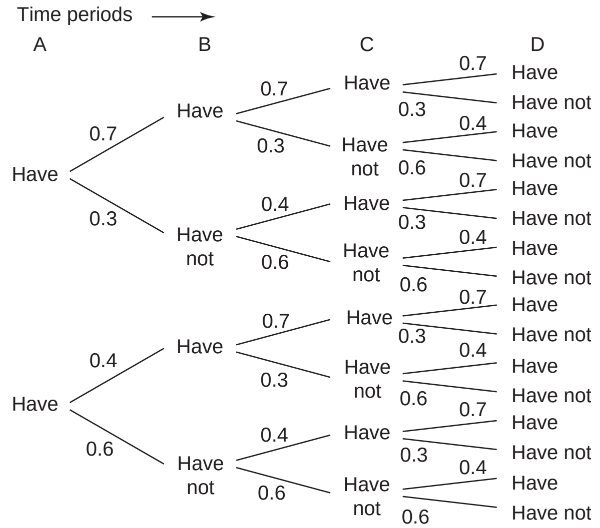

Figure 7 A Markov chain (numbers refer to probabilities of change from one state to another).

mid 1970s, based on pioneering research by Rene Thom, Alan Wilson, and Benoit Mandelbrot.

Catastrophe Theory

Research on discontinuities has been termed bifurcation or catastrophe theory, and focuses on critical values of system parameters at which some unusual behavior of the system can occur. The simple catastrophe is the cusp catastrophe which involves one variable (x) and one (u) or two (u_1 and u_2) parameters or control variables. The maxima, minima, and points of inflexion of the potential function (F) are known as stationary points. Where these points coalesce a singularity occurs, or a point at which the value of F can jump (i.e., the catastrophe). Maxi mizing or minimizing F determines the state of the sys tem. The equilibrium of such a system can be determined using quadratic equations, and the system can be por trayed graphically, providing a framework within which control variables may be investigated. Potentially, sudden changes in otherwise gradually evolving systems can be examined mathematically using these theories.

Chaos Theory

Chaos theory has been developed from the recognition that apparently simple physical systems which obey de terministic laws may nevertheless behave unpredictably. Nonlinear systems can converge to an equilibrium (steady state) or there can be a stable oscillation (periodic behavior) or there can be chaotic change. These three regimes are known as the logistic map and can be rep resented through simple mathematical equations so that a deterministic equation can portray nonrandom chaos. However, to model chaotic systems, Mandelbrot utilized a new form of geometry, fractal geometry. Fractals are irregular but with the same degree of irregularity on all scales. This can be seen in nature in the tendency toward repeated self similarity in ferns, cauliflowers, and broc coli (the same basic structure created in ever decreasing scale with a detail twice as fine as that in the preceding scale). In addition to linear self similarity, fractals can also be random or nonlinear. Using advanced computer graphics random fractals have been applied to model coastlines, mountains, and cloud formations. Equations used to represent turbulence in liquids and weather patterns are nonlinear and display deterministic chaos. They have been used to model urban morphology and unpredictable aspects of behavior within economic and social systems. Complex measures have been developed to assess changes in state of such systems and computer packages are now available to calculate some of these measures.

Statistical Analysis and GIS

The development of computer packages has transformed the ability of geographers to conduct statistical analyses, with standard packages, such as Excel, SPSS, SAS, Minitab, and GLIM, becoming standard elements in the geographers' 'toolbox'. These have also been incorpo rated into GIS: integrated computer systems for the input, storage, analysis, and output of spatially referenced data. The use of GIS has generated wider applications of statistical methods and models, promoting a new techno scientific approach to analyzing spatial data. The in corporation of complex statistical operations, such as measurements of spatial autocorrelation, into GIS pack ages has greatly enhanced researchers' ability to apply sophisticated statistical techniques to problems. As a consequence, statistical analysis of geographical data has become commonplace within many businesses and in the public sector. Yet, there are ongoing concerns about how these GIS based applications of spatial statistical analysis may be at variance with current trends within human geography emphasizing qualitative methods and espous ing theory driven investigation as opposed to the more data led enquiries using GIS. Nevertheless, the flexibility and scope for statistical analysis by geographers using GIS and computer packages has transformed the use of statistics within geography while constantly developing new technical advances, especially in the areas of spatial statistics and statistical modeling.

The evolution of human geography over the last half century has seen the role and status of statistical methods in the discipline alter dramatically. From the dominance of such methods during the quantitative revolution of the 1950s and 1960s, subsequent critiques of statistical tech niques and the adoption of new philosophies, such as postmodernism and structuralism, have produced a much more diverse discipline in which statistical and other quantitative techniques are often subordinate to the use of qualitative approaches. These have grown in promi nence as interest in the cultural, social, and political di mensions of geographical problems has risen. Important questions have been raised concerning the applicability of classical statistical hypothesis testing to geographical data, the representation of people by numbers, and the sim plification of complex processes in statistical models. Such concerns and the changing nature of human geography subject matter have produced several different types of approach to geographical research, some rejecting the use of statistics entirely, others embracing new technological advances, and employing complex methods of statistical analysis. There is evidence, though of a growing body of work that incorporates both quantitative and qualitative methods, as championed first by human geographers working from a feminist perspective. In this area of human geography, statistical methods have been used to docu ment the spatial and social dimensions of women's lives at a variety of spatial scales. Statistical methods can provide the power of generalizations that may then be coupled with more qualitatively based studies of the particular. This has helped advance work employing multilevel modeling to analyze relations at different scales and the display and overlay capabilities of GIS.

Therefore, one key challenge for future development of statistics in human geography will be to produce techni ques that can be readily adopted within mixed approaches that mesh the quantitative and the qualitative. Another ongoing challenge will be to respond to computing ad vances associated with the creation of cyberspace and new revolutions in computing and information technology.

See also: Categorical Data Analysis; Chaos and Complexity; Factor Analysis and Principal-Components Analysis; Geographically Weighted Regression; Hypothesis Testing; Markov Chain Analysis; Modifiable Areal Unit Problem; Monte Carlo Simulation; Neighborhood Effects; Point Pattern Analysis; Q Method/ Analysis; Quantitative Methodologies; Regression, Linear and Nonlinear; Sampling; Segregation Indices; Spatial Autocorrelation; Spatial Data Mining, Cluster and Pattern Recognition; Spatial Data Mining, Geovisualization; Spatial Filtering/Kernel Density Estimation; Spatial Interaction Models; Statistics, Descriptive; Statistics, Inferential.

Further Reading

Bailey, T. C. and Gattrell, A. C. (1995). *Interactive Spatial Data Analysis*. Harlow: Addison Wesley Longman.

Birkin, M., Clarke, G. and Clarke, M. (2002). *Retail Geography and Intelligent Network Planning*. Chichester: John Wiley.

Dorling, D. and Simpson, S. (eds.) (1999). *Statistics in Society: The Arithmetic of Politics*. London: Arnold.

Fotheringham, A. S., Brunsdon, C. and Charlton, M. (2000). *Quantitative Geography: Perspectives on Spatial Data Analysis*. London: Sage Publications.

Haining, R. (2003). *Spatial Data Analysis*. Cambridge: Cambridge University Press.

Hair, J. F., Black, B., Babin, B. and Anderson, R. (2007). *Multivariate Data Analysis* (6th edn.). Upper Saddle River, NJ: Prentice Hall.

Johnston, R. J. (1978). *Multivariate Statistical Analysis in Geography: A Primer on the General Linear Model*. London: Longman.

O'Brien, L. (1992). *Introducing Quantitative Geography: Measurement, Methods and Generalised Linear Models*. London: Routledge.

Robinson, G. M. (1998). *Methods and Techniques in Human Geography*. Chichester: John Wiley.

Rogerson, P. A. (2006). *Statistical Methods for Geography: A Student's Guide* (2nd edn.). London: Sage Publications.

Romesburg, C. (2004). *Cluster Analysis for Researchers*. Morrisville, NC: Lulu.com.

Roy, J. R. (2004). *Spatial Interaction Modelling: A Regional Science Context*. Berlin: Springer Verlag.

Thomas, R. W. and Huggett, R. J. (1980). *Modelling in Geography: A Mathematical Approach*. London: Harper and Row.

Walford, N. S. (1995). *Geographical Data Analysis*. Chichester: John Wiley.

Wheeler, D., Shaw, G. and Barr, S. (2004). *Statistical Techniques in Geographical Analysis*. London: David Fulton Publishers.

Wilson, A. G. (2000). *Complex Spatial Systems: The Modelling Foundations of Urban and Regional Analysis*. Harlow: Prentice Hall.

Relevant Websites

http://www.geoda.uiuc.edu
GeoDa: an introduction to spatial data analysis, developed at the Spatial Data Analysis Laboratory, University of Illinois, Urbana Champlain, Illinois.

http://www.icpsr.umich.edu
Inter University Consortium for Political and Social Research: CrimeStat spatial statistics program for analysis of crime incident locations, National Institute of Justice, Washington, DC.

http://www.office.microsoft.com
Microsoft Office Excel: a spreadsheet application containing statistical facilities.

http://www.minitlab.com
Minitab statistical software originally developed at Pennsylvania State University.

http://www.nag.co.uk
Numerical Algorithms Group: GLIM, a generalized linear interactive modeling package developed by a working party of the Royal Statistical Society, London.

http://www.sas.com
SAS statistical analysis software.

http://www.spss.com
SPSS statistical package for the social sciences, first released in 1968.

http://www.mrc bsu.cam.ac
WinBUGS statistical software developed by Medical Research Council Biostatistics Unit, University of Cambridge, Cambridge, UK.

Statistics, Spatial

D. J. Unwin, Birkbeck, University of London, London, UK

Glossary

Centrography The name given to early methods that attempted to describe patterns in point event data.

Complete Spatial Randomness (CSR) The benchmark process used in much spatial statistical analysis.

Event Data Data that locate a point object or occurrence simply by its spatial location.

Field, Field Data In geographic information science it is used to denote a spatially continuous variable, such as air temperature, which has a value everywhere.

First-Order Variation Variation in the density of events that results from variation in some background control.

Joins Counts A test for spatial autocorrelation (q.v.) based on counts of the number of joins in a binary (0/1, black/white, here/not here) indicator variable collected over an area lattice.

***k*-Color Map** A map of some qualitative attribute or attributes using different colors to represent each type of entity recognized, for example, a map of land use.

Lattice Data Data that refer to discrete areas such as counties, countries, and other census enumeration districts.

Local Statistics A relatively recent development in which almost any standard statistic is calculated for a small region (the kernel) around the place of interest and then mapped to show where there is spatial variation that might be of interest.

Mean Center A centrographic (q.v.) measure of the middle of a pattern of point events.

Modifiable Areal Unit Problem The problem that arises in the analysis of lattice data when it is almost always the case that any results are conditional on the often arbitrary choice of spatial units used.

Quadrat Analysis A method for the analysis of point event data based on counts of the numbers in small subregions known as quadrats.

Scatter Plot, Moran A plot of the value for a zone in a lattice against the average of its neighbors used to identify locally unusual observations.

Second-Order Variation Differences in the spatial density of point events that result from some form of interaction between them such as attraction or repulsion.

Standard Distance A centrographic measure of the dispersion of point events around their mean center.

Introduction: Motivations and Justifications

The term spatial statistics refers to the application of statistical concepts and methods to data that have a spatial location attached to them, and in which this locational element is used as an important and necessary part of the analysis. Typically such analysis is descriptive and exploratory, but it might also use the full power of statistical theory to test and/or calibrate models of the spatial processes held to be operating. The term spatial analysis is sometimes used to cover this approach together with deterministic, mathematical procedures applied to location data as in optimal facility location, area coverage problems, the traveling salesman problem, and so on. This article addresses only those spatial analytical methods that rely upon statistical concepts and logic.

At first sight, classical statistical analysis, which is used to infer properties of whole populations of objects from suitable samples or to calibrate models of how the objects of interest are thought to behave, might seem to have little relevance to geography. After all, if we have a map showing the entirety of some unique phenomenon, what is the meaning of statements about, for example, the population mean of these data? Similarly, geographic theory is seldom sufficiently well found to allow for easy model formulation and the almost ubiquitous presence of spatial autocorrelation makes any calibration using conventional methods at best hazardous.

The key concept in the statistical analysis of any spatial data is to regard them, and the map patterns they create, as an outcome, or realization, of a spatial stochastic (random) process. In this interpretation any single mapped 'geography' is regarded as but one of a set of possible outcomes of some hypothesized process that itself creates some spatial differentiation. Typically, but again not always, the benchmark process used is one that has no geography inherent in it, called complete spatial randomness (CSR). Spatial statistical analysis is used to determine how unusual any mapped pattern is relative to this standard. A process that is CSR makes two assumptions. First, it is assumed that everywhere in the study region has the same chance of having a given property; this is the assumption of equal probability. Second, it is assumed that there are no dependencies between places; this is the property of independence. If our Earth had a human or physical geography that was truly CSR, not only would it be a very strange place, geography as we know it – including this volume – would have absolutely no interest whatsoever. Only when there are departures from CSR can we speak of there being any geography of interest.

Departures from CSR are the result of two general types of variation, referred to as first or second order, and these occur when one or other assumption is relaxed. These variational effects are best understood by reference to a pattern of individual point 'events' making up a simple dot map, such as the distribution of cases of a disease. First, variations in the receptiveness of different parts of the study area may mean that the assumption of equal probability cannot be upheld. For example, points representing the individual cases of the disease will cluster in more densely populated areas that are more likely to have infected persons in them. Thus, we have a process that takes place in space, but does not contain within itself any explicit spatial ordering. The results show variation of the first order. Second, it may also be that the second assumption, that event placements are independent of each other, cannot be sustained. For example, some processes involve aggregation or clustering, where the occurrence of one event, such as the existence of somebody with an infectious disease, at a particular location increases the probability of other events being located nearby. The resulting pattern will be more aggregated/clustered than random. Alternatively, if the existence of an event at one place makes it less likely that other events cluster around it, this gives a tendency toward uniformity of spacing. An example might be the distribution of market towns, each of which for its survival requires a hinterland of some minimum area that does not contain a similar town. In both cases we have a process that is not simply taking place in space, but is a truly spatial process that will create a pattern even if the study region is itself homogeneous. The result is second order variation. Processes that combine both sources of variation can also be envisaged and are a common occurrence.

A major weakness of any spatial statistical analysis is that observation of just a single realization of a process, for example, a simple dot map, is almost never sufficient to decide which of these effects is operating. Departures from CSR can be detected using a variety of statistical tests, but it will almost always be impossible to say whether this is due to variations in the environment (first order) or to interactions between point events (second order). It is, however, in the very nature of geography, formalized by many under the name of Tobler's law, or as the so called first law of geography, that spatial processes are never CSR, which in turn means that both first and second order variation are common.

If we have a map that seems to show some pattern, it is natural to want to explore further and characterize this pattern using descriptive statistics. A great many spatial statistics have been assembled to deal with point 'event' data, 'flow' information through a network of linear objects, 'lattice' or area data, and what is often called 'geostatistical' data resulting from the sampling of a continuous 'field' of data such as the earth surface relief or the pattern of temperatures across a city.

Methods for Event Data

With a dot map showing a distribution of some defined features, called the 'events', such as specific landforms, factories, or shops, it is natural to attempt to characterize the pattern that the dots create. The first approach, called centrography, attempted to define the 'center' of such a distribution by quantities such as the mean center, median center, and point of minimum aggregate travel. The spread around such a center was measured using a spatial equivalent of the univariate standard deviation called the standard distance. These measures remain useful for tracking the change in a point pattern over time, or for comparing the distributions of different phenomena, but they do not allow inferences about either first or second variation in the pattern, which is tackled in a number of ways.

The simplest density based approach, called quadrat analysis, uses a fixed grid of small areas that completely covers the study region and counts of the number of grid squares, or quadrats, into which fall 1, 2, …, n events. In effect this is a map of the spatial variation in point event density. Alternatively, and by analogy with a standard field technique in ecology, use is made of a small quadrat thrown randomly and repeatedly onto the pattern of events to provide the same counts. Either way, the frequency distribution of numbers, called a quadrat count, provides some guidance as to the nature of the pattern. Nowadays, with increased computer power, similar maps of the spatial variation in the spatial density are created using varieties of kernel density estimation. In this a spatial window, or kernel, is passed over the entire area visiting every possible place in turn to generate an estimate of the event density at that place. The result is a spatially continuous surface of density estimates that can be contoured, much as one would a surface of ground altitude. Locally high values of the spatial density are revealed as peaks on this surface and indicate the locations of clusters of events. Several variations on this basic theme are possible, and the method has rapidly gone from research frontier to being a standard GIS visualization technique.

As an alternative, patterns of events are also characterized using methods based on the between point distances, of which the mean and standard deviation of the distribution of distances to the nearest neighbor of each and every point event was the first to be used in 1954 by Clark and Evans as an indicator in ecological populations. Over the subsequent half century, their so called R index of randomness has been much used and abused.

More recent work now uses a series of plots of the cu mulative frequency distribution of various distances. All are more informative than the simple approach using the mean distance to nearest neighbor, but to date there have been surprisingly few sustained examples of their appli cation in substantive studies. The simplest, the $G(d)$ function, shows the cumulative distribution of the nearest neighbor distances as a function of distance. Rather more complex, the $F(d)$ function finds the nearest neighbor distances to actual events in the pattern from a series of randomly located locations in the same study region. At some cost in additional computation, the $K(d)$ function uses concentric circles of radius d placed at each of the events in the pattern and counts the number of other events falling into each. The plot of the mean density for each circle radius, d, against distance gives the function. Note that all of these approaches to a greater or lesser extent sensitive to edge effects at the boundaries of the usually arbitrary study regions selected.

All the above methods provide summary statistics that in some sense characterize the pattern of events. More formal spatial statistical analysis tries to determine whether or not they are typical or unexpected relative to a specified stat istical model. As noted above, most work uses CSR as its testing benchmark. It is relatively easy to derive the ex pected distributions generated by the assumptions CSR for the various summary statistics, and thus to assign a prob ability to any observed pattern under this hypothesis. In biology, astronomy, and some experimental sciences, this kind of spatial analysis can be conducted using more real istic process models that modify the basic assumptions, but in geospatial problems this has only seldom been attempted and there are severe analytical problems in differentiating between possible, but competing, models. First, there is an inherent circularity in using a map pattern to estimate key model parameters at the same time as using these self same data to validate that model. Second, although capable of discriminating between CSR and first /second order ef fects, with the available data these methods often are in capable of discriminating between otherwise similar, but alternative process models. Finally, with the exception of kernel density estimation and some recent experiments using agent based approaches, none of these methods dir ectly tackle the problem of cluster detection, determining where significant clusters occur, which is frequently uppermost in practical point pattern analysis in disciplines such as epidemiology and criminology.

Event data are extremely common in human geog raphy and such data have been increasingly available through our present use of post code and similar lo cational systems as a rough spatial point coordinate. Nowadays, methods such as kernel density estimation and plot of the various functions of distance outlined above are routine in, for example, spatial epidemiology, retailing, geodemographics and crime pattern analysis.

Methods for Flows in Networks

Geographic objects such as roads, railways, or river channels are often linear in nature. In practice, the spatial analysis of such line data most frequently represents them as topological graphs in which the networks under study are summarized by matrices in which rows and columns represent 'nodes' (sources or junctions) and the 0/1 entries record whether or not a link is present. At only a slightly higher level of sophistication, the 0/1 can be replaced by entries that represent some measure of distance, flow, or other interaction, between the nodes. An interesting example of this topological representation of line objects that emphasizes connection rather than pure distance is provided by Henry Beck's justly famous 1933 visualization of the London Underground.

In network analysis a basic distinction can be made between studies that describe the pattern of connection *per se*, and those that in some sense take the network as given and analyze the pattern of flows through it. Interest in the structure networks *per se* developed from work in geology where the interest lay in the characterization of natural stream networks for use in hydrological fore casting. Analysis was based on the idea of treating the tree like networks created by river channels as topo logical graphs, within which individual streams could be labeled by their position in the graph as '1st, 2nd, …, *n*th order'. The descriptive regularities that emerged were collectively known as Horton's laws of stream com position. In fact, as numerous workers have shown, these descriptive laws seem to be a general property of more or less all similar tree like natural structures, including the branching of real trees. Similarly, when judged against simple models of random junction formation, the ob served patterns have been found to be extremely com mon realizations. More generally still, networks with closed loops within them have been analyzed using methods from standard network analysis, for example, to compute the shortest path from one network node to another. A well known, but hard to solve, example from algorithmics is traveling salesman problem, but there are many others, including the consequences of the so called 'small world' phenomenon. Describing the structure of geographical networks in this way draws attention to the fact that the network is a very general data structure that is encountered in many other, nonspatial, contexts and this is a field that has developed greatly in recent years.

Methods for Lattice Data

Area, or 'lattice', objects are of three basic types. First, they can arise naturally, as for example the area of a lake, or forest stand. Second, they also arise as imposed and es sentially arbitrary regions created by human agencies, such as census enumeration districts, counties, and countries. In both cases when they are coded into GIS, the term polygon is often used to characterize such data. Third, they are often created by investigators, or the sensors that have been used, as some sort of regular space filling tessellation of the area of interest such as a grid of squares or hexagons. In a GIS, these are usually referred to as a raster of data. Any data collected for these second and third types of lattice, even the appearance of any maps that are drawn using them, are conditional on the boundaries chosen: this is the crux of the so called modifiable areal unit problem (MAUP). In dealing with lattice data, interest focuses on the attribute values assigned to the lattice and not on the patterning of the zones themselves which is known and assumed to be unchanging. The objectives of any spatial statistical analysis are to understand the spatial arrangement of these attribute values, detect patterns in them, and to examine any relationships among a set of attributes measured over the same lattice, while at the same time allowing for any spatial effects that might be present.

The attribute values assigned to any lattice can also show considerable variety. Sometimes these values are simple categorical assignments that refer to the whole of each area, such as lake/dry land, a land use category, or labor/conservative/liberal voting. Such data are usually visualized using distinct color shading to give what have been called *k* color maps, where *k* is the number of colors (categories) used. Notice that these attributes are not themselves a function of the physical area of the zones to which they relate. Alternatively, the numbers might be interval or ratio scaled counts, such as the populations of the areas, which arise as aggregates that are clearly themselves functions of the spatial areas to which they refer. It is vital to note the consequences of the MAUP for such data; they are inevitably related to the sizes of the areas used, such that in both mapping and analysis it is usual and sensible to standardize them in some way, by creating an areal density (population per square kilometer) or other population ratio (% of zone population with some property or other). As with patterns of point events, a number of statistical descriptors of the patterning in area objects have been developed and used. Again the objective is usually to detect and describe departures from CSR in the numerical values assigned to these areas, the phenomenon known as spatial autocorrelation. It cannot be stressed too highly that, in addition to the MAUP and unless it is explicitly incorporated into the analysis, the almost ubiquitous presence of spatial autocorrelation in lattice data will have potentially serious implications for the results of almost any statistical analysis applied to such data.

The simplest test for spatial autocorrelation is a two dimensional analog of the Runs Test in serial data used for situations in which the areas have been coded into two states, black (B) or white (W), according to the presence or absence of some property to give a two color map. In the Joins Count test for spatial autocorrelation, the numbers of joins of different types (B/B, W/W, and B/W) across area borders are compared to those that would be expected under the hypothesis of random coloring.

For interval and ratio scaled attributes, testing for spatial autocorrelation uses two statistics, called Moran's *I* and Geary's contiguity ratio, *C*. Of these Moran's *I*, a simple translation of the standard nonspatial correlation coefficient, is the most popular. It is worth examining its composition in full:

$$I = \left[\frac{n}{\sum_{i=1}^{n}(y_i - \bar{y})^2}\right] \times \left[\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}\right]$$

The important part of the calculation is the second fraction whose numerator is

$$\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})$$

This is a covariance. The subscripts i and j refer to different areal units or zones in the study, and y is the data value in each. By calculating the product of two zones' differences from the overall mean ($\bar{y}$), this quantity provides a measure of the extent to which they co vary. These covariance terms are multiplied by w_{ij}. This is an element from a weights matrix **W**, which switches each possible covariance term on or off depending on whether we wish to use it in the determination of *I*. In the simplest case, $w_{ij} = 1$ if zone i and zone j are adjacent, otherwise $w_{ij} = 0$, so that **W** is an adjacency matrix for the study region.

The rest normalizes the value of *I* relative to the number of zones being considered, the number of adjacencies in the problem, and the range of values in y. The divisor $\Sigma\Sigma w_{ij}$ accounts for the number of joins in the map. The multiplier

$$\frac{n}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

is actually a division by the overall dataset variance, which ensures that *I* is not large simply because the values and variability in y are high. As for a conventional, nonspatial correlation coefficient, a positive value indicates a positive autocorrelation, and a negative value a negative or inverse correlation. The value is not strictly in the range -1 to $+1$, although there is a simple correction that can be made to scale it appropriately. Geary's *C* is formulated in a similar way, but instead of using the covariance between zones it is based on the squared differences in adjacent zone values.

As with point pattern analysis, these measures can be used at the global level in formal statistical hypothesis tests in which the entire mapped pattern is tested against an independent random model. Such tests are an essen tial first step, if only to prevent the eye/brain asserting that some pattern exists when there is no objective evi dence for this. As with commonly used measures in point pattern analysis, they do not identify where over the entire study region such patterning is prevalent. How ever, recent work has moved away from this use as global measures and concentrated more on their utility as what are called local indicators of spatial association (LISA). For example, plots of individual zone scores relative to the average of their surrounding zones, called Moran scatter plots, provide a rapid exploratory method of identifying zones that are unusually high or low in value relative to their immediate surroundings.

For reasons to do with the need to protect individual privacy, aggregated lattice data are extremely common in human geography are possibly the most used and abused of all data types, particularly in any work involving population and related census data.

Methods for ‘Geostatistical’ Data

The major characteristics of field (surface or geostatis tical) data are those of spatial continuity, whereby everywhere has an associated value, and thus also self definition. A familiar example is the relief of the earth’s surface, but many other scalar quantities, such as tem perature or the mean annual rainfall, are also represented as fields. The property of continuity means that in almost all non earth surface relief examples we can never rep resent fields in their entirety and must rely on some data recording scheme that in some sense samples the field as a pattern of discrete, exact values. Examples include spot heights, contour lines, local or global mathematical functions that give the field height, z, as a function of the spatial location (x, y), or a regular grid of exact point values in a digital elevation matrix (DEM). Such fields can be visualized in many ways, but, since the nineteenth century the most common approach has used contours, or, more generally, isolines that plot the locus of points that have the same z values. Note again that it is only possible to draw such maps and analyze these data be cause the phenomena under investigation exhibit spatial autocorrelation.

Possibly the simplest statistical approach to the anal ysis of geostatistical data uses a variation of standard multiple linear regression to calibrate models that ex press the field height, z, as some function of its position:

$$z = f(x, y)$$

Since it is almost never the case that such a function would exactly honor all the data, the surface predicted by any such function is called a trend surface. There is a separate entry that deals in more detail with this approach.

Numerous methods have been used to interpolate from a discrete sample of heights of the field into the ‘unknowable’ spaces between them. The geostatistical approach estimates for the field height at un sampled locations using a simple weighted sum of the surrounding sampled values:

$$\hat{z}_j = \sum_{i\quad 1}^{m} w_{ij} z_i$$

but informs the choice of weights and neighborhood by computation and modelling of a so called experimental semi variogram, estimated from the sample data as

$$2\hat{\gamma}(d) = \frac{1}{n(d)} \sum_{d_{ii}\quad d} (z_i \quad z_j)^2$$

The experimental semi variogram, $\gamma(d)$, is thus a func tion of distance, d, based on the average sum of squared differences in attribute values for all pairs of control points that are a distance d apart. The number of pair of points at separation d is $n(d)$. Note that although tech nically this is the semi variogram and the quantities are the semi variances, in the literature and where the con text makes it obvious, these terms are often shortened to ‘variogram’ and ‘variance’. It is also worth noting the similarity between this measure and Geary’s C measure of spatial autocorrelation. In essence, the semi variogram is an application of exactly the same idea to geostatistical data, with the additional provision that we wish to esti mate its value at a series of distances. Numerous models for the semi variogram have been proposed, as have a variety of methods of fitting such models to the experi mentally determined one. **Figure 1** shows one of the most common, which is the spherical model, defined as

$$\gamma(d) = c_0 + c_1 \left(\frac{3d}{2a} \quad 0.5 \left(\frac{d}{a} \right)^3 \right)$$

The important property of this type of model is that it specifies a continuous function that can be used to esti mate the semi variance at any distance, d. Provided we have the correct model for the experimental semi variogram, ordinary kriging solves a system of linear equations to provide values of the field height that are in some sense optimal. There are other forms of kriging that make different assumptions about the underlying surface and the nature of the data used. Universal kriging is used when there is drift in the mean height of the field. In dicator kriging is used where the field variable is a binary,

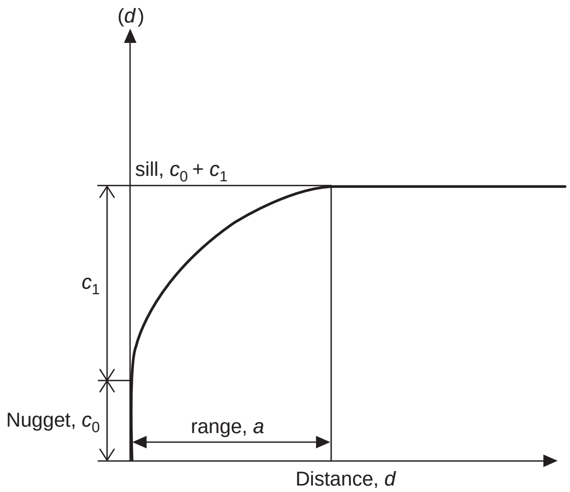

Figure 1 A typical fitted variogram.

0/1, indicator. Disjunctive kriging gives the probability of exceeding a predefined threshold z value and co kriging extends the analysis to two or more variables considered at the same time.

Although introduced primarily as a way of interpolating geostatistical data, readers should be aware that there is a rich theory associated with this method of analysis. This has been developed mostly by applied statisticians and it finds major application in the mineral exploration industry and environmental science. Applications in human geography are rare, but many GIS now offer geostatistical analysis options, and there is a great deal of public domain software, such as GSLIB, that implements and extends these basic notions.

Conclusion

Spatial statistical analysis was introduced into human geography during the so called quantitative revolution of the 1960s and 1970s. At that time computational facilities were primitive and data were generally scarce. What was also often unappreciated was that appropriate methods for spatial data simply had not been developed, with the result that many reported applications lacked a sound basis in statistical theory. One consequence has been the total rejection of this approach by many human geographers working in the 1980s and 1990s. Yet, as the above review shows, our understanding of the problems associated with the statistical analysis of spatial data has developed greatly. Similarly, stimulated by the emergence of GIS and the discovery by other numerate disciplines that geographical location actually 'matters', the available toolkit of methods has been expanded greatly. Although many contemporary human geography theoreticians may think otherwise, the plain fact is that there is now much greater use of these methods in research than at any time during the entire history of academic geography.

See also: Kriging and Variogram Models; Quantitative Revolution.

Further Reading

Anselin, L. (1995). Local indicators of spatial association: LISA. *Geographical Analysis* 27, 93 115.
Bailey, T. C. and Gatrell, A. C. (1995). *Interactive Spatial Data Analysis*. Harlow: Longman.
Chiles, J. P. and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. New York: Wiley.
Clark, P. J. and Evans, F. C. (1954). Distance to nearest neighbour as a measure of spatial relationships in populations. *Ecology* 35, 445 453.
Cressie, N. (1993). *Statistics for Spatial Data, revised edn*. New York: Wiley.
Deutsch, C. V. and Journel, A. G. (1997). *GSLIB: Geostatistical Software Library and User's Guide* (2nd edn.). Oxford: Oxford University Press.
Diggle, P. (1983). *Statistical Analysis of Spatial Point Patterns*. London: Academic Press.
Fotheringham, S., Brunsdon, C. and Charlton, M. (1999). *Quantitative Geography: Perspectives on Spatial Data Analysis*. London: Sage.
Gatrell, A. C., Bailey, T. C., Diggle, P. J. and Rowlingson, B. S. (1996). Spatial point pattern analysis and its application in geographical epidemiology. *Transactions, Institute of British Geographers, NS* 21, 256 274.
Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University Press.
Harel, D. (2000). *Computers Ltd. What They Really Can't Do*. Oxford: Oxford University Press.
Openshaw, S., Charlton, M., Wymer, C. and Craft, A. (1987). Developing a Mark 1 geographical analysis machine for the automated analysis of point data sets. *International Journal of Geographical Information Systems* 1, 335 358.
O'Sullivan, D. and Unwin, D. J. (2003). *Geographic Information Analysis*. Hoboken, NJ: Wiley.
Watts, D. J. (1999). *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton, NJ: Princeton University Press.
Webster, R. and Oliver, M. (2001). *Geostatistics for Environmental Scientists*. Chichester: Wiley.

Relevant Websites

http://www.csiss.org
Center for Spatially Integrated Social Science.
http://www.icpsr.umich.edu
CrimeSat®*III*(version 3.1), ICPSR.
http://www.spatialanalysisonline.com
Geospatial Analysis Spatial and GIS Analysis Techniques and GIS Software.
http://www.ncgia.ucsb.edu
National Center for Geographic Information and Analysis.

Storper, M.

P. Sunley, University of Southampton, Southampton, UK

Glossary

Flexible Specialization An economic system in which districts of specialized and complementary small firms respond rapidly to changes in demand.

Relational Assets Networked and collaborative relationships between economic actors which act as forms of social capital and are beneficial for economic activity.

Technological Trajectories Routes of technological change that follow a course channeled by collective and incremental learning and past experience.

Untraded Interdependencies Relationships between economic actors that are not market contracts and actions, and which increase the capabilities of these actors.

Worlds of Production Varieties of industrial production system defined by differing conventions and common understandings, and hence different types of economic behavior.

Storper, Michael (1954–)

US geographer Michael Storper was born in New York in 1954, and educated in sociology and history at Berkeley in the 1970s (**Figure 1**). A prolific and esteemed economic geographer, Storper's international eminence is evidenced by the fact that he currently holds professorial posts at the London School of Economics, Sciences Po, Paris, and the University of California, Los Angeles. He has been a leading figure in the development of an institutionalist and relational approach to local and regional development. At its core this approach has attempted to explain why regional and urban economies have proved to be, not just resilient but, resurgent entities in the context of the globalization and rapid falls in the costs of communication and transport. Indeed, he has argued that the principal challenge facing economic geography is to explain why regions and cities keep reemerging as sites of agglomerated economic growth in an integrating global economy. Storper describes the contemporary economic landscape as a complex mosaic created by the outcome of tensions between two forces; the dispersal of routine and standardized production operations and a set of agglomerative forces that causes production to concentrate in leading regions. His research mainly focuses on the latter.

Storper's earlier work combined a critique of political economy with an emphasis on the importance of flexible specialization. In particular he argued that Marxian political economy provided a framework for analysis but was too conservative and unable to address micro scale and concrete issues and thus was unable to understand the geographical variability of capitalism and its radically changing nature. Much of his work has attempted to address these more concrete issues by looking at the intersection of what he termed a 'holy trinity' of technology, organization, and territory. In addressing these concrete issues, Storper has drawn insights from institutional, evolutionary economics, economic sociology, and the French school of economic conventions. His expertise and originality lie in synthesizing these ideas and demonstrating how they explain the causes of industrial clustering and co location. There are several key themes in this work.

One of the foundations of Storper's work is the argument that industrialized economies have crossed a major divide from mass production to a post Fordist style of growth. Together with Allen Scott, Storper became a leading figure in the 'Californian School' of economic geography which emphasized the growth and significance of territorial complexes of small firms. Complexes of specialized but complementary firms benefit from economies of scope and are marked by positive feedback loops between agglomeration and vertical disintegration. In his account, film production in Hollywood and the demise of the large studios represent a paradigmatic case of this

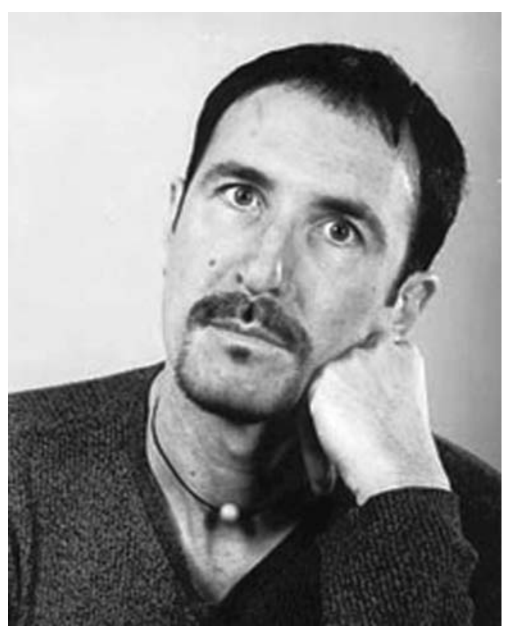

Figure 1 Professor Michael Storper.

change. While some initial explanations emphasized that co location lowered transaction costs, Storper argued that such explanations were insufficient as they were unable to distinguish economically dynamic regions and districts from those that were stagnant and locked in to outdated production methods. They therefore gave in sufficient attention to learning.

A second theme in much of Storper's writing has been the importance of technologically based product learning. Here he has argued that the most competitive technology districts are based on a strong form of competition based on selling products that are differentiated not so much by their price as by their quality. Thus the growth of intra industry trade has seen the appearance of districts that are spe cialized in particular types of product, so that, in Europe for example, trade integration is not leading to regional specialization by industry similar to that of the United States. Storper has explained some of the key character istics of product based technological learning and, in par ticular, he has argued that it is subject to increasing returns and that it shows path dependent, technological trajec tories. He suggests that when new technological trajectories emerge their location may be indeterminate so that there may be windows of locational opportunity. Storper's key original contribution, however, lies in arguing that this technological learning is strongly shaped by apparently 'noneconomic factors' and untraded interdependencies.

One of the most important theses in Storper's work is that learning processes and knowledge creation and transfer are dependent on underlying conventions. He has argued that the key economic problem is one of co ordination between actors and that in the context of un certainty such coordination is dependent on the stabilization of conventions between actors. These are cognitive frameworks that increase the predictability of behavior and so allow collective action. High rates of re flexive learning depend on the presence of the appropriate conventions and relationships. These conventions are not the same in different types of production systems, how ever, but are seen to vary according to characteristics of the product. Together with Robert Salais, Storper dis tinguished four main types of industrial 'worlds' based on different types of conventions relating to whether products are generic or dedicated, specialized or standardized. Such worlds of production are realized in various combinations in different national and regional economies.

Storper's more generic argument that untraded inter dependencies are the key to successful and dynamic re gional economies has been highly influential. In essence, it argues that particular types of extramarkets relation ships are required in order to allow dynamic and reflexive learning in economic systems. For instance, dense face to face relationships generate trust between actors which allows the exchange of tacit knowledge. Storper's inter pretation of urban buzz proposes that regular face to face contact is an efficient means of communicating complex knowledge, as well as a screening mechanism for judging the effort and commitment of others, and it also has a performative effect in motivating and stimulating others. Thus he argues that variations in face to face contact within interpersonal network can help to explain why some cities are more creative and dynamic than others. While these relational arguments have been seen as a move toward a more cultural approach to economic geography, Storper has also been highly critical of cul turalism and postmodernist relativism. He has argued, instead, for a humble and pragmatic type of modernist rationality. His work continues to search for some of the micro foundations of resurgent and emergent urban economies and has theorized the intersections of indi vidual preferences with urban structures.

See also: City-Region; Globalization, Economic; Innovation; Learning Regions; Planning, Urban; Political Economy, Geographical.

Further Reading

Reimer, S. (2004). Michael Storper. In Hubbard, P., Kitchin, R. & Valentine, G. (eds.) *Key thinkers on space and place*, pp 282 287. London: Sage Publications.
Scott, A. J. and Storper, M. (2003). Regions, globalization, development. *Regional Studies* 37, 579 593.
Storper, M. (1992). The limits to globalization: Technology districts and international trade. *Economic Geography* 68, 60 93.
Storper, M. (1995). The resurgence of regional economies, ten years later. *European Urban and Regional Studies* 2, 191 221.
Storper, M. (1996). Innovation as collective action: Conventions, products and technologies. *Industrial and corporate change* 5, 761 790.
Storper, M. (1997). *The regional world: Territorial development in a global economy*. New York: The Guildford Press.
Storper, M. (2003). Regional 'worlds' of production: Learning and innovation in the technology districts of France, Italy and the USA. *Regional Studies* 27, 433 455.
Storper, M. and Christopherson, S. (1987). Flexible specialization and regional industrial agglomerations: The case of the US motion picture industry. *Annals of the Association of American Geographers* 77, 104 117.
Storper, M. and Salais, R. (1997). *Worlds of production: The action frameworks of the economy*. London: Harvard University Press.
Storper, M., Chen, Y. C. and De Paolis, F. (2002). Trade and location in the location of industries in the OECD and European Union. *Journal of Economic Geography* 2, 73 107.
Storper, M. and Venables, A. J. (2004). Buzz: Face to face contact and the urban economy. *Journal of Economic Geography* 4, 351 370.
Storper, M. and Walker, R. (1989). *The capitalist imperative: Territory, technology and industrial growth*. Oxford: Basil Blackwell.

Relevant Websites

http://www.cso.edu/cv equipe.asp?per id=85
Centre de Sociologie des Organisations.
http://www.spa.ucla.edu/dept.cfm?d=up&s=faculty&f=faculty1.cfm&id=257
UCLA School of Public Affairs, Department of Public Policy, Los Angeles.

Street Names and Iconography

M. Azaryahu, University of Colorado at Boulder, Boulder, CO, USA

Glossary

Commemoration The cultural production of the past in different media aimed at facilitating remembrance. Commemoration is a two-tiered process that involves both the representation of the past (content) and its presentation through different media (form). Commemoration in space includes media such as monuments, historic sites, street names and cemeteries.

Semantic displacement The change of meaning that occurs when a historical street name becomes a geographical designation. As a result of semantic displacement, the geographic meaning of a street name take over while the historical meaning becomes increasingly obscure to most users of the city.

Semiotics The study of signs and sign systems as modes of communciation. A semiotic analysis explores how signs facilitate the communication of meaning and how messages are encoded and decoded.

City-text A system of commemorative features in the cityscape. Constantly written and rewritten, a city-text represents not only the past but also the commemorative priorities of former periods.

Introduction

Place names combine geography and language, politics and culture, vernacular traditions and official policies, and contemporary ideologies and visions of history. The shaping or reshaping of sets of toponyms is a profound example of the connection between political and cultural processes.

Ostensibly mundane and seemingly obvious, street names are actively involved in the spatial organization and semiotic construction of the city. Street names designate locations and pronounce certain thoroughfares as distinct urban units. They also conflate urban space and the symbolic realm of cultural signification. Geographical street names (of other cities, rivers, mountains, etc.) confer prestige or conjure romantic associations. When they serve commemorative function, place names and street names in particular are also employed and involved in the semiotic construction of history.

Traditionally, street names were vernacular and designated either by geographical orientation or referred to a certain peculiarity associated with local topography and history. The modern practice of bestowing nonlocal names was introduced in Paris in the eighteenth century. As a prerogative of elected or nominated authorities and affected by administrative and political agendas competing for control over the public domain, naming streets is an expression of authority. Once affixed, street names belong to the structures of power. This becomes especially evident with the conventional use of street names for commemorative purposes.

Following the example set by the French Revolution, commemorative street naming has become a widespread feature of modern political culture. At a practical level, street names distinguish between different streets, provide the users of the city with spatial orientation, and secure administrative control over the city. As a vehicle of commemoration, street names introduce official representations of the ruling sociopolitical order and especially a canonized version of history into the cityscape and into social communication. Commemorative street names provide for the intersection of hegemonic ideological structures with the practices of everyday life. Potentially contested and possibly challenged, they conflate the discourse of history with the geography of the city and the politics of public memory.

Addressing History: The Politics of Commemorative Street (Re)Naming

Prominent squares and streets may commemorate political abstractions or concepts such as President or King, Revolution, Republic, and Independence. Yet the vast majority of commemorative street names in the cityscape perpetuate the memory of historical figures and events found worthy of public honoring by the authorities in charge of street naming. Commemorative street names are a conventional element of the urban texture. Embedded into language, they are constantly evoked in contexts that are commonly detached from the sphere of ideology and the realm of politics. In this semiotic capacity they introduce an official version of history into mundane settings of everyday life.

The significance of street names as historical commemorations becomes apparent when the political role of history, as both a version and interpretation of the past, is considered. On the one hand, political regimes and elites utilize history to legitimate their dominance and augment their authority. On the other hand, the sense of a shared past is crucial to the cultural viability and social cohesiveness of ethnic communities and nation states. Commemorative activities figure prominently during national revivals and stages of state formation by a postcolonial or a revolutionary regime. Officially promoted commemorations belong to the symbolic foundations of modern statehood. In particular, street naming patterns reflect the needs of communities and regimes to assert ‘their’ vision of history in commemorative terms. In the late nineteenth century, Berlin street names commemorated the ruling Hohenzollern dynasty and Prussian military victories. London had 11 Queen Streets and a dozen King Streets. However, since the late nineteenth century commemorating historical heroes and events by street names was not limited to celebrating dynastic glory only, but was increasingly associated with nation building measures and the efforts of political elites to cast national identity in the mold of a national history.

The association of commemorative street names with nation building measures became paramount in the twentieth century in numerous cases of successful ethnic revivals and postcolonial state formations. The street names that appeared in Stockholm in the late nineteenth century, for instance, also commemorated Swedish national history. Introduced in, and expressive of, a stage of national revival, the street names of Tel Aviv represent a Zionist version of Jewish history. Street naming in post independence Singapore was one aspect of the attempt to mold a sense of national identity. Affixed in the early 1990s, the street names of Arab–Israeli cities governed by the Islamic front celebrated the glory of early Islamic history.

Aware of the power of symbols as political statements, authoritarian regimes have utilized commemorative street names to consolidate and to legitimize their authority. Street names in Nazi Germany celebrated a National–Socialist version of German history. The extensive use of street names (as well as names of ships, hospitals, schools, and towns) to commemorate Soviet leaders was a prominent feature of Soviet political culture. This practice was introduced after the death of Lenin in 1924, and reached extreme proportions in Stalin’s cult of personality. As long as the communist regimes of Eastern Europe were in power, the street names of communist cities represented an ideologically correct version of national history that corresponded to the officially promoted Marxist–Leninist interpretation of history.

Socially constructed and culturally formatted, public memory is continuously reevaluated as the consequence of changes in cultural codes, social values, ideological premises, and political power relations. In an open society, where public debate is the norm, commemoration of persons and events is potentially controversial. A case in point is the controversies over the commemoration of Martin Luther King Jr. in the American South, where issues pertaining to evaluation of historical greatness and the scaling of memory in terms of selecting an appropriate street to be named after King have figured prominently.

The intentions of national and local elites and the popular support for these elites notwithstanding, challenging the legitimacy of certain names as well as the historical traditions they represent belongs in the politics of commemorative street naming. Where political opposition is legitimate, the rejection of commemorative names becomes part of public debate. In the late 1980s, a campaign was waged in West Berlin against street names that commemorated the colonial past of the German Empire. Though this campaign targeted specific street names, it provided a focal point for public discussion on a fundamental issue: what constitutes German historical heritage?

The influence of political majorities in the city council is evident in the city’s commemorative traditions. Former communist majorities are still reflected in the street names of Parisian suburbs and Italian towns. A pertinent issue is how a change of local administration impacts on the existing register of commemorations. This is especially relevant when the commemorative priorities of the new administration differ from those practiced by its predecessors. When continuity and stability are considered a priority, a possible solution is the commemorative naming of streets in newly built neighborhoods in order to compensate for alleged or real past commemorative deficiencies.

That commemorative street names belong to the symbolic foundations of the ruling sociopolitical order makes them, together with other symbolic expressions of power, vulnerable to changes in the course of a revolutionary phase of national history. As a measure of historical revision, renaming streets in the context of a regime change recasts the presentation of history in a modified commemorative mold. Renaming the past is a twofold procedure: first comes the decommemoration of persons and events associated with the old regime, followed by commemorating the heroes and events that represent the new regime and its version of history in their stead.

Renaming the past figured prominently in nineteenth century France. With the collapse of the imperial regime in France in 1814, *rue* Napoleon was renamed *rue de la* Paix. The renaming of streets figured prominently in Paris after the establishment of the Third Republic in 1870. The purge of Nazi street names in German and Austrian cities followed immediately after the surrender of the Third Reich in May 1945. In April and May of 1951, 159 streets in East Berlin were renamed in a well orchestrated renaming operation aimed at decommemorating the Prussian tradition from the street signs of the capital of the communist German Democratic Republic.

In the Soviet Union, decommemorations indicated major political changes. The end of the Stalinist era in the Soviet Union, as well as in other East European communist states, was sealed in 1961 with the purge of

Stalin from public memory, including the renaming of streets and cities and pulling down monuments. At the height of *perestroika*, the Central Committee of the Communist Party decided to strip Leonid Brezhnev of the public honors bestowed upon him after his death in 1982 and to rename a large square in Leningrad, together with a famed Moscow suburb, and a city whose names perpetuated the memory of the former Soviet leader.

Notwithstanding local differences, the cases of Moscow, Bucharest, Budapest, and East Berlin demonstrate that renaming the past figured prominently in the transition from a communist to a post communist regime. In these cases, beyond decommemorating heroes of the communist state, the question was also whether and to what extent traditional names erased by the communist regime should be restored. Decommemorating the communist history was an aspect of the democratization of political life, which also meant debating issues pertaining to public memory in the public arena.

Relatively simple and comparatively cheap, renaming streets and other public spaces in the course of a regime change has an immediate effect on language and space. The transformation of the city's political geography introduces the political change and the ideology of the new regime into mundane spheres of urban experience and even intimate levels of everyday life. As an assertion of discontinuity in the perception and experience of urban space, renaming streets is bound to produce a sense of dissonance and confusion among users of the city. When conducted in the context of a regime change, the commemorative street renaming, together with the more spectacular pulling down of monuments, is an act of substantial resonance since it communicates a powerful political message. Renaming streets asserts that a radical restructuring of power relations in society and an ideological reorientation have indeed been accomplished, or are under way.

The Power of Commemorative Street Names

Spatial commemorations formulate memory and history in terms of geographical knowledge: when invested with commemorative functions, monuments and street names are constitutive elements of particular geographies of public memory and a culturally shared vision of history. Spatial commemorations merge the commemorated version of history with the landscape. The effect of this is twofold. On the one hand, the landscape is invested with culturally shared symbolic meanings. On the other, the commemorated past is reified by its integration into the physical environment and its subsequent identification with what is considered the natural order.

Commemorative street naming demonstrates the effect that administrative political processes have on semiotic procedures. Though the authorities decide upon what should be commemorated and where, they can hardly regulate the different contexts in which the name is used, the extent and context of its evocation, or the responses to the past thus evoked. As the result of its commemoration, the past is a concrete semiotic reality rather than a platonic idea or a mere mental abstraction. A primary effect of commemoration is that the encounter with the past, that is, an authorized version of history, becomes an optional shared experience. Of course, the influence of such experiences and individual responses to them are to a large extent governed by preconditioned ideological dispositions and political attitudes, and are dependent upon the context in which the evocation of a particular name takes place.

The main virtue of street names as a commemorative vehicle is that everybody uses them but hardly anyone pays attention to their specific historical meaning and to the fact that they belong to the structures of power. The encounter with a street name is casual. Unlike commemorative monuments, street signs do not induce pathos and are not charged with the sacred. On the other hand, the utilization of street names for commemorative purposes enables an official version of history to be incorporated into spheres of social and spatial practices of everyday life that seem to be totally detached from political contexts or communal obligations, and to be integrated into intimate realms of human interactions and activities.

Street names are not intended to fulfill any cultic purposes. They are intended to assert and to reaffirm the validity of the history they reify as a constituent of everyday life. Their power lies in their ability to make a version of history an inseparable element of reality as it is constantly constructed, experienced, and perceived on a daily basis. History is interwoven with daily life and thus gains the appearance of naturalness, a highly desirable effect in light of the function of history as a legitimating factor.

Semantic Displacements

The name given to a street also invests the street with the official meaning assigned to the particular historical referent. However, the meanings of a street name, anchored as they are in reputations, mythologies, images, and sociospatial practices, are dynamic. Though often determined by personal experiences, they may further be differentiated according to social, ethnic, gender, and generational divisions. This notwithstanding, the importance of the official name is paramount. Contested as it may be politically, and despite the inherent instability of its meanings in the course of time and for different audiences, the official name provides a fixed point of reference in the official geography of the city and, in most cases, also in its folk geography. The official meaning originally assigned to the name, in its capacity as a historical referent, figures prominently throughout the naming procedure and dominates the decision making process that leads to the particular commemoration.

The semiotic operation of commemorative street names is based upon semantic displacements that disconnect the meaning of the name from its original historical referent without, however, denying the existence of such a referent. The success of commemoration by means of a street name becomes evident in how the name loses its historical meaning and becomes a geographical designation, and as such an integral part of the language of the urban landscape. The history behind the name is known to only the few who are familiar with the commemorated history. Semantic displacement is an inevitable aspect of the history of street names. The official commemoration refers to the history represented by the name. However, with the passage of time the name becomes increasingly entangled in the meanings assigned to it as a result of the geographical context of its use, and the name increasingly becomes associated with a location; history becomes geography, and in this geographical capacity it is integrated into narratives of the city.

When names are transported from history to geography, they become susceptible to sharing the fortunes of places. When obscure streets gain in urban significance, their namesake heroes gain in prominence and fame. The opposite is also true – when famous heroes are associated with the reputation of notorious slums or derelict neighborhoods.

In a few cases, a street name – in its capacity as a geographical designation – also becomes a metaphor for a social milieu, cultural phenomena, modes of behavior, and a certain way of life that, at least in their early evolution, were connected with the location. The geographical designation provides the appearance of a specificity and objectivity and solidifies a set of human activities and notions about these activities into a powerful cultural reference that transcends the physical location itself. Dizengoff Street, a central thoroughfare in Tel Aviv, is named after the first mayor of the city whose foundation and development was considered a major achievement of the Zionist enterprise. In Israeli popular culture, the activities centered upon the smart shops and coffee shops on Dizengoff Street, in the 1950s, introduced and featured values of a new consumption and leisure culture in Israel. The type of activities associated with the street provided a model of a new urban experience, for which Dizengoff Street became a celebrated cultural metonym.

Reading History: City-Text

The spatial configuration of commemorative street names in a city defines a particular city text as a semiotic feature of the city that presents a particular mapping of space and time. Constantly written and eventually overwritten, city texts are the sum of additions and erasures. Writing a city text is a prolonged process that conflates urban developments, ideological considerations, and political contingencies. As a particular geography of public memory, a city text represents not only a version of history but also commemorative priorities and hegemonic discourses in different periods.

Reading street signs or referring to street names, as is practiced repeatedly in the context of everyday life, mainly involves their practical function as markers of location rather than their commemorative function. Moreover, a city text is not intended to be read as an entirety, and its eventual reading as a text does not involve any obligation to a prescribed order. In principle, a city text is a spatially configured register of historical figures and events. It is a disorderly representation of history rather than a coherent historical account. However, the semiotic structure of a city text offers a measure of historical analysis that is pertinent to reading a city text.

The correspondence between the significance assigned to a historical figure or event and the spatial status of its commemoration is a prominent feature of historical consciousness. The semiotic structure of a city text provides a built in scale which enables the reader to determine the relative status of the commemorated events and heroes. The status hierarchy in an ideal type text is reflected through the configuration of elements in the urban fabric, for which the underlying rule is that the urban importance of a thoroughfare and the prestige of the associated commemoration are positively correlated, and that urban and historical significance conform to each other. In 1925, the republican authorities named a distinguished street in the historical center of the German capital after Friedrich Ebert, the Social Democrat first president of the ill fated Weimar republic. The resolve of the authorities to accord Ebert an honorable commemoration was also manifest in the selection of an appropriate thoroughfare at the center of Berlin.

Yet, deviations from this pattern often occur as the result of the combined effect of urban dynamics and reevaluations of historical significance. One problem is the evaluation of urban prestige. It seems obvious that, from an urban perspective, a central boulevard is more significant than a small lane on the margin of the city, and that the measure of public exposure of the first is much greater than that of the latter. The evaluation of urban prestige, however, also has to take into account the fact that different parameters may have contradictory effects on the production of prestige and status. A newly built highway may be of supreme importance from the point of view of transportation, yet the context in which its name is used is specific and may even reduce its appeal as a vehicle for commemoration in spite of the public ex posure it guarantees to the name.

The fact that a city text is a product of a protracted historical process is a source of further distortions as regards reading a city text. The history of the text has a twofold effect. One is the reevaluation of the significance of historical figures and events, which is an indispensable element of a societal discourse of history. This process is independent of urban developments. The other is the effect of urban dynamics on the prestige of streets and neighborhoods, in general, which is independent of the academic and political discourse of history.

Commemorations reflect prevailing notions of the period concerning the historical significance and great ness of events and persons, respectively. Unless renam ings were undertaken, a city text would conserve notions and priorities that prevailed in different stages of its emergence. The urban expansion of Berlin in the late nineteenth century and the tight control of the Prussian authorities produced a situation in which numerous streets of the former West Berlin further commemorated obscure Prussian generals and officials.

Urban dynamics is a factor to be reckoned with whenever the relative status of commemorations is de duced from the significance of the thoroughfare in the urban fabric. An example is provided by the history of Herzl Street in Tel Aviv, where, as the result of the ex pansion of Tel Aviv to the North, the urban prestige of Herzl Street gradually diminished.

The version of history represented in a city text is biased in favor of local history; this is evident in the names of local dignitaries, most notably mayors, and local events that are commemorated in the street signs. The recurrence of particular commemorations in different city texts is a measure of the historical significance as signed to them. The popularity of a commemoration in different city texts that belong to the same national context is a marker of status and an indicator of repu tation. From the practice of street naming in France, it can be deduced that Jeanne d'Arc, Hugo, Pasteur, Jaurès, Clemenceau, and Blum belong to the French national consensus, Napoleon and Richelieu are highly contested, while Danton, Robespierre, and Napoleon III – all salient protagonists of modern French history – have a bad reputation.

In contrast to textbooks, which present both heroes and villains of history and an evaluation of their roles in national history, a city text mentions heroes and glorious events only. The guiding principle is that villains are not entitled to the place in eternity ostensibly guaranteed by commemorative street names. Hitler and Pétain are ab sent from contemporary German and French city texts, respectively, in spite of their paramount significance in twentieth century German and French history. The heroes/villains dichotomy becomes apparent when, as a result of political changes and the reevaluation of history that follows, 'heroes' and 'villains' change roles and eventually even their places on the street signs. After the establishment of a socialist–communist administration in Leipzig in 1945, the name of August Bebel, the founding father of German Social Democracy replaced that of Kaiser Wilhelm, his historical archrival, on the street signs of the Saxon metropolis.

The spatial organization of street names does not produce any significant linear order, and the intersection of streets does not necessarily imply any temporal or thematic relationship between their names. Such inter sections, however, may suggest imaginary encounters that are sometimes weighed down with irony. For instance, the intersection of the Boulevard Voltaire with *rue* Saint Sebastien is ironic in the light of the notorious anticlerical disposition of the eighteenth century philosopher.

A fundamental property of a city text is that it lacks a built in time arrow, and hence chronology. Historical figures and events coexist simultaneously, and one can not make a distinction between before and after. A city text does not provide its readers with a chronological narrative of history, but rather with an authorized index of a putative narrative. Notwithstanding the afore mentioned distortions that characterize the semiotic structure of this text, the similarity between the list of street names attached to a city map and the index of a history textbook is illuminating. The index of a city map is an exact reproduction of the city text as it mentions all the names that appear on the map. The alphabetical order of this index implies neither spatial nor temporal order, but is essential for using street names as geo graphical designations that provide for spatial orien tation, which is, after all, considered to be their primary function.

Conclusion

The growing interest in the critical study of street names reflects the understanding that they belong to the his torical and cultural geographies of urban landscapes and spaces and to the political geographies of public memory and historical heritage. Current studies direct attention to the role and performance of street names in the pol itical and administrative shaping of urban space. In par ticular, the analysis of the social context and ideological dimensions of commemorative naming procedures highlight the extent to which street names pertain to the societal discourses of national and local history.

See also: Art and Cartography; Landscape Iconography; Memorials and Monuments; Urban Morphology.

Further Reading

Alderman, D. H. (2000). A street fit for a king: Naming places and commemoration in the American South. *Professional Geographer* 52(4), 672 684.

Alderman, D. H. (2003). Street names and the scaling of memory: The politics of commemorating Martin Luther King, Jr. within the African American community. *Area* 35(2), 163 173.

Azaryahu, M. (1986). Street names and political identity: The case of East Berlin. *Journal of Contemporary History* 21, 581 604.

Azaryahu, M. (1996). The power of commemorative street names. *Environment and Planning D: Society and Space* 14, 311 330.

Azaryahu, M. (1997). German reunification and the politics of street names. The case of East Berlin. *Political Geography* 16(6), 479 493.

Azaryahu, M. and Kook Doron, B. (2002). Mapping the nation. Street names and Arab Palestinian identity: Three case studies. *Nations and Nationalism* 8(2), 195 213.

Gill, G. (2005). Changing symbols: The renovation of Moscow place names. *The Russian Review* 64, 480 503.

Ferguson, P. (1988). Reading city streets. *The French Review* 61, 386 397.

Foote, K., Toth, A. and Arvay, A. (2000). Hungary after 1989: Inscribing a new past on place. *Geographical Review* 90(3), 301 334.

Holscher, S. and Alderman, D. H. (2004). Memory and place: Geographies of a critical relationship. *Social & Cultural Geography* 5(3), 347 355.

Light, D. (2004). Street names in Bucharest, 1990 1997: Exploring the modern historical geographies of post socialist change. *Journal of Historical Geography* 30, 154 172.

Light, D., Nicolae, I. and Suditu, B. (2002). Toponymy and the communist city: Street names in Bucharest, 1948 1965. *GeoJournal* 56, 135 144.

Milo, D. (1986). Le nom des rues. In Nora, P. (ed.) *Les lieux de memoire II: La nation*, pp 28 315. Paris: Gallimard.

Palonen, K. (1993). Reading street names politically. In Palonen, K. & Parvikko, T. (eds.) *Reading the political: Exploring the margins of politics*, pp 103 121. Tampere: The Finnish Political Science Association.

Stump, R. W. (1988). Toponymic commemoration of national figures: The cases of Kennedy and King. *Names* 36(3 4), 203 216.

Yeoh, B. S. (1996). Street naming and nation building: Toponymic inscriptions of nationhood in Singapore. *Area* 28(3), 298 307.

Structural Adjustment

G. Mohan, The Open University, Milton Keynes, UK

Glossary

Balance of Payments The difference between total payments to foreign nations arising from imports of finance, goods, etc. and total receipts from foreign nations from exports of finance, goods, etc.

Civil Society Those organizations that form part of the political sphere, but which lie outside the state and market.

Cold War The period between the end of World War II and the late 1980s when the world was divided between the two superpowers of the US and USSR.

Conditionality The practice by the major international lenders of attaching policy conditions to loans.

Governance A way of organizing policy making and rule through networks of actors rather than simply by the institutions of state.

Informal Sector The area of economic activity which lies outside of the state-regulated system and is characterized by self-employment and small scale.

Keynesian(ism) Economic policy named after the economist John Maynard Keynes, which favors state intervention in correcting market failures and was practiced in the immediate post-World War II period.

Neocolonialism The persistence of external control in the formerly colonized world by foreign governments and transnational corporations.

Privatization The sale of once state-owned enterprises to capitalist firms.

Introduction

Structural adjustment sounds like a highly technical idea, which appears benign, even boring. However, it is a generic term that refers to a set of programs that seek to restructure the economies of countries in the developing world or in transition; literally 'adjusting' the 'structure' of national economic life. These programs are comprehensive and affect everything from government economic policy to ownership of firms and the price of essential services. They are based on a set of neoliberal economic principles that were first practiced from the 1970s to 1980s in a number of 'developed' economies before being exported to the developing world. For these reasons, this apparently unassuming concept affects us all, but has its greatest impact on the poorer sections of society across the world. Structural adjustment programs, or SAPs, are orchestrated by the major international lenders – the International Monetary Fund (IMF) and World Bank – in which policy conditions are attached to loans to indebted and impoverished countries. In this way the financial power of these institutions is central to policy making in developing countries and continues a long history of external interventions in the developing world, which raises vital questions about sovereignty in a globalizing world.

Understanding the dynamics of structural adjustment reveals two important geographical themes. The first concerns long standing interdependencies across space. A persistent feature of the international system has been the way in which capitalist expansion and political intervention have generated uneven development. Within this system, through both formal colonization and less obvious neocolonialism, countries in the global South have been subject to unequal terms of trade and political interference in their affairs of state. While not exonerating these countries' leaders, they have in effect been 'hemmed in' within the global economy, leaving little space for maneuver. Structural adjustment, therefore, reflects the latest phase in this unequal relationship.

The second issue concerns the uneven geographical expression of neoliberalism. Although neoliberalism is based on a set of universal principles and is a term bandied around as if the neoliberal condition were uniform, the realization of neoliberalism is shaped by differences within and between states and societies. So, while it is possible to discern general trends, the real world geography of neoliberalization is uneven and deeply political. The key, and crudely stated, difference between the neoliberalization occurring in the global North and that in the South is that in the former it has been driven by domestic political elites within nominally democratic systems, whereas in the latter it has been forced on impoverished populations by the same political elites acting through multilateral organizations alongside their cronies in these developing countries.

The article begins by tracing the emergence of structural adjustment before outlining what makes up a typical SAP. From here we disaggregate the major effects of SAPs and then examine how the politics of the policies operate and are contested.

The Neoliberal Turn and the Underlying Principles of SAPs

SAPs are based on a neoliberal set of assumptions. In this subsection the genesis of SAPs from the 1970s right up to the current period is examined, where, despite changes in name and emphasis, SAPs are still very much in evidence.

SAPs are based on a belief that free markets can deliver the best route to prosperity and development. In terms of economic policy the post war period up to the early 1970s was broadly characterized as one of Keynesianism, which saw a role for the state to balance out economic cycles and protect the marginalized. By contrast, neoliberalism believes in a minimalist state and a faith that open and unfettered economies can generate growth, which will eventually trickle down to the poorest.

While Keynesian policy was focused on national economies, international economic regulation in the post war period was through a system called the 'gold standard', which pegged currencies to the value of gold and prevented wild speculation. Additionally, a system of international economic institutions was created to avoid volatility. The architecture of this was laid down at a conference in 1944 at the Bretton Woods ski resort in the USA, so that this complex of regulatory bodies became known as the Bretton Woods Institutions and included the IMF, the International Bank for Reconstruction and Development (more usually called the World Bank), and the General Agreement on Tariffs and Trade (GATT), which later became the World Trade Organization (WTO). In different ways these institutions sought to regulate international economic flows and were governed by representatives of the countries that contributed finances.

By the early 1970s pressures had mounted which saw a gradual dismantling of the Keynesian regulatory system. The gold standard was abandoned and currency movement became easier, in many developed economies inflation and labor unrest were distressing governments, and an oil crisis forced an international recession. This was set against Cold War polarities, which saw aid and (often covert) military support for regimes that professed support for one of the superpowers.

In September 1973, a US backed coup d'etat in Chile ousted the left leaning Salvador Allende and replaced him with General Pinochet. Pinochet and his allies in the US put in place an economic model that removed labor regulations and brutally suppressed dissent. The intellectual support came from a group of economists based in the University of Chicago and led by Milton Friedman. These 'Chicago Boys' were the architects of the neoliberal policies in Chile, which acted as a test bed for ideas that would soon come to dominate policy.

Internationally the oil crisis of the early 1970s generated a recession but also huge profits for oil producers. With vast amounts of money in the banks, but fewer productive outlets in the capitalist core, bankers looked for other avenues for profitable investment. These so called petrodollars were lent to developing countries at relatively low interest rates. In the late 1970s, as one of the earliest neoliberal attacks on inflation in the core economies, interest rates were raised rapidly. For those developing countries with loans from Northern banks this created a massive and unsustainable debt problem with Mexico famously declaring itself bankrupt and unable to repay in 1982. For the banks this was a huge (and embarrassing) issue that could threaten the legitimacy of the international financial system and the privileges of wealth it secured. Hence, a mechanism was needed to repay the debt, which ensured that these economies did not decline further so that funds were available for debt servicing. It is at this critical point that structural adjustment first appears.

Given that state financing was available commercially, the Bretton Woods Institutions had to 'reinvent' themselves as development agencies whose main work was in the developing world, and from the late 1980s the former Soviet Bloc as well. Their role became that of ensuring the steady repayment of debt and the inculcation of neoliberal economic 'openness', which they did by attaching policy conditions to the concessional loans they granted to developing countries. It was following the neoliberal turns in the USA under President Ronald Reagan, in the UK under Prime Minister Margaret Thatcher, and in Germany under Chancellor Helmut Kohl that the Bretton Woods Institutions formally introduced SAPs in the early 1980s.

Although set up in Bretton Woods, these institutions are all based in Washington, DC. The orthodoxy that came to dominate their relationships with the developing world became known as the Washington Consensus and enforced a dogmatic reading of neoliberal policy as part of their conditions. This started in the early 1980s and was in full flow by the middle of that decade. These controversial policies had mixed impacts that are discussed below, but in the subsequent quarter of a century they became, and remain, a major feature of international development despite undergoing some changes in emphasis.

As a result of limited impacts, theoretical critique, and active resistance, the 1990s saw a softening on the hard neoliberal line. The discourse shifted away from growth at any costs liberalization to poverty eradication by a more diverse set of means, although never displacing the centrality of the market mechanism. This saw structural adjustment programs falling out of circulation as a term to describe neoliberal policies and being replaced by a new, but very similar, initiative called Poverty Reduction Strategies (PRSs). In order to fully appreciate how structural adjustment operates and the difference that geographical context makes, it is necessary to understand what constitutes a typical SAP.

What Is Structural Adjustment?

At the heart of SAPs is liberalization of the economy. It is based on the persuasive discourse of freedom and self sufficiency, that human beings are by nature driven by self interest, and that anything which prevents the real ization of this potential is wrong. For the neoliberals the main target was the state, which they believed impeded national and international markets and created depend ency among the poor on state welfare.

The 1970s saw the intellectual refinement of this at tack on the state with full blown SAPs emerging in the early 1980s. A typical SAP contains the following elements:

- Tariff barriers, which had been set high to protect local production, are removed to promote competition from imports with a view to kick starting local pro ducers into efficient production.
- Export promotion is also encouraged by pushing for key commodities, especially agricultural goods, to be produced, which diverts efforts away from production for local consumption.
- Devaluation of the national currency, which is often overvalued, to make exports cheaper on the global market and imports more expensive, which en courages exports over imports to correct the balance of payments deficit.
- Financial liberalization to allow freer inward and outward flows of international capital, as well as a removal of restrictions on what foreign businesses and banks can own or operate.
- Subsidies are removed in order to remove market distortions and overcome the dependency of the poor on state welfare.
- User fees are levied on key services as a way of 're covering' the costs and reducing the burden on tax revenue. On the other side of the fiscal equation ser vices are cut back to reduce government spending.
- Bureaucracies are trimmed down and other state workers laid off to reduce the government's wage bill. Bureaucracies are also to be made more efficient and less corrupt under the 'good governance' initiatives that followed SAPs.
- Privatization is introduced as a way of enhancing competitiveness and efficiency, and reducing the burden on the state of sapping and corrupt state owned industries (SOIs). With liberalization, heavily defended national economies are now open to possible inward investment to buy up these SOIs.

Ideally, the upshot of stabilization and liberalization is that the economy opens up and the state reduces spending, which provides a conducive environment for foreign firms and releases finances for debt repayment.

In terms of implementing a SAP there was a division of labor between the IMF and World Bank, with the former following its historical remit of macroeconomic stabilization and adjustment and the latter focusing more on sectoral problems, such as industrial strategy, and supply side bottlenecks, like poor infrastructure. In practice, the two institutions worked together closely. What is crucial about IMF agreements is that they sig naled to other lenders that a particular country was loan worthy. While total IMF lending to a country may not have been huge, it was necessary in order to access other concessional lending.

The high point of SAPs was the mid–late 1980s whereby 64 countries had an SAP in place. Ongoing resistance within recipient governments to SAPs saw the addition of a political dimension to these economic programs in the shape of 'good governance' initiatives. As corruption and mismanagement were seen to be im peding SAPs, attention turned in the early 1990s to re forming the bureaucracy and giving greater voice to potentially pro market elements within 'civil society'. At the same time, changes in government in the UK and USA ushered in a social democratic model, which was still very much pro market but conceded that for those excluded from the benefits of markets so some form of market based support was needed to protect them.

A further development was around debt. After over a decade of SAPs it became apparent that most countries were still mired in debt and barely keeping up with repayments. Concerted international lobbying forced a move toward debt reduction and further concessions to indebted countries, which became tied into the adjust ment agenda. The Heavily Indebted Poor Countries (HIPC) initiative started in the mid 1990s and was aimed at those countries in severe debt, but who also showed the necessary commitment to neoliberal reforms. Once these countries demonstrated a 'rational' plan for im plementing neoliberal policies and made the necessary moves toward 'good governance', they passed the HIPC 'decision point' when further concessions would be made.

In terms of international policy all this produced the 'post Washington Consensus' with a shift away from hard line conditionality and an exclusive growth orien tation toward greater 'participation' in policy making by both recipient governments and their people. The re sponse was PRSs that were introduced in the late 1990s. According to their initiators in the Bretton Woods Institutions, PRSs are country driven, results oriented, comprehensive, partnership oriented, and long term. They are seen to reverse the draconian conditionality of SAPs, which made a set of policies conditional on taking a loan. By contrast, PRSs favor 'smart' conditionality, which assesses each country's institutional and financial worthiness and then awards a loan accordingly. One important change is that rather than giving funds for specific projects, such as the revamping of a port facility, they are given directly to the national treasury to be spent in accordance with the home grown adjustment program set out in the Poverty Reduction Strategy Paper (PRSP). This so called 'direct budget support' appears to give greater control to recipient governments, but despite this rhetoric of ownership all PRSs contain the same elements as SAPs in terms of liberalization and cost recovery.

Impacts and Effects

Although neoliberalism is usually talked about in universal terms, its elaboration and effects are locally specific. In this subsection we review some of the key impacts of SAPs, while trying to capture these complex geographical patterns. Analysis is difficult for various reasons. First is the counterfactual problem, because it can never be known what would have happened to the economy had SAPs never been implemented. The usual approach is simply to take 'before' and 'after' snapshots of key indicators to try and capture what impacts SAPs have had. Second, SAPs focused on domestic conditions and constraints, whereas economic performance is also affected by global conditions so it is hard to separate out which effects were produced by what factors. Third, data for many developing countries are poor so calculating impacts is always difficult. However, qualitative data is often used to get behind the aggregate trends. We begin with aggregate trends before disaggregating the impacts of major reform areas.

Table 1 shows the performance of countries classified as 'strong adjustors' meaning a commitment to SAPs and receiving many loans. It shows that contrary to the objective of 'adjustment with growth', the recipients of adjustment loans had the same near zero per capita growth rates as the overall developing country sample. In the worst cases, there were very poor macroeconomic outcomes, although the best cases showed that growth was possible. The best performers tended to stabilize their economies, which could have gone further awry with SAPs, but then failed to kick start growth. However, there were no cases where growth was reasonable and all macroeconomic imbalances were under control for the adjustment lending period. For example, by 1992 Uganda had shown good growth, but erratic and high inflation, despite having received 14 adjustment loans.

Table 1 Performance indicators for a sample of adjusting countries

Country	*Average annual per capita growth rate (%), from first SAP to 1999*
Africa	
Niger	– 2.3
Zambia	– 2.1
Madagascar	– 1.8
Togo	– 1.6
Cote d'Ivoire	– 1.4
Malawi	– 0.2
Mali	– 0.1
Mauritania	0.1
Senegal	0.1
Kenya	0.1
Ghana	1.2
Uganda	2.3
Other developing countries	
Bolivia	– 0.4
Philippines	0.0
Jamaica	0.4
Mexico	0.4
Argentina	1.0
Morocco	1.1
Bangladesh	2.4
Pakistan	2.7
Minimum in the top 20	– 2.3
Average top 20	0.1
Maximum in the top 20	2.7
Average all developing countries	0.3
Transition countries	
Ukraine	– 8.4
Russian Federation	– 5.7
Kyrgyz Republic	– 4.4
Kazakhstan	– 3.1
Bulgaria	– 2.2
Romania	– 1.2
Hungary	1.0
Poland	3.4
Albania	4.4
Georgia	6.4
Minimum	– 8.4
Median	– 1.7
Maximum	6.4

From Easterly, W. (2002). What did structural adjustment adjust? The association of policies and growth with repeated IMF and World Bank adjustment loans, Center for Global Development, Institute for International Economics, August 2002, www.cgdev.org/content/publicatons/detail/2779.

This strong adjustment lending group includes some notable disasters. Zambia received 18 adjustment loans but had sharp negative growth, large current account and budget deficits, high inflation, massive overvaluation of the currency, and a negative real interest rate. In other regions, there were also problem cases. After the initiation of adjustment lending, Bolivia had hyperinflation, negative real interest rates, and overvaluation. Even when Bolivia stabilized inflation in 1987, growth was poor, interest rates swung from excessively negative to excessively positive, and overvaluation remained.

The ex communist 'transition' countries only received adjustment loans in the 1990s after the fall of the Berlin Wall and the breakup of the USSR. Their average growth was −1.7% per annum, showing an overall decline, but there were more successes than in Africa. Six of the countries had negative per capita growth and four had positive growth after the initiation of SAP lending. Only Albania, Georgia, Poland, and Hungary are clear success stories.

It would be wrong to argue that SAPs have had no benefits, but what is crucial is the distribution of these benefits between and within countries, which brings the logical rider of the distribution of their costs. At one level SAPs have been successful in ensuring debts are repaid. This net transfer of finance from the poor to rich countries can be viewed as a form of tribute, not dissimilar to the actions of colonial governments in the nineteenth and twentieth centuries. In development speak adjusting countries service their debt, which when measured against economic activity (GDP) gives a 'debt–service ratio'. The higher the ratio, the greater percentage of a country's income goes on paying off loans. An example of the typical impacts of SAPs is Ghana. In 1983, at the start of the first SAP, its debt–service ratio was 0.81% but jumped to 10.82% in 1987 meaning that one dollar in nine went on repayments to the lenders. It was the recognition of this crippling debt servicing that forced HIPC and related moves.

Trade liberalization aims to enhance exports and encourage competition from imports to promote domestic industries. All countries with an SAP put such provisions in place and, in general, their exports did rise. In keeping with theories of comparative advantage countries were encouraged to produce in sectors where they had an advantage, which often meant relying on a narrow range of products, such as one or two commodity crops. As a result, adjusting economies rarely diversified to develop intersectoral linkages or a more robust portfolio of different activities. The rise in exports was set against a rise in imports, which meant that the overall balance of payments position did not improve. Many of these imports are finished consumer goods rather than intermediate capital goods that would be used in domestic manufacturing, which means they usually displaced comparable local products that may be more expensive or poorer quality. The net effect of this is deindustrialization where the number of firms declines, production falls away even more, unemployment rises, and purchasing power in the local economy is reduced. Sectorally there has been a shift away from manufacturing to services, but local firms often lack the expertise, which opens the market up to foreign firms. In addition, as larger and more formal firms close, this releases labor, which cannot be absorbed and is forced to enter the informal sector and rely on a range of risky survival strategies. So, those worse affected by trade liberalization are smallholder farmers who cannot produce for export, very small manufacturing enterprises, and semiskilled labor.

In terms of financial liberalization the SAPs aimed to reform interest rates, ease the flow of funds in and out of countries, and make credit easier to access. These are basically about de regulating the financial sector. A key outcome is that while these regulations have been relaxed, only those people and businesses with a sound financial footing have been able to capitalize on these opportunities, leading to concentration of ownership and the further enrichment of the already rich at the expense of the poorer section of society. Particularly badly hit have been small and medium sized enterprises that lacked sufficient collateral for loans. Many of these are women run, so the effects of financial liberalization were felt worse by them. The wealthiest who could gain credit often used a portion of this on nonproductive consumption, which raised demand for foreign imports over local goods, thereby exacerbating the effects of trade liberalization.

While competition from imports, among other things, undermined local producers, other changes such as privatization and labor market reforms have also impacted on the poorest in adjusting countries. In countries with a well established manufacturing base, jobs that were created were often in 'flexible' and semiskilled industries, which meant that labor has low pay, few rights, and little job security. Again, this changed labor market was gendered with women usually taking the jobs in manufacturing and subject to harsh labor conditions. The export drive of SAPs also pushes the economy into specialization, which renders some sectors unviable. For example, where agricultural production is not for export, it becomes obsolete and generates rural unemployment and a concomitant rise in rural–urban migration, which undermines rural communities. And as unemployment rises, households are forced into securing income and resources from wherever they can, which can mean forcing children into informal and risky jobs (**Figure 1**).

Privatization was intended to enhance the competitiveness of the economy and reduce the liability on the state of inefficient enterprises. SAP countries tended to produce an audit of their SOIs, get them into reasonable financial shape, and then offer them for sale on international markets. Typically these included hotels, heavy industry and commodity processing plants, and infrastructures like telecommunications, water, and electricity. Privatization has had mixed results, with efficiency rising in some and a reduction in the state's obligations coupled with a rise in taxable income. In general, the best SOIs were cherry picked by foreign investors which meant that profits leave the country. Socially, the move toward efficiency resulted in

Figure 1 A cartoon showing the limited impacts of SAPs. http://live.newint.org/issue236/curious.htm.

Figure 2 A cartoon showing the social impacts of SAPs. http://live.newint.org/issue269/curious.

streamlining practices and getting rid of unwanted labor, which adds to unemployment levels. Additionally, in most cases, the prices for privatized services like phones rose after privatization pushing them further out of reach of the poor. In this respect, water has been the most difficult and contested as it is obviously vital to life and costing it out of reach of the poor makes humane living near impossible (**Figure 2**).

Cutbacks in social spending and the repricing of government services are designed to balance the government's books and free up funds for debt repayment. In practice, this means reduced funding for education and health in particular, which clearly has major knock on effects for society and economy. The introduction of user fees for these basic services puts them out of reach of the poorest in society, which further deepens their exclusion and poverty. These effects were so visible in the early days of adjustment that a landmark publication called '*Adjustment with a human face*' was published by the United Nations Children's Fund (UNICEF) in 1987, which highlighted the impacts on the poor and vulnerable and called for the introduction of social safety nets to protect people during the harsh, but supposedly temporary, adjustment process. The social protection schemes were rolled out with much fanfare, but donor funding was quite limited and by making access conditional upon certain workfare practices they also served to reinforce a pro market logic.

Agricultural promotion was also geared to exporting cash crops at the expense of locally consumed food crops. Additionally, to enhance the market competitiveness of the sector subsidies on key inputs such as fertilizer were cut. This has major livelihood and environmental consequences. Poorer farmers who could no longer afford inputs saw yields decline, which reduced income and food security. It also meant that larger farmers who had the initial capital or could access credit were able to consolidate their position and expand at the expense of smaller farmers. This often meant taking over the best land and forcing the poorest onto even more marginal land, which was more prone to environmental degradation, thereby fuelling a vicious cycle of poverty and degradation. The focus on cash crops also encouraged mono cropping, that reduced biodiversity and with it an increase in the susceptibility of crops to disease. Similar trends were seen in mineral extraction as mining enterprises expanded operations at the expense of agricultural land, not to mention the pollution problems associated with under regulated heavy industry. Overall, food production for local consumption was negatively affected by these reforms, which saw a consolidation and specialization of production.

While it is difficult to generalize about the aggregate impacts of SAPs, compared to their stated aims there was limited or negative economic growth, increased social polarization, and enhanced foreign ownership. However, debt had been serviced. Within these general impacts women and children were the worst affected. So the economic pie did not increase much in size, but the richest peoples' share of the pie got bigger.

As we have seen, the PRSs promise a big change from the SAPs of the 1980s and 1990s. While it is a little early to say with any precision what the effects have been, there is a lack of clarity over how they will contribute to poverty reduction as opposed simply to stabilization and growth. By and large, the strategies still focus on economic growth without really addressing how any growth is to be redistributed to the poor. At their heart are the same macroeconomic prescriptions of the old SAPs with a continued emphasis on privatization, liberalization, and a reduced role for the state.

The Politics of Sovereignty and Accountability

The economic effects of SAPs are clearly far reaching, but so too are their political impacts. In this subsection we examine the politics of SAP intervention, which affects national sovereignty and promotes accountability to the donors despite a countervailing rhetoric of democratization and self determination.

A key paradox of structural adjustment is that the creation of 'free' markets unencumbered by state regulation requires a great deal of state based politicking, encompassing both the use of raw power and the less obvious deployment of theoretical discourses. The Bretton Woods Institutions are bound, in theory, to noninterference in the politics of sovereign countries, yet conditionality shows that they are central to the decision making of many countries. At the national level we see the state heavily involved in the withdrawal of the state from regulation and service delivery. Rather than disappearing, the state is restructured to form a parallel government run by insulated and unaccountable technocrats, most often located in the finance ministries. On the other hand, much is made in the neoliberal armory of devolution to enhance 'choice' at the local level which runs counter to the massive centralization of power that the implementation of SAPs requires.

This polemic of devolution within SAPs suggests that local governance is to be a key arena for enhancing political choice. This localism has seen a whole host of political experiments centered on 'civil society', 'participation', and 'empowerment'. These positive and radical sounding concepts promise local self determination, but prove to be about marketizing services such as education, bypassing the central state which has in many cases proven obstructive to liberalization, and as a vehicle for promoting certain civil society organizations that are known to be pro market. As a result, 'choice' through devolution is part of the wider move to create markets and weaken opposition to SAPs, while all the time couching it in a language of empowerment.

This localism has proven to be one of the mainstays of the 'good governance' programs already mentioned. After nearly a decade of SAPs it was realized that the economic impacts were limited, which donors put down to political and bureaucratic bottlenecks. In many cases well placed political elites were able to benefit from the new inflows of money, often through corrupt re routing of funds for personal or party gain, and so needed to be stopped or bypassed. This led to a whole raft of political reforms such as downsizing bureaucracies, removing key individuals, insulating decision making, and using external consultants to provide 'objective' analysis of what needed to be changed. These programs were always couched in terms of accountability and transparency, and were tied to democratization agendas that also promised accountability. Such attempts to remove blockages to the reforms were not only a donor driven agenda, but were supported by domestic citizens keen to see an end to corruption. Where the tension arose was the sometimes untransparent and heavy handed way in which good governance reforms were implemented. People noted the irony of

Figure 3 A protestor in South Korea in 1997. Mark Henley PANOS.CO.UK.

these moves toward good governance being forced on countries as part of the loan conditionality, which runs against any definition of democratic politics, and can only ever be justified through an assertion of the moral superiority of Western political models and a concomitant demonization of developing countries as unilaterally irrational and corrupt.

This tension around excessive conditionality has been addressed in rhetoric at least with the PRSs, in which this upward accountability to donors is supposedly reversed, as recipient countries now 'own' their own adjustment programs. As we have seen, despite the promise of local input to program design, which should produce different policy mixes, all PRSs look stunningly similar. The shift to supposedly smart conditionality and direct budget support has not reversed the accountability to donors who still pull the strings and undermine the domestic democratization they work so hard to champion. Rather than brutal conditionality the language has shifted to compliance, yet a whole system of checks are in place – most notably Country Policy Institutional Assessments – to ensure that recipient countries are trustworthy before loans are approved and that annual approvals are made in order for funds to be released. While this makes much sense and ensures the funds supplied by taxpayers of the North are not misspent, it is still not a process in which free rein is given to countries to determine their own development pathways.

However, these harsh economic conditions and political meddling have not gone unopposed. As the neoliberal model creates authoritarian, centralized, and unaccountable politics, ordinary people are often forced into direct action. Dubbed 'IMF riots' these protests combine genuine anxiety over austerity driven economic hardships with a frustration at the way in which national sovereignty has been usurped by the Bretton Woods Institutions and their cabal of donors. These have been occurring since the start of the SAP era, and more recently fed into the worldwide 'antiglobalization' movements (**Figure 3**).

Conclusion

Structural adjustment is a critical issue for the lives of many in the developing world. It is a process linked to older trajectories within the international political economy, which are about uneven interdependencies across space. The developing world is further hemmed in by SAPs and complex multileveled governance sees national sovereignty being usurped by international organizations acting in the name of the poor. They also espouse localism as a way of enhancing choice and freedom, but this proves to be a smokescreen for destatization. By understanding this complex economic and political geography we get an insight into how neoliberalization operates so that it appears everywhere and nowhere at the same time.

Since the so called 'war on terror' began, the economic well being of the global poor has taken on a new significance since poverty is believed to breed grievances, as the poor in the South witness with growing envy the lifestyle of the North. Hence, the poverty focus of adjustment programs is now also about maintaining the security of the North by preventing these grievances building up. Under this war footing the neoliberal governance prerequisites become even more important, because they signal a commitment to like minded (read US administration) values. Muddying the waters further is China's rise in global economics and politics, which potentially provides an alternative to structural adjustment. The Chinese government is ready to lend to developing countries and enter partnerships and joint ventures unencumbered by the same ideological baggage and the burden of debt repayment. All this means that structural adjustment is set to continue, but is likely to change yet further.

See also: Africa; Civil Society; Cold War; Communist and Post-Communist Geographies; Debt; De-Industrialization; Dependency; Governance; Governance, Good; Informal Sector; International Organizations; Neocolonialism; Neoliberalism; Neoliberalism and Development; Privatization; Sovereignty; Third World; Transnational Corporations in Developing Countries; Uneven Development.

Further Reading

Bracking, S. (1999). Structural adjustment: Why it wasn't necessary and why it did work. *Review of African Political Economy* 80, 207 226.

Chossudovsky, M. (1997). *The globalisation of poverty: Impacts of IMF and World Bank reforms*. London: Zed Books.

Cornia, G., Jolly, R. and Stewart, F. (1987). *Adjustment with a human face, vol. 1: Protecting the vulnerable and promoting growth*. Oxford: Clarendon Press.

Cornia, G., Jolly, R. and Stewart, F. (1987). *Adjustment with a human face, vol. 2: Country case studies*. Oxford: Clarendon Press.

Craig, D. and Porter, D. (2003). Poverty Reduction Strategy Papers: a new convergence. *World Development* 31(1), 53 69.

Easterly, W. (2002). What did structural adjustment adjust? The assocaition of policies and growth with repeated IMF and World Bank adjustment loans, Center for Global Development, Institute for International Economics, August 2002. www.cgdev.org/content/publications/detail/2779.

Elson, D. (1995). Male bias in macro economics: The case of structural adjustment. In Elson, D. (ed.) *Male bias in the development process* (2nd edn.), pp 164 190. Manchester: Manchester University Press.

Green, D. (2003). *Silent revolution: The rise and crisis of market economics in Latin America*. London: Latin American Bureau.

Harvey, D. (2005). *A brief history of neoliberalism*. Oxford: Oxford University Press.

Mohan, G., Brown, E., Milward, B. and Zack Williams, A. (2000). *Structural adjustment: Theory, practice and impacts*. London: Routledge.

Schuurman, F. (1997). The decentralisation discourse: Post Fordist paradigm or neo liberal cul de sac? *European Journal of Development Research* 9(1), 150 166.

Stewart, F. and Wang, M. (2003). Do PRSPs empower poor countries and disempower the World Bank, or is it the other way around? *Queen Elizabeth House Working Paper Series, WP* 108.

Williams, G. (1994). Why structural adjustment is necessary and why it doesn't work? *Review of African Political Economy* 60, 214 225.

World Bank (1994). *Adjustment in Africa*: *Reforms*, *results and the road ahead.* New York: Oxford University Press/World Bank.

Relevant Websites

http://www.brettonwoodsproject.org
Bretton Woods Project.

http://www.essentialaction.org
Essential Action.

http://www.globalissues.org
Global Issues, Social, Political, Economic and Environmental Issues That Affect Us All.

http://www.imf.org
International Monetary Fund.

http://www.jubileedebtcampaign.org.uk
Jubilee Debt Campaign UK.

http://www.odi.org.uk
Overseas Development Institute.

http://www.saprin.org
Structural Adjustment Participatory Review International Network.

http://www.worldbank.org
The World Bank.

http://www.tjm.org.uk
Trade Justice Movement Right Corporate Wrongs.

http://www.wdm.org.uk
World Development Movement Justice for the world's poor.

http://www.50years.org
50 Years Is Enough Network.

Structural Equations Models

P. L. Mokhtarian, University of California – Davis, Davis, CA, USA
D. T. Ory, Federal Transit Administration, Washington, DC, USA

Glossary

Confirmatory Factor Analysis An approach to analyzing relationships among variables in which multiple indicators of one or more postulated underlying (latent, or unmeasured) constructs are observed, and the hypothesized relationships of those indicators to the latent constructs are tested.

Covariance Structure Analysis A broad class of statistical methods involving the investigation of patterns in how a set of measures vary together. Includes as special cases regression, analysis of (co)variance, canonical correlation, seemingly unrelated regressions, confirmatory factor analysis, structural equations modeling (SEM), and time series analysis.

Effect (Direct, Indirect, Total) A factor indicating the magnitude of a change in the target variable arising from a unit change in the source variable. Direct effects occur without any intervening variables ($X \rightarrow Y_2$) and are simply coefficients in the model, whereas indirect effects are those involving one or more intervening variables ($X \rightarrow Y_1 \rightarrow Y_2$). The total effect of one variable on another is the sum of the direct effect and all distinct indirect effects.

Identification The issue of whether a set of parameters in a model can be uniquely determined. Depending on the number of observations, the number of parameters to be estimated, and the structure of the model, a model could be under- (or not-) identified, just identified, or overidentified.

Latent Variable An unmeasured variable, generally (in this context) referring to a hypothetical construct for which several observed indicators exist.

Maximum Likelihood Estimation A technique for estimating unknown parameters, consisting of finding the values that maximize the joint probability (or density) of obtaining a given set of sample observations.

Measurement Model A special case or subcomponent of a full structural equations model that specifies the relationships between latent endogenous or exogenous variables and their observed indicators (contrast 'structural model').

Multivariate Normality The condition that a set of random variables jointly follow the multivariate normal distribution. Many parameter estimation techniques, such as maximum likelihood and generalized least squares, require that the variables of interest be multivariate normal.

Nonrecursive A model containing closed loops or paths of influence, such as $A \rightarrow B \rightarrow A$.

Path Analysis An early precursor of modern SEM. Today, the term can be used synonymously with SEM, but generally connotes the use of a graphic portrayal of an SEM to analyze the magnitudes and significances of direct, indirect, and total effects among variables.

Recursive A model containing no closed loops or paths of influence.

Structural Model A special case or subcomponent of a full structural equations model, in which all variables (aside from disturbance terms) are observed (contrast 'measurement model').

Introduction

Suppose we are interested in assessing the impact of information and communications technology (ICT) on travel. We plan to model transportation demand (measured, say, in vehicle kilometers traveled, or VKT), using cross sectional data (with longitudinal data some considerations change, but the basic concepts are still appropriate). Following conventional wisdom, we assume VKT to be a function of socioeconomic and demographic (SED) factors (such as income, auto ownership, and household size), travel costs (TC), and potentially land use (LU) characteristics such as density. We also assume it to be a function of ICT demand (measured, say, in minutes spent communicating electronically).

A logical approach would then be to estimate the coefficients of the following equation using ordinary least squares (OLS) regression:

$$\mathrm{VKT} = \beta_{12}\ \mathrm{ICT} + \gamma_{11}\ \mathrm{SED} + \gamma_{12}\ \mathrm{LU} + \gamma_{13}\ \mathrm{TC} + \varsigma_1 \quad [1]$$

where β_{12} and the γ's are the unknown coefficients to be estimated, and ς_1 is a residual error or disturbance term (each variable could be expanded into multiple measures, the individual case subscript is suppressed for simplicity, and the notation and lack of constant term are clarified later).

However, considerable evidence suggests that the relationship between ICT and travel demand is not unidirectional as indicated by eqn [1]. Rather, it is conceptually, theoretically, and empirically plausible that transportation also affects the demand for ICT.

Specifically, we could also model:

$$\mathrm{ICT} = \beta_{21}\ \mathrm{VKT} + \gamma_{21}\ \mathrm{SED} + \gamma_{22}\ \mathrm{LU} + \gamma_{24}\ \mathrm{CC} + \varsigma_2 \quad [2]$$

where CC = (tele)communication cost.

This pair of simultaneous equations is illustrated in **Figure 1**, where transportation demand and ICT demand are called endogenous variables (i.e., influenced by other variables within the system of interest), and the others are called exogenous variables (determined outside the system). The arrows indicate the expected direction of influence of one variable on another: any variable with an incoming arrow is endogenous, while variables that only have outgoing arrows are exogenous. The classification of a variable as one or the other partly depends on the time frame of the study (the longer the time frame, the more variables there are that should be considered endogenous), and partly on the judgment of the analyst with respect to the nature of the relationships within a given time frame. Note, however, that an explanatory variable can be either endogenous or exogenous.

If both eqns [1] and [2] are true, but we only model eqn [1], we will violate a key assumption of OLS regression: the assumption that the observed explanatory variables are independent of the unobserved variables captured in the disturbance term ς_1. In particular, ICT is a function of ς_1, because it is a function of VKT (eqn [2]), which is a function of ς_1 (eqn [1]). This condition is called simultaneity bias (one form of endogeneity bias, referring to any situation in which an explanatory variable is correlated with the disturbance term in the same equation). The consequences of using OLS to estimate the coefficients of eqn [1] in the presence of endogeneity bias can be severe: in econometric parlance, the resulting estimates will be biased and inconsistent. Conceptually, it can be seen that if both eqns [1] and [2] are true, modeling only eqn [1] will give an untrustworthy estimate of β_{12}: it could be partly reflecting the influence of VKT on ICT, as well as the converse. (To see this even more clearly, imagine that eqn [2] reflected the only correct direction of causality, and that VKT had a significant influence on ICT. If we only estimated eqn [1] instead, we would probably 'find' a significant influence of ICT on VKT, but it would be entirely spurious, reflecting instead the opposite direction of causality.) To obtain a trustworthy estimate of β_{12}, we need a way to sort out the influences in multiple directions.

Structural equations modeling (SEM) is the most appropriate technique for doing that. Equations [1] and [2] above constitute an SEM, whose parameters can be simultaneously estimated through several different approaches described briefly below. A fully general SEM is a system of two or more equations, for which at least one endogenous (left hand side, or LHS) variable is at least once also an explanatory (right hand side, or RHS) variable, and whose disturbance terms are allowed to be correlated. SEMs have regression, analysis of (co)variance, confirmatory factor analysis, seemingly unrelated regressions, and other general linear models as special cases.

Variables in SEMs can be of two types: observed or latent. The latter case arises when we postulate a conceptual construct, such as travel demand, which is not or cannot be measured directly, but whose essence can be captured through two or more observed indicators. For example, we might envision a latent travel demand construct to be represented by observed variables such as VKT, number of vehicle trips, and total travel time. Similarly, we might model a latent ICT demand construct as represented by the observed indicators minutes of electronic communication, kilobytes of information exchanged, number of voice calls made/received, and number of email messages sent/received. In behavioral studies, latent variables are often attitudinal constructs (such as pro environment), which are represented by

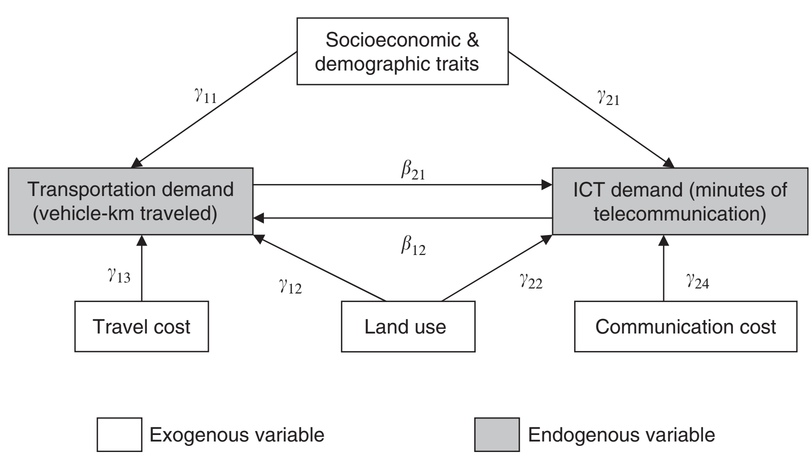

Figure 1 A structural model of transportation and ICT demand. Source: Modified from Choo and Mokhtarian (2007).

observed responses to several related statements or questions.

SEMs in which all explanatory variables are latent are referred to as measurement models, also known as confirmatory factor analysis models. SEMs in which all model variables are observed are called structural models. The two types of models can be mixed. In mixed models, when the measurement submodel is relatively simple, it is common to combine structural and measurement submodels into a single system and estimate all parameters simultaneously. For more complex latent constructs, however, it is also fairly common to conduct an exploratory factor analysis in the first stage, and then use the factor scores obtained from the first stage as observed variables in a structural model. Due to length limitations, this article focuses on structural models such as the one illustrated by **Figure 1**.

SEMs can also be classified as recursive or nonrecursive. A recursive SEM is one with no closed loops among variables (and no correlated disturbance terms between two directly linked endogenous variables), whereas a nonrecursive SEM has one or more such loops. The model in **Figure 1** is nonrecursive, since it has the loop VKT → ICT → VKT. Recursive models are simpler to handle, but the more general case, including nonrecursive models, is addressed below.

In the remainder of this article, we first present a generic structural model and discuss its interpretation. We next describe various estimation approaches, and address the issue of identifiability. We then present the most commonly used measures of goodness of fit (GOF), and finally address the key assumption of multivariate normality.

A Generic Structural Model and Its Interpretation

In matrix form, a generic structural model containing n_Y observed endogenous variables and n_X observed exogenous variables can be written as follows:

$$\mathbf{Y} = \mathbf{BY} + \mathbf{\Gamma X} + \boldsymbol{\varsigma} \qquad [3]$$

where $\mathbf{Y}$ is the n_Y dimensional column vector of endogenous variables, $\mathbf{X}$ is the n_X dimensional column vector of exogenous variables, $\mathbf{B}$ is the $n_Y \times n_Y$ matrix of coefficients showing the direct effects of endogenous variables on other endogenous variables, $\mathbf{\Gamma}$ is the $n_Y \times n_X$ matrix of coefficients showing the direct effects of exogenous variables on endogenous variables, and $\boldsymbol{\varsigma}$ is the n_Y dimensional vector of disturbance terms. In keeping with typical practice, $\mathbf{Y}$ and $\mathbf{X}$ are assumed to be mean centered (expressed as deviations from the means of the raw variables), and as such there will be no constant term in each equation.

For example, the $n_Y = 2$ equation system shown in eqns [1] and [2] can be represented in matrix form as:

$$\begin{bmatrix} \mathrm{VKT} \\ \mathrm{ICT} \end{bmatrix} = \begin{bmatrix} 0 & \beta_{12} \\ \beta_{12} & 0 \end{bmatrix} \begin{bmatrix} \mathrm{VKT} \\ \mathrm{ICT} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & 0 \\ \gamma_{21} & \gamma_{22} & 0 & \gamma_{24} \end{bmatrix} \begin{bmatrix} \mathrm{SED} \\ \mathrm{LU} \\ \mathrm{TC} \\ \mathrm{CC} \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} \qquad [4]$$

Notice that a variable like SED affects VKT in two different ways: through its direct effect, represented by the coefficient γ_{11}, and indirectly through its effect on ICT, which in turn affects VKT. The sum of the direct and all indirect effects is referred to as the total effect of one variable on another. The direct effect can be thought of as the immediate reaction of one variable to a change in another (holding all other variables constant), while the total effect can be viewed as representing the final, longer term overall effect of one variable on another, after intermediate reactions and counterreactions take place and the system settles back into equilibrium. Direct and total effects can be dramatically different from each other in magnitude, and even be of opposite signs. The ability to distinguish and compare direct and total effects is one key feature of SEM that marks it as a major improvement over single equation regression. For example, this capability may identify unintended consequences of a certain policy or trend, in which the direct effects show the desired direction, but are more than counteracted by indirect effects having the opposite impact. Alternatively, an effect which appears to be insignificant in a single equation regression may actually be the outcome of two significant effects of similar magnitude working in opposite directions – quite a different outcome than that of no effect at all.

In the matrix notation of eqn [3], the direct, indirect, and total effects are given by the expressions presented in **Table 1**.

In empirical SEM studies, it is generally of interest to ask which explanatory variable has the greatest (direct or total) impact on a given $\mathbf{Y}$ variable. However, if the $n_Y + n_X$ variables of the system are measured on differing scales, the answer cannot be deduced directly from the raw $\mathbf{\Gamma}$ and $\mathbf{B}$ coefficients, because those coefficients are

Table 1 Matrix decomposition of effects in an SEM

	Direct	*Indirect*	*Total*
Effects of *X* on *Y*	$\mathbf{\Gamma}$	$(\mathbf{I} \quad \mathbf{B})^{1}\mathbf{\Gamma} \quad \mathbf{\Gamma}$	$(\mathbf{I} \quad \mathbf{B})^{1}\mathbf{\Gamma}$
Effects of *Y* on *Y*	$\mathbf{B}$	$(\mathbf{I} \quad \mathbf{B})^{1} \quad \mathbf{I} \quad \mathbf{B}$	$(\mathbf{I} \quad \mathbf{B})^{1} \quad \mathbf{I}$

Source: Based on the data from Mueller (1996).

sensitive to scale (e.g., if the ICT variable were measured in hours, its coefficient would be 60 times larger than that for ICT measured in minutes). To facilitate the com parison of effects, it is common to use so called stand ardized coefficients, that is, the coefficients obtained when all **X** and **Y** variables are standardized (subtracting their means and dividing by their standard deviations, so that the transformed variables all have mean 0, variance 1, and are expressed in terms of standard deviations from the mean of the corresponding original variable), similar to so called beta coefficients in a regression equation. The magnitudes of standardized coefficients are directly comparable, and thus can be used to identify the relative importance of each variable in an equation (with the result, however, that the units of measurement have be come standard deviations rather than the more natural original units).

Path Analysis

The term path analysis can be used synonymously with SEM, but generally connotes the use of a graphic por trayal of an SEM to analyze the magnitudes and sig nificances of direct, indirect, and total effects among variables. Path analysis is best introduced in the context of a recursive model, so we present a new example: a simple formulation illustrating the role of attitudes in residential location and travel behavior. In the conceptual model of **Figure 2**, attitudes toward travel modes and various types of neighborhoods (AT) affect the choice of residential location (RL, probably operationalized as one or more built environment characteristics representing the density and diversity of the neighborhood) and of travel behavior (TB) given RL. Residential location probably also affects travel behavior directly, and it is still an open research question: is the influence of residential location on travel behavior really due to the prior self selection of individuals into neighborhoods consistent with their attitudes and SED characteristics, or does residential location exert a separate influence of its own? SEM offers one powerful approach to answering that question.

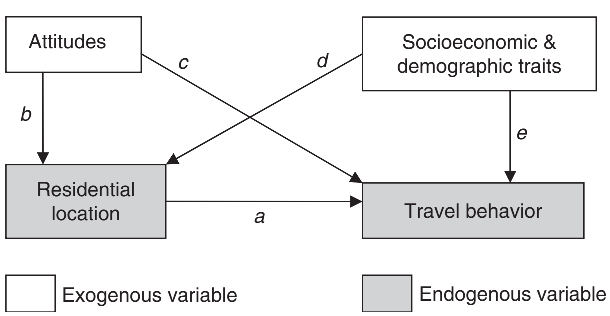

Figure 2 Paths in a recursive model of travel behavior.

The links in **Figure 2** have been labeled with the letters *a*–*e*, representing the direct effects of each variable on the relevant others. Indirect effects can be computed as the product of the coefficients on each link in the path. For example, the indirect effect of AT on TB through RL is $b \times a$, and the total effect of AT on TB is the sum of the direct and indirect effects: $c + ba$. Similarly, the total effect of SED on TB is $e + da$, and the total effect of RL on TB is equal to its direct effect, *a*. Just as in a single equation regression, if the direct effect *a* remains statis tically significant even after AT and SED are included, we would conclude that RL has an autonomous effect of its own on TB. With a single equation model, however, the intertwined nature of the effects of AT and RL on TB would not be made explicit, and either or both of their total effects on TB might be misrepresented by the es timated coefficients. Further, as indicated in the intro duction, a single equation model could not properly capture feedback loops (not shown in the figure), as would occur if one's travel behavior influences one's at titudes over time.

Path analysis for nonrecursive models is more chal lenging, because the presence of closed loops (such as that between VKT and ICT in **Figure 1**) means that there is an infinite process of back and forth influence among the variables in the loop. Thus, in a complex nonrecursive model it can be difficult to decompose a total effect into all of the component paths. However, direct and total effects are readily computed using the formulas of **Table 1**, and the difference between them constitutes the sum of all indirect effects.

For some values of coefficients, a nonrecursive model can be unstable – the infinite back and forth process does not 'damp down' and converge to a finite total effect, but 'blows up'. As its name suggests, the stability index is an indicator of the stability of such a model, and should be evaluated for any nonrecursive model. Some SEM soft ware packages will compute and report the stability index; values less than 1 are generally taken to indicate a stable model.

Estimation

There are several different methods for finding the 'best' set of **B** and **Γ** coefficients in an SEM, all of them based on a general approach called covariance structure an alysis. The idea is that the relationships among the **Y** and **X** variables are manifested in how those variables covary: taking other relationships into account, two variables will have a high magnitude covariance if they are strongly (linearly) related, and a low one if they are not. The true population covariance structure is represented by the $[(n_Y + n_X) \times (n_Y + n_X)]$ covariance matrix $\mathbf{\Sigma}$, for which the observed sample covariance matrix **S** is an unbiased estimator. The hypothesized model, on the other hand, implies a certain set of covariances among the **Y** and **X** variables, and so a natural goal is to find the set of parameters that minimizes the discrepancy between the population covariance matrix (as estimated by **S**), and the covariance matrix implied by the model (which is a function of those unknown parameters).

The unknown parameters include the elements of **B** and **Γ**, as well as the true population variances and covariances of the exogenous variables and of the disturbance terms. Define those covariance matrices as **Φ** ($n_X \times n_X$) and **Ψ** ($n_Y \times n_Y$), respectively. For expository purposes, it is customary to array all unknown parameters in a single vector, called **θ**. Then, for a structural model with no latent variables, it can be shown that the model implied covariance matrix for the observed variables **Y** and **X**, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, can be written in the following $(n_Y + n_X) \times (n_Y + n_X)$ block matrix form:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{YY}}(\boldsymbol{\theta}) & \boldsymbol{\Sigma}_{\mathbf{YX}}(\boldsymbol{\theta}) \\ \boldsymbol{\Sigma}_{\mathbf{XY}}(\boldsymbol{\theta}) & \boldsymbol{\Sigma}_{\mathbf{XX}}(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} (\mathbf{I} \quad \mathbf{B})^{-1}(\boldsymbol{\Gamma\Phi\Gamma}' + \boldsymbol{\Psi})[(\mathbf{I} \quad \mathbf{B})^{-1}]' & (\mathbf{I} \quad \mathbf{B})^{-1}\boldsymbol{\Gamma\Phi} \\ \boldsymbol{\Phi\Gamma}'[(\mathbf{I} \quad \mathbf{B})]^{-1}]' & \boldsymbol{\Phi} \end{bmatrix} \quad [5]$$

where **X** and **Y** are mean centered ($E[\mathbf{X}] = E[\mathbf{Y}] = 0$), and exogenous variables and disturbances are uncorrelated ($E(\mathbf{X}\boldsymbol{\zeta}') = 0$).

Thus, estimating a structural equation model involves finding the parameter estimates that minimize a scalar discrepancy or fitting function, $\mathrm{F}[\mathbf{S}, \boldsymbol{\Sigma}(\hat{\theta})]$, where F indicates the difference between the sample covariance matrix **S**, and the model implied variance–covariance matrix for the estimated parameters $\hat{\theta}$, $\boldsymbol{\Sigma}(\hat{\theta})$. The three most common approaches to estimating SEMs are maximum likelihood (ML), generalized least squares (GLS), and asymptotic distribution free (ADF), a type of weighted least squares (WLS) method. The first two approaches assume the observed variables to be multivariate normally (MVN) distributed, while the ADF method requires no distributional assumption (the same is true of a fourth approach, bootstrapping, which is still experimental). When the model is correctly specified and the data are MVN, all three approaches are asymptotically equivalent, that is, give similar parameter estimates for large samples. The MVN assumption is addressed in more detail in a later section.

Identifiability

Degrees of Freedom

Users of linear regression are accustomed to understanding the degrees of freedom (df) in an OLS model to be the sample size minus the number of estimated parameters. The calculation of df differs in SEM. Because SEM involves the analysis of the covariance structure of the endogenous and exogenous variables, those covariances (including variances) constitute the observations in the dataset. For $n_Y + n_X$ variables, the number of nonredundant covariances (since covariance matrices are symmetric) is $(n_Y + n_X)(n_Y + n_X + 1)/2$, and that quantity minus the number of estimated parameters (the length of the **θ** vector defined above) is the df of the model. Note then that df in an SEM are not a function of the sample size, but only of the number of variables. Increasing the sample size does not increase df; only increasing the number of variables, or decreasing the number of estimated parameters (generally through restricting some to fixed values such as zero), can do so. However, a larger sample is still better in that it increases the precision with which parameters can be estimated.

Testing for Identification

Identification is the assessment of whether the unknown parameters of an SEM can be estimated or not. This assessment has three possible outcomes: exact (or just) identification, overidentification, and underidentification. If any of the equations in an SEM are underidentified (also called not identified), the model cannot be estimated. If the model is exactly identified (also referred to as saturated), all parameters can be estimated and the model will fit the data perfectly (there are zero df), but such a model will be hostage to the idiosyncratic nature of that particular sample, and will probably not be very transferable to other samples. Thus, just as in a single regression equation we want far more cases than parameters, SEM overidentification is preferable to exact identification. If either of those two conditions is true, however, we say the model is identified.

Fortunately, recursive models are always identified, assuming the number of parameters to be estimated does not exceed the number of nonredundant covariances. For nonrecursive SEMs, two methods are frequently used to determine their identification status: the order and rank conditions. The order condition, a necessary but not sufficient requirement (for models allowing all possible disturbance correlations), is applied equation by equation. It states that the number of variables (endogenous and exogenous) excluded from a given equation must be greater than or equal to the total number of endogenous variables (i.e., equations) in the system, minus 1. The rank condition, which is a sufficient requirement, is more complex, but can be tested by the estimation software. These are tests for structural identifiability; models can also be empirically nonidentifiable due to multicollinearity or other data problems.

If the model is not identified, its specification needs to be modified, generally either by applying some restrictions (e.g., constraining some parameters to equal zero, or to be proportional to others), or by adding some exogenous variables to the system. The quest for identifiability often necessitates the imposition of restrictions that are undesirable from a conceptual standpoint (in essence assuming relationships that should in fact be statistically tested). The only way around this is to have an ample supply of relevant exogenous variables, which can be difficult in some applications. However, note that identifiability can and should be ascertained for the conceptual model, before the data collection stage. For example, it can easily be seen that the model of eqn [4] (if each element stood for just one variable) barely meets the order condition, since there are two endogenous variables in the system and each equation excludes one (exogenous) variable.

Goodness-of-Fit Measures

In OLS regression, the R^2 GOF measure represents the proportion of variance in Y that is explained by the model. In SEM, by contrast, model fit refers to how close the model implied covariance matrix $\mathbf{\Sigma}(\boldsymbol{\theta})$ is to the true population matrix $\mathbf{\Sigma}$, as estimated by the observed sample covariance matrix $\mathbf{S}$. A high overall GOF for an SEM does not necessarily mean that individual endogenous variables are predicted well, nor that exogenous variables are strongly or significantly related to endogenous ones.

The foundational measure of GOF in SEM is the model χ^2 statistic, generally written as $(N-1)F_{\mathrm{ML}}$, where N is the sample size and F_{ML} is the discrepancy function used in ML estimation. The larger the discrepancy between $\mathbf{S}$ and $\mathbf{\Sigma}(\hat{\theta})$, the larger the model χ^2, so we want the measure to be as small as possible and should reject the model, as inconsistent with the data, if it is too large. However, the χ^2 measure increases with sample size, so the same large sample that is required for the theoretical properties of the statistics to apply, makes it virtually impossible to fit a model with a small χ^2. For this reason among others, the χ^2 statistic is not exclusively relied upon, but it is always reported, and it is involved in most of the other measures of GOF that have been developed.

GOF measures can be classified into various families, based on their underlying concept. Although space does not permit an extensive discussion of each family, let alone each indicator, **Table 2** summarizes the most commonly used measures in each category (values of these indicators are generally computed by the estimation software). Since no single measure fully addresses every aspect of a model, it is customary to review and report measures from multiple families in evaluating a model's GOF (be aware, however, that in general these measures rely on the assumption that the data are MVN, and that the sample size is reasonably large). Measures that are most commonly reported include χ^2, χ^2/df, root mean square error of approximation (RMSEA), and comparative fit index (CFI). It is also advisable to estimate and compare the GOF of several equivalent models, representing different hypotheses about directions of causality, in addition to the model embodying the analyst's preconceptions.

The Assumption of Multivariate Normality

As mentioned earlier, the use of either the ML or GLS estimation methods requires that the observed variables in the model have the multivariate normal distribution. Evidence indicates that parameter estimates are fairly robust even with serious departures from normality, but in such instances, (1) standard errors will tend to be underestimated, so that parameters will be incorrectly inferred to be significant, yet (2) the model χ^2 statistic will tend to be overestimated, so that the model may be wrongly rejected.

Software packages often check for normality by assessing the univariate normality (UVN) of variables individually. This is done by calculating the skewness (departure from symmetry) and kurtosis (degree of peakedness) for the distribution of each variable. However, UVN is only a necessary, but not sufficient, condition for MVN: a set of MVN variables are each UVN, but each of a set of variables could be UVN without being jointly MVN. MVN can be tested using the Mardia coefficient of multivariate kurtosis, which is asymptotically distributed as standard normal. A set of variables is considered to be MVN with 95% confidence if the critical ratio of Mardia's coefficient to its standard error is smaller than 1.96.

It is quite common to find that the observed variables are not MVN. In this case, several remedies are possible. First, one could transform variables – generally just the variables departing most markedly from UVN. Selective log, square root, or Box–Cox transformations often bring a set of variables much closer to MVN, with a penalty on the straightforwardness of interpretation of the resulting coefficients. The sum or average of a set of conceptually related variables is often more approximately normal than the individual variables (though again, complications in interpretation can result).

Second, one could remove outliers from the sample, that is, discard cases that conform the least to an MVN distribution. These cases are typically identified as those having the greatest Mahalanobis distance (a measure of how distant, in standard deviation units, a vector of observed variable values is from the vector of sample means). In practice, however, achieving MVN this way

Table 2 Summary of common goodness-of-fit measures

Measure	*Description*	*Recommended values*	*Typical values in operations research: mean (range)*	*Typical values in marketing: mean (range)*
χ^2 (N 1) F_{ML}, where N sample size	Discrepancy between observed and model-implied variance covariance matrices; low values are better	Depends on df want to fail to reject H_0 of no difference between Σ and $\Sigma(\theta)$		
χ^2/df	Reduces the sensitivity of χ^2 to sample size	<5	1.82 (0.02, 4.80)	1.62 (1.19, 2.26)
Standardized root mean square residual	Square root of average squared difference between cells in the observed and estimated correlation matrices	<0.1	0.052 (0.01, 0.14)	0.05 (0.03, 0.06)
Root mean square error of approximation (RMSEA)	Amount of error of approximation per model degree of freedom, correcting for sample size and penalizing model complexity	<0.08	0.058 (0.00, 0.13)	0.06 (0.03, 0.08)
Absolute fit indexes (comparison to no model)				
Goodness-of-fit index (GFI)	An absolute fit index that estimates the proportion of variability explained by the model (similar to R^2 in regression models)	>0.9	0.93 (0.75, 0.99)	0.95 (0.90, 0.98)
Adjusted goodness-of-fit index (AGFI)	GFI penalized for model complexity	>0.9	0.89 (0.63, 0.97)	0.91 (0.84, 0.95)
Incremental or comparative fit indexes (comparison to baseline independence model)				
Normed fit index (NFI)	Proportion of baseline (independence) model χ^2 explained by the model of interest	>0.9	0.91 (0.72, 0.99)	
Relative fit index (RFI)	NFI corrected for degrees of freedom	>0.95		0.85 (0.78, 0.91)
Incremental fit index (IFI)	Incremental improvement of the model of interest over the baseline (independence) model	>0.9	0.94 (0.88, 0.98)	0.95 (0.91, 0.97)
Comparative fit index (CFI)	Assumes a non-central χ^2 distribution for the baseline model discrepancy	>0.9	0.96 (0.88, 1.00)	0.95 (0.91, 0.97)
Predictive fit indexes				
Akaike information criterion (AIC)	Balances the discrepancy against complexity	Smaller is better		
Browne Cudeck criterion (BCC)	Penalizes complexity more heavily than the AIC	Smaller is better		
Expected cross-validation index (ECVI)	Discrepancy between the estimated model-implied covariance matrix and its expected value	Any value possible; smaller is better when comparing models		

Source: Data from Baumgartner Homburg (1996) and Shah and Goldstein (2006).

often requires discarding an uncomfortably high pro portion of the data, with the result that the analyst may have to choose between a theoretically conforming model which is estimated on a sample that is no longer representative of reality, and one estimated on the full sample but which does not meet a key assumption of the estimation method. These two undesirable outcomes are often balanced through removal of a few of the most egregious outliers, transformation of some variables, and acceptance of a certain degree of departure from MVN. Some studies suggest that violating the MVN require ment is less consequential the larger the sample size.

A third remedy for non normality is to use test stat istics and estimates of standard errors that are corrected for this condition, and hence considered robust. However, these improved estimates require large samples, and are not available or well documented in all software packages.

A final remedy for non normality is to use the ADF estimation method. However, this method is most effec tive for relatively simple models and large samples (1000 cases or more), which restricts its usefulness. For small samples or when non normality is not severe, the ADF method can in fact be inferior to the others.

See also: Factor Analysis and Principal-Components Analysis; New Urbanism; Regression, Linear and Nonlinear; Selection Bias; Telecommunications; Transportation and Land Use.

Further Reading

Baumgartner, H. and Homburg, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing* 13, 139 161.

Bentler, P. M. and Dudgeon, P. (1996). Covariance structure analysis: Statistical practice, theory, and directions. *Annual Review of Psychology* 47, 563 592.

Bollen, K. A. (1989). *Structural Equation Models with Latent Variables*. New York: Wiley.

Byrne, B. M. (2001). *Structural Equation Modeling with Amos: Basic Concepts, Applications, and Programming*. Mahwah, NJ: Lawrence Erlbaum Associates.

Golob, T. F. (2003). Structural equation modeling for travel behavior research. *Transportation Research B* 37, 1 25.

Hoyle, R. H. (ed.) (1995). *Structural Equation Modeling: Concepts, Issues, and Applications*. Thousand Oaks, CA: Sage Publications.

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling* (2nd edn.). New York: The Guilford Press.

Lei, M. and Lomax, R. G. (2005). The effect of varying degrees of nonnormality in structural equation modeling. *Structural Equation Modeling* 12(1), 1 27.

Long, J. S. (1983). *Confirmatory Factor Analysis: Preface to LISREL*. Beverly Hills, CA: Sage Publications.

MacCallum, R. C., Browne, M. W. and Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Physiological Methods* 1(2), 130 149.

Mueller, R. O. (1996). *Basic Principles of Structural Equation Modeling: An Introduction to LISREL and EQS*. New York: Springer.

Rigdon, E. E. (1995). A necessary and sufficient identification rule for structural models estimated in practice. *Multivariate Behavioral Research* 30(3), 359 383.

Shah, R. and Goldstein, S. M. (2006). Use of structural equation modeling in operations management research: Looking back and forward. *Journal of Operations Management* 24, 148 169.

Relevant Websites

http://www.statmodel.com
M*plus*.

http://www.mvsoft.com
Multivariate Software, Inc.: EQS.

http://www.eviews.com
Quantitative Micro Software: EVIEWS.

http://www.ssicentral.com
Scientific Software International, Inc.: LISREL.

http://www.spss.com
SPSS, Data Mining, Statistical Analysis Software, Predictive Analysis, AMOS.

Structural Marxism

A. Kent, University of Manchester, Manchester, UK

Glossary

Agency The capacities and potentialities human beings possess to re/make history and geography given certain limits and pressures.

Contradiction A relational or dialectical ontology (i.e., theory of reality and existence) holds that each body or thing internalizes myriad social relations which not only support but also undermine one another, and thus set in motion change.

Determination The process whereby internal relations set limits to and exert pressures on behavior.

Economism A school of thought that holds revolution to be inevitable for the contradictions in capitalism will lead to its destruction not because of, but despite people.

Essentialism A school of thought that holds that some bodies and things in the world are essentially capitalist (or whatever) through and through.

Humanism A school of thought that holds the will, choice, intention, experience, and consciousness of men and women to be central to the re/making of everyday life.

Overdetermination A 'cause without a cause' in determinist Marxism, the economy is both (over)determined as well as (over)determining vis-à-vis politics and ideology in overdeterminist Marxism.

Relative Autonomy Rather than being reducible to the needs and wants of capitalism, the state, the legal system, the education system, the family, etc., have a historically and geographically specific effectivity in relation to it.

Structure The invisibly real and really invisible processes and relations through which social life is constrained and enabled whether we acknowledge it as such or not.

Introduction

Everyone knows that Marxism is dead. Dumped on the ash heap of history, Marx and the Marxists are well and truly over, made obsolete by a triumphalist capitalism. "There is," as Mrs. Thatcher once said, "no alternative." Capitalism is so interwoven with our lived experiences that cultural critics like Fredric Jameson and Slavoj Žižek have remarked that it is probably easier to imagine the end of the world than to imagine the end of capitalism. It relentlessly penetrates ever more bodies and things; it inexorably colonizes each and every place, space, and geographical scale.

We live, it seems, in a 'postpolitical' world, beyond Left and Right, Labour and Conservative, Democrat and Republican. In these apparently neoliberal times, once divergent ideologies and policies are becoming increas ingly convergent. The state rolls back and the market rolls out, touted as the solution to myriad problems – everything from healthcare and education to develop ment and the environment. The discourse of global capitalism saturates societies such as ours so broadly and so deeply that it constitutes a sense of reality for most men and women, most of the time.

However, step outside, turn on a television, or open a newspaper; uneven development and poverty are terribly – even painfully – obvious. Furthermore, the aggressive spread of the new imperialism is undeniable, and ex ploitation and alienation are plainly facts of everyday life. To be sure, these world historical matters of concern exist not despite but precisely because of capitalism, and thus it is more than a little troubling that few people are willing and/or able to think within, against, and beyond it. The need for a (re)turn to Marxism, then, is grave and great.

Capitalism is all too easily taken for granted; as it becomes ubiquitous, we become blind to its being in the world. Given that capitalism is something we live day after day, we tend to uncritically normalize or naturalize it. In other words, inasmuch as it is everywhere, it is nowhere, and this is the beauty (*sic*) of capitalism – it is invisibly real and really invisible. Rather than a tangible object to be identified and made accountable, it is something spectral and slippery that we cannot easily lay our hands on or pin down. Karl Marx famously declared that philosophers (and, indeed, social scientists) have only interpreted the world in various ways; the point is to change it. Structural Marxism, then, is all about under standing the world differently, illuminating the peculi arity and sociality of capitalism, and in so doing taking the first step down the road to revolutionary social change. Rather than an end in itself, it is a 'way of seeing' that transforms both how we participate in as well as know the world – a move toward building survivable futures.

The Marxist philosopher Henri Lefebvre famously argued that space was social through and through, that it was not given once and for all, but that it was produced and reproduced day in, day out, in and through our practice, our labor. The question asked by structural

Marxists, then, is whether or not we wish to go on re/producing the places, spaces, and geographical scales of uneven development, poverty, the new imperialism, exploitation, alienation, etc., for there is nothing normal or natural about these unjust geographies.

This article proceeds in three sections. First, it presents an introduction to structural Marxism in general (focusing especially on the work of the Marxist philosopher Louis Althusser), and Marxist human geography more specifically. Second, it presents an overview of the major critiques of both Althusserian Marxism and the work of human geographers such as David Harvey and Neil Smith. Finally, it presents an analysis of the limited and limiting nature of these critiques.

Before moving on, a health warning. Marx once said that there is no royal road to science – its luminous summits demand the fatiguing climb of its steep paths. What follows, in places, operates at a high level of abstraction, and thus may be less than easy reading at times. However, complaints should be addressed neither to the author nor to Marx but, rather, to capitalism. To be sure, if Marxism is all about revolutionary politics, then abstract theory (as opposed to concrete practice) seems to be a strange point of departure. Turning this common sense upside down, abstract theory is a rare and valuable starting point for it allows us to reconstruct capitalism in thought, to see it from the inside as if from the outside. It is a way of seeing the wood for the trees, a mode of theorizing that makes the opaque transparent. As such, Marxian theory is political through and through.

Structural Marxism and Marxist Human Geography

Before all else, Marxism is a materialist philosophy. In the oft quoted 1859 preface to *A Contribution to the Critique of Political Economy*, Marx argued that it is not that consciousness determines social existence, but that social existence determines consciousness. In other words, it is an ontological fact that human beings are first and foremost material animals. It is in virtue of the sort of bodies we have that the realm of production and reproduction sets limits to and exerts pressures on the domain of ideas, thoughts, and feelings. Abstracting the particularities of history and the specificities of geography (i.e., actually existing time spaces) out of focus, Marx reminds us that society is determined, in the last instance, by the economic. In the final analysis, or, in the long run, the mode of production of material life represents the universal condition of possibility for any and all other happenings in the world. After all, the dour world of social reproduction (food, clothing, housing, or whatever) is the primary matter of concern for most people, most of the time.

However, the dominant mode of production – namely, capitalism – is making and remaking a world, it seems, that is an unjust time–space of freedom for the few (emancipation from basic needs and wants) and necessity for the many (immiseration in spite of technoscientific advances). Moreover, the few's emancipation is predicated upon the many's immiseration; as the human geographer Doreen Massey argues, we ought to develop a "global sense of the local," a "progressive sense of place," for progress for one people in one place all too often means regress for another people in another place. For example, think about the socioecological network of bodies and things, processes and relations, set in motion each and every time you exchange money for some good or service. The myriad unknown (if not, perhaps, unknowable) costs and benefits for different people and places – (great) costs and (small) benefits for which the consumer is, in part, responsible – are masked or veiled by the commodity form itself, and thus our own world making complicity and compliance goes unacknowledged. To be sure, to interpret the world is not enough, we badly need change. Marx envisioned the present as humanity's 'prehistory', its 'pre geography', for another world is not only possible but also essential given the reality of capitalism – wealth, uneven development, consumerism, poverty, and all – at the beginning of the twenty first century.

Besides economic determination in the last instance, two concepts are central to structural Marxism. The first is the idea of structure and agency. Marx once said that men and women make their own history (and, of course, geography), but not of their own will or intention; not under conditions of their own choosing, but under the given conditions with which they are confronted. In other words, our individual freedom and liberty to make and remake the world is illusory, and yet by no means are we docile dupes dancing to the tune of a ruling class. On the one hand, agency (the capacity or potential to act) always exists in the presence of the structures that constrain and also enable it. On the other, structure (ontologically real but nonobservable processes and relations) never exists in the absence of the agencies that reproduce and also transform it. The first point is that structures exist outside of and prior to us. Social relations are more or less spatially and temporally durable, and thus the processes constitutive of society stand before and above us as 'ready mades'. It follows, then, that structures are, in the last instance, all about coercion rather than consent in that we cannot 'be in' and 'become through' some or other social relations. To see and to say, to make and to do, and so on we need myriad resources – the very social necessary conditions for life itself. That said, and this is the second point, structures are thoroughly dependent upon the activities and conceptualizations of people. Social relations exist not despite, but precisely because of what we are and what we do as creative social beings. As Roy Bhaskar (a philosopher true to the spirit if not the letter of Marxism) argues, the form and content of so ciety may be something people like us are not personally responsible for, but it is just so in virtue of our activities and conceptualizations. For example, if we do not go to work each and every day to create the capitalist economy, then neither do most people, most of the time, go con sciously to sustain (nevermind undermine) it. Never theless, the capitalist economy is at once the necessary condition for, and unintended consequence of, work.

The second concept central to structural Marxism is the idea of totality. For Marx, capitalism is more than the economy understood as workaday laborers, machines, and natures. It is a complex social whole – a 'whole way of life'. In social formations such as ours, the economic 'base' of so called productive forces and social relations of production determines an extra economic (political, ideological, etc.) 'superstructure' of institutions and knowledges, representations, imaginaries, etc., necessary but not sufficient for social reproduction. The economic determines the political (the state, the legal system, etc.) and the ideological (the education system, the family, etc.) in the sense of setting limits to and exerting pressures on their development. The base explains not the genesis but the evolution of the superstructure; the former represents the latter's historically and geo graphically specific conditions of possibility; it conditions and shapes it in particular (but not total) ways. To sim plify rather, thinking in terms of a totality means thinking in terms of the relationships between the seemingly separate – seeing the economic as inextricably 'in' the political, ideological, etc., and, importantly, vice versa. Interactions, connections, and dependencies, rather than discrete, bounded things, are – as the human geographer Erik Swyngedouw argues in this encyclopedia – the bread and butter of Marxism, a dialectical mode of theorizing. Think, for example, of how the leftist slogan "think global, act local" developed out of a growing understanding of the myriad forces, relations, and knowledges both attached to place and reaching out across space that constitute globalization. As localities become more and more interconnected and reciprocally dependent, and 'time–space compression' (the term is Harvey's) and the creative destruction of place based meanings, values, expectations, etc., proceeds apace, a dialectical mode of theorizing becomes more than ever pertinent.

Althusser's Marxism

Against immodest economism and overconfident humanism

Outside human geography, structural Marxism is today primarily (but not exclusively) associated with the work of Althusser. Nothing if not controversial (and more than a little obscure), Althusser burst onto the Anglophone scene in the late 1960s and early 1970s with the trans lation and publication of two groundbreaking works – namely, *For Marx* and, with his student Etienne Balibar, *Reading Capital.* Althusserian Marxism pivots around the ideas of structure and agency and totality, representing an extended critique of two uses of the concepts popular in 1950s and 1960s France.

First, before the Russian premier Nikita Khrushchev's revelations about and denunciations of Stalinism at the Twentieth Congress of the Communist Party of the So viet Union, French Communist Party philosophers like Althusser were expected to walk the party line and swallow vulgar Stalinist dogma hook, line, and sinker. After Khrushchev's great speech, Althusser sought to distance Marxism from the one sided economism of crude doctrine. Economism represents a misunder standing of Marx's idea of totality – the reduction of society as a complex social whole to the economy 'pure and simple'. The political and the ideological, for ex ample, are taken to be mere epiphenomena or simple effects; produced but not productive. The economy is seen as the one social force driving the making of history and geography, and revolution, the transition from cap italism to communism, is seen as inevitable. Economism takes the contradiction between the inherently dynamic productive forces and the inherently static social re lations of production to be the immutable law of motion for capitalist societies, operating behind the backs of people. An insatiable, profit driven mode of production, in capitalism competition between capitalists forces them to continually make and remake laborers, machines, and natures to work harder, faster, and better (and this is the point) no matter what the socioecological consequences. The crisis tendencies of capitalism, so the thesis goes, will invariably extensify and intensify as its internal processes and relations not only support but also undermine one another.

Second, Marxism in post war France had been influ enced by Maurice Merleau Ponty's phenomenological and Jean Paul Sartre's existentialist philosophies. Brush ing orthodox Marxism against the grain, Merleau Ponty and (especially) Sartre insisted that people in capitalist societies are not passive objects, caught in a trap of their own making, but active subjects, able (if not yet willing) to consciously remake the world. Rather than the 'iron laws' of capitalism, phenomenology and existentialism emphasized the creative behaviors, lived experiences, and consciousness of human beings – their innate, actually existing capabilities for 'praxis' (conceptualization and activity). In the wake of phenomenology and existen tialism, Althusser sought to distance Marxism from the subjectivist humanism of Merleau Ponty and Sartre. This humanism (it is, of course, inaccurate to talk about humanism in the singular rather than humanisms in the plural; there is, as we shall see, much variety both within and beyond Marxism) represents a misunderstanding of Marx's idea of structure and agency, as an emphasis on the capacities and potentialities of men and women invariably descends into an overemphasis. Focusing on the ephemerality and volatility (agency) of real individuals on the ground, rather than the fixity and permanence (structure) of capitalist processes and relations, these humanists argue that subjective will and intention is more important, in the final analysis, than objective social forces shaping and conditioning each and every choice we do and do not make. Merleau Ponty's and Sartre's subjectivist humanism, in sum, takes the movement from the capitalist present to the post capitalist future to be of our human nature, and thus inevitable.

Althusser sought to craft a Marxism that rejected both faith in a single social force, an immutable law of motion, and hope in peoples' innate, actually existing capabilities for praxis. In other words, he sought to craft a Marxism that avoided dangerously optimistic notions of inevitable revolutionary social change. This is not to say that Althusser was a pessimist. Far from it, he held Marxian philosophy to be a guiding thread key to socialist strategy. Marxian philosophy struggles to not only interpret the world but also to change it in the sense of transforming both how we participate in as well as know it. For Althusser, economism and humanism were poor 'cognitive maps' of what actors in societies such as ours have to cope with and respond to. They are ideological inasmuch as they represent what he calls the imaginary relationship of individuals to their real conditions of existence. As conceptual apparatuses, their silences are symptomatic, signifying their absences and exclusions. Economism's immodesty blinds it to the myriad other (political, ideological, etc.) structures that are constitutive of as well as constituted by the economic in all manner of ways. Humanism's overconfidence blinds it to the less than obvious objective determinants constraining and enabling action and belief in capitalist societies.

Before moving on, it is worth noting that Althusserian Marxism was undoubtedly influenced by, but is not the same as, structuralism in the sense of the work of Ferdinand de Saussure in linguistics, Claude Levi Strauss in anthropology, and Jacques Lacan in psychoanalysis, among others. To simplify rather, Saussure and fellow travelers sought to deconstruct taken for granted surface appearances, and in so doing illuminate underlying realities. What lies beneath are logics or structures through which apparently discrete, bounded beings in the world are positioned relative to, and situated in relation to, one another. For example, Saussure studied language, Levi Strauss studied culture, and Lacan studied the unconscious as systems in which seemingly separate objects are, in reality, interconnected and reciprocally dependent.

A cognitive map of where and when to strike back

Althusser sought to make sense of what, exactly, Marx meant by totality, on the one hand, and structure and agency, on the other. First, society as a complex whole is dynamic rather than static – never made, always in the making. Structures ought to be understood dialectically – not as 'things', but as processes and relations. In other words, the economic, the political, the ideological, etc., exist neither outside of nor prior to the processes and relations that create, maintain, or destroy them. Seemingly absolute, structures are, in reality, relative and relational in nature – open rather than closed, internally heterogeneous rather than homogeneous. Hence, they are essentially contradictory; internal relations simultaneously support and undermine one another, and thus change is very much a part of what structures are. For example, think of yourself as a contradictory social whole of identities – class, gender, 'race', and so on – that come together in and through you. You are not a product of your class, gender, or 'race' alone, or, to put it another way, you are 'underdetermined' by your class, gender, or 'race' alone. A single structure does not determine absolutely who you are; indeed, it cannot for a single structure is but an analytical distinction (i.e., class is not class without gender, 'race', and the like).

Rather than thinking dualistically in terms of the determining and the determined, the independent and the dependent, or cause and effect, one should think in terms of 'overdetermination'. Each structure internalizes every other structure in a social formation. That is, each structure is (over)determined and (over)determining vis à vis every other structure or, each structure is overdetermined by and plays a part in the overdetermination of every other structure. The whole is 'in' the parts and the parts are 'in' the whole, and thus society is not simple (an expression or reflection of the dominant contradiction of the forces and relations of production only), but is complex. The political, the ideological, etc., are constitutive of as well as constituted by the economic – they are interconnected and reciprocally dependent. Each structure, then, has a certain kind of effectivity (a specific effectivity) and a certain degree of autonomy (a relative autonomy). Simply put, society as a complex whole is 'decentered' (that is, without essence). Marx, for example, referred to it as a unity of the diverse – a concentration of not one but many determinants.

Second, Althusser looked to V. I. Lenin's and Mao Tse tung's writings on the Russian and Chinese revolutions respectively for clues on how to put the idea of overdetermination to work, intellectually and politically speaking. He argued that historically and geographically specific analyses of the articulations, dearticulations, and rearticulations of the structures (economic, political, ideological, etc.) that constitute capitalism can identify moments when a contradictory social whole comes together ('condenses', 'fuses'), forming a so called 'ruptural unity'. Interconnected and reciprocally dependent, different structures in different places at different times unfold and spin out in particular ways. For example, the plural rather than singular nature of your identity explains much of your creative behavior – it is your hybridity rather than purity that is at the heart of your becoming as a human being. For Althusser, it is at moments of ruptural unity, when serious frictional problems associated with the interactions within and between the economic, the political, and the ideological emerge, that the conditions of existence (in other words, opportunity) for revolutionary social change crystallize.

The interactions of economic, political, ideological, etc., structures occasionally come together in a moment charged with radical potential; a moment where and when working men and women can and should heed Marx and Engel's call in *The Communist Manifesto* of 1848 to unite (the proletarians, so the saying goes, have nothing to lose but their chains). The lesson Althusser learnt from Lenin and Mao was that revolutionary social change is all about not only being prepared to change the world, but also being in the right place at the right time. Revolution, then, is at once a happening irreducible to the agency of a single social group (*contra* humanism's faith in the working class) and irreducible to a single independent causal mechanism (*contra* economism's hope in the economic). The economy is, perhaps, determinate in the long run, yet it is definitely both overdetermined as well as overdetermining in the here and now. Similarly, class struggle is, perhaps, the motor of historical and geographical change, yet it is definitely causally efficacious only at certain times in certain places. History and geography, then, are processes with neither a *telos* (logic or rationality) nor a subject.

Marxist Human Geography

Inside geography, Althusser's Marxism has never been big. For example, Harvey's groundbreaking *Social Justice and the City* – an early classic in Marxist human geography – largely ignores the contribution of Althusser to Marxism after Marx. The urban theorist Manuel Castells' pioneering *The Urban Question* (as well as the work of Massey, among others, in the 1970s), is the exception that proves the rule, putting Althusserian ideas to work systematically. That said, and the details of yesterday's Althusserian geography notwithstanding (the 1980s saw both Castells and Massey moving on from Althusser), books like Harvey's germinal *The Limits to Capital* and Smith's seminal *Uneven Development* – books that have been and remain foundational – are arguably true to the spirit if not the letter of Althusser's Marxism. They are works of structural Marxism broadly defined, or in a strictly improper sense, only.

Without getting lost in the details, Harvey's and Smith's Marxism exhibits some important family resemblances with structural Marxism proper. In general, both take Marx's mature works such as the *Grundrisse* of 1857 and *Capital* of 1867 to be suggestive of a way of seeing the signals in the noise, a mode of theorizing that makes the invisible visible. The intellectual labor that is envisioning capitalism theoretically is taken to be essential for the political labor of dismantling and overthrowing it practically. Marxism is, then, to borrow Althusser's words, a revolutionary weapon – class struggle in theory. More specifically, like Althusser, Harvey and Smith embody a historical–geographical materialist sensibility. Revolutionary social change is a matter neither purely of the agency of a single social group nor simply of a single independent causal mechanism. Pace humanism and economism, human agency must encounter opportunity – the historically and geographically specific economic, political, ideological, etc., crisis tendencies of and contradictions in capitalism as a totality. It is neither structure nor agency in and of itself that drives historical and geographical change; like Marx, we should think dialectically in terms of 'both/and' rather than 'either/or'.

This is not to say that Harvey's Marxism, for example, lacks distinction. Far from it, as the human geographer Noel Castree argues. (1) It is a 'classical' Marxism that refines and extends Marx's original work rather than the work of Marxists such as Lukács and Marcuse, or, like Althusser, non Marxists such as Bachelard and Spinoza; (2) it is less philosophical than scientific, more about explaining the historical geography of capital accumulation than about clarifying, like Althusser, the ontological and epistemological principles of Marx's substantive analyses; (3) it is 'relational' in ways Althusserian Marxism is not. It tells us, for example, something about the relationships between society and nature, time and space, etc., insofar as it understands these things to be unified by the processes and relations of capitalism; (4) it argues for a 'geographical imagination' that, pace Marx, takes seriously the limits and the possibilities that the material geographies of production, distribution, exchange, and consumption represent for the survival of capitalism; (5) it restlessly illuminates capitalism's being in the world. It was there in the eighteenth and nineteenth centuries, and, lest we forget in these 'after Marxist' times, it is here in the twenty first century; (6) books like *The Urbanization of Capital* and, especially, *Consciousness and the Urban Experience* ground Harvey's Marxism somewhat (making it less 'theoreticist' than Althusser's) as he puts abstract concepts and categories to work explaining concrete conjunctures; (7) it is in some ways very much accessible. Recent books on imperialism and neoliberalism comprehend and condemn present geopolitical and geoeconomic realities plainly, speaking to audiences beyond the 'ivory tower'; (8) it is revo lutionary theory through and through. Pivoting around a notion of dissensus, it is, not unlike Althusser's, am bitiously within and against a whole way of life; (9) it is most unusually consistent. Whereas an older Althusser proclaimed a crisis not only 'in' but also 'of' Marxism, Harvey tirelessly argues that Marxism is more than ever relevant; (10) it is important to note that whereas the language of structural Marxism served Althusser well in the 1960s and 1970s, it is the language of dialectical Marxism that Harvey mobilizes in today's politico in tellectual environment (on which more anon); and (11) like Althusser, Harvey is a researcher and a teacher in universities. The point is that the movement from theory to practice, from the Marxist intelligentsia to popular agitation and organization, has never been and will never be easy.

Readers could be forgiven for thinking at this point that if structural Marxism proper has played only a minor role in the development of Marxist human geography over the last 30 or so years, then why revisit *For Marx* and *Reading Capital*? In the first instance, it is because critics of Althusserian Marxism have significantly shaped the de bate on and around today's Marxist human geography. Many of the more well known criticisms of Marxist human geography turn on the family resemblances be tween Althusser and geographers like Harvey, Smith, Richard Walker, and Richard Peet, among others. Criti cisms of Althusser penned outside geography have long been imported, warts and all, into the discipline, and used as ammunition against leading figures such as Harvey *et al.* Hence, a basic understanding of the principles of structural Marxism is needed if one is to make sense of the last three decades of debate inside geography.

Structural Marxism and Its Critics

Presented as the gospel of Marx according to Louis, Althusser's clinical and uncompromising Marxism has, unsurprisingly, been neither universally nor uncritically embraced. Two critiques stand out, the first of which concerns the question of agency and class, and the second of which concerns the question of structure and capit alism. First, it has been argued that Althusser rudely shunted aside human agency in conceptualizing capital ism as a totality of interacting social processes and re lations. This begs the question of whether the working class makes and remakes itself in spite of capital as much as it is made and remade by and for capital. Second, it has been argued that if structures are always already both overdetermined as well as overdetermining, then the economic cannot be determinate in the last instance – after all, there is no such thing as the economy in and of itself. This then begs the question of whether capitalism as a mode of production is more invention of the theorist than historical–geographical reality. The problem, for some critics, is that Althusser's Marxism at once goes too far (from humanism) and not far enough (from econo mism), and in so doing produces, in thought, something that does not exist in reality – capitalism and class. To put it another way, the problem is all about how and with what consequences capitalism and class are represented in and through the academic work of structural Marxists.

On the one side, the Marxist historian E. P. Thompson penned an (in)famous polemic against Althusser's rep resentation of class in 1978 (which has proven to be in fluential among some geographers seeking to bring people's 'life worlds' (their creative behaviors, lived ex periences, and consciousness) back in to the discipline). On the other side, the Marxian political economists Stephen Resnick and Richard Wolff penned an important reappraisal of Althusser's representation of capitalism in 1987 (which has proven to be influential among some geographers seeking to make sense of an actually existing more than capitalist world). These criticisms will be unpacked below.

Structural Marxism beyond Human Geography

Thompson: O agency, where art thou?

Thompson's thesis is long and, at times, hard to follow, yet the question of agency continually surfaces; it is the glue that holds the argument together. Thompson the historian is suspicious of Althusser the theorist. For Thompson, categories such as class tend to become rigid in the hands of self styled philosophers and scientists. Unable and/or unwilling to get their hands dirty and 'mess up' their theories with historical evidence, writers like Althusser conceal at least as much as they reveal. Thompson argues that class understood as the social relations of capitalist production is never anything more than a generalization that cannot capture the infinite complexity at large in the world. Wonderfully compli cated and superbly obstinate, real individuals always hold out the possibility of surprising the analyst. General izations plunge actually existing capitalism and its con tingencies (i.e., its ceaselessly changing nature) into darkness, while historical evidence of the thoughts, feelings, loyalties, (in)actions, and the like of real indi viduals casts light upon it. Thompson puts flesh on the bones of this thesis by writing off Althusser's idea of history and geography as a process without a subject as ridiculous. That individual men and women are but 'bearers' or 'supports' for interacting structures is taken to be self evidently absurd. Drawing on his earlier, now classic, work, *The Making of the English Working Class*, Thompson argues that the working class (his case study is eighteenth and nineteenth century England) made itself as much as it was made, that class and history and geography as process is less about inhuman structures and more about human experience and consciousness.

We are, Thompson insists, the voluntary agents of our own involuntary determinations. Rather than a mere bearer or simple support for economic, political, and ideological structures, class is something that happens in particular places, at particular times. In other words, class is not; it becomes. Given a common experience in historically and geographically specific social relations of production, some people feel and articulate particular identities and interests in relation to others, and thus class consciousness emerges. Historical and geographical change is not driven by the condensation or fusion of contradictory structures into a ruptural unity behind the backs of the 'pawns' on the ground, but is a process driven by class struggle – the continuous ascendancy and descendency of different sociopolitical forces (think, e.g., of the 1984 miners' strike in the UK and the 1999 battle in Seattle in the USA). Given certain productive relations, people experience exploitation, identify antagonistic interests, struggle (think, value, act) around these issues, and in so doing they achieve consciousness – they become a class. For Thompson, then, class is consciousness, and consciousness is struggle. In terms of what is to be done, Thomson argues that Marxists should set out to map the islands of worker resistance in a sea of capitalist power.

Resnick and Wolff: It's capitalism, but not as we knew it

Resnick and Wolff's thesis is nothing if not bold, yet much more sympathetic than Thompson's. Rather than reproduce Althusser's work verbatim, Resnick and Wolff seek to refine and extend it, for an ambiguity, or, better, hesitancy runs like a red line through it. Plainly, the primacy or privilege of capitalism has been and remains a thorn in the side of Marxism after Marx; economic determination in the last instance sits uneasily alongside the idea of overdetermination. The former implies the closedness of an 'impermeable', 'pure' capitalism, and the latter implies the openness of a 'porous', 'hybrid' capitalism. Plainly, if, on one level, capitalism is a totality of economic, political, and ideological structures where each is overdetermined by and participates in the overdetermination of every other, then, on another, it itself must be overdetermined and overdetermining. The logic that there can be no necessary cause, no cause of fundamental significance, precisely because there can be no independent determinant holds both within capitalism and between capitalism and other social wholes.

In other words, given that capitalism is always already and everywhere more than economic, it follows logically that social life has never been and will never be 'essentially' capitalist. What we are dealing with here, then, is not a singular and global cause (namely, Althusser's capitalism) but plural and local effects – multiple modes of production that are at least as much noncapitalist (think, e.g., both of slavery and feudalism, as well as of the informal sector, cooperatives, etc., in the world today) as capitalist. Simply put, all noncapitalist 'others', 'exteriors' and 'outsides' are implicated in each capitalist 'self', 'interior' and 'inside', and thus Resnick and Wolff's antiessentialist Marxism is all about the ceaseless exploration and elaboration of how the former constitute the latter. Hence, rather than something that explains, capitalism is taken as something to be explained.

Structural Marxism within Human Geography

Duncan and Ley: Docile working class bodies?

Drawing on Thompson's work, the geographers James Duncan and David Ley published a humanist critique of structural Marxism and human geography in 1982 (see also the work of Marwyn Samuels, Yi Fu Tuan, David Seamon, and J. Nicholas Entrikin, among others). Arguing that human geographers like Harvey, Smith, Walker, and Peet can be defined broadly as structural Marxists, Duncan and Ley insist that Harvey *et al.*'s work does not represent a radical 'epistemological break' with Althusser's apparently Hegelian–Marxist way of seeing the world. There is some debate in Marxist circles regarding the extent to which Marx was influenced by the idealist philosopher G. W. F. Hegel. Hegel argued that *Geist* ('reason', 'the idea', or 'world spirit') is transcendental in that it is 'in' everybody and everything – it is actualized through any and all material objects in the world. Hegel insisted that the surface appearance of chaos that is everyday life is merely the cunning of reason, for an underlying reality is causally efficacious in the apparent confusion of bodies and things. *Geist*, then, is a formal (true, active) cause and human beings are merely an efficient (false, passive) cause, and in spite of what appears to be chaos and confusion, it unfolds through contradictions toward its predetermined goal.

For Duncan and Ley, Marxist human geographers represent capital as a formal cause and people as an efficient cause. Capital replaces the idea or world spirit as the mysterious force flowing in and through us as it unstoppably spins out toward a given end (namely, communism). People, then, are reduced to means – agents in, but not of, history and geography. Conceptualizing the whole as more than the sum of its parts, Marxist human geographers attribute power, activity, and, it seems, will and intention to capitalism, and only passivity and little, if any, choice to human beings. This is little more than functionalism – people exist precisely because they fulfill the needs of the system, and they fulfill the needs of the system precisely because they exist. For Duncan and Ley, Marxism amounts to the argument that people exist because they are functional and they are functional because they exist.

Gibson–Graham: The end of capitalism

Drawing on Resnick and Wolff's work, the geographers Katherine Gibson and Julie Graham (writing as J. K. Gibson–Graham) published an antiessentialist critique of Marxist human geography in 1996 (see also the work of Andrew Leyshon, Colin Williams, Roger Lee, and Kevin St. Martin among others). Gibson–Graham seek to envision capitalism not as a monstrous unity but as a network of localities – something not only constitutive of but also constituted by myriad others, exteriors, or outsides. Gibson–Graham wish to remap the relationship between capitalism and noncapitalism, not in terms of power on the one side and resistance on the other, but in terms of hybridity; thinking of capitalism as a sort of 'cyborg' that is always more than capitalist, never pure. What makes their thesis novel is that it troubles the first rule of structural Marxism: if one is willing to change capitalism, then one must be able to understand it – to see it as something structural, global, and consequential. Such representations, Gibson–Graham argue, are performative in that they construct (in part) the anticapitalist imagination. Social movement activism depends largely on what activists think is and is not possible, and what they think is and is not possible in the current conjuncture depends largely on what they imagine capitalism to be and not to be. If capitalism is always colonizing/penetrating/expanding and never colonized/penetrated/contracting, then revolutionary social change is largely unimaginable and, as such, is doomed to inertia.

Marxist human geographers take concepts like capitalism and class for granted and in so doing represent them in thought as less open and more homogeneous than they are in reality. Hence Gibson–Graham seek to deconstruct the metanarrative of big C Capitalism (a 'driverless juggernaut' or 'Leviathan') and see in its place the narratives of little c capitalisms (as much constructed by as constructive of their others – slavery, feudalism, the informal sector, cooperatives, etc.). How we think and talk about capitalism (our discourses of it) really matters for it all too often obscures capitalism's 'constitutive outsides'. Rather than empowering discourses of both capitalism and noncapitalism (entangled in myriad ways), of diverse economies and economic difference, we have so called 'capitalocentric' discourses of either capitalism or noncapitalism where the former stands before and above the latter. The problem with discourses of the capitalist economy and economic sameness is that they are very much disempowering; they appear to verify Mrs. Thatcher's maxim – "there is no alternative," capitalism is the end of history and geography. Radical scholarship, surely, is all about falsifying it, and demonstrating that another world is not only possible, but also actual. In sum, there is no capitalist core or center that is independent of its noncapitalist peripheries or margins. The latter existing conditions are the former's conditions of existence, and thus capitalism is at once less (homogeneous) and more (open) than Marxist human geographers see.

Marxian Counter-Critiques

Doing Dialectics: Not Either/or but Both/and

The editor of *New Left Review*, Perry Anderson, responded in a measured fashion to Thompson's polemic in 1980. Anderson's argument is as long and, at times, as hard to follow as Thompson's, yet it is at heart devastatingly simple. Thompson's insistence on the parity between voluntary agents and involuntary determinations is a hypothesis that is never tested. Thompson sets up something of a dualism, with working class suffering and indignity on the one side and struggle and consciousness on the other. The problem, it seems to Anderson, is that his account of immiseration is in no way equivalent to an account, like Althusser's, of the objective determinants that set the limits and exert the pressures constraining and enabling working class behavior in particular historical–geographical contexts. To be sure, Thompson talks about shaping and conditioning in opposition to human agency – necessity in opposition to freedom – and unproblematically judges them to be equal (the working class made itself as much as it was made, etc.). However, the rich Althusserian notion of a totality of determinants nowhere plays a part in his analysis, and thus his judgment remains, in the final analysis, unqualified. What is to be done, then, is not less but more conceptualization. Thompson's account of class consciousness and struggle in eighteenth and nineteenth century England may be the exception to the rule – most people, most of the time, neither articulate nor feel class identities and interests, for the crisis tendencies of and contradictions in a social formation condense/fuse in time–space only occasionally. Hence the need for theoretical analyses in which to situate historical evidence of class formation.

For Anderson, Thompson's weaknesses are Althusser's strengths, and vice versa. The former is weak on how and with what consequences people set their given conditions of existence in stone, and the latter is weak on the messy, indeterminate struggles to make history and geography that people engage in. Thompson's lucid accounts of the shared sense of strains, grievances, obligations, entitlements, etc., that develop within and between working class communities have proven to be paradigmatic. To be sure, Thompson and others (especially Richard Hoggart, Raymond Williams, and Stuart Hall) were central to the movement on the Left to take seriously – following the Marxian theorist Antonio Gramsci – the idea of culture, and the attitudes and identities of ordinary men and women, in relation to revolutionary social change. However, without theoretical analyses of interacting structures framing it, a rich cultural materialism (the term is Williams') becomes a poor 'culturalist' voluntarism as troubling as determinism. In other words, like structural Marxism, cultural studies is a knife edge path with high theory on the one side and the micro politics of everyday life on the other, and it is – as the human geographers Derek Gregory (writing about Thompson in 1981) and Nigel Thrift (writing about Williams in 1983) tell us – all too easy to fall into the abyss now and then. (On the possibilities and the limits of cultural materialism in human geography, see also Peter Jackson's wonderful *Maps of meaning* and Don Mitchell's superb *Cultural geography*.) In sum, then, Althusser's Marxism is all about opening up not closing down space for class struggle. The Marxist dialectician Bertell Ollman injects some much needed clarity into the debate when he argues that Althusser and detractors such as Thompson work at different levels of abstraction. When Althusser abstracts political–economic structures into focus, he necessarily abstracts the cultural life of people out of focus. Similarly, when Thompson makes visible real individuals, he necessarily makes invisible political–economic structures. In other words, both Althusser's and Thompson's theses are necessary but not sufficient, and thus something of an 'academic division of labor' is set up. In the flux and flow of elegant and forceful writing, though, cooperation all too easily descends into conflict, and productive epistemological conversations and political solidarity are, sadly, lost.

In the late 1980s, Marxist human geographers like Walker and Peet cited Anderson's counter critique approvingly, with Smith wisely noting that properly political academic practice is all about taking seriously both the limits as well as the possibilities of human agency. It is, he insists, one thing to investigate the structural and the structured, and another thing altogether to be "structuralist." As Althusser once said, to be structuralist is to be ideological – to conceal at least as much as one reveals (and this is the point) in non self conscious or unknowing ways. To be sure, far from a so called 'totalizing' discourse that closes down the possibility of people struggling for radical historical–geographical change, Marxism opens up the possibility by enabling people to find their bearings in the topsy turvy universe of capitalism. However, to place center stage what actors in a society such as ours have to cope with and respond to is to place beyond the pale how, exactly, different actors in different places at different times cope and respond in progressive and/or regressive ways. What Althusser called the "detour of theory" and what Harvey calls the "theoretical imperative" is, then, necessary but not sufficient – the costs (concealing the contingencies that are the lived experiences of suffering and indignity as well as struggle and consciousness) must be weighed against the benefits (revealing the necessities that are the tendencies of the capitalist mode of production). To be sure, keeping these two viewpoints (not mutually exclusive but complementary) in creative dialectical tension is no mean feat – it is all too easy to improperly mediate between one way of seeing and the other (as Harvey noted in an exegesis of Marx's method in 1989). However, Thompson's and Duncan and Ley's intemperate critiques are very much de as opposed to reconstructive. Althusser and Harvey *et al.*, respectively are seen to be simply wrong rather than self consciously or knowingly leaning to one side of the dialectic rather than the other. This is clearly one case of mistaken identity.

Envisioning Capitalism, Making Connections

Another case of mistaken identity is Gibson–Graham's antiessentialist critique of Harvey *et al.* – a critique that in many ways crowns a canon (of considerable weight) of postmodernist, post structuralist, and feminist writing critical of Marxist human geographers' apparently structuralist ways of seeing the world. For example, true to the spirit (if not, perhaps, the letter) of Resnick and Wolff's earlier and Gibson–Graham's later antiessentialist critiques, the feminist theorist Rosalyn Deutsche's attack in 1991 on Harvey's *The Condition of Postmodernity* for eschewing noncapitalism as a matter of concern is a case in point. Harvey's 'grand theory' of social life – of the underlying reality of capitalism beneath the surface appearances of lived experiences – is labeled authoritarian, disinterested, and disembodied for he seemingly performs the 'god trick' (to appropriate the socialist feminist Donna Haraway's felicitous phrase) of seeing everybody and everything from nowhere. Harvey is authoritative precisely because he is unattached from the world, seeing society as essentially capitalist from the perspective of a free floating third person (from 'above', 'up there') rather than the perspective of a real individual on the ground (from 'below', 'down here'). Far from unproblematic, this sanitizing, exclusionary gaze and its exorbitant knowledge claims is diagnosed as symptomatic of a modernist–masculinist 'will to power' – a need to tame, and thus master, the intractable world. The world is represented, it seems, as it really is, or, as in a mirror, without reference to the 'positionality' of the observer. Feminists and others take it to be self evident that the observer's location in social space affects what they do and do not see, and thus that knowledges are always already 'situated'. The known (capitalism) becomes something global – a monstrous unity – only because the knower's (Harvey's) partial perspective as a middle class white male is denied. The question, then, for Deutsche, is "whose reality counts?", for social life is seen to be all about capitalism, and thus the lived experience of difference (sexism, racism, production and reproduction in the home, etc.) is reduced to something irrelevant (a so called 'false consciousness'), rather than something as much constitutive of as constituted by capitalism.

Arguably, no human geographer has done more to engage with Marxism's critics than Harvey. Always at the cutting politico intellectual edge, Harvey it seems anticipated antiessentialist critiques in 1982 when he made clear the unpretentiousness of Marxian theory. It is not about explaining everybody and everything that is and is not, that has ever been and will ever be, but is about crafting a framework for understanding, or, a conceptual apparatus for making sense of, a few significant processes and relations at large in the world. The question of significance is the crux of the matter. Following Haraway, Harvey noted in a rejoinder to Deutsche in 1992 that epistemology is about knowing the difference between more significant and less significant differences. There are multiple significant differences in the world that bind human beings together, including class, gender and 'race', and these articulate in many different ways in many different places. It is, on the one hand, perfectly right and good to argue that significant differences articulate in different ways in different places to produce uniqueness (Manchester is not Mumbai is not Melbourne is not Minneapolis, etc.). On the other hand, it is more than a little problematic to see uniqueness as absolute otherness. In other words, we have certain differences in common in the world today, and thus it is not uniqueness in and of itself but the dialectic of commonality and difference that ought to be taken as a matter of concern. Antiessentialists like Gibson–Graham arguably bend the stick too far back from seeing commonality 'in' difference (capitalism here, there and everywhere) to seeing irreducibly specific localities only (the end of capitalism (as we knew it)).

The insistence on locality's sensuous specificity is not only intellectually but also politically problematic. In response to an emphasis on how and with what consequences different people in different places have certain differences in common, Gibson–Graham shade toward an emphasis on the radical difference that marks each and every being in the world. To emphasize, so the saying goes, is to overemphasize, and Gibson–Graham's reply to Marxist human geographers is, to be sure, polemical in the best sense – self consciously. However, it is one thing to argue that capitalism (and, indeed, class) is more plurality than singularity, and another thing altogether to argue this at the expense of all 'capitalism talk'. The problem is that if we cannot make strong claims about the nature of a few significant processes and relations such as capitalism, then we know more about what it is that divides us than what it is that unites us, and thus alliance formation between localities becomes virtually impossible. For example, a struggle for indigenous rights in Mexico, a struggle against dam construction in India, and a struggle over labor law in France have little in common on the surface – unique places seemingly give rise to unique problems. The underlying reality, though, is that all three struggles are anticapitalist through and through, and thus could be strengthened (rather than weakened) by the ability to see capitalist processes and relations in otherwise unique contexts, as Harvey argues in both *The New Imperialism* and *A Brief History of Neoliberalism*.

If reality is the unity of the diverse (a concentration of not one but many determinants), then the power of abstraction must be put to work in order to make sense of infinite complexity. In other words, it is precisely because, in practice, social life is underdetermined by capitalism (it does not exist in a pure state) that it is intellectually and politically useful to sketch out, in theory, its 'essential' characteristics. Abstractions produced in and through historical–geographical materialist analyses are deployed to identify commonalities within and between different people in different places, their other differences notwithstanding. Abstractions serve to isolate significant differences constitutive of (and, of course, constituted by) unique places. Hence, rather than thinking about unique places as irreducibly particular localities marked by absolute otherness, Marxist human geographers like Harvey envision them as 'permanences' – more or less stable articulations of a few significant processes and relations. To isolate one significant difference but not another is, then, a tactical move. To foreground one and background another is but a moment in a scientific exploration, and not a reflection or expression of the analyst's modernist–masculinist will to power, blindness to positionality and situated knowledges, or whatever.

Before moving on, it is worth noting that the structure agency and commonality difference questions have been troubling not just for Marxists. With great elan, human geographers such as Gregory (in his *Regional Transformation and Industrial Revolution*) and Allan Pred (in his *Place, Practice and Structure*) looked to social theorists like Pierre Bourdieu and Anthony Giddens in the 1970s and 1980s for a way out of the problems, not in terms of levels of abstraction or creative dialectical tension, but in terms of the 'duality of structure'. Giddens' structuration theory, as explained in *The Constitution of Society* and elsewhere, holds that structures are both the medium and the outcome of social practices, while Bourdieu's theory of practice, as explained in *Outline of a Theory of Practice* and elsewhere, holds that structures are both structuring of and structured in social practices. In other words, global structures and commonalities are both produced and productive vis à vis local agencies and differences; the global is 'in' the local, and vice versa.

Conclusion

This article proceeded in three sections. First, it presented an introduction to structural Marxism in general (focusing especially on the work of the Marxist philosopher Althusser), and Marxist human geography more specifically. Marxism is a materialist philosophy that takes the dominant mode of production of material life to be both unsettled and unsettling. Turning on the concepts of structure and agency and totality, and rejecting both economism and humanism as poor cognitive maps for finding our bearings in today's topsy turvy world, structural Marxism represents the intellectual labor necessary but not sufficient for the political labor of dismantling and overthrowing capitalism. While Althusserian Marxism has played only a minor direct role in the development of Marxist human geography, it has played a major indirect role in that its critics have significantly shaped the debate on and around the work of Harvey *et al.*

Hence second, the article presented a two part overview of the major critiques of both Althusserian Marxism and the work of Marxist human geographers. First, Thompson's work outside, and Duncan and Ley's work inside, Anglophone geography concerns the question of agency and class. Thompson argues that the working class has always made itself as much as it was made, and thus that Althusser's abstract theory is never suitable to capture the infinite complexity that is class consciousness and struggle. In a similar vein, Duncan and Ley criticize Marxist human geographers like Harvey and Smith for reducing the sensuous particularity of real individuals to docile bodies in and through which capital circulates. Second, Resnick and Wolff's work outside, and Gibson–Graham's work inside, Anglophone geography concerns the question of structure and capitalism. Resnick and Wolff argue that capitalism is less impermeable and more porous, hybrid and cyborg than Althusser's (albeit subtle) thinking allows. It itself is overdetermined by and plays a part in the overdetermination of all other social wholes. Similarly, Gibson–Graham criticize Harvey *et al.*, for representing capitalism as always colonizing/penetrating/expanding and never colonized/penetrated/contracting by thinking and talking about historical and geographical sameness but not difference.

Both critiques, in the last analysis, can be found wanting, and thus third, the article presented an analysis of their limited and limiting nature. Marxists like Anderson argue that Thompson's weaknesses are Althusser's strengths, and vice versa; the former is weak on the objective determinants setting limits to and exerting pressures on social life, and the latter is weak on the idea of culture, and the attitudes and identities of ordinary people on the ground. Human geographers such as Smith, Walker, and Peet too see an academic division of labor as salutary, for when a writer abstracts into focus what people have to cope with and respond to in capitalist societies he or she necessarily abstracts out of focus how, exactly, different people in different places at different times cope and respond in different ways. Furthermore, Harvey insists that Marxist human geographers are not unknowingly blind to anything and everything noncapitalist, but are seeking to craft a framework for understanding a few significant processes and relations at large in the world. It makes good intellectual and political sense to theoretically sketch out capitalism's 'essential' characteristics, even though in practice it does not exist in a pure state. To emphasize how and why different people in different places have certain differences in common is to hold out the possibility for alliance formation, and a stronger politics that is both attached to place and reaching out across space. A politics of this sort ("something geographical," as Smith would say) demands that the power of abstraction be put to work in order to make sense of infinite complexity, seeing the wood for the trees and the signals in the noise.

See also: Determinism/Environmental Determinism; Dialectical Reasoning and Dialectical Materialism; Historical-Geographical Materialism; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Posthumanism/Posthumanistic Geographies; Poststructuralism/Poststructuralist Geographies; Structuration Theory; Structurationist Geography.

Further Reading

Althusser, L. (1969). *For Marx*. London: Allen Lane.
Althusser, L. and Balibar, E. (1970). *Reading Capital*. London: New Left Books.
Anderson, P. (1980). *Arguments within English Marxism*. London: Verso.
Benton, T. (1984). *The Rise and Fall of Structural Marxism: Althusser and His Influence*. London: Macmillan.
Bhaskar, R. (1989). *Reclaiming Reality: A Critical Introduction to Contemporary Philosophy*. London: Verso.
Castree, N. (2007). David Harvey: Marxism, capitalism and the geographical imagination. *New Political Economy* 12, 97 115.
Deutsche, R. (1991). Boys town. *Environment and Planning D: Society and Space* 9, 5 30.
Duncan, J. and Ley, D. (1982). Structural Marxism and human geography: A critical assessment. *Annals of the Association of American Geographers* 72, 30 59.
Elliott, G. (1987). *Althusser: The Detour of Theory*. London: Verso.
Gibson Graham, J. K. (1996). *The End of Capitalism (As We Knew It): A Feminist Critique of Political Economy*. Oxford: Blackwell.
Gregory, D. (1981). Human agency and human geography. *Transactions of the Institute of British Geographers NS* 6, 1 18.
Harvey, D. (1982). *The Limits to Capital*. Oxford: Blackwell.
Harvey, D. (1987). Three myths in search of a reality in urban studies. *Environment and Planning D: Society and Space* 5, 367 376.
Harvey, D. (1989). *The Urban Experience*. Oxford: Blackwell.
Harvey, D. (1992). Postmodern morality plays. *Antipode* 24, 300 326.

Massey, D. (1991). A global sense of place. *Marxism Today* June, 24 26, 28 29.
Ollman, B. (2003). *Dance of the Dialectic: Steps in Marx's Method*. Urbana, IL: University of Illinois Press.
Resnick, S. A. and Wolff, R. D. (1987). *Knowledge and Class: A Marxian Critique of Political Economy.* Chicago, IL: University of Chicago Press.
Thompson, E. P. (1978). *The Poverty of Theory and Other Essays*. London: Merlin.
Thrift, N. (1983). Literature, the production of culture and the politics of place. *Antipode* 15, 12 24.

Relevant Websites

http://www.marxists.org/
Marxists.

Structuralism/Structuralist Geography

R. G. Smith, Swansea University, Swansea, UK

Glossary

Antihumanism The emphasis in structuralism on explanation through 'structural causality' necessarily downplays the role of purposive action from individuals and groups. Structuralism stands in stark contrast to humanism in that its leading proponents – Lévi-Strauss, Barthes, and Lacan – all call into question the theory of the subject and consequently its central role within social theory.

La Langue The whole social linguistic system into which an individual is born. A master system of differences between signs – a higher principle of organization – that regulates all sign production above and beyond linguistics alone.

Parole Individual acts of speech, or how language is actually practiced by a linguistic community.

Saussurean Semiology/Peircean Semiotics The study of the life of signs within society (linguistics, fashion, rules of conduct, etc.). Semiology is the name given to such study by those following Saussure (primarily in Europe), while the term semiotics is associated with Charles S. Peirce and the American tradition of semiological analysis.

Sign In Saussurean linguistics a term to describe a unit of meaning with two components: the signifier and the signified existing in an arbitrary relation. Consequently, the sign has a dyadic (two-sided) structure with both material and immaterial aspects.

Signified The mental concept of the sign that is arbitrarily bonded to the signifier to form an internal relation. The signified is engendered by the signifier.

Signifier The material aspect of the sign. For example, when our vocal chords vibrate sounds are produced that are material in nature.

Structuralism A twentieth-century European (mainly French) school of thought that applied the insights of structural linguistics to social and cultural life. The approach briefly appealed to some human geographers in the 1970s because it was overtly anti-empiricist – being concerned with below-surface logics that have a causal efficacy over surface events – and consequently different to spatial science.

Synchronic/Diachronic Structuralism is, broadly speaking, an approach that stresses the synchronic rather than the diachronic; in other words, an approach that is antihistoricist – not seeking historical antecedents or causes – because it is concerned with discovering the universal 'timeless' structures of human societies and human minds.

Introduction

Structuralism is a philosophy and method that developed from insights in the field of linguistics in the mid twentieth century to study the underlying patterns of social life. In the social sciences the structuralist mode of inquiry sought not simply to identify structures or relationships *per se*, but rather to look behind or beneath the visible and conscious designs (beliefs, ideas, behaviors) of active human subjects (surface manifestations) to expose or unearth how those designs are in fact outputs, effects, consequences, products generated by underlying causes, hidden mechanisms, or a limited number of 'deep' structures that are universal to the human mind; structures which while not directly visible or knowable – with no grounding in subjectivity – are nevertheless absolute, autonomous, and only accessible theoretically through the techniques of a structuralist analysis. In other words, any explanation of social life, of observed patterns, can only be found in the general mechanisms, structures, schemas, or systems that are assumed to underpin all observable events; these structures cannot be touched and measured and are not directly observable in the events, phenomena, or patterns themselves, but must be deduced from them to reveal deeper logics.

Structuralism, the structural, stands in contrast to a reductionism because it holds that all forms of cultural expression – be they the domains of art, architecture, cookery, dress, human self perception, kinship relations, language, literature, music, etc. – cannot be understood in isolation, as somehow separate, but rather must be understood as positions within a structure or system of relations. Indeed, structuralism is holistic (anti individualistic and anti empiricist) because of its insistence that while observable phenomena are present, they are also absent precisely because any object's being is determined by its relationships, its relation to the whole structure to which it belongs; a structure that while not apparent is present in each of its observable parts.

The structuralist approach (across the human and social sciences) claimed a fundamental importance to the task of identifying and analyzing (implicit and hidden) 'deep structures' which – as with Bhaskar's realist philosophy – are theorized to underlie and generate (explicit and obvious) 'surface' or observable phenomenon. While structuralism resonates with the philosophy of realism in that it challenges the positions of empiricism and positivism, it does not agree – as in realism – that there is a knowable real world 'out there'; for structuralism what is knowable is structures. In other words, 'structuralism' is a term used in general to denote any kind of analysis that is concerned with exposing structures and relations, with finding orders rather than actions. More specifically and ambitiously however, structuralism was promoted as a philosophy with a worldview (a *Weltanschauung*), a universal understanding of reality and knowledge. Indeed, there were high hopes that structuralism could provide a general framework – a solid common structural basis – for 'rigorous' and 'serious' work across all the human and social sciences.

The hold of phenomenology and especially Jean Paul Sartre's existential Marxism in post war France was the intellectual context from which structuralism emerged as critique in the 1950s. In the 1960s and 1970s structuralism came to replace existentialism as the dominant intellectual movement and paradigm in France; an acceptance that can also be attributed to a broader general optimism as to the universality of science, with structuralism claiming to be a new science (albeit one that offered a structural analysis rather than the causal analysis common to the natural sciences). Structuralism represented a challenging critique, a radical break from previous philosophical traditions and theoretical methods/models (including those focused around beliefs in human intention, understanding, and consciousness such as phenomenology, humanism, and existentialism), with its rejection of metaphysics, its indifference to the human subject (including individual and collective action), and its interest in discontinuity rather than continuity and flux, or sociohistorical context. In general, structuralism was a method that was applied extensively in the study of language, society, art, and literature. However, whether or not structuralism can be described as a 'movement' is a subject of some debate given the differences between the so called 'first generation structuralists' (i.e., Lévi Strauss, Barthes, Lacan, Foucault). Indeed structuralism has also been described as an ideology, a debate, a method, a 'creative activity', and an intellectual fad precisely because it was never consciously formed to be a specific school of thought.

Structuralism became a highly influential strand of post war French philosophy impacting – to varying degrees – a host of disciplines including (but not limited to) anthropology, linguistics, literary criticism and the sociology of literature, aesthetic theory, Marxism, mathematics, psychology, sociology, history, architecture, and human geography. The works of Lévi Strauss in anthropology, Roland Barthes in literary theory, Jean Piaget in psychology, Jacques Lacan in psychoanalysis, Michel Foucault in philosophy, and Louis Althusser in Marxism were particularly influential (and controversial) in their respective fields pioneering the new approach in their respective disciplines and spawning a legion of followers. Indeed, structuralism had a number of independent leaders who – while often addressing common concerns – were also quite different in their specific enquiries. However, it can be said that it was above all the anthropological work of Lévi Strauss which was most responsible for making 'structuralism' both an intellectual fashion and almost a household word.

The structuralist approach did not, of course, appear in the 1950s/1960s out of a vacuum, somehow spontaneously appearing on the French intellectual scene. Indeed, many of the themes of structuralism had precursors, earlier antecedents, in the works of Bachelard, Bakhtin, Canguilhem, Cavaillès, Freud, Marx, Mauss, Merleau Ponty, etc. In fact, the basic idea of structuralism can be traced back at least as far as 1725, to the writings of Giambattista Vico, and in many ways owes a debt to the core continental tradition of rationalist philosophy that was advanced by René Descarte, Baruch Spinoza, Gottfried Leibniz, and Immanuel Kant. However, the true intellectual source of modern structuralism (even though the term postdates him) was the work of the Swiss linguist, Ferdinand de Saussure (1857–1913). For while the works of structuralism differ considerably by author, Saussure provided the core lexicon – the technical vocabulary with terms such as semiology, *la langue*, *parole*, anti essentialism, signified, signifier, arbitrary sign, synchronic, diachronic, paradigmatic, syntagmatic, etc. – common to many (but not all) structuralists (and post structuralists). Indeed, it was also Saussure who called for the development of a new science of the study of signs (a semiology) that would be applicable not only in the field of linguistics but also directly or indirectly applicable to all aspects of human cultures and social institutions as systems of signs.

Structuralist Thinkers

The Swiss linguist Ferdinand de Saussure is widely credited as the founder or father of modern linguistics. The inventor of structural linguistics Saussure also made significant contributions to comparative and historical studies of linguistics through his work on Sanskrit and Indo European languages. However, it was the posthumous publication (in 1916) of his *Cours de linguistique générale* (Course in General Linguistics) that led to the development of both structuralism and a radical new methodology for the study of linguistics. That text was constructed by his students from a course in general linguistics that he taught at the University of Geneva from 1907–11. Saussure's development of a structuralist theory of language led him to propose both a profound transformation in our conceptualization of language and a new science of the study of signs within society; a science which Saussure was to christen 'semiology' from the Greek *sēmeîon* 'sign'.

One of Saussure's key insights was the structure of the linguistic sign. He noted an arbitrary relation between the 'signifier' (signifiant) (word or sound image) and the 'signified' (*signifié*) (concept). Arbitrary because Saussure demonstrates that there is no natural or obvious con nection between signifier and signified, only a social re lation established by tradition and convention. In other words, the sign does not unite a thing and a name as in the process of nomenclature, but rather a concept and a sound image. Thus, signification (meaning) is a product of relations between signifiers and signifieds, and there fore not the product of an essential linguistic bond be tween word and thing.

Saussure's radical and important discovery was that languages function as sign systems that are relatively autonomous from reality precisely because any sign is radically arbitrary; with any signification only being de termined by the historically constituted systems of con ventions to which any sign belongs. Consequently, this discovery enables Saussure to draw a distinction between *la langue* (language) and *parole* (speech). The former de scribes the conventional relations between signs that are in the same system at any given time. In other words, *la langue* describes language as it exists as a structure without positive terms. In contrast, the latter is individual speech acts, or in other words, *parole* describes language as a process, how language is actually practiced by a community of speakers. For Saussure, *parole* (individual speech acts) cannot be understood in isolation from the system of conventions to which such acts belong (*la langue*). This distinction between *la langue* and *parole* is important for structuralists because it suggests a general model for research across a number of disciplines in the social sciences and humanities.

Saussure uses the analogy of the game of chess to illustrate how languages function as differential totalities where what matters is the relationships between the various pieces and not the intrinsic value or essence as such of individual pieces. Because there is no 'natural' correspondence between word and thing, then any of the pieces on a chess board can in fact be substituted by any other object without changing the structural basis – the rules – of the game (i.e., a coin could be a bishop, or a matchstick, anything). What is more, the game of chess also helps illustrate how Saussure's structural linguistics emphasizes a synchronic, rather than a diachronic (the view of historical linguistics), approach to the study of languages and societies. In chess it is the present con figuration of pieces on the board that matters to the players (nothing is to be gained from knowing how the pieces came to be organized as they are at any moment), and that is precisely how Saussure contends that lan guages should be studied if we are to grasp how language functions. In other words, Saussure downplays the im portance of individual autonomy and the role of practice to focus on the differential value of signs within totalities. It was the generalization of this new way of thinking that came to be so influential in the social sciences and hu manities in the mid twentieth century.

In 1967, the French Literary magazine *Quinzaine Littéraire* published a cartoon by Maurice Henry that subsequently became extremely famous through its widespread reproduction as a synopsis of the increasingly influential 'structuralist movement'. The cartoon became known as 'The Structuralist's Lunch Party' and featured the leading figures of French structuralism, namely Claude Lévi Strauss, Roland Barthes, Jacques Lacan, and Michel Foucault, and consequently we will briefly con sider their contributions to that particular strand of social and cultural theory in the remainder of this section.

The anthropologist, Claude Lévi Strauss (born 1908) is perhaps the world's most famous anthropologist and certainly one of the twentieth century's master thinkers influencing some of that century's most famous thinkers, such as Jacques Lacan, Roland Barthes, and Michel Foucault. The inventor of structural anthropology, Lévi Strauss's innovative approach meant that his doctrine and influence also spread far beyond academia, France, and the Francophone world. His contribution to the devel opment and reputation of anthropology is immense, but perhaps even more so has been his unswerving com mitment to applying the insights of structural linguistics to social science. Indeed, in the wake of Saussure, Lévi Strauss is widely considered as the founder or 'high priest' of modern structuralism for in addition to ad vancing anthropology as a science (ordering masses of ethnological data), rather than merely an activity of de scription, he also developed an approach and model that others – working across the humanities and social sci ences – could follow and develop.

Lévi Strauss directly introduced ideas from Saussure's structural linguistics into his anthropological studies to invent a structural anthropology that stood in sharp contrast to the functional anthropology developed by Bronislaw Malinowski (1884–1942) with its specific utilitarian empirical fieldwork based studies. Lévi Strauss applied the principles of Saussure's structural linguistic analysis to a range of cultural and societal phenomena: kinship, myths, rites, religion, ideology, cooking practices, ceremonies, totemic systems, the 'savage mind', etc. Analyzing both his and other anthro pologists' field data Lévi Strauss noted how fundamental symbolic structures underpin social and cultural life, structures that are extraordinarily universal and homo geneous across different people, cultures, and ages. In other words, Lévi Strauss was unswerving in his insist ence that the most apposite means to understand human societies is to investigate the fundamental structures of their organization. Lévi Strauss's structural anthropology was distinctive because it was antifunctionalist (i.e., not finding explanation in the intrinsic nature or value of phenomenon); antibiologist (i.e., not grounding social life in biology but in symbolic structures and exchange); and, anti empiricist (i.e., 'facts' are not deemed to speak for themselves) because its concern was with differential relations and exchange; with how different elements combine together to form universal structures. For Lévi Strauss the starting point for structural analysis was methodological, to find differences so as to define any social and cultural phenomena of interest as a relation between two or more terms (cf. a causal analysis that would operate on a one by one basis), the possible per mutations of which could be tabulated to eventually re veal a complete system, such as kinship or myth.

Lévi Strauss's work is rightly famous because of his identification and novel analysis of several fundamental underlying symbolic structures across different human societies (e.g., the incest taboo as a basis for kinship systems that enable the formation of social alliances). However, it is Strauss's conceptualization of structures as ternary (the product of at least three elements) and consequently internally dynamic (not simply empirical, skeletal, architectural, an observable reality), open to the historical and contingent, which is fundamental to grasping why and how universal symbolic structures self perpetuate. Indeed, Strauss has often been incorrectly critiqued – probably because his work emerged at a time when existential Marxism was the dominant intellectual paradigm in France – for interpreting structures as static systems of differences (as synchronic) and consequently in opposition to diachronic analyses. However, this cri tique is not true as such because in his key works Strauss sees structures as always containing an 'empty' third heterogeneous element – a floating signifier (like the phallus in Lacanian psychoanalysis) – that is perpetually open to take on new meaning and consequently intro duces the historical and contingent (*parole*) into his analyses of such structures as kinship, myth, and face painting.

Another key figure of the structuralist movement was the French semiotician, essayist, author, cultural critic, and literary theorist, Roland Barthes (1915–80) who in addition to his wide ranging interests and international influence was particularly famous within literary criti cism, primarily through his development of literary semiology, or a structuralist 'science of literature'. In his early career Barthes' interest in culture and language was influenced by the Marxism and existentialism – centered around Sartre – that was popular in France in the mid twentieth century. However, he subsequently moved toward an interest in semiology and Saussure's theory of the sign, making a substantial contribution to 'structur alism' by extending Saussure's semiological analysis of meaning. For example, Barthes developed Saussure's understanding of language as a system of internal op positions or differences to study fashion as a system or language (*langue*) where what is crucial is the relations between signifier and signifier (the clothing sign) and not the arbitrary relation between signifier and signified famously identified by Saussure. Overall, Barthes defined structuralism himself as a way of analyzing cultural ar tifacts through methods that were originally developed within the study of linguistics, precisely because Barthes' concern was with defining a number of nonlinguistic languages – such as fashion – through a series of struc tural analyses.

Barthes is perhaps most famous for his proclamation of 'the death of the author' and 'the birth of the reader', arguing for the disentaglement of texts rather than the deciphering of authors. Barthes' own texts served to undermine the idea that works of literature or critical thinking have a single 'message' or 'meaning', a uni vocality as perhaps intended by an original author (cf. the 'naturalist' view of language). Instead, Barthes demon strates how any author is not a God over meaning because 'texts' are multidimensional spaces where a multitude of writings, none of them original, collide and merge. In other words, Barthes presides over a shift in power from author to reader, so that the reader becomes the central figure in literary criticism precisely because – through the eyes of literary semiology – any author is no longer understood as the source and arbiter of meaning: Barthes shatters the illusions of individualism and originality.

Barthes is also famous for his numerous critiques of the – especially French – modern bourgeois social order. Within literary criticism he undertook structural analyses of some of the key texts from the canon of French lit erature – such as Balzac's novella *Sarrasine* – to reveal their intertextual construction where no single 'theolo gical' meaning is evident. However, while these literary studies were highly provocative, challenging, and con troversial it was Barthes more popular critiques of bourgeois society that made him almost as famous as Sartre with the French public. In his *Mythologies* (1957) Barthes famously deconstructs a number (28) of appar ently innocent instances of French culture and society to demonstrate their ideological contamination: from wrestling matches and striptease to advertisements for cleaning products and Einstein's brain. In his analyses Barthes contrasts myth to ideology (moving beyond simple ideological critique) to demonstrate the import ance of analyzing semiotically 'how' any myth (message) is communicated – rather than with what is said or hidden by a message as with ideology – and so consequently how myth is a product of *parole* (speech), rather than *langue* (the structure of language). Perhaps Barthes most famous demystifying essay was his analysis of a cover of the French magazine *Paris Match* that featured a young black soldier looking up to and saluting the French flag. Barthes' analysis questioned the 'naturalness' or presumed 'innocence' of this image, the unquestioned relationship between signifier (the image on the front cover) and signified (Frenchness, militarism) to produce the total sign (i.e., loyal black French soldier), reading it instead as a mythical attempt by the French mass media to justify French imperialism and more generally a bourgeois worldview by ignoring historical fact. For Barthes there is in fact a second order semiological system in play where the total sign (the loyal black soldier) is only the signifier, so that the signified is in fact France as a great empire served faithfully by all its subjects whatever their color or creed, and the total sign is the ideology of colonialism and imperialism.

Jacques Lacan (1901–81) was a French psychoanalyst who exposed the field of Freudian psychoanalysis to Saussure's structuralist approach to language. Lacan highlighted the structural aspects of Freud's thought, and crucially noted how Freud was concerned with explicating the structures of the unconscious. Lacan transforms Freud's approach through reference to the terms and concepts of Saussurian linguistics (and to some extent Jakobson) to argue that the unconscious is structured just like a language and consequently can be analyzed as such. In other words, while Freud was concerned with the structural organization of the unconscious mind, Lacan was concerned with linking language and the unconscious to argue that "the unconscious is structured like a language" (a famous Lacanian aphorism).

Lacan's psychoanalysis is concerned with the complex relationship between the subject and the signifying system. Lacan makes explicit the antihumanism of the structuralist approach through his argument that while any human subject is divorced from signification, he or she is nevertheless constituted by that very system of representation (Lacan famously developed the theory of the 'mirror stage' in infants). Lacan points out that for a subject to have a place in the social world he or she must also take a place within its language: to become a social subject any human must enter into the preexisting system of signification. Any subject is dominated by the signifier, by the internal differences of *langue*, precisely because the process of subjectivity is one of being entangled within the infinite web of signification. Thus, for Lacan the task of psychoanalysis is to examine the chain of signification.

We have learnt how Saussure was concerned with the arbitrary internal vertical relation between signifier and signified. However, Lacan's concern was with the division of the sign through an impenetrable bar so that the relations between signifiers are primary. For Lacan, the subject is the subject of the signifier precisely because of his insistence as to the impenetrability of the Saussurean bar, effectively separating signifiers from signifieds. The autonomy of the signifier is guaranteed precisely because no signifier ever comes to settle on any signified. Thus, for Lacan the Symbolic order is the realm of the signifier, and it is within that order – of signs, symbols, images, representations, and significations – that any individual is produced as a subject.

Through the publication of an astonishing series of books the French philosopher and historian of ideas, Michel Foucault (1926–84) transformed the European Anglo American intellectual landscape. Foucault's 'structuralism' was not as inspired by the ideas of Saussure as that of Lévi Strauss, Barthes, and Lacan (although the work of Saussure did have a very significant influence upon his earlier studies). Instead, Foucault's overall approach was more a kind of 'quasi structuralism' with no specific theoretical or methodological unity as such, but rather a concern with specific problems and methods at different times in his working life (Foucault did not have one all consuming model with which to explain everything). The approach evident in his early studies of psychiatry (*Madness and Civilization*, 1962), the social sciences (*The Order of Things*, 1966), and clinical medicine (*Birth of the Clinic*, 1963) was that of an antihumanist 'archaeology of knowledge' that understood these various systems of thought as 'discursive formations' independent of the beliefs and intentions of individual subjects. In other words, the legacy of Saussure is clear to see in early Foucault with different discourses all being understood as both arbitrary and autonomous. However, the approach evident in his later writings has far more to do with the ideas of the German philosopher, Friedrich Nietzsche than Saussure. Foucault shifted his approach to analyze different discourses, or models/*régimes* of 'truth' (*épistèmes*), not just as languages (following Saussure) but as genealogies (following Nietzsche). This shift to a 'genealogical approach' – being in the spirit of Nietzsche in that it signaled Foucault's lack of interest in developing totalizing or comprehensive explanatory schemas – enabled Foucault to explain changes in systems of discourse by linking them to changes in the nondiscursive practices of social power structures. Foucault first deployed his genealogical approach in his study of the discourse of penology (*Discipline and Punish*, 1975), a study of the psychological and sociological knowledge upon which modern prisons have been founded, and consequently effectively demonstrated the connection between knowledge and power; with how bodies of knowledge become systems of social control.

The brief explication, outlined in this section, of some of the key ideas of perhaps the most famous structuralist thinkers was informative because it demonstrates how structuralism is best understood not as a consciously formed school, a readymade answer (self enclosed and deployable as a fixed approach), but as a series of loosely connected relatively open interrogations across a number of disciplines which strive to introduce an explanatory order among seemingly incoherent phenomena. We have seen how while some structuralists followed Saussure's ideas closely, others such as Foucault did not as such, preferring to develop a more idiosyncratic approach. Indeed, to end this section it is perhaps interesting to note just how detached Piaget's definition of structuralism is from Saussure's lexicon.

Jean Piaget (1896–1980) was a Swiss psychologist, philosopher, and biologist who authored numerous articles and more than 50 books. He is most famous for his research on cognitive development in children; an analytical structuralist investigation into the structures that regulate the development of intelligence in humans. In his book entitled *Le Structuralisme* (1968), Piaget famously defined structuralism as a method of inquiry common to, not just linguistics and anthropology, but a wide range of disciplines including biology, mathematics, philosophy, physics, and psychology. Piaget noted three ways of identifying a structure: (1) wholeness, (2) transformation, and (3) self regulation.

Through the idea of 'wholeness' Piaget demonstrates how a structure differs from an aggregate: namely, any part of a structure has no independence outside of the structure in the form that it would otherwise have within the structure. In other words, wholeness for Piaget emphasizes the internal coherence of structures, so that structures are more than simply a sum of their parts precisely because all parts conform to a set of intrinsic laws which determine both their nature and the structure itself. Through the idea of 'transformation' Piaget notes how structures are not static, but have the ability to transform through their intrinsic laws: 'structures are also structuring'. For structuralists structures are not passive, but rather have a transformational capacity to process new things passing through the system. For example, a living language produces new words and expressions without changing its essential structure. Piaget also notes that structures are 'self regulatory' in their transformational procedures, valid in themselves in that they are closed. A structure such as language is 'sealed off' from any need to reference 'reality', or other structures or systems, because the structure of language has its own internal and self sufficient rules.

Piaget's transdisciplinary definition of structuralism is interesting because it outlines an understanding of 'structuralism' that is not directly connected to the 'structuralist project' initiated by Saussure and developed by researchers such as Lévi Strauss, Barthes, and Lacan for whom it made no sense to distinguish between structuralism and semiology precisely because their work made such serious recourse to the nomenclature – signifier, signified, synchrony, diachrony, etc. – of signification. This difference is significant because it is indicative of a drift away from a strict definition of structuralism where the ideas and lexicon of Saussure are fundamental to a far more general definition of structuralism as simply any kind of analysis that emphasizes the importance of structures and relations: 'among the 'structuralists' some are more structuralist than others'. And it is this distinction which is undoubtedly fundamentally important for understanding why 'structuralist geography' is in fact not strictly structuralist at all, but is instead a Marxist project that was above all concerned with how to couple together description and explanation (something that spatial science was failing to achieve in the 1960s): namely, with demonstrating how an understanding of underlying 'structures' allows for an understanding of 'surface' geographies.

Structuralist Geography?

Structuralism entered Anglophone human geography in the early 1970s in a very limited way – with citations from the humanist geographer, Yi Fu Tuan in 1972 and the Marxist geographer, David Harvey in 1973 – as a part of the humanist and Marxist projects of critiquing the empiricism and positivism of a spatial science that had come to dominate geography in the 1960s. While some human geographers subsequently went further to discuss the work of structuralists – for example, Lévi Strauss was mentioned and briefly discussed by Derek Gregory in his 1978 book *Ideology, Science and Human Geography* in which he attempted to relocate geography's search for spatial order to Lévi Strauss' conception of structure – the truth is that the work of structuralists has not been engaged with in any sustained way by human geographers; instead, their interest has been piecemeal and sometimes rather idiosyncratic.

A few of the geographers who have flirted with the work of structuralists include: James Duncan who discussed Barthes' structural and post structural interpretations of landscapes as texts, systems of communication, object systems, where ultimately – for Barthes – signs come to be emptied of all cognitive meaning; Hugh Matthews who, through his work on children, cognition, and mental mapping, added, among other things, to a critique of the ideas of Piaget and the constructivist approach to understanding cognitive development in children; and, perhaps most significantly, Gunnar Olsson who since the publication of his famous book *Birds in Egg* (and then *Birds in Egg/Eggs in Bird*) in the mid 1970s has – among other things – worked within the horizon of Saussure, engaging with the work of many structuralists, to explore the limits of language and representation. However, despite these noteworthy individual efforts (and an interest by some human geographers in semi otics), a coherent structuralist geography has not been produced in any canonical sense.

An examination of some undergraduate level intro ductory texts about the philosophy of human geography is telling. In a selection from the 1990s one finds either no reference at all to Saussure, Lévi Strauss, Barthes, or Lacan (as in Cloke, Philo, and Sadler's 1991 *Approaching Human Geography*); or a discussion of their ideas for no other reason than to note how human geographers have not significantly engaged with their ideas (as in Unwin's 1992 *The Place of Geography*, or Peet's 1998 *Modern Geo graphical Thought*). Furthermore, this neglect has con tinued into the twenty first century in edited collections. The book *Key Thinkers on Space and Place*, edited by Hubbard, Kitchen, and Valentine, published in 2004 collects essays by geographers on 52 'key thinkers', ran ging from Benedict Anderson to Iris Marion Young, but contains only one chapter on a so called structuralist, namely Michel Foucault. Similarly, the 2006 *Approaches to Human Geography* edited by Aitken and Valentine neither contains a dedicated chapter on structuralism, nor is there any mention of Saussure, Lévi Strauss, Barthes, Lacan, Piaget, or even Althusser; again only Foucault is mentioned in the book. Indeed, Foucault is by and large the only 'structuralist thinker' whose work a number of geographers have seriously engaged with (perhaps most notably by Felix Driver, Chris Philo, and Stuart Elden who between them introduced 'Foucault's geography', his 'spatial histories').

The reason for this lack of interest in, and engagement with, Saussurean structuralism is because in the 1970s many human geographers were becoming interested in Marxism. While one of the earliest citations of a struc turalist in human geography was that of Piaget – by Harvey who attempted to demonstrate certain con vergences between the genetic structuralism of Piaget and the historical materialism of Marx – it was the antihumanist 'structural Marxism' of Louis Althusser which was to become popular, and widely (mis)under stood as 'structuralism', within Anglo American human geography.

Louis Pierre Althusser (1918–90), the Algerian born Marxist philosopher and communist, is perhaps most famous for the circumstances of his later life which was ruined by bouts of mental illness (probably a manic de pressive psychosis), and his murder of his wife (Helene Rytman). However, it was Althusser's work which was of interest to human geographers in the 1970s because it was a variant of Marxism, a philosophy which, after falling out of fashion in France after the events of 1968, became – rather belatedly it must be said – highly popular in 1970s Anglo American human geography. While within some fields of study structuralism was brought into dialog with the works of Marx(ism) – e.g., Marx (alongside Freud) had some influence on the de velopment of Lévi Strauss' structural anthropology and consequently conveyed a certain Marxist orientation to anthropological studies – it was the work of Louis Althusser which is most significant in this regard because Althusser tried to synthesize his reading of Marx with a broadly structural viewpoint (or perhaps more accurately to read Marx anew as a 'science' along structural lines). It was Althusser who most directly (and controversially) introduced a structural perspective onto the central tenets of Marx's oeuvre, namely the modes and relations of production, the relations between different levels of society (economic, political, and ideological), and the structural contradictions.

Althusser's work gained an international audience in the 1960s with the publication of two key works, *For Marx* and *Reading 'Capital'* (the later was coauthored with Éti enne Balibar). Althusser developed a structural Marxism that challenged and stood in stark contrast to the He gelian (historicist) and humanist readings of Marx that were influential at that time. Most famously, Althusser identified an 'epistemological break' in Marx's oeuvre between a young Hegelian Marx and a mature scientific Marx. According to Althusser, the later Marx espoused a theoretical antihumanism in that he rejected a universal human essence or nature to conceptualize history as 'a process without a subject'. Althusser's so called 'return to Marx' was an attempt to counter what he saw as the distortion of Marx's writings by the official communist movement to reveal the science of Marxism, historical materialism.

For the purposes of our discussion here however it is enough to note that Althusser is only a structuralist thematically in that his 'structural Marxism' is anti em piricist, antihumanist, anti individualist, and anti historicist, and consequently on a very superficial level 'structuralist'. However, above all, Althusser is a Marxist precisely because he has little interest in Saussure's approach to language and its subsequent deployment and development in fields such as psychoanalysis and anthropology. Indeed, 'Althusser turns structuralism on its head' because his approach is fundamentally different to that of structuralists such as Lévi Strauss. Althusser sought to develop theoretical abstractions of underlying 'structures' from which to then approach and explain 'surface' realities; whereas Lévi Strauss started with his, and other anthropologists', empirical field data to then theorize the existence of underlying structures (cf. Derek Gregory's attempt in the 1970s to link together Lévi Strauss and Althusser as practitioners of 'structural explanation').

In fact, it is the post structuralist, Jean Baudrillard who can be truly described as a 'structural Marxist' precisely because, unlike Althusser, he does read Marx through Saussure. In his third book, *For a Critique of the*

Political Economy of the Sign (1981), he famously goes against Althusserian Marxism to argue that ideology is a form that traverses material production and sign pro duction, in other words, ideology is not just content, the product of an infra superstructural relation. To reach this brilliant insight Baudrillard reinterprets and combines both Marx's description of the commodity form and Saussure's description of the linguistic sign to show how ideology is present in both the very internal logic of the commodity form, in the relations between exchange value and use value, and in the sign form in the relation between signifier and signified: the latter terms – use value and signified – are the capitalist system's ideo logical guarantee. Consequently, through recourse to Saussure Baudrillard places back on the theoretical agenda in France the articulation of culture and economy from the roots of the commodity form (the structural relation between commodity and sign forms). His astute observation is that in late capitalism commodity has fused with sign (the political economy of the sign), and that it is precisely the homologous relation of the com modity sign form which now best describes the field of general political economy.

Anyhow, the main reason why Althusser came to prominence in human geography was because of the influence of his 'structural Marxism' on the work of the urban sociologist, Manuel Castells who was concerned with developing a Marxist understanding of cities, an approach that had a profound influence on the devel opment of urban geography. However, the structural Marxist geography that subsequently developed after Althusser and Castells – and which through critique by geographers such as Duncan and Ley came to lose its specificity, to be associated far more generally with 'structures' *per se*, anonymous 'forces', or 'logics' un touched by human agency – has little to do with either structuralism or with approaching what a structuralist geography would be after Saussure. Indeed, it is true to say that 'structuralism has not taken place in human geography in any integrated, systematic, or paradigmatic sense'. And perhaps one reason for that lack is because of how structuralism came to be associated in a general sense with 'structures' to the detriment of 'agency', and so human geographers moved away from structuralism to become interested in those approaches (e.g., structuration theory, realism) that promised to reconcile the perceived split between structure and agency. In effect, the possi bility of a fully fledged structuralist geography was strangled at birth, unjustly abandoned due to a supposed (and ungrounded) association with, on the one hand Althusser, and on the other hand, certain clumsy econ omistic versions of Marxism. However, at least from the 1990s onward a number of geographers have come to encounter structuralism retrospectively through their interest in first, postmodernism, and then second, post structuralism. Indeed, this is precisely how Foucault (and Baudrillard) came to be received in human geog raphy, not as structuralists, but as postmodernists and post structuralists.

Conclusion: Ex-Structuralists and Post-Structuralists

Based on applying the insights of Saussure's structural linguistics to all social and cultural phenomena struc turalism developed from the 1950s to the 1970s to be come a theory of human reality that is based on the argument that content (appearances) is reducible to form (structures). In addition, the focus of structuralism was upon relationships – the relations between terms, not the terms in themselves – that create their own autonomous spheres precisely because there are no extra linguistic referents or realities (Platonic forms or essences) avail able to human cognition. However, in the 1960s and 1970s (the heyday of structuralism across the human and social sciences), human geography remained largely ab sent from the debate, in part because at that time human geography was not the interdisciplinary subject that it is today.

Interest in the 'structuralist movement' in France proved to be short lived, as intellectual life moved on in the late 1960s/early 1970s toward a post structuralist approach. And some previously major structuralist figures – such as Barthes and Lacan – became ex struc turalists, proto post structuralists, or simply post struc turalists (much of post structuralism was in fact already present in structuralist writings) dissatisfied with the confines of Sausserian linguistics (the foundation of structuralism) to destabilize it and shift their position (others also moved on, with Foucault rejecting not only Marxism and psychoanalysis, but also eventually his particular version of structuralism). Alongside the shift of ex structuralists there was also the emergence of a new generation of post structuralists whose work began with and subsequently developed from a critique of the structuralist approach (e.g., Jacques Derrida's famous 1967 critique) to undermine, destabilize, and make undecideable our understandings of – among other things – language, meaning, social institutions, and the self.

There is a final point to be made. The writing of this article inevitably reflects the particular demands of this encyclopedia. In other words, while this article has been written in a way that highlights and summarizes certain key structuralist thinkers there are other ways that this article could have been written, and consequently other ways that structuralism and its significance may be understood and gauged within human geography. Indeed, in his essay on 'How Do We Recognize Structuralism?'

Gilles Deleuze eschews the method of recognizing certain people as structuralist or not, the very method that has been followed here. Instead, Deleuze structures his essay around a question: 'What do we recognize in those that we call structuralists?' As an answer Deleuze identifies seven criteria for identifying structuralism (e.g., the symbolic; the serial; the empty square), and his essay is undoubtedly particularly useful as a guide for reading human geography today, for revealing how contemporary human geographers are engaging with structuralism retrospectively through post structuralism. In other words, the 'crisis of representation' and the 'cultural turn' that took place in human geography in the 1990s has served to place much of human geography within the horizon of Saussure's splitting apart of the sign, and Deleuze's essay is instructive for helping us in recognizing that fact.

See also: Poststructuralism/Poststructuralist Geographies; Semiotics; Structural Marxism; Structuration Theory.

Further Reading

Althusser, L. (1969). *For Marx*. Harmondsworth: Penguin Books.
Barthes, R. (1973). *Mythologies*. St Albans: Paladin.
Baudrillard, J. (1981). *For a Critique of the Political Economy of the Sign*. St. Louis: Telos.
Caws, P. (1988). *Structuralism: The Art of the Intelligible*. Atlantic Highlands, NJ: Humanities Press.
Culler, J. (1975). *Structuralist Poetics: Structuralism, Linguistics, and the Study of Literature*. London: Routledge.
Deleuze, G. (2004). How do we recognize structuralism? In *Desert Islands and Other Texts, 1953 1974*, pp 170 192. New York: Semiotext(e).
Derrida, J. (1976). *Of Grammatology*. Baltimore, MD: John Hopkins University Press.
Driver, F. (1990). Discipline without frontiers? Representations of the Mettray Reformatory Colony in Britain, 1840 1880. *Journal of Historical Sociology* 3, 272 293.
Duncan, J. and Duncan, N. (1992). Ideology and bliss: Roland Barthes and the secret histories of landscape. *Writing Worlds: Discourse, Text and Metaphor in the Representation of Landscape*, pp 18 37. London: Routledge.
Duncan, J. and Ley, D. (1982). Structural Marxism and human geography: A critical assessment. *Annals of the Association of American Geographers* 72(1), 30 59.
Elden, S. (2001). *Mapping the Present: Heidegger, Foucault and the Project of a Spatial History*. London: Continuum.
Foucault, M. (1973). *Madness and Civilization: A History of Insanity in the Age of Reason*. New York: Vintage/Random House.
Gregory, D. (1978). *Ideology, Science and Human Geography*. London: Hutchinson.
Harris, Z. S. (1951). *Structural Linguistics*. Chicago, IL: University of Chicago Press.
Harvey, D. (1973). *Social Justice and the City*. Oxford: Blackwell.
Hawkes, T. (1977). *Structuralism and Semiotics*. Berkeley, CA: University of California Press.
Lacan, J. (1977). *Ecrits: A Selection*. London: Tavistock.
Matthews, M. (1984). Environmental cognition of young children: Images of journey to school and home area. *Transactions of the Institute of British Geographers* 9, 89 105.
Olsson, G. (1975). Michigan Geographical Publication No. 15: *Birds in Egg*. Ann Arbor, MI: Department of Geography, University of Michigan.
Philo, C. (1992). Foucault's geography. *Environment and Planning D: Society & Space* 10, 137 162.
Piaget, J. (1971). *Structuralism*. London: Routledge & Kegan Paul.
Robey, D. (ed.) (1973). *Structuralism: An Introduction*. Oxford: Clarendon Press.
Saussure, F. de (1983). *Course in General Linguistics*. La Salle, IL: Open Court.
Strauss, L. (1963). *Structural Anthropology*. New York: Basic Books.
Sturrock, J. (ed.) (1979). *Structuralism and Since: From Lévi Strauss to Derrida*. Oxford: Oxford University Press.
Tuan, Yi. Fu. (1972). Structuralism, existentialism and environmental perception. *Environment and Behaviour* 3, 319 331.

Structuration Theory

R. Lippuner and B. Werlen, Friedrich-Schiller-University Jena, Jena, Germany

Glossary

Agency Conceptualization of human activity emphasizing the capacity of an individual actor to control its own actions instead of being subjected to external forces and conditions.

Duality of Structure Core concept of Giddens' theory of structuration. In opposition to classical social theory, which is characterized by a dichotomy of structure and agency, structuration theory argues that social structures are the medium and the outcome of human agency. All social practices are carried out under specific structural conditions. While at the same time structures are maintained and reproduced only through these social practices.

Habitus Schemes of perception, interpretation, and conduct. The habitus of an individual is a product of incorporating the logic of practice in a specific field of social space and is therefore shaped by the structures of a society or group. Then again it is the generative principle of practices and therefore a basis for the production of structures, that is, differences (fields) of social space. Thus habitus can be called a structured and structuring structure.

Practical Consciousness/Sense of Practice Tacit knowledge about the circumstances of social actions. In the vast majority of social situations which they encounter in everyday life, human individuals know what has to be done. They routinely draw upon this knowledge without questioning its foundation. Since they are usually not asked to discursively express what they do or why they do certain things, most social situations, respective expectations, and utterances are taken for granted.

Praxis The entirety of individual and collective human activities that maintain and transform social life, that is, the manner in which all aspects of social reality are constituted. Praxis encompasses conduct and interaction in face-to-face situations, as well as complicated and extensive types of social collectivities and relations between people in remote places.

Reflexivity Circular reproduction of a process (of cognition). In the social sciences reflexivity particularly refers to the ability of an individual to reflect upon (the conditions of) its doing and thinking and to intentionally govern its activities. In structuration theory, reflexivity is also regarded as a characteristic feature of life in late modernity, as well as an indispensable element of social scientific research.

Resources Disposal of the activities of other people (authoritative resources) and availability of material objects or artifacts (allocative resources) to achieve aims or further interests. Resources are involved in the generation of power and as such implied in all social practices.

Rules In the philosophical tradition either directives of conduct or regularities of praxis. From a structurationist point of view, rules are techniques or generalized procedures social actors routinely employ to cope with social situations. Rules have a normative aspect – they refer to legitimation and sanction – and a constitutive character – they refer to signification and communication.

Social Space In Bourdieu's theory of practice, a synonym for society or the social. According to the notion of social space, society is a relational space, in which each position is defined only by its relations to all the other positions. Differentiating principles are economic capital (money, property, estate, etc.), social capital (social relations, allegiances, contracts, etc.), and cultural capital (education, skills, taste, etc.). Social space is divided into various subspaces that represent general fields of social practice (politics, economy, science, art, education, etc.).

Structure In structuralist tradition of social theory, a set of stable and stabilizing patterns of social life that shape human agency. In structuration theory, structures are seen as rules and resources inherent in social practices.

William James, one of the chief proponents of American pragmatism, held that every theory is easy to understand, once the reader has grasped the core of its perspective. Concerning structuration theory, this core expresses the idea that social reality is being continually (re)produced and structured in space and time by competent actors in their everyday praxis. Society is constituted in and through human agency, which must be regarded as both structured and structuring. Consequently, the basic assumption of structuration theory is that social sciences must not explore societies or social life in structural categories, but rather in concepts of action and institutionalized practices. To put it succinctly, not structure is the main object of social research, but the acts of structuration. With that, structuration theory wants to avoid a dichotomization of agency and structure and bridge the fundamental gap that hitherto stretched in social theory: the gap between individualistic voluntarism and structuralistic determinism, as well as between micro and macroanalysis. This dichotomy is, as Pierre Bourdieu states in the introduction of his 1980 book, *Le sens pratique* (The logic of practice), the most funda mental and fatal opposition that artificially splits up so cial sciences.

The social theories of the twentieth century find themselves between two paradigms: the individualistic and the structuralistic line of reasoning. The indi vidualistic approach is represented by phenomenology (Husserl, Merleau Ponty, and Schütz), Berger and Luckmann's theory of the social construction of reality, ethnomethodology (Garfinkel), as well as by Wittgen stein's philosophy and symbolic interactionism (Mead, Blumer). Despite the remarkable differences among these theories, they all center on thinking, talking, or acting subjects. All individualistic theories focus their attention on the actors' knowledge and experiences and ask how they make the social world intelligible to themselves and others.

The structuralistic paradigm is first of all based on the work of Durkheim and on de Saussure's strucural lin guistic. Thanks to structural anthropology (Levi Strauss) and Marxist structuralism (Althusser), structuralism is one of most influential social scientific approaches of the twentieth century. In contrast to individualism, structuralism assumes that there are social and cultural structures within society, that is, superindividual distri bution patterns and collective symbol systems, that shape individual acting, thinking, and talking. Hence, the at tention of structuralistic social science does not focus on individual motives and interpretation but on material and symbolic structures that, in Durkheim's terms, represent the mold of social praxis.

In between these two poles, structuration theory holds a mediating position. It does not deny that super individual social circumstances provide the frame for the actors' individual actions. However, structuration theory also points out that social structures have to be con tinually produced and reproduced, that is, maintained, in actual everyday practices. That is the reason why on the methodological level, structuration theory aims to over come the dichotomy of agency and structure, as well as the opposition of micro and macro level.

The chief proponent of structuration theory is doubtlessly the British sociologist, Anthony Giddens. His theory received great attention also in geography. Besides Giddens, the French sociologist, Pierre Bourdieu, is an other important proponent of structuration theory. Since the 1960s authors other than Giddens and Bourdieu, the two outstanding authors of the structuration theoretical approach, tried to solve this dichotomy in steering clear from the determinism of structuralistic–objectivistic explanations, as well as from the voluntarism of individualistic–subjectivistic descriptions. In the famous article 'On the determination of social action in space and time' from 1983, Nigel Thrift even speaks about a structurationist school. Proponents of this school are, among others, the British philosopher, Roy Bhaskar; the historian, Edward Palmer Thompson; the cultural sci entist, Raymond Williams; and the German sociologist, Ulrich Beck.

The following remarks are confined to presenting the theories of the two main proponents of structurationist thinking: the first part of the article deals with the core ideas of Giddens' theory of structuration; in the second part Bourdieu's theory of practice will be addressed. Structuration theory was and is one of the most in fluential contemporary social theories not only in social and cultural sciences but especially also in human geography. The human geographical interest in struc turation theory stems from Giddens' and Bourdieu's concentration on the concept of space and the spatiality of social practices. Adapting structuration theory in the 1980s and 1990s, enhanced the development of a struc turationist approach in human geography. This struc turationist geography is not discussed in this article but addressed in a separate one.

Giddens' Theory of Structuration

Claiming that no classical sociological theory approaches the social in a way that is still valid today, Giddens holds that a new perspective is needed. With his theory of structuration, this new perspective is put into concrete terms in the 1970s and 1980s. The outlines of the theory of structuration are given in Giddens' main work *The Constitution of Society* (1984). The theory's broad appeal allows theoretical and empirical research in all areas of the social sciences, thus promoting interdisciplinarity to a high degree.

Giddens' theory of structuration is important to geography in many ways. Like no other theory of the social, it integrates spatial aspects of social life system atically into its argumentation. Unsurprisingly, struc turation theory found its way into Anglo Saxon human geography, particularly into new regional geography. In German speaking human geography it was conceived as an elaboration of classic action theory or action centered approach toward a 'social geography of everyday regionalizations'.

Duality of Structure

Central to the theory of structuration is the idea of the duality of structure. Duality means that social structures are constituted by human agency while, at the same time, being the means of this constitution. Structure and agency are neither the same, nor can they be regarded as dualistic. Rather they must be received as two diverse aspects of one and the same thing, that is, social praxis. Structures are both medium and the outcome of praxis. Agency and structure are therefore characterized by an immanent dialectic relation. Of course, these assumptions do not only affect the relation between agency and structure but also agency and structure itself. Agency in this sense is to be apprehended as both structured and structuring. Structure is regarded as the outcome of ac tions and, at the same time, as a means for generating actions. One of the consequences of this notion is that every social situation can be looked at from two different perspectives – from the viewpoint of agency, as well as structure. The analysis of praxis and the analysis of structural settings can therefore be seen as two compat ible concepts of social research.

However, within the theory of structuration, only the acting individual is endowed with a consciousness. Merely the authority generating the action, that is, the human individual, is capable of structuring. Structures have no consciousness and do not act. Consequently, Giddens' approach attempts to dissolve the duality be tween structure and agency (that has shaped large parts of social theory) by emphasizing the subjective side of this relation.

Agency and the Flow of Human Practice

One main assumption of structuration theory is that the social world is generated in and through social praxis. Social praxis comprises the conditions and consequences of all historically and spatially localized human actions and interactions. Therefore, praxis can be regarded as synonymous with the constitution of social life, that is, the manner in which all aspects of social life are gener ated in and through the performance of social conduct. In stressing the significance of praxis, Giddens implies a conception of agency contrary to classical sociological action theory, as, for instance, proposed by Max Weber. In contrast to classical action theory and analytical action philosophy, Giddens does not conceive social practices as sequences of distinct activities but rather as a flow or series of activities. Neither is this flow clearly delimited nor is it fixed. Consequently, 'intention' cannot be re garded as the essential characteristic of human action. What is of central importance is, according to Giddens, the ability to act in social context, or, put differently, the capability of shaping the flow of activities in a way that makes social life possible at all.

The theoretical presumption that human agency must be considered as a continuous flow has diverse impli cations. First, it implies that social praxis might be ana lytically differentiated into reasons, motives, and intentions, etc. However, these analytical categories cannot – contrary to assertions in traditional action theory or analytical action philosophy – be analyzed separately. Since they are inseparable in social praxis, they shall be scientifically analyzed in context, that is, as being related to one another.

Second, Giddens' concept of praxis entails a specific notion of intentionality. The term 'intentionality' forms the very core of a definition of action proposed by the philosophy of language and traditional action theory. The philosophy of language, but also the sociological action theory of Max Weber, Alfred Schütz, or Jürgen Habermas, commonly regard intentionality as the guiding sequence of conduct. Since Giddens, however, refuses atomistic descriptions of human agency, he has a different outlook on intentionality. Giddens rejects the view that only those activities must be called action that are considered in tentional. Instead he claims that, first of all, intentionality cannot be conceptualized as a delimited sequence, and that second, all activities bearing unintended con sequences must be regarded as actions as well. Agency therefore refers not to the intentions actors have in doing things but to the capability of changing the trajectory of their doing at any moment in a given sequence of conduct. Intentionality is seen as a process rather than a completed sequence. In this sense, intentionality is always open to change or revision during the course of action.

The third crucial implication of Giddens' conception of practices refers to the reflexivity of action. Reflexivity allows the continuous revision of the direction of social conduct. Describing agency as being reflexive therefore stresses the freedom of human praxis – at least to a certain degree. That is to say, human agency cannot be determined in a causalistic sense. Human individuals can always act differently than they actually do or have done. Nevertheless, this freedom to act is limited. Since agency also involves following rules, it not only comprises a transformative but also a normative dimension.

It should not be inferred from what has been said up to this point that the consequences of actions do always coincide with the effects the agent attempted to bring about. The fact that actions are reflexive does not ne cessarily mean that all of our activities bear the con sequences actually intended. Quite the reverse. Actions often have consequences that were not intended by the agent. However, the unintended outcomes are results of human agency and not products of a social system. Sys tems, to put it directly, do not act in any way. They do not have any needs. Only human individuals are capable of acting and bringing about intended and unintended outcomes that in turn become the conditions of suc ceeding actions.

Agent and Agency

The different levels of consciousness in human agency are described in a three stage model that not only differs from the classic action theory but, despite some similarity, also from Freud's psychoanalysis. Modifying Freud's distinctions, Giddens differentiates between the unconscious, practical consciousness, and discursive consciousness. The unconscious is separated from the other two by mechanisms of repression. The border between practical consciousness and discursive consciousness, however, is not clearly definable but a dynamic one. Much of what actors know about the world and the domain of their actions is known to them in an unarticulated manner, as a part of practical consciousness. They are able to do things that require this kind of knowledge, even though they are unable to report outright on it. That is to say, actors possess this knowledge without being able to give account on it (see **Table 1**).

The three stages become differently relevant for several aspects and types of agency. Motivating oneself for an action is primarily bound to the unconscious. Hence, motivation is most remote from the actor's reflexive monitoring. However, the unconscious is not regarded as a hotbed of dark forces. Instead, motives are rather interpreted as needs or desires. They merely refer to the potential of action, not to action itself. From the fact that human beings do actually have needs and desires, Giddens does not conclude that actions are determined by them. Needs and desires solely indicate a general course of action, which is perpetually open to interpretation by the acting individual. Thus, there is no deterministic causality in the social world.

Practical consciousness and rationalization of action relate to one another as most of our daily activities are highly routinized and do not require further explanation or justification. Giddens points out that, at all times, actors 'know', albeit they cannot directly report on it. This stresses the actors' 'knowledgeability' in contrast to the structuralists' claim that society and institutions or social structures operate behind the back of the actors. Even though actors are not always able to describe what they actually do, they are not unaware. Social actors rather possess a practical consciousness consisting of unquestioned, rationally undoubted knowledge, which forms the basis for routines and traditions. They hold a great deal of knowledge about the social world, which they apply in the production and reproduction of everyday social encounters. Nonetheless, the vast bulk of such knowledge is practical rather than theoretical in character. Yet, this practical consciousness can be raised to the level of discursive consciousness, if actors ask themselves what they actually do and how they do it: they are, at least potentially, able to discursively account for their actions (see **Figure 1**).

Table 1 Types of consciousness and agency

Type of consciousness	*Reference to action*
The unconscious	Motivation of action
Practical consciousness	Rationalization of action
Discursive consciousness	Reflexive monitoring of action

Only discursive consciousness allows reflecting on, describing, consciously controlling, and rationally reporting on actions. Here, the agents' 'knowledgeability' reaches its climax. Actors are permanently able to correspond to the given and altering situations of action. Although these premises might be given occasionally, they are not necessarily given at all times.

Giddens argues that in modernity, those parts of life proliferate that are either expressions of the discursive consciousness or reflexively monitored action. Late modern societies can thus be characterized by the reflexivity of social processes – even though this reflexive constitution is not accompanied by an increased rational control over all areas of life. While, on the one hand, our world is increasingly constituted through reflexively applied knowledge, this knowledge is, on the other hand, ever changing. All this results in a highly complex and, for the most part, problematic relationship between reflexivity and ontological security, which, above all, affects the basis of trust.

Structure and Structuration

Within the strucuralist approach, structures are considered to be the crucial part of the social world. They are attributed with generative or causal power. Giddens strictly opposes this view in assigning the main constitutive power of the social world to the agency of human individuals, who, in turn, relate to structures in their social praxis. Another divergence from structuralism is Giddens' redefinition of the term 'structure', which differs fundamentally from its hitherto use in social sciences.

Giddens initially approaches the term 'structure' in a negative way: structures are neither groups, nor organizations, nor subjects. This negative definition elucidates two aspects of what structures are not. First, structures do not exist outside agency. They are not external to action,

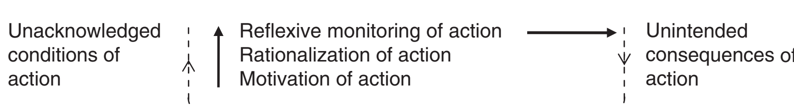

Figure 1 Stratification model of the agent. From Giddens, A. (1984). *The Constitution of Society: Outline of a Theory of Structuration*, p 5. Cambridge, London: Polity Press.

as structural functionalists or structuralist Marxists claim. Therefore, structures cannot be conceived as being deterministic, that is, as externally governed human agency. Second, structures are no patterns or systems of interactions as has been partly stated by Parsons. Structure and system must not be equated; it must rather be assumed that systems have structural properties.

Structures can be envisaged similar to language. Language represents an abstract ability of a community of speakers. Analogously, structures are the abstract capability of a group of actors. Like language, structures only exist in a virtual sense and neither has a subject nor the ability to act. As language becomes real only in speech, structures are solely realized in action. But where do structures exist, when they are not presently being realized in actions?

Giddens' answer to this question is, at first glance, rather confusing. He claims that structures are present in their absence. That is to say structures cannot be found beyond the capability of acting individuals. Structures are not present as something observable but as something invisible, that is, the capability of agents. They find expression only in the realization of these capabilities. Specifying the capabilities in question, Giddens outlines what they consist of. He claims that structures exist as knowledge about how things are to be done, said, or written. Social praxis is based on this knowledge and at the same time the actualization of it. The knowledge in question can be located in the discursive or practical consciousness and in the unconscious of an actor. In this respect, structures can also be seen as parts of consciousness or traces of memory. In order to gain a deeper understanding of the concept of structure, it seems necessary to clarify what this knowledge is made of.

Structure comprises two aspects: rules and resources. That is to say, the flow of human action is embedded in certain rules and resources that are, at the same time, produced and reproduced by human action. Or, put more accurately: human agency refers to specific semantic and normative rules, as well as to allocative and authoritative resources (see **Table 2**).

Rules and Processes of Structuration

Actors always interpret social situations in accordance with certain rules of interpretation. The compliance of standards ensures that certain scopes of interpretation are not being transgressed. This can be called the normative aspect of structure. Rules of social practice have not only a regulative, but also constitutive character for social practices. This can be illustrated by the offside rule in football. According to this rule, no player of the attacking team must, at the moment the ball is played by his teammate, be nearer to his opponent's goal line than both the ball and the second last opponent. This offside rule is constitutive for the attacking and defending strategies of the two teams. It is also regulative because violating the rule will be penalized (through a free kick against the attacking team). Since rules are both regulative and constitutive, they cannot be characterized as being exclusively one of both.

Table 2 Rules and resources

Rules	*Resources*
Interpretative schemes (semantic rules)	Allocative resources
Norms (rules of sanction)	Authoritative resources

Giddens considers rules as procedures of action. In this view, social rules are a kind of technique that is applied in the flow of action. Nonetheless, the applicability of rules is not restricted to verbally expressed or written rules (laws, bureaucratic rules, rules of a game, etc.). Formulated rules are thus codified interpretations of rules rather than rules as such. Giddens additionally points out that most rules are part of the practical consciousness. Hence, rules are best conceived as generalized procedures of action that enable the agents to routinely reproduce their actions, to coordinate them with others' actions, and to change situations. Rules are regarded as techniques in the sense that they play a constitutive role for meaning and forms of social conduct. Furthermore, they are also constitutive in shaping action. They do not, however, generate action as structuralism claims.

Resources and the Reproduction of Power

Within structuration theory, the term 'resource' does not refer to something material, like natural resources or raw materials. Resources rather refer to power over people and materials. Or in other words: a resource is not understood as a certain state or thing but rather as a capability of transformation. It denotes the range of things an actor is able to do. Obviously, this notion of transformative capacity refers particularly to the capacities commonly referred to as power. Hence, Giddens defines power as a part of social encounters. That is to say, power is not regarded as external to social practices but as implicit in all interactions and social relations.

Giddens distinguishes between two types of transformative capacities: the transformative capacity of (1) authorization and (2) allocation. Authoritative resources refer to the capacity of controlling the humanly created social world, that is, controlling the actors themselves. In contemporary societies, the dominant authoritative resource is the power to organize: the social time–space of human individuals (the temporal–spatial constitution of society), the production and reproduction of the human body, as well as constitution of chances of self development and self expression. Allocative resources refer to the capability of controlling the natural basis of human life, that is, steering the appropriation and use of material objects. Therefore, allocation resources involve a transformative capacity generating command over material phenomena.

Social Systems and Institutions

The organization of access to material goods by means of authoritative resources and the regulation of activities of individuals in the sense of authoritative resources point to established patterns of social practice, that is, to institutions and social systems. According to Giddens, the term social system refers to social relations, interdependencies, and networks between actors and collectives that are reproduced as regular social practices. Thus, Giddens dissociates himself from newer systems theory that relates the term social system to the autopoietic self-production of society or parts of society. He rather regards social systems as systematic social relations that arise from the practical reproduction of structures (rules and resources) over space and time. That is the reason why social systems contain moments of structure, that is, institutionalized aspects of social practice that stretch in space and time. Such moments of structure can refer to sequence of regulated interaction (e.g., a seminar) or to operational mode of organizations (e.g., a university) just as to the organization of societal totalities.

According to Giddens, societal totalities, that is, societies, are themselves social systems that stand out in bold relief against other systemic relations. In continually reproducing specific structural principles, a system of institutions is created by means of which societal totalities contrast with other system relations. Thus, the notion of 'system' opens up the possibility to study the reproduction of societal reality on the scale of larger and long term relations, that is, by means of practices that are continually reproduced over time and space. In this regard, various structural properties of social systems can be identified: structures of signification (semantic rules), structures of authority (resources), and structures of legitimation (moral rules). Each dimension can be linked to corresponding social institutions, which are apprehended as expressions of institutionalized praxis. On the one hand, the respective institutions (modes of discourse, political institutions, economic institutions, legal institutions) monitor these domains, and on the other hand they are responsible for the production of rules of interpretation and sanction, as well as the attainment, management, and enforcement of the resources. Accordingly, the different institutional spheres can be related to the specific domains of scientific theory (see **Table 3**).

Table 3 Structures, theoretical domains, and institutions

Structures	*Theoretical domain*	*Institutional order*
Signification	Theory of coding	Symbolic orders/modes of discourse
Domination	Therory of resource authorization	Political institutions
	Theory of resource allocation	Economic institutions
Legitimation	Theory of normative regulation	Legal institutions

From Giddens, A. (1984). *The Constitution of Society: Outline of a Theory of Structuration*, p 31. Cambridge, London: Polity Press.

Space, Time, and Structuration

Giddens' theory of structuration aims to analyze social praxis in both spatial and temporal context. Ignoring these contexts is, according to Giddens, one of the fundamental flaws of previous social theory. This is particularly true for the spatial aspect, which has hitherto been excluded from sociology fearing a lapse into some kind of geographical or environmental determinism.

Giddens develops a concept of regionalization that opens up the door for a lively discussion of structuration theory within geography. Unlike earlier proponents of social theory, Giddens refers positively to the theory of geography. Conceptualizing social actors as bodily beings, Giddens discovers Hägerstrand's time geography – a concept that catches the essential conditions of spatial existence. In social sciences, the time geographical method is, above all, used to depict the constraints of coordination, which occur in the time–space sequences of everyday life. It shows that social interactions always take place at specific places that form the reference frame of these interactions. However, in trying to grasp the contextuality of social interactions, Hägerstrand's time geographical models turned out to be too mechanistic. For this reason, Giddens develops his own spatial terminology employing Goffman's insights into human behavior in public space. One core element of this spatial nomenclature is the term 'place', for which Giddens prefers the expression 'locale'. In doing so, he wants to point out that places are more than mere arrangements of structures or material conditions for interactions. Physical elements of places are loaded with specific (individual or collective) meanings that are incorporated by the actors as part of the interaction.

When determining the meaning of locales in more detail, it is crucial to note that locales are regionalized, that is, they are subdivided into different regions. In other words: physical elements of social interaction's context are loaded with social content through symbolic markings and thus become meaningful context elements of social praxis. An important dimension of regionalization is the division of space into front regions and back regions. This division encompasses all spatial scales and encopasses the positioning of the body in face to face interactions, the division of countries into privileged and disadvantaged areas, or the geopolitical distinction between center and periphery. Even though Giddens attaches high importance to the significance of spatiality in his theory of structuration, he did not enhance the spatial terminology in his later work. Nonetheless, in the 1980s and 1990s, his theory evoked the geographer's interest and had a significant impact on geographical theory development.

Discussion

Giddens' work is criticized mostly for its syncretism and eclecticism. And quite legitimately if you look at the numerous theories that are fused in his structuration theory. The potpourri of theoretical references is espe cially suspicious, if you are used to following one specific line of argumentation and assign every theory to an al ready established tradition of thinking. Giddens, however, rejects this objection. On the contrary, he considers the comfort of established views to be some kind of intel lectual sloth. According to Giddens, it is more important to set up a frame in which vital and enlightening ideas from other contexts can be brought to bear. Although there is a certain point to such a pragmatic view, it surely does not erase all logical and methodological contra dictions that are bound to a theoretical eclecticism. For instance, the objection remains that, in the end, Giddens' theory offers only a limited insight into the relation of agency and structure as critics do not consider the dualism to be successfully overcome. In this respect, geographers like Thrift criticize that the actors are con ceptualized as being too individualistic rather than social. Moreover, it is said that Giddens underestimated the efficaciousness of structures. These objections are the reason why Thrift, Gregory, and Pred back off from structuration theory, even though in the 1980s they modeled their reformulation of regional geography on it.

Another important problem of human geography concerns structuration theory's conception of space. In using the conception of space outlined by Hägerstrand's time geography, Giddens conceives space (in the New tonian sense) as given and absolute, sometimes even as a determining factor of human agency. Assuming however, as Giddens does, that every societal reality is a consti tuted reality, this must also be true for spatial realities. The significance spatial conditions have for human agency, are constituted in social praxis and thus depend on human actions.

Besides these objections, which have partly been re futed by geography's adaptation of structuration theory, Giddens' theory of structuration is one of the most vital theoretical foundations of human geography up to the present. The discussion of the theory in the 1980s con tributed to a social scientific reorientation of human geography while at the same time providing a suitable basis for it.

Bourdieu's Theory of Practice

Along with Anthony Giddens, the French sociologist, Pierre Bourdieu is one of the most cited contemporary social theorists. His work reaches back into the 1950s and is exceedingly voluminous. Besides a multitude of studies concerning culture and social life in Northern Africa and France, his work contains an elaborated outline of a theory of praxis. In contrast to the significance Bourdieu gained in sociology, cultural studies, and anthropology, his works received rather minor attention in geography. Unlike Giddens' theory of structuration, Bourdieu's theory of praxis has never been used systematically to develop a specifically geographical approach. This is even more surprising as Bourdieu develops an interesting notion of social space and of the social appropriation of physical space.

Theoretical Foundations and References

The theoretical foundations of Bourdieu's theory of praxis are mainly the works of Durkheim, Marx, and Weber. From Durkheim, Bourdieu adopts the notion that the social world is a reality for its own and therefore an independent research topic of social sciences. In addition, he follows Durkheim's maxim that the social can only be explained by social issues. With his theoretical alignment, Bourdieu tries to take hold of what becomes apparent as an independent social reality beyond subjective thinking and beyond the biological nature (of humans). In his line of reasoning, however, he turns away from Durkheim claiming that Durkheim's arguments and methods are based on a naive realism. This naive realism eventually leads to a social mechanism that reduces social reality to categories (classes and groups) constructed by social theorists. Studying Marx and Weber helped Bourdieu to overcome Durkheim's 'rules of sociological method'.

Following Marx, Bourdieu argues that the observable order of the social world at a given point in time is not the result of a causal mechanism but the historically contingent product of practical constructions. Social differences (classes, groups, ethnicities, cultures, and gender) resemble the current state in the struggle for distinction – a struggle about and with the help of pic tures, imaginations, and symbolic representations. Con sequently, the analysis of social processes and the formation of current social classifications require a his torical perspective. The production of the social world is conceived by Bourdieu as a meaningful process, that is, a process of producing meaning. He draws this idea from Weber's famous definition of sociology according to which sociology should analyze the subjective meanings of social actions in order to interpret and explain it. Bourdieu's theory of praxis is a theory of the meaningful production of the social world, too. It has, however, to be distinguished between what seems to be the meaning of actions from a socioscientific perspective and what is in fact the practical meaning of actions and utterances. Bourdieu basically agrees with Weber in maintaining that actors allocate a subjective meaning to their actions. At the same time, it has to be taken into account that actors do not reflect this meaning component at all times but practically master it. That is to say, actors reproduce meaning even though they do not discursively reflect on it.

The theoretical foundations of Bourdieu's theory can be summarized as follows: (1) social reality, the research topic of social sciences, is not to be reduced to either individual states of mind or to natural determinants; (2) the social world is the historically contingent product of an ongoing struggle for representation; it is the result of the symbolic production of differences and distinctions; (3) constructing and implementing principles of distinction is an attempt to determine the legitimate view of the social world; and (4) social sciences include themselves as a part of their research object.

Basically, Bourdieu holds a rather traditional understanding of science that stands out from postmodernist reactions to the crisis of representation. Of course, he assumes that truth is not found but made. That, however, does not lead him into an uncritical relativism. On the contrary, he concludes that if truths are being made, the means of truth production, that is, the scientific praxis of observation and representation, have to be scrutinized closely. That is the reason why, according to Bourdieu, socioscientific research is fundamentally reflexive. He repeatedly stresses that social sciences have to direct their findings also to themselves. Socioscientific methods of investigation are to be applied to the social praxis of science as well. That is to say, social scientists have to make themselves the object of their research. In doing so, they grant themselves an epistemic privilege: the position of an observer who observes himself while observing and who is capable of questioning and steering his observation practice. What makes socioscientific observation special is, thus, not the fact that social scientists observe others but rather that they include themselves in the observation. They constantly question the unquestioned plausibilities of their own descriptions.

Social Space, Capital, and Field

Bourdieu's theory of praxis bears certain similarities to Giddens' theory of structuration. Nonetheless, there is a major difference between the two. Giddens puts his emphasis on social praxis – on what actors do and think in their everyday life – and on the constraints and freedoms of individual action. Giddens' theory of structuration is an attempt to overcome the opposition of voluntarism (freedom) and determination (constraint) by regarding both aspects as moments of praxis. Bourdieu, for his part, discusses the opposition of objectivism and subjectivism, which is not exactly the same since objectivism and subjectivism are not aspects of social praxis but moments of its analysis. He is less interested in delimiting the scope of individual freedom of acting or finding out how strong agency is restricted by social constraints. What he wants to foreground and eventually overcome is the opposition of two types of socioscientific observation of social praxis. He holds that both modes of social scientific analysis, subjectivism and objectivism, depict social praxis one sidedly and distortedly. It is Bourdieu's aim to unveil these distorted scientific viewpoints in order to eventually establish a scientific understanding of social practice and of practical understanding. Criticizing objectivistic and subjectivistic modes of insight is, thus, the starting point of Bourdieu's conception of a theory of praxis.

Social space

Among the concepts that form the core of Bourdieu's theory, social space is the most vital term. Especially in geographical contexts, it has to be clarified immediately what the term does not mean. Bourdieu does not use it to describe (a part of) geographical space. Social space is neither a territory nor the 'habitat' of individuals or groups. Social space cannot be found as a structure on Earth's surface. The term is used to designate the social world or society in general. However, it does not refer to a container or some kind of box but rather to a relational formation of positions. Social space is a relational space, that is, an arrangement of social positions that are determined by relations of proximity and distance. The multiple dimensions of social reality are displayed by means of different forms of capital. Besides economical capital Bourdieu also introduces social and cultural capital. These three forms of capital constitute the main coordinate axis of social space, that is, the determinants by means of which positions in social space are defined and allocated.

Capital

Bourdieu's usage of the term 'capital' does not correspond to the meaning of the word in colloquial language or other social theories where it generally denotes monetary or material value. In Bourdieu's sense, it rather designates structural principles that are brought to bear in interactions between social actors in different fields of social space. Above all, it is used to describe the actors' abilities deriving from the structural conditions of their situation, especially the means of power at their disposal, that is, the power over products, persons, and meanings. This notion of capital includes also the conventional economic meaning (money, income, property) of the term capital.

Social capital can be conceived as an authoritative resource in Giddens' sense. It is, as a form of controlling persons and their actions that rests on established social relations, like employment contract or amicable commitments. Cultural capital designates knowledge and skills that are acquired through education. However, cultural capital covers more than school and university education. It encompasses all sorts of competences in dealing with signs and semantics. That is to say, it includes the ability to participate in specific language games in various contexts and to fulfill the corresponding expectations. Besides these main determinants of social space, Bourdieu occasionally mentions other forms of capital that can become relevant in specific social contexts: for example, symbolic or scientific capital. According to Bourdieu, the different types of capitals are only partly interchangeable among one another. In addition, there is a struggle for the value of capital in social space and for the exchange rate of various sorts of capital. This struggle is closely bound to the division of social space into different fields.

Fields

According to Bourdieu, the social space is divided into different fields with each field displaying its own social microcosm. Each social field has, as Bourdieu stresses, its own logic. It is characterized by specific principles of differentiation and representation that define how to ascribe positions, interpret events, and deal with problems. Each field has a specific mode of labeling positions, that is, a specific logic of distinction. That is the reason why a field appears to be a more or less autonomous social universe. With the conception of social fields, Bourdieu's theory of praxis takes the differentiation of social systems into account. The social fields correspond to the subsystems of society, that is, economy, politics, science, art, or religion. Each field also represents a specific viewpoint from which social reality is being observed and constructed.

Habitus and Sense of Practice

According to Bourdieu, the principles of distinction within a given field are reflected in the minds (and bodies) of the actors within this field. According to the field's logic of distinction, actors develop schemes of perception and of acting. These schemes of perception, thought, and action are called the actor's habitus. Bourdieu argues that the habitus is a product of conditionings that are bound to a certain positioning in social space. Consequently, the habitus is as differentiated as the positions whose products they are. Nevertheless, it has to be added that the habitus is also differentiating as it contains those interpretation schemes that help to comprehend the meaning of utterances and objects and reproduce social practices. Hence, the habitus has two sides. On the hand, it is a system of structured expectations and a scheme of interpretation, on the other it is the generative principle of symbolic practices that reproduce social structures.

It has to be noticed that the habitus is acquired in and through praxis and not through reflected learning. The habitus evolves from processes of incorporation beyond consciousness and discursive thinking. The application of the habitus, its realization, also takes place without reflexive contemplation. The majority of our powers of discernment, our competences in dealing with social situations, our comprehension of the meaning of utterances and objects is routinely applied. We have a practical mastery of the valid rules and the social habits. This bears a certain similarity to the practical consciousness in Giddens' sense. That is to say, the habitus is a kind of knowledge or skill that does not have to and often cannot be readily articulated. In this regard, Bourdieu speaks of an almost bodily anticipation of the requirements in social situations within certain fields. Thus, social differentiation does not only show up in the consciousness of the actors as schemes of thought and acting but is also more or less directly transferred to the body as it manifests itself in the posture and gesture.

With the concept of habitus, Bourdieu turns against the idealistic conceptions of agency as outlined by certain action theories (e.g., rational choice or ethnomethodology). These theories conceive actions as products of calculation or conscious planning and decisions. Bourdieu on the contrary holds that there is a practical sense and practical mastery of social praxis that is not connected with this kind of rationality. Bourdieu metaphorically calls this practical sense a sense for the game. Here, game designating the (serious) social life in general.

The fields of social space are, in a way, the playing fields with their corresponding rules. The habitus enables actors to intuitively play the game and to reproduce it (mostly rather unquestioned). According to Bourdieu, there is a direct connection between field and habitus. The habitus is adjusted to the principles of distinction that prevail in a field. It represents the social conditions under which it is acquired. Vice versa, structures of social space (the fields and their logics) are produced and reproduced by means of symbolic practices that are generated by the dispositions of this habitus. It must not be concluded, however, that social conditions are always reproduced uniformly. People who are involved in a social field do not perceive its presets and constraints as an outer constraint. The practical meaning is also a kind of a practical belief in the game. It ensures that the majority of what happens within the field is taken for granted. That does not imply that social actors agree with everything that happens in their social surrounding. It just expresses that they are truly involved in what they do.

Social Space and Geographical Space

Bourdieu's discussions of geographical space and specific geographical issues of the relation between society and space are manifold. He discusses this topic (without referring to geographical research) in his early studies about the kabylic society and rural life in Béarn (France). However, Bourdieu's notion of geographical space is less elaborated than that of comparable geographical studies. The concept of habitus defined as a structure of the social world that is inscribed in the bodies could provide access to a more detailed discussion of Bourdieu within geography. This view implies that social life is, in a fundamental way, body centered. Thus, it points to the spatial dimension of social praxis.

In a short paper entitled 'Effets de lieu' (effects of place), Bourdieu explicitly discusses the notion of geographical space. In this paper, Bourdieu analyzes the relation of social and geographical space. He notices that social space inscribes itself more or less exactly into physical space. Social space manifests itself in certain positionings of actors and properties in physical space. That is to say that, for example, the societal distribution of economic and cultural capital becomes apparent in the geography of a city, be it in the concentration of rare goods or their possessors at certain places or in the evolvement of regions where the poorest live (suburbs, ghettos). In social praxis, physical space always appears as a socially appropriated space.

Besides economic capital, the appropriation of space always requires social and cultural capital. Vice versa, the place occupied or appropriated by a social actor (or a group) can in turn yield a 'space profit'. Such profits are expressed by, for example, control over space that allows to keep disturbing or unwanted things or persons at a distance. Profits can also arise from the proximity to rare and treasured goods or facilities. Hence, the appropriation of desired spaces does not only require capital, it also pays it back to all those who gain excess to the space endowing them with additional symbolic capital. Reversely, this 'club effect' works as 'ghetto effect', for example, discriminating people because of their origin.

Another consequence of permanently inscribing social reality into the physical world is the tendency to naturalize social relations. Spatializations of social and cultural differences can mistakenly give the impression that social and cultural differences stem from the very nature of things. That is the reason why spatializating social and cultural issues represent traps for scientific observers. If the rash observer takes them for granted he inevitably gets caught in a substantialistic or realistic line of reasoning that denies the social production of differences and entities. Bourdieu additionally warns against an unnoticed inversion of cause and effect, which happens whenever the actor's position in social space is inferred from their localization in physical space. The scientific observer who ignores the social construction of the meaning of places becomes entangled in a naturalizing discourse.

Discussion

The critical discussion of Bourdieu's approach is just as extensive as his work itself. The objections raised against single arguments and concepts of his theory derive from different subjects and are thus manifold. The main critique, however, aims at the concept of the habitus. For quite some time, it is (repeatedly) argued that the idea of the habitus is founded on a reductionistic view of social praxis. The close connection of structures of social space and schemes of action and thought (habitus) has raised the objection that Bourdieu has a deterministic understanding of praxis. Other critics, however, claim quite the opposite. Especially Marxist scholars object that Bourdieu's concentration on symbolic practices of distinction (culture) underestimates actual socioeconomic settings and ignores structural conditions. Bourdieu, however, wants to emphasize that structure and agency cannot be separated. Indeed, social space and habitus cannot be thought of without the other. This is well illustrated by the game metaphor. On the one hand, we see the 'game', that is, the social world with its rules and conventions, whenever we analyze the social space with its fields. On the other hand, we discover the intentions and the strategies of players, if we focus on the social actors with their knowledge and attitudes. However, the two sides cannot, strictly speaking, be considered independently. Without the game, that is, without rules and institutions, there are no players and without the players applying the rules and playing the game, there are no games. That is the reason why the actual object of social sciences is, according to Bourdieu, neither the social actor (the individuals) nor the social world (the collectives and the groups) or its structures but the relation between the habitus and social space.

See also: Cultural Capital; Deconstruction; Determinism/Environmental Determinism; Ethnomethodology/Ethnomethodological Geography; Evolutionary Algorithms; Functionalism (Including Structural Functionalism); Habitus; Homelessness; Longitudinal Methods (Cohort Analysis, Life Tables); Military Geographies; Nature-Culture; Phenomenology/Phenomenological Geography; Place; Poststructuralism/Poststructuralist Geographies; Pragmatism/Pragmatist Geographies; Quantitative Data; Rational Choice Theory (and Rational Choice Marxism); Regionalisations, Everyday; Society-Space; Space II; Structural Marxism; Structuralism/Structuralist Geography; Structurationist

Geography; Symbolic Interactionism; Time Geography; Transport, Rural; Tuan, Y.-F.

Further Readings

Bourdieu, P. (1972). *Esquisse d'une théorie de la pratique: Précédé de trois études d'ethnologie kabyle*. Genève: Librairie Droz.

Bourdieu, P. (1980). *Le sens pratique*. Paris: Les éditions minuit.

Bourdieu, P. (1993). Effets du lieu. In Bourdieu, P. *et al*. (eds.) *La misère du monde*. Paris: Éditions du Seuil.

Bourdieu, P. (1994). *Raison pratique: Sur la théorie de l'action*. Paris: Éditions du Seuil.

Bourdieu, P. and Wacquant, L. J. D. (1992). *Réponses Pour une Anthropologie Réflexive*. Paris: Editions du Seuil.

Bryant, C. G. A. and Jary, D. (eds.) (1991). *Giddens' Theory of Structuration: A Critical Appreciation*. London: Routledge.

Bryant, C. G. A. and Jary, D. (eds.) (1997). *Anthony Giddens: Critical Assessments*, 4 Vols. London: Routledge.

Cohen, I. J. (1989). *Structuration Theory: Anthony Giddens and the Constitution of Social Life*. London: Macmillan.

Craib, I. (1992). *Anthony Giddens*. London: Routledge.

Giddens, A. (1976). *New Rules of Sociological Method*. London: Hutchinson & Co.

Giddens, A. (1979). *Central Problems in Social Theory: Action, Structure and Contradiction in Social Analysis*. London: Macmillan Press.

Giddens, A. (1984). *The Constitution of Society: Outline of a Theory of Structuration*. Cambridge, London: Polity Press.

Giddens, A. (1990). *Consequences of Modernity*. Stanford: Stanford University Press.

Held, D. and Thompson, J. (eds.) (1989). *Social Theory of Modern Societies: Anthony Giddens and His Critics*. Cambridge: Cambridge University Press.

Kaspersen, L. B. (2000). *Anthony Giddens: An Introduction to a Social Theorist*. Oxford: Blackwell.

Lippuner, R. (2005). *Raum Systeme Praktiken: Zum Verhältnis von Alltag, Wissenschaft und Geographie*. Stuttgart: Franz Steiner Verlag.

Painter, J. (2000). Pierre Bourdieu. In Crang, M. & Thrift, N. (eds.) *Thinking space*, pp 239 259. London: Routledge.

Thrift, N. (1983). On the determination of social action in space and time. *Environment and Planning D: Society and Space* 1, 23 56.

Werlen, B. (1995). *Sozialgeographie alltäglicher Regionalisierungen. Band 1: Zur Ontologie von Gesellschaft und Raum*. Stuttgart: Franz Steiner Verlag.

Werlen, B. (1997). *Sozialgeographie alltäglicher Regionalisierungen. Band 2: Globalisierung, Region und Regionalisierung*. Stuttgart: Franz Steiner Verlag.

Relevant Websites

http://www.theory.org
Anthony Giddens, Theory.Org.

http://www.lse.ac.uk/Depts/global/stafflordgiddens.htm
Centre for the Study of Global Governance Lord Giddens, London School of Economics and Political Science.

http://www.hyperbourdieu.jku.at
Pierre Bourdieu.

http://www.hyperbourdieu.jku.at/
Pierre Bourdieu Sociologue Énervant, Le Magazine de l'Homme Moderne, Bienvenue.

Structurationist Geography

B. Werlen, Friedrich Schiller University of Jena, Jena, Germany

Glossary

Geography-Making The core concept of praxis-centered research addresses all forms of constitutional processes of geographical realities and related geographical imaginations of the world. It encompasses the analysis of all kinds of singnificative/symbolic, normative, and rationale appropriations of spatial facts as well as the transformation of nature and the productions of spatial patterns of material artifacts (infrastructures, settlements, urban networks, etc.) on the basis of this different forms of appropriation and power relations in form of rules (schemes of meaning attribution and interpretation) and (allocative and autoritative) resources.

Locale/Setting It denotes a spatial context of action as a setting for action constituted by material elements as well as by sets of social norms and culturally shared values, to be understood as a material, socio-economic and socio-cultural constellation of action with intersubjective shared meaning contents. Therefore, it is an action-related concept that cannot be turned into an objective fact or a generalized as a social category, having the same meaning for members of a society (in a certain region).

Place It denotes a position in a spatial continuum with a certain specific meaning, expression of processes of appropriation. These processes can, for instance, be based on economic actions, including value attribution in the sense of the 'central place theory', on normative actions related to territorial regulations or to symbolic-emotive appropriations, expressed by feelings of belonging.

Regionalization It is traditionally understood as an academic practice of spatial delimitation of natural, social, economical, cultural, or political spheres of reality. According to the perspective of structurationist human/social geography, it is understood as an everyday practice, not for delimitating the 'world' spatially, but rather using spatial references for the economical, political, cultural etc. structuring of social realities. Therefore, the key issue is world-binding for the performance of human or more precisely: social agency as forms of appropriations on the basis of spatial concepts.

Spatial Disembedding The ontological pre-condition of globalization, based on agency or action orientations strictly separated from time in the sense that the 'where' and 'when' of acting is open to negotiations and not anymore tied to each other by tradition. The interrelated 'lifting out' of late-modern institutions and of social relations in general from local contexts and their re-formation across indefinite spans of time-space is called time-space distanciation. As major disembedding mechanisms are regarded symbolic tokens (like writing, money, credit cards, etc.) and expert systems (infrastructure of transportation, electronic communication etc.), enabling actions over distance in real time, provoking as well an enormous acceleration of speed of life and social transformation as a crisis of the nation-state.

Spatial Embedding The ontological condition primarily of traditional societies and cultures, based on the unity of space and time, enclosing a local and regional anchoring of social institutions and religious practices, very often linked to powerful processes of reification mythification. One of the major characteristics lies in the dominance of face-to-face communication, and the nearly complete absence of space-time distanciation. These two characteristics encompasses a rather narrow spatial clustering of traditional societies and cultures, expressed by quite clear spatial borders of local and regional traditions and cultures, study object of traditional regional geography. Under conditions of globalization 'spatial embedding' is the product of re-embedding activities of in principal dis-embedded social and cultural relations.

World Binding Key concept of action/agency-centered human/social geography, based on the shift from a space-centered geographical imagination to the analysis of the everyday geography-making. It expresses the basic assumption, that the postulation of the study of societies and cultures in space or the study of space includes to a certain extend the acceptance of a (pre-modern) concept of an absolute and material (container) space. An action/agency-centered perspective develops a worldview in which the embodied subjects form the nucleus of geographic research, analyzing how they bring the 'world' to themselves under spatially dis-embedded conditions (space-time distanciation/globalization), on the basis of everyday regionalizations, specific forms of appropriations and specific power relations.

Zoning It denotes further differentiation of regions for specific activities in spatial as well as temporal respect. Spatial specifications of the region 'house', for instance, include the distinction of 'living room' (meeting other persons), 'kitchen' (cooking, eating), the temporal distinction, for instance, the separation of daytime (living room and kitchen in the basement) and nighttime (sleeping rooms on the first floor). Zoning involves activity-related differentiations at all possible scales, from one single room to global scope and time unites, from seconds to day, seasons, years, etc.

Structuration Theory and the Spatial Contexts of Social Agency

In his theory of structuration, Giddens develops a specific vocabulary to overcome one of the most problematic shortcomings of the classic social theories of Marx, Durkheim, Weber, and others, thereby integrating space and spatiality into the theory of social practice. The three spatial key concepts are: 'locale/setting', 'region', and 'zone'.

The first term is 'locale'. Giddens explicitly differentiates between this and the geographical concept of 'place', in order to show that the contextuality of actions, not the physical realm, is central. 'Locale' refers to a piece of space used for a specific activity that is already organized by pregiven material interactions. In short, we can understand 'locale' as a material context, or the constellations of actions which acquires a specific social meaning intersubjectively as well. Giddens occasionally uses the term 'setting' as in 'setting for interaction' for this. But setting is by far, not just a spatial parameter or a physical containment of social interaction. 'Setting' has rather to be understood as an element that is mobilized as an inherent part of interaction.

An example is a house. A locale can only be called a 'house' if we observe that this material context has the furnishings for living in and is treated in that way. According to the context of action, a locale can be a house, a street corner, a city, or the territory of a state. Consequentially, the determination that a constellation is a 'locale/setting' does not depend on the size or spatial dimensions, but rather from the way in which actions are carried out.

Every locale is 'regionalized' in space and time. A house, for example, has floors, hallways, and rooms. Regionalization thus means a first specification of the 'social definition' of certain spatial slices or locales 'in relation to certain ways of acting'. This does not foreground the physical, worldly expression or the material characteristics of a region in the (traditional) geographic sense. It is, again, a combination of the social and spatial categories or characteristics.

Seen in this light, a region should be understood within a 'locale/setting' as a social part of the situation or the context of action delimited by symbolic marking which can be ascertained by physical and material relations (walls, lines, rivers, valley, etc.). A region within a locale refers 'to aspects of the settings which are normatively implicated in systems of interaction, such that they are in some way 'set apart', for certain types of activities.' 'Regionalization' can be further differentiated according to type, length of time, and character. Determining the 'form' of the regionalization refers to the type of borders that define the region and can be determined by physical or symbolic markers. Seen physically, this can mean the walls between individual rooms or mountain chains between larger pieces of space. Seen socially, these translate into understandings that one does not sleep in the kitchen, or that the borderlines of regional differentiation represent significant social characteristics of spatially positioned persons of a population, such as between the 'North' and the 'South' of England. This means that regions can differ in their spatial and temporal extents.

Consequently, the intersection of 'spans' of space and time may be variable. But the social maintenance of regions of a bigger span (like nation states, international markets, etc.) necessarily depends – and that is a consequence of the social nature of regions and borders as elements of social action – in general upon a high degree of institutionalization. Only on that basis, a stabilization in the form of social reproduction is possible.

By the 'character' of regionalization, Giddens refers to 'the modes' in which the time–space organization of locales is structured within an embracing social system. 'Modes' can be understood as the overarching spatial organizing principle, which governs a certain encompassing social system. As an example, preindustrial economic systems rarely led to sharp divisions between familial and work life. Already in industrial societies, there is however a clear differentiation between living space and workspaces. The corresponding organizing principle here can be characterized as a 'functional division' between and within production and reproduction zones.

'Zoning' refers to the further differentiation of regions. This means that, for example, specific rooms are only defined for specific activities: the living room is for meeting and receiving guests, the kitchen is for cooking and perhaps also for eating, and so on. Within production facilities, this could be zone of production, administration, and rest.

Regionalization and zoning are mostly combined with time, such that time–space regionalization and zoning result, which can be seen in an English house, is as follows: the upstairs includes the rooms which are used for activities at night, the ground floor for daytime activities. Temporal zoning, however, is also central in a more encompassing manner. The clear division into day and night can be held for the fundamental differentiation of social life into zones of exertion and relaxation in all societies, and the needs of the human organism for regular periods of sleep are governed by it. In this way, 'nighttime' forms a very clear 'border' for social activities as much as any spatial boundary. In other theaters, zoning occurs according to weekly, monthly, or even yearly rhythms.

The Integration of the Theory of Structuration into New Regional Geography

The first attempt to integrate the spatially enlarged social theory of structuration into the geographical worldview has been undertaken by Derek Gregory in his project of a critical new regional geography. The main characteristic of that attempt lies rather in the integration of a social theoretical perspective into regional research than in a critical transposition of a metadisciplinary position into a new disciplinary framework.

Unlike traditional geography, Gregory's regional geography – based initially on Habermas' writing and later on Giddens' critical theory – is not founded on physical geographical conditions but on criteria that are derived from social or structuration theory respectively. Starting with the question of how societal realities are regionally formed and articulated, Gregory then wants to pave the way for appropriate regional political or scientific interventions. The major thesis of the project is that spatial structures are somehow incorporated in social structures and vice versa.

The empirical regional study 'Regional transformation and industrial revolution' is the first to encompass a systematic geographical discussion of Giddens' structuration theory (see **Figure 1**).

The transformation comprises three steps of decoding: In the 'first step', the central stages of transformation and the transformation power of regional realities are identified. The changes in the wool industry are regarded as expressions of structuration processes, that is, as results of transforming social reality by means of rule governed human action. Being differently equipped with authoritative and allocative resources, actions are seen as the basis for the reproduction of social structures.

In the 'second step', the transformation of labor into capital and the consequences of this structuration process are empirically analyzed. Changes in the labor process encompass, above all, technical innovations and evolving routinization of work steps, gender specific division of labor, chronic unemployment, vanishing of the inviolability of landowners, etc. All together, these processes are to be seen as expressions of the oncoming convertibility of private property, which turns labor into a commodity and thus working hours into objects of exchange. To explore these changes in their regional appearance is the second task of the 'new regional geography'.

The 'third step' commences with an analysis of the time–spatial consequences of the latter two aspects. Subsequently, it should be reconstructed how time–spatial aspects of action help to 'prevail' what has been singled out in the first two steps of decoding, that is, the identification of the transformative power and the regional transformation analysis. Analyzing the changing rhythms of daily and annual activities within the spheres of household and workplace (which are separated under the capitalist regime) is a central aspect of structuration processes and geographical regional research. It is assumed that this casts a light on the regional transformations of societal realities.

These three steps together with an analysis of the theoretically inferred link between the three subject areas should allow two things: first, a deeper understanding of uneven regional developments, second, an insight into the 'mechanisms' of the spatial differentiation of society, especially the varying regional constitution of class relations.

Alan Pred's reference to Giddens' structuration theory is embedded in his interest in Hägerstrand's time geography. Above all, Pred wants to reconstruct the historical becoming of places and regions as expressions of structuration processes. Consequently, a given region has to be

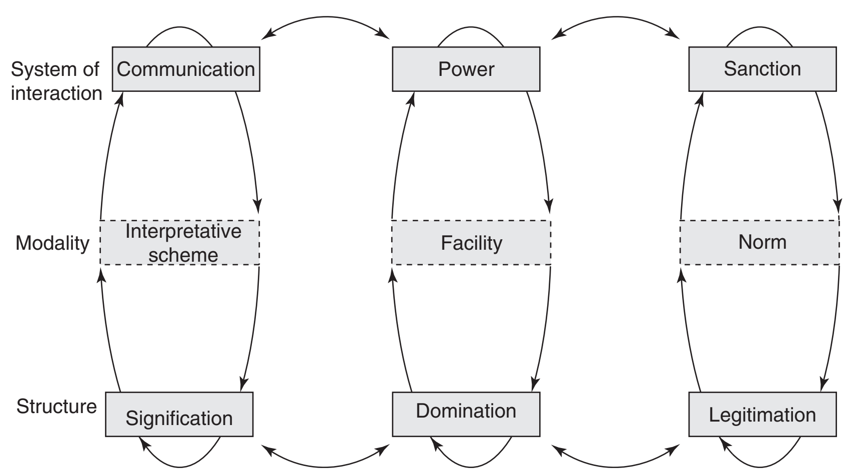

Figure 1 Gregory's reference to the structuration theory. From Gregory, D. (1982). *Regional Transformation and Industrial Revolution: A Geography of Yorkshire Woollen Industry*, p 17. London: Macmillan.

understood and analyzed as the result of transformation processes. The objective is to examine regions within the analytical frame of a structurationist 'theory of place'. In this context, the structurationist comprehension of 'place' and 'region' becomes important.

For Pred, the becoming of 'place/region' always im plies two social processes: the appropriation of 'space' and the transformation of 'nature'. Every place or region is an expression of these two processes (**Figure 2**). The historical becoming of a 'place' or 'region' by means of appropriating 'space' and transforming 'nature' is, as Pred stresses, inseparably bound to the reproduction and transformation of society in space and time. Accordingly, the process of societal transformation and reproduction is bound to the becoming of 'place/region' by means of the appropriation of 'space' and the transformation of 'nature'.

Consequently, the distinctiveness or uniqueness of every place and every region is an expression of the agent's historically variable interpretation of these pro cesses under historically specific, contingent conditions.

The 'theory of place' comprises three main points: first, the 'institutional projects'. Their shaping has an influence on daily routines and paths of the participants of these institutions. This shaping of routines and paths has strong implications for the place; additionally, they leave traces in the landscape and they influence power relations. As a consequence, social structures in Giddens' sense become particularly important.

Second, Pred suggests examining the formation of biographies as an expression of structuration processes 'within' a place; that is, to say, the influence of the place/region on a person's life. Thus, place/region is regarded to play a key role in the formation of biog raphies. Regional studies should disclose this constitutive power.

Third, it is assumed that there is a meaning of place that – although it does not exist as an independent entity – is part of the growing consciousness of an individual: according to Pred, the meaning of place is to be treated as an object of regional research just as the constitution of biographies and the consequences of regional institutions are. At this point, it becomes even more obvious than in the two preceding points that what is analyzed as re gionally constituted cannot be the object of regional re search *per se*. If the meaning of place is said to be a mental issue, it can just not be material at the same time and therefore not be located topographically. It is a mental or social entity that can be extremely relevant for regional processes without having a regional existence. Con sequently, it can be addressed as a crucial element for the process of regionalization, but not as a regional object. This shows that the application of structuration theory and social theory in general to social geography cannot lead to regional geography, but rather to the investigation or processes of regionalization, what in fact is the original intention of Pred, but is not what he suggested.

In the early 1980s, Nigel Thrift successfully in corporated structuration theory into an outline of con textual regional research. The key concepts of this reorientation of regional research are 'locale' and 'social action'. With reference to Giddens, locale designates a certain portion of physical space that already displays a specific pattern of material objects and persons. That is, to say, locale is the material context, the given constel lation for action and interaction. This constellation is also labeled 'setting'. In order to scientifically demarcate a region as a spatial interaction structure, an in depth understanding of interaction is necessary. Interaction is interpreted as an activity linking social and spatial structure by means of institutional interrelations. Con sequently, a 'region' cannot be regarded as a pre given

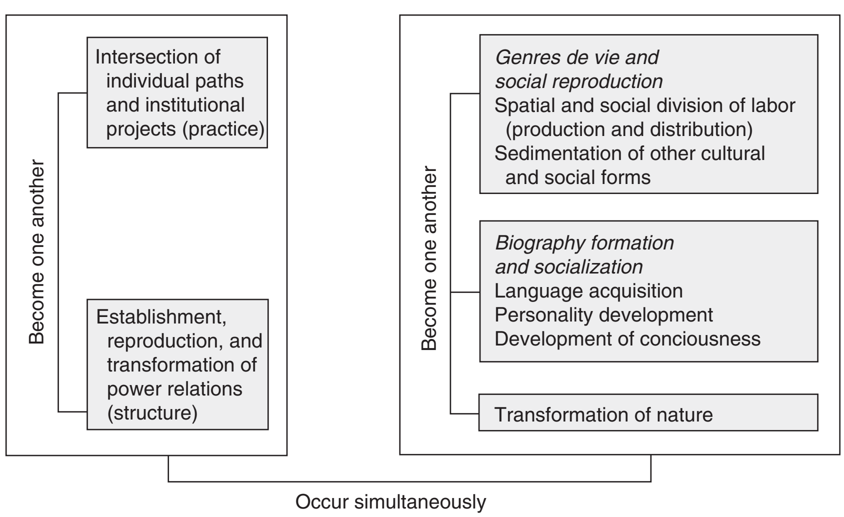

Figure 2 Components of 'place' and 'region'. From Pred, A. (1986). *Place, Practice and Structure*, p 11. Cambridge: Polity Press.

place or corporeal entity. Instead, demarcating a region should be the objective of scientific research. It is to find out which 'settings' of interactions are to be joined in order to form what then can justifiably be called a 'region'.

Every region is, above all, to be understood as enabling and restricting actions. Both options decisively influence the life paths of the individuals living there. Consequently, the structure of the arrangement pattern of 'locales' is regarded as an expression of the social structure. In this regard, the varying importance of different settings is crucial. The organization of production determines which settings become dominant. They are characterized by the fact that time 'must' be allocated to them.

Discussion

The aim, however, is to outline a conception of geographical research that, on the one hand, is not only social theoretically informed but in a consistent way compatible with the basic principles of social research and, on the other hand, aims at a more thorough incorporation of spatial and temporal contexts of social action. This aim is not fully achieved by first attempts to integrate structuration theory in geographical world perspective. The main problems are insufficient theoretical reflections of the geographical concept of Earth space and a misapprehension of modern and late modern social, ontological conditions. These, however, have to be taken into account, if human and social geography wants to adequately represent late modern geographical conditions and consequently be relevant for everyday life. Before the problematic aspects of the adaptation of structuration theory to geography are to be discussed, it should also be mentioned, that some of the identified problems are integral to structuration theory itself.

The main problem of this adaptation lies in the fact, that 'places' as well as 'regions' are considered in the existing shape, or even regarded as historically pregiven and fixed. Human actions are analyzed 'in' regions and not as the region building forces. That is to say, action is not considered as a constitutive power for the social formation of place and region. In the course of establishing a new regional geography, the core principle inspired by structuration theory: 'people produce history and places' moves toward a quite different governing principle: 'people are produced by history and places' implying that times, spaces, or places can constitute meanings and have a socializing effect. Consequently, Pred suggests that places/regions do exist as such – but only as a constant flow of human practice. The latter, however, would imply that place/region can only be seen as an element of social praxis/action, but not existing as such in the form of a spatial reality in itself.

Giddens proposes that analyzing the processes of regionalization could be the specific, most important, or even legitimizing task of contemporary geography. This notion is reformulated in Gregory's, Pred's, and Thrift's adaptations to a research program of the constitution and transformation processes of historically pregiven political regions. In doing so, they miss recognizing the social meaning of regionalization for the reproduction of social realities. Then the program shows clearly that, still, the explanation of a given geographical phenomenon forms the center of scientific interest. All together, the Anglo-Saxon reception of structuration theory lacks a comprehensive embeddedness in social theory. Instead, structurationist geography should rather interpret the signification of geography making as a core dimension of the constitution process of social and cultural realities as well as of the reproduction of power. Moreover, it should reflect social theoretical implications of space conceptions and late modern living conditions.

In this regard, it is not sufficient to apply structuration theory to regional analysis. But also for a productive application to the social scientific analysis of everyday geography making, the theory has to be revised radically. Its internal contradictions have to be overcome for appropriate geographical research. Above all, this excludes regarding 'space' and 'time' as special powers of the constitution of society. Giddens' problematic incorporation of 'space' has to be revised in the sense that, it can be made compatible with the core principles and assumptions of structuration theory. In short, a more radical implementation of the structurationist perspective into human geographical research has, first of all, to clarify the current geographical condition, the current ontology spatial and temporal relations and its relevance for the constitution of sociocultural realities; second, to theorize 'space' in the context of social action or agency; third, to put the processes of regionalization in the center of geographical research interest, and finally, to redefine processes of regionalization under globalized conditions of local actions.

Late-Modern Ontology and Structurationist Geography

If we start from the premise that all kinds of spatial relations are constituted historically under certain social and cultural conditions, we have, above all, to spell out the basic principles of that interconnection. For this, the construction of ideal types should allow to identify certain characteristic aspects in the interrelation of the sociocultural and spatial. The first question, then, is: what conditions have to be given or have to hold, if the spatial description of societies and cultures should be possible, as required by geographical methodology? The first hypothetical answer to this question is that traditional life forms come very close. The most important spatio temporal characteristics of traditional life forms and regional societies can – in ideal typical form – be summarized as shown in **Table 1**.

Stability over time or embeddedness in a temporal respect is based on the domination of local traditions. Traditions connect past and future and are the central frame of reference for action, orientation, and legitimization in daily praxis. They set narrow bounds for individual decisions. Social relations are predominantly ruled by kinship , tribe , or rank relations. Depending on birth, age, or sex, clearly defined social positions are attributed to persons.

Narrow spatial limitations, or 'embeddedness' in a spatial sense, lie in the technical standards of transportation and communication. The predominance of walking and the limited significance of writing restrict social and cultural expressions to the local and regional level. Face to face interaction is almost the only possible situation for communication. Additionally, production processes have to respect natural conditions because of technological development. Economies are, consequently, highly adapted to the prevailing physical conditions.

In addition – as many anthropological studies teach us – temporal, spatial, and sociocultural aspects of everyday praxis are closely bound together. For traditional life forms, it is important not only to carry out certain activities at a certain time, but also in a certain place and sometimes even with a certain spatial orientation. In this way, social regulations and activity patterns are reproduced and enforced by unreflexive spatiotemporal commitments.

The unity of sociocultural and spatiotemporal dimensions of activities becomes the basis for extremely powerful reification processes. In this way, for example, places of worship are identified with the act of worship. Only in this way is it possible to claim that somebody who puts his or her feet on a certain place also desecrates that place. But this can only appear as a meaningful phrase, if there is no distinction made between significance and place. To put it another way: only when the significance is seen as a quality of the place itself and not as a product of the subjects' constitution process, only then is it possible to talk about desecrating places. Based on this process of reification, 'space' and 'time' are loaded, are filled up with specific meanings. Signification appears as a quality of things, deeply rooted in them and embedded in the territory of a given culture. Under such conditions, regional geography consequently appears a meaningful scientific project. Under late modern conditions, however, the project of traditional as well as 'new regional geography' aiming at spatial representations of life conditions and of life forms is confronted with serious limitations.

'Late modern life forms' are the basis and expression of the globalization process. Here, traditions are not at the center of daily social praxis. Social orientations and social actions need – as a consequence of modernity – discursive justification and legitimization. The dominant life contexts are spatially and temporally 'disembedded'. The disembedding mechanisms are grounded in the history of modernization. These mechanisms are at the core of the transformation of the space–society nexus, leading today to the globalization of our life conditions.

Temporal stability is replaced by constant social transformation. Late modern everyday actions are not dominated by local traditions. It is rather routines that sustain ontological security. For individual decisions, a wide field of possibilities stays open. Globally observable cultures, life styles, and life forms (**Table 2**) – very often linked to a specific generation – become much more important. A person's social position is determined by production and valued work and – following the principles of the enlightenment – not by birth or age, nor by sex or race.

The spatial clustering and embeddedness of traditional social life forms is replaced by global interconnections and disembedding mechanisms. The actual and potential reach of actors is stretched to a global dimension. The most important disembedding mechanisms

Table 1 Ideal type of traditional life forms and regional societies

1. Traditions intertwine past, present, and future
2. Kinship organizes and stabilizes social relations over time
3. Birth, age, and sex determine social positions
4. Face-to-face situations dominate communication
5. Small amount of interregional communication
6. The local village constitutes the familiar life context

Traditional life forms are temporally and spatially embedded.
From Werlen, B. (1995). *Sozialgeographie alltäglicher Regionalisierungen. Band 1: Zur Ontologie von Gesellschaft und Raum*, p 104. Stuttgart: Franz Steiner Verlag.

Table 2 Ideal type of late-modern life forms and globalized societies

1. Everyday routines sustain ontological security
2. Globally observable cultures, life forms, and lifestyles
3. Production and valued work determine social positions
4. Abstract systems (money, writing, and expert systems) enable mediated social relations over enormous distances
5. Worldwide communication systems
6. Global village as anonymous context of experience

Late modern life forms are spatially and temporally disembedded.
From Werlen, B. (1995). *Sozialgeographie alltäglicher Regionalisierungen. Band 1: Zur Ontologie von Gesellschaft und Raum*, p 134. Stuttgart: Franz Steiner Verlag.

are money, writing, and technical artifacts. Means of transportation enable a high level of mobility. Together with individual freedom of movement, this produces a mixing of formerly locally fixed cultures. This multi cultural mixing, combined with global communication systems, enables a diffusion of information and infor mation storing that is not dependent on the corporeal presence of the actors. Of course, face to face interaction still exists as an important situation of communication, but the most substantial part of communication is me diated. Consequently, a spatial characterization of cul tures and life forms is becoming more and more problematic, if not even highly irritating.

Spatial and temporal dimensions do not determine the content of social actions. These dimensions are merely the formal aspects of human activities. They are the basis for standardized metrification and calculation of material facts and the succession of events. Both standardized metrification and calculation form, together with the recognition of the difference between concept and object, the nucleus of the rational constitution of late modern worlds and the end of mystical interpretations of nature.

From 'Space' to 'Action'

'Space', just as all other terms, has undergone a manifold reinterpretation in the course of time. Without entering into a discussion of the history of spatial understandings in philosophical and natural scientific terms, it can be claimed that they too have changed from a rather reifi catory objectivist (Aristotle, Newton) to formalistic subjectivist (Leibniz, Kant, Husserl) understandings.

This change, however, has not been recognized suf ficiently by the social sciences. Although it would be logical to eliminate the objectified choric spatial concept from theories of modern society, it would be equally fatal to completely eliminate 'space' from social theory for it is apparent that numerous aspects of social realities have spatial connotations. With their ambition to take the spatial dimension of human lives seriously, structuration theory as well as time geography were referring at least implicitly to the reificatory objectivist conception of space by Newton. This has a number of problematic impacts.

The basic shortcomings of Giddens' structuration theory is based on the assumption, that 'space' is – in contradiction to 'society' – not a theoretical concept but a pregiven fact, free of any need for theoretical consider ations. This is expressed in Giddens concept of 'locale' and 'region' and is basically the result of his reference to Hägerstrand's time geography and its Newtonian con ception of Earth space, standing in total contradiction of structuration postulation of agent's capability of agency.

In order to do away with this 'blind spot' and the lack of care in the integration of spatial categories into social theory, one must first assure that the concepts of space are compatible with the categories of social scientific analysis. For the purposes of arriving at a starting point for research in theories of social praxis, in general, and the structurationist geography, in particular, this neces sitates the provision of a concept of space which is compatible with action. What, then, can be understood under the term 'space' within the framework of praxis oriented research? To avoid the problematic implication by using 'space' for the construction of somehow pregi ven 'region' or 'locale', it is necessary to redefine it. There are some good reasons to understand 'space' as nothing more than a concept. If so, however, it is admittedly a very unique one.

If one were to take the arguments, which have been brought together in the philosophy of the last few cen turies seriously, then 'space' would consequently appear to be neither an empirical concept nor simply an *a priori* one, but rather a formal classificatory. It cannot be an empirical concept, because 'space' is not an object. It is formal, because it does not refer to any specific, topical aspects of material objects. It is classificatory, because it enables us to describe a certain order of material objects in respect of their specific dimension.

'Space' is not simply an *a priori* concept, because it is actually based on experience. It is not based on the ex perience of a specific, mysterious object known as 'space', but rather on the experience of one's own corporeality, its relation to other realities (including the corporeality of other subjects), and its significance in enabling and re stricting actions.

The claim that different theoretical approaches to action correspond to different approaches to everyday reality implies, that we can distinguish three main types of interpretations in formal and in classificatory respects (**Table 3**) a rational, a normative, and a communicative interpretation. Accepting this, the next step can be pos tulated: The different constitutions of 'space' are

Table 3 Action and space

Action	*Formal*	*Classification*	*Examples*
Rational	Geo-metric, absolute	Calculation	Land market; location theory
Normative	Geo-metric, Body centered	Normative prescription	Nation-state; back-/front-region
Communicative	Body centered	Relational signification	Regional/national identity; regional symbol

From Werlen, B. (2000). *Sozialgeographie. Eine Einführung*, p 329. Bern: Haupt Verlag.

expressions of different interpretations of its formal and classificatory dimensions.

Therefore it may be assumed that, depending on the type of action, both the formal and the classificatory aspect of the concept of space can have a correspondingly unique connotation. This means that it is to be expected that, depending on one's motivations, both the orientation and the classificatory order can vary greatly.

We must now apply both the distinction between different types of action or agency and the consideration of the relative level of significance of spatial dimensions in each respective field of practice to the examination of everyday 'geography making' under various socio ontological and sociohistorical conditions.

Regionalization as World Binding

Paasi has demonstrated how political regions are to be conceptualized from a structuration theoretical perspective: not as container like entities of concentrated power as postulated in traditional political geography but as institutional realities. The institutionalization of a region is to be understood as a process over which spatial–temporal dimensions of acting and thinking are established and *qua* individual and institutional practices constantly culturally, legally, educationally, economically, politically, etc. reproduced. Thus, 'region' becomes conceivable and researchable as an aspect of social action, as an institution, that is, as a part of social reality that is produced and reproduced through regular action, as a facet of the constitution of societal reality.

Instead of socially explaining the spatial as the 'new' regional geography does, Paasi wants to elucidate the meaning of socially appropriated space for the constitution of societal realities. Processes of regionalization or transformation processes of regions can occur at the same time in different spatial and temporal dimensions. The territorial shape of an administrative region is the result of localized political–administrative practices. It is through these practices that a region is demarcated. That is to say, the border is defined by institutional arrangements. Thus, the emergence of (administrative) institutions regulates the reproduction of the societal reality 'region' *qua* regulated interrelations of action. With this approach, structuration theory is successfully employed for analyzing processes of territorialization, but not yet for other aspects of social reality under globalized conditions.

For outlining the different dimensions of globalization, the implications for regionalization processes should be noted first. In geography, regionalization has traditionally been understood as a scientific practice of spatial classification. On the level of everyday actions, 'regionalization' usually refers to a process of political appropriation or demarcation. Both understandings, scientific and everyday, are centered around the spatial demarcation of a 'region'.

The borders of political regions normally consist of physical or symbolic markers. But 'physical markers' are, in a social sense, nothing more than representations of the symbolic demarcation of the reach of normative standards. Accordingly, they can never be social forces, only the normative standards they are. Spatial aspects can consequently never be the causes of or reasons for human actions. They are only socially relevant to the extent that they are utilized in actions: as a means of either social categorization or symbolic representation.

All sociocultural aspects of action that are not bound to territory by means of institutions or organizations may nevertheless not be presented as regional realities in a geographical sense. Their significance for local experience may best be revealed and clarified by analyzing the forms of action. If we accept this claim then it is also true that, under globalized living conditions, a new understanding of 'regionalization' becomes necessary.

In terrestrial terms, regionalization on the everyday level is the expression of a (usually political) production of the ordering system of competences. However, this process is also tied to more comprehensive ordering systems of meaning. The political territorial order in the form of nation states, countries, districts, etc. is a symbolizing expression of the regulation of competences. The symbolic ordering systems of meaning are connected to specific territorial areas.

If one were to generalize this principle, every form of regionalization could then be understood as a sort of 'world binding.' It takes the basic principle of the modern worldview into consideration, which assigns the central roles to the perceiving and acting subjects. 'Regionalization' as a practice of 'world binding' also means a practice of 'reembedding' by which the subjects, under globalized conditions, define the connection between themselves and the world. 'World binding' in this sense represents the following: the social control of spatial and temporal horizons in order to control one's own actions and the practices of others. This implies practices of allocative appropriation of material goods; practices of authoritative 'appropriation,' or control of subjects over a distance; and the symbolic appropriation of objects and subjects on the basis of the available inventory of knowledge.

It is not the 'formation of space', but rather the various forms (calculative/economic, normative/social, and symbolic/cultural) of appropriating the world of corporeal realities, that is, terrestrially classified objects and bodies that are of central interest.

Therefore, the specificity of geography making as globalization process may be defined as follows: a process of appropriation over distance or in absence, which may proceed differently for differing life forms and in dif fering life worlds.

Summary

In this way, an action centered understanding of geo graphical realties aims to reconstruct everyday region alization of the life world. It critically examines the taken for granted geographical representations of the world, which are so often mobilized, region centered representations of the world, and politically mobilized by regionalist and nationalist discourses with a strong ten dency to strengthen fundamentalist attitudes. At the same time, it should become obvious that the analysis of re gionalization processes should not be limited to the re construction of actually existing political regions, as this is the case for the new regional geography. The utiliza tion of structuration theory for geographical investigation of sociocultural universes has rather to focus on processes of regionalizing the world through actions as forms of world binding and the significance of spatial relations for the production and reproduction of sociocultural real ities. In this sense, geographical representations of late modern spatial relations has to focus on the agencies of world binding and to give insights into the ways agents 'live the world' and not how they live 'in' regions and (living) spaces. This can be considered as one of the challenging duties of structurationist geography and contemporary human geography in general.

See also: Cultural Geography; Habitus; New Regionalism; Regional Geography I; Regionalisations, Everyday; Situated Knowledge, Reflexivity; Social Geography; Society-Space; Structuralism/Structuralist Geography; Structuration Theory; Subjectivity.

Further Reading

Giddens, A. (1979). *Central Problems in Social Theory. Action, Structure and Contradiction in Social Analysis*. London: Macmillan Press.

Giddens, A. (1981). *A Contemporary Critique of Historical Materialism, Vol. 1: Power, Property and the State*. London: Macmillan Press.

Giddens, A. (1984). *The Constitution of Society*. Outline of the Theory of Structuration. Cambridge: Polity Press.

Giddens, A. (1990). *The Consequences of Modernity*. Stanford: Stanford University Press.

Giddens, A. (1991). *Modernity and Self Identity. Self and Society in the Late Modern Age*. Cambridge: Polity press.

Gregory, D. (1978). *Ideology, Science and Human Geography*. London: Harper Collins.

Gregory, D. (1981). Human agency and human geography. *Transactions of the Institute of British Geographers, NS* 6, 1 18.

Gregory, D. (1982). *Regional Transformation and Industrial Revolution: A Geography of Yorkshire Woollen Industry*. London: Macmillan Press.

Gregory, D. (1984). Space, time and politics in social theory. An interview with Anthony Giddens. *Environment and Planning D. Society and Space* 2(2), 123 132.

Gregory, D. (1989). Presences and absences: Time space relations and structuration theory. In Held, D. & Thompson, J. (eds.) *Social Theory of Modern Societies. Anthony Giddens and His Critics*, pp 185 214. Cambridge: University Press.

Gregory, D. and Urry, J. (eds.) (1985). *Social Relations and Spatial Structures*. London: Macmillan Press.

Paasi, A. (1986). The institutionalisation of regions: Framework for understanding the emergence of regions and the constitution of regional identity. *Fennia* 164(2), 105 146.

Paasi, A. (1991). Deconstructing regions: Notes on the scales of spatial life. *Society and Space* 23, 239 256.

Paasi, A. (1992). The construction of socio spatial consciousness. Geographical perspectives on the history and contexts of Finnish nationalism. *Nordisk Samhällsgeografisk Tidskrift* 15, 79 100.

Paasi, A. (1996). Inclusion, exclusion and territorial identities the meaning of boundaries in the globalizing geopolitical landscape. *Nordisk Samhällsgeografisk Tidskrift* 23, 3 18.

Paasi, A. (2002). Bounded spaces in the mobile world: Deconstructing regional identity. *Tijdskrift voor Economische en Sociale Geografie* 93(2), 137 148.

Pred, A. (1981). Social reproduction and the time geography of everyday life. *Geografiska Annaler* 63B, 5 22.

Pred, A. (1984). Place as historically contingent process. *Annals of the American Association of Geographers* 74(3), 279 297.

Pred, A. (1984). Structuration, biography formation and knowledge: Observations on port growth during the late mercantile period. *Environment and Planning D. Society and Space* 2(3), 252 276.

Pred, A. (1985). The social becomes the spatial, the spatial becomes the social: Enclosures, social change and the becoming of places in the Swedish province of Skåne. In Gregory, D. & Urry, J. (eds.) *Social Relations and Spatial Structures*, pp 337 365. London: Macmillan Press.

Pred, A. (1986). *Place, Practice and Structure.* Cambridge: Polity Press.

Thrift, N. (1983). On the determination of social action in space and time. *Environment and Planning D. Society and Space* 1, 23 56.

Thrift, N. (1985). Bear and mouse or bear and tree? Anthony Giddens' reconstruction of social theory. *Sociology* 19(4), 609 623.

Thrift, N. (1985). Flies and germs: A geography of knowledge. In Gregory, D. & Urry, J. (eds.) *Social Relations and Spatial Structures*, pp 366 403. London: Macmillan press.

Werlen, B. (1987). Gesellschaft, Handlung und Raum. Grundlagen handlungstheoretischer Sozialgeographie. Erdkundliches Wissen, Heft 89. Stuttgart: Franz Steiner Verlag.

Werlen, B. (1993). *Society, Action and Space. An Alternative Human Geography*. London: Routledge.

Werlen, B. (1995). Sozialgeographie Alltaglicher Regionalisierungen. Band 1: Zur Ontologie von Gesellschaft und Raum. Stuttgart: Franz Steiner Verlag.

Werlen, B. (1997). Sozialgeographie alltaglicher Regionalisierungen. Band 2: Globalisierung, Region und Regionalisierung. Stuttgart: Franz Steiner Verlag.

Werlen, B. (2000). *Sozialgeographie. Eine Einführung*. Bern: Haupt Verlag.

Werlen, B. (2007). Sozialgeographie alltaglicher Regionalisierungen. Bd. 3: Ausgangspunkte und Befunde empirischer Forschung. Erdkundliches Wissen, Heft 121. Stuttgart: Franz Steiner Verlag.

Relevant Websites

http://www.cultural geography.net
Cultural Geography.

http://www.social geography.net
Social Geography.

Subaltern

C. McEwan, Durham University, Durham, UK

Glossary

Alterity Literally meaning 'otherness', this refers to the systematized construction of difference, which is also hierarchical, between groups of people on the basis of such factors as race, ethnicity, or culture.

Contact Zone Social spaces where cultures meet, clash, and come into contestation with each other; often in contexts of highly asymmetrical relations of power, such as colonialism, slavery, or their aftermaths as they are lived out in many parts of the world today.

Discourse Social practices through which the world is made meaningful and intelligible, including narratives, concepts, ideologies, and signifying practices.

Epistemic Violence Violence exercised through knowledge regimes and the epistemological suppression of the other; the violently appropriative colonizing practices that are deeply ingrained in Western textual production.

Hegemony The dominance of one group or class in society, achieved not through force but rather through the consent of other groups. Consent is achieved through the dominant group associating itself with moral and intellectual leadership in a society.

Historiography The study of the way that history has been and is written, or the history of historical writing. This is not the study of events of the past directly, but the changing interpretations of those events in the works of historians.

Orientalism Western constructions of the East as an opposite, working ideologically through a series of binary oppositions and underpinning imperialism.

Othering A way of defining and securing one's own positive identity through the stigmatization of an other; the self-identity of the colonizer and its culture is defined as superior through the identification of an inferior, opposite, colonized other.

Subalternization A term first used by Gayatri Spivak to refer to the systematic denigration of colonized peoples, largely through colonial discourse, which renders not only peoples and cultures subordinate but continues to create a world in which power (political, cultural, economic, and military) is still concentrated in the former colonial powers and the global North.

Transculturation One of the key cultural processes that operate between hitherto sharply differentiated cultures and peoples who are forced (usually by the processes of imperialism or globalization, and primarily through migration) to interact. This interaction often takes place in profoundly asymmetrical ways in terms of relative power between different groups.

Origins and Definitions

The meaning and usage of the term subaltern is complex and there are two distinct but related ways in which it has been understood and deployed. First, it originates as a military expression to refer to junior officers derived from its literal meaning – 'subordinate'. In the British army, for example, subalterns are commissioned officers below the rank of captain. These ranks are subordinate to the superior officers, but serve a particular role in connecting the lowest ranks with the commanding ranks and thus ensuring that the chain of command is maintained. Second, the term is used more broadly to refer to any person of inferior station or rank, again drawing on its literal meaning of 'subordinate'. Since the 1980s, this meaning of the term has become widely used, primarily within postcolonial theory and also by geographers, historians, and cultural theorists, to mark the subordinate positions of former colonized peoples who remain marginalized primarily on the basis of race and class.

This broader usage of subaltern to refer to peoples subordinated by relationships of power is adopted primarily from the writings of Italian philosopher Antonio Gramsci (1881–1937), who examined political and cultural issues from a Marxist perspective. Most of Gramsci's meditations on the subject were written while he was imprisoned for dissidence and, since he was obliged to encrypt his writings to get them past prison censors, his original usage is covert and at times opaque. Confusion sometimes arises because both usages of the term appear in his various essays. In *The Intellectuals*, for example, he uses the term 'subalterno' (subordinate) to draw explicit analogies between urban intellectuals and subaltern officers. The intellectuals – the subaltern classes – are the deputies to the dominant group, exercising the subaltern functions of social hegemony and political government. They thus give legitimacy to the dominant group and perform a mediating function in the struggle of class forces between the dominant group and the masses. However, in most of Gramsci's writings, subaltern is used surreptitiously to refer to the 'proletarian', whose voice could not be heard, being structurally written out of capitalist bourgeois narratives.

What is significant about Gramsci's theorization of the subaltern classes is his analysis of the ways in which they are subordinated by hegemony – the dominance of one group or class in society – achieved not through force but through the consent of other groups. Consent is achieved through the dominant group associating itself with moral and intellectual leadership in society, usually through civil society structures. Gramsci argued that the problem for the subordinated working classes is that they are not unified and their inability to unify means that hegemonic relationships of power remain unchallenged. However, he argued that the working classes are capable of developing from within their ranks their own organic intellectuals and developing links with certain sections of the traditional intelligentsia, through which they could then move toward hegemony.

The Subaltern in Postcolonial Theory

Both meanings of subaltern – the military analogy and the general meaning of subordinate – have been particularly significant within postcolonial theory. Drawing on Gramscian notions, Edward Said and other postcolonial theorists have examined the processes by which elites in colonized countries internalize Eurocentric ideas of the inferiority of their own cultures and ensure the dominance of the European powers. This dominance was achieved largely through processes of colonialism, whereby imperial power was grafted onto existing mechanisms of power. Westernization in India and the African colonies, for example, was a tactic for forming an administration and military from local elites, whose access to power would be limited to a subaltern role. Rather like the subaltern officers in the British army, these elites were essential to the chain of command and in maintaining order amongst the masses; the hegemony of the British colonial power was maintained through the consent of these subaltern administrative and military ranks. Other, more subordinated groups (for example, proletarian and peasant groups) were not drawn into these structures but were left to renegotiate their position relative to these local elites.

Subaltern, both as a term of reference and as an object of enquiry, was popularized within academic literature by the Subaltern Studies collective. This was a group of South Asian scholars (orginally Gayatri Chakravorty Spivak, Gyan Prakash, Ranajit Guha, Partha Chatterjee, Shahid Amin, David Arnold, David Hardiman, Gyanendra Pandey, and Dipesh Chakrabarty) interested in the postcolonial and postimperial societies of South Asia in particular. Their use of the term subaltern was somewhat broader than the military analogy, referring to wider processes of subordination, particularly in the writing of histories of independence struggles and postcolonial nation building. The Subaltern Studies scholars appropriated Gramsci's term in order to locate and reestablish a voice or collective locus of agency in postcolonial India that moved beyond the role of elites. Beginning in the 1980s, they attempted to formulate a new narrative of the history of India and South Asia. Although clearly influenced in various degrees by Marxist approaches, they were critical of traditional Marxist and elitist narratives of Indian history, both of which focused on the political consciousness of elites in mobilizing the masses, ending the domination of Britain as the colonial power, and building an independent nation state. Instead, the Subaltern Studies group focused on the role of nonelites – subalterns – as agents of political and social change, concentrating specifically on the role of peasants. In particular, they explored the discourses and rhetoric of emerging political and social movements, in contrast to only highly visible actions like demonstrations and uprisings. The important point in the early writings of the collective was that subaltern groups were also in a position to subvert the authority of those who had hegemonic power, and they thus played a more significant role in historical change than had previously been acknowledged. Historical narratives thus need to reflect this in contrast to the grand, heroic narratives of independence and nation building that erase the historical agency of subaltern groups.

Inspired by the work of the Subaltern Studies collective, the term subaltern is now more popularly used, particularly within postcolonial theory, to describe groups who are excluded and do not have a position from which to speak, for example, peasants and women in postcolonial societies. Thus in current philosophical and critical usage, subaltern describes specifically a person or groups of people rendered voiceless and without agency by their social status. Gayatri Spivak's 1985 essay *Can the subaltern speak?* has been a particular influence on this broader deployment, especially in its deconstruction of gender and representation in India. The idea of the subaltern has been central to postcolonial theory, particularly the exploration of the processes that have rendered the peoples of colonized countries (and continue to render them in postcolonial contexts) as subordinate, inferior, and without agency or voice. These processes are overwhelmingly discursive. For example, Edward Said's *Orientalism* (1978) referred to the rendering of groups of people as subaltern as 'othering' and critiqued the constellation of false assumptions underlying Western attitudes toward other parts of the world. Subaltern contains within it the notion of alterity, literally meaning 'otherness', which refers to the systematized and hierarchical construction of difference between groups of people on the basis of such factors as race, ethnicity, or culture. Said exposed the subtle and persistent Eurocentric prejudices against non European peoples and their cultures, which he argued had served as an implicit justification for Europe's colonial and imperial ambitions. The denigration of oppressed, colonized peoples and their rendering as subaltern subjects was crucial to the self definition and assumed superiority of the colonizers. Thus, postcolonial theory has examined the centrality of alterity to imperialism and colonialism, global geopolitics and development theory and practice, with the construction of First World/Western/Northern/advanced/developed/modern 'self' and Third World/Southern/backward/developing/traditional 'other' framing interventions in both colonial and postcolonial contexts.

Since subaltern can refer to any person or group of inferior rank and/or status, it might thus be employed in discussions of race, ethnicity, class, gender, sexuality, religion, and so on. However, some care needs to be exercised in its usage. Spivak, for example, objects to careless use of the term and its appropriation by other marginalized groups who are not specifically subaltern. She asserts that subaltern is not just another word for the oppressed or marginalized nor a claim to disenfranchisement within a system of hegemonic discourse. Rather, it signifies very specifically a group of people whose voices cannot be heard or that are wilfully ignored in dominant modes of narrative production. Thus, as it is used in postcolonial theory, subaltern refers to those who have no access to processes of cultural imperialism, which locates them in a space of difference or alterity. In this sense, then, subaltern is not simply the oppressed. For example, working classes in the West can be said to have been oppressed, but they are not subaltern by the fact that they have historically had access to cultural imperialism and are positioned within hegemonic discourse. The project of Subaltern Studies has been to create intellectual space to allow the subaltern – in this case, Indian women, workers, and peasants – to speak rather than always being spoken for by either elites or colonizing Western representatives. It is this use of subaltern – as an individual or group rendered subordinate by cultural imperialism and without agency and voice – that has come to be used within historical, cultural, political, and development geographies.

Subaltern Historiographies of Geography

The critique of elitist histories of Indian independence by the Subaltern Studies group and the emphasis on recovering the role that subaltern groups played has had an influence on the ways in which histories of geography are written, particularly since the early 1990s. Histories of European geography and imperialism are deeply intertwined. However, while histories of imperial exploration have been at the core of the historiography of European geography, the role of subalterns in the creation of geographical knowledge is often written out of histories of the discipline. In order to counter these hegemonic histories, geographers have begun to unsettle notions of the modern geographical enterprise being solely the product of metropolitan energies and initiative and to open up spaces to reveal multiple agents in the production of imperial geographical knowledge.

The notion of countering Eurocentrism in histories of geography is obviously inspired by postcolonial critiques, which demand an interrogation of the history of geography as it has been constructed as a putatively 'Western' tradition. The narration of the history of European geography tends to suggest that the origins of these knowledges are located solely in the West, and thus ignores the agency of subaltern knowledges in their formulation. Not only did European knowledge appropriate extensive local knowledges but Eurocentric geographical knowledge was formed through the effacement of these alternative ways of knowing. Thus, famous explorers such as Burton and Speke constructed their geographies of Africa by appropriating local indigenous knowledges, but in positioning these knowledges as scientific and rational they actively dismissed and erased from their publications those subjective, local knowledges upon which they drew. This simultaneous appropriation and erasure of indigenous knowledges is symptomatic of colonial discourse, and is evidence of the epistemic violence that continues to write out the agency of subalterns in the production of geographical knowledge.

The writing of histories of geography, therefore, tends to reprivilege Europe as the site of the production of knowledge. In reality, empires were contact zones in which (at least) two cultures came together, each influencing the other, albeit against a backdrop of inequality. Thus, notions of transculturation are an important departure from the European expansionist perspective in allowing for the possibility of agency from within the colonies. However, the possibilities of recovering colonized voices from within these contact zones are still problematic. For example, there are very few written accounts from Africa of interactions with and responses to Europeans during the nineteenth century. It is not always possible to retrieve subordinated voices from imperial archives because they are not present to be emancipated. Instead, inspired by Subaltern Studies scholars, geographers have begun to explore the epistemic and real violence that suppressed the subaltern voice, the evidence of this in the imperial archives, and the subsequent denial of the role of subalterns in the production of geographical knowledge about the empire.

In addition, while the subaltern voice may be irretrievable from the archives, it is only within these archives that evidence of agency and resistance by colonized peoples in British and European imperialism can be recovered. In *Can the subaltern speak?* Spivak argues that the absence of subaltern testimonials and written histories does not preclude the uncovering of the agency of the colonized subject. Rather, the colonial archive is shaped not only by the will of the colonizer, but also by the will of the colonized; the texts of counterinsurgency and elite documentation provide evidence of the con sciousness of the subaltern. Therefore, official documents, exploration narratives, diaries, travelogs, missionary re ports, and such like contain evidence of the will of the colonized, and of the extent to which the subaltern may have played a constitutive rather than reflective role in colonial discourse. Deconstructing the colonial archive also allows for a more subtle analysis of the production of geographical knowledge, and the possibility of de cen tering a putatively Western tradition by viewing the production of these knowledges as a complex process of cultural exchange and negotiation, in which the subaltern played an active part.

Deconstructing the hegemonic discourses can allow, to some extent and not unproblematically, for the re covery of subaltern agency and resistance in histories of imperialism and geography. Exploration narratives also contain accounts of insubordination on the part of local guides and carriers, suggesting that racialized imperial authority was challenged in subtle ways by subaltern agency. The imperial archive is also revealing of resist ance to European imperialism in a wider sense. Com muniqués between colonial officials illustrate indigenous resistance to colonization and court and military records document small scale as well as large scale insurgencies. Of course, there is also the added problem of silence as an act of resistance, since this raises questions about whether the subaltern voice wishes to be emancipated and who performs this emancipation. The subaltern does not require an advocate to speak for her. With this in mind, recent historical geographies of imperialism have attempted to follow the direction of the Subaltern Studies group as an example of how critical histories of the discipline can be written not to give the subaltern voice but to create space in which the subaltern can speak and be heard.

Subaltern Historical Geographies: History from Below

The approach of the Subaltern Studies scholars is often referred to as 'history from below', focusing more on what happens among the masses at the base levels of society than among the elite. The critique of elitist his tories and the focus writing subalterns into history and presenting them as on the agents of political and social change has had influence in both history and historical geography. Rather than an emphasis on grand narratives of historical geographical change, the emphasis is now very much on more limited and specific accounts of particular events and incidents. Furthermore, the con cern of Subaltern Studies and other postcolonial scholars (primarily Edward Said) with not only political acts (such as rallies, demonstrations, and uprisings), but also with political rhetorics and discourses, has also shaped what now constitutes historical geography. Recent years have seen an outpouring of work on colonial discourses, for example, and how these were used to create particular geographical imaginaries. These geographical imagin aries were Orientalist in nature, representing colonized peoples and landscapes as barbaric and inferior, and being deployed as justification for imperialist inter ventions in far off lands.

The methodologies of historical geography have also been influenced by the broader interest in subaltern histories. In particular, oral histories provide access to the 'histories from below', the informal and the subaltern since they extend beyond official discourses. One of the strengths of oral history, then, is that it provides access to politically and economically marginalized groups. It ex tends beyond the limitations of written documents of fering a voice to the 'rank and file'. Central here is the problematic nature of knowledge as a social and histor ical construct and an acceptance that written evidence is only one side of a version of history. Thus, historical and cultural geographers are now using the recorded ex periences of subaltern groups in order to write post colonial historical geographies of, for example, the experiences of minority ethnic communities in European cities, or the personal accounts of the perilous journeys made by refugees, or collective and individual memories of colonialism in places as diverse as Ireland, the Caribbean, India, and Australia.

Space, Place, and the Subaltern

In addition to helping to shape new approaches to both historiographies of geography and to historical geog raphies, the subaltern has become an increasingly sig nificant presence in geographical analyses at a range of different scales. This has largely been a consequence of a 'cultural turn' within the discipline and the focus on subaltern groups. Class and labor in critical geographies have been largely supplanted by identity politics and the impact of postcolonial theory meant increased ques tioning of conventional, Eurocentric forms of theorizing. The result was that culture, rather than being a neglected afterthought, became a central issue for geographers. A renewed and reinvigorated cultural geography has thus incorporated postcolonial and feminist theories in its challenge of the meaning of culture. No longer is culture seen to be constituted by a uniform set of attitudes, be liefs, and material artefacts, since this is seen as the ideology of a dominant group seeking hegemonic power over what it regards as 'other' to itself. Rather, the cul tural integrity of those traditionally excluded from discourse (women, the formerly colonized, the materially dispossessed, and other subaltern groups) is recognized, and the theory of cultural politics is seen as central to understanding the constant struggle for power. Thus, geographers have explored topics as diverse as the politics of place naming in postcolonial Ireland, the lived experiences of diasporic communities around the world, and aboriginal rights claims in both Canada and Australia, topics that place subaltern groups at the center of analysis and retain the hearing of their stories as the key political motivation.

At a different scale, political geographers have also been influenced by this cultural turn and are increasingly interested in the role of subalternization in global geopolitics. The systematic denigration in colonial discourse of other peoples and cultures as processes of subalternization has, according to postcolonial scholars such as Spivak, left a legacy that is buried in the psyche and assumed knowledge about the world that continues to have problematic effects today. These include the continued power of Western conceptualizations of development and modernization and a continued deference to the authority of Western discourses. Political geographers have drawn on such arguments to explore how this relationship between power, knowledge, and development continue to shape geopolitical relations throughout the world. Subalternization renders not only peoples and cultures as subordinate but creates a world in which power (political, economic, and military) is still concentrated in the 'core' (the advanced economies of the former colonial powers, the USA and, increasingly, China) at the expense of the 'periphery' (mainly the former colonies). Some commentators claim that the postcolonial world is still kept in a state of subordination because it is subsumed under capital and integrated in the global order of capital.

At a smaller scale, political geographers are also interested in reformulating an understanding of the public sphere that might provide more space for subaltern politics. The public sphere is seen as important to democracy and citizenship because it exists outside the institutions of government, providing a space where an active citizenry can offer critique and thus influence formal politics. Feminists, in particular, have drawn on postcolonial theory to challenge the notion of a public sphere shaped by exclusions based on class, race, and gender and have called for a more plural conceptualization. The idea of multiple, overlapping public spheres is significant in societies characterized by systemic inequalities, and subaltern counterpublics are seen as vital spaces from which marginalized groups can formulate public critique and discourse. This is one of many examples where postcolonial theories of the subaltern have meshed with feminist theories to provide a politics concerned with creating spaces in which subaltern groups can be heard.

Concerns such as these were perhaps first voiced within feminist development geographies, where the requirement to understand the lived experiences of women in the South acquired increased urgency from the 1970s onward, again inspired by both feminist and postcolonial theories. It is now recognized that women, in the South face multiple challenges, including poverty, unemployment, limited access to land and credit, legal, social and cultural discrimination, sexual abuse, and other forms of exploitation and violence. While these challenges might be similar to those faced by all women, the specificities of history, political economy, and culture make them differentially oppressive to women in the South. In Spivak's words they are 'doubly subaltern'. Development geographies have acknowledged both these realities, as well as the fact that women of the South are not victims. Rather, as subaltern counterpublics they meet these challenges and are active in confronting and overcoming them, often in remarkably creative and empowering ways. Development, then, is not something that is 'done to' the Third World; rather, people of the South play an active role in contributing to the discourses and practices of development. These notions of subaltern agency and voice have become increasingly significant within critical development geographies.

Related to these geographies have been studies that have explored contemporary subaltern resistance. Of interest to geographers has been the resurgence of interest in the concept of place as an important object of struggle in the strategies of subaltern social movements. Critical geographers have begun to explore the significance of place in subaltern political strategies aimed at resisting the detrimental effects of globalization. The defense of place by subaltern social movements is often constituted as a rallying point for both theory construction and political action. Thus place based struggles by indigenous peoples, around issues such as environmental destruction, resource extraction, water privatization, and destruction of livelihoods might be seen as multiscale, network oriented subaltern strategies of localization. In such cases, subaltern groups are consciously building translocal solidarities as a means of achieving local aims and ambitions. Of equal significance are struggles over land rights claims and resurgences in localized political identities based around indigeneity, and geographers are increasingly exploring such issues.

Conclusion

Despite the embracing of postcolonialism in many aspects of geography and a political impetus to acknowledge and hear marginalized voices, the spectral presence of the subaltern continues to haunt the discipline. Northern geographies still struggle to move on from

Eurocentric visions, and there are still important ques tions to be asked about who is heard speaking about economic, political, or cultural geography. The chal lenges raised by the Subaltern Studies scholars have yet to be fully embraced, let alone overcome. In concluding that 'the subaltern cannot speak' critics have accused Spivak of not recognizing or allowing the subaltern to speak. However, her point was not that the subaltern does not speak in various ways, but that speaking is a trans action between speaker and listener. The question is does anyone hear when the subaltern speaks? In the absence of listeners, subaltern voices cannot achieve a dialogic level of utterance. This lesson is salutary for contemporary human geographies. There are still major challenges in overturning hierarchies of representation by recognizing colonial histories and power relations and the spaces through which colonial practices continue to reproduce themselves in the contemporary world. There remains a need to recognize the position from which people speak and to de center analysis. It is a source of irony, for ex ample, that the influence of Subaltern Studies in coun tries of the South is mediated by the academic filter of the North, pointing to the persistence of unequal global relations of knowledge production that erase or fail to acknowledge subaltern knowledges and voices.

See also: Colonialism I; Colonialism II; Discourse; Eurocentrism; Feminism and Work; Feminism, Maps and GIS; Feminist Geography, Prehistory of; Gender and Rurality; Gender, Historical Geographies of; Geopolitics; Imperialism, Cultural; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Other/Otherness; Postdevelopment; Race; Social Class; Third World.

Further Reading

Castree, N. (2004). Differential geographies: Place, indigenous rights and 'local' resources. *Political Geography* 23, 133 167.
Clayton, D. (2000). *Islands of Truth: The Imperial Fashioning of Vancouver Island*. Vancouver: UBC Press.
Escobar, A. (2001). Culture sits in places: Reflections on globalism and subaltern strategies of localization. *Political Geography* 20, 139 174.
Featherstone, D. (2005). Atlantic networks, antagonisms and the formation of subaltern political identities. *Social & Cultural Geography* 6(3), 387 404.
Gramsci, A. (1992). *Prison Notebooks*. Buttigieg, J.A. and Callari, A. (trans.). New York: Columbia University Press.
Gregory, D. (2004). *The Colonial Present*. Oxford: Blackwell.
Guha, R. (ed.) (1982). *Subaltern Studies*. New Delhi: Oxford University Press.
Jacobs, J. (1996). *Edge of Empire. Postcolonialism and the City*. London: Routledge.
McEwan, C. (1998). Cutting power lines within the palace? Countering paternity and eurocentrism in the 'geographical tradition'. *Transactions of the Institute of British Geographers* 23(3), 371 385.
Mohanty, C. (1991). Under Western eyes: Feminist scholarship and colonial discourses. In Mohanty, C., Russo, A. & Torres, L. (eds.) *Third World women and the politics of feminism*, pp 51 80. Bloomingdale: Indiana University Press.
Said, E. (1978). *Orientalism*. London: Peregrine.
Slater, D. (2004). *Geopolitics and the Post Colonial: Rethinking North South Relations*. Oxford: Blackwell.
Spivak, G. (1988). Can the subaltern speak? In Nelson, C. & Grossberg, L. (eds.) *Marxism and the interpretation of culture*, pp 271 313. Urbana: University of Illinois Press.
Spivak, G. (1988). Subaltern studies: Deconstructing historiography. In Guha, R. & Spivak, G. (eds.) *Selected subaltern studies*. Oxford: Oxford University Press.
Yeboah, I. (2006). Subaltern strategies and development practice: Urban water privatization in Ghana. *Geographical Journal* 172(1), 50 65.

Relevant Websites

http://wsf2007.org/info/
About World Social Forum, Nairobi 2007. An example of subaltern counter publics and their anti globalization networking activities.
http://www.lib.virginia.edu/area studies/subaltern/ssmap.htm
Bibliography of the writings of the Subaltern Studies group, 1988 2000.
http://www.heinemann.co.uk
Link to Heinemann publishers and their African Writers Series and Caribbean Writers Series.
http://www.bamako themovie.com.
Website of the movie Bamako (2006) including video clips; a fascinating example of a counter discourse from the perspective of the subaltern.

Subalternity

V. Gidwani, University of Minnesota, Minneapolis, MN, USA

Glossary

Aporia A rhetorical term for 'doubt' or 'difficulty in choosing' (from ancient Greek, *a* 'without', *poros* 'way', or 'passage'); in Jacques Derrida's usage, 'a non-road'.

Contrapuntal Any music that contains two or more voices heard simultaneously; in the literary context (and closely associated with the work of Edward Said), a narrative relating to or marked by 'counterpoint'.

Catachrestic Paradoxical, forced; within postcolonial criticism, an 'ab-usive' strategy of reading that seizes a given discourse to reverse and displace its effects.

Hegemony In V.I. Lenin's formulation, an achieved unity of working class and other social fractions that enables the proletariat to capture the apparatuses of State and civil society. Antonio Gramsci extends Lenin's description, characterizing hegemony as a dynamic 'ethico-political' terrain in which a group becomes 'directive and dominant' in society through "the combination of force and consent variously balancing one another, *without force exceeding consent too much*."

Historical Bloc A term used by Gramsci to designate the dialectical unity of base and superstructure, theory and praxis, and intellectuals and masses. It is more than a mere alliance of social forces.

Singular A figure that is unique, as in the proper name, which resists characterization as merely a particular instance of a general phenomenon; in short, an entity that is irreducible.

Sublation English translations have struggled with G.W.F. Hegel's concept of *Aufhebung*, a process of simultaneously lifting up, conserving, and negating the 'other' within the self-movement of the dialectic. 'Sublation' captures these contradictory meanings better than the alternative 'supersession'.

Supplement Jacques Derrida plays with slippages in the noun and verb 'supplement' (Fr. *supplément*) and 'to supplement' (Fr. *suppléer*) to describe those unseen forces, which 'add what is missing' or 'supply a necessary surplus' to texts and their narratives, and are viral in that they resist domestication or arrest even as they erupt a text's desire for self-adequacy. The Derridean seam of postcolonial studies conceives the subaltern as that untamable specter, which haunts and operates all master narratives.

Telos An ultimate end, or the end of a goal-oriented process.

Introduction

Subalternity is, simply put, the state of being subaltern. But, what does it mean to be subaltern? The *Oxford English Dictionary* notes that the noun, subaltern, is an appellation for "an officer in the British army below the rank of captain, especially a second lieutenant." As an adjective, it implies a person 'of lower status' or 'of inferior rank'. The term appears to have first come into use in thirteenth century France to denote a person who is a 'subordinate'. Its current coinage and popularity in academic writing owes to the influence of Subaltern Studies historiography. Launched in 1982 under the intellectual leadership of Ranajit Guha, the Subaltern Studies collective of historians sought to write histories from 'below' – of the subordinated masses, as it were – that would be contrapuntal to hegemonic Liberal and Marxist narratives of colonialism, nationalism, and modernity in India. Subaltern Studies crystallized the groundswell of discontent that had been building within a subset of the academic Left in the 1960s and 1970s over the marginalization of peasant and tribal movements by the State and intellectual orthodoxy. Discontent became rupture in the wake of political events in India from the late 1960s onwards. In 1967, the Maoist Naxalbari peasant uprising in West Bengal was brutally crushed by government forces. In a savage twist of history, the cadre of the Communist Party of India (Marxist), which had been elected to a majority in the state legislature, was at the forefront of repression. Naxalbari and subsequently the suspension of civil liberties by Indira Gandhi's regime during the 1975–77 'Emergency' years laid bare the authoritarian tendencies and fissures within Indian democracy, and the failure of Nehruvian socialism as a nation building program.

Gramsci and Subalternity

Given that India in the mid 1970s bore many of the hallmarks of fascism it was logical for Subaltern Studies to turn for theoretical inspiration, initially, to the writings of the Italian revolutionary, Antonio Gramsci. Gramsci, a founding member of the Italian Communist Party, was thrown into prison in 1926 for his opposition to Mussolini's Fascist government, which feared his organizational and intellectual acumen. His health failed in prison and he died in 1937 at 46. But between 1916 and 1935, based on his experiences as a communist organizer and strategist, and wide ranging engagement with political and philosophical tracts, Gramsci authored a set of remark able commentaries on class, popular culture, folklore, literature, philosophy, ideology, the state, and political strategy (many of these are in the 29 notebooks, some unfinished, that comprise his prison writings).

Once described by Eric Hobsbawm as 'par excellence the philosopher of political praxis', Gramsci was deeply mindful of the 'peasant question' – which was, in Fried rich Engels' paradigmatic formulation, how to enroll a conservative peasantry for a proletarian revolution led by urban working classes. Like V.I. Lenin before him, Gramsci recognized that a politically effective class al liance centered about working class radicalism and unity could not be taken for granted merely because a certain deterministic Marxist orthodoxy proclaimed it. Neither the class alliance, nor working class radicalism, nor working class unity was preordained. It was the task of political praxis to produce these, which meant that Marxist intellectuals would have to understand and en gage, rather than dismiss or be diffident toward, peasant culture and consciousness. Although the term 'subaltern' has become a more elastic descriptor since – a (some times indiscriminate) referent to persons and groups hierarchically positioned as subordinates or inferiors within nation states, capitalist production relations, or relations of patriarchy, race, caste, and so forth – in Gramsci's usage, which was the proximate inspiration for Subaltern Studies, subaltern classes (It. *subalterno*) were invariably contrasted to a 'dominant and directive social group', that is, a group which had been able to rise to 'the phase of ethico political hegemony in civil society, and of domination in the state'. It is likely that Gramsci used the word 'subaltern' in place of 'proletarian' in an effort to evade prison censors: thus, syndicalists and peasants in southern Italy were among the groups whom Gramsci identified as subaltern, contrasting them to the Fascists and the northern bourgeoisie who were at the time dominant in Italy.

Writing about the 'common sense' of the 'masses' – and by masses he meant the subaltern groups who 'must be made ideologically homogeneous' through the polit ical work of hegemony – Gramsci asserted that such common sense was a 'spontaneous philosophy of the multitude' that was non identical in 'time and space' and took 'countless different forms'. Fragmentary, even in coherent, it was nevertheless a 'conception of the world', which possessed a 'formal solidity' and a 'consequent imperative character' that 'produced norms of conduct'. While autonomous from the 'great systems of traditional philosophy' and 'high culture', which in Gramsci's esti mation had 'no direct influence' on the multitude's 'way of thinking and acting', ordinary people's common sense was not thereby in necessary opposition to ruling class ideologies. Instead, elite conceptions of the world oper ated on 'the popular masses as an external political force, an element of cohesive force exercised by the ruling classes and therefore an element of subordination to an external hegemony'. Here Gramsci appears to make a claim about subalternity that is both ontological and methodological: subalternity resides in the crevices of common sense. Or more exactly, popular consciousness as the accreted experience of being underclass, as well as a realm of unsystematic and officially disqualified knowledge, is a living record of subalternity.

The telescoped implications of this claim are pro found. Politically, it implies that popular consciousness is the lived ideology of the masses that any group or party must accept as a point of departure if it is to forge a 'historical bloc' and achieve 'ethico political hegemony' in civil society and state. Intellectually, it prods scholars and activists to recognize spaces of subaltern politics whose forms of mobilization, organization, and operation differ from and are relatively independent of elite modes of politics.

Critical to note here is the grammar of geography that saturates terms like 'subaltern' and 'hegemony' – key words in Gramsci's political arsenal. Gramsci's injunction to reconstruct histories of popular initiative and resist ance that were absented in official records (that is, in the archives of the ruling classes) was innovatively seized in the 1960s by scholars of the English New Left, notably Perry Anderson, Tom Nairn, Raymond Williams, Stuart Hall, Eric Hobsbawm, Rodney Hilton, Christopher Hill, and E.P. Thompson. Hobsbawm, Hill, and Thompson, among others, relied on non conventional sources of evidence such as popular protest pamphlets and posters, workers' diaries, folk songs and lore, and court testi monies to produce contrapuntal narratives of the his torical experiences of workers, peasants, and 'common people'. Their 'history from below' approaches trans formed the study of history the world over.

The themes pioneered by Gramsci and extended by English cultural Marxism were repeated with difference in the first volume of Subaltern Studies, in Ranajit Guha's now canonical essay on subaltern historiography. In staking out as this alternative historiography's point of departure the 'structural dichotomy' that has differen tiated elite politics from subaltern politics, Guha declared that there "were vast areas in the life and consciousness of the people which were never integrated into [the colonial or nationalist bourgeoisie's] hegemony…. Such dichot omy did not, however, mean that these two domains were hermetically sealed off from each and there was no con tact between them. On the contrary, there was a great deal of overlap arising precisely from the effort made from time to time by the more advanced elements among the indigenous elite, especially the bourgeoisie, to integrate them." Guha's manifesto, subsequently elaborated in his path breaking monograph, *The Elementary Aspects of Peasant Insurgency in Colonial India* (as well as a later book, *Dominance Without Hegemony: History and Power in Colonial India*) anticipates many of the theoretical and methodo logical innovations, as well as problems, that mark the Indian strand of Subaltern Studies.

Subaltern Studies and the Geopolitics of Representation

Contiguous with Gramsci, the early volumes of Sub altern Studies emphasized the 'independent initiatives' of rural subaltern classes, peasants and tribals, whose con tributions as subjects of history were effaced in Liberal (Cambridge School) and Marxist (nationalist) historio graphies of India (see David Arnold's perceptive essay on Gramsci's relevance to studies of subalternity in India). *Contra* Gramsci, in a move that was methodologically convenient but did no favors to the sophistication of Hegel's dialectical thought, Ranajit Guha and his acolytes theorized subaltern consciousness as, quite simply, a 'negative consciousness' whose "identity was expressed solely through an opposition, namely, its difference from and antagonism to its dominators." While the imprint of the 'history from below' approaches of Hobsbawm and Thompson on Indian Subaltern Studies was marked – both wanted to recover the agencies of 'people without history' – their battles were ultimately different. Sub altern Studies acknowledged the global reach of capital but challenged Eurocentric accounts that reduced mul tiple pasts and forms of alterity to particular instances of a general scheme (such as the stages of development of capitalism). Not even the best of the English Marxist historiographers was exempt from this reflex, which as sumed Europe's history to be the universal template: thus Hobsbawm's acclaimed study, *Primitive Rebels*, actively participated in the *telos* of Europe's Progress in charac terizing rebellious subalterns as 'primitive', 'backward', 'prepolitical', and 'anachronistic'. Subaltern Studies re jected such evolutionist descriptions of subaltern con sciousness and politics. Ranajit Guha, for example, insisted that when insurgents from the subaltern classes attempted to annex or annihilate symbols and insignias of enemy power during rebellions, their actions had to be read as an avowedly 'political struggle' to 'abolish the marks' of their subalternity. Inversion, a 'principal mo dality' of uprisings, was a deliberate strategy to force a crisis or semiotic rupture in ruling class discourses.

The more wide ranging criticism here, subsequently elaborated in the writings of Shahid Amin, Partha Chat terjee, Dipesh Chakrabarty, and Sumit Sarkar among others, was that peasants and various 'Others' effaced within structures of authority (colonialism, capitalism, patriarchy, caste, racism, nationalism, etc.) required nei ther preparation nor liberation to become political sub jects ready for modernity. They were *already* political and coeval. Such arguments were *contra* Liberal apologists for empire (like John Stuart Mill), who papered over liber alism's illiberalism in the colonies by claiming that 'na tives' in Britain's dominions had to be raised to political maturity before they could take up 'self rule'; and *contra* nationalist elites, both bourgeois and Marxist, who, misled by a parallel developmental conceit, repeatedly failed to recognize and ally themselves with politicized peasantries during anti colonial struggles (Frantz Fanon was par ticularly fierce in his denunciations of such national elites in *The Wretched of the Earth* and *Black Skin, White Masks*).

Thus, Indian Subaltern Studies historiography saw the task of 'provincializing Europe' – critiquing stagist, cap ital centered accounts in which Europe operated as silent empirical referent, site of Reason, and epistemic space of power – as intertwined with the project of dislodging elitist narratives of Indian nationalism, which "portrayed nationalist leaders as ushering India and its people out of some kind of precapitalist stage into a world historical phase of 'bourgeois modernity', properly fitted out with the artifacts of democracy." Contending that the domain of politics of 'the people' (a term that Guha used syn onymously with 'subaltern classes') functioned differ ently from the domain of elite politics – the latter secular, legalistic, and constitutional in orientation, based upon 'vertical mobilization' of constituents; the former, non secular, often violent in its modality, and involving 'horizontal affiliations' of 'kinship and territoriality' – Subaltern Studies attempted to erupt, both, the unitari ness of 'the nation' and its proclaimed trusteeship by a nationalist bourgeoisie. They argued, with justification, that the problem of subalternity operating within the integuments of colonial and nationalist historiographies not only disrupted the possibility of totalizing historical accounts (a claim keenly evidenced in Partha Chatterjee's *The Nation and Its Fragments* and Shahid Amin's *Event, Memory, Metaphor*), but also exposed the relations and geographies of power that sustain disciplines like History (a charge readily extended to Anthropology and Geog raphy). In its critique of imperial and capital centered teleologies, the nation form, and the historical enterprise itself, Subaltern Studies displaced and radically advanced the contrapuntal approach of English Marxist histori ography. These differences became increasingly salient after 1985 as postcolonial studies took up the question of subalternity – an engagement that prompted, both, renewed appreciation of Subaltern Studies and renewed scrutiny of its accumulated contradictions.

From Subaltern Studies to Postcolonial Criticism

The most serious complaint issued against Subaltern Studies by critics hostile and sympathetic to the project was, briefly stated, its unwitting complicity with Europe's episteme due to an inconsistent theory of subalternity. (See, for instance, the regional critiques of Subaltern Studies historiography in the 1994 mini forum on Subaltern Studies published in *The American Historical Review* **99**(**5**). Whereas Gyan Prakash evaluates the operations of subaltern studies as postcolonial criticism, Frederick Cooper and Florencia Mallon take stock of the insights and limitations of Indian subaltern studies for African and Latin American historiography, respectively.)

The mid 1980s was a watershed in this respect. Three articles – two by Gayatri Chakravorty Spivak in 1985 and 1988, and another by Rosalind O'Hanlon in 1988 – took Subaltern Studies to task in no uncertain terms for its failure to adequately theorize the 'subject'. While acknowledging substantial differences between members of the collective, they observed nevertheless that Subaltern Studies had veered perilously close to objectifying 'the subaltern', a set of heterogeneous actors, into an undifferentiated, humanist, and implicitly male subject agent. This was doubly ironic, because not only did this subaltern man resemble the bourgeois individual of liberal political theory, which Subaltern Studies sought to dislocate; but, like that archetypical 'man' who was completely legible to theory, 'the subaltern' also remained under the control of the knowledge that claimed to 'restore versions of causality and self determination to him'. The upshot of O'Hanlon's as well as Spivak's critique was that Subaltern Studies, with few exceptions, had absented women as subjects and, thanks to the procedural dualism and crypto positivism of Guha's canonical formulations, undercut its central mission – namely, to retrieve resistant presence in ways that would 'preserve the alienness and difference in the figure of the subaltern' rather than domesticate it to the historian's 'transcendent subject position'.

Other critics observed that Subaltern Studies' penchant for binary schemes was pernicious. Consider, for example, the vexed issue of violence. Taking cue from Guha's monograph, *The Elementary Aspects of Peasant Insurgency in Colonial India*, the early volumes of Subaltern Studies repeatedly foregrounded acts of insurrection by dominated groups whose modality was 'violent' and 'inversionary'. Whether motivated by methodological convenience (violent protests leave the most visible trace on colonial archives and thus the possibility of recovering the subaltern perspective by 'reading against the grain'), dialectical expedience (if colonial archives represent the colonizer's consciousness, then subaltern consciousness must be the inverse of this totality), or a certain theoretical inheritance (Hegel, Bakhtin, Gramsci) there is little denying the dangers of this move given the persistent antinomies, in Eurocentric scholarship, between the 'violence' and 'irrationality' of the colonized as against the 'civilized' and 'rational' conduct of colonizers. Similarly, Guha's conspicuous refusal to cede to subaltern insurgents the historical 'maturity' of consciousness and, by extension, the corporate capacity to "develop…a fully fledged struggle for national liberation" unintentionally nourished a series of worn colonial and bourgeois nationalist prejudices: namely, that Europe and its modernity were on the side of culture, reason, history and development, while subaltern classes were precisely those who 'lacked' these attributes.

Furthermore, if subaltern struggles were ultimately prone to 'political failure' – not because of the historical immaturity of subaltern classes, but more plausibly because of coercion by hegemonic groups, who stood in relations of superordination – it begged a fuller answer to the question: In what experiential and discursive sense were subalterns 'autonomous agents'? Indeed, by trucking in a more or less transparent figure of the subaltern as the dialectical negative of elite 'man', Subaltern Studies voided the more radical possibility of subaltern as that heterogeneous other which, although subordinated, defies sublation into hegemonic systems of understanding.

These were precisely the directions urged by postcolonial criticism. Gayatri Chakravorty Spivak's essay *Can the Subaltern Speak?* was, without doubt, the most provocative and widely noticed of various interventions. Her argument enacts a series of displacements. Spivak begins by noting the failure of Gilles Deleuze and Michel Foucault, two of the most ardent critics of the 'sovereign subject', to decenter this apparition. What initially appears as an idiosyncratic move – Spivak hones in on a conversation between Deleuze and Foucault, rather than their formal writings – ends up being emblematic of the undetected and viral workings of ideology. Spivak argues that Deleuze and Foucault's aversion to a theory of representation that would take account of the intellectual's location and of the geopolitical determinations that operate him serves, unpardonably, to "conserve the subject of the West, or the West as Subject."

The criticisms staged in the opening encounter with Deleuze and Foucault frame the remainder of Spivak's intervention. Her methodological aim is to advocate a radical practice of textuality and reading that will account for the politics of representation in its double sense of proxy ('speaking for') and portrait ('speaking of'), rather than glibly reincarnate the transcendent, paternal subject of authority and knowledge. Spivak disrupts the intellectual's complicity in an epistemic violence that persistently constitutes "the Other as the Self's shadow." This remark, which is directed as much at Subaltern Studies as at Western radical intellectuals, paves the way for a stark and simple injunction: "To confront them [subaltern Others] is not to represent them (*vertreten*) but to learn to represent (*darstellen*) ourselves." A lengthy discussion of the Hindu practice of *sati* (widow sacrifice) and the paradoxes that attend its abolition by British colonial rulers – Spivak's point is not that 'the killing of widows' is acceptable but rather that women's subjectivity is repeatedly foreclosed, first by the epistemic violence of Hindu law and later by colonial and postcolonial law – culminates in the now notorious answer to her article's title and provocation. Her terse response: "The subaltern cannot speak."

Several critics, even sympathetic ones, have objected to Spivak's 'muting' of subaltern voices (see, e.g., Fernando Coronil's article 'Listening to the subaltern'). They misconstrue the implications of Spivak's essay, which has been formative for the articulation of Subaltern Studies with postcolonial criticism. There are at least five overlapping readings of it, all with radical implications for studies of subalternity. First, the essay is tautological. The title is a red herring. The subaltern by definition is that colonized subject who cannot speak. She is always 'spoken for'. Second, the subaltern is that subject who cannot speak because her speech is disqualified. Here 'speech' carries the political valence of being the property of that person who is properly recognized as citizen of the *polis* – and hence a 'political being'. Third, the subaltern is that singular figure who, although exploited and marginalized within hegemonic formations, defies dialectical integration. She is the figure of the 'radically other' who marks off a cryptic, secret 'space of withholding' within the territorialized ambits of modernity, which dreads her precisely because she represents an internal margin that resists coding and, hence, the Enlightenment desire to know in order to control. Fourth, the subaltern names that figure of undecidability whose reality cannot be known ('spoken of'): as such, she forces an ethical decision around how and why to substitute for her ('speak for'). Subalternity pits the two senses of representation – portrait (*darstellen*) and proxy (*vertreten*) – into an aporetic relationship. Fifth, even if the subaltern could speak for herself – or as Spivak puts it, "invest her own itinerary" – she would still be within the domains of ideology and politics, and be operated by them. In other words, she would come to presence via the work of 'supplements': the network of signifiers, which precede and constitute her. The rub? There is no escape from the politics and ethics of representation.

Subalternity and Its Futures

Dipesh Chakrabarty's book *Provincializing Europe* (the first part especially) is one of the most sustained attempts to extend Spivak's discussion of subalternity. Chakrabarty employs the analytic device of History 1 and History 2s – a strange double geography – to build his arguments. To write within the episteme of History 1, according to Chakrabarty, is to produce the 'universal and necessary history we associate with capital', or equivalently, the history of 'the reproduction of the logic of capital'. Here, following Spivak, 'capital' stands in for both, 'the subject of the West' and 'the West as Subject'. By contrast, History 2s are those 'multiple possibilities' or 'subaltern pasts' that 'may be under the institutional domination of the logic of capital and exist in proximate relationship to it' but do not belong to its 'life process'. The task of History 2s, then, is both epistemological and ontological: on the one hand, to function as a strategy of writing that constantly interrupts 'the totalizing thrusts of History 1' and on the other hand, to write those pasts and presents that reveal the human bearer of labor power as having the capacity to enact 'ways of being in the world' other than as the bearer primarily of labor power – that is, use value for capital.

Chakrabarty's argument has profound implications for a text like David Harvey's *Limits to Capital* that has achieved canonical status within radical geography for its acute analysis of the historical and geographical dynamics of capitalist development. *Provincializing Europe* shows how an argument like *Limits*' operates firmly within the epistemic space of History 1, with the result that it writes back dispersed geographies of life into the expansionist narrative of capital's becoming – as variations in a singular, relentless process of capitalist development. In so doing it effectively recoups a diversity of struggles and modes of existence into the unifying dialectic of capital. Chakrabarty's claim is that History 1's effacements, which amount to a systematic erasure of certain subject effects within intellectual and shop floor production, can be gathered under the name 'subaltern'. This usage remains faithful to Gramsci to the extent that it designates subordinated elements of society. But since these subordinated elements are heterogeneous and always fragmented, the signifier 'subaltern' becomes not the container of a presence but the placeholder of an absence – a catachrestic, anti universal signifier that urges an ethic of responsibility to the dominated.

Chakrabarty relies heavily on theoretical openings in Marx's *Grundrisse* and the displacement of Hegel's dialectic in Derrida's filial notions of 'différance', 'supplement', and 'trace' to convey his argument about subalternity. He contends that although capital strives to sublate 'difference' (Chakrabarty variously invokes 'History 2s', 'concrete labor', 'real labor', 'subaltern pasts' as examples) into itself, these continue to 'inhere in capital' and 'interrupt and punctuate the run of capital's own logic'. Derrida's concept of 'trace' provides an important corrective to the logic of double negation that is both the power and failing of Hegel's dialectic of identity. For Chakrabarty the 'trace' is that ineffable element – the *Grundrisse*'s 'unconquered remnants', Derrida's 'radically other within the structure of difference that is the sign', Spivak's 'subaltern', or his own 'History 2s' – that constantly challenges from within capital its 'claims to unity and universality'. As such, his argument offers an implicit critique of the Hegelian *Aufhebung*, particularly its commonplace interpretation as a philosophical historicism that negates difference and leaves no room for its affirmation outside the protocols of European modernity. *Provincializing Europe* ultimately demands an ethics and politics of representation that does not translate variegated forms of life into the abstract, formalized categories of capital in lockstep with the transactional logic of positivism, but rather stages subalternity in ways that resist the pull of equivalence.

Caribbean and Latin American Subaltern Studies have been considerably more attentive than its Indian counterpart in exploring how colonialism and racism have intertwined with capitalism to produce forms of subordination and exclusion that are manifestly 'subaltern' (*The Latin American Subaltern Studies Reader* surveys this rich seam of research). It has also gone farther, most vividly in critical scholarship around the Black Atlantic, in spatializing subalternity and revealing how people and places in imperial margins have enabled the (epistemic and economic) center that is 'Europe'. In the work of scholars like Enrique Dussel, this ingenious displacement of 'Europe' and its modernity as effect rather than autochthonous cause of world history renovates 'the subaltern' (a term that Dussel uses interchangeably with 'the poor') as the excluded Others of modernity who are owed an insuperable debt by it.

Geographers have extended this line of argument, which amounts to a spatial critique of the core's hegemony, in vital ways. David Slater, for example, endorses the postcolonial turn as a means to 'challenge the history and geopolitics of metropolitan theory'. Matthew Sparke's *In the Space of Theory* embraces this summons. Thinking spatially through Gayatri Spivak's deconstructive Marxism, he advocates a disruptive 'geo graphy': writing the earth in ways that both expose the imperial and territorial underpinnings of metropolitan narratives (such as those around 'globalization' or Iraq) as well as make space for 'other graphings, other geographies' from the margins. David Featherstone, advancing the insurrectionist imaginary in C.L.R. James's account of the Black Jacobins, shows how the early radical conception of 'man' espoused by the London Corresponding Society in the 1790s was formed within transnational political networks involving ex slave, free black activists. Such networks, he argues, nurtured forms of 'subaltern cosmopolitanism' and allowed for the circulation of narratives that were often counter hegemonic of establishment positions. In these various endorsements of contrapuntal geographies, 'the subaltern' is that figure who vitally enables western modernity but is (necessarily) effaced from its self narratives.

Linking the issue of effacement to questions of political economy, Melissa Wright's ethnographies of the devaluation of women's work and femicide in the *maquiladora* cluster of Ciudad Juarez, Mexico, bear stark witness to the gendered violence of this effacement, and how subalternity services capital accumulation. Sharad Chari also examines the dynamics of subalternity and capital accumulation, but in the knitwear cluster of Tiruppur, south India, that is dominated by owners from the Gounder caste, many who rose to that position from the ranks of the working class. These owner entrepreneurs of modest peasant and working class origins rely on the exploitation of labor for capital accumulation, but continuously efface this in their self presentations. Calling themselves 'self made men', they narrate their class mobility and industrial success as the outcome of a propensity to 'toil' rooted in an agrarian past of hardship and labor. The lesson for subaltern representation, Chari says, is to never forget that subalternity is a differentiated condition, and that it is as important to translate subaltern difference as it is to translate the difference between elites and subalterns.

The question of subalternity receives a fitting (and yet another geographical) twist in Timothy Mitchell's 'Can the mosquito speak?' – a clever send off of Spivak's original provocation. Noting that stories of the nation, of modernity, or of capital are always written around the 'actions and intentions' of people, he asks how nonhumans enable these stories. Is it not in fact the case that human agency is parasitic on – "draws its force by attempting to divert or attach itself to" – other kinds of energy or logic whose contributions, like those of subalterns, are erased from conventional accounts? If so, what are its implications for who we represent and how? In a familiar yet unexpected move, Mitchell stages the nonhuman as that (subaltern) actor who must be continuously effaced in order to conserve modernity's story of self movement, but whose subject effects keep disrupting this fiction.

See also: Difference/Politics of Difference; Eurocentrism; Hegemony; Imperialism, Cultural; Neocolonialism; Orientalism; Postcolonialism/Postcolonial Geographies; Representation, Politics of; Subaltern.

Further Reading

Arnold, D. (1984). Gramsci and peasant subalternity in India. *Journal of Peasant Studies* 11(4), 155 177.

Chakrabarty, D. (2000). *Provincializing Europe: Postcolonial Thought and Historical Difference*. Princeton, NJ: Princeton University Press.

Chakrabarty, D. (2002). A small history of subaltern studies. In *Habitations of Modernity*, pp 3 19. Chicago, IL: University of Chicago Press.

Chari, S. (2004). *Fraternal Capital: Peasant Workers, Self Made Men, and Globalization in Provincial India*. Stanford, CA: Stanford University Press.

Chaturvedi, V. (ed.) (2000). *Mapping Subaltern Studies and the Postcolonial*. London: Verso. (in association with *New Left Review*).

Coronil, F. (1997). Listening to the subaltern: The poetics of neocolonial states. *Poetics Today* 15(4), 643 658.

Dussel, E. (1998). Beyond eurocentrism: The world system and the limits of modernity. In Jameson, F. & Miyoshi, M. (eds.) *The Cultures of Globalization*, pp 3 30. Durham: Duke University Press.

Featherstone, D. (2007). The spatial politics of the past unbound: Transnational networks and the making of political identities. *Global Networks* 7(4), 430 452.

Forgacs, D. (ed.) (2000). *The Antonio Gramsci Reader: Selected Writings 1916 1935*. New York: New York University Press.

Gidwani, V. (2008). *Capital, Interrupted: Agrarian Development and the Politics of Work in India*. Minneapolis, MN: University of Minnesota Press.

Guha, R. (1982). On some aspects of the historiography of colonial India. In Guha, R. (ed.) *Subaltern Studies I: Writings on South Asian History and Society*, pp 1 8. Delhi: Oxford University Press.

Mitchell, T. (2002). Can the mosquito speak? In *Rule of Experts: Egypt, Techno Politics, Modernity*, pp 19 53. Berkeley, CA: University of California Press.

O'Hanlon, R. (1988). Recovering the subject: Subaltern studies and histories of resistance in colonial south Asia. *Modern Asian Studies* 22(1), 189 224.

Rodríguez, I. (ed.) (2001). *The Latin American Subaltern Studies Reader*. Durham: Duke University Press.

Slater, D. (1998). Post colonial questions for global times. *Review of International Political Economy* 5, 647 678.

Sparke, M. (2005). *In the Space of Theory: Postfoundational Geographies of the Nation State*. Minneapolis, MN: University of Minnesota Press.

Spivak, G. C. (1985). Subaltern studies: Deconstructing historiography. In Guha, R. (ed.) *Subaltern Studies IV: Writings on South Asian History and Society*, pp 330 363 (reprinted in Spivak, G. C. (1988). *In Other Worlds*: *Essays in Cultural Politics*, pp 197 221. London: Routledge). Delhi: Oxford University Press.

Spivak, G. C. (1988). Can the subaltern speak? In Nelson, C. & Grossberg, L. (eds.) *Marxism and the Interpretation of Culture*, pp 271 313. Chicago, IL: University of Illinois Press.

Wright, M. (2004). From protests to politics: Sex work, women's worth and Ciudad Juárez modernity. *Annals of the Association of American Geographers* 94(2), 369 386.

Subjectivity

J. P. Sharp, University of Glasgow, Glasgow, UK

Glossary

Agency The ability to act.
Discourse A discourse can be thought of as a temporally and spatially specific lens through which the world is interpreted, thus defining the boundaries of what can be known.
Epistemology The theory of what can be known.
Hermeneutics The theory of understanding.
Situated Knowledge The idea that there is no possibility of universalized knowledge, but that knowledge is always produced from a specific vantage point.

Introduction

All forms of human geographical thinking presuppose some theory of subjectivity. However, not all approaches have acknowledged this. Positivistic approaches – most closely associated with spatial science – denied that the subjectivity of the researcher had any influence and that the scientific nature of such research produced a purely objective account of the world. Post positivistic geography has focused on the effects of the subject, with debates ranging over the extent to which the individual subject has agency over actions (or whether action is constrained by structures or greater powers of one form or another), the extent to which actions are known (or knowable) to the subject, and the extent to which subjectivity should be centered purely around humans. Each debate has an impact upon methodology in terms of the epistemological value given to the knowledge created.

The Origin of the Knowing Subject

It was with the rise in the eighteenth century of the European Enlightenment that the human subject became the center of knowledge, replacing models which were centered on God as all knowing. The rational Enlightenment subject was regarded as bounded and self knowing. The body contained the subject and the mind, separating the knowing subject from external objects in the world. The Enlightenment sought to produce knowledge through the scientific method – visualizing, naming, and classifying (and in geography, also mapping) the world: capturing the world as an array of facts, and ordering these into a knowable taxonomic structure. It was the subject – perceived as separate from the natural world of objects through the reason afforded by rational enlightenment – which could extract the order of things in the world.

The idea of visualizing and naming the world's objects reinforced the notion of the subject remaining outside of the world which was being ordered and understood. The feminist theorist Donna Haraway has famously talked of this archetypal enlightenment subject as playing 'the God trick' of simultaneously being everywhere (seeing everything) and nowhere (having no limiting perspective on what is being seen). The Cartesian representation of the map is perhaps the best example of this God trick. This highlights the epistemological and ontological beliefs of Enlightenment knowledge: the world existed in a particular order and it was the role of the enlightenment scholar to uncover that order. This naturalistic epistemology suggested that so long as the rational, scientific method was adopted, the order would be revealed the same way to any viewer of it.

Thus, knowledge of the world was something considered to be external to individual subjects. With the development of the European Enlightenment, faith in the progressive nature of science meant that the development of human knowledge was considered to be through a process of the world revealing its secrets to the knowledgeable scholar. The highest achievements of science and philosophy were presented as objective. Any indication of subjectivity (emotions, biases, cultural perceptions, and so on) was banished from this process.

However, despite this apparently universalizing language, objective knowledge was not attainable by all. Women, the working classes, racialized others and children were seen as being unable to achieve objectivity as their minds were not seen to have triumphed over their bodies (women were regarded as too closely aligned to natural patterns due to their role as mothers, through the cycles of menstruation, and through 'womanly' problems such as hysteria and fainting; the working classes and racialized others were believed to be lacking rational intellect and were driven by bodily passions rather than higher goals; and children were not yet sufficiently developed to achieve control and reason). Thus, the knowledge produced by these subjects was invalidated as it was regarded as hopelessly biased, emotional or irrational. There were even discussions about the usefulness of attempting to educate women and racialized subjects as there were beliefs that they would never be able to attain enlightenment and reason. In geography, women were excluded from intellectual societies. Women were not elected as members of the RGS, for instance, until 1913, and their accounts of travel to distant parts of the world were treated as entertainment rather than as scientific knowledge as were the accounts of their male counterparts.

In the postwar years, the social sciences sought to emulate the objectivity of Enlightenment science using techniques that minimized the subjectivity of researchers through methodologies such as spatial science, the use of large data sets and statistics, and the adoption of regularized data collection in questionnaires.

Subjective Knowledge

However, in the last 30 or 40 years, the scientific method has faced serious challenge, in geography and elsewhere. The rise of humanistic geography sought to foreground the human experience of place over the analysis of spatial form and thus could not be objective. This put human experience and understanding of the world at the center of geography. Such an approach negated the possibility of objectivity, replacing scientific methods of analysis and prediction with one based around interpretation (hermeneutics), empathy, and understanding.

In addition to drawing the human as the subject of geography, humanists also turned to examine the subjectivity of the geographer. Humanistic geography's emphasis on interpretation and intersubjectivity meant that the researcher could not be seen as standing outside the research process but instead the researcher had to acknowledge her or his role in the process of interpretation and the production of knowledge. This necessitated the adoption of ethnographic and participatory research methodologies. Rather than trying to erase the influence of the subject in the creation of knowledge, humanistic geography recognized subjectivity as part of the research process, even viewing the subject of the researcher as a tool in the research process. Scholars sought to insert themselves into the outcomes of their work. Rather than using the detached scientific writing of positivism, postpositivistic writers abandoned the passive voice and located agency, wrote in the first person, offered extended quotation from research participants, and experimented with literary devices drawn from the arts.

The recovery of subjectivity also had more political motivations to restore subaltern knowledges that had been hidden from bourgeois history and geography. Marxist and radical theorists are critical of the implicit support by social science of state capitalism, pointing instead to the need to construct alternative working class histories and geographies. Such approaches still often presented an objective knowledge, but one where objectivity was grounded in the class position of the subject.

Feminists argued that there is more than one subject of knowledge. Women have been excluded because of the masculinism of Enlightenment knowledge privileging objectivity over subjectivity, and dismissing women as incapable of true reason. While some feminists challenged this distinction as being based in untenable stereotypes, others accepted the binary distinction but inverted the values, arguing that there exist specifically women's ways of knowing that were based on distinctive 'feminine' characteristics (collaboration, empathy, and intersubjective understanding). This latter position insisted that women were incapable of objectivity due to their embeddedness within networks of care and dependence. This challenged the binary division of subject and object and highlighted the necessity of ontological and methodological subjectivity, but was often critiqued for accepting the binary distinction of essentially masculine and feminine characteristics.

This celebration of the duality of knowledge sometimes suggested that the subaltern view was more objective. Dominant views were supported by power structures within society (whether based around class, race or patriarchal power) and so such views would lack critical rigor. On the other hand, subaltern views could see how power worked and their views would be free of the ideological distortions of social power and therefore more objective. Others were wary of romanticizing the subaltern. From views that recognized the entangled and pervasive nature of power relations, there could never be a simple binary of dominant–subaltern. Class, race, gender, sexuality, and other social identities would be crosscutting. The rise of interest in identity politics based around these aspects of cultural identity offered a model of subjectivity that went against simple binary models, causing some Marxist and feminist theorists to fear a concomitant loss of political efficacy.

Various attempts have been made to produce situated knowledges, arguing for the veracity (and, sometimes, even objectivity) of knowledge but only from one perspective (rather than being totalizing). These partial perspectives (including approaches such as feminist standpoint theory and situated knowledge) sought to avoid the relativism that might arise from suggesting that all views were necessarily incomplete by arguing for the situated and partial nature of all knowledge.

Postcolonial theory makes attempts to recover subaltern subjectivity more complex still. The notion of hybridity is central to postcolonialism, challenging any attempt to divide up knowledge into discrete categories. Such a viewpoint clearly challenges any notion of a distinct subject and object. However, postcolonial critics are attuned to the power relations involved in how this intersubjectivity operates. Most prominently, Gayatri Spivak has examined the ways in which the voice and knowledge of the subaltern has been heard by those who are dominant. Even when apparently expressing their own views however, the subaltern cannot speak, because she/he must use the words, phrases, and cadences of Western thought in order for it to be heard. In order to be taken seriously – to be seen as knowledge and not opinion or folklore – the knowledge of the subaltern is to be translated into the language of science, development, or philosophy, dominated by Western concepts and Western languages. Thus, the subaltern must always be caught in translation, never truly expressing her/himself, but always already interpreted.

This recognition of the influence of subjectivity in academic work required the adoption of methods which recognized how the positionality of the author would inevitably effect the research from the initial con ceptualization of the project (and the motivations behind this), through the response of the research participants to the researcher, methods in which information is analyzed, to what is done with the research outputs. Various strategies have been adopted, to varying degrees of suc cess, to foreground this process, including writing the author into the research narrative, collaboration with research participants in terms of formulating, conducting and disseminating the research, and various writing strategies to emphasize the partial nature of the representation.

Structure/Agency

Theories of power are central to understandings of subjectivity in terms of considering the agency of the subject, in other words, the ability of the subject to act independently. Agency is not only about intentions to act, but also capabilities to act. Ongoing debates about the existence of social structures enabling or con straining human action are at the heart of theories of subjectivity. In the early twentieth century, geographers paid little attention to individuals' actions as these were viewed as being the result of structures operating at a level beyond the individual. For instance, both en vironmental determinism and traditional forms of cul tural geography assumed that people's actions could be explained by recourse to the overwhelming power of the environment or of culture. Marxist approaches also re garded subjectivity to be the product of location within a systemic economic logic that transcended the individual.

Other theories allowed for some exercise of individual subjectivity. Possibilism allowed a range of possible ac tions to emerge from any situation. Spatial science rec ognized the agency of the individual subject but only in as far as it followed the logic of neoclassical economic modeling. Behavioral geography sought to provide a more developed understanding of the psychology behind human agency, to account for what spatial scientists would see as irrational action. Such work included focus on values, aspirations, experiences, and, in some cases, imagination, in their models of decision making (al though much of this work has been characterized by measurable, quantitative methodologies). However, due to focus on conscious motivations, behavioral geography has been criticized for having an underdeveloped understanding of psychoanalysis.

Structuration theory and realism sought to explain how social structures were both outcome and medium of the agency that constitutes them. Both avoid the ex tremes of arguing that society is comprised of individual acts or is determined by social forces by holding these two extremes in tension with one another. It is through repetition that the acts of individual agents reproduce social structures such as institutions, moral codes, norms, and conventions. Nevertheless, this approach still relies upon a binary understanding of society – structuration theory seeks to bridge the individual – structure gap, but it still imagines that it exists.

As already stated, humanistic geography emerged in the 1970s with the stated goal of reanimating geography, to put people and their thoughts, emotions, and beliefs at the heart of the discipline. Focusing on issues such as dwelling, lifeworld, and rhythms, humanistic geographers believed that social life was constructed through human actions. This approach faced criticism for paying too little attention to structures that might limit human agency, even moving into voluntarism, and making methodologies particularist.

Post-Subjectivity

Post structural theory turned preexisting theories of subjectivity on their head. Rather than assuming a subject which has agency to think and act, post structural ac counts see the subject – and its subjectivity – as an epi phenomenon of discourse or performance: the discourse or act creates the sense of a knowing and capable actor. This approach comes from the work of Michel Foucault who has produced a history and geography of the pro duction of subjects and subjectivity. For Foucault, there is no subject prior to knowledge, power, and discourse; instead subjectivity is a product of these. For example, discourses about homosexuality as an identity preexist homosexual subjects, rather than vice versa. As a result, there can be no knowable subject which exists outside of history (and geography): always the subject must be historicized. Foucault proclaimed the 'death of man', by which he meant that emerging trends of Western thought at the end of the twentieth century (including post structuralism) have de invested 'man' as a stable category. Whereas for theorists such as Hegel and Marx subjectivity was a necessary aspect of achieving freedom, for Foucault subjectivity represented enslavement to self identity. Now that it has been so thoroughly decentered, the subject cannot serve as an absolute ground of knowledge, as a transcendental historical figure, or as an ultimate justification for moral theories.

Foucault's ideas have been developed by a number of feminist and queer theorists. Perhaps most significantly, Judith Butler has critiqued stable understandings of subjectivity to argue that gender identity is constituted not in any essentialist characteristics or through a founding act, but through a regulated performance. The performance of identity establishes the fiction of a subject who is performing. Butler argues that heterosexuality is only established as the norm through a performance of heterosexual acts which give this identity the image of authenticity. Dominant sexualities are no more 'normal' or 'natural' than alternatives, but the impression of natural, stable identities (and deviance from this) is established through a set of normalizing performances that have been invested with a great deal of social and economic power.

Rather than seeing it consisting of the dramatic moments in life, most have understood performance constituting subjectivity through mundane, day to day improvisations. Micro scale analysis has turned attention to bodily knowledge, the ways of getting by each day, rather than performances which are coherently thought through and reflected upon. Some have pointed to the body as the most immediate space for geographers to examine. Feminist and queer theorists have turned attention to the body as the site of meaning, where discourses of power, and of divisions between inside (subject) and outside (object) are inscribed. Theories of the body have considered how dominant power writes particular subjects through microdisciplinary practices such as fashion, diet, and exercise. However, such approaches also consider the body as a site of resistance to these dominant powers through performances such as cross dressing, piercings, tattoos, and self harm. This has led to debates upon the extent to which these performances are transgressive or actively resistant.

This focus on everyday minutia has been critiqued as ignoring important relations of politics as it can result in an understanding of subjectivity devoid of any attention to power relations. However, feminist and queer theorists have highlighted the fact that performances are not simply freely chosen. Examples of fashion, clothing, and the display of symbols of family life in the workplace have allowed study of performances of dominant body cultures, compulsory (hetero)sexuality and the masculinism of work.

Importantly for geographers, the concept of performance requires an understanding of space: it is important where subjectivity is performed as this influences the outcome and meaning of the performance and so the meaning and effect of subject identity.

The psychoanalytic turn in geography similarly sought to decenter the knowing subject of knowledge. Freud and those who followed examined, in different ways, the establishment of the individual as a subject through the break with the mother (from whom the newborn child is initially unable to distinguish itself). This break is seen, for instance, as the result of patriarchal power for Freud, the symbolic patriarch for Lacan, or the realization of abjection for Kristeva. Psychoanalytic theory insists that the unconscious is inaccessible to the subject but that its dynamics persistently intervene in more conscious states such as dreams, thoughts, speech, and precognitive practice. This therefore insists that subjectivity cannot be completely known, the subject cannot be completely mapped. The subject must always be to an extent unknowable and unstable. This presents a critique of feminist research which situates the subject through self reflexive methodologies as such situated knowledge assumes the knowability of the subject to establish its location. Clearly then the psychoanalytic turn has impacts on methodologies.

For many, psychoanalytic approaches are regarded as essentializing. Some geographers have thus tried to situate psychoanalysis into the particularities of context, for instance, considering the social processes of exclusion and abjection. Studies of agoraphobia and other conditions have linked the interior spatiality of the psyche to the spatiality of everyday life.

The rise of nonrepresentational theory similarly attempts to get at knowledge without an entirely conscious theory of subjectivity, using precognitive or bodily knowledges or responses to mundane situations rather than adopting discursive methods. This has led to further questions about assumptions concerning subjectivity in geography being limited to human subjects. Human subjectivities are supplemented and maintained through IT and communications networks, medical implants, and genetic manipulation, and human social life is also populated by nonhuman animals as companions, combatants, and food. Here agency is liberated from the subject. It is regarded as an external quality, owned by the network or collective rather than being internal to the individual subject.

Methodologically, post subjectivity has proved very troubling. The radical challenge to objectivity has led some to consider all knowledge to be no more than language games, concepts that cannot be translated because of the dismissing of any claims to transcendental meaning. Science becomes one knowledge alongside so many others with no recourse to explaining its superiority. Nonrepresentational and psychoanalytic approaches raise challenging questions about methodological approaches to get at the nondiscursive and subconscious and unconscious aspects of subjectivity.

There are also profound political and ethical questions about the responsibilities of agency and the politics of academic work. Some feminists have reacted more powerfully to postmodern and post structural pronouncements of the 'death of the subject' wondering whether this had occurred just when the male, white, subject might have had to share its status with those formerly excluded from subjectivity.

However, there are elements of transgression and resistance possible in the post structural understandings of subjectivity. The performative theories of subjectivity provide for the possibility of politics and of transformation. If there is always a compulsion to repeat then identity is never complete, further performance is always required. As identity is only secure when performed 'correctly', this offers a great potential for subversion.

See also: Embodied Knowing; Enlightenment Geography; Ethnography; Feminist Methodologies; Foucauldianism; Humanism/Humanistic Geography; Masculinism; Non-Representational Theory/Non-Representational Geographies; Performativity; Poststructuralism/Poststructuralist Geographies; Psychoanalysis; Resistance; Situated Knowledge, Reflexivity; Spatial Science; Structuration Theory.

Further Reading

Bondi, L., Avis, H., Bankey, R. *et al*. (2002). *Subjectivities, Knowledges, and Feminist Geographies: The Subjects and Ethics of Social Research*. London: Rowman and Littlefield.

Butler, J. (1990). *Gender Trouble: Feminism and the Subversion of Identity*. London: Routledge.

Cupples, J. (2002). The field as a landscape of desire: Sex and sexuality in geographical fieldwork. *Area* 34, 382 390.

Foucault, M. (1979). The History of Sexuality. Volume 1: An Introduction (first published 1976) Hurley, R. (trans.). Harmondsworth: Penguin.

Gregson, N. and Rose, G. (2000). Taking butler elsewhere: Performativities, spatialities and subjectivities. *Environment and Planning D: Society and Space* 18, 433 552.

Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies* 14, 575 599.

Limb, M. and Dwyer, C. (2001). *Qualitative Research Methods for Geographers*. London: Arnold.

Mascia Lees, F., Sharp, P. and Cohen, C. B. (1989). The postmodern turn in anthropology: Cautions from a feminist perspective. *Signs* 15, 7 33.

McDowell, L. and Court, G. (1994). Performing work. *Environment and Planning D: Society and Space* 12, 27 50.

Nash, C. (2000). Performativity in practice: Some recent work in cultural geography. *Progress in Human Geography* 24, 653 664.

Philo, C. and Parr, H. (2003). Introducing psychoanalytic geographies. *Social and Cultural Geography* 4, 283 293.

Pile, S. and Thrift, N. (eds.) (1995). *Mapping the Subject: Geographies of Cultural Transformation*. London: Routledge.

Spivak, G. C. (1994). Can the subaltern speak? In Williams, P. & Chrisman, L. (eds.) *Colonial Discourse and Post Colonial theory*, pp 66 111. New York: Columbia.

Whatmore, S. (2002). *Hybrid Geographies: Natures, Cultures, Spaces*. London: Sage.

Suburbanization

A. Mace, University of Westminster, London, UK

Glossary

Fordism Taken from the Ford motor company's model of mass production, the term is used to describe an economic system driven by centralized large-scale industrial production for a mass consumer market. This economic model then influences political and social arrangements in order to maintain itself.
Micro-Generation The use of solar panels, wind turbines, and other sources of power generation on a small scale sometimes attached to domestic and commercial buildings.
Post-Industrial City A term used to describe former industrial cities where the economy is now driven by service industries including financial services, education, leisure, and tourism.
Radial Transport Generally used in relation to public transport, these are links that connect the center of a settlement to outlying areas including suburbs.
SUV or Sports Utility Vehicle Much criticized for their high petrol consumption and for their marketing as a heavily defended private environment on the public highway – they have been referred to as a gated community on wheels; one range of SUVs is named 'The Suburban.'
The Chicago School From the University of Chicago and influential for thinking on the nature of the modern-day city, especially during the first-half of the twentieth century. The ecological model of the city, associated with the Chicago School, sought to explain social competition for space within the city in terms of a Darwinian struggle for advantage and survival.
Tupperware The trade name for a range of products associated primarily with food storage. Culturally, the product is often linked with an idealized version of the consumption-based suburban family life of the 1950s and 1960s. That is, a married couple with dependent children where the woman homemaker, or housewife, did not have paid employment.

The Roots and Spread of Suburbia

Edge of city development has been around in one form or another for as long as cities. This development around the fringes of the city has encompassed both exclusiveness and exclusion. Dotted around first century Rome were *villae sububanae*, more in the nature of the modern day second home. They provided Rome's elite with respite from the demands of the city. At the other extreme a quite different type of suburban development had attached itself to cities – the zone of overspill to which the poor and undesirable were exiled. Southwark to the south of the city of London was, in Shakespeare's day, infamous as the location for various entertainment and business activities which were sometimes disreputable and which were formally excluded from the city proper. In more recent times, Toronto's early suburbs were often the product of self build by incomers who could not afford to live in the city and who struggled to make ends meet on the suburban fringe.

The low density town/country suburban form, that is the focus of this article, was profoundly influenced by the 'garden city' movement, attributed to Ebenezer Howard, and which advocated urban dispersal from overcrowded industrial cities. This informed a policy of urban dispersal and low density housing in Britain throughout much of the twentieth century. Both the US and the UK lay claim to being the birthplace of suburbia. Robert Fishman sets out the case for the UK, arguing that fundamental changes in family form led by London's and later Manchester's Georgian merchants gave rise to the classic town/country form of suburbanization. This model is closely associated with a new economic division of labor within the family. Women were gradually excluded from their hitherto central role in the family business and cast as full time mothers tending to the children and domestic management. In this reading, suburbia has been implicated from its earliest days with a patriarchal family structure that has oppressed women.

Increasingly, the Georgian inner city was viewed as morally corrupting and an inappropriate residential location for women and children. Thus Georgian merchants gradually separated off the business premises from the family home moving the latter out of the city. The male city was separated off from the female domestic suburb. Robert Fishman argues that, this was first fully realized in Manchester (England), where the complete residential withdrawal from the city center by the merchant class led to the world's first central business district. Such examples are more than just historical detail. This configuration of Manchester carried through into the late twentieth century. By the mid 1980s only a few hundred people lived in the city center, so creating a serious challenge for those seeking to regenerate the post industrial city by bringing a residential population back from the suburbs to the center.

Meanwhile, Kenneth T. Jackson sets out the case for the US being the true home of mass suburbanization. The economic downturn in the UK's economy toward the end of the nineteenth century limited the growth of suburbia there while it took off in the USA. Further, in the UK with less space and higher land values the typical suburban form became the semidetached family house; in the US and Australia the detached house on its own plot more fully realizes the rural/urban dream. It seems reasonable to assert that the US led the world in mass suburbanization and has developed it more fully. If suburbia has reached its zenith in the US then it is also here that it is most closely associated with sprawl. Veronica Strong Boag *et al.* claim that the suburbs are more defining of national culture and identity in the US than is the case in either Canada or the UK. Flowing from this, Fishman has argued that the US suburban lifestyle, serviced by gas guzzling SUVs, is a defining political issue in a world concerned with environmental matters. The idea of a distinct suburban culture is developed further in the next section. First, more objective or scientific explanations for the growth of suburbia are considered.

Suburbanization is inextricably linked with the growth of mass public transit systems, and car use. Clearly there is a relationship between the development of transit technology and the growth of suburbia. Changes in transportation have made it possible to develop housing at ever greater distances from the city center as transport systems have allowed commuters to cover greater distances to work in the center more easily and economically. Horse drawn buses, trams, and motor buses and then trains and metro systems have all expanded the reach of suburbia. However, there is not always a simple relationship between transit systems and suburban development. There are instances where new transport infrastructure did not lead immediately to suburban development and also, of where suburban development preceded transport links to the center.

The relationship between the middle class suburb and mass transit has long been an ambivalent one. As Robert Fishman notes, the exclusive suburb needed good transport links with the city center, so that residents could commute to their work in the center. But to sustain its status, its exclusiveness, it was essential that those transport links should not provide for the in migration of working class households. Suburban transport had to make the center accessible but keep the suburb exclusive. The cost of public transport served to discourage the development of housing for working class households in many cases. In other instances, mass transportation was used to facilitate the movement of the working class out of the center. An example is found in the UK where, toward the end of the nineteenth century, central government required the private train companies to provide cheap workers' fares. This allowed for the development of working class suburbs, for example in Tottenham, North London, where former upper middle class estates were broken up and developed with rented workers' housing. The effect of this cheap fares policy is still imprinted on the social pattern of London's suburbs today. In addition to transport systems, the location of employment has also influenced the social mix within suburbia.

The social and transport models discussed so far tend to assume that suburbia represents, in the first instance, a purely residential expansion of the city. However, this is challenged by Richard Walker and Robert D. Lewis who argue that suburban expansion has typically been accompanied by the outward movement of industry sometimes encouraged by local and national policy. The transport infrastructure developed in suburbia was attractive to industry – particularly if suburban rail networks were part of a national system or where new road infrastructure was provided. Therefore, not all suburban development has been based on commuting into the center from an exclusively dormitory or residential suburb. The suburbs have developed complex commuting patterns which reflect the location of employment in the periphery as well as in the center of cities.

Beyond these more particular accounts, suburban growth has been explained as the product of both social and economic competition. The Chicago School sought to develop an analogy between the Darwinian competition in nature and the way in which communities vie for space and favored locations in cities. In their description of early twentieth century American cities, migrants move into declining inner city areas which are vacated by upwardly mobile residents moving out to suburban areas. These new suburbanites were seeking competitive advantage; they were responding to the perceived socioeconomic advantages that suburban home ownership could bestow upon them.

An overarching economic logic is offered by David Harvey who provides an analysis of suburbanization as being the spatial expression of capital seeking out new investment opportunities. When investment in the inner city appears less attractive (often because of the poor environmental conditions that capital itself has created) new investment opportunities become apparent around the city fringe where cheaper land and development costs combine with a better environmental quality. The unspoilt environment of the city fringe serves as a free good to the investor, further encouraging investment in suburban development. In this model, the suburbs can be viewed as the product of capital exploiting an investment opportunity; it has an economic inevitability about it. Certainly, detailed studies of London's explosive suburban growth during the mid twentieth century indicate that a combination of local, regional, and national eco nomic factors fueled the expansion. Cheap land and building materials, a willingness by landowners to sell, the increasing availability of mortgages, and a relative scarcity of accommodation to rent, along with the long term shift of employment and population to the southeast of England, all combined to fuel a suburban housing boom.

However, explanations of suburbia which view it as being shaped solely by unavoidable structural forces are rejected by Doreen Massey. She recognizes the need to accommodate the economic imperative but seeks to go beyond this, arguing that there is more to places than the dictates of capital. In short, while the logic of capital accumulation may explain something of suburbanization, we also need to understand the role of human agency if we are to better understand suburbia as a place or more accurately, as a variety of places. Suburbia is often rep resented visually by aerial photographs which contain implicit judgments about the regularity of this built form and its land hungry nature. However, as Lars Lerup observes, in his graceful account of Houston, sprawl looks quite different on the ground. In the next section the focus is on the internal dynamics of suburbia and on the external relationship with the city.

The Relationship with the City

Suburbanites have been criticized for their conformity to the prevailing demands of capital and its associated social norms; for blithely accepting the mass consumer society, traditional domestic roles for women, and the promotion of the individualistic and privatized family unit. In his study of comfortable middle class suburbia, Robert Fishman concludes that successful bourgeois suburbia has always been more about what it excludes than what it includes.

Concern for the social impact of suburbanization has reflected national concerns. In America the focus has been on race and suburbanization, while in the UK the emphasis has been on class. In both cases though there is a sense of the calculating suburban type. Suburbia has been characterized as containing privileged social groups seeking to entrench their advantage through spatially segregating themselves from those perceived to be of inferior or undesirable status. Once in suburbia the competition continues. Harold J. Dyos describes how aspiring groups would seek to move to where a better class of person lived. However, if the incomers were seen to be lowering the status of the neighborhood, the higher status residents might move to a more exclusive suburb. Through this process it is important to note that suburbs can and do decline in status in the same way as inner city areas.

Criticism of suburbia as socially exclusive tends to suggest that suburbia has been successfully shaped by a single suburban type. Furthermore, this classic subur banite is often defined as being in opposition to, or apart from, the city resident. The suburbanite is everything that the city resident is not. This approach is encouraged by the existence of a city/suburban binary which sets the two in opposition. Both the detractors and supporters of suburbia have often resorted to a quite crude city/suburb opposition in order to reinforce their point. However, as Cloke and Johnston have argued, the use of geographical binaries such as city/suburb can undermine, first our understanding of the suburbs themselves and second our understanding of the relationship between the city and the suburb. This is because, once established, binaries tend to underestimate differences within each of the categories (differences within the city and within sub urbs) and overestimate the differences between the two.

Much writing on mass suburban development of the mid twentieth century, including from the medical pro fessions, associated the suburbs with a form of female neurosis. The bored stay at home housewife with young children found herself isolated in her new home, away from the support of the traditional network of family and friends. To some extent these studies reflected their time, when women were less likely than now to be in em ployment or to have access to a car. Writers in the Feminist tradition such as Robyn Dowling have sought to challenge the one size fits all view of women in suburbia. Rather, the case is made for the need to understand the variety of women's experience there. Similarly, in the UK, studies of working class suburbanization have sought to demonstrate that the process of urban dispersal has been responsible for the breaking up of traditional working class neighborhoods and their associated bonds and social networks. In these studies the working classes found themselves divided and alienated in neighborhoods dominated by the middle classes and middle class values.

Suburbia undoubtedly has excluded, and has contained examples of traditional or conservative values. However, it is probably reasonable to assert that suburbia is still only one place among many where these views are expressed; the suburban experience reflects as much as it propagates social values. Isolated young mothers, a changing working class, and continuing racial division are all found in suburbia, but they are not the product of suburbia nor are they exclusive to it. Suburbia cannot be understood in isolation from the society that it is a part of.

Moreover, suburbia is not simply the opposite of the city. This point is developed further through considering the suburbs as a site of consumption. The new technol ogy of the home in the middle of the mid twentieth century meant that mass suburbanization was associated with the spread of domestic appliances such as the fridge, telephone, the radio (or wireless), washing machine, and of course, the car. Both the housing and the appliances within reflected the Fordist economic model, standard ized products, centrally produced for a mass market. In the UK, the cognoscenti were quick to criticize the standardized nature of the houses and the crassness of mass consumption which was associated with poor taste and a lack of culture. Famously, in the 1950s the architect Ian Nairn summed up the sneering attitudes of the elite in the term subtopia. Nevertheless, as Jane Jacobs ob served, there are numerous means through which priv ilege is captured within the urban fabric. Changes in housing markets in the inner city and the suburbs may be blurring the traditional differences between the center and periphery; the city may, in some respects, be be coming suburban.

The private suburban culture of house parties and Tupperware is traditionally contrasted with the public life and high culture found in the city. Even if this en during binary were ever entirely true, it appears to be progressively more unsustainable. This point is illus trated by changes in the gentrification process. Early gentrification was associated with individual's self im proving properties in neighborhoods which they valued for the social mix. The growth of supergentrification has seen the influx to the inner city of the hyperwealthy who appear not to value the diversity of their neighborhood in the way that the original gentrifiers did. Rather, as Diane Perrons has observed, the inner city is becoming about diversity only insomuch as it encompasses the superstars and their servants.

Gated communities, the urban village movement, and a more general influx of like minded middle classes into areas may all be producing in the inner city the type of socially homogeneous communities (and built form) that suburbia was criticized for perpetuating. This blurring of city and suburb is integral to the urban village movement which seeks to make the urban more inviting as a place to live by invoking the imagery of rural settlement and geographic community within the city. Arguably, urban villages are designed to appeal to someone more clearly identifying with stereotypical traditional suburban val ues. They are not targeted at the demographic that would be attracted by the traditional grittiness of the city, as was the case with early gentrification.

While changes in the inner city housing market may be tempering the richness and diversity that led Jane Jacobs to extol the inner city, in suburbia major demo graphic change may be producing a reverse effect. In London, some suburban boroughs are now characterized by an ethnic mixing rivaling some inner London bor oughs; the suburbs are home to mosques, temples, and other ethnic infrastructure as minority ethnic groups migrate out from the inner city. More recently the tra ditional pattern of immigration into city centers followed by dispersal out to suburbia has been challenged. In the US there is evidence of a direct move into suburbia by, for example, Korean immigrants.

The academic literature on the suburbs has long ac cepted that suburbia is not homogeneous. There is rec ognition that there are a wide variety of family forms in suburbia, also that women's experience of suburbia is multivariate and that, in terms of class and ethnicity, suburbia is diverse. However, suburbia still tends to be subject to the external gaze. Relatively few studies give a voice to those who actually live in suburbia. Suburbia is commonly framed and studied in terms of its deficiencies, what it should be, rather than being better understood for what it is. There is now a focus by policy makers on suburbia as an unsustainable urban form in need of remediation. In the final section, consideration is given to the future of suburbia.

Suburban Futures/Issues

Any suburban development which seeks to combine the rural and the urban is, by its very nature, always in danger of becoming entirely urban. London's Victorian semidetached villas, once apart from the city are now very much a part of it. Later developments, although much farther from the center, have been profoundly in fluenced by increasing car ownership and use and the suburbanization of industry. As cities have sprawled, so have subcenters of commerce and entertainment de veloped in suburban areas. Indeed as Ed Soja argues, based on the experience of LA, it may be that polycentric suburbia is the new city. In an attempt to capture some of the economic and social changes that are occurring in suburbia, terms such as technoburb, exurb, and edge city have been employed. These are united by the concept of the suburb as becoming progressively more separate from its old, core city center and more urban in its own right.

With the rise of the polycentric region of multiple networked subcenters, it may be that the traditional model of the dormitory suburb feeding workers into a central business district simply no longer holds. Again, the fortunes of suburbia cannot be understood in isol ation from wider social and economic restructuring. In this context one may question what sort of place suburbia is becoming in an increasingly urban and global world. Lefebvre offers the opportunity to conceptualize sub urbia as part of an urban dynamic rather than in isolation. One of his enduring interests was how cities simul taneously spread out physically while concentrating power at the center. Under these conditions what benefits or disadvantages does a suburban residential location confer? Why do people choose to live there? And how do they shape suburbia? To some extent the answers depend on which country the suburb is located in. Suburban ization may still represent either the rejection of the city by those seeking economic or social benefit (as has been the case in the US and UK), or a rejection by the city of those who do not have the financial wherewithal to live in the center (France, Italy). Though, even for the subur banite in the UK and the US, it still depends on which part of suburbia she or he lives in. While some suburbs continue to do well others are struggling.

The theoretical concepts of Pierre Bourdieu offer the potential to build on Doreen Massey's call to understand places as shaped both by broad economic forces and the people who reside there. Through his concept of 'habi tus,' Bourdieu offers those studying suburbia the oppor tunity to move beyond the traditional focus of class and race through an alternative focus, culture. In a study of suburban Manchester, Mike Savage *et al.* developed a study of suburbia that builds on Bourdieu's concept of habitus to develop an account of the way in which sub urbanites construct and influence the suburban experi ence in an increasingly interconnected city, regional, and global context. The term 'elective belonging' is employed to capture the dynamic sense of how people come to construct a sense of place in modern suburbia.

While parts of suburbia may be becoming more urban, for others it is still not urban enough. In the UK the attempt to rehabilitate city center living through the urban renaissance debate means that suburbia remains out of favor with policy makers. The urban renaissance debate is driven by esthetic judgments that favor the supposed social qualities of city life (inclusion and di versity) as opposed to low density suburbia (exclusivity and homogeneity). However, the urban renaissance de bate also draws much of its power from the environ mental imperative – the need to achieve more sustainable lifestyles. Suburbia now finds itself out of favor with policy makers because of its land hungry, low density form and its dependence on the car.

Although the suburbs are more resource hungry than the city center they do also offer some potential in en vironmental terms, including the possibility of increasing biodiversity, micro generation of power, harvesting rainwater, and growing vegetables. Furthermore, as sub urban areas increasingly develop their own subcenters of employment, it is not necessarily the case that com muting times or distances will increase as suburbia spreads. Indeed, Tokyo is an example of a high density city with long commute times and distances. However, polycentric suburbia, with commuting between suburban centers, is not conducive to public transport which favors radial routes into city centers. Therefore, suburbia is likely to remain open to criticism for being relatively unsustainable in comparison to the city, not least because of its relative defendence on the car.

To conclude, although numerous academic studies have highlighted the difference that exists within suburbia and the change that is taking place there, the perception of suburbia remains static. To be suburban is still frequently employed as a shorthand term for safe, cautious, and small minded. Popular culture in both sides of the Atlantic and in Australia continues to perpetuate a view of suburbia as a place of secure domesticity; of sitcoms, and farce. A common feature of suburban studies and policy inter vention is that they seek to interpret suburbia from without rather than understand it from within. As a consequence, suburbia is typically defined for what it is not rather than for what it is. It may be that there is some catching up to do by the media and policy communities; to paraphrase Ed Soja, "Suburbia is no longer what it used to be."

See also: Culture; Gender in the City; Gentrification; Housing; Industrial Location; Informational City; Polycentricity; Transport, Public; Urban Village.

Further Reading

Clapson, M. (2003). *Suburban Century: Social Change and Urban Growth in England and the USA*. Oxford: Berg.
Dyos, H. J. (1961). *Victorian Suburb: A Study of the Growth of Camberwell*. London: Leicester University Press.
Fishman, R. (1987). *Bourgeois Utopias: The Rise and Fall of Suburbia*. New York: Basic Books.
Jackson, K. T. (1985). *Crabgrass Frontier: The Suburbanization of the United States*. Oxford: University Press.
Lerup, L. (2000). *After the City*. Cambridge, MA: MIT Press.
Richards, L. (1990). *Nobody's Home: Dreams and Realities in a New Suburb*. Oxford: University Press.
Savage, M., Bagnall, G. and Longhurst, B. (2005). *Globalization and Belonging*. London: SAGE.
Sunders, W. S. (ed.) (2005). *Sprawl and Suburbia*. Minneapolis, MN: University of Minnesota Press.
Silverstone, R. (1997). *Visions of Suburbia*. London: Routledge.
Soja, E. W. (2002). Sprawl is no longer what it used to be. In GUST (Ghent Urban Studies Team) (eds.) *Post Urban*, *Ex Fragmentations*, *Sub and Dis Constructions*, pp 76 88. Rotterdam: 010 Publishers.
Strong Boag, V. (1999). What women's spaces? In Harris, R. & Larkham, P. J. (eds.) *Changing Suburbs: Foundation, Form and Function*, pp 168 186. London: E & FN Spon.
Thompson, F. M. L. (ed.) (1982). *The Rise of Suburbia*. Leicester: University Press.
Tonkiss, F. (2005). *Space, the City and Social Theory: Social Relations and Urban Forms*. Cambridge: Polity.
Walker, R. and Lewis, R. D. (2001). Beyond the crabgrass frontier: Industry and the spread of north American cities, 1850 1950. *Journal of Historical Geography* 27, 3 19.

Superpower

J. O'Loughlin, University of Colorado, Boulder, CO, USA

The definition of a superpower is like that of an elephant; it is hard to describe but most people have a fairly clear notion of what it entails. At a minimum, a superpower is a state that has a global interest, a worldwide reach, is an economic or military giant (usually both), has an influ ence on world affairs through pursuit of its own interests, and promotes its values. What is frequently debated is the point that a 'great power' becomes a 'superpower' and when a 'superpower' becomes a 'hegemon'. This defin ition from the historian, Paul Dukes, that "a 'superpower' must be able to conduct a global strategy including the possibility of destroying the world; to command vast economic potential and influence; and to present a uni versal ideology" encapsulates the elements of superpower.

Superpower Rivalry

The term 'superpower' reached public consciousness in the early years of the Cold War as the rivalry between the United States and the Soviet Union heated up. A search of Google News Archives showed a peaking of the word 'superpower' in news stories in the late 1980s to early 1990s as the Soviet Union imploded and the US was left astride the world stage as the only superpower. The concept of superpower has implications not only for the daily relations between states but also for the concept of the stability of the international system itself. While the United Nations (UN) encompasses all states within its ambit, the body was structurally constituted by the winning powers of World War II with a controlling authority in the Security Council for the five super powers of the day (US, United Kingdom, Soviet Union, China, and France). When these five divided into two ideological camps in the late 1940s, and as other great powers (re)appeared on the world stage (Japan, Germany, India, and European Union), the calls for a revision of UN, the key institution that facilitates cooperation in world affairs, and advance international norms, have grown.

Like most geopolitical writings, discussion of super power status takes on a strongly ethnocentric character. How does one's own country measure up to opponents and other states in the world system? As early as 1835, Alexis de Tocqueville spoke of the United States and Russia as 'two great nations' on a collision course as if "called by some secret design of Providence one day to hold in its hands the destinies of half the world" (quoted by Paul Dukes). Research by the historians Paul Kennedy and George Modelski, political scientists William Thompson and Joshua Goldstein, among others, has documented the cycles of power since the beginning of the modern world system around 1500 using a variety of economic, military, and geographic territorial indicators. Superpowers (at various times, including the Spanish Empire, Portugal, Britain, The Netherlands, France, Prussia/Germany, Japan, Russia/Soviet Union/ Russia, and China) have been through cycles of growth and declines, experienced periods of global competition at times of relative equality in power status, built alliances to promote or protect their positions, and promulgated ideological and cultural values to go along with their foreign activities and settlements. By Jack Levy's calculations, over 60% of all years between 1600 and 1980 saw wars between the superpowers of the day.

Superpower Rivalries

Though they share a cyclical view of superpower rivalry with regular phases of growth and decline for global contenders and accepting that transitions between peri ods of global leadership are times of uncertainty and contention, the world systems models of George Mod elski and Immanuel Wallerstein differ dramatically in the length of the cycles (50 years for Wallerstein, 100 years for Modelski), the forces driving and underlying global power ambitions (economic leadership for Wallerstein and naval–military leadership for Modelski), the process of transition from one superpower leadership to another (global military conflict for Modelski and loss of eco nomic leadership for Wallerstein) and the prospects for continued US superpower leadership in the twenty first century (strong for Modelski and bleak for Wallerstein). Joshua Goldstein in his book *Long Cycles* goes a long way toward integrating these (apparently) conflicting inter pretations of the last half millennium with his presen tation of detailed annual measures of economic and military status. Though there is widespread acceptance by both camps that economic and political–military leadership are strongly connected (there is only one world system, an integrated economic–political logic as argued by Christopher Chase Dunn and by Colin Flint and Peter Taylor in their textbook), major differences are obvious in the sources of evidence used to document past phases of leadership and to make predictions for the future.

The experience of Britain and its superpower status from about 1600 to about 1945 is central to understanding the focus on the prospects for further American superpower status. The American nineteenth century geopolitical author, Alfred Thayer Mahan, argued that the best hope for a rise to global prominence for the US was to replicate the British experience. Like Britain, the US was an offshore country with advantageous physical and economic geographic characteristics favorable for the development of a dominant navy. Mahan promoted a naval strategy for the US, including the occupation and development of strategically located islands such as Hawaii, The Philippines, and in the Caribbean. The beginnings of the analysis of seapower–landpower dichotomy in world affairs are visible in Mahan's key work, the *Influence of Seapower upon History, 1660–1783* published in 1890; his legacy is continued by Modelski and Thompson in their arguments about the significance of naval power for global leadership. There are further resonances of Mahan in Halford J. Mackinder's heartland model and later in George Kennan's famous containment strategy that was implemented as a US policy toward the Soviet Union's heartland location.

The Cold War

In the Cold War, 1948–88, the US and the Soviet Union were the only superpowers, though their contest was always highly unequal. The US bloc had about four times as many allies as the Soviet one, and the cumulative economic disparity was of a similar nature. What evened the competition somewhat was the military element after the mid 1950s when the Soviet Union developed the ability to launch nuclear missiles to all parts of the US alliance and maintained a strong presence in Eastern Europe face to face with the North Atlantic Treaty Organization (NATO) forces. The standoff had a final phase in the introduction of middle range nuclear missiles (up to 2500 km) by both sides into Europe in the early 1980s. In his statistical analysis of the US–Soviet relations during the Cold War, Joshua Goldstein showed that both superpowers reciprocated each other's actions within a short time frame of up to about 2 months. In the first two decades, the US reacted consistently to Soviet actions, but in the last two Cold War decades, the Soviets responded predictably to US actions. Typical reciprocal actions were increasing support to allies, moving troops to flashpoints, signing or abrogating treaties or arms agreements, initiating a new weapons program or significantly increasing expenditures, or bilateral meetings and negotiations. Measuring and plotting the level of cooperation–conflict between the two superpowers shows a deep freeze in relations in the first decade of the Cold War, a period of quick oscillation from cooperation to conflict between the mid 1950s and late 1960s, just over a decade of *détente* to the late 1970s, a 'second cold war' during late Carter and early Reagan presidencies, and an improvement after the accession of Mikhail Gorbachev to the Soviet Communist leadership in 1985.

The US–Soviet superpower contest was also marked by a distinctive political geography. Though the term 'cold peace' is sometimes used to describe their bilateral relations (through actions in supporting allies and actively engaging in third party wars), the superpowers were also engaged in 'hot wars'. Jan Nijman has mapped and analyzed the changing geography of these surrogate actions and his maps clearly show a diffusion of the proxy wars from the margins of the heartland (the Soviet Union and its Eurasian neighbors) in the early years to the Third World by the mid 1970s. Early flashpoints included Berlin (1948), Turkey (1950s), Southeast Asia after the withdrawal of France in the mid 1950s, Korea (1950), Iran (mid 1940s to early 1950s), and Greece (1947). By the 1960s, the Soviet Union could project its power beyond the Rimland, bordering its vast territorial extent, and Cuba (1963) became the stereotypical Cold War standoff after Soviet nuclear weapons were delivered to the island.

During the long wars in Southeast Asia (Vietnam, Laos, and Cambodia), the Soviets provided significant aid to the Viet Cong and to the North Vietnamese regime. The Reagan administration accused Soviet President Leonid Brezhnev of using the *détente* with the US to expand Soviet influence beyond the ring of containment and advocated a policy of 'rollback' to counter Soviet support for left wing regimes in Africa, Asia, and Latin America. These included newly independent Portuguese colonies (Guinea Bissau, Mozambique, and Angola), Somalia (but later Ethiopia in a reversal of earlier superpower clientelism), Syria, Vietnam, Nicaragua, and rebels in El Salvador. American support for the 'contras' in Nicaragua was designed to overthrow the leftist government of President Daniel Ortega, justified by arguments about the geographic proximity of this Central American country to the US, as well as appeals to the long adherence to the Monroe Doctrine of 1823, in American policy toward Latin America.

The denouement to this geographic contest in the late 1980s was the strong American support to the enemies of the Soviet backed government in Afghanistan. After a decade there, the Soviet Union withdrew in 1989, and within 5 years, the Taliban took control of Kabul from their bases in the south and in Pakistan where they had received indirect US support with Pakistani government assistance. In turn, the Taliban were removed from power in 2001 by the US and a coalition of Afghan regional opposition movements also supported by Russia who feared a radical Islamist state near its southern borders. The events in Afghanistan, repeated elsewhere as rebels gained control of capitals and former government forces became rebels often with bases in neighboring countries, provide clear evidence that superpower involvement in local conflicts can backfire when power shifts, or regional alliances collapse. However, any salutary lessons of the dangers of regional entanglements during the Cold War competition are lost on the United States which, according to Chalmers Johnson and James Carroll (in 2004) owned or rented 702 overseas bases in about 130 countries and has another 6000 military bases in the United States and its territories.

Hyperpower: The Endgame of Superpower or Empire by Another Name?

Of all of the actual and putative superpowers of the past 500 years, none achieved the leadership over rivals or the heights of cumulative and dispensable power that the United States attained after 1945. After a half century of superpower rivalry and tit for tat military and political actions in the Cold War, the past two decades have been an era of adaptation to a 'new world order' in which putative American dominance has become undermined by its own actions and those of state and nonstate actors. By the end of the Cold War in the early 1990s, Paul Kennedy noted that the American naval supremacy dwarfed all the other navies in the world combined. In 2007, the US was on the verge of equaling the military expenditure of 'all' other countries in the world combined.

What really distinguished America's period of superpower leadership (or hegemony) was the cultural power that the US brought to the global scene, especially after 1945. Blending a focus on freedom and democracy with free enterprise and open markets, successive American governments differed from previous superpower leaderships in promoting a seductive mix of political, economic, and cultural norms modeled on the American experience. A rapid development in the technology of mass media (print, radio, telecommunications, television, films, and the Internet) over the past half century allowed the 'American dream' to be diffused globally through the actions of major corporations (Disney, TV networks, music companies, and movie companies) and brought American life and values to every corner of the globe. The combination of these multiple leads in 1941 provoked Henry Luce, publisher of Time Magazine, to dub 1900–2000 as the 'American Century'.

Commentators from all across the ideological spectrum have searched for historical analogies to try to explain the concentration of power in the US after the end of the Cold War, and to predict its continuance or decline. Niall Ferguson, the British historian, as well as Joseph Nye, Immanuel Wallerstein, and of course, George Modelski, have all compared the US rise to hyperpower status to that of the United Kingdom in the nineteenth century. But, by almost any measure, the US concentration of power after 1989, consequent to the (temporary) disappearance of the Soviet superpower had no historical comparison. The US was no longer a superpower, one of many on the world stage, and new terms had to be employed to describe such a concentration. A hegemon in world affairs is usually defined as a state with overwhelming dominance in military and economic power that precludes effective challenge from a state or a combination of states. From the end of the Cold War to the present, American superpower became hegemonic. Hegemony suggests both a hyperconcentration of power and no evident contender on the world stage. It also has an association with aggressive or expansionist policies that are designed to ensure the continued dominance of the hyperpower. For Wallerstein, however, the US hegemony ended by the early 1970s as it was challenged both on the political–military front from the Soviet Union and on the economic front by Japan, the European Union, and later China.

Niall Ferguson has recently promoted the idea of the US as a reluctant empire by showing that it shares many characteristics with Britain in the nineteenth century. The key difference between the two states is that while Britain openly proclaimed its imperial status and role, including its 'civilizing mission' for territories in Africa and Asia, no American political leader would see imperial power as a worthy aim or imperial control as a political end. In fact, the political rhetoric of American foreign policy is wrapped up with the concepts of liberty, freedom, democracy, entrepreneurship, and enterprise, none of which figured prominently in the kind of imperial control pursued by the great European power before World War I. But Ferguson argues that imperial control did not necessarily mean a complete lack of political expression in the empire, and that US actions in Iraq and Afghanistan, including the installation of puppet regimes bolstered by an American military presence, does not differ much from the British model. Contrast this to the statements quoted by Niall Ferguson; Sandy Berger, President Clinton's national security adviser, declared in 1999 that the United States is the "first global power in history that is not an imperial power." In 2000, then candidate George W. Bush echoed his words, arguing, "America has never been an empire. … We may be the only great power in history that had the chance, and refused." Adding the component of 'preventive war' to its vast military reach, Johnson takes the opposing view that "Like other empires of the past century, the United States has chosen to live not prudently, in peace and prosperity, but as a massive military power athwart an angry, resistant globe."

Current US leadership is still predicated on 'hard power' (the ability to command) and measured by military indicators, that in turn rely on a large and powerful economy. In a realist perspective on world affairs, where putative enemies abound and no permanent rules control the depredations of states, only hard power can guarantee the defense of the state and the projection of power abroad. Combined with a judicious alliance strategy, large powerful states can bend smaller, weaker states to their will. In a world of competition and uncertainty, only hard power matters in the realist perspective.

In contrast to realism is the liberal perspective, derived from the idealist notions of building world institutions to dampen competition and to uphold values and norms that are universal, such as democracy and liberty. For superpowers, and especially hegemonic states in a liberal world order, 'soft power' (the ability to persuade) is paramount. Joseph Nye has been advocating a soft power strategy for the US since the end of the Cold War, both as a device to reduce that country's military spending and as a more effective long term strategy to structure American security. Key to the soft power ideal is the use of American leadership to encourage global agreements on major challenges that threaten the security of all – climate change, terrorism, drugs and crime, financial instability, and economic recession. Nye believes that traditional American ideological and political values have broad appeal and rather than instituting them via forceful means (removing a dictator and installing a pro American leader), which generates a hostile opposition, the US is more successful by working with like minded states to manufacture stable global institutions that enshrine US principles. The UN and its various agencies are paramount in the success of this strategy as is a change in attitude in the Washington, D.C. political establishment (and for part of the American public) that values cooperation and compromise rather than pursuit of purely American selfish interests. But as Gary Wills has noted, the American rhetoric of human idealistic achievements promulgated by the self described 'leader of the free world' has all too often degenerated into quite the opposite policies – especially during the Cold War when regional expediencies frequently overtook hyped up claims that supposedly supported human rights and democracy.

After about two decades of overwhelming American hyperpower, it is now clear that a belated response is under way among other major powers. Observers are noticing the rise or return to global prominence of Russia, China, and the European Union. In the case of Russia, the second term of President Vladimir Putin has been marked by a reassertion of traditional Russian geopolitical interests in the territories on, and adjoining, the former Soviet Union. With a resource economy of gas and oil exports thriving in a time of high commodity prices, Russia is rebuilding the strength of its armed forces on a scale not seen for a quarter century. For a brief period after the terrorist attacks of 11 September 2001, it seemed as if the geopolitical interests of Russia and the United States were coincident on the Eurasian landmass with a joint fear of Islamist radicalism emanating from its southern rim (Afghanistan, Iran, and Arabian peninsula). But the US invasion of Iraq showed the long standing differences in strategic goals while the presence of American forces on the territories of the former Soviet Union (Kyrgyzstan, Georgia, Azerbaijan, and up to 2005 in Uzbekistan) and close political relations with Washington of many former members of the Warsaw Pact has increased Russian worries about its encirclement by a belt of hostile states.

The Shanghai Cooperation Council (SCO – founded in 1996 and incorporating Russia, China, Kyrgyzstan, Kazakhstan, Uzbekistan, and Tajikistan) has a broad range of aims from individual state security to regional stability in Central Asia (meaning a check on militant Islam) and economic cooperation. However, joint military exercises in 2005 between Russia and China amid common language about the dangers of 'hegemonism' in world affairs have definitely added to the impression that a new 'Great Game' between rival superpower alliances (SCO and the US led one) is underway in Central Asia. The 1990s have seen Chinese economic growth rates averaging 10% and continued economic growth points to projections of a gross domestic product larger than America's by 2020. This huge spurt has generated an enormous demand for energy imports and China is now engaged in signing agreements with energy exporters from Central Asia to central and southern Africa, especially Sudan. These agreements with regimes that have traditionally linked their economies to Europe and the United States is setting off a scramble for long term energy security through locking inflows to the huge energy consumers, China, Japan, the US, and the EU.

Two other potential superpowers, Japan and the European Union, also deserve mention. In the 1980s, Japan's economic prosperity and foreign area investments through massive trade and savings surpluses worried the US that this former enemy was in a position to replace the US as the world's economic superpower. But by 1990, this fear had vanished as the Japanese economy imploded after the collapse of the 'financial bubble', a situation that is still not completely reversed almost two decades later. By contrast, economic growth in the European states was slow and, at times, stagnant, the so called 'eurosclerosis'. As the political project advanced with the accession of 21 states over the next half century to the original six members of 1957, Europe's global political weight has not been commensurate with its cumulative gigantic economic weight, well in excess of the US and China. Much of this lack of presence is due to differences between the larger European states in aligning with the US (the

United Kingdom is the most reliable ally and most of the new (2004–07) members have also positioned themselves as enthusiastic members of NATO, which they joined shortly before EU membership). Efforts to deepen the European institutional structures to advance a common foreign policy and bring highly varied national domestic policies into line have been rebuffed by the European publics. Until the 'democratic deficit' (the gap between elite and public interests) is resolved, the projections for Europe as a superpower are undermined.

Conclusions

Any indications that the US is going to allow its super power lead to slip, at least on the military side, are easily dispelled by the latest *National Security Strategy of the United States* document from 2006. The central premise of this document which lays out the framework for American policy in the immediate future is that Ameri can "forces will be strong enough to dissuade potential adversaries from pursuing a military build up in hopes of surpassing, or equaling, the power of the United States." In order to accomplish this, the document states that the US military must: "a) assure our allies and friends; b) dissuade future military competition; c) deter threats against U.S. interests, allies, and friends; and d) decisively defeat any adversary if deterrence fails." Clearly, this is not a statement of disengagement from world affairs and a retreat to something like "Fortress America." Instead, the document is firmly focused on the potential for fur ther attacks on US soil and its citizens from international terrorists and argues strenuously for continued American military bases and involvement in all parts of the globe. Such a global vision is part of what Luiza Bialasiewicz and colleagues have called the 'performance of security', whereby imaginative geographies and popular geopoli tical representations of the US and its enemies are combined to produce the effects that they name. The end product is an imaginary geography that mirrors and bolsters support for the US led 'war on terror'.

One of the memorable Yogi Berra (a baseball player famous for his mangled colloquialisms) quotes encapsu lates the problem of trying to anticipate the development of superpower rivalries – "The future ain't what it used to be" – suggests that the era of competition has to end because an all out military confrontation would result in destruction of all sides. The mutual assured destruction (MAD) arrangement of the Cold War lasted four decades and while preventing direct military attacks, it did not dampen proxy wars or competitions in other spheres (economic, political, alliance building, sports, and sci ence). Thomas Barnett, as advisor to the Pentagon and popularizer of a new American geopolitics, has taken a pessimistic perspective on the global future, a dystopia filled with threats to American interests and values from much of the world's poor periphery, an area that he calls 'the gap'. US strategists with this perspective have to 'mind the gap' and be especially cognizant of the kinds of social and economic changes in these states that lead to the entrenchment of movements inimical to US interests. If this perspective is adopted wholesale, the seeds for another round of superpower rivalries with local and regional proxies is already *in situ*.

See also: Cold War; Critical Geopolitics; Empire; Geopolitics; War; World-System.

Further Reading

Barnett, T. P. (2004). *The Pentagon's New Map: War and Peace in the Twenty First Century*. New York: G. P. Putnam's Sons.

Bialasiewicz, L., Campbell, D., Elden, S., Graham, S., Jeffrey, A. and Williams, A. J. (2007). Performing security: The imaginative geographies of current US strategy. *Political Geography* 26, 405 422.

Chase Dunn, C. (1989). *Global Formation: Structures of the World Economy*. Oxford: Basil Blackwell.

Ferguson, N. (2003). Hegemony or empire. *Foreign Affairs* 82, September October, 154 161.

Flint, C. and Taylor, P. J. (2007). *Political Geography: World System, Nation State and Locality* (5th edn.). New York: Pearson.

Goldstein, J. (1991). Reciprocity in superpower relations. *International Studies Quarterly* 35, 195 209.

Johnson, C. and Carroll, J. (2004). *The Sorrows of Empire: Militarism, Secrecy, and the End of the Republic*. New York: Henry Holt.

Kennedy, P. (1987). *The Rise and Fall of Great Powers: Economic Change and Military Conflict from 1500 to 2000*. New York: Random House.

Keohane, R. O. (1984). *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton, NJ: Princeton University Press.

Levy, J. S. (1987). Declining power and the preventive motivation for war. *World Politics* 40, 82 107.

Modelski, G. (1987). *Long Cycles in World Politics*. Seattle, WA: University of Washington Press.

Modelski, G. and Thompson, W. (1988). *Seapower in Global Politics, 1494 1993*. Seattle, WA: University of Washington Press.

National Security Council (2006). *National Security Strategy of the United States 2006*. Washington, DC: Office of the President of the United States.

Nijman, J. (1993). *The Geopolitics of Power and Conflict: Superpowers in the International System, 1945 1992*. London: Belhaven Press.

Nye, J. S. (2005). *The Paradox of American Power: Why the World's Only Superpower Can't Go It Alone*. New York: Oxford University Press.

Paul, D. (2000). *The Superpowers: A Short History*. London: Routledge.

Wallerstein, I. (2003). *The Decline of American Power: The United States in a Chaotic World*. New York: New Press.

Wills, G. (1999). Bully of the free world. *Foreign Affairs* 78, March April, 50 59.

Surrealism/Surrealist Geographies

D. Pinder, Queen Mary, University of London, London, UK

Glossary

Avant-Garde Derived from a military term for a small force sent in advance of an army, it has been used since the nineteenth century to refer to a group of artists at the forefront of developments, opposed to traditional practices and usually seeking to change not only art's techniques and forms but also – in the case of the, surrealists, among others – its institutions and the wider world.

Dreams Of fascination for surrealists for their poetic, fractured, and spontaneously associational qualities, dreams were explored by them in terms influenced by Freud as symbolic of hidden desires and fears, but more for their productive imagination in relation to reality than for their therapeutic meaning.

Found Object Exerting a strange attraction or psychic charge for the discoverer, it often takes the form of an old, discarded, or seemingly useless thing chanced across at a flea market or other urban location; for the surrealists, its charge relates to the ways in which it manifests unacknowledged or repressed desires, or in social, historical terms to its outmoded status, as standing for a now superseded mode of production.

Marvelous Aspired to by surrealists, it involves the eruption of wonder and revelation within the everyday and ordinary.

Objective Chance An understanding of chance as a positive rather than neutral force and as determined by both 'exterior necessity', in the form of objective processes in the world, and also by forces of desire in relation to the interior psyche.

Uncanny Refers to a feeling of dread, horror, disturbance, or strangeness generated through the co-presence or intermingling of the familiar and unfamiliar; a theme central to much surrealist art and activity, it was understood by them in Freudian terms as connected to the return of materials than had been repressed, whereby feelings of the uncanny (*unheimlich*) may suddenly erupt within even in the most apparently familiar or homely (*heimlich*).

Introduction

Surrealism is a revolutionary movement that emphasizes imagination, desire, dreams, eroticism, and the unconscious in its activities and efforts to change the world. It is an attempt to find new ways of living that are free from the shackles of capitalism and bourgeois morality, and that subvert what its advocates see as the deadening influence of the Enlightenment rationalism and the dominant intellectual traditions of Western civilization. Originating as an avant garde group of artists, writers, and radicals in 1920s' Europe, with an epicenter in Paris but with groups emerging in many parts of the world, its multiple forms have had an extensive and widely acknowledged influence on modern culture. Yet, the term is frequently understood as a style of art or literature, with the focus falling on individuals and their works rather than on collective endeavor and practice. The word 'surreal' has meanwhile lost specificity through casual usage for a whole range of activities, including the most banally 'strange' or 'weird'. In the process, surrealism has become a respectable part of the modern art and literary canon, and has lost much of its early shock value. It has been turned into the stuff more of museums, prestigious art collections, and reading lists than of street revolt, and has been recuperated by consumer culture, with its images frequently used to sell commodities rather than undermine their fetishism.

In this context, it is important to stress the radical political as well as artistic intent behind surrealism. It is also necessary to underline surrealism's collectivism and its adventurous and experimental approach that brings together practices associated with many different fields of cultural activity, such as literature, essay writing, poetry, painting, photography, film, architecture, sculpture, and performance, while at the same time seeking to go beyond their institutionalization as individualistic modes of art. As an oppositional and collective avant garde, surrealism expresses insubordination to the existing state of things, to the routinization and bureaucratization of everyday life, and to the commodification of more of the social realm. It reaches toward the possibilities that lie within the everyday, seeking to overturn those forces that prevent their realization. Inspiration is taken both from Karl Marx's injunction to 'transform the world' and from Arthur Rimbaud's call to 'change life'; as noted by André Breton, the movement's leading figure until his death in 1966, for the surrealists those watchwords were one.

Geographical issues have been important in much surrealist activity, which has often taken cities and their spaces as a focus for imaginative investigation and exploration. Many surrealist texts and artistic practices as well as performances, games, and political actions engage with everyday spaces and streets, with Paris a frequently favored locale. They challenge dominant codes of space and time through a variety of strategies, interventions, and poetic and transgressive acts. Surrealists have also addressed spatial issues through a long overlooked yet significant engagement with architectural theory and practice. Further, running through the surrealist movement are concerns with resisting nationalistic and imperialist geographical imaginations, and with supporting anti imperialist, anticolonial, and antifascist struggles.

Despite these interests, surrealism's influence within the discipline of geography has been largely subterranean and indirect. Influence has come in particular through intermediary and primarily French writers who have become key figures in spatial and social theory, and who were themselves influenced by or in critical dialog with surrealism. Among them are Gaston Bachelard, Roland Barthes, Jean Baudrillard, Walter Benjamin, Hélène Cixous, Michel de Certeau, Guy Debord, Michel Foucault, Julia Kristeva, and Henri Lefebvre. Also significant has been the so called 'dissident surrealism' of Georges Bataille. The importance of surrealism for these thinkers has typically been downplayed in their incorporation into human geography, however, perhaps due to the ways in which it sits uncomfortably with geography's more rationalist and social scientific traditions. Explicit discussions within geography of surrealist ideas and activities remain rare.

That said, connections and related concerns are apparent between surrealism and certain geographical work since the 1970s that has explored a poetics of place and space, and, in so doing, strained at the limits of disciplinary frameworks and conventions. Surrealism has also garnered increasing direct attention among geographers in recent years, as both a subject for research and also an inspiration for developing critical perspectives. Discussion of surrealist geographies has been opened up as part of a wider interrogation of the historical geographies of modernity and modernism, which has emphasized the multiple and contested nature of the twentieth century modernisms and the different geographical imaginations that inform them, and which has attended to the ways in which artists and writers investigated the modern by exploring its spatialities. As such, it has been accompanied by investigations of the geographical concerns of other modernist and avant garde groups, including the situationists. The radical activities of both historical and contemporary surrealists have further been drawn upon and discussed in relation to utopian visions of cities, to critiques of the spaces of everyday life, and to explorations of urban lived experience that have given prominence to issues of emotion, desire, the unconscious, and the uncanny. The next section introduces surrealism and its historical emergence before the following sections outline selected contours of surrealist geographies. They center, first, on surrealist concern with the spaces of everyday life, and with the possibilities of their transformation; and, second, on architecture and spatial construction. Brief remarks are then offered on surrealism's presence or absence within the discipline of geography, and on the potential significance of surrealist geographies.

Threads of Revolt

The immediate origins of surrealist movement lay in dada, an international avant garde initially based in Zürich. The dadaists themselves reacted to the carnage and inhumanities of World War I, and vigorously rejected cultural norms associated with nationalism, patriotism, militarism, capitalism, religion, and the family. Their performances, visual works, manifestoes, and other publications scorned such institutions and values, and favored spontaneity, chance, and a kind of creativity through destruction. They denigrated art in particular for its ideological role in papering over the barbarities at the heart of Western civilization, and turned artistic or anti artistic means, such as 'ready mades', spatial assemblages, cabaret performances, and linguistic pyrotechnics into weapons to shock and disrupt systems of meaning, and to serve the struggle for individual liberty. While the most overtly politicized of the dada manifestations developed in Berlin, where members eventually joined the German Communist Party, the dada group in Paris around Tristan Tzara contained a number of members who would soon break to establish the surrealist movement, among them Breton, Louis Aragon, Paul Eluard, and Philippe Soupault.

The term *surréalisme* was adopted in homage to the poet Guillaume Apollinaire, and Breton announced the new movement's presence in his first *Manifesto of Surrealism* of 1924. That year also saw the opening of the *Bureau de recherches surréalistes* and the founding of the journal *La Révolution surréaliste.* Most attention has focused on the tightly organized 'historic' surrealist group based around Breton in Paris during the 1920s and 1930s, about which an extensive literature has developed especially in art history and literature. Further commonly identified phases of surrealism in France are 1939–45, when the attention of members increasingly turned to issues of myth and the occult while Breton was in exile in the United States, and a number of members remained in occupied France; and 1946–69, when new members joined, staged exhibitions, and participated in political actions. Historians have often depicted the latter stages as ones of decline or empty repetition, if not dismissing them altogether, before the dissolution of the Paris group by Jean Schuster in 1969. Yet, such accounts have been challenged by those asserting the significance of postwar surrealists for radical art and politics in France, leading up to their participation in the revolts of May 1968, and for surrealism's global diffusion in the postwar period.

The contention of surrealism's 'end' has also been dis puted, with numerous groups developing surrealist ac tivity in more recent years not only in Paris but also in other cities worldwide, currently including Chicago, Prague, London, Leeds, Madrid, Athens, San Paulo, and Santiago. Surrealists have indeed often evoked the idea of a surrealist spirit or current that surpasses the formal movement and its periodizations, and that looks not only forward but also backward to celebrated precursors that include Apollinaire, Baudelaire, Fourier, Lautréamont, Nerval, Rimbaud, and Sade.

In his first manifesto, Breton addressed themes cen tral to surrealism. Notable among them were resistance to the 'realistic attitude' and to positivism, and the positive embrace of dreams and the imagination. Breton asserted that the latter "alone offers me some intimation of what 'can be'," and suggested that it "is perhaps on the verge of reasserting itself, of reclaiming its rights" (1972: 5, 10). Breton's interest in psychoanalysis and the writings of Sigmund Freud came through working as a medical orderly with shell shock patients during World War I, although his adoption of Freud's ideas never gained the approval of the man himself, just as his reading of Marx found little favor with the self pro claimed guardians of Marxist orthodoxy within the French Communist Party. Breton defined surrealism as "psychic automatism in its pure state" (1972: 26), and he and his colleagues sought to approach this state through automatic writing and through dreams, which they col lected and published. They also explored other ways of loosening social and psychological constraints and ha bitual influences through cultivating trance like states, startling juxtapositions, and games of free association as well as the pursuit of convulsive beauty and *amour fou* or mad love, so that artistic creation became a means of liberating instincts and of challenging reason and re pression through creative play.

Everyday life should be transfigured through the realization of the marvelous possibilities it contains here and now, according to the surrealists. Ordinary life pro vides the basis within which the extraordinary must be sought. To this end, surrealists questioned the separation between dream life and waking life, believing not that the first should subsume the second but that, like poetry and the everyday, they should suffuse one another in the manner of 'communicating vessels'. In his first manifesto, Breton looked toward the potential resolution of "these two states, dream and reality, which are seemingly so contradictory, into a kind of absolute reality, a *surreality*, if one may so speak" (1972: 14). The aim was therefore to deepen or plumb the real rather than to transcend it. Importance was given to sexuality and to what Freud referred to as Eros, or life force, although less as some thing that must be controlled for the maintenance of civilization as Freud averred, than in terms influenced by the Marquis de Sade as a drive or disruptive force that could be unleashed for radical ends.

Surrealism's commitment to revolution and to the be lief that art could serve revolutionary ends – signaled, for example, by the group changing the title of its review to *Surréalisme au service de la révolution* in 1929 – led to many attempts to forge alliances with left wing political organ izations, in particular, the French Communist Party. The surrealists' emphasis on the unconscious, on free creativity, and on their own independence made these efforts strained, however, and Breton opposed Stalinists in the party to ally himself with Trotsky. After World War II, the surrealists in Paris severed ties with the Communist Party and denounced the Soviet suppression of the Hungarian uprising in 1956, while continuing to take positions on numerous political issues, especially those involving re sistance to imperial and colonial powers, for example, in Algeria and in Vietnam. In recent years, surrealists have been drawn to anarchists and Marxists of an unorthodox persuasion in relation to a range of protests and struggles from opposition to capitalism and to the war in Iraq, to defending the rights of migrants and *sans papiers* (people without legal documents).

Surrealist Urban Dreams

Reconsidering the relationship between humans and the environment has long been important to surrealists. They have opposed the destructive and antagonistic stance toward nature they see as characteristic of Western ra tionalism and capitalist social relations, and have stressed the need to 're enchant' the natural world and to learn from different non Western knowledge systems. Such a displacement and recentering of worldviews is suggested by the group's map of the world, a work published in a special issue of the journal *Variétés* on surrealism, in Brussels in 1929 (see **Figure 1**). In a deliberately pro vocative move that mocks norms of geographical repre sentation, the anonymous author of the map expands certain land masses enormously – notably Easter Island, Mexico, Alaska, the west coast of Canada, and Africa – while diminishing or erasing others altogether, such as the United States and much of Europe. Surrealist geog raphy in this way connects with anti imperialist and anti colonial politics, while at the same time valorizing the so called 'exotic' and 'primitive' in relation to creativity, something that is also evident from many of their own practices of collecting art works and objects. Their interest in analogical methods, metamorphoses, and al chemical transformations as part of a poetic engagement with matter further prefigures aspects of radical eco logical and geographical thinking that opposes a Car tesian separation between the material world and the mind, as well as rigid distinctions between human and

Figure 1 The world at the time of the surrealists. Source: Variétés (June 1929), pp 26 27.

nonhuman, and attends to the liveliness of material things and nature.

Surrealism has also been rooted in and engaged critically with urban environments. Walter Benjamin remarked in his essay on the surrealists in 1929 that "the most dreamed about of their objects" was "the city of Paris itself" (1999: 211). Arcades, cafés, hotels, monuments, parks, and above all, streets played a significant role in the group's activities, and its explorations of Paris have been influential in shaping understandings of surrealist geography and in inspiring much creative urban exploration since. Breton wrote of his passion for city streets and for wandering, describing the street as his true element. It was in the streets, he believed that one can become open to chance encounters and coincidences, to the flow of events, to signs and messages that speak of hidden desires, and the mysterious qualities of modern urban experience. It is there that the hold of habit and routinized behaviors might be loosened, the boundaries between interior and exterior interrupted, and the possible revealed as one enters into charged relationships with other people, places, and objects and learns to perceive reality afresh with renewed vision.

Classic prose works of surrealist Paris include Breton's *Nadja* and *L'Amour fou* (*Mad Love*), and Aragon's *Le Paysan de Paris* (*Paris Peasant*). Based on the author's experiences in the city especially while walking its streets, often at night, where the practice itself becomes a poetic act comparable with other forms of creativity, the texts document both mundane and extraordinary events as they interweave accounts of places, encounters, and coincidences with inventories of objects and images, photographs, or records of what was there as well as flights of philosophical speculation. In the process they explore experience itself, identity and urban space, the last being not simply a physical container but the medium through which adventures and experiences unfold. Spaces are suffused with desire, at times appearing as a series of discrete sites in a personal geography of emotion and subjective experience, at other times in flux, as the city itself becomes dreamlike and fluid.

Urban spaces contain a forest of signs to be deciphered and felt by the surrealist wanderer, for whom they often spark emotional and apparently irrational associations or reveries. The trigger might be the juxtaposition of words on a street or shop sign, or a surprise encounter or a strange coincidence: these serve to release the marvelous from within the mundane, opening windows onto worlds beyond the brute empirical. Central to the eruption of these moments of revelation and wonder is the operation of what the surrealists termed 'objective chance'. That last phrase, taken by Breton from Friedrich Engels, emphasizes the materialist dimension of chance as a positive rather than neutral or negative force, by relating it to what Engels termed 'exterior necessity'. In this view, it is the objective qualities of reality itself that provide the conditions for – or convulse into – what might otherwise appear as an inexplicably coincidental event, encounter, or sign. At the same time, Breton emphasizes how the event, encounter, or sign also seems to arise from the subject's unconscious desire, to the extent that she or he appears to be impelled toward it or to somehow constitute it. The notion of objective chance thus cuts across simple divisions between an exterior objective world and an interior psychic realm, depending instead on the intersection of forces of determination related to each. Such moments of objective chance run through Breton's 'novel' *Nadja*, including his encounters with the woman of that name that she apparently and strangely foresees. They are also evident in Breton's discovery of enigmatic objects whose significance is often only discerned later, for example, the wooden slipper spoon that in *Mad Love* he recounts finding at the Marché aux Puces flea market in Paris, and that resonates both on a subjective level, representing his desire for love, and on a sociohistorical level, as an outmoded object created through peasant craft that now stands for a displaced mode of production and social formation.

As repositories for objects that were old, discarded, forgotten, and outmoded, flea markets were favored surrealist haunts. So too were other urban sites that were out of the way or marginal in relation to the monumental and spectacular city. For Breton, these sites included Les Halles, Place Dauphine, and Tour Saint Jacques in addition to less well defined 'poles of attraction' to which he found his feet drawn while out on strolls without necessarily knowing the reasons. The attraction was not simply physical but could come from a vague sense of expectation that something would happen there, or from the personal associations of place that, as with the meaning of dreams, might only become manifest later. Such was the case with Breton's draw to Place Dauphine, which he came to recognize in terms of its erotic connotations. But the charged quality of places also came from their historical associations, in terms of what happened in particular locations and the roles those places played in the revolutionary and bohemian pasts of the city. As discussed by Margaret Cohen, a ghostly topography can thus be mapped of surrealist Paris, featuring places whose pasts threaten to burst into the present during surrealist strolls.

It was Benjamin who, while critical of many surrealist claims, asserted the political significance of what he called the 'extraordinary discovery' by Breton of the revolutionary energies that appear in the outmoded, in "the first iron constructions, the first factory buildings, the earliest photos, objects that have begun to be extinct." Benjamin wrote that the surrealists "bring the immense forces of 'atmosphere' concealed in these things to the point of explosion" (1999: 210). His argument has been developed by, among others, the art historian Hal Foster who, from a psychoanalytic perspective, relates the outmoded to notions of the uncanny. The surrealists understood the latter in terms borrowed from Freud as referring to the return of materials that were previously hidden or repressed, and that have been made strange by that repression. It involves the eruption of the unfamiliar within the familiar, such as ordinary urban space or setting. Attending to the operation of the uncanny in this manner demonstrates its potential as a disruptive force as part of surrealism's efforts "to win energies of intoxication for the revolution" and to spark moments of "profane illumination," as Benjamin put it (1999: 215). But for Foster it also directs attention to surrealism's 'dark side', which necessitates looking beyond the pleasure principle and the emphasis on love and liberation by surrealism's advocates to consider the role of trauma, the compulsion to repeat, and the death drive.

Architecture, Space, and Fantasy

As developed within the historic surrealist movement of the 1920s and 1930s, surrealist geography thus gave prominence to lived experiences, to encounters on the streets, and to the possibilities of the marvelous as well as the surge of the uncanny within ordinary urban spaces. Its utopianism centered on realizing the potential that lay within present sociospatial conditions rather than on projecting grand plans of ideal futures. The early surrealists nevertheless fantasized about how they might transform the spaces of Paris according to their desires, outlining 'some possibilities for the irrational embellishment of a city' in 1933. This collective inquiry or game contested the symbolic geographies of monuments and statues in the center of the city through proposals that were often driven by feelings of disgust or hate. Breton's contributions included detonating the Arc de Triomphe after "burying it in a mountain of manure," razing the Palace of Justice and letting "the site be covered by a magnificent graffiti to be seen from an airplane," and transforming the Opéra into "a fountain of perfumes" (1978: 95–6).

Surrealism also addressed space through its relationship with architectural theory and practice. The surrealist movement has contained few practicing architects, and on first brush its involvement with architecture seems limited. Certainly, it has been neglected within histories of modernist architecture as well as in accounts of surrealism itself. Until the recent appearance of a volume edited by Thomas Mical in 2005, the most extensive survey in English remained an issue of *Architectural Design* edited by Dalibor Vesely in 1978. Given surrealism's emphasis on the fluidity of fantasy, dream, and the unconscious, this might seem understandable: how could a surrealist room, building, let alone city not prove disappointing or even a travesty of revolutionary desire once represented and materialized in a spatial form? Much surrealist engagement with architecture has been critical and textual, involving discussions and polemics over architectural styles and traditions, especially in relation to modernism during the 1920s and 1930s. Thus, Breton, Tzara, and others attacked what they saw as the technocratic and rationalist approaches of modernists, such as Le Corbusier, with Tzara calling for a sheltering, comforting, and maternal 'interuterine' architecture that looked for inspiration toward caves and other nonrectilinear forms. Salvador Dalí meanwhile wrote enthusiastically of the delirious qualities of *art nouveau* architecture and its 'terrifying and edible beauty', al though interestingly both Tzara and Dalí chose to dwell in clean lined modernist interiors rather than in such spaces themselves.

The surrealists celebrated a number of 'found' struc tures that embodied alternative values and interests. Among the most treasured were buildings by Antoni Gaudí, praised for their dreamlike and fantastic qualities, and the Palais Idéal that was built single handedly by the rural postman Ferdinand Cheval in the Drôme area of southeast France, according to his own untutored vision. Spatial constructions by surrealists are more rare. They include exhibition spaces by the likes of Salvador Dalí, Marcel Duchamp, and Man Ray that immersed partici pants in intense sensory experiences as well as stage sets, film sets, and, in Dalí's case, commercial designs for de partment store window displays (not for nothing did Breton castigate his commercial interests through the nickname 'Avida Dollars'). But architecture's more gen eral significance for the surrealists has been as a site in which the psyche, fantasy, dream, and space intersect and are explored. The often haunting and psychically charged nature of such investigations are apparent in the interiors depicted in texts by Leonora Carrington and collages by Max Ernst, in architectural proposals by Roberto Matta Echaurren and Frederick Keisler, and in the often fan tastical reimagining of architectural and urban spaces in the writings of Aragon, Breton, and other surrealists through to the present.

Surrealism was therefore embroiled in a complex critical relationship with modernist architecture and ur banism. Surrealism can certainly be viewed as a significant counterblast to the contemporaneous modernist calls to order promoted by Le Corbusier and other modernist urbanists in early twentieth century Europe. Yet, placing surrealism back alongside those modernist dreams of order also reveals shared concerns, tensions, dialog, and en tanglements, especially around the monstrous, the magical, the outmoded, and the uncanny. As David Pinder argues, these disturb orthodox views of the rationalism and sup posedly pure spaces of modernist urbanism, highlighting desires and anxieties through which those spaces were constituted, and the ways in which they remained haunted by elements that they could never expunge. Of particular interest in this regard are the writings of Bataille, who, as surrealism's self described 'old enemy from within', es tablished the dissident surrealist journal *Documents*, and who most vigorously assaulted categorical, hierarchical, and regulatory thinking as part of an anti architectural 'base materialism'. He conceived of the 'formless' as an operation that brings things down in the world and that interrupts efforts to impose spatial and social codes. The term has since been influential in identifying subordinate tendencies within esthetic modernism as well as in more recent art practice.

Engaging Surrealist Geographies

Surrealist geography continues to be of considerable interest, exerting an influence in numerous fields. As indicated above, however, surrealism has until recently received relatively scant attention within the discipline of geography. Something of its spirit, if not necessarily with accompanying references, may nevertheless be discerned in certain poetic and experimental excursions in geog raphy since the 1970s. The most direct example involves the remarkable writings of Gunnar Olsson, who, since the middle of that decade, has drawn inspiration from modernist and avant garde figures that include Breton, Aragon, and especially Duchamp. Olsson once described himself as being among "the late surrealists," and his struggles with the prison house of language and his "critique of cartographic reason" reveal a long standing interest in surrealist writings and art works that informs his approach to the study and writing of geography. His references to surrealism in the later essays of his *Birds in Egg/Eggs in Bird* came as he wrestled with positivist and spatial science approaches, which he portrayed as far from innocent or neutral, arguing that they assault am biguity and imprison meaning, helping to perpetuate a "descriptive/truth functional/thingified language" that serves to bolster the *status quo*. He contrasted these ap proaches with that of the poet, quoting with approval Octavio Paz on how "words behave like capricious and autonomous beings" (Olsson, 1980: 10e). Noting that the poet's aim is not to contain that capriciousness but to liberate it, not to eradicate ambiguity but to appreciate and create from it, he gave particular significance to surrealist efforts to transcend the limits of their language and by extension the limits of their world.

Dwelling within the prison house of language remains unavoidable, Olsson acknowledged, but he argued that this makes all the more compelling the surrealists' at tempts to assault that language, to create new modes of poetry and syntax through words and materials at hand, and in the process to reach toward another way of life and a fuller reality. Hence, he advocated the continuing vi tality of surrealism, including for social science, which he suggested ought to become "as surreal as the reality it pretends to portray" (1980: 17e). Much of his work since then has sought to chart new dialectical approaches to social science and a "heretic cartography" that might, if not shatter the ceiling of language, at least raise it higher. An almost continuous intellectual companion on this journey has been Duchamp, whom Olsson describes as a "kindred spirit" who "devoted his life to the exploration of the unknown territory of the in between, a reconciler of opposites" (2007: 368). Through extended readings of Duchamp's works alongside those of other modernists, Olsson seeks to challenge norms and conventions of social scientific language and to open up other ways of writing and conducting geography. His texts have not been uncontroversial, with some criticizing their obscuritanism as well as their gender politics. Yet, as a form of modernist experimentation in the writing and doing of geography, they have carved a highly distinctive and influential path.

Surrealism has at times also come into focus as a subject of study within human geography and has been referenced in the context of debates on geography and literature, with Brian Robinson, for example, addressing the challenge it poses to realist–humanist approaches that long held sway in geographical writings in that field. Recently, however, Jill Fenton has argued most explicitly and fully for the continuing relevance of surrealist perspectives for critical geographical studies and activism. In essays based on ethnographic research with contemporary surrealists in Paris, she discusses their *géographie passionnelle* and the ways in which they rework ideas and practices in the context of the changing city and associated political struggles. In the process, surrealism is revealed as a significant subject for historical and contemporary geographical research as well as a radical perspective that can inform critical approaches to geographical issues, within and beyond the discipline, for example, among those concerned more widely with spatial politics, utopianism, and imagining alternative urban futures.

Any contemporary confrontation with surrealism comes in the wake of the considerable scrutiny that this movement has been subjected to over the years. Sources of critique have included other movements such as the situationists, who during the 1950s and 1960s sought to build on the constructive possibilities of the early surrealist movement, while attacking many of its positions and practices. The situationists opposed in particular the surrealists' reliance on chance and what they saw as their reactionary slide into mysticism, magic, and the occult after World War II. Feminists have also critically considered the position of women and the feminine in both surrealist works and the movement itself. Many have addressed the ways in which male surrealists have objectified women and the female body, and have portrayed women as mysterious, intuitive, and enlightening muses, hence as holding a key to renewed perception of space and place. More positively, some have also explored the oft neglected contributions made by female surrealists who, at least in comparison to their counterparts in other twentieth century avant gardes, have been unusually numerous. Others have traced the importance of surrealism for the later experimental writing practices of French feminists during the 1970s.

The theoretical and political climate within geography and other social sciences and humanities in recent decades, especially in Britain and North America, has proved more receptive to some aspects of surrealism than others. The writings of Bataille, who has been a major influence on much post structuralist theory, and who clashed with Breton whose work was more tied to dialectical thought and Hegel, have gained ascendancy and have helped to refigure understandings of surrealism itself. The repercussions of Bataille's challenging work on excess, expenditure, eroticism, and transgression will continue to be felt and reflected upon in much future work in geography and related spatial studies. Yet, also demanding and beginning to receive more attention are the varied surrealist geographies associated with Breton, Aragon, and others from early twentieth century Paris through to present practitioners. Such surrealist geographies have significantly influenced critiques of everyday life and lived experience that similarly take space seriously as part of a French tradition that includes Lefebvre, Debord, Certeau, and others. They have also informed and helped to inspire urban explorations and imaginative geographies of cities, for example, through interest in practices of walking. There they take their place in a long line of poetic urban wanderers, although their urban strolls differ markedly from the famed esthetic practices of the nineteenth century *flâneur*, being more active, engaged, and convulsive than the latter predominantly male figure; and also from the later *dérives* of the situationists, who in turn were more sceptical of chance and the unconscious, and more focused on analysis and collective political intervention.

Returning to surrealism and seeking a fuller appreciation of the complexities, diversities, and profound influences of this revolutionary movement has much to offer human geography. This is not only in relation to understandings of the historical geographies of modernity, modernism, and the avant gardes, and of the development of social and spatial theory, especially currents of Marxism and anarchism, but also in terms of current creative and critical thought and practice. Surrealist explorations of urban spaces and environments may be regarded as of particular interest at a time when some geographers and others are increasingly concerned with space as a realm of multiplicity, openness, and encounter with the unforeseen, and with investigating emotions, dreams, the magical, the uncanny, and the ghostly as intrinsic elements of what Steve Pile refers to as 'real cities' and the 'phantasmagorias of city life'. They may further repay attention from those wishing to reassess the potential of utopian geographical imaginations, both past and present, for relighting the fires of revolt and for seeking paths beyond the crushing weight of the *status quo*.

See also: Anarchism/Anarchist Geography; Avant-Garde/Avant-Garde Geographies; Cultural Politics; Psychoanalytic Theory/Psychoanalytic Geographies; Situationism/Situationist Geography; Utopian Cities.

Further Reading

Aragon, L. (1980). *Paris Peasant,* Watson Taylor, S. (trans.). London: Pan Books (originally published in 1926).

Benjamin, W. (1999). Surrealism: The last snapshot of the European intelligentsia. In Jennings, M., Eiland, H. and Smith, G. (eds.) *Walter Benjamin*: *Selected Writings*, *Volume 2*, *1927 1934*, Livingstone, R. *et al.* (trans.), pp 207 221. Cambridge, MA: The Belknap Press of Harvard University Press (originally published in 1929).

Bonnett, A. (1992). Art, ideology and everyday space: Subversive tendencies from Dada to postmodernism. *Environment and Planning D: Society and Space* 10, 69 86.

Breton, A. (1960). *Nadja,* Howard, R. (trans.). New York: Grove Weidenfeld (originally published in 1928).

Breton, A. (1972). *Manifestoes of Surrealism,* Seaver, R. and Lane, H. R. (trans.). Ann Arbor, MI: University of Michigan Press. (originally published in 1962).

Breto, A (1978). *What is Surrealism? Selected Writings*. New York: Pathfinder.

Cohen, M. (1993). *Profane Illumination: Walter Benjamin and the Paris of Surrealist Revolution*. Berkeley, CA: University of California Press.

Durozoi, G. (2002). *History of the Surrealist Movement,* Anderson, A. (trans.). Chicago, IL: University of Chicago Press.

Fenton, J. (2004). 'A world where action is the sister of dream': Surrealism and anti capitalism in contemporary Paris. *Antipode* 36(5), 942 962.

Fenton, J. (2005). Space, chance, time: Walking backwards through the hours on the left and right banks of Paris. *Cultural Geographies* 12, 412 428.

Fenton, J. (2006). Contemporary surrealist imaginary geographies of utopia and dystopia. In Emden, C., Keen, C. & Midgeley, D. (eds.) *Imagining the City, vol. 1: The Art of Urban Living*, pp 273 293. Oxford: Peter Lang.

Foster, H. (1993). *Compulsive Beauty*. Cambridge, MA: MIT Press.

Mical, T. (ed.) (2005). *Surrealism and Architecture*. London: Routledge.

Olsson, G. (1980). *Birds in Egg/Eggs in Bird*. London: Pion.

Olsson, G. (1991). *Lines of Power/Limits of Language*. Minneapolis, MN: University of Minnesota Press.

Olsson, G. (2007). *Abysmal: A Critique of Cartographic Reason*. Chicago, IL: University of Chicago, Press.

Pile, S. (2005). *Real Cities: Modernity, Space and the Phantasmagorias of City Life*. London: Sage.

Pinder, D. (2005). *Visions of the City: Utopianism, Power and Politics in Twentieth Century Urbanism*. Edinburgh: Edinburgh University Press.

Pinder, D. (2006). Urban encounters: Dérives from surrealism. In Adamowicz, E. (ed.) *Surrealism: Crossings/Frontiers*, pp 39 64. Oxford: Peter Lang.

Robinson, B. (1987). The geography of a crossroads: Modernism, surrealism, and geography. In Malloy, W. E. & Simpson, P. (eds.) *Geography and Literature: A Meeting of the Disciplines*, pp 185 198. Syracuse, NY: Syracuse University Press.

Vesely, D. (ed.) (1978). *Surrealism and architecture*. *Architectural Design* 2 3.

Surveillance

M. Henry, Massey University, Palmerston North, New Zealand

Glossary

Code Unique strings of numerals and letters generated by software programs.

Disciplinary Power Techniques of governance which accustom individuals to the normalizing effects of constant examination and surveillance.

Machine-Readable World Application of codes to all objects through the use of media able to be read and acted on by software.

Omnioptic Networks of surveillance based on the premise of the many watching the many.

Panopticon System of surveillance based on the isolation of individuals and their visibility to a central watcher.

Total Institution Institutions within which the lives of individuals are completely regulated and separated from the outside world.

Introduction

Our daily spaces seem to have become inexorably populated by surveillance devices which range from the cookies on our computers, closed circuit television (CCTV) in buildings and on our streets, and the proliferation of bureaucratic forms requiring personal disclosure from our insurance companies and their ilk. These surveillance devices range in technological sophistication from the mind boggling to the banal; and range in interest from the intensively individual to the extensive, aggregate acts of populations. Expressed in metaphors such as 'Big Brother', the dense ecology of surveillance seems to threaten cherished notions of privacy and individuality; but equally, such devices seem to provide security and coordination as we interact with our social worlds. In this milieu, geographers have started to explore the historical and contemporary logics and practices of surveillance, and in doing so have contributed to understanding surveillance as consisting of contingent, embedded practices rather than the totalitarian 'will to power' which informs much of the cultural imagination of surveillance.

However, what is surveillance? Surveillance can be defined as the organized monitoring of the activities of actors in order to produce personal data. As such surveillance is not practiced in isolation, but rather is linked to efforts to (re)order the conduct of actors. Surveillance is a feature of all societies, but has been most closely associated in the work of geographers with the rise of state and capitalist, bureaucratic apparatuses. The systematic collection of information about births, baptisms, marriage, and deaths has had a long ecclesiastical history, but as nascent nation states coalesced in Europe during the sixteenth to nineteenth centuries, a synergistic relationship developed between states' needs for more information in order to govern and states' capacity to produce information from their subjects. The increasing collection of individual information was initially tied to more intensive forms of state policing. Indeed the use of surveillance as a tool of intensive state policing has remained an overt feature of authoritarian and colonial regimes, as well as an implicit tactic in ostensibly liberal democracies. However, in the emergent liberal democracies, states' interest in their citizens gradually shifted from the maintenance of power implied in state policing to an interest in ensuring 'national' improvement and progress. Simultaneously animated by questions of control and improvement, contemporary states and commercial enterprises have become reliant upon the production, collection, storage, and coordination of personal data for their decision making.

The implications of the gradual embrace of actors within networks of governmental and corporate surveillance are ambiguous. In seeking to understand the practice of surveillance and its ambiguities, geographers have travelled along two paths: one path which has emphasized an understanding of the technical aspects of surveillance (especially in the intersection of geographic information systems (GIS) and surveillance); and a second path which has sought a critical understanding of the logic and grammar of surveillance practice. A characteristic of the second, critical path of surveillance studies in geography has been the effort to position the practice of surveillance within the networks of spatial control associated with modernity, and in this effort geographers have tapped into a deep tradition of social science research.

Surveillance and the Social Sciences

A common point of reference in understanding the connection between modernity and surveillance has been via the emergence of the new forms of industrial and administrative organization associated with modernity. The specific treatment of surveillance within the social sciences has been framed by the presuppositions of different intellectual traditions. Within Marxist inspired strands of research the creation of surveillance assemblages has been traditionally connected to the imperatives of capital accumulation. Marx, for instance, argued that in the capitalist factory the need to subordinate workers to their machines meant the emergence of a barracks discipline that involved the creation of a system of surveillance that divided workers into operatives and overlookers. It is the character of the conflict between these groups over the control of the workplace that has dominated Marxist inspired considerations of surveillance. From Comte and Durkheim the practices of bureaucratic surveillance were understood as artefacts of industrial culture. Within this tradition, surveillance practices were seen to be pursued because they represented technical solutions to the problems of administering complex, differentiated societies and territories. In extending this strand Weber highlighted the rationalization of social life as an integral feature of modernity, and argued that the spread of bureaucratic forms of administration was an effect of these broader processes of rationalization, and the technical efficiency of bureaucracies vis à vis other means of administrating large scale societies.

It is this second strand which has arguably been the most influential entry point for geography's specific interest in surveillance. Drawing on Weber, bureaucratic administration assumes a certain stance toward rationality. It bases decision making on the collection, storage, and distribution of knowledge, and it frames the exercise of decision making within a disciplinary matrix characterized by hierarchical authority and an adherence to rule governed conduct. The effect of this mixture of knowledge and discipline has been to create an administrative logic (both state and capitalist) which increasingly seeks to obtain finer grained levels of calculability through the collection of individual information. These themes of rationality, administration, and calculation find deep echoes in Foucault's work on disciplinary power, and it has been this work which has provided the immediate frame for geographers' critical understanding of surveillance. Foucault's attraction for geographers (particularly his work in *Discipline and Punish*) has been the resolutely spatial way in which he framed questions of surveillance.

Panoptic Visions

According to Foucault, disciplinary power revolves around the accumulation of information about individuals, the use of that information to govern the conduct of individuals, and the acceptance of that personalized governance by individuals. In governing individuals, disciplinary power transforms space from a medium of activity into a calculative technology in its own right. To illustrate the spatiality of disciplinary power Foucault drew on Jeremy Bentham's (1748–1832) model for a panopticon as the archetypical expression of a surveillance technology. Designed around a ring of cells open to a centralized viewing point, the panopticon simultaneously ensured the total visibility of its inmates and the unknowability of the surveillance enacted over them. In the uncertainty between visibility and unknowability, the panopticon was intended to inculcate a ceaseless self discipline amongst its inmates as they refashioned themselves according to the moral norms transmitted via punishment and reward from the unknown judge.

The figure of the panopticon has been deeply influential as geographers have sought to uncover and understand the power relations embedded in the architecture and operation of institutional spaces such as prisons, schools, and factories. However, geographers have qualified their understanding of the intersection of surveillance and panoptic logics. This has involved a shift in the debate away from regarding the panopticon as an actual surveillance technology. One effect of this shift from the literal to the metaphorical panoption has been the development of more nuanced understandings of the spatiality of actually existing surveillance practices. A feature of which has been a movement in attention away from total institutions and their sequestered inhabitants toward an attempt to reveal the pervasive embedding of surveillance into the daily routines of 'normal' individuals.

While the development of increasingly intensive surveillance practices has been closely linked to the emergence of European modernity, it has been persuasively argued that it is in Europe's colonial regimes (both historically and in neocolonial regimes) that surveillance practices have been most thoroughly developed and extended into the daily lives of colonial subjects. Here geographers have revealed the extensive surveillant assemblages that linked together the minute ordering of individual spaces with the articulation of a colonial power aligned along fundamentally uneven axes of race, class, and gender, and designed to maintain these relationships. In these myriad colonial spaces, geographers have shown that the practice of surveillance differed markedly from the enclosed spaces described by Foucault. In particular the mobility of colonial populations (especially indigenous peoples, but not confined to them) resulted in the development of surveillance techniques centered upon the creation of spaces of temporary immobility within which the rituals of identification were a necessary adjunct to the movement of individuals. In tracing the use of spatial controls embedded in places such as native reserves, the point has been demonstrated that the order created by surveillance and colonial bureaucracies was not simply a technical, neutral abstraction, but rather was intimately shaped by logics of ordering which were irreducibly framed by racism and economic exploitation.

CCTV has been often cited as an example of the creeping pervasiveness of panoptic logics and disciplinary power into the spaces of everyday life. Here, urban geographers in particular have traced the explosive rise in the use of these electronic technologies as tools of urban ordering. Notwithstanding the disquiet of geographers, CCTV systems are now a ubiquitous, indeed banal, feature of urban public places, private workplaces, shops, and increasingly rural spaces. In understanding the implications of CCTV on public spaces, geographers have argued that CCTV itself is only one manifestation of broader processes of (re)ordering. In particular the deployment of CCTV systems has occurred in contexts marked by concerns over the existence of city spaces (especially inner city spaces) as economic spaces. In promising improved security CCTV has been touted as a necessary adjunct to the economic revitalization of urban spaces. However, questions over the value and fairness of CCTV systems have been subsumed under the assumption of effectiveness supported by the widespread publicity given to CCTV's forensic successes. Despite such success, the research increasingly indicates there are significant questions to be asked regarding the unevenness of CCTV surveillance, and in particular the judgments made by CCTV controllers as they mobilize agents of ordering such as the police and security guards in response to the images on their TV screens.

The surveillance of the spatially mobile, which has been a feature of colonial surveillance regimes and CCTV mediated urban landscapes is continually marked by the slippage between a will to watch and the practical ability to watch. In bridging this gap geographers have pointed to the deployment of surveillance techniques beyond those which rely on the direct visibility of individuals. It is not that an intensive concern with individuals has been eliminated, but that the means of addressing their visuality has been linked with an extensive concern with populations. Cartography and statistics have provided the paradigmatic technologies in the production of these broader visions. Geographers have become particularly sensitive to cartography's entanglement with surveillance and power (especially in colonial contexts), and increasingly more sensitive to the use of statistics as a means of surveillance. In particular, contemporary surveillance practices which draw on statistical tools have been amplified by the addition of actuarial logics to the classic techniques of survey and census. Actuarial logics draw on the vast databases of information collected by state and nonstate agencies to predict the actions of individuals, where those actions remain opaque to forms of direct surveillance. In the use of actuarial logics we can discern a deeper shift in surveillance practices from the visuality of certainty in which the immediate role of surveillance is to categorically link individuals to spaces and actions; toward a visuality of risk in which individuals, spaces, and actions are linked together in networks of possibility framed by the management of risk.

In tracing surveillance practices geographers have shown that literal accounts of the panopticon have been flawed in their failure to appreciate the limits imposed on surveillance practice by issues such as scale and technological capacity. For example, as the scale of the space and population to be surveilled increases so do the difficulties in ensuring the total visibility of panoptic logic. Indeed, even at the architectural scale envisioned in Bentham's panopticon, and actually practiced in total institutions such as prisons, the logic of total visibility has provided to be an aspiration rather than a reality. Moreover, Foucault's account of the panopticon was predicated upon the spatial enclosure (and indeed isolation) of individuals. As indicated in the growing use of CCTV, and in colonial surveillance regimes, most individuals do not lead lives marked by confinement: rather the lives of these individuals are characterized by degrees of mobility which in turn creates moments of anonymity. It is a qualified anonymity which marks the experience of surveillance for most people, and under these conditions the metaphor of the panopticon as used by geographers has become problematized because of the image it conjures up of a centralized, omniscient gaze. Rather than being centralized in tall watchtowers or vast computer databases the information collected by contemporary surveillance systems is generally stored in small, decentralized repositories, and used not by a central authority (Orwell's 'Big Brother') but rather by small (relatively), dispersed governmental systems. Thus, in contrast to the simple image of the prisoner subjected to the intensive gaze of the unseen jailer, the experience of 'normal' individuals is of an extensive embrace with myriad surveillance systems each of which has its own calculative rationalities and desire to order. The result is both a fragmented visibility assembled from a wide, and often disparate, number of different surveillance systems, and an equally fragmented landscape of control enacted over individuals. In tracing this fragmentation, geographers' understandings of the limitations of surveillance practice have provided a prophylactic against some of the dystopian fears associated with the growth of surveillance technologies.

Omnioptic Visions

Notwithstanding the recognition of the fragmentation of actuality existing surveillance, panoptic inspired understandings of surveillance still rely at some level on a categorical, hierarchical distinction between watcher and watched. A further strand of surveillance studies in geography has counterpointed a panoptic distinction between watcher and watched with a sense of the omnioptic character of much surveillance. Rather than being based on the logic of the one viewing the many, omnioptic surveillance is based on the many viewing the many. Perhaps the paradigmatic articulation of such omnioptic logic comes from Jane Jacobs' seminal *The Death and Life of Great American Cities*. Here Jacobs argues that safe city streets are places that are marked by the presence of people and a pervasive surveillance engendered by the presence of everyday 'eyes on the street'. Work in this vein has pointed to the enactment of surveillance through dense webs of demotic sociability, rather than through the imposition of technoarchitectural programs.

Geographers have recognized that these omnioptic logics are deeply embedded in governmental spaces. In part this recognition has dovetailed with an understanding of the link between power and vision found in institutions of enclosure such as prisons, asylums, and factories. Alongside these institutions of enclosure geographers have paid attention to the emergence of public institutions centered upon ensuring both the visibility of individuals and the world. Exemplified in institutions such as the museum, the library, art gallery, gym, and zoo, a doubled gaze has been argued to operate which positions both objects and individuals in a visual relationship. On the one hand objects are systematically, but selectively, framed for the viewer. In viewing framed objects, the individual patron is placing it under surveillance. Conversely, on the other hand the act of viewing in such institutions is in itself a public performance in which the conduct of the viewing patron is simultaneously framed by the gaze of fellow patrons. In this doubled gaze in which individuals are simultaneously the object of surveillance and a surveilling subject there can be seen operating omnioptic logics of dispersed, demotic surveillance.

Such omnioptic visions are not confined to public institutions but can be increasingly discerned in the explosion of public archives of images, videos, and discussion in blogs and on websites such as 'YouTube'. Such collections have been enabled by the proliferation of the same digital technologies that have become synonymous with contemporary panoptic visions. Traced back to the emblematic images of the Rodney King beating in Los Angeles in early 1991, the countersurveillance enacted through such technologies has been argued to provide a check on the abuse of bureaucratic and media power, and increasingly a means of spreading a popular calculative matrix over the mundane services and products which individuals and groups use. It is an irony that notwithstanding the dystopian imagery which accompanies surveillance in the cultural imaginations of many people and the fact that whilst we have all become the targets of surveillance, many of us have also become enthusiastic agents of surveillance.

Automatic Surveillance

Whilst surveillance has been characterized by fragmentation and limitation, the ongoing story is of the increasing connection of different surveillance networks. This has been made possible through the electronic mediation of contemporary life in which individuals' activities increasingly require the use of electronic credentials, and the consequent storage of the electronic traces that are generated through transactions and encounters. Geographers have recently started to examine the ways in which the intricate spaces of social life have become subject to an automatic surveillance conducted through the intersection of individuals and a machine readable world. Automatic surveillance has been enabled through the creation of a dense ecology of codes – serial numbers, barcodes, id numbers, tracking numbers, and passwords – which are embedded in the objects we use, and which are required for them to be used. These codes do not simply enable a finer grained appreciation of the movement and use of specific objects, but have a profound, but largely hidden ordering role in shaping the spatiality of social life.

The embeddedness of codes in the technological infrastructure of daily life is linked to an enduring concern with classification and sorting. Nonetheless the capture and collation of individual information has shifted attention from the enactment of a direct gaze upon individuals towards the construction of evermore elaborate classification schemes in which our informational shadows perform as our proxies in complex simulations of social relations. Two processes can be seen amplifying the shift toward automatic surveillance. First, following September 11 (2001), and the subsequent (and potentially permanent) 'war on terror' there has been a rearmouring of travel spaces such as airports and train stations through the use of surveillance measures such as biometric passports, CCTV with facial recognition, and extended electronic eavesdropping of communications. Second, increasingly intersecting with the securitization of surveillance has been the commercialization of surveillance. In particular the rise of geodemographics, and the ability to both store and combine the consumption patterns of individuals recorded in myriad electronic transactions has enabled the production of a sheath of commercial opportunities (or risk) around individuals. In both these cases, individuals are increasingly enmeshed in a matrix of calculative possibilities constructed from the recombination of multiple sources of information. The power of such technologies is their ability to be used to sift and sort individuals in increasingly finer ways and to amplify the putative precision of management systems which attempt to expedite the flows of those lives and activities calculated to be riskless (in both a security and a commercial sense), and which aim to simultaneously hinder the flows of the risky.

Geographers have highlighted the sorting effects of the banal technologies which shadow our movements through space. At one level the use of these machine readable codes to monitor and sort objects and individuals ever more finely is simply a quantitative extension of those individualizing processes which have proliferated as a consequence of the state's interest in taxing individuals, mobilizing armies, and securing populations. At another level the practice of automatic surveillance suggests a qualitative change in the nature of judgment away from humans and toward nonhuman actors. In the codes and the centeres of calculation buried in lines of software, qualitatively different modalities of surveillance emerge in which individuals are only ever partially aware of the extent of their machine generated visuality, and the automatic decision making algorithms which are con stantly adjusting access to services and spaces.

Toward a Surveillance Assemblage

Foucault indicated that his understanding of power was productive in the sense that power not only constrained, but that it also enabled new subjects and spaces to be created. Geographers' accounts of surveillance have ten ded to focus on the former rather than the latter. However, we have increasingly come to view the practice of sur veillance as framed by the tension between control and freedom. Thus, on the one hand surveillance system render us and our actions visible and consequently amenable to the calculative grids of governments and businesses; whilst on the other hand those self same sys tems enable us the convenience of electronic banking or access to social welfare benefits. Yet, in identifying the ambiguity of surveillance practices we need to be con stantly mindful that the balance between control and freedom is both contingent and uneven. Thus, whilst surveillance can be enabling, it is not enabling for all. Whilst omnioptic surveillance can act as a check on the unrestrained power of governmental agents, the automatic, mundane surveillance of everyday life goes on invisibly enabling a subtle grid of possibilities to envelope and shape our life experiences. And unfortunately for some, the experience of surveillance is both direct and brutal.

In this world of multiple lines of surveillance, the traditional moments of hierarchical, architectural obser vation have been supplemented, and in many cases supplanted, by extensive networks of both omnioptic and automatic surveillance. Whilst older traditions of sur veillance were enacted upon the suspected object, the exercise of automatic surveillance positions every indi vidual, object, and space as potentially risky. No longer is surveillance enacted solely over society's outsiders, ra ther we are all suspects, if not for past events, then for present and future events. Surveillance conducted in this latter vein disappears into the very rhythm and fabric of life. Its practice, distinctions, and decisions have become both more dispersed and more connected, and both more intimate and more extensive. The importance of space has not lessened in this changing surveillance environ ment, but rather the questions have shifted from those of enclosure to those of control, permeable access, and the calculation of risk.

See also: Crime/Fear of Crime; Defensible Space; Foucauldianism; Gated Communities/Privatopias; Law and Law Enforcement; Public Space; Public Spaces, Urban; Regeneration to Renaissance; Visuality.

Further Reading

Adey, P. (2004). Surveillance at the airport: Surveilling mobility/mobilising surveillance. *Environment and Planning A* 36, 1365 1380.

Crush, J. (1994). Scripting the compound: Power and space in the South African mining industry. *Environment and Planning D: Society and Space* 12, 301 324.

Curry, M. (1998). *Digital places: Living with geographic information technologies*. London: Routledge.

Dandeker, C. (1990). *Surveillance, power and modernity: Bureaucracy and discipline from 1700 to the present day*. Cambridge: Polity Press.

Dodge, M. and Kitchin, R. (2005). Codes of life: Identification codes and the machine readable world. *Environment and Planning D: Society and Space* 23, 851 881.

Foucault, M. (1977). *Discipline and punish: The birth of the prison*. London: Penguin Books.

Graham, S. (1998). Spaces of surveillant simulation: New technologies, digital representations, and material geographies. *Environment and Planning D: Society and Space* 16, 483 504.

Graham, S. and Wood, D. (2003). Digitizing surveillance: Categorization, space, inequality. *Critical Social Policy* 23, 227 248.

Haggerty, K. and Ericson, R. (2000). The surveillant assemblage. *British Journal of Sociology* 51, 605 622.

Hannah, M. (1997). Imperfect panopticism: Envisioning the construction of normal lives. In Benko, G. & Strohmayer, U. (eds.) *Space and social theory: Interpreting modernity and postmodernity*, pp 344 359. Oxford: Blackwell Publishers.

Hannah, M. (1997). Space and the structuring of disciplinary power: An interpretive review. *Geografiska Annaler* 79B, 171 180.

Lyon, D. (1994). *The electronic eye: The rise of surveillance society*. Minneapolis: University of Minnesota Press.

Marx, G. (2002). What's new about the 'new surveillance'? Classifying for change and continuity. *Surveillance & Society* 1, 9 29.

Norris, C. and Armstrong, G. (1999). *The maximum surveillance society: The rise of CCTV*. Oxford: Berg.

Staples, W. (2000). *Everyday surveillance: Vigilance and visibility in postmodern life*. Lanham, MD: Rowman and Littlefield.

Relevant Websites

http://www.aclu.org
American Civil Liberties Union.

http://www.liberty human rights.org.uk
Liberty Human Rights.

http://www.anu.edu.au/people/Roger.Clarke/DV/
Roger Clarke.

http://www.statewatch.org
Statewatch.

http://www.surveillance and society.org
Surveillance and Society.

Surveying

U. F. Paleo, University of Extremadura, Caceres, Spain

Glossary

Control Network A set of well-spaced and interconnected points with very accurately known horizontal and vertical positions covering a large area.
Coordinate Value indicating the spatial position of an object or a point located in the Earth surface in a two- or three-dimensional space.
High-Definition Surveying Surveying completed with a laser scanner providing a cloud of points with coordinates of the target area.
Theodolite Instrument used to measure accurate horizontal and vertical angles when aiming at a certain distant point, relative to the position of measurement.
Triangulation Calculation of coordinates of points in a triangle using angle measurements, distances, and trigonometric functions.

History

Surveying has been a fundamental technique since the beginning of recorded history for purposes of measurement and division of the land into plots, to record land ownership, 'and facilitate taxation, using basic geometrical calculations'. Maps were important for the requirements of military operations, with a need for a good topographic coverage. The civilian maps lacked the detail required for moving troops, the construction of road and railroad networks connecting fortifications or planning campaigns, so that military pioneered the works in many geographical areas.

The exploration of newly discovered lands by European countries and emerging states led to an effort of ensuring territorial domination, expansion, and consolidation of national or continental borders. The search for natural resources and further claim to ensure exploitation was the motive for large scale efforts to survey, describe, and map large territories in Africa, Southern Asia, Australia, Russia, and North America in the nineteenth century, notably impulsing the scientific and technological advance of the discipline. Railroad development, driven by the need to open land trading routes, required the support of route surveying in construction.

Several paired scientific efforts were conducted to accurately measure the shape and dimensions of the Earth. In 1735, the French Academie Royal des Sciences commissioned two scientific expeditions to the equator – the Spanish–French Geodesic Mission – and Lapland to measure the length of an arc of 1° of a meridian and determine the roundness of the Earth. Another major surveying endeavor was the Great Trigonometric Survey of India. The survey began on 10 April 1802 with the measurement of a baseline near Chennai and a second baseline measured near Bangalore in 1804. Colonel William Lambton, in charge of the operations as the superintendent of the survey until his death in 1823, and succeeded by George Everest, proceeded firstly southward to Cape Comorin and then northward with the triangulation survey across the whole Indian subcontinent. The survey was completed in 1866 although its application as a support to large scale surveys was limited due to its prolonged duration. The most symbolic achievement was the measurement by Radhanath Sikdar of the elevation of Peak XV in the Himalayas, later renamed Mount Everest, and its identification as the world's highest peak in 1852.

Methods

Various approaches may be adopted according to the data type to be collected, the accuracy to be reached, and the consideration of the true spherical shape of the Earth.

Geodetic surveys cover large areas of the Earth's surface with measurements over long distances, where curvature is significant and taken into account to reach a high accuracy. The objective is to locate highly accurate points on the terrain and establish triangulation networks of control points, used to support further plane and photogrammetric surveys.

However, the most commonly practiced method is plane survey, a local data collection covering much smaller areas, so that the curvature of the Earth's surface can be ignored and considered a flat plane, without compromising the accuracy of the positions obtained.

Surveys which only measure the horizontal position of points and features, in order to establish its geographical location, are identified as horizontal surveys, obtaining just coordinate pairs. While those determining the elevation relative to a reference surface, or the difference in heights between vertical positions, are termed 'vertical surveys'. The most common reference surface used is the mean sea level at certain coastal locations, commonly indicated in topographic maps. However the most common occurrence is the combined measurement of both positions.

Echo sounding is the method used in hydrography to survey the bottom of water masses and courses. An application of active sonar is that it sends acoustic pulses from the surface down to the floor and then records the return signal. The distance to the bottom is calculated as a function of the time for the signal to travel from the instant of transmission to the pulse echo reception. Additionally, the changes in the shape of the reflected wave help to characterize the properties of the material and texture type of the bed.

Instruments

A range of instruments have been utilized in surveying history, succeeding each other in the search for a higher accuracy and efficiency.

Distances were determined with chains and flat steel tapes of a known length, but they introduced substantial error in measurement until they have given way to the electronic distance measuring (EDM) device. The EDM measures distance by sending an infrared light beam to the target – commonly a reflecting prism – which is reflected back to the instrument and calculating the time invested by the light beam to travel.

The basic instrument for measuring horizontal angles is the surveying compass, but commonly both vertical and horizontal angles are measured using the same instrument. The alidade, dumpy level, transit and, finally, the theodolite came to replace all the other telescopic devices. The introduction of the telescope in instrumentation allowed to precisely locate the target points at great distances. The theodolite has been the most used instrument for measuring horizontal and vertical angles since, at least, the sixteenth century when it was first described. It consists of a telescope, mounted on a tripod that rotates 360° around a vertical axis and vertically around a horizontal axis, which are perpendicular. With level control, the set of graduated circles allows us to read the two rotation angles when aiming at a certain position, relative to a fixed horizontal position and the horizontal plane when pointing to the surveyed target. In the electronic theodolite, readout is done by electronic means assuring a higher accuracy, and data recording brings the capability of digitally storing large volumes of data. The integration of an electronic theodolite, a distance measuring device and processing capability in a single instrument led to the total station.

Global Positioning System (GPS) surveying has not only introduced a new technology but also specific techniques. 'GPS is a radio based real time navigation and positioning system with a global coverage which allows to know the exact position of a point on the Earth's surface in almost any meteorological condition. The geostationary satellites of the system emit a continuous signal with information on the exact time it has been sent out, that allows a receiver to calculate the distance to each satellite by estimating the time invested to arrive. When at least three satellites of the constellation are simultaneously visible the receiver calculates, by triangulation, its local position in geodesic coordinates (*x*, *y*, and *z*) in the WGS84 Geodetic datum. The distance to a fourth satellite facilitates to calculate the height or vertical position of the point'. Three dimensional (3D) measurements can be completed at any accessible point without visibility limitations, and a high accuracy can be achieved with differential GPS. The conventional static mode data capture is enhanced with the kinematic measurement using a moving receiver, and when data are transmitted and promptly processed a real time kinematic (RTK) GPS surveying is feasible.

The application of light detection and ranging (LIDAR) to terrestrial laser scanning has initiated 'a new phase' of the field, high definition surveying (HDS), by using laser scanner technology or HDS scanner. This conflating technology between the fields of photogrammetry and surveying facilitates automatic data capturing and the generation of large data sets of 3D position coordinates in the form of a dense point cloud. The instrument scans the target area with a narrow beam of pulsed laser in a short time producing a detailed physical model of reality, cutting down the need of revisits and making it appropriate for remote or inaccessible areas. The point density allows an accurate and precise description of features which, combined with surface characteristics of objects obtained from the strength of the laser return signal, renders a photorealistic image using grayscale or hue based color maps.

In hydrographic surveying, the sidescan or side looking, sonar system is used as an underwater imaging system. It emits narrow fan shaped pulses perpendicular to the path of the survey vessel and continuously records the strength of the return signal. The multibeam sonar system adopts the same technology but focuses on measuring the time for the signal to travel, in order to produce data about depths instead of just images.

Techniques

Triangulation is a type of survey which starts at a baseline joining two positions with a known distance and grows by adding sides to form a triangle, measuring the angles formed – always exceeding 20° – and shaping a network of connected triangles whose sides have 'calculated' distances. Triangulation may adopt various patterns in the search for a higher coverage or accuracy: a single chain of triangles, a chain of quadrilaterals, or a central point arrangement. The mode of operation in 'trilateration' is the calculation of angles from the measured lengths of the triangle sides instead of determining them directly.

A traverse survey is formed by a series of concatenated lines between points whose bearing and distances are measured. This configuration is optimal to adapt to terrain and obstacles. The traverse may reach the de parture point or remain open. The open traverse starts at a point of known position and ends at a point whose relative position is measured, so that it is functional to survey routes, such as roads or water courses. The closed traverse starts and ends at the same known point, a procedure which is applicable to polygons.

Vertical surveys determine the vertical position of points using differential and trigonometric leveling techniques. Through differential leveling elevation is determined as a difference between heights, and in tri gonometric leveling that is estimated as a function of vertical angles and horizontal distances.

According to the coordinate system used, there are two modes of positioning – relative and absolute. Relative or differential positioning is a mode of indicating the position of a point with respect to another, taken as the origin of a local coordinate system. In the second mode, the position of a point is referred to a universal coordinate system with a distinctive origin such as the geocenter in the geocentric coordinate system. In GPS surveying, the latter mode is generally common while in terrestrial surveying the for mer is more frequently used, assigning coordinates of new stations to an arbitrarily origin station.

Applications

The different applications entail specific methods and techniques. Topographic surveying requires both hori zontal and vertical control to determine the location and elevation of selected features of both the natural and built environment, as well as terrain elevation. The objective is to obtain the data needed for producing detailed and accurate topographic maps showing contour lines, ele vation points, cultural, and natural features, which are true to scale.

Control surveying, also known as geodetic or trigo nometrical surveying, deals with the establishment of a series of highly accurate horizontal and vertical control points, or stations, according to specifications and standards, using a defined datum. The points are solidly fixed to the ground and provide a network of permanent positions used to reference further survey measurements.

The elaboration of nautical charts is the result of hydrographic surveying, which precisely maps the depths of water bodies in coastal sea areas, lakes, reservoirs, and rivers. It is required for the purposes of safe surface navigation and support of coastal engineering works or monitoring dynamic processes.

Cadastral surveying is the set of operations leading to the compilation or updating of a legal cadastre, a register of the ownership of land parcels. Land and boundary surveys, closely related, are used for property or real state purposes and provide a description of the land parcel, and data to prepare plots detailing the boundary lines and calculate dimensions and area.

Surveying applied to construction, engineering, and mining is an application whose goal is the data collection to precisely represent the site to facilitate the structure design. Subsequent phases of the survey include placing the design on the ground, monitor and register the ad justment of the work to the original plans.

See also: Landscape.

Further Reading

Anderson, J. M. and Mikhail, E. M. (1998). *Surveying: Theory and Practice* (7th edn.). Boston: McGraw Hill.
Bannister, A., Raymond, S. and Baker, R. (1998). *Surveying* (7th edn.). Harlow: Longman.
Danson, E. (2001). *Drawing the Line: How Mason and Dixon Surveyed the Most Famous Border in America*. New York: John Wiley and Sons.
Edney, M. (1997). *Mapping an Empire: The Geographical Construction of British India, 1765 1843*. Chicago: University of Chicago Press.
Irvine, W. H. (1995). *Surveying for Construction* (4th edn.). London: McGraw Hill.
Johnson, A. (2004). *Plane and Geodetic Surveying. The Management of Control Networks*. London: Spon.
Kavanagh, B. F. (2005). *Surveying: Principles and Applications* (7th edn.). Upper Saddle River, NJ: Prentice Hall.
Keay, J. (2000). *The Great Arc: The Dramatic Tale of How India Was Mapped and Everest Was Named*. New York: Harper and Collins.
McCormac, J. C. (2004). *Surveying* (5th edn.). New York: John Wiley and Sons.
Moffit, F. H. and Bossler, J. D. (1998). *Surveying* (10th edn.). Upper Saddle River, NJ: Prentice Hall.
van Sickle, J. (2001). *GPS for Land Surveyors* (2nd edn.). New York: Taylor and Francis.
Schofield, W. and Breach, M. (2007). *Engineering Surveying* (6th edn.). Oxford: Butterworth Heinemann.
Uren, J. and Price, W. F. (2005). *Surveying for Engineers* (4th edn.). London: Palgrave Macmillan.
Whyte, W. S. and Paul, R. E. (1997). *Basic Surveying*. Oxford: Laxtons.
Wolf, P. R. and Ghilani, C. D. (2005). *Elementary Surveying: An Introduction to Geomatics* (11th edn.). Upper Saddle River, NJ: Prentice Hall.

Relevant Websites

http://www.fig.net
International Federation of Surveyors.
http://www.gmat.unsw.edu.au
Principles and Practice of GPS Surveying.
http://americanhistory.si.edu
Smithsonian National Museum of American History.
http://www.surveyreview.org
Survey Review Journal.
http://www.fig.net
The Meridian Arc Measurement in the Equator.

Sustainability

T. Marsden, Cardiff University, Cardiff, UK

Introduction: Origins

Recent decades have witnessed a significant shift in social and geographical thought as various disciplines in the social sciences have attempted to incorporate eco logical awareness, actions, and vulnerability into their theories and paradigms. The steady recognition in the second half of the twentieth century that the modern industrialization and urbanization process – unleashed first in the North and then in the South – was creating increasing environmental vulnerability; and that it was humans who were both the cause and a significant consequence of this, created a fertile intellectual seed bed on which notions and debates about 'sustainability' were to be built.

Human geography, with its long tradition of linking the natural and cultural landscape to human settlement and activity, was at the forefront of tracing the growing 'metabolic rift', which was being witnessed between seemingly unstoppable economic processes and their environmental problems. This had first been recognized as early as the mid nineteenth century as artificial fer tilizers were introduced to 'solve' the problems and needs of increasing agricultural output for a growing industrial and urban population, within the intensive agricultural systems of North America and Europe. It was then re inforced in the post war period of the twentieth century as the state induced 'intensive food regime' became more dominant and associated with the ecological effects of pesticides and herbicides on local human and natural ecologies; but it quickly spread to an array of industrial and consumption processes as the effects of pollution, contamination, soil depletion, and water quality were first defined and then measured.

From the 1960s onward there were many conferences and debates concerning the impending 'environmental crisis'. This was not just associated with the 'limits to growth' assumptions concerning the gradual burning up of finite fossil fuels. Neither was it simply about society's avaricious appetite for consuming finite natural re sources, like coal, oil, and a host of other minerals. Ra ther, it began to question whether the now rapid and intensive processes of economic and social development themselves were indeed severely debilitating the ca pacity of natural ecologies and resource structures to renew themselves at all. Hence sustainability, and its more policy related and normative variant – sustainable development – began to focus on the vulnerability and potentially irreversible denudation of both nonrenewable and formerly renewable (e.g., fish stocks, animal breeds, forests, soils, and biodiversity) resources.

At the same time sustainability became a conceptual question – sustainability? – in that it raised both ques tions about the historical development of capitalism as a severe 'metabolic rift' between nature and society on the one hand; and more pressingly it raised questions of 'intergenerational equity' – that is, the ability or other wise to pass on to future generations the renewable capabilities of ecologies, on the other. Marx, following Liebig, a century before, had recognized the 'rift' that was occurring as a result of the onset of capitalist indus trialization, urbanization, and intensive agriculture, with the removal of soil fertility aided by the growing geog raphies of trade and the removal of human labor through technologies. Now, however, for the first time, a wider scientific and civil movement was developing around a broader set of concerns with regard to the maintenance of the socioecological metabolism after a century of pro gress in the capitalist modernization project.

At that time the wider macro global concerns about climate change and global heating had yet to be fully recognized. Nevertheless, a growing number of com mentators began to consider the implications that pre sent generations were, indeed, playing a sort of 'Russian roulette' with the natural and ecological heritage. Human geographers, having conveniently demarcated a sort of intradisciplinary binary divide between 'human' and 'physical' geography as the twentieth century had progressed, now, by the 1980s, were recognizing the need to reassemble and to integrate their discipline once more around the growing 'environmental agenda'. A major problem for geography in the latter stages of the twentieth century was to maintain and reactivate their longstanding interest in *Staffwechsel*, or the social–eco logical metabolism, and to do so with much more ur gency as the other social sciences, such as ecological economics and environmental sociology began to de velop their interest in this from a much lower intel lectual base.

Finding Sustainability

The 1980s and 1990s were a formative time for this in tellectual ferment. They were a period when three key external features to the discipline began to embed themselves in public and policy concern. First, it was a time when the environmental problems of intensive agriculture in the advanced world at least began to be measured and assessed. It was clear that there was not simply one 'metabolic rift' now occurring but many. Often adopting the more authoritative vocabularies of the economists, it was clear that conventional agricultural production was creating a cocktail of largely un controllable environmental 'externalities'. These, as the Organization for Economic Cooperation and Develop ment (OECD) finally recognized in 1986, included: eu trophication of water courses, landscape and biodiversity loss, water and ground pollution, air pollution, animal and plant diseases, and human health concerns (not least bovine spongiform encephalopathy (BSE)).

Second, especially after Chernobyl, Bhopal, and Three Mile Island, as well as a host of smaller oil spills and pollution incidents, it was becoming clear that des pite the growth of tertiary sector industries and new pollution saving 'clean' technologies, the global as well as regional environmental risks of pollution of advanced industrial development were significant, and likely to affect the more economically and socially vulnerable in society. Third, and indeed emanating from these, was the growing recognition of carbon emissions and its effects on global climate change.

All three of these strands created a major intellectual and research agenda for human geography, and very much continue to do so. They have led to large inter disciplinary research programs in which human geog raphers are playing a central role. Moreover, they have led to the development and reinvigoration of significant subfields, such as political and cultural ecology; rural studies; urban and political ecology; and environmental, economic, and cultural geography. Further, they have led to significant paradigmatic contestation within the dis cipline, for instance, between political ecology and economy and postmodernity, or between structuralism and behaviorism; productionism and consumerism.

Governing Sustainability

If the onset of sustainability debates became a 'call to arms' for scholars it also, by the late 1980s, became a major element in government and policy debates. The World Commission on Environment and Development in 1987, and especially Bruntlands' much quoted definition of sustainable development as 'development which meets the needs of the present without compromising the ability of future generations to meet their own needs', immediately broadened, and some may say diluted, the sustainability problematic by ushering in notions of de velopment, intergenerational equity, the precautionary principle, and a host of social and economic (as well as environmental) concerns. For the next two decades we have witnessed a period when the concept of sustainability has thus simultaneously deepened, broad ened, and become more of a contested discursive site upon which different environmental interests (coming from the private, public, and civil arena) have acted and articulated their strategies.

Whereas academics focused their prime energies on discovering what was not sustainable about the late modern world (i.e., finding and measuring unsustain ability), policymakers had another related agenda: to develop societies which were more rather than less sus tainable (i.e., to progress sustainable development), but to do so by retaining the benefits of an advanced capitalist and increasingly globalized economy. There were many assumptions about how advances in technologies could enable capitalism to innovate itself out of its smoke stacked past. Under this definition it was important that sustainability was at least linked, if not made synonymous with sustainable 'development'; for it was not just to be seen as a strictly environmental or natural problem, or to be left to environmental scientists alone. Rather, it was to have social and political objectives – in short it had to be politicized to be progressed.

If academics were now to practice both critical and managerialist perspectives with regard to their 'sustain ability science', then government and governance debates began to take a much more technicist and ecologically modernizing approach. With the coincidence of the end of the 'Cold War', which created a hegemony of belief with regard to the sustainability of capitalism (if not with its social ecology), the furthest government bodies could go was to embrace an ecologically modernizing approach. Supported by selected European scholars who had been studying advances in the chemical and other high pol luting industries in the 1980s, and by many (softly sus tainable) nongovernmental organizations (NGOs) who were all too pleased that at least governments and cor porate firms were at last prepared to share a table with them, the 'post Bruntland' period has been characterized by an 'ecologically modernizing approach' by many governments and leading corporate firms.

This approach assumed, rightly or wrongly – and it is still an open question – that through a variety of socio technological and market based means it is possible to turn, as one leading environmental modernization (EM) scholar typified it, the lumbering 'industrial caterpillar' into an 'ecological butterfly'. In other words, to attempt a new transition of progressive economic and social mod ernization which recreates in a far more technologically sophisticated way a new *Staffwechsel* – a new social–eco logical metabolism – which is capable of meeting the grand aims of the wider definitions of sustainable de velopment and becoming legitimate to a far more en vironmentally aware public.

Ecological modernization thus poses a key set of questions and challenges to both scholars and policymakers alike: 'To what extent are we seeing the arrival of a more ecologically modernizing process operating in advanced societies?' 'If we are, how is it being progressed and what sorts of relationships and organizing frameworks does it imply between public, private, and civil society sectors?'

Analytical Challenges in Understanding and Creating Sustainability

Scholars and policymakers have, in beginning to address these challenges, faced some key conceptual and methodological obstacles in recent years. First, is the growing 'recognition of the complexity' of the problem of sustainability itself. While earlier scientific work pointed to the multitude of differently sourced environmental externalities emanating from the modernization project of the twentieth century; governance debates have compounded this complexity by significantly broadening its ambit to cover social, economic, and environmental justice/exclusion concerns. The reaction of most natural scientists to this complexity conundrum has been to increasingly focus on the specific, and to conduct micro studies of complex biological phenomena, leaving climate change modelers to deal with the macro abstractions and assumptions of predicting climate change. Social scientists and many policymakers have alternatively recognized the need for more 'holistic' and publicly deliberative approaches, however difficult or complex these may be to validate. Human geographers have only recently seen it as one of their roles to focus on both the problems of real and analytical scale in sustainability debates; and to move (between these scales) from the detail of very rich case studies of unsustainability to broader aspects of theorizing.

What is clear is that for the human geographer and the policymaker the 'problems of scale' (both as a physical, governance, and social phenomenon) are as significant as the problems of temporality and intergenerational change in the study of sustainability. The framing of 'the sustainability question' thus becomes not only how can more sustainability be achieved and over what time period, but where, by whom, and at what spatial scale is it most likely to be successful? Is the local the best place to start? How should strategic national and regional sustainability concerns link with local concerns?

A second analytical problem for human geography relates to how it deals with what we might call the critical assessment 'of the real sustainability question'. This calls into question debates about realist versus constructivist debates. Sustainability, as we have seen, creates a question; indeed a normative question as much as it does a critical question. How might a more sustainable society develop? If we see expressions of this, how far are they related to deeper structural shifts in the economy or polity? What might its new human geography look like? These have been difficult questions for many human geographers to address having been trained in a more critically interpretative, as opposed to critical–normative stance and perspective. This raises questions about sustainability planning, visioning, and building as much as it does about the identification, measurement, and continuance of unsustainable practices or their connected sets of power relations. In dealing with this conundrum it is necessary to interrogate the articulation and the practice of sustainability in the context of at least three spatial frames: uneven development, articulation and appropriation, and production and consumption systems. We will deal with these three framings in turn.

Uneven Development

Sustainability research in human geography and in the broader social sciences more generally, has raised important questions concerning spatial inequality and uneven development. Some of the more radical writers see this inequality as both a cause and a consequence of growing environmental vulnerability and unsustainability. This works at different spatial scales. Clearly, late capitalism and its more recent neoliberal and globalizing tendencies have exposed many countries and regions to new forms of environmental exploitation as part of the race toward more capitalist and globalized 'development'. Northern advanced countries and their corporate firms may advocate more ecological modernization at home (e.g., the raft of European Union (EU) legislation on pollution, waste recycling, and agricultural land protection), while basing this upon the importation of 'cheaper' goods, services, and commodities produced in more unsustainable conditions.

Thus environmental vulnerability and unsustainability can be conveniently exported in the name of neoliberal trade regulation and privatized supply chain management. This is most clearly seen in the contemporary sense, for instance, in patterns of trade (for such 'environmental goods' as 'fresh' foods, timber products, biofuels) from (as well as between) the newly industrializing countries of Southeast Asia, including most notably China and Latin America. This multicentric reproduction of dominant and dependent states and spaces as part of the global capitalist logic and the relatively cheap carbon based transport and patterns of trade on which it relies, is seen by many scholars as a root systemic cause of uneven unsustainability. The creation, recreation, and redistribution of rights and regulations associated with dominant, dependent, and what we might term 'void spaces' (areas of the world (e.g., parts of sub-Saharan Africa) which have an absence of any resources needed by the dominant powers), is a major feature and condition of contemporary sustainability. The current increasing politics and economics of the scarcity of oil reserves on the one hand, and the realization of global climate change on the other, serve to intensify these uneven development tendencies and their geopolitical controversies and manifestations. These have been clearly expressed both inside and outside of the current round of world trade (World Trade Organization (WTO)) talks and world environmental summits. Such global arenas become sites of major contestation, not only between void, dependent, and advanced nations and re gions, but also between politicians, corporate firms, and a variety of globally linked NGOs.

Of course, these patterns of uneven resource ex ploitation as part and parcel of the (declining) carbon based economy are also played out at the lower spatial scales, within states, and between cities and regions. There has been a recent and growing literature in many of the new environmental and planning journals about the processes and expressions of territorial environmental deprivation and justice. This was clearly expressed most vividly by the Hurricane Katrina disaster which com bined both the realization and denial of climate change with the uneven processes of human poverty, vulner ability, and its effects.

Even at the city level we see more affluent parts enacting a more ecologically modernizing path (through, for instance, locally based recycling and organic food box schemes), while on the other side of the tracks there remain 'food deserts', high risk air pollution, and lower levels of life expectancy. Hence, the onset of sustain ability debates has served to rightly reinvigorate and infuse debates about spatial inequalities at the multiple scales of city, region, and nation state. At the same time these spatial inequalities are now partly expressed through deepening disparities in living conditions and consumption patterns.

Contested Articulation and Appropriation

Many early critics of the rise of concern about sustain ability point to the fact that its usage and application can be far too widely interpreted and therefore devalued. This lack of precision and thus theoretical power has itself been challenged by many scholars who argue that it is necessary to adopt a more holistic and systemic ap proach to both why sustainability seems to be declining, and how and under what conditions it may be recreated. Clearly however, it has become a major task for scholars to adopt a critically interpretative approach to the usage and articulation of 'sustainable development'. This has become more urgent in recent years as major state agencies, and corporate firms have developed their own usage and articulation of sustainability discourses. In short, the term or concept is prone to different types of 'linguistic appropriation' which can lead to contested arguments concerning the real effects of state and cor porate actions. These tensions clearly surround the changes in strategies and 'straplines' of major corporates (like BP's 'Beyond Petroleum' label, for instance); and they also need to call into question state based commit ments to sustainable development.

One consequence of the now widespread insinuation of sustainability into both the boardrooms and corridors of power of both the private sector and the state is that a major intellectual challenge becomes how to decipher 'real' forms of progress and action in sustainable devel opment from what might be regarded as 'greenwash'. In short, does it become just another branch of corporate social responsibility (in corporate firms) and 'duty of care' in state agencies, or, more importantly, a real force for ecological modernization and systemic change? For in stance, the UK government and its devolved bodies (Scotland, Wales, and Northern Ireland), have recently adopted a set of shared sustainability priorities which are to be measured and enacted across government depart ments. These include four key areas of action: sustainable consumption and production, climate change and energy, natural resource protection and environmental en hancement, and sustainable communities. The govern ment embraces all three strands of sustainable development – environment, economy, and social inclu sion. The jury is still out, of course, about how these objectives will be achieved and early research suggests, not surprisingly, more progress on some of the softer targets (like more sustainable government buildings and waste recycling) than on the hard nuts of advancing re newable energy and reducing carbon emissions. Inertia and backlash politics are also a critical part of the sus tainability debate, especially during the midst of eco nomic downturns.

Nevertheless, it is clear that concepts of sustainable development are now a major dimension of both pol icymaking and monitoring; whether they become com pletely absorbed into a 'managerialist measurement' and 'indicator culture', or lead to real systemic change is now becoming a key governance question for human geographers.

Practices and Planning: Sustainable Production and Consumption in Communities

Taking a more critically normative stance in geo graphical research becomes a necessity we can argue with regard to the question of sustainability. Moreover, despite the avoidance of some scholars to tread along paths which may suggest policy implications of their findings, it is clear that this cannot really be avoided in sustainability research. Of course, the development of what some scholars are now calling interdisciplinary 'sustainability science' has to be rooted in rigorous, critical, and theo retically and empirically based research. This research, however, also needs to suggest solutions and possible ways of achieving systemic shifts toward more sustain ability. For it is inherently a human and geographical question.

A major challenge for us concerns how to link our basic and applied research to the question of sustain ability planning. For instance, a critical area here – as indicated by major EU studies of the life cycle analyses of major commodities, goods, and services – concerns how scholars can track existing production and con sumption supply chains, assess their degree of sustain ability (or otherwise), and then find ways of adjusting or influencing sets of behavioral, regulatory, and market based adjustment factors which could lead to more sus tainable production–consumption systems. For instance, it is now evident that energy, transport, food, and household goods and services are four of the key con tributors to carbon emissions and global heating. How we adjust these key production–consumption systems and in what ways will be a key question for sustainability science.

One geographical aspect of these key systems of course is that of their relative spatial embeddedness. Production–consumption chains, while dependent upon increasing levels of mobility and transport, are also rooted in spaces and nodes which increasingly transgress national and regional boundaries. Does it make sustain ability sense to 'short circuit' these chains and make them more localized? This relocalization process is now oc curring in some major European cities around food chains. For example, in Copenhagen, Frankfurt, and Malmo the provision of relocalized organic food has now become mainstreamed, representing a real form of eco logical modernization rather then just another middle class 'niche' activity. The role of greener forms of city and regional governance have yet to be fully explored yet they will be crucial in shifting production and con sumption chains in a more sustainable and less carbon dependent direction. Sustainability planning needs therefore research which can understand the complexity of city and regional governance and the obstacles con sumers and producers face in achieving more sustainable supply chains.

City and city regional planning, indeed, needs to be more multidimensional by not just dealing with the two dimensional frameworks of more sustainable land zoning and the built environment of cities and towns. Rather, it needs to consider both urban and rural spaces, and the eco economic interactions between them as sets of complex networks, supply chains, and flows of goods, services, and knowledge. The current emerging policy and academic concern regarding 'sustainable com munities' is a vital development in this regard which could serve to link more basic sustainability research to sustainability planning. Clearly the engagement of more vibrant local communities of place, as well as the dis persed communities of interest needs to be incorporated into sustainability planning.

Conclusions: From Recognition and Appropriation toward Experimentation

The past three decades have seen the steady, contested, but nevertheless profound rooting of sustainability de bates and concepts into the political and economic fabric of both advanced societies and some developing coun tries. At the same time this has promoted and encouraged human geographers to reconnect with many of their earlier origins, associated with the quest to understand the 'social–ecological metabolism'. This has also led to sustainability debates becoming a major intellectual magnet for those scholars practicing in the more estab lished subdisciplines of economic, social, cultural, and political geography. Indeed, an important intellectual trend has been to recognize that sustainability theory needs to integrate and be critically interwoven with broader and challenging spatial theories of the economy and polity.

However, today, the paradox is that after a century and a half of the application of a carbon based modernization project, the globalized, resource intensive and un sustainable world in which the human geographer em ploys these approaches is far more complex, vulnerable, and potentially irreversible. This requires us to address many of the analytical challenges identified in this article, and its needs to invigorate both a 'critically interpretative and a critically normative approach' which links basic research with how to create systemic and transformative shifts in resource allocation and regulation across and through space. To do this it will need to be open to new epistemologies and conceptual vocabularies which can guide a more effective understanding of how real and progressive ecological modernization can be constructed, not least in engaging communities of space and com munities of interest.

Sustainability arose initially as a question. And it sparked a fragmented agenda for the recognition and realization of its very antithesis – 'unsustainablity'. It's inception in the post war period arose out of the gradual realization that the inherent renewable resource depleting tendencies in capitalism needed to be harnessed and monitored by a more 'precautionary' approach to economic and social development. This has, more recently, due to a plethora of scientific and NGO effort and evidence, forced this onto the mainstream political, corporate, and social agendas in the guise of (what we might regard as a form of 'soft') ecological modernization.

However, we need to recognize that this is a continuing story. Sustainable development is now by definition a constantly moving, dynamic, and increasingly contested intellectual and political location. Toward the end of the first decade of the twenty first century it becomes a much more potentially emancipating and creative force. It is one which is beginning to unleash the capabilities of social and actor network based experimentation in both spatial communities and communities of interest by creating new ethical markets, and production and consumption networks around heretofore sectoral concerns: such as transport and mobility, waste, energy, social and ecological enterprise, amenity services, household goods and services, and relocalized food provenance and consumption practices.

A growing lacunae of research begins to show that 'room for maneuver' can be made to allow these communities to experiment and innovate with diverse and more contextually dependent combinations of human and ecological resources in ways which could mainstream sustainable living systems both at the city and the regional levels. The human geographer's role in providing and comparing the knowledge bases, as well as demonstrating the wider value of these experiments will be a critical component in the degree to which the struggle for sustainability can be really embedded in societies; most of which are still inertly organized around sclerotic regulatory structures which continue to deny, hide, and indeed spatially displace the true costs of unsustainable practices. In these ways the critical question of 'sustainability' is likely to grow in its maturity as a key and challenging concept central to the development of human geography, as well as contribute, more broadly to interdisciplinary 'sustainability science'.

See also: Agriculture, Sustainable.

Further Reading

Drummond, I. and Marsden, T. K. (1999). *The Condition of Sustainability*. London: Routledge.
Forsyth, T. (2003). *Critical Political Ecology: The Politics of Environmental Science*. London: Routledge.
Harvey, D. (1974). Population, resources and the ideology of science. *Economic Geography* 50(3), 256 277.
Harvey, D. (1996). *Justice, Nature and the Geography of Distance*. Oxford: Blackwell.
Marsden, T. K. (2004). The quest for ecological modernisation: Re spacing agri food and rural studies. *Sociologia Ruralis* 44(2), 129 147.
Marsden, T. K. (2006). Denial or diversity? Creating new spaces for sustainable development. *Journal of Environmental Policy and Planning* 8(2), 183 198.
Marsden, T. K. (ed.) (2008). *Sustainable Communities: New Spaces for Planning, Participation and Engagement*. London: Elsevier.
Mol, A. (2000). The environmental movement in an era of ecological modernisation. *Geoforum* 31, 45 56.
O' Riordan, T. (2004). Environmental science, sustainability and politics. *Transactions of the Institute of British Geographers* 29, 234 247.
Owens, S. and Cowell, R. (2002). *Land and Its Limits: Planning and Sustainable Development*. London: Routledge.
Redclift, M. (ed.) (2000). *Sustainability: Life Chances and Livelihoods*. London: Routledge.
Sachs, W. (1999). Globalisation and sustainability. In *Planet Dialectics: Explorations in Environment and Development*, pp 129 155. New York: Zed books.

Sustainability, Urban

M. Whitehead, Aberystwyth University, Aberystwyth, UK

Glossary

Collective Consumption An idea developed by urban theorists to describe the role of cities as concentrated centers for the delivery of key public services (including water supply, healthcare, housing, and other key facilities) to large numbers of urban residents.

Ecological Footprint A term developed to describe the surface land area and associated ecosystems which are required for and affected by a given city's resource requirements and pollution outputs.

External Relations of the City The various social, economic, and environmental relations which connect metropolitan areas with places which are located beyond the official boundaries of a city.

Internal Relations of the City The complex of social, economic, and environmental interconnections which unify the different residential, commercial, and industrial districts of a city.

Structured Coherences A framework devised by the Marxist geographer David Harvey to describe the ways in which the social, economic, and environmental infrastructures and ideological belief systems of urban areas are marshaled to create spaces which are compatible with continued capitalist accumulation.

Urban Cramming The increasing spatial concentration of people, industrial activities, and services in a limited number of established urban areas.

Introduction – On Darwin, Geography, and Urban Sustainability

In a recent article for the *Guardian* newspaper Jonathan Watts claimed that 2006 would be remembered as the year when human society finally proved the tried and trusted theories of Charles Darwin wrong. This intriguing assertion is based upon figures produced by the United Nations (UN), which indicate that at some point during 2006 (and for the first time in human history) more people will live in UN designated urban areas than in rural districts. According to Watts then, our collective refutation of Darwin stems from the simple fact that rather than adapting to suit our habitat, through the assiduous construction of cities, humanity has altered and irrevocably transformed its habitat to meet our own needs. This profound shift in the geographical constitution of human society has seen the city becoming an ever more scrutinized category of, and for, theory building, policymaking, and planning. A key corollary of this process has been the realization that the form and quality of urban social, economic, and environmental relations will increasingly define the form and quality of human life itself. It is in this context that the idea of urban sustainability has become such an intriguing, alluring, and heavily debated concept. As this article shows the notion of a sustainable city is a multifaceted and highly confusing ideal, its appeal, however, derives from what it appears to offer – namely a solution to some of humanity's most pressing socio environmental problems.

This article has two primary goals: first, to explore the impacts that the notion of urban sustainability has had on urban geography; and second, to consider the role urban geographers have played within the ever evolving construction of what urban sustainability involves. The remainder of this article unpacks the different theoretical and practical implications associated with urban sustainability. It begins by considering the origins of sustainable urbanism. Space is then given to consider the impacts which sustainable thinking has had on urban geography and the role which urban geographers continue to play in challenging and transforming the ways in which sustainable cities are understood and pursued. Analysis then moves on to consider the practical applications of sustainable urban policies within two very different geographical contexts, the Greenwich Millennium Village in London and Ibadan city, Nigeria. These case studies are used to reveal the great geographical diversity that exists both within and between cities regarding what sustainable urban development represents and involves.

Metropolitan Utopia and the Origins of Urban Sustainability

By far, the clearest and most detailed summation of the principles and practices associated with urban sustainability was developed in the Turkish city of Istanbul in 1996. It was in this place, and at this time, that the UN held its second international conference on urban issues (the first was convened in 1976 in Vancouver). Representatives of 170 countries attended the Istanbul Conference (which is also referred to as the Cities Conference or Habitat II) and collectively agreed on 100 commitments and 600 recommendations concerning the form and nature of urban sustainability. While there is neither time nor space to discuss all of the principles of urban sustainability agreed in Istanbul in this short article, it is possible to summarize its recommendation in three points. First, the Istanbul conference made it clear that sustainable urbanism is, at least in part, concerned with ensuring that cities are ecologically sustainable – that is to say that urban areas' uses of environmental resources and their production of environmental pollution should not exceed nature's ability to reproduce and absorb such metropolitan inputs and outputs. Second, the Cities Conference also asserted that urban sustainability is about addressing issues of social sustainability within cities – ensuring that the living and working conditions found in cities are tolerable (and particularly so for the most vulnerable in urban society). Finally, Habitat II emphasized that urban sustainability draws particular attention to the ways in which social, environmental, and economic processes interact within urban space. Consequently, while being concerned with social and environmental sustainability, sustainable urbanism recognizes the important role of the urban economy in supporting social need and environmental protection. It is in the context of this final consideration that urban sustainability has become synonymous with an integrated vision of urban policies and planning, within which the social, economic, and environmental conditions and needs of cities are considered together.

When trying to define what sustainable urbanism is, it has become an almost axiomatic reaction to search for an appropriate UN quote like the one provided above. This is perhaps because of the fact that the specific goal of sustainable urban development has been pursued most noticeably by two UN agencies: the UN Centre for Human Settlements (UNCHS) (established in 1978) and the UN Environmental Programme (UNEP) (who in partnership with the UNCHS run the UN's Sustainable Cities Programme). In 2002, the growing significance of sustainable urbanism was recognized by the UN when it upgraded its settlement agencies into a full program of the UN – known now as UN HABITAT (United Nations Settlement Programme). UN HABITAT acts as the coordinating organization for the delivery and promotion of UN sponsored programs of sustainable urbanism at a global level. The association between sustainable urbanism and the UN is indicative of the geopolitical origins of the sustainable cities ideal, but, as we will see, it is necessary to go beyond the UN in order to develop a comprehensive account of sustainable urbanism. International nongovernmental organizations (NGOs), geopolitical organizations (such as the European Union), and different nation states have all actively promoted differing visions of urban sustainable development. There are also many examples of more locally inspired visions of urban sustainability, which have either adapted UN ideas about sustainable development to suit particular local circumstances, or instigated their own definitions of what constitutes sustainability, in particular geographical milieu. Some see the diversity surrounding the notion of sustainable urbanism as a virtue (see following section), but it is also clear that such eclecticism carries with it some intellectual and practical dangers.

Urban Geography and the Sustainable City: Space, Scale, and Posterity

Before moving on to consider some examples of urban sustainability in practice, it is important to reflect upon the relationship between the sustainable city and urban geography. Over the last 15 years it is clear that more sustainable ways of thinking about urbanization have challenged established geographical conceptualizations of urban space. At the same time, however, it is also clear that urban geographers' conceptualizations of scale, locality, and politics have contributed to changing understandings of the sustainable city. This section explores this evolving relationship and the impacts, which it has had on geographical approaches to urban questions.

Provisional Dialogues: Spatial Forms and Sustainability

The connections between geographical thinking and discussions of urban sustainability are essentially as old as discussions of urban sustainability itself. Drawing inspiration from the work of urban geographers on the capitalist form of the industrial metropolis, early scholars of the sustainable city analyzed the spatial logics of modern cities in order to understand why urban areas had become unsustainable. In this context, analysts claimed that cities were not inherently unsustainable things, but that it was the spatial logics of capitalist urbanization – replete with its mega suburbs, sprawling commuter belts, and heavily congested commercial districts – that made cities so socially, economically, and environmentally unsustainable. The unregulated spatial expansion of cities has destroyed rural landscapes; seen an increasing social dependence on automobiles (with an associated escalation in air pollution levels); and resulted in the neglectful construction of unsafe and unhealthy places for urban citizens to live (particularly in developing world cities). While such unregulated (or at least under regulated) forms of urban development have, in part, been justified on the grounds of necessary economic expansion, urban congestion and inadequate affordable housing have meant that the spatial form of many industrial cities is now increasingly seen as economically, as well as socially and environmentally, unsustainable.

Geographical sensitivities to the spatial forms of urbanization informed debates concerning how urban areas could be made to be more sustainable. At one end of the spectrum, writers such as Peter Hall have argued that the key to sustainable urban development is to transform the spatial fabric of cities so that they provide more intensively used and inhabited districts. Peter Hall's point is that greater urban cramming and the tighter spatial integration of people's home, work, and leisure life is the key to a less energy intensive, and more environmentally sustainable future. On Hall's terms then, the geographical form of cities, when appropriately planned, offers a spatial fix for the creation of a more sustainable society. Others, however, are far less optimistic about the potential of cities to offer sustainable spatial contexts for low energy lifestyles. Joan Martinez Alier, for example, argues against Hall's position, claiming that cities are so hard wired into the logics of capitalist development that they can never provide a sustainable geographical context for life.

Sustainable Challenges for Urban Geography: Space, Time, and Relational Thinking

Beyond debates concerning the most appropriate spatial form of the sustainable city, early geographical writings on urban sustainability focused on the practical difficulties of implementing sustainable development policies in urban areas. Gradually, however, it became clear that thinking about cities on sustainable terms did not only set practical conundrums for urban geographers to solve, but also challenged the intellectual grounds upon which geographers understood and conceived of urban areas. At one level of course, urban sustainability suggests the need to think more carefully about the complex relations which make up urban space. Visions of urban sustainability (at least those promoted by the UN) emphasize that urban economic, social, and environmental sustainability cannot be secured in isolation, but require integrated strategies of urban development. An appreciation of the integrated nature of urban space is not new to urban geographers (geographers have written widely, e.g., on theories of collective consumption and the structured coherences of urban areas). What is perhaps unique about visions of urban sustainability is the particular emphasis it places on the role of environmental consideration within the operation of urban systems. A concern with the environmental sustainability of cities has transformed geographical thinking about urban space in two main ways. First, it has drawn greater attention to the role of everyday urban environments in contributing to the operations of urban systems as a whole. Consequently, whether it be in terms of workplaces, homes, sidewalks, or parklands, it is clear that the urban environment has a major impact upon the physical and psychological health of metropolitan residents and a key role in defining the overall quality of urban life. The increasing awareness of urban environmental considerations within debates about urban sustainability has consequently been mirrored by a growing concern among urban geographers with questions of urban environmental justice. Second, discussions of the environmental sustainability of cities have brought attention to the role of cities within broader forms of regional and global sustainability. Cities are vast consumers of environmental resources and enormous producers of pollution. The point is that the resources which cities consume and the pollution they produce are neither sourced, nor wholly contained, within the territorial boundaries of the city. The ecological footprints of cities now reach around the world, from Antarctica to the Sahara, and from coral reefs to the ozone layer.

An awareness of the environmental reach of cities challenges closed, territorially bounded visions of the city, and encourages a more open, fluid, and process based account of urban life. This vision of the city is, however, not confined to environmental concerns alone. The desire to create more sustainable cities is consequently leading many city authorities to consider their extended socioeconomic as well as environmental relations (this is perhaps evidenced most clearly in the growing proliferation of fair trade cities and urban food networks). The connectivity of cities with global socioeconomic and environmental processes has been explored and developed by a number of geographers who write on the relational dynamics of cities and space. While it would be erroneous to assume that theorists of relational space have been directly influenced by work on urban sustainability, it is clear that both traditions are now facilitating a much richer account of urban space than had previously been thought possible. What perhaps differentiates work on urban sustainability from work on relational space, however, is its insistence on an 'integrated' and 'relational' account of cities – an analysis of cities, which recognizes the complex socioeconomic and environmental relations, which come together in urban space.

Perhaps the most original challenge, which notions of urban sustainability offer to existing work within urban geography, however, relates not to questions of metropolitan space, but to issues of urban time. Time has featured in urban geography in two main ways: first, in relation to studies of the historical emergence of urban centers, and second, in relation to the impact of the urban experience on the social experience of time in the present. Urban sustainability, however, draws immediate attention to issues of intergenerational justice and the impacts of urban development on as yet unborn urban and nonurban residents. The 'stretching' of temporal considerations within work on sustainability, to allow for a greater appreciation of future time, represents a clear challenge to the often narrow timeframes within which urban geographers situate their work.

Geographical Challenges for Urban Sustainability

While notions of urban sustainability have challenged some of the key spatial and temporal assumptions of urban geographers, it is also clear that urban geographers have offered significant and concerted critiques of established visions of urban sustainability. At one level, geographers have been at the forefront of work exploring the politics associated with urban sustainability. Analyses of the politics of urban sustainability have challenged the idea that sustainable urban development simply involves the programmatic imposition of an accepted set of technological solutions for urban development. In this context, the work of urban geographers has revealed that the acceptance of urban sustainability, and the form which sustainable urban development takes, are dependent on local political cultures and the vagaries of geographical context. Related work within urban political geography has also started to question the governmental structures through which sustainable cities are being developed and managed. In their sustained analysis of the relationship between cities and climate change policy, for example, Harriet Bulkeley and Michele Betsill have questioned the often preeminent role given to metropolitan authorities in the delivery of sustainable urban development policies. Given the complex local, regional, national, and global relations that constitute global warming, the ability of urban authorities to address climate change in a comprehensive way is obviously diminished. It is in this context that Bulkeley and Betsill reveal the important role of appropriately networked, but multiply scaled institutions, operating at a local, urban, regional, national, and global level, in implementing and managing urban sustainability. While an appreciation of the multiple scales of political decision making which impact upon urban sustainability was recognized in the Brundtland Report itself, Bulkeley and Betsill's intervention has introduced an important critical awareness of geographical scale into the sustainable city debate.

The final challenge offered by geography to conceptualizations of the sustainable city was developed as part of what must count as the largest and most comprehensive engagement of geographers with the principles of urban sustainability. At a workshop convened in 1998, and supported by the National Science Foundation (of the USA), a group of 18 geographers from North America and the UK considered what a geographical perspective on urban sustainability could actually offer. In their final report, 'Towards a comprehensive geographical perspective on urban sustainability', this collection of geographers critically questioned the understanding of the 'urban' promoted within dominant visions of urban sustainability. The workshop argued that while addressing certain issues of urban sustainability at a citywide level had clear benefits, implementing a vision of sustainability at a metropolitan scale failed to recognize the multifarious spaces, communities, and cultures that constitute urban areas. Accordingly, it was argued that to talk about the sustainable city was misleading because it tended to neglect the fact that what is deemed to be sustainable in one part of a city, may be very different to the sustainability issues that are prioritized in other urban districts. It was in this context that the Workshop on Urban Sustainability claimed that policies for urban sustainability should focus on the issues facing urban localities – or those intermediate geographical scales, which while operating as part of wider urban systems, have their own histories, neighborhood cultures, and socio environmental pressures. In this context, the Workshop on Urban Sustainability sought to bring a heightened sense of geographical difference to work on urban sustainability and to reveal the importance of developing visions of urban sustainability which start with the priorities and needs of local communities, not the abstract goals of a technocratic plan for urban sustainable development.

Sustainable Cities in Practice

In light of the reflections that have so far been made on urban sustainability, this final section explores the development of different forms of sustainable urban development in two geographical contexts. The first geographical case study is provided by London's newly formed Millennium Village; the second example focuses on the implementation of a UN Sustainable Cities Programme in Ibadan, Nigeria. These two case studies are used to reveal the complex decision making processes that inform programs for sustainable urban development, and the value of a geographical perspective on sustainable cities.

Case Study I – England's Urban Renaissance and the Millennium Village Project

Since the election to power of Tony Blair's New Labour Government in the UK in 1997 the principles of sustainable development have become an increasingly important part of British urban policy. On coming to power the New Labour government was confronted with a set of urban problems which are not uncommon in many developing world cities. At one level, urban economic expansion and growth (especially in and around cities such as Birmingham, Bristol, Cambridge, London, and Manchester) was having two socio environmental effects: first, urban growth was resulting in cities expanding onto valued rural land (or 'green field sites'); second, with increasing demand for urban housing, the costs of living within cities were becoming prohibitively expensive for many people, and especially for so called key workers (including teachers, nurses, and various blue collar workers). At a second level, however, the New Labour government was also confronted with a series of urban communities that had serious levels of poverty and social disadvantage, and were finding it hard to attract jobs and investment. It was in the context of such paradoxical urban problems, that New Labour turned to the ideas of sustainable urban development as a strategy for tackling the problems associated with urban Britain.

While it is possible to trace the emergence of sus tainable development thinking within some of New La bour's earliest urban policy initiatives, the idea of creating more sustainable cities in the UK was only ar ticulated fully following the publication of the Urban Task Force Report, 'Towards an urban renaissance'. The Labour government commissioned the Urban Task Force to develop a futuristic vision of urban Britain. Led by the prominent architect Lord Richard Rogers of Riverside, in its first section the Urban Task Force established a vision of the British sustainable city – an integrated space of social justice, economic development, and environmental care. Although the Urban Task Force report has been heavily criticized for being an elitist vision of urban Britain, it established the foundation for the first urban white paper to be produced in Britain for 23 years. The urban white paper, or 'Our towns and cities: The future – delivering an urban renaissance', established the parameters around which more sustainable urban com munities should be developed throughout England.

Running parallel to the shift in British urban policy toward the principles of urban sustainability has been the establishment of the Millennium Village on the Green wich Peninsular, London (see **Figure 1**). The Greenwich Village project was designed to provide a model of urban sustainability for the new millennium. Based upon a 121 ha site in east London, the Millennium Village combines innovative and energy efficient architecture, an ecology park, affordable housing, and a fully integrated set of community services (see **Figure 2**). In keeping with the government's desire to encourage brown field de velopments and to foster sustainable urban communities, the Millennium Village has been promoted as a safe and healthy place for people to live in the city. Its sustainable credentials are based at a social level on the affordability of the Village's housing, the strong sense of community identity promoted in the area, and the availability of a range of urban services (including a healthcare center and school). At an environmental level, the sustainability of the Village is based upon its good links to public transport networks (particularly the recently completed Jubilee Extension Line), its eco architecture, and its links to an adjacent riparian nature reserve on the banks of the Thames. The image of the Millennium Village as a safe, ecologically sound urban development has become a central element within its promotion and marketing to potential residents (see **Figure 3**).

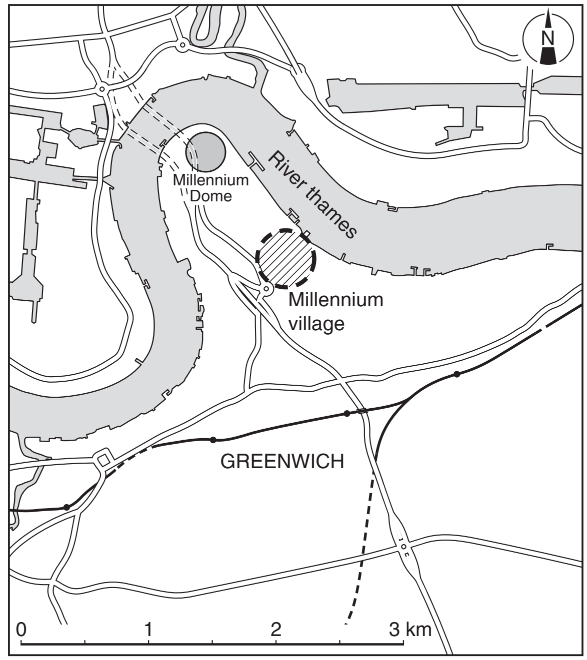

Figure 1 Map of the Greenwich Peninsular, showing location of the Millennium Village (currently being redrawn).

There are features of the Greenwich Millennium Village which mirror the classic forms associated with the ideals of the UN's sustainable city vision. It is, for ex ample, based upon a carefully integrated form of urban development, which has been specifically designed to minimize the environmental impact and associated en ergy demands of the settlement. Despite this, however, it is clear that the Millennium Village is not merely the product of an internationally sanctioned blueprint for sustainable urban development. Its particular focus on promoting sustainable urban communities, in a previ ously deprived and derelict area of a city, reflects an interpretation of sustainability which has been con ditioned by the particular urban pressures being ex perienced in the UK. In this sense, while clearly sensitive to issues of environmental sustainability, the Millennium Village (and for that matter England's broader urban renaissance program) is based upon an interpretation of urban sustainability which is focused on the ability of urban centers to continue to function as affordable and safe places for people to be able to live and work.

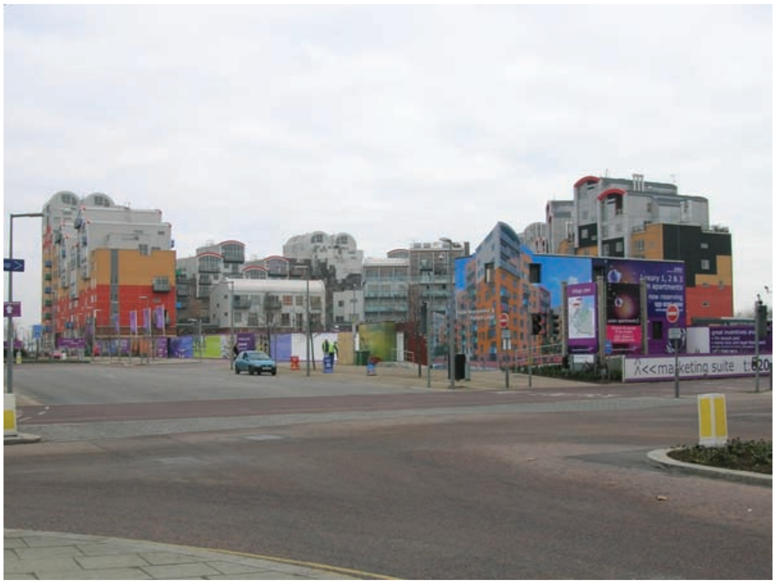

Figure 2 The Greenwich Millennium Village, London (Photograph by author).

Figure 3 Environmental narratives and the promotion of the Millennium Village as a sustainable community.

Case Study II – Urban Sustainability in the Developing World, Ibadan Nigeria

The second case study of urban sustainability in practice addressed in this article is taken from a very different geographical context than that of the Greenwich Millennium Village, and serves to illustrate the great diversity that exists in programs for urban sustainability. Ibadan city is located in Nigeria's southwest region and is known for being the most populous indigenous city in Tropical Africa (see **Figure 4**). As with many cities in less economically developed countries, the factors preventing Ibadan from becoming a sustainable place to live take a more pressing and immediate form than they do in cities of the developing world. Its strategic positioning between Lagos and Abuja has seen Ibadan grow in economic importance (particularly in relation to retail trade and public administration), and become a significant destination for rural migrants seeking work. The 1991 census revealed that Ibadan's population had swelled to nearly two million people (it is now believed to be in excess of three million). The purported economic success of Ibadan has, however, placed great stress on the city's fragile and under-resourced municipal infrastructure. One of the most

Figure 4 Ibadan, Nigeria (currently being redrawn).

worrying aspects of this process has been an inability of Ibadan's municipal waste and water supply networks to meet the needs of the rapidly expanding population.

The main issues defining the sustainability of Ibadan are consequently related to human health and the broader environmental implications of waste disposal and clean water supply. The supply of clean water to urban centers is a prerequisite of urban sustainability. Without clean water households cannot function, businesses can not operate, and human health rapidly deteriorates. At the same time the widespread presence of untreated waste can lead to a range of human health problems as well as untold environmental damage. In the context of Ibadan, the need to dispose of untreated urban waste, for example, poses a major threat to the ecological integrity and sustainability of surrounding woodlands and savan nah areas. As is often the case with cities facing basic threats to their social and environmental sustainability, however, the problems of Ibadan are not only the product of an underdeveloped urban infrastructure, but also poor planning and inappropriate municipal management structures.

It was in the context of the complex and intractable problems facing Ibadan that in 1994 the UN launched a Sustainable Cities Programme for the metropolis. While focusing on the provision of key water supply and waste disposal services, what is most interesting about the sustainable Ibadan initiative are the ways in which it has sought to deliver urban sustainability through active forms of local community engagement. The sustainable cities initiative in Ibadan has challenged the top down, bureaucratic orthodoxies of previous urban development plans, and tried to encourage much greater participation of people at a local scale in the decision making sur rounding the nature of future urban development. Community development was initially encouraged through a public consultation program at the beginning of the Sustainable Cities Programme. This consultation process was designed not only to try and uncover local sustainable development priorities, but also to try and promote a greater sense of community ownership over sustainable development projects in Ibadan. Following this community consultation process the citizens of Iba dan have also been encouraged to take an active role in the implementation of new sustainable development initiatives. These initiatives have ranged from spring water and borehole water supply projects, to waste management and local environmental improvement schemes. By involving local residents in such schemes, it is hoped that people can acquire both the skills and the requisite motivation to ensure that such projects are sustainable themselves.

The case of Ibadan serves to remind us of the great geographical diversity that exists when defining what urban sustainability involves and what should be prioritized within urban sustainable development projects. Ibadan also emphasizes that sustainable urbanism should not only be about sustainable development planning at large urban scales, but also involves the active engagement of local communities. It appears that only through a policy process which is sensitive to the needs and constraints of particular urban areas, turfs, and neighborhoods will it be possible to recognize and implement appropriate long term solutions to unsustainable social, economic, and environmental conditions within cities. Reports reflecting on the impacts of Ibadan's urban sustainable development programs show that while great progress has been made, resistances still remain to the practices associated with urban sustainable development within the urban authorities and key interests groups of the area. As a threat to existing patterns of urban development and management, however, perhaps the greater concern would be if urban sustainable development projects did not meet with opposition from established urban interest groups and authorities.

See also: Ecology; Locality Debates; Mega-Cities; Planning, Urban; Regeneration to Renaissance; Slums; Social Justice, Urban; Sustainability; Sustainable Development; Third World Cities; Urbanization.

Further Reading

Bulkeley, H. (2001). Governing climate change: The politics of risk society? *Transactions of the Institute of British Geographers* 26, 430 447.

Bulkeley, H. (2005). Reconfiguring environmental governance: Towards a politics of scales and networks. *Political Geography* 24, 875 902.

Bulkeley, H. and Betsill, M. (2003). *Cities and Climate Change: Urban Sustainability and Global Environmental Governance*. London: Routledge.

Bulkeley, H. and Betsill, M. (2005). Rethinking sustainable cities: Multilevel governance and the urban politics of climate change. *Environmental Politics* 14, 42 63.

Girardet, H. (1999). *Creating Sustainable Cities*. Dartington: Green Books.

Harvey, D. (1996). *Justice, Nature and the Geography of Difference*, ch. 13. Oxford: Blackwell.

Jonas, A. E. G., Gibbs, A. C. and While, A. (2004). State modernization and local strategic selectivity after Local Agenda 21: Evidence from three northern English localities. *Policy and Politics* 32, 151 168.

Lee, K. N. (2006). Urban sustainability and the limits of classical environmentalism. *Environment and Urbanization* 18, 9 22.

Rijkens Klomp, N., van de Lindt, M., van Asselt, M. B. A. and Rotmans, J. (2003). Integrative policy making for the improvement of the quality of urban life. In Girard, L. F., Forte, B., Cerreta, M., De Toro, P. & Forte, F. (eds.) *The Human Sustainable City: Challenges and Perspectives from the Habitat Agenda*, pp 89 106. Aldershot: Ashgate.

Swyngedouw, E. (2004). *Social Power and the Urbanization of Water Flows of Power*. Oxford: University Press.

Watts, J. (2006). Invisible city. *The Guardian* 15.3.06, 8 13.

Wheeler, S. M. and Beatley, T. (2004). *The Sustainable Urban Development Reader*. London: Routledge.

Whitehead, M. (2003). (Re)Analysing the sustainable city: Nature, urbanization and the regulation of socio environmental relations in the UK. *Urban Studies* 40, 1183 1206.

Workshop on Urban Sustainability. (2000). Towards a comprehensive geographical perspective on urban sustainability. *Final Report of the National Science Foundation Workshop on Urban Sustainability*. New Brunswick, NJ: National Science Foundation.

Relevant Websites

http://www.greenwich village.co.uk

The Greenwich Millennium Village Project.

http://www.unhabitat.org

UN Habitat Agenda and its associated declaration. Sustainable Ibadan Program: provides links to a range of other sustainable city programs. It is interesting to compare the different types of sustainable urban development projects that are being pursued in different regions throughout the world. General information about the UN Sustainable Cities Programmes.

Sustainable Development

J. A. Elliott, University of Brighton, Brighton, UK

Glossary

Environmentalism The ideas, language, and practices which stem from a concern for human–environment relations.

International Monetary Fund (IMF) A financial institution created in 1944 to regulate the international monetary exchange system. Lends money to member states to alleviate short-term balance of payment problems.

Millennium Development Goals (MDGs) Set of eight goals for international development with time-bound targets and measurable indicators agreed in 2000 by the United Nations General Assembly.

Neoliberalism A set of ideas about political and economic development associated with free-market economics, removal of trade barriers, and withdrawal of the state.

Nongovernmental Organizations (NGOs) An umbrella term encompassing a range of diverse organizations at various scales which are fundamentally outside government and generally not-for-profit.

Poverty Reduction Strategy Paper (PRSPs) Since 1999, the strategic documents around which decisions regarding loans and debt relief are made. Administered jointly by World Bank and International Monetary Fund. PRSPs are written by national governments, required to set out coherent plans for reform focused on poverty reduction and identify financing needs.

The Precautionary Principle An element for policy decisions concerning environmental protection and management. It is applied where there are reasonable grounds for concern that an activity does or could cause harm but where there is uncertainty about the probability of the risk and degree of harm. It was endorsed internationally within Agenda 21, for example, where Principle 15 states: "In order to protect the environment, the precautionary approach shall be widely applied by States according to their capabilities. Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation."

Transnational Corporation (TNC) Organizations with investments and operations in many different countries, with no clear 'home-base' and with factories, offices, subsidiary companies, etc., integrated at the international scale.

World Bank (WB) Common name for the International Bank for Reconstruction and Development (IBRD), which together with the International Development Association (IDA), the International Finance Corporation (IFC), and the Multilateral Investment Guarantee Agency (MIGA) comprise the 'World Bank Group'. Lends money to governments often on concessional terms and its activities are increasingly central to the lending decisions of other public and private institutions.

Introduction

The term sustainable development is a well used one and is probably familiar to many within and beyond academia, certainly in the more developed parts of the world. It is a term that we come across in arenas ranging from door step recycling initiatives to media explanations of global security issues. Within human geography, it informs research extending from social exclusion within cities of the United Kingdom to outcomes of environmental transformations in rural Africa. Indeed, some consider that there is none so relevant a discipline as geography to contribute to the sustainable development debates given its ability to marry the science of the environment with an understanding of economic, political, and cultural change, that is, development. The pursuit of sustainable development is now stated as a principal policy goal of organizations and institutions across all scales of public life and the field of academic and practical enquiry around sustainable development is a diverse and expanding one.

It is generally acknowledged that the report of the World Commission on Environment and Development (also known as the Brundtland Commission) published in 1987 did much to bring the term 'sustainable development' into the popular consciousness and onto public agendas. This commission, established by the United Nations (UN), comprised people drawn from member states of both the more developed and less developed worlds and was charged with identifying the long term environmental strategies for the international community. Its definition of sustainable development, as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs" has become the most widely cited expression of the term. The fundamental notion that development today should not be at the expense of that in the future has found widespread allegiance.

As the term sustainable development reaches both further into daily lives and becomes bound up with ever larger movements of the modern world, academics and practitioners are increasingly aware of the need to reflect critically on the fundamental principles encapsulated within the term as it evolves. In addition, close examination is needed of what is trying to be achieved and how, in the name of sustainable development, to encompass the multiple and often competing agendas being pursued and to interpret changes within dynamic local and global contexts. To that end, this review accords substantial detail to the origins and development of the notion of sustainable development and how the complex interdependencies of economic, social, and environmental development processes and their outcomes are being revealed in the pursuit of sustainable development.

The concept of sustainable development has gained some degree of notoriety including for its 'slippery nature' (the multiple definitions that it has), its ambiguities (the various interpretations that flow from those definitions), and its fundamentally oxymoronic character (the suggested opposition between the two encapsulated terms). This review details a number of frameworks that have been forwarded for handling the diversity and dynamism associated with the notion and points to the principal ongoing divisions within the field of enquiry. For some, the way in which the notion of sustainable development has been redefined so many times and in relation to so many aspects of society–environment relationships undermines its usefulness. For others, it is the contestations over the direction of social and economic development into the future (the discord of modern politics) that are the substance of sustainable development and as such, the utility of the idea lies precisely in the debate and compromise that it challenges researchers and practitioners to engage in.

Two particular literatures, those of environmentalism and of development, are considered to be particularly important in understanding the origins of sustainable development. The first use of the term 'sustainable development' is acknowledged to have been within the World Conservation Strategy of 1980 that was drawn up by the International Union for Conservation of Nature and Natural Resources. For the first time, development was forwarded as a means for achieving conservation bringing the two literatures closer together. However, the 1980s are also understood to have been an era of 'impasse' within both the theory and practice of development. Past theories were upturned and seemed to offer little in terms of explaining the current experiences of development and underdevelopment (let alone into the future). It was also a period when the failures of 'development' on the ground were increasingly evident including the environmental impacts of the mounting debt crisis and of the solutions implemented to solve it.

In the globalized era of the early years of the twenty first century, environmentalism is considered to be thriving, particularly as it has adapted to changing scientific evidence and has been informed by the ideas of related social movements such as ecofeminism. Whilst it can be considered that discourses remain dominated by environmental sustainability concerns, the work of geographers is proving important in placing human needs and rights more centrally into these agendas. In turn, development studies (and development geography) is considered to have moved beyond its impasse to be characterized by lively debate within which environmental, social, and economic sustainability are a central concern. Whilst this review provides evidence of the substantial work of geographers, particularly in uncovering the nexus of poverty–environment relationships, there is continued concern as to how far this work is impacting on the literature and practices of sustainable development.

One way in which sustainable development can now be considered a concept that has come of age is through its position as a primary policy goal of many of the major institutions of the world, including the UN and the World Bank (WB). In particular, finding new approaches to poverty alleviation is currently considered to be on a new and superior roll. This review analyses how this consensus has development and considers how the policy prescriptions that flow from it intersect with local and global environmental agendas.

Discussions of the idea and practices of sustainable development are centrally concerned with the future of the Earth and its inhabitants' relationships, and policy challenges that are the long standing traditional concerns of geographers. This review considers the contribution of human geography in exposing the inherently political and conflictual endeavor that is sustainable development, in particular through the work within political ecology.

Framing Sustainable Development

Given that almost all disciplines potentially have something to say in terms of analysis of the environment, of development, or with regard to policy for sustainable development, it is no surprise that there are a very large number of definitions of the term. **Figure 1** lists a small number of definitions that point to the range of different 'starting points' that can be identified regarding the capacities and functioning of the Earth and the requirements of society, now and in the future, locally and globally.

The notion of sustainable development is often portrayed in terms of three pillars or corners of a triangle as illustrated in **Figure 2**. Such frameworks encompass the idea that the arenas are interdependent and the objective of sustainable development is to maximize the goals across all three systems. Whilst there is no agreement as to the detail within the pillars/corners, these kinds of

A better quality of life for everyone now and for future generations to come.

The net productivity of biomass maintained over decades to centuries.

International resource development that is socially desirable, economically viable, culturally appropriate, and ecologically sustainable.

Nondeclining per capita utility.

The process of improving the living conditions of the poorer majority of mankind while avoiding the destruction of natural and living resources, so that increases of production and improvements in living conditions can be sustained in the longer term.

The amount of consumption that can be sustained indefinitely without degrading capital stocks, including natural capital stocks.

Figure 1 Definitions of sustainable development.

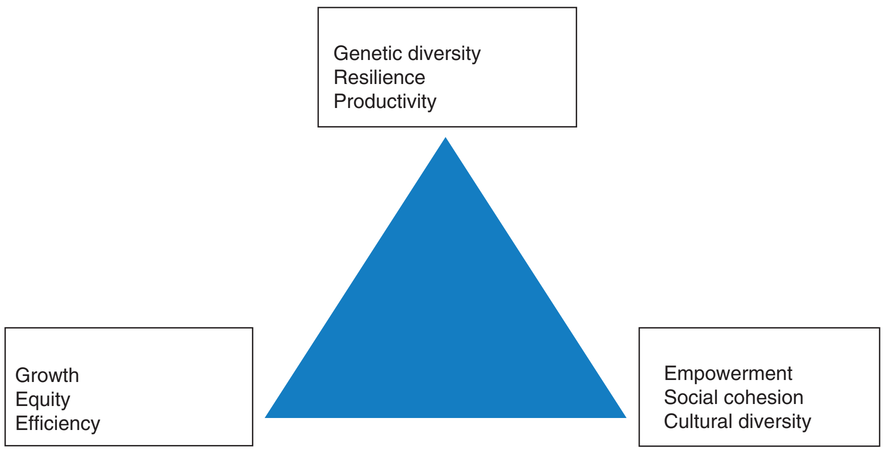

Figure 2 The objectives of sustainable development.

models are important for stressing how system goals are socially constructed and how sustainable development in practice requires choices and trade offs to be made (at particular points in time and at particular scales) regarding what is being pursued, how and at what costs to particular groups of people and interests.

Diversity and contestation over the meaning of sustainable development is revealed in the consideration of 'how' to conjoin human demands of the Earth without endangering the systems on which survival depends. A predominant way of organizing the varied interpretations of sustainable development has been to label them along a set of continua (often used interchangeably or in combination) as shown in **Figure 3**. Such frameworks are useful for exposing at each end, the very different views of human relations with the Earth, and of the nature and sources of change required for sustainable development. For those that articulate a weaker view of sustainability, the focus is on the ways in which nature can be adapted to meet human needs, through enhancing the stock of resources via developing renewables and creating substitutes, making more effective use of existing resources (enhancing efficiency and reuse), and technological solutions that will reduce problems of depletion and pollution. In this weaker view, there is no need to fundamentally alter the predominant narrative of a human centered view of nature or the existing discourse on what constitutes economic progress. In contrast, strong sustainability is a discourse that prioritizes the Earth itself, requires radical changes to what constitutes development, and forwards very different ideas in terms of interventions and the roles for the state, markets, legislation, and for individual citizenship in practice. In reality, these are spectra rather than dualisms such that distinctions are often blurred and individuals themselves may find favor with different elements at different times and in relation to the various roles they play in society at any point.

The meaning of sustainable development can also be considered in terms of 'what' it is anticipated to achieve. Target setting (and the associated search for appropriate indicators) has become increasingly popular as the interests in sustainable development have widened, as businesses and public authorities are being required to

	Expand stocks of resources	⟷	Revise human demands placed on the Earth
Views of nature and human relations with it	Desire to manipulate nature inherent to human condition Humans as separate from nature	⟷	Earth nurtures humanity's existence Human society as immersed in totality of Earth's systems
Degree of empathy with need to respond to constraints set by natural system within environmental management	Reformist More responsible institutions Individual self-regulation	⟷	Radical New, devolved political structures needed Self-reliant communities

Figure 3 Typologies of sustainable development.

account for their activities and as results based management has become a dominant approach to management within all kinds of organizations. Whilst many authors remain sceptical about the utility of such exercises (that are fundamentally reductionist) for reconciling inherently complex problems of sustainable development in practice, supporters point to the transformative potential of accessible information that usually means simple and quantifiable measures against which performance can be measured.

In 2000, the UN General Assembly adopted eight MDGs (with time bound targets and measurable indicators) against which international agencies, donors, and countries would be judged. Whilst international development targets have been set before, the MDGs (**Figure 4**) are considered unprecedented for their comprehensive nature, for the broad support that they have garnered and their pivotal role in shaping international development policy. One MDG goal refers to environmental sustainability and many others are central to sustainable development more widely for the way in which they demand better outcomes in many arenas that affect poorer groups of people.

The meaning of sustainable development can also be considered in terms of the practices being implemented to achieve it and it is in this regard that most diversity is revealed. As the range of organizations and interests aligning themselves to the notion has expanded, the practices of sustainable development include the efforts of local groups to define and shape a sustainable future for their communities of scientific academies in developing technologies that enhance sustainability within and across particular sectors, and the actions of social movements that operate transnationally to foster change in complex arenas of debt rescheduling, labor rights, or genetic engineering. In consequence, there are no blueprints for what sustainable development means in terms of practice, rather it is the outcome of ongoing dialog, conflict, and compromise in specific places and time.

1. Eradicate extreme poverty and hunger.
2. Achieve universal primary education.
3. Promote gender equality and empower women.
4. Reduce child mortality.
5. Improve maternal health.
6. Combat HIV/AIDS, malaria, and other diseases.
7. Ensure environmental sustainability.
8. Develop a global partnership for development.

Figure 4 The millennium development goals.

Mainstream Sustainable Development

An understanding of the contemporary challenges of sustainable development and opportunities in practice requires consideration of the historical antecedents from which the concept has developed. Many of the current conflicts surrounding sustainable development can be considered a product of the substantially separate nature of the literatures surrounding the environment and that of development in the past. The following section illustrates how debates on sustainable development have substantially reshaped understanding in both arenas to bring modern environmentalism and development studies closer together.

Modern Environmentalism

Whilst it is acknowledged that an intellectual concern for the conservation of nature goes back at least to the late eighteenth century (and indeed that religious ideas about the relationship between people and the environment have much longer standing), most accounts point to

Rachel Carson's 1962 book on pesticide pollution, *Silent Spring*, as a landmark for founding the popular movement known as modern environmentalism. It was through the 1960s that a coherent philosophy and language surrounding the environment developed and environmentalism started to influence public policy. As such, modern environmentalism is distinguished from longer standing literary appreciations of nature or the scientific analyses of species, for example. The subsequent history of environmentalism is necessarily complex, wide ranging, and varied in different countries. The focus here is to consider the 'mainstream' or broad consensus identified within major publications and international conferences, for example, concerning the predominant views of nature, the characteristics of prescribed conservation interventions, and the role of political economic change. However, like other social movements, environmentalism embraces substantial diversity from individuals through to ideologies as seen in **Figure 3** such that a consideration for diversity and dynamism has to be retained through any selective review.

In the 1960s, environmentalism was largely a movement reflecting European and American, white, middle class concerns. After a period of two centuries of industralization and urbanization, the undesirable effects of such development were beginning to be seen. Environmentalism became overtly activist and political and received substantial media support. The combination of actual changes in the environment and people's perceptions, generally, at this time brought widespread public support for the environmental movement for its focus on environmental problems and as a means for expressing anti establishment views and lobbying for political change more widely.

It was not solely local outdoor environments which were perceived to be under threat, but human survival itself, and the environmentalists' cause was enhanced by a number of influential 'global future studies' published in the early 1970s. These included texts such as *The Limits to Growth* (also known as the Club of Rome Report) by D. H. Meadows and co workers. In short, these modeled (often in some detail and complexity), the ever expanding human population and the mounting demands of society on a fundamentally finite resource base. To avoid the catastrophic consequences of crossing the physical limits of the Earth's resources, urgent conservation actions (particularly population control in the developing world) and 'zero growth' in the world economy were required. These global future predictions gave little attention to the social, technological, or institutional factors which affect the relationship between people and resources. They were also ahistorical in the sense that they gave no attention to how or why the world is divided into rich or poor and were apolitical in considering the future of the Earth (rather than people) as the overriding and paramount concern. There was no consideration of how the solutions advocated would favor some nations or groups over others. Within such understanding of the environment as the stocks of substances found in nature and of pollution and environmental deterioration as the inevitable consequences of industrial development, development and conservation at this time were portrayed as incompatible such that the environmental movement found little support within nations of the developing world keen to assert sovereign control over resources and to deliver further industrialization and growth.

The UN Conference on the Human Environment held in 1972 in Sweden, is considered a key event for putting environmental issues on the international political agenda. One hundred and thirteen nations participated at Stockholm and a new international institution for monitoring global environmental change, the United Nations Environment Program (UNEP), was a direct outcome of the conference. Whilst the impetus for and central activities of the conference remained centered on northern environmental priorities (i.e., the environmental impacts of industrialization), the term 'pollution of poverty' was used for the first time to refer to the environmental concerns of the poor, such as lack of clean water or sanitation, which were the development problems that threatened life itself for many in the developing world. Notions of sustainable development were widened to embrace understanding that a lack of development (as well as industrial development) could cause environmental degradation. However, there were no detailed discussions of the links between poverty and environmental degradation or the causes of poverty. There was continued mistrust on behalf of the developing nations regarding the extent to which ideas of development and environment were indeed understood as interdependent and of truly 'shared' problems of global resource management. Whilst proceedings within the conference confirmed that environment and development should be integrated, there was no plan of action as to how that integration should occur or be fostered. In short, whilst the previously separate arenas of environment and development had started to move together and become understood as interconnected concerns, the practical challenges of sustainable development remained ill defined.

In 1977, UNEP commissioned the International Union for the Conservation of Nature and Natural Resources (IUCN) to consider and report on the conservation problems of the world and to identify effective solutions to these. The World Conservation Strategy was published in 1980 and was envisaged as a comprehensive action oriented program for political change. The objectives of conservation were expanded from principally those of nature preservation to include aspects of human development through the ways in which human survival and future needs in science, medicine, and industry are linked to ecological processes and genetic diversity. In forwarding 'development' (in the form of 'appropriate' environmental management) as the means to achieving conservation (rather than an obstruction), the strategy can be considered important in the development of ideas around sustainable development. For many, the strategy shared many of the weaknesses of earlier global future publications, where the priority continued to be the problems for the environment raised by economic development and the root of degradation attributed to the escalating numbers of people and short sighted approaches to natural resource management. For a global agenda that stressed global responsibility and strategy, the insufficient attention given to the political realities of resource use or for implementing solutions was particularly problematic for many commentators.

In 1987, the report of the World Commission on Environment and Development (WCED) did much to insert human development concerns and the challenge of overcoming poverty. By showing how environmental problems were inextricably linked to economic and social inequality, the WCED brought development issues into environmental debates and exposed the environmental impacts of core North–South relationships in trade and industrialization as well as aid. The notion of sustainable development was extended through the recognition that the environment did not exist as a sphere separate from human actions and ambitions and the practical challenges of sustainable development were opened up through the Commission's consideration of the political and social requirements of sustainable development. The report was underpinned by a strong ideology that nature and the Earth could be managed through reviving economic growth, such that there was no specter of economic disaster portrayed by the WCED. Deteriorating environments were also seen to be in opposition to continued development such that tackling the broader factors underlying world poverty and international inequality was fundamental to overcoming environmental problems. Inequality was seen as a global environmental problem and seeking distributive justice was part of the solution.

The combination and interdependence of environment and development concerns of the WCED are evident in the critical objectives of sustainable development identified in **Figure 5**. Within the substance of the report, the WCED also challenged a number of dominant narratives on development–environment relationships that had pervaded modern environmentalism. For example, whilst ensuring a sustainable level of population was one of the critical objectives, the WCED recognized the differentiated impact on resources of current international patterns of consumption and considered that these patterns and processes of consumption were 'as important as numbers of consumers' in the conservation of resources. The commission also recognized varied responsibilities of different countries in implementing solutions and finding the less energy intensive forms of economic growth. The WCED has been accused of 'comfortable reformism' for the ways in which it suggests that economic growth can be reconciled with environmental conservation without any significant adjustments to the prevailing capitalist market system and for how it did not challenge patterns of consumption in the North. It is praised for identifying the substantial political and economic changes required in a range of arenas to achieve sustainable development and criticized for not going far enough in terms of identifying the barriers or obstacles to change in these sectors.

In 1992, the UN Conference on Environment and Development, the 'Earth Summit', was held in Rio de Janeiro, explicitly to consider progress on environment and development since the WCED report. At the time, it was the largest ever international conference held with over 170 governments represented and a further 2500 nongovernmental organizations (NGOs) participating in a parallel Global Forum. The central aim of the summit was to identify the principles of and an agenda for action towards sustainable development in the future. The challenge was seen to require consensus at the highest level, such that, for the first time, heads of state gathered to consider the environment. The outputs were the Rio Declaration on Environment and Development (27 Principals), the Agenda 21 document (a 40 chapter action plan), the Framework Convention on Climate Change, and the Convention on Biological Diversity.

There have been many and often opposing evaluations of the Rio conference, but most acknowledge the need to consider it as a process rather than an event. It

Reviving growth.
Changing the quality of growth.
Meeting essential needs for jobs, food, energy, water, and sanitation.
Ensuring a sustainable level of population.
Conserving and enhancing the resource base.
Reorientating technology and managing risk.
Merging environment and economics in decision making.

Figure 5 The critical objectives of sustainable development, WCED, 1987.

can be considered to have prompted new ways of thinking regarding the links between poverty and the environment and placed politics more centrally into the environmental arena that endures. Agenda 21 continues to carry political authority in framing the actions of local, national, and international development institutions and the Framework Convention on Climate Change has been supplemented by the Kyoto Protocol. For others, the agenda remained consistent with that of the earlier UN conference; one dominated by northern priorities focused on issues of the global commons such as climate warming and biodiversity that are more amenable to political and technical solutions rather than environmental concerns associated with poverty or global inequality. **Figure 6** summarizes the ways in which Agenda 21 is considered to have reinforced, rather than challenged, mainstream ideas around environment and development.

In 2002, the World Summit on Sustainable Development (WSSD) was held in South Africa with the stated aim of reinvigorating at the highest political level, the global commitment to a North–South partnership to achieve sustainable development. It is considered to have been a more inclusive summit than Rio for the way in which more stakeholder groups were brought into formal meetings (including a bigger presence for business), but also for the extensive informal participation of many individuals and groups in parallel civil society events. There is a widespread view that the period between Rio and Johannesburg has been one in which any consensus around sustainable development fell away. There were substantial difficulties within the preparatory meetings (PrepComs) in arriving at the statements intended to provide the framework for subsequent debates. Opportunities for constructive dialog were compromised by the US and the UK engagement in war on Iraq (without UN sanction). The absence of an international consensus on the challenges of sustainable development were also evident within the proceedings; illustrated in the failure to establish new targets for proportional increases in renewable energy use or for mandatory standards on corporate accountability.

The official outcomes of WSSD were the Declaration on Sustainable Development and the Plan of Implementation. The Declaration (notably shorter than the Rio Principles) is reproduced in **Figure 7**. The Declaration confirms again that poverty, inequality, and the environment are interdependent concerns within mainstream international thinking. Some evolution of sustainability thinking is also evidenced in the reference to the interdependent pillars of sustainable development and to collective responsibility in reinforcing these, suggesting that notions of development had expanded beyond economic development and social concerns (and to include the implications of globalization for the first time).

The Plan of Implementation details the next steps for action for sustainable development identified by the UN. It was poorly received by international nongovernmental organizations, particularly in the way it restated existing commitment and failed to identify funding nor mechanisms for how actions were to be realized. For others, the lack of fresh commitments or vision was not unexpected, given that world leaders had already committed to pursuing sustainability and suggested that the WSSD was essentially about implementing and monitoring progress.

Evidently, ideas surrounding sustainable development have influenced modern environmentalism and mainstream understandings of the environment and its relationship to the concerns of development, particularly in less developed regions of the world. In short, understandings of basic opportunities and needs in development, of persistent poverty, and distributive issues have come closer into environmental debates. For some commentators, this has led to a diffusion of ideas within environmentalism and confusion surrounding sustainable development. The suggestion is that environmental concerns have become secondary to issues of power and politics (e.g., as business interests become key actors within policy debates) and sustainable development

Agenda 21 predicated on economic growth as within WCED.
Emphasizes the familiar environmental issues and management of earlier reports such as the World Conservation Strategy.
Heavily technocentrist including the means through which sustainable development will be achieved, building on science, information, and sound technologies.
Assumes change will arise through mutual interests of North and South and of current and future generations.
Calls for participation on behalf of rainbow coalition' of women, children, young people, indigenous people, trade unionists, business, industry farmers, local authorities, and scientists.

Figure 6 Mainstream environmentalism as forwarded within Agenda 21. Compiled from Adams, W. M. (2002). Sustainable development? In Johnston, R. J., Taylor, P. J. & Watts, M. J. (eds.) *Geographies of global change: Remapping the world,* pp 412 426. New York: Wiley.

> We recognize that poverty eradication, changing consumption and production patterns, and protecting and managing the natural resource base for economic and social development are overarching objectives of, and essential requirements for sustainable development.
>
> The deep fault line that divides human society between the rich and the poor and the ever-increasing gap between the developed and developing worlds pose a major threat to global prosperity, security, and stability.
>
> The global environment continues to suffer. Loss of biodiversity continues, fish stocks continue to be depleted, desertification claims more and more fertile land, the adverse effects of climate change are already evident, natural disasters are more frequent and more devastating and developing countries more vulnerable, and air, water, and marine pollution continue to rob millions of a decent life.
>
> Globalization has added a new dimension to these challenges. The rapid integration of markets, mobility of capital, and significant increases in investment flows around the world have opened new challenges and opportunities for the pursuit of sustainable development. But the benefits and costs of globalization are unevenly distributed, with developing countries facing special difficulties in meeting this challenge.
>
> We risk the entrenchment of these global disparities and unless we act in a manner that fundamentally changes their lives, the poor of the world may lose confidence in their representatives and the democratic systems to which we remain committed, seeing their representatives as nothing more than sounding brass or tinkling symbols.

Figure 7 The Johannesburg declaration on sustainable development.

becomes aligned with broader concerns of justice and governance. For others, mainstream environmentalism remains centered on questions of the natural environment, of species extinction, and on ways of reducing consumption of natural resources, and therefore, disengages from sustainable development. Whilst this literature describes the issues of environmental degradation and resource futures as affecting everyone, the discourse continues to pursue the rights of the environment rather than human rights and the identified responses continue to differentiate the more developed from the less developed countries of the world.

Work in the field of environmental justice can be considered to be bringing the concerns of environmentalists and those with interests in inequality closer together. Whilst a defining feature of this grassroots movement as it originated in the United States has been the way in which issues are brought to public attention by people themselves rather than environmentalists, work in this arena reaffirms how poor people often live in poor environments (a significant step in the development of ideas of sustainable development revealed within the WCED) and how the costs of pollution and the benefits of environmental protection are regularly uneven (a principal case of many Southern governments in international environmental agreements). In these ways, the field of 'just sustainabilities' is considered to be creating new spaces for both environmentalism and for sustainable development.

Development Thinking

Development is often discussed in relation to 'developing countries', but is a concept which relates to all parts of the world at every level, from the individual to global transformation. Whilst ideas about the best means to achieve human aspirations are potentially as old as human civilization, the study of development and formal development planning originated after World War II. It has been suggested that development studies 'barely' made it to the twenty first century, whilst the object of study made it rather less effortlessly. The following section traces some of the principle changes in thinking regarding the ideologies and strategies of development over this history and identifies some of the ways in which ideas of sustainable development have impacted on these through to the current period. Although these shifts are considered chronologically, existing theories are rarely totally replaced, rather that new ones find relative favor and contestation over the prescriptions are a continuous feature.

From the late 1950s, development thinking was firmly centered on the potential of economic growth and the application of modern scientific and technical knowledge as the route to prosperity in the less developed countries. In short, the global development problem was conceived as one in which these nations needed to enter the modern age of capitalism and liberal democracy and to 'catch up' with the West. A modernization thesis dominated mainstream theories of development to the early 1970s within which development was modeled as a series of stages along an unconstrained linear path. Modernization was equated with the characteristics of westernization (in terms of economic prosperity, but also in terms of aspects of society, culture, and politics). It was an optimistic time in which underdevelopment could be overcome through imitative processes and the transfer of finance, technology, and experience from the West to less developed countries and from urban centers to rural areas. All countries were considered to have an equal chance of development. It was an era in which the contribution of neoclassical economics was to the fore, in which understandings of development were largely confined to the history and experience within Europe, and when an almost unquestioned faith in urban based industrial growth as the driver for development persisted. The discourse of development at this time was largely separate from that of the environment; the constraints of development were seen as internal to those countries in terms of an insufficiently developed industrial and commercial base and insufficient levels of investment and national savings. Strengthening the material basis of society was the key to becoming mature, developed economies and societies. Too little industrialization, rather than too much, was the dominant message coming from development thinking at this time.

Into the 1980s, the optimism of a speedy end to underdevelopment faded with rising levels of debt and the problems for oil importing countries in the context of the oil crisis. Rising economic inequalities and rural–urban differences (rather than any understanding of environmental impacts of development) led to growing dissatisfaction with ideas of development as modernization. During the 1970s, a radical critique of mainstream development thinking emerged (particularly through scholars from regions strongly linked to the US such as Latin America and the Caribbean) in which politics were to the fore. The 'dependency school' forwarded explanations of underdevelopment (framed in terms of the socioeconomic structures as well as the economic conditions of these countries) as outcomes of the exploitative/dependent relations with other parts of the world. The root cause of underdevelopment was modeled as the structural disadvantages external to underdeveloped countries and regions and the processes of colonialism in the past and of the capitalist economy generally that linked the periphery with the core. The radical critique found wider support within Europe at this time, where there was a reinvigorated interest in the work of Marx and an emerging 'New Left' movement that linked with the struggles of the Third World anticolonial movements. In consequence, rather than seeing the US and Europe as the source of solutions, dependency theorists saw the role of these regions as actively creating the problems of underdevelopment.

Despite some core differences between modernization and dependency theories, both encompass a common notion of linear progress and a shared belief in the role of the state to realize that progress (although they disagreed on the nature of that role). From the late 1970s, however, greater attention started to be given to how development should occur rather than with theorizing social change. Dependency theory moved out of fashion as a broad sweep of changes in thinking regarding the meaning of development and how best to achieve it emerged (commonly captured under the umbrella term of 'another' or 'alternative' development). Whilst economic growth remained important within development ideas, phrases such as 'growth with equity' emerged and encapsulated the recognition that it was critical to ensure that the benefits do not fall solely to a minority of the population. Furthermore, development itself became conceived as a multidimensional concept encapsulating widespread improvements in the social as well as the material well being of all in society. In turn, the strategies forwarded for achieving development became diverse and multiple rather than single and top down and were considered to require investment in all sectors, including agriculture as well as industry. It was asserted that development needed to be closely related to the specific local, historical, sociocultural, and institutional conditions, focused on mobilizing internal natural and human resources, appropriate technologies, and giving priority to basic needs. Rural based strategies for development were particularly important amongst those promoting 'development from below'.

In clear contrast to development thinking to that time, development was to be more inclusive, with individual and cooperative actions and enterprises becoming the central means for development rather than the state. Strong notions of participatory development emerged in recognition of the shortcomings of top down, externally imposed, and expert oriented research and development practice. It was understood that development needed to be sustainable (in encompassing not only economic and social activities, but also those related to population, the use of natural resources, and the resulting impacts on the environment) and a consensus emerged surrounding the characteristics of interventions that were more likely to be sustainable. These challenged both academics and practitioners to make a number of 'reversals' in their work including; putting people's priorities first, combining the strengths of both indigenous and scientific knowledge, and moving from a blueprint to a learning process approach to planning. The centrality of security of resource rights and tenure were exposed as were the benefits of locally based, smaller scale initiatives and the capacities of NGOs to foster these orientations in development.

However, the experience through the 1980s of many developing countries (with the exception of the 'Asian Tiger' economies) was of previous gains being lost and in many cases being reversed. By the mid 1980s, the sister institutions of the WB and the International Monetary Fund (IMF) were concerned about the threat of the severe balance of payment difficulties being experienced by many developing nations to the international financial system as a whole. Comprehensive, long term solutions were considered to be required to address the debt crisis, based on packages of broad policy reforms known as structural adjustment programs (SAPs). SAPs became the requirement for lending from these institutions and increasingly became conditions for access to bilateral assistance and private investment, such that SAPs increasingly defined the entry of many developing nations into the global economy. Whilst each package in theory was tailored for the particular country, SAPs generally included many common elements as shown in **Figure 8** and reflected strongly the ideals of neoliberalism that emerged as the predominant idea in defining development policy and practice through the 1990s. From the late 1980s, dissatisfaction with the record of state involvement in the economy and social life more broadly had taken root within the governments and policies of the North. Neoliberalism is an approach to development that heralds the free market as the best way to initiate and sustain economic development, such that typical policy responses involve removing the influence of the state in the domestic economy and external markets as shown in **Figure 8**.

By the mid 1990s, private capital flows to the developing world exceeded official aid and markets appeared triumphant. However, the 1997–98 Asian crisis revealed how rapidly such capital could be removed and progress dismantled. Debt burdens rose rather than fell through the decade and whilst modifications to the adjustment paradigm were made in response to the evidence of widening socioeconomic differentiation and environmental degradation under reform packages, there was little challenge made to the fundamental, neoliberal rationale. By the end of the decade, rising dissatisfaction with conventional models of development was increasingly being articulated on the ground through popular struggle and the activities of new social movements. Protest within countries of the South coalesced around the combined issues of the failures of the state and the market to deliver prosperity or well being, around prominent environmental problems, and the hardships created by the debt crisis (both its impacts and those of the solutions designed to resolve it). The legitimacy of the International Financial Institutions was also questioned by mass demonstrations on the streets of Seattle and Davos, for example, around meetings of the World Trade Organization and the G8 Finance Ministers. Questions over the utility of existing models for development were also raised by the collapse of communism that undermined the strength of Marxist analyses and the 'postmodern' critique within the social sciences more widely was challenging fundamental notions of modernity. In addition, the rise of globalization was changing the position of the nation state and national governments across economic, social, and political spheres. A number of 'post' and 'anti' development versions of development thinking emerged in response to these varied concerns; in short, questioning the whole discourse of development for the way in which it served Eurocentric interests. A post development era depended on breaking the 'holds of westernization' be it as organized by the aid industry or activities of western private capital and 'defending the local' (through ecological, women's, and peoples organizations) against the forces of globalization. Heated debates over policy also emerged, including within the IFIs, where prominent figures conceded that neoliberal reforms were failing. At the turn of the century, the arena of development as a discipline, as an institutional practice, and as a popular struggle was considered to be in substantial and pervasive ferment.

The Poverty Consensus: Current Engagement of Development Policy with Local and Global Environments

It is evident through the review above, that ideas concerning sustainable development have been a significant

Internal policy reforms designed to increase the role of the market in the domestic economy including:

- Privatization of public enterprises.
- Reduction and removal of state subsidies.
- Reducing government workforce and public spending.
- Improvements to tax system.
- Removal of wage controls.

External policy reforms designed to raise exports and increase foreign investment including:

- Currency devaluation.
- Trade liberalization through removal/reduction of tariffs and quotas.
- Remove state control of exports.

Figure 8 The characteristics of structural adjustment programs.

factor towards integrating the discourses of modern environmentalism and of development studies. By the start of the twenty first century, both arenas had been shaped by the restructuring of global capitalism, by the emergence of unprecedented environmental changes (including those associated with maturing TNCs and also with collapse of communism), and by the challenges of new social movements and civil society (nationally and globally) looking to impose regulation over governments, business, and international institutions. It is considered to be a time where the costs and ambiguities of any development intervention are now being more readily acknowledged and a context within which more meaningful and inclusive debates around sustainable development can occur.

Whilst poverty has been a familiar word in development since the start of formal development planning in the 1950s, there is an acknowledged consensus currently around poverty reduction spanning multilateral agencies, government, and development NGOs that is unprecedented and different. This is attributed in particular to the move of the WB to a Poverty Reduction Agenda in the late 1990s and the centrality of poverty across the MDGs (**Figure 4**). There is also widespread agreement around how to operationalize poverty reduction strategies at a country level (via Poverty Reduction Strategy Papers, PRSPs) and in relation to new mechanisms for development assistance based on partnerships between donor and host country to improve the management and delivery of aid.

Critics question whether the current poverty focus is fundamentally changing the practice of development or constitutes a real shift beyond neoliberalism. Elements for concern include how the macroeconomic prescriptions remain centered on rolling back the state and engaging with the globalized economy as the solution to poverty. Furthermore, the prescribed solutions are suggested to remain focused on government and society of the South as responsible for the causes and outcomes of poverty (and view external factors such as trade, international finance, or donor activities as fundamental to overcoming poverty rather than part of the dynamics of poverty creation). This arena of new partnerships in development is also under critical review including for the way in which budgets continue to be set by criteria (and audited) by donors and the required consultation between governments and civil society in practice often means with select members of professional groups. Whilst many diverse actors have bought into the MDG consensus of international development as a measurable, monitorable (and moral) goal, critics point to the way MDGs imply rather than direct necessary policy change. The MDGs are noted as stronger on the material aspects of poverty and deprivation than regarding the nonmaterial aspects of freedom such as empowerment and citizenship that are identified as important ends as well as means within development. The following sections consider the current poverty consensus explicitly in terms of its engagement with the environment as evidenced within the framework of MDGs and within PRSPs as instruments in development.

MDG7 (**Figure 4**) commits nations to take action to ensure environmental sustainability and the specific targets and indicators developed in relation to this goal are shown in **Figure 9**. It has been suggested that, to date, the MDGs contribute little in enhancing the complex links between poverty reduction and environmental sustainability or the role of ecosystem integrity in supporting pro poor growth.

Target 9 is seen to refer to country policies, but the indicators accompanying this target track largely issues of global concern with only one (no. 29) being directly relevant to the way in which the poor use the environment. Whilst other indicators under this target are issues of importance to the poor, they do not assess the ability of the poor to access key ecosystems as a source of livelihood, to protect those ecosystems from outside interests nor assess those resources in terms of a route out of poverty. Whilst the importance of country policies is accepted widely, there is a concern that such a focus can detract from more challenging questions such as those regarding global and corporate dimensions to resource use currently and in future that are being shown by recent research to be most important in shaping sustainability. For example, a 'new scramble for Africa' on behalf of transnational corporations is becoming evident as the continent becomes the source of 'non Islamic oil' and other strategically important minerals, the current exploitation of which is shown to be a primary source of conflict and instability within this region and to be failing to meet the development needs of the majority of people within African countries. There is no recognition of corporate co responsibility in such failings of African governance. Whilst MDG8 concerns global partnerships for development and refers to transnational issues, there is no mention of the environment. Whilst mounting research exposes the impacts of global climate changes and the processes of northern industrialization and consumption on local environments of the South, there is little evidence that such research is impacting on the policy framework of the MDGs.

Targets 10 and 11 can be considered to be more directly relevant to the interests of the poor, but there is still no consideration of the wider issues determining the access to resources by the poor (in the physical sense nor in terms of information or their participation in environmental decision making). For example, in relation to numbers of people with access to basic water supply and sanitation, there is no consideration of issues such as the reliability of service or pricing in relation to income that are governance

Integrate the principles of sustainable development into country policies and programs and reverse the loss of environmental resources.	25. Proportion of land area covered by forests. 26. Ratio of area protected to maintain biological diversity to surface area. 27. Energy use per 1 GDP. 28. Carbon dioxide emissions (per capita) and consumption of ozone-depleting chloroflurocarbons. 29. Proportion of population using solid fuels.
Halve, by 2015, the proportion of people without sustainable access to safe drinking water and sanitation.	30. Proportion of population with sustainable access to an improved water source(urban and rural). 31. Proportion of population with access to improved sanitation.
Have achieved, by 2020, a significant improvement in the lives of least 100 million slum dwellers.	32. Proportion of households with access to secure tenure.

Figure 9 MDG 7: Global targets and indicators.

Biodiversity	Encourage by 2010 the application of the ecosystem approach. Establish representative marine protected area networks by 2012. Achieve by 2010 a significant reduction in the current rate of loss of biodiversity.
Fisheries	Maintain or restore fish stocks to a level that can produce a sustainable yield by 2015.
Water	Develop integrated water resources management and water efficiency plans by 2005.
Chemical pollution	By 2020, minimize significant adverse effects on human health and the environment associated with the production and use of toxic chemicals, via use of transparent, science-based risk assessment and risk management procedures, and taking account of the precautionary principle.

Figure 10 Additional targets for environmental sustainability (United Nations, 2002).

issues that directly impact on access of the poor to these resources. Indeed there is no MDG relating to governance.

At the World Summit on Sustainable Development, additional targets relating to environmental sustainability were created (**Figure 10**). In the commitment to apply an ecosystem approach in biodiversity conservation and in restoring fish stocks to levels that can support sustainable yields, these targets are steps towards the incorporation of greater pro poor elements related to sustainable de velopment. Encouragement was also given at WSSD to countries to modify MDG7 to suit local conditions and to establish country specific targets and indicators, which is further source for optimism concerning engagement with local environments.

In 1999, the IMF announced that PRSPs would re place SAPs as the strategic documents around which future loans and decisions over debt relief would be made and are administered jointly with the WB. The core principles of the approach are shown in **Figure 11**. PRSPs are written by national governments, required to set out coherent plans for reform focused on poverty reduction and identify financing needs. Governments are also 'encouraged' to consider environmental factors in their PRSPs.

A number of assessments have revealed serious shortcomings in relation to the mainstreaming of en vironmental issues into PRSPs, even suggesting the need for an 'environmental overhaul' if the central role of ecosystems in the lives of the poor and their potential to reduce rural poverty is to be secured. To date, environ mental health targets such as safe drinking water and sanitation have been found to receive more attention than issues of natural resource management as shown in **Figure 12**.

A study by the Environment Department of the WB itself found that many of the PRSPs that were

PRSPs should involve broad-based participation by civil society and the private sector at all stages, including formulation, implementation, and outcome-based monitoring.
PRSPs should focus on outcomes that will benefit the poor.
: PRSPs should recognize the multidimensional nature of poverty and the scope of actions needed to effectively reduce poverty.
PRSPs should involve the coordinated participation of development partners, including bilateral and multilateral agencies and nongovernmental organizations.
PRSPs should recognize that sustained poverty reduction will require action over the medium and long terms as well as in the short run.

Figure 11 Core principles of the PRSP approach. Compiled from World Resources Institute, (2005). *World Resources Institute 2005: The wealth of the poor.* Washington, DC: World Resources Institute.

A study of forest-related issues in 36 PRSPs found treatment of forest issues to be weak, particularly in their analysis of the links between poverty and forest resources and the role of natural resources and ecosystems in shaping well-being. Despite this, many went on to set forest-related policies and programs in their agendas for action.
An assessment of biodiversity-related themes within 15 PRSPs revealed that the majority analyzed declines in biodiversity, only one developed a policy prescription that integrated biodiversity conservation and poverty reduction.
A study of 10 PRSPs revealed inadequate and inconsistent incorporation of water issues. In particular, the need for close links between strategies for delivering additional water supply and sanitation infrastructures and those for managing water supplies for productive use by the poor were not recognized.

Figure 12 Issues of natural resource management within PRSPs. Compiled from World Resources Institute (2005). *World Resources Institute 2005: The wealth of the poor.* Washington, DC: World Resources Institute.

particularly weak in their attention to environmental issues in fact came from countries where poverty–environment linkages were strongest (in terms of heavy dependence on natural resources for rural livelihoods, dependence on traditional fuels, and low levels of basic sanitation). Few PRSPs made any reference to the indicators of MDG 7 (other than access to safe water and adequate sanitation). The degree of mainstreaming of the environment generally within PRSPs was found to be influenced by the nature and degree of civil society participation in their preparation that in turn was driven more by donor concerns rather than political priorities.

In addition to these weaknesses in terms of mainstreaming of the environment, analyses of poverty within PRSPs have revealed an emphasis on technical solutions to poverty–environment issues rather than for the more politically challenging issues of access, control, and rights to resources and how these impact on the capacity of the poor to secure sustainable livelihoods. Many PRSPs to date, fail to consider the impact on sustainability of their proposed growth strategies; incentives for high input, export agriculture still dominate, for example, but fail to analyze the impact on natural resource depletion, on waste generation, or on small scale farmers and their ability to manage local resources. In short, PRSPs currently give insufficient attention to the centrality of ecosystems in the lives of poor people, to the need to enhance the ability of the poor to govern these and indeed, to the negative impacts of wider economic growth strategies on the livelihoods of the poor as well as ecosystems.

Evidently, whilst there are positives emerging from the current poverty consensus, there are shortcomings in terms of the degree of mainstreaming of the environment within both the framework of MDGs and the strategies that are considered central to delivering poverty reduction. Optimistic signs for the future include the way that the MDG approach gives civil society organizations opportunities to monitor the activities of governments, to force further accountability for inadequacies, and to assert more effective pressure. Similarly, optimism can be drawn from the evidence that some of the best examples of environmental mainstreaming within PRSPs were found in those where civil society engagement had been most extensive (although concerns remain as to how far such examples were driven by the donor community). To date, however, it is suggested that the conception of the poverty–environment nexus remains the ‘mainstream view', that is, of the poor caught in a vicious cycle with environmental degradation as both victims and continued agents. There remains little nuanced understanding of the possibility of the poor pursuing livelihood options that are sustainable or for the arenas and practices of international aid or trade, for example, in creating and perpetuating poverty. There is a danger that attention has been diverted away from global and macro level policies and practices that influence global poverty trends and are the context within which poor people operate. Whilst the obstacles the poor face in terms of environ ment and development often manifest themselves at the local level, it is in these wider arenas that many suggest the roots of their problems lie. The suggestion is that whilst 'alternatives' in development theory may have seemingly been incorporated into the mainstream as identified here within development policy instruments of international organizations and the donor comm unity, they have been distorted in the process of such incorporation.

Geography and the Poverty–Environment Nexus

It has been suggested that sustainable development is not only an important concept for the future of the world, but also for developments in geography. Given that many of the issues central to contemporary sustainable develop ment debates have a long standing history of study within the discipline, many point to the substantial knowledge as well as empirical and theoretical tools that exist within the community to ensure that geographers go on making a contribution to a positive future for life on Earth. As scientific understandings of environmental change develop, for example, geographers are engaging in the required detailed field studies and assessments, in highlighting the limitations of science and exposing the social construction of the meanings of resource use and degradation. Geographers are also fostering under standings of the spatial, temporal, and social nature of outcomes on the ground of conservation and develop ment interventions for people, cultures, and ecologies. They are also well placed to expose the ideologies underpinning the moves towards (and interests within) the mounting and multiple ways in which sustainable development is currently being taken up. In having context and place at the heart of their tradition, geog raphers are also considered to have the potential to overcome the suggested formulaic and parochial under standings of how to foster sustainable development and to fill the gaps between theory and practice of sustain ability in the future.

The sections above have revealed the nexus of poverty and environment interactions that are now central to understandings of sustainable development as a notion and in terms of policy and practice. Geographers working in the field of political ecology can be considered to be particularly active in the critical investigation of the complex realities of poverty and environmental degrad ation and the integrated scales and processes of political, economic, and environmental development that under pin the varied outcomes experienced by particular peo ple and places. Political ecology as an area of geography developed rapidly in the 1980s, particularly as a response to the understood apolitical nature of mainstream en vironmental literature at that time. For example, in ex posing the political and economic interests and conflicts that lie beneath 'proximate' causes of environmental problems such as poverty or population growth, they have been part of the challenge to mainstream under standings evident in this review. However, whilst early work in political ecology was rooted in a critique of policy failures such as in conservation and in hazard management, it has been suggested that political ecolo gists have subsequently been better at critiquing dom inant narratives than in generating counter narratives. They are also considered to be falling short currently in engaging with and influencing policy. It has been sug gested that work in political ecology has overemphasized the local case study that can be charged as anecdotal and case specific such that further comparative and broad scale study is required. As development geographers more widely have come to emphasize local outcomes and initiatives, there is a need for close consideration of scale in each case to avoid assumptions that the local neces sarily means more sustainable. Most importantly, greater intradisciplinary work, including between geographers who work in the North and in the South can be con sidered essential for understanding and engaging in the challenge of sustainable development.

See also: Aid; Development I; Development II; First World; Neoliberalism and Development; Nongovernmental Organizations; Postcolonialism/ Postcolonial Geographies; Postdevelopment; Poverty; Second World; Third World.

Further Reading

Adams, W. M. (2001). *Green development: Environment and sustainability in the Third World* (2nd edn.). Abingdon: Routledge.
Adams, W. M. (2002). Sustainable development? In Johnston, R. J., Taylor, P. J. & Watts, M. J. (eds.) *Geographies of global change: Remapping the world*, pp 412 426. New York: Wiley.
Agyeman, J. (2005). Where justice and sustainability meet. *Environment* 47(6), 10 23.
Barbier, E. B. (1987). The concept of sustainable economic development. *Environmental Conservation* 14(2), 101 110.
Black, R. and White, H. (eds.) (2004). *Targeting development: Critical perspectives on the millennium development goals*. Abingdon: Routledge.

Chambers, R. (1983). *Rural development: Putting the last first*. London: Longman.

Cornwall, A. and Brock, K. (2005). What do buzzwords do for development policy? A critical look at 'participation', 'empowerment' and 'poverty reduction'. *Third World Quarterly* 26(7), 1043 1060.

Elliott, J. A. (1996). *Introduction to sustainable development*. Abingdon: Routledge.

Meadows, D. H., Meadows, D. L., Randers, J. and Behrens, W. W., III. (1972). *The limits to growth*. New York: Universe Books.

O'Riordan, T. (1996). Environmentalism on the move. In Douglas, I., Huggett, R. & Perkins, C. (eds.) *Companion encyclopedia of geography*, pp 449 478. Abingdon: Routledge.

Redclift, M. (2005). Sustainable development (1987 2005): An oxymoron comes of age. *Sustainable Development* 13, 212 227.

Schuurman, F. J. (2000). Paradigms lost, paradigms regained? Development studies in the twenty first century. *Third World Quarterly* 21(1), 7 20.

Simon, D. (2003). Dilemmas of development and the environment in a globalising world: Theory, policy and praxis. *Progress in Development Studies* 3(1), 5 41.

United Nations (2002). Johannesburg Plan of Implementation, http://www.un.org/esa/sustdev/documents/WSSD POI PD/English/POIToc.htm (accessed on Nov 2007).

World Commission on Environment and Development (1987). *Our common future*. Oxford: Oxford University Press.

World Resources Institute (2005). *World resources, 2005: The wealth of the poor*. Washington, DC: World Resources Institute.

Relevant Websites

http://www.un.org/esa

Home page of the Division for Sustainable Development within the Department of Economic and Social Affairs at the United Nations. The Division is responsible for providing leadership and expertise on sustainable development within the UN. In addition to facilitating coordination within the UN, it supports the work of the Commission on Sustainable Development in facilitating intergovernmental negotiations, for example, and also provides technical assistance and advice to particular countries.

http://www.iied.org

Website of the International Institute for Environment and Development a research institute and NGO working for more sustainable and equitable processes of global development. Based in London but has partners throughout the world. Rooted in a commitment to partnership working and providing a voice for vulnerable groups. A good source for recent research, project findings, resources to support practice and links to events and networks.

http://www.worldbank.org

Web site of the World Bank Group that provides access to a range of research publications, policy and project documentation, data sets, country information and recent initiatives. Can access PRSPs for particular countries and can examine data concerning progress towards MDGs, for example, through this site.

http://www.wri.org

Web site of the World Resources Institute, a Washington based research and development organization committed to sustaining ecosystems for human needs, enabling public access to information regarding decisions over natural resources and to expanding economic opportunity whilst conserving environments. Hosts a major data base 'EarthTrends' tracking environmental, social, and economic changes as well as widely cited publications such as the World Resources volumes.

Symbolic Interactionism

V. J. Del Casino, Jr. and D. Thien, California State University, Long Beach, Long Beach, CA, USA

Glossary

Ethnography The study of cultural and social practice in everyday context with particular attention to the broader cultural and social meanings human individuals and social groups develop over time and space.

Functionalism This body of theory challenges behavioral-based social theories, arguing that psychologically each individual's psyche is invested with certain cultural and social traits that are essential to human interaction.

Nonrepresentational Theory This much debated philosophical framework seeks to extend 'beyond' representation and values knowledges as they are practiced or performed, instead of interpreted or interpretable.

Participant Observation This methodological technique suggests that researchers must participate in the mundane, day-to-day interactions among a social group or people to understand the cultural and social processes at work in a particular time and space.

Post-Structuralism A theoretical framework that developed out of a critique of structuralism, suggesting that linguistic practices, discourses, and texts have no essential characteristic but are constructed temporarily through the deployment of power.

Pragmatism This philosophical framework proposes people act on the basis of the knowledge they have at hand – that which is most useful. The intellectual focus, on practical application of linguistic conventions as opposed to deeply embedded structures. This approach also suggests that theories must be pluralistic and relevant to the particular context in which various ideas are employed and developed.

Social Identity A term that is often attached to a number of social categories (e.g., gender, sexuality, age, race, ethnicity, nationality, etc.), which are themselves based in the construction of broader sociocultural and political–economic processes.

Social Role This concept, often associated with structuralist thinking, suggests that certain sets of practices and behaviors (i.e., social roles) are regulated in and through particular social situations.

Structuralism A theoretical framework that developed in the study of linguistics and cultural anthropology, which suggests that individual linguistic practices (signifiers) represent the deeper, underlying sociocultural and political–economic contexts and processes (signified) in which they are produced.

Structuration Theory A social theory that proposes the intertwining of behavior and structure, as opposed to the predominance of either structure or individual agency.

Introduction

The development of symbolic interactionism in Western social theory is informed by a general interest in the investigation of subjective experience arising, in part, from a critique of structural functionalism. The term was coined by US sociologist, Herbert Blumer who – following the theories of his mentor, George Herbert Meade – argued that common understanding was developed through the mundane acts of daily interaction. More specifically, Meade methodologically concentrated on language and talk between individuals. Symbolic interactionism thus entails the study of face to face interactions to investigate how social meanings and a common sense of belonging emerge in and as part of societies.

Contrary to macro theoretical approaches to the study of 'the social' and 'society', which investigate social systems and processes, symbolic interactionism concentrates on the micro scale of interactions among and between social actors. In many ways, early interactionist theorists challenged the assumptions of Marxist and other structuralist thinkers, who chose to focus their attention on how capitalism, for example, mediates all social relations. Put simply, through macro theoretical approaches Marxists argue that relations of capitalism structure social interactions; collective meaning is constituted through the larger system or process at work in society at a given time and place. In turning attention toward subjective experience, symbolic interactionists seek to investigate collective social meaning from the ground up, suggesting that societal meanings are constructed and reconstructed through practice. At the same time, critical to the development of a social interactionist model is the idea that humans do not simply react to other's prompts. Rather, individuals are pragmatic actors who proactively employ social and cultural symbols in their interactions. Unlike a strict structural Marxist perspective, power is not theorized as domination; rather, symbolic interactionist philosophy is premised on the notion that humans have agency in the construction of their symbolic culture and society.

Even though symbolic interactionism appears to privilege the individual over society in its logic, the intent is actually to consider the individual and society as mutually constitutive: all social action is interactive between individuals and thus it must be thought of as complexly intertwined with and productive of collective meaning. Within the field of sociology, various scholars engaged with these notions of interactionism to develop methodological frameworks for investigating urban life (i.e., the Chicago School) or social roles (i.e., Erving Goffman). In the case of urban studies, scholars em ployed ethnographic methodological approaches to examine communication in and between various social groups. In the case of role theory, Erving Goffman suggested that through detailed observation it is possible to investigate how social roles are developed inter actively. In constructing these roles, people employ various symbolic meanings that are mutually 'accepted', as they perform and present their 'self' in relation to other 'selves'.

In the context of twenty first century social theory, it is less common for scholars to explicitly claim 'allegiance' to symbolic interactionism. At the same time, the broader tenets of this philosophical approach can be commonly found at work in a number of fields and disciplines. In geography, questions of how symbolic meanings of place are constructed are increasingly important in critical human geography. Geographers are also very much concerned with the mundane and profane, as well as the daily interactions at work in the constitution of both social space and identity. Moreover, geographers remain interested in how theories of both 'the symbolic' and 'social interaction' inform understandings of not only human relations but also human–nonhuman relations. In the case of the nonhuman, geographers are theorizing interactions between objects from the so called natural environment, as well as human products and artifacts, such as maps or buildings.

Symbolic Geographies

The development of human geography throughout the late nineteenth and early twentieth centuries established the discipline's focus on the material landscape. Cultural and social geographers, for example, traced the diffusion and distribution of the artifacts of social and cultural systems across local and regional spaces. Early on, how ever, some geographers also insisted that the study of cultural and social worlds demanded attention not only to material artifacts but immaterial processes – for ex ample, religious, linguistic, ethnic – as well. These geographers suggested that psychological geographies, how people imagine the world, were critical to how they might understand cultural and social worlds. Although these geographers did not necessarily engage with the conceptual or theoretical language that now facilitates geography's study of the immaterial world, they certainly understood that these processes were important.

Thus, as the discipline of geography developed an expanded repertoire of methodological tools to investi gate the spatial organization of the material social world, there also emerged a number of new approaches to cope with the complexity of immaterial geographies of meaning and feeling. The conceptual development of *genre de vie* (or way of life) in French human geography provided geographers with a conceptual basis to in vestigate not only human landscapes but also human social behavior in place. As part of this move, geog raphers developed a sense of how places produced a unique set of meanings and feelings. Each area or region was (and is) unique because of the ways that social interactions take place in and through them.

Other geographers, particularly Marxists, considered broader social processes and 'structures' that regulated social life and human behavior. Importantly, these geographers insisted that they needed to study how various systems of power, economic or social, construct spaces of control. It is here that geographers turned their attention, to the power and importance of cultural symbols, and particularly the 'symbolic landscape'. Central to this study of 'the symbolic' in Marxist geog raphy has been an analysis of class relations and the power of the elite to construct spaces that structured how and where people move on a day to day basis. In this rendition of the landscape, the capitalist economy is the conceptual lens through which an understanding of symbolic meaning and space is theorized. Turning at tention to symbolic meaning means that geographers have developed methodological techniques and concepts to examine this new object. The symbolic landscape has thus been conceptualized as a space that can be 'read' like other cultural texts for the subjective meanings invested in them.

Importantly, humanistic and feminist geographers struggled with the centrality of purely economic models of power. Humanistic geographers, on the one hand, thought of the symbolic not simply as a system of power and authority wielded by elites. Instead, these geog raphers suggested that the symbolic meaning of places are tied to the mundane, everyday feelings that people ascribe to their 'place in the world'. Symbols – and more broadly representations – are a critical aspect of understanding how we, as humans, understand the rela tionships between our 'feelings' and places. As such, cultural symbols help us both represent our 'sense of place' and make meaning of it. Feminist geographers, on the other hand, remained concerned with broader issues of social power and authority but argued that purely economic theories failed to explain the symbolic nature of that power, a result not only of the economy but also of patriarchal social systems. Feminist geographers are, therefore, also critical of both humanistic and cultural theories because they privilege masculinist 'ways of seeing' the symbolic landscape; they argue that geographers have, in the past, failed to examine the importance of the symbolic power of gendered identities and subjectivities.

These debates about the symbolic landscape in geography complicated the discipline's understanding of the relationship between social identities and spaces. This paralleled the entry of 'critical race' and 'queer' theories into geography, as well as post structuralist theories of the symbolic all of which began to deconstruct the relationship between class, gender, and space, and other relationships between age, development, ethnicity, nationalism, race, and sexuality. In a post structuralist reading of 'the symbolic' spaces are conceptualized as relational, constructed through a web of intertextual power relations that privilege certain identities over others. Spatial representations are thus important empirical nodal points in a post struturalist analysis. However, unlike broader structuralist or humanist theorizations of the symbolic, the notion that a symbolic system is somehow underpinned by an essential character, such as sense of place, or by a deep underlying structure, such as capitalism or patriarchy, is largely suspect. Instead, post structuralist geographers have examined how symbolic spaces and meanings may be temporarily fixed through relations of power and authority, giving the appearance of a 'real' world that structures everyday social interactions. Such symbolic spaces are given meaning and can be 'sutured' to particular identities in ways that mediate daily social interactions. In this way, our symbolic geographies become 'real' and quite powerful as both a process of control, as well as a site of resistance, both overt and mundane.

Over the course of more than 100 years of academic geography, geographers have developed a conceptual language for studying the symbolic, and cultural and social symbols, particularly through their study of various representations of power and authority. The turn toward 'the symbolic' has suggested that purely objective studies fail to account for how meanings are ascribed to various places and spaces or to the dynamics of power that might structure the symbolic world, even temporarily. While the 'vantage point' of humanists, Marxists, feminists, queer, or post structuralist theorists in geography might be different, all suggest a focus on subjective experience and human interaction.

Interactive Geographies

As the interest in symbolic geographies has become more complex so has the theoretical and conceptual language that geographers have used to understand social interactions and how humans define and comprehend their 'place in the world'. Geographers have become particularly interested in the interrelationship between place and community, space and identity. In the case of the study of identity, geographers investigate how social actors use various symbolic systems and spaces in the construction of their identities, reciprocally investigating how human interactions with particular spaces also help frame social identities and practices. Geographers have thus explicitly turned their attention to human interaction, the mundane acts of engagement social actors 'go through' on a day to day basis as they negotiate the city, the country, or their home. To do this, geographers have needed a number of new theoretical and methodological frameworks to examine the geographies of human interactions, which have included Giddens' structuration theory, Husserl's phenomenology, and also the interpretative sociology of Erving Goffman and symbolic interactionism more generally.

To understand how people perform socially in the various regions of everyday life, sociologist, Erving Goffman sketched famously detailed observations of face to face encounters, carefully considering the spaces of these interactions. In his dramaturgical approach, people variously assume social roles as performers, audience members, or outsiders in an ongoing drama devoted to projecting and receiving particular information regulated by rules socially determined through practice. The spaces where these activities of daily life take place are critical: the 'front', 'back stage', or 'outside' regions. As with a stage set, the front is where social performances take place. This spatial designation is not determined by space alone. For example, the front is not only the dining room to a back stage of a restaurant kitchen, but may also be determined in symbolic ways, such as by appearance and manner of the occupants, including such fleeting events as facial expressions. Through the performance of social roles, individuals manage the front stage through both the material production of places and through the symbolic organization of cultural meanings. The symbolic nature of social interactions is translated through geography's sensitivity for the complex relationship between symbolic and material space; it is possible to trace how spaces are constituted to support certain sets of symbolic practices.

More than simply suggesting that space was divided between front and back, however, Goffman also argued that different actors manage different spaces. The distance between performer and audience is a managed space that enables people to limit and regulate what is available for perception by others; that is, both performers and audience can control their use of information in particular spaces. These spatial strategies include managing discrepancies of interaction – appearing out of one's social 'role' or showing the wrong 'face', for example – to avoid disruption and eliminate uncomfortable emotions such as embarrassment, shame, or the discrediting of an individual's self conceptions. Goffman's notion of 'civil inattention', the process of acknowledging then avoiding the presence of others as tactic to preserve privacy, is well known. The potential for unintentional disruptions (incorrectly performing, or presenting in the wrong space) serves to highlight both the interactive and symbolic aspects involved in the expression of the self within an everyday social milieu. Goffman's work is rich with insights into human encounters.

Humanistic and posthumanistic geographers, with their attention to the symbolic nature of being 'in place' and 'out of place', have resonance with this form of symbolic interactionism. In the 1970s, for example, David Ley suggested that geographers pay close attention to the intersubjective nature of place meaning, arguing that they consider the mundane aspects of everyday life as essential to their understanding of the place meaning relationship. Although his contributions to the discipline do not draw directly from symbolic interactionism, he pushed geographers to consider the theoretical and methodological value associated with the study of social interaction in their studies of place, validating the concerns of social and cultural geographers more broadly that wanted to engage in the study of subjective experience. Even more generally, the 'arrival' of symbolic interactionism and other theories of subjective experience in the context of human geographic theory and practice paralleled the growing interest in human agency, which has been seen as a corrective to the overly structuralist geographies that failed to account for the subtlety, nuance, and difference found in everyday social interaction. In the context of rural geography, Sarah Harper has suggested that positivist and structuralist approaches fail to appreciate the subjective relationships that emerge between people and their place in the world. Turning her attention directly to symbolic interactionism, she has argued that through the use of participant observation it is possible to better understand the complexity of everyday life in rural spaces.

Within human geography, then, the emphasis on both the everyday and on the centrality of space remains a key interest, as is the notion that identity is a process of social and spatial intertwining. Moreover, Goffman's symbolic interactionism has pushed some geographers to consider just how people manage ourselves and identities in particular kinds of places, as well as in the face of social roles and rules. Phil Crang, for example, developed an extended analysis of a restaurant as a 'workplace geography' employing Goffman's body of work to examine the imagined and spatial structures of interaction which constituted this particular place. Crang's research highlights the interconnections between forms of interaction and identity politics, intersections which are also deeply infused with relations of power. Feminist geographers have remained interested also in how the everyday practices of managing gendered roles within social spaces remain regulated through the wider symbolic systems of patriarchy and sexism, for example.

Symbolic Interactionism and Contemporary Human Geography

Thus, despite certain reservations, symbolic interactionism still figures in human geography, though twenty first century geographers seldom claim a direct relationship with the label 'symbolic interactionism'. The traces can be identified in the ongoing concern with symbolic meanings of place and how these are produced; the continued emphasis on the everyday; and the search for ways to explain identity as a process of social and spatial intertwining. Methodologically, the qualitative boom of the 1990s necessitated a deeper understanding of the micro scale interactions that take place in interviews and participant observation encounters. Theoretically and empirically, many cultural, feminist, social, and urban geographers, for example, remain preoccupied with developing sociospatial accounts of human encounters in social spaces, accounts that reveal the implicit or explicit traces of a symbolic interactionist analysis. Such work investigates the extent to which people attempt to manage ourselves and identities, in particular kinds of places, both 'in' the face of or 'as' the face of social roles and rules. Erving Goffman's work, in particular, has provided a basis for both theoretical and methodological investigations of symbolic meaning as employed in the 'presentation of self'. Such investigations consider how symbolic orders and meanings are produced and reproduced, examine how bodies are presented, and assess the negotiation of emotions in social encounters.

At the same time, some geographers have become dissatisfied with the binary logics of the front and back stages and have tried to expand on Goffman's notion of 'self presentation' through alternative theories of social interaction. Turning their attention to the performative theory of Judith Butler and the deconstructive methodologies of Jacques Derrida, geographers have reconsidered the relationships that are constructed through social interactions as they occur between mutually constituted front and back stages. In this framing, social identities, and the roles that people attach to them, are constructed through relational social interactions that rely on an 'other' to make them 'real'. Front stage performances do not hold any meaning without the performances that supposedly occur back stage; there is no front and back, then, only performance, which makes it appear as if the world is structured into discreet spaces of 'inside' and 'outside', or 'us' and 'them'. Through de construction, geographers may examine how various so cial interactions put in place temporary barriers between front and back, inside and outside; the so called 'man agement' of the distance between these two spaces col lapses, however, when researchers understand how social categories and identities are performed relationally. Such deconstructive tactics also allow geographers to examine the notion of identity or the 'self' within a symbolic interactionist tradition. For example, Nicky Gregson and Gillian Rose suggest that within the 'Goffmanesque tra dition' performance is staged by knowing actors, who craft their 'play' with intent. Offering Judith Butler's theory of performativity as an alternative, they critique some geographic analyzes of presentation or 'perfor mance' for too readily taking up Goffman's concepts of agency, and thus not attending to the multiple ways in which power informs performances, an awareness that informs Butler's theory.

The keen interest in the production of selves in particular spaces, such as in Goffman's work, also pres ages the more recent interest in embodiment as the basis of relational engagement. In Goffman's attention to feelings, he also foregrounds a additional preoccupation with emotion and affect as a consequence of self space interactions. The expansion of qualitative research within geography has also inspired additional attention to face to face encounters. However, neither Goffman, nor symbolic interactionism more generally, is taken up wholesale in contemporary human geography very often. This is because symbolic interactionism is premised on the notion of a conscious actor interacting with others in spaces that can be interpreted through an individual's ability to produce and decode socially symbolic mean ing. Ontologically, this claim presupposes a social world as a product of social interaction between pragmatic actors who interpret the actions of others and behave accordingly. Information can be consciously withheld or controlled by individual social actors; that is, individuals are imbued with agency. People choose, for example, to display feelings to produce, conceal, or reveal infor mation. Thus, symbolic interactionism is much less concerned with questions of broader systems and pro cesses of power and resistance that are central to the current work of many human geographers today.

And yet, by working in the arena of symbolic spaces and interactions, human geographers have deepened the social interactionist notion of the social self, extending the work of Goffman and others, to create an explicitly sociospatial concept of identity. This has helped ground some urban geographic studies in the empirical analysis of political claims to space, such as rights to city spaces and the construction of 'local' identities. Post positivist geographers have also drawn upon and expanded sym bolic interactionism in a few useful ways. First, they consider more explicitly the relationships people have 'with' the places of their encounters (e.g., in positioning themselves as 'local' or not), as opposed to detailing their performances 'within' certain places. Second, they acknowledge more explicitly the relationships people hold within and/or of themselves in the psychic life of encounters (e.g., considering self perceptions or experiences of mental health). Finally, they critically reexamine the notion of a conscious actor in light of developed psychoanalytically inspired notions of sub jectivity, and feminist and cultural theorizations about identity including performativity, gender, and the 'other'.

In these ways, Goffman's theorizations of both self and space have been employed and critically moved forward by human geographers. For example, Joyce Davidson has worked with Goffman's ideas of management in order to understand and detail agoraphobic experiences of anxiety within social spaces. His explanation of front and back stage spaces has been extended by Deborah Thien to investigate how dominant understandings of intimacy have a particular spatial logic symbolized by the prox imity of face to face encounters. While this logic recruits people to particular roles and rules, people may con sciously or unconsciously resist and redefine their iden tities in relation to the particular mundane spaces of their daily affairs. As such, symbolic distinctions of 'front' or 'back' are collapsed or rendered unstable. Eric Laurier and Chris Philo also draw from Goffman's rich resources as inspiration for their ethnomethodological investi gations of gestures between strangers in urban cafes. Extending Goffman's ideas of impression management, they note that while gestures as part of a 'body's work' can always elude a manager's control and so thwart concealment of private intent, such bodily moves may also be an expression of shared presence. Such gestures may contribute to constituting places as sites of sharing a more positive reading than Goffman's self management analysis allows.

Geographers also remain interested in how theories of both 'the symbolic' and 'social interaction' inform understanding of not only human relations but human–nonhuman relations as well. In the case of the nonhuman, geographers are theorizing interactions between objects from the so called natural environment, as human products and artifacts, such as maps or buildings. In much of this work, symbolic interactionism, which itself began as a critique, is subject to critical revisioning by human geographers who have incorporated and extended this work in productive ways. At the same time, the (im) possibilities of symbolic representations have also been much debated, particularly within cultural geography. Nonrepresentational theory has emerged out of a struggle to escape representation, to acknowledge what is unspeakable or unwriteable. Some feminist theorists have also argued for the impossibility of representing a feminine 'other' within a masculine symbolic framework. A focus, therefore, on 'symbols' and 'the symbolic' is being rethought in various ways through human geography's interest in emotion and affect theory.

The enduring attention within human geography to the symbolism of interaction as constitutive of daily, 'located knowledge', such as expressed by Angus Laurier, Eric Whyte and Kathy Buckner, and an abiding methodological concern for assessing social and spatial identities through assessing encounters, can at least in part be traced to the ideas and methods that make up the symbolic interactionist tradition. While symbolic interactionism may now seem little more than an historical, conceptual artifact, lingering only in traces of contemporary geographers' conceptual and methodological frameworks, these traces register a continued, considered thinking about social interactions in and of space through detailed engagements with the symbolic and material aspects of geography. Such engagements extend human geography's theoretical and methodological conceptualization of what is meant by symbolic space, human social interaction, and human/nonhuman relations.

See also: Emotional Geographies; Ethnography; Ethnomethodology/Ethnomethodological Geography; Feminist Methodologies; Humanism/Humanistic Geography; Lamarck(ian)ism; Non-Representational Theory/Non-Representational Geographies; Participant Observation; Phenomenology/Phenomenological Geography; Posthumanism/Posthumanistic Geographies; Regional Development, Endogenous; Structurationist Geography; Subjectivity.

Further Reading

Blumer, H. (1969). *Symbolic Interactionism: Perspectives and Methods*. Englewood Cliffs, NJ: Prentice Hall.

Crang, P. (1994). It's showtime: On the workplace geographies of display in a restaurant in southeast England. *Environment and Planning D: Society and Space* 12, 675 704.

Davidson, J. (2003). Putting on a face: Sartre, Goffman, and agoraphobic anxiety in social space. *Environment and Planning D: Society and Space* 21, 107 122.

Duncan, J. (1978). The social construction of unreality: An interactionist approach to the tourist's cognition of environment. In Ley, D. & Samuels, M. (eds.) *Humanistic Geography: Prospects and Problems*, pp 269 282. Chicago, IL: Maaroufa Press.

Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Garden City, NY: Anchor.

Goffman, E. (1963). *Stigma: Notes on the Management of a Spoiled Identity*. Englewood Cliffs, NJ: Prentice Hall.

Gregson, N. and Rose, G. (2000). Taking Butler elsewhere: Performativities, spatialities and subjectivities. *Environment and Planning D: Society and Space* 18, 433 452.

Harper, S. (1987). A humanistic approach to the study of rural populations. *Journal of Rural Studies* 3, 309 319.

Laurier, E. and Philo, C. (2006). Possible geographies: A passing encounter in a cafe. *Area* 38, 353 363.

Laurier, E. and Philo, C. (2006). Cold shoulders and napkins handed: Gestures of responsibility. *Transactions of the Institute of British Geographers* 31, 193 207.

Laurier, E., Whyte, A. and Buckner, K. (2001). An ethnography of a cafe. *Journal of Mundane Behaviour* 2, 195 231.

Ley, D. (1977). Social geography and the taken for granted world. *Transactions of the Institute of British Geographers* 2, 498 512.

Ley, D. (1978). Social geography and social action. In Ley, D. & Samuels, M. (eds.) *Humanistic Geography: Prospects and Problems*, pp 41 57. Chicago, IL: Maaroufa Press.

Pile, S. (1993). Human agency and human geography revisited: A critique of 'new models' of the self. *Transactions of the Institute of British Geographers* 18, 122 139.

Pile, S. (1991). Practicing interpretive geography. *Transactions of the Institute of British Geographers* 16, 458 469.

Smith, S. J. (1984). Practicing humanistic geography. *Annals of the Association of American Geographers* 74, 353 374.

Thien, D. (2005). Intimate distances: Considering questions of 'us'. In Davidson, J., Smith, M. & Bondi, L. (eds.) *Emotional Geographies*, pp 191 204. Ashgate: Aldershot.

Symbolism, Iconography

M. Denil, Washington, DC, USA

Glossary

Conventional Signs Something considered able to stand for something other than itself by virtue of its place in a convention of meaning.
Depiction A form of mimesis or imitation relying upon visual resemblance of some aspect of an object.
Generalization A process wherein more detailed information is made understandable by elimination or modification of aspects of its representation while maintaining the aspects of its character considered significant.
Gestalt The resolution of the interaction of the various elements visual field.
Gestalt Psychology The study of phenomena concerning the visual field.
Graphic (Axis) The axis of a triadic model of map symbol typology tending toward nonlinguistic abstraction in representation of the object in question.
Graphic Redundancy A means of ensuring that different map symbols can be received as different by variation in two or more aspects of the symbol.
Iconography A set of specified or traditional symbolic forms associated with a subject or theme.
Interpretant A code or convention of understanding brought to a potential sign by the reader.
Interpretive Community Description of a group sharing one or more conventions of understanding.
Linguistic (Axis) The axis of a triadic model of map symbol typology tending toward specific verbal or written names for the objects represented.
Mimetic (Axis) The axis of a triadic model of map symbol typology tending toward depiction of some (often visual) aspect of the object represented.
Practics Issues concerned with the practice of theory.
Qualitative Differences in kind.
Quantitative Differences in amount or value.
Representamen The physically present trace of the sign. It represents the object through the working of the interpretant.
Representation The act of standing in for something else; another object or concept.
Semiology The science of the signs (Saussurian school).
Semiosis The process of a sign's acquisition of meaning.
Semiotics The science of the signs (Peircian school).
Signified The part of the sign which is represented.
Signifier The part of the sign which is present.
Symbolism The use of signs to represent features on a map.
Syntagam The composed statement made up of constituent signs; the map as an argument.

Theoretic Overview of Symbolism and Iconography

A map is a mediation between some situation, relationship, or condition and a user who is able to interpret that mediation. A map is a sign, constructed of constituent signs, composed to persuade the map user of its capacity to represent something other than itself. As with any sign, there is nothing inherent in the map itself which holds meaning; meaning is something that can only reside in the perceived relationship of the sign with some external concept accessible through a paradigm or system of understanding. These systems can be thought of as languages, as they utilize vocabularies, grammars, and canons of conventional usage. The potential user brings a system of their choice to the potential map and judges the potential map's fitness for recognition as a map, and, should it qualify, he or she then reads meaning into it. If the user is satisfied with the validity of the reading s/he has made, s/he declares the artifact to be a good map.

The map mediates between a territory (a situation, relationship, or condition) and a user by standing in for that territory – by representing it to that user in some limited way. The map is not the territory, but instead it persuasively suggests some aspect of the territory in a useful and usable manner. By reducing the territory to an abstraction, and presenting it as a configuration of categorized and conventionalized signs, the map allows and facilitates a broad conceptual grasp of the territory. A map is not a reconstruction of its represented territory in all specific detail, but is instead a selected, generalized, rhetorical presentation of one way of conceptualizing that territory. The map's argument hides some specificities and submerges others in generalities while preserving or making explicit still others. The abstract map symbols function in much the same way, but, obviously, not to the same extent, that algebraic symbols abstract a mathematical equation from the specific to the general.

Our understanding of how map symbols function and acquire meaning derives from the science of signs, which is variously known as semiotics or semiology. The use of two terms for this science arises from the originally independent development of the two main schools of thought about sign function. The latter school derives from the teachings of the Swiss linguist Ferdinand de Saussure while the former refers to the system developed by the American polymath, physicist, philosopher, and geodesist Charles Sanders Peirce. Each school is grounded in a different model of sign relationship, and each model is useful and informative.

The Saussurian semiological model views the sign as a two part entity consisting of a signifier and a signified with an essentially arbitrary relationship between the parts (**Figure 1**). The signifier is the part present: for example, the letters C, U, and P, for instance, written on a paper. The signified is the concept of cup ness which groups together everything which might be thought of as having some cup like quality. It is the structure of the language or code system that permits the linking of the two factors (the combination of letters and the concept). This sign is then available to take part in a further act of semiosis; that is to say, it can itself become the signified of a new signifier–signified pair (a new sign). This mechanism of progressive construction of meaning is clearly displayed by the semiological model. For example, a stylized rendering on a sign board might be recognized as denoting a cup, while a sign of a cup can connote a café, and a café can connote a restroom facility. Each reading builds on the previous one.

The Peircian semiotic model is triadic: it consists of an object (the thing signified), a representamen (that which is present), and the interpretant (the code which provides the link (**Figure 2**). The utility of this model is that it allows a switching of interpretant; it shows how a sign can take on various meanings dependent upon context, expectation, and/or user experience. This is important, because the interpretant or code is brought to the semiotic by the interpretor (the reader, viewer, or hearer), and may or may not bear any resemblance to the interpretant expected by the author. Without an interpretant, the object and representamen have no relationship and the representamen would not be recognizable as a sign: it would not, in fact, be a sign. With an unexpected interpretant, the representamen may be taken as a sign for some quite unintended object. This model demonstrates quite clearly that the representamen (be it a text, a map symbol, or whatever) is not a self contained repository of meaning; but, rather, that meaning is negotiated each time the sign is read.

Herein lies an issue at the heart of all symbolization and iconography: how can the understanding of a sign be assured, or at least anticipated? If interpretants can be chosen or switched, or if they are, in any event, brought to the sign by the interpreter, how is it that reading or interpretation is not wholly idiosyncratic and individual? It is not idiosyncratic because readers are not autonomous and unconstrained; rather, a reader is created by the interpretive community to which the reader belongs. In fact, it is membership in one (or more) communities for which a given representamen exists as a possible sign (with denotations and connotations) which allows someone to exist as a reader. An analogy might be drawn to a global positioning system (GPS) receiver: the device recognizes a particular signal pattern from a satellite, picking it out of the background of electromagnetic noise (most of which is meaningless) because it accepts that signal pattern as a possible legitimate and meaning bearing pattern.

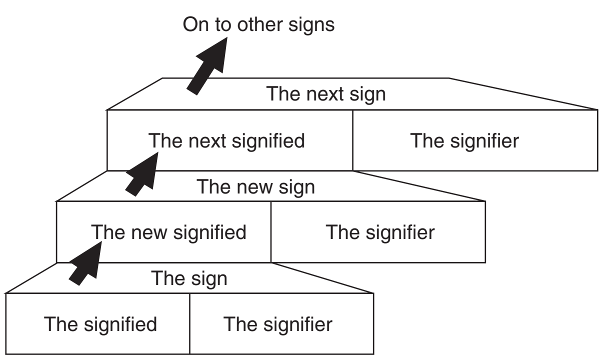

Figure 1 The Saussurian semiological model views the sign a two-part entity with an arbitrary relationship between the parts. The signifier is the part present, and this is linked to a signified object, concept, or sign. The resulting sign is then available to take part in a further act of semiosis by itself becoming the signified of a new signifier signified pair (a new sign).

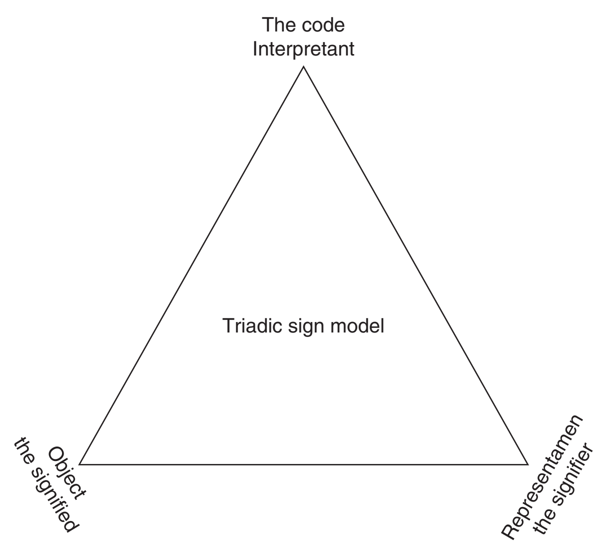

Figure 2 The triadic Peircian semiotic model consists of an object (the thing signified), a representamen (that which is present), and the interpretant (the code which provides the link). It is the interpretant which gives meaning to the representamen, and allows one to leap from the representamen to a signified object. The object (and thus the meaning of the sign) is negotiated between the representamen provided by the message composer (map maker) and the interpretant brought by the reader (map user).

The pattern, recognized as being amongst a vocabulary of legitimate patterns, is read as meaning "I am this satellite." A satellite, perhaps one from a different system, which emitted a different pattern, would not be recognized as bearing a legitimate meaning. Conversely, a legitimate pattern received from an entirely different source would be assumed to be the satellite. It is (for the purposes of this analogy) membership in the interpretive community of devices that can recognize the signals of legitimate GPS satellites that define the GPS receiver. Conversely, a text like Joyce's *Finnegans Wake* is considered difficult because it requires a reader with access to a tremendously wide and varied set of interpretants; the reader must be conversant with the usages of a large number of interpretive communities, applying several interpretants and recognizing interplay between the codes, to begin to make sense of the text.

Thus it can be seen that a map, as both a sign itself and an assemblage of constituent signs, must either exploit interpretant paradigms accepted as natural and legitimate to its target interpretive communities, provide independent access to interpretant paradigms utilized in its composition, or (usually) both.

Normally, it is through a legend, key, or list of conventional signs that a map makes explicit its symbolization scheme. Often a legend on a map will be incomplete in that it is restricted to signs that are deemed especially significant to the theme of the map, or that outline a specific hierarchy, or both. Generally speaking, the more complete a legend provided on the map itself, the less the map user or reader is required to rely upon their own expectations and assumptions. The legend, therefore, is an important feature affording access to the map.

Some map user communities (by definition, a type of interpretive community) have conventionalized forms, formats, symbolization regimes, etc., which serve to both define and to make usable maps for that community. Competitive orienteering maps, geological maps, and marine navigation charts are three instances of highly conventionalized maps. Often the conventions for such maps are so standardized and ubiquitous that understanding is assumed; in these cases, if a legend is published at all it is often as a separate document altogether.

Mechanics of Symbolization

As mentioned above, symbolization is one of the mechanisms of generalization, although, in effect, all generalization can also be seen as a process of symbolization. This is because generalization is a process where one representation is made to stand in not only for the configuration of the feature but possibly for another representation; it is the substitution of a sign for some aspect of a feature or of a sign. All maps require some form of generalization/symbolization; it is one of the primary means of directing and focusing attention upon aspects deemed important by the map's rhetorical argument and of rendering transparent or invisible aspects or features deemed to be extraneous.

Much of the mechanics of perception and understanding of map symbolization may be understood by application of the tenets of Gestalt psychology, or the psychology of form. Factors such as proximity, similarity, figural stability, good continuation, closure, and others make up the Gestalt laws of organization in perceptual forms, and explain how individual symbols are perceived on the page, and how they are perceived to interact with and among other symbols. While understanding of the mechanics of perception have moved beyond the relatively basic *gestalt* laws, it is still useful to conceptualize the visual field as governed by these rules. It is the *gestalt* of the image that, for instance, allows the individual marks that make up a dashed or dotted line to be understood as a single continuous entity. It is, as well, the *gestalt* that establishes the figure–ground relationship in the map image and permits multiple visual levels to develop, thus facilitating the reading of multiple intersecting configurations of symbolized features (**Figure 3**).

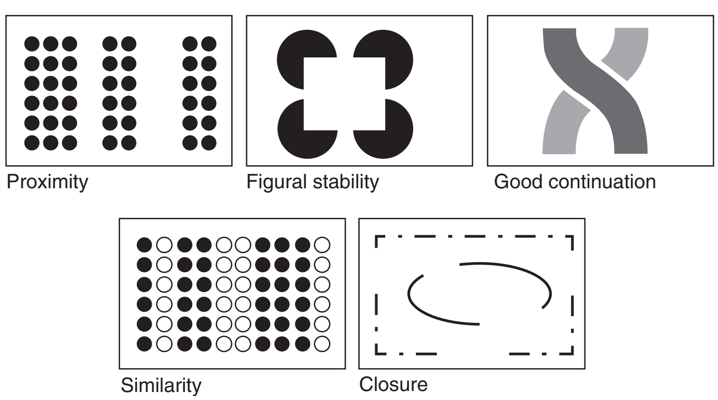

Figure 3 An illustration of some of the basic Gestalt laws.

An important distinction to keep in mind is that be tween representation and depiction. Depiction is simply one dimension or possible means of representation. A photo or sketch may depict a feature more or less realistically, and the depiction can be used to represent the feature, but the two activities are fundamentally different. A drawing depicting the Brandenburg Gate may, depending upon the context, represent an urban space, German Imperialism, postwar politics, or con temporary German reunification or it may represent a landmark or tourist destination. In any event, while it depicts the Brandenburg Gate, context governs what it may or may not represent.

It is useful to conceptualize a triadic model of sym bolization. The three axes of such a model would be the mimetic, the linguistic, and the graphic. The extreme of the mimetic axis would be a realistic rendering, while at the linguistic extreme would be a name. The graphic apex would be a geometric shape, pattern, or color tone. As one moves away from the mimetic pole, and toward the linguistic, the graphic, or both, the symbol becomes more abstract. Away from the graphic pole, the symbol becomes more specific. Precision or nuance is lost with distance from the linguistic peak; naming accesses powers of context and allusion not available to the other poles (**Figure 4**).

Taking Washington, DC as an example, the rendering of the US Capitol dome would be a mimetic, a label ‘Washington, DC’ a linguistic, and a circled star a graphic, symbolization. A tinted area (mimicking either the political or urban area of the city) would fall some where between the graphic and mimetic poles, with some edging toward the linguistic. Normally, some combin ation of symbol type is used: a labeled dot, for instance.

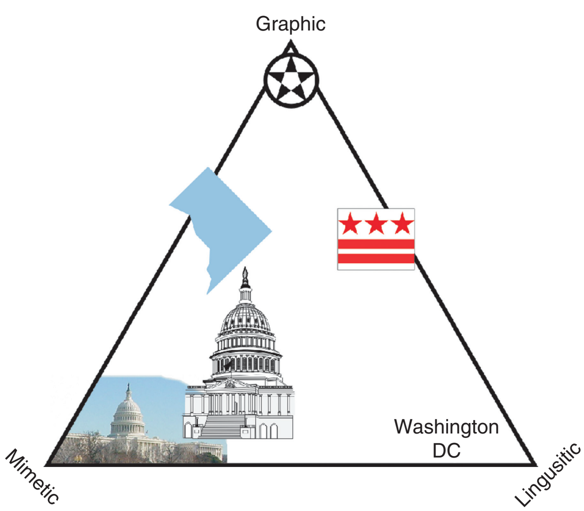

Figure 4 An example of various possible symbols for Washington, DC plotted on a triadic symbolization model.

This graphic redundancy multiplies the accessibility of the symbols.

Practics of Symbolization

A vocabulary of symbols and a grammar for their use is required. A vocabulary establishes the paradigm of meaningful representamen from which a selection can be made, while a grammar governs the acceptable proper usage to help allow orthodox interpretation. At the most basic level, map symbols are used to represent locatable features, and specific attributes of the locatable features. Representation of these features fall generally into three types: points, lines, and areas, and the attributes may be either qualitative or quantitative, or a combination thereof.

Point features are those which are locatable on the map (at that map’s scale) as a single location, or as a constellation of individual or representative locations. A series of connected points defines a line feature; usu ally the phenomena, described by the line, are assumed to exist equally at all locations along the line. An area is defined by a closed line: the enclosed zone is the repre sentation of the extent of the phenomena. Volumes can also be shown, but except in three dimensional models, volume depiction relies on perspective and perceptual conventions, the understanding of which may or may not be problematic.

The differentiable aspects of the map symbols can be categorized as symbol variables. Location, size, value, texture, color (value, saturation, and hue), orientation, shape, pattern, and the tripartite element of clarity (composed of the three dimensions of crispness, reso lution, and transparency) are the basic dimensions. Some types of symbols are well matched to some types of data but awkward or even wholly unsuitable for others. The accompanying tables may be of assistance: derived from MacEachren, they outline some example symbol vari ables and suggest a grammar of usage (**Figures 5** and **6**).

Selection of the symbol type most appropriate for a particular use is governed by many factors, such as map scale, the availability of graphic variables (such as color), and the representations used for other map information. In some cases, a point symbol may be chosen to stand for an area, or an area may be defined by use of a repre sentation of its boundary (a line feature) alone. Many times, text alone may be used to identify an area, espe cially if that area has no definite or uncontested bound ary. A mountain range comes to mind as an example, or an archipelago, where in fact the actual edge of the area is not significant or decidable in all instances.

Identification of the attribute upon which to base the representation is integral to symbol selection. As men tioned above, the attributes selected may be qualitative,

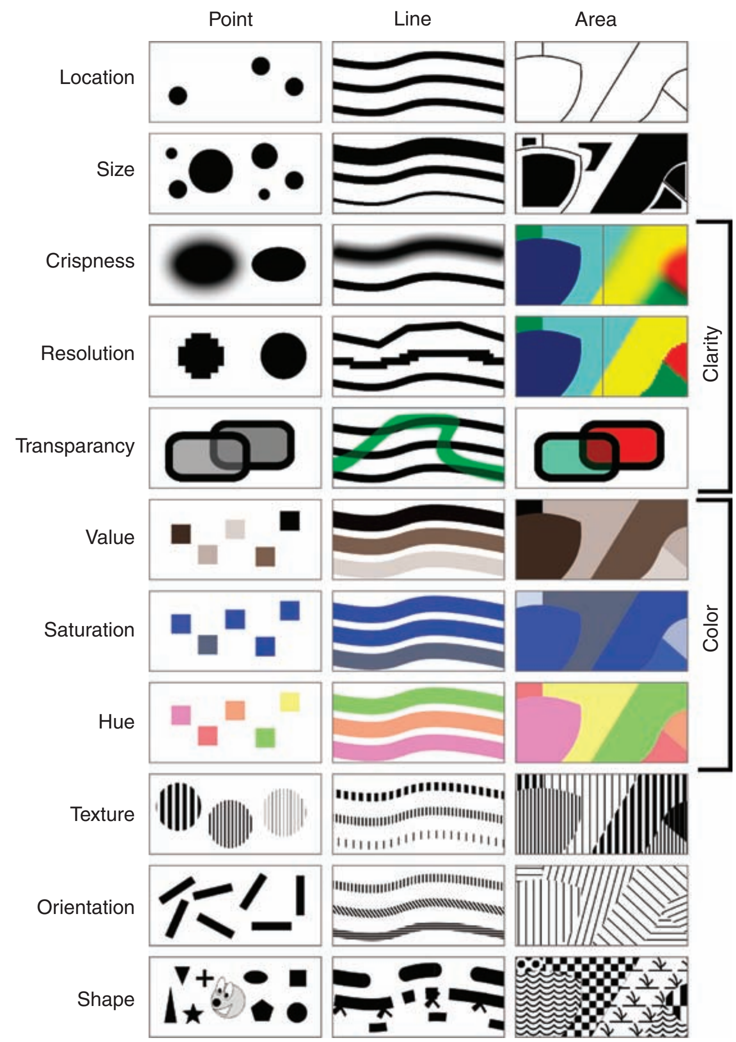

Figure 5 A basic vocabulary of graphic symbol variables.

quantitative, or a combination of both. Similarities be tween features are usually indicated by similarities in the feature symbolization, and vice versa.

Qualitative differences are usually represented by qualitative differences in the symbols employed; with quantitative differences between similar features employing differences in quantitative aspects of similar symbols. A symbolized road, for example, has its physical linearity represented by the employment of a line sym bol; but other qualitative attributes may be indicated by the exact specification of the line used. A particular road may be classified, for example, by surface type, traffic type, one or two way traffic, access restrictions, etc., and these various classes symbolized using different color hues, line weights, or complexities. These qualitative differences can be incorporated into a qualitative hier archy of road types and thus indicate information im portant to navigation of the road network. Roads in a network could also, or alternately, be classed by quantitative attributes, for instance, traffic capacity or demand. Quantitative differences in a line symbol may be expressed by line weight or color value, and are amen able, by their very quantitative nature, to being in corporated into hierarchies of their own.

Perhaps most critical in establishing a symbol set for a particular map is the issue of differentiability and asso ciativity: the ease with which the various symbols can be identified unambiguously as different, where differences are important, and as similar, where similarity is of interest. In both cases, it is the identification of significant differences between symbols that is the key. Conversely, it is important to avoid gratuitous variation: that is, dif ferences between symbols that do not carry symbolic significance but can distract from interpretation. Many factors can affect what qualifies a difference as significant or perceivable. Differences between shades of color which may seem obvious in large adjacent patches can become problematically differentiable in small patches

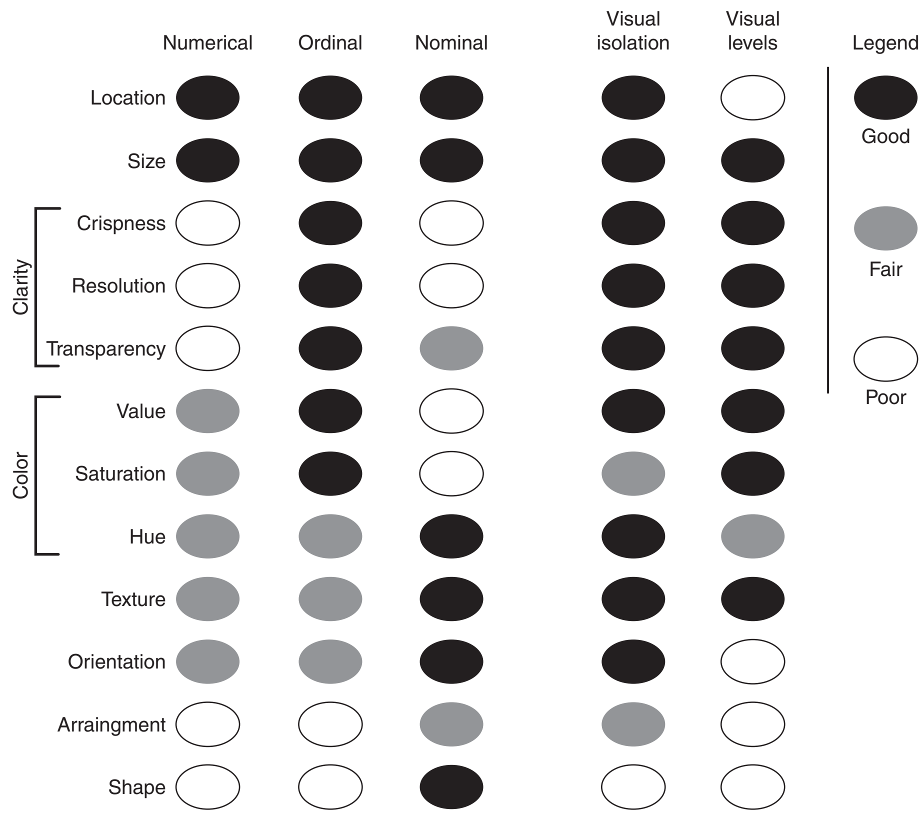

Figure 6 A basic grammar for use of graphic symbol variables.

widely separated on a map. The same concerns apply to small, iconic point symbols, or other types of related signs. As the symbol size decreases, and as symbols appear more widely separated on the map, more significance is required in their differences. The factor of graphic redundancy can be a key in overcoming problems of this sort. By altering two variables, the symbol is more easily seen as different.

Without doubt, the most critical practical aspect of symbolization is ensuring that the code system is accessible to the map user. As has been mentioned above, it is possible for a potential user to bring almost any interpretant to the map, and without guidance the map user's interpretation cannot necessarily be anticipated, let alone guaranteed. Provision of a key or legend, explicating the conventional signs employed on the map is, therefore, of great importance. Whatever form the legend takes, its function as a bridge to disambiguation remains central.

Legends or keys can take a variety of forms. The list type is very often seen: in a list, map symbols are directly paired with definitions and hierarchies are set out in order. This type of legend can be quite efficient and compact, although abstraction of individual symbols from the general map context to the neutral ground of the key may interfere with identification. Sometimes, this problem is addressed by the use of a so called natural legend, or sample map. This is a small section of map containing examples of all the symbols used on the map with each symbolized feature labeled. Although a small section of the actual map itself may be used for this, it is relatively seldom that any conveniently sized section of the map actually contains all the required features. More often, an idealized sample is created for use as a key. It is also possible to create legends intermediate between the neutral list and the natural sample map, which bring a ground mimicking the map context into the list.

It is very desirable that all symbols appear in a legend exactly as they appear on the map; including the exact same symbol size, color, and orientation. Occasionally, production or design constraints can make this impossible, but every effort should be made to allow as little difference between the symbols as shown on the map and in the legend. Introduction of random variations into the symbol vocabulary key is counter to the rationale for the existence of the key (**Figures 7** and **8**).

Maps and Map Language

Cartographic symbolism and iconography has been explained above as the application by a potential map user of a code system, something commonly understood as a language (either graphic, verbal, or both), to a set of signs in order to imbue them with meaning. It is important to keep a clear conceptual distinction between the system(s) of meaning (languages) utilized or leveraged by maps and the body of works (the individual maps) that take on meaning when interpreted through one or another of these systems. This distinction has not always been well understood, and has at times contributed to debate in the cartographic literature.

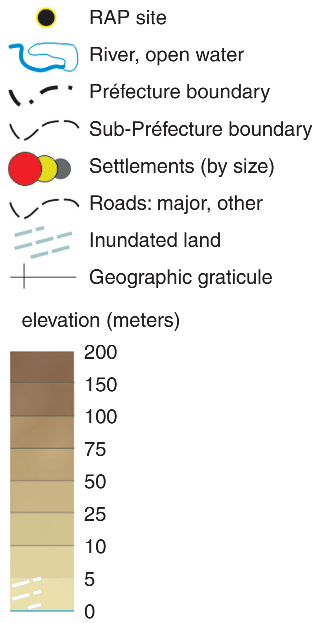

Figure 7 A list type map legend. *Préfecture de Boké dans le nord-ouest de la Guinée*. From Wright, H. E., McCullough, J. N. and Diallo, M. S. (eds.) (2006). A rapid biological assessment of the Boké Prefecture, Northwestern Guinea. RAP Bulletin of Biological Assessment 41, p13. Conservation International, Washington, DC.

Since the 1970s, there has been a good deal written about 'map language' or 'maps as language', with much of this discussion utilizing an essentially mechanistic flow model of communication first developed in the late 1940s. Although some authors (e.g., Koláčný) produced very complex diagrammatic elaborations on that basic theme, most working cartographers found the models to be of little utility.

Similarly, the occasionally recurring interest of some cartographic writers in the structure and mechanics of the eye is also of marginal utility. The eye is the tool of vision, but the eye has evolved to facilitate vision; vision did not evolve to make use of the eye.

Other writers attempted to draw on research in the linguistic, semiotic, and/or psychological fields in order to account for the multifaceted nature of cartographic practice. These borrowings, however, often met with resistance. In discussing linguistic theory as applied to cartography in 1976, Robinson and Petchenik were so distracted by superficial differences between verbal texts and graphic texts that they were unable to grasp the essential identicality of all semiotic activity. They were thus forced to reject parallels to language and to instead claim special powers for what they termed 'mapping communication' in order to set it apart from other reading. Head's 1984 discussion of using the analogy of natural language as a paradigm for understanding map communication approached the issue rather less doctrinally

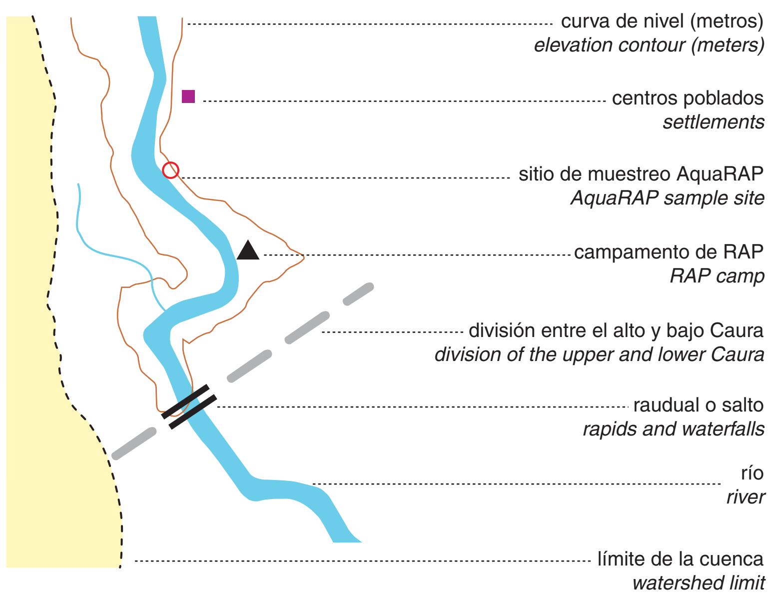

Figure 8 A natural-type map legend. *AquaRAP Caura River Watershed*: *State of Bolivar*, *Venezuela*. From Chernoff, B., Machado-Allison, A., Riseng, K. and Montambault, J. R. (eds.) (2003). *A Biological Assessment of the Aquatic Ecosystems of the Caura River Basin*, *Bolívar State*, *Venezuela*. RAP Bulletin of Biological Assessment 28. Conservation International, Washington, DC.

and more usefully, although Schlicthtmann's critique of that work from a semiotic standpoint is quite pertinent.

Certainly, as has been mentioned above, the code systems leveraged by maps can be thought of as languages (because they in fact are languages; usually graphic lan guages). We also know that meaning is not resident in the map; the map is made up only of signs, which must be interpreted through a language, because language/code systems are the only systems through which meaning is possible. A code system through which an artifact can be interpreted as a map can be termed a map language.

Obviously, there can be no one 'map language'. Map languages are cultural artifacts, and although various such systems of understanding can inform and influence each other, it is entirely possible that a map recognizable through the aegis of one system may be nonsensical under another.

Although a cartographer can provide a legend or key to a map (i.e., to suggest a system for reading and understanding), it is still up to the user to choose and apply a system. The signs (which can potentially be read as a map) are presented to a potential user, and it is up to that potential user to select and apply an interpretant (also called a code system or language) in order to generate meaning: in order, that is, to make the thing a map and make themselves readers.

See also: Color, Mapping; Map Interactivity; Map Types; Mapping, Philosophy; Maps; Qualitative Spatial Reasoning; Quantitative Data; Symbolic Interactionism.

Further Reading

Belbin, J. (1996). Gestalt theory applied to cartographic text. In Wood, C. H. & Keller, C. P. (eds.) *Cartographic Design: Theoretical and Practical Perspectives*, ch. 18, pp 253 269. Chichester: Wiley.

Bertin, J. (1983). *Semiology of Graphics*. Madison, WI: University of Wisconsin Press.

Caivano, J. L. (1990). Visual texture as a semiotic system. *Semiotica* 80(3/4), 239 252.

Campbell, J. (1993). *Map Use and Analysis* (2nd edn.). Dubuque, MS: Wm. C. Brown.

Chernoff, B., Machado Allison, A., Riseng, K. and Montambault, J. R. (eds.) (2003). *A Biological Assessment of the Aquatic Ecosystems of the Caura River Basin*, *Bolívar State*, *Venezuela*. RAP Bulletin of Biological Assessment 28, p 17. Conservation International, Washington, DC.

Denil, M. (2003). Cartographic Design: Rhetoric and Persuasion. *Cartographic Perspectives* 45, 8 67.

Foucault, M. (1983). *This Is Not a Pipe*. Berkeley, CA: University of California Press.

Head, C. G. (1984). The map as natural language: A paradigm for understanding. *Cartographica* 21(1). Monograph 31.

Hoffman, D. D. (1998). *Visual Intelligence: How We Create What We See*. New York: W.W. Norton.

Imhof, E. (1975). Positioning names on maps (trans. McCleary, G. F.). *The American Cartographer* 2, 2.

Jacob, C. (2006). *The Sovereign Map: Theoretical Approaches in Cartography throughout History* (trans. Conley, T.; Dahl, E. (ed.)). Chicago, IL: University of Chicago.

Koláčný, A. (1977). Cartographic information A fundamental concept and term in modern cartography. *Cartographica* 14(supplement 1). Monograph 19.

MacEachren, A. M. (1995). *How Maps Work: Representation, Visualization, and Design*. New York: Guilford Press.

McCloud, S. (1994). *Understanding Comics: The Invisible Art*. New York: Harper Collins.

Roninson, A. and Petchenik, B. (1976). *The Nature of Maps: Essays toward Understanding Maps and Mapping*. Chicago, IL: University of Chicago Press.

Rock, I. and Palmer, S. (1990). The legacy of Gestalt psychology. *Scientific American* 263(6), 84 90.

Schlicthtmann, H. (1984). Discussion of C. Grant Head 'The map as natural language: A paradigm for understanding'. *Cartographica* 21(1). Monograph 31.

Wood, D. and Fels, J. (1986). Designs on signs: Myth and meaning in maps. *Cartographica* 23(3).

Wright, H. E., McCullough, J. and Diallo, M. S. (eds.) (2006). A rapid biological assessment of the Boké Prefecture, Northwestern Guinea. RAP Bulletin of Biological Assessment 41, p 13. Conservation International, Washington, DC.

Systems

G. P. Chapman, Lancaster University, Lancaster, UK

Glossary

Complexity Theory Abroad spectrum of ideas, relating to dynamic nonlinear systems which exhibit unpredictable, nonrepetitive, complex emergent behaviors

Degrees of Freedom The degrees to which system components or whole systems are free from constraint on their behavior.

Entropy In thermodynamics a measure of the unavailability of energy for work. At maximum entropy, energy is uniformly distributed, there are no gradients (potentials) so no work can be done. In statistics, entropy is a measure of uncertainty, and is maximized when all probabilities are equal.

Feedback The measure by which a system detects its behavioral distance from a target state.

General Systems Theory The theory that postulates that isomorphisms between systems of different materials and scale may be found at an abstract and behavioral level.

Godel's Theorem This says that a complete formal system cannot demonstrate its own completeness. If an axiomatic system can be proven to be consistent and complete from within itself, then it is inconsistent. This is about the limits of reflexivity in formal systems.

Laws of Thermodynamics These are laws about the conservation of energy, and about its availability for work (see the term 'entropy') and about the fact that the universal tendency is for things to run down–that is, energy dissipates toward an equilibrial state of unavailability.

Most of us have an appreciation that our environment is composed of distinguishable objects or things. We also learn quite early in our lives that these objects are made of other objects, that they have parts, whether the leaves of a tree or the grains of sand in a sandpit. But these objects do not all have equal ontological status. It is quite evident that a pile of sand is simply an aggregation of sand particles, whereas a tree is not simply and only an aggregation of leaves. The word system is used to describe those phenomena where the parts are related to the whole in a manner which suggests that there are some organizing principles involved in the assembly of the whole.

At this level the idea is intuitively obvious, and we can appreciate that a wristwatch is a system, and that an animal can be viewed as a system. Both of these are contained within a region of three dimensional space, and are, therefore, volumetric wholes. But the idea of a system can very rapidly be extended to assemblages that are not volumetrically whole. We talk glibly about social security systems, broadcasting systems, relief delivery systems, etc., which we thereby identify as holistic in some sense, and as having organization. The basic question to be asked of systems theory is: what is to be gained by saying that all these things are systems?

To be useful, we have to know what distinguishes systems, and what we can say about them from basic abstract principles. Then we will know when we are on firm ground with sound conclusions, and be clearer about when we are on shaky ground with more tentative conclusions. If we are not clear about these things then the value of the concepts is undermined. Indeed, uncritical and uninformed (if enthusiastic) 'applications' of systems theory have led some critics to observe that the idea of 'system' is nothing but common sense.

The Origins of Systems Theory

The origins of systems theory are to be found in the application of mathematics to physical problems, such as Newton's formulation of the laws of gravity and the deepening understanding of the solar system. The development of powered machinery also prompted new thinking at an abstract level about how work could be abstracted from heat, and the limits that governed the performance of 'thermodynamic systems'. The machinery itself employed new principles such as governors, which were 'negative feedback' devices governing the speed of rotation of a steam engine.

Systems thinking of the classical kind, to be described in this article, was most forcefully articulated in the period 1940s–80s, a period when Ludwig von Bertalanffy promoted the vision of 'general systems theory' (GST). This was founded on the belief that understanding was more to be found in the structure and organization of wholes than in the physical distinctiveness of their parts. Thus, it was the isomorphic structure of materially different systems that was of most general interest. By this reckoning, there is no difference between Newton's laws of gravity applied to the solar system and the application of the 'gravity models' to human migration between towns. Both can be shown to have a base in probability theory and both can be described as 'entropy' maximizing – that is the most probable solution within known constraints. Another example is that the oscillation of a spring and the oscillation of a pendulum can both be described by the same differential equations characterizing sine wave.

The coupling of parts within an organized whole can be revealed empirically by statistical techniques such as correlation (crude and linear, or polynomial and more subtle) or modeled in terms of minimum information requirements. One of the reasons why GST accelerated in the post war period was the development of formal 'information theory' and its application within 'cybernetics'. Just as thermodynamic theory had shown what the theoretical limits were to the capture of energy, so information theory could show what the limits were to the 'degrees of freedom' of the parts within the whole.

Until the advent of computers, the only way to know theoretically how a system would behave was to find an algebraic solution to the equations which summarized the transformation of the variables that represented it. If the system did not go to a stable state, or if there were lots of numerical constants involved, it was practically and sometimes theoretically impossible to find a solution. However, with the arrival of computers, these constraints were broken. GST coincided with these developments. To begin with systems theorists used computers to break through the practical barrier, and to iterate calculations until a stable and terminal state was found for some arbitrary system. An example in geography is the derivation of a 'Von Thunen' landscape from any arbitrary starting point when farmers compete with each other.

However, the programmers also discovered that some systems never go to equilibrium; that they behave chaotically in an unrepeating trajectory. The theories developed so far had not expected results of this kind – even if the French mathematician Poincaré had glimpsed the possibility of this behavior in the nineteenth century. The discovery of these new dynamics is the point when systems theory as a comprehensive framework begins to fade into the background. Its ideas are still around, but now subsumed within the exploding field of 'complexity theory'. The blurring of the terms is well illustrated by the recent foundation of the European Complex Systems Society.

The basic understandings of systems theory are nevertheless still relevant, and are the subject of the rest of this article. There is however one extremely important caveat to stress. Systems theory began with physical systems in a physical world, and understands physical limits. There are no perpetual motion machines, and never will be. But much of systems thinking in the social sciences forgets these constraints, and often argues implicitly by analogy. If the analogy is exposed, the conclusions can sometimes be shown to be very dodgy. For example, in classical systems theory the word 'equilibrium' has a thermodynamic origin. It is then possible to show how an open system may be far from equilibrium, and how if it loses its inputs, it has to subside to its ground state of equilibrium. Some social theorists speak of social systems, without thinking of physical constraints, as being 'far from equilibrium'. Since they cannot or do not specify what equilibrium is, that is, the ground state that equilibrium is said to be far from, the analogy may be meaningless.

Isolated, Closed, and Open Systems

The Earth, viewed from space, and conceptually imagined as a lifeboat, is a single system. But most of the examples given so far in this article are of examples of systems to be found on the Earth. So 'the' system is composed of many systems: indeed Brian Berry authored an article entitled 'Cities as systems within systems of cities'. This begs a large question. The fundamental truths of systems theory can only apply if the systems are 'not' arbitrarily defined. How do we define the boundaries of a system, if in the end everything is related to everything?

Starting from thermodynamics, theorists define three kinds of systems. The first is an 'isolated system', which has a border that is crossed by neither energy nor matter. We know of no such thing in the current universe, though we can hypothesize a perfectly insulating thermos flask. The nearest is a black hole, in the sense that it can send us neither material nor energy (although it takes both from our universe), so there is nothing we can know or say about the conditions within the black hole.

A 'closed system' is one which is closed to the import and export of material, but open to an exchange of energy. Planet Earth is almost a closed system; it isn't quite, because it receives an input of dust and meteoroids on a daily basis. For conceptual purposes, it is near enough closed. It is open with respect to energy, receiving short wave radiation from the Sun, and transmitting long wave radiation back into space. If the energy were turned off (the Sun stopped shining), the Earth, like a switched off kettle, would cool to the background temperature of space. The trend toward the universal equilibrium temperature is one of the ways in which we can express the idea of entropy – the end of differentiation. So, the fact that the Earth is not at that background temperature is a statement that it is not at equilibrium, and it maintains its distance from equilibrium by the throughput of energy. The energy throughput creates pathways – for example, wind patterns in the atmosphere – so by dissipating energy, structures are created. Closed systems are dissipative structures – where the structures are recognizable by differentiation. 'Von Thunen's isolated state' was not actually an isolated one. It was a closed one. The material content of his plain was fixed, but the Sun shone to grow the crops and the woods, so that people and animals could find the energy to work and the fuel to keep themselves warm in winter and cook their food; and the waste products were presumed to be recycled.

The differentiation creates opportunities for other systems to develop, which are also dissipative in their own way. 'Open systems' are those which exchange both matter and energy with their environment – usually the one bound up with the other. So, for example, plants use solar energy to convey water from the ground to the atmosphere in transpiration, thereby also conveying the materials out of which they construct themselves, and in those materials energy is locked up. This is sometimes called a local reversal of entropy, or the production of negentropy – order out of disorder.

Human beings are open systems. Although humans can absorb heat energy from the Sun to warm themselves, they actually absorb virtually all their useful energy by consuming food, which means they cannot dissociate their exchange of energy with the environment from their exchange of material. The same concepts can be used to describe current industrial civilization. For the moment, leaving aside the small contributions made by nuclear and wind/hydro energy, our industrial systems consume energy by breaking down coal, oil, wood, etc. Our cities are far from a thermodynamic equilibrium, and are therefore dependent on this continual through put. As a thought experiment, the question to be asked is: can our urban concentrations of population, living off food delivered from supermarkets survive long if all in dustrial energy supplies failed?

Where is the 'boundary of the open system' of modern civilization? What defines the distinction between us and our environment? It is reasonable to think of this human economic–social system as a separate subsystem of planet Earth 'only so long as' the essential outputs from the subsystem, like waste heat, waste gases, waste materials, do not change the nature of the inputs we can receive from the environment. When that is no longer the case, and there is feedback, then we are no longer separate from our environment, and had better pay heed to the totality of the Earth as one system.

Degrees of Organization

In a system, the degrees of freedom of the whole are less than the sum of the 'degrees of freedom' of the parts. Simple mechanical analogies can illustrate the principles. Suppose you buy an Ikea wardrobe flatpack containing 40 parts. When each part comes out of the pack on its own, it has six degrees of freedom. It can be placed anywhere in three dimensions of space, and at whichever location it happens to be in, it can be rotated around three different axes. When the contents of the box are tipped out, there are $40 \times 6 = 240$ degrees of freedom. The task is to as semble the parts in the right connection with each other, and if you are successful, when you have finished you will have a single wardrobe, with only six degrees of freedom (plus a little extra because the two doors can rotate on their hinges). You can place the wardrobe anywhere in the three dimensions of space, and rotate it at any angle. But whatever you do to the whole, you have also done to the all the parts (except for the doors if they are not locked and swing open in your face). By building a wardrobe you have absorbed 234 out of the 240 degrees of freedom. This is such a staggering task that Professor Andrew Ng of the Computer Science Department at Stanford University in 2006 made it his target to get his team of 30 scientists to build a robot that can do this, within the following decade.

In any nontrivial system the parts need to have the ability to be in alternative states. The state they actually adopt is dependent at least to some degree on their membership of the system. In a 'central place landscape' the idea of the isotropic plain is that any part of the plain has the same potential to be farmland, a small settlement, a town, or a city as any other part of the plain. But once the willingness to travel is fixed and a central city de marcated (this is arbitrary), all other places are selected to be what they are, from farmland to town. The *a priori* ensemble of possibilities is reduced to an *a posteriori* en semble with a single member. Now, the degree of or ganization is expressed as the degree to which the entropy of the *a posteriori* ensemble is less than the en tropy of the *a priori* ensemble. Before we put the parts together, anything could be anything, and so there was maximum uncertainty. But once they have been put together, there is only one member of the ensemble possible, and entropy (uncertainty) is zero. So the model has absorbed all uncertainty, and has achieved 100% organization. Each part is 'exactly' and 'totally' what it is by virtue of its membership of the system.

The perfectly organized system cannot evolve – not from within itself. This was a lesson learned the hard way by the planners of the former Soviet Union. The basis of their planning was a rational input–output matrix, which specified the quantities that linked any one industry with all others. Such a system could be used for scaling up production across all industries, but it did not allow for innovation, for new industries, for new technologies that changed efficiencies, etc.

Regulation and Variety

'One' way of defining a system and its environment is to look for key variables that may be held constant in the face of environmental variability. Warm blooded animals have to keep a constant body temperature in a variable environment, and some can do so by sweating and shivering, others by plumping their feathers, etc. (among other responses.) To physiological responses human beings have added the extra responses of variable clothing and shelter, which may have variable heating and cooling. This is regulation – given a formal definition by the cybernetician W. Ross Ashby's 'law of requisite variety'. This says, in linguistic terms, that only variety can destroy variety. If the animal cannot shiver more or less, or sweat more or less, it cannot counter the environmental fluctuations. Since its formulation, there have been many 'applications' of the law in the social sciences which are nonsense – because they have ignored its precise mathematical formulation. If the quantities 'environmental variability' and 'variability of response' are not properly defined, we can end up with the proposition that one witch doctor's spells can confound another's spells.

If the variety in the environment exceeds the capacity to respond, then change will occur in the subsystem. Temperatures can be reached which human beings cannot survive, even by sweating. Thus parts are linked to wholes in ways which show nonlinear responses, some of them even discontinuous, or 'threshold' responses. The modeling of nonlinear relationships before the age of the computer was too difficult for more than rather simplistic two variable attempts to be made (e.g., the famous Volterra–Lotka equations, looking at the phase space of shrimps versus sharks). After the advent of computers it proved possible to investigate these – and it is at this point that the deterministic, nonrepeating behavior of chaotic systems is discovered, and out of systems theory chaos and complexity science are spawned.

The Application of Systems Theory in the Social Sciences

The idea of 'systems theory' has been adopted with enthusiasm by many members of the social sciences, but whether the resulting theories are actually compatible with the 'original' systems theory grounded in physical science is sometimes open to doubt. Arguing by analogy, in economic systems money becomes 'energy'. But money is not energy: energy can be found in power lines and fuel depots and human stomachs. Money is one of many regulators that regulate how energy and materials flow. Some social systems are said to be 'far from equilibrium', which has an implicit feel of the potential for incipient instability, crisis, and fall from grace. Whereas a thermodynamicist knows the direction of the ground state, the social theorist does not. Is random social anarchy the ground state? Is this truly the most probable state – the Hobbesian nightmare imagined in Leviathan? Despite the contemporary fears about failed states, in most of the Earth there are some sorts of ordered social states, even if some social orders do not like other social orders. Mankind has, after all, evolved as the most socialized animal on the planet, and even where there are failed states societies can be found.

Is a central place landscape far from equilibrium? Above it was suggested that any urban system is far from equilibrium – in a 'thermodynamic sense'. In cybernetics there is another sense in which a system can be said to be closed. Such systems can be transformed from any initial state to any output state, but there is no output state which is not included in the input conditions. So the system cannot go to a state from which we cannot predict its future. It is 'cybernetically closed'. Repeated transformations of such systems lead to a stable end state or end cycle. An example in economics would be a cobweb diagram that converged on a market equilibrium. The same can be said for the central place system, in that it is the equilibrium result of the postulates. But this sense of 'most probable' is the opposite of the thermodynamic 'most probable'. Just as Von Thunen's landscape needed sunshine, so the inhabitants of the central place world need energy to shop infrequently in high order cities, and more frequently in local stores. We have come back to the same point: and it does bear reiterating. In an abstract conceptual world 'most probable' may mean exactly the opposite of 'most probable' in an actual physical world.

Self-Image and Reflexivity

If there is a system, there is an environment, and the system behavior is independent of that environment at least in some minimum respects. According to some authors this means that the system must in some sense have some image of itself and its environment, to maintain its independent state. This image may be as simple as a thermometer reading used by a thermostat, controlling a room heating system. The environment then is constituted simply by the variation of temperature – no other feature of the environment exists so far as the thermostat is concerted. A little more sophistication could be added if there are combined temperature and humidity controls. Then one can hypothesize a very sophisticated system which also included visual assessment of the state of pot plants in the room, and perhaps in the future there will be such systems. In human society a more complex example would be the understanding that a person has of his or her social standing and the reasons for it. But no system can have a perfect understanding of itself and its environment – whatever that would mean – because the information specifying the totality cannot be compressed in a subsidiary part. This is one of the implications of 'Gödel's (incompleteness) theorem'. This can be re expressed as the problem of 'reflexivity', in that if parts were to have perfect images of other parts, these perfect images would encompass the images that those other parts held of the first part – and this becomes an infinite reiteration. An example of this is the 'reality' of the world's longest running radio drama soap opera series, the BBC's *The Archers*. This has now run for several generations, dramatizing the lives and relationships of a fictional village 'Ambridge', in rural England. Many fans treat it as a real depiction of a real place, and it enshrines the idea of society reflexively examining itself. However, it cannot be real, because it lacks one essential feature of the lives of people in rural England. At least some of the people of Ambridge should from time to time be heard listening to *The Archers*. For obvious reasons, this never happens.

All of these problems became apparent in the systems models publicized by Jay W. Forrester (*Urban Dynamics* and *World Dynamics*). Despite this, the Club of Rome com missioned Forrester's student Donald Meadows to provide the simulation basis for *The Limits to Growth*. The variables selected for these models were essentially chosen on the basis of best guesses, and the relationships between the variables were also specified on the same basis. So a part of the world system (D. Meadows) is exhibiting its ignorance of the world system. It uses the computer systems model simply as a way of calculating results it could not previ ously have achieved. That in most scenarios the world 'collapsed' was a function of the form of the nonlinear relationships, something at the time that was very poorly understood, and the time frame within which collapse occurred was arbitrary. Despite the specious objectivity of these projects, they had huge popular impact, and can be said to be one of the beginnings of the modern environ mentalist movement. What they achieved can best be summarized by a word coined by Forrester: the behaviors exhibited were often 'counter intuitive'. However poorly constructed, their systems had shown 'surprising' behavior. In that lay the seeds that would flourish with the 'new' fields of chaos and complexity.

Since Meadows work, there has been little more on this kind of social–environmental world dynamics. However, in the field of atmospheric modeling, the pro ponents of global circulation models believe that they have produced reasonably robust models which can simulate the gross dynamics of the atmosphere over decades. These models do not include the reflexive im pact of the models on the behavior of their human op erators nor the wider society. There are other ontological and epistemological issues too. A critical realist would say that there is a real atmosphere we can detect, but the aspects we chose to detect and the scales of resolution at which we do so – both of which are critical to model behavior and outputs – are conditioned by our per ceptions. Further, in epistemological terms, since model outputs are predictions of the future, they can never be tested except after the event. Hence the validity of these systems rests not in correspondence, but in coherence. 'Truth' is then confirmed when all the experts sing in unison from the same hymn sheet.

Related Fields

Systems theory influenced, and has been influenced by many other fields, most notably management science and optimization techniques. This field, otherwise known as 'operations research', has its origins in statistical model ing and other optimization techniques, developed during World War II, to minimize Atlantic convoy losses and Royal Air Force (RAF) bomber losses. The modern field includes techniques often employed within industrial and construction projects, such as critical path review (pro gram evaluation and review technique), queuing theory, game theory, probability and risk, linear programming and agent based modeling. The essential key to all of these is that the system is the paramount level of analysis, but usually seen from the perspective of external control responsible for optimization. This is somewhat different from the other perspective that developed out of systems theory, namely the understanding of self organization and complex dynamics, reviewed in 'chaos and com plexity'. The difference is not absolute, because opti mization could well involve allowing systems to become self organizing. This is a point lost on most hierarchically minded managers.

See also: Spatial Science; Systems Theory.

Further Reading

Benenson, I. and Torrens, P. (2004). *Geosimulation*. New York: John Wiley.

Chapman, G. P. (1977). *Human and Environmental Systems: A Geographer's Appraisal*. London: Academic Press.

Chapman, G. P. (2007). Evolutionary psychology, complexity theory, and quantitative social epistemology. *Futures* 39, 1067 1083.

Chisholm, M. (1967). General systems theory and geography. *Transactions of the Institute of British Geographers* 42, 42 52.

Huggett, R. (1980). *Systems Analysis in Geography*. Oxford: Oxford University Press.

Skyttner, L. (2001). *General Systems Theory: Ideas and Applications*. River Edge, NJ: World Scientific Publishing.

Systems Theory

D. Straussfogel, Pennsylvania State University, DuBois, PA, USA
C. von Schilling, University of Northern British Columbia, Prince George, BC, Canada

Glossary

Autopoiesis The process by which a system regenerates itself through the self-reproduction of its own elements and network of interactions.

Cybernetics The study of feedback and derived concepts such as communication and control in living organisms, machines, and organizations.

Dissipative Structure A dissipative structure is a thermodynamically open system which is operating far from thermodynamic equilibrium in an environment with which it exchanges energy, matter, and/or entropy.

Environment The context within which a system exists. It is composed of all things that are external to the system, and it includes everything that may affect the system, and may be affected by it at any given time.

Feedback A process by which information concerning the adequacy of the system, its operation, and its outputs are introduced into the system.

General Systems Theory The concepts, principles, and models that are common to all kinds of systems and the isomorphisms between and among various types of systems.

Holism A nonreductionist descriptive and investigative strategy for generating explanatory principles of whole systems. Attention is focused on the emergent properties of the whole rather than on the reductionist behavior of the isolated parts.

Reductionism One kind of scientific orientation that seeks to understand phenomena by breaking them down into their smallest possible parts or to a one-dimensional totality.

System A group of interacting components that conserves some identifiable set of relations with the sum of their components plus their relationships (i.e., the system itself), conserving some identifiable set of relationships to other entities (including other systems).

Wholeness In reference to systems, the condition in which systems are seen to be structurally divisible, but functionally indivisible wholes with emergent properties.

Emergence of Systems Thinking in Contemporary Science

Synopsis: Overview of Systems Theory

Systems theory, emerging in western science in the middle of the twentieth century, encompasses principles derived from the defining characteristics of systems regardless of their discipline of origin (i.e., neurology or ecology) or their scale of analysis (i.e., microscopic or macroscopic). Systems theory allows lessons learned about one system to be applied to understand, manage, or construct another. For example, greater understanding of how neural net works work in the brain could provide insights about ef ficiencies in telephone communication systems.

Systems theory does not represent just one theory, *per se.* Rather, it co evolved with a set of related intel lectual streams concerned with the nature and character istics of systems. These include information and game theory, cybernetics and chaos theory, theory of autopoiesis, complexity theory, and dynamic systems theory. Disci plines of application are as diverse as engineering, biology, ecology, geography, sociology, psychiatry, and neurology.

The origins, context, and dimensions of systems the ory will be outlined and explained in the following sec tions. Although they are, for the most part, presented in chronological order, it should be understood that these events and theoretical developments have become interwoven in such a way as to define the broad, inter disciplinary, and multidimensional face of systems sci ence as it is known today.

What Is Systems Thinking?

'Systems thinking' holds to the adage that the whole is greater than the sum of its parts. This means that the qualities that emerge from a whole are different from the qualities found in any of the individual parts or even the summation of those parts. Stated another way, what characterizes a system is not the nature of its parts alone, but the interrelationships between the parts and between the system and its environment. In this way, systems thinkers are expansionists, who would seek to understand a system by examining its processes and relationships as opposed to the reductionists of classical science, who would seek to understand a system by first examining its elementary parts and then by adding them together. While 'the study of wholes' is not mutually exclusive of 'the study of parts', systems scholars maintain that sys tems are better comprehended in their entirety.

Classical Science and the 'New' Science

The intellectual heritage of the paradigm shift in science that occurred at the beginning of the twentieth century extends to the Age of Reason that took place in Europe from approximately 1500–1700 corresponding with the end of the Middle Ages. Engendered by Nicolaus Copernicus (1473–1543) in the mid sixteenth century, and René Descartes (1596–1650) at the turn of the seventeenth century, the Age of Reason launched Europe out of the Middle Ages and marked a decline in the ideas pervasive in medieval thought, which focused on evidence of the divine in mind and matter and marked a rise in the meaning and power of the human intellect. If the Cartesian worldview heralded the power of human reason to 'know' the universe through deduction, Isaac Newton's (1643–1727) work demonstrated it. He posited a model of the universe that could be known by reducing it into ever smaller parts. Even at the cosmological level, planets could be reduced to particles of matter found at the microscopic level, which he conceived as the building blocks of all matter. Moreover, regardless of scale, he argued that all parts, be they atoms or planets, could be described by their location, movement, and the laws of attraction (i.e., gravity). The mathematics of 'Newtonian mechanics' would be developed over time by numerous individual achievements and described a linear, causal, and therefore predictable universe, rather like a complex mechanism.

Early in the twentieth century a team of scientists that included Albert Einstein (1879–1955), Werner Heisenberg (1901–76), and Erwin Schrödinger (1887–1961) developed the conceptual framework of quantum theory to explain several important but puzzling findings made in physical science at that time. Subatomic particles were not particles at all; they were probabilities of wave like and particle like behavior. Moreover, the tendency toward one or the other behavior appeared to be linked to the observer. These discoveries challenged the Newtonian worldview that the world was made of matter, was deterministic, and predictable. Moreover, it appeared that 'energy' replaced matter at the subatomic level. 'Probability' of location replaced spatial reference; 'uncertainty' replaced predictability. Moreover, an observer at the macroscopic level could instantaneously affect qualities observed at the microscopic level. Together, these discoveries pointed to a universe that could not be understood by reducing it to its parts and then by adding up those parts. Rather, it was a universe that seemed to be comprised of relationships between scales (i.e., atoms, molecules, cells, or objects, etc.). This concept pointed to a world that appeared to be better understood as an embodied whole. Thus, some scholars sought a science of holism rather than a science of parts. This marks the emergence of systems thinking in contemporary western science.

From Systems Thinking to Systems Science

Origins and Overview of General System Theory

The individual most credited with catalyzing systems theory in contemporary western science, is Austrian born biologist Ludwig von Bertalanffy (1901–72). He first presented his ideas at the University of Chicago in 1937. In 1968 he published his work in the book *General Systems Theory* (GST), for which he is perhaps best known. In it he presented two important concepts, a vision of one 'science of wholeness' and a detailed description of 'open systems'.

Bertalanffy's contemporaries included economist Kenneth Boulding (1910–93), physiologist Ralph Gerald (1900–74), and mathematician Anatol Rapoport (1911–2007), who, in collaboration with psychologist James Grier Miller (1916–2002), would institute the Society of General Systems Research in 1956. They were motivated by a vision of a general theory of all systems that could help scientists share information about commonalities across their disciplines without replacing discipline specific theories.

General System Theory and Open Systems

General system theory is a doctrine of principles about the commonalities of all systems, regardless of type and scale. The purpose of its genesis was to provide scientists from different disciplines with a theoretical framework from which to conduct research and share findings. As such, a general systems theory would reduce redundant research between disciplines and reduce the risk that scientists in highly specialized fields could overlook seemingly insignificant phenomena. Bertalanffy explicated the 'isomorphies' (structural similarities) and 'correspondences' (relationship with the system's environment) that systems appeared to have in common.

As a biologist, Bertalanffy was particularly interested in the paradox of how complex systems, such as living organisms, could manifest complexity (order) against a universal tide of entropy (disorder). This appeared to contradict the second law of thermodynamics (conceived by Sadi Carnot (1796–1832) and reworked by Rudolf Clausius (1822–88)). The second law of thermodynamics states that energy, while it may not be created or destroyed (first law of thermodynamics), will dissipate over time; that is, entropy increases toward its maximum in an isolated system. However, Bertalanffy noted that, contrary to being isolated or closed systems, living organisms were in fact 'open systems'. Bertalanffy explained that open systems can garner energy resources from their environment to build internal order despite the universal tide of entropy. The 'steady state' of a physical system or 'homeostasis' in living organisms refers to self regulating mechanisms maintaining system stability despite a constant flow through of energy in a fluctuating environment.

Defining Systems

Bertalanffy clarified that, most fundamentally, a system is comprised of at least two or more parts; each part is connected with at least one other part through observable relations; the integrity of the whole system depends upon the parts; and the emergent quality of the whole is not found in any of the parts. Bertalanffy emphasized that systems are not comprised merely of a sum of parts de fined by linear, causal relationships and described by Newtonian mechanics. In fact, he argued that a New tonian type, mechanistic system could only exist when the influence of the parts on each other were minimal. The systems Bertalanffy described, on the other hand, are comprised of interrelated parts defined by both linear and nonlinear relationships.

Advances in general system theory demonstrated that certain principles and laws governing the function of one system could be applied to others (isomorphism). Growth and competition are examples of isomorphic character istics in systems. Take, for example, a population of badgers and a system of railway lines. The unit – badger or railway line – is 'immaterial'. The exponential growth of, for example, the population of badgers or the system of railway lines is 'material'. It is material because the behavior of each system is determined according to a systemic 'relationship', which is further defined by its internal organizational structure, energy inputs, and the boundary of the system. Badger populations grow at a certain rate because of food energy input, while railway lines might grow similarly because of other energy inputs (i.e., materials, service needs of growing populations).

Systems within systems have also been characterized as 'hierarchies of systems'. At each level, a set of prin ciples and laws would appear that are unique from those found at lower levels. For instance, subatomic particles comprise atoms, which make up cells, which comprise humans, which live within societies, and so on. Moreover, unique qualities emerge at each level. Atoms obey principles of quantum theory, organisms obey principles of genetics, and societies obey principles of population growth, etc.

From Systems Thinking to Systems Theory to Systems Science

Together with Kenneth Boulding, Ralph Gerard, James Grier Miller and Anatol Rapoport, and Ludwig von Bertalanffy instituted the Society of General Systems Research in 1956. In keeping with the goal of general system theory, the society set out to further the unifi cation of the sciences and research into the isomorphies and laws governing systems of different types and scales. The society was renamed the International Systems Science Society (ISSS) in 1988 after the dialog about general systems broadened to what is better known today as 'systems science'. Systems science remains inter disciplinary in scope; however systems research has ar guably become more applied, and therefore, absorbed into individual disciplines and, as such, has lost some of its momentum as a single intellectual movement driven by theory, as it was in the 1950s.

On the whole, the work regarding general system theory carried the systems thinking movement to a new level. Relatively new theories developed, such as Laszlo's 'general evolution system theory'. Especially in the social sciences, developments include 'critical systems theory', 'liberating systems theory', and 'unbounded systems thinking'. In the natural, social, and physical sciences, systems theory is successfully applied in areas of 'systems engineering', that is, to construct organizational systems in a given environment; 'system dynamics', that is, to understand change in natural systems; or 'human sys tems', that is, to manage change or realize vision in value laden human systems. Through this development, from theory into practice, methodological approaches evolved, including 'hard' systems and 'soft' systems approaches. A detailed discussion of specific applications can be found in the section titled 'Systems analysis'.

'Hard' and 'Soft' Systems

'Hard systems' science is a diverse field concerned pri marily with controlling, managing, or constructing sys tems. Emerging during the World War II, it is probably best known for its application in 'systems engineering' and 'operations research'. 'Systems engineering' is con cerned with constructing, manipulating, and controlling physical, organizational, and informational systems for desired outputs in human society. For example, systems engineering has been applied to developments in tele phone systems. Attempts have also been made to apply systems engineering to social systems. As in all systems, the focus of systems engineering is on the relationship between the parts or elements, be they animate, like the integration of physical components, or inanimate, like the flow of information between individuals. Today, systems engineering focuses on systems of processes, products, or services and all their required inputs a situational context in order to accomplish a prescribed task or solve a problem. For example, systems engineering is applied to systems such as military systems, transportation systems, software engineering, and consumer product systems (i.e., automobiles or cameras). The purpose of applying sys tems engineering includes, for example, improving the safety, reliability, or the efficiency of a system, whatever that system is defined to be (i.e., a car or a telephone network).

'Operations research', on the other hand, is concerned with attaining an optimal efficiency or solution for any operation or organization. Similar to systems engi neering, operations research dates to World War II when the US and British military personnel engaged re searchers to seek optimal efficiencies and strategies for their military operations. Today, operations research is more closely linked to business and its management; yet it appeals to a wide arena of application, including communications, manufacturing, and public services. It is typically used to facilitate decision making, manage risk, and weigh alternatives, but above all, to increase efficiency. Both systems engineering and operations research focus on closed type systems.

By the 1970s, some scholars recognized that the quantitative and analytical approaches applied to closed systems were not adequate for inquiry involving human systems. For example, C. West Churchman (1913–2004) argued against the propensity of analytical systems planning to isolate and solve for perceived problems like inefficiency or productivity as opposed to focusing on human systems as integral, moral, and value laden wholes, wherein lie greater societal concerns (i.e., malnutrition or environmental degradation). In contrast to 'hard' systems, which tend to be evaluated or constructed according to single goals, human systems involve multiple and often competing objectives and interests.

Responding to this dilemma, Peter Checkland (born 1930) introduced a 'soft systems approach' for social systems research. It is an iterative approach to understand and then react to an identified systemic problem by learning about its facets, deciding on which actions would pose better alternatives, and then evaluating the outcomes. Checkland argued that, with the soft systems approach, further learning, deciding, and acting would take place because perceived problems in human systems are never solved for long, since humans systems are dynamic. Moreover, the iterative nature of the soft systems approach acknowledges the dynamic human system as fundamentally value laden by seeking meaning and solutions from the system itself, rather than imposing analytical solutions upon it.

Social Systems

Sociologists Talcott Parsons (1902–79) and Niklas Luhmann (1927–98) are widely credited with major advances in the general theory of social systems. Both sought a theoretical framework to encompass social interaction. Parsons saw social systems interwoven with cultural and personality systems to comprise the 'social action system', wherein human behavior exists within an environment comprised of physical and organic components. Like other systems thinkers he viewed social systems as open systems, with internal self regulating mechanisms that would retain the systemic integrity when faced with perturbations or disturbances (i.e., deviant behavior from within, or conflict from without).

Luhmann focused on communication as the mechanism binding human systems. He posed that communication within the human system was characteristically simpler than the total available information. The 'reduction of complexity' was made possible through the value laden human construct of 'meaning' such that the meaning humans attribute to information in the environment drives the process of its assimilation and further communication. Furthermore, the process of information selection gives the social system its identity.

Cybernetics

Origins and Overview of Cybernetics

Evolving concurrently with but independently from general systems theory, and motivated by the demand for more sophisticated technology in World War II, 'cybernetics' grew out of the desire to understand the role that sensory perception, memory, communication, and circular causality (feedback mechanisms) played in the purposeful behavior of living organisms, and to apply these concepts to make better machines. While the use of control systems was not new, the interdisciplinary field of cybernetics would go well beyond the study of artificial systems and their control and into the study of goal oriented living systems, including human knowing, or 'epistemology'. It would make significant advances in engineering, computer technology, cognition, and information science.

In the 1940s and 1950s key scholars of the growing cybernetics movement included neurophysiologist Warren McCulloch (1898–1969), engineer Jay Forrester (born 1918), biologist Humberto Maturana (born 1928), mathematician John von Neumann (1903–57), anthropologist Gregory Bateson (1904–80), electrical engineer and mathematician Claude Shannon (1916–2001), psychiatrist W. Ross Ashby (1903–72), and physicist Heinz von Foerster (1911–2002). Yet, it was Norbert Wiener (1894–1964), mathematician and philosopher from the Massachusetts Institute of Technology, who first defined and applied the name 'cybernetics' to the emerging area of study. Being derived from the Greek word kubernετες, meaning 'steersman', the fundamental tenet is that, as a steersman reacts to messages from the environment to stay on a given path, so an organism receives messages from its environment to which it reacts. In this manner, Wiener argues, communication and control are the most important factors that allow organisms to maintain homeostasis (stability) within a greater tide of growing entropy in the universe. The key mechanism of homeostasis is 'negative feedback', whereby a system utilizes information from the environment to limit the effects of change. Positive feedback mechanisms, by contrast, are the processing of environmental information as a catalyst for change.

Initially, cybernetics focused on these self regulatory mechanisms inherent in organisms and the application of this knowledge to exact better control over machines.

Interest hinged around questions like, 'how' does this operate and how can this operate better? This would come to be called, 'first order cybernetics'. By contrast, second order cyberneticists asked, 'why' is it like this and what does it mean? It focuses on aspects of the human experience such as language, communication, purpose, cognition, and epistemology. Advances in second order cybernetics are best known in education, management, cognition, education, and epistemology.

Contributions of Cybernetics

Today, the first and second orders of cybernetics are less clearly delineated. In total, the accomplishments of cy bernetics include 'intelligent machines' such as navi gational equipment, robotics, artificial intelligence, and, more recently, a deeper understanding of feedback mechanisms in economics. In addition, cyberneticists developed theories that have been subsequently applied to problems or issues in many disciplines. Some of these theories include information theory, game theory, and decision theory.

Proposed in 1948 by Claude Shannon, 'information theory' is concerned with communication – its quantity (how much information is heard) not its quality (the meaning of the message). Information theory holds that the transmittal of data follows the law of entropy, such that a minimum of information is required for the ori ginal message to be understood by the receiver, despite its dissipation during transmittal due to noise.

From the work of John von Neumann and Oskar Morgenstern in 1944, 'game theory' allows the behavior of competing interests in the context of limitations to be modeled mathematically. It is applied today in many disciplines of inquiry including political science, geog raphy, economics, and sociology. 'Decision theory' refers to a variety of analytical tools that may be applied to facilitate decision making in business and government. It factors in variables such as accurate information, risk, and alternatives.

Complexity Science

Origin and Overview of Complexity

'Complexity science' is the study of complexity in open systems. One of the early contributors to complexity theory was Belgian chemist and 1977 Nobel Prize winner Ilya Prigogine (1917–2003) who developed the 'theory of dissipative structures'. A comprehensive explanation of his work was written and published together with Belgian philosopher Isabelle Stengers (born 1948) in English in 1984 in the book titled, *Order out of Chaos*. Like Berta lanffy, Prigogine noticed that the biosphere is comprised of a great many systems that build complexity. Prigogine demonstrated that chemical systems can change spon taneously when internal fluxes become amplified by positive feedback and reach a critical point which he called the 'bifurcation point'. Prigogine demonstrated that, under a certain threshold, any perturbation will be absorbed by the system. That is, internal fluctuations occur but will not change the integrity of the system as a whole. Over a certain threshold, the internal fluxes amplify until the system reaches the bifurcation point. At this point, the system spontaneously reorganizes into an unpredictably different system, defined by different in ternal mechanisms. The flow of a stream provides a simple example. A pebble that is tossed into a stream does not change the nature of its flow; however, if the disturbance is great enough or accumulates to a certain threshold, the flow becomes turbulent. In tandem with important developments in nonlinear mathematics, in cluding dynamical systems and chaos theory, the theory of dissipative structures represents an early contribution to complexity science.

Complex open systems are self organizing, meaning they may adapt to environmental changes; self regu lating, meaning they remain characteristically stable while exploiting energy from their environment; and they exhibit emergent properties that are unique to the entire system as a whole. Moreover, the parts and pro cesses of which they are comprised are highly inter connected. As such, system complexity is defined by the amount of information that is required to describe the system. The more interdependent its parts and processes, the greater is the system's complexity. The science of complexity employs the methodological tools of com puter models, mathematics, conceptual frameworks, and theory to study the patterns, networks, and dynamics of systems.

Chaos Theory

Chaos theory describes the qualities of the point at which stability moves to instability or order moves to disorder. For example, unlike the behavior of a pendulum, which adheres to a predictable pattern a chaotic system does not settle into a predictable pattern due to its nonlinear processes. Examples of chaotic systems include the be havior of a waft of smoke or ocean turbulence. Chaotic systems are characteristically sensitive to initial con ditions. Therefore a slight change in the initial condition could have a vast impact on the outcome, as exemplified by the 'butterfly effect' and in the work of meteorologist, Edward Lorenz (born in 1917).

Advances in chaos theory and its mathematics are owed to physicist and mathematician Jules Henri Poin care (1854–1912), who used topological techniques to visualize mathematics. Chaos mathematicians in the 1960s would map the trajectories, for example, of a simple pendulum. This map or depiction would be called a phase space, corresponding to the coordinates of the movement. A simple pendulum swing would have a two dimensional phase space of velocity and angle. Once the movement is represented or mapped on the coordinates, a pattern appears. This pattern is called an ‘attractor’. The attractors of chaotic systems are called ‘strange attractors’.

At the time that René Thom (1923–2002) was developing the related ‘catastrophe theory’, a French mathematician Benoît Mandelbrot (born 1924) developed a geometric representation of everything from natural phenomena to the strange attractors of chaos, using ‘fractal geometry’. Some striking features of fractals include their remarkable similarity with patterns found in nature such as the crystal shapes of snowflakes, or the patterns in a fern frond. Another striking characteristic is that the shapes embedded in the fractal patterns can reappear at smaller and smaller scales.

Autopoietic Theory

The complexity that Chilean biologists Humberto Maturana (born 1928) and Francisco Varela (1946–2001) studied was different from Prigogine’s. Prigogine’s dissipative structure theory described stability and spontaneous change in living and nonliving entities. Maturana and Varela’s ‘autopoiesis theory’ described the characteristic processes that were fundamental to self organized, living organisms. Derived from the Greek words auto, meaning ‘self’ and poiesis, meaning ‘creation’, autopoietic systems are ones that are comprised of self creating processes. In other words, networks or relationships between components that are self referential and create the complexity of living organisms. Moreover, these processes serve the same function in the body (i.e., a cell) as in the mind (i.e., cognition). Components are part of self creating circular networks because they create other components in order to maintain themselves and the structure in its entirety. By contrast, systems that utilize energy to create complexity that is unlike the system itself is called ‘allopoiesis’, such as an assembly line whose components make cars, not itself. Maturana and Varela argued that the function of the organism was characterized by the nature of the self referential and self creating processes and networks. Working to better understand neural networks Maturana found that the same process applied. Maturana argued that cognition is a self referencing system whereby understanding is built upon previous understanding. What is understood by the mind through what the eye sees is not the absolute reality of the ‘outside’; rather, it is the expression of the brain’s neural networks that builds the experience of interpretation and understanding.

Resilience Theory

Ecologist C.S. Holling defined resilience as the ability of a system to remain functionally persistent by absorbing externally generated disturbances. His work led him to argue that it is not an inherent quality of stability – the ability to return to the original state following a disturbance – that defines the health of an ecosystem; rather, it is its ‘resilience’. This is an important distinction because, as demonstrated in the theory of dissipative structures, systems, in this case ecosystems, are comprised of interconnected self regulatory feedback mechanisms and fluctuations that occur naturally within them. It is these mechanisms and their inherent fluctuations that absorb change or disturbances. Positive feedback loops may amplify to the point that the integrity of the ecosystem is lost. At this point, the ecosystem is said to have lost its resiliency in the face of change. While ecosystem collapses might be caused by large disturbances, they are more typically caused by incremental disturbances or perturbations. Beyond its threshold, the ecosystem reorganizes into a state or ‘regime’, which is similarly characterized by internal self regulatory mechanisms. Moreover, the new regime is ‘persistent’, meaning, it is difficult to change back into the ecosystem it once was. A simple example is the regime shift of a savanna ecosystem that, due to the accumulative effects of overgrazing on the natural drought cycles experienced naturally, reorganizes into a desert ecosystem.

More recently, resilience theory has been applied to super systems of social, economic, and ecological systems. Some research suggests that regime shifts (collapse or irreversible change) in one system can cascade into other systems, affecting their resilience. It is possible for one system to collapse and cause the collapse of several others. Attempting to model human systems with ecological systems is challenging because it is recognized that human systems share some fundamental characteristics that natural systems do not: they are purposeful, they can look into the future to decide on an action today, and they imbue their experience with meaning. Researchers working within this paradigm are seeking new and innovative conceptual frameworks with which to study issues such as ‘sustainability’ in environment–society systems.

Contribution of Systems Theory and Complexity Science to Geography

Systems theory, viewing the social and physical worlds as an interconnected hierarchy of parts and processes exchanging matter, information, and energy, provides a set of conceptual and methodological tools for research in human geography. For example, L. Douglas Kiel reviews the contribution of, specifically, the theory of dissipative structures to the social sciences. And W.J. Coffey outlines the progression of research methods and the contribution of systems theory in urban systems research from static statistical analyses to more recent analyses of dynamic systems using chaos, complexity, and bifurcation theories.

Others have applied the concepts of systems theory directly to specific systems of interest in human geog raphy. For example, the research of Peter M. Allen, Guy Engelen, Michele Sanglier, and Robert Crosby furthers our understanding of spatial and evolutionary aspects of urban environments. Alan Wilson applied systems theo retic constructs to the study of transportation systems. Additionally, systems theoretical approaches have been undertaken by researchers such as Richard N. Adams and Debra Straussfogel, whose work with systems theory in human geography reveal, respectively, the relationship of energy to society and the systemic underpinnings of world system theory as applied to problems of develop ment. (Other noteworthy contributions of systems oriented research to human geography include early work by Robert Bennett, Richard Chorley, Graham Chapman, and R. Huggett.)

Yet, the advent of systems theory as a lens through which to examine issues related to human geography has also been criticized. For example, Peter Gould cogently argued that the theory of dissipative structures does not account for the breadth of variability of the human ex perience. And systems theory in the research of human behavior has been criticized for being used analogously without algorithmic substantiation. As Sunny Auyang aptly states, "The more crudely we look at a system, the more likely we are able to discern patterns" (1998: 328). However, this being said, with the advent of increasingly advanced computer modeling techniques, the ability to test pioneering theoretical frameworks continues to ex pand our understanding of the form and function of geographic systems.

Conclusion

Systems theory, the theory that open systems are com prised of interconnected elements that, as a whole, demonstrate emergent properties, and that as such, ad here to isomorphic laws and principles, emerged in contemporary science in the middle of the twentieth century. In a time of technological advancement, insti gated by World War II and in concert with discoveries in quantum physics that shook the foundations of New tonian mechanics, systems theory was embraced by many scholars who espoused the advantages of systems think ing rather than reductionism. Systems theory has been applied by scholars from diverse backgrounds – from those seeking efficiencies in their production network or organization to those seeking answers about the human condition by exploring value laden human systems. Consequently, over the breadth of systems inquiry in the last seven decades, an array of related theories emerged. Yet, these theories are held collectively by systems thinkers whose worldview is that the parts of a whole cannot be reduced to better understand the whole, making systems science the multifaceted and inter disciplinary science that it is today.

See also: Chaos and Complexity; Complexity Theory, Nonlinear Dynamic Spatial Systems; Networks, Urban; Neural Networks; Systems; Urban Modeling.

Further Reading

Adams, R. N. (1982). *Paradoxical Harvest*. Cambridge: Cambridge University Press.

Alisch, L., Azizighanbari, S. and Bargfeldt, M. (1997). Dynamics of children's friendships. In Eve, R. A., Horsfall, S. & Lee, M. E. (eds.) *Chaos, Complexity, and Sociology*, pp 163 181. Thousand Oaks, London, New Delhi: Sage Publications.

Allen, P. M. (1997). *Cities and Regions as Self Organizing Systems*. The Netherlands: Gordon and Breach.

Ashby, R. (1952). *Design for a Brain*. New York: Wiley.

Auyang, S. (1998). *Foundations of Complex System Theories in Economics, Evolutionary Biology, and Statistical Physics*. Cambridge: Cambridge University Press.

Banathy, B. H. (1996). Systems inquiry and its application in education. In David, H. J. (ed.) *Handbook of Research for Educational Communications and Technology*, ch. 3. New York: Macmillian.

Bateson, G. (1972). *Steps to an Ecology of Mind*. New York: Ballantine.

Boulding, K. (1956). *The Image: Knowledge in Life and Society*. Ann Arbor, MI: University of Michigan Press.

Capra, F. (1982). *The Turning Point*. New York: Simon and Schuster.

Checkland, P. B. (1981). *Systems Thinking, Systems Practice*. New York: Wiley.

Churchman, C. W. (1971). *The Design of Inquiring Systems*. New York: Basic Books.

Coffey, W. J. (1998). Urban systems research: Past, present, and future. A panel discussion. 'Urban Systems Research: An Overview'. *Canadian Journal of Regional Science,* 21(3), 327 364.

Engelen, G. and Allen, P. M. (1986). Modelling the spatial distribution of energy demand for the province of Noord Holland: Towards an integrated approach. *Systemi Urbani* 8, 241 261.

Gleich, J. (1987). *Chaos: Making a New Science*. New York: Penguin Books.

Gould, P. (1987). A critique of dissipatives structures in the human realm. *European Journal of Operational Research* 30(1), 211 221.

Hillier, F. and Liebermann, G. (1995). *Introduction to Operations Research* (6th edn.). New York: McGraw Hill.

Holling, C. S. (1973). Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics* 4, 1 23.

Kiel, L. D. (1991). Lessons from the non linear paradigm: Applications of the theory of dissipative structures in the social sciences. *Social Science Quarterly* 72(3), 431 442.

Laszlo, E. (1972). *Introduction to Systems Philosophy: Toward a New Paradigm of Contemporary Thought*. New York: Gordon and Breach.

Laszlo, E. (1987, 1993). *Evolution: The Grand Synthesis*. Boston: Shambhala.

Luhmann, N. (1995). *Social Systems*. Stanford: Stanford University Press.

Parsons, T. (1951). *The Social System*. Glencoe, IL: Free Press.

Prigogine, I. and Stengers, I. (1984). *Order out of Chaos*. New York: Bantam Books.

Sage, A. (1992). *Systems Engineering*. New York: Wiley.

Sanglier, M. and Allen, P. M. (1989). Evolutionary models of urban systems: An application to the Belgian provinces. *Environment and Planning A* 21, 477–498.
Straussfogel, D. (1997). Redefining development as humane and sustainable. *Annals of the Association of American Geographers* 87(2), 280–305.
Varela, F. G., Maturana, H. R. and Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization, and a model. *Biosystems* 5, 187–196.
von Bertalanffy, L. (1968). *General System Theory*. New York: George Braziller.
Weiner, N. (1948). *Cybernetics or Control and Communication in the Animal and the Machine*. Cambridge, MA: MIT Press.
Wiener, N. (1950, 1956). *The Human Use of Human Beings*. New York: Doubleday and Company Inc.

Relevant Websites

http://www.complexsystems.org/
Institute for the Study of Complex Systems.
http://www.incose.org/
International Council of Systems Engineering.
http://isss.org/world/index.php
International Society for the System Sciences.
http://pespmc1.vub.ac.be/SERVER.html
Principia Cybernetica Web.
http://www.resalliance.org/1.php
Resilience Alliance.
http://pespmc1.vub.ac.be/ASC/INDEXASC.html
Web Dictionary of Cybernetics and Systems.

T

Taylor, G.

J. M. Powell, Monash University, Melbourne, VIC, Australia

Glossary

Environmental Determinism The idea that the natural environment controls human activities.

Physiography In the context of this article, a loose, early term for descriptions of nature or physical geography - occasionally so loose as to permit references to environmental controls on economic development.

Possibilism In the context of this article, the notion that the physical environment allows human communities to select from a range of alternative responses.

Resource Management Decision-making for the determination and implementation of policies concerning the use and conservation of natural resources, embracing multidisciplinary and interdisciplinary approaches.

Social Darwinism The application of Darwin's evolutionary principles to economic, political, and social contexts.

Sustainability In the context of this article, concerning the adoption of modes of development capable of meeting current needs without jeopardy to environmental health and the needs of future generations.

Taylor, Griffith (1880–1963)

Taylor was born in Walthamstow, England, but his closest family origins were in the Lancashire's cotton districts, near Manchester. His name commemorated his maternal grandfather, Thomas Griffiths, using 'Griffith' as a fore name. His father, James, was a well traveled chemist and mines manager. In 1883 Griffith Taylor emigrated with his family to Sydney and continued his education at elite private schools and at Sydney University. Following successful studies in science and mining engineering, he undertook postgraduate training in geology with the celebrated Antarctic explorer, Professor Edgeworth David, and was subsequently awarded an 1851 Science Research Scholarship to Emmanuel College, Cambridge. Increasingly exposed to the influence of internationally prominent geologists, geomorphologists, and regional geographers, he commenced an ambitious writing trajectory, including *Australia in Its Physiographic and Economic Aspects* (1911), and made a distinguished personal contribution to the scientific work of Captain Scott's renowned, ill fated *Terra Nova* expedition to Antarctica. Robust and wiry, of above average height, staunchly teetotal, already balding, Taylor was recognized as an all round scientist and enterprising team player, who seemed inclined to the occasional rush of blood. Above all, he was highly valued for his courage, endurance, and quirky humor, characteristics that would be repeatedly tested after his return to Australia.

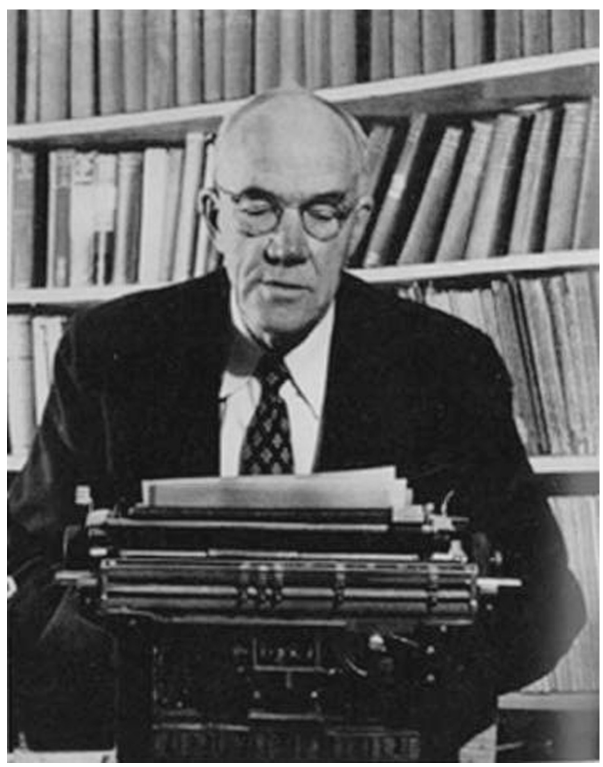

The Cambridge and Antarctic adventures led directly to his crucial appointment to Australia's new federal weather service as a major environmental researcher (officially, as a 'physiographer'). The demanding brief was heavily influenced by wartime exigencies, nationalist–imperial anxieties, and an emerging focus on national development policy, but the energetic, pugnacious Taylor seemed well suited to the challenge. His rapidly expanding research schedule included the significance of environmental limits on Australia's ability to accommodate a large increase in total population, the inadvisability of massive new investment in tropical settlement, and the promotion of a national resources atlas. It would also accelerate Griffith Taylor's return to academia. In the interim, although his forthright pronouncements were famously caricatured as unpatriotic, pessimistic 'environmentalism' in Britain and Australia, they resonated very well with North American luminaries, such as Isaiah Bowman. The *Geographical Review* provided a very respectable outlet for Taylor's work, and in the 1920s published a series of his papers on subjects as diverse as the geography of race, culture, and language.

At the age of 40, after applying unsuccessfully for a range of senior university positions in meteorology, geology, and geography, Taylor welcomed (Sir) Edgeworth David's inspired maneuvers to bring him back to Sydney as the founding head of the new department of geography. He took up his appointment in 1921. Inevitably, given Taylor's national notoriety, the new Sydney interlude was tempestuous. Taylor was a lively and enthusiastic teacher, whose unflinching commitment to geography's underestimated potential for civic engagement collided with powerful campus and political interests. While carefully attending to the presentation of his Antarctic work, he interrogated key intellectual and political positions of the day – on racial origins and dispersals, the potential targeting of northern Australia by reportedly overcrowded Asian communities, and aridity as a primary obstacle to settlement intensification. He never gave quarter to disciplinary purists: certainly not to threatened classicists who branded him a philistine, or to what he came to regard as naïve 'possibilism'. Furthermore – and dangerously, as it would prove – he chose to air some aggressive opinions in the popular media, rather than to confine himself to peer scrutiny in academic journals.

Taylor tended to exaggerate his status as lone crusader. His difficulties partly derived from naïve enthusiasm, excess of opportunity, and a late arrival's anxiety to make a mark in academia. Vaguely acknowledging social Darwinism, he repeatedly sought to clarify 'Nature's Plan' and to define a role for science in public affairs. Antarctic fame and an exceptional employment experience had encouraged his appetite for individualistic pioneering; on the other hand, the deceptively plum government position had set enormous tasks which were incapable of solution from any specialist viewpoint. He may have felt entitled to make use of other researchers' work as circumstances demanded, but although he inserted a patina of novelty that seemed for a time instructive, detractors found more art than science.

In particular, his excursions into racial origins and distributions reprised Ellsworth Huntington's deterministic ideas in human geography, the cephalic index work of W.Z. Ripley, and the global dispersal themes introduced by palaeontologist and zoogeographer, W.D. Matthews. Taylor preferred to extrapolate by means of enterprising adaptations of a series of familiar diagrammatic techniques, drawn mainly from meteorology (isobars), and geology (stratification). This gave rise to a celebrated 'zone and strata' (or Migration Zone) theory, introduced in a number of succinct articles before its ultimate elaboration in *Environment and Race*. It is worth noting, however, that he also exhibited a distinctive, arguably innate facility for freehand draughtsmanship. It had been honed by his undergraduate courses in geology and mining engineering, during Antarctica's arduously remote and testing silences, as a facet of his relentless advocacy of field observation, and in coping with a seriously understaffed and under resourced teaching milieu.

Taylor's opponents attacked his purportedly 'deterministic' stance on Australia's potentials and his (remarkably accurate) forecast of a total population of about 20 million by AD 2000. They also undervalued critical, supplementary explications of a rational sequence of regional development, particularly within his 1923 address to the Australasian Association for the Advancement of Science. Increasingly vilified in the press and by campus enemies, in 1928 he resigned his Sydney appointment and left for the University of Chicago.

The brief Chicago interlude strengthened his contact with Bowman, Barrows, and many other influential geographers, provided increased opportunities for extensive fieldwork in the US, Europe, and Scandinavia, and generally increased his exposure to geography's emerging presence as a university subject. His major writings maintained broadly familiar emphases on Antarctica, Australia, and race. In 1935 he was appointed founding professor of geography at the University of Toronto; on his retirement in 1951 he returned to Sydney.

More comfortable with Canadian ways and institutions, Taylor settled down to build a fine teaching and research department, improved the subject's standing in Canadian schools, continued his recent Chicago work on environmental influences in cultural and urban geography, and resumed his older studies of climatic controls in regional development. His Toronto inaugural lecture, *Geography – The Correlative Science* (1935), reflected a few of the philosophical and pedagogical fragments gleaned at Sydney and Cambridge, but also borrowed from an under remarked appreciation of the liaison role fash ioned for geography by Chicago's sophisticated ad ministrators. On the whole, the late career evidence finds for modulated speculation, more focused, less ana chronistic bibliographical referrals, and gradual relax ation into established academe, especially during wartime and in the early postwar period.

Retrospectively, *Environment and Nation* (1936) has been considered another wild foray into European his tory, but both *Urban Geography* (1949) and (his edited) *Geography in the Twentieth Century* (1951) seem compara tively mainstream. *Our Evolving Civilization* (1946) was generally familiar too, except that it was proffered as a timely introduction to 'geopacifics', or peace studies in formed by a human (or humane) geography, a theme to which he would return in his retirement. More indica tively, *Australia* (1940) had recommended 'stop and go determinism', visualizing human intervention as a form of traffic control – altering the rate, rather than the dir ection, of progress. Similarly, *Canada* (1947), a very marketable text, actually disputed a prevailing mood of pessimism in that country.

Until the end of his life, Taylor continued to picture himself as an 'enfant terrible', but he had never been as brazenly deterministic as his enemies charged. Somewhat unkindly, it could now be said that his initial adoption of a deterministic stance mainly provided a launching pad for flailing attacks on themes which interested him. Further, his indulgence in controversy contributed to a much wider determinism–possibilism debate which consumed a great deal of expensive attention at all levels of the subject. If his early career exemplified pioneering adaptability, his notoriety ensured that his adopted sub ject contrived simultaneously to struggle against powerful detractors while rejoicing in his achievements. Taylor cast a particularly large shadow in both Australia and Canada, but arguably the Australian legacy is the more problematical.

He encountered more sympathetic company in the far larger pool of North America, and as geography itself found secure institutional footings around the world his authorial persona became more conventional. In 1940 he was elected the first non American President of the As sociation of American Geographers. His writings became more patient than provocative, except insofar as they retained a predilection for startlingly idiosyncratic dia grams, including the usual 'griffograms', much to the chagrin of conservative readers and editors. Before leaving Toronto in 1951 for his retirement in Australia, he was elected President of the new Canadian Associ ation of Geographers. In 1954 he was made a Fellow of the Australian Academy of Science, the first geographer in that fraternity. He was also elected founding president of the Institute of Australian geographers, in 1960; the Institute now awards a major medal named in his honor. Sydney University awarded him an Honorary D.Litt., 30 years after he had left that campus in some despair. Colleagues and graduate students alike generally found 'Grif' (also 'Griff') accessible, interesting, and constructive, with a low tolerance for trivialities that barely changed down the years: apparently, when asked to recommend a good hotel, he would invariably reply – 'take the first left outside the station, then the first right, and it's the first one you'll see'. Throughout his career he championed the role of geographers in public policy, and in Canada, as in Australia, his reinforcement of a tradi tional environmental bias in school geography assisted the retention of British as opposed to American repre sentations of the subject. By the early twenty first cen tury, some of the controversial Australian works of Griffith Taylor were applauded by environmentalists, national planning advocates, and geography teachers as prescient examples of academic geography's subsequent engagement with resource management, environmental management, and sustainable development.

See also: Bowman, I.; Darwinism (and Social Darwinism); Determinism/Environmental Determinism; Environmentalism; Possibilism; Resource Management, Rural; Sustainable Development.

Further Reading

Powell, J. M. (1979). Thomas Griffith Taylor 1880 1963. *Geographers: Bibliographical Studies* 3, 141 153.

Powell, J. M. (1993). *Griffith Taylor and 'Australia Unlimited'*, The John Macrossan Memorial Lecture, 1992. St Lucia, QLD: Queensland University Press.

Powell, J. M. (1991). *A Historical Geography of Modern Australia. The Restive Fringe*, pp 121 149. Cambridge: Cambridge University Press.

Sanderson, M. (1988). *Griffith Taylor. Antarctic Scientist and Pioneer Geographer*. Ottawa, ON: Carleton University Press.

Strange, C. and Bashford, A. (2008). *Griffith Taylor: Visionary, Environmentalist, Explorer.* Canberra: National Library of Australia Press.

Taylor, G. (1922). The distribution of future white settlement. *Geographical Review* 12, 375 402.

Taylor, G. (1935). Geography The correlative science. *Canadian Journal of Economics and Political Science* 1, 535 550.

Taylor, G. (1936). The zones and strata theory A biological classification of races. *Human Biology* 8, 348 367.

Tomkins, G. S. (1967). *Griffith Taylor and Canadian Geography*. Unpublished Ph.D. thesis, University of Washington, Seattle, WA.

Winlow, H. (2006). Mapping moral geographies: WZ Ripley's races of Europe and the United States. *Annals of Association of American Geographers* 96, 119 141.

Taylor, P.

J. Glassman, University of British Columbia, Vancouver, BC, Canada

Glossary

Hegemony In general, a form of domination; after the Italian social theorist Antonio Gramsci, a form of leadership in which a particular social group exercises power as much by gaining consent from subordinate groups as by exercising repressive force.

Modernization Theory An approach to development that emerged in the 1950s and 1960s, especially within fields such as economic history, political science, and sociology, which suggested that every society can and should go through a more or less teleological process of change from largely rural, agrarian, and premodern to urban, industrial, and modern.

Neoliberalism A doctrine, based loosely on an updating of classic nineteenth-century liberalism and emerging in the 1970s and 1980s, according to which the best political and economic outcomes can be procured through minimizing the role of states and allowing market activities considerable scope, including through encouragement of global free trade.

World Cities A notion developed in the 1980s, suggesting that particular cities, such as New York, London, and Tokyo, play crucial roles in articulating the global economy, thus serving as far more than national economic centers.

World Systems Analysis A form of analysis developed by historical sociologist Immanuel Wallerstein in the 1970s, built around the claim that contemporary societies can only be understood if they are analyzed as part of a global capitalist system, one that has developed through very long-term historical processes, and is inherently hierarchical.

Taylor, Peter (1944–)

Peter James Taylor is an influential political geographer and world systems theorist whose work has been read widely in geography and other disciplines, both in his home country of England and internationally. Taylor was born in 1944 and grew up in Calverton, near Nottingham, England, receiving a BA in Geography from Liverpool University in 1966 and a PhD in Geography from Liverpool in 1970. He taught at Newcastle University from 1968 to 1995, and since then he has taught at Loughborough University, where he co founded the Globalization and World Cities research network (GaWC) in 1999 and served as editor of the prestigious journal *Political Geography*. During 1990 Taylor was a research associate for a MacArthur Foundation project on comparative hegemonies at the Fernand Braudel Center of the State University of New York, Binghamton, a project headed by Immanuel Wallerstein, founder of world systems analysis. In recognition of his scholarship, Taylor was elected as a Fellow of the British Academy in 2004. He has also served as a local councilor for Labour in the Northeast of England, and he has always displayed a strong interest in the political implications of his scholarly work.

Taylor's academic work can be seen as both contributing a world systems perspective to geography and as bringing a more explicit geographical perspective to world systems analysis. Where the latter is concerned, Taylor has developed arguments about states and territoriality – including their changing forms over time – that have helped elaborate the framework developed by Wallerstein. The notion that the world system is composed of two types of entities – global commodity chains that are nonterritorial but transnational in scale, and territorial states that are national in scale – owes much to Taylor's writings.

Where Taylor's development of a world systems perspective within geography is concerned, three major areas of influence can be mentioned. Taylor's early work was in fields such as electoral geography, and as he became interested in world systems approaches, Taylor applied insights from this analysis to politics and elec tions. As one example, he developed an argument that in order to understand differences in the ways electoral processes and political party structures work in different parts of the world, it is necessary to understand the positions of various national polities within global pro cesses of capital accumulation. Political elites within countries of the Global South, down the global hierarchy of capitalist accumulation from the wealthiest capitalist countries, lack the resources necessary to effectively buy off political opposition groups such as labor parties – at least to the same extent as elites in the Global North. As such, political systems of the Global South are much less likely to have liberal democratic structures and will in stead tend toward authoritarianism and/or corruption, not because of the failures of leaders or polities but be cause of the differing structural position of countries in the global hierarchy. Such a view is very much at odds with the arguments characteristic of modernization the ory and neoliberalism, which see political structures as formed largely by forces internal to the state and society in question.

A second area in which Taylor's work has strongly influenced geography is in his use of world systems ap proaches to reconfigure arguments regarding the struc tures of states and their roles in international relations and the international political economy. Taylor insists that states are formed not only by the internal processes at work within them but by the external relations in which they are embedded – what can be called their 'interstateness'. This, again, makes it necessary to exam ine the structure and functioning of states, including their varying forms of territoriality, within the context of the world system. Taylor has taken this project forward by examining the rise and decline of states that in different periods have exercised global 'hegemony', a concept he adapts and develops from the Italian social theorist An tonio Gramsci, as well as from contemporary neo Gramscian scholarship.

The third area in which Taylor's work has strongly influenced geography is in his arguments regarding 'world cities' – cities that play crucial roles in global commodity chains and transnational flows of capital. Taylor has linked the analysis of world cities directly to a world systems framework, and has also worked to over come what he and others have characterized as the lack of adequate empirical evidence regarding the structure of the world cities network. Taylor has carried this last project forward through the formation of the GaWC, an international, collaborative scholarly venture based at Loughborough. Taylor and other GaWC contributors – notably Jon Beaverstock – have developed large data sets documenting various types of connections between branches of firms across the world city network, as well as global urban connections for organizations such as non governmental organizations. Taylor has used these data to support his argument that the recent phase of 'glob alization' has tended to undermine the territorial integ rity and governability of nation states, without however making national states irrelevant to processes of global capital accumulation.

Along with these contributions to scholarly work in geography, Taylor has been an active participant in projects for transforming approaches to higher education. Taylor was a member of the Gulbenkian Commission, headed by Wallerstein and given the mandate of exam ining historical changes and current needs in the social sciences. The Gulbenkian Commission's 1996 report, *Open the Social Sciences*, argued that the conventional, historically developed distinctions between social science disciplines are becoming increasingly obsolete and need to be replaced by approaches that are more transdisci plinary – and indeed geographical, including transcend ing the traditional nation state centrism of the social sciences. The report also suggested that any hard and fast distinction between the physical and the social sciences is becoming problematic and should be reconsidered in light of recent advances in the physical sciences. This, too, seems to highlight the potential of a geographical perspective, with its strong emphasis on human–en vironment interactions. Though Taylor and the other authors of the report note that geography became a poor cousin to other social science disciplines in the twentieth century, the implications here seem to be that a revived geographical perspective has much to contribute to a more robust transdisciplinary approach for the twenty first century.

See also: Electoral Geography; Hegemony; Political Geography; World/Global Cities.

Further Reading

Taylor, P. J. (1981). Political geography and the world economy. In Burnett, A. D. & Taylor, P. J. (eds.) *Political Studies from Spatial Perspectives*, pp 157 172. New York: Wiley.

Taylor, P. J. (1982). A materialist framework for political geography. *Transactions of the Institute of British Geographers NS* 7(1), 15 34.

Taylor, P. J. (1986). An exploration into world systems analysis of political parties. *Political Geography Quarterly* 5(4) (supplement), 5 20.

Taylor, P. J. (1991). Political geography within world systems analysis. *Review* 14(3), 387 402.

Taylor, P. J. (1994). The state as container: Territoriality in the modern world system. *Progress in Human Geography* 18(2), 151 162.

Taylor, P. J. (1995). Beyond containers: Internationality, interstateness, interterritoriality. *Progress in Human Geography* 19(1), 1 15.

Taylor, P. J. (1996). *The Way the Modern World Works: World Hegemony to World Impasse*. London: Wiley.

Taylor, P. J. (1999). *Modernities: A Geohistorical Interpretation*. Cambridge: Polity and University of Minnesota Press.

Taylor, P. J. (2000). Havens and cages: Reinventing states and households in the modern world system. *Journal of World Systems Research* 6(2), 544 562.
Taylor, P. J. (2000). World cities and territorial states under conditions of contemporary globalization. *Political Geography* 19(1), 5 32.
Taylor, P. J. and Flint, C. (2000). *Political Geography: World Economy, Nation State and Locality* (4th edn.). Harlow: Prentice Hall.
Taylor, P. J. (2004). *World City Network: A Global Urban Analysis*. London: Routledge.
Wallerstein, I., Juma, C. and Keller *et al*. (1996). *Open the Social Sciences: Report of the Gulbenkian Commission on the Restructuring of the Social Sciences*. Stanford: Stanford University Press.

Relevant Websites

http://www.lboro.ac.uk
Peter J. Taylor's Website at Loughborough University, Department of Geography.

Technological Change

G. Norcliffe, York University, Toronto, ON, Canada

Glossary

Invention The act of 'invention', which is often portrayed in film and fiction to consist of the lone experimenter having eureka moments in a macabre laboratory, is today an increasingly commercialized and instrumental process as corporations and research institutions search for specific solutions to prioritized problems. There still are some serendipitous discoveries, but for the most part invention is the result of a structured and commercialized process.

Innovation It occurs when inventions are commercially applied. This, too, has become a progressively more rational process as an innovator tries to assess the costs of deploying new technology versus the increase in the revenue stream that it is expected to generate.

Imitation and Diffusion Processes whereby successful technologies are copied – sometimes quite legally through franchises, and various technology-sharing arrangements, or because they are public knowledge not protected by patents – and sometimes illegally in clandestine operations which pirate other patented technologies. Imitation and diffusion are more important, economically, than invention and innovation.

Learning This is essential to technical change. For instance, learning how to write machine code is important to developing new software, as much as understanding gene maps is to bioengineering.

Codified Knowledge Technical information compiled in written or recorded form as text, equations, and tables, which make it easy to transfer, if necessary, over long distances.

Tacit Knowledge Knowledge, accumulated by people through experience, is difficult to codify. Sometimes called 'know-how', people who possess tacit knowledge may be only vaguely aware of it. For instance, the author has been stuck on page 1 of a juggling manual (i.e., codified knowledge) for many years, patiently waiting for an expert to pass on to him the tacit knowledge needed to master the art. The transfer of tacit knowledge normally requires trust and close personal connections.

Neighbourhood Effect This effect appears to influence technical change in a traditional closely knit rural society, and also (perhaps more surprisingly) in modern networked society, so that persons located close to an innovator are more likely to adopt a new technology than those located further away. Trust established by face-to-face contact may be important for this process; indeed science and technology parks are partly based on this premise.

Networks In modern industrial society, technology – which may be privately owned and protected by patents – diffuses via social and corporate networks of suppliers, branch plants, project teams, joint-ventures, venture capital initiatives, trade shows, trade magazines and the media. Important technical information may be exchanged face-to-face (especially if trust is an issue), and also via the web, phone networks, and printed media. University and public research institutions also connect with these networks, which have both global and local dimensions. The global dimension is illustrated by China, which is eagerly entering into international joint ventures that bring the latest foreign technologies into its many new science parks. Locally, many technopoles and metropolitan areas have become hotbeds of technological innovation due to (1) new technical knowledge moving through dense contact networks; (2) skilled labour moving locally from firm to firm; (3) the fission and fusion of small firms located within the technopole (Silicon Valley is the archetype of this model); and (4) complex networks of suppliers and sub-contractors.

Technological change is crucial to economic development, with innovation occurring in clusters that generate long waves of growth concentrated in specific regions where invention, innovation, imitation, and learning – four distinct stages of technological change – are promoted by access to codified and tacit knowledge. Protecting intellectual property has allowed industrialized nations to generate large technology rents. Technology is both socially and geographically constructed in favoured 'spaces of science' where users and creators of technology interact intensively. Technological change occurs in spaces with good infrastructure, intellectual capacity, and possibilities for networking aided by proximity and neighborhood effects. Extensive recent privatization of technology allows capital to reap substantial benefits. In a neoliberal age, the state needs to play an active role to narrow technology gaps between poor and rich regions, and manage powerful new technologies that allow humans to alter nature in a profound way.

Introduction and Overview

Although other living creatures have developed simple tools, *homo sapiens* is the only species to have developed sophisticated technologies that in recent years have become increasingly hierarchical and interconnected as one technology builds on, and interrelates with another. It is the development of technology, including both the 'tools' used to produce goods and services, and the 'crafts' or processes applied to the production process that most profoundly distinguishes humans from all other species. The increasing speed of innovation and the sophistication of contemporary technology provide humans with the power extensively to reconstruct nature; indeed – and paradoxically – we now have the capacity to engineer the demise of the human species.

In the premodern era, technology was developed locally and in a somewhat haphazard way, such that, in the absence of means to codify knowledge, a number of important breakthroughs were lost, only to be reinvented centuries later. The development of written script permitted the (rather laborious) storing of knowledge, but it was the invention of moveable type by Johannes Gutenburg in 1438 that initiated the printing, codifying, and dissemination of technical knowledge. From that point on, technological innovation has progressed at an accelerating rate, particularly during the Modern Era.

Technological Change and Economic Growth

In the 1930s and 1940s, Joseph Schumpeter stressed the importance of technological innovation to economic growth, coining the famous phrase 'creative destruction' to describe how economic growth results from entrepreneurs and corporations being willing to discard old technologies and innovate. Robert Solow's neoclassical growth model, which attributed about 80% of US economic growth to technical change, and the remainder to increases in inputs of labor and capital, broadly confirmed Schumpeter's ideas. It is now widely accepted that technical change is central to economic growth and the production of wealth in the modern world, although Solow's model assumed that new technology was exogenous to the system and available to producers, whereas – as will be stressed later – much new technology today is an endogenous good, held privately by firms who seek to extract large technology rents from their proprietary knowledge.

Technological changes do not, however, appear to arrive evenly, but rather (in some economic historians view) to be clustered around interconnected sets of innovations, known as long waves or Kondratieff waves lasting from 45 to 60 years. The nature of long waves is considered to vary with time and place as laggard regions catch up, and each region plays to its comparative advantages, but they are all marked in the first half of the wave by a rising economic boom based on new technologies which eventually results in overinvestment and overcapacity, coupled with market saturation, leading to a major recession. The pattern seen to describe western economic development since the 1780s recognizes five waves (**Table 1** and **Figure 1**).

Table 1 Long waves: Dates and key industries

First Wave	1780 1840	iron; steam engines; cotton
Second Wave	1840 1890	steel; railways; steamships; machine tools
Third Wave	1890 1940	electrical engineering; automobiles; chemicals
Fourth Wave	1940 1990	petrochemicals; aviation; computers
Fifth Wave	1990 ?	information technology; biotechnology; nanotechnology

The concept of a 'carrier wave', as propounded by Hall and Preston, relates closely to the theory of long waves. They see key technologies leading economic booms and carrying economic development forward for two or more decades, until the market is saturated and the boom runs out of steam. Such carrier waves are geographically concentrated; good examples including the technology led cotton boom of the late eighteenth century in Manchester, UK, the railroad boom of the nineteenth century centered on Pittsburgh, and the space technology boom of the late twentieth century concentrated in Orange County, California. One of a number of criticisms leveled at the long wave model is that it tends to compartmentalize technological change, whereas Norcliffe stresses continuity between waves, as bridging innovations link one wave with the next. The bicycle, for instance, was launched in the age of machine technology, but in turn – as stamping and pressing technologies were developed – laid some of the groundwork for the mass production of automobiles in the following long wave. Similarly, microprocessors, developed late in the fourth long wave, have had their main impact during the current fifth wave.

Intellectual Property

In the premodern world, technology was a public good generally shared in a haphazard way among those who happened to become familiar with the new tools and methods. The first recorded patent, granted by Henry VI to John of Utynam in 1449 for stained glass windows, set a precedent: the state could create a monopoly protecting an entire industry, or protecting an inventor's technique. From these beginnings, the system of patents spread through the Western world. The first US patent act dates

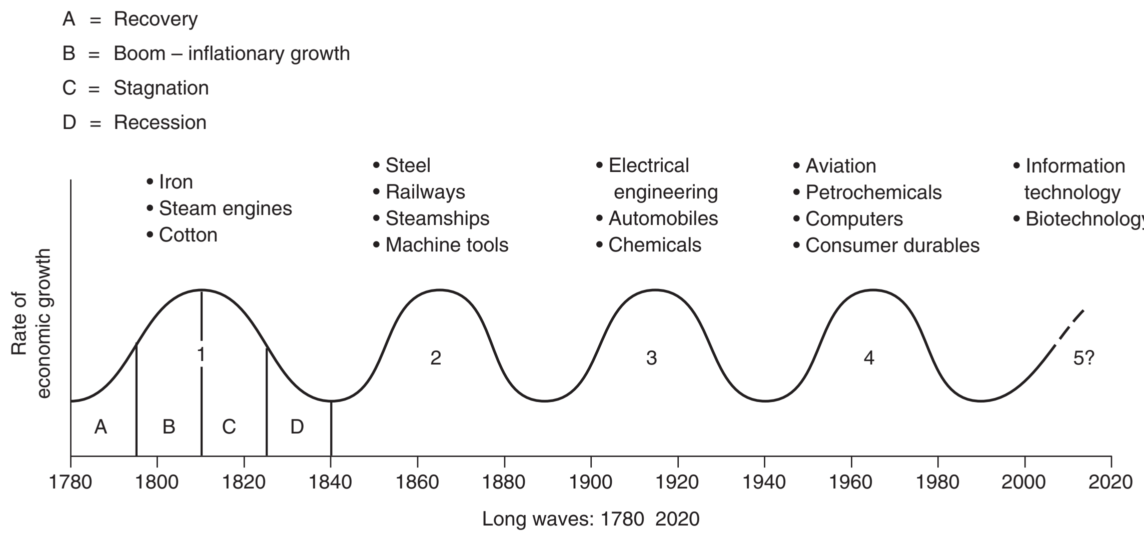

Figure 1 Long waves of economic development 1780 2020.

to 1790. As patent legislation evolved, it ceased to protect whole industries, and was restricted to specific written descriptions of processes or products, protected for the inventor during a statutory time period. The earlier requirement that a patented good be broadly beneficial to society was also dropped – wisely so since the best application of many technical innovations is not always immediately obvious.

In the modern industrial age, holding patents gives the patent owner a monopoly for a specified time period. Fortunes have been created by those who hold key patents, with family names like Nobel (inventors of dynamite and detonators) achieving wide recognition, partly through applying their wealth to philanthropic purposes. It follows that owning key technologies through patents and copyright, or simply through secrecy, may yield huge economic rents because the owners may set the price of technological change. Bill Gates, for instance, who is reported to be the world's richest person, has a major interest in Microsoft, which currently holds the key patents for the world's most popular computer operating systems.

In the current industrial era, the origins of patents are not randomly distributed, but are concentrated in industrial regions. This in turn reflects on the geography of technical innovation, which also tends to be concentrated in innovative industrial regions. In a country such as Canada, far fewer patents have originated in agricultural and resource regions than in major urban areas in the industrial heartland. This geographical concentration of patent activity has evolved in the twentieth century, with the research laboratories of major firms and universities, and smaller enterprises located in technopoles and science parks, creating an increasingly large share of intellectual property.

In recent decades, the ownership of intellectual property has been an immense advantage for the industrialized nations, which have been able to protect a vast array of technological innovations that are exported to the Global South and Newly Industrialized Countries at high prices, including (controversially) AIDS retroviral drugs. The recent stubbornness of the income gap between rich and poor countries appears to have been materially affected by the substantial economic rents that the rich industrialized countries derive from the intellectual property they own, and the high prices that the poorer nations often have had to pay to use the newest technologies. Of relevance here is the Marrakesh Agreement concluding the Uruguay Round (1986–94) of the General Agreement on Tariffs and Trade (the GATT), which included language on the protection of intellectual property, and penalties for countries that break these copyright protections.

The Social and Geographical Construction of Technology

To simplify a rather complex debate, it was widely assumed until the middle of the twentieth century that new technologies were created in the minds of great inventors who were almost exclusively male, and possessed remarkable powers of insight. Over a long period, many countries have been boosting this version of history by placing statues of their own famous inventors in visible places, by naming streets, schools, and universities for them, and placing their images on stamps and bank notes. In effect this is a version of Whig history that envisages great inventors creating a better modern world by their own efforts. This position was espoused by a group of historians whom Wiebe Bijker calls 'internalists', which is to say, those who focus on technological details internal to an innovation. On the other hand stand the rival camp whose followers argue for the social construction of technology (SCOT), focusing their attention on the social settings in which specific technologies have been developed and the ways users have often adapted innovations for their own purposes. Bijker refers to this group of social historians as 'contextualists' because they are interested in the many nontechnical elements that surround technology. A recent contribution to this approach by Oudshoorn and Pinch, entitled *How Users Matter*, moves the focus quite explicitly from the inventors to the users of technology.

Key aspects of the social construction of technology include: (1) the prioritizing of social needs – for instance, the current search for technologies that minimize vehicle CO_2 emissions to meet the mandatory limits that signatories to the Kyoto Protocol must achieve and (2) the adaptation of technologies to uses that the original inventors had not anticipated – such as ceramic glass which was originally developed for the mirrors of astronomical telescopes because its thermal expansion coefficient was close to zero as temperatures soared up or down, but was subsequently recognized to make excellent cook tops that minimize domestic clean up chores. In addition, a common setting for the actual act of invention is a social one, with a team of researchers and assistants, networking and sharing knowledge with related research groups, collectively working toward their research objectives.

The Geographical Construction of Technology (G COT) suggested by Norcliffe seeks to reconcile the internalist and contextualist camps by adopting an anti essentialist position where the development and use of technology are viewed as geographically contingent. In any given geographical space at a certain moment in history, there may come together creative persons engaged in the development of new technology, and others with the ability to find utility or symbolic value in that technology. These creators and users may then interact creatively with each other, as happened, for instance, in the case of Silicon Valley and in many other technopoles during the latter part of the twentieth century. This is also a part of Bijker's (1995: 4) understanding of the process since he writes "values, skills and goals are formed in *local* cultures."

Spaces of Science and Technological Change

In recent years a number of historians of science have explored the 'places' of innovation, countering the earlier idea that "scientific ideas floated free in the air, as historians gazed upon them in wonder and admiration" (Ophir and Shapin 1991: 3). To suggest that science has a geography challenges the notions of universality attached to science; it "goes against the grain", as Livingstone notes, because knowledge has no bounds and is certainly not local. For this reason, many scientists have seen technological research as essentially a placeless activity; yet progress in the development and application of science and technology evidently takes place in the minds of creative researchers who work in specific contexts of time and place so, as Gieryn (1998: 248) notes, "all knowledge begins as local knowledge." Constructivists take this argument further by insisting that all knowledge is socially constructed so that technical knowledge is "a human product, made with locally situated cultural and material resources, rather than as simply the revelation of a pre given order of nature" (Golinski 1998, ix). Local contexts may not always be conducive to research, but the existence of innovation hearths subsequently designated as World Heritage Sites (e.g., Coalbrookdale in the UK), of 'shock cities', of technopoles, science parks, learning regions and science laboratory complexes all suggest that various forms of clustering promote technological innovation, which in turn contributes to regional growth.

David Livingstone also addresses this geographical dimension of science and technology in his recent book *Putting Science in Its Place*. He points to three important geographical components of science. First, science and technology are produced in recognized spaces – the laboratory, the museum, the workshop, the incubator, the science park, each with a microgeography, an internal spatial organization with defined boundaries, and internal and external spatial relations. Second, scientific knowledge is unevenly distributed as a result of the clustering of knowledge creation, and a diffusion process that favors certain well connected hubs that produce and exchange technical and scientific knowledge, while other remoter places remain largely out of the loop. Scientific knowledge is often exchanged during face to face encounters such as lectures, laboratories, seminars, and conferences or, as Shapin stresses, by the acts of seeing and observing, which create a strong neighborhood effect. Even in this age of unparalleled global communication, proximity is important to technological change, as is demonstrated by the growth of urban and regional clusters of science and technology. Third, what is accepted as science depends upon place and time so that, for instance, 'intelligent design' has officially been set beside evolutionary biology in schools and university curricula in parts of the southern United States.

Livingstone's approach to the development of science emphasizes three geographical aspects – the site, the region, and circulation (movement), all of which are important to the present argument because the impact of place on technology is inescapable; this is precisely the point of the argument for the 'geographical construction of technology'. In the past quarter century many geographers have recognized this line of thinking by contributing to a growing literature on technologically innovative spaces, on an evolutionary 'path dependency' as regions accumulate intellectual capital in a particular technological field, and in some cases become 'locked in' a particular suite of technologies that eventually contribute to regional decline as other technologies supersede them.

A new vocabulary has been developed that is captured in many of the contributions to this encyclopaedia. These innovative activities are often found in 'new industrial spaces', as Allen Scott has noted, although rejuvenated metropolitan regions have also often become technopoles. Within these new technology intensive spaces, there develop 'associational economies', as Cooke and Morgan phrase the idea, or 'transactional economies' in Storper's version of territorial development. Such spaces are occupied by interconnected 'knowledge intensive industries' where the creative class plays a key role. Economists likewise stress the role of 'knowledge spillovers' or 'regional spillovers' where new technologies developed endogenously within the firm through investment in R&D often lead to external (exogenous) developments as new ventures are created to operationalize the technologies at nearby locations. In the same vein, Michael Storper identifies 'relational assets', which he envisages becoming embedded in regional culture and fostering a 'learning economy', although Yeung cautions that relational economic geography is under theorized, with insufficient attention being paid to power relations, and to the role of key actors. Storper and Venables use the term 'buzz' to capture this notion of excitement about new technologies that persons close to the action experience. The upshot, from a geographical perspective, is that many geographers see proximity as a key factor in technological change.

Technology and Capital

The preceding discussion of new industrial spaces may give the impression that technology is the key driver of regional economic growth, and that capital will follow promising new technologies. But technology needs capital invested in research and development, in state of the art plants, and in infrastructure if it is to promote economic growth. Doreen Massey uses the metaphor of sediments to describe the accumulation of capital in particular regional settings. What this model suggests is that at different times, different towns and regions have seen major investment in new technologies, triggering rounds of regional economic growth. Today, for instance, massive private investment in new oil sand technologies is leading to rapid growth in Canada's Fort McMurray region, while a mix of public and private investment has created – in less than a decade – the 13 Science Parks located in Zhongguancun in the northwest suburbs of Beijing. The point to these examples is to suggest that capital is taking an increasingly active role in technological change. This is not to suggest that the link between capital and technology did not previously exist – it did, but some of the technological research both in universities and in the private sector was centered on pure rather than applied matters, and the commercialization of technology was not always a priority. In the neoliberal era, capital has become strongly focused on technological innovation, with venture capital actively searching for new technologies with commercial promise, and with capital markets developing a category of funds invested in technology intensive industries. Thus, capital is searching out technology, rather than technology securing capital.

Public capital directed by the state is also an important element in creating technology led regional growth, although now given less mention than in the Keynesian era. Yet Harrison's critique of the discourse on Silicon Valley, long held up as an example of private entrepreneurship, stresses the key contribution of the state (i.e., federal defense expenditures and the NASA aerospace program) to the rapid growth of the technopole. Likewise China's current explosive rate of economic growth and technological progress has been helped by the active role of the national and local state in promoting business partnerships, and providing infrastructure.

As capital aggressively pursues technological change, so nature is increasingly being put at risk. The impact of technology on nature has become so pervasive that we no longer talk of nature as an original condition, but as a condition produced by humans consciously and unconsciously as a by product of our industrial activities. Technological change has not only transformed both industry and domestic production, it has transformed nature to such an extent that all landscapes need to be viewed as human modified. In consequence, the management of nature has increasingly become a political issue, resulting in the rise of the new field in human geography called political ecology.

Futures of Technology: A Concluding Comment

There is no shortage of gurus predicting the future impact of technologies, some preaching gloom, and others proclaiming a brave new world. From a geographer's perspective, in the short run it seems likely that four principal current trends will continue:

1. New labor saving technologies plus resource depletion (which often results from using powerful

'nature changing' technologies) are leading to de population of the resource peripheries of the indus trialized world (e.g., interior of Australia, mid north of Canada).
2. New technology intensive industries and services (this includes cultural production and financial ser vices) show increasing concentration in the world's major metropolitan centres, some of which are growing at a remarkably rapid rate.
3. By aggressively buying best technologies, or copying or importing them via supply/subcontractor ar rangements, and by increasing their own capacity to develop technology, newly industrialized countries (especially in East Asia) have narrowed the economic gap with rich countries.
4. The global south remains highly dependent on im ported technologies, many of which are embedded in high value added goods and services including med ical supplies, luxury goods, military equipment, the media, and financial services. Lack of technological capacity in these and related fields make it difficult for these underdeveloped states to make rapid economic progress.

In the longer run, future geographies of technological change will depend largely upon shifts in world politics. The unfettered application of new technology to indus try, to military adventurism, and to resource exploitation raises concerns about the wholesale destruction of nature, and human survival itself. If current neoliberal policies prevail, then it seems certain that new technologies will remain largely in the private sphere, and the associated polarization of incomes and economic development be tween North and South, core and periphery, and rich and poor will continue. Newly industrialized states – now including China – have demonstrated that a technolo gical catch up is feasible, but that depends upon an active 'dirigiste' role for the state.

See also: Growth Poles, Growth Centers; High-Tech Industry; Industrial Restructuring; Innovation; Learning Regions; Neoliberal Economic Strategies; Neoliberalism and Development; Regional Development and Technology; Regional Innovation Systems; Technology and Regional Development; Technology Industries; Transnationalism and Technological Transfer; Venture Capital.

Further Reading

Acs, Z. J. and Varga, A. (2005). Entrepreneurship, agglomeration and technological change. *Small Business Economics* 24, 323 334.
Bijker, W. (1995). *Of Bicycles, Bakelites, and Bulbs: Towards a Theory of Sociotechnical Change*. Cambridge, MA: MIT Press.
Castells, M. and Hall, P. (1994). *Technopoles of the World: The Making of Twenty First Century Industrial Complexes*. London: Routledge.
Cooke, P. and Morgan, K. (1998). *The Associational Economy: Firms, Regions and Innovation*. Oxford: Oxford University Press.
Cooke, P. and Piccaluga, A. (2006). *Regional Development in the Knowledge Economy*. London: Routledge.
Florida, R. (2005). *Cities and the Creative Class*. New York: Routledge.
Freeman, C. and Louca, F. (2001). *As Time Goes By. From the Industrial Revolutions to the Information Revolution*. Oxford: Oxford University Press.
Golinski, J. (1998). *Making Natural Knowledge: Constructivism and the History of Science*. Cambridge, MA: Cambridge University Press.
Hall, P. and Preston, P. (1988). *The Carrier Wave: New Information Technology and the Geography of Innovation, 1846 2003*. London: Hyman Unwin.
Harrison, B. (1994). *Lean and Mean: The Changing Landscape of Corporate Power in the Age of Flexibility*. New York: Basic Books.
Heertje, A. (1977). *Economics and Technical Change*. London: Weidenfeld and Nicolson.
Hsu, J. Y. and Saxenian, A. (2001). The Silicon Valley Hsinchu connection: Technical communities and industrial upgrading. *Industrial and Corporate Change* 10, 893 920.
Livingstone, D. N. (2003). *Putting Science in Its Place*. Chicago, IL: University of Chicago Press.
Massey, D. (1995). *Spatial Divisions of Labor: Social Structures and the Geography of Production*. New York: Routledge.
Norcliffe, G. (2001). *The Ride to Modernity: The Bicycle in Canada, 1869 1900*. Toronto: University of Toronto Press.
Norcliffe, G. (2005). The rise of the Coventry bicycle industry, and the geographical construction of technology. In Van der Plas, R. (ed.) *Cycle History 15: Proceedings of the Fifteenth International Cycle History Conference*, pp 41 58. San Francisco, CA: Cycle Publishing.
Norcliffe, G. (in press). G COT: The geographical construction of technology. *Science*, *Technology and Human Values.*
Ophir, A. and Shapin, S. (1991). The place of knowledge: A methodological survey. *Science in Context* 4, 3 21.
Oudshoorn, N. and Pinch, T. (2003). *How Users Matter: The Co Construction of Users and Technologies*. Cambridge, MA: MIT Press.
Rigby, D. (2000). Geography and technological change. In Sheppard, E. & Barnes, T. J. (eds.) *A Companion to Economic Geography*, pp 202 223. Oxford: Blackwell.
Saxenian, A. (1996). *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.
Schumpeter, J. (1942). *Capitalism, Socialism and Democracy*. New York: Harper.
Scott, A. J. (1988). *New Industrial Spaces*. London: Pion.
Solow, R. (1956). A contribution to the theory of economic Growth. *Quarterly Journal of Economics* 70, 65 94.
Shapin, S. (1994). *A Social History of Truth: Civility and Science in Seventeenth England*. Chicago, IL: Chicago University Press.
Storper, M. (1997). *The Regional World: Territorial Development in a Global Economy*. New York: Guilford.
Storper, M. (1997). Regional economies as relational assets. In Lee, R. & Willis, J. (eds.) *Society, Place, Economy: States of the Art in Economic Geography*, pp 248 259. London: Arnold.
Storper, M. and Venables, A. J. (2004). Buzz: Face to face contact and the urban economy. *Journal of Economic Geography* 4, 351 370.
Yeung, H. W. C. (2005). Rethinking relational economic geography. *Transactions, Institute of British Geographers* 30, 37 51.

Relevant Websites

http://inventors.about.com/od/timelines/Timelines of Invention and Technology.htm
http://www.scimaps.org/
Places & Spaces: Mapping Science.
An exhibition created to demonstrate the power of maps to understand, navigate and manage not only physical places, but also abstract information spaces.

Technology and Regional Development

J. Zhang, University of Minnesota, Tacoma, WA, USA

Glossary

Absorptive Capacity The ability of a firm or an individual to scan and monitor relevant technological and economic information, to identify technical and market opportunities, and to acquire knowledge, information, and skills needed to develop technologies.

Collective Learning The cumulative creation of common knowledge that takes place among a community of actors in a locality by interactive mechanisms based on shared rules, norms, organizations, and procedures.

Knowledge Base The set of information inputs, knowledge, and capabilities that inventors draw on when looking for innovative solutions.

Innovation The first attempt to introduce a new product or process into the economy.

Invention The first occurrence of an idea for a new product or process.

Social Technologies Structures and forces mold the way individuals and organizations interact to get things done, which are commonly referred as institutions.

Introduction

The central challenge to economic geographers is to search the determinant of economic development, thus to explain the 'forging ahead', 'catching up', and 'falling behind' of regions and nations. The level of a region's economic development can be measured by the average performance of its profit seeking firms, without considering its procedural justice and distributional effects. Clearly, technology has been identified as the primary determinant of firm productivity and competitiveness. Technological change, or the 'creative destruction' of older ways of doing things, is largely understood as the driving force of economic development after Schumpeter. However, it is important to differentiate between 'immediate' sources of economic development and causation at a deeper level. It is reasonable to identify technology as the immediate source of economic development. The more fundamental question is what deeper causes lie behind it. At the most general level, 'institutions' presently is the name many geographers and economists give to these deeper causes. For geographers, the key question is why firms in some regions have repeatedly been the site of path forming innovations or new industrial sectors, but many others have never spawned any major new technology or industrial sector. Moreover, to what extent can successful practices be replicated in other less fortunate locations?

In order to answer these questions, we need to first develop our conceptualization of technology. Production involves a set of actions or procedures that need to be done, for example, as specified in a recipe for making a cake. In this view, technology can be defined as a recipe describing a detailed program by which inputs can be transformed into outputs. A recipe is anonymous regarding any division of labor, but most economic activities involve a division of labor plus a mode of coordination among multiple actors. Nelson and Sampat propose to call the recipe aspect of an activity its 'physical' technology, and the way work is divided and coordinated its 'social' technology. 'Social technologies' here are defined as structures and forces mold the way individuals and organizations interact to get things done, commonly referred as institutions. In this view, technology is inseparable from institutions, and technological change often coevolves with institutional change.

When technology is concerned, an important distinction is normally made between invention and innovation. Invention is the first occurrence of an idea for a new product or process, while innovation is the first attempt to introduce it into the economy. A time lag of several decades or more between invention and innovation is not uncommon, because many inventions require complementary inventions and innovations to succeed at the innovation stage. While inventions may be carried out anywhere, for example, in universities, innovations occur mostly through entrepreneurs in business firms.

Technological knowledge is embodied in machinery and equipment, but technology cannot be reduced to machinery, a set of well defined blueprints, or operating manuals. The fundamental knowledge, often tacit, is embodied in individuals and organizational procedures, entailing the formation of an 'epistemic community' with a shared cognitive frame, or codebook, among practitioners. Technology is also 'knowledge in action', because invention and innovation proceed through sequences of problem solving activities. Beyond inventive discovery and patenting, technological learning often involves imitation, reverse engineering, adoption of capital embodied innovations, learning by doing, and learning by using. Consequently, technological learning is continuous and cumulative; current technological development often relies on the 'absorptive capacity' developed through past experiences of production and innovation.

Whatever the primary source of scientific invention, it is the successful introduction of product, process, and organizational innovations that allows firms to override the preexisting conditions of markets and industries, and to grow and gain market shares at the expense of non innovating firms. Hence, innovation is at the core of Schumpeter's theory of economic development. Innovation allows firms to build up monopolistic rents which tend to be progressively eroded alongside the imitative diffusion of new products and processes.

Regions, as repositories of knowledge and facilitators of learning, are becoming focal points for innovation in the new, global, knowledge intensive, capitalism. Region specific enablers and constraints for knowledge production and diffusion can feature prominently in explaining the systematic differential innovative capability of firms between regions. Successful regions often have a well functioning 'regional innovation system' in place, with co located firms and supportive technological infrastructure such as public and private research laboratories, universities and colleges, technology transfer agencies, and vocational training organizations. In order to understand the relationship between technology and regional development, we need to explore the geographic sources and mechanisms of technological innovation.

Geographic Sources of Technological Innovation

In the nineteenth century, the greatest invention – the method of invention, or the process of transforming intellectual and physical capital into new knowledge and new technology – became institutionalized through the establishment of research laboratories and universities in the West. The linkages between science and technology have been tight ever since; and the emergence of major new technologies has frequently been directly linked with major scientific breakthroughs.

As a result, the conventional wisdom regarding the source of innovation is embodied by the science led 'linear model'. The innovation process is described as a 'linear' process in the sense that increased investment in research and development (R&D) leads to greater innovative output which results in increased economic growth. The main focus of this model is inventions that are largely exogenous to existing firms and the creative entrepreneurs who take the risks involved in turning the inventions into commercial innovations. Based on such a topdown model of innovation, high flows or stocks of both public and private R&D and academic research, measured by patents or innovation counts, are believed as the major source of firm innovation. Hence, the regional distribution of R&D institutions becomes the major explanation for regional economic performance.

The policy implication of this linear model is also clear. Due to 'market failure' in basic research, government must engage in or subsidize basic research and protect intellectual property rights (IPRs) through patents and other means, so that the 'appropriability' of knowledge producers can be ensured. Accordingly, local innovation can be advanced if the state improves its public R&D system.

The linear model of innovation fails to embrace the continuity of innovation and the variety of means by which individual firms and groups of interlinked firms and related institutions generate new technologies. The relevance of scientific knowledge and the mechanisms through which such knowledge is transmitted vary greatly across scientific disciplines, technologies, and industries. It has been shown that science is directly relevant to industrial R&D only in a limited number of industries – typically agriculture, chemicals and pharmaceuticals, electronics, and precision instruments. In most contemporary developed economies, one typically observes quite a few institutions, together with a multitude of firms, sharing in different combinations the tasks of scientific explorations and search for would be technological applications.

Firms, especially large ones, do undertake a significant amount of R&D activities. Back in 1940s, Schumpeter argued that the increasingly scientific base of economic activities had caused innovation to become more and more costly, due to indivisibilities and significant economies of scale and scope. In the presence of barriers to entry and weak appropriability conditions, large firms and *ex ante* monopolistic power might be more conducive to innovation than fully competitive markets populated by small firms. In this case, the strong positive feedback loop between successful innovation and increased R&D activities can lead to increased market concentration through vertical integration and to the establishment and persistence of R&D activities in particular cities.

Since the late 1980s, some geographers have come to argue that increasing market uncertainty has led to flexible specialization, characterized by vertical disintegration of large firms. As a result, spatial agglomeration of smaller firms and localized division of labor replaced organizational division of labor of large firms to achieve economies of scale and scope. However, increasing uncertainty could lead to either rising or declining internal economies of scale. While it is true that many firms continue to restructure themselves to reap the benefits of specialization in their core business, the resulting structures do not always exhibit the characteristics of flexible specialization. Some industries such as oil and gas in the early 1980s, pharmaceuticals in the late 1980s, and telecoms and banking in the 1990s have been characterized by large scale merger and acquisition. The large scale R&D programs of the multinational companies (MNCs) and their key role in knowledge creation and transmission have provided evidence to the continuing significance of large firms in technological innovation.

The spatial pattern of these large companies is ultimately determined by organizational decision making. The locations of their headquarters and R&D activities, often in the same region, strongly influence the location of innovation. They also create possibilities for small and medium enterprises (SMEs) located in the same areas to benefit from knowledge spillovers and to exploit new ideas and niche markets for innovation. To the extent that such endogenous knowledge creation and transmission within large firms matter, opening up the technological blackbox entails opening up the organizational blackbox and its geography.

In the recent geographic literature, it is neither public research, nor large firms, but industrial communities that is considered as the all important site of knowledge formation. The key members of such communities are defined differently by different authors. Commonly they may consist of engineers, managers, and entrepreneurs when inter firm networks are stressed (e.g., in the literature of industrial districts), but can be extended to also include scientists and financiers when linkages with knowledge generating organizations (i.e., R&D institutes and universities) and investors (e.g., venture capitalists) are treated seriously. Here the emphasis is placed on social relationships within autonomous communities and rules of behavior embedded in those relationships. It is argued that informal communities can solve knowledge problems that escape governmental and market solutions. Acknowledging the key role of communities in the production of knowledge means that the mere presence of isolated firms and/or research institutions is insufficient for knowledge creation. It also indicates the need for institutional norms that can support communities. Consequently, public policy must emphasize collective cooperation rather than just individual appropriation.

In sum, there is not yet a definite answer to the sources of technological innovation. A moderate conclusion is that governmental research institutes, R&D laboratories of large firms, networked SMEs, as well as their interconnections can all be important sources of innovation. The concrete pattern is an empirical question, depending on the unique 'knowledge base' of the industry in question and the stage of its life cycle. A more demanding question, for which we do not yet have a good answer, is that where do these sources come from in the first instance. Moreover, we have to explain the mechanisms, in addition to the sources, leading to geographically uneven technological innovation.

Geographic Mechanisms of Technological Innovation

Geographers usually explain regional unevenness of technological innovation by invoking a set of localized structural externalities. This tradition goes back to Alfred Marshall who identified three sets of localization economies: (1) economies of specialization due to local access to specialized suppliers, (2) labor market economies due to local access to a pool of specialized workforce, and (3) knowledge spillover due to the advantage of localized informal communications. Economies of specialization and labor market economies are generally referred as pecuniary externalities, and knowledge spillover technological externalities. Pecuniary externalities allow co located firms to reduce costs and uncertainties through market interactions. Technological externalities, in contrast, materialize through nonmarket interactions within the local community and, in principle, are accessible to all members of the local community.

Theoretically, there is no *a priori* reason to exclude either pecuniary externalities or technological externalities as drivers of innovation in regional clusters. Empirically, it is difficult to discern direct and indirect contributions to knowledge production. Consequently, some authors maintain that the old fashioned Marshallian pecuniary economies are still relevant to learning and innovation in the contemporary setting. In the course of time, however, local knowledge spillover (LKS), without explicit consideration of pecuniary externalities, appears to have gained ascendance in academic discourse vis à vis Marshallian pecuniary externalities as the principal explainer of regional innovative capacity.

Local knowledge spillovers can be defined as knowledge externalities bounded in space, which allow companies operating nearby important knowledge sources to introduce innovations at a faster rate than rival firms located elsewhere. The ascendance of LKS approach comes from the accredited importance of learning in the knowledge based economy. Many economic geographers also claim that traded 'pecuniary assets' becomes less important to regional knowledge creation in the globalizing knowledge economy, while 'relational assets' or 'untraded interdependencies' are seen as increasingly important.

A linkage between learning and geography usually is made through the notion of collective learning: the cumulative creation of common knowledge that takes place among a community of actors in a locality by interactive mechanisms based on shared rules, norms, organizations, and procedures. It is argued that co location of economic agents provides multiple opportunities to benefit from inter firm linkages, high density communication, interactive learning, collective problem solving, and enables firms to develop a shared cognitive frame (code book) and common understanding. In large enterprises, organizational learning is enabled through the creation of common rules and routines, imposed by hierarchy and control. Within a regional innovation system, the litera ture suggests that organizational learning is achieved through user–producer relationships, formal and in formal collaborations, inter firm mobility of skilled workers, and the spin off of new firms from existing or ganizations (firms, universities, and public research cen ters) through the formation of new enterprises.

In the geographic literature of collective learning, emphasis has been placed on the role of 'tacit' as against 'codified' knowledge, in that the former is seen as more conducive to rapidly changing markets, and flexibly specialized innovative activity, whereas codified know ledge is equated more with standardized mass production type activity. In a world in which codified knowledge is becoming increasingly ubiquitously available, tacit knowledge attains a higher premium in deriving com petitive advantage owing to its uniqueness. Direct, face to face interactions enabled by geographic proximity are considered as crucial for the exchange of tacit knowledge.

However, economic geographers emphasize that it is not geography *per se* that matters for learning and in novation, but it is the embeddedness of firms in localized networks that facilitate the diffusion of knowledge and enhance collective learning in clusters or regional in novation systems. First, knowledge spillovers are con sidered to occur mainly through direct interactions on the individual level. Second, to enable and foster these interactions, it is argued that the existence of social networks and their perceived stability and reliability are of pivotal importance. In this sense, tacit learning is a form of social learning. It depends not only on spatial proximity, but also on social embeddedness or 'social capital', which is built upon particular relational con ditions, common cultural understandings, and shared norms of trust and reciprocity. The origin and sources of such social capital, or regional culture, however, are not yet adequately understood.

Knowledge that spills over through social networks has also been argued as a local public good. As collective learning processes develop, it is suggested that knowledge becomes diffused pervasively in an unstructured manner through the channel of business networks, turning into a public good within the local, or regional, boundaries. In this view, all firms within a region or an industrial district are assumed as relatively homogeneous. Knowledge cre ation and regional development, however, continues to be very much shaped by decision making firms as private, profit seeking agents. Firms in industrial districts or clusters are usually characterized by heterogeneous and asymmetrically distributed knowledge bases and ab sorptive capacities. Embodied scientific and technical knowledge (individual tacit) also largely remains a private good, unless sharing agreements turn it into common property or a club good. Meanwhile, social networks are not necessarily fair or democratic; rather they are often characterized by strategic behavior and unequal power status. Moreover, the communication of tacit knowledge can be willfully manipulated to serve as a key ex clusionary mean, so that certain actors (even local ones) can be prevented from understanding the content of scientific and technical messages. In light of these, recent literature has come to emphasize that the local diffusion of innovation related knowledge is selective and uneven, a property that reflects the different internal capabilities and willingness of firms to transfer and absorb knowledge.

Another challenge to the KLS approach is the lack of empirical evidence. The empirical studies so far do not offer any definite answer to the question of 'the role of geographical distance in the economics of knowledge transmission'. Networking, which is said to be a central feature of local cluster and new industrial districts, is as much an international phenomenon linking actors in the global economy as it is a characteristic of regional pro duction nodes. Moreover, it has been recognized that tacit knowledge can also be shared within physically dispersed 'epistemic communities' as long as the decoding cap ability of the addressees can be adequately developed.

As a result, economic geographers have increasingly acknowledged the fact that geographic proximity *per se* is not sufficient to explain processes of localized learning and innovation. The knowledge spillover argument thus has been reformulated. The assumed correlation between learning and geographic proximity is now relaxed, while the association between learning and networking re tained. Now it is largely agreed that the creation of new knowledge must be based on the combination of close and distant interactions. The importance of local net works remains emphasized, but nonlocal linkages are considered as vital in order to pump novel ideas and technologies into the local knowledge system to increase its dynamics and to reduce the risk of lock in. However, when the LKS approach is increasingly questioned, the role of geography becomes blurred and a viable alter native theory is yet to emerge.

In any case, the notion of cluster seems to be central to many of the geographic accounts of technological in novation. Many high tech clusters, or technopoles, have been identified to support various geographic accounts of technological innovation. These include the traditional machinery and automotive districts of Baden Württem berg in southwestern Germany, emerging information and communication technology (ICT) clusters in Beijing, China, and Banglore, India, as well as Silicon Valley, the 'grand daddy' of them all (**Figure 1**). The evidence of a positive association between clustering and innovation is not consistent, however. Far from being the general rule, the benefits realized from geographical clustering appear to be specific to certain industries at certain stages of

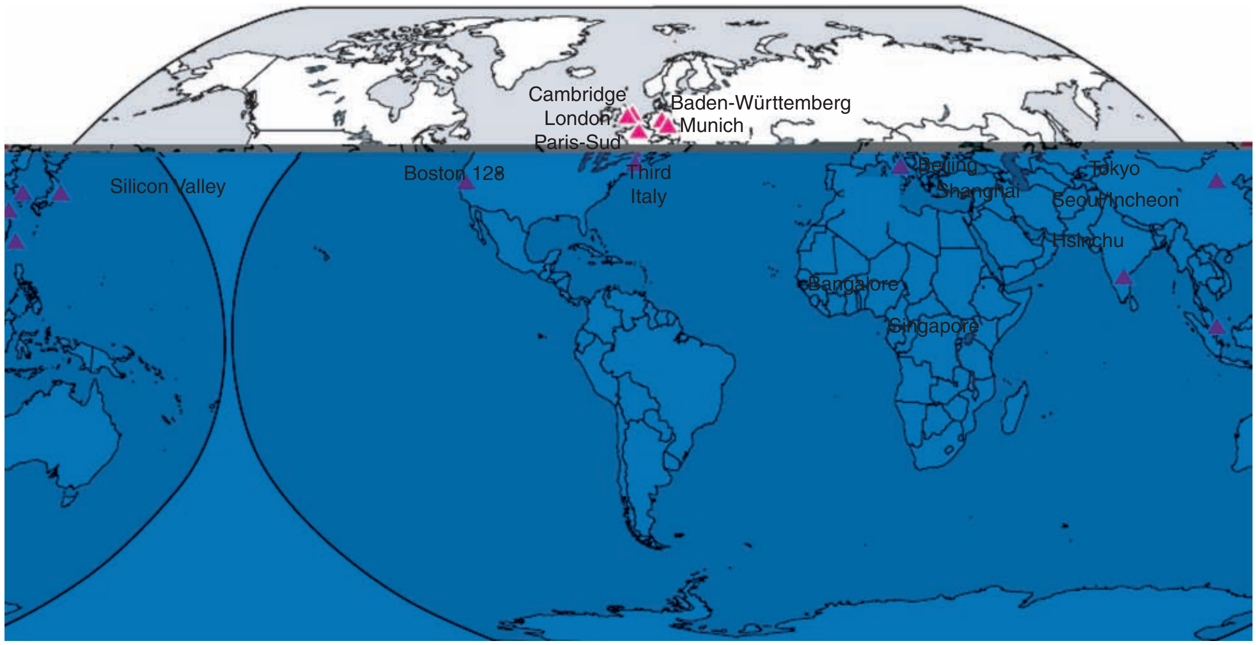

Figure 1 Major technopoles and high-tech regions of the world.

development in certain places, and are only realized under particular conditions. Particularly, an emphasis on the desirable geographic externalities always faces the challenge to explain how such externalities arise in the first place. Apparently, knowledge is absorbed relatively easier in regions that already have a relatively higher productivity level and a larger stock of knowledge. The meaning, origin, and function of historical endowment to regional innovation are not yet clear, neither the effect of regional competition based on highly uneven historical endowments in a politics ridden playing field. Last but not least, the emphasis on extra firm factors and local environment has led to a lack of any serious analysis or theory of the internal learning mechanism of business enterprises. The mechanism of collective learning remains ambiguous and its interactions with firm based learning are nearly left completely unexamined. Further research efforts need to be devoted to exploring how knowledge is actually transmitted, among whom, at what distance, and on the basis of which codebooks. Above all, it is highly questionable whether a parochial focus on firms as the only actor of regional technological innovation can offer a comprehensive understanding of the dynamics of regional development.

Actors of Innovation and the Scale(s) of Explanation

Recent geographic accounts of technological development have been centered on the actor of firms, although the emphasis is largely placed on inter firm linkages rather than intra firm organizations. Consequently, the cause of regional development is largely explained at the local scale as the capacity or responsibility of local economic agents and/or local culture, even when the importance of nonlocal connections is recognized. In the process of regional technological development, however, multiple actors other than firms residing at multiple geographic scales play a crucial role. In the evolution of the computer industry in the US, for example, six types of actors, each of them highly heterogeneous in terms of competences and behavior, have been identified to play a major role: (1) firms; (2) universities and public laboratories; (3) the government (especially the Ministries of Defense, Industry, Trade, and Science and Technology); (4) suppliers (of semiconductor components, disk drives, etc.); (5) users (consumers, banks, insurance companies, manufacturing firms, etc.); and (6) financial institutions (such as venture capitalists and banks).

Incorporating more actors into the analysis of regional technological development can be disruptive to existing geographic approaches, but it may well be a desirable 'creative destruction'. It implies that the current conceptualization of 'resources' and 'institutions' must be reformulated. With regard to resources, the current focus on knowledge and relational resources need to be broadened to take pecuniary and political resources more seriously. At the same time, immediate resources (or causes) need to be distinguished from deeper resources. When institutions are concerned, the present focus on institutions as "the ways things are done" (i.e., norms, routines, and conventions) must also be broadened to treat institutions simultaneously as 'the basic rules of the game' and 'governing structures'.

Institutions shall be treated as a system, consisting of at least all the three aforementioned layers. Some insti tutions, such as 'respect for the rule of law', affect broadly a wide range of economic activities. Other institutions are sector or activity specific. To be useful, analysis of how institutions affect economic growth and productivity needs to explain the complementarity and inter connection of all these layers of the institutional system at work, as well as to get into the details of how such in stitutions affect behavior. From this point of view, it is a mistake to search for a small set of institutions that are necessary and sufficient for economic productivity and progress. Many different but complementary institutions are needed at the same time, and the institutions that are effective are very context dependent. Such a con ceptualization of institutions means that geographers have to treat seriously both culture and politics, as well as the implied interactions of multiple geographic scales therein.

Due to the centrality given to knowledge and learning, current geographic literature generally assumes that purposeful economic agents will pursue knowledge and learning automatically above any other goals. Under such an assumption, the incentive shaping public institutions are frozen and thus excluded from the analysis. However, there is no guarantee that firms will devote resources to invest in skills and knowledge that will increase prod uctivity. Schumpeter reminded us half a century ago that innovation activities are costly, risky, and uncertain. Profit seeking firms will invest in skills and knowledge for innovation if, and only if, this is perceived as the best way to obtain profit. The absorptive capacity of a firm is determined not only by its cognitive capacities, but also by its willingness to incur costs of learning new know ledge. The direction of organizational investment in skills and knowledge will reflect the underlying incentive structure defined by the institutional matrix. If the highest rate of return in an economy is to piracy we can expect that the organizations will invest in skills and knowledge that will make them better pirates. Inefficient political institutions and mechanisms of corporate gov ernance frequently act as barriers that prevent the ap propriation of knowledge spillovers in many economies. In other words, the willingness of firms to absorb knowledge about new production technologies is quite likely to differ between regions and countries depending on the actual pay off structure that they face.

The institutional environment, while durable, also evolves through time and is itself shaped by local, na tional, and international politics and power struggles. Organizations are not just constrained by institutional environments. In fact, law, legitimacy, and political outcomes all somehow reflect the actions taken by or ganizations to modify their environments for their inter ests of survival, growth, and certainty. Our present level of understanding of regional technological dynamics still remains woefully underdeveloped, with seemingly little or no room for power and politics. Without a holistic theory of uneven regional development, much of the current work tends to provide snapshots of successful regions, overlooking the richer set of factors and con ditions that account for the geography of innovation. It is an urgent task for our interpretation of regional techno logical dynamics to be modified to incorporate purposeful human agency as well as the structural preconditions and contextual influences that shape the emergence and evolution of firms and industries.

See also: Embeddedness; Industrial Districts; Institutionalism/Institutional Geographies; Knowledge Economy; Learning Regions; Network Regions; Regional Innovation Systems; Technological Change.

Further Reading

Amin, A. and Cohendet, P. (2004). *Architectures of knowledge: Firms, capabilities and communities*. Oxford: Oxford University Press.
Asheim, B. T. and Gertler, M. S. (2005). The geography of innovation: Regional innovation systems. In Fagerberg, J., Mowery, D. & Nelson, R. (eds.) *The Oxford handbook of innovation*, pp 291 317. Oxford: Oxford University Press.
Bathelt, H., Malmberg, A. and Maskell, P. (2004). Clusters and knowledge: Local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography* 28, 32 56.
Breschi, S. and Francesco, L. (2001). Knowledge spillovers and local innovation systems: A critical survey. *Industrial and Corporate Change* 10, 975 1005.
Breschi, S. and Malerba, F. (eds.) (2005). *Clusters, networks and innovation*. Oxford: Oxford University Press.
Castells, M. and Hall, P. (1994). *Technopoles of the world: The making of twenty first century industrial complexes*. London: Routledge.
Doring, T. and Schnellenbach, J. (2006). What do we know about geographical knowledge spillovers and regional growth? A survey of the literature. *Regional Studies* 40, 375 395.
Gertler, M. S. (2003). Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). *Journal of Economic Geography* 3, 75 99.
Gertler, M. S. (2004). *Manufacturing culture: The institutional geography of industrial practice*. Oxford: Oxford University Press.
Malecki, E. J. (1997). *Technology and economic development: The dynamics of local, regional and national competitiveness* (2nd edn). Essex: Longman.
Malerba, F. and Orsenigo, L. (1996). The dynamics and evolution of industries. *Industrial and Corporate Change* 5, 51 88.
Maskell, P. and Malmberg, A. (1999). Localised learning and industrial competitiveness. *Cambridge Journal of Economics* 23, 167 186.
Nelson, R. and Sampat, B. (2001). Making sense of institutions as a factor shaping economic performance. *Journal of Economic Behavior and Organization* 44, 31 54.
Schumpeter, J. A. (1959/1912). *The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle* (trans. Opie, R). Cambridge, MA: Harvard University Press.
Scott, A. J. (1988). *New industrial spaces: Flexible production organization and regional development in North America and Western Europe*. London: Pion.
Simmie, J. (2005). Innovation and space: A critical review of the literature. *Regional Studies* 39, 789 804.
Storper, M. (1997). *The regional world: Territorial development in a global economy*. New York: Guilford Press.

Technology Industries

C. G. Alvstam, Göteborg University, Gothenburg, Sweden

Glossary

Process Technology The provision of modern machinery and equipment to suppliers, technical support in product planning, quality management, inspection and testing, as well as advice on tooling, maintenance, production layout, and operations.

Product Technology The provision of proprietary product know-how, for example, through patents or licenses, the transfer of product designs and technical specifications, technical consultations with suppliers to help the partners to master new technologies, and by giving regular feedback on product performance.

Research and Development An activity within the firm in which new physical products, services, and processes are invented and developed into commercial applications.

Technological Innovation A learning process, within which knowledge is continuously accumulated and developed.

Technology Industry A branch of industry with a high value added in research and knowledge production.

Technology Transfer The transfer of technical skills from one economic actor to another.

Basic Concepts and Definitions

The notion of technology in economic theory is often related to the works of the Austrian economist Joseph Schumpeter (1883–1950), who in a classical text, published 1912, emphasized the crucial impact of innovation and technological change in the continuous, evolutionary process of economic development. In Schumpeter's terms, technology and innovation are seen as part of a general entrepreneurial process. He coined the expression of 'gales of creative destruction' to describe the process within capitalism in which radical innovations and technological breakthroughs make established competence and physical machinery and equipment suddenly obsolete.

Technological innovation can in this context be viewed upon as a learning process, within which knowledge is continuously accumulated and developed. The development of knowledge is localized in the sense that it is considered to be produced in specific places. The preconditions for creation of knowledge, or how to construct and to cultivate an environment of creativity, can therefore be regarded as the most crucial location factor in contemporary economic geography, surpassing traditional factors like access to raw materials, capital, and proximity to final markets. There is usually a clear distinction made in the literature between technological knowledge that is codified, that is, explicitly expressed, and tacit knowledge, which is implicitly possessed and communicated in informal contexts. Furthermore, there is often a distinction made between embodied and disembodied technologies, related to the physical versus the nonphysical character of technological transfers, as well as between product technology and process technology, the latter being broader and encompassing a wider range of factors related to production efficiency.

The technology concept can cover every aspect from simple and routine procedures on the shop floor level to advanced and complex operations within a production system. In this sense it is often related to the process of learning. It does also incorporate every stage from basic research, via applied research, the breakthrough of an innovation and the continuous product development to the commercialization and marketing of a physical product to various distribution and after sales activities. Infrastructure related to research and development (R&D) of a product – office and laboratory space, machinery, apparatus and equipment, various service functions connected with the main activity, etc. – are normally also included in the R&D budget. In order to grasp such an elusive element of production, it is common to measure the impact of technology in terms of expenditures directly related to R&D within the firm, in proportion to sales or gross output value. These expenditures may be variously described in the accounting systems of different firms in different countries, and it is therefore implicitly stated that the pecuniary estimate of the level of technology in a private firm or in a country involves an arbitrary element, and should be used and compared with great cautiousness, especially when it comes to planning and implementation of a technology policy in the local or national economy.

A second commonly used definition of the technology level in different manufacturing subsectors is to measure the number of highly educated technical employees, scientists, engineers, and professional specialists as a percentage of the total workforce, thus estimating the technical inputs into production in addition to R&D. A definition of a high technology industry would then be that these groups are over represented, compared with the industrial average.

A third way of approaching the technology concept is to describe changes of productivity in a company, for example, in terms of manufacturing value added per employee, implying that an increase in the measured productivity is *ceteris paribus*, an effect of improved technology. The impact of technology is largest in the initial phases of the product life cycle, in which the costs of technological innovation are high, production techni ques are rapidly changing, the immediate commercial yield is limited, and the economies of scale, as well as the capital intensity, are low.

A fourth approach to describe the level of technology intensity is to measure the number of patent applications and approvals. A large number of patent applications per capita in a country is considered as a sign of high level of research activity. A methodological problem by using the geography of patent applications as a measurement of the concentration or dispersal of technology is related to different traditions of legislation regarding intellectual property rights between countries, and also to the glob alization of R&D, with the effect that the country where a patent is applied for, may not be the locus of the underlying research activity itself. Rather, it may reflect the need to protect the commercial distribution of a product within a certain market.

The usually adopted industrial classification nomen clatures are aggregates of different types of firms and workplaces. Accordingly, a single branch of industry within the statistical group of classification contains all different technical levels from advanced basic research to standardized routine work. One example is that the re tailing of electronic articles, where a large number of individuals are employed, is often classified as belonging to the electronic industry in general, and is in such cases grouped within a high technology segment, even though this activity as such does not represent a high level of technical knowledge. Therefore, by using a classification of different levels of technology intensity between dif ferent branches of industry, one should take into account that such a classification encompasses a broad range of inner variations. The Organisation for Economic Co operation and Development (OECD) categorization of various levels of technology, using the measure of R&D intensity (R&D expenditure/output), is widely used, but should, accordingly, be interpreted with cautiousness (**Table 1**).

Table 1 Classification of manufacturing industries based on technology intensity

Industries	*1991*	*1999*
High-technology industries	9.4	8.7
Pharmaceuticals	9.4	10.5
Aircraft and spacecraft	13.9	10.3
Medical, precision, and optical instruments	6.6	9.7
Radio, TV, and communications equipment	7.9	7.4
Office, accounting, and computing machinery	10.9	7.2
Medium-high-technology industries	3.1	3.0
Electrical machinery and apparatus n.e.c.	4.2	3.6
Motor vehicles, trailers, and semitrailers	3.7	3.5
Railroad equipment and transport equipment n.e.c.	2.9	3.1
Chemicals excluding pharmaceuticals	3.4	2.9
Medium-low-technology industries	0.9	0.7
Building and repairing of ships and boats	0.9	1.0
Rubber and plastic products	1.0	1.0
Basic metals and fabricated metal products	0.7	0.6
Coke, refined petroleum products, and nuclear fuel	1.2	0.4
Low-technology industries	0.3	0.4
Wood, pulp, paper, paper products, printing, and publishing	0.3	0.4
Food products, beverages, and tobacco	0.3	0.3
Textiles, textile products, leather, and footwear	0.2	0.3

OECD (2005). Science, technology and industry scoreboard, annex 1.1 2, p 182f.
Note: Figures are based on 12 OECD countries.

The aerospace industry is, not surprisingly, considered to be among the most technologically advanced indus tries by adopting this measure. However, it is also an example of the fact that a significant part of the tech nologically most intense industries is often related to military expenditure. Apart from the measurement problem that the distinction between the defense budget in general and investments in military R&D is not clear cut, it is even more difficult to assess in advance which breakthroughs of military research may lead to appli cations in civil production and therefore should be an ticipated in a comparison between countries of investments in technology. Accordingly, one may argue that a not insignificant share of investments in techno logical development is concealed in general defence ex penditure, particularly in the world's largest economic and political powers, and, thus, are underestimated when one compares the technology intensity between coun tries. On the other hand, it is also possible to argue that a broad range of military activities can be classified as applied research, and should therefore not normally be incorporated among general technology investments.

The most technology intense industry, according to the OECD, the pharmaceutical sector, is an illustrative example of the significance of risk taking when making investments in R&D. The imbalance between R&D ex penditure and costs of manufacturing the resulting commercial product in mass production is extremely high. The winner in the 'patent race' between pharma ceutical firms will enjoy the entire yield of investment, while the losers will face only sunk costs. It is therefore natural that competing firms tend to share the technology risk by merger and acquisitions, and by making strategic alliances and other forms of research collaboration.

In the bottom of the list of technology intensive in dustries one finds sectors of industry that represent later stages in the product life cycle, for example, food, tex tiles, and forest products. It would, however, be a too precipitate conclusion that there is little or no techno logical development within these sectors. Rather, it can be sign of a different institutional setting, compared with those industries that report a high level of R&D. There is a tendency in mature sectors that R&D is outsourced and concentrated to separate research institutes, which are separately classified from manufacturing itself.

Technology Industries and Location

One particular field in which the technology factor has changed the basic theoretical assumptions in economic geography is its time–space shrinking impact, for example, the decreasing costs to move raw materials and manu factured products over longer distances, due to more ef ficient transport and handling systems, and the opportunity to marginalize the distance factor within many service sectors as a result of the development of new information and communication technologies. New sub sectors within the service industry have emerged based on new technology, for example, call centres, while a number of traditional service sectors have developed into entirely new spatial patterns. These forces of transition can be acting in a direction of either geographical concentration or dispersal. The managerial control over technological inputs, for example, R&D, design, quality control within the globally organized corporation has tended to become more concentrated in the geographical sense, while the operational activities have become more dispersed. R&D activities, that previously were concentrated to a few lo cations within the transnational corporation (TNC), may today often be geographically dispersed in a pattern of 'islands' of smaller research centers in a global 'archi pelago', while at the same time closely connected in a dense intranetwork, substituting physical exchange of in formation and communication to the use of advanced in formation and communication technologies. In both cases, there is a regionally uneven distribution of industries characterized by using advanced technology and em ploying a high level of scientists. Michael Storper has pointed out three 'schools' of explanation: First, those who emphasize the research university spin off process, high quality of life, and good infrastructure, exemplifying with the Silicon Valley concentration and Route 128. Second, the 'regional politics' approach, that holds that regional coalitions push for the transfer of high technology re sources. Third, those scholars who put forward the 'milieu' of innovation and creativity. The milieu is described as a system of regional institutions, rules, and practices which lead to innovation and as a network of actors: producers, researchers, politicians, etc., in a region. Richard Florida has, in his research on the bases of creativity, focused on the combination of 'technology', 'talent', and 'tolerance' as significant factors to understand why certain geographical locations are more successful and competitive as others when it comes to develop new knowledge intensive pro duction. Michael Porter, finally, has put technological strength within a larger context of factor conditions, which, in their turn, form an integrated part of a 'diamond' of competitive advantage of a firm or a nation.

With Silicon Valley serving as a major model, many policy makers have tried to combine the three 'schools' as explained above, to put together the advantages of proximity to scientific researchers, access to venture capital, good physical infrastructure, excellent avail ability to newest information and communication tech nology, and a favorable geographical location in order to 'create' a milieu of technological innovations and to en hance employment in technology intensive industries, by establishing 'technopoles', 'technology districts', or 'technology parks'. This policy has been particularly popular in emerging markets in Asia, beginning with Tsukuba in Japan, and copied in Hsinchu in Taiwan, Inchon in Korea, Cyberjaya in Malaysia, Xingwang in Beijing, China, and Bangalore in India. The success of national or local policies to stimulate localized tech nological knowledge varies, but seen in longer time perspectives, it is evident that these initiatives have contributed to an acceleration of local, as well as national technological development and industrial upgrading through the concerted public action. For example, Cindy Fan and Allen Scott have recently investigated the role of industrial agglomeration and economic development in Chinese regions and found a positive correlation between agglomeration and economic performance and that many manufacturing sectors are characterized by strong posi tive relationships between spatial agglomeration and productivity.

In its larger theoretical context, the notion of tech nology can, very simplified, be seen as having been de veloped from being an implicit exogenous factor in the neoclassical production function, to being given a more explicit and endogenous role in the 'new growth theories' developed mainly in the 1990s. Here, the conventional production function is expanded to incorporate invest ments in R&D, in infrastructure, and in human capital. The 'new growth theories', for example, described by R. R. Nelson and others, are in many ways inspired by the evolutionary setting in its Schumpeterian sense, but they are at the same time criticized for their lack of attention to institutions and institutional change. Recent theore tical contributions have aimed at developing the original

Schumpeterian models in order to more explicitly identify the evolutionary process of 'localized' technological change. Technological change, and the location of technical innovation is in Michael Storper's terms seen as 'path dependent', that is, the time sequence of the development of technological knowledge is crucial to understand the decision to locate a certain type of production to a certain place.

Technology Transfer as a Spatial Process

At the same time, the technological skill is not stuck in a certain location. It can 'spill over' and be transferred through formal (codified) exchanges, and through untraded interdependencies. The transfer of technology between firms and between countries is seen to be a decisive factor in understanding economic growth and industrial transformation in developing countries. Since many developing countries still possess a limited capacity to generate new technology themselves, the power to import technology from abroad, and to combine this ability with domestic human resources and know how, becomes crucial for their long term economic growth. In the literature more specifically focused on technological change in developing countries we can also observe a growing interest in the role played by agglomerations and industrial clusters. However, the role of geographical proximity is often ignored or unclear in most of these studies. For example, in the 'technological capability' approach, originating in the 1980s by researchers inspired by evolutionary theories of technological change and following many disappointing experiences with international technology transfer projects, it is shown that the import of new technologies from advanced countries are insufficient for enhancing productivity and inducing self sustaining industrialization in developing countries. The access to foreign technology does not necessarily imply mastery over it, since high entry barriers to the effective implementation of new technologies exist, mainly because of the tacit nature of knowledge and because foreign technologies are often unsuitable for needs of the importing country. Thus, substantial investment in time and resources is needed to assimilate, adapt, and improve known technologies imported from abroad, ultimately leading to the capacity to create new technologies independently. While technological improvements in individual enterprises are often seen to be enhanced by being part of innovation systems or networks of actors and institutions involved in complementary activities, there is a lack of systematic conceptual and empirical treatment of how and why spatial proximity between actors in such networks or innovation systems can contribute to improve the technological capability within individual firms.

One important source of technology imports is foreign direct investments, where TNCs are transferring technology, know how, and information through their foreign located affiliates to local suppliers. The importance of interfirm linkages for technology learning and upgrading is especially evident in modern growth theory literature. A key perspective in this literature is that capability building is to a significant degree built through relations between firms and institutions, and that it involves efforts at all firm levels and not only the R&D departments, for example, at the shop floor, engineering, quality management, and procurement. Thus, technological learning in a firm does not take place in isolation; rather the process is rife with externalities. As claimed, by Sanjay Lall, these external linkages are often so dense and crucial for technology development that it can be questioned whether they should be called 'externalities' at all.

Moreover, this literature also shows that the process of learning and technical change in successful industrial enterprises is not an easy task, and may involve considerable economic risks, costs, requiring conscious decisions by firms rather than the mere accumulation of enterprises with enhanced experiences of production. For example, although in theory the hardware is equally available to all countries and firms, for example, through imports, the disembodied elements of technology cannot be bought and transferred like physical products. The embodied elements cannot be used at best practice levels unless they are complemented by a number of tacit elements that have to be developed locally. Therefore, as is showed by many students of technology transfers, technological learning calls for conscious, purposeful, and incremental efforts to collect new information, to create new skills, and to strike new external relationships. This process must be located at the production facility and be embodied in the institutional and organizational setting of the manufacturing enterprise.

United Nations Conference on Trade and Development (UNCTAD) has developed a classification of various types of technological assistance, in order to give advice regarding the most efficient ways of using technology transfers as a tool for enhancing economic growth and industrial transformation in developing countries. Broadly speaking, technological assistance between a globally organised TNC and a domestic local partner can be distinguished between two categories: those related to product technology and process technology, respectively. The main types of product related technology transfers include the provision of proprietary product know how, for example, through patents or licenses, the transfer of product designs and technical specifications, technical consultations with suppliers to help the partners to master new technologies, and by giving regular feedback on product performance. Major areas of process technology transfer comprise the provision of modern machinery and equipment to suppliers, technical support in product planning, quality management, inspection and testing, as well as advice on tooling, maintenance, pro duction layout, and operations.

In addition to these types of technology transfers, TNCs can assist their suppliers by transferring organ izational and managerial know how, for example, inven tory management and the use of various delivery and logistical systems. Another important area of technology transfer is related to different types of training programs offered by TNCs in order to upgrade the human re source base of their suppliers.

A more detailed classification to distinguish between different types of technological assistance between the global TNC and its local supplier, developed within UNCTAD, and used with an empirical material by Ivarsson and Alvstam in a study covering technology transfer by a foreign TNC to local suppliers in China, Brazil, India, and Mexico, is the following (**Table 2**).

One important issue, in order to provide a better understanding of the geography of technology industries, is to what extent spatial proximity can affect technology transfer from TNCs to their suppliers. Are suppliers located at a short geographical distance from the pro duction plants in a better position, compared to those located at longer distances from the plants, to absorb and learn from the technological assistance that the TNC provides the suppliers as a part of an on going business relationship? Empirical studies indicate that the TNC provides the suppliers with technological assistance in a number of important areas, including both 'embodied' forms of assistance, especially through product designs and technical specifications, but also through the pro vision of materials and components. It does also give the suppliers technological assistance through the transfer of disembodied technology, where all are given regular feedback in order to improve existing product technol ogy as well as process technology, both through written documents and personal meetings. Comparing the form, extent, and effects regarding the technological assistance between local and nonlocal suppliers, geographical proximity seems to have a role to play. Although there are rather small differences in the extent and forms of technology assistance, taking geography into account, in particular aspects, for example, support to obtain raw materials and components, as well as machinery and equipment and assistance with quality assurance systems, a significantly larger share of local suppliers are provided with technological assistance, compared to nonlocal suppliers. It seems also that geographical proximity is more crucial for opportunities among the suppliers to absorb external technology. This is to a large extent an effect by the commitment and resources that the sup pliers have had to invest in order to learn from their interaction with the foreign TNC. For many smaller, domestic companies, low transaction and communication costs and an opportunity to collect valuable information through regular and spontaneous inter firm meetings, seem to be an important determinant for the successful absorption of external technology.

Table 2 Types of technological assistance between companies

Product technology
Provision of product designs and technical specifications
Provision, advice, or financial assistance to obtain raw materials and components
Regular feedback on product performance to improve existing product-technology
Technical consultations on product characteristics to master new product technology
Organised R&D collaboration in product-related areas
Production technology
Provision, advice, or financial assistance to obtain machinery and equipment
Technical support to improve existing production technology
Technical consultations on machinery operation to master new production technology
Advice concerning production layout and organisation
Assistance with quality assurance systems (e.g., ISO certification, TQM, etc.)
Training programmes for suppliers' personnel
In-plant training for managers and technicians
Training for managers and technicians in the host country or abroad
In-plant training for workers
Training for workers in the host country or abroad

From Ivarsson, I. and Alvstam, C. G. (2004). International technology transfer to local suppliers by Volvo tracks in India. *Tijdschrift voor Economische en Sociale Geografie* 95, 27 43.

The future challenges in technology industry research are to further understand the intangible assets in tech nological development, and to describe and explain the factors that affect the geographical pattern of new pro fessional business service sectors that arise as a result of technological change.

The complex relationship between technology trans fer and knowledge transfer should also be given more attention. Technology transfer refers normally to a transfer between two well defined economic units, while in a knowledge based economy, it becomes more difficult to define the contribution of each separate element in the value chain. Thus, the spatial divisions of expertise, that is, the importance of proximity to intellectual assets, or 'core competences' will become even more emphasized for an understanding of the intricate factors behind the creation and diffusion of tomorrow's technology industries.

See also: Economic Geography; Knowledge Economy; Technological Change; Technology and Regional Development.

Further Reading

Amesse, G. and Cohendet, P. (2001). Technology transfer revisited from the perspective of the knowledge based economy. *Research Policy* 30, 1459 1478.

Bryson, J. R. and Rusten, G. (2006). Spatial divisions of expertise and transnational 'service' firms: Aerospace and management consultancy. In Harrington, J. W. & Daniels, P. W. (eds.) *Knowledge Based Services, Internationalization and Regional Development*. Aldershot: Ashgate.

Buzás, N. (2005). From technology transfer to knowledge transfer: An institutional transition. In Alvstam, C. G. & Schamp, E. W. (eds.) *Linking Industries Across the World: Processes of Global Networking*. Aldershot: Ashgate.

Dicken, P. (2007). *Global Shift: Mapping the Changing Contours of the World Economy* (5th edn.). London: SAGE.

Florida, R. (2005). *The Flight of the Creative Class*. London: Harpercollins Publishers.

Freeman, C. (1994). The economics of technical change. *Cambridge Journal of Economics* 18, 463 514.

Ivarsson, I. and Alvstam, C. G. (2004). International technology transfer to local suppliers by Volvo trucks in India. *Tijdschrift voor Economische en Sociale Geografie* 95, 27 43.

Ivarsson, I. and Alvstam, C. G. (2005). The effect of spatial proximity on technology transfer from TNCs to local suppliers in developing countries: The case of AB Volvo's truck and bus plants in Brazil, China, India and Mexico. *Economic Geography* 81, 83 111.

Lall, S. (2000). Technological change and industrialization in the Asian newly industrializing economies: Achievements and challenges. In Kim, L. & Nelson, R. R. (eds.) *Technology, Learning & Innovation: Experiences of the Newly Industrializing Economies*, pp 13 68. Cambridge: Cambridge University Press.

Malecki, E. J. (1997). *Technology & Economic Development* (2nd edn.). London: Longman.

Nelson, R. R. (1994). The co evolution of technology, industrial structure and supporting institutions. *Industrial and Corporate Change* 3, 47 63.

OECD (2005). *OECD Science, Technology and Industry Scoreboard*. Paris: OECD Publishing.

Porter, M. E. (1990). *The Competitive Advantage of Nations*. London: Macmillan.

Schumpeter, J. A. (1934). *The Theory of Economic Development*. Oxford: Oxford University Press (original title in German: 'Theorie der wirtschaftlichen Entwicklung', 1912).

Storper, M. (1995). The resurgence of regional economies, ten years later: The region as a nexus of untraded interdependencies. *European Urban and Regional Studies* 2, 191 221.

UNCTAD (2001). *World Investment Report: Promoting Linkages*. United Nations Conference on Trade and Development. Geneva: United Nations.

Telecommunications

B. Warf, Florida State University, Tallahassee, FL, USA

Glossary

Digital Divide Social and spatial inequalities in access to the Internet.

Space of Flows The society and geography typified by hypermobile capital and information.

Time–Space Compression The shrinking of relative distances between places when measured in terms of travel or communication times.

In the massive rounds of time–space compression char acteristic of post Fordist capitalism, telecommunications play a central role in the organization and transformation of economic activity. As production systems have become stretched over ever larger distances, including the world wide spaces of the global economy, and as the division of labor has become markedly more information intensive, the need and ability to transmit vast quantities of data have grown accordingly. Telecommunications have been critical to this process for more than 150 years, acceler ating the flow of information across distance and bringing places closer to one another in relative space.

The origins of telecommunications can be traced to the invention of the telegraph in the early nineteenth century, which initiated the process by which com munications became detached from transportation. The telephone, starting in 1876, elevated this process to in clude aural information, and until the 1960s, tele communications was essentially synonymous with simple telephone service. The telephone quickly became critical to the growth of industrial city systems, allowing multi establishment firms to centralize their headquarters' functions in the cores of large cities while they spun off branch plants to smaller towns. Even today, despite a plethora of sophisticated digital technologies, the tele phone remains by far the most commonly used form of telecommunications for businesses and households.

In the 1980s and 1990s, as the cost of computing capacity dropped and its power increased rapidly with the microelectronics revolution, new technologies, particu larly fiber optics, and satellites, drastically increased the capacity of telecommunications. With the digitization of information in the late twentieth century, telecomm unications companies formed integrated information networks, most spectacularly through the Internet. Fiber optics lines, which now comprise the backbone of the global telecommunications industry, allow enormous quantities of data to be transmitted securely and almost instantaneously. Electronic Data Interchange and wireless services are becoming increasingly popular. A substantial literature has demonstrated the importance of tele communications in the globalization of producer services, particularly finance. Within cities, digital networks have contributed to an ongoing reconstruction of urban space.

This article begins by dispelling some popular myths about information systems, particularly simplistic post industrial, technological determinist views that argue telecommunications entail a uniform decentralization of economic activity and the 'end of geography' by ob literating the importance of proximity. Second, it dwells on the role of telecommunications systems in financial markets, particularly global banking and securities. Third, it examines the decentralization of clerical jobs at the local, regional, and global scales. Fourth, it focuses on urban infrastructural investments in this area, including telecommuting. Fifth, it addresses the origins and impacts of the Internet and social discrepancies in access to it.

Popular Misconceptions about Telecommunications

There exists considerable confusion about the real and potential impacts of telecommunications, largely as the result of early advocates of 'postindustrial' theory such as Daniel Bell and Alvin Toffler. This perspective, which has long been remarkably widespread among many aca demics and planners, hinges upon a simplistic, utopian technological determinism that ignores the complex, often contradictory, relations between telecommunications and local economic, social, and political circumstances.

For example, postindustrial theorists offered repeated proclamations that telecommunications would eventually allow everyone to work at home via telecommuting, dispersing all functions and signifying the eventual ob solescence of cities through the 'death of distance'. Yet such predictions have fallen flat in the face of the per sistence of growth in densely populated urban areas. In fact, telecommunications are generally a poor substitute for face to face contacts, the medium through which most sensitive corporate interactions occur, particularly when the information involved changes quickly, is ir regular and context dependent, proprietary, and un standardized in nature. Thus, the vast bulk of high wage, white collar, administrative command and control func tions remain clustered in downtown areas. In contrast, telecommunications are ideally suited for the transmis sion of routinized, standardized forms of data, facilitating the dispersal of functions involved with their processing (i.e., back offices) to low wage regions.

In short, the notion that telecommunications inevit ably lead to the dispersal or deconcentration of economic functions is erroneous; rather, information systems facili tate the simultaneous concentration and deconcentration of economic activities. While the costs of communications have decreased, other factors have risen in importance, including local regulatory regimes, agglomeration econ omies, and the cost and skills of the local labor force.

Telecommunications and Finance

The digital revolution in telecommunications generated the greatest impacts in financial markets. Banks, insur ance companies, and securities firms – all highly infor mation intensive industries – aggressively promoted the construction of an extensive worldwide network of leased and private fiber optics lines in the 1990s and early 2000s. A primary purpose of these systems is electronic funds transfer systems, which form the nervous system of the global financial economy, allowing banks to move capital around a moment's notice, arbitrage interest rate differ entials, take advantage of favorable exchange rates, and avoid political unrest. With the breakdown of the Bretton Woods Agreement in 1971 and the shift from fixed to floating exchange rates, electronic trade in national cur rencies skyrocketed. By some estimates, the world's banks move more than $1.5 trillion around the globe every day. In securities markets, telecommunications systems have facilitated global trading of stocks, bonds, and other equities through computerized trading programs. In es sence, under the influence of digitization, information and capital have become two sides of the same coin.

Moving through fiber optics lines at the speed of light as streams of digital zeros and ones, global money crosses the boundaries of nation states with such ease and in such volumes that national bank interventions have become increasingly problematic, making national monetary controls over exchange, interest, and inflation rates ever harder to sustain. It is simplistic to conclude that the nation state has no impact on this process, for national regulatory controls still exert some, if diminishing, degree of leverage. The simplistic notion of a perfectly seamless global economy has yet to materialize, if it ever will.

One arena in which telecommunications have power fully affected the financial industries is through the growth of offshore banking. Typically in response to highly favorable tax laws implemented to attract foreign firms, offshore banking has become an important source of revenue to many microstates in the Caribbean, Western Europe, the Middle East (e.g., Bahrain), and the South Pacific. As the technological barriers to moving money around internationally have steadily fallen, legal and regulatory ones have increased in importance, and financial firms have found the topography of regulation to be of the utmost significance in choosing investment lo cations. Generally such facilities are very capital inten sive in nature and generate relatively few jobs, although this aspect varies widely. Because such facilities are fre quently used by merchants of illegal commodities such as drugs and weapons to 'launder' money, offshore banking centers often operate under a cloud of suspicion by 'le gitimate' financial centers and governments in the eco nomically developed world.

Telecommunications and Global Cities

One of the most significant repercussions of the inter nationalization of financial markets has been the growth of 'global cities', particularly London, New York, and Tokyo. Global cities act as the 'command and control' centers of the world system, serving as the home to massive complexes of financial firms, producer services, and corporate headquarters of multinational corporations. In this capacity, they operate as arenas of interaction, facilitating dense networks of face to face contacts, access to repair and ancillary services, political connections, artistic and cultural activities, and elites to rub shoulders easily (e.g., through easy access to meetings). While other cities (e.g., Paris, Toronto, Los Angeles, Osaka, Hong Kong, and Singapore) certainly can lay claim to being important national cities in a global economy, the trio of New York, London, and Tokyo has played a dis proportionate role in the production and transformation of international economic relations in the late twentieth century.

At the top of the international urban hierarchy, global cities are simultaneously: (1) centers of creative innov ation, news, fashion, and culture industries; (2) metrop oles for raising and managing investment capital; (3) centers of specialized expertise in advertising and mar keting, legal services, accounting, computer services, etc.; and (4) the management, planning and control centers for corporations and nongovernmental organizations (NGOs) that operate with increasing ease over the entire planet. At their core, global cities allow the generation of spe cialized expertise upon which the management of the current global economy heavily depends. Each city is tied through vast tentacles of investment, trade, migration, and telecommunications to markets around the world. All three metropolises are endowed with enormous tele communications infrastructures that allow corporate head quarters to stay in touch with global networks of branch plants, back offices, customers, subcontractors, subsidiaries, and competitors. This phenomenon again illustrates how geographic centralization can be facilitated by telecomm unications.

However, telecommunications simultaneously threaten the agglomerative advantages of urban areas, particularly those obtained through face to face communications. This trend is particularly evident in finance, the bread and butter of global cities. For example, the National Association of Security Dealers Automated Quotation system (NASDAQ) has become the world's largest stock market by volume of shares; unlike many other exchanges, NASDAQ lacks a trading floor, connecting a million traders worldwide through fiber optics lines. Numerous stock markets have recently abolished their trading floors in favor of screen based trading.

Telecommunications and Back Office Relocations

The flip side of global cities is back office functions, another industry profoundly reshaped by the telecommunications revolution. Back offices essentially perform standardized functions such as data entry of office records, telephone books, or library catalogs, payroll and billing, bank checks, insurance claims, and magazine subscriptions. The jobs they generate are typically filled by semiskilled workers, primarily women. Unlike corporate headquarters, with their extensive backward and forward linkages to other sectors, back offices have linkages to other sectors. They tend to be energy intensive and require extensive data processing facilities, reliable sources of electricity, and high capacity telecommunications facilities.

Until the 1980s, back offices were clustered in the same building as headquarters' activities in downtown areas to facilitate close management supervision. However, as central city rents rose in the late twentieth century, a force evident in the corporate reclamation of many downtown regions and associated rounds of gentrification, many large firms detached their back offices from the headquarters, moving the former out of the urban core to lower cost locations in the suburbs, for example, in office parks. Suburban areas typically offer easier access via the automobile, lower rents, and better skilled workers. As the international telecommunications infrastructure grew by leaps and bounds, many firms began to relocate their back offices on much broader spatial scales. For example, many financial and insurance firms and airlines moved their back offices from expensive sites in metropolises such as New York, San Francisco, and Los Angeles to relatively low wage labor markets in the Sunbelt. At the global scale, the relocation of back offices took the form of the overseas offshore office. Often the industries involved were subject to enormous competitive pressures brought on by de regulation and globalization, and were intensely interested in the cost savings generated by relocations to lower wage regions. In the 1990s, for example, numerous American insurance firms established back office facilities in Ireland with the active encouragement of the Irish Development Authority. Likewise, US airlines utilized offices in Anglophone Caribbean countries such as Jamaica and Barbados, and to a lesser extent, the Dominican Republic. Manila, in the Philippines, has emerged as a back office center for British firms, with wages 20% of those in the UK. The Indian city of Bangalore has developed a famous comparative advantage in call centers, which provide customer assistance to clients of firms such as computer software, telecommunications, banks, airlines, and international hotel chains. Thus, in addition to fueling the growth of high skilled jobs in metropolitan cores, the telecommunications revolution simultaneously encouraged the flight of many low wage, low value added jobs to the developing world. Far from being a simplistic 'end of geography', this process has led to new rounds of centralization and decentralization.

Telecommunications and Urban Space

Telecommunications have had profound effects on the organization of urban space. Although large cities typically have much better developed telecommunications infrastructures than do small ones, the technology has rapidly diffused through most national urban hierarchies. In the future, therefore, the competitive advantages based on telecommunications will diminish, forcing competition among localities to occur on the cost and quality of labor, taxes, and local regulations. Regions with an advantage in telecommunications generally succeed because they have attracted successful firms for other reasons.

One increasingly important effect of new information systems is 'telework' or 'telecommuting', in which workers substitute some or all of their working day at a remote location (almost always home) for time usually spent at the office. Telework is most appropriate for jobs involving mobile activities or routine information handling such as data entry or directory assistance. Proponents of telework claim that it enhances productivity and morale, reduces employee turnover and office space, and leads to reductions in traffic congestion, air pollution, energy use, and accidents. The extent to which telecommuting has become popular is unclear, but some estimates hold that roughly 5% of the US labor force engages in this activity at least once per week. Future growth of telecommuting will encourage more decentralization of economic activity in suburban areas.

However, the growth of teleworking may ironically increase the demand for urban transportation services rather than decrease it, as was widely expected. While telecommuters travel to their workplace less frequently than do most workers, many have longer 'weekly' commutes overall, leading to a rise in aggregate distances traveled. In addition, the time released from commuting may be utilized traveling for nonwork related purposes. Thus, while telecommuters may have different reasons for traveling, the frequency or volume of trips may not change. Moreover, even if telecommuting did reduce some trips, the associated reductions in congestion could simply induce other travelers onto the roads. In short, whether or not an actual trade off between telecomm unications and commuting occurs (their substitutability rather than complementarity) is not clear.

Another potential impact of information systems on urban form concerns transportation informatics, in cluding smart metering, electronic road pricing, auto mated toll payments and turnpikes, automated navigation and travel advisory, remote traffic monitoring and dis plays, and improved traffic management and control systems. Such systems are intended to minimize con gestion and optimize traffic flow, especially during the critical rush hour peak travel times. The rapid growth of wireless technologies, particularly cellular phones (which constitute the majority of phones on the planet today), allows commuters stuck in traffic to use their time more productively. Such systems do not so much comprise new technologies as the enhancement of existing ones.

The widespread use of telecommunications has led to important changes in the nature of urban and regional planning. Desperate for investment capital and jobs, many localities vie for one another with ever greater concessions to attract firms, forming an auction that re sembles a zero–sum game. The effects of such a com petition are hardly beneficial to those with the least purchasing power and political clout. Left to sell them selves to the highest corporate bidder, localities fre quently find themselves in a 'race to the bottom' in which entrepreneurial governments promote growth – but do not regulate its aftermath – via tax breaks, subsidies, training programs, looser regulations, low interest loans, infrastructure grants, and zoning exemptions. As a result, local planners have become increasingly less concerned with issues of social redistribution, compensation for negative externalities, provision of public services, and so forth, and more enthralled with questions of economic competitiveness, attracting investment capital, and the production of a favorable 'business climate'.

The Internet

By far the largest and perhaps most important tele communications network is the Internet, which con nected an estimated 1 billion people in more than 200 countries in 2006, or roughly 15% of the planet's population. From its military origins in the US in the 1960s, the Internet emerged upon a global scale through the integration of existing telephone, fiber optics, and satellite systems, a process made possible by the tech nological innovation of packet switching, in which indi vidual messages maybe decomposed, the constituent parts transmitted by various channels, and then re assembled instantaneously at the destination. Spurred by declining prices of services and equipment, the Internet is the most rapidly diffusing technology in history, allowing any individual with a microcomputer and modem to 'plug in' to cyberspace. However, given rela tively low incomes, personal computers in poorer coun tries are often difficult to afford; in such contexts, cybercafes often provide a major point of access.

The Internet has had significant economic and social impacts. For well educated elites, it offers a means to be connected to the global economy while becoming in creasingly disconnected from the local environments of their own cities and countries. Online retail shopping, or 'e tailing', has grown exponentially, particularly for large firms (e.g., Wal Mart) and given birth to entirely virtual stores such as Amazon.com. The Internet is central to reservation and ticket systems for airlines and hotels. The higher levels of competition that the Internet has facili tated have reduced costs for consumers. Many schools and universities use it for online courses, allowing stu dents who otherwise might not have access to their ser vices to partake. Most newspapers offer online versions, greatly facilitating access to the news and entertainment. Downloading of music files and films has become a big business, raising concerns over intellectual property rights. Internet based telephony (e.g., Skype) has grown rapidly. e Business, or Electronic Data Interchange, allows firms to advertise, recruit, submit contracts and bills, monitor in ventories, and perform other functions quickly and cheaply, reducing transactions costs and raising product ivity. The use of the Internet for political purposes, in cluding weblogs or 'blogs', has increased the domain of popular political participation. Graphical interfaces have made the Internet a major source of recreation, including online games, while others browse for pornography, rais ing widespread concerns about its impacts on traditional mores, especially with regards to children. For many users, cyberspace allows for the creation of 'communities with out propinquity', that is, groups of users who share com mon interests but not physical proximity. Political groups of various types and marginalized groups such as gays have used the technology to great advantage. The extent to which digital relations can substitute for 'real world' face to face interactions, however, is open to debate.

However, the Internet may be used for illegal, un savory, or illegitimate purposes. 'Spam', or unsolicited e mail advertising, is estimated to comprise the bulk of electronic traffic on the Internet. Online gambling has grown rapidly. Electronic systems are used to monitor everyday life, including credit cards, visas and passports,

Figure 1 Internet hosts per 100 000 persons, January 2006.

tax records, medical data, police reports, telephone calls, utility records, automobile registration, crime statistics, and sales receipts. The growth of Internet fraud and identity theft raises serious concerns about privacy. Thus, telecommunications can be used against people, as well as for them. The unfortunate tendency in the popular media to engage in technocratic utopianism largely obscures these power relations. *Contra* this perspective, social categories of wealth and power and geographical categories of core and periphery continue to be reinscribed in cyberspace.

Despite utopian expectations that the Internet would obliterate the importance of social and spatial differentials, there remain significant discrepancies worldwide in terms of access to the Internet. In all countries, access to computers linked to the Internet, either at home or at work, is highly correlated with income, educational level, and employment in professional occupations. Because they are often not well served by telecommunications firms, rural areas often have slow or substandard connections. The elderly often find the Internet to be intimidating, although they comprise the fastest growing demographic group of users. Difficulties in procuring the skills, equipment, and software necessary to get Internet access threaten to exclude economically and socially marginalized groups from the benefits of cyberspace, a phenomenon that reflects the growing inequalities throughout industrialized nations generated by labor market polarization. Thus, the microelectronics revolution has accentuated the digital divide between groups that are proficient with computer technology and those who neither understand nor trust it, reinforcing and deepening the inequalities in broader social relations.

At the international level, inequalities in access to the Internet reflect the long standing bifurcation between the First and Third Worlds (**Figure 1**). The best connected nations are in Scandinavia, Canada, Australia, and Singapore, where more than 70% of the population is connected. In the US, the Internet penetration rate was 55% in 2005 (with substantial internal variations); however, due to the enormous size of the American economy, 80% of all international Internet traffic is either to or from the US. Outside of the economically advanced First World, the vast bulk of the world's people, that is, the Third World, have comparatively little access, particularly in Africa. Low incomes, inadequate infrastructures, and unhelpful or oppressive governments play major roles in this context.

The Oligopolization of Telecommunications

Telecommunications, an industry characterized by high fixed costs and low marginal ones, has long experienced waves of corporate consolidation. Over the last three decades, two overlapping and intersecting industries, telecommunications and the media, have been dramatically transformed by technological change and deregulation. The microelectronics revolution annihilated geographic and industry boundaries by making it possible for firms in one sector to produce outputs easily used in

Table 1 World's largest seven international telecommunications carriers, 2005

Company	*Origin country*	*Assets ($ billions)*	*Employees (thousands)*
AT&T	USA	252.3	156.7
Nippon Telephone & Telegraph	Japan	170.3	215.2
Versión	USA	164.7	258.3
France Telecom	France	118.2	188.9
Deutsche Telekom	Germany	115.2	211.9
WorldCom	USA	98.9	51.5
SBC Communications	USA	98.6	209.8

Source: www.forbes.com.

another; digital convergence – the blurring of the traditionally separated industries of telephone, cable, and computers – allows conglomerates to provide more than one service over the same medium. In the United States, the Telecommunications Act of 1996 ushered in significant changes in the industry's regulatory environment, changing foreign ownership controls, allowing telecommunications firms to acquire and control foreign firms, and relaxing cross industry ownership rules that prohibited firms in one telecommunications sector to operate in and control firms in another. Large telecommunications firms relied on horizontal integration to enhance their competitiveness and penetrate new markets, leading to a tsunami of mergers and acquisitions.

The result has been an increasingly oligopolistic market structure. Across the planet, the market for media services has become dominated by a few giants that have established powerful distribution and production networks (**Table 1**). Corporate control over the world's production and distribution of information includes the five largest media conglomerates, AOL Time Warner, Walt Disney, Viacom, News Corporation, and Vivendi, behemoths that own numerous information related sectors, including book publishing, cable and network television, radio, movie studios, and music companies. Such firms tend to be led by well known media CEOs, including, for example, Michael Eisner of Disney, Cable News Network's Ted Turner, Italy's Silvio Berlusconi, and Thomas Middelhoff of Bertelsmann, the robber barons of the digital age. These companies' size, financial resources, infrastructure, and human capital effectively give them control over a vast share of the world's media, the channels through which billions of people get their news and entertainment, and thus collectively play a hugely influential role in shaping the outlooks and views of their viewers and subscribers.

Conclusions

Economic activity today is overwhelmingly dominated by the production of intangibles, in which knowledge, information, and communications are critically important. The geography of the world economy rests heavily upon invisible flows of data and capital, binding places unevenly to the world system. Telecommunications has also transformed the ways in which economic landscapes are constructed, suturing localities to global 'spaces of flows' that move at the speed of light and giving firms unprecedented mobility and flexibility. Under the impetus of the microelectronics revolution, telecommunications ushered in a vast round of time–space compression, linking places together to a historically unprecedented degree. Markets, labor processes, transportation routes, planning policies, and spatial structures have changed accordingly.

Cities and regions search for a competitive advantage in this world in several ways. In large metropolitan areas with dense complexes of firms bound together by face to face interactions, telecommunications have left skilled, high wage functions largely intact, but not unchallenged – as the electronic trading of stocks suggests. By allowing firms to stay in contact with operations around the world, telecommunications have contributed to the centralization of key activities in global cities such as New York, London, and Tokyo, which rely upon their extensive connections to the global telecommunications infrastructure to serve as nerve centers of the world economy. For such tasks, telecommunications have been largely unable to substitute for face to face interactions. Other services that process routinized information, however, and rely upon unskilled, low wage labor, are highly vulnerable to substitution by new telecommunications systems and have decentralized (e.g., back offices). A key theme in understanding the new economic landscapes of the information economy is that telecommunications tend to reinforce the agglomeration of high wage, high value added, white collar functions while decentralizing low wage, low value added, blue (or pink) collar ones.

In short, telecommunications have a variety of impacts upon cities and regions, both positive and negative, local and global, centralizing and decentralizing. To appreciate the complexity of these effects requires a step back from the simple utopianism and technological determinism that tends to pervade much public opinion and policy making. Telecommunications systems are a social product, interwoven with relations of class, race, gender, and power. Given the rapid rate of technological change in the late twentieth century, predictions about the future of telecommunications are hazardous at best. Capitalism has had a very long history of technological and economic changes that periodically refashion local and global landscapes; the information revolution is but the latest chapter in this story.

See also: Cyberspace/Cyberculture; Information Technology; Internet, Economic Geography; Space II.

Further Reading

Cairncross, F. (1997). *The death of distance: How the communications revolution will change our lives*. Boston: Harvard Business School Press.

De Sola Pool, I. (ed.) (1977). *The social impact of the telephone*. Cambridge, MA: MIT Press.

Gibbs, D. and Tanner, K. (1997). Information and communication technologies and local economic development policies: The British case. *Regional Studies* 31, 768 774.

Graham, S. and Aurigi, A. (1997). Virtual cities, social polarization, and the crisis in urban public space. *Journal of Urban Technology* 4, 19 52.

Graham, S. and Marvin, S. (1996). *Telecommunications and the city: Electronic spaces, urban places*. London: Routledge.

Kurtzman, J. (1993). *The death of money*. Boston, MA: Little, Brown and Co.

McChesney, R. (1999). *Rich media, poor democracy: Communication politics in dubious times*. Champaign, IL: University of Illinois Press.

McChesney, R. (2001). Global media, neoliberalism, and imperialism. *Monthly Review* 52, 165 178.

O'Brien, R. (1992). *Global financial integration: The end of geography*. Washington: Council on Foreign Relations.

O'Kelly, M. and Grubesic, T. (2002). Backbone typology, access, and the commercial Internet, 1997 2000. *Environment and Planning B* 29, 533 552.

Schiller, D. (1999). *Digital capitalism: Networking the global market system*. Cambridge, MA: MIT Press.

Shields, R. (ed.) (1996). *Cultures of Internet: Virtual spaces, real histories, living bodies*. London: Sage Publications.

Solomon, E. (1997). *Virtual money: Understanding the power and risks of money's high speed journey into electronic space*. Oxford: Oxford University Press.

Warf, B. (1995). Telecommunications and the changing geographies of knowledge transmission in the late 20th century. *Urban Studies* 32, 361 378.

Warf, B. (2003). Mergers and acquisitions in the telecommunications industry. *Growth and Change* 34, 321 344.

Wheeler, J., Aoyama, Y. and Warf, B. (eds.) (2000). *Cities in the telecommunications age: The fracturing of geographies*. London: Routledge.

Relevant Websites

http://www.webmergers.com/

http://opennet.net/
Collaborative Partnership of Four Leading Academic Institutions.

http://faculty.washington.edu/krumme/550/resources.html
Geography of the Information Economy.

http://www.internetworldstats.com/
Internet World Stats, Usage and Population Statistics.

http://www.sibis eu.org/
Statistical Indicators, Benchmarking the Information Society.

http://www.ejisdc.org/ojs2/index.php/ejisdc
The Electronic Journal of Information Systems in Developing Countries.

Territorial Production Complexes

B. Domański, Jagiellonian Univeristy, Krakow, Poland

Glossary

Industrial Complex Analysis A method of analysis of the linkages between industries in the industrial complex developed in regional science.

Input–Output Analysis A method of analyzing interindustry connections in the economy, where material and financial flows between sectors are presented as rows (outputs) and columns (inputs) of a matrix (table).

Least-Cost Location Theories Conceptualizations of industrial location based on the assumption of the minimization of costs of inputs, especially materials and labor, rather than the maximization of profit.

Technical–Economic Indicators Indices measuring the amount of materials, energy, labor, and transportation inputs per ton of the final product and/or per worker.

Vertical Integration The centralization of productive units which represent various stages of production process, including upstream suppliers and downstream buyers.

A Basic Explanation of the Concept

The territorial production complex (TPC) is a concept introduced in Soviet planning of spatial organization of the economy – industrial production, in particular.

The classic definition given by M. K. Bandman (1976: 15) defines a TPC as "an aggregate to be created in accordance with the plan of persistently interrelated and proportionally developing units of national economy sectors (industry, agriculture, construction, transport, non productive sphere) which…are concentrated over a relatively limited and compact (not disunited) area." In other words, the TPC is a planned economic entity consisting of various economically linked industrial and nonindustrial activities in a particular geographical area.

Two types of TPCs can be distinguished on various geographical scales:

1. industrial centers (towns)
2. industrial regions or vast interregion areas (macro scale TPCs); they may cover several local TPCs.

There are three major groups of interrelated components of the TPC which have to be harmonized: production, infrastructure, and local resources. The key element is specialization industries, the development of which provides a rationale for the planning of the TPC. These industries are accompanied by complementary productive activities, which are auxiliary to the specialization of the complex or service to the local population. Infrastructure is understood in a very broad sense as everything which serves production, from railways and power transmission lines to housing, education, and healthcare facilities. Local resources (inputs) include natural resources and manpower (see **Figure 1**).

The TPC is a deliberately planned structure created in order to achieve a rational spatial organization of economic activities and infrastructure. Its aim is illustrated by the following quotation from Bandman:

> Optimization of TPC creation is referred to as determination of the best (from the viewpoint of the national economy) variant of utilizing all kinds of resources, of time limits, rates, scope and proportions of development; of production structure and spatial organization of all elements of the economy as a complex, its external and internal relationships on the condition that there should be maximum savings in social labour input in order to fulfil the task of producing the output of specialization industries and providing fixed living conditions for the population. The basic function of the TPC is thus to produce specific quantities of goods (and necessary services) as efficiently as possible, with minimization of labour expended as the operative criterion. (Bandman, 1976: 19 20)

Thus, the task is to design the optimum form for the TPC, which means selection of appropriate proportions of various components (economic activities, infrastructure, labor force, and natural resources) and their spatial distribution, together with linkages of all the elements. This can only be achieved by central planning.

The main rationale for planning the TPC is making savings in investment outlays on infrastructure (including social infrastructure) and transportation costs. Analytical measures used in calculating the optimization of the TPC include the following: reduced use of land for factories and new residential housing, decreased length of railway lines and other infrastructure (e.g., electricity lines), lower expenditures on social infrastructure, fuller use of local natural resources, byproducts, and manpower, both male and female. Probst and Secomski emphasized that the aim is to maximize the use of all the stock of infrastructure. Savings are derived from the economies of

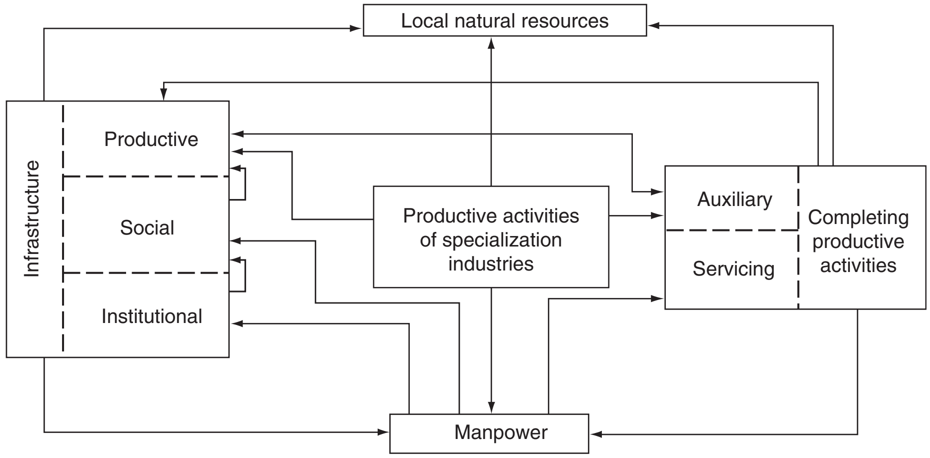

Figure 1 Interrelationships among elements of a territorial production complex affecting its planned development. Reproduced from Bandman, M. K. (1976). Content, succession, and tools of optimizing creation of territorial production complexes. In Bandman, M. K. (ed.) *General questions of modelling territorial production complexes*, p. 33, Chart 2. Novosibirsk: USSR Academy of Sciences, Siberian Branch, Institute of Economics and Organization of Industrial Production, by permission of the Russian Academy of Sciences.

scale of large production and infrastructure units, agglomeration economies of better utilization of capacity of interlinked plants and nonproductive facilities and reduced transportation costs.

A set of technical–economic indicators all play an important role. They include indices measuring the amount of various material and energy inputs per ton of the final product, similar to Alfred Weber's material indices and/or interindustry coefficients expressing linkages between various industrial branches in input–output analysis, as well as inputs and output per worker and transportation effort. The technical–economic indicators provide the basis for planning the composition of and linkages within the TPC. In addition, they may reveal the significance of various location factors in planning the TPC.

The planning of a TPC has two successive stages: the economic structure is determined first, and the spatial (territorial) structure is decided second. Thus, at the beginning the basic proportions of industrial branches in the structure of production are specified. Then, the functional units of the economy in the form of aggregates of specialization industries, their supporting activities and infrastructure, transport facilities, and manpower are assigned to specific territorial units.

The concept of a TPC was primarily elaborated in the planning of the industrial development of the undeveloped lands of Siberia. TPCs were projected here on a macro spatial scale and could cover vast territories of more than 1000 km in diameter. The core industries were those utilizing the natural resources of the region.

Since the TPCs were mainly established in sparsely populated lands of the inhospitable eastern frontier, the shortage of labor was one of the major issues here. Wages were used as an economic incentive to encourage migration from the European core regions, and forced labor could also have substantial significance. This also explains the strong emphasis on the maximum savings of manpower input in the optimization of the TPC.

The most famous and widely discussed example is the Urals–Kuznetsk Combine created in the 1930s. It was originally meant to unite the iron ore of Magnitogorsk in the Urals with the vast coal reserves of Kuznetsk Basin (Kuzbas) in Western Siberia, some 2000 km away by rail (**Figure 2**). Later formulations led to the integration of various economic activities of broader regions of the Bashkir Autonomous Republic and Kazakhstan. Later applications of the TPC include the Sayan TPC in the Angara–Yenisey region of Siberia related to the huge hydroelectric power stations at Bratsk, Ust' Ilimsk and Krasnoyarsk, Rudnyy Altay TPC oriented to nuclear weapons components and products for the aviation industry (Semipalatinsk and Ust' Kamenogorsk), and the Norilsk region for nickel and related metals.

The major intellectual center where the concept and mathematical modeling of TPCs were developed in the 1970s and 1980s, was the Institute of Economics of the USSR Academy of Sciences in Novosibirsk, headed by M. K. Bandman. The idea of TPCs was introduced to American and Western European geographers, regional scientists, and economists by the translation of N. N. Kolosovskiy's article in 1961, but aroused broader interest in the decades that followed.

The Characteristics of a TPC in Relation to Other Concepts and Approaches

The notion of a TPC avoids focus on a single plant and/or firm, which is a weakness of classic location theories of the Weberian type. The concept of the TPC

Figure 2 Location of selected territorial production complexes in the Soviet Union.

applies to all sorts of economic activities and various geographical scales at the same time. In addition, it is not just concerned with the question of location of economic activities, but also takes into account linkages between these activities and between the places where they are situated. Consequently, the TPC approach provides a far more comprehensive perspective on the spatial organ ization of economic activities than many other view points. In fact, the TPC may be regarded as a spatial system: a large, complicated entity consisting of a number of elements and the relationships (linkages) between them. The popularity of systems analysis in the heyday of the development of the TPC concept in the Soviet Union may explain the fact that the latter bears several prop erties of a system, even though it has rarely been ex plicitly discussed as such. It can also be argued that the TPC might be interpreted in the categories of a network.

The TPC has to be distinguished from a combine (Russian and German *Kombinat*) – another typical elem ent of state socialist industrial organization. The combine is a large organization usually comprising vertically in tegrated production activities. It can be seen as a socialist equivalent of the capitalist corporation with several en terprises under centralized control (one decision making body). The *Kombinat* can cut through divisions between various industries and combine activities normally belonging to different ministries and/or industrial associations in a single organizational unit. In general, the combine is basically an industrial organizational entity and the territorial dimension is not a constituent element of its formation, in contrast to a TPC. A typical TPC may include several combines.

The TPC can be seen as one of the attempts to link the perspective of industry and place. It belongs to the family of approaches which perceive spatial location and linkages by and large as technology driven. As a result, on an analytical level, the TPC shows some resemblance to the industrial complex analysis developed by Walter Isard and other regional scientists in the 1950s. They share the use of a sort of input–output analysis in studying the links between industries in industrial complexes.

There are several important differences between the concept of TPC and the notion of cluster. Among others, TPCs are constructed mainly in newly developed regions with low population density and involve production chains between different industries, especially producer oriented heavy industries and machinery, in contrast to cluster firms specialized in production of buyer oriented goods and services, where information flows and con nections with Research and Development (R&D) and other supporting institutions are crucial.

The essential feature of the TPC, making it funda mentally different from all the concepts conceived in

Western geography and economics, is that it was de veloped for the practical purposes of determining and creating the territorial organization of production, rather than simply analyzing or explaining it.

At the advanced stage, especially in the form de veloped by the Novosibirsk researchers grouped around M. K. Bandman, the TPC rested on the use of sophisti cated econometric modeling. It aimed to optimize re gional and local economic structures according to the economic goals established on the national level.

The TPC was a concept meant to bridge the basic division into economic and territorial (spatial) planning and decision making in the centrally planned economy. It was believed to provide a tool for both economic plan ning and regional development at the same time.

How Is the Concept Embedded in the Socialist Economic and Planning Ideas?

The idea of TPCs is based on a belief in the necessity and possibility of designing and controlling economic life at the macro spatial level. This belief was one of the fundamental assumptions of the centrally planned economy of the Soviet Union and other state socialist countries.

Views on location criteria of industry had evolved in the Soviet Union. Until the early 1930s, there were many voices supporting dispersal of production and population as a direct spatial expression of the egalitarian aims of socialism. TPCs manifested the departure from the idea of dispersal and towards the seeking of advantages of spatial concentration. It can be argued that the TPC represents the most advanced and comprehensive con cept of the spatial organization of economic activities, and industrial production in particular, in centrally planned economies. At the same time, it inevitably re flects the limitations and weaknesses of these economies and their underlying ideology.

Serious doubts exist as to whether the TPCs were really implemented, due to the often cited lack of co ordination between autonomous 'functional' ministries in the Soviet Union. The major inherent contradiction in the central planning system was a permanent conflict between economic ministries and the territorially or ganized administration. The former were always more powerful as they controlled the basic assets. Probst, along with other Soviet researchers of spatial organization of production and TPCs, emphasizes strongly that the TPCs required creation of organizational conditions, where the TPC was designed as a single entity from the beginning, and then the creation of a common governing body. This rarely happened, as the entire political and economic system was built on branch industrial minis tries responsible for individual sectors. The only form of 'coordination' of economic and territorial planning was the subordination of the latter to the former. Thus, the solution to this problem in the case of the Urals–Kuz netsk TPC was an attempt to create the Urals–Kuznetsk combine. Nevertheless, the typical TPC would consist of dozens, if not hundreds, of huge enterprises and com bines, which could be controlled by many ministries. As a result, the TPC was probably too large and complex a structure to be managed as a single *Kombinat*. In addition, the TPC as a combine would have reduced the power of branch ministries, which were the pillars of the Soviet administrative system.

In general, the significant practical and theoretical deficiency of the TPC was that it did not allow for the recognition and subsequent reconciliation of conflicting interests and objectives of various agents involved in economic activity: ministries, combines and enterprises, regional and local authorities, not to mention various social groups. In fact, socialist planners faced immense difficulty in resolving contradictory goals and actions, which are endemic to any large complex social system (organization) that has to cope with scarcity. They struggled to find a formula for resolving these conflicts in the immense system of the centralized state.

Two central attributes of TPCs could be subject to criticism: the top down approach and the subordination of social to economic objectives.

The notion of TPCs was developed under the ideo logical system resting on the conviction that the state performs the mission of creating a new, better social order. This led to the seeking of large scale, top down solutions to problems of regional and local development. The idea of a central decision making body, ignoring local agency, sharply contradicts the contemporary trend to promote the bottom up approach and emphasizes the role of indigenous agents, local institutions, civil society, and social participation.

Another important issue is the question of the aims of TPCs. Bandman's definition makes it clear: "The basic function of the TPC is thus to produce specific quantities of goods (and necessary services) as efficiently as possible, with minimization of labor expended." This is in ac cordance with the overriding objective of central socialist planners concerned with the growth of industrial pro duction perceived as an element of historical necessity. In other words, there is little doubt that the TPCs had to develop according to productive rather than social in vestment. Social benefits (aims) and costs were sub ordinated to economic objectives. This is expressed by the fact that fixed living conditions for population appear as constraints, not as goals in the optimization modeling of the TPC. Moreover, services for population are re garded as supporting ('completing') activities only, which are needed as long as they serve the functioning of the specialization industries. The criterion for optimizing the

TPC is the minimization of labor required for a given set of outputs. All this reflects an instrumentalist approach to society reduced to manpower consumed by production facilities and assigning socialist towns the function of storage place for the labor force. Critics attribute this to the economism and reductionism of Marxian political economy.

Other Limitations and Weaknesses of the Concept

It is striking that the TPC borrows two important ideas from Alfred Weber's location theory:

1. The quest for minimization of costs of input (materials and labor), and
2. The vital role of transportation.

The affinity with the least cost location theories rather than market area approach (e.g., Lösch) is understand able in the case of supply side socialist economy. The disadvantage is that demand (consumption) is absent here, except for the output targets determined in an unspecified way by central decision makers. At the same time, the basic difference remains that Weber's analysis focused on the location of a single plant, while the TPC concerns entire industrial complexes of activities toge ther with the accompanying wide ranging infrastructure.

The use of a form of Weberian locational analysis, where physical properties of materials and products are considered in relation to transportation costs, can be explained more easily if we realize that one of the major objectives of industrial location in the Soviet Union was to ease tightening bottlenecks in freight movements on railways.

The TPC is strongly oriented towards the utilization of natural resources and fits commodity chains linked by vertical integration. This may be the reason why in other socialist countries outside the Soviet Union the concept of TPC was mainly present in the academic world, with little practical impact on industrial planning.

The emphasis is on technical and economic linkages, whereas social and institutional relationships are ignored. This limitation is shared with industrial complex analysis. This also makes the TPC different from clusters, where numerous independent enterprises are concentrated, linked by various sorts of cooperation and competition.

Environmental issues were practically neglected in the planning of TPCs, except for some lip service paid in the late writings of M. K. Bandman and his group.

Conclusion

TPCs represent an attempt to create an integrated sys tem of linked economic activities which are theoretically planned to ensure the coordination of productive and infrastructure investment. They were perceived as an opportunity to carry out the changes in the spatial or ganization of industrial production and to plan the ra tional distribution of other elements, such as infrastructure and labor force, according to the needs of the former. They were basically a planning concept grounded in the overall system of the centrally planned economy of the Soviet Union under state socialism. At the same time, they stemmed from the need for new planning and management mechanisms and aimed to overcome one of the critical weaknesses of the central planning and decision making system, namely the di vergent aims and actions of ministries responsible for various industries and territorial planning. The lack of such coordination underlay practical difficulties in the implementation of TPCs.

There have been attempts to search for parallels with planning systems of transnational corporations. We can also find some analogy with industrial complex analysis at the technical level, as this too is characterized by its interest in interindustry linkages. However, the approach of TPCs fundamentally differs from any theoretical and planning concepts developed and/or applied in market economy: first by its comprehensive (total) character, and second by its practical orientation towards creation of, rather than intervention in, the spatial organization of economic activities.

TPCs led to spatial concentration of economic activities on a local and regional scale, but they were able to facilitate dispersal on an interregional scale, promoting growth in non European regions of the Soviet Union.

See also: Industrial Location; Input-Output Analysis.

Further Reading

Aganbegyan, A. G. and Bandman, M. K. (1984). Territorial production complexes as integrated systems: Theory and practice. *Geoforum* 15, 25 32.

Bandman, M. K. (1976). Content, succession, and tools of optimizing creation of territorial production complexes. In Bandman, M. K. (ed.) *General questions of modelling territorial production complexes*, pp 5 100. Novosibirsk: USSR Academy of Sciences, Siberian Branch, Institute of Economics and Organization of Industrial Production.

Bandman, M. K. (ed.) (1985). *Regional development in the USSR: Modeling the formation of the Soviet territorial production complexes*. New York: Pergamon Press.

Baransky, N. N. (1956). *Economic geography of the USSR*. Moscow: Foreign Languages Publishing House.

Barr, B. M. (1974). The changing impact of industrial management and decision making on the locational behavior of the Soviet firm. In Hamilton, F. E. I. (ed.) *Spatial perspectives on industrial organization and decision making*, pp 411 446. London: John Wiley and Sons.

Domański, B. (1997). *Industrial control over the socialist town: Benevolence or exploitation*. Westport, CT: Praeger.

Holzman, F. D. (1957). The Soviet Ural Kuznetsk combine: A study in investment criteria and industrialization policies. *The Quarterly Journal of Economics* 71, 368 405.

Kolosovskiy, N. N. (1961). The territorial production combination (complex) in Soviet economic geography. *Journal of Regional Science* 3, 1 25.

Koropeckyj, I. S. (1967). The development of Soviet location theory before the Second World War. *Soviet Studies* 19, part one 1 28; part two 232 244.

Linge, G. J. R., Karaska, G. J. and Hamilton, F. E. I. (1978). An appraisal of the Soviet concept of the territorial production complex. *Soviet Geography* 19, 681 697.

Pilipienko, I. (2005). *Clusters and territorial industrial complexes Similar approaches or different concepts? First evidence from analysis of development of Russian regions*. Paper at the 45th Congress of the European Regional Science Association, Amsterdam.

Probst, A. (1971). *Voprosy razmeshcheniya socyalisticheskoy industrii*. Moscow: Izdatyelstvo Nauka.

Probst, A. (1974). Certain questions on the general theory of the location of industry. In Demko, G. J. & Fuchs, R. J. (eds.) *Geographical perspectives on the Soviet Union*, pp 352 364. Columbia: Ohio University Press.

Rutt, S. (1986). The Soviet concept of the territorial production complex and regional development. *Town Planning Review* 57, 425 439.

Smith, D. M. (1981). *Industrial location: An economic geographical analysis*. New York: John Wiley and Sons.

Territory and Territoriality

D. Delaney, Amherst College, Amherst, MA, USA

Introduction

Territory and territoriality are among the most basic and significant terms in human geography. Territories and territorial ensembles are fundamental elements of human sociospatial organization. Although most commonly as sociated with the modern nation state, this is merely one – albeit, highly significant – manifestation of terri toriality. Territoriality is much more pervasive. Indeed, it is inescapable. Its operations have become increasingly recognized in other 'nonpolitical' contexts.

The present era can be regarded as something of a golden age of reflection on territory and territoriality within human geography and other disciplines. This, in part, is the result of increasing theoretical pluralism and interdisciplinarity. How territory looks from one theo retical or disciplinary perspective may differ markedly from how it looks from other perspectives. Scholarly cross fertilization has been productive of novel under standings. The increasing attention to territory is also the result of profound perceived transformations in the worlds of experience associated with, among other things, globalization and the emergence of the Internet. In short, the ways in which territory and territoriality had ap peared to work have changed sufficiently to require fundamental rethinking. This reinvigorated scholarship highlights what may be the central paradox of the topic. Long regarded as a social device for simplifying and clarifying the scope and limits of power through its in side/outside, either/or dichotomizing effects, territori ality has been revealed to be complex, subtle, often ambiguous and unstable. Still, it continues to 'work', that is, to have effects, even if many of these effects are not intended by anyone.

This article examines some of the complexities of territory and territoriality. First, it touches on some preliminary and definitional issues. This is followed by an analytic examination of the topic in terms of the con nections with space, power, and meaning. Finally, the topic is situated historically vis à vis a discussion of the relationship to time and notions of modernity.

Preliminaries Issues

Basic Definitions

A territory is a bounded, meaningful social space the 'meanings' of which implicate the operation of social relational power. It is, in a sense, an expression of the fusion of meaning, power, and social space. The proto type territory may be that of the territorial nation state. Here, territory delimits the spatial scope and limits of sovereignty, jurisdiction, administration, and citizenship. But there are innumerable other generic forms and ex pressions of territory. Many directly implicate govern ance, for example, municipalities, voting and school districts, indigenous reserves, police precincts, and na tional parks. However, many exist in and give spatial expression to the so called private realm – residential lots, the micro spaces of work places, dioceses, and sales territories of businesses. In some situations even the human body can usefully be regarded as a kind of ter ritory. Some territories are formal and relatively enduring (e.g., Bolivia). Some are informal and relatively ephem eral (my seat in a diner). Some aspire to general com prehensive application (the nation state); some are very specific (the protest buffer zone around an abortion clinic).

Among the most pervasive and significant, if over looked, territories are those organized by 'the public/ private distinction' itself. Regardless of the form it takes, one never encounters a territory in isolation. Rather, territories are organized in relational ensembles or mo saics that have the effect of differentiating segments of social space. Again, this pertains as much to the smaller territorial constitution of rooms, cells and neighborhoods as to the global grid of the international system of states, as well as to intermediate configurations such as those associated with federal political structures. This differ entiating is not simply a matter of sorting. Rather, it entails the inscription of meaning (e.g., in the form of rules, or the terms of access, entry, confinement, or ex clusion) and the spatialization of power of various kinds.

The term 'territoriality' is used in a number of senses. It has often been noted that these senses do not neces sarily add up to a coherent definition, and that the term is frequently used in vague or ambiguous ways. Con ventionally, if 'territory' refers to a bounded space, 'ter ritoriality' refers to behaviors related to the establishment and defense of territories. Such behaviors include de marcating or partitioning space; classifying spaces as territories; communicating that a territory exists and where the boundaries are; communicating the conditions for entering or leaving the territory; communicating the consequences for not abiding by the conditions; and following through with the implicit or explicit threats.

‘Territoriality’ may also refer to the involvement of territorial ensembles with other social phenomena such as power or identity. In this sense of the term one might say that the international system of states represents the territoriality (or territorialization) of sovereignty; or that South African apartheid gave expression to the territoriality of ruling racial ideologies. Similarly, the territoriality of labor can be seen to involve various components that include border control and the segmentation of tasks on the shop floor. Territoriality in this sense often implicates specific projects and practices such that verb forms such as ‘territorialize’, ‘de territorialize’, or ‘re territorialize’ become useful.

Territoriality can also refer to these projects and practices themselves. In many treatments of the topic it is common to focus attention on those actors who engage in creating territories or enforcing territoriality in order to control others through the control of territories. These sorts of practices are, of course, significant. Many such practices, however, concern not only the establishment of territories but the activities bearing on already existing lines and spaces. At least as significant as the practices of the controllers, territoriality is also performed by those who would be controlled. This can be seen in the simple compliance or conformity with the asserted rules of space. Or it may be seen in the evasion or overt resistance. Squatters, smugglers, intruders, fugitives, and asylum seekers participate in territoriality no less than do owners, customs officials, and prison guards (**Figure 1**). Also, not all expressions of territoriality are centrally concerned with hierarchy or the maintenance of domination and subordination. Other kinds of social relations (e.g., those based on equality, cooperation, solidarity, and mutual respect) may also be given territorial expression. These might be implicated in common property spaces predicated on the right not to be excluded or in the constitution of sanctuaries. Regarded holistically, territories are artifacts that give spatial and material expression to a wide range of social relations which are themselves instantiated through a range of socially significant (or even trivial) practices and projects. What makes these relations and practices ‘territorial’ is their involvement with the dynamic interplay of power, meaning, and bounded space.

Another preliminary issue concerns the naturalness of territoriality. There is a long tradition of conceiving human territoriality as ‘natural’ in some nontrivial sense. This may involve pseudo biological claims about territorial instincts that are attributed to individuals or collectives (tribes and nations). In this tradition human territoriality is continuous with the territorial behavior of other primates, or, indeed, birds and insects. Of course, the distinctive characteristics of each species – language, flight, and olfactory repellent – are recognized as contributing to the distinctive territorializations. The naturalness claim may be supported by the apparent cultural universality of human territoriality. However, even if one were to grant the innateness of territoriality to human beings as such, one would not get very far in understanding its operations. Language is also a human cultural universal but there are (or have been) tens of thousands of languages. Just as one cannot argue for the naturalness of ‘King Lear’ or ‘Smells Like Teen Spirit’ from the innateness of language, one cannot derive the naturalness of the modern territorial state, historical forms of gender segregation, or the existence of supermax prisons from the apparent cultural universality of human territoriality (**Figure 2**). The issue matters, however, because ‘naturalizing’ territory and territoriality may have the effect (intended or otherwise) of naturalizing what is associated with it or what it is giving spatial–material expression to. For example, the state, identity, private property, and various sorts of hierarchies that have been territorialized (such as those associated with racism, colonialism, male supremacy, or homophobia) have all commonly been naturalized. In these contexts ‘naturalizing’ may imply the denial of contingency, and the de politicization (and thereby, the tacit justification) of inherited or proposed territorializations. Naturalizing power also commonly facilitates justifying opposition to proposed changes. Naturalizing territory also facilitates simplifying territory.

Figure 1 Squatters united against global gentrification. Source: www.indymedia.org.uk/en/2006/02/333663.html.

In contrast to the innateness tradition are more ‘constructionist’ understandings that locate the practical operation of territoriality and its profusion of forms and

Figure 2 Supermax prison. Source: www.polilceinteractions.com/homepage.html.

expressions in the contingencies of history, culture, pol itics, ideology, and agency. Constructionism locates ter ritoriality in 'social' as opposed to 'natural' processes. In a sense this framing posits a false dichotomy. In any case, seeing territories or territoriality as 'natural' or as 'so cially constructed' (and therefore, perhaps, as contingent, arbitrary, or illegitimate) are not simply descriptive acts. They may represent political–normative positions. As such, they may be not only views about territory, but interventions in the politics of territory. To that extent they might better be understood as internal to the practice of territoriality itself.

Another related preliminary issue concerns the func tions of territoriality. Territory is commonly discussed in terms of the social or psychological functions it serves. It is seen as a device for accomplishing certain ends. Among these may be control of resources, efficiency or conveni ence, stability or security, and clarity of definition. There is no doubt that territory is functional insofar as it may help to accomplish desired ends. But too narrow a focus on functionality may have the effect of concentrating ex clusive attention on 'the controllers' (those who are em powered to impose their will on others) and the facile identification of 'function' with their desires and inten tions. As already mentioned, this tends to exclude from consideration the 'others' who act in relation to the con trollers. Functional analysis also tends to privilege rational calculation over other modes of behavior that, in a given context, may be as or more significant. Often territoriality is used as a means toward 'irrational' ends such as the expulsion or confinement of feared others. Finally, func tional views may be relatively blind to the unintended effects and consequences of territorialization. In sum, while particular instances of territoriality may accurately be assessed in terms of functions (whether successfully realized or not), functionality cannot be attributed to territoriality as such or to complex territorial configur ations. Indeed, its dysfunctional features may be as or more salient.

Analytics

By definition, territory and territoriality involve the practical interrelationships among 'space', 'power', and 'meaning'. Each of these refer to complex social phe nomena in their own right. In combination they account for much of the complexity of territory and territoriality. In this section we examine territory and territoriality through an analysis of these three components.

Space

'Space' is a basic vocabulary item in English. One might imagine that it does not need to be defined. Space simply is. But some features and distinctions should be noted. Our concern here is with social space – space as a fun damental element of social formations, social relations, and social life. Space often refers to two dimensional extent. It may be conceptualized in terms of distance and proximity; betweenness; distributions of phenomena across space; or motion through it. Not all that is spatial (or 'sociospatial') is territorial, but territoriality neces sarily implicates partitioned social space.

By definition, territories are bounded spaces. Even the simplest territory, then, is a composite of at least three components: (1) the boundary or line that defines the edge of the territory, (2) the enclosed space, and (3) 'the outside' to which 'the inside' is set in relational contrast. Territories need not be entirely closed, though typically they are. 'Public space' can be understood as a kind of territory even though it may be defined by its relative openness. Even so, as a territory, public space may be highly regulated. Insofar as it is, it works like other kinds of territories. What is necessary is that there be a differentiation that establishes an in/out or off/on orientation.

Often, 'like' territories (territories of the same general kind) are organized in horizontal complexes. Political or administrative territories border each other. People and things can pass from one to another, for example, at an international border. Instances of passing (of crossing the lines) may be governed by complex rules of ingress and egress, and the passage itself may be consequential in ways that motion within a given space is not. The same is true with respect to other kinds of spaces: property lots, rooms, or cells, may border each other and this bordering has effects. As this list suggests, different kinds of terri tories are used to accomplish a wide variety of different social objectives. A prisoner's cell, the designated critical habitat of an endangered species, and a gated community are established for different purposes. The social world is also composed of and structured by heterogenous en sembles of territories. These may be arranged in more complex configurations. For example, private property lots abut wildlife refuges and military bases. Movement across these lines is also movement across kinds. These territories are also embedded within jurisdictions, districts, and zones of various kinds. While to some extent a prison cell is a prison cell, it might matter a great deal whether that cell is also located within the territorial spaces called China, Sweden, or Guantanamo Bay Naval Station.

The signs of partitioned space are ubiquitous. The lines and spaces that constitute territoriality are rarely questioned. The spaces themselves can appear to be self evident, quasi natural 'containers', or compartments within which social life takes place. Indeed, this apparent self evidentness is important to how territory works. Contemporary human geographers (and other scholars), however, have developed alternative understandings of social space that render such assumptions problematic. Among the many recent 'retheorizations' of social space that have a direct bearing on territory four are of particular significance: (1) the mutual constitutivity of 'society' and 'space', (2) space as product and process, (3) the production or construction of scale, and (4) the geographical imagination and its relation to identity.

One important strand of recent sociospatial theory has shifted attention away from 'spaces' as such and turned attention toward a concern for 'spatiality' or 'the spatiality of social life'. Among the fundamental premises of this conceptual reframing is the idea that 'space' is not simply a backdrop or context against which social events take place. Rather, the spatial and the social are understood as inextricably bound to each other. Spatiality is a fundamental dimension of 'the social' through which the social becomes as it is and operates as it does. As Edward Soja has put it, "As a social product, spatiality is simultaneously the medium and outcome, presupposition and embodiment of social action and relationship." Thinking about social space this way means that 'space' or any given space or spatial arrangement is not self evident or quasi natural. Therefore, understanding spatiality (and territoriality) requires a deeper (i.e., more sophisticated, noncommonsensical) understanding of the social order of which it is a constitutive element or ingredient. For example, the whys and hows of spatiality (and therefore territoriality) under conditions of modernity, capitalism, or liberalism may differ markedly from their dynamics under other conditions. This assumption then directs attention toward the ways in which spatiality is 'produced', and the historical conditions and social dynamics of its production. It also raises questions about the social and experiential consequences of how it is produced and transformed.

This is a second strand of retheorization that has influenced theorists of territoriality. Territories are increasingly understood less as inert things ('containers'), and more like activities or events. Prominent theorist of territoriality Anssi Paasi (2003: 109–122) describes "…territories as processes in which social space and social action are inseparable." The insistence on the complex 'mutual constitutivity' of the social and the spatial through the dynamics of territoriality has also facilitated closer scrutiny of the various dimensions of social life (e.g., economic, political, cultural, etc.) that contribute to the dynamic spatialities of social life, including the specifically territorial. Territory, then, is less like a grid and more like choreography.

Another development in sociospatial theory with direct relevance for understanding territory has been increased interest in the social – especially political – construction of scale. In contemporary human geography scale commonly refers to the spatial extent or 'level' of operation of some relevant phenomenon. As with other aspects of 'space' scale has conventionally been regarded as a quasi natural 'objective' framework that sets the scope and level of detail of analysis. Typically, 'the global', 'the national', and 'the local' are seen as the most useful scales of analysis or description. For present purposes, 'scale' is used to identify processes and forces that operate on or through territoriality at different 'levels'. Such processes may have different effects or characteristics when examined through different scalar lenses, or when examined through interrelations among scales. Labor migration, for example, looks different (different aspects are highlighted or obscured) when situated at different scales of analysis. However, social actors may have an interest in asserting or denying the relevance or appropriateness of one or another posited scale over contenders. For example, a particular event (say, an instance of violence against women, or the detention of 'unlawful combatants') may be conceptualized as implicating human rights and international law (global scale), or as implicating civil rights and sovereignty (national scale), or as implicating privacy rights or local control (local or corporeal scale). As with other aspects of spatiality, contemporary geographers have come to see 'scale' not as a quasi natural, objective framework for analysis but as an interpretive construct. The '(political) construction of scale' suggests that claims about the appropriate scale of analysis are not necessarily disinterested. Another implication of this is that the sociospatial dynamics associated with territoriality can be seen to unfold not only with respect to two dimensional space – but also as may be commonly represented on maps. Instead, territoriality as process takes place within multiple, perhaps overlapping or interpenetrating spaces. Much of the practical work of territoriality, then, concerns the 'vertical' relationships between or among these overlapping, mutually conditioning 'levels' – again, commonly, but not only, 'the global', 'the national', and 'the local'.

A fourth recent development in geographical theory that pertains directly to territory is the attention paid to 'the geographical imagination'. This notion addresses the ways in which images, concepts, dispositions, or ideologies about social space condition forms and expressions of self and social consciousness, and thereby inform the practices associated with the production of space. One implication of this is that space generally, and territory more specifically, are understood not as external to human thought, imagination, and language. One might say, in fact, that territories are primarily 'imagined'. This is not to say that they are 'subjective' or 'not real', only that they necessarily implicate aspects of social and self consciousness. Of special importance for understanding territory are the connections among material territorial spaces, geographical imaginaries, and aspects of 'identity' (nationality, citizenship, race, and gender). Themes of identity and difference are commonly expressed through the territorialization of elements of self and social consciousness. Consider how images of territory stabilize (or potentially de stabilize) categorical identities such as 'American' and 'Mexican', 'citizen' and 'alien', 'legal' and 'illegal', and 'rights holder' and 'rightless' and how these reciprocally stabilize conceptions of territory. Other examples of territorial imaginaries may come into play in contexts involving racial or gender segregation, privacy, or ownership.

As can be seen, 'space' has come to refer to much more than the apparently self evident, 'objective' features of the external world. It also implicates the dynamic relationship between 'material' spaces (such as states, lots, rooms, and zones) and the ideologies that inform and are reinforced through them. We will return to this theme below. For the present it is sufficient merely to note that there are different ways of understanding the spatial, and that which view among competing views is taken as more accurate or explanatory will have consequences for how territory and territoriality are understood.

Power

By definition, territory and territoriality implicate social relational power. But, as with 'space', this concept is, in its own right, among the most fundamental, if complex and controversial, terms of social analysis. One can take a rather simple, unexamined understanding of power and draw inferences from it, or one can avail oneself of any number of more sophisticated theories and problematize it. As is the case with 'space', different conceptions of power are likely to yield different understandings of territory and territoriality.

By 'power' one often simply means the capacity to act or the capacity to impose one's will against the will of others. Power may be expressed, exercised, and experienced in countless ways. Scholars identify and distinguish, for example, political, economic, military, ideological, familial, and other forms or sources of power. Each of these, by itself or in complex combinations, may be territorialized or may condition the workings of territoriality in a wide variety of ways. Military power may be effected through a territoriality of 'theaters', fronts, war zones and free fire zones, invasions, and occupations, whereas economic power may more commonly implicate trade zones, trade barriers, investment zones, and real property. Also, different social formations ('cultures', historical eras) exhibit differing sources, bases, and expressions of power. The workings of power in, say, tenth century Hopi culture, fifteenth century France and contemporary France all differ significantly from one another. The dynamics of power under conditions of modernity operate differently than under other conditions.

Here simply mention some common limiting assumptions about power. Then I sketch more complex elements as they pertain to contemporary scholarly understandings of territory and territoriality. First, although in ordinary speech we might say that power is something that persons or institutions 'have' and 'use', one important feature of most theoretical conceptions of social power is that it is relational. Second, discussions of power are frequently short hand for 'power over' (A 'has power over' B). This formulation privileges certain forms such as domination and coercion within hierarchical relations. But power can also be manifest as 'power to' in the absence of domination. As mentioned earlier, power might also reflect relations of cooperation, solidarity, protection, and altruism. As reference to protection suggests, 'power over' and 'power to' are not mutually exclusive. Indeed, they may be mutually dependent. Third, discussions of power are frequently cast in bilateral (often zero sum) terms (again, A and B, ruler and ruled). But, while bilateral power relations have an undeniable experiential significance, it is also useful to consider power relations in terms of complex networks or configurations. Fourth, power can usefully be analyzed in more or less structural terms, as intrinsic to any social order. Power as it relates to capitalism, patriarchy, racism, caste, etc., can usefully be thought about in structural terms. This is especially important for understanding (and misunderstanding) territoriality. One highly significant feature of power is the degree to which it is institutionalized. Indeed, many expressions of territory are themselves institutionalized structures of power. Power is often most clearly discerned through examining enduring impersonal role relations that an agent (this or that person) contingently occupies or engages (think of the supervisor, the bishop, the teacher, the border patrol agent; **Figure 3**). At the same time, power can be analyzed in more situational terms. Power is realized, performed, and experienced by and through agents in actual contexts in relation to specific others. Examining power situationally draws attention more precisely to the practical strategies and tactics, accommodations, and negotiations through which it is realized. That is, while

Figure 3 A border patrol agent frisks a would-be crosser. Source: www.desertexposure.com/200707/200707 immigrat.

Figure 4 Territory as sanctuary. Source www.svsu.edu/clubs/gsa/left.htm.

power often inheres (or is 'deposited') in institutions, organizations, and broader social structures, it is realized (exercised, experienced) in concrete situations.

Institutionalized power can be territorialized in various ways. The territorialization of forms of power through the institutions of 'sovereignty' and 'ownership' (via the state and property) may be most obvious. But there are many other significant power projects that also implicate territoriality. Referencing a different sense of 'institution' one might consider how sexism and/or homophobia, for example, are territorialized through notions of 'separate spheres' for women and men, or through the construction and deconstruction of the material spaces of 'the closet'. Often these other dimensions of structural (and situational) power work through the fundamental liberal spaces of governance and property, as, for example, the exclusion of gays and lesbians under some immigration laws or the insulation of forms of private power (domestic violence, workplace discipline) from democratic intervention (**Figure 4**).

Social power can be understood as relatively static (as corresponding to set roles within a hierarchy) or as more dynamic. The operations of power are seen not only in discrete events, they also condition the unfolding of subsequent chains of events. Power is not only discernable in explicit acts. Rather, we – collectively or individually – are never not engaged with, partaking of, participating in the structures and situations of power. Part of what exercising power means is the capacity to exert control over future events. Each of these aspects of power have important implications for territory and territoriality.

That territoriality and power are intertwined is incontestable, but the practical details of such intertwining are complex and subtle. One might begin with analysis of the power to 'territorialize' and inquire about the sources of this power. One might ask: who 'has' the power (the authority, the rights) to draw the lines, to create the borders, to set or change the rules of entry and exit, of inclusion and exclusion? Who has the power to partition, or annex? Who has the power to disregard the desires of others who may be effected: to invade, trespass, or occupy? Who is obligated – or obliged – to acquiesce? (Again, by 'has the power' here we mean, 'who is positioned within networks of power in relation to whom such that these acts may be effectively performed'.) One point of connection between power and territory involves the structures of power that position different social actors with respect to the power to engage in authoritative (or otherwise effective) acts of territorialization. Although all social actors 'do' or 'perform' territory simply by virtue of being alive, some people do it in ways that are more consequential than others.

Territorializing is often the prerogative of state agents. The source of this power is usually ideologically derivable from notions of sovereignty. In the early twentieth century, for example, Great Britain (which is to say, agents such as High Commissioner Sir Percy Cox) participated in the invention of 'Iraq' by partitioning and reassembling portions of the Ottoman Empire. Likewise, just weeks before the British were to leave India, the Viceroy established the Bengal Boundary Commission. Its Chairman, Sir Cyril Radcliffe, partitioned India and created Pakistan, the eastern portion of which became Bangladesh in 1971. Boundary disputes may be resolved by diplomats aided by cartographers. Municipal governments (again, actual persons employed by these institutions) create and revise school districts, voting wards, and policing precincts. Other state agents may be empowered to incorporate municipalities, set city limits, annex territory; to establish wilderness areas, Native

Reserves, free trade zones, or free fire zones. Judges may issue restraining orders that effectively create territories around specific people or create buffer zones around abortion clinics. The police may establish boundaries around protesters at demonstrations or order a particular space cleared of people. But it is not only official state actors who may be empowered to territorialize (or de territorialize). Property owners, employers, parents, gangs, and even children may set rules along a line, effectively 'communicate' these rules, 'enforce' compliance, and punish noncompliance. It is no exaggeration to say that one of the common incidents of 'authority' (of whatever form) is the power to territorialize within one's space of authority. Obviously, this power to territorialize is often deployed with the aim of achieving power over others.

Territory is also commonly used to demarcate the scope and limits of power itself (again, in various forms and expressions). Property and jurisdiction tell the world: my/our power (rights, authority) is presumptively supreme and legitimate and requires deference within these boundaries; it is presumptively illegitimate and imposes no obligations just over the line. The boundaries of territory confine the practice of jurisdiction, and define the edge of sovereignty or regulation. They may mark the limit of legitimacy, membership, participation, and voice. Territory is also important in establishing the spatial scope of rule application. It delineates what is allowable, forbidden, or mandatory where. That is, territory provides communicable limits to spatial fields of action and to the sociospatial applicability of power. Again, this applies not only to political or governing spaces but to households, workplaces, public space, personal space, and other territories.

As will be discussed in detail in the following section, any such performance of territorial authority may be assessed in terms of its 'legitimacy'. Legitimacy entails an evaluation of the power claimed against possible sources or justifications. Occasionally, assessments of the legitimacy of a given territorialization can be negative. Others – particularly the others over whom power is wielded, or those whose actions are to be constrained by territory – may not recognize the legitimacy of the territorializing power and with it the obligations imposed through territoriality. The denial of legitimacy, in turn, may provide justification for the deployment of other forms of counter territorializing power in resistance. Ultimately, though, claims of 'legitimate authority' (to territorialize) may be put aside and replaced by physical coercion and the violent imposition of will. But even this may be insufficient to effect the desired territorialization. For example, in the 1950s, the United States government (again, some actors within that institutional structure) and some Vietnamese collaborators asserted the existence of 'the Republic of Vietnam' (South Vietnam) as a sovereign space separate and separated from what they called North Vietnam at the 17th parallel. Political actors throughout Vietnam repudiated this fiction and, in 1975, Vietnam was unified. This asserted partition proved no more successful than the invention of the Confederate States of America or many other secessionist projects have been.

Territory effects other aspects of power (authority, legitimacy, obligation, sanction) in various ways. Territorial spaces are frequently created in order to amplify, intensify, or consolidate various forms of power. They may simultaneously disempower, subordinate, or injure others. Territory is commonly employed to divide and conquer, to dilute opposition, to confine, exclude or deprive others of access to necessities, to impose burdens, to create relations of dependency or vulnerability, and to dispossess. At the same time it may be a device for creating the conditions for security, self determination, or sanctuary. That is, it is frequently an instrument of power (to and/or over) that further conditions the operations and relations of power.

As with space and spatiality more generally, it is important to look beyond a single space/line or a specific territorializing event and examine more extensive ensembles or configurations. Territory has the effect of channeling the circulation of power through the social world. Territorialized regimes such as apartheid (with its 'group areas', its 'Bantustans', its 'influx control', and its political prisons) or classic American Jim Crow (with its separate schools, seating assignments in busses and waiting rooms, and racialized hospital wards) represented the complex hyper territorialization of racialized power. In practice, the territorialization of racialized power provided innumerable occasions and opportunities to 'enforce', to perform whiteness as dominance, to compel submission, to punish those who refused to submit. These are extreme examples, but analogous arrangements structure more ordinary territorial mosaics (**Figure 5**).

Social power is not static, it is political. 'Politics' here refers to unfolding and shifting contests for and about power. Just as discussions of territory are frequently confined to (by) the territorial state, so 'politics' is often restricted to contests for formal state power. Viewed more expansively though, contests about power take place in many other contexts. The old adage that "war is politics by other means" and the feminist credo that "the personal is the political" open up some of these other possibilities for locating the political in contexts ranging from the international to the interpersonal. Political objectives can be accomplished through violence, voting, negotiation, persuasion, bureaucratic rule making, legal argument, economic transactions, and many other means. Territorial politics, always about space and power, are often focused on the politics of meaning and interpretation.

Figure 5 Signs of racialized territories. Source: www.ferris.edu/jimcrow/traveling2/THEM/34.htm. Ferris State University.

Meaning

Conventional treatments of territory recognize that in order for it to 'work' – to do what is expected or intended – something has to be 'communicated'. At minimum some agent, typically the controller, has to communicate to relevant others that a territory exists or is being claimed; that entry (or exit) is prohibited or conditional upon certain specifications; and that there are consequences to be expected if one disregards these claims. The image that this calls to mind, perhaps, is that of a gate keeper or bar bouncer. No doubt territoriality does seem to work this way on occasion. But whether this is the usual way, and whether thinking about territory and territoriality this way is most useful, are debatable propositions. As with the components 'space' and 'power', communication – or what the author refers to more broadly as 'meaning' – is much more complicated and has undergone rather profound rethinking in contemporary investigations of territoriality. 'Meaning' here refers to distinctively human social symbolic capacities involving consciousness, language, and communication. It implicates thought, discourse, ideology, feelings, and practical significance. This should give some suggestion as to the potential complexity of its relationship to space and power through territoriality. 'Keep Out' means (presupposes, entails) much more than keep out. The increased concern with questions of meaning in human geography reflects, in part, the heightened influence of the humanities, especially Continental philosophy, in this and other social disciplines. Taking questions of meaning, meaning production, and signification more seriously – and more problematically – reveals a greater degree of complexity and subtlety of territoriality than is usually acknowledged. Of specific concern here are meaning's interconnections with 'power' and 'spatiality' that constitute territory. The discussion here proceeds from rather simple to more complex and ambiguous connections.

Perhaps the simplest and most obvious role of 'meaning' in connection with territoriality implicates territories as 'meaningful' spaces or spatial 'containers' of social meanings. They are claimed, defined, distinguished, and differentiated from other spaces. Lines and boundaries are established as defining, differentiating, signifying features of the landscape. Often in modern, literate contexts, territories are named (Bolivia, Manitoba, Ludlow State Prison, Pine Ridge Indian Reservation), or numbered (714 Second Street, Apartment 4 G, Mining Concession Block 1240). They are identified, mapped, and made known. In the modern world there are innumerable generic territories (the state, the refugee camp, the wilderness area) that designate an area as a particular kind of space. Within each category there may be innumerable instances of specific territories (Australia, Jabalia Refugee Camp, Weminuche Wilderness Area). Through the establishment of generic kinds and specific instantiations of territories, 'meaning', at this most basic level, is 'inscribed' or projected onto social space. The resultant territorializations can then meaningfully differentiate locations and events in space. This sort of meaning is, of course, able to be communicated. In the modern hyper literate world this level of meaning is often communicated through both physical structures and linguistic markers. Our world is saturated with signs announcing, directing, prohibiting, allowing, and mandating behaviors with respect to partitioned space. Perhaps the most salient effect of these communicative events is their dichotomization: the inscription of either/or; in/out; on/off; mine/not mine; public/private; and domestic/foreign (and so on) categories onto social space at territorial boundaries.

As we know from daily experience, territorial(ized) meanings are frequently propositional. Propositions signify, for example, what it means to cross the defining line. Sometimes these are implicit. People in any social order are habituated into the territorial regime that constitutes that order. In our world a key component of this is seen in the territorialization of privacy and publicity. But often the propositional aspect of territorialized meaning is explicit. Meanings are encoded on signs that state, for example, the conditions of entry, or the rules of acceptable, unacceptable, or punishable behavior. These territorial meanings can be 'performed' in a number of ways, for example, through the rituals of asking, granting, or withholding permission to enter or leave; by knocking and waiting; by demanding and presenting papers at the border; or by being picked up and bodily thrown back across the line.

As discussion of the propositional content demonstrates, the meanings of territory are frequently encoded in rules. Rules are a special form of proposition; they convey a special sort of meaning. Again, these (or some small subset) may be 'posted' at the point of ingress or around the perimeter. Rules, like territory itself, 'work' best to the extent that they are taken to be simple and unproblematic. However, many rules that appear simple and clear from one perspective can appear to be much more complex or ambiguous when seen from a different, perhaps less deferential, point of view. Moreover, when one pulls on a rule (even the simplest, clearest of rules such as 'Keep Out') one often finds that it is entangled in a potentially endless cluster of other rules, principles, exceptions, restrictions, syllogisms, ideologies, fears, desires, etc. These constitute what might be called the 'deeper' meanings of territory and territoriality.

Questions concerning rules (and ruling, being ruled) connect most directly to questions concerning power. Rules of territory are much more than direct propositions about what it means to cross a line. A given territorial rule implicates other rules about the authority to rule (to inscribe meaning). Not all commands to Keep Out issue from recognized authorities. As we saw in our discussion of power, questions about legitimacy may be raised. Sometimes the appropriate response to Keep Out is 'who says?' (Were you to encounter a sign reading: 'You Are Now Entering The Republic of Bonzonia: Underwear is Prohibited', you might reasonably disregard it. But if the sign said 'Private Property: Trespassers will be Shot on Sight' you might inquire into the existence of the authority to claim or to shoot.) Again, behind the question of authority is the question of legitimacy of that authority. An agent may, in fact, be formally authorized to condition territoriality. For example, during the period of classical American Jim Crow segregation conditioning entry and exit on the basis of 'race' was not only allowed by law (presumptively legitimate authority) but also mandated by State Constitutions. Nonetheless, the legitimacy of that authority was also repudiated by relevant others. Pushing still further, one may challenge the source of the asserted legitimacy of that claim to authority. Significantly, claims to legitimate authority are often referred back to territory itself by way of the (by no means simple) notions of jurisdiction, sovereignty, or ownership: it is simply, self evidently 'the law of the land'. However, a particular territorial (territorializing) rule may and often does raise questions about the obligation to follow it, that is, to conform one's actions to the asserted meanings inscribed through territory. This, in turn, may and often does raise questions about the authority or legitimacy to compel compliance with the rules laid down.

As a particular kind of proposition, a rule is typically prescriptive. Prescriptive rules have a syllogistic (IF>THEN) structure which, again, may raise questions concerning the consequences (sanctions, immunities) associated with not following the rule (>THEN). That is, 'meaning' as encountered in 'rules' implicates potentially dense, open ended questions about power relations. This potential open endedness or indeterminacy contributes further to the complexity of territory. It reveals a dimension of possible (and possibly useful) ambiguity. It opens up the further possibility of multiple interpretations. With the possibility of multiple plausible interpretations comes the possibility of instability and of transformation through re interpretation or re inscription. These possibilities are commonly ignored in conventional treatments of territory. The issue of interpretivity may be compounded in literate social orders where much (by no means all) meaning is textualized, not just in 'signage' (or what is called 'official graffiti') but in treaties, deeds, contracts, constitutions, policies, regulations, doctrines, etc., the interpretation of which is required to stabilize the sometimes slippery meanings of power in social space.

In whatever manner they are communicated or stabilized, rules are just rules: words, propositions, categories, syllogisms. But territorial rules are fundamentally lived. As noted above, territorialized meanings may be internalized, often as common sense dispositions toward social power and space. We learn, and usually conform to, and thereby 'perform' territoriality. Occasionally, though, we disregard, resist, or reject these ruling meanings. This is to say that 'meaning' is not free floating or untethered to lived experience. Neither is it simply communicated unequivocally. Like space and power, meaning is relational. It is meaning for subjects and agents. This may have implications for 'meanings' of a difference sort: meanings as affective significance. Here consider the meanings of 'home' or 'homeland', and associated feelings of belonging and attachment, or fear and aversion. These kinds of meanings may align with more formal propositional or prescriptive meanings (as 'home' may align with property, or 'homeland' with citizenship and sovereignty), but they may also be in tension with these formal propositional expressions. These kinds of 'meanings' and their connections to space and power through territoriality are no less important to understanding how territory works.

Digging still deeper beneath the apparent clarity of territory there are other (in some sense subtler) contributions to the meaningfulness of territories that implicate space and power. We can gather these under the headings of ideology and discourse. For present purposes by 'ideology' is meant a more or less elaborate, internally coherent, shared set of beliefs, dispositions, images about power. Discourses can be thought of as conventional or institutionalized ways of thinking; the categories through which thinking (and doing and being) happens and which determine the limits, in any given context, of what can be thought or of what is acceptable thinking. Ideologies are represented through discourses. Ideologies tend to be justificatory, but (according to which among many theories of ideology one accepts) they need not be. What is important for present purposes is how ideologies can inform and condition territory and territoriality. Like wise, territorializing projects can give social–material expression to ideologies. One might say that territoriality is *per se* ideological insofar as any territory reflects ideological conceptions of space and power. It has proven useful to examine various territorial arrangements through the interpretive lens of ideology.

A given ideology may be said to be hegemonic, or taken as common sense for nearly all participants in a social order – those who rule as well as those who do not. One example of this is ideologies of statism. In our era, claims about globalization notwithstanding, 'the state' and more specifically 'the territorial state' is such a taken for granted feature of the social world that questioning it or identifying the belief in it as 'ideological' is tanta mount to a request not to be taken seriously in most political discussions. In contrast, an ideology may be said to be dominant (but not hegemonic) insofar as a large number of people (or, at least a large number of the most powerful people!) take it to be common sense, even though it may be challenged, or resisted by a significant number of others. Relevant examples of this include ideologies of colonialism or racism that inform the crafting and maintenance of colonial and/or racial(izing) territorial structures. Likewise, territorially significant ideologies may be seen as insurgent or emergent. At one moment in time marginal or heretical, they may, over time, become more plausible or acceptable. Examples here include anti colonial, nationalist ideologies, and ideologies of racial separatism or desegregation. It may be useful to see these contending ideologies as giving expression to rival geographical or territorial imaginaries of how social worlds should be put together or how social space should be organized so that desirable worlds might result and undesirable worlds be avoided. These terri torial and territorializing ideologies are at once a kind of mapping, tools for making meaning, and tools for (re)making worlds.

The meaning of 'meaning' in the context of territory, then, is much richer, more complex, and subtler than 'No Trespassing' or 'Now Entering … ' may suggest. Both the complexity (and the obscurity of this complexity) are important to grasp. The dynamic, intertwining, mutually conditioning relationships of power, space, and meaning (and, therefore, their connections to lived experience) are much more complicated than is commonly acknow ledged. This, again, is one of the central paradoxes of territory and territoriality. The apparent (and func tional?) simplicity and clarity of lines and spaces is belied by the complexity and potential ambiguity and in determinacy. Recognizing this is indispensable to understanding how territory actually operates in the world. Indeed, at least under conditions of modernity one of the most important forms that the politics of terri toriality takes is to be seen in efforts oriented toward (selective) interpretive restructuring. Rival projects often operate primarily – if not exclusively – at the register of meaning: propositions, rules, ideological commitments, and discourses, in order to effect the re signification of space. Through the operations of the territorial imagin ation spatialized meanings may be taken as self evident. Old meaning may be repudiated, new meanings re imagined, asserted, inscribed, and imposed. Spatialized power – understood relationally – is continually re configured both structurally and situationally.

Temporalities

Because territory is regarded primarily as a spatial phe nomenon, and because space (and geography) is com monly contrasted with time (and history) in social thought, territory may too readily be understood as relatively atemporal or static. Insofar as they implicate relatively enduring spatial structures that give order to elements of social flux, territories can easily be under stood as effectively fixing aspects of social order in space. This view is also facilitated by conventional cartographic representations. The preceding discussion suggests something different. Space, spatialities, and spatializa tions are now frequently seen as more dynamic. Territory as process; power as necessarily concerned with the un folding of events in time; the making and remaking of meaning; the increased prominence in theoretical writ ings of process verbs such as 'territorialize' and 'de or re territorializing' all suggest that temporality is an ir reducible dimension (or co participant) in the workings of territory and territoriality. Just as scale allows us to comprehend a multiplicity of interpenetrating spaces, so time can be parsed into time – frames of various inter related durations from the momentary to the deep historic.

As a practical matter, there seems to be an inherent temporal dimension insofar as territory implicates mo bility or immobilization. It is often concerned with (controlling) access, entry, exclusion, expulsion, or con finement. Disjunctures in space (in/out; here/there) often cause disruptions in the smooth flow of experiential time. Vectors of motion often slow down or come to a halt at the edges of territories. We stop and knock, we show our passports, our passes, and tickets. We are moved into or out of spaces. While many territories are relatively en during and are marked by borders, fences, or doors that are fixed to the landscape, others are more fleeting and fluid. They emerge and dissipate. Their contours shift. Others may be described as rhythmic or characterized by periodic openings and closings over the course of a day or year. A special case may be that of spatially defined curfews by reference to which a territory may exist only during darkness or the duration of an emergency. One particularly dramatic example of such temporal terri tories can be seen in the Israeli policy of 'closures' in the Occupied Palestinian Territory (**Figures 6** and **7**).

Figure 6 Boundary of the Pine Ridge Indian Reservation. Source: www.fmep.org/…/checkpoint kalandia.html. Foundation for Middle East Peace.

Just as attention to scale allows one to examine the spatial 'nestedness' or embeddedness of territories, at tention to the range of durations allows one to examine the different temporalities that condition territories and territorial projects. Territorial configurations change over different spans of time. These changes may register pri marily in the domain of meaning, rules, and interpret ation. Here, the spaces remain relatively constant but the meanings and associated power relations change. Con sider, for example, the decades long projects of segre gation and desegregation, or the vicissitudes of immigration laws of different states, or long term chan ges in landlord–tenant relations. Other large scale transformations may entail the spatial reconfigurations of governance. Examples of this include the spatial expan sion of the United States during the nineteenth century, the consolidation of the European Union in the late twentieth and early twenty first centuries, and the dis integration of the Soviet Union. Or consider the local dynamics of subdivision, 'development', and suburban municipal incorporation and metropolitan fragmentation.

Taking into account even larger spans of time, terri toriality as such is commonly said to have a history. If one subscribes to a naturalistic theory of human territoriality, then the 'hows' and 'whys' of it might be understood as temporally contiguous with primate forms and the spe cifics of human territoriality could be linked to theories of evolution. Without this naturalistic assumption it is possible to link transformations of human territorial practices to those large scale 'stage' transformations associated with history as such. Archeological and ethnographic evidence may be drawn on in support of explanatory claims about transformations in human

Figure 7 Checkpoint Kalandia the most notorious checkpoint between Israel and the occupied territories. Most traffic between Jerusalem and Ramallah has to go through here. Source: www.youthbuild.org/site/c.htIRI3PIKoG/b.14090.

territorializing practices. The differences among, say, hunter–gatherer, nomadic, and sedentary social orders (or cultures), and the transformation of a given social order from one type to another may be suggestive of such an historical understanding. This kind of evidence also supports the broader claim that territoriality is an inte gral aspect of every and any social order. Depending upon how a given social order is organized (e.g., in terms of dominant mode of production, practices and insti tutions of governance, gender and social reproduction, technology and material culture, or conceptions of self hood), certain territorial expressions will be possible and more or less serviceable and others will be less likely. These differences are commonly conceptualized as 'stages', and these stages are conceptually organized around ideas of progress, development, or advancement. This understanding of time or history suggests that sedentary cultures engaged in agricultural production are more 'advanced' than either nomadic or hunting–gath ering peoples and that the modes of territorializing are also more advanced. Drawing on our earlier discussion of the importance of the geographical imagination, such thinking facilitates the belief that 'modern' forms and expressions of territory are the most advanced or, as it is sometimes put, 'mature'.

It is with respect to the problematic of modernity that territoriality is most commonly historicized. Modernity, as a frame of reference, historicizes and temporalizes territory and territoriality in at least two important senses. First, it posits a radical disjuncture within the whole of human history. As a frame of reference it pos itions most social orders not only as other than modern but as premodern. To invoke the premodern – or what is still often referred to as the primitive or the ancient – is to gather some thousands of ways of organizing human collective life under one rubric that emphasizes their imagined commonality, ignores the profound differences among them, and positions them 'before' the advent (or local arrival of) modernity. Thus, as different as they may be in other respects, the modes of territorializing asso ciated with inhabitants of the Kalahari Desert, Lapp landers (Suomi) and eleventh century Pueblo peoples such as the Zuni or Hopi are all taken as instances of premodern territoriality. It follows that reconfiguration of these spatial arrangements to conform to the require ments of capitalism and modern statist forms of sover eignty constitutes 'modernization', and maturation or development.

There is a common temporal narrative that describes and explains the local emergence and global diffusion of distinctively modern forms, expressions, and practices of territory and territoriality. It is necessarily the case that 'premodern' social formations (even those that continue to exist) change and that these changes are evinced in territorial reconfigurations: territorial systems of land tenure, settlement patterns, and household organization change over periods of time; more or less formally ter ritorialized state structures emerge, coalesce, and dis integrate; empires rise and fall. But the distinctiveness of modernity implicates temporality in a different sense. Modernity as such is defined in part by its allegedly novel modalities of change as these are understood in relation to 'progress' and 'development'. As a frame of reference modernity is associated with continuous transformation and continuously intensifying and accelerating transfor mations in virtually all aspects of social life. These transformations are reflected in and effected by territory and territoriality. One can see them registered in the emergence and diffusion of the (modern) territorial state, including the colonial and postcolonial state; in 'land policy' and 'indigenous policy'; in patterns of trans con tinental and trans national migrations; in urban form; and in the territorial aspects of rights.

Until late in the twentieth century, 'modernity' and 'modernization' were most commonly taken as self evident (and self evidently positive) facts and processes in Western (that is to say, modern) thought. While there were debates about the most appropriate path to mod ernity, modernization as such was rarely questioned. In the late twentieth century there arose an increased skepticism about modernity and about any of the 'paths' that might lead to it. In intellectual circles some of this criticism took the form of postmodernism understood as a suspension of generally dominant understandings about the West in itself and in 'its' relation to the rest of the world. But in whatever form, this skepticism put into problematic perspective 'modern' and 'modernist' understandings of territory and territoriality. There are three dimensions of this that warrant closer attention. First, under the impetus of the various processes asso ciated with globalization many have come to see the withering away of the modern territorial state. The as serted 'erosion' of sovereignty as a spatialized form of power and the attendant diminution of regulatory cap acity within and beyond territorial boundaries entails (seemingly by definition) the arrival of a de territor ialized world. Some observers celebrate this while others decry it. In either case distinctively modern forms of territoriality are seen as passing into oblivion. However, other observers criticize this de territorial thesis as sim ple minded. The argument here is that 'the state' – or at least some states and their associated capacities – may exist in new relationships with other forms of economic, cultural, and political power, but they (or some subset of developed states) are fully capable of maintaining meaningful sovereign and regulatory power within and beyond territorial boundaries. Moreover, even if the as serted de territorializing processes are in play, they only pertain to this specific expression of territoriality. While borders may no longer be understood as absolute barriers, the processes associated with globalization also entail re territorializations at other scales and in other contexts.

Second, the related emergence and increasing significance of the internet and cyberspace is also seen as a harbinger of the demise of modern territoriality. The disassociation of actions and events from 'real' spaces and places is imagined as entailing the erosion of the territorial conditions of not only sovereignty but privacy as well. While it is incontestable that the cyber revolution has had profound consequences for territorial practices, the extent to which this signals the passing of the distinctively modern is by no means obvious.

Third, 'postmodern' modes of social analysis and criticism more explicitly question the seemingly timeless verities of modernism's self understandings. They are also productively applicable to assessments of globalization and the cyber revolution. This kind of analysis has been especially prevalent in the (Anglo American) spatial disciplines of human geography, international relations and architecture, as well as in anthropological and sociological examinations of territory and territoriality. Some of the fruits of these endeavors have been mentioned throughout this article, for example, in the shift in assumptions from territory understood as clear and simple to increasingly ambiguous and complex; the understanding of territory as 'process' and not fixed container; the novel emphasis on the construction or production of scale; and the invigorated attention to interpretation. These are all, in part, the results of a turn away from modernist assumptions about space. Not only would these newer assumptions have been disputed by earlier scholars, they would have been quite unintelligible. But note: some of these claims seem to be about the world of territory and territoriality while others seem to be more about the concepts or interpretive conventions through which scholars describe and explain that world. These are not identical. But neither are they unrelated. The novel practices and processes associated with 'globalization', the cyber revolution and postmodernist modes of analysis are all instances of and ingredients in novel ways of 'doing' territory and territoriality. They all implicate the intricate mutual involvement of time in the workings of territoriality and the interplay of space, power, and 'meaning' that this entails.

Finally, just as many territorial projects are informed by (sometimes competing) elements of a spatial or geographical imaginary, these may be inseparable from elements of (sometimes competing) historical narratives that project the significance and meaning of time or history onto spaces of power in order to justify the deployment or resistance of power in concrete situations. These may be 'backward looking' as is frequently the case with nationalist discourses and ideologies. But they may also be forward looking, asserting the inevitability of a sociospatial outcome. Examples of this are the American ideology of Manifest Destiny and, some suggest, the contemporary discourse of globalization.

See also: Critical Geopolitics; Geopolitics and Religion; Ideology; Place, Politics of; Political Boundaries; Political Geography; Sovereignty; State; State Theory.

Further Reading

Agnew, J. and Corbridge, S. (1995). *Mastering Space: Hegemony, Territory and International Political Economy*. London: Routledge.
Allen, J. (2003). *Lost Geographies of Power*. Malden, MA: Blackwell.
Brenner, N. (1999). Beyond state centrism? Space, territoriality, and geographical scale in globalization studies. *Theory and Society* 28, 39 78.
Cox, K. (2002). *Political Geography: Territory, State and Society*. Oxford: Blackwell.
Delaney, D. (2005). *Territory: A Short Introduction*. Oxford: Blackwell.
Gregory, D. (1994). *Geographical Imaginations*. Waldon, MA: Blackwell.
Newman, D. (2003). Boundaries. In Agnew, J., Mitchell, K. & Ó Tuathail, G. (eds.) *A Companion to Political Geography*, pp 123 137. Malden, MA: Blackwell.
Paasi, A. (2003). Territory. In Agnew, J., Mitchell, K. & Ó Tuathail, G. (eds.) *A Companion to Political Geography*, pp 109 122. Malden, MA: Blackwell.
Sack, R. (1985). *Human Territoriality: Its Theory and History*. Cambridge: Cambridge University Press.
Sassen, S. (2006). *Territory, Authority, Rights: From Medieval to Global Assembladges*. Princeton, NJ: Princeton University Press.
Shapiro, M. and Alker, H. (eds.) (1996). *Challenging Boundaries: Global Flows, Territorial Identities*. Minneapolis: University of Minnesota Press.
Soja, E. (1985). The spatiality of social life: Towards a transformative retheorization. In Gregory, D. & Urry, J. (eds.) *Social Relations and Spatial Structures*, pp 91 127. New York: St. Martin's Press.
Storey, D. (2001). *Territory: The Claiming of Space*. Harlow: Pearson Education.
Taylor, P. (1995). Beyond containers: Internationality, interstateness, interterritoriality. *Progress in Human Geography* 19, 1 15.
Walker, R. (1993). *Inside/Outside: International Relations as Political Theory*. Cambridge: Cambridge University Press.

Terrorism

D. Mustafa, King's College London, London, UK

Glossary

Critical Geopolitics It is a rejection of the traditional geopolitics whereby politics are deemed to be influenced by physical geography of places. Instead, critical geopolitics critiques conceptions of places and spaces produced by those in power to facilitate practice of politics and statecraft.

Discourse A regulated system of knowledge to which authority is ascribed to the extent that it becomes the arbiter of truth within a certain context. A discourse in this sense does not just produce knowledge but also the reality that it is describing.

Place A geographical space in which social relations and identities are constituted. Place is formed when geographical space intersects with human experience, memory, and social relations.

Political Ecology An intellectual perspective within human geography that has focused on how environmental quality and resource use is influenced and changed by political economic factors.

Post-Structuralism An intellectual perspective that rejects the dualistic conception of reality, for example, between nature and culture, mind and body, etc. Instead the perspective posits that what is relevant is reality that is created by humans through language. Reality, in other words, cannot be separated from the language used to describe it.

Simulacrum A representation or image of something.

Introduction

In the first decade of the twenty first century, few terms have come to dominate the public and policy discourse as terrorism. This popular and policy attention to terrorism – bordering on obsession – is primarily driven by the events of 11 September 2001 when a group of hijackers crashed commercial airliners into the twin towers of the World Trade Center in New York and the Pentagon building in Washington, DC. Although the events of 11 September 2001 were the most dramatic, the term had been gaining a higher profile among the Western policy elites with reference to the violence in the Middle East, in the last three decades of the twentieth century. Despite its increasing importance to policy, the phenomena of terrorism had received relatively scant attention within the discipline of geography until the events of 11 September 2001, and the ongoing 'war on terror' in its aftermath. Geographers' pre and post 9/11 engagement with the issue of terrorism has been along three major themes: clarifying the contested meaning and implications of the labels, 'terror', 'terrorist', and 'terrorism'; unpacking the dominant discourses on terrorism and thereby clarifying the issue of the root causes of terrorism; and lastly through an engagement with the spatiality of terrorism linking it to wider concerns with war, violence, and state oppression. This article discusses the geography of terrorism with reference to these three thematic concerns.

What and Who of Terror, Terrorist, and Terrorism

Terrorism does not have a universally accepted definition. The term entered the Western popular and scholarly lexicon with reference to the excesses of the Jacobin regime in the aftermath of the French Revolution at the end of the eighteenth century – though some have used the term to refer to secret societies of assassins in first century Palestine and eleventh century Persia. In fact, through most of the early nineteenth century the term was mostly used with reference to state violence. In mainland Europe and Russia in the second half of the nineteenth and early twentieth century, the terms terror and terrorism were mostly applied to substate left wing anarchist and revolutionary movements. The concept had a brief realignment with the state during the Fascist/Nazi era in Germany and Italy, but it reverted back to being understood as a substate phenomenon during the second half of the twentieth century.

Despite concern on part of many for lack of an acceptable definition of terrorism, geographers have not been very attentive to the definitional issues surrounding terrorism. This is partially because they have tended to approach it as a discourse rather than an absolute conceptual category to be defined, and secondly because of their concern with a critical engagement with the dominant understanding of the concept and its implications. Even beyond the discipline of geography, the exact definitions of terrorism are contested. Webster's dictionary defines terrorism as "the act of terrorizing; use of force or threats to demoralize, intimidate, and subjugate, esp. such use as a political weapon or policy." Other major organizations such as the US Federal Bureau of Investigation (FBI) and the US State Department define terrorism as follows:

> Terrorism is the unlawful use of force or violence against persons or property to intimidate and coerce government, the civilian population or any segment thereof, in furtherance of political or social objectives. (Federal Bureau of Investigation)

> Premeditated, politically motivated violence perpetrated against noncombatant targets by sub national groups or clandestine agents, usually intended to influence an audience. (US Department of State 2000)

Predictably, the dictionary definition is broad enough to include almost all types of violence because it does induce fear in the victims of violence. The FBI and State Department definitions, because of the nature of the organizations, also predictably emphasize the legal and state centric perspectives respectively. Within the discipline of geography, many have defined terrorism as violence that could be perpetrated by state, nonstate, as well as criminal elements to instill fear. Following is a more specifically geographical definition of terrorism:

> Terrorism is an act of violence, different from other acts of violence, e.g., genocide, war, war crimes, political assassinations etc. in that it is (1) a spectacle directed towards a wider audience than the immediate victims, (2) directed towards place destruction and/or (3) place alienation. (Daanish Mustafa)

Even though these other types of violence may also have a spectacular aspect to them, the difference between these types of violence and terrorism is twofold. First, in say an assassination, the prime motivation is to kill the target, with accompanying possible publicity as a secondary objective. The person in an assassination is killed not because of the message it sends but because of the attributes of the person, and what they were doing or had done. Second, the definition links terrorism to place annihilation/alienation. Without the latter, it is argued that the act will not qualify as terrorism. Besides, there is no hierarchy implied in the label terrorism. Genocide, or war crimes or other acts of violence are at times just as morally reprehensible as terrorism, if not more.

The definition stresses the theatrical aspects of terrorist violence and goes on to draw attention to the spatial focus of the perpetrators of terrorism, both state and nonstate. The definitional exercise on terrorism is concerned with delimiting the scope of terrorism, that is, all violence is not terrorism, and the need to distinguish it from other types of violence, for example, genocide, assassinations, war crimes, etc. for which there are established legal definitions. In talking about place destruction and place alienation, the above definition draws upon earlier work by geographers who draw attention to the traumatic psychological impact of destruction of places from aerial bombing and environmental extremes on war and disaster victims. This attention to the spatiality of the terrorism phenomena has been a prominent theme in the geographical literature on terrorism.

Spatiality of Terror: Globalization, Scale, and State Violence

Geographers have been concerned with the activity, policy, and perceptual spaces of terror which form the context for understanding where the terrorists operate, the government policies that impact the patterns of terrorism, and the perceptions of spaces that make them iconic or mundane places in the minds of the terror victims and the terrorists.

The spatiality of terror and social violence has been a subject of research in many terrorism hot spots, from Northern Ireland to Israel to the global terrorist network of Al Qaeda. In the case of Northern Ireland, terrorists have targeted everyday spaces of pubs, public transportation, dance halls, etc., to engender a climate of fear and terror. The spatial variation in terrorist targets closely matches the spatiality of underlying sociocultural and political differences in the Irish society whereby religious groups were systematically excluded from everyday spaces. Similar theme of targeting of mundane spaces by Palestinian suicide bombers is also evident in case of Israel. Mundane spaces of bus stops, cafes, and restaurants formed the bulk of suicide bombings targets, to the almost complete exclusion of iconic buildings and spaces of Israeli society. Here again the spatiality of Israeli occupation that denies the Palestinians access to everyday spaces of markets, cafes, roads, and means of mobility and in fact engenders those spaces with the terror of being searched, detained, and even shot, seems to match the Palestinian suicide bombers' target selection in Israel proper.

Al Qaeda leadership on the other hand has different geographical concerns and ambitions that lead it to target a different class of spaces from the Palestinian suicide bombers and the Irish Republican Army. Al Qaeda's geographical concerns are hierarchical and based principally on perceptions of sacred spaces, for example, the holy places of Mecca, Medina, and Jerusalem. Consequently, Al Qaeda's targeting bias is also toward what it perceives as icons of Western civilization, for example, the World Trade Center in New York, the Pentagon, US naval vessels, and embassies.

Policies directed toward defending everything and anything from water purification plants to harbors to television stations from terrorist attacks are clearly wasteful, misguided, and likely to be ineffective in the face of evidence regarding the targeting biases of different terrorist groups. Most international terrorists of the Al Qaeda variety seemingly have a very good grasp of the symbolic values of different spaces for Western society and they target them accordingly. Consequently, a values vulnerability nexus as suggested by hazards geographer Ken Mitchell is likely to be more effective in anticipating terrorist violence, where a reflection on values attached to spaces by a society, and the spatial imaginations of the terrorists is matched with the likely vulnerability of those spaces to violence. Much of this research lends credence to the geographical definition of terrorism suggested above whereby the geography of perceived injustice, as well as the terrorists' geographical imaginations bear upon target selection, in addition to the need for a spectacle for furtherance of their cause.

Unlike the government sponsored research, geographical research on terrorism has not limited itself to subnational perpetrators of terrorism. In fact, nation state scale terror and its spatiality has been a prominent research theme within geography. The state centric scale of General Agusto Pinochet's geographical ruminations for example was integral to his highly successful coercive remaking or destruction of places in Chile under his dictatorship. The complex of the Chilean prison system, torture chambers, and penal colonies as well as the urban streets and plazas under curfew and military guard created and maintained very well known geographies of the spaces of terror. All of that terror was justified in the name of the nation state and the supreme national interest.

The geography of state terror is not gender neutral or race blind either, for example, in case of the Guatemalan military's excesses against Mayan women, there were very specific racialized and gendered geographies of violence. In case of Guatemala, as in other cases of state terror, persons became victims of state violence by virtue of where they were and their family and social connections, rather than what they were doing or had done. Another common denominator of state terror is its rather limited spectacle only at the nation state or even local scale. The Israeli violence against the Palestinian population instilling an atmosphere of terror in the Palestinian population is meant to be a spectacle only for the Palestinian population and not for the world at large, while the Palestinian suicide bombings against Israel are meant as a spectacle for a global audience. This theme of only a national level theater of state violence is found in most instances of state terror, for example, Pinochet's Chile, Saddam Hussein's Iraq, Russia's war in Chechnya, and so on.

Nation states are not necessarily the direct perpetrators of terror, their policies coupled with wider geopolitical contexts and foreign policy compulsions can spawn types of social capital and civil society actors which are decidedly antisocial and uncivil by virtue of their extremist violent agendas. The spatiality of Pakistan's role in the Afghan war, for example, in the 1980s contributed heavily to the development of violent religious movements within the country, which were subsequently to go global in their operations. It is also not the case that violent movements once spawned by state policies have the social sphere to themselves in the society. Democratic and inclusive forms of social capital and civil society can also form in opposition to state and nonstate violence with very specific geographies of operation. Institutional analyses of terrorist related as well as democratic and peaceful civil society organizations using a social capital lens is an under explored theme within geography, but a theme that is likely to pay rich dividends in terms of analytical insights.

The spatiality of terror and terrorism is not just limited to the traditional enclosed territorial spaces at the nation state or subnation state scales. There has been considerable unease with the geopolitics of conceptualizing, defending, and targeting terrorism, terror targets, and terrorists, respectively, in terms of countries and places. Network spatiality of terror networks like Al Qaeda and their fluid mutable sense of place and potential targets has been a challenge for governments and counterterrorism experts worldwide. The implications of a move away from the Cartesian space of latitudes/longitudes, toward a network based spatiality and spatial imagination have profound implications for counterterrorism. Instead of a focus on locations and countries harboring terrorists, from a network perspective, the emphasis in counterterrorism will have to be on conditions and social relations that foster terrorism, therefore making intelligence about networks, not just about locations and specific people but a sociospatial inquiry.

A sociospatial inquiry of the type called for above, for example, will perhaps go a long way toward showing that poverty and the poor of the world, far from being the cause of terrorism, are, in fact, its most numerous victims. While the West, quite legitimately, asks for sympathy from the rest of the world when it becomes the victim of terror, it has little sympathy to offer to the millions of the poor of the world who have been victims of state and nonstate terrorism for much longer and much more intensely. The sociospatial inquiry would also further show that international terrorism is, in fact, a deeply middle class phenomenon carried out by professionals who in today's world of strong border controls against the poor crossing into the rich countries have the wherewithal to secure visas to the West. Last but not the least, empathy for the daily violation of the life spaces of poor may just help form alliances with a promise for a lasting peace.

Unpacking the Discourses of Terrorism

In the aftermath of the events of 9/11, there was a sense of urgency within the scholarly community in general, including geographers about how to explain the events, and to produce research that could contribute toward preventing such events from happening in the future. Potential contributions of geographers range from discourse analyses, to critical geopolitics to vulnerability analyses to GIS and remote sensing based techniques. In fact, much of the attraction of geography as a useful discipline in the sphere of counterterrorism has been on account of its strong spatial analysis techniques. Human geographers on the other hand have preferred to adopt a more critical approach. Some of the prominent research on human geographical themes in a critical mode with regard to terrorism include: (1) the importance of geohistorical context and US role as the prime hegemonic power in understanding the role of contemporary terrorism; (2) the spatiality of terrorist networks (3) the negative outcome of the so called 'war on terrorism'; and (4) unpacking the varied discourses of terrorism and counterterrorism.

Critical geopolitics has been the most commonly used analytical lens to unpack the discourses of terrorism and counterterrorism. First, the discursive engagement with terrorism from a critical geopolitical perspective has challenged the geographical specifications of American power by the media and the policy elite, definition of the spatiality of its enemies, and war as an appropriate response to the terrorist acts of 11 September 2001. Second, the critical geopolitics has focused on how the official rhetoric about terrorism has emphasized the 'evil' character of the 'Other' thereby legitimizing the unjust and inhumane state responses in the name of justice and humanity. Finally, the critical geopolitical perspective has drawn attention to the links between terrorism and globalization. Globalization is the main structure, through which geographies of inclusion and exclusion are determined – the prime modernity or hegemonic power of the US can be a cause of considerable resentment on part of those who are excluded or perceive to be threatened by globalization. The need is for an empathetic attitude toward those who are excluded in order to understand their particularities in the face of the universalizing tendencies of globalization.

Analyses of terrorism discourses have not been limited to the critical geopolitics perspective. Historical reflexivity, postcolonial theory, political ecology, and even post structuralist and postmodern perspectives have also been applied to unpack the hegemonic discourses on terrorism. An immediate response to the terrorist attacks of 9/11 and its aftermath was a reflection on the historical parallels to the events. David Alexander drew a comparison between the cataclysmic impact of the 1755 Lisbon earthquake and the 11 September attacks on two very cosmopolitan and commercial hubs of their times. He further drew a parallel between the extreme reaction of the inquisition in response to the earthquake and the sheer futility of that reaction and the present day 'war on terrorism'. In both cases he argued that the events signified the apparent victory of chaos over order, but in the long run the philosophical and cultural reflections instigated by the Lisbon earthquake led to beneficial cultural and material contributions to the European society. The question of whether similar emancipatory developments could be expected to emerge out of the violence of the 'war on terror' remains an open question and something of a faint hope for many.

The postcolonial perspective, in the context of terrorism in the Middle East has shown that Islamic identity is as unstable as a European, US, or Israeli identity and all of them are subject to (at times theologically driven) attempts at consolidating political authority and power. Construction of idealized geographies of the Middle East – on part of the West and the Israeli government, or of the West on part of the Palestinian suicide bombers or Al Qaeda – are attempts at creating a demonized Other. The politics of the demonic other are deeply implicated in perpetuating the cycles of violence characterizing the region's social landscape.

Political ecologists have also contributed to the analyses of terrorism and social violence. Although some political ecological work on violence has been missing the terrorism label, its insights have nevertheless been pertinent. In terms of resource wars over precious metals and stones, mostly in Africa, the political ecologists have argued that the perceived criminal character of rebel movements and corrupt governments' inclusion in the global commodity market responds to an exclusionary form of globalization.

In response to terrorism, the Western democracies have been metaphorically and literally treating it as a disease whose outbreak must be isolated, controlled, and finally eliminated. But political ecological analyses have pointed out that terrorism unlike a biological process is innately human in its causes and consequences. Terrorism and the spaces vulnerable to it are humanly constructed with the litany of emotions, symbolism, ideas, and values bound up with those spaces. Consequently, bottom up human participation and input in building resilience against terrorist acts would be key to terrorism response. Similarly, human centered research on understanding terrorism's cause and potential will provide the best guides for action.

The post structuralist inspired research on terrorism has problematized the surveillance ethos of the 'war on terror' as well as the complicity of the Western media in the same. Electronic borders are becoming a ubiquitous frontier in the war on terror. These electronic borders purport to distinguish between legitimate mobilities, for example, leisure and business from illegitimate mobilities such as terrorism and illegal immigration. The efficacy of the electronic borders is predicated upon electronic identities as a source of prediction and prevention. The

Figure 1 A signboard of an Islamist student organization on the campus of the University of Engineering and Technology, Lahore, Pakistan. Photograph by James L. Wescoat Jr.

electronic borders are particularly disturbing because they hand over the traditionally state role of surveillance to private corporations. But the electronic surveillance and borders are not exactly watertight because they are based upon false assumptions about airtight boundaries between legitimate and illegitimate boundaries and correspondence between electronic and real human identities.

Post structuralist analyses have shown that it is not that the so called free Western media has been appropriated by the state in the war on terror but rather that the 'form' of the media necessarily leads it to distort the nature of reality itself. Electronic media coheres around a simulacrum, where there is no longer the image of the event but rather the event of an image. Human communication is not just verbal or written; it is the totality of body language, environment, facial expressions, feelings, etc. Electronic media while pretending to facilitate communication by its very nature cannot capture the totality of human communication. Consequently what seems like human communication mediated through electronic mediums such as the Internet, television, radio, etc., is actually 'mediatized' communication. In this mediatized communication, the referent events and actors become subordinated to their media representations. Since in a post structuralist sense the self is constituted through a web of communication, the mediatized communication about terror in fact, not just defines the self (as Western, democratic, enlightened, humane) but also the Other (as violent, fanatic, terrorist). The social interaction then in this world becomes really a clash of fundamentalisms (**Figure 1**) in a philosophical sense, that is, two antagonistic beliefs in unmediated static reality violently trying to subdue each other. One may ask why should there not be a war of/on terror in a world like this.

Conclusion

Geography like any other social science has to engage with the issues of the day and engage in research that not only helps elucidate reality but also provides guides for thoughtful action. Geographical engagement with the phenomena of terrorism has been varied conceptually and topically. Beyond the broader categories of problem definition, spatiality, and discourse analyses, geographers have used multiple intellectual perspectives, including critical geopolitics, political ecology, historical parallels, and post structuralism toward the study of terrorism. Topically, geographical research has dealt with obvious issues such as the legitimacy, efficacy, and morality of the war on terror, the spatiality of terrorism, terror, terrorists and terror victims, state terror versus substate actor terrorism, globalization, resource induced wars, elec tronic borders, and the role of the media.

One consistent theme in the geographical research on terrorism has been confusion at the contradiction between the modernist pretensions of the West with its belief in reason, Enlightenment and search for explanation and the complete refusal of the contemporary policy elites to en gage in the type of reasoned explanation for terrorism. The almost teleological construction of terror and ter rorists as evil and motivated by evil has been deemed to be a root cause, not just of the war on terror but also of the ongoing terrorist threat.

Violent movements and terrorists are likely to be around for some time. Victory in the war on terror in terms of physical elimination or apprehension of ter rorists is unlikely, if not impossible. The closest re quirements for reducing and perhaps even eliminating terrorism would be eliminating the perceptual distortions about the other; recognition of the spatiality of injustice – real and perceived – that spawns violent responses; and reflexivity about state policies that create spaces of terror, and empathy for the victims of that terror. Attention to the processes producing spaces of justice and injustice, spatial imaginaries, and cultural fundamentalisms may yet be the greatest contribution of human geography to not just counter terrorism but a just and peaceful world.

See also: Critical Geopolitics; Discourse; Political Ecology.

Further Reading

Agnew, J. A. (2001). Not the wretched of the Earth: Osama bin Laden and the "clash of civilizations". *The Arab World Geographer* 4(2), 85 88.

Alexander, D. (2002). Nature's impartiality, man's inhumanity: Reflections on terrorism and world crisis in a context of historical disaster. *Disasters* 26(1), 1 9.

Amoore, L. (2006). Biometric borders: Governing mobilities in the war on terror. *Political Geography* 25, 336 351.

Coleman, M. (2003). The naming of 'terrorism' and evil 'outlaws': Geopolitical place making after 11 September. *Geopolitics* 8(3), 87 104.

Cutter, S. L., Richardson, D. B. and Wilanks, T. J. (eds.) (2003). *The Geographical Dimensions of Terrorism*. New York, NY: Routledge.

Dalby, S. (2003). Calling 911: Geopolitics, security and America's new war. *Geopolitics* 8(3), 61 86.

Ettlinger, N. and Bosco, F. (2004). Thinking through networks and their spatiality: A critique of the US (public) war on terrorism and its geographic discourse. *Antipode* 36(2), 249 271.

Flint, C. (2003). Political geography II: Terrorism, modernity, governance and governmentality. *Progress in Human Geography* 27(1), 97 106.

Flint, C. (2003). Terrorism and counterterrorism: Geographic research questions and agendas. *Professional Geographer* 55(2), 161 169.

Hanlon, C. N. and Shankar, F. (2000). Gendered spaces of terror and assault: The testimonio of REMHI and the commission for historical clarification in Guatemala. *Gender, Place and Culture* 7(3), 265 286.

Hewitt, K. (2001). Between Pinochet and Kropotkin: State terror, human rights and the geographers. *The Canadian Geographer* 45(3), 338 355.

Hobbs, H. J. (2005). The geographical dimensions of Al Qa'ida rhetoric. *The Geographical Review* 95(3), 301 327.

Jones, S. H. and Clarke, D. B. (2006). Waging terror: The geopolitics of the real. *Political Geography* 25, 298 314.

Kliot, N. and Charney, I. (2006). The geography of suicide terrorism in Israel. *GeoJournal* 66, 353 373.

Le Billon, P. (2001). The political ecology of war: Natural resources and armed conflict. *Political Geography* 20, 561 584.

Mitchell, J. K. (1979). Social violence in Northern Ireland. *Geographical Review* 69(2), 179 200.

Mustafa, D. (2005). (Anti)social capital in the production of an (un)civil society in Pakistan. *Geographical Review* 95(3), 328 347.

Mustafa, D. (2005). The terrible geographicalness of terrorism: Reflections of a hazards geographer. *Antipode* 37(1), 72 92.

Sidaway, J. D. (1994). Geopolitics, geography, and 'terrorism' in the Middle East. *Environment and Planning D: Society and Space* 12, 357 372.

Relevant Websites

http://denver.fbi.gov
Counterterrorism, Denver Division, Federal Bureau of Investigation (FBI).

http:///www.state.gov
Patterns of Global Terrorism, U.S. Department of State.

Text and Textuality

J. M. Smith, Texas A&M University, College Station, TX, USA

Glossary

Deconstruction A hermeneutic that purports to show that nothing stands behind texts – no author, no referents – and that all presence is illusory.

Hermeneutic Principles and presumptions brought to the act of reading or interpretation that complete or settle the meaning of an indeterminate text.

Interpretive Community A social group that shares common interests and hermeneutic practices, and so produces a characteristic reading of a text in the act of reading.

Intertextuality A network of mutual reference between texts, and between texts and the linguistic and metaphysical systems they rely on. May be seen as the sum of all referents, but as simulating reference to entities outside the text.

Scholia Marginal commentary typographically segregated from a text.

Text Formerly a writing specimen deemed authoritative; now frequently any coherent signifying system or configuration of signs.

Textual Indeterminacy The proposition that the meaning conveyed in any text is incomplete, and is only completed, in one of may possible ways, in the act of reading through interaction with the interests and hermeneutic practices of an interpretive community.

Textuality Formerly a principle of literal interpretation; now to be a coherent system of signs or to possess the qualities of a text.

Introduction

Geographers began to make frequent use of the concepts of text and textuality in the 1980s, as part of the widespread hermeneutic, interpretive, or cultural turn then underway in all of the social sciences. In geography, as elsewhere, the social scientists who took this turn were a motley collection of humanists, Marxists, feminists, and postmodernists who shared little beside an aversion to positivist social science and an intuition that a more satisfying human geography could be fashioned after the likeness of the humanities. Geographers who took this turn wished to practice geography as a hermeneutic discipline in the manner of literary critics, not as a positivist science in the manner of physicists or biologists.

The appeal of the literary critic as a role model was hardly the same for every geographer drawn in this direction, but for most it lay in the geographers' perception of the literary critic as a scholar who responded to meaning, changed meaning, produced meaning. Although a scientist may study meaning, explain how it is generated and what social function it fulfills, he should not be himself moved by those meanings or motivated to produce them. As Max Weber famously put it, the scientist is disenchanted. His concern is to explain the numinous power of language and ritual, not to fall under their spell or to take them up as instruments of social manipulation.

For geography to be practiced as a hermeneutic discipline with mental habits and scholarly practices like those of literary criticism, the concept of the text would have to move to the center of geographic thought. A hermeneutic is, after all, an interpretive strategy or way of reading a text, and literary criticism criticizes literature. The concept of the text was, moreover, assured of a central place in the thought of anyone concerned with meaning by the fact that, just before geographers made the hermeneutic turn, a revolution in literary criticism had produced and promulgated a new theory of text that made the word text applicable to many things that were not literary works or verbal artifacts, and that also appeared to have in it power to change the world.

This new theory posited what might be called the textuality of human experience – its pervasive text like quality – and it gave geographers who took the hermeneutic turn an alternative to positivist science that was, nevertheless, an alternative reassuringly similar to science in its universal pretensions, abstraction, seeming rigor, and promise of instrumental power. Before proceeding any further into this new theory of text and textuality, we must, however, review the traditional concept of the text and its relevance for geography.

Text as Authoritative Edition

Historically and in vernacular usage, text is written words, and a text is a discrete unit of written words such as a book, poem, or letter. Textual scholarship is the study of this sort of text as a verbal artifact. It is not only concerned with the meaning encoded in the text, but also with problems such as assigning the probable date of composition, attributing authorship, and identifying and excising corruptions, such as errors unwittingly introduced by typesetters and copyists and emendations intentionally inserted by redactors and commentators. The original purpose of textual scholarship was to produce authoritative editions of important verbal artifacts, editions that textual scholars believed most perfectly embodied the final intention of the person or community that wrote it. This sort of textual schol arship does not attempt to stop controversy over the proper interpretation of the words, only over the words themselves.

Biblical scholarship is the prototype of all Western textual scholarship. Given that the Bible was generally believed to be inspired, authoritative, canonical, and the word of God, it is hardly surprising that serious efforts were made to recover and then preserve the words of scripture in their original form. The character of text and the basic function of textual scholarship can be seen by briefly considering one book of the Bible, the Gospel according to Mark. Textual scholarship suggests that the story told in Mark (and the other gospels) began as an oral tradition in the early Christian community. This oral tradition combined eyewitness accounts of the words and acts of Jesus with commentary added when the com munity recursively interpreted these words and acts in light of the crucifixion and resurrection. So long as this remained an oral tradition, this description and com mentary was an 'open text' subject to emendation and deletion. When it was written down sometime around AD 64, however, the text was closed and the words were more or less fixed for all time.

Commentary did not stop when the words were fixed, however. It moved to the margin of the page, as scholia and footnotes, to the front or back of the book, as preface or endnotes, or into separate texts such as homilies and commentaries. This may be the defining character of a text understood as an authoritative edition: the separation and subsequent typographic segregation of the original, edited, and now closed text (normally placed at the center of the page or volume) from subsequent com mentary on that text (pushed to the margins, ends, or outside the covers). Separation and segregation mark the two basic functions of textual criticism: editing, which judges which available version is the best expression of authorial intention, and annotation, which judges which interpretation of this intention is the best understanding of authorial intention.

Annotators and commentators who perform this sec ond function make the meaning of a text accessible. Their work has three basic aspects. The first is proposing the hermeneutic, interpretive strategy, or way of reading appropriate to the work at hand. In the case of early Biblical commentators, this came down to urging literal or allegorical readings, although in time Biblical her meneutics multiplied to include typologically, anagogi cally, and moral readings. The second aspect is exegesis of cryptic, difficult, or recalcitrant passages intelligible only to a reader skilled in the hermeneutic. The third aspect, which becomes important as the closed text ages, is historical explanation of the increasingly alien cir cumstances of the author and intended audience.

Beginning in the Renaissance, an ever growing num ber of texts were thought to deserve textual treatment. Scholars edited authoritative editions of classical authors and resumed the work of classical annotators. By the nineteenth century, scholars produced authoritative edi tions of, and copious commentaries on, works by living authors. It must be noted that the several functions of textual criticism assume or loose importance when tex tual scholarship shifts its attention from ancient to modern texts. The editorial function does not disappear, as there remain questions of which draft or edition to accept as the best expression of authorial intention, as well as questions about the degree to which the text expresses intentions of editors and collaborators rather than the author, but it ceases to be an exacting philo logical exercise. Annotation and commentary at the same time become largely a matter of proposing the appro priate hermeneutic and exegesis of cryptic passages, for the circumstances and language of the author are, when the author is modern, very often the circumstances and language of his audience.

It is hardly surprising, therefore, that literary criticism in the twentieth century became very largely a matter of proposing alternative readings of cryptic texts. We may suppose that if the critic of modern literature is to be employed, it will have to be at interpretation of cryptic material, for the works of modern authors are seldom badly corrupted or rendered inaccessible by archaism. At the same time, modern authors who wish to attract critical attention will have to write inaccessible works that give critics room to exhibit interpretive virtuosity. This is not the whole story, but it goes some way toward explaining the difficulty of much modern literature, and why twentieth century literary criticism, which geog raphers making the hermeneutic turn took as their model, was almost entirely a hermeneutic enterprise.

Before describing the character of this hermeneutic enterprise, we should note some examples of authori tative editions in geography. Richard Hakluyt, who be came the first academic English geographer when he was appointed to a post at Oxford in 1617, was an editor of what might well be called an authoritative edition. Hakluyt's *Voyages*, released in three volumes between 1598 and 1600, collected accounts written by European mariners. Few geographers since have taken on this editorial task, a notable instance being the collection of ancient, medieval, and early modern geographic texts edited by George Kish in 1978.

Geographers have more often engaged with texts as commentators, although here, again, the tradition is weak because the field acknowledges no great and indispens able work from the hand of a past geographer. Nothing in the geographical literature is thought to possess the enduring value that the other social sciences grant to works by scholars like Adam Smith, Karl Marx, Sigmund Freud, Max Weber. It is telling that what may be the most admired commentary by a geographer, David Harvey's *Limits to Capital*, elucidates a text by the political economist Karl Marx. Because geographic literature is largely modeled on the natural sciences, where books are rare and the half life or articles is short, and because the aim of the academic geographical has so frequently changed, geographers have limited experience with text in the sense of authoritative edition.

Textuality

Until about 1970, textuality was a rather arcane legal term denoting a principle to be followed when inter preting constitutions, laws, and treaties. Following the principle of textuality, such documents were to be in terpreted almost exclusively on the basis of the actual words in which they were written. This principle of textuality was opposed to the principle of contextuality, or the teleological approach, which maintained that ju dicial interpretation of legal documents should attempt to recover and follow the intentions of the parties who wrote the constitution, law, or treaty.

An electronic search of academic papers shows that, beginning in 1970, the word textuality became increas ingly, and in time overwhelmingly, a term of academic jargon concentrated in literary criticism but present throughout the humanities and social sciences. The term climbed to a peak of popularity in 1994, and thereafter fell into a notable but by no means terminal slump. In this literary usage the word textuality has several rival definitions, all of which come down to possessing the properties of a text. The popularity of the term textuality was a direct consequence of a radical change in what academics meant by the word text. Carried out under the heading of post structualism, this change had two parts. The first was that 'text' no longer denoted only verbal artifacts, but now denoted any system of signs possessed of textuality. Thus it became possible to speak of a building, a dream, a social situation, a landscape, or anything else not utterly meaningless as a text. Indeed, because systems of signs referenced one another in a vast 'intertextual' network, which embraced the conceptual apparatus of language and metaphysics, as well as par ticular verbal and nonverbal works, the entire world appeared possessed of textuality, and one could describe world of human experience as a text.

The second part of the change was a presumption of textual indeterminacy. The meaning of a text was no longer thought to have completely resided in the author's intentions, or to completely reside in the words on the page or the signs in the system. The meaning, indeed for some the text itself, was now thought to be produced by a relationship between a system of signs and an inter pretive community. Post structural meaning was not conveyed from author to reader by the written word (or other sign system); it was produced in the act of reading or interpretation that occurred when certain hermeneutic presumptions and practices were brought to bear on a particular sign system.

Semiotic Anthropology and Humanistic Geography

The expanded definition of text as any system of signs had immediate appeal for anthropologists who followed Max Weber in the belief that humans desire meaning above all else, and who began to speak of rituals, ordinary behavior, and even the whole of culture as a text. The most influential and eloquent proponent of this semiotic approach was Clifford Geertz, but he was part of a broad consensus that spread beyond anthropology into fields like geography. This consensus maintained that culture is essentially expressive, so that the behavior of any culture should be understood, not as a functional adaptation to the environment, but as an expression of the conscious ness and subjective experience of persons in that culture. Culture is not a tool, as functionalists had argued, but a collective work of art that externalizes the unique sens ibility of the group, much as a poem externalizes the sensibility of the poet. Perhaps the most famous example of this interpretive approach is Clifford Geertz's essay, *Deep Play: Notes on the Balinese Cockfight* (1972), which ar gues that the popular but, to the casual eye, mundane cockfight is, for the Balinese, "a Balinese reading of Balinese experience." Geertz explains this reading as a subtle meditation on the ambivalent feelings toward rage and the fear of rage that lie at the heart of Balinese sensibility. Externalizing these emotions in the cockfight serves to communicate this sensibility to young Balinese, who must be socialized to the mores of the group, and as a means of catharsis for adults already socialized to the Balinese style of feeling. Thus culture is, in this semiotic account, very largely a matter of edification, instructing youth in the sensibility (or hermeneutic) that is the culture, and permitting adults to reflect on this under standing. Much the same approach was taken in geog raphy by writers like J. B. Jackson, Pierce Lewis, and Wilbur Zelinsky who spoke of "reading the landscape." In this work, best exemplified in the essays collected in *The Interpretation of Ordinary Landscapes*, there is an assumption that the landscape is expressive, as well as functional, and that what it expresses is the unique sensibility of the people who made and inhabit it.

This approach is often criticized for overinterpreting the ordinary. The proffered interpretation may fit the facts, but it is also very often a *post hoc* construction made by the interpreter with little resemblance to any message actually sent or received by members of the group whose behavior or landscape is being interpreted. The Balinese cockfight can be interpreted as something rich and suggestive as a symbolist poem – this much Geertz has shown – but there is no evidence that anyone but Geertz and his students intends or sees these meanings. Geographers who read ordinary landscapes as expressive texts are open to this same criticism. The meaning they discover in the landscape is, in fact, often a meaning largely supplied by their own subtle, educated, and articulated intelligence.

The charge of overinterpretation is not altogether devastating if one is working with post structuralist assumptions, though, for if meaning is produced in the act of reading, then every interpretation is an overinterpretation. To say that a reading overinterprets is simply another way of saying that the text alone is indeterminate, and does not by itself force any single meaning onto every reader.

The Hermeneutics of Suspicion

Critics of literary and other texts have various responses to the problem of textual indeterminacy, authorial intention being the most common solution. For others, indeterminacy reduces a text to something like a blank page onto which any reader may project his or her interests. Reader response theory expresses this in the formula: 'you can only read what you've already read' (how you read anything in the first place being a troubling enigma). More commonly, critics recognize that, while many readings are possible, some readings are better, more powerful than others. The better reading is often said to be the reading that is most comprehensive, that accounts for most of what there is in the text, and perhaps context, to be accounted for. Reading C. S. Lewis's *The Lion, the Witch, and the Wardrobe* as a Christian allegory is, for instance, more comprehensive than reading it as the adventures of three children in a magic land because the first reading finds meaning in actions that are, to the second reading, arbitrary and accidental. Comprehensiveness does not settle disputes between interpretive communities, though, since the interpretation that impresses me as most comprehensive is the interpretation that takes into account those aspects of text and context that my interpretive community tells me ought to be taken into account. A materialist critic would not be satisfied to take Lewis's religious views into account, for instance, and would likely insist that a powerful interpretation comprehend his sexual behavior or class interests.

Beginning in the 1960s, the hermeneutic disciplines were increasingly dominated by an interpretive community that believed any comprehensive interpretation must take in the concealed intentions and hidden agendas that the community presumed to stand behind all texts. Paul Ricoeur (1970) called this reading strategy "the hermeneutics of suspicion," and described it as "a method of interpretation which assumes that the literal or surface level meaning of a text is an effort to conceal the political interests which are served by the text," and that "the purpose of interpretation is to strip off the concealment, unmasking those interests." This hermeneutic suffers the flaw of the liar's paradox, for writing about writing describes itself, but it has been very popular among geographers, who practice it in the forms of deconstruction and the power/knowledge analysis of Michelle Foucalt.

Deconstruction

Summarizing deconstruction is more difficult than doing deconstruction, because paraphrasing Derrida is difficult as lifting a haystack with two hands: the problem being not the weight, but the incoherence. Equivocation and the portentous jargon impress some readers as profundity; others feel only irritation. The central claim of deconstruction is that 'presence' is an illusion conjured by the text. Presence denotes being, or that which exists outside the text as independent reality. Deconstruction maintains, roughly, that everything referred to in every text has the ontological status of a character in a novel, so that words like Sun, I, or truth can no more be said to have a real world referent than the name Oliver Twist can be said to have a real world referent. To think of a world of beings independent of the text is, therefore, deluded in the same way it would be deluded to think that the characters in a soap opera continue their affairs after the program has ended.

This argument rests on what many see as two rather obvious category mistakes. The first is the mistake of reading nonfiction as if it were fiction, a misreading surely as disastrous as reading a novel believing that the characters, places, and events it describes are real. The second is the error of treating concepts, and even things, as if they were words. In the linguistics of Ferdianand de Sassure, from which deconstruction is derived, a sign is said to have two parts: the signifier and the signified. The signifier is the sound with which the word is vocalized, or the mark with which it is written; the signified is the concept or meaning communicated by that word. Now, as we can see by the fact that humans speak many languages, signifiers are arbitrary and conventional. Deconstruction takes this fact, true of signifiers, and extends it to the concepts they signify, and holds that these, too, are arbitrary conventions. One reason to doubt this is that human languages, though many, are translatable.

Deconstruction is a critical technique suited to works of fiction, and appealing to scholars who believe, or wish to believe, that the world itself is fictional, a cultural construction that was made up after the fancy of one imagination, and that might, therefore, be revised to suit the fancy of another imagination. Critical geography rightly supposes that deconstruction underwrites claims for the cultural construction of reality, and wrongly supposes that it supports progressive politics.

Knowledge/power Interpretation

In deconstruction the textualized and fictional world is the mask of nothing; in the hermeneutic propounded by Michelle Foucault it is the mask of power. In the first reading, texts conceal the abyss of a cosmos that is as nonexistent as Oliver Twist. In the second reading, texts conceal a conspiracy that legitimates and naturalizes an interpretation of the world that serves the interests of powerful people. Foucault's analysis rests on the two basic features of the post structuralist text. First, it sees the world as an object of interpretation rather than scientific investigation. It is a text, but an indeterminate text construed by 'discourses' and open to many interpretations, no one intrinsically deserving privilege over any other. Second, it sees the meaning of this text as created in the act of reading by a particular interpretive community, for whom the readings is a tool serving the interest of the community. Dominant or hegemonic groups are able to impose their interpretation on subordinate groups, and so trick them into working against their own interests. The mighty are aided in this ideological swindle by the fact that the discourses that make up their interpretation of the world are embodied in language, law, science, and landscape, and so come to be naturalized and accepted as normal.

This sort of power/knowledge analysis has much in common with deconstruction, which is hardly surprising since both grew out of the same mix of Marxist existentialism and the counterculture of the 1960s. Foucaldian readings have been more popular in geography, no doubt because the hermeneutic is more obviously suited to social analysis, and because it seems, at first, to offer aid to marginalized minorities in their struggle against power. This last impression is almost certainly mistaken, since reducing knowledge to power deprives the weak of truth, historically their best weapon against power.

Conclusion

Geographers who practice geography as a hermeneutic discipline typically work with one of these three views of textuality. Humanistic geographers who take the semiotic approach have generally interpreted landscapes as expressive sign systems; however this approach has fallen out of favor as the view of culture as consensus has been replaced by a view of culture as contested and conflict ridden. Deconstructionists have generally deconstructed the concepts and ontological presumptions of scientific geography, the text they criticize being the conceptual apparatus of language and metaphysics. Foucaldians have generally directed their power/knowledge analysis at particular embodiments of hegemonic knowledge that, they believe, permeate everyday practices and the written word as minutely as 'capillaries'. Exposing the self serving assumptions of colonial literature has been especially popular.

The concepts of text and textuality were undoubtedly central to the transformation of a part of human geography into a post positivist hermeneutic discipline. The existence of today's 'critical geography' is proof that geography can be practiced after the fashion of literary criticism – at least literary criticism of one sort – and not only after the fashion of positivist science. The great limitation of text, textuality, and the critical mode, and their great hazard, is that they tempt geographers to exaggerate the similarities between the sign systems of landscapes or nonfictional geographic writing and the sign systems of novels and poems, to mistake things for words.

See also: Artificial Intelligence and Expert Systems; Creativity; Critical Geography; Cultural Studies and Human Geography; Culture; Deconstruction; Discourse; Humanism/Humanistic Geography; Landscape; Landscape Iconography; Language; Media; Nature, History of; Oral History; Orientalism; Polyvocality; Representation and Re-presentation; Representation, Politics of; Resistance; Semiotics; Text, Textual Analysis; Travel and Travel-Writing.

Further Reading

Barnes, T. J. and Duncan, J. S. (eds.) (1992). *Writing worlds: Discourse, text and metaphor in the representation of landscape*. London: Routledge.

Curry, M. R. (1996). *The work in the world: Geographical practice and the written word*. Minneapolis: University of Minnesota Press.

Duncan, J. S. (1990). *The city as text: The politics of landscape interpretation in the Kandyan kingdom*. Cambridge: Cambridge University Press.

Grant, R. (2003). *Imagining the real: Essays on politics, ideology, and literature*. New York: Palgrave Macmillan.

Hanks, W. F. (2000). Texts and textuality. *Annual Review of Anthropology* 18, 95 127.

Meinig, D. W. (ed.) (1979). *The interpretation of ordinary landscapes: Geographical essays*. New York: Oxford University Press.

Ricoeur, P. (1970). *Freud and philosophy: An essay on interpretation*. New Haven: Yale University Press.

Schneider, M. A. (1987). Culture as text in the work of Clifford Geertz. *Theory and Society* 16, 809 839.

Sosnoski, J. J. (1977). The use of the word "text" in critical discourse. *College English* 39, 121 137.

Steiner, G. (1989). *Real presences*. Chicago: University of Chicago Press.

Text, Textual Analysis

D. Wilson, University of Illinois at Urbana–Champaign, Urbana, IL, USA

Glossary

Intertextuality The shaping of meaning in a text by other texts whose usage may be advertent or inadvertent.

Metaphor A term used, or regarded as being used, to represent another object or process to suggest a resemblance or similarity.

Metonymy A figure of speech in which one word or phrase is substituted for another to represent or depict attributes of an object or process.

Sanitary Coding The representing of strong or potentially objectionable ideas or meanings in text by rendering them more acceptable.

Text The content of spoken words, written language, or built physical form that communicates ideas, concepts, and realities to interpreters.

The concept 'text' has become important in human geography in the last 20 years. Commonly understood as the content of written language, spoken words, or symbolic content embedded in built physical form, it is examined by geographers as a complex form of communication that often seamlessly builds understandings and realities. It is seen to exude standard and conventionalized communicative symbols and rhetorical tropes and power plays (hidden positionalities, imposed presences and silences, metonymies, metaphors, and sanitary codings) that structure and organize the communication process. Even within the most seemingly innocent and neutral symbolic ensembles rely upon this 'troping' to communicate.

What geographers have identified as text is enormous: city growth coalition discourses, mayoral and politician pronouncements, family storytelling, linguistic utterances of planners, community architectural styles, everyday newspaper reportage, informal banter on the streets, federal codifying of housing policy, and many other things. Perhaps, what now passes as text threatens to burst the bounds of a simple unitary category. But in being an intuitively resonant and easily grasped focus, textual analysis now approaches the status of quasi paradigm. It has spun out whole debates around issues such as 'the crisis of representation', 'the politics of ontology', and 'the power of the discursive versus the material' which solidifies the institutionalization of this focus. Heightened sensitivity to the power of text and discourse, to David Harvey, now makes many writers and analysts acutely sensitive to two key things: how the world is represented and the authorial place of textual composers.

In this context, the notion 'text' has ascended to confront the foundational notions that have historically anchored conceptions of knowledge generation. Two central influences have been Francis Lyotard's *The Postmodern Condition* and Jacques Derrida's *Of Grammatology*. Lyotard suggests that science continuously produces and draws on texts of authenticity that corroborates their privileged standing, a text termed 'philosophy'. Lyotard asserts and popularizes a prominent thrust of textual analysis: that scholars should be skeptical of metanarratives and representations in purveying realities. Textual analysis, to Lyotard, should be sensitive to the inseparability of power and representation. A fruitful kind of knowledge, to Lyotard, is one that is paralogical, that is, one that searches for and creates instabilities in dominant worldviews. Derrida provides the conceptual ground for Lyotard, arguing that seemingly solid and objective text acquires this only via a mix of supplements, plays, and chains of signifiers. Derrida's view of the history of dominant ideas is a history of metaphors and metonymies.

Yet, there are multiple interpretations of how text made and used is powerful. This article briefly discusses two prominent takes which are extremely influential today: neo Marxian and neo Foucaultian. Since *c.* 1990, 'the representational turn' has greatly influenced neo Marxist analysis. It has posited intimate connections between text and the deploying of class power in diverse and variegated capitalist settings. Here, text helps construct understandings of the world that builds social realities (e.g., social relations, class categories, gender fractions, and racial groupings), state formations (e.g., economic programs, fiscal policies, and government redistributive schemes), and material realities (e.g., landscapes, cities, neighborhoods, and nations). For example, text in the form of technical documents and planner pronouncements helps sanitize and legitimate efforts of city and regional planners to balkanize space and segregate classes of people that assist capital accumulation projects. At the core of this, offers of emotion invoking 'realities' frequently center for common consumption two themes: the truths of dangerous and encroaching streets and dark and foreboding inner cities which need to be isolated or annihilated, and the need to concentrate upgrading in the service of cities which need to 'go global'.

Catherine Kingfisher's recent work personifies this focus. She unearths the crucial role of text in producing and maintaining understandings of homeless people as undeserving citizens in Canada. Kingfisher's starting point is the power and imperatives of neoliberal governance. To regulate people, spaces, and economies in distinctive economic times, texts offer elaborate understandings of the homeless as 'a people' devoid of what localities now need: rugged, staunchly individual, entrepreneurial beings. Kingfisher weaves a persuasive set of connections between capitalist realities and needs, forms of political governance, strategies of making and regulating people, and texts that do the bidding of political governances and capital. Here, social constructionism, textual formation and maintenance, patterns of human management, and capitalist realities deftly interconnect within one conceptual frame.

Text here is a profoundly relational element. Text always needs to be understood against the realities of societal organization, prevailing class dynamics, and the imperatives of capital accumulation. Offers of truth are hardly a simple effect of the rules of a text; they are rather outgrowths of this formation's relational dynamics. Understanding text becomes displaced to a key domain: how texts are constituted and used within the totalizing reality of current political–economic forces. Texts, from this position, are typically cast less as the essential ground that defines social life than as tools whose usage provides the crucial injection of 'realism' into a turbulent, class stained world. Text is never neutral or passive in shaping social and spatial outcomes: capital put it in the service of classes to pursue their materialist desires and aspirations.

Paralleling this neo Marxist development has been neo Foucaultian accounts of the power and prowess of text. Foucault's notion of text ('discourse'), methodologically laid out in *Archeology of Knowledge*, offers a critique of how modern societies manage and discipline populations in sanctioning the claims and practices of the human sciences. Texts here are systems of possibility for knowledge, that is, fields of knowledge formulated via sets of discursive rules. Through these subjectively constituted fields, 'regimes of truth' are always specific to them. Systems of possibility for suggesting and labeling truth and falsity – carried in the text's behind the back rules for producing 'logical' assertions – make these entities potent sources of power. To Foucault, texts are lynchpins of power rather than mediums through which power is asserted. He presents these as complex, organic units whose usage by anyone puts power in the service of their renditions and statements. Power and prowess, in the final analysis, embeds in the potentiality of text to manage, control, and discipline people.

But geographer's use of this framework has often been oblique and tentative. Usage has emphasized two key tenets: texts as power laden entities unto themselves and texts as fields of political assertions reflecting the desires of individuals to manage and control. Power, then, is seen to lie in the local, in the individual, and in the flow of periodic and unstable moments. Texts ultimately orchestrate worlds seemingly unconnected to broader scale political and social dynamics. They rise, live, and fall in the micro scale spaces of the everyday: in workplaces, school rooms, hospitals, welfare offices, public parks, private board rooms, planning meetings, government offices, street encounters, and youth centers.

In the process, some important Foucaultian aspects of text have been neglected. First, discursive rules that structure and enable the formation of statements have been all but ignored. We thus know little about the rules and principles that permit certain statements to be made (the behind the backness of text) and how they come to be and evolve. Second, the complex and multitextured mediation of texts against which they are constructed and meanings extracted have been greatly simplified. To Foucault, all texts are mediated in many ways: by language, cultural systems, conventions of writing genre, intertextuality (the infiltration of meanings from one text into another), and acts of deciphering meanings. This point, as much as any in Foucault's arsenal, places the status of truth in the realm of subjective and relative.

Yet these investigations in geography have proved fruitful, particularly with their sensitivity to the spatialities of the human condition, that is, the production of space, the constitution of places, and the construction of scale. These studies have examined how texts codify and regularize patterns of thought and practice around city and regional planning strategies, provision of public housing and social welfare, state actions, and the choices of people and populations to choose residential locations and imagine patterns of acceptable residential structuring. Texts here are akin to powerful prisons without walls; they lock people into ways of seeing and understanding as systems of imposed possibility that are fluid but grounded sources of power in their own right. There is much subtlety to this work. Thus, it widely distinguishes between the dual spheres of grammar and rhetoric as operative forces. Grammar is identified as the syntagmatic protocol that serves as the vessel for communicating, and rhetoric as the intertextual system of signifiers that fill the grammar with content. Moreover, texts are seen as hybrid constructs that draw on other texts to imbue it with meaning. With this point, texts become patched together mosaics whose origins are multiple and variegated.

What, then, unifies these different frames? Most fundamentally, text is seen as a creative human offering that enables power, structures, and interests to be reproduced at diverse scales in diverse settings. While the issues of sources of this power and who wields it are contested by these frames, texts nevertheless are identified as central sites for the naturalizing of power and political striving. Texts may be turfs for the genesis and application of power, or simply fragments of context that are capillaries in a circuitry of power wielding, but both acknowledge their crucial role in purveying authority and regulation. As Loretta Lees notes, the power for text may lie in formations of enabling rules rather than the text, or in passions and impulses displaced into these formations, but texts are always recognized as locations which render a world transparent, ordered, and knowable. Set against this, texts become potentially powerful coercive terrains that are at the interface of poli tical drives and necessary to influence human beings.

See also: Discourse; Foucauldianism; Text and Textuality.

Further Reading

Derrida, J. (1974). *Of Grammatology*. Baltimore, MD: Johns Hopkins University.

Harvey, D. (1996). *Justice, Nature, and the Geography of Difference*. Oxford: Blackwell.

Hubbard, P. (2000). Desire/disgust: Moral geographies of heteroesexuality. *Progress in Human Geography* 24, 191 217.

Kingfisher, C. (2007). Spatializing neoliberalism: Articulations, recapitulations, and (a very few) alternatives. In England, K. & Ward, K. (eds.) *Neoliberalization*, pp 195 223. Oxford: Blackwell.

Lees, L. (2002). Rematerializing geography: The "new" urban geography. *Progress in Human Geography* 26, 101 112.

Lyotard, J. F. (1984). *The Postmodern Condition*. Minneapolis, MN: University of Minnesota.

MacLeod, G. (2002). From urban entrepreneurialism to a revanchist city. *Antipode* 34, 602 624.

McCann, E. (1999). Race, protest, and public space: Contextualizing Lefebvre in the US city. *Antipode* 31, 163 184.

Mumby, D. (1991). *Narrative and Social Control*. Newbury Park, CA: Sage.

Smith, J. (1996). Geographical rhetoric: Modes and tropes of appeal. *Annals of the Association of American Geographers* 86, 1 20.

Theocracy

Nick Megoran, University of Newcastle upon Tyne, Newcastle upon Tyne, UK

Glossary

Clergy The class of people constituting the formal leadership and functionaries of a particular religion.

Iranian Revolution The 1979 overthrow of the Iranian monarchy that eventually led to the establishment of an Islamic theocracy under Ayatollah Khomeini. Khomeini died in 1989, but the status of Iran as an Islamic Republic persists to this day.

Islamism The political ideology that modern states should be reconstituted constitutionally, economically, and judicially in accordance with what is posited as a return to authentic Islamic practice.

Laity Followers of a particular religion outside of the formal clergy.

Majority Catholic/Muslim States Countries where the majority of the population formally identify themselves with these particular religions, whether or not that influences constitutional government.

Permanent Observer Mission (at the United Nations) An entity that is not a member state of the United Nations, but which has received a standing invitation to participate as observer in the sessions and the work of the General Assembly and maintain permanent observer mission at the Headquarters. These include entities such as the Holy See, the Palestinian Authority, the African Union, and the Community of Portuguese-Speaking Countries.

Religion A set of beliefs about the divinity, practices emanating from these beliefs, and institutions that represent them.

Secularization The name given to a process whereby a previously highly religious society becomes increasingly less religious.

Taliban Political movement emerging from Islamic schools in Pakistan that ruled most of Afghanistan from 1996 to 2001, enforcing a strict form of Islamic Shari'a law.

Vatican City The world's smallest state, governed by the Pope as Bishop of the See of Rome.

Voluntarism Voluntary action for the common good.

Introduction

Theocracy is the exercise of political power by the clergy of a particular religion, usually (although not necessarily) claiming to be acting primarily on behalf of a divinity and governing according to its principles and requirements. Although the worldwide influence of religious political ideologies has grown over the past three decades, there are very few theocratic states, or theocracies, in the world. Nonetheless, those that do exist can have profound and highly idiosyncratic impacts on development.

Three Definitions of Theocracy

Theocracy at first appears to be a simple concept – clerics ruling a state instead of, say, the general population via a professional class of politicians (democracy) or hereditary potentates advised by appointed specialists (monarchy). Based on this definition, there are hardly any theocracies in the world today. Even states that enshrine religious legitimization or codes in their legal existence or practice of statehood fall short of this standard. The monarch of the United Kingdom is crowned by the country's senior cleric (the Archbishop of Canterbury), and some other clerics have seats in the upper chamber of Parliament (the House of Lords), but the clergy's actual political power is negligible. Saudi Arabia enforces a strict form of adherence to a variant of Islamic Shari'a law, even to the extent of forbidding non Islamic worship, yet is ruled by an extended royal family, not a class of clerics. Likewise, the Je Khenpo, the highest Buddhist religious official of Bhutan, is traditionally considered to be the king's most important adviser, but the state is formally a monarchy. Such a strict definition leaves few states in the world as theocracies.

However, this traditional definition is more problematic than it first appears due to complications regarding the meaning of the word religion. Although the term religion is an old one, its contemporary usage is heavily influenced by medieval Christian conceptions of separate spheres of power exercised by the church and the civil authorities, and enlightenment classifications of epistemic communities. The word religion is therefore an attempt to demarcate a specific sphere of belief, practice, symbols, and institutional organization, separate from other such religious systems or from what therefore becomes defined as the secular – the realm of law and government of a specific jurisdiction.

If the disciplining of knowledge implied by the traditional definition of religion has been called into question, then religion (and the sacred) can be seen as a more diffuse set of practices, beliefs, and symbols, and social formations that embrace a wider range of phenomena. It also permeates supposedly secular civil society in fluid ways, rather than being hermetically sealed from it. Such an understanding has recently provided the basis for a reworking of the idea of theocracy. Rather than being simply formal clerical rule, it can be considered as the multiple patterns of the intertwining of religion in the language, practices, and substance of the politics of modern statehood. From the 1960s and 1970s, for a var iety of complex reasons, the increasing importance of religion in national politics can be seen in polities as diverse as the USA, Israel, Lebanon, and the Palestinian territories – a phenomenon that Giles Kepel has mem orably termed, 'The revenge of God'.

A number of authors have thus seen fit to describe the USA under the presidency of George W. Bush as a theocracy. In the longer term, this reflects the adaptation of Republican Party policies to appeal to an increasingly important conservative Christian support base. This process advanced under George W. Bush as the most personally deeply religious president for many decades. Under the Bush Presidency, the influence of Christianity in the White House grew markedly, from prayer meet ings to formal and informal links with notable Christian leaders. George W. Bush's personal religiosity arguably influenced both domestic and foreign policy. Some scholars and commentators have seen fit to label these processes as instances of theocracy.

A certain school of political thought has des cribed some Western European and North American states as secular theocracies. According to this argument, these societies are therapeutic states that attempt to re design social conscience by legislating morality. This is performed by moral crusades to re educate populations away from what become constructed as social pathologies such as sexism, racism, homophobia, and the questioning of multiculturalism. This is a feature of contemporary Western governments that can be understood as part of the historical trajectory of secularizing Protestantism. Just as protestant cultures dwelt upon the individual finding divine grace by confessing one's sins and cleansing one's soul, the therapeutic state works by allowing the majority to dwell on its collective sins of the past yet feel per sonally righteous by expiating them in enforcing toler ance. This governmental practice arose within the cultural milieu of secularizing Protestantism and is thus deemed, oxymoronically, secular theology.

However, there are analytical disadvantages to re defining theocracy to mean any trace of the influence of religious discourse or practice in the field of formal state politics. It would be difficult to identify a society where the influence of religion or the sacred, widely defined, would be entirely absent from political culture. Fur thermore, genuine theocracies in the classical sense of the term where clerics have captured state power pro duce modalities of governance startlingly different to those of absolute monarchies and democracies merely influenced by religious ideas, and map different geog raphies of development. The failure of many European and American thinkers to grasp the significance of the Iranian revolutionary theocratic movement headed by Ayatollah Khomenei, which took power in 1979, was arguably due to their inability to comprehend the im portance of the religious context. Paradoxically, re defining theocracy could reduce our sensitivity to certain intersections of politics and religion, not increase them.

Theocracy and Geography

The geography of religion has been a recurring, if somewhat marginal, topic of cultural geographical an alysis over many decades. Traditionally, geographers mapped phenomena such as the spatial incidence of re ligions measured by demography, and temporary migra tory patterns of pilgrimage. More recent analysis has afforded greater sensitivity to the multiple intersections of religion/the sacred in constructions of space and place, and to geopolitical imaginings of the world.

Theocracy, however, has been of marginal interest to geographers. The founder of modern political geography, Friedrich Ratzel (1969, 1896: 17–28), wrote that, "All the ancient states and all states on lower cultural levels are theocracies," dismissing analysis of them as an unworthy task for the modern discipline. However, presently existing theocracies (or movements to create them) present fascinating examples of the reconfiguration of the political geographical relationships between territory, sovereignty, and (religious) identity. They are increas ingly hard to overlook in debates about the re making of the world map and the emerging shape of contemporary patterns and processes of territory and sovereignty. This article considers two examples in relation to develop ment: the Holy See, and movements for the restoration of a Caliphate/the construction of Islamic states.

Modern Christian Theocracies

Christianity emerged in the context of Roman imperial oppression of first entury Palestine. Unlike contemporary violent Jewish nationalist movements, its establisher, Jesus Christ, did not seek to capture state power and create a theocratic state. His immediate followers adop ted the same position, which has been described by twentieth century theologians as pacifism or non violence. However, by the fourth century, this position was reversed as Christianity came to play an increasingly important role in the Roman Empire. A contemporary inheritor of this legacy is the Holy See.

Figure 1 The Papal States in 1861. Copyright Nick Megoran.

Roman Catholicism and the Holy See

The largest branch of Christianity, the Roman Catholic Church, meaning in effect all churches in communion with and under the spiritual authority of the Bishop of Rome, is unique among world religions by virtue of its relationship to the territorial expression of the Holy See in the form of the State of Vatican City. Crucially for this article, the association of these entities with various Catholic relief and development agencies, and the alli ance of these agencies and the Holy See in international lobbying, creates a network able to have profound and unique impacts on the development process.

The Holy See is a nonterritorial entity composed of the Roman Catholic College of Cardinals and the Bishop of Rome (the Pope). The two bodies are not to be confused – the Roman Catholic Church is composed of its members around the world, while the Holy See is the supreme organ of government of the Roman Cath olic Church. The Holy See has acquired a juridical personality which permits it to sign treaties and receive and send diplomatic representation with states – in ef fect, it has achieved the juridical equivalence of statehood.

Although the Holy See has engaged in diplomatic exchanges with rulers and states at least as far back as the late fifteenth century, part of the reason that it has been able to accrue this unique power is its territorial ex pression in the State of the Vatican City (Vatican City). This is a legacy of the Pope's territorial possessions in Central and North Italy, the Papal States (see **Figure 1**), held until the creation of the modern united kingdom of Italy in 1861.

Vatican City, an enclaved state in the city of Rome, is the world's smallest microstate with a mere 0.44 km^2 territory and a population under 1000, entirely male. It was created in 1929 by the Lateran Treaty between the Holy See and the Kingdom of Italy. The Treaty also recognized extraterritorial rights of noncontiguous properties deemed necessary for the functioning of the Holy See, such as the Pope's summer residence of the Palace of Castel Gandolfo and administrative buildings in other parts of Italy. The Pope, who is elected by the College of Cardinals, governs as 'Sovereign of the State of Vatican City', with full legislative, judicial, and executive powers, through a commission of cardinals nominated by him. Vatican City's system of government, which is highly anomalous, may thus be regarded as both an elective monarchy and an elective theocracy (although its representatives would be unlikely to accept that label). The Pope is head cleric in the Roman Catholic Church as incumbent of the Holy See and sovereign of Vatican City. This dual role of head of the world's largest religion and sovereign of the world's smallest state can be con fusing (especially as, colloquially, the central authority of the Roman Catholic Church is shortened to 'The Vati can'). Nonetheless, understanding it is crucial to grasp the remarkable juridical position that the Holy See oc cupies as a state. Indeed, since 2002, when Switzerland joined the organization, it has been technically the only nonmember state to maintain a Permanent Observer Mission at the United Nations Headquarters in New York, with all the privileges and opportunities that this entails.

The nexus of various state, quasi state, religious, and NGO structures of Roman Catholicism/the Holy See/ the State of the Vatican City and nonstate development agencies such as the Catholic Agency for Overseas De velopment (CAFOD) has created a unique political geographical network. As the next two examples show, this can result in range of different forms of engagement with development.

The Holy See at the Cairo and Beijing conferences

The Holy See's engagement at two United Nations de velopment conferences, the 1994 International Confer ence on Population and Development in Cairo and the 1995 Fourth World Conference on Women in Beijing, exemplify the ambiguities of this theocracy cum religion cum nonstate actor. The Cairo conference was intended to stabilize population growth. The Holy See, reflecting the Catholic Church's long standing opposition to both contraception and abortion, lobbied strenuously against language in the draft and final documents that legitim ized forms of both practices. Before the conference, it did this in official pre conference meetings and through bi lateral diplomatic negotiations with individual states. During the conference, it fought over wording at every stage, allegedly attempting to form an alliance of majority Catholic and Muslim states (although the Holy See itself denied this). The final document represented a com promise between the militant pro abortion position ad vocated by the USA under President Clinton, and the militant anti abortion policy of the Catholic Church under Pope John Paul II. Pro abortion delegates not merely criticized the Holy See's negotiating position (criticism which was unsurprising), but also its presence at the conference as a state rather than as a religious NGO. It was questioned whether a microstate with zero natural growth rate, indeed with no permanent population, women or children, had a legitimate voice at a conference of states on population growth.

Likewise, at the Beijing Conference the Holy See's position was controversial. It allegedly lobbied to exclude certain religious NGOs, such as 'Catholics for a Free Choice', from attending – an example of different elem ents of Roman Catholicism at odds over a contentious issue. Along with certain majority Muslim states, it also challenged references in the final document to abortion and sexual orientation. Objections to its policy position are part of the normal process of debate, but the conference again raised questions about the Holy See's participation as a state. Essentially, the Holy See derives its authority from its position as the world's largest religious organ ization, yet it uses its association with a tiny enclaved state to act as a state, obtaining privileges in negotiation that no other religious body possesses. Commentators such as Yasmin Abdullah have questioned whether this unusual arrangement ought to be allowed to continue.

Liberation theology and development in Latin America

A very different form of impact on development can be seen in the Catholic Church's 'Basic Christian Com munities' in demographic strongholds, Latin America. The 1960s and 1970s saw the growth of liberation theology in the continent, a Marxist informed structural critique of poverty and political repression in Latin America combined with the praxis of a traditional Christian commitment to assist the poor. At a time when much of the continent was ruled by repressive dictator ships, grassroots resistance and pressure for change was often conducted through the movement for Basic Christian Communities (*Communidade Eclesial de Base*). These were small grassroots groups of Catholics who met for Bible study and to reflect and act on pressing social problems – illegal arrests and land expulsions, lack of basic services, etc. Activism – hiring lawyers to defend rights, pressing for healthcare facilities, building schools, founding trade unions, forming the basic constituencies for democratic opposition politicians – arose from these bodies, which numbered in their tens of thousands across the continent. Commentators such as Thomas Bruneau thus argued that these bodies had a profound potential for grassroots led development in the continent at a time when formal church structures were relatively weak compared to the violence of oppressive regimes.

Basic Christian Communities, although grassroots organizations, were often initiated and encouraged by the Roman Catholic priests and nuns disseminating politic ally radical theological ideas and practices under the gaze of state apparatuses. Although far from the formal structures of the Holy See and the theocracy of the State of the Vatican City, this grassroots development initiative was, ultimately, part of the same network. The Roman Catholic Church can thus be conceived of as a network of diverse entities, incorporating often conflicting nonstate actors, theological structures, grassroots communities, and a theocratic state. This multiplicity opens up nu merous spaces for multiple forms of intervention in processes of development.

Modern Islamic Theocracies

The unusual case of the Holy See notwithstanding, there exist no formal Christian theocracies today, nor import ant or widespread movements advocating their creation. The same is not true in the majority Muslim world, where the relationship between religion and state has been one of the major topics of debate among Muslim intellectuals for the past two centuries. Unlike Chris tianity, the establisher of Islam, the Prophet Muhammad, sought to lay the foundations of a polity based on re ligious teaching to displace other forms of political power, a polity that was to be defended, where necessary, by military force. Although this historical precedent is very important, the recent intellectual situation can be explained to a significant degree by the struggles of Muslim peoples against Western colonialism, and dis appointment with forms of self rule that replaced formal foreign sovereignty with often corrupt, authoritarian, and incompetent patrimonies supported by ongoing forms of Western political and military intervention. The failure of secular political paradigms such as nationalism and Marxism to deliver tangible benefits in majority Muslim countries has intensified this process since the 1970s, particularly in promoting Islamism.

Hizb ut-Tahrir: Resurrecting the Caliphate

The Caliph was the title given to the leader of the Islamic community, although since the death of the Prophet Muhammad in the year 632 there was no unanimous agreement at any one time about who that should be. A key moment in debates about modern Islamic theocracy was the abolition of the Caliphate in Istanbul by Mustafa Kemal secularist republican Turkish government in 1924.

This move prompted many differing reactions. Some Muslims sought to adopt European and American models of secular governance, while others devised new concepts of the Islamic state. Some argued for the restoration of the Caliphate as being the only theologically legitimate political goal for Muslims.

An active group that advocates this latter position today is Hizb ut Tahrir (Party of Liberation). Founded by the Palestinian jurist Muhammad Taqiuddin al Nabhani in 1953, it seeks to re establish the Caliphate as a single state sovereign over all lands where Muslims form a majority of the population. It considers the modern map of the Middle East to be the work of Western imperialist powers intent on dividing the Islamic community by supporting corrupt and un Islamic rulers, and considers it the duty of Muslims to work, using political means, for their abolition. It is thus repressed in most majority Muslim states where it represents a threat to the in cumbent regimes. It is also banned in some countries without majority Muslim populations.

Development is an important concern of Hizb ut Tahrir. A key appeal of the Caliphate to its modern ad vocates is that it would redistribute wealth from the rich to the poor, in sharp contrast to rulers who have often lived lives of indulgence and opulence while ordinary Muslims have suffered in poverty. Opponents of this group argue that Muslims have never agreed on a single legitimate Caliph and would be unable to do so now, but the appeal of this form of theocracy persists for many Muslims.

It is worth observing that Hizb ut Tahrir itself rejects the label that it is advocating a theocracy. It argues that the Caliph is a representative not of God but of the people, as he would be elected by them, and that ultimate sovereignty belongs not to the clergy but to God.

Islamism in Egypt and the Horn of Africa

Another major reaction to the abolition of the Caliphate was the emergence of Islamist ideas in Egypt associated with the Muslim Brotherhood. Islamism is a political ideology which posits that modern states should be re constituted constitutionally, economically, and judicially in accordance with what is held as a return to authentic Islamic practice. Founded in 1928, the Muslim Brother hood sought to establish a modern Islamic theocratic state by political means. It wields considerable influence in Egypt although it is formally banned and repressed by the state, wary of a group whose members have assas sinated, among others, President Sadat in 1981, and at tempted to assassinate President Nasser in 1954. Its ideas have subsequently spread into neighbouring Sudan and Somalia, as well as elsewhere.

Abdel Salam and Alex de Waal argue that the project of Islamism has failed in Egypt and the Horn of Africa. They point not only to the failure of an Islamist leadership in Sudan to maintain a competent, peaceful, and popularly legitimate system of rule, but also of Islamists in Egypt to gain formal state power. This 'impasse', they suggest, is an indication of the intellectual failure of the project to move from generalizations and resistance to designing structures for establishing secure and popular successful rule.

This impasse has had two consequences. The first is a tendency to violence and repression of dissent, by those in power, or the intimidation and murder of anti Islamist Muslim intellectuals. The second impact of the failure of political Islamism is the growth of nonstate voluntarism which has led to a culturally specific form of development. Islamism sees political and social action as inseparable, and the failure to secure power in Egypt and elsewhere has channelled energy into the nongovernmental field.

In the field of education, Islamists in Egypt have provided free public transport to ferry students to uni versity. They introduced segregated seating in lectures to protect women from sexual harassment, and distributed subsidized veils and long robes to women for the same reason. They have established hospitals and healthcare provision, and taken over professional associations sup porting the causes of groups such as engineers and medical professionals. Multiplying Islamist NGOs have supported commerce and become involved in disaster relief. The strength of Islamist voluntarism was demon strated following a devastating earthquake in 1992 that killed over 500 people. While the state response was slow and ineffective, Islamist volunteers swiftly set up tents, and distributed food, blankets, and medicine. All of these activities increased support for the Islamists by embed ding them in civil society.

Such strategies have given Islamists unprecedented control over many areas of life in Egypt. However, de Waal and Salam argue that they are still unable to get beyond the impasse – they nurture intolerance and re inforce practices of censorship that deepen authori tarianism – allowing the Egyptian state to continue to repress them. Furthermore, the exploitation of Islamic NGOs by violent terrorist organizations such as Al Qaeda has, since 2001, brought the activities of these organizations under the tight scrutiny of criminal in vestigators throughout the world.

Conclusion

From Bhutan to Britain, there are some states in the world in which the clergy play a formal role in politics and from the USA and Russia to Israel and Egypt, many where explicitly religious political ideologies are increasingly important. Some thinkers would describe these processes as 'theocracy'. However, there are very few theocracies, as classically understood in the sense of the formal exercise of political sovereignty by the clergy. In the case of the State of the Vatican/the Holy See, this appears a unique historical aberration – for revolutionary Iran, Taliban Afghanistan, and Sudan, a highly contested project whose medium term durability remains uncertain.

Nonetheless, the mobilization of Islamist ideologies, with their promise of social renewal and economic just ice, has made theocracy an increasingly realistic alter native to other forms of government in some parts of the world. The general inability of these projects to capture states and establish popular legitimate and stable and peaceful systems of government has led many commen tators to proclaim the failure of modern Islamist variants of theocracy. However, in spite of their relative rarity on the global stage, theocracies have been able to influence geographies of development at different scales.

See also: Political Geography; State.

Further Reading

Abdullah, Y. (1996). The Holy See at United Nations Conferences: State or Church? *Columbia Law Review* 96(7), 1835 1875.
Agnew, J. (2006). Religion and geopolitics. *Geopolitics* 11(2), 183 191.
Bayes, J. and Tohidi, N. (eds.) (2001). *Globalization, gender and religion: The politics of women's rights in Catholic and Muslim contexts*. Basingstoke: Palgrave.
Bruneau, T. (1980). The Catholic Church and development in Latin America: The role of Basic Christian Communities. *Development* 8(7 8), 535 544.
Enayat, H. (2005). *Modern Islamic political thought: The response of the Shi'i and Sunni Muslims to the twentieth century*. London: I.B. Tauris.
Gottfried, P. E. (2002). *Multiculturalism and the politics of guilt: Towards a secular theocracy*. London: University of Missouri Press.
Kepel, G. (1994). *The revenge of God: The resurgence of Islam, Christianity and Judaism in the modern world*. Cambridge: Polity.
Proctor, J. (2006). Religion as trust in authority: theocracy and ecology in the United States. *Annals of the Association of American Geographers* 96(1), 188 196.
Ratzel, F. (1969, 1896). The laws of the spatial growth of states. In Kasperson, R. & Minghi, J. (eds.) *The structure of political geography*, pp 17 28. London: University of London Press.
Ryall, D. (2001). The Catholic Church as a transnational actor. In Josselin, D. & Wallace, W. (eds.) *Non state actors in world politics*, pp 41 58. Basingstoke: Palgrave.
Salam, A.H.A. and Waal, A.d. (2004). On the failure and persistence of jihad, In de Waal A.(ed.), *Islamism and its enemies in the horn of Africa*, pp 21 70. London: C. Hurst and Company.
Secor, A. (2001). Islamist politics: Antisystemic or post modern movements? *Geopolitics* 6(3), 117 134.
Strozier, C. and Swiderski, K. (2005). The pyschology and theocracy of George W. Bush. *Journal of Psychohistory* 33(2), 102 116.
Sykes, S. (2006). *Power and Christian theology*. London: Continuum.
Yorgason, E. and Robertson, D. (2006). Mormonism's raveling and unraveling of a geopolitical thread. *Geopolitics* 11(2), 256 279.

Relevant Websites

http://www.hizb.org.uk
Hizb Ut Tahrir (UK).
http://www.vatican.va
The State of Vatican City.

Therapeutic Landscapes

W. M. Gesler, University of North Carolina, Chapel Hill, NC, USA

Glossary

Cultural Ecology The study of the interactions between a group with specific material and symbolic practices and the natural environment.

Cultural Landscape The landscape that results when a culture group interacts with a natural environment.

Humanism A body of ideas and approaches that emphasizes the importance of human awareness, consciousness, and creativity.

Post-Structuralism A body of thought that stresses the importance of language as constituting reality; the formation or identity of subjects; power and its intimate connection with knowledge; an opposition to thinking in terms of binary opposites; and an insistence that there are no facts, only interpretations.

Structuralism The theory that surface appearances can be explained only by hidden causal mechanisms that produce divisions in society such as class, gender, ethnicity, sexual orientation, and age.

Therapeutic Landscape A conceptual framework used to analyze physical, social, and symbolic environments as they contribute to physical and mental health and well-being in places.

Introduction

Therapeutic landscapes is a conceptual framework that has been developed mainly by health geographers for the analysis of physical, social, and symbolic environments as they contribute to physical and mental health and well being in places. The term is a geographic metaphor for how the healing process works itself out in specific sites such as the home, a physician's office, or a hospital ward. It has its origins in the 'new cultural geography' of the 1980s and the reform movement within medical geography in the 1990s. The concept is aligned with the socioecological model of health and the holistic health paradigm. It integrates scholarship from the social sciences, humanities, medicine, and public health from a geographic perspective. Use of the concept extends the study of healthcare beyond the prevailing emphasis on physical environments, clinical studies, and quantitative assessments. The framework has been used to examine places with lasting reputations for healing, well being, and health maintenance in places, and a variety of Western and non Western settings such as gardens, healing retreats, deprived neighborhoods, and fictional places.

Origins of the Concept

The therapeutic landscape concept borrows heavily from the notion of a cultural landscape developed throughout the twentieth century by cultural geographers. It is based on three strands of thinking. The first, arising from traditional cultural geography, is cultural ecology, which investigates human interactions with and transformations of natural environments. The emphasis is on the material aspects of culture such as agricultural practices or the building of cities. This aspect of therapeutic landscapes is operationalized as the physical environment. It includes healing elements found in nature such as water, medicines from plants, and scenic beauty, as well as aspects of built environments such as the material used in and the layout of hospital wards.

The second and third strands of thinking behind therapeutic landscapes derive from a broadening of the cultural landscape idea through the application of social and cultural theory. From structuralism and post structuralism came ideas about the social construction of beliefs and institutional practices, the importance of power in human social relations, social exclusion, legitimization, marginalization, and resistance to attempts to control others. This aspect of therapeutic landscapes is operationalized as the social environment. It encompasses such aspects of healing situations as patient–doctor interactions, labeling people as deviant mentally or physically, Michel Foucault's notions about governmentality and surveillance, and healthcare consumerism.

Humanism provides a theoretical background to the third aspect of therapeutic landscapes. Here the focus is on mental constructs, on people's perceptions and experiences, on what a landscape or place is interpreted to mean. Physical and abstract symbols, myths, and stories are important to this perspective. This aspect of therapeutic landscapes can be operationalized as the symbolic environment. In healthcare situations this environment might include the language used by a healer in talking to a patient, the color of a room in a hospital, or stories told by a *shaman* during a healing session.

It should be noted that the three environments described above are merely convenient categorizations. In practice, aspects of a healing situation may fall into more than one category. For example, a discussion about a patient between a nurse and a doctor in an operating theater could have both symbolic meaning and reveal social hierarchies. Furthermore, the content and relative importance of the three environments will vary greatly by time and place.

Applications of the Concept

The first use of the therapeutic landscape framework was to investigate places that achieved lasting reputations for healing. An example is the sanctuary at Epidauros in Greece where the healing God man, Asclepius, was believed to heal people in dreams over a period of 1000 years. The remoteness of the site, the presence of water and trees, the involvement of sacred animals, and the construction of several magnificent buildings that blended in with the natural surroundings were important aspects of the physical environment. Aspects of the social environment conducive to healing included a degree of equality in patient–doctor relationships, access of doctors and patients to Asclepius as a personal God, and the many activities that took place in the sanctuary. The symbolic environment encompassed location on sacred ground, dream healing, and myths about Asclepius' birth and healing powers.

Over time, the use of the therapeutic landscape concept was extended beyond famous healing places to include places associated with the maintenance of health and wellness. The concept has seen international applications. A wide variety of places and subjects from countries around the world have been examined: a national forest in England, a wilderness area in Alaska, a children's hospital and children's health camps in New Zealand, home care workers and declining urban neighborhoods in Canada, First Nations Peoples in Canada and Native Americans in the US, a mental hospital in East London, mental health user collectives in Nottingham, the chronically homeless in the UK, a rural respite care center in southwest England, and fictional settings in France and the US. Reacting to the Western bias of most studies, research was carried out using non Western examples such as the Navajo of the Southwestern United States, the First Nations Peoples of Canada, and Havana, Cuba.

Ongoing critiques of the therapeutic landscape concept have enhanced its theoretical underpinning and its usefulness. Some researchers have emphasized its application to everyday pursuits of health and well being. It has been suggested that the negative aspects of healing environments be emphasized, as well as the positive ones. Several scholars have pointed out that environmental features that are therapeutic for one person may not be for others.

Engagement with the therapeutic landscape concept facilitates interactions with other health related disciplines. Perhaps most obvious are connections with environmental psychology, sociology, medical anthropology, public health, and biomedicine. Recent studies have incorporated such fields of study as nursing, comparative literature, psychoanalysis and psychotherapy, history, politics, and environmental perception.

Scholarship based on the therapeutic landscape concept has been one of the key developments in the transition from medical to health geography and one of the principal components of the identity of the latter. At the same time, it has helped to broaden the scope of health geography by connecting it to the broader fields of cultural and social geography. Overall, therapeutic landscapes have proven to be an important contribution by health geographers to the study of health and well being in places. The idea continues to evolve and can be applied to new locales and types of healthcare situations. It promises to inform an increasing body of literature within health geography and beyond.

See also: Cultural Geography; Difference/Politics of Difference; Gardens and Gardening; Health Geography; Humanism/Humanistic Geography; Land Change Science; Sense of Place; Social Geography; Structuralism/Structuralist Geography; Wellbeing.

Further Reading

Cosgrove, D. and Jackson, P. (1987). New directions in cultural geography. *Area* 19, 95 101.

Gesler, W. M. (1992). Therapeutic landscapes: Medical issues in light of the new cultural geography. *Social Science and Medicine* 34, 735 746.

Gesler, W. M. (2003). *Healing Places*. Lanham, MD: Rowman and Littlefield.

Gesler, W. M. (2005). Therapeutic landscapes: An evolving theme. *Health and Place* 11, 295 297.

Gesler, W. M., Bell, M., Curtis, S., Hubbard, P. and Francis, S. (2004). Therapy by design: Evaluating the UK hospital building program. *Health and Place* 10, 117 128.

Gesler, W. M. and Kearns, R. A. (eds.) (2002). *Culture/Place/Health*. London: Routledge.

Kearns, R. A. (1993). Place and health: Towards a reformed medical geography. *Professional Geographer* 45, 139 147.

Kearns, R. A. and Gesler, W. M. (1998). *Putting Health into Place*. Syracuse, NY: Syracuse University Press.

Palka, E. J. (2000). *Valued Landscapes of the Far North: A Geographical Journey through Denali National Park*. Lanham, MD: Rowman and Littlefield.

Williams, A. (1998). Therapeutic landscapes in holistic medicine. *Social Science and Medicine* 46, 1193 1203.

Williams, A. (ed.) (1999). *Therapeutic Landscapes: The Dynamic between Place and Wellness*. Lanham, MD: University Press of America.

Wilson, K. (2003). Therapeutic landscapes and First Nations People: An exploration of culture, health and place. *Health and Place* 9, 83 93.

Thiessen Polygon

L. Mu, University of Georgia, Athens, GA, USA

Glossary

Delaunay Triangulation Triangulate a finite set of points, if all the triangles satisfy the empty circle criterion, we call this triangulation a Delaunay triangulation.

Dual Graph The dual graph of a planar graph with tessellated polygons is formed by taking the center of each polygon as a vertex and joining the centers of adjacent polygons.

Empty Circle Criterion Triangulate a finite set of points, if the circumcircle of a triangle in the triangulation is a circle and contains no other points than the three nodes, the triangle satisfies the empty circle criterion.

Euclidean Plane The two-dimensional Euclidean (or Cartesian) space that is parameterized by coordinates, so that each point is located based on its position with respect to two perpendicular lines, the coordinate axes.

Triangular Irregular Network (TIN) A 2.5-dimensional surface model of digital elevation. A TIN is created based on the construction of a Delaunay triangulation. Each node is raised to its *z*-value (elevation or another numeric attribute).

Thiessen polygons are named after American meteorologist, Alfred H. Thiessen, for his contribution of developing a method for estimating average areal climate conditions from data collected at scattered points representing weather stations. Thiessen used the method to define regions around each weather station, so that every location in a given region has the enclosed station as the nearest among all stations. These regions, known as Thiessen polygons, define the territory that is closer to a particular point (in this case, a weather station) than to any other point. In Thiessen's original work, the area of each Thiessen polygon was used as the weight of that weather station in estimating the average rainfall for the whole study area. The term Thiessen polygon is widely used in geography, hydrology, and meteorology. These polygons are also commonly called Voronoi polygons in fields of mathematics, computer science, computational geometry, and astronomy, as the Russian mathematician George Voronoi is believed to be the first to formally introduce the concept in mathematics. In this article, the terms Thiessen polygons and Voronoi polygons are used interchangeably.

Definition

The formal definition of Thiessen polygons derives from topological relationships between points in a two-dimensional (2D) space. Given a set of two or more distinct points in the Euclidean plane, we associate all locations in that space with the closest member of the point set with respect to Euclidean distance. The result is a division (tessellation) of the plane into a set of the regions associated with members of the point set. We call this tessellation the planar ordinary Voronoi diagram generated by the point set. The regions associated with each point are called ordinary Voronoi polygons (Thiessen polygons).

Mathematically, let P be a finite set of points in the Euclidean plane, $P = \{p_1, \ldots, p_n\}$, where $2 \leqslant n < \infty$. Let x be any location in the plane. The Euclidean distance between x and p_i is $d_e(x, p_i)$. Let $T(p_i)$ denote the Thiessen (Voronoi) polygon of the point p_i, that is, p_i's influence region in P, then

$$T(p_i) = \{x \,|\, d_e(x, p_i) \leq d_e(x, p_j),\ \text{for all } j, j \neq i \text{ and } i, j \leq n\} \quad [1]$$

In **Figure 1**, the Thiessen polygon of point p_1 is the shaded area. Any location within that polygon has a shorter distance to p_1 than to all other points (p_2, p_3, p_4, and p_5).

Discoveries and Rediscoveries of Thiessen Polygons

Mathematicians and scientists discovered and rediscovered Thiessen polygons many times, and often referred to them using different terms. Among those, Voronoi diagram/polygons and Thiessen polygons are the most well known. However, a dozen other terms have been used to describe Thiessen polygons, and these terms have been associated with the development of theories, algorithms, and applications in various fields. The earliest examples of Thiessen polygons can be traced back to the seventeenth century, when Descartes discussed cosmic fragmentation of the 'heavens' in 1644. He divided the sky into polygons. Each polygon has a star in the center with concentric circles representing orbits, and the interaction of all the orbits from different stars form the boundaries of the polygons. Those polygons are actually Thiessen polygons. The process Descartes used in creating the polygons is consistent with the 'growth model' which is used today for constructing Thiessen polygons and is discussed further in this chapter. **Table 1** summarizes the names associated with Thiessen polygons found in the literature.

Thiessen polygons are not only named differently, but also viewed and understood from different perspectives. Thiessen polygons draw attention from researchers across many fields because their spatial patterns can be found in natural situations such as crystal growing and describing the potential areas from which trees draw moisture and nutrients from soil. There are interesting and surprising mathematical properties of those polygons, and they are robust enough to solve many seemingly unrelated com putational problems. In 1991, Aurenhammer gave a comprehensive bibliography on Voronoi diagrams that summarized various viewpoints and concerns of Voronoi diagram/polygons from different fields.

In geography and related fields, researchers have made great progress in developing methods and applications using Thiessen polygons in varied forms. Because Thiessen polygons integrate considerations of location, distance, influence range, and topological relationship, they are viewed and used by geographers in many dif ferent ways. They serve as a proximity method that can be used to assign school children to the nearest school, for example. They are also used as an automated topology for point data to define neighbors in geostatistics, and as a basic data structure in geographic information systems (GIS). Thiessen polygons have been applied in de lineating influence regions or service areas for points such as fire stations and newspapers, in defining neighbors for spatial analysis, and in calculating weighting factors as in the original Thiessen study.

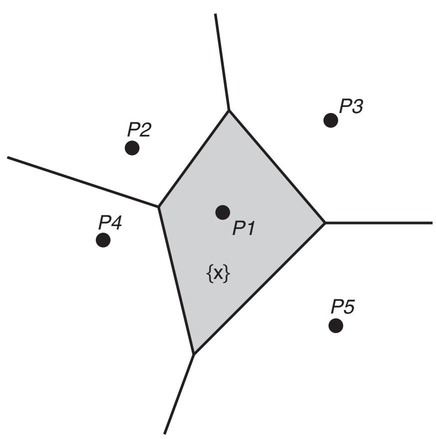

Figure 1 A Thiessen polygon $T(p_1)$ $\{x \mid d_e(x,p_1) \leqslant d_e(x,p_2), d_e(x,p_3), d_e(x,p_5), d_e(x,p_4)\}$. Source: Modified with permission: Okabe, A., Boots, B. N., Sugihara, K., and Chiu, S. N. (2000). *Spatial Tessellations, Concepts and Applications of Voronoi Diagrams, Wiley Series in Probability and Statistics*, pp 44. New York: Wiley.

Methods of Constructing Thiessen Polygons

Conceptually, the construction of Thiessen polygons can be achieved by two types of spatial processes: growth and assignment. The growth process is usually simulated using a buffer like growing model (**Figure 2**). The buf fers expand around each point at the same speed. With the expansion of the circular buffers, their frontlines intersect with each other and form the boundaries of Thiessen polygons.

In contrast, the assignment process is implemented mathematically based on topological relationships among the points. For the assignment process, there is a wide variety of methods and algorithms ranging from very straightforward and naive to quite complex yet robust. Five representative methods are discussed in the fol lowing section. Since many of these methods derive from mathematics and computer science, the term Voronoi is used more often.

Plane intersect. For a given point p_i, generate half planes for each pair of points p_i and p_j (all other points) by bisecting a line between the two, then intersect all half planes. The resultant space is the Voronoi polygon of p_i.

Table 1 Alternative terms for Thiessen polygons in different fields

Date	*Term*	*Author(s)*	*Field*
1644	Heavens	Descartes	Astronomy (cosmic fragmentation)
1850	Dirichlet tessellation	Dirichlet	Mathematics
1908	Voronoi diagram/polygons	Voronoi	Mathematics
1909	Area of influence polygons	Boldyrev	Geology
1911	Thiessen polygons	Thiessen	Meteorology and geography
1927	Domains of action	Niggli	Crystallography
1933	Wigner-Seitz regions	Wigner and Seitz	Chemistry and physics
1958	Atom domains	Frank and Casper	Physics
1965	Areas potentially available	Brown	Ecology (intensity of trees in a forest)
1966	Plant polygons	Mead	Ecology
1967	Blum's transform	Blum	Biology and visual science
1985	Capillary domains	Hoofd et al.	Anatomy
1987	Voronoi foam	Icke	Astronomy

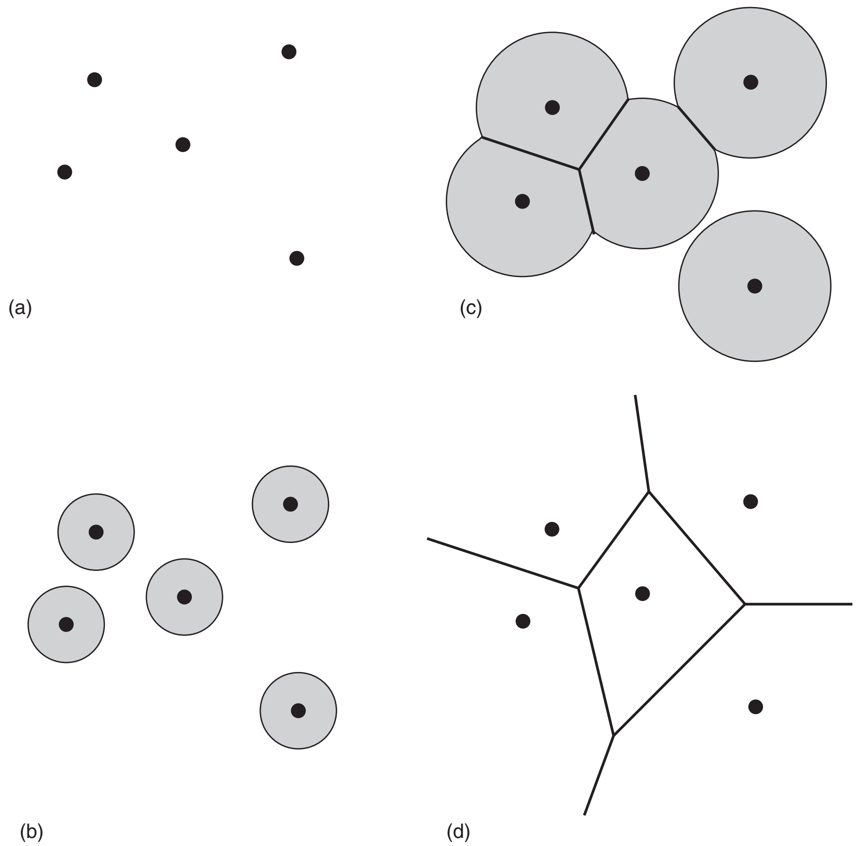

Figure 2 Growth process to construct Thiessen polygons.

This method is simple, but it is a direct translation of the definition of Voronoi polygons. In **Figure 3**, the shaded area is the Voronoi polygon of p_1, which is obtained by intersecting half planes of p_1 p_2, p_1 p_3, p_1 p_5, and p_1 p_4.

Incremental. The method starts with constructing a simple Voronoi diagram with a subset of original points (referred to as generator points), and updates it by adding other points one by one.

This method is robust and efficient, thus it is one of the most practical methods. In **Figure 4**, a simple diagram is constructed from points p_2, p_3, p_5, and p_4. An additional point p_1 is inserted, and the new Voronoi polygon of p_1 (dot lines) is updated. The process continues for other points.

Divide and conquer. All generator points are recursively divided into smaller subsets ($\approx$ the integer of $n/2$) until the number of points is $\leqslant 3$. Voronoi diagrams are constructed for each subset and the subsets are then merged recursively into the final Voronoi diagram.

This is another popular method. **Figure 5** shows a simple example. The left subset constructed from p_1, p_2, and p_4 (dashed lines) is merged with the right subset constructed from p_3 and p_5 (shaded lines). The thick line is the merged boundary composed of segments of bisecting lines of near boundary points.

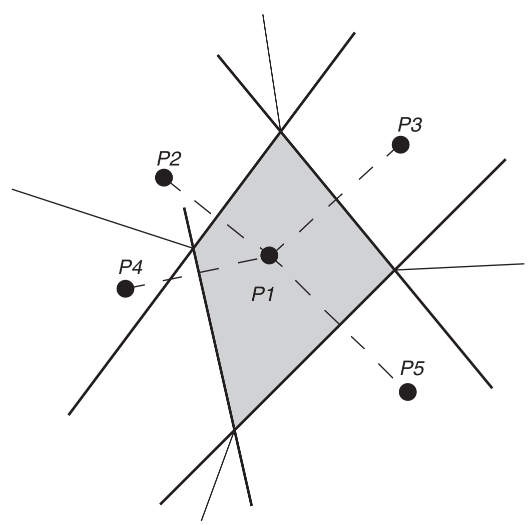

Figure 3 Plane intersect method. See text for details.

Plane sweep (*Fortune's algorithm*). An imaginary sweep line is moved over the plane from top to bottom. A parabola is formed between each generator point (focus) above the sweep line and the line itself (directrix).

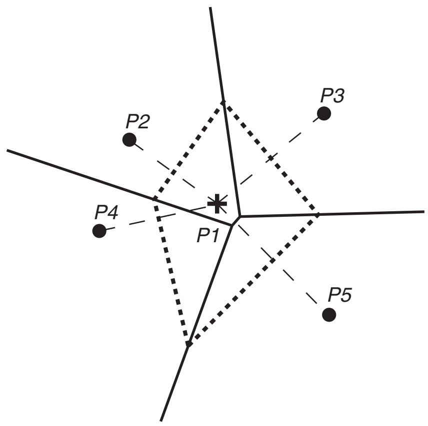

Figure 4 Incremental method. See text for details.

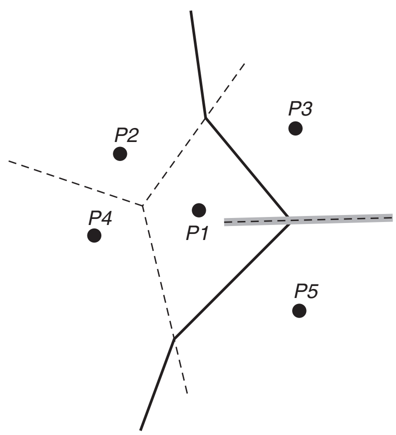

Figure 5 Divide-and-conquer method. See text for details.

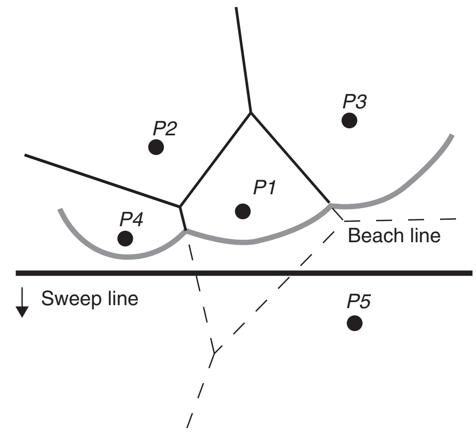

Figure 6 Plane sweep method. See text for details.

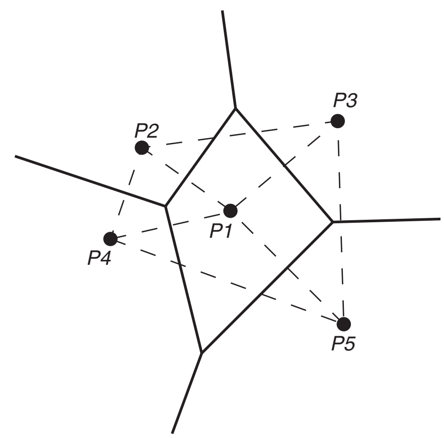

Figure 7 The dual graph of the Delaunay triangulation method. See text for details.

The collection of all intersection points of the parabolas during the sweeping constructs the Voronoi diagram.

This method offers an important and fundamental algorithm in computational geometry. In **Figure 6**, with the move of the sweep line (straight, thick, black line), a 'beach line' (parabolas: curved, thick, grey line) forms the final Voronoi diagram.

The dual graph of the Delaunay triangulation. All points are input to create a triangular irregular network (TIN) that meets the Delaunay criterion. The Delaunay criterion, also called empty circle criterion, states that a circle drawn through the three nodes of a triangle will contain no other point. Each TIN edge is perpendicularly bisected to form the Thiessen polygons. The centers of TIN polygons are vertices of Thiessen polygons. The Thiessen polygons are the dual graph of the Delaunay triangulation.

This method is used in many GIS programs because the TIN model is an important model for representing surfaces. In **Figure 7**, the dash lines show the TIN edges, and the solid lines are the boundaries of Thiessen polygons.

Variations of Thiessen Polygons: Generalized Voronoi Diagrams

People have long been aware of the limitations of ordinary Thiessen polygons. The primary concerns are: first, the shape of the polygons is completely dependent on the locations of the sample data points, and does not reflect differences in attributes of the points. For example, some points may be larger or more important than others, and one might expect the areas of influence for these larger points to extend further than those for smaller points. These differences in attribute values are not considered in constructing ordinary Thiessen polygons. Second, in using the polygon areas as weighting factors, the weight value for each polygon is estimated from a sample of one point which may lead to inaccuracies, errors, and difficulties in estimating weights for points other than the generator points. Finally, Thiessen polygons do not assume that points closer together are more similar than points far apart, and thus ignore spatial autocorrelation.

Thus, ordinary Thiessen polygons are restricted in terms of point location, weight, rules for assignment or growth, and space, and so on. Complex, real world environments often cannot meet these simplified assumptions. In recent years, scientists have relaxed these assumptions in research on generalized Voronoi diagrams. An important extension is to introduce weights for each point so that the extent of the Thiessen polygon for each point is a function of both distance and weight. For instance, consider the problem of identifying service areas for hospitals. Let the capacity of each hospital be the weighting factor. **Figure 8** shows the capacity weighted Thiessen polygons for the same set of points as in **Figure 1**. The diagram demonstrates that when a weighting factor is considered, the derived weighted Thiessen polygons (weights in parentheses: solid, black lines) show a very different spatial pattern as compared to the simple Thiessen polygons (all weights = 1: gray, dashed lines). Other extensions of Voronoi diagrams include the use of different distance metrics, such as network distance, and the incorporation of barriers to movement in generating Voronoi polygons.

Applications of Thiessen Polygons in Geography

In addition to using Thiessen polygons to determine weighting factors for calculating areal averages for point data, there are many other uses of this method in geographic information science and human geography. In geographic information sciences, researchers are using Thiessen polygons to revisit traditional GIS operations like buffer zone and polygon overlay and to extend the algorithmic foundations of GIS. The Voronoi diagram has also been applied as an 'automated' topology for point data to define neighbors, and as a basic data structure in GIS. Spatial weights, used in identifying 'nearby' areas in spatial analysis methods, have been generated based on Voronoi methods. Finally, Thiessen polygons also inspire profound thinking about proximity relationships, the meaning of 'neighbor' and the issues of human versus computer models of space.

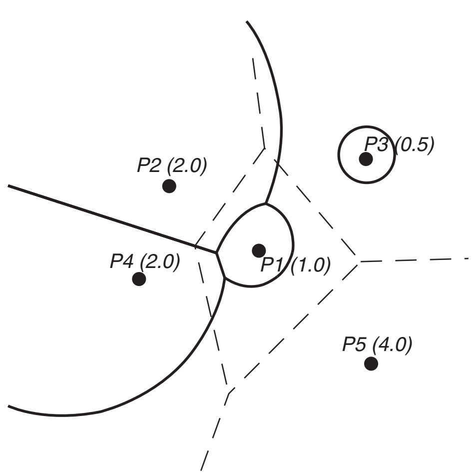

Figure 8 Weighted Thiessen polygons. See text for details.

Applications in human geography include identifying market areas for retail stores and service facilities such as hospitals and libraries. In fact, Voronoi diagrams are most famous for solving the 'post office problem' which involves delineating postcode/delivery boundaries based on known post office locations. Thiessen polygons provide a convenient method for identifying territories or 'areas of influence' around points. This is an important topic in research on the geographies of retail and service activities, media outlets, and sociocultural institutions – all activities whose influence extends over nearby territories. The diagrams have also been used to study the changing hierarchy of cities in the United States and to analyze central place settlement patterns. Applications in the social sciences and humanities include using the Voronoi model to understand linguistic structure and conducting qualitative analyses with the Voronoi model of space, and, from the viewpoint of a social scientist, the Voronoi model of space has its advantages because it can consider location, distance, and influence range and relationship. **Table 2** provides a glimpse of some research of the applications and development of Thiessen polygons.

Summary and Conclusion

Thiessen polygons are widely used in geography, computer science, and other fields. The major application of Thiessen polygon in human geography is to delineate dominant regions or service areas for point data, such as stores or hospitals. Thiessen polygons are models of spatial processes, nonparametric techniques in point pattern analysis, organizing structures for displaying spatial data, and information theoretical approaches to point patterns. Although ordinary Thiessen polygons are used most commonly in human geography and other fields due to their simplicity, their limitations have hindered applications. Such limitations include using point location as the only determination factor to construct the model, lack of weight consideration for each point, and inadequate consideration of spatial relationships among points. Research developments in weighted Thiessen polygons and Thiessen polygons for lines, polygons, and three dimensional (3D) objects have effectively overcome the limitations of ordinary Thiessen polygons. Future research of Thiessen

Table 2 Some examples of Thiessen/Voronoi research

Author(s)	*Date*	*Topic*	*Study area*
Hyson and Hyson	1950	Economic law of market areas	
Gambini *et al.*	1967	Market place properties	
Illeris	1967	Functional regions of urban centers	Denmark
Huff and Jenks	1968	Urban systems	
Hubbard	1970	Functional regions	Jamaica
Beckman	1971	Market potential	
Hogg	1971	Archaeology site territory define	England
Boots	1973	Subdivision of space	Great Britain
Huff	1973	Urban spheres of influence	
Boots	1975a	Patterns of urban settlement	
Boots	1975b	Structure of socioeconomic cellular network	Great Britain
Wood	1974	Functional regions	Kenya
Cox and Agnew	1974	Theoretical counties	Ireland
Fraser	1977	Forest sampling	
Getis and Boots	1978	Spatial processes	
Jones	1979	Economic law of market areas	
Huff and Lutz	1979	Urban hierarchy	Ireland
Edelsbrunner *et al.*	1983	Voronoi diagram, alpha-shape, and Delaunay triangulation	
Aurenhammer and Edelsbrunner	1984	Constructing the weighted Voronoi diagram	
Vincent and Daly	1990	Weighted Voronoi diagrams in GIS	

Reproduced from Mu, L. (2004). Polygon characterization with the multiplicatively weighted Voronoi diagram. *The Professional Geographer* 56(2), 223 239.

polygons in human geography may investigate issues of Thiessen polygons of multidimensions (e.g., multiple socioeconomic variables), Thiessen polygons of dynamic phenomena (e.g., transportation network and human movement), and Thiessen polygons of real time inputs (e.g., wireless communication).

See also: Point Pattern Analysis; Spatial Data Models.

Future Reading

Aurenhammer, F. (1991). Voronoi diagrams: A survey of a fundamental geometric data structure. *ACM Computing Surveys* 23(3), 345 405.

Boots, B. N. (1980). Weighting Thiessen polygons. *Economic Geography* 56, 248 259.

Boots, B. N. (1987). *Voronoi (Thiessen) polygons*. (*Concepts and Techniques in Modern Geography*). Norwich, UK: Geo Books.

de Berg, M., van Kreveld, M., Overmars, M. and Schwarzkopf, O. (2000). *Computational Geometry: Algorithms and Applications*. New York: Springer Verlag.

Edelsbrunner, H., Kirkpatrick, D. and Seidel, R. (1983). On the shape of a set of points in the plane. *IEEE Transactions on Information Theory IT* 29(4), 551 559.

Edwards, G. (1993). The Voronoi model and cultural space: Applications to the social sciences and humanities. In Frank, A. U. & Campari, I. (eds.) *Spatial Information Theory: A Theoretical Basis for GIS*, pp 202 214. Springer Verlag: Berlin.

Fortune, S. (1987). A sweepline algorithm for Voronoi diagram. *Algorithmica* 2, 153 174.

Gambini, R., Huff, D. L. and Jenks, G. F. (1967). Geometric properties of market areas. *Regional Science Association: Papers* 20, 85 92.

Gold, C. M. (1992). The meaning of 'neighbour'. *Theories and Methods of Spatio Temporal Reasoning in Geographic Space*, pp 220 235. Springer Verlag: Berlin.

Gold, C. M., Remmele, P. R. and Roos, T. (1997). Voronoi methods in GIS. In van Kreveld, M., Nievergeld, J., Roos, T. & Widmeyer, P. (eds.) *Algorithmic Foundations of GIS*, pp 21 35. Springer Verlag: Berlin.

Griffith, D. A. (1983). The boundary value problem in spatial statistical analysis. *Journal of Regional Science* 23, 377 387.

Huff, D. L. (1973). The delineation of a national system of planning regions on the basis of urban spheres of influence. *Regional Studies* 7, 323 329.

Okabe, A., Boots, B., Sugihara, K. and Chiu, S. N. (2000). *Spatial Tessellations, Concepts and Applications of Voronoi Diagrams, Wiley Series in Probability and Statistics*. New York: Wiley.

Third World

J. Connell, University of Sydney, Sydney, NSW, Australia

Glossary

CARICOM Caribbean Community and Common Market.

COMECON The Council for Mutual Economic Assistance was an economic organization of communist states and a kind of Eastern Bloc equivalent to, but more inclusive than, the European Economic Community.

Cold War The protracted geopolitical, ideological, and economic struggle that emerged after World War II between the global superpowers of the Soviet Union and the United States, supported by their military alliance partners.

Dependency School A school of thought, initially associated with Latin American economists, that argued that development problems were primarily a function of exploitative links between less-developed countries and those of the West.

G8 The 'Group of Eight' had its origins in the 1973 oil crisis and subsequent global recession. It links eight nations, the last of which (Russia) joined in 1991, that have a combined two-thirds of the world's economic power.

Millennium Development Goals Eight wide-ranging development goals agreed to by all members of the United Nations to be achieved by 2015.

NICs Newly Industrialized Countries.

OPEC Organization of Petroleum Exporting Countries.

Sahel The savanna area on the southern border of the Sahara desert characterized by considerable poverty and other crucial development issues.

Introduction

Broad consensus exists that the term 'Third World' was first coined by the French demographer Alfred Sauvy in 1952 at the height of the Cold War, between the largely Western capitalist and Eastern socialist blocs. It was then used to apply to less developed countries that remained outside these two blocs, then respectively centered on the United States and the Soviet Union (USSR). The first sense of the phrase was actually the French *tiers monde* that was usually glossed as 'third force', indicating nonalignment rather than 'underdevelopment'. The concept thus emerged in a primarily political sense.

The original political dimension was emphasized at a conference in Bandung, Indonesia in 1955 when a group of African and Asian nations sought to assert their unity, common interests, and independence from both the capitalist and socialist world, identifying themselves as a nonaligned alternative world. Leading members of this group were Yugoslavia (under Tito), India (under Nehru), and Indonesia (under Sukarno). Subsequently the group lost identity and purpose as new nations became caught up in the spheres of influence of the two rival blocs.

Over time, notably in the 1960s, as more colonies became independent and the Cold War weakened, the phrase increasingly took on an economic rather than political sense. This reflected a new distinction, if not confrontation, between relatively rich and relatively poor nations, the former sometimes being referred to as industrialized states. That distinction had new geographical connotations. Rather than there being a confrontation between East and West, the growing division was between North and South, as most of the emerging Third World states were in the Southern Hemisphere, especially around the tropics. The original confrontation between East and West was retained in the growing use of First World to refer to the industrialized market economy states (the capitalist countries) and the centrally directed, socialist states, which became known as the Second World.

Around the start of the 1960s, the term 'Third World' essentially became a synonym for 'underdeveloped countries', or even as one geographer Norton Ginsburg had called them: "retarded countries." Such derogatory terms later disappeared largely, to be replaced by 'developing' or 'less developed', or 'two thirds world' and 'Global South', but the connotation of some sort of failure never entirely disappeared from the concept of Third World.

By the 1970s the term had acquired global usage though there was no general agreement over what or where it was. Some authors and studies took a geographical approach, where the Third World was essentially Africa, Asia, and Latin America (and later, by extension, Caribbean, Indian Ocean, and Pacific Islands), and where there was ambivalence about parts of the Middle East (such as Israel and Saudi Arabia), Japan, and South Africa. Others had a more structural approach that centered on poverty and the lack of significant industrialization. In either case early Eurocentrism was evident in relatively poor and less developed counties (such as Portugal and Greece) being placed in the First World, and the world being seen in primarily ethnic terms. In both perspectives socialist states remote from Eastern Europe, and with limited industry (such as Laos and Cuba), were usually regarded as part of the Second

World, hence the political perspectives partly survived. Equally, most countries grouped as Third World were also members of the Non Aligned Movement that came into existence in 1961, reflecting both – some sense of shared identity and some unity of both political and economic purpose – despite growing ties with one or the other of the northern blocs (**Figure 1**).

As most colonies eventually became independent states, such general groupings were increasingly found wanting. Much smaller countries also became in dependent, a process that continued until decolonization had almost reached its finale. By the end of the 1970s there were some 97 members of the United Nations from Africa, Asia, and Latin America alone, and thus well over 120 countries that might loosely be grouped within the Third World. Within this group there were evident and significant differences.

As smaller countries, and especially small island states, became independent the differences between countries like Tuvalu (with barely 10 000 people scattered over eight coral atolls in the central Pacific), recently created city states such as Singapore, and India were evident. India, for example, had human and natural resources, fertile land, large domestic markets, access to overseas markets and capital, while Tuvalu had none of these. Singapore had a growing financial service and manu facturing sector. Their colonial histories and cultures were also different. It made little sense to consider either the status of development, or the prospects for future development, in the same way for such disparate places.

Beyond obvious distinctions of geography (location, physical structure, and resources), culture, and history, there were other key distinctions, centered on natural resources and industrialization. Some countries were able to achieve a greater degree of economic growth through the export of mineral resources, especially oil. The am biguous status of Middle Eastern oil producers was evi dent early on, but after the establishment of OPEC in 1960, and declining oil production in the United States, the first 'oil shock' led to a rapid increase in oil prices and equally rapid economic growth amongst the Gulf states especially. To some extent the very success of these states meant further challenges for poorer states that were oil importers.

Other countries achieved less dramatic success with different commodities, such as copper in Chile, but economic growth was not always sustained, as in the case of Nauru, which went from great wealth to extreme poverty after the extinction of its phosphate in the late 1990s. Similarly rapid economic growth did not always equate with development in other arenas such as edu cation and health.

Industrialization was successful both in what became known as the Newly Industrializing Countries (NICs),

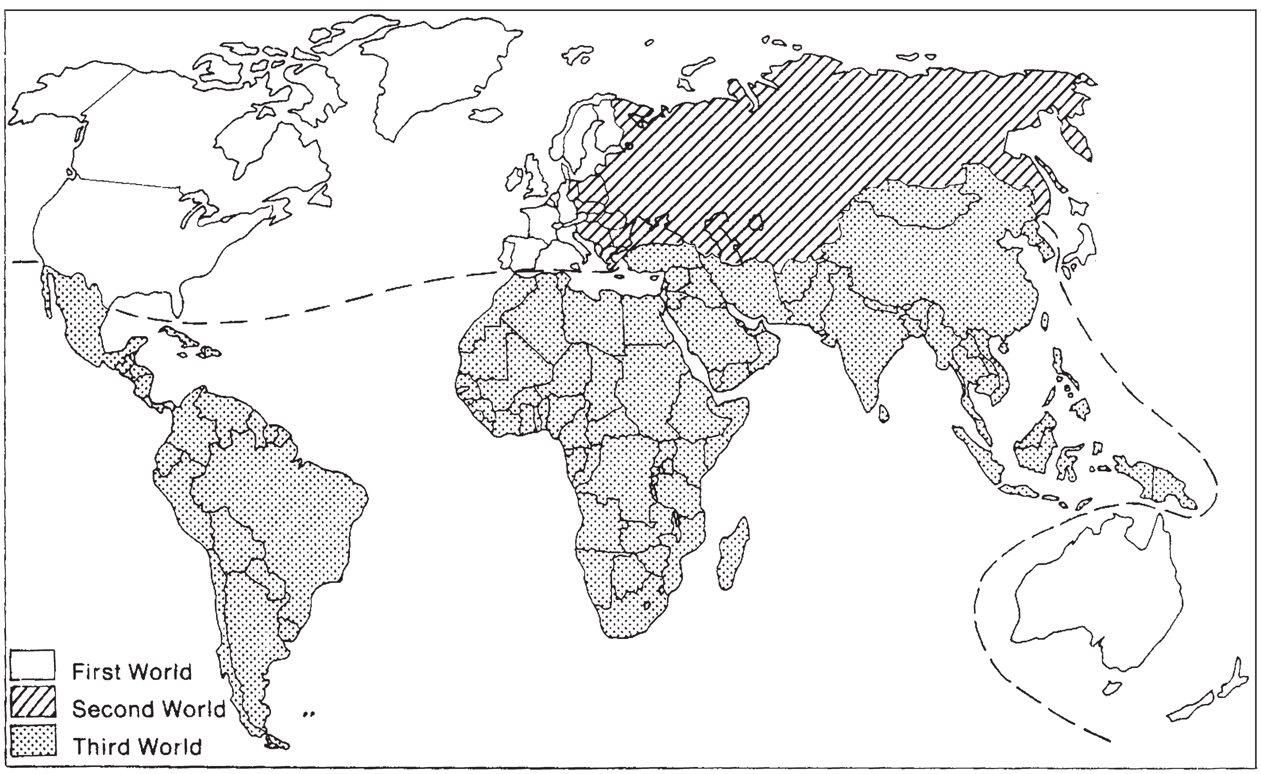

Figure 1 First, Second, and Third Worlds and the North South division (indicated by the broken line). Source: Barke, M. and O'Hare, G. (1984). *The Third World*, p 1. Edinburgh: Oliver and Boyd.

notably four in Asia – Taiwan, South Korea, Hong Kong, and Singapore – and Mexico and Brazil in Latin America. Based on economic performance alone, there were substantial differences within the evolving Third World.

UNCTAD and the Fourth World

As significant changes occurred within what was still labeled the Third World, the United Nations, through the United Nations Commission on Trade and Development (UNCTAD), distinguished a further group of countries, those that appeared to experience particular development challenges. Beyond general and unspecified notions of poverty, for the first time a set of basic indicators was developed to enable classification. These three indicators, first set out in 1968, were gross domestic product (GDP) per capita, the percentage share of manufacturing in the GDP, and the extent of adult illiteracy. For the first time this provided a simple measure of development centered on one social and two economic variables.

While the resultant classification excluded otherwise impoverished countries such as India, where there had been significant industrialization, it resulted in a list of 25 countries, 16 of which were in sub Saharan Africa, and seven in Asia. One (Haiti) was in the Caribbean and one (Western Samoa) in the Pacific. A particular concentration of such states existed in immediately sub Saharan Africa, the Sahel region stretching from the Atlantic coast to Ethiopia and Somalia in the east, where poverty was associated with intermittent drought and natural hazard. The intent of the classification was to distinguish countries that were in particular need of assistance whether through new overseas aid strategies, investment priorities, or trade arrangements. Over the next two decades other countries were added to the list, mainly in sub Saharan Africa, but none were removed.

By 2004 some 50 countries had been formally designated by the UNCTAD as 'least developed countries' in comparison with the 'less developed countries' of the Third World. Again most were in sub Saharan Africa (**Table 1**). The list is reviewed every three years. Three broad criteria were used to designate the least developed countries: low income, weak human assets (based on nutrition, health, education, and literacy), and economic vulnerability (based on instability in agricultural and industrial production, trade, diversity, and size). This group, with their seemingly greater problems, became known to many as the Fourth World. (The concept of the Fourth World has also been used, but much less frequently, to refer to either women or distinct ethnic minorities within the Third World.)

Differentiation within the Third World was paralleled by obvious divisions within the Second and First Worlds. In the First World, there were vast differences in structures and levels of economic development between countries such as the United States, Ireland, and Australia, the last retaining a largely nonindustrial economy. Much later these became evident in groupings such as the G8 countries (USA, France, Japan, Canada, UK, Italy, Germany, and later Russia), self selected as the leading global economies.

Similarly, in the Second World, the COMECON countries of Eastern Europe had pursued largely industrial development strategies, while the socialist states of Asia (and Cuba) had more mixed, but primarily agricultural economies. Moreover there were tensions between the two main socialist powers, China and the Soviet Union.

The Meaning of Development

Even by the time the UNCTAD had established its classification in 1968, it was evident that there were a series of problems attached to notions of the Third World. First, there were disagreements on what might be defined, and also measured, as parameters that constituted 'development'. Second, there were objections to these terms as neocolonial imperialist phrases that either had no meaning or were to be rejected as slurs on the status of people in many countries. Indeed such objections, largely following from the arguments of what was in the 1970s the emerging dependency school, emphasized that problems in developing countries were a function of active underdevelopment by rich world countries. Conceptualizing One World, in which countries were interlinked through relationships of trade, aid, investment, migration etc., seemed a more appropriate perspective.

Indicative of other perspectives, in 1974 Mao Tse Tung, the Chinese leader, said, "In my view, the United States and USSR form the First World. Japan, Europe and Canada, the middle section, belong to the second world. We are Third World…With the exception of Japan, Asia belongs to the Third World, the whole of Africa belongs to the Third World and Latin America too." In the midst of a new Cold War between China and the Soviet Union, Mao was arguing that the principal exponents of imperialism and capitalism were the United States and the Soviet Union. Political perspectives might therefore be quite different, and there was no consensus on how the world might be perceived.

As development problems appeared more acute, further thought went into the conceptualization of development, and two broad perspectives emerged. The first, which emphasized economic growth, centered on historical models, and primarily on that of Rostow, argued that the success of northern countries as they moved

Table 1 Bas c data on the east deve oped countr es (per cap ta GDP and popu at on: eve s and growth)

Country	*Per capita GDP (in 2002 dollars)*			*Annual average growth rates of per capita real GDP (%)*			*Level (millions)*	*Population Annual average growth rates (%)*		
	1980	*1990*	*2002*	*1980–1990*	*1990–2000*	*2000–2002*	*2002*	*1980–1990*	*1990–2000*	*2000–2002*
Afghan stan	..	..	..	..	..	..	22.9	–1.3	4.5	3.5
Ango a	987	934	863	0.5	–1.2	6.5	13.2	2.8	2.9	3.2
Bang adesh	220	246	329	1.1	2.4	2.7	143.8	2.6	2.3	2.1
Ben n	339	329	410	–0.5	1.7	2.4	6.6	3.0	3.0	2.7
Bhutan	279	466	698	5.4	3.4	4.4	0.9	2.1	3.0	2.8
Burk na Faso	172	182	225	0.8	1.5	2.5	12.6	2.7	2.9	3.0
Burund	125	143	109	1.2	–3.6	0.7	6.6	3.2	1.0	2.6
Cambod a	..	204	266	1.6	1.8	2.8	13.8	4.1	3.0	2.5
Cape Verde	..	956	1390	3.8	3.7	1.5	0.5	1.9	2.2	2.1
Centra Afr can Repub c	338	293	282	–1.0	–0.3	1.4	3.8	2.5	2.4	1.4
Chad	168	216	232	3.4	–0.8	6.5	8.3	2.6	3.1	3.0
Comoros	434	425	343	–0.3	–1.8	–0.4	0.7	3.1	3.0	2.9
Dem. Rep. of the Congo	330	269	111	–1.3	–7.3	–2.1	51.2	2.9	2.6	2.7
Dj bout	..	1209	861	–6.7	–3.6	–0.4	0.7	5.1	2.3	2.0
Equator a Gu nea	..	973	4517	–0.8	19.0	–1.9	0.5	5.0	2.6	2.7
Er trea	..	..	146	..	3.9[a]	5.6	4.0	2.8	1.7	3.7
Eth op a	..	76	87	–2.1	1.6	3.7	69.0	3.2	3.0	2.5
Gamb a	282	281	280	–0.1	–0.3	–0.2	1.4	3.7	3.4	2.9
Gu nea	..	328	380	1.7	1.3	2.5	8.4	2.6	2.9	1.5
Gu nea-B ssau	151	191	149	1.5	–1.8	–4.8	1.4	2.4	3.0	3.0
Ha t	751	568	437	–2.6	–2.1	–2.6	8.2	2.4	1.5	1.3
K r bat	513	457	512	–1.4	1.5	0.7	0.1	2.2	1.6	1.5
Lao PDR	..	199	304	1.0	3.9	2.9	5.5	2.6	2.5	2.3
Lesotho	233	295	405	2.5	2.8	3.5	1.8	2.1	1.3	0.4
L ber a	712	194	174	–8.2	0.4	–0.2	3.2	1.4	3.5	4.9
Madagascar	427	341	267	–1.7	–0.9	–6.1	16.9	2.8	2.9	2.9
Ma aw	173	141	158	–1.9	2.0	–2.0	11.9	4.6	1.8	2.2
Ma d ves	..	..	2000	..	3.8	–0.2	0.3	3.2	3.0	3.0
Ma	255	210	251	–1.6	1.3	2.4	12.6	2.5	2.8	3.0
Maur tan a	317	296	350	–0.5	1.5	1.8	2.8	2.3	2.7	3.0
Mozamb que	149	135	211	–1.0	3.3	9.8	18.5	0.9	3.0	1.9
Myanmar	..	..	..	..	..	..	48.9	1.8	1.6	1.4
Nepa	137	173	223	2.3	2.5	–0.3	24.6	2.3	2.4	2.3
N ger	298	216	188	–3.2	–1.0	1.3	11.5	3.2	3.5	3.7
Rwanda	232	216	210	–0.9	–1.6	4.4	8.3	3.1	1.3	3.5
Samoa	..	..	1482	..	3.0	2.7	0.2	0.3	0.8	0.9
Sao Tome and Pr nc pe	..	340	321	–0.4	–0.8	0.4	0.2	2.0	2.6	2.5
Senega	434	444	501	0.2	1.1	1.6	9.9	2.9	2.5	2.4
S erra Leone	318	277	166	–1.8	–5.6	1.9	4.8	2.4	0.7	3.9
So omon s ands	561	766	517	3.1	–1.0	–9.7	0.5	3.4	3.2	3.0
Soma a	..	..	..	..	..	..	9.5	0.8	1.9	4.3

Sudan	265	264	410	−0.2	3.0	6.3	32.9	2.5	2.4	2.3
Togo	418	338	288	−1.6	−0.6	0.2	4.8	3.3	2.8	2.6
Tuva u[b]	..	..	..	..	..	..	0.0	1.6	1.4	1.3
Uganda	..	161	235	−0.6	3.9	2.2	25.0	3.4	3.0	3.2
Un ted Rep. of Tanzan a	..	241	259	1.9	0.0	3.6	36.3	3.3	3.0	2.0
Vanuatu	1113	1231	1133	0.5	−0.5	−3.5	0.2	2.4	2.8	2.5
Yemen	..	482	538	..	1.6	−0.5	19.3	3.9	4.2	3.5
Zamb a	481	389	344	−2.2	−1.9	2.6	10.7	3.2	2.4	1.3
A LDCs	..	253	281	−0.1	0.9	2.4	699.6	2.6	2.6	2.4
A deve op ng countr es	761	901	1195	1.6	3.1	1.5	5018.5	2.1	1.7	1.5
Deve oped market economy countr es	18813	23832	28388	2.6	1.7	0.2	871.4	0.6	0.7	0.5
Countr es n Centra and Eastern Europe	..	3160	2781	1.0	−2.1	4.3	335.1	0.7	−5.5	−0.4

[a] 1993–2000

[b] Population 10466 and area 26 km^2

Source UNCTAD *Handbook of Statistics 2003* World Bank *World Development Indicators* online data

Note GDP per capita data are based on World Bank data on GDP and population data are based on United Nations/DESA/Population Division

Data for Ethiopia prior to 1992 include Eritrea Population data for Bhutan is from national sources through the industrial revolution provided an appropriate model for other countries to follow (Walt Rostow was an American economic historian and national security advisor, whose book *The Stages of Economic Growth* (1960), subtitled. 'A Non Communist Manifesto', provided strong advocacy for capitalist development). If developing countries adopted modern democratic and education systems, constructed adequate transport infrastructure, and invested in industry they could achieve economic growth. Primacy was continually extended to industrialization, rather than agricultural development. The basic idea was persuasive, since Western countries certainly had evolved and hence suggested that national decisions could influence progress.

This was criticized, initially by the dependency school, who pointed out that many countries not only lacked the preconditions to achieve economic growth but were actively held back by processes emanating from the rich world, including unequal exchange in trade and transfer of profits from peripheral countries through large corporations (that gave rise to generalizations about 'core' and 'periphery' countries, with the former group in the capitalist West exploiting the others). Moreover, growth was not necessarily linked to distribution of wealth or services and thus to ideas of development that had some relationship to notions of equity, while a sole focus on growth had potentially negative environmental consequences.

While emphasis on economic growth and trade had resulted in indicators of development that focused on industrial development and income generation, subsequent emphasis was placed on more distributive aspects of development, including access to health, housing, nutrition, employment, and education, and issues such as ecological stability and equity. Though this resulted in a wider range of development indicators, that came to be known as Basic Needs, some of these could be contentious and were often difficult to measure. Thus equity might be seen in terms of income levels, which were relatively easy to measure, but could also be perceived as relating to gender issues, which varied enormously across cultures, and were largely immeasurable.

For a quarter of a century broad ideas of basic needs, centered on health, nutrition, employment, education, freedom, housing, ecological stability, and some notion of equity (at national, regional, and household levels), informed most institutional perspectives on development. Over time they became more subtle, and were regularly assessed through the United Nations' global 'Human Development Index', undertaken irrespective of the political status of countries. By the end of the century this had effectively evolved into the 'Millennium Development Goals' (MDGs), which gave greater emphasis to poverty and hunger, education, gender equality, and educational sustainability. The MDGs, developed by the United Nations, both specified particular goals (such as the reduction of child mortality by two thirds between 1990 and 2015) and emphasized that a 'global partnership' was required to achieve the goals, since individual countries could not achieve the task on their own, and that trade reform and greater and more effective aid delivery were crucial to change and development.

While basic needs, and also MDGs, tended to be used at national levels they enabled international comparisons that both stressed that over time there had been little general shift in the manner in which development problems remained highly concentrated in sub Saharan Africa, and that in many counties in this region there was deprivation on multiple scales. Since there could be no ultimate consensus on which variables were most critical, and which should be included, or how they might be measured and compared, it made little sense to draw up league tables of success and failure and seek to identify particular worlds as demarcated by these indicators. The maps of the worlds that had accompanied early studies usually became things of the past, as development became more complex and subtle, even though the broad parameters were changed little, especially in terms of poverty, malnutrition, and access to basic needs.

While the evolution of the basic needs and MDGs gave a more subtle and nuanced perspective on development they remained essentially ahistorical indicators that did not necessarily imply particular policies or strategies to meet them, at national or international levels, or in what circumstances these goals had not been met. They therefore stood somewhat outside history and development in a broader sense and, rather like Rostow's model, implied a critical role for national development. They were primarily of value in focusing the attitudes and policies of international agencies, and emphasizing the needs of particular countries.

New Geopolitics and State Failure

Further geopolitical changes brought new global structures. The most dramatic was the collapse of the formal structures of communism, with the disintegration of the Soviet Union in 1991, symbolically marked by the destruction of the Berlin Wall (between the then separate nations of East and West Germany) in 1989, and consequently the effective disappearance of the Second World, though several countries, including China, Cuba, and North Korea, continued to maintain avowed socialist development strategies. In Europe, formerly socialist states entered a period of postsocialist reconstruction toward privatization.

In other parts of the world, countries formed regional groupings, such as CARICOM in the Caribbean, often designed to improve regional and international trade structures. Small island states grouped themselves into the Association of Small Island States (AOSIS) recognizing their own particular needs which had much to do with environmental matters and exceptionally limited economic diversity. Though the Non Aligned Movement continues to meet, it has had decreasing relevance to the diverse needs, and particularly the political concerns, of the countries it sought to involve. Any sense of significant global blocs was thus dwindling.

A contemporary manifestation of the Third World, with its implication of both failure and vulnerability, is the growing recognition of failed states, that can no longer perform basic functions such as education, security, or governance, usually due to divisive violence or extreme poverty, or more generally control the state, collect state revenue, and police state borders. Within this power vacuum, people fall victim to competing factions, crime, and hunger, and sometimes the United Nations or neighboring states have intervened to prevent humanitarian disaster. However, states fail not only because of internal factors. Foreign governments can also destabilize states by fueling ethnic warfare. The World Bank has identified around 30 'low income countries under stress', the United Kingdom's development agency (Department for International Development) has named 46 fragile states of concern, and the journal *Foreign Policy* has developed an annual global ranking of weak and failing states, using 12 social, economic, political, and military indicators to classify 60 failed or potentially failing states (**Figure 2**). **Table 2** lists the indicators of instability for these countries. The analysis excludes island microstates in the Pacific, though both Nauru and Solomon islands have been seen as failed states.

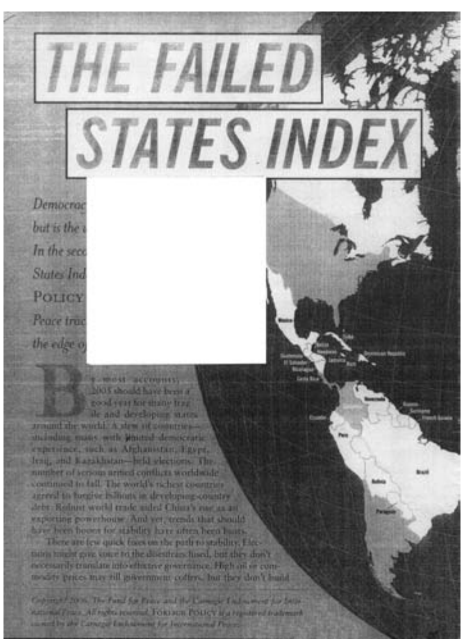

Figure 2 Weak states. Source: *Foreign Policy* (2006). The failed states index. *Foreign Policy* 154, July, 50 59.

Most of the identified failed states, and those considered most at risk, are again in sub Saharan Africa, including Congo, Somalia, Sierra Leone, and Ivory Coast, but also include such countries as Iraq and Afghanistan, and smaller island states such as Haiti and Nauru. In many of these states ethnic divisiveness often indicates irrelevant state boundaries, and several have experienced sometimes continuing internecine warfare, where external intervention has propped up the state. Uneven development is particularly significant in almost all the states that are high in *Foreign Policy*'s failed state index, suggesting that inequality rather than poverty is a critical factor in instability. Similarly important is the criminalization of the state where state institutions are regarded as corrupt, illegal, or simply ineffective. Demographic factors, stemming from refugee movements and environmental degradation, are further problems.

One World

Over time greater subtlety of analysis reduced the value of simplistic and ideological terms such as Third World. Around 200 countries are too complex to be reduced to a simple three or fourfold classification, and within the nominal Third World, there was extraordinary variety in culture, history, and contemporary development. Homogeneity was spurious. No single organization effectively linked them.

Values are crucial to any approach to development, hence development is partly a question of ideology, and is necessarily relative. Assessments of development may be both intuitive and biased, and linked to questions of power. All countries, even the most reclusive and isolated, are interdependent and external links, such as trade and aid, are often critical in influencing national development. This has intensified with new manifestations of globalization, including migration. Moreover, economic and even political power may more obviously reside in the hands of large multinational corporations, rather than countries, suggesting that examining development solely in terms of state issues and values is of limited relevance. Interdependence both imposes costs and enables gains, whether through trade, investment, migration, or other linkages. While in smaller, weaker countries, external influences may be particularly important this does not necessarily demarcate a group of countries with particular characteristics.

While in the post Second World War years, numerous countries had identified socialist models of development as their objectives, almost all countries now exist within a 'one world' structure of 'Western' capitalism, although the means of achieving a more elaborate capitalist

Table 2 nd cators of nstab ty

Rank	Total	Country	Demographic pressure	Refugees and displaced persons	Group grievance	Human flight	Uneven development	Economy	Delegitimization of state	Public services	Human rights	Security apparatus	Factionalized elites	External intervention
1	112 3	Sudan	9 6	9 7	9 7	9 1	9 2	7 5	9 5	9 5	9 8	9 8	9 1	9 8
2	110 1	Dem Rep of the Congo	9 5	9 5	9 1	8 0	9 0	8 1	9 0	9 0	9 5	9 8	9 6	10 0
3	109 2	vory Coast	8 8	7 6	9 8	8 5	8 0	9 0	10 0	8 5	9 4	9 8	9 8	10 0
4	109 0	raq	8 9	8 3	9 8	9 1	8 7	8 2	8 5	8 3	9 7	9 8	9 7	10 0
5	108 9	Zimbabwe	9 7	8 9	8 5	9 0	9 2-	9 8	8 9	9 5	9 5	9 4	3 5	8 0
6	105 9	Chad	9 0	9 0	8 5	8 0	9 0	7 9	9 5	9 0	9 1	9 4	9 5	8 0
6	105 9	Somalia	9 0	8 1	8 0	7 0	7 5	8 5	10 0	10 0	9 5	10 0	9 8	8 5
8	104 6	Haiti	8 8	5 0	8 8	8 0	8 3	8 4	9 4	9 3	9 6	9 4	9 6	10 0
9	103 1	Pakistan	9 3	9 3	8 6	8 1	8 9	7 0	8 5	7 5	8 5	9 1	9 1	9 2
10	99 8	Afghanistan	7 9	9 6	9 1	7 0	8 0	7 5	8 3	8 0	8 2	8 2	8 0	10 0
11	99 0	Guinea	7 5	7 2	8 1	8 4	8 0	8 0	9 1	9 0	8 1	8 1	9 0	8 5
11	99 0	Liberia	8 0	9 3	7 0	7 1	8 6	8 9	7 8	9 0	7 2	7 3	8 8	10 0
13	97 5	Central African Republic	9 0	7 7	8 8	5 5	8 5	8 1	9 0	8 0	7 5	8 9	8 0	8 5
14	97 3	North Korea	8 0	6 0	7 2	5 0	9 0	9 5	9 8	9 5	9 5	8 3	8 0	7 5
15	96 7	Burundi	9 0	9 1	7 0	6 7	8 8	7 8	7 2	8 5	7 5	7 3	7 8	10 0
16	96 6	Yemen	7 8	6 7	7 0	8 2	9 0	7 8	8 8	8 2	7 2	9 0	9 4	7 5
16	96 6	Sierra Leone	8 5	7 9	7 1	8 9	8 7	9 0	8 0	8 0	7 0	7 0	7 7	8 8
18	96 5	Burma	8 9	8 8	9 0	6 0	9 0	7 1	9 2	8 2	9 8	9 0	8 0	3 5
19	96 3	Bangladesh	9 0	5 8	9 5	8 5	9 0	7 0	9 0	7 5	7 8	8 3	8 9	6 0
20	95 4	Nepal	8 5	4 8	9 2	6 0	9 2	8 5	9 2	6 2	9 1	9 0	9 0	6 7
21	94 5	Uganda	8 0	9 2	7 8	5 7	8 4	7 5	3 8	8 0	8 0	8 5	7 9	7 5
22	94 4	Nigeria	8 0	5 9	9 1	8 5	9 0	5 4	9 0	8 3	7 1	9 2	9 0	5 9
22	94 4	Uzbekistan	7 7	5 8	7 5	7 5	8 1	7 0	9 3	7 0	9 3	9 1	9 1	7 0
24	92 9	Rwanda	9 5	7 0	9 0	8 2	7 2	8 0	8 7	6 9	7 7	5 0	8 9	6 8
25	92 4	Sri Lanka	8 0	8 2	9 1	6 7	8 0	5 7	8 6	7 0	7 2	8 5	8 9	6 5
26	91 9	Ethiopia	9 0	7 6	7 0	7 5	8 5	8 0	7 6	6 2	8 0	7 5	8 7	6 3
27	91 8	Colombia	7 0	9 1	7 4	8 5	8 5	3 2	8 7	6 5	7 6	9 0	9 2	7 1
28	90 3	Kirgistan	8 0	6 6	7 0	7 5	8 0	7 5	8 3	7 3	7 9	8 3	7 9	6 0
29	89 8	Malawi	9 0	6 0	6 0	7 0	8 8	8 8	8 0	9 0	8 0	5 5	6 7	7 0
30	89 7	Burkina Faso	9 0	5 9	6 5	6 6	8 8	8 2	7 8	8 4	6 5	7 6	7 7	6 7
31	89 5	Egypt	8 0	6 0	8 5	6 0	8 0	7 0	9 0	7 3	8 0	6 5	7 7	7 5
32	89 2	ndonesia	7 5	8 2	6 3	8 3	8 0	6 8	6 7	7 2	7 5	7 5	7 9	7 3
33	88 6	Syria	7 0	7 1	8 0	6 8	8 9	6 5	9 0	5 5	9 8	7 5	7 1	6 2
33	88 6	Kenya	9 0	7 1	6 7	8 0	8 0	6 8	7 3	7 2	6 9	7 0	7 6	7 0
35	88 5	Bosnia and Herzegovina	6 5	8 5	8 6	6 0	7 3	6 2	8 1	5 8	5 3	7 5	8 7	10 0
36	88 4	Cameroon	6 5	6 8	6 5	8 0	8 7	6 0	8 5	8 0	7 2	7 6	7 9	6 7
37	88 3	Angola	8 0	8 5	6 3	5 0	9 0	4 9	8 8	7 6	7 8	6 8	8 0	7 6
37	88 3	Togo	7 0	5 8	6 0	6 5	7 5	8 0	8 7	8 1	8 1	8 1	7 8	6 7
39	87 9	Bhutan	6 0	8 1	7 0	6 7	9 0	8 0	8 4	6 0	8 6	5 0	8 4	6 7
39	87 9	Laos	8 0	5 9	6 3	6 6	5 9	6 5	7 9	8 0	8 2	9 0	8 9	6 7
41	87 8	Mauritania	9 0	5 9	8 5	5 0	7 0	7 8	7 1	8 2	7 1	7 6	7 9	6 7
42	87 7	Tajikistan	7 0	6 6	6 2	6 5	7 4	6 8	8 9	7 5	8 6	7 5	8 7	6 0
43	87 1	Russia	8 0	7 2	8 0	7 0	8 0	3 7	8 2	6 9	9 1	7 5	9 0	4 5
44	87 0	Niger	9 4	4 3	8 5	6 0	7 2	9 0	7 9	8 5	6 5	6 7	6 0	7 0

45	86 l	Turkmenistan	7 0	4 2	5 2	6 0	7 2	8 0	9 1	7 2	9 7	8 5	8 0	6 0
46	85 4	Guinea-Bissau	7 0	4 9	5 5	7 0	9 3	7 4	7 8	8 0	7 9	7 5	6 5	6 6
47	85 0	Cambodia	7 5	6 5	7 0	8 0	7 2	6 0	7 8	7 5	6 9	6 7	7 5	6 4
47	85 0	Dominican Republic	7 8	7 0	6 5	8 5	8 0	6 0	6 2	8 0	7 1	7 0	7 4	5 5
49	84 6	Papua New Guinea	8 0	2 5	8 0	8 0	9 0	7 0	7 8	8 0	6 1	7 0	6 7	6 5
50	84 5	Belarus	9 0	5 1	5 5	3 5	8 5	6 3	9 0	7 5	7 3	6 8	8 0	8 0
51	84 3	Guatemala	8 7	6 0	7 1	6 7	8 0	7 1	7 5	7 1	7 1	7 5	6 0	5 5
52	84 0	Equatorial Guinea	7 0	2 0	6 7	7 5	9 0	4 0	9 0	8 0	8 5	8 3	8 0	6 0
52	84 0	ran	6 5	8 7	6 9	5 0	7 5	3 0	8 1	6 1	9 1	8 0	8 8	6 3
54	83 9	Eritrea	8 0	7 2	5 4	6 0	6 0	8 0	8 0	7 3	6 8	7 2	7 5	6 5
55	83 8	Serbia and Montenegro	5 7	8 5	8 6	5 5	8 0	6 5	7 8	5 0	5 6	6 5	8 6	7 5
56	82 9	Bolivia	7 5	4 0	7 0	7 0	8 8	6 2	7 0	7 8	6 7	6 5	8 4	6 0
57	82 5	China	8 5	5 1	8 0	6 6	9 2	4 5	8 5	7 3	9 0	5 5	8 0	2 3
57	82 5	Moldova	7 0	4 7	7 3	8 0	7 5	7 5	7 4	7 0	6 8	5 5	6 8	7 0
59	82 4	Nicaragua	6 5	5 5	6 4	7 1	9 0	8 5	7 3	7 2	5 7	6 5	7 0	5 7
60	82 2	Georgia	6 0	6 8	7 4	6 1	7 0	5 5	7 7	6 3	5 6	8 1	7 1	8 6

development necessarily vary considerably. However, they are often subsumed in the increased universality of policies based on neoliberalism and the primacy of markets, sustained by external interventions from such international institutions as the World Bank, the Asian Development Bank, and the International Monetary Fund.

At the same time there are significant differences within states as, on the one hand, capitalism and material success become evident even in parts of nominally so cialist states, such as Shanghai in China (sometimes re ferred to as 'First Worlds within the Third World') while on the other, pockets of material deprivation and poverty exist even in the wealthiest states. Uneven development is universal, as global complexity increases, evident particularly in significant rural–urban distinctions in some less developed countries.

Further, enormous cultural distinctions between, and within, countries and regions remain important. These cultural distinctions suggest that critical global divisions are now rather more cultural, as once expressed in Samuel Huntington's concern over the "clash of civil isations," rather than in terms of broad economic orien tation. Once again, culture and politics may be more important than economics in shaping crude global divisions.

Ultimately the Third World has largely become a journalistic handle that may effectively focus attention on parts of the world that experience high levels of multiple deprivation, both wide regions such as the Sahel or parts of many nations, such as remote inland Australia, that are not necessarily in the 'conventional' Third World. It is at this partly metaphorical level that the concept can be most effective and dramatic rather than in now archaic attempts to distinguish world regions, from a primarily Eurocentric perspective. In countries that see themselves labeled as Third World the term is regarded as outdated, colonialist, inaccurate, and essentialist, and a denial of any notion of change and progress. The term represents a particular form of Orientalist discourse that has come to be known as 'developmentalism', whereby countries supposedly different from those of the West become essentialized, homogeneous entities that should evolve toward the status of the countries of the West.

See also: Aid; Dependency; Development I; Development II; Eurocentrism; First World; Neoliberalism; Neoliberalism and Development; Orientalism; Second World; Transnational Corporations in Developing Countries.

Further Reading

Berger, M. (1994). The end of the Third World. *Third World Quarterly* 15, 257 275.

Corbridge, S. (1999). Development, post development and the global political economy. In Cloke, P., Crang, P. & Goodwin, M. (eds.) *Introducing Human Geographies*, pp 67 75. London: Edward Arnold.

Dodds, K. (2005). *Global Geopolitics. A Critical Introduction*. Harlow, UK: Pearson.

Escobar, E. (1995). *Encountering Development The Making and Unmaking of the Third World*. Princeton: Princeton University Press.

Foreign Policy (2005). The failed states index. *Foreign Policy* 149 July August, 56 65.

Lee, R. (2000). Third World. In Johnston, R., Gregory, D., Pratt, G. & Watts, M. (eds.) *Dictionary of Human Geography*, pp 827 830. London: Blackwell.

Power, M. (2003). *Rethinking Development Geographies*. London: Routledge.

Relevant Websites

http://www.shambles.net
Third World Countries.

http://www.twnside.org.sg
Third World Network.

www.undp.org
United Nations Development Programme.

www.worldbank.org
World Bank.

Third World Cities

S. Lloyd-Evans and R. B. Potter, University of Reading, Reading, UK

Glossary

Desakota A term commonly used in Southeast Asia to refer to zones that are neither rural or urban in character.

Dualism A theory that characterises Third World cities as comprised of two distinct economic sectors (traditional and modern).

Extended Metropolitan Region (EMR) A term given to rural areas surrounding cities that are adopting urban economic characteristics. These boundary areas may extend up to 100 km from the city core.

Informal Sector A collective name given to all remunerated work that is not recognized, regulated, or protected by existing legal or regulatory frameworks and all nonremunerated work undertaken within enterprises.

International Labour Organization Founded in 1919, the ILO is a specialized agency of the United Nations that deals with labor issues.

Kampungs Traditional villages in Southeast Asia.

Mega-cities A term that defines cities with population in excess of 8 million inhabitants.

Nongovernmental Organizations A term given to a heterogeneous group of organizations that are not part of the structure of government. It is commonly used to refer to voluntary sector, civil society, grassroots, and nonstate organizations.

South or Global South Collective name given to the poorer countries of the world by the Brandt Commission. These are generally situated to the south of the globe. This term is now used in preference to the then common collective term the 'Third World'.

Third World A term coined in the 1960s to pinpoint those, primarily, former colonial countries that were emerging as independent states, in contrast to the free-market First World and the state-socialist Second World.

Third World City A collective name given to cities in the Third World to differentiate them from cities in the so-called First World.

United Nations Human Settlements Programme (UN-HABITAT) Established in 1978, it is the United Nations agency concerned with human settlements.

World Cities A concept that refers to cities that undertake major economic, social, and political roles in the global accumulation of capital.

Introduction

The Validity of the 'Third World City' Concept

Third World cities have presented many challenges to those who have tried to explain their distinctive nature. The term 'Third World city' has been widely applied in urban and development geographies since the 1960s, and stems from a notion that there are distinct differences between cities in the developing and developed worlds. Its heyday as a popular geographical term was between the 1960s and 1980s, before criticisms were voiced from postcolonial writers attempting to deconstruct what they saw as an essentialist ideology based on the 'othering' of urban forms in the so called 'Third World'. However, geographical interest in 'Third World' or 'developing world' cities continues as a result of heightened concern over the immense problems associated with unequal development and urban poverty in the twenty first century, and as a backlash to the Western centricity of much contemporary urban geographical discourse.

Although Third World cities are dynamic and complex urban systems, legacies of colonialism, the sheer pace and scale of urbanization, and processes of unequal global development have led many commentators to suggest that there is some convergence in which social, economic, and political processes produce similar spatial patterns in housing, employment, and urban form. Many geographers believe that although each city is located in specific historical, cultural, and geographical space, Third World cities are shaped by a unique combination of local, national, and global processes that differ from elsewhere. While appreciating the contradictions that arise from generalization, it is argued that Third World cities reflect unequal urban development and poverty on a scale that is unprecedented in the developed world. Furthermore, structural inequalities visibly manifest themselves in different ways in terms of urban form, economic activity, and sociospatial diversity when compared to cities of the Global North.

On the other hand, some critics have argued that growing diversity between cities in Africa, Asia, Latin America, and the Caribbean has served to weaken the validity of the 'Third World city' concept. Is it still possible to identify common processes in so called 'Third World cities' that make them distinct? Indeed, attempts to categorize all cities in the Global South as similar can be charged with adhering to neocolonial traditions that essentialize Third World cities as backward and poor 'others'. Hence, the extent to which cities in the 'developing' or

'Third World' are different from 'Western' or 'developed world' cities continues to be widely contested.

Recent Trends in Third World Cities

The world is entering an historic phase of urban transformation, and the latest projections suggest that over the first two decades of the twenty first century, the number of city dwellers in the developing world will increase from 3 to 4 billion. For the first time in history, the world will be predominantly urban, with most population growth occurring in towns and cities in the Global South. According to recent projections, 61% of the world's population will be living in cities by 2030, with cities in the South growing four to five times faster than in the developed world.

It is predicted that rapid urbanization across Africa, Asia, Latin America, and the Caribbean will see 25 out of 30 of the world's largest cities situated in the Global South with mega cities such as Mumbai, Delhi, and Mexico City heading the league tables by 2030. In fact, some developing world cities are growing so rapidly that they are merging together to form large compound urban systems referred to as 'mega cities', a term that defines cities in excess of 8 million inhabitants. A predominant reason for this rapid growth has been migration by people in search of improved livelihood opportunities, standards of living, and services. The sheer scale of this growth presents immense challenges to local and national governments.

The majority of this new growth will be absorbed by cities of less than 1 million inhabitants in the developing world where comparatively little research has been undertaken. In 1950, there were only 31 million cities in developing countries; by 1985, there were 146; and today there are around 400, with 75% located in low and middle income countries. In the writings on Third World cities, it is the largest cities such as Sao Paulo, Mexico City, Mumbai, or Lagos that have received most attention. However, socioeconomic conditions in smaller cities, particularly those with a population under 100 000, are frequently far worse than in the mega cities, as they often lack basic services such as piped water, waste disposal, electricity, accessible education, and health services. So, while globalization and economic liberalization may have improved lives within wealthier urban communities, they have generally resulted in greater inequalities and worsened the everyday living conditions for the most vulnerable.

Deconstructing Third World City Research

There is a vast body of literature that has explored different aspects of Third World cities over the last 50 years. Broadly categorized, there are six main literatures that make claims about cities in the Third World, namely those dealing with (1) dualist interpretations of 'Third World city' structure; (2) cities as theaters of accumulation and centers of diffusion; (3) models of 'Third World city' structure; (4) globalization and Third World cities; (5) urban poverty, livelihoods, and sustainable cities; and (6) spatial diversity within Third World cities. Each of these is now briefly considered in turn.

Dualist Interpretations of 'Third World City' Structures

Early debates on cities focused on the structure of Third World urban economies. In the 1950s, it was argued by economists such as Arthur Lewis that Third World cities could be characterized by a dualist structure comprising of two separate economies: a 'traditional', 'backward', and 'unproductive' economic sector; and an 'advanced', 'modern', and 'productive' sector. The heyday of the dualistic approach was the 1970s, when advocates regarded the traditional economic sector as the major barrier to development and argued that the backward economy would be eradicated once prosperity increased by industrial modernization. Dualist interpretations also provided a framework for the informal and formal sector dichotomy that focused attention on the existence of a large informal sector providing jobs and services for the urban poor. Dualist perspectives have since been criticized for being oversimplified and failing to recognize the ways in which the different economic sectors are inextricably linked.

Cities as Theaters of Accumulation and Centers of Diffusion

There exists an empirically grounded body of 'Third World city' research that stems from the same period, and that focused on the consequences of rapid rural–urban migration. In the 1960s and 1970s, cities and urban growth were largely regarded in the context of development and underdevelopment, and development research was accused of 'urban bias'. The vast body of data collected on the economies and livelihoods of city residents drew attention to poverty and vast disparities in social equity in Third World cities. Research heightened our understanding of the economic structure of cities, the informal employment sector, squatter settlements and slums, environmental conditions, participatory planning, and the lives of the working poor. A distinguishing feature of Third World cities was their variegated appearance with mixed land uses typified by poor spontaneous settlements juxtaposed with high rise commercial clusters. In the 1980s, a useful initial framework devised by Armstrong and McGee to examine the changing role of cities in the developing world identified them as 'theaters of accumulation' and 'centers of diffusion'. Cities were identified as national symbols of modernity, centers of capital accumulation and political power, and responsible for diffusing Western culture and consumption patterns. The extent to which cities in the developing world are converging with or diverging away from the 'West' is still under debate.

Models of 'Third World City' Structure

One way of attempting to identify common processes operating within Third World cities has been through the construction of models of 'Third World city' structures. Since the 1960s, a significant body of work has focused on developing models of Third World cities in order to identify them as distinct from 'First World' cities (see **Figure 1**). These models witness some appreciation of the heterogeneity of city forms across the developing world, while at the same time suggesting that Third World cities are different.

The preindustrial city

In 1960, Sjoberg argued that in origin, Third world cities were quintessentially preindustrial. The concept of the preindustrial city as a distinct urban form identifies three main aspects of land use: the urban center tends to be preeminent over the periphery as elites and administrative power is centred close to the central core with disadvantaged groups occupying the periphery; broad social areas are differentiated into distinct quarters according to wealth, status, or kinship; and there exists a lack of functional differentiation in urban land use patterns, where plots are used for multiple functions such as housing and work. It is widely argued that today's Third World cities do share some common attributes expressed in this model, particularly smaller Asian and African cities, but that processes of industrialization and globalization have transformed preindustrial city structures.

The colonial city

Horvath classified the colonial city as a distinct urban form resulting from the domination of an indigenous civilization by colonial settlers. Colonial cities were seen as political, military, economic, and social entrepots between the colonizers and the colonized. In respect of internal structure, the colonial settlers retained their position through the development of culturally and ethnically distinct quarters. The colonial city model was seen to be particularly relevant in explaining the development of cities such as Delhi with its old and new quarters.

The African city

It has been argued that many African cities conform closely to the preindustrial model, an assumption that underplays the postcolonial development of African cities. The 'African city' model was developed by the United Nations in 1973, and was based on an indigenous core and the organization of different ethnic groups through density gradients in a pattern that ascribed low density housing to the colonial elites and high density to indigenous populations. The most extreme form of social segregation and economic division could be seen in South Africa's 'apartheid cities'. Criticisms of the African city model center on the postcolonial transformations that have restructured African contemporary cities.

The Southeast Asian city

Terry McGee was the first to identify Southeast Asian cities as distinct and unique within the 'Third World city' paradigm. Of great importance was the mixed zoning of residential areas between modern and traditional types of housing and social structure. For example, in Jakarta, spontaneous as well as traditional villages known as *kampungs* occur throughout the city. Although there are zones of squatter settlements on the periphery, new industrial zones and agriculture are also located in a semirural periphery known as *desakota* zones. *Desakota* regions, which are perceived to be neither 'urban' nor 'rural' in any conventional sense, are transforming rural–urban interactions, rural livelihoods, and lifestyles in Southeast Asian cities. In this way, the core, inner, and outer zones typical of contemporary Southeast Asian cities are seen to constitute an Extended Metropolitan Region (EMR). Today, just how fast Bangkok, Jakarta, and Manila are growing depends on where you draw the city boundaries. EMRs have been mostly examined in the Southeast Asian context where rural–urban interaction has produced a blurring of traditional rural–urban boundaries with rural economies and lifestyles becoming increasingly diverse. While some scholars have typified Southeast Asian cities as 'Third World', others have argued that globalization had made the paradigm of the 'Third World city' obsolete in Southeast Asia.

The Latin American city

Primacy, the prominence of mega cities, and rapid growth due to rural–urban migration are all typical of Latin American cities. According to Griffin and Ford, the dominant characteristic of the Latin American city is a commercial spine or sector which extends from the Central Business District and which acts as the focus of the city's more important economic, social, and cultural amenities, and a substantial proportion of high income residences. Away from the spine, there is a series of concentric zones:

- a zone of maturity, which refers to residential areas that once housed the elite but have filtered down the social scale;

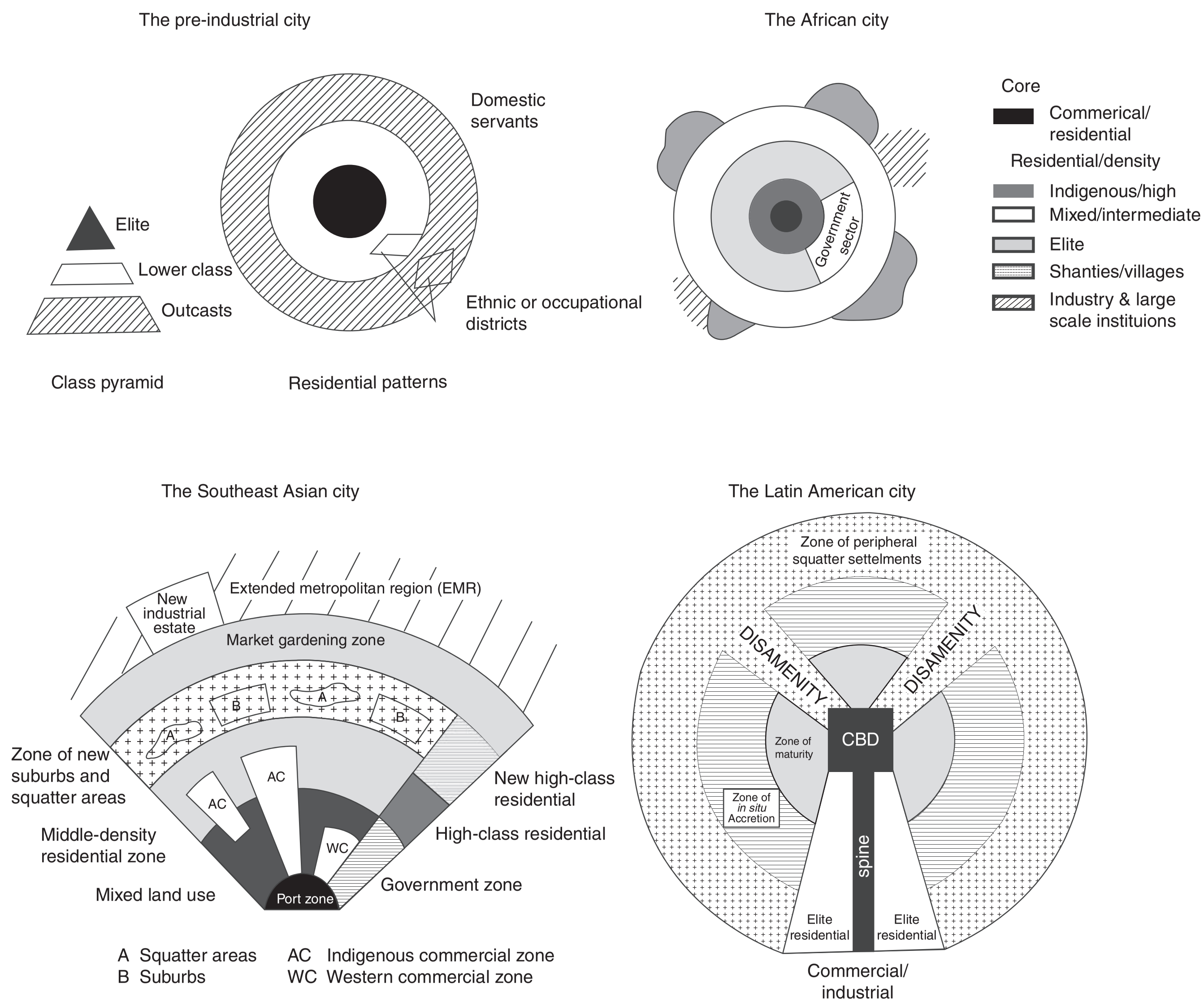

Figure 1 Mode s of Th rd Wor d c ty structure. Source: Adapted from Potter, R. B. and L oyd-Evans, S. (1998). *The City in the Developing World*. Har ow: Longman.

- a zone of *in situ* accretion where there is constant change as residents move in and out; these are zones of gradual improvement where self built housing has consolidated over time;
- a zone of peripheral squatter settlements; and
- zones of disamenity that refer to slum areas which are unlikely to be upgraded such as the *favelas* of Rio de Janeiro.

Critics of the Latin American model highlight the fact that since the 1980s, Latin American cities have been restructured by processes of suburbanization of the elites, and extended metropolitan development, in a way similar to cities in Southeast Asia. Mexico City provides an excellent illustration of how the vast urban subsystem is now encompassing the population of Toluca.

Most spatial models are now outdated and criticized for being oversimplified and lacking in explanation of what exactly it was that was being modelled. Although models have been useful in highlighting processes, they give preeminence to structure and form rather than the social processes and groups that make up cities. Since the 1980s, many developing countries have undergone structural adjustment, liberalization, and the opening of their economies and such processes have had fundamental repercussions on city forms. Indeed, traditional paradigms fail to acknowledge the 'First World' elements of cities and there is a need for new debates that integrate 'Third World' cities within First World debates.

Globalization and Third World Cities

Globalization has produced a changing urban geography in the developing world and the internal structures of cities must now be understood in relation to their evolving roles in the wider global economy. The world city literature, which explores the development of a global hierarchy of cities that form the basic networks of the global economy, has viewed some of the most prominent Third World cities, such as Singapore, as playing key roles in linking developing countries with global commodity markets. As cities are dynamic entities, undergoing contact transformations according to the prevailing macroeconomic and social conditions, it is not surprising that the question of how globalization affects cities in the developing world has received attention. Scholarly research on the impacts of globalization on Third World cities has been limited, partly as a result of the Western bias of globalization debates. Commentators have highlighted a gap between contemporary global theory and the empirical studies that characterized 'Third World city' research in the 1960s and 1970s.

There are common perspectives that argue that as a result of globalization, Third World cities are becoming more like their Western counterparts, particularly in respect of their patterns of consumption, popular culture, and the built environment. New urban economies based on Foreign Direct Investment (FDI) and global capital have created new urban spaces for information technology services which are replacing the traditional manufacturing functions of Southeast Asian cities in particular. The new international division of labor, in particular, the offshoring of data processing and call centers, has profoundly impacted upon internal migration patterns and the labor markets in cities like Mumbai.

Other scholars argue that global cities are not constituted by corporate power alone, as they are also home to large concentrations of immigrants and undervalorized 'others', such as street children and the homeless. Concepts of 'transnationalism' and 'ethnoscapes' that identify cities as constituted by flows of people articulated within specific local urban geographies, have helped scholars move towards a more postcolonial understanding of the processes shaping cities in the developing world.

While this literature has highlighted the important role played by some mega cities in Africa, Asia, and Latin America, the majority of cities in the developing world have been ignored. Globalization is highly uneven with few African cities benefiting from FDI in manufacturing and services, unlike those in Asia, Latin America, and the Caribbean. Smaller cities that cannot attract FDI may find it increasingly hard to compete for national resources and this is likely to result in further geographical differentiation in the twenty first century between large wealthier Third World cities and poor cities with smaller resources.

Urban Poverty and Livelihoods: Toward Sustainable Cities

The growing concern with Third World cities and urban poverty has resulted in a plethora of academic studies and policy reports from international organizations such as UN HABITAT the World Bank, and nongovernmental organizations (NGOs). In the recent past, poverty reduction strategies have tended to focus on rural areas but the growing concern for urban poverty has focused much attention on the internal and external processes affecting Third World cities and their residents. The challenges facing these cities are enormous, given the sheer number of people requiring housing, services, opportunities for making a living, good governance, and civil society. It has been argued that there are four key underlying reasons for urban poverty: lack of access to basic services such as health, water, and education; lack of decent work and livelihood opportunities; a lack of power by local communities to access their rights and entitlements; and poor governance by city authorities.

Central to these debates is a growing body of litera ture that examines the relationship between sustainable environmental development, livelihoods, governance, and civil society. Particular environmental areas of concern include the provision of safe water supplies, sanitation, waste and garbage treatment, safe shelter and work en vironments, infrastructure and good governance, and the eradication of disease. A number of initiatives, including the World Bank's Brown Agenda and the Millennium Development Goals have attempted to prioritize en vironmental problems in Third World cities. They dis cuss the extent to which pro poor solutions can help build 'liveable cities', and examine the roles undertaken by new partnerships between NGOs, the state, and the private sector to encourage local civil society. A range of contemporary solutions that aim to integrate poor resi dents into what the UN defines as 'the fabric of urban society' include a focus on small business and the in formal sector, recent debates over legal frameworks, newly emerging City Development Strategies (CDSs), policy, and good governance. The commitments made by governments at the Millennium Summit in 2000 have not resulted in improving the living conditions among the majority of the urban poor. It has become increasingly recognized that urban poverty is different from rural poverty, and a better understanding of its nature is re quired in order to tackle the problems facing Third World cities in the twenty first century.

Spatial Diversity in Third World Cities

While conceptual frameworks and models have served to enhance our understanding of transformation in the structures of Third World cities, contemporary research has highlighted the important spatial aspects of diversity within them. Focusing on issues of equity, urban well being, and urban social interaction, research has drawn attention to how people live through an appreciation of local everyday geographies and livelihood trajectories.

Central to understanding Third World cities is the appreciation of the marked socioeconomic inequalities in the urban experiences and lifestyles of the diverse com munities living within their boundaries. Almost all cities include elite neighborhoods and modern commercial centers that are well served by public infrastructure, education and health services, retail and leisure facilities, as well as poor neighborhoods characterized by slums and lacking in infrastructure and opportunities for decent livelihoods. These positive and negative elements of city life are intermixed as cities are sites of intense social and economic interactions and opportunities in terms of so cial networks, education, and jobs as well as disbenefits of congestion, underemployment, or crime. Some of the most important issues within contemporary Third World city debates include housing, livelihoods and the informal sector, gender, and cities for children.

Urban poverty and housing

Housing is a key economic resource for low income urban residents. Third World cities have traditionally been characterized by a substantial proportion of their populations living in slums, shanty towns, and squatter settlements. As cities grow, so do their slum populations. Slums already account for 1 billion of the world's urban population, with the figure set to rise to 1.5 billion by 2020. Between 1990 and 2001, 65 million people moved into slums in Africa with 72% of its urban population living in slums. If achieved, the Millennium Develop ment Goal 'to improve the quality of the life of 100 million slum dwellers by 2020' will barely scratch the surface (see **Table 1**). The latest UN HABITAT 'State of the World Cities' Report 2006–07 states that slum dwellers in Third World cities are worse off than their rural counterparts.

Shanty towns and squatter settlements are usually defined by the poor materials used to construct some kind of shelter, while squatter settlements refer to spontaneous housing built illegally on land without title to ownership. Although it is generally agreed that slums are physically inappropriate places for communities to live, there has been some controversy over the extent to which squatter settlements can present an answer to the housing problems of Third World cities. For migrants, they have been seen as 'slums of hope', places where new arrivals to the city can adapt to city life and/or centers of citizen solidarity and participatory power. However, spatial segregation limits the life course possibilities for poor urban families and the separation from dynamic service industries, professional workers, and political elites can serve to increase socioeconomic inequalities

Table 1 Geographical variations in slum populations (in millions)

Year	*South Asia*	*East Asia*	*Sub-Saharan Africa*	*Latin America/ Caribbean*	*North Africa/ Middle East*	*Industrialized countries*
1990	199	200	101	111	50	42
2005	276	272	199	134	68	47
2015	345	355	313	153	85	50

Source: UNDP International Poverty Centre (2005). Poverty and the city, *In Focus*, August 2005.

within cities. There is evidence to suggest that informal settlements can become consolidated over time and be come gradually linked to the city's infrastructure and services. Well established squatter settlements often house middle class communities, including teachers and doctors, and they are embedded in local social capital and networks. NGOs have focused on empowering low in come communities to fight for their rights and participate in local planning decisions. Projects such as the Feder ation of Low Income Groups in Mumbai and the Self Employed Women's Association in Ahmedabad have helped raise the political visibility of poor urban resi dents, often women, and helped them to create solutions through participatory 'self help' housing, land tenure, and microcredit projects.

Urban livelihoods and the informal sector

Since the 1960s, research has also highlighted the di versity of economic activity in Third World cities through explorations of the informal and formal sectors of the urban economy. Commentators such as Santos, Hart, and the International Labour Organization (ILO) drew attention to the dualistic structure of Third World cities that they argued was distinct from cities in the developed world. The existence of both modern and traditional economic sectors was recognized, referred to as the 'formal' and 'informal' sectors by Keith Hart in 1973. The informal sector refers to a heterogeneous range of activities, which can be broadly divided into four subsectors (see **Table 2** and **Figure 2**):

1. subsistence activities (e.g., urban agriculture);
2. small scale producers and retailers (e.g., street traders);
3. petty capitalists (those who operate small scale pro duction units that operate outside regulations and often employ casual labor); and
4. the criminal sector (e.g., prostitution, slavery, and drug trafficking).

As such, the informal sector performs a variety of roles within society, encompassing wealth, productivity and inefficiency, and exploitation and liberation. It is largely

Table 2 Urban street vendors in developing countries (various years 1990 2000)

Country	*No. of street vendors*	*Women (%)*
Tunisia (1997)	125 619	2
Benin (1992)	45 591	81
Kenya (1999)	416 292	33
India (1999 2000)	3 881 700	14
Brazil (1991)	1 445 866	30
Costa Rica (1997)	13 685	18
Guatemala (2000)	259 203	55
Mexico (2000)	1 286 287	44
Venezuela (1997)	318 598	32

Source: International Labour Office (2002). *Women and Men in the Informal Economy: A Statistical Picture.* Geneva: ILO.

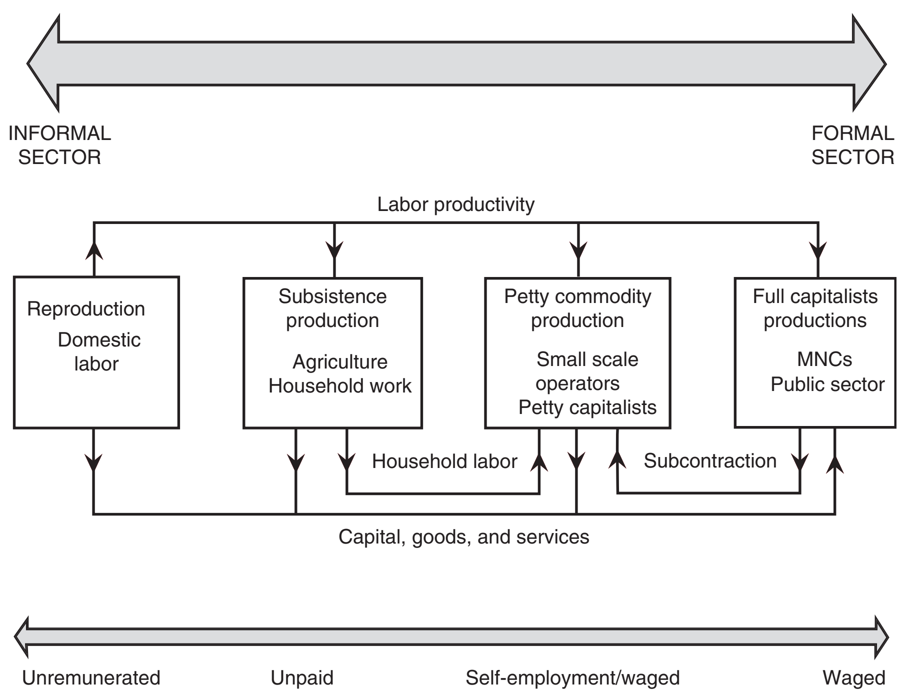

Figure 2 The urban informal sector continuum. Source: Adapted from Potter, R. B. and Lloyd-Evans, S. (1998). *The City in the Developing World*. Harlow: Longman.

regarded as a continuum. Due to high unemployment rates and a lack of decent formal sector jobs, many cities are characterized by large informal sectors that provide a diverse range of livelihood opportunities for low and middle income communities. Recent figures suggest that around 60% of urban residents are employed within the urban informal economy of many Third World cities, with street vending constituting important income earning opportunities (see **Table 1**). The informal sector is a particularly important provider of livelihoods for women and children, many of whom are concentrated in the lower echelons of the economy as a result of unequal access to decent work, education, and resources.

Gender and urban space

The gendering of urban space is central to understanding sociospatial diversity in Third World cities. The in creasing poverty of many women witnessed in the last 30 years has been attributed to their unequal position within the labor market, their subordination under patriarchal social systems, and the undermining of their status and power through capitalism. Structural adjust ment programs and global economic trends have ex acerbated all these factors. Thus, cities are influenced by gendered and racialized processes that shape women and men's differential access and entitlement to work, shelter, capital, and resources.

External forces have affected women's ability to meet their basic needs and the gender division of labor in the urban economy is such that women are frequently engaged in low paid, exploitative, and arduous work, while also maintaining their households and communities. The 'triple burden' of reproduction, production, and community par ticipation often stretches women's ability to cope in a crisis. Transformations in the global cheap labor economy have resulted in a feminization of rural–urban migration pat terns where young women are increasingly becoming new urban migrants. Although migration can be empowering for women as it can accord economic independence and educational opportunities, it can also further exploit their labor power and stretch their coping mechanisms as they adjust to city life. In the past, the literature on Third World cities has been criticized for being gender blind and urban policies accused of 'male bias'. But this imbalance is starting to be redressed by enlightened research that highlights the importance of empowering women within urban develop ment programs.

City spaces for children

Children in urban poverty are unquestionably at high risk. Cities are places of social turmoil for children with the capacity of adults to function as caregivers dwindling in many urban neighborhoods. One of the most hotly debated topics related to Third World cities has been the increasing incidence of child labor and street children in cities of the Global South. In the 1990s, heightened concern over the future welfare of millions of the world's poorer children largely developed from widespread media coverage of child related issues, such as the murder of Brazilian street children by police death squads, and increased documentation on child work by NGOs and international institutions such as the United Nations Children's Fund (UNICEF), the World Bank, and the ILO. The ILO estimates that 218 million children aged 5–17 can be classified as child laborers with the 5–14 age group constituting 166 million.

Child labor takes on many forms, from paid work in factories and other forms of waged labor like street trading, to bonded domestic labor and prostitution. The incidence and visibility of child labor are more acute in the city due to increasing household vulnerability and the availability of waged work in urban industries such as glassware, carpet weaving, and rag picking. Societal and public views on child labor and related issues will depend on the meanings people attach to public 'spaces' and what they see as unnatural and inappropriate spaces for chil dren. In Latin America, millions of street children are perceived as criminal and a threat to family and social order. New geographies of childhood have highlighted the heterogeneity of childhood experiences in Third World cities and research on 'other childhoods' has sought to critique constructions of childhood that dom inate the understanding of child labor in the Global South.

The United Nations' Convention on the Rights of the Child has been identified by many as the most appro priate framework for creating cities that are supportive of children. This movement has resulted in a more child focused development agenda which is linked to the focus of the Millennium Development Goals on the provision of 'decent work' for youth and universal primary edu cation, and the promotion of the largest global anti child labor program (International Programme on the Elim ination of Child Labour IPEC) by the ILO since the 1990s.

Conclusions – Cities within Cities

The extent to which there are 'Third World cities' de pends on whether their characteristics are distinctive from today's cities in the developed world. Despite claims that globalization is leading to homogenization between so called 'First' and 'Third World cities', there is com pelling evidence from the latest 'State of the World's Cities' report' by UN HABITAT to suggest that there are, in fact, two cities within one city, comprising one entity that has all the benefits of urban living and another where the urban poor live in worse conditions than their rural counterparts. Furthermore, the report also suggests that the world's billion slum dwellers are more likely than ever to die earlier, experience more hunger and disease, and have less chances of gainful employment than urban residents living outside of slums. Such claims are not new and stem in part from dualistic interpretations of the urban economy mentioned previously, but they do highlight the need to reappraise 'Third World city' de velopment in an era of rapid globalization and urban ization. If nothing else, the existence of marked social inequalities in terms of access to wealth, urban services, employment, and adequate shelter tell us that Third World cities are not following exactly the same trajec tories as Western cities. Quite simply, the numbers of poor people involved and the socioeconomic transfor mations that are required are of a different magnitude.

See also: Mega-Cities; Planning, Urban; Third World.

Further Reading

Armstrong, W. and McGee, T. G. (1985). *Theatres of Accumulation: Studies in Asian and Latin American Urbanization*. New York: Methuen.

Bartlett, S., Hart, R., Satterthwaite, D., de la Barra, X. and Missair, A. (1999). *Cities for Children: Children's Rights, Poverty and Urban Management*. London: UNICEF and Earthscan.

Devas, N., Amis, P. and Beall, J. *et al*. (2004). *Urban Governance, Voice and Poverty in the Developing World*. London: Earthscan.

Drakakis Smith, D. (2000). *Third World Cities* (2nd edn.). London: Routledge.

Gilbert, A. (1994). Third world cities: Poverty, employment, gender roles and the environment during a time of restructuring. *Urban Studies* 31, 605 633.

Grant, R. and Nijman, J. (2002). Globalization and the corporate geography of cities in the less developed world. *Annals of the Association of American Geographers* 92(2), 320 340.

International Labour Office (2002). *Women and Men in the Informal Economy: A Statistical Picture*. Geneva: ILO.

Montgomery, M. R., Stren, R., Cohen, R. and Reed, H. E. (eds.) (2004). *Cities Transformed: Demographic Change and Its Implications in the Developing World*. London: Eathscan.

Potter, R. B. and Lloyd Evans, S. (1998). *The City in the Developing World*. Harlow: Pearson Prentice Hall.

Robinson, J. (2002). Global and world cities: A view from off the map. *International Journal of Urban and Regional Research* 26(3), 531 554.

UNDP International Poverty Centre (2005). Poverty and the city, *In Focus*, August 2005.

Relevant Websites

http://www.citiesalliance.org
Cites Alliance.

http://www.eldis.org
Eldis.

http://www.ilo.org
International Labour Organization (ILO).

http://www.unifem.org
United Nations Development Fund for Women (UNIFEM).

http://www.unhabitat.org
United Nations Habitat.

Thrift, N.

U. Strohmayer, National University of Ireland, Galway, Republic of Ireland

Thrift, Nigel (1949–)

The task of critically assessing the contribution an indi vidual has made to the rich tapestry that informs the practices of a particular discipline is no easy endeavor. Amalgamating the personal with the theoretical, such an undertaking all too often makes subtle demands on both author and reader while at the same time leaving little space for nuance. Hence, biographies or biographical sketches often tend to obscure as much as they help to illuminate. Accepting these limitations, we can access the contribution of Nigel Thrift to geography more con fidently. The relevant key biographical cornerstones can be summarized relatively quickly: born in Bath, Somerset in 1949, Thrift received his BA from the University of Wales in Aberystwyth and his PhD from the University of Bristol, before taking up a research post at the Aus tralian National University in Canberra. Subsequently, Thrift returned to Wales to pick up his first teaching post, this time to that once hotbed of creative energy within UK geographical practices, the Department of Geog raphy at the University of Wales in Lampeter. In 1987, he joined the Department of Geography at Bristol Uni versity, where he became a Professor in 1990. The new millennium saw Thrift – eager to accept challenges of a more administrative kind, successfully devoting his con siderable energy first to reinvigorating 'geography' at the University of Oxford (as 'division head'), before accepting the post of Dean of Life Sciences at the same university. Since 2006, Thrift has been the Vice Chancellor of the University of Warwick (the equivalent, in US and Irish terms, of a University Presidency); in this latest capacity, Thrift is thus one of few human geographers to head an institute of higher education (**Figure 1**).

Hidden beneath these raw places and dates is one of the more involving and interesting intellectual trajec tories within the world of human geography of the last half century. Like many of his generation, Thrift appears to have been driven to act by what he believes in, be it in academic or administrative contexts. He furthermore shares with those interested in 'progressive' or 'emanci patory' forms of geographic enquiry the double bind that came to characterize the post war generation: first, the desire to distance radical thought and praxis from the legacy of Stalinist aberrations, and, second, the ongoing struggle with structuralist temptations in the provision of geographical explanations. If the former characterized the 1960s and 1970s, the latter has been central to geo graphical discourses in the 1980s and 1990s.

In addition, his work has also been motivated by a genuine desire not to abandon theory where theory ar guably matters most; in using it to shed light on material conditions of human existence, Thrift has been nothing short of prolific in his desire to dismantle any pre conceived borderlines between 'theory' and 'practice'. Both his writing and his administrative engagements are entirely consistent with this principled stance. Unlike other, no less 'visible' geographers of his and the pre ceding generation, however, Thrift has not been associ ated with the consistency of a lifelong project or set of beliefs. His writing is thus a more adaptable undertaking compared, say, to the writing of David Harvey, Yi Fu Tuan, or Denis Cosgrove.

Thrift was making his entrance onto the world of academia with a number of publications on 'time geog raphy', including the decisive three volume edited work *Timing Space and Spacing Time* (1978), he was sub sequently to publish the seminal theoretical assessment of time geography in the launching issue of *Environment and Planning D: Society and Space* (1983) – after having been centrally involved in the founding of this journal. For

Figure 1 Nigel Thrift.

many writing in geography today, the existence of this journal sums up what is one of the more lasting legacies of his work within the discipline: the creation of spaces for theoretically informed work beyond the narrow confines of what others would and did deem to be 'proper geography.' The 'biographical spaces' of personal de cisions and chance moments alluded to above here be come entwined in a complex manner that in turn creates conditions of possibility for others; the 'human' and the 'nonhuman' link up and thereby change the scope of opportunity for both. And while this may be truism of sorts, there is no better 'biography' than Thrift's to ex emplify the often breath taking capacity for insights thereby created. In this sense, then, Thrift is perhaps the most effective of 'gatekeepers', working in geography today, both in the editorial sense of the term and in designating someone working both at and on the human/ nonhuman nexus. Thrift's work on 'actor network theory' (about which more further on in this article) thus 'makes sense' on a biographical, theoretical, and professional level, weaving together supervisory roles, mentoring, and international collaborations along the way.

It thus comes as no surprise to anyone that Thrift's name can be found repeatedly among those of the most cited geographers or social scientists of his generation. As would furthermore be expected from the biographical sketch above, Thrift's influence manifests itself in the numerous panels, review committees, and fellowships to which he was elected, appointed, or which were awarded to him over the years. 'Influence' here becomes a tightly woven but flexible network of global proportions cen tered around different places in the UK that he is or was associated with.

Thrift's ostensibly empirical work has centered mainly around the (mostly urban but increasingly virtual) spaces of finance, money, and globalization. Many – in fact most – of the publications that make up Thrift's curriculum vitae are written in a joint capacity with other academics, including Michael Taylor, Paul Glennie, Andrew Ley shon, Kris Olds, Ash Amin, and scores of other well known geographers. A teamworker 'par excellence', Thrift continues to maximize the surface of possible encounters with interesting material; as an all too early assessment of an ongoing project, Thrift's work has been nothing if not original and often eclectic in orientation and breadth.

Take, for instance, his work on the development of the financial services and the anchorage of this work in Thrift's livelong passion for the City of London. Char acteristic for the overall thrust of his work is the active role accorded to 'space' throughout Thrift's writings. This attribution of agency to space differs markedly from an often criticized 'fetishization' of space precisely because 'space' no longer remains a stable entity that could be fetishized according to some inherent quality. In the less specialized context of work on 'cities' (and working with Ash Amin), this particular emphasis translated into the incorporation of such broad ranging terms and processes like 'rhythm analysis', 'flow', 'escape', or 'software' into the analysis of urban phenomena and politics, all adding a decidedly novel quality to 'space' and its (possible) effects.

If this elevation of 'space' to the role of an active ingredient in the construction of geographies – rather than being conceptualized as a passive, Euclidean back drop – is *de rigeur* in today's discipline, readers ought to remember that it was geographers like Nigel Thrift who were instrumentally involved in the turn toward a more dynamic and complex form of geographically informed construction of knowledge. It is hence no surprise to see Thrift's work – although ostensibly classified under the rubric of 'economic geography' – stretching well beyond the bounds associated with the subdiscipline to cover aspects like the role of the media, citizenship, and the role of clocks alongside more traditional themes such as innovation, service industries, or consumption. Just as important, though, is the fact that each of these topical fields is approached consistently without the metaphorical crutches of a stable ontology that would automatically privilege particular routes towards ex planations. Neither 'agency' nor 'structure' – to use poles that have provided such a stable ontology in the past – continue to function in Thrift's work in quite the same way they did in the past. If they 'function' at all, such concepts function momentarily only: within the precise context for which they were coined and in which they are used. Explanation here becomes a performed event that is recognizable as such without affording the means by which to translate or transpose it onto other contexts.

Mention of 'performance' leads us to what is by far the most controversial of Thrift's theoretical endeavors during the last decade and a half: his discovery of 'non representational theory', following from his earlier en gagements with 'actor network theory'. Perhaps the most conventional of responses to the challenges borne of postmodernism and post structuralism, 'nonrepresenta tional theory' – at the same time has proven to inspire a sizable cohort of students to engage with geographical realities in a novel and different way. Key to theorizing the 'nonrepresentational' is the shift from a perceived static world of representations – the main target of the post structuralist critique of language often summarized under the heading 'crisis of representation' – into the realm of performances and their associated rhetorics. The inherent lack of a fixed identity lends to performances that which had rendered traditional anchors for under standing problematic: an identity entirely constructed through temporally and spatially nonfixed events. The orizing such events had been an ambition central to the hermeneutic strand of philosophy that had arguably culminated in the work of Deleuze and Guattari. Thrift's particular accomplishment has been to take these often abstract ideas and concepts and put them to work in material contexts.

While the present circumstances do not afford suf ficient space for a critique of 'nonrepresentational the ory', it is nonetheless important to point toward the unresolved epistemic status of its knowledge claims. The question that must be asked, in other words, concerns the (continued) existence (or not) of a mimetic impulse at the heart of 'nonrepresentational theory': if such theoretical labor is thought capable of 'mirroring' what makes per formance performative and thus capable of representing their identity, such theorizing would not have escaped the epistemological binds of old. But what if it did not set its aim in this manner and onto such a traditional target? Could we – can we – write geographies beyond the es tablished norms that have governed academic modes of reasoning at least since the Enlightenment? Attempting to find an answer to this question is perhaps the lasting legacy of 'nonrepresentational theory'; independent of such future ruminations, however, Nigel Thrift's contri bution to the debate has been outstanding indeed.

See also: Actor-Network Theory/Network Geographies; Financial Exclusion; Non-Representational Theory/Non-Representational Geographies; Performance, Research as; Time and Historical Geography.

Further Reading

Amin, A. and Thrift, N. (2002). *Cities: Reimagining Urban Theory*. Cambridge: Polity.

Carlstein, T., Parkes, D. N. and Thrift, N. (eds.) (1978). *Timing Space and Spacing Time* (3 vols.), *Vol. 1: Making Sense of Time. Vol. 2: Human Activity and Time Geography. Vol. 3: Time and Regional Dynamics.* London: Edward Arnold.

Leyshon, A. and Thrift, N. (1995). Geographies of financial exclusion. *Transactionsof the Institute of British Geographers NS* 20, 312 341.

Peet, R. and Thrift, N. (1989). *New Models in Geography: The Political Economy Perspective*, 2 Vols. London: Unwin Hyman.

Thrift, N. (1983). On the determination of social action in space and time. *Environment and Planning D: Society and Space* 1, 23 57.

Thrift, N. and Leyshon, A. (1992). Liberalisation and consolidation: The single European market and the remaking of European financial capital. *Environment and Planning A* 24, 49 81.

Thrift, N. (1996). *Spatial Formations*. London: Sage.

Thrift, N. (2000). Performing cultures in the new economy. *Annals of the Association of American Geographers* 90, 674 692.

Thrift, N. (2004). Performance and performativity: A geography of unknown lands. In Duncan, J. S., Johnson, N. C. & Schein, R. (eds.) *A Companion to Cultural Geography*, pp 121 136. Oxford: Blackwell.

Thrift, N. (2005). From born to made: Technology, biology and space. *Transactions of the Institute of British Geographers, NS* 30, 463 476.

Time and Historical Geography

M. Kurtz, Open University, Ottawa, ON, Canada

Glossary

Diachronic Characterized by a developmental approach to time, or concerned with changes in an area over an extensive, noncyclical period of time.

Synchronic Characterized by a snapshot approach to time, or concerned with a select area at a given moment or cycle in time with little reference to its antecedents or development.

Temporality A specific construction or framing of time (e.g., circular or linear temporalities). When the word is used in the social sciences in a manner akin to 'spatiality', it especially accentuates the social contexts in which time is ordered and produced in a particular way.

Time Consciousness A conception of time held by an individual or group. Since the work of E. P. Thompson, it is closely associated with the daily practices, shared experiences, and popular articulations through which one conception of time gained meaning and import.

Time-Geography A method of analysis associated with Torsten Hägerstrand, characterized by the use of diagrams to represent the particular and converging paths of a small number of individuals and objects across a closed unit of space and time.

Time–Space Compression Processes through which the modern world is characterized by a constantly accelerating pace of life under capitalism. It is often understood to be driven by competitive laws of accumulation that require ever-shorter time horizons for production or transportation in order to return a profit on an investment.

Introduction

"'When I saw them lying there, I knew it. That there is not just one time, that there must be different sorts of time, all existing at once.' She was now talking so quietly that I had to lean forward. … 'I want to study it scientifically,' she said. 'We're going to touch time.'" In movies, time is a common motif (think of *Back to the Future* or *Memento*), and in fiction, writers often experiment with sequencing, folding time over on itself to create provocative temporalities and juxtapositions. The novel just quoted, for instance, *Borderliners*, opens with a question from St. Augustine – "What is time?" – which then troubles its young protagonists throughout the book as its author, Peter Høeg, moves the story back and forth in time.

In contrast to novelists, "different sorts of time" have not troubled too many geographers much. There is a wealth of work offering sophisticated analyses of space and spatiality, but to see time and temporality given equal treatment is unusual. In Anglophone contexts, this tendency is poorly explained by reference to the 'nature' of geography, just as it would be mistaken to look to the discipline of history for an abundance of theorizations of social time or studies of different temporalities. Such work appears more often in philosophy, feminist studies, sociology, and anthropology.

The article is organized into three sections. The opening section looks to the historical geographies of time. It constitutes a selective review of work that takes the rhythms, measures, and experience of time as the object of analysis. The subsequent section traces ways that geographers have understood and handled a temporal dimension in recent decades, reviewing common notions of time and temporality in their analyses. The final section explores the difference that one representation of time can make. The analysis assumes that notions of time and the subject of historical geography were both socially constructed. It looks particularly at a 1954 textbook to illustrate how specific understandings of time and temporality were used to distinguish and legitimate historical geography as an academic field.

Historical Geographies of Time

While time and temporalities are not common objects of analysis in the discipline, two geographers have put time consciousness at the forefront of their research: Nigel Thrift and Paul Glennie. Indeed, Thrift has been writing about time for over three decades, and a recent essay he co wrote with Glennie, titled 'Revolutions in the times', is an effective starting point with which to examine contributions to the historical geographies of time offered in the discipline. Their paper is noteworthy insofar as it operates at two levels. First, it disputes a widely held proposition that popular perceptions of time saw rapid change in the Industrial Revolution. In a classic essay from 1967, the historian E. P. Thompson depicts a dramatic shift in the eighteenth century England, from a widespread task oriented sense of time in society to one where the measured times of employment and labor were predominant. For Thompson, the shift was not simply the result of industrialization, as the revolution in time consciousness involved broader cultural changes as well. Glennie and Thrift, however, offer us a different temporal model than the rapid sea change which Thompson describes: less that of revolution, more that of sequential diffusions of new technologies and interrelated practices for keeping adequate time. In short, their nar rative uses a model of time that explicitly challenges Thompson's notion of a singular period of revolution.

At a second and more mundane level, Glennie and Thrift offer a schematic account of time keeping practices in everyday life in England between 1300 and 1800. First they sample the diffusion of mechanical clocks across the country in the late medieval era, as the first devices in select public locations began to replace other means of dividing the day into equal hours for a variety of purposes: vespers, market bells, the dinner hour, appointments, etc. Then the authors trace a technological revolution in watchmaking that enabled greater precision in keeping minutes as well as hours, while arguing that uses for such precision were both highly selective and that needs for it in the sixteenth century may have preceded the technical achievements in minute keeping in the seventeenth cen tury. Lastly, they outline interactions between user groups whose need for accurate clocks and watches varied, and among whom a market for precision instruments leaked downward over the eighteenth century. Throughout, Glennie and Thrift offer an account driven, not by technological advance or by changing dispositions of quotidian practice, but by the circular reinforcements between technologies and practices.

In geography, a more prominent narrative about changes in time consciousness is David Harvey's account of time–space compression. Like Glennie and Thrift, he assumes time is a social construct: different societies produce different conceptions of time, and many people in the modern world accept clock time as an objective fact in daily life. Yet, in *The Condition of Postmodernity*, Harvey asserts that capitalism is the fundamental force shaping current understandings of time. In that regard, he shares some affinity with E. P. Thompson. Harvey sug gests that time consciousness is now marked by rapid and often disorienting accelerations in the pace of life. These follow as communication and transportation industries overcame more frictions of distance, a process that Donald Janelle described in 1968 with his notion of time–space convergence. Harvey extends the idea, arguing that the annihilation of space by time (captured with his image of progressively smaller globes) was the outcome of logics of accumulation that required competitive re ductions in the time it took to turn a profit on an in vestment. While ridden with contradictions, time–space compression – the quick delivery of news and com modities, the sense of a shrinking world, the experience of an accelerating pace of life, and not incidentally, a speed up in the production of new ideas – is seen to be driven fundamentally by the fact that, in capitalist soci eties, time costs money.

In historical geography, there is now a related body of work on the creation of 'standard time' in the nineteenth century when the coordination of telegraph and railway operations was a concern across a number of nation states. There is also a body of scholarship on the history of navigation, time, and estimation procedures for lon gitude. Others have written about the delivery of Royal Mail, tracking increases in the speed of communication in early modern England. Yet others have traced the unequal access to standard mail among different classes and the creativity with which postage was circumvented. The last topic serves to complicate the uniform trajectory of Harvey's notion of time–space compression, and it can be used to argue, like Doreen Massey, that this 'speeding up' has been uneven as various social groups are situated in very different ways relative to spatial connections and temporal accelerations. Massey calls up a memorable contrast in this regard: consider at one end, a male ex ecutive on a jet between London and Tokyo, and at the other, a woman in sub Saharan Africa who spends hours each day collecting water on foot.

Among the more elaborate empirical responses to the time–space compression thesis is the work of Jeremy Stein. He records the changing experience of time in the textile town of Cornwall in the nineteenth century On tario while arguing that the drama of time–space com pression is often exaggerated by scholars who draw from elite accounts. As he traces the incremental reduction of travel time between Montreal and Cornwall, he suggests that the majority of the town's residents probably felt the benefits of time–space compression only gradually and indirectly through greater access to outside goods, at the same time that communication technologies increased commonplace contrasts between the metropolitan cen ters with their new capital intensive technologies and the relative 'backwardness' of other areas. Stein also finds that the temporalities of social discipline in Cornwall by passed a wholesale revolution: factory rhythms inter sected with other schedules as a factory manager, for instance, adjusted work schedules to accommodate existing religious routines. Multiple routines and times continued to coexist in the Ontario town, suggesting that accounts of time–space compression elsewhere may be overdrawn as well.

This points to a general issue in the geographies of time and temporality: if time is held to be a social con struct, and if the experience of time varies across dif ferent societies (much as Harvey suggests), then some geographers hold that the means by which a particular sense of time is constructed, and how it comes to frame common understandings and practices, are also multiple and dynamic. The editors of *Timespace: Geographies of Temporality*, Jon May and Nigel Thrift, make such an argument. They outline four domains over which a sense of time can develop (1) natural rhythms and cycles, such as seasons; (2) systems of social discipline, such as waged labor; (3) its relation to devices and technologies, such as clocks; and (4) its relation to new conceptualizations and influential texts. Whereas Harvey's account privileges a system of social discipline, contributors to the volume trace the varied impact and reach of these four domains upon time consciousness in different places. The result is less a singular sense of time that comes to stretch across space, more that of multiple temporalities in which the quality and importance of each domain varies from place to place. For instance, new production and transportation technologies may gradually reconfigure temporality in a nineteenth century Canadian town, much as Stein indi cates. But elsewhere, as Jenny Shaw, John Frow, and Karen Davies suggest, natural life cycles significantly affect the pace of activities in a British retirement com munity; a major report about the separation of aboriginal children from their parents makes the past altogether present in postcolonial Australia; and a different system of social discipline, an ethics of care, creates other tem poral priorities than that created by an ethics of contract within health and social work in Sweden.

Time in Historical Geography

If the previous material illustrated two strands in the geographies of time and temporality, it did so with a considerable emphasis on contemporary work in geog raphy. In that regard, the last section displayed what has been called 'chronocentricity', a prejudice toward the present as an era worth particular attention. That tem poral focus may well be representative of geography more generally. Rhys Jones documents such a trend in Anglophone geography over the past two decades. His concern is not the varied notions of time in geography, rather the different time periods (or centuries) discussed in geographical research. He shows that, relative to work in the discipline 50 years ago, articles in the flagship English language geography journals have been focusing more on the twentieth century and the contemporary world to the neglect of earlier centuries. Jones offers five possible explanations, among them the potential signifi cance of wider demands for more 'applied' research, the 'cultural turn' in human geography, and the growing interests in the history of geography.

If the latter move is chronocentric, this section will not be exempt. The outline is highly schematic, and while the material covered was not necessarily the most salient in historical geography, it was symptomatic of various conceptualizations of time. As such, it will broaden the review beyond contemporary studies of the geography of time. Recall, for instance, how Glennie and Thrift explicitly reconfigure Thompson's singular revo lution in the history of time consciousness into a narrative of phased and gradual change. As their analysis draws on a number of ecological formulations (e.g., en vironments of information, communities of practice, and their growth and diffusion) to derive three successive dynamic phases in a history of time consciousness over the course of five centuries, the wording and staging of their sketch becomes reminiscent of plant succession in biology.

Such a stimulus would not be the first. In the *Annals* in 1929, Derwent Whittlesey suggested that 'sequent occu pance' could be a foundation for work in the discipline, viewing geography "as a succession of stages of human occupance [that] establishes the genetics of each stage in terms of its predecessor." He advocated a human geog raphy energized by visions of recording the legacy of the past in shaping a contemporary landscape. More than a 'purely descriptive' endeavor, he recommended that re gions be treated dynamically, with each stage of human occupance clearly linked to its predecessor in sequence. (The 'analogy' between plant succession and sequent occupance, he thought, would be 'apparent to all'). Like Sauer and the Berkeley School (whose renown would last far longer), Whittlesey favored a geneticism that traced the history that produced the current landscape. The temporality invoked by such work was diachronic, a historical trajectory of changes over time that lead to the present and that presumably help to explain it.

It was with Sauer rather than with Whittlesey, how ever, that a genetic approach began to represent an American vision for historical geography, distinct from the endeavors of H. C. Darby in Great Britain. Con vinced that he had to stake a claim for the field, Darby had consistently defined historical geography (as Oxford geographer, J. N. L. Baker, had before him) as "the re construction of geographical conditions of past times." Ralph Brown's work at Minnesota would bear strong similarities with Darby's series of historical cross sections in England. Like Darby, Brown was stimulated by the meticulous regional histories that Vidal de la Blache had produced in France, and like Vidal, both held geography to be a series of contingencies. Yet, Brown's early death and Darby's link to the distribution networks of Cam bridge University Press, along with the magnitude of his multivolume project, helped to ensure that Darby's syn chronic approach to geography – with its reconstructions of a region in a given period – would later be held as the British parallel to the diachronic and genetic approach taken by many other American geographers before the war.

By the 1950s, Darby was starting to recommend a diachronic approach, while work such as *The Making of the English Landscape* brought popular attention to a more genetic historical geography. In North America, the new intellectual leadership of Andrew H. Clark in the 1950s also helped reconcile the diachronic and synchronic approaches in the field. A student of Sauer, Clark's first book had traced the invasion of people, animals, and plants in New Zealand. Ten years later, his book on Prince Edward Island combined a sequence of cross sections, linked by chapters that trace changes in culture, settlement, and land use in the intervening years. More than simply enrolling Darby's cross sections into motion, the temporalities in Clark's work transformed the passing of time into an evolutionary schema whose object for explanation could become, not contemporary phenomena whose history the geneticists traced, but the very processes of geographical change over time.

Three fairly distinct temporalities informed the Anglophone traditions in historical geography at that point: some traced the history of a present landscape in a genetic fashion; many strove to record and contextualize the historical geography of a past period; and others worked a more evolutionary scheme concerned with the diverse processes of geographical change. The principal interests in Britain included agrarian transformation, industrialization, and regional environmental change, while scholars in North America documented the effects of European colonization on regions around the continent. In the early 1950s, such research was closely intertwined with economic history and economic geography, and historical geographers often published their work in journals of economic geography and history. Yet, a quantitative revolution soon cast economic and historical geography at opposite ends of the methodological spectrum as the former came to symbolize the new nomothetic analysis, while the latter was often cast as a refuge for 'outdated' regional geographers and an idiographic approach. By the 1970s, a spate of publications began to appear that looked explicitly for 'new' theoretical directions in the field. In that context, structuration theory and a new humanism took root in historical geography.

More than a reaction to the quantitative spatial science, humanist geographers highlighted human creativity and agency in ways that resonated with the individualized experimentalism that was popular among many students. A focal point for humanism was lived experience: not just an awareness of place, but also temporal experience. Yi Fu Tuan, for instance, wrote about diverse experiences of time in relation to space as motion, and to place as a pause. Elsewhere, Anne Buttimer sought a new vocabulary to ask how space and time 'speak' in the everyday life world, and David Lowenthal explored select American attitudes to the past, describing people who felt themselves "uniquely untrammelled by the past." Perhaps this future oriented quality of much American experience at mid century suited the contemporary foci of nomothetic spatial analysts in geography as well, as Cole Harris once remarked. But before humanism had taken root, before some looked afield to Marxian–humanist historians such as E. P. Thompson and Raymond Williams, many historical geographers of Clark's days had largely disregarded lived experience in favor of disembodied narratives of regional change. By the 1980s, scholarly interests in subjective lived experience had rippled across the field, and geographers such as Thrift and Harvey could speak acceptably (though from very different theoretical orientations) about the changing everyday geographies of time consciousness.

If humanists struggled to evoke a stronger sense of creativity in cultural and historical geography, structurationists were more equivocal. Structuration theory drew schematic links between human agency, structures, and social change. In its formulation, daily practices both structured, and were structured by, wider social configurations. The theory dovetailed with the techniques of time geography (first developed by Torsten Hägerstrand, later modified with Buttimer's input), where the actions and course of an individual were represented as a continuous 'path' across a unit of time–space, subject to innumerable, corporeal interactions and constraints. Among Anglophone historical geographers, work toward a structurationist time geography such as that of Don Parkes and Nigel Thrift was deemed too present oriented and formalist. Rather, Allen Pred's work on urban family life in the nineteenth century America was often considered exemplary. While his work seemed to deploy a notion of time as a linear dimension through which people move, its greater significance, like that of humanism, was its focus on 'the event' – rather than genetic, contextual, or evolutionary time – as the crucial temporality in analysis. Agency was thus understood as a decisive moment, an event (or if unsuccessful, a non event) that served as a potential rupture in the residues of history.

To date, work on gender and individual time budgets demonstrates the most practical implications of a structurationist time geography. A number of feminist geographers have explored how women and men are positioned differently with respect to labour markets and how their work is valued differently in national accounting systems. Accordingly, less than one third of men's work worldwide is devoted to unpaid and undocumented labor, whereas for women, such labor on average accounts for two thirds of their work. Others have used time geography to depict the daily paths of individual women in select locales. As Gillian Rose argues, time geography problematically erases passion and bodily difference when it reduces each subject to an object on a path. Yet, insofar as the technique treats time as a resource central to the production of privileged social networks, scholars such as Jackie Tivers and Isabel Dyck have used it to reveal the daily constraints and unnaturally limited options their informants faced as mothers, and to show, through a series of events, how the women negotiated and changed the conflicting demands of motherhood and paid work. Still others, leaning more toward historical than social geography, such as Jeanne Kay and Mona Domosh, made use of the inroads of humanist geographers in the 1980s (rather than those of time geographers) to reinterpret traditional forms of historical geographic narrative from feminist per spectives, in order to highlight the significance of women's lived experience, domestic household econ omies, and the gendering of identities in the shaping of North America.

Since then, feminist theory has become one of the three major strands of thinking – along with Gramscian–Marxist and postcolonial thought – that guide a vast new body of work on collective memory. Those three theo retical developments helped to reshape an understanding of memory in geography. Conceived as a social activity, memory was seen less as passive cognitive phenomena like perception, more as shared recollections that con tinue to forge a sense of identity, place, and social dif ferentiation. Yet the act of remembrance is instrumental as well, since people recall the past in support of their own agendas and understandings, and since the past that one calls up at any moment is necessarily selective: recollecting some things inevitably entails the neglect of other things. In the 'historic' village of Stockbridge in rural Massachusetts, for instance, popular memory tells of many national leaders, writers, and artists who lived in this posh, quiet community. Yet when residents consider the character of the place they now wish to endorse, it makes a difference whether they remember the large theater built on Main Street in 1888 to promote the arts, or the small working class homes built just behind the theater (secluded from view and easily forgotten), as that which made this historic village work.

Along a similar vein, many historical geographers have been writing about the politics of memory. Among them, there is considerable work on symbolic landscapes, war memorials, relics, and museums. Like the time budget analyses, such work carries potent implications about what is valued in the molding of social memory. Many invoke Walter Benjamin, and his concern with arresting the ostensibly 'progressive' flow of history in the 1930s (Kevin Hetherington describes Benjamin's as a more 'kairological' notion of time) resonates with the narratives of counter memory that historical geographers often offer. Here, a disjunctive temporality is sometimes deployed, one where time is folded over. The past is understood to be hauntingly present, active in the 'now' of structural forces, landscapes, practices, and objects, all while people work on the past through contemporary interpretation. From this perspective, time is seen less as a cyclical or progressive movement, more as the winding strands of a double helix, where the present and the past intertwine and change their interlocking course in un predictable ways.

Time and Historical Geography

It may be productive not only to call upon Benjamin and the unsettling temporalities he provokes, but also to ask why his conception of time has now become attractive in historical geography. Yet this article's remaining space is devoted, instead, to suggesting how a different notion of time came to predominate with the emergence of his torical geography. This question enables one to render historical geography, like time, as an unsettled object, a configuration subject to contestation. If a title like 'time in historical geography' prefigures the certainty of sub disciplinary space, then the editors' chosen title for this article, 'Time and historical geography,' affords more room for symmetry, so as to consider notions of time and the subject of historical geography both as social con structions. One might accordingly turn to a decade when boundaries of the subject were far less certain than they are today.

In 1953, H. C. Darby published one of the first methodological treatises about historical geography, and Andrew H. Clark was just composing his 35 page in ventory that was to give the subject some shape and lineage. That was the context in which Jean Brown Mitchell's book, titled *Historical Geography: A Student's Guide*, was published in 1954. As a lecturer at Cambridge, Mitchell's work was a noted contribution to the insti tutionalization of that which is called 'historical geog raphy' in the Anglo American academic world. Hers was a small volume (356 pages, with the dimensions of a field guide) and it was the last volume in a series called *Teach Yourself Geography*. The need to define the field was introduced at the beginning of the first chapter, titled 'What is historical geography?' Seven substantive chap ters followed, outlining first the sources that could be used, and then the processes of interest for a potential student: the peopling of England and Wales, the growth of farms and villages, and changes in the countryside, industry, and transportation.

Mitchell ended each chapter with a dozen mapping exercises that "the reader may like to make of his area." She apologized for her treatment of material that "may often seem local, even parochial, in scope." Yet that focus was strategic: Mitchell's was a deliberate effort in adult education for "readers outside the University Schools." Its aim was not only the production of a more learned populace in Britain (building on the 'citizens outdoors' movement), but also the promotion of the new sub discipline while using a popular interest in local history as its springboard. As historian Raphael Samuel notes, organized historical landscape walks were 'in the air' with the nostalgia of post war England, enlisting energy from thousands of amateur historians in 'do it yourself' pro jects that honored the hardships of regular people in Britain. Many universities were adding adult education courses in local history to their offerings, and by 1952, local historians in Britain had even started their own journal. Yet, before W. G. Hoskins and M. W. Beresford became prominent in that movement, *Historical Geography* had showed interested adults how they might use family names in their community to trace the histories of immigration through which their predecessors had come to Britain. In doing so, Mitchell capitalized on post war national sentiments and a wave of interest in local history in order to bring the emerging subdiscipline to broader public attention.

If we are to better understand this emergence however, the context in which 'historical geography' was promoted is not the only critical component. The predominant temporality with which Mitchell configured this new subdisciplinary formation was also significant. As anthropologist Carolyn Greenhouse suggests, propositions and assumptions about the nature of time should be understood as situated ideas about the sources and circuitry of agency. Greenhouse argues that specific representations of time are deployed in order to legitimate an account, in order to render its distributions of relevance, responsibility, and agency credible within the political contestations of an institution. If Mitchell's book, *Historical Geography*, is read with that in mind, then the primary temporality with which she operates is not simply as another development in the many theorizations of time in historical geography (that seemingly 'progressive flow of history' again). Rather, her major temporal framing can then be understood as a political move, one that fortified the terms of accountability within which a subdisciplinary formation could be assembled and legitimated.

Consider, for instance, Mitchell's distance from the temporality of 'the event'. She positioned its examination as an exercise 'best left to the historian', whose primary concern lies with a civilization and the individuals who, to many historians, constitute the primary sources of agency in constructing that society. Historical geographers, according to Mitchell, instead were to embrace the temporal frame of 'development'. In contrast to 'the event', her developmental notion of time effectively distributed agency across the complex and "complete assemblage" of relationships between people, the environment, and technologies, an assemblage that in turn constituted a place or region. Configuring time in this way was advantageous, in that it distinguished the work of historical geographers from the work of historians. Moreover, it conformed to the diachronic temporalities that Darby and Clark were starting to advocate, and with such notions of time, Mitchell strove to legitimate her new subdiscipline as a developing assemblage in the evolution of knowledge.

But this was not the only temporality configured in Mitchell's book. At the end of her second chapter, after reviewing a number of sources and methods for historical geography, she summarized its purpose. If we understand the subject as a formation in construction, then that moment might become one that we should pay attention to, for there Mitchell (1954: 40) wrote: "Equipped now with maps, note book, stout boots and a stout heart the historical geographer is ready to pursue his investigations o'er fell, field, and fen, down macadamed road, up cobbled street with eyes open and mind alert to see and appreciate the visible landscape as the present phase of an ever changing pattern indissolubly linked to its past and irrevocably the foundation of its future." Here the dominant temporality was again diachronic, from the past to 'the present phase' in the evolution of a landscape, which she then posited as the basis for its future. Here she represented time as a progressive continuum in which an 'ever changing' landscape appears, and she distributed its sources of agency widely across an 'assemblage' of quotidian relationships in place, between technologies, social groups, and the environment. Yet a minor temporality was also at work, one that cultivated a moment of preparation and potential, a moment of anticipation on the part of readers 'ready' to pursue their studies "o'er fell, field, and fen." This other sense of time in Mitchell's book was a decisive moment, now effaced from most disciplinary histories, when time and its circuits of agency were briefly framed so as to configure an open historical geography constituted through a fecund democratization of learning and knowledge production.

See also: Historical Geography; Memory; Modernity; Postcolonialism/Postcolonial Geographies.

Further Reading

Bassin, M. and Berdoulay, V. (2004). Historical geography: Locating time in the spaces of modernity. In Benko, G. & Strohmayer, U. (eds.) *Human Geography: A History for the 21st Century*, pp 64 82. London: Arnold Publishers.

Glennie, P. and Thrift, N. (2005). Revolutions in the times: Clocks and the temporal structures of everyday life. In Livingstone, D. & Withers, C. (eds.) *Geography and Revolution*, pp 160 198. Chicago, IL: University of Chicago Press.

Gould, S. J. (1987). *Time's Arrow, Time's Cycle: Myth and Metaphor in the Discovery of Geologic Time*. Cambridge, MA: Harvard University Press.

Green, D. B. (ed.) (1991). *Historical Geography: A Methodological Portrayal*. Savage, MD: Rowman and Littlefield.

Greenhouse, C. J. (1996). *A Moment's Notice: Time Politics across Cultures*. Ithaca, NY: Cornell University Press.

Harvey, D. (1989). *The Condition of Postmodernity*. Cambridge, MA: Blackwell.

Hetherington, K. (2008). The time of the entrepreneurial city: Museum, heritage, and kairos. In Cronin, A. & Hetherington, K. (eds.) *Consuming the Entrepreneurial City*, pp 273 294. New York: Routledge.

Hoelscher, S. and Alderman, D. H. (2004). Memory and place: Geographies of a critical relationship. *Social and Cultural Geography* 5/3, 347 355.

Jones, R. (2004). What time human geography? *Progress in Human Geography* 28(3), 287 304.

May, J. and Thrift, N. (eds.) (2001). *Timespace: Geographies of Modernity*. London: Routledge.
Mitchell, J. B. (1954). *Historical Geography: A Student's Guide*. London: English Universities Press.
Morin, K. and Berg, L. D. (1999). Emplacing current trends in feminist historical geography. *Gender, Place, and Culture* 6(4), 311 330.
Pred, A. (1981). Production, family, and free time projects: A time geographic perspective on the individual and societal change in nineteenth century US cities. *Journal of Historical Geography* 7(1), 3 36.
Thompson, E. P. (1967). Time, work discipline, and industrial capitalism. *Past and Present* 38, 56 97.
Tuan, Y. F. (1978). Space, time, place: A humanistic frame. In Carlstein, T., Parkes, D. & Thrift, N. (eds.) *Making Sense of Time*, pp 7 16. London: Edward Arnold.

Time Geographic Analysis

M. Dijst, Utrecht University, Utrecht, The Netherlands

Glossary

Authority Constraints Regulation of access to activity places in the form of social rules, laws, financial barriers, and power relationships.

Base Location Activity places in which activities are conducted that act as pegs around which the rest of the activity and travel pattern is organized.

Capability Constraints Biological, mental, and instrumental limitations.

Coupling Constraints All limitations which arise when people, equipment, and materials are joined together at given times and places.

Potential Action Space The area which encompasses all activity places that can be visited in a certain time window.

Prism Three-dimensional space–time construct which contains all potential paths a person may choose given that person's spatiotemporal constraints.

Project Entire series of tasks necessary to the completion of any goal-directed behavior.

Station Activity places with diverse physical extensions.

Time Window A block of time in which travel and nonbase location activities can be conducted.

Travel Time Ratio A share of travel time in the sum of travel time and activity duration in a certain time window.

Introduction

The origins of time geography can be traced to the mid 1960s. It forms a part of the human ecological tradition in human geography, one that exceptionally has been developed by geographers alone. The founding father Torsten Hägerstrand (1916–2004) sees it as a four dimensional (4 D) view of the world. Although there are various versions of time geography, in this contribution the focus is on classical time geography, which was mainly developed in the 1960s, 1970s, and early 1980s, with some later extensions.

Hägerstrand redefined the domain of this geographical discipline on the basis of his observations of approaches in human geography in the 1960s. His main problem with the typical geographer's thinking and practices at that time was that these were unable to catch the conditions which circumscribe human action. His criticism of human geography was largely driven by his strong motivation to make geographical knowledge and skills more useful for policymakers and planners. For that purpose it was necessary to develop a conceptual and instrumental framework that could capture the connected human activities that are taking place in time and space bound situations. To address these shortcomings of human geography, Hägerstrand proposed a new theory based on three pillars: space, time, and the identity of mankind.

Hägerstrand reproached quantitative geographers, asserting that they were more interested in the distribution of human beings and other material entities across space than in their space consuming properties and the consequences of these for their ordering in space. Space is not only a matter of distance between material entities, but also of the room taken up by these entities. A housefly needs just a tiny piece of 3 D space, but an international airport consumes a relatively large amount. This notion of space as provider of room implies, at least in some relationships, competition for available room between human beings, animals, plants, roads, buildings, and other entities. Some sites in space offer better opportunities than others for reaching one's goals. Once a site has been occupied by a certain entity that space is no longer available for other objects, which are then forced to search for another site in space.

In addition to this constraining notion of space, the time dimension also sets limitations on relationships between entities. Time is not a budget like money that can be saved up for later use. Just like a clock, time in an individual's life ticks away. Moreover, at every moment in time an individual is located somewhere in space. As a consequence, events in time, like the birth of a child, calling a dentist, buying groceries in a store, and traveling to a destination, are not discrete, but are located and connected within space–time blocks. Consequently, one event might have consequences for another event, like the birth of a child that triggers the move to another, larger dwelling, which in itself might disconnect the household from the local community.

The third pillar of time geography forms the identity of humankind. Hägerstrand stated that a human being is treated as a 'dividual' rather than an 'individual'. By this he means that in human geography human beings are studied as shopper, husband, father, colleague, or sportsman separately, without taking account of the fact that a person has to integrate the activities which belong to these various roles in space and time. Many of these activities have specific durations, and times and places at which they can be carried out. For example, a full time job as a teacher binds an employee to a school for a long duration during daytime hours. This location in time and space limits a teacher's opportunity to participate in other activities.

Basic Ideas of Classical Time Geography

Time geography is based on the idea that the life of an individual, but also of other organisms and material ob jects, describes an uninterrupted 'path' through time and across space. The timescale of these paths can vary from a day to the whole life span. Every organism or thing is constantly in movement, even when they are themselves at rest. In this manner one can see society as built up from a large number of webs or networks composed of uninterrupted paths that have been drawn by people, other organisms, and nonliving elements through time and space. These individual paths cross each other in 'stations', also called 'bundles'; these are activity places like a dwelling, an office, a shop, or a beach where an individual participates in activities. These activities are called 'projects' and can range from preparing a pie to performing surgery and includes people, resources, space, and time.

In **Figure 1**, segments of paths are placed in a 3 D 'aquarium' in which the vertical time dimension is placed on a 2 D representation of space. The left side of this figure shows that each member of a household follows an individual path. From the perspective of a station (right side of figure), in this case a school, various paths of pupils and teachers come together within a certain time span to attend or to give lessons. The slope of the lines in the figure represents the speed of traveling: the steeper the slope, the slower or more time consuming is the transport mode moving across space.

In time geography, participation in activities in pro jects is not a matter of choice, but is subject to three types of constraints. The first type is capability constraints: biological, mental, and instrumental limitations. For ex ample, people sleep on average 7–8 h a night at home, have certain intelligence and skills to carry out activities, and have transport modes at their disposal which enable them to travel through time and across space at various speeds. Coupling constraints, the second type, are en countered by people when they come together at a cer tain time and location with equipment and other materials for joint activities. Different time schedules and different locations of origins can complicate this coupling of individual paths. Finally, authority constraints regulate the access of individuals to activity places through social rules, laws, financial barriers, and power relationships. Business hours, the price of admission, and curfews are some examples of these regulatory constraints.

At the foundations of these constraints lie eight fun damental interrelated restrictions that people cannot escape. These are given as follows:

1. Human beings, and many other living and nonliving entities are indivisible.
2. Human beings, and many other living and nonliving entities have a limited life span.
3. Human beings, and many other living and nonliving entities have a limited ability to take part in more than one task at a time.
4. All tasks have a certain duration.
5. Movement across space consumes time.
6. Space has a limiting packing capacity.
7. Each territory has a boundary, like a house, a city, a country, or the Earth as a whole.
8. Each situation is inevitably rooted in past situations linked by an individual's paths through time and across space.

In time geography, it is assumed that some activities such as work and home activities are fixed in time and in locations called 'bases' (e.g., home and work location) and determine the opportunities to conduct more flexible activities such as shopping or recreation in time and space. These flexible activities can be pursued in 'time windows', also called time intervals, which represent blocks of time in which travel and relatively flexible ac tivities can be carried out. The constraints mentioned above define for a certain time window a 3 D 'prism',

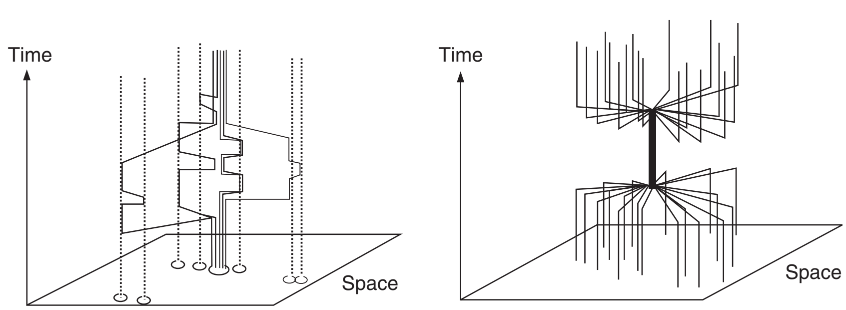

Figure 1 Bundles: on the left a household as a bundle, on the right a school. From Thrift, N. (1977). *An Introduction to Time Geography.* Norwich: Geo Abstracts, University of East Anglia (CATMOG-13).

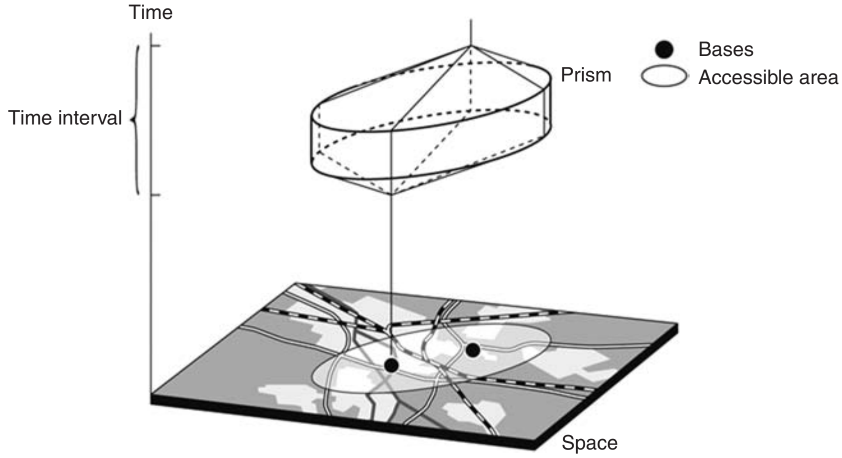

Figure 2 Prism and potential action space. From Lenntorp, B. (1976). *Lund Studies in Geography, Series B, Human Geography, no. 44: Paths in Space-Time Environment: A Time Geographic Study of Possibilities of Individuals*. Lund: The Royal Society of London, Department of Geography.

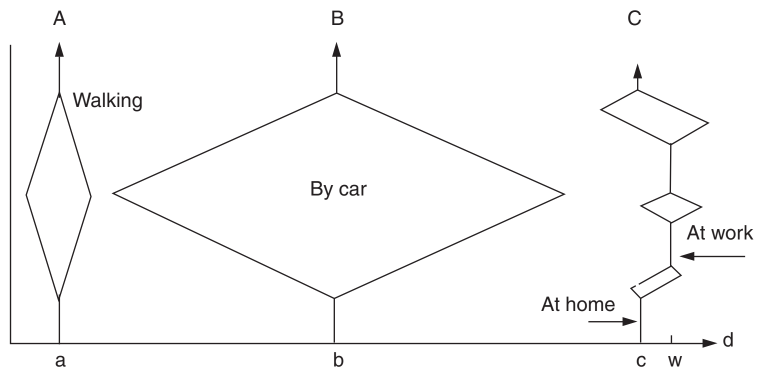

Figure 3 Cross-sections of prisms. From Parkes, D. N. and Thrift, N. J. (1980). *Times, Spaces, and Places: A Chronogeographic Perspective*. Chichester: John Wiley.

which embrace the set of opportunities for traveling to activity places and to participate in activities (**Figure 2**). Each step taken in space and/or time transforms the prism into a path.

The volume of the prism depends on the type of transport mode chosen (**Figure 3** (a to b)). Faster trans port modes increase the size of the prism, which is also shown by the shallower slopes of the prism. **Figure 3** (c to w) also shows that time spent at various stations cuts the original prism into several smaller prisms that all nest within the original one.

The projection of a prism in a 2 D or 3 D space des ignates the 'potential action space', also called the 'activity area' or 'potential path area'. This area contains all the activity places that can be visited within a certain time window: in other words, it represents the accessible area. Each individual carries around a personal potential action space like an air balloon, which expands and contracts under the influence of changes in the three constraints mentioned and behavioral decisions concerning the course of the daily path. An action space has not only a plastic, but also a fluid character, as shown in **Figure 4**. This means that when an individual moves across space, other people and material objects move in and out of the po tential action space because of the change in the indivi dual's spatial position. Conversely, when a person is fixated in space, others with their own potential action spaces move in and out of that person's action space.

Potential action spaces can be grouped into three basis categories according to their ideal shape: circular, linear, or elliptic (**Figure 5**). Circular action space results when a journey starts and ends at the same base, like shopping trips from home or in the lunch break from work. A linear action space starts in one base and ends in another and the available time is just enough to cover the distance between the two bases. This is often the case for daily commuting trips. After subtracting the time ne cessary to travel between the bases, some extra time re mains for visits to other stations en route; the resulting action space is elliptical. Of course, these three basis categories of action spaces are ideal forms. In reality, because of variations in spatiotemporal configurations of

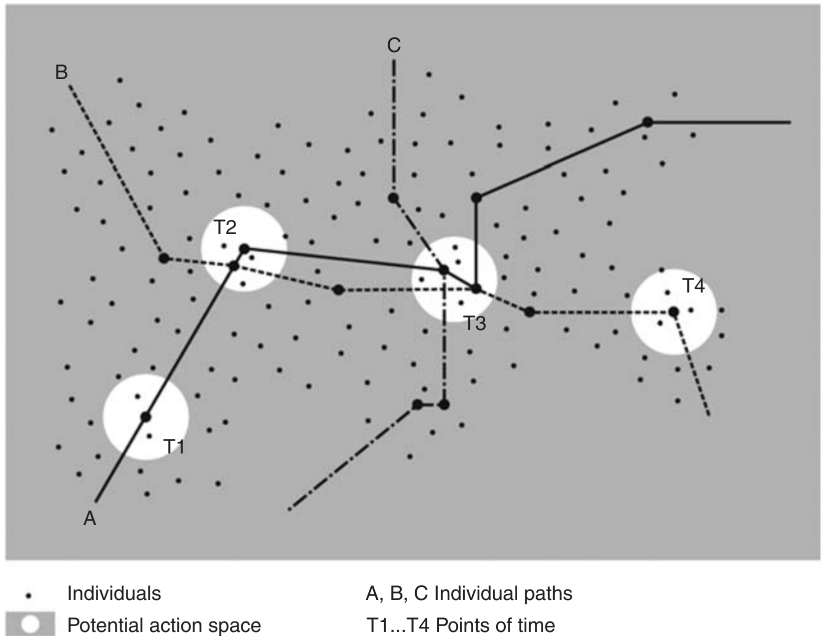

Figure 4 Composition of potential action space dependent on temporal and spatial location. From Dijst, M. (2006). *Stilstaan bij Beweging: Over Veranderende Relaties tussen Steden en Mensen.* Utrecht: Faculty Geosciences, Utrecht University.

road networks, for example, potential action spaces can take fanciful shapes (**Figure 6**).

An individual's prism or potential action space has in principle three temporal choices: to spend the entire available time budget on travel without spending any time in an activity place; to stay in the base location without traveling outside; or to spend time on travel as well as on activities in one or more activity places. It is assumed, however, that individuals will balance travel and activity duration, as reflected in the definition of the 'travel time ratio': the ratio between travel time and the sum of travel time and activity duration. In principle, this ratio lies somewhere between 0 and 1. For example, a travel time ratio of 0.45 for shopping means that 45% of the available time window in which a visit to a shop is included will be spent on traveling to and from the shop. Although no comprehensive studies are available, empirical evidence suggests that the ratios for commuting, maintenance, and discretionary activities are on average 11%, 45%, and 25%, respectively. Application of the travel time ratios to prisms means that the volume of the prism shrinks.

Application of Time-Geographic Analysis

Hägerstrand attached great value to the application of knowledge in planning practice. For that purpose, time geographers have developed simulation models to assess the effects of planning measures on the choice opportunities that individuals of various types have. Two examples of these models are Program Evaluating the Set of Alternative Sample Path (PESASP) and its improved version Model of Action Spaces in Time Intervals and Clusters (MASTIC). These models facilitate the assessment of the potential impact of time policies (e.g., the business hours of shops, flexible working hours, and adjusted public transport schedules), transport policies (e.g., new road constructions and new bus stops), and spatial policies (e.g., changes in the density or mixture of activity places) on the opportunities offered to people to participate in their desired activities.

The MASTIC model includes four variables that determine the size, shape, and location in time and space of potential action spaces. These determinants are: the distance between the bases, the length of the available time window, the speed of the main transport mode, and the travel time ratio. The characteristics of potential action spaces do not depend solely on the individual. Not only the supply and quality of transport systems, but also the spatiotemporal configuration of the activity places influence the quantity and quality of available opportunities within the action space. Application of this model is dependent on the availability of accurate data on at least the individual departure times from origins and arrival times at destinations, the transport modes used, and the street addresses of all types of activity place visited. In addition, a digital geographic database of the study area is necessary that includes information on activity places (like type and business hours) and a detailed street network for all transport modes.

The MASTIC model has been applied to the Dutch city of Zoetermeer. This new town is located 12 kilometers from The Hague and accommodates over 100 000 inhabitants and 40 000 jobs. The backbone of the

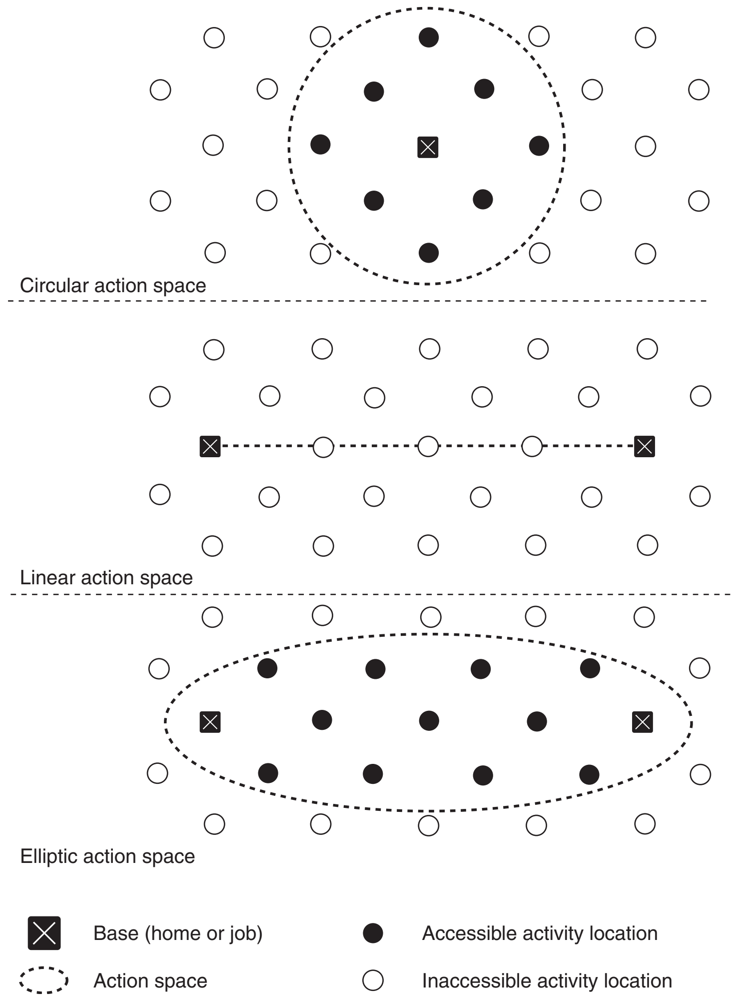

Figure 5 Three basis forms of action spaces. van Eck, R. J., Burghouwt, G. and Dijst, M. (2005). Lifestyles, spatial configurations and quality of life in daily travel: An explorative simulation study. *Journal of Transport Geography* 13, 123 134 with permission from Elsevier.

city is formed by a railway line with 12 stops within the city and connects Zoetermeer with The Hague. A sample of 222 respondents living in one neighborhood com pleted a travel diary for two successive days. They wrote down all the trips made, the departure and arrival times, the transport modes used, and the type and exact address of the activity places visited. These diaries gave infor mation on 'activity programs' or the activities carried out during the various time windows on two days. Data on the household composition and employment situation were collected in a complementary survey; this resulted in the definition of five types of couples and families: single earner households without children, single earner households with children, dual earner households with out children, two earner households with children, and retired households without children. During the simu lations the distance between the bases and their location, the timing and length of time windows, the travel time ratios, and the activity program for each time window, which comprises all activities in which the individual participated, were not changed. Only the order in which activity places were visited, the location of activity pla ces, and transport mode could change.

Various scenarios can be defined to show the impact of transport and spatial policies. The scenarios in **Table 1** differ in the density of dwellings, namely 15 and 35 dwellings per hectare. It is made clear that these spatial configurations have different impacts on people's op portunities to participate in activities outside the bases. As expected, a high suburban density offers relatively the best opportunities; on average, four fifths of all activity programs can be realized. In contrast, in a low suburban density with its longer travel times, only two thirds of all activities can be carried out. In this scenario everybody in low density areas, especially those who walk or cycle, and members of two earner households, encounter substan tial difficulties in their daily lives.

Spatial configurations can be assessed not only in terms of the feasibility of activity programs, but also in terms of travel distance and travel time efficiency

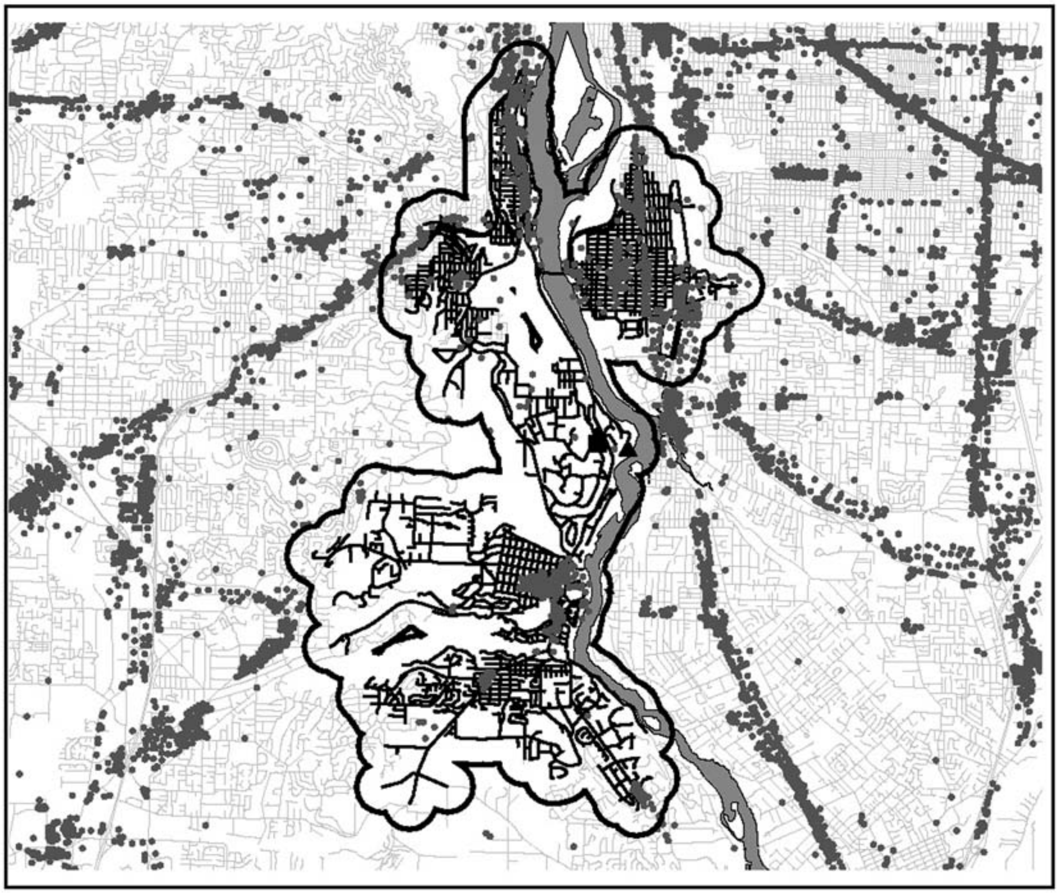

Figure 6 A potential path area (PPA) of 20 minutes travel duration between a pair of fixed activities. From Weber, J. and Kwan, M.-P. (2003). Evaluating the effects of geographic contexts of individual accessibility: A multilevel approach. *Urban Geography* 24, 647 671.

(**Figure 7**). The absolute differences in distance traveled by car users in the low and high density scenarios are between 2 and 5 kilometers or relatively between 15% and 40%. For travel time, the differences between the two scenarios are much smaller, but still substantial: between 3 and 6 minutes or relatively between 15% and 20%.

Time geographic simulations could be valuable in facilitating the assessment of the impact of policies on people's daily activity participation. The approach could also be valuable in assessing the impacts of policies on the use of various facilities. **Figure 8** shows that a con sequence of seriously pushing back the use of the private car in a community to the benefit of public transport, walking, and cycling is that consumer flows and the economic basis of many stores will change.

Geographic Information System, Three-Dimensional Geovisualization and Time-Geographic Analysis

The enormous increase in the power of personal computers, the availability of rich georeferenced individual level data, and improvements in geographic information systems (GISs) facilitate the extension and enrichment of the analysis of human activity and travel patterns in time and space within a time geographical framework. This development has largely been made possible by GIS based geocomputation, which refers to a set of computer based techniques, such as data mining, genetic algorithms, cellular automata, fractal modeling, and visualization; as a result, geographical variations in phenomena across scales can be analyzed. Algorithms have been developed within the time geographical framework to refine the analysis of the spatiotemporal opportunities available for individuals. For example, procedures are available to take into account the spatial and temporal variations in travel speeds and business hours of activity places, the impact of traffic rules (like one way streets and turning prohibitions), minimum ac tivity duration, and maximum travel time budgets.

The visualization of geographic information (geovi sualization) in particular offers ample opportunities to create, view, and understand the spatiotemporal images of individual paths in real geographic space of particular

Table 1 Percentage of activity programs realized in different spatial configurations, by household type and main in transport mode Zoetermeer-Rokkeveen, The Netherlands

Density/household type	*Car*	*Public transport*	*Walk/ bicycle*	*Total*
Suburban density scenario: high density[a]				
Single-earner households with child(ren)	87	100	62	80
Single-earner households without child(ren)	83	100	100	84
Two-earner households with child(ren)	82	93	68	80
Two-earner households without child(ren)	83	86	61	78
Retired households without child(ren)	100	100		100
Total	83	89	65	80
Suburban density scenario: low density[a]				
Single-earner households with child(ren)	72	100	41	64
Single-earner households without child(ren)	73	100	50	72
Two-earner households with child(ren)	69	87	55	68
Two-earner households without child(ren)	74	72	34	64
Retired households without child(ren)	88	100		83
Total	72	78	41	67

[a] All differences significant at the $p < 0.01$ level.
The dash in the table represents that the number of cases are too small.
From van Eck, R. J., Burghouwt, G. and Dijst, M. (2005). Lifestyles, spatial configurations and quality of life in daily travel: An explorative simulation study. *Journal of Transport Geography* 13, 123 134.

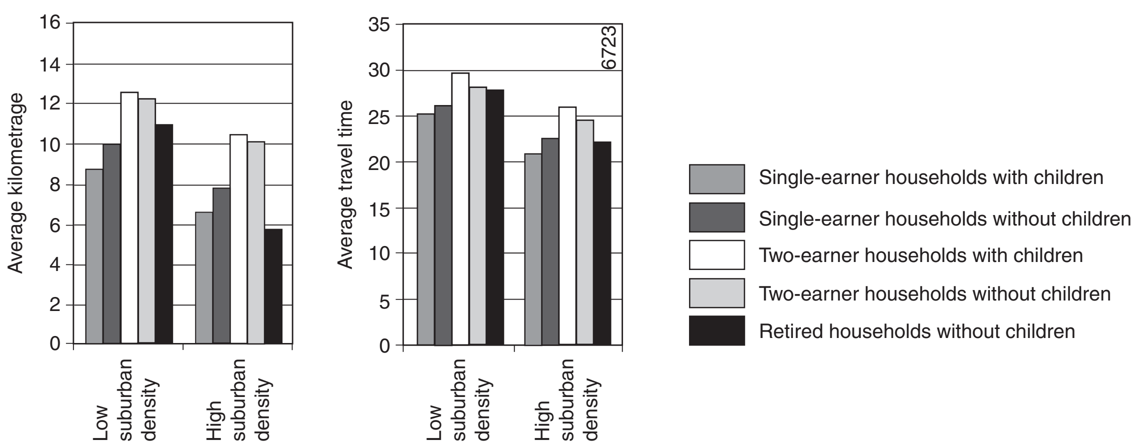

Figure 7 Average car distance per journey (left) and car travel time per journey in minutes (right) for low and suburban densities by household type. Source: Ritsema van Eck, J., Burghouwt, G. and Dijst, M. (2005). Lifestyles, spatial configurations and quality of life in daily travel: An explorative simulation study. *Journal of Transport Geography*, 13, 123 134.

study areas. In the past, 2 D maps and geographic methods were applied to depict the web of paths created by individuals. 3 D offers not only 3 D complex and realistic representations of the (urban) environment, but also the opportunity to see immediately the effect of the user's manipulation of the determinants of individual paths, such as changes in travel speeds, the location of activity places, and business hours. Finally, navigation tools, such as fly through, zooming, planning, and dynamic rotation together with multimedia capabilities, allow users to create their own virtual images of the real world.

Individual paths through time and space in GIS are represented as a collection of vertical and diagonal trajectories in a space–time aquarium (**Figure 9**). The vertical axis is the time dimension and the horizontal plane the geographic space of the study area. On this plane, the administrative boundary of the study area (Metropolitan Portland), rivers, and major car routes can be clearly distinguished. As the enormous bundle of paths in daytime shows, most individuals from the African American, Hispanic, and Asian American ethnic groups work in the downtown area. Many maintenance and leisure activities are undertaken by these individuals within and to the east of downtown Portland. By using the interactive capabilities of 3 D GIS, categorizations of paths can be displayed and readily interpreted. For

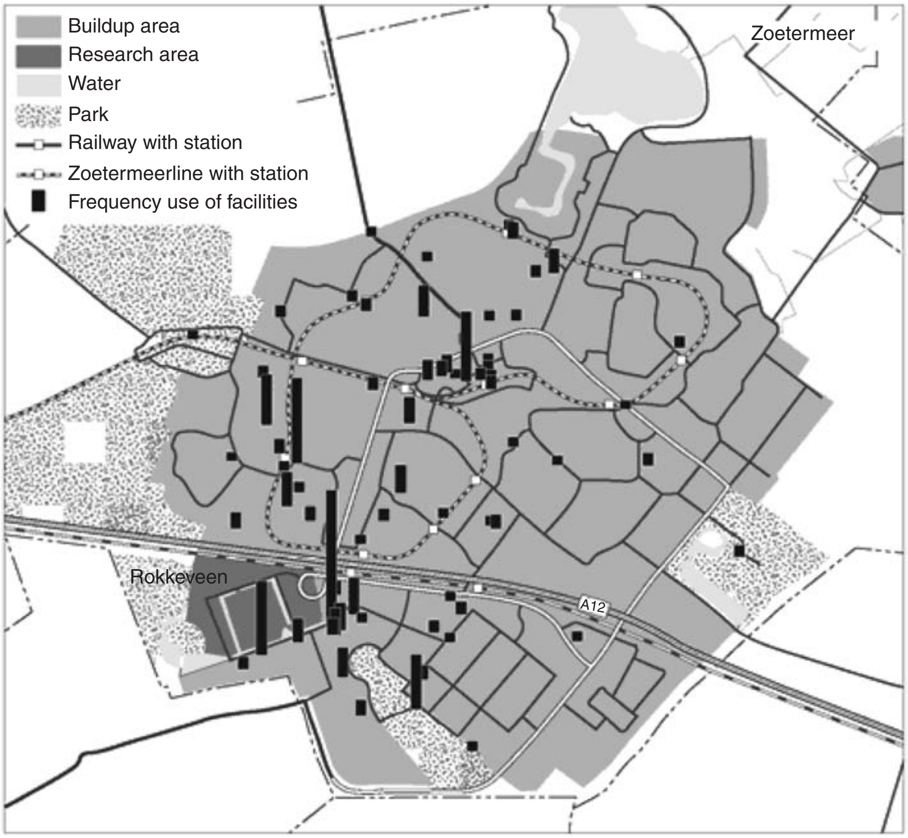

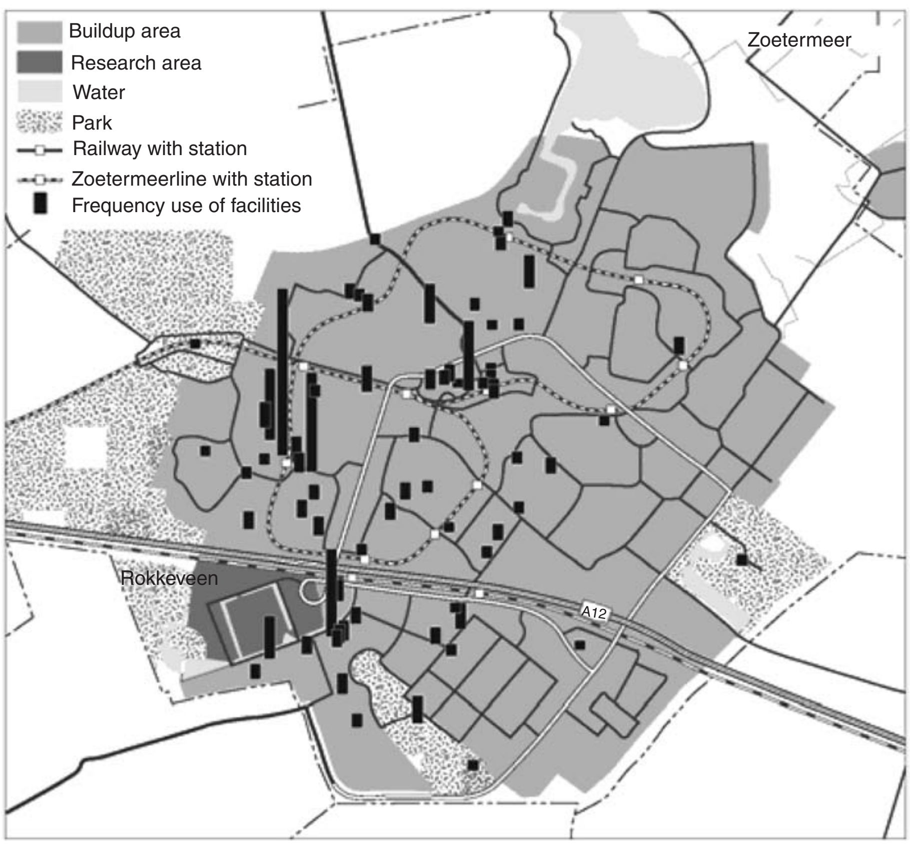

Figure 8 Actual use of facilities (upper) and without car use (lower) within Zoetermeer, according to location and frequency. From Dijst, M., de Jong, T. and van Eck, R. J. (2002). Opportunities for transport change: An exploration of a disaggregated approach. *Environment and Planning B* 29, 413 430.

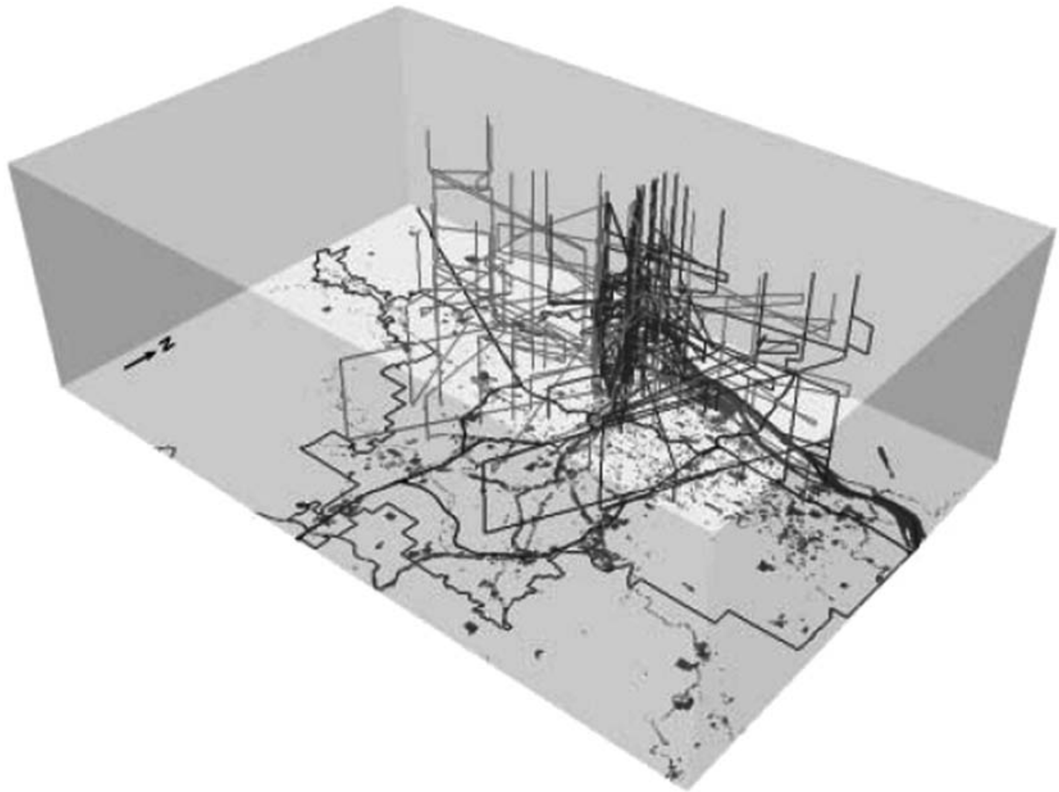

Figure 9 Space time aquarium showing the space time paths of African and Asian Americans in the sample. From Kwan, M.-P. and Lee, J. (2004). Geovisualization of human activity patterns using 3-D GIS: A time-geographic approach. In Goodchild, M. F. & Janelle, D. G. (eds.) *Spatially Integrated Social Science*. New York: Oxford University Press.

example, it can be shown that full time employees have a less fragmented space–time path than others; and, compared with the other two ethnic groups, African Americans have a much more compact spatial pattern of paths.

Zooming in on this 'aquarium' brings up a more detailed representation of the study area and the individual paths it contains (**Figure 10**). The image clearly shows the Willamette River and the 'loop' around downtown Portland. The heights of the buildings on parcels are shown and the use of color codes for specific types of urban functions, such as orange brown (commercial buildings) and green (industries), gives a good impression of the spatial opportunities offered in the study area. In combination with the portions of some space–time paths, the user can search on fine spatial scales for specific types of individual and the use of time and space and the opportunities to adjust these paths to changed individual or contextual circumstances.

GIS based geocomputation also offers methods to extract patterns from the large number and on first sight highly diverse space–time trajectories drawn in the 'aquarium'. One application is shown in **Figure 11** for employed individuals from the three ethnic minority groups mentioned above. In this figure all home locations in the 'aquarium' have been moved to the origin (0, 0) and the diverse home work axes have been rotated until it becomes the positive x axis. This standardization of space–time paths is depicted by the home work plane. In an interactive session it can be shown that many work activities are carried out in bundles in proximity of the home and that a considerable amount of nonwork activities are carried out during the day often toward the home work axis.

With the help of Global Positioning Systems (GPSs) it is feasible not only to show the origins and destinations on an individual's path as in the former figures, but also to trace the actual travel routes followed in an urban environment. **Figure 12** portrays the actual routes traveled by female car drivers without children under 16 years of age in the Lexington area in Kentucky. These car drivers' trips mainly followed highways and major arterials.

3 D visualization can also be extended to the use by individuals of information and communication technologies (ICT). People can communicate with others without moving across space with a personal computer at home or at work and with mobile communication devices on the road. **Figure 13** represents an individual's space–time path and ICT activities on three geographical scales: global (upper); regional (lower: 15 states in northeastern region of the USA); and local (middle: Franklin County, Ohio, USA). As the space–time path depicts, the person lives and works in Franklin County. She visits a

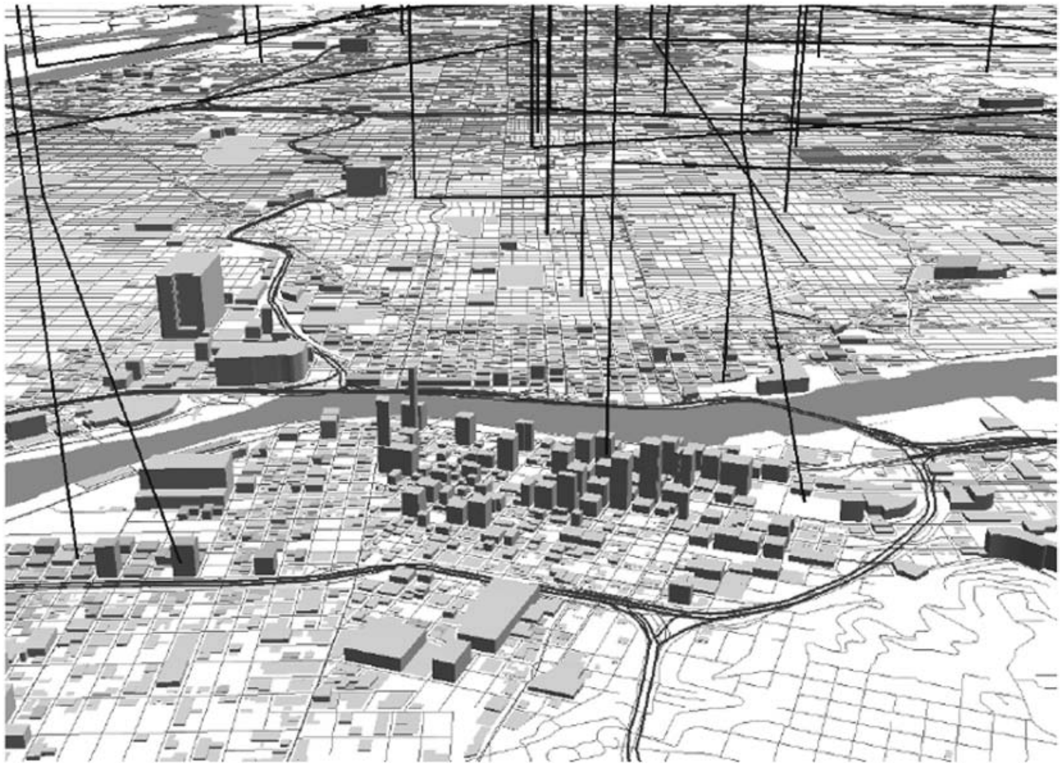

Figure 10 A close-up view of downtown Portland. From Kwan, M.-P. and Lee, J. (2004). Geovisualization of human activity patterns using 3-D GIS: A time-geographic approach. In Goodchild, M. F. & Janelle, D. G. (eds.) *Spatially Integrated Social Science*. New York: Oxford University Press.

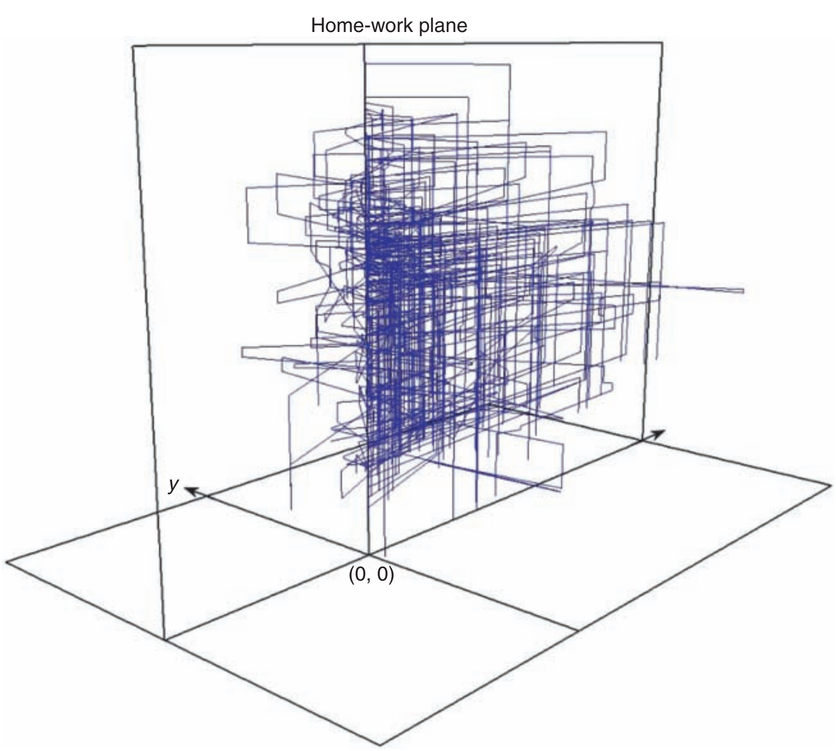

Figure 11 Standardize space time paths. Source: Kwan, M.-P. (2000). Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: A methodological exploration with a large data set. *Transportation Research C* 8, 185 203.

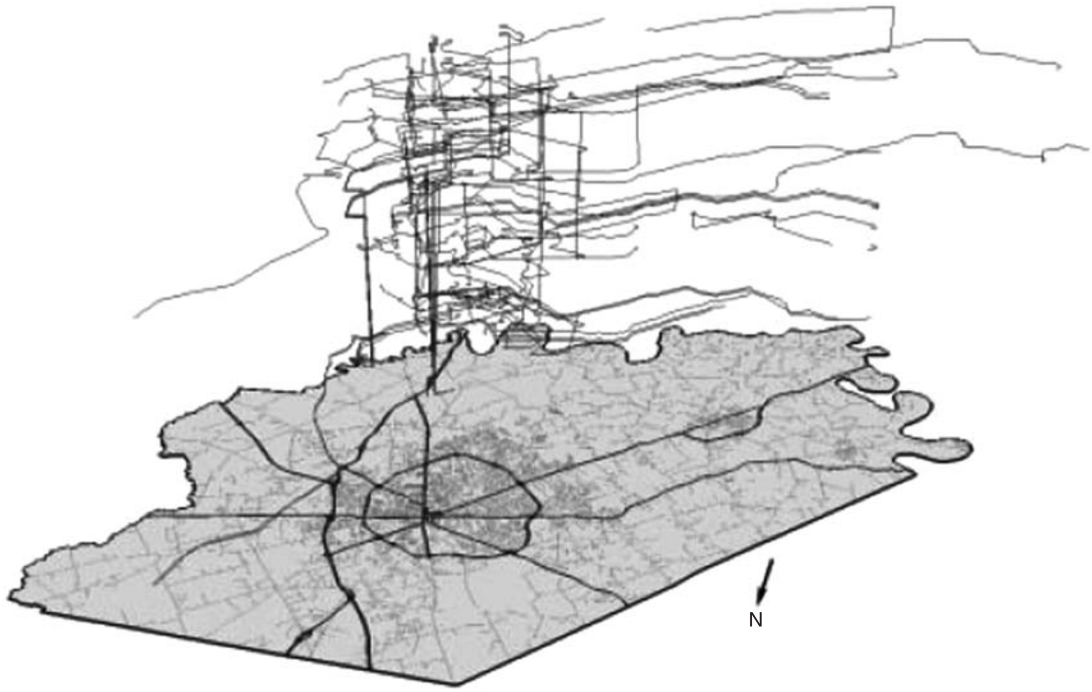

Figure 12 Space time paths based on GPS data collected in Lexington, Kentucky. From Kwan, M.-P. and Lee, J. (2004). Geovisualization of human activity patterns using 3-D GIS: A time-geographic approach. In Goodchild, M. F. & Janelle, D. G. (eds.) *Spatially Integrated Social Science*. New York: Oxford University Press.

restaurant at walking distance during a lunch break at work (see in **Figure 13**). She also engages in ICT ac tivities with other locations in the northeastern states of the USA, in South Africa, and Japan. At two moments at work she participates in nonpersonal communications. In the morning she subscribes to a webcasting service (a) where news from Yugoslavia, South Africa, and Nashville is forwarded to her web browser and in the afternoon she browsed web pages in New York, Charlotte, and An chorage in Alaska (d). At other moments of the day she participated in personal communications. In the morning she sent some emails to three friends living in Hong Kong, Chicago, and Vancouver (b). The difference in slopes shows that the friends read the email with different time lags. The friend in Hong Kong was the last to read the email but replied immediately (e). She read this email a couple of hours later at home (g). In the evening at home she chatted with friends in Tokyo, Melbourne, Memphis (Tennessee), and Dublin (Ohio) (f).

Analysis of the impact of ICT on human activity and travel patterns leads also to the introduction of new time geographical concepts such as ‘portal’, ‘message window’, and ‘digital box’. A portal is a type of station where in dividuals can access appropriate communication services. A message window is a time interval in which individuals can interact with portals. Finally, a digital box is a space–time block within which use of ICT devices and services is allowed.

Criticisms on Time Geography

Criticism on time geography has been dominantly for mulated by supporters of structuration theory, feminist geographers, and those who embrace a relational per spective on geography.

Anthony Giddens, the founding father of structuration theory, criticized time geography for its treatment of individuals as independent of their social settings in daily life. Hägerstrand did not discuss the origins of people’s projects, stations, domains, and constraints. His emphasis was more on behavioral routines instead of the essentially transformational character of human action. “All types of constraints .... are also types of opportunity, media for the enablement of action” (Giddens, 1984: 117). As a consequence, time geography has a very weakly de veloped theory of power.

The feminist geographer Gillian Rose opposed the unproblematized universality of bodies and spaces in time geography. Although corporeality is a basic notion in time geography there is little discussion on the body. The theory considers the body as undifferentiated and actually reduces it to a path through time and space. Based on rationalist considerations paths mesh at certain spatiotemporal locations but they never merge. This minimalist account of the body denies differences be tween sexuality, gender, and race which together with associated emotions all have a differentiating impact on

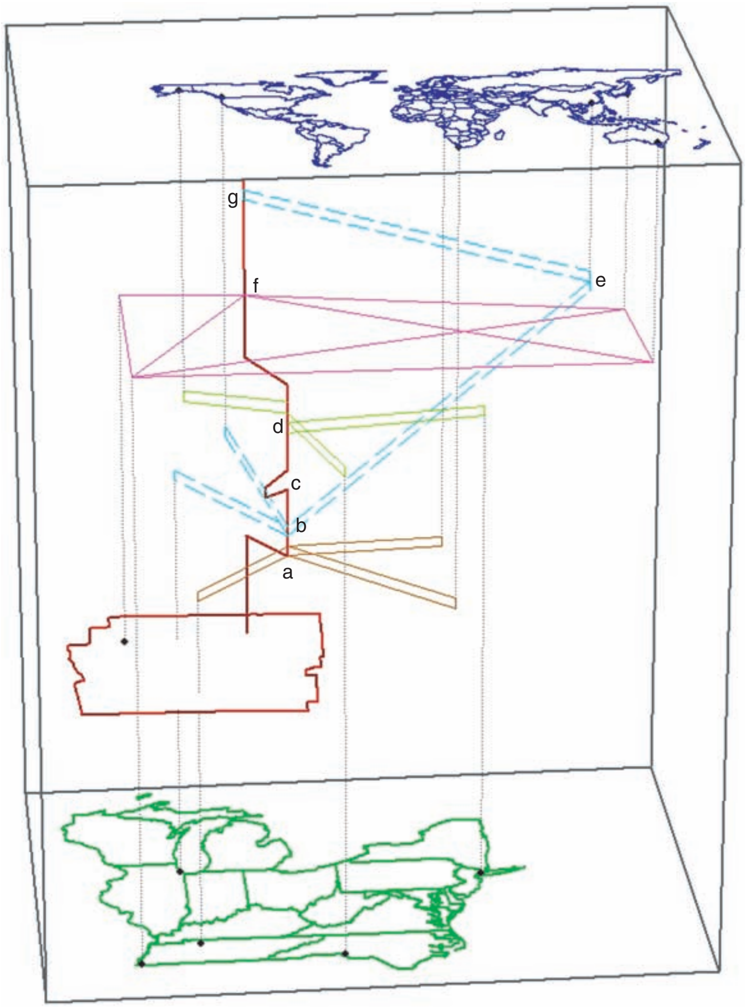

Figure 13 Travel and ICT activities in space time using 3-D GIS. From Kwan, M.-P. (2000). Human extensibility and individual hybrid-accessibility in spacetime: A multi-scale representation using GIS. In Janelle, D. G. & Hodge, D. C. (eds.) *Information, Place and Cyberspace: Issues in Accessibility*, pp 241 256. Berlin: Springer-Verlag.

consciousness and agency. For example, childbirth, mothering, and domesticity create a specifically different subjectivity and sociality for women than men who in daytime work in the office and return home after work.

A related criticism of feminist geographers is on time geographer's conceptualization of space. 3 D space is seen as infinite, unbounded, and transparent. Although in time geography participation in activities is constrained by constraints, it denies the feelings associated with these constraints. For example, women often experience fears caused by sexual harassment. The same is true for some homosexual men and nonwhites for violence in the streets. Comparable with bodies there is no universal unproblematized space but there are more spaces dif ferentiated by subjectivity and sociality.

Differentiation of time and space is also the subject of relational perspectives on geography, like post structur alism, nonrepresentational theory, and actor network theory. Time geographers have an absolute perspective on time and space. They use a 3 D time–space or 'aquarium' (**Figure 1**), in which individual paths are placed. It is a representation of: "*…a singular or uniform social time stretching over a uniform space…*" (May and Thrift, 2001: 5). However, in a relative conceptualization of time–space this entity is understood as a network of relationships in which consciousness of temporal and spatial distances are measured in terms of processes and activities. These processes and activities could be related to nature (like temporalities of seasons and years), re ligion (like church attendance and feast days) and economy (mechanical clock and business hours).

Future Issues

Time geographic analysis offers ample opportunities to analyze human activity and travel patterns. Although this approach can be applied at the individual, station, and population/activity system level, analysis on the first level dominates. For example, the spatiotemporal prob lems of women, household types, and ethnical minorities have been studied extensively. However, the other two levels also deserve more attention. Mapping out the prisms or potential action spaces of individuals will de pict the size and composition of the group of people who can visit a station and as a consequence the economic or social basis of an activity place. At a societal level a time geographic analysis could enhance our insights into total time investments in activities and travel and in inter actions and communications of various population categories.

Although the development and spread of new ICT means and services can be said to be revolutionary, the knowledge on this subject is highly inconclusive and even contradictory. So far, the opportunities offered by the use of communication means and services have only been marginally developed in time geography. The integration of conceptual and methodological ICT in the time geographical framework and the analysis of its impact on activity and travel patterns will be a challenge.

Finally, classic time geography does not integrate into the theoretical framework the meaning of cognitive processes, such as opinions and feelings, for the devel opment of paths through time and across space. In this respect the approach is reductionist situational. One can wonder whether this exclusion of the mental part of human activity and travel patterns remains a preferred approach or whether insights from other approaches from sociology and psychology should be integrated with time geography.

See also: Geovisualization.

Further Reading

Dijst, M. (2004). ICT and accessibility: An action space perspective on the impact of new information and communication technologies. In Beuthe, M., Himanen, V., Reggiani, A. & Zamparini, L. (eds.) *Transport Developments and Innovations in an Evolving World*, pp 27 46. Berlin: Springer.

Dijst, M. (2006). *Stilstaan bij Beweging: Over Veranderende Relaties tussen Steden en Mensen*. Utrecht: Faculty Geosciences, Utrecht University.

Dijst, M., de Jong, T. and van Eck, R. J. (2002). Opportunities for transport mode change: An exploration of a disaggregated approach. *Environment and Planning B* 29, 413 430.

Dijst, M. and Vidakovic, V. (2000). Travel time ratio: The key factor of spatial reach. *Transportation* 27, 179 199.

Giddens, A. (1984). *The Constitution of Society: Outline of the Theory of Structuration*. Berkeley, CA: University of California Press.

Hagerstrand, T. (1970). What about people in regional science? *Papers of the Regional Science Association* 24, 7 21.

Hagerstrand, T. (1973). The domain of human geography. In Chorley, R. J. (ed.) *Directions in Geography*, pp 67 87. London: Methuen.

Kwan, M. P. (2000). Human extensibility and individual hybrid accessibility in spacetime: A multi scale representation using GIS. In Janelle, D. G. & Hodge, D. C. (eds.) *Information, Place and Cyberspace: Issues in Accessibility*, pp 241 256. Berlin: Springer Verlag.

Kwan, M. P. (2000). Interactive geovisualization of activity travel patterns using three dimensional geographical information systems: A methodological exploration with a large data set. *Transportation Research C* 8, 185 203.

Kwan, M. P. (2004). GIS methods in time geographic research: Geocomputation and geovisualization of human activity patterns. *Geografiska Annaler B* 86, 267 280.

Kwan, M. P. and Lee, J. (2004). Geovisualization of human activity patterns using 3 D GIS: A time geographic approach. In Goodchild, M. F. & Janelle, D. G. (eds.) *Spatially Integrated Social Science*, pp 48 66. New York: Oxford University Press.

Lenntorp, B. (1976). *Lund Studies in Geography, Series B, Human Geography, no. 44: Paths in Space Time Environment: A Time Geographic Study of Possibilities of Individuals*. Lund: The Royal University of Lund, Department of Geography.

May, J. and Thrift, N. (eds.) (2001). *Timespace: Geographies of Temporalities*. London: Routledge.

Miller, H. J. (2005). A measurement theory for time geography. *Geographical Analysis* 37, 17 45.

Miller, H. J. (2005). Necessary space time conditions for human interaction. *Environment and Planning B* 32, 381 401.

Parks, D. and Thrift, N. (1980). *Times, Spaces, and Places: A Chronogeographic Perspective*. Chichester: John Wiley.

Raubal, M. and Miller, H. J. (2004). User centered time geography for location based services. *Geografiska Annaler B* 86, 245 265.

Rose, G. (1993). *Feminism and Geography: The Limits of Geographical Knowledge*. Cambridge: Polity Press.

Thrift, N. (1977). *An Introduction to Time Geography*. Norwich: Geo Abstracts, University of East Anglia (CATMOG 13).

van Eck, R. J., Burghouwt, G. and Dijst, M. (2005). Lifestyles, spatial configurations and quality of life in daily travel: An explorative simulation study. *Journal of Transport Geography* 13, 123 134.

Weber, J. and Kwan, M. P. (2003). Evaluating the effects of geographic contexts of individual accessibility: A multilevel approach. *Urban Geography* 24, 647 671.

Time Geography

M. Gren, Hólar University College, Sauðárkrókur, Iceland

Glossary

Conceptual Apparatus Time-geographic concepts with meanings and definitions which meet the requirements of the matter-realistic ontology.

Corporeality In time-geography the everyday realm consisting of bodies (or material things) with the quality of being physical.

Matter-Realism A philosophical basis for time-geography, which means that the conceptual apparatus and time-geographic notation should, in terms of representation, correspond to the everyday world of matter and corporeal existences.

Notation System A system of representing or depicting by signs (or symbols).

Time-Geographic Diagram A system of graphic representation that depicts by lines in relation to the coordinates of time and space the simultaneous movement of corporeal objects or material events on the reference plane of the Earth.

Time–Space Corporeal objects and material events have both spatial extension and temporal duration and therefore time and space always come together in time-geography.

Trajectory A graphical depiction of the movement of corporeal objects (or material events) in both time and space. The trajectory is the simplest and most fundamental concept in time-geography.

Introduction

Time geography is a conceptual geographic framework. It is founded on a matter realistic ontology that explicitly takes departure from the material eventness of geographical phenomena and processes, in their dynamic time spatial situatedness on Earth. An essential characteristic of time geography is thus that conditions of the physical world are conceptually acknowledged and incorporated at the outset. Most importantly, it means that matter with the dimensions of time and space is conceived of as fundamental in the constitution of human corporeal and geographical life, as well as for nonhuman configurations. Time geography includes a notation system and a conceptual apparatus which in combination offers a time geographic methodology. By this methodology, the time spatial embeddedness of human earthly life, as well as various material events and their geographical orderings, can be studied regardless of their disciplinary belongings.

As a research program on its own, time geography has increased the conceptual and empirical knowledge of time–space material conditions of and for human corporeal and geographical existence. Over the years it has been applied in numerous areas in human geography, as well as in a large number of studies in a variety of different disciplinary settings. One of its most successful applications has been in studies of mobility, accessibility, and everyday geographies. As an ontological approach, rather than a theory in a formal sense, time geography has also enabled through its methodology the utilization of other theoretical and empirical knowledges. For example, feminist scholars have drawn on time geography in order to investigate social divisions of gender. This has meant that time geographic methodology has also been used for purposes and in contexts that as such may have had little to do with time geography *per se*.

On the History of Time-Geography

The central ideas of time geography were initially developed in the late 1960s and the early 1970s by Torsten Hägerstrand (1916–2004), the most influential Swedish geographer of our time. Especially during the final 10 year period of his tenure, when Hägerstrand was holding a personal professorship (1971–82) at the University of Lund, he tried to distill and elaborate upon his own fundamental geographical ideas so as to give them a more consistent and systematic form. This development eventually came to include also contributions from a number of associates that together were to become known as 'the Lund school'.

Many of Hägerstrand's geographic ideas that were to be formative for time geography had been with him for a long time, in a professional sense, at least since the 1940s. For example, doing research in the area of demography acquainted him by necessity with problems and opportunities regarding the graphical representation of statistical data. When investigating migration patterns in the population of Asby in rural southern Sweden his collected data concerned 10 000 individual biographies, yet he managed to condense these into one small diagram.

In terms of Hägerstrand's own biography, he grew up in the small village of Moheda in the Swedish countryside, which, according to him, produced in his mind a 'holistic' geographic map of a place on Earth where it was visible where the milk, meat, and the firewood were coming from. It was a kind of a visible vertically enclosed geography that was soon about to be changed through various processes of modernization, leading to more horizontal, and increasingly invisible, geographies of and through society and space. In that modern world, the earthly trajectories of various materialities would proliferate into complex and wide spanning geographies, now increasingly at a global scale. At the same time, Hägerstrand came to argue that they were prone to disappear in the dematerialized language of growing modern social science. Such a verbal understanding of the social enabled one to move freely around in thought and tell stories about economy, culture, society, and the like, but without the necessity to establish an apparent and clear reference to the time–space embeddedness of material events and technological artifacts on Earth.

The discipline of geography was also to be transformed during Hägerstrand's career, and by the 1960s, it was no longer a synthesizing 'Earth discipline'. A modern division had left physical geography to take care of nature and the world of matter and set human geography with the task of establishing itself as a social science. Hägerstrand, having been schooled in geography as a discipline that preceded the nature–culture divide and which included a significant German Earth tradition (*Erdbeschribung* or *Erdkunde*) done by great synthesizers like Alexander von Humboldt, Carl Ritter, and Alfred Hettner, always remained skeptical about the modern trajectory of the discipline. In that respect, his time geographic project was not an attempt to construct a version of human geography as social science. Instead, it was a way of holding onto, and to further develop, some of the fundamentals of a unifying geography. As an attempt to counterbalance the fragmentation of disciplinary knowledge by trying to keep nature, society, and technology within a single conceptual framework, time geography is then quite reminiscent of an 'old geography' interested in the transformation of habitats over time that included not only its people and artifacts but also its natural base.

Time geography was also motivated and influenced by Hägerstrand's involvement in regional policy studies underway in Sweden in the 1960s, for example, in a major research project funded by the Bank of Sweden Tercentenary Foundation. A characteristic of Swedish modernity thus was that social scientists, to a large extent, were involved in governmental commissions and research projects that were meant to provide scientific evidence on which political decisions about the improvement and future betterment of society could be made. A central aspect of the policy studies that Hägerstrand participated in was to compare the living conditions in various parts of Sweden and to figure out ways of equalizing these conditions with respect to access of employment, education, healthcare, cultural resources, and recreation. Thus, in focus were various locational and spatial issues, for example, how to delineate new municipal districts, or how to plan for the supply of various public facilities and services such as transportation.

One central representational problem Hägerstrand then identified, which cuts into the very heart of time geography, was that there did not yet exist a 'neutral frame of reference' for describing and measuring the impacts of processes in society in such a way that their geographic implications and consequences for its people as individuals became visible. Eventually, he drew the conclusion that it was one thing to reason about social phenomena in society, for example, to make various calculations on the basis of dematerialized statistical aggregates of living conditions, but quite another to place them and their consequences in actual space–time settings on Earth.

As a response to that representational problem, time geography was an attempt to create a combination of demography and geography. One of the original principal aims was to develop a notation system that, through a set of descriptive concepts, would make it possible to register and handle observations in their developing space and time dependent relations. Furthermore, such a system should make possible the meeting of different nonadditive and disparate knowledges in such a way that the time spatial effects and impacts upon individuals could be grasped and measured, for example, an individual's possibility for actions somewhere. In the context of policy studies, the proper yardstick would then be how the development of society affected the biographies of its people as individuals who are always living geographically specific lives rather than in a dematerialized social realm.

The first publication by Hägerstrand, which most explicitly presents the aim and postulates of 'time geographic description', was published in Swedish 1974. He there declared that the task was to counteract the processes of differentiation in geography and further develop the art of putting the various materials of specialists and subfields together into a coherent and unified worldview. He also proposed the development of a 'method of description' which should fulfill four general representational requirements:

1. It should be easy to realize to what the description corresponds to in reality.
2. The description must have a wide scope of applicability.
3. It should have the ability to generate questions which one could not pose without it.
4. It should admit conclusions and calculations whose reality correspondence does not have to be verified by observations.

The method should make it possible to put together, and hold on to, observations in such a way that data were not being pulled out from its place and time dependent context. It was to be developed into the notation system, which has become something of a hallmark of time geography. Time geographic ideas were subsequently further formalized by Hägerstrand himself, and increas ingly so also by his associates in Lund and other scholars, into a conceptual apparatus. In the 1970s time geography began to diffuse throughout human geography and social science, not least because of Hägerstrand's prolific pub lication in a broad spectrum of journals and with the help of internationally renowned scholars like Allan Pred and Nigel Thrift. After several decades time geography cannot be reduced to a conceptual entity that comes with only one unambiguous identity, nor to the initial geo graphical ideas of Hägerstrand. Yet, the significance of his ontological skeleton prevails underneath the added flesh.

The Notation System and the Conceptual Apparatus

Time geography includes a conceptual apparatus that consists of concepts that have in common that they have been given meanings and definitions in order to meet the requirements of the matter realistic ontology. These are by now almost too numerous to list; therefore, only a few of the most important concepts will be touched upon here. It seems, however, appropriate to begin with 'the notation system', since it has been so much equated with time geography and because its most important feature is the underlying ontology. In combination with the con ceptual apparatus the notation system provides the po tential to carry out analyses and make calculations that do not need to be tested empirically since the concepts are neither derived or decided nor affected by ephemeral contexts or situations. This illustrates the universal claims of the applicability of time geography.

The development of time geographic notation has always been part of an attempt to overcome the limits of verbal language. Hence, an essential idea behind the graphic representation in time geography is that it should make it possible to hold on to an earthly material situatedness which verbal language too readily departs from. In time geography the importance of nonverbal graphic description must then be kept as a guiding ontological and methodological principle however dif ficult this matter realism may be to execute in practice. One could, for example, think of how to represent every person's movements in a city during a day.

Time geographic notation has also been motivated by a need to overcome a limit of the traditional map, as it represents a spatial configuration frozen in time. By contrast, a central characteristic of notation in time geography is that time and space always come together in order to keep track simultaneously of both spatial and temporal dimensions of material geographical events. As long as a 'corporeal' object exists, it will be somewhere and it will endure in time. In other words, it will have and be a time geography. This means that positions of an object may change in reference to time only, or to both time and space (**Figure 1**).

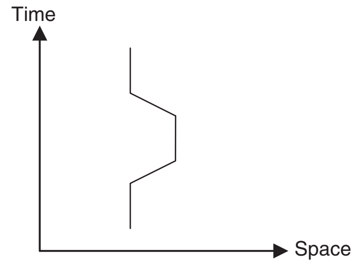

Figure 1 A time-geographic diagram.

The time geographic diagram illustrates, in principle, fundamental time spatial conditions of material life and how these can be made observable by means of graphic representation. In the diagram an object is represented as a 'trajectory' (or path), which is both the simplest and the most fundamental concept in time geography. The tra jectory above graphically depicts an object that is first located somewhere enduring only in time. Then it moves in both time and space to another location, where it stops for a while, before returning to its original position. The trajectory thus internalizes the level of generality de manded by the conceptual apparatus by depicting the continuous movement of a corporeal material object, such as a car or a human, by reference to both time and space. It also erases the distinction between human and nonhumans because it graphically represents both as matter. Any physical object can in principle be repre sented by a continuous trajectory, as long as the object is indivisible and has a positive direction in reference to the time axis. That physical objects cannot move backward in time, or make discontinuous jumps into the future, are examples of basic time spatial constraints which time geography claims to be universal in the everyday cor poreal realm.

In the everyday time geographical life of corporeal material entities there are three principal phases: 'coming into being', 'survival', and 'death/destruction'. Matter/energy is continuously changed, moved, and transformed on Earth so that corporeal entities may arise, or be cre ated, and thereafter embark on their own time geog raphies. The inhabitants of the Earth are thus corporeal humans, objects, trees, birds, technological artifacts, and the like, on their respective roads from cradle to grave. These corporealities constitute together various popu lations and, in turn, their aggregated time geographies.

Every form of human, as well as nonhuman, geog raphy is thus preceded and succeeded by space–time regularities due to Earth bound conditions and con straints laid down by matter. The concept of 'pockets of local order' refers to the time spatial coordination of various resources, people, and artifacts. The home is thus a local pocket of order embedded in the wider web of flows of matter. In time geography the concept of local pockets of order is not only restricted to the human world but includes also the wider ecologies of nature. Plants, like humans, do not only need local pockets of order for their maintenance, they are their own local pockets of order during their time geographies of existence.

The activities of humans in their geographies of everyday life, the most important scale of inquiry in time geography, necessarily demand a time spatial co ordination of various materialities. For example, work place, school, and home must be brought forth, arranged, and continuously maintained in order to serve their functions. In the conceptual apparatus, 'domain' and 'station' denote such time spatial locations and the or derings they imply. Like local pockets of order, domains and stations have to be negotiated when we need, or wish, to carry out projects either on our own or together with others in the choreography of earthly coexistence.

The concept of 'project' used in time geography de notes a consciously goal directed undertaking which could range from short term ones, like reading this art icle, to long term projects such as getting an education. The successful execution of any project demands that one be able to time spatially establish and negotiate a series of local pockets of order that are advantageous to the goal in question. The realization of any project is subject to interactive constraints that in time geography often have been divided into three sets:

1. 'Capability constraints' are those that limit activities of individuals due to their own bodily capabilities, or needs like food, shelter, and sleep. Also included are nonhuman facilities and technologies that individuals are able to draw upon.
2. 'Coupling constraints' are those that define where, when, how, and for how long an individual must couple with others, which could also be tools and materials, in order to complete a task or fulfill a project.
3. 'Authority (or steering) constraints' are those of a more dematerialized kind that impose restrictions on access and modes of conduct, for example, legal and political regulations or praxis in a workplace. The concept of domain often denotes these more 'invisible' kinds of constraints.

Together, these constraints define a set of possible time–space trajectories that individuals in principle are able to realize within their own specific time–space settings. In a certain sense 'the prism', which has been applied espe cially in studies of accessibility and potential movements, is in fact a much more important graphic representation than the trajectory. The reason is that the prism graphically discloses the constraints that circumscribe a time–space within which trajectories must progress (**Figure 2**).

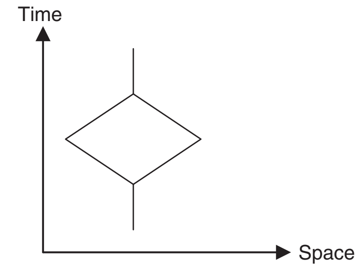

Figure 2 The prism.

The prism depicts that, under given conditions like a certain access to transportation, one could only travel so far if one has to be back at a specific time. Thus, it also illustrates that time geography is primarily concerned with the identification and analysis of fundamental con ditions and constraints of geographical life.

Although crucially important, the graphic represen tations should not be equated with time geography *per se*. Furthermore, the abstraction of the graphic representation used, with the two axes of time and space, has most likely contributed to an impression that time geography simply, or only, operates with a Newtonian like conception of time–space as a kind of absolute container in which ma terial events occur. There is some additional complexity and ambiguity attached to the usage of time and space in time geography than what readily may meet the eye.

In time geography the concepts of time and space always come together or, preferably, as one time–space. Objects and material events are often accounted for as being situated 'in time–space'. Of crucial importance, however, is that both concepts arise in time geography foremost as means of representing and keeping track of corporeal objects and events. In principle, then, time and space are conceptualized in time geography as intrinsic dimensions of matter itself, having both spatial extension and temporal duration. This means that neither time nor space is to be understood as absolute since they primarily refer to, and have no existence outside of, matter. In the conceptual apparatus these have been called 'embedded time' and 'embedded space'.

Since this conceptualization of time and space is de rived from the workings of the material world, it is also to some extent at odds with ideas that treat space and time as merely individual or mental categories deriving from dematerialized social understandings. Nevertheless, in the conceptual apparatus of time geography one also finds 'symbolic time' and 'symbolic space'. These con ceptualizations refer to space and time as abstract or dering devices, for example, clock time and administrative spatial boundaries. In relation to time geographic diagrams, one could here imagine existence

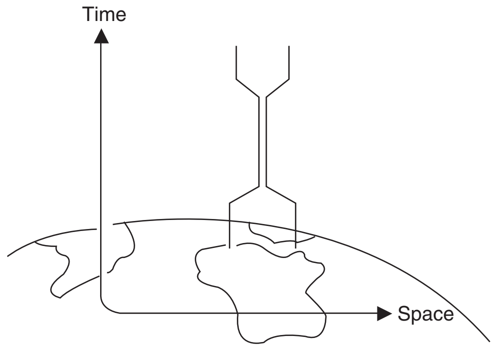

Figure 3 The Earth as plane of reference.

inside the trajectory as embedded time and space, whereas the external axis of time and space is of a symbolic kind. Indeed, the concepts of time and space are themselves symbolic constructions which only humans have access to because they operate in language.

When it comes to reconceptualizations of time and space that may have arrived with time geography, it is therefore highly important to now point out and include one decisive element that, due to the economics and problems of representation, actually has been missing in both **Figures 1** and **2**: the Earth (**Figure 3**).

In accordance with the matter realistic ontology of time geography, there is on Earth an almost infinite number of dynamic time geographies of corporeal objects and events in circulation and constant transformation. Any notational system is only able to represent limited samples of those time geographies. A time geographic diagram which uses time spatial notation should thus not be regarded as ignoring the geographical properties. It is the Earth that is the real plane of reference in time geography and its notation system, and not time and space either as separate or combined as time–space, however necessary and valuable these concepts may be.

Time-Geography Then and Now

Time geography has been distinguished as one of few truly original geographical approaches. It is clear that over the years it has been influential on a number of dimensions, both inside and outside human geography. Even though it is not, and was never meant to be, a social theory, it has nevertheless been drawn into discourses of the social in various ways. Of particular significance was when Anthony Giddens in the 1980s made time geography known to a wider audience operating inside social theory, by his usage of time and space in his theory of structuration. In that he drew on time geography and thereby came to acknowledge, among other things, that the physical mileux of social action is more than a passive backdrop. Time geography was thus taken up in the attempts to spatialize social theory. It was recognized for having something profound to say about the material and time spatial grammar of everyday life, and it was also perceived as filling a gap for those interested in studying it, not least because it offered a firm methodology. Especially many feminist projects thus came to valuably incorporate time geographic method in order to time spatially map gender practices and divisions of labor.

Yet, at the same time, this encounter with social theory also opened up room for critique as well as criticism. Through the years that followed time geography continued to be criticized for being 'too physical and mechanistic', echoing Anne Buttimer's remark that time geography was a "dance macabre" that left no space for mindful and passionate human beings, reduced as they were to mere trajectories in the bubble chamber of a transparent and unproblematic time–space. More broadly, many came to argue that time geography overemphasized the material, and because of that it was more likely to be just another variant of social engineering in disguise. In short, time geography lacked a sufficient substantial theorizing of the social. It could, for example, map and describe how daily life of individuals unfolded in time–space, but say very little about how stations and domains were socially produced.

The notation system and the graphical representation, or more specifically the time geographic diagrams, also came under scrutiny. They came to be disclosed not merely as a privileging of the visual in time geography, but as an expression of a 'masculine bias' or 'masculinist gaze'. This was shown by their conceptualization of time–space as a supposedly objective, transparent, and neutral frame of reference in which human bodies appeared as elementary universal particles erased from social and cultural markings such as gender. In retrospect this was something of a paradox because in few social areas, other than in studies of gender, has time geography been so successfully applied. To this, one could add that the criticisms were usually also dependent on a delineation of time geography which effectively excluded those that were situated in both time geography and social theory. Of particular importance in that respect were works by Allan Pred, who was always intimately aligned with time geography. In his theory of spatial practice he not only firmly acknowledged the time spatial and corporeal aspects of the human condition, also included the social through relationships of power and human systems of meaning.

So it is, that every time and space often comes with quite specific observations and conversations. By now much has changed in human geography, as well is in social science broadly conceived. There is now a much wider recognition and understanding of material and nonhuman agency in social theory. Instead of equating time geography with 'naturalism', it is possible to consider it to be a forerunner to this 'material turn'. Time geography effectively tried to sidestep social con structivism by emphasizing the physical and corporeal conditions as both constraints and possibilities on a human action always situated in wider networks of competing time spatial opportunities. Those diagrams then criticized for being signs of a 'physical reductionism' may now also be appreciated for attempting to in corporate nonhumans and humans in a common popu lation of time geographies, not that different from what Bruno Latour and others in actor network theory have referred to as 'the Collective'. Although time geography was never intended to be a social theory, it has always acknowledged that humans make their geographies together with nonhumans and that society and the social include technology and artifacts.

To this, one could add that time geography shares ingredients that appear in material variants of post structuralism as well in geophilosophy. In the former one could find an interest for the pragmatics of material events, their embeddedness and eventual unfolding in topologically complex time spatial networks, and a re fusal to reduce the world to language and texts. In the latter it is actually the Earth itself that appears as a primary plane of reference. Consequently, a geophilo sophically informed geography must somehow return to the discipline as an 'Earth science'. So now, as well as then, there is still a challenge posed to those geographers who sit uncomfortably in between the two modern pillars of the discipline and would like to have some kind of synthesizing discipline and theory of geography re invented. In that respect, time geography still provides an ontological window from which one may see not so much of society and space, nature or culture, but instead the time geographies of matter on the reference plane of the Earth, however delimited.

See also: Hagerstrand, T.; Human-Nonhuman; Maps; Philosophy and Human Geography; Physical Geography and Human Geography; Representation and Re-presentation; Space I; Space II; Space-Time; Urban-Rural Continuum.

Further Reading

Adams, P. C. (1985). A reconsideration of personal boundaries in space time. *Annals of the Association of American Geographers* 85, 267 285.

Carlstein, T. (1980). *Time, Resources, Society and Ecology*. Lund: Meddelanden fran Lunds universitets Geografiska institution avhandlingar LXXXVIII.

Carlstein, T., Parkes, D. and Thrift, N. (eds.) (1978). *Timing Space and Spacing Time II: Human Activity and Time Geography*. London: Edward Arnold.

Dyck, I. (1990). Space, time, and renegotiating motherhood: An exploration of domestic workplace. *Environment and Planning D* 8, 459 483.

Gren, M. (ed.) (1994). *Torsten Hägerstrand's geography of the visible* In *Earth Writing: Exploring Representation and Social Geography In between Meaning/Matter*. Gothenburg: Department of Human and Economic Geography, University of Gothenburg.

Gren, M. (2001). Time geography matters. In May, J. & Thrift, N. (eds.) *TimeSpace: Geographies of Temporality*. London: Routledge.

Hagerstrand, T. (1970). What about people in regional science? *Papers of the Regional Science Association* 24, 7 21.

Hagerstrand, T. (1974). Tidsgeografisk beskrivning. Syfte och postulat. *Svensk Geografisk Arsbok* 50, 86 94.

Hagerstrand, T. (1985). Time geography: Focus on the corporeality of man, society and environment. In Aida (ed.) *The Science and Praxis of man, society and environment*, pp 193 216. Tokyo: The United Nations University.

Hagerstrand, T. (1989). Reflections on What about people in regional science? *Papers of the Regional Science Association* 66, 1 6.

Kwan, M. P. (2004). GIS methods in time geographic research: Geocomputation and geovisualization of human activity patterns. *Geografiska Annaler* 86B(4), 264 280.

Lenntorp, B. (1976). *Paths in Time Space Environments: A Time Geographic Study of Movements Possibilities of Individuals*. Lund: Meddelanden fran Lunds universitets Geografiska institution Avhandlingar LXXVII.

Pred, A. (1981). Social reproduction and the time geography of everyday life. *Geografiska Annaler* 63B(1), 5 22.

Rose, G. (1994). *Feminism and Geography: The Limits of geographical knowledge*. Cambridge, MA: Polity Press.

Schwanen, T. (2007). Matters(s) of interest: Artefacts, spacing and timing. *Geografiska Annaler* 89B(1), 9 22.

Tivers, J. (1985). *Women Attached. The Daily Lives of Women with Young Children*. London: Croom Helm.

Wallin, E. (1980). *Vardagslivets Generativa Grammatik*. Lund: CWK Gleerup.

Time Series Analysis

G. M. Robinson, Kingston University, Kingston upon Thames, UK

Glossary

Autoregression The correlation between observations at different time–distances apart in a time series. This is a measure of 'self-correlation' between one data point in the series and another data point in the same series to see if successive or near-neighbor observations are correlated.

Correlogram A graphical plot depicting autocorrelation in a time series at various leads and lags.

Cross-Spectral Analysis Comparing whether time series for two locations are in phase or whether one series leads or lags another. This can establish the intensity of the relationship between them for various lags and leads, enabling change at one location to be used to predict change elsewhere.

Expansion Method Investigates the variation or drift in the temporal growth of a disease, in response to population density.

Fourier Analysis Named after the French scientist and mathematician Joseph Fourier. A time series may be expressed as a series of fundamental harmonics, combinations of which are referred to as Fourier series.

Hamer–Soper Model A process-based model for predicting the spread of a disease based on analyzing the stocks of susceptibles, infectives, and removals in a population.

Harmonics A harmonic of a time series is a component frequency that is a multiple of the fundamental frequency. If the frequency is f, the harmonics have frequency $2f$, $3f$, $4f$, etc.

Power Spectra A plot of the percentage of the total variance accounted for versus frequency. 'Power' represents the percentage of total variance accounted for at each frequency. Different power spectra indicate different types of time series.

Space–Time Forecasting Models The evolution of variables over time and space can be forecast using statistical models based on regression analysis, the variables' past values, lagged spatial diffusion effects, and lagged exogenous or explanatory variables. Models are generally known by their acronyms, for example, STAR, STIMA, STARMA, STARIMA.

Spatial Forecasting Analyzing spatial patterns based on past trends to use this information predictively.

Spectral Analysis This distinguishes three components of a time series: the trend (long-term change), cyclical or oscillating functions, and random or irregular components. It breaks down the series into aggregate or simple waveforms described by a series of sine and cosine waves.

Stationary Series A time series with no systematic change in the trend, no systematic change in variance, and removal of strictly periodic variations.

Time Series A sequence of data points for a variable, typically measured at successive times, usually at uniform time intervals. Time series analysis comprises methods to understand such time series and to make forecasts (predictions) of their future pattern.

Wave Train A succession of waves in a time series, usually spaced at regular intervals.

Introduction

Geographical research on spatiotemporal change has focused both on the development of spatially located variables over a relatively short time period and on forecasting the future using various techniques of time series analysis, such as Fourier analysis and spectral analysis. The time series analyzed by geographers are usually affected by a diverse range of factors and so cannot be predicted exactly. Future values of a time series are regarded as having a possible range of values that are conditioned by knowledge of past values. These time series are said to be stochastic, that is, their prediction contains a random element which can be estimated using probability theory. If exact prediction is possible, the time series is said to be deterministic.

Time series analysis has four basic objectives:

1. description of the main properties, usually after plotting the data in graphical form;
2. explanation – when observations are taken on two or more variables over time it may be possible to use variation in one time series to explain variation in another series, perhaps by using regression techniques;
3. prediction, by using time series to forecast future values; and
4. control, as in quality control of a manufacturing process by using time series for monitoring purposes.

Spectral Analysis

Simple graphical description of time series can be performed by calculating a moving mean that smoothes out sudden changes in values and portrays longer term

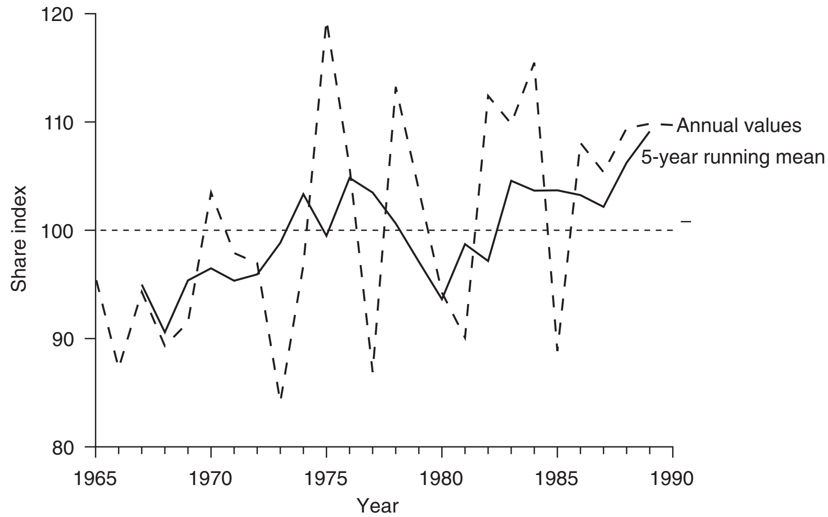

Figure 1 A moving mean.

trends (**Figure 1**). If a time series is stationary, that is, with no systematic change in the trend, no systematic change in variance, and removal of strictly periodic variations, then spectral analysis can be employed to examine the components of the time series.

Spectral analysis distinguishes three components in particular:

1. The long term or trend component: this is the broad, smooth undulating motion of the series over a relatively long period of time. This can often be partially isolated using a moving mean.
2. Cyclical or oscillating functions: these move about the trend, often exhibiting a seasonal effect or local variations.
3. Random or irregular components: these are revealed when cyclical and trend functions are removed from the time series. This is essentially a residual component (**Figure 2**).

These components generally represent disaggregations of a series into its trend, cyclical and seasonal elements, and random or white noise elements. A typical example is a time series of unemployment data, which may have a clear cyclical function reflecting seasonal factors. This type of discrete series (usually based on regular monthly data) can be modeled by breaking down the series into aggregate or simple waveforms described by a series of sine and cosine waves. Therefore, a time series may be expressed as a series of fundamental harmonics, combinations of which are referred to as Fourier series. These series can be fitted to a time series using least squares techniques. This enables the variance in the time series to be disaggregated into harmonics of different frequencies, each of which accounts for a proportion of the total variance. Principally, this process examines the extent to which variance can be attributed to waves of various frequencies.

Further description of time series data may be performed using a diagram of the power spectrum: a plot of the percentage of the total variance accounted for versus frequency. Here 'power' represents the percentage of total variance accounted for at each frequency. Different power spectra indicate different types of time series (**Figure 3**):

1. If the time series are a set of random values, a horizontal power spectrum should be generated.
2. For a cyclical time series a cosine angle with a frequency corresponding closely with the main periodicity in the time series accounts for the largest proportion of variability in the series. For example, in work by Haggett and associates on unemployment data in Southwestern England, power spectra revealed periodicities of 3, 4, 6, and 12 months corresponding to different seasonal cycles. The power spectra have a declining or damped waveform. The highest peak in the spectrum corresponds with the basic periodicity or fundamental wavelength. Other peaks correspond to waves with frequencies that are harmonics of the fundamental wavelength. Low frequencies produce waves with long periodicities while high frequencies produce waves with short periodicities.
3. Power spectra can be calculated for autoregressive time series (series where prediction about a location at time t is dependent only on its previous history at time $t-1$). It is possible to measure the correlation between observations at different time–distances apart in the series. Therefore, this is measuring 'self correlation' between one data point in the series and another data point in the same series to see if successive or near neighbor observations are correlated. This can be

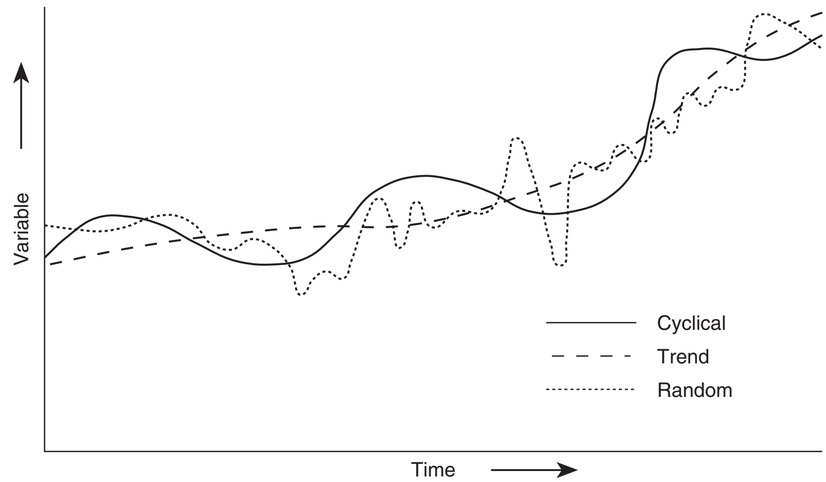

Figure 2 Three components distinguished using spectral analysis.

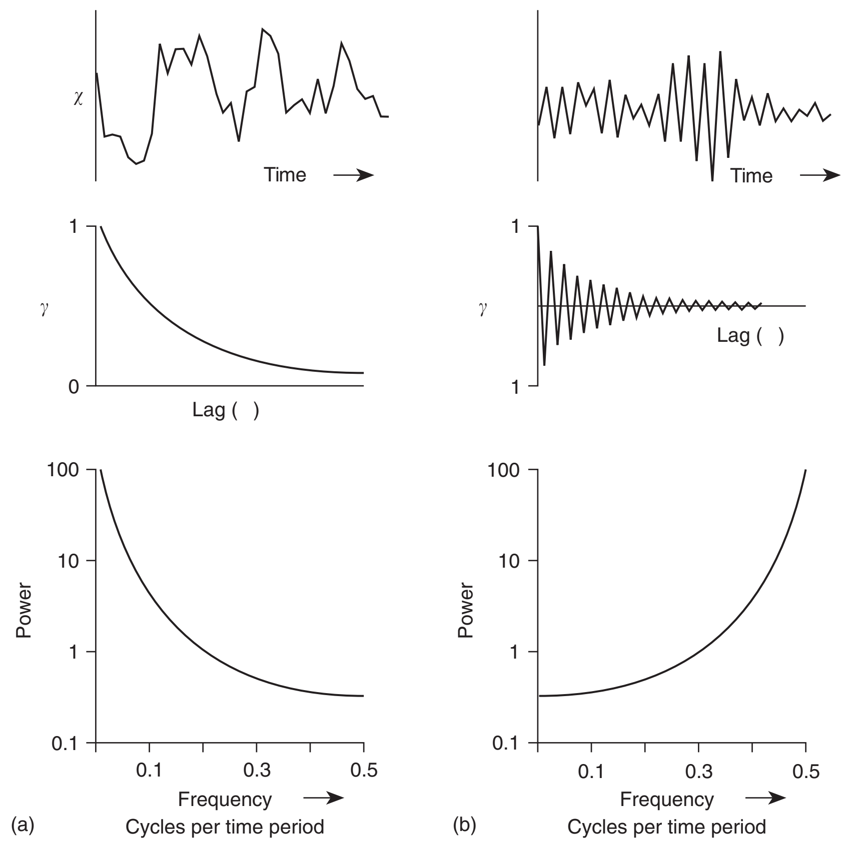

Figure 3 Power spectra for autoregressive time series: (a) series positively autocorrelated and (b) series negatively autocorrelated. From Haggett, P., Cliff, A. D. and Frey, A. (1977). *Locational Analysis in Human Geography*. London: Edward Arnold.

calculated in the same manner as the Pearson product–moment correlation and generates an autocorrelation coefficient (r_k), where $k =$ the 'lag'. If t_1 to t_{10} has the series of values 17, 27, 23, 31, 42, 31, 37, 48, 49, 46, and the lag is 3, then these values would be correlated with 31, 42, 31, 37, 48, 49, 46… when calculating r_3. Autocorrelation for a time series may be plotted on a correlogram, which shows distinctive forms for the various types of series (**Figure 4**). For autoregressive series, if r_k is large and positive this suggests that adjacent values are similar and the time series varies smoothly through time. This smooth variation is approximated most closely by waves of low frequency. If r_k is large and negative the time

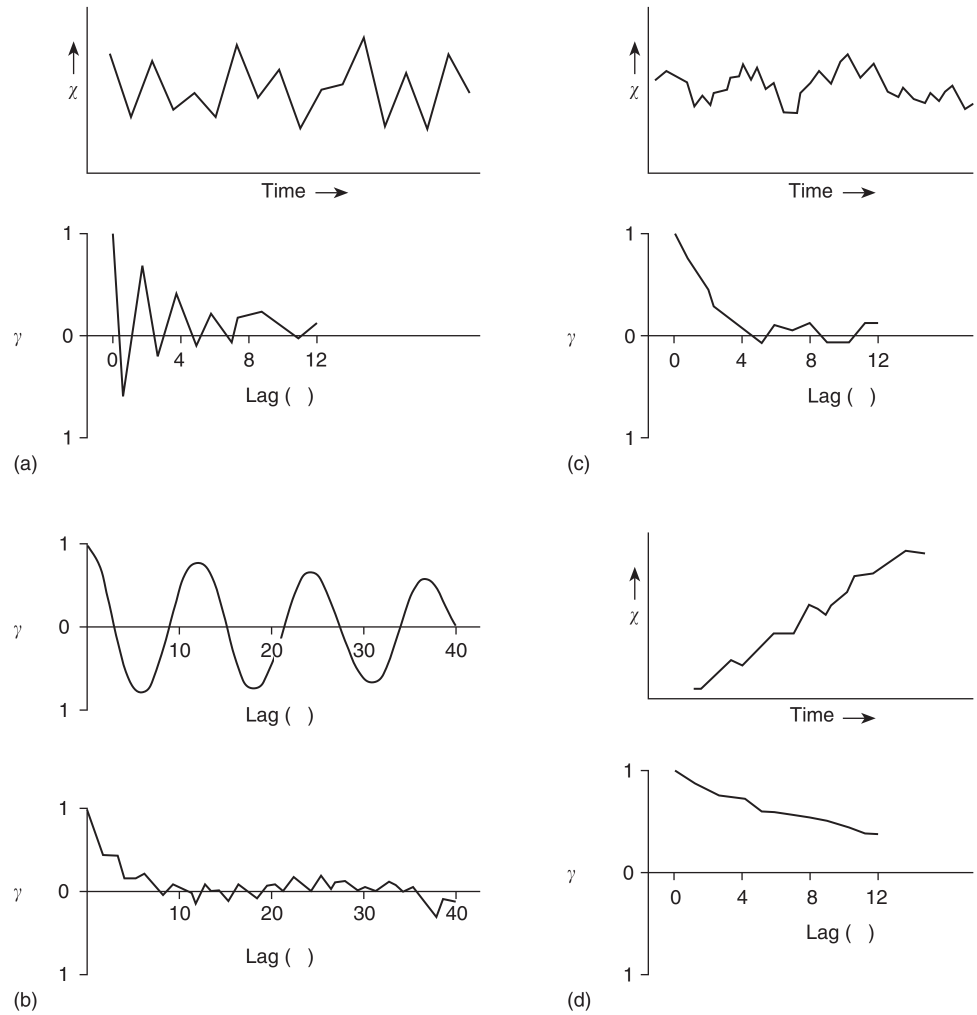

Figure 4 Correlograms: (a) alternating time series; (b) cyclical series, top part: unadjusted and bottom part: seasonally adjusted; (c) time series with short-term correlation; and (d) nonstationary time series.

series alternates rapidly, producing a spiky appearance. These short oscillations can be fitted most closely by waves of short wavelength. Positive autocorrelation is associated with power spectra at low frequencies, and negative autocorrelation is associated with power spectra at high frequencies.

In modeling a time series, a curve can be fitted to the data by making use of the basic forms of power spectra: flat for a random series (or white noise), smooth for a positively autoregressive series, and spiky for a series with cyclical components.

Spatial Forecasting

One of the principal aims of time series analysis when performed by geographers is to make forecasts of future spatial patterns, ranging from just a few days ahead in the case of meteorological phenomena to perhaps 5 or 10 years in the case of economic geography. This requires a clear understanding of the processes that produce the spatial patterns, and then identifying patterns based on past trends to use this information predictively. In predicting values of a variable for a particular location, the past record of that variable at other locations can be taken into account in various ways. In human geography the chief applications have been regional forecasts for unemployment and the incidence of disease.

Pioneering work on regional forecasting of unemployment was performed by Peter Haggett and associates at the University of Bristol in the 1970s. This involved the development of procedures for estimating how trends in time series data would progress through time and space. The focus of this approach is $x_i(t)$, a time series for variable x at the ith location, which also has values for that variable for other locations. A set of time series for different locations can be compared using cross spectral analysis. For example, by comparing whether the time series for two locations are in phase or whether one series leads or lags another. This can establish the intensity of the relationship between them for various lags and leads. Behavior at a location that consistently leads others can be used as a predictor of behavior in other locations.

This largely descriptive and exploratory approach can be extended into the area of modeling by using the equation

$$y_t = B_k x_{t-k} + e$$

where y_t = time series in a region x_t, $t = 1, \ldots, T$, x_t = the national series, e = a random disturbance term, and the vector of B_k coefficients records the regional responses at each lead or lag (k). Autocorrelation can affect B_k and test statistics, but this can be calculated and incorporated into the model.

The pattern of the variable being analyzed can be disaggregated into three components for which equations could be applied to each location, deriving coefficients using multiple regression analysis. For research on regional unemployment patterns, the three components are:

1. Aggregate cyclical component: that part of a region's unemployment rate resulting from cyclical variations at the national level. Sensitivity $(a_{ij}) = 1$ if the aggregative regional component behaves in the same way as the national pattern of variation. More prosperous areas with low unemployment rates tend to be cyclically insensitive and vice versa.

$$A_{jt} = a_j U_t = l_j$$

   where A_{jt} = aggregative cyclical component in region j at time t, U = national unemployment rate, l_j = length of lead or lag between the national and regional series, and a_j = sensitivity of region j to national cyclical variations.
2. Structural component: this is the component peculiar to each region, which reflects long term disturbances in the labor market.

$$S_{jt} = C_j + b_{jt} + d_{jt}^2$$

   where S_{jt} = structural component of regional unemployment in region j at time t, which is measured in terms of a quadratic time trend, C_j = structural component in the initial time period, and b_j and d_j = coefficients of the quadratic time trend.
3. Regional cyclic component: this is given by

$$R_{jt} = A_{jt} + S_{jt} + R_{jt}$$

   where R_{jt} = level of unemployment U_{jt} in region j at time t. If $a_j = 1$ and $C_j = b_j = d_j = R_{jt} = 0$, then no regional unemployment problems exist.

In theory these models enable areas to be classified on the basis of their performance with respect to the three components and there are applications in terms of regional planning and forecasting. However, the models assume that macroeconomic conditions are stable over a period of time, which is not true in any period of economic instability. To allow for this, more complex rules can be created and added to the basic equations, for example, using techniques that model directly the changing statistical relationships in a time series. Nevertheless, lack of stability between interregional lead–lag relations in one economic cycle and another can make forecasting very difficult. Hence, geographers have tended to use predictive models in other applications, notably in forecasting epidemic waves for diseases.

Space–Time Forecasting Models

The evolution of variables over time and space can be forecast using statistical models. Using regression analysis, the future values of a variable are predicted on the basis of its own past values (i.e., it is autoregressive), lagged spatial diffusion effects, and lagged exogenous or explanatory variables. A simple space–time model, taking account of both the spatial dependencies of a variable and its past behavior, takes the form

$$x_{i,t} = f(x_{i,t-k},\ x_{j,t-k})$$

where $k = 1, 2, \ldots$ and is a predefined time interval, t = time, $x_{i,t}$ = a variable's current value at location i at time t, and $x_{j,t-k}$ = a variable's past value at location j at time $t-k$. In this case a variable's present value at a particular location is defined as a function of its past value at that location and its past value at another location (**Figure 5**).

Several variants of space–time forecasting models have been developed to forecast the evolution of variables over both time and space. They are generally known by their acronyms.

1. STAR, a basic space–time autoregressive model: this considers n areal units for T time periods, analyzing a variable y. For area i at time t, the variable $= y_{it}$. The model expresses the dependence of the values of y in area i at time t upon values of the variable in other areas in the previous time period (i.e., dependence that is first order autoregressive in time). Alternatively, a regressor variable can be used, for example, if a regressor variable has the same value for each area it could be used to isolate local effects.

$$E(Y_{it}|\text{past history}) = ay_{i,t-1} + w_{ij}y_{j,t-1}$$

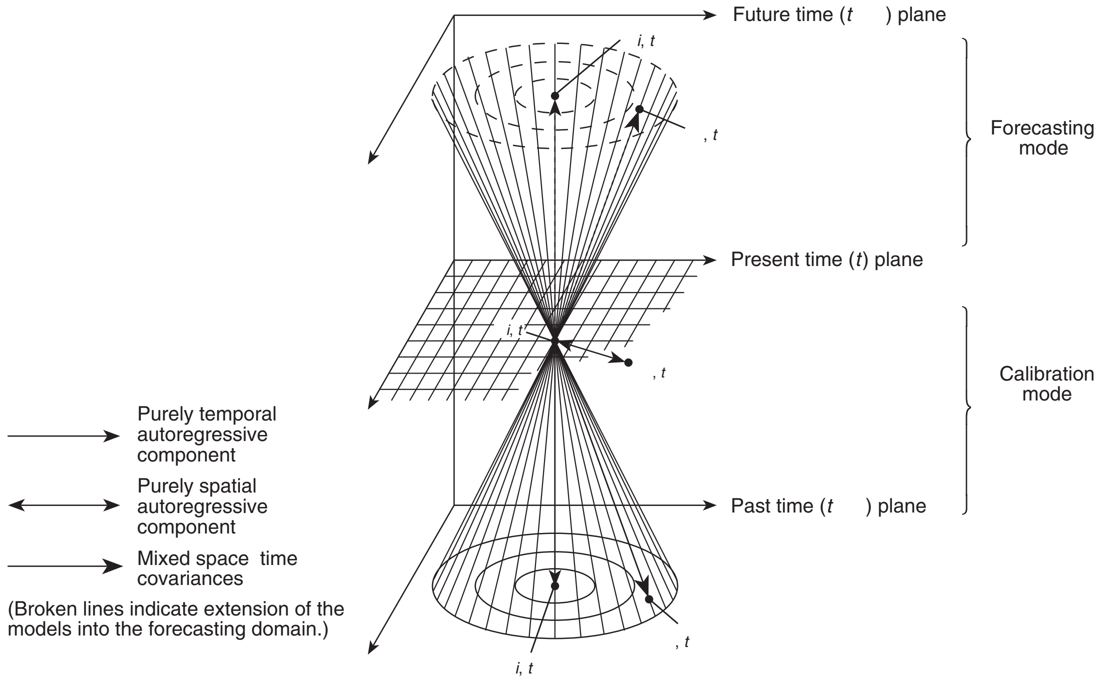

Figure 5 A space time Model. From Haggett, P., Cliff, A. D. and Frey, A. (1977). *Locational Analysis in Human Geography*. London: Edward Arnold.

where $E(Y_{it}\,|\,\text{past history})=$ the expected value of Y_{it} conditional upon its past history, and $w_{ij}=$ structural weights as employed when calculating a spatial auto correlation coefficient.

2. STIMA, a space–time integrated moving average model, which induces random 'shocks' into the fore cast, for example, 'chance' closures of factories. Changes in the variable under consideration (y_{jt}) are viewed as a function of a series of shocks to the system arriving randomly in time and space.
3. STARMA, a space–time autoregressive moving average model, which is based on the equation:

$$Y_{jt} = a_1 y_{j,t-1} + a_2 y_{j,t-2} + e_{jt} + b_1 e_{j,t-1} + b_2 e_{j,t-2}$$

where a variable (y) in a region j at time t is predicted by regression on its own earlier values and by its delayed response to the impact of random shocks e_{jt}, $e_{j,t\ \ 1}$ and $e_{j,t\ \ 2}$; and b_1 and $b_2=$ moving average coefficients.
4. STARIMA, a space–time autoregressive integrated moving average model, which has been extended to include seasonal behavior. It expresses $z_i(t)$, the ob servation of a random variable at site i, $i=1, 2, \ldots, N$, and time t as a weighted linear combination of past observations and errors which may be lagged both in space and time.

Applications of these models are restricted by their complexity and decisions regarding which orders of temporal and spatial lags to include. A common aim is to have as few parameters as possible to estimate in a model providing this gives an adequate representation of the time–space process.

Predicting Epidemic Return Times

For nearly two decades geographers' probabilistic simu lations of the spread of disease have been an important contribution toward gaining an understanding of the genesis and spread of disease. Based initially on spatial diffusion models developed in the 1950s by Swedish geographer Torsten Hagerstrand, increasingly it has ac companied research on cultural, social, and political contexts for various diseases, illustrating the multifaceted approaches to the same fundamental problem that can be made from a broadly geographical perspective. This re search has combined description, prediction, and inter diction (intervention to control diffusion of a disease).

Work on the spread of diseases in human communities by Andrew Cliff, Peter Haggett, and Matthew Smallman Raynor developed a detailed understanding of the rela tionship between size of community and temporal spacing of epidemics. Three types of community can be distinguished (**Figure 6**):

1. a community large enough to sustain the virus and so having a continuous record of infection (with regular advance and decline of an epidemic, known as a Type I wave train);

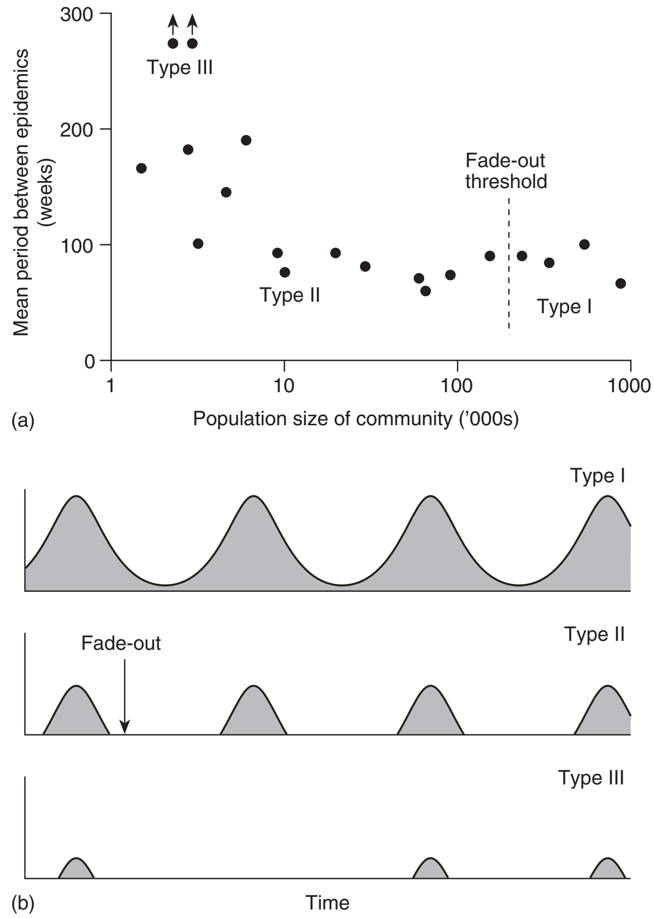

Figure 6 Epidemics and community size: (a) impact of population size of towns on the spacing of measles epidemics and (b) characteristic epidemic wave trains for town types I III in (a). From Cliff, A. D. and Ord, J. K. (1995). Estimating epidemic return times. In A. D. Cliff, P. R. Gould, A. G. Hoare & N. J. Thrift (eds.) *Diffusing Geography*. Institute of British Geographers Special Publications Series, 31, pp. 135-167. Oxford and Cambridge, MA, Blackwell Publishers Ltd.

2. a community too small to sustain the virus (epidemics can completely fade out on a regular basis: Type II waves); and
3. a community with a very small population (irregular recurrence of epidemics that complete fade out: Type III waves).

In the latter two cases, epidemics can only recur when the virus is reintroduced by the influx of infected indi viduals from reservoir areas of Type I. This has impli cations for the development of strategies to restrict the disease, based on vaccination procedures, but the strat egies rely on an ability to forecast the space–time oc currence of epidemic outbreaks. Accurate forecasts can enable targeted vaccination to reduce the susceptible population in particular localities so that it falls below the threshold at which an epidemic can occur and spread.

Various forecasting models have been developed, both for a single city or region and for multiple regions. They fall into two categories:

1. Process based models: these take the stocks of sus ceptibles (members of the population who might catch the disease in question), infectives (those who have the disease), and removals (those no longer able to transmit the disease). They also formalize the transi tions that can occur between these subgroups and use these as the basis for studying the size and spacing of epidemics. They include the Hamer–Soper model, which detects the start of epidemics quite accurately but tends to give inaccurate estimates of total cases, making it difficult to estimate the amounts of vaccine and levels of healthcare provision required. These inaccuracies come from the deterministic nature of the model which damps successive epidemic waves, which is not observed in practice.
2. Time series methods: the past history of an epidemic in the area is used to predict the future with an autoregressive/moving average (ARMA) framework. However, these tend to miss the start of epidemics thereby restricting their practical utility.

Pioneering work on forecasting epidemics, using the case study of measles epidemics in Iceland, demonstrated that significant gains in accuracy could be achieved by adding extra complexity to the models. This led to the development of a process based model based on size of the susceptible population (*S*), the infected population (*I*), and the removed population (*R*) or those who have contracted measles but can no longer transmit it to others because of recovery, isolation, or death: and hence the term SIR model.

The SIR model allows for four types of transition:

1. a susceptible becomes a removal,
2. an infective becomes a removal,
3. a susceptible 'birth' occurs as children are born or in migration occurs, and
4. an infective in migrates (no out migration is assumed).

In these transitions it is assumed that the infection rate is proportional to the product, *SI*; that the removal rate is proportional to *I*, and the birth and immigration rates are constant.

If transition *i* occurs at the rate of $r_i (i = 1, 2, 3, 4)$ then in a small time interval $(t, T + \delta t)$ the probability of transition *i* occurring is $r_i \delta t + o(\delta t)$, where $o(\delta t)$ is a term of smaller order than δt (and assuming that all events are independent and depend on only the present state of the population).

This model can be developed to estimate the probability of the extinction of an epidemic and the time between epidemics. Simulation experiments can be used to evaluate the performance of estimators in the model, for example, in work by Cliff *et al.*, this was based on the 16 main measles epidemic waves that have affected Iceland since 1945. This revealed certain weaknesses in the model, such as its sensitivity to accurate interpretations of epidemic commencement and ending, but it did reflect the general 3 to 5 year cycle of measles epidemics in Iceland. There was some underestimation of epidemic return times, especially for very small settlements where return times are more variable and reflect the chance arrival of the measles virus at times when the susceptible population has grown sufficiently to sustain an epidemic. It was concluded that incorporating seasonal and inter regional components would improve forecasting of return times. Such estimation has become increasingly significant as the World Health Organization's Expanded Programme of Immunization pushes measles into a diminishing part of the world's surface. The research also revealed the importance of a threshold value for a susceptible population. Below the threshold an infectious disease becomes naturally self extinguishing. So if vaccination can be employed to reduce the susceptible population below a critical mass then biological processes may work to end an epidemic. The critical threshold population for measles is between 300 000 to 500 000 people.

In recent years, geographers have focused attention on the spread of acquired immune deficiency syndrome (AIDS), again frequently using relatively isolated island communities as 'laboratories' to study the epidemic, for example, concentrating on the Caribbean where the majority of AIDS cases reported have also been mainly on some of the smaller islands. This research has combined time series analysis with application of spatial diffusion models. The latter are based on probabilistic simulations of the spread of disease through direct person to person transmission, known as expansion or contagious diffusion. In simulating the diffusion of the disease through time and space, key variables can be incorporated based on results from multiple regression analysis. Typical variables are employment characteristics, incidence of drug use, migration rates, incidence of sexual diseases, and numbers of tourist visitors. Simulation can then be applied to the logistic growth curve of the disease.

One major advance in developing these simulation models from the late 1980s is the expansion method, which investigates the variation or drift in the temporal growth of a disease, such as AIDS, in response to population density. This technique can be applied to orderly introduction of complexities into simpler mathematical formulations, modeling variations in selected parameters, and investigating whether models of the spread of a disease vary in different contexts.

The expansion method permits various hypotheses to be tested, related to the distribution of disease with respect to population density, noting that different mixes of mechanisms propagating diseases operate in different geographical environments. For AIDS this has enabled variables on levels of urbanization, housing types, crime levels, ethnic minorities, employment characteristics, student population, male headed households, and single person households to be investigated. For example, AIDS 'hotspots' in San Francisco have been shown to be related to the proportion of males aged 20–44 years, multiple male households, a poverty index, and homosexual intravenous drug users (IVDUs). The spatial pattern is closely associated with age, though with significant variation between ethnic groups. The expansion method enables predictions of the spread of the disease to be made, showing a division between the rapidly evolving epidemic in the east of the city and the slowly evolving epidemic in the West. In the east a continued intensification of the heterosexual IVDU epidemic is predicted, with a secondary focus to the south, underlain by a growth rate of 170 new AIDS cases per month. Ethnicity and low socioeconomic status are important factors in AIDS among heterosexual IVDUs.

Conclusion

Attempts by human geographers to combine the study of change through both time and space have led to im portant advances in understanding the spread of disease. Applications to economic variables have often been less successful. The need for stability in broad economic parameters in order for time series analysis methods to be viable is one restriction. In addition, successful pre diction requires the principal influences on a system to be known, explanations to be complete rather than par tial, and for the system in question to be closed. Many human systems do not fulfill these requirements and hence space–time models have to cope with lack of completeness and closure, making prediction more problematic. The 'solution' with respect to predicting the occurrence and spread of disease has been to incorporate stochastic approaches into disease modeling. The most successful forecasting has tended to occur at a local (micro) level when information on individuals is avail able, and understanding can be enhanced through use of both quantitative and qualitative information.

The incorporation of time series analysis in the rap idly evolving area of geographic information systems (GISs) has been relatively limited, partly because of the complexity of handling both temporal and spatial infor mation simultaneously. The greatest advances have ten ded to come from spatial simulation modeling, building on initial work on spatial diffusion processes in the 1950s. However, GIS based applications of spectral and Fourier analysis to relatively short term forecasting in the retail sector have produced effective and commercially valu able predictions.

See also: Disease Diffusion; Monte Carlo Simulation; Regression, Linear and Nonlinear; Simulation; Space-Time Modeling; Spatial Autocorrelation; Spatial Expansion Method; Spatially Autoregressive Models; Statistics, Overview.

Further Reading

Chatfield, C. (2003). *The Analysis of Time Series* (6th edn.). London: Chapman and Hall.
Cliff, A. D., Haggett, P., Ord, J. K. and Versey, G. R. (1981). *Spatial Diffusion: An Historical Geography of Epidemics in an Island Community*. Cambridge: Cambridge University Press.
Cliff, A. D., Haggett, P. and Smallman Raynor, M. R. (1993). *Measles: An Historical Geography of a Major Human Viral Disease from Global Expansion to Retreat, 1840 1990*. Oxford: Blackwell.
Cliff, A. D., Haggett, P. and Smallman Raynor, M. R. (1998). *Deciphering Global Epidemics: Analytical Approaches to the Disease Records of World Cities, 1888 1912*. Cambridge: Cambridge University Press.
Cliff, A. D., Haggett, P. and Smallman Raynor, M. R. (2000). *Island Epidemics*. Oxford: Oxford University Press.
Gould, P. R. (1993). *The Slow Plague: A Geography of the AIDS Pandemic*. Cambridge, MA: Blackwell.
Haggett, P. (1994). Prediction and predictability in geographical systems. *Transactions of the Institute of British Geographers* 19, 6 20.
Haggett, P., Cliff, A. D. and Frey, A. (1977). *Locational Analysis in Human Geography*. London: Edward Arnold.
Peuquet, D. (2003). *Representations of Space and Time*. New York: Guilford Press.
Robinson, G. M. (1998). *Methods and Techniques in Human Geography*. Chichester: John Wiley.
Shumway, R. H. and Stoffer, D. S. (2006). *Time Series Analysis: Its Applications with R Examples* (2nd edn.). New York: Springer Verlag.
Smallman Raynor, M. R. and Cliff, A. D. (2001). Epidemic diffusion processes in a system of US military camps: Transfer diffusion and the spread of typhoid fever in the Spanish American War, 1898. *Annals of the Association of American Geographers* 91, 71 91.
Smallman Raynor, M. R., Cliff, A. D. and Haggett, P. (1992). *The London International Atlas of AIDS*. Oxford: Blackwell.
Smallman Raynor, M. R., Haggett, P. and Cliff, A. D. (2004). *World Atlas of Epidemic Diseases*. London: Hodder Arnold.
Warner, R. M. (1999). *Spectral Analysis of Time Series Data*. New York: Guilford Press.

Relevant Websites

http://www.statoek.wiso.uni goettingen.de/veranstaltungen/zeitreihen/sommer03/ts r intro.pdf

http://www.r project.org/
A free software environment for statistical computing and graphics that runs on a wide variety of UNIX platfoirms, Windows and MacOS. For time series analysis with R, see:

http://www.statsoft.com/textbook/sttimser.html
A general introduction to time series analysis, including ARIMA models and spectral analysis

http://www.statistik.mathematik.uni wuerzburg.de/timeseries/
This is a first course on time series analysis with a solution of statistical procedures used in general practice, including the statistical software package SAS Statistical Analysis System.

http://www.omatrix.com/stsa.html
This is a time series analysis toolbox.

http://www.lse.ac.uk/collection/cats/
This is an introduction to the Centre for the Analysis of time series at the London School of Economics.

Time-Space Diaries

T. Schwanen, Utrecht University, Utrecht, The Netherlands

Glossary

Activity Episode A discrete activity participation characterized by a unique combination of activity dimensions for which information is collected.

Memory Jogger A simple and small-sized form that diarists can take along during the recording period and on which they can scribble information which can later be used to fill out the actual diary.

Reporting Fatigue The decline in the completeness and accuracy of the information during the recording period. Indicators of fatigue are lower numbers of reported activities or trips in later stages of the recording period and an increase over time in attrition or the tendency among respondents to quit the study.

Representativeness The degree to which the respondents who completed and returned the time–space diary resemble the research population with respect to one or several characteristics.

Response Rate The ratio of people who completed and returned the time–space diary to the number of people in the sample or originally approached.

Time Geography An approach that seeks to understand how the physical world evolves developed by Torsten Hägerstrand from the 1960s onwards as an alternative to the increasing fragmentation and specialization of (human) geography.

Time–Space Path A concept from time geography that denotes the uninterrupted movement of a human body or any other physical entity through space and time.

Introduction

Time–space diaries have long been used in various disciplines, including sociology, anthropology, urban planning, and geography. The earliest examples can be found in sociology, where they were first employed in the early twentieth century. Urban planners started using time–space diaries in the 1960s to better understand the functioning of urban systems. While there are some examples of time–space diaries in geography before 1970, their use expanded enormously after the introduction of Hägerstrand's time geography to Anglo American geography. Through concepts such as the space–time path and prism, time geography provided a conceptual apparatus for the analysis of, and could not do without, time–space diaries. Time–space diaries have played a role in geography ever since, but there has been an upsurge in interest in this research method since the early 1990s. To a certain extent this reflects a shift in transportation policies from a focus on supply to demand side measures, which required new and detailed information on travelers' time–space trajectories. This revival was also due to advances in geographic information system (GIS) and computational capabilities that greatly facilitated the analysis of time–space diaries. Recently, cultural geographers have (re)discovered the time–space diary as a means for capturing the continuous flow of everyday practices.

The most important reason for the enduring utilization of time–space diaries is that they offer a window on everyday life. They provide information on what people do, where, and when, which can be used to derive indicators of quality of life and social and economic well being and which constitutes basic input for forecasts of the use of facilities and services in transportation, retail, recreation, and other sectors. At least five specific uses of time–space diaries in human geography can be distinguished. At the aggregate level of city populations, time–space diaries provide essential information about temporal variations in the social geography and ecological structure of cities. Geographers have also utilized time–space diaries to explore the relations between individual behavior, accessibility, and the physical environment. For instance, time–space diaries provide invaluable information for policymakers' attempts to curtail urban sprawl and reduce auto travel. More recently, geographers have turned to time–space diaries to investigate the ramifications of new information and communication technologies (ICTs) for everyday behavior. Additionally, time–space diaries enable the reciprocal relations between individual action and social structures to be explored. Some of the best examples of this use can be found in feminist geography, where time–space diaries have effectively revealed how gender regimes are bound up with the everyday lives of men and women in different localities. Finally, time–space diaries can be used to delegate power in the research process to respondents and to stimulate reflexivity in subsequent phases of a research project.

Nonetheless, it is rather difficult to define what a time–space diary is, what it measures or should measure. This is because a wide variety of diaries have been developed and employed. The remainder of this article attempts to systematize this variety by outlining a series of design dimensions determining what and how information on individuals' paths through time and space is collected. In doing so, it will also discuss a series of issues pertinent to the use of existing sets of time–space diaries or the collection of new diaries.

Design Dimensions

In general terms, a time–space diary is a document in which an individual has recorded a series of events and acts in which she has been involved and/or her impres sions thereof during a certain period of time. As this definition implies, only documents where respondents have actively recorded information are addressed in this article; fully researcher administered or computer generated records are not considered, in part because they have not been applied frequently in geographical inquiry. Self administered time–space diaries come in many variants, which can be classified on the basis of the following six interdependent dimensions: function, con tents, medium, format, respondents, and time frame. Researchers seeking to answer a set of research questions are encouraged to consider each of these dimensions when selecting an existing diary data for analysis or collecting their own diaries.

Diary Function

Two ideal typical approaches to time–space diaries, which to some extent function as the extremes of a continuum, can be distinguished on the basis of the function(s) diaries have. On the one hand, there is a more extensive approach underpinned by traditional research epistemology: time–space diaries constitute the main data collection instrument, although brief questionnaires or structured interviews are normally used to collect supplementary information; diaries are used to extract more or less objective representations of an external reality; contact between researcher and researched is relatively limited; the respondent has a more passive role because the researcher tends to control what and how information is recorded; and the information contained in the diaries is processed and analyzed with quantitative techniques (including GIS) to detect regularities and systematic relations between variables.

On the other hand, there is a nascent, more intensive approach to time–space diaries, primarily associated with recent developments in cultural geography and rooted in feminist epistemologies and nonrepresentational ap proaches to 'doing' research. Here, the use of time–space diaries is embedded in a mixed methods endeavor and complemented with interviews or other methods relying on intensive contact between researcher and researched. The function of time–space diaries is also different: they are not so much used to extract objective representations to be generalized across respondents, but instead to emphasize the plurality and multiplicity of experiences of everyday life. Additionally, they function to trigger respondents' reflexivity about less than conscious ex periences and feelings and so open up spaces for con templation and dialog to be explored and inhabited during subsequent phases in the research process. Time–space diaries also appeal to cultural geographers because they allow the role of researchers in what and how things are recorded and selected for further analysis to be minimized. Participants thus retain much more control over the research process, although the inter respondent comparability of the collected information tends to be rather low. It should be noted that the use of time–space diaries in cultural geography is still in its infancy; the conventional, more extensive approach to time–space diaries is more widespread in contemporary human geography. However, the use of diaries in cultural geography is expected to grow in the near future, as this method affords a closeness between actual practices and experiences and their recording that interviews alone do not allow.

The distinction between these two approaches should not be considered absolute. There are some examples of hybrid approaches in the literature, whereby traditional researcher directed diaries are combined with, and used as input for, interviews to provide thicker descriptions of the conduct and meanings of everyday practices.

Diary Contents

The question of what information should be collected through diaries has attracted most attention among re searchers associated with extensive approach identified above and more specifically in the transportation and time use research literature. A first important distinction drawn there is that between the trip diary and the activity diary. In the former, a trip or movement between two spatial locations is the basic event for which information is recorded. Such characteristics of trip as their purpose, origins and destinations, departure and arrival times, and transportation mode(s) used are collected, but no de tailed information about the activities people engage in on specific locations is recorded. In contrast, activity diaries gather information about activity episodes. While they continue to be used widely in transportation ana lysis, trip diaries have increasingly been displaced by activity diaries, in part because the latter are more suc cessful in capturing short, non home based trips. This article therefore concentrates on activity diaries.

At a minimum, activity diaries collect information on the following dimensions:

1. what: the activity content;
2. when: the clock time(s) that an activity episode starts (and ends);

3. where: the spatial location where the activity takes place; and
4. Technology use: this usually concerns the transpor tation mode(s) used to access the place where the activity occurs. More recently the use of ICTs to perform an activity episode is also being recorded more frequently.

There are several additional dimensions for which in formation is often or sometimes collected:

1. with whom: the person(s) involved in the event;
2. the amount of money spent during the activity episode;
3. space–time flexibility: the extent to which the episode could have been conducted at other times and places;
4. the degree to which activities are planned or not; and
5. the respondent's feelings and emotions during epi sodes, such as the levels of enjoyment or stress experienced.

Of these dimensions, activity content has racked the brains of researchers most intensely. Many different classification schemes for activity content exist, but al most all are based on assumptions about the goals people will attain through completing a set of acts that together make up an activity. Geographers and transportation analysts frequently adopt very general activity categories like paid labor, shopping, childcare, and social visits and have often failed to differentiate between activities con ducted within the home. In contrast, time use researchers have employed a much more detailed, harmonized clas sification scheme since Szalai's multinational time use study in the 1960s that, in updated form, continues to be used in contemporary studies throughout the world. This scheme comprises 41 aggregated activity types, 31 of which are differentiated into 176 more detailed codes. Time use analysis is also useful because it draws atten tion to polychronicity (multitasking) and asks not only about the primary activity type but also about secondary and sometimes tertiary time use. Nevertheless, the use fulness of most time use data for geographical analysis is limited, because only information about generic locations (home, workplace, etc.) tends to be collected. So far only very few diary time use studies have collected infor mation about street addresses.

Diary Medium

Self administered time–space diaries traditionally in volve paper and pencil documents. However, the ways these documents are transferred between researcher and the researched have always varied (**Table 1**). A com pletely mail out/mail back procedure (with written in structions) involves the least resources and yields high levels of inter respondent comparability but often also (very) low response rates, low levels of accuracy and completeness, and rather stringent limitations on the complexity of the questions that can be asked. This complexity and also response rates can be (somewhat) larger if respondents are provided with verbal in structions and motivations via telephone and especially via face to face contact with researchers prior to (and during) the period for which respondents have to record information. Even more preferable from the perspective of response rate, complexity of questions, and accuracy and completeness of the provided information is a pro cedure whereby the paper and pencil diaries are directly collected after the recording period by the researcher, who can check, correct, and/or complete them during a conversation with the respondent. Another procedure involving fewer resources is to let respondents return the diary by surface mail and then have research assistants verify the information in a telephone call. However, be cause of the different social relations between researcher and researched in situations of physical co presence and communication at a distance, the commitment to (de tailed) completion of the diary on the part of respondent is likely to be lower when there is only telephone contact.

In recent years, digital media have become increas ingly important for the construction of time–space diaries. First, web based diaries for logging everyday activities are becoming more popular. Some of these strongly resemble traditional paper and pencil diaries; others exploit the flexibility web based procedures offered for collecting detailed information about the activity scheduling process underlying time–space paths. The Computerized Household Activity Scheduling (CHASE) software developed by Sean Doherty and others is the most well known example of this approach.

Second, wired and wireless technologies are employed more and more to construct virtual and physical time–space paths. Researchers increasingly utilize log files of websites visited and phone calls made to create repre sentations of people's communication patterns. Ad ditionally, such technologies as Global Positioning Sys tem (GPS) and Global System for Mobile Communi cation (GSM) are applied to automatically trace participants' geographical position in the physical world at specified intervals. Respondents just have to carry a mobile device – their own mobile phone or personal digital assistant (PDA) or one provided by the re searcher(s) – and are normally requested to add sup plementary information about activity episodes at a later stage. Research indicates that automatically traced in formation tends to be more accurate and complete and can be obtained with fewer resources than paper and pencil diaries (**Table 1**). However, this mode of con structing time–space diaries requires a certain level of computer and mobile phone literacy on the part of re spondents, making automatic tracing especially attractive in studies of younger people. Because the technology can

Table 1 Qua tat ve assessment of t me–space d ary var ants on se ected cr ter a

	(Traditional) paper-and-pencil diary				*Digital diary*				*Flexible diary with minimum requirements (cultural geography)*
	Fully mail-out/mail-back diary	*Phone/face-to-face contact prior to recording*	*Phone/face-to-face contact upon recording*	*Phone/face-to-face contact prior to & upon recording*	*Web-based diary*	*Logging websites & phone calls*	*Automatic tracing of positions in physical world*	*Photo/video diary*	
Response rate	—	–	–	–/+	–	–/+	+	+	+
Comp ex ty of quest ons to be asked	–	–	–	–/+	–	n.a.	n.a.	n.a.	n.a.
Comp eteness & accuracy of co ected nformat on	–	–	–	–	–	++	++	—	–
Pr vacy & confident a ty ssues	+/–	+/–	+/–	+/–	+/–	—	—	+/–	+/–
Respondent burden	–	+/–	+/–	+	–	+/–	+/–	+	++
nter-respondent comparab ty	+	+	+	+	+	++	++	–	–
Su tab ty for record ng un que or persona exper ences	–	–	–	–	–	–	–	+++	+++
Power & contro n hands of the respondent	–	–	–	–	–	–	–	+++	+++
Techno ogy teracy	+	+	+	+	–	–	–	–	+/–
Resources per record (t me, staff, costs)	++	+	+	+/–	++	+	+	–	–

Not all criteria are equally applicable or appropriate to every study This depends on the nature of the research question(s) and one s epistemological premises — through +++ represent an ordinal scale —=very unfavorable and +++=very favorable n a =not applicable

Adapted and extended from Ettema D Timmermans H and Van Veghel L (1996) *Effects of Data Collection Methods in Travel and Activity Research* Eindhoven E RASS trace each and every place visited (provided that tele communication networks function properly), the privacy of respondents constitutes an important matter of con cern. The literature therefore recommends that re spondents are offered the possibility to switch off recording devices or delete certain information from generated records after the collection period.

Third, the proliferation of digital (and disposable) cameras allows for time–space diaries that do not (only) rely on text and language but (also) foreground alter native experiences of places that may be difficult to ex press in words. They also enable a redistribution of power in the research process, promoting the participation and empowerment of the respondent relative to the re searcher. However, potential disadvantages of self directed photography/video include less precise accounts of the continuity of time–space paths, lower levels of inter respondent comparability, and exclusion of people who cannot operate cameras. For these reasons, (audio)visual diaries have been used in combination with other methods.

Diary Format

This design dimension embodies two related aspects: whether the diary has open or closed questions, and issues involving the layout of the diary. Both aspects are, however, most relevant to the more traditional, extensive approach; cultural geographers tend to restrict diary in structions to a minimum to allow respondents maximum freedom to express their own views and experiences. Cultural geographers' minimal use of instructions bears nonetheless some resemblance to the discussion of adopting open or closed questions among transportation geographers and others. This is because the main reason for using an open question format lies in the enhanced opportunities for letting respondents describe what they are doing in their own words. This tends to reduce the complexity of questions and respondent burden. In con trast, the advantage of using a pre coded scheme for re porting information on activity episodes is that it allows for quicker processing of the diaries by the researcher. Some researchers have, therefore, combined these pro cedures, employing a classification into general categories (e.g., leisure for activity content) and then asking diarists to specify the chosen category using their own words (e.g., watching a movie or exercise in the park).

The closed/open distinction also matters to the re cording of activity times. Time use studies in particular use fixed intervals of 5, 10, or 15 minutes for which re spondents have to write down information on other ac tivity dimensions. An important advantage of using fixed time slices is that the continuity of the record is guar anteed; there will be no time gaps in the diary. Un fortunately, very short activity episodes – many of which belong to the realm of care work – tend to get under reported. Fixed time intervals also make it difficult to differentiate true multitasking (e.g., phoning when driv ing a car) from short consecutive activity episodes (as, e.g., with leisure outings with small children), which may impede the researcher's understanding of the impli cations of ICTs or the gendered nature of caregiving.

With regard to diary layout, various issues can be identified, including the sequencing and phrasing of questions in the diary, as well as the use of fonts, style, color, and symbols. Few general rules can be given other than that the layout should make keeping the diary as easy, clear, and unambiguous as possible for participants. Extensive pilot studies and discussions with test phase participants before actually implementing the diary are, therefore, strongly recommended. For a paper and pen cil diary, researchers can also choose between regular or larger sized paper sheets versus a booklet that can be put in a pocket or purse. The advantage of using booklets is that respondents can easily record activities and experi ences on the move, thereby minimizing the time lag between actual experience and reporting and thus the impact of memory retention. There is some evidence available that the use of such booklets indeed results in more activity reporting. Researchers advocating larger sized sheets have sometimes used memory joggers to assist in the completion of the actual diary at a later stage.

Respondents

Decisions about whom and how many respondents should be recruited depend on one's research questions and epistemological premises. In particular, for those working within mainstream epistemological frameworks, representativeness of the respondents of a larger popu lation is a major issue. These issues have been discussed most extensively in the transportation and time use lit erature; at least three topics can be distinguished. The first of these concerns the number of respondents. There is some consensus in the time use literature that about 2000 randomly selected respondents are required to ob tain outcomes that are at least to some extent represen tative of the larger population. Second, there is some discussion as to which householders should participate in the study. It is now widely accepted that participation of all household members capable of keeping a diary should be targeted so that within household interactions in time–space paths can be investigated. An age based cri terion is often used to determine who is requested to record activities, whereby the minimum ages of 12, 15, and 18 years old seem to have been adopted most fre quently. The drawback of asking all rather than one householder to participate is that respondent burden increases and response rates may drop.

The third issue pertains to the selectivity of the recruited respondents. The most and the least mobile persons or households are usually under represented in time–space diary data. Various strategies to induce very mobile and time pressured individuals to participate are discussed in the literature, including the establishment of a personal bond between researcher and the researched; the use of prompting; (financial) incentives; and sample stratification. One example of the last strategy is to segment on the basis of residential location and then oversample neighborhoods known to house many people with socioeconomic characteristics that correlate positively with mobility (e.g., high average income or many dual earner households). Inducing the least mobile to participate in a time–space diary study seems to be even more difficult; few strategies appear very successful in this respect.

Time Frame

Various time related issues have concerned researchers working with time–space diaries. In addition to the open/closed time interval, these include the collection about information about yesterday or tomorrow, recall period, and the time of year. As the discussion so far has already suggested, the 'tomorrow approach' whereby diaries are given or sent to respondents prior to the day(s) for which information is recorded and the minimization of the recall period are strongly preferred to approaches asking respondents what they have done or experienced some time ago. With regard to time of year, time use researchers have recommended the collection of diaries throughout the year; if that is not possible, recording in October–November (outside holiday season) is preferred because time use in these periods corresponds closely to annual averages.

However, the time aspect discussed most extensively in the literature is the time frame or the duration of the recording period, as this is intimately linked to respondent burden. Although time use researchers often select the week as recording period, geographers and transportation analysts tend to adopt one or two days as the basic time frame, thereby implicitly assuming that this day or these days render(s) an accurate representation of persons' overall time–space paths. This assumption has been criticized by Hanson and colleagues using seven week time–space diary data from Uppsala, Sweden collected in the early 1970s and recently by Axhausen and colleagues with German data (Mobidrive). Both have contended that respondent fatigue effects can be overcome when a personal bond between researcher and the researched exists. While the use of long duration studies enables more detailed insights into the variability and rhythms of everyday life, breadth is traded for depth, as the number of participants is much smaller and the burden placed on respondents considerably larger than in more conventional time–space diary studies. These factors also contribute to the increasing use of GPS and other ICTs in multi day studies, which is likely to grow further in the near future. These technologies may not only reduce respondent burden and resources required for longer duration studies, but may also be used to extend recording periods.

Some Salient Issues

In the preceding discussion, several issues pertinent to the collection and usage of time–space diaries have been identified and discussed. These include concerns about the representativeness of a larger population; completeness and accuracy of the information; the resources needed to collect, process, and analyze time–space diaries; the trade off between universality and interpersonal comparability on the one hand and particularity and suitability for recording personal experiences on the other; the distribution of power and control over recording and selecting information between researcher and the researched; and ethical issues with respect to privacy and confidentiality in general and automatic tracing in particular. In various ways, there appears to be a shift toward more intensive approaches of using time–space diaries. This is evidenced not only in the (re) discovery of such diaries by cultural geographers but also in the use of researcher directed diaries collecting information on more dimensions of everyday life and increasing popularity of longer duration diaries. Consequently, difficult issues of respondent burden and ethical dilemmas about obtrusiveness, confidentiality, and privacy are likely to become even more important than nowadays.

In light of the issues above, researchers seeking to answer a specific set of research questions might want to consider the use of existing diary data as an alternative to new data. The collection of new time–space diaries specifically tailored to their research questions is likely to be the most preferred strategy for those researchers with a particular interest in the multiplicity and particularity of people's experiences of everyday life and/or those keen on giving participants as much control over the research process as possible. The same is probably true of those with ample resources and those wanting to address issues for which existing data do not suffice. In other cases, however, researchers may well be able to answer their questions with existing data. Fortunately, previously collected datasets are increasingly made available to the research community. The number of websites bringing together information about various datasets available is growing (some examples are provided below). Alternatively, one can easily contact those who originally collected a dataset to get access to the data. It is antici pated that the exchange of data and the experiences of working with them will advance the state of the art in the collection and design of time–space diaries.

See also: Diaries (Video, Audio or Written); Feminist Methodologies; Intensive/Extensive Research; Mixed and Multiple Methods; Mobility; Performance, Research as; Quantitative Data; Quantitative Methodologies; Time Geography.

Further Reading

Alaszewski, A. (2006). *Using Diaries for Social Research*. London: Sage Publications.
Axhausen, K. W. (1995). *Travel Diaries: An Annotated Catalogue* (2nd edn.). Innsbruck: Institut fur Strassenbau und Verkehrsplanung, Leopold Franzens Universitat.
Axhausen, K. W., Lochl, M., Slich, R., Buhl, T. and Widmer, P. (2007). Fatigue in long duration travel diaries. *Transportation* 34, 143 160.
Doherty, S. T. and Miller, E. J. (2000). A computerized household activity scheduling survey. *Transportation* 27, 75 97.
Ettema, D., Timmermans, H. and Van Veghel, L. (1996). *Effects of Data Collection Methods in Travel and Activity Research*. Eindhoven: EIRASS.
Gershuny, J. and Sullivan, O. (1998). The sociological uses of time use diary analysis. *European Sociological Review* 14, 69 85.
Hanson, S. and Hanson, P. (1993). The geography of everyday life. In Garling, T. & Golledge, R. G. (eds.) *Behavior and Environment: Psychological and Geographical Approaches*, pp 249 269. Amsterdam: Elsevier Science Publishers B.V.
Hanson, S. and Huff, J. O. (1988). Systematic variability in repetitious travel. *Transportation* 15, 111 135.
Kenyon, S. (2006). The 'accessibility diary': Discussing a new methodological approach to understand the impact of Internet use upon personal travel and activity participation. *Journal of Transport Geography* 14, 123 134.
Kwan, M. P. (1999). Gender, the home work link and space time patterns of nonemployment activities. *Economic Geography* 76, 370 394.
Kwan, M. P. (2000). Human extensibility and individual hybrid accessibility in space time: A multi scale representation using GIS. In Janelle, D. G. & Hodge, D. C. (eds.) *Information, Place, and Cyberspace: Issues in Accessibility*, pp 241 256. Berlin: Springer Verlag.
Latham, A. (2003). Research, performance, and doing human geography: Some reflections on the diary photograph, diary interview method. *Environment and Planning A* 35, 1993 2017.
Stopher, P. and Jones, P. (eds.) (2003). *Transport Survey Quality and Innovation*. Oxford: Elsevier Science Ltd.

Relevant Websites

http://www.timeuse.org
Centre for Time Use Research, University of Oxford.
http://www.surveyarchive.org
Metropolitan Travel Survey Archive, University of Minnesota.
http://www.ivt.ethz.ch
Mobidrive survey, Institute for Transport Planning and Systems, ETH Zurich.
http://www.stmarys.ca
Time Use Research Program, Saint Mary's University, Canada.
http://www.travelsurveymethods.org
US Transportation Research Board, ABJ40 Travel Survey Methods Committee.

Tourism

T. Edensor, Manchester Metropolitan University, Manchester, UK

Overview of History of Tourism

Tourism is often cited as constituting the largest industry in the world, and there can be no doubt of its expansion throughout the twentieth century to the point where it is a major constituent of global flows of people. The venues for tourist related activities have expanded to encompass a vast range of attractions, sites, landscapes, and routes, and the scale at which tourism is practiced by most people ranges from the local to global, and temporally, it encompasses both the everyday and the festive. Similarly, activities understood as touristic have proliferated so that tourism cannot be understood as one kind of practice but incorporates a host of leisure based endeavors emerging out of market segmentation, speciality holidays, and identity based pursuits.

The Grand Tour, the Romantic Gaze, and Colonial Travel

While there is little doubt that tourism is a largely modern phenomenon, there are antecedents from the ancient world – for instance in the travels undertaken throughout the Roman Empire – which should con textualize the widespread urge to tour other unfamiliar places. Nevertheless, the forms of mass tourism which continue to characterize much contemporary practice emerged with the rise of rail and later, air travel, making cheaper travel available to an ever growing numbers of people. Prior to this expansion, early modern travel primarily centered around the 'Grand Tour', through which the offspring of the affluent undertook prolonged periods of travel around the vestiges of the European classical world, notably in Italy, informed them about the values of the Enlightenment and Renaissance which led to them valuing the art, architecture, and philosophy of these earlier times.

The destinations of these middle class European tra velers were extended by the rise of eighteenth century Romantic ideas and tastes which emerged from the recontextualization of certain forms of rural landscape as 'sublime' and 'picturesque'. Such notions emerged out of the popular and influential work of romantic poets, such as William Wordsworth, and expanded to encompass artists and art theorists who foregrounded the com position of ideal 'natural' views, ideas which found a ready audience in the residents of the growing industrial towns and cities of Britain and their nostalgia for a largely imaginary preindustrial rurality. Soon enough commercial excursions were developed, and through the growth of rail travel and tourist infrastructures, it became possible for those in search of romantic and sublime scenes in previously disregarded areas such as the Lake District, popularized by Wordsworth, to satiate their desires by walking, painting, and gazing. This longing for the romantic expanded to incorporate 'wild' attractions in Europe and beyond, and it maintains a persistent hold on tourist motivation and marketing. The organization of these further flung tours were the prelude to, and ac companied the rise of mass commercial travel, organized by travel companies, notably Thomas Cook, who made travel available to a much wider clientele, including working class people.

In addition to this mass tourism, earlier forms of tourism may also be noted in colonialism. Along with the anthropological, geographical, and historical accounts of the colonizing powers, which 'expertly' classified, de coded, and exhibited the natural histories and cultures of the colonized, tourism provided numerous accounts where Western travelers authoritatively reported their experiences whilst traveling through colonized space. These travel accounts and the foci of their interest are, in many cases, strikingly similar to the narratives found in contemporary, postcolonial guidebooks where tourists are advised to look at particular spectacles or 'exoticisms'. Thus the representation of the non Western 'other' in contemporary tourism appears to be informed by these traditions of colonial depiction which continue to re produce colonial assumptions about cultural backward ness and unassimilable difference as well as fear and desire for an imagined alterity. Others point to other resonances between the present and the colonial era in the iniquitous social and economic relations that persist between the West and the rest.

Industrialization and Mass Tourism

The emergence of working class excursions to the British seaside was typically part of this growth of mass tourism. During their still limited but longer holiday periods, towards the end of the nineteenth century large numbers of laboring classes from the north and midland industrial heartlands of the United Kingdom traveled in huge numbers to seaside resorts including Blackpool, More cambe, and Scarborough and those from the Southeast headed for Brighton and Bournemouth, with smaller resorts also being established wherever sand and sea were found. Whereas these venues had typically served a middle class clientele with bathing facilities, promenades, parks, and other pursuits that centered upon healthy recreation and the virtues of taking sea air, they quickly transformed in providing cheap accommodation facilities, large amusement parks, pleasure gardens, drinking venues, and all manners of cheap diversions and entertainments. Thus 'respectable' forms of leisure promoted by the middle classes were squeezed out in favor of the cheap and cheerful pleasures favored by the working classes, as part of the massive development of an often vilified mass popular culture. This development, in turn, drove the middle classes to seek other destinations to maintain distinction, a still prevailing trend in tourism whereby the often maligned popularity of particular destinations produces desires by certain tourists to seek out new and 'unspoilt' locations. In the later half of the twentieth century, a similar phenomenon involving mass tourism arose at a different geographical scale, this time with the advent of cheap air travel which made continental tourism available to the poorer classes who had never before been afforded such opportunities. Consequently, large numbers of Northern European tourists migrated to Mediterranean resorts in Spain and France and more recently, have been attracted to a widening range of beach locations across and beyond Europe, in search of sun, sea, sand, sex, and other hedonistic pleasures, pursuits, and places. Once more, these have been labeled as being of a lower status by middle class tourists in pursuit of distinction, in contradistinction to their own quest for the 'unspoilt', the more culturally refined, and different. These distinctly modern forms of mass tourism continue to attract large numbers of tourists on package holidays and excursions, yet they have been accompanied by the rise of an extraordinary variety of touristic places and practices at various scales and times, postmodern tourism which will now be discussed.

The Globalization of Tourism and the Increasing Diversity of the Tourist Product

The Globalization of Tourism

The range and breadth of tourist destinations have expanded so that we may now appropriately address the subject at a global scale. The proliferation of travel agencies and other tourist retail outlets, much cheaper air travel and accommodation, the diversification of package tours, intensified place promotion and marketing, improved standards of hospitality, standardization of facilities and other technological developments in airport procedures, access to finance and telecommunications make international travel much easier. The time–space compression involved in this growing interconnectedness brings distant spaces and cultures closer together, although such proximity is always under threat by political volatility and economic instability through which destinations may instantly lose their cachet.

The expansion of tourism implicates not only the compression of space but also time, so that periods of tourist activity may occur at any time of the year without the conventional holiday temporal demarcations of Christmas, Easter, and summer, so that, for instance, short breaks and extended travel to a range of destinations are increasingly common. Nevertheless, the European share of the market remains by far the largest both in terms of flows of tourists to and from the continent. Certain countries predominate on the basis of well established patterns of travel, such as the French and Spanish Mediterranean hot spots, although these beach destinations are being challenged by new attractions in countries such as Turkey and Egypt, and those more 'exotic' and luxurious venues requiring long haul travel such as the Maldives and Mauritius. Nevertheless, global flows of tourists are highly uneven, with certain areas of Africa, Asia, and South America characterized by relatively static visitor figures over recent decades. However, the huge expansion of tourists from non Western countries has also added to the volume of travel between and within a growing range of destination countries.

The globalization of tourism must take account of the expansion of global travel companies and also the advent of the international regulation of tourism. Less developed countries are liable to be exploited by the operations of the transnational tourist corporations (TNCs) who invest and build an infrastructure but subsequently repatriate much of the profit accrued from their operations. On the other hand, tourism requires less investment than other industries and provides an option for development; moreover, state management may obviate some of the worst excesses of the TNCs. Indeed a recurrent problem in the literature has been the prevalence of assumptions that less developed forms of tourism always lead to exploitation and entirely disembed culture from its local context, a set of expectations which minimizes the agency of cultural interpreters and intermediaries, local businesses, and hybridizing cultural entrepreneurs and posit tourist practices and places as static rather than dynamic.

Certain kinds of tourist venture, notably the prevalence of package tours at famous destinations in less developed countries, can act to freeze out smaller tourist operators. For instance, in Agra, the city of the Taj Mahal, local wealthy operators and large transnational travel businesses form alliances which package tourist experience, confining it to specific attractions, forms of accommodation, retail outlets, and dining facilities. The stranglehold on the mobility and expenditure of tourists means that the economic circuits through which tourist money flows are tightly controlled. Smaller traders, ho teliers, craft merchants, guides, and restaurateurs must struggle to attract custom, primarily from the in dependent tourists such as backpackers. With their ex panding income, the large concerns exert a powerful influence on tourist planning, continually jockeying to maximize their profit and isolate weaker competitors.

Tourist sites, however, particularly those that are of considerable architectural, historical, or environmental renown, are liable to be caught up in global conservation process that extends beyond to enfold places into the 'universalization' of heritage. For instance, the World Heritage List drawn up by UNESCO is motivated by a mission to protect cultural and national heritage world wide, and accordingly identifies those sites it considers outstanding, a much desired designation by local and national tourist authorities that confers status on places. However, the monitoring of the appropriate measures of due care for such sites mean that they must also exert responsibility for their upkeep and management.

The Increasing Diversity of Tourism

It is certainly the case that very specific kinds of tourism and particular spaces of tourist destination have been discussed as if they are in some way synonymous with broader social and spatial configurations of tourism and the analytical perspectives that have emerged from such studies have been utilized to generalize about tourism *per se*. However, the global expansion of spaces and practices of tourism renders such approaches and subjects even less widely applicable than they were, for tourist attractions now seem almost innumerable, and there seems to be a venue, outlet, or activity available for every imaginable enthusiasm. For tourism has proved such a lucrative ex ercise, or at least the only economic strategy in areas of deindustrialisation and development, that numerous places have attempted to attract the tourist dollar. This is particularly noticeable in cities where, especially in the increasingly postindustrial West, strategies have been devised to attract the important groups of tourists, shoppers, middle class inhabitants, and investors. Here, the tourist product has become a key element utilized to mark difference in the construction of place images, where a unique characteristic, cultural form of practice, sporting or conference facility, shopping center, historical attraction, museum, landscape, or activity has been identified to create brand distinctiveness within a highly competitive market. For instance, the city I presently live in, Manchester, markets itself using a heady mix of dis tinctive features: an early industrial heritage; a trendy youth culture comprising music, clubbing, and fashion; a prominent gay community; the globally famous sporting institution of Manchester United; and a mix of shopping facilities, among other things. Needless to say, these resources are unequally existent, and it is easier for some cities than others to produce and market their uniqueness.

One key element in these urban plans has been the marketing of all heritage in all manner of forms. A common statistic capturing the rise of British heritage has been the opening of a new museum every two weeks during the 1980s; a similar expansion may be noted in numerous other destination countries. Heritage thus may be offered because of the prominence of some specialized local industry and the remnants of its production sites, a famous event or historical figure ranging from scientist to military commander to entertainer, its architectural or artistic heritage, or simply because there are well pre served vestiges of lower class domestic dwellings. There are now a remarkable range of heritage walks and other tours, booklets, and guides; dramaturgical re enactments; and historical festivals, even where there is little physical evidence of the past. This emphasis on the past has been decried as obsessively nostalgic, and often, unduly se lective and politically conservative; because of the given requirement to produce easily consumable stories and reenactments, it merely recounts the reified histories of the powerful. On the other hand, there is no doubt that hitherto ignored events, lives, and processes have also been represented in recent attractions based upon slavery and racist oppression, and feminist and working class versions of history, and it may now be more appropriate to consider the ambiguities, contestations, and sophisti cation of the total mix of heritage provision in any one city or region.

This intensified variety of attractions is also present in rural areas where there has been a sharp decline in agriculture, glaringly illustrated by the fact that in the United Kingdom, tourism is more or less twice as im portant as farming to the rural economy. In the post productivist countryside, tourist attractions have also proliferated. Country breaks and the expansion of the mix of accommodation, working farms, country trails, and footpaths of varying length, together with the growing popularity of a large variety of action sports at suitable rural locations, have meant that the countryside is increasingly and primarily a realm of leisure and tourism. This has led to growing contestation between diverse activities and particularly the aforementioned notion that the rural is a site for romantic gazing, so that for instance, the quest for tranquillity on a solitary walk might be disrupted by hang gliders or mountain bikers in pursuit of an adrenaline rush, or the calm pursuit of sailing is shattered by the power boat community. These claims on the countryside and the competition they in stigate testify to the social pressure upon rural space, and this is complemented by the deleterious effects of erosion and pollution.

The proliferation of numerous attractions and itiner aries, and practices and forms of pleasure at a global level has entailed the diversification of tourist space wherein any space has the potential to become attractive to tourists. For instance, wine tours in Argentina and Aus tralia are growing in popularity, aficionados of modernist architecture visit East European housing estates from the communist era, The Bureau of Atomic Tourism pro motes sites of nuclear weapon development in the USA, and visitors to New York are drawn to the disaster site of Ground Zero. These multifarious forms of tourist activity and destination reveal that any simple categorization of tourism is defeated by its proliferation and diversity. Yet certain broad tendencies within tourism continue to produce distinctive forms of space – in some cases, forms of serial space. This is most evident in the kind of resort space reserved for generally upmarket package tourists that assumes a somewhat generic appearance and func tion worldwide.

These international standard resorts and hotels, with their familiar, home from home environments and mix of themed spaces, spectacle, and shopping provide a sealed space within which disruptive sounds, tactilities and smells, as well as local people and scenes, are kept away. Local color is usually alluded to with carefully placed local paintings, textiles, and sculptures, and in displays and shows of folk dance and music shows, but the largely purified, rationalized space reduces design and function. Such enclavic spaces have been criticized as 'placeless', 'nonplaces', or 'blandscapes' which erase local cultural specificity and may appear to echo the wider production of functional differentiation and single purpose spaces in which specialized realms are assigned for play, work, and reproduction in accordance with the bureaucratic and commercial imperatives of the tourist industry. Despite these serialities, enclavic resort complexes such as those in Cancun, Mexico, offer an increasingly varied range of activities and excursions which permit tourists to tailor their vacations within more flexible limits. How ever, the delivery of predictable pleasures which induce relaxation and comfort have been overlooked and such vilification seems to reproduce much of the snobbery found among tourists themselves where notions of superficiality, lack of engagement with local people, and lack of authenticity persist. These criticisms of highly productive forms of cultural capital are often mobilized by what are often cited as the package tourists' polar opposite, backpackers, whose touristic pursuits also pro duce distinctive forms of tourist space.

Often outside the West, the destinations and itiner aries of backpackers satisfy a different series of motiv ations and tastes, and in general they are produced through different modes of spatial organization to pro duce more variegated, heterogeneous spaces in which tourist practices and facilities mix with local business, domestic, educational, leisure, and bureaucratic activities. The larger tourist spaces in which backpackers circulate around cheap hotels, restaurants, and bars differ from the purified spaces identified earlier in their more contingent aesthetics, looser regulation, and multiple functions, and they generally ensure that backpacking tourists cannot remain inured to interaction with local people and in stitutions, along with the growing numbers of domestic tourists who pursue desires at variance to those of Western package tourists.

Nevertheless, Western backpackers usually follow prescribed routes, and share 'alternative' values con cerning self development and the 'authentic' engagement with difference. They are drawn to backpacker 'honey pots' in which they may find cheap facilities, other backpackers, as well as identifiable, often conditional, forms of 'otherness'; they are also focused on establishing cultural capital by moving towards more 'exotic' places off the beaten track where they might seek local culture, spiritual awareness, self development, and a fuller en counter with the 'other'. Irrespective of their desires, the growing popularity of talking a gap year among Western students and the expansion of the itineraries and at tractions available to the backpacker are part of the wider commercial institutionalization of such tourism, with the consequence that it is becoming more difficult, but also for some, more important, to diverge from the track beaten by other backpackers.

Early Theoretical Explanations of Tourism

Until the 1990s, tourism research tended to be locked into somewhat of an academic ghetto and seemed to be largely untouched by developments elsewhere in the social sciences and humanities. Early theories about tourism can be further criticized in a number of ways, not least for their tendency towards functionalist explanation, their overgeneralization in their lack of consideration for the varied forms of tourism besides sightseeing and cultural tourism, and above all, for their ethno centric assumptions that tourism is exclusively a Western European, Australian, and North American phenomenon. Accordingly, totalizing conceptions of tourists and tour ism emerged, although by and large, particular kinds of tourists were investigated as they engaged in specific kinds of holidays in particular kinds of places. However, certain theories from this era were highly influential in accounting for the social and cultural origins of modern tourism.

Daniel Boorstin and the Spectacle

In an early and rather dystopian, proto postmodernist account of tourism, Daniel Boorstin maintains that mass tourism is typified by the experience of the 'pseudo event', an occasion which gives the illusion that it is 'real' but is, in fact, a staged production put on for the benefit of tourists. Hence tourism cannot provide any opportunity for experiencing local culture and meeting local folk in unmediated fashion but, rather, isolates tourists from the world within an insulated environment, a tourist bubble within which inauthentic spectacles are consumed. To be sure, there are numerous examples within tourism of simulated events and places and simulacra of existing buildings, environments, and people, most notably in themed spaces such as Disneyland. Yet to conclude that tourists are passively duped into consuming such ersatz spectacles erases their agency and neglects the multiple ways in which they might interpret and experience such attractions. More obviously, such endeavors form only a small percentage of the huge diversity of activities that can be termed touristic.

Victor Turner: Liminality

The anthropological studies of Victor Turner developed an analysis that suggested that tourism was a specific occasion upon which the normal duties and constraints of everyday life were suspended for a period, providing a social safety valve through which people could throw off the ordinary shackles of home and work and engage in playful, transgressive behavior. Like the medieval carnival, that liminal occasion when the ordinary order of things was inverted and sense of *communitas* was temporarily established, tourism similarly involves a movement away from mundane discipline, though in a less threatening, less extreme, and more conventional fashion, which pertains to the carnivalesque but is pursued through rather more organized ways, and can thus be termed 'liminoid' rather than antistructural and liminal. The functional benefits of the holiday were thereby conceived as accommodating people to the stress and repetition of everyday industrial modernity by moving from the familiar context to the unfamiliar realm and then back again. It is possible to identify certain forms of tourism that veer towards the carnivalesque, for example, in Club 18–30 holidays and in certain forms of beach and rave tourism, yet to suggest that tourists are merely carrying out a social function to ensure the health and stability of the social organism again minimizes their agency and their myriad understandings and practices. Furthermore, the explanation draws too powerful a distinction between the everyday and the festive, for as we will see, in contemporary times in particular, the two realms overlap and fuse much of the time.

Dean MacCannell: The Quest for Authenticity

The most influential account of tourism within these earlier studies has been Dean MacCannell's depiction of the tourist as a Western subject in a doomed search for authenticity. Dwelling in an alienated, superficial world of frenetic change and one in which social relations, everyday living, and media and popular culture seem far removed from the past and deeply inauthentic, the tourist is motivated by a quest for authenticity elsewhere. By implication, the tourist, like a pilgrim, is drawn to less developed parts of the world, where the 'real' continues to exist. Sadly however, such authentic experience is rarely satisfied for instead, it is packaged and sold in the mediated form of 'staged authenticity'. Following Erving Goffman, MacCannell alleges that the craved for backstage 'authenticity' of everyday life is elusive, replaced by the contrived productions on the 'frontstage'. The notion that tourism is motivated by a search for authenticity has provoked many accounts that have explored the various ways in which 'authenticity' might be experienced, manufactured, and contested. However, once more the notion that tourists may be defined by their quest for the authentic delimits the exploration of other kinds of tourism and makes assumptions about tourists' motivations that are unsupported by any empirical research. While MacCannell's theory is interesting, the most that can be said is that it applies to some tourists sometimes in certain contexts but as a generalization it proves to be of little use. For instance, it might be relevant for certain backpacking tourists who want to experience the 'real' India by getting off the 'beaten track' or are looking for spiritual enlightenment. Indeed, the notion that some practices and places are more 'authentic' than others is another means through which distinction and status can be, and are frequently, asserted, and provides a form of rhetoric through which backpackers distinguish themselves from package tourists. In critiquing MacCannell's conception, Maxine Feifer argues that tourists – or what she terms 'post tourists' – may actually enjoy consuming inauthenticity, for they are self reflexive individuals increasingly aware of the whole provision of tourist products, aware of the games that tourists must play, and the constructed nature of the itineraries, guide accounts, and attractions they experience.

Later Theoretical Explanations of Tourism

Erik Cohen: Tourist Typologies

Following MacCannell, Cohen insisted that rather than just a singular kind of tourist, there are a variety of different kinds of tourists and practices. Basing his explanation on the degree of immersion in local culture and space, or conversely the inhabiting of the environmental bubble, Cohen distinguishes between 'experiential', 'experimental', and 'existential' tourist types in terms of the degree to which they are searching for authenticity, an idea that inspired other researchers to construct their own tourist typologies and investigate the specificities of distinctive modes of travel, motivation, and cultural practice. Again, however, although this typology provides a step beyond the singular versions of tourist theory, it nevertheless entrenches notions that a limited range of motivations and experiences characterize tourism and, moreover, suggests that distinctive kinds of tourists and tourism can be identified rather than conceiving tourism as constituted by a fluid and diverse series of sensations, practices, and desires.

Valene Smith: Hosts and Guests

Another approach that explored the effects of the global growth of tourism focused upon the social, economic, cultural, and ecological effects that contacts between visitors and residents (or hosts and guests) produce. In Smith's edited collection, and in work that followed this volume's publication, attention was drawn to the frequently exploitative and unstable nature of labor conditions, along with the frequent lack of financial benefit which devolved to the local economy because of the repatriation of most of the profits made by large, nonlocal travel, hotel, and tour companies. Also much discussed was the influence of wealthy Western tourists upon the local cultural context. This, it was argued, may result in 'demonstration effects' wherein locals try to emulate high status fashions and lifestyles of tourists through acculturation, producing conflict where local youth becomes the object of parental concern as 'traditional' values become threatened by their preoccupation with Western mores and tastes. In addition, the conspicuous display of tourists' wealth has been argued to have brought an increase in crime, usually directed at tourists themselves. Alternatively, locals may react angrily towards the presence of relatively wealthy tourists in poor locales or what may be assessed as culturally offensive behavior (for instance, nudity and alcohol or drug use in public spaces). In investigating factors which might minimize the potential for conflict, some have identified a few of the key cultural brokers or intermediaries, typically guides or acculturated locals, who mediate and translate between the two groups. A further strand of research has continued to explore the sexual relationships that emerge between guests and hosts, including sex tourism and sex work, along with the kinds of often ambiguous and complex relationships that emerge outside economic exchanges.

Part of the work on sex tourism has evolved from feminist perspectives and, among other topics, has considered the stereotypical representation of women as objects of the tourist gaze, where they are presented and consumed as exotic and erotic, devoid of agency and sophistication, mirroring many of the sexist assumptions that developed out of colonial representations. More broadly, feminist analyses identify the ways in which the sexual division of labor is reproduced by the restricted range of employment opportunities for women which tend to be of low wage and seasonal, although others have pointed to the potential for female empowerment through the greater economic and cultural autonomy, where women may work outside the realm of conventional employment but market their skills of domesticity and care.

Other avenues of research have contrasted the benefits of economic development with the crass commercialization of valued cultural artefacts, heritage, rituals, and dress. With regard to this latter issue, it was contended that the products of cultural tourism, the 'traditional' music and dance shows delivered for tourists, the ethnic clothing worn by tourism workers, the handicrafts on sale, and visits to their fabrication units were responsible for diluting the meanings of local cultures, providing consumable forms and styles tailored for tourists which emptied out the rich meanings and diversity of local use. However, others have argued that tourism may actually give a boost to local culture and reinvigorate dying traditions, bringing back forgotten modes of skillful manufacture and performance. On the whole, despite their usefulness in identifying important social and cultural processes, these analyses suggest that both tourists and their hosts are somewhat homogeneous, lacking agency, and tend to conceive destinations, communities, and tour groups as culturally static entities rather than dynamic.

Concerns about the unequal and potentially exploitative relationships between hosts and guests have led to the rise of various forms of alternative tourism, most evidently, 'ecotourism' or sustainable tourism, where greater emphases are placed upon local economic control, small scale development, environmental conservation and damage limitation, poverty alleviation, and a more equitable distribution of the economic rewards of tourism, female empowerment, and the fostering of more meaningful relationships between tourists and locals. If successful, however, such schemes may become liable to fall victim to the phenomenon of the 'resort cycle'.

The Resort Cycle

The extraordinary global expansion of tourism, where whole nations, cities, and regions compete with each other on a global scale, with a corresponding volatility in the popularity of destinations has prompted scholars to point to the phenomenon of the 'resort cycle'. The resort cycle is produced by the desire for many tourists to escape the large scale tourist resorts, with their crowded beaches, diluted culture, and mass entertainment facilities and head to somewhere 'off the beaten track'. There are therefore, opportunities for unfamiliar venues lacking modern tourist infrastructure to become fashionable by virtue of their attributed 'untouched' nature, for they are sought out by more adventurous travelers on a quest for the 'unspoilt' and 'untouristy'. As word spreads that there is an untarnished beach, trail, or city to visit, local entrepreneurs then identify the potential for tourist development, there is investment in tourist infrastructure, more tourists arrive, and finally, the formerly quiet locale becomes a large tourist resort. However, the new scale of popularity subsequently threatens the resort's reputation for it now lacks the very qualities that led to its emergent attractiveness in the first place, and it may be variously considered to be overcrowded, too expensive, inauthentic, too modern, and too homogeneous. The resort cycle highlights the tenuous position of those currently fashionable tourist destinations, where development can herald their demise, not to mention any sign of instability or political violence. Nevertheless, venues can also regain a measure of former popularity by promoting festivals, developing 'unique' attractions, and simply because they may regain a vanished trendiness once more after time passes.

John Urry: The Tourist Gaze

A great breakthrough in sociocultural understandings about contemporary tourism arose with the publication of John Urry's *The Tourist Gaze* in 1991, which developed the perspective that tourists are characterized above all, by their desire to gaze upon that which is different and unusual, and can be distinguished from the everyday. Moreover, they are collectors of sights, literally sightseers, as they travel around the realm of the 'other', gazing upon people, cultural shows, landscapes, and historic locations, as part of the broader sociocultural development of an intensified visual culture generated by the dense and proliferating flows of global signs which has apparently reinforced the dominance of the visual according to assessments of the importance of the senses in Western culture. In addition, Urry maintains that tourists are semioticians; that is, they interpret the sights they consume according to symbolic frameworks whereby the objects of their gaze come to stand for some other general quality. For instance, renowned spectacles such as the Eiffel Tower connote Paris, and smaller 'typical' sights, such as the city cafe or breadsticks, symbolize Frenchness. Typically, tourists are also concerned with 'capturing' the scenes upon which they gaze through photography, and latterly, the video camera. This touristic disposition to gaze upon the unfamiliar largely derives from the aforementioned romantic desire to experience the visual pleasure associated with the sublime and is associated with predispositions to gaze in particular ways and via specific technologies. The more protean, expanded lust for gazing upon difference in contemporary tourism is similarly identified as the 'romantic' gaze, although Urry allows for other modes of gazing, including the 'collective' gaze where large crowds gaze upon popular spaces and each other. Whatever the form of gaze, however, the importance of considering the visual in contemporary tourism is reinforced by the massive increase in the number of places, people, attractions, and landscapes upon which to gaze, from the mundane to the spectacular. This proliferation of objects available for the gaze of tourists is compounded by the enormous global expansion of flows of mediatized images of other sites that would previously have been considered unsuitable tourist venues. Despite the widespread influence of his contentions, Urry's claim that tourism is preeminently concerned with visual consumption has been strongly challenged by writers who assert that such a focus neglects the numerous other ways in which other senses are foregrounded and mobilized in tourist experience. Other embodied sensations are most particularly evident in the tactile, auditory, and gustatory pleasures implicated in other forms of tourism, as we will see. Ultimately, it must be questioned whether an explanation of tourism which privileges visual experience as central is not also guilty of over generalizing about the motivations, experiences, and underlying generation of tourist experience, consumption, and production, especially in the absence of any empirical evidence.

The Recent Broadening of Theories about Tourism

While much scholarly work on tourism in the major academic journals concerned with the subject remains fixated on notions of tourist motivation, relationships between hosts and guests, quantitative depictions of tourist flows, models for resort development, and marketing and management strategies, many researchers have moved away from old conceptual models and theoretical preoccupations and developed approaches that are informed by some of the most influential new ideas emerging out of the social sciences. This article briefly summarizes these more recent approaches, focusing upon understanding tourism as constituted by accelerating and proliferating global mobility, as embodied, a form of performance, increasingly mediated, and part of everyday life.

Tourism as an Effect of Enhanced Global Mobility

The reconceptualization of places as continually reconstituted by the processes which flow through and center upon them, especially in an increasingly mobile world through which places become intertwined with people, information, companies, technologies, and other places at a rate never before experienced implicates understandings of tourism. Tourism brings together building materials and architectural planes, company expansions, transport networks, images and brands, temporary and permanent workers, and multiple other elements in complex entanglements which contribute to the ongoing production of place. In such a conception, there is no essential characteristic of place, no authentic spirit of place, or heritage, despite the tendency of tourist marketing to suggest such fixity in the quest for identifiable, distinct place images. Rather, places are characterized by their contingency so that stability resides in the propensity for tourist workers to reproduce regular narrative versions of place and expected standards of service for the tourist infrastructure to remain reliable, for tourists to continue to visit, and upon the repetitive interpretations and practices of tourists. Yet the fluid 'ethnoscapes', the global flows of people of which tourists form a large proportion, are part of a mobile world which brings together different kinds of people and spaces at accelerating rate, incorporating ever more places and landscapes into potentially volatile networks of movement through which a place might become temporarily stabilized but also threatened by chaos. Under conditions of volatile flow, elements of tourism are apt to escape from the networks within which they serve particular functions, and enter into other circulating flows to appear in other geographical contexts. For instance, the woollen hats and sweaters popular among travelers to the Andes are now widely accessible in the retail outlets of Western cities.

Embodied, Phenomenological, and More-Than-Representational Accounts of Tourism

Growing social science research in phenomenology, the body and embodiment, and the senses have recently exerted a strong influence on theories of tourism. While early academic ideas suggested that tourists were somewhat disembodied subjects, Urry's notion of the 'tourist gaze' discussed earlier developed the notion that tourists were embodied and that the senses were important in the practices of tourism. In recent times, however, scholars have pointed out that this is an excessively ocularcentric conception of tourism which privileges vision to the exclusion of the other senses. This is to be expected in Western enlightenment culture where vision has long been held to be the most important in hierarchical understandings of the senses. It is therefore important to identify the cultural specificity of such claims about senses and further note that the senses are culturally shaped and do not provide an unmediated access to the world as purely 'natural' tools. We perceive according to cultural values which prescribe what is sensually desirable and acceptable and we are trained to sense the world according to these values. Accordingly, new ways of theorizing the relationship between the body and tourism have evolved which focus attention upon ideas about physical fitness and health, sensual connections and immersion, sexuality, and visceral sporting experience.

Most obviously, the rise of muscular tourism, the proliferating collection of adventurous pursuits based upon white water rafting, canoeing, mountain biking, surfing and skateboarding, snowboarding, bungee jumping, hang gliding, and orienteering have supplemented the older activities of walking, skiing, swimming, and climbing, and have ramped up the values of risk, thrill, and competition. It is arguable whether any of these pursuits foreground the visual as the dominant sense.

Conceptions of embodiment and tourism have also emerged because of the critique of binary constructions of nature and culture, where the body is conceived as separate from nature and that which is gazed upon. More than representational theory has instead asserted the idea that bodies, nature, and all other elements (for instance, within the systems of tourism) are inextricably connected, part of assemblages which draw together humans, technologies, information, souvenirs, minerals, vegetables, and animals to constitute (temporary) networks through which the elements are endowed with particular attributes that contribute to the system as a whole and connect specifically with other parts of the system. Tourism, bluntly, is not solely concerned with the cognitive understanding of places and with the consumption of representations. Within this scheme of thought, bodies are always already involved with the world in precognitive fashion, constantly attuning to new environments, sensing stimuli, and acting in unreflexive ways towards the world. We apprehend the world in ways that accord with the qualities of place and the capacities of our bodies which go way beyond the semiotic decoding enacted by a solitary, gazing tourist. Moreover, the human body is not primarily a text itself but is immersed in changing and diverse sensations.

Yet we should be wary of conceiving of tourism as solely a movement through unfamiliar realms for the embodied practices of tourists are more frequently involved with habitual immersion in familiar spatial and social circumstances, where they experience regular, homely tactilities, smells, sounds, tastes, and textures. As such, tourist bodies develop practical dispositions through sensory experience, except where they are overwhelmed by social and sensory unfamiliarity. What becomes important to explore then, is the relationship between bodies and tourist space. The experience of an amusement park might foreground the sensation of rapid motion, giddiness, and disorientation as the body hurtles around a white knuckle ride. A stroll down an Indian bazaar in heterogeneous tourist space is likely to provoke strong resistance and attraction to a welter of sensations, varying in intensity and combination, as movement is impeded and obstacles avoided, delicious and atrocious aromas inhaled, sweet and jarring noises received, and curious and familiar sights beheld, composing a sensual experience that includes the auditory, tactile, visual, aromatic, and gustatory at once, and in succession. And while a coach tour may well prioritize the consumption of visual spectacle, there is always the prospect that other sensations will intrude upon the tourist gaze: the bus is too hot, travel sickness arises, hunger gnaws, and the seats are too cramped and uncomfortable.

In addition to bodies, and informed by more than representational theories, tourism has also recently dir ected attention towards materiality, the objects of tour ism, which possesses an integral role in the ongoing constitution of tourism. Such nonhumans are part of the relationalities which produce the ordering of tourism in different ways and are crucially not separate from humans but inextricably bound up with them so as to produce human–object hybrids that foster the distinct qualities of tourist experience and practice. There are those objects which travel from place to place, the pass ports, suitcases, cameras, and items of clothing which tourists carry with them, the souvenirs which tourists bring back from their holiday destinations, and the technological objects of tourism itself, the vehicles, loudspeakers, and recording equipment that guide tour ists around heritage sites. Such objects possess particular affordances, qualities which enable and prevent action, provide pleasure, and facilitate interaction and move ment. Some functional traveling objects move between places without changing their meaning, but others, such as souvenirs, are removed from their place of origin and recontextualised in tourists' home settings, often pro viding signs of status and identity, or prompts for nar ratives about holiday experiences.

Tourism as Staged and Performed

A particularly productive sociological approach has considered how tourism might be conceived as staged and a form of performance. Tourist stages are many and varied as are the kinds of performances enacted by tourist workers and participants in spectacles. Highly organized ceremonies such as royal and state pageants tend to be rigid and reproduce the distinction between actor and onlooker, because they are required to convey particular ideological messages and thus minimize ambiguity, for such contestations and failings must be continually held at bay. Alternatively, more carnivalesque festivals allow for more convivial, sensual, improvisational, and playful performances.

The production of heritage dramas and reenactments tends to decenter the authority of historical experts and provide a more sensual, dramatic version of history than the dry accounts of the authoritative, an affective, ex periential experience that often require tourists to par ticipate in role playing as opposed to the improving acquisition of useful knowledge. There is also the on going proliferation of themed spaces in which stage managers attempt to produce a cultural and physical environment encoded with clear visual cues and codes through the use of scenography. Another line of dis cussion has investigated how tourist productions stage nature through safari parks, wildlife safaris, and sea life shows, foregrounding the spectacular, the anthropo morphic, and the attractive and eliding the uncomfort able, noncuddly denizens of the animal world. On the whole, we may discern a situation whereby intensified attempts to design and theme space coexist with the promiscuous proliferation of forms of tourist space.

Most tourist stages depend for their stability and ef ficacy on the ability and disposition of workers to perform in ways that induce comfort, service, and predictability. Accordingly, such workers are trained to express defer ence, eagerness to please, and friendliness. In the case of Disneyland, such personnel are conceived as 'cast members', and as in other settings, required to wear outfits and expressions that harmonize with the theme environment. In addition, tour companies and hotels organize cultural displays of local dancing and music, and here, performers enact selective aspects which typically titillate tourists without alienating them by confronting them with complex cultural meanings. Not only do tourist workers directly serve tourists but also manage them, in a role comparable to choreography and stage management. For instance, tour guides provide scripted commentary and choreograph tourists' movement along prescribed routes, reinforcing notions of what should be looked at and how tourists should behave.

I have already inferred that patterns of tourist be havior partly emerge out of dispositions that evolve around class, gender, ethnicity, and sexuality, practices which are grounded in the specific habitus of tourists, their unreflexive ways of being. Moreover, tourists re produce shared conventions about what should be seen, done, and communicated. These understandings about what to do and how to be a tourist move between the self aware performance in front of other tourists and habitual dispositions which are rarely questioned. The apprentice tourist may well become inculcated in the procedures whereby performative competence in fol lowing recognizable conventions is attained in knowing what to photograph, how to exclaim upon witnessing the 'exotic', and how to record experience.

Such conventions need to be continually enacted to retain their power, and the prescriptive conventions and values that are inherent in them are rarely disrupted if they are performed unreflexively. Yet where reflexive improvization and a critical disposition are mobilized, the resultant ambiguity can threaten a sense of well being, as can sudden immersion in stages that contain unfamiliar fixtures and people acting in ways that confound expectations and cultural values. Although the consent of tourists is required for the successful staging of events, guided tours, sightseeing, and other forms of consumption, there are also those participants identified earlier as 'post tourists', who revel in the artificiality of the tourist product. Enacting cynical performances, they undermine the staging of 'authenticity', yet though they display their distance from the normative roles expected of them, they are ultimately unwilling to challenge the drama in which they are involved. There are also those who refuse to participate at all in the tourist product and strike off on their own into unfamiliar settings, where they might hope to self reflexively shed the normative roles they ordinarily inhabit, in search of spiritual enlightenment, drug induced experiences, extreme sports, and other forms of 'limit experience'.

Different tourist ventures are carried out upon particular stages – on beaches and mountains, in cities, heritage sites, museums, and theme parks. These settings are distinguished by their boundedness, whether physical or symbolic, and are often organized or stage managed to provide and sustain common sense understandings about what activities should take place. Indeed, the coherence of most tourist performances depends on their being performed in specific 'theaters'. In such realms, tourist performance is socially and spatially regulated to varying extents. Indeed, the organization, materiality, and aesthetic and sensual qualities of tourist space influences the kinds of performances that tourists undertake, varying from the predictable collective enactments of package groups under supervision and surveillance to the improvisational performances of individuals moving through differently regulated, unfamiliar environments, from the predictable routines of those in serial, 'enclavic' space to those amidst heterogeneous tourist space. Yet the stability of any tourist stage depends upon the repetitive performance of familiar routines and procedures and where this is challenged or incompetent, the normative purpose and meaning of the space itself come under threat.

However, rather than being fixed, performance is an interactive and contingent process which succeeds according to the skill of the actors, the context within which it is performed, and the way in which it is interpreted by an audience. Even the most delineated social performance must be reenacted in different conditions and its reception may be unpredictable. Each performance can never be exactly reproduced and fixity of meaning must be continually achieved. Yet because tourism involves the process of accelerating mobilities through which people continually bump into forms of difference, the normative, habitual social performances which they ordinarily enact with groups of peers are made manifest, brought to the surface, and can therefore be challenged, inducing a self critical reflexivity about the assumptions which underlie their behavior or a defensive retreat into certainty. It is evident that at those venues where flows of people from different geographical and cultural locations meet and mingle, there are frequent contestations about what constitutes proper touristic dispositions, habits, and practices. At the Taj Mahal, for instance, Western package tourists often become irritated at the tourist performances of large parties of domestic tourists who quickly move around the sight, chatting noisily, failing to take photographs, and enact a romantic gaze towards the mausoleum, and otherwise displaying a lack of reverence.

The Tourism and Media Nexus

The expansion and intensification of global flows of images (via 'mediascapes') has encouraged some scholars to investigate the growing nexus between tourism and the media. Their arguments focus upon a number of issues. First, they highlight the ways in which tourist destinations, and the people, sights, and cultures foregrounded in tourist marketing campaigns circulate through a variety of global and national media forms, contributing to growing mobility between places, not through the actual movement of tourists but through virtual representations. Images of tourist places and landscapes feature in television travel programs, advertisements in print and screen, on internet sites, and in books and newspapers, informing the tourist imagination and fuelling the anticipation, expectations, and desires of those who will subsequently visit them. Crucially, these objects of the virtual tourist gaze tend to rely on familiar, repetitive images in much the same ways as colonial guidebooks and travel accounts have tended to reinforce the significance of particular sites and landscapes, cultural habits, and exotic forms of 'difference', and tourist encounters with locals. For instance, travel programs have a tendency to rehash the notion that places outside the West are 'timeless', a denial of their modernity which testifies to the desire to escape to another place and time. The catalog of dancing girls, palm lined shores, and lined faces full of character is only new in its expansion. These television programs, combining advice, information, and imagery, also frequently rely on the individual travel accounts of celebrities.

It has been a contention that the recycling of familiar tourist imagery is also generated by tourists' own photographic reproduction of familiar images, as they position themselves within familiar scenes of the Statue of Liberty or Petra. These images may well be utilized in post holiday recollection but they are also increasingly apt to appear, together with tourists' individual accounts, on internet websites and travel blogs. This addition to the flow of global tourist images testifies to a mediation which is democratizing the flow of information and opinion about tourist destinations, and in the case of backpacker travel, for instance, is part of the reflexive engagement which their journeys constitute and that forms a basis forum for intersubjective communication.

In addition to the systematic siting of tourist places within global mediascapes, there are now a plethora of destinations which have become associated with mediated imagery in film and television, with celebrities, particular cinematic scenes, and landscapes. Recently, fans of the *Lord of the Rings* film series have flocked to New Zealand to witness the scenery featured in the movies, and others visit Tunisia to behold the place where sections of the *Star Wars* movies were filmed or go in search of the seaside setting of the film, *The Beach*, filmed in Thailand. In some cases a strong association with film and television assists in the tourist promotion of an area, as in the case of Stirling, which has become labeled as *Braveheart*. Yet the connection between the virtual world of television and the phenomenological world encountered *in situ* is not a matter of simply witnessing familiar scenes. For example, the welter of cinematic representations of Los Angeles may serve to assist recognition of that city's space for a visiting tourist but might also clash with the unexpected sights of everyday life and squalor that are confronted, the unfilmed places of the city.

The link with media is especially close in the case of the numerous themed spaces which actively seek to combine space with image and celebrity. The most famous of these sites is, of course, Disneyland, which continually updates its attractions to keep pace with modish and popular films and celebrities. Familiar cartoon, film, and television programs have similarly attracted large numbers of tourists to Universal Studios in Los Angeles, the Asterix theme park in France, and the now defunct Granada Studios tour in Manchester, as well as a host of attractions in theme parks, shopping centers, and festival marketplaces.

Moreover, the tourist experience and the presentation of place is itself increasingly mediated through the application of audiovisual technology. This tendency seems particularly prevalent at heritage attractions where slide shows, short videos, *son et lumiere* displays, and animatronic figures engage the tourist audience. This provides a usually melodramatic narrative where little actual vestige of historical import exists such as at Bannockburn Heritage Site in Scotland, where disappointment at the dearth of any trace of the medieval battle between Scottish and English armies is assuaged by the mediated spectacle of an animated video of the event.

Tourism as Everyday and Mundane

Part of the discomfort with the conception of tourism as an extraordinary period has recently been addressed by accounts which consider that the division between the everyday and the festive has become blurred. First, as already mentioned, tourism is replete with unreflexive, mundane, and rigid conventions, with culturally coded habits and routines which shape the particular practices and experiences of tourists and embody common sense understandings of how to be a tourist. And though suffused with notions of escape from normativity, tourists carry quotidian habits and responses with them along with their luggage and such habits are accommodated and understood in the tourist spaces they come to occupy. Part of the reason for notions that tourism is extraordinary has persisted is because tourists have been so frequently misconceived in general terms as Western solitary travelers amidst realms of otherness. This neglects the social nature of tourism, whether this includes the duties that are carried along when families travel together or the obligations that are performed when visiting old friends or hosts in familiar places.

Second, because tourism now extends across an increasing vast range of times and places, it has become part of the everyday apprehension of the world, with normative technologies and aesthetic dispositions. In a spectacular society, bombarded by signs that saturate everyday worlds, we can be tourists in our everyday travels, whether actual or virtual. For instance, the regeneration of Western urban economies involves the production of proliferating shopping 'experiences' with themed pubs and restaurants, festival marketplaces, amusement parks, museums, heritage attractions, art spaces, cultural quarters, spectacles, and signature buildings – all spaces in which the highlighting of selective cultural and historical attributes has become routine. While primarily intended to attract outsiders, for the workers, inhabitants, and shoppers of cities, such sights become mundane.

Moreover, because of the pace through which global cultural flows promiscuously expand, a previously obscure otherness is ubiquitous in commodities ranging from music, food, and fashion from around the world, and those people and cultures formerly considered 'exotic' and worthy of a visit are commonly confronted as part of daily life in multicultural, multiethnic cities as part of a sort of mundane cosmopolitanism, not to mention the numerous virtual travel we engage in on the internet, television, and cinema. As cities utilize festivals, cultural events, demonstrations, and exhibitions to prove their cosmopolitanism, the penetration of the exotic into everyday lives and banal urban spaces renders it mundane, diluting its power to confound normativity. Accordingly, tourism has lost much of its power as a practice through which the everyday might be transcended via a confrontation with otherness.

Third, the capacity of the conventional scheduling of tourism to provide occasions for confronting and consuming otherness is further diluted by the sheer proliferation of occasions for leisure that are now available to increasing numbers of people. As work routines for many become more variegated and flexible, there are growing occasions when we may take afternoons off work or 'long weekend' breaks, and numerous leisure opportunities available at night and the weekend. The diverse availability of holidays offered throughout the year similarly blurs notions that certain times are appropriate for a holiday and other periods should be spent working. The specialness of the holiday in contradistinction to 'ordinary' life, and the swelling sense of release following a sustained period of consistent work without a significant leisure break, is becoming less common, although the distribution of leisure opportunities remains highly unequal.

These notions that tourism is the opposite of the quotidian say much about how the everyday is conceived, as a realm of repetition, duty, boredom, and habit. In fact, while the bureaucratization of the life world and industrial rhythms may have produced such mundane experiences, such understandings neglect the potential of the everyday in its temporal fluidity to also provide opportunities for dreams and fantasies, moments of creativity, and unexpected happenings.

See also: Ecotourism; Tourism, Rural; Tourism, Urban; Travel and Travel-Writing.

Further Reading

Baerenholdt, J., Haldrup, M., Larsen, J. and Urry, J. (2004). *Performing Tourist Places*. Aldershot: Ashgate.

Boorstin, D. (1964). *The Image: A Guide to Pseudo Events in America*. New York: Harper.

Britton, S. (1991). Tourism, capital and place: Towards a critical geography of tourism. *Environment and Planning D: Society and Space* 9, 452 478.

Cohen, E. (1988). Traditions in the qualitative sociology of tourism. *Annals of Tourism Research* 15, 29 46.

Crick, M. (1989). Representations of international tourism in the social sciences: Sun, sea, sights, sex and servility. *Annual Review of Anthropology* 18, 393 444.

Crouch, D., Jackson, R. and Thompson, F. (2005). *Tourism and the Media Imagination: Converging Cultures*. London: Routledge.

Desmond, J. (1999). *Staging Tourism: Bodies on Display from Waikiki to Seaworld*. Chicago, IL: University of Chicago Press.

Edensor, T. (1998). *Tourists at the Taj*. London: Routledge.

Enloe, C. (1990). *Bananas, Beaches and Bases: Making Feminist Sense of International Politics*. Los Angeles, CA: University of California Press.

Feifer, W. (1985). *Going Places*. London: MacMillan.

Franklin, A. (2003). *Tourism: An Introduction*. London: Sage.

Hall, C. and Tucker, H. (eds.) (2004). *Tourism and Postcolonialism: Contested Discourses, Identities and Representations*. London: Routledge.

MacCannell, D. (1976). *The Tourist: A New Theory of the Leisure Class*. New York: Schocken.

Meetham, K. (2001). *Tourism in Global Society*. Basingstoke: Palgrave.

Minca, C. and Oakes, T. (eds.) (2006). *Travels in Paradox: Remapping Tourism*. London: Rowman and Littlefield.

Sheller, M. and Urry, J. (eds.) (2004). *Tourism Mobilities: Places to Play, Places in Play*. London: Routledge.

Smith, V. (ed.) (1989). *Hosts and Guests*. Philadelphia, PA: University of Pennsylvania Press.

Smith, M. (2003). *Issues in Cultural Tourism Studies*. London: Routledge.

Turner, V. and Turner, E. (1978). *Image and Pilgrimage in Christian Culture*. New York: Columbia University Press.

Urry, J. (2002). *The Tourist Gaze* (2nd edn.). London: Sage.

Wood, R. (1998). Tourist ethnicity: A brief itinerary. *Ethnic and Racial Studies* 21, 218 241.

Tourism, Rural

M. Cawley, National University of Ireland, Galway, Republic of Ireland

Glossary

Commodification A process whereby objects take on an exchange value over and above their use value and may be traded.

Gaeltacht Legally defined areas in the Republic of Ireland, located for the most part in the west, northwest, and southwest, where the Irish language is the first language of the majority of the bilingual population.

Partnership A collaborative agreement between two or more interest groups from the state, private, and voluntary sectors for the purpose of achieving a defined aim.

Tourist Gaze A term coined by sociologist John Urry, to describe the process of consuming tourism experiences which differ from the everyday.

Defining Rural Tourism

Reference to rural tourism inevitably raises issues relating to the rationale, if any, for distinguishing between tourism and recreation and between urban and rural destinations. Any definitions that are used are subject to discussion. Furthermore, although the countryside is a major resource for recreation internationally and tourism is advocated as a method of supplementing declining rural incomes, many countries do not provide statistics for rural tourism as distinct from rural recreation. A number of criteria are, however, widely recognized that may be used to frame a discussion of rural tourism according to associated features, theory, impacts, and policy.

Tourism and the Rural

For purposes of international statistical comparison, tourism is usually defined according to the World Tourism Organization criterion as involving a stay of at least 24 h away from home, with recreation involving more regular shorter sojourns. Tourism and recreation share many similarities, however, in terms of the activities undertaken and their use of the resource base. Thus boating, golfing, horse riding, and walking may be practiced as regular recreational leisure activities or during a longer holiday and may not differ appreciably in their resource demands and environmental impacts. It may be difficult to distinguish between a day visit to a heritage site by a local family and that of a tourist family on holiday. The similarities in the use of the resource base by recreationists and tourists are recognized in this discussion and features of tourism more specifically such as accommodation and development policies are highlighted as pertinent. The definition of the rural *per se* is also contested, as reflected in an extensive literature, and many tourists divide their holiday between urban and rural locations. However, rural tourism is usually distinguished from urban tourism by a number of features including the type of experience anticipated by the visitor, the scale of infrastructural development, the use of local natural resources, and the involvement of local people. To some extent it may be more appropriate to refer to tourism and recreation that take place in the countryside while recognizing that the countryside may include small towns and villages, as well as cultivated, uncultivated, and forested landscapes and areas of water. A distinction that is long established within rural tourism is that between truly remote wilderness landscapes, as in northern Scandinavia and Canada, and locales that are less isolated. The United States of America Wilderness Act 1964 defines wilderness in terms of an absence of other than a transitory human presence and retention of a primeval character and influence. In more accessible locations rural tourism experiences often include contact with both nature and culture through passive and active recreation, cultural presentations, and artifacts relating to country lifestyles and heritage. Rural tourism is recognized by the European Union (EU) as providing development opportunities for declining economies.

Tourists and Recreational Activities

Surveys in many countries reveal that demand for rural recreational opportunities has grown during recent decades. This growth has been associated with increased leisure time in industrialized states; concerns relating to health, physical, and mental well being; antipathy to the standardization characteristic of mass tourism destinations and awareness of the pressures imposed on their environmental resources; and a postmodernist quest for contact with older cultures and ways of life as an antidote to the anomie of contemporary urban society. Changing value systems and associated change in the experiences or gaze sought by tourists, as described by John Urry, involve a transition to more environmentally sympathetic forms of soft, green and ecotourism but can also include physically demanding adventure sports such as hang gliding, surfing, and white water rafting. Traditional foods, drinks, craft products, and cultural practices form part of the experience sought and provide commercial opportunities for local producers. The search for the (increasingly elusive) non Westernized and exotic is re flected in the growth of ecotourism among backpackers and organized groups to rural locales in developing areas of Africa, Asia and South America, the Arctic, and Antarctica.

In addition to domestic and international visitors, returning emigrants or their descendants form an im portant stream of, sometimes repeat, visitors who have affinities with reception areas in parts of France, Ger many, Greece, Ireland, Japan, Norway, Sweden, and Spain, for example. Some of these visitors may be second home owners. Second home ownership may be a con tentious form of tourism, however, if it contributes to inflating house prices for locals. When ownership is associated with weak local ties among national or inter national immigrant owners it may in extreme circum stances lead to open antipathy and the destruction of property, as occurred in Wales during the 1970s and the 1980s.

Rural Accommodation and Farm Tourism

Many types of accommodation formerly associated with the countryside (even camping) are now found as niche products in towns and cities. Nevertheless, rural ac commodation continues to be characterized as being small in scale and dominated by provision in private homes, small hotels, guesthouses, camping and caravan sites, and self catering in vernacular buildings (best represented, perhaps, by the French gîte which began to be marketed as a distinctive form of rural accommodation in the early 1950s). In some countries, farm accom modation in isolation or in combination with recreational activities is marketed under a specific label; for example, ‘Farm Holidays’ in Austria (Urlaub am Bauernhof) and ‘Staying on a Farm’ (Bo på Lantgård) in Sweden. Even when not marketed jointly farm holidays often form a clearly defined product. Opportunities for involvement in agricultural tasks may be provided, as well as an in creasingly diverse range of outdoor activities and on farm shopping for local foods and crafts. Rural accom modation has expanded also to include spas, religious and spiritual retreats to appeal to emerging tourist tastes. The sauna is an integral part of a rural stay in Scandinavia. The state, at levels from the international to the local, has promoted the introduction of on farm tourism as a form of income supplement, since the late 1980s, following agricultural restructuring in Australia, Canada, the EU, New Zealand, and the USA.

Theoretical Considerations

Rural tourism and recreation research have been criti cized in the past as having a weak theoretical base which was attributed to a concern with practice arising, in part, from a reliance on funding by government and other agencies and, in part, from an excessive concern with case studies. Harvey Perkins’s use of structuration theory to conceptualize the rural leisure locale as a human inter action space, which derives meaning from the amounts of time devoted to it, is sometimes cited as an exception from the 1990s. A number of key theoretical concepts are increasingly being invoked by geographers to analyze the construction of rural tourism and its experience by tourists and recreationists, some of which have been applied more extensively than others. These relate to: (1) the commodification of the countryside for con sumption; (2) tourism as performance by and for tourists and embodiment as a method of analyzing active en gagement with nature in place of passively gazing upon nature; and (3) issues of power associated with contacts between tourism, tourists, local residents, and the re sources on which tourism and recreation depend.

Commodification

The commodification of the countryside forms part of the transition from high input high output productivist agriculture to its postproductivist opposite that has taken place to varying degrees in different locations since the early 1990s. Farm labor, investment, buildings, and land have attained a new exchange value in being allocated to tourism in response to growing consumer demand. Rural landscapes and particular features have been commodi fied through their incorporation into promotional media and products sold as souvenirs to visitors. Thus, Uluru (Ayers Rock) has become an iconic symbol that attracts large numbers of visitors annually to the desert interior of Australia. Regionally labeled foods and drinks are the basis for specialist tours and trails. Associations with the constituents, conditions, and locale of production which form the basis for Appellation d’Origine Contrôlée (AOC) labeling of wines and cheeses in France, and its equivalent in several other European countries, have been adapted to a wider range of food and drink products in EU Protected Designation of Origin (PDO) and Pro tected Geographical Designation (PGD) labels. The re gional imagery used to promote places and as a basis for festivals and workshops may include associations with writers or fictional characters (Hardy Country in Eng land, Anne of Green Gables on Prince Edward Island), film and soap operas (the Wizard of Oz which the small town of Sedan, Kansas, evokes in its yellow brick road, Heartbeat which attracts tourists to rural north York shire), artists (Pont Aven in Lower Brittany), and even outlaws (Glenrowan, New South Wales where the no torious Ned Kelly was captured is the center of Kelly Country). Commodification and regional labeling are being funded through the EU Liaison entre Actions de

Développement de l'Economie Rurale (Leader) rural development program as a form of cultural economy whereby economic value is allocated to cultural resources. Labels may be regulated to ensure quality.

Performance and Performativity

Another area of theory that is increasingly invoked in studies of rural tourism experiences relates to performance and performativity for and by visitors. Tourist locales may become stages where actors in traditional dress act out domestic chores and conduct agricultural or industrial tasks in open air or indoor museum settings. Many heritage and historical sites which are recreated and performed for visitors are in rural settings, for example, Valley Forge in Pennsylvania where the Continental Army's winter stay of 1777–78 is commemorated. Analysis of performance for and by tourists also involves issues relating to the authenticity associated with the act of seeing. Studies among native peoples in Australia, Hawaii, and New Zealand reveal back spaces which local residents reserve from the tourist in an attempt at self protection and front spaces where encounters take place and which may be staged to meet tourist expectations; hence giving rise to the concept of staged authenticity. More generally the tourist becomes a performer when on holiday in adopting out of the ordinary behavior, dress, and methods of observing. Performativity represents a move beyond John Urry's tourist gaze, focusing instead on embodied (and often nonrepresentational) aspects of tourism, for example, in adventure tourism. Thus, recreationists may co construct places with animals, as illustrated by Paul Cloke and Harvey Perkins's study of cetacean performance and tourism in Kaikoura off the South Island of New Zealand, which uses concepts relating to actor networks and relational agency to analyze embodied performances associated with whale watching and swimming with dolphins.

Power Relationships

Rural tourism incorporates a wide range of stakeholders including tourists and recreationists, tour operators, businesses, policy institutions, resource providers, and local residents among whom power relationships exist to varying degrees relating to both public and private resources. Access to land for hiking can be contentious, as illustrated in the United Kingdom during the 1990s prior to the introduction of new legislation relating to the right to roam in the countryside. There is a tradition of access to the uncultivated countryside in Scandinavia and Germany, for example, whereas in the Republic of Ireland few rights of way exist and private property rights are defended actively. In 2007, a recommendation by a group of legal experts that access might be provided to the countryside for recreation, under stipulated provisions, was met by vigorous opposition by farm representative groups. Underexplored scope exists to invoke theories of power, following Michel Foucault and others, to conceptualize the relationships present between different rural tourism actors. In developing countries, postcolonial and feminist critiques are sometimes applied to analyze unequal relationships between international capital, tourists, and local providers.

Tourism Impacts and Implications for Sustainable Rural Systems

The impacts of tourism and recreation in the countryside are generally conceived as being interrelated and threefold: economic, sociocultural, and environmental.

Economic Impacts

Economic benefits arise from tourism but also disbenefits because at least some of the economic returns may accrue outside the area which the tourists and recreationists visit (and many of the sociocultural and environmental disbenefits may occur within the area). Visitors who stay in urban centers and tour by car or coach may contribute to congestion on roads and crowding at honey pot sites but expend little locally. If rural accommodation is externally owned, leakage of expenditure takes place from the local area. Coach tour companies are often based in major cities and, increasingly, large European tour operators use their own coaches for international touring in Europe thereby retaining greater shares of profits. Tourists who travel in camper vans and caravans often bring supplies of food, drink, and alcohol and make limited purchases in the locales visited. The returns to the rural economy are likely to be greatest when the accommodation, food, and other products used, and the recreational activities engaged in, are owned and produced locally. Hence, the increased attention being given to the refurbishment of underutilized rural buildings for accommodation, the promotion of speciality foods, and the development of food, drink, and historical trails. The Camino de Santiago which follows an historical pilgrim route from France across the Pyrenees through northern Spain, to the supposed tomb of St. James, the apostle, at Santiago de Compostela in Galicia, illustrates the way in which heritage can provide an opportunity to attract and disperse large numbers of hikers into quite remote rural areas and thereby support local economies. Studies of farm tourism in many countries reveal that tourism is usually a supplementary rather than a main source of income. Nevertheless, the aggregate contribution to the wider rural economy can be substantial, as the loss of income arising from restriction of access to the countryside during the foot and mouth crisis in the UK in spring and early summer 2001 illustrated.

Sociocultural Impacts

Tourism and recreational provision are sources of replacement and new employment in areas experiencing economic restructuring. Seasonality and the limited career opportunities that arise for locals are well known weakness of rural tourism employment. Winter ski resorts can benefit from summer tourism but many summer destinations are dependent on occasional festival celebrations to attract low season visitors. When present, full time employment may be held by immigrant managers rather than local people. In the absence of alternatives, seasonal employment is usually valued by local students and underemployed people. However, new farm accommodation providers often receive relatively low returns to their labor and capital inputs initially when demand is limited. In developing areas of the world where wage rates have been traditionally low, local labor may be subject to exploitative relationships.

Tourists' visits are often reported as engendering a sense of pride and self worth in rural societies. However, tourism may be socially disruptive. The apparent affluence of visitors can lead to feelings of inadequacy among local residents and especially among teenagers. Teenage dress and sexual mores may change under the influence of nontraditional norms and alcohol consumption often increases. Major interruption of settlement patterns have been reported in developing areas: of both permanent and temporary settlement in Nepal, to meet the growing demand for accommodation in the environs of Mount Everest, and the displacement of whole communities in Botswana to make way for tourism infrastructure.

The cultural impacts of tourism in rural areas, like the economic and social impacts, are often mixed. Visitor interest in local cultures may lead to a revitalization of traditional music and dance, as among some Aboriginal communities in Australia and New Zealand, although concerns have been expressed about cultural degredation. The annual arrival of large numbers of second level students to spend a 3 week period of language study in the Irish Gaeltacht, during the summer vacation, contributes to the increased use of the Irish language. Vernacular structures have been refurbished as tourist accommodation in many parts of France, Italy, and Spain. Threats to local cultures may, however, arise in association with tourism development. Thus the familial links encoded in local dress in Sámi cultural areas in northern Finland are said to be ignored in the generic costumes introduced through commercialization of tourism. In some cases the type of tourism developed lacks any link with the traditional culture or scale of settlement, as the gaming casinos favored in Indian reservations in the USA illustrate.

Environmental Impacts

Discussion of tourism and recreational impacts in the countryside is closely implicated with discourses relating to environmental conservation and sustainability. Growth in visitor numbers and motorized vehicles on land and water, expansion of infrastructure and associated waste disposal, all pose threats to the physical environment. Most countries have had experience of surface erosion of popular mountain pathways and interruption or destruction of wildlife habitats as numbers of hikers have grown. Even less visually intrusive activities can have unforeseen consequences for ecological systems. Food left by tourists in the Cairngorm Mountains in Scotland was identified during the 1990s as having attracted foxes which preyed on the indigenous ptarmigan and grouse. The single cell organism giardia, which causes gastrointestinal illness in humans and animals, has spread to remote areas in western USA, the Australian Alps, and New Zealand, apparently through unhygienic toilet practices among hikers. Examples such as these illustrate the need for principles of good practice to be followed by countryside visitors, such as those advocated by the Leave No Trace program in the USA.

Policy Concerns

Rural tourism and recreation rely heavily on the physical landscape and on built archaeological and historical structures, as well as on local people and cultures. Institutional policies and structures have been developed to promote tourism and to manage the use of the countryside and its resources for tourism and recreation.

The State as Tourism Resource Provider and Manager

In most countries, the state controls key resources used for recreation and tourism through its ownership of extensive areas of conserved landscapes (national parks in the USA and Canada, for example), the coastline and inland waterways, historical sites, buildings and museums, its investment in transport infrastructure and other public services, and through its grant aiding of entrepreneurial activities. The use of both the public and private resource base is regulated through state legislation, land use, and physical planning provisions. The state at international level also designates areas for protection which have potential tourism and recreational uses: for example, EU Natura 2000 sites, Ramsar Wetland sites (arising from the convention signed at Ramsar in Iran in 1971), UNESCO World Heritage sites. Local communities are being represented increasingly in management structures for protected areas, as recommended by the International Union for the Conservation of Nature. Local Agenda 21, which arises from the 1987 United Nations Commission on Development, provides for community consultation in local government programs which potentially include rural tourism. In Australia, native title rights are recognized over traditional land areas in Crown lands and national parks and Aboriginal Tourism Australia has responsibility for indigenous tourism more generally.

State Support

The Organization for Economic Co operation and Development (OECD), the EU, and national govern ments have all promoted tourism as an alternative source of income in rural areas of developed countries under going economic restructuring since the late 1980s, as has the World Bank and national governments in developing countries. The former command economies of Central and Eastern Europe have identified rural tourism as a sector with particular economic potential. Assistance in the form of grant payments, training, and skills is usually available from the state in recognition of the difficulties inherent in adapting from primary activity to service provision. Private sector banks and other financial insti tutions may also provide special supports. Because of the small size of many rural tourism businesses, their dispersed nature, and a weak tradition of entrepreneur ship, partnership formation has been advocated as a method of developing and promoting assemblages of local tourism products, as part of the EU Leader pro gram, for example.

Integrated Rural Tourism

Recent research findings suggest that a more integrated approach to rural tourism development and management is required in both developed and developing countries to take greater account of the links between tourism, the resource base, other economic sectors, and the com plexity of actors implicated in producing, exchanging, and consuming tourism and recreational experiences. Partnerships and networking are central to this approach but a process of ongoing negotiation is required between the multiplicity of interests involved to ensure that multidimensional sustainability is achieved.

See also: Cloke, P.; Cultural Economy; Economic Development, Rural; Ecotourism; Feminism/Feminist Geography; Heritage and Economy; Leisure; National Parks; Natures, Postcolonial; Non-Representational Theory/Non-Representational Geographies; Performativity; Postmodern City; Postmodernism/Postmodern Geography; Post-Productivist and Multifunctional Agriculture; Resource Management, Rural; Second Homes; Structuration Theory; Sustainability; Tourism.

Further Reading

Baerenholdt, J., Haldrup, M., Larsen, J. and Urry, J. (2004). *Performing Tourist Places*. Aldershot: Ashgate.
Bramwell, B. and Lane, B. (eds.) (2000). *Tourism Collaboration and Partnerships*. Clevedon: Channel View Publications.
Church, A. and Coles, T. (2006). *Tourism, Power and Space*. London: Routledge.
Cloke, P. and Perkins, H. C. (2005). Cetacean performance and tourism in Kaikoura, New Zealand. *Environment and Planning D: Society and Space* 23, 903 924.
Crouch, D. (2006). Tourism, consumption and rurality. In Cloke, P., Marsden, T. & Mooney, P. (eds.) *Handbook of Rural Studies*, pp 355 364. London: Sage.
Hall, C. M. and Boyd, S. (eds.) (2005). *Nature Based Tourism in Peripheral Areas: Development or Disaster?* Clevedon: Channel View Publications.
Hall, C. M. and Page, S. J. (2006). Rural recreation and tourism. In Hall, C. M. & Page, S. J. (eds.) *The Geography of Tourism and Recreation* (3rd edn.), pp 223 252. London: Routledge.
Hall, D., Kirkpatrick, I. and Mitchell, M. (eds.) (2005). *Rural Tourism and Sustainable Business*. Clevedon: Channel View Publications.
Hall, D., Roberts, L. and Mitchell, M. (eds.) (2003). *New Directions in Rural Tourism*. Aldershot: Ashgate.
Perkins, H. (1993). Human geography, recreation and leisure. In Perkins, H. & Cushman, G. (eds.) *Leisure, Recreation and Tourism*, pp 116 129. Auckland: Longman Paul.
Pigram, J. J. and Jenkins, J. M. (2006). *Outdoor Recreation Management* (2nd edn.). London: Routledge.
Sharpley, R. and Telfer, D. (2007). *Tourism and Development*. London: Routledge.
Shaw, G. and Williams, A. M. (2002). *Critical Issues in Tourism: A Geographical Perspective* (2nd edn.). Oxford: Blackwell.
Smith, M. K. and Robinson, M. (eds.) (2006). *Cultural Tourism in a Changing World*. Clevedon: Channel View Publications.
Urry, J. (2002). *The Tourist Gaze* (2nd edn.). London: Sage.

Relevant Websites

http://www.iucn.org
International Union for the Conservation of Nature.
http://www.ec.europa.eu/agriculture/ruraldev
Leaderplus Provides Examples of Rural Tourism in Practice, EU LEADER Program.
http://www.lnt.org
Leave No Trace Program.
http://www.oecd.org
OECD: Has Useful Publications Relating to Rural Tourism.
http://www.ramsar.org
Ramsar Convention on Wetland Sites.
http://www.unesco.org
UNESCO Man and Biosphere Reserves; UNESCO World Heritage List.
http://www.world tourism.org
UN World Tourism Organization.

Tourism, Urban

T. Hall, University of Gloucestershire, Cheltenham, UK

Introduction

Reviewing the literature on urban tourism leads to the conclusion that it is a somewhat peripheral concern within human geography. It is relatively rare, although not un known, for human geographers to write exclusively about urban tourism, but urban tourism is often acknowledged as an important element within broader discussions of con temporary urban change and development, particularly where these focus on urban regeneration. However, in being subsumed within these broader discussions, urban tourism remains somewhat underexplored and theorized within the context of human geography. This mirrors, to an extent, a neglect of urban tourism in disciplines such as tourism studies and urban studies, as noted by Gregory Ashworth in his two seminal reviews of the subject. Ash worth noted a failure by tourism researchers to acknow ledge the contexts within which tourism occurs and by urban studies researchers who have consistently failed to recognize the significance of tourism to the economies, not to mention the cultures and landscapes, of towns and cities. The relative invisibility of urban tourism has long been a weakness of these disciplines, but is particularly so in the context of the postindustrial economy where such a volume of diverse tourist activities now take place in urban areas and where tourism is emerging as an increasingly significant driver of urban development in a growing range of urban areas.

Having said this, since the early 1990s, there has been a growth in the literature on urban tourism. However, this literature has been somewhat diffused in disciplinary terms and only serves to confirm that urban tourism has not emerged as a distinctive research focus within human geography in its own right. It is instructive to note, for example, that urban tourism has no dedicated inter national journal. Of the literature that has emerged we can recognize three prevailing themes. First is a focus on the nature, characteristics, and diversity of urban tourism destinations, which typically recognize the diversity of groups who use them. Second is a consideration of the interrelationships between the supply of tourist facilities, the demands of tourists, and urban tourism policies. Fi nally, urban tourism has been seen as a characteristic of the postindustrial economy and examined in the contexts of structural changes in late capitalism, a shift toward entrepreneurialism in city governance, and the emer gence of postmodern processes of urbanization. The last of these is of most interest from the perspective of critical human geography, but does tend to again subsume urban tourism within wider discussions of urban change. While this can be seen as appropriate, in that it seeks to con textualize urban tourism and make links to other aspects of urban development, it has tended to compound the failure of urban tourism to emerge as a distinctive focus of research both within human geography and cognate disciplines. This article concludes by discussing three growing areas of urban tourism research of particular interest to human geography and in which human geographers are well placed to contribute both empiric ally and in terms of the methodology and theoretical development or urban tourism research.

Historical and Contemporary Patterns

Urban tourism, defined as visits to settlements not linked to their service functions or for military reasons, can be traced back to the Middle Ages to the practice of re ligious pilgrimage. The development of cities such as Canterbury in England or Santiago de Compostella in Spain reflects this historical function. Although pilgrims would not have recognized themselves as tourists, their journeys share sufficient characteristics with modern definitions of tourism for them to be classified thus. Pil grimage remains a significant form of urban tourism to a range of destinations such as Lourdes in France, and Mecca. The eighteenth century was also a significant period in the history of urban tourism that saw the emergence of a range of tourist practices, all largely confined to the wealthy, which were predominantly centered on urban areas. These included travel to spa towns such as Cheltenham and Bath in the southwest of England for health and medical reasons, travel to capital cities for cultural entertainment and social engagements during the so called ‘season’, and the grand tours undertaken around European cultural centers by young aristocrats. Travel motivations began to diversify in the mid nineteenth century and early forms of business and events tourism emerged in the form of large exhibitions of arts, crafts, industry, and technology held in large cities. The first of these was held at Crystal Palace in 1851. Later examples included those held in Paris in 1855 and 1889 and Barcelona in 1888 and 1929. These were often significant vehicles both in the promotion and de velopment of the cities in which they were held. They also heralded the ongoing importance of events of various kinds in the histories of urban tourism and urban development. Business travel, more generally, expanded over through the twentieth century with the growth of international trade and the globalization of business.

The industrialization of Europe and beyond from the mid nineteen century is significant because it saw the formal recognition of leisure time, as distinct from that of work, and the emergence of set periods within the cal endar which were deemed holiday time. A consequence of this was the development of mass tourism centered predominantly in seaside resorts such as Blackpool in the north of England which were accessed by an expanding railway network from nearby industrial metropolises. The late twentieth century has seen the continued evo lution of mass tourism alongside the emergence of a variety of niche tourist markets. The growth in people's leisure time and increasingly flexible patterns of working has seen people take more frequent holidays, often of shorter duration, accessing a greater diversity of destin ations. Many of these niche markets are centered on some form of urban tourism; the growing popularity of 'city breaks' in the early twenty first century has been evi dence of this.

Classifying Urban Tourism

Attempts to define and classify urban tourism reflect the complexity of the process. Classifications of urban tour ism destinations have sought to reflect a diversity of elements including the range of tourist facilities they possess, the origins of visitors, the popularity of destin ations, the processes of tourist development, and the nature of places. They have typically sought to either categorize the supply side of urban tourism (destinations and their facilities) or the demand side (tourists). Judd and Fainstein, for example, recognize a threefold classification which distinguishes between tourism ur banization (planned or unplanned specialized resorts), tourism historic cities/ancient cities and converted cities while Jansen Verbeke has classified the facilities within tourist destinations recognizing a distinction between primary elements, secondary elements, and additional elements (**Table 1**). Burtenshaw *et al.* offer a rare ex ample that has sought to combine both the demand and the supply side within a single model (**Figure 1**).

Table 1 The elements of tourism[a]

Primary elements	
Activity place	Leisure setting
Cultural facilities	*Physical characteristics*
Concert halls	Ancient monuments and statues
Cinemas	Ecclesiastical buildings
Exhibitions	Harbors
Museums and art galleries	Historical street pattern
Theatres	Interesting buildings
	Parks and green areas
Sports facilities	Water, canals, and river fronts
Indoor and outdoor	
Amusement facilities	*Socio-cultural features*
Bingo halls	Folklore
Casinos	Friendliness
Festivities	Language
Night clubs	Liveliness and ambience of the place
Organized events	Local customs and costumes
	Security
Secondary elements	*Additional elements*
Hotel and catering facilities	Accessibility and parking facilities
Markets	Tourist facilities: information offices, signposts, guides, maps, and leaflets
Shopping facilities	

[a] Based on Jansen Verbeke, M. (1986). Inner city tourism: Resources, tourists, and promoters. *Annals of Tourism Research* 13(1), 79 100.

Urban Tourism Data

One thing that has bedevilled urban tourism research has been the lack of international statistical data to allow comparative analysis. The major sources of international tourism data, the World Tourism Organization (WTO) and the Organization for Economic Cooperation and Development (OECD) do not collect data specifically on urban tourism. The urban tourism data that does exist is either produced by a range of national and regional tourism organizations, local authorities, commercial or ganizations, or academics, or is derived indirectly from WTO and OECD data. The former sources raise po tential issues of comparability related to the different methods through which data is collected and the cat egories used to collect and organize data. While statistics from these sources cover many aspects of urban tourism including numbers of arrivals, volumes of trips, origins of visitors, types of tourism and tourist motivations, trip duration, tourist activities, use of services, and expend iture, currently there are no standard international data sources for urban tourism that allow systematic, com parative analyses of tourism between different cities.

Tourism and Urban Development

Prior to the 1980s it was rare for policy makers, local authorities, or commercial interests in urban areas, apart from specially developed tourist resorts or those with long traditions of tourism, to recognize tourism as a significant economic activity, either actual or potential. However, the coincidence of widespread de industrial ization of urban economies in the older industrialized

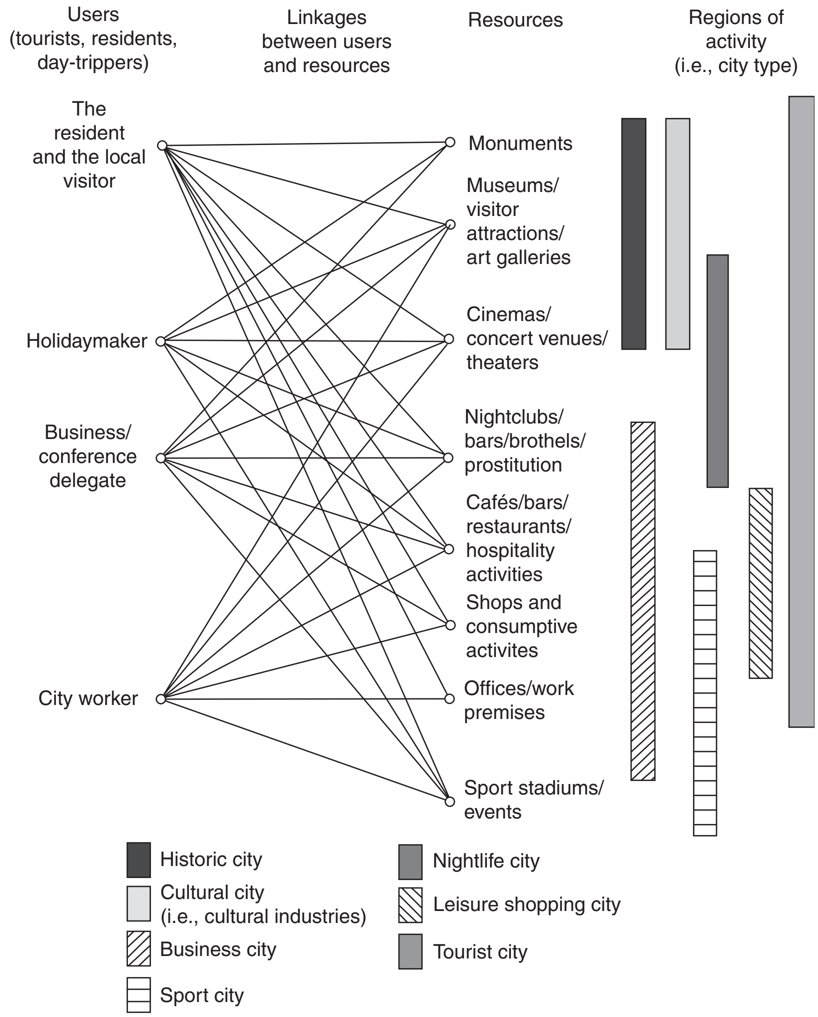

Figure 1 Functional areas in the tourist city.

countries and the steady growth of world tourism saw an increasing range of urban areas exploring the potential of tourism as a way to address the interrelated problems of economy, image, and environment. The model of tourism based on urban regeneration pioneered by Baltimore in the USA, who in the late 1970s redeveloped an extensive area of its derelict Inner Harbor by creating a number of leisure attractions, museums, a convention center, hotels, a festival shopping center, and later sports stadiums, became a blueprint adopted by many cities internationally.

Typically the resultant developments were mixed use (to cater for leisure, retail, cultural, and business tourists), were often anchored around one or more major 'flagship' regeneration projects, combined new developments with the reuse of historic urban fabric, were often based around derelict waterfronts, received extensive public funding, and were accompanied by programs of place promotion designed to address often negatively perceived urban images. Increasingly cities also competed for temporary events or festivals such as Olympic Games or European 'City of Culture' celebrations, or developed their own events, and used these as catalysts for regeneration. This emphasis on tourism based programs of economic development has been recognized as symptomatic of a shift toward an entrepreneurial city politics in which cities have become more innovative, risk taking, speculative, and competitive than had been the case in the past. Commentators have recognized that the priorities of city governments have shifted from a managerial concern with social welfare achieved through the distribution of resources among urban populations toward an entrepreneurial concern with capturing external sources of investment, tourists and their spending key among them.

Two cities that exemplify the increasing importance of urban tourism are Birmingham in the UK and Barcelona in northern Spain. Birmingham's regeneration, initiated in the mid 1980s, included the building of a convention center, opened in 1991 at the cost of £180 million, largely financed through public funding, and associated hotel developments, extensive redevelopment of its central area, and subsequently a major retail re development. These were a response to the de indus trialization of its predominantly manufacturing economy in the late 1970s and early 1980s, which saw over 300 000 jobs lost in the city between 1980 and 1982. Associated problems that regeneration sought to address included derelict industrial capital adjacent to the city center, an unpopular post war redevelopment of the city center, and a poorly regarded external image. The regeneration of the city has led to a significant increase in tourist spending within Birmingham. For example, it has been estimated that the 1998 International Rotary Meeting held at the International Convention Center attracted 23 000 delegates resulting in spending of almost £20 million in the city during the meeting, much of which remained in the local market. There is also evidence of economic diversification within the city's economy as sociated with its growing tourist economy as well as less tangible positive impacts on civic pride and the city's image. Despite this some critics have argued that this has done little, if anything, to improve levels of social welfare among the city's most deprived populations or reduce social exclusion.

While apparently sharing many similarities with Bir mingham, in that it has targeted cultural and business tourists through flagship developments, Barcelona has adopted a somewhat different approach to the develop ment of its urban tourist economy. Some critics have accused Birmingham of what Greg Richards (1996) called "a tendency for culture to be produced for tourist consumption". Diane Dodd has argued that Barcelona, by contrast, pursued tourist development both as a way of promoting economic development and as a platform for the re invigoration of its regional culture and identity. The pursuit of cultural, and to a lesser extent business, tourism was in part a reaction to the negative impacts of mass beach tourism in surrounding resorts, in part a way of addressing the economic problems stemming from the city's de industrialization, and a recognition of the growing significance of these sectors of the world tourism market. The conscious positioning of local culture alongside urban regeneration was crucial in gaining public support for the catalyst event, the hosting of the 1992 Olympic Games in the city. This, and subsequent initiatives, saw massive infrastructural improvements throughout the city, numerous major landscaping pro jects, and the provision of a number of new cultural fa cilities such as the Museum of Contemporary Art, Barcelona. Barcelona is widely regarded as a model of successful tourist based urban regeneration. There is no denying the impacts on the physical and cultural infra structure of the city and its tourist economy. However, some critics have voiced concern over the impacts of gentrification associated with the development of cultural facilities in marginal neighborhoods and overinvestment in the hard infrastructure of cultural provision rather than the creative process, citing the lack of funding for local and community arts groups, for example. Finally, the rapid growth in stag and hen party tourism in the early twenty first century, and the popularity of Barce lona as a destination, is threatening to taint the experi ence of those tourists who travel to the city for more refined cultural activities.

These examples illustrate the heightened recognition among local officials and policy makers of the signifi cance of tourism to urban development. It is common for the media to talk about both of these cities in terms of an apparent transformation since the mid 1980s. While one might dispute the substance of such journalistic claims, it is impossible to deny the role of tourism in fueling the changes that have occurred within both cities.

Issues

There are a number of issues within urban tourism where a paucity of research has been identified or which are areas of current interest across a range of cognate dis ciplines. Three are identified and discussed below. These are all areas with which human geographers have en gaged or are currently undertaking research within. These constitute a research agenda into which human geographers could contribute and through which cross disciplinary dialogs might profitably develop.

The Urban Tourism Experience

Studies of the urban tourism experience include exam inations of tourists within the tourist environment. They might encompass studies of the relationships between a number of elements including: the motivations of tourists and the importance of their prior knowledge of en vironments, tourists' perceptions and cognition of the environment, the values or importance that tourists at tach to different elements of the environment, and the behavior and activities of tourists.

Despite well developed traditions in both social psychology and behavioral geography of researching people's perceptions of the environment, there have been relatively few studies of the ways in which tourists make sense of the environments of tourist destinations. This is surprising given the obvious value of such research to the planning, development, policy, and management of urban tourism destinations and, indeed, to all types of tourist destinations.

A pioneering study of the urban tourism experience was that conducted by Jensen Verbeke comparing the values that tourists and residents attached to different elements of city center environments finding both dif ferences in the motivations underlying their visits and similarities in the elements of the environment they valued. Despite the value of Jensen Verbeke's contri butions, a number of changes that have occurred since the original research was conducted make it timely to revisit and update the study. Since Jensen Verbeke undertook his study many city centers have seen con siderable increases in their residential populations and considerable growth and diversification of their tourist economies. It would be instructive, for example, to con sider the impacts of the re imagination of urban areas in the West as sites and landscapes of leisure and the in creasing promotion of city center lifestyles through touristic rhetoric and discourses. Rather than simply re running Jensen Verbeke's study it might be profitably expanded to compare the values and attitudes of different groups of tourists, for example, leisure and business tourists, new populations of wealthy city center resi dents, established residents of less affluent inner city areas, and outer city residents. This is one of many areas where human geographers might contribute to studies of the urban tourist experience.

The Impacts of Urban Tourism

Tourists have a number of impacts on destinations; often of the most concern to destinations are the economic impacts associated with expenditure on travel, accom modation, food, and entertainment. Tourist spending has a multiplier effect through local economies. Tourists spend money directly, for example, on travel, hotels, food, and drink. This in turn generates indirect expenditure as tourist facilities pay staff salaries, bills, and taxes, and buy supplies. Finally induced expenditure refers to the spending of those employed in the tourist industry. However, there is also a leakage effect as money leaves the local economy through taxation, purchases from external suppliers, or as profits to external companies who own and operate tourist facilities within destinations. Generally speaking, the larger the spatial scale under consideration, the greater the potential for self sufficiency and hence the greater the chance that tourist expenditure will circulate throughout the local economy rather than leak out of it.

In addition to the positive economic impacts of tourism, there are potential economic costs. Successful tourist developments can have inflationary pressures, for example, on the cost of land. This can have negative impacts on local populations who might be displaced by tourist led gentrification of residential areas. Tourism can also bring with it additional costs associated with mar keting and the maintenance of facilities. Tourism is also a very vulnerable and potentially unstable industry in a number of ways. External threats to tourist destinations include competition from other destinations and new developments, shifts in patterns of tourist activity and changes in taste, reductions in the numbers of international travelers, and concerns over the perceived safety of specific destinations stemming from media re ports of crime or from terrorist attacks, currency ex change rates making certain destinations relatively expensive, and events such as natural disasters.

In the longer term, tourist destinations have been recognized as possessing a life cycle. This is characterized by phases of growth and maturity, with rising and then steadying visitor numbers, followed by a period of de cline as tastes change or technology and price changes make other destinations more attractive. The rise and subsequent decline of many 'traditional' British seaside resorts fit this model well. The costs of bucking these life cycle trends and the risks of these inherently speculative ventures are very high.

Another potentially negative aspect of tourism is the criticism that many jobs in the tourist industry are low paid and low skilled. Many destinations also have a distinctive 'season' during which tourists visit. The problems of lack of tourist income out of season are obvious as are the prob lems associated with quality of life for residents of tourist destinations when the majority of facilities are closed down. Finally, some critics have recognized the potential conflicts between tourists and residents, for example, as sociated with congestion at peak visitor times and sites.

Tourism also has social and cultural impacts on its host communities. A number of studies have attempted to devise indexes or matrixes that capture the attitudes of local residents to tourism or their relationship with them. These include Doxey's Index of Tourist Irritation (Irri dex). Doxey argued that host communities pass through four stages in terms of their attitude toward tourists (euphoria, apathy, annoyance, and antagonism). Such devises offer useful frameworks within which to think about the social and cultural relationships between hosts and tourists but they are not subtle enough to capture the truly complex nature of these relationships. Within any one city there is likely to be a wide range of feelings toward tourism and, indeed, individuals may hold a range of attitudes toward tourism. For example, while certain residents might welcome cultural or business tourists, and the investment in and development of associated cultural and business facilities, they might feel less wel coming toward sports tourists and weekend tourists as sociated with stag and hen parties or they might express moral concerns over gambling tourism. Again recognition of the complexities of attitudes and behavior and the diversity of the tourist economies of towns and cities is important to appreciate the nature of the social and cultural impacts of urban tourism.

Tourism and the Postmodern City

Urban geographers since the 1980s have argued that we have been witnessing the emergence of the new urban forms that they have characterized as postmodern cities. They agree that these new cities are more fragmentary in their form, more chaotic in structure, and are generated by different processes of urbanization than earlier cities. Rather than being a single coherent entity, postmodern cities are typically described as consisting of a number of large, spectacular, residential and commercial develop ments with large environmentally and economically de graded spaces in between. They are said to resemble a pattern of stars floating in space rather than the unitary metropolitan development growing steadily outward from a single center. Another element of the post modernization of cities has been the increasing in corporation of fantasy and hyperreality into urban landscapes. This is perhaps most developed in certain specialist tourist resorts such as Las Vegas or the Disney resorts.

The case of urban tourism highlights limitations in debates about the postmodern city. While tourism has been recognized as a characteristic of the postmodern city there has been little discussion of the role of tourism *per se* in the development of postmodern city forms and urban landscapes, save some discussions of the fantasy landscapes of specialist destinations such as Las Vegas. This is less a reflection of the significance of tourism to the postmodernization of cities, more a reflection of the limited empirical foundations upon which debates rest. Tourism is a significant aspect of the economies of cities such as Los Angeles, London, and Tokyo. It is not, however, a significant driver of the postmodernization of these cities relative to other elements of their economies such as concentrations of corporate headquarters or their roles as international financial markets.

By contrast though, as one moves down the urban hierarchy to major provincial cities the centrality of tourism to the processes of their postmodern urban ization become much more apparent. These cities have tended to lack the sorts of links to the global economy that have driven the recent development of cities such as Los Angeles and Tokyo. Rather cities such as Birming ham, Glasgow, and Barcelona have had to forge new links to international circuits of capital. It is an almost ubi quitous response of cities of this kind to do so through the development of their tourism economies. Post modernism as a process of urban development has been written into the landscapes of a growing number of major provisional cities, and it is done so, to a significant degree in these cities through major tourist developments such as Birmingham's International Convention Center or Bilbao's Guggenhiem Museum, or events such as 'City of Culture' celebrations, or the hosting of Olympic Games.

Studies of urban tourism highlight both the limitations characteristic of discussions about postmodernism in cities and the failure to situate discussions of urban tourism in wider theoretical debates. Human geographers are ideally placed to expand discussions about the post modern city beyond the currently limited range of ex amples. In doing so they should aim to highlight the broader significance of urban tourism to ongoing pro cesses of urbanization.

See also: Heritage and Economy; Postmodern City; Regeneration to Renaissance.

Further Reading

Ashworth, G. J. (1989). Urban tourism: An imbalance in attention. In Cooper, C. P. (ed.) *Progress in Tourism, Recreation, and Hospitality Management*, vol. 1, pp 33 54. London: Belhaven Press.
Ashworth, G. J. (1989). Urban tourism: Still an imbalance in attention. In Cooper, C. P. (ed.) *Classic Reviews in Tourism*, pp 143 163. Channel View: Clevedon.
Burtenshaw, D., Bateman, M. and Ashworth, G. J. (1991). *The City in Western Europe* (2nd edn.). Chichester: Wiley.
Dodd, D. (2003). Barcelona: The making of a cultural city. In Miles, M., Hall, T. & Borden, I. (eds.) *The City Cultures Reader*, pp 177 182. London: Routledge.
Doxey, G. V. (1975). A causation theory of visitor resident irritants: Methodology and research inferences. pp 195 198, *Proceedings of the Travel Research Association Sixth Annual Conference*, San Diego, CA: Travel Research Association.
Hall, C. M. and Page, S. J. (2006). *The Geography of Tourism and Recreation* (3rd edn.). Abingdon: Routledge.
Jansen Verbeke, M. (1986). Inner city tourism: Resources, tourists, and promoters. *Annals of Tourism Research* 13(1), 79 100.
Jansen Verbeke, M. (1988). Leisure, Recreation and Tourism in Inner Cities. Explorative Case Studies, *Netherlands Geographical Case Studies 58*, Amsterdam: Koninklijk Nederlands Aardrijk skundig Genootschap.
Judd, D. and Fainstein, S. (1999). *The Tourist City*. New Haven, CT: Yale University Press.
Kotler, P., Asplund, C., Rein, I. and Haider, D. H. (1999). *Marketing Places Europe: How to Attract Investments, Industries, Residents, and Visitors to Cities, Communities, Regions, and Nations in Europe*. London: Pearson.
Law, C. M. (2002). *Urban Tourism: The Visitor Economy and the Growth of Large Cities*. London: Continumm.
Page, S. (1995). *Urban Tourism*. London: Routledge.
Richards, G. (1996). *Cultural Tourism in Europe*. Walingford: CABI.
Shaw, G. and Williams, A. M. (1994). *Critical Issues in Tourism: A Geographical Perspective*. Oxford: Blackwell.

Relevant Websites

http://www.tourisminbarcelona.com
Barcelona.
http://www.visitbirmingham.com
Birmingham.
http://www.sogaer.ex.ac.uk/geography/tourism/gltrg/welcome.html
Geography of Leisure and Tourism Research Group.
http://www.tourismconcern.org.uk
Tourism Concern.
http://www.ratztamara.com/tourism.html
Tourism Research.
http://www.world tourism.org
World Tourism Authority.

Trade Blocs

J. P. H. Poon, University at Buffalo-SUNY, Buffalo, NY, USA

Glossary

Bilateralism Trade or political relations between two states.
Comparative Advantage The ability of a country to produce a good at a lower opportunity cost compared to another country.
Free Trade Area When two or more countries agree to eliminate trade barriers between themselves while keeping national trade barriers against nonparticipating countries.
Friction of Distance A term to describe the negative effect of distance between places.
Multilateralism When two or more states coordinate national policies and relations among themselves according to some principle.

Introduction

The role of international commerce and trade has long been of interest to human geographers. One of the earliest publications on the topic may be found in Chisholm's 1889 *Handbook of Commercial Geography* which became an important guide book for British merchants and traders. For contemporary human geographers, international commerce and trade constitutes an important inquiry because of the inherent spatiality of trade flows, particularly the influence of distance on the level of trade between countries. In general, trade is highest between geographically proximate countries, for instance, one of the world's largest bilateral trade flows occurs between the United States and Canada. This positive relationship between trade interactions and geographical proximity has spurred countries to enhance economic interactions through regional arrangements such as the 1994 North American Free Trade Area (NAFTA), the 1965 European Economic Community, now the European Union (EU), and the 1992 Association of Southeast Asian Nations (ASEAN) Free Trade Area (AFTA). As of 2005, countries of the EU consist of the original European Economic Community (EEC) countries including Belgium, Denmark, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, and the United Kingdom. Austria, Finland, and Sweden joined the EU in 1995 and ten more countries were added since 2004, that is, Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia, and Slovenia, bringing the total number to 25. Members of AFTA comprise the ten Southeast Asian countries of Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand, and Vietnam. The benefits of regional trade arrangements, however, have been hotly debated because of the fear that such regional arrangements will assume bloc like characteristics that exclude the participation of nonmembers. The term 'trade blocs' has thus evoked not only positive and negative arguments in the academic community but its realization is also contested among civic groups including labor and environmental groups.

Trade Blocs: Pros and Cons

Broadly speaking, the benefits and disbenefits of a trade bloc may be evaluated in terms of two schools of thought. The first school led by liberal economists and international relations scholars argues that trade blocs are a suboptimal solution to govern economic interactions and relations compared to free trade and multilateralism. Both free trade and multilateral principles are based largely on the logic of comparative advantage that state that a country is better off specializing in the production of a good that it does best than producing all of its own goods. This nation should then trade for another good that it does not produce even if it possesses absolute advantage in the production of both goods. The main idea here pertains to that of the country's relative advantage and the benefits of specialization arising from such advantage. Noncoercive trade between countries in turn is more likely to create multilateral economic exchanges.

In international relations, multilateralism facilitates the coordination of policies among groups of countries. Under the multilateral principle of international trade, trade relations are organized around the principle of nondiscrimination, namely the most favored nation (MFN) treatment that is embodied originally in the General Agreement of Tariffs and Trade (GATT), the latter of which has since been subsumed by the establishment of the World Trade Organization (WTO) in 1995. Regional blocs, on the other hand, are thought to be discriminatory since they involve some form of preferential treatment to members of the arrangement. Under GATT, provisions for regional trade agreements (RTAs) that underscore trade blocs are speltled out in Article XXIV. These provisions enable countries to form free trade agreements (FTAs) under certain conditions such as the facilitation of trade between FTA members such that they do not raise barriers to the trade of other contracting parties. With the creation of the WTO, a Committee on Regional Trade Agreements (CRTA) has been formed to provide a means for reconciling regional trade initiatives to the organiza tion's multilateral trading principle. However, advocates of multilateralism consider regional arrangements to be a less desirable spatial vehicle for organizing international trade, since they potentially compartmentalize trade interactions into closed intrablocs and discourage infor mation flows between regions thereby increasing the possibility for interbloc conflicts. Indeed for political analysts, the strongest argument for multilateralism and free trade is that they provide the basis for international growth and political stability.

According to the GATT/WTO, there are close to 150 regional trading agreements as notified by its members, and another 100 in force since 1948. Most WTO mem bers are participants of at least one of such agreements though the share is much higher for European (60%) than developing countries (15%). The proliferation of RTAs suggests that despite strong arguments for non discriminatory multilateral trade, regional governance remains a popular mode of geographical coordination. This may be due to the perceived difficulties of achieving free trade among a large number of countries, for ex ample, the demands of environmental and labor stand ards by industrialized countries in international trade practices which are viewed to be a form of disguised protectionism by developing countries. More recently, long standing trade deficits between the US, as well as Europe, and Japan, and growing trade deficits with newcomers like China, have led to demands for fairer trade. These demands reflect a concern that mercantilist countries in Asia are not driven by static comparative advantages so that economic solutions need to be sought at the regional than global level. Growing support for trade arrangements in the form of geographical blocs has in turn generated new arguments for regional than multilateral cooperation and thereby the relevance of bloc making. The possibility of trade blocs as an efficient spatial mechanism to manage and regulate economic and political spaces underscores the second school of thought which is elaborated further below.

First, there is the argument that geographically proximate trade is efficient while small is beautiful. Krugman has shown that countries trade more with their neighbors than with distant countries arising from lower transport and transaction costs. Countries that form trade blocs on the basis of 'natural' geography are thus likely to benefit from such an arrangement. On the other hand, when non natural blocs drive trade arrangements, this is less beneficial since such blocs are more likely to create trade diversion from nonmembers. In other words, non natural blocs are negatively regarded in economic orthodoxy because of their welfare reducing implications. Second, a larger number of countries also increase the complexity of the nature of economic interactions so that regulatory frameworks are much more difficult to formulate. An increasingly popular re sponse in the regulation of cross country trade in the context of large number problems has been the down sizing of international economic relations to minilateral relations.

Large number problems may be illustrated by the number of countries that are members of the WTO, that is, 149 from less than 30 when GATT was formed in 1948. Tariff reductions under GATT occurred in a series of trade rounds, the last of which was associated with the Uruguay Rounds (1986–94). That the trade rounds were contentious, protracted, and often difficult, and that they did not result in reductions in barriers on trade in agri culture or services have led to the belief that trade ne gotiations could be more effective when applied to a smaller number of countries. In turn, the world trading system could be more effectively managed under a sys tem of regional blocs where trade liberalization goals are more modest so that the overall outcome is that of minilateralism than multilateralism.

Other economic arguments for minilateral blocs may be found in the new or strategic trade theory. The new trade theory suggests that contrary to conventional wis dom, comparative advantage is arbitrary. States that intervene strategically, particularly in certain technolo gical industries, could exploit regional spillovers from economies of scale, thereby creating comparative ad vantage in these nations. The new protectionist impli cations of the strategic trade theory saw a shift in the United States away from its post war free trade policies in the late 1980s. Geographer Richard Grant notes that domestic institutional changes through interactions be tween international and national politics determine US trade policies. Hence while US hegemony was strengthe ned by free trade policies in the post war years, under pressures from increased competition from Japan and Germany, regionalism became an increasingly popular foreign policy tool in the 1980s. The Clinton adminis tration, for instance, was much more persuaded by ar guments of the nonbenefits of free trade and the advantages of strategic trade policies. To some com mentators, US departure from its post war policy of trade liberalization and free trade unleashed a domino effect as other countries began to convert to a more regional mode of economic operation in trade.

In addition, from a political and international relations point of view, trade that is conducted between members of a regional institution (e.g., preferential trade agree ment) can help to reduce the likelihood of military conflicts. Similar to the argument of minilateralism, some political economic analysts believe that multilateral in stitutions are best achieved at a smaller geographical scale, and international stability is more attainable through the pursuit of peaceful regional neighborhoods. Regional trade agreements facilitate the lowering of transaction costs through the increase in informational flows and institutional linkages. This in turn fosters greater economic interdependencies that are conducive for resolving collaboration problems between countries. John Maynard Keynes, for example, anticipated such a positive relationship for Europe, as well as Turkey, Egypt, and India in the 1950s with his proposal for a free trade union among these countries.

The Geography of Trade Blocs

From a human geography perspective, regional trade blocs more closely approximate Krugman's natural blocs. In her study of world trade flows, Poon found that most intercountry trade interactions are organized along continental lines and that such regionalization has strengthened over time from 1980 to 2000. Similarly, O' Loughlin and Anselin found evidence of bloc stability around the three cores of the United States, Germany, and Japan from 1968 to 1992. This is because inter national trade is influenced by the friction of distance, that is, both absolute and relative distance. This reflects the costs of economic exchanges across space. However, whereas economic analysis tends to examine geo graphical costs typically in terms of absolute costs (e.g., higher transport costs with greater physical distance) through gravity like model formulations, human geog raphers are just as interested in relative costs that are associated with cultural and political factors.

The US–Israel FTA as well as Korea's FTA with Chile illustrate the expansion and integration of economic space that is based not so much on lower physical costs from natural geography as in political factors: the US displays a strong geostrategic interest in Israel, while political opposition from Korea's farmers and historical sensitivities associated with neighboring Japan have forced the Korean government to establish a trade agreement with a distant than nearby country. Political economic factors also help to explain the continental economic integration of Canada and US that sub sequently developed into the present NAFTA. Despite attempts by Canada to reduce its dependence on the US market for its staples and resources, including the control of foreign direct investment, continental regionalism proceeded because of the possibility of a regulatory framework for resolving long standing trade disputes (e.g., U.S. protectionist tendencies toward Canada's lum ber industry) through the establishment of dispute mechanisms.

Similarly, cultural factors can also play a role in motivating bloc production. Cultural factors are an important explanatory variable in Asia's growing intrar egional trade and investment among overseas Chinese firms which have established dense interfirm trade net works through advantages of language and ethnic knowledge and interrelations. This has led to calls to coordinate the increase in density of interstate trade interactions through regional arrangements. However, for reasons that are elaborated below, bloc production has progressed only very slowly in the Asia Pacific.

Trade Blocs and Developing Countries

Among developing countries, trade blocs are formed even when intraregional trade levels are low. AFTA's share of intraregional trade is less than 20% suggesting that most countries in Southeast Asia are trading with non Southeast Asian countries. Moreover, intraregional trade is dominated by bilateral flows between Singapore and Malaysia, and Singapore and Indonesia. Similarly, intraregional trade among the MERCOSUR countries of Brazil, Uruguay, Paraguay, and Argentina is only about 15%. In contrast, over 60% of trade in the EU is intraregional suggesting that a majority of its members are trading with one another. One explanation for the difference in trade patterns between blocs from de veloping countries and those of industrialized countries is that developing countries associate bloc making with a potential loss of political sovereignty in the former.

In the Asia Pacific, scholars have pointed to the dearth of regional institutions despite rising levels of intra regional trade and investment, and the articulation of a localized economic space. Intraregional trade is pre dominantly driven by two sources: the outward direct investment of Japanese firms in Asia and the latter's concomitant positive effect on increased intrafirm trade, and the economic activities of overseas Chinese firms. Economic regionalism is particularly strong among the more industrialized countries of Japan, South Korea, Singapore, Taiwan, and Southern China. In addition to the activities of overseas Chinese firms, Yeung also points to the integrative forces arising from the regional net works of transnational corporations. Despite these posi tive integrative trends, East Asia remains one of the least regionalized areas in terms of formal institutional gov ernance: there is no sign of an East Asian free trade area being formed anytime soon, for example. Meanwhile, AFTA took nearly 30 years to form and its tariff elim ination schedule was moved forward to 2003 from 2008 in order to compete for foreign direct investment that is increasingly being directed to China. The most common explanation for the apparent lack of bloc production in the Asia Pacific is associated with the sovereignty sensitivities of the nations here. Regional institutions require a set of rules that govern economic exchanges, but countries here are averse to power delegation at a supraregional scale. Historical difficulties associated with Japan's World War II activities hinder the most industrialized core in the region from organizing Asian countries in the same manner that the US is organizing Latin America in the proposed free trade area of the Americas. China is an emerging and potential core in the future but its role is being contested by the United States which has remained lukewarm in regional initiatives by Asian countries.

Suspicions regarding supraregionalism are also manifested in narratives of Asian cultural values that suggest that legalization and formal regulation accompanying Western forms of institutional organization are irrelevant to Asian regionalism because the latter is characterized by more informal forms of negotiations and political convergence around mutual consensus than formal rules. These suspicions have meant that institutional linkages have been slow to form. ASEAN, for instance, was largely under institutionalized until the recent implementation of its free trade agreement. Similarly trade liberalization is voluntary than mandatory among member countries of the pan Pacific Asia Pacific Economic Cooperation (APEC). Instead more formal regional trade relations are being proposed or formed on the basis of bilateralism, for example, the free trade agreements between Singapore and the US, Singapore and Australia, and Thailand and Australia. Critics see the proliferation of bilateral trade agreements in Asia as a reflection of the difficulties of larger regional deals associated with political sensitivities.

Gunning too notes the difficulties of implementing regional trade blocs in Africa. Like many Asian countries, intraregional trade is low so that "free trade within the bloc would generate no trade." Yet Africa has pursued several regional arrangements (e.g., the Common Market for Eastern and Southern Africa or COMESA, and the South Africa Zimbabwe FTA) under the assumption that this would enable the exploitation of economies of scale with increased competition and enlarged markets. While the progress and economic performance of these blocs have been dismal, noneconomic objectives such as improved security in the context of Southern African Development Community (SADC) appear to have fared better. Gunning further concludes that bloc production in sub Saharan Africa may do better to emphasize infrastructural improvements than trade liberalization, since many countries here suffer from high transport costs and thereby investment diversion arising from poor locational incentives.

While the discussion above suggests that bloc formation is driven by rather different motivations among developing countries, it has also become apparent that trade blocs are being formed between countries from both developed and developing countries. NAFTA is a good example of this and there is a possibility of expanding NAFTA to the rest of South America in the proposed Free Trade Area of the Americas. Meanwhile, regional enlargement is also being contemplated in the EU with its eastward and southward drift. Part of such increased north–south integration may be explained by the new logics of geographies in bloc formation. Developing countries in the 1960s, for instance, were much more inclined to form trade blocs on the assumption that this would enlarge a large domestic market that would promote local infant industries and force an industrialization process away from historical core–periphery dependency on the West. These undertakings did not turn out to be successful in part because the lessons of import substitution indicated a loss in international competitiveness arising from rent seeking. As export promotion became increasingly popular in the 1990s, led by several East Asian countries, reintegration of developing countries to the North came to center on the search for export markets in industrialized countries, and new comparative advantages or complementarities.

Conclusion

This article has examined the interrelationship between trade blocs and geography. A central theme is that most trade blocs are regional in nature because of the advantages of geographical proximity including lower transport costs. However, not all trade blocs are motivated by considerations of absolute costs and natural geography. Several regional FTAs, for example, consist of countries that are non neighbors and are motivated by noneconomic factors such as cultural or political relationships. This suggests the possibility of multiple scales of economic and potentially political and institutional governance as nations confront the increasingly global scale of economic activities including financial integration and social movements (e.g., environmental and labor), not just trade and investment. On one end of the spatial spectrum, the most common regional arrangement occurs at the bilateral level, and earlier paragraphs point to this scale as becoming increasingly privileged in contemporary international regionalism. At the other end of the spectrum lies the continental scale of integration as has been attempted in Europe or the pan Pacific APEC. Globalization also leads to new geographies in the form of hub and spoke regional arrangements such as the FTAs being proposed for ASEAN and China, ASEAN and Korea, ASEAN and India, EC and South Africa, EC and Chile, etc.

One way to interpret the multiple scales of regional arrangements is to see bloc making as not being solely constituted from interstate trade relations but an expanded recipe of regional civic relations arising from concerns surrounding security, social, environmental, and human rights problems. Trade as a leading indicator of internationalization may have been a handmaiden for spatial testing grounds regarding interstate regulation and supranational forms of governance. However, such spatial experiments have also created more socially constitutive forms of regional institutional building tying local, national, and regional groups across multiple scales of operation. The political economic space of the EU, for instance, is characterized by a bottom up process. Dis cursive political analyses of EU regionalism point to the importance of a myriad of both local and transnationally oriented business and civic groups actively constructing EU regional policies. In other words, internationalization through trade and investment engenders qualitative changes in economic and political processes through transformation of state and territorial capacities, so that such qualitative changes also create the need for forms of governance or coordination beyond that of international trade issues.

Finally while regional trade blocs may have become popular in the organization of international trade, their implementation remains a difficult task. Both France and the Netherlands recently rejected the EU Constitution which sets out EU's values and political objectives, and that paves the way for deeper European integration. Regional projects are increasingly complex reflecting the expansion of agendas beyond the economics; regional arrangements are also becoming multiscalar. Taken together, they are transforming economic and social re lations at the local, regional, and global levels. Human geography provides a lens with which to study the spatial constitution of these relationships.

See also: Export Processing Zones; Foreign Direct Investment; Trade, International; Trade, Transport and Communications, Historical Geographies of.

Further Reading

Abraham, F. and Van Hove, J. (2005). The rise of China: Prospects of regional trade policy. *Review of World Economics* 141, 486 509.

Checkel, J. (2003). 'Going native in Europe'? Theorizing social interactions in European institutions. *Comparative Political Studies* 36, 209 231.

Dicken, P. (1994). The roepke lecture in economic geography: Global local tensions firms and states in the global space economy. *Economic Geography* 70, 101 128.

Dicken, P. and Yeung, H. (2001). Investing in the future: East and Southeast Asian firms in the global economy. In Olds, K. (ed.) *Globalization and the Asia Pacific: Contested territories*, pp 107 128. London: Routledge.

Frankel, J., Stein, E. and Wei, S. J. (1995). Trading blocs and the America: The natural, the unnatural and the supernatural. *Journal of Development Economics* 47, 61 95.

Grant, R. (1993). Trading blocs or trading blows: The macroeconomic geography of United States and Japanese trade policies. *Environment and Planning A* 25, 273 291.

Gunning, J. W. (2002). Trade blocs: Relevant for Africa? *Journal of African Economies* 10, 311 335.

Hayter, R. and Holmes, J. (2000). Continentalism in an era of globalisation: A perspective from Canada's resource periphery. In Barnes, T. & Gertler, M. (eds.) *New Industrial Geography: Regions, Regulations and Institutions*, pp 176 201. London: Routledge.

Hudson, R. (2000). One Europe or many? Reflections on becoming European. *Transactions of the Institute of British Geographers* 25, 409 426.

Jones, E. M. and Smith, M. L. R. (2002). ASEAN's imitation community. *Orbis* 93 109.

Kahler, M. (2000). Legalization as strategy: The Asia Pacific case. *International Organization* 54, 549 565.

Krugman, P. (1991). The move towards free trade zones. *Symposium on Policy Implications of Trade and Currency Zones*, pp. 7 42. Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming.

Mansfield, E. D. and Pevehouse, J. C. (2000). Trade blocs, trade flows and international conflict. *International Organization* 54, 775 808.

Mansfield, E. D. and Synder, J. (1995). Democratization and the danger of war. *International Security* 20, 5 38.

McConnell, J. E. (1986). Geography of international trade. *Progress in Human Geography* 10, 471 483.

O'hUallachain, D. and Wasserman, D. (1999). Vertical integration in a lean supply chain: Brazilian automobile component parts. *Economic Geography* 75, 21 42.

O' Loughlin, J. and Anselin, L. (1996). Geo economic competition and trade bloc formation: United States, German and Japanese exports 1968 1992. *Economic Geography* 72, 131 160.

Oyejide, A., Elbadawi, I. and Collier, P. (eds.) (1997). *Regional integration and trade liberalization in sub Saharan Africa. Volume 1: Framework, issues and methodological perspectives*. Basingstoke: Macmillan.

Poon, J. P. H. (2001). Regionalism in the Asia Pacific: Is geography destiny? *AREA* 33, 252 260.

Poon, J. P. H., Thompson, E. R. and Kelly, P. F. (2000). Myth of the triad? Geography of trade and investment blocs. *Transactions of the Institute of British Geographers* 25, 427 444.

Poon, J. P. H. (1997). The cosmopolitanization of trade regions: Global trends and implications. *Economic Geography* 73, 288 302.

Ravenhill, J. (2003). The new bilateralism in Asia Pacific. *Third World Quarterly* 24, 299 317.

Ruggie, G. (1995). The false premise of realism. *International Security* 20, 62 70.

Yeung, H. (2001). Organizing regional production networks in Southeast Asia: Implications for production fragmentation, trade and rules of origin. *Journal of Economic Geography* 1, 299 321.

Relevant Websites

http://www.wto.org
World Trade Organization.

Trade, International

R. Grant, University of Miami, Coral Gables, FL, USA

Glossary

Comparative Advantage Theory of free trade and perfect markets whereby differences between places should be exploited to the maximum through specialization to achieve economic development.

Global Value Chains (GVCs) Activities that are divided among multiple firms and spread over geographical space; each activity accounts for differing values in terms of trade rewards.

Gravity Model Model that predicts trade flows according to the economic size and distance between two countries or units.

Intrafirm Trade Trade that takes place within a firm (buyer is an affiliate of the firm and not an individual).

Intraindustry Trade Trade that takes place within an industry (buyer is another firm and not an individual).

New Trade Theory Geographical economics theory contending that in a world of imperfect competition, international trade is driven as much by increasing returns and external economies as by comparative advantage.

Off-Shoring Arm's length international trade that can entail international sourcing from unrelated suppliers and/or migration abroad of some of the activities conducted within a multinational company (MNC).

Regional Trading Agreements (RTAs). Free trade agreements among regional member countries.

Trade Barriers Tariff and non-tariff barriers (NTBs) impede the free flow of international trade.

World Trade Organization (WTO) Multilateral organization that superseded the General Agreement on Tariffs and Trade (GATT) in 1995. The WTO regulates international trade and aims to ensure that trade is free and fair.

Introduction: The Theory of Comparative Advantage

International trade is a key driver of spatial extent of the world economy. Significant trade expansion took place from the 1400 s onward after Europe expanded overseas and a modern trade system emerged. The study of patterns of trade and specialization is most widely associated with the formulation of comparative advantage, which itself is based on neoclassical assumptions of free trade and perfect markets. David Ricardo is credited with systemizing the study of international trade with the theory of comparative advantage in his 1817 book *The principles of political economy and taxation*. He formulated his theory using an example involving the exchange of one good for another between England and Portugal. While it was cheaper to produce cloth in Portugal than in England, it was economically beneficial for Lisbon to produce surplus wine and to export it to London and in turn to import English cloth. In Ricardo's time, goods were produced in a single location because fragmentation of production was uneconomic: the coordination of production in remote locations was virtually impossible.

Patterns of trade and specialization are linked in comparative advantage at all geographical scales and connect places together economically. Layers of trade geographies are thus created at a variety of spatial scales: global, world regional, national, and subregional.

Although the theory of comparative advantage has provided a framework for understanding trade patterns and specializations, it has limitations. The low priority given to the geographical components means that in practice it is an aspatial approach, emphasizing static notions of location and distance. Other important concerns relate to core assumptions that capital is immobile and that competition is absolutely perfect, assumptions less meaningful since the 1970s. Almost two centuries after the development of comparative advantage theory, the core of international trade theory continues to be dominated by production and exchange of complete goods. However, revolutionary change in communications and information technologies has enabled a historic (and ongoing) breakup of the production process. Countries such as England and Portugal still produce some goods from start to finish, but increasingly they participate in global supply chains in which the many tasks required to manufacture goods are performed in several different geographical locations. Although there are many practical reasons for trade theorists and governments to continue to adhere to free trade theory, there is also growing opposition to free trade assumptions and the promises of free market trade policies that never materialize in virtually all parts of the world. Antiglobalization movements, the development community, and the tilt to the left in Latin American governments are good examples. It is widely accepted that the current system of global trade bears little resemblance to the neoclassical ideal. Free trade proponents even acknowledge that as much as 40% of global trade is not free.

Gravity Model

The gravity model has been put forward as a corrective to the aspatial bias in comparative advantage and represents a particular geographic view of trade. It shows that the elasticity of trade with respect to distance is in the range of -0.9 to -1.5; this means that a 8000 km distance chokes off over 90% of trade that could take place within a 1000 km distance. The gravity model has been an empirical success in estimating macro patterns in international trade. Variables such as language, tariffs, contiguity, and colonial history have been added to refine the basic model and enhance its predictions. However, many of the model's relationships are not predicted by the theory of comparative advantage. For instance, comparative advantage theory fails to explain fully two salient spatial patterns: (1) the three axes of international trade (North American–EU, Japan–EU, and Japan–North American) that involve countries with similar levels of income trading more with each other and often trading similar goods (i.e., one country sell autos to another country from which it also imports different models of autos); and (2) the spatial clustering of rich and poor countries in international trade.

New Trade Theory

The core free trade theory of international trade treats countries as points with no spatial structure either between them or within them. New trade theory emerged in the 1980s to explain differences between free trade theory and real world patterns. Because there is no widely accepted alternative conception of international trade, many seek to improve comparative advantage theory with better understandings of economic geography. Paul Krugman's work in particular has been influential in assigning key importance to the role of domestic economic geography in determining the trading performance of a state's industries. Furthermore, external economies are more likely to be realized at the local and regional scales than at the international level. Accordingly, international trade can be improved by understanding the processes leading to the local and regional concentration of production. Krugman draws on a range of industrial geographic ideas such as agglomeration economies, location theory, cumulative causation, and regional specialization. Within the discipline of geography, attention has also focused on what changes take place in the internal economic geographies of two countries once trade develops between them. Industrial geographers have welcomed the turn toward geographical economics, but many argue that the legacy of comparative advantage, the treatment of technological externalities, and the empirical bias toward the US and Europe to illustrate trends limit new trade theory. Other critics argue that new trade theory is no more than 'old wine in new bottles'.

Challenges to Understanding Contemporary International Trade

The complexity of the geography of international trade has led researchers to specialize in aspects of the trade system (e.g., country, commodity, and regional studies), and few researchers (some new trade theorists and global value chain (GVC) researchers are exceptions) synthesize or integrate their findings. There are four challenges in accounting for trade patterns that conventional trade theory cannot explain.

The first set of challenges involves explaining the sectoral composition of international trade and the shifts in types of trade that predominate in the international system. Different trade concentrations and individual countries' particularities require different explanations. For instance, trade in primary commodities, such as petroleum, responds to very different conditions than trade in manufacturing goods or trade in services. Furthermore, the spatial organization of trade is no longer solely on a horizontal country–country trade of finished commodities (whereby countries fully produce and trade final goods). Instead, vertical specialization (whereby countries specialize and trade in one stage of the production process) is spatially reorganizing trade, accounting for 20–25% of total trade.

Accounting for the shifting terrain of barriers is the second set of challenges for trade theory. At the establishment of the General Agreement on Tariffs and Trade (GATT) in 1947 free trade primarily referred to exports and imports of goods and restrictions to trade were mainly in the form of tariff barriers at the borders. This conception of trade barriers no longer suffices. Trade policies have grown to include NTBs, which are much more difficult to measure and quantify.

The third set of challenges involves incorporating evidence from industrial geographic research into paradigms about the shifting contours of international trade. Incorporating the specifics about the fragmentation of production, as well as about the particular niches within various value chains of production better captures the dynamic nature of international trade.

Fully accounting for the tilt toward regionalism in international trade constitutes the fourth set of challenges. Various positions are taken on impacts of widespread trade bloc formations, ranging from conceptualizing them as building blocs to seeing them as stumbling blocs in the organization of international trade. Some even suggest that regionalism has emerged as an alternative form of trade to counter the more aggressive stance of free trade espoused by the World Trade Organization (WTO). The WTO is supposed to vet blocs for conformity with multilateral rules, but regional groupings consider it is in their best interest to keep bloc criteria and procedures vague. Re gardless of the position taken, 42% of trade now takes place within regional trade agreements (RTAs).

Changing Sectoral Composition and International Trade

The volume, value, and structure of trade have under gone fundamental changes over time, particularly in the past two decades. Economic value and the dominant types of trade accounting for most of the value added in cross border trade have proceeded through distinctive eras, with technology as the driving force. In the first phase of trade expansion (1450 onward), primary com modities (i.e., raw materials, minerals, and food products) dominated, and economies were organized spatially for local needs. In the second phase (1840–present), manufacturing goods in standardized ways accounted for the largest share of international trade and became the most innovative sector of trade. Manufacturing trade was organized regionally and internationally. For instance, colonial policies ensured that territories did not compete with colonial home countries but instead became sites for importing manufacturing commodities. In the third phase (1970–present), services became global commod ities, accounting for a growing share of cross border trade, and these became differentiated from trade in manufacturing goods because they emphasize trade in customized commodities. The globalization of markets has meant that there are fewer geographical limits to the spatial organization of international trade. As far as the leading sector in the next trade era is concerned, some analysts envision a new energy source substituting petroleum and driving trade shares for a short but ac celerated time period. Others envision the next com petitive battleground beyond the services trade will be in staging and trading experiences (e.g., virtual travel), predicting that commodity trade will be further up graded and that the more lucrative rewards from trade will involve the selling of lifelike interactive or virtual reality experiences globally.

In 1980 the largest components of international trade were primary products and medium technology manu factures (**Figure 1**) By 2000, primary products accounted for the smallest component of international trade, being superseded dramatically by exports of high technology products. In high technology industries, knowledge is a prime source of competitive advantage, there is above average spending on R&D and employment is dominated by scientists and engineers. Good examples include the aircraft, electronics, and pharmaceutical industries. The

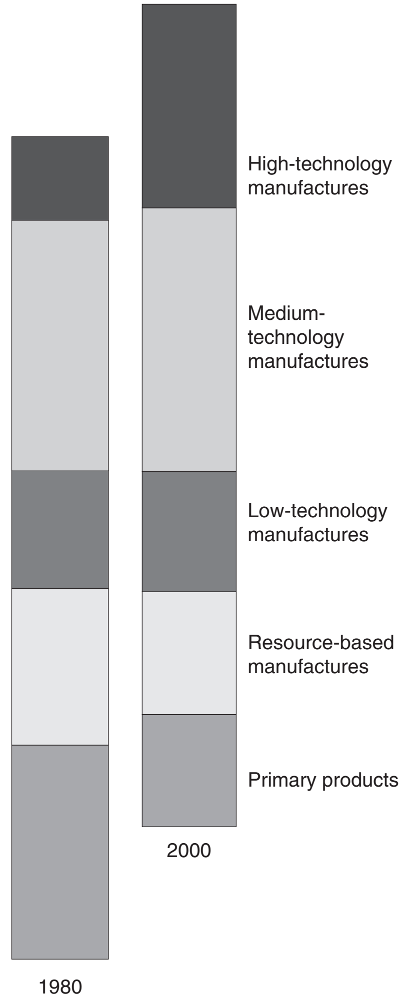

Figure 1 Sectoral composition of international trade, 1980 and 2000. From WTO (2005) International Trade Statistics. http://www.wto.org/english/res e/statis e/statis e.htm.

share of high technology goods in the world trade rose from 10% in 1980 to 25% in 2000.

Concentration in high technology trade is also highly correlated with export growth shares. Thus, the highest export growth rates are recorded in the high technology category, and the lowest overall growth rates are in primary products. In Latin America, for example, primary product exports grew 5% between 1980 and 2000 while high technology manufactures expanded by 21%. There are important macroeconomic geography implications to this trend. The most dynamic traders with the highest export growth rates occurred in countries that concentrated in medium and high tech manu facturing. A select group of traders (China, Mexico, South Korea, Singapore, and Taiwan) registered export growth rates of more than 10% between 1980 and 2000, which were more than double those of the US and EU. The China factor has become significant in trade: China is the fourth largest international trader (counting the EU as a single unit), and its exports in 2005 exceeded the combined exports of Africa, the Middle East, and Latin America.

The recent globalization of production also has altered the types of commodities that are traded and the way the trade is being conducted. Firms influence the economic geography of trade through both intra industry and intracorporate trade. Combined, these types of trades account for 30–50% of Organisation for Economic Co operation and Development (OECD) trade. They exhibit complex interconnections, often operate under con ditions of imperfect competition, and shape new trade geographies. First, a distinguishing feature of intra in dustry trade is that the buyers are other firms rather than individuals; thus, items meet firms' quality, durability, and serviceability needs rather than individual consumer tastes and/or fashion. Second, an increasing proportion of international trade is intracorporate trade (with multi national corporations (MNCs)) rather than interstate. Intracorporate trade is less controlled by 'arms' length' price/cost determinants of regular trade; transactions between affiliates are least likely to be market deter mined; and they are sensitive to international decisions of MNCs. This type of trade also responds differently to changes in economic conditions compared to trade be tween unrelated parties.

Because of the changing role of firms in international trade, a significant portion of trade involves component parts rather than finished goods. Vertical specialization is becoming more common and takes place when countries acquire expertise in particular stages of the production process: a country may import a good from another country to use it for the production of its own good and then export that good to the next country. The sequence only ends when the final product reaches its destination. The geographical value of the production of an 'Ameri can' car, for example, is apportioned as follows: 37% of the value of production is generated in the US, 30% goes to Korea for assembly, 17.5% to Japan for components and advanced technology, 7.5% to Germany for design, 4% to Taiwan and Singapore for minor parts, 2.5% to the UK for advertising and marketing services, and 1.5% to Ireland and Barbados for data processing.

Changing Policy Terrains and New Barrier Architecture

Trade liberalization has been consistently promoted under the GATT/WTO framework. Average industrial tariff rates in 1947 were 40%, and they were 3.8% in 2003. Reductions of tariffs have fostered growth in all types of trade, especially vertical trade because it in volves multiple border crossings. Beyond industry, mar ket access is very much an unfinished business: only 55% of imports enter markets duty free. One basic computa tional problem in quantifying barriers and comparing across countries is the variance in the ways that countries utilize tariff lines (a single item as defined in a country's tariff schedule): most countries apply 5000–8000 tariff lines but Turkey, Malaysia, Brazil, and Mexico use more than 10 000 lines. Four sectors – textiles and clothing; leather, rubber, footwear, and travel goods; transport equipment; fish and fish products – receive special treatment and account for the highest tariff rates.

The proliferation of NTBs since the 1970s represents a counterliberalizing trend in most countries and for most sectors; in 1986, 50% of all products traded were subject to NTBs in some form. NTBs are discretionary barriers applied often to specific commodities but are not limited to a form of technical measure, custom rules and procedures, competition related restrictions to market access, or quantitative import restrictions. Compared to tariffs, for which there are quantitative databases that make it possible to measure, analyze, and monitor, NTBs are more difficult to identify because of national no menclatures and varying degrees of transparency. NTBs display significant geographical variation by country and by commodity. For example, one third of countries pro hibit certain types of trade in used products: China and Latin American countries have banned imports of used car parts for safety reasons, whereas Argentina has enacted a more comprehensive ban on imports of used tires, car parts, clothing, and medical equipment. Given the fungibility of NTBs, more efforts are required to investigate how health, safety, and environmental regu lations and standards affect international trade.

Shifting Geographies of Production

Besides diverting trade via barriers, governments have become more active in creating trade over and beyond what is typically considered in comparative advantage. Governments influence the geography of production's organization by joining RTAs and by emphasizing strategic trade policies. Governmental intervention (and ultimate influence on comparative advantage) ranges from specific export promotion and import prohibition to broadly drawn infrastructural policies, R&D, targeting of particular industries for special attention (e.g., the aircraft industries in Europe) and fostering clusters in technology and industrial specialized districts. Government nonac tion can also be used to support particular sectors and specialized districts, for example, tolerating clandestine workforces. Governments and the private sector in the developing world are attempting to forge partnerships and promote niche export processing zones that spe cialize in one or two sectors, such as information tech nology and global offshoring. Over the last decade, offshoring activities have increased at a rapid rate as the cost of international transactions continues to decline as entrepreneurs in low wage countries learn how to produce to global standards and the tasks formerly undertaken in one country are now being performed abroad. The variety of services offshored has grown in complexity over time and now involves customer service calls, X ray readings, tax preparation, and heart surgery.

GVC research is a newer approach that studies the spatial organizational aspects of international trade and within a context that goes beyond trade alone; it includes the entire range of activities from primary production to final consumption to linkages binding them. The activities that constitute a value chain can be contained within a single firm or more typically divided among different firms. Value chains can be contained within a single geographical location, but more commonly they are spread internationally. GVC research examines how lead firms build, coordinate, and control the linkages and flows of products between raw material suppliers, processors, primary traders, wholesalers, retailers, and consumers. The advantage of GVC is that it unpacks trade by identifying clusters and showing who is benefiting the most and where, particularly with upstream activities that reap most of the rewards from international trade.

The global landscape of manufacturing trade is marked by enormous diversity from place to place and across industries. Nonetheless many sectors of this trade are organized in important ways around the great agglomerations that constitute the main spatial anchors. Agglomeration and trade are mutually reinforcing phenomena. The classical example of a leading industrial district is the clothing industry in northern Italy, where there are dense concentrations of independent small and medium sized companies connected to auxiliary industries and knowledge services (e.g., fashion related publishing and design centered on Milan). Some centers of production (e.g., clothing, footwear, and furniture) exhibit a tendency toward more and more spatial concentration of production. International trade among centers of clothing, footwear, and furniture shows three broad types of flows: (1) direct exports of high quality final products among advanced economies; (2) intrafirm trade from low wage plants in the developing world to North America,

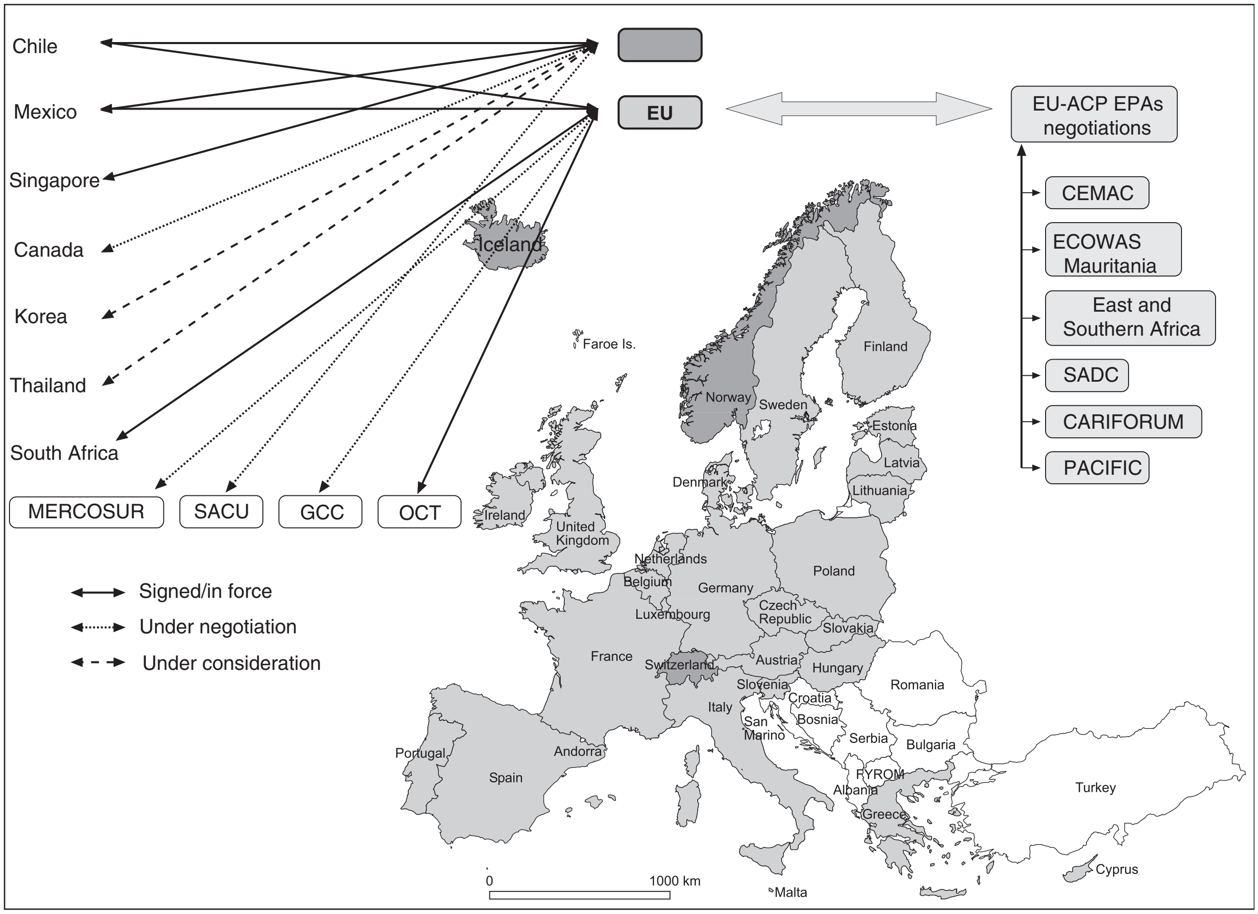

Figure 3 European RTA network. Adapted from Crawford, J. A. and Fiorentino, R. (2005). The changing landscape of regional trade agreements. WTO Discussion Paper no. 8. http://www.wto.org/english/res e/booksp e/discussion papers8 e.pdf (accessed Sep. 2007) (2005:9).

Western Europe, and Japan; and (3) outsourcing rela tionships from high wage countries to independent firms in low wage countries. These three features combine to create an international mosaic of trade flows and pro duction sharing arrangements. However, the spatial or ganization of industrial production is always in flux. For instance, the termination of the Multifibre Arrangement on 1 January 2005 means that the global apparel industry is undergoing a transition from one regime to another; subsequently China and India have been emerging as major low cost producers.

Regional Trade Agreements

RTAs are a major part of and perhaps an irreversible trend in international trade. Both the number of RTAs, as well as the amount of trade conducted within regional trade associations have steadily increased. Forty two percent of international trade is affected by preferential agreements. Some 312 RTAs have been notified to the GATT/WTO between 1947 and 2005: in force currently are 170 RTAs; many of the original agreements were modified and superseded by other agreements over time. The trend will intensify as countries across the world are making RTAs the centerpiece of commercial policies: 20 RTAs are pending ratification and another 70 RTAs are in a proposal/negotiations stage of formation. RTAs can be viewed as the clash of liberalizations: regional liber alization as opposed to global liberalization.

RTA activities have intensified across all geographical regions of the world, and especially in the Western Hemisphere and Asia Pacific Region. **Figure 2** shows the establishment of the main international regional trading blocs. Some RTAs are much more important than others: at the country level the proportion of international trade covered by an RTA is more important than the total number of RTAs in which a country participates. There is considerable range in the scope of activities covered by RTAs, from extensive to more narrow coverage, but the trend is toward increasing complexity. There are also RTAs (e.g., EU–South Africa and US–MERCOSUR) that cover the so called 'Singapore issues' (investment, gov ernment procurement, and competition), a set of issues

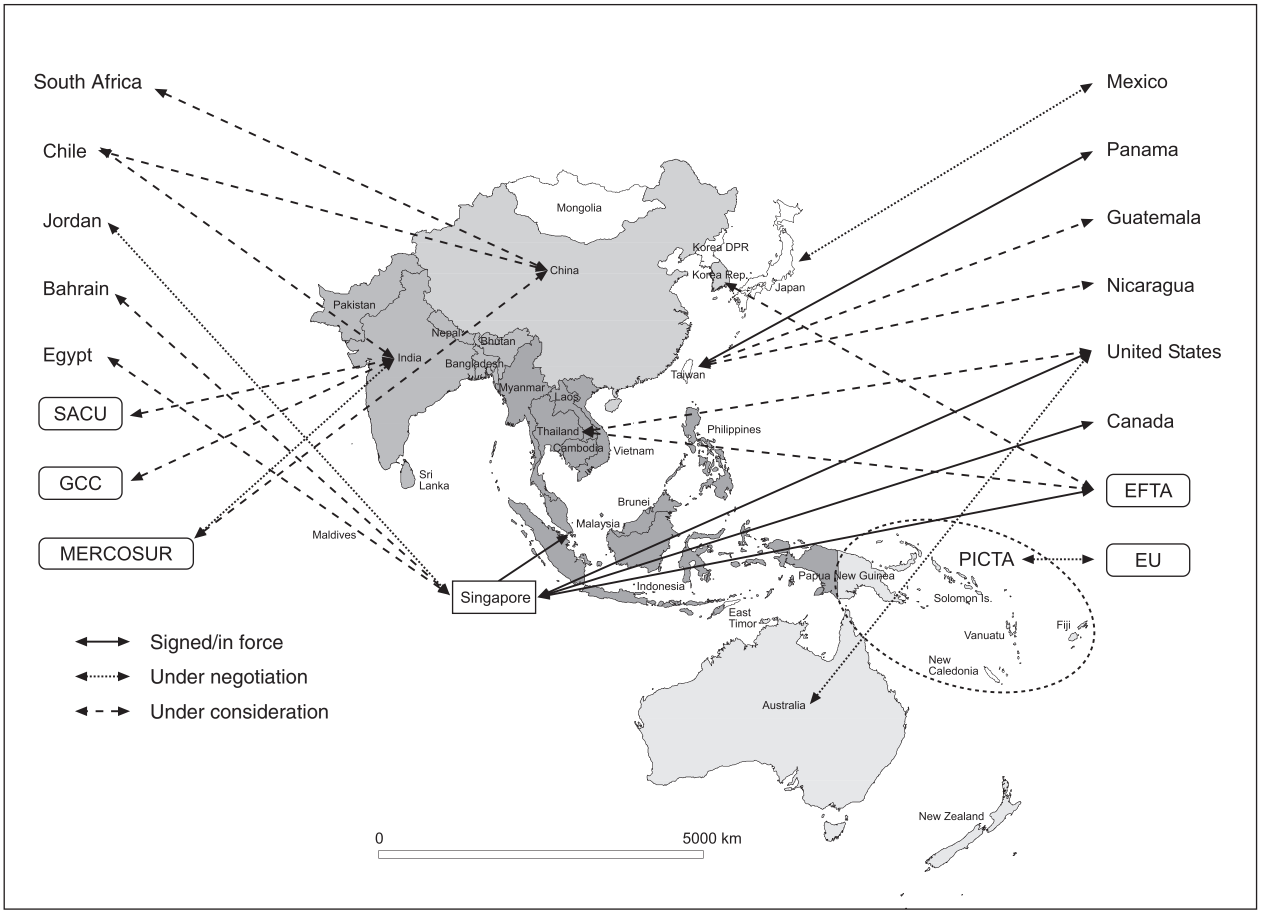

Figure 4 Asia-Pacific RTA network. Adapted from Crawford, J. A. and Fiorentino, R. (2005). The changing landscape of regional trade agreements. WTO Discussion Paper no. 8. http://www.wto.org/english/res e/booksp e/discussion papers8 e.pdf (accessed Sep. 2007) (2005:12).

that did not secure final WTO approval. Traditionally, RTA formation has occurred between proximate trading partners, most often contiguous countries with well established trading patterns. The EU and the North American Free Trade Agreement (NAFTA) countries provide good examples. Indeed most countries sign their first trade agreement with one or several neighboring countries (**Figure 2**).

Most regional blocs have a strongly defensive char acter: they represent an attempt to gain size advantage by creating larger markets for their products and by pro tecting them, at least in part, from outside competition. Consequently, the largest trade blocs have considerable influence on international trade. Regional trade blocs affect the geography of trade by diverting trade (most countries within a bloc import goods from each other rather than from nonmember countries, as done previ ously) and by increasing the level of intraregional trade. These blocs boost intraregional trade at least by half (e.g., the EU) and at maximum by fivefold (e.g., Association of Southeast Asian Nations (ASEAN)). In the process they reduce trade with the rest of the world.

Complex geographies of RTAs are emerging with the spatial configuration of trade agreements becoming more complicated with overlapping RTAs and networks, as well as a pending wave of RTAs that will span across world regions (see **Figures 3–5**).

RTAs are increasingly being concluded among non contiguous countries. New spatial patterning is linking various countries around the globe in cross regional preferential agreements. A new development, anticipated by 2011, is the emergence of a new spatial category of agreement, namely supra RTAs: each of the parties is a distinct RTA itself and consolidation takes place among larger spatial units. Examples include the EU–GCC (Gulf Cooperation Council) in Europe (**Figure 3**); India–PICTA (Pacific Island Countries Trade Agreement) in Asia (**Figure 4**); and CARICOM (Caribbean Community and Common Market)–EU in the Western Hemisphere (**Figure 5**).

Because members of each RTA trade more with each other than they do with nonmember countries, trade is becoming more geographically concentrated. RTA pro liferation is undermining transparency and predictability

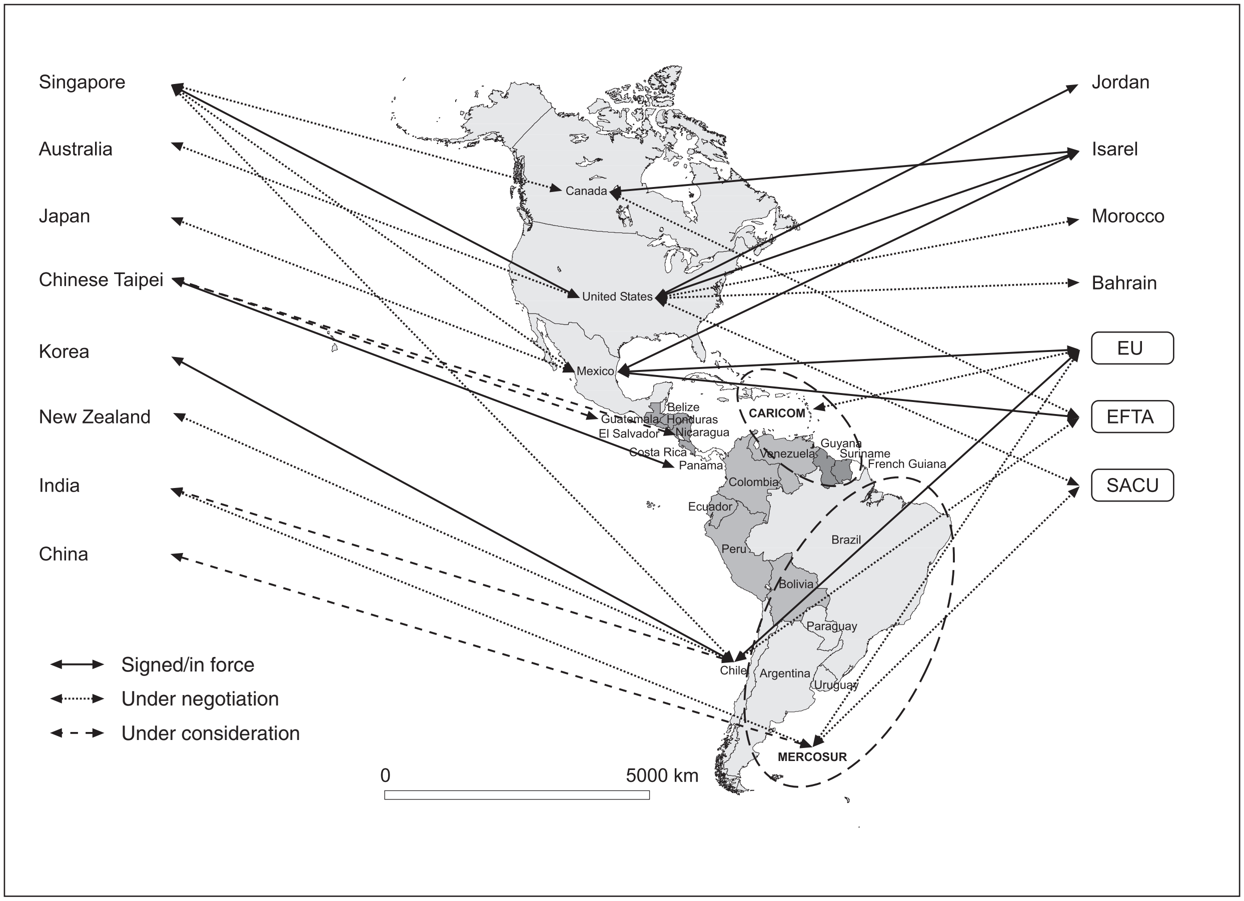

Figure 5 Western Hemisphere RTA network. Adapted from Crawford, J. A. and Fiorentino, R. (2005). The changing landscape of regional trade agreements. WTO Discussion Paper no. 8. http://www.wto.org/english/res e/booksp e/discussion papers8 e.pdf (accessed Sep. 2007) (2005:11).

in international trade. RTAs appear to have an adverse effect on the multilateral trade system. The inability of the WTO to solve current trade issues, evident in the limited outcome of the Doha Round, seems to be en couraging countries to look for regional solutions. This new trend of cross regional RTAs will create another spatial layering of international trade.

Issues

Trade is not wine for cloth anymore. The theory of comparative advantage is still the most widely employed framework to explain and understand trade, despite its limitations and critics. New trade theory, gravity models, GVCs, and RTAs offer alternative frameworks to understand how trade works in the twenty first century. Although the insights offered by each of these approaches are valuable, none has thus far been capable of organizing findings into a comprehensive international trade theory with significant explanatory power. These alternative frameworks do shift the ground in the social sciences to some degree toward collecting trade data based on units other than countries (e.g., firms, networks, regions) and emphasize local and regional variations (as opposed to national) and niches within value chains, in contrast to the earlier aspatial neoclassical approach.

See also: Foreign Direct Investment; Global Production Networks; Globalization and Transnational Corporations; Trade Blocs.

Further Reading

Begg, B., Pickles, J. and Smith, A. (2003). Cutting it: European integration, trade regimes, and the reconfiguration of East Central European apparel production. *Environment and Planning A* 35, 2191 2207.

Crawford, J. A. and Fiorentino, R. (2005). The changing landscape of regional trade agreements. WTO Discussion Paper no. 8. http://www.wto.org/english/res e/booksp e/discussion papers8 e.pdf (accessed Sep. 2007).

Gibb, R. and Michalak, W. (eds.) (1994). *Continental trading blocs: The growth of regionalism in the world economy.* New York: Wiley.

Hanink, D. M. and Cromley, R. G. (2005). Geographic change with trade based on comparative advantage. *Annals of the Association of American Geographers* 95, 511 524.

Grant, R. (2000). The economic geography of global trade. In Sheppard, E. & Barnes, T. (eds.) *A companion to economic geography*, pp 411 431. Malden: Blackwell.
Krugman, P. (1992). *Geography and trade*. Cambridge: MIT Press.
Martin, R. and Sunley, P. (1996). Paul Krugman's geographical economics and its implications for regional development theory: A critical assessment. *Economic Geography* 72, 259 292.
Organization for Economic Cooperation and Development (OECD) (2005). *Looking beyond tariffs: The role of non tariff barriers in world trade*. Paris: OECD.
Ricardo, D. (1817). *The principles of political economy and taxation* (Reprinted 1996. New York: Prometheus Books.)
Scott, A. J. (2006). The changing global geography of low technology, labor intensive industry: Clothing, footwear, and furniture. *World Development* 34, 1517 1536.
WTO (2005). International Trade Statistics.

Relevant Websites

http://www.aseansec.org
AFTA and FTAs, Association of Southeast Asian Nations.
http://europa.eu
External Trade, Europa The European Union On Line.
http://www.globalvaluechains.org
Global Value Chain Initiative Duke University.
http://www.ustr.gov
North American Free Trade Agreement, Office of the United States Trade Representative.
http://www.wto.org
World Trade Organization (WTO).

Trade, Transport and Communications, Historical Geographies of

P. J. Hugill, Texas A & M University, College Station, TX, USA

Glossary

Geopolitics The attempt by states to control earth space for the resources it contains or the strategic military advantage it confers over other states.

The historical geographies of trade, transportation, and communication are intimately interlinked, both with each other and in relation to geopolitics and to the historical geographies of warfare, production, and urban development. There are two links to geopolitics, one in the traditional sense, one with regard to the literature on the difference between trading and territorial states and polities. As late as the 1890s authors such as Mahan argued that sea power was the key to history. Sea power had rebalanced global geopolitics in favor of the seaborne trading polities of Western Europe from AD 1431 onward which gives much of this account an unavoidably Eurocentric flavor. However, in 1904 Mackinder pronounced the 'Columbian epoch' over and saw the world returning to a land based geopolitical system, now based on the ability of railroads and telegraphs to integrate continental sized territorial polities, he believed able militarily to dominate first the world island (Eurasia and Africa), then the world. In 1919 Mackinder identified two polities within the German state, one seeking territory to the east, the other trading opportunities to the west. The internal competition between the two polities was acute and, eventually, tore Germany apart, with the result that it lost its hegemonic struggle with Britain, a highly focused trading state. However, in 1919, because the naval power of the trading states could not penetrate far into the continental interiors, the future seemed likely to lie with the territorial and against the trading states. By the early 2000s, however, it has become clear that extremely accurate air power capable of penetrating well into the continental interiors has become the cynosure of military strength, returning the global balance of power to the trading polities who are more interested in controlling flows and nodes than territory.

Before the European seaborne expansions Eurasia was often dominated by nomadic cultures pulsing out of central Eurasia. Only two areas resisted: a Chinese polity that signified successful invaders, defeating them culturally; and European polities that, from about AD 800 onward developed superior military technologies. The link to the geographies of warfare and production is through the military technologies that allow a given polity or set of polities to dominate certain Earth spaces and, on occasion, that create the possibility of a new geopolitics. The link to the geography of production is that military technologies have become increasingly expensive over time, thus requiring more and more technologically sophisticated production facilities.

Early Webs of Interaction

It makes sense to think of the historical geographies of trade and transport as land, sea, and air based systems and of communications as either systems based on the physical movement of information or systems based on electronic movement. Neolithic agriculture developed in a few favored riverine areas of Eurasia (Tigris–Euphrates, Nile, Indus, and Huang Ho). Agriculture allowed population increase, cities, and the formation of simple states. Communications became vital because of the need to manage increasingly complex economies, polities, bureaucracies, social systems, religions, and military structures. As the McNeills note, this was the first major human web of interaction, a concept applicable through the present, human webs constantly expanding spatially as transport improved and thickening through better communications. The formation of simple states in these earliest human civilizations was greatly facilitated by river transport, and state management by the ability to trade agricultural surplus up or downstream to regions in need. The critical symbiosis between land and water based transport systems was born. However, as states expanded geographically they encountered different peoples and more complex polities were born to manage different human cultures, usually within a single imperial system. Also, from 1700 BC onward, formal military structures arose in organized response to attacks on the productive riverine economies of settled agriculturalists by predatory charioteers.

The Rise and Decline of Islam

The development of coastwise navigation in the Bronze Age extended the range of this early transport, trade, and communications web to multiple human cultures across the entire Mediterranean Basin, an almost entirely geographically closed sea. The Roman Empire came to epitomize the most successful phase of this, using sea based communication and transport to bring under its long term political control the entire Mediterranean Basin and even on into the North Sea to Britain and the Rhine Valley.

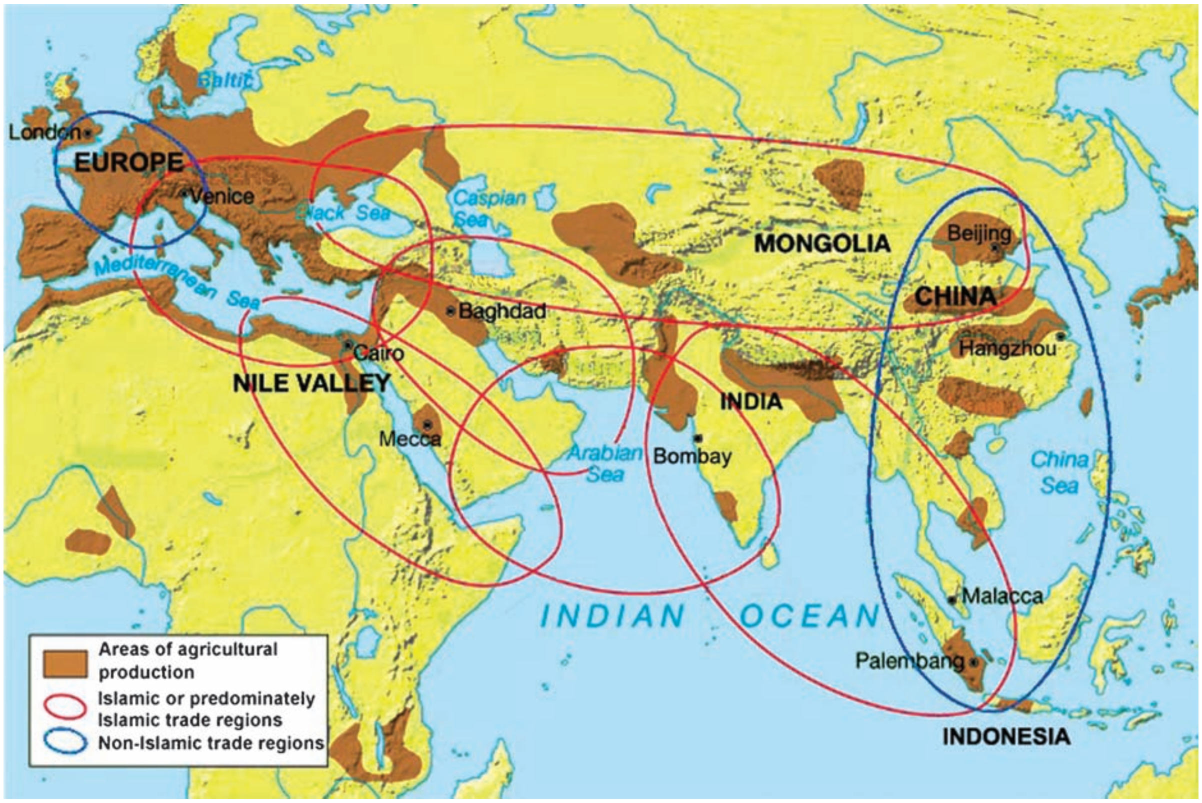

Figure 1 Trade regions before the Columbian expansions (by the author, after Abu-Lughod).

Polities east of the Mediterranean enjoyed important bouts of expansion, but lacked effective sea based trans port systems to hold them together. Three forces brought change to the region. First, around the time of Christ traders discovered that highly reliable monsoon winds blew ships east to India at one time of year and back at another. Second, traders made increasing use of camels as pack animals to push trade inland, at least for pre ciosities, in regions of oasis agriculture. Third, the region became united ideologically by the rise of an aggressive and highly successful traders' religion, Islam. Islamic traders and raiders then pushed out of the Middle East converting as they went, contesting Christian control of the Mediterranean sea trade to the east and Gujarati and even Chinese strength in the Indian Ocean to the west. By the mid 1400 Islam stretched from the Atlantic to the Pacific in a broad band across North Africa, through the Middle East and India, on through the Indonesian archipelago and into the Pacific at the Philippines. Only China and Europe held out against this relentless ex pansion (**Figure 1**).

One fatal flaw for Islam was, however, its failure to develop a productive local farm economy of the sort developing in China and Europe. By AD 610 China was using the Grand Canal to unite its northern wheat region with its southern rice region and integrate a single, large, territorial polity. A complex network of local canals completed a job not begun in Europe for another 1000 years. Europe was a 'peninsula of peninsulas', and Roman style coastwise navigation improved incre mentally but steadily after Roman withdrawal. From about AD 1200 on, and in the absence of any single central political authority such as existed in China, self defending trading ships of the Hanseatic League moved goods between the relatively independent feudal cities of the North Sea, the Baltic, and the Atlantic coast, con tributing mightily to the explosion of European blue water navigation after AD 1431. The hardening of formal states in the region post 1400, in particular an increas ingly powerful France and England, reduced the power of the cities embedded in those states, encouraged mer chants to demand formal, state funded navies to protect trade, and ended the power of the Hanse around AD 1441.

The European Trading Polity

Thus was the scene set for the European expansions. New techniques of blue water navigation, developed in Portugal and Spain but improved on in Holland, Eng land, and France, and new ships designed to cope year round with the harsh seas of the Atlantic and to fight their way out of any trouble they could get into, allowed Europeans to begin a seaborne expansion around the world island in AD 1431. Their goal was to reach China and the Spice Islands (Indonesia) direct, thus reducing the substantial transfer costs (doubling at each boundary) imposed upon long distance trade across the world is land by the eight circuits of trade described by Abu Lughod. In AD 1486 the Portuguese navigator Bartolo meo Diaz circumnavigated Africa and entered the Indian Ocean, proving a sea route east to the Indies possible. The Spanish, seeking an alternate route around the world island to the west, encountered the Americas. Two 'new worlds' came into existence: one, the newly en countered lands of the Americas; the other, a world made whole, as Agnew has shown, at least in the European imagination, by the expansion.

Previously unstoppable, Islam collapsed quickly for three reasons and Europe rose for a fourth. First, much of Islam's wealth came from transfer costs imposed at trade circuit boundaries, six of which were wholly or almost wholly controlled by Islam. Second, Islamic local agri culture was nowhere near so productive as that of Western Europe or China, and could not support either large manufacturing cities or social spending on ships and navies. Third, Islamic societies failed to adopt European shipping technologies, staying with oared warships and naval warfare based on boarding when Europeans had moved on, creating fast, maneuverable, cannon armed ships that could both kill at a distance and carry a profitable cargo. Fourth, Europeans sought to control flows of trade and the crucial nodes or choke points through which such trade had to pass rather than territory because the chief concern of their maritime trading polities was to generate taxes through trade, not through control of territory.

For nearly 400 years after the Portuguese pioneered blue water expansion in AD 1431 the Europeans, con stantly competing to develop better seaborne technolo gies and better ways of killing each other, applied those technologies to dominate a world they saw united rather than divided by the oceans. In the early 1800s, however, two new transport technologies were developed, steam boats and railroads. These allowed movement of bulk goods through the continental interiors and thus the promise of powerful territorial polities, the first of which was America.

Inland Transport and the Territorial State

In 1815 Shreve developed steam boats powerful enough to push up the fast flowing rivers that drained the American South into the Atlantic and the Gulf of Mex ico, opening vast new lands. With a viable crop and enough slaves, 'King Cotton' created the world's first genuinely global production system. The South displaced all previous competitors, shipping cotton to mills around Manchester, England, to be spun and woven into cheap textiles, then marketed globally. Manchester developed as a 'new' style of city, one based on massive technical innovation and a totally new industry. However, as Hall notes, such cities have a brief 'life'. Technologies diffuse, competitors emerge, and overconcentration ensures that when the industry declines, the city declines too.

The second inland transport technology, developed in England, but with its most serious impact in America, was the railroad. Canals needed an absence of rugged topography and much water and capital. By the late 1840s powerful 4 4 0 'American' class steam locomotives could climb substantial grades: a sophisticated suspension allowed cheaply laid track. By the 1850s the Appalachian barrier to movement inland from the Atlantic coast had been conquered, and by the 1860s railroads united the Atlantic and Pacific. Alongside the railroads marched the electric telegraph. Telegraphs allowed single track working over long distances, keeping capital inputs to the new railroads low, and informed farmers and brokers of prices in different markets. In America cities such as Chicago created inland empires based on rail nets thrust aggressively into their hinterlands, though, as Cronon has shown, Chicago's brilliant innovation of the futures market for trade in commodities ensured its long term rise to primacy within America's heartland.

The First Global Web and Its Problems

Coincident with this creation of a new territorial state and economy in America's continental interior was a renewed explosion of European seaborne expansion, also driven by steam and electrical telegraphy. The first steamships crossed the Atlantic in the late 1840s but were at the limits of their range, expensive, and attracted only well heeled passengers. By the 1870s, however, the triple expansion engined steamship had been developed. This was much cheaper to run, hauled cargo as profitably as passengers, and made Glasgow, where such ships ori ginated, the world's next great city based on technolo gical innovation. Alongside the global network of submarine telegraph cables installed by British capital after the success of the transatlantic cable in 1866, such steamships reinforced Britain's centrality to the increas ingly global economy of the late Victorian era and made London, by 1900 the world's first global 'capital', the first 'information economy' city. Certainly, by 1900 the first truly global 'web' had formed, its trade moving almost entirely on British steamships fueled by British coal from a network of British coaling stations around the globe and

Figure 2 Global telecommunications, 1938 (by permissions, Cable & Wireless, London).

managed from London by British capital directed by British submarine cables connected to land based cable systems (**Figure 2**). Some land based systems were continental in scale and controlled by increasingly powerful cities, New York, in the American case.

Although this system was global it was thin. Goods and people moved at the speed of steam or slower, and speeding things up was expensive. Railroads pushed un evenly into continental interiors, needing increasing amounts of capital to pay for heavier, more expensive track. Demands for higher speed and greater capacity brought forth larger, more powerful locomotives and ships, all of which cost more. Although telegraphy moved information at the speed of light, capacity was limited and range was subglobal, with repeater stations needed at some 2000 mile intervals with human operators rekeying messages onward. Telegraphy required a compact written language, communication could easily be garbled, espe cially when rekeyed by fallible humans, and the important levels of meaning conveyed by spoken and visual communication were absent.

Between the late 1800s and about 1920 all this shifted. Long distance telephony became possible, first using line loading, then electronic amplification of signals. Wireless allowed telegraphy and telephony to begin to break free of the tyranny of the fixed routes of the cable systems. More flexible, lower capital cost public transportation systems were developed to carry smaller numbers of people in the form of the electric streetcar and gas electric interurban. Bicycles, automobiles, and trucks also offered a break from the fixed and expensive route tyr anny of the railroads. However, they needed hard surfaced roads that turned out to be as expensive as railroad tracks: the difference was that roads could be funded out of a tax on fuel, thus opening a huge new source of capital for infrastructural development. Elec trification, then dieselization, of railroads offered the possibility of higher speeds on fixed routes on land without requiring more expensive trackage, and dieselization of shipping offered more reliable, more easily managed, more fuel efficient ships than those using triple compound engines. Finally, it was obviously going to be possible to take to the air, which offered both the potential for higher speed and freedom from the planet's last crucial piece of geographic tyranny, the need for transfer points at land–sea interfaces, with their resultant costs.

New Technologies and Hegemonic Shifts

Almost all these shifts between the late 1800s and 1920 either developed because of or were intensified by the wars of the period, and the wars of the period focused on the tectonic shifts in hegemony as transportation and communications technology showed their potential for very powerful territorial states and the power of the maritime trading states slipped. America, Imperial Germany, Czarist Russia, and Imperial Japan began to rise. America, Germany, and Russia seemed like territorial states of the sort Mackinder predicted in 1904. Russia's defeat by Japan in the Russo Japanese War of 1904–05 only seemed to reinforce the territorial nature of the awakening Russian giant. Japan developed initially as a maritime trading polity, actively emulating Britain, to which it was formally allied between 1902 and the Washington Naval Treaty of 1922, when America *de facto* forced Britain to abandon the alliance. Japan's territorial polity would not manifest itself until its invasion of China in the 1930s. America and Germany were mixed states, containing territorial and trading polities. In Germany, the Prussian Junkers seeking to push east for more territory vied for control over the state apparatus with the Hamburg traders seeking to push out into the Atlantic. America was, at one level, more complex: from the Civil War on the protectionist and isolationist wing of the Republican Party vied for control of the Party with East Coast traders best represented through Teddy Roosevelt's imperialist, progressive impulses. At another level, America was much simpler: no neighboring power opposed its dominance of North America. In the late 1930s the American trading polity became resurgent, a position it retains to the present. The geopolitical problem was not, in any case, in America, but in Eurasia, where German and Russian polities vied for control of territory Mackinder saw as the key to world domination.

Here, technological improvements in transport and communications that seemed beneficial to liberal Victorians took a lethal turn. Steam turbines were first used in transatlantic passenger liners, but by 1914 packs of turbine engined dreadnought battleships roamed the seas armed with guns up to 15" caliber capable of delivering lethal punches to dreadnoughts 25 miles away. Marconi's first sale of wireless was to Lloyd's of London, to ease the rescue of ships in distress at sea, such as *Titanic* in 1912. Two years later wireless proved its massive worth in causing distress to ships at sea, allowing fewer dreadnoughts to patrol more space and be quickly repositioned to address new threats, such as that posed by von Spee's squadron, sunk by British battle cruisers in the first battle of the Falklands in 1914. Railroads allowed fast movement over fixed lines of communication, but at the Western front in 1914 they turned lethal for millions of men. The heavy railroads that delivered millions of tons of artillery shells from the factories were vulnerable to opposing artillery, so had to be many miles from the front. Close to the front, shells had to be delivered on human powered tramways, which froze the front into a long line some 20 miles away from but parallel to the main line railroads. Coupled with innovations such as barbed wire and machine guns, war between industrial powers froze in space and became a matter mostly of defense. Generals trained for wars of mobility that favored the offence ensured appalling slaughter when they persisted in their offensives. By 1918 the automobile, its cousin the tractor, the tracked tractors that had been developed for soft soil conditions in California agriculture, and the military descendant of such tractors, the tank, offered a return to a war of mobility.

Aircraft offered a new challenge. Wells argued in 1908 that air war would plunge the world into a new Dark Age. Although this failed to happen, in World War I airplanes developed from frail devices into powerful, heavily armed military machines designed to carry thousands of pounds of bombs hundreds of miles, "bloody paralysers" in the words of one British general. By late 1918 the British had created the doctrine that Germany would use to perfection against Poland in 1939 and call *Blitzkrieg*, a rapid advance by combined tank and air arms coordinated by wireless telecommunications giving the advantage back to the offensive.

At the end of World War I the military forcing of new transport and communications technologies paid off in civilian life. Trucks and caterpillar tractors built for the war but never shipped to France were distributed free to every county in America to aid in the explosive growth of hard surfaced roads after 1920, on which would run ever larger numbers of trucks and automobiles. War surplus bombing aircraft everywhere were the foundation of primitive air services: such services quickly focused on mail, which could survive fiery crashes in fireproof compartments. The demands of proto *Blitzkrieg* had forced rapid development of wireless telephony in 1918, the training of huge numbers of wireless technicians, and mass production of vacuum tubes sold cheaply as war surplus, in short, all the necessary components for the immediate, rapid growth of radio entertainment after

World War I, and the foundation for television and radar technologies in the 1930s.

America entered the automobile age from about 1915 onward and Detroit became the next city to rise on the basis of manufacturing an exciting new technology. Streetcars, automobiles, trucks, and buses made new, spatially extensive city forms possible, initially and most spectacularly in Los Angeles, a city that also embraced an innovative communications technology, film. At their margins all American cities began to resemble Los Angeles: first streetcars, then buses, provided public transport, trucks brought goods from shop to house, and automobiles allowed individual commuting. Such local transport improvements allowed much more complex city forms in which people could live richer lives with access to many more goods, services, and experiences.

Communications, the Dominant Technology

Although it was not obvious in 1914, between then and 1945 communications technology came to dominate both trade and warfare. Britain was able to read all cable traffic passing through London, coded or not, and ensured America's entry into World War I with the notorious 'Zimmerman telegram'. In the 1920s Britain's need to communicate cheaply around its spatially extensive Empire and its even more extensive global trade routes led to massive innovation in shortwave wireless, which in the mid 1930s led the British to radar as a way of resisting the renewed German threat of air attack. American innovations leading toward television as an entertainment medium were seized on quickly by Britain in the late 1930s as even more useful for the display of radar signals in night fighters. Britain's code breakers continued their communications intelligence success by breaking Germany's 'enigma' machine code. After World War I Britain and America entered a long struggle to dominate global communications, which America 'won' only in the 1970s. The first of many American submarine telephone cables crossed the Atlantic in 1956, and in the 1970s the global network of communications satellites run by Intelsat finally gave America a clear communications advantage.

The Rise of Air Power

The second great inter war development was in air transport and warfare. In the mid 1930s all metal, twin engined air transports started to appear, in particular in America, which were profitable and safe carrying passengers as well as mail and used sophisticated radio communications to navigate after dark and in bad weather. By the late 1930s the great ocean barriers to commercial air transport were also falling. Such long range transports promised long range bombing of an opponent's cities. Naval aviation using aircraft carriers developed alongside land based aviation. World War II developed as the world's first real air war, two of its four 'turnover' battles, Britain and Midway, being totally decided by air power. Of the other two, El Alamein and Stalingrad, only the latter resembled the battles of immobility from World War I. What won the crucial war in the air was primacy in the struggle for supremacy in electronic communication that surged around radar. By 1941 Britain and America were slowly but surely gaining the upper hand in the technology, its application, and its mass production, easily against Japan, with difficulty against Germany. At the end of World War II Berlin was the first high technology city to be 'removed from the map', its pre war abilities in electronics 'lost' to its competitors, in particular London and Boston.

Communications have dominated the period since World War II. American primacy was brief, satellites being rapidly overtaken in the 1980s by fiber optic cables, though they have remained vitally important as fixed reference points for the Global Positioning System (GPS). As they recovered from World War II the European countries and then Japan experienced a rapid rise in automobile ownership akin to that of America after 1915 and had to create new infrastructure and rebuild cities. Similar forces now seem at work in the rest of East Asia, especially China. Jet aircraft have radically altered the ability to move people and high value freight around the planet, many airlines being as much about freight as passengers. The web of human interaction is now both truly global and greatly thickened by the success of the Internet. The most profound change has, however, been in global geopolitics. The combination of GPS with air power, especially precision guided munitions (PGMs), has allowed the trading states to re exert the global authority Mackinder believed they had lost to the territorial states early last century. GPS and PGMs allow control of flows and nodes even well inside opposing territorial states without the need to occupy territory, never the long suit of trading states with their dislike of expensive standing armies. Of course PGMs of a cruder sort are now deployed against the trading states and their allies, as the rise of suicide bombing and the events of 9/11 demonstrated. At a geopolitical level this does not seem like a struggle of the sort Mackinder feared, a struggle that ended with the collapse of the last serious territorial state at the end of the Cold War, merely an attempt by dissident groups to resist the return of the trading states to the position of global dominance their constant attention to control of flows and nodes, thus of trade, transportation, and communications, has given them.

See also: Capitalism and Division of Labor; Migration, Historical Geographies of.

Further Reading

Abu Lughod, J. L. (1989). *Before European Hegemony: The World System A. D. 1250 1350*. New York: Oxford University Press.
Agnew, J. (2003). *Geopolitics: Re visioning World Politics* (2nd edn.). New York: Routledge.
Boyne, W. J. (2003). *The Influence of Air Power Upon History*. Gretna, LA: Pelican.
Cronon, W. (1991). *Nature's Metropolis: Chicago and the Great West*. New York: Norton.
Hall, P. (1998). *Cities in Civilization*. New York: Pantheon.
Hall, P. and Preston, P. (1988). *The Carrier Wave: New Information Technology and the Geography of Innovation, 1846 2003*. London: Unwin Hyman.
Hugill, P. J. (1993). *World Trade since 1431: Geography, Technology, and Capitalism*. Baltimore, MD: Johns Hopkins.
Hugill, P. J. (1999). *Global Communications since 1844: Geopolitics and Technology*. Baltimore, MD: Johns Hopkins.
Hugill, P. J. (2003). Technology, its innovation and diffusion as the motor of capitalism. *Comparative Technology Transfer and Society* 1, 89 113.
Hugill, P. J. (2005). Trading states, territorial states, and technology: Mackinder's unexplored contribution to the discourse on state types. In Blouet, B. W. (ed.) *Global Geostrategy: Mackinder and the Defence of the West*, pp 108 125. London: Frank Cass.
McNeill, W. H. and McNeill, J. R. (2003). *The Human Web: A Bird's eye View of World History*. New York: Norton.
Mackinder, H. J. (1904). The geographical pivot of history. *The Geographical Journal* 23, 421 444.
Mackinder, H. J. (1919). (1941). *Democratic Ideas and Reality: A Study in the Politics of Reconstruction*. New York: Henry Holt.
Mahan, A. T. (1890). (1957). *The Influence of Sea Power Upon History, 1660 1783*. New York: Hill & Wang.
Meinig, D. W. (1986 2000). *The Shaping of America, 4 vols*. New Haven: Yale.

Relevant Websites

http://www.nrm.org.uk
British National Railway Museum & Railroad Museum of Pennsylvania.

http://www.porthcurno.org.uk
Cable & Wireless s historic site in SW England.

http://www.infoage.org
Excellent on history of wireless from a US perspective.

http://www.nationalmuseum.af.mil
USAF museum site. Links to other major air power museums around the world (especially the Royal Air Force Museum & Canada Aviation Museum).

Transatlantic

M. Sheller, Swarthmore College, Swarthmore, PA, USA

Transatlantic

The contemporary field of transatlantic geographies rests on a translocal, postnational conception of regional or area studies. It has emerged out of recent developments in several disciplines, including colonial history, postcolonial literature, cultural anthropology, cultural studies, postcolonial studies, sociology of race, and cultural geography. All emphasize the interconnectedness of metropole and colony and the importance of transoceanic networks in the making of a transnational and transimperial Atlantic world. Postcolonial geographies have de centered Europe by showing its complex and intimate relation with the colonial world, disrupting the geographical imagination of nationally bounded units of analysis by bringing multifaceted transactions and relations with the periphery into view. Transatlantic studies thus seek to connect and combine the various and hitherto artificially separated histories of Western Europe, Africa, the Americas, and the Caribbean, and to demonstrate how the most important processes of social, political, and economic change can be understood only through a transnational frame of reference. An emphasis on transatlantic travels, exchanges, networks, and cultures fundamentally challenges the traditions of Eurocentrism within national historiographies and the social sciences by showing how the very idea of Europe, and related ideas such as whiteness, the West, modernity, and freedom were constructed in relation to, and in interaction with, the wider Atlantic world, including its indigenous Native Caribbean, Native American, and enslaved and free African populations.

Crossing the Atlantic

Transatlantic histories usually begin with Genoese 'caravels' finding their way through the straits of Gibraltar in 1277, and to the Canary Islands in the mid fourteenth century, opening up what had been until then a largely Mediterranean centered subsystem linked overland to the larger trade routes of Asia. This was followed by Portuguese voyages to the Madeiras, Azores, and Cape Verde Islands in the 1420s, 1440s, and 1460s, respectively. Settlements on these islands positioned the Portuguese to begin trading for gold on the West Coast of Africa, facilitated by the invention of the astrolabe in 1497 by Abraham Zacuto, a Jewish astronomer. In 1482 the Portuguese established the trading fort known as Elmina on the edge of Akan country (in modern Ghana, now a major heritage tourism destination), first used to store gold and later to hold enslaved captives prior to transport across the Atlantic. With support from the rival rulers of Castile and Aragon, Isabella and Ferdinand, the Genoese navigator Christopher Columbus famously set sail in 1492 in search of a western route to Asia, reaching the Bahamas 33 days after leaving the Canaries. He is now credited with the 'discovery' of the New World, although it is generally accepted that Viking ships reached North America in the late fourteenth century, and recent evidence (from chicken bones) surprisingly suggests that Polynesian ships may have reached the Pacific coast of South America even earlier. But it was only in the sixteenth century that a steady flow of ships and communication was established across the Atlantic, especially after the Spanish conquest of the Aztec (1519–21) and Inca (1533–35) empires.

This transoceanic connectivity initiated an era of exchange in which not only people made the Atlantic crossing, but also plants, animals, diseases, cultures, and technologies. Smallpox is estimated to have killed up to one half of the native population of Hispaniola in 1518–19, with only 2000 natives left in 1542. Disease brought from Europe is also estimated to have caused Central Mexico's population to decline from 5–10 million (1519) to 2.7 million (1568), and Peru's from 9 million (1532) to 1.3 million (1570). Up to 90% of some groups of New England native peoples died in 1616–20. The natural world was reconfigured in other ways. As Alfred Crosby has shown, maize, the common bean, potatoes, tomatoes, squash, and sweet potatoes all moved by ship from the Americas to Europe, while wheat, rye, oats, and Old World vegetables, along with commercial crops such as sugar cane, coffee, rice, and oranges entered the New World. And, as Judith Carney stresses, Africans carried Guinea corn, yam vine, gourd vine, okra, black eyed peas, watermelons, and red husked rice into their American gardens. Domesticated animals such as goats and sheep were released on some islands to become feral populations, causing heavy damage to insular species, while several West Indian islands were rapidly deforested leading to soil erosion and dessication, and initiating European environmental consciousness according to Richard Grove.

Human migration to the New World was also extensive, with the largest component being the coerced migration of enslaved Africans. For example, from the

British Isles approximately 380 000 people emigrated to the Americas in the seventeenth century, and 270 000 in the eighteenth century, with more than one half settling in the Caribbean. In comparison, coerced migration from Africa, according to David Eltis, totaled 607 000 from 1580 to 1640, 829 000 from 1640 to 1700, 2 846 000 from 1700 to 1760, 4 325 000 from 1760 to 1820, and 2 296 000 from 1820 to 1880, when slavery was finally abolished in Cuba and Brazil. In addition to the displacement of in digenous populations, the massive coerced migration of Africans, and the voluntary migrations of Europeans into the New World, other significant migration streams around the Atlantic world included the expulsion of the Jews from Spain and Portugal in the fifteenth century, the migration of indentured contract laborers from India and China into the Caribbean in the nineteenth century, the movements of people within and beyond the Caribbean in search of employment especially in the mid to late twentieth century, and the ongoing flow of migrants from Central America and Mexico into the United States today.

Transatlantic Historiography

Human geographers approach these large scale processes of transatlantic change in conversation with several other disciplines. Transatlantic voyages were crucial to the history of cartography, and to the ongoing relation be tween maps and colonization. Study of the early Euro pean maps of the Atlantic and their comparison with earlier world maps has been a central theme in research on the geographical imagination and forms of geographic representation. The transatlantic provides one of the major foci for studies of migration, island development, colonialism, the formation of diasporas, race and racial formation, and geographies of food (see other entries in this work). However, the concept of the transatlantic ei ther as a region or as overlapping transoceanic networks has perhaps been best developed within the new Atlantic history. David Armitage describes three different ap proaches within recent Atlantic history: Circum Atlantic (transnational in scope), Transatlantic (comparative international or inter imperial studies), and Cis Atlantic (the study of specific localities as parts of the whole). Each of these is contributing to an emerging field of Transat lantic Studies in which there are now graduate programs, centers of study, and numerous international conferences.

The term 'Trans Atlantic', however, has an older ge nealogy, having first come into use in the late eighteenth century among English writers referring to events taking place on the other side of the Atlantic, such as the American War of Independence. By the early nineteenth century a genre of self consciously transatlantic travel writing was emerging, with numerous writers, from Captain J. E. Alexander (1833) to Henry James (1875), titling their works *Transatlantic Sketches*, encompassing journeys between Britain, the West Indies, the coastal regions of South and Central America, and North America. By the early twentieth century the Atlantic Ocean was understood as forming the heart of Western civilization, with a shared Judaeo Christian heritage (interpreted as European in origin) linking the 'civilized' (understood as 'white') peoples of Northern Europe and North America (Bonnett). This worldview influenced the historiography on the *Age of Democratic Revolution*, such as R. R. Palmer's influential work of this title, which focused on the French and American revolutions in comparative perspective (problematically excluding the Haitian revolution from discussions of the origins of democracy and modernity).

In the World War II era this idea of a democratic and civilized North Atlantic culture supported the notion of a 'transatlantic partnership' between Europe and the United States, which became formalized under the North Atlantic Treaty Organization (NATO) and shaped transatlantic geopolitical relations in the Cold War era. The legacies of this partnership can be seen today in the field of Trans atlantic Studies within International Relations (IR), insti tutionalized in bodies such as the Center for Transatlantic Relations at the Paul H. Nitze School of Advanced International Studies (SAIS) at Johns Hopkins University. Scholars and policy makers within this field are mainly concerned with the evolving relationship between the European Union and the United States, and what is sometimes referred to as the 'transatlantic crisis' and the future of 'transatlantic cooperation'. The term is fre quently used in this sense by politicians and in the media.

The Black Atlantic and Beyond

Yet this North Atlantic perspective excludes Africa, Latin America, and the Caribbean as part of the Atlantic world. Thus, a second strand in the lineage of transatlantic geographies must be traced back to critical work on the African slave trade, including that of sociologist W. E. B. DuBois in the 1890s and gaining ground with the debates generated by Eric Williams' thesis in *Capitalism and Slavery* (1944), which proposed that slave based econ omies in the Americas generated the capital underpin ning the industrial revolution in Europe. Based on earlier understandings of a 'triangular trade' between Europe, Africa, and the New World, studies of the transatlantic slave trade emphasize that from the fifteenth to the nineteenth century there was an interconnected trans oceanic economy connecting the markets for slaves from West Africa, plantation commodities such as sugar, rum, coffee, salt, and indigo from the West Indies, goods such as timber, salted cod, tobacco, and cotton from the North American colonies, and early industrial trade products such as guns, cloth, beads, and pottery from Britain and Northern Europe, as well as tools, furniture, woolen cloth, linens, and silk, once colonies were established. By the 1980s there was heated debate among historians concerning the significance of the economic growth associated with the transatlantic slave trade and the possible economic causes and consequences of its abolition. New data continue to be collected, and there are numerous online resources concerning the transatlantic slave trade (see below).

A transatlantic perspective has been crucial to the emergence of African Diaspora studies and the conceptualization of the Black Atlantic as a dynamic zone of migration, communication, and cultural formation. Within cultural studies and postcolonial studies, theorists of the Atlantic world such as Paul Gilroy and Joseph Roach insist on foregrounding the diasporic histories of Africa and the 'genocidal' history of indigenous peoples in North and South America and the Caribbean, as crucial aspects of the culture of modernity. A Black Atlantic perspective offers a better understanding of the ship borne interconnections between African origin cultures in North and South America and the Caribbean, including shared heritages noted in styles of speech, music, dance, cuisine, play, and performance. It has also contributed to the theorization of creolization as the conflicted process of cultural interaction, resistance, appropriation, and transformation through which African, European, and native indigenous people's languages, material cultures, and cultural practices were remade and amalgamated in diverse 'contact zones' throughout the Atlantic world. Studies of the Black Atlantic have been crucial to generating new understandings of comparative formations of race and new transnational perspectives on histories of racism and antiracism.

Transoceanic perspectives have also taken inspiration from Fernand Braudel's landmark study, *The Mediterranean*; indeed, the circuits of Atlantic exchange extended to Baltic and Mediterranean states, as well as the Levant. Spanish and Portuguese historians long recognized that their transatlantic expansion and colonization, in particular New World bullion imports, were inseparable from their modern political and economic development. A preponderance of work on the North Atlantic is now being balanced by new research on the Tropical Atlantic and the African Atlantic, as well as more interest in the Hispanic and Portuguese Atlantic among Anglophone scholars. There is even an awareness of transpacific connections, exchanges, and networks, which also may be incorporated into the transatlantic perspective. Historical and literary studies of European transatlantic cultures now include not only the British Atlantic, which has predominated until recently, but also a growing body of work on the Hispanic, French, Dutch, Irish, Welsh, and Scottish Atlantic. It also includes histories of the mobile and multiethnic working class and dispossessed of the Atlantic world – the sailors, pirates, apprentices, indentured servants, slaves, domestics, proletarians, and vagabonds brought to attention especially by the work of Peter Linebaugh and Marcus Rediker. The Atlantic's age of revolution, Rediker shows, mobilized these multiethnic, multiracial, multicultural, and transnational masses in ways that arguably changed global politics in decisive, irreversible ways.

Transatlantic Cultural Geographies

The initial emphasis on systems of production, commodity markets, and laboring classes has also expanded into a greater interest in systems of consumption, material cultures, and visual cultures of the Atlantic world. Studies of the circulation of specific food commodities such as sugar (Mintz) and rice (Carney), as well as new histories of colonial botany and plant collecting, have demonstrated the extensive transatlantic networks through which plants, seeds, and knowledge of cultivation techniques and botanical properties of specific plants have traveled, as well as recovering the crucial contributions of African and indigenous Americans to the production of such systems of knowledge and practical know how. Abolitionism was also a thoroughly transatlantic movement that worked through transoceanic networks of antislavery print and communication, and religious affiliations such as Quakerism, while mobilizing consumers in England, France, and the United States to boycott products grown by slaves. Such movements for ethical consumption are the forerunners of more recent movements for fair trade and debates over the sustainability of certain forms of agricultural production, tourism, and development that link together northern Atlantic consuming publics with poorer producers or service providers in the Caribbean and Central and South America. Geographies of food suggest how everyday forms of shopping, provisioning, and ingestion can be used to educate consumers about global food networks.

The study of material culture and consumption in the Netherlands, Britain, and France has also highlighted the extent to which products produced in the Americas entered the domestic realm through everyday rituals such as taking tea or coffee with sugar, purchasing mahogany furniture, using tobacco and rum, or making clothes from cheap slave grown cotton. Wealth from the West Indies transformed not only domestic life in Europe, but also civic culture, urban public space, and such iconic national spaces as country estates and their gardens. Studies of the visual cultures of the Atlantic world are also beginning to examine how European art once viewed through a solely national and ethnocentric lens is thoroughly imbued with transatlantic themes such as slavery, racial difference, racial equality, the noble savage, and moral reflections on luxury goods. By tracing the movements of material objects around the Atlantic world, contemporary cultural geographies also emphasize how the movement of goods stitches together a transatlantic region, but does so in complex and uneven ways (Cook and Harrison).

Transatlantic studies in their 'cultural' turn also encompass the making of transnational identities and forms of translocal belonging, shedding new light on current concerns with globalization, transnationalism, multisited ethnographies, and similar attempts to broaden the scope of anthropological and sociological inquiry. One entrée into this is through the biographies of particular individuals who used transatlantic crossings to construct multilocal identities, ranging from Olaudah Equiano, Benjamin Franklin, and Marcus Garvey, to Gertrude Stein, James Baldwin, and C. L. R. James. Social historical approaches consider the wider formation of transatlantic diasporas, circuits of transmigration, and the recent phenomenon of transatlantic heritage tourism which builds tourist circuits around historical sites connected with the slave trade or with the vernacular histories of European migration.

An analytical focus on the Atlantic region as an historical field integrating Europe, Africa, and the Americas furthermore changes the way in which modernity (and hence postmodernity or late modernity) can be conceptualized. Global modernity is now often viewed through the prismatic lens of 'discrepant' or 'alternative' modernities, plural, dispersed, and polyvalent. New historical understandings of the Atlantic world can contribute to resituating debates about contemporary globalization by showing how various regions have long been interconnected. The identification of the Atlantic as a special sphere of cultural, economic, and social relations articulates with and depends integrally on particular ideas of the East, or the Orient, and thus connects to studies of Orientalism and colonial/imperial relations within Asia, as well as an emerging understanding of the idea of the West as generated from multiple locations, with different meanings in different periods (Bonnett).

In contrast to the structural constraints of world systems theory or theories of dependency and underdevelopment that focus on economic relations and use national units of analysis, a transatlantic perspective opens up possibilities for an analytics of interdependencies and world relations that occur at a variety of scales and in both economic and noneconomic relations. Transcending disciplinary boundaries, conventional subcategories, and accepted periodizations, the transatlantic draws new connections between the local, intimate, bodily scale; the urban, regional, interlocal meso level; and the transnational, macro level, global scale. Thus the analysis of formations of race, gender, class, and sexuality at the level of intimate and day to day interbodily contact play a central part in unpacking transnational power relations and dynamics that occurred at a transoceanic scale.

Transatlantic Mobilities

Transatlantic studies call attention, above all, to relations, transactions, networks, and mobilities. Rather than emphasizing a bounded region, it emphasizes the extended circuits of human, floral, faunal, capital, visual, and informational movements that constitute transoceanic networks of production and consumption. Thus transatlantic geography intersects with recent theoretical developments within actor network theory and mobilities research by focusing on the interconnected 'hybrid' mobilities of people, commodities, texts, visual images, material cultures, and natures, as well as the infrastructures of transport and communication that channel and enable such mobilities (and immobilities) of humans and nonhumans. These infrastructures have of course been massively transformed since the mid twentieth century by air travel, the laying of transatlantic cable, satellite communications, the internet, and high speed freight shipping. Cultures of travel have also been transformed by new forms of transmigration, border control, surveillance, and by uneven access to forms of mobility.

Transatlantic mobilities are not only about physical travel and movement, but also about imaginary travel, imaginary worlds of connection, and imaginary geographies. What we might call the transatlantic imaginary not only informed the acts of exploration, mapping, colonization, migration, and botanizing that have occurred over several centuries, but also flourished in the forms of travel writing, visual representation, and area studies that have come to define the transatlantic as a field of study. Thus, transatlantic geography also is concerned with the invention of the idea of the transatlantic, what it has come to mean and to do in a performative sense, what role it has played in different eras and contexts, and how it operates as a particular system of knowledge and representation. The intersection of the physical geographies of transatlantic crossing and connectivity and the ideational geographies of transatlantic representation and imagination, and how these have together changed over time, remain a rich and expanding field of geographical research.

See also: Cartography, History of; Colonialism I; Colonialism II; Diaspora; Eurocentrism; Heritage; Island Development; Migration; Migration, Historical Geographies of; Modernity; Natures, Postcolonial; Postcolonialism/Postcolonial Geographies; Race; Transnationalism.

Further Reading

Armitage, D. (2002). Three concepts of Atlantic history. In Armitage, D. & Braddick, M. (eds.) *The British Atlantic World, 1500 1800*, pp 11 30. Houndmills: Palgrave Macmillan.

Bailyn, B. (2005). *Atlantic History: Concept and Contours*. Cambridge, MA: Harvard University Press.

Bonnett, A. (2004). *The Idea of the West: Culture, Politics and History*. Basingstoke: Palgrave Macmillan.

Carney, J. (2002). *Black Rice: The African Origins of Rice Cultivation in the Americas*. Cambridge, MA: Harvard University Press.

Cook, I. and Harrison, M. (2003). Cross over food: Re materialising postcolonial geographies. *Transactions of the Institute of British Geographers* 28(3), 296 317.

Crosby (1972). *The Columbian Exchange: Biological and Cultural Consequences of 1492*. Westport, CT: Greenwood Press.

Cunliffe, B. (2001). *Facing the Ocean: The Atlantic and Its Peoples, 8000 BC AD 1500*. Oxford: Oxford University Press.

Curtin, P. (1990). *The Rise and Fall of the Plantation Complex*. Cambridge: Cambridge University Press.

DuBois, W. E. B. (1896). *The Suppression of the African Slave Trade to the United States of America, 1638 1870*. New York: Dodo Press.

Eltis, D. (1987). *Economic Growth and the Ending of the Transatlantic Slave Trade*. Oxford and New York: Oxford University Press.

Gilroy, P. (1993). *The Black Atlantic: Modernity and Double Consciousness*. London: Verso.

Grove, R. (1995). *Green Imperialism: Colonial Expansion, Tropical Island Edens and the Origins of Environmentalism, 1600 1860*. Cambridge: Cambridge University Press.

Klooster, W. and Padula, A. (eds.) (2005). *The Atlantic World: Essays on Slavery, Migration and Imagination*. Upper Saddle River, NJ: Pearson Education.

Linebaugh, P. and Rediker, M. (2000). *The Many Headed Hydra: Sailors, Slaves, Commoners, and the Hidden History of the Revolutionary Atlantic*. London: Verso.

Mintz, S. (1985). *Sweetness and Power: The Place of Sugar in Modern History*. New York: Viking.

Morrison, M. A. and Zook, M. S. (eds.) (2004). *Revolutionary Currents: Nation Building in the Transatlantic World*. Lanham, MD: Rowman and Littlefield.

Quilley, G. and Kriz, K. D. (eds.) (2003). *An Economy of Colour: Visual culture and the Atlantic World, 1660 1830*. Manchester: Manchester University Press.

Roach, J. (1996). *Cities of the Dead: Circum Atlantic Performance*. New York: Columbia University Press.

Sheller, M. (2003). *Consuming the Caribbean*. New York: Routledge.

Thornton, J. (1992). *Africa and Africans in the Making of the Atlantic World, 1400 1680*. Cambridge: Cambridge University Press.

Williams, E. (1964). *Capitalism and Slavery*. (first publ. 1944). London: Andre Deutsch.

Relevant Websites

http://hitchcock.itc.virginia.edu
Exploring Cultural Landscapes, Digital Media Lab, University of Virginia: The Atlantic Slave Trade and Slave Life in the Americas: A Visual Record (Virginia Foundation for the Humanities). A collection of 1200 images from a wide range of sources, most of them dating from the period of slavery.

http://www.fas.harvard.edu
Harvard University: The International Seminar on the History of the Atlantic World was established at Harvard University in 1995 by Bernard Bailyn, with the support of the Andrew W. Mellon Foundation. Members of the Seminar are drawn from the nations of Western Europe, Africa, and Latin America, joined by US and Canadian scholars who are also at an early stage of their careers, for presentation of work in progress, discussions of the theme of the Seminar, and exchange of views with senior scholars. The site includes Working Papers, syllabi, and bibliographic resources.

http://www.h net.org
H Net: H Atlantic is an international online discussion list for Atlantic World History from 1500 to 1800. This an interdisciplinary list for scholars who study colonial North America and the United States, Europe, West Africa, the Caribbean, and South America in a transatlantic context.

http://www.star.ac.uk
The Star Project: Scotland's Transatlantic Relations Project (STAR) provides a focus for international circumatlantic research with a Scottish dimension. Through a combination of seminars, conferences, publication ventures, and other collaborations, it fosters interchange and discussion; it also makes relevant information and resources available to a wider audience via this website.

http://portal.unesco.org
UNESCO: Slave Trade Archives. This website aims to outline the history of the Transatlantic Slave Trade; to present the various archival funds and the document typology according to places where they are preserved; and to provide access to a database of digital archives relating to the transatlantic slave trade.

http://www.ucalgary.ca
University of Calgary: Peopling North America: Population Movements and Migration. This site offers an overview of migratory movements and diasporas within Canada, the United States, Mexico, and the Caribbean from Europe, Asia, and Africa.

http://www.umich.edu
University of Michigan: The Atlantic Studies Initiative (ASI) focuses on the circum Atlantic flow of peoples, cultures, goods, and capital. It explores the interaction and interdependencies of Atlantic cultures from Africa to Europe and across the Americas and the Caribbean.

Transcripts (Coding and Analysis)

M. Cope, University of Vermont, Burlington, VT, USA

Glossary

Axial Coding Coding along a predetermined axis that represents the interests of the researcher and informs the basis of the research questions.

CAQDAS (Computer-Aided Qualitative Data Analysis Software) A generic name for a wide range of computer programs that aid in storing, organizing, coding, and analyzing qualitative data.

Coding The process of identifying, assigning, and analyzing codes based on inductive review of transcripts and other data for the purpose of revealing commonalities and disjunctures, investigating patterns and themes, and producing new knowledge and insight. Coding is both a data reduction process and a recursive, fluid, powerful analytical technique. Different types of coding include *in vivo*, axial, and open coding; these represent a spectrum from descriptive to analytical.

***In Vivo* Codes** Codes originating from expressions, phrases, or terms used by research participants.

Open Coding Coding without preidentified themes, goals, or categories; remaining open to what the data convey.

Recursive Analysis The process of going back cyclically in multiple iterations to re-evaluate data and previous analysis in light of new insights.

Transcription The operation of transferring audio- or video-taped material into verbatim text documents, which are usually then coded and analyzed. This involves typing up audio- or video-recorded interactions (focus groups, interviews, etc.), researchers' fieldnotes, participants' products (travel diaries, or other handwritten materials), or archival data for the purpose of digital data storage, data sharing, and analysis.

Introduction

Transcripts are text documents that can originate from several sources in the process of research. Researchers commonly choose to transcribe the audio recorded (or video taped) verbal accounts, expressions, exchanges, and interactions that occur in research activities such as focus groups, interviews, and oral histories. In addition, scholars working on archival data may transcribe text from materials and historical ephemera they are examining (e.g., handwritten diary entries, letters, signs, and text in photos) as a way to integrate them more seamlessly into digitally stored data. Finally, many researchers use transcription for their own fieldnotes, whether the notes were originally audio recorded or written in a shorthand style while in the 'field'. The common element is that materials from various sources are 'typed up', that is, transformed into text, which can then more readily be systematically cataloged and analyzed.

There are both challenges and strengths to the transformation of various materials into text documents. The challenges of transcription lie both in the practical and conceptual realms. First, the act of transcribing itself is often tedious (and/or expensive), though there are also benefits to doing it oneself, which are examined below. Second, transcripts cannot convey the same meanings, emotions, reactions, humor, or hesitations that can be heard, seen, and experienced in the real time, real life experience of, say, a focus group or interview. Even inserting comments for 'laughter' or 'subject hesitates' in brackets in the text is only a small representation of what could have been a significant moment in the research. Thus, transcripts are always acknowledged as a partial and imperfect data source, though there are ways to minimize errors.

The primary strength of transcription is that our world is increasingly digital, including not only our data (text, maps, images, and numbers), but also the software that helps us search, code, and analyze text, such as computer aided qualitative data analysis software (CAQDAS). A mixed methods research project may generate or collect data from multiple activities, such as interviews, participants' travel journals, focus groups, the fieldnotes produced by one or more researchers, newspaper accounts of local events, digitized maps, census or other secondary data, and informational documents produced by relevant social/cultural organizations. As scholars become increasingly strong users of electronic databases to manage their projects and data, this multitude of data becomes easier to organize and analyze if fully digitized. New advances in CAQDAS also allow users to include nontext data such as photographs, audio and video files, and maps in the database, allowing richer processes of analysis (see **Figure 1**).

Practical Issues: Transcription in Research Projects

Long before the transcription machine is hauled out, researchers planning on using verbal data collection methods should think ahead to the practicalities of transcription, both in terms of equipment and strategy.

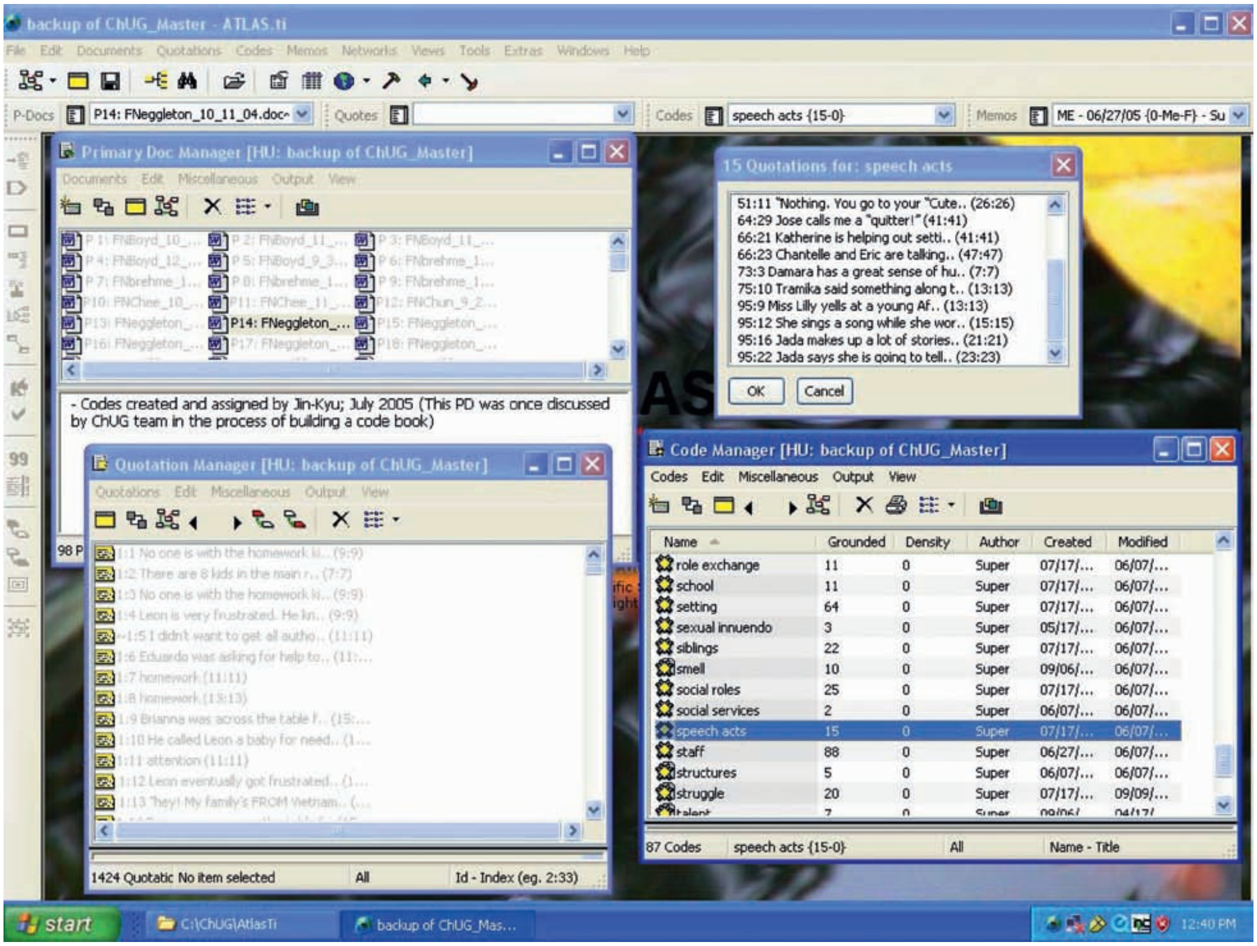

Figure 1 Screen capture of coding in Atlas.ti. Source: Author.

First, it is often startling to new qualitative researchers just how long it takes to arrange focus groups or interviews or other research involving real people, then to conduct these, and finally to transcribe the proceedings. Many grant proposals reflect this ignorance in overly ambitious research timelines. One of the most important insights for those engaging in qualitative research for the first time is that greater volume of data does not (necessarily) mean a better or more 'rigorous' project. Rather, a more realistic proposal that allows both the subjects and the researcher appropriate time to talk, listen, process, reflect upon, and analyze the discussions and observations will result in richer, more detailed, and ultimately more solidly grounded interpretations. Brief surveys can give a superficial sense of what a large number of people think about an issue, and they have a lot of value in that sense, but for deeper understandings of 'context' and 'process', a smaller number of in depth interviews (or life histories, or a similar other method) can be more valuable. However, these take a great deal of time, skill, and attention both to perform and to turn into usable 'data' through transcription.

The second common mistake is to assume that transcription of recorded talk is a straightforward, mechanical (or electronic) process. The digital age has indeed improved and smoothed transcription processes to some degree. Digital audio (and video) recorders can store hundreds of hours of material without the need for audio tapes (with their frailties and failures), and some researchers are using MP 3 players to record and organize voice files. With a digital device, voice files can easily be transferred to a computer for data storage, backup purposes, data sharing, and transcription. At this time, voice recognition software is still fairly rudimentary; there are some programs that translate voice files to text, but they require 'training' by the speaker to allow the software to understand that person's intonations and pronunciations. This can be useful for certain applications, such as when a researcher records audio fieldnotes and wants them transcribed, but in other 'talk' situations where there are multiple speakers, cross talk, and so on, such software is of limited use.

This means that someone still has to listen to verbal exchanges and type them up. Sometimes this is the researchers themselves, other times it is a process that is outsourced to a transcription service; each scenario has benefits and problems. The benefits to researchers doing their own transcription are several: (1) transcribing forces one to listen and review the talk based material, providing another round of reflection and analysis; (2) researchers are likely to understand more of the verbal accounts because they have already heard it and because they are very familiar with the context, specific names, and acronyms that might be used; and (3) researchers will also recall (and hopefully have written notes on) more of the nonverbal elements of the recorded episode, such as a moment of humor or tension, someone entering or leaving the room, facial expressions and body language, or other cues to emotion such as tears, nervous behavior, smiles, or frowns. All of these contribute to more accurate texts that are 'thicker' with description and supporting information. The primary drawback for researchers doing their own transcription lies in its heavy time commitment.

These, conversely, are as the main drawbacks and benefits to having a transcription service perform the work – it can be more cost effective and time efficient for researchers to have transcription done, but the nuances and context are fewer, the accuracy is lower, and they miss out on the value of hearing and processing the information themselves. These differences (between doing transcription oneself and outsourcing) are reflected in the 'rules of thumb' for transcription time: doing it oneself likely will take up to 4 hours to transcribe 1 hour of material (less for fieldnotes, more for focus groups or interviews), while outsourcing usually takes approximately 2 hours per 1 hour of material. The time difference lies in the added value the researcher gives to the final product (richer, more accurate, and contextualized transcriptions) as well as the fact that researchers inevitably begin some of the coding and analysis process while transcribing, such as jotting down themes, pulling out interesting phrases, taking notes, identifying patterns, and so on.

Coding and the Rigorous Analysis of Transcripts

The primary purpose of transcribing research materials into text is to be able to systematically examine and analyze them in order to produce new knowledge, which in turn should help answer the research questions initially posed in a project. One method for getting a rough sense of what is contained in the data is content analysis. Content analysis is essentially a quantitative analysis method involving counting the instances of a particular phrase or term in the body of the text. For example, in a series of interview transcripts, a researcher might want to explore how many different subjects raised a particular topic, such as 'the media'. Counting the incidence of phrases such as newspaper, radio, TV, internet, etc., and who used them more or less frequently will give a general sense of how often the topic of 'media' is raised. Content analysis is fairly simple and even keyword searches in a word processor can be used to find specific terms and how often they occur. However, content analysis is not nearly as powerful as 'context analysis', in which themes, patterns, correspondences, and inconsistencies are identified and built upon through careful reading, systematic coding, and recursive critical analysis.

Coding refers, at its simplest, to a data reduction process of creating categories based either on inherent qualities of the data or on elements predetermined by the researchers to be of particular interest. These categories (codes) can then be used as shorthand to qualify sections of text, visual segments in a photograph or video, or other primary data. For example, a researcher might have an interest in people's use of transportation; she could construct codes for different modes of transport, different ways people feel about those modes, common problems people encounter, benefits people see or experience in using transportation, and categories based on people's locations and access to different modes.

The process of coding is not complicated, but it does require attention and care, and a coding structure can quickly become complex. The first step is to go back to the research questions and identify any codes that are implicit in the topic, such as the above list regarding transportation. These are entered into a 'codebook', which literally was a book in the past, though it is generally a computerized list, spreadsheet, or set of entries in a CAQDAS format. The second step of coding is to begin reading through the transcripts slowly and thoughtfully. The information revealed by line by line analysis is very valuable, but so is the more general insight gained by reading through a whole paragraph, several pages, or an entire interview.

Coding proceeds through the identification of codes of several types. The first is *in vivo* codes, which means words or phrases actually used by the subjects and appearing in the transcript. In the transportation example, we might discover the term 'transfer' used frequently among public transit users and notice from their accounts that discussion of transfers is often accompanied by expressions of hassle and stress. 'Transfer' would thus become an *in vivo* code because the term came from within the words of the subjects, but it would also take on a contextual meaning in its frequent association with difficulty; when combined with demographic information, it might become apparent that transfers are especially difficult for older people, or perhaps for those living in a particular location (marked by race and class) whose transit connections are poorly scheduled and often result in long waiting periods.

A second type of coding is known as 'open coding'. As its name implies, it is a process in which the researcher is open to whatever seems to pop out while reading the transcripts. It might be something that the researcher noticed during the interview (or focus group, archival review, etc.) stage – people kept mentioning a word, idea, or phenomenon – or it might be something that went unnoticed during the data collection phase but becomes obvious while working through transcripts. As with other types of codes, open codes do not necessarily have to be based on something that is widespread or occurs re peatedly; indeed, one might have a code for a fairly rare instance or an exception to the general findings in order to pursue commonalities between these rare instances.

The third most common type of code is an 'axial code'. This is the formal name for themes, patterns, or concepts that the researcher had identified early in the process and which may even be part of the research questions. Axial coding involves coding along a selected axis, for example, continuing with the transportation project, the researcher may have predetermined that s/he was interested in the challenges faced by the elderly, disabled, or people with small children using public transport. These challenges would form an axis along which the material would be coded.

Any given code may emerge from several of these origins, and there is no particular need to keep the three types separate. For instance, 'challenges in using public transit' might have started as an axial code, but the subcode 'transfers' would be an *in vivo* code, having emerged from people's comments. The process of open coding might reveal a common coping mechanism for dealing with the challenges of bus transfers. This dem onstrates the ways that codes are in fact fluid, overlap ping, inform each other, and sometimes bump into each other. Coding is often perceived to be a task of drudgery, but in fact it can be one of the most exciting moments in research because it is here that new correspondences, contradictions, ideas, concepts, and knowledge are pro duced (see **Figure 2**).

Another teenage African-American boy is flirting with a Latina girl right in front of me. She must be 13 or 14, very made-up, hair pulled back in ponytail, lots of gold jewelry. He is sort of	Race Gender Sexual innuendo Age	Teens are out of place' here at this time of day, but few options for them are teens always and everywhere out of place'?	Teens aren't supposed to be at the program until the evening – why were this boy and girl here? Cold outside? Where do teens go in the winter after school in harsh climates, anyway?
leaning into her, teasing her verbally and physically at the same time. At one point he sort of grabs his crotch and sort of pumps his hips in her direction – not directly at her, but	Physical appearance Speech act, teasing, comportment	Sex and pop culture saturate kids' comments and actions	We've seen a lot of these expressions of sexual themes, even in younger preteens. Notice that many Latina and African-American girls wear make-up, jewelry, and revealing clothing, especially when they hit puberty, sometimes even before.
sort of implying it. It's blatantly sexual, then I think he catches me looking, and sort of turns it into more of a dance by swinging his arms a bit and nodding his head. The girl is smiling slightly. I can't tell too much from her expression. I wouldn't say she's pleased or annoyed – more enigmatic.	Comportment, sexual innuendo ChUG effect	Observing the influence of our presence on kids	Would the boy have reacted differently to an observer who was not a white male in his 30s? Think about power dynamics
	Facial expression Emotion	Playing it cool? Perhaps masking her intimidation?	

Figure 2 Sample of coding with text, codes, themes, and notes. Field observation by graduate student, Christopher Brehme; *Children's Urban Geographies Project*, Nov. 29, 2004.

Coding is also often mistaken to be a discrete process – one builds the codes, codes the transcripts, and then the analysis is complete. However, the practice of coding is usually much more contingent, slippery, and often messy. First, in building the coding structure, there is a tendency to create enormous numbers of codes as a way to capture every detail of the primary material. However, even in a large research project, if there are more than about 50 codes the codebook becomes unwieldy, some codes go unused or only used a handful of times, there is too fine a level of detail to serve the purpose of 'data reduction' through categorization, and researchers are unable to hold so many ideas in their heads at one time. Some ways to avoid this within a CAQDAS program are to use features like Atlas.ti's 'memo' option; there may be a passage that is particularly interesting in the transcript and one wants to return to explore it more, but it does not necessarily fit existing codes. Rather than designing a new code (or sev eral) to capture the essence of this one passage, it may be better to assign a memo to it; this is basically a note to oneself about the interesting features of the passage and one's thoughts on how it connects to other themes. Other useful features of CAQDAS programs include ways to group codes together ('code family' in Atlas.ti), have sub codes within a grander code, and define the relations be tween codes such as 'is associated with' or 'co occurs with'. These allow researchers to use coding for very powerful analytical procedures and to get as much as possible out of the data they have so painstakingly collected.

A second source of fluidity in the coding process is the fact that, to be rigorous and thorough, coding must be a recursive and iterative process. For example, one often identifies new themes in material that reflect back on already coded transcripts, requiring the re coding of the earlier documents. It is entirely possible for a theme to be revealed in the course of coding the very last transcript that the researcher realizes is actually present through most of the others; s/he would need to go back and re read the earlier material with the new theme (and associated codes) in mind. Conceptually and methodo logically, the fluidity and 'messiness' of coding can, therefore, actually serve as the strengths of the process.

Final Thoughts

Because much of the work done by geographers involves multiple processes (social, economic, political, and en vironmental) as well as multiple positions and identities of subjects (race, class, gender, sexuality, disability, re ligion, and so on), it can be challenging to juggle so many balls at once. However, rather than further complicating this multiplicity, coding should allow us to generate new insights and knowledge by helping us construct well grounded themes, categories, and analysis. This is the heart of 'grounded theory': new theoretical devel opments are solidly grounded in rigorous inductive re search. By transcribing primary data materials carefully and embracing the fluidity and contingencies of the coding process for their potential power, geographers can not only use these methods to create new understanding in geography, but can also help inform research in other disciplines through the unique contributions of the spa tial perspective.

See also: Focus Groups; Grounded Theory; Interviews: In-Depth, Semi-Structured.

Further Reading

Conradson, D. (2005). Focus groups. In Flowerdew, R. & Martin, D. (eds.) *Methods in Human Geography: A Guide for Students Doing a Research Project*, pp 128 143. New York: Pearson Prentice Hall.

Cope, M. (2003). Coding transcripts and diaries. In Clifford, N. & Valentine, G. (eds.) *Key Methods in Geography*, pp 445 459. Thousand Oaks, CA: Sage Publications.

Cope, M. (2005). Coding qualitative data. In Hay, I. (ed.) *Qualitative Research Methods in Human Geography*, pp 223 233. South Melbourne: Oxford University Press.

Crang, M. (2001). Filed work: Making sense of group interviews. In Limb, M. & Dwyer, C. (eds.) *Qualitative Methodologies for Geographers: Issues and Debates*, pp 215 233. New York: Oxford University Press.

Crang, M. (2005). Analysing qualitative materials. In Flowerdew, R. & Martin, D. (eds.) *Methods in Human Geography: A Guide for Students Doing a Research Project*, pp 218 232. New York: Pearson Prentice Hall.

Jackson, P. (2001). Making sense of qualitative data. In Limb, M. & Dwyer, C. (eds.) *Qualitative Methodologies for Geographers: Issues and Debates*, pp 199 214. New York: Oxford University Press.

Peace, R. and van Hoven, B. (2005). Computers, qualitative data, and geographic research. In Hay, I. (ed.) *Qualitative Research Methods in Human Geography*, pp 234 247. South Melbourne: Oxford University Press.

Silverman, D. (2000). *Doing Qualitative Research: A Practical Handbook*. London: Sage.

Van Hoven, B. (2003). Using CAQDAS in qualitative research. In Clifford, N. & Valentine, G. (eds.) *Key Methods in Geography*, pp 461 476. Thousand Oaks, CA: Sage Publications.

Relevant Websites

http://www.qualitative research.net/
General qualitative research site (international).

http://online.hud.ac.uk/Intro QDA/preparing data.php
Qualitative Data Analysis site.

http://www.uvm.edu/~mcope/QRSG/index.html
Qualitative Research Specialty Group of the Association of American Geographers site.

Transitional Economies

J. Round, University of Birmingham, Birmingham, UK

Glossary

Command Economy Economic system developed in the Soviet Union through which the state controlled all aspects of the economy.
Post-Socialist Countries/Regions Refer to the areas that were under the influence of the Soviet Union but were not within its borders.
Post-Soviet Countries Countries which gained independence from the Soviet Union when it collapsed in 1991.
Privatization Within structural adjustment programs, this refers to the selling off of state enterprises.
Shock Therapy The simultaneous unleashing of market reforms.
Structural Adjustment Programs Originally referred to the movement away from a state-led economy but came to refer to the conditions imposed by the World Bank and the International Monetary Fund (IMF) on reforming countries in return for continuing assistance.
Washington Consensus The term given to policy advice given by Washington, DC-based institutions such as the World Bank and the IMF.

Introduction

Transitional economies are considered to be countries which are undertaking macroeconomic reforms in an attempt to alter the ways in which their economies are managed. Traditionally it implies that the country is making a structural adjustment from a state run economy toward a more market led system. The use of the term transition suggests that there is a start and endpoint to the reforms and that there is a set of policies that can be followed to achieve this. The term came into popular usage when South American countries, such as Argentina and Brazil, moved from military dictatorships toward democratic governance in the 1980s. Under military rule their economies were tightly governed by those in power, with little integration into global markets, and were often extremely inefficient. Therefore, economic problems, such as inflation and unemployment, often contributed to a regime's downfall, meaning that democratically elected leaders were faced with immediate economic crisis. Forced to look for outside assistance, states typically turned to institutions such as the World Bank and the International Monetary Fund (IMF). These groups offered loans to enable a country to avert economic collapse, and technical advice on how to implement economic reform. In order to qualify for financial assistance governments had to agree to undertake the policy advice and further installments were conditional on their continuous implementation. This advice evolved to become a 'recipe' for transitional economies to follow and as the headquarters of the World Bank and the IMF are in Washington, DC, the process became known as the 'Washington Consensus'.

At approximately the same time as the bestowing of this sobriquet the countries of Eastern Europe were beginning to break free from communist rule. The economies of these countries were dictated to by Moscow with almost every aspect centrally planned. The type and number of goods produced, their price, and location of sale were, for example, controlled by the state. The poor performance of the economy and the lack of consumer goods were some of the reasons why people wanted reform, therefore it was no surprise that the new leaders wished to develop market economies. Given their experience of effecting economic change it was logical that these states turned for assistance to the policy experts who had worked in South America. The policy prescriptions of the Washington Consensus were quickly adopted. Soon after these events the Soviet Union collapsed and the newly independent states turned to the same sources for advice and financial assistance. It was now this region of the world which became synonymous with transitional economies, though Southeast Asian states were also following similar advice to liberalize their economies.

Such polices, and the term itself, are not unproblematic. Given the very high social costs and the persistent poverty that former Soviet states have endured since embarking on this reform path, many have questioned the thinking behind the advice. The main criticism is that a transition policy 'recipe' ignores the geographical legacies of the regions undertaking them, and that too much emphasis is placed on withdrawing the state from the economy at the expense of institutional reform. Such debates challenge traditional econometric viewpoints for planning/researching economic change, as geographical approaches highlight how reform is mediated differently across regions. Therefore, a single policy recipe of economic transition in order to create a market economy, and the belief that subsequent growth will lead to increases in prosperity for the population overall, must be questioned. In other words, when examining structural economic reform geography matters. Such debates have vast importance for how governments/international organizations construct development policies and whether importance is placed on macro or micro processes. Through a case study of Russia, the discussions below demonstrate how such debates led to the development of a 'post Washington Consensus', and the idea that economies, rather than undergoing transition, are in fact in transformation.

The Washington Consensus

The theory of neoliberal macroeconomic reform, which came to be termed the 'Washington Consensus', originated from the World Bank's and IMF's responses to the economic crises faced by Latin American countries in the 1980s. The resultant development paradigm was one of a withdrawal of the state: fiscal discipline and liberalization bound up in the structural adjustment programs that came attached to financial aid supplied by the World Bank/IMF. The consensus suggested that economically struggling countries needed to follow a set of 'rules' in order to promote stable, sustainable growth. By following these rules the country would be able to progress toward a market economy and experience long term economic growth. These rules became known as the 'four pillars of transition' as countries had to liberalize, stabilize, privatize, and internationalize their economies. Within each category the following was to be achieved;

Liberalization

- The freeing of the economy from state control and political influence. The state can no longer dictate what, where, and how much is produced, or where it can be sold.
- Prices are to be determined by the laws of supply and demand. Subsidies on goods must be removed and prices for services must be brought into line with world levels.
- Regional subsidies must be curtailed.
- The use of barter between enterprises should be eliminated.

Stabilization

- As the state withdraws from the economy, it must also curb its expenditure. This entails a reduction in the size of the state apparatus and a decrease in spending on social safety nets.
- The state should avoid excessive borrowing and living beyond its means.
- It should implement a strict monetary policy to control inflation and create a stable currency.
- A myriad of laws is implemented for the whole system to work, making it an extremely complex and time consuming process.

Privatization

- The state must divest itself of ownership of the means of production. This is to be achieved through privatization schemes, open to both foreign and internal competition.
- The state should create conditions for the formation of new private enterprises.
- It is preferable that employees have the opportunity to become stakeholders in their enterprises through voucher privatization schemes.
- Monopolies are to be dismantled.

Internationalization

- The domestic economy must be opened up to international competition.
- Conditions must be created to encourage foreign investment, such as legal norms, property rights, and a 'trust' culture.
- International standards for accounting, custom duties, and tax to be developed.
- A stable, convertible currency is to be introduced.

Furthermore, these reforms had to be implemented simultaneously, in what became known as a 'big bang', in order for them to work effectively. Therefore, for example, a country could not liberalize prices and maintain state monopolies on production. This process is also known as 'shock therapy', as the initial reforms are so far reaching the market are shocked into action. This is why China, which has undertaken some of the above, is not referred to as a transition economy as the reform process was introduced gradually and the state retains some control. Fundamental to the theory of transition is the notion that the 'market knows best' and that government interference stymies economic development. Thus the Chinese model of reform was not seen as a viable alternative reform path for post socialist states as the state remained too prominent an influence.

The geopolitical aspects of these reforms must also be considered. By implementing such wide ranging reforms in a very short space of time, a very decisive break with the past has been made. For example, if a market economy is implemented and its subsequent growth leads to increased prosperity for the general public it is much harder for a military led government to regain power. Similarly, if a command economy is dismantled it becomes almost impossible for communists to return. As Fukuyama famously stated, the collapse of the communist system signaled the 'end of history', as it demonstrated the superiority of the Western neoliberal economic model and therefore, signaled the end of the Cold War. Jeffery Sachs, a leading advisor to many former Soviet countries, stated that "The thinking of the day was that 'we' had a once in a lifetime window of opportunity to effect change." This suggests that the set of policy reforms was concerned with far more than just the economic. This is not to suggest that political change was not desirable in the countries undergoing such reforms but it highlights how the reforms were, at least partially, an ideological construct. Furthermore, the opening up of these states had economic benefits to Western countries as it provided access to new markets, cheaper labor, and natural resources.

Transition in the Former Soviet Union

Using Russia as an example it is clear that while the above reforms seem relatively straightforward on paper they were incredibly difficult to implement in practice. The first major obstacle was the complexity of the reforms and the context in which they were taking place. At the collapse of the Soviet Union at the end of 1991, every single aspect of the Russian economy had to be reformed. Given that the country was also moving from a one party system to a democracy and was contending with the loss of its empire it is unsurprising that a sense of uncertainty prevailed. With the assistance of foreign experts it was decided that shock therapy was the correct path for Russia to take. However, the 'big bang' was not ignited, as debilitating political turmoil ensured that price liberalization was, initially, the only reform the federal government could introduce effectively. Even this process could only be partially executed as some regional administrations retained control of local prices, and basic goods and services remained subsidized across the country. Privatization – the largest and, in theory, irreversible step to the market – began in early 1992. By 1993 over 60 000 enterprises were privately owned, compared to only 70 in 1991. By the end of 1993, two thirds of all state assets had been transferred to private enterprises. However, this did not lead to significant increases in efficiency. The government, more concerned with transferring ownership to private hands in case it lost power, did not concern itself with who it was selling to or that it was receiving only a fraction of the market value for its assets. Liberalization occurred without stabilization, at both macro and micro levels, resulting in inflation levels reaching over 2500% in 1992, as the monetary overhang, accrued during the Soviet period, entered the economy. Attempts at macro stabilization were flawed, with money supply constricted simply by not paying state employees, suppliers, or pensions, with the failure to enforce hard budget constraints allowing bankrupt enterprises to continue operating. Two years after the process began foreign trade remained heavily regulated, any benefits of foreign aid were negated by the financing of imports by export credits, and foreign investment regulations were at best opaque.

The increasing unpopularity of the transition project, due to its ever growing social costs, led to the previously compliant communist politicians, sensing that they might be able to regain power, withdrawing their support for the process, plunging the government into crisis. Several prominent reformers were dismissed and by the end of the first year it appeared that the reform process would grind to a halt. It was only revived when Yeltsin won a nationwide referendum on his position and policies in 1993, though gaining only 53% support for the reform process. Despite this, gross domestic product (GDP) still continued to fall, leading the government to near bankruptcy. The Russian government, facing a crippling budget deficit, was offered loans by Russian entrepreneurs in return for letting the lender 'manage' the enterprises exploiting the country's resources. If the government was unable to repay the loan then it would be converted into shares, passing control to the lender. As the government could not repay the loans entrepreneurs were able to take control of key enterprises for a fraction of their true worth and an oligarch class emerged.

Despite such problems, institutions such as the European Bank for Reconstruction and Development (EBRD), the European Union (EU), and many economists, believe that Russia has completed its transition to a market economy as it has privatized production, stabilized the economy as inflation has fallen, the lack of price controls means it is liberalized, and foreign firms can operate in the region. However, many believe that the form of economy that has developed is far removed from that envisaged by the reformers. Rather than the growth of a neoliberal market, it is argued that systems of 'economic involution', 'chaotic capitalism', or a 'virtual economy' have developed instead. This, coupled with the social costs of transition detailed below, led many to question the validity of the whole transition process and the Washington Consensus.

The Social Costs of Transition

Even the most enthusiastic free market reformers acknowledged that the social costs arising from transition would be high. Each of the four pillars has a negative impact on the general population. As shown above liberalization leads to very high levels of inflation. This means that savings are wiped out overnight and the real worth of wages cannot keep pace with price rises. Privatization leads to unemployment as enterprises seek to increase efficiencies and those that cannot compete go into bankruptcy. As governments attempt to stabilize their economies, state workers often go unpaid, or the real worth of their wages falls dramatically, pension and unemployment benefits fall to match the cost of living, and other social support packages are withdrawn. Many economists note that the burdens of transition are disproportionately felt by the poorest members of society. Furthermore, some see increasing social dislocation as an indicator that the reforms are working.

What was not envisaged in the former Soviet Union is the depth and longevity of these social costs. Reformers assumed that managers and workers, now free of the constraints of the command economy, would become aware of the new market opportunities and transmute into rational actors in the new flexible labor market. This would mean, in theory, that for the majority the pain caused by the transition would be brief and beneficial as they now had the opportunities to fulfill their potential. Thus it was expected that as macroeconomic growth occurred after the initial expected collapse, wealth would trickle down to the advantage of the majority. Very few reformers gave much attention to how this was going to happen; it was just expected. By the end of 1992, 65% of Russians were considered to be living on incomes below subsistence levels and by 1995 real incomes had declined by over two thirds. While there is much debate over the causes, Russia has also seen its population decline by approximately 7 million people and male life expectancy is around 59 years. This figure is the lowest in the Northern world, and is a full 14 years below female life expectancy. In 2007 it is estimated that 20% of the population, around 30 million people, received incomes below the subsistence minimum. While this figure is a lot lower than that in 1992, it does not reveal the full story. For many more people everyday life is extremely uncertain as employment is often insecure and low pay means that saving for retirement is difficult. As the subsistence minimum figure does not reveal the true costs of everyday life, for example, it only allows for one pair of shoes every 5 years, many people know they will not be able to retire as pensions are so low. In response to this many people have developed informal practices to generate extra income, such as undertaking extra work, growing food for sale and domestic consumption, or setting up micro enterprises. While this provides a certain standard of living they do not provide long term security.

Social safety nets offer little protection. Pensions, as noted above, are low and unemployment benefits are very hard to obtain. Therefore, while unemployment rates remain relatively low, and are used to show that the reforms are working, they underestimate nonemployment. At the turn of the century it was estimated that the true level of unemployment was more than double the official 11% level. Senior citizens are entitled to free medicines. However, as the pay for government workers is so low medical staff often ask for a bribe before they issue or honor prescriptions. Corruption has become endemic in many countries that underwent transition. A 2006 report suggested that over 70% of Ukrainians have had, in the past year, to pay a bribe to access goods or services they were entitled to. Corruption has become so prevalent in these regions not just because of low wages, but due to the lack of institutional development, such as robust legal systems, during the initial reform period.

Institutions and Geographies Matter

As it became clear that the long term consequences of the shock therapy approach were far worse than expected, the rationales behind the Washington Consensus came under scrutiny. The World Bank began to publish reports with titles such as 'Institutions matter', and it was acknowledged that it had been too simplistic to expect fully functioning market economies to form without state guidance. As Joseph Stiglitz, then the World Bank's chief economist, stated in 1998,

> The policies advanced by the Washington Consensus are not complete, and they are sometimes misguided. Making markets work requires more than just low inflation; it requires sound financial regulation, competition policy, and policies to facilitate the transfer of technology and to encourage transparency. States can improve their capabilities by reinvigorating their institutions. This means not only building administrative or technical capacity but instituting rules and norms that provide officials with incentives to act in the collective interest while restraining arbitrary action and corruption. (Stiglitz 1998)

This was a dramatic reversal in thinking and spelt the end of the Washington Consensus. It was argued that such statements were a plea to economists to accept that the world is more complex than their macroeconomic policies had assumed. It was simply not enough to expect the market to develop the necessary institutions as the state withdrew from the economy. This has led to the World Bank taking a much more microeconomic approach, with an emphasis on reducing corruption and strengthening the institutions which guide market development. However, this is proving problematic in post Soviet states, as corruption has become so entrenched it is proving very difficult to combat.

Missing from these debates, however, is the role that geography played in the transition process. As shown above, Eastern European states and the former Soviet Union countries have experienced rather different outcomes as a result of shock therapy. One of the reasons for this, it is argued, is because the transition process failed to incorporate, or even acknowledge, the differing geographies of post socialist regions. Countries had differing levels of economic development before they entered the Soviet sphere of influence. These can be split into four groups (see **Figure 1**):

1. The states involved with Europe's Industrial Revolution, such as what are now known as the Czech Republic and Slovenia.
2. Those which lagged behind group one in the industrialization process, such as today's Hungary, Poland, and Slovakia.
3. Those which were still predominantly rural as they entered into the command economy, such as Belarus, Russia, and Ukraine.
4. The 'underdeveloped' regions of Central Asia.

The countries from the first two groups above had established some form of market economy prior to their subjugation into the socialist space. Thus after its demise they were able to revive existing legal frameworks, such as property rights, almost immediately thereby avoiding the vacuum that existed in groups 3 and 4. While such legislation was not ideal, it did provide a basis around which institutions, such as a banking code, could be developed. Therefore, these countries were able to adopt a market framework very quickly, whereas in Russia such institutions had to be developed from scratch. This would be difficult enough under normal circumstances but given the political turmoil, economic collapse, and the developing oligarchy who wished to steer institutional development to their benefit, it is no surprise that the process is still not complete.

It was also assumed that the Soviet sphere of influence was a homogenous economic and political space, with mediations from Moscow accepted universally across it. The reality was markedly different, with interpretations of the plans differing widely from the macro country level to the actions of individual plant managers and workers. Some countries were able to interpret state socialism in differing ways, for example, within Hungary some small scale private enterprises were permitted. Thus upon moving toward a market economy there was some entrepreneurial culture, as well as relevant legislation, which could be called upon. Furthermore, there were artificial divisions within the Soviet Union as sector planning meant that certain sectors of the economy were concentrated in individual republics. For example, Belarus was the site of a considerable amount of light industrial production and in Ukraine agricultural development was prioritized. Also, one industry towns were created, often based around one enterprise. Therefore, countries and regions simply did not have the infrastructure in place within which a market economy could easily develop. Geography also played an important role for the countries of East and Central Europe as the Soviet Union collapsed. These countries,

Figure 1 The geographies of economic legacies.

such as Poland and the Baltic States looked immediately toward Western Europe with the hope of joining the EU. This meant these states had institutional structures which they could work toward implementing.

Transition or Transformation?

The very notion of transition suggests single start and endpoints. Using this theoretical base, economists de veloped a singular policy approach to be applied in countries wishing to move toward a market economy. This assumes that all the regions are starting from the same base point and that policy advice will be mediated in a uniform manner. By ignoring individual geographies, those leading the reform process were unable to foresee the problems that individual countries would have. Such a catch all policy also ignores the fact that the market economy they were trying to implement only existed in textbooks. Every country has its own version of a market economy, for example, the Swedish model is very dif ferent to the German form which in turn differs signifi cantly from the Anglo American variations. Therefore, it can be seen that as well as differing start points, there are also a multitude of endpoints. Thus the question must be asked as to what point transition economies start from and at what point is the process complete?

Using the EBRD's scorecards, most post socialist states have completed the transition process. Therefore the reform process in Russia and Ukraine is considered by some to be complete as their economies are, in theory, privatized, liberalized, and open to international com petition; a popular view amongst economists. However, as the above has shown, macroeconomic indicators fail to reveal the economic complexity in these regions. Many social scientists working in this area dispute that the reforms are complete, arguing that the form of economic development that has occurred in these countries is very different to that envisaged by those who steered the re form process. If macro level differences were not in corporated into the transition process then clearly micro scale experiences could not be articulated within such strategies. As these are the spaces where the momentous political and economic changes were mediated, inter preted, experienced, and lived it is understandable that the results of the orthodoxy have seen widely different outcomes across the post Soviet sphere of influence.

The fact that there are so many outcomes problem atizes the term 'transition' even further. This has led to an increase in the use of the term 'transformation' when referring to these economies. Transformation suggests that there is not a singular starting point and that the process is fluid and the outcomes unclear. While this might seem like a semantic argument, this distinction demonstrates the diversity of the region and encourages institutional development that is relevant to specific countries. Fur thermore, it recognizes that the economic forms that might eventually develop in the region might be very different to those that were expected and that forcing them, through policy and financial assistance, to conform to a textbook market economy is counterproductive.

See also: Communist and Post-Communist Geographies; Economic Geography; Economics and Human Geography; Geopolitics; Globalization, Economic; Liberalism; Neoliberal Economic Strategies; Neoliberalism and Development; Poverty; Privatization; Structural Adjustment.

Further Reading

Åslund, A. (1995). *How Russia Became A Market Economy*. Washington, DC: Brookings Institution.
Berglof, E. and Roland, G. (eds.). *The Economics of Transition*. Basingstoke: Macmillan.
Bradshaw, M. and Stenning, A. (eds.) (2003). *East Central Europe and the Former Soviet Union: The Post Socialist Economies*. Harlow: Pearson/DARG Regional Development Series.
Burawoy, M., Krotov, P. and Lykina, T. (2000). Involution and destitution in capitalist Russia. *Ethnography* 1(1), 43 65.
Burawoy, M. and Verdery, K. (1999). *Uncertain Transition*. Lanham: Rowman & Littlefield Publishers.
Gros, D. and Steinherr, A. (1995). *Winds of Change: Economic Transition in Central and Eastern Europe*. London: Longman.
Gustafson, T. (1999). *Capitalism Russian Style*. Cambridge: Cambridge University Press.
Klein, L. and Pomer, M. (eds.) (2001). *The New Russia: Transition Gone Awry*. Palo Alto: Stanford University Press.
Lane, D. (2000). What kind of capitalism for Russia? A comparative analysis. *Communist and Post Communist Studies* 33(4), 485 504.
Lane, D. and Myant, M. (eds.) (2006). *Varieties of Capitalism in Post Communist Countries*. Basingstoke: Palgrave Macmillan.
Lavigne, M. (1999). *The Economics of Transition: From Socialist Economy to Market Economy*. Basingstoke: Macmillan.
Pickles, J. and Smith, A. (1998). *Theorising Transition: The Political Economy of Post Communist Transformation*. London: Routledge.
Smith, A. (2007). Articulating neo liberalism: Diverse economies and urban restructuring in post socialism. In Sheppard, E., Leitner, H. & Peck, J. (eds.) *Contesting Neoliberalism: The Urban Frontier*, pp 204 222. New York: Guilford.
Smith, A. and Stenning, A. (2006). Beyond household economies: Articulations and spaces of economic practice in post socialism. *Progress in Human Geography* 30(2), 190 213.
Stiglitz, J. (2003). *Globalization and Its Discontents*. London: Penguin.
Stone, R. (2002). *Lending Credibility: The International Monetary Fund and the Post Communist Transition*. Princeton, NJ: Princeton University Press.
Wedel, J. (1998). *Collision and Collusion: The Strange Case of Western Aid to Eastern Europe*. Basingstoke: Macmillan.
World Bank (2002). *Transition The First Ten Years: Analysis and Lessons for Eastern Europe and the Former Soviet Union*. Washington, DC: World Bank.

Relevant Websites

http://www.ebrd.com
European Bank of Reconstruction and Development.
http://www.imf.org
International Monetary Fund.
http://www.worldbank.org
World Bank.

Translation

F. M. Smith, University of Dundee, Dundee, UK

Glossary

Back Translation A technique employed to ensure concept equivalence in translations which involves translation of a text (such as a questionnaire) into the target language and reverse translation to the original language by another translator. Versions are then compared to identify aspects requiring further clarification.

Concept Equivalence A technique in translation which seeks to ensure equivalent values are being interrogated, for example, in a questionnaire survey delivered in different languages. Often combined with back translation.

Domesticating Translation The translation strategy dominant in technical and scientific circles where the translated text is produced with emphasis on transmission of meaning between the two languages and fluency for the target audience. Although widespread, it is critiqued for reinforcing ethnocentric readings of the translated material as it renders the translation process invisible and may close down the varieties of meanings and styles present in the original text.

Foreignizing Translation A translation strategy which highlights the translated nature of the text by retaining the original syntax and style as much as possible, or by including terms from the original, for example. By flagging the ethical and political choices made in translation, it can decenter the ethnocentrism of dominant translation strategies, which tend to make the act of translation invisible and may prioritize the cultural values of the target language.

Interpreting The rendering of speech from one language to another requiring consideration of the types of translation strategy used, the positionality of the researcher and the interpreter, and the ethics of language usage in particular contexts.

Translation The rendering of text from one language to another, requiring decisions about the relative emphasis on meaning and style, on fluency, fidelity to the original text, and freedom of translation.

Introduction

Although translation is an extensive practice globally, and geographers are extensively involved in translation in their research, geographical writing has given only limited attention to translation practices as a methodological issue. Translation practices are typically relegated to footnotes or ignored entirely in written research accounts. This relative invisibility is surprising given the extensive consideration in human geographical research of power relations, working with 'others', the positionality of the researcher and questions of representation. Paying more attention to the practices and the politics of translation is crucial to the development of critical translation practice in geography, making an important contribution to cross cultural research. Geographers need to explore the theoretical underpinnings of various approaches to translation alongside the practical choices made in diverse translation strategies. These include decisions about the balance between translation as a search for equivalence and as a meaning making process, and between domesticating or foreignizing approaches to translation. Reflection on the situated practices of translation, both in the field and in the analysis and representation of research, needs to take account of the agency of the translator and their relation to the researcher, the political implications of translation and the uneven cultural and institutional power relations influencing the creation and representation of geographical knowledges. This helps to make more visible the often invisible spaces of translation in human geography research.

Translation (rendering a written text from one language to another) and interpretation (rendering speech from one language to another) are skilled professional practices, essential to the functioning of cross cultural communication. They operate in myriad, often uncommented, everyday practices – among tourists, migrants and diaspora communities, as well as in law, business, diplomacy, or international development. Such is the significance of translation in popular cultural practices that Héctor Tobar describes bilingual Hispanic communities in the contemporary USA as the *Translation Nation*, while the Italian scholar, Umberto Eco goes so far as to argue that translation is the one 'true' language of the multinational and multilingual European Union. The impossibilities and limitations of translation are also reflected in popular culture (in Sophia Coppola's film, *Lost in Translation*, for example) while paranoia about the inability to translate 'foreign' language worlds and anxiety at the power of the interpreter is portrayed in Sydney Pollack's film, *The Interpreter*, and highlighted in cases of criminal prosecution of translators engaged by the legal teams of people accused of terrorist acts. Tied up with these processes are unequal geographies of cultural, economic, and political dominance which mark some languages as lesser in status than others. Thus, acts of translation are never simply about meaning and communication; they intersect with a broader politics of translation and are in turn subject to challenge and contestation.

Diverse approaches to translation wrestle with these challenges. Whether the translated text can or should adhere closely to the style or meaning of the original, and whether this can produce a text in the translated language which communicates meaning clearly to the target audience, has long been subject to debate. Writing in 1923, Walter Benjamin described the complex negotiation of the translator between 'fidelity' to the original and the necessity for 'freedom' of the translator from the original text in order to communicate those meanings fluently to the audience of the translated text. For geographers seeking critical approaches to translation practices, relevant concerns emerge in relation in data collection and field work (In which languages are interviews conducted?), data analysis (Are the interviews analyzed in the original language, or in a translation of these? Who carried out the translation, and what effect does this have on the meanings conveyed?), dissemination and publication (Are issues of translation mentioned, with ambiguities and difficulties explained, and insights gained?), and ethical and political issues (Are particular social groups excluded because of their language (in) ability? Are sensitive topics and concepts carefully dealt with?).

Geographers and other social scientists have begun to develop critical translation practices to address these issues, drawing on work from social theory and translation studies. This is outlined below before diverse practices around translation in quantitative and qualitative research methodologies are discussed. Finally, development of critical translation practices needs to be contextualized within the politics of translation by taking account of the uneven geographies of translation and the institutional practices of the representation of geographical knowledges.

Toward a Critical Theorization of Translation

Lawrence Venuti argues that the dominant approach to translation and interpretation in international professional contexts (including business, technical and scientific fields, and geographical research) is one which Bogusia Temple describes as a 'transmission' model of translation. This emphasizes the fidelity of the translation to the meaning of the original speaker or writer and also the need for fluency for the target audience. The role of the translator is to translate meaning into the target language as accurately as possible, without the process of translation adding or subtracting from the original text, while also rendering the meaning transparent in the target language. Professional codes of ethics for translators therefore emphasize the neutrality of the translator or interpreter, making a commitment to translate that which was said or written, without omission or addition, and to do so without the influence of personal views. The translator is professionally invisible and for their influence on the text to be visible is considered an indication of a lapse in professional competency. In research, this positions the researcher and the translator epistemologically as objective instruments where elimination of bias or any effect on the data collected, other than to facilitate comprehension, is the key concern. Broadly positivist in approach, these translation strategies view translators as ideally invisible: their function is to render the meaning of the original as closely and neutrally as possible into the target language in such a way as to make the translated text appear to be the same as the original without a translator's intervention. In social science contexts, this leads to translation being seen as a logistical challenge to be dealt with by employing competent translators or interpreters with adequate linguistic skills to allow as close as possible a transmission of meaning between language contexts.

Positivist approaches have, however, been subject to criticism from those advocating more 'critical' strategies of translation in geography and elsewhere. These critical approaches, broadly social constructivist in orientation, draw on postcolonial, post structuralist, and feminist critiques of translation from writers such as Derrida, Venuti, and Spivak. Critique centers around two themes – translation and meaning, and translation and agency – and situates practices of translation within a wider politics of translation.

In thinking about translation as the communication of meaning, critical approaches emphasize the challenges, ambiguities, limitations, and possibilities of translation between linguistic and cultural contexts, and the political implications of these practices. While there can be some level of simple equivalence between languages (*la chaise*, the chair, *der Stuhl*), the challenge of transporting or mapping meaning from one frame of language reference to another is seldom so straightforward. This is particularly the case with the kinds of concepts which may be involved in human geographical research. Terms such as nation, community, home, space, or place are not simple even in a single language such as English and the opportunities for diverse meanings and terms multiply when working cross culturally. While most translators (both positivist and social constructivist) debate the need to consider whether the approach to translation adopted is one which accepts source language or target language bias, and stress that no translation can avoid such bias in some form or another, critical translation differs to some extent from positivist approaches in challenging the assumption that this choice can be judged in terms of how invisible the process of translation remains and highlights instead the implications of different translation strategies.

In human geography, explicit concern with translation has been limited. However, some scholars have given consideration to this question of meaning and translation. Martin Müller demonstrates the difficulties of working across languages, particularly in key terms where there may be a very rich set of connotations in one cultural context, but which may be obscured in translation, or of which the researcher may not be aware. This may occur where the original term is much richer than that which can be conveyed in the target language, as he explores in relation to a number of Russian terms (*derzhava*, *vlast'*, and *sila*) which all translate in English as 'power'. Searching for an equivalence of meaning requires further clarification as each of these terms carries particular implications, especially in political contexts. Müller argues that greater considerations of these varied cultural connotations would enrich a number of studies on Russia and geopolitics published in English which gave little sense of the diverse notions of power being employed in the original texts. Nevertheless, he argues that even such concern fails to take account of the extent to which the collapse of meaning differences involved in translation constitutes a political moment when the range of meanings in the original language is elided by the choice of terms in the target language. Conventional translation, which may leave such choices uncommented, therefore is depoliticizing as it neutralizes translation as a seemingly invisible practice. Even where translators employ different terms to indicate different original terms, the politicized or ideological nature of terms may be lost in translation. The apparently neutral translation can become a hegemonic representation of the original, or may circulate independently of the original, while losing the political implications of the original. For researchers, this raises serious questions about working with material in the original or in translation, the interrogation of diverse and complex meanings, and the representation of the research material in papers and reports.

In a parallel argument, Venuti describes the 'neutral' approach to translation as a 'domesticating' approach to translation where 'fluency' and 'freedom' in translation are employed to enforce dominant canons around clarity of expression (making the act of translation at once invisible) and yet also allowing for changes to the original (freedom) to render the text readily interpretable within the target language. Freedom and fluency are judged in terms of the cultural and linguistic norms of the target language so that the domestication of the foreign text into the target language constitutes an act of violence. This is not to privilege the original text over the target one, but to indicate that the effect of this approach is to render the process of translation and the difference of the foreign text invisible, placing the cultural frame of reference of the target language as the dominant one while leaving its effect unquestioned. Given the dominance of English in global contexts, this tends also to reinforce the dominance of English language cultural norms. As an alternative, Venuti proposes 'foreignizing translation' strategies which aim to retain a sense of the difference of the foreign text, of cultural references and linguistic patterns foreign to the translation language. Likewise, for Müller, the issue is not to create 'better' translations through producing greater equivalence, but to destabilize and denaturalize the hegemony of the translated text. In this he argues fundamentally for the need for language training among geographers, since those who work only from already translated material will have little awareness of the issues raised above. However, even where the researcher works across languages, problems of representation remain. Possibilities such as leaving original terms untranslated and footnoting them, or deliberately working with the style of the original text as much as possible are all strategies for making the politics of translation more visible.

Giving attention to the politics of translation also entails discussion of the agency of the translator in a form which is usually elided from traditional approaches to translation. Instead, language abilities should be discussed alongside more commonly discussed aspects of positionality such as gender, class, race, sexuality, or caste while the translator's agency is recognized. The choices of the translator around particular translation strategies become part of a set of political choices in which translation is a meaning making process. Gayatri Chakravorty Spivak, discussing translating poems by Bengali women writers into English, describes how she starts by avoiding thinking about the audience for the translation and concentrates more on the meaning, and particularly the style and language of the original. She situates these as specific choices within the unequal language hierarchies between Bengali and English, in relation to a critique of postcolonial cultural practices, and in relation to the gender politics of the writers concerned. While this may seem to have less relevance in a social science context, in discussing her research on domestic workers in Kenya and Tanzania, Janet Bujra considers the implications of a very accurate, literal translation compared to a translation which gets at issues of slang, local references, or other themes. The latter may not necessarily produce a polished final text, but may give greater insight to the terms in which the research participants articulated their views.

Translation is therefore a meaning making process, while it also seeks to transmit meaning in the greatest possible equivalence between contexts. It operates in the tension between what Jacques Derrida described as both the necessity and impossibility of conveying fully the meaning and style of the original text and its wider cultural resonances in the translated text. However, al though something is always lost in translation, reflecting both the incommensurability of translation between two languages, once translation is viewed as a meaning making process, insights can be gained from close at tention to the ambiguities, disjunctures, neologisms, and hybrid meanings which emerge from translation. This allows reflection on the limitations of the target language and its cultural meanings, and on the plurality of meaning in the 'original' text, which is itself often not unproblematic. Thus, Fiona Smith reflects on how to translate terminology from a study of community activist groups in Germany where those involved in fact use terminology with greater emphasis on citizenship and civic values (*Bürgervereine*, *Bürger*) than in the English term 'community'. Crucially, this not only raised issues about translation, but also forced the denaturalization of the notion of 'community' activism in English. Working through moments of rupture or ambiguity, the researcher can identify themes requiring further interrogation, ex ploring, for example, what it means to use such terms for these social practices in the German context as well as how concepts from the English notion of 'community' might be implicated but worked out in different cultural or political registers. Considering the diversity of meanings in both source and target language gives rise to the possibility of translation as a hybrid form – neither wholly the original nor wholly separate from it, but working between and across language registers to high light the politics of translation and the politics of dis course in both language worlds.

Exploring Translation Practices in Human Geography Research

The themes outlined above are negotiated in diverse ways in the practices of translation in human geography research. Three sets of typical practices are considered here: questionnaire surveys; working with interpreters; and working as a bilingual researcher.

Questionnaire Surveys

Cross national or multilingual questionnaire surveys need to focus on identifying equivalent ideas, measures, and concepts between different cultural and linguistic contexts. A broadly positivist approach to translation is commonly advocated where translation needs to be both accurate and appropriate. As explored by Behling and Law, this typically involves the search for 'concept equivalence' or 'concept validity' between languages. A common strategy for ensuring this is to translate the questionnaire into the target language and then have another translator do a back translation from the target to the source language. The back translation and the ori ginal are compared to identify inconsistencies and am biguities, allowing meanings to be clarified and questions reformulated. Importance is placed on clear and un ambiguous translation and on the comparability of con cepts between languages.

Creating that equivalence of meaning requires cultural knowledge alongside language proficiency. A word for word translation may not be helpful if it produces am biguous, insensitive, or offensive translations. The original concept might need adjusted to a register appropriate for the target language, creating challenges in ensuring that equivalent concepts or qualities are being measured in diverse contexts. For example, a cross national study ask ing students to evaluate the role of professors in geo graphical education at university needed careful translation between American English (where 'professors' is commonly taken to be most academic teachers) and German (where 'professors' are a specific senior section of the academic staff). Thus while the term 'professor/Pro fessor' is straightforward to translate, concept equivalence is more difficult to achieve. Awareness of appropriate frames of reference helps to elicit information from re spondents in a manner relevant to their cultural terms. In the development studies context, David Simon emphasizes sensitivity to different calendars, to units of measurement, and to using significant local events as reference points. Thus, quantitative survey research requires as much sen sitivity to the politics and ethics of translation as all cross cultural research. Despite these difficulties, even a basic effort at translation can develop further understanding of the issues being researched, and of how language abilities are implicated in social relations. Fiona Smith outlines how, in using an English and a Spanish version of their questionnaire which allowed them to survey those with lower level English skills, British students investigating the local labor market in a Spanish tourist resort came to understand how language abilities allowed some people access to more highly paid employment, while others lacking these language skills were excluded from these jobs. The questionnaire itself is not the only crucial element in survey research. It is also necessary to consider the positionality (gender, class, language ability, and in sider or outsider status) of those administering the survey and their influence on the research, while attention needs to be paid to issues of recruitment and training of inter preters or those administering the questionnaire in the field. These issues are discussed in more detail below.

Working with Interpreters

Some writers suggest the ideal for cross cultural research is where the researcher speaks the language of the re search participants, taking bi lingualism as the 'gold standard' and suggesting the use of interpreters distances the researcher from the research. Certainly, cross cultural research benefits enormously from greater linguistic communication, but situations inevitably arise where the use of interpreters and translators is necessary, or even desirable. In this case, a number of practical and technical issues require consideration, as do the social relations of the interpreter to the research process.

Finding interpreters with suitable language skills, who are available and affordable, is a key challenge. While this may involve skilled professional translators, in practice it often utilizes local informants with a range of language skills. Actual choices depend to a great extent on funding levels and the time available for recruitment. Associated with the question of recruitment is the technical question of the kinds of translation or interpretation strategies which are possible or desirable. Researching domestic labor in Kenya and Tanzania, Janet Bujra explored a number of technical questions about translation strategies. What terms were appropriate to use to ask the respondents about the practices being discussed? The politics of naming diverse domestic labor practices and the social roles involved with them meant careful choices were made about the terminology being used. (Sensitivity to the impact of translation on research participants is explored further by Kevin Simms.) Bujra also decided about the kind of translation most suitable in that context, suggesting a rougher translation retaining local terminologies, and expression was preferable to rendering the talk into standard Swahili or English. In Venuti's terms, a 'foreignizing translation' was adopted, rather than one which domesticated expression and meaning to the target language context. Working with interpreters therefore ideally involves careful attention to recruitment and training, which should at a minimum involve discussion with interpreters of the purpose of the research, ethical issues, and the translation strategy preferred.

Thinking about the social relations of the interpreter to the research and the researchers involves discussion of the positionality of the interpreter and their agency in the research. Considering the positionality of the interpreter, Bujra and Simon both raise a series of issues. If interpreters are outsiders to the research context, will their translations be more objective, or tend to miss important issues of local interpretation? Are insider interpreters more likely to be able to act as gatekeepers to access local populations? Or might they introduce bias to the research by interviewing those they already know, or social groups more similar to themselves? In some particularly divided or factional contexts, using an interpreter closely associated with one group may reduce the chance of accessing other social groups in the research. Likewise, the gender, class, or age position of the interpreter may require consideration so as not to offend or exclude particular groups of research participants. Researchers should also understand that those able to act as interpreters may well be more affluent or educated than the research participants and should consider how this shapes the social relations of the research, rather than assuming interpretation is a neutral conduit of meaning and information between research participants and researcher. Allied to this are questions about what material is being collected, recorded, or analyzed in the research. Is it appropriate to record the whole translation process or will the researchers only record the answers translated by the interpreter? Either scenario raises questions about what is being analyzed (the views of the translator, or of the participants) and how. Is the interpreter engaged in the analysis and understood epistemologically as a participant in a meaning making process or viewed as an ideally invisible transmitter of information between researcher and participants?

Work by Twyman, Morrison, and Sporton on society–environment links in rangeland management in Botswana illustrates how a focus on the processes of meaning making within a research project can reveal key issues about how participants, interpreters, and researchers all negotiate meaning in a research setting. The researchers, from a British university, employed a number of skilled interpreters to allow them to interact with the wide range of languages used by the research participants. Rather than recording the answers to questions once they had been translated into English, the process of communication back and forwards between the research participants, including the translators, was taped and transcribed. The research team then investigated the whole research process as one of complex meaning making where it was very difficult to map the ideas and meanings of the research participants on to English language terms (and vice versa). This led the researchers to rework the concepts with which they were researching. Furthermore, a focus on the processes of interpretation and translation of meaning between research participants illustrated that many research participants were already speaking in what was not their first language. In their everyday negotiations around rangeland management, translation was part and parcel of the practices of the research participants as they worked between a variety of languages and concepts to communicate their attitudes toward and practices of livestock and land stewardship. Attention to the processes of meaning making in translation and interpretation can therefore provide a nuanced appreciation of how concepts are mediated and utilized in cross cultural contexts. It decenters the conceptual and linguistic assumptions of the researcher(s) and furthers understanding of how negotiations of meaning, translation, and interpretation play a key role in a range of socio spatial processes.

Working as a Bilingual Researcher

Arguably, working as a bilingual researcher, where the researcher speaks both languages fluently, does reduce the distancing effects of using interpreters and may result in the researcher being more aware of the challenges and choices involved in translation. However, bilingual re searchers cannot ignore decisions about translation strategies or the ethics and politics of translation. Kathrin Hörschelmann outlined her approach in research in the former east Germany which drew on qualitative, in depth interviews about the cultural and social consequences of German unification. Central to it is consideration of her own positionality and of her strategy in working between two languages (German and English). The interviews were conducted in her native German and carefully transcribed in German, then analyzed through reading, coding, and note taking. Only the quotations selected for use in the final paper were translated into English. She adopted a 'literal' approach to translating the quotations. This produced a rather 'artificial' construction resulting from the 'inevitable failure' of 'imitating German speech in written English' but it preserved as much as possible of the style and emphasis of the original interview extracts. This foreignizing strategy also involved retaining original terms holus bolus – the term *die Wende* (the change, or turning point) was not translated but was retained in German to indicate the highly politicized and contested nature of this term which refers to the processes of rupture and transformation in Germany in 1989/90. Hörschelmann notes how the research involved shifts between various forms (conversations to selected quota tions, spoken speech to written text, as well as a shift from German to English). As a result, the whole research process entails diverse translations, not only in the narrow linguistic sense but also in a wider sense. Informed by actor network theory, it involved an attempt to preserve and communicate meaning but also a loss of contextual and nonverbal information which was compensated for to some extent by the discussion of the texts and inclusion of wider contextual meanings and implications.

For bilingual researchers, foreignizing translation strategies may help to highlight and represent diverse and divergent meanings. Academic writing adopts diverse textual practices including parallel language texts, ex tensive use of footnotes, keeping the register of the ori ginal language, or use of the original terms mixed in with the translation. Using two or more languages without translation of every line is another strategy, used by Altha Cravey in her research on Mexican communities in the USA. Analyzing translation as meaning making, reflecting carefully on the translation strategies entailed in bilingual research and its representation decenters the ethno centrism of the research process and signposts the pol itical moments involved in all language usage.

The Politics of Translation

As is evident from the above discussion, the politics of translation relate fundamentally to questions of meaning and terminology and to the positionality of the re searcher and/or translator. However, it is necessary to add a further context within which the politics of trans lation must be considered, namely that of language hierarchies and the structural conditions of knowledge production and circulation. A contemporary example from Nina Laurie *et al.* demonstrates how the linguistic politics of working between indigenous languages and cultural values within Latin America is often highly contested and intertwined with transnational strategies of development, the professionalization of indigenous groups, and the roles and values of development funders and of state agencies. Thus, translation and intercultural practices are entwined with wider uneven socio spatial relations.

Of particular relevance here are the structural con ditions under which geographical knowledges are pro duced and circulated. Commercial translation of literature translates many more titles from English into other lan guages than vice versa, reflecting and reinforcing the dominance of English language cultural production at a global scale. In technical and scientific publishing, in cluding academic geography, the trend is even more strongly toward publication in English. This creates its own tensions and difficulties for researchers and students who find themselves working always in translation as they study and produce geographical knowledges. This prob lematic positioning of geographical work conducted in languages other than English in relation to academic publishing and professional networks, and the dominance of 'Anglo American geography' have been raised by a number of geographers, such as Maria Dolors Garcia Ramon. Developing geographical knowledges, reading and writing in translation and working with the often unstated normative assumptions of English language geographical knowledges and practices, all serve to shape particular practices of geographical knowledge production and to center or marginalize particular experiences and know ledges. The implications for academic researchers are discussed by Helms *et al.*, and Desforges and Jones discuss the experiences of bilingually educated students. Given the dominance of English as a published academic language around the globe, this raises questions of how geographical knowledges are produced and represented, since even those who pay attention to translation have not often produced bilingual (let alone polyvocal) texts, although a number of journals have now made efforts toward publishing at least translations of article abstracts. In this context, it is ironic that this article is written in English. However, the work discussed in this article seeks to take seriously questions of translation and interpretation, and to situate this within the broader pol itics of language hierarchies. It challenges ethnocentric assumptions about the invisibility of cross cultural com munication and stands alongside discussions of the range of geographical research and knowledge produced globally in contexts and languages other than English. For geog raphers, the spatiality of translation and of the politics of language hierarchies at a range of scales from the local to the global are of significance. Thus, questions of cultural politics, situated knowledges, positionality, representation, and the construction of geographical knowledges are all intimately related to the strategies of translation adopted in geographical research and analysis.

See also: Cross-Cultural Research; Fieldwork; Globalization, Cultural; Language and Research; Polyvocality.

Further Reading

Behling, O. and Law, K. (2000). *Translating Questionnaires and Other Research Instruments: Problems and Solutions*. Thousand Oaks, CA: Sage.

Bujra, J. (2006). Lost in translation? The use of interpreters in fieldwork. In Desai, V. & Potter, R. B. (eds.) *Doing Development Research*, pp 172 179. London: Sage.

Cravey, A. (2003). Toque una ranchera, por favour. *Antipode* 35, 603 621.

Desforges, L. and Jones, R. (2001). Bilingualism and geographical knowledge: A case study of students at the University of Wales, Aberystwyth. *Social and Cultural Geography* 2, 333 346.

Eco, U. (1993). *La ricerca della lingua perfetta nella cultura europea*. Rome: Laterza.

Garcia Ramon, M. D. (2003). Globalization and international geography: The questions of languages and scholarly traditions. *Progress in Human Geography* 27(1), 1 5.

Helms, G., Lossau, J. and Oslender, U. (2005). *Einfach sprachlos* but not simply speechless: Language(s), thought and practice in the social sciences. *Area* 37, 242 250.

Horschelmann, K. (2002). History after the end: Post socialist difference in a (post) modern world. *Transactions of the Institute of British Geographers* 27, 52 66.

Laurie, N., Andolina, R. and Radcliffe, S. (2003). Indigenous professionalization: Transnational social reproduction in the Andes. *Antipode* 35, 463 491.

Muller, M. (2007). What's in a word? Problematizing translation between languages. *Area* 39, 206 213.

Simms, K. (ed.) (1997). *Translating Sensitive Texts*. Amsterdam: Rodopi.

Simon, D. (2006). Your questions answered? Conducting questionnaire surveys. In Desai, V. & Potter, R. B. (eds.) *Doing Development Research*, pp 163 171. London: Sage.

Smith, F. M. (1996). Problematizing language: Limitations and possibilities in 'foreign language' research. *Area* 28, 160 166.

Smith, F. M. (2003). Working in other cultures. In Valentine, G. & Clifford, N. (eds.) *Key Methods in Geography*, pp 179 193. London: Sage.

Temple, B. (2002). Crossed wires: Interpreters, translators, and bilingual workers in cross language research. *Qualitative Health Research* 12, 844 854.

Tobar, H. (2005). *Translation Nation: Defining a New American Identity in the Spanish Speaking United States*. New York: Riverhead Books.

Twyman, C., Morrisson, J. and Sporton, D. (1999). The final fifth: Autobiography, reflexivity and interpretation in cross cultural research. *Area* 31, 313 326.

Venuti, L. (ed.) (2000). *The Translation Studies Reader*. London: Routledge.

Relevant Websites

http://www.languageline.com
Language Line Services: Interpreter Code of Ethics.

http://www.languagemarketplace.com
Language Marketplace: Code of Ethics for Interpreters and Translators.

http://translationjournal.net
Translation Journal (online journal for translators about translation issues).

Transnational Corporations in Developing Countries

C. Berndt, Johann Wolfgang Goethe-Universitaet Frankfurt, Frankfurt am Main, Germany

Glossary

Equity Capital Equity capital is the foreign direct investor's purchase of shares of an enterprise in a country other than that of its residence.

FDI Inflows and Outflows FDI inflows and outflows comprise capital provided (either directly or through other related enterprises) by a foreign direct investor to an FDI enterprise, or capital received by a foreign direct investor from an FDI enterprise.

FDI Stock FDI stock is the value of the share of their capital and reserves (including retained profits) attributable to the parent enterprise, plus the net indebtedness of affiliates to the parent enterprises.

Foreign Direct Investment (FDI) The components of FDI are equity capital, reinvested earnings, and other capital (mainly intracompany loans). Countries do not always collect data for each of these components; reported data on FDI are therefore not fully comparable across countries.

Intracompany Loans Intracompany loans or intracompany debt transactions refer to short- or long-term borrowing and lending of funds between direct investors (parent enterprises) and affiliate enterprises.

Reinvested Earnings Reinvested earnings comprise the direct investor's share (in proportion to direct equity participation) of earnings not distributed as dividends by affiliates or earnings not remitted to the direct investor. Such retained profits by affiliates are reinvested.

Introduction

> A transnational corporation is a firm that has the power to coordinate and control operations in more than one country, 'even if it does not own them'.
>
> (Dicken, 2003: 198; author's emphasis).

Scholars across the disciplines can subscribe to this broad definition which hints at important changes since the academic discussion of the multinational/transnational corporation (MNC/TNC) started in the 1950s and 1960s. First, this refers to the question of exclusive ownership of assets. While it was crucial for firms to possess the capacity to internally produce and organize proprietary assets and match these to market needs, the emphasis today is less on monopoly possession of scarce and unique resources than on the capacity to organize knowledge flows on a global scale.

Second and following from this, there is the question of how firms organize their core competencies. In a global economy, firms build complex webs of alliances, engage in collaborative projects, and constantly reinvent them selves, investing and divesting at ever more breathtaking speed. TNCs are responding to major upheavals which have occurred during the 1980s and 1990s, most notably the rise of the informational economy, the increasing role of knowledge in determining the competitive position of firms, the integration of international financial markets, and the continuing opening of national markets.

Third and most importantly for the purpose of this article, these processes are inherently geographical, not so much because they take place somewhere, but because TNCs produce complex geographies, entangling and disentangling places at great distance with profound implications for our lives. In this context it is often argued that the crucial aspect of TNC power is the ability to take advantage of geographical differences. To this, one has to add the performative capacity to reproduce these differences, an aspect which is particularly salient with regard to relations between the Global North and the Global South.

The Changing Geography of the TNC

As with any aspect of globalization, the organization of cross border production networks involves complex and often contradictory processes. Two dimensions can be distinguished here. The first concerns the centralization/decentralization dichotomy. There is a discussion in the literature which assigns 'global' status only to those firms that grant peripheral sites little autonomy and centralize knowledge and core competences. This is opposed to more decentralized blueprints where worldwide operations are organized as a portfolio of national businesses. Globalization appears to imply centralization and limited autonomy ever more for peripheral sites in corporate networks. Yet, if one looks more closely at TNCs, the distinction is far less straightforward. A number of activities may have to be standardized across different sites in the TNC network, for instance, the creation and availability of core knowledge or the formulation of a common corporate vision and mission. Yet other areas may require more sensitivity towards local difference. These include different consumer preferences, state regulations and laws, and what is often referred to as cultural differences: lifestyles, work practices, political cultures, etc. It is precisely this complexity of the contemporary TNC which demands careful consideration, above all in a North–South context.

The second divide in the literature revolves around the benefits accruing to TNCs when they invest in locations abroad. TNCs are generally believed to tap into the comparative advantages offered by regions and nation states. Scholars such as Michael Porter argue that the role of place specific assets has profoundly changed since the first wave of globalization, where companies predominantly pursued benefits by shifting activities to low cost locations. Regarding global sourcing only as a second best solution compared to the competitive advantages which arise from concentrations of highly specialized skills and knowledge, institutions, rivals, and sophisticated customers in a particular nation or region, Porter (1998) advances the localized cluster as the main locational asset for regions intending to attract FDI in today's global economy.

Discussions like these stress the virtues of stability and spatial fixity. For it is the ability to stabilize the flows of capital, goods, ideas, and people which distinguishes a successful region in the global economy. But it is equally important to take account of the crucial role of mobility in sustaining the corporate network and in providing the sort of environment in which TNCs are able to reap the benefits of their global strategies. Seen from a North–South perspective in particular, and this is extended further below, this implies, first, that sites and places of production are made by those movements and, second, that these places obtain their value for the TNC in relation to other sites in the corporate network.

Such a position offers an approach to the ongoing renegotiation of the spatiality of economic relations which avoids some of the pitfalls of a scalar perspective. First, this concerns the stylized opposition between advocates of all encompassing homogenization and those who insist that national or regional practices continue to diverge in the global economy. The second problem is the relationship between national and regional differences and management practices. According to the widespread culturalist logic, corporate decision makers setting up firms in other national contexts are faced with different institutions and cultures and are seen as having to adapt to or change these 'ways of doing things'. The idea of relatively stable 'host' and 'guest' cultures being involved in cultural struggles over adaptation and assimilation underlying this logic has been criticized for its neglect of translocal forces and hybrid forms.

From an alternative perspective, TNCs are therefore conceptualized as highly specific organizational setups which often develop spontaneously; as hybrids, which combine elements of what is more commonly termed global, multinational, and international, or ethnocentric, polycentric, and geocentric model of TNC organization, respectively. Most importantly, the stress is less on quantification (e.g., FDI data, numbers of units, and employees abroad), but rather on connectivity and interrelatedness between different sites in TNC networks and the implications this may have for the Global South.

TNCs and the Global South

In his contribution to a book edited by Jagdish Bhagwati in the early 1970s, Stephen Hymer (1972) asked about the role of the MNC at the end of the twentieth century. Arguing that it represented "an important step forward over previous methods of organizing international exchange" by bringing about an ever more extensive and productive international division of labor, he predicted the continuing expansion of the multinational form of industrial organization as a key element of global capitalism.

Indeed, corporate cross border activities have increased dramatically. The worldwide inward stock of cross border FDI, for instance, has risen 18 fold from 561 billion USD in 1980 to 10 129 billion USD in 2005 (Source: UNCTAD 2006). Of course, the global distribution of FDI has always been a highly uneven affair. Over the 15 year period 1990–2005, developed countries attracted almost 70% of FDI inflows, developing countries and transition economies accounting for 28% and 2%, respectively. What is more, if one breaks the data down to 5 year intervals the overall share of so called developing countries has fallen, from 30.9% (1990–94) to 25.7% (2000–04; see **Table 1** and **Figure 1**). And the cartography of global FDI flows gets even more unbalanced in the latter group, if the disproportionate share of countries such as China, Mexico, or Brazil and the almost negligible flow to African countries are taken into account (**Figure 2**). These disparities are slightly offset by the growing role of southern TNCs and FDI originating in the Global South, but the noticeable growth of southern global players such as Lenovo or Tata in recent years does not change the overall picture.

There has always been considerable controversy about the precise role of FDI from the perspective of the Global South. The underlying arguments do not seem to have changed much during the last 50 years or so. In a recent special issue of *World Economy,* Jagdish Bhagwati (2007) writes a robust defense against what he sees as oversimplified criticism. On the one hand, he refutes claims of tax competition and a 'race to the bottom', downplays the power of large TNCs in smaller countries, comments positively on the role of global media as guardians against political intrusion by foreign TNCs, and puts claims of systematic exploitation of workers, payment of low wages and sweatshop conditions into doubt. On the other hand, he stresses external economies to the sites of investment, the spillover of know how and

Table 1 Global FDI inward flows according to geographical regions, 1990 2005

	1990 2005		*1990 1994*		*1995 1999*		*2000 2004*	
	Billion USD	*%*	*Billion USD*	*%*	*Billion USD*	*%*	*Billion USD*	*%*
Developed economies	633 376 720	69.72	68 733 038	68.37	211 987 762	69.86	297 063 217	71.96
Developing economies	256 126 762	28.19	31 012 086	30.85	87 020 879	28.68	106 026 772	25.68
Economies in transition	18 914 498	2.08	786 042	0.78	4 433 444	1.46	9 727 076	2.36
Total	908 417 980	100	100 531 166	100	303 442 085	100	412 817 066	100

Source: UNCTAD, 2006.

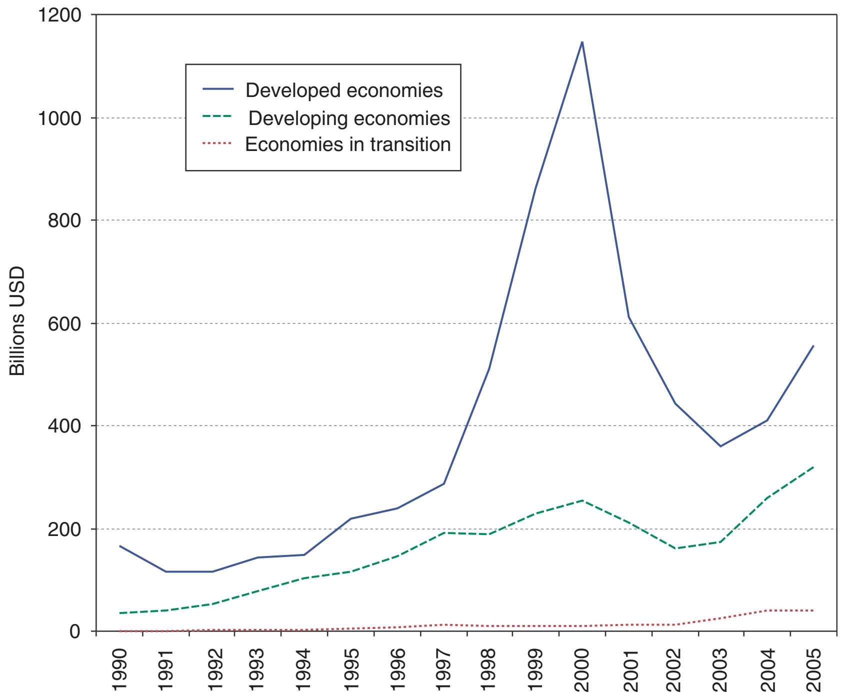

Figure 1 Global FDI inward flows according to geographical regions, 1990 2005. Reproduced from UNCTAD, 2006.

technology as well as the learning effect for workers hitherto not familiar with modern industrial production. He reminds readers of the crucial role of domestic elites, putting much of the blame for the shortcomings of FDI led growth strategies on endogenous problems and stressing the necessity of supportive state policies.

These arguments are well known to students of 1950s and 1960s development theory. Modernization theory rose to prominence during roughly the same time when the vertically integrated MNC came to be regarded as the dominant form of the business enterprise. It was little wonder therefore that scholars, at that time, looked at the large industrial corporation as the catalyst of development processes. In line with the dominant economic paradigm of that time, hopes were predominantly directed at the establishment of domestic industries, but scholars such as Walt W. Rostow always took the international dimension seriously, for instance, the benefits of FDI inflows to a developing country's capital stock or of technological spillovers to the domestic sector. Following the neoliberal paradigm shift, direct investment by MNCs/TNCs moved fully at center stage and the import substitution model was replaced by export led industrialization.

From the 1950s and 1960s, when dependency theorists attacked the ideas of Rostow and others, until today, critics have advanced two lines of argument. The first may be termed 'bypass problem' and refers to the widely acknowledged fact that large parts of the globe remain blind spots on the global map of corporate investment decisions. The case in point is Africa which received only 7.4% of the FDI attributed to developing countries in 2005.

The second and more interesting argument for the purpose of this article may be termed 'exploitation problem' and refers to problems which arise because of the presence of foreign investment. Dependency theorists argued that international capital produced export enclaves in host countries, created few external economies, and had minimal effect on national economies as a whole. Accordingly, TNC investment only intensified the

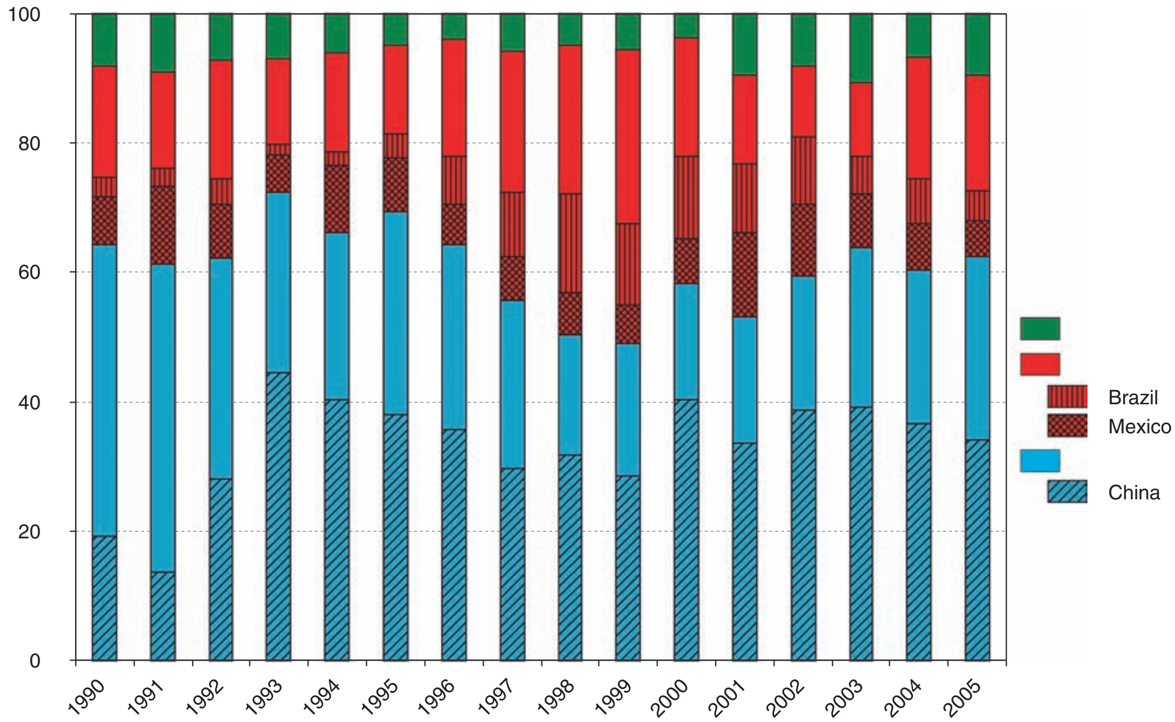

Figure 2 Developing economies, FDI inward flows according to geographical regions and key receiving countries, 1990 2005. Reproduced from UNCTAD, 2006.

structural heterogeneity of the economies, corrupted domestic elites and further exacerbated the financial and technological dependence on the North. In the 1970s, critics developed these arguments further and pointed to the role of TNCs in shaping and exploiting an inherently uneven geography. At the forefront of this stance was Stephen Hymer (1972), who predicted the continuing expansion of the multinational form of industrial organ ization ("law of increasing firm size"). In addition to this, he identified the hierarchical division of labor between geographical regions, which corresponds to the vertical division of labor within the firm, as a key for the uneven character of the modern world economy ("law of uneven development"). This was referred to as new international division of labor between company headquarters in the North and different types of offshore 'global factories' in the South.

In sum, for critical geographers in particular, the TNC/MNC epitomized the contradiction between physical fixity and mobility at the heart of the uneven geography of capitalism. Scholars such as David Harvey or Neil Smith regarded TNCs decision makers as powerful prisoners in a deeply ambivalent game, fever ishly searching for more advantageous conditions of production while contributing to geographical equal ization at the same time (Smith, 1989). For these critics, the ongoing search for geographical solutions to the capitalist dilemma, that is the 'spatial fix' (Harvey, 1981), on the part of transnational capital put different regions on and off the map, promising individual countries modernization and progress, only to drop them like a hot potato shortly afterwards. Or as Hymer (1972: 133) put it: "As [the TNC] crosses international boundaries, it pulls and tears at the social and political fabric and erodes the cohesiveness of national states. [...] It creates hierarchy rather than equality, and it spreads its benefits unequally."

The role of the state, critically commented upon in this quote by Hymer, has always been a major issue in the TNC literature. And the extent to which southern gov ernments are able to influence investment decisions and to ensure that TNC investment in the Global South has positive effects continues to spark controversial dis cussions. In a stylized way, positions may again be traced back to discussions during the heyday of development economics. On the one hand are those who point to the decisive role of state policies in providing the sort of regulatory environment which allows for longer term benefits of TNC activities and which differentiates the winners from the losers in the competition for northern FDI. Following the positive performance of Southeast Asian economies, emphasis is put on investments into national education systems and the research and devel opment (R&D) infrastructure as preconditions for the knowledge spillovers associated with 'export promoting' or 'outward oriented' trade strategies. Current emphasis on 'good governance' by multilateral development organizations is a variant of this argument (Shah, 2006). Neoliberal apologists of TNC investment often give this argument a different twist and blame “bad domestic policies for harmful effects” of FDI (Bhagwati, 2007: 221–222).

One sided positions like these have regularly provoked opposition from more critical scholars, Hymer (1972), for instance, arguing in similarly ideological vein that MNCs “weaken political control because they span many countries and can escape national regulation” (p. 127) and “reduce the sovereignty of all nation states” (p. 129). This was also the position of *dependistas* such as Osvaldo Sunkel, Theotonio dos Santos, or Andre Gunder Frank. However, Frank’s (1996: 34) admission to have “perhaps underestimated [the East Asian tiger economies’] capacity for technological upgrading and new participation in the international division of labor” reflects the gradual emergence of less ideological and unforgiving positions during the 1990s. As a consequence of this, the debate on the relationship between the (Northern) TNC and the (Southern) state has become more differentiated and sophisticated. Crucially, this also includes a more self critical tone in parts of the mainstream literature: Take the recent exchange about the implications of offshore outsourcing for northern and southern economies between Paul Samuelson on the one side and his former students Jagdish Bhagwati and Gregory Mankiw on the other, or the Stiglitz–Williamson debate on the Washington Consensus, which includes the liberalization of FDI inflows amongst its ten propositions.

It is obvious that there are a number of factors that intervene differently in a state’s capacity to positively influence the effects of TNC investments: the size of the country in question, strategic interest and motivation of TNC activities, economic structure, history, etc. In this context China is often cited as an example of the extent to which governments in the Global South have bargaining power and are able to dictate their terms of investment. China’s sheer size and political power notwithstanding, recent debates about the stalled implementation of a stricter labor law and the successful pressure against even basic labor rights exerted by northern capital (i.e., US companies via the American Chamber of Commerce) speaks a more sobering language. It is often overlooked that northern states and northern TNCs act in collusion, pushing through their interests against southern actors by forming “protection networks” as Immanuel Wallerstein (2005) would have it. This serves as a reminder that one should be wary of one sided arguments in the rugged terrain of state TNC relations.

Performing Production Systems and Market b/Orders

The transnational corporation is the driving force behind the configuration of global commodity chains (GCCs) and global value chains (GVCs). As outsourcing and offshoring processes gained momentum during the 1980s and 1990s, cross border trade in intermediate products (inter and intrafirm) has risen enormously. This was enabled by the global dispersion of knowledge and capabilities, by the establishment of new transportation and communication technologies, and by specialization in modular production networks. The crucial question from the point of view of southern actors is whether they are able to improve their position within far flung global production systems, a process referred to as ‘upgrading’ in the literature.

Although constituting an important step forward, scholars argue that the GCC/GVC approach may still not go far enough. It is argued that assignment of more or less stable positions in the capitalist world system to individual production sites (i.e., belonging to the center, the semiperiphery, the periphery) does not chime well with the heterogeneous experiences of production sites in the North and the South. This can be illustrated with the example of a US production plant in northern Mexico, which is analyzed in more detail elsewhere (Berndt, 2003, 2004).

The plant is owned by a US corporation under the auspices of the Maquiladora program and produces electronic switches and control panels for the appliance and automotive industries. Production has taken place in the border city Ciudad Juárez since 1986. During recent years and in line with the experience of other Maquiladoras, the plant operated in a volatile environment, reflected in the constant fluctuations of production volume (between 30 and 40 million USD) and personnel (between 300 and 450). Since the mid 1990s, the plant has experienced a slow upgrading process, symbolized by the relocation of advanced production from sister plants in the US. This process culminated in the transfer of membrane switch production from the US to Mexico, which involves the application of surface mount technology (SMT). SMT allows the construction of electronic circuits by mounting components directly onto the surface of printed circuit boards. Stakeholders in the Mexican plant, managers as well as workers, regarded these changes as proof of modernization and a further step towards ‘first world’ production, while decision makers in affected US plants understandably felt peripheralized. The newly won self confidence, however, was soon withering again, when the company announced the opening of another production plant in Nanmien City in the Donguang region, China, in 1999 and started to move production away from Mexico. This included not only basic labor intensive but also more advanced technology. At the same time workers in Mexico were told that it was becoming increasingly difficult to compete with their Chinese colleagues, not only with regard to wages, but also because of better productivity and quality in Nanmien City. Hopes for catching up were further dashed when membrane switch production was suddenly sold to another company on quality and prod uctivity grounds in 2006. Yet at the same time, the Mexican plant received high tech automotive products from a sister plant in Raleigh, North Carolina, which was closed down and production redistributed between Mexico and China.

Economic change, whether analyzed at the plant level or at the level of regional or national economies, does not take place in a linear way with a clearly identifiable aim. Rather it is a chaotic process, occurring in leaps forward and backward, and being characterized by elements of chance. Yet, this does not mean that there is no logic and structure to the workings of TNCs in the Global South. In order to come to terms with this logic, it is paramount not to look at any location in isolation, whether it is the individual production plant, the subnational region, or the national economy. Downgrading and upgrading is a constant feature of sites in the North and the South, selective changes even dividing and fragmenting shop floors, cities, and whole countries in a way that it is in creasingly difficult to talk of homogenous entities. In this system, pressure is without any doubt exerted by the movements of whole plants from place to place. It is more important for decision makers, however, to wield lo cational choice and to maintain organizational flexibility. This refers to the ability to control and discipline indi vidual production units, plants, and respective locations in relation to other sites in the corporate production network rather than reverting to a 'strategy of torched earth'.

Differences in factor costs and endowments, state regulations, consumer taste, etc., play a crucial role in determining investment decisions. At the same time the inscription of difference presupposes a common frame of reference – of standards and norms which allow the classification of different parts of a system. Decisions over movements of goods, capital, and people in global pro duction systems therefore demand constant measurement and qualification. This includes practice disciplines and sociotechnologies such as benchmarking, transfer pricing, or quality management, and also attempts to implement social and environmental standards. This is why, from the point of view of managers, the value of any single pro duction site is always more than the sum of localized place specific assets, whether these are old style natural resources or cheap wages, or the trendier specialized skills and the tacit knowledge celebrated by the region alist literature.

Understood in this way, flexible cross border pro duction systems are both plurilocal and translocal at the same time. They are 'plurilocal' because they bring about dense networks and entanglements, connecting pro duction sites in the South and in the North, and relating each element within global production networks to every other. They are 'translocal' because any individual site in these networks assumes different qualities depending on which location it is put in relation to. With their actions decision makers therefore never only exploit existing differences, they actively produce and reproduce differ ence, and the specific geography associated with this (e.g., constructions of distance/proximity, removal and repro duction of borders). The result is a confusing and blurred picture, a b/order regime which is deeply contradictory, but which suits capital interests well.

Framing their work and everyday life from an out ward oriented global rather than an inward oriented national perspective, TNC executives are powerful dri vers of this b/ordering regime. Yet, as productive as it may be for the accumulation of profit, the global eco nomic order is not simply a product of intentional and strategic decisions on the part of TNC decision makers or other members of the transnational class. Without any doubt, executives, managers, and other TNC mobiles play a powerful role in shaping global production systems and markets. With their everyday practices they connect and disconnect, and format a flexible global economic geography. But they only rarely do so strategically. De cisions and actions more often are trial and error re sponses to global pressures – problems TNC managers obviously helped to create themselves. This provides a more realistic picture of what are in fact deeply contra dictory practices. At the same time this is also a more optimistic perspective. The insight that TNCs and their decision makers are not as powerful as many critics would have them and the acknowledgment of the con tradictions inherent in everyday business practices allow alternative representations and may open ways to contest an ultimately effective global division of labor.

Such a position, however, requires an approach that focuses on the social practices of the actors involved and the way managers, executives, and also workers perform complex market and production orders. This refers to the way they literally order production steps, isolating ac tivities and assigning them to different locations, calcu lating value added, quality, and productivity, and giving complex flows and networks quasi linear form. Seen from this angle, only the idea of linear, chain like production allows the assignment of more and less advanced pos itions to different nodes in the network. And only when there is difference between more developed and less developed links in a chain is it possible to define pro cesses of upgrading and downgrading. What remains out of sight are the contradictions and ambivalences, for in stance, the existence of deeply fragmented production realities which are transformed with breathtaking speed and involve confusing positional changes depending on the product and production process in question. This does not chime well with representations that ascribe plants in the US core, in Mexico semiperipheral, and in China peripheral positions in global production net works. It is in this way that the actors involved in global production networks practically perform linear value chains. As long as these are taken as granted and given, it is possible to veil the contradictions necessary for their reproduction (e.g., the extent to which clandestine movements of goods, people, or capital are an integral part of global production systems). It is a crucial task for economic geographers to lay open these hidden geog raphies of global TNC activities.

Conclusion

Roughly five decades after the phenomenon raised the attention of development economists, the MNC/TNC continues to be at center stage, if one intends to grasp the dynamics of the ever expanding division of labor that characterizes global network capitalism. Much of today's debate is still surprisingly close to the uncompromising stand off between generally sympathetic positions in line with classical modernization theory and critical scholars portraying the TNC as the key driver of inequality and uneven development at the global scale. This stalemate needs to be unlocked. To be sure, in the light of the sheer scale of TNC activities and the inherently uneven character of capitalism, it would be foolish for geog raphers not to continue to critically engage with these issues. However, given the complexity and fluidity of globalized modernity, it is paramount in my view to adopt a more careful and modest approach when dealing with the TNC in a development context. As powerful as the northern TNC may be, its key actors – the trans national bureaucrats managing its operations, the prac titioners of practice disciplines such as accounting, auditing, or benchmarking working for global business service firms, or the representatives of business schools disseminating the latest managerial gimmick to even the tiniest supplier in multi tier production networks – are themselves acting in a (self inflicted) environment of volatility and uncertainty, more often reacting to un intended consequences of their decisions rather than strategically shaping the world. To be sure, this insight makes little difference to the daily life of the 16 year old *operadora* in a maquiladora plant in Ciudad Juárez, subject to the extremely problematic 'translation' of global management tools, or the male migrant from rural China, working and living in a sweat shop in the Donguang region. However, investigations into the consequences of TNC activities in the Global South, regressive or pro gressive, need to include other actors, for instance the contradictory role of the state both in the Global South and the Global North, the self reproducing burgeoning nongovernmental organization and development aid sector, or the Janus faced northern consumer. This is the rationale behind the plea for a more differentiated and also modest approach informing this article.

See also: Development I; Development II; Global Commodity Chains; Global Production Networks; Globalization, Economic.

Further Reading

Amin, A. and Cohendet, P. (2004). *Architectures of knowledge: Firms, capabilities, and communities*. Oxford: Oxford University Press.

Bartlett, C. A. and Ghoshal, S. (1998). *Managing across borders: The transnational solution*. New York: Random House.

Berndt, C. (2003). El paso del norte… modernization utopias, othering and management practices in Mexico's maquiladora industry. *Antipode* 35(2), 264 285.

Berndt, C. (2004). *Globalisierungs Grenzen: Modernisierungsträume und Lebenswirklichkeiten in Nordmexiko*. Bielefeld: Transcript.

Bhagwati, J. (2007). Why multinationals help reduce poverty. *The World Economy* 30(2), 211 228.

Bhagwati, J., Panagariya, A. and Srinivasan, T. N. (2004). The muddles over outsourcing. *Journal of Economic Perspectives* 18(4), 93 114.

Dicken, P. (2003). *The global shift: Reshaping the global economic map in the 21st century, New York*. London: Guilford.

Dos Santos, T. (1996). Latin American underdevelopment: Past, present, future. In Chew, S. C. & Denemark, R. A. (eds.) *The underdevelopment of development: Essays in honor of Andre Gunder Frank*, pp 149 170. Thousand Oaks, CA: Sage.

Dunning, J. H. (2000). The eclectic paradigm as an envelope for economic and business theories of MNE activity. *International Business Review* 9, 163 190.

Frank, A. G. (1969). The development of underdevelopment. In Frank, A. G. (ed.) *Latin America: Underdevelopment and revolution*, pp 3 17. New York: Monthly Review Press.

Frank, A. G. (1996). The underdevelopment of development. In Chew, S. C. & Denemark, R. A. (eds.) *The underdevelopment of development: Essays in honor of Andre Gunder Frank*, pp 17 55. Thousand Oaks, CA: Sage.

Harvey, D. (1981). The spatial fix: Hegel, von Thunen, and Marx. *Antipode* 13, 1 12.

Hedlund, G. (1986). The hypermodern MNC: A heterarchy. *Human Resource Management* 25(1), 9 35.

Hymer, S. (1972). The multinational corporation and the law of uneven development. In Bhagwati, J. N. (ed.) *Economics and world order: From the 1970s to the 1990s*, pp 113 140. London: Macmillan.

Jones, A. (2005). Truly global corporations? Theorizing 'organizational globalization' in advanced business services. *Journal of Economic Geography* 5(2), 177 200.

Navaretti, G. B. and Venables, A. J. (2004). *Multinational firms in the world economy*. Princeton, NJ: Princeton University Press.

Pitelis, C. (ed.) (2006). Special Issue: Honoring the Work of Stephen Hymer. *International Business Review* 15(2).

Samuelson, P. A. (2004). Where Ricardo and Mill Rebut and confirm arguments of mainstream economists supporting globalization. *The Journal of Economic Perspectives* 18(3), 135 146.

Sapsford, D. and Greenaway, D. (eds.) (2007). Special Issue: Honoring the Work of Indian Economist Balasubramanyam V. N. *World Economy* 30(2).

Shah, A. (2006). *Local governance in developing countries*. Washington: World Bank Publications.

Sklair, L. (2001). *The transnational capitalist class*. Oxford: Blackwell.

Smith, N. (1989). Uneven development and location theory: Towards a synthesis. In Peet, R. & Smith, N. (eds.) *New models in geography: The political economy perspective*, pp 142 163. London: Unwin Hyman.

Stiglitz, J. E. (2004). The Post Washington Consensus Consensus, Initiative for Policy Dialogue Working Paper Series, Columbia University, November 2004.
Sunkel, O. (1972). Undervelopment in Latin America: Toward the year 2000. In Bhagwati, J. N. (ed.) *Economics and world order: From the 1970s to the 1990s*, pp 199 231. London: Macmillan.
Teece, D. J. (2006). Reflections on the Hymer thesis and the multinational enterprise. *International Business Review* 15(2), 124 139.
UNCTAD (2006). World investment report 2006 FDI Statistics (http://stats.unctad.org/FDI/ReportFolders/ReportFolders.aspx; 13 April 2007).
Wallerstein, I. (2005). Protection networks and commodity chains in the capitalist world economy. Paper given at the *Conference on Global Networks: Interdisciplinary Perspectives on Commodity Chains*, 13 14 May 2005, Yale University.
Williamson, J. (1999). What Should the Bank Think about the Washington Consensus, Paper prepared as a background to the World Bank's World Development Report 2000. Washington: Institute for International Economics (http://www.iie.com/publications/papers/williamson0799.htm; 11.2.04).

Relevant Websites

http://unctad.org
Comprehensive and up to date geographical FDI data from the UNCTAD World Investment Report; interactive.

http://www.globalpolicy.org
Part of the UN Global Policy Forum section Social and Economic Policy exploring how TNCs dominate the global economy and exert their influence over global policymaking.

http://www.corporatewatch.org.uk
Website of UK based organization (no affiliation with the above) providing detailed profiles of some of the world's largest corporations and overviews of each major industry sector.

http://www.corpwatch.org
Website of US based organization providing information on corporate violations of human rights, environmental crimes, fraud and corruption around the world.

http://www.csreurope.org
Website run by the European business network for corporate social responsibility with over 60 leading transnational corporations as members.

Transnational Elites

K. Olds, University of Wisconsin-Madison, Madison, WI, USA

Introduction

On 23 September 2007, the *New York Times* ran a story about the lives (and spaces) of a couple who should be known to most human geographers – Saskia Sassen and Richard Sennett (**Figure 1**). The story included this passage:

> Until this year, when Ms. Sassen gave up a chair at the University of Chicago to return to Columbia, the couple inhabited three "global cities" a term that happens to have been coined by Ms. Sassen (whose book "The Global City" was published in 1991). They would spend six months in London followed by six months divided between New York and Chicago. "It seems so much more reasonable to be in just two cities," Ms. Sassen said. "It's terribly cozy."
>
> Her most recent book, "Territory, Authority, Rights: From Medieval to Global Assemblages," was published by Princeton University Press last year.
>
> "I am nomadic," admitted Ms. Sassen, who was born in the Netherlands, grew up in Argentina and Italy and speaks six languages fluently. "I make myself at home wherever I go." For her, "home" means one thing: "a place where I can work."
>
> "When I go to giant conferences," she said, "I go to the press room. All I need is coffee, a computer table and people who will just leave you alone. My work is my anchor. It is how I keep the rhythm of my daily life."
>
> Mr. Sennett, meanwhile, prefers to write at a neat, trim desk tucked into one corner of the house's grandest space, a double height living room on the second floor that is illuminated by casement windows and a skylight as wide as the room.
>
> "I've spent a lot of my life in this room," Mr. Sennett said. "I've written all my books here, and I've had the experience for a long time of this light. And the silence, which is so rare."

The lives and practices associated with this couple are hardly ordinary; there is a huge disjuncture between the practices and technologies, and the social and symbolic capital, of Saskia Sassen and Richard Sennett, and those of the average person in New York or London, let alone the average person in Kalamazoo in the USA, Upper Slaughter in the UK, or Medan in Indonesia. Sassen and Sennett move across global space more frequently and at greater speeds; they contribute to the production of discourses that circulate across space and time, and in doing so they shape the nature of discursive fields not only in academia, but also in relation to select public policy debates; and they are affiliated with prominent and well resourced institutions (in their cases the London School of Economics and Political Science, New York University, and Columbia University), with global reach, that are based in global cities. In short Sassen and Sennett are 'transnational elites'.

Of course there are many types of transnational elites, and the *New York Times* could just as easily profiled the lives and spaces of the 26 year old Venezuelan conductor

Figure 1 Saskia Sassen and Richard Sennett, New York, September 2007. Source: *New York Times*, 23 September 2007.

Gustavo Dudamel (the new director of the Los Angeles Philharmonic Association), Eva Egron Polak (Secretary General of the International Association of Universities), Dipak Patel (the Zambian trade minister who is feted by numerous developing country trade negotiators for his insightful analysis of how the World Trade Organization (WTO) operates), or Tade Akin Aina (professor of sociology at CODESRIA, the University of Dar es Sa laam and the University of Cocody Abidjan, and former Ford Foundation officer). Sassen and Sennett are simply exemplars of what is really a very heterogeneous and multilayered transnational elite social formation.

In this article I explore some of the above themes, highlighting how human geographers and closely aligned analysts have made sense of transnational elites. Of course, transnational elites have existed for hundreds of years; one only has to think of, for example, the thir teenth and fourteenth century merchant explorers so insightfully analyzed in Janet Abu Lughod's now classic text *Before European Hegemony: The World System AD 1250–1350*. But due to space limitations, and the relatively recent focus on transnational elites in human geography, this article is of a contemporary nature, focusing in on the emergence of the concept in the 1980s and 1990s, its subsequent development, and some of the key contri butions that have been made in the literature.

The Evolving Nature of the Transnational Elite Concept in Economic Geography

While the term 'transnational' was utilized in the middle of the twentieth century, it was not until the 1990s that its usage became commonplace in the social sciences, in cluding in human geography and the subfield of eco nomic geography. The concept transnational emerged in the context of a broad structural transformation in the world space economy, as well as an associated analytical shift, to making sense of this transformation.

The structural shift that underlies the emergence of the transnational elite concept in human geography was technological and ideological change (i.e., liberalism) that enabled manufacturing firms, especially in the West, to stretch their production chains out across global space. In the process, various segments of the production system were progressively stretched out across space, with new activities laid out within new territories. Research and development (R&D), for example, was most often re tained in the West, though production and assembly were most often relocated to the so called developing (or in dustrializing) world. Clearly this is a process that needed, and still needs to be, surveyed and coordinated. Why? Because while capitalism is a dynamic force it has to be constituted, and reacted to, by firms – the primary 'mover and shakers' of the global economy.

Firms operating at a global scale were initially deemed multinational corporations (MNCs), though the term transnational corporation (TNC) is now the com monplace term of use, at least in economic geography. This discursive shift reflected the emergence of global scale 'flexible' forms of production, and improved understandings that global production chains were in creasingly 'governed', though not in formal 'vertical' sense. Nike's global production chain is the case exem plar, with the US based firm not owning one factory, yet responsible for the governance (via subcontracting re lations) of the practices of hundreds of thousands of workers in dozens of countries in Asia and Latin America. In short, TNCs are firms that can control and coordinate operations in multiple countries, though the firms may not formally own the overseas affiliates.

Regardless of whether we are speaking about pro ducer driven or buyer driven production chains, all forms of global production chains need to be coordinated and managed, and it this functional requirement that led to the emergence and use of the term 'transnational elite'. Transnational elites were first conceived of, then, as the people who bring global production chains into being, and who then shape their nature in a governance sense. On the basis of this definition the social formation in cluded not only chief executive officers (CEOs) and chief financial officers (CFOs), but also the upper echelon of senior management; the 'command and control' officers of the private sector dimension of the world economy.

In conjunction with the restructuring of the world economy, an analytical transformation took place, with the emergence and rapid dominance of political economy approaches in human geography. The Marxian influ enced political economy approach inspired research on the changing nature of the mode of production, capital–labor relations, and the spatial structure of the (world) economy.

In the 1980s and 1990s the spatiality of such trans national elites was examined, though really indirectly inferred via an analysis of the locational patterns asso ciated with MNCs. These types of studies typically mapped the global geographies of such firms, often using a variety of proxy measures. The most influential meas ure was the Fortune 500 list annual ranking of America's largest 500 companies. This measure was combined with a variety of other analytical tools, leading to the associ ation of MNCs and particular types of cities – world cities.

The very different work of University of California, Los Angeles (UCLA)'s John Friedmann was among the most influential of this era, leading to the identification of a 'world system of cities' in which MNCs, and therefore transnational elites, were based. John Friedmann and Goetz Wolff (1982: 310) noted that such world cities have become tightly "interconnected with each other through decision making and finance," and they now "constitute a world wide system of control over production and mar ket expansion." Cities of course, cannot control global production chains; it is the transnational elites working within such firms that do so.

Friedmann's work on the nature of the global system reverberates with research and writing by economic/ development geographers like T. G. McGee and War wick Armstrong who sought to make sense of how the transnational capitalist class shaped the world system (using dependency parlance) so as to extract resources from the Third World, creating 'theaters of accumulation' (aka cities). Armstrong and McGee, foreshadowing sub sequent developments in the emergence of the term, fixed the spatial locations of elites – both global and national – within cities. **Figure 2** nicely captures the structural logic of their argument.

Thus, 'strategic decision making' takes place in 'global centers', while 'tactical decision making' and 'political control' take place in 'national' metropoli. A functionalist take on culture is certainly evident in much of the ana lyses of the 1980s and 1990s, with these elites perceived to be 'disseminating' their 'cultural influence' below, to society.

In a relatively known narrative, the structuralist ana lytic of Friedmann *et al.* was critiqued and extended. People like Peter Taylor, but especially Saskia Sassen, the sociologist cum planner and honorary geographer, worked with a more institutionally aware framework to determine the nature and function of what she deemed

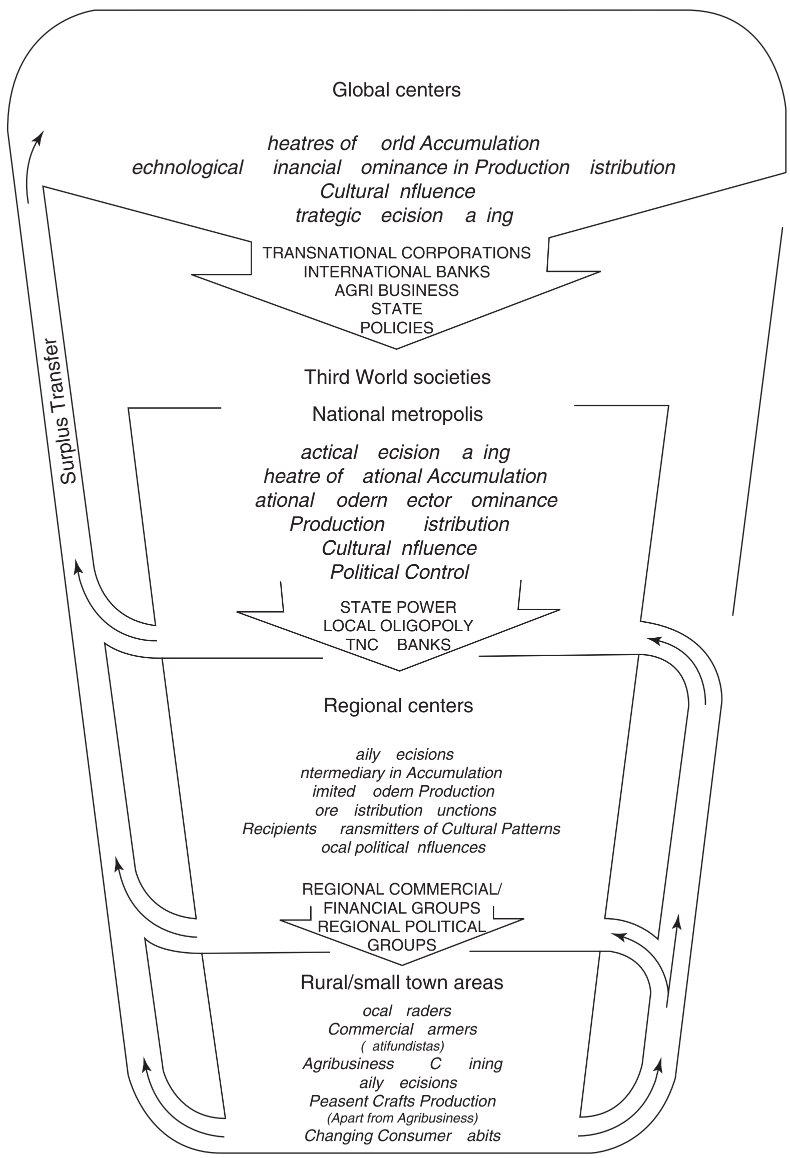

Figure 2 Hierarchical model of economic organization in world economic system. Source: Armstrong, W. and Mcgee, T. G. (1985). *Theatres of Accumulation*: *Studies in Asian and Latin American Urbanization*. London: Methuen. Reprinted by permission of authors.

'global cities'. Sassen's now obligatory reference to the global city is worth citing, though in this case note the ways that she infers that there is a complex of inter dependencies between knowledgeable people within and between geographically dispersed cities:

> [A] combination of spatial dispersal and global inte gration has created a new strategic role for major cities. Beyond their long history as centres for international trade and banking, these cities now function as centres in four new ways: first, as highly concentrated command points in the organization of the world economy; second, as key locations for finance and specialized service firms, which have replaced manufacturing as the leading eco nomic sectors; third, as sites of production of innovations, in these leading industries; and fourth, as markets for the products and innovations produced. These changes in the functioning of cities have had a massive impact upon both international economic activity and urban form: cities concentrate control over vast resources, while fi nance and specialized service industries have re structured the urban social and economic order. Thus, a new type of city has appeared. It is the global city. (Sassen, 2001: 3 4)

Her analysis, and the research that has been influ enced by it, identifies the socioeconomic foundations of the global city – the managers and producer service (e.g., accountancy, law, advertising, and political risk consult ants), workers. Many (not all) of these people are the transnational elites who do the 'work of globalization' in the firms that are based on the nodes where the global system is constructed and reconstructed on a real time basis.

The insights and gaps associated with work on MNCs and global cities spurred on two concurrent innovations with respect to transnational elites in economic geography.

First, analysts pursued a much broader framework, effectively opening up the black box of 'the' transnational elite. Thus, the social formation was viewed in a more wholistic or collective sense, with an awareness that: (1) the social field made up of transnational elites was viewed with greater sensitivity such that the myriad of transna tional elites that exist were identified and studied; and (2) the dependencies of transnational elites upon other social groups within the spatial units that they use as a base (which is most often global or globalizing cities).

Second, they focused on the unique features associ ated with particular types of transnational elites, in cluding the nature of sectorally specific labor markets (e.g., accountancy, law, architecture, and finance), or the unique features of transnational elites that depend upon nonmarket actors (e.g., academics and the media) for validation, legitimacy, and power.

Broadening Out and Focusing In

The analysis of elites has commonly recognized diversity. C. Wright Mill's classic work – *The Power Elite* – for ex ample, examined not only 'chief executives' and the 'corporate rich', but also 'warlords', the 'political direct orate', and celebrities. Such "professors of power and wealth and celebrity" have provided food for fodder in both academic and especially nonacademic outlets for ages. In human geography, though, few researchers have examined elites beyond those associated with firms and the polity.

In sociology, Leslie Slair's work is representative of this type of work. Sklair, now an emeritus professor at the LSE, put over a decade of work into developing the concept of the transnational capitalist class (TNCC), culminating in his 2001 book *The Transnational Capitalist Class*. In Sklair's view, which is partially derived out of Wallersteinian work on the world system, the global ization process has unfolded through the interdependent efforts of four 'fractions' of a coherent social formation:

1. those who own and/or and control the major TNCs and their local affiliates (corporate fraction);
2. globalizing politicians and bureaucrats (state fraction);
3. globalizing professionals (technical fraction); and
4. merchants and media (consumerist fraction).

In which way, though, is the TNCC transnational? Sklair notes that the TNCC is transnational in five core dimensions:

1. The economic interests of its members are increas ingly globally linked rather than exclusively local and national in origin.
2. The transnationalist capitalist class (TCC) seeks to exert economic control in the workplace, political control in domestic, international, and global politics, and culture ideology control in everyday life through specific forms of global competitive and consumerist rhetoric and practice.
3. Members of the TCC have outward oriented global rather than inward oriented local perspectives on most economic, political, and culture ideology issues.
4. Members of the TCC tend to share similar lifestyles, particularly patterns of higher education, and con sumption of luxury goods and services.
5. Finally, members of the TCC seek to project images of themselves as citizens of the world as well as of their places and/or countries of birth.

In this schema, transnational elites are characterized not only by class position, but also via the associated (it is assumed) capacity to exert near complete control over their foreseeable environment, their increasingly (it is assumed) nonlocal (or deterritorialized) disposition and imaginaries, and their reflexivity with respect to a sought after 'cosmopolitan' identity and reputation.

While this broadening out process helps shed more light on the nature of the 'movers and shakers' of broad structural transformations, it initially retained a some what distanced and 'top down' (or satellite like view) character. While such top down approaches to the pol itical economy of economic and urban change have been subsequently critiqued for their functionalist assump tions, and the limitations of this level of abstraction, it is important to note that such analysts laid important groundwork for the more nuanced exploration of trans national elites that followed.

In the development of analyses about globalization, TNCs, and the creation and governance of global net works, geographers gradually reframed their level of abstraction, and became more aware of the geographic ally and historically specific nature of transnational elites, of the nature of state–TNC relations, of the contin gencies associated with the construction of global net works, and of the critically important role of transnational political and policy elites in the de nationalization process. In short, a more process oriented and institutionally aware form of political–economic analysis emerged; one that rarely conceived of repre sentations such as that presented in **Figure 2**.

For example, Nigel Thrift's work provides a good example of the broadening out drive in work on trans national elites. Thrift's work has frequently examined what Peter Dicken calls the 'movers and shakers' of the global economy. Thrift (often with Michael Taylor and Dean Forbes) began one thread of his career in the 1980s and early 1990s career examining the nature of MNCs, including those active in the Pacific Basin. Again, while the focus is ostensibly on MNCs, the texts can be read as a contribution to making sense of the geographies of transnational elites. A conceptual shift away from polit ical economy toward what Thrift has deemed 'non representational theory' has led to a continued focus on transnational elites, though from a level of abstraction that pays much greater attention to the sociotechnical networks, practices, governing power, and governability, of such elites. For example, Thrift has sought to shed light on the emergence of 'fast capitalism', and globally influential management practices. Though he does not use the term transnational elite, his work effectively fo cuses on such elites. Take the case of globally active and highly ranked business schools. Thrift suggests that we need to think about business schools as part of the 'cultural circuit of capitalism' – the fluid and highly interdependent network made up of management con sultancies, management gurus, business media, and business schools – and their associated infrastructures and technologies. Thus, business schools, which train up new generations of transnational elites, and which employ some transnational elites, fundamentally depend upon an underlying media infrastructure that represents the schools, but that also governs them (especially via benchmarking and ranking schemes). Such an approach to transnational elites pays much closer attention to variegated nature of the transnational elite social for mation, and in the process we acquire a more processual understanding of the creation and governance of these elites versus treating them as ephemeral and all knowing masters of the universe.

Thrift's work resonates well with that of Saskia Sassen in her latest epic – *Territory, Authority, Rights: From Medieval to Global Assemblages* – which includes a focus on the emergence of what she calls four 'global classes'. These classes (and class is used in a loose designative sense) are:

1. transnational professionals and executives;
2. transnational networks of government officials (in cluding experts, judges, and immigration officials);
3. global civil society activists; and
4. disadvantaged low wage workers, "including members of transnational immigrant communities and house holds (Sassen, 2006: 298–299)."

She sees each of these classes as being interdependent with each other, and underlain by a "variety of economic, political and subjective structures." that are insti tutionalized and shape the activities of these transna tionals. She carries on:

> Global networks with variable degrees of formalization and institutionalization are evident for each class. Fur ther, contexts such as the global corporate economy or the international human rights regime also play critical roles in the proliferation of global networks. These and other globalizing dynamics have weakened the exclusive authority, both objective and subjective, of national states over people, their imaginaries, and their sense of be longing. This facilitates the entry of non state actors into international domains once exclusive to national states. Economic, political, and civic processes largely confined to the national sphere can now go global, even when this is only an imaginary or a subjective disposition rather than a daily reality for many of these actors. (Sassen, 2006: 299)

Saskia Sassen, in conjunction with those of like minded geographers such as Beaverstock, Larner, Ley, McDowell, Olds, Thrift, Yeoh and Chang, Waters Yeoh, Huang, and Lam, and Yeung – a relatively small group of people relative to the size of the subfield of economic geography, and the impact of transnational elites upon society and economy – have sought to develop more nuanced accounts of the geographically and historically specific nature of transnational elites. In *Territory, Authority, Rights: From Medieval to Global Assemblages* Sassen is resolute in arguing that all of these transnational social formations, including the elites, are “embedded, in often unexpected ways, in thick and localized environments: financial and business centers, national governments, the localized microstructures of daily civic life and struggles, and the translocal insertions of immigrants.” Kris Olds, for example, shed light on the complicated relationship between elite networks of architects (that he deemed the ‘global intelligence corps’, following Peter Rimmer’s earlier work) and an array of institutions situated in global cities than enable such architects to perform on two levels that simply have to be brought together for elites to become and/or maintain their elite status: the client–architect level, and the ‘intelligentsia’ architect level. Thus, architects classified as transnational elites need to work at both levels, acquiring contracts around the world, and also creating and disseminating ideas that are consumed in the ‘right’ institutions (e.g., Harvard’s School of Design) that are scattered in select countries.

Geographers have sought to also bring in a concern with the ways in which the state is governing transnational elites (vs. assuming they are all powerful). Wendy Larner’s work, for example, focuses on the development and effects of state led expatriate strategies for national development. In her 2007 paper she identifies the rationality and technologies associated with the development process; she sees transnational elites from New Zealand (her focus) as being enrolled in state projects, and in the process the state is constructing new categories of elites, and in doing so contributing to the construction of what we take for granted as ‘globalization’.

Other geographers have also sought to both broaden and deepen their frame of analysis with Jonathon Beaverstock’s work exemplary in this regard. Over the course of the last 16 years he has built up a voluminous, coherent, and insightful body of work on transnational elites (especially with respect to the producer services sector, and finance, law, and accountancy more specifically). Beaverstock, now based at the University of Nottingham, has shed light on the structural forces underlying the presence and power of such elites, the ways that they govern and are governed, the fundamentally important role of networks and institutions (from exchanges to clubs) to their *modus operandi*, and the role of ‘translocal’ cities in shaping the opportunities that such elites have should they choose to pursue them.

Concluding Comments

‘Elite’, ‘elites’, and ‘elitism’. Words that carry with them significantly varying baggage depending on what society and era one is speaking of. The ‘tall poppy’ cultures Australia and Canada have traditionally viewed elites with disdain if not outright hostility. At the other end of the spectrum, countries like France and Singapore have cultivated their elites, placing them on a pedestal, resting much hope that they would effectively ‘steer’ their countries through difficult challenges. Thus, the concept of ‘elite’ carries with it a value statement, foreshadowing the troubled and contested history of the concept of transnational elite, and perhaps (as it is noted later) the lacunae of research on transnational elites in economic geography (and human geography more generally).

‘Transnational elites’ are elites that engage in activities, including the production of discourses, that involve border crossings. Their bodies and ideas move across national borders on a regular basis, though each movement, and the totality of these movements, has a particular geography; a very uneven geography that is reflective of uneven development processes at a global scale.

Over the last three decades a small coterie of economic geographers, and several honorary geographers like John Friedmann, Aihwa Ong, Saskia Sassen, and Leslie Sklair, have sought to shed light on the nature of such transnational elites. Research approaches to such elites have shifted away from a more abstract form of political economy to a closer to the ground approach that has blended the insights derived out of a blend of approaches – institutionally focused political economy through to post structuralism, including Foucauldian inspired governmentality approaches. These approaches to transnational elites have partly drawn inspiration from social scientists interested in the ‘underbelly’, the excluded, the poorly paid, and often exploited migrant workers, many of whom are transnationals themselves. Analysts working in this area – an area we have paid scant attention to for reasons of brevity, have examined the increased presence and impacts of international migration, especially with respect to migrants who sought to retain involvement in ‘both’ sending and receiving nations; a transformation that coincided with ethnographic cum hermeneutic oriented analyses that were sensitive to the imaginaries and worldviews of relatively mobile but poor people.

In economic geography research has increasingly pointed out that there is an institutional political economy to the lives of transnational elites. Their status and power is fundamentally dependent upon a myriad of institutions – public, private, and nonprofit – that support these people and their network building activities. Transnational elites and their ideas could not cross space and time were it not for the state apparatus, universities, think tanks, clubs, nongovernmental organizations (NGOs), the media, and so on, that enable them to be creative in whatever it is that they do. Transnational elites thus depend upon these institutions to construct the sociotechnical networks that sustain them. But the spread of such institutions across space is also highly uneven, and many people as creative as Sassen and Sennett, whom we started this article with, have been stranded mid career, ground down, forgotten about, or even murdered, in less hospitable climes.

In closing it is worth deliberating about the paucity of work on elites in economic geography, and even more so in human geography. Human geographers have devoted far more attention to the excluded and marginalized than they have elites. This is in part because of problems of access (in a methodological sense), and also because of the 'pro gressive' politics of human geography, which supports a general inclination to practice 'critical geography'. Given this disposition critical human geographers have tended to focus on the nature of the marginalized, and the forces underlying their marginalization, and in doing so they have shed less light on some of the elites who are integrally involved in creating, directly and indirectly, the marginal ization process. To be sure, research on elites, and even more so transnational elites, is difficult, especially with respect to issues of access. However, the discipline cannot continue to effectively ignore the primary architects of the systems that we all have to adapt to, and seek our liveli hoods in, whatever forms of work and play we engage in.

See also: Financial Centers, International; Globalization, Economic; Transnationalism; Transnationality.

Further Reading

Abu Lughod, J. (1989). *Before European Hegemony: The World System A.D. 1250 1350*. Oxford: Oxford University Press.
Armstrong, W. and Mcgee, T. G. (1985). *Theatres of Accumulation: Studies in Asian and Latin American Urbanization*. London: Methuen.
Beaverstock, J. (1991). Skilled international migration an analysis of the geography of international secondments within large accountancy firms. *Environment and Planning A* 23(8), 1133 1146.
Beaverstock, J. (2004). 'Managing across borders': Knowledge management and expatriation in professional service legal firms. *Journal of Economic Geography* 4(2), 157 179.
Beaverstock, J. (2005). Transnational elites in the city: British highly skilled inter company transferees in New York City's financial district. *Journal of Ethnic and Migration Studies* 31(2), 245 268.
Brenner, N. and Keil, R. (eds.) (2006). *The Global Cities Reader*. London: Routledge.
Dicken, P. (2004). Geographers and 'globalization': (Yet) another missed boat? *Transactions of the Institute of British Geographers* 29(1), 5 26.
Dicken, P. (2007). *Global Shift: Mapping the Changing Contours of the World Economy* (5th edn.). London: Sage.
Friedmann, J. and Wolff, G. (1982). World city formation: An agenda for research and action. *International Journal of Urban and Regional Research* 6, 309 344.
Godfrey, B. and Zhou, Y. (1999). Ranking world cities: Multinational corporations and the global urban hierarchy. *Urban Geography* 20(3), 268 281.
Johnston, R. J. and Sidaway, J. D. (2004). *Geography and Geographers: Anglo American Human Geography since 1945* (6th edn.). London: Arnold.
Knox, P., Agnew, J. and McCarthy, L. (2003). *The Geography of the World Economy*. New York: Oxford University Press.
Larner, W. (2007). Expatriate experts and globalizing governmentalities: The New Zealand diaspora strategy. *Transactions of the Institute of British Geographers* 32, 331 345.
Levitt, P. and Jaworsky, B. N. (2007). Transnational migration studies: Past developments and future trends. *Annual Review of Sociology* 33, 129 156.
Ley, D. (2003). Seeking homo economicus: The Canadian state and the strange story of the business immigration program. *Annals of the Association of American Geographers* 93(2), 426 441.
McDowell, L. (1997). *Capital Culture: Gender at Work in the City*. Malden, MA: Blackwell.
Olds, K. (2001). *Globalization and Urban Change: Capital, Culture and Pacific Rim Mega Projects*. Oxford: Oxford University Press.
Olds, K. and Yeung, H. W c. (2004). Pathways to global city formation: A view from the developmental city state of Singapore. *Review of International Political Economy* 11(3), 489 521.
Peet, R. and Thrift, N. (eds.) (1989). *New Models in Geography: The Political Economy Perspective*, vols. 1 and 2. London: Routledge.
Rimmer, P. (1991). The global intelligence corps and world cities: Engineering consultancies on the move. In Daniels, P. (ed.) *Services and Metropolitan Development: International Perspectives*, pp 66 106. London: Routledge.
Sassen, S. (2001). *The Global City* (2nd edn.). Princeton, NJ: Princeton University Press.
Sassen, S. (2006). *Territory, Authority, Rights: From Medieval to Global Assemblages*. Princeton, NJ: Princeton University Press.
Sklair, L. (2001). *The Transnational Capitalist Class*. Oxford: Blackwell.
Stutz, F. P. and de Souza, A. R. (1998). *The World Economy: Resources, Location, Trade, and Development* (3rd edn.). Upper Saddle River, NJ: Prentice Hall.
Taylor, M. J. and Thrift, N. J. (1982). *The Geography of Multinationals*. New York: St. Martin's Press.
Taylor, P. (2004). *World City Network: A Global Urban Analysis*. London: Routledge.
Thrift, N. (1999). The rise of soft capitalism. In Herod, A., Ó Tuathail, G. & Roberts, S. (eds.) *An Unruly World? Globalization, Governance and Geography*, pp 25 71. London: Routledge.
Thrift, N. (2005). *Knowing Capitalism*. London: Sage.
Waters, J. (2006). Geographies of cultural capital: Education, international migration and family strategies between Hong Kong and Canada. *Transactions of the Institute of British Geographers* 31(2), 179 192.
Wright, M. C. (1956). *The Power Elite*. New York: Oxford University Press.
Yeoh, B. S. A. and Chang, T. C. (2001). Globalising Singapore: Debating transnational flows in the city. *Urban Studies* 38, 1025 1044.
Yeoh, B. S. A., Huang, S. and Lam, T. (2005). Transnationalizing the 'Asian' family: Imaginaries, intimacies and strategic intents. *Global Networks* 5(4), 307 315.
Yeung, H. W. c. (2004). *Chinese Capitalism in a Global Era: Towards Hybrid Capitalism*. London: Routledge.

Relevant Websites

http://www.ft.com/home/us
Financial Times.
http://www.lboro.ac.uk/gawc/
Globalization and World Cities Research Network.
http://international.metropolis.net/index e.html
International Metropolis Project.

Transnational Ethnic Networks

Jinn-Yuh Hsu, National Taiwan University, Taipei, Taiwan

Glossary

Diaspora It refers to ethnic groups whose sizable parts have lived outside their country of origin for at least several generations, while maintaining some ties to the homeland.

Overseas Chinese Business Networks Chinese business people can exploit their privileged ability to recognize worldwide market differences, or knowledge arbitrage. From this perspective, market differences can be translated into geographical market differences, and then into regional comparative advantage.

Social Network Analysis Social network has been used as a metaphor to connote complex sets of relationships between members of social systems at all scales. But social network analysis needs to move from being a suggestive metaphor to an analytic approach to a paradigm. It takes account of the structure of a network, the relations among network members, and the location of a member within a network as critical factors in understanding social behaviors.

Transnationalism The processes by which immigrants build social fields that link together their country of origin and their country of settlement.

One of the most promising contributions to recent debates about globalization is the literature in which Cox *et al.* analyze the interplay between the various scales of economic activity, from the local to the global. These theorists recognize that the process of globalization constitutes of the complex relationships between multiple, different local systems – that we are in an era of glocalization, rather than globalization. The key challenge for this perspective is to understand how producers cope with the interface between the local and the global, or the governance mechanism in the glocalization process. Some writers take the multinational corporation as the core economic actor in a global hierarchical fiat system, others like Scott and Dicken conceive of the global system as a network of regional worlds. Within this network governed economy, social ties between communities in the interacting regions are a central concern. Rather than taking the large corporation as the major actor, the network approach focuses on the social embeddedness of the ostensibly profit oriented business world. Scholars adopting this network approach to globalization have focused on the role of ethnic ties, in particular, to explain the accelerated growth of cross border economic transactions. They see social networks, not asocial economic rationality, as the basis for the emergence of economic transnationalism.

Transnationalism has become a key field of study in international migration since the later part of the 1990s. Across a range of disciplines, academicians sought to define and trace the development of transnational communities and practices, and examine the ramifications for identity and citizenship in an increasingly globalized world. Glick Schiller *et al.* (1992) define transnationalism as "the processes by which immigrants build social fields that link together their country of origin and their country of settlement." The transnationalist approach is based on the observation that expatriates keep close personal and financial relationships with the source country and they retain a strong national identity even years after migrating. Driven by shared national identity, migrants establish social networks with other expatriates from the same source country. This common identity perpetuates a commitment to the source country that drives cross border linkages among nationals and decisions to return. A migrant has thus a natural gravitation toward the home country and repatriation takes place once enough resources, whether financial or knowledge, have been gathered and when the social and economic conditions at home are perceived sufficiently favorable. Hence, the main motivating factor, according to Portes, is not one of personal utility but an identity marked by an attachment to one's birthplace.

The operation of social networks from below, through the mechanisms of transnational migration, would tackle the issue of the articulation of people's everyday lives (both of elites and ordinary ones) with the macro forces of sociocultural and political–economic institutions in different geographical scales. Smith (2001) suggests that the globalization–localization process articulated in the global city is manifest in 'transnational social space' or the 'translocality', through the agency of 'transnational migrant networks'. Among these networks, the transnational connection made possible by a community of people, no matter whether elite or refugees, is the key agent in the emergence of transnationalism. In particular, the international migration of skilled people and wealthy people was identified as the actor in cross fertilizing both the sending and hosting cities with the traveling back and forth of highly specific knowledge, capital, and professional networks. Evidence of the transnational hypermobility of this highly skilled and capitalist class is the development of a growing population of 'astronauts' who work in 'both' places and spend much of their lives on airplanes. While their families may be based on either side of the Pacific (most often they stay in California, Vancouver, or Toronto because of the lifestyle advan tages), these wealthy migrants join their families for a couple of weeks every 3 or 6 months, taking advantage of the opportunities to play middlemen bridging the cross Pacific regional economies.

Central to concept of transnationalism is the recon figuration of the geographical space and territorial boundaries to accommodate intricate and flexible inter actions engaged in by migrants at a global, regional, and local level. Such personalized networks used to be dis missed as outdated or marginal modes of transactions under modernism. Now they command new respect under the condition of late capitalism, globalization, and flexible accumulation.

As demonstrated by Portes (1998), the transnational immigrant communities often actively bridge the donor and host regions, and constitute another network for capital, information, and people flow, besides, although not necessarily contradictory to, the mainstream channels of foreign direct investment and international trade. Some of the members of the diaspora simply settle abroad and become progressively integrated into local ways and constitute an enclave economy; others cultivate their networks across space, and travel back and forth in pursuit of reciprocal ventures. The reciprocal collaboration be tween the diaspora and the homeland can be in different forms: technology transfer and financial contributions, including both remittances and investments. While the former move along the course of assimilation, the latter refuse to be confined in either one space or the other. It is the latter that links their members with their countries of origin and turn the brain drain into a brain gain.

Conventionally, the human capital paradigm refers to a substantialist view of skills as a stock of knowledge and abilities embodied in the individual and takes the concept of brain drain to denote the phenomenon whereby a country suffers an outflow of its educated elite, on a scale threatening the needs of national development in the long term. In contrast, the network approach, adopting a connectionist viewpoint, claims that skills cannot be understood without their social definition, construction, and integration, and consequently, individuals should be presented as being involved in knowledge intensive ac tivities, deeply rooted in their networks, with their own skills historically and physically contextualized. More often, a transnational technological community can pro vide an alternative and potentially more flexible and responsive mechanism for long distance transfers of skill and know how, particularly between very different business cultures or environments.

Mohann (2002) classifies three types of linkages of diaspora and homeland development: development 'in', 'through', and 'by' the diaspora. The first one, de velopment in the diaspora relates to the pooling of fi nancial capital, intellectual capital, and political capital by members of an ethnic community for the purpose of growing wealth in the homeland country. Lots of sending governments have sponsored programs to get access to the pooling for development. For example, the Chinese government runs the *Chun hui* Program which grants fi nancial and manpower support to entrepreneurs who have returned, to start ventures in China. A number of 'returning students venture parks' were established to target returning entrepreneurs.

The second type, development through diaspora means development as a result of networking within and between diasporas of the same ethnic group in different parts of the world. A classic example involves the networks of overseas Chinese from Taiwan, Hong Kong, Singapore, and the U.S. who contribute to economic development after China's economic reform in late 1970s (see below). Another well noted example comes from the investigation made by Michell and Olds on the Pacific Place in Vancouver, a property project invested in by the Hong Kong–based property tycoon Li Ka Shing who took advantage of a long term North American base for his family and friends to build a cross Pacific Chinese business network in the pursuit of property development opportunities.

Finally, the third type, development by diaspora refers to the diaspora working chiefly in helping the homeland development. For instance, many Jewish associations have been established to work with the Israelis to ar ticulate the needs of Israel to Washington, raise and contribute funds to elect congressional members who support Israel, and help defeat those who do not.

The Overseas Chinese Business Networks (OCBN): An Illustration

The dominant example of the transnationalist approach is the analysis of the Overseas Chinese Business Net works (OCBN). The OCBN argument claims that Chi nese business people can exploit their privileged ability to recognize worldwide market differences, or what Kao (1993) refers to as 'knowledge arbitrage'. From this per spective, market differences can be translated into geo graphical market differences, and then into regional comparative advantage, as Borrus (1997) argues for Asian production networks. It is referred to as the 'global web' of Chinese business and sees the close ties between the overseas Chinese and local Chinese communities as a central mechanism for economic cross fertilization in the Pacific Rim.

By the same token, it has been argued that there is a distinctive and crucial synergy between overseas Chinese investors and local governments in Mainland China, one not found to the same extent or in the same form with non Chinese FDI. Yeung (1998) also demonstrated, based on his research about the investments made by Hong Kong based Chinese firms in the Southeast Asia, the Chinese cultivation of personal relationships institutionalized the trust, loyalty, reciprocity, and reputation that facilitated efficient business operations. Synergy is seen among the specific skill set of Chinese investors (experience with labor intensive exports to the West and with managing to achieve high quality and productivity from unskilled labor), cultural similarities and propensities (shared culture and language, ability to form trustworthy relationships, values of authority and hard work), knowledge of how to get things done in contexts of policy uncertainty while minimizing transaction costs, and industrial structures that tolerate higher levels of regulatory uncertainty and risk.

In this view, global economic transactions are enhanced by the advantage of blood bonds. Ethnic ties render the utilization and coordination of resources among firms of the cross border regions flexible and economical, reinforcing their competitiveness. Such a system is seen as particularly advantageous in the current economic environment as it allows the members of the network to quickly identify complementary assets and build close partnerships at a global scale.

Moreover, today's learning economy involves more than purchasing technology and includes social dimensions such as the absorption of tacit knowledge which is embodied in technical staff. Consequently, the key to success in the rapidly changing market lies in capabilities to identify the right people (know who), and accordingly fix the right technologies and products (know how), as most innovations are human embodied and involve teamwork.

More recently, growing empirical evidences suggest that highly skilled engineers and professionals are returning in large numbers. In many cases they have become key players in generating new technological practices and capacity in their home countries. Saxenian (2006) has found active engagement of Chinese and Indian engineers in the technological startups in Silicon Valley, California and noted the attempts of these émigré to contribute to technology development in their home countries. Saxenian and Hsu (2001) also discovered that the international circulation of technical expertise has become institutionalized by the local ethnic associations in California, that promotes the transnational ties and activities of émigré entrepreneurs or professionals in their home countries. These findings led to the term: 'brain circulation' that describes the phenomenon of skilled immigrants being of increasing benefit to both home and hosting nations.

All of these possible links are established smoothly not on an individualistic base, but with the mediation of overseas organizations. Take the Silicon Valley–Taiwan connection as an example. One of the most important overseas organizations for high technology industries is the Monte Jade Science and Technology Association. It was formed to promote cooperation and mutual flow of technology and investment between Taiwan and Silicon Valley. The activities of the Monte Jade include monthly dinner meetings, which encourage and promote networking among members, special topic seminars that are put on in cooperation with other professional organizations, and social events and entertainment. Through Monte Jade's efforts, many networking opportunities are created and proliferated.

Similar to the Silicon Valley–Taiwan connection, China's link became institutionalized by establishing a transnational community. One of the key associations, the Hua Yuan Science and Technology Association was formed in Silicon Valley in 2000 to 'promote the technological, professional, and scientific development of the Chinese business community'. Hua Yuan and other Chinese professional associations also sponsor regular business tours to China, receive government delegations, and serve as conduits for Chinese firms recruiting in the US. Through the forums provided by these organizations, two way channels were established for Silicon Valley's high technology capital and China's emerging market and manpower.

Instead of focusing on personal linkages and 'natural' gravity towards home, the institutionalization argument emphasizes the importance of voluntarily created linkages that are targeted on specific objectives. Expatriates are not passive respondents to social and economic conditions in the home country, but social actors that are motivated by their linkages with home institutions and social capital from which they benefited before migrating. Not only are acceptable social and economic conditions an important factor for returning, but also opportunities that the home country provides for making use of the migration experience and established networks. Expatriates are thus regarded as well informed agents that gather information about context and opportunities in their countries of origin.

Cautionary Remarks

Critics have pointed out a number of flaws in the transnational economic network model which presumes that shared language, culture, and history facilitates the construction of shared identities among ethnic diaspora. First, it tends to assume both the continuation of cultural commonalities and the power of shared identity to facilitate trustworthy business networks which can then operate efficiently with lower transaction costs. This culturalist position ignores internal differentiation among diaspora and takes identity as an unproblematic given. Second, it assumes that even while social connections are used for business purposes, their basis in solidarity and commonality resolves the problems of conflicting economic interests. Third, even if business ties are more easily constructed and maintained among ethnic diaspora, and even if they do generate effective and stable cooperation, the approach still risks oversocializing economic behavior that is rooted in business and technological considerations and thus assumes that social relationships determine economic transactions and outcomes. While economic relationships are clearly socially and culturally embedded, markets and industries have their own dynamics that reward certain types of organizations and practices and drive others toward bankruptcy and failure.

In fact, the 'Chinese business system' should be examined in a dynamic and transformational way. Based on his case studies of the internationalization of a large number of Chinese firms from Singapore and Hong Kong, Yeung (2004) clearly illustrated that their business networks have been broadened in the globalization process. The growing involvement of nonfamily, or even non Chinese members in the entrepreneurial activities of key actors in Chinese business networks showed that dynamic transformation process (e.g., professionalization) can propel these ethnocentric firms to survive in the relentless competition of world economy.

Moreover, the role played by ethnic ties varies by sectors. Rather than arguing that ethnic social capital won't work in the technology intensive industries, the critics contend that its role and style is changed. For example, Smart and Hsu (2004) find that ethnic social capital was crucial for reducing transaction costs (or indeed even making investment possible at all in certain circumstances) and enhancing mutual trust in the early stages of overseas Chinese investments in Mainland China. After the business environment became clearer and technology intensive investment became more sought after, ethnic ties became more important as collective assets which assisted in the identification and recruitment of appropriate talents among overseas Chinese in the United States and across the Taiwan Strait. In brief, the economy is not reducible to interpersonal relationships, but composed of multiple production worlds that are defined by product configuration, market principles, and technology and production process. Dense social ties cannot substitute for the sophisticated managerial and technological learning that is required to compete in a particular sector in spite of the fact that the social dimension of learning is critical.

Another controversy with the thesis of transnational ethnic network involves the role of the state. The students of developmental state attack against the ignorance of the state leadership in the transnationalization process. Rather than counting for the rise of transnationalism as 'globalization from below', they argued that it is the state, particularly in the latecomer industrialization, which initiates and fosters the cross border networking, and sponsors the returnee program to take advantage of the diaspora connection in industrial upgrading.

Such a developmental statist argument reveals a partial truth that the late industrializing state pushes aggressively in encouraging spinoffs from public labs to take shape from dense social–technical networks, and recruiting overseas experts back to start up new ventures. However, the key challenges met by the developmental states are the construction of network development to participate in the new economy which becomes imperative to integrate local, regional, and global structure of innovation into a system. In other words, if the developmental state is argued to be a top down, bureaucratic rationality based governance mechanism, how can it build up and articulate with the supposedly bottom up, trust based social networks? How can the potential tension between the top down and the bottom up be settled?

As a matter of fact, while making a distinction from the argument of 'globalization from above', transnationalism does not contradict with the role played by the state in controlling borders and regulating migration, as critiqued by Waldinger and Fitzgerald (2004). In contrast, transnational social relations are usually anchored in, while also transcending, one or more nation states. The transnationalist discourses insist on the continuing significance of national borders, state policies, and national identities, while simultaneously crossing over them, and constituting a hybrid social space. Instead of maintaining a kind of zero sum assumption of exclusiveness of nation states and globalization, the transnationalist discourses take them as engaging in a process of mutual construction.

Finally, the idea of transnational ethnic network bases its conceptualization on the network governance approach which revolves around Granovetter's embeddedness notion. As observed by Peck (2005), the Granovetterian reading of embeddedness epitomizes a highly versatile key concept in the economic geography of the so called 'cultural turn'. Invariably, economic geography until more recently turns networks into shorthand for enduring, trust based ties, and focuses on the indeed very human side of family, friendship, and kin in economic relations. Consequently, network, in most economic geography literature, is used as a metaphor or governance form, rather than an analytic tool.

In contrast, as suggested by Grabher (2006), economic geographers can alternatively take social network analysis seriously and systematically investigate the behavioral consequences of network configuration. In light of the analysis, the critical role of network position and structure fundamentally departs from the cohesion fixated ideas of networks. Burt's (2001) social network analysis invites an understanding of arbitrage and innovation that sharply contrasts with the trust based relations prevailing in the study of transnational ethic networks in particular, and economic geography in general. From the standpoint of network analysis, the arbitrage advantage and lock in disadvantage of transnational ethnic network can be analyzed by exploring the interdependencies between accounts on structure and context.

See also: Embeddedness; Ethnic Economies; Global Production Networks; Networks; Transnationalism and Technological Transfer.

Further Reading

Amsden, A. and Chu, Wan wen. (2003). *Beyond Late Development: Taiwan's Upgrading Policies*. Cambridge: MIT Press.

Beaverstock, J. (2002). Transnational elites in global cities: British expatriates in Singapore's financial district. *Geoforum* 33, 525 538.

Borrus, M. (1997). Left for dead: Asian production networks and the revival of U.S. electronics. In Naughton, B. (ed.) *The China Circle: Economics and Technology in the PRC, Taiwan, and Hong Kong*, pp 139 163. Washington, DC: Brookings Institution Press.

Burt, R. (2001). Structural holes versus network closure as social capital. In Lin, N., Cook, K. & Burt, R. (eds.) *Social Capital: Theory and Research*, pp 31 56. New York: Aldine De Gruyter.

Coe, N. and Bunnell, T. (2003). 'Spatializing' knowledge communities: Towards a conceptualization of transnational innovation networks. *Global Networks* 3(4), 437 456.

Coe, N., Kelly, P. and Olds, K. (2003). Globalization, transnationalism, and the Asia Pacific. In Peck, J. & Yeung, H. (eds.) *Remaking the Global Economy*, pp 45 60. London: Sage.

Cox, K. (1997). *Spaces of Globalization: Reasserting the Power of the Local*. New York: Guilford Press.

Dicken, P. (2003). *Global Shift: Transforming the World Economy* (4th edn.). New York: Guilford Press.

Dicken, P. and Yeung, H. (1999). Investing in the future: East and Southeast Asian firms in the global economy. In Olds, K., Dicken, P., Kelly, P. F., Kong, L. & Yeung, H. (eds.) *Globalisation and the Asia Pacific: Contested Territories*, pp 111 122. London: Routledge.

Grabher, G. (2006). Trading routes, bypasses, and risky intersections: Mapping the travels of 'networks' between economic sociology and economic geography. *Progress in Human Geography* 30(2), 163 189.

Haque, N. and Kim, S. (1994). Human capital flight: Impact of migration on income and growth. IMF Working Paper. Washington, DC: International Monetary Fund.

Hsing and You Tien (1998). *Making Capitalism in China: The Taiwan Connection*. New York: Oxford University Press.

Hsing and You Tien (2002). Ethnic identity and business solidarity: Chinese capitalism revisited. In Laurence, J. C. & Cartier, C. (eds.) *The Chinese Diaspora: Space, Place, Mobility, and Identity*, pp 221 236. Lanham: Rowman and Littlefield.

Hsu, Jinn Yuh and Saxenian, A. (2000). The limits of *guanxi* capitalism: Transnational collaboration between Taiwan and the US. *Environment and Planning A* 32(11), 1991 2005.

Kao, J. (1993). The worldwide web of Chinese business. *Harvard Business Review*. March April: 24 36.

Ley, D. (2004). Transnational spaces and everyday lives. *Transactions of the Institute of British Geographers* 29, 151 164.

Lin, J. (1998). *Reconstructing Chinatown: Ethnic Enclaves, Global Change*. Minneapolis, MN: Minnesota University Press.

Low, M. (1997). Representation unbound: Globalization and democracy. In Cox, K. (ed.) *Spaces of Globalization*, pp 240 280. New York: Guildford Press.

Lundvall, B. A. (1996). The social dimension of the learning economy. DRUID Working Paper: 96 1. Aalborg: Aalborg University.

Meyer, J. B. (2001). Network approach versus brain drain: Lesson from the diaspora. *International Migration* 39(5), 91 110.

Mitchell, K. and Olds, K. (2000). Chinese business networks and the globalization of property markets in the Pacific Rim. In Wai chung Yeung, H. & Olds, K. (eds.) *Globalization of Chinese Business Firms*, pp 195 219. Basingstoke: McMillan Press.

Mohann, G. (2002). Diaspora and development. In Robinson, J. (ed.) *Development and Displacement*, pp 77 139. London: Open University Press.

Olds, K. (2001). *Globalization and Urban Change*. Oxford: Oxford University Press.

O'Riain, S. (2004). *The Politics of High Tech Growth: Developmental Network States in the Global Economy*. Cambridge: Cambridge University Press.

Patterson, R. (2006). Transnationalism: Diaspora homeland development. *Social Force* 84(4), 1891 1907.

Peck, J. (2005). Economic sociologies in space. *Economic Geography* 5, 127 176.

Portes, A. (1998). Globalisation from below: The rise of transnational communities. ESRC Transnational Communities Program Working Paper No. 1. Oxford

Powell, W. and Smith Doerr, L. (1994). Networks and economic life. In Smelser, N. & Swedberg, R. (eds.) *The Handbook of Economic Sociology*, pp 368 402. Princeton, NJ: Princeton University Press.

Pries, L. (2005). Configurations of geographic and social spaces: A sociological proposal between 'methodological nationalism' and the 'space of flows'. *Global Networks* 5(2), 167 190.

Redding, G. (2002). The Chinese business system of China and its rationale. *Asia Pacific Journal of Management* 19, 221 249.

Saxenian, A. (2002). Transnational communities and the evolution of global production networks: The case of Taiwan, China and India. *Industry and Innovation* 9(3), 183 202.

Saxenian, A. (2006). *The New Argonauts: Regional Advantage in a Global Economy*. Cambridge, MA: Harvard University Press.

Saxenian, A., Hsu and Jinn Yuh (2001). The Silicon Valley Hsinchu connection: Technical communities and industrial upgrading. *Industrial and Corporate Change* 10(4), 893 920.

Schiller, N. G., Basch, L. and Szanton Blanc, C. (1992). Transnationalism: A new analytic framework for understanding migration. In Schiller, N. G., Basch, L. & Szanton Blanc, C. (eds.) *Toward a Transnational Perspective on Migration*, pp 1 24. New York: New York Academy of Sciences.

Scott, A. (1998). *Regions and the World Economy: The Coming Shape of Global Production, Competition and Political Order*. Oxford: Oxford University Press.

Smart, A., Hsu and Jinn Yuh (2004). The Chinese diaspora, foreign investment, and economic development in China. *The Review of International Affairs* 3(4), 544 566.

Smith, M. (2001). *Transnational Urbanism: Locating Globalization*. Oxford: Blackwell.

Storper, M. and Salais, R. (1997). *Worlds of Production: The Action Frameworks of the Economy*. Cambridge, MA: Harvard University Press.

Vertovec, S. and Cohen, R. (1999). *Migrations, Diasporas and Transnationlism*. Cheltenham: Edward Elgar.

Waldinger, R. and Fitzgerald, D. (2004). Transnationalism in question. *American Journal of Sociology* 109(5), 1177 1195.

Yeung and Wai Chung (1998). *Transnational Corporations and Business Networks: Hong Kong Firms in the ASEAN Region*. London: Routledge.

Yeung and Wai Chung (2004). *Chinese Capitalism in a Global Era: Towards Hybrid Capitalism*. London: Routledge.

Yeung, Y. (2000). *Globalization and Networked Societies: Urban Regional Change in Pacific Asia*. Honolulu, HI: University of Hawaii Press.

Zhou, M. (1992). *Chinatown: The Socioeconomic Potential of an Urban Enclave*. Philadelphia, PA: Temple Universiy Press.

Zhou, Y. and Tseng, Y. (2001). Regrounding the 'ungrounded empires': Localization as the geographical catalyst for transnationalism. *Global Networks* 1(2), 131 153.

Relevant Websites

http://www.montejade.org/homepage/
Monte Jade Science and Technology Association.

Transnationalism

D. Ley, University of British Columbia, Vancouver, BC, Canada

While there are disagreements about the beginnings of globalization – a dominant current paradigm across a number of interdisciplinary research fields – no one discounts the significance of the neoliberal innovations of the 1980s that accompanied the pro market policies of President Reagan and Prime Minister Thatcher. Transnational corporations were the agents of ever more intensive global economic activity, as they increasingly controlled raw material, manufacturing, financial, and consumer sectors of the economy. So it is scarcely a surprise that in conjunction with these global flows of capital, what has been called a transnationalism from above, scholars in the 1990s were soon recognizing a corresponding transnationalism from below, the accelerating global movement of labor. But these migrant workers did not necessarily follow the conventional role of settlers to a new land. A much more flexible model of movement saw repeated journeys between homeland and destination countries, with variable periods of residence at each end. So prevalent has this style of migration become that by 2005 the UN's Global Commission on International Migration was identifying a new norm, with the former paradigm of permanent migrant settlement giving way to a new pattern of temporary and circular migration.

This is a significant reconceptualization of international migration. The conventional wisdom envisages a simple account of departure, arrival, settlement, and assimilation as the migrant undergoes a status passage from an *émigré* of the old world to an acculturated citizen of the new world. An extension of this model includes return migration, undertaken by a minority, as a completion of a migration cycle, with the same sense of finality and closure once the return move to the homeland has been completed. Transnationalism complicates that view as migrants move repeatedly across national borders, and 'here' and 'there' remain, for them, part of a single social field. Such a transnational social space is sustained by the availability of cheap and fast communications and transportation. The ever more affordable real cost of air travel permits relatively frequent visits among members of households and communities who are fragmented on opposite sides of national borders. Access to cheap telephone services, the Internet, and television satellite dishes allow for daily or weekly renewal of ties with friends, families, and cultures elsewhere.

In addition to the maintenance of old world social networks, researchers quickly discovered that money as well as good wishes were sent through communications networks, giving rise to the important study of the relationship between migrant remittances and economic development at the origin. In a few cases economic flows moved in the other direction, from homeland economic activity to relocated family members overseas. In addition, some migrants have sustained homeland political connections, particularly with the expansion of dual citizenship, permitted by a growing number of nations in part to retain influence over their diasporas; Italy has taken this process furthest, creating four overseas electoral constituencies for its diaspora. Research has also noted the development of transnational organizations and social movements as well as the activities of formal political participation. Studies of transnationalism have proliferated by region and by focus, and a number of special issues of journals have reviewed and extended the subfield.

Transnationalism by Discipline

The initial contributions to the now abundant literature on transnationalism came from anthropologists who noted the density of migrant transactions between destinations in the United States and origins in the Caribbean islands and Central America. These studies were commonly ethnographic in method, with a rich description and interpretation of activities that spanned national borders. The entry of sociologists to the field led to methodological disagreement, with Alejandro Portes and his co workers in particular expressing concern at what they regarded as a lack of rigor in some of the anthropological work. They criticized the uneven treatment of evidence. By only citing those cases where transnational activities could be demonstrated, and disregarding those cases where they were absent, researchers were only 'sampling on the dependent variable', thereby exaggerating the extent of the phenomenon. Ideally a sample should be randomly selected to be more representative of a larger population of immigrants, including those with and without significant transnational activities. In place of the qualitative evidence produced in the ethnographies, the critics also urged the necessity for standardized and measurable evidence from large surveys that would demonstrate the significance of transnationalism and permit comparative studies in different regions.

By and large, geographers have sided with the qualitative approach of anthropology rather than the quantitative preferences of sociology, though they have typically employed the compromise of semi directed interviews of immigrants rather than lengthy ethno graphic studies. This approach provides some methodo logical uniformity in data collection while at the same time there is room for exploration of individual experi ences. One emphasis has been to challenge the argument that globalization is associated with the 'end of geog raphy'. The compression of time and space associated with faster travel and cheaper communication un doubtedly diminishes the friction of distance. It has been argued that this ease and frequency of movement has led to a 'deterritorialization' of transnational migrants, as they tend toward a cosmopolitan and placeless identity. Moreover, the power of transnationalism from above, the strength and reach of private corporations, is said to di minish the role of national government, and thereby di lute the distinctiveness of individual places.

A strong argument could be made that neither of these challenges to the continuing significance of place and space holds. First, cosmopolitans also have particular loyalties and place affiliations. It has been noted that rather than being transnational many migrants might better be described as translocal, for their movements are between a small number of distinct localities, for ex ample, between a Mexican village and a co ethnic neighborhood in a city in the United States, or between the central neighborhoods of a global city like London and equivalent expatriate neighborhoods of some other world city, perhaps Hong Kong or Sydney. Second, time–space compression by collapsing distance actually deep ens the significance of geographical variation. For, the spatial extension of the migrant's life world permits ac cess to a range of geographical conditions. A popular saying among wealthy migrants to Canada from Hong Kong has been, "Hong Kong for making money, Canada for quality of life." Such a statement acknowledges not only that households are able to partake of quite different geographies but also implies that the art of transnational behavior is to maximize net household benefits across these diverse geographical settings. More than this, at different stages in the life cycle one setting may rise to prominence over the other. In the case of wealthy East Asian migrants to Canada or Australia, East Asia takes priority during the years of career and equity building, but its attributes diminish during childrearing when Western educational opportunities are preferred, and again at retirement when quality of life considerations prevail. Indeed, the variable attributes of different geo graphical places at particular life stages are those that motivate repeated movement.

We should not assume, however, that transnational migrants are always masters of spatial variation. The wealthy migrants from Hong Kong, Taiwan, and South Korea who arrive in North America, Australia, and New Zealand have considerable business experience. While this stock of human capital facilitates their entry status, for some visa types there is also the necessity that they develop a business in their place of landing. Despite their successful track record in their homeland, empirical study has shown that they rarely repeat as positive a performance as migrants. A typical reason is their un familiarity with the regulated economic culture of their destination countries. The labyrinthine complexity of labor, environmental, and business regulations is con fusing, and in addition tax rates are much higher than the 15% or so they are used to paying in East Asia. It is because geography still matters that successful business people in East Asia are often unable to match their performance in North America or Australasia.

Dimensions of Transnationalism

Relatively few studies exist that document the extent of transnational behavior, because such data are difficult, and therefore expensive to collect, requiring a survey with a large sample undertaken in a number of languages. **Table 1** shows a profile of transnational activities undertaken by a sample of close to 1500 immigrants interviewed by telephone survey in Vancouver, Canada. Interviews were conducted in the respondent's choice of English or one of the leading five Asian languages spoken at home. Given this constraint, a random sample of re spondents was contacted in five parts of the city that varied by socioeconomic status and age of housing. Contact was made with an adult household member, who might or might not be the principal wage earner, con tributing to a significant fraction of respondents who were not in the labor force.

Almost all respondents continue to have friends or family in their country of origin and not surprisingly a high level of contact is maintained with them, at least weekly for 45% (**Table 1**). Cheap international rates meant that the telephone was overwhelmingly the most popular means of contact, though half also used e mail, while only a fifth maintained contact through the postal service. Two thirds of immigrants travel back to their country of origin – a lengthy trans Pacific crossing for most – and of these over two fifths make the trip at least once a year. Traffic also moves in the other direction, with more than 60% of immigrants receiving visitors from their home country, 40% of these at least once a year. We may conclude from this survey that transna tional contacts are sustained with some regularity be tween origin and destination.

The survey also probed the economic dimensions of transnational contacts. Surprisingly, less than 15% ac knowledged sending remittances to families back home, a rather low figure that may be influenced by the pattern of immigration to Vancouver that includes few sending countries, other than the Philippines, that are reputed as

Table 1 Basic enumeration of transnational activities

Question	Yes #	Yes %	No #	No %	Refused	Total
Family or friends in pre-migration country?	1332	90.1	143	9.7	4	1479
Do you keep in touch with them?	1271	95.4	60	4.5	1	1332
How often?						
Daily	87	6.8				
Weekly	486	38.2				
Monthly	524	41.2				
Yearly	139	10.9				
Less than once a year	23	1.8				
Don't know/Refused	12	0.9				
Total	1271					
How do you keep in touch?[a]						
E-mail	652	51.3				
Telephone	1167	91.8				
Postal mail	253	19.9				
Visits	302	23.8				
Refused	3	0.2				
Total	1271					
Do you travel to that country?	1008	68.2	468	31.6	3	1479
How often?						
More than once a year	201	19.9				
Once a year or so	238	23.6				
Less than once a year	550	54.6				
Don't know/Refused	19	1.9				
Total	1008					
Does your job require you to travel to that country?[b]	48	7.9	555	91.7	2	605
Do you own property there, or have a home?	329	22.2	1123	75.9	27	1479
Do you run a business there?	47	3.2	1411	95.4	21	1479
Do you provide financial assistance to people there?	212	14.3	1247	84.3	20	1479
Do family or friends from that country visit you here?	903	61.1	567	38.3	9	1479
How often?						
More than once a year	123	13.6				
Once a year or so	237	26.2				
Less than once a year	514	56.9				
Don't know/Refused	29	3.2				
Total	903					

[a] Respondents were allowed two answers to this question; percentage figures therefore exceed 100.
[b] 676 out of 1479 immigrant respondents are currently employed (another 155 are self employed).
Reproduced from Hiebert, D. and Ley, D. (2006). Characteristics of immigrant transnationalism in Vancouver. In Satzewich, V. & Wong, L. (eds.) *Transnational Identities and Practices in Canada*, pp 71 90. Vancouver: University of British Columbia Press.

receiving high levels of remittances. In contrast, more than a fifth still own property in their homeland, a figure possibly inflated by former farmers from India in the sample who have not sold their holdings. But only 6% travel to their homeland for job related purposes, and half as few again own a business there. These more dis criminating indicators indicate a more slender role for economic transnationalism than the near ubiquitous pattern of social traffic. While it is difficult to know how generalizable these results are, the general profile – broad social traffic, much more limited economic linkages – agrees with other surveys undertaken in immigrant gateway cities in the United States and Canada.

The Vancouver data do however, because of the par ticular profile of the respondents, produce lower esti mates of remittances than has been shown in other studies. All data on remittances are estimates, for many economic transactions do not enter public records, while some official sources, notably the International Monetary Fund, have a very expansive definition of the term. Nonetheless, we do know that remittances are a signifi cant part of the foreign earnings of small countries and even of larger nations including India, Mexico, China, and the Philippines. As such they are potentially an im portant component of economic development in these countries, particularly in villages and small towns where funds from individual senders may rebuild family homes, while the targeted projects of overseas hometown associations complement community infrastructure – paying for a school, a road, a medical clinic, or clean drinking water – and thus may play important local roles.

Despite measurement problems there is agreement that the scale of remittances is large and increasing – a World Bank estimate suggests that flows are now two thirds the size of foreign direct investment in developing countries. Moreover, remittances are less cyclical and more dependable than market flows. World Bank estimates suggest that in 2005 India received $23.5 billion in remittances from overseas workers, China $22.4 billion, and Mexico $21.7 billion. Such large numbers represent the sum total of many individual cases of self denial and hardship by migrants, for example, Ghanaians or Pakistanis who endure severe overcrowding and long work hours in gateway cities in the global North to maximize the funds sent home, or Filipinas who as domestic workers sacrifice autonomy and sometimes dignity in order to support families back home.

Political Aspects of Transnationalism

Partly because of the economic impact of rising levels of remittances, developing countries have been attempting to maintain their connections with their global diasporas. A second reason, notably in China and India, and also to a growing extent in countries with emerging shortages of skilled labor, such as Japan or Australia, is to encourage the return home of nationals who are skilled workers in foreign countries. As mentioned earlier, in its last federal election, Italy established four overseas constituencies for members of its diaspora to elect deputies to the Italian Parliament. Other emigrants play an even more direct role in national politics. In a recent election in an African state, competing candidates for one seat were both residents of New York City. Such transnational political identities are facilitated by the growing tenure of dual citizenship, so that home country political identities and passports are maintained even while new passports and political rights are collected from the receiving country.

The holding of dual passports certainly permits a practice of 'flexible citizenship' to be practiced, with households maximizing their rights in different nations for different purposes. During the Israeli bombardment of Lebanon in 2006, numerous expatriates contacted their missions in Lebanon requesting safe evacuation. As the immigrant receiving nations of Canada and Australia began their preparations, they were surprised to learn the size of their expatriate populations. Canada discovered the existence of 40 000–50 000 passport holders living in Lebanon, Australia some 25 000 citizens. A majority were living in Lebanon continuously and many held jobs there, but they had lived in their receiving country long enough to acquire a passport and citizenship rights. The evacuation of its nationals cost Canadian taxpayers over $80 million and sparked political discussion about the potential abuse of citizenship privileges, particularly as a number returned to Lebanon once hostilities had ended. Meanwhile perhaps 250 000 Canadian passport holders are living in Hong Kong, having returned following a typically brief residential tenure in Canada.

The broader issue here is the dilution of citizenship responsibilities with transnational movement. To what extent is national belonging, and individual and household contribution to the national project compromised by dual citizenship and split loyalties? It would seem as if transnational migrants might be opportunistic in their deployment of citizenship obligations, and that is certainly the sense of Aihwa Ong's reference to "flexible citizenship" to describe the behavior of wealthy trans Pacific movers. But an American study, undertaken by Portes and his co workers, did not support this expectation. Among a large sample of Colombian, Salvadorean, and Dominican immigrants in US cities, those engaging in cross border economic and political activity tended to be longer settled, of higher status, satisfied with life in the US, and naturalized citizens. In contrast, the Vancouver study mentioned earlier (**Table 1**) found exactly the opposite. While agreeing that transnational economic ventures are limited to a small proportion of immigrants, the Vancouver data showed that such residents tended to be more recent immigrants, and had landed in Canada with business class visas. They were less likely than other migrants to be naturalized Canadian citizens and also much less likely to feel a subjective affinity with Canada. The difference between the two studies is striking, and alerts us to the diversity among immigrants. The contemporary scale of transnational behavior includes households with diverse characteristics and thus a range of measurable outcomes.

Criticisms of Transnationalism

In less than 20 years, transnationalism has become the new migration paradigm in the definitive words of the UN Global Commission on International Migration. Inevitably, it has suffered from overstatement and overuse; we have already noted challenges to anthropological work along these lines. But a range of additional disagreements have also been launched by less sympathetic critics.

A repetitive challenge is that transnational behavior is not new but has been practiced by earlier migrants including the old European migration of the 1880–1930 period to the United States. It is claimed that the Jews from Eastern Europe and the Italians who arrived in large numbers during this period also maintained contact with home – indeed one of the early classics in social science examined the letters home from Polish peasants in the United States. It has been observed as well that between a quarter and a third of migrants from this first migration wave returned to Europe, so that departure from the sending country in no way implied a final rupture. While this argument has merit, the scale and intensity of cross border transactions today, plus the expansion of dual citizenship, have created a qualitatively different en vironment, facilitating not only a different order of magnitude in transnational contacts, but also a re definition of what it means to be a migrant. The ubi quitous presence of television satellite dishes on the outer walls of apartments in European cities, picking up homeland television signals, indicates that many migrants live each day in the cultural and informational environ ment of home even as they are physically located in a foreign land.

A second criticism challenges the degree of agency imputed to migrants in transnational approaches. While globalization narratives often introduce strong economic constraints around individual actions, there is a danger that transnational emphases move unduly in the opposite direction, exaggerating the freedom of migrants to pursue their own agendas. Indeed, Katharyne Mitchell has taken the argument further, challenging the tendency of some authors to celebrate transnational migrants as political actors who transgress, even subvert, state boundary maintenance. This position seems out of kilter with the empirical record. On the basis of his survey data, Portes' team claimed that the Latin American migrants with the most extensive transnational economic ties are those who are most assimilated in US society, scarcely *provocateurs*. Wealthier trans Pacific migrants similarly have no ar ticulated political project; their objectives are economic and familial, and if they transgress national boundary control it is in pursuit of these household aims.

Even if the characterization as *provocateurs* is removed, there is still an impression of footloose agency in the lit erature that needs to be qualified. To have access to mo bility is to have the freedom to escape local constraints, but in fact uneven geographies continue to exert their effects. For one thing, transnational movement may be a result of business failure in the receiving country which rekindles the attraction of the homeland; certainly, this is a finding from research on the return of skilled workers to the People's Republic of China after a failure to equal their pre migration professional status in North America. In deed, it is common to find downward economic and social mobility, especially for male household heads, following immigration, a status deficit that readily motivates return migration or transnational movement, with the household head replenishing the income stream from business ven tures in China, while family members continue to benefit from Western education and quality of life. The regular 'Pacific shuttle' that brings temporary reunion to such fragmented 'astronaut' families does not address the trauma of family separation, challenging in different ways to men, women, and children in such households. As narrated to the author by one such family: "When he leaves we cry in Vancouver, and he cries in Taiwan."

Conclusion

In an ever more mobile and interdependent world, transnational populations comprise a significant and growing share of many nation states. Their flexibility and predilection for movement bring a series of new chal lenges to the receiving nation state. Whereas a policy such as multiculturalism, or assimilation for that matter, is aimed to contain and perpetuate a population of na tional citizens – transnational migrants may or may not allow themselves to be corralled by a single nation state. The extension of dual citizenship and the innovation of diaspora politics is an attempt by both sending and re ceiving nation states to come to terms with their wan dering citizens. Sending nations are anxious to maintain contact with skill sets and remittance flows that bring palpable benefits, while receiving nations wish to con solidate their hold on valued human capital that has entered their borders and taken up citizenship. In dem onstrating the pervious nature of national borders, transnational migrants add weight to globalization the ories that announce the erosion of the borders and the authority of the nation state. Yet transnational migrants remain also subject to national political and economic regimes and vulnerable to (as well as benefiting from) the stresses of geographical separation and the uncertainties of spatial variation. Their mobility for economic ad vancement is commonly secured at the cost of a fraught emotional geography.

See also: Emigration; Immigration I; Population Geography; Segregation.

Further Reading

Ali Ali, N., Black, R. and Koser, K. (2001). The limits to 'transnationalism': Bosnian and Eritrean refugees in Europe as emerging transnational communities. *Ethnic and Racial Studies* 24, 578 600.

Bailey, A. (2001). Turning transnational: Notes on the theorization of international migration. *International Journal of Population Geography* 7, 413 428.

Basch, L., Glick Schiller, N. and Szanton Blanc, C. (1994). *Nations unbound : Transnational projects, postcolonial predicaments, and deterritorialized nation states*. Langhorne, PA: Gordon and Breach.

Beaverstock, J. (2005). Transnational elites in the city: British highly skilled migrants in New York City's financial district. *Journal of Ethnic and Migration Studies* 31, 245 268.

Global Commission on International Migration (2005). *Migration in an interconnected world: New directions for action*. Geneva: Report to the United Nations of the Global Commission on International Migration.

Hiebert, D. and Ley, D. (2006). Characteristics of immigrant transnationalism in Vancouver. In Satzewich, V. & Wong, L. (eds.) *Transnational identities and practices in Canada*, pp 71 90. Vancouver: University of British Columbia Press.

Jackson, P., Crang, P. and Dwyer, C. (eds.) (2004). *Transnational spaces*. London: Routledge.

Ley, D. (2004). Transnational spaces and everyday lives. *Transactions of the Institute of British Geographers NS* 29, 151 164.

Ley, D. and Kobayashi, A. (2005). Back to Hong Kong: Return migration or transnational sojourn? *Global Networks* 5, 111 127.

Mitchell, K. (2004). *Crossing the neoliberal line: Pacific rim migration and the metropolis*. Philadelphia, PA: Temple University Press.

Ong, A. (1999). *Flexible citizenship: The cultural logics of transnationality*. Durham, NC: Duke University Press.

Portes, A., Haller, W. and Guarnizo, L. (2002). Transnational entrepreneurs: An alternative form of immigrant economic adaptation. *American Sociological Review* 67, 278 298.

Vertovec, S. (1999). Conceiving and researching transnationalism. *Ethnic and Racial Studies* 22, 447 462.

Waldinger, R. and Fitzgerald, D. (2004). Transnationalism in question. *American Journal of Sociology* 109, 177 195.

Waters, J. (2002). Flexible families? Astronaut households and the experiences of lone mothers in Vancouver, British Columbia. *Social and Cultural Geography* 3, 117 134.

Relevant Websites

http://international.metropolis.net/International Metropolis Network on Immigration and Urbanisation.

http://www.migrationpolicy.org/ Migration Policy Institute.

http://www.iom.int/International Organization for Migration.

Transnationalism and Labor Geography

P. F. Kelly, York University, Toronto, ON, Canada

Glossary

Labor Geography The study of labor markets, work processes (both waged and unwaged), and labor mobilization as they are constituted in places, across space, and at various scales.

Labor Regulation The institutional and cultural means by which contestation within labor markets and labor processes is stabilized.

Remittances The monetary transfers that migrants make to their places of origin.

Reproductive Labor The unwaged domestic work that is necessary for the sustenance of a current workforce and the raising and socialization of a future workforce.

Segmentation The pattern of job allocation among workers in a labor market, usually referring to the over- and under-concentration of certain groups (e.g., by gender or ethnicity) in particular kinds of employment.

Transnationalism The phenomenon of multistranded economic, political, cultural, and social connections maintained by international migrants with their places of origin.

Labor Geography

Labor geography comprises the study of labor markets, work processes (both waged and unwaged), and labor mobilization as they are constituted in places, across space, and at various scales. Although issues of labor supply have always featured in economic geography, it is only since the early 1990s that a wider sense of the issue has emerged as a distinctive field.

In classical formulations of economic geography, labor was simply incorporated as one component of the production process. Thus, when locating a factory or retail store, or when explaining processes of urban and regional development, labor was an input with specific endowments of skill and a particular cost. Marxist approaches to economic geography, emerging in the 1970s, viewed labor in class terms. That is, not simply as an input for production, but as one side of a contested relationship between owners (and managers) of firms and their workers. Labor was thus conceptualized as part of a structural process and subject to the imperatives or logics of a capitalist system.

In the 1990s several conceptual developments saw a shift in the way that labor was viewed. First, through the influence of regulation theory, the 'labor market' came to be seen as a social regulated and embedded institution rather than something that operated through the 'natural' workings of the market, or the abstract dictates of capitalism (although a structural understanding of capitalism still largely underpinned this regulationist approach). To this, geographers contributed an understanding of labor markets as regulated 'locally', through spatially uneven institutions and practices. Attention to the particularities of place added new depth to labor market studies.

Second, the work of feminist economic geographers differentiated the labor market experiences of men and women and sought to dissolve the traditional distinctions made between spheres of 'production' and 'reproduction'. In this way, gendered expectations concerning unwaged labor in the home were seen as central to understanding differences in sectoral and occupational concentrations in the waged labor market, and uneven levels of pay and mobility.

Third, but also emerging in part from feminist approaches, closer attention has been paid to the ways in which workers are not simply robotic components of a production process, but are instead embodied actors with gender, race, ethnicity, sexuality, religion, citizenship, and other identities. These axes of difference are important in understanding how individuals are allocated to different kinds of jobs and how discipline in the workplace is achieved. Furthermore, embodied workers don't just carry out functions in the labor market, they also 'perform' their roles in bodily ways that draw upon and reinscribe these various forms of difference. In service sector work in particular, it has been shown that 'looking' and 'acting' the part is inseparable from the objective 'skills' required in determining how jobs are allocated.

Finally, the emergence of labor geography has been most closely associated with a drive to give appropriate recognition to the role that organized labor plays in shaping the economic landscape. Thus, when unions tackle employers locally, regionally, nationally, or even transnationally, they actively contribute to the spatial structure of economic activity in a way that was hardly recognized even in radical approaches to economic geography. This might take the form of forcing employers to structure the geography of their operations in a particular way (e.g., when a North American auto union bargains for new investment in a particular plant during collective agreement negotiations), or it might involve the creation of a certain culture of industrial relations in a locality that drives away certain employers but attracts others.

These various branches of labor geography have together placed greater emphasis on how labor markets, work processes, and industrial relations actually operate in the real world. To gain this insight, studies have generally focused on the specificity of particular places, and have therefore provided a corrective to the universalizing tendencies of both classical economic theory and its Marxist adversary.

Transnationalism

Despite the importance of understanding labor markets, work processes, and industrial relations within localities, it is important to remember that they are also shaped by various flows across space. Of particular interest here are those flows that cross national borders. These might be flows of investment capital, traded commodities, or organizational processes in transnational firms. But they might also be transnational flows and connections established by workers themselves. In recent studies of migration, it is these latter flows – the social, economic, political, cultural, and organizational connections established by migrants – that have spurred a renewed interest in transnationalism. In many respects they are the inverse of the transnationalism represented by large corporations and intergovernmental organizations, representing instead a form of transnationalism 'from below'.

Migration has always been a part of human history, but the magnitude and geographical scope of contemporary flows rival anything that has been seen before. Between 1975 and 2005, the estimated number of international migrants around the world increased from 84.5 million to around 196 million. This represented an increase from 2.2% to 3.0% of the world's population. But international migrants do not simply leave their countries of origin and set up a new life elsewhere. They often maintain a variety of social, economic, cultural, and political ties with their home country. The most tangible of these are financial remittances sent to households back home. In many countries these now represent a significant contribution not just to their families and communities, but also to the national economies in which they are spent. Remittances sent through formal channels alone (and the real amounts would be larger still) increased from US$46 billion in 1990 to an estimated US$276 billion by 2006. To get a full sense of the transnational ties maintained by migrants, however, we would also have to note the many strands of connection that migrants maintain – information relayed, cultural expectations transmitted, political alliances and support networks forged, family and household relations stretched, and citizenship rights dislocated.

This article will examine four ways in which economic geographers have demonstrated the intersections between migrant transnationalism and the various dimensions of the workplaces and labor processes described above. The first concerns research that has explored how migrant transnationalism affects the working of labor markets, especially those in major gateway cities where migrants have tended to settle, as well as other cities where temporary migrant workers form a significant component of the labor force. The second relates to the effects of migrant transnationalism on the operation of workplace labor processes. The third explores how transnational ties often stretch households across global space and can affect the way in which unwaged labor is allocated within the home. The fourth examines how workers have themselves organized transnationally in a number of instances.

Transnationalism and Labor Markets

How do migrants' transnational connections affect the operation of the labor market and their experiences within it? The first point concerns the way in which information about the labor market circulates. Such information might include intelligence about employment opportunities arising, necessary training and credentials, contacts inside potential employers, or regulatory issues around welfare unemployment benefits. On the employer's side it might include knowing who is available with the skills to do a particular job and knowing how to assess foreign credentials.

It is often assumed that markets, including those for labor, are efficient at circulating such information between buyers and sellers, but in fact information travels through particular channels and the most effective of these channels in the case of labor markets are the social networks of existing employees. But where employees are immigrants, especially recent immigrants, flows of information are not just among social networks within an urban labor market, but also to migrant source areas around the globe. Thus, in the case of Taiwanese professionals recruited to work in the electronics and software industry in California's Silicon Valley, it is frequently trans Pacific social networks that supply the talent pool. The same is true in less secure and less well paid jobs. In Vancouver, Canada, for example, immigrants from India's Punjab dominate the airport taxi sector, largely because social networks extending back to India have become the prime channel through which recruitment occurs. In the UK, Ireland, and elsewhere, Filipino nurses are present in increasing numbers in hospital and clinics, in part because information flows have been established with sending areas in the Philippines. The result in the Philippines has been the massive expansion of nursing training programs (as they become viewed as a ticket to emigration) and a crisis in the Philippine healthcare sector as so few experienced nurses remain. In this way, then, transnational flows of labor market intelligence drive labor market processes at both ends of the migration process. An important corollary of this is that migrants from particular places become destined for certain kinds of jobs in certain sectors, even before they have made their journey to their host societies.

Once employed, migrants often have transnational commitments and obligations to their family members back home – to repay money loaned, to support a household, or to finance the immigration applications of family members left behind. These commitments place an intensified pressure on the migrant to participate quickly in the host city's labor market (and without being too choosy about which job to accept). Transnational obligations may also require that a migrant intensifies her participation in the labor market by holding down several jobs. The result, then, is an intensification of participation in the workforce and curtailed job searches in which immediate rather than appropriate employment is the imperative upon arrival. Given that such immediate employment is likely to be in low paid precarious service sector jobs, and is likely to be found through networks of migrant compatriots, the result is a perpetuation of patterns of segmentation in the labor market.

While labor market information flows and participation decisions may both incorporate a transnational component, an even more significant issue concerns the ways in which a transnational migrant's status as non citizen (and possibly as a racialized minority as well) shapes the conditions of their integration into the labor market. In contexts where transnational migrants are incorporated into the labor force as temporary workers rather than immigrants, there is a dislocation between their citizenship rights at 'home' and their labor market participation in a 'host' society. Singapore's labor market is predicated upon such distinctions to an unusual extent. In 2000, almost 30% of Singapore's work force (and about one quarter of its total population) were noncitizens. Furthermore, the city state implements a system whereby the rights of foreign workers are dictated by a graduated system of visas and work permits, so that highly paid foreign workers in areas such as financial services and academia can be accompanied by their families and can exercise most of the same rights as citizens. Low paid temporary workers in domestic service work and construction, on the other hand, are regulated by a strict set of conditions that curtail their rights under the constant threat of deportation. While Singapore is an extreme example, the threat of deportation that hangs over non citizens in every immigrant city creates a potentially exploitable and marginalized workforce.

Issues of citizenship rights and visa conditions present clear and tangible ways in which a transnational life (living and working in one place, but rooted and enfranchised with citizenship in another) can lead to particular labor market outcomes. But individuals also make subjective decisions about jobs, careers, work hours, etc., based on cultural expectations. Hence, it is important to consider how the construction of identity is forged in the labor market through a transnational consciousness. What this means in practical terms is that jobs, salaries, class, and status are often evaluated by migrants in relation to distant comparisons and expectations. For deprofessionalized and subordinated transnational migrant workers, this means that when salaries are translated into purchasing power back home, when personal status is gauged based on the respect accorded in one's home rather than one's host community, and when compatriots have gone through similar degrading experiences, expectations within the labor market may be adjusted.

Transnationalism and Workplace Processes

The preceding section established some of the ways in which the transnational connections maintained by migrants affect their integration into host labor markets. But transnational ties may also influence the micro politics of the workplace and thereby shape the ability (or inability) of migrant workers to assert their rights, demand better conditions of employment, or seek advancement up the managerial hierarchy.

For many migrant workers in low paid and insecure employment, it is precisely their transnationalism that undermines their capacity to assert rights in the workplace. As citizens of one territory but workers of another, they may only be protected in theory by labor laws. In practice, the precariousness of their migrant status and their transnational obligations to family members back home, mean that there would be a heavy price to pay for pushing back against abusive employers or the authorities of host states. In Canada, research on the working conditions of Filipina domestic workers (many employed as 'live in' caregivers for young children) has shown that women, whose employment, visa, and even food and shelter, are all dependent on their employers, are vulnerable to abuse of various forms – from inadequate provision of private living space, to compulsory overtime, to physical abuse. It is frequently the obligations of such workers to their transnational households and their lack of citizenship rights that forces them to endure such abuses.

A less tangible form of transnational connection may also inhibit the advancement of immigrants in their workplaces. In addition to their financial obligations to family members 'left behind', immigrants also bring with them a different set of cultural codes and norms of workplace practices. These might include expectations in terms of teamwork or individualism, culturally acceptable levels of self promotion, and norms with respect to requesting promotions, pay increases, or time off. These are practices learned in one cultural context (the source country) but implemented and evaluated by others, in another context. Hence, a transnational juxtaposition of cultural codes can create circumstances where an immigrant's workplace practices are read as being inappropriate for advancement and upward mobility. In the immigrant receiving countries of the 'Global North' these are frequently cultural barriers that exist for racialized minority groups in particular. While not rooted in tangible transnational ties, they are very much the product of immigrant workers being seen as 'of' some other place.

Transnational ties may, however, have the opposite effect and can, in limited ways, empower employees in their workplaces. An example is provided by research on the information processing industry in Jamaica which provides services such as data entry and software code modification. The workforce in this sector is largely young, female, and nonunionized. The worker–manager relation is one of constant tension. While few workers have overtly resisted work conditions, many have engaged in acts of everyday resistance that may hamper the ability of firms to meet contractual deadlines for their clients. Examples include absenteeism, working to rule, rejecting overtime when offered, and 'finger dragging' (i.e., deliberately slowing the work rate when basic targets have been met). These are activities that can actually reduce the wages that workers receive, and may indeed put their employment in the industry in danger. The viability of these strategies – and the psychological boost that workers may receive from them – can only be understood in terms of the transnational household connections of workers in this Jamaican context. Remittances from overseas relatives provide a cushion against the low wages and harsh discipline of the information processing industry, and embolden workers to engage in workplace resistance that would otherwise be too risky.

Transnationalism and Reproductive Labor

Transnational migrants at various levels in the labor market are frequently separated from their immediate families. This sets up a situation in which the reproductive work of running a household, maintaining a home, and raising children, is often displaced from the place of work of the household's principal earner. This transnational household arrangement often implies a significant restructuring of the division of unwaged domestic labor.

In one study, researchers examined the household dynamics involved in transnational labor migration by professional and managerial Singaporeans into China. Intra household gender power dynamics are heavily impacted by this process, but so too are state ideological frameworks around household definitions and gender roles. The reproductive activities of professional and managerial families must be negotiated and frequently the burden of making this adaptation falls to wives who, whether they engage in waged work or not in Singapore, still take on the lion's share of such duties. In another study, female Ghanaian immigrants to Canada engage in 'transnational mothering' as a result of women migrating to work in Canada but deciding to leave their children behind in the care of relatives.

Researchers have also found examples where domestic reproductive work is located in the immigrant host society, while waged work is conducted in the sending country. This occurs in the case of so called 'astronaut' families. Governments in Canada and Australia have established 'business immigrant' programs in which individuals are admitted based on the investment capital they have available and their ability to establish and run entrepreneurial ventures. In practice, many such immigrants (most of who have come from, Hong Kong, South Korea, and Taiwan) find it difficult to run successful businesses in a different cultural and institutional environment. As a result, many return to their places of origin to continue business careers there, but leave their families (meaning, in nearly all cases, wives and children) to live in Vancouver, Melbourne, etc. In such circumstances, waged work in one place is linked with unwaged work in another place through a transnationalized household structure.

Transnational Organizing

As labor geography emerged as a distinctive field in the 1990s, a core concern was the ability of organized labor, in the form of trade unions, to shape the geography of capitalist development. In this literature, transnationalism refers less to the movement of migrants and the ties that they maintain across space (as in the preceding three sections), and more to one possible scale at which labor movements might conduct campaigns and even constitute themselves organizationally.

For some researchers, transnational organization has been shown to be a necessary response by trade unions to the global spread of corporate power. Certainly it is true that transnational unionism has yielded some important successes for the labor movement. In a dispute in 1996–98, for example, Liverpool dock workers forged international relationships to undermine the Merseyside Dock and Harbour Company during a dispute concerning casualization. Union activists organized several successful international days of action in 1996 and 1997 that saw Swedish, French, Danish, Canadian, American, and Australian workers taking action to target ships and operators connected with Liverpool. International sup port proved to be a potent weapon for the dockworkers in their dispute, increasing the scale of the dispute and directly threatening the employer's profits.

In another example, a local of the United Steelworkers of America (USWA) in Ravenswood, West Virginia, took on the metals trading empire of the international finan cier Marc Rich. The dispute arose in 1989 when an aluminium plant was purchased by investors who over saw the immediate sacking of 100 workers, increased the pace of work in the factory, and disbanded a joint worker–manager safety committee. Over the course of the next 18 months, five workers were killed and several injured in the Ravenswood plant. In a subsequent round of bargaining, union workers were locked out and de cided to take up the fight through their national union (the USWA) and the American Federation of Labor Congress of Industrial Organizations (AFL CIO). The campaign began to spread when it was discovered that the new investors were lead by Rich, a notorious financier living in Switzerland and wanted in the US on tax fraud charges. This led to a transnational campaign that in volved worldwide lobbying of over 300 end users of Ravenswood metals to stop buying the product. There after, in a series of actions involving unions in Western Europe, attempts were made to disrupt economic activ ities associated with Rich. By the spring of 1992, anti Rich activities had been organised in 28 countries on five continents, with the result that in June of that year, the management was forced to concede defeat and to re instate the sacked workers.

Studies tracing the transnationalization of unions and their campaigns emphasize, however, that the optimal scale for prosecuting labor disputes need not necessarily incorporate transnational networks. In many circum stances and in many sectors, local and national scales are the key pressure points for asserting the demands of workers.

Research in labor geography has also shown that transnational labor activism need not necessarily come in the form of organized labor. Other examples of transna tional mobilization of, or for, labor include activist or ganizations that assert the rights of workers across national borders, either through advocating for those workers collectively, or through consumer campaigns that seek to influence labor processes in distant pro duction sites. Numerous examples of international labor advocacy organizations exist outside of the formal trade union movement, including some that focus explicitly on the plight of migrant workers. One such organization, Migrant Forum Asia, was formed in 1994 as a network of regional nongovernmental organizations. Its purpose is to forge "concerted action to address discriminatory laws and policies, violence against women migrants, unjust living conditions, unemployment in the homeland and other issues affecting migrant workers." Significantly, the organization's mandate includes the issue of unemploy ment in sending countries, which drives transnational migration in the first place.

Another case of transnational labor activism is found in consumer campaigns to ensure labor standards are upheld in distant production sites. The Clean Clothes Campaign, for example, focuses on improving working conditions in the global garment and sportswear indus tries. With an international secretariat in Amsterdam, it also has local campaigns in 11 Western European coun tries: Austria, Belgium, France, Germany, Italy, the Netherlands, Norway, Spain, Sweden, Switzerland, and the United Kingdom. A similar campaign, with a more North American focus is the Maquila Solidarity Network (MSN), based in Toronto but focusing on working con ditions in the garment sector in Mexico, Central Amer ica, and Asia. The MSN defines its aim as being: "to strengthen the capacity of civil society organizations in the North and South to challenge the negative impacts of industry restructuring in the global garment industry. In a global economy, groups in the North and South must work together for employment with dignity, fair wages and working conditions, and healthy workplaces and communities." In both cases, transnational internet based fundraising and campaigns form an important channel for placing pressure on employers, retailers, and govern ments to address workplace exploitation.

Conclusion

This article has explored the intersections between transnationalism and labor geography in various ways. In each case, transnationalism has referred to the ways in which workers establish transnational ties and networks to create form of globalization 'from below'. The first three sets of examples related to the transnational ties maintained by migrants, while the last related to trans national organizing by, or on behalf of, workers.

In the case of labor market processes, the transnational circulation of information shapes both migrant flows supplying the labor market, and in some cases the nature of training infrastructure in sending countries as well. Financial obligations to family members back home may also affect a migrant's participation in the labor market. In cases where migrant are permitted to work on tem porary visas only, then the transnational dislocation of working lives from citizenship rights creates a vulnerable position for migrant workers in the labor market.

This vulnerability can be translated into the work place, where obligations to home, and a lack of citizen ship rights, can leave transnational migrants open to abuse. In addition, the enactment of embodied cultural practices learned from home contexts can frequently leave migrants marginalized, underestimated, and dis criminated against in their new workplaces. In some in stances, though, transnational sources of livelihood can support a higher degree of labor resistance, as the ex ample of Jamaica illustrated.

Labor market studies often relate to work that is sold to an employer for a wage, but every waged employment relationship is underpinned by a set of unwaged labor processes that sustain and reproduce the workforce, primarily in domestic spaces. In some instances, these reproductive forms of labor are conducted within households that are stretched transnationally, so that waged productive labor and unwaged reproductive labor are dislocated. The consequence of this is a reordering of divisions of reproductive labor within the household.

Finally, labor geography's core concern with the agency of workers has a transnational dimension in the case of disputes and campaigns that expand out of local scales to encompass global networks of allies. Examples indicate some successes when labor asserts itself trans nationally in this way.

In all of these processes, our understanding of labor geographies has been deepened through attention to the ways in which local processes are constituted, in part at least, by their connections across space with distant places. Clearly this does not imply that labor geographies should be analyzed, nor labor activism engaged, in a way that prioritizes transnational processes. But it does imply that local and national scales alone are not sufficient to develop a full understanding of the geographies that workers create.

See also: Immigration II; Labor Market; Remittances; Transnationalism.

Further Reading

Bauder, H. (2005). Habitus, rules of the labor market and employment strategies of immigrants in Vancouver, Canada. *Social and Cultural Geography* 6, 81 97.

Bauder, H. (2006). *Labor Movement: How Migration Regulates Labor Markets*. Oxford: Oxford University Press.

Castree, N. (2000). Geographic scale and grass roots internationalism: The Liverpool dock dispute, 1995 1998. *Economic Geography* 76, 272 292.

Herod, A. (2001). *Labor Geographies*. New York: Guilford.

Kelly, P. F. and D'Addario, S. (2007). Filipinos are very strongly into medical stuff: Labour market segmentation in Toronto, Canada. In Connell, J. (ed.) *The International Migration of Health Workers*. London: Routledge.

Kelly, P. F. and Lusis, T. (2006). Migration and the transnational habitus: Evidence from Canada and the Philippines. *Environment and Planning A* 38, 831 847.

McDowell, L., Batnitzky, A. and Dyer, S. (2007). Division, segmentation, and interpellation: The embodied labors of migrant workers in a Greater London hotel. *Economic Geography* 83, 1 26.

Mullings, B. (1999). Sides of the same coin? Coping and resistance among Jamaican data entry operators. *Annals of the Association of American Geographers* 89, 290 311.

Pratt, G. (1999). From registered nurse to registered nanny: Discursive geographies of Filipina domestic workers in Vancouver, BC. *Economic Geography* 75, 215 236.

Sassen, S. (1995). Immigration and labor markets. In Portes, A. (ed.) *The Economic Sociology of Immigration*, pp 87 126. New York: Russell Sage Foundation.

Saxenian, A. (1999). *Silicon Valley's New Immigrant Entrepreneurs*. San Francisco, CA: Public Policy Institute of California.

Waters, J. (2002). Flexible families? 'Astronaut' households and the experiences of lone mothers in Vancouver, British Columbia. *Social and Cultural Geography* 3, 117 134.

Wills, J. (1998). Taking on the cosmocorps: Experiments in trans national labor organization. *Economic Geography* 74, 111 130.

Willis, K. and Yeoh, B. (2000). Gender and transnational household strategies: Singaporean migration to China. *Regional Studies* 34, 253 264.

Wong, M. (2006). The gendered politics of remittances in Ghanaian transnational families. *Economic Geography* 82, 355 381.

Yeoh, B. (2006). Bifurcated labor: The unequal incorporation of transmigrants in Singapore. *Tijdschrift voor Economische en Sociale Geografie* 97, 26 37.

Relevant Websites

http://cleanclothes.org
Clean Clothes Campaign.
http://www.iom.int
International Organization for Migration.
http://en.maquilasolidarity.org
Maquila Solidarity Network.
http://www.mfasia.org
Migrant Forum Asia.

Transnationalism and Technology Transfer

J.-Y. Hsu, National Taiwan University, Taipei, Taiwan

Glossary

Industrial Upgrading As more and more developing country producers are integrated into global markets, there is downwards pressure on the prices of both agricultural and manufactured products. For producers to maintain or increase incomes in the face of this pressure, they must either increase the skills content of their activities or move into market niches which have entry barriers and are therefore insulated to some extent from these pressures. The shifts in activities which sustain higher incomes are taken as upgrading.

Knowledge Knowledge is viewed fundamentally as a heterogeneous resource that firms value in different manifestations. Mostly the main types of knowledge can be distinguished as codified knowledge versus tacit knowledge, and individual knowledge versus collective knowledge.

Overseas Diaspora It refers to ethnic groups whose sizable parts have lived outside their country of origin for at least several generations, while maintaining some ties to the homeland.

Structure Hole Burt defines it as "a gap between two individuals with complementary resources or information." When the two are connected through a third individual as entrepreneur, the gap is filled creating important advantages for the entrepreneur. Competitive advantage is, therefore, a matter of access to these 'structural holes'.

Technology Transfer Students of technology transfer make an important distinction, separating 'material transfers' and 'design transfers' from 'capacity transfers'. While technology transfer involves management and investment, it is difficult to rely exclusively on the transfer of machines and blueprints. Therefore, the migration of engineers must be considered an essential element in the effective transfer of technology.

Transnational Technical Community A community that spans borders and boasts as its key assets share technical knowledge and ethnic networks.

There has been a resurgence of interest in the implications of innovation and learning for the creation and maintenance of economic competitiveness since the new economy, or knowledge based economy, became the buzzword in the past two decades. Two types of knowledge have been identified by the students of knowledge management: on the one hand, codified knowledge is necessarily explicit, formal, or systematic, and can be expressed in words and numbers, scientific procedures, or universal principles. The codified type of knowledge is easy to transfer, store, recall, and valorize. On the other hand, tacit knowledge is extremely difficult to transfer, and is gathered from the accumulation of practice, the ability to communicate, and wisdom. Johnson *et al.* raised a critical assessment of the dualism of codified/tacit knowledge used by Cowan *et al.* They argued the dichotomy was problematic since it was rare that a body of knowledge could be transformed into codified form without losing some of its original characteristics and that most forms of relevant knowledge were mixed in these respects.

Under the influence of new information technology and global knowledge economy, codified knowledge becomes ubiquitous, and the creation of unique capabilities and products hinges on the production and use of tacit knowledge. It is asserted that knowledge creation and diffusion should be taken as a situated practice embedded in distinctive communities and actor networks in which the powers of context, spatial and temporal, have to be placed at the center. Accordingly, the combination and composition of tacit and codified knowledge depend on the context within which agents or organizations manipulate knowledge. While codified knowledge can be disseminated through impersonal means, tacit knowledge is closely associated with meaning and understanding, and personal communication is indispensable for its transfer. In the times we now live, when communication technology is highly developed and information is widely available, interpersonal connections have become more, rather than less, important for the exchange of knowledge.

But we recognize that knowledge has the properties of being embedded in an individual, who is part of an organizational context. The consequence of combining personalized knowledge and organizational embeddedness is that knowledge can be at either end of the tacit codified continuum and may therefore be straightforward or almost impossible to transfer.

In fact, the mobility of personnel as the most effective way to overcome the problem of tacit dimension of knowledge transfer has been well identified. One of the major goals of technology transfer is to cultivate local technological capabilities. (According to Bell and Pavitt (1993: 163), technological capabilities consist of the resources needed to generate and manage technical change, including skills, knowledge and experience, and institutional structures and linkages.) Students of technology transfer make an important distinction, separating 'material transfers' and 'design transfers' from 'capacity transfers'. Material transfer is characterized by the import of new materials and techniques. Local adaptation is not conducted in an orderly and systematic fashion. The local adaptation of borrowed technology and the development of new machines tend to occur primarily as a result of trial and error, often termed 'learning by doing'. Design transfer is primarily carried out through the transfer of certain blueprints, formulas, books, etc. The knowledge contained in these design materials is coded and explicit, and must be adapted to local conditions. Capacity transfer means the transfer of scientific knowledge, which leads to the production of locally adaptable technology, based on technology prototypes existing abroad. A critical element in the process of capacity transfer is the migration of scientists and engineers, as most of the innovative knowledge is human embodied and diffuses through personal contact and association. While technology transfer involves management and investment, it is difficult to rely exclusively on the transfer of machines and blueprints. As Almeida and Kogut track the movements of over 400 engineers and show their patterns of mobility influence the inter and intraregional patterns of knowledge flow.

It is particularly relevant to the knowledge economy in which the access to knowledge becomes the key for the construction of competitive advantage. According to Lundvall, the learning economy includes mastery over the know what of facts and information, the know why of principles and theorems, the know how of competence and skills, and the know who of knowledge in networks of collaboration and communication. Moreover, as Lundvall argued, the learning process involves more than purchasing technology, and includes social dimensions such as the absorption of tacit knowledge, which is embodied in technical staff. As a result, the know who became the key job in identifying appropriate technologies to transfer.

The importance of the know who illustrated a critical 'gatekeeper' role in technology transfer. Studies using a social network approach to innovation and product development determined that strategically positioned individuals facilitate information dissemination which in turn facilitates innovation. Individuals with more informal contacts outside the organization, or 'gatekeepers', were critical for importing novel information and linking the organization with its environment. These gatekeepers effectively serve as the primary link to external sources of information and technology.

The role of gatekeeper well epitomizes why some companies seem to be able to adopt technologies earlier than others. This could be related to their absorptive ability to screen their environment and identify new potential solutions from external sources. One way to achieve this is to have contacts in the supporting scientific community. These boundary spanning individuals, or gatekeepers, are most likely well connected and informed. They have insight into the related technologies and underlying sciences.

The Rise of Transnational Technical Community and Technology Transfer

Based on a perspective of social network, a growing body of transnationalism, or globalization from below, tried to decode the technology diffusion and transfer which were usually engaged exclusively with the transnational corporations and nation states. It contended that a transnational community of engineers has coordinated a decentralized process of reciprocal industrial upgrading by transferring capital, skill, and know how to the source region and by facilitating collaborations between specialist producers in the two regions. The transnational community thus allows local companies from latecomer industrializing countries to avoid the problems that many corporations face when they establish operations in the technology hub such as Silicon Valley. Foreign firms need to be able to integrate into the region's social networks to gain access to up to date technology and market information, while simultaneously maintaining the ability to communicate quickly and effectively with decision makers in the headquarters. It argues that the multinational corporation may no longer be the advantaged or preferred organizational vehicle for transferring knowledge or personnel across national borders, but a transnational technical community provides an alternative and potentially more flexible and responsive mechanism for long distance transfers of skill and know how – particularly between very different business cultures or environments.

As demonstrated by Amin and Cohendet, two major advantages are exploited by the governance mode of community. The first is that communities 'freely' absorb the sunk costs associated with building the infrastructure needed to produce or accumulate knowledge, usually in a completely nondeliberate manner embedded in their daily practices which render the codification of tacit knowledge easy and costless. The second is that communities do not need a visible or explicit central authority to control the quality of work or enforce compliance with any standard procedure, as communities monitor the behavior of their members and render them accountable for their actions. By the same token, these advantages apply to the transnational technical communities in the current global economic system.

First, an increasing specialization of production and a deepening social division of labor generates entrepreneurial opportunities for innovation in formerly peripheral regions. By exploiting these opportunities in their home countries, a transnational technical community can build independent centers of specialization and innovation, while simultaneously maintaining ties to the technology hubs such as Silicon Valley to monitor and respond to fast changing and uncertain markets and technologies. The community was well positioned to establish cross regional partnerships that facilitate the integration of their specialized components into end products. In the transnational community, ethnic ties and interpersonal relationships can facilitate collaborations and reduce the uncertainty of economic deals. The members of the transnational ethnic network are likely to be better informed on the capabilities and requirements of domestic labor and the sort of training local labor requires. In other words, such social ties fulfill the need of ‘know who’ in the learning economy in which the social dimension is the key and often ignored issue in the constitution of competitiveness. These transnational relationships support technology transfer by supporting joint problem solving and complementary innovation. Close, trust based relationships among the transnational community of engineers are thus an essential precondition for the flexible collaboration needed to adapt and survive in today’s fast paced competitive environment. The case study of the Silicon Valley–Hsinchu (Taiwan) connection vividly verifies the point. In addition, the development of the Indian software industry owes much of its growth to the Indian diaspora in the Silicon Valley, as shown by Kapur.

Second, accounting for technical upgrading in late industrializing regions is a contested issue. In fact, as discovered by Saxenian, the links of the key technological latecomers such as Taiwan, India, and Ireland, with the Californian technology hub unfold in several ways: the latecomer companies recruit overseas engineers, they set up listening posts in Silicon Valley to tap into the brain power there, or successful overseas engineers return to the homelands to start up their own businesses. All of these possible links are established smoothly not only on an individualistic basis, but largely through the mediation of overseas organizations. In spite of ethnic ties that facilitate cross regional technological cooperation, the technical community benefits more from integration with broader business networks. It is clear that the overseas diaspora helped transfer technology and business models back to the homelands. It is particularly true that while the latecomer firms in both regions had to rely on ethnic ties with mainstream businesses during their embryonic stages, ethnic ties could ease information collection across the globe. Trust and reciprocity incubated from primary ethnic bonds and informal personal relationships facilitate cooperation between these regions, and broaden the scope of network building. While some top down accounts, such as that of Alice Amsden suggest that the developmental state and key big companies (national champion) should be put at center stage in the process of late industrialization, other accounts focus on global production networks (GPNs) and argue that late development benefits from its insertion into global value chains. Both are partially true, but fail to take seriously the embedded institutions in transborder connections between technologically leading and following regions. In fact, a new strand of GPNs perspective has more detailed and nuanced analysis of the social and developmental dynamics of contemporary capitalism at the global–local nexus, which becomes the mandate for the transnational corporations to survive and prosper in the interconnected economic system. The GPN perspective asserts that strategic coupling between GPNs and regional assets, or an interface mediated by a range of institutional activities across different geographical and organizational scales will be critical for the development of each region in the global economy. At the same time, the transnationalist discourses insist on the continuing significance of national borders, state policies, and national identities, while simultaneously crossing over them, and constituting a hybrid social space. Instead of maintaining a kind of zero sum assumption of exclusiveness of nation states and globalization, the transnationalist discourses take them as engaging in a process of mutual construction. In light of the GPN perspective, the transnational technical community plays a critical mediating role in the globalizing process of regional development.

The Limit of Transnationalism in Technology Innovation

As Amin and Cohendet warns, governance by community does not come without limits. One of the major causes of failure is the risk of parochialism, discrimination, or vengeance on other communities. A second limit is the risk of lack of variety. These cautious remarks apply to the mediating work of transnational technical community in technology innovation and transfer.

In fact, as shown by social network analysis, networks with an abundance of structural holes, a gap between two individuals with complementary resources or information, by situating people at the confluence of different social domains, create opportunities for the novel combination and recombination of ideas. These same networks, however, pose a problem for acting on such ideas. Structural holes pose an action problem because the dispersed, unconnected people found around structural holes are inherently more difficult to mobilize or coordinate, especially around novel ideas. Dense networks present the optimal conditions for the exchange of the complex information necessary for innovation in complex organizations but present an idea problem because of the redundancy of information circulating within the network.

Individuals who are active in introducing dissimilar others and facilitating action among previously tied alters (or people in one's social network) will be more involved in the combinative activity that leads to innovation. The information advantage Burt associated with networks with structural holes roughly corresponds to the advantageous position occupied by the gatekeepers in the earlier social networks/innovation literature to the extent that both imply boundary positions. Extensive social knowledge about the personnel and differing styles of critical departments across the organization resulting from informal ties and potentially shared design experience contribute critically to cross boundary innovation efforts. In other words, knowledge heterogeneity was a significant predictor of both overall managerial performance and innovation performance.

As a result, a number of limits should be raised about the relations of transnationalism and technology transfer. First, the transnationalist argument risks oversocializing economic behavior that is rooted in business and technological considerations. This is the blind spot in the discourse on social constitution, which assumes that social relationships determine economic transactions and outcomes. The economy is not reducible to interpersonal relationships, but composed of multiple production worlds that are defined by product configuration, market principles, and technology and production process. In other words, dense social ties cannot substitute for the sophisticated managerial and technological learning that is required to compete in a particular sector in spite of the fact that the social dimension of learning is critical.

Second, close ties sometimes will become blind trust, and make firms unconscious of the exterior technological breakthroughs or new business opportunities. To make things worse, ethnicity embedded system occasionally compels people to compete based more on the thickness of tie than the depth of capability. In other words, a socially overembedded industrial world without cautious monitoring might lead to unproductive situations, rather than a healthy and efficient production system. The peril of lock in is usually put aside by the transnationalist (as well as networked firm) argument.

In light of the discussion above, two interesting threads, among others, could be elaborated in the future research. The one is a comparative study between the performances of transnational technical communities in divergent 'home countries' which demonstrate contradictory industrial structures and absorptive capabilities. Another thread lies in the combination of GPN perspective and transnationalist discourse by the use of social network analysis. By doing so, a number of typologies of global–local interface could be built up for the advanced study of transnational technology transfer.

See also: Transnationalism.

Further Reading

Almeida, P. and Kogut, B. (1999). Localization of knowledge and the mobility of engineers in regional networks. *Management Science* 45(7), 905 917.

Amin, A. and Cohendet, P. (2004). *Architectures of Knowledge: Firms, Capabilities, and Communities*. Oxford: Oxford University Press.

Amsden, A. (2001). *The Rise of the Rest: Challenges to the West from Late Industrialization Economies*. Oxford: Oxford University Press.

Bell, M. and Pavitt, K. (1993). Technological accumulation and industrial growth: Contrasts between developed and developing countries. *Industrial and Corporate Change* 2(2), 157 210.

Bunnell, T. and Coe, N. (2001). Spaces and scales of innovation. *Progress in Human Geography* 25(4), 569 589.

Burt, R. (2001). Structural holes versus network closure as social capital. In Lin, N., Cook, K. & Burt, R. (eds.) *Social Capital: Theory and Research*, pp 31 56. New York: Aldine De Gruyter.

Coe, N., Hess, M., Yeung, H., Dicken, P. and Henderson, J. (2004). 'Globalizing' regional development: A global production networks perspective. *Transactions of the Institute of British Geographers* 29, 468 484.

Cohen, W. and Levinthal, D. (1989). Innovation and learning: the two faces of R&D. *Economic Journal* 99, 569 596.

Connell, N., Klein, J. and Powell, P. (2003). It's tacit knowledge but not as we know it: Redirecting the search for knowledge. *Journal of the Operation Research Society* 54(2), 140 152.

Cowan, R., David, P. and Foray, D. (2000). The explicit economics of knowledge codification and tacitness. *Industrial and Corporate Change* 9(2), 211 253.

Dicken, P., Kelly, P., Olds, K. and Yeung, H. (2001). Chains and networks, territories and scales: Towards a relational framework for analyzing the global economy. *Global Networks* 1, 89 112.

Grabher, G. (1993). *The Embedded Firm: On the Socioeconomics of Industrial Networks*. London: Routledge Press.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology* 78, 1360 1379.

Hobday, M. (2001). The electronics industries of the Asia Pacific: Exploiting international production networks for economic development. *Asian Pacific Economic Literature* 15(1), 13 29.

Johnson, B., Lorenz, E. and Lundvall, B. A. (2002). Why all this fuss about codified and tacit knowledge? *Industrial and Corporate Change* 11(2), 245 262.

Kapur, D. (2001). Diasporas and technology transfer. *Journal of Human Development* 2(2), 265 286.

Lall, S. (1987). *Learning to Industrialize: the Acquisition of Technological Capability by India*. London: Macmillan.

Lundvall, B A. (1996) The social dimension of the learning economy. *DRUID Working Paper* 96 1.

Maskell, P. and Malmberg, A. (1999). Localised learning and industrial competitiveness. *Cambridge Journal of Economics* 23(2), 167 185.

Sabel, C. (1989). Flexible specialization and the re emergence of regional economies. In Hirst, P. & Zeitlin, J. (eds.) *Reversing Industrial Decline?* pp 17 70. New York: St. Martins.

Saxenian, A. (2006). *The New Argonauts: Regional Advantage in a Global Economy*. Cambridge: Harvard University Press.

Saxenian, A. and Hsu, J. Y. (2001). The Silicon Valley Hsinchu connection: Technical communities and industrial upgrading. *Industrial and Corporate Change* 10(4), 893 920.

Smith, M. (2001). *Transnational Urbanism: Locating Globalization*. Oxford: Blackwell.

Storper, M. and Salais, R. (1997). *Worlds of Production: The Action Frameworks of the Economy*. Cambridge: Harvard University Press.

Tsoukas, H. and Vladimirou, E. (2001). What is organizational knowledge? *Journal of Management Studies* 38(7), 973 993.

Tushman, M. and Scanlan, T. (1981). Boundary spanning individuals: Their role in information transfer and their antecedents. *The Academy of Management Journal* 24(2), 289 305.

Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly* 42, 35 67.

Transnationality

S. Huang, National University of Singapore, Singapore

Glossary

Networks These may be conceived structurally as webs through which information, values, capital, people, and goods flow. Social networks (often based on familial, friendship, or ethnic ties) are particularly important in mediating the volume and direction of transmigrant flows.

Simultaneity It reflects the everyday geographies that incorporate activities, routines, and institutions located both in a destination country and transnationally, given that transmigrants' embeddedness in multisited transnational social fields encompass those who move as well as those who stay behind.

Transnationalism from Above/Below It refers to transnational activities and enterprises initiated and carried out by structural forces and institutions (such as states and transnational corporations) and those resulting from grassroots initiatives (including those) by individuals respectively.

Transnationality This is a condition of sustained interconnectedness experienced by transmigrants whose everyday geographies span social fields located in two or more nation-states. Transnationality may be best embodied in the person of the multiple passport holder who circulates easily between two or more nation-states, building a life history rooted in and affected by multiple national affiliations.

Transnational Social Spaces/Fields These move the idea of migration spaces as fixed and contained within the boundaries of a nation to one that conceptualizes space as more relational and tied to opportunity structures and migrant experiences. Social fields are thus interlocking networks of social relationships through which ideas, practices, and resources are unequally exchanged, organized, and transformed, and exist above and beyond the social contexts of national societies.

Introduction

Conceptually, the landscape of transnationality is a variegated and evolving one, researched by scholars from a variety of disciplines, especially cultural studies, anthropology, and increasingly geography, and still lacking a well defined theoretical framework. A review of the literature reveals some differing conceptualizations about what transnationalism and transnationality encompass and at times, ambiguity in how the terms (especially 'transnationality') are utilized. The idea of transnationality was perhaps first expressed by Randolph S. Bourne in an essay on 'Trans national America' published in the *Atlantic Monthly* in July 1916, which called for the acceptance of a new cosmopolitan ideal characterized by dual loyalties and the "free and mobile passage of the immigrant between America and his native land."

While the term transnational'ity' (as opposed to transnational'ism') is more often than not implied rather than explicitly employed in much of the current transnationalist scholarship, it continues to be used primarily in the context of immigrant transnationalism to emphasize the state of simultaneous embeddedness in more than one society that many transmigrants in the era of globalization experience due to changes in the nature of capitalism. The resultant restructuring of the global economy has in turn brought about new migration strategies which hinge on flexibility in terms of location and allegiances. Discussions of transnationalism almost never fail to cite the work of anthropologists Nina Glick Schiller, Linda Basch, and Cristina Szanton Blanc explicating the multistranded networks (social, cultural, familial, political, economic, and religious) that are sustained between societies of origin and settlement under conditions of contemporary international migration.

In other words, transnationality may be first and foremost thought of as a condition of sustained interconnectedness experienced by transmigrants whose everyday geographies span social fields located in two or more nation states. It is generally agreed that this simultaneity of transnational social fields has been fostered by significant technological advancements in transportation and communications – including rapid air flights (from budget to first class), electronic transmissions over the Internet and e mail, and inexpensive and even free telephone connections – which have facilitated regular and nearly instantaneous cross border flows of goods, remittances, information, and people in the contemporary period. Some scholars have even attributed causal weight to new technologies in accelerating international linkages and transforming the form of migration from older linear and circular patterns, to its current morphology in which migrants are said to move easily and participate actively in the social, business, political, and familial affairs of two or more nations, thus necessitating the development of new transnational frameworks to understand migration.

Earlier theorizing on transnationalism, mainly by cultural anthropologists, centered around the ideas gen erated by Glick Schiller, Basch, and Szanton's work, and highlighted a transnationality of multiply located lives that contemporary transmigrants participate in, as they, unlike their predecessors, continue to actively engage in the societies from which they emigrated even as they invest in building new lives in their new societies. In this conceptualization, transnationality may be best embodied in the person of the multiple passport holder who cir culates easily between two or more nation states, build ing a life history rooted in and affected by multiple national affiliations. Thus, the transnational is often one who has multiple (perhaps ambivalent and even con flicting) loyalties as opposed to exclusive allegiances. However, transnationality cannot be limited to an in terrogation of the state–citizen relationship and simply confined to formal conferment of citizenship. As Aihwa Ong has argued, the form and meaning of citizenship has been reworked by globalization, and many traveling subjects may also employ the locational strategies and tactics of 'flexible citizenship' to capitalize on and re spond opportunistically to different nation state regimes by locating investments, work, and family in different geographical sites across the world.

Such flexible citizens include those of the elite cap italist class (such as entrepreneurs, managers, techno crats, and professionals) who often set up 'astronaut families' in which the wife and mother is planted in one country to anchor a home for the children so as to take advantage of the social, cultural, and educational facilities of the destination, while the husband and breadwinner undertakes the transnational shuttle between workplace in the country of origin and family at the destination. The envelope of actors who may experience multilocal lives under the dynamics of transnationality also includes more marginal migrant 'others' (such as contract workers in domestic service or the construction and shipping industry) whose sustained links with home through the regular exchange of tangible and intangible goods creates transnational social fields that both maintain their own cultures as well as allow them to embrace those of the host society. Thus, more recent understandings of transnationality have moved from the more exact, as defined by transnational citizenship involving rights and enforceable claims on the state, to the more evocative where the simultaneity of cultural rather than legal citizenship is assumed to be the salient feature.

The Logics of Transnationality

Transnationalism is argued to force a rethinking of the role of geography in the formation of identity. Rather than viewing transmigrants as agents who move easily between bounded and separate worlds, geographers have been especially interested in how transmigrants' social spheres may be conceived as the same society differen tially located in space but tightly woven across borders. As such, the boundaries of social fields do not necessarily coincide with the territorial space of the nation state. It is in these transnational social spaces, or translocalities, where the everyday activities and geographies of trans national communities occur and accumulate to enable transmigrants who share some form of identity to find and develop dense networks and social relations with one another in host societies, as well as to create and maintain links with home and other translocal spaces in other parts of the world in a multidirectional manner. These social spaces may be abstract entities such as business networks, but also include real geographical spaces, whether for malized in the form of social clubs and expatriate or ganizations, or informally convened such as migrant contract workers' weekend enclaves. These networks and social spaces can significantly influence transmigrant adaptation and flows across global space which, in turn, mediate flows of information, finance, commodities, cul tural values, and the like. Traditionally bounded notions of community and locality based on fixed spatialities are thus argued to be rendered obsolete by such a relation ally constituted sense of 'citizenship'.

The concept of social fields is an important one as it allows an incorporation of nonmigrants – those who do not themselves move or who move infrequently – into the domains of transnationality. Some scholars have pointed out that individuals may participate in transna tionality differentially, with the scope and intensity of transnational practices ranging along a continuum of different forms of transnational social practices: from core transnationalism/narrow transnationality (involving indi viduals whose daily lives are integrally tied up with transnational social fields) to expanded transnationalism/ broad transnationality (involving individuals who may engage infrequently but intensely in transnational prac tices that are not well institutionalized). Others, however, have noted unequivocally that the notion of transnation ality (and its attendant terms) should be restricted only to activities and flows that involve a high intensity of exchanges on a regular and sustained basis. In trying to understand transnational social fields, emphasis has moved from a focus on global narratives of structural conditions (a kind of transnationalism 'from above'), which enable the fluidity and seemingly effortlessness of movements in an increasingly borderless world, to the quotidian and mundane practices that sustain – and resist – transnational mobilities and transnationality (transnationalism 'from below').

While the former approach often ignores the con textual nature of transnational linkages and adopts a homogenizing account of capitalism's transnational workings, downscaling the analytical lens from the global to the everyday reveals how the logics of transnationality work at smaller scales, such as that of the household, to reveal micro geographies of how capitalism is practiced and played out in different transnational contexts, and leading to, for example, a difference between the transnationalism experienced by a Hong Kong businessman based in Vancouver, and another based in Sydney or Shanghai. Recent work has drawn attention not only to the significant efforts, resources, and organization but also to the emotional and mental costs that are an inherent part of living life transnationally. Such efforts (through telephone calls, e mails, overseas flights, and so on) and costs (monetary, homesickness, displacement, possibilities of strained or broken marriages, and so on) are well exemplified in the dynamics of maintaining transnational families stretched across space but whose constituent members continue to function economically and emotionally as a collective. It is also clear that even for the individual elite transmigrant with a hypermobile career, globetrotting lifestyle, and cosmopolitan consciousness, the actual overcoming of transnational space is not unproblematic.

While discussions of transnationalism tend to underscore the persistence of linkages and connections between transnational social spaces, how these affect the way people think about their senses of citizenship and belonging as well as their practice of transnationality is often less clear in the literature. Is transnationality a reflection of deterritorialization and the transcendence of the nation, or about being embedded, located, and incorporated in multiple localities? On the one hand, transnationals, because of their multilocal lives, are often argued (in a rather clichéd way) to be 'neither here nor there', always in a space of flows, routed but rootless. It is said that they never quite arrive at their destination because they have never quite left home and hence are never really home. Yet, on the other hand, transnationals have also been said to be both 'here and there', multiply routed between and rooted in the fabric of two or more social fields, whether through business dealings, cultural practices, social processes, or political engagements. Thus, the spatial stretching that is seen as the core of transnationalism is at times theorized as a form of deterritorialization, disrupting and challenging the notion of singular, territorialized identities and loyalties and at other times as a respatialization, creating and nurturing the global citizen who is culturally embedded and politically engaged somewhere but also at home in other locales.

Both these viewpoints demonstrate that transnational maneuverings have brought about new modes of constructing identity and subjectification that cut across political borders and which give rise to tensions between nationally defined versus transnationally assumed identities. Have transnational interconnectivities rendered any strictly bounded sense of community or locality obsolete? Do nation states still continue to define, discipline, control, and regulate the transnational movement of people and the differential incorporation of migrant groups within their borders? Does being 'trans'national mean simply a crossing of national borders, a total transcending of the nation or something in between?

Scholars who propose the loss of viability of the nation state frame contemporary migration patterns as part of the larger process of deterritorialization. They argue that while citizenship historically marked not only a belonging and commitment to a specific place, but also that as a citizen, one performed one's rights and responsibilities within the context of the nation state; additionally, transnationality reflects a weakening of conventional associations of citizenship with territorial referents of the nation state. With legal and economic freedom to move across borders and between cultures, transnationals have not only their feet, but also their hearts and minds, in more than one nation. Because of their transnational orientations and positions, and even multiple citizenships, their loyalties are divided between country of origin (from which they may derive their identity) and country of residence (from which they may derive their rights). Observers also point out how despite the attempts of states to control interaction across borders, borders have also grown increasingly 'leaky'.

Dissenting voices to the notion that transnationality fosters deterritorialized, free floating citizens argue that the transnational must not be seen in direct opposition to the national, and highlight the tension in the word 'transnational' which always keeps the national in focus. On one level, from the perspective of the 'state', scholarship has noted how transnationality has induced governments to govern dispersed subjects via 'graduated sovereignty' (Aihwa Ong's term) whereby even as the state continues to maintain territorial control of subjects within its borders, it also employs a differential system of power to control and regulate its subjects abroad. It has also been demonstrated that an increasing number of states now actively extend their territorial boundaries to reclaim emigrant members by offering continued social and political membership to those residing outside their territories. For example, to promote transnational participation, Singapore has not only introduced overseas voting privileges to Singaporeans abroad, it has also established an Overseas Singaporean Unit in 2006 to connect and engage with the 140 000 or so Singaporeans working and living overseas and has also organized 'Singapore Days' in major cities like New York and London to encourage overseas citizens to relive, for one day, Singapore culture and life (in the form of food, cultural performances, and the like). While efforts on the part of more economically well off countries like Singapore are directed at luring its local talent home in the longer term (in the light of an aging population, below replacement level birthrates, and the global war for talent), the projects of transnational reincorporation of their citizens abroad – through *inter alia*, health and welfare benefits, as well as property and voting rights – for less economically well off nations such as the Phil ippines, Mexico, the Dominican Republic, and El Salvador, are tied to ensuring that remittances are sent 'back home'. On another level, from the perspective of the 'citizen', transnational location may enhance rather than diminish the transmigrant's sense of national iden tity, as individuals reference their national memberships to establish a clearer sense of place and identity in the receiving nation. Transnationality can also cultivate ab sentee patriotism and long distance nationalism in cases where transmigrants are not fully validated by their host society and therefore continue to seek recognition and validation in their countries of origin.

Thus, scholars argue that rather than seeing transna tionality as a 'post'national cartography (i.e., the demise of the nation state), we should view transnationality as ne cessitating a reconceptualization of the way we tradition ally think about the nation state as a bounded, territorial container, and citizenship rights and respon sibilities as also granted and contained within this crucible. Instead, the boundaries of the state and the rights and responsibilities of transnational citizenship may be viewed as evolving – sometimes expanding and sometimes con tracting – with the changing tides and currents of trans migration flows. Indeed, we need to tread cautiously in equating transnational connections with disloyalty to the nation state; it is more useful to regard transnationality as an expression of the way in which nation states have been reconstituted to enable transnationals to assert local loy alties while participating in global values and lifestyles. It also needs to be pointed out that in spite of transnation ality, the state continues to assert – with increasingly re strictive immigration policies – control and authority of the population and resources within its territories, and also the flows across its borders. Indeed, technical advances have significantly increased the state's capacity to monitor not only what people say and do, but also its capability to monitor and regulate its borders.

The Uneven Scape of Transnationality

Beyond eyeing with some suspicion the liberatory po tential and claims of transnationalism as dismantling nationalisms and state structures, we should also chal lenge any assumptions that the hypermobility and easy transgression of national borders are emancipatory for all individuals alike. Because of the continued power that nation states wield to exclude undesirable alien 'others' and differentially incorporate migrant groups into their folds, transnationality is not an even playing field. Not everyone can participate in transnationality; neither do those who participate in it all experience it equally or uniformly. Increasingly too, transnational mobilities are brokered by agencies and organizations that mediate the types and directions of transnational flows, as in the case of labor agents and head hunters that recruit labor, whether lower skilled foreign domestic workers, higher skilled IT workers, or elite professionals.

In terms of class, communities, and individuals with greater economic resources and human capital and those 'more fortunately geographically positioned' are able to register higher levels of participation and assert different degrees of autonomy over the nature of their engagement with transnational transactions and processes that mi grants maintain between their new and old 'home' countries. For example, mobile entrepreneurs sustain far more elite and high level forms of transnationality (e.g., through aspects such as financial investments, conference calls, frequent jet setting, and political and cultural in volvement) as compared to the far more modest forms of transnationality associated with low or unskilled migrant contract workers (e.g., through short message services (SMSs), weekly telephone calls, and bimonthly remit tances home). Additionally, while the hypermobile international class of talented cosmopolitans has states eager to transform them into citizens, many states have in place policies to ensure that low skilled contract migrant workers remain in permanent transience on the circuits of the global labor market.

In failing to distinguish between mobile and less (or non)mobile subjects, accounts of transnationality also often fail to acknowledge the gendered dimensions of transnationality. Fortunately, a small but growing body of feminist literature has begun to highlight how transna tional movements can have enormously different effects for men and women. This body of work has established that in valorizing mobile masculinity and localized femininity in terms of transnational strategies, the ex perience of living and working in transnational social fields tends to reinforce male authority and is thus not emancipatory for most women. Male transnationals more often than not.

> feature as entrepreneurs, career builders, adventurers and breadwinners who navigate transnational circuits with fluidity and ease, while women are alternatively taken to be missing in action from globalized economic webs, stereotyped as exotic, subservient or victimized, or relegated to playing supporting roles, usually in the domestic sphere or as a trailing spouse. (Yeoh, 2005: 151)

Thus, the transnational field continues to reproduce the patriarchal binaries between male and female, with men remaining on the economically productive and public side of the divide, and women on the socially reproductive and private half, regardless of whether they leave with their high flying men folk to establish new homes overseas, or stay on in their home countries to protect hearth and home in support of the men's trans national journeyings. Others have also highlighted how racial aspects can make transnational movements a difficult and even terrifying experience for racial minorities.

Whither Transnationality Research?

Despite some 20 years of discussion and debate, the field of transnational studies is still a relatively fragmented one, studied by a range of disciplines using a variety of ap proaches. Somewhat overused but not always explicated, the idea of transnationality (and related concepts) remains by many accounts, rather 'conceptually muddled'. Its prodigious use exacerbates the fact that too much writing on transnationality is often at the abstract level rather than grounded in everyday geographies. In calling for grounded empirical work on transnationality, scholars have high lighted the need for the research to have more geo graphical specificity to enable better and comparative understandings of transnationality's place specific vari ations. As already observed, not every one is able to ex perience, or experiences, it the same way; neither is it manifested the same way everywhere. Unfortunately, while there is general agreement that more empirical research is necessary for tighter conceptualization, and that new methodological and conceptual tools are needed, the methodologies and scope by which to study transnation alism/transnationality are still unagreed.

Nonetheless, there appears to be a growing chorus of voices that suggest that any serious effort to develop in depth and comparative empirical work on transna tionality must move in the direction of multisited ethnographic research that employs a translocal frame work. Work also needs to adopt a longitudinal per spective as the dynamics of transnational migration change over time, ebbing and flowing with crisis and opportunity. Some have also argued that for transnational research to move forward conceptually, collaborative multidisciplinary work is vital. In constructing the transnational as a field of enquiry, it is also vital that more attention is paid to migrants' multiple positions and focus on the very different and unequal ways in which the register of social actors covered by transnationalism ex perience transnationality as a major feature of globalizing forces. Research needs to distinguish the jet setting elite from the less powerful players on the transnational field; it also needs to acknowledge the gendered, racialized, and classed discourses that are too often ignored in global narratives of transnationality.

Finally, in examining issues of sojourning versus set tling, routes versus roots, and ephemerality versus em placement, most studies of transnationality have conceived of space simply as the backdrop to the practice of trans national social relations and the evolution of transnational social forms. More recently, geographers such as Peter Jackson, Philip Crang, and Claire Dwyer have contended that to fully appreciate the contours of transnational spaces, research needs to recognize that spaces are not simply containers waiting passively to be filled by trans migrants but instead, they must be seen as constitutive of transnationality in all its different dimensions.

See also: Diaspora; Identity Politics.

Further Reading

Crang, P., Dwyer, C. and Jackson, P. (eds.) (2004). *Transnational Spaces*. London: Routledge.
Deforges, L., Jones, R. and Woods, M. (2005). New geographies of citizenship. *Citizenship Studies* 9, 439 451.
Glick Schiller, N., Basch, L. and Szanton Blanc, C. (eds.) (1992). *Towards a Transnational Perspective on Migration: Race, Class, Ethnicity and Nationalism Reconsidered*. New York: New York Academy of Sciences.
Hannerz, U. (1996). *Transnational Connections*. London: Routledge.
Kofman, E. (2005). Citizenship, migration and the reassertion of national identity. *Citizenship Studies* 9, 453 467.
Levitt, P. and Glick Schiller, N. (2004). Conceptualizing simultaneity: A transnational social field perspective on society. *International Migration Review* 38, 1002 1039.
Ley, D. (2004). Transnational spaces and everyday lives. *Transactions, Institute of British Geographers* 29, 151 164.
Mitchell, K. (2002). Cultural geographies of transnationality. In Anderson, K., Domosh, M. & Pile, S. (eds.) *Handbook of Cultural Geography*, pp 74 87. London: Blackwell.
Ong, A. (1999). *Flexible Citizenship: The Cultural Logics of Transnationality*. Durham: Duke University Press.
Portes, A., Guarnizo, L. E. and Landolt, P. (1999). The study of transnationalism: Pitfalls and promise of an emergent research field. *Ethnic and Racial Studies* 22, 217 238.
Smith, M. P. and Guarnizo, L. E. (eds.) (1998). *Transnationalism from Below. Comparative Urban and Community Research*, vol. 6. New Brunswick: Transaction Publishers.
Vertovec, S. (2001). Transnationalism and identity. *Journal of Ethnic and Migration Studies* 27, 573 582.
Waldinger, R. and Fitzgerald, D. (2004). Transnationalism in question. *American Journal of Sociology* 109, 1177 1195.
Westwood, S. and Phizacklea, A. (2000). *Trans Nationalism and the Politics of Belonging*. London: Routledge.
Yeoh, B. S. A. (2006). Mobility and the city. *Theory, Culture and Society* 23, 150 152.

Relevant Websites

http://canada.metropolis.net
Metropolis Canada (clicking on 'partners' will bring you to links to specific research partners in North America that look specifically at issues of immigration and integration). A project that aims to improve policies for managing migration and cultural diversity in major cities by enhancing academic research capacity, focusing academic research on critical policy questions, options, and delivery mechanisms, and developing effective ways to use research in decision making.

http://www.compas.ox.ac.uk
The Centre on Migration, Policy and Society (COMPAS). The center aims to provide a strategic, integrated approach to understanding contemporary and future migration dynamics across sending areas and receiving contexts in the UK and EU.

http://www.transnational studies.org
The Transnational Studies Initiative. This initiative attempts to bridge foster conversations on transnational studies between academics and practitioners working on a variety of topics, from a variety of disciplines, around the world.

http://www.transcomm.ox.ac.uk
Transnational Communities Programme. A research program that looked into the human dimensions of globalization and aspects of emerging transnationalism. Although ended, the website provides descriptions of the program's research projects, events, and publications (including the Global Networks journal and over 80 downloadable Working Papers) and links with other research institutes.

Transport and Accessibility

J. Gutiérrez, Universidad Complutense de Madrid, Madrid, Spain

Glossary

Hub and Spoke Structures Air-transport structures that concentrate the movements on a reduced number of routes converging in one or several large airports.
Impedance Resistance to movement (cost of transport).
Impedance Function Function used in accessibility indicators to measure the cost of transport.
Space–Time Prisms Two-dimensional representation of the area that can be reached by an individual. The horizontal axis represents a planar space and the vertical one measures time in a 24 h scale. The slopes of the prism depend on the speed of the travel.
Transport Generalized Cost The sum of the monetary (fare on a public transport journey or fuel, parking charge, toll, etc., in a car journey) and nonmonetary (travel time converted to a money value) costs of a journey.

Accessibility Concept

Accessibility is one of the key concepts in transport geography. This complex notion appears in the scientific forums of the 1950s, under the auspices of Hansen, who then defined accessibility as the potential of opportunities for interaction. Since then, several definitions have emerged that relate accessibility to the concepts of proximity or facility for spatial interaction in relation to economic and social opportunities. To put it more precisely, the term 'accessibility' expresses the facility with which activities may be reached from a given location by using a certain transport system, or in other words, the opportunities available to individuals and companies to reach those places in which they carry out their activities. Therefore, in contrast to the term 'mobility', which refers to actual movements (movements of passengers or goods over space), accessibility refers to a characteristic of places or individuals (facility to reach a destination or ensemble of destinations).

Why Is Accessibility Important?

There is a well known relationship between transport infrastructures, accessibility, and regional development. The improvement of infrastructures produces an increase in accessibility that positively affects the competitiveness of the economic system and favors the appearance of specialization benefits and scale economies. In fact, regions with high levels of accessibility to the locations of input materials and markets tend to be more productive and competitive than peripheral and remote regions. The effects of transport infrastructures on regional development are weak when there is already high accessibility in the area (highly developed and efficient transport networks), but greater when there are accessibility problems in the region (underdeveloped transport networks, bottlenecks, etc.).

Accessibility is not only important from the economic point of view. It is also of high value from the perspective of social welfare. Accessibility to work, health, education, and services is one of the key components when it comes to evaluating the level of welfare of the population. Remote rural areas generally have serious problems in access to services, which is a determining factor in migratory processes and the depopulation of such territories. On the other hand, there is a broad offer of employment and services in large cities (both in quantity and in quality), which is considered as a factor of attraction for the population.

Accessibility is usually seen as a positive factor, but it may also have negative implications. High value natural spaces may be in danger if a new highway is built in the area to facilitate access for the population. Maintaining the inaccessibility of certain natural spaces is a means of protecting them.

Transport Networks and Accessibility

The spatial distribution of accessibility depends to a great extent on geographic location. Within one and the same geographic region, there are areas that take a central position from which the main activity centers can easily be accessed. By contrast, there are other areas that have disadvantages derived from their peripheral location, far from the main markets and centers of activity.

However, accessibility does not only depend on geographic position. Transport networks also play a determining role. On the one hand, places crossed by efficient infrastructures, such as motorways, have greater accessibility and development potential (corridor effect) than those located in interstitial spaces between corridors (shadow effect). On the other hand, high speed transport systems produce relevant changes in the distribution of accessibility. In the case of the high speed train, the reduction of stopping places produces the so called tunnel effect: the new transport system efficiently links large cities with one another, but not so with regard to the small towns situated along the line. The effects are therefore similar to those of air transport, which is highly concentrated around the main airports. From this point of view, geographic position is becoming less and less important, whereas the connection to high speed networks is becoming more relevant.

The structure of transport networks decisively influences the distribution of accessibility. Radial transport networks tend to reinforce the accessibility of central areas, while the decentralized networks distribute accessibility more homogeneously and favor decentralization. This idea may be applied to different spatial scales, as, for example, national motorway networks or metropolitan road networks. In the case of metropolitan areas, radial networks favor the consolidation of monocentric systems, while decentralized networks favor polycentric structures.

Transport systems are in constant evolution, modifying accessibility conditions. Changes are not only due to infrastructures but also due to the organization of the services. Thus, for example, hub and spoke structures in air transport systems reinforce the accessibility of the main airport centres to the detriment of the smaller ones.

Relationships among Transport Systems, Accessibility, Land Use, and Mobility

Accessibility is closely related to other geographic concepts. It can be modified by changes in transport infrastructure (e.g., the building of a new motorway) or in land use (e.g., the building of a new hospital or shopping center). Yet on the other hand, an improvement in accessibility favors changes in land uses, and these in turn influence mobility.

There are complex relationships between transport infrastructure, land use, accessibility, and mobility. In **Figure 1**, these relationships may be followed beginning with the transport infrastructure. The improvement of transport infrastructure brings about an increase in accessibility. Within the more accessible areas there may appear changes in land use (e.g., new residential areas or new facilities). As a result, the number of trips increases because there are now new areas of generation and attraction of trips. But accessibility not only acts indirectly through land uses, it also does so directly: the better accessibility conditions created by the improvement in infrastructures may modify the modal split (more trips in private transport, as a result of the building of a motorway), alter the routes of car users, or it may even generate induced demand (trips that would not be made if the new infrastructure did not exist). Finally, the existence of greater mobility in the area may, in mid or long term, produce congestion problems, which logically entail a loss in accessibility and worse conditions in the area for attracting more activities (land uses changes). Congestion problems can be solved by enlarging the capacity of the transport system, so that there would be an improvement in accessibility in the area bringing about new changes in land use and mobility.

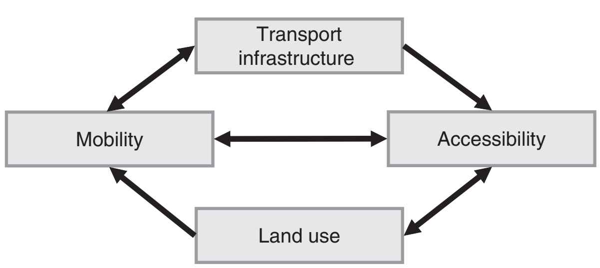

Figure 1 Relationships between transport infrastructure, land use, accessibility, and mobility.

How to Measure Accessibility?

The spatial distribution of accessibility can be calculated by means of accessibility indicators (see the next section) using the geographic information system (GIS). The results are usually displayed in the form of maps (although also in table and graphic form).

There is a wide variety of accessibility indicators, for they may be used for very different purposes. Yet most of them combine two basic elements in a single measure: the cost of transport to centers of activity and the capacity of attraction of these centers:

- The cost of transport is a measure of the friction of space. This may be expressed in units of distance, time, money, or transport generalized cost, bearing in mind possible restrictions on the network (one way directions, prohibited turns, etc.).
- The attraction capacity of the destinations expresses the offer of activities (both quantitative and qualitative). But activities may be treated in a more detailed way, particularly in studies on small urban or rural areas: for example, the accessibility to health or educational services can be measured. This not only affects the activities considered in the destinations, but also population groups or companies taken into account in the origins. Thus, for example, tourists are interested in tourist attractions, but consumers in shopping centers; for their part, consumer oriented firms consider accessibility to the population, whereas business oriented firms the accessibility to other companies.

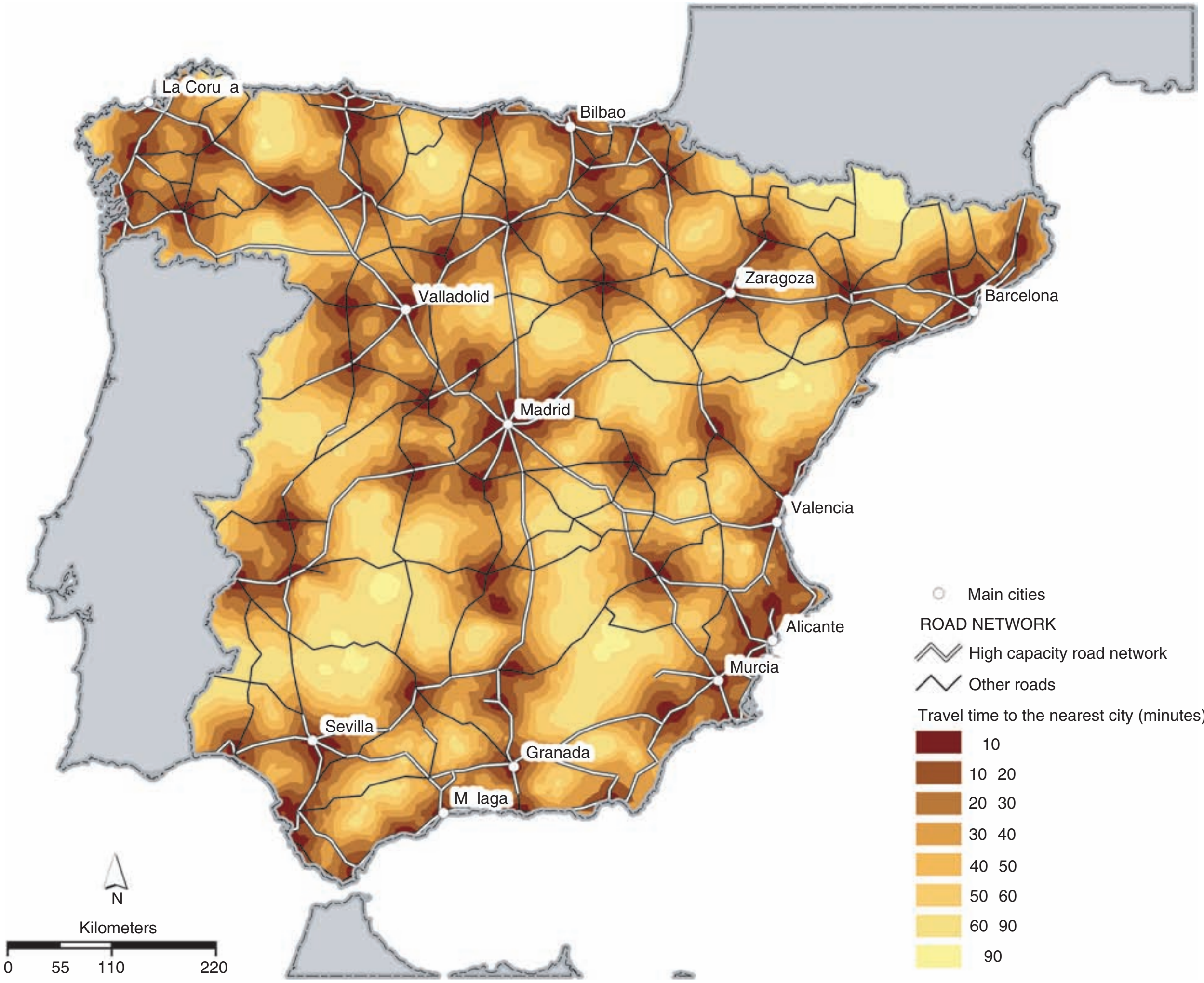

Figure 2 Travel times by motorway to the nearest city in Spain.

Accessibility indicators may be calculated in different ways, depending on the modes taken into consideration:

- Unimodal accessibility: Only one transport network is considered, such as, for example, the road network.
- Multimodal accessibility: The results from two or more unimodal accessibility indicators are added together (e.g., roads + railways).
- Intermodal accessibility: An intermodal network is used which enables transfers between modes (e.g., between roads and railways), by using the most fa vorable mode on each stretch of the trip.

Some Accessibility Indicators

There is a wide variety of accessibility indicators, which respond to the different aims of the researches. Some of the indicators used in human geography, from different perspectives, are as follows.

Location-Based Accessibility Measures: Weighted Average Travel Times

The simplest location based accessibility measures are obtained by calculating the travel cost to particularly relevant destinations, for example:

- travel time to a previously selected destination (e.g., a hospital or the Central Business District (CBD) or
- travel time to the nearest of the destinations con sidered (e.g., the nearest hospital or, in **Figure 2**, ac cess time to the nearest city).

A more complex formulation is that of the weighted average travel times (or location indicator). This consists of calculating the average access times with regard to the destinations considered, taking as weight the volume of activity in the destination, according to the following expression:

$$A_i = \frac{\sum_{j=1}^{n} t_{ij} m_j}{\sum_{j=1}^{n} m_j} \qquad [1]$$

where A_i is the accessibility of node i, t_{ij} is the travel time by the minimum time route in the network between node i and destination j (in minutes), and m_j is the mass (volume of activity) of the destination j. The mass (attraction capacity) of the destination center is used to weight in order of importance the minimal time routes.

This indicator reflects quite well central and peripheral locations of the demarcation area with respect to the destinations considered. **Figure 3** shows a distribution of accessibility in which the core peripheral rings are distorted by the presence of motorways, which offer the possibility of traveling at higher speeds. This measure is not a gravity based indicator, so that it does not place the emphasis on short distances (impedance function 'b' in **Figure 7**). This fact seems to be critical in order to evaluate the infrastructure effects at the national level.

Infrastructure-Based Accessibility Measures

Some measures are used to describe the functioning of a transport system (in terms of sinuosity, congestion, etc.). One of these measures is the network efficiency indicator. It offers a measure in terms of the relative ease of access according to the network efficiency. Most of the location based measures (as the location indicator) highlight differences between peripheral and central locations. Network efficiency indicator neutralizes the effect of geographical location, and the ordinary notion of costs is replaced by the ease of access in relative terms (network efficiency), as follows:

$$A_i = \frac{\sum_{j=1}^{n} \frac{t_{ij}}{\hat{t}_{ij}} m_j}{\sum_{j=1}^{n} m_j} \qquad [2]$$

where A_i measures the accessibility of node i, t_{ij} and m_j are defined as above, and $\hat{t}_{ij}$ is the ideal time spent between the nodes i and j, assuming the hypothetical existence of a high speed infrastructure (motorway or high speed train, according to the mode) in straight line (Euclidean impedance). The weighted average of the ratio $t_{ij}/\hat{t}_{ij}$ is calculated according to the mass of the destinations (impedance function 'b' in **Figure 7**), but other weight

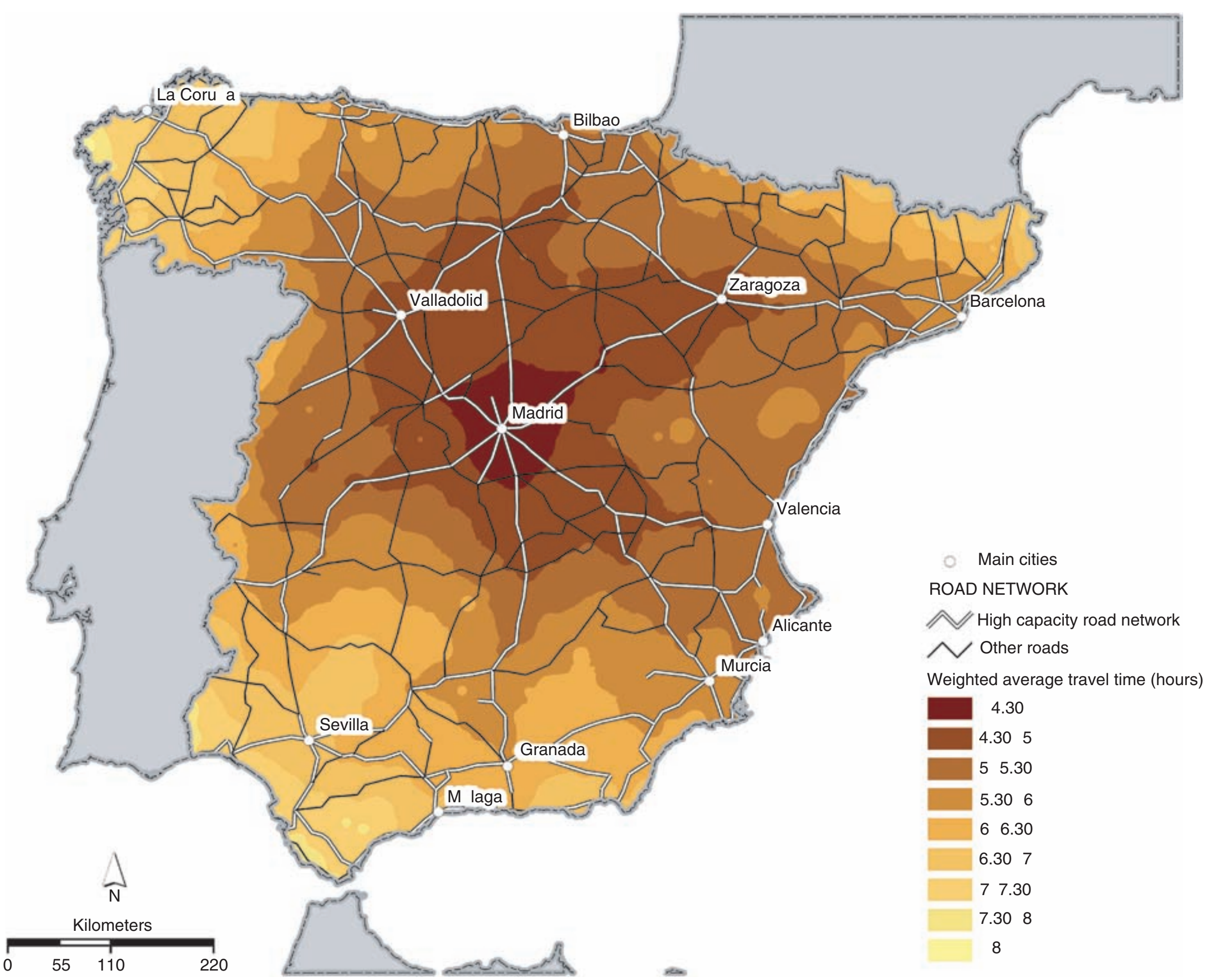

Figure 3 Weighted average travel times by motorway to the main Spanish cities.

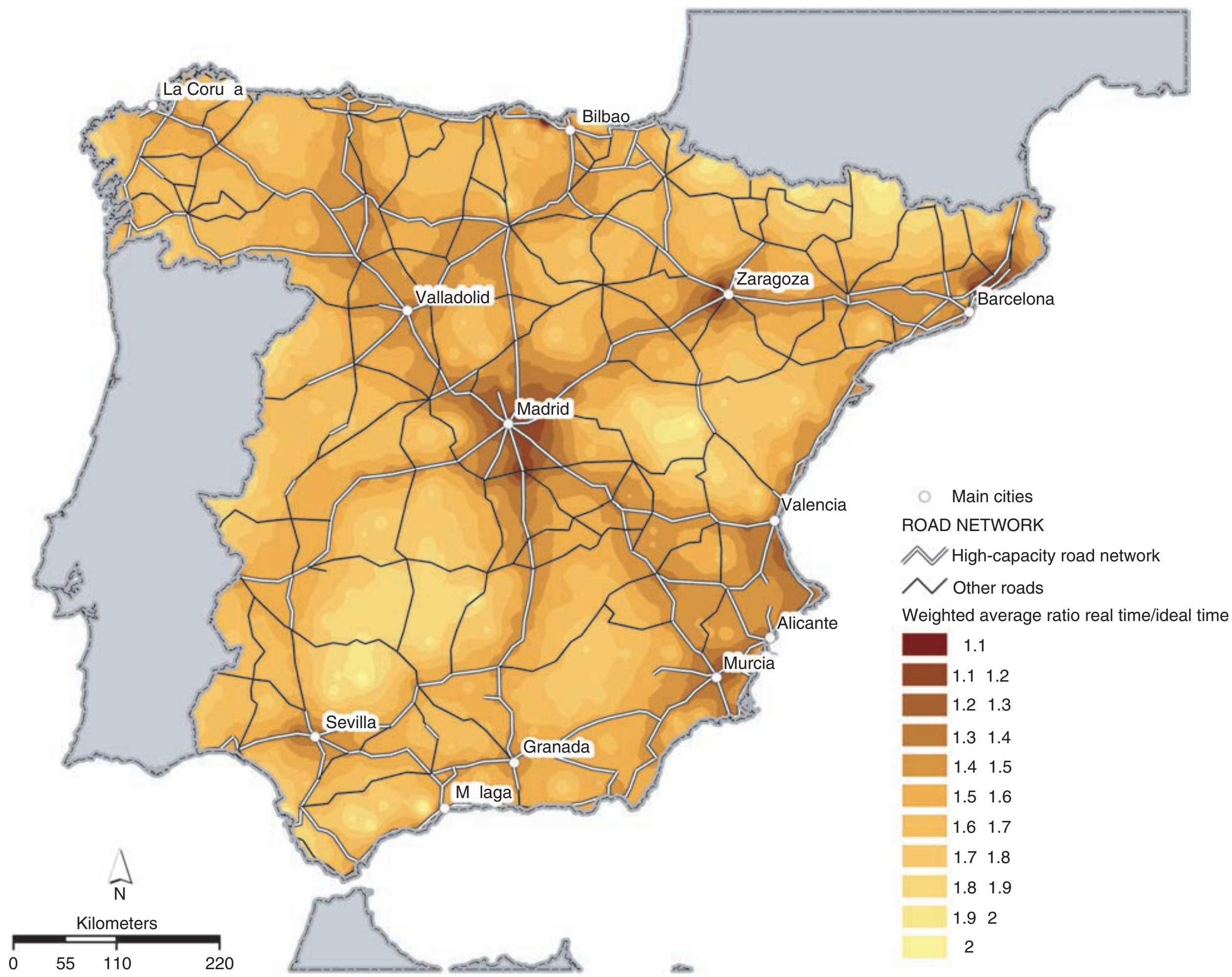

Figure 4 Network efficiency: Spanish motorway network.

factors can be used as, for example, the real transport demand in each city pair (impedance function 'a' in **Figure 7**).

This indicator should be interpreted from an infra structural viewpoint. It measures in relative terms the efficiency of the infrastructure in the access to the rele vant destinations. The results are given in meaningful units: how much network impedance surpasses Euclidean impedance. For example, a value equal to 2 means that the network impedance doubles the Euclidean one and so forth. This indicator allows us to identify not only high accessibility corridors (along the motorways) and shadow areas (those apart from the corridors), but also nodal places in the corridors (the nodes at which motorways converge) (**Figure 4**).

Gravity-Based Accessibility Measures: Economic Potential

Economic (or market) potential is a gravity based measure that has been extensively used in accessibility studies. According to this model, the level of opportun ities between two nodes increases with the mass of the destination node and declines with the distance (time) between both nodes (impedance function 'a' in **Figure 7**). Its classical mathematical expression is as follows:

$$P_i = \sum_{j=1}^{n} \frac{m_j}{t_{ij}^{x}} \qquad [3]$$

where P_i is the economic (market) potential of node i, m_j and t_{ij} are known terms, and x is a parameter that reflects the effect of the distance decay function. In the majority of accessibility studies the value of the parameter x is 1. If $x>1$, it means that more weight is given to short distances.

The results provided by this indicator must be carried from an economic point of view. It measures the close ness of potential economic activity to a particular node. Main urban agglomerations tend to have the highest economic potential, while remote rural areas face prob lems of lack of accessibility (**Figure 5**).

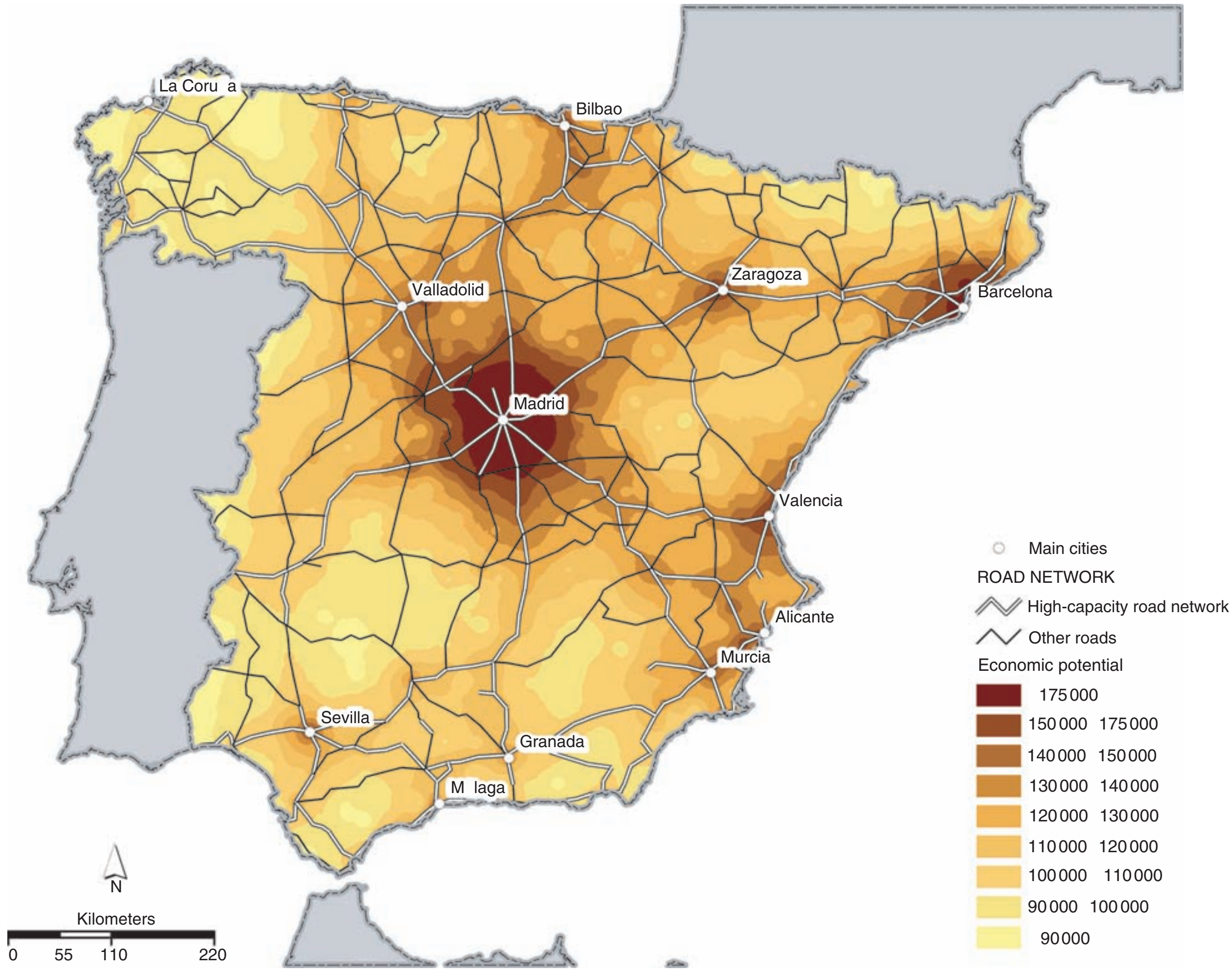

Figure 5 Economic potential: Spanish road network.

Daily Accessibility

This indicator develops the notion of 'contact networks' calculating the amount of population or economic ac tivity that can be reached from each node within a limit of travel time (as an indicator of the position of a city in the urban system). The threshold figure is usually es tablished between 3 and 4 h, so that it is possible to go and return within the day and carry out some activity at the destination city. Its mathematical expression is as follows:

$$\mathrm{DA}_i = \sum_{j\ 1}^{n} m_j \delta_{ij} \qquad [4]$$

where DA_i is the daily accessibility indicator of node i, m_j is the mass of the destination center j (e.g., population), and δ_{ij} takes the value of 1 if t_{ij} is less than 3 h, and 0 otherwise. This index can be viewed as an extreme case of a potential market indicator, because the distance decay function takes the discontinuous form of all or nothing depending on the threshold of travel time considered (**Figure 6** and impedance function 'c' in **Figure 7**). The results are of the following type: from a node i, within a travel time limit of 3 or 4 h, 5 million inhabitants can be reached. Areas located near several main cities tend to have the highest accessibility values.

This measure is particularly useful for calculating accessibility for business and tourist trips, for the need to stay overnight in the destination city means an important extra cost for both firms and individuals.

Individual-Based Accessibility Measures: The Space–Time Geography Approach

The space–time geography approach puts the emphasis on individuals, not on locations. It considers accessibility in terms of the possibilities for individuals to take part on activities, incorporating spatial and temporal constraints. Space–time prisms are used to identify the potential areas of opportunities that can be reached by each person

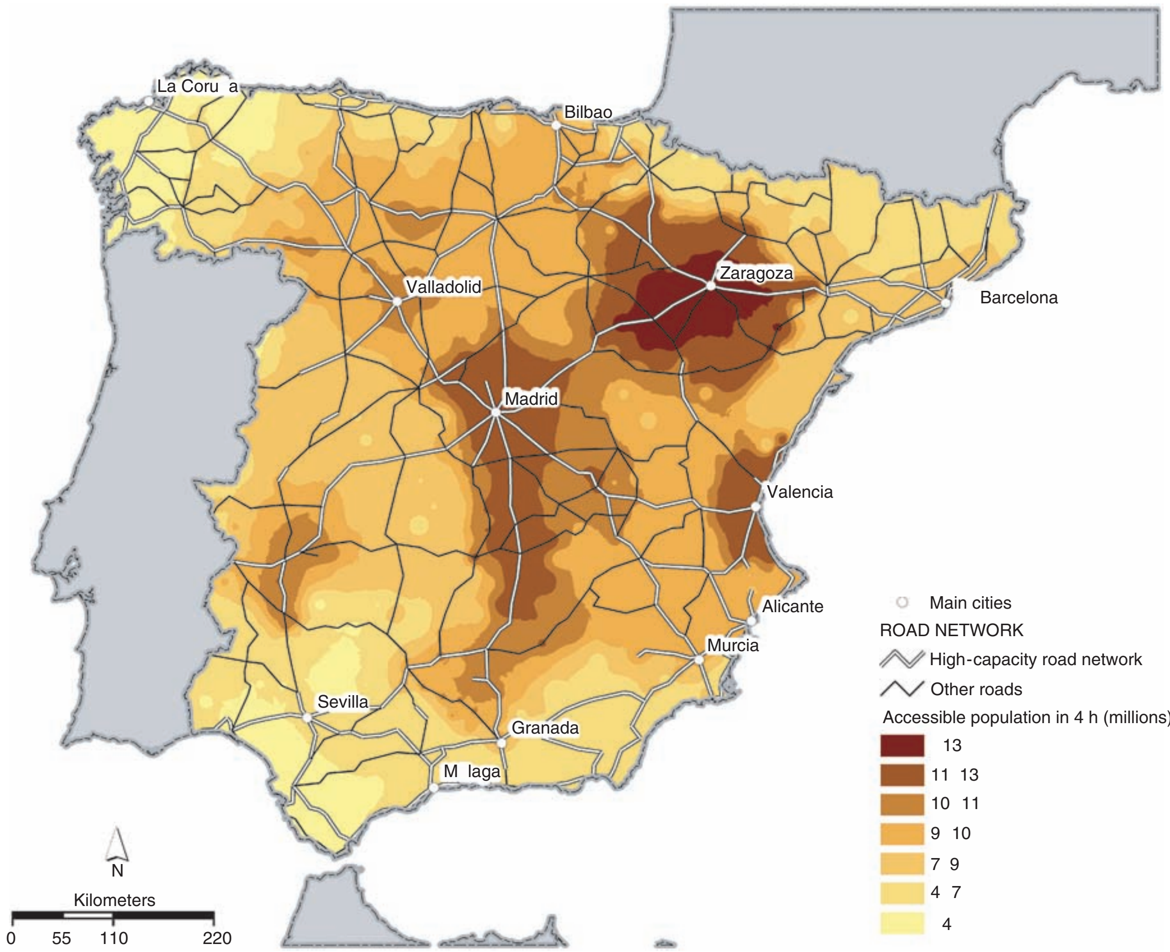

Figure 6 Daily accessibility: Spanish motorway network.

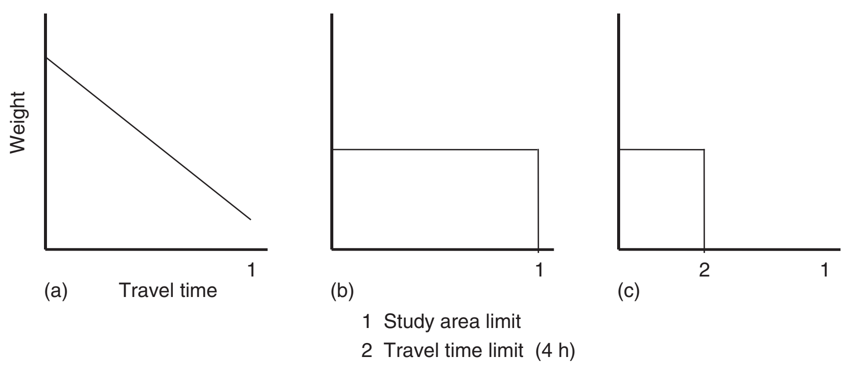

Figure 7 Accessibility impedance functions. (a) Economic potential all relationships within the study area are taken into account, but inversely to the travel time, so that short-distance routes contribute very heavily in measurements of potential but very little those over long distances (distance decay). (b) Weighted average travel time all relationships within the study area are considered too, but routes on short distances do not contribute more than others in accessibility calculations, since there is not distance decay. (c) Daily accessibility indicator exclusively are considered the centers within a certain travel time limit, but not the rest of the centers within the study area. Within this time limit there is not distance decay too.

taken into account its time constraints. For example, a shop assistant who works in the CBD is near a large number of shops (good spatial accessibility) but is at work most of the time that the shops are open (temporal constraints). It is certainly a detailed approach allowing many subsets of individuals to be considered. But con sequently large amounts of data are required and ana lyses are only possible for small areas.

Accessibility in Transport Infrastructure Planning

One of the main aims in transport policies is to improve accessibility. The building of new infrastructures is to a great extent justified by the increase in accessibility (improving access to markets from peripheral regions, facilitating access to services, solving the access problems to the CBD, etc.).

In order to simulate improvements in accessibility brought about by a transport plan, accessibility indicators can be used considering two scenarios: the scenario with plan (which includes the infrastructures proposed in the plan), and the scenario without plan (which only takes into account existing infrastructures). If one calculates the distribution of accessibility in both scenarios and differences are obtained between the two maps, the re sulting map shows the areas that will obtain more benefits as a result of the building of the proposed infrastructures.

On the basis of the accessibility values, dispersion measures (such as the variation coefficient) can be ob tained in order to know the degree of spatial equity in the distribution of accessibility on that particular region. By comparing both scenarios (with and without plan), it is possible to know whether the plan increases or reduces spatial equity. Thus, for example, the development of a centralized network of radial motorways tends to favor the center in detriment to the periphery and therefore increases spatial imbalances; yet if it is later converted into a decentralized network by building transversal motorways, it will reduce the spatial imbalances between center and periphery.

Accessibility indicators are also used in urban and regional planning. Analyses of accessibility allow one to know those areas that are not properly served by a cer tain type of facilities and to choose the best locations to build new facilities (such as health centers, schools, and shopping centers). They will likewise enable one to identify those areas poorly covered by the urban system.

See also: Aviation; GIS and Cartography; Transport, Public; Railways; Regional Connectivity; Space-Time; Transport Geography; Transportation and Land Use; Welfare Geography.

Further Reading

Farrington, J. and Farrington, C. (2005). Rural accessibility, social inclusion and social justice: Towards conceptualisation. *Journal of Transport Geography* 13, 1 12.
Frost, M. E. and Spence, N. A. (1995). The rediscovery of accessibility and economic potential: The critical issue of self potential. *Environment and Planning A* 27, 1833 1848.
Geurs, K. T. and van Wee, B. (2004). Accessibility evaluation of land use and transport strategies: Review and research directions. *Journal of Transport Geography* 12, 127 140.
Hagerstrand, T. (1970). What about people in regional science? *Annals of the Regional Science Association* 24, 7 21.
Handy, S. L. and Niemeier, D. E. (1997). Measuring accessibility: An exploration of issues and alternatives. *Environment and Planning A* 29, 1175 1194.
Hansen, W. G. (1959). How accessibility shapes land use. *Journal of American Institute of Planners* 25(2), 73 76.
Jones, S. R. (1981). *Accessibility Measures*: *A Literature Review* (*TRRL Laboratory Report 967*). Crowthorne: Transport and Road Research Laboratory.
Koenig, J. G. (1980). Indicators of urban accessibility: Theory and applications. *Transportation* 9, 145 172.
Kwan, M. P. and Weber, J. (2003). Individual accessibility revisited: The implications of changing urban form and human spatial behavior. *Geographical Analysis* 35, 341 353.
Morris, J. M., Dumble, P. L. and Wigan, M. R. (1978). Accessibility indicators for transport planning. *Transportation Research A* 13, 91 109.
Moseley, M. J. (1979). *Accessibility: The Rural Challenge*. London: Methuen.
Pirie, G. H. (1979). Measuring accessibility: A review and proposal. *Environment and Planning A* 299 312.
Reggiani, A. (ed.) (1998). *Accessibility, Trade and Locational Behaviour*. Andershot: Ashgate.

Relevant Websites

http://people.hofstra.edu/geotrans/
The Geography of Transport Systems.
http://www.vtpi.org/tdm/tdm84.htm
Victoria Transport Policy Institute.

Transport and Deregulation

P. White and A. Sturt, University of Westminster, London, UK

Glossary

Deregulation The. process of removing controls which may limit the quantity of transport service provided, price levels, routes to be served, etc.

Economies of Scale An association between the scale of an economic activity and unit costs, leading to lower unit costs as scale of production increases (e.g., costs per vehicle-km run falling as total vehicle-km increase).

Low-Cost Carrier (LCC) A form of airline concentrating on a low-price, low-cost market, typically by minimizing passenger service on board, airport turnround time, etc., and utilizing internet booking. Major examples include Ryanair and easyJet.

Privatization The transfer to private ownership and/or control of assets and operations previously in the public sector.

Quality Regulation Regulation of aspects of transport service not directly related to the quantity of service provided, nor price – typically, rules intended to improve safety levels and/or reliability of service to the user.

Regulation The process of applying controls to the quantity, price, etc., of transport services.

Re-Regulation The reintroduction of regulations (not necessarily in their original form) after a period of 'deregulation'.

Yield Management A system of pricing based on maximizing use of capacity, by varying price according to capacity available, and demand, in real time, using passenger pre-booking, initially adopted by 'low-cost' airlines but now in wide use for long-distance travel by all public modes.

Basic Concepts

The concept of 'deregulation' has come into widespread use in transport and many other industries in the last 20–30 years. However, it is often used in a vague form, covering many distinct processes.

The first half of the twentieth century saw the development of extensive motorized transport systems, greatly increasing mobility and supplanting other modes for most journeys other than very short distance. However, this rapid development was often associated a number of problems, including poor safety records, financial instability in the industries concerned, and the need to protect customers. Indeed, the railways in the nineteenth century had raised similar problems. Regulation took three main forms:

1. *Quantity*. The total quantity of service provided was regulated, so as to ensure stability of production, and enable the supposed effects of excessive competition to be avoided. In some cases this took the form of area monopolies (such as that of London Transport from 1933), in others the licensing of individual routes (as in the regulation of bus services outside London from 1931), although in practice the degree of competition permitted under the latter was often small. A direct example of 'quantity' regulation is that which limits the number of licensed taxicabs in a defined area, still common in many countries.
2. *Price regulation*. Prices charged by operators were restricted, to avoid unreasonable increases. An element of consumer protection was also introduced through requiring display of prices (e.g., the specification of fare scales charged by metered taxicabs) enabling the user to see what the maximum permitted charge would be.
3. *Quality regulation*. This related mainly to specifying minimum standards for vehicles, drivers, and infrastructures, primarily to improve safety levels in the industry concerned – for example, through driver and operator licensing. Under franchising or tendering systems, this may be extended to monitoring of service quality (e.g., to minimize the percentage of services running later than a defined range), but as such is not generally applied directly by the regulatory body.

In the case of the railways, nineteenth century regulation had been concerned firstly with giving consent to build new infrastructure, subsequently with safety, and control of price levels. New modes which developed in the twentieth century initially experienced a period of minimal regulation (e.g., the bus and coach industry in Britain in the 1920s), but this was subsequently followed by a national system. Even the airline industry was regulated from an early stage, and more heavily so in the period after World War II. Much of the regulation affecting road transport was developed and applied in the 1930s, affected by the general economic recession at that time, which may have influenced thinking in the direction of protection and stability, in addition to factors specific to the transport sector.

However, such regulation has often applied more strictly to forms of public transport than other modes.

Growth in car ownership and use has meant that in most countries this mode now accounts for the great majority of person km traveled. In general, however, it has been subject to much less regulation than other modes, al though increasingly strict controls are now applied to improve its safety. Attempts to regulate within the public transport market in order to protect existing operators (e.g., railways against express coaches) may have limited effect given the major role of the car.

Most regulation has been applied at the level of na tional governments. However, international organizations (such as the European Union) play an increasingly im portant role. In some cases, notably the taxi and private hire vehicle market, regulation continues to be applied at a much more local level.

Regulation and Ownership

The process of 'deregulation' can be seen as reaction to excessively strict forms of regulation, especially in terms of quantity and price control, which may have done little to achieve their objectives (e.g., regulating prices at low levels during an inflationary period could result in low profitability for an industry, discouraging investment, and hence to the long run disadvantage of its users).

It often stems from the same political philosophy as 'privatization', that is, the transfer of publicly owned operations to private ownership. Very often the modes of transport protected (or restricted) by regulation were also in public ownership (e.g., national rail systems in most European countries). However, the two processes do not necessarily coincide. For example, there was a period of intensive competition in Britain between long distance rail and express coaches, following deregulation of ex press coach services in 1980, yet the dominant coach operator (National Express) was in public ownership until its sale in 1988, and the corresponding rail services were not privatized until the late 1990s.

It may also be the case that privatizing businesses which retain a monopoly element may result in stricter (or at least, more explicit) forms of regulation than ap plied before. For example, the rail industry in Britain has been subject to far more controls over prices, and service specifications for particular routes, than applied under the more informal relationship between the British Railways Board and the government prior to privatization.

Examples from Experience in Britain

Some examples from the British case illustrate the principal outcomes of regulation. Under the Conserva tive government of Mrs Thatcher elected in 1979, the first major element of deregulation was that of long distance coaches in 1980. While Britain already contrasted with other large European countries in per mitting some express coach operations in competition with rail, competition within that industry was very limited, a dominant position having been built up by the state owned operator National Express. Removal of quantity and price controls enabled many new operators to enter the market, and for coach to compete more in tensively with rail than before. Operators simply notified the routes they wanted to run, and were free to charge prices they wished. Initial impacts included a large number of new entrant operators, and a rapid growth in total express coach use of about 50% in the first 5 years, coming from both diverted rail passengers and new de mand. However, National Express remained dominant, contrary to economic theories which suggested that in a market with low entry costs, a large number of competing operators might be expected. While there is little evi dence for simple economies of scale in direct operating costs, a large operator may be in a much stronger position in terms of marketing, especially through providing a comprehensive interconnecting network. From the late 1980s until 2003, there was little competition for Na tional Express within the coach sector. It is also note worthy that the major new competition which has emerged – in the form of the 'Megabus' network, based entirely on price yield management with pre booking – has been developed by the large Stagecoach Group (see below), not smaller operators.

Express coach deregulation appears to have func tioned well in terms of widening user choice, giving both a greater supply to service (by frequency, and route coverage) and intensive price competition, especially in the early period of deregulation. Hence, it is reasonable to infer that net user benefits were produced (these would also apply to some rail users remaining with that mode, whose fares might be lower than otherwise, due to the need for rail to respond to price competition from coach). This outcome may well have influenced coach deregulation in Norway and Sweden.

In the British case, express coach deregulation was followed by that of local buses in 1986 (outside London and Northern Ireland). The objectives here may have been different, in that the major concern of government was with the rapidly growing public expenditure for support of the bus industry. In addition to enabling higher levels of service and/or lower prices than would otherwise be the case, there was concern that inefficient methods of operation were being supported.

Network wide financial support was prohibited, and instead the aim was that most services would be com mercially registered (i.e., covering all their costs from revenue). It was recognized that some services, such as those in low density rural areas, might not be com mercially viable, and in such cases a mechanism for support was introduced, in which competitive tendering (rather than funding of the incumbent operator) was applied.

In contrast to the effects of express coach deregulation, that of local buses did not produce a growth in ridership (despite a large rise in vehicle km run). This was due in part to increased fares when large subsidies were removed (notably in some large metropolitan areas). However, there is also evidence that price competition functions less effectively for local, short distance markets than in long distance markets. In the latter, users tend to plan ahead, often prebooking journeys, and the mix of purposes makes the market as a whole more price sensitive. In the latter, taking for the first bus or train to arrive is a logical response on frequent services. Extending waiting time to allow for the use of a cheaper service may involve a greater user cost overall (when value of waiting time is quantified), and hence is often unattractive. The taxi market also displays this characteristic, the 'search time' for a cheaper taxi often offsetting the benefit of lower fares.

Contrary to some fears, deregulation of the coach and local bus sectors did not result in a worsening of the safety record. In the longer run, the casualty rate (killed and seriously injured per thousand million passenger km) has continued to fall. In part, this may be attributable to stricter 'quality licensing' introduced from 1981, which required each operating business to hold a separate 'Operator Licence'. This in turn is subject to curtailment or cancellation where a poor maintenance record has been found, for example.

Airline Deregulation in Europe

The USA and Europe are the world's only geographically extensive examples of airline deregulation. In the USA full deregulation was achieved almost overnight in 1978. Here a state of market maturity now exists, with one large low cost carrier (LCC), that is, Southwest Airlines, together with several medium sized LCCs accounting for a fairly stable 25% of the air passenger market. In the European Union (EU) by contrast, deregulation was more recent and phased in over the period 1987–97. As a result the EU now has a rapidly growing (7–8% pa) and changing low cost sector. Emphasis in this review here is on EU air deregulation and its impacts. However, analysis is complicated by the fact that many EU air passenger services are provided by airlines whose networks extend beyond, or are based outside, the EU. British Airways is a global player and several major EU carriers are now embedded in global alliances each involving several non EU operators.

Prior to 1987 European airlines were subject to the international regulatory arrangements laid down in the immediate post World War II years to foster growth in the nascent civil aviation industry. The core principle was that each nation had sovereignty over its own air space. For each airline route, a bilateral Air Service Agreement had to be negotiated between the countries involved. These reciprocal agreements specified the airports to be served and the capacity (usually frequency) to be offered. It was left to individual countries to designate the airline chosen to operate their share of the route. In practice, the designated carrier was usually the State owned 'flag carrier' (BA, KLM, Lufthansa, etc.). Nominally national governments also controlled fares, but in practice these powers were usually delegated to the International Air Transport Association (IATA) where representatives of the world's airlines gathered in an annual conference to agree fares. The upshot of this regulatory framework was a complete absence of market forces in international air transport. Each airline route was a duopoly in which the two airlines involved effectively had powers to fix fares at whatever level they deemed necessary to meet their costs. The situation was further complicated by the fact that, for prestige and other reasons (some unclear), many governments heavily subsidised their flag carriers. The air traveler was faced with a technically reliable and geographically extensive but high cost air service network. This suited the needs of the business dominated air transport market of the 1950s and 1960s, but was ill adapted to the potential mass leisure market of subsequent years. The result was the emergence of a parallel low cost Charter airline sector with seasonal routes linking northwest Europe to (mainly Mediterranean) vacation destinations. Charter airline seats could be bulk purchased by tour operators, but under IATA regulations could not be sold on a 'seat only' basis to individual travelers. Although by 1980, it accounted for almost half of the European internal air transport market, Charter offered little real competition to Europe's state flag carriers.

In the mid 1980s, a few EU countries (notably UK, Irish Republic, and The Netherlands) attempted to promote airline competition by negotiating liberal Bilateral Air Service Agreements (greater fares flexibility, relaxed capacity sharing rules, etc.) on a few of Europe's busiest routes. Full EU wide deregulation was achieved over the period 1987–1997 with particular importance attaching to the so called EU 'Third Package' measures of 1993. For EU based airlines operating intra EU routes, EU wide general multilateral arrangements replaced the need for route by route bilateral agreements. As a result, any EU airline can now theoretically operate on any chosen EU route, free of capacity and fares controls. This statement slightly oversimplifies the real situation. In fact, several key EU international gateway airports, including London Heathrow, and London Gatwick have no slots available for new operators, thus perpetuating the advantages of existing airlines (and swelling airport profits). Furthermore, several EU member states (including Greece, Italy, and France) found ways to subsidize their flag carriers despite EU rules to the contrary. This again distorted free market competition, to the disadvantage of not only new operators, but also the more cost effective 'old' operators.

The initial impact of deregulation on the European airline scene was rather modest. There were some signs of competition between flag carriers on heavily trafficked city pair routes: the London–Amsterdam and London–Dublin routes both experienced falling fares and rapid traffic growth. EU flag carriers in general pursued cost reductions by improved labour productivity. Several major EU operators began a significant spatial restructuring of their straggling route networks into operationally more efficient (but also more delay prone) hub and spoke configurations: the KLM operations focused on Schiphol are a good example of this. However on much of the EU airline network until the late 1990s, it was largely a case of 'business as usual'.

The main ongoing impact of deregulation has been the emergence of a dynamic LCC sector. Two airlines, in particular, spear headed this development. Ryanair founded as a conventional airline in 1986, became Europe's first LCC in 1992. EasyJet operated as an LCC from its inception in 1997. The spectacular success of these early LCCs resulted in the foundation of numerous other EU LCCs. Some of these, including GO, BUZZ, and BMI Baby, started life as subsidiaries of existing carriers (BA, KLM, and BMI, respectively).

Most EU LCCs to date have been loosely modeled on an enduringly successful American low cost airline, namely Southwest Airlines. Although no two LCCs are alike, most EU examples share a common set of cost saving characteristics. Single class cabins and high seating densities are the norm. LCCs are often described as no frills airlines; a reference to the fact that they provide no free in flight catering. This saves not only the cost of meals, it also permits cabin crew numbers to be reduced to the minimum required for safety reasons. LCCs use a simplified pricing system based on one way fares rather than the complex price discrimination strategies and associated restrictions for lowest fares (e.g., that a stay must include a Saturday night) which have traditionally been associated with air travel. The simple LCC point to point route patterns permit more intensive use of aircraft and crews than is possible with the complex hub and spoke plus interlining operations used by established airlines. The use of regional airports, rather than major international gateways, results in both lower landing charges and fewer delays. Indeed, the existence of a surplus of airport capacity in the form of a high density of regional airports each competing for new users has been a key factor in the success of EU deregulation. There has been considerable disquiet that some of the very low airport charges on offer amount to subsidization of the low cost airline sector out of public funds. The European Commission recently and successfully took Brussels South Charleroi airport to court for such behaviour.

Recent studies suggest that the production cost per seat km of an efficient European LCC is less than half that of a conventional short haul carrier. These dramatically lower unit operating costs are further reinforced by marketing cost savings mentioned earlier. The result for the passenger has been attractively low fares together with a rapidly widening range of destinations. Not surprisingly, the LCC sector has experienced explosive growth in demand for its services. In 1994, the LCC sector (effectively Ryanair alone) accounted for a tiny proportion of intra EU air travel – some 3 million trips. By 2000, LCCs already accounted for 18 million trips. In part, this LCC growth has been at the expense of conventional airlines; however, most of it has been newly generated traffic. In addition to their success in leisure travel, LCCs are now increasingly penetrating the business market.

Over the period 1999–2004, there were in effect two groups of airlines employing quite different business models competing in the growing EU air travel market. They were the LCCs described above and the full service network carriers (FSNCs) as the remaining conventional EU (ex flag carrier) airlines came to be known. While both groups have had their business failures, there can be little doubt that the LCC format is increasingly winning the battle for leisure market share. Recent figures suggest that there are over 60 LCCs operating in Europe and that between them they now account for around 25% of the intra EU air travel market. The most recent development in the deregulated EU air travel market has been the partial adoption of the LCC business model by the one time FSNCs.

The rapid growth of the EU's low cost sector has interesting spatial dimensions. The airline map of Europe has been transformed. LCCs have preferred to grow their markets by adding new destinations rather than by adding capacity to long established routes. In 1996, just over 200 city pairs were served by non stop flights.

By 2004, this number had risen to around 430 and this growth is ongoing. Substantial research effort is now devoted to the geographical impacts of EU air deregulation. It may be that deregulation's consequences have been as great for Europe's airports and their hinterlands as they have been for the airline industry itself. In particular, deregulation is extending the economic stimulus of tourism to regions located far from the main corridors of the pre deregulation airline network.

Viewed overall the far reaching impacts of EU airline deregulation have significantly benefited the consumer.

What the future holds for Europe's air transport industry is rather unclear. The large annual European air traffic growth levels resulting in part from deregulation are now threatened by factors exogenous to the airline industry and to the EU. These include rising fuel costs and rapidly mounting concern about air transport's contribution to global warming.

Issues in Analysis of Deregulation Effects

The strongly held views on both sides of the debate often result in somewhat subjective analysis of the impacts of deregulation. It is important to consider what other factors may apply in the time period over which impacts are being assessed, and to avoid very short periods after the introduction of deregulation, which may give atypical results. Some changes may not be the result of deregulation *per se*, but of other factors changing during the same period. Examples include growth in car ownership (which would be expected to reduce bus ridership in any case), national economic growth (which tends to stimulate rail and air use), and the condition of the labor market (e.g., when there is substantial unemployment, it may be possible to reduce wages and worsen working conditions as a means of reducing unit costs – this applied to the local bus industry in Britain in the 1980s. It may also have influenced the ability of new low cost carriers to recruit aircrew at various periods). However, it may subsequently become necessary to raise wage and improve working conditions to recruit sufficient staff in a tighter labour market, as can now be seen in Britain.

Spatial Variation in Impacts

Spatial variation may also be an important factor, even within a single country such as Britain. For example, scope for lower wages at the time of local bus deregulation in the 1980s was greatest vis à vis previously standard national rates in areas with lower living costs and greater unemployment, such as the South West. Population trends will also affect total demand, irrespective of the regulatory system in place – this is an element in comparing the growth in bus use in London with decline in other major cities, for example. At an international level, these variations will be much greater. For example, per capita car ownership varies greatly, affecting the potential total market for public transport. Overall population density will affect the scope for rail services.

In some respects, deregulation impacts may be greater on higher density flows, where there is scope for additional carriers and a wider variety of service options to be introduced – for example, on trunk intercity routes, offering a range of rail, coach and air services with a wide mix of fares, frequencies and service quality. Conversely, lower density areas may experience little impact, or even a worsening of service where cross subsidy previously applied.

An effect which runs counter to some economic theories has been the continued role of large operators in markets subject to extensive competition. Indeed, the role of such groups continues to grow in many cases. Where there are no substantial economies of scale, one might expect a large number of small competing operators to characterize a market. One example already mentioned is that of the express coach industry in Britain, where an operator able to offer a well marketed national network may be in a much stronger position than smaller organizations. Even in local bus operation, where a large network may be of limited value, a process of consolidation has also been very evident, notably through groups such as Stagecoach, First, and Arriva which in total provide about 60% of local bus services by turnover. Among the factors accounting for this may be some economies of scale in vehicle and materials purchasing (even through the largest single element of costs, labor, is not so affected), concentration of management expertise, and access to capital (e.g., by floating on the stock market). In some cases, a common 'brand', such as the 'Megarider' travelcard of Stagecoach, may be more easily marketed as a single product.

Likewise, in the airline sector, while many new operators have emerged, many have been short lived, and some individual companies have become much more successful, securing growing market shares (notably Ryanair).

Similar concentrations may also be seen in a market characterized by franchising (in part because the minimum size of franchises requires a relatively large organization to make effective bids), such as that for passenger rail services in Britain. The same groups dominant in the bus sector (along with National Express) play a similar role in rail. Some groups are developing international activities – for example, Arriva of Britain has many subsidiaries elsewhere in Europe, and French operators play a growing role in the British market.

International forms of ownership are also developing in the airline industry, and the traditional 'flag carrier' is being replaced by larger groupings (such as the Air France/KLM merger). However, ownership restrictions applied by the USA may limit this process on a wider scale.

Looking over a wider range of countries, it is evident that many European states retain relatively strict forms of regulation and ownership, notably larger states such as France and Germany, which continue to restrict the degree to which coach and rail can compete, for example. However, there is also growing concern about the high cost of supporting some publicly owned or regulated transport services, resulting a wider interest in com petitive tendering (although not deregulation *per se*). This can already be seen in Scandinavian countries in par ticular, and is also evident in The Netherlands, for ex ample. A common approach to provision of competitively tendered services is now being adopted by the European Union, albeit with much debate and revision to the initial proposals. This may result in greater competition where currently a single monopoly provider is found. However, it would not necessarily result in greater direct com petition for passengers over the same routes by different operators.

See also: Aviation; Transport, Public.

Further Reading

ANON (2005). Can Ryanair stick to the winning formula? *Aviation Strategy*.

Bradshaw, W. and Lawton Smith, L. (eds.) (2000). *Privatisation and Deregulation of Transport: Proceedings of a Conference at Oxford University July 1998*. London: Macmillan.

Commission of the European Communities (2005). Proposal for a Regulation of the European Parliament and of the Council on public passenger transport service rail and by road. COM(2005) 319 Final. Brussels.

Dennis, N. (2007). Stimulation or saturation? Perspectives on the European low cost airline market and prospects for future growth. *Transportation Research Board Research Record* 2007, doi: 10.3141/2007 07.

Dobruszkes, F. (2006). An analysis of European low cost airlines and their networks. *Journal of Transport Geography* 14, 249 264.

Doganis, R. S. (2001). *The Airline Business in the 21st Century*. London: Routledge chs. 2 and 6.

Estache, A. and Gomez Lobo, A. (2005). Limits to competition in urban bus services in developing countries. *Transport Reviews* 25(2), 139 158.

European Conference of Ministers of Transport (ECMT) (2007). (De)Regulation of the taxi industry. Round Table 133. Paris.

Fan, T. (2006). Improvements in intra European inter city flight connectivity: 1996 2004. *Journal of Transport Geography* 14, 273 286.

Freeman, R. and Shaw, J. (eds.) (2000). *All Change British Railway Privatisation*. London: McGraw Hill.

Hensher, D. A. (ed.) (2005). *Competition and Ownership in Land Passenger Transport: Selected Refereed Papers from the 8th International Conference (Therdbo 8), Rio de Janeiro, September 2003*. Oxford: Elsevier chs. 1, 9, 14, 15, 20, 22 and 28.

Hibbs, J. (2006). *The Railways, the Market and the Government*. London: Institute of Economic Affairs.

Shaw, J. (2001). Competition in the UK passenger railway industry: Prospects and problems. *Transport Reviews* 21(2), 195 216.

Trethaway, M. W. (2004). Distortions of airline revenues: Why the network airline business model is broken. *Journal of Air Transport Management* 10, 3 14.

Van Der Veer, J. P. (2002). Entry deterrence and quality provision in the local bus market. *Transport Reviews* 22(3), 247 265.

White, P. R. (1997). What conclusions can be drawn about bus deregulation in Britain? *Transport Reviews* 17(1), 1 16.

Relevant Websites

http://www.caa.co.uk
Civil Aviation Authority.

http://www.cpt uk.org
Confederation of Passenger Transport UK.

http://www.rail reg.gov.uk
Office of Rail Regulation.

http://www.pteg.net
Passenger Transport Executive Group.

Transport and Globalization

K. O'Connor, University of Melbourne, Parkville, Australia

Glossary

Containerization The use of a box with internationally agreed dimensions for the shipment of goods.

Gateway City The first place of contact of a transport network in a nation or region.

Global Cities Concentrations of the production of high-level services and the focus of international commercial links between nations.

Hubs Concentrations of passenger and/or freight traffic.

Intermodal Movement of freight between transport modes.

Logistics The comprehensive organization of the movement of freight.

Production Systems The ways that firms arrange the collection of raw materials and/or components and the delivery of final goods.

Supply Chains Links within and between firms for the movement of goods.

Transport Systems The organization of transport by routes and modes.

Globalization involves the dispersal of production across the globe. Transport systems of various kinds have enabled that dispersal to take place. This can be seen in historic examples as production took tentative steps from a small known world to a global network. It is apparent today in the tightly organized production systems that involve collection and distribution to and from sites spread within and between continents on regular timetabled services. These services tend to favor hub cities, which now stand out as the key players within most transport systems. These changes have reorganized spaces within cities, as airports have emerged as new centers of activity, seaports and airports have been relocated, and new transport infrastructure built to cope with the movement of freight and passengers. Change continues with some smaller cities, and the fringes of cities, feeling the impact of new transport technologies and methods.

Introduction

Transport is perhaps the *sine qua non* of globalization. In a very broad sweep of history, the stages in the spread of production across the globe are closely linked to eras in transport systems. These eras appeared following the efforts of explorers and then matured as technicians refined the ships, trains, trucks, and eventually planes to carry people and freight. Through their evolution and maturation, these methods of transport extended the mobility, speed, and reach of goods and services of firms, merchants, and also of people. Taking advantage of these gains enabled firms to expand by accessing new and often cheaper sources of raw materials, as well as by supplying larger markets. Realizing those opportunities needed new management skills in trade and transport operation. Those with the requisite skills in effect managed the global economy, and reshaped the size, location, and internal organization of cities where they worked. The pressure of those forces is still being felt with some new cities playing a role in global transport, new skills being applied in old cities, and new patterns emerging within them to accommodate new needs.

A Historical Perspective on Globalization and Transport

The global economy emerged in stages. For a time it was expressed in small amounts of trade around the Mediterranean Sea, and into the agricultural communities of the Middle East and beyond. A series of cities emerged along the transport routes across what was a small part of that small current world. Better organization allied to higher standards in shipping and land transport spread the reach of this trade and eventually very long supply chains connected Western Europe to much of Asia. Many of these supply chains were organized via Venice. Merchants there, familiar with a range of sea and land transport systems, were able to connect sources of materials and supply markets that were long distances away to their east and also to their west. Venice became in its time what we now call a 'global city'. Later, sea transport improvements would redefine the scale and scope of the global economy in terms of ocean frontages rather than inland caravan routes; the capacity to manage and direct this new set of supply chains fell to Atlantic coast rather than Mediterranean cities. Amsterdam was especially prominent in this era.

The importance of ocean frontages was then enhanced by land transport improvements. First canals, then rail, and eventually road linkages enabled the goods from the sea to be carried to and from the resources and markets inland. What followed was not only industrial expansion on a massive scale, most obviously in northwestern Europe, but also the accelerated expansion into new territories associated with colonialization. In the colonies, ocean based points of contact provided the foundation for subsequent urban settlement of continents, and eventually nations. The colonial cities dotted along the US, Canadian, Australian, African, Latin American, and Asian coastlines provided the local infrastructure for supply chains using global sea transport and regional (and local) land transport to connect farmers, and miners in many places across the globe to factories and con sumers in Europe. Return voyages bought consumer goods and the equipment needed for colonial manage ment and development. Shipping was the core element of these supply chains and provided the exotic as well as the low cost supplies to Europe, while also providing access to large new markets. Steady technical improvements in shipping meant that transport that was once scarce, dif ficult, and unpredictable became more reliable. The na tion (Britain), and the city (London) that managed much of the new arrangements in shipping gained most from this era. In the colonies, a gradual sorting out of the importance of places took place as shipping service began to select some ports over others. Places that acted as gathering points or key stops along the way had special prominence. Singapore owes its emergence as major port to that effect. The management of the colonial networks eventually required better contact and electronic com munication systems were developed. These snaked ten tatively across the globe, following some of the sea routes and focusing on the large colonial seaports. The skeletal structure they created has been replicated in the complex webs of optic fibers that now span the globe.

In a more recent period, and in another part of the globe, the location and growth in economic activity was also closely linked to the utilization of sea transport. A new focus of economic activity in Asia emerged once transport innovations allowed large scale shipping of coal, iron ore, and oil. Those innovations were eventually reflected in the global domination of the steel mills of Japan and later Korea. That steel was later built into global exports of cars, ships, and industrial machinery.

Around the same time the mobility of all sorts of in dustrial goods, and the role of shipping, was increased substantially once a simple standard container was adopted by the world's shipping lines. This piece of equipment meant that goods could be packed in one part of the world and moved easily to any other part of the world, lifted from ship to train to truck before arriving intact at a customer's factory or warehouse. The scope and significance of this simple physical move was ex panded substantially when digital information was added to the box, and the data transmitted by global tele communications systems. The optically scanned code on the side of the box, connected to the shipping manifest, showing delivery arrangements at the destination enabled finely articulated logistics systems to be developed and run on predetermined schedules. Today, much of the stock for sale in the retail stores in the US or Western Europe was packed into a container on an industrial es tate somewhere in China several weeks earlier. The order for the container load might have come from a US warehouse computer connected both to the factory in China, and to a transport company (with sea, rail, and road connections) that would arrange the pickup and delivery of the container.

This new reliability and flexibility in intermodal trans port has enabled major redirections in the organization of production. Globalization is now expressed in production systems that fragment production into separate processes located across the world. It is now common for the finance and management, design, component production, and as sembly of a good or service to be located in many different places. Final deliveries are made to a wide array of des tinations. These global production systems only work when predictable global sea and air logistical services are avail able. This logistical capacity can have a direct role in production systems where delivery times, and their flexi bility, are built into production planning. In fact, in some firms corporate structure and organization has been re shaped to coordinate production in time as well as in space.

This ease of transport has changed the pattern of trade. Rather than one nation's goods or resources mov ing to another nation where they are not available, today's trade is made up of complex and multidirectional movement of the same goods and services between in dividual firms. For example, car parts and cars cross the Atlantic and the Pacific in containers in both directions as producers based in the US, Japan, and Europe draw upon supplies from, and arrange assembly in, places as diverse as Brazil, Mexico, and Thailand. Similarly, trade in ser vices is greatest between North America and Europe, the two leading sites of service production. That exchange accounts for the very substantial amount of business travel on aircraft between those two continents.

This new level of mobility of goods and services has been grafted on to an international trading economy that emerged over the previous 400–500 years. That graft has not been a simple match however, as the focus of traffic has changed and new ports, new air transport hubs, new rail, and road links have emerged. Once again, those capable in the management and organization of the new transport tasks have become prominent. Today that in cludes firms associated with the ports of Singapore, Hong Kong, and Dubai, as well as a shipping company from Copenhagen (Maersk) and an air freight company cen tered at Memphis (Federal Express). At the same time some of the old cities have retained their role in the management of the new network. For example, the Baltic Exchange (established in the eighteenth century) still provides a global market place for sea and air freight from its location in the City of London. Hamburg, too, remains as an influential operator in transport and logistics.

Geographic Impacts of Global Transport

The examples cited above show how the evolution in transport technology has been associated with the steady and evolving globalization of the world economy. That evolution has had some distinctive impacts. Many of these emerge from the geographically selective character of transport services. That can be seen initially in the linear connection between an origin and a destination that is provided by a mode of transport. This linearity creates a geographic selectivity first in which points are connected, and second in the benefits it offers places along the way. Hence, the caravan routes across Central Asia, the shipping routes across the Mediterranean, the shipping routes along the coasts of Africa and India, the rail net across the US, and airline routes from Asia to Europe all have a linearity, and all have created narrow corridors of impact across vast continents. Some places along these routes emerged as markets, production sites, or centers of administration. Those outcomes narrowed the local and regional impact of globalization and privileged some places over others.

That privilege is felt especially in 'gateway cities'. Gateway cities are places where global transport systems make local contact. Their role is best seen in the port cities on colonial coastlines, but their effect would have been felt in the caravan staging points of Central Asia, can be seen at rail junctions and termini, and now in the role played by major airports in cities. The transport privilege of a gateway city emerges because it offers first access to global goods and services and may have frequent services. It can be a place of management of the transport system, requiring effort in the organization of physical goods and information, dispersing inward supplies and assembling outgoing goods; that means it is a place of local trading, as inbound and outbound goods and services are sold by local merchants. Finally it has a set of functions (offices, storage facilities, technical servicing, and so on) not needed in other cities, which enriches its commercial activity. The gateway city effect was important in the development of New York, Montreal, San Francisco, Melbourne, Sydney, and Mumbai. Today, Hong Kong and Shanghai draw much of their vitality from their role as gateway cities in the Pearl River Delta and the Yangste Delta, respectively. These experiences illustrate how the operation of the transport system itself creates a locally uneven impact of globalization.

The Changing Geographic Impact of Transport

That uneven impact has been refined and changed as each transport system has evolved and in some cases replaced. This evolution has usually been expressed in both size and speed of vehicles. As ships, planes, and trains became larger and technically advanced in terms of speed and fuel consumption, the initial short stages between linear or coastal patterns of transport were gradually extended. Some earlier stopping points were bypassed and some original gateway cities were no longer needed as new technologies were adopted. Later, new standard roads would bypass some smaller towns, while high speed rail services operated with fewer stops.

This outcome is well illustrated in the concentration of sea traffic at a series of major nodes. Steady increases in the capacity and speed of container ships have changed the networks of shipping. Today, loads are gathered from regional supply as well as local sources for transshipment from small to large vessels in ports such as Hong Kong and Singapore before being moved across oceans (bypassing smaller ports enroute) to markets. In air transport, increased speed and size were provided by the introduction of jet engine, and the creation of the Boeing 747 aircraft. Both represented sudden major changes from existing technology. The first lifted the speed and range of aircraft considerably as the jet engine had much greater speed and lower fuel needs of the old piston engines. The Boeing 747 was a very major step up in size and range, enabling many more passengers (and freight) to be flown between airports. The first effect of these changes was felt in a substantial reduction in the number of stopping points along international routes. As an example, in the petrol engine era there were up to five or six stops between Singapore or Hong Kong and London; the Boeing 707 reduced these to three or four; later the Boeing 747 would eliminate stops on this route altogether. Across the Atlantic, steady improvement in aircraft speed and capacity obviated the need for technical stops at Shannon (Ireland) and Gander (Canada). In the northern Pacific, jet aircraft were able to use great circle routes between the US and North (and later Southeast) Asia. Initially these routes involved Anchorage as a technical stop over, but soon the routes were flown nonstop. The technical capacity to fly great circle routes across the polar regions (facilitated by the adoption of global positioning systems) had another important consequence: northern and eastern US cities like Minneapolis, Chicago, Detroit, and New York became as close to Asia as San Francisco and Los Angeles. Up until this time the latter cities had been the main gateways to Asia, fed by traffic across the central Pacific via Hawaii and Guam. Flying times to Europe were also reduced, and daily nonstop flights between Tokyo and London, Paris, Amsterdam, and Frankfurt became common. Simultaneously, places like Honolulu, Anchorage, Mumbai, and Bahrain had less daily through traffic.

The geographic effect of these technical developments in the jet aircraft network (especially the increase in the size of aircraft) was not limited simply to the bypassing of places along a route. As was occurring in the concentration of sea freight in containers, there was a gathering of larger numbers of passengers at a smaller number of places, labeled 'hubs'. The large numbers of passengers were needed in part to justify the use of larger aircraft; this approach had a significant impact on the relative size of airports in Southeast Asia, with competition between Bangkok, Singapore, and Hong Kong for the regional hub role. In the US, the hub concept was used by airlines to reorganize their domestic services. Hub airports were fed by short links from surrounding cities, many of which had once been stops along the old trunk routes. As a result, hubs had more frequent and more diverse airline connections, so they began to assert some of the effects of the early gateway cities, providing good opportunities for business (and in many cases cheaper fares). Some were creations of the old transport system (such as London and New York), others reflected new patterns and capacity (Los Angeles and Singapore), others the home base of airlines (Chicago, Dallas–Fort Worth, and Atlanta). In this way, the air transport system has redirected the urban effects of globalization. In fact, the hierarchy of global cities is now mirrored in the hierarchy of busy airports. Today's large airport cities can be seen as a spatial expression of the working of the global economy.

It is possible that outcome may be changing a little as airline services are increasing at mid sized cities. That reflects the development of smaller long haul aircraft that can provide point to point services to pairs of nonhub cities. Debate about whether the future will involve more very large aircraft flying from hub to hub, or more services between mid sized cities on smaller aircraft, is ongoing.

Similar change can be seen in the technical development in sea transport. Steady expansion in the size and speed of shipping reduced the need for stops on long routes, so removing the servicing and management roles from many places, and also their regular delivery of goods. The adoption of containers accelerated this selectivity; that change led to a steady rise in ship size, so that the hub system created for large and long distance aircraft has been replicated by large and long distance shipping services. Hub ports were fed by delivery from other cities; in some cases rail and road freight services facilitated the gathering and dispersal of containers to and from a hub, replacing earlier coastal services. The scale of container ships is continuing to rise, so that the hub to hub service seems likely to remain significant in the future.

The data displayed in **Table 1** provide an indication of just how powerful the hub system has become in global transport. Almost one half of all the containers moved through the 100 busiest port cities are handled by those cities ranked in the top ten; the top 20 ports handle almost two thirds of the total. Air freight is similarly concentrated: the share of freight moving through the top ten airports as not quite as significant, but the top 20 account for 60% of the tonnage handled. Air passenger movements are less concentrated: one has to include the top 30 cities to account for 60% of all passenger movements. These outcomes illustrate that mix of size, speed, and reach of ships and planes has consolidated and strengthened hub operations.

Table 1 Shares of transport activity at top 100 airports and seaports 2002

Cities ranked in	*Sea ports*	*Air ports*	
	Container movements TEUs	*Passenger number*	*Air freight tons loaded*
Share of traffic			
Top 10	48.0	30.5	38.7
11 to 20	14.6	15.0	22.3
21 to 30	9.0	13.3	12.4
31 to 40	7.0	9.5	7.3
41 to 50	6.4	7.7	5.6
51 to 60	4.8	6.6	4.1
61 to 70	3.5	5.6	3.3
71 to 80	2.5	4.6	2.4
81 to 90	2.2	4.0	2.3
91 to 100	2.0	3.2	1.6
	100.0	100.0	100.0

Source: Containerisation International 2002.
Airports Council International 2002.
Note: Data for cities with more than one port or airport have been merged.

Although hubs are prominent, there is a well developed technical capacity to manage the transshipment of goods from large to small ships, from ships to rail or road, from large planes to small planes, and from planes to trucks. That technical capacity (based on information technology) ensures that the transport system is well equipped to collect from, and deliver to, small places. Hence, perhaps paradoxically, the logistics system in modern transport has allowed the links associated with globalization to be felt well beyond the larger sea and airport cities of the world. That has made the sea container (and on some continents the parcel delivery system) almost ubiquitous.

Global Transport and Change within Cities

Transport systems have long provided one cornerstone in the structure of a metropolitan area as seaports stimulated commercial zones nearby and rail stations generated passenger and freight handling facilities. Today, the airport is the key focus of movement of people into and out of a city. At the same time airfreight has increased, so that the warehouses and storage facilities that once sur rounded the seaport have been replicated (at a larger scale) around airports. In addition, the construction of hotels and business parks has made the airport a new commercial center in a number of cities.

At the same time the hub role of some cities, which involves handling of freight at a much greater physical scale than in the past, has challenged the fortunes of the original seaports and rail yards. The traffic associated with these new transport systems, and the storage space needed, has created many local planning problems. In some cities, airports, seaports, and rail terminals have been relocated to the outer fringe. Special local solutions, including new road or rail links dedicated to freight movement across a city, have been adopted to accom modate (and redirect) the freight traffic. Where new capacity cannot be provided cities have managed to ex pand and retain activity through physical extension, along with innovative management, of existing facilities: Singapore and Hong Kong illustrate that outcome with seaport planning, while London and Chicago show the way in airport planning. In many cities, however, old port and rail yards remain unused and derelict, reminders of the forces of change in the global transport system. Only a few have become potential redevelopment sites (no tably the redevelopment of London Docklands).

The relocation of transport facilities to suburban and even distant fringe locations has contributed to the spread of urban development in large metropolitan areas. That outcome reflects the power of logistical systems to connect producers to one another and to consumers, which in effect has spread the benefit of some agglom eration economies across the whole of a metropolitan area rather than being available just at its core. In effect, some suburban or fringe sites are now effective locations for many businesses, even those serving global markets.

In overview, then, globalization is the long shadow of the transport system. Over time, transport has changed in technical terms, and in its management and organization. As it has done so it has usually involved more of the world (although concentrated in some parts), more of the cities (although concentrating on some at the expense of others), and has often shifted the focus of metropolitan development to outer and fringe locations. That complex set of outcomes means an understanding of the impact of globalization is difficult to comprehend without understanding the current stage and trends within transport.

See also: Global Commodity Chains; Global Production Networks; Globalization and Transnational Corporations; Transport Geography.

Further Reading

Burchardt, A. (1971). A hypothesis about Gateway Cities. *Annals of the Association of American Geographers* 61, 269 285.
Goetz, A. (1992). Air passenger transportation and growth in the US urban system 1950 1987. *Growth and Change* 23, 217 238.
Gordon, P., Richardson, H. W. and Yu, G. (1998). Metropolitan and non metropolitan employment trends in the US: Recent evidence and implications. *Urban Studies* 35, 1037 1058.
Hesse, M. (2004). Logistics and freight transport policy in urban areas: A case study of Berlin Brandenburg/Germany. *European Planning Studies* 12, 1035 1053.
Ivy, R. I., Fik, T. J. and Malecki, E. (1995). Changes in air service connectivity and employment. *Environment and Planning A* 27, 165 179.
O'Connor, K. (1995). Airport development in Southe East Asia. *Transport Geography* 3, 269 279.
O'Connor, K. (1996). Airport development: A Pacific Asian perspective. *Built Environment* 22, 212 223.
O'Connor, K. (2003). Global air travel: Toward concentration or dispersal? *Journal of Transport Geography* 11, 83 92.
O'Connor, K. and Scott, A. (1992). Airline services and metropolitan areas in the Asia Pacific region 1970 1990. *Review of Regional and Urban Development Studies* 4, 240 253.
Rodrigue, J. P. (1994). Transportation and territorial development in the Singapore extended metropolitan region. *Singapore Journal of Tropical Geography* 15, 56 74.
Smith, D. S. and Timberlake, M. (1998). Cities and the spatial articulation of the world economy through air travel. In Cicantell, P. & Bunker, S. (eds.) *Space and Transport in the World System*, pp 213 240. Westport, CT: Greenwood Press.
The Economist. (2006). The Physical Internet. *A Survey of Logistics,* 379(8482), (June 17).

Relevant Websites

http://www.logisticsquarterly.com
Ideas for Logistics and Transportations Magazine.
http://outsourced logistics.com
The Global Supply Chain authority.
http://www.tli.gatech.edu/
The Supply Chain and Logistic Institute.
http://siteresources.worldbank.org/INTTLF/Resources/lpireport.pdf
The World Bank.
http://www.fhwa.dot.gov/freightplanning
U.S. Department of Transportation.
http://www.worldcargoalliance.com
World Cargo Alliance.

Transport and Social Exclusion

J. Hine, University of Ulster, Newtownabbey, UK

Glossary

Reverse Commute Problem Travel from inner cities to suburban locations.

Social Exclusion Nonparticipation in normal activities of society.

Transport Disadvantage Poor access to transport.

Introduction

The link between transport provision, policy, and social exclusion is complex, but it is also clear that there are powerful synergies with poverty, sustainable development, and the regeneration of communities. Social exclusion is closely associated with 'social justice' and 'accessibility', terms that have been cited in the literature for a considerable time now. More recently, studies in the fields of human geography and transport studies have demonstrated that the link between transport and social exclusion is essentially related to differential levels of access to modes of transport, often referred to as transport disadvantage, and poor access to goods and services. More recently, research efforts focused on linkages with well being and aspects of quality of life which make it difficult for people to participate fully in society. A number of factors are seen to contribute to social exclusion; it is clear that access to transport is central to all these. These include differentials in education, training opportunity and attainment, socioeconomic circumstances, local environment as well as access to information and physical accessibility to a wide range of opportunities, including employment, shopping, and recreation.

Internationally, attention has been focused on low income neighborhoods and where social and economic changes have resulted in joblessness; the concentration of vulnerable people in deprived neighborhoods; family breakdown; poor core public services and public service failure; and the declining popularity of social housing. The problem in these areas has been compounded by a lack of attention to links between poor neighborhoods in urban centers and local and regional economies, and the links between planning and economic development which can accentuate the barriers to work, education, and childcare. In the US, a significant form of social exclusion can be found in the isolation of poor inner city residents from jobs that are increasingly located in the suburbs and the urban fringe. This has resulted in the reverse commute problem among these groups. Transport disadvantage can also be acute in informal settlements and favelas where poor transport links emphasize economic and social dislocation.

A greater understanding of the geographical/spatial nature of social exclusion is critical to the delivery of transport solutions. The use of geographic information system (GIS) and accessibility measurement has become an important research and policy development area. New research areas have developed from this, including the analysis of activity spaces and social networks, and assessments of the impact of new information technology on travel.

Defining Social Exclusion

Social exclusion has come to be accepted as a term that refers to the loss in the ability (by people or households) to fully participate in activities and society and connect with many of the jobs, services, and facilities. There is broadly an agreement in the literature that social exclusion represents a conceptual shift away from the traditional forms of explanation and should not be considered equivalent to older terms and definitions previously applied to individuals, groups, and processes considered to exist and operate outside a certain social norm – such as poverty, deprivation, and the underclass.

Despite no common definition of social exclusion there is also no common definition of the dimensions and factors involved in it. But in both cases, the approaches taken by various authors, though different in detail, broadly overlap. Typically, these dimensional frameworks refer to the key areas of daily life. It is also recognized that the ability of a group or individual to participate across these dimensions could be affected by a number of factors. These include the individual's own characteristics, life events, and characteristics of the area resided in and social, civil, and political institutions of society. In sociology and studies of time use, this has been referred to as the structure of social obligations that often make travel necessary (e.g., legal, economic, and familial obligations).

In terms of transport shaping patterns of social exclusion, it is clear that there are a variety of processes that can impact on the synergies of this relationship. These processes can include: (1) the nature of time–space organization in households; (2) the nature of the transport system; and (3) the nature of time–space organization of the facilities and opportunities individuals are seeking to access. The nature of these will differ according to gender, age, cultural background, level of ability, and economic circumstances. The seven categories of ex clusion suggested by Church *et al.*, connected to trans port, are:

- 'physical exclusion': refers to a situation where phys ical barriers inhibit the accessibility of services which could be experienced by mothers with children, eld erly, or frail, those encumbered by heavy loads or those who do not speak the dominant language of the society;
- 'geographical exclusion': refers to a situation where poor transport provision and resulting inaccessibility can create exclusion not just in rural areas but also in areas on the urban fringe;
- 'exclusion from facilities': refers to the distance from facilities – for example, shopping, health, leisure, and education – from people's homes, especially from those with no car, makes access difficult;
- 'economic exclusion': refers to a situation the high monetary or temporal costs of travel can prevent or limit access to facilities or jobs and thus income;
- 'time based exclusion': refers to a situation where other demands on time such as caring restrict the time available for travel;
- 'fear based exclusion': refers to a situation where worry, fear, and even terror influence how public spaces and public transport are used, particularly by women, children, and the elderly; and
- 'space exclusion': refers to a situation where security and space management strategies can discourage so cially excluded individuals from using public trans port spaces.

Of course, it can be argued that transport exclusion mechanisms can be peripheral to non transport mech anisms but what is important to recognize is that trans port, or rather the lack of access to it, can compound these factors. At a local level it is clear that there is a need for the identification of a selection of indicators that reflect the processes linked to social exclusion and in particular, the role of transport in that exclusionary process. Indeed, the idea of introducing dimensional frameworks to the debate on the links between transport and social exclusion highlights the need for approaches that can identify this range of experience. Research studies have now started to address this using travel diaries combined with social surveys and GISs.

Social Exclusion and Patterns of Transport Disadvantage

Internationally, research has identified that those groups most likely to experience transport disadvantage are those on low incomes, women, elderly and disabled people, and children. Evidence also indicates clearly that multiply deprived households are highly intercorrelated with other factors such as low incomes, low levels of car ownership, and public sector housing. Essentially, these groups are those with traditionally lower levels of access to cars. Race is also an important indicator of transport access and social disadvantage.

Low-Income Groups

For Great Britain, the National Travel Survey, for ex ample, shows that over the period 1989–91 to 2004 although the number of households in the lowest real income quintile with no car has declined by 20%, a substantial proportion in this income group have no car – 54% in 2004 (**Table 1**). This is despite evidence which indicates that the availability of bus services has changed little over this period in urban areas and has actually increased in rural areas. (In the National Travel Survey this is measured as households within a 13 minute walk of an hourly bus service.) Low income families are moving into car ownership as a response to rising public transport fares and poorer levels of public transport ac cessibility to different labor and housing markets in creasingly in peripheral locations.

Walking and public transport are the dominant modes of transport for people from households on low incomes, but in particular for non car owning households in the lowest income quintile.

Table 1 Household car ownership by income band, 1989/91 and 2004

	1989/91			*2004*		
	None	*One*	*Two or more*	*None*	*One*	*Two or more*
Lowest real-income quintile	73	24	3	54	38	7
Second quintile	48	44	8	37	47	15
Third quintile	25	55	20	20	52	28
Fourth quintile	12	55	33	11	46	43
Highest real-income quintile	7	46	47	8	40	52

Source: Department for Transport (2005) National Travel Survey.

Women

Women also experience exclusion in a number of ways as a result of poor public transport services. There is evi dence from many parts of the world that travel patterns differ by gender, including, for example, the United States, Sweden, Australia, and most of Africa. There are clear issues affecting women's transport which relate to patterns of travel, patterns of employment, income, car ing responsibilities, and access to forms of travel (par ticularly access to cars). There are also differences among women in terms of the experiences of specific groups (e.g., older women, disabled women, women from ethnic minorities, and women living in rural areas and lone parents). As with older people and the disabled, the de sign of the infrastructure can mitigate against the use of a local transport system. Women with young children are perhaps hardest hit in this respect. Personal safety when using or trying to access transport infrastructure is also a major consideration for this group.

Disabled

Disabled people are a group that also features in dis cussions surrounding the link between transport and social exclusion. They suffer because for a variety of reasons they find it difficult to access public services. These reasons include: low incomes, physical layout of infrastructure and design of vehicles, and location of stops. The restructuring of bus services to the edges of residential and commercial areas on main transport corridors could potentially have a profound effect on this group. There is, however, little evidence to suggest the extent and nature of such service restructuring exercises on this group.

The Impact of Transport Disadvantage and Social Exclusion

Barriers to Employment

Poor transport access is an important barrier to em ployment opportunities. For young people in rural areas, transport is vital to holding down a job; public transport is often seen as unreliable with timetables that do not match up with work schedules. Cost and availability of childcare, lack of knowledge of the local job market, and an unwillingness to travel outside the locality can also be barriers to employment. In disadvantaged housing areas, the economic reality is sharply defined by the resource constraints of childcare, travel to work time, cost, and availability. Particularly, when confronted by low wages and an increasing proportion of part time work, travel to work costs may make taking a low income job an un economic option.

Exclusion from Services

Lack of readily available transport, whether car or public transport, has a clear impact on whether particular goods and services can be accessed. In the case of public transport the problem, especially for communities with low levels of car ownership, is that services are more likely to be located on transport corridors. This when combined with timetables that do not accommodate new forms of employment (e.g., shift work) means that access can be problematic and that temporal barriers to job markets have been created.

Lack of personal transport and poor public transport also have been found to reduce the ability of people, living in disadvantaged areas, to access financial services, and creates difficulties in accessing health facilities for women. The cost of public transport is often cited as a barrier to accessing further education.

Fear and Perceptions of Safety

Inadequately managed and structured public transport systems can be responsible for influencing perceptions of safety and fear. Perceptions of safety and fear can have significant effects on levels of personal mobility. Older people, women, and those from ethnic communities are more likely to fear crime while using public transport. Behaviors found include avoidance of making a trip, es pecially in the dark. Fear of interchange facilities and stations in the dark and at off peak periods, have high lighted the need for a security presence in these locations.

The consequence of this fear is that trips are either not made or that alternative arrangements are made where it is possible to avoid these situations. Switches from bus to car are often a result of these fears and the inability of the services to meet shift patterns. Other work has found that taxis can play a very important role in these circumstances. For younger people, work has found that the anxieties experienced when using public trans port are similar to those for adults. Young women in particular feel very unsafe after dark when using public transport.

Policy Responses

Increasingly, local services and activities are located in inaccessible places for noncar owners, new sites for em ployment and housing are also located on the edge of towns and cities. Public transport networks often do not adequately serve these out of town/edge of town lo cations. In many towns and cities during the off peak (early morning and late evening) periods, buses can be a rare commodity. Typically, social exclusion and its re duction through improved public transport are treated as a general policy aim at the local government level. This may reflect a lack of control over public transport op erators, following de regulation, with regard to the price and quantity of public transport; however, policies are in place for concessionary travel and the buying in of so cially necessary public transport services.

Policy responses, typically, are based around ap proaches aimed at: (1) improving levels of personal mobility; (2) providing mobile key services; and, (3) nontransport solutions. Traditionally, policy inter ventions for the transport disadvantaged in this area have emerged from the specialist transport provider who has sought to address those gaps in provision not filled by mainstream transport providers. Also, other interventions have resulted from the need to introduce concessionary travel for older and disabled people and the growth of guidance and legislation on these matters.

Improving Levels of Personal Mobility

Bus policies

Public transport network coverage is a key public policy issue. The commercialization of local bus services, seen as essential by operators for future business growth, has resulted in the development of Metro type urban bus routes or high frequency corridors in many towns and cities. In rural areas, the focus has often moved to the provision of fewer higher frequency or limited ser vice routes. In these situations, the emphasis is not on route subsidy from public funds, although in the UK where operators control a large part of the network there are examples of cross subsidy between routes. A con sequence of the trend toward high frequency corridors in urban areas has been a movement away from the pro vision of socially necessary services especially in the off peak periods. These are often in or adjacent to areas with high proportions of public sector housing.

In rural areas, network shrinkage has been endemic partly in response not only to higher levels of car own ership in these areas but also due to cost cutting meas ures associated with a reduction in revenue support for public transport. In other European countries, compared to the UK, this trend has been less pronounced due to higher levels of subsidy. In the US, the emphasis has tended to be on capital projects for rail and the devel opment of public transport to address the reverse com mute problem for low income inner city residents trying to access jobs in the suburbs and welfare to work schemes.

Nonetheless, funding support for innovative rural public transport schemes is very unusual; funding tends to only cover more regular bus operations. Despite this there is growing evidence of rural operators beginning to work more closely with community transport operators. For example, in the Messary Valley in Southern Crete, the local municipality decided to operate a regular service for transporting students to school and local in habitants to shopping and health facilities in the nearby town. In Galway, Ireland, the coordination of private bus operators, voluntary transport operators, and non emergency health transport operators has been used to provide a network of flexibly routed and demand re sponsive services. In England, increased funding of rural public transport in England through the Rural Bus Subsidy Grant and the Rural Bus Challenge (RBC) has encouraged innovation particularly with regard to the development of demand responsive systems and com puterized booking. In terms of the impact on community transport operators, it has been clear that consultation between local authorities and community transport op erators has increased and that there has been a fostering of community transport through the joint use of com puterized booking systems. There is also evidence that in these RBC schemes there has been sharing of vehicles with community transport schemes; these arrangements not only maximize vehicle usage but also provide add itional revenue for the overall RBC scheme. In Northern Ireland, a rural transport fund has been established to integrate community transport operations with Ulsterbus services and assist with the development of social car schemes.

An established method of improving access to bus services is through a general or targeted subsidy. Until recently, there has been limited evidence to suggest the benefits of this approach for low income groups. Work has suggested that the cost of subsidies and fare reduction are less than has been previously thought. A fare re duction in metropolitan areas in money and time savings to passengers and other road users would outweigh the costs of the subsidy.

During the rise of neoliberal policies around the world, pressure to reduce subsidies grew. For instance, in the UK the 1985 Transport Act, which de regulated bus services in Great Britain, heralded the end of low fare policies; in other words, passenger transport authorities and local authorities could no longer subsidize bus ser vices except those that were deemed to be socially ne cessary and unprofitable. De regulation in Merseyside and resultant fare increases have been cited as the main cause of reduced use.

Targeted subsidies is another approach that is used to grant concessionary travel to pensioners, the disabled, children under 16, and students aged up to 18 years in full time education. Research indicates that these schemes encourage travel – those with concessions travel more often and further. Nonetheless, it is possible for other groups to be included in a concessionary scheme on a voluntary basis. Improving the accessibility of ser vices is another aspect of bus policy. The UK Disability Discrimination Act 1995, which legislates for mainstream public transport to become accessible to the disabled and wheelchair users, when combined with local transport strategies and quality partnerships will ensure a fleet of accessible vehicles. The adaptation of street infra structure, including bus boarders and raised platforms, upgrading of bus shelters, and the enforcement of parking in bus lanes and around bus stops, is also an important component of this approach.

Taxis and specialist transport services

Specialist services are typically provided by the volun tary sector. These services typically consist of: group hire bus services; dial a ride services, and voluntary car schemes. The objectives of dial a ride services were originally to provide a demand responsive service serv ing low density suburban areas. As a concept they have been around in the UK since the early 1970s. Initial experiments with this form of transport found that the services were expensive and failed to cater to dispersed trip patterns. Despite this, the approach remains an ac cepted method of delivering services to the elderly and disabled, a section of the population where transport needs can be expensive. The efficiency of dial a ride systems has been improved by computerized scheduling packages that in effect provide the operator of services with a reservation system for services. A variety of dial a ride services are now offered by community transport operators.

Another specialist transport concept is the ‘service route’. This is concerned with bringing the bus service closer to the residents and as such represents a move away from traditional route planning, which is based on radial routes coming into a city or town center. The service route is a regular route network but the route is based on where the proportions of elderly and disabled people live and important destinations, such as health centers, hospitals, and shops. These schemes have been implemented internationally in Denmark, Finland, Norway, The Netherlands, Canada, and the US. Available evidence indicates that they are cost effective and reduce the need for specialist transport.

Taxis are the most flexible transport service and are a popular alternative to other modes of transport even though they are the most expensive form of transport in the UK. On average taxis are 5 to 7 times more expensive than other modes per passenger mile while those groups in the population that tend to use taxis the most also tend to be on lower incomes, although there has been a growth in their usage by the general population for leisure purposes. To combat the high cost taxi card schemes exist as a subsidy for travel by this mode. Two forms of taxi operation have developed. Essentially, there are those taxis that are run through voluntary driver schemes and taxis operated by commercial firms. Voluntary car schemes have been concerned with transporting people for social services, health, and education purposes; however, this role has expanded to shopping and leisure based trips. These schemes have been effective although funding and volunteer resources do dictate their avail ability, which is restricted according to specific eligibility criteria.

Virtual Mobility and Nontransport Solutions

Potentially, many transport and accessibility problems can be addressed by non transport solutions. These so lutions encompass the use of mobile services for rural communities, including banks, doctors, and dental sur geries. The use of information communication technol ogy (ICT) now also offers opportunities for these types of resources and services to be offered remotely as in the case of banking or booked in advance of use. Virtual mobility has been seen as a means to reduce work related travel, but it also offers transport poor households op portunities to access work, goods, and services. ICT can allow home based working enabling people from ex cluded groups with low levels of mobility to enter the labor market. New ICT approaches also allow more flexible hours of working for those with caring re sponsibilities. Nonetheless, at the time of writing, more data are required to assess the impact of online services in the areas of health, learning, and leisure before the impact of this new technology can be assessed.

Conclusion

Improving levels of personal mobility through transport investment can play an important role in reducing social exclusion. Patterns of transport disadvantage are present in many towns and cities and represent a key challenge to transport policy. Human geography has a key role to play not only in the development of new solutions but also in the spatial analysis of these transport problems. The problem with social exclusion, associated with its com plexity, is how it can be used to develop policy goals. Nonetheless, the ideas of basic needs and well being are powerful when the experiences of those on low incomes and in deprived communities are examined. Indeed, it is the examination of these transport experiences that should prompt closer examinations of the allocation of resources and the distributional consequences of these policy decisions.

Trends in transport use and the operation of public transport networks however seem to downplay the im portance of accessibility to goods and services for the transport disadvantaged; instead, the policy has been developed favoring mode shift objectives which chal lenge habitual car use. This often obscures the basic needs function of public transport and specialized com munity transport operations. Analyses in transport and human geography need to be extended to provide nar ratives and spatial analyses of the transport choices of car poor households where travel time budgets are clearly of a different structure.

See also: Governance, Transport; Transport Geography.

Further Reading

Axhausen, K. (2007). Activity spaces, biographies, social networks and their welfare gains and externalities: Some hypotheses and empirical results. *Mobilities* 2(1), 15 36.

Cervero, R. (2004). Job isolation in the US: Narrowing the gap through job access and reverse commute programs. In Lucas, K. (ed.) *Running on Empty*. Bristol: Policy Press.

Church, A., Frost, M. and Sullivan, K. (2001). Transport and social exclusion in London. *Journal of Transport Policy* 7(3), 195 205.

DETR (2000). *Social Exclusion and the Provision and Availability of Public Transport*. London: DETR.

Gaffron, P., Hine, J. and Mitchell, F. (2001). The Role of Transport in Social Exclusion in Urban Scotland: Literature Review, Scottish Executive Central Research Unit, Edinburgh.

Hine, J. and Grieco, M. (2003). Scatters and clusters in time and space: Implications for delivering integrated and inclusive transport. *Transport Policy* 10(4), 299 306.

Hine, J. P. and Mitchell, F. (2001). The Role of Public Transport in Social Exclusion, Scottish Executive Central Research Unit, Edinburgh.

Hine, J. P. and Mitchell, F. (2003). *Transport Disadvanatge and Social Exclusion: Exclusionary Mechanisms in Transport*, Transport and Society Series. Aldershot: Ashgate Publishing Limited.

Lucas, K. (2004). *Running on Empty Transport, Social Exclusion and Environmental Justice*. Bristol: Policy Press. Bristol.

Raje, F., Grieco, M., Hine, J. P. and Preston, J. (2004). *Transport, Demand Management and Social Inclusion*. Aldershot: Ashgate Publishing Limited.

Social Exclusion Unit (2003). Making the Connections: *Final Report on Transport and Social Exclusion*, Cabinet Office, London.

Relevant Websites

http://www.eclipse eu.net
ECLIPSE (European Cooperation and Learning to Implement Transport Solutions to combat Exclusion) is one of 24 projects funded under the European Commission 2nd Transnational Exchange Programme (TEP) which forms part of the Community Programme.

http://www.matisse eu.com
MATISSE (Methodology for Assessment of Transport Impacts of Social Exclusion) project is a preparatory action funded by the European Commission's Directorate for Employment and Social Affairs. The key objective of the project is to compile and validate an evaluation tool to assess the impact of transport related policy interventions on social exclusion.

http://www.foe.co.uk
UK Friends of the Earth study using GIS to assess accessibility levels of the bus network in the City of Bradford, UK.

http://www.socialexclusionunit.gov.uk
UK Government Social Exclusion Unit provides policy statements and reviews of policy.

Transport and Sustainability

K. Button, George Mason University, Fairfax, VA, USA

Glossary

Acid Rain The acidic precipitation that has harmful effects on plants, aquatic animals, and infrastructure.
Global Warming The increase in the average temperature of the Earth's near-surface air and oceans.
Greenhouse Gases Gases present in the Earth's atmosphere and warm near-surface temperatures through the greenhouse effect.
Kyoto Protocol An international agreement with the objective of reducing greenhouse gas emissions.
Polluter Pays Principle Principle that polluting party pays for the damage it causes to the natural environment.
The Café Standards US federal regulations intended to improve the average fuel economy of cars and light trucks.

Introduction

There is mounting concern about the long term ability of the natural environment to absorb the activities of mankind. While environmental concerns have taken a variety of forms over the years, as information has expanded and as rising incomes have led people to become more concerned about their natural surroundings, so a variety of issues have emerged. These have ranged from quite local matters of protecting the productive capacity of land to concerns, often with religious overtones, that certain actions will lead to Armageddon. The emergence of modern science, improved data, and the ability to carry through complex simulations at the end of the twentieth century has provided a more systematic basis upon which to examine this issue.

The notion of "Space ship earth," expounded by Kenneth Boulding (1966) and the subsequent work of the United Nations' Commission on Environment and Development (1987), with its focus on sustainable development, has introduced a more holistic and structured approach to the environmental debate. The basic argument of these studies is that current generations need to leave a legacy for future generations that does not diminish the resource base of the planet. There is no prescribed path for attaining this, indeed there may be many potential paths. This work led to more operational concepts that involve economic sustainability, environmental sustainability, and social sustainability to ensure that the overall objective is met.

Although the concept of sustainability is very much a holistic one, there has been some sectoral digression from this comprehensive approach as notions such as sustainable energy, sustainable transport (sometimes couched as sustainable mobility), and sustainable agriculture have emerged. This type of stove piping moves away from the original idea of sustainability that essentially allowed transferability between activities provided the overall environmental threshold was retained. In other words, there is nothing inconsistent within the holistic approach in having more atmospheric damage resulting from society using additional carbon fuel for heating as long as there is at least a comparable reduction in its use in other activities such as using the motorcar. The narrower definition of sustainable transport that emerged in the late 1980s has, however, led to the seeking of ways to move the existing volume of people, goods, and information so as to limit its impact on the environment, the economy, and society without fully considering possible social trade offs with other activities.

The situation is also a dynamic one. The nature and scale of the challenges confronting policymakers is continually changing as new scientific evidence emerges on both the local health and social effects of transport produced pollution, and as a better understanding of global environmental trends materialize. At the same time, while challenges are continually shifting, so new policies and institutions, as well as market induced feedbacks, are emerging to provide potential ways of reconciling the various social and environmental costs and benefits of transport.

The Challenges

Transport takes various forms and there are many aspects to it (**Figure 1**). From an environmental perspective, the issues are not simply the amount of transport but also who uses it and the particular mode favored. The environmental damage associated with transport depends as much on whether it is used for passenger movements or freight (e.g., the former is highly concentrated and generally goes close to environmentally sensitive urban areas) as the amount moved. It also depends on the mode of transport used (e.g., a ton of freight moved by rail is likely to have a different impact than a ton moved over the same distance by truck) and the intensity of that use (e.g., four people making a trip by car leave a smaller environmental impact than if they all went in their individual vehicles). However, within each mode there are

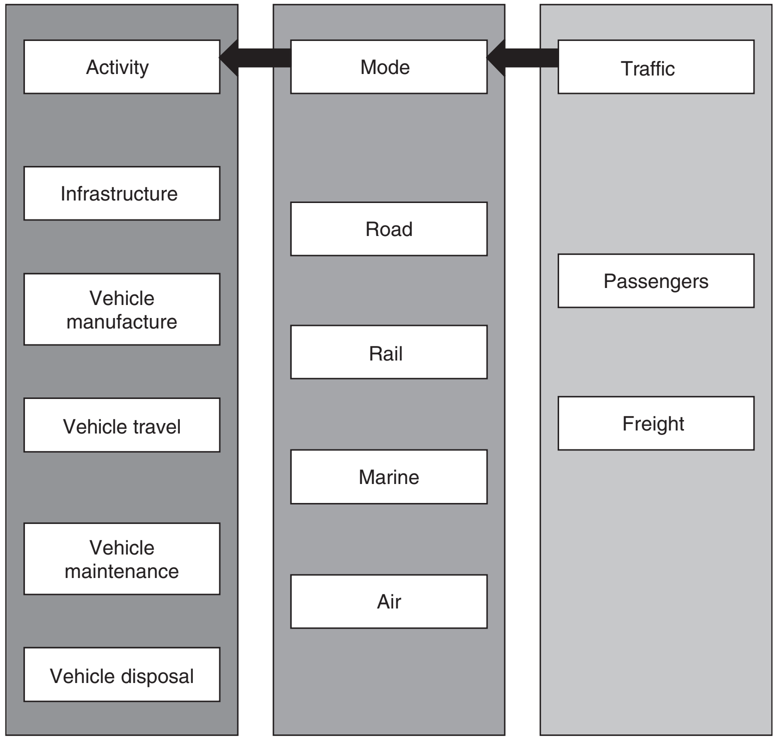

Figure 1 The various ways in which the transport and sustainable interaction can be viewed.

further breakdowns to consider; there is not just the implications associated with the direct movement of people and goods but also the environmental impacts stemming from producing the hardware required, including the infrastructure. From an analytical perspective and policy formulation perspective, all these stages need to be embraced when considering sustainability.

Transport also poses particular environmental challenges somewhat different to many other activities. It is by definition largely a mobile source of pollution; it has been described as "industry on wheels" by Thomson (1974). This can make the measurement of environmental damage difficult and can pose problems when it comes to responsibility for remedial policy formulation. It also relies on a large, immobile infrastructure, it is responsible for a diverse range of environmental effects, and it is a major generator of numerous individual types of pollution such as the global warming gas, and carbon dioxide (CO_2) to which it contributes 14%.

Figure 2 offers a very stylized impression of the implications of some of the most important environmental effects of transport in terms of their geographical spread from source and the duration over which they have an impact on the ecosystem. For example, nitrogen oxide (NO_X) can contribute to acid rain that has both a relatively long term impact on plant life and water quality that extends considerable distances from its source. Other forms of environmental damage such as lead emissions that cause brain damage to

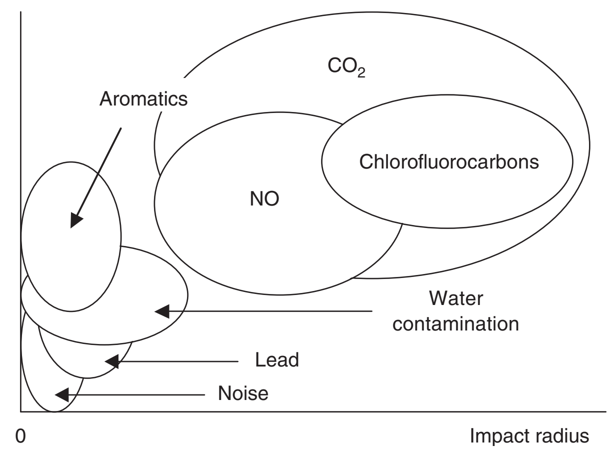

Figure 2 The relative spatial range and temporal duration of transport-associated environmental damage.

children are very local in their impact and are soon washed away.

The impacts imposed on the environment also vary according to mode of transport and across countries. The US' transport system, for example, is estimated to contribute about 37% of global CO_2 emissions stemming from transport use, mainly because of the country's very high level of automobile ownership and the low fuel efficiency of its car park. Road transport is the main direct global contributor to these greenhouse gas emissions, accounting for 76% of transport's contribution, with rail transport contributing another 2%, water

Table 1 Contributions of individual transport modes to overall transport CO_2 emissions and to total CO_2 emissions in the UK

Natural logarithm of million tons of carbon (Ln MtC)	*(Ln MtC)*	*As % of total domestic transport*	*As % of total UK CO_2 emissions*
Road transport			
Passenger cars	32.6	93.1	21.4
Light duty vehicles	19.4	55.5	12.8
Buses	1.0	2.8	0.6
Heavy goods vehicles	7.6	21.6	5.0
Mopeds and motorbikes	0.1	0.3	0.1
Railways	0.7	1.9	0.4
Civil aircraft	0.6	1.9	0.4
Shipping	1.0	2.9	0.7
All domestic transport	35.0	100	23.0
Total UK emissions	152.5		100
International aviation	9.0		5.5
International shipping	1.6		1.0
All transport emissions	45.7		28.0
UK CO_2 emissions	163.1		100

transport 10%, and international air transport 7%. These figures do not include the emissions associated with the electricity generated for railway use and with the pro duction and maintenance of transport vehicles and infrastructure.

Table 1 offers another form of breakdown, and also highlights deviations from the global average. It details the case of the UK. This shows the importance of indi vidual modes, both domestic and international to the overall effects of transport on the environment and also as a proportion of CO_2 emissions in the country. What this type of table omits, however, is the increasing en vironmental burden associated with international trans port. While much of this stemming from surface modes is captured in the domestic data, the increasing use of air and maritime transport outside of national borders is adding to atmospheric and sea pollution. The emissions of CO_2 at high altitude by aircraft, for example, would seem to have a greater impact on global warming than similar amounts emitted at surface levels.

Given that more localized effects of other forms of environmental damage associated with transport are more site specific, it is difficult to both measure the overall scale of the problem and attribute it to specific modes. Ambient levels of noise, for example, differ between locations, making the implications of traffic difficult to isolate. Similarly, the damage done by acid rain chemicals such as NO_X depends very much on re gional geography. While estimates are sometimes made of the total amounts of some of the emissions into the atmosphere or pollutants into the sea, the actual impacts of pollution and the damage it causes depend on local conditions. Nevertheless, they can have important social and economic implications for local communities and thus their longer term sustainability.

The implications of greenhouse gas emissions are largely to be felt in the long term, and thus assessment of their potential magnitudes requires forecasts of traffic. Transport forecasting is, however, notoriously difficult. In general traffic levels are closely correlated with income growth because of the derived nature of the demand for its use. Forecasts based upon existing policies indicate that transport use will double from their 1990 levels by 2050 but that emissions of CO_2 will grow even faster because of the increasing use of synfuels produced from coal and gas, which are twice as polluting as conventional oil. It is predicted that there will be a disproportional increase in the greenhouse gas emissions associated with air transport.

While the global warming effects of CO_2 emissions pose direct environmental effects, sustainable develop ment also involves local environmental and social and economic considerations. Local pollution associated with transport not only poses concerns about the health and welfare of those affected but also raises issues of social equity: it is often argued that serious inequalities in welfare are economically inefficient and politically un sustainable. The US federal government, for example, requires that any transport initiative involving federal funding integrate consideration of environmental justice in the decision process.

Sustainable development also means that economic performance is efficient over the long term, indeed the report of the World Commission spends considerable time on this. Historically, large parts of national transport systems, ranging from the roads built by the Romans, through the mountain footpaths for Inca runners in South America, through to the maritime coaling stations of the British Empire and the transcontinental railroads of Canada and the US, were constructed to sustain pol itical systems and to create social cohesion. In the higher income countries of the world, transport accounts for about 8% of national incomes and serves an important intermediary role in facilitating high quality supply chain inputs to industrial and agricultural production.

Transport also has a particularly important role to play in sustainable economic development in lower income countries; the World Bank devotes some 14% of its an nual lending to transport initiatives. Many of these emergent economies, however, while enjoying higher material standards of living are also generating significant amounts of pollutants and contaminants. China, for example, has 16 of the 20 most polluted cities in the world and is the second largest emitter of CO_2. Although the widespread use of coal in industry is the primary contributor, part of the problem is due to the rapid increase in car ownership and use. In 2000, vehicular CO and NO_X emissions were, for instance, 30 and 3.8 million tons respectively, and vehicles consumed 65.6 million tons of oil. The problem, under current policies, would seem likely to worsen; in the 1980s there were virtually no private cars in China but by 2003 there were 14 million, and it is forecast that by 2015 there will be 150 million.

Policy Issues

Most policies regarding sustainability are sector based – hence the idea of 'sustainable transport' – rather than being holistic. This more localized approach is often because of the difficulties of defining global policies for such things as carbon emissions. One of the problems with the 1997 Kyoto Protocol to the United Nations Framework Convention on Climate Change, for example, has been the difficulty in reaching a consensus across a multitude of nations on a common approach to reducing global warming gas emissions. Further, because in many cases there are considerable local variations in impact and cause, this may quite logically require more bespoke actions. This type of approach, for example, can be seen in the initiatives of the European Union (European Commission, 2001) and at the individual country level (UK Department for Transport, 2006).

There are also issues concerning the way to approach sustainability and at which point in the transport supply chain to act. Policy may be initiated at the root cause of the concern – in terms of CO_2 emissions this may be done through a blanket carbon tax that is set to limit CO_2 emissions irrespective of how carbon is burned. Transport uses can thus be traded off with uses of carbon fuel for other purposes. The UK pursued this type of policy in the 1980s. As a second best option the policy may be aimed at transport use of carbon fuels – for example, specific taxes on fuel used for motoring – or even more indirectly by limiting in absolute terms the overall growth and pattern of passenger and freight transport use. A modal approach, in this context, is geared more toward switching traffic to the least environmentally intrusive modes and often employs subsidies to the cleaner alternative. A wider approach, see again **Figure 1**, is to aim policy at the entire range of activities associated with transport. The latter, often called a 'whole life' approach, explicitly recognizes the wide diversity of ways in which transport affects the environment.

There are generally trade offs between environmental benefits and economic and social effects; to use economic jargon there is seldom a Pareto improvement whereby some parts of society gain but no one loses. In effect there is always some sort of payment involved. The Organization for Economic Cooperation and Development in the 1960s obtained agreement among its members to adopt the 'polluter pays principle'. This does not mean that direct payments are always needed but rather the onus of responsibility lies with the perpetrator. Working out what this payment should be is a challenging task. It is important however, because many transport policies in the past have been 'government failures' in the sense that actions to achieve a particular objective have not taken into account the full costs that they can impose, and in particular not the full environmental costs.

Sustainable development is holistic but physical measures of economic, environmental, and social effects are in diverse units. There has thus been extensive work seeking to put monetary values on the various elements of these three dimensions of sustainable development so that a common unit of measurement would allow trade offs to be made. Valuing these items is far from easy, and this applies to transport as much as it does to other sectors. **Figure 3** provides details of the various stages that have to be gone through before the actual monetary determination stage is reached. Most are complex and

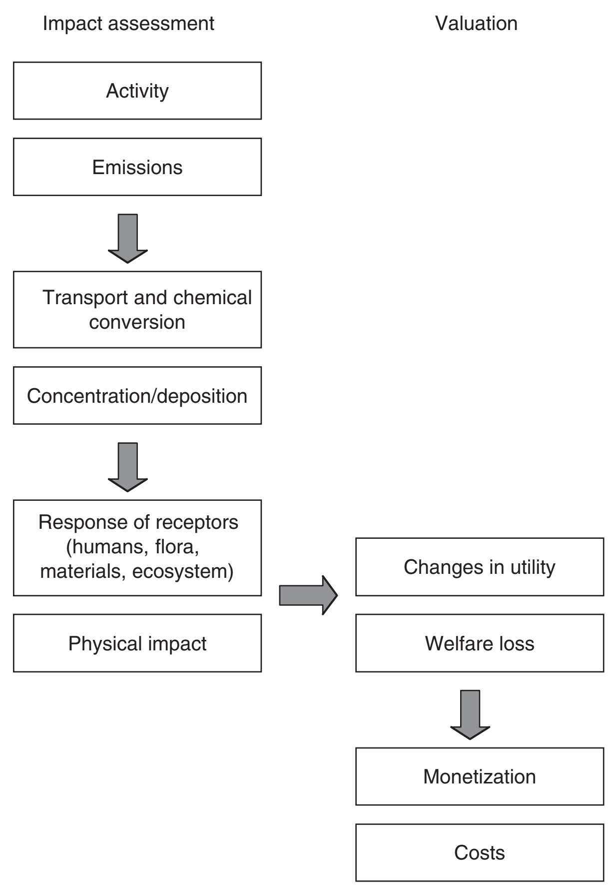

Figure 3 Valuing the environmental impacts of transport.

none without controversy. They combine scientific in formation with economic and social information; inevit ably there is a degree of subjectivity involved.

The measures that have been adopted to make pol luters pay have been varied. The economic efficiency of policy is not always the primary objective when the outcome results in an unfavorable impact on income distribution or certain groups in society. Regulations and standards are common methods to control emissions. These are sometimes applied directly on the emissions (**Figure 4** gives as an example, the European Union controls levels over particulates from diesel fuel and NO_X outputs) but may also be indirect on such things as the fuel efficiency of cars (as with the US Corporate Average Fuel Economy or CAFE standards). These types of policies are found attractive because they are relatively easy to understand and cheap to administer and enforce. In other cases, emissions charges may be imposed that allow transport to be used but only if an environmental charge is paid equal to the costs imposed; for example, many airports charge more for noisy aircraft to land and take off, and countries such as Sweden have both carbon taxes and car owners have to pay a disposal fee when buying their vehicles. Taxes and regulations are often imposed at the local level, rather than as national pol icies, to reflect specific geographical circumstances.

There are also more direct economic instruments that seek to bring environmental considerations directly within markets. They involve allocating property rights to environmental assets and allowing these to be traded. Tradable permits were, for example, used in the US to allow for the efficient removal of lead from gasoline. Essentially, the US government limited the lead available to refineries and gradually reduced this amount. Refiners were given initial quotas in proportion to their prior use, and were allowed to buy and sell among themselves with the aim that the least efficient refineries would be forced from the market as their costs rose leading to production being focused in the most efficient and less environ mentally damaging plants. This is very much the idea behind the carbon trading schemes that have been initi ated between some countries since the Kyoto Protocol.

In a number of respects transport policy has been successful, at least in developed countries, in meeting some of the criteria that many feel are relevant for wider sustainability. Many forms of local air pollution have been reduced with the use of unleaded fuel, the reduction in particulate emissions from diesel engines, and the adop tion of catalytic converters to deal with NO_X. A number of waterways and ports have had contaminates removed or significantly reduced. Economic de regulation of trans port in many countries – the removal of price and market entry controls – has resulted in greater efficiency and with this increased outputs of goods and services to help achieve economic sustainability. There has globally been an increase in mobility that has added to social cohesion. However, success in these areas has caused additional emissions of greenhouse gases as traffic has grown, espe cially in developing countries, and carbon fuel use has risen. This has brought forth new assessments of policy that are embracing a more holistic approach to CO_2 emissions with transport as only one element in them.

Conclusions

Interest in sustainability came to the fore in the late 1980s when scientific evidence really began to build up

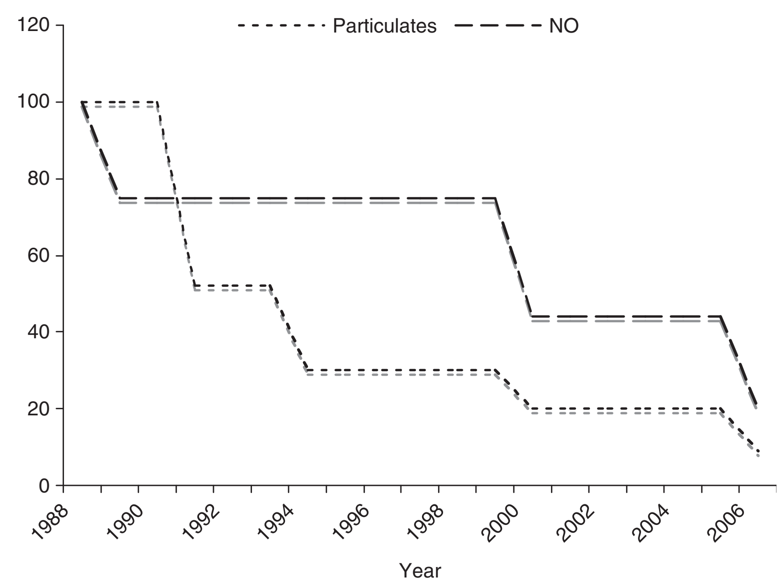

Figure 4 European Union emissions standards for cars.

regarding global warming and the contribution that transport was making to greenhouse gas emissions. Prior to that the environmental damage done by transport was certainly a concern, and most countries had, to varying degrees, initiated policies to limit noise nuisances and local pollution problems such as smog and lead. Acid rain emissions, that can have wider geographical effects, were also reduced as catalytic converters became compulsory on automobiles. The notion of sustainability had ori ginally been thought of in the context of sustainable economic development but broadened out as it became appreciated that excessive environmental damage was itself unsustainable.

Transport is a major contributor to environmentally damaging pollution but also serves important roles in the economy and in ensuring social and political cohesion. Inevitably trade offs have to be made if the transport is to fit appropriately within a wider sustainable structure. To date, however, there has been a tendency to treat sectors in isolation from each other with the result that policy is often aimed at seeking 'sustainable transport' rather than treating transport as part of a more holistic system. This has led to some major improvements in many local en vironments, albeit mainly in developed countries, with less NO_X, lead, and particulate emissions but with more limited success at tackling global challenges such as greenhouse gas emissions. The rapid economic growth of China, India, and other physically large economies, and the concomitant growth in their vehicle fleets is posing particular burdens on efforts to reduce the environmental footprint of transport.

See also: Climate Change; Transport Geography.

Further Reading

Banister, D. and Button, K. J. (eds.) (1993). *Transport, the Environment and Sustainable Development*. London: E & F.N. Spon.

Black, W. R. and Nijkamp, P. (2002). *Social Change and Sustainable Transport*. Bloomington, IN: Indiana University Press.

Boulding, K. (1966). The economics of the coming spaceship Earth, presented at the Sixth Resources for the Future Forum on Environmental Quality in a Growing Economy in Washington, DC.

Commission on Environment and Development (1987). *Our Common Future*. Oxford: Oxford University Press.

European Commission (2001). *European Transport Policy for 2010: Time to Decide*. Luxemburg: European Commission.

European Commission (2005). *Doing More with Less: Green Paper on Energy Efficiency*. Luxemburg: European Commission.

Hahn, R. W. and Hester, G. L. (1989). Marketable permits: Lessons for theory and practice. *Ecological Law Quarterly* 16, 380 391.

Organisation for Economic Cooperation and Development (1992). *Market and Government Failures in Environmental Policy: The Case of Transport*. Paris: OECD.

Schweitzer, I. and Valenzuela, A. (2004). Environmental injustice and transportation: The claims and the evidence. *Journal of Planning Literature* 18, 383 398.

Swedish Environmental Protection Agency (1997). *Environmental Taxes in Sweden Economic Instruments of Environmental Policy*, Report 4745. Stockholm.

Thomson, J. M. (1974). *Modern Transport Economics*. Harmondsworth: Penguin.

UK Department for Transport (2006). *The Eddington Transport Study: Transport's Role in Sustaining the UK's Productivity and Competitiveness*. London: The Stationary Office.

UK Treasury (2006). *The Economics of Climate Change*. London: UK Treasury.

World Business Council for Sustainable Development (2004). *Mobility 2030: Meeting the Challenges to Sustainability*. Geneva: WBCSD.

Relevant Websites

http://www.euractiv.com/en/transport/transport infrastructure environment/article 168342
European Union Information Website.

http://www.ipcc.ch/
Inter Governmental Panel of Climate Change.

http://www.euro.who.int/transport/hia/20021009 2
World Health Organization.

Transport Geography

R. D. Knowles, University of Salford, Greater Manchester, UK

Glossary

Accessibility The ease with which activity sites can be reached for goods, services, employment, education, and recreation. Transport is the main but not the only factor; the Internet provides instant, space-free access to certain services.
Death of Distance The misconception that as telecommunications have annihilated space, they have made the physical transport of people unnecessary.
Deregulation The abolition of government regulations controlling, for example, the issuing of transport route licenses and transport fares.
Differential Collapse in Time–Space The process of time–space convergence is geographically uneven as investment in transport innovations varies and creates a misshapen world in terms of accessibility.
Gender Gap A difference in behavior or performance attribute between men and women.
Intermediacy The spatial attribute of locations *en route* between important origins and destinations and geographically well located in relation to traffic flows.
Intermodal Transport Where goods, containers, or vehicles can transfer quickly and cheaply between different modes of transport enabling the use of two or more modes of transport. An example is a landbridge route where a container can be transferred from ship to rail and back to ship between Western Europe and China or Japan via a North American rail route and compete with a longer shipping route via the Panama Canal.
Mobility The ability of individuals or groups of people to move around.
New Mobilities Paradigm A major recent upturn in the significance of the study of mobility in social sciences to examine how, why, and where people move.
Peak Oil When global oil production peaks and becomes more expensive to use and alternative types of fuel become necessary.
Social Exclusion Where people have lost the ability to connect with jobs, services, and facilities that they need to participate fully in society.
Telecommuters People whose jobs allow them to work from home using computers with Internet access.
Time–Space Convergence The idea that as a result of faster transport places move closer together in terms of travel time.

Introduction

Transport has been described by David Keeling as 'quintessentially geographic'. This is because it plays a crucial role in defining the function and character of spaces and places while facilitating the flows and interactions within and between them. Transport and geography share a mutual interrelationship; while transport is central to the study of geography, geography is also central to the study of transport. Transport is an enabling technology; it makes possible a variety of geographies of human interaction. Specialization in production and the division of labor, accessing raw materials and energy supplies, distribution and trade, personal mobility, and accessibility to employment and other activity sites all require effective transport systems.

Transport geography examines spatial aspects and impacts of transport systems at local, national, and global scales. It encompasses the development, location, and structure of transport networks and analysis of their use by people and freight; impacts of transport systems and policies on the economy, society, and environment; and issues of personal mobility, accessibility, and social exclusion. In the interdisciplinary study of transport, transport geography complements the contributions of economists, civil engineers, logistics and supply chain managers, planners, psychologists, and sociologists and adds important spatial perspectives.

The Development of Transport Geography

Transport geography emerged as a leading subdiscipline in the 1960s and 1970s when it was at the forefront of the quantitative–theoretical revolution in human geography. Its largely positivist approach and a focus on spatial organization and interaction and network analysis led to a lasting interest in modeling and analytical–empirical research which was boosted later by the development of geographical information systems. However, this narrow emphasis on mechanistic models and network analysis resulted in transport geography becoming rather detached from socioeconomic and political realities and therefore somewhat isolated from the rest of human geography. At the same time many human geographers saw cheap transport and personal mobility as a norm and largely discounted transport in their examination of social and economic patterns and systems.

Peter Rimmer was in the vanguard in calling for a more humanistic, nonpositivist transport geography in the late 1970s but progress was slow. More transport geographers began to engage in the analysis of transport policy, socioeconomic aspects of decision making, and transport's impacts on people. Behavioral approaches to transport required more focus on individual and house hold activity patterns and decision making. Analysis of transport welfare problems led to more exploration of accessibility, mobility and gender issues, and transport needs but this was sidelined for more than a decade by the market based transport policies of the 1980s. Wide spread deregulation and privatization of transport how ever has provided a wealth of opportunities for research into the organization of transport systems and networks and the spatial and economic outcomes of policy change. At a time when urban, political, and economic geog raphers were exploring Marxist social theory, transport geographers were engaging more with other transport disciplines.

A renaissance in transport geography in the 1990s saw the development of new methodologies and theories stimulated in particular by globalization, sustainability issues, and transport–societal relationships and enabled by spatial analysis technologies. The *Journal of Transport Geography* provided an international focus for research in the subdiscipline from 1993 accompanied by an array of new transport geography texts. Engagement with eco nomic geography started to revive with its emphases on 'new economic geography' and 'new trade theory' and transport geography's renewed interest in freight trans port systems and freight flows. Interdisciplinary links with logistics and supply chain management are of growing importance. The current major opportunity for transport geography to re engage more fully with social and cultural geography and with other social sciences lies in the common interest in mobilities. David Keeling re cently highlighted the importance for transport geog raphers of theorizing and conceptualizing relationships between people and places focused on accessibility and mobility as the core of human interaction.

Differential Collapse in Time–Space and the Increased Importance of Location

Transport helps to shape patterns of development and accessibility and presents future opportunities. Innov ations that provide faster and cheaper transport and mechanized handling technologies have reduced the barrier to interaction that distance presents. This has led many geographers to ignore the importance of transport and take it for granted. However, the reductions in time–space and cost–space are very unequal, and relative dis parities have widened between core and peripheral areas at global and national scales. The differential collapse in time–space is a product of unequal investment in trans port within as well as between areas. Access to transport also varies enormously and is limited by personal income, age, mobility impairment, and gender as well as by lo cation. Richard Knowles showed that this differential collapse has enhanced the importance of location by creating a misshapen world in terms of accessibility.

Time–space convergence is the term coined by Don Janelle to describe the effect of a succession of innov ations in faster transport from steam railways, steamships, and motorized transport to high speed trains and pro peller, jet, and supersonic aircraft. More than 150 years of time–space convergence have been followed in the last 30 years by a plateau in the general speed of transport modes although local convergence can still occur, for example, with new bridges and tunnels, intermodal ser vices, and high speed trains. William Black described the withdrawal of Concorde's supersonic jet services as a rare example of time–space divergence. On the contrary there is now increasing incidence of time–space divergence caused by traffic congestion, airline security checks, reduced speed limits, and the withdrawal of a range of transport services particularly to peripheral lo cations and smaller settlements. Time–space divergence presents a range of important contemporary geographical research issues which have been largely ignored apart from work identifying the local pain of transport service withdrawal, for example, by Andrew Goetz on US air transport and Robert McCalla on global intermodal seaport terminals.

Telecommunications and the 'Death of Distance'

Telecommunications have transformed global com munication since the 1960s and particularly in the last 10 years. Global positioning systems have revolutionized navigation and enabled operators of all transport modes to develop real time information systems. Mobile (cell) phones and laptop computers using wireless internet have removed the restriction of fixed locations for per sonal communication. They have changed the perception of travel time from wasted to useful time and transformed travel space into working or leisure space. However, Frances Cairncross' notion of the 'death of distance' is not correct. Electronic communication rarely provides a total replacement for the physical movement of people or goods. Most people still prefer to meet in person than over a video link. Current research shows that only a small minority of jobs are likely to be undertaken fully from home. Although there are now more tele commuters, most still sometimes travel to a workplace and also tend to live further away than daily commuters.

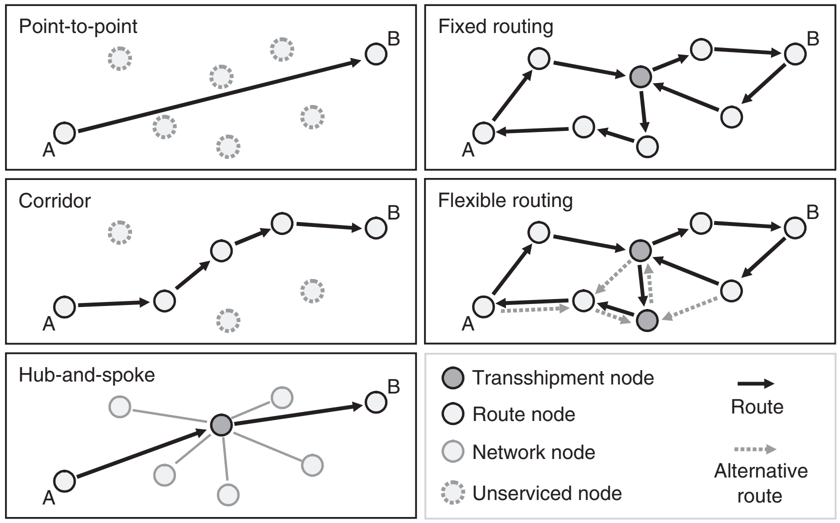

Figure 1 Freight distribution and network strategies. From Hesse, M. and Rodrigue, J. P. (2004). *Journal of Transport Geography* 12, 179.

Internet shopping still accounts for a small minority of transactions and itself requires transport for home de livery. Transport geographers are well placed to identify and explain why activity patterns and locations are changing under the influence of these new technologies.

Transport Innovation and Globalization

Probably the most significant transport revolution in the last half century has been in sea transport. Major in novations in shipbuilding technology have allowed much larger bulk carriers and container ships to be built which together with mechanized port handling facilities have sharply reduced maritime freight transport costs. These shipping innovations have allowed raw materials to be sourced globally with coastal areas benefiting more than interior locations. Globalized industrial production takes advantage of lower wage rates in less developed and newly industrialized countries and is enabled by cheap containerized transport and lower or zero import tariffs. Economies and societies are becoming more enmeshed and interdependent as the world shrinks with time–space and cost–space convergence.

The development from the 1950s of quick and low cost mechanized handling of unitized freight in standard containers removed the financial and time barrier created by traditional methods of manual handling of freight in ports. This enabled intermodal transport routes and in land loadcentres to be developed and has allowed shorter rail/ship landbridge and minibridge routes, particularly in North America, to compete with all water routes. How ever, many bulk carriers are now too large to use the Panama and Suez Canals. Since 1988, the largest, post Panamax, container ships have also been too large for the Panama Canal, although the latter is now being enlarged. Jean Paul Rodrigue has called for transport geographers to focus much more research on global and national freight transport systems which underpin so much eco nomic development. There are vast opportunities for research into maritime and land based freight flows, route networks, and terminals (**Figure 1**). More geog raphers need to recognize the powerful role of logistics providers and supply chain managers in choosing routes, terminals, and freight transport operators and understand the importance of their decision making processes (**Figure 2**).

The location and turnover of the largest container ports provides a good measure of the extent and spread of globalization (**Table 1** and **Figure 3**). The world's biggest ports until recently were located mainly in developed countries. In 1975 New York and Rotterdam were the largest ports while seven of the top 20 container ports were in the USA, five in both Europe and Southeast/East Asia and two in Australia, handling a total of 10.6 million TEU (twenty foot equivalent unit). By 2003, the 20 largest container ports handled 144.8 million TEU, a more than twelvefold increase; Hong Kong and Singa pore were now the largest ports and the rapid growth of China's manufacturing exports showed with three of the world's top four ports. Overall 11 of the top 20 ports were located in Southeast/East Asia, five in Europe, three in the USA and one in the Middle East. Hub and feeder transit container ports such as Singapore and Gioia Tauro in Italy have developed on major global trading routes.

Transport and Economic Development

Transport investment usually facilitates economic de velopment but the relationship is complex and two way

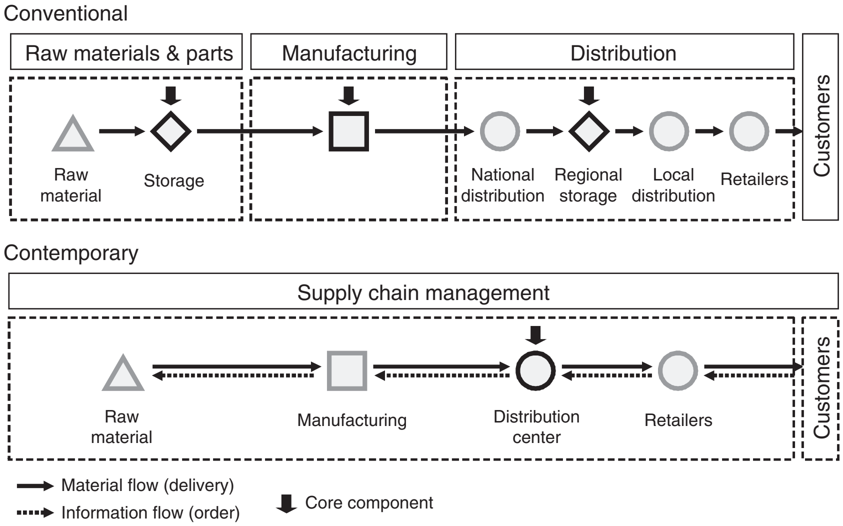

Figure 2 Conventional and contemporary arrangement of freight flows. From Hesse, M. and Rodrigue, J. P. (2004). *Journal of Transport Geography* 12, 176.

Table 1 Container traffic ('000 TEU) at the world's 20 largest ports

Rank	*Ports*	*Container traffic (1975)*	*Ports*	*Container traffic (2003)*
1	New York	1730	Hong Kong	20 449
2	Rotterdam	1078	Singapore	18 100
3	Kobe	904	Shanghai	11 280
4	San Juan	877	Shenzen	10 615
5	Hong Kong	802	Busan	10 408
6	Oakland	522	Kaohsiung	8 840
7	Seattle	481	Los Angeles	7179
8	Baltimore	421	Rotterdam	7107
9	Bremen	409	Hamburg	6138
10	Long Beach	390	Antwerp	5445
11	Tokyo	368	Dubai	5152
12	Melbourne	364	Port Klang	4840
13	Yokohama	328	Long Beach	4658
14	Hamburg	326	Qingdao	4239
15	Antwerp	297	New York/New Jersey	4068
16	Hampton Roads	292	Tanjung Pelepas	3487
17	Sydney	262	Tokyo	3314
18	London	260	Bremerhaven	3190
19	Keelung	246	Laem Chabang	3181
20	Le Havre	231	Gioia Tauro	3149

TEU = 20 foot equivalent unit; NJ = New Jersey; Brhvn = Bremerhaven.
Source of data: Containerisation International.
From Knowles, R. D. (2006). Transport shaping space: Differential collapse in time space. *Journal of Transport Geography* 14, 407 425.

and there are sometimes unintended negative effects. Impacts of transport investment are spatially uneven and favour some sections of the population. Additional transport investment in already developed countries can yield fewer benefits and tends to generate additional traffic. David Banister and Joseph Berechman identified the need for positive economic externalities, good in vestment factors, and a favourable political environment to be in place for transport investment to assist regional economic growth.

More research is needed to measure the role of transport investment in economic development whether in regenerating former industrial towns, particularly derelict industrial or dockland sites, or stimulating the growth of peripheral areas. Isolating the role of transport investment is easier where new bridges and tunnels are built across straits and tidal estuaries and tunnels are cut through mountains; these fixed links create measurable time–space convergence and reconfigurate local eco nomic and social space.

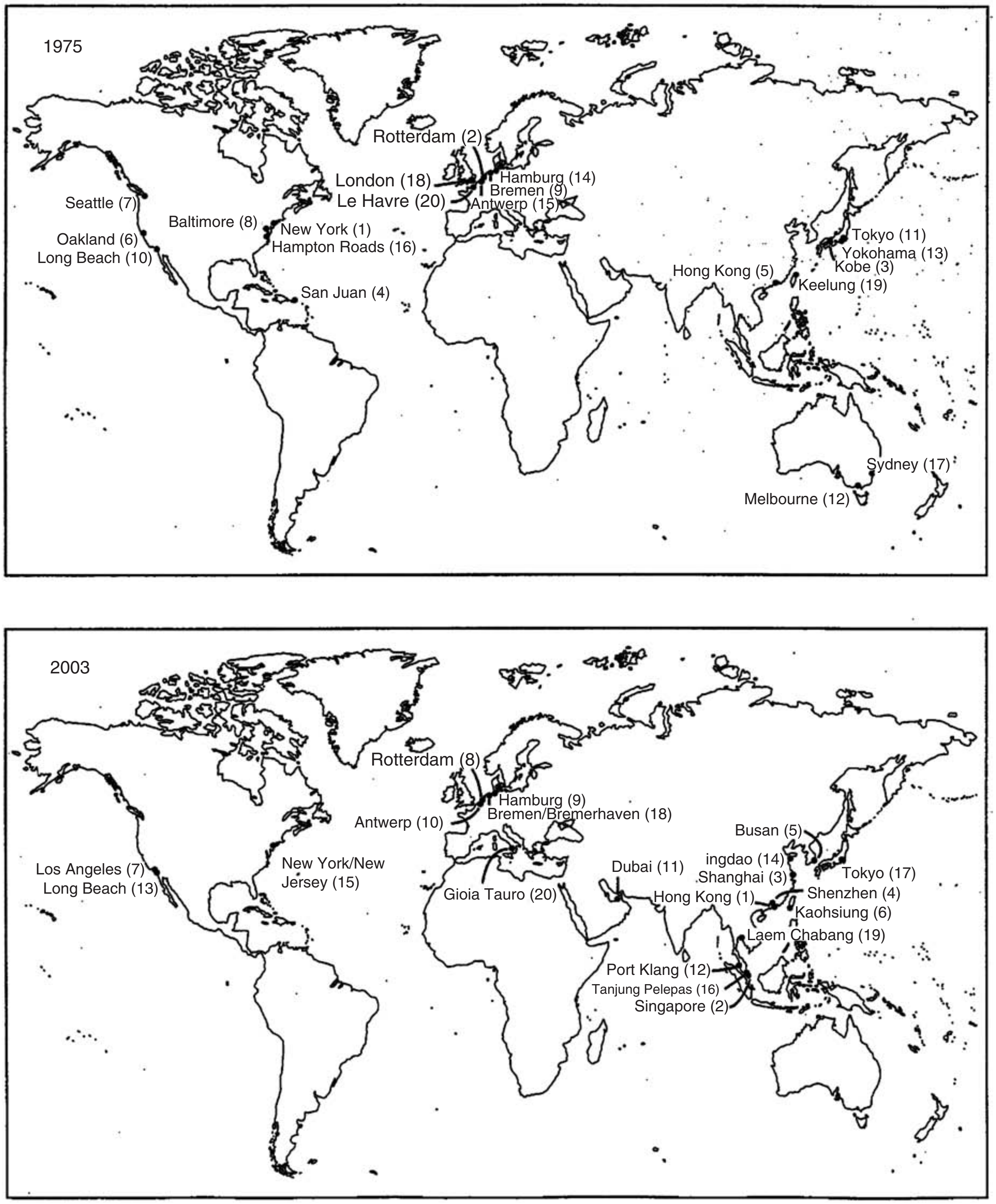

Figure 3 World's largest 20 container ports 1975 and 2003. From Knowles, R. D. (2006). *Journal of Transport Geography* 14, 414.

Centrality, Proximity, and Intermediacy

Centrality within the central business district (CBD) of large urban areas is longer than the undisputed optimal locational goal for activity sites and this is reflected in polycentric, rather than monocentric, peaks in urban land values. Locations that are in proximity to large markets can offer quicker and cheaper access in congested urban areas than central locations, hence the extensive use of motorway and ring road junctions for activity sites and of secondary airports by low cost airlines in deregu lated airline markets. However, proximity can also be a negative attribute if a place is too close to and falls under the economic or traffic shadow of a larger node. Douglas Fleming and Yahuda Hayuth invented the term 'intermediacy' to describe the spatial property of lo cations between important origins and destinations; these include hubs, route junctions, break in bulk points, and gateways. The geographical importance of intermediate locations, situated close to principal transport flows be tween major markets, has grown for in transit inter change of traffic in deregulated transport markets. Intermediacy is also important in determining advan tageous locations for intermodal transfers of traffic and for refueling stops.

Although intermediacy is an important geographical concept it is little recognized in academic research out side transport geography. Globalization has led to a rapid re evaluation of intermediate locations by entrepreneurs especially in the transport, logistics, and supply chain sectors. This is an area of research potential for trans port geographers, economic geographers, and business communities.

Transport Deregulation, Privatization, and Network Adaptation

Widespread liberalization or deregulation of transport services from government control, through gradual or outright abolition of route licensing and controls on fares or freight rates, can change the shape of transport net works and often reduce the cost of transport. The theory of contestable markets assumes that there will be a competitive transport market among privately owned transport companies whereas many transport markets are dominated by monopolies or near monopolies. Deregu lated transport is market orientated, follows traffic demand flows, and tends to enhance metropolitan advan tage. Some markets, such as European air transport and British express and local bus transport, have been dis torted by the presence of state or municipally owned companies which potentially have access to subsidies.

Deregulated air transport has become a very active research field for transport geographers initially in the USA after domestic air transport deregulation in 1978, later in Europe with air transport liberalization, and more recently elsewhere. Human geographers studying world urban systems also use airport passenger totals and flows as functional measures of importance. New airlines, mergers and bankruptcies, and organiza tional changes in the deregulated airline industry and good data availability have enabled transport geog raphers to research extensively into hub structures, air ports, changes in route networks and fares, and into competition between 'legacy' airlines and low cost car riers. In the USA, deregulation of the domestic passenger airline market led to the development of hub and spoke networks by large incumbent 'legacy' airlines. Hubs in large cities are linked with smaller cities by regular feeder and distributor flights and are connected with other hubs by frequent direct flights. Hub airports tend to be located *en route* along main traffic flows in intermediate locations rather than in larger but more peripherally located cities. Passengers usually benefit from more flights and cheaper fares but at the expense of longer, more circuitous journeys if via a hub. Low cost competitors to legacy airlines within both the USA and Europe, following a model pioneered by South West Airlines, charge lower fares by offering 'no frills' services, usually on point to point routes with no timetabled connections. They also often reduce airport charges and turn round times by using smaller, secondary airports close to urban markets.

Andrew Goetz showed that there have been some negative outcomes for passengers from US air transport deregulation; he identified 'pockets of pain' where communities lost air services or where legacy airline dominance resulted in higher fares. European air trans port deregulation since 1997 has also allowed low cost competitors, led by Ryanair and easyJet, to cut fares, compete with legacy airlines, and generate new traffic (**Figure 4**). However, shorter distances between cities, high speed rail competition, continuing state ownership of some legacy airlines, and allegiance of some passengers to nationally based airlines have created a less even pattern of change than in the USA (**Figure 5**).

These dynamic processes of change spreading from the USA and Europe to other airline markets provide major research opportunities in a field which is well connected with the expansion of leisure travel and the wider 'new mobilities' agenda.

The perceived success of air transport deregulation has not been replicated in all transport markets. De regulation of local bus services in Great Britain outside London, followed by piecemeal privatization of state owned and most municipal bus companies, resulted in very little competition, much higher fares, a collapse of integrated services and fares, and a continuing reduction in passengers in all regions and most towns and cities. Geographical analysis shows a market oriented focus on heavily trafficked, profitable urban corridors, enhanced in many areas by bus quality partnerships between private bus companies investing in better services and local au thorities giving buses priority on the roads. In contrast, there has been reduced focus on peripheral and rural areas and on evening and weekend services. In an in creasing mobile society based largely on private car use, the substantial minority of people reliant on buses to access increasingly distant activity sites have suffered a substantial reduction in both personal mobility and accessibility.

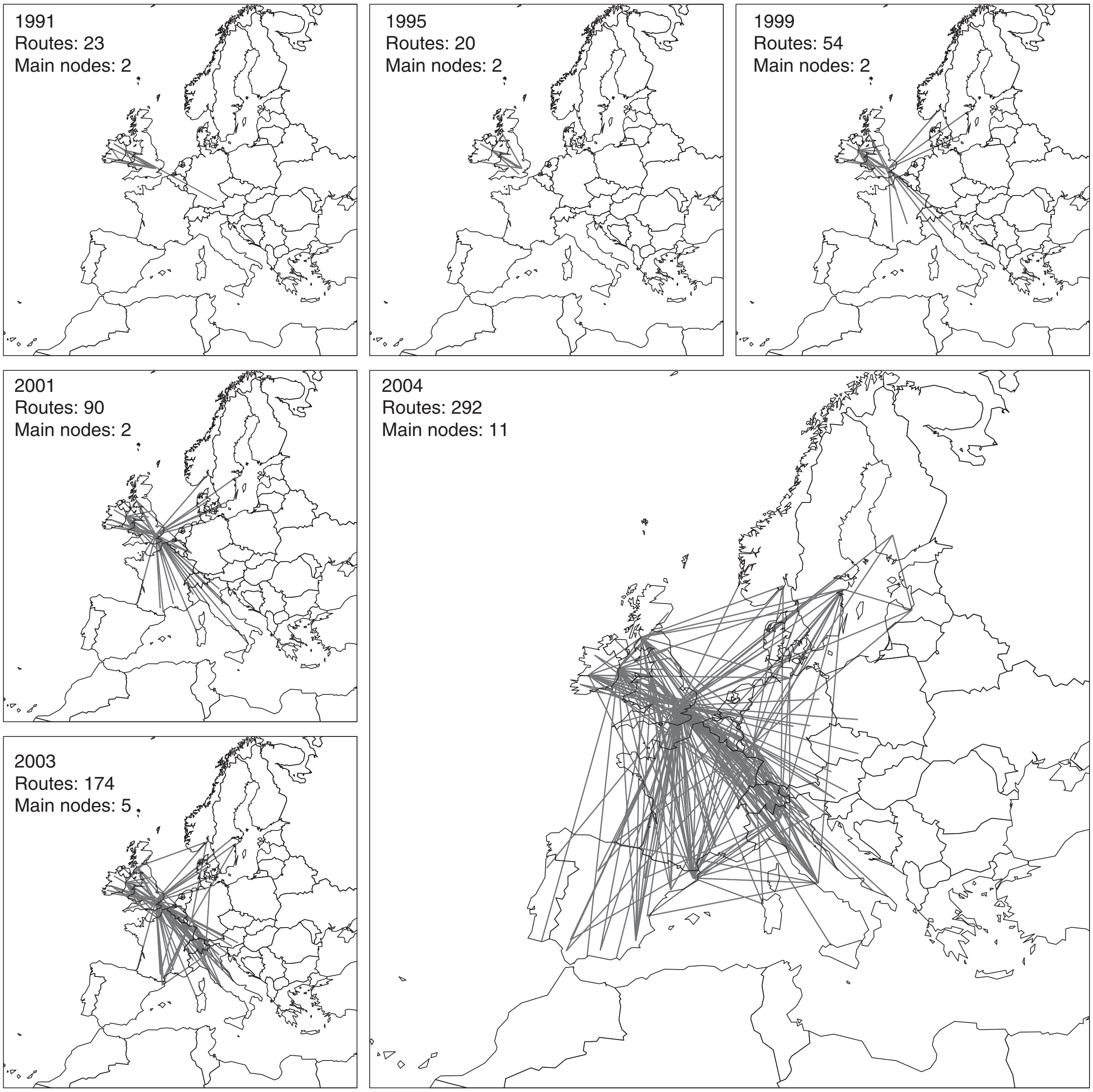

Figure 4 The growth of the Ryanair network 1991 2004. From Dobruszkes, F. (2006). *Journal of Transport Geography* 14, 260.

Urban Expansion, Transport Technology, and Planning

Urban expansion and transport technology have been closely interconnected since the development of mech anized transport. From the compact and circular form of the pre industrial city based on walking short distances, a pronounced star shaped urban form developed in re sponse to development opportunities aligned to tram and commuter rail routes but focused strongly on one CBD. Mass car ownership, cheap oil, and cheap land sub sequently enabled and encouraged low density urban expansion and infilling between radial routes unless they were constrained by planning restrictions. In a few European countries, the public transport accessible city was deliberately strengthened in the mid twentieth century by land use and transport development plans that limited urban sprawl. Copenhagen's 1947 Finger Plan is a well known example with new planned high density suburbs connected to its CBD along five electrified commuter rail lines (**Figure 6**).

Limited access radial and orbital highways have fur ther facilitated urban expansion and a refocusing of jobs, shops, and leisure facilities at activity sites remote from the CBD. In most American cities, CBDs have declined as jobs and higher order retailing moved out to 'suburban

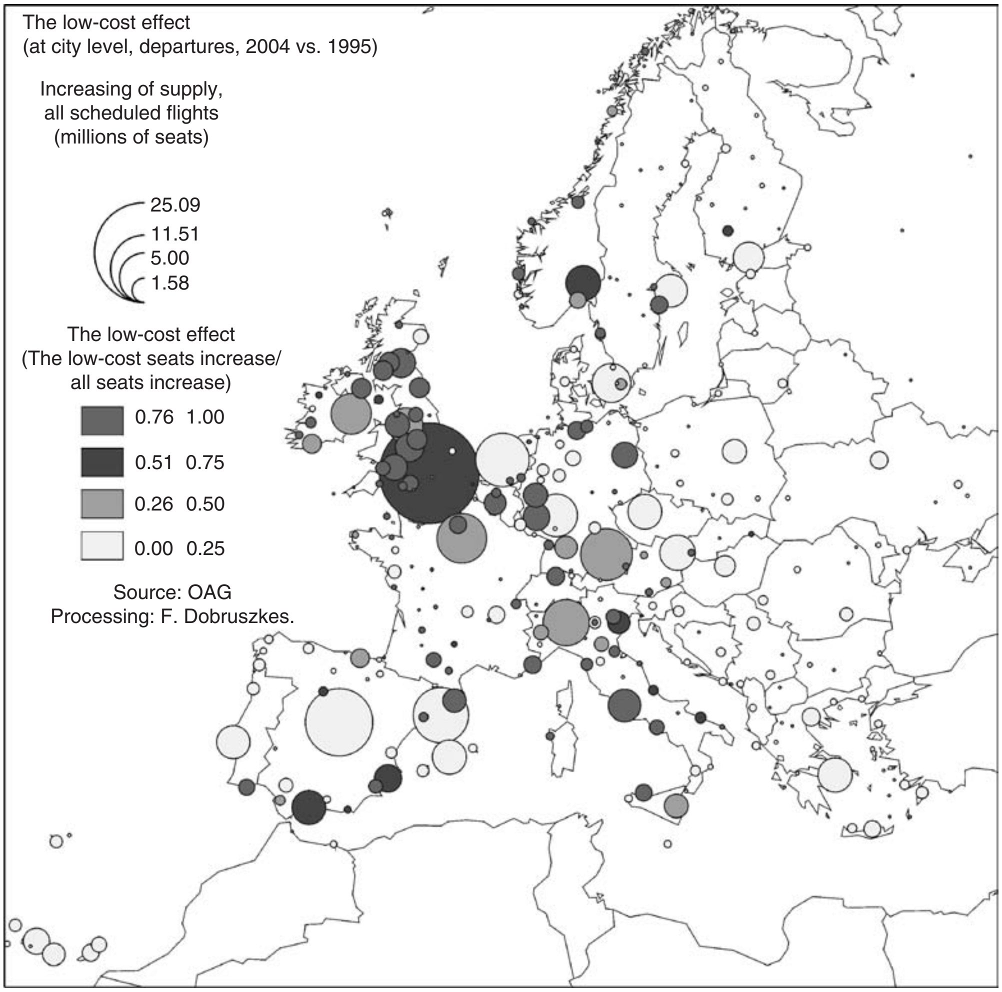

Figure 5 The low-cost effect (city level departures 2004 versus 1995). From Dobruszkes, F. (2006). *Journal of Transport Geography* 14, 262.

downtowns' or 'edge cities' which are difficult to serve by public transport. In Europe decentralization is less pronounced as lower car ownership, fewer limited access highways, tougher planning controls, dearer land prices, and better public transport have restricted urban sprawl and allowed multifunctional CBDs to survive.

Transport geographers are well placed to contribute to research into recent processes of repopulating and regenerating cities, further stimulated by the growing debate about sustainable urban areas, global warming, and dependency on quickly depleting fossil fuel resources.

Unsustainable Energy Supplies and the Consequences of Peak Oil

William Black drew attention to the unsustainability of oil dependency for transport systems in particular and the global economy in general. With global demand for oil increasing yearly by over 2% and the likelihood of global 'peak oil' production being reached soon, the price of oil based fuels will at that stage increase sharply and quickly become too expensive to power motor vehicles. Global warming and climate change are linked to the increase of greenhouse gas emissions with private cars, an important contributor. Both issues highlight the urgent need to develop alternative renewable fuel sources with less harmful emissions such as hydrogen, derived from water, and biofuels, which compete for land with food crops, both of which cost much more than the cheap twentieth century oil. Much higher fuel prices would have far reaching consequences for economies and societies predicated on cheap oil and low cost personal mobility and cause time/space divergence. Demand for transport would reduce but the spatial structure of society would take many years to adapt.

Transport geography has the potential to make sub stantial contributions to examining the consequences of the demise of cheap oil and the search for more sus tainable but dearer energy supplies.

Mobility and Accessibility

The significance of transport and travel in geographical inquiry is undiminished. We are in an age of un precedented personal mobility at local, national, and international scales throughout the developed world. The movement of people is now at the core of social science research exemplified by Mimi Sheller and John Urry's 'new mobilities' paradigm. The increasing significance of mobilities and transport flows and spaces provides a golden opportunity for transport geography to reposition itself back into the core of human geography.

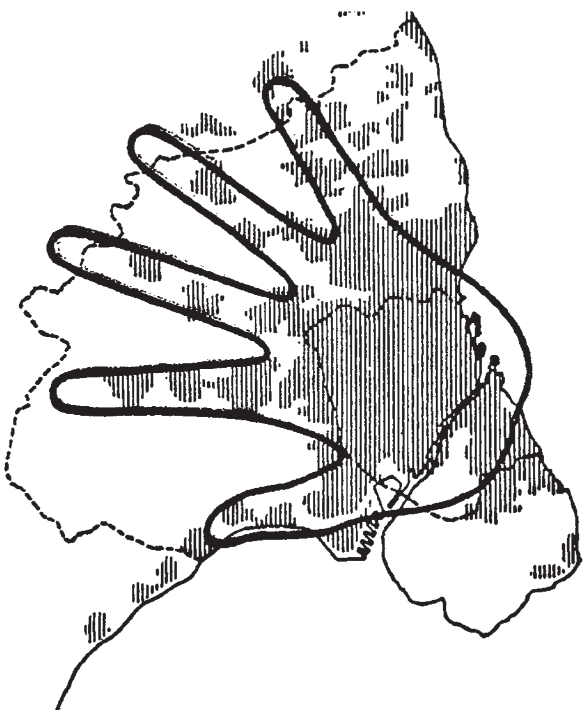

Figure 6 Copenhagen's Finger Plan 1947. From Knowles, R. D. (2006). *Journal of Transport Geography* 14, 418.

Mass car ownership has transformed personal mobil ity and permitted and encouraged the dispersal of ac tivity sites for employment, retailing, warehousing, and leisure to car accessible locations especially near to junctions on high capacity roads. Average yearly distance traveled has increased sharply in developed countries with leisure travel exceeding work journeys and domin ated by car and air travel. However, people without ac cess to cars suffer increasing social exclusion and reduced accessibility as more activity sites move beyond walking distance. This is compounded by a well above inflation increase in bus and train fares and most people who are dependent on public transport have low incomes.

Car ownership and availability is far from universal and varies with geographical location as well as by in come, age, gender, and disability. The need to own a car increases in rural and smaller urban areas where public transport services are more limited or nonexistent (**Table 2**). This is reflected in much higher proportions of households with regular use of cars in rural areas and smaller towns, in all income groups and accounting for a much higher share of household income. The ability to drive also varies by age and gender. There are still sig nificant albeit declining differences by age, especially for those over 70 years old. The gender gap between males and females is particularly pronounced in the age groups over 50 years old (**Table 3**). In an increasingly mobile world, access or lack of access to cars provides the main distinction between people who are highly mobile and enjoy good accessibility and those who are increasingly isolated. Locational differences add to a substantial re search agenda which is at the heart of human geography and social science inquiry.

The failure of the 'predict and provide' policy, of trying to build new road space to accommodate the growing volume of car traffic, led to Phil Goodwin's espousal of a 'new realism' in transport policy which prioritizes demand management and improved public transport. In Great Britain, local authorities have been given a new re sponsibility to undertake accessibility planning of medical,

Table 2 Households with regular use of cars by type of area in Great Britain 2005

Type of area	*Cars per household*	*No car (%)*	*One car (%)*	*Two or more cars (%)*
Greater London	0.83	39	43	18
Provincial conurbations	0.99	32	41	27
Other urban areas:				
Over 250k	1.14	23	45	32
25 250k	1.13	25	43	31
10 25k	1.13	23	47	30
3 10k	1.24	20	43	37
Rural areas	1.59	11	37	52
Great Britain	1.15	25	43	32

Source of data: Transport Statistics 2006. London, Department for Transport.

Table 3 Full car driving licence holders by age and gender in the UK 1975/6 2003

Age Group	*Male (%)*		*Female (%)*		*Gender gap (M F) (%)*	
	1975/6	*2003*	*1975/6*	*2003*	*1975/6*	*2003*
17 20	36	31	20	24	16	7
21 29	78	74	43	62	35	12
30 39	85	88	48	77	37	11
40 49	83	91	37	78	46	13
50 59	75	90	24	69	51	21
60 69	58	87	15	59	43	28
70+	32	69	4	27	28	42
All adults	69	81	29	61	40	20

Source of raw data: Transport Statistics 2004. London, Department for Transport.

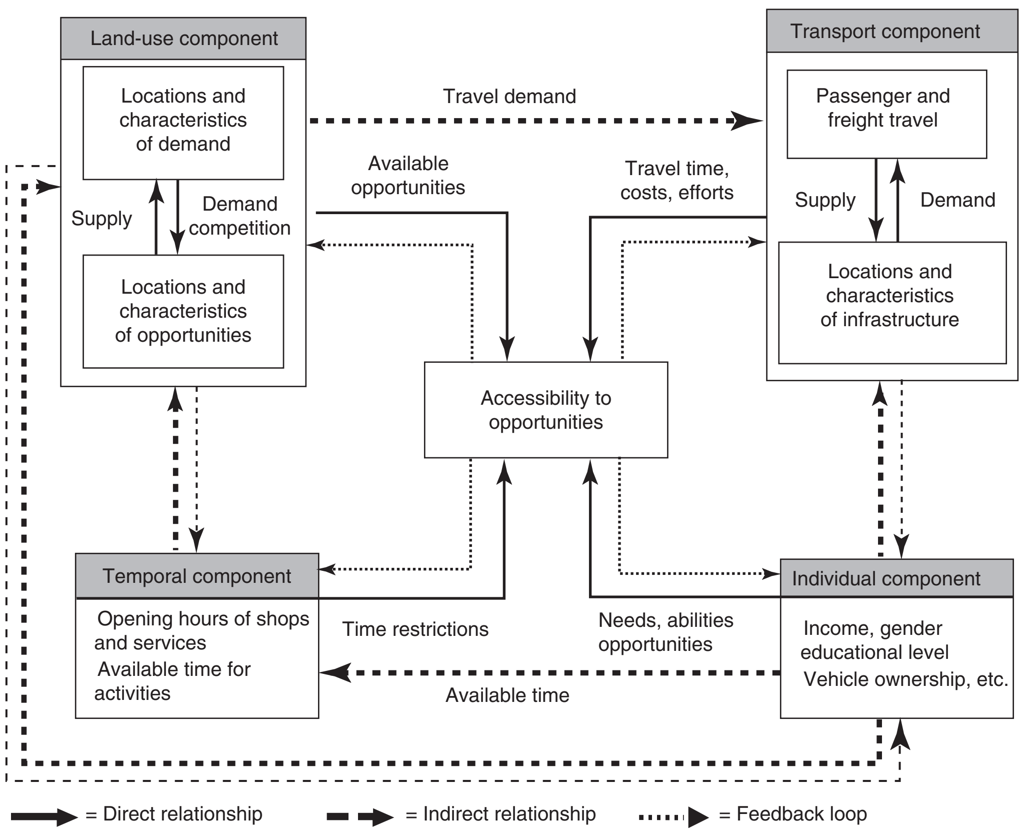

Figure 7 Relationships between components of accessibility. From Geurs, K. T. and van Wee, B. (2004). *Journal of Transport Geography* 12, 129.

retailing, and educational facilities. This has helped to revive interest among human geographers and other social scientists in the concept of accessibility as a means of tackling mobility deprivation (**Figure 7**).

Conclusion

This chapter has traced the development of transport geography as a subdiscipline of human geography and explored its changing relationships with other transport disciplines and social sciences. An earlier focus on positivist and quantitative methodologies has shifted to include nonpositivist and qualitative methodologies. The differential collapse in time–space and cost–space has enhanced the importance of location contrary to a widely held view in human geography that transport has become relatively unimportant. Telecommunications have trans formed communication but have not eliminated the im portance of distance. Transport innovations, especially in freight shipping, continue to play a central role in glob alization processes. Relationships between transport and economic development are important but complex and difficult to measure. As well as centrality, locational qualities of proximity and intermediacy have become increasingly important in more flexible and deregulated transport markets. Urban expansion continues to be enabled by transport innovations but constrained by planning restrictions and more expensive fuel. The sustainability of future transport fuel supply will become even more important after peak oil production has been reached. In an increasingly mobile world, personal mobility and accessibility are central to transport geography inquiry and should reposition the subdiscipline at the core of human geography.

See also: Human Geography; Mobility; Transport and Accessibility; Transport and Deregulation; Transport and Globalization; Transport and Social Exclusion.

Further Reading

Black, W. R. (2003). *Transportation: A Geographical Analysis*. New York: Guilford Press.

Fleming, D. K. and Hayuth, Y. (1994). Spatial characteristics of transportation hubs: Centrality and intermediacy. *Journal of Transport Geography* 2, 3 18.

Goetz, A. R. (2006). Transport geography: Reflecting on a subdiscipline and identifying future research trajectories. The insularity issue in transport geography. *Journal of Transport Geography* 14, 230 231.

Goetz, A. R., Ralston, B. A., Stutz, F. P. and Leinbach, T. R. (2004). Transportation geography. In Gaile, G. L. & Wilmott, C. J. (eds.) *Geography in America at the Dawn of the 21st Century*. Oxford: Oxford University Press.

Hoyle, B. S. and Knowles, R. D. (eds.) (1998). *Modern Transport Geography* (2nd edn.). Chichester: Wiley.

Keeling, D. J. (2007). Transportation geography: New directions on well worn trails. *Progress in Human Geography* 31(2), 217 225.

Knowles, R. D. (2006). Transport shaping space: Differential collapse in time space. *Journal of Transport Geography* 14, 407 425.

Knowles, R. D., Shaw, J. and Docherty, I. (eds.) (2008). *Transport Geographies: Mobilities, Flows and Spaces*. Oxford: Blackwell.

Preston, J. and Raje, F. (2007). Accessibility, mobility and transport related social exclusion. *Journal of Transport Geography* 15, 151 160.

Rodrigue, J. P., Comtois, C. and Slack, B. (2006). *The Geography of Transport Systems*. New York: Routledge.

Sheller, M. and Urry, J. (2006). The new mobilities paradigm. *Environment and Planning A* 38, 207 226.

Tolley, R. S. and Turton, B. J. (1995). *Transport Systems, Policy and Planning: A Geographical Approach*. Harlow: Longman.

Transport, Public

J. Preston, University of Southampton, Southampton, UK

Glossary

Elasticity The ratio of the proportionate change in one variable with respect to the proportionate change in another. An example of an own elasticity is the elasticity of the demand for bus with respect to its own price. An example of a cross elasticity is the elasticity of the demand for bus with respect to the price of car.

Interconnection The extent to which transport networks are connected in physical terms both within and between modes.

Intermodality Passenger (or goods) movements that consist of a combined chain from origin to destination involving at least two different modes (excluding walking for passengers).

Interoperability The ability of national and geographically defined transport networks to provide operations and services across national borders and across physical and technical barriers.

Returns to Scale The proportionate change in output divided by the proportionate change in inputs (and hence with fixed input prices, the proportionate change in costs). It is the inverse of the cost elasticity with respect to output.

Value of Time The ratio of the marginal utility of time to the marginal utility of money. It is used to convert time into money units.

Introduction

Public transport may be defined as any mode of transport available for hire and reward. In other words, any form of transport available for use by the general public. This includes not only bus and rail, but also taxis, as well as air and sea services. It can include both passenger and freight services, although own account freight services (where a company moves its own goods) are not normally thought of as public transport. In practice, the term public transport (transit in the US) has tended to refer to land based passenger transport and, especially, bus and rail services.

Even where the definition of public transport is limited to bus and rail services (and variants thereof), there is a spectrum of modes that may be considered can be distinguished in terms of costs and capacity (see **Figure 1**). Low capacity and low cost public transport systems include taxis, mini micro buses and paratransit. Higher cost variants include demand responsive transport (DRT) and personal rapid transit (PRT) systems. Examples of PRT include the ParkShuttle system in use at Schiphol Airport (Amsterdam) and the ultra light transport (ULTra) system being developed for Heathrow Airport (London).

Medium capacity public transport systems include various bus and tram technologies including single deck, double deck, and articulated buses/coaches and trolley buses. In addition, there are traditional trams (streetcars in the US) and more modern light rapid transit (LRT) systems. Bus systems may have their own road space (bus lanes and busways) and their own guidance (either curb, electronic, or optical). Guided bus (O bahn in Germany) and LRT systems are sometimes referred to as intermediate public transport as they may be seen to combine features of both bus and rail systems. However, conventional bus systems, when operated with articulated vehicles in convoys with dedicated stops and rights of way, can reach high capacities. An early example of such a system was Sao Paulo (Brazil). More recently these systems are being formalized as bus rapid transit (BRT) systems, as exemplified by Bogota (Colombia).

High capacity public transport systems are normally associated with heavy rail technology. A distinction can be made between urban, suburban, and inter urban services. Urban rail services (U bahn in Germany) are typified by short station spacings, high frequencies, electric traction, and underground routes in the central area. Examples include the London Underground, the Paris Metro, and the New York Subway. Suburban services (S bahn in Germany) have longer spacings between stops, less frequent services, run predominantly on the surface, and can include diesel traction (at least outside Germany). Inter urban services have long gaps between stops, relatively infrequent services, and can involve a

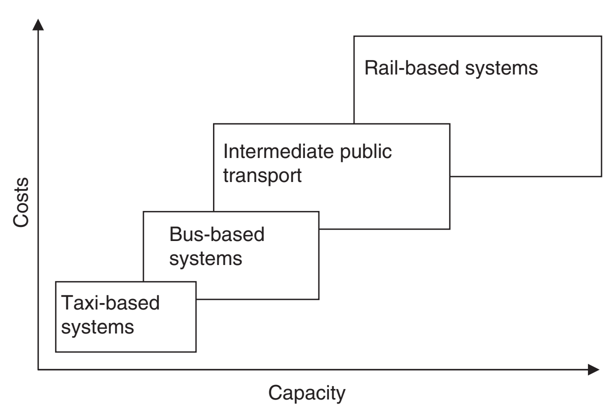

Figure 1 Schematic taxonomy of public transport modes.

mix of diesel and electric traction. A distinction can be made between conventional rail services and high speed services (the definition of which usually requires top speeds in excess of 250 km h^{-1}) such as the Shinkansen in Japan, the train à grand vitesse (TGV) in France, and the InterCity Express (ICE) services in Germany.

Interesting technologies that do not easily fit into these categories include monorail systems (such as the early suspended system in Wuppertal, Germany, and the more recent straddle system in Sydney, Australia), mag lev (such as the Shanghai Airport service), and flywheel technology (such as the Parry People Mover system in Stourbridge, England). Hybrid track sharing systems have been developed in Germany, with light rail and heavy suburban rail sharing tracks (Karlsruhe) or even guided bus and rail sharing tracks (Essen).

There is no such subdiscipline as public transport geography and whatever work has been done by geog raphers has tended to be interdisciplinary, carried out in conjunction with engineers, economists, and planners, among others. Academic work on public transport has focused on three main themes. The first theme is that of public transport operations as spatial systems worthy of study in their own right. The second theme, driven by transport economists as much as geographers, focuses on the economics of public transport systems and has tended to provide more analytical accounts, with an emphasis on elasticities, returns to scale (RTS), and cost–benefit op timization. The third theme highlights the role of public transport as a derived demand. People do not usually demand public transport in, and of, itself (although in some cases, such as heritage steam railways, they clearly do) but from the derived benefits of being able to engage in economic and social activity. There has therefore been interest in ways that public transport interacts with the wider economy and promotes social inclusion, as well as the impact of public transport on the environment, its interaction with private transport, and its contribution to sustainable development.

Operations

One area that geographers have focused on is the role of public transport operations across the globe. A useful distinction here can be made between developed, de veloping, and transitional economies. In developed countries, public transport use often peaked during or just after World War II and since the early 1950s has exhibited a long decline, particularly for bus transport. The United States represents an extreme example where, outside the major cities, bus and rail use is negligible. Countries such as Australia and Canada exhibit similar trends. The decline has been less extreme in Western Europe and in the developed economies of East Asia.

Table 1 Passenger transport shares in selected countries (2002)

	Car (%)	*Bus and coach (%)*	*Rail (%)*	*Passenger kms per capita*
Great Britain	88.1	6.4	5.5	12 430
France	86.7	4.7	8.6	14 283
Germany	82.7	8.9	8.4	10 276
Poland	77.0	13.5	9.5	5691
Japan[a]	61.5	7.0	31.5	9592
USA[a]	96.2	3.5	0.3	23 388

[a] 2001 data.

Source: UK Department for Transport. (DfT) (2005). Transport Stat istics Great Britain. London: DfT. Tables 10.1 and 10.6.

Indeed in certain countries (Switzerland, Japan), absolute levels of public transport use have increased, although the relative importance of public transport has declined with the rise of the motor car. **Table 1** shows that even among major developed nations there are large variations in mobility levels and modal shares, with public transport only having a 4% share of the passenger market in the US but a 38% share in Japan.

Whereas the problems for public transport in de veloped countries are associated with insufficient and declining demand, in developing countries problems tend to be associated with growing demand and insufficient supply. Public transport use is often high, rail provision is limited, and conventional bus services are overcrowded, unreliable, and slow. Bus fleets are often characterized by high failure rates, due to inadequate investment in ve hicles and maintenance facilities and a shortage of spare parts (which often have to be imported). Paratransit services of various types and various degrees of legality have often emerged to satisfy unmet demand. Examples include Jeepneys in the Philippines, Dolmus in Turkey, KombiVans in South Africa, and Matatu in Kenya. These modes are sometimes referred to as informal or un conventional public transport. With respect to long distance passenger transport, rail plays an important role in countries such as China and India but has a less important role in much of Latin America and Africa.

The role of public transport is different again in the transitional economies of the Former Soviet Union and the countries of Central and Eastern Europe. Under Communist rule, public transport had very high market shares, with particularly high use of tram and trolleybus technology. The transition to market economies saw a rapid decline in public transport use and a corresponding increase in motorization. Market shares have since sta bilized somewhat (albeit still with year on year losses of share for public transport), and remain higher than in Western Europe. A major challenge is the funding of the investment required by the dilapidated public transport infrastructure.

One area of active study is the determination of the appropriate public transport mode in different locations. The emphasis has been on radial corridors in major cities in the peak hour and peak direction, where the choice has between bus, heavy rail, and intermediate public transport technologies (see **Figure 2**). However, geographers have also highlighted that suburb to/from central business district movements no longer dominate urban transport patterns in the way they once did. Suburb to suburb movements have correspondingly become more important. These circumferential (or orbital) movements are more difficult to serve by conventional public transport due to their diffuse nature and the low density of flows. Emphasis has been placed on examining the scope for more flexible forms of public transport such as dial a ride and web based versions that are collectively referred to as demand responsive transport. The scope for these services to feed into conventional public transport at key interchange transfer points has also been examined. Quantitative variants of this work have involved the developments of public transport network modeling tools such as the Volvo Interactive Planning System (VIPS – now incorporated in the VISUM suite of programs) and the use of operations research tools to optimize the use of vehicles and staff.

With respect to inter urban transport, a key area of study has been examination of the evolution of public transport networks. Initial work was stimulated by Taaffe, Morill, and Gould's deterministic model of the development of transport networks in West Africa. Behavioral and probabilistic variants were also developed, such as Kansky's work on the development of the railroad network in Sicily. More generic work on the evolution of transport systems highlighted the role of historical, technological, physical, economic, political, and social factors. With respect to the morphology of transport systems, one area of research has been efficacy of fully connected networks compared to sparser networks based on hub and spoke principles. The bulk of research has focused on air passenger transport and container shipping, but there have been applications to express coach and urban buses. Quantitative variants of this work use graph theory and accessibility indices to measure the impacts of new public transport links.

There has also been interest in the interrelationships between different scales and types of (public) transport. The European Commission has referred to this as the 'three Is'. The first I is the physical interconnection between local (urban and rural), national, and international networks. Trans European Networks (TENs) are being

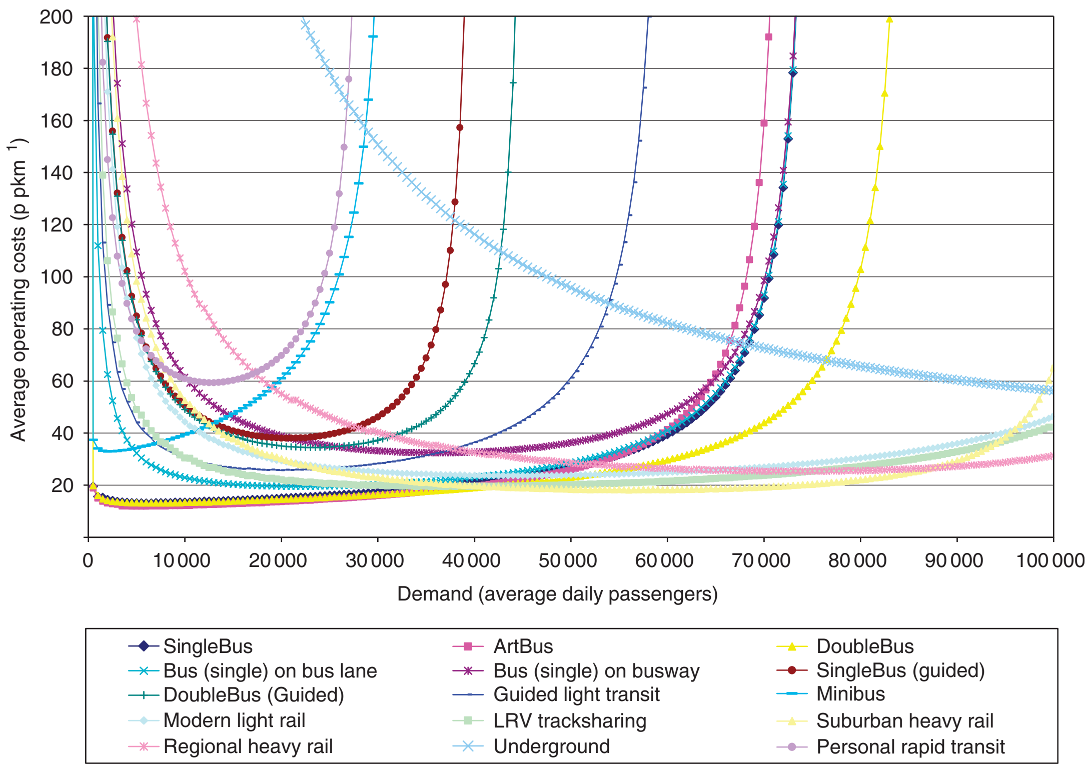

LRV Light rail vehicle.

Figure 2 Average operating costs as a function of demand for a 12 km urban public transport route (in pence per passenger kilometer, 2000 prices). Source: www.tsu.ox.ac.uk/research/test.

developed to improve interconnectivity. The second I is the interoperability between different transport systems. For example, although there are physical connections between Europe's rail systems, differences in track and loading gauges, electric traction, train control, signaling, and operating practices, make it difficult for services to operate across national systems. European Rail Traffic Management System (ERTMS) has been developed to overcome such technical aspects of the interoperability issue. The third I is intermodality. This is normally framed with respect to road services providing feeder services for the movement of containers by rail but is also relevant for public transport, particularly the develop ment of the parkway concept for rail services.

Economics

Academic work on public transport has been somewhat dominated by transport economists who tend to down play the spatial nature of their work. The key contri bution of geographers has been to highlight the spatial nature of economic concepts such as public transport demand, supply, prices, investments, regulation, and ownership.

With respect to demand, economists note that public transport is like most other economic goods, in that de mand is a function of the price of the good, the price of rival goods, income, and tastes. However, demand is also a function of the amount supplied, as waiting times and walking times decrease as service is increased. This leads to the Mohring Effect whereby average user cost de creases as more service is provided. Demand is also a function of journey quality and in particular speed. Price and time variables are often brought together into a generalized cost measure.

In public transport demand analysis, the key indi cators are elasticities, which measure the ratio between the proportionate change in one variable and the pro portionate change in another. The most widely used elasticity is the elasticity of demand with respect to price, which is the proportionate change in demand divided by the proportionate change in price. For example, for local bus travel an elasticity of –0.4 is often quoted which suggests that if bus fares increase by 10%, a decrease in demand of 4% can be expected. By contrast, for many long distance journeys by rail the price elasticity is greater than 1 (in absolute terms). This is an important threshold because when the elasticity exceeds 1 (in ab solute terms), a reduction in fares can lead to increases in revenue. It is in such elastic markets that budget fares are offered. As well as for price, elasticities have been esti mated for journey time, wait time (or proxies such as headway, frequency, or vehicle kilometer), access/egress time, and reliability. Compilations of the available evidence on elasticities are provided by a number of bodies, such as the Transport Research Laboratory (TRL) in the UK (see also **Figure 3**), the Transportation Research Board (in the US), and the Bureau of Transport and Regional Economics (BTRE) in Australia. Some researchers have conducted meta analyses of these studies. In absolute terms, elasticities tend to be higher: in the off peak (than the peak); for optional, leisure jour neys (rather than mandatory, business and commuting journeys); for inter urban journeys (compared to local journeys); and in situations where public transport has competition from other modes. High income travelers (such as business people) tend to be time sensitive with high service quality elasticities but low price elasticities. Low income travelers (such as students and the elderly) tend to be price sensitive with high price elasticities and low service quality elasticities.

Besides own elasticities – the elasticity of one aspect of a particular good (say rail demand) with respect to another aspect of that good (say rail price) – cross elasticities have also been heavily researched. They are the elasticity of one aspect of a particular good (say public transport demand), with respect to another aspect of another good (say private transport price). They will be positive where goods are substitutes (car and bus) but negative where goods are complements (car and rail based park and ride). They are heavily influenced by market shares. In advanced economies, where public transport has relatively low market shares, the cross elasticity of car demand with respect to public transport price is often very low (substantially less than 0.1). By contrast, the cross elasticity of public transport demand to car price can be quite high as can cross elasticities within the public transport market.

Another measure that is widely used is the value of time. This is the ratio of the marginal utility of time to the marginal utility of money. A number of meta analyses have been undertaken of values of time. For travel during the course of work, empirical studies suggest the value of time is broadly equivalent to the wage rate plus over heads (superannuation, national insurance, etc.). For other travel (including travel to and from work), it is often found to be between a quarter and a half of the wage rate. The value of time is therefore closely corre lated with income. The value of travel time for rail is usually higher than that for bus for two reasons. First, rail has a higher proportion of business travelers. Second, rail users tend to have higher incomes than bus users. Similar measures have been developed to determine the values placed on comfort (e.g., the level of overcrowding), in formation provision (pretrip, at stop, and on vehicle), station facilities, etc.

With respect to the supply of public transport, economists note that this is a function of the price of public transport, the price of rival products, input prices

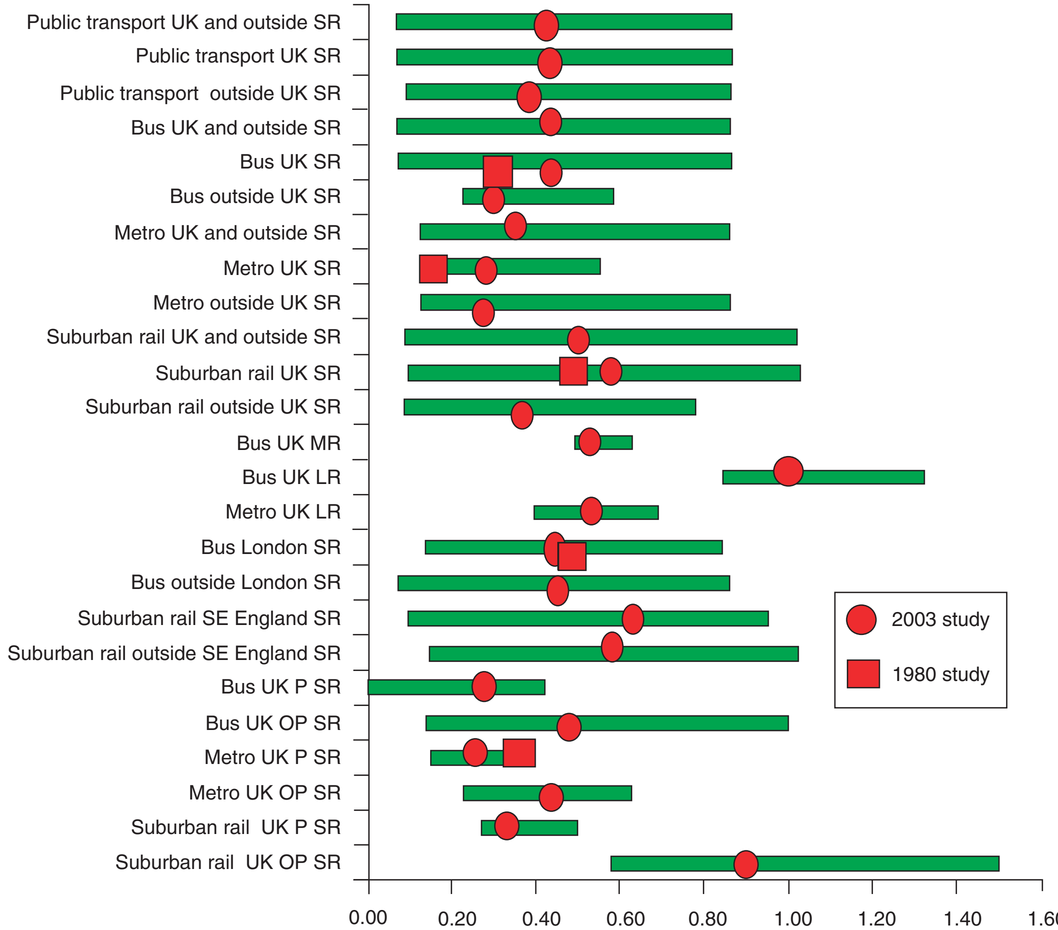

Figure 3 Summary of mean values and ranges of fare elasticities. SR Short run, MR Medium run, LR Long run, P Peak, OP Off peak. Source: Balcombe, R. (ed.) (2004). *The Demand for Public Transport: A Practical Guide*. TRL Report 593. Crowthorne: TRL, Figure 3.1.

(and in particular the wage rate and the price of capital – the interest rate), the objectives of producers, and the level of technology. In advanced economies, where capital is relatively cheap and labor expensive, capital intensive forms of production are employed such as the automated Val system in Lille (France). In developing economies, where labor is cheap and capital expensive, labor intensive forms of production are employed such as the hand drawn rickshaws of Kolkotta (India). Cost and productivity studies have been deployed to determine the optimal conditions for public transport production and a number of reviews are available in the literature. Studies show that the cost of provision in the peak is often much higher than the off peak (due to lower utilization of staff and vehicles) and, for bus travel, provision in urban areas is more costly than in rural areas, largely due to the greater peakiness of demand and more congested roads. Publicly owned firms may have output maximizing rather than profit maximizing objectives and will tend to produce more service than privately owned firms. Moreover, the costs of publicly owned firms are often higher than privately owned firms, while firms facing competition often have lower costs than firms protected from competition.

In studies of public transport supply, emphasis has been placed on RTS – the proportionate change in output divided by the proportionate change in inputs (which, with fixed input prices, is the inverse of the cost elasticity with respect to output). Where RTS are greater than 1, the firm exhibits increasing returns (economies of scale) and would benefit from being larger (see **Figure 4a**). Where the RTS are less than 1, the firm exhibits decreasing returns (diseconomies of scale) and would benefit from being smaller. Where the RTS are 1, the firm exhibits constant returns and is of optimal size (at least in terms of technical efficiency). For the bus industry, econometric studies suggest that firms exhibit constant returns across a broad range of outputs from 100 vehicles to 1000 vehicles or more. However, this assumes constant input prices. In practice, smaller firms may benefit from the use of cheaper, nonunionized labor, while larger firms can exploit purchasing power

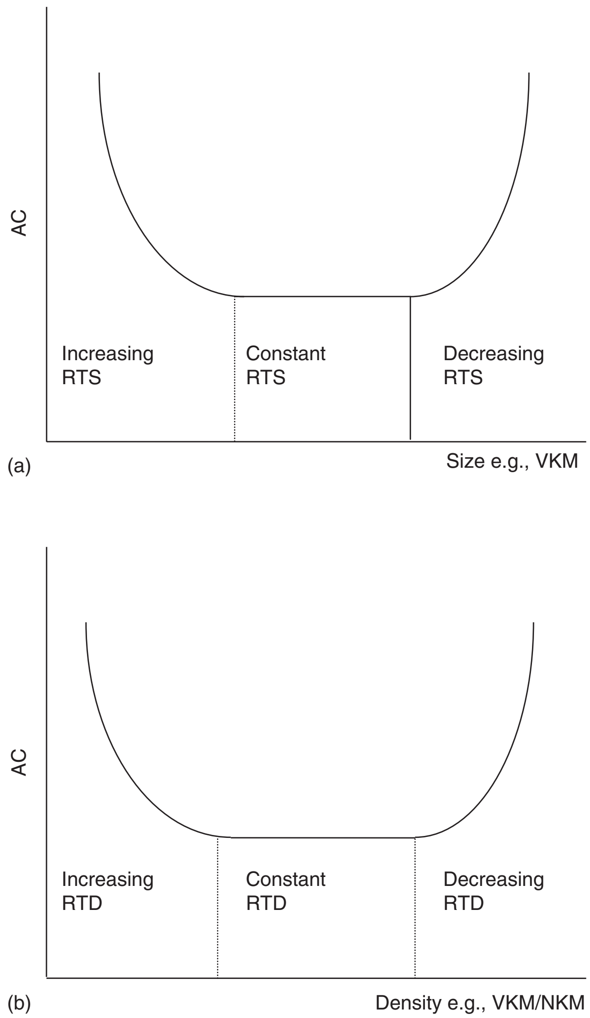

AC Average cost NKM Network kilometers RTD Returns to density RTS Returns to scale VKM Vehicle kilometers

Figure 4 (a) Economies of scale; (b) Economies of density.

advantages for fuel, vehicles, and capital. It has been argued that the 'big five' stock exchange listed public transport companies that have emerged in the UK (First Group, Stagecoach, Arriva, Go Ahead, and National Express) try to simultaneously mimic being big (e.g., purchasing done centrally) and small (operations devolved to the depot level). For the rail industry, econometric studies in Europe suggest that there may be economies of scale for small railway companies such as in the Republic of Ireland and diseconomies for large railway companies such as Deutsche Bahn (DB, Germany) and Société Nationale des Chemins de fer Français (SNCF, France). Optimal sized companies might include the railways in Belgium and Norway. However, most railways exhibit increasing returns to density (RTD) in that costs decrease less proportionately than density (train kilometers per route kilometer) (see **Figure 4b**). This is particularly true of the lightly trafficked Nordic railway (Finland, Norway, Sweden). By contrast, the congested system in the Netherlands exhibits decreasing returns.

With respect to fares, simplified systems based on flat fares, zonal fares, or distance related fares, with or without a taper with distance, are the norm for local public transport. For longer distance public transport, more complex fare regimes exist, partly because of the prevalence of prebooking and off vehicle sales. In such circumstances, public transport firms may try to introduce perfect price discrimination in which prices vary by the amount of travel purchased and by the type of traveler (first degree price discrimination). For example, discounts may be offered to regular travelers through the provision of season tickets, network cards, and smartcards (such as Octopus in Hong Kong or Oyster in London) or through brand loyalty schemes (similar to the frequent flyer programs offered by airlines). This is referred to as second degree price discrimination. In addition, different fares may be offered to different types of travelers. For example, many train companies charge higher fares for services believed to be used by business travelers and lower fares for services believed to be used by leisure users, based on the time and direction of travel, the degree of flexibility required, and the time of booking. This is referred to as third degree price discrimination and may be compounded by product differentiation (e.g., the distinction between first class and standard rail services). The outcome of such revenue yield systems is a geography of public transport prices that is highly variegated by both space and time.

Similar variations exist in terms of patterns of public transport investment and subsidy. Cost–benefit analyses show that the best returns from public transport investments tend to occur in congested urban areas and on heavily used inter urban routes, although recent work suggests that the appraisal of mega projects, particularly rail, may exhibit systematic upward biases. Investments in smaller urban areas and in rural areas, at least in advanced economies, can rarely be justified on narrowly defined cost–benefit grounds but can be justified using more qualitative multicriteria analysis.

In many parts of the world, public transport is subsidized and there are a number of reasons why this should be so. The Mohring effect provides a first best rationale for subsidizing high frequency urban public transport. Operator economies of scale (or more precisely density) may provide some basis for subsidizing certain inter urban services. The fact that car users do not pay the marginal external costs they generate (in terms of congestion, accidents, and environmental costs), provides a second best argument for subsidizing public transport. Other arguments for subsidizing public transport include its option value as a standby for when private transport is not available and its merit good characteristics in providing basic levels of mobility for disadvantaged groups (such as the young, the elderly, and those on low incomes). Cost–benefit studies tend to show that public transport subsidies exhibit the highest returns in the big cities. However, in practice rural services are often heavily subsidized through revenue support, dispensation on fuel tax, and user subsidies.

Another area of study has been that of ownership and competition in public transport. The classic public transport model has been one of public ownership and control. This model still exists in much of North America and Western Europe. The alternative model is one of private ownership and deregulation of fares and service levels. This is the model that was adopted for local buses in Great Britain (outside London) in the mid to late 1980s and was subsequently adopted by some other countries (most notably Chile and New Zealand). This is sometimes referred to as competition on the road. Yet another model, referred to as competition off the road, involves the public authority specifying services and private firms competing for the tenders to provide the services. Pioneers of this model form include San Diego (USA), London (UK), and Copenhagen (Denmark). Similar models exist in the rail industry but with a further distinction in terms of the extent of vertical integration between operations and infrastructure. When Japan corporatized and then privatized its railways, they remained vertically integrated. By contrast, when railways were privatized in Great Britain, it was against a backdrop of a vertically fragmented industrial structure. Transport economists relate public ownership and control to public interest theories that emphasize the role of the state in controlling market failures. Economic geographers relate such polices to the Fordist/Keynesian consensus and regimes of mass accumulation. Similarly, transport economists relate the privatization and deregulation of public transport to neoliberal public choice theories that emphasize regulatory failure as a result of processes of capture and creep. Economic geographers relate this to post Fordist regimes of flexible accumulation.

Integrating Issues

There are a number of issues that link public transport with other sectors of socioeconomic activity. The first issue is the impact that public transport can have on the wider economy. Geographers have long been interested in the interactions between land use and transport. Work by the International Study Group on Land Use and Transport Interaction (ISGLUTI) showed how public transport systems can lead to the centralization of urban employment and the decentralization of population. There have been extensive studies of the impact of new urban public transport systems on land values, historical studies of the impact of railways on the development of national economies, and on the regional impacts of high speed rail. Recently the new economic geography has focused on the extent to which public transport (particularly rail) can promote agglomeration benefits.

A second issue is the extent to which public transport can promote social inclusion. Studies have shown how cheap, frequent public transport can enhance the accessibility of facilities and promote participation in civil society. However, these same studies have also shown how certain groups, such as the elderly, the disabled, and ethnic minorities, find existing public transport systems difficult to use. Research has therefore been undertaken to re engineer public transport so as to promote usability.

A third issue is the extent to which public transport can promote environmental betterment. Per passenger kilometer, buses and trains tend to consume less energy and land, emit less air pollutants, make less noise, and contribute less to climate change than cars. However, this green advantage is being eroded as cars are designed to be more eco friendly and trains, for example, are tending to become heavier and more powerful.

A fourth issue is the extent to which public transport can contribute to integrated transport. This has a number of aspects which are sometimes referred to as rungs on the integration ladder. The bottom rung is the integration between different modes of public transport, including coordinated timetables, integrated fare structures, and comprehensive information provision. The Verkehrsverbund of West Germany such as in Hamburg were early pioneers of this form of integration. Middle rungs involve the integration of public and private transport policies. An example is the use of the congestion charge in central London to finance improvements to the bus system. Upper rungs of integration involve the coordination of transport activities with those in education, healthcare, and social services, while also integrating transport policy with policies concerning land use, the environment, and regional development.

The fifth issue is the extent to which public transport can contribute to sustainable development. Sustainability could be thought of as a high level form of integration which brings together appropriate economic, social, and environmental criteria. International studies have shown that cities with high population densities and high public transport shares tend to be more sustainable, for example, in terms of energy consumption per capita. Examples include Hong Kong and Singapore. The development of public transport is seen as a key strategy in the promotion of smart growth. The role of public transport in promoting sustainable communities is likely to be crucial for its future prospects.

See also: Transport and Deregulation; Transport, Urban.

Further Reading

Balcombe, R. (ed.) (2004). *The Demand for Public Transport: A Practical Guide*. TRL Report 593. Crowthorne: TRL.

Berechman, J. (1993). *Public Transit Economics and Deregulation Policy*. New York: North Holland.

Davison, L. J. and Knowles, R. D. (2006). Bus quality partnerships, model shift and traffic decongestion. *Journal of Transport Geography* 14(3), 177 194.

De Borger, B., Kerstens, K. and Costa, A. (2002). Public transit performance: What does one learn from frontier studies? *Transport Reviews* 22(1), 1 38.

Flyvbjerg, B., Bruzelis, N. and Rothengatter, W. (2003). *Megaprojects and Risk. An Anatomy of Ambition*. Cambridge: Cambridge University Press.

Hanson, S. and Guiliano, G. (eds.) (2004). *The Geography of Urban Transportation*. New York: The Guilford Press (in particular see Chapter 8, Pucher, J., Public transportation).

Hay, A. (1993). Equity and welfare in the geography of public transport. *Journal of Transport Geography* 1(2), 95 101.

Iles, R. (2005). *Public Transport in Developing Countries*. Oxford: Elsevier.

Kansky, K. J. (1963). *Structure of Transportation Networks: Relationships between Network Geometry and Regional Characteristics*. Department of Geography Research Paper 84. Chicago, IL: University of Chicago.

Knowles, R. D. (1996). Transport impacts of Greater Manchester's metrolink light rail system. *Journal of Transport Geography* 4(1), 1 14.

Mohring, H. (1972). Optimisation and scale economies in bus transport. *American Economic Review* 62(4), 591 604.

Nash, C. A. (1981). *The Economics of Public Transport*. London: Longman.

Newman, P. and Kenworthy, J. (1999). *Sustainability and Cities. Overcoming Automobile Dependence*. Washington, DC: Island Press.

Office of the Deputy Prime Minister (ODPM)/Royal Institute of Chartered Surveyors (RICS) (2003). *Land Value and Public Transport*. London: ODPM.

Oum, T. H., Waters, W. G. and Yu, C. (1999). A survey of productivity and efficiency measurement in rail transport. *Journal of Transport Economics and Policy* 33(1), 9 42.

Pratt, R. H. *et al*. (2000). *Traveler Response to Transportation System Changes*. TCRP Web Document 12, Washington DC: TRB.

Shaw, J. (2000). *Competition, Regulation and the Privatisation of British Rail*. Aldershot: Ashgate.

Taaffe, E. J., Morril, R. L. and Gould, P. R. (1963). Transport expansion in underdeveloped countries: A comparative analysis. *Geographical Review* 53, 503 529.

UK Department for Transport (DfT) (2005). Transport Statistics Great Britain. London: DfT. Tables 10.1 and 10.6.

Wardman, M. R. (1998). The value of travel time: A review of British evidence. *Journal of Transport Economics and Policy* 32(3), 285 316.

Webster, F. V., Bly, P. H. and Paulley, N. J. (1988). *Urban Land Use and Transport Interaction Policies and Models*. Report of the International Study Group on Land Use Transport Interaction. Aldershot: Gower.

White, P. R. (2002). *Public Transport: Its Planning, Management and Operation* (4th edn.). London: UCL Press.

Relevant Websites

http://www.dynamic.dotars.gov.au
Bureau of Transport and Regional Economics (BTRE) Transport Elasticities Database.

http://www.cpt uk.org
Confederation of Passenger Transport (CPT).

http://www.eltis.org
European Local Transport Information Service (ELTIS).

http://www.its.usyd.edu.au
International Conference on Competition and Ownership in Land Passenger Transport (the Thredbo Conference Named after the Location of the First Conference in 1989), Institute of Transport and Logistics Studies (ITLS).

http://www.uitp.com
International Union of Public Transport (Union Internationale des Transports Publics UITP).

http://www.uic.asso.fr
International Union of Railways (Union Internationale des Chemins de fer UIC).

http://www.tsu.ox.ac.uk
Transport Studies Unit, University of Oxford.

Transport, Rural

D. Banister, University of Oxford, Oxford, UK

Glossary

Accessibility The ability to reach needed services and facilities.

Accessibility Paradigm Builds on the uniqueness of place and the ability to reach needed services, but explicitly includes accessibility as a relative concept that has strong social and distributional implications.

Demand Responsive Transport Public transport that does not operate on a fixed route or times, but 'responds' to the requirements of its passengers.

Mobility The amount of travel undertaken.

New Mobilities Paradigm Places are not fixed but implicated within complex networks which bring together certain performances in certain places at certain times.

Introduction

The car dominates the transport scene in rural areas and this situation is likely to remain so, with the USA leading the way as over 97% of rural households own a car and over 91% of all trips are made by car. The pattern of usage across all ages, incomes, and race, is the same, and mobility levels are higher than those in urban areas due to the longer distances traveled. In one sense the rural transport problem has been solved and there is no debate in the USA, as those without cars are politically invisible as their numbers are so small. Similar patterns of car dependence can be found in other parts of the world, but to a lesser degree of domination. However, these headline figures disguise the fact that there is a sub stantial minority that cannot drive or does not have access to a car.

Nationally in the UK, about a third of households have two or more cars (2005), but in rural areas this figure is 52%. About 75% of UK households have a car, but in rural areas this figure rises to 89% (2005). Within these headline figures, there is a substantial minority of people, including those under 17 years of age (nearly 20% of the rural population), households that are in the lowest income groups (35% of those in the lowest quintile are in non car owning households), and those that are too old to drive (19% of adults do not drive). In each of these groups (too young, too poor, or too old), mobility depends on others (lifts), public transport, and the availability of services and facilities within walking distance. The total figure for those without access to car as a driver is about 32%. This figure is derived from those who are too old or living in non car owning households – about 12% – and those too young to drive – about 20%. Getting around in rural areas is not a problem if you have the exclusive use of a car as a driver, but it is for the one third of the population that is dependent on others for their transport.

Physical accessibility is described within the context of rural development and it is argued that social con structs should be seen as being as important. This in cludes the levels of car availability and its affordability. Alternative conceptualizations of accessibility are pre sented in terms of the remoteness and income trade offs, the importance of social justice and social inclusion, and the essentially aspatial arguments presented as part of the new mobilities paradigm. The contrasting situation in developing countries is covered as rural migration and abstraction may take place out of the rural areas to the cities, leading to hardship and greater poverty in the rural areas. Finally, some future prospects are summarized.

Transport and Rural Development

Rural development provides the wider context within which transport facilitates the means to gain access to work, leisure, and other opportunities, but the uniqueness of location must also be recognized. Each situation is different and there is no universality in solutions, as illustrated in **Table 1**. Much of the debate has revolved around physical concepts of accessibility, with less attention being paid to social and distributional issues such as affordability, employment opportunities, gender differences, and educational and skills levels. Remoteness is not a matter of poor infrastructure in rural areas, as the quality and extent of the road network are both good, but a problem of income, social constraints, and lack of services for those without cars. It seems that social justice has strong spatial dimensions in rural areas – this is the poverty of access.

Addressing this problem requires a wide range of innovative thinking and services, including promoting better use of available car and taxi services, such as the taxicab innovation resulting from the deregulation of bus services in rural Sweden. There is a considerable trans port resource already available, and at a modest cost it would be possible to use new technology to encourage lift sharing, car pooling, and flexibility in travel times to raise occupancy levels. Many of the travel demand

Table 1 Transport and rural development

Evidence	*Comments*
Economic issues	
1. Higher levels of employment are observed in rural areas, with increases in commuting distances and willingness to travel	Restricted to the better-paid and well-qualified population; problems for low-skilled and older (>50 years) unemployed, and for women, part-time employed, and others (e.g., students)
2. Informal networks are effective at finding jobs in buoyant local labor markets where there are reasonable levels of accessibility	Possession of a driving licence is a good indicator of employability and motivation to get a job
3. Perceptions of the rural area important to employers and inward investment good quality transport infrastructure often quoted as an important factor	Transport costs are a small proportion of overall business costs transport infrastructure enables labor mobility rather than reducing distribution costs
4. With new infrastructure, competition increases across labor market areas and may reduce wages in rural areas even further	Wage rates for rural residents may also rise if they are prepared to travel to work elsewhere net effect is a widening difference in rural income levels
5. There is limited evidence on specialization and increases in productivity brought about by transport investment	Need here for more monitoring and ex-post studies; transport costs important at the margin and need to be seen in relation to profits, as well as costs
6. Migration to rural areas through better transport may raise rural house prices	Limited evidence of this effect, and it may also put upward pressures on rural wage levels
7. Abstraction effects of transport investments, and the opportunity for businesses to serve regions rather than localities	Debates about increases in agglomeration and economies of scale from larger production units, and the resulting increases in travel distances this depends on the costs of transport
8. Logistics is more important than just cost minimization	Distribution networks are based on a mix of cost, time, reliability, and flexibility criteria
9. Limited evidence of road-induced relocation and expansion	Importance of roads in perceptions of competitiveness and image of area
Social issues	
10. Social exclusion is reduced through subsidy to the service and to the user	The modes used are more important than the infrastructure hence the tendency to invest in capital projects means that social exclusion may be increased, not reduced
11. Fuels costs are important to the rural poor as there are no alternative forms of transport to the car	The impacts of higher fuel prices can be over-estimated, as rural drivers travel further but more economically
12. Accessibility in rural areas has been used primarily in a physical transport context combining income and remoteness	Accessibility needs to be broadened as a concept to include institutional and organizational barriers; poverty (and lack of skills and motivation) is also a structural issue

From Banister, D. (2005). What are sustainable rural communities? Thought piece for the commission for rural communities, Cheltenham, October, p 38.

patterns in rural areas are small in scale and diffuse, and ideal for private transport. There may also be environmental benefits in the sense that a full car or taxi is better than an empty bus.

The main aim of public transport policy should be to increase effectiveness through transforming bus companies from service providers to managers of a range of mobility services. This would include better co-ordination of all services currently operated by a variety of different agencies (e.g., Demand Responsive Transport), the introduction of rural transport brokers (already operating in some local authorities in the UK), much closer linkages with all service provision (e.g., pensions, postal services, shops, health centers, and education facilities), and a new role of facilitator bringing together car sharing schemes, the voluntary and community sectors, social services transport, and the delivery of goods and services to needy households. Transport would become fully integrated with rural service provision and it would concentrate on providing for those without access to a car, as well as improving the use of all available forms of transport.

The broader issue raised here is whether public policy should continue to subsidize the higher costs of providing a whole range of services and facilities in rural areas. The longer distances make it more expensive and harder to provide the same level of provision as in urban areas, and this problem is magnified by the lower levels of demand in rural areas. The provision of affordable housing in more accessible locations might encourage people to move from the remote areas to the more accessible areas, where services and employment opportunities are easier to provide. The debate here concerns the promotion of more traditional approaches to public responsibility for social welfare, as compared with the more recent expectations of individualism and self reliance, while at the same time recognizing the importance of cultural differences and accepting that there is no universal solution.

Social Constructs

Accessibility planning is concerned about coordinating transport with business priorities and the need to maintain social inclusion, so that those at risk can be identified and access to employment increased. In the past, little attention seems to have been placed on ensuring accessibility to key services and employment, either through transport or location policies, and this has increased social exclusion, with the common perception that the lack of access to private transport is an instrumental barrier. The lack of an integrated approach to the location of employment and key services, combined with poor public transport provision, have again reinforced social exclusion, and it is often those same excluded people that are affected by the negative externalities from transport due to the higher levels of traffic noise and pollution. This argument was powerfully presented by the UK Social Exclusion Unit (2003), but reflects the wider concerns in the European Union (EU), Australia, and the USA about migration from rural areas and the perception that rural areas are becoming the domain of rich car drivers seeking a higher quality of life (**Figure 1**).

Specific groups at risk have been highlighted. An analysis of secondary data sourced mainly from UK government statistics, found that poor transport limits women's employment and quality of life, as women rely more upon trip chaining and off peak travel (owing to part time work). Rethinking public transport to meet the needs of women would have commercial, inclusion, and environmental benefits. In addition, evidence shows that women can face constraints in their travel patterns, and hence employment opportunities, as a result of their household responsibilities.

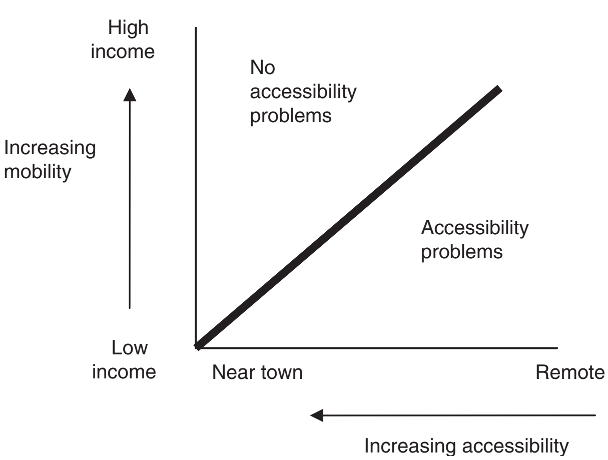

Figure 1 The remoteness and income trade-offs.

The Select Committee on Education and Employment (1998) reported that lone parents were at a disadvantage in getting to work as only 35% have access to a car, compared with 90% of couples with children. Similarly, other research has found that travel problems, connected with childcare, low wages, and part time work can act as a barrier to employment and that some employers also discriminated on travel to work grounds, particularly if the employee did not have their own car.

Residents of rural areas are at risk from exclusion. A survey of unemployed job seekers in the north of Scotland highlighted the problem in that despite the majority of these job seekers holding current driving licenses, relatively few had access to their own transport. The cost of owning and running a vehicle was cited as the main reason for this. For 54% of the long term unemployed, the costs associated with private transport were viewed as an important barrier to work. Socio economic factors, such as gender, having dependent children, and low education levels, have been found to be more influential in determining shorter potential travel times of job seekers (and hence increasing the chance of gaining employment) than the provision of public transport accessibility and access to private transport.

The evaluation of the rural transport partnerships (RTPs) in England has examined about 100 of the more than 600 schemes funded by the Countryside Agency in collaboration with other rural agencies. Their main benefits have been to help lever additional levels of funding for innovative transport schemes, and this in turn has benefited local businesses through increased local spending by residents and tourists. The RTP projects have also opened up new opportunities for those without alternative transport, and over a third of projects were considered to have contributed to the regeneration of deprived areas. Half the projects also impacted positively on rural employment, through providing direct jobs, as well as encouraging higher levels of rural spend.

Reductions in social exclusion can be achieved through enhanced accessibility provided by improvements to the planning and delivery of local transport and the location of employment and key services in accessible locations. Existing planning policy guidelines are designed to promote the location of employment in accessible areas by ensuring that new business developments which employ large numbers of people are located in urban centers and/or near public transport interchanges.

New Conceptualizations

Transport problems in rural areas can be summarized along two dimensions, remoteness and income levels (**Figure 1**). Along the diagonal there are few problems, as those on high incomes have a car available and can afford to live in the remote locations. Less affluent households can live in or near to the urban centers where there are better public transport facilities and where travel distances are shorter. Those locations or income groups above the diagonal also have no problems as the income axis outweighs the remoteness axis.

It is in those households that fall below the diagonal where problems lie, as their income remoteness situation is not in balance. This is often not a physical problem of lack of transport infrastructure, but a social and economic problem of lack of income and suitable public transport or other service provision. This can be addressed through subsidy and service provision (both transport and non transport alternatives) to reduce the levels of remoteness or raise their effective income levels so that they move toward the diagonal – this is a mobility solution to increase their propensity to travel through subsidy. Alternatively, they could move away from the remote area to again reach the threshold situation (i.e., the diagonal in **Figure 1**) at a given level of income. This is an accessibility solution to increase their proximity to jobs, services, and facilities.

The responsibility for accessibility planning cuts across most public sector activities in rural areas, suggesting that a 'structural' perspective has been used that assumes the means to tackle social exclusion is essentially aspatial. Poverty and low income, lack of qualifications and skills, low levels of economic activity, as well as age and gender, are all structural factors used to explain the lack of opportunity, but equally important is the issue of poverty of access.

Accessibility is central to social justice, but not all areas can have or should have the same levels of accessibility. People make residential location choices as to what is more important to them, whether it is acces sibility or a higher quality environment. In this new accessibility paradigm it is argued that accessibility is not an absolute concept but one that has strong relative dimensions in terms of needs and rights (in the Rawlsian tradition). By setting accessibility as a general goal, it provides the necessary impetus for policy sectors to work together to achieve greater social justice and social inclusion.

This contrasts with the new mobilities paradigm that moves away from spatiality and scale, arguing that places are not fixed but implicated within complex networks by which

> hosts, guests, buildings, objects and machines are con tingently brought together to produce certain perform ances in certain places at certain times. (Hannam *et al.*, 2006: 13)

Here it is argued that places are not rural but they are dynamic and create co presence in a variety of ways. In both paradigms thinking has gone way beyond tradi tional physical concepts of remoteness and income (**Figure 1**), to cover social inclusion and social justice on the one hand, and social capital and the sociology of networking, including the existence of small worlds that are spatially independent, on the other.

The new more abstract notion of social capital and the nonexistence of space ignores the spatial imperative and the real difficulties that poor residents in rural areas have as a result of their physical remoteness. There needs to be an acceptance that social capital and networks are dependent on accessibility, and that limited mobility will in turn limit social capital. This seems to be a reformu lation of the old debate between mobility, as the amount of travel and networking (quantity), and accessibility, as the ease with which needed facilities and activities can be reached (quality).

Developing Countries

The provision of rural transport services is often cited as being central to development in developing countries, as it is argued that there is a strong link between economic growth and transport. However, the evidence is not clear as transport provides a strong complementarity to other economic and social factors. Transport is again linked to rural development, and the role it plays in mitigating poverty through providing employment (in construction and operation), reducing levels of household expenditure, improving economic efficiency in moving goods and ser vices, in providing access to public services (e.g., health care and education), and in improving health, quality of life, and the environment. Research has also provided a clear example of the differential impacts of transport in providing access to education, in terms of getting children to school, and getting them to stay in class.

In developing countries, rural transport is primarily concerned with poverty reduction. Low income groups receive the lowest quality transport options, and at the same time these groups are exposed more than others to transport's negative impacts. There is much evidence to suggest that the poorest spend a higher percentage of their income on transport in both developed and developing nations. The poor also receive the brunt of the risks and negative externalities of motorization, despite the fact that few of the poor own motorized transport themselves. Low income groups tend to live at the periphery of employment centers and thus have greater travel distances. Gender and age inequities create mobility challenges for large parts of the population, and women face specific restrictions on movement due to the complex nature of their activity patterns.

The same issues are raised here as in developed countries, but with a much greater degree of polarization as the levels of poverty are correspondingly much higher. There is also an insatiable demand for more travel and rapid motorization, with likely further inequalities arising, as well as dramatically increased levels of resource consumption and emissions. Nevertheless, there are some well documented examples of innovation that have clear implications for transport elsewhere, as they combine high quality with low cost solutions. These include bus rapid transit (mainly in and around cities), cycle priorities, a range of unconventional paratransit and informal transport services, and car free areas, as well as nonmotorized initiatives (in Africa). All of these can substantially raise rural accessibility, and many of these innovations could be transferred to developed countries.

Future Prospects

The future of rural transport should combine technologies that allow the best use of cars to raise occupancy levels in developed countries, with innovative nonmotorized and collective transport ideas from developing countries, to provide mixes of modes that respond to the needs of current patterns of travel and communication in rural areas. Conventional rail and bus services will be limited to corridors of higher demand between larger urban centers, where development can be concentrated at the key locations along these routes so that high levels of accessibility can be maintained. House prices may well have captured the quality of the transport network through the higher prices paid for properties in some more accessible rural homes. The network effects can be both positive and negative (two way road effect), and the difficulty is in isolating real net increases in the local economy (additionality) in terms of new jobs, inward investment, and tourism, as compared with a redistribution effect from one place to another.

Further increases in car ownership in rural areas in all countries will gradually marginalize those without access to a car, but there are many opportunities to reduce rural inaccessibility in both physical and social terms. The fundamentally different geographies, economies, and social priorities outside the USA suggest that innovation in terms of management and organization of existing services, looking at transport as part of a range of rural activities, and experimenting with new modes and ideas could substantially benefit those living in rural areas without access to a car.

See also: Distance; Mobility; Poverty; Poverty, Rural; Rural Geography; Transport and Accessibility; Transport and Social Exclusion; Transport Geography.

Further Reading

Banister, D. (2005). What are sustainable rural communities? Thought piece for the commission for rural communities, Cheltenham, October, p 38.

Banister, D. and Wright, L. (2005). *The Role of Transport in Supporting Sub National Growth*. Report prepared for the Department for International Development, May, p 42.

Brake, J., Nelson, J. D. and Wright, S. (2004). Demand responsive transport: Towards the emergence of a new market segment. *Journal of Transport Geography* 12(4), 323 337.

Commission for Rural Communities (CRC) (2005). *The State of the Countryside 2005*. Wetherby: Countryside Agency Publications.

Countryside Agency (2004). *Evaluation of the Rural Transport Partnerships*, Countryside Agency Research Note, CRN78, June.

DHC (2004). *Relationships between Transport and Rural Economies*. Report on Literature Review and Proposed Methods, prepared for the Countryside Agency by Derek Halden Consultancy, November.

Farrington, J. and Farrington, C. (2005). Rural accessibility, social inclusion and social justice: Towards conceptualisation. *Journal of Transport Geography* 13(1), 1 12.

Gray, D., Shaw, J. and Farrington, J. (2006). Community transport, social capital and social exclusion in rural areas. *Area* 38(1), 89 98.

Lindsay, C., McCracken, M. and McQuaid, R. (2003). Unemployment duration and employability in remote rural labour markets. *Journal of Rural Studies* 19(2), 187 200.

Nutley, S. D. (1996). Rural transport problems and non car populations in the USA: A UK perspective. *Journal of Transport Geography* 4(2), 93 106.

Nutley, S. D. (2003). Indicators of transport and accessibility problems in rural Australia. *Journal of Transport Geography* 11(1), 55 71.

Pucher, J. and Renne, J. L. (2005). Rural mobility and mode choice: Evidence from the 2001 National Household Travel Survey. *Transportation* 32(2), 165 186.

Rosenbloom, S. (2003). Facing societal changes: The need for new paradigms in rural transit service. *Journal of Public Transportation* 6(1), 1 18.

Social Exclusion Unit (2003). *Making the Connections: Final Report on Transport and Social Exclusion*. London: ODPM. http://www.cabinetoffice.gov.uk/social exclusion task force/~/media/assets/www.cabinetoffice.gov.uk/social exclusion task force/publications 1997 to 2006/making transport 2003%20pdf.ashx (acessed in June 2008).

Vasconcellos, E. A. (1997). Rural transport and access to education in developing countries: Policy issues. *Journal of Transport Geography* 5(2), 127 136.

WBCSD (World Business Council for Sustainable Development) (2001). *Mobility 2001: World Mobility at the End of the Twentieth Century*. Geneva: Atar Roto Presse.

Relevant Websites

http://www.rural transport.net
Actions on the Integration of Rural Transport Services.

http://www.cfit.gov.uk
Rural Transport: An Overview of Key Issues, Published by the UK Commission for Integrated Transport (2001).

http://www.ruralcommunities.gov.uk
The UK Commission for Rural Communities.

http://www.worldbank.org
The World Bank's Homepage for Rural Transport.

http://www.transport links.org
Transport for Development.

Transport, Urban

B. P. Y. Loo, University of Hong Kong, Hong Kong, People's Republic of China

Glossary

Land Use–Transport Interrelationship The close and complex set of relationships between land use and transport. Traffic is seen as a joint consequence of land-use activities and transport capabilities.

Metro System An off-street, high-speed, heavy railway system which operates on exclusive rights-of-way and in the urban areas of cities.

Paratransit It includes public transit services which have flexible routes, timetables, and/or fares. Sometimes, these services may be provided to particular market segments only.

Sustainable Urban Transport An urban transport system which minimizes adverse environmental impact, ensures the long-term economic and financial feasibility, and promotes social equity.

Transit-Oriented Development A type of development which promotes higher public transit usage. It is often characterized by a compact and mixed land-use pattern near major transit stops.

Urban Freight Transport The movement of goods within cities and is essential to support the urban economy. Over time, a larger share of the urban goods movement is organized and carried out by logistics companies.

Introduction

An urban transport system considers all elements of transportation, including the infrastructure, rolling stock, and traffic flows, in a city as a functional entity. The interrelationships among different system components are emphasized. Generally speaking, the most notable characteristic of an urban (vs. rural, regional, national, or international) transport system is the high concentration of human activities within a limited space, especially at the commercial and business districts (CBDs), of the cities. The results are severe traffic congestion and other associated air quality, parking, and safety problems at the CBDs and the major arteries leading to it, especially during the morning peak hours when people commute to work. In many cities, the average traffic speed at the CBD fell below 15 km per h during the morning peak hours (usually 08.00–10.30 a.m).

Strategies to solve the urban traffic congestion problems can broadly be divided into the supply oriented and demand oriented ones. The former focuses on the provision of extra infrastructure, such as downtown suburb expressways, urban ring roads, and car parks. The latter refers to transport demand management (TDM) measures, which range from simple ones like traffic signals, intersection controls, and stopping restrictions to the more advanced ones like flexitime, congestion pricing, and high occupancy lanes. In an urban context, it is obvious that the scope of providing extra infrastructure is often limited because of the dense, built up areas and the high land price. Moreover, the provision of more transport infrastructure would arguably lead to newly generated transport demand, called induced demand, because of the expanded transport capacity and the shorter travel time.

Various TDM measures were adopted by cities at different development levels. Congestion pricing is one of the TDM measures advocated by economists for a long time but has been receiving renewed interests in the twenty first century. The essential idea is that the road users should pay for the external costs that they impose on the other road users by entering the congested urban areas. The congestion pricing scheme can vary from a uniform charge per day to variable pricing depending on the season, day of the week, time of the day, or even actual congestion level of the road or area (defined by cordons). Congestion pricing differs from the other road pricing schemes, such as road and tunnel tolls, whose primary aim is to raise revenue. Experiences of congestion pricing and its various impact have been closely examined in London (United Kingdom), Oslo (the Netherlands), Singapore, and Stockholm (Sweden). Nonetheless, despite all these efforts, urban traffic congestion is still prevalent in all large cities worldwide. The world's first urban traffic congestion pricing scheme was implemented in Singapore in 1975. Its 'area licensing scheme' (ALS) had undergone significant extensions in coverage (in terms of the types of vehicles, area, and time) before it became fully automated and integrated into the electronic road pricing in 1998. In Europe, the London congestion charge has attracted wide attention because it was one of the largest congestion pricing schemes in the world and was introduced by a democratically elected city mayor. Like the experience of Singapore, both the charge and the spatial extent of the London congestion charge have increased since its inception in 2003.

Urban Passenger Transport

Different Urban Passenger Transport Modes

Passenger transport often represents the largest volume of urban traffic. In policy analysis, it is customary to distinguish between private and public transport. The former includes nonmotorized modes of walking and cycling, and motorized modes of motorcycles and auto mobiles. The latter includes conventional mass carriers, such as the metro systems, light rails, trams (streetcars), and buses, and paratransit modes, such as taxis, share ride taxis, jitneys, minibuses, and other public transit modes which have flexible routes, timetables, and/or fares. Many low technology and slow paratransit modes are common in Southeast Asia. Examples include the 'tuk tuk' in Thailand and India, 'tricycles' in the Philippines, and 'becaks' in Indonesia. Other paratransit modes are provided for certain market segments only. School buses, demand responsive buses for the transport disadvantaged (particularly the disabled and the elderly), and residents' coach are some cases in point. Residents' coach, offering exclusive and high quality service to middle class resi dents living in low density suburbs, has had some success in replacing private car use. Overall, there is a broad range of paratransit both in terms of technology and standard of service.

Figure 1 shows the major means of urban passenger transport in terms of their carrying capacity and the environmental impact (particularly air pollution). In the figure, it is assumed that all on road vehicles use diesel or gasoline. Technological advancement means that alternative fuel vehicles (AFVs), including liquified petroleum gas (LPG) and electric vehicles, have become more widely used over time. Attempts to introduce AFVs have been made in many cities, to reduce the reliance on fossil fuels and to minimize the environmental impact caused by the heavy, urban vehicular traffic. Monetary incentives (such as a reduction of the 'first registration tax' and other forms of subsidies) and legislative amendments have been made to encourage the replace ment of private automobiles by AFVs and the use of AFVs to provide public transport services. Some notable examples include California in the United States and Hong Kong in the People's Republic of China. In California, the California State Government has actively encouraged the use of AFVs. Following the amendments to the Clean Air Act in 1990, a new classification of vehicles based on tailpipe emission was introduced. Plans and legislation were formulated to facilitate the transition to AFVs. The California Air Resources Board oversees that vehicle manufacturers in California meet mandatory low emission vehicle sales quotas. In addition, electric or natural gas (particularly LPG or CNG) dual fuel vehicles were promoted in Canada, Japan, New Zealand, and the Netherlands. In Hong Kong, LPG taxis were made mandatory; and there were government support schemes for purchasing electric or LPG minibuses.

As a city develops, people desire higher personal mobility. One way to achieve this is through a high reliance on motorization. This has to be supported by massive investment on the construction of expressways and the road network, an expanding automobile industry, and the heavy consumption of nonrenewable energy. A second way to achieve higher personal mobility is through the development of the urban public transit system. Overall, the relative importance of motorized private transport differs substantially across different cities, even with similar development levels. This can range from as low as 10% of the trips made by private transport in Hong Kong to more than 90–95% in most American cities.

Although the reasons for the varying importance of public transport vary in different cities, most large cities

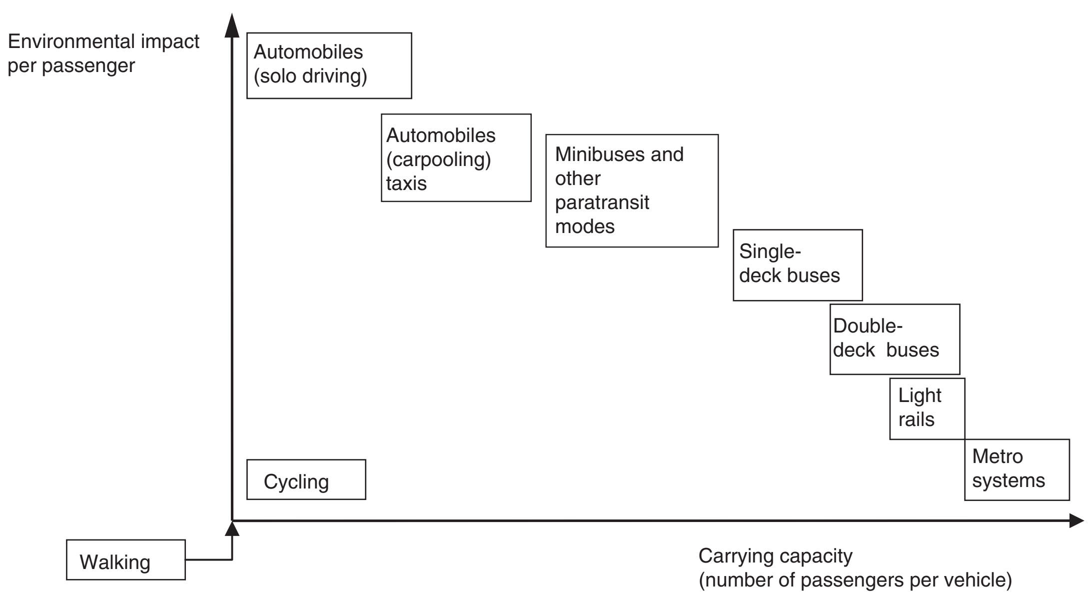

Figure 1 Different passenger transport modes in an urban transport system.

with high shares of public transport were served by metro systems. Metro systems refer to the heavy railways with exclusive rights of way serving the urban areas of the cities. However, the building of metro systems requires a very long lead time (typically 7 years or more) and heavy capital investment (ranging from US$60 million to US$0.3 billion per km). However, the building of the metro system *per se* would not guarantee that it could generate all the expected beneficial effects of encouraging higher public transport usage and fulfilling other policy objectives such as a compact urban form. Hence, very often, the introduction of the metro systems requires coordination with other mass transit modes to ensure sufficient railway patronage. Examples include the cancellation of long distance bus routes running parallel to the railway lines and/or the realignment of bus routes which compete directly with the railways. So far, evi dence from different cities with high public transport patronage suggests that a well integrated multimodal public transport system is crucial for public transport to be attractive vis à vis automobiles. These attempts include the introduction of the integrated public fare collection system, transfer discount, and timetabling coordination among various public transport modes. Intermodal cash cards, such as the Metrocard in New York and the Oyster Card in London, are common in major global cities. In the case of Hong Kong, the Octopus Card is used not only as a means of payment for all mass transit modes but also at car parks, shops, and restaurants. The use of advanced travelers' information systems based on the geographic information systems (GIS) represents another way of improving the attract iveness of the public transit system. The provision of timely and reliable public transit information is believed to enhance the attractiveness of public transit vis à vis private cars.

Land Use–Transport Interrelationship

Urban form also has an important role to play in affecting the patterns of urban passenger transport. In particular, a mixed land use pattern is conducive to higher public transport patronage and shorter average travel distance. Different land uses have different traffic generation and attraction rates, which can be input into the four stage transport models. In wake of the calls for sustainable urban transport, new initiatives have been introduced in many cities to reduce the cities' dependency on automobiles and to promote public transport and nonmotorized transport, namely cycling and walking. Though these new initiatives differ in their emphases and, hence, their policy recom mendations, they can be broadly classified as transit oriented development (TOD) and be contrasted with the automobile dependent development, as listed out in **Table 1**.

From **Table 1**, one sees that this broad definition of TOD includes elements of the smart growth strategy, the location efficient strategy, and the walkable com munities. While these various initiatives have slightly different emphases, they all aim to reduce the automobile dependency of the community and to encourage the de velopment of a sustainable urban transport system in the long run. In particular, the smart growth strategy was first proposed as an alternative to urban sprawl. Hence, the strategy puts an emphasis on the smart use of the existing urban areas and the slowing down of the spatial expan sion of cities. The location efficient strategy likewise emphasizes the possibility of in fill development in cities, but it also puts much importance on housing affordability. Lastly, the walkable communities advocate walking as the primary mode of passenger transport in cities. In addition, it is worth mentioning that there have also been renewed interests in developing paratransit modes in the United States, as they are more flexible and adaptable to the low density and dispersed suburban ization process.

Urban Freight Transport

Types of Urban Freight

A discussion on the urban transport systems would not be complete without addressing the urban freight issues. In the first place, it must be recognized that the move ment of goods supports the economic vitality of the urban economy. An efficient urban freight system is essential to support the high spatial agglomeration of people and economic activities in cities. Generally, urban freight movements may be classified as intra urban goods movement and through traffic. Intra urban goods move ment is primarily associated with the wholesaling and retailing activities of the city. Most cities do not have exclusive freight railway to handle intra urban goods movements. Hence, most of them take place in the form of road transport, especially container trucks, lorries, and light goods vehicles. As a result, city center delivery and curbside loading and unloading activities often aggravate the severe urban traffic congestion. In many cities, there are restrictions for goods vehicles to enter the CBD during the morning peak hours. Off peak delivery is therefore common. In addition, recent changes in supply chain management, such as the rising importance of just in time (JIT) deliveries, lower inventory levels, and more frequent deliveries, have caused new challenges to urban freight planning and management in urban areas. In medium and large cities, the setting up of urban dis tribution centers (UDCs) and integrated freight centers (IFCs) is often worth serious considerations. Moreover, in the light of the more stringent demands on urban freight movement, more and more urban freight is being handled

Table 1 Characteristics of automobile and transit-oriented development

	Automobile-oriented development	*Transit-oriented development*
Road network	Expressway-oriented road system with hierarchical street patterns designed to maximize traffic volume and speed.	Few expressways. Highly interconnected streets with traffic calming measures.
Parking availability	Abundant, easily accessible parking lots at low or no cost.	Parking management in force. Limited and multi-story parking charged at relatively high cost.
Public transport services	Very poor and/or restricted to paratransit only.	Very good, in terms of comfort, frequency reliability, and speed.
Pedestrians and cyclists	Not friendly to pedestrians and cyclists, no pedestrian walkways on some streets.	Friendly to pedestrians and cyclists with high accessibility to all major attractions. In some circumstances, the multi-exit railway station forms a well-connected and segregated pedestrian walking system.
General land-use pattern	Segregated and dispersed land-use pattern at the urban fringe, with little in-fill development.	Dense, compact, and a balanced mixed land-use pattern around transit stops, particularly railway stations.
Housing types and density	Single-lot, single-household detached houses in a spread out manner.	A variety of affordable housing types with high-rise and mixed residential and commercial buildings near the transit stations acting as the centers of towns or suburb communities. Residential density declines progressively outward.
Public and private facilities	Major city facilities, such as shopping malls, public services, schools, stadiums and other leisure venues are located at major expressway exits requiring automobile access. Large-scale single-story or low-rise retail malls away from the city center, surrounded by free parking lots.	The CBD remains the main transit hub and the center of major business, entertainment, government/institutional, and retail activities. Medium-scale retail shops on the lower floors of the commercial and resident buildings near transit stations. Schools, cafes, and neighborhood stores on the ground level of residential buildings to help create attractive street life.
Public space	Limited. Private club facilities, large private yards, and gated communities are common.	Abundant and well-planned. Provision of public parks, sports centers, libraries, and other public facilities.

From Elmwood Renaissance. (2007). Walkable vs. automobile oriented communities. http://www.freewebs.com/elmwoodrenaissance/walkablevsautooriented.htm; Fehr & Peers. smartgrowthplanning.org: The transportation perspective. www.smartgrowthplanning.org; Reconnecting America. Reconnecting America: Center for Transit oriented Development. http://www.reconnectingamerica.org; Victoria Transport Policy Institute. (2007). Online TDM Encyclopedia. http://www.vtpi.org/.

by third party logistics companies. In the future, changes in supply chain management will continue to drive changes and innovations in urban freight movements.

In addition, through traffic is important in the urban freight system of cities which serve as gateways, major manufacturing centers, and/or important transport and logistics hubs. Amsterdam (the Netherlands), Hong Kong, New York (USA), and Singapore are some notable examples. In these cities, freight transport to the ports and airports can constitute an important part of the urban transport systems. The segregation of such through traffic from the intra urban freight transport and the urban passenger transport is sometimes practiced for both se curity and traffic management reasons. Moreover, vehicle tracking technology (supported by the 'global positioning system', GPS) has been increasingly employed by large logistics companies to manage their fleet. In some large countries, such as Canada, China, and the USA, long distance freight movement should preferably be carried on the national freight rail system for both environmental and economic reasons. Hence, there have been calls to encourage more freight rail movements in both the USA and Europe.

Problems of Mixed Traffic

In situations where on road mixed traffic (between pas senger and goods vehicles) is prevalent, several transport problems are particularly serious in an urban context. First, heavy goods vehicles pose greater danger to the other road users (particularly pedestrians and cyclists), when a crash happens. While research has shown that heavy goods vehicles may not be associated with higher accident rates, it causes much concern because crashes involving heavy goods vehicles usually result in a higher number of casualties to the other kinds of road users than the heavy goods vehicle users themselves.

Second, the transportation of dangerous goods, such as chemical wastes, petroleum, and LPG, needs careful planning to avoid critical urban transport links, tunnels, and bridges, to prevent catastrophic events in case of traffic crashes. Third, there are concerns about the higher road maintenance costs related to the use of the heavy goods vehicles, especially when they are loaded. In Germany, trucks over 12 tons have been charged on a kilometer basis since 2003 to partly reflect the more severe pavement damage caused by these heavy vehicles. Overall, it is argued that the current heavy goods vehicle charging systems do not fully reflect the extra wear and tear that these heavy goods vehicles caused to the road system.

Last but not least, heavy goods vehicles, especially if they are loaded, generate higher levels of air pollution, noise pollution, and roadside vibration. These problems are particularly acute in the dense urban environment. Furthermore, the efficiency of fuel energy consumption among heavy goods vehicles is another area of concern. Currently, AFVs available in the commercial market are mainly passenger rather than goods vehicles. The technologies for developing heavy goods vehicles running on alternative fuels are much more complicated and less mature. Hence, the reliance of the heavy goods vehicles on diesels is unlikely to fall substantially in the near future. All of the above affect the long term sustainability of the urban transport system and raise important issues that need to be addressed by the government.

Conclusions

Overall, the key for developing a sustainable urban transport system lies in the best possible combinations of transport and other (notably, land use) policies that meet the personal mobility needs of the people and the goods mobility needs of the economy. In all the major arenas of urban passenger and freight transport, government policies can play a critical role in affecting the urban transport system not only through its public transport policy but also other key public policies such as land use and planning policies.

See also: Governance, Transport; Transport and Social Exclusion; Transport Geography; Transport, Public; Transportation and Land Use.

Further Reading

Allsop, R. E. (2005). Some traffic safety implications of movement of goods by road. *Transportmetrica* 1(1), 3 12.
Banister, D. (ed.) (1998). *Transport Policy and the Environment*. London: E & FN Spon.
Banister, D. (2005). *Unsustainable Transport: City Transport in the New Century*. London: Routledge.
Banister, D., Button, K. and Nijkamp, P. (eds.) (*c*. 1999). *Environment, Land Use and Urban Policy.* Cheltenham: Edward Elgar Publishing.
Bernick, M. and Cervero, R. (1997). *Transit Villages in the 21st Century.* New York: McGraw Hill.
Buchanan, C. D. (1963). *Traffic in towns: A study of the long term problems of traffic urban areas. Reports of the Steering Group and Working Group appointed by the Minister of Transport.* London: HMSO.
Button, K. J. and Pearman, A. D. (1981). *The Economics of Urban Freight Transport*. London: Macmillan.
Cervero, R., Murphy, S. and Ferrell, C. *et al.* (*c*. 2004). *Transit Oriented Development in the United States: Experiences, Challenges, and Prospects: Transit Cooperative Research Program Report 103.* Washington DC: Transportation Research Board of the National Academies.
Hanson, S. and Giuliano, G. (eds.) (2004). *Geography of Urban Transportation*. New York: Guilford.
Hau, T. D. (1992). *Economic Fundamentals of Road Pricing: A Diagrammatic Analysis*. Washington DC: Transport Division, Infrastructure and Urban Development. World Bank.
Hesse, M. (2005). *Freight Transport and Logistics in an Urban Context: Cities under Pressure of Structural Change and Supply Chain Dynamics*. n.p.: Ashgate.
Knowles, R., Shaw, J. and Docherty, I. (eds.) (2008). *Transport Geographies: Mobilities, Flows, and Spaces*. Malden: Blackwell.
Loo, B. P. Y. and Li, D. Y. N. (2006). Developing metro system in the People's Republic of China: Policy and gaps. *Transportation* 33, 115 132.
National Research Council, Commission on Behavioral and Social Sciences and Education (*c*. 1994). Curbing gridlock: peak period fees to relieve traffic congestion. *Report of National Research Council, Commission on Behavioral and Social Sciences and Education*. Washington DC: National Academy Press.
Vuchic, V. R. (1981). *Urban Public Transportation: Systems and Technology*. Englewood Cliffs: Prentice Hall.
World Bank (2002). *Cities on the move: A World Bank urban transport strategy review*. Washington DC: World Bank.

Relevant Websites

http://www.freewebs.com
Elmwood Renaissance. Elmwood Renaissance: Working toward pedestrian friendly, mixed use, smart development.
http://www.smartgrowthplanning.org
Fehr & Peers, the transportation perspective.
http://juts.janes.com
Jane's Information Group. Jane's Urban Transport Systems.
http://www.reconnectingamerica.org
Reconnecting America: Center for Transit Oriented Development.
http://www.smartgrowth.org
Sustainable Communities Network (SCN). Smart Growth Online.
http://www.vtpi.org
Victoria Transport Policy Institute. Online TDM Encyclopedia.

Transportation and Land Use

S.-L. Shaw, The University of Tennessee, Knoxville, TN, USA

Glossary

Discrete Choice Model A mathematic model that predicts an individual's choice based on the utility of a finite set of competing alternatives.

E-Commerce Commerce conducted with electronic transactions over the Internet.

Monocentric City A city with most employments concentrated in one core area.

New Urbanism An urban design movement that promotes walkable and transit-oriented neighborhood designs with mixed-use land development.

Smart Growth A concept that advocates policies of encouraging land-use patterns that are compact, walkable, transit-oriented, mixed-use development, and providing quality housing for all income groups.

Spatial Interaction Model A model used to estimate the amount of flows between different places.

Teleworking Working at home or locations other than the office using information and communications technologies.

Urban Sprawl Spread of urban developments to nearby undeveloped land around a city.

Virtual Mobility Substitution of electronic exchanges and online communications for physical transport activities.

Introduction

In today's society, most of us live at one location, but work, shop, and conduct other activities at separate locations. In other words, we carry out our activities at different locations to fulfill our physiological, economic, social, and other needs. The spatial distribution of these activity locations for all individuals and businesses collectively gives the land use pattern of an area. Transportation serves as the means to move people and goods among different activity locations. In this regard, it is evident that transportation systems and land use systems are mutually related to each other.

Up to the early twentieth century, most cities were developed with compact and mixed use neighborhoods. As transportation technologies improved, urban development pattern began to change. After World War II, due to rapid growth of private automobiles and adoption of modern zoning regulations, a new urban development pattern with significant suburban growth became widespread across the United States and many other countries. This is known as 'urban sprawl', which encourages spatial separation of various activity locations and increasing dependence on private automobiles.

By the 1970s, some researchers began to advocate new urban design and urban planning approaches to help reduce the negative impacts (e.g., traffic congestion, air pollution, and urban sprawl) caused by the widespread use of private automobiles. The basic idea behind these approaches is to encourage more compact and mixed use developments that encourage the use of nonmotorized travel (e.g., walking and biking) and transit systems. These urban design and planning approaches have been called under different names such as mixed land use development, transit oriented development, smart growth, and new urbanism. They all reflect an idea that a fundamental change in land development patterns may be a promising way to change people's travel behavior and therefore change their travel patterns and reduce the automobile induced problems.

Transportation and land use interaction received renewed interests in recent years due to the concept of sustainable development. Sustainable development suggests development strategies that meet the needs of the present generation without compromising the ability of future generations to meet their needs. In the United States, legislations such as the Clean Air Act Amendments (CAAA) of 1990, the Intermodal Surface Transportation Efficiency Act (ISTEA) of 1991, the Transportation Equity Act for the 21st Century (TEA 21) of 1998, and the Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (SAFE TEA LU) of 2005, all call for considerations of the likely effects of transportation policy decisions on land use in order to achieve more sustainable development plans. In Europe, concerns about sustainable development also led to many studies sponsored by the European Commission's Framework Programmes to examine transportation and land use issues. One key reason that transportation and land use interaction has been an active subject pursued by geographers, planners, and policymakers is because transportation and land use policies often involve significant investments and can influence the evolution of urban development patterns.

Transportation and Land-Use Interaction

Transportation and land use interaction is often described in the literature as a mutual and dynamic relationship (see the inner box in **Figure 1**). Spatial distribution of transportation infrastructure and services shapes the accessibility levels of different locations. Different accessibility levels then influence the land development and land use decisions. Changes in land use patterns in turn lead to new activity and travel patterns in space and over time. New activity and travel patterns will then affect the decisions on planning and design of future transportation systems, which in turn will lead to future changes in accessibility and the next round of transportation and land use interaction.

In addition to the mutual and dynamic relationships between transportation and land use, many other processes also influence the evolution of transportation and land use systems in an urban environment (see the outer box in **Figure 1**). Land development decisions are affected not only by transportation systems but also by factors such as the amount of developable land in an area, availability of utility infrastructure (such as sewer and water), regional economic trend, and local zoning regulations, among others. By the same token, transportation investment decisions also are influenced by factors other than land use patterns. For example, advancements in information and communications technologies (ICTs), such as the Internet and cellular phones, have made teleworking, e commerce, e learning, e government, and other virtual mobility more feasible and convenient. They are expected to lead to changes in activity and travel patterns that have direct implications to future configurations of transportation and land use systems. As a result, it is important to realize that transportation and land use interaction cannot be easily isolated from other economic, technological, political, and social changes in an urban system.

There are in general two ways to study transportation and land use interaction. One is through empirical studies and the other is to use models. Empirical studies can help us identify the change patterns observed in real world. However, we need to keep in mind that transportation and land use interaction takes place in a complex and dynamic manner. Effects of transportation system changes on land use, and vice versa, occur at varying spatial and temporal scales. For example, construction of a new rail transit line tends to have greater impacts around the rail stations and along the transit corridor than on the urban area as a whole. This suggests that we could reach contradictory conclusions regarding the effects of transportation system changes on land use if we have conducted empirical studies of the same urban area at different spatial scales. On the other hand, building of a new expressway may lead to immediate land use changes in one area, but a similar expressway project in another area may not induce significant land use changes until many years after its completion. The different outcomes could be due to the influences of other processes shown in **Figure 1** that evolve at different rates over time in different areas. Consequently, an ideal comprehensive model of transportation and land use interaction should consider all of these processes and their mutual, dynamic, and complex relationships. In practice, a model is often based on some simplifying assumptions to make it feasible and manageable. Since empirical studies *per se* cannot forecast future land use and travel demand patterns, models are necessary to help us derive the forecasts.

Theories and Models

Changes in transportation and land use systems are often considered as the outcomes of location choices. Land developers make decisions on where to build a residential subdivision or a commercial development. Transportation planning agencies must decide where to add a new rail transit line or to improve an existing highway. These location choice decisions influence each other in space

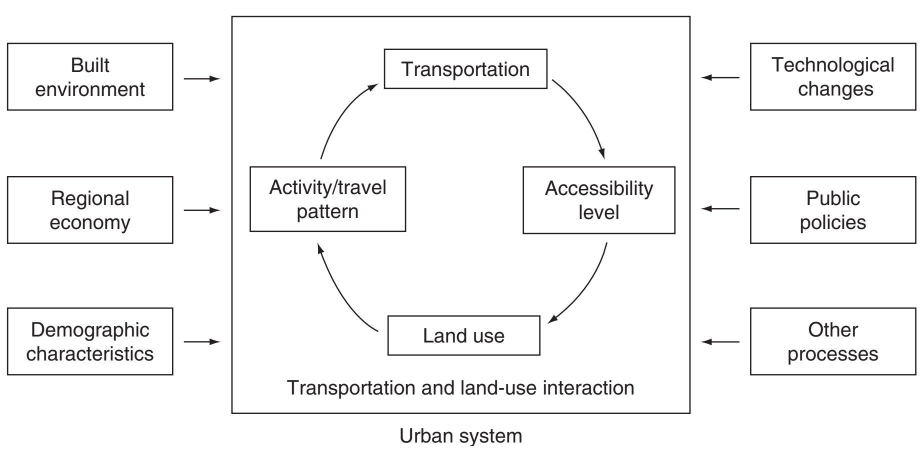

Figure 1 Transportation and land-use interaction within an urban system.

and over time. Many theories and models have been developed to describe the relationships between trans portation and land use. In general, they can be organized into three groups based on their emphasis on influence of transportation on land use, influence of land use on transportation, or an integrated transportation and land use approach.

Influence of Transportation on Land Use

The earliest land use theory can be traced to von Thünen's agricultural land rent theory of 1826, which is based on a concept of balancing between transportation cost and land rent. Locations closer to market incur a lower transportation cost and therefore have a higher land rent. Under the assumptions of an isolated and self sufficient state with completely homogeneous land and one central market, von Thünen's theory suggests con centric zones of agricultural activity from the market. High value and perishable agricultural products (e.g., dairy products and fresh vegetables) are located closest to the market while low value goods (e.g., grain and live stock) are located in outer zones. Residential location theory was developed later on based on a concept similar to von Thünen's agricultural land rent theory. It suggests that people make trade offs between commuting cost and housing price when choosing their residential location. A decreasing commuting cost therefore encourages people to live farther away from the urban center and leads to population decentralization.

For employment location theory, Christaller's central place theory has been used to describe the locations of market centers. Christaller examined the urban settle ments in southern Germany to derive the rules that determine the size, number, and distribution of towns. Central place theory suggests a hierarchical development of settlements or market centers. The spatial arrange ment and size of these market centers depend on the minimum market threshold level and the range that consumers are willing to travel for goods or services. As transportation cost declines, we can expect a more dispersed pattern of larger market centers. Weber's industrial location theory, on the other hand, focuses on goods movements rather than people movements. Weber's theory suggests that the best industrial location is at the place where the total transportation cost of shipping raw materials and finished product is min imized. The above location choice theories are based on simplifying assumptions such as an isotropic surface, uniform transportation cost in all directions, a mono centric city, and single worker households. A funda mental principle of these theories is that agricultural, residential, employment, and industrial location choices are mainly shaped by transportation costs, which are treated as an exogenous variable. Therefore, the emphasis of these theories is on the influence of transportation costs on land use patterns.

Influence of Land Use on Transportation

To address the influence of land use patterns on trans portation, the most widely used approach is the four step travel demand models, which include a trip generation step to estimate the total number of trips generated by each zone, a trip distribution step to estimate the number of trips between each pair of origin and destination zones, a modal split step to estimate the market shares of dif ferent travel modes, and a trip assignment step to esti mate the traffic flows on individual highway segments or transit route segments. The main purpose of a four step travel demand model is to forecast traffic flows in a transportation system under a given land use pattern. Land use, therefore, is treated as an exogenous variable. Various theories and models have been used over the years to implement the four step travel demand models. Earlier travel demand models used approaches such as simple growth factors, gravity models, regression meth ods, and modal split curves that focused on the prediction of traffic flows rather than on the understanding of travel decisions. In the 1970s, behavioral approaches such as the random utility theory and discrete choice models were introduced into travel demand models. Random utility theory employs a utility function with a random variable to assess the utility of competing choice alternatives. The random variable is included to reflect factors such as imperfect knowledge and uncertainties involved in decision making processes. Random utility theory has been widely adopted in discrete choice models to determine an individual's choice among competing alternatives. Logit model is the most frequently used mathematic form to implement the random utility theory in a discrete choice model. It should be noted that a conventional travel demand model focuses on the impact of land use on transportation and usually does not con sider the feedback effects of transportation system changes on land use.

Integrated Transportation and Land Use Approach

A number of integrated transportation and land use models have been proposed over the years to address the shortcoming of modeling only one half of the circular relationship between transportation and land use illus trated in **Figure 1**. The Lowry model, which was originally developed to model spatial distributions of residential and service development, was among the first operational models to implement the transportation and land use feedback cycle. The Lowry model employs the economic base theory to simulate the interactions among the basic sector (i.e., export oriented employment), the nonbasic sector (i.e., employment to meet the local demands), and the residential sector. It starts with an exogenously given spatial distribution of the basic sector employment within a region. Employees of the basic sector are allocated to residential zones in the region according to a gravity model type of spatial interaction model. It then computes the incremental residential population and the additional nonbasic employment needed to serve the incremental residential population. The increased nonbasic employment in turn is allocated to different workplace zones. Workers in the nonbasic sector also require places to reside and will generate additional nonbasic employment. An iterative procedure is implemented to model this circular and cumulative process. The procedure stops when the increments of employment and population between two consecutive iterations become insignificant.

The Lowry model has stimulated many other studies on the subject. In general, the theoretical approaches employed in these studies of transportation and land use interaction can be categorized into technological approach, economic approach, social science approach, and behavioral approach. The technological approach suggests that the growth or decline of a place is related to its locational advantages in the transportation systems. Locations with a higher accessibility level are more likely to be chosen for land development. As transportation technologies evolved over time, accessibility at different places changed accordingly. The history of rising and declining cities during the waterway dominant era, the railroad dominant era, and the highway dominant era supports this technological approach. In more recent decades, air transportation and ICTs have played an increasingly important role in shaping the accessibility levels. Janelle's time–space convergence concept offers a framework of understanding relative changes of accessibility levels at different places due to technological advancements. In practice, spatial interaction models are frequently used to estimate accessibility levels at different locations. Although we can calibrate a spatial interaction model to match an observed flow pattern and then use the calibrated model to forecast future changes, spatial interaction models fall short of providing a theoretical basis of explaining the change process.

Agricultural, residential, employment, and industrial location choice theories discussed above fall under the economic approach. They are based on an assumption that households and firms choose their locations through trade offs between accessibility and land price. This assumption is closely related to the microeconomic concept of utility maximization. Alonso's bid rent theory, which assumes that firms and households choose the location where the land price they are willing to pay (i.e., bid rent) equals the rent asked by the landlord, serves as the foundation of many land choice models. It is assumed that interactions between supply and demand lead to a market equilibrium. If we take into account other factors such as internal economies of scale, economies of agglomeration, labor force, and the shift from manufacturing based economy to service based economy, location choice decisions become much more complex than simple trade offs between accessibility and land price. Although mathematical models based on the concept of equilibrium provide a useful method of balancing supply and demand in order to forecast future travel demands and future land use patterns, some researchers have argued that urban areas are constantly adjusting themselves to the changing environments. As a result, urban areas may be never in equilibrium.

There are also social science approaches to studying transportation and land use interaction. John Adams examined cities in the Midwest of the United States and identified the changing spatial form of urban areas in relation to changes in transportation technology. He reported that spatial form of urban areas during the walking horse era (1800–90) and the recreational automobile era (1920–50) were characterized by compact and concentric form. However, urban areas developed into radial or star shaped spatial form during the electric streetcar era (1890–1920) and the freeway era (1950–present) due to developments along the radial transportation routes. Concentric urban form was first observed by sociologist Ernest Burgess in the 1920s. In the 1930s, economist Homer Hoyt reported a sectoral urban form as the outcome of a residential filtering process between different income groups. Geographers Chauncy Harris and Edward Ullman in the 1940s identified a multinucleated urban form that has multiple competing centers. Another important theoretical contribution under the social science approach is time geography developed by Torsten Hägerstrand in the 1960s. Time geography examines human activities under different types of constraints in a space–time context. It emphasizes that both space and time should be considered when examining an individual's activity participation and accessibility. Space and time are connected through the concept of space–time path, which depicts the sequence of an individual's activities at various locations over a time period (see **Figure 2a**). Transportation is used as the means to trade time for space since movements in physical space take time. The possible locations that an individual can reach within a given time window forms a continuous space known as a space–time prism (see **Figure 2b**). With the growing interests in activity based travel demand modeling and the effects of ICTs on activity and travel patterns, time geography has received increasing attention in recent years as an alternative approach to studying transportation and land use relationships. For example, there have been suggestions to

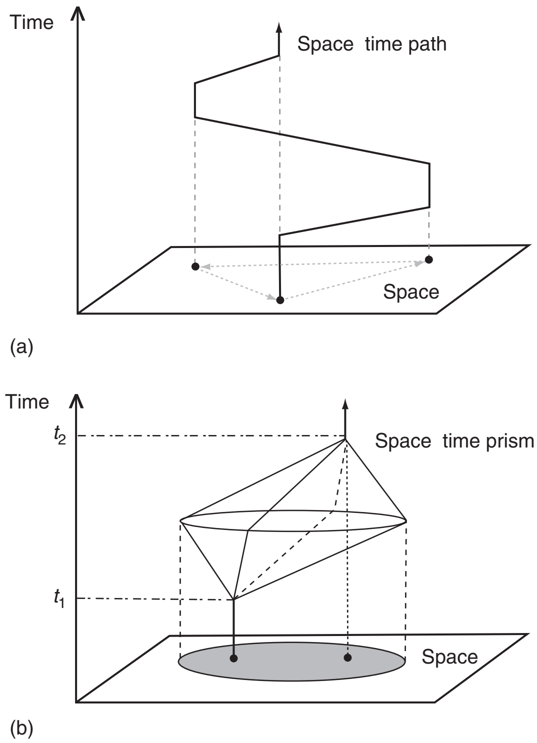

Figure 2 Time-geographic concepts of (a) space time path and (b) space time prism.

replace gravity model based accessibility measures by space–time accessibility measures based on the time geographic concepts.

The behavioral approach, on the other hand, focuses on the understanding of decision processes underlying choice behavior rather than on the prediction of outcomes of choice behavior. Most operational integrated transportation and land use models are implemented as mathematic models or computer simulation models to forecast future land use and travel demand patterns. The behavioral approach seeks new concepts, theories, data collection methods, and analysis frameworks to enhance the behavioral foundations of these models. It argues that the aggregate patterns of traffic flows and urban growth are the results of individual choice decisions. It is therefore important to gain a better understanding of choice decision processes at the disaggregate level. For example, the utility maximization theory may be limited in representing behavioral processes because people also use other decision rules (e.g., a threshold rule rather than a maximization rule) to reach their decisions. In addition, utility assessment can be dynamic and dependent on the context within which a decision is made. A particular activity considered by an individual is often constrained by other activities of the individual as well as the environment (e.g., available activity time, available travel means, open hours of activity locations, etc.) when the individual makes decisions on if, when, where, and how this activity will be carried out. Furthermore, reduced travel and telecommunication costs and additional flexibilities gained from the use of the Internet and cellular phones have expanded our activity opportunities both spatially and temporally. How these changes affect the location choices of individuals and businesses remain an open research question.

Lessons Learned and Future Prospects

We can expect that transportation and land use will remain a major policy issue as long as we continue to desire that transportation and land use policies can help us achieve a more sustainable environment. Previous studies have taught us that it may be more challenging than we think to untangle the complex and dynamic relationships between transportation and land use in order to identify effective policies leading to sustainable development.

First of all, we have learned that the automobile dominant transportation system is not sustainable. The question then is how we can assess different policies in terms of their potential of leading to sustainable development. Although sustainable development presents a good concept, it is not always consistent with other goals of developing a vibrant urban system. As population and economy in an urban area grow, we need to provide transportation systems that will meet the increasing demand. One solution is to increase the transportation supply with more efficient transportation systems that offer faster speed and better accessibility. Unfortunately, we have not been very successful in finding transportation systems that are efficient, safe, environmentally friendly, socially equitable, and in the meantime attractive to most individuals and firms. Another alternative is to come up with sustainable land use systems that will reduce the need for travel. However, successful stories of this nature are also rare. Given that sustainable development and economic development often have conflicting goals, criteria used to judge good transportation and land use policies depend on the perspectives of different stakeholders and may change over time.

Another important lesson we have learned is that the mutual interaction between transportation and land use cannot be isolated from many other related processes that constantly change and evolve in a dynamic and complex urban system. Contemporary urban environments are very different from the cities in the past. Most urban areas today have multiple centers along with decentralized residential and economic activities. In addition, they are increasingly dominated by service oriented and information based economy that changes the nature of accessibility and creates a more mobile and dynamic environment to both firms and individuals. As suggested by Genevieve Giuliano, with the many factors affecting the location and travel choices of individuals and firms, it is difficult to identify and almost impossible to predict the interaction between transportation and land use in contemporary urban areas.

We also have learned from empirical studies that efforts of implementing various transportation and land use policies to achieve sustainable development have resulted in mixed outcomes. In general, transportation policies that make private automobile travel less at tractive appear to be more effective than land use pol icies that attempt to reduce the need for travel through approaches such as mixed land use development. In the meantime, we need land use policies to create a less automobile dependent urban form in the long run. Overall, our understanding of transportation and land use interaction is still very limited. We need more comparative empirical studies that employ consistent methodology and comparable data to systematically study this subject.

With regards to operational integrated transportation and land use models, many of them have been developed with various theoretical and modeling approaches to simulating transportation and land use interaction. Although significant progress has been made to enhance these models, there are many aspects of these models that can benefit from further improvements. For example, recent developments of location aware technologies such as the Global Positioning System (GPS) and information technologies such as geographic information systems (GIS) now provide more efficient methods to collect and analyze detailed location data at the individual level. Coupled with research progress on activity based travel studies and time geographic studies, there is a good potential of gaining additional insight of individual location and travel choice behaviors. In addition, ad vancements in microsimulation techniques, along with the availability of high performance computers, make it feasible to represent and model many agents and processes involved in the complex and dynamic system of transportation and land use interaction that can help us assess various policy scenarios.

See also: Central Place Theory; Land Rent Theory; Transport and Accessibility; Transport and Sustainability; Transport, Urban.

Further Reading

Adams, J. S. (1970). Residential structure of midwestern cities. *Annals of the Association of American Geographers* 60, 37 62.

Alonso, W. (1964). *Location and Land Use*. Cambridge, MA: Harvard University Press.

Boarnet, M. G. and Crane, R. (2001). *Travel by Design: The Influence of Urban Form on Travel*. New York: Oxford University Press.

Chan, Y. (2005). *Location, Transport, and Land Use: Modelling Spatial Temporal Information*. Berlin: Springer.

Christaller, W. (1966). *Central Places in Southern Germany* (translated by Baskin, C. W.). Englewood Cliffs, NJ: Prentice Hall.

Giuliano, G. (2004). Land use impacts of transportation investments: Highway and transit. In Hanson, S. & Giuliano, G. (eds.) *The Geography of Urban Transportation* (3rd edn.), pp 237 273. New York: The Guilford Press.

Hagerstrand, T. (1970). *What about people in regional science? Papers of the Regional Science Association* 24, 1 12.

Harris, C. D. and Ullman, E. L. (1945). The nature of cities. *Annals of the American Academy of Political and Social Science* 242, 7 17.

Hoyt, H. (1939). *The Structure and Growth of Residential Neighborhoods in American Cities*. Washington, DC: US Government Printing Office.

Janelle, D. G. (2004). Impact of information technologies. In Hanson, S. & Giuliano, G. (eds.) *The Geography of Urban Transportation* (3rd edn.), pp 86 112. New York: The Guilford Press.

Lee Gosselin, M. and Doherty, S. (eds.) (2005). *Integrated Land Use and Transportation Models: Behavioural Foundations*. London: Elsevier Ltd.

Miller, E. J. (2004). Integrated land use/transport model requirements. In Hensher, D. A., Button, K. J., Haynes, K. E. & Stopher, P. (eds.) *Handbook of Transport Geography and Spatial Systems*, pp 147 165. London: Elsevier Ltd.

Park, P. E. and Burgess, E. W. (1925). *The City: Suggestions for Investigation of Human Behavior in the Urban Environment*. Chicago: University of Chicago Press.

Paulley, N. and Pedler, A. (2000). *TRANSLAND: Integration of transport and land use planning*. Deliverable 4 of the project TRANSLAND of the 4th RTD Framework Programme of the European Commission. Transport Research Laboratory, Berkshire, England.

Putman, S. H. (1983). *Integrated Urban Models: Policy Analysis of Transportation and Land Use*. London: Pion Limited.

Weber, A. (1929). *Theory of the Location Industries* (translated by Friedrich, C. J.). Chicago: University of Chicago Press.

Wegner, M. (2004). Overview of land use transport models. In Hensher, D. A., Button, K. J., Haynes, K. E. & Stopher, P. (eds.) *Handbook of Transport Geography and Spatial Systems*, pp 127 146. London: Elsevier Ltd.

Wegner, M. and Furst, F. (1999). *Land use transport interaction: State of the art*. Deliverable 2a of the project TRANSLAND of the 4th RTD Framework Programme of the European Commission. Institute fur Raumplanung, Univeristat Dortmund.

Relevant Websites

http://www.fhwa.dot.gov/pubstats.html
U.S. Department of Transportation, Federal Highway Administration's Electronic Reading Room.

http://www.urbansim.org/
UrbanSim is a simulation model that incorporates the interactions between transportation, land use, and public policy for integrated planning and analysis of urban development.

http://www.vtpi.org/
Victoria Transport Policy Institute offers an online Transportation Demand Management (TDM) Encyclopedia with information about "Land Use Impacts on Transport".

Travel and Travel-Writing

R. Phillips, University of Liverpool, Liverpool, UK

Glossary

Royal Geographical Society (RGS) Founded in 1830, the London-based RGS supported and provided a forum for explorers and map-makers in the nineteenth and early twentieth centuries, and remains Britain's leading geographical society.

Travel, defined in Waite's *Little Oxford Dictionary* as to 'go from one place to another; make journey(s), esp. long or abroad', has been applied to a wide variety of experi ences: daily and periodic journeys to work, shopping, pleasure trips, pilgrimages, voyages of discovery, and more. Travel writing is an equally inclusive and expan sive category, as the burgeoning travel supplements in weekend newspapers and travel shelves of bookshops testify, with their numerous accounts of and guides to travel.

This entry explains why geographers are interested in travel and its literature, before defining them with ex amples from well known travel writers including Jan Morris, and some more obscure travellers such as Vic torian women emigrants. It then reviews the geographical literature on the political and theoretical significance of travel and travel writing, with reference to the medium's special significance for constructions of gender, imperi alism, and finally class (**Figure 1**).

Why Geographers Are Interested in Travel-Writing

There was a time when academic geographers preferred to leave travel and its literature to others, concentrating upon more specific and ostensibly more serious forms of movement and writing. In nineteenth century Britain, for example, geographers did not 'travel'; they mounted 'expeditions'. If they lacked the funds or the energy to do so themselves, they studied the findings of others, a select group they referred to as 'explorers'. The Royal Geo graphical Society (RGS) funded expeditions, provided explorers with a forum in which to present findings, both orally (in the RGS lecture theater) and in printed form (in its journals, and in the contents of its reading room and map library), and awarded medals to those whose work and methods it endorsed, as shown by Livingstone in 1992. In the twentieth century, when the age of ex ploration was largely over, academic and institutional geographers devoted themselves to smaller expeditions, detailed survey, and cartographic work, and also to for mal regional descriptions and analyses; they continued to distance themselves from the ostensibly unscientific practice of travel and (worse still) tourism, and the popular geographical medium of travel writing.

All this has changed, for two main reasons. First, it is now recognized that this narrow approach to the practice and literature of geography was exclusionary. Most ex plorers, surveyors, and cartographers were white men, funded and recognized by other white men, within the geographical and social parameters of metropolitan in stitutions such as the RGS. Women were formally ex cluded from the RGS and were expected to confine themselves to forms of writing that were considered both more feminine and 'lower' than those of their formally geographical counterparts. Felix Driver (1995: 411) argues that "opening up a space for excluded voices and counter traditions – extending the realm of 'legitimate' geographical knowledge" is "a necessary task for any critical history of geography." The voices to which Driver alludes are not just those of women, but others excluded from the narrow frame of establishment geographies, past and present, for instance by their race, ethnicity, na tionality, religion, social class, sexuality, or age. Serious attention to travel and travel writing has emerged from a critique of the discipline, and also from broader attempts to forge more critical and inclusive geographies.

Second, the recent geographical 'discovery' of travel and travel writing was also driven by the cultural turn within the discipline in the 1990s, which focused atten tion on representation in written and more generally textual forms. Writing, as is increasingly recognized, is central to the conception of geography. Derived from the Greek *geo*, meaning earth, and *graphien*, meaning 'to write', geography is etymologically 'earth writing', ac cording to Barnes and Duncan. Questions about words and their place in the world have increasingly been raised and discussed in human geography. In the early 1990s, Trevor Barnes and James Duncan (1992: 1) could ac curately complain that "very little attention is paid to writing in human geography." This is no longer the case: in recent years geographers have become increasingly conscious of the way in which they write – the form and significance of geographical writing – and have also paid a great deal of attention not only to textual sources but also to the textuality of geography. Moreover, as this article will explain in more detail, it is recognized that geographical writing in general and travel writing in particular is not simply a neutral process of description.

Figure 1 Canadian Pacific Steamship poster, 1963. Source: Choko, M. H. and Jones, D. L. (1988). *Canadian Pacific Posters 1888 1963*. Montreal: Meridian Press (1988, 75).

Introducing a book in 1999 about geographies of travel writing, James Duncan and Derek Gregory (1999: 2) stressed that travel is a politicized medium: as they put it, "travel and its cultural practices have been located within larger formations in which the inscriptions of power and privilege are made clearly visible."

Travel and Travel-Writing Defined

Travel is best defined, not by dictionaries or literary critics, but by the protagonists and authors of travel books and articles, including contemporary and recent figures like Jan Morris, Bill Bryson, and Bruce Chatwin, and their historical counterparts such as Mary Kinglsey, Richard Burton, and Isabella Bird. In this section, travel is introduced with illustrations from the work of Jan Morris, before the (rather narrow) definition this affords is complicated and broadened, this time with illustrations from more obscure travellers: women emigrants and male tramps. This explains some of the characteristics of travel writing, and also fleshes out the contrast between this and other, more formal and conventionally scholarly geographical writing.

Travel writers such as Jan Morris have a number of things in common. First, they move and write in the first person, playing down the company and practical assistance provided by cooks, porters, guides, and others (**Figure 2**). Duncan and Gregory acknowledge the popularity of "*personal*, avowedly imaginative accounts of travel" (added emphasis, Duncan and Gregory 1999: 1). These lone travellers are modern, diluted descendents of a "Romantic view of travel," which "positions the traveller as the subject of urgent desires, impulses and inner needs" (Chard 1999: 213), always distancing the traveller from the herd, perhaps in order to intensify the experience, or at least to respect literary convention.

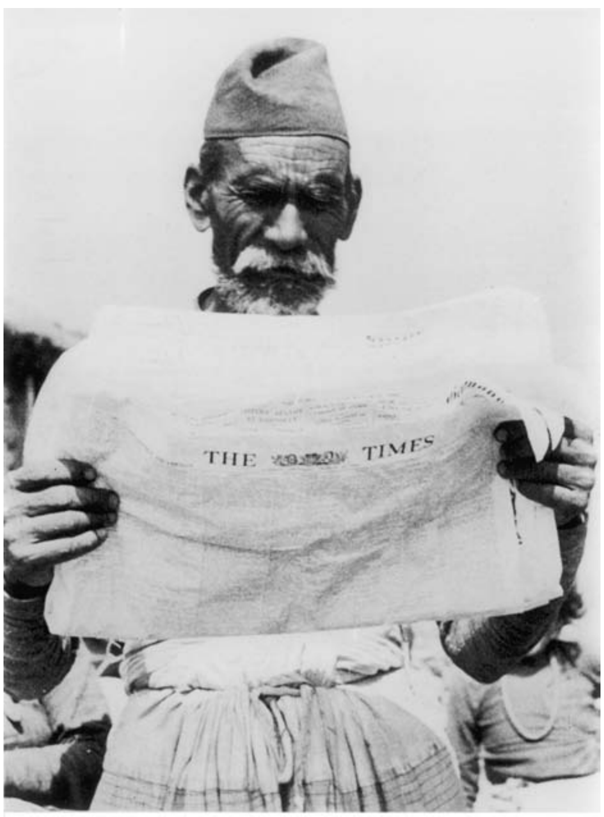

READING 'THE TIMES'

Figure 2 Travelers rely for assistance and guidance on local people such as this man, who accompanied the Everest Expedition in 1953. He is reading a copy of the British newspaper in which Morris' travel journal first appeared. These dispatches were compiled in book-length form in *Coronation Everest* (1958). Source: Morris, J. (1958). *Coronation Everest, an Account by the Special Correspondent of the Times of the Everest Expedition, 1953*. London: Faber.

Second, Morris' travel articles and books – with titles such as *O! Canada* (1992), *City to City* (1990), and *Coast to Coast* (1962) – all describe voluntary travel. The expression of free will, curiosity, and a degree of privilege, journeys are not lavishly funded but money is never a worry, and neither generally are passports or visas. The picture of travellers as a kind of peripatetic middle class is reinforced in much of the literature on the subject, for example Cheryl McEwan's 2000 study of women travellers in West Africa. McEwan pigeonholes the various travellers in terms of their class, and though not all are labelled middle class – some are aristocratic and others, such as Mary Slessor, the missionary in southern Nigeria, working class – the overall impression is that of people who, their basic needs met and often far exceeded, are at liberty to travel for pleasure or to pursue an interest or project. These travellers are driven, but not by necessity. Alison Blunt puts this in more nuanced terms in 1994 in her account of Mary Kingsley's travels in West Africa, arguing that Kingsley occupied and indeed forged an ambivalent social position – on the one hand metropolitan, white, and upper middle class, a privilege that freed her to travel; on the other, feminized and excluded from exclusively masculine, higher status geographical practices such as exploration and scientific writing. In this hierarchy, travel and travel writing are accorded lower status than some more formal geographical practices, but higher status than others. In particular, travel is more prestigious and exclusionary than, say, tourism, which is sometimes associated with, as Chard (1999: 216) puts it, "trivial, pointless and innocent gratification."

Travel writing has a similar mixed status in literary criticism. Typically, Morris' books have done well in the bookshops, if not always with the critics. The literary establishment did once nominate her dystopian novel, *Last Letters From Hav*, for a Booker Prize (a prestigious literary award), but she has generally been relegated to the popular genre, to the kind of writing that is not entirely neglected by academic researchers but is nevertheless beyond the scope of most literary and, until recently, geographical criticism. By beginning to take seriously the travel and travel writing of figures such as Morris, geographers have begun to throw new light on popular geographies – practices and texts – which respond to, represent, and actively shape geographical worlds. Before considering what these studies have shown, it is necessary to broaden the way in which travel and travel writing are understood.

Towards a Broader Understanding of Travel

The preliminary understanding of travel and travel writing, provided by figures such as Mary Kingsley and Jan Morris, must now be complicated and broadened to take account of a wider range of people and practices.

First, as acknowledged in the previous section 'Travel and travel writing defined', travel is not always as solitary as it generally appears in popular travel books. Anthropologists James Clifford and Caren Kaplan developed the idea of collective travel, Clifford extending the term to include migration, immigration, and homelessness, Kaplan writing of travelling as a pervasive and more or less universal social condition. "For many of us," she argues, "there is no possibility of staying at home in the conventional sense – that is, the world has changed to the point that these domestic, national, or marked spaces no longer exist" (Kaplan, 1996: 7). As a condition, travel is not always or necessarily concerned with temporary movement, with merely visiting other people and places, as it appears in Morris' accounts. Rather, it is a way of life, one which is increasingly common but not entirely new. In the past, for instance in the British Empire, travel was a way of life for many different people, from the English colonial administrators whose careers took them between different colonies to the Kru sailors who divided their shore leave between Freetown and Liverpool. Today, according to Kaplan, travel is the apparently permanent condition of many more.

As a social condition rather than an individual choice, travel need not be quite as voluntary as popular travel books imply. In contrast with Mary Kingsley, whose wealth and status enabled her to travel in West Africa for pleasure and fulfillment, many of her less privileged countrywomen had practical reasons for travel. Rather than escaping from the responsibilities and banalities of home, they were travelling in the hope of securing positions – mostly with long hours and little pay – as domestic servants. Their experiences are harder to trace than those of popular authors, but archives of female emigration societies and of organizations such as the Travellers' Aid Society (TAS) do present some valuable records of travel by ordinary people, in the form of reports sent in by TAS workers and volunteers and, less commonly, letters sent in by female travellers and their guardians. Case histories, reported in the Society's annual reports, acknowledged that most of its clients were young women in desperate search of work.

Tim Cresswell develops the idea of involuntary travel, driven not by privilege or pleasure but necessity, in his 2001 study of American tramps. Tramps, he argues, were an underclass kept – by the hostility they met in particular places and by their constant attempts to scavenge a livelihood – in a state of motion. Cresswell argues that the American fascination with and sometimes hostility towards tramps speaks of a cultural ambivalence towards travel and mobility more generally. On the one hand, the mobility of tramps was seen as a threat to "respectable society"; on the other, it "slotted quite nicely into … American national identity" (Cresswell, 2001: 14). In Britain and Ireland, according to David Sibley, there is less ambivalence about 'travellers' – those formerly known as 'gypsies' – whose mobility is widely regarded as a threat to those with more settled ways. Other forced travellers include slaves, refugees, asylum seekers, and others in exile of one kind or another, as mentioned in the writings of Bammer in 1994, and Broe and Ingram in 1989.

Broadening still further the idea of travel, it is possible to think of nonhuman forms of travel: the movement of ideas and objects. Edward Said (1984: 226) argued thus: "Like people and schools of criticism, ideas and theories travel – from person to person, from situation to situation, from one period to another." This process may take tangible human form, as for example in Carl Sauer's 1952 analysis of the dispersal of cultural traits from origins or "cultural hearths." But the movement of ideas and objects is distinct from that of people, and is organized not around human subjects but rather transport and communications technologies. John Law in 1986 developed this point in a study of the flows of information and goods – over long distances and through networks – that formed the basis for imperialism and arguably proto globalization in the early modern period. The idea of nonhuman movements and interactions, driven and structured by technologically defined and materially substantive networks, was taken further by Bruno Latour in 1987, whose 'Actor Network Theory' examined the interrelations between human and nonhuman agents within broad processes of sociospatial interaction. A full review of this body of theory is beyond the scope of this article, but its brief mention does usefully complicate an understanding of what travel and travel writing are, while at the same time underlining the politicization of travel – the subject of the next section titled 'Politics of travel'.

Politics of Travel

This section reviews geographical literature on the political and theoretical significance of travel and travel writing, with reference to the medium's special significance for constructions of gender, imperialism, and finally class.

Though it makes sense to discuss travel and travel writing in the same general context, the two should be carefully distinguished. For Duncan and Gregory (1999: 5), they are necessarily connected: travel writing is the trace of "corporeal subjects moving through material landscapes." But, as the eighteenth century English journalist, travel writer, and novelist Daniel Defoe once commented, the best travellers do not always make the best travel writers, and vice versa (**Figure 3**). Indeed, some of the most successful published works of travel are either free with the experiences they ostensibly chronicle, or are entirely invented. For instance, Defoe's own bestselling travel journal, *A New Voyage Round the World by a Course Never Sailed Before* (1725), was a work of fiction. Duncan and Gregory are right to assert the fundamental materiality of travel writing, however, because whether nonfiction or fiction, it responds in some way to, and in turn shapes, the material world.

Like other forms of geographical representation, travel writing should be understood not as the mere depiction of the world, but as part of it . This means that, as Edward Said argued in 1978, geographical representations are not simply copies or misrepresentations of some "original"; more actively, they are always for something, and against or instead of something else. For instance, European representations of the Orient, according to him, should not be seen as the "misrepresentation of some Oriental essence," but rather as purposive formations, which intervened in "specific historical, intellectual, and even economic" settings (Said, 1978: 273). This fundamentally political way of thinking about representation informs geographical and other recent interest in travel writing. Consequently, this section concentrates not on the formal or aesthetic qualities of travel writing but rather its material and political significance. It concentrates, as Duncan and Gregory (1999: 2) put it, on locating "travel and its cultural practices … within larger formations in which the inscriptions of power and privilege are made clearly visible."

For instance, as explained earlier, travel was and is an important medium for the inscription of gender. Victorian women, excluded from other geographical practices and media, were permitted to travel and write travel books, though they were expected to do so in certain ways, which were variously spelled out and hinted at in a mixture of guidebooks, women's emigration societies, unwritten social codes, and so on. Travel and travel writing were therefore important for the reproduction of constructions of femininity and masculinity. But they also opened up space in which to contest and shift constructions of gender. Mary Kingsley's efforts to respect conventions of feminine style were recognized by some critics, such as the newspaper reviewer of *West African Studies* who praised the "fresh, vigorous, breezy style," which befitted "a lady … so light hearted and gay" (*Nottingham Daily Guardian* 1899, quoted by Blunt, 1994: 121). And yet, Kingsley negotiated and pushed the boundaries of contemporary femininity where she judged it possible to do so. She had gone to parts of West Africa that had not been visited or mapped by other Europeans, and she had maintained a seriousness of purpose that contradicted the apparent frivolity of her style, making scientific observations and trading with Africans as she went. Blunt concludes that Kingsley's travels and travel writing contested and reshaped conventions of gender, both in the colonial contexts of travel, and also closer to home.

Other travel writers, in contrast, have used the medium of travel writing to contest contemporary constructions of sexuality. For example, in 1885–86, Richard Burton used travel books and essays (as well as translations) to contest Victorian prudery and argued that men who had sex with men "deserve, not prosecution but the pitiful care of the physician and the study of the psychologist" (Burton, 1885–86: x.209). Traversing a series of sexual cultures, accumulating a picture of diversity, he used travel writing to assemble a case against the moral universalism of his time, and for a more differentiated and sympathetic sexual culture.

Kingsley and Burton both travelled in an age of empire and were, in different ways, imperial travellers. Travel writing could, as Said (1978: 219) put it, be used to chart "geographical space to be cultivated, harvested, and guarded" by themselves. The relationships between travel and imperialism were and are varied and complex, sometimes subtle but sometimes – as, for example, in the work of Richard Burton – direct and crude. Burton was an outspoken advocate of racially structured imperialism. He used travel books about West Africa to argue, for example, that "the Negro" would be best "held to labour" by "better and wiser men than himself" (Burton, 1864: ii.204). On the other hand, the colonial subjects of travel texts were often presented more positively, as, for example, in the Canadian Pacific Steamship poster reproduced in **Figure 4**, though these representations could still be imperialist, legitimating colonial relationships.

Imperial travel writing, in its many forms, is not confined to the past. Derek Gregory traces echoes of Orientalist texts – their colonial tropes and imaginative geographies – in the foreign policy discourse of the United States and its allies in relation to Afghanistan, Palestine, and Iraq. He shows, for instance, how Western travellers in Egypt continue to see their own journeys through the lens of other books, written by past generations, which they take with them and read along the way. This establishes continuity between travel books and travel experiences, blurring boundaries between imaginative and real travel, textual and material geographies. So, even though writers often try to distance themselves from their colonial predecessors, the latter remain important points of reference. Gregory (2004: 18–19) argues that the "categories, codes, and conventions" of contemporary Orientalism "shape the practices of those who draw upon it, actively constituting its object in such a way that this structure is as much a *repertoire* as it is

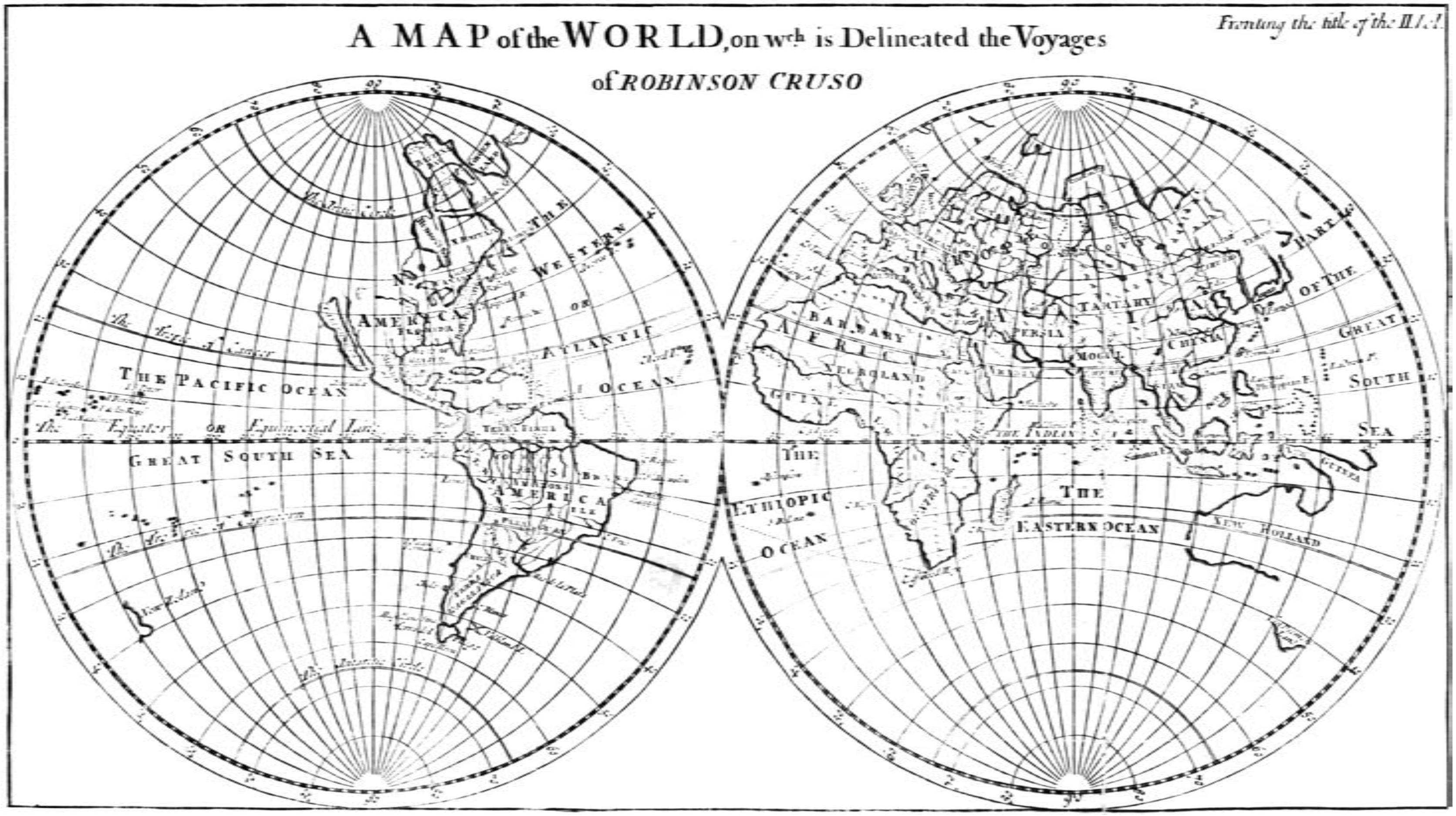

Figure 3 Trave -wr t ng as fict on. A Map of the Wor d on wh ch s De neated the Voyages of Rob nson Crusoe.' Rob nson Crusoe was a fict ona trave book, though not ent re y an mag nary one, s nce t drew upon the nonfict on narrat ve of a trave er named A exander Se k rk, and ts geograph ca sett ngs and h stor ca context were c ose y researched.

Figure 4 A Canadian Pacific Steamship poster advertising travel to the Orient. This poster depicts European civilization in the form of a gleaming modern steamship, contrasted with Oriental exoticism, sensuality, and primitivism, in the form of the junks and the boatman in the foreground. This image of travel animates an imaginative geography neatly divided between modern and primitive, colonizer and colonized. Source: Choko, M. H. and Jones, D. L. (1988). *Canadian Pacific Posters 1888 1963.* Montreal: Meridian Press.

an archive." This repertoire has surfaced, for example, in Presidential speeches justifying war in Iraq, and a range of commentaries on Islam in the context of the 11 Sep tember attacks and the 'war on terror' in Afghanistan and Iraq. Addressing a joint session of Congress on 11 Sep tember 1990, on the crisis in the Gulf following the Iraqi invasion of Kuwait, George Bush Senior stated that: "Out of these troubled times, a new world order can emerge," one in which "the rule of law supplants the law of the jungle." Gregory argues that Bush's speech revived a colonial discourse of "wild untamed spaces" in which "civilisation" was menaced by a reversion to "barbarism." Though this kind of imaginative geography and, more specifically, through the inertia of its language, the past weighs heavily upon the present, and travel's "inscrip tions of power" (Duncan and Gregory, 1999: 2) are reified.

Conclusion

This article began by explaining that geographers have paid increasing attention to travel writing because, in so doing, they are able to open up what Driver (1995: 411) calls "space for excluded voices and counter traditions" and also extend the realm of "legitimate" geographical knowledge. But this brief review of gender and imperi alism in travel writing warns that this medium should not be identified with any particular politics. Travel, not necessarily the preserve of the excluded or the dissident, can also be a hegemonic medium. Moreover, the politics of travel writing and any given travel writer or text are debatable; their inscriptions of power are never more than claims to power. Their words, like anyone else's, may be ignored, contested, twisted, and interpreted. Burton was taken to task and contradicted by some of the people he portrayed. William Rainy, the articulate Creole barrister who once successfully prosecuted him when he was visiting Freetown, later attacked the Englishman's representation of Sierra Leone. In an 1865 pamphlet, which he dedicated "To the African People," (Rainy, 1865: 3). Rainy dismissed Burton's travel writing. De nouncing it as "a vile national slander," Rainy (1865: 5) made it clear that a travel writer could not simply say what he or she wanted and get away with it, or expect people to believe and act on what was said.

This discussion of travel writing begs the question: what if anything is distinctive about the medium, whether as a form of geographical writing or a means of political intervention, or both. This question is, at very least, complicated by the eclectic and contested nature of travel and travel writing as categories – in which one person's travel is another's displacement, one person's travel journal another's exploration narrative. Still, different travel nar ratives do generally have one thing in common: to recall the Dictionary definition of travel, they describe journeys as "from one place to another." At worst, as Rainy argued of Burton, the brevity of travel breeds a superficial and ill informed engagement with place. He dismissed Burton's travel journal as the quick judgements of an "ignorant" and "superficial observer" who in a matter of "moments" had arrived at unsound "conclusions upon the morals, habits, and social qualities of nations" (Rainy, 1865: 5). At best, however, travel brings a freshness of perspective to places, which it also positions within wider frameworks of com parison and connection. A final example of travel writing illustrates this point, and with it the power of the medium.

Friedrich Engels visited Manchester and the neigh boring city of Salford in 1844, and described his obser vations in a travel journal. His was very much the description of an outsider, to whom the pungent en vironmental and social condition of mid nineteenth century northern England remained unfamiliar, and re tained the power to shock.

At the bottom the Irk flows, or rather stagnates. It is a narrow, coal black stinking river full of filth and garbage which it deposits on the lower lying right bank. In dry weather, an extended series of the most revolting blackish green pools of slime remain standing on this bank, out of whose depths bubbles of miasmatic gases constantly rise and give forth a stench that is unbearable even on this bridge forty or fifty feet above the level of the water.' (Engels, 1958,1845: 60)

As a visitor from Germany he had a broader geo graphical perspective, which enabled him to locate this part of northwest England within a wider geographical, eco nomic, and political frame. Describing a particular place, which he compared and contrasted with Birmingham and other towns and cities he had recently passed through, Engels conveyed the sense that a new form of industrial society was emerging, a new set of relations between property owners and workers. With this, as his vivid geo graphical descriptions made clear, came new levels of social and environmental degradation, new forms of alienation and exploitation. What he saw was not only new to him, but also relatively new to the world, and it was precisely in his efforts to represent that newness that Engels groped towards an understanding of capitalism, which he would develop with Karl Marx in more general and theoretical terms. Engels' travel narrative bore out the most funda mental promise of geographical description, which is for "empirical" writing to assume "heightened theoretical significance" (Marcus and Fisher, 1986: 15). It showed that, at best, travel writing need not simply depict places or passages through them, but that it may open new avenues for understanding and political intervention.

See also: Colonialism I; Colonialism II; Literature.

Further Reading

Bammer, A. (ed.) (1994). *Displacements: Cultural Identities in Question*. Bloomington, IN: Indiana University Press.

Barnes, T. and Duncan, J. S. (eds.) (1992). *Writing Worlds: Discourse, Text and Metaphor in the Representation of Landscape*. London: Routledge.

Berman, M. (1982). *All That/Solid Melts Into Air: The Experience of Modernity*. New York: Simon and Schuster.

Broe, M. L. and Ingram, A. (1989). *Women's Writing in Exile*. Chapel Hill, NC: University of North Carolina Press.

Blunt, A. (1994). *Travel, Gender and Imperialism: Mary Kingsley and West Africa*. New York: Guilford.

Burton, R. F. (1864). *Mission to Gelele, King of Dahomey.* 2 vols.London: Tinsley.

Burton, R. F. (1885 1886). *The Book of the Thousand Nights and a Night: A Plain and Literal Translation of the Arabian Nights Entertainments*. 10 vols. Benares: Kama Shastra Society.

Chard, C. (1999). *Pleasure and Guilt on the Grand Tour: Travel Writing and Imaginative Geography*. Manchester: Manchester University Press.

Choko, M. H. and Jones, D. L. (1988). *Canadian Pacific Posters 1888 1963*. Montreal: Meridian Press.

Clifford, J. (1992). Travelling cultures. In Grossberg, L., Nelson, C. & Treichler, P. (eds.) *Cultural Studies*, pp 96 112. New York: Routledge.

Clifford, J. (1997). *Routes: Travel and Translation in the Late Twentieth Century*. Cambridge, MA: Harvard University Press.

Cresswell, T. (2001). *The Tramp in America*. London: Reaktion.

Defoe, D. (1719). *The Life and Strange Surprising Adventures of Robinson Crusoe* (4th edn). London: W. Taylor.

Domosh, M. (1991). Towards a feminist historiography of geography. *Transactions of the Institute of British Geographers* 16(1), 95 104.

Driver, F. (1995). Submerged identities: Familiar and unfamiliar histories. *Transactions of the Institute of British Geographers* 20(4), 410 413.

Duncan, J. and Gregory, D. (eds.) (1999). *Writes of Passage: Reading Travel Writing* London: Routledge.

Engels, F. (1958). *The Condition of the Working Class in England in 1844 (first publ. 1845)*. Oxford: Blackwell.

Frost, D. (1992). *The Kru in Freetown and Liverpool: A Study of Maritime Work and Community during the 19th and 20th Centuries*. Liverpool: Liverpool University Press.

Gregory, D. (2004). *The Colonial Present: Afghanistan, Palestine, Iraq*. Oxford: Blackwell.

Hudman, L. and Jackson, R. (1999). *Geography of Travel and Tourism*. New York: Delman.

Kaplan, C. (1996). *Questions of Travel: Postmodern Discourses of Displacement*. London: Duke University Press.

Lambert, D. and Howell, P. (2003). John Pope Hennessey and the translation of 'slavery' between late nineteenth century Barbados and Hong Kong. *History Workshop Journal* 55, 1 24.

Latour, B. (1987). *Science in Action*. Milton Keynes: Open University Press.

Law, J. (ed.) (1986). *Power, Action and Belief: A New Sociology of Knowledge*. London: Routledge.

Lefebvre, H. (1991). *The Production of Space*. (trans. Nicholson Smith, D.) Oxford: Blackwell.

Livingstone, D. N. (1992). *The Geographical Tradition: Episodes in the History of a Contested Enterprise*. Oxford: Blackwell.

McEwan, C. (2000). *Gender, Geography and Empire: Victorian Women Travellers in West Africa*. Aldershot: Ashgate.

McLynn, F. (1990). *Snow upon the Desert*. London: John Murray.

Marcus, G. E. and Fisher, M. M. J. (eds.) (1986). *Anthropology as Cultural Critique: An Experimental Moment in the Human Sciences*. Chicago, IL: University of Chicago Press.

Morris, J. (1958). *Coronation Everest, an Account by the Special Correspondent of the Times of the Everest Expedition, 1953*. London: Faber.

Morris, J. (1989). *Pleasures of a Tangled Life*. London: Barrie & Jenkins.

Ogborn, M. (2006). Mapping words. *New Formations* 57, 145 149.

Phillips, R. (1999). Writing travel and mapping sexuality: Richard Burton's Sotadic Zone. In Duncan, J. & Gregory, D. (eds.) *Writes of Passage: Reading Travel Writing*, pp 70 91. London: Routledge.

Phillips, R. (2001). Decolonizing geographies of travel: Reading James/ Jan Morris. *Social and Cultural Geography* 2, 5 24.

Phillips, R. (2006). 'Heterosexuality, respectability and the travellers' aid society. *Acme: An E Journal of Critical Geography* 5, 163 190.

Phillips, R. (2006). *Sex, Politics and Empire: A Postcolonial Geography*. Manchester: Manchester University Press.

Pratt, M. L. (1992). *Imperial Eyes: Travel Writing and Transculturation*. London: Routledge.

Rainy, W. (1865). *The Censor Censured; or, the Calumnies of Captain Burton on the Africans of Sierra Leone*. London: George Chalfont.

Rappaport, E. D. (2000). *Shopping for Pleasure: Women in the Making of London's West End*. Princeton, NJ: Princeton University Press.

Said, E. (1978). *Orientalism: Western Representations of the Orient*. New York: Pantheon.

Said, E. (1984). *Travelling Theory. The World, the Text and the Critic*, pp 226 247. London: Faber and Faber.

Sauer, C. O. (1952). *Agricultural Origins and Dispersals: The Domestication of Animals and Foodstuffs*. Cambridge, MA: MIT Press.

Sibley, D. (1995). *Geographies of Exclusion: Society and Difference in Society and Space*. London: Routledge.

Urry, J. (2002). *The Tourist Gaze*. London: Sage.

Waite, M. (1998). *Little Oxford Dictionary*. Oxford: Oxford University Press.

Walker, K. (ed.) (1964). *Love, War and Fancy: The Customs and Manners of the East from Writings of the Arabian Nights*. London: William Kimber.

Trend Surface Models

D. J. Unwin, University College London, London, UK

Glossary

Dependent Variable The unknown quantity to be estimated in a statistical analysis.
Eastings The distance east (along) from some origin point in a pair of map grid references.
Field In geographic information science used to denote a spatially continuous variable, such as air temperature for which everywhere has a value.
Georeference A description of the location of a point on the Earth's surface, such as a grid reference, latitude/longitude, or postal/ZIP code.
Independent Variable The known quantities used, often in conjunction with others, to estimate the values of some unknown dependent variable.
Interpolation, Spatial The process of threading lines of equal value, such as contours or more generally isolines through a field of sample numbers.
Isolines Lines of equal value of the height of a continuous field.
Least Squares Minimizing the sum of squares of the differences between model values and observed data, much used as a criterion in regression analysis.
Leverage In regression analysis, a measure of the influence of an individual datum on the overall result.
Northings The distance north (up) from some origin point in a pair of map grid references.
Polynomial In trend analysis a mathematical expression of the form $z \quad b_0 + b_1x + b_2y + \cdots + b_nx^my^m$.
Regression In statistics, a method of calibrating models using observed data.
Residual The difference between an observed value of the height of a field and the value predicted at the same point by a trend surface model.
Trend Component The value predicted at some point by a trend surface model.
Universal Kriging A method of spatial interpolation used when there is a change in the mean value across a field.

Introduction: Representing Scalar Fields

In circumstances when a spatially distributed variable either is, or can be assumed to be, continuous across space, such that it is possible to draw an isoline (contour type) map, we can express this using the notion of a mathematical function of the general form given in

$$z_i = f(x_i, y_i) \qquad [1]$$

This defines what is known as a scalar field, and relates surface height (z) at a location i, to its position in space specified by a pair of (x, y) coordinates. In the notation of eqn [1], f is simply some as yet unknown function. If we knew the form of the function, specification of a location in space by its (x, y) coordinates would be all that was needed to give a predicted value for the surface height at this point.

By way of illustration, suppose that the spatially distributed variable is altitude above sea level, and that the (x, y) are the northings and eastings grid coordinates of the base map. Although this uses the height of the land surface, it is important to realize that the z values could be almost any spatially distributed phenomenon that is continuous, for example, temperature, or rainfall. In human geography we frequently imagine that variables of interest are spatially continuous, as, for example, in the concepts of population density and potential. For ground height, the familiar contour map is a graphical way of specifying the linking function, f, but in geographical information systems fields are often specified and represented in other ways, for example, by regularly or irregularly distributed point samples, polygons, or a network of triangles. These methods and graphical constructions are used in preference to specification of some regular mathematical function for two reasons. First, almost always the function will be complex, incapable of being expressed by any simple mathematical form and, second, in many applications we want our representation of the field to ensure that heights are always correct. On a topographic map we expect the contours to be in their correct positions, such that the mapped surface 'honors' all the available height information. In these circumstances, the usual approach taken would involve some form of spatial interpolation to create a contour map representation of the field.

Trend Surfaces and Their Residuals

In contrast, a trend surface deliberately generalizes such a field into its major feature, or trend, defined as any systematic change that extends smoothly and predictably from one map edge to the other. Geographic examples of such systematic trends might be the dome of atmospheric pollution over a city, the dome in population density over a city, a north–south trend in mean annual temperatures, and so on.

All that a trend surface analysis does is to define and fit surfaces to observational data by specifying a simple mathematical form for the function using conventional least squares regression. Since it is unlikely that any such relatively smooth mathematical function will correctly reproduce all the observed data heights, it follows that there will be local departures from the trend, called residuals. Mathematically, we denote this division of any height on a scalar field into a trend plus/minus a residual as in eqn [2]:

$$z_i = f(x_i, y_i) \pm \varepsilon_i \qquad [2]$$

That is, the surface height at the i th point is made up of the fitted trend component, $f(x_i, y_i)$, plus or minus a residual, or error, ε_i. It remains to examine possible forms for the function, f, and the details of how the fitting is accomplished.

The first issue we have is to decide on is the appropriate functional form for the trend, since there is an enormous range of candidates. The simplest trend surface imaginable is an inclined plane, or linear trend surface, specified as in

$$z_i = b_0 + b_1 x_i + b_2 y_i + \varepsilon_i \qquad [3]$$

This sort of trend might be assumed to be a reasonable model for a simple, regular change in the value of some variable across a region of interest.

To calculate values for the trend part of eqn [3], we need to know the coefficients b_0, b_1, and b_2 together with the coordinates of points of interest. For this linear function, the coefficients have direct, physical interpretations as shown in **Figure 1**. First, b_0 represents the height of the surface at the map origin, where $x_i = y_i = 0$. The second, b_1, is the surface slope in the x direction and the third, b_2, is the same in the y direction. The linear trend surface is the shaded inclined plane passing through a series of data points, each shown as a circle. Some of the observed data points, those colored white, lie above the trend surface and so have a positive error or residual, while others, shaded gray, are below the surface and so have a negative residual.

Having specified and computed a trend surface, we have two spatial distributions of interest. In some work it might be the hypothesized form of this trend that is of interest. In others, interest could center on the distribution and size of the local residuals. From the previous equations, it is obvious that these can be calculated from:

$$\varepsilon_i = z_i \quad (b_0 + b_1 x_i + b_2 y_i) \qquad [4]$$

That is, the residual at each point is given by the difference between the observed surface height at that point

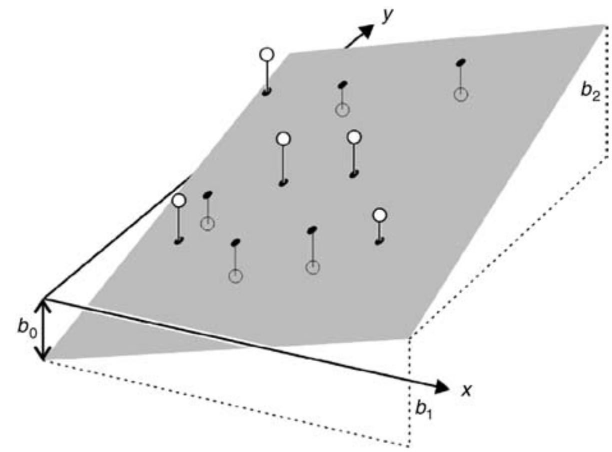

Figure 1 A simple linear trend surface.

and the value predicted by the fitted surface. Maps of residuals are a useful way of exploring data to suggest the operation of strictly local processes on the variable of interest.

Although it is the simplest of the functions that might be used to specify a trend surface, the linear model is limited. No matter how the values of the b coefficients are changed, the result is always a simple inclined plane threaded through the observational data. If geographical theory leads us to expect a different shape then other more complex surfaces can be fitted. Exactly the same technique is used, but, as we will see, the calculations become extremely lengthy. Suppose that we wish to create a trend surface that specifies a dome or trough like trend. The appropriate function to use is a quadratic polynomial, giving the model specified by:

$$z_i = f(x_i, y_i) = b_0 + b_1 x_i + b_2 y_i + b_3 x_i y_i + b_4 x_i^2 + b_5 y_i^2 + \varepsilon_i \qquad [5]$$

This is still a trend surface, but there are now six coefficients, b_0 to b_5, to be found and, instead of an inclined plane, it describes a tilted surface that has been warped along one axis giving a dome or trough like surface. In some of the literature this type of surface is referred to as a polynomial of the second order. Notice that no additional information has been provided, all the new terms in this equation are manipulations of the original (x, y) data.

Addition of further terms specifies ever more complex surfaces of 3rd, 4th, etc. orders, but in practice these have seldom been used. Other types of surface, including oscillatory ones specified by a double Fourier series and surfaces in three dimensions specified by (x, y, z) locational coordinates, have also been experimented with. Whether any of these more complex trend models can be matched to appropriate geographic theory is doubtful. The danger is of 'overfitting' the trend surface. In fact, if we have n observations to fit, it is almost always possible to get a perfect fit by introducing the same number of coefficients.

Fitting Trend Surfaces

The second issue concerns how, once defined, such surfaces are calculated. Returning for illustration to the linear model of **Figure 1**, it should be apparent that different choices for the coefficients b_0, b_1, and b_2 will specify different trend surfaces and hence different values for the associated residuals. The decision as to which is the best combination of values is made using the least squares criterion of goodness of fit in which the specific function chosen for the trend must minimize the sum of the squared residuals from the trend. This is the standard criterion used in ordinary least squares multiple linear regression. As we have noted, a linear trend surface is thus exactly the same as a conventional multiple linear regression in which the dependent variable is surface height, z, expressed as a function of two independent variables, the locational coordinates (x_i, y_i). The addition of terms to create a dome or trough like quadratic surface, as in eqn [5], simply adds an additional three $(x_i, y_i, x_i^2$, and $y_i^2)$ supposedly independent variables to the regression equation.

It is easily shown that the vector of required coefficients, **b**, that provides the least squares solution for the forms given in eqns [3] and [5] can be found from the matrix equation:

$$\mathbf{b} = (\mathbf{X}^{\mathrm{T}}\,\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{z} \qquad [6]$$

In this, the matrix **X** is given by:

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & y_1 \\ \vdots & \vdots & \vdots \\ 1 & x_n & y_n \end{bmatrix} \qquad [7]$$

The surface height, z, is the dependent variable, whose values form the column vector **z** and the solution vector, **b**, contains our required coefficients, b_0, b_1, and b_2 for eqn [3] for example. To fit alternative models, all that needs to be done is to change the definition of this **X** matrix. Thus for a quadratic, we would solve eqn [6] using a 6 column X matrix of the form:

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & y_1 & x_1y_1 & x_1^2 & y_1^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_i & y_i & x_iy_i & x_i^2 & y_i^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & y_n & x_ny_n & x_n^2 & y_n^2 \end{bmatrix} \qquad [8]$$

The so called sums of squares and cross products matrix, $\mathbf{X}^{\mathrm{T}}\mathbf{X}$, is now 6 by 6, its inverse $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$ is harder to find, and the vector of coefficients, **b**, now has six values, b_0–b_5. As the number of coefficients to be calculated increases, so do both the amount of arithmetic involved and the chance of error. In practice it would be unusual to go further than using the squares or cubes of the locational coordinates and the calculations are almost always done by computer.

During the 1960s a number of authors published elaborate programs, often in the FORTRAN language, specifically to calculate and map trend surfaces and their residuals. Nowadays, it is much easier to use standard statistical packages, such as Statistical Package for the Social Sciences (SPSS) or MINITAB, to do the basic calculations. The ability to calculate and map trend surfaces is included in many geographic information systems. However, workers tempted to fit polynomials with very many terms using commercial off the shelf software should be aware that eqn [6] can in some circumstances be hard to solve using a digital computer whose word length is fixed. This arises both from the size of the numbers involved when very high powers of the x and y coordinates are used and from linear dependence between the rows of the matrix $\mathbf{X}^{\mathrm{T}}\mathbf{X}$, which makes the inverse, and hence the solution for **b**, very unstable to small changes in the input data.

Some Statistical Issues

The above exposition has considered trend surface analysis from a strictly mathematical viewpoint, neglecting the fact that a statistical interpretation is also appropriate. In statistical terms the process described above uses some observed values of the field, z_i, at a series of locations (x_i, y_i) to estimate a set of unknown coefficients, **b**, defining a specified trend surface that in the least squares sense best fits these data.

Clearly, it would be useful to have some measure of how well the surface fits these observed data. This is provided by comparing the sum of squared residual values for the fitted surface to the sum of squared differences from the simple mean for the observed z values, called the percentage reduction in sums of squares. This quantity is better known as the coefficient of determination used in standard regression analysis and is given by the square of the coefficient of multiple correlation, R^2:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \varepsilon_i^2}{\sum_{i=1}^{n}(z_i - \bar{z})^2} = 1 - \frac{\mathrm{SSE}}{\mathrm{SS}_z} \qquad [9]$$

In this, SSE is the sum of squared errors and SS_z is the sum of squared differences from mean. The statistic indicates how much of an improvement the fitted trend surface is, compared to simply using the data mean to predict unknown values.

In almost all applications of the method, the observed data will be a sample taken from an underlying continuous field of surface heights. It follows that the coefficients, **b**, although calculated by a defined mathematical procedure, can also be viewed as least squares estimates of some unknown population parameter values, which are denoted using the Greek characters as the vector $\boldsymbol{\beta}$. The question that naturally arises is to what extent these values are statistically significantly different from zero, which would indicate that in the population there is no trend of the type specified in the fitted model. As in conventional multiple linear regression, whether or not this fit is statistically significant can be tested using an F ratio statistic

$$F = \frac{R^2/\mathrm{df}_{\mathrm{surface}}}{(1 - R^2)/\mathrm{df}_{\mathrm{residuals}}} \qquad [10]$$

In which $\mathrm{df}_{\mathrm{surface}}$ is the degrees of freedom associated with the fitted surface, equal to the number of coefficients used, less one for the base term b_0 and $\mathrm{df}_{\mathrm{residuals}}$ is the degrees of freedom associated with the residuals. This is the total number of observed data points, n, less the number of coefficients in the fitted equation. Formally, this is a test of the null hypothesis that the population parameters are all equal to zero:

$$H_0 : \beta_0 = \beta_1 = \beta_2 = \cdots = \beta_n = 0 \qquad [11]$$

The alternative is that one or more of the parameters is not zero:

$$H_1 : \text{not all } \boldsymbol{\beta} = 0 \qquad [12]$$

If the computed value exceeds the tabulated of F, the null hypothesis is rejected in favor of the alternative, and it is inferred that there is evidence of a real trend of the specified functional form in the real field. A complication that arises in the use of higher order, polynomial models is that they include many more terms in the model and many more parameters to be estimated. Here the appropriate question to ask at each step is whether or not the additional terms contribute significantly to the definition of the trend. This is done by taking the difference between the surface fitted and that of the next lower order, divided by the difference in degrees of freedom between the two, dividing this in turn by the mean square due to the residuals of the higher order surface. The resulting F ratio thus allows a test of the significance of the added terms.

If these tests reveal that the null hypothesis cannot be rejected, that there is no significant trend in the data, several explanations might be adduced. One possibility is that there really is no trend of any sort across the surface. Another is that there is a trend in the underlying surface but that the sample size, n, is too small to detect it. Yet another is that the wrong functional form has been specified for the trend.

Calibration using least squares makes some assumptions about the data that in practice might be hard to sustain. It is simply impossible to solve eqn [6] if the number of observed data values is less than the number of coefficients in the trend specification, thus setting an absolute lower limit on the sample size, n. If formal statistical testing is appropriate, then n also sets limits on the number of degrees of freedom associated with the residuals, and hence the applicability of the F ratio tests. However, because we are dealing with spatial data two issues that are present, but often hidden, in conventional nonspatial least squares regression become very evident. The first is that when mapped there is a tendency for surfaces of order higher than the linear to give pronounced and unwelcome edge effects in which predicted values can be very high or low along the margins of the area over which sample values have been used. When possible, many workers will protect against this by forming a buffer region around the area of interest into which these effects can be concentrated and then ignored. The second, related, effect is for individual, isolated data points to have excessive influence on the form of the fitted surface, a phenomenon that in conventional regression analysis is called their leverage. In general, the leverage of a datum increases away from the center of the mapped area and is most for points at the very edges of the distribution. Such points may, or may not, be well fitted, but it is wise to be aware of their potential influence on both the form and fit of any trend surface. Formal calculation of the leverage exerted by each data point has been shown to be a very useful diagnostic in analyzing such effects.

Finally, calibration using least squares also makes the usual assumptions about the unknown error terms in the underlying population, which are assumed to have constant variance over the entire studied area, be normally distributed, and independent of each other. It is the independence assumption that is the most troublesome, since in a very large number of applications this will simply not hold and will be evidenced by the phenomenon of spatial autocorrelation in the residuals. If spatial autocorrelation is present, there is some remaining spatial pattern in the observed data that has not been properly modeled by the fitted surface. The wrong model for the trend has been fitted. Autocorrelation in the residuals also means that in significance tests the effective degrees of freedom is less than assumed, which in turn leads to computed F ratios higher than they should be and thus a greater chance of committing a Type 1 error, that is, of inferring a statistically significant trend when one cannot be justified. As with conventional multiple linear regression, statistical purists might well argue that the analysis and associated tests are flawed, but the fact remains that in numerous applications the results of trend surface analyses that have broken these assump tions have been found to be scientifically useful.

Applications in Human Geography

Trend surface analysis was introduced into geography from its origins in geology, and especially stratigraphy where it was used to model the depth of specific rock horizons of interest, or variations in quantities such as mean grain size within a formation. In physical geog raphy it was used to study large scale spatial variations in, for example, rainfall and atmospheric pollution. It has also been used as a means of locally estimating the gra dient of the land surface, the skiers 'fall line', by fitting linear or quadratic surfaces to the values in a kernel around each point and then computing both the magni tude and direction of the maximum slope through that point. Similarly, low order trend surfaces are routinely used to remove 'drift' in the method of spatial inter polation known as Universal Kriging.

There are a number of reasons why trend surface an alysis has been less effectively used in human geography. Probably the most important is the fact that, unlike physical geography in which such analysis is the norm, human geographers only seldom attempt to model spa tially continuous fields. Instead, the objects of interest tend to be discrete points (e.g., a factory marked with as its z value the output or number of workers), lines (e.g., a railway network, also marked with the frequency of ser vice), or areas (e.g., a census tract, marked with its popu lation voting in a particular way). In no sense are these objects samples of some underlying real, continuous phenomenon. At first sight this limits the applica bility of trend surface models in human geography, yet there have been several ways by which this has been cir cumvented. Consider, for example, that we have some strictly discontinuous information that relates to discrete points across a region such as the number of workers in a series of factories distributed around a city. In no sense is there a continuous field of factory employment across the region. However, effectively ignoring the problem, work ers have fitted surfaces to such data and argued that the results provide a useful separation of a spatial trend from purely local effects. An alternative, used in the same case, has been to argue that these data are a sample of what in some sense is a potential for employment (indeed, this can be formalized in the notion of a population potential). Others have noted that, although the individual objects are discrete, it is possible to conceptualize continuous surfaces of their spatial density and thus to apply these types of surface models to them. Where the objects of interest are discrete areas, such as census tracts, workers have applied the trend surface method using as the locations of their dependent variable some sort of reference points such as the centroids of the areas involved. Provided the areas are uniformly small and that there is a large number of them across the region of interest this seems legitimate, but it is a device that should be used with extreme caution. Finally, experiments have been undertaken in which surfaces have been generated through discrete, binary (0/1, yes/no), point valued data. By definition, fitting such surfaces cannot use ordinary least squares methods. Instead, a logit model approach is taken using different methods of esti mation for the coefficients and the resulting predicted values are interpreted as probabilities of a specified binary response across the region.

In conclusion, trend surface analysis has been a much used and abused technique, but it remains one that is useful in spatial statistical analysis.

See also: Edge Effects; GIScience and Systems; Maps; Quantitative Methodologies; Regression, Linear and Nonlinear; Spatial Autocorrelation; Spatial Interpolation; Statistics, Spatial.

Further Reading

Bailey, T. C. and Gatrell, A. C. (1995). *Interactive spatial data analysis*. Harlow: Longman.

Chorley, R. J. and Haggett, P. (1965). Trend surface mapping in geographical research. *Transactions of the Institute of British Geographers* 37, 47 67.

Davis, J. C. (2002). *Statistics and data analysis in geology* (3rd ed.). New York: John Wiley & Sons.

Norcliffe, G. B. (1969). On the uses and limitations of trend surface models. *Canadian Geographer* 13, 338 348.

O'Sullivan, D. and Unwin, D. (2003). *Geographic information analysis*. New York: John Wiley & Sons.

Unwin, D. J. (1975). An introduction to trend surface analysis. *CATMOG*, 5, Norwich, Geo Abstracts.

Wrigley, N. (1977). Probability surface mapping: A new approach to trend surface mapping. *Transactions of the Institute of British Geographers, New Series* 2, 129 140.

Wrigley, N. and Unwin, D. (1987). Towards a general theory of control point distribution effects in trend surface models. *Computers and Geosciences* 13, 351 355.

Relevant Websites

http://www.wun.ac.uk/ggisa/elearning.html
Global GIS academy comprehensive introduction to regression methods.

http://www.spatialanalysisonline.com
Simple regression and trend surface modeling.

http://www.kgs.ku.edu
Trend surface analysis.

http://www.geog.ucsb.edu
Trend surface analysis, basic concepts.

Triangulation

A. Nightingale, University of Edinburgh, Edinburgh, UK

Glossary

Methodological Framework A framework for gathering and analyzing information within a research project. It links together conceptual or theoretical ideas with the methods used.

Participant Observation A method of data collection used in qualitative research to gather information on daily practices. The method involves active participation in activities with research participants in order to gain insight into those activities and the social dynamics surrounding them.

Positionality The position from which a researcher views the research objectives. It generally refers to various aspects of a person's social status including gender, race, class, educational level, etc. It also includes factors that create bias or a particular perspective.

Postmodern Theories A body of conceptual thought that developed out of critiques of structuralism. It most often includes the idea that knowledge is not fixed or independent, multiple truths can coexist, and that meaning is a contested process created in context. It causes a shift from searching for causes to questions about how things happen.

Validate The process by which research methods are evaluated for the extent to which they reflect the research questions asked. Validation in quantitative methods involves the calculation of statistical errors, whereas in qualitative methods validation methods include cross-checking or triangulation with other methods; theoretical saturation, or the point at which no new information is gained; and adherences to a code of conduct believed to ensure accurate results such as avoiding leading questions.

Triangulation

Triangulation is an analysis technique used in multimethod research designs. Many research projects utilize more than one data collection method, leading to the development of different datasets. Datasets might be those collected from a quantitative survey or participant observation, for example. The results from the datasets are analyzed independently, but they also need to be compared to each other in some way. How they are compared depends on the methodological framework used. Triangulation is one technique to combine datasets, and three different kinds of triangulation can be distinguished: convergence, complementarity, and divergence or dissonance. These three kinds of triangulation are discussed in detail below.

Triangulation is a term that was borrowed from surveying, where it refers to the collection of different compass bearings (usually three, hence triangulation) in order to pinpoint a location on a map. In research, there is no fixed number of datasets that need to be compared, although in many cases, researchers will use two main datasets. The datasets are compared for convergence, complementarity, and divergence. If the results lead to the same conclusions, then the methods help to validate each other. Some researchers will also use triangulation to validate results by comparing the results from different observers of the same phenomenon. In other words, it is a form of cross checking.

Convergence

Triangulation has most commonly been used to ensure that methods were done properly and the data are robust by looking for convergence in the results. A research project that has used participant observation and interviews, for example, would check that responses given in interviews are consistent with observations made by the researcher. If the two datasets agree with each other, they are said to converge. This kind of validation requires that researchers believe that the methods will produce the same kind of understanding of reality. If the question is about forest harvesting practices, for example, it is assumed that there is a set of forest harvesting practices that can be discussed and observed and that verbal accounts and practices will mirror each other. In this kind of understanding, if the datasets do not converge, then the quality of the data is drawn into question and the researcher would seek to discover the error. This is an important point, as in other kinds of triangulation the lack of convergence would signal theoretical errors or be treated as new data rather than data collection errors.

Convergence can also be sought when comparing quantitative and qualitative data. In these instances, often a quantitative survey will be used to validate the results of qualitative interviews, or more rarely, the other way around. Again, this kind of analysis assumes that the two datasets are evaluating the same phenomenon and will produce the same kind of knowledge about that phenomenon. Similarly, when looking for convergence of results from two different observers, the assumption is that the results are independent of the observer and repeatable; therefore, a lack of convergence would signal problems with the data collection.

Complementarity

Convergence is the most common form of triangulation but two other methods of triangulation also need to be highlighted. As mentioned above, triangulation for convergence assumes that the datasets produce the same picture of reality. Theoretical developments in the social sciences since the 1970s, however, have challenged the idea that knowledge production is so straightforward. Instead, these theories focus on the importance of the positionality of the researcher when analyzing research results. In this kind of paradigm, datasets may not produce the same understanding of the phenomenon under investigation and therefore triangulating for convergence is not appropriate. Some theories propose that unobservable phenomenon or structures shape observable phenomenon and therefore the results from two different kinds of datasets may not perfectly converge. Other theoretical paradigms go further and contest the existence of one truth and rather propose that truth is situated depending on the observer – these are often called postmodern theories. Therefore, datasets would not be independent of the research context and as such not necessarily repeatable. In both instances, researchers would not expect convergence. Rather they may look for complementarity.

Complementary triangulation seeks to produce a fuller picture of the research questions by combining information from different methods or different observers. In this kind of analysis, quantitative and qualitative datasets can be combined together. The results are not expected to be the same, but rather to make sense in relation to each other and to help create a fuller picture of the research problem by creating more complete information about a topic. Commonly, this kind of analysis will use a phased methodology, where researchers will use one method in order to have a background understanding for another kind of data collection. For example, a quantitative survey might be used to collect statistical information on a population in order to then conduct in depth qualitative interviews with people from groups identified by the survey. Similarly, qualitative interviews might be used to help develop questions for a quantitative survey. The analysis of the first dataset thus informs the development of the second dataset.

Complementary triangulation does not necessarily need to be phased, however. In fact, triangulation is often used to understand a research problem more fully by using datasets both of which provide similar information. To return to the forest harvesting example, a researcher may use a quantitative survey, participant observation, and interviews, to gather information on harvesting practices. The survey would be designed to gather information on quantities of harvest and the range of practices used, while the participant observation would try to understand how harvesting practices are implemented on a day to day basis and the interviews to get descriptions and people's feelings or beliefs about harvesting. All of these provide important information about harvesting, some of which would be expected to converge, but a lot of it would be complementary. For example, the interviews may help to explain why some households harvest less than others, a finding that was revealed by the survey. Here, the important methodological issue is that triangulation is not only trying to validate the methods used, but rather is used to enrich the analysis. If complete divergence was the result, then the researcher may draw into question the validity of the methods as some convergence is expected, but it is not the only purpose of comparing the datasets.

Divergence

Divergence of datasets presents a different set of methodological challenges. When triangulation is used within a research project that assumes the different methods used are providing information on the same phenomenon, then divergence indicates problems with the data. Using the above example, the researcher would check that how they asked questions in the harvesting survey were tested properly to ensure respondents understood what was asked if the results diverged dramatically from the qualitative interviews.

Many researchers who work with postmodern theories, however, are interested in how knowledge itself is produced and thus find complementarity, dissonance, and gaps between datasets interesting. In this kind of methodology, the researcher would not necessarily expect convergence or even complementarity but rather would seek to interrogate the implications of both consistency and inconsistencies between datasets. The inconsistencies do not signal problems with the data collection but rather problems with the theoretical assumptions behind the object of research. In the forest harvesting example, dissonance between the interviews and participant observation can provide important insights into which practices people are willing to publicly disclose (record on a cassette) and which they choose to keep more hidden. This may provide important information on people's willingness to cooperate with a management plan. In cases where there is no management plan to contest, it is also possible that people are not conscious of all the harvesting techniques they use or the importance of them. In this case, dissonance can provide insights into which aspects of harvesting are considered most im portant to the participants. Here, theoretical assumptions about why people chose to reveal certain aspects of their harvesting practices would need to be carefully exam ined. It would also signal the need to collect more data to help verify if the new theoretical assumptions are more accurate. Methodologically, the important issue is that the gaps and silences between datasets are used as data in themselves and interrogated to produce a richer picture of the research questions.

Finally, a few researchers use dissonance to present different understandings of a phenomenon. This is not considered triangulation, as the datasets are deliberately not compared, but rather their divergence is held in a creative tension to highlight issues around the production of knowledge.

Studies That Have Used Triangulation

Many geographical studies use triangulation. A common way is to look for convergence between interviews and participant observation. For example, a study of a wildlife management area and associated community manage ment schemes in Botswana used in depth, semistructured interviews as the main method, complemented by ob servation, participation in activities, and secondary sources. Thus, the different methods employed all lent insight into the dynamics of community management. Triangulation of the methods was primarily designed to help validate the interviews (convergence) and perhaps add some new information (complementarity). There was no overt attempt at analyzing any areas of dissonance or inconsistencies between the datasets. This study in combination with another study, also used a variety of other methods to understand the overall situation, but the triangulation strategies employed were primarily to look for convergence and complementarity.

Other ways of using triangulation are also evident in the geographical literature. A study of community for estry in western Nepal used oral histories and aerial photo interpretation to build an understanding of his torical forest change. Importantly, in this study the pur pose was not to validate the oral histories with the aerial photo interpretation, but rather to treat them both as equally valid ways to understand forest change. Thus, when the oral histories indicated that forests had im proved significantly and the aerial photo interpretation showed only slight improvement, this dissonance be tween the datasets was explored. By analyzing the dis sonance between datasets as a new form of data, the study was able to better explain how the participants under stood 'forest conditions' to be both the ecological con ditions and the management arrangements. In other words, the use of triangulation was able to demonstrate how the participant's world view did not separate the social from the ecological. The comparison of these two datasets also lent insight into how particular parts of the forest were more important than others, something which the interviewees had been reluctant to express. They had an interest in retaining control over the entire forest and therefore were careful to emphasize in interviews that the entire stand was all equally important. Yet the aerial photos and participant observation demonstrated that not all parts of the forest were used regularly. Certain parts were more important for the gathering of resources used daily, such as firewood and leaf litter, while other parts were preferred for timber and the more remote parts were rarely used. Thus the dissonance between the datasets lent insight into which parts of the forest are the most valuable, at least in the short term.

This kind of analysis was also consistent with the theoretical aims of the project which sought to interro gate different forms of knowledge. Working from the literature, the study was interested in how different kinds of methodologies helped to reveal different understand ings of the world. The aerial photos thus provide one way of gaining knowledge of the forest while the oral histories provide another. Triangulation then was used analytically to explore the ways in which both methods can produce only a partial understanding. Combining them produces a richer picture but researchers need to be careful of assuming triangulation can produce a 'complete' picture.

See also: Postmodernism/Postmodern Geography; Quantitative Methodologies; Questionnaire Survey; Reliability and Validity.

Further Reading

Denzin, N. K. (1988). *The Research Act: A Theoretical Introduction to Sociological Methods* (3rd edn.). Englewood Cliffs, NJ: Prentice Hall.

Denzin, N. K. and Lincoln, Y. S. (eds.) (1998). *Collecting and Interpreting Qualitative Materials*. Thousand Oaks, CA: Sage Publications.

Erzberger, C. and Prien, G. (1997). Triangulation: Validity and empirically based hypothesis construction. *Quality & Quantity* 31, 141 154.

Krueger, R. A. (1998). *Analyzing and Reporting Focus Group Results*. Thousand Oaks, CA: Sage Publications.

Nightingale, A. (2003). A feminist in the forest: Situated knowledges and mixing methods in natural resource management. *ACME: An International E Journal for Critical Geographers* 2, 77 90.

Robson, C. (2002). *Real World Research* (2nd edn.). Oxford: Blackwell.

Rocheleau, D. (1995). Maps, numbers, text and context: Mixing methods in feminist political ecology. *Professional Geographer* 47, 458 466.

Shurmer Smith, P. (2002). Methods and methodology. In Shurmer Smith, P. (ed.) *Doing Cultural Geography*, pp 95 112. London: Sage.

Silverman, D. (2000). *Doing Qualitative Research: A Practical Handbook*. London: Sage Publications.

Twyman, C. (2001). Natural resource use and livelihoods in Botswana's Wildlife Management Areas. *Applied Geography* 21, 45 68.

Waysman, M. and Savaya, R. (1997). Mixed method evaluation: A case study. *Evaluation Practice* 18, 227 237.

Yin, R. (1994). *Case Study Research: Design and Methods*. Thousand Oaks, CA: Sage Publications.

Relevant Websites

http://www.sociology.org.uk
Sociology Central: teaching notes.

http://www.voicewisdom.co.uk
Voice Wisdom: triangulation in qualitative research.

Tropical Geography

M. Power, University of Durham, Durham, UK

Glossary

Eurocentrism The privileging of European culture, beliefs, values, and religions, to the extent that these have been naturalized, often through imperialism.

Genealogy Through genealogical study, Michel Foucault examined how discourses and social practices manifest the limitations of traditional historical study; the process discloses the human power plays of domination, imposition of order, subjugation, and struggle through resistance.

Modernization Modernization theory is a socioeconomic theory, encompassed within 'development theory', which highlights the positive role played by the 'developed world' in modernizing and facilitating progress in 'underdeveloped' nations, often contrasted with 'dependency theory'.

Orientalism Edward Said's term for an entire discourse through which the colonial 'other' is represented by the West as subordinate, thus providing an intellectual foundation for material domination.

Postcolonialism This term refers to a variety of theoretical positions and historical conditions associated primarily (though not exclusively) with the aftermath of colonialism. Postcolonialism seeks to critique Western ways of knowing and traditions of thought. It also involves a retracing of the impacts of colonial processes with a particular concern for marginal or subaltern peoples.

Tropicality 'Tropicality' is a term that refers to the potent Western discourses (which were deeply implicated in colonialism) that construct non-European otherness in environmental terms.

Introduction

The concept of 'the tropics' has been seductive, powerful, and enduring. Whether considered a sublime landscape, malignant wilderness, or the endangered site of environmental conflicts, the tropics are largely a construct of American and European imaginations. European ideas about the tropics that developed as a result of exploration, conquest, and colonization from the fifteenth century onward were very diverse, however. Images of the tropics in the nineteenth and twentieth centuries were conveyed through drawings, paintings, photographs, literature, and travel writings. From the earliest photographic attempts to represent tropical hybrid races to depictions of disease in new tropical medicines, there were thus multiple practices through which the tropics were known, practical and bodily, as well as intellectual and discursive. The term 'tropical' is still routinely attached to a whole range of issues, everything from resorts to beaches, from rainforests to rainstorms, from urbanization to soils. 'Tropicality' on the other hand is a term that refers to the potent Western discourses that construct non European otherness in environmental terms and is therefore deeply implicated in colonialism. The identification of the northern temperate regions as the normal, and the tropics as altogether 'other' – climatically, geographically, and morally – demonstrate that the tropics was a conceptual and not just a physical space. Tropical nature is here produced as 'other' where through the prism of climate, acclimatization, and disease the temperate world is seen as the environmental 'norm' against which the tropical world is perceived and evaluated. As we shall see however, discourses of tropicality were highly ambivalent, with 'the tropics' seen in both positive and negative terms (as a space of abundance and fertility, as well as of poverty and disease). By the beginning of the nineteenth century it was commonplace to proclaim that the luxuriance and heat of the tropical zone made its indigenous peoples wild and indolent and therefore ill equipped to harness the natural riches of the tropical world or to become 'civilized'. In many ways a false singularity was ascribed to the tropics as 'other' where climate was used to make sense of cultural difference and in the process to project moral categories onto global space.

The notion of a 'tropical geography' has a somewhat complex history through the nineteenth and twentieth centuries and became more formalized and more widely recognized after World War II. The French geographer Pierre Gourou (1900–99) was one of its most famous exponents, but there were also important American, British, Dutch, Portuguese, and Spanish geographical formations of 'tropicality' (although there were also many skeptics who disagreed that this consituted a separate branch of the discipline at all). Tropical visions and representations extended far beyond particular national research schools and traditions, while debates about tropical geography drew in contributions from geographers based in a wide variety of locations. Work on the genealogy of tropical geography is still in its infancy, but longer term histories of tropical science, travel, and medicine have been a fertile area of research. To date, more attention has been given to celebrated traveling naturalists such as Charles Darwin or Alexander von Humboldt (who once described himself as 'made for the tropics'), and there has also been more work on tropical nature, the role of European scientific institutions and global networks in shaping the emergent fields of tropical botany, tropical medicine, and tropical climatology. From 1953 new possibilities began to emerge for the publication of geographical scholarship on the tropics with the creation of the *Malayan Journal of Tropical Geography*. Pierre Gourou's *The Tropical World* (an exemplary text in tropical geography) was also published in English in the same year. A distinct field of geographical enquiry had thus begun to emerge, supported by new conferences, journals, and funding possibilities. In a way, the notion of the tropics was crucial to the very creation and professionalization of the discipline of geography, imbuing it with a sense of itself as a 'sternly practical science'. The notion of tropicality was also bound up with the development of other disciplines such as anthropology, medical science, and natural history. Geographers thus 'invented' or constructed areas and regions of study that would concentrate their 'expertise' – frequently in the service of colonial rule and hence many geographers did shape colonial practices in tropical spaces. Geographical practices and knowledges thus provided a set of lenses through which the tropics were known, understood, and represented. What began to emerge was a new spatialized domain of intellectual enquiry and imagined constructions of otherness around which crystallized a distinctly 'modern' set of truths, assumptions, and hierarchies. In order to problematize the notion of 'the tropical' in geography it is firstly necessary to open a historical enquiry into the term and its enrolment in a variety of scientific, aesthetic, and political projects.

Orientalism, Geography, and the Tropics

Western colonial projects were, despite being far from homogenous, partly based on an imagination of the world that legitimized and supported the power of the West to dominate 'others'. By representing other societies as 'backward' and 'irrational', the West and Europe in particular construct themselves as 'mature', rational, and objective. Here the notion of the tropics was invented alongside a range of other constructs (e.g., 'the Orient') such that certain societies are envisioned as 'other' and different from Europe or the 'West'. This difference was often seen as something which could be interpreted and represented by rational, even handed social scientists in a supposedly 'purely objective' and 'scientific' manner.

The tropics then can be understood as an imaginative geography, as a particular kind of Orientalism (following the work of Edward Said) or as an imperial vision of particular places and subjects (that displaces other voices) which had important material consequences. As landscapes of desire, the tropics were understood as a domain of largely untamed nature that served, by contrast, to demonstrate the moral and material 'superiority' of northern climates, civilizations, and 'races'. For Said, the discourse of Orientalism transcends individuals or institutions and represents a kind of collective archive of images and statements. In an important way this provided a common language for knowing a particular region and its peoples. 'The Orient' was thus established in the European geographical imagination as an image, an idea, a personality, an experience, thereby reinforcing a sense of 'difference' between 'the Orient' and the rest of the world. Europeans sought knowledge about the Orient and through their Orientalist discourse they produced particular forms of domination, inventing the authority and legitimacy to subjugate and oppress non Western peoples. The concept of the tropics illustrates that Europe possessed more than one sense of otherness with the tropical seen as a twin to the temperate: the opposite of all that was civilized, modest, and enlightened. What we are concerned with then, following this recasting of Said's *Orientalism*, are the Eurocentric discourses which enframe the 'tropical world' and the way they 'contain' and 'normalize' them through a certain kind of discursive hegemony. Through such discourses, acceptable ways of speaking about and representing the tropics are established which have been powerful and enduring.

In the nineteenth century, the tropical was enfolded within racial imaginaries, especially in debates about the prospects for white settlement, 'degeneration', evolution, and 'races' that were rendered 'scientific' by the cloak of Darwinism. Geography participated in this via studies of climate and land use, rendered through theories of 'determinism' and 'possibilism'. Despite the many opportunities presented by colonial possessions and emerging debates about colonial development, few British geographers had worked overseas by 1939. Prior to 1945 it was geographers in a number of other European countries who first began writing tropical geographies. In France much research focused on Southeast Asia through the work of geographers like Pierre Gourou. In each case the connections between geography, geographers and empire were always close (even if this was not always acknowledged) as the maintenance of European empires created a number of new opportunities for geographical research. Most of this work was conceived of as part of the quest for 'enlightenment' and more scientific knowledges of colonial territories with rich potential for further colonial progress and development. Particular geographical ideas, to do with land use and agrarian change, population growth and mobility and environmental conservation, run through much of the early thematic approaches of these kinds of geographies. At the end of World War II many geographers could easily and immediately see the case for 'colonial geography' (geography in the service of colonialism) and were consequently often uncritical of the impact of colonial rule on indigenous societies. Some geographers even dedicated large sections of their texts to 'colonial achievement' and (in general) viewed colo nialism as something of a material success.

World War II was a key turning point in that it had led to much wider, unexpected, and in some cases involun tary foreign travel by geographers. A number of British geographers such as Charles Fisher (who was posted to Singapore) and B.H. Farmer (who served in India, Ceylon, and Singapore) found themselves in the service of the military (often the Royal Engineers or sometimes the Inter Services Topographical Department) and even in some cases became prisoners of war. Other geographers worked in England for the Naval Intelligence Division (NID) that produced a series of geographical handbooks intended to provide commanding officers in the Navy with information on countries they might be called upon to serve in. These handbooks were widely drawn upon by the first generation of post war British geographers (many of whom were returning from war) and served to inspire geographers to enter the brave new world of re construction in the colonies and the newly independent states of India and Pakistan in particular. The idea of 'geography militant' is important here: many tropical geographies were written by former military personnel and in some cases former prisoners of war and they were inspired by war experiences and geographical handbooks produced by military institutions. Each expedition and exploration of the tropics or its constituent parts also often represented a kind of militarized crusade dedicated to progress and modernity, the realization of 'tropical potential', and to the enlightenment of tropical spaces.

Tropicality, Regional Geography, and 'Development'

The curiosity aroused by war travels and experiences, coupled with the provision of new funding opportunities to conduct research overseas, thus began slowly to in crease the interest of geographers in other, distant worlds. Additionally, the emerging notions of underdevelopment and underdeveloped areas, in conjunction with the cre ation of new national and international development agencies, also provided a stimulus to geographical re search. Geographers soon staffed new universities in the colonies such as the University College of East Africa, University College of the Gold Coast, and the Uni versities of Ceylon, Colombo, and Malaya in Asia. In France and Germany the end of World War II also cre ated a new wave of geographical research on the tropics, among which Pierre Gourou's work was particularly influential. Gourou offered a powerful framework for geographical writing, which impacted on the post war work of a number of European geographers. Gourou's texts in many ways tell us as much about a collective northern centered, even Eurocentric, worldview as they do about the 'tropical world' they seek to depict. Gourou devoted considerable attention to tropical diseases, soils, plantations, and population densities, the importance of scientific knowledges for tropical 'development' and the potential for white settlement and the relations between lesser and greater civilizations. Echoing the ideas of nineteenth century philosophers, Gourou sees the West as the epitome of 'civilization', with India and China in secondary roles and the rest more or less outside history – the subjects of 'tropical geography'.

The influence of physical factors predominates in many tropical geographies in a way that frequently ver ges on environmental determinism. Climates rather than capitalism impoverish the people of the tropics in these accounts. New technologies and new agricultural tech niques were seen as the way forward for rural economies in tropical regions. Where colonialism is seen as relevant the focus is on white settler states like Rhodesia and the needs of white colonists, rather than with indigenous stories and experiences of colonial rule. For the next 25 years after the end of World War II numerous bulky, regional geographies of the non Western world were published by geographers like Dudley Stamp, stimulated in large part by the naval handbooks. Although more information was available for post war 'regional geog raphy', there was limited methodological or theoretical progress (methodology was largely intuitive and few at tempts were made to elaborate a theoretical context). As such, a great deal of this work was characterized by a 'dreary' descriptive nature or by a deadening stylization. The regions selected for study were not pre given and there were no formal rules for recognizing, defining, delimiting, or describing the region or its broader pos ition within the imagined 'tropics'. Some of these in herent contradictions meant that by the end of the 1960s (and in the context of a move to more quantitative techniques) this kind of regional geography had become increasingly unpopular.

Debates revolved around whether regional synthesis was a valid academic study given the tendency to divide the world into unique and seemingly unrelated chunks and the tendency not to seek to identify more general laws governing progress and 'development' in all regions. These kinds of simplified images of difference, which assumed away much of the geographic variation within the world, became less popular partly because (despite being very normative) they offered no immediate ethical and socially responsible solutions to social ills. For some geographers, what was lacking was a dedicated special ization by area. Thus the implication was that work conducted before the early 1970s was characterized by a lack of specialization and depth, by an unsatisfying un reality. Area specializations were also to become in creasingly important during the Cold War. A lack of area experts to inform policy was identified by the Armed Forces during the 1940s as World War II came to an end. In the 1950s six area studies associations were formed in the US as social scientists took up the mantle laid down by President Truman's injunction to reach out to the 'backward' regions of the world.

The post 1945 creation of area studies centers in some Western countries was intimately bound up with Cold War geopolitics and impacted upon a wide range of social sciences. Geographers soon became active in many such centers, and also found roles in newly established 'schools of development studies'. Some critics had fundamental objections to the very conception of these area studies centers, arguing that they had been 'colonially conceived' and that the appropriate centers of Asian or African studies should be in Asia or Africa. Thus specialist research into 'backward' regions began to increase in volume as more disciplinary generalists entered the debate around the crude classification of 'Third World' studies. During this early period of the Cold War then, development and area studies programs became in creasingly initiated by US state security discourses, which began a certain kind of relationship between scholarship and policy making, centered on discourses of 'development' and 'modernization'. These attempts could thus be viewed as designed to produce knowledge that would enable the maintenance of political control over societies that threatened the institutional capacities of certain states.

The Degeneration of Tropical Geography

Pretensions of 'modernizing' tropical regions structured many of the early post war assumptions about modernity and how to progress away from conditions of under development. This way of viewing non Western regions began to break down and degenerate after 1945 with the widespread desire to overcome European notions and idioms of imperialism such as 'the white man's burden', the 'civilizing mission', and the idea of 'civilization'. In this way social scientists in various different Western countries began to adopt much less avowedly procolonial notions of 'modernity' and 'development' and began to engage with more radical streams of social theory. Add itionally, these geographers (perhaps more than many of their contemporaries) were influenced directly by de bates and publications in other associated disciplines (such as economics, sociology, anthropology, and the emerging fields of development studies and international relations). Not surprisingly, much of the geographic scholarship of the 1950s and 1960s was framed by some variant of modernization theory, or the presumption that processes of modernity were shaping indigenous insti tutions and practices. Many geographers interested in Africa sought to model modernization surfaces and at tempted to map patterns of modernity by charting the diffusion of indices of modernity (such as schools or mailboxes) through the settlement pattern. This work at best raised only limited questions about the legacies of colonial transport systems or the character of African urbanism. The more influential work was increasingly focused on 'modernization', seen less as a tropical phe nomenon and more as part of a 'universal' process of 'development'. Many of the contributions to what was becoming known as 'development geography' still lacked any major theoretical undercurrents however and often concentrated on broad, traditionally defined, questions. During this time the number of geographers based in the South began to increase, in countries like Nigeria, for example, partly as a consequence of the establishment of new universities. These geographers had an increasingly important impact on the formation of national develop ment policy (particularly in the area of planning) from the late 1960s and early 1970s onward. As the numbers of geographers from non Western regions began to increase, the theoretical impoverishment of development geog raphy was soon brought into stark relief.

Major shifts in development thinking were thus beginning to influence researchers in a number of disci plines and the way they viewed the hallowed 'develop ment process'. In the 1960s this began to change the nature of the discipline as geographers were propelled into a heightened state of social awareness by political and economic events around the world. Dependency theorists like Andre Gunder Frank had begun to offer popular structuralist explanations of the causes of 'underdevelopment' and the idea of development as 'economic growth' had been subjected to radical critique by theorists who sought to locate the production of dependent relations within the structure of the capitalist world economy. These shifts in development thinking during the 1960s and 1970s had a profound impact on geographers, particularly after the establishment of the radical journal *Antipode* in 1969, which soon began pub lishing important articles about geography and uneven development. This journal was founded, upon principles of radical geography, by a group of faculty and students at Clark University in Worcester, Massachusetts in 1969. In the United States many of these geographers met from 1967 at the annual geography conferences, but the dis cipline of geography in the US was slow to respond to their radical discussions, dominated as it was until the early 1970s by 'the quantifiers' and a kind of ephemeral spatial science. The philosophical and theoretical foun dations for this radicalism were found in the work of

Marx, which radical geographers began to engage with more and more in the late 1960s and early 1970s.

Instead this work called for a critical approach, arguing that most of the geographers of the Western world had been educated in the 'liberal' ethos that development is a necessity. In this way many geographers criticized the modernization approach for its reassertion of the primacy of the 'American–European experience' and argued that it was important to remember that other regions had different trajectories. Whilst these interventions were pointing to other directions that would only emerge more fully in the 1990s amidst debates about 'post colonialisms', 'postdevelopment', and (alternative) 'modernities', the critique of geographies of modernization proved effective. By the 1980s, renditions of 'development geography' could not credibly ignore power relations, dependency, and imperialist legacies. At the same time, 'modernity' and 'modernization' were increasingly being seen as multiple, hybrid, and heterogeneous. Radical geographers had pointed to resistance, disjuncture, and difference amidst the enduring power relations of domination and dependency.

Conclusions

Much of the geographical work on issues of development in the early post war period was either a more or less direct link to tropical geography, or part of the vanguard of positivism and quantification. Notwithstanding some work on geographies of modernization that gave this a special character, the legacy of descriptive tropical geography and its association with determinism and other passé modes of geographical thought and analysis has proved enduring. As the 1960s moved on and revolutionary pressures in the South accelerated (epitomized by the insurgencies in Vietnam and the Portuguese colonies and the lurch into Mao's 'cultural revolution' in China), some radical geographers began to embrace them, years before a wider development geography (or indeed wider human geography) was recast as 'radical'. The interventions of radical geographers were beginning to show that the prescriptions issued to 'tropical' and 'backward' countries on how to 'catch up' were not a neutral, rational enterprise that was untouchable, but rather something that was very much caught up with the strategic political interests of the US during the Cold War.

For a few short years development geography was infused with the sort of spatial algebra and 'mathematical mystification' then driving the 'new geography', and some development geographers did begin to map out the optimistic logic embodied in the linear stage theories of the modernization school. However, this marriage of statistics and space was short lived as it became apparent that this perspective was blind to the historical reproduction of systems and processes of inequality. In many ways conventional understandings of the emergence of spatial science and other subdisciplinary areas of geography are interesting for all the places not included (e.g., Africa, Asia, or Australasia). Thus standard histories of the discipline focus on the movements of texts, ideas, and individuals between key nodes in Europe and North America but focus much less on the making of the discipline in Ibadan or Durban, Singapore or Rangoon (for example). Even the brief and partial account offered here indicates the consequences and merits of beginning to take into account the margins in histories of twentieth century geography.

Unravelling the history of dominant ideas of the tropics reveals a process of circulation in which plants, people, and ideas migrate from one part of the world to another. Thus an understanding of Western modernity and its economic, cultural, and political geography can be creatively 'deepened' by engagement with the norms of what are usually scripted as its margins. Considering the role of marginal or peripheral spaces and peoples in allowing geographers to represent the world and its changing political and economic systems can therefore enhance the story of tropical and later development geography. More work on the geography of the idea of tropicality is also required. This involves greater attention to the ways in which European conceptions of the tropics may have been shaped by interactions with indigenous peoples and places. Typically the story of tropical geography has focused largely on white, male, Western academics and much less has been said about the role of women or non Western geographers in shaping debates in geography about the tropics and development and in defining the nature of its methods and agendas. Thus there remains considerable scope for more sustained and direct engagements with feminist readings of the history of these subdisciplinary specialities and with 'postcolonial' approaches to thinking about the collaborations and 'partnerships' that were formed between Western and non Western geographers in the new universities of Africa, Asia, and Latin America from the 1960s onward. Additionally, it is worth recognizing that in the formulation of European visions of the tropical world experiences of disorientation, uncertainty, and novelty had their place.

It is necessary then to understand how geography and geographers created a space in which only certain things could be said or even imagined about tropical 'others'. It is this notion of the unique and distinctive space(s) and processes of the tropics that is the ghost of development geography today. Those geographers who specialize in what are known as 'peripheral' areas are often seen (even today) to belong to some distant relative of the discipline, to area specialities, which mark their contributions out from other parts of geography. Thus many themes of interest to contemporary development geographers (e.g., globalization, neoliberalism, and postcolonialism) should also be seen as fertile grounds for subdisciplinary inter action. Many of the academic researchers who went to study the 'problems' of the tropics located themselves in what we might refer to as the realm of the 'nonproblem'. In this way they located themselves in privileged epi stemological positions outside the problems they studied. Thus, few geographers saw themselves as implicated in colonialism and the underdevelopment they sought to write about and map. With the radicalization of geog raphy in the 1960s and 1970s geographers began to adopt different and more critical ways of looking at colonialism and its legacies in the 'Third World'. In trying to excavate the genealogies of tropical and development geography, it is therefore necessary to understand the geopolitical, economic, and social traumas of decolonization.

The biographies of geographers involved in the rad icalization of the discipline during this period can act as a mirror to disciplinary (and other) histories and the study of movements, places, and departures and as a counter point to the more familiar big stories (metanarratives). There is a danger however of representing this radical ization of geography as a straightforward and simplistic undertaking when actually it was a slow process given the 'ostrich like' behavior of some (liberal) geographers who buried their heads in the sand and refused to engage with politics or the politicization of geography. Thus, writing in 1977, one radical geographer, Keith Buchanan, noted how the term 'development' had become a 'dirty word' and how geographers were increasingly seeking to build critiques of the generation of 'formula fixated' and 'model making technocrats' based in the 'white North'. What was important about the work of radical geog raphers was that earlier, descriptive geographical ap proaches to tropical areas were taken to task for the way in which they sought to 'explain away' the supposed backwardness of tropical zones in an apolitical fashion which was at core, deeply racist and shot through with the language of earlier imperial endeavors. Rather than an objective, detached intellectual endeavor, schol arly discourses on tropical regions, and 'development' were imbued through and through with the imperial representations that preceded them. The concept of the tropics while not always geographically accurate (any more than the terms 'Non West or the South'), thus re calls a telling complex of metaphors, assumptions, and representations inherited from the empire.

See also: Eurocentrism; Modernization Theory; Orientalism; Third World.

Further Reading

Arnold, D. (2000). 'Illusory riches': Representations of the tropical world 1840 1950. *Singapore Journal of Tropical Geography* 21(1), 6 18.

Bowd, G. and Clayton, D. (2005). French tropical geographies: Editors introduction. Singapore Journal of Tropical Geography 26(3), 271 289.

Buchanan, K. M. (1977). Reflections on a 'dirty word'. In Peet, R. (ed.) *Radical geography: Alternative viewpoints on contemporary social issues*, pp 6 30. London: Methuen and Co Ltd.

Claval, P. (2005). Colonial experience and the development of tropical geography in France. *Singapore Journal of Tropical Geography* 26(3), 289 303.

Driver, F. (2001). *Geography militant*. Oxford: Blackwell.

Driver, F. and Martins, L. (eds.) (2005). *Tropical visions in an age of empire*. Chicago: University of Chicago Press.

Driver, F. and Yeoh, B. S. A. (2000). Constructing the tropics: Introduction. *Singapore Journal of Tropical Geography* 21(1), 1 5.

Farmer, B. H. (1973). Geography, area studies and the study of area. *Transactions of the Institute of British Geographers* 60, 1 15.

Farmer, B. H. (1983). British geographers overseas, 1933 1983. *Transactions of the Institute of British Geographers* 8, 70 79.

Forbes, D. K. (1984). *The geography of underdevelopment*. London: Croom Helm.

Gourou, P. (1940). *La Terre et L'homme en extrême Orient*. Paris: Armand Colin.

Gourou, P. (1953). *The tropical world*, Laborde BD (trans). London: Longman.

Livingstone, D. (1992). *The geographical tradition: Episodes in a contested enterprise*. London: Blackwell.

Peet, R. (1977a). Introduction. In Peet, R. (ed.) *Radical geography: Alternative viewpoints on contemporary social issues*, pp 1 4. London: Methuen & Co Ltd.

Power, M. (2003). *Rethinking development geographies*. London: Routledge.

Power, M. and Sidaway, J. D. (2004). The degeneration of tropical geography. *Annals of the Association of American Geographers* 94(3), 585 601.

Pratt, M. L. (1992). *Imperial eyes: Travel writing and transculturation*. London: Routledge.

Said, E. (1978). *Orientalism*. London: Penguin.

Schwartz, J. and Ryan, J. R. (eds.) (2003). *Picturing place: Photography and imaginative geographies*. London: I B Tauris.

Tuan, Y.-F.

T. Cresswell, Royal Holloway, University of London, Egham, UK

Glossary

Topophilia Specifically the love of place. More generally, the emotional connections between human beings and the physical environment.

Tuan, Y.-F. (1930–)

Yi Fu Tuan was born on 5 December 1930 in Tientsin, China. He was the son of a Chinese diplomat who formed part of a small middle class cohort of well educated and well travelled men in China at the time. In 1941 his family moved to Australia and Tuan attended an English speaking school modeled on an English public school. In 1946 the Tuan's were posted to London. Yi Fu Tuan was admitted to Oxford in 1948. In 1949 the People's Re public of China was established and the Tuan family – supporters of Chiang Kai shek's Republic of China (Taiwan) – found themselves effectively exiled. In 1951 Tuan moved to Berkeley to begin his postgraduate career. Here he was exposed to the work of, among other people, Carl Sauer. He received his PhD in 1957. Tuan's first academic job was at Indiana University (1956–58). This was followed by stints at the Universities of Chicago (as a postdoctoral fellow in statistics), New Mexico (1959–65), and Toronto (1966–68) before arriving at the University of Minnesota in Minneapolis in 1968 where he was to spend 15 years before moving to the University of Wis consin–Madison in 1983. In 1985 he became both Vilas Research Professor and John Kirkland Wright Professor. He retired in 1998 and became Professor Emeritus.

Tuan's early interests were in desert landforms and, particularly in pediment formations. His first published paper was 'Types of pediment in Arizona' which was published in the *Yearbook of Pacific Coast Geographers* in 1954. This was followed by the first of many book length manuscripts, *Pediments in Southeastern Arizona*, in 1959. Tuan's interest in desert landscapes was far from simply scientific, however, and from a very early point in his career he began to ruminate over the esthetics of the desert (a landscape he was later to confess a deep and abiding affection for in his 1999 autobiography, *Who Am I*: *An Autobiography of Emotion*, *Mind and Spirit*). Interspersed with scientific writing on pediments was a paper on the 'Use of simile and metaphor in geographical description' published in the *Professional Geographer* in 1957. Highly unusual at the time, this work signaled the future dir ection of Tuan's work. By the mid 1960s Tuan's published work had taken a decidedly reflective and humanistic turn including papers published in the maverick geographer J. B. Jackson's influential journal *Landscape*.

During his time in Toronto Tuan developed a dis tinctive and original voice in geography, ruminating on philosophical issues of environmental perception that would lead to some of his most influential writing. His paper 'Geography, phenomenology and the study of human nature' published in the *Canadian Geographer* in 1971 was one of the key moments in the development of 'humanistic geography' in the 1970s. Along with geog raphers such as Anne Buttimer, Edward Relph, David Seamon, Donald Meinig, and David Lowenthal, Tuan – inspired by the philosophical approaches of phenomen ology and existentialism – sought to counter an overly scientific and quantitative spatial science (then hege monic) with a more human centered approach focused on what it is like to be a human in the world. The human senses (not just sight), emotion, imagination, and feeling were all central to this endeavor.

Tuan's two most influential books were published in the 1970s – *Topophilia* in 1974 and *Space and Place: The Perspective of Experience* in 1977. In these books Tuan revealed the depth of human attachment to the world. The term topophilia describes the various emotional at tachments between human beings and the natural world. It is this attachment that Tuan would later point to as the central issue for all of geography in his paper 'A view of geography' in 1991. Here, he described geography as the story of the Earth as the home of people. For Tuan, the central term is home – a word which, to him, describes the way people make the world meaningful and become attached. This attachment is also apparent in *Space and Place* where place is defined as a center of meaning and field of care – a pause in an endless field of space. This definition exceeds conventional notions of place as lo cation or locale and suggests that place can be found at all scales from a favorite chair to the whole planet as seen from space. This definition of place has become hugely influential in human geography and well beyond. Phil osophers, artists, performers, and literary theorists, as well as social scientists have drawn on and developed the notion of place defined by Tuan. It is his most important contribution to geography. In these books and earlier papers Tuan developed a unique writing style that re flects on philosophical issues while remaining accessible to a large, educated lay audience.

The sense of connection between people and world continued to inspire Tuan's work. In *Dominance and Af fection: The Making of Pets* (1984) Tuan prefigured later work inspired by Michel Foucault by reflecting on how humans twist the natural world into recognizable form through simultaneously dominating it and having af fection for it. Examples include pets, gardens, and water. In *Passing, Strange and Wonderful: Aesthetics, Nature, Culture* (1993) Tuan considers the role of esthetics in the re lations between people and place. He notes that even the most dull, utilitarian, and functional object (a drainpipe or a paper cup) has been made human by the addition of some small decorative motif. This he sees as evidence for the central role of esthetics as a driving force for human life – in many ways more important than even food and shelter. Most recently Tuan has considered his own life as a Chinese boy who became a well known cosmo politan Chinese American scholar in his autobiography *Who Am I* (1999). Here he reflects on his love of geog raphy (his 'salvation') born in the deserts of the American Southwest as well as the emptiness he perceives at the heart of his personal life. His collection of letters, *Dear Colleague: Common and Uncommon Observations* (2002) re veals Tuan at his most unbuttoned and unconventional. Many of the themes of his work are succinctly sum marized in *Place, Art and Self* (2004), an essay accom panied by photography that considers the role of the arts (from painting to dance) in developing the attachment between self and place. This book continues his dedi cation, noted in the final line of *Space and Place* to 'in crease the burden of awareness'.

During his career Tuan has received numerous honor and awards including the LaurŽat d'Honneur of the International Geographical Union, the Cullum Geo graphical Medal of the American Geographical Society, and the Journal of Geography Award of the National Council for Geographic Education. He is a fellow of the British Academy, the American Academy of Arts and Sciences, and of the American Association for the Ad vancement of Science. He has been a John Simon Gug genheim Fellow, a Fulbright Hays Senior Scholar to Australia, a Distinguished Visiting Professor at the Uni versity of California at Davis, and Harry Lyman Hooker Distinguished Visitor at McMaster University, Canada. In 1992 he was awarded the title of Best Professor by the Wisconsin Student Association.

The author of 16 books and numerous papers and contributions, his work has been translated into eight languages.

See also: Humanism/Humanistic Geography; Literature; Phenomenology/Phenomenological Geography; Place; Sense of Place; Space I.

Further Reading

Tuan, Y. F. (1957). Use of simile and metaphor in geographical description. *Professional Geographer* 9, 8 11.
Tuan, Y. F. (1971). Geography, phenomenology and the study of human nature. *The Canadian Geographer* 15, 181 192.
Tuan, Y. F. (1974). Space and place: Humanistic perspective. *Progress in Geography* 6, 211 252.
Tuan, Y. F. (1974). *Topophilia: As Study of Environmental Perception, Attitudes, and Values*. New York: Prentice Hall.
Tuan, Y. F. (1976). Humanistic geography. *Annals of the Association of American Geographers* 66, 266 276.
Tuan, Y. F. (1977). *Space and Place: the Perspective of Experience*. Minneapolis: University of Minnesota Press.
Tuan, Y. F. (1984). *Dominance and Affection: The Making of Pets*. New Haven: Yale University Press.
Tuan, Y. F. (1989). *Morality and Imagination: Paradoxes of Progress*. Madison: University of Wisconsin Press.
Tuan, Y. F. (1989). Surface phenomena and aesthetic experience. *Annals of the Association of American Geographers* 79, 233 241.
Tuan, Y. F. (1991). A view of geography. *Geographical Review* 81, 99 107.
Tuan, Y. F. (1991). Language and the making of place: A narrative descriptive approach. *Annals of the Association of American Geographers* 81, 684 696.
Tuan, Y. F. (1993). *Passing Strange and Wonderful: Aesthetics, Nature, Culture*. Washington, DC: Island Press.
Tuan, Y. F. (1999). *Who Am I? An Autobiography of Emotion, Mind, and Spirit*. Madison: University of Wisconsin Press.
Tuan, Y. F. (2002). *Dear Colleague: Common and Uncommon Observations*. Minneapolis: University of Minnesota Press.
Tuan, Y. F. (2004). *Place, Art, and Self*. Santa Fe, NM: Center for American Places and University of Virginia Press.

U

Uncertainty

M. F. Goodchild, University of California, Santa Barbara, CA, USA

Glossary

Adjustment Theory The theory of errors in survey measurements.

Frequentist Holding to the view that probability must be defined through experiment, as a proportion of events.

Fuzzy Set A set defined by membership in a specific class, with membership allowed to be partial.

Gaussian (or Normal) Distribution A description of the relative sizes of errors in repeated measurements, a mathematical function resembling a bell.

Geostatistics The theory of regionalized variables, or spatial variables subject to spatial dependence.

Monte Carlo Simulation Simulation of multiple realizations of a statistical model, for example, by tossing a coin.

Rough Set A set defined by a mix of full and partial membership.

Subjectivist Holding to the view that probability can be defined subjectively, as a value that determines individual behavior.

Taylor Series Expansion A form of approximation of a nonlinear function.

Tobler's First Law Nearby things are more similar than distant things, a widely observed tendency of geographic information.

Introduction

As humans we experience only a very small fraction of the geographic world. While seasoned travelers may claim to have visited all of the 50 US states, or a large proportion of the world's countries, or all of its continents, even they will have sensed directly only a fraction of the Earth's 500 000 000 sq km of surface. In the temporal domain our experience is even more limited, as none of us can expect to live much longer than 100 of the Earth's 5 000 000 000 year history. All other knowledge about the geographic world must come to us through representations of various kinds, in the form of text, speech, images, digital data bases, or physical models. Moreover, such representations must inevitably be generalized, approximated, abstracted, or subject to measurement error. It follows then that virtually all knowledge of the geographic world is subject to uncertainty. No representation can provide more than an approximation to the truth, forcing geographers always to look at the world through a fuzzy, clouded, and dis torting lens. The advent of digital technology and powerful geographic information systems (GISs) has changed this not one bit, though the precision of com puting devices often encourages a false sense of accuracy on the part of their users.

It is clearly in the interests of users of representations to understand something of their inherent uncertainty. Researchers are interested in knowing something about the uncertainty of the results of their research; decision makers should be aware of the uncertainties surrounding their decisions, particularly if they expect to be held le gally or administratively accountable; and the general public should be aware of the mistakes that can result from overly precise interpretation of geographic infor mation. Thus, research on uncertainty has tended to address four specific issues:

1. the measurement of uncertainty, expressed as prob abilities, confidence limits, or qualitative judgments;
2. the modeling of uncertainty within the theoretical frameworks of fuzzy and rough sets, spatial statistics, and geostatistics;
3. the visualization of uncertainty, in order to convey knowledge about uncertainty to the user of infor mation in intuitively straightforward ways; and
4. the propagation of uncertainty, tracking the effects of uncertainty as information is processed and manipu lated, and expressed as measures of confidence in results.

Humans are inveterate explorers, whether as hunter gatherers searching for new food sources and reporting their discoveries to the band, or as earlier Europeans setting out to explore and conquer new worlds, or as modern cavers searching for previously unexplored passages. The notion that exploration can be complete, that every needle in the geographic haystack can eventually be found, has surfaced at various times in human history. In the late nineteenth century, for example, it was widely believed that the few remaining areas of the planet – the interior of Africa, the Antarctic continent – would eventually be explored and mapped; by the 1960s space was being described as the final frontier; and some years later the US Geological Survey completed its national topographic program of mapping the 48 contiguous states at 1:24 000. Of course, all of these efforts leave representations that are uncertain; but it is far more compelling to think about exploring the surface of Mars, or finding unexplored cave passages, than to update or improve the detail of existing mapping. Thus, the US is currently anticipating a massive program of human exploration of Mars, while allowing its national topographic maps to become increasingly out of date; their inherent uncertainty about the world they purport to represent is increasing rather than decreasing.

The following sections follow the outline above: first, methods for describing and measuring uncertainty; second, theoretical constructs that frame the modeling of uncertainty; third, recent developments in visualizing uncertainty in geographic information; and fourth, methods for the propagation of uncertainty.

Describing and Measuring Uncertainty

Two methods are often used to express uncertainty in common parlance: probabilities, as in such statements as 'there is a 40% chance of rain today'; and confidence limits, as in 'there is a 2% margin of error on this opinion poll result', or 'this thermometer has an accuracy of 0.5 degrees'. Both are grounded in extensive conceptual and theoretical frameworks. Statements of probability may be interpreted in one of two ways: in a frequentist context as the results of real or imagined experiments; and in a subjectivist context as relative expressions of likelihood. For example, a frequentist might interpret a 40% chance of precipitation as "on 2 out of 5 days with atmospheric conditions similar to this it rained," or as "40% of the area covered by the forecast will experience rain." A subjectivist, on the other hand, would understand that rain is more likely than if the probability was given as 20%, and less likely than if it was given as 60%.

Confidence limits are associated with measurements, and express the uncertainty resulting from the use of a specific instrument. The theory of errors asserts that under very general conditions, repeated measurements of the same phenomenon using the same instrument will follow a Gaussian distribution defined by a mean and standard deviation. If the mean is different from the true value, the measurement is said to be biased, and the root of the mean squared error is known as the standard error. These concepts are routinely applied to the measurement of position using the Global Positioning System (GPS), to the measurement of elevation, and to many other common forms of measurement. The theory of errors also addresses the impact of errors on the results of calculations from measurements.

Several issues make this simple picture inappropriate for geographic information. First, and perhaps most importantly, the equating of uncertainty and error assumes the existence of a truth, since error is defined as the difference between a measurement and its true value. While this may be a reasonable assumption in some domains, it is rarely appropriate in a geographic context, particularly in human geography. Even a task as apparently straightforward as the measurement of latitude and longitude with a GPS unit is fraught, because of widespread variation in the definitions of those quantities, variations in the Earth's axis, and the effects of tectonic movements and Earth tides. The classification schemes often used in mapping land use or land cover are based on inherently vague definitions, such that two observers, however experienced, cannot be assumed to share precisely the same definitions of urban, suburban, or rural, and are almost certain to produce different mappings of the same area.

Although the early literature on uncertainty in geographic information was dominated by concepts of error and accuracy, the literature of the past two decades has increasingly avoided these terms, preferring uncertainty as the umbrella term, and making frequent reference to imprecision, vagueness, and fuzziness. Traditional approaches to mapping have emphasized so called Boolean classification, in which every point on the landscape must be assigned to exactly one soil class, or exactly one land cover class. However, more recent work has opened the way for partial or fuzzy membership, by assigning degrees of membership to several classes. For example, a point near the boundary between the urban core and the suburbs might be assigned a 40% membership in the urban class and a 60% membership in the suburban class; the boundary between the two classes would no longer be an infinitely thin line, but a zone of transition.

Second, the theory of errors assumes, quite reasonably, that repeated measurements are made independently. While this is a reasonable assumption in measuring temperature with a thermometer, it flies directly in the face of the pervasive property of geographic data described as Tobler's first law: nearby things are more similar than distant things. Consider, for example, the measurement of a street's location using a GPS unit, and suppose that a vehicle equipped with the GPS is driven along the street, recording location every second. The horizontal accuracy of the GPS is known to be 5 m. If each second's measurement is independently subject to an average error of

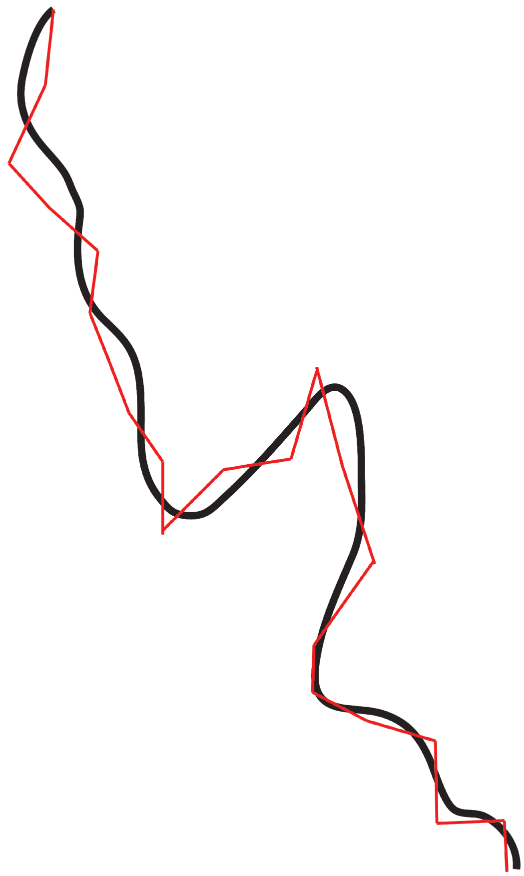

Figure 1 An illustration of the effects of independent errors in the representation of a complex geographic feature such as a road.

5 m, the result would be a very jagged track, much longer than the true length of the street (**Figure 1**). In reality, errors in GPS measurements persist over extended periods of time, and GPS tracks are automatically smoothed by the unit's own software. As a result, the shape of the track bears an acceptably close resemblance to the street's true shape, and its length is only slightly longer than the true length. This property of positively correlated errors is very common in geographic data; if it were not, many of the tasks routinely performed by GISs and other technologies would be much less successful than they are. Together with mean and standard error, spatial correlation ranks as one of the most important measures of uncertainty in geographic information.

Much effort has been expended in recent years in devising standard ways for describing the uncertainties that are known to be present in geographic datasets. By the early 1990s five dimensions of quality had been identified, and written into standards by various national and international bodies:

1. positional accuracy, summarizing the differences between positions recorded in the data and the corresponding true positions on the Earth;
2. attribute accuracy, summarizing the differences between recorded properties of geographic features and corresponding true properties;
3. logical consistency, a measure of the degree to which the contents of the dataset match the rules used to define it (e.g., are the boundaries of area features truly closed?);
4. currency, a measure of the degree to which the dataset is complete and up to date, or the degree to which its contents match the real world at the stated time of validity; and
5. lineage, a history of the actions and processes by which the dataset was compiled.

The emphasis in these standards is typically on truth in labeling: the statements should simply record what is known, rather than referencing some threshold standard of quality. What is known might be quantitative, but it might also be qualitative and to some degree subjective. Typically such statements will be assembled by the producers of data, who are likely to be staff members of the government agencies that have traditionally dominated the production of geographic data. But the development of powerful tools over the past two decades has meant that increasing quantities of geographic information are now available from local agencies and even individuals, who may be much less likely to spend the time needed to determine and document quality.

These standards for the description of data quality have become an accepted part of metadata, or the data used to describe datasets and to assess their fitness for a particular application. Today many large websites provide catalogs of geographic datasets, allowing users to specify requirements, search the catalog, and in many cases to retrieve and use selected datasets in their own work. Data quality plays a very important role in this process, by allowing the searcher to assess whether levels of uncertainty are above or below the appropriate thresholds for a given application. Contemporary tools allow metadata to remain attached to a dataset as it is retrieved, providing an effective means of documentation.

Modeling Uncertainty

A large number of models of uncertainty in geographic information have been proposed over the past four decades. In essence, these models allow the results of limited analyses to be generalized, as for example, when the manufacturer of a thermometer asserts that the accuracy of the instrument is 0.5 degrees, based on a small number of sample measurements. Models also allow for the calculation of related properties, such as the impact of an uncertain slope measurement on predictions of soil loss. Most usefully, models allow for the simulation of uncertainties, through the generation of a number of alternative realizations of the model, each of which might be the truth, but showing a degree of variation that matches the known level of uncertainty. Such simulations have become a powerful basis for visualizing uncertainties (see next section), and for investigating the influence of uncertain data on the results of analysis (see final section).

A simple example is shown in **Figure 2**. Each of the four corner points of this parcel of land has been surveyed, with an uncertainty that is a small fraction of the length of each side. Uncertainty can be modeled by assuming that both coordinates of each surveyed point are disturbed by an error drawn from a Gaussian distribution. The figure shows two simulated realizations of this model.

In surveying, the well developed body of theory known as adjustment allows practitioners to make use of known levels of measurement error in their instruments, and to assess the accuracy of positions determined from those measurements. Adjustment theory is based on the theory of errors and the Gaussian distribution. More broadly, however, this rigorous approach is problematic for geographic information for the reasons discussed earlier. Geostatistics, or the theory of regionalized variables, addresses directly the spatial dependence that is endemic to geographic information, and has provided a very powerful basis for modeling uncertainty.

Finally, the theory of fuzzy and rough sets has found an appropriate home in the study of uncertainty in geographic information, particularly in addressing issues related to vaguely defined classifications. The notion of replacing Boolean or hard classification with measurements of membership is very appealing, and has been adopted enthusiastically in several areas. Rough sets provide something of a common ground between Boolean and fuzzy classification, and are also finding applications in this domain. The lack of a rigorous way of dealing with spatial dependence in this framework is a major barrier to greater progress, however.

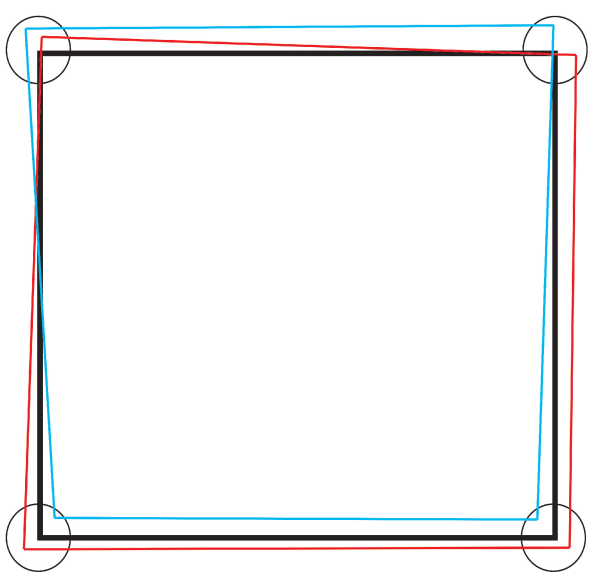

Figure 2 Two realizations of a model of uncertainty in the four surveyed corner points of a simple land parcel.

Visualizing Uncertainty

Cartographers have long been concerned with the portrayal of certain specific types of uncertainty on maps. The depiction of mythological beasts on early maps and globes and white areas in the center of Africa are instances of practices that over the years have become less and less common as the geographic world has become better known. Nevertheless, dashed lines still denote uncertainty in some national boundaries (e.g., in the Arabian Peninsula) and in intermittent watercourses.

Today's paper maps and atlases show little regard for the uncertainty that is almost universally present in geographic information. The width of contours on topographic maps reflects the width of the pen that drew them, not uncertainty over their positions. Maps showing classifications almost always employ Boolean methods, and show boundaries between classes as a single pen width. While this might be acceptable as long as printed maps had to be drafted using pens, the digital world is far more flexible in the options it provides for cartographic design. Colors and shading can be continuously graded, line widths can be varied, and assorted methods can be used to convey the uncertainty message.

Over the past two decades researchers have investigated several methods. Maps have been produced with varying line widths and continuous variations of shading and color. Features have been blurred to indicate uncertainty about their positions, and colors have been grayed to indicate uncertainty about classifications or measured attributes. All of these methods are based on assessments of uncertainty for individual features, however, and fail to address the spatial dependence that is almost always present. For example, blurring of points may indicate uncertainty in their positions, but may mislead the user who is interested in knowing the distance between the points, because if both points are misplaced by the same error in the same direction, the distance computed between them will be perfectly accurate.

Addressing this problem has proven particularly difficult. While many models exist that incorporate spatial dependence, including geostatistical models, it makes little sense to attempt to describe uncertainty by specifying the parameters of such models, when few users will have a sufficient statistical background to interpret them appropriately. Instead, researchers have experimented with various forms of animation. In the case of positional error, for example, animation allows the features to be shown dancing either independently or in a correlated fashion, clearly demonstrating the difference between absolute and relative positional error, and its implications for analysis.

Propagating Uncertainty

As tools for processing geographic information have become more sophisticated, it has become possible to evaluate the extent to which the results of analysis are affected by known uncertainties in the underlying data. What, for example, is the impact of known errors in census population counts on forecasts of future school populations, or analyses of social deprivation in neighborhoods? Such questions are particularly fraught when they involve changes of scale, or changes in the zones used to report statistics and to conduct analysis.

Significant research progress has been made in our understanding of the propagation of uncertainty, using tools ranging from differential calculus to Taylor series expansion to Monte Carlo simulation, and powerful packages are now available to analyze the effects of uncertainty on many standard manipulations of geographic information. Some progress has also been made on automating the modification of metadata as datasets are manipulated, in order to create data quality statements for the output of operations.

This work on uncertainty propagation has led to some useful conclusions. First, many of the operations performed on geographic information are essentially nonlinear in their responses, leading to surprisingly large impacts from comparatively modest uncertainties. Missing a key feature in a geographic dataset, for example, can sometimes have dramatic consequences. Second, in other instances the impacts of uncertainty are remarkably low, a point that has already been made in the context of Tobler's first law and the estimation of such simple properties as distance or shape – the distance between two features may be estimated very accurately despite large uncertainties about the features' positions.

Conclusion

Uncertainty is in some ways the Achilles heel of modern geographic information processing. Despite the apparent precision of computing systems, and the use, for example, of double precision (15 significant digits) in the representation of many aspects of the geographic world, the actual accuracy of geographic information and the products of processing may be alarmingly low. Several instances have surfaced in recent years of the potentially disastrous and expensive consequences of uncertainty, ranging from the severing of a cable in the Italian Alps by an aircraft (the cable was not marked on the maps available to the pilot) to disputes over land ownership that result from poor quality surveying. Some of these have resulted in expensive court settlements, when it became apparent that the users of geographic information technologies had not taken adequate note of the known uncertainties in their data.

Despite extensive research over the past two decades, this message still has not filtered through into the design of maps and GIS, and comparatively simple steps to reveal and evaluate issues of uncertainty have not been taken. It would be simple, for example, to provide estimates of the uncertainties associated with measurements of area, when these are made from representations with known levels of positional error. In reality, uncertainty is a messy concept that decision makers would often rather sweep under the carpet, and its successful analysis often requires an understanding of sophisticated concepts of probability and correlation. Nevertheless, any knowledge of uncertainty, however crude, is better than none, and the consequences of ignoring it can be severe.

See also: Error (Propagation and Modeling); Fuzzy Set and Fuzzy Logic; Global Positioning/GPS; Representation and Re-presentation.

Further Reading

Guptill, S. C. and Morrison, J. L. (1995). *Elements of Spatial Data Quality*. Oxford: Elsevier.
Heuvelink, G. B. M. (1998). *Error Propagation in Environmental Modelling with GIS*. Bristol, PA: Taylor and Francis.
Isaaks, E. H. and Srivastava, R. M. (1989). *Applied Geostatistics*. New York: Oxford University Press.
MacEachren, A. M. (1995). *How Maps Work: Representation, Visualization, and Design*. New York: Guilford.
Monmonier, M. S. (1996). *How to Lie with Maps*. Chicago, IL: University of Chicago Press.
Shi, W., Fisher, P. F. and Goodchild, M. F. (eds.) (2002). *Spatial Data Quality*. New York: Taylor and Francis.
Zhang, J. X. and Goodchild, M. F. (2002). *Uncertainty in Geographical Information*. New York: Taylor and Francis.

Underclass

C. Haylett, The University of Manchester, Manchester, UK

Glossary

Classifying Practices A way of ordering the social world by positioning a subject in a wider structure of social meaning, status, and value.

Discourse A social practice in which selective words, images, and ideas are used to represent a subject and establish particular parameters of understanding.

Postindustrial Society A stage in the development of advanced Western societies in which manufacturing industries no longer form the primary base of economic and social organization.

Power/Knowledge Related to discourse, a concept originated by the French philosopher Michel Foucault to explain how power works by creating expert knowledge of a subject.

Sociospatial Polarization A social process in which inequality works through social space to create geographical distance, dislocation, or boundaries between social groups.

Working-Class Cultures Everyday ways of being that are shared by working class groups of people living in conditions of social and economic inequality.

Definitions/Introduction

Underclass is a concept rather than a fact. This means, it is an idea or series of ideas about certain groups of people rather than an uncontested truth. There is no general consensus as to whether underclass is an appropriate concept to describe social groups that are diversely composed and could otherwise be regarded as part of the working class in advanced Western societies. These groups include the poor and unemployed, those dependent on state benefits, single parent families, and inhabitants of poor urban areas, where these characteristics are long term or intergenerational in duration. Contemporary debates in fields of politics, academia, and the media question whether an underclass exists, how to define it, how to account for the characteristics with which it is associated, and how to reform or remove it from society. These questions are extensive and contested. Many people regard underclass as a disparaging term that casts certain groups of people and the places they inhabit as less worthy and human than others. They argue that the term obstructs real understandings about the causes and consequences of class inequality and poverty. In this vein, it is suggested that the use of underclass by powerful social groups (especially politicians, academics, and the media) says more about their own class position and perspective than it does about the social groups it is meant to describe. Other people regard underclass as an appropriate term to describe people and places exhibiting characteristics of long term socio economic dislocation and cultural entropy, or a collapse in acceptable social norms and values. Given the debates and disagreements surrounding underclass, it should be treated with some equivocation.

In empirical social science, underclass refers to a class stratum in advanced Western societies made up of social groups at the base of the working class, beneath skilled and unskilled manual workers. It is a concept associated with urban areas that have undergone long term socio economic changes since the 1970s, resulting in the loss of traditional working class employment in heavy and manufacturing industries and their forms of community. These areas are often described as postindustrial. The main characteristic of underclass groups is their marginal employment status (long term unemployment and underemployment or employment in informal or black economy sectors) which is statistically related to social and cultural outcomes of single parenthood, dependence on state benefits, antisocial or criminal behavior, sub cultural value systems, and residential immobility. These social groups are said to constitute a permanent subclass, that is, a group of people on the very margins of the class system of capitalism for whom there is little to no likelihood of social or spatial mobility.

Within various definitions and debates surrounding underclass, there is common interest in discussing how poverty relates to culture, in particular, certain values, attitudes, and ways of life that are deemed problematic for the rest of society, such as teenage pregnancy. This raises a question about the relationship between class disadvantage and cultural difference: are the characteristics of so called underclass groups matters of culture or matters of inequality, or both? This question can be seen as the essence of the underclass debate with consequences for how these social groups are socially regarded and targeted by policy. The contentious nature of the question arises because it is partly constructed through moral signifiers, which impute the legitimacy, decency, and respectability of certain ways of life. There is a suggestion that the underclass groups are responsible for the ongoing nature of their poverty due to the poor culture and morality of their group culture and environment. This aspect of the debate has been popularized by media and political representations which are less concerned with how certain groups are disadvantageously caught up in processes of economic and social change than with using underclass as a means to diagnose and dramatize social and cultural ills. Pejorative notions of worklessness and unemployability, anti social and criminal behavior, wel fare dependency, teenage pregnancies, fatherless families, council estate or ghetto criminality, and an absence of traditional morality, proliferate in such representations. In reference to these issues, the term underclass has slipped into casual usage and common knowledge, working as shorthand for social decline and crisis in which the cultural failings of poor, working class people can be extended to meet any requirements for social horror mongering.

Debates about underclass then are debates about cultures associated with different class groups. They raise questions about value, status, and taste, and how different class groups are culturally represented. Discussion of the concept should therefore be accompanied by questions about its use and its usefulness in particular times and places. The next section examines some of the historical and geographical contexts through which underclass has been produced and used as a concept. Through this contextual approach, underclass can be seen as a dis course about class rather than a fact about class; in other words, it is a way of narrating class issues in particular times and places. The following sections will consider the two main explanations of underclass offered in social scientific approaches: first, behavioral explanations and second, structural explanations. The conclusion considers underclass as a discourse produced through historical and geographical relations of power, in particular, powers of representation.

Historical and Geographical Contexts

The provenance of underclass as an idea is historical. It shares its cultural history with the 'dangerous classes' of mid nineteenth century industrial England, the 'casual poor' and 'undeserving poor' of the late nineteenth century, and Karl Marx's 'lumpenproletariat' or residuum of capitalism's class structure. Contemporary use of the concept modernizes these cultural labels in important regards, but continues to position the social groups to which they refer as outside, beyond, or beneath normal society and to cast them as a cultural or economic drag on the traditional working classes. There are also his torical continuities in the way underclass is constructed in the popular imagination as an unknown entity, an object of fear and fascination and of political intervention and reform. In geographical terms, underclass remains a strongly Western and urban concept: the modern in dustrial city is the place of underclass. Despite national variation in definitions and use of underclass, the idea commonly carries stigma for poor groups and has greatest currency in countries with an industrial history of class relations, welfare state development, and prob lems of socioeconomic change associated with cycles of capital disinvestment. In general, this has meant the underclass debate has been conducted in Europe and North America with North America constructing the debate in terms of behavioral versus structural explan ations of underclass and in many ways setting the polit ical agenda as an issue of welfare reform. Within this basic conceptual and political framework, geographical variations are significant. North American debates con centrate almost exclusively on race and urban issues of black inner city cultures; Britain was the first country in Europe to widely use the concept from the 1970s onward with reference to long term unemployment and welfare dependence; in France, the debate is about the segregated outer estates of large cities and their predominantly ethnic and immigrant populations. While serious debates pertain in relation to these issues, popularized versions of the underclass debate reduce such issues to emotive and insubstantial charges. Here, the geography of underclass is largely figurative. Its territory is 'the slum', 'the street', 'the ghetto', 'the estate', 'the inner city', and 'no go areas' whose misfortunes may be tied to those of urban econ omies but whose symbolism works independently of economics. Commonly, these urban areas are depicted as a problem of urban space *per se* for the threat of social disorder it brings. The threat is about class and the possibility of its suppressed social and psychological conflicts spilling over, represented by an unruly working class city encroaching on civilized, bourgeois urban forms. Since modern industrial cities first developed, commentators have posed 'the urban question' in terms of what cities 'do' to human forms of habitation and development. In this regard, underclass is a powerful image of urban dysfunction used to crystallize and dramatize a range of urban ills (degeneration, destitution, depravity, deviance, and disorder) which diagnose the city and its poor subpopulations as ecologically and pathologically 'sick'. It is an enduring and powerful geohistorical imaginary.

The identification of historical continuities in ideas of underclass is important to avoid too much 'presentism' and being caught up in moral panic and crisis. Following the French philosopher Michel Foucault's archaeological approach to powerful discourses of the social, the dis course of underclass can instead be regarded as 'histor ically present'. This means that both behavioral and structural explanations for the contemporary rise of underclass should be put in geographical and historical context.

A Behavioral Underclass?

In the nineteenth century, ideas of underclass were inextricably part of the psychic or imaginary construction of the newly industrial city. Attempts to order the apparent chaos and threat of massive urban expansion by planning the built environment and socially managing its burgeoning population required both technical expertise and cultural representations of its inhabitants. Cultural historians have described this project of urban social management as a 'will to order' which was, in part, a psychic ordering to manage uncontrollable emotions of fear, hate, desire, fascination, pity, and estrangement in the bourgeois psyche as it confronted the urban poor in its imagination and less commonly in the city street. The sociospatial distance between the different classes of newly industrializing cities was vast and full of latent social conflict as extremes of poverty and wealth developed apace. In the geographical imagination of bourgeois elites, urban space was both uncontrolled and threatening, requiring new knowledge and maps for its navigation. In the mid to late nineteenth century, the urban poor came to dominate social commentaries of bourgeois society across popular print media (such as travel writing), political campaigns for social control and reform, and in emerging social researches. The industrial slums that had developed in the transition to capitalism from the end of the eighteenth century onward became the focus of burgeoning practices of social science. In England, the documentary writings of Henry Mayhew (1861, *London Labour and the London Poor*), Andrew Mearns (1883, *The Bitter Cry of Outcast London*), Charles Booth (1902, *Life and Labour of the People of London*), and William General Booth (1890, *In Darkest England*) were exemplary of the genre. Their narratives were conducted in terms of exploration, discovery, and enlightenment, paralleling colonial stories about the dark continent of Africa and similarly depicting the inhabitants of these unknown territories as savages belonging to another race, and a burden to the western bourgeoisie's march of progress.

The scientific identification and naming of new social groups created by urban industrial development in the nineteenth century was central to social transformations underway. In mid nineteenth century social documentations of the urban industrial revolution, at which England was the heart, those people previously referred to as 'the lower orders' emerged as the 'working classes'. The working classes were divided into two separate classes using a distinction between the respectable and nonrespectable based on a differential account of behavior–environment relations between the two classes. Social observations claimed scientific status using empirical methods of first hand experience and induction and asked whether poor urban environments were creating false behaviors or showing people for what they were. This was an ecological approach to human behavior. If a bad environment was merely obscuring the nature of an otherwise good type of person, then that person was respectable, deserving, and redeemable. In contrast, the character of the nonrespectable was not obscured by a bad environment but revealed by it. Such people chose their way of living and that choice was the product of their nature. They were pathologically given to behave in particular ways. The respectable poor could be observed as clean, industrious, sober, evenly tempered, disciplined, and law abiding, while the nonrespectable were seen to be dirty, slovenly, loud mouthed, undisciplined, and criminal. This kind of categorization was foundational in developing the status of a 'social' science, asserted through scientific modes of empirical observation and discourses of human nature. From the mid nineteenth century onward, ecological and pathological explanations of the relationship between the behavior and environments of the poor became embedded in social scientific modes of inquiry into urban space. They remain powerful in contemporary discourses of underclass.

If we consider nineteenth century representations of urban industrial space as the early part of a modern history of underclass, we can put twentieth century representations of postindustrial urban space in a fuller context. The most authoritative accounts of these urban social transitions are deemed to be those provided by social scientific empirical inquiry. The Chicago School's paradigm of urban sociology has been most influential in this regard, creating a lens through which 'race' and neighborhood became the means of discussing urban social problems at a human level. Moreover, their approach became part of establishing the North American city as a model for assessing present urban conditions and predicting urban futures. In the 1920s and 1930s, the University of Chicago established a tradition of urban research in which the observed practices of marginal social groups became the way to understand the city's problems. This was a time of great social upheaval as immigrant groups and former slaves from the Southern states of America jostled for living space in Northern cities. Theories of urban ecology were used to posit causal relationships between human behaviors, environmental conditions, and spatial organization. Ecological perspectives remain part of contemporary social scientific explanations of underclass in postindustrial urban environments although they are widely criticized for their environmental determinism. Commonly, they are incorporated in multiaspected explanations of underclass, which are given credence by using socioeconomic data and seeming to foreground the importance of culture in understandings of urban poverty. In fact, there is much blurring of understanding about the different roles of environmental, cultural, and economic factors in the creation and perpetuation of urban poverty, and the discourse of underclass has been extremely useful to those who wish to obscure understandings about poverty and its causes in the interests of power. This can be seen by tracing the use and uptake of the term by political, media, and academic elites.

The term underclass was first used by the Swedish economist Gunnar Myrdal in 1962 to describe the position of poor Americans permanently forced beyond the margins of the labor market in postindustrial economies. Although Myrdal's conceptualization was structural, in that it located the cause of the underclass' condition in economic and social forces, it quickly developed into a cultural identifier of racial and ghetto location. By the end of the 1970s, the term had widespread media currency in North America and was primarily behavioral and ecological in meaning: it was about black cultures and the ghetto way of life. Social scientific efforts to anchor the term in sociostructural explanations from the late 1970s onward have been unable to shake off 'the ghetto' as a powerful myth and reality of North American urban history. For example, the work of Professor William Julius Wilson on underclass talks about the 'concentration' and 'isolation effects' of the North American urban ghetto and of 'ghetto specific culture'. His conceptualization of relations between 'race', residence, and sociostructural disadvantage in globalized economies has influenced media and political agendas and is usually seen as a powerful corrective to behaviorist or victim blaming accounts of underclass. However, Wilson's work has also maintained a central idea of spatially bound and generated cultures of poverty, albeit from a liberal or left wing perspective and this has this has proved to be an attractive combination for many social commentators. For others, Wilson's work in the late 1970s and 1980s is specifically about 'race' and segregation in North American cities where racism and sociospatial exclusion have taken virulent form in relation to the country's political–economic history of slavery and urban immigration. For this reason, it is argued that the concept should not be applied freely elsewhere. More contentiously, the work of American writer Charles Murray in the 1980s identified underclass as a group created by liberal social welfare policies encouraging behaviors and attitudes of welfare dependence. This notion had political currency beyond North America, particularly in Britain, in the ultraconservative politics of Thatcherism. Murray's ideas coincided with a political period of heightened welfare retrenchment in North American, Antipodean, and Western European welfare states and the time was ripe for it to become a powerful discourse. Notions of behavioral and environmental poverty (intergenerational, subcultural, internally generated, and/or perpetuated) had been in circulation from the 1960s onward, often referred to as 'the culture of poverty' or 'cycles of deprivation' but they did not have the political ascendancy granted to them by ultraright wing governments in the 1980s in their quest to transform welfare states into market economies. Underclass is therefore a deeply political issue about welfare governance with real consequences for poor groups in welfare states subject to the demands of a globalizing economy. To say underclass is a discourse therefore makes it no less real than if it were a fact in terms of the issues it raises. These are given an alternative perspective in structural explanations of underclass, discussed in the next section.

A Structural Underclass?

Since early modern classifications of the urban poor as 'the casual poor' attempts to stigmatize poor groups as deficient in themselves have been accompanied by attempts to explain their condition as a product of historical development. Historical analysis of the human costs of social and economic transformations has largely discredited behaviorist explanations of poverty. Historical accounts focus on the social imperatives of particular kinds of economic organization and the social outcomes they give rise to, beyond the control of relatively powerless individuals and groups. From the late nineteenth century, Marxist traditions of historical analysis have been most influential in explaining why poverty develops and particular groups are disproportionately affected by it. By studying the workings of capitalist accumulation, Marxist structural analysis shows how poverty is not a matter of happenstance, personal misfortune, or failure, but is inextricably part of the way capitalism works. Capitalism systematically creates inequality and poverty in order to be able to create wealth (capital accumulation). Structural analysis points to the power of economic structures to stratify groups of people in social relations, positions, and situations of inequality which are called classes, most commonly divided into upper, middle, and working classes.

The social sciences have developed basic structural understandings of how capitalism works in various forms of class analysis, which aim to provide advanced western societies with a means to understand their diverse and unequal social outcomes. These outcomes are historically and geographically specific because processes (or rounds) of capital accumulation, commonly called wealth creation, position different classes, regions, and countries in specific and contingent relations with each other. These relations change over time in the quest for new profits. Cyclical crises are an intrinsic part of such processes in which capitalism creates itself anew by casting aside 'old' labor markets and human forms of community. This understanding is central to explanations of underclass as a structural class position.

Since early industrialization, the Marxist notion of a sub working class stratum or 'lumpenproletariat' has described an economic group or 'reserve army of labor' which can be taken up or dropped in times of economic boom or bust. This group represents 'disorganized labor' as opposed to organized (trade unionized) labor and functions to undermine the position of the traditional working classes by threatening to displace them and undercut their wages and conditions of work. They are only ever transiently, informally, or weakly attached to the labor market, but are a crucial part of how capitalism works through processes of stratification. The relevance of this analysis to contemporary ideas of underclass is clear but structural theorists are generally skeptical of the notion and prefer to focus on the social and economic processes that give rise to marginal employment and new forms of poverty. Their starting point is that con temporary changes in the accumulation of capital asso ciated with globalization have created a period of 'flexible accumulation' with new patterns of investment and in dustrial production. While radical changes in the struc ture, mobility, and power of capital have degraded conditions and forms of employment for working class groups, their political management by governments in the West has largely been about accommodating and encouraging 'flexible' employment practices, privatizing much of the public sector, and restructuring the welfare state. The outcome of this political–economic process has been long term unemployment; increased casual, in secure forms of employment; and reduced entitlements to social protection from the welfare state.

Processes of welfare and capital restructuring are lived as conditions of social decline and insecurity for poor working class groups. They are marked by social charac teristics of family and community instability and break down, single parenthood, street crime, marginal social participation, drug abuse, and long term claims on welfare. Structural analysis relates the rise in such social and cul tural characteristics to a radically diminished structure of economic opportunity for some working class groups and sees them as the result of heightened economic margin alization, not its cause. Similarly, the social problems as sociated with poor urban areas are not internally generated as suggested by underclass images of ecologic ally degenerate places populated by morally degenerate people. Rather, they are the expression of processes of sociospatial polarization in which new or intensified social divisions work themselves out through space, for example, through disinvestment in the public sector especially housing and community facilities, and through govern ment failure to intervene in market led abandonment of places. These processes disproportionately affect poor, working class groups but process based approaches are skeptical of the notion of underclass as a distinct social class, having developed as a result of these processes. By focusing on social process and change rather than static distributions of characteristics, there is seen to be much movement, complexity, and change over time within and between working class groups, including poor and welfare claiming groups. This points to the importance of classi fying practices in determining how different class group ings are named, explained, and seen to relate to each other. Classes are not objective, social phenomena waiting to be observed by social scientists as models, patterns, and dis tributions. They are a product of particular ways of seeing, ordering, and representing the social world. In other words, they are produced through discourse. The final section concludes by considering underclass as a discourse.

Conclusion

Much of the debate over whether and why underclass exists rests on an appeal to the authority of empirical observation and positivistic social explanation – if an underclass exists, we will be able to find it, describe it, account for it, and change it. It ignores the extent to which underclass already exists as a regime of meaning and material practice. Underclass is produced by a range of people, institutional practices, and historically formed languages of the social. It is also materially productive, as particular ways of seeing poor groups become the basis of dominant policy formulations, for example, in welfare reform. In this regard, it is important to see that the discourse of underclass is not simply about negative representations of poor people and places but processes of individual and social 'becoming'. The French phil osopher Michel Foucault has talked about these pro cesses as modes of subjectification, whereby subjects are incorporated into expert or powerful knowledges about them. Underclass can be seen as the product of such power/knowledge relations as poor groups are formed and reformed by historical conditions and of work and welfare and the classifying practices on which they are based. These historical forms of production are also al ways geographical or spatial, created by someone, somewhere, in relation to others elsewhere. By studying the historical and geographical relations of the discourse of underclass, we can see how they are power relations and how the power of representation is a productive power with serious consequences for poor groups.

See also: Discourse; Ghettos; Inequality; Representation, Politics of; Social Class; Welfare Reform.

Further Reading

Dikec, M. (2006). Badlands of the Republic? Revolts, the French state, and the question of banlieues (Guest editorial). *Environment and Planning D: Society and Space* 24(2), 159 163.

Dickec, M. (2006). Two decades of French urban policy: From social development of neighbourhoods to the republican penal state. *Antipode* 38(1), 59 81.
Driver, F. (1987). Moral geographies: Social science and the urban environment in mid nineteenth century England. *Transactions of the Institute of British Geography* 13, 275 287.
Finch, L. (1993). *The Classing Gaze*. St. Leonards, Australia: Allen and Unwin.
Foucault, M. (1977). *Discipline and Punish: The Birth of the Prison*. London: Penguin.
Gans, H. (1993). From 'underclass' to 'undercaste': Some observations about the future of the post industrial economy and its major victims. *International Journal of Urban and Regional Research* 17, 327 335.
Giddens, A. (1973). *The Class Structure of the Advanced Societies*. London: Hutchinson.
Hall, P. (1996). The city of the permanent underclass, *in his Cities of Tomorrow*, pp 362 400. Oxford: Blackwell.
Jackson, P. (1983). Social disorganisation and moral order in the city. *Transactions of the Institute of British Geography* 9, 168 180.
Mann, K. (1992). *The Making of an English 'Underclass'?* Buckingham: Open University Press.
Peck, J. (2001). *Workfare States*. New York: Guilford.
Smith, D. (ed.) (1992). *Understanding the Underclass*. London: Policy Studies Institute.
Stedman Jones, G. (1984). *Outcast London*. London: Penguin.
Wacquant, L. (1999). Urban marginality in the coming millennium. *Urban Studies* 36, 1639 1647.
Williams, R. (1973). *The Country and the City*. London: Chatto and Windus.
Wilson, W. J. (1996). *When Work Disappears: The World of the New Urban Poor*. New York: Vintage Books.

Relevant Websites

http://www.chavtowns.co.uk
Chav Towns.
http://www.dwp.gov.uk/welfarereform/
Department For Work and Pensions.
http://www.hbo.com/thewire/
Home Box Office: The Wire.
http://sociology.berkeley.edu/faculty/wacquant
Loic Wacquant Homepage.

Uneven Development

B. Christophers, The University of Auckland, Auckland, New Zealand

Glossary

Comparative Advantage An economic theory explaining why it can be beneficial for two parties to trade, even though one may be able to produce every item more cheaply than the other; the theory relies on the ratios between how easily the two parties can produce different goods, rather than absolute costs of production.

Perfect Competition A hypothetical market form in which no producer or consumer has the market power to influence prices.

Returns to Scale Refers to a technical property of production that predicts what happens to output if the quantity of all input factors is increased by a given amount – if output increases by that same amount, there are constant returns to scale, while if output increases by more than that amount, there are increasing returns to scale.

Introduction

Uneven development is one of economic geography's central themes and can be interpreted very literally: it means, quite simply, that the development of the economy is uneven. This unevenness is seen to exist in various dimensions – different sectors of the economy, for instance, developing to varying degrees and at different rates – but, for geographers at least, 'uneven development' ordinarily implies uneven *geographical* development. This unevenness can be and has been established on the basis of any number of different metrics (employment rates, income levels, rates of economic growth, and so on), and it has been described at all geographical scales – from intra urban disparities all the way through subnational regional differences to uneven international development.

Further to identifying and delineating such unevenness in its various forms, researchers' primary concern has been to seek to explain it. Within the academic discipline of human geography, uneven development has most commonly been understood as an inevitable – or at the very least, highly likely – manifestation of the capitalist mode of production. Drawing to one extent or another on Marxist political economy, writers in this tradition see uneven development not so much as an exception but as the norm, with development in one place directly and ineluctably linked to underdevelopment in another. This article focuses primarily on such work. First, however, it is important to register another tradition of thinking on uneven development (a tradition against which those offering Marxian interpretations have often been writing), which, by contrast, sees uneven development more as a departure from, or exception to, 'normal' economic development. This tradition has made much less of a mark within human geography, but it has been very influential in policy circles and elsewhere in the academy, and many of its practitioners within contemporary economics explicitly style themselves 'economic geographers'. Its roots are in classical political economy, while today it resides at the heart of neoclassical trade and growth theory.

Uneven Development in Classical Political Economy and Neoclassical Economics

Adam Smith and David Ricardo, the two great British political economists of the late eighteenth and early nineteenth centuries, shared a deep concern about the obvious economic disparities between nations. Thus, their respective inquiries into national wealth – its nature and the causes of its growth – became at the same time investigations of uneven international development. Both writers saw contemporary trade flows as the immediate and identifiable cause of such disparities (Smith was optimistic that free(r) trade would ultimately lead to international convergence; Ricardo's prognosis, meanwhile, was more gloomy). However, for both writers, contemporary trade was essentially just exacerbating *existing* inequalities that neither one sought to ascribe to the underlying dynamics of modes of economic production. Instead, these inequalities were pegged to the differing internal characteristics of the territories in question: for Smith, inferior development reflected inferior physical geography, policy, or demography; for Ricardo, it was a function of being a producer and exporter of subsistence goods. In both cases, trade occurring in the context of these existing differences (on 'absolute advantage' grounds for Smith and 'comparative advantage' grounds for Ricardo) was serving merely to amplify those differences.

From the mid nineteenth century until relatively recently, uneven development largely slipped off the radar of mainstream Western economics, as the latter came to be dominated by neoclassical (or marginalist) economic approaches. This lacuna was not surprising: the persistence of uneven development simply does not square with neoclassical growth theory, which predicts eventual income convergence. However, in the last 30 years, a number of economists, dissatisfied with the ability of neoclassical theory to account for uneven de velopment, have sought to address this phenomenon directly. Two broad 'schools' can be identified.

First, in the 1970s and 1980s, economists such as Ronald Findlay revisited the same questions about international trade that Smith and Ricardo had asked two centuries earlier: did it tend to reduce or increase international inequalities? As it had been for Ricardo, the answer proffered was largely pessimistic; the main dif ference was in the methodological approach, which was now highly mathematical, with researchers explicitly building models of so called 'North–South' growth and trade. However, one key similarity remained: uneven development – a basic asymmetry in the stage of devel opment between two regions – was assumed *already* to be in evidence. Few economists asked why such asymmetries had arisen in the first place, and where they did, their appeal to histories of empire and colonialism went largely unnoticed by most of their peers.

Second, since the mid 1980s, an alternative set of economic models has been advanced, in this case pro ceeding from the assumption of spatial symmetry. This is an important difference: these models aim to answer the question of why countries *starting from a comparable stage of development* might develop at different rates. Again, however, the approach is highly mathematical; and again, the focus is on trade. This modeling tradition has grown rapidly in the last two decades, and has thrown into question traditional neoclassical arguments regarding the long term (convergence) benefits of international trade. The main theoretical departure has been the loosening of traditional assumptions of perfect competition and con stant returns to scale. Instead, these models incorporate imperfect competition and, most importantly, increasing returns – the idea that economic processes may be self reinforcing rather than self limiting, which has been seen to be a powerful tool for understanding the emergence of international inequalities. Some of the main researchers in this area, such as Paul Krugman, claim they are practicing economic geography (or a 'new geographical economics'), but their work has had relatively limited influence within human geography.

There is, however, one thread of research within human geography that does bear important parallels and discernible links to the work of Krugman *et al.* Krugman has explicitly sought to tie international trade not to theories of comparative advantage, but to the tendency for individual nations to pursue different economic specializations based on the presence of increasing re turns. However, Krugman – and alongside him, most notably, Michael Porter – has also observed that such specialization has geographical manifestations *within* territories, in the form of clustering of production in specific cities or regions. This, then, is another facet of uneven development for which economists have offered theories: the tendency not only for some countries to develop more rapidly than others, but for certain cities or regions to 'get ahead' and stay there.

Economists have used terms such as 'economies of agglomeration' or 'economies of localization' to explain the geographical clustering of economic activity in spe cific places. In doing so, they have typically appealed to the late nineteenth century work of Alfred Marshall on 'industrial districts', which identified a critical trinity of labor market pooling, supply of intermediate goods, and knowledge spillovers. Economic geographers have car ried out research along similar lines in the last two decades, albeit with a much greater emphasis on histor ical–geographical empirics, as opposed to abstract mathematical modeling. Amongst the main contributors have been Ash Amin and Nigel Thrift – who refer to 'neo Marshallian nodes' – and Michael Storper and Allen Scott. For these writers, debates on development and (therefore) uneven development are incomplete unless cities, regions, and 'city regions' are understood to be active, causal economic entities. Like Porter and Krug man, they claim that such agglomerations maintain competitive advantage due to positive internal (local) economies (though they further insist that economistic readings are not in themselves sufficient, but must be allied to consideration of local institutional, cultural, and political contexts). In addition, they argue, crucially, that globalization has – perhaps paradoxically – only served to intensify such local economies, thus redoubling the significance of regions and aggravating patterns of uneven geographical development.

Uneven Development in Marxist and Other Critical Political Economy

Much of the writing on uneven development that has occurred either within human geography, or which has most strongly influenced human geographers, constitutes a direct critique specifically of the quantitative neo classical models sketched out above – such models, the critique says, are inherently ahistorical; they assume overly simplistic conceptions of geographic space; they downplay structural interrelationships between different regional economies; and they fail to explain why capital in general creates such uneven geographical develop ment, focusing instead on individual capitals and the rarefied market mathematics of exchange, supply, and demand.

In view of the fact that uneven development comes in an infinite variety of forms and exists at all geographical scales, it is not surprising that a wide range of more holistic and historical interpretative approaches have been advanced. These various approaches defy easy differentiation and categorization; they have been developed alongside one another, have frequently informed one another, and share many common themes (as well as differences). However, it is fair to say that most are inspired to one extent or another by Marxist political economy (as opposed to classical political economy or neoclassicism). Moreover, from the perspective specifically of human geography, they can be usefully split into two types. First, there are explanations in which uneven geographical development is effectively identified as a key *outcome* of the underlying social, political, and economic processes and structures that such theories elucidate. Second, there has been work – primarily by human geographers – that seeks to make the geography a far more central component of the theoretical framework: wherein uneven development is seen less as an outcome, and more as an integral, constitutive element of the capitalist mode of production. These two sets of approaches are now considered in turn.

Uneven Development as Geographic Outcome of Capitalism

Much of the most influential thinking on uneven development can be traced to the extensive international debates on development and ('Third World') underdevelopment that ensued from the late 1960s. Inevitably, therefore, such thinking deals almost exclusively with uneven development on an international scale. The antagonist against which this thinking crystallized was mainstream modernization theory deploying multistage economic 'takeoff' models developed by the likes of Walt Rostow. Modernization theory looked at countries largely in isolation and sought to map out future development trajectories for 'underdeveloped' territories based on the 'successful' blueprint of paths previously followed by 'developed' Western nations.

Two main critiques of modernization theory soon developed. The first, 'dependency theory', is perhaps most closely associated with Andre Frank. This argued that development and underdevelopment were in fact two sides of the same coin (uneven development): rich ('core') countries *need* a group of poorer ('peripheral') countries in order to remain rich; and those rich countries perpetuate this state of dependency through their economic, military, and cultural policies. A second key theory was that of 'unequal exchange'. As formulated by Arghiri Emmanuel in particular, this argued – somewhat similarly to the more pessimistic of the classical and neoclassical trade models discussed above – that international trade would tend to exacerbate uneven development. However, it did so from a Marxian perspective, and through a direct critique of Ricardian 'comparative advantage'. The central thesis was that when products are traded between high and low waged countries, there will always be a transfer of value from the latter to the former.

Out of both dependency and unequal exchange theories – but with critical perspectives on each – there emerged from the late 1970s what has come to be known as 'world systems' theory. Its main proponents, who include Samir Amin, Immanuel Wallerstein, and Giovanni Arrighi, have argued that dependency and unequal exchange theories assumed too much passivity on the part of less developed countries. Furthermore, the singular driver of unequal exchange theory (wage differentials) was increasingly seen to be a *result* of uneven development, not its underlying cause; and dependency theory, meanwhile, has been taken to task for oversimplification of the world into a bimodal system with only 'cores' and 'peripheries'.

Thus, world systems theory is perhaps best known – but also too easily caricatured – for its analysis of the world's uneven development in terms not only of an economic core and periphery, but also a 'semiperiphery'. This semiperiphery plays a key role in the world economy: it is a stabilizing entity between the 'core' and 'periphery', but it is also where the most intense restructuring and conflict occurs. World systems theory places great emphasis on specific analysis of the relationships *between* societies (and places), as distinct from more generalized models of social change. To this extent it diverges from classical Marxism, and it also rejects the developmentalism assumed not just by Marxism but by modernization theory (the assumption of unidirectional movement through 'stages' of development). However, Wallerstein, for one, sees himself as writing largely in a Marxist vein. And indeed, one of the key features of the three way economic stratification posited by this theory (core, periphery, semiperiphery) is that it is deemed to be fluid and dynamic: positioning in this hierarchy is not fixed, but rather is influenced by the effects of periodic crises in capital accumulation, and the resulting periods of stagnation and growth (described as 'waves'). The geographies of such crisis effects, however, have not been explicitly theorized.

Human geographers have, nonetheless, been influenced by world systems theory in exploring uneven development and have demonstrated that geographical positioning in the world system deeply shapes development possibilities and paths of differential economic progress. The work of Jim Blaut has been central in this regard – especially on the Europeanization of capitalism from the sixteenth century – and has been well received by world systems theorists. Other human geographers who have sought to draw on the insights of world systems theory (such as its rejection of developmentalism) include Peter Taylor, who has written widely on 'world hegemonies' and has worked with Wallerstein.

There is, too, another body of work that has strongly influenced geographers' thinking on uneven development, and which has explicit connections to the world systems literature. This is (in very broad terms) research into the international division of labor and global commodity chains. (World systems theorists rely on both of these concepts to give concrete, descriptive definition to the capitalist world economy; they argue that variation in the operation of the different institutions implicated in such chains and labor markets, at different geographical 'nodes', serves to reproduce patterns of uneven development. And yet for world systems theorists, global commodity chains and the international division of labor are essentially supporting concepts, not the primary objects of analysis.) This literature has been very influential within human geography in the last decade, guiding much of the work on 'global production networks' by economic geographers, particularly those of the 'Manchester school' most closely associated with Peter Dicken. One key figure in this original literature is Stephen Hymer, whose work on multinational corporations has been pivotal. Again, it would be wrong to simply label this work 'Marxist'. Nevertheless, Hymer himself did gravitate over time from a more mainstream economic perspective to an avowedly Marxist one, which increasingly saw him focus his work explicitly on the relationships between multinational corporations and uneven development.

Hymer's argument about uneven development was relatively simple (and indeed, has been criticized as overly simplistic). He claimed, in short, that multinational corporations create a world 'in their own image' (an echo of Marx and Engels), where the internal corporate spatial division of labor (paradigmatically, with headquarters in Western cities and production activities increasingly outsourced to 'developing' countries) is reproduced in the international division of labor more widely. Economic geographers' work on global production networks – and, specifically in relation to such networks, on uneven development – has been much less categorical. Arguing for a more accretive, bottom up form of 'theory', this work suggests that we can only illuminate the mechanisms of uneven development at all scales by examining in detail the places comprising such networks, the flows (of capital, knowledge, and labor) between them, and the ways in which these flows and places co construct one another. Most research in this area has been conducted on an industry sector specific basis, and therefore understands uneven development in relatively narrow terms. This work also, for the most part, has taken geographers quite a way from Marx. In both of these respects, it differs substantially from the strand of economic–geographical thinking on uneven development that I discuss next and in most detail. This work appeals directly to classical Marxism; indeed, it seeks to put the geography 'back' into Marx; and a theory of uneven development has been a central component of this 'reassertion' of space in Marxian political economy.

Uneven Development as the Geographical Motor of Capitalism

As Marxism became a driving force with human geography in general and economic geography in particular in the late 1970s and early 1980s, the concept of uneven development emerged as pivotal. The work of David Harvey and Neil Smith was, and remains, seminal (Dick Peet was another important early contributor). Both – and Smith most specifically – have sought to theorize uneven geographical development at a macro, abstract level and expressly in terms of the core conventions of Marxist economics. That said, and as Smith has indicated more than once, Marx himself did not say a great amount about uneven geographical development (although he did of course write about many other dimensions of unevenness in capitalist development). It is more in the writing of Vladimir Lenin and Rosa Luxemburg, and especially their writing on imperialism, that economic geographers have found direct inspiration for their theorization of this phenomenon.

The fullest such theorization has been offered, and subsequently refined, by Smith (though both he and, most recently, Harvey have insisted that the theory remains very much work in progress and will perforce continue to evolve as capitalism evolves). Smith argues that at the heart of capitalism there are contradictory geographical tendencies toward 'differentiation' and 'equalization'. Geographical differentiation is seen to occur at the level both of individual capitals and of specific sectors of the economy; meanwhile, capital is also, paradoxically, a 'leveler', tending to equalize across space the conditions of production and the level of development of productive forces. For Smith, these two opposed tendencies, and the contradiction between them, are what cause uneven development.

However, Smith emphasizes that this process is inherently dynamic; in other words, it does not result in a static disparity in levels of geographical development. Instead, places that are development 'laggards' at one moment may surge ahead at a later moment. This is because, in the development of one place, capital creates there many of the very conditions that militate against further development, manifested in a falling rate of profit – and vice versa for places of underdevelopment. Hence, in articulating a theory of uneven development, Smith asserts that capital is apt to oscillate (or 'seesaw') geographically from developed areas to undeveloped areas and then, later (once conditions have reversed), back again. But he admits that such a 'seesawing' is limited in practice by a variety of factors, including the level of mobility of capital. Hence, oscillation is highly evident at the urban scale, less so at the national scale, and hardly at all at the international scale (although the experience of the so called 'newly industrializing countries' arguably stands as an example of such international seesawing occurring).

These latter references to scale foreground a critical component of the concept of uneven development, which Smith has fleshed out in a variety of ways since developing his initial theoretical architecture. This is that uneven development not only occurs at a variety of different scales, but is itself intimately bound up with the production and reproduction *of* geographical scale. Thus, human geography's burgeoning 'scale literature' of the last two decades can in part be traced to the conceptualization of uneven development in Marxist economic geography. Uneven development, in this tradition, entails a fundamental restructuring of the geographical scales that frame – and are framed by – the production and circulation of surplus value. Most specifically, it is argued, recent decades have seen erosion of the materiality of the nation state as a scale of economic influence, with a complicated nexus of global and local scales increasingly coming to the fore. Others writing in the same political–theoretical tradition as Smith – such as Erik Swyngedouw – have sought to theorize and explore this rescaling, and its links to processes of uneven development, in more detail.

This high level Marxist theorization of uneven geographical development has been very influential both within human geography and outside it. However, it has not been without its critics. John Browett, who has also strongly criticized Frank's dependency theory, has suggested that Smith overstates the degree to which uneven development is inherent to capitalism and is systematically weaved into its fabric; contingent forces – and especially, he believes, 'start off' conditions in different territories – may constitute a sufficient explanation for the development of economic–geographical inequalities. Gordon Clark has been equally critical of the 'necessity' model of capitalist regional inequality. And from a very different theoretical and political perspective, meanwhile, Gillian Rose has offered a feminist critique of the Marxist theory of uneven development.

These critiques notwithstanding, a large number of economic geographers have been influenced by the work of Smith and Harvey in seeking to account for patterns of uneven development at all manner of geographical scales. Some of this work has sought to provide more precise analytical flesh to what were broadly theoretical claims – Michael Webber and David Rigby have both carried out research in this area, offering quantitative spatial analyses of, in particular, rates of profit and industrial development in order to buttress key elements of the uneven development thesis. Other work has been less mathematically oriented, but equally concerned to understand and explain uneven development in its historical–geographical particulars. On occasion this research has retained, and even deepened, the theoretical sophistication found in Harvey and Smith (most notably Patrick Bond's work on uneven development in Zimbabwe). More commonly, it has taken the theory and sought to examine the extent to which it is evidenced – or otherwise – in specific empirical situations. There has, for instance, been a large amount of geographical work on regional economic development in Europe that is influenced, to varying degrees, by the Marxian theorization of uneven development. Mick Dunford, Costis Hadjimichalis, Ray Hudson, Andres Rodriguez Pose, and Adrian Smith have all been active contributors in this area.

In addition to working with theories of uneven development in particular historical–geographical contexts, economic geographers have, lastly, tried to build what can best be called midlevel theoretical 'bridges' between generalized theory and those tangible historical geographies. The rationale for such efforts has been that Marx's (and in turn, Smith and Harvey's) thinking, while extremely suggestive, is at once too abstract and too rigid to provide an adequate framework within which to carry out detailed empirical research and to grasp the significance of specific empirical findings. (Smith, in turn, suggests that these other 'theories' frequently lose theoretical purchase and power altogether by lurching too far, and too exclusively, toward particularities.)

Perhaps most influential in this regard has been Doreen Massey's parallel work on uneven regional development, especially in the UK context (her first findings were published several years before Smith first wrote on uneven development). While also rooted in Marxist political economy, Massey's work uses a different terminology – her preference is 'spatial divisions of labor' – and she explicitly states that she does not see this as an explanatory concept. In discussing UK regional disparities (and in particular the congregation of economic control functions in the southeast, while direct production activities are found proportionately more elsewhere), she uses this phrase in reference to the spatial organization of relations of production, and to signal the importance of three interrelated sets of changes: changes in the national division of labour; the changing relation of the UK economy to international divisions of labor; and changes in modes of technical and industrial organization. Massey's key arguments are that the reasons for uneven development vary over time and space, and that different sectors of the economy exploit geographical unevenness in different ways, thus (re)producing different kinds of uneven development. Storper and Dick Walker have made not dissimilar claims (though with a greater emphasis on growth dy namics), their model of 'geographical industrialization' being equally indebted to Marx, equally grounded in production relations, and equally insistent on 'bridging' abstract theory and nuts and bolts empirics.

A final important direction taken by economic geog raphers in regard to the political economy of uneven development has been to seek to incorporate the insights of the regulation theory of, among others, Michel Aglietta. Adding nuance specifically to the type of the oretical structure erected by Harvey and Smith has, again, been a principal objective here. Jamie Peck and Adam Tickell followed this route in analyzing the in tensification of UK regional disparities under Thatcher ism. For them, regulation theory – with its twin emphasis on the historical specificity of capitalism's institutional forms, and on the role of processes of social regulation in averting crisis tendencies – helps us to better understand the temporal evolution of capitalism. They admit, how ever, that in its original form, regulation theory does not begin to broach the spatial complexities of uneven geo graphical development. By integrating the analysis of regional uneven development within a regulationist framework, Peck and Tickell thus sought to bridge high level Marxian theory and the regional particulars of UK historical geography. They have continued to pursue this work in different ways up to the present day, focusing in recent years on shifting inter urban patterns of UK un even development (a regional example of Smith's 'see saw'), specifically in the context of the political economy of neoliberalism.

Conclusion

One of the most interesting and revealing features of the concept of uneven geographical development has been its durability: in economic geography as in other disciplines and subdisciplines, tides inevitably change, bringing with them changing philosophies, methodologies, and sub stantive concerns. However, uneven development re mains a core political, theoretical, and empirical thematic for economic geographers (and indeed for many human geographers more generally) – notwithstanding 'post ' interventions of various stripes, the 'cultural turn' of the 1990s, and indeed the tentative dialogs between eco nomic geography and 'new geographical economics'. This is not to suggest that there is anything approaching a consensus on the nature of, or reasons for, uneven de velopment; a central objective of this article has been precisely to demonstrate the range of approaches to this phenomenon. Doubtless such variation will continue to feature in both research and debate. What the history of economic geography does suggest, however, is that no matter how it is delineated and explained, uneven de velopment will remain a central disciplinary topic for as long as it persists on the ground. The Marx inspired approaches to uneven development that continue to exert a major influence within the discipline would suggest that such persistence will be into perpetuity, so long, that is, as the world economy remains largely a capitalist world economy.

See also: Agglomeration; Capital and Space; Capitalism; City-Region; Core-Periphery Models; Dependency; Economic Geography; Economics and Human Geography; Global Commodity Chains; Global Production Networks; Globalization and Transnational Corporations; Industrial Districts; Labor Market; Local Economic Development; Local Economic Development, Politics of; Marxism/Marxist Geography I; Modernization Theory; Neoliberal Economic Strategies; North-South; Radical Political Economy; Regional Inequalities; Scale; Spatial Division of Labor; Third World; Trade, International; Uneven Regional Development.

Further Reading

Amin, S. (1977). *Imperialism and unequal development*. New York: Monthly Review Press.

Bond, P. (1998). *Uneven Zimbabwe: A study of finance, development and underdevelopment*. Trenton, NJ: Africa World Press.

Darity, W. A. Jr. and Davis, L. S. (2005). Growth, trade and uneven development. *Cambridge Journal of Economics* 29, 141 170.

Harvey, D. (2006). Notes towards a theory of uneven geographical development. In Harvey, D. (ed.) *Spaces of global capitalism: Towards a theory of uneven geographical development*, pp 69 116. London: Verso.

Hymer, S. H. (1972). The multinational corporation and the law of uneven development. In Bhagwati, J. N. (ed.) *Economics and world order*, pp 113 140. London: Macmillan.

Krugman, P. (1991). *Geography and trade*. Cambridge, MA: MIT Press.

Massey, D. (1984). *Spatial divisions of labour*. London: Macmillan.

Peck, J. and Tickell, A. (1995). The social regulation of uneven development: 'Regulatory deficit', England's South East, and the collapse of Thatcherism. *Environment and Planning A* 27, 15 40.

Scott, A. J. and Storper, M. (2003). Regions, globalization, development. *Regional Studies* 37, 579 593.

Smith, N. (1984). *Uneven development: Nature, capital and the production of space*. Oxford: Blackwell.

Smith, N. (2000). The restructuring of spatial scale and the new global geography of uneven development. *Jimbun Chiri (Japanese Journal of Human Geography)* 52, 51 66.

Storper, M. and Walker, R. (2003). *The capitalist imperative: Territory, technology and industrial growth*. Oxford: Blackwell.

Wallerstein, I. (1979). *The capitalist world economy*. Cambridge: Cambridge University Press.

Uneven Regional Development

R. Hudson, Durham University, UK

Glossary

Centrifugal Forces They are factors that encourage spatial de-concentration. They include labor immobility, lower land and property prices, and rents and external diseconomies, such as congestion costs.

Centripetal Forces They are factors that encourage agglomeration and spatial concentration. They include market size, cooperative and functional linkages among firms, dense and heterogeneous labor markets, and external scale economies, such as knowledge spillovers.

Cumulative Causation It is the process whereby inequalities between regions become self-reinforcing, so that growth is increasingly concentrated in some regions while in others it declines.

Division of Labor It refers to the ways in which work is divided between people. It can refer to the organization of work in individual workplaces (technical division of labor), between companies (social division of labor), or between places (spatial division of labor). The latter is of particular relevance in relation to regional uneven development.

Political Economy This refers to a variety of approaches to understanding economies and their geographies, including evolutionary, institutional, and Marxian variants. While these differ in important ways, they share a rejection of the assumptions and static equilibrium approach of mainstream neoclassical economics.

Introduction

The concept of uneven development originally developed within Marxian political economy to describe – *inter alia* – relations between political classes, between the forces and relations of production, and between companies. Subsequently, it became generalized across a wide range of approaches in the social sciences and substantive domains, not least that in regional uneven development. Uneven development is endemic to capitalist development at all spatial scales, including the regional (understood here to mean spatial units below the scale of the national). The empirical evidence is so widespread that, with the exception of the proponents of neoliberalism, there is a general acceptance of the fact of regional uneven development and that this is an endemic feature of capitalist development and is not something that can be eliminated. Whether it can be managed and its effects ameliorated so that it remains within politically and socially acceptable limits is a much more contentious issue.

A corollary of this widespread agreement on the inevitability of regional uneven development, however, is that a variety of approaches to understanding and explaining regional uneven development have been constructed, some of which are compatible and complementary, others which are competitive and incompatible. Life is further complicated by those approaches that also seek to problematize the very notion of 'development' itself, though for reasons of space they are largely ignored here. Before reviewing some of the major perspectives on uneven development, however, the article begins with a brief consideration of neoliberal approaches that assume that uneven development would cease to be an issue only if people behaved as rational economic actors and allowed market forces to establish equilibrium conditions and thereby abolish uneven development.

The Denial of the Necessity and Inevitability of Uneven Development

The essence of neoliberal approaches is an assertion that freeing up markets, so that capital and labor can flow to where they are most efficiently deployed, will lead to an equalization of factor returns and convergence between regions and/or to a balanced pattern of regional development. Labor and capital, it is claimed, will move from areas of surplus to areas of shortage in response to the differential in wage rates and profit rates, respectively. Thus, labor should move from poorer, less developed to richer, more developed regions; conversely, capital should move in the opposite direction until wage and profit rates are equalized across regions, resulting in an efficient and balanced pattern of development.

There are a number of difficulties with this perspective, both empirical and theoretical. Empirically, there is very little if any evidence of actual patterns of regional development following the path to strong convergence; and in cases where there is some evidence of weak convergence, as in the European Union (EU) in the 1970s and 1980s, this can be attributed more to the effects of state regional policies than to market forces. Theoretically, for markets to operate in the unfettered way assumed by neoliberal perspectives, it is necessary to make a series of sweepingly implausible assumptions about the knowledge and motives of actors (much in the same way that the neoclassically inspired location theories that were seen as cutting edge in economic geography in the 1960s required a similar suite of heroic assumptions).

Cumulative Causation and Uneven Development

In contrast to those who argued that regional uneven development could be eliminated if only market forces were allowed to operate in an unfettered manner, others saw the play of market forces as setting in motion processes of cumulative causation that would reinforce uneven development, as economically 'successful' and 'unsuccessful' regions became locked into their respective trajectories of growth and decline. As concepts of cumulative causation and polarization are discussed fully in elsewhere in the encyclopedia, they will only be discussed briefly here.

Myrdal directly challenges the view of self equilibrating markets and that movements of capital and labor are in opposite and self regulating directions. In contrast, "the play of the forces in the market tends to increase, rather than decrease, the inequalities between regions" (Myrdal, 1957: 26). The pattern of spatial interaction between growing and stagnating regions is therefore an important aspect of Myrdal's explanation of differential regional growth and decline. A variety of 'backwash effects' emerge that militate against growth in peripheral regions while sustaining it in the growing economic core regions. Hirschman independently developed an interpretation that was remarkably similar to Myrdal's except that in Hirschman's view, if an imbalance between regions emerges as a result of 'polarization effects' during the early stages of growth, counterbalancing forces will in time come into being to restore an equilibrium position If counterbalancing forces will 'in time' come into being; however, the issue of how much time will pass before this is the case becomes quite a critical one. To paraphrase Keynes, in the long run we are all dead and the prospect of regional convergence at some point in an unknown future is not necessarily an attractive one.

Some four decades after the pioneering work of Hirschman and Myrdal, Paul Krugman also set out to explore the rationale for cumulative causation and the dynamics and processes of uneven development. Krugman's 'new economic geography' (perhaps more accurately described as 'new geographical economics') involves quite a formal mathematical analysis of endogenous growth within an otherwise neoclassical framework that remains committed to methodological individualism and a very thinly socialized explanatory account of the processes of cumulative causation. Krugman's approach focuses upon the balance between centripetal and centrifugal forces in determining the extent and form of regional concentration of economic activities. Centripetal forces, which tend toward spatial concentration, include market size, cooperative and functional linkages between firms, dense labor markets with a diversity of available skills, and external economies of scale, such as knowledge spillovers. Centrifugal forces are those that tend toward spatial de concentration, notably labor immobility, lower land and property prices and rents, and external diseconomies of various sorts such as congestion. Within the range of centripetal and centrifugal forces, Krugman emphasizes economies of scale and transport costs – variables more amenable to quantitative measurement and as such consistent with his methodological preference for building mathematical models. Within these models, then, the tendency for spatial clustering is positively correlated with economies of scale and negatively correlated with transport costs. Growing regional divergence and a core periphery pattern of economic development is a result of specialization leading to increased efficiency, comparative advantage, and cumulative growth in 'core' regions because firms there benefit from cost savings and/or revenue increases as a result of mutual interaction and intraregional cooperation. Krugman acknowledges that regional specialization can evolve accidentally but, having done so, economies of scale and external economies will lead to cumulative growth, leading to lock in and path dependency, thereby reinforcing regional unevenness and uneven development as a consequence of processes internal to the workings of capitalist economies.

In contrast to the main emphases of the new economic geography/new geographical economies inspired by Krugman's work, other economic geographers, and others (such as the business economist Porter, with his emphasis upon clusters and regional development) have in parallel sought to develop explanations of regional uneven development much more in cultural and institutional terms. As Allen Scott has pointed out, "a strictly economic logic will only take us so far in understanding industrial organizational processes …[T]ransactional systems are always and of necessity embedded in historically determinate social conditions." In seeking to explicate this process of social embedding and the sociocultural constitution of economies, cultural and institutional analysts have drawn upon notions of knowledge, innovation, and learning; of 'untraded interdependencies' and the non economic relationships that underpin regional economic success; or of institutional capability and 'thickness' as the reasons for increasingly divergent regional economic performance. Thus, successful regions are seen as those that possess apposite institutional structures, rich patterns of supportive social relationships beyond the workplace and the boundaries of the company that facilitate successful economic practices and flows and systems of exchange of both codified and tacit knowledge. These latter help ensure that knowledge of 'best practice' diffuses within the regional economy and that knowledge flows and interaction underpin regional innovation sys tems. While clearly such issues have always been crucial in underpinning regional economic success, they are claimed to be of added salience in the context of con temporary 'knowledge based economies' (although how 'any' economy could exist without being knowledge based is difficult to conceive). Conversely, poorly per forming regions are seen to be locked into institutional structures and forms of knowledge that may once have been relevant to regional economic success but are now obsolete, rendered redundant by wider processes of economic change and changes in spatial divisions of labor. This remains the case despite policy induced at tempts to change the institutional and knowledge bases of these poorly performing regions. However, there is a distinct spatial selectivity in the regions chosen as case studies because proponents of such approaches typically focus upon case studies of particular regions, generally – though not always – the successful ones. As a result, while providing often rich and detailed insights into the reasons for the economic success or failure of individual regions, such studies lack the capacity to provide insights into the wider overall patterns of regional uneven development and the reasons for them and the interrelationships be tween economic 'success' in some regions and 'failure' in others.

Marxian Political Economy and Uneven Development

The approaches discussed so far, in various ways, all seek to account for uneven regional development and set out the reasons why growth or decline becomes a cumulative and self reinforcing process. They all see regions be coming locked into their respective trajectories of growth or decline once these have been established, an essen tially – thought implicitly – evolutionary perspective on regional development, in which the past strongly con ditions, though does not rigidly determine, the future. There are certainly many regions that can be described and interpreted in these terms. However, the historical geography of capitalist development is also replete with examples of regions – the northeast of England, the Ruhrgebiet, Wallonia, Nord Pas de Calais, the industrial mid west of the USA (those often described as 'old in dustrial regions') – that were once centers of capital ac cumulation and cumulative economic growth but which 'flipped' and became regions of decline, characterized by capital flight, devalorization, and disinvestment. Sub sequently, to varying degrees, they have experienced a limited degree of renewal, based upon flows of new capital into these regions. Such dramatic reversals from trajectories of growth to decline to renewed growth of a different type and scale require a different sort of con ceptual and explanatory approach.

This is provided by Marxian political economy, with its emphasis upon the inner dynamics of the capitalist mode of production (i.e., the particular combination of social relations and technologies that define capitalist economies as capitalist), with competition among and within the structurally defined classes of capital and labor as the driving force of the capitalist economy. Individual companies seek to compete in a range of ways – via innovative products and production processes, for ex ample – and also, of particular relevance in the context of regional uneven development, by seeking out locations that are particularly profitable and so favorable for pro duction. These can include both locations in which existing products can be produced more profitably with existing production technologies and locations in which new innovative processes and or products can be intro duced. Thus as an integral part of their competitive strategies, companies are regularly scouring the globe for more locations that will enable them to produce more profitably and so gain a competitive edge over their rivals. Consequently, they both routinely transfer value between regions via exchange relations within and between companies and routinely devalorize capital and disinvest from some regions while investing in others, either re gions previously not penetrated by capitalist relations of production or regions that have been abandoned by other companies (or national states) as locations for profitable production. Thus, regions may sequentially experience successive waves of investment and disinvestment ex pressed as sequences of regional industrialization, de industrialization, and re industrialization, as part of the processes of combined and uneven development that are structurally inscribed as integral to capitalist develop ment, with spatially uneven development seen as a product not just of the uneven distribution of natural resources and the influences of nature on economic geographies but as arising of the social relations of capital.

While Marx's own work contains suggestive comments and hints about regional uneven development and its significance for capitalist development, these were not fully or systematically developed. Subsequently, others such as Gramsci and Lenin, working in various strands of the Marxian tradition, further developed Marx's insights and the analysis of uneven development at various spatial scales. It was not, however, until the 1960s and the work of Ernest Mandel that regional uneven development began to be more integrally related into Marxian political economy, extending the reach of that approach in the process. Mandel specifically recognized the centrality of intranational uneven development and the production of 'regional problems' to the accumulation process. For example, he argued that "unevenness of development as between different parts of a single country" is an essential pre condition for capital accumulation and that its significance had been greatly underestimated in previous Marxian analyses. It is also worth noting that other social scientists were also paying increasing attention to issues of spatially uneven development as a part of this reinvigoration of Marxian scholarship.

However, the development of Marxian political economy to take account of regionally uneven development as a structurally necessary feature of capitalist development owes most to David Harvey, especially in his magisterial account of *The Limits to Capital* (1982), in many ways best regarded as a reconstruction of a historical geographical materialism. Harvey located the issue of regional uneven development within the context of his 'third cut' at crisis theory and the way in which capital both produces and uses spatial differentiation as part of its repertoire of tactics to offset falling profitability. Capitalist development was therefore explicitly conceptualized as necessarily and unavoidably uneven, simultaneously encompassing regions of growth alongside those of decline and integral to the crisis prone process of capital accumulation.

Harvey's approach was further elaborated by Neil Smith with his concept of 'a see–saw theory of uneven development' as a way of seeking to grasp the dialectical relations between development and underdevelopment and the inevitability of capital eroding the place specific conditions of profitable production as part of the developmental process. In response to differences in profitability and those things that determine it – such as labor market conditions or pollution regulation – capital flows into and out of regions, in the process generating growth or decline as well as helping (un)make regions as socio material ensembles. As Smith (1984: 149) puts it: "Underdevelopment, like development proceeds at every spatial scale, and capital attempts to move geographically in such a way that it continually exploits the opportunities of development without suffering [the] economic costs of underdevelopment." That is, capital attempts to see–saw from a developed to an underdeveloped area, then at a later point back to the first area which is by now underdeveloped, and so forth. Capital seeks not an equilibrium built into the landscape but one that is viable precisely in its ability to jump landscapes in a systematic way. This is the see–saw movement of capital, which lies behind the larger uneven development process. Thus the 'see–saw theory' represents capital's attempt to secure a 'spatial fix' via systematic mobility and a sort of dynamic equilibrium rather than fixity and a static equilibrium and helps grasp the rationale for the constant ebb and flow of capital into and out of regions that lies at the heart of processes of regional uneven development.

However, this still leaves the question of which regions experience which sorts of trajectories of growth and decline rather open and progress on this front was primarily a result of Doreen Massey's seminal work, brought together in *Spatial Divisions of Labour* (1984). Rather than see the fate of regions as simply the end product of the decisions of capital, as layers of investment and disinvestment are sedimented sequentially in a region at capital's behest, with regions as little more than passive objects resulting from the logic of capital, Massey emphasized the need to take account of both the natural and socially produced attributes of regions, the activities of people in their place seeking to make and defend the economic and social viability of their regions, for example, as critical in understanding which regions experienced which sorts of growth, decline, and revival. In short, the economic success or failure of regions was a result of the interplay between regionally specific attributes and processes and wider systemic forces shaping flows of capital. It is worth pointing out that subsequently she was to develop more sophisticated conceptions of 'regions' and of the way in which they became intertwined via processes of combined and uneven development. In this sense, there are points of convergence between the sort of explanatory approach that Massey developed and more recent cultural and institutional perspectives on why some regions 'succeed' while others 'fail', provided they can be connected with more systemic explanations of uneven development.

The Role of the State in Managing Regionally Uneven Development: Managing Tensions and Avoiding Crises

Regions may develop a 'structured coherence', generating a sense of regional identity and interest shared by a range of social groups and forces, expressed via a particular 'structure of feeling'. Such a structure of feeling and attachment to a region can, when linked to uneven regional economic development, often become the trigger for a variety of regionalist campaigns as different alliances of social groups come together to defend or promote a shared regional interest. This might involve a campaign to protect existing economic activities or to attract new ones to peripheral regions. It might involve pressures to reduce income transfers from economically successful to peripheral regions, to increase central state resource allocations from economically successful to peripheral regions, or to establish more devolved forms of regional governance that give more powers and resources to regions (or, more precisely those empowered to speak and act on behalf of the region) as active subjects that can act 'for themselves' and have greater degree of influence over their own economic well being, although such moves may be contested within the region itself. In one form or another, however, the political effects of regional uneven development are to generate pressures to alter patterns of resource allocation via the state and keep state expenditures within acceptable limits and/or to restructure the state itself in an effort to smooth the path of regional economic growth, secure the legitimacy of state action, or to secure the territorial integrity of the national state and avoid various potential forms of crisis.

Beginning in the UK in the late 1920s, then, regional uneven development ceased to be seen simply as an inevitable and unremarked feature of industrial capitalism and became seen as a potential political problem, a potential source of social unrest and dissent and a brake on national economic growth, to which the state 'had' to respond. Subsequently, such regional policies became developed in more sophisticated ways and became part of the policy repertoire of virtually all national states as they sought to manage the map of regional uneven development. Consequently, such policies must seek to defuse the unavoidable tensions and latent conflicts that arise as a result of a region being simultaneously a socially produced place with multiply dimensional meanings and attachments for a variety of people and a part of a socially produced space in which capitals seek to make profits. For much of the time, this conflict remains latent, submerged, at worst simmering, as the tensions remain within tolerable limits. But the tendency toward boiling over and erupting into regional crisis never disappears; it always is (and must be) immanent in the social relations of capital. In most regions in a capitalist economy, their relationships to the ebb and flow of the accumulation process in processes of uneven development, is critical to their (re)production as 'places'.

There are, however, limits to the capacity of the state to act so as to reduce the extent of regional uneven development. These arise because of the links between the state and the dominant social forces and relationships within capitalist societies. Consequently the state is not simply a neutral instrument amenable to capture by particular sociopolitical forces but is itself both an arena for and product of political conflict. As a result, there are definite selection mechanisms that serve to focus its policy choices and the socioeconomic interests that these serve and limits to the sorts of policy choices that it can make precisely because it is a capitalist state.

Conclusions

Few would deny the reality of regional uneven development as a chronic feature of capitalist development, and as such it is something that will certainly be with us for the foreseeable future; the weight of empirical evidence, along with the insights provided by a variety of theoretical positions, renders any other conclusion both unreasonable and implausible. That said, there is clearly a variety of theoretical perspectives that seek to explain why this is so rather than consensual agreement on matters of theory. While some of these perspectives at least suggest that regional uneven development can be avoided, the point more generally at issue is which of the variety of competing perspectives as to the persistence of regional uneven development is the most appropriate explanation. However, because of the ways in which regional inequality impinges upon daily life for so many people, regional uneven development long ago ceased to be seen simply as a feature of capitalist development of interest to human geographers and other social scientists and became a political issue of some significance, one that national states have felt obliged to address for the last 80 or so years. Thus in seeking to manage the tensions arising from pressures that arise precisely because of the regionally uneven character of economic development, national states have devised and implemented a repertoire of regional polices 'for' regions as objects of central state policies and/or devolved power and responsibility 'to' regions to become active subjects and deal with their own developmental issues. Through these varied policy responses, national states seek to ensure that their national territories remain (or become) key nodes in global capital flows, containing the tensions between the imperatives of accumulation and the sociopolitical pressures that might build up to challenge the legitimacy of the state. It is a sobering comment upon the limits to the capacity of national states to shape the trajectories of uneven development that regional inequalities and regional uneven development remain as chronic and persistent features of capitalist economies, both the most advanced and the most backward, and will continue to be so in the future.

See also: Cumulative Causation.

Further Reading

Allen, J., Cochrane, A. and Massey, D. (1998). *Re Thinking the Region*. London: Routledge.

Amin, A. and Thrift, N. (eds.) (1994). *Globalization, Institutions and Regional Development in Europe*. Oxford: Oxford University Press.

Castree, N., Essletzbichler, J., and Brenner, N. (Convenors) (2004). Symposium: David Harvey's *The Limits to Capital*: Two decades On. *Antipode*, 36, 401 549.

Dunford, M. (1994). Winners and losers: The new map of economic inequality in the European Union. *European Urban and Regional Studies* 1, 95 114.

Dymski, G. (1996). On Krugman's model of economic geography. *Geoforum* 27, 439 452.

Grabher, G. (1993). The weakness of strong ties: The lock in of regional development in the Ruhr area. In Grabher, G. (ed.) *The Embedded Firm: On the Socio Economics of Industrial Networks*, pp 255 277. London: Routledge.

Habermas, J. (1976). *Legitimation Crisis*. London: Heinemann.

Hadjimichalis, C. (1987). *Uneven Development and Regionalism: State, Territory and Class in Southern Europe*. London: Croom Helm.
Harvey, D. (1982). *The Limits to Capital*. Oxford: Blackwell.
Harvey, D. (1985). The geopolitics of capitalism. In Gregory, D. & Urry, J. (eds.) *Social Relations and Spatial Structure*, pp 128 163. Basingstoke: Macmillan.
Hirschman, A. (1958). *The Strategy of Economic Development*. New Haven, CT: Yale University Press.
Hudson, R. (1989a). Labour market changes and new forms of work industrial regions: Maybe flexibility for some but not flexible accumulation. *Society and Space* 7, 5 30.
Hudson, R. (1989b). *Wrecking a Region: State Policies, Party Politics and Regional Change*. London: Pion.
Hudson, R. (1994). Institutional change, cultural transformation and economic regeneration: Myths and realities from Europe's old industrial regions. In Amin, A. & Thrift, N. (eds.) *Globalization, Institutions and Regional Development in Europe*, pp 331 345. Oxford: Oxford University Press.
Hudson, R. (1999). The learning economy, the learning firm and the learning region: A sympathetic critique of the limits to learning. *European Urban and Regional Studies* 6, 59 72.
Hudson, R. (2001). *Producing Places*. New York: Guilford.
Hudson, R. (2004). Conceptualising economies and their geographies: Spaces, flows and circuits. *Progress in Human Geography* 28, 447 471.
Hudson, R. (2006a). From knowledge based economy to … knowledge based economy? Reflections of changes in the economy and development policies in the north east of England. Paper presented at the annual conference of the Association of American Geographers, Chicago, March 7 10.
Hudson, R. (2006b). Regional devolution and regional economic success: Myths and illusions about power. *Geografiska Annaler B* 88, 159 171.
Hudson, R. and Sadler, D. (1983). Region, class and the politics of steel closures in the European Community. *Society and Space* 1, 405 428.
Krugman, P. (1991). *Geography and Trade*. Cambridge, MA: MIT Press.
Krugman, P. (1998). What's new about economic geography? *Oxford Review of Economic Policy* 14, 7 17.
Mandel, E. (1963). The dialectic of class and region in Belgium. *New Left Review* 20, 5 31.
Mandel, E. (1968). *Marxist Economic Theory, Vols. 1 and 2*. London: Merlin.
Mandel, E. (1975). *Late Capitalism*. London: New Left Books.
Martin, R. and Sunley, P. (2006). Path dependence and regional economic evolution. *Journal of Economic Geography* 6, 395 637.
Massey, D. (1984/1995). *Spatial Divisions of Labour* (1st and 2nd edns). London: Macmillan.
Morgan, K. (1997). The learning region: Institutions, innovation and regional renewal. *Regional Studies* 31, 491 503.
Myrdal, G. (1957). *Economic Theory and Underdeveloped Regions*. London: Duckworth.
O'Connor, J. (1973). *The Fiscal Crisis of the State*. New York: St. Martins Press.
Porter, M. (2000). Location, clusters and company strategy. In Clark, G., Gertler, M. & Feldman, M. (eds.) *Oxford Handbook of Economic Geography*, pp 253 274. Oxford: Oxford University Press.
Porter, M. (2003). The economic performance of regions. *Regional Studies* 37, 549 578.
Poulantzas, N. (1978). *State, Power, Socialism*. London: New Left Books.
Smith, N. (1984). *Uneven Development*. Oxford: Blackwell.
Storper, M. (1995). The resurgence of regional economies, ten years later: The region as a nexus of untraded interdependencies. *European Urban and Regional Studies* 2, 191 222.
Williams, R. (1989). *The Politics of Modernism*. London: Verso.

Urban Architecture

P. Kraftl, University of Leicester, Leicester, UK

Glossary

Critical Geography of Architecture A recent approach to studying architecture that moves beyond symbolism and iconography, to explore how 'ordinary' inhabitants and users of buildings make meaning through their actions, speech, and emotions. It draws on geographies of consumption and nonrepresentational theory (NRT).

Mega-Structures Very large buildings that may incorporate skyscrapers, shopping malls, enormous office complexes, distribution warehouses, and self-contained 'cities' within single buildings. These are often built in the 'International Style', either in city centers or their outskirts. They are usually highly controlled in terms of environment and security, and sometimes, built to include ambiguous 'public' spaces.

Modern Architecture A diverse period or style of architectural production characterized by efficiency, a 'machine' esthetic, straight lines, and notions of 'health', which ran from around 1900 to 1970. However, the precise nature, historical periodization, and even the very existence of Modern architecture are subject to intense debate.

Typology It is an attempt to conceptualize objects or processes (in this article, urban architecture), by grouping together similar entities into 'ideal types'. This process allows researchers and practitioners to find patterns and to work from typologies (e.g., of building style) when planning. It can be criticized for generalizing and modeling the world relatively simplistically, but it is a useful tool when used as a creative spark rather than a static model.

Introduction

The study of urban architecture transcends many boundaries. Buildings in urban areas have been the focus of study for urban, cultural (of various persuasions), social, and economic geographers. Buildings in urban areas have been studied as individual entities; or, they have been divided into their micro geographical, constituent parts; or, they have been incorporated into larger scale analyses of 'the built environment', 'the cityscape', or 'the neighborhood'. Buildings in urban areas have, therefore, been the focus for research not only by geographers, but by urban sociologists, architects, planners, anthropologists, landscape architects, and urban designers.

As a result of such complexity, this article shall concentrate principally on studies of individual buildings that have significantly informed geographers' approaches in researching urban landscapes around the world. This means that, in places, the work of researchers who may not define themselves as 'geographers' has been included. Moreover, certain seminal works that do not deal solely with individual buildings – but which have affected geographers' study thereof – also appear in this article.

This article begins by tracing the historical development of urban architectural research in geography – outlining the main theoretical and methodological approaches that have been adopted. It is important to remember that this development has not been simple, linear, or sequential: some research methodologies developed during the early part of the twentieth century continue to affect the study of urban architecture today. The rest of the article outlines how those approaches have enabled geographers to broach five broad themes in the study of urban architecture. For each theme, examples of geographically orientated work from the Majority and Minority Worlds demonstrate the variety of work in the field of urban architecture, which continues to develop today.

The Historical Development of Urban Architectural Research in Geography

The roots of architectural research in geography began in the 1920s in the USA. The early pioneers, in fact, constituted the first identifiable cultural geographers (who formed part of the so called 'Berkeley School'). At the large scale of the 'culture region', they mapped the material manifestation of cultural and religious traits in the US landscape. Concentrating almost exclusively on rural areas in the USA, their geography of artifacts enabled them to explore how cultural diversity gave rise to different house styles, land uses, and transportation systems. Although concerned with the materialization of culture in general, this early work represented the first sustained attempts to engage with architectural forms (such as agricultural barns), and to map their cultural geographical variety and significance. The simplicity of the Berkeley School's approach is attractive: in fact, it still informs the study of built environments by some US geographers today. Nevertheless, during the 1980s, the theoretical and methodological naiveté of the Berkeley School was vehemently critiqued by geographers who advocated more nuanced, critical programs for studying the symbolic and iconographic meaning of landscapes (see below). In particular, they were admonished for treating culture as an entity, rather than a process; they were also criticized for their assumption that culture emerged organically, whereas in many cases, cultural processes and forms could be actively manipulated to serve certain interests. Moreover, despite their impact on cultural geography as a whole, the Berkeley School did little to further the study of architectural geography in urban areas.

Meanwhile, from the 1930s onward, a disparate group of theorists and social commentators produced a number of seminal studies that were to inform not only the academic study of buildings, but also their design and configuration in the urban landscape. Loosely grouped under the banner of the 'built environment' tradition, their work considered the condition of the urban environment (largely in the Minority World). Often, their work exceeded the scale of individual buildings to consider the interrelationships between architectural forms, the design of street spaces, and the vitality of whole cities. Nevertheless, four major elements of built environment research have provided particularly important precedents for geographers' study of urban architecture. All four elements provided the foundations for more critical, politicized approaches to the study of urban architecture.

First, built environment scholars such as Lewis Mumford attempted to characterize the problematic conditions of urban areas during the first half of the twentieth century. They criticized the poverty, pollution, and poor planning of industrialized cities; and in particular, bemoaned the condition of the built environment. In effect, these scholars did not (merely) view architectural form as the manifestation of culture, but began to contextualize urban architectural design within the unequal social, political, and environmental conditions of industrial towns and cities. Second, theorists of the built environment promoted pragmatic theories and methodologies for the improvement of urban architectural design – and hence the improvement of urban life in industrialized cities. Influenced by Modern design dictates, they argued that better buildings and better streetscapes were a crucial part of planning for better cities. Third, in order to achieve their pragmatic (and sometimes, idealistic) aims, built environment scholars, such as Christopher Alexander and Gordon Cullen, produced 'typologies' of architectural design. For instance, typologists identified ideal types of pathways, shapes, spaces, and boundaries that would promulgate healthier, safer, more vibrant, and more sociable cities. Hence, their pragmatism was imbued with a psychological, quasi scientific desire to categorize both extant and possible planned architectural forms. Fourth, from the 1960s onward, built environment scholars extended their research methodologies beyond challenging the conditions of urban environments and beyond the formal designation of typologies. On the one hand, stinging critiques (of which Jane Jacobs' is perhaps the most polemical) of the loss of socialization, the loss of vitality, and the loss of public space in cities led to a reconsideration of the founding precepts of Modern architecture and planning. Herein began geographers' engagement with the issue of public space; moreover, herein were sown the seeds of distrust in universal, meta theoretical modes of conceptualizing and designing the city. On the other hand, the work of behavioralist academics sought, for the first time, to assert the importance of urban inhabitants' experiences and perceptions of city spaces. Behavioralists, too, sought to improve the vibrancy of urban areas through design and the production of typologies: yet their use of in depth interviews and graphical techniques ('mental maps') with city residents represented early attempts to engage with the users of buildings and streets – not just the producers. In sum, the built environment tradition – although design orientated – provided the early stimuli for more contextualized, politicized, critical, and engaged approaches to the study of urban architecture.

The 1970s saw the radicalization of human geography and, in particular, an increasing turn to (Marxist) political economic studies of urban geography. The work of the built environment school was one of a number of social and academic influences on this turn. Yet, far from signaling a departure from 'cultural' geographies of architecture, the work of Harvey, Castells, and others began to demonstrate how the production of built form was contingent upon the varied political economic power relations that structure urban spaces. In particular, it became clear that urban architecture symbolized, to some extent, the power relations that inhered in its making. For instance, it was argued that throughout history, the most iconic and impressive buildings (such as skyscrapers; **Figure 1**) symbolized the most acceptable version of urban 'meaning' that resulted from the negotiation process between different actors and social groups. Yet, in practice, it was argued that these edifices merely symbolized the ideological and economic interests of the dominant or ruling class. Hence, buildings – and differences in architectural form and function – could be read together as a 'landscape of power'.

An interest in the symbolism and materialization of power through built form has pervaded a large proportion of research on urban architecture since the 1980s. During the 1980s, there were impassioned debates between (cultural) humanist geographers and Marxist (social) geographers – the former preferring the explanatory logic of individual human agency, the latter of broader, structural social relations. 'Radical' or 'new' cultural geographers sought to reconcile these differences by considering: the interplay of structure and agency; the contemporary and the historical; the urban and the

Figure 1 Architectural ornamentation and symbolism in skyscrapers: Bryant Park, New York City. Author's photograph.

rural; and, the relationship between dominant and subordinate classes. Two strongly related approaches to urban architectural forms emerged. First, it was argued that, following the Marxists, landscapes could be read as 'texts'. In contradistinction to the Berkeley School, landscapes were conceptualized as configurations of signs and symbols that required interpretative (rather than morphological) methodologies. It was argued that social meaning was inscribed ('written') into built form, and that those forms could be read for their meaning, in order to uncover the explicit, implicit, and hidden negotiations of power that constituted their making. Second, it was argued that, rather than providing a text, a landscape could be considered as an image, or a 'way of seeing' the world. This 'iconographic' approach retained the notion that cultural artifacts, such as buildings, could symbolize meaning: yet there was acceptance that images were not transparent windows onto a world of reality. Instead, images were contextualized and always acted in interplay with one another, such that academic enquiry could not always 'get to the bottom of things' in terms of what an image symbolized. In the case of either approach, there followed a huge variety of work that explored how and why buildings in urban areas (especially religious institutions, commemorative national monuments, political buildings, and commercial buildings) symbolized and served specific ends.

Many variations of Marxist, symbolic, iconographic, and visual approaches to architectural geographies emerged during the 1990s. Despite such diversity, very recent work by architectural geographers has roundly criticized these textual and visual methodologies. Some critics have argued that in their search for textual or visual meaning, the 'critical', the political edge of geographical research (in other words, that which connects cultural and political analysis) has disappeared. More broadly and perhaps more importantly, critics have also suggested that reading landscapes in terms of their signification and production ignores the many and varied meanings that users and inhabitants – 'ordinary people' – make on an everyday basis. Newer, 'critical' arguments draw, instead, on both contemporary consumption research and the burgeoning interest in nonrepresentational theory (NRT). Hence, these 'critical geographers' advocate approaches which engage with the active, ongoing, diverse, and everyday inhabitation and consumption of architectural spaces (**Figure 2**). By focusing on the perceptions, actions, and emotions of 'ordinary' users of buildings (and the 'affective' import of those spaces for their inhabitants), one can explore the multifarious ways in which architectural spaces are both produced and experienced, beyond the intentions of architects. Although the methodological implications of such critical geographies have yet to be fully considered, geographers have used ethnographic methods, such as in depth interviewing and participant observation, to move beyond purely symbolic accounts of buildings.

As with all geographical research, geographers do not tend to work solely from one methodological approach. Although the insights of NRT have questioned the neat commentaries of textual and visual methodologies, those earlier approaches have not been forgotten. Some of the best contemporary geographical work on architecture

Figure 2 Grand Central Station, New York City. Recent geographical research has focused upon people's diverse use of urban buildings such as the different movements and flows of passengers in this image. Author's photograph.

explores how meaning, action, and emotion are co produced by and through architectural spaces in their attendant complexity – for instance, in exploring how buildings' symbolism interacts with their use, and with their socioeconomic context. Accordingly, geographers' work depends also on the individual buildings under consideration: the specific methods used to research a skyscraper in New York would be very different from those used for a shopping mall in Australia, or a public library in Vancouver.

Rather than generalizing the rich diversity of geographers' recent research on urban architecture, it is more instructive to consider the various empirical themes that such work has addressed. The rest of this article provides a critical overview of five of the most important themes.

Theme 1: Machines

One of the most important developments in urban architectural design was the advent of architectural Modernism during the twentieth century. Interestingly, despite a few seminal texts, there has been relatively little specifically geographical research on architectural Modernism in its diverse incarnations. It should also be noted that such research focuses almost exclusively on the Minority World. Nevertheless, extant research has highlighted the, sometimes, stringent practical and symbolic doctrines of Modernism and its repercussions for contemporary urban space – including the identification of buildings as healthy, efficient 'machines' for living in, and thus representative of a technological future.

On the borderlines between architectural history and historical geography, the diverse personalities (and attendant ideologies and practices) that constituted the patchy but global spread of architectural Modernism have been documented. These have largely focused on some of the best known architects (such as Le Corbusier and the Bauhaus) and their work, tracing the development of Modern design philosophy and the techniques (such as the refinement of reinforced concrete) that enabled those ideas to be realized. With specific stories to tell, such histories often lack a critical edge – rarely referring either to users' experiences of their buildings, or to the socioeconomic inequalities that urban architectures sometimes compounded. Nevertheless, these histories do highlight the importance of such personalities and ideologies to twentieth century urban design, and to debates which have affected cities to the present day (from the political practical discourses of the built environment tradition, to the questionable utopianism of Le Corbusier, Frank Lloyd Wright, and Ebenezer Howard).

More or less symbolic and iconographic readings of Modern buildings in urban settings have more fully interrogated the ideological underpinnings of architectural Modernism. For instance, despite Modern architects' disregard for ornamentation (the dictate that 'form follows function'), it has been shown that the technological, machinic esthetic was enrolled in the creation of certain sociospatial effects. The new technological esthetic was to symbolize health and efficiency – to engender healthier bodies, healthier communities, and ultimately, healthier nations. Thus, it could be shown that certain elements of architectural design (such as white walls and straight lines) were symbolically involved in the re production of urban spaces. This and similar machinic themes have been identified in geographical work at sites as diverse as zoos, social housing blocks, and office buildings.

Geographers have also demonstrated that a more 'critical' approach to architecture is possible in an historical context, by comparing in depth interviews with past residents of Modern communal housing with detailed analyses of the symbolic and practical elements of Modern housing design. It has, therefore, been shown that elements of architectural design at different scales (from the efficient kitchen to the gleaming tower block) have been critical in the production of modern cities.

Theme 2: Power

Geographers' long standing interest in the symbolism of certain interests (whether artistic, historical, political, or economic) in buildings has, in particular, highlighted the importance of power in architectural form. As buildings 'frame' places, they mediate and display certain interests to the detriment or exclusion (and occasionally the explicit oppression) of others. Thus, although some forms of architectural symbolism may be relatively innocent – perhaps referring to a certain artistic style – in many cases, buildings are designed to exert and/or represent the interests of those with the power to build (very often the dominant or ruling class in a given society). It could be argued that one of the most important contributions of textual and iconographic approaches to architecture is in this process of pulling apart the seeming inertness and innocence of buildings, and stressing the key role that they play in shaping power relations in urban space.

Architectural geographers have explored the differential operation of buildings' symbolic power in numerous urban environments, in different periods of history, and with numerous ideological impetuses. For instance, in Renaissance Italy, architects revived Roman architectural design and intended that their grand, symmetrical, and columned buildings express advances in scientific and religious thought (as those buildings represented both geometrical perfection and humanist morality). Similarly, in the 1930s National Socialist (Nazi) Germany, many large buildings were erected as displays of Nazi power, through which the myths of Nazi ideology could be constructed and perpetuated. Geographical research has stressed the complex and contextual relationship in Nazi thought between German identity and Germany's rural landscape. The mythical timelessness of German culture, through a lens of rural nostalgia, was expressed in the deliberately neoclassical style of the Nazis' most spectacular buildings, and in references to 'folk' or vernacular buildings, such as farmhouses and more modest structures (such as Adolf Hitler's farmhouse in Berchtesgaden, Germany). Finally, tall buildings and skyscrapers built during the twentieth century – especially in New York, Chicago, and London – have attracted attention from geographers, as they represent in some ways the ultimate expression of capitalist financial power in urban space. Although skyscrapers may differ in their symbolic content, the need to build higher has been connected with an array of desires amongst the wealthy elite, which range from advertising and the expression of corporate identity to personal statements of power and physical competition with other elites. Today, the process of skyscraper construction continues apace in many major cities around the world.

Theme 3: Play

With influences from consumption, leisure, and tourism research, geographers have turned increasingly to leisure in urban spaces (once more largely in the Minority World). Urban geographers and theorists have debated the overt symbolism of postmodern, 'hyper real' spaces of consumption in touristic and fantasy cities, such as Las Vegas. In terms of architecture itself, researchers have focused almost exclusively on the shopping mall as the key site at which postmodern spaces of consumption are affected most conspicuously by architectural design. Drawing on Marxist and structuration theory, geographers have shown that the shopping mall – like other built environments – both mirrors and reproduces particular social relations, geared in this case to the mass purchase of consumer goods. These architectural geographers fuse material historical and symbolic analyses of shopping malls to show that the shopping mall operates as a complete environment in which to live out (white, middle class, middle aged, and family biased) leisure fantasies. Herein, the basic physical structure of the mall, its controlled atmosphere, and constant visual references to fantasy landscapes (such as tropical beaches and Parisian arcades) are all designed to subtly and not so subtly persuade visitors to buy. The most extreme and well cited (but much maligned) example is Las Vegas – where the mall is pushed to its most spectacular incarnation in conjunction with the casino: here, the fantasy is often so wildly a contextual that reference to external reality is deliberately foregone (**Figure 3**). Indeed, the almost alien landscape of Las Vegas has been likened to a 'heterotopia', wherein the patchwork of themed casinos lining the 'Strip' operates as the most extreme reflection and ironic reinterpretation of the greatest excesses of contemporary consumer society – yet, located in the middle of the vast Nevadan desert. Although, on the whole, they have failed to engage with the ordinary users

Figure 3 New York, New York: a super-casino in Las Vegas. Las Vegas includes many examples of 'playful' urban architecture. Author's photograph.

of such malls themselves, geographers have critically contrasted the symbolism of the mall's interior with both the overt surveillance and security measures used to exclude 'undesirables', and its inward lookingness with respect to its suburban context.

The notion of 'play' has adopted a somewhat different and more critical role in historical and cultural geographies of 'alternative' social movements, such as those of the Dadaists, Surrealists, and the Situationists. The Situationists' modes of urban exploration (begun during the 1960s), in particular, have attracted attention from geographers; moreover, the Situationists' work exemplifies forms of playful architectural experimentation that have characterized various 'radical' architectural practices. Although very few of these alternative designs were built, their implications for urban space and the theoretical linkages with urban theorists such as Henri Lefebvre have concerned geographers since the 1990s. Within the context of these linkages, the alternative visions of the city propounded by the Situationists and others have been critical elements in urban geographers' contemporary theorizations of city life. Employing approaches somewhat other than iconographic or ethnographic research (and in fact closer to historical geographies of Modern architecture), urban geographers have asserted the importance of alternative, artistic, and ironic interventions into urban space which attempt to subvert the consumerist and undemocratic cities through which global capitalism operates today. In a less formal yet more participatory sense, geographers are also becoming actively engaged with radical artists, designers, and activists who – drawing on the Situationists and others – are seeking to invigorate and democratize urban spaces.

Theme 4: Globalization

Perhaps the most explicit marriage of political economic geography and architectural geography has emerged in research on the global city. Although the effects of globalization are patchy and contestable, in a very simplistic sense, a direct effect of globalization has been the emergence of a so called 'International Style' of architecture (**Figure 4**). This style is a continuation of both Modern architecture and skyscraper construction – typified by the monolithic, glass, steel, and concrete mega structures of corporate headquarters, office blocks, and hotels found today in most major cities, in both the Majority and Minority Worlds.

Despite such apparent homogeneity, the implications of these mega structures are diverse, with linkages to all of the other themes outlined in this article. For some commentators, the presence, size, and form of such structures can be ascribed to the patchy investment of capital in cities, such that urban spaces, such as Docklands in London, Tokyo, Sydney, Rio de Janeiro, and Mexico City, are interpreted as the outcomes of late modern capitalism and attendant developmental modernization impulses. Similarly, others have suggested that the convergence of Postmodern, historicist architectural styles (including ornamentation and *faux* references to local context) with the International Style simply

Figure 4 An example of the International Style(s) in Boston, USA. Author's photograph.

reasserts the importance of late modern global capitalism in a symbolic sense. Herein, buildings such as the AT&T Building in New York City are merely surfaces, inscribed with meanings which relate to other architectural styles rather than their own context and function – and rendering their 'textuality' somewhat illegible. The design of such mega structures has also been criticized via related debates about the loss of 'public space' in cities, wherein seemingly public (but privately owned) spaces in urban offices and shopping malls are controlled via security gates, guards, fences, and closed circuit television (CCTV) systems, with a range of associated implications.

Of the many developments affecting cities on a global scale, one further process is particularly pertinent to architectural geographers. The advent of gentrification has entailed substantial investment in former working class housing in larger cities around the world. Although the precise nature, causes, and value of gentrification are still disputed, geographers have demonstrated that the material effects of such investment include the improvement (at least on a surface level) of the built environment. This improvement may entail the refurbishment and inhabitation of existing housing or industrial building stock by the upper middle classes in some areas (such as New York City's Greenwich Village), or the construction of new housing, shopping malls, and office blocks in International or Postmodern styles. Although architecture has not always been the principal focus for research, gentrification highlights the ways in which buildings and associated practices operate as compounding agents of what are, for many commentators, uneven and socially iniquitous changes in urban areas.

Theme 5: Nostalgia

The relatively recent advent of nostalgia (and especially 'neotraditionalism' or the 'new vernacular') in urban architectural design has proven as important and controversial as the development of shopping malls, mega structures and gentrified buildings. The roots of nostalgic design are complex, in that newer building forms in many cultures tend to refer to those of older forms, for diverse reasons. It can be argued, for instance, that the design of suburban housing in the UK often draws on earlier British Tudor and Georgian architectural styles. Yet the nostalgic drive in urban architecture goes beyond an occasional reference to a particular style in, otherwise, relatively homogeneous mass housing schemes, and beyond simple esthetic concerns. For instance, by combining a political economy approach, iconography, and community studies, some geographers have shown that even established suburbs are riven with deeply moralized and mythicized precepts about who should live there, how inhabitants should act, how buildings should look, and how the landscape should refer to imagined historical notions of 'nature' or 'rurality'.

More recently, Postmodern, neotraditional styles of building and urban design – that knowingly rework historical styles – have attracted attention from geographers. High profile new towns such as Seaside and Celebration in Florida have been heralded by conservatives as exemplars of a return to long lost community values. However, many critical geographers have decoded the sanitized, historicist facades of the houses – and symbols of 'community' in their surrounding landscapes – to argue that neotraditional housing developments offer very particular, narrow, and often exclusionary notions of community. In some cases, the symbolism and icon ography of such nostalgic developments has been linked to associated advertising material to demonstrate that such esthetic concerns are, in part, a relatively subtle marketing exercise, designed to appeal to the relatively conservative lifestyles of middle class target consumers. In other cases, the addition of more overt security measures, such as perimeter fences, CCTV, and con trolled entrances (giving rise to the so called 'gated community'), has led geographers to reassert the power of urban and architectural design to facilitate the creation of controlled enclaves for certain groups: usually white, middle class, traditional, nuclear families. Finally, through ethnographic research, geographers have ques tioned the presence of implicit 'rules' in some neotradi tional developments (such as the color one may paint one's house), that serve to reinforce particular social and economic norms via daily practice and inhabitation. From these critiques, and drawing on the insights of some academics from the built environment tradition such as Jane Jacobs, geographers have been able to raise both theoretical and practical suggestions as to how more socially inclusive (especially in terms of race and gender), vibrant, and open urban spaces may be designed.

Conclusion

Geographers' study of urban architecture has followed a number of theoretical and methodological trends in urban geography and cultural geography. It would be difficult to argue that a true subdiscipline of 'urban architectural geography' has ever existed – as geog raphers who study urban architecture regularly draw on issues as diverse as urban planning and design, icon ography, social difference, local historical context, and everyday practice in the course of their work. Yet certain discrete (but not separate) strands in urban architectural research can be discerned: from the Berkeley School through the built environment tradition and behavioral ism, through Marxism and political economy, through textual and iconographic approaches, and into critical, material, and nonrepresentational research. Geographers of urban architecture have considered themes as varied as (but not limited to) machines, power, play, globalization, and nostalgia.

There are many thematic and methodological issues that geographers of urban architecture have yet to ex plore, or have only recently begun to explore. Four sig nificant issues stand out, although the reader may identify many more. First, the relationship between urban archi tecture and certain social groups (especially young peo ple) requires greater and more nuanced consideration. Second, disparate trends in many countries toward eco logical and sustainable design offer critical geographers of urban architecture the opportunity to debate the socio spatial implications of 'green' modes of urban building. Third, recent work by geographers on virtual reality and the Internet has yet to be fully explored by geographers of urban architecture as architects increasingly move to embrace such technologies in their designs. Finally, re search on urban architecture in the Majority World re mains patchy; yet the unprecedented growth of countries such as China signals many, diverse, and potentially problematic urban processes to whose study geographers of architecture can and should contribute.

See also: Berkeley School; Gentrification; Public Space; World/Global Cities.

Further Reading

Al Hindi, K. F. and Stadder, C. (1997). The hidden histories and geographies of neotraditional town planning: The case of Seaside, Florida. *Environment and Planning D: Society and Space* 15, 349 372.
Borden, I., Jerr, J., Pivaro, A. and Rendell, J. (eds.) (1996). *Strangely Familiar: Narratives of Architecture in the City*. London: Routledge.
Chaplain, S. (2000). Heterotopia deserta: Las Vegas and other spaces In Borden, I. & Rendell, J. (eds.) *Intersections: Architectural Histories and Critical Theories* (1st edn.), pp 203 219. London: Routledge.
Cosgrove, D. and Jackson, P. (1987). New directions in cultural geography. *Area* 19, 95 101.
Cuthbert, A. R. (ed.) (2003). *Designing Cities: Critical Readings in Urban Design*. Oxford: Blackwell.
Domosh, M. (1989). A method for interpreting landscape: A case study of the New York World Building. *Area* 21, 347 355.
Ellin, N. (1996). *Postmodern Urbanism*. Oxford: Blackwell.
Goss, J. (1988). The built environment and social theory: Towards an architectural geography. *Professional Geographer* 40, 392 403.
Gruffudd, P. (2003). Building sites: Cultural geographies of architecture and place making. In Blunt, A., Gruffudd, P., May, J., Ogborn, M. & Pinder, D. (eds.) *Cultural Geography in Practice* (1st edn.), pp 238 254. London: Arnold.
Hosagrahar, J. (2005). *Indigenous Modernities: Negotiating Architecture and Urbanism*. London: Routledge.
Jacobs, J. (2006). A geography of big things. *Cultural Geographies* 13, 1 27.
King, A. (2004). *Spaces of Global Cultures: Architecture, Urbanism, Identity*. London: Routledge.
Lees, L. (2001). Towards a critical geography of architecture: The case of an ersatz Colosseum. *Ecumene* 8, 51 86.
Llewellyn, M. (2004). 'Urban village' or 'white house': Envisioned spaces, experienced places, and everyday life at Kensal House, London in the 1930s. *Environment and Planning D: Society and Space* 22, 229 249.
Pinder, D. (2005). *Visions of the City*. Edinburgh: Edinburgh University Press.
Soja, E. (2000). *Postmetropolis: Critical Studies of Cities and Regions*. Oxford: Blackwell.

Relevant Websites

http://www.architecture.com
The website of the Royal Institute of British Architects (RIBA), which contains information, images, and links to other architectural sites.

Urban Design

E. Street, King's College London, London, UK

Glossary

Broken Windows Theory A theory that has its origins in a 1982 paper by Kelling and Wilson. It suggests that poor maintenance of the urban fabric can lead to high rates of crime, particularly vandalism, and may result in the breakdown of community.

Environmental Determinism It is premised upon the idea that physical environmental conditions determine cultural and social outcomes. The theory, now largely discredited, has been used to understand differences between different social groups.

Modernism A style of architecture and approach to urban design that was popular in the early twentieth century. Modernist styles are characterized by clean lines, open spaces, and the use of new technologies and materials.

New Urbanism An approach to urban planning and design that originated in the US during the 1980s. New urbanists consider the physical environment key to creating economically, socially, and culturally successful urban areas.

The Garden City A planned, self-contained development originally proposed by Ebenezer Howard as a way to alleviate urban overcrowding and poor living standards in late nineteenth-century Britain.

Introduction

The built environment is composed of an intricate web of roads, pathways, open spaces, and buildings that, collectively, form part of the physical urban fabric. Although the practice of intervening in the built environment to shape and guide the appearance and function of settlements has occurred for millennia, the term 'urban design' emerged relatively recently. Kevin Lynch, whose theories about city form have been highly influential, suggests that urban design emerged as a profession in the 1960s as a response to a crisis in city planning. Lynch was critical of what he saw as attempts to design the entire city based on a single image, and regarded modernist architects and planners as being out of touch with the realities of city residents' lives. Lynch argues that values shape cities, and that, therefore, imagination and plurality of thought is required in attempting to understand and design cities.

The design of cities mediates our experience of urban environments in a series of complex ways. The shape of a building can influence the esthetic form and qualities of places, and can contribute to creating a certain sense of place. This is particularly so with signature architecture, or landmark buildings, such as the Guggenheim Museum in Bilbao, designed by Frank Gehry. Likewise, physical design can shape personal, emotional and practical, experiences of the city. For instance, a building with a stepped entrance may exclude a disabled person from ease of access to the built environment, while large scale road systems often serve to divide places up and render pedestrian movement between them difficult. Urban design is also at the forefront of government policies to improve the livability of cities, and create places characterized, increasingly, by green and open spaces, high density development, and mixed land uses. The practices of urban design are primarily concerned with place making, which has captured geographical imaginations. The study of urban design is an interdisciplinary affair and, reflecting this, the work of a range of scholars from several disciplines has shaped geographical thought about the design of city spaces.

Re(making) Urban Design

Allan Cochrane has suggested that urbanization has long been characterized by attempts to count, map, and record. The most powerful expression of this, he argues, has been the emergence of the profession of urban design, which encompasses the practice of architecture and urban planning. According to Cochrane, one of the primary aims of the practice of urban design is to shape, simplify, and manage the complexity of urban life. However, despite this, he stresses that urban form and design remain open to interpretation. City residents should be seen as active participants in processes of urban design since they are continually (re)interpreting, (re)translating, and (re)defining the urban language. For example, pedestrians may, over time, subtly rewrite planned urban thoroughfares by stepping off pavements and creating new walkways. Spaces and buildings can also be used by residents in ways beyond their intended purpose. For example, civic squares may be claimed, and perhaps eventually redefined, as sites of protest or festival.

Ley and Cybriwsky describe how, during the 1960s and 1970s, graffiti was used, primarily by black youths, as a tool to mark and symbolically reclaim parts of the city outside of the ghetto. Popularized by the rap industry during 1980s and 1990s, today works by the illusive urban graffiti artist Banksy hang in the homes of style conscious collectors as well as sprayed on walls in several major cities (**Figure 1**). Eradicating graffiti has been central to several high profile urban crime initiatives, many of which have drawn upon scholarly debates that emphasize the role urban environments play in creating safe communities. Kelling and Wilson's 'broken windows theory' is based upon the premise that poor maintenance of the physical fabric, such as leaving a broken window unrepaired, conveys the message that nobody cares about the area. This, it is argued, can lead to an increase in vandalism, the degeneration of community spirit, and set in motion a cycle of decay. Kelling and Wilson's ideas have been influential in policy circles and were perhaps most famously taken up by Rudolph Giuliani, then mayor of New York, during the 1990s, in response to rising crime rates in the city. Giuliani, along with his Chief of Police, William Bratton, instigated a 'get tough' approach to crime, clamping down on minor crimes, such as graffiti which they considered to be a sign of disorder, as well as more serious felonies.

The UK government has also made links between the quality of the physical urban fabric and the safety and vitality of cities and towns. Much of the origins of current British urban policy can be traced back to the publication of *Towards a New Urban Renaissance*. Authored by architect Lord Richard Rogers, the report suggests that poor quality urban environments have contributed to an exodus of people from English cities and towns. Key to attracting these populations back, the report claims, is a commitment to quality and creativity in the design of buildings, public spaces, and transport networks. Scholars such as Loretta Lees have critiqued the government's adoption of a renaissance rhetoric, suggesting that the use of 'neutered terms' such as 'renaissance' serve to downplay the class conflicts likely to emerge from attempts to attract middle class resident back to cities.

British urban policy continues to emphasize the importance of urban design in achieving economic success, social cohesion, and cultural vitality. There is some evidence to suggest that design can increase the economic viability of developments and deliver enhanced social benefits. However, defining what good design is can be problematic. Conversely, as Alice Coleman has argued, urban design can be a powerful negative force. In *Utopia on Trial*, Coleman presents the results of a research program where design features such as building heights were tested to establish which ones were associated with antisocial activities such as litter dropping. The researchers identified 15 design variables present in blocks of flats which, they claim, adversely affected residents' behavior. For example, entranceways with closed doors were ten times more likely to be graffiti free than those where doors had been removed. By contrast, the research found that the houses surveyed showed far less sign of social breakdown. Coleman utilized the findings to launch an attack on the utopian visions of designers behind high rise housing estates, which she claims, have been ineffective and costly.

While the extent to which urban design can deliver social and economic benefits remains contested, iconic architecture is being used as a way to 'boost' cities' economic, cultural, and social capital in an increasingly competitive global investment market. Statement urban design can have a powerful impact. Striking designs like the Sydney Opera House have helped to provide cities with globally recognizable images. More recently, buildings such as the Wales Millennium Centre in

Figure 1 Graffiti in New York, USA.

Cardiff, UK, have been used as tools to stimulate processes of regeneration and renewal (**Figure 2**).

As Donald McNeill points out, often, this has much to do with the 'star architect' behind designs. As a result, urban administrations are courting top designers to bring creative flair to cities. David Harvey has situated these developments within a broader critique of urban governance. He claims urban governments' preoccupation with fostering local development and employment has arisen from attempts to recover from de industrialization and the subsequent rise in unemployment experienced in many cities. This has been coupled with what Harvey terms the 'rising tide' of neoconservatism, with its preachings of market rationality and privatization. For Harvey, the resulting entrepreneurial approach to urban governance marks a significant reorientation of patterns of governance away from the managerial approach of the 1960s and 1970s where the provision of services and welfare was a key concern.

Designing and Planning the City

Government intervention in the form and appearance of the city, often to serve specific political functions, is not a recent phenomenon. In nineteenth century Europe, urban design was used as a political tool to control the expanding urban populations associated with rapid industrialization. Gerges Eugene Haussmann's redesign of mid nineteenth century Paris is one of the most striking examples of the way that design can be utilized to produce a regularized urban spatial form. The urban design of cities was also seen as an essential component in fostering public health reforms. However, beyond this, design was used by governments as a way to enforce controls on the population based upon the perceived social, moral, and political dangers associated with growing slum areas. Throughout the early to mid twentieth century, there were major slum clearance programs in operation in most major cities. Many residents were moved to purpose built estates, often on the periphery of urban areas that were designed to offer a healthier lifestyle. As Robert Beauregard has suggested, this period was characterized by a sense of governmental optimism, and officials often considered the ills of urban life to be opportunities for reform.

The rapid changes associated with industrialization presented many challenges to those charged with governing cities. Designers were tasked with creating spaces that would accommodate these new and unfamiliar urban futures. Writing at the start of the twentieth century, Ebenezer Howard expressed alarm at the movement of people from the country into already overcrowded cities. However, rather than send people back to the countryside, his solution was to design spaces that combined the beauty of the countryside with the energy, activity, and job opportunities of the city (**Figure 3**). The resulting 'Garden City', was a planned, self contained community consisting of wide boulevards leading to a central garden and surrounded by a range of public buildings. Howard's vision was based upon an idealized image of a place that would bring some of the 'good life' to all Garden City residents.

Howard was not the only designer striving to create spaces that would improve living standards. Disillusioned with modern cities, Frank Lloyd Wright, one of twentieth century America's most influential architects,

Figure 2 Wales Millennium Centre.

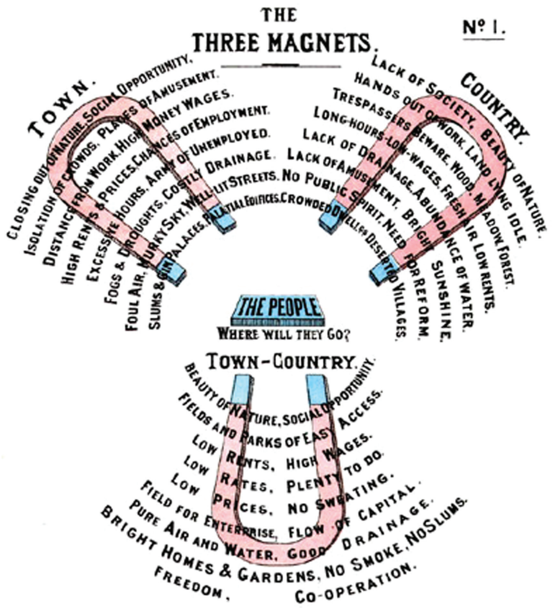

Figure 3 Howard's 'three magnets' diagram showing the benefits and difficulties associated with town and country living.

envisaged a new form of city living he called 'Broadacre City'. Broadacre was spread over 16 square miles and integrated public, commercial, and private buildings. It was divided into four areas, each containing a mix of housing types and facilities resulting in a self sustaining community. In addition, Lloyd Wright proposed that all of Broadacre's 1400 families receive a one acre plot of land per family member to use productively. Wright's design was born out of his desire to give residents greater individual control over their environment and engender a sense of community integration.

Very different, but similarly utopian in nature, were Le Corbusier's ideas about the power of design. Widely held to be the twentieth century's most influential architect, Le Corbusier emphasized standardization, geometric lines, and embraced modern technology in his work. In *The City of Tomorrow*, Le Corbusier (1929: 18) equates straight roads with civilization and order, arguing that, without them, "cities sink to nothing and ruling classes are overthrown." Underpinning the ideas of both Lloyd Wright and Le Corbusier was the understanding that spatial order had the capacity to produce a new moral and social order. This notion was at the core of modernist town planning and design and contributed toward the production of spatially segregated or 'zoned' cities.

Le Corbusier's ideas were radical, and attracted admiration and condemnation in almost equal measure. One of his most notable critics was the prominent architectural commentator and urban theorist Lewis Mumford. Mumford was particularly scathing of Le Corbusier's *Plan Voisin*, in which he proposed to tear down and rebuild the historic core of Paris. Mumford described Le Corbusier's imagination as "working like a bulldozer on an urban renewal project" (Mumford, 1968: 120), and suggested that his image of the city combined the three major mistakes of the twentieth century. For Mumford, Le Corbusier, along with other modernist urban designers, were guilty of overvaluing mechanization and standardization without respect for the human purpose to be served. He was also critical of what he saw as designers' readiness to dispose of the past, and their tendency to separate out and overspace facilities which, he argued, needed to be in close proximity for daily use. Interestingly, while Mumford's critique of modernist design and planning often echoed urban author and theorist Jane Jacobs's views, he claimed her ideas were "equally unsound" (Mumford, 1968: 122).

Jacobs's 1961 treatise, *The Death and Life of Great American Cities*, outlined particular conditions she saw as essential to the creation of vibrant urban environments. Jacobs was critical of rigidly controlled modernist planning techniques, and emphasized the importance of organic urban growth. She called for multifunctional districts in cities, as well as the mixing of new and old buildings. Jacobs also suggested that street blocks should be short, and populations densely concentrated. Today, Jacobs's ideas are undergoing something of a renaissance, much of this associated with the popularity of the new urbanism. Emerging in the US during the 1980s as a reaction to the decline of inner city areas and the expansion of urban sprawl, the new urbanism emphasizes the importance of mixed use buildings, social diversity, and compact and pedestrian friendly neighborhoods, to maximize community interaction.

Drawing parallels between the new urbanism and modernist theories of urban design, David Harvey has questioned whether they also perpetuate the idea that spatial form can be used to influence moral order. Harvey's point can be situated within a broader set of debates leveled at urban design theory, namely that attempting to shape social behavior through the design of the urban fabric is a form of environmental determinism. Susan Fainstein has questioned whether the new urbanist aim of achieving diversity through planning is in fact an oxymoron. As Fainstein highlights, a criticism commonly leveled at new urbanism is that it simply produces another form of suburbia, rather than eliminating social segregation in the city. She suggests that while Jacobs's work has fed into new urbanist approaches to design, Jacobs would in fact be critical of attempts to control what she believed should be spontaneous.

Urban Design and Spatial and Social Segregation

The suggestion that urban design can (re)produce patterns of social segregation has not only been aimed at new urbanist communities. The design of cities is char acterized by boundaries and borders that serve to de marcate and separate places. Physical barriers can be used as political tools to segregate and control popu lations. The Berlin Wall, which separated East and West Berlin from 1961 to 1989, has become perhaps the most famous example of the way urban design can be used to divide populations. In the UK, the Cutteslowe Walls became notorious as an example of the ways that urban design can reproduce and reinforce social barriers. Erected in 1934, the walls ran between two Oxford housing estates, one accommodating council tenants, the other, consisting of mainly owner occupied housing, owned by the Urban Housing Company. On hearing that the council estate was to house ex slum dwellers, the company erected two, metal spiked walls between the estates, apparently in response to their residents' fears about living in close proximity to slum clearance tenants (**Figure 4**). Today, urban design continues to be used as a tool to segregate spaces. The West Bank barrier, which Israel began constructing in 2002, runs for almost 700 km through the West Bank occupied Palestinian territories, and continues to be expanded.

Barriered spaces also exist in less obvious, but no less significant, guises than walls and fences. Mike Davis has described how, in Los Angeles, gated districts use in telligent design to keep out unwanted elements, while protecting white collar office workers and middle class tourists. During the 1992 Los Angeles riots, Davis re counts how the city's downtown buildings formed a virtual bunker as security staff mobilized bullet proof steel doors, severing pedestrian walkways and immobil izing escalators. He has termed these intelligent build ings, 'architectural anti heroes', their advanced sensory systems designed to screen and identify human popu lations through motion, sound, temperature, and hu midity, and able to eject undesirables without the need for human intervention.

Segregation in the urban environment can be drawn along many lines, including gender, sexuality, disability, and race. Feminist scholars have engaged with what they perceive to be the masculinized nature of the built en vironment. *Matrix*, a collective of female designers active during the 1980s, argued that the urban design process was male dominated, and that the built environment was therefore man made. Dolores Hayden has argued that stereotypical portrayals of gendered spaces, such as the home as a woman's place and the city as a man's domain, produce an architecture of gender. This, she asserts, re mains at the center of American house building today.

The history of urban design has also been permeated by racialized segregationist policies. The developers be hind the Levittown housing estates, built during the 1950s and early 1960s, wrote a clause into every home owners contract stating that homes at the Long Island development could only be purchased by white people (**Figure 5**). In South Africa, racialized urban segregation was at the heart of policies of apartheid. According to Christopher, the Housing Act (1920), which provided financial assistance to municipalities for house building in poorer sections of the community, dictated racial homogeneity as an administrative requirement. Although

Figure 4 The Cutterslowe Walls.

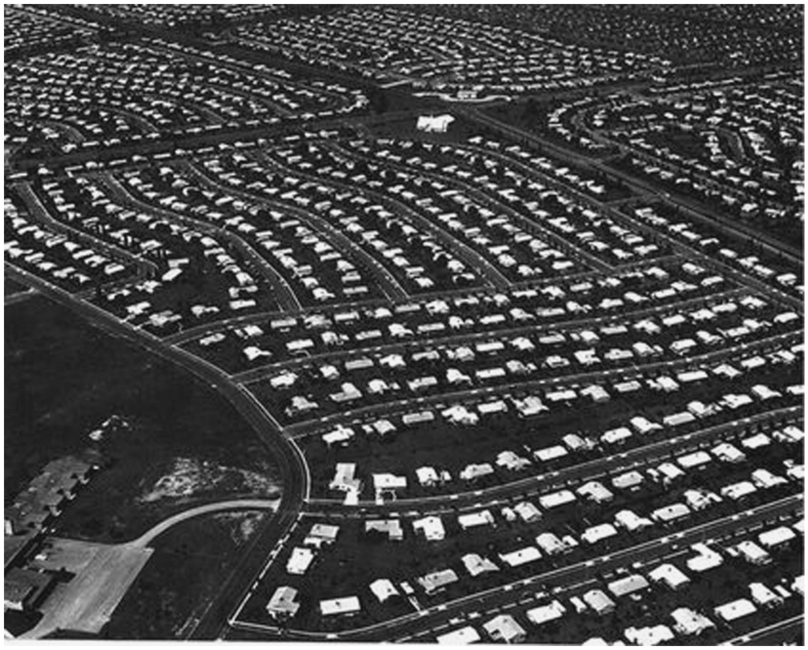

Figure 5 Aerial view of Levittown, Pennsylvania, USA, 1959.

legal apartheid ended in 1991, in a 2001 paper, Chris topher describes how, despite some positive trends, levels of racial segregation remain high.

The design of the city can also produce barriered and bounded spaces for those with physical and/or mental impairments, producing what Rob Imrie has termed, an architectural apartheid. He points to urban features such as underground stations, which, without lift access, can be no go areas for mobility impaired people. As Imrie has suggested, the notion of bounded and barriered space is often complex and hidden and is, in part, reflective of the symbolic and cultural encoding of the city. Such codes are often based upon ideas about 'civilized' or 'normal' bodies. As a result, the disabled body is often absent from theories and practices of urban design. Brendan Gleeson has emphasized that designing barrier free spaces re quires more than adapting the built environment to make it accessible, for example, by using specialized fittings. Rather, he argues, barrier freedom must be won by transforming the political, economic, institutional, and cultural forces that shape urban spaces.

John Turner has also sought to challenge exclusionary views in relation to urban design, calling for an approach to housing that enables residents to contribute toward the design, management, and construction of their own homes. Turner draws on his experiences in less developed countries in his theories. He describes how in the mid 1950s, Pedro Beltran, then Peruvian President of the Commission for Housing and Agricultural reform, visited the *urbanizaciones populares* in Arequipa, Peru, a vast self improving squatter settlement. Beltran perceived the settlement to be a shanty town and slum. However, his views were at odds with many residents' who, according to Turner, took great pride in their homes. Turner questions who has the right to decide who lives where and how, and suggests that the design and provision of housing should be a more autonomous and inclusive process.

These analyses highlight the need to think of urban design as a political process. In this respect, physical structures should be seen as incorporating political and social assumptions and messages. Iris Marion Young has argued that the realization of urban designer's plans can produce abstract spaces of efficiency that can dominate and displace what she terms the "lived space of human movement and interaction." She is critical of what she sees as the increased segregation of urban space, and suggests that the separation of functions, such as work places and residential areas, can lead to the oppression of certain social groups. While it is important to be aware of the barriers and boundaries that urban design can pro duce, it is also important to recall that people are able to resist and transgress boundaries and spaces of exclusion. For example, spaces of tolerance and sexual freedom in the city have emerged in areas such as Soho in London.

Technology and the Design of the City

While experiences of city spaces are mediated by our own personal histories, they are also influenced by ex ternal factors such as technologies that are designed into the urban fabric. One example of this is presence of surveillance technologies such as CCTV, which is now ubiquitous in most major western cities. However, the impacts of such technologies have been felt unevenly. Drawing on Michel Foucault's theories of discipline and punishment, Hille Koskela has argued that surveillance technologies are being used to target and 'normalize' groups commonly perceived to be different, deviant, or suspicious. As a result, she has claimed the homeless, ethnic, and sexual minorities, youths, and activists are all more likely to be subject to the cameras' controlling gaze.

Predating the rise of CCTV technology, Oscar Newman's work considered the ways that design technologies could be utilized to maximize surveillance and defend against crime. Newman suggested the social mechanisms that once kept crime in check, were absent from the modern city. He felt cities had become anonymous places, where people were strangers to one another. Newman called for an architectural model of corrective prevention. This involved strategically positioning windows, and building to medium densities to maximize natural surveillance, reduce opportunities for crime, and improve residents' sense of security.

Other technologies such as the motor vehicle influence how we behave in, and navigate, urban spaces. Roadways are designed technologies, and as we take to them instead of traversing the city by foot, we experience and 'read' the urban environment in a profoundly different way. Susan Robertson has suggested that the view from the car windscreen has become a hegemonic way of experiencing town and country. However, Sheller and Urry argue that despite this, social scientists have failed to grasp the impact of the car in transforming the 'time space scapes' of the urban dweller. They claim that, whereas classic urban architecture was designed from the perspective of (and for) the pedestrian, postmodern architecture emphasizes quickly read signs and surface quotation to be read at high speed from the car (**Figure 6**). Distinct forms of design have grown up around urban routeways. Martin Pawley has termed the large warehouses and chain stores that line roadsides, 'big shed' architecture. He suggests that urban theorists have so far neglected these commercial and retail centers, despite his assertion that they are often more economically powerful than traditional urban centers.

According to Marc Auge, the increased intensity at which we consume and traverse urban spaces has produced 'nonplaces'. These spaces, such as motorway cafes and airport lounges, are characterized by generic, nonplace specific design. Rather than being rooted in the locale, they leave visitors with the impression that they could be anywhere in the world. However, Peter Merriman has argued the landscapes that have grown up around roads are complex and varied. He highlights how in Britain during the 1950s and 1960s, motorway landscapes were considered exciting, modern places to visit. Indeed, Merriman describes how landscape architects involved in designing the M1 motorway attempted to bring modernist ideals of flow and functionality to its construction. Merriman contests the conception of environments such as service stations as nonplaces, arguing that they are significant spaces for users such as lorry drivers and tourists.

The Role of Geography

Geographers have made important inroads in attempting to make sense of the ways in which urban design shapes

Figure 6 Roadside signage design to be read at speed.

cities, and our experiences of them. In recent years, scholars have begun to unpack and problematize the ways that urban design can produce or reinforce social inequalities according to, among other things, class, gender, sexual orientation, disability, and ethnicity. Geographical theorizing, on the role of power, place, and politics, is of central importance to (re)thinking patterns and processes of urban design. However, there remains much more to do in exploring such dialectics. In par ticular, urban design cannot be seen as a technical or value neutral process, but instead is infused by ethical and moral standpoints about what the 'good city' is or ought to be. The esthetics of a place reveal much about social and political structure and process and, in turn, social and political structure and process are revealed, in part, through the form and texture of the built environment.

See also: Disability and Chronic Illness; Gender in the City; Governance, Urban; Housing; Modern City; New Urbanism; Planning, Urban; Segregation, Urban; Urban Architecture.

Further Reading

Augé, M. (1995). *Non Places: Introduction to an Anthropology of Supermodernity*. London: Verso.
Beauregard, R. A. (2003). *Voices of Decline: The Postwar Fate of US Cities*. New York: Routledge.
Carmona, M., de Magalhaes, C. and Edwards, M. (2002). What value urban design? *Urban Design* 7, 63 81.
Cochrane, A. (1999). Urban design. In Pile, S., Brook, C. & Mooney, G. (eds.) *Unruly Cities?* pp 103 138. London: Routledge.
Coleman, A. (1985). *Utopia on Trial*. London: Hilary Shipman.
Davis, M. (1993). *Ecology of Fear*. New York: Vintage.
Fainstein, S. (2000). New directions in planning theory. *Urban Affairs Review* 35, 451 478.
Harvey, D. (1989). From managerialism to entrepreneurialism: The transformation in urban governance in late capitalism. *Geografiska Annaler, Series B, Human Geography* 71(1), 3 17.
Hayden, D. (2002). *Redesigning the American Dream: Gender, Housing and Family Life*. New York: W.W. Norton and Company.
Howard, E. (1902). *Garden Cities of Tomorrow*. London: Swan Sonnenscein & Co.
Imrie, R. (2001). Barriered and bounded places and the spatialities of disability. *Urban Studies* 38(2), 231 237.
Jacobs, J. (1961). *The Death and Life of Great American Cities*. New York: Random House.
Kelling, G. L. and Wilson, J. Q. (1982). Broken windows: The police and neighborhood safety. *The Atlantic Monthly* 249(3), 29 38.
Le Corbusier (1929). *The City of Tomorrow*. London: Architectural Press.
Lynch, K. (1981). *A Theory of Good City Form*. Cambridge, MA: MIT Press.
Mumford, L. (1968). *The Urban Prospect*. London: Secker and Warburg.
Newman, O. (1972). *Defensible Space*. New York: Macmillan.
Sheller, M. and Urry, J. (2000). The city and the car. *International Journal of Urban and Regional Research* 24(1), 737 775.
Turner, J. F. C. (1976). *Housing by People*. London: Marion Boyars.
Young, I. M. (1990). *Justice and the Politics of Difference*. Princeton, NJ: Princeton University Press.

Relevant Websites

http://www.cnu.org
Congress for New Urbanism.
http://www.rudi.net/
Resource for Urban Design Information.
http://www.architecture.com
Royal Institute of British Architects.
http://www.aia.org/
The American Institute of Architects.
http://www.udg.org.uk
Urban Design Group.

Urban Growth Machine

S. Rodgers, The Open University, Milton Keynes, UK

Glossary

Exchange Value The monetary return that a given piece of property generates. Adapted from Marxian economic theory.

Idealism The principle that reality is constituted wholly or in large part through human thought or perception, rather than objective material reality. Sometimes used as a pejorative label.

Rentier Someone who derives a monetary return from property holdings.

Thesis A proposition put forward for further discussion and empirical consideration.

Use Value The personal or social utilities a given piece of property generates (e.g., as a home and recreational site). Adapted from Marxian economic theory.

Value-Free Development The doctrine that free markets should determine land use and/or that urban development is inherently a collective good.

Voluntarism The principle that deliberate human will or action is the fundamental force constituting the social world. Sometimes used as a pejorative label.

Overview

Urban growth machine is an influential thesis of urban politics that suggests the objective of growth unites otherwise pluralistic interests in relation to a city. The thesis is situated within a broader theory about the commodification of place, where place is understood to be socially and economically valued land. Its key premise is that coalitions of actors and organizations (i.e., growth machines), all sharing an interest in local growth and its effects on land values, compete with growth machines elsewhere for scarce mobile capital investment, while simultaneously attempting to gain the tacit support of local publics for such urban growth.

In the mid 1970s, three areas of writing dominated American theories of urban politics: (1) the longstanding community power debates, polarized into theories of elite power and democratic pluralism; (2) work in the Chicago School tradition of human ecology; and (3) the rising influence of Marxian political economy. For Harvey Molotch, an American sociologist, the first area represented a largely stagnant debate, overly focused on identifying the agents of political power without questioning their motivations. The latter two areas did get at some of this missing question, but were, respectively, overly focused on the cultural peculiarities of urban place, or the determinate nature of capital relations. Molotch sought a middle ground between all of these areas, placing a strong focus on the activities of people and institutions in shaping and contesting what he argued to be central to a specifically urban politics: land, and its political, economic, and social construction as place. So important was this politics of place, suggested Molotch, that it compelled otherwise pluralistic elites to join together in their collective interest for urban growth; they became growth machines.

This portrait of an American urban politics driven by growth machines was later elaborated by Molotch, writing with John Logan, in the 1980s as part of a more general theory about the commodification of place. Yet, most writing, including that of human geographers, has largely focused on the growth machine thesis specifically. One possible reason for this is that the thesis offered a framework that, particularly for those disillusioned with abstracted theories of economy or capital, presented a new way to apprehend the agency of people and institutions in urban politics. In joining such an agency focus with an emphasis on power, the growth machine thesis made a partial, qualified return to the agency centered 1950s elite theories of urban politics, held in check by a more specific focus on the politics of property and growth.

Subsequent sections detail the key concepts, as well as issues and debates, connected to the growth machine thesis. However, it is worth noting here the four main ways that the thesis has been received in the academy, and particularly by human geographers. First, in focusing upon disputes and action related to land property, the thesis has been seen as specifying an analytical frame for investigating a politics genuinely directed to an urban object (it has also been asserted as an authentic urban sociology). Second, the thesis has been taken as a trenchantly critical assessment of elite power in cities, and the forms, motivations, and effects through which elites pursue their narrow collective objectives, even as such objectives are promoted as the wider public good. Third, the critical edge to the thesis has in turn suggested the empirical existence of and possibility for resistance to growth machines, the most researched of which is probably the rise of the environmental movement in American cities. Finally, after some keen initial interest internationally, use of the thesis has waned, increasingly being seen as too particular to the urban situation in the United States. While this is not to suggest that it has become completely outmoded, most urban politics research has spun off toward a range of newer, and putatively more sophisticated and flexible, theories of urban politics.

Key Concepts

The term growth machine (or growth coalition) has been widely deployed in writing on urban politics, though not always with close fidelity to the conceptual apparatus or intellectual traditions on which the thesis depends. Nevertheless, the thesis rests on a set of fairly explicit and identifiable key concepts.

Use Value, Exchange Value, and Place

To begin with, the thesis relies fundamentally on place as a concept, and particularly how place becomes commodified. It assumes that in market societies, and especially in the United States, place is not only a basis for carrying on life but also an object from which to derive wealth. Place is defined here in a quite particular way: as the outcome of social activities and constructions seeking to stake out a living, or otherwise a monetary return, from a piece of land. Therefore, there is a fundamental conflict over urban places, which the growth machine thesis operationalizes with reference to the Marxian notions of exchange value and use value. Individuals or groups seeking exchange value hope to generate some form of rent from real estate, while for those seeking use value, the same real estate might form the basis for everyday social life. The somewhat amorphous category of residents is seen as primarily concerned with use, while narrower political coalitions of individuals and organizations – growth machines – are oriented toward exchange values.

It is important here to avoid viewing exchange value as an abstract or predetermined economic relation. Rather, the growth machine thesis posits exchange value as a product of concerted activities on the part of various actors to make money off real estate. Since land property, like labor, is not produced but something existing in finite amount, the market for land is intrinsically monopolistic. One parcel of land does not usually perfectly substitute for another. But by working to change the content of their property (e.g., by influencing zoning), and ensuring that certain qualities or conditions exist in relation to it (e.g., transport, services, policing, and the uses of surrounding properties), land owners commodify place and therefore enhance the possible rent they can derive from their particular slice of the property market. It is this exchange seeking activity that designates a particular type of actor hypothesized to be at the core of place commodification: the place entrepreneur.

Place Entrepreneurs

While the activities of a whole series of different individuals and groups are oriented to deriving exchange values from land, place entrepreneurs have a particularly central role in this arena. Sometimes called modern rentiers, in a loose analogy to feudal land holders, place entrepreneurs are those directly involved in the exchange of, and collection of rents from, land. Broadly speaking, place entrepreneurs fall into one of three types. First are those that gain land holdings through inheritance or some other chance situation. These place entrepreneurs are basically passive when it comes to developing their property. They are land owners who simply collect rents or sell their property in ways characteristic of the traditional rentier. A second type of place entrepreneur is more active in the sense that their involvement in buying and selling land is conducted with some understanding of urban change and thus of possible future land uses. These place entrepreneurs rely on a variable proficiency to predict and estimate future changes in land values, so as to make a profit by strategically buying and selling real estate.

Finally, the most active and perhaps most important place entrepreneurs are the structural speculators. Unlike the other land speculators described above, these land owners do not merely estimate future land values, but intervene in the wider sphere of decision making, regulation, and investment outside of their direct control, yet affecting their holdings. They seek to produce a particular set of conditions and relationships to increase the value of their property. These place entrepreneurs, often embodied by more complex organizations, make specific and targeted efforts at influencing an array of relevant decision makers. At the same time, these modern rentiers are most likely to form the core of broader political coalitions seeking to encourage a more general objective: urban growth.

Growth Machines and Their Allies

Perhaps the most noted aspect of the growth machine thesis – and where it intersects with some other theories of urban politics (such as urban regime theory) – is its suggestion that elite coalitions tend to have an overwhelming influence on the politics of cities. The growth machine thesis is distinct, however, by not only pinpointing particularly powerful actors and organizations, but also placing special emphasis on their common motivation for urban growth. So, while acknowledging that a plurality of interests make up political power in cities, growth is forwarded as that which binds them together: they become growth coalitions.

This collective drive for growth is situated in the broader theory discussed above about place commodification. Growth coalitions (or machines) are driven by those who have the time and resources to participate in local politics; have particular interests in political decisions around property; and share a desire for urban growth and development. Unsurprisingly, place entrepreneurs are seen to be the core figures of such growth coalitions, along with others connected to deriving rent from property, such as property investors, developers, financiers, and so on. While this core group – sometimes collectively labeled a rentier class – is seen as central to driving urban growth in a given locality, they are nevertheless typically seen as closely associated with at least three other major affiliates.

First are local politicians, who, despite internal differences, are all seen as under tremendous pressure to fall in line with a general consensus for growth, not least for their political survival. More importantly, however, local politicians are empowered in relation to the local government apparatus, a jurisdiction with the most significant influence on decisions related to land. Second are local media, and particularly the metropolitan newspaper, important because their preferences for growth are seen as less particularistic than many other coalition members. Newspapers are conceived as a kind of mediator, acting both publicly, by pronouncing on what is good growth, and outside of the public eye, through informal social relationships between the newspaper proprietor and coalition members. As will be noted below, newspapers are also seen to be crucial in promoting growth coalition objectives to wider urban publics. Finally are utilities, such as water or public transport agencies, which similar to local media are less particularistic about growth, often taking on a mediating role.

While the above actors are seen to form the core of the coalition, its influence is seen to often extend even further, to a wide range of allies that variably promote and support the common objective of growth. These include professional sports teams, organized labor, small retailers, corporations, universities, and cultural institutions. Exactly why growth coalitions and their allies are compelled toward such collective action is explained with reference to a twin orientation: on the one hand to mobile capital, and on the other hand to urban publics.

Organizing for Mobile Capital

An important factor in uniting diverse interests around urban growth is the limited amount of mobile capital, and the presumed necessity of attracting such investment to the locality around which a growth coalition is oriented. In this sense, growth coalitions are essentially regarded as territorially organized collectives that see themselves as competing with similar collectives elsewhere for mobile capital investment. Of course, this opens up the possibility that growth machines might organize around territories that are not urban (however defined), for example, regions or nations. Yet, as already noted, the growth machine thesis is at heart about the urban. Although cities might be argued to be amorphous, or not obviously delineated, the assumption made implicitly in the thesis is that growth machines largely organize around the jurisdictional space of local government in pursuing mobile capital.

Promoting Growth as a Public Good

Equally important for growth coalitions is gaining the tacit support of wider urban publics. To do this, growth coalitions are said to propagate an ideology of urban growth as value free. In other words, they de emphasize the exchange value benefits of growth for narrower groups, and emphasize growth as an inherent collective good that will enhance the lives of regular people. Engendering public support for growth into the foreseeable future is particularly important in relation to the image that growth coalitions might be able to project to mobile investors.

Exactly how such ideologies are circulated or fostered is complex. In general, however, the growth machine thesis suggests that widely held local identities and civic pride are tied in various ways to urban growth as an inherent good. This potentially crosses a number of spheres, from the ways in which local history is taught in school curricula to the boosting and supporting of local sports teams. Local newspapers are suggested to be particularly central in instilling local ideologies, since they are often seen as casting various urban development projects as coincident with the wider good of the city or region, usually emphasizing urban pride and greatness.

Issues and Debates

Thus far, the notion of urban growth machines has been presented as a relatively coherent thesis of urban sociology, as set out especially through the work of Harvey Molotch and John Logan. However, it is a thesis that, particularly among geographers, has been an object of extensive discussion and critique, both of which have contributed to the remit of the thesis for studying and theorizing the political geographies of cities.

The Property Focus

The most direct critique of the growth machine thesis is the central importance placed on land property. To begin with, concerns have been raised about the broader theory of place commodification in which the growth machine thesis is situated. Many Marxian writers have critiqued how the theory deploys a distinction between use value and exchange value solely as they relate to matters of place (or socially constructed land). So called residents, for example, can also be argued to pursue exchange values, not least because many must sell their labor. Tying people solely to their use values in the city is therefore argued as too narrowly cast, creating a ro manticized image of residents battling growth machines (and vice versa).

Others have pointed to the changing constitution of property markets, as well as new strategies for attracting mobile investment, and argued that the notion of locally tied, single purpose, powerful place entrepreneurs is increasingly problematic. The property industries have become more and more complex; there has been a rise of international property firms, and property investments are used by non real estate organizations to spread risks. At the same time, contemporary strategies of local au thorities or partnerships, for example, toward mobile investors tend to go well beyond concerns of property development alone, focusing on matters such as regional research and training, quality of life, and increasingly the elusive notion of creativity.

The broadest critique around the property focus, however, is simply that it leaves too many other forms of urban politics out of the picture. One glaring omission, considering its longstanding debate in European litera ture on urban politics, is the politics surrounding the collective consumption of various welfare services, con tested particularly through urban social movements. In addition to this, there have been arguments made for other potentially urban political concerns – for example, citizenship, gender, ethnicity, security, and more – that, even if sometimes connected to property development, are hardly reducible to such matters.

The Human Agency Focus

Another area of scrutiny is the deliberate focus that the growth machine thesis places on human agency – the activities and social constructions of individuals and groups – which writers, particular from Marxian or structuralist traditions, have accused of both voluntarism and idealism. The principal target for such critiques is the emphasis the growth machine thesis places on how property entrepreneurs and others take action, organize, dream, and desire in relation to cities, or urban places. Many have seen this approach as theoretically deficient, developing its theses based on the so called level of ap pearances: what actors are empirically seen to do. For those writing from a structuralist standpoint, as important as what human actors do are the social relations that give rise to such forms of action in the first place. Without a theorization of such social relations, it is argued, the growth machine thesis paints a portrait of particular elites with an improbable scope of power in local politics.

In addition to accusations that the growth machine thesis is poorly theorized, others have also suggested that its assertion of human agency is not backed by a particu larly clear methodology. It has been pointed out that, first, the original thesis was based not on original research but a synthesis of many previous studies, often with contrasting agendas. How these studies were assembled to construct the thesis is unclear, and moreover, given the reliance on past studies, some argue that the thesis might primarily point to a somewhat outmoded empirical situation. Second, others have observed that subsequent research building on the growth machine thesis has most often avoided the direct study of coalitions, instead conducting case studies on the relationships between, for example, development projects or urban policy strategies and growth coalitions. Therefore, the complex sets of agency supposedly making up growth coalitions remain in empirical practice a largely unexamined, independent variable.

Difficulties in International Application

Questions around the agency focus in the growth machine thesis – made on both methodological and theoretical grounds – have also been used to suggest that the thesis is poorly contextualized, leading to problems in inter national application and comparison. While the pro genitors of the growth machine thesis strongly emphasized its relevance to the urban situation in the United States first and foremost, they and others have often hinted at its possible use in other contexts with private property markets. Indeed, there has been no shortage of attempts to deploy the thesis, or to debate and critique its application, in non American contexts, in cluding cities in Australia, Britain, Canada, China, France, Germany, Israel, Italy, Japan, and the Netherlands.

The international setting in which the growth machine thesis has been most frequently applied, and certainly most debated, has been cities in Britain. Into the late 1980s, Britain had seen almost a decade of market led reform brought on by successive Conservative governments. As part of such reform, new area based economic develop ment agencies, public–private partnerships, and re structured local government all increasingly engaged in various property led urban programs. In hindsight, it is likely that such unprecedented change served as a major impetus for many British urban scholars to embrace American urban political economy, such as the growth machine thesis, as a highly salient heuristic device for understanding what was happening on the ground. Quickly following a spate of early studies, however, were critiques of attempts to use such imported theories to understand British urban politics. Only at a metaphorical or relatively superficial level, many argued, could things like rising business involvement in urban governance be labeled as growth machines. The most forcefully put reason for this was that the rising private sector involvement in local British politics arose less from local business activism, but largely through explicit central government policy.

The debates around the relevance of the growth machine thesis in Britain and elsewhere have somewhat ironically led to a clearer identification of several peculiarities of American cities. These include: a strong private sector – and correspondingly weak state – presence in property acquisition, development, ownership, and servicing; a relatively regionalized banking and finance industry (if less and less the case from the mid 1990s onward); relatively autonomous local government, particularly in property regulation; the tradition of business involvement in local politics, often directly as politicians; the importance of private financial contributions to local election campaigns, especially for mayors; and the low (or nonexistent) profile of social democratic or labor parties in most cities. Thus, to the extent that such features are accepted for American cities, the notion of growth machines is more plausible, at least empirically (if not as a thesis). At the same time, they increasingly have been seen as key limitations in using the language of growth machines in international studies.

Scale and Local Dependency

Although the central deployment of place might seem the most obvious concept in the growth machine thesis where geographers would focus critique, for the most part, attention has instead been directed toward the thesis' conceptualizations of local dependency and scale. Outlined earlier were critiques of both the focus on land property and human agency, and how both are sometimes argued as overly empirical and thus undertheorized. Refracted through a geographical lens, these foci have also been critiqued as painting a simplistic and *a priori* picture of local dependence, whereby a more or less specific range of actors and institutions – basically those associated with the rentier class and local government – are assumed to be dependent on a locality and its growth. For some, this has been seen as lacking a sophisticated theorization of the contingent and relative spatial mobilities and immobilities of different actors and institutions. While relative immobility might sometimes form the basis for coalitions around specifically local political concerns, such collective action may unfold around a range of issues, and is unlikely to always be directed to urban growth.

If some have suggested that the growth machine thesis poorly theorizes the crossroads between local dependency and political coalitions, others have suggested that it also neglects the multiscalar nature of urban governance, and particularly the effects of heteronomous forces, actors, or organizations. Critiques here have ranged from simply pointing out that regional, national, and sometimes supranational activities all have a hand in urban politics to more theoretically dense arguments about the contingent and strategic deployment of scalar constructions in political projects of various origins. This wider body of writing would suggest that growth machines are about a politics at certain scales, when by contrast what is most important is the politics of scale.

Arguments around local dependency and scale, rather than place, are less surprising when one considers that it has primarily been urban and political geographers problematizing the growth machine thesis. Above all else, such writers have critiqued the assertion that the thesis frames a specifically urban politics and sociology. As many have argued, simply focusing on those particular actors seen empirically as operating at the so called urban scale or level does not produce a convincing theory of how the spatialities of political action relate specifically to cities.

Urban Political Projects and Local Feeling

A final area of debate is perhaps the most long lasting (if, in some respects, latent) aspect of the growth machine thesis: that organized political projects for growth in some way manipulate or influence more widely held feelings of local attachment. It is important to underscore here that early writing around the growth machine thesis did not necessarily imply that the elite propagation of local feelings (or so called ideologies) is completely seamless or always effective vis à vis urban publics. Rather, elites are characterized as at best partially conscious that they are manipulating local pride to promote their particular agenda, and urban publics are seen as often skeptical of city boosterism. Nevertheless, the tendency of most subsequent studies to simplistically portray the connections between elite political projects and wider local feeling has led to two major criticisms.

First, it has been pointed out that there tends to be little distinction made between organized, strategic urban or territorial projects and the messier, more distributed sentiments, feelings, memories, and discourses related to a locality. This distinction is important since, while the former indicates points of consensus that may be relatively attainable for narrow elite groups, the latter concerns a much more complex and uncertain political field that is not so easily made coherent. Second, research into the connections between the political projects of growth coalitions, and the more widely held sentiments or feelings of urban publics, has been noted for its distinct one sidedness. Largely, a focus has been placed on elite strategies or representations, with an unstated assumption that such political projects are in some way effective or hegemonic in relation to urban public life. By directing most empirical attention to studies of elite behavior, discourses, or representations, the supposed subjectification of urban publics through such projects has more often been a matter of theorization, or even just speculation.

Most research in this area has been via studies of discourse and representation, in a somewhat belated and selective incorporation of early 1990s work in new cul tural geography. Although the use of these methods seems to have diminished more recently, similar ques tions about the link between elite power and wider urban publics continue to be broached, if now through the lens of various new approaches or theories, such as those re lated to governmentality, social practices, and even emotion and effect. In pointing early on to such questions in a unique – if limited – way, the growth machine thesis can be viewed (at least) as a harbinger of still enduring concerns for contemporary studies of urban politics.

See also: City Marketing; Cox, K.; Governance, Urban; Land Rent Theory; Local Economic Development, Politics of; Neighborhood Change; Neighborhoods and Community; Place, Politics of; Urban Policy; Urban Regimes.

Further Reading

Clark, G. L. (1990). A realist project: Urban fortunes: The political economy of place. *Urban Geography* 11, 194 199.
Clarke, S. E. (1990). ‘Precious’ place: The local growth machine in an era of global restructuring. *Urban Geography* 11, 185 193.
Cochrane, A., Peck, J. and Tickell, A. (1996). Manchester plays games: Exploring the local politics of globalisation. *Urban Studies* 33, 1319 1336.
Cox, K. R. and Mair, A. (1989). Urban growth machines and the politics of local economic development. *International Journal of Urban and Regional Research* 13, 137 146.
Harding, A. (1995). Elite theory and growth machines. In Stoker, G. & Judge, D. (eds.) *Theories of Urban Politics*, pp 25 53. London: Sage.
Harding, A. (1999). North American urban political economy, urban theory and British research. *British Journal of Political Science* 29, 673 698.
Jonas, A. E. G. and Wilson, D. (eds.) (1999). *The Urban Growth Machine: Critical Perspectives Two Decades Later*. Albany, NY: State University of New York Press.
Lake, R. W. (1990). Urban fortunes: The political economy of place: A commentary. *Urban Geography* 11, 179 184.
Logan, J. R. and Molotch, H. (1987). *Urban Fortunes: The Political Economy of Place*. Berkeley, CA: University of California Press.
MacLeod, G. and Goodwin, M. (1999). Space, scale and state strategy: Rethinking urban and regional governance. *Progress in Human Geography* 23, 503 527.
Molotch, H. (1976). The city as a growth machine: Toward a political economy of place. *American Journal of Sociology* 82, 309 332.
Molotch, H. (1993). The political economy of growth machines. *Journal of Urban Affairs* 15, 29 53.
Molotch, H. and Vicari, S. (1988). Three ways to build: The development process in the US, Japan and Italy. *Urban Affairs Quarterly* 24, 127 143.
Swanstrom, T. (1985). *The Crisis of Growth Politics: Cleveland, Kucinich, and the Challenge of Urban Populism*. Philadelphia, PA: Temple University Press.
Ward, K. G. (2000). From rentiers to rantiers: ‘Active entrepreneurs’, ‘structural speculators’ and the politics of marketing the city. *Urban Studies* 37, 1093 1107.

Relevant Websites

http://www.citymayors.com/
City Mayors Network.
http://www.commurb.org/
Urban And Community Sociology Section, American Sociological Association.
http://www.uli.org/
Urban Land Institute.
http://www.hud.gov/
US Department of Housing and Urban Development.
http://www.chamberfind.com/
World Wide Chamber of Commerce Guide.

Urban Habitats/Nature

J. Byrne, Griffith University, Gold Coast, QLD, Australia
J. Wolch, University of Southern California, Los Angeles, CA, USA

Glossary

Anthropocentrism Human-centered value system, humans have prime importance.
Biocentrism Value perspective that holds all life as sacrosanct.
Biosphere/Ecosphere The sum total of all life on Earth.
Brownfields Former sites of industrial activity, oftentimes contaminated.
Ecological Restoration The repair of degraded habitats to restore ecological functionality.
Environmental Justice Inequitable exposure to environmental harm and/or inequitable access to environmental benefits.
Habitat Fragmentation The breakup of habitats through land-use change.
Political Ecology Perspective that links environmental degradation with economic inequity, social marginality, and vulnerability.
Trans-Species Urban Theory Perspective incorporating animals and plants in dynamics of urbanization.
Wildlife Corridor Pathways allowing animals to move between segregated habitat patches.
Zoöpolis A city of people and animals coexisting in urban life spaces to their mutual benefit.

Introduction

Most humans inhabit cities. Cities appropriate, metabolize, and transform 'natural' environments, reconstituting them through the application of technology and labor. As socionatures – simultaneously natural and cultural spaces – cities have variously been termed chimeras, cyborgs, or hybrids by contemporary urban theorists. But this theoretical view has not always prevailed. Until recently, cities were considered the habitat of humanity, the apotheosis of civilization, and inherently 'un natural'. Modernist thought denied the organic origins, functions, and corporeal impermanence of cities, regarding them instead as bastions of culture, reason, and progress, predicated on the conquest of nature and an ontological dualism separating culture and nature. This anthropocentric, utilitarian perspective has profoundly affected both human and nonhuman lives within cities.

Cities as 'Nature'

Upon closer inspection, carpets, gasoline, coffee makers, and freeway interchanges are 'wholly natural', derived as they are from the tissue of once living organisms, and from the raw materials of life. Moreover, the lengths to which we go in keeping rats out of our garbage, bacteria and mould out of our bathrooms and kitchens, and weeds out of our gardens belie the idea that cities are artificial. Our own bodies host a multitude of microbes, fungi, lice, and even occasionally fleas and ticks. Thus there is nothing inherently unnatural about cities. In cities we transform water, minerals, nutrients, and organic material into information, toasters, streetlights, sofas, and automobiles, and we re route complex biogeochemical cycles to suit our needs and wants. In cities air is respired, wastes are generated, transported, and excreted; heat is vented, water is consumed, and all manner of materials are moved through vast interconnected biophysical networks.

Many of these networks now span a range of scales and not all of them are human centered. For instance, global flyways link countless cities, incorporating them into the genetic circuits of migratory birds and avian influenza. International travel and commodity exchange also connect urban centers into a matrix of genetic flows. Potentially pandemic diseases like SARS can cross the globe in days, from live animal markets in Vietnam or China to homes and boardrooms of North America, Europe, and Australia. The ballast tanks and containers of ocean going vessels carry exotic organisms, translocating them to new habitats where they may outcompete endemic species. Global cities, their hinterlands, and the human and nonhuman bodies within them have thus proven to be remarkably porous.

Despite their local particularities, or perhaps because of them, cities of both the industrialized north and developing south have become an interconnected global urban socionatural system of genetic, chemical, material, information, and energy exchange.

Transformation of Nature in the City

This is not to say that cities do not alter so called 'natural' environments – for they do, usually in profound ways. Urbanization precipitates countless biological and physical changes. Most cities are characterized by high biodiversity due to their patchwork of land use patterns, for example, although this same mosaic creates other effects, isolating endemic (locally native) plants and animals within relict habitats, leading to potential extirpation. Animals moving from one habitat patch to another are also vulnerable to vehicle collisions, attack, and disease. And habitat patches are at risk from fire, weeds, air and water pollution, herbicides and pesticides, and invasion by exotic species. Within habitat patches, plants and animals are vulnerable to edge effects, increased predation, increased competition with other species, increased ambient noise levels, and increased artificial night lighting. Recreation in habitat fragments can also lead to serious ecological disturbance including soil compaction, reduced infiltration, increased erosion, and changes in soil moisture, temperature, and fertility.

Urbanization radically modifies hydrological and biogeochemical cycles, creating fluxes of temperature, nutrients, runoff, and soil acidity. Cities also face environmental extremes (i.e., floods and drought), as development reconfigures the hydrology of wetlands, streams, and rivers through aquifer depletion, stream channelization and diversion, increased surface runoff, and increased sedimentation. Water bodies experience increases in turbidity, temperature, nutrients (e.g., phosphorus and nitrogen), salinity, and changes to water chemistry, and can also display lower oxygen levels due to eutrophication (nutrient enrichment), and suffer from increased contamination from heavy metals, pesticides, sewage, hydrocarbons, and endocrine disruptors. These modified regimes can cause extirpation of aquatic flora and fauna and modify the behavior, phenology, and phenotypes of surviving organisms.

When terrestrial areas are cleared of endemic vegetation, many animals lose their nesting, denning, and foraging sites. Clearing woodlands can leave isolated trees prone to wind throw, and can raise groundwater levels resulting in salinization. Human modified terrestrial environments may exhibit changed fire regimes, since fire suppression strategies can lead to decreased fire frequency and also to more intense fires when areas with high fuel loads eventually burn. This reduces plant and animal biodiversity and encourages the proliferation of weeds. Heat island effects may alter the spawning, nesting, and hatching times of a variety of insects, fish, amphibians, and vertebrate animals, and flowering and leafing times of some plants. Changes to biogeochemical cycles can render urban plants and animals susceptible to disease, and some migratory patterns may be severely disrupted (e.g., channelizing urban streams can disrupt the migration of lampreys or trout). In addition, air pollution can weaken trees, leaving them vulnerable to parasites, increasing wildfires and displacing animals dependent upon these trees for food and shelter (e.g., spotted owls, bark beetles, and Californian coniferous forests).

Coastlines are increasingly urbanized, and changes to coastal process are equally profound. Sediment transport is interrupted through groynes, jetties, and sand pumping; coastal wetlands are drained and filled, and dunes are denuded and developed. Drawdown from coastal aquifers may cause saltwater intrusion, and septic systems can leach nutrients into both groundwater and near shore coastal waters. Some of these changes may exacerbate the effects of seasonal storms leaving urban residents, plants, and animals exposed to flooding, landslides, and coastal erosion.

Urban Habitats

But cities are not 'dead zones'. Nature suffuses the city. In Los Angeles, flocks of feral parrots roam the suburbs, and coyotes prey on pets even in downtown Beverly Hills. New York's falcons dine on pigeons; Londoners share their city with sparrows, foxes, deer, and the occasional badger; while on Australia's Gold Coast residents share their roof spaces with pythons, possums, and fruit bats. Human–animal interactions in the city are common. But the ability of urban wildlife to prosper depends upon the time, place, and scale of human–animal interactions. Issues of seasonality – such as breeding cycles – and the duration, intensity, and predictability of interactions are important. Outcomes are a function of the type of animal involved, and the animal's health, body size, behavioral adaptability, social group size, age, and sex.

Cities provide diverse habitats. Opportunistic species in particular usually prosper. Our houses, backyards, parks, and landfills create numerous opportunities for plants and animals to flourish. Although many opportunistic species face eradication efforts, urban environments may actually provide a greater diversity of opportunities for such species than wildlands. Such opportunities may be supplemented by habitat creation by community restoration groups that fence off open space patches for protection and conduct regular weeding and revegetation. But even some ornamental plant species found in urban and suburban gardens can be beneficial to many bird species and supplementary feeding of birds and other wildlife may have short term benefits including enhanced reproductive success.

Some animals, like mosquitoes, rats, and cockroaches are our 'familiars' – commensal species that have travelled with us over the eons. Commonly reviled and persecuted, they nonetheless retain their hold on urban space. We also categorize some plants as 'weeds' – targeting them for eradication – but they too hold on tenaciously, undermining our efforts to control nature in the city, even flourishing because of air pollution. Others animals have the misfortune of being our prey. We crowd chickens, pigs, cattle, and fish into industrial farms;

ultimately they reappear on urban dinner plates or as commodities like soap, paintbrushes, and perfume. But for some animals and plants, urban life is good. Cats, dogs, goldfish, and philodendrons share our living rooms, bedrooms, and backyards.

Unfortunately cities appear to have significant negative impacts on urban wildlife through exploitation, disturbance, habitat modification, and pollution. Exploitation results in death from hunting, trapping, fishing, or collection. Disturbance may be either unintentional (accidentally scaring a nesting bird) or intentional (frightening a deer to get a good photograph). Habitat modification typically results from vegetation clearing or damage, the introduction of invasive plant species, or the release of predators or competitors, and pollution.

Other negative impacts include electrocution from overhead powerlines, poisoning from insecticides, avicides, or rodenticides, collision with vehicles, and collision with glass windows. Disease is also a hazard; animals like pigeons can spread an array of avian diseases not normally found in wild species. Some animals may also modify their behavior in cities, for example, coyotes and bobcats may become more nocturnal in the presence of humans – a strategy to avoid contact.

Urban habitats and their resident wildlife also impact people – in both positive and negative ways. Negative impacts include animal attacks on individuals and/or pets (e.g., nesting magpies swooping people or coyotes eating domestic cats); the spread of disease (e.g., Lyme disease); nuisance activities such as animals foraging through trash, defecation on cars, benches etc.; damage to houses caused by nesting; garden damage through insect attack, animal foraging, and/or trampling; and annoying vocalizations (e.g., North American mockingbird calls in the middle of the night). Urban trees can also cause property damage (e.g., sap on cars, buckled footpaths, severed powerlines). Urban trees may block out light, cause allergies, transmit disease, and attract nuisance wildlife.

But access to urban nature also confers benefits for urban residents. These are economic (such as tourism, increased property values, energy savings), recreational (e.g., bird watching), and ecological (e.g., seed dispersal, pollination, shade). A wide variety of beneficial ecosystem services include pollution abatement, noise attenuation, and carbon sequestration. Many species also have esthetic value (for those delighted by the chance to observe and/or photograph wildlife or trees), spiritual value, provide stress relief, and foster socialization. Educational benefits accrue to children who learn about nature close to home.

Research on Urban Nature

Diverse scientists have recently begun to investigate the complex ecological patterns and processes that occur in urban space. In Baltimore and Phoenix, for example, research funded by the National Science Foundation's Long Term Ecological Research (LTER) program is currently underway. These two urban LTER programs focus on urban ecosystems to better understand cities as disturbance regimes. They include humans and human activities as integral components, and investigate, for example, humans as geomorphic agents, nutrient flow through highly modified urban environments, interactions between exotic and endemic species in urban parklands and their influence on ecological succession, land use change and ecological processes, and the ecological consequences of socioeconomic perturbations.

Increasingly research on urban nature focuses on ecosystem services. The valuation of such 'nature's services' is part of a broader move to quantify how the environment creates or protects economic value, and thus warrants investments for conservation, preservation, and/or remediation. Parks, for example, can help regulate ambient temperatures, improve air quality, purify runoff, buffer ambient noise, provide habitat for pollinators, protect biodiversity, infiltrate stormwater, and help recharge aquifers. Each of these 'services' can either enhance values or avoid costs; for example, homes near parks or in proximity of green cover have enhanced market value, while energy costs for homes shaded by trees are lower.

There are still large gaps in our knowledge of urban ecosystems. Geographers have begun to problematize nature–culture and human–animals relations in urban environments. Cultural ecology, human ecology, political ecology, and environmental justice – to name a few perspectives – suggest that how we relate to nature shapes how we relate to each other; and that the 'nature of cities' and 'nature in the city' are complexly intertwined.

Urban Nature and Urban Theory

Over the past two decades there has been an efflorescence of writing about human–environment relations within the city from both inside and outside geography. Scholars from environmental history including William Cronon, Alexander Wilson, Jennifer Price, and others have challenged what we understand 'nature' to mean, and in so doing, have reformulated views of the city. Exploring changes in ideas of nature, they reveal how city building has generated elaborate new interrelationships between society and nature. Within geography, scholars like Yi Fu Tuan, Denis Cosgrove, and Val Plumwood have reconfigured understanding of concepts like landscape, patriarchy, and difference, and their relation to nature. And geographers have focused on transformations in scale – how, for example, hog farming in North

Carolina has been altered by industrialization of agri culture and the emergence of global agro industrial corporations. These are urban phenomenon, since com modity supply chains link these new forms of animal intensive industrial agriculture to vast urban markets.

New concerns have spurred renewed interest in ani mal geography. The rise of biotechnology and genetic engineering, concerns about animal rights, and scientific insights into animal behavior have led scholars to 're animate' cultural and urban geography. The new animal geography encompasses studies of how animals shape identity, the role of animals in place making, and the ontological and epistemological challenges that are pre sented by including animals within the domain of human ethical concern. These challenges include the ways we might have to re think cities if animals are granted not only moral considerability but subjectivity and agency.

Reanimating the City: Geography and Urban Nature

Underpinning the geography of urban nature are ques tions including 'what is nature?', 'how is it socially (re) produced?', 'how are ideas of nature used to legitimize particular lifestyles, norms, and power relationships?', and 'how does nature configure space and place?'. By problematizing the nature of animals in the city, animal geography scholars like Jody Emel, Chris Philo, Kay Anderson, and others have provided insights into the places where people and animals coexist or conflict, es pecially with regard to symbolic and material contesta tions about animal rights and animal subjectivity. Spaces like gardens, parks, homes, alleyways, and zoos have proved to be fruitful sites of inquiry. And the study of animals in urban geography has illuminated racism, sexism, and speciesism underlying geographic thought, potentially leading us to what Whatmore, Braun and others have termed a 'more than human' geography.

Building on ecofeminism, critical race, and post colonial theory, Jennifer Wolch, Sarah Whatmore and Kay Anderson have noted that many animal practices are simultaneously political practices, connected to power, identity, and control. Moreover, the human–animal div ide is historically, culturally, and geographically contin gent. For example urban zoos are not just stationary animal exhibits; they are central to the formation of local and even national identity. Certain animal practices may simultaneously legitimize and naturalize racial op pression in urban settings and some animals become coded with pejorative meanings, bound up in the politics of human identity. For instance, Chris Philo has shown how, in the nineteenth century, working class neighbor hoods near abattoirs became coded as 'unclean', 'un hygienic', or even 'wild' and 'savage' like the animals destined for slaughter, while Wolch and her colleagues have shown that certain immigrant groups in the U.S. have been maligned for animal practices that transgress white cultural norms, such as dog eating, stigmatizing them as 'other', 'beastly', and even 'inhuman'.

But human–animal relations in the city extend beyond the discursive and ideological to material concerns too. Animals have been instrumental to transportation, war fare, fashion, religion, entertainment, communication, and sustenance. Animal are transformed into food, clothing, cosmetics, medicines, biologicals, spare body parts, handbags, shoes, furniture, tools, and even musical instruments. Human–animal interactions in the city are thus characterized by antinomy and affection, utili tarianism, as well as commitments to the intrinsic value of nonhuman others. How to develop cities that are sensitive to and respectful of the needs and desires of animals is the challenge.

Nature, the Ecological City, and Sustainability

Much thinking about nature–society relations in cities is now linked to questions of sustainable development since contemporary cities enfold vast hinterlands with com modity chains spanning the globe. The expansive reach of cities – their 'ecological footprint' – rooted in urban consumption means cities are directly implicated in ecosystem degradation, species endangerment, climate change, ozone depletion, and fisheries collapses around the world.

Phil McManus, Keith Pezzoli, David Sattherwaite, and others have identified a sustainable city as one that meets current human and ecosystem needs without comprom ising the welfare of future generations of plants, animals, and people, in ways that are economically sufficient, so cially just, and respectful of environmental limits. Other commentators have suggested this might occur through the reintegration of cities with the ecosystems that sus tain them. Strategies to do so include reducing car de pendence, shifting to more compact, mixed use urban forms, and increasing rooftop gardens, parks, and nature preserves. Underpinning such suggestions is the need for us to modify our patterns of production and consumption.

Ecological modernization initiatives such as ecological restoration, bioremediation of contaminated sites, waste to energy programs, industrial ecology, and water con servation are beginning to help cities become more sus tainable by re routing the material and energy pathways within them. Such initiatives are indicative of efforts to 'green' cities. Some urban geographers have begun to map out a pertinent research agenda, focusing, for ex ample, on the impacts of climate change on cities, the impact of natural hazards on urban populations, and how urban harms and benefits – including access to nature and exposure to pollution – are distributed throughout the city.

Conclusion: Toward a 'More Than Human' Urban Geography?

Future research on urban nature needs to address an array of socioecological and sociospatial processes across broad temporal and spatial scales, and must focus on the transformative dimension of social action, politics, and decision making with respect to access to resources and management of pollution. Urban political ecology has begun to embrace this challenge, For example, the work of Paul Robbins on elk and lawn chemical ecologies, and Nick Heynen on urban forests and the political ecology of urban hunger hints at how our sites of research will have to expand if we are to transcend the anthropo centrism characteristic of urban geography.

Specifically, we will need to be able to characterize the city and its economy both as a constellation of socio political and economic activities and as a set of energy and materials flows; to understand how animals perceive the urban environment, cope with urban life, and shape our emotional geographies; to map the social, insti tutional, and physical challenges posed by ecological restoration or projected impacts of climate change; and to deconstruct industrial commodity chains that fuel ur banization and explore how they can be reforged to create a more sustainable future.

In practice, the challenge of genuinely recognizing and coexisting with urban nature – what has been termed zoöpolis – is large. New practices must include plants and animals within the circle of moral considerability – an ethics of caring and respect based on the recognition of mutual interdependence and kinship but also of dif ference. Changes in land use, building design, transpor tation systems, energy generation, urban forestry, wildlife corridors, watershed management, and waste systems will take us some of the way toward zoöpolis. But we will also need to better understand the ideologies of nature that permeate place making which have for so long allowed animals and plants to be simultaneously eradicated and mobilized to sell urban places for profit.

See also: Animal Geographies; Biodiversity; Conservation and Ecology; Gardens and Gardening; National Parks; Nature, History of; Nature, Social; Nature-Culture; Natures, Charismatic; Natures, Gendered; Plant Geographies; Political Ecology; Sustainability; Wilderness.

Further Reading

Anderson, K. (1997). A walk on the wild side: A critical geography of domestication. *Progress in Human Geography* 21, 463 485.

Braun, B. (2005). Environmental issues: Writing a more than human urban geography. *Progress in Human Geography* 29, 635 650.

Braun, B. (2007). Biopolitics and the molecularization of life. *Cultural Geographies* 14, 6 28.

Cronon, W. (1996). *Uncommon Ground: Rethinking the Human Place in Nature*. New York: W.W. Norton & Co.

Davis, M. (2002). *Dead Cities and Other Tales*. New York: The New Press.

Douglas, I. (1981). The city as an ecosystem. *Progress in Physical Geography* 5, 315 367.

Elder, G., Wolch, J. and Emel, J. (1998). Race, place and the bounds of humanity. *Society and Animals* 6, 1 20.

Gandy, M. (2003). *Concrete and Clay: Reworking Nature in New York City*. Cambridge, MA: MIT Press.

Gunn, S. and Owens, A. (2006). Nature, technology and the modern city: An introduction. *Cultural Geographies* 13, 491 496.

Heynen, N., Kaika, M. and Swyngedouw, E. (eds.) (2006). *In the Nature of Cities: Urban Political Ecology and the Politics of Urban Metabolism*. New York: Routledge.

Isenberg, A. C. (ed.) (2006). *The Nature of Cities: Culture, Landscape and Urban Space*. Rochester, NY: University of Rochester Press.

McHarg, I. (1992). *Design with Nature* 25th anniversary edn. New York: John Wiley and Sons.

Philo, C. and Wolch, J. (1998). Through the geographical looking glass: Space, place and society animal relations. *Society and Animals* 6, 103 119.

Platt, R., Rowntree, R. A. and Muick, P. C. (eds.) (1994). *The Ecological City: Preserving and Restoring Urban Biodiversity.* Amherst: University of Massachusetts Press.

Whatmore, S. and Thorne, L. (1998). Wild(er)ness: Reconfiguring the geographies of wildlife. *Transactions of the Institute of British Geographers* (NS) 23, 435 454.

Wilby, R. L. and Perry, G. L. W. (2006). Climate change, biodiversity and the urban environment: A critical review based on London, UK. *Progress in Physical Geography* 30, 73 98.

Wolch, J. (1996). Zoopolis. *Capitalism Nature Socialism* 7, 21 48.

Wolch, J. (2002). Anima urbis. *Progress in Human Geography* 26, 721 742.

Wolch, J., Pincetl, S. and Pulido, L. (2002). Urban nature and the nature of urbanism. In Dear, M. (ed.) *From Chicago to L.A.: Making Sense of Urban Theory.* Thousand Oaks: Sage Publications pages 367 402.

Relevant Websites

http://www.beslter.org
Baltimore Ecosystem Study.

http://www.lternet.edu
Central Arizona Phoenix Urban LTER CAP.

Urban Modeling

M. Batty, University College London, London, UK

Glossary

Agent-Based Models (ABMs) A class of models developed since the 1980s which are based on representing objects and populations at an elemental or individualistic level which reflects behaviors of those objects through space and time. These models operate from the bottom up and sometimes generate emergent spatial and temporal patterns at more aggregate levels.

Calibration The process of dimensioning a model in terms of finding a set of parameter values that enable the model to reproduce characteristics of the data in the most appropriate way. Calibration is not the same as validation which seeks to optimize a model's goodness of fit to data, but often, these processes are equivalent.

Cellular Automata (CA) A class of spatially disaggregate models, often pictured as being formed on a two-dimensional lattice of cells, where each cell represents a land use and embodying processes of change in the cellular state are determined in the local neighborhood of any and every cell. Such models can be seen as simplifications of agent-based models where the focus is on emergent spatial patterns through time.

Complex Systems Systems that show surprising and unanticipated or 'emergent' behaviors as shown in patterns that arise at the aggregate level from the operation of system processes at the micro or agent level. Such systems are intrinsically unpredictable in an overall sense but can be fashioned in such a way that makes knowledge of them useful and certain. Cities are the archetypical example, but so too is the economy.

Discrete Choice A development of computable microeconomic theory in which individuals maximize a utility, subject to constraints on their choices which can be tailored to reflect how decisions are made in complicated situations. Such models have been applied extensively in transportation modeling and have strong links to more aggregate maximization models as derived from spatial interaction and social physics.

Land-Use Transport (LUT) Models A class of models that focus primarily on the way populations and employments locate in urban space consistent with the spatial interactions between different locations of these activities. These models usually simulate the city at a cross section in time and as such, bundle urban dynamics into equilibrium behaviors.

Social Physics The application of ideas from classical 'Newtonian' physics to social systems usually in the form of analogies with Newton's laws of motion as reflected in the concepts of potential energy and gravitational force. This lies at the heart of spatial interaction modeling but more recently such physics has been extended to embrace notions of complexity as reflected in scaling, self-organization, and the dynamics of far-from-equilibrium systems.

Spatial Interaction Movements of goods, people, and information between different spatial locations, often referred to as origins and destinations, theorized and simulated using analogies with gravitational laws in physics. Such models form the basis of standard methods for describing and modeling interactions ranging from trip making in cities to long-range migration between cities.

Urban Dynamics Representations of changes in urban spatial structure through time which embody a myriad of processes at work in cities on different, but often interlocking, time scales ranging from life cycle effects in buildings and populations to movements over space and time as reflected in spatial interactions.

Urban Economics The development of microeconomic theory at the urban scale, following the tradition of the von Thunen model in which location and land rent are hypothesized as a function of distance or travel costs from some market center. The development of these ideas in the 1960s led to this branch of economics being called the New Urban Economics and more recently it has been extended using growth and trade theory.

Urban Modeling The process of identifying appropriate theory, translating this into a mathematical or formal model, developing relevant computer programs, and then confronting the model with data so that it might be calibrated, validated, and verified prior to its use in prediction.

Urban Models Representations of functions and processes which generate urban spatial structure in terms of land use, population, employment, and transportation, usually embodied in computer programs that enable location theories to be tested against data and predictions of future locational patterns to be generated.

Online Resources

There are many online resources that can be used to extend the material of this article. Land use transporta tion (LUT) models tend to be large and uniquely tuned to particular applications and although software might be downloaded, there are few, if any, online demonstrations. However the state of the art is represented by the UrbanSim class of models which are now being fashioned into the Open Platform for Urban Simulation (OPUS) which is located at http://www.urbansim.org/. Cellular automata (CA) models are more manageable and soft ware is available to download to generate simple ex amples. The Dynamic Urban Evolutionary Model (DUEM) developed in London and Michigan can be accessed at http://www.casa.ucl.ac.uk/software/duem.asp and used to develop simple demonstrations of cellular growth. More complex models developed for simulating urban growth in North America have been developed by United States Geological Survey (USGS) as part of their project Gigalopolis and these SLEUTH models can be downloaded and viewed at http://www.ncgia.ucsb.edu/projects/gig/. Extensive and up to date information about agent based models (ABMs) is available at http://gisagents.blogspot.com/. Simple demonstrations of such models are available at http://www.genesis.ucl.ac.uk/ and at http://www.complexcity.info/.

Defining Models

Models are simplifications of reality – theoretical ab stractions that represent systems in such a way that es sential features crucial to the theory and its application are identified and highlighted. In this role, models act as a vehicle to enable experimentation with theory in a pre dictive sense, and to enhance understanding which may be prior to predictions of situations as yet unrealized, for example, in the future. This role of experimentation is usually through an environment somewhat different from the laboratory sciences where manipulation of the phe nomena in question is direct and controlled. Models in the context here, are invariably implemented in com puter environments which act as a surrogate for the la boratory where this use of the term 'model' has gained considerable currency over the last 50 years with the rise of computing in the social sciences. Urban models are thus essentially computer simulations of the way cities function which translate theory into a form that is tes table and applicable without experimentation on the real thing. Computers act as the laboratory for experimen tation on phenomena which is represented digitally, with its manipulation being virtual. Urban modeling, the subject of this article, is the activity of defining, building, and applying such digital models for specific purposes which, traditionally, have been in physical planning. These applications increasingly extend to other social and human geographies built around location theory and spatial analysis in commercial as well as public decision contexts.

In this article, we first define an urban model, and then chart the process of building such a model from the as sembly of data through calibration to validation and prediction and thence into forecasting, design, and planning. Over the last 50 years since computer models were first developed in the urban domain, several distinct types have emerged, and we will classify these in such a way that the various techniques and modeling styles as sociated with them are clarified. In particular, we will identify distinct generations of model, beginning with static and aggregate LUT models which had their heyday in the 1960s, dynamic variants of these same models, more recent bottom up styles of model which are called agent based and are dynamic and disaggregate, and urban models which focus on the temporal dynamics of ag gregate populations. All these types still exist today as the field continues to develop and proliferate.

Urban Theory, Models, and the Scientific Method

Science begins with theory that is translated into a form that enables it to be compared with reality through the process of making predictions. If the predictions are good, the theory has withstood the test and confidence is gained in its relevance. As our abilities to model dif ferent and more richer realities are enhanced, it becomes increasingly unlikely that our theory can be tested in the controlled conditions of the laboratory, and this is where the computer plays an essential role. Theories are thus translated into a form that enables them to be repre sented as mathematical or logical models, with the computer acting as the laboratory in which simulation of the reality takes place.

In terms of cities, the kinds of urban theory that are basic to the development of computer models are those that are traditionally called location theories: theories that propose mechanisms that enable industries, services, and households to locate in space within economic constraints of income and profitability. In turn, these economies are conditioned primarily by distance (often as a proxy for travel cost) between land uses associated with these activities which depend upon a range of market conditions essentially underpinned by trade. Thus, distance and movement are central to such theory which by the mid twentieth century had more or less resolved itself into three styles: an aggregate theory in which space was differentiated according to principles of social physics which loosely revolved around energy and potential, an aggregate theory of macroeconomic rela tionships between various types of production and con sumption, and a more disaggregate theory based on the microeconomics of competing land uses in which trans port cost, land rent, and spatial profit dominated location.

Alonso, in 1964, was perhaps the first to fashion a formal statement of urban economic theory which came to be called the 'new urban economics', although it was Isard *et al.* who presented a catalog of more practical methods based on spatial interaction ideas from social physics and macroeconomic models such as input–output analysis. The first generation of urban models were based on applying such techniques by treating the urban system as a static entity whose land uses and activities were to be simulated at a cross section in time and whose dynamics were largely regarded as self equilibrating. Thus, the early models were in the tradition of comparative statics. Movement at a cross section in time either as long term migration and/or as routine, diurnal transportation through trip making was central to these models de veloped in analogy to gravitational and potential theory where flows were simulated in inverse proportion to travel cost or distance between 'origin' and 'destination' places. Much of the theory that passed muster in those early days is still with us although the influence of movement has lessened; other factors, particularly eco nomic, governing housing choice have come onto the agenda, and the strong macroeconomic focus through input–output and economic base theory has weakened.

The link from theory to model as a vehicle to test hypotheses has also fractured somewhat as traditional models have loosened their link to theory. In essence, theories of the city system were found wanting in that they did not reflect the diversity and heterogeneity that was very evident in modern cities, nor did they reflect the comparative volatility of urban dynamics suggesting that this dynamics could be absorbed within a wider equi librium. During the 1970s and 1980s, the aggregate static approach to theory and modeling began to switch around to more bottom up decentralized dynamics. As these styles of modeling gained ground, their data require ments exploded to the point where it became impossible to even calibrate, never mind validate, such models in their entirety. This of course reflected the move to relativism, postmodernism, and a style of social science that was diametrically opposed to the traditions of urban economics and social physics. No longer are models vehicles for testing hypotheses. Urban models are more likely to be frameworks for assembling relevant infor mation, frameworks for formal and informal dialogs where they are essential tools in much more consensual and participative processes of decision support.

Model Principles, Types, and Styles

There are a limited but fundamental set of principles for building computer models which are generally accepted by those involved in the field. In essence, modeling is a process of simplification where paraphrasing Einstein's famous dictum, models "…should be as simple as pos sible but not simpler." This reflects the difficulties in abstraction where there is always a tension between how much to leave in and how much to leave out of any theory and its model. We will divide this discussion into approaches to abstraction and implementation, dealing with these in turn.

Scientific Abstractions

The first issue involves simplifying the spatial system in terms of articulating spatial structure at a cross section in time – in 'equilibrium' – or as a dynamic sequence of change. From one perspective, cities can be seen as largely unchanging in terms of their land uses and transport structures with marginal change far less im portant to that which exists in totality. This might almost be a matter of taste for there is another view that sees structure as being continually transformed even though it remains in place and, in this sense, any model should be dynamic reflecting such processes of ceaseless transfor mation. 'Statics' versus 'dynamics' is thus a central and often contentious issue.

This issue of time also relates to 'aggregation' and 'scale'. In general, the finer the spatial scale and the shorter the time period, the greater the dynamic in that as we aggregate activities from their elemental form, we tend to average, thus reducing 'heterogeneity'. Again, the degree to which the model should reflect heterogeneous activity depends on what is being modeled at what scale with this trade off part of the process of simplification. In some models, there are consistent schemes for illustrating what actually happens when data and model elements are aggregated but, in general, the process is *ad hoc* and dis tinct differences of theory and style exist between the aggregate and disaggregate, between 'macro' and 'micro': models based on economic theory are the classic example.

'Representation' of the key elements of urban struc ture whether they be as individuals comprising various populations or aggregates thereof involves key problems of definition and classification. In the models considered here, spatial representation is critical for the scale in terms of size of the areal unit as well as its configuration directly relating to the way the system's elements are defined. Individuals exist at point locations but many models simplify these by aggregating basic units into groups which are associated with distinct areal units, often called zones or regions. This issue involves the previous two: temporal dynamics and heterogeneity, and clearly the greater the degree of aggregation, the simpler the model. When urban models were first developed in the 1960s, almost all were highly aggregated in terms of their representation, whereas now a new class of indi vidual or ABMs have appeared which seek to represent the urban system in much more disaggregate and heterogeneous terms. In terms of the actual spatial units defined, there is a recurrent argument as to whether densities or counts should be the way of representing elements – densities implying a regular spatial system of 'cells' on a lattice in contrast to variable vector reporting units usually based on 'administrative units' or 'zones'. All these issues involve trade offs involving scale which in turn are determined by more pragmatic concerns, such as available resources of data and computation.

Mathematical Implementations

To introduce the various styles of mathematics used to construct urban models, we define population as P and employment as E, location by subscripts i and/or j, and time as t. We will present six kinds of model which reflect the main types used to date: urban economic base models, social physics models that distribute activity according to gravitational hypotheses and are widely referred to as spatial interaction models, and rent and population density models based on microeconomic theory which trades off the demand for space against cost or distance traveled usually to workplaces. These are all static spatial models which are in stark contrast to models that deal with time. Dynamic models can be at an aggregate level, such as those based on population dynamics, or more micro in terms of mobility, and there are various models that deal with long term dynamics that reflect discontinuous change in the urban system. The last class of models deals with the supply of land and reflects developer dynamics and decision making. We will now sketch these types to give some sense of how urban models are constructed.

Urban economic base models divide aggregate employment E into basic employment B which drives demand for nonbasic or service employment S through the multiplier effect where $E = B + S$. We relate demand for services to total employment as $S = \beta E$ from which it is easy to demonstrate that $E = B(1-\beta)^{-1}$, where $(1-\beta)^{-1}$ is the multiplier. If we then consider that population can be generated by applying an activity rate α to employment as $P = \alpha E$, we have the rudiments of a generative sequence that has been widely used in input–output modeling on the one hand and in cross sectional urban modeling on the other. If we now consider that employment and population are related spatially through movements or interactions called T_{ij} between work i and home j, we can articulate these as

$$T_{ij} = E_i \frac{R_j d_{ij}^{-\lambda}}{\sum_k R_k d_{ik}^{-\lambda}} \text{ where } P_j = \alpha \sum_i T_{ij} \qquad [1]$$

where d_{ij} is the distance or travel cost from zone i to j, λ is a friction of distance parameter, and R_j is some measure of attraction at residential location j. If we were also to relate the location of services to population through another gravitational model of the same form as eqn [1], we would have a scheme tying population and employment together through the economic base relation at the spatially aggregate level and through gravitational hypotheses at the spatially disaggregate level. In fact, this is the model first developed by Lowry, in 1964, which is still the most widely applied of all operational urban models and which has been elaborated in various ways, particularly in relation to the transportation sector.

This kind of modeling does not take account of the supply of land or other infrastructures and thus there is no market clearing. Consequently, prices for land are not determined. It is quite easy, however, to embed theories of the urban economy into these structures using theories of land rent first extended to the urban housing market by Alonso in 1964. This clearly demonstrates that rent ρ_j at location j is an inverse function of distance (or travel cost) from some central location $i = 0$ which can be simulated as $\rho_j = K d_{0j}^{\gamma}$ where K and γ are parameters of the function. This theoretical result emerges from utility maximization subject to constraints on travel costs and other goods which are realized through an income constraint set for a typical individual. The focus on budgets enables the model to be integrated into gravitational models through various kinds of utility/entropy maximization associated with constraints on travel and housing costs.

Dynamic models have usually focused on longer term temporal dynamics which involve demographic change and mobility through migration. Standard population accounting methods composed of births, deaths, and migration components are typically used to forecast future aggregate change in the urban system with spatial disaggregation often accomplished using the spatial model types noted above. Constrained population growth reflecting both exponential change and capacity which, in turn, reflect densities and congestion are simulated using various kinds of logistic growth. For example, aggregate population change from time $t-1$ to t defined as $\Delta P(t)$ can be modeled as

$$\Delta P(t) = \eta P(t)[P_{\max} - P(t)] \qquad [2]$$

where η is a composite growth/change rate and $P_{\max}$ is the maximum population that the system can take. If we examine eqn [2], when population $P(t)$ is small, then the term $[P_{\max} - P(t)]$ makes little difference and the system grows exponentially at an increasing rate with positive feedback. When population $P(t)$ is large relative to capacity, the growth is dampened by negative feedback, with total population increasing at a decreasing rate as capacity is reached when the change is zero. This is classic logistic growth that appears to occur in constrained systems which mirror human populations in contrast to Malthusian exponential growth. Extended urban models of this kind have been applied to city systems beginning with Forrester, in 1969, and have been developed to deal with more complicated dynamics where singularities and catastrophes can occur as distinct breaks in continuous growth. Wider systems of coupled equations extend the nonlinearities in such systems enabling cycling of various kinds to take place as well as bifurcations that generate novelty and surprise in growth and change. Models of this kind have been developed by Allen, in 1997, and Wilson, in 2000, among others and map surprisingly well onto recent developments in complexity theory where systems with multiple positive and negative feedbacks generate emergent structure.

There are many mathematical developments but we have sketched those of most importance with the exception of a rather different style of modeling that has recently become popular. More aggregate modeling has slowly given way to models which are articulated from the bottom up, rather than the top down, reflecting the increasingly popular paradigm of decentralization which appears to offer much better explanations of the way human systems function. This also reflects the move to dynamics from statics and to process rather than product. Models in which agents or actors (individuals or groups) are central, are now being widely developed. These models focus on various dynamics in the urban system embodying fast to slow processes, from local movement, such as travel, to migration over different time scales through to changes in individual life styles and cycles in the built form itself. These models may include more classical simulation techniques, such as those already sketched, but they are usually focused on decision rules in which agents behave in response to their environment and in response to each other. In this sense, the models simulate emergent phenomena and are capable, at least in principle, of embodying novelty, surprise, and innovation in the system.

Typical urban models in this style have so far focused on rather simple mathematics of bottom up processes, such as CA. The system is partitioned into physical zones or cells which are small enough to reflect individual characteristics, such as distinct land uses. Cells change their state – their land use – dependent upon what happens in their neighborhood which essentially embodies the decision rules as to what land uses are compatible with one another at various spatial distances between them. These CA models can be articulated as rules of the following kind:

$$\text{if } C_j = \text{empty, and if } N_j^\ell = C_k^\ell,\ \forall k \in \Omega_j, j \neq k, \\ \text{then } C_j^\ell = \text{developed } as\ \ell \qquad [3]$$

Essentially eqn [3] mirrors the decision process. A cell at location j is converted to development – to land use ℓ –; if there is a certain number # of cells N_j^ℓ with land use ℓ within a neighborhood of j called Ω_j. This is usually restricted to a few cells around the location in question. If the neighborhood is restricted in this way, then local rules lead to global emergence as in classic CA models, such as the Game of Life. Imagine this kind of system being extended to all zones with many different extents of neighborhood and many land uses with multiple counting, majority, and physical decision rules and this gives an idea of the sort of urban CA models that have been developed. If one adds individual actors to this environment, then one has urban models which are agent based. Many CA but few ABMs of this kind have been developed to date. The advantage of this style of modeling is that it is dynamic and behavioral and can easily extend to both demand and supply sides of the development process. Its disadvantage is that it has enormous data requirements, and does not fit easily into more top down processes that drive the urban system. Because heterogeneity in agents is often introduced by local randomness, it does not generate the sorts of deterministic prediction that are usually needed in operational urban modeling.

The Model-Building Process

When computer models were first developed, they were seen as embodiments of theory translated into compatible media that enabled their testing against data. This process of testing in its purest sense is one where hypotheses as theories are matched against 'the facts' and as such, represents the process whereby theory is falsified or confirmed. This is the process of validation which is distinct from calibration and verification. That a model be validated – that it pass various tests which ensure that it replicates the phenomena of interest in an acceptable way – is necessary before such a model can be used for making predictions which are then acted on in some way, for example, by professional or political decision makers.

This process is dominated by deduction in which models enable outcomes in the form of patterns and processes to be derived from theory and tested under experimental conditions setup as computer laboratories. In fact, the scientific processes of developing theory which makes good predictions, even laws, revolves around a sequence of induction and deduction through historical time and in practice; it takes a stretch of imagination to see a particular model building process as mirroring this wider activity. In fact, most models are first calibrated or fine tuned to data which simply ensures that they meet certain dimensional constraints on the system of interest. Calibration provides values for unknown parameters. Sometimes these are also chosen to optimize some goodness of fit criterion, such as how close pre dictions are to observed data, and in this sense calibration merges into validation, with the line between the two being blurred. Even if calibration is considered a separate process, validation takes place immediately after, the difference being that parameter values are often chosen using criteria different from those used to validate the model.

This rudimentary sketch is one that most appli cations loosely ascribe to. It is paralleled by the process of verification which accords to testing the model for internal consistency and is often separate from testing how good the model's predictions are. A key issue is to ascertain under what conditions the model is judged acceptable for further use; in short, what the conditions are for accepting the model as confirming the various theoretical constructs which it embodies. A strict test is to validate the model in an empirical situation different from that for which its parameters are calibrated but this is rarely done literally. Other prior applications of the model which have withstood this process are often assumed to be sufficient. However, the degree and strength of confirmatory evidence is difficult to ascer tain. There are very few models that exist that can be tested on all their dimensions and the newer class of ABMs, which are much richer in terms of the hypoth eses they frame and the data required to calibrate and validate them, will never meet all basic scientific cri teria. This suggests that the rules for testing theory in this domain are changing and that the role of models is no longer entirely for this purpose: models are being developed as much for their exploratory and discursive value in a wider participatory process of developing robust but contingent knowledge than for their ability to generate good theory.

Urban Models and Their Applications

We have already introduced various modeling styles and, together with problem applications, these serve to define three main classes. These classes can be further sub divided into those which mainly have theoretical import and are not focused on empirical applications or pol icymaking in contrast to those that are, with the later subclass being considerably more pragmatic in structure than their theoretical equivalents.

Land-UseTransportation Models

This first class is built around the aggregate static models of economic and spatial interaction. Their the oretical pedigree is rooted largely in regional economics, location theory, and the new urban economics which represent the spatial equivalents of classical macro and microeconomics, and perhaps in social physics insofar as this can be said to embody social theory. The most co herent recent statement in this vein is based on appli cations of trade theory to the urban economy as reflected in the work of Fujita, Krugman, and Venables in 1999 but there is a long heritage of empirical models in the Lowry tradition of 1964 which continues to be built. These models now incorporate the four stage transpor tation modeling process of trip generation, distribution, modal split, and assignment explicitly and they are consistent with discrete choice methods based on utility maximizing specifically in their simulation of trip making.

These models have slowly been adapted to simulate dynamic change although they still tend to generate the entire activity pattern of the city. They remain parsi monious in that the assumption is that all the outcomes from the model can be obtained in terms of their good ness of fit. They have also become more disaggregate and now there are links to physical land use although they still remain at the level of activity allocation despite their nomenclature as LUTs. In short, this class of models is the most operational in that newer styles tend to be less comprehensive in their treatment of urban activities and transportation. Probably the most highly developed of these models currently is the UrbanSim model, although the MEPLAN, TRANUS, and IRPUD models best seen in the EU PROPOLIS (2004) project also represent the state of the art.

Urban Dynamics Models

Very few aggregate dynamically temporal urban models have been applied empirically. After Forrester's early attempt in 1969, the focus has been on theoretical de velopments of nonlinear growth and change which gen erate discontinuities through coupled nonlinearities, threshold effects, or random perturbations. Allen, in 1997, was the first to show how bifurcations could be generated through random perturbations of nonlinear structures at the micro level. Wilson, in 2000, did much the same except that he formally embedded spatial interaction models into the nonlinear dynamics of their independent variables. Various attempts have been made to link such models to ecological dynamics, building on Lotka–Volterra models of predator and prey, while at tempts have also been made at fusing the chaos of the logistic and other maps into spatial dynamics. But in one sense, all these forays into dynamics were rather aggre gate and consequently less applicable to the kinds of urban processes that are characteristic of cities. As such they were simply the path to more micromodels and have been eclipsed somewhat by spatial simulations of dy namic processes whose scale is at a much more indi vidualistic level, as embodied in ABM.

Cellular Automata (CA), Agent-Based Models (ABMs), and Microsimulation

The last class which is attracting the greatest attention at present involves models built around representing the actions and behavior of individual agents located in space. As might be expected, there have been various predecessor models in this vein; for example, Chapin and Weiss's work in 1968 in North Carolina and Ingram, Kain, and Ginn's housing market models in 1972 were constructed around individuals, market processes, and developer decisions which are the meat of the new generation of ABMs. In fact, the most popular type of model which has been applied empirically but has not been used much for policymaking is that based on CA where agents are, in fact, cells which change their land use cell state. There have been a substantial number of applications but few have been used to test urban policies in that transportation is handled rather crudely or even excluded in such models. The main focus, of course, is on urban growth, which in a contemporary manifestation is sprawled with these models tending to be indicative rather than predictive. The other issue in such models is that they are primarily physicalist in scope and as such, largely ignore features of the spatial economy such as house prices, wage rates, and transport costs.

There are some ABMs at the land use or activities level which enable predictions of future urban patterns but the main focus is at the very micro level where local movements in terms of traffic are being simulated. Several models that approach the agent based ideal originate from other areas. TRANSIMS is a hybrid in that its roots are in agent based simulation of vehicles but it has been scaled to embrace urban activities and even UrbanSim has been interpreted through the agent paradigm. A parallel but significant approach to individualistic modeling is based on microsimulation which essentially samples individual behavior from more aggregate distributions and constructs synthetic ABMs linked to spatial location.

Applications: Using Models in Urban Policymaking

To conclude, it is worth noting that urban models span both theory and practice and that their rationale depends on developing new theory as well as their use in policymaking and planning. This tends to confuse and conflate their development as the same class of model is often used for both. Traditional LUT models are the most applied and the most parsimonious and are still the dominant model used in practice. This is largely because they attempt to be comprehensive in simulating location and interaction, land use, and transport. But in this they sacrifice detail and process. They are largely nondynamic and in a world where change is to the fore, this limits their applicability. The rise of more microdynamic CA models and ABMs clearly attempts to meet this challenge and insofar as these models are being applied, they tend to concern more particular processes in cities, such as segregation, housing market policies, pedestrian movement, and related behaviors. What has happened, however, is a broadening of styles and model types. It is worth noting too that, at the edge of this domain, there are many computer methods, in GIS, for example, that in some circumstances might be considered as 'models' which blur into the model types reviewed here.

See also: Choice Modeling; Location Theory; Scientific Method; Simulation.

Further Reading

Allen, P. M. (1997). *Cities and Regions as Self Organizing Systems: Models of Complexity.* London: Taylor and Francis.
Alonso, W. (1964). *Location and Land Use*. Cambridge, MA: Harvard University Press.
Batty, M. (1976). *Urban Modelling: Algorithms, Calibrations, Predictions*. Cambridge, UK: Cambridge University Press.
Batty, M. (2005). *Cities and Complexity: Understanding Cities with Cellular Automata, Agent Based Models, and Fractals*. Cambridge, MA: MIT Press.
Batty, M. and Longley, P. (1994). *Fractal Cities, a Geometry of Form and Function*. San Diego, CA and London: Academic Press.
Ben Akiva, M. and Lerman, S. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.
Benenson, I. and Torrens, P. M. (2004). *Geosimulation: Automata Based Modelling of Urban Phenomena*. London: Wiley.
Castle, C. J. E. and Crooks, A. T. (2006). Principles and concepts of agent based modelling for developing geospatial simulations. *Working Paper 110*. London: University College London, Centre for Advanced Spatial Analysis.
Chapin, F. S. and Weiss, S. F. (1968). A probabilistic model for residential growth. *Transportation Research* 2, 375 390.
Clarke, G. (ed.) (1996). *Microsimulation for Urban and Regional Policy Analysis*. London: Pion Press.
Dendrinos, D. S. and Mullally, H. (1985). *Urban Evolution: Studies in the Mathematical Ecology of Cities*. Oxford: Oxford University Press.
Forrester, J. W. (1969). *Urban Dynamics*. Cambridge, MA: MIT Press.
Fujita, M., Krugman, P. and Venables, A. J. (1999). *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge, MA: MIT Press.
Harris, B. (1965). Urban development models: A new tool for planners. *Journal of the American Institute of Planners* 31, 90 95.
Ingram, G. K., Kain, J. F. and Ginn, J. R. (1972). *The Detroit Prototype of the NBER Urban Simulation Model*. New York: National Bureau of Economic Research, Columbia University Press.
Isard, W. *et al*. (1960). *Methods of Regional Analysis*. Cambridge, MA: MIT Press.
Lowry, I. S. (1964). *Model of Metropolis*. Santa Monica, CA: Memorandum RM 4035 RC, Rand Corporation.
Lowry, I. S. (1965). A short course in model design. *Journal of the American Planning Association* 31, 158 165.
Nagel, K., Beckman, R. J. and Barrett, C. L. (1999). *TRANSIMS for Urban Planning*. Los Alamos, NM: LA UR 984389, Los Alamos National Laboratory.
PROPOLIS (2004). *PROPOLIS (Policies and Research of Policies for Land Use and Transport for Increasing Urban Sustainability)*. Final Report for the Commission of the European Communities. Helsinki, Finland: LT Consultants.
Waddell, P. (2002). UrbanSim: Modelling urban development for land use, transportation and environmental planning. *Journal of the American Planning Association* 68, 297 314.

Wegener, M. (2005). Overview of land use transport models. In Henscher, D. A. & Button, K. (eds.) *Handbook in Transport, Vol. 5: Transport Geography and Spatial Systems*, pp 127 146. Kidlington, UK: Pergamon/Elsevier Science.

Wilson, A. G. (2000). *Complex Spatial Systems: The Modelling Foundations of Urban and Regional Analysis*. London: Pearson Education.

Relevant Websites

http://www.fractalcities.org/
Fractal Cities: A Geometry of Form and Function.

http://www.genesis.ucl.ac.uk
GENESIS Project.

http://gisagents.blogspot.com
GIS and Agent Based Modeling.

http://www.ncgia.ucsb.edu
Project Gigalopolis: Urban and Land Cover Modeling, NCGIA.

http://www.casa.ucl.ac.uk
The DUEM Cellular Automata; Cities and Complexity: Understanding Cities with Cellular Automata, Agent Based Models, and Fractals (Book), Centre for Advanced Spatial Analysis, University College London.

http://www.urbansim.org
UrbanSim.

Urban Morphologies, Historical

T. R. Slater, University of Birmingham, Birmingham, UK

Glossary

Agents of Change People and institutions that enable building or rebuilding to take place in towns. They include landowners, architects, developers, estate agents, town planners, local interest groups, and many others.

Building Bye-Laws: The regulations promulgated by local government for the construction of new buildings which concern matters such as foundations, drainage, light, the circulation of air, and the minimum dimensions of rooms. They often followed national model bye-laws from the mid-nineteenth century onward.

Building Cycle: The process which links the variability in the number of new buildings constructed each year to the general business or economic cycle of the region or country.

Burgage The urban strip plot held by a burgess in a medieval borough town and usually charged with a fixed annual rent to the communal tax of the town. It could contain buildings, stables, yard, gardens, and orchard and could be bounded by a fence, hedge, wall, or ditch.

Burgage Cycle The cyclic process of building development on the central plots in historic towns in response to changing socioeconomic demand between the late-medieval period and the present.

Fixation Line A linear feature such as a town wall, road, or railway line that prevents development beyond and marks the edge of the town for a considerable period. Land uses seeking a peripheral location tend to locate beside it and lead to the development of a fringe-belt.

Fringe-Belt A zone of often larger plots containing land uses that seek a peripheral location because they require large areas of cheap land, or they are noxious or noisy. During periods of slow development, such land uses concentrate at the edge of the city leading to a distinctive pattern of both land use and plot sizes.

Metrological Analysis Evaluation of data, such as the length, breadth, and area of urban plots, using mathematical measurement.

Orthogonal Planning The laying out of urban plans using a geometrical grid of streets at right angles to one another.

Plan-Unit An area of streets, plots, and buildings which is distinct from its neighbors and endowed with a measure of morphological homogeneity. It usually represents an area that was developed at a particular time.

Speculative Housing Houses built by a builder in the expectation that they will be sold immediately on completion for a profit.

Town-Plan Analysis The technique of evaluating the complexity of street systems, plot patterns, and building block plans to provide a coherent developmental sequence of the whole.

The academic study of urban forms (literally 'urban morphology') began in the later part of the nineteenth century in the universities of German speaking central Europe. There were two fields of study where research began to develop a body of ideas, models, and materials: first was cultural geography, where practitioners took their lead from colleagues researching in geomorphology, and second, there was the beginning of academic and practical town planning studies.

The originator of the first was Otto Schlüter who postulated a morphology of the cultural landscape as the object of research in cultural geography to match the geomorphological study of the physical landscape. Within the cultural landscape, Schlüter regarded urban landscapes as preeminent. As a consequence, morphological studies came to dominate central European urban geography in the first half of the twentieth century. Geomorphological practice suggested to cultural geographers that the careful cartographic plotting of urban data would lead to the classification of urban forms and then on to the explanation of those forms (and it is important to note that explanation was always the endpoint, never mere description).

An example of the beginnings of this type of work is Hugo Hassinger's *Kunsthistorischer Atlas der Reichshaupt und Residenzstadt Wien* (Art historical atlas of Vienna) published in 1916, with its large scale, building by building plotting of building age in each district of the city, illustrated with photographs and building plans of characteristic examples of each category. Hassinger was concerned with the preservation of the historical townscape of an important city under pressure from the transformative processes of the Industrial Age. His atlas was therefore a resource for those seeking conservation of the building fabric, rather than an explanation of the patterns mapped. However, a later study of Basel (published in 1927) paid more attention to patterns of land use and residential densities thereby providing the threefold framework for explanation of form, function, and developmental history.

More significant than Hassinger in the post 1918 years, was one of Schlüter's graduate students at the University of Halle, Walter Geisler. His thesis on Gdansk (Danzig) (1918) not only investigated the form of the inner city in much more detail than hitherto, but also presented the data on land use and building height in detailed color maps. This was followed, in 1924, by a book aiming to provide a comprehensive comparative study of German town sites, plan forms, and building types. This book, *Die deutsche Stadt*, established Geisler as the central figure in inter war urban morphological studies. Unfortunately, the scale of this undertaking (matched by another such survey by Martiny 4 years later) pushed these authors, and others, into simplistic and almost wholly descriptive (morphographic) classificatory terminologies. A large number of regional urban studies began to use such descriptive classifications of town plans in the inter war period in Germany. For experienced researchers, such broad descriptive surveys were a useful prelude to tackling research questions exposed by variations in the mapped forms, such as the difference between the regular, planned urban forms of eastern Germany and the much more irregular looking town plans of the western parts of the country first observed by Fritz as early as 1894.

A second major problem for these broad regional studies was their use of small scale town plans and the devising of classifications based entirely on street plans. The enormous variability of street plans meant that most larger towns seemed to be unique in their plan characteristics, while the small scale plans meant that the temporal compositeness of most historic town plans, which is revealed in the details of plot boundaries, remained unseen. Poorer studies of this kind also failed to consult historical town plans so that even substantial plan change related to particular periods remained unidentified. Finally, a failure to engage in cross disciplinary discussion with the literature of urban history and urban planning meant that substantial advances in these fields were often ignored for long periods. It was not until the very end of the inter war period, for example, that geographers began to take notice of German constitutional historians' discovery that most older towns had not simply grown from villages but had been deliberately founded by means of legal charters from secular and ecclesiastical lords, such as the famous group of towns founded by the Dukes of Zähringia, studied by Hamm and published in 1932.

In the English speaking world, most of the interest in historical urban plans and their development through time came from town planners writing text books for the training of new entrants to their profession. In providing a history for planning, they naturally lighted upon the regular street grids of Roman cities, medieval new towns such as Salisbury, and renaissance and absolutist planning in places such as Turin and Karlsruhe. An exception was two studies of towns in the Baltic region, by American geographer J. B. Leighly, which used the term 'urban morphology' as early as 1928, but other American urban geographers did not follow this lead. It was not until R. E. Dickinson began writing in the 1940s, that knowledge of German scholarship on the form of cities began to penetrate the English speaking world, while in France it was a historian, Pierre Lavadan, who provided the major text on medieval urban forms in his *Histoire de l'Urbanisme* (1933); but, as with German work, this concentrated on street plans.

Post-War Research

After World War II, there were further detailed studies of urban plan developments published in the German speaking world but already the very different ideas published by Christaller in the 1930s were providing a new focus for research in urban geography. Thus, when M R G Conzen, a German emigré geographer to Britain in 1933, published his seminal work on the town plan analysis of the small Northumbrian town of Alnwick in 1960, his ideas were both out of place (they would perhaps have received a wider hearing in Germany) and out of time (since the socio economic model building derived from Christaller's research was now gathering pace). Nonetheless, the monograph on Alnwick proved to be of enormous long term significance for studies in urban morphology. Conzen sought to provide morphological studies with both a carefully defined terminology and a set of conceptual ideas of broad application. All this was derived from the extremely detailed study of the developmental growth of an English market town surrounded by an aristocratic estate which provided a rich array of source materials, including an all important sequence of large scale, accurate town plans from the 1770s onward. At the same time, Conzen was testing the applicability of his concepts on the much larger city center where he was based – Newcastle upon Tyne, and this was published 2 years later. Conzen was well aware that plan analytical studies needed to be combined with studies of land use and the social geography of the city since he had begun such studies of Alnwick; but they were to remain uncompleted. He also retained an active interest in the applicability of morphological ideas to the conservation planning of modern cities, since he had trained as a town planner after his arrival in England.

Urban Fringe-Belts

The first of his conceptual ideas to be taken up in further research by others was that of the urban fringe belt. This posited that towns developed physically in temporal pulses of rapid growth, usually marked by extensive house building, and periods of stasis. These pulses reflect the longer term variations in the national and regional economy. During periods of economic depression, the edge of the townscape is marked by the establishment of land uses which require larger plots of land and which are unable to compete with house builders in boom times. The most distinctive of these uses are institutions of various kinds and large mansion houses. These distinctive land use belts (with equally distinctive plot patterns) are also prone to develop and consolidate around physical impediments in the townscape (fixation lines) such as town walls and transportation lines. J W R Whitehand published a body of work in the late 1960s and 1970s which went a long way to working out the details of precisely how economy, rent theory, house building cycles, and institutional development interacted in the formation and development of urban fringe belts. These ideas were tested in the context of northwest Glasgow and west London. Other scholars looked to explore the sequential development of institutional plots, seeking to explain their longevity in the townscape, the influence of family life cycles on the development of urban mansion houses, the functional linkages that developed between fringe belt users over time, and the interaction of building cycles with the cyclical characteristics estab lished by new building bye laws. All these studies sought to integrate the fringe belt concept into wider aspects of urban geographical research but their temporal concern has been largely with the nineteenthth and twentieth centuries, given the need for detailed datasets; few re searchers have sought to investigate earlier periods of fringe belt formation or transformation.

The Burgage Cycle

Conzen's concept of the burgage cycle has received generally less attention from historical urban morph ologists. The burgage cycle posits that urban plots in the medieval cores of towns undergo a cyclical process of increasing building coverage and building density through the industrial era. This leads to an increasingly poorer fit between building and land use, a propensity to poorer building maintenance, and so to either re furbishment, or demolition and rebuilding, often with a period of what Conzen termed 'urban fallow' in between. This process might also be accompanied by plot amal gamation to provide larger areas for modern town center buildings. In the modern era, the cycle from new building to demolition and rebuilding might well be as short as 25 years but for earlier historical periods, the full develop ment cycle might take several hundred years, depending upon the destructibility of the building materials and the adaptability of the buildings. The burgage cycle was later recognized as a specific case of a wider urban plot development cycle. Historically, specific studies have been undertaken of the Polish city center of Łódź in the nineteenth century, by Koter, and of a sample of urban blocks in the American city of Omaha, by M P Conzen. Archaeological work in a number of European cities has revealed the historical developmental sequence for much earlier periods, most notably in Lübeck, Germany, but publications have not couched these transformations in conceptual terms (**Figure 1**).

The Transformation of Historic Towns

A continuing theme in research over the past 50 years has been a concern with the origin, form, and transformation through time of historic towns. This is characteristic of almost all European language realms, though the specific use of the techniques of Conzenian town plan analysis has been much more restricted to historical geographers. Increasingly, however, the benefits of interdisciplinary teamwork is being recognized and architects, archae ologists, geographers, and historians are working together to explore the long term changes in the townscapes of historic cities, especially those of the medieval period. As a consequence, the period compositeness of medieval town plans is now clearly recognized. Key to this understanding was Conzen's concept of the 'plan unit', a morphologically homogenous area in terms, especially, of its plot pattern. Plan units can be recognized at different scales and there has been no development of a common methodology for their discernment, though Conzen's own methodology on the plan analysis of the market town of Ludlow can be implied. Ludlow has been an important laboratory for town plan analysis with suc cessive researchers providing analyses which build on Conzen's. Despite the lack of a specific methodology, research by Slater, Lilley, Baker, and others has provided evidence which has established beyond doubt that medieval towns, too, grew in developmental pulses with each phase leaving its distinctive evidence in the town plan as a plan unit (**Figure 2**). This type of research has also provided extensive evidence for the practices of medieval town planning showing that planning was characteristic of a very large proportion of towns large and small and not just in specific regions, such as Gas cony or eastern Poland, where grid street plans and rectangular market places made the recognition of town planning obvious. Whitehand and Cai have recently utilized the technique to analyze the development of a Chinese city with its very different sociocultural framework.

The Alnwick study includes a metrological analysis of the plot dimensions in the town center. This suggested an average plot width of some 28–32 English feet (8.5–9.75 m) which, it was suggested, equated with a two bay

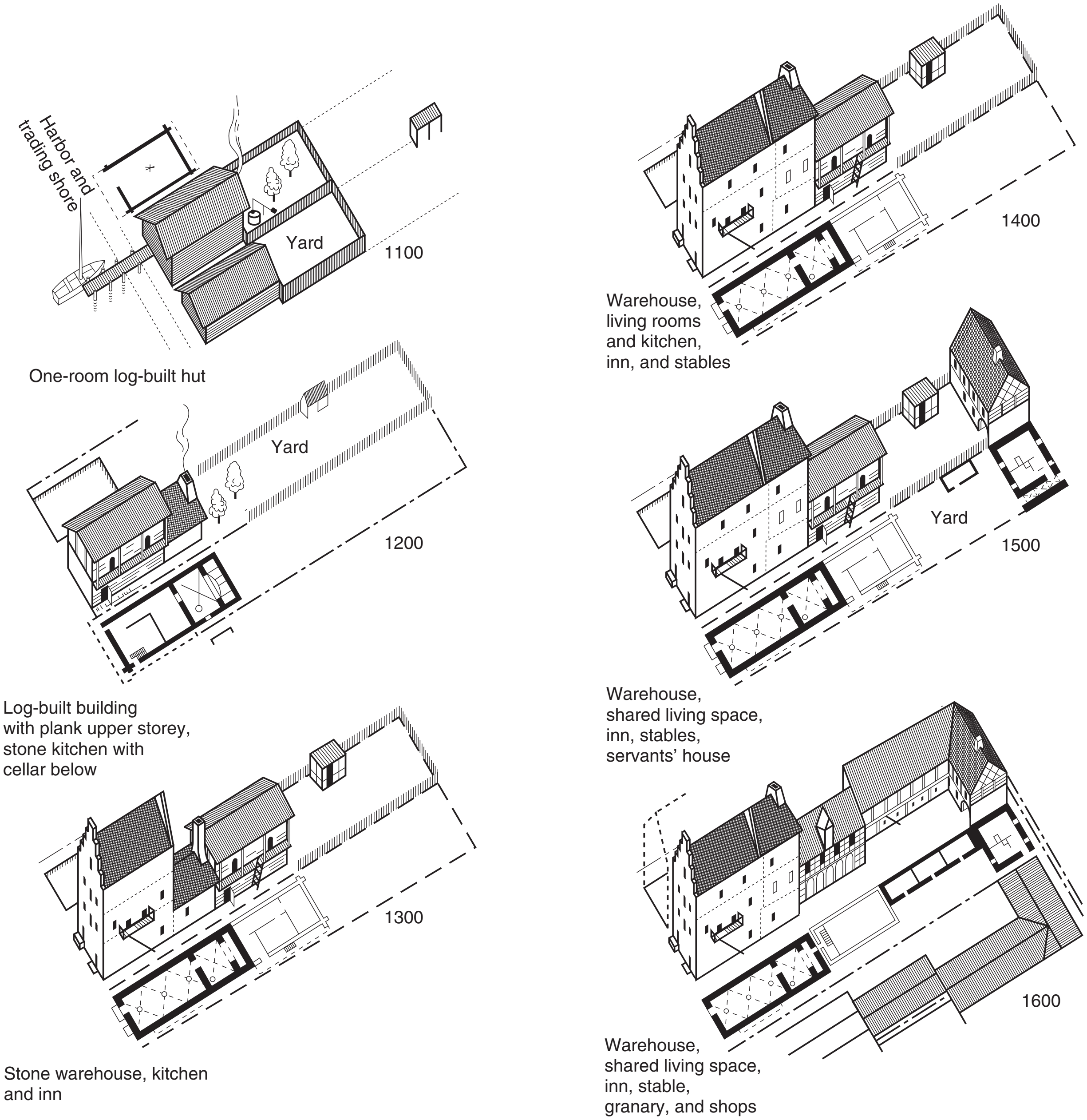

Figure 1 The hypothetical burgage cycle of one plot in a medieval Baltic-Sea port town between 1100 and 1600 AD. The diagram shows the plan form and the building types of a single plot. Adapted from Griep, H. G. (1985). *Kleine Kunstgeschichte des Deutschen Bürgerhauses* Darmstadt.

timber framed building aligned crossways on the plot. At much the same time, two Polish geographers, Gola chowski and Pudełko, were using metrological techni ques in their analysis of the planning of medieval towns in Silesia. A decade later, Slater made extensive use of metrology to gain a better understanding of medieval English town planning, beginning with Stratford upon Avon where there were important documentary sources. It became clear that in most English planned towns, lords used the metrology of the statute perch or the statute acre to lay out urban areas. In Germany, Lafrenz was similarly using metrological analysis to gain a better understanding of the early modern port town of Friedrickstadt.

More recently, these plan analytical studies have been extended from small towns to large medieval cities, Lil ley's study of Coventry, and Baker and Holt's analysis of Gloucester and Worcester being especially notable. Archaeological findings were incorporated into these studies and archaeology has been important, too, in proving Conzen's hypothesis of the longevity of urban plot boundaries. The excavations of York's Coppergate

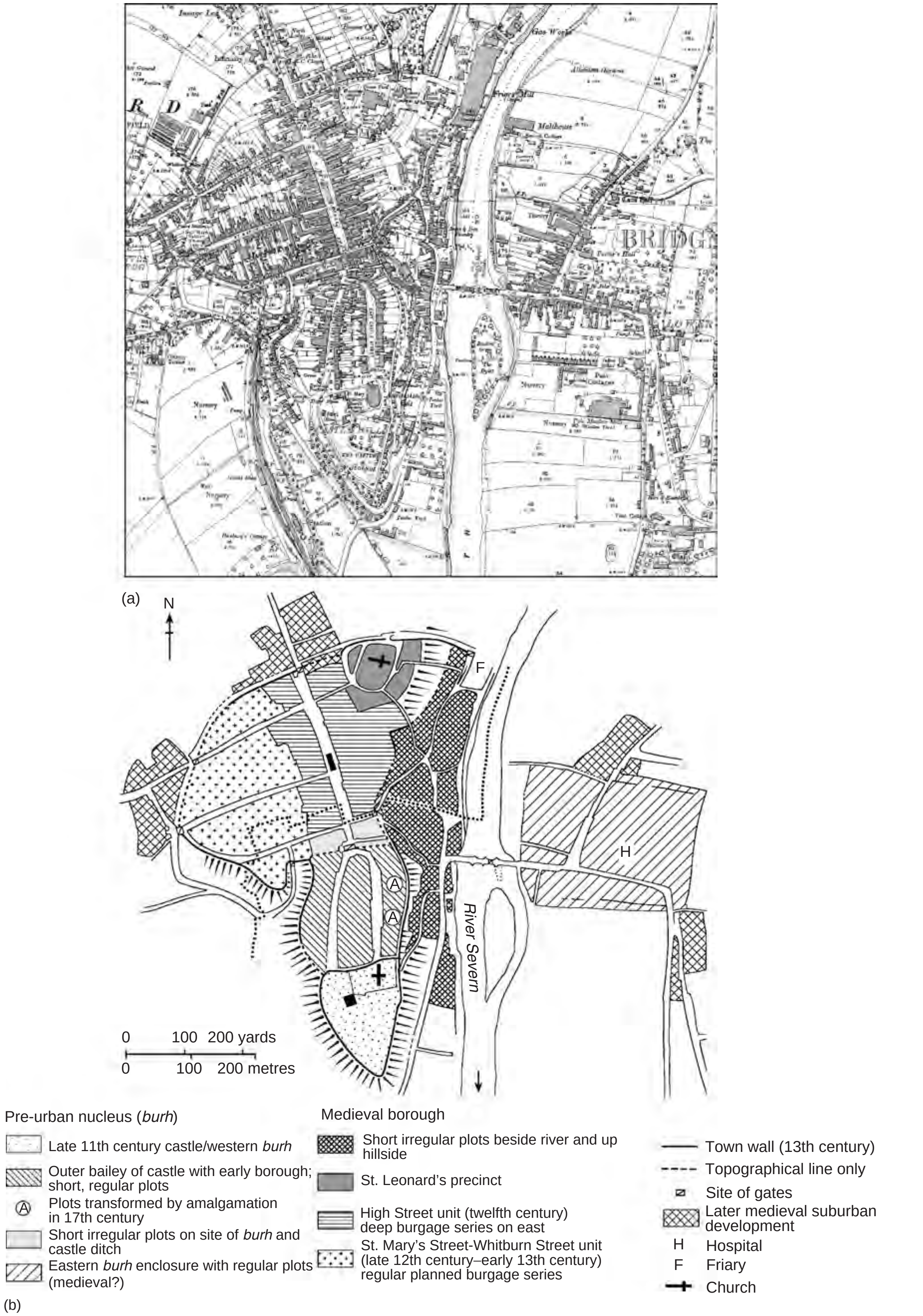

Figure 2 Town-plan analysis of Bridgnorth, Shropshire. First Edition 1:2500 Ordinance Survey County Series (a) and the plan units of the medieval town. (b) From Slater (1990).

site and of central Dublin showed that, even in the tenth and eleventh centuries when buildings required rebuilding every few decades, plot boundaries were remarkably stable. Archaeological evidence has also been crucial in defining the late twelfth century as the period when timber frame building techniques transformed towns by giving them a long lasting built fabric and a multistorey townscape which allowed more people to live in the same unit area. This is a characteristic of the whole of northern Europe as recent excavations in a number of German cities have begun to demonstrate. Architectural historians meanwhile have begun to analyze the considerable number of high status stone built dwellings that survive in historic towns from the eleventh and twelfth centuries in central and southern Europe. Buildings provide a third dimension in historical urban morphology and provide the most visual manifestation of urban forms. However, morphologists are more interested in buildings *en masse* than with innovative or architecturally distinguished structures by famous architects. Some historians have worked in this way on the building types of the nineteenth and twentieth century city including, for example, Muthesius's studies of terraced housing and of blocks of flats. Trowell's work on mid nineteenth century terraced housing in Leeds has shed light on the involvement of architects in the construction of ordinary housing while Whitehand and Carr have done similar research on inter war speculative housing in Britain. In North America, Gadd and Holdsworth have published a number of studies on the development of the office building from the late nineteenth century onward, linking it to both changing business practice and technological innovations such as the elevator and the typewriter. There have also been both architectural and humanistic analyses of the skyscraper as a building type.

It is important to note work on urban decline and plan shrinkage in the later medieval period due to the forces of economic change and, especially, of population decline. Historian Derek Keene's monumental study of Winchester was amongst the first detailed work to bring the physical consequences of economic and social decline to wider notice while an interdisciplinary volume edited by Slater explored these consequences over a much broader timespan and in a number of geographical areas. There has also been an important international debate on the significance and characteristics of orthogonality and geometrical knowledge in medieval grid planned towns. It is important because architectural historians have published a number of studies suggesting that the layout of some towns in Europe was based on extremely sophisticated geometrical knowledge. However, the only group of towns where documentary evidence supports such a hypothesis is the Florentine new towns developed by the city state in the late thirteenth century. Friedman's analysis of these town plans and documents is exemplary and shows that early Renaissance mathematical knowledge was being applied in the layout of at least some of these places, most notably at San Giovanni Valdarno. There is no doubt about the orthogonality of European colonial towns and they, too, have been researched, usually to discern the influence of the variety of regulatory regimes that governed their establishment. Examples include Williams' analysis of early plans in South Australia and Reps' monumental surveys of the towns of colonial North America. The majority of such studies content themselves with detailed description of the initial plan only and do not analyze their subsequent development. The grid plans of such towns are remarkably flexible and their built forms and land use patterns can undergo processes of change very quickly as work by Scheer on Cincinnati, Conzen on Omaha, and Moudon on San Francisco residential areas have shown.

Agents of Change

Another major strand of research in historical urban morphology over the past half century has been that investigating the agents and agency responsible for urban landscape change. Some of this work has been on medieval towns. Here the town founding activities of both secular and ecclesiastical landholders, both episcopal and monastic, have been investigated, while Lilley has drawn attention to the overwhelmingly Christian ethos of European medieval society and therefore to the underlying precepts that sometimes were grounded in urban plans and buildings. The majority of research on agents and agency, however, has been concerned with the industrial and modern periods. The theme which has been researched over the longest time period has been concerned with the influence of land ownership on urban development and this has been the subject of research by both historians and geographers. Dyos' seminal study of Camberwell, Cannadine's work on Edgbaston, and Beresford's detailed analysis of central Leeds are amongst the best known early studies by historians while Roger's recent analysis of nineteenth and twentieth century individuals and institutions involved in Edinburgh's development demonstrates that there is still much to learn about the variability of urban land ownership patterns. Mortimore's study of Bradford is one of the earliest by a historical geographer, while Springett's more detailed work on West Yorkshire textile towns is another study in the same vein.

For twentieth century cities, a major source of information on agents of change is the building control records and the planning applications submitted to local authorities. The 1980s saw a plethora of investigations, by Whitehand and a group of doctoral students at

Birmingham, using these sources. However, there are problems with these data which need to be recognized when they are used to try to reconstruct the making of urban landscapes through a decision making approach. Freeman was one of the first to explore the range of agents active in the development of commercial town scapes in the period after World War II. Subsequent studies have explored the influence of particular agents. These include estate agents in residential development; architects in commercial office development; and local councilors in the planning process. This research theme has also introduced the ideas of innovation diffusion and distance decay into urban morphological work. It has also led to significant debates on the management of change in historic towns and the esthetics of new buildings in historic settings. Punter's study of office building in Reading showed clearly that esthetic considerations were the first thing to be sacrificed to the economics of speedy development and mainstream architectural fashion dic tated by the office letting market.

Urban Maps and Atlases

Since large scale maps are such an essential source for historical urban morphology, it is important to take note of the research which has led to both urban historical atlases and the European Historic Towns Atlas project. The latter was begun in the late 1950s with the intent of producing national atlases, to a common specification, of a representative collection of historic towns as they stood at the beginning of the industrial era so as to enable comparative research to be undertaken. An accurate map at a scale of 1:5000 was the key product. The project continues to expand to encompass new national and re gional atlases but many of the early atlases continued to make comparative work almost impossible with their very different specifications. The multivolume German atlas, and a number of German regional town plan at lases, means that more central European towns are mapped in this way than for any other region, at least partly because there is a long tradition of this type of research dating back to the 1930s. German historical geographers have also led the way in sociocultural topographic mapping of occupational, wealth, and social class data for all periods from the late medieval period onward. Denecke's mapping of towns in Lower Saxony, especially of Göttingen, in the early modern period is notable. The plethora of national historical atlases that characterize the second half of the twentieth century, most notably those of Canada and New Zealand, include some maps which reflect the morphological development of towns in different periods, and there are also historical atlases of large cities such as London and New York, but the concepts and terminology of historical urban morphology rarely find a place there.

See also: Feudalism and Feudal Society; Historical Geographies, Urban; Medieval Historical Geographies.

Further Reading

Baker, N. J. and Holt, R. (2004). *Urban Growth and the Medieval Church: Gloucester and Worcester*. Aldershot: Ashgate.
Conzen, M. R. G. (1960). *Institute of British Geographers Publication 27: Alnwick, Northumberland: A Study in Town Plan Analysis*. London: George Philip.
Friedman, D. (1988). *Florentine New Towns. Urban Design in the Late Middle Ages*. Cambridge, MA: MIT Press.
Larkham, P. J. and Freeman, M. (1988). Twentieth century British commercial architecture. *Journal of Cultural Geography* 9, 1 16.
Larkham, P. J. (2006). The study of urban form in Great Britain. *Urban Morphology* 10(2), 117 141.
Lilley, K. D. (2000). Mapping the medieval city. *Urban History* 27(1), 5 30.
Lilley, K. D. (2004). Cities of God? Medieval urban forms and their Christian symbolism. *Transactions of the Institute British Geographers* NS29, 296 314.
Roger, R. (2001). *The Transformation of Edinburgh. Land, Property and Trust in the Nineteenth Century*. Cambridge: Cambridge University Press.
Slater, T. R. (1987). Ideal and reality in English episcopal medieval town planning. *Transactions of the Institute British Geographers* NS12, 191 203.
Slater, T. R. (ed.) (1990). *The Built Form of Western Cities*. Leicester: Leicester University Press.
Slater, T. R. (2000). Understanding the landscape of towns. In Hooke, D. (ed.) *Society for Landscape Studies Supplementary Series 1: Landscape, the Richest Historical Record*. 97 109.
Whitehand, J. W. R. (1967). Fringe belts: A neglected aspect of urban geography. *Transactions of the Institute British Geographers* 56, 39 55.
Whitehand, J. W. R. (1987). *The Changing Face of Cities*. Oxford: Blackwell.
Whitehand, J. W. R. (1992). *The Making of the Urban Landscape*. Oxford: Blackwell.
Whitehand, J. W. R. and Carr, C. M. H. (2001). *Twentieth century Suburbs: A Morphological Approach*. London: Routledge.

Relevant Websites

http://www.intbau.org
International Network for Traditional Building, Architecture and Urbanism.
http://www.planninghistory.org
International Planning History Society.
http://www.urbanform.org
International Seminar on Urban Form.
http://udal.org.uk
Urban Design Group.

Urban Morphology

K. D. Lilley, Queen's University Belfast, Belfast, UK

The Place of Urban Morphology in Human Geography

Not all urban morphologists are geographers, and indeed urban morphology has been subject to marked fluctuations in popularity among Anglophone human geographers. As a broad sketch, it enjoyed most academic weight within the UK and US urban geography during the 1950s and 1960s, but with the 'quantitative revolution' of the later 1960s and 1970s, and the 'cultural turn' of the 1980s and 1990s, there was a marked downturn in its fashionability among urban geographers. It still appears in urban geography text books as a field of enquiry, but its presence is diminished. In these studies, urban morphology is defined in terms of a concern with studying the material built form of urban landscapes and identifying different types of urban landscape based upon their morphological characteristics, as well as explaining variations in urban form through historical processes and social agency. It is by its very nature a subject that relies on the use of cartography, both as a medium of analysis and presentation, as well as architectural study. Besides appealing to urban geographers then, urban morphology has attracted the attention of architects, planners, and urban designers, who have also contributed to its development and character. It is also a subject that features in historical studies of towns and cities, carried out by both historians and archaeologists, who make use of the inherent conservatism of urban landscapes to deduce patterns and processes of historical development. In conceptualizing the disciplinary place of urban morphology, then, it may be said to cross three overlapping areas: geography, architecture, and history.

Within geography, four key themes characterize the work of urban morphologists. The first is the use of street patterns to derive typologies of urban form, using a classification system to identify different types of town plan. This approach uses the study of maps of towns, particularly those of small scale (1:10 000 +), and focuses principally on the overall layout of streets at the scale of the entire town or city. Using a key criterion, the principle here is to reduce to a range of symbols or types the great variety of urban forms that exist. For example, regularity varies between street patterns, so degree of regularity is a means by which towns can be classified into types, those that are highly regular in form as opposed to those that are highly irregular. Street plans also fall into different types in terms of their overall configuration as 'grid' or checker plans, 'linear' or axial plans, 'radial' or concentric plans, and such like. Once a classification system is derived from studying the layouts of street patterns for a group of towns or cities (e.g., within a certain region, or for a certain historical period), it is commonplace to examine their geographical distribution pattern by mapping them out, and using this as a means to identify spatial variations in urban form. The idea here is to use mapping and classification as an explanatory tool. But there are two important criticisms of this approach. The first is that by its nature it is highly reductive. Street patterns are morphologically complex and do not lend themselves to simple categorization based upon types. So reducing a whole town or city to one single plan type is misleading and oversimplistic. It also overlooks the composite nature of urban street patterns and does not accommodate temporal changes in form. The second criticism derives from the first. Since the typology is itself suspicious, being insufficient to recognize internal spatial and temporal variations in urban form, it follows that any distribution maps based upon them are also suspect, and that their patterns are neither historically meaningful nor geographically significant.

The second theme in geographical studies of urban form recognizes the problems of the typological approach and is in a sense a response to those criticisms leveled against it. Rather than reducing complex urban forms to single types, it is possible to use morphological complexity as a basis to analyze the formation and transformation of urban landscapes and map out over time how particular towns and cities took shape. The principle here is that urban landscapes are a palimpsest – a record of their own evolution. The patterns of streets and lots that make up an urban landscape fossilize the forces that originally made them, making it possible to trace their morphogenesis even in remote periods of history when contemporary accounts are not so forthcoming. In looking closely at the patterns of streets and lots that make up a town's plan it is possible to see variations in their form, as new areas of streets and lots were laid out in the past. These areas are thus often morphologically distinct, and collectively are what gives town plans compositeness in form which makes it difficult to reduce them to a single type. Analyzing the period compositeness of town plans requires large scale mapping (1:500–1:5000) which shows, in detail, streets, lots, and buildings in plan form. It is possible to trace morphogenesis through historic cartography, but in most cases detailed mapping of towns does not predate 1800. The inherent conservatism of urban landscapes, especially in street and lot patterns, means, however, that modern town plans contain evidence for urban features surviving from earlier historical periods. These inherited built forms often date back to the time when a town or city first came into being, and so it is perhaps of no surprise that studying urban forms this way is not just of interest to historical geographers but also to historians and archaeologists.

Besides identifying from plan analyses how urban landscapes have changed over time, urban morphologists in geography are concerned, thirdly, with explaining these observed changes in form. For this, they refer to the role of 'agents', both direct and indirect, who have an influence in processes of urban formation and transformation. The concern here is with those who have agency in the urban landscape, and not only identifying simply who is involved but how much influence they exercise in matters of decision making, either individually or as groups. Those 'direct' agents of change who have attracted particular attention are landowners, developers, architects, and planners, and whose activities can be examined by reference to documentary sources such as planning and development control submissions kept in public and private archives, as well as by conducting interviews of those concerned. In this regard, urban morphologists are engaging with the architectural fabric of urban landscapes and not just plan forms, recognizing that the changes occurring in built environments are often small scale and cumulative, such as façade alterations or building extensions, and not always large scale and comprehensive. Seeing the urban landscape as an architectural space, as well as fashioned through the ideas and practices of agents such as architects and planners, has brought closer contact between academic urban morphologists – including geographers – and practitioners of urban morphology – including architects and urban designers. In historical studies of urban form, too, this concern for agency is important for revealing the cultural and social processes that shaped urban landscapes in the past, and for linking particular periods of urban formation and transformation to certain agents. In looking at agents and agency, then, urban morphologists seek to connect the study of urban forms with the people that make urban landscapes. There is more scope, however, to examine not just how urban landscapes are produced but also how they are consumed, and how agency concerns not just decision makers but all those who inhabit built environments.

Finally, urban morphologists have shown interest in the meaning and symbolism of urban landscapes, particularly their 'iconography' in terms of power and authority. This stems from an interest in the agents who shape urban landscapes, and their ideologies and ideals expressed in urban form. The principle here is that forms convey and construct social and cultural meanings, so that urban landscapes reproduce and reinforce social divisions and inequalities, for example, or project the ideas and visions of state leaders, philanthropists, or architects, fashioning an urban landscape in their own image. The forms of urban landscapes are thus inscribed and imbued with meanings and not neutral or inert, and may be read or interpreted as cultural 'texts'. These urban landscapes are not just present on the ground either, but exist in the pages of books or in photographs in magazines, or in paintings and maps, questioning the perceived gap between 'real' and 'imagined' geographies. Indeed, this interrelationship between urban imaginary and materiality is how urban landscapes construct and convey cultural meanings: it is how they work. While urban morphologists have examined such themes in contemporary settings, or for the more recent times, the iconography of past urban forms is an underexplored area.

A Historical Geography of Urban Morphology

Urban morphology has formed a part of Anglophone human geography since the early 1900s. The aim here is to summarize its development since then, and to identify particular 'schools' or traditions of urban morphology that emerged over this time. These 'schools' are an important dimension of morphological study internationally and reflect different strands of urban morphology as well as the influence of key thinkers at particular times. The most thorough surveys of these schools and their development are those by Whitehand. He divides urban morphology internationally into four main traditions, comprising 'morphogenetic geographers', 'Conzenian geographers', 'other English speaking geographers', and 'urban designers'. Since 1992, and thanks to the work of the International Seminar on Urban Form (ISUF), greater recognition has emerged of the diverse research traditions in urban morphology, especially those beyond the Anglophone world, summaries of which appear in the journal *Urban Morphology*, itself a sign of greater international dialog among urban morphologists of different traditions.

As far as Anglophone urban morphology is concerned, its early influences came from continental Europe, Germany, and Italy in particular, during the middle decades of the twentieth century. The more dominant influence on those Anglophone urban morphologists who were geographers (rather than architects) was Germanic (rather than Italian), and derives from a strong tradition of settlement study and landscape morphology in central European geography, beginning with the work of

Schlüter in the 1890s at the University of Hälle. This earliest work of German speaking urban morphologists was characterized by a concern for creating typologies of historic urban forms, especially medieval towns and cities of continental Europe. For instance, a study of West German towns by Klaiber identified six plan types, while Geisler attempted to classify all German town plans according to their street systems in his book, *Die Deutsche Stadt.* By the 1940s, this 'morphographic', classificatory approach was giving way to a 'morphogenetic' approach based upon principles of plan analysis in which the period compositeness of town plans was recognized and put to use in mapping out not typologies but histories of urban form. This had begun in the inter war period with studies by Dörries, Bobek, and Rörig, and developed through a series of German historic town atlases that used large scale town plans and also identified phases of historical urban development.

It was this work by German speaking urban morphologists that gradually became known by English speaking geographers in the UK and USA. R. E. Dickinson had introduced German urban morphology to a geography audience in Britain in the 1930s, and also undertaken a typological study of town plans in eastern England very much like those of Klaiber and Geisler. In his *West European City* he also referred to the then more recent (and sophisticated) work of Bobek and Rörig. But Dickinson's influence on Anglophone geographers was overtaken in the early 1960s by a German émigré geographer, M. R. G. Conzen, who used his grounding in German geography to develop an approach to urban morphology that has become known as the 'Conzenian tradition'. In the US, it was the work of Leighly in the 1920s and 1930s that brought urban morphology into cultural geography, in particular, the Berkeley School, but the focus there was rather more on rural settlement than urban, and subsequently it was Conzenian urban morphology again that came to influence North American historical geographers' work on built environments and urban landscapes.

The 'Conzenian geographers', whose intellectual origins stem back to the 'morphogenetic geographers' of continental Europe, share a common belief that urban landscapes are dynamic yet at the same time conservative. This principle guides their thinking on the methods as well as applications of urban morphology. It concerns the historicity of the urban landscape, and the values attached to this. Conzen initially dealt with problems of method, of how to interpret town plans 'morphogenetically', and of how to disaggregate the morphological complexity of urban landscapes both in terms of their plan form and built fabric. He devised an approach, 'town plan analysis' to identify 'plan elements' (streets, plots, and buildings) that together constituted a town's plan, and used this to define areas of morphological homogeneity, which he called 'plan units', each indicating a phase of urban morphogenesis as he saw it. Conzen conceptualized formation and redevelopment processes through his close analysis of streets, plots, and buildings, and paid particular attention to cyclical changes in form, for example, the formation of 'fringe belts' and the internal transformation of building lots ('the burgage cycle'). These were discussed in his monograph on Alnwick, a small historic market town in northeast England. In this study he also paid attention to what else gave a townscape its character, not only its 'town plan' but also local 'building fabric' and 'land utilization'. With building fabric, he sought to focus on architectural style and building materials and façade features, mapping building types in detail, and likewise with types of land utilization. Such detailed study required fieldwork and not only borrowed on Conzen's training as a geographer in Berlin between the wars but also his early profession as a town planner. The ways that these three 'form complexes' – town plan, building fabric, and land utilization – all combined gave rise, he suggested, to the *genius loci*, or spirit of place, achieved through long term historical and morphological development and redevelopment processes.

Conzen's methodological and conceptual ideas, as well as his thinking on historical townscapes, has formed a critical school of thought in urban morphology that touches not only geography but also history, archaeology, architecture, and planning. In UK geography in particular, Conzen's work was further developed in the 1970s and 1980s, primarily at the University of Birmingham under the auspices of the Urban Morphology Research Group (UMRG) set up by J. W. R. Whitehand. This fostered the Conzenian tradition and helped to disseminate Conzenian ideas across Britain, as well as Ireland. Two areas in particular were developed, one historical in focus concerning the use of 'town plan analysis' to reconstruct the early histories of European towns and cities, and the other more contemporary in focus concerning the role of agents and agency in urban redevelopment and development control. Through this activity, the 'Conzenian school' has gained identity and international recognition, with new links fostered between the UK and urban morphologists working in Spain and Poland. It has also revealed the presence of cognate schools of urban morphology, notably the 'Italian school' that had emerged in the 1950s and 1960s at the same time that Conzen was writing, through the work of Muratori and Caniggia. In the US too, the Conzenian tradition was fostered by an out migration of UK geographers, including M. P. Conzen and Deryck Holdsworth. But neither in the UK nor the USA has urban morphology gained ascendancy in human geography as a whole. It remains a fairly peripheral part of the discipline. Indeed, by the 1990s, some UK geographers were beginning to suggest that urban morphologists' view of 'landscape' was too narrow, too empirical, particularly in the light of a more theorized approach to landscape that had begun to root in Anglophone historical and cultural geography which was more concerned with iconography than morphology.

With the foundation of ISUF, in 1994, and the pub lication of its journal *Urban Morphology* (since 1997), the various 'schools' of urban morphology present both within and between academic disciplines and national intellectual traditions have begun to cross fertilize, making it perhaps now more difficult than 20 years ago to map out a genealogy of urban morphology as Whitehand did in the 1980s. The subject continues to attract com ment, as it always has done, not just by geographers but also by others interested in understanding built forms and managing urban landscapes.

Topics of Debate and Controversy in Urban Morphology

Three controversial topics merit particular attention: theory, method, and application. On the whole, 'critical' theory has remained virtually absent from Anglophone urban morphology. This is despite a more theorized agenda among human geographers on urban matters as evidenced by the contributions of David Harvey, Derek Gregory, and Ed Soja, for example. Theorization of the idea of landscape, and the power of landscape, has also been generally neglected by urban morphologists, so the ideas of Henry Lefebvre, Michel Foucault, and Jacques Derrida, rarely feature even though they would have much to offer urban morphologists' understanding of how urban landscapes work. There are also epistemo logical matters to be addressed, again drawing on recent critical thinking in the social sciences and humanities, on the way urban morphology is done, on the reading of 'texts', and on their duplicity of meaning, for example. Apart from the need for an epistemology of urban morphology, there are practical matters of method that have raised debate, especially for those seeking to re construct past urban landscapes from later cartographic sources. The idea of using modern representations of urban landscapes as proxy evidence for previous urban landscapes is fraught with difficulty. Recently, though, the questions that were being raised by historians as to the validity of morphological methods have begun to be addressed through archaeological excavation work and by geographers and archaeologists working together. Lastly, the potential application of urban morphology has also attracted debate, especially among those concerned with its uses in contemporary townscape management. The transfer of ideas stemming from academic urban morphology for use in urban design and planning as practical guidance is in part made difficult by the dif ferences in approach between geographers and architects in studying urban form. With further interest in the role of agents and agency, there is now more dialog between academic studies of urban form and architectural appli cations of urban form. In effect, these three issues – theory, method, and application – are each inter connected, and no doubt will continue to provide a focus for debate both within and between geographers engaged in the study of urban form, in urban morphology.

See also: City-Region; Historical Geographies, Urban; Housing; Landscape; Landscape Iconography; New Towns; Planning, Urban; Urban Architecture; Urban Design; Urban Morphologies, Historical.

Further Reading

Cataldi, D. (2003). From Muratori to Caniggia: the origins and development of the Italian school of design typology. *Urban Morphology* 7, 19 34.
Darin, M. (1998). The study of urban form in France. *Urban Morphology* 2, 63 76.
De Alvarenga Pereira Costa, S. (2006). Brazilian urban morphology. *Urban Morphology* 10, 142 144.
Gauthiez, B. (2004). The history of urban morphology. *Urban Morphology* 8, 71 89.
Gilliland, J. and Gauthier, P. (2006). The study of urban form in Canada. *Urban Morphology* 10, 51 66.
Kealy, L. and Simms, A. (2008). The study of urban forms in Ireland. *Urban Morphology* 12, 37 45.
Larkham, P. J. (2006). Urban morphology in Great Britain. Urban Morphology 10, 117 41.
Marzot, N. (2002). The study of urban form in Italy. *Urban Morphology* 6, 59 73.
Moudon, A. V. (1997). Urban morphology as an emerging interdisciplinary field. *Urban Morphology* 1, 3 10.
Satoh, S. (2008). Urban morphology in Japan: researching castle towns. *Urban Morphology* 12, 5 10.
Siksna, A. (2006). The study of urban form in Australia. *Urban Morphology* 10, 89 100.
Whitehand, J. W. R. and Larkham, P. J. (eds.). (1992). *Urban Landscapes: International Perspectives*. London: Routledge.

Relevant Websites

http://www.urbanform.org/
Urban Form.

Urban Order

M. Jayne, University of Manchester, Manchester, UK
D. Bell, University of Leeds, Leeds, UK

Introduction

The term 'urban order' implies a number of different things at once; it refers to the ways in which cities are ordered, that is, made orderly (or alternatively disorderly). Urban order is therefore about ordering, about producing order – through planning, for example, which often aims at producing an orderly landscape. So cities are subject to various processes of ordering (and counterprocesses of disordering or reordering), and some key writers have spent much of their time considering the forms of order and disorder in cities, and the tension between forces for order and disorder at work in various times and places. But 'urban order' also refers to the urban hierarchy, to a comparative, evaluative ordering of cities, a ranking based on certain valued or devalued characteristics. Urban order here means sorting different cities into some sort of order, classifying them, and differentiating between them. So, we might think of population size or growth as one way of assessing urban order: bigger cities at the top, getting smaller as we move down the list. Or we might order on the basis of wealth or power – so often these two are hard to separate. Part of this form of ordering is about assessing success and failure – cities on the up, or those that are spiralling downward. The idea of urban order is therefore normative. And, of course, this form of comparative ordering occurs at different spatial scales: there are global urban orders, national urban orders, regional urban orders, and local urban orders. Attentiveness to scalar context is therefore important to any talk of urban order. Our first form of ordering, however, is temporal, or historical, rather than spatial: it seeks to order (in both senses) cities through time, to classify different types of cities as characteristic of (or symptomatic of) different periods in history.

The Epochal City

One of the most enduring pursuits of urban studies has been to conceptualize urban life and urban form at particular periods of history (see **Table 1**). Depicting the political, economic, social, cultural, physical, and spatial changes associated with cities at different times – variously described as premodern, modern, or postmodern; preindustrial, industrial, or postindustrial; and so on – has been one key aim of urban theorists. For over a 100 years, writers have sought to produce generalizable templates, yet given that work has tended to concentrate on those cities that best establish and highlight particular arguments, urban research has been dominated by studies of big cities in the global North.

While such an approach is perhaps understandable as theorists seek to promote breakthroughs in critical debate, just a few cities such as New York, London, Paris, Berlin, Los Angeles, and Tokyo have been used to develop visions of archetypal urbanism. As such, other cities that are of supranational, national, regional, or local significance have tended to be relatively ignored. In a similar vein, when high profile urbanists talk about new and transformed urbanity, they invariably do this via what they consider to be the most high profile urban practices, processes, spaces, places, and actors. In these terms the department store, shopping arcades, the flaneur, the prostitute, gated communities, the ghetto, urban quarters, loft living, the yuppie, spectacular shopping malls, edge cities, entrepreneurial governance, and public–private partnerships are just a few of the numerous landmarks that have come to represent the city in different historical epochs. The depictions of people, places, and practices in order to characterize cities at different periods of history has been a vital factor in aligning the study of cities firmly at the center of social science knowledge.

The political, economic, social, cultural, and spatial stories generated by urban geography have thus created a benchmark against which theorists judge the success and failure of cities. The presence and concentration of landmarks of the epochal city have also become a measure against which cities are deemed worthy of sustained attention (or not). Indeed, the impact of the epochal city approach is so profound that it also pervades critiques of these archetypes. The pull of the global metropolis is so dominant that even when studies address national capitals, regional or provincial cities, it is most often the case that researchers are seeking to debunk theories developed for different versions of the epochal city, rather than seeking to develop specific theories about those 'lesser' cities. Thus, while such work has been useful in identifying that not all cities share the same trajectory as epochal cities, and that very specific local conditions come into play, researchers have not fully developed a vision of the cities they are actually studying. This represents a failure to critically approach cities as

Table 1 Conceptualizations of epochal urbanism

The **modern city** is characterized by its homogenous zoning of business and residential activity with a dominant commercial centre and a steady decline in land values away from the centre. Urban governance in the modern city is based on a managerial approach and involves the redistribution of resources for social purposes, and public provision of essential services. The economy is dominated by industrial mass production and based on economies of scale. The city features functional architecture and a mass production of styles and is planned in its totality with space being shaped by social needs. There are strict class divisions, with a large degree of homogeneity within class groups.
The **postmodern city** is structurally chaotic, and hosts highly spectacular events in contrast to the presence of large areas of poverty. There is also an archipelago of distinct urban areas such as hi-tech corridors and postsuburban development. Urban governance is entrepreneurial and resources are used to lure mobile international capital and investment. The public and private sector work in partnership, and there is a market provision of services. The economy is service based with flexible production aimed at niche markets, based on economies of scope. The most important economic drivers are the globalized telecommunications and finance industries and the city is orientated around consumption. Planning is undertaken in a piecemeal way, and urban design based on aesthetic rather than social ends. Architecture is eclectic and made up of a collage of styles that is often spectacular, playful, ironic, and produced for specialist markets. Culture in the postmodern city is highly fragmented and characterized by lifestyles divisions. There is a high degree of social polarization and social groups are distinguished by their consumption practices.
The **industrial city** emerged as a result of manufacturing and production focused capitalism. The nation state holds close control of city power, and local authorities focus on development and maintenance of collective consumption infrastructure in order to service the needs of capital accumulation (and hence underpins class conflict). The physical organization of the city is based around factory production, and the separation of the household from production. In order to increase levels of production and productivity, urbanization is also predicated on improvements in provision of collective consumption and social welfare such as the provision of mass education, literacy, health, and housing. It is industrialization and not broader capitalist activity that determines the physical, political, economic, and social trajectory of the city. The social structure of the city is dominated by extended family and friendship based on residential proximity.
The **postindustrial city** emerged during the latter part of the twentieth century and is characterized by a declining dependence on manufacturing and industry and a rise in the importance of service industries. There is also a particular emphasis on the role of knowledge in production, consumption, and leisure. This includes an increased importance of higher education in order to stimulate a knowledge economy based on innovation. The social organization of the postindustrial city is characterized by a dominance of knowledge based professional, managerial, and business occupational groups.
The **Fordist city** is organized around industrial production that is low cost and standardized. The city is structured by residential and industrial zones in close proximity to each other, a central business district, and affluent residential areas, especially in suburban locations. The city is characterized by conflictual industrial and social relations. Workers tasks are fragmented, repetitive and the labour force is occupationally divided and management is centralized. The production process is dominated by Henry Ford's mass manufacturing techniques and standardized products that generate mass consumption. Forsism is based on capital intensive, large-scale plant, inflexible production, rigid bureaucracy, scientific management, and semiskilled labour. Political power and social structure is determined by this dominant mode of production and the social divisions of capitalists, managers, and workers.
The **post-Fordist city** is characterized by a production landscape that has experienced the decline of the old manufacturing and 'smoke stack' industries and rise of new computer-based technology and more flexible decentralized labor processes and work. The city is dominated by the autonomy and needs of multinational corporations and global process of capital production that produce new international division of labor. Production in the post-Fordist city is underpinned by the economic possibilities of microchip technology, computers, and the exchange of information and commodities. Economic activity is dominated by large numbers of small- to medium-sized enterprises, catering for segmented markets through flexible production of specialized goods and services. The power and spending of local government is restricted by central government and nonelected bodies often determine economic development and planning policy. Associated social and economic transformations have led to a growth in the number of low paid insecure work and the promotion of consumption cultures pursuing the concept of individual taste and chosen lifestyles distinctiveness. These economic, social, and cultural processes are mapped onto the city in terms of spectacular urban developments contrasted with areas of economic poverty.

Jayne, M. (2005). *Cities and Consumption*. London: Routledge.

diverse and not simply measured in relation to the global metropolis.

The Urban Hierarchy

This theoretical and empirical partial sightedness also resonates through attempts to characterize cities in terms of a hierarchical order (see **Table 2**). As early as 1922, Christaller described seven types of urban areas in Germany. Based on both measurement of population size of the settlement, and the population of a broader market area, small settlements were described as having a market area population of less than 3500, and the other cities were characterized in terms of having 3.5 million people living within their market area. More recently, the Global Analysis of World Cities Group (GaWC) has portrayed a global hierarchy of cities based on concentrations of

Table 2 Conceptualizations of the urban hierarchy

The Christaller central place system		*GaWC inventory of world cities*	*World city hierarchy (based on air connections)*	*Hall's (2004) – unnamed*	*A geographical – functional characteristics (Europe)*
  Marktort Amtorst Kreisstadt Bezirkstadt Gaustadt Provinzstadt Landstadt (Christaller 1996)	Population (market area) 3500 11 000 35 000 100 000 350 000 1 million 3.5 million	*A. Alpha world cities* 12: London, Paris, New York, Tokyo 10: Chicago, Frankfurt, Hong Kong, Los Angeles, Milan, Singapore *B. Beta world cities* 9: San Francisco, Sydney, Toronto, Zurich 8: Brussels, Madrid, Mexico City, Sao Paulo 7: Moscow, Seoul *C. Gamma world cities* 6: Amsterdam, Boston, Caracas, Dallas, Düsseldorf, Geneva, Houston, Jakarta, Johannesburg, Melbourne, Osaka, Prague, Santiago, Taipei, Washington 5: Bangkok, Beijing, Rome, Stockholm, Warsaw 4: Atlanta, Barcelona, Berlin, Buenos Aires, Budapest, Copenhagen, Hamburg, Istanbul, Kuala Lumpur, Manila, Miami, Minneapolis, Montreal, Munchen, Shanghai *D. Evidence of world city formation* *Di. Relatively strong evidence* 3: Auckland, Dublin, Helsinki, Luxembourg, Lyon, Mumbai, New Delhi, Philadelphia, Rio de Janeiro, Tel Aviv, Wien *Dii. Some evidence* 2: Abu Dhabi, Almaty, Athens, Birmingham, Bogotá, Bratislava, Brisbane, Bucharest, Cairo Cleveland, Köln, Detroit, Ho Chi Min City, Kiev, Lima, Lisbon,	London Frankfurt Paris New York Amsterdam Zurich Miami Los Angeles Hong Kong Singapore Tokyo Seoul Bangkok Madrid Wein San Fransisco Chicago Dubai Osaka Brussels (Smith and Timberlake 2000)	Global cities: typically with 5 million and more people within their administrative boundaries and up to 20 million within their hinterlands, but effectively serving very large territories: London, Paris, New York, Tokyo Sub–global cities: typically with 1–5 million people and up to perhaps 10 million in their hinterlands, performing global service functions for certain specialized services (banking, fashion, culture, media) and an almost complete range of similar functions for more restricted national or regional territories: a European capitals apart from the global cities, together with commercial capitals (Milan, Barcelona) and major provincial cities in large nation states (Glasgow, Manchester, Lyon, Marseilles, Hamburg, etc.) Regional: population 150 000 to 1 million, some of these have characteristics of evidence of world city Provincial population (1 000 000–250 000)	Central high-level service cities: major cities (national capitals) and major commercial cities in the so-called Pentagon', London, Paris, Milan, Munich, Frankfurt, Hamburg, Amsterdam, Brussels, Luxembourg Gateway cities (subcontinental capitals) national capitals and major commercial cities outside the Pentagon' acting as high-level service centers for major parts of the European space: Madrid Barcelona, Rome Athens, Vienna, Berlin, Copenhagen, and the canddtate 3 capitals at Prague, Warsaw, Budapest. They also include some larger commercial cities: Manchester, Lyon, Stuttgart, Leipzig Smaller capitals and provincial capitals: these are smaller equivalents for the previous case, commanding less extensive space in terms of population and GDP, in many cases they are at the periphery of the European space: Dublin, Edinburgh, Lisbon, Helsinki, Stockholm, and also smaller, remoter capitals of the extended EU: Bratislava, Ljubljana, and in the future Sofia. They also include smaller commercial centers controlling provincial territories: Bristol, Bordeaux, Grenoble, Strasbourg, Hanover, Bologna County towns: this describes the typical rural administrative and service centers for a surrounding area typically 40–60 km in radius of which hundreds exist in Europe (Hall 2004)

(*Continued*)

Manchester, Montev deo,
Os o, Rotterdam, R yadh,
Seatt e, Stuttgart, Den Hagg,
Vancouver

Diii. Minimal evidence

1: Ade a de, Antwerp, Arhus,
Athens, Ba t more, Banga ore,
Bo ogna, Bras a, Ca gary,
Cape Town, Co ombo,
Co umbus, Dresden,
Ed nburgh, Genoa, G asgow,
Gothenburg, Guangzhou,
Hano , Kansas C ty, Leeds,
L e, Marse e, R chmond, St
Petersburg, Tashkent,
Tehran, T juana, Tor no,
Utrecht, We ngton
(Beaverstock, Tay or and
Sm th 1999)

Adapted from Hall P (2004) European cities in a global world   n Eckhardt F & Hassenpflung D (eds) *Urbanism and Globalization* Frankfurt Peter Lang economic activity. The focus of GaWC has firmly been on seeking out the presence of advanced producer services like financial services, law and accountancy, or on airport connectivity. The counting and analysis of such measures of 'world cityness' has generated a schema of the global urban hierarchy that describes cities in terms of their alpha, beta, and gamma world cities status. This was followed by the labeling of cities in terms of 'relatively strong', 'some evidence', and 'minimal evidence' of world city formation. Of course, in pursuing this research agenda, work has overwhelmingly focused on proving that the schema works with reference to the biggest cities with alpha, beta, and gamma status.

In a similar vein, urbanist Peter Hall suggests a four level hierarchy of cities – including global cities (with 5 million people or more), sub–global cities (with between 1 and 5 million people), regional cities (with a population of between 250 000 and 1 million) and provincial cities (with populations of between 100 000 and 250 000 people). However, upon realizing that his work had failed to account for urban diversity in different parts of the world, Hall then redrafted his vision of the urban hierarchy. This time with a specific focus on Europe, Hall sought to marry geographical and functional approaches and in doing so describes four levels of cities – central high level service cities (major cities and national capitals), gateway cities (subcontinental capitals), smaller capitals and provincial capitals, and finally county towns. This revised version is underpinned by measures that seek to unpack by both the geographical reach of cities, based on the presence of agglomerated economic and cultural activity, as well as air and rail connectivity.

Despite this redrafting, Hall ultimately reproduces deficiencies present in previous attempts to conceptualize the urban hierarchy. Firstly, the versions of the urban hierarchy developed by Hall (and those before him) are illustrated almost exclusively with reference to big cities. It is those cities at the top, 'successful cities', that are deemed to be the pinnacle of urbanity (and all that is considered to be urban). Secondly, cities are judged in terms of a limited number of measures relating to population size and the presence of particular types of economic activities. Thirdly, such visions fail to account for the very different nature of cities and their place in specific supranational and national urban hierarchies. While attempts to depict cities in terms of an ordered hierarchy have played a vital role in highlighting the importance of cities as nodes which attract flows of capital, culture, and people, studies have failed to account for the diversity of urbanity in different places throughout the world. This is in fact a prominent side effect of urban ordering: that it works with a set definition of 'cityness' that precludes some places from the outset; there is a clear order to the research sites that urban geographers study, and upon which they then base their own urban orders. The so called world city, relatively uncontested as the acme of urban development, routinely sits at the top of urban orders, therefore.

Cities and Regions

Attempts to describe the formation and classification of world (or later global) cities have to a certain degree sought to overcome some of the limitations associated with study of the urban hierarchy. This body of work can be picked up in the writing of, among others, Peter Hall, Manuel Castells, John Friedman, and Saskia Sassen, and has had a significant impact on academic debate. Research into cities and regions has importantly described patterns of urbanization in terms of globalization and economic restructuring. This work has had a focus on global cities as centers of government, trade, financial services, insurance, scientific knowledge, mass media, consumption, culture, entertainment, and so on. Cities with the greatest concentrations of such economic and cultural activities are presented as the powerhouses of the global economy.

For instance, researchers have argued that during the 1980s, economic globalization unfolded most rapidly in cities at the top of the urban hierarchy. Measurement of flows of capital, culture, and people was undertaken in terms of the presence of global multinational or national financial sector institutions, the presence of political command and control functions, and the presence of culture industries and the tourist sector. Global city research thus depicts the rapid economic restructuring and associated speeding up of processes of globalization during the 1980s in these cities. In doing so, however, writers inevitably focused only on the biggest urban areas and on what they considered to be most numerically significant economic and cultural indicators found in those cities. Global cities are therefore a particular elite in the urban order, and far from typical of cities around the globe.

In contrast, the theoretical successor of the world city thesis – work that looks at the global city region – is underpinned by the idea that the ordering of cities is more complex than previously identified by the global city research, and therefore demands a more nuanced approach than simply focusing on the biggest of cities. Theorists argue that the study of the global city region should be concerned to look at the relationship between cities, and so has commonly depicted city regions as spatial nodes, made up of specific and unique political, economic, and cultural milieux. However, while one of the core aims of the global city region thesis was to show that not all global cities are synonymous with large metropolitan areas' the global city region research agenda has rarely been pursued beyond the largest and the most well known city regions.

For example, while studies have identified around 300 city regions with populations of around one million, and twenty with excess of 10 million, most of the research has focused on the latter. It is clear that, while work focusing on global city regions offers a clear rationale for pursuing a greater understanding of cities (and the interrelationship between very different cities), research has ultimately failed as yet to fully achieve this agenda. Such limitations can be argued to be underpinned by a mindset that considers theorization of 'the city', and hence the ordering of urbanity, as being most relevantly pursued by concentrating on the largest concentration, no matter what unit of measurement is at hand. Thus, for all of the positive elements of the study of global city regions and its goal to understand the global nature of urban hierarchies in terms of regionally based economic and political organizations, it is overwhelmingly the largest global city regions that have received attention. Again, the research produces its own orderings, its own hierarchies, which then shape the trajectory of subsequent studies – either to compound or to contest the initial ordering.

Ignored Urbanism and Other Orders

Critiques of the ways in which cities have been measured and ordered have focused on the failure to address the diversity of cities throughout the world. This has ensured that cities other than those of the large 'global' variety are at best poorly represented, and at worst completely missing from urban studies research. For example, writers have identified the dominance of studies of global cities and older industrial cities in the north. Others have noted the limited scope of cities designated as 'global' and have sought to stimulate interest in cities that are 'remote'. 'Small cities', 'nonglobal cities', 'second tier cities', and 'would be' cities have also been rightly identified as largely missing from the urban studies agenda.

Other writers have taken a very different approach to questioning the deficiencies measuring, categorizing, and ordering cities. For example, the works of Ash Amin and Steve Graham and more recently Jennifer Robinson show that all cities should be considered as 'ordinary'. They argue that while it is important to understand that cities are important nodes of globalizing political, economic, social, cultural, and spatial dynamics (and some arguably more than others), there is a real need to look at local practices, processes, identities, and autonomies. So it may be more productive to think of cities in terms of the wide variety of their influence and reach, rather than just focusing on their population size, density, or economic activities. Different lenses produce different orderings, and there is a clear emerging agenda to rethink the units of assessment used to produce hierarchical orderings. Why value some activities or resources more highly than others? Studying cities in this way demands an approach founded on the unpacking the heterogeneity of structural and everyday city activity. Such research can, of course, continue to compare cities using the kind of classifications and units of measurements noted above, but must also strive to encompass political, economic, social, cultural, spatial, and physical elements that have not been a key element of research to date.

So far we have only talked about academic orderings of cities. It is important to remember, however, that there are other orders, both official and unofficial – academics have a vested interest in establishing certain measure of 'cityness', but these should be set alongside orders and hierarchies produced by other key actors, and those that circulate outside of academia. This issue relates closely to 'the uses of order': hierarchical orderings have become a key weapon in the competition between cities – itself an increasing feature of urban life in these postindustrial times. The shift to so called entrepreneurial governance has obliged cities and the people who manage them to think about how ranking in an urban order may be a tradable asset in conditions of intense interurban competitiveness – competitiveness for flows of people, capital, jobs, and so on. So ordering matters intensely, not just to academics, but to city managers, boosterists, and arguably to those who live there, too. A good place on the hierarchy – providing the ranking is of positive attributes – is a valuable promotional tool, something to boast about and make the most of. Finally, we should note the circulation of 'unofficial' or 'popular' urban orders, alternative rankings based on less boosterized characteristics or assets, that appear in media accounts – such as 'Britain's fattest cities' or 'sex capitals of the world'.

Beyond the examples given in this discussion, there are, of course, various other ways that urbanity has been theorized, measured, and ordered. For example, research relating to the postcolonial city, the informational city, the city of difference, the revanchist city, and so on has generated much debate, as well as providing a different set of measurements about the ways that cities are ordered. We would suggest, however, that the arguments outlined here regarding the deficiencies of urban research could be relevantly pursued in terms of these other bodies of literature. Thus, while the measuring and categorizing of cities has added much to our understanding of urban life and highlighted the centrality of cities to unpacking broader political, economic, social, cultural, and spatial issues, it is clear that much more needs to be achieved in terms of understanding ordering of cities beyond a limited range of cities and measurements that have dominated urban research to date.

See also: Central Place Theory; Mega-Cities; Modern City; Postmodern City; Urban Growth Machine; Urban Village; Urbanization; World/Global Cities.

Further Reading

Amin, A. and Graham, S. (1997). The ordinary city. *Transactions of the Institute of British Geographers* 22, 411 429.

Beaverstock, J. G., Lorimer, H., Smith, R. G., Taylor, P. J. and Walker, D. R. F. (1999). A roster of world cities. *Cities* 16, 445 458.

Bell, D. and Jayne, M. (eds.) (2006). *Small Cities: Urban Life Beyond the Metropolis*. London: Routledge.

Castells, M. (1989). *The City and the Grassroots*. Berkeley: University of California Press.

Christaller, W. (1966). *Central Places in Southern Germany*. London: Prentice Hall.

Friedmann, J. (1986). The world city hypothesis. *Development and Change* 17, 69 83.

Hall, P. (2004). European cities in a global world. In Eckhardt, F. & Hassenpflug, D. (eds.) *Urbanism and Globalization*, pp 31 46. Frankfurt: Peter Lang.

Jayne, M. (2005). *Cities and Consumption*. London: Routledge.

Robinson, J. (2005). *Ordinary Cities*. London: Routledge.

Sassen, S. (1991). *The Global City*. Princeton: Princeton University Press.

Scott, A. J., Agnew, J., Soja, E. W., and Storper, M. (2001). Global city regions. In Scott, A. J. (ed.) *Global City Regions*: *Trends, Theory, Policy*, pp 11 32. Oxford: Oxford University Press.

Smith, D. A. and Timberlake, M. (1995). Conceptualising and mapping the structure of the world's city system. *Urban Studies* 32, 287 302.

Relevant Websites

www.citymayors.com/features/quality survey.html
City Mayors.

www.lboro.ac.uk/gawc
Globalization and World Cities Research Network.

Urban Planning and Human Geography

J. Dodson and B. Gleeson, Griffith University, Brisbane, QLD, Australia

Introduction

In Western nations, modern urban planning as a distinct discipline had its origins in the development of the industrial city during the nineteenth century. The evolution of planning as a set of policies and practices for the development of cities has relied heavily on the insights of social and spatial science. The 'science' of planning has been largely derivative of thought from other disciplines, though the endemic scholarship of planning has grown in strength since World War II as dedicated planning programs were established in Western universities. In English speaking countries, human geography has provided urban planning with a substantial proportion of its intellectual insights since the late nineteenth century. The relationship between human geography and planning is perhaps best conceived through a 'structure–superstructure' metaphor. Geography has long provided the underlying structure of concept and method for planning, and evolving geographic thought has played a key role in the development of planning theories and practice. Geography's intellectual importance to a discipline frequently dominated by practical and professional imperatives cannot be doubted. For its part, geography has not tended to instate planning thought and practice as a core concern, though leading geographers, such as David Harvey, have had much to say about planning ambition and purpose. There would be much to gain – both for planning and geography – if spatial scientific thinking were to more closely analyze urban policy and practice. Planning is, as Patsy Healey would have it, a 'spatial strategy' that plays an important role in shaping social geographies, especially human settlements.

Modern Urban Planning – Origins

The 'age of discovery' expanded the horizons of Western human geography through the broader global extension of trade, scientific exploration, and political empires. The widening social scientific and geographic lens fixed on new social and economic phenomena. In particular, geographic analysis turned its focus to the rapidly developing nineteenth century industrial metropolis. The engineering marvels and civil monuments of the industrial city were a source of both wonderment and fascination to the public and scholars alike. The urban working class had burgeoned during the nineteenth century as the acceleration of industrial capitalism combined with changes in rural social structures brought rapid population growth in European cities. Much of the teeming new industrial population was the source of mystery, obscurity, and enthrallment for social science. Large tracts of the early industrial cities were foul places, where extremely poor and often undernourished populations dwelt in overcrowded, dilapidated, and unsanitary housing conditions. Social investigators who sought to understand the very rapid processes of urbanization and the emerging new urban forms and social patterns began to add numerical and spatial components to their analyses as they ventured into the dark slums and laneways.

Most human geographic thinking in the mid nineteenth century was dedicated to understanding the role of physical environmental processes in human societies. The majority of research was focused on biophysical processes at the regional scale, rather than on the urban realm and its murky spaces and places. Consequently, the initial investigators of urban environmental and social conditions were not necessarily formally trained human geographers. Many were middle class amateurs or lay persons who, with professional 'slum journalists' like London's Henry Mayhew, investigated the conditions of the industrial city and reported them to the wider public through novellas, pamphlets, public lectures, newspaper accounts, and more formal published books and reports. Gradually, statistical enumeration and physical description entered the slum watchers' analytical toolkits. The size, condition, and form of dwellings were a particular fascination, as was the concentration and composition of individuals and families within these dwellings. The simple Euclidean geography of population per unit of area, described as 'dwelling density', emerged as a key measure of urban conditions and could be easily calculated through simple dwelling surveys.

Much of this fascination was driven by public health concerns. The early industrial city lacked basic urban infrastructure for clean water supply and waste removal. Contagion was a pervasive risk with epidemics of diseases such as dysentery and typhoid also afflicting the middle classes. The industrial city was the first modern 'risk society' where environmental threat ignored class boundaries and could not be contained spatially. Under prevailing medical beliefs, such contagion was thought to be due to atmospheric 'miasmas' produced by effluent, wastes, and stagnant air and water. Dense living environments were also considered to produce degenerate moral behavior among their inhabitants. Medical geog raphy and allied public health knowledge emerged as the statisticians cataloged and mapped the outbreaks of dis ease and surmised connections to urban social and physical conditions. Medical administrators played a key role in the development of urban social statistics by drawing on their increasingly detailed geographic docu mentation of mortality and life expectancy. The linking of data on disease outbreaks with those on urban environ mental conditions was a key element in this new socio spatial statistical thinking. Increasing volumes of this so cial geographic reportage brought housing conditions and the plight of urban populations to the forefront of public thinking.

By the late nineteenth century, the stark human geography of the industrial city in many Western nations had stimulated a desire among politicians, government officials, and the broader public to bring about a reform of urban conditions. The central concern was the pro vision of sanitary and spacious housing for the masses at a level of rent that could be afforded on the relatively meagre wages earned by most factory workers. This in tention gave rise in turn to the modern town planning movement as public debate turned to identify solutions to the urban housing 'question'.

Models, Codes, and Formal Practices

The informal urban geographies produced by slum ex plorers and reformists had illuminated and attempted to explain (with doubtful success) the ills of the industrial city. In the wake of this endeavor, and propelled by new public reform agendas, modern town planning emerged as a codified set of practices which sought to impose order and safety in urban environments. While town planning began to take shape as a small but expanding set of regulatory provisions (building and infrastructure stand ards, zoning, etc.), it was shadowed by a larger conceptual imagination which sought to define good urban conditions comprehensively. From the late nineteenth century, the initial chroniclers of urban conditions were joined by new thinkers who proposed new, often utopian, models and ideas upon which the development of cities should be founded. Many were civically minded individuals who believed that deliberate and well reasoned spatial and social arrangements could generate new forms of human well being. Again a human geographic perspective was implicit in this new urban utopian imaginary. Getting space and place right could improve not just living con ditions but modern society as a whole.

The leading exponent of this new urban intellectual imagination was Ebeneezer Howard, an English planner, who proposed an idealized model for a comprehensively planned 'garden city' in which a network of town centers would be dispersed within a wide rural landscape, linked by express railways. Howard's vision encompassed not only the physical and spatial framework for the garden city but also promoted new social relations of land tenure, collective control of productive activity and a communi tarian societal structure. The 'garden city' concept was widely heralded in both the UK and Europe and to a limited extent in North America and Australasia during the first two decades of the twentieth century. The social dimensions of the model were typically set aside, how ever, as the physical and design dimensions were pursued variously by private, municipal, or state developers. The model's largest and most successful realization in the twentieth century was the Australian capital city, Can berra. In the UK Howard's advocacy resulted in the for mation of the Garden City Planning Association in 1899 which provided the template for the emergence of pro fessional planning associations elsewhere, such as the American Institute of Planners which was formed in 1917.

The emergence and widespread acceptance of the garden city model by the nascent town planning move ment coincided with an acceleration in the outward growth of urban development and infrastructure. This outward surge had commenced in the mid nineteenth century as elites and later middle classes moved from older, and usually crowded, urban cores into new lower density 'suburbs'. Initially, suburbanization occurred as wealthier groups used their economic power to escape the unpleasant conditions of the industrial city by pur chasing land on the existing urban fringe and taking advantage of coach or omnibus and subsequently rail services for travel into the central business and com mercial districts. Such transport was relatively expensive – only the wealthy could afford such services. The ma jority of the population had little option other than to get to their factory employment by foot or bicycle, which constrained the distance between home and work and contributed to crowded living conditions.

The activities of the public health advocates and the emergence of new town planning ideas accelerated the rate of suburbanization. In some British and European cities, municipal governments took on a planning role as land developers by building new rail transport networks to open up access to lower price ex urban land that could accommodate cheaper, relatively good quality housing. In the United States and Australia, the provision of access to fringe urban land for housing development often oc curred through a mix of public and private investment in transport infrastructure and land subdivision.

Human geography was not extensively involved in the early promotion of planning as a governmental activity. From the first decade of the twentieth century planning developed both as an activity of government and sub sequently as a distinct professional practice. Inter nationally, the development of public planning was uneven, emerging most strongly as a 'town planning' endeavor in Britain and its European colonies, such as Australia and New Zealand. By 1920 most jurisdictions in the English speaking world had instituted some form of legislation to govern the construction of housing and some were beginning to plan the basic layout of new urban development. This transition was anticipated by the British (largely English) experience in which slum clearance legislation such as the 1885 *Housing of the Working Classes Act* gave way to the more comprehensive *Housing* [*and*] *Town Planning Act* of 1909. Governments also began to establish commissions that were dedicated to assessing the need and desirability for greater levels of planning of cities during this period. Examples included Royal Commissions for the *Improvement of the City of Sydney* (1909) and for *Melbourne and Its Housing* (1913–15).

Throughout the first half of the twentieth century planning continued to pursue improvements to the social and environmental conditions within cities. The task of improving housing beyond mere safety and sanitary issues began to receive attention from architects who conceived new modernist designs for housing based on 'cellular' internal layouts drawing on mass construction techniques. The private motor car emerged in the early decades of the twentieth century and became increasingly important as a mode of transport. As road systems became increasingly subject to congestion from private motor traffic, transport planning gained in importance as an urban discipline deploying new roads in the form of the American urban 'parkway' and the German regional 'autobahn'.

The urban utopian imagination was at work again, this time infused with new technocratic and scientific ambition. New utopian urban schemes appeared in the 1920s. Some of them were impractical (and unlovely) but nonetheless influenced the course and the esthetic form of modern urban development. Others, especially those using the suburban form as substrate, exercised a much more general influence on planning and development. One design impractical as a whole, yet influential in technique, was Le Corbusier's notion of the 'radiant city' in which skyscrapers housing elite social groups were to be superimposed on an existing urban fabric. They would be served by radial motorways and rail lines connecting workers to garden suburb residences. Lloyd Wright's 'broadacre city' conceived of an urban form of broad grid of wide streets among which dispersed and decentralized low density development on wide lots would be distributed. Soria Y Mata's 'linear city' proposed an urban form based on an extended single longitudinal transport corridor along which high rise development would be sited. Few of these schemes were realized in their entirety, although Wright's conception was taken up in schematic form by much twentieth century suburban development, especially in the USA. Le Corbusier's physical architectural modernism exerted a heavy influence on urban redevelopment schemes from the 1950s onwards even if the social underpinnings of such urban modernism were set aside.

The City as Geographic Object

Human geography began to develop a more dedicated and comprehensive understanding of cities throughout the early decades of the twentieth century as urban populations grew and as the internal demographic and spatial composition of cities changed and expanded. European scholarship aroused much of this new attentiveness, through, for example, research into some of the smaller cities of Europe, such as Rouen and Grenoble, and with the appearance of the journal *La Vie Urbaine* in 1919. The focus of much early urban geography was on describing and comprehending individual cities as singular objects while also mapping and differentiating the functional roles of districts within them. By the early 1920s dozens of cities across the globe had been subjected to the functional mapping method.

Urban development processes accelerated in New World cities from the mid 1920s onward, such as in Chicago, which exemplified the stark urban modernity of the "skyscraper, the subway, the department store and the daily newspaper" – as Burgess described it – accompanied by continuous waves of immigration from Europe. This population growth wrought great expansion and transformation within these cities whose complexity and human variability often astounded and enraptured both geographers and planners. An initial research focus on urban form and function gave way to human geographers' comprehending these new urban processes as reflecting a form of 'human ecology'. Competition, symbiosis, evolution, segregation, and amalgamation were identified as the forces of urban change. This approach is most recognizable in the methods and concerns of Chicago based researchers such as Robert Park and Ernest Burgess, who sought to develop detailed socio spatial understanding of urban conditions and the lives of individuals and social groups. Social change was linked to changes in land use and residential activity as the city grew, expanded, and developed. Spatial zoning of city precincts for specific land uses had already been introduced by planners in the early twentieth century. Academic geography was now reflecting this idea in its analysis of urban change by identifying, delineating, and differentiating urban socio economic zones.

From the 1920s, geographers and planners together began to comprehend the city not just as a singular entity but as a focal point nested within a broader spatially interconnected region. Regionalism had been a fascination of geographers since the nineteenth century but had earlier focused mainly on the environmental conditions conducive to different forms and scales of human settlement. The expansion of cities and the intensification of their linkages with their wider regions opened the horizons of interest and scholarship of both planners and geographers. As cities grew planners aspired to shape and control their internal dynamics and also the relationships between cities and their surrounding periurban and rural zones. Regional planning and regional geography drew on the preceding regional geographic perspective in the initial decades of the twentieth century and combined it with newer conceptions of urban growth, including the capacities of new technologies, such as the automobile and telephone, to produce a distinctive regional view of urbanization, often with a strong anarchist flavor. Regional plans were proposed – such as that produced for New York in the late 1920s – which contained elements of the regionalist perspective including an emphasis on decentralization and decongestion of the city into regional subcenters, although as with the garden city concept the social elements of regional planning were largely set aside.

While the regional planning perspective captured the planning imagination in many jurisdictions, including in the USA, Australasia and the UK, regional geography evolved to become progressively less concerned with the initial social and political objectives of regionalism. Regional geography strengthened its focus on the scientific and methodological dimensions of regional spatial analysis. From the 1940s regional science became a prominent geographic subdiscipline influenced by economics, sociology, and planning. Regional science was typified by highly quantified and statistically driven approaches to urban and regional questions and problems, with particular attentiveness to economic functions and spatial efficiencies. Regional science also promoted the role of formal spatial models within geographic thought and deployed quantitative methods to derive and propose spatial mechanisms and processes by which regions developed and evolved socially and economically.

The formal scientific claims of regional science and the apparent validity and predictive power of regional science models proved highly attractive to urban and regional planning. There was no clearer instance of geographic science shaping planning thought and practice. The models' mesmerizing appeal lay in their claim to predict with high levels of certainty how a region might develop and change into the future. They seemed to offer a tantalizing – and often linear – certainty against which planners could shape policy and craft regional strategies to achieve or avoid 'scientifically' predicted scenarios. From the 1950s onward, this quantitative human geography also found collegial and intellectual affiliation with developments in the fields of transport engineering and transport planning.

After World War II, increasing economic prosperity, combined with technological and productivity improvements, enabled urban populations to purchase automobiles. These in turn competed with trams and buses on city streets, causing traffic congestion. Planners sought new methods to support the growth in automobile traffic by connecting residential zones and employment areas with high capacity road networks. Quantitative geographic methods proved highly suited to supporting this task by combining regional economic models with roadway network models to identify likely future patterns and volumes of travel demand. These regional transport analyses were typically formulated around a variant of a 'gravity model' based on Newtonian physics. In a gravity model the level of travel demand between two locations was calculated as a function of their relative population sizes and the inverse square of the distance between them. The scale and complexity of the calculations soon saw such models being undertaken on new computing technology.

The scientific mantle of quantitative geographic methods and the apparent physical analogy of the gravity model calculated by futuristic computer systems provided such transport analyses with an extraordinary degree of technical legitimacy within urban planning. The city could now be conceived as a regional 'system' in which economic and social functions, linked by transport networks, could be planned at a massive scale by technical and scientific experts. As early as the 1950s such 'predict and provide' methods were being used to conceive and construct massive freeway schemes which required the destruction or displacement of existing neighborhoods through which the roads would pass. Such plans emerged in the USA, starting with Detroit in 1955 and Chicago in 1959, but quickly spread to Canadian, Australasian, and some European cities. They were often complemented by urban regeneration schemes in which older, usually central, areas were characterized as 'blighted' and slated for demolition and redevelopment along modernist lines. In many cities inner urban residential zones were demolished and replaced with austere Corbusian 'towers in a park'.

Modernity in Question

Transport and regeneration schemes justified on rational, scientific grounds were rolled out across Western, especially Anglophone, cities in the 1960s. The scale of their impact on residents and communities soon generated vociferous resistance, however, among affected locals and eventually the broader public. The financial costs of the schemes were very high and meant that most metropolitan freeway plans would take many decades to complete. In many European cities this community and political resistance strengthened in the 1970s and re sulted in a shift away from 'predict and provide' freeway planning and toward much greater emphasis on public transport and nonmotorized travel. In most Canadian and some Australian cities 'freeway revolts' also emerged in response to such transport schemes and by the 1970s many of plans for extensive freeway grids had been shelved indefinitely. Many regeneration programs had also provoked community resistance and generated broader disrepute. This was frequent because the sup posed social and economic expectations of some of the early examples, such as the 'Pruitt Igoe' towers in St Louis, were never realized and, indeed, produced depredations of a new and entirely modern kind.

While often bitter and hard fought, the urban strug gles over modernist planning and transport schemes were not the only urban battles engaged in the 1960s. Broader questions of race and ethnicity, gender, social class, and environmental quality had emerged not only in social science but across the Western political sphere. When linked to urban questions, these new social scientific concerns provided stimulus to a new wave of critical urban scholarship in which progressive planning per spectives and politically honed human geographic ana lyses became closely intertwined. This scholarship sought to comprehend the enormous institutional, pol itical, and ideological power of 1960s modernist urban planning, its social consequences, and its provenance in scientific quantitative geography. David Harvey's work on social justice and the city exemplifies these efforts.

Many critical human geographers in the 1970s es chewed formal quantitative spatial scientific methods, regarding them as captive to social forces underpinning and driving the urban spatial expression of capitalist modernity and its faults. The process by which 'produced urban space' was created under capitalism became the concern of many urban geographers. (This critical interest in the production of space was later broadened by at tention to other formative social modalities, such as patriarchy and racism.) New social scientific methods were introduced which called in to question the object ivity of quantitative geographic science and sought to reveal the ideological basis for contemporary urban an alysis and urban planning. Much of this effort drew on new interpretations of the nineteenth century analyses of capitalism developed by Karl Marx and sought to update these for the urban conditions of the late twentieth cen tury. Capitalism and the state became key points of en gagement for human geographers and planners seeking to understand the social and economic forces shaping cities.

The new found fascination with capital and space that developed among human geographers' during the 1970s and 1980s also coincided with profound shifts in the organization of cities. Western industrial capitalism had faced structural problems such as labor process rigidities and constrained markets which limited profitability from the 1960s to the 1980s. From the 1970s, firms across the West began relocating production to exploit cheaper labor in less industrialized nations while also seeking new global markets for their products. In the West rising in comes produced new urban consumption practices while architects and urban designers began to move away from strict modernist style toward more varied and eclectic urban designs. Social scientific critiques of modernist urbanism, industrial restructuring, and new consumption practices were heralded in the 1980s by urban geog raphers such as Michael Dear as representing a new 'post modern' era in which analytical confidence in conventional – often universal – notions of science, social structures and the formal composition of cities would be re inscribed with uncertainty, contingency, and ambigu ity. By the late 1980s Los Angeles was lauded, briefly, in the human geographic imagination as the epitome of the emerging postindustrial postmodern city where extreme urban social inequality was mixed with new technologies of exclusion in support of a poorly planned, disorganized if highly productive form of advanced capitalism.

Planning and Doubt

While many human geographers and planners were pondering the ability of spatial social science to com prehend the reconfigurations of late capitalism, urban planning itself experienced profound changes from the 1980s onwards. In Anglophone nations doubt over the capacity of modernist master schemes gave way to a marked shift toward greater use of market processes by governments, resulting in planning becoming less dir ective and more managerial. Urban services and infra structure were reconceived as private consumption commodities rather than public goods and city planners were urged to become more 'entrepreneurial' to attract spatially footloose development capital. Regulatory frameworks in many jurisdictions were reconstituted to emphasize the role of the private sector, albeit sometimes accompanied by greater environmental attentiveness. The problems of modernist blueprint planning at the spatial scale of the city or region gave way to an emphasis on smaller, local scale 'place making'. Place making has frequently sought to rework urban precincts abandoned under industrial restructuring. These shifts are ex emplified by the international proliferation of urban 'docklands' redevelopments which have reconfigured spaces such as precincts of consumption rather than production.

Planning thinkers and theorists have however found it difficult to negotiate the intellectual challenges created by the new urban regimes and landscapes of post industrial cities. There has been a shift away from conceiving planning as a means of achieving particular social outcomes in favor of identifying the right set of practices according to which planning should be undertaken. In many jurisdictions planning to achieve a specific physical or spatial arrangement has given way to concerns with 'communicative' methods through which the desires and needs of diverse urban communities might be negotiated and mediated. This shift has also been accompanied by a pronounced disdain on the part of many planners for older concerns with urban structure and its distributive social and economic effects. Planning thinkers have increasingly drawn on continental social theories, with the work of Habermas, Foucault, and, latterly, Deleuze featuring prominently in many texts. Geographers too have drawn on these areas of thinking to comprehend cities with the 'governance' of urban areas and associated issues of 'governmentality' assuming prominence in scholarly thought. In terms of spatial metaphor and ambition, the latter period has been signified by a transition from concepts of structures to notions of networks and nodes organized by markets rather than the state.

New Horizons for Planning and Geography

The first decade of the twenty first century has seen new challenges, and questions emerge for urban planning and human geography. Geographers are continuing to grapple with the conceptual challenges thrown up by contemporary urban spatial and social change yet have also found many new territories and dimensions for investigation and analysis as cities are transformed and reconfigured by accelerating global flows of capital and people. Rather than determined and structured by capital, some geographers now see cities as constituted by multiple networks of social and economic actors that are also splintering under an increasingly fluid modernity. New urban systems of unprecedented demographic and spatial complexity are emerging as the globe's inhabitants urbanize. By 2008 the majority of the world's population is predicted to be urban. Older models of city development derived from Western intellectual traditions that have formed the prevailing body of geographic and planning thought are now charged with the task of explaining, charting, and responding to a panoply of new urban structural and social formations being created in cities beyond the traditional foci of Europe and North America. The task of widening the scholarly urban geographic and planning horizon to encompass cities throughout the globe has been underway since at least the mid twentieth century. This imperative hastened after the 1970s under conditions of globalization, and must accelerate as shifts in global political economic and cultural power mean that the economic 'periphery' of modernization theory increasingly usurps the 'core'.

Planning has struggled to cope with the urban transformations wrought over recent decades. While geography is charged mainly with the task of comprehending urban for, urban planners are expected to devise means for designing and managing cities. There is a spreading appreciation that the future management of urbanization will occur under conditions marked by not only fragmentation of public and private realms but also severe environmental constraints, including unreliable water supplies, the threat of climate instability, and increasingly insecure urban energy supplies. The scale of these challenges, even threats, suggests that planning will assume renewed importance in the quest for urban sustainability and security. To play this role and meet renewed expectations, planning will need to become empowered and emboldened after three decades of timid intervention in cities and growing uncertainty of scholarly and professional purpose. There are some indications that planning is responding to the challenge of the times, carried by some renewal of support in its wider institutional and political bases. In Australia, for example, the past decade has seen a revival of metropolitan planning in all the major capital cities. These halting steps have been criticized by urban geographers who have charge that the new plans risk failure because they have inadequately comprehended the character, scale, and pace of contemporary urban change. Planning might retort that human geography itself now struggles to provide a comprehensive framework for understanding contemporary urbanization let alone identify socially commensurable points of policy and market leverage upon which new urban plans might hinge. While human geography and urban planning retain a close intellectual allegiance, there remains extensive opportunity for renewed engagement and collaboration in comprehending and acting upon the urban world that is constantly being (re)created.

See also: Housing; Regional Geography I; Regional Geography II.

Further Reading

Dear, M. (1986). Postmodernism and planning. *Environment and Planning D: Society and Space* 4, 367 384.

Fincher, R. and Jacobs, J. (eds.) (1998). *Cities of Difference*. New York: Guilford Press.

Graham, S. and Marvin, S. (2001). *Splintering Urbanism: Networked Infrastructures, Technological Mobiities and the Urban Condition*. London: Routledge.

Haggett, P. (1965). *Locational Analysis in Human Geography*. London: Arnold.

Hall, P. (2002). *Cities of Tomorrow: An Intellectual History of Urban Planning and Design in the Twentieth Century*. Oxford: Blackwell.

Hall, P. (2002). *Urban and Regional Planning* (4th edn.). London: Routledge.
Harvey, D. (1973). *Social Justice and the City*. London: Edward Arnold.
Healey, P. (2006). *Urban Complexity and Spatial Strategies: A Relational Planning for Our Times*. London: Routledge.
Howard, E. (1946). *Garden Cities of To Morrow*. London: Faber and Faber.
Johnston, R. J. (1987). *Geography and Geographers: Anglo American Human Geography since 1945*. London: Arnold.
Le Corbusier (1929). *The City of Tomorrow and Its Planning*. New York: Payson & Clarke.
McGee, T. G. (1971). *The Urbanization Process in the Third World: Explorations in Search of a Theory*. London: G. Bell and Sons.
McLoughlin, B. (1969). *Urban and Regional Planning: A Systems Approach*. London: Faber and Faber.
Park, R., Burgess, E. and McKenzie, R. (1967). *The City*. Chicago, IL: University of Chicago Press.
Smith, D. (1996). *Third World Cities in Global Perspective: The Political Economy of Uneven Urbanization*. Boulder, CO: Westview Press.
Watson, S. and Gibson, K. (eds.) (1994). *Postmodern Cities and Spaces*. Cambridge, MA: Blackwell.
Wohl, A. (1977). *The Eternal Slum: Housing and Social Policy in Victorian London*. London: Edward Arnold.

Relevant Websites

http://www.gpean.org
Global Planning Education Association Network.
http://www.planninghistory.org
International Planning History Society.
http://www.tcpa.org.uk
Town and Country Planning Association.
http://www.unhabitat.org
United Nations Habitat Human Settlements Program.

Urban Policy

A. Cochrane, The Open University, Milton Keynes, UK

Glossary

Area-Based Initiatives Public policy initiatives targeted on particular geographical areas, rather than particular client groups.

Keynesian Welfare State The particular combination of economic and social policy that was dominant in most countries of the global North in the middle decades of the twentieth century, and in which welfare spending was mobilized to correct some of the inequalities generated by the market.

Multiple Deprivation The notion that if a household is disadvantaged in one way, then it is likely to be disadvantaged in several others, which reinforce each other; this is sometimes translated by analogy into the notion of multiply deprived areas.

Urban Competitiveness The notion that cities (as well as firms) are engaged in competition between each other for employment, investment, and development.

Introduction

At first sight, the task of defining urban policy seems a reasonably straightforward one. Governments and administrations of all stripes and in most countries of the world have been launching urban initiatives of one sort or another on a regular basis for the past 50 years or so. The term is simply taken for granted in much policy and academic discourse, so it is hard to avoid the conclusion that urban policy is no more and no less than the sum of all these initiatives, all either targeted on problems understood as 'urban' or taking advantage of opportunities associated with urban development.

The problem is that, however tempting it is to draw such a conclusion, and however accurate it may be as a summary of policy in practice, it does not help us in identifying what makes urban policy distinctive – what defines it as a practice or set of practices. Nor does it help us in assessing the significance of the initiatives or the ways in which they have changed over the years.

The nature of the 'urban' which is identified as the object of policy intervention has varied dramatically according to the policy context in which it has emerged. In some cases, the focus of attention has been on local communities or neighborhoods and sometimes it is placed on capacity building and self help; in others, the emphasis has been on policing and crime control or on the regeneration of housing stock and forms of 'gentrification'. Sometimes cities are imagined as dangerous places, whose populations are liable to threaten the social order and so need to be managed carefully. In some cases, by contrast, stress is placed on the importance of revitalizing urban economies, making cities competitive, because they are seen as drivers not only of national and regional prosperity but also of the individual prosperity and well being of their residents. Sometimes, the issue is a different one and the regeneration of physical and commercial infrastructure becomes a policy target in its own right, because it offers the prospect of making urban land economically productive once again, whether in the form of office, retail, or consumption development. Cities may themselves become reimagined through forms of place marketing and 'rebranding' as attempts are made to take advantage of their perceived status as cultural hubs.

This brief (and by no means comprehensive) listing of the ways in which the 'urban' has been incorporated into particular policy agendas also helps to confirm that there is little agreement about the nature of urban problems, or about the policy potential of cities. There does not appear to be a central continuing professional or statutory core around which urban policy might cohere in any of the contexts in which it has developed. Indeed, one of the difficulties in reaching a clear definition of urban policy is that it occupies a space that seems to be shared with other policy areas, such as housing, urban planning, and even architecture, each of which has some claim to longer professional legitimacy and a more continuous policy identity. It is not organized around any shared 'problem definition'. On the contrary, although each new policy initiative seems to emerge from the failure of the last, there is rarely any explicit reference back to explain (or justify) the latest twist.

Basic Features

Nevertheless, at a descriptive level, it is possible to identify some very basic features that underpin notions of urban policy as a distinctive sphere.

Perhaps the most obvious of these is that initiatives that attract the 'urban' label all have a territorial or area focus, rather than (explicitly at least) a focus on a particular client group, the provision of a particular service, or the distribution of benefits (whether universal or means tested) and of course, the territories are all understood to be within or to constitute urban areas. Although the urban has on occasion been drawn widely, for example, to include small towns and suburbs (particularly in response to 'new urbanism'), the problems of the rural poor or rural economy remain outside urban policy's remit. Urban social policy is concerned with places and spatially delimited areas or the groups of people associated with them. The size and shape of these areas may have varied significantly, but it is the territory that gives the policy its identity as 'urban' policy – whether defined through neighborhood, community, slum, zone, ward, inner city, *banlieue*, *favela*, city, or city region.

But urban policy also incorporates a distinctive approach to social policy, suggesting that it is able to mobilize a wide range of (innovative) policy tools and to have holistic ambitions that go beyond those of particular professional specialisms, precisely because of its area focus. This implies that urban policy practitioners are capable of taking an overview, developing policy initiatives that take account of the ways in which a range of social forces interact to shape the ways in which people live their lives in cities. Instead of cutting social life in accord with fields of professional expertise (planning, housing, social work, education, health, waste disposal etc.), urban policy promises the ability to see how the various services overlap, interconnect, and work with (or against) each other. It also offers the prospect of understanding how they relate to other development processes, from business and employment to tourism and cultural facilities.

This tentative set of understandings remains tentative, however, and urban policy remains an area in the process of construction, constantly subject to challenge and reimagination, not least because in practice different 'urban' policies define their areas or territories differently. Although an area focus provides a useful starting point for analysis, it also masks a very wide range of policy initiatives, concerned with a dramatically different definition of the problem faced either 'by' urban areas or by those living in them. This may – ultimately – be true of all areas of social policy, but urban policy is particularly fluid – a policy area in search of both 'problem definition' and tools of policy intervention.

Moving beyond the Keynesian Welfare State

There is a long history of policy intervention targeted on those living in cities – for example, in the form of slum clearance, in various policing initiatives, in town planning and design, as well as in sanitary reform and municipal endeavor. But a key entry point for understanding the development of urban policy as a practice in its own right lies in the responses made by the US federal government to the Civil Rights movement and then to the urban riots that cut through the African–American parts of major US cities toward the end of the 1960s. The histories of urban policy since then are complex and uneven, but it is possible to identify two major strands that are interwoven through its development. The first relates to the role of urban policy in the reimagination and restructuring of the national welfare states of the global North and the second, to the rise of urban policy as a global phenomenon expressed both in the policy initiatives of global agencies but, probably more important, in the new global common sense of urban competitiveness.

The origins of urban policy as a distinctive policy field are best understood as early responses to the perceived crisis of the Keynesian welfare state in the countries of the global North. In the US, urban policy represented an attempt to bring the previously excluded black urban populations into the political mainstream, offering them a share (albeit a subordinate one) in the 'Great Society', while in Britain it was presented as a means of positively managing a much smaller 'immigrant' population. In both cases, it was a response to the failure of the post war boom and the welfare states built around it to solve the problems of urban poverty and 'race', as well as the political grievances associated with that failure.

Equally important, it seemed to offer a means of resolving some of those problems without extending the welfare state as a universal system by shifting the way in which the problems were defined. Instead of targeting particular individuals or even particular groups (although in practice the urban became a coded reference to 'race'), geographical areas could be identified and targeted for intervention – an alternative to either the provision of universal services or individually means tested benefits. Spatially delimited clusters of deprivation (multiple deprivation) were identified, within which it was believed communities were weak or dysfunctional. As well as generating different targets for intervention, this approach also implied that no single professional specialism could tackle the multiple interacting forms of place based deprivation. It also supported a vision in which self help or community empowerment rather than the provision of services by the state offered a way forward. It reinforced the notion that what was needed was community action or the launch of area based initiatives which would enable (or 'empower') people to find some way of improving their collective lot with the help of new sets of professionals (such as community workers).

It has been strongly argued that there was a fundamental shift from what were described as the 'social pathology' or community based approaches of the early years to more structural, economy based initiatives in the late 1970s and early 1980s. Alongside this shift, the 'territory' of the urban also changed – now, it was the economic base of cities or large areas of cities that became the focus of attention since it was believed either that the problem of the cities was that they had been deserted by capital or that they provided object lessons in state failure so that what was needed was to open them up to the market. The language of regeneration came to be defined in terms of partnership, the reclamation of derelict land, and the provision of small factory units. Enterprise zones, in which planning controls and local taxes were waived, and urban development corporations, which opened up declining urban areas to the opportunities of property development, became emblems of the new approaches. Across the globe, waterfront development seemed to become the symbol of urban success, feeding into what became the excited bandwagon of urban entrepreneurialism and megaprojects, place marketing, cultural promotion, the creative industries, and 'urban renaissance'. By the start of the twenty first century, the importance of the 'competitiveness agenda', in which the priority for cities was not only to position themselves as attractive to business investment but also to provide an environment within which enterprise would prosper, was taken for granted.

The danger of a single narrative which points unequivocally in this direction, however, is that it fails to capture some of the other aspects of urban policy, as it has been popularly understood and actively pursued. So, for example, alongside this apparently inexorable shift toward the economic, it has been hard to ignore the stubborn survival (and revival) of discourses which emphasize the importance of community, whether tied up with notions of social capital or the moral basis of urban life which may be translated back into policy (e.g., in forms of 'neighborhood renewal'). Within most urban areas it is also increasingly hard to avoid the impact of forms of urban policy which focus on the management of crime and social disorder – or antisocial behavior. The 'broken windows' thesis famously asserts that the accumulation of apparently small acts of incivility can come together to provide an environment within which crime and more serious disorders may flourish. The policy consequences of such an understanding have included initiatives directed toward revitalizing community engagement as well as more explicit ones involving new (sometimes punitive) forms of policing and surveillance, as well as more informal uses of private and semiprivate security.

It is, of course, these shifts in focus that make it so difficult to define urban policy as a single object of study whose shared ambitions can neatly be summarized. The extent to which urban policy has so regularly and consistently been reinvented reflects wider shifts of the policy and political landscape but, in their different ways, all of the various shifts have highlighted the possibility of reimagining welfare in terms that are very different from those which underlay the settlements that defined the Keynesian welfare state.

Global Competitiveness

If urban policy emerged in the context of the crisis of national Keynesian welfare states, it has been transformed by the impact of, and has helped shape the politics of, (neoliberal) globalization. The new emphasis on urban competitiveness, whether expressed in the development of megaprojects, cultural initiatives, place marketing, or the provision of social and economic infrastructure, has been part of the process of redefining 'welfare' in terms that stress the importance of economic success for cities and access to employment for urban residents. These have been reflected in the emergence of forms of globalized and global urban policy, as cities (and sometimes the nations of which they are a part) seek to position themselves in accordance with the new policy environment and global agencies seek to influence that environment.

If one way of thinking about urban policy is to locate it within the wider crisis and consequent restructuring faced by the welfare states of the North, it is also increasingly clear that its development reflects the rise of quite different global agendas – and may even provide the basis on which a specifically global social policy is capable of being constructed. The core message of contemporary urban policy is an explicitly globalized one, precisely because it is rooted in an emphasis on notions of economic competitiveness, or what has been critically described as the 'new conventional wisdom'. According to this new global urban policy common sense (which some have labeled 'neoliberal'), because economic development is the basic underpinning of collective local prosperity, the key tasks of urban policymakers are: to foster urban competitiveness and enable the competitiveness of business within their territories; to build effective and market friendly forms of governance, with a strong role for nonstate actors in the delivery of services; and to maintain social cohesion, whether by fostering forms of community self discipline or, if necessary, through punitive policing. Environmental sustainability has also been identified as an aspect of this policy vision both because it makes cities more attractive to the knowledge workers and managerial strata they wish to retain or bring in, and because it is likely to underpin continued growth, enabling cities to move into a virtuous circle of self sustaining growth.

This language has also been reinforced as key international agencies have sought to develop their own urban policies or strategies for competitiveness. The message for the cities of the West and North is relatively simple – unless cities position themselves more effectively to be more competitive, they will inevitably decline, to be superseded by leaner and hungrier replacements (in India or China). In addition, the message for the (mega)cities of the Third World is equally clear, and distinctive, albeit still within an overarching framework of competitiveness. There, stress is placed on the need to remove any state based restrictions to the operation of the market, but also to build the social infrastructures capable of taking advantage of the creativity and informal networks that are said to be located in the urban slums.

In this context, the World Bank has explicitly sponsored the development of urban policies which stress the role of cities as drivers of wider development, despite currently being constrained by state regulatory policies and failures of governance. Instead of being seen as 'problems', it is suggested that if the constraints are removed, growth is possible – even in the previous problem sites (the slums) seen as the source of dynamism and entrepreneurial vitality. In this context, too, the coming together of economic and community endeavors is expressed quite explicitly, since the challenge is defined as being to find ways of actively drawing the urban poor into the processes of growth, 'empowering the people of the slums' through a process of slum upgrading.

A similar rhetoric informs the tentative emergence of a European urban policy. Urban policy is not one of the competencies granted to the European Commission by the European Union's member states (and explicit urban programs have relatively small budgets), but it is nevertheless possible to identify some features which are coming to define such a policy. Emphasis in policy documents is placed on the bringing together of wealth creation (through the building of economic competitiveness) and the promotion of social cohesion (and even environmental sustainability). Here the ways in which urban policy is framed in economic terms is particularly clear. The emphasis is placed on the ways in which what might traditionally have been seen as a social policy intervention – intended to help some disadvantaged group or individual – is instead justified because it is expected to deliver economic benefits, helping to make the cities (and so the countries) of the European Union more competitive in the global market. In other words, the old model of welfare is more or less explicitly inverted, so that urban policy has the task of providing the social infrastructure required to sustain the competitiveness of local business.

The extent to which these broad understandings and pressures have been able to construct a global template for urban social policy may remain questionable, but there can be little doubt that the process of policy development is now globally framed, rather than being best understood through a national lens.

Urban Policy beyond the Policymakers

Seeking for watertight (scientific) definitions of urban policy is ultimately a fruitless task, if only because it is always inflected through the dominant understandings – of the urban, of the social, of the economic, of social policy – of each particular moment. Urban policy is whatever it is said to be by those pursuing it at any particular time. Although often framed by particular (changing) understandings of the role of cities in social and economic life, urban policy itself rarely explicitly reflects on these, instead focusing on the here and now, delivering apparently practical solutions to immediate problems. Paradoxically perhaps, it is this that makes the study of 'urban policy' so useful as a way into debates about cities and how people live in them, since contemporary forms of urban policy both reflect and help shape how urban life is experienced. At the same time, precisely because it is not central to or deeply embedded in the traditional professions of social policy, it also offers a powerful means of interrogating some of the ways in which those professional understandings have been challenged in practice. In other words, excavating and exploring how urban policy is defined or understood at any particular time makes it possible to explore both how the urban is being defined and interpreted as well as the ways in which social policy is being reshaped and rethought.

There is nevertheless a danger of defining urban policy solely through the words of the policymakers, because there are also strong alternative visions of the urban, of what it is and what it might be. Instead of being somehow incidental to the achievement of other policy objectives, from these perspectives, it is the city itself which might be the object of policy, precisely because of what the city is, how it is defined by its residents, and by the social relations that cut across it. So, for example, Lefebvre suggests that urban politics should be defined by the demand for citizens to have "a transformed and renewed *right to urban life*." Others have argued strongly that, because cities are defined by the juxtaposition of, and interaction between, a wide range of different people through a series of overlapping networks of social interaction, they are also spaces within which opportunities for developing positive citizenship may be created. As a result, it becomes possible to imagine rather a different sort of urban policy – one in which what matters is to find ways of encouraging forms of interaction and negotiation which draw on the richness of the multiple identities experienced in the city – so that the city has the potential of delivering democratic engagement and emancipation. This opens up the possibility of an urban policy across which a range of interests and political engagements flow, connect, and collide as attempts are made to determine the basis on which we negotiate everyday ways of living together as well as to explore (and struggle over) issues of social justice.

See also: City Marketing; City-Region; Cultural Economy; Gentrification; Governance, Urban; Local Development; Neoliberalism and Development; New Urbanism;

Regeneration to Renaissance; Slums; Social Justice, Urban; Sustainability, Urban; Urban Growth Machine; Utopian Cities; Waterfront Development; Welfare Geography.

Further Reading

Amin, A. and Thrift, N. (2002). *Cities. Reimagining the Urban*. Cambridge: Polity.
Atkinson, R. and Moon, G. (1994). *Urban Policy in Britain. The City, the State and the Market*. London: Macmillan.
Barnekov, T., Boyle, R. and Rich, D. (1989). *Privatism and Urban Policy in Britain and the United States*. Oxford: Oxford University Press.
Brenner, N. and Theodore, N. (2002). *Spaces of Neoliberalism: Urban Restructuring in North America and Western Europe*. Oxford: Blackwell.
Buck, N., Gordon, I., Harding, A. and Turok, I. (eds.) (2005). *Changing Cities. Rethinking Urban Competitiveness, Cohesion and Governance*. Basingstoke: Palgrave Macmillan.
Cochrane, A. (2006). *Understanding Urban Policy: A Critical Approach*. Oxford: Blackwell.
European Commission (2003). *Partnership with the Cities: The URBAN Community Initiative*. Luxembourg: Office for the Official Publications of the European.
Hall, P. and Pfeiffer, U. (2000). *Urban Future 21: A Global Agenda for Twenty First Century Cities*. London: Spon.
Higgins, J., Deakin, N., Edwards, J. and Wicks, M. (1983). *Government and Urban Poverty. Inside the Policy Making Process*. Oxford: Blackwell.
Imrie, R. and Raco, M. (eds.) (2003). *Urban Renaissance? New Labour, Community and Urban Policy*. Bristol: Policy Press.
Kelling, G. and Coles, M. (1996). *Fixing Broken Windows: Restoring Order and Reducing Crime in Our Communities*. London: Martin Kessler Books.
Lees, L. (ed.) (2004). *The Emancipatory City? Paradoxes and Possibilities*. London: Sage.
Lefebvre, H. (1996). *Writings on Cities*, Kofman, E. and Lebas, E. (trans. and eds.). Oxford: Blackwell.
Marris, P. and Rein, M. (1972). *Dilemmas of Social Reform. Poverty and Community Action in the United States* (2nd edn.). London: Routledge and Kegan Paul.
Moulaert, F., Rodriguez, A. and Swyngedouw, E. (eds.) (2003). *The Globalized City. Economic Restructuring and Social Polarization in European Cities*. Oxford: Oxford University Press.
Rogers, R. (1999). *Towards an Urban Renaissance. Final Report of the Urban Task Force*. London: E. and F. N. Spon.
Savitch, H. and Kantor, P. (2002). *Cities in the International Marketplace. The Political Economy of Urban Development in North America and Western Europe*. Princeton, NJ: Princeton University Press.
UN Habitat (2003). *The Challenge of the Slums: Global Report on Human Settlements*. London: Earthscan.

Relevant Websites

http://www.citiesalliance.org
Cities Alliance is an international partnership organization which brings together cities, aid, and intergovernment agencies with a focus on the development of poverty reduction strategies or ‘slum upgrading’.
http://www.urban.org
The website of the Urban Institute in the US which brings together an extensive variety of research and publication on urban policy issues.
http://communities.gov.uk
UK’s Department of Communities and Local Government.
http://www.un habitat.org
UN Habitat is the United Nations Human Settlements Program and its website provides and brings together a series of new stories and publications relating to urban policy, particularly in the cities of the global South.
http://hud.gov
US Department of Housing and Urban Development.

Urban Regimes

M. Lauria, Clemson University, Clemson, SC, USA
R. K. Whelan, University of Texas, Arlington, TX, USA

Glossary

Fordism It is a regime of accumulation where the state helps stabilize patterns of production and demand (via Keynesian economic management), consumption through providing appropriate forms of service provision, fiscal policy, labor relations, and productive infrastructure.

Pluralism A theoretical understanding that politics and decision making are located mostly in the governmental framework, but many nongovernmental groups use their resources to exert influence. Lines of conflict are multiple and shifting. There may be inequalities but they tend to be distributed and even out.

Urban Regime A stable governing coalition of governmental, business, and civic interests that develops a coherent policy agenda in order to overcome the inability to govern as described by pluralism.

Regime theory has been, in large part, a response to, and a critique of, pluralist theories of urban politics that were themselves a response to and critique of elite theories of urban politics in the US. In particular, regime theory has responded to the New Haven school of pluralism. There were many events in the US during the 1960s and 1970s which could not be subsumed by the pluralist paradigm. These events included the rise of African American politicians, the increased importance of neighborhood organizations, the dominating nature of government bureaucracies in urban politics, and the growing influence of public sector unions in cities. Pluralist assumptions that were questioned included the view that power is widely distributed in communities that public officials and formal governmental bodies make decisions in a community, and that slack political resources in the system allow for social change. Thus, it should be noted that regime theory is US centric in both the empirical basis it was theorized from and the community of scholars it was engaging.

In the 1970s, political scientists increasingly explored the political economy. This exploration went in several different directions. The public choice school, emphasizing the free market economy and *laissez faire* approaches to governance, was applied to urban politics. Bish and Ostrom's short book on metropolitan governance, *Understanding Urban Government*, was an early indication. The Ostroms' workshop at Indiana University was the major center for published works on urban politics and service delivery, utilizing the public choice approach. At the same time, European scholars in Marxian political economy turned their attention to the city. Landmark works such as Castells' *La Question Urbaine*, Harvey's *Social Justice and the City*, and Pahl's *Whose City?* were published in the 1970s. In addition to later works by Castells and Harvey, other scholars made invigorating contributions to this literature. They include Katznelson and Saunders. Much of this work is nicely summarized in Katznelson's *Marxism and the City*. One sign of paradigm ascendancy is the publication of textbooks which reflect its ideas. Judd and Swanstrom's text is clearly constructed around a political economy perspective.

The early 1980s was a propitious time for the development of regime theory, for several reasons. First, the Republicans were in power, and conservative ideas were in fashion. National urban policy was epitomized by President Reagan's statement that urban dwellers could "vote with their feet" in choosing places of residence. The President's statement was a popularized version of the work of the public choice school of urbanists. Second, Marxian approaches to urban politics attained more academic respectability. Third, the publication of Paul Peterson's influential book, *City Limits*, motivated responses from urbanists with different views. Indeed, Stone and Sanders (1987) begin *The Politics of Urban Development* by saying "this volume can be considered an alternative to Peterson's *City Limits*: it departs in particular from his notion that a unitary interest guides the making of development policy." In contrast to Peterson's argument about cities' pursuit of development policies for economic reasons, Stone argues that politics matters.

As to be expected, disciplinary context shapes the respective approaches taken in this study of urban politics. By and large, the reputational studies (based on interviewing key economic, political, and social elite actors and asking who was influential in particular decisions, etc.) which support the elite theory of urban politics were conducted by sociologists. Similarly, political scientists conducted the overwhelming majority of studies which support pluralist theories of urban politics. However, geographers began using regime theory in their work and have made some very exciting contributions to the literature. In general, geographers have done a better job (than political scientists or sociologists) of recognizing and theorizing the extraregional political economic connections important for local regimes.

The original focus of urban regime theory was to abstract models of urban politics from historical epochs (that some) connected to changes in the structure of the world economic system. What is interesting about urban regime theory is not the typologies it inspires, but rather the structural features, the axes, or the defining focus that scholars use to describe local politics and abstract to typologies. Fainstein and Fainstein argue that the character of urban regimes is derived from two structural features of the political economy: first, that local government organizationally is dependent on property taxes for its fiscal solvency; thus, municipalities must maintain their revenues by enhancing existing property investments and attracting more such investment. At the same time, low income populations are a drain on local revenues. Second, a basic feature of a capitalist economy is the private control of production. The population at large is dependent upon private investment and profit for employment. Thus, the state must "facilitate accumulation in order to advance the material interests of its citizens" (Fainstein and Fainstein, 1983: 245–282). These structural features explain business dominance in local politics.

Fainstein and Fainstein, in their analytic summary of the case studies on the political economy of urban redevelopment in the US, describe three successive post World War II urban regimes: directive (1950–64), concessionary (1965–74), and conserving (1975–?). In the directive regime, the urban regime planned large scale redevelopment that was directly sponsored by the local government. The difference between this regime and the subsequent concessionary regime was not a change in business dominance but rather the addition of socially and politically forced concessions made to lower class urban residents. The conserving regime is capital striking back: retracting concessions and conserving political and social control while maintaining fiscal stability in a tenuous national and world economy. Neil Smith later termed this more broadly as the Revanchist city. Regime succession, according to Fainstein and Fainstein, is dependent upon the character of political struggles (social movement pressures) and the extent of fiscal or economic constraint. The post World War II economic expansion in the US lasted approximately until 1973. The political pressures of the civil rights movement and the welfare rights movement forced concessions on local political regimes, but it was the post World War II economic expansion that allowed these concessions both fiscally and politically. Thus, the succession from a directive to a concessionary regime was politically and economically expedient. The end of the post World War II economic boom forced an end to these concessions and spurred the development of the conserving regime.

Elkin (1987) argued that the structural features that define an urban regime stem from what he calls the "division of labor between market and state," specifically the definition of the respective prerogatives of the "controllers of productive assets" (capital) and public authorities, the organization of public authority, and the external relations of public authorities to private controllers of productive assets. While these structural features correspond well to the analysis of Fainstein and Fainstein, Elkin adds a closer examination of the bureaucratic politics of urban service provision. These structural features were politically constructed during the urban (governmental) reform movement of the late nineteenth and early twentieth centuries. They suggest three defining axes of urban politics: public/private growth alliances, electoral politics, and bureaucratic politics. According to Elkin, urban political economies are defined by the particular constellation of these axes, that is, the particular ways in which the various land interests and politicians ally, electoral coalition strategy, and the structure of bureaucratic service provision. He discusses three such constellations that vary temporally and geographically: pluralist (Northeast and Midwest, 1950s through early 1960s), federalist (Northeast and Midwest, mid 1960s through late 1970s), and entrepreneurial (post World War II Southwest) political economies and their respective regimes. Although the focus and thus the names are different, Elkin's first two regime types (pluralist and federalist) are essentially the same as the analysis of Fainstein and Fainsteins (directive and concessionary). It is in the third regime type that the two schemes depart, with Elkin's (entrepreneurial) being more pertinent to the 1980s and early 1990s. Fainstein's subsequent scholarship would seem to indicate concurrence with our assessment. Harvey has argued that this entrepreneurial form is no longer confined geographically.

Responding to this periodicity, urban political scholars began to indicate that the cities they study do not demonstrate the expected constellation during the specified epoch. In other words, there is more geographical and temporal variation than the Fainsteins or Elkin had suggested. The various articles in Stone and Sanders exemplify this variation. This geographic variation was interpreted by some to suggest that the connections between urban regimes and structural changes in the world economy were extremely complex if not spurious. Stone does not suggest that these connections are spurious, but rather he assumes them to be mediated by local political and economic actors. Stone, in his analytic summary of his edited book, developed a new typology of urban regimes that used governing coalition structure and development outcomes as their defining feature: corporate, progressive, and caretaker.

In his second book on Atlanta, Stone fully explicates his version of regime politics. It is through this source that one clearly sees the theoretical roots of Stone's ap proach. The basis and focus of Stone's approach mirrors Banfield's *Political Influence* albeit less formally and his torically richer. Banfield's main problematic was that since government and economic power were so frag mented in metropolitan areas (the pluralist assumption), it was a wonder anything was ever accomplished in cities like Chicago. So Banfield's goal was to uncover how this formal decentralization of power (again political and economic) was politically overcome (temporarily cen tralized) to achieve a specified end. Banfield used a microeconomic investment model of human behavior to explain how actors used their various forms of influence to develop political coalitions that either promoted or stymied large scale development projects with the ul timate goal of accumulating more political power. His descriptive generalizations stressed the importance of organizational actors in development politics. Un certainty played a major role in actors' evaluations of the costs and benefits of taking a political position, exerting influence, or maintaining informal semiautonomous partially centralized structures of influence.

Stone promotes a social production model that is based on a similar question, "how, in a world of limited and dispersed authority, actors work together across in stitutional lines to produce a capacity to govern and to bring about publicly significant results." Stone's coali tion building process is informal, built on the uncertainty of the various forms of influence, and his actors are also concerned with the cost and benefits of maintaining those coalitions. According to Stone, "Elite power in Atlanta is therefore constrained, not so much by the countervailing power of others outside the coalition as by the main tenance needs of the governing coalition itself." Fol lowing Banfield, Stone focuses internally on local coalition building, theoretically neglecting the con nections to the larger world economy.

Regime formation and transition are based on the stochastic microeconomic investment calculations of in dividual actors involved in coalition building and main tenance. Connections to external (state and national) politics and the larger world economy are only con sidered insofar as they affect those individual actors' calculations. This exclusion began what some have called a volunteerist return in urban politics (witness Stone's attempt to distinguish his work from pluralist theory). As indicated above, Stone does not deny these structural constraints; in fact he assumes them, but by focusing solely on the local individual and organizational actors involved in coalition building he risks slighting or neg lecting to evaluate the respective effects of those con straints. Consequently, the abstraction of theoretical insights becomes confined to behavioral microeconomic, and possibly pluralistic, explanations of the social pro duction of cooperation and political coalition building.

Much of the work on urban regime theory is based on urban politics in the United States. Stone's Atlanta and Elkin's Dallas typify the business dominated regime. Turner's and Kerstein's work on Florida cities also ex emplify business dominated regimes. The progressive regime is less common, but San Francisco, as character ized by DeLeon, has many elements of a progressive regime. Clavel's study, although not explicitly utilizing a regime framework, indicates other cities that may rep resent progressive regimes. Kalamazoo, Michigan, as observed in Stone and Sanders' case study, personifies a caretaker regime, while Whelan, Young, and Lauria focus on racial politics and attempts at regime transition in New Orleans (caretaker to corporate). Finally, Ramsey uses an urban regime approach to analyze the politics of economic development in two small towns in Maryland. She places a greater explicit emphasis on the norming role of culture in explaining the politics of economic development. In so doing, she unveils a newly recog nizable type of regime, one deeply embedded in a culture of subsistence.

In the 1990s, some urban regime theory researches began to focus on cross national and comparative study. Canada is the nation with the greatest similarities to US urban politics. Leo's study of Edmonton demonstrates the usefulness of regime theory to the study of economic changes in that city. Regime theory has been utilized in studies of several UK cities. Despite major differences in central government role, and the structure of subnational governments, students of British politics have applied regime theory to some advantage. DiGaetano and Klemanski's study of Birmingham was conducted by two US academics, but recent work by Axford and Pinch, Harding, Peck, Peck and Tickell, and Stoker indicates the high interest among British urbanists Body Gendrot's study of the thirteenth arrondissement in Paris and Levine's research on Lille indicate that regime theory might be employed in the French urban context. John and Cole's analysis of Leeds and Lille proved fruitful also. Regime theory has been diffusing further: Low for Aus tralia, Vicari and Molotch for Italy, and Terhorst and Terhorst and van de Ven for the Netherlands. This dif fusion did not stand without Davies' critical review.

We are aware of a few major international comparative case study projects; DiGaetano and Klemanski com paring Boston and Detroit in the US with Brimingham and Bristol in the UK takes the most explicit regime theory approach. They argue that the structure of the governing coalition determines the governing agenda. Regime stability and/or transition is the result of a combination of contextual conditions (the political economy of the city and its relationship with the global economy, local institutions and political culture, and specific leadership in local decisions). Sellers comparing cities in France, Germany, and the USA also argues that local politics matters and utilizes an empirical methodology similar to urban regime theorists. Finally, Savitch and Kantor comparing ten cities throughout North America and Western Europe use many of the insights of regime theory but take a larger political economic perspective. Analytically they focus on the bargaining resources that cities bring to the international marketplace. Subsequently, they provide an excellent summary of the methodological difficulties of comparative analysis and their attempts to surmount them.

Until recently, most published studies using regime theory have used a case study approach. Some of these are comparative case studies – either in the sense of comparing more than one city or in the sense of examining a city at different points in time. As far as methodology, we believe that analysts of urban politics proceed in similar fashion, whether they are proponents of regime theory, pluralist theory, or elite theory. That is, analysts read the newspapers of the city they are studying, examine relevant documents, read previously published material, etc. Moreover, they conduct interviews, which generally fall under the rubric of elite interviews. In the appendix to *Regime Politics*, Stone notes that he interviewed 97 people between 1982 and 1988 for his research. This is probably understated, since he interviewed many people in Atlanta during research for his earlier book. If there are differences, the differences are in the kinds of questions asked by the regime theorists and the framework they use in analyzing political situations. Regime theorists are empirical in approach. Certainly this is the case; although we have observed above that many of the regime theorists were part of the postbehavioral critique of social science in the late 1960s. Moreover, regime theorists have demonstrated evidence of nondecisions and agenda setting; it was said by pluralists that one could never observe such phenomena.

At the same time, some scholars have criticized the explanatory power of urban regime theory. Other scholars with this in mind have been attempting to embed urban regime analysis in more abstract explanatory frameworks. Lauria, and the scholarship within, focused on regulation theory. Here, the state and local politics are part of the mode of regulation within a particular regime of accumulation. According to Goodwin, Duncan, and Halford, the state is part of the mode of regulation of a Fordist regime of accumulation because it helped stabilized the patterns of production and demand (via Keynesian economic management), consumption through providing appropriate forms of service provision (via the various components of the welfare state), fiscal policy, wage relations (via collective bargaining and corporatist politics), and productive infrastructure. This Fordist mode of regulation and the state's role corresponds to the epoch that comprises Elkin's pluralist and federalist and Fainstein and Fainstein's directive and concessionary urban regimes. The recent economic restructuring and transition from a Fordist to a post Fordist regime of accumulation have been used to explain the rise of an entrepreneurial urban regime that facilitates privatization and the dismantling of collective services. Although regulation theorists argue that the transition from one mode of regulation to another will be uneven and spatially differentiated, this is explanation by caveat in terms of the differentiated structure of urban regimes.

One solution is to look for helpful explanations at a lower level of abstraction (see below). For example, since we know that economic restructuring occurs unevenly, we can hypothesize that differing external economic constraints may begin to explain the differentiation in specific urban regime structures and policies. Second, we can hypothesize that the structure of capital and representation of fractions of capital locally will affect the nature of the governing coalition, potential urban regimes, and their urban development strategies.

Lauria has argued elsewhere that the growth machine (Harvey Molotch's term for landed interests dominated development oriented regimes) is but one possible constellation of urban politics and that urban regime theory's more robust conceptualization allows for alternative governing coalitions. However, urban regime theory has critical shortcomings of its own. Its rational choice theoretic basis of causality, its truncated theoretical mechanisms, and its myopic preoccupation with the local have made it difficult for urban regime theory to explicate transformation.

Lauria has also argued that embedding urban regime analysis within a regulationist theoretic approach can provide a solution to these two shortcomings. So, what does regulatory context add to regime theory interpretation? First, it provides an opening for a nonbehavorialist approach. In so doing, and second, it stresses the importance of discourse and hegemony in governance. Third, it suggests the importance of including a focus on supraeconomic and extralocal connections of governing coalition participants. Finally, it provides a contextual understanding of the institutional arrangements and their role in governance. In a neoliberal mode of social regulation, the development of quasi public development corporations is the Schumpeterian component of the newly developing workfare state.

A few authors have begun to evaluate urban regimes comparatively through a secondary analysis of case studies extant in the literature. Kilburn uses qualitative comparative analysis to evaluate the contextual conditions (market and democratic political) under which different regimes emerge. de Socio argues that different configurations of business elite representing different economic activities are likely to produce different policy perspectives and thus affect the types of regime coalitions that evolve. He uses factor analysis to characterize the structure of interlocking directorates in order to evaluate whether regime orientation (from extant case analysis) corresponds to the different configurations of business elites he describes. Both authors argue that the case study method is essential for understanding urban politics but that more extensive methods are necessary to consolidate and interpret more broadly the contextual configuration these case studies provide. As one might expect, this movement between intensive and extensive empirical re search will continue to be an iterative process.

See also: Governance, Urban; Urban Growth Machine.

Further Reading

Banfield, E. C. (1961). *Political Influence*. New York: Free Press.

Clavel, P. (1986). *The Progressive City*. New Brunswick, NJ: Rutgers University Press.

Dahl, R. A. (1961). *Who Governs?* New Haven, CT: Yale University Press.

DeLeon, R. E. (1992). *Left Coast City: Progressive Politics in San Francisco, 1975 1991*. Lawrence, KS: University of Kansas Press.

de Socio, M. (forthcoming). Business community structures and urban regimes: A comparative analysis. *Journal of Urban Affairs* 29(3).

DiGaetano, A. and Klemanski, J. S. (1999). *Power and City Governance: Comparative Perspectives on Urban Development*. Minneapolis, MN: University of Minnesota Press.

Elkin, S. L. (1987). *City and Regime in the American Republic*. Chicago, IL: University of Chicago Press.

Fainstein, S. S. and Fainstein, N. I. (1983). Regime strategies, communal resistance, and economic forces. In Fainstein, S. S. *et al*. (ed.) *Restructuring the City: The Political Economy of Urban Redevelopment*, pp 245 282. New York: Longman.

Hackworth, J. (2007). *The Neoliberal City: Governance, Ideology, and Development in American Urbanism*. Ithaca, NY: Cornell University Press.

Harvey, D. (1989). From managerialism to entrepreneurialism: The transformation in urban governance in late capitalism. *Geografiska Annaler* 71B(1), 3 17.

Kantor, P. and Savitch, H. V. (2005). How to study comparative urban development politics: A research note. *International Journal of Urban and Regional Research* 29, 135 151.

Katznelson, I. (1981). *City Trenches: Urban Politics and the Patterning of Class in the United States*. New York: Pantheon Books.

Kilburn, H. W. (2004). Explaining US urban regimes: A qualitative comparative analysis. *Urban Affairs Review* 39(5), 633 651.

Lauria, M. (ed.) (1997). *Reconstructing Urban Regime Theory: Regulating Urban Politics in a Global Economy*. Thousand Oaks, CA: Sage Publications, Inc.

Lauria, M. (1999). Reconstructing regime theory: Regulation and institutional arrangements. In Jonas, A. & Wilson, D. (eds.) *Twenty One Years After: Critical Perspectives on the Growth Machine*, pp 183 205. Albany, NY: State University of New York Press.

Leo, C. (1995). How is global change mediated by local politics? Economic decline and the local regime in Edmonton. *Journal of Urban Affairs* 17(3), 277 299.

Logan, J. R. and Molotch, H. L. (1987). *Urban Fortunes: The Political Economy of Place*. Berkeley, CA: University of California Press.

Logan, J. R. and Swanstrom, T. (eds.) (1990). *Beyond the City Limits*. Philadelphia, PA: Temple University Press.

Molotch, H. L. (1976). The city as a growth machine: Toward a political economy of place. *American Journal of Sociology* 82, 309 332.

Peterson, P. (1981). *City Limits*. Chicago, IL: University of Chicago Press.

Ramsey, M. (1996). *Community, Culture, and Economic Development: The Social Roots of Local Action*. Albany, NY: State University of New York Press.

Savitch, H. V. and Kantor, P. (2002). *Cities in the International Marketplace: The Political Economy of Urban Development in North America and Western Europe*. Princeton, NJ: Princeton University Press.

Sellers, J. M. (2002). *Governing from Below: Urban Regions and the Global Economy*. Cambridge: Cambridge University Press.

Smith, N. (1996). *The New Urban Frontier: Gentrification and the Revanchist City*. New York: Routledge.

Stone, C. N. (1989). *Regime Politics: Governing Atlanta, 1945 1988*. Lawrence, KS: University Press of Kansas.

Stone, C. N. (1993). Urban regimes and the capacity to govern: A political economy approach. *Journal of Urban Affairs* 15(1), 1 28.

Stone, C. N. and Sanders, H. T. (eds.) (1987). *The Politics of Urban Development*. Lawrence, KS: University Press of Kansas.

Turner, R. S. (1992). Growth politics and downtown development: The economic imperative in sunbelt cities. *Urban Affairs Quarterly* 28, 3 21.

Vicari, S. and Molotch, H. (1990). Building Milan: Alternative machines of growth. *International Journal of Urban and Regional Research* 14, 602 624.

Whelan, R. K., Young, A. H. and Lauria, M. (1994). Urban regimes and racial politics in New Orleans. *Journal of Urban Affairs* 16(1), 1 21.

Urban Representation/Imagination

P. Kraftl, University of Leicester, Leicester, UK
J. Horton, University of Northampton, Northampton, UK

Glossary

Anti-Urbanism An enduring, far-reaching and deeply ingrained contemporary imagining – noted in Anglo-centric culture since at least the mid-nineteenth century – of urbanity as profoundly and inherently 'negative'. Anti-urbanism can be understood as part and parcel of the widespread contemporary imagining of rural places and lives as 'idyllic' in this cultural context.

Empire Cities Cities of huge significance to the political–economic power and identity-formation of empires. Often, empire cities used large-scale architectural schemes, exhibitions, and other forms of representation to assert their supremacy. Many contemporary cities – all over the world – still display signs of their imperial pasts in their landscapes and in popular imaginations thereof.

Flâneur A French word which roughly translates as 'loafer', 'idler', 'wanderer', or 'one who saunters'. The term was famously used in the French poet Charles Baudelaire's 1860 essay, '*The Painter of Modern Life*,' to describe the exploratory, somewhat detached, strolling of artists in search of inspiration in contemporary Paris. This notion of the '*flâneur*' figured prominently in the early twentieth-century work of the German philosopher Walter Benjamin. For Benjamin, '*flâneurie*' was both an analytical category (i.e., a new lifestyle constituted by modern urban life) and a methodology (thus Benjamin adopted observant strolling as a means of observing the easily overlooked, everyday characteristics of modern urbanity). Benjamin's notion of the '*flâneur*' is widely cited, and methodological variations upon '*flâneurie*' are numerous, in much contemporary – especially philosophically inflected – human geography.

Urbanity A complex term used to refer to the lifestyles, cultures, and politics of urban places. Rather than describing a set of easily measurable phenomena, urbanity denotes a more subjective interpretation of the conditions that render a place 'urban', even though individual cities may be hugely different from one another. Forms of representation/imagination are therefore key to the condition of urbanity.

Representing and Imagining the Urban: Two Points of Departure

1. What do you think of, as you read the following words: city; town; urban area; 'the urban'?
2. What comes to mind, as you read each of the following words: Athens; Baghdad; Beijing; Berlin; Cairo; Johannesburg; London; Los Angeles; Mumbai; Moscow; New York; Paris; Sao Paulo; and Sydney?

Introduction

This article critically evaluates human geographers' work regarding the representation and imagination of the urban. As may be apparent from consideration of the two questions above, this is by no means a neat, singular subject matter.

Take part of the first question, for example: what thoughts does 'the urban' conjure up? Maybe buildings, routeways, vehicles, machines, or technologies? Maybe iconic images or classic texts? Maybe industry, commerce, or culture? Maybe places to live, work, congregate, or consume? Maybe materials (concrete, tarmacadam, glass); maybe abstractions (light, speed, sounds, rhythms, feelings, smells)? Maybe histories, economies, communities, subcultures, practices, or ecologies? Maybe big issues; maybe fleeting, personal concerns, anecdotes, or memories? Maybe critical reflections about the status and usefulness of the term 'the urban' *per se*? Maybe all of the above, all at once; maybe a million and one things besides?

Or, take an example from the second question: what first comes to mind when one pauses to think about New York City? One could readily summon up manifold images and imaginings (whether or not they are 'true' is another matter; indeed, here, they unavoidably betray the authors' Anglo centric positionality). For example, the following come to mind: 'The Big Apple'; the Brooklyn Bridge; Woody Allen; George Gershwin; yellow taxis; steaming sewer vents and alligators in sewers; New York Central Park; the Empire State Building; the Statue of Liberty; 9/11; black and white photographs of Grand Central Station, and builders sitting on a girder; Ellis Island, immigration, and the American Dream; Harlem, the Bronx, Queens, Manhattan; Wall Street; abstract expressionism; the Velvet Underground; Broadway, bagels, delis; Spiderman; Susan Sontag, Sharon Zukin, Michel de Certeau's '*Writing in the City*'.

The preceding examples are intended to suggest just some of the diffuse, complexly interrelated issues and geographies relating to urban representation and imagination. The former example perhaps suggests the profusion of ways in which the representation and imagination of 'the urban' is open to investigation: at multiple spatial and temporal scales; mobilizing diverse intellectual and methodological techniques (from theoreticist to empiricist; from philosophical to applied); focusing upon diverse substantive interests. The latter example suggests the sheer number and diversity of representations and imaginations which come to be attached to particular cities (and thus ways in which 'urbanity' comes to be represented and imagined, in and of particular urban/spatial/historical contexts).

The work of human geographers has, variously, concerned itself with practically all of the diverse aspects of urban representation and imagination mentioned in the preceding paragraphs. This article seeks to provide an introductory overview of this work, by outlining three 'levels' of engagement between human geographers and the representation and imagination of urbanity. First, it introduces the significant ways in which human geographers have represented and imagined urban geographies, and have themselves contributed importantly to the popular representation and imagination of 'the urban'. Second, it introduces geographers' efforts to study 'the urban' as manifest in numerous styles of representation, and especially literary texts, cinema, the arts, and urban design. Third, it introduces geographers' engagement with theories and theorists of urban representation. In conclusion, the entry suggests some aspects of urban representation and imagination which may have been hitherto overlooked in the literature thus reviewed.

Urban Geographies as Urban Representations/Imaginations

In common with academics from most disciplines, urban geographers sometimes have the power to manipulate the ways in which we perceive our world. They may not only intervene in 'real world' towns and cities with policy recommendations and the like, but may also produce new ways of representing and imagining urban environments, which are then (differentially) adopted by lay readers.

Many chief histories of the discipline of geography position (perhaps problematically) eighteenth and nineteenth century explorers who – as part of European colonial expansion – charted lands previously unknown to Europeans as the earliest 'modern' geographers. In many of these regions (especially South America, China, and southern Asia), city forms and cultures were highly developed. Recently, urban geographers have documented the proliferation of these 'preindustrial' or 'precolonial' urban areas. Yet, in general terms, the early exploratory geographers represented such cities through two constructs. First, the explorers presented romanticized images of cities like Delhi – of oriental architecture, scented canals, colorful markets, and cultured (but nonetheless not yet modern) city life. Second, and conversely, those cities were equally imagined as places in need of colonial reconstruction and consequent inhabitation; as unfinished, harboring the chaotic, unhealthy, and disorganized poor; and therefore as spaces for the addition of colonial urban appendages. The complex and often problematic effects of both representations are still evident today in many 'majority world' urban areas (and representations thereof), and form the focus of ongoing critique from postcolonial geographers.

For generations of (especially Anglo American) school and university students, the work of the Chicago School provides the most widespread example of the intimate relationship between research and the production of urban representations. E.W. Burgess' concentric zone model of Chicago is perhaps the most recognizable emblem of urban geography, and has affected not only common perceptions of urban geography, but the whole discipline. Although most readers recognize that real cities do not correspond neatly to the models, they continue to impact upon our perception of urban places. For instance, many of us still have the tendency to imagine cities in 'zones', 'areas', or 'ghettos'; we imagine sharp divisions between zones along ethnic or economic lines; and 'inner cities' in the 'minority world' have been represented as the poorer and more degraded relations of the richer suburb.

Certain scholars (although perhaps not labeled 'urban geographers' at the time) affected urban representations that were particular to their own period of history. For instance, Lewis Mumford, who wrote on city form, function, and characterization during the first half of the twentieth century, held at the time an almost iconic status outside the academic study of cities. His ideas – along with those of his academic contemporaries – held great sway over professionals' representations of cities, and particularly over the emergent town planning profession. Taken together, these impulses (based on imaginative commentary as much as 'good science') profoundly affected the theorization, imagination, and consequent planning of urban areas in Europe and beyond.

Finally, more recent work by structural and especially Marxist urban geographers (the best known of whom is David Harvey) has led to more radical representations of urban space. Marxists have highlighted differences in the socioeconomic organization of urban spaces, and particularly the inequalities that characterize different urban areas, and that result from the workings of the capitalist economy. Although perhaps not so familiar to non geographers, Marxist urban geographers drew on critiques of exclusion and social injustice to fundamentally reorient geographers' imaginations and representations of urban spaces. No longer were towns and cities the spaces for pure, inductive (scientific) quantification; instead, the problematic social and spatial processes that constituted urban space needed to be uncovered and critically interrogated, so that more emancipatory – even revolutionary – solutions to social injustice could be imagined. Human geographers, at least, were able to represent urban spaces anew, and during the 1970s and 1980s in particular, students of urban geography were encouraged to imagine urban spaces as sites for critical intervention by academics. In some senses, it can be argued that these approaches were symptomatic of a period in European history in particular, when students were becoming radicalized and when developments in the global economy (with their urban manifestations) were under intense scrutiny. Hence Marxist representations of urban space extended beyond the ivory towers of academic research into protest movements and concrete political practices geared optimistically toward challenging global capitalism.

Geographical Research on Urban Representations/Imaginations

Drawing on theoretical and methodological approaches from a number of disciplines, geographers have critically evaluated the role of written and visual media in the formation of urban representations and imaginations. Geographers of different theoretical persuasions have demonstrated the key role of such media in constituting and reflecting societies' relationships with urban places. Four types of media have been particularly significant to such work: literary texts, cinema, visual art, and urban design. Many researchers do not concentrate exclusively on one of these media, but investigate the intertextuality of multiple representations of urbanity.

Literary Texts

For centuries, towns and cities have been represented and imagined in novels, poems, quasi scientific documents, political and planning discourses, and even 'lay' articulations of urban space. Texts such as these can be powerful tools for the dissemination of prevailing idea(l)s about urbanism.

The forms and processes of historical cities have proven particularly stimulating for textual and archival research by urban geographers, in a number of ways. First, contemporaneous debates about the ideal or planned city, especially in Athens, London, Paris, Rome, Vienna, and Stockholm, provided material for the sustained geographical study of the ideological, scientific, and philosophical foundations of modernity – a key concept in the development of urban places in Europe and beyond.

Second, in relation to debates about modernity, critical and Marxist geographers highlighted and problematized diverse representations of the squalor of working class areas of cities such as nineteenth century London and Paris. Reflecting bourgeois perceptions of the city, these areas were imagined simultaneously as fields of desire, disgust, and vice by novelists and social commentators – and later on, as areas ripe for cleansing and modernizing by planners and philanthropists. In particular, geographers have scrutinized the questionable 'moral geographies' of such representations. More recent research has moved from a rather narrow Marxian focus on class, to question the importance of other social categories (especially gender, ethnicity, and age) in the negotiation of these moral geographies.

Third, cultural geographers have charted the rise of the 'empire city' in and through the ideologies of architects and city politicians to construct specific built forms (especially in London), again through archival research. In fact, during the nineteenth century, many important cities throughout Europe contained buildings around which imaginations 'beyond' the city could be orientated: ultimately, a city could be placed at the center of its empire via a selective narrative based upon its own built form.

From the 1980s onward, geographers paid increasing attention to cities' representations of themselves, in particular through 'place promotion'. These and similar ideas concerning the 'urbanization' of society as a whole have attracted interest from urban geographers keen to explore the spatial pervasiveness not merely of urban form, but of urban lifestyles and attitudes – of 'urbanity'. Urban authorities have begun to market certain narratives and urban lifestyles (be they historical, playful, or environmental) in order to compete with other similar cities. Therefore, geographers have highlighted a discursive shift from imagining the abstract, modern, organized city (with a broad emphasis on social welfare), to representing a specific city, whose imaginary status has been achieved and then packaged (with an emphasis on globalization, privatization, and consumption). However, the coherence of such representations – which tend to assert the primacy of certain financial interests, and identities – has been criticized by geographers in cities as diverse as Singapore, Hong Kong, Glasgow, and Los Angeles. In particular, geographers' ongoing research on less 'formal' texts associated with alternative political movements – which range from political banners to radio broadcasts, and even to literal urban texts in the form of graffiti – has successfully exposed the partial, inconsistent, contestable, and divisive nature of place promotion and its problematic role in the contemporary (re)presentation of towns and cities.

Cinematic Texts

Since the 1990s, a growing field of research at the intersection of human geography, urban studies, and film studies has begun to investigate hitherto overlooked relationships between cinema and the imagination and representation of urbanity. Broadly, this work has had two emphases.

First, a number of human geographers have explored cinematic representations of cities. To date, this work has tended to focus on the following (perhaps limited; cer tainly overwhelmingly Americo centric) genres of representations:

Journeying to the city. Individuals' journeys from 'small town life' to 'the big city' were a common motif in popular Hollywood movies of the 1920s and 1930s. These journeys – and the often comic juxtapositions of 'small town' values vis à vis 'big city' life that result – are oft read as metaphors for the transformative onset of modern urbanity. Watch: *Safety Last* (1923); *The Crowd* (1928); *Mr. Deeds Goes to Town* (1936); and *Mr. Smith Goes to Wash ington* (1939).

Modernist cities. 'Abstract modernist ' and 'street ' films made in Weimar Germany are considered revolutionary in their signification of the rhythmic dynamism of modern cityscapes. Watch: *Berlin, Symphony of a City* (1927).

Noir cities. The 'low life' inhabitants and dark, corrupt, claustrophobic, urban milieus of 1950s film noir and gangster movies are widely interpreted as pivotal in twentieth century representations of the urban, and symptomatic of 'Cold War' era cultural and political anxiety Watch: *The Asphalt Jungle* (1953).

Affirmative cities. The opening sequence of Woody Allen's *Manhattan* (1979) – a black and white, widescreen rendering of iconic New York scenes, to the accom paniment of George Gershwin's *Rhapsody in Blue* – is widely cited as a loving, life affirming homage to the city. Likewise, a number of upbeat Hollywood musicals are interpreted as joyful, heart warming affirmations of urban life. Watch: *42nd Street* (1933); *On the Town* (1949); *An American in Paris* (1951); and *Singing in the Rain* (1952).

Unsettling cities. Alfred Hitchcock's cinematography is considered iconic in its rendering of urban geographies as profoundly unsettling, and rife with neuroses, anxieties, and guilty secrets. The capacity of urban life to induce alienation, psychosis, disgust, and horror is explored via films in which individuals are 'tipped over the edge', and resort to violent retribution. Watch: *Rear Window* (1954); *Taxi Driver* (1976); and *Seven* (1995).

Divided cities. Human geographers concerned with the exclusionary striations of urban spaces have often made use of films in which division along lines of 'race' or gang allegiance is an axiomatic device. Watch: *Do the Right Thing* (1989); *Bonfire of the Vanities* (1990); *Boyz 'N the Hood* (1991); and *La Haine* (1995).

Future cities. A significant body of criticism has been devoted to filmic depictions of (typically dystopic) future urbanity. *Metropolis* (1926) and *Blade Runner* (1982; 1992) are unfailingly cited in human geographers' engagements with cinema, and lauded for their prescient depictions of urban futures which are, respectively, repressively machinic, and corporately dominated/environmentally damaged/technocentric.

Second, elsewhere, work by cultural geographers has explored how cinematic visuality and urban represen tations/imaginations have been fundamentally inter twined and mutually constitutive. On the one hand, this has entailed empirical investigations into, for example, the major, iconic role of film production in the urban econ omies of Los Angeles (Hollywood), Mumbai (Bollywood), and Vancouver (location filming), or the development of movie going at various historical moments in modern European cities. On the other hand, drawing upon the work of the philosophers Frederic Jameson, Walter Ben jamin, and Jean Baudrillard, cultural geographers have contributed to theorizations of the relationships between geographies of modern urbanity and contemporary cin ema. From Jameson and Benjamin are derived, respect ively, the notions that cinema represented a space where city dwellers might pause for thought to make sense of the new stimuli, forms, and norms of modern urbanity, and that the forms, scenarios, and stimuli of modern urbanity fundamentally constituted the language of cinema. Bau drillard's conception of cityscapes as 'screenscapes' is oft quoted by geographers who, in interpreting his work (perhaps partially), have attempted to evoke ways in which urban streets and sights are frequently, already imbued with meaning, mythos, and *déja vu.* Thus, the boundaries between 'the discursive' and 'the real' (or 'reel') become somewhat blurred as a result of decades of cinematic representations of urbanity.

Art

Human geographers with an interest in urban represen tation and imagination have engaged with 'the arts', past and contemporary, in diverse ways, refracting different and changing levels of faith in representation, and es pecially the representability of urbanity.

First, and predominantly, visual artworks – especially paintings, photographs, and posters – have been ap proached, simply, as images and icons of urbanity. Here, human geographers have cataloged recurring motifs, subjects, scenes, and vistas in artworks which have come to be iconic and/or evocative of particular cities, at particular times. Oft cited examples include: Renaissance paintings of Florence and Rome; black and white photographs of New York's skyscrapers in the 1930s; the depiction of industrial and slum settlements in the work of Mexican social realist scene painters and muralists of the 1930s; and contemporary photographs of Las Vegas. A converse, common, use of artistic representations of cities is found in geographers' work regarding the

Anglo centric notion of 'rural idyll': here, widespread use is made of nineteenth century illustrations and paintings of cities – typically replete with visualizations of squalor, noise, danger, and disease – within attempts to under stand and historicize a prevailing anti urbanism in con temporary Anglo centric cultures.

Second, geographers have contributed to research on art movements and scenes, and their relationship to the representation and imagination of cities. For example, geographers have explored the ways in which urban centered art 'movements' (e.g., the Florentine Renaissance) and subcultures (e.g., pop art and abstract expressionism in 1960s–70s New York) come to be integral to the sub sequent imagination and mythos of 'host' cities, as well as constitute a set of novel representations in and of them selves. This work has contributed to understandings of how particular urban contexts, artistic movements, and broader contemporary idea(l)s and tropes of visuality are mutually constitutive.

Third, as part of wider attempts to locate critical re theorizations of urban representation and imagination (see below), a number of human geographers have en gaged with key urban artistic interventions. Here, the artistic and theoretical legacies of the 'Situationists' have come to be particularly important. Many geographers have found in this group's ideas and strategies (not least their site specific 'happenings' and methods of 'urban meandering', and, especially, their now iconic sub versions and sloganeering during the Paris riots of March 1968) the beginnings of languages and methods with which to re imagine, re present, and re enliven urban settings. In so doing, the representability of urbanity *per se* has been called into question, in numerous ways. Else where, a number of geographers have engaged with everyday, localized instances of urban artistic practice – such as graffiti ing, mural painting, or various forms of 'outsider art' – to think through the representation and imagination of the urban as manifest in forms without the urban canon and logics of art.

Urban Design

The field of urban design (incorporating architecture and planning) has been a key medium for imagining and representing cities. In fact, urban design may draw toge ther many of the above mentioned media. Historical geographers have explored – through textual analysis and archival research – how through time, urban designs have reinforced and produced dominant beliefs and ideologies. For instance, European discourses about (utopian) city form and organization were critical to representations and new imaginations of city life during the Middle Ages, especially the Renaissance. Herein, it has been possible to critique and complicate traditional histories of modernity, and trace the idea of the 'modern city' to well before the Enlightenment period.

The production of ideal, modern city plans perhaps reached a climax during the early twentieth century, when planners and architects such as Le Corbusier, Ebenezer Howard, and Frank Lloyd Wright imagined cities as efficient, gleaming, healthy, and green places. For geographers, such (more or less) concretized utopias were both the material expression and representation of new, modern attitudes to urban and especially suburban space, and (sub)urban lifestyles, in countries as diverse as Brazil, the USA, Egypt, the UK, and France. The plans, designs, maps, and images of this era (and in particular of UK town planning) were imbued with the power to imagine and produce ostensibly new, modern cities wherein the old (pre modern) era could be replaced. Despite their various successes and failures, the tropes of early twentieth century urban design continue to haunt contemporary urban landscapes.

Geographers have also attended to more contemporary urban design, exploring the power of design to imagine diverse cities in very different ways. First, urban design in contested spaces such as post apartheid South Africa and postcolonial Singapore has been related to continuing questions of race relations and morality, to demonstrate that planners' representations of urban space may enable – or disable – very particular conceptions of social justice. Second, 'heritage landscapes' in urban areas (such as gen trified and leisure orientated waterfront re developments) have been critiqued for reproducing dominant – often imagined – histories of particular cities: this trend has been most apparent in North America, Europe, and Australasia. For many commentators, the drive to selectively revitalize old industrial or working class areas has led to a prob lematic historicization and romanticization of those places, characteristic of capitalist competition between cities who vie for tourism and investment. However, similar processes have been observed more recently in Africa (for instance in the Zanzibar Achipelago), allowing researchers to critically evaluate the emergent role of globalization in the re im agining of local urban histories. Finally, the controversial but almost global rise of 'defensible city' design – in office districts, 'gated' residential communities, and shopping malls – has been well documented. Both radical and poststructural geographers have debated the relative merits and problems of urban districts policed by armed security guards, razor wire, closed circuit television, and controlled entrances. Herein, the city outside the walls is imagined as a landscape of fear and violence, an image perpetuated by marketing materials for such developments; inside, nos talgic, neotraditional urban design offers residents a re vived but largely symbolic, sanitized, and exclusionary sense of community which relies upon standardized rep resentations of historical cities which (perhaps) never existed.

Geographers' Engagements with Theorists of Urban Representation/ Imagination

There is no singular, coherent geographical theorization of urban representation/imagination. Rather, human geographers have engaged with a multiplicity of theorists of the urban (or, more typically, with multiple theoretical moments, keynotes, aphorisms, and motifs from their work). These engagements have fundamentally constituted human geographers' styles of envisioning and prioritizing what matters in relation to urbanity.

First, many geographers position cities as fundamentally important nodes in meta theorizations of global economic and political processes. Typically developing Marxian interpretations of capital(ism) and urban labor relations, and Weberian understandings of the relations between urban institutions and broader sociopolitical processes, historically changing urban representations and imaginations have often been read as manifestations of shifting international economic and political geographies. Accounts of nineteenth century Paris by Henri Lefebvre and David Harvey are typical of this representational style, approaching cities as spatial/historical materializations of the development of modernity, urbanity, and capitalism. Similarly, developing the work of recent social scientific theorists of global economics and urbanism, geographers have contributed significantly to the understandings, terms, and visualizations of the role of cities in relation to recent and ongoing developments in the capitalist world system. The prominent work of human geographers in mapping new and unprecedented forms of urban capitalism, and in articulating notions such as 'world cities', 'postmodernism', 'uneven development', and 'post Fordism', have come to be profoundly important in framing both academic and contemporary popular understandings and imaginings of the urban.

Second, cities are, in effect, envisaged by geographers as assemblages of spaces, and associated social geographies. Drawing, often, on notions of the 'public sphere', the 'public realm', the 'production of space', or 'habitus' and 'social capital', human geographers have contributed to the understandings and contemporary imaginings of the social spaces of cities. Here, geographers' accounts are typically theoretically inflected with profound pessimism or optimism in relation to urban social spaces. On the one hand, many geographers' representations of urban social life can be read as continuations of much longer running pessimistic discourses regarding urbanity, such as the 'fortress city' and 'the city of panic', and understandings of life in capitalist cities as characterized by, and constructed in relation to, continual dread and anxiety. Michel Foucault's notion of 'governmentality' and metaphor of the 'panopticon' are frequently mobilized in such representations to evoke the surveillance, anxiety, and alienation said to characterize urban life. On the other hand, and often in deliberate opposition to such negative accounts, a number of geographers have sought to develop alternative, more positive and hopeful understandings and representations of urban society. Here, more optimistic theorizations of the city as a key achievement and manifestation of civilization and an egalitarian space for the 'being together of strangers' are keys. Celebratory representations of cities as 'ethnoscapes' (where ethnic diversity produces new and vital forms of urban life and culture), as counterpoints to small town tradition and private sphere repression, or as places of liberation, resistance and/or self realization, have become significant points of departure for such optimism.

Third, in purposeful contrast to the 'bigger pictures' produced by the preceding styles an increasing number of human geographers take personal, everyday, subjective experiences as a point of departure for accounts of the urban. Here, the figure of the individual in the city has become commonplace and significant. Following (sometimes literally) in the footsteps of theorists such as Walter Benjamin, Michel de Certeau, and Roland Barthes – the experience and practice of 'walking' in the city is increasingly figured as a revelatory counterpoint through which it is possible to encounter and creatively and/or critically engage terrains, happenings, and issues which go unrecorded in many totalizing, 'top down' accounts of 'the urban'. Walter Benjamin's figure of the '*flâneur*', and accounts of strolls through early twentieth century Berlin and Paris, have become particularly widely cited in this context, as have situationist *dérives* through European cities.

Conclusion

Human geographers have been concerned with many and diverse forms of representing/imagining urban places. Drawing on a substantial range of media, this work can be generalized into three 'levels' of engagement, which this article has outlined. Nevertheless, there remain some nascent (and some as yet unexplored) styles of urban representation/imagination that warrant greater attention. First, most research in this area remains focused on the minority world. Re dressing the balance in favor of the majority world is an important step; however, as majority world cities develop apace, and become larger and more important than those in the minority world, geographers may need to apprehend completely new forms of representation/imagination. Second, despite a recent surge of interest in selected popular and vernacular culture, much research remains focused on elitist modes of representing/imagining the urban. In particular, there has been very patchy attention to the roles of diverse social groups (such as children, the elderly, tra velers) in shaping urban representations/imaginations. Finally (relatedly), geographers' current interest in the 'everyday' and in emotion is ideally suited to exploring the less spectacular, hard to describe, mundane ways in which urban space – and urbanity – is represented and imagined. Importantly, such individual perceptions of, and emotional responses to, urban places are not merely interesting research topics; they may also be potent political tools that go largely unnoticed, but which at the same time may be manipulated by those in power. In sum, this signifies that the (politicized) relationship be tween diverse forms and scales of urban representation/imagination still requires constant critical attention from geographers.

See also: Capitalism; Chicago School; Gentrification; Landscape Iconography; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Modernity; Situationism/Situationist Geography; Street Names and Iconography; Urban Architecture; World/Global Cities.

Further Reading

Al Hindi, K. F. and Stadder, C. (1997). The hidden histories and geographies of neotraditional town planning: The case of Seaside, Florida. *Environment and Planning D: Society and Space* 15, 349 372.
Bridge, G. and Watson, S. (eds.) (2000). *The Blackwell City Reader*. Oxford: Blackwell.
Bridge, G. and Watson, S. (eds.) (2003). *A Companion to the City.* Oxford: Blackwell.
Clarke, D. B. (ed.) (1997). *The Cinematic City*. London: Routledge.
Cuthbert, A. R. (ed.) (2003). *Designing Cities: Critical Readings in Urban Design*. Oxford: Blackwell.
Driver, F. and Gilbert, D. (eds.) (1999). *Imperial Cities: Landscape, Display and Identity.* Manchester: Manchester University Press.
Gold, J. R. (1997). *The Experience of Modernism: Modern Architects and the Future City*. London: E & FN Spon.
Gold, J. R. and Ward, S. V. (eds.) (1997). *Place Promotion: The Use of Publicity and Marketing to Sell Towns and Regions*. Chicago: Wiley.
Harvey, D. (2002). *Spaces of Capital: Towards a Critical Geography*. London: Routledge.
Knox, P. and Pinch, S. (2006). *Urban Social Geography: An Introduction* (5th edn.). Harlow: Pearson.
Miles, M. (1997). *Art, Space and the City: Public Art and Urban Futures*. London: Routledge.
Myers, G. (1996). Naming and placing the other: Power and the urban landscape in Zanzibar. *Tijdschrift voor Economische en Sociale Geografie* 87, 237 246.
Pile, S. and Thrift, N. (2000). *City A Z: Urban Fragments*. London: Routledge.
Pinder, D. (2005). *Visions of the City*. Edinburgh: Edinburgh University Press.
Soja, E. (2000). *Postmetropolis: Critical Studies of Cities And Regions*. Oxford: Blackwell.
Winchester, H. (1992). The construction and deconstruction of women's roles in the urban landscape. In Anderson, K. & Gale, F. (eds.) *Inventing Places*: *Studies in Cultural Geography* (1st edn.), pp 139 156, London: Longman Cheshire.

Relevant Websites

www.imbd.com
The Internet Movie Database.

Urban Village

G. Crow, University of Southampton, Southampton, UK

Glossary

Cosmopolite Middle-class, inner-city dweller attracted to live there by cultural vitality and diversity.

Gentrification Redevelopment of previously rundown, inner-urban areas brought about by in-migration of middle-class, owner-occupiers.

Peasant-Workers Individuals who combine urban industrial employment with agricultural production in the countryside where they live, a pattern produced by state socialist industrialization.

Rural–Urban Continuum The idea that community types can be located along a linear scale with rural and urban at the two extremes, based on the assumption that population scale and density determine social relationships.

Introduction

An urban village is an oxymoron, if one associates villages with rurality and if rurality is regarded as the opposite of all things urban. There is a long tradition in social thought of highlighting the contrast between villages and towns and between rural life and urban life. In this tradition, rural life is stable, traditional, and centered on established family and community ties, in contrast to urban existence which is characterized by a shallower way of life lived at a faster pace, reflecting dynamic commercial and technological forces that make social relationships fleeting and impersonal. The notion of urban villages, created and inhabited by urban villagers, is compelling because it subverts this rural–urban dichotomy in which types of social relationships are associated with particular sorts of place. The notion of an urban village is thus immediately engaging and challenging because of its paradoxical character, linking modern city life with deep seated patterns of supportive social relationships that are more readily identified with established rural arrangements. Urban villagers appear to have one foot in the city and the other in the countryside, or one foot in modernity and the other in a past world.

History of the Concept

The concept of the urban village has its origins in the work of the sociologist Herbert Gans. His study was based on ethnographic fieldwork among Italian Americans in Boston, USA. It portrayed the people of the West End of that city (prior to its redevelopment) as cohesive, family centered, and street centered, but with relatively few points of connection to the world beyond their ethnic urban enclave. Their way of life prized experiences as part of the group and continued to bear many of the hallmarks of the intense but restrictive social relations which had characterized the communities in southern Italy and Sicily from which their ancestors (and a few of the older residents themselves) had migrated. Gans coined the term the peer group society to describe this way of life in which the individual is dominated by the group. Group members are necessarily familiar to each other because of their shared social and physical environment. Gans' account emphasizes the importance of the poverty of this environment in which people come together because they all have to confront problems of low income and poor education, but urban villages have an ethnic as well as a class dimension as people seek to adapt their culture and traditions to the urban environment. In a related publication, Gans uses the term ethnic villagers rather than urban villagers, and it follows from this that urban villages may vary according to the originating culture of the migrants (as Gans suggested in his discussion of how Italian Americans compared to Americans from Irish, Jewish, or Negro backgrounds). Gans' article is more analytical and wide ranging than his ethnography of Boston, and in it he characterizes inner city villages in terms of the emphasis that they place on kinship and other primary groups. There are correspondingly few secondary group contacts, and formal organizations are weak, as are connections beyond the neighborhood. Gans also makes the point that there is nothing inevitable about an inner city area becoming an urban village; it may just as easily approximate to an alternative ideal type, the urban jungle.

Gans' work marked an important development on previous urban sociology because it sought to portray life in the urban village in a sympathetic way without sentimentalizing it. The particular target of his criticism was the attitude of planners and social workers who saw only social problems in inner urban areas that they perceived to be slums (but which were not necessarily perceived to be slums by their inhabitants). Gans offered an explanation to these professionals of why opposition existed to their urban redevelopment schemes. In his view, they embodied an ill informed, middle class bias in favor of suburban lifestyles and values that the inhabitants of urban villages simply did not share. Gans' analysis is strengthened by his frank admissions of how his fieldwork shook him out of his own preconceptions as an outsider. His methodology allowed him to learn about the people he studied by spending time with them as they went about their everyday lives. Gans moved on from Boston to research the new community of Levittown, New Jersey, which he also studied ethnographically. There, he discovered that life in suburbia did not conform to outsiders' conceptions of it any more than life in urban villages did, and on the basis of these findings he developed a critique of spatially deterministic explanations of people's ways of life. This argument challenged the previously influential idea associated with Louis Wirth that city dwellers' lives were made anonymous and impersonal by the size and density of urban populations. In doing so, he contributed to the demise of the rural–urban continuum and thereby helped pave the way for more nuanced, sociological accounts of both city and village life.

Gans was not alone in developing these arguments. The idea that migrants to cities continue to live according to the social norms that they (or their ancestors) brought with them from the countryside, can be found in studies at least as far back as William Thomas and Florian Znaniecki's classic study of Polish peasants' migration from Europe to America. The related recognition of the centrality of ethnicity to urban ecology was a central plank of the Chicago School of sociology which flourished in the first half of the twentieth century. By the time Gans was building on these ideas, other researchers were arriving at similar conclusions. Michael Young and Peter Willmott's analysis of family and kinship patterns in East London, in fact, predates Gans' Boston study and is mentioned at various points in the book. This is unsurprising, because it reported on fieldwork in another poor, inner city area characterized by a history of settlement by successive ethnic minorities. Urbanization had long ago swallowed up what were in earlier centuries distinct villages, but Young and Willmott claimed that these village identities lived on in people's minds. In another parallel with Gans, Young and Willmott highlighted the continued importance of close knit, family based community relationships in large cities, noted how this confounded the expectations that such locations would be characterized by social disorganization and lack of community, and offered a critique of well intentioned but sociologically ignorant schemes of urban redevelopment.

The Continued Relevance of the Concept

Gans' work can be interpreted as a lament for a doomed and disappearing way of life contained in urban villages. In the specific case of Boston's West End, the community of Italian Americans had literally disappeared by the time of the publication of his research, as the book's photograph of the area following its demolition testifies. The book treated this outcome as the product not only of the politically driven physical redevelopment of the area's housing, but also as the indirect result of broader social and economic processes. Notable among these processes were the growing opportunities for social and geographical mobility offered by the development of an affluent society and service sector employment, changes to the traditional order of gender relations, and the increasing intrusion into everyday life of the mass media. Gans' message is that the days of the community he studied were numbered well before the coming of the bulldozers. Yet Gans' work has continued to provide a point of reference for subsequent researchers because the idea of an urban village has continuing relevance. This is for two reasons, empirical and theoretical. The first is that subsequent studies have reconfirmed the importance of the rural cultures of origin of migrants to urban settlements, and this on an increasingly global scale. It has been noted, for example, that all but 5% of British Pakistanis come from rural origins. Many studies reveal that urban villages can continue to be found in a wide variety of settings. The second reason for Gans' continuing relevance is that he identified a number of functions performed by urban villages, of which mutual insurance against the hardships of poverty was just one. The congregation in a particular area of people of a shared ethnic minority status can help offset some of the discriminatory processes to which they are typically subject, and in the process support a positive sense of individual as well as group identity. To this it can be added that urban villages also make it easier to maintain connections with wider kin who remain in the migrant group's country of origin.

A good example of a contemporary study that is very much in the tradition of Gans' classic study (but which surprisingly does not make this fact explicit, despite the echo in its title) is Peggy Levitt's research. Based on research among migrants to Boston from the village of Miraflores in the Dominican Republic and among people who remained in this village, Levitt's argument is that migrants' lives in Boston are recreated around much that is familiar from their home village. Contrary to common sense expectations that migrants will leave behind the social world from which they have come and pursue assimilation into their adopted homeland, Levitt reported little evidence of people having to give up their existing connections. Rather she found it to be the case that transnational migrants integrate to an extent into the countries that receive them without having to jettison contacts with the countries they have left. The flows of monetary and other remittances sent back to their home village by migrants are vital to this continuing link with their place of origin, but political and social as well as economic factors are responsible for the recreation of familiar community relationships by the migrants. The cultural connections are so strong that it can appear that village life is transported more or less in its entirety by the migrants.

Levitt claims that the lives of the people she studied are replicated by many migrants to the USA from villages across Latin America and the Caribbean, and she suggests that there is no sign of any decline in the phenomenon of migration leading to the creation of urban villages in migrants' destination countries in the global North. Other studies support this argument. Fred Halliday's study of Yemeni migrants to Britain charts how they created urban villages in the industrial cities in which they settled, and reports the surprising finding that in these new community settings, the degree of their segregation from mainstream British tended to increase rather than decrease. Their urban communities were so self contained that they had the character of remote villages. A similar narrative is developed by Pnina Werbner in her account of Pakistani migrants to Britain, based on her research in Manchester. Arguing that migrants take their cultures with them when they travel to take up work opportunities in new settings, she recounts how cultural traditions were drawn upon as this immigrant population grew in the second half of the twentieth century and relocated south of the city center to an ethnic residential enclave. This enclave had a sufficient degree of concentration of Pakistanis that it might be called a ghetto, although (like Gans) she is keen to point out that ghettos are not necessarily slums. What developed there was a network of relatives and friends whose maintenance of shared cultural practices provided support and security in an environment that could be hostile. These arrangements helped immigrants and their families survive disadvantage and resist discrimination and racism.

An important aspect of contemporary studies is that they highlight the continuing mobility of migrants. Permanent return migration of immigrant groups is regarded as sufficiently rare for some writers to treat it as a myth, but although aspirations to return are not always realized, such language nevertheless overstates the case. Historical statistics indicate that as many as one third of certain waves of immigrants to the USA later returned to their countries of origin, and these include Italian migrants among whom would have been some of Gans' urban villagers and their ancestors. This is unsurprising given that the majority of these migrants did not regard their journeys as involving permanent relocation. Keeping open the possibility of return migration is an important aspect of the lives of migrants. Levitt reports that transnational villagers in Boston are reluctant to consider their move to be permanent, and most state an intention to return. The creation of urban villages by migrants thus serves the function of helping to keep this option alive far more than a strategy of assimilation would. Moreover, as global travel has become increasingly affordable, it has facilitated short term visits that allow social networks to be sustained through face to face encounters. A similar argument can be made that developments in communications technology enable migrants to combine elements from geographically distant cultures and thus feel at home in more than one location. People who are part of global population flows thus have something of a cosmopolitan character, and can continue to regard themselves as having a place in two cultures rather than having to choose between them.

Wider Significance

It is instructive that Gans' identification of five types of inner city residents put ethnic villagers into a separate category from cosmopolites. He identified the latter as people who are attracted to the city by its cultural life, and included among them groups such as students, creative artists, and professionals. In Boston's West End, Italian Americans made up 42% of the resident population, and other ethnic groups a further 32%, but located alongside them were limited numbers of bohemians and middle class individuals whose cultural affiliations made them stand out. Gans distinguished these residents from the urban villagers who were the main focus of his Boston study because in his view, the latter's inward looking character and their relative lack of resources led them to have few connections with cosmopolites, beyond those links with professionals whose formal roles necessarily brought them into contact with the urban villagers. This portrayal has been revisited by more recent researchers who have sought to follow up Gans' suggestion that the distinctiveness of ethnic enclaves in inner city areas has a particular appeal for young adults seeking to escape from the mainstream of the dominant culture embodied in suburbia. As Graham Fennell has argued, these locations provide opportunities for students and young, single people to escape from mainstream culture and participate in alternative and unconventional lifestyles. Tim Butler and Garry Robson echo this view in their suggestion that movement to the inner city is prompted by the association of suburbia with boredom, together with a rather nostalgic view of what inner city neighborhoods used to be like.

The research by Butler and Robson into how middle class in migration has transformed certain areas of inner London shows that the cosmopolitan ethos which prizes contact with other cultures is, in practice, at odds with the frequently stated desire of these gentrifiers to live among people like themselves. A matching sense of social distinction and separateness informed Gans' respondent who referred to our kind. Butler and Robson's use of the term enclave to describe the middle class settlements that they studied is an appropriate indication of the exclusivity of the actual patterns of social interaction that these people create. As they note, although living in the context of otherness is said to be highly valued, in practice, middle class, inner city dwellers have very limited interactions with different social groups in their everyday lives. They also note the irony that as gentrification proceeds, the poorer urban villagers (whose cultural distinctiveness constitutes part of the attraction) may find themselves displaced as they are priced out of the local housing market. There is a remarkable parallel here between the unintended impact of these middle class, in migrants' pursuit of an imagined community (what Butler and Robson call Brixton of the mind, or Hackney in the mind) and Ray Pahl's analysis of how the movement of city dwellers to the countryside eroded the very type of community they sought to join.

A particular threat to the existence of the sort of urban villages described by Gans is posed by in migration and gentrification associated with the emergence of so called new, artistic quarters. It has even been suggested by David Bell and Mark Jayne that the distinctive quarters of cities that have been created to service the consumption patterns of the service class, constitute the most notable contemporary expression of the urban village. However, urban villages can be found all over the world and no one characterization adequately captures their diversity. Rather, what is needed is an approach to the analysis of urban villages that is sensitive to their various political, economic, social, spatial, and cultural contexts. The cultural appeal of rural life has very deep historical roots, and so the prospect of a village like existence in an urban setting has understandable attraction to those for whom the city represents an unnatural, alienating, and dangerous entity, even in highly urbanized societies. In other parts of the world, the concept of an urban village retains a more literal applicability. In Latin America, for example, Bryan Roberts has argued that complex patterns of rural–urban migration continue to mean that it is difficult to draw a hard and fast distinction between rural and urban phenomena, making it appropriate to continue to refer to cities of peasants. Urban–rural relations are arguably even more difficult to unravel in contemporary China where millennia of peasant traditions are being overturned by patterns of industrialization and urbanization that are historically unprecedented in their pace and form. And in the countries of the former Soviet Bloc, things are different again because the legacy of under urbanization which characterized the communist period has been the location of the equivalent of urban villages, populated by peasant workers, in the countryside.

Gans' concept of the urban village is thus of far more enduring interest than the specific community of Italian Americans in 1950s Boston from which it originated, but which has long since disappeared. New research continues to add to the picture of life in urban areas settled at some point in the past by migrants from rural backgrounds. For example, the historical work of Christine Harzig and her colleagues has deepened understanding of the specific contribution of women to the communities into which they were transplanted. Gender also figures prominently in Talja Blokland's analysis of contemporary ways of life in the inner city district of Hilleslius in Rotterdam, where women from diverse backgrounds both within and beyond the Netherlands are more prominent in the neighborhood than are men. What is striking about contemporary work in the field of urban studies is not only how many authors continue to make explicit references to Gans' work, but also the extent to which their research findings confirm the veracity of his analysis of the underlying structure and operation of urban social relations despite the substantial transformations which cities have undergone in the meantime. The demonstration by Mike Savage and his colleagues that the forces of globalization are as likely to reinforce people's local attachments as they are to undermine them, follows in the tradition of Gans analysing empirical evidence to identify and then resolve an apparent paradox. It is not at all surprising that sales of *Urban Villagers* placed it among the top 20 bestsellers by American sociologists in the second half of the twentieth century.

See also: Community; Cosmopolitanism; Ethnography; Gentrification; Migration; Transnationalism; Urbanization.

Further Reading

Bell, D. and Jayne, M. (eds.) (2004). *City of Quarters: Urban Villages in the Contemporary City.* Aldershot: Ashgate.

Blokland, T. (2003). *Urban Bonds: Social Relationships in an Inner City Neighbourhood*. Cambridge: Polity.

Butler, T. with Robson, G., (2003). *London Calling: The Middle Classes and the Re Making of Inner London.* Oxford: Berg.

Fennell, G. (1997). Local lives distant ties: Researching community under globalized conditions. In Eade, J. (ed.) *Living the Global City: Globalization as Local Process*, pp 90 109. London: Routledge.

Gans, H. (1962). *The Urban Villagers: Group and Class in the Life of Italian Americans*. New York: Free Press.

Gans, H. (1962). Urbanism and suburbanism as ways of life: A re evaluation of definitions. In Rose, A. (ed.) *Human Behavior and Social Processes: An Interactionist Approach*, pp 625 648. London: Routledge and Kegan Paul.

Halliday, F. (1992). *Arabs in Exile: Yemeni Migrants in Urban Britain*. London: I.B.Taurus.

Harzig, C. (ed.) (1997). *Peasant Maids City Women: From the European Countryside to Urban America*. Ithaca, NY: Cornell University Press.

Levitt, P. (2001). *The Transnational Villagers*. Berkeley, CA: University of California Press.

Pahl, R. (1968). The rural urban continuum. In Pahl, R. (ed.) *Readings in Urban Sociology*, pp 263 305. Oxford: Pergamon Press.

Roberts, B. (1995). *The Making of Citizens: Cities of Peasants Revisited*. London: Arnold.
Savage, M., Bagnall, G. and Longhurst, B. (2005). *Globalization and Belonging*. London: Sage.
Thomas, W. and Znaniecki, F. (1918 20). *The Polish Peasant in Europe and America*. Boston, MA: Gorham Press.
Werbner, P. (1990). *The Migration Process: Capital, Gifts and Offerings among British Pakistanis*. Oxford: Berg.
Young, M. and Willmott, P. (1957). *Family and Kinship in East London*. London: Routledge and Kegan Paul.

Relevant Websites

http://www.holbeckurbanvillage.co.uk/about; http://www.bilstonurbanvillage.co.uk/
UK Urban Village.

Urbanism

G. Bridge, University of Bristol, Bristol, UK

Glossary

Anomie The experience of alienation or separation from others in the city.

A Public A group of people that constitutes itself as a political actor and lies between the state and civil society.

Cosmopolitanism Openness to difference in the city.

New Urbanism A movement that seeks to design urban villages.

Performativity Acting out; doing prior to or apart from representation.

Public Realm A space of political activity conducted by a public.

Public Space Spaces into which all people have access and are able to mix.

Rationality The ability to deal with information in a clear, balanced, and consistent way.

Social Disorganization Coined by Chicago School researchers to make a claim for the state of flux in social norms experienced by immigrant groups arriving in the city.

Urbanism The distinctive features of everyday experience and interactions in the city.

Urbanity An attitude of sophisticated sensitivity in interactions with others.

Urbanization The growth in the proportion of a country's population living in urban areas.

Urban Growth The growth in the physical extent of the city.

Introduction

Urbanism describes the distinctive features of the experience of everyday life in cities. It should therefore be distinguished from urbanization (the growth in the proportion of a population living in cities) and urban growth (the physical expansion of cities). In the development of systematic thinking about cities there have been a number of approaches that define urbanism in different ways.

Urbanism as the Polis

In classical Greece there were conceptions of urbanism that were very much tied to the city as a political unit (or polis). The focus of the city was the agora, the central amphitheatre in which the key political issues of the day could be debated in a structured but open way. Tied to this were ideas of rationality, the ability to work through the issues in an orderly and open way to come to a considered conclusion. These ideas rested, however, on deep social inequalities. Those able to participate in debate were male citizens, excluded were women and slaves who were by implication considered less than rational. Urbanism here was strongly associated with the idea of cosmopolitanism – the city as a space that helped constitute knowledgable actors who were able to distance themselves from everyday localized passions and debate through with their knowledgable equals the key political issues to arrive at well considered conclusions. But this was an urbanism that came at the cost of huge social division.

Urbanism as the Community of Commerce

It was the economic activities of the late medieval cities from about 1000 CE onward that defined another form of urbanism. The guild system of organization of artisans and the growing productivity and wealth of medieval cities resulted in their increasing independence from the rural feudal system. A dense network of economic and social interrelations organized via the guilds and the churches promoted the growth of strong civil societies whose political independence matched the physical distinctiveness of these dense, fortified settlements.

Urbanism as the Experience of Urbanization

From the classical Greek city onward the idea of the city providing a more diverse set of experiences and a more sophisticated life has persisted. However interpretations of the moral worth of urban sophistication have been long divided between ideas of pro and anti urbanism. One set of anti urban sentiments were realized in the analysis of the transition from rural to urban living that came with industrialization and consequent urbanization of populations in the industrial revolution from the eighteenth century onward. Many scholars of the city saw this transition as resulting in the breakdown of those close and multilayered ties that were found in small villages. Urbanism was the experience of this loss of social and geographical belonging that came with the migration to cities for work. In a village social ties are closed and overlapping, everyone knows everyone else and they all know each other in their different roles in life (e.g., people may know each other in their roles as parent, neighbor, work life, and in leisure pursuits). In the city these overlapping social networks start to break down and people know each other in more singular or spe cialist ways (as either neighbor, work colleague, leisure colleague). This expansion and thinning out of social ties has a number of consequences for urban scholars writing in this perspective. First there might be a loss of feelings of belonging and rootedness. Combined with this is the greater fluidity of the city in terms of social and ethnic diversity, brought together in high density living. The researchers of the Chicago School who framed the first modern systematic analysis of the city saw these con ditions as 'social disorganization.' Using the analogy of plant ecology for the urban social environment the Chicago School saw the mix of different social groups (like plant species) in an environment that was unstable and where no one group was fully established as a kind of ecological disequilibrium. In the growing US metropolis at this time (early twentieth century) there was mass immigration, from various countries of northern, then southern Europe, as well African–American migration from the south. The neighborhoods that received immi grants were usually in the center with high density rental apartments in which immigrants gained a foothold in the city. These were termed zones in transition by Chicago researchers.

As well as economic marginality and social instability the Chicago School characterized zones in transition in the moral sense. Social disorganization was seen as a loss of prevailing moral codes in these areas. Different social conventions operated within their own little worlds. In fact the mix of different social conventions and the loss of various forms of moral sanction that is easier to enforce in village communities, led, according to the Chicago School, to a range of new social norms within these neighborhoods, some of them seen as 'deviant' or 'pathological.' Thus the Chicago School research eth nographies investigated the vice and crime they saw to be associated with social disorganization. This was not a wholly negative perspective though. The different moral worlds that existed in these neighborhoods were seen as social adaptations to the pressures of the conditions of the city – they were ways of getting by in the difficult social and physical and economic conditions in these parts of the city. Thus the Chicago School's idea of ur banism was the range of social behaviors and conventions that arose out of the scale and complexity of the city itself. It is a complex but rich ecology.

In contrast to the ideas of social disorganization other urbanites found new forms of social organization in the growing city. A range of studies in the early to mid twentieth century found forms of sociability that had been more associated with rural life. These 'urban vil lages' had dense social networks that were more often founded on common ethnic identity or based on working class sociability. In the present day context in many cities of the global south clusters of associations based on clan or village relations are common features, as are sustained links between certain urban neighborhoods and the villages from which many of the urban settlers came.

Between the ideas of decline of community and dis covery of community in the understandings of the dis tinctive qualities of everyday life in cities, are what might be called limited community relations. These suggest that the experience of social relations in cities is neither dispersed and disorganized nor one of strong urban vil lages but one in which neighborhood based ties are more about personal involvement and preference and that urban communities have limited demands on their resi dents. These experiences might be divided between working class residents who have more place related ties and needs and more mobile middle class professionals. Such contrasts have been drawn out strongly in the ex perience of neighborhood change through gentrification in many cities across the globe. The relationship of middle class residents to their urban neighborhood has been seen as ascriptive, a form of elective belonging where the ambience of particular neighborhoods pro vides the lifestyle trappings that different fractions of the middle class identify with, and which they gravitate to as part of the confirmation of the social identities. In con trast the choices and lifestyles of working class residents are disrupted through displacement or loss of affordable amenities.

Urbanism as Rationality (and Irrationality?)

Rather than urbanism being defined as the richness of urban life (allowing for the difficulties of the environ ment) other views of urbanism see it as comprising a set of forces that are simplifying and singular: an emptying out, rather than filling up of experience. This relates to the two factors: first, the nature of the urban economy and second, the effect of the richness of the city on the mental attitude of city dwellers. There is the idea that social relations became more singular and specialist in cities (in contrast to rural living). Ferdinand Tonnies has argued that modern social relations are characterized by more singular and self interested (what he called gesellschaft, or association) rather than more multi layered and trust based relations oriented toward the community (gemeinschaft). In the modern metropolis association prevailed over community. This approach intersects with the view of the city as a huge economic machine in which people took on more and more spe cialized roles in their work, but that these narrow spe cialisms were more and more interdependent in the overall functioning of the urban economy. This was an economic interdependence that was accompanied, iron ically, by social relations that were less about unconscious interdependencies and more and more contractual, in strumental, and transient. Relationships were made not because of the intrinsic qualities of the other person but because they were seen as a means to an end in terms of the ambitions of the individual. This is also a result of the nature of the psychological effect of urban life. In this view urbanism was associated with a particular type of mentality. The city is full of a range of stimulation. It is impossible for the urban dwellers to take in and process all of the sources of stimulation all the time. They would be overwhelmed emotionally. There is a closing down of the individuals as they take one form of stimulation to stand in for all the others. The result is the blasé attitude, a form of indifference in which the social interactions between people in the daily routines of urban life become flat and undifferentiated. This is a way of coping with the overstimulation of the city. The result is that social re lations between people in the city become the equivalent of monetary relations.

Indeed Georg Simmel, who developed these ideas based on his experience of living in Berlin at the turn of the twentieth century, saw monetary exchange as pri marily a psychological relationship in which the urban dweller could keep at bay the potential overstimulation and chaos of the city. Social, psychological, and economic relations became interchangable in this view of urbanism, one that resulted in 'anomie,' a more anonymous and alienated experience of the city. That very anonymity is something celebrated in the poetry of Charles Baudelaire and the ability of the *flaneur*, the gentleman stroller of the streets, to move unhindered through the crowds and spaces of the city and witness its diversity of human life and experience. This freedom is highly gendered and sexualized though – it is the male figure that can move unaccompanied through the nineteenth century Paris streets. The only unaccompanied female is the prostitute, symbol of the gendered and sexualized nature of urban social and economic relations.

As well as arguments about urbanism as an effect on conscious mental life and a form of rationality in social interaction there are explorations of urbanism as the ef fect of city life on the subconscious feelings and sens ibilities. This draws some of the claims of psychoanalysis into urban studies. The juxtapositions and mix of stimulus in the city is seen to evoke feelings of discomfort or disruption. Memories, dreams, and fantasies can have presence in the everyday experience of the city. The realm of the imagination is a source of creativity and disruption. These experiences of the invisible city and its effect on imagination and affect are encapsulated in ideas of the uncanny or drawing on Freund, the unhomely. Urbanism encompasses both the familiar and unfamiliar, the rational and irrational.

Urbanism and Capitalism

The relationship between the social and the economic, or rather the impact of the economic on the social is registered most strongly in Marxist urban geography. At first reading the idea of urbanism might seem to be part of a cultural and social superstructure and therefore relatively unimportant to the main concerns about the economic relations and position of cities in the circu lation of global capital. However urbanism from the Marxist perspective can be understood as the urban ization of consciousness with all its inherent contra dictions. Marx and Engels believed that the concentration of large numbers of poor workers in cities to drive in dustrial production, and the conditions of exploitation that prevailed, would result in revolution. Urbanization in this sense would be part of the formation of working class consciousness in itself and for itself. Marxist scholars who focused their attention on the urban process in capitalism from the 1970s onward looked to the ways in which the social and economic conditions in cities worked to divide working class consciousness or confine examples of working class protest to certain times and spaces. Cross cutting loyalties of class and ethnicity, the division of the city neighborhoods according to certain trades, and the range of policing and other exercises of power and au thority by the state that could be deployed in the city became the focus of Marxist interpretation of the urban experience. These urban divisions, physical and social, fragmented working class identity and the possibilities of mobilization. They were manifestations in the urban environment of the divisions in the production process and the separation of workers from the product of their labor, the production process, their fellow workers, and their common humanity (or species being) that Marx had portrayed in his early work so evocatively in his interpretation of the idea of alienation. The ethnic and social divisions and the manifestation of the effects of specialized labor through the physical deformity of bodies that Engels noted in his observations of Manchester in the mid nineteenth century were stark portrayals of urban ism in capitalism.

Another development in Marxist inspired scholarship that has had a profound influence on the thinking about cities and the idea of urbanism is the work of the Frankfurt School on the effects of capitalism on culture and everyday life. Frankfurt School scholars argued that the sorts of exploitation and alienation that Marx had observed in the process of the production of commodities now extended to all areas of life. The key terms here are commodity fetishism and the culture industries. Com modity fetishism, Marx noted, was where in capitalism human relations became more and more like the relations between objects, and commodities took on more and more the quality and relations of human beings. Key to this transposition was a range of culture industries in which various aspects of cultural life got taken up by capitalist production, routinized and commodified, and then circulated for the purposes of profit making. The city is the hub of the culture industries and the production of images, advertising, and media messages that bombard the senses and, according to this Marxist analysis, keep people in a state of distraction from the real conditions of their lives and of other people's lives. This is a modern, mediatized version of false consciousness. It is an understanding of urbanism as the experience of alienation in capitalism, realized or not.

Alienation realized through the unequal way that capitalism transforms nature has come to prominence most recently in the Marxist inspired new political ecology approach. This sees cities and urbanism as part of the nature that is being transformed and explains the role that cities might play in contributing, or helping avoid, global ecological disaster.

Postmodern Urbanism

Following on from the insights of the Frankfurt School, the proliferation of images and the collapsing of the distinction between original works and their copies is at the heart of postmodern understandings of urbanism. Rather than an overwhelming sense of the presence of the city in its various stimulations to the senses (the experience of the modern city), the postmodern metropolis produces a sense of nonplace, of ramifying connections, and refracted images. This gives the urban dweller a sense of playfulness, contingency, and flow. Such a sense of flux often masks deeper urban divisions but those advocating a postmodern urbanism point to the productiveness of contingency and flow in terms of a politics less bound by traditional senses of identity and location. Los Angeles (LA) is often seen as the postmodern city par excellence: a constellation of different cultural and social identities, a vast, sprawling urban form with no real center. The LA economy is underscored by Federal defense spending in a city that is also host to the most famous image factory of all – Hollywood. Postmodern urbanism suggests that there is no distinctive urban experience: urbanism is a fragmented, and perhaps ultimately unconnected, set of experiences.

Urban Policy and ‘New Urbanism’

In recent urban policy there have been claims to try to achieve a more coherent and manageable urban experience. Urban revitalization, urban regeneration, or urban renaissance – are all terms that are variously used in policy discourse in cities that have experienced growth and then decline. The drive is to ‘turnaround’ areas where there has been economic decline. The aim of policies range from laudable strategies for social mix and affordable housing and employment opportunities to forms of state sponsored gentrification that encourages middle class residents into the urban core (to benefit local taxes and/or consumption and services) often at the expense of existing working class or poor residents. Themed spaces (named ‘quarter,’ or ‘village,’ or ‘neighborhood’) are often part of this strategy. Lying behind many of these initiatives is an idea of urbanism as involving social mix and mixed use developments, walkability, and a sense of community through physical design. Some of these ideas have been behind the ‘new urbanism movement in the US, the practices and ideas of which have spread to many other countries. The aim is to create village like neighborhoods, based around primary schools at higher densities than traditional suburban sprawl.

One famous example of the new urbanism movement is the Disney Corporation sponsored settlement of Celebration in Florida, constructed wholly using new urbanism design principles. There has been a good deal of enthusiasm for new urbanism among policy circles, particularly in the desire for social mix, reduced car usage, and higher density developments. However examples of new urbanism have been critiqued for being socially exclusive, lacking the dynamism of ordinary city neighborhoods, and for being in many cases covered by restrictive covenants over what modifications can be made to the houses (even down to the color of front doors), how public space is used, and (more implicitly and insidiously) the types of people that are encouraged to buy into the neighborhood. In these ways they have much in common with the gated communities (or ‘common interest developments’) that are rapidly increasing in number all over the world. These developments point toward an increasingly separated and fractured urbanism. Although sometimes invoked as the original inspiration for new urbanism Jane Jacobs arguments *Death and Life of the Great American Cities* pointed to the importance of daily rhythms of vital neighborhood life but open to the wider currents and energies of the big city as a whole. Jacobs's vision captures nicely the tensions of the idea of urbanism as a distinctive pattern of everyday life. These tensions are now realized most emphatically in the rapidly growing cities of Africa and East Asia.

Urbanism, the Public Realm and Public Space

Questions at to what constitutes action, politics, and participation brings us back to one of the core concerns within the idea of urbanism and that is the idea of an urban public realm and public space. The degree to which cities produce a distinctive set of circumstances that permit the constitution of a public between state power and civil concerns in the local community has long fascinated urbanists. The city as a space that allows the coming together of strangers is significant. But as we have already seen that the encounter between strangers in the public spaces of the city is more likely to encourage reserve and indifference, a closing down of communicative possibilities behind the mask of rationality. Yet the city has also been portrayed as the space of urbanity – the place of an open and cosmopolitan attitude, at ease with difference. The range of experience of the city as the place for a self constituting public seems to be somewhere between indifference and openness. Some discussions of the urban public realm have pointed to the importance of rationality – either as a mask behind which people hide to protect themselves from the potential chaos of the city or as a form of urban encounter in which differences can be expressed in conversation or debate.

The rise of mediatized communication as well as spaces in which the issues of the day could be expressed by people from different walks of life is why the influential political theorist Jurgen Habermas pointed to the coffee shops of Paris and London in the eighteenth century as key locations in which 'a public' started to appear. Interpretation since, especially through Richard Sennett's work, points not to the rationality of these spaces but their performativity: the way in which they allowed people from different classes to communicate across their differences by adopting the communicative conventions of the theater. Theatrical communication allowed for the expressing of opinions and exchange of opinions in ways detached from social rank and civil, localized loyalties. The city permits a wide range of performative gestures to take place supported by a hum of activity in ways that in smaller settlements would look peculiar or threatening.

The way that the city can bear different types of expression and appearance has long been considered an important part of the understanding of urbanism as a distinctive form of everyday life. The spaces in which these broader encounters can take place are of interest here. They need to be open, secure, but unregulated. City spaces such as parks, streets, and public buildings as well as less obvious spaces are the potential sites of the coming together of strangers. That is why there has been a good deal of concern over threats to these kinds of public spaces, especially through the privatization of those spaces. This can occur, for instance, when a shopping street loses out commercially to a rival indoor mall, the covered space of which is private and often policed by security services. The more unregulated space of the street is lost and the types of behavior permitted more circumscribed, as people are encouraged to circulate through the shops for the purposes of consumption.

Discussions of the threats of the privatization of public space are accompanied by wider concerns about the nature of that privatization leading to the production of themed spaces. These themed consumption spaces can be found in most large (and many not so large) urban areas today. Especially noticeable are the sorts of spaces that are developed along waterside frontages. Festival marketplaces, waterside villages, heritage sites drawing on the industrial past – all of these districts tend to combine shops, restaurants, and leisure facilities with high end gentrification (through reclaiming old loft buildings or themed new build). These spaces most intensively, but the wider urban area by implication is argued to become like one vast theme park. This is especially significant in the branding of cities in forms of place marketing. Cities are increasingly in competition with each other for inward business investment and tourism, and redevelopments that badge the city in certain areas are seen as significant. This is part of the wider commodification of the city that we have already explored – distinctive ways of life sold as marketing tools in a kind of commodified urbanism.

Transhuman Urbanism

One criticism of all the discussion of public space and the public realm (and traditional discussion of urbanism itself) is that they are anthropocentric: they overprivilege human activity. Yet cities are complexes of machines, animals, software, as well as human activities. All these act together in certain ways to produce the ensemble that is urbanism of which human activity is a part. The city can be seen as an ecology of acting things in ways that actor network theorists have begun to explore. Included here are ideas of hybrids – ecologies of action that are human and non human. Automatic doors, traffic signals, crowd routers, GPS, software spreadsheets, bacteria – all act in certain ways to produce effects and intersections of action whose summative effect is the city itself. The city is more than human in its effects but urbanism also transforms animals and technologies, creating new hybrid forms. The question here is what kind of politics can be built out of these understandings? Ideas such as the democracy of things are beginning to redirect the way we think about action and experience in the city and more widely.

See also: Anti-Urbanism; Flâneur, The; New Urbanism; Postmodern City; Public Spaces, Urban; Urbanization.

Further Reading

Amin, A. and Thrift, N. (2002). *Cities: Re imagining the Urban*. Cambridge: Polity.

Buck Morss, S. (1989). *The Dialectics of Seeing: Walter Benjamin and the Arcades Project*. Cambridge, MA: MIT Press.
Bridge, G. (2005). *Reason in the City of Difference: Pragmatism, Communicative Action and Contemporary Urbanism*. London: Routledge.
Bridge, G. and Watson, S. (eds.) (2000). *A Companion to the City*. Oxford: Blackwell.
Certeau, M. de (1984). *The Practice of Everyday Life*. London: University of California Press.
Davis, M. (1990). *City of Quartz: Excavating the Future in Los Angeles*. London: Verso.
Gans, H. (1962). *The Urban Villagers: Group and Class in the Life of Italian Americans*. New York: The Free Press.
Graham, S. and Marvin, S. (2001). *Splintering Urbanism*. London: Routledge.
Hannerz, U. (1996). *Transnational Connections: Culture, People, Places*. London: Comedia.
Harvey, D. (1989). *The Condition of Postmodernity*. Oxford: Blackwell.
Jacobs, J. (1961). *Death and Life of Great American Cities*. New York: Random House.
Jameson, F. (1984). Postmodernism, or the cultural logic of late capitalism. *New Left Review* 146, 53 92.
Lefebvre, H. (1996). *Writings on Citi*es, Kofman, E. E. and Lebas, E. E. (trans) (eds.). Oxford: Blackwell.
Mitchell, D. (1995). The end of public space? People's park, definitions of the public, and democracy. *Annals of the Association of American Geographers* 85, 108 133.
Park, R. E. (1926). The urban community as a spatial pattern and moral order. In Burgess, E. W. (ed.) *The Urban Community*, pp 3 18. Chicago: University of Chicago Press.
Park, R. E., Burgess, E. W. and McKenzie, R. D. (1925). *The City*. Chicago: University of Chicago Press.
Pile, S. (2005). *Real Cities: Modernity, Space and the Phantasmagorias of City Life*. London: Routledge.
Robinson, J. (2006). *Ordinary Cities: Between Modernity and Development*. London: Routledge.
Ross, A. (1999). *The Celebration Chronicles: Life, Liberty and the Pursuit of Property Values in Disney's New Town*. New York: Ballantine.
Sennett, R. (1974). *The Fall of Public Man*. New York: Norton.
Simmel G. (1950) (1903), The metropolis and mental life. In Wolff, K. (trans) (ed.) *The Sociology of Georg Simmel*, pp 409 424. London: Free Press.
Soja, E. W. (1989). *Postmodern Geographies: The Reassertion of Space in Critical Social Theory*. Oxford: Blackwell.
Tonnies, F. (2001). (1887) *Community and Civil Society*, Harris, J. (ed.) Cambridge: Cambridge University Press.
Watson, S. (2006). *City Publics: The (Dis)Enchantments of Urban Encounters*. London: Routledge.
Wirth, L. (1938). Urbanism as a way of life. *American Journal of Sociology* 44, 1 24.

Relevant Websites

http://www.unhabitat.org
United Nations studies on cities.
http://www.urban age.net
Urban age seminar series.
http://www.urbandesign.org
Urban design website.

Urbanization

P. Knox, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

Glossary

First Modernity The economic, social, political, institutional, and cultural changes that were associated with the Industrial Revolution and that were framed within national territorial states.

Glocalization The simultaneous shift from the traditional institutions and agencies of governance and development upward to the global economy and downward to local organizations.

Megacities Very large cities characterized by both primacy and a high degree of centrality.

Overurbanization Urban population growth that is significantly in excess of the growth of jobs and housing.

Polycentric Urbanization The development of functionally integrated networks of cities and towns, clustered around one or more larger central cities.

Second Modernity The reassertion of the economic and spatial logic of capital at the global scale and involving transnationalism, supranationalism, and cosmopolitanism.

Splintering Urbanization The fragmentation of the economic, social, and material fabric of cities as a result of the selective impact of new technologies and networked information and communications infrastructures.

World Cities Cities in which a disproportionate share of the world's most important business – economic, political, and cultural – is conducted.

Introduction

At the most fundamental level, urbanization can be defined as the relative concentration of population in urban areas – the towns and cities – of a given territory. It is usually expressed in terms of the proportion of urban populations relative to total populations; or as the rate at which the proportion of urban populations is increasing over time. Urbanization may also be defined in terms of the concentration of population in the larger urban settlements of a given territory; and in terms of the increasing density of population within urban settlements. These definitions, however, depend in turn on the definition of exactly what constitutes an urban area. Such definitions vary considerably from one country to another, and often include multitier classifications or taxonomies that differentiate urban areas according to population size and, sometimes, population density, commuter catchment areas, and employment structure. In the United States, for example, the Bureau of the Census recognizes urbanized areas, urban clusters, micropolitan statistical areas, metropolitan statistical areas, and core based statistical areas. The utility of such definitions rests, in turn, on the size and consistency of the areal units – usually administrative districts of some kind – that are used as the framework for data collection. Thus, while most countries have adopted some minimum definition of what constitutes an urban settlement, cross country comparisons must be treated with caution.

Levels and Rates of Urbanization

At the global scale, the overall level of urbanization is now estimated to be about 50%. In general, there is a strong correlation between levels of economic development and levels of urbanization. Most countries of the 'Organization for Economic Co operation and Development' (OECD) have levels of urbanization of 70% or more, the highest levels reaching more than 89% in Belgium, Germany, Iceland, and the UK. The lowest levels of urbanization – less than 25% – are found in Africa (Burundi, Eritrea, Ethiopia, Malawi, Rwanda, and Uganda) and in parts of South and Southeast Asia (Bhutan, Cambodia, Nepal, and Sri Lanka). The rate of urbanization, on the other hand, is inversely correlated with levels of economic development. Most of the world's less developed countries, with low levels of urbanization, are currently experiencing unprecedented rates of urbanization. The highest average annual rates of growth in levels of urbanization between 2000 and 2005 exceeded 5% in parts of Africa. In contrast, some highly urbanized countries such as Italy, Sweden, and Switzerland experienced rates of urbanization that averaged less than 0.2% annually over the same period.

Urbanization as Process

Underlying these basic patterns of urbanization are several demographic processes. Migration from rural to urban areas, and from smaller to larger urban areas, is a principal contributor to increasing levels and rates of urbanization. Emigration and immigration can also be important processes, affecting levels and rates of urbanization within both sending and receiving countries. Natural increase in urban populations may also be important. High rates of natural population increase in cities may be the result of lower death rates associated with higher incomes, better healthcare, and better nutrition; or as a result of higher birth rates – usually associated with high levels of fertility among recent in migrants.

These demographic processes are interdependent with several other kinds of processes, as depicted in general terms in **Figure 1**. Urbanization is driven by a series of interrelated processes of change – economic, demographic, political, cultural, technological, environmental, and social. It is also modified by locally contingent factors such as topography and natural resources. The most important of these interdependent processes is structural economic change. Towns and cities have been the foci of the production, distribution, and exchange processes at the heart of successive modes of production, from feudalism through the various phases and stages of capitalism. These complex relationships between urbanization and structural economic change have been the focus of seminal theoretical works in urban geography, notably by Manuel Castells, David Harvey, Ron Johnston, and Edward Soja.

Urbanization as Outcome

The outcomes of this broader set of interdependent processes contribute to the character of urbanization that is associated with particular modes of production and particular stages of economic development. At the scale of urban systems, processes of economic, demographic, political, cultural, technological, environmental, and social change interact to determine the relative size, spacing, and functional specialization of urban settlements. Within towns and cities, they cause changes in patterns of land use, in social ecology, in the built environment, and in the nature of urbanism as a way of life. Some aspects of these outcomes precipitate important feedback effects as people's values, attitudes, and behavior, government policies, and city planning respond to the changing character of urbanization, thus resulting in changes (sometimes unanticipated) that in turn affect the dynamics that drive the overall processes of urbanization.

These dynamics and feedback effects encompass many of the most important aspects of urban geography. Four aspects of urbanization as outcome are particularly important. First is the mobilizing function of urban settlements. Urban settings, with their physical infrastructure and their large and diverse populations, are places where entrepreneurs can get things done. Cities, in other words, provide efficient and effective environments for organizing labor, capital, and raw materials, and for distributing finished products. Second is the decision making capacity of urban settlements. Because urban settings bring together the decision making machinery of public and private institutions and organizations, they come to be concentrations of political and economic power. Third are the generative functions of urban settlements. The concentration of people in urban settings makes for much greater interaction and competition, which facilitates the generation of innovation, knowledge, and information. Fourth is the transformative capacity of urban settlements. The size, density, and variety of urban populations tend to have a liberating

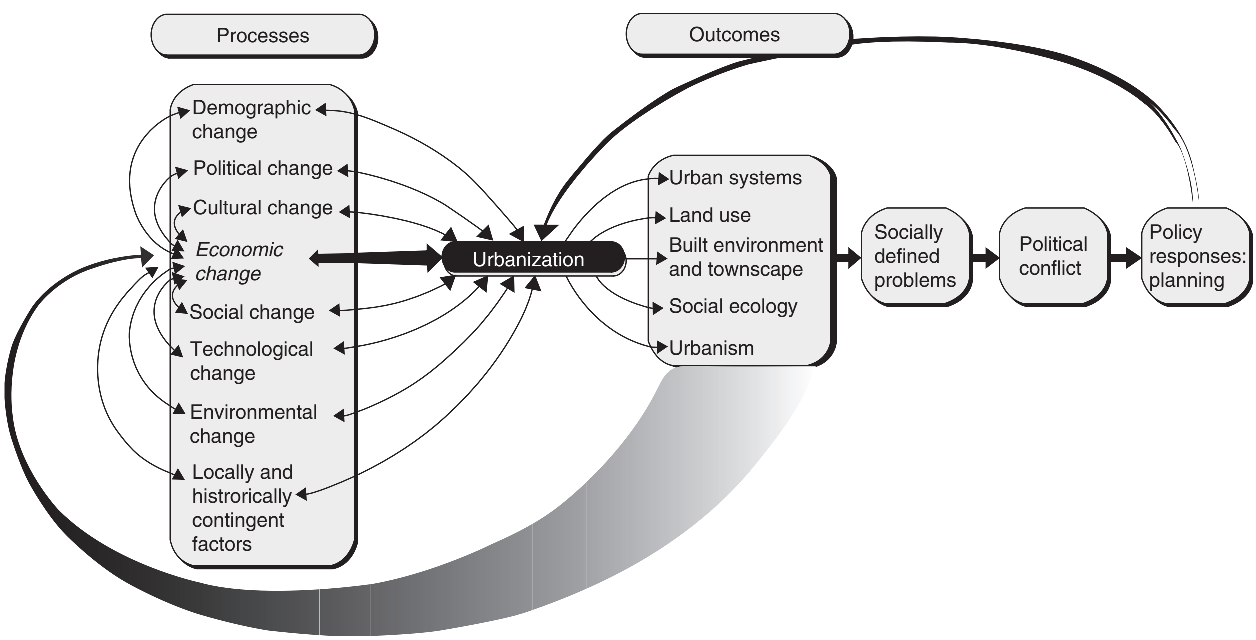

Figure 1 Urbanization: processes and outcomes. After Knox, P. L. and McCarthy, L. (2005). *Urbanization* (2nd edn.), p. 9. Upper Saddle River, NJ: Prentice Hall.

effect on people, allowing them to escape the rigidities of traditional, rural society and to participate in a variety of lifestyles and behaviors.

Changing Processes and Patterns of Urbanization

The earliest evidence for urbanization comes from five regions: Mesopotamia, Egypt, the Indus Valley, northern China, and Mesoamerica. Each of these regions seems to have been an independent setting for a transition from hunter gatherer, subsistence societies to urbanized soci eties based on agricultural production. Although it would have been difficult for cities to emerge in the absence of certain basic environmental, demographic, social, and other preconditions, there are various explanations for these first cases of urbanization. These explanations in clude the need for more centralized structures of social organization in order to administer an agricultural sur plus; the need for occupational specialization and cen tralized social organization in order to construct and maintain the irrigation projects on which early agri culture depended; the pressure of population growth; the emergence of long distance trade in luxury products; the need for defensive strongholds; and the need for cen tralized physical structures and social organization as sociated with a powerful religious elite.

Urbanization slowly spread out from the five regions of urban origin so that by about CE 1000, successive generations of city based empires – including those of Greece, Rome, and Byzantium – had emerged in Southwest Asia, China, and parts of Europe. But urban expansion was a precarious and uneven process. Not until the eleventh century did the regional specializations and long distance trading patterns emerge that provided the foundations for a significant new phase of urbanization based in Europe on a new mode of production: merchant capitalism.

Theories of Preindustrial Urbanization

Merchant capitalism not only led to the resurgence of towns and cities within Europe as points of production, distribution, and exchange but also to gateway ports, entrepôts, and market towns in territories colonized by Europeans. The associated patterns of urbanization have been theorized by James Vance Jr. in terms of five distinct stages: exploration, the harvesting of natural resources, the emergence of farm based staple production, the es tablishment of interior depot centers, and finally eco nomic maturity and central place infilling. Jane Jacobs pointed out that once the dynamics of trade had been initiated, the key to urban growth was a process of import substitution. Although some things were hard to copy because of the constraints of climate or basic resource endowment, many imported manufactures could be copied by local producers, thus increasing local em ployment opportunities, intensifying the use of local re sources, and increasing the amount of local investment capital available. Cities that replaced imports in this way could then afford new types of goods being produced in other cities. The newly imported innovations, in their turn, might then be replaced with local production, opening up the market for still more innovations from elsewhere. As a result, patterns of trade and urban growth were very volatile; and long term local success within the new economic order became increasingly dependent on sustained improvisation and innovation, repeated epi sodes of import substitution, and the discovery and control of additional resources and new kinds of re sources. Colonization and the expansion of trade around the world eventually allowed Europeans to shape the world's economies and urban societies. Subsequently, the Industrial Revolution generated new kinds of cities and led to dramatic increases in the rate and level of ur banization, not only in Europe but also in European colonies and beyond.

The First Modernity and Industry-Driven Urbanization

The Industrial Revolution was the catalyst for an un folding of economic, social, political, institutional, and cultural changes that were, for the most part, framed within national states. Together, these changes constitute the 'first modernity'. As indicated in **Table 1**, the first modernity was dominated by three broad phases in the nature of capitalism, each associated with distinctive processes and outcomes in terms of urbanization. The significance of this evolution for urbanization and urban geography is fundamental. Each new phase of capitalism, drawing on successive innovations and technology sys tems, saw changes in what was produced, how it was produced, and where it was produced. These changes called for new kinds of cities, while existing cities had to be modified. At the same time, of course, cities and their populations played important roles in the transformation of capitalist enterprise.

The first phase, lasting from the late eighteenth cen tury until the end of the nineteenth, was a phase of competitive capitalism, the heyday of free enterprise and *laissez faire* economic development, with markets char acterized by competition between small family businesses and with few constraints or controls imposed by gov ernments or public authorities. Important in this phase of urbanization were principles of initial advantage – based on external economies derived from existing labor mar kets, existing consumer markets, and existing frameworks

Table 1 Urban zat on and econom c deve opment

Major phases of capitalism	First modernity				Second modernity		
	Competitive		*Organized*		*Advanced*		*Globalized*
Technology systems			Coal-powered steam engine		nternal combustion engine		
	Water power Steam engines		Steel Railways		Oil and plastics Electrical engineering		Microelectronics Digital telecommunications
	Cotton textiles ron works		Machine tools World shipping		Automobiles aircraft Radio Telecommunications		Biotechnology nformation systems
Infrastructure development	Canal building	First railway boom	Second railway boom	Streetcar boom	First automobile boom	Motorways and airports	Broadband and satellite systems
Role of central government in urban development	Negligible	ncreasing regulator	Strong (direct) manager/regulator		Strong (indirect) partner/facilitator		Decreasing broker
Phases of urban governance and politics	Laissez-faire	Municipal socialism and machine politics	Boosterism and reform	Cities as growth	Fiscal crisis	Entrepreneurial politics	Neoliberalism
Epochs of urban system development	Mercantile	Expansion and realignment	Organization of industry	Fordist spatial decentralization	Neo-Fordist informational cities	World cities and international networks	
Evolution of urban form	Mercantile city	Early ndustrial	ndustrial city	Suburban infill	Sprawl	Polycentric metropolises and edgeless cities	

of fixed social capital – and of localization economies (cost savings that accrue to particular industries as a result of clustering together at a particular location).

The second phase, beginning at the end of the nineteenth century, was a phase of organized capitalism, characterized not only by Fordism and Taylorism but also by the growth of national labor unions and by expanded roles for central and local governments. Important in this phase were the agglomeration effects of economic linkages and interdependencies, together with processes of cumulative causation, resulting in the emergence of dominant metropolitan regions.

The third phase, beginning in the 1970s, was a phase of advanced capitalism as the more developed countries experienced a shift away from industrial production and toward services, particularly sophisticated business and financial services, as the basis for profitability. This shift began to transform urban occupational structures, sparking de industrialization and fiscal crisis in many cities. Meanwhile, agglomeration diseconomies and negative externalities prompted a loss of jobs and population from many central city areas. In several countries, this contributed to counterurbanization – a net loss of population from cities to smaller towns and rural areas.

Colonial Urbanization and Overurbanization

The Industrial Revolution generated an enormous demand for food and raw materials that intensified colonial and neocolonial development in less developed regions. The development of transportation networks in these regions was critical to urbanization and economic development, with a generalized sequence typically beginning with the growth of port cities that had been established during the era of mercantile colonialism, followed by the development of transportation linkages to interior urban centers and then the emergence of secondary feeder centers that eventually became interconnected as a national urban system. Political independence in the 1950s and 1960s left most former colonies vulnerable to processes of economic underdevelopment and coincided with the onset of the demographic transition. The result was that large increases in population occurred well in advance of any significant levels of industrialization or rural economic development.

Emigration provided one potential safety valve for fast growing rural populations, but most of the more developed countries have put up barriers to immigration. The only option for the growing numbers of impoverished rural residents has therefore been to move to the larger towns and cities, where, at least, there is the hope of employment and the prospect of access to schools, health clinics, piped water, and the kinds of public facilities and services that are often unavailable in rural regions. Natural disasters, environmental degradation, and civil wars have also impacted highly populated rural regions in less developed countries, causing rates of rural–urban migration to increase dramatically. Because these migration streams have been comprised disproportionately of teenagers and young adults, exceptionally high rates of natural population increase have followed. The consequence of all this urban population growth has been described as overurbanization, which occurs when cities grow more rapidly than the jobs and housing they can sustain. In these circumstances, urban growth produces instant slums and squatter settlements, built illegally by families who are desperate for shelter. These urbanization processes have also produced megacities – very large cities (10 million or more in population) characterized by both primacy (i.e., where the population of the city is disproportionately large in relation to other cities) and a high degree of centrality (i.e., the functional economic dominance of the city within the national urban system).

The Second Modernity: Contemporary Processes and Patterns of Urbanization

Transnational corporations, increasingly dominant in the world economy since the 1970s, have successfully outmaneuvered the national scope of both governments and labor unions by moving routine production and assembly operations to lower cost, less developed parts of the world as part of a new international division of labor. Together with the associated expansion of global consumer markets and the expansion of the world economy that resulted from the end of the Cold War and the economic liberalization of China, this globalized framework has allowed the reassertion of the economic and spatial logic of capital, this time at the global scale and involving transnationalism, supranationalism, and cosmopolitanism – all of which are at odds with the managed capitalism and planned modernity of the twentieth century. These changes form the basis of what Ulrich Beck has theorized as a 'second modernity'.

In this second modernity, towns and cities everywhere are subject to an intensification of modernity while the structures and institutions of nineteenth century urbanization are both deconstructed and reconstructed. Cities are "wrestling free from the cocoon in which the managed capitalism and planned modernity of the postwar era had tried to contain them" according to Swyngedouw. One result of this is increasing 'glocalization' – the simultaneous shift from the traditional institutions and agencies of governance and development upward to the global economy (via, for example, the European Union and the North American Free Trade Association) and downward to local organizations. Meanwhile, the wide spread deployment of new information technologies within this new political economy has contributed to what Stephen Graham and Simon Marvin have charac terized as splintering urbanization: an intense geo graphical differentiation, with individual cities and parts of cities engaged in different and rapidly changing ways in everbroadening and increasingly complex circuits of economic and technological exchange. The uneven evo lution of networks of information and communications technologies is forging new urban landscapes of innov ation, economic development, and cultural transforma tion, while at the same time intensifying social and economic inequalities between cities in developed and less developed regions.

The chief beneficiaries of splintering urbanization are the world cities that provide an interface between the global and the local. World cities contain the economic, cultural, and institutional apparatus that channels na tional and provincial resources into the global economy, and that transmits the impulses of globalization back to national and provincial centers. They are the sites of most of the leading global markets for commodities, com modity futures, investment capital, foreign exchange, equities, and bonds; of clusters of specialized, advanced business services; of concentrations of corporate head quarters, international headquarters of trade and pro fessional associations, and the leading nongovernmental organizations and intergovernmental organizations that are international in scope; and of the most powerful and internationally influential media organizations and cul ture industries.

Polycentric Urbanization and Megapolitan Development

The intensified economic and spatial logic of the second modernity has also begun to alter other patterns and processes of urbanization in developed countries. Urban regions have been stretched and reshaped to accom modate increasingly complex and extensive patterns of interdependency, while the political economy of metro politan regions has been reshaped in response to socio economic realignments and cultural shifts. In both Europe and North America, polycentric networks of up to 50 cities and towns, physically separate but function ally networked, clustered around one or more larger central cities, and drew enormous economic strength from a new functional division of labor. Key to this polycentric structure is clusters of decentralized em ployment. One of the most striking components devel opments have been edge cities – nodal concentrations of retail and office space typically located near major highway intersections – but in fact, most decentralized office employment is located in outer suburban and ex urban areas, in what Robert Lang has dubbed edgeless cities. The challenges of characterizing evolving out comes of urbanization have in fact invoked a great variety of neologisms: postsuburbia, exurbia, exopolis, boom burbs, generica, satellite sprawl, for example; but the most distinctive attribute of polycentric urbanization is its scale. Bound together through urban freeways, arterial highways, beltways, and motorways, polycentric urban regions are now beginning to coalesce functio nally into megapolitan regions that dominate national economies.

See also: Agglomeration; Central Business District; Colonialism I; Colonialism II; Counterurbanization; Emigration; Fordism; Immigration II; Mega-Cities; Migration; Port-Industrial Complexes; World/Global Cities.

Further Reading

Beck, U. and Lau, C. (2005). Second modernity as a research agenda: Theoretical and empirical explorations in the 'meta change' of modern society. *British Journal of Sociology* 56, 525 557.
Castells, M. (1977). *The Urban Question*. London: Edward Arnold.
Champion, A. and Hugo, G. (eds.) (2004). *New Forms of Urbanization: Beyond the Urban Rural Dichotomy*. Burlington, VT: Ashgate.
Gottmann, J. (1961). *Megalopolis*. New York: Twentieth Century Fund.
Graham, S. and Marvin, S. (2001). *Splintering Urbanism*. London: Routledge.
Hall, P. and Pain, K. (2006). *The Polycentric Metropolis. Learning from Mega City Regions in Europe*. London: Earthscan.
Harvey, D. W. (1985). *The Urbanization of Capital: Studies in the History and Theory of Capitalist Urbanization*. Baltimore, MD: Johns Hopkins University Press.
Jacobs, J. (1970). *The Economy of Cities*. New York: Vintage.
Johnston, R. J. (1989). *City and Society.* London: Unwin Hyman.
Knox, P. L. (2008). *Metroburbia, USA. New Brunswick,* NJ: Rutgers University Press.
Knox, P. L. and McCarthy, L. (2005). *Urbanization* (2nd edn.). Upper Saddle River, NJ: Prentice Hall.
Lang, R. (2003). *Edgeless Cities*. Washington, DC: Brookings Institution Press.
Soja, E. (2002). *Postmetropolis: Critical Studies of Cities and Regions*. Oxford: Blackwell.
Swyngedouw, E. (2005). Exit 'post' The making of 'glocal' urban modernities. In Read, S., Rosemann, J. & van Eldijk, J. (eds.) *Future City*, pp 21 40. Spon Press: London.
Taylor, P. J. (2004). *World City Network: A Global Urban Analysis*. New York: Routledge.
Vance, J. Jr. (1970). *The Merchant's World*. Englewood Cliffs, NJ: Prentice Hall.

Relevant Websites

http://www.le.ac.uk
Centre for Urban History at the University of Leicester, contains research and hyperlinks related to European urban history.
http://www.inta aivn.org
International Network for Urban Development: Established in Paris in 1974, INTA is a nonprofit international network that encourages the exchange of information, experience and best practices on urban development and revitalization worldwide.

http://www.albany.edu
Lewis Mumford Center for Comparative Urban and Regional Research.

http://www.megacities.nl
Megacities Foundation: The Megacities Foundation in the Netherlands was established as a result of an initiative by UNESCO to focus attention on the problems of explosively growing megacities around the world.

http://www.mi.vt.edu
Metropolitan Institute at Virginia Tech: The Metropolitan Institute's research on urban and metropolitan affairs is available on its website, including publications on suburbanization, boomburbs, edge cities, and world cities.

http://whc.unesco.org
UNESCO: The United Nations Educational, Scientific and Cultural Organization's World Heritage List includes information and images of cities and towns of global importance from around the world.

http://www.unhabitat.org
UN Habitat (United Nations Human Settlements Program) Sustainable Cities Program: The UN Habitat Sustainable Cities Program (SCP) was established to build capacities in urban environmental planning and management as a basis for achieving sustainable urban development and growth around the world.

http://www.unchs.org
United Nations Human Settlement Programme: (UN Habitat) UN Habitat, the United Nations agency for human settlements, attempts to promote socially and environmentally sustainable towns and cities with the goal of adequate shelter for urban residents.

http://www.intoronto.com
Virtual Cities Directory, In Toronto.Com: The Virtual Cities Directory is a collection of links to 3 D representations of real cities around the world including Boston, Helsinki, London, New York, San Francisco, Sydney, Tokyo, Toronto, Vancouver, and Warsaw.

Urban–Rural Continuum

K. H. Halfacree, Swansea University, Swansea, UK

Glossary

Continuum An unbroken gradation from one state to another.

Dualism The idea that some element of reality consists of one of two basic opposing elements (either/or).

Environmental Determinism A theory that (some element of) the environment (physical and/or human) controls human character, behavior, cultures, and societies.

Ethnocentrism Overarching concentration on the people, experiences, values, etc., of one ethnic group, usually also implying its superiority.

Introduction

> The city and the country may be regarded as two poles in reference to one or the other of which all human settlements tend to arrange themselves. (Louis Wirth 1938: 3)

This quote, taken from the classic paper that forms the heart of this article and is discussed more fully below, captures the essence of what is meant by the concept of the urban–rural continuum (or rural–urban continuum; the terms are interchangeable). It is usually applied, as by Wirth, to represent a continuous gradation in the ways of life (or some aspect thereof) between people living at one end of the continuum – the urban – and those people resident at the other – the rural. The more detailed content of this contrast will be outlined after the contextual origin of the continuum idea is considered. Following this, the concept is subjected to some quite major criticisms before its, perhaps consequently quite surprising, resilience today is reflected upon in the conclusion.

Origin of the Concept

To appreciate the origins of the urban–rural continuum we must go back well over a century to note the response of academics and other commentators to the massive spatial and social upheavals brought about by the industrialization and urbanization that arguably reached its zenith in the late nineteenth century. For example, contrast the dramatic change in the geography of Britain's population density between the censuses of 1801 and 1911, as shown in **Figures 1a** and **1b**.

Clearly, the substantial transformation of the physical and human landscape that is implied in **Figure 1b** had a number of major cultural impacts and consequences, many of which are represented throughout this encyclopedia. One of these was to reinvigorate academic interest in the long standing distinction between town and country, or urban and rural, that is so deeply ingrained and recurrent within Anglo Saxon culture. Moreover, this distinction was not value free but predominantly expressed an anti urbanism, an evaluation which, whether intended or not by writers, inevitably impressed itself upon any categorization of urban (and rural) that was developed.

The pioneering study that led to the urban–rural continuum concept was the German sociologist Ferdinand Tönnies's 1887 book *Gemeinschaft und Gesellschaft*. Tönnies's central thesis was that human societies had changed through history from being characterized by *Gemeinschaft* associations – communities built around kinship, neighborhood, cooperative behavior, and place attachment – to becoming *Gesellschaft* associations – societies of impersonal and instrumental relations founded on formal contract and exchange. Crucially, the study also displayed a sense of social geography, with the switch from *Gemeinschaft* to *Gesellschaft* not only signifying historical change but also mapping a spatial dualism between rural and urban, respectively. This enabled *Gesellschaft* to immediately accrete the aforementioned, established, negative 'city' associations – artifice, noise and disruption, sinfulness, pollution, stress – and *Gemeinschaft* to draw to itself the similarly well established, positive cultural associations of 'country' – nature, peace, innocence, purity, and relaxation. Positive cultural associations with city and negative associations with country were largely lost in this process of association.

Although Tönnies's work made a major contribution to contemporary understanding of the then not inconsiderable novelties of the modern industrial city, its mix of sociology and geography also set in train a series of often equally seminal works that, for all their merits, confused and conflated, in particular, the effects of urbanization/urbanism and capitalism. These included Louis Wirth's 'Urbanism as a way of life', widely regarded today as one of the most influential sociological articles ever written. Wirth tried to present a generic account of the now ubiquitous urban culture of the global North. His key thesis was that the large size of cities combined with their

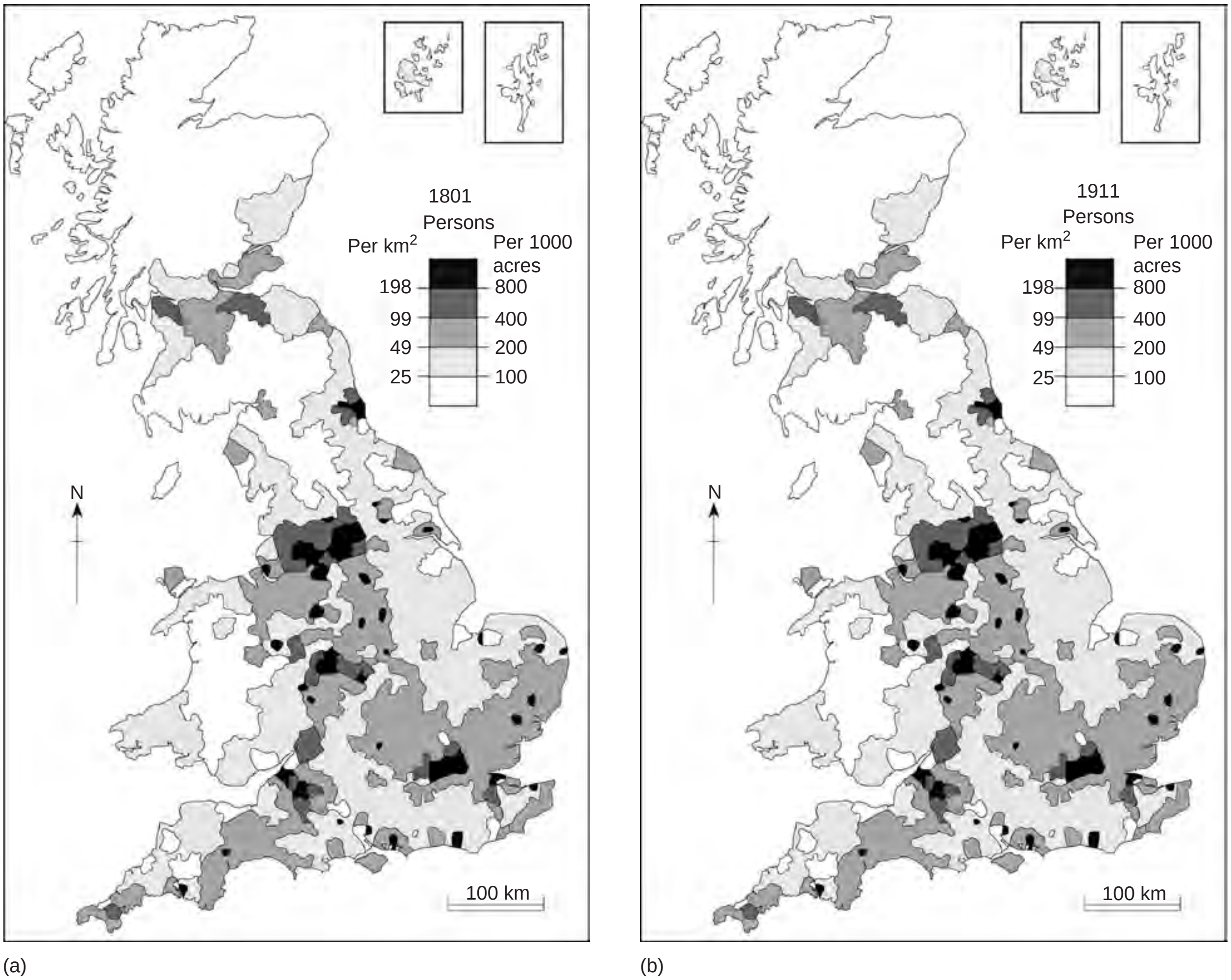

Figure 1 (a) Population density, Britain 1801. (b) Population density, Britain 1911. Source: Lawton, R. (1986). Population. In Langton, J. & Morris, R. (eds.) *Atlas of Industrializing Britain 1780 1914*, Figures 2.1 and 2.3 on p 11. London: Methuen.

dense and heterogeneous sets of social relations to establish a distinctive cultural form that centrally stressed themes such as isolation and social disorganization. Crucially, this focus on urbanism left capitalism very much as the idiomatic elephant in the room.

The studies noted thus far largely operated, as we have seen, through establishing a contrast based on a dualism between an urban and a rural category. However, pioneered by anthropological work on Mexican villages in the 1930s and 1940s by Robert Redfield, by the mid twentieth century it was apparent that such a dualistic formulation was oversimplistic and the seemingly more flexible continuum idea became popular. In this formulation, instead of seeing 'real' places as either urban or rural, the degree to which they expressed urban (or rural) characteristics allowed them to be placed on a scale that stretched from the totally urban at one pole to the totally rural at the other. Perhaps the key work here was Ronald Frankenberg's compilation of studies, *Communities in Britain*. This took a number of 'community studies', most from the 1950s, and sought to place them along the continuum, as shown in **Table 1**. Similar exercises were undertaken in the USA and elsewhere.

Conceptualization of a continuum added flexibility to the categorization of settlements, as it explicitly allowed 'rural' characteristics to be found in largely 'urban' places and 'urban' characteristics to occur in 'rural' places. In part, this reflected the changes being experienced by many rural communities, in particular, by the mid 1960s that could no longer be ignored by researchers. Indeed, it was from research into such communities that a major critique of the urban–rural continuum concept was launched.

Critique of the Urban–Rural Continuum Concept

Although there were numerous precursors (not least in the pages of the *American Journal of Sociology*), a pioneering and highly influential critique of the urban–rural

Table 1 Community studies placed along an urban rural continuum

Author[a]	*Date*	*Community*	*Location on continuum*
Arensberg & Kimball	1940	County Clare (Ireland)	Truly rural
Rees	1950	Llanfihangel (Wales)	Truly rural
Williams	1956	Gosforth (England)	Truly rural
Frankenberg	1957	Glynceiriog (Wales)	Village in the country
Dennis, Henriques & Slaughter	1956	Ashton (England)	Town that is a village
Birch	1959	Glossop (England)	Small town
Stacey	1960	Banbury (England)	Small town
Young & Willmott	1957	London (England)	Community in a conurbation
Lupton & Mitchell	1954	Liverpool (England)	Urban housing estate
Hodges & Smith	1954	Sheffield (England)	Urban housing estate
Mogey	1956	Oxford (England)	Urban housing estate
Young & Willmott	1957	London (England)	Urban housing estate
Durant	1959	London (England)	Urban housing estate

[a] For full bibliographic details, see original sources.

Sources: Frankenberg, R. (1966). *Communities in Britain*. Harmondsworth: Penguin; Phillips, D. and Williams, A. (1984). *Rural Britain: A Social Geography*, Table 1.4. Oxford: Blackwell.

continuum concept came through from the work around a PhD thesis by Ray Pahl, a study of class, community, and social cohesion in villages located in rural Hertfordshire (England) but, crucially, also within commuting distance of London. Pahl was scathing of the attempt to read off 'ways of life' or any social patterns from spatial milieus since, for him, these patterns have to be related directly to social structure rather than geography. Indeed, he saw efforts to link social relationships to geography as a worthless endeavor. Instead, Pahl stressed the need to pick apart the elision of spatial and social that Wirth's formulation had encouraged or, at the very least, set in train. Thus, for Pahl, what was going on sociologically in his Hertfordshire villages was far more to do with sociological developments within a capitalist society than with the supposed influences or ambiences of village England.

Drawing on the interventions of Pahl and numerous other commentators, critiques of the urban–rural continuum idea can be split into two families, those that show how the idea does not work practically and those that engage more conceptually with the idea. From the first perspective, even by the time of Frankenberg's synthesis in 1966, there was plenty of evidence accumulating that population density, population size, and the general social and physical environment did not coordinate any specific form of society. The specific type of society found in a specific place, in terms of things such as sense of community or alienation, cannot be read off from the position of this place along the continuum. For example, Pahl himself found 'urban' people in villages firmly located at the 'rural' end of the continuum and studies in both the UK and North America described 'villagers' in settlements very much toward the 'urban' pole; there were 'communities' in the city and isolated individuals in the country. Such 'mismatches' continue to be found today, of course, Furthermore, as pioneered by the Chicago School of sociology (of which Wirth, perhaps somewhat ironically, was a member) in the 1920s and 1930s, cities hardly exhibited homogenous 'urban cultures' but instead were typically richly shaded by varied subcultures, each as 'urban' as the next. Lastly, the details of what was to be found at the rural end of the continuum was largely neglected and assumed prior to Pahl, as researchers' central interest was with the urban; rural life in many of these formulations always seemed to be 'by implication' rather than considered via direct evidence.

The empirical salience of the urban–rural continuum also relies upon a fairly clear gradation from urban to rural as we go out from the city center. Such a gradation, if it ever was very clear cut, has been progressively destroyed as cities have evolved. In particular, the extensive growth of the suburbs and then, even more radically, the emergence of polycentric urban clusters – especially in North America – disrupt any simple gradation fundamentally with their displacement of the gravitational dominance of a clear urban core. For example, the term 'edge city' was coined in 1992 by the American journalist Joel Garreau to suggest the development of more coherent urban centers of residence, employment, and services than the residentially orientated suburbs or the residentially and service orientated mall suburbs, located on the fringes of these existing suburbs. Crucially, though, the cultures and lifestyles expressed in these places do not 'map' their apparent city status but instead suggest a combination of 'urban' and 'rural' lifestyles. Responding in part to the provocation of Garreau's concept, other authors have increasingly traced the characteristics of, and named with countless neologisms, the immense heterogeneity and hybridity of 'urban' forms in the twenty first century. Within this, although we must be wary of the hype surrounding ideas such as the edge city, any clear urban–rural gradation that does still seem to hold appears almost as a relic feature.

More conceptual critiques touch on a number of overlapping issues. First, and perhaps most fundamentally, the urban–rural continuum idea is strongly rooted in spatial or environmental determinism, now largely seen as an erroneous conceptualization of the relationship between space and society, a topic widely debated in the 1980s within geography. Environmental determinism, as expressed in the urban–rural continuum idea, presents an absolute conception of space, with this space itself possessing causal powers. Consequently, properties arguably inherent to the urban environment (such as population density) were thought to produce a distinctive urban character. In contrast, adopting a less absolute concept of space can position space as intim ately intertwined with the (re)production of social forms and consequently draws at least equal attention to the (re)production of space by human agents and/or social structures. From such a perspective, for the urban–rural continuum to be a valid representation, space and society would need to mutually reinforce one another in such an ordered and orderly manner that seems largely im possible, especially when considered in the light of the following two additional criticisms.

Second, the urban–rural continuum suggests strongly a series of largely self contained communities (*sic.*). This is because although the very continuum idea – as opposed to the extremely rigid urban–rural dichotomy – indicates settlement change with the infinity of possible positions that the continuum throws up, this dynamism is always *a priori* locating of any specific community on the con tinuum. Frozen on the continuum, a place becomes en tirely endogenously defined. This, of course, spectacularly fails to allow for the irreducible dynamism of both people and place, as expressed through the flows of people (e.g., commuters, tourists) and goods (e.g., globalized com modities) that continuously rescript even the most bucolic rural place today, at least in the global North. This sense of constant motion and interlinkage is perhaps best cap tured through Doreen Massey's famous expression of place as being the meeting point of diverse social, cultural, and spatial networks. From such a perspective, it is clear that the urban–rural continuum idea is far too static.

Third, and reflected in almost everything written in this article thus far, the urban–rural idea can be criticized for its introverted ethnocentrism. Almost all of the work used to support and illustrate the concept in the classic studies came from the global North, and from the USA and UK in particular. Notwithstanding all the diversity that is found within 'urban cultures' in, for example, a country such as the USA alone, the spatial form/social form elision fails to give any meaningful independence for cultures associated with cities in the global South where similar spatial forms are found to varying degrees. As any putative 'global' concept, the urban–rural continuum is thus found sorely wanting and one sided.

The Urban–Rural Continuum Concept Today

In spite of all the practical and conceptual problems with the urban–rural continuum, it is a concept that is far from dead. A central reason for this is that the desire to map societal variation by spatial variation – spatializing the social – even in countries so thoroughly urbanized as the USA and the UK, is immensely strongly embedded, even in academic circles. Taking just one example, there is considerable work in the social science of medicine examining suicide rates using, often explicitly, an urban–rural continuum. Some of the results of two such studies in this tradition are shown in **Tables 2a** and **2b**. While neither shows an unambiguous relationship between suicide and position on the urban–rural continuum, a general pattern of suicide rates increasing with levels of rurality in England and Wales and for US males does come across. Moreover, and as just suggested by these two suicide studies, the continuum idea does seem to 'work' to some extent and is thus clearly worth pursuing in a research context. In short, and in spite of the spatial homogenizing tendencies of capitalism, not least in its 'globalization' manifestation, spatial categories from 'urban' through to 'rural' remain important social categories.

One of the reasons that the urban–rural continuum can still be useful in research terms is that it is an aca demic concept with a very strong cultural presence, as has already been indicated; indeed, one could say that for

Table 2a Change in suicide rates per 100 000 person years between 1981 83 and 1991 93 by quartiles of population density, England and Wales

	Age	*Quartile*			
		1 Most urban	*2*	*3*	*4 Most rural*
Males	15 24	4.3	5.8	6.9	6.3
	25 44	2.0	4.6	3.7	4.0
Females	15 24	– 0.5	0.8	0.3	1.7
	25 44	– 2.0	– 1.7	– 1.5	– 1.2

Source: Based on Middleton, N., Gunnell, D., Frankel, S., Whitley, E. and Dorling, D. (2003). Urban rural differences in suicide trends in young adults: England and Wales, 1981 1998. *Social Science and Medicine* 57, 1183 1194, Table 1.

Table 2b Age-adjusted suicide mortality rates 1970 94 by rural urban continuum, USA

	1970 74	1980 84	1990 94
Males			
Most urban	19.8	19.2	18.7
	20.3	20.1	20.7
	20.1	20.8	21.6
	19.9	20.7	22.2
	20.8	20.2	22.3
	22.1	22.1	25.8
	20.7	20.9	23.9
	20.8	21.7	25.0
	20.1	21.5	24.4
Most rural	20.7	23.2	25.7
Females			
Most urban	8.7	6.0	4.6
	6.3	5.0	4.2
	6.8	5.8	4.8
	6.3	5.3	4.6
	5.9	4.9	.4
	6.4	5.2	5.0
	5.4	4.5	4.3
	5.3	4.7	4.5
	5.4	4.5	4.1
Most rural	4.1	4.6	4.1

Source: Based on Singh, G. and Siahpush, M. (2002). Increasing rural urban gradients in US suicide mortality, 1970 1997. *American Journal of Public Health* 92, 1161 1167, Table 1.

most of us it is probably lodged within our common sense, lay understanding of the world. In other words, there is something of the self fulfilling prophecy in the idea that spatial form is associated with social form: people's expectations of what they will find in particular places, and subsequently, how they will act and be expected to act in such places can be seen to be related to the cultural representation of that place. The cultural weight of all that has been written about life in the city and how it differs from that in the country is so great that it almost cannot help but influence our expectations and behavior. This weight was, of course, expressed and critically analyzed so well academically by Raymond Williams in his 1973 book, *The County and the City*. However, from the perspective of how these as sociations are reinforced at a mundane, trivial, yet persistent, level within popular culture, there is no better manifestation than the television dramas that track the trials and tribulations facing 'urban' people adopting new, supposedly more relaxed, lives 'in the country', as inspired by such best selling memoirs as Peter Mayle's *A Year in Provence*. Moreover, this deeply culturally embedded perceptual sense of urban–rural difference can, to some extent, be seen to be reinforced further by, and indeed also informs, key institutional practices, such as land use planning, where a desire to retain a clear spatial demarcation between town and country is re flected in such things as Green Belt policies and the generally strict regulations preventing urban 'sprawl' into the countryside. In spite of edge cities, such institutional practices help keep the urban–rural continuum alive, at least spatially.

Furthermore, and again as **Tables 2a** and **2b** suggest, it seems that in spite of the best endeavors of social science to reject the entwining of society and space, there is still something about space that helps to 'explain' social form. From a contemporary theoretical angle, this comes through, for example, in actor network theory (ANT), with its depiction of agency as fundamentally relational (contextual), via the network, rather than compositional, via the (constructed) individual, thereby encapsulating nonhuman objects as well as people. To understand this a bit further, while hopefully not succumbing to the omnipresent risk of spatial determinism, we note briefly the work of Claude Fischer, who argues that urban living facilitates – but does not determine – the already noted proliferation of subcultures and identities because the diversity, density, and heterogeneity of 'options' in the urban environment allows (and – recognizing that this is not always a matter of choice but often stems from per secution and prejudice – even forces) such identities to be practically realized. The critical mass necessary for a sustained presence is possible, and its condensation through overlapping networks and repeated meetings leads, in cities across the world, thus encompassing, fi nally, the global South, to the emergence of strong group identities such as those of sexual minorities, youth gangs and tribes, *avant garde* artists and bohemians, religious denominations, marginal political factions and, of course, strongly ethnically distinctive communities. This is not to say that such realizations are impossible in rural areas, or even necessarily that much harder to achieve, or that such realizations will actually take place in even the largest of cities if economics, ideologies, politics, etc., serve to suppress them (or, in terms of otherwise per secuted groups, to make them unnecessary) but simply that more of the logistics are in place and the ground work done, so to speak, in such locations to facilitate these realizations.

In conclusion, though, the author first follows Peter Saunders, who approvingly cited a 1960 paper by the American sociologist Richard Dewey that assessed the urban–rural continuum as "real but relatively un important". Nuancing this though, while it may be the case overall that the urban–rural continuum is a poor heuristic tool, detailed empirical work can show it to be of much greater apparent significance in certain places and for certain phenomena, and study of these mani festations is clearly worthwhile. Thus, maybe Louis Wirth and his legacy still have a place in twenty first century urban and rural social science.

See also: Anti-Urbanism; Urbanization.

Further Reading

Benet, F. (1963). Sociology uncertain: The ideology of the rural urban continuum. *Comparative Studies in Society and History* 6, 1 23.

Cochrane, A. (2005). Cities. In Daniels, P., Bradshaw, M., Shaw, D. & Sidaway, J. (eds.) *An Introduction to Human Geography. Issues for the 21st Century* (2nd edn.), pp 213 228. Harlow: Pearson.

Dewey, R. (1960). The rural urban continuum: Real but relatively unimportant. *American Journal of Sociology* 66, 60 66.

Fischer, C. (1982). *To Dwell Among Friends: Personal Networks in Town and City*. Chicago, IL: University of Chicago Press.

Fischer, C. (1995). The subcultural theory of urbanism: A twentieth year assessment. *American Journal of Sociology* 101, 543 577.

Garreau, J. (1992). *Edge City: Life on the New Frontier*. New York: Random House.

Halfacree, K. (1993). Locality and social representation: Space, discourse and alternative definitions of the rural. *Journal of Rural Studies* 9, 23 37.

Pahl, R. (1968). The rural urban continuum. In Pahl, R. (ed.) *Readings in Urban Sociology*, pp 263 293. London: Pergamon.

Pile, S. (1999). What is a city? In Massey, D., Allen, J. & Pile, S. (eds.) *City Worlds*, pp 42 48. London: Routledge.

Saunders, P. (1981). *Social Theory and the Urban Question* (2nd edn.), ch. 3. London: Unwin Hyman.

Savage, M., Warde, A. and Ward, K. (2003). *Urban Sociology, Capitalism and Modernity* (2nd edn.), ch. 5. Basingstoke: Palgrave Macmillan.

Wirth, L. (1938). Urbanism as a way of life. *American Journal of Sociology* 44, 1 24.

Relevant Websites

http://www.dailykos.com/story/2006/12/7/04913/9030/204/278472
America Before Pearl Harbor Early Kodachrome images.

http://www.cartoonstock.com/
Cartoons of Urban and Rural Life.

http://www.bartlomiejspeth.com/bartlomiej speth urban rural photography.html
Urban to Rural in Eastern Europe.

Utopian Cities

G. Brown, University of Leicester, Leicester, UK

Glossary

Dystopia An imagined place or state in which the condition of life is extremely bad, and frequently marked by deprivation, inequality, and an oppressive or repressive social order.

Utopia An ideal place or state, often a small-scale city-state; or a visionary spatialized system of social and political life.

The Spaces of Utopias

Cities have been seen both as sites imbued with emancipatory potential and as the location of alienation and despair. Over the centuries, many utopian proposals for alternative ways of living in urban space have been teased out of this paradox. This article examines the development of utopian thinking and propositions in relation to cities, charting utopias through fictive accounts of future cities, radical town planning proposals, and ongoing theoretical debates concerning the ethical principles that motivate utopian thinking.

The term 'utopia', as originally conceived by More, had an ambiguous, double meaning from the start, as it is a contraction of two different, but related, concepts. A utopia is both a good, happy, or fortunate place (eutopia); and 'no place' (outopia). This dual meaning is suggestive of utopia as somewhere that is perfect, and also a place that does not exist and perhaps never can exist. More's *Utopia* was a satirical critique of the English society in which he lived. Satirical and potentially subversive political critique has been a feature of much utopian fiction ever since. It serves as a means to measure society as it is against society as it could (or should) be, with the aim of inspiring radical change. Although this is the original use of the term, it has long ceased to apply purely to works of fiction, and has come to refer to practical experiments of putting utopian ideas into practice, for example, through the creation of intentional communities, or the building of 'new towns'.

Spaces, spacing, and architecture are central to utopias – just think of the root of the word, 'topos' (Greek for 'place'). Rather than simply issuing a manifesto of demands, utopian writers offer a visionary sociospatial blueprint of their alternative society. The question of scale is important here, with many utopian thinkers restricting their proposals to the scale of the city (and frequently advocating a return to some form of city state). This is explicit in the names given to many utopias – the Heavenly City, New Jerusalem, the Garden City, and the Radiant City. Changes in urban space and urban life are bound up with each other in many utopian proposals, with transformations in one entailing changes in the other. Mumford theorized that the archetypal ancient city was the original utopia, in that was established in the name of God, as a sacred place, enclosed from the rest of society, designed in an orderly manner that reflected the dominant cosmology – a heaven on Earth. Most utopias share a common presentation of their end goal as something 'settled' and comforting, even if their proposals are (currently) unsettling and the processes of achieving them are uncomfortable.

Baeten has argued that, in the late twentieth century, the spatial and the urban lost their centrality to utopian projects. Later this article discusses some tensions between utopias of spatial form and utopias of process, as well as the search for a more dialectic form of utopian thinking (proposed by Harvey) that attempts the dynamic synthesis of both approaches. For now, however, it is important to note that although the urban has been central to many utopian proposals, the rural has never been completely absent from utopian writing. *News from Nowhere*, a book by Morris, is typical in portraying a vision of nature 'taking back' the city and making space, on which new (smaller scale) societies can be built. Many rural utopias (and 'ecotopian' visions) rely on the 'comforting' materiality and imagery of nature, connection with the Earth, and nostalgia for pastoral, vernacular dwellings. Whether through revolution, civil war, plague, or natural disaster, fictional accounts of utopian futures have frequently relied upon the narrative device of creating new sociospatial structures on ground that has been razed as the consequence of an apocalyptic event. This not only produces space for the new society, but also symbolizes the decay and ruination of the extant society that they are subjecting to critique.

As the example of these fictional accounts illustrates, discussions of utopian proposals that pay attention to the form, content, or function of the imagined society in isolation from each other are insufficient to adequately capture the definition and value of utopianism. Levinas proposed that a more analytical approach is required that evaluates what desires are enrolled in any utopian project. This analytical approach asks what aspects of society are repressed, controlled, or erased in the proposed utopian society. This highlights the importance of considering the ethical principles that motivate utopianism, and this is an issue that is addressed throughout the rest of this article. It is suggested that all utopias contain their own commitment to an ethical, affective, or emotional 'good' (or, at the very least, a relational 'better'). They contain an emergent sense of what 'should be'. However, whether that is an optimistic or a pessimistic vision strikes at the core of the utopian paradox as simul taneously a 'good place' and 'no place'.

It is worth remembering that many advocates of al ternative visions of society have shunned the term 'uto pia' because it has become associated with the 'unrealistic' and 'unachievable'. The representatives of the established order are prone to discrediting their critics (and those who proffer alternative ways of organizing society) as 'utopians' to infer the naïve impossibility of their proposals. As a consequence, these visionaries prefer to stress the achievability of the transformations they propose. Pinder has suggested that one of the more productive functions of utopianism lies in the challenges it poses to existing understandings of the possible and the impossible. In his studies of utopian urbanism in twentieth century Europe, he has focused on utopian imaginings that engaged with the existing spatial and social relations of their time in order to expand perceptions of the 'possible' with a view to radical change. It is to some of these modernist (and postmodernist) urban utopian projects that attention is now turned, offering a brief, critical history of utopian ideas of the city since the late nineteenth century. This article then examines some key theoretical debates that cut across these temporally and spatially specific imaginings.

Planning Utopias

The late nineteenth and early twentieth century, as a period of unsettling social and economic transformations and rapid urbanization, saw an intense flurry of utopian thinking and manifestation. Many of these changes were perceived as threatening to plunge humanity into chaos and destruction – a threat that was reflected in some utopian writing of the period (which harks back to 'simpler' times), but other thinkers sought to harness this dynamic and push it further. The next few paragraphs chart the development of modernist plans for the de velopment of new forms of urban living in response to this period of rapid industrialization and increases in urban density.

Like William Morris, Ebenezer Howard proposed a restorative utopia that sought to restore values of small scale harmonious communities from an idealized past. Their visions of clean, fresh, open, and light cities were a response to the pollution of Victorian cities and an en gagement with contemporary debates about social inequality, slum housing, pollution, and public health. Despite their visions, and the subsequent interpretations of them, they did not perceive themselves to be anti urban, but attempted to offer alternative modes of urban life. Howard's proposals for the Garden City were in large part a synthesis of other's ideas, including the anarchist geographer Kropotkin's ethics of mutual aid and co operation. Howard envisioned urban space composed of separately zoned areas for industry, housing, and small scale agriculture, through which he hoped to achieve a unification of city and country. Adapted versions of his plan for the Garden City were implemented in the building of Welwyn Garden City and Letchworth outside London, and had an influence on the planning of the small company towns built by paternalistic (frequently Quaker) industrialists for their workers (e.g., New Ears wick built by Joseph Rowntree). His ideas also had a lingering influence on the building of inter war suburban developments in the UK. Jane Jacobs sharply denounced utopian plans like Howard's as 'nice' towns for people with no plans or ambitions of their own, where the only permissible plans, of any significance, were controlled by the planners themselves. Nevertheless, Morris and Howard perceived urbanism as a means of changing so ciety for the better. Contemporary anarchist and socialist ideals inspired their visions of urban life, even if these were largely erased from and sanitized in the subsequent implementation of elements of their plans.

In contrast to the restorative utopian town plans ad vocated by Howard, Le Corbusier was inspired by the new technologies and materials emerging at the begin ning of the twentieth century, as well as the increasing pace of urban life. His was very definitely a pro urban vision harnessing the symbols and materials of the modern, the new, and the industrial. Le Corbusier at tacked the 'disorder' of large cities and sought means to transform their spaces, in order to 'cure' them. Le Cor busier, in his (interior and exterior) designs, favored Euclidean geometry, static order, and smooth, clean lines – much removed from the tactility of ornate decoration and organic shapes. However, lest Le Corbusier's work be wholly associated with high modernist geometries of straight lines and right angles, his proposals for the Ra diant City emphasized the 'basic pleasures' of sun, space, and greenery. If Howard proposed horizontal garden cities, Le Corbusier sought to develop vertical garden cities that celebrated these basic 'natural' pleasures without recourse to preindustrial nostalgia. His was a vision of 'order as a task', a technical procedure, rather than a social relationship. Le Corbusier's technical ap proach poses important questions about whose interests were served by this conception of social order. Certainly, when he devised an urban plan for Algiers, the archi tectural designs and town models stood in direct op position to Islamic culture and were perceived to promote an urban order that served the extension of French colonial interests. In 1923, Le Corbusier famously posed the choice between 'architecture and revolution', claiming revolution could be avoided through the reshaping of the urban built environment in ways that could come to terms with the demands of industry and the modern age. This statement was hotly debated at the time (and continues to be controversial). Some commentators interpret his comment as advocating a strategy for the prevention of Communism, while others believe he was identifying architecture as a radical act that could circumvent the need for revolution. The fact that he went on to establish links with Mussolini and the Vichy government in France has often added weight to the former interpretation.

Although Le Corbusier's huge high rise buildings contrast starkly with the cottages envisioned by Howard, ideas and images derived from both their visions continue to influence contemporary urban planning. Town planning envisions better (or, at the very least, 'good enough') cities, it is a practice imbued with hope and optimism, even if it is inherently too rational to ever be fully utopian. MacLeod and Ward argue that utopian thinking has inspired and shaped many contemporary cities, precisely through processes of official town planning, creating "an intensely uneven patchwork of utopian and dystopian spaces," from sink estates to luxurious gated communities. This article now focuses more fully on utopian influences in the contemporary postmodern city.

Postmodern Utopias

The postmodern city (with Los Angeles problematically servicing as its synecdoche) has been typified as a decentered urban landscape littered with 'edge cities', traditional suburbs, gentrified inner city enclaves, gay and ethnic cultural quarters, and multiple zones for leisure and consumption. Hannigan has proposed that the more fantastical of these urban consumer spaces (Las Vegas and megamalls) are the new utopias. If they are, then for Harvey, they are 'degenerate utopias'. Yet, the sanitized, secure, enclosed, and controlled spaces of gated communities and privatized resorts may also have more in common with the visions of many classical utopias (from More onward) than many contemporary commentators might care to acknowledge. Both seek a contained and controlled urban space in which to enact new and idealized modes of living. The vision of that ideal may have changed, but the belief that changes in urban form can enable altered social practices has not. The emancipatory potential of post modern cities, and the paradoxical possibilities for freedom and innovation that exist within them, continues to be much debated, as a recent collection edited by Lees illustrates.

In recent decades there has been a shift away from urban political priorities shaped by rational managerial planning toward the reshaping of the urban landscape by entrepreneurial developers. Within this political context, in Britain and the United States, gentrification as a mode of urban regeneration has been posed as the 'medicine' for urban social ills. Commenting on recent proposals from the British Urban Task Force and its American counterparts, MacLeod and Ward believe that a new urban utopianism is emerging out of the blueprints for environmentally sustainable and culturally inclusive cities. In contrast to fragmented postmodern cityscapes, 'new urbanism' is proposing more holistic urban forms based on small towns and organic urban villages. In this endeavor, it is drawing selectively on Howard's plans for garden cities without addressing his belief in the social ownership of land.

In the last decade, several geographers, such as Baeten, Harvey, and Pinder, have challenged contemporary 'critical urbanists' for offering trenchant critiques of existing urban life, but timidly refusing to offer alternatives for how urban processes and forms might be radically different. Many critics locate this timidity in a response to postmodern skepticism about 'grand narratives' and the political legacy of the failure and authoritarian degeneration of so many modernist utopian projects (e.g., state socialism). For Baeten, the dominant concepts used by contemporary urban geographers (e.g., globalization, world cities, underclass, and social capital), while descriptive of current conditions, are 'surprisingly neutral' in tone and diminish critical geographers' ability to offer alternatives.

Other scholars and urban activists have found inspiration for new utopian projects within the Situationists' orientation to the open ended, ludic aspects of urban lived space (something they share with Lefebvre who proclaimed himself a 'partisan of possibilities'). The Situationists envisioned a city constituted by grand situations between which its inhabitants continually drift. These psychogeographical drifts through the city are attuned to the possibilities contained within urban life and resonate with more recent interest in the molecular politics of affect and performance. In this spirit, Doron has drawn attention to the 'dead zones' of the city, highlighting how spaces that mainstream representations present as dystopian can be lived as utopian spaces of escape, refuge, and entertainment for some 'marginalized' social groups. These are urban utopias that celebrate and create space for playfulness, urban erotics, imaginative spontaneity, and the festive. However, Pinder, writing from a perspective that is very sympathetic to the Situationists, acknowledges that their attacks on the fixity of urban structures and celebrations of open and continuous change take on a very different connotation in a period when the advocates of neoliberal free markets demand 'flexibility' from the labor force and the built environment. More harshly, Tafuri charged the Situationists with complicity in the (modernist) capitalist project of rationally planned urban development on the basis of their very commitment to the project of utopia and the importance of altering the esthetic form of the city.

Beyond Utopias of Place and Process

Debord and the Situationists did not envisage utopia as an endpoint that would harmoniously resolve all social and spatial conflicts, but as an ongoing process of journeying toward an aspired set of social relations. They championed a form of utopianism that moved beyond building 'cities on a hill', in which fixed social and spatial order were imposed at the exclusion of any forms of spontaneity or disorder that might get in the way of the achievement of that (singular) plan. An appeal to the subversive meanings and practices of everyday life in urban space can refocus the attention of utopian thinking away from prescribed visions of future spaces in favor of a more embodied and process oriented approach to changing urban space, architecture, and urbanity.

This highlights the question of closure in utopian projects. For many thinkers, such as Sennett, inspired by anarchist ideas and heterodox Marxism, planning's obsession with imposing order and closure debilitates the potential for urban communities to make their own decisions about their lives and negates the generative potential of encounters with strangers and other cultures. In contrast, Harvey stresses that utopianism must be a dialectical spatiotemporal project rooted in a critique of existing capitalism (and the possibilities it leaves open for alternatives) at the same time that it advocates different trajectories for society. He has critiqued the openness of 'utopias of process' on two levels: first for failing to recognize that some form of closure (however temporary) must take place in order for utopian spaces to be realized; and second, for overlooking the traditions and institutions that utopian projects (will) accumulate over time. This situation demands that the advocates of new utopian spaces must confront issues of power and authority in the realization of those spaces. Harvey poses the challenge that the Left must confront the issue of what is legitimate (and, by implication, progressive) authority. He argues that to materialize any one sociospatial design shuts down the possibility of materializing others. The dialectical utopianism Harvey advocates operates with an either/or, rather than a both/and, dialectic – and he challenges progressive thinkers both inside and outside the academy to confront the choices available to them.

Baeten suggests that an alternative dialectic is at work in which "the construction of utopian futures has fuelled dystopian critiques and vice versa." A feminist utopia is likely to be dystopian for a moral conservative and vice versa, but both reactions can inspire new utopian aspirations for the other group. Feminist critiques of classic utopianism, and their own alternative utopias, recognize that all utopian visions are partial and partisan. In this respect, utopias help expose the major divisions of interest and ambition within a given society. Kraftl has proposed that this utopian/dystopian dualism needs deconstructing. He argues that the process of building utopias involves both disruption and creation; and that despair and hope come together in providing inspiration for many utopian ambitions. As one possible solution to the partiality of utopian claims, Sandercock has proposed a vision of participatory planning as an open ended process that could never be fully finished. Her hope was that the participatory process could mobilize the visionary desires of all sections of the population – inevitably, incommensurate desires that would need to be constantly negotiated in their realization.

Those who promote open processes of utopian experimentation can fall into the trap, in articulating understandable and justified critiques of the enclosure and privatization of public space in the contemporary city, of nostalgically yearning for a lost urban public realm, without paying attention to multiple exclusions from 'traditional' public space around axes of gender, ethnicity, or religion. Equally, they overlook the experience of encountering 'difference' in public space, which can be experienced as uncomfortable, risky, and painful for many, even though it remains at the core of many (recent) 'utopian' visions of 'better' cities. Those who advocate open, nomadic urban life potentially overlook the desire for fixity and simplicity held by many (possibly most) people in societies marked by rapid change and increasing speed. Indeed, for Harvey, the increasing speed of contemporary life actually precludes the time to imagine alternatives. This article concludes by charting recent approaches to utopianism that rely on modest practices of hope as a means of anticipating potential utopias.

Practices and Principles of Hope

As suggested throughout this article, many utopian visions rely on nostalgia for the 'comfort' of the past (and elsewhere), whether that past is childhood, a previous 'golden age', or a rural idyll (itself usually located in an imagined past). Nostalgia is at the core to the restorative utopias proposed by Morris and Howard, among others. In contrast and drawing on the writings of Bloch, Anderson has stressed the immanence of utopia and its infusion of everyday practices of hope (music listening, in his case study). He attempts to move beyond an understanding of utopia as coming into existence through its qualitative difference from existing (nonutopian) social arrangements; focusing instead on the excess of the utopian in a world always still in a process of becoming. In this way, utopianism is understood as an anticipatory rather than restorative process. Bloch's theoretical–empirical practice offers a utopian ontology. It perceives utopia as existing in most aspects of quotidian life, and offers a means of discerning the utopian moments that have already taken shape within the mundane practices of the present. He offers "an account of an experimental, enchanted world full of the not yet in suspense," (Anderson, 2002: 217).

If many utopias have been written or designed in response to a reflection upon the desires that remain unfulfilled within existing social relations, working with Bloch's notion of the immanence of utopian hope offers new possibilities for the utopian imagination. Rather than working negatively from a lack or absence in the present, what might our utopias look and feel like, if they were written instead, with hope, as Anderson advocates?

Writing from a similar perspective, Kraftl has suggested that the good place/no place paradox at the heart of the concept of utopia can induce anxiety, as it poses difficult questions about whether hopeful ambitions can ever be achieved. For him, utopias are often predicated on a sense of yearning for the 'not yet'. He proposes that the ethical and emotional reach of utopias might be extended if their imagery, desires, and practices were more unsettling and discomforting in themselves and paid more attention to the unsettling impacts they have traditionally offered. To explore this idea further, Kraftl examines the case of communes and intentional communities as examples of existing spaces in which participants have attempted to realize their utopian hopes and ambitions. He recognizes that such experiments are intended to comfort rather than unsettle those who participate in their creation, but at the same time, they frequently alienate and unsettle (albeit unintentionally) those who live nearby but do not share the social vision of the communards. Nevertheless, he acknowledges that "the *intention* of any utopia is … a large measure of its function" (Kraftl, 2007: 123).

Kraftl's example of intentional communities is timely, as they and other prefigurative political practices have recently attracted the attention of geographers. In their writing on 'autonomous geographies', Pickerill and Chatterton have examined a range of experiments in social autonomy (e.g., social centers, convergence spaces, and intentional communities), where people have experimented with bringing into existence their desire for noncapitalist, egalitarian forms of social and political organization that are based on mutual aid and solidarity. Although these are spaces of resistance, they are also creative spaces in which the possibilities for what might be are played out. Within this constellation of sites, small scale workable alternatives to existing social relations are tried out, and the tensions between individual freedom and collective self rule are continually negotiated in practice. As such, autonomy is always incomplete, and in the process of becoming, it emerges out of the process of acting autonomously. These prefigurative spaces, by attempting to put into practice ethical principles of hope, generosity, dignity, and respect, are also attempting to imagine and enact spaces that are not entirely knowable or thinkable within the conventions of contemporary society. This is a performative politics that attempts to bring hope into being through experimental practices. In the process, they are "opening out the affectual potentialities of hope" (Kraftl, 2007: 125). In this respect, they transcend the hope–critique dualism found within many accounts of utopian thinking, as they experiment with hope while also critiquing existing social relations. How open they can remain to multiple futures and how much they can resist the temptation to impose their own visions in a way that forecloses other possibilities remains to be seen. However, in making space for utopian thinking at a time when this imagination has frequently been sidelined by the pressures of everyday life, they offer hope for the future.

See also: Anarchism/Anarchist Geography; Gated Communities/Privatopias; Planning, Urban; Situationism/Situationist Geography; Situationist City.

Further Reading

Anderson, B. (2002). A principle of hope: Recorded music, listening practices and the immanence of Utopia. *Geografiska Annaler* 84B(3 4), 211 227.

Baeten, G. (2002). Western utopianism/dystopianism and the political mediocrity of critical urban research. *Geografiska Annaler* 84B(3 4), 143 152.

Bloch, E. (1986). *The Principle of Hope,* vols. 1 111. Plaice, N., Plaice, S. and Knight, P. (trans.). Oxford: Blackwell.

DETR. (1999). *Towards an Urban Renaissance: Sharing the Vision*. London: Department for the Environment, Transport and the Regions.

Doron, G. (2000). The dead zone and the architecture of transgression. *City* 4, 247 263.

Hannigan, J. (1998). *Fantasy City: Pleasure and Profit in the Postmodern Metropolis*. London: Routledge.

Harvey, D. (2000). *Spaces of Hope*. Edinburgh: Edinburgh University Press.

Jacobs, J. (1961). *The Death and Life of Great American Cities*. New York: Random House.

Kraftl, P. (2007). Utopia, performativity, and the unhomely. *Environment and Planning D: Society and Space* 25, 120 143.

Lees, L. (ed.) (2004). *The Emancipatory City? paradoxes and possibilities*. London: Sage.

Levitas, R. (1990). *The Concept of Utopia*. Oxford: Phillip Allen.

MacLeod, G. and Ward, K. (2002). Spaces of Utopia and Dystopia: Landscaping the contemporary city. *Geografiska Annaler* 84B(3 4), 153 170.

Miles, M. (2008). *Urban Utopias: The Built and Social Architectures of Alternative Settlements*. London: Routledge.

More, T. (1988). *Utopia* (Logan, G. M. and Adams, R. M. (eds.); first publ. 1516). Cambridge: Cambridge University Press.

Morris, W. (1993). *News from Nowhere and Other Writings* (first publ. 1891). Harmondsworth: Penguin Books.
Mumford, L. (1966). *The City in History: Its Origins, Its Transformations, and Its Prospects*. Harmondsworth: Penguin.
Pickerill, J. and Chatterton, P. (2006). Notes towards autonomous geographies: Creation, resistance and self management as survival tactics. *Progress in Human Geography* 30(6), 730 746.
Pinder, D. (2002). In defence of utopian urbanism: Imagining cities after the 'end of Utopia'. *Geografiska Annaler* 84B(3 4), 229 241.
Pinder, D. (2005). *Visions of the City*. Edinburgh: Edinburgh University Press.
Sandercock, L. (1998). *Towards Cosmopolis*. Chichester: Wiley.
Sennett, R. (1996). *The Uses of Disorder. Personal Identity and City Life* (first publ. 1970). New York: Faber and Faber.
Starhawk (1994). *The Fifth Sacred Thing. A Visionary Novel*. New York: Bantham Books.
Tafuri, M. (1976). *Architecture and Utopia: Design and Capitalist Development*. La Penta, B. L. (trans.). Cambridge, MA: MIT Press.

Relevant Websites

http://www.christiania.org/index.php
Christiania 'Free City', Copenhagen.
http://www.fondationlecorbusier.asso.fr/
Foundation Le Corbusier.
http://www.letchworthgc.com/aboutletchworth/index.html
Letchworth Garden City, designed by Ebenezer Howard.
http://www.urbanpermacultureguild.org/
Urban Permaculture Guild, utopian urban design for environmental sustainability.
http://www.utopia britannica.org.uk/
Utopia Britannica, a comprehensive, historical listing of utopian communities in Britain.

V

Venture Capital

C. Mason, University of Strathclyde, Glasgow, UK

Glossary

Bootstrapping Creative ways in which financially strapped entrepreneurs can minimize or eliminate the need for money to access resources needed for business development: for example, use of personal credit cards, barter arrangements, working from home.

Business Angels Wealthy private individuals – often successful entrepreneurs – who use their own money to invest in new or recently started businesses. Typically they will also provide hands-on support to the businesses in which they invest. Business angels invest at an earlier stage in the development of a business than venture capital firms and their investments are much smaller. Indeed, it is quite common for business angels to provide the first round of external finance to get the company established, and for venture capital funds to provide subsequent rounds of financing. Some commentators have used the metaphor of the relay race to describe this complementary relationship between business angels and venture capital funds.

Debt Finance Finance which a firm raises by borrowing from a bank or other lenders. The borrower will repay the original amount borrowed (the principal) plus an agreed amount of interest over a set period. Bank lending is normally secured. In other words, the borrower has to provide an asset of equal or greater value to the loan (termed collateral) which, in the event of the borrower defaulting on the loan, is seized by the borrower and sold to repay the loan.

Equity Finance Finance that is raised by companies selling shares (stock) to individuals and institutions who become part owners. There are various classes or shares (e.g., ordinary, preference) which have different rights. Holders of equity are entitled to a share of any profits, via a dividend. However, investors in unquoted companies normally seek a return in the form of a capital gain, with the performance of the company enabling the investor to sell the shares that they own for a higher price than they paid for them when they originally invested.

Management Buyout (Termed a Leveraged Buyout in the USA) The purchase of a company, or part of a company (e.g., a subsidiary or division) by its management, with the financial backing of a private equity firm. This occurs in a variety of situations including family-owned companies where there is a succession problem; large companies which are restructuring their business; as a means of privatizing state-owned companies; and companies that are in bankruptcy.

Private Equity This relatively new term is now used to cover all forms of medium-to long-term finance which is provided to companies in exchange for an equity stake. It therefore covers investments at the seed, startup, and early growth stages of emerging entrepreneurial companies through to the much larger types of transactions involving the restructuring of established companies, such as management buyouts and public-to-private transactions. In Europe the terms private equity and venture capital are often used interchangeably. However, in the USA the term venture capital continues to be used to refer to investments in emerging entrepreneurial companies.

Public-to-Private Transactions The purchase of the shares of a company that is listed on a stock market (i.e., a publicly quoted company) by a private equity firm with the intention of taking it out of the quoted sector and turning it into a privately owned firm. This would normally occur when the investor thinks that the company's shares are undervalued by the stock market and that it would perform better as a private company.

Stock Markets Institutions which facilitate and regulate the buying and selling of company shares and other financial instruments. While this may be a physical place (e.g., Wall Street), most transactions are now made electronically. Companies whose shares are traded in this way are termed public, or publicly listed, companies. Most large companies are publicly listed. However, stock markets also include numerous smaller companies who achieve a listing as a means of raising finance and giving their existing shareholders (which may include their employees) an opportunity to trade their shares. A company which joins a stock market is said to be making an initial public offering, or IPO.

Introduction

Venture capital emerged as a specialist form of business finance designed to meet the needs of emergent firms, particularly in technology sectors, which are pursuing significant growth opportunities. The financing needs of such firms typically exceed their capability of generating funds internally, while their ability to attract bank loans (debt finance) is restricted by their lack of collateral and negative cash flows. Indeed, the faster a firm grows the more voracious is its appetite for cash to invest in research and development (R&D), product development and testing, recruitment of key team members, premises, specialized equipment, raw materials and components, sales and distribution capability, and inventories. Venture capital is intended to fill this gap in the supply of finance so that such firms can achieve their growth potential.

Venture capital can be defined as professionally managed money that is invested on a medium to long term basis in unquoted companies in exchange for an 'equity stake'. Investors will share in the upside, obtaining their return in the form of a capital gain on the value of the shares at a 'liquidity event' which normally involves either a 'stock market' listing (often termed an initial public offering, or IPO), acquisition by another company, or the sale of the shares to another investor, but will lose their investment if the business fails. Venture capital investors therefore restrict their investments to businesses which have the potential to achieve rapid growth and gain a significant size and market position because it is only in these circumstances that they will be able to achieve both a liquidity event and a capital gain. However, very few businesses are capable of meeting these demanding investment criteria. Venture capital backed firms therefore represent an out of the ordinary phenomenon.

Although the number of companies that are successful in raising venture capital is small, they have a disproportionate impact on economic development, for example, in terms of innovation, job creation, R&D expenditures, and export sales. The injection of money and support enables venture capital backed companies to grow much faster than the proceeds from sales revenue alone would allow. Moreover, this superior growth rate is sustained over the long run. Venture capital backed companies are faster in developing products and bringing them to market, pursue more radical and ambitious product or process innovation, and produce more valuable patents. It is of interest to economic geographers because venture capital investments within countries are highly spatially concentrated; hence, the economic benefits which flow from such investments are restricted to a small number of favored regions. Moreover, venture capital is one of the key drivers in the growth of technology clusters. It is also an area of increasingly active public policy with governments attempting to stimulate or create venture capital funds as a means of promoting economic growth in less favored regions.

Growth and Evolution

The first venture capital firm, American Research and Development (ARD), was formed in Boston in 1946. However, it was not until the late 1970s that venture capital took off in the USA as a result of a combination of factors operating on both the demand and supply sides, including regulatory changes which removed restrictions on pension funds from investing in venture capital, reductions in capital gains tax, which stimulated entrepreneurship, technological advances, particularly in the communications and information technology and life sciences sectors, which has boosted the level of technology entrepreneurship, fueling demand for venture capital, and healthy stock markets which increased investment returns. Venture capital spread to Western Europe and Israel in the 1980s, and over the past decade has become a truly global activity, with significant growth in China and India in particular. However, the USA, along with Canada, Europe, and Israel continue to dominate venture capital activity, accounting for 93% of global venture capital by value in 2005, with China and India accounting for most of the balance.

The growth of venture capital investment activity has been cyclical on account of the variability in investment returns. For example, the late 1990s was characterized by a huge surge of money flowing into venture capital, attracted by the high returns made by investors who had invested earlier in the decade in the immediate aftermath or the previous downturn. However, this led to increased competition for deals, driving up valuations. The increasing number of business failures prompted a significant downturn in investment activity in 2001, bringing an end to the boom, from which the industry has only recently recovered.

As venture capital expanded – both globally and also in terms of the number of venture capital firms and the amount of money under management and invested – so its investment focus has broadened from its initial concern with young, rapidly growing entrepreneurial companies to include more mature and later stage deals involving ownership change and restructuring of large companies (e.g., management buyouts, family business ownership transitions and public to private deals). This has given rise to a new term – private equity – which includes both the traditional venture capital investing and also investments in larger, established companies.

How Venture Capital Works

Types of Venture Capital Firms

Venture capital firms are financial intermediaries which raise money from investors and then invest it either in young, growing businesses which offer the prospect of high return or in later stage businesses where there is an opportunity to restructure to create value. There are several types of venture capital firm. Most are independent limited partnerships which raise investment funds from financial institutions (banks, pension funds, insurance companies), large companies, wealthy families, and endowments (termed 'limited partners') for fixed life investment vehicles ('funds') with a specific investment focus (location, technology, stage of business development). The function of the fund managers (the 'general partners') is to identify promising investment opportunities, support them through the provision of advice, information, and networking and ultimately exit from the investment. Under this limited partnership model the venture capitalists have discretion over the management of the fund which is normally established with a 10 year life. The majority of the cash is invested over the initial 3 years, with the rest held back for follow on investments. These investments are then harvested in the later years so that by the end of the fund's life it can be liquidated and the proceeds (initial sum and profits) returned to the limited partners. For performing this role the venture capital firm normally receives an annual fee (2–3% of the value of the fund) – which covers running costs – and a 'carried interest', or profit share (between 20% and 30% of the profits generated) which is distributed among the general partners. The general partners would normally seek to raise a new fund some 2–4 years into the life of an existing fund so that they always have at least one fund which is in investing mode.

Some venture capital firms are subsidiaries of financial institutions (termed 'captives') and obtain their investment funds from their parent company. A few large nonfinancial companies, particularly technology companies, have their own venture capital subsidiaries which invest for strategic reasons to complement their own internal R&D activities (corporate venture capital). A few venture capital firms have been established as publicly traded investment vehicles which raise their money through the stock market, in a similar fashion to mutual funds. Most countries also have government backed venture capital funds. These are essentially of two types. The first are public sector venture capital funds established by national and state/provincial governments to invest in specific territories where commercial venture capital is lacking. The second is investment vehicles that governments have established to fill investment 'gaps', particularly the lack of smaller investments, which offer tax and other incentives to private investors. Examples include the Small Business Investment Companies in the USA, the Labor Sponsored Venture Capital Funds in Canada, and Venture Capital Trusts in the UK. Finally, community development venture capital firms emerged in the 1990s to use the tools of venture capital to create jobs, entrepreneurial capacity, and wealth in low income and economically distressed communities.

The Investment Process

From a theoretical perspective the venture capitalist's investment process is guided by the need to minimize two kinds of risk:

1. adverse selection: because of information asymmetries the entrepreneur knows more about their abilities and their business than the potential investor does, creating the risk for the investor of investing in low quality businesses;
2. moral hazard: the risk that having provided the entrepreneur with money s/he will undertake behavior that is in their interests rather than in the best interests of the investor.

Venture capital firms rarely invest in basic innovation or in the business start up stage. Businesses typically fund these stages through one or more of the following: 'bootstrapping', money from the 3Fs (founder, family, friends), government grants and soft loans, and 'business angels'. Venture capital funds will only consider investing in businesses that are at the commercialization stage where many of the uncertainties about the abilities of the management team, market acceptance of the product, size of the market, and profitability of the business have been reduced and where funding is needed to create the infrastructure to grow the business.

Venture capital firms receive investment opportunities from two main sources. The first source is unsolicited deal flow. Venture capital firms are very visible, being listed in various directories and websites. The second source of deal flow comes through the personalized networks of venture capitalists such as bankers and lawyers, entrepreneurs, and other venture capital funds. Most venture capitalists ignore the unsolicited deal flow and focus on those investment opportunities which come through their networks of trusted intermediaries. The quality of this deal flow is superior because it has been filtered by these intermediaries who know what kinds of deals will be of interest. A recommendation from such a source is seen by the venture capitalist as a positive signal about its quality.

Venture capital funds operate a two stage evaluation process, comprising an initial screening process followed by a detailed evaluation process for those deals which pass the initial screening. The initial screening, which is undertaken quite rapidly and intuitively, involves an investor specific screen to ensure that the investment proposal fits the investment focus of the fund (e.g., location, stage, sector, size) and a generic screen to search for features in the proposal which would indicate that closer investigation is likely to be worthwhile. Those deals which pass the initial screening stage are scrutin ized in detail. An attractive investment proposal will rate highly on three components – the concept, management, and returns – but research suggests that it is the quality of the entrepreneur that ultimately determines the invest ment decision. Most of the businesses which seek venture capital do not meet the investment criteria of investors, hence rejection rates are high (over 95%).

Having decided to invest, the next stage is for the venture capitalist and the entrepreneur to negotiate the terms and conditions of the investment. A typical investment agreement for a venture capital fund will involve the following elements:

1. It will give the investor control over key decisions.
2. It will give the investor involvement in the company, typically in the form of one or more seats on the board of directors.
3. It will specify a compensation scheme for the man agement team to align their interests with that of the investor, typically by means of a combination of low salaries and stock options.
4. The investor will use investment instruments which give both downside protection and provide them with a favorable position to make additional investment if the company is successful.
5. Rather than providing the entire sum of money at the outset, it will be staged over time as various milestones are met. This provides the investor with the option to revalue or abandon the investment if new information emerges which suggests that the business appears unlikely to be successful, thereby reducing losses from bad investment decisions. However, it also enables the investor to increase the capital committed in the light of positive information on the company's prospects.

Venture capitalists spend around half their time moni toring and supporting the companies in their portfolio. They use their expertise and networks to make a variety of 'hands on' contributions to their investee companies:

1. recruiting key individuals in the firm;
2. identifying suppliers and customers;
3. help in establishing tactics and strategy;
4. help to raise further finance; and
5. help to structure any transactions (e.g., mergers and acquisitions) that the company might make.

They also provide emotional support to an entrepreneur. Entrepreneurs place most value on the role of their in vestors as 'sounding boards'. However, the actual degree of involvement varies, partly on the basis of investor ethos and partly on account of contingent factors, notably stage, degree of technology innovation, distance, experi ence, and the performance of the investee business.

It is fairly common for venture capital investments to be syndicated between several venture capital funds, especially in second and subsequent funding rounds. There are several reasons for this. First, it avoids over committing a fund to a small number of investments in situations where large follow on funding rounds occur and so enables diversification. Second, it provides the original investor with a second opinion and helps to establish a fair price for the next round of funding. Third, it provides complementary sources of value added. However, any syndicated investment requires a lead investor – usually the initial investor – to perform the key monitoring and support tasks.

The final stage in the investment process is to secure profitable exits for their investments. Venture capitalists do not expect every investment to be successful. Rather, the objective is to maximize the returns of the fund. Indeed, there is a rule of thumb in the industry which expects that out of every ten investments, two will fail, six which will be economically self sustaining but fail to achieve expected levels of growth or exit opportunities (the so called 'living dead'), and a further two will be spectacular successes. It is the presence or absence of these winners which determines the performance of the fund. In Europe, the average returns have been higher for funds specializing in later stage and, in particular, man agement buyout funds than early stage funds (which typically have a technology focus), whereas in the USA the pattern is the exact opposite. As noted earlier, venture capital returns are cyclical, hence the returns are also sensitive to the year in which the fund was launched.

There are three main mechanisms for harvesting investments:

1. A 'public offering' in which the shares of the venture capitalist, and potentially other shareholders, are sold on the 'stock market'. This generally produces the highest returns but this is not unconnected with the fact that only the best companies in a venture capitalist's portfolio go down this route. However, stock markets are cyclical and so cannot always be relied upon to produce the most profitable exit. Moreover, there is normally a 'lockup' period which prevents existing shareholders from selling their shares for a prespecified period after the flotation (usually six months).

2. A ‘trade sale’ in which the entire business is sold to another company. This type of exit has the advantage that investors can sell all of their shares immediately (indeed, are likely to be required to do so) and get paid straight away (in cash or shares of the acquiring company). A trade sale may also be an option for poorly performing investments since the knowledge assets of such firms may have value to a competitor.
3. A ‘private placement’ which involves the purchase of the venture capitalist’s share by another investor. This is less common although there are some secondary purchase funds which specialize in buying the portfolios of existing venture capital funds rather than making their own investments.

Geographical Aspects of Venture Capital

Whereas venture capital firms raise their funding from a wide range of geographical sources, the location of their investments are highly concentrated in space. Information on venture capital investing is collected by, or on behalf of, national venture capital associations and some of this information is publicly available, for example, on the Money Tree^{TM} website. In the case of the USA, venture capital investments are concentrated in California, New England, and New York, and further concentrated at the intrastate level. Equally, there are large swathes of the USA, including much of the south and midwest, which have attracted relatively little venture capital investing. In Canada venture capital investments are disproportionately concentrated in Ontario and Quebec, and within these provinces in the Greater Toronto Area, Ottawa, and Montreal, with the Atlantic and Prairie provinces having the smallest amounts of activity. Turning to Europe, the geographical distribution of venture capital investing in the UK favors London and the Southeast. A number of other West European countries, notably France and Germany, also exhibit high levels of geographical concentration of venture capital investments in just a few regions. Moreover, these patterns are fairly stable over time. In the USA the investment ‘bubble’ of the late 1990s – caused by a large inflow of capital into the venture capital sector, resulting in more, and larger, investments – did lead to a short lived spatial diffusion in investment activity as the increased competition for investments forced venture capital firms to look further afield for investment opportunities. However, in the subsequent investment downturn venture capital firms quickly reversed this geographical expansion in investment activity to refocus on investments closer to home.

This uneven geographical distribution of venture capital investments arises from the combination of the clustering of the venture capital industry in a relatively small number of cities, and the localized nature of venture capital investing.

Venture capital firms are clustered in a small number of cities, typically major financial centers and cities in high tech regions. In the USA venture capital offices are concentrated in San Francisco, Boston, and New York. In Canada, over half of all venture capital firms have their offices in Toronto, while in the UK, nearly three quarters of venture capital firms have their head offices in Greater London.

By itself the clustering of venture capital offices need not necessarily lead to the uneven geographical distribution of venture capital investments. However, making local investments is one of the ways in which venture capital firms can reduce risk and uncertainty. This can be seen at each of the stages in the venture capital investment process: deal flow generation, deal evaluation, and post investment relationships.

At the deal flow stage, venture capitalists rely on their local networks of individuals and organizations, with whom they have built up a high level of trust through repeated interactions, to evaluate those investment opportunities that have passed the initial screening stage. Venture capitalists use these contacts to research the background of the entrepreneurs, the viability of the market, likely competition already in place, or on the horizon and protection of the intellectual property. Considerable emphasis is placed on the people. It is easier and quicker for a venture capitalist to check an entrepreneur’s résumé if he or she is local by using their own personal knowledge and local connections. The quality of information is also likely to be better.

The local focus becomes even more important once an investment is made. Venture capitalists do not only provide finance; they also monitor the performance of their investee companies to safeguard their investment, usually by taking a seat on the board of directors, and support them with advice and mentoring in an effort to enhance their performance. Supporting and monitoring their investments – which is an important part of managing the risk – also emphasizes the importance of proximity. Even though some forms of support do not require close contact there will nevertheless be many occasions when face to face contact is required and the venture capital firm will incur high costs each time a nonlocal firm is visited. Geographical proximity therefore plays an important role in both the level and quality of support that businesses are able to obtain from their venture capital investors. Indeed, it is precisely because venture capital is more than just the provision of capital that geographical proximity is important.

A further consequence of the localization of venture capital firms and their investment activity is that knowledge of venture capital investing varies from place to place. This, in turn, has implications for the demand for venture capital. In areas where venture capitalists are concentrated knowledge and learning about venture capital is diffused through the local business community. Thus, both entrepreneurs and intermediaries, including accountants, bankers, lawyers, and advisers, will have a greater understanding of the role and benefits of venture capital, what types of deals venture capitalists will con sider investing in, and the mechanics of negotiating and structuring investments. And, as noted earlier, the con nections that lawyers, accountants, and others have with venture capital firms means that the businesses that they refer for funding will be given serious consideration. The overall effect is to raise the demand for venture capital in such locations. Conversely, in areas which have few or no venture capital firms, knowledge about venture capital among entrepreneurs and the business support network will be weak and incomplete, intermediaries will lack connections with venture capital firms and, perhaps most significantly of all, will be less competent in advising their clients on what it takes to be 'investable'.

It is important to recognize that long distance in vesting within countries also occurs. However, the effect of long distance investing is actually to reinforce the geographical clustering of venture capital investments, rather than producing a more dispersed pattern of investing, because it mainly flows to areas where venture capital firms are already concentrated. The reason for this is that long distance venture capital investments typically occur in the context of the syndication of investments between nonlocal and local investors. Because of the presence of a local lead investor distance is not important to the co investors who themselves can either be local or nonlocal. Co investors are willing to trust the local venture capital fund to undertake the deal evaluation, monitoring, and support functions, including taking a seat on the board, leaving them to take a purely passive role. If the long distance investors do contribute value added functions then they are of a type that do not require close contact with the investee business.

Although most venture capital firms focus on in vesting within their home country, or region, and will continue to do so, a small number, mainly US based, have adopted a global strategy, investing in foreign companies with significant offshore operations and also directly in foreign markets. However, because of the local nature of venture capital investing, global expansion has typically involved partnering with local venture capital firms. This is the way in which much of the recent growth of venture capital activity in China and India has occurred.

Public Policy Issues

The strong geographical effects that characterize venture capital investing contradicts the economist's concept of perfectly mobile capital markets. Although venture capital firms can, and do, raise their investment funds from anywhere, there are strong geographical constraints on where they make their investments. Given the positive effect that venture capitalists have on new firm formation and growth, as both capitalist and catalyst, the effect of the geographical clustering of their investments, in turn, contributes to uneven regional economic development. Access to sources of venture capital gives young busi nesses substantial first mover advantages, which trans lates into fast growth, profitability, and job creation. For these reasons, governments see venture capital as an essential ingredient in their efforts to promote technol ogy led economic development in lagging regions. However, simply making venture capital available will not create the conditions under which high technology entrepreneurship will flourish. Venture capital needs to be combined with talented individuals who can start and grow companies on the basis of innovative ideas. More over, providing money is only part of the role of venture capitalists. It is difficult for governments to replicate the value added skills of venture capitalists. Trying artificially to create a regional pool of venture capital is therefore unlikely to be effective as a means of generating economic development in lagging regions. However, venture capital will flow to areas with world class science and technology. Policymakers should therefore concen trate on developing the region's technology base, en courage business start up and growth, and enhance the business support infrastructure.

See also: Entrepreneurship; Uneven Regional Development.

Further Reading

Florida, R. and Kenney, M. (1988). Venture capital, high technology and regional development. *Regional Studies* 22, 33 48.
Florida, R. and Smith, D. F. Jr. (1991). Venture capital formation, investment, and regional industrialisation. *Annals of the Association of American Geographers* 83, 434 451.
Fried, V. H. and Hisrich, R. D. (1994). Toward a model of venture capital investment decision making. *Financial Management* 23(3), 28 37.
Gompers, P. and Lerner, J. (2001). *The Money of Invention: How Venture Capital Creates New Wealth*. Boston, MA: Harvard Business School Press.
Green, M. B. (2004). Venture capital in the United States 1995 2002. *The Industrial Geographer* 2(1), 2 30.
Landstrom, H. (ed.) (2007). *Handbook of Research on Venture Capital*. Cheltenham: Edward Elgar.
Martin, R., Berndt, C., Klagge, B. and Sunley, P. (2005). Spatial proximity effects and regional equity gaps in the venture capital market: Evidence from Germany and the UK. *Environment and Planning A* 37, 1207 1231.
Mason, C., Cooper, S. and Harrison, R. (2002). Venture capital and high technology clusters: The case of Ottawa. In Oakey, R., During, W. & Kauser, S. (eds.) *New Technology Based Firms in the New Millennium, Volume II*, pp 261 278. Oxford: Pergamon.

Mason, C. M. and Harrison, R. T. (2002). The geography of venture capital investments in the UK. *Transactions of the Institute of British Geographers* 27, 427 451.

Murray, G. C. (1998). A policy response to regional disparities in the supply of risk capital to new technology based firms in the European Union: The European Seed Capital Fund scheme. *Regional Studies* 32(5), 405 419.

Powell, W. W., Koput, K. W., Bowie, J. I. and Smith Doerr, L. (2002). The spatial clustering of science and capital: Accounting for biotech firm venture capital relationships. *Regional Studies* 36(2), 291 305.

Sorsenson, O. and Stuart, T. E. (2001). Syndication networks and the spatial distribution of venture capital investments. *American Journal of Sociology* 106(6), 1546 1588.

Sunley, P., Klagge, B., Berndt, C. and Martin, R. (2005). Venture capital programmes in the UK and Germany: In what sense a regional problem? *Regional Studies* 39, 255 273.

Venkataraman, S. (2004). Regional transformation through technological entrepreneurship. *Journal of Business Venturing* 19, 153 167.

Zider, B. (1998). How venture capital works. *Harvard Business Review* Nov Dec, 131 139.

Zook, M. A. (2004). The knowledge brokers: Venture capitalists, tacit knowledge and regional development. *International Journal of Urban and Regional Research* 28(3), 621 641.

Relevant Websites

http://www.avcanet.com
African Venture Capital Association.

http://www.asianfn.com
Asian Venture Capital Journal.

http://www.avcal.com.au
Australian Private Equity and Venture Capital Association.

http://www.privateequitymedia.com.au
Australian Venture Capital Journal.

http://www.bvca.co.uk
British Venture Capital Association.

http://www.cvca.ca
Canadian Venture Capital Association.

http://www.ey.com
Ernst & Young: 2006 report.

http://www.evca.com
European Venture Capital Association.

http://www.evcj.com
European Venture Capital Journal.

http://www.nasvf.org
It is the website of the National Association of Seed and Venture Funds, a US organization for public sector and not for profit venture capital funds.

http://nvca.org
National Venture Capital Association: Deliotte 2006 report.

http://www.pwc.com
PriceWaterhouseCoopers, Deloitte, and Ernst & Young all produce annual reviews of global venture capital activity.

http://www.realdeals.eu.com
Real Deals (UK and Europe).

http://www.pwcmoneytree.com
The Money Tree^{TM}: this provides quarterly statistics on venture capital investments in the USA. It also provides a link to statistical information on venture capital investments in Israel.

http://www.cdvca.org
The website for the Community Development Venture Capital Alliance (CDVCA), the trade association for the community development venture capital industry.

http://www.canadavc.com
Thomson Financial: it is a source of statistics and other information on venture capital in Canada.

http://www.venturecapitaljournal.net
Venture Capital Journal (US).

Vichianism (After Vico)

D. Kunze, Pennsylvania State University, University Park, PA, USA

Glossary

Anamorphosis The (usually) visual effect of concealing an image so that it is visible from one viewpoint only. More generally, anamorphosis covers conditions of form-within-form, concealed form, and disjunctive form, even when the form is nonvisual.

Enlightenment The post-Renaissance period of consolidation of rational philosophy and its application; secularization, and emphasis on transparent methods of inquiry.

Ideal Eternal History Vico's three-part schema of human development, realized in historical, psychological, and epistemological forms. The age of gods is dominated by the subject's unwitting metaphorical projection of human nature onto the external environment; the heroic age secularizes the severe religion of the mythic age, paving the way for representative government in the age of men, dominated by conceptualism and technology. The barbarism of the final age leads to a possible 'ricorso', or new cycle.

Positivism The general philosophical movement begun in the nineteenth century, an outgrowth of Enlightenment rationality and a precursor of modern scientific method. Positivism privileges objective truth (the signified – see below) and counts the subject (and the signifier) mainly as a source of potential error.

Projective Metaphor The comparison of knowledge to a light projected from a source (embodying reality) onto various kinds of representational screens (languages, theories, pictures, etc.). In this model, subjectivity is portrayed as distance, culture as distortion or refraction, and falsehood as shadow.

Signifier/Signified Ferdinand de Saussure (1857–1913) divided the function of the sign into signified (the mental idea of something) and signifier (language's phonetic and subsequent graphic means of indicating this mental idea by distinguishing it from other signifiers). Vico and later Jacques Lacan expressed doubt about any positive existence of the signified, and claimed that meaning was based on relationships among signifiers.

Who Was Giambattista Vico?

Giambattista Vico (1668–1744) was an original philosopher of culture who contributed in diverse ways to the intellectual life of Europe from the eighteenth century onward. His theories about culture and human development prefigured those of Hegel, Spengler, Marx, Piaget, and even Freud and Lacan. His idea of sequential stages of history and thought anticipated the metatheories of Hayden White, Stephen Pepper, and Karl Mannheim, where periods of history, schools of thought, political attitudes, and modes of perception are determined by the forms of metaphor. He was promoted by major figures such as the French historian Jules Michelet, the Italian humanist critic Benedetto Croce, the Irish novelist James Joyce, and the British philosopher and critic Sir Isaiah Berlin. Yet, partly owing not only to his style of writing but also to his attempt to frame his theories within a mixture of poetic imagery and autobiographic involvement, Vico's works have been among the most misread and misunderstood of any philosopher. Vico's conception of the human imagination, based on a dynamic model of signification, is perhaps the most original and complex ever conceived. His ability to relate linguistics, psychology, environmental factors, and philosophy within a single theory of culture remains, at least in its ambition and scope, has been unparalleled.

Vico was born in Naples, Italy, the son of a bookseller. He received his early education from local grammar schools and Jesuit tutors. Although he graduated from the University of Naples in 1694 as a doctor of civil and canon law, he characterized himself as a lifelong autodidact. His self instruction began at the age of 7, during his convalescence from a concussion. The attending physician predicted that the boy would either die or grow up to be an idiot. Vico mobilized this omen as a scholarly 'leitmotif'. He characterized his personality as split by the dry humors of melancholy and choler. He argued that history was a parallax view afforded by the twin eyes of geography (contingency) and chronology (necessity). More prophetically, Vico purposefully contrived his major work, the *New Science* (1744), to strike readers as ingenious or idiotic.

Vico's first important works, one on methods of humanistic study and a second on truths hidden in ancient etymologies, appeared in 1709 and 1710, respectively. A major study of universal right was completed in 1720–22, and the first two parts of his autobiography were finished in 1725 and 1731. Vico's major work, the *Scienza Nuova*, or *New Science*, was first published in 1725, with a completely revised edition, the *Scienza Nuova Secunda* (*Second New Science*) appearing 5 years later. A third, slightly revised version was published in

1744, just before his death. This version is also called the *Second New Science*, but it is regarded as the definitive edition. The English translation made in 1948 by Thomas Goddard Bergin and Max Harold Fisch has served as the main reference point for a revival of interest in recent times among a broad range of scholars.

Vico as an Anti-Enlightenment Philosopher of Culture

Vico is typically portrayed as the first important anti Cartesian. He attacked René Descartes' rationalism as abstract and unconcerned with cultural invention, human nature, or historical specificity. However, Vico was a Cartesian in his own way. While Descartes conceived of a subject secure in its own self recognition (*cogito ergo sum*), Vico proposed an equal but opposite subject, an empty place holder unable to witness his/her own nature except by creating networks of symbolic relations with others and nature. Keenly concerned with the mind's dynamic and self constructing nature, Vico defined a necessary logic within and through the contingencies of everyday experience, a necessary sequence of develop mental stages ('ideal eternal history'), which focused on the question of human origins and how the first human thought and language were distinguished from animal cognition and sign use.

Vico's original theory of mythic thought was based on the idea of a *universale fantastico* (imaginative uni versal) by which the first human subjects unknowingly disguised their own natures as demonic elements of the external world. Because the human mind was at first unable to form abstractions, it used metaphors involving the body and the senses. Because of this, Vico held that the first perceptions of all cultures were structurally consistent, a means of thinking through things ('bri colage', as the anthropologist Claude Lévi Strauss would call it).

Vico saw that the creation of culture involved two kinds of inversion of 'near and far'. The first was an exchange of what was spatially near and far. The (unconscious) sense of the body found its antipode in the sky, where ether was the 'animus' of a god as robust and sensual as the first humans, whose voice was thunder and whose indecipherable messages were written in constellations and other celestial phenomena. The sec ond inversion was a moral inversion of private and public. Borrowing from Bernard Mandeville's 'thesis of the bees' (1705), Vico held that humans civilized themselves by inadvertently producing public goods (institutions, customs, protections, etc.) by acting out of private self interest. For example, by grouping together into cities to strengthen the hold of the families over the plebs, the aristocrats instituted a collective civic religion that protected all within the city, laying the foundation for later secularization and representative government.

In ways reinforced through references to mirror images, irony, and metaphor, Vico seems to hold that each cultural institution, act, and object is 'anamorphic' in that it embodies these two exchanges of near and far. 'Anamorphy' is a term typically reserved to describe visual images that are concealed within ordinary images, visible only from a specific viewpoint. For readers of Vico, the idea of anamorphy can be extended to cover (1) Vico's account of the composite nature of human creations and (2) Vico's methods – some of them strik ingly optical – for discovering and deciphering their complexity. Because inversion began with perception, where the extended world of nature was seen in terms of bodies with demonic intentionalities, Vichian com prehension can retrace this process in reverse by paying attention to a topological rather than projective (Car tesian) use of dimensions and distances. As Vico points out, for the first humans the heavens were no further away than the tops of nearby mountains; Hades was as close as the bottom of the furrow. Near was construed as far. Distances were constructed as ritual relations.

Vico's theory of history itself was an anamorph – a 'parallax view' produced by the 'twin eyes' of geography and chronology. Geography stood for the immediate contingent conditions that gave rise to the great variety of human cultures; chronology was the necessary sequence of cultural stages through which all cultures (as well as all individuals, institutions, and even events) had to pass. Each stage was defined in terms of a form of metaphor. In the first, the mythic 'age of gods', humans unknowingly back projected their robust sensuality onto the screen of external nature, regarding physical appearances as divine signs in need of translation. Mythic mentality was purely metaphorical but unaware of metaphor as such. It saw nature filled with literal mes sages from gods to humans. The practice of divination, the first form of knowledge, concealed the human authorship of natural signs.

Divination gave rise to the development of writing, the natural sciences, linear time (through genealogy), and prosaic rather than poetic language. As demonic nature and the cyclopean family were supplanted by models of individual autonomy (the fable of Odysseus' encounter with the Cyclops is a signature of this transition), a heroic age replaced the mythic one. Metaphor yielded to the more logical forms of metonymy and synecdoche. Family based theocracies were secularized into representational governments with written laws. Finally, the power of language to distinguish objects from attributes laid the ground for a modern mentality able to abstract and objectify nature. Human relations could be con ceptualized. Truth could serve logic rather than divine will. But, Vico's account of modern thought returns to the theme of barbarity. Where the violence of the age of the gods came from the body, modern violence is mental, the result of irony and the inversion of the principle of 'pubic goods from private vices'; conscious attempts to do good inadvertently but unavoidably cause suffering.

Vico's historical stages applied also to individual development and even individual experiences, but because of this final ironic downturn, Vico did not privilege modernity as did Jean Piaget, whose scheme of developmental mentality resembles Vico's. Mythic thought was barbaric through the domination of the senses; modern rationalism was barbaric in its addiction to technique and conceptualization. Thought began in poetic wildness and ended in logically justified madness. Yet, there was hope in this symmetry for a return or rebirth ('ricorso'). Clearly, Vico intended the *New Science* as a preparation for this rebirth. The Irish author James Joyce exploited the sequence of the 'ideal eternal history' and ricorso rather clearly in his novels, *Ulysses* and *Finnegans Wake*. The theme of 'ontogeny recapitulates phylogeny' underscores the dynamics and scale independence of Vico's three part schema.

How Was Vico Received?

Vico has meant different things to different ages and ideologies. Modern philosophers have used Vico to counter positivism and technological domination. Academics have used Vico to protest the tunnel vision of disciplinary practices. While German National Socialists in the 1930s extolled Vico for his theory of nations, Marxists in the 1920s and 1950s applauded his account of the rise of the proletariat. The panorama of the *New Science* embraces a wide array of subjects. Therefore it was not surprising that, from the 1960s to the 1980s, prominent scholars from a variety of fields – humanists, philosophers, and social scientists – met in a series of conferences aiming to restore Vico to his rightful place in Western culture.

Geographers typically pride themselves on their ability to synthesize across disciplinary lines, but none participated significantly in these Vico revivals. This is difficult to explain in light of several timely and comprehensive introductions, published in the 1980s, which specifically addressed Vico's value as a philosopher of place. The first was made by British scholar William J. Mills. In 1982, Mills positioned Vico opposite the positivists and focused on the importance of Vico's principle, *verum ipsum factum* (*est*), interpreted variously as 'the made and the true are convertible', or 'humans can have true knowledge of what they have created'. Donald Kunze, an American geographer, responded to Mills by showing how the imaginative universal was an even more useful principle for creating a Vichian philosophy of place. The Mills–Kunze exchange constituted an 'index' of how geographers might employ Vichian ideas in cultural analysis. Key terms in this debate include (1) understanding how human cultures are 'self constituted', (2) the role of the imagination in the formation of mythic, representational, and conceptual thinking, and (3) the role of place as both a result and 'constitutive principle' of these three forms of thought. Where Mills emphasized the contrast of Vichian thought with Cartesian and positivist thought, Kunze focused on the role of Vico's 'imaginative universal', the means by which the first humans unconsciously saw the external world in subjective terms. This ironic ignorance that concealed the artificial element of the natural world from its true human makers was, by Mills, evident in the cultural landscape, but Kunze argued that the Vichian geographic imagination was active especially in the formation of ideas of the wilderness and the cosmos. Kunze emphasized that geography was as Vico said: one of the 'twin eyes of poetic history', which with chronology created a 'parallax view' of every object and event. Thus, geography is poetic even when it is trying to be scientific: first because it, as all thought, must use metaphor in its otherwise logical constructions; second because its awareness in general is structured.

These complementary introductions seemingly had much to offer humanist geographers responding to the growing use of quantitative methods. Humanists perennially faced charges of pastiche because they drew on a bewildering variety of resources. They might have found in Vico the basis for comprehensive, grounded theory. This did not happen. Vico, important for nearly every humanistic and social science, played no significant role in the development of human or humanistic geography. One geographer, John Pickles, wrote persuasively of Vico's potential for providing a revolutionary view of the imagination's role in scientific as well as everyday thought, but his monograph did not gain the attention it deserved. Kunze published a book on Vico as a philosopher of place in 1987; but, as the geographer Yi Fu Tuan commented in an otherwise positive review, the difficulty of Vico's ideas and language puts him and his commentators at a considerable distance from the norms of geographic thinking. The years following these introductions yielded no more than scattered short references by a few sympathetic scholars. Ironically perhaps, the most successful study employing Vico as a philosopher of place is the Shakespearean scholar John Gillies' *Shakespeare and the Geography of Difference* in 1994. Gillies has created a new geographic paradigm by showing, through Vico's *New Science*, how art can serve as *prima facie* evidence of the sense of place and place's role in culture.

Why Vico Did Not Fit into the Humanist Program: Two Versions of Metaphor

Beginning with C. P. Snow's famous distinction of 'the two cultures' in 1956 and Herbert Simon's equally popular *Models of Man* in 1957, the tendency in Anglo American scholarship has been to view metaphor from the perspective of a 'normal' brand of thought that is ethically obliged to follow rules of logic and the aspirations of universal reason. The modern view holds that metaphor use is 'motivated' and, like imagination in general, suspected of a willful indifference to the need for objective universality. This modern view of metaphor, significantly, appears at the last stages of thought's development, after rationality has established its hold as the 'common outlook', and sees metaphor as a mode of faulty thinking, useful only for embellishment.

Vico's originality was to show how metaphor in ancient thought was a necessity and, in many senses, the opposite of modern metaphor. Where modern metaphor forces particulars to conform to a universal, ancient metaphor adjusted the universal to fit the particulars of the cultural and geographical site, literally a means of 'thinking through things'. As the Vico scholar Andrea Battistini put it, Vico inverted the logic of antonomasia, *individuum pro species*. Instead of the particular individual representing the general concept or value (the particular king or queen 'instantiates' the monarchy), the mental idea was adjusted to the perceptual matter at hand. The (unique) individual becomes available as a universal idea, as in 'He was a Rasputin' (meaning that someone exerted uncanny influence over a head of state).

This distinction lay at the center of Kunze's critique of Mills' two introductions, to a geographic audience, of Vico as a champion of metaphor. While humanist geographers familiar with writings of Christian Norberg Schultz, Anne Buttimer, and Yi Fu Tuan were ready to embrace metaphor as a poetic summation of sense data, they were far from ready to accept Vico's more radical model of reversed antonomasia. Popular concepts such as 'mental map' and 'sense of place' served as academic metaphors, where Vico would have accurately predicted that ideas were adjusted to the data at hand. This immunized such ideas from Karl Popper's famous test of the *modus tolens*, refutability. Once in place, these concepts worked as master signifiers that, instead of reducing ambiguity, absorbed and organized the polysemy generated by their various applications. On this level, the terms worked (as Vico might have predicted) as ideological constructs. Scholars unaware of Vico could thus debate the validity and meaning of the weak versions of metaphor used by others (novelists, artists, local inhabitants, etc.) while using the strong version uncritically in their own conceptualizations. Scholars such as Kevin Lynch, David Lowenthal, Roger Downs, David Stea, Peter Gould, and Douglas Pocock, who depended on some version of a 'mental image' of place, continued the empiricist tradition of regarding metaphor as a 'representation' colored by subjectivity. Vico's notion was, if anything, not that.

Does Vico Have a Future in Geography?

Typically, Vico revivals have taken the form of pairing the (unfamiliar) Vico with a (more familiar) contemporary figure. This has forced Vico to be the scholar's silent partner, a Doppelgänger, a secret source who makes things happen but never gets proper credit. Vico could instead be used to reconstruct existing theory. The Slovenian scholar Slavoj Žižek has referred to this process as one of making 'impossible short circuits' that discover, in familiar works, 'more in them than they themselves'. Vico is capable of bringing new dimensions even to thinkers who are seemingly remote from him; but, more significantly, he is able to address the issue of the signifier (the word, image, symbol, or other material indicator of something) through an ingenious, if barely acknowledged method, which could be called an 'optical proof'.

In the semiological terms of signified (idea) and signifier (material indicator), positivistic science reinterprets and emphasizes the signified as an objective reality, against which thought, language, and theory are subordinated as interpretive systems. In this view, one aims to reduce subjective error and thereby increase objective clarity and predictive power. Critics of positivistic science have defended the autonomy of the signifier as such by showing how the idea of authenticity is itself constructed by culture. Central to this debate has been the metaphor of knowledge as a light radiating from some external objective reality toward perceiving subjects. Refracted by culture, diffused and dimmed by subjectivity, and eclipsed by unreason, this model forever puts authority at an all too literal distance. Philosophers such as Michel Foucault have exposed the geometry of this metaphor as ideological appropriation of knowledge for the sake of power ('panopticism'). A few, including Gilles Deleuze, Martin Jay, and Jacques Lacan, have confronted the issue of optics directly by unpacking its presuppositions. Like Lacan, Vico employed optical strategies to counter projective optical metaphors. The frontispiece Vico inserted into the second edition of the *New Science*, just before the book went to press in December 1730, is a case worth examining in detail (**Figure 1**).

Mario Papini, Margherita Frankel, Donald Phillip Verene, and others have provided extensive historical–philosophical accounts of this image, but there is still the need to see specifically how this Rosicrucian style image constitutes an optical proof. Known as the

Figure 1 The frontispiece from Vico's *New Science*, known as the 'dipintura', showing rays from the divine eye of providence reflected off a jewel on the breast of metaphysics onto a statue of Homer, the first poet of humankind. In the clearing, modeled after the first clearings made for ritual observances and also the origin of agriculture, divination, religion, the arts, etc., are shown an altar with symbols of marriage and burial, a plow, a rudder for the birth of commerce, a funerary urn, a tablet with an alphabet, fasces, scales, and other items. Vico discusses all of the elements but one, the helmet of Hermes in the lower left corner, token of the reader's relationship to the text and the image and, like the image used for the *Table of Cebes*, capable of generating madness or wisdom. Reprinted from G. B. Vico, *Principj di scienza nuova d'intorno alla commune natura delle nazione*, 1744.

'dipintura', the frontispiece shows the gaze of the divine eye of providence shining through an azure opening in a cloudy sky, onto a clearing ('eye') in the forest environing the first humans, Vico would seem to prefigure the philosopher Jacques Derrida with this obvious presence of an authoritarian gaze; but, by distinguishing distinctive and separate parts of this gaze, and by relating the divine eye to the eye of the reader of the *New Science*, Vico strikes nearer to Lacan's theory of three gazes, derived from Edgar Allan Poe's 1845 work *The Purloined Letter*. It is constructive to review this short story briefly.

A letter that is potentially compromising for the Queen of France is delivered while the queen is in the presence of the king and the minister D—. The queen immediately realizes the danger; the minister, seeing her anxiety about the letter, realizes it; but the king notices nothing. The minister, wishing to exert power over the queen through possession of this potentially scandalous correspondence, replaces the letter with the one he happens to be carrying and takes the object of scandal to his apartment, where he puts it in a letter bag con spicuously hanging from the mantel. The police, working on behalf of the queen, are convinced that such an important letter will certainly be cleverly hidden and overlook the unconcealed letter when they search the minister's residence. The detective Dupin deduces that the letter will be concealed by not being concealed and, visiting the minister's apartment, exchanges it with a fake while the minister is distracted. That is to say, the king and the police, symbols of our networks of symbolic relationships, take in the scene while allowing certain important elements to remain invisible. The queen and the minister see the so called invisible elements but are caught in the impasse of power relationships maintained by the unopened letter. Dupin perceives the nature of the structure maintained by the 'invisible' letter and is able to act decisively.

In theoretical terms, the first gaze describes the naïve perception of the reader as well as dogmatic application of preexisting protocols (the frontispiece as a memory image, an illustration; but also perception in its primary dependence on concealment). The second gaze corres ponds to traditional theoretical accounts that could benefit from Vichian short circuiting (the visual ray connecting the statue of Homer with the jewel worn by metaphysics, but also the ironic reversal of private vices into public goods; Vico gives an example of this in his revisionistic theory of 'the true Homer'). The third type of gaze empowers any 'new science' that, like Vico's, undertakes an account of the subject without relying on a template of the signified – objective history – as an authoritative ideological source (the ray linking the eye of providence to the jewel; but also the ability of Poe to place important clues beneath the eyes of the reader using exactly the same trick the minister plays on the police and, later, Dupin uses on the minister). Poe and Vico are, in these terms, birds of a feather.

Vico's frontispiece, like the scenery of a stage play or film, can be conceived as a room whose fourth wall is occupied by the apparatus of production and, later, consumption. The viewer's eye encounters another eye, however – the divine eye of providence. Is the viewer's eye, like providence's, also split by reflection? It should be if we follow the recommendation of the *verum ipsum factum* principle. The best way to know something is to make or do it. To pursue this line of advice, the fourth wall must be converted into a screen with a small eye piece for the viewer. On the reverse of the screen, hidden from the viewer, is the image of the frontispiece painted in reverse (**Figure 2**). Instead of a stage set, the viewer sees a mirror, which serves in the role of a metaphysics – a study guide – for the *New Science*. The place of the subject is, in effect, emptied out. The objects of the ideal eternal history dislodge it; as a subject, it no longer exists. Like the first humans, the viewer is at first trapped within the naiveté of the first gaze, then led to a realization of a second, reflected gaze, and finally made aware of a quasi divine, providential gaze useful to theory. Providence becomes predictive power, the *sine qua non* of any science.

This optical narration is a demonstration but not yet a proof. To establish the higher level of validity, Vico provides an element of overdetermination. The reflected scene, viewed naïvely, contains an emblematic inventory of the ideas of the *New Science*, arranged in order and explained one by one in the introduction. One object, however, is not explained: the helmet of Hermes. This textual absence makes this the second *punctum cæcum* (blind spot) of the frontispiece, the point at which, as Lacan would say, the subject is inscribed into the gaze. Was this omission an accident? Or, was it a purloined letter?

Reading the image in the traditional way, the fron tispiece is an exercise in the Classical Renaissance method of memory places as described by Quintillian, Cicero, or the *Ad Herennium*. A place with distinctive parts – a garden, a building, or a landscape feature, for example – is memorized beforehand and words or ideas to be remembered are turned into objects or personifi cations and placed inside or around the memory place. This straightforward view must count the missing de scription of the helmet as an oversight. Perhaps this symbol for commerce, communications, or any of the other functions assigned to Hermes by the classical tra dition was simply forgotten. But, Homer's statue clearly gestures and even looks toward the helmet, and it would be difficult to accept an explanation of accidental neglect. Rather, it seems more likely that, given the esoteric quality of the image as a whole, along with its possible use of the viewer's eye as the eye of providence, a more important role is assigned to Hermes. The reader–viewer has already encountered one *punctum cæcum*, the divine

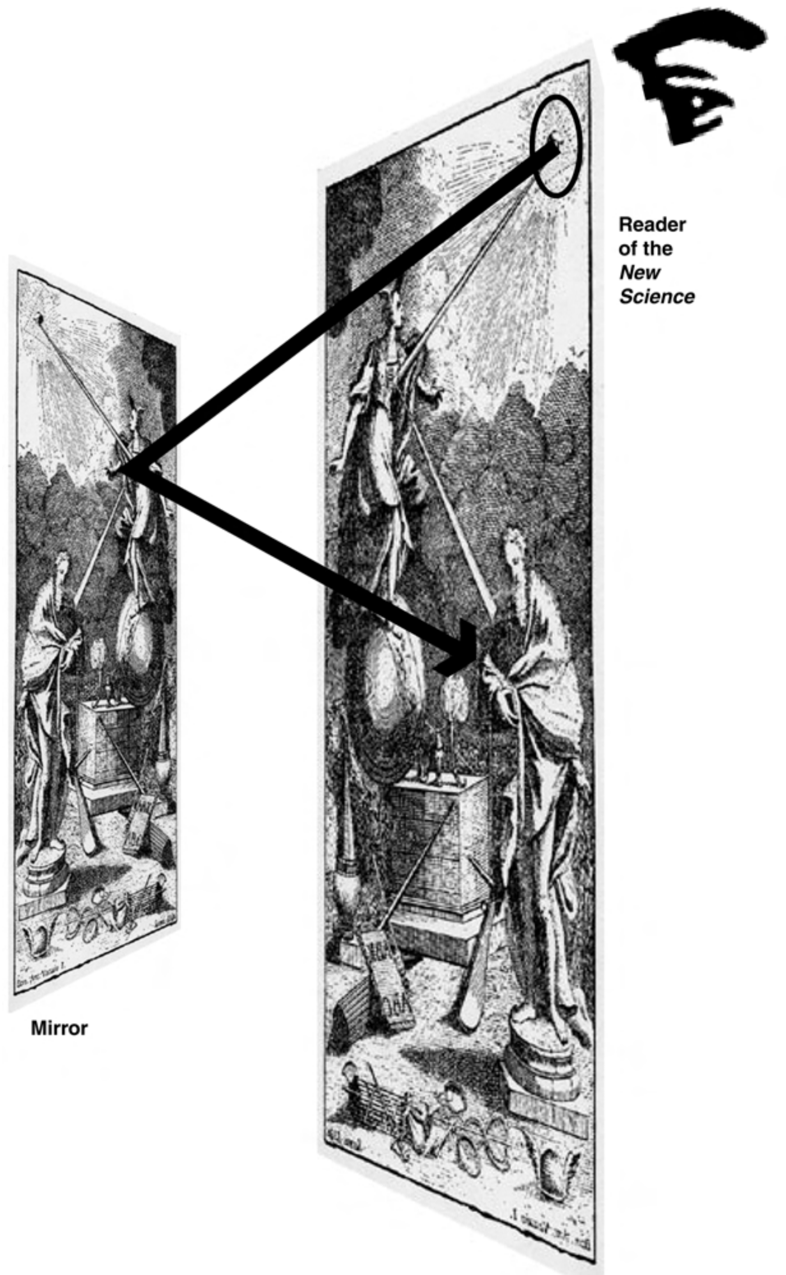

Figure 2 The frontispiece as structured by the three-part gaze. The viewer is reconstituted as a screen upon which the cultural self's component parts are displayed in their ideal order. The viewer takes the place of providence, following Vico's directions in the *New Science* that the reader must narrate the work by realizing the necessary order of his/her own mind. Collàge by author.

eye, which, like the king's gaze in *The Purloined Letter*, sees nothing. The text designates this second blind spot by saying nothing about it.

Together with the mirror action, the silent Hermes adds a dimension of dimensions, so to speak, by referring to the most esoteric (and, in Vico's day, well known) of memory treatises, Giulio Camillo's *Gran Theatro Delle Scienze* (great theater of knowledge). This theater, which some reputed eye witnesses claimed was a full scale architectural device, was detailed in a posthumously published work in 1550, *L'idea Del Theatro*. Camillo aimed at creating a 'universal memory' by simulating a death journey taken by the user. Hermes, the conductor of the souls of the dead, is also the emblem for that journey common to all cultures. Like Vico, Camillo used a three part gaze to explain the logic of his theater. His model was the three part soul of the 'Kabbalah'. The 'Nephes' was the soul closest to the flesh and hence that part of vision that tolerates and even requires blindness; the 'Ruach' was the middle or rational soul corresponding to the theoretical outlook; the 'Nessamah', the highest soul, required separation from the lower two souls in order to act decisively to reach the divine. Camillo drew on ancient traditions linking vision with wisdom in a schema of threes, such as the story of the three Gracchi sisters, whose prophecies came from the single eye and tooth that they passed from one to another. This differentiation and ultimate separation of the three part soul lay the ground for the narrative of discovery of a universal memory. By allying the universal eschatology of the 'journey between the two deaths' to a theory of expanded visuality, Camillo like Vico, was able to create an internal gap within his own theory, a gap to serve as the site of what Lacan would identify as a 'repetitive automatism'. This Freudian symptom is, Lacan explains, embodied by subjects' desire to model their very being on the signi fying chains that run through them – but, this is nothing other than Vico's 'ideal eternal history', which (Vico advises the reader of the *New Science*) will be proved when the reader fuses his own internal mental narrative order with the history's threefold nature.

The journey between the two deaths, commonly portrayed as a fractal like meandering path, as repetitive automatism, is the conceptual heart of the *New Science*. The prototype Vico himself acknowledged specifically, the *Table of Cebes*, reveals the structure and purpose of this automatism. This widely distributed publication was well known throughout Renaissance Europe. The illus trations that usually accompanied it describe a series of enfiladed spaces, sometimes portrayed as a mountain which a pilgrim in search of truth must navigate in order to reach a temple of wisdom on the top center. The prototype of this trial is an even more ancient kind of journey, the 'katabasis' or descent into Hades, is por trayed throughout classical literature. It is the interval between the hero's descent into Hades and return. The Virgilian version, taken from Book VI of the *Æneid*, a source that was extremely familiar to Vico, contains the (James) Joycean theme of an encounter between father and son. It is also the trial of the reader in search of the dead and misunderstood Vico. It is, as the place of the *New Science*, the model of place in its most fun damental form.

Vico's version of Hermes is the figure metaphysics (*Metafisica*), shown alighting a celestial sphere in the frontispiece but again in an image on the title page un available to most readers of the Bergin–Fisch English translation, which shows only the frontispiece. This emblem, known as the 'impresa', portrays a female figure similar to metaphysics (and Hermes) in that she has winged temples and repeats the mirror theme of the frontispiece (**Figure 3**). It was added at the 1744 printing of the *New Science*, apparently taking advantage of the larger page size of that edition.

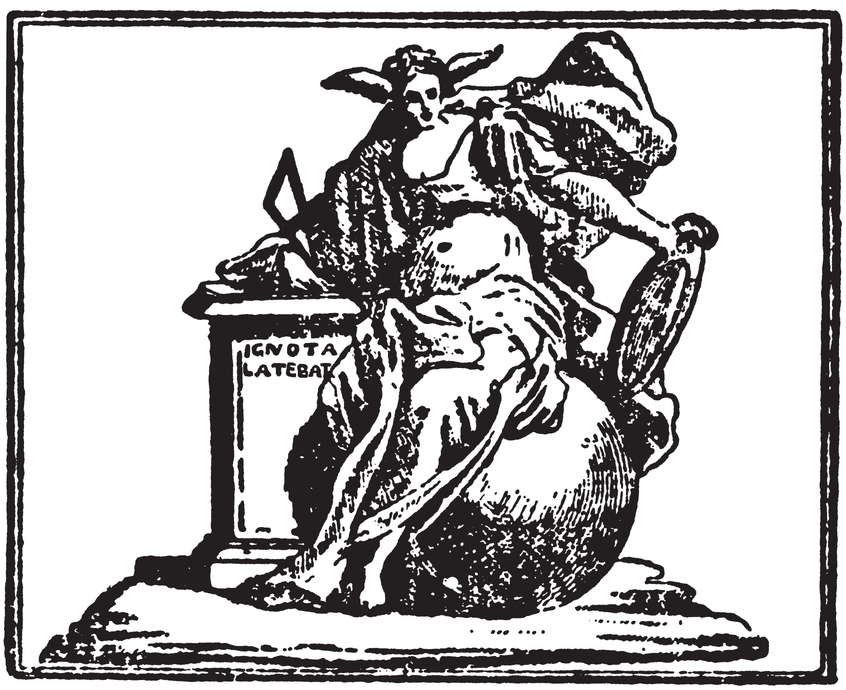

Figure 3 The image known as the 'impresa', showing metaphysics sitting on a sphere, leaning against a plinth or altar, regarding a carpenter's square via a mirror. The inscription on the plinth reads *Ignota Latebat* ('she lay hidden'). This emblem suggests that the truth of the human world, *il vero*, can be viewed only through an angular reflection. Yet, metaphysics grasps the square as well as the mirror, suggesting another parallax relationship (stereognosis) between what Michael Polanyi would characterize as tacit knowing and reasoned, critical articulation. Reprinted from G. B. Vico, *Principj di scienza nuova d'intorno alla commune natura delle nazione*, 1744.

Donald Phillip Verene has, in a detailed commentary on this unattributed image, opined that the impresa is a kind of footnote added to clarify the meaning of the dipintura. It is, he suggests, the initial condition of a before–after comparison: the triangle metaphysics holds in her hand is merged with the circle of the mirror in a virtual, or constructed sense. We see seeing in a way that instructs us how to see the dipintura, that is, through the mirror of anamorphosis, where the twin 'eyes' of geography (contingency) and chronology (necessity) combine to produce a parallax view. The inscription on the plinth/altar next to the sphere where the winged figure sits, *Ignota Latebat* ('She lay hidden') refers both to the anamorphic condition of the *New Science* and to the relation of the divine gaze corresponding to that of the king in Poe's *The Purloined Letter*: seeing without seeing. Working backward, Vico's science is a way of knowing, of building a human world that is known in some latent or tacit sense but not yet known in a way that can serve as the basis for theoretical action. The two remaining parts of the three part Lacanian–Vichian gaze are supplied by the dipintura.

The frontispiece emblematizes the study of place through the creation of a scholarly place, where the optics of scholarly understanding (re)navigate the structure that they wish to understand. This method opens geographical inquiry up to the extensive resources of art, literature, and architecture without the positivist bracketing of these as 'merely subjective'. Vico, thus, having lacked a past or present in the scholarship of academic geography, may have a future based not on inserting him into contemporary work as a silent partner but by recognizing, in him as well as in those who forgot him, the potential of a *New Science* lying hidden.

See also: Humanism/Humanistic Geography; Place; Positivism/Positivist Geography; Semiotics; Sense of Place.

Further Reading

Berlin, I. (1976). *Vico and Herder: Two Studies in the History of Ideas*. London: The Hogarth Press.

Grassi, E. (1990). *Vico and Humanism: Essays on Vico, Heidegger, and Rhetoric*. New York: Peter Lang.

Kunze, D. (1983). Commentary on 'Metaphorical vision: Changes in Western attitudes to the environment'. *Annals of the Association of American Geographers* 73, 153 156.

Kunze, D. (1983). Giambattista Vico as a philosopher of place: Comments on the recent article by Mills. *Transactions of the Institute of British Geographers* 8, 237 248.

Kunze, D. (1987). *Thought and Place: The Architecture of Eternal Places in the Philosophy of Giambattista Vico*. New York: Peter Lang.

Lacan, J. (2006). *Écrits, The First Complete Edition in English*, trans. Fink, B. London: W. W. Norton & Company.

Mazzotta, G. (1999). *The New Map of the World: The Poetic Philosophy of Giambattista Vico*. Princeton, NJ: Princeton University Press.

Mills, W. J. (1982). Metaphorical vision: Changes in Western attitudes to the environment. *Annals of the Association of American Geographers* 72, 237 253.

Mills, W. J. (1982). Positivism reversed: The relevance of Giambattista Vico. *Transactions of the Institute of British Geographers* 7, 1 13.

Pickles, J. (1986). *Concepts and Techniques in Modern Geography*, 44: Geography and humanism. Norwich (UK): Geo Books.

Tagliacozzo, G. and Verene, D. P. (eds.) (1976). *Giambattista Vico's science of humanity*. Baltimore, MD: The Johns Hopkins University Press.

Verene, D. P. (1981). *Vico's Science of Imagination*. Ithaca, NY: Cornell University Press.

Verene, D. P. (1987). Vico's 'Ignota Latebat'. *New Vico Studies* 5, 77 98.

Vico, G. B. (1944). *The Autobiography of Giambattista Vico*, trans. Fisch, M. H. and Bergin, T. G. Ithaca, NY: Cornell University Press.

Vico, G. B. (1968). *The New Science of Giambattista Vico*. Ithaca, NY: Cornell University Press.

Relevant Websites

http://www.us.es
(In Spanish) Centro de Investigaciones sobre Vico: This is the official website of the Spanish Center of Vico Research (Sevilla, Spain). It is a comprehensive guide to materials in Spanish on Vico and Vico scholarship.

http://www.vicoinstitute.org
Institute for Vico Studies, Emory University: This is the official site of the Institute, with links to other sites, information on New Vico Studies (Journal), and information about how to use the institute for research.

Vidal de la Blache, P.

G. Mercier, Laval University, Quebec City, QC, Canada

Glossary

***Genre de vie* (Way of Living)** The concept of *genre de vie* (way of living), as used by Vidal de la Blache, includes all activities, practices, and techniques that characterize the adaptation of a human group to the *milieu*, that is, to the natural surroundings found where the group has settled. By way of the *genre de vie*, the *milieu* acquires a specifically human dimension which, according to Vidal de la Blache, the *paysage* (landscape) will never fail to reveal.

Geographic Personality For Vidal de la Blache, geographic personality, a fundamental concept in the *Tableau de la Géographie de la France* (portrait of French geography) (1903), refers to the ingenuity shown by each human group – and, more specifically, each people or nation – in taking full advantage of the resources drawn from the *milieu* in which it lives.

Vidal de la Blache, Pual (1845–1918)

Paul Vidal de la Blache was the leading figure of the French school of geography that exerted worldwide influence until World War II (**Figure 1**). Along with Friedrich Ratzel in Germany, William Morris Davis in the United States and Halford Mackinder in Great Britain, he was a key architect of geography's rise at the beginning of the twentieth century to the rank of a highly regarded discipline, not only in academic circles but also in everyday social life.

Although Vidal de la Blache was long celebrated for his exceptional contribution to the advancement of geography, his legacy was the target of widespread criticism from the 1960s onward as new notions of the discipline emerged and the 'old masters' came under heavy fire. Little by little, however, criticism gave way to a rather widespread indifference to the lifework of Paul Vidal de la Blache. Indeed, nowadays very few people read his key works *Tableau de la géographie de la France* (portrait of French geography) (1903) or his *Principes de géographie humaine* (principles of human geography) (1922). Hence, in order to describe the thought of Paul Vidal de la Blache, it usually seems sufficient to repeat the few ideas that are summarized and repeated practically by rote in geography textbooks. Nevertheless, despite this relative neglect, Vidal de la Blache's work is currently the focus of solidly documented and more balanced studies that shed new light upon his contribution to the rise of geography at the turn of the twentieth century.

Paul Vidal de la Blache was born on 22 January 1845 in Pézenas, a small town in Languedoc, a French Mediterranean region close to Spain. Sent to Paris at the age of 13 to complete his high school studies, he was then admitted to the *École Normale Supérieure* (national teacher training college) where he studied arts and history. He specialized in ancient Greek history at Athens's *École française*, which directly led to his obtaining a doctorate from the Sorbonne in 1872. In spite of his training in history, he then turned to geography, accepting a position at the *Université de Nancy* that very same year.

Vidal's conversion to geography was tied in with France's recent defeat at the hands of Germany. This defeat, which many – including Vidal de la Blache – interpreted as a sign of France's moral and intellectual decline, culminated in the Treaty of Frankfurt (1871) that saw the winner amputate all of the Alsace region and part of Lorraine from the rest of France. According to the general consensus, French patriotism needed to be regenerated in order for the nation to get back on its feet.

Figure 1 Vidal de la Blache.

As such, the teaching of French geography appeared to be essential as a means of strengthening people's attachment to their national territory. But in order for this to materialize, French geography as a discipline would have to live up to the hopes placed on it. Whereas Germany, through Humboldt, Ritter, and others, had raised geography to the rank of a science, France had fallen seriously behind in this and other scientific fields, a cause of general consternation and dismay. Vidal hence insisted that France should participate actively in the building of a new geography. Thus, Vidal de la Blache, who was appointed Chair of Geography at the *Université de Nancy* in 1872, had the opportunity to take a direct part in the strongly desired renewal of France.

Vidal de la Blache's first works were greatly influenced by Germany's Carl Ritter. Like Ritter, he saw geography at this point in his career as the study of regions whose goal was to discover how each region is marked by the specific contribution it has made to the advancement of humanity. Again, like Ritter, Vidal adhered to the idea that each human society has its own way of transforming nature in order to produce whatever is useful to it, never excluding a capacity for improvement that could lead any given society toward a higher level of civilization. In this context, therefore, the art of geography consists of understanding regional dynamics, which, in part, involves understanding how each region deals with the resources at its disposal. How are they extracted? What products are derived from them? How are they used? How are they traded? On the other hand, geography should also study how each region has improved, and also could improve upon, the ways natural resources are developed. In this sense, it also becomes incumbent upon geography to explain how these regional differences – and the interregional relations resulting from them – condition the distribution and migration of human beings across the surface of the Earth.

Although Vidal de la Blache would never break with this Ritterian heritage throughout his career, he would nonetheless gradually develop his own geographical doctrine in which the concepts of *milieu* (natural surroundings), *genre de vie* (way of living), and *paysage* (landscape) would eventually serve as complements to his notion of the region. As such, Vidalian geography became a method for studying how each human group, through its activities, practices, and techniques, adapts to its *milieu*, that is, the natural surroundings as they exist in the region where the group has settled. This adaptation constitutes a *genre de vie* (way of living) which, in return, leaves its mark upon the group's *milieu*. Hence, by way of the *genre de vie*, the *milieu* acquires a specifically human dimension which, according to Vidal, the *paysage* never fails to reveal.

Another abiding feature of Vidalian geography lies in its naturalism. Although he came from a Catholic family, Vidal de la Blache was an agnostic who saw the human being purely as a component of nature. He also recognized that every human group, regardless of the level of civilization it has reached, can never completely escape the restrictions imposed by its *milieu*, that is, the natural surroundings upon which it depends. However, once this principle had been established, it was still necessary to describe and measure how these restrictions were imposed as a function of time and place. Suddenly, Vidal was faced with two fundamental problems. On the one hand, he felt that it was important to discover which natural elements influence ways of living and landscapes most directly. Concerning this point, Vidalian geography evolved as well. In his earliest works, Vidal gave most weight to climatic conditions, after which he emphasized geological formations, and lastly he considered mainly biological factors, thus drawing his school of thought closer to the form of ecology that was gaining strength at that time.

On the other hand, a more philosophical question was raised as regards the position and role of the human being within nature. In this respect, Vidal de la Blache distanced himself from Ritter by objecting to the latter's teleologism. Whereas the German geographer interpreted human enterprise as the playing out of a divine plan in which, by following the path of progress, the human being becomes the Earth's true conscience, Vidal saw human endeavor as the expression of a simple dialectic of dependence and freedom based on technical skills and abilities. Vidal argued that although human beings are permanently dependent upon nature, we are by no means condemned to be purely a production of the conditions of the *milieu* (natural surroundings) within which we evolve. One of the fundamental human characteristics, according to Vidal, is, in fact, the capacity to improve how we produce and distribute useful things. Thus, by increasing technical abilities, human beings would gradually break free from the restrictions imposed by our natural surroundings. With greater technical skills, humans would be able to take better advantage of natural conditions that previously had to be endured. Thus, by gradually attaining technical prowess, we human beings would construct our own freedom, which does not, however, release us from the natural order, but rather offers us the possibility of becoming active agents of the general causality that energizes nature as a whole (from whence the term 'possibilism' that is often used to characterize Vidalian geography). This said, the aforementioned possibility is by no means the goal of human history. Vidal certainly did not downplay the benefits of human freedom acquired through technical skills; in fact he wanted these benefits to spread as extensively as possible. Nevertheless, according to Vidal, the progress of civilization cannot result from the premeditations of some sort of providence; it will simply be accessible to human beings as we improve how we organize our societies, produce goods, and trade them. From this perspective, humans are nothing other than what we make of ourselves, such that freedom, according to Vidal, is only a contingency, that is, a reality that may or may not materialize, and, depending on the rate of technological development, may either grow, stagnate, or decline.

For Vidal de la Blache, the acquisition of technical skills, the primary factor of freedom, would not be possible without human ingenuity. Moreover, he contended that this acquisition also stems from human will which, in his view, would have little impact if it were not projected by political authorities. This explains why, from his first writings onward, Vidal, following the example of Germany's Friedrich Ratzel, paid such great attention to the State. According to Vidal, the State is not a purely arbitrary product but rather a full fledged geographical entity. He believed that this form of political organization would result from the desire of a people making up a nation to control the territory that it is actually developing. For Vidal de la Blache, state territory is made up of regions that, over the centuries, have established powerful trade relations such that all those living therein have a feeling of belonging to a larger whole: the nation. This is why, in his view, nationalism is not to be condemned in itself, so long as it truly serves to help a people – and through that people all of humanity – progress. But this nationalism, Vidal repeated, needed to draw upon its grassroots, that is, regional dynamism. He also remained forever attentive to factors of regional modernization that in his opinion stem from the spread of large scale industry, the development of transportation modes and networks, the intensification of trade and exports, the influence of cities, and the strengthening of regional institutions. Moreover, Vidal felt, nationalism must not be harmful to other peoples since all human societies, in all ages, have benefited from the contacts and exchanges that have brought them closer together. This is the main reason why he recognized the legitimacy of international cooperation and colonialism, which, in his view, made it possible to transfer the means of progressing more rapidly to less technically advanced population groups.

It should be noted that this view of the State played a key role in Vidal's 1917 analysis of eastern France. Here, Vidal in fact argued that the German occupation of Alsace and a part of Lorraine was illegitimate, since, on the one hand, the inhabitants of this region had long belonged to the French nation and, on the other, because German colonialism had not contributed to the progress of the region. More broadly, on a theoretical level, it should be recalled that Vidal actually saw political authority as the extension of human enterprise, especially in terms of what resulted from the projection of this authority for a region or a people as regards organizational efficiency. This is why Vidal sometimes despaired of French politicians who, he contended, did not always understand that the mission of the State and of other political authorities – especially regional institutions – is fundamentally to promote technical and economic development.

Throughout his career, Vidal de la Blache accordingly built a geographical lifework that, while becoming more complex and balanced, never turned away from the inspiration originally drawn from German geography. Throughout the decades, his works and their influence attested to the fact that France would henceforth fully participate in the renewal of geographical science. As such, he won the confidence and esteem of the authorities and his peers. Thus, from his position at the *École Normale Supérieure*, where he was appointed as of 1877, and then at the Sorbonne, as of 1898, he gradually became an incontestable intellectual leader, training numerous disciples and creating institutions, such as the *Annales de géographie* (annals of geography) in 1892, that promoted the development of the discipline in France.

Although Vidal de la Blache had been able to acquire, in academic, research, and political circles, the reputation of being a leading proponent of the new French geography, it was only in 1903, with the publication of his *Tableau de la géographie de la France* (portrait of French geography) that he gained public fame. It is true that, up until that time, his subtle, erudite, and detailed work was able to meet the intellectual standards of scholars and satisfy the curious, but it had not yet seduced a more general public that was not necessarily interested in the discipline of geography for the storehouse of knowledge it harbored since what the public was seeking in geography was, above all, a compelling idea, the conviction that an essential truth was being expressed through the discipline itself. Another factor could have been that Vidal de la Blache's geography was too close to his German model which, in his eyes, embodied a certain scientific ideal. He was in fact reproached for this. Thus, while not sacrificing the scientific character of geography to the imperatives of patriotism, Vidal clearly managed to reform French geography, but he had not yet imbued this discipline with a full 'patriotic value'. This certainly stems, at least in part, from the themes addressed by Vidalian geography. During the first decades, Vidal in fact neglected France itself, concentrating especially on other parts of the world, the continents (in the first place Europe), and large countries such as China, India, and Russia. Was this not a mistake considering the patriotic mission assigned to geography at the time of his hiring at Nancy? Clearly his compatriots were longing for a geography worthy of its name, but was not a geography of France their top priority? In 1903, with the *Tableau*, the cycle was completed: Vidal had finally given France a geography, not only in order to position the country at the cutting edge of science (of geographic science, in any case) but also so that France itself would be the subject of this geographic science, even its very best illustration. And in 1917, when he published *France de l'Est* (eastern France), just a few months before his death and while his country was facing Germany in a war that had cost him his own son, it became clear that geography was hence forth at the service of the French nation.

Reading the geography of Paul Vidal de la Blache a century after it was formulated, one cannot fail to be surprised by two astonishing paradoxes. The first stems from the fact that the much discussed French school of geography that owes so much to Vidal de la Blache is not, in the final analysis, that different from other national schools of geography that were flourishing in Germany, Great Britain, the United States, and elsewhere at the time. Of course, French geography and the other national geographies each took on a specific coloring. Neverthe less, what we have here are variations on the same theme since, behind this mosaic looms a single fundamental question that at that point crystallized the entire chal lenge of building a scientific geography: how to view the relationship – both in terms of unity and diversity – that is established between human societies and the various regions around the globe? Although the national schools of geography easily reached a consensus when it came to establishing the unity of this relationship based on the principle of humankind's dependence on nature, it was difficult for them, in a context where numerous existing or emerging states were seeking to affirm their legitimacy (by way of economic competition, war, or otherwise) to get away from the idea that each people has, in this regard, its own spirit, or to borrow Vidal's term, a personality.

This is why the geographic science of the period with this issue at the heart of its concerns was so easily able to accommodate the diverse national schools which, by their very differences, corroborated the premise imposed by the circumstances that prevailed. Here we see the second paradox of Vidalian geography. Is it enough to argue, as Vidal did, that the State exists as a means of ensuring regional development, the affirmation of the national personality, and the progress of civilization? Is this not at the very least a naive approach that overlooks the com plexity of human societies, the power issues that brew within them, and the divisions that cut across them? In addition, although Vidal de la Blache was sensitive to the intrinsically political dimension of human insti tutions, he managed to bypass any analysis of this element by making the people the ultimate, even the only, player on the political scene. Even as he broke down this scene into regional, state, colonial, or international components, his doctrine, carrying the stamp of natur alism and distrustful of the fledgling discipline of sociology, in fact remained incapable of characterizing society otherwise than as a fascinating national whole.

See also: Francophone Geography; Landscape; Regional Geography I; Ritter, C.

Further Reading

Andrews, H. F. (1986). The early life of Paul Vidal de la Blache and the making of modern geography. *Transactions of the Institute of British Geographers* 11(2), 174 182.

Archer, K. (1993). Regions as social organisms: The Lamarckian characteristics of Vidal de la Blache's regional geography. *Annals of the Association of American Geographers* 83(3), 498 514.

Berdoulay, V. (1995). *La formation de l'École française de géographie*. Paris: Éditions du Comité des travaux historiques et scientifiques.

Claval, P. (1998). *Histoire de la géographie française de 1870 à nos jours*. Paris: Nathan.

Mercier, G. (2001). The geography of Friedrich Ratzel and Paul Vidal de la Blache: A comparative analysis. *Geography Online* 2(1). http://www.siue.edu/GEOGRAPHY/ONLINE/mercier.htm (accessed June 2008).

Robic, M. C. (1994). National identity. In Hooson, D. (ed.) *Vidal's Tableau de la géographie de la France: From Political Geography to Human Geography. Geography and National Identity*, pp 58 70. Oxford: Blackwell.

Sanguin, A. L. (1993). *Vidal de la Blache, 1845 1918. Un génie de la géographie*. Paris: Belin.

Vidal de la Blache, P. (1889). *États et nations de l'Europe autour de la France*. Paris: Delagrave.

Vidal de la Blache, P. (1896). Le principe de la géographie générale. *Annales de géographie* 5(20), 129 142.

Vidal de la Blache, P. (1898). La géographie politique. À propos de écrits de Frédéric Ratzel. *Annales de géographie* 7(32), 97 111.

Vidal de la Blache, P. (1902). Les conditions géographiques des faits sociaux. *Annales de géographie* 11(55), 13 23. http://classiques.uqac.ca/classiques/vidal de la blache/conditions geo faits sociaux/conditions geo faits soc texte.html (accessed June 2008).

Vidal de la Blache, P. (1903). *Tableau de la géographie de la France*. Paris: Hachette. http://gallica.bnf.fr/ark:/12148/bpt6k898391 (accessed June 2008).

Vidal de la Blache, P. (1904). Les pays de France. *La réforme sociale* 8, 333 344.

Vidal de la Blache, P. (1910). Régions françaises. *Revue de Paris* 6, 821 849.

Vidal de la Blache, P. (1911 1912). Les genres de vie dans la géographie humaine. *Annales de géographie* 20, 193 212 and 289 304.

Vidal de la Blache, P. (1913). Des caractères distinctifs de la géographie. *Annales de géographie* 22(124), 289 299. http://gallica.bnf.fr/ark:/12148/bpt6k89837b (accessed June 2008).

Vidal de la Blache, P. (1921). *Principes de géographie humaine*. Paris: Armand Colin. http://gallica.bnf.fr/ark:/12148/bpt6k89838p (accessed June 2008).

Vidal de la Blache, P. (1926). *Principles of Human Geography*. London: Constable Publishers.

Relevant Websites

http://classiques.uqac.ca
Collection 'the author(s) classic'.

http://gallica.bnf.fr
Daily Press in Gallica.

http://www.siue.edu
The geography of Friedrich Ratzel and Paul Vidal de la Blache: a comaparative analysis.

Visuality

F. MacDonald, University of Melbourne, VIC, Australia

Glossary

Anti-Ocularcentrism A movement in continental philosophy which has sought to describe and thereby undermine the ocularcentric foundations of Western thought.

Oculacentrism The idea that vision, among all other senses, has enjoyed a cognitive primacy throughout Western modernity such that the faculty of sight is uniquely tied up with the production of knowledge.

Phenomenology A branch of philosophy concerned with the intuitive experience of phenomena, based on the principle that reality consists of objects and events as they are perceived in human consciousness rather than being independent of human consciousness.

The Gaze A shorthand term that indicts the faculty of sight as part of the apparatus for the expression or enactment of social power. Most commonly, 'the gaze' has been used to describe the social dynamics of tourism ('the tourist gaze'), imperialism ('the imperial gaze') and, most famously, patriarchy ('the male gaze').

Visual Culture A label that refers to the many expressions of visuality – across perception, cognition, and representation – as manifest in the ever-changing realm of aesthetics, media, and technology.

Visuality A term used to describe the myriad ways in which vision is culturally as much as biologically configured.

Visuality refers to the acculturation of sight. It starts from the premise that vision is cultural as much as biological; as Martin Jay has pointed out "there is no 'natural' vision prior to cultural mediation." How we acquire, interpret, and transform ocular data is always contingent on the cultural and historical context of the observant subject. An optometrist might still refer to a person's defective 'vision' but in the humanities and social sciences this word has largely been displaced by 'visuality', a term which is used to think about the experiential, discursive, and necessarily subjective character of the human senses. Nicholas Mirzoeff has detailed how the earliest appearance of the term can be found in the writings of the Scottish essayist and historian Thomas Carlyle (1795–1881). Carlyle coined 'visuality' as a historiographical device with which events separated by time ('a succession of vivid pictures') could be ordered within a single analytical frame. Its more recent origins, however, are to be found in a philosophical reflection on the status and function of the visual realm throughout human history, a move that has taken place over the last three decades in a nexus of academic fields that includes philosophy, art history, film studies, and cultural studies. Visuality is one of a number of key terms (alongside 'the gaze', 'scopic regime', 'ocularcentrism') in the emerging hybrid field of 'visual studies' which – with its attendant *Journal of Visual Studies* – has sought to bring together old and new ways of handling visual culture, from established domains such as art and architectural theory and history through to photography and new media, as well as drawing insights from social sciences such as sociology and anthropology.

At first glance, geography might seem at the margins of this new field. The 'big names' in visual studies do not, on the whole, identify as geographers. However, from its earliest origins as a formal discipline, geography has consistently engaged with questions of visuality. Indeed, the connection between the faculty of sight and the practice of geography is so strong as to have attained the status of a rather banal truism. Over a 100 years ago, Halford Mackinder – one of the reputed 'founders' of disciplinary geography – wrote that "to be able to visualise is of the very essence of geographical power." So frequently has this claim been made throughout the succeeding twentieth century that it has started to sound a little too glib. Take, for example, the following repetitions of and variations on this basic position. Susan Smith held that "geography is *after all* a quintessentially visual enterprise, traditionally using observation as the route to knowledge, and regarding sight as the measure of truth." In a similar vein, Don Mitchell has remarked that "geographers *for very obvious reasons*, tend to focus on the visual." Daniel Sui has gone further still, claiming that it is "*almost trivial to point out* that geography is [to use Susan Smith's exact phrase] a quintessentially visual enterprise." So safe is this association between geography and the visual that, until recently, it was in danger of becoming beyond serious consideration. Gillian Rose has since brought this train of certainty to a shuddering halt in her provocative paper entitled 'On the need to ask how, exactly, is geography "visual"?' This question, together with its counterquestion (how, exactly, is visuality geographical?), will doubtless set the terms for a great deal of future work in both geography and visual studies.

Geography's Visuality

Let's return for a moment to the hyperbole about geography's own 'way of seeing', the idea that a particular visuality is internal to the disciplinary practice of teaching and research. It is not difficult to find eye-minded geographers. "I suspect most good geography of any stripe, begins by looking" wrote J. Fraser Hart in 1981. In 1979, Yi Fu Tuan went so far as to say that "blindness makes a geographical career virtually impossible." At the same time, however, visuality took on a dangerous, even seductive guise. Tuan worried that using pictures in the classroom could introduce 'bias' or, worse still, 'mislead'. "Visual media may not be able to educate at an analytical level" he wrote, "but there is no doubt that they do entertain. They entertain and amuse for the same reasons that they fail to educate." Visual media, he concludes, "even as they open our eyes, blind us to other realities." From a similar era, Douglas Pocock considered that "geography is to such an extent a visual discipline that, uniquely among the social sciences, sight is almost certainly a prerequisite for its pursuit." This was the sort of tough talk that preceded the cultural turn: a casual certainty about the relationship between geography and the visual unaccompanied by much empirical enquiry.

The advent of the 'new' cultural geography in the late 1980s changed all this, ushering in a greater reflexivity about the precise character of the discipline's engagement with the visual. Central to this movement was the influential historical scholarship on landscape and representation by Denis Cosgrove and Stephen Daniels. In his classic monograph on the idea of landscape since the Renaissance, *Social formation and symbolic landscape*, Cosgrove gave primacy to vision suggesting that geographers were interested "in the argument of the eye." Much of his subsequent work and that of Daniels – encapsulated in their deservedly famous edited collection *The iconography of landscape* – has been concerned with the nature of this relationship. Cosgrove and Daniels used a broadly historical materialist approach to draw attention to the role that Western art has played in the dominant construction of a landscape aesthetic. A prominent theme in this work is, according to Cosgrove, the attempt to link:

> the coevolution in the modern West of spatial experience and conception and of the techniques and meanings of seeing. Cultural landscape may be regarded as one of the principal geographical expressions of this coevolution, whose critical examination is a current preoccupation with cultural geography (Cosgrove).

He considers that there is "a profound connection … between the modern usage of landscape to denote a bounded geographical space and the exercise of sight or vision as the principal means of associating space with human concerns." Throughout this work, it is worth pointing to the undoubted inspiration of art historian and novelist John Berger, whose BBC TV series and book *Ways of seeing* not only brought a critical sensibility to art history but traced the connections between painting and more everyday practices of looking. However, it was the uncompromisingly art historical focus of Cosgrove and Daniels' work that attracted criticism from anthropologists such as Tim Ingold who challenged their definition of landscape as "a cultural image, a pictorial way of representing or symbolising surroundings" on the grounds that it was essentially static, generalizing one experience (representation) into landscape *tout court*. And as we shall later discuss, Ingold's more phenomenological interest in the senses has since acquired a wider currency in geographical treatments of landscape. However, at a time when current research agendas are shifting away from representation, it is worth acknowledging the abiding significance of Cosgrove and Daniels' early work, which not only brought a careful art historical scholarship into geography, but also, more importantly, drew attention to the political status of the observant subject.

If landscape was one major field of empirical geographical work on visual culture, then another was a reflection on the primacy of vision in the history and philosophy of geography. Derek Gregory has suggested that geography "continued to privilege sight long after many other [discipline]s became more – well, circumspect." His book *Geographical imaginations* is, among other things, an attempt to explore geography's own 'empire of the gaze'. At the heart of this thesis is an attentiveness to particular forms of visual practice at work in geographical teaching and research. For instance, another geographer, Joan Schwartz has shown how photography and the ability to represent photographic images through the magic lantern (the precursor to the slide projector) was an essential means of acquiring, ordering, and disseminating geographical knowledge. Amateur, professional, and commercial photographers gave visual form to the empire, making the photographic plate a key referent to be used as art, record, data, or propaganda. It has already been mentioned that Halford Mackinder forged a longstanding association of geography and visual culture, a role that has attracted much recent commentary from scholars in the history of geography. James Ryan has documented how Mackinder, as an early advocate for 'visual instruction' in the teaching of geography, supervised the British Colonial Office Visual Instruction Committee (COVIC) from 1902–11, a body that distributed lantern slides of India for use in British schools. Ryan suggests that Mackinder's geographical gaze had truly imperial intent, allowing citizens "the power of roaming at ease imaginatively over the vast surface of the globe," the 'real geographer' being the 'man' "who sees the world drama as *he* reads *his* morning paper." In a similar vein, Gearóid Ó Tuathail sees Mackinder as an exemplar of the Cartesian model of viewing the world that is both detached and Olympian. For Ó Tuathail, Mackinder's way of seeing stands in for an entire philosophical approach to geopolitics in which the intellectual is independent of, and set apart from, a preexisting social reality along the lines of the simple Cartesian binary between an 'in here' mind/self/consciousness and an 'out there' world of objects. The gaze of the geopolitical theorist is thus neutral and dis embodied and, by implication, blind to issues of personal subjectivity.

The last three decades of geographical scholarship have included an important seam of reflexive work examining the discipline's engagement with visual cul ture, both contemporary and historical. However, it is also the case that much of this literature has not been adequately concerned with 'actual' practices of looking, either empirically in terms of archival or ethnographic work, or – until recently, at least – in terms of geography's engagement with visual art practice. Indeed, as James Ryan has argued, the recent history of geographical thinking on vision has tended toward the iconoclastic, with a mistrust of the visual as a mode of geographical knowledge, centered upon the fear of being 'seduced', 'misled', or of introducing 'bias'. Images have certainly been surrounded by more semiotic hazard lights than other forms of evidence, a mistrust of the visual which is, as Ryan points out, part and parcel of visuality within geography. Such a stance is significantly informed by debates about visuality in philosophy and cultural studies, particularly those surrounding the alleged 'ocularcentr ism' of Western thought. Indeed, much of geography's treatment of visuality could be characterized as part of a wider 'anti ocular' turn that has drawn inspiration from feminism, psychoanalysis, and post structuralism.

Ocularcentrism and Anti-Ocularcentrism

The concept of 'ocularcentrism' describes a tendency in Western modernity to ascribe particular primacy to vision above the other human senses. But why should the modality of sight be singled out for this cognitive pri ority? Take, for example, the sentence: 'I see'. In this, the most ordinary of statements in the English language, there is an alignment of 'the eye' (vision) with the 'I' (the Cartesian *cogito*) to signify rational knowledge. This interplay between the 'I' and the eye has become one of the hallmarks of Western modernity. However, the one thing we cannot see is sight itself. The 'problem' of perception – whether by vision, hearing, or any other sense – is one of the most ancient and profound of philosophical questions, as it has a bearing upon the very constitution of the human subject. The Cartesian con ception of the observer as a seat of awareness, bounded by the skin and set against an exterior world has proved to be a remarkably durable model. It is concerned with the crossing of boundaries and the translation across ideas of 'in here' and 'out there', and poses such basic questions as 'how is the external world rendered intelligible to the self?' Our ability to see has thus become messily entangled with our ability to know. Seeing has a bearing on how we know what we know. The question of how visual evidence structures the stories we tell is one of the foremost problems of epistemology.

This association between sight and sense making is at the heart of the idea of ocularcentrism. The term was popularized by the work of Martin Jay in his book *Downcast eyes* in which he identifies vision as a philo sophical buttress to Western thought that was only chal lenged by an 'anti ocular' turn in twentieth century continental philosophy. To be ocularcentric is to adhere to an objective world independent of and external to human consciousness, to which the observer gains admission by the authority of the eye. Jay argues that this 'Cartesian perspectivalism' – the form of perception that represents space, and the subjects and objects in that space, according to the rules of Euclidean geometry – succeeded as "the reigning visual model of modernity" because "it best expressed the 'natural' experience of sight valorized by the scientific world view." In the last century, European (and particularly French) scholarship has mounted a thoroughgoing critique of ocularcentrism, though argu ably this recent 'denigration' of vision has paradoxically served to reaffirm its centrality. Jacques Derrida, for instance, holds that the binary of darkness and light is the founding metaphor of Western philosophy: "the entire history of our philosophy is a photology, the name given to a history of, or treatise on, light."

Many of Jay's examples are taken from the latently (or explicitly) French Marxist intellectual tradition in which vision is considered to have an undesirable pres ence in Western culture and epistemology. In Marx's own writing the camera obscura is used as a metaphor for the inversion of the truth of appearances. As Jonathan Crary notes, "the very apparatus that a century earlier was the site of truth becomes a model for procedures and forces that conceal, invert, and mystify truth." Vision, according to anti ocularcentrism is the one modality of perception that leads us to objectify our environment, regarding it as a repository of things that exist outside of and alien to our own bodies, and over which we exercise domination, enabled and often heightened by technical prostheses. In this analysis, the eye (or at least a particular historical manifestation of visual experience) has given birth to modernity's bastard twins, positivist science and capitalist popular culture, while other sensory registers such as the ear have enjoyed only virtuous association. This critique of the primacy of vision has gained con siderable purchase within human geography; indeed, one might argue that much of the geographical canon on visual culture has adopted, to a greater or lesser degree, this basic position. As a consequence, by the end of the twentieth century the geographer Daniel Sui was able to identify a transition "from the eye to the ear in geographic discourse"; that is to say from visual metaphors of 'perspective' or 'point of view' to aural ones such as 'voices', 'silences', and 'polyphony'. In line with the anti ocularcentric turn, Sui (2000: 322–343) notices "an increasing disenchantment with visually evocative metaphors," a *fin de siécle* tendency that has been paralleled by a predilection for the aural as it is expressed in terms of 'conversation' and 'dialog'. Under this theoretical regime, the human senses have been ascribed certain moral characteristics. Vision, not surprisingly, has come out of this rather badly: it is the fallen sense associated with abstraction, objectivity, masculinity, coldness, and detachment. Hearing by contrast, is the very embodiment of virtue: it is concrete, subjective, warm, and inclusive. This odd individuation of the senses based on apparent moral characteristics is what the historian of sound Jonathan Sterne has described as the 'audiovisual litany', a model that posits history as something that happens 'between' the senses. The stock of one sense (hearing) rises, as another (vision) falls. There are of course numerous problems with this argument, but it has been the dominant position of the geographical literature in the last 20 years. And the newly discovered affinity for the aural and the haptic is in keeping with this anti ocular trend rather than a departure from it. One of the reasons that this model has persevered for so long is surely the undoubted influence of a feminist critique of vision which has constructed 'the gaze' as a metonymic expression of patriarchal power.

Gendering the Gaze and Situating Vision

This feminist critique is most notably and most memorably present in the work of Gillian Rose who, on reading the Marxist aligned work of Cosgrove and Daniels, identified a "blind spot in the geographer's way of seeing," namely that the analytic of class eclipsed that of gender in their art historical treatments of landscape. In a much cited essay on 'Geography as a science of observation' and subsequently in her book *Feminism and geography*, Rose examines the ways in which the visual landscape has been feminized in the geographical tradition, such that nature and its exploration are conflated with the body of a woman. Rose develops John Berger's critique of the masculine gaze at the nude within Western art to reconsider the pleasurable gaze of the geographer in the field. In doing so, she reveals the gendered social relations between the implicitly male observer and the landscape as woman.

By translating John Berger's critique of the masculine gaze at the nude within Western art to the pleasurable gaze of the geographer in the field, Rose reveals the gendered social relations between the implicitly male observer and the landscape as woman. She identifies "a masculine position [which] is to look actively, possessively, sexually and pleasurably, at women as objects." But for Rose, it is the practice of geographical fieldwork rather than landscape painting which is at the heart of her critique. Rose highlights the ocularcentric nature of fieldwork in geography, as a process of visualization where "the eye [holds] the landscape together as a unit" … "selecting [those] … features requiring elucidation." In the figure of the geographical fieldworker, she discerns a complex tension between "the distanced, disembodied objectivity of science" and the sensitive, artistic, desirous appreciation of scenery. It is a contradiction latent in geography's treatment of landscape, between the production of knowledge on the one hand and the disruptive pleasures of looking that might 'seduce' or compromise the masculine integrity of such scientific 'truth' on the other.

In order to challenge "geography's white, heterosexual masculine gaze, a gaze torn between pleasure and its repression," Rose employs the feminist psychoanalytic theory of Laura Mulvey, who in turn draws upon the work of Sigmund Freud, Jacques Lacan and Simone de Beauvoir. Although this is a formative (and in certain respects, a compelling) critique, Mulvey's theory – which is a significant basis for Gillian Rose's work – has since been troubled by subsequent feminist readings. In particular, Mulvey's basic binary:

> Woman as image/man as bearer of the look
> Female passive object of the gaze/male active viewing subject (Mulvey)

has been challenged on the basis that it unwittingly reifies the dominant constructions of gender that it seeks to critique, leaving women powerless to break from a marginal perspective (as long as men are held to be the chief agents in the construction of the visual field). The most significant attempt to develop a feminist approach to landscape that celebrates the agency of women as observant subjects has been undertaken by Catherine Nash. Nash discusses the work of two women artists who construct the male body as landscape, thus offering an alternative schema of association between visual pleasure and the gaze. Rather than an outright dismissal of the visual pleasures of landscape (whether in oil painting or in geographical field practice), Nash has questioned the way in which the female figure is singularly constructed as passive object of the gaze rather than as the observant viewing subject. "Arguments based upon Mulvey's initial critique" she argues "have tended to ignore the issue of women's spectatorship and reproduce the dominance of heterosexism in understanding gender and visual pleasure." For Nash, then, the dominant paradigm of vision and landscape is open to critical interrogation and practical disruption. Rose's work, like that of Cosgrove and Daniels, has been developed and critiqued in various empirical, theoretical, and political ways. However, it would be unwise to consider their contributions as being safely 'closed', as simply an old part of the 'new' cultural geography. There remains an abiding significance to this work in the sense that it has focused attention on the construction of the looking subject; as Rose points out, it forcefully reminds us that the ways in which the world is made visible cannot be separated from the positions of power and privilege that the observer may occupy.

This stance in turn takes us to another strand of feminist theory often associated with Donna Haraway but which is also at the forefront of many recent geo graphical treatments of visuality. It concerns the inevit ably situated character of vision. In contradistinction to the model of the Cartesian observer, this idea insists that vision 'takes place' through particular situated bodies. When we see, we have a point of view and a field of vision. As Martin Jay noted, "there is no 'view from nowhere' for even the most scrupulously 'detached' ob server." But the consequence of embodying the eye is resolutely geographical: as Haraway has outlined, a feminist politics of vision that argues for an embodied knowledge is also a call for a situated knowledge. That is to say that the grounding of vision in some body is also to locate a view from somewhere. By promoting situated and embodied knowledges, she is arguing against "vari ous forms of unlocatable, and so irresponsible, knowledge claims." Another consequence follows from thinking of sight as embodied and situated: that vision cannot therefore be understood as an abstract or free floating discourse but rather must be examined as an embodied practice occurring in particular places and in specific historical contexts. Geographers have had to think beyond the abstraction of ocularcentrism and the now familiar visual exegeses of art historical scholarship, to more carefully (and empirically) address what it means to see. In other words, the future agenda for visuality in geography will surely be concerned with the messy business of looking itself. This, one would argue, is where some important contemporary work is heading.

Looking Lively: Observant Practice

One of the niggling problems with visuality both within and beyond geography, is that it often requires us to isolate vision from other sensory modalities. The litera ture on ocularcentrism, for instance, treats vision pri marily as a discourse rather than as a practice and neatly distinguishes sight both from other sensory discourses and a wider bodily sensorium. If vision is regarded as the hallmark of modernity, the distinction between vision and other senses is for the most part considered to be unproblematic. And the feminist critique of vision too assumes a safe individuation of the senses, such that the eye can neatly stand in for masculinity; "the eye has a penis," in the preposterous terms of Arthur Kroker. However, there remains a problem here. While much of this work recognizes the discursive nature of vision, it does not also acknowledge that the idea of vision as a discrete modality is itself culturally specific. It is one thing to argue that sight has been given a certain dis cursive priority in the Western configuration of the senses. It is another thing altogether to accept the terms upon which this honor has been conferred; namely, that the senses can be individuated as having a discrete rather than a connective function and that this distinction lies outside the realm of discourse.

It is in this context that some new writing in a phe nomenological (or post phenomenological) vein has made an impact. Especially significant is the work of the an thropologist Tim Ingold. He demands that we consider the wider perceptual system in its totality instead of isolating stimuli specific registers of experience. Under the influence of Maurice Merleau Ponty, Ingold is interested in 'looking and listening' as aspects of one bodily activity rather than two individuated actions. He describes it as a kind of scanning movement, accomplished by the whole body – albeit from a fixed location – and which both seeks out, and responds to whole modulations or inflections in the environment to which it is attuned. As such, per ception is not an 'inside the head' operation, performed upon the raw material of sensation, but takes place in circuits that crosscut the boundaries between brain, body, and world. For Ingold (2000), vision and hearing "are virtually indistinguishable: vision is a kind of hearing and vice versa," an argument which leads him to reject Jay's (1993) thesis that the Western fixation with 'objectivity' is a function of the privileged place of vision. He emphatically dismisses the binary which constructs hearing as "warm, connecting and sympathetic" and defines the self socially in relation to others, while vision is held to be "cold, distancing and unfeeling" defining the self individually in opposition to others. Ingold also dismisses the idea that anthropology can meaningfully differentiate cultures ac cording to the relative weighting of the senses through which people perceive the world around them. The dis tinction between hearing and vision, he argues, is itself a division drawn according to the specific culture of Western intellectual enquiry. "It is hard to avoid the suspicion," he writes, "voiced by Nadia Serematakis, that in the imput ation to non Western 'Others' of heightened auditory (along with tactile and olfactory) sensibilities, they are being made to carry the burden of sensory modalities exiled from the sensory structure of Western modernity on account of the latter's attribution to the hegemony of vi sion." The implications of Ingold's reassessment of vision are at once abstract and profound. Following his critique of

Jay, on the basis that the latter is overly invested in re ducing vision entirely to the realm of discourse, Ingold instead advocates the detailed empirical consideration of actual practices of looking.

Ingold's agenda has gained at least a toehold in some new work in geography. One thinks here of John Wylie's accounts of the body subject in a landscape which depart significantly from earlier work in the geographical tradition. For instance, in an essay on walking the English South West Coast path Wylie explores how "various affinities and distanciations" of self and land scape emerge through the experience of walking. In this account, vision is scarcely divisible from other tactile and sonorous relations; it certainly does not involve the 'cleaving of self and world' that characterize earlier models of landscape. "Elemental encounters with land, sea and sky" he argues "are less a distanced looking at and more a *seeing with*." Elsewhere Wylie reconfigures the visuality of landscape by supplementing the later phenomenological account of Merleau Ponty with a Deleuzian critique. Contra John Berger, Wylie argues that landscape "is not a way of seeing the world. Nor is it something seen, an external, inert surface. Rather the term 'landscape' names the materialities and sensibilities with which we see." The conventional spectatorial view of the gazing subject is thus replaced by an examination of the ontological processes (of depth and of fold) which enable the actualization of landscape. This sort of work, although highly theoretical and abstract, is nevertheless a serious engagement with what it means to see. It is concerned with rewriting the conditions of the gaze.

Very different in style, though sharing certain similar intents, is a recent ethnomethodological investigation by Eric Laurier and Barry Brown, who reject the idea of vision as the cognitive skill of a lone individual and instead conceive of it as the learned outcome of certain communities of practice. For them, the ability to appre hend certain things is learned for specific purposes. Using the example of learning how to see fish under water (a skill acquired, practiced, and transmitted by fly fish ermen), they argue for a detailed and specific inquiry into mundane practical activities, in this case, learning 'how' to see. Not content with writing about their own learning to see fish, Laurier and Brown insist that the reader attempt to acquire this skill through reading the paper – that is, being instructed – an experience which itself generates the 'data' for their research.

In this way, the familiar mode of deconstruction and visual exegesis since the cultural turn – so commonly applied to maps, images, and to geographical fieldwork – is giving way to a more diverse set of inquiries into the sociality of looking, both in theory and in practice. When T. J. Clark dismissively referred to art history's penchant for "hauling … visual images before the court of political judgement," this rather terse description could perhaps have been levelled at human geography in the 1990s. Now, however, the picture is rather different. How these emerging agendas on 'observant practice' will fold back into the key concerns of human geography is as yet unclear, but what has been achieved in recent times is to seriously question what visuality itself means. This has entailed no loss of interest in many of the key themes that might fall under the label of 'visual culture' nor, for that matter, in the political dimensions of visuality. New geographical scholarship is being produced on surveillance culture and urban space; on the proliferation of screens in late capitalism; on new modes of visualization from Google Earth to web based video telephony; as well as more lively historical geographies of past optical instruments such as diaromas, panoramas, and camera obscuras. Geography's engagement with the visual still runs broad and deep.

See also: Cultural Turn; Discourse; Feminism/Feminist Geography; Film; Geopolitics; Landscape; Modernity; Psychoanalysis; Sensorium; Subjectivity.

Further Reading

Berger, J. (1972). *Ways of seeing*. London: BBC/Penguin.
Cosgrove, D. (1984). *Social formation and symbolic landscape*. Madison: University of Wisconsin Press.
Cosgrove, D. and Daniels, S. (eds.) (1988). *The iconography of landscape*. Cambridge: Cambridge University Press.
Gregory, D. (1993). *Geographical imaginations*. Oxford: Blackwell.
Haraway, D. (1991). *Simians, cyborgs and women: The reinvention of nature*. London: Free Association Books.
Ingold, T. (2000). *The perception of the environment: Essays on livelihood, dwelling and skill*. London: Routledge.
Jay, M. (1993). *Downcast eyes: The denigration of vision in twentieth century French thought*. Berkeley: University of California Press.
Mirzoeff, N. (2006). On visuality. *Journal of Visual Culture* 5(1), 53 79.
Nash, C. (1996). Reclaiming vision: Looking at landscape and the body. *Gender, Place and Culture* 3, 149 169.
Rose, G. (1992). Geography as a science of observation: Landscape, the gaze and masculinity. In Rose, G. & Driver, F. (eds.) *Nature and science: Essays on the history of geographical knowledge*, pp 8 18. Cheltenham: Historical Geography Research Group.
Rose, G. (2003). On the need to ask how, exactly, is geography 'visual? *Antipode* 35(2), 212 221.
Ryan, J. R. (1997). *Picturing empire: Photography and the visualization of the British Empire*. London: Reaktion Press.
Sui, D. Z. (2000). Visuality, aurality, and shifting metaphors of geographical thought in the late twentieth century. *Annals of the Association of American Geographers* 90(2), 322 343.
Wylie, J. (2005). A single day's walking: Narrating self and landscape on the South West Coast Path. *Transactions of the Institute of British Geographers* 30, 234 247.
Wylie, J. (2006). Depths and folds: On landscape and the gazing subject. *Environment and Planning D: Society and Space* 24, 519 535.

Relevant Websites

http://earth.google.com
Google Earth.
http://www.sagepub.com
Journal of Visual Culture, SAGE Publications.

Visualization, Feminist

M. Pavlovskaya, Hunter College – CUNY, New York, NY, USA

Glossary

Cartographic Visualization or Geovisualization A set of techniques for representing spatial data for visual examination in a computerized environment.
Exploratory Spatial Data Analysis (ESDA) Techniques for multiple and interactive graphic display of data to facilitate its visual examination.
Feminist Geography Part of human geography and focused on representing adequately women's worlds and transforming research practice in accordance with principles of feminist scholarship.
GIS (Geographic Information Systems) Computer-based technology for storage, management, display, and analysis of spatial information.
GISci (Geographic Information Science) Body of knowledge about representing and analyzing spatial objects and processes in a computerized environment.
Visuality Use of visual information and images in production of knowledge and other social practices.

Geography is a fundamentally visual discipline. Geographers have long studied and understood landscapes, people, and their experiences through observation (using a whole range of devices from the human eye to satellites) and have represented their knowledge of them using images – maps, pictures, and photographs. As a feminist geographer Gillian Rose argued, visuality plays a crucial role in the production of geographic knowledge, although seldom recognized or reflected upon. Because of the substantive and methodological diversity of geography, its strategies of visualization today encompass such varied and even contrasting practices as geographic information systems (GISs) map display, plotting a regression line, tornado activity modeling, photographs of urban mural art, diagrams of theoretical relationships, or mapping qualitative interview data. With the rapid growth of GIS in recent decades, visualization research has become one of the key directions in geographic information science (GISci). It seeks new ways to represent and visually examine digital spatial information. In this GISci context, visualization is variously referred to as cartographic visualization, geographic visualization, or geovisualization.

The trenchant post structuralist, postcolonial, and especially feminist critiques of the role of vision and looking in Western science have fundamentally changed practices of knowledge production in the last two decades. Moreover, feminist scholars including geographers have relied on the intensive use of images in their work that tended to look into aspects of human experience not captured by quantitative methodologies. But it is within the context of GIS based visualization that feminist visualization gained particular prominence in recent years. Inspired by the feminist critiques of both visuality and technology, it attempted to reinvent computer based visualization – and mapping – as a tool for feminist geography. This article focuses on feminist critiques of visuality in geography and feminist GIS based visualization.

In the next section I turn to feminist analyses of vision and looking as a way of knowing. Then I consider visuality in geography and identify a number of feminist visualization strategies. Lastly, I focus on feminist GIS based visualization. The article concludes by discussing examples of feminist visualization from the recent literature.

Feminist Critiques of Vision

Social theorists and human geographers explicitly acknowledged and also profoundly criticized the power of vision, sight, and looking in Western culture and scientific practice. Most critics exposed the privileged position of vision over other senses that facilitated the separation of the object from the subject and helped to produce the neutral, detached, and objective observer – a modern scientist and a controlling master. For example, postcolonial and post structuralist theorists have analyzed vision as a central mechanism in establishing power relations. Michel de Certeau famously exposed the sense of mastery over the city one gained by observing New York City from the top of the World Trade Center. Edward Said examined the colonial gaze – Orientalism – as a set of practices to represent and rule colonial subjects, while Michel Foucault analyzed the role of vision in the rise of Western surveillance practices.

Feminist scholars, however, made a unique contribution. They, too, emphasized the significance of looking as a way of knowing in Western culture but focused on knowledge/power linkages that determine how different subjects use their vision. Not all looking is the same. People can, do, and are allowed to look as well as to be seen differently. Specifically, feminist scholars showed how the gendering of visual practices facilitated the rise of masculinist knowledge production in Western science.

In a number of works, Donna Haraway has argued that the authority of science rooted in its claims to objectivity stems from the construction of a scientist as someone capable of creating a disembodied vision of the world, the view from nowhere. Such a neutral (or objective) way of seeing the world occurs once there is a distance between the knower and the known (or the subject and the object of research) because this distance allows for a detached examination. Various technologies of visualization, from satellites to microscopes, consolidate the illusion of infinite vision – "the god trick" in Haraway's words – and the possibility of objective knowledge. Post structuralist and feminist scholars, in contrast, have argued that knowledge production practices are not neutral but in Western societies are related to the quest for power and express the power of the privileged such as, for example, white middle class men. Furthermore, disguised as scientific objectivity, Haraway shows, a masculinist vision of the world escapes critical examination and responsibility for the outcomes of the practices it legitimates.

Feminists also challenged the related construction of women and other disempowered groups as having embodied subjectivities with no claims to objectivity and, therefore, authority and power. Donna Haraway elaborated the concept of situated knowledge centered around seeing and viewing practices that always originate in a particular location and are produced by embodied subjects. With transcending views no longer possible, all knowledges must be treated as responsible accounts of the world. As a result, masculinist knowledge is no longer objective either but situated as that of a particular ('master') subject – white, middle class, and heterosexual man. As such, it has lost its absolute authority while subjugated knowledges (e.g., those of women or colonial subjects) are enabled to create their partially objective – situated and responsible – visions that can converse with each other and work together. Thus, reflexivity and critical self examination are important aspects of feminist scholarship.

More recently, practices of looking and visual representation have received renewed attention in the humanities and social sciences as visuality studies. They focus on the role of vision and visualization in maintaining and challenging cultural practices. Images of different sorts are considered sites for the construction of difference and hierarchy – as in diasporic visualities in art or in scientific practice.

The Visual in Geography and Its Critiques

The visual in geography is found in several locations. First, geographers have long used vision as a primary method to gain knowledge. Second, geographers have used photographs and other visual means to communicate their knowledge. Finally, geographers have specifically used maps as a key tool to represent territories and the people within them. The feminist visualization draws upon the feminist critique of vision and science (see previous section) and the related feminist critiques of vision in geography. It involves a widespread use of (noncartographic) imagery in feminist research as well as the appropriation by feminist geographers of mapping, mainly GIS based mapping that is both quantitative and qualitative.

While the 'new' cultural geographers such as Denis Cosgrove and Stephen Daniels have problematized seemingly objective representations of landscape as expressions of hegemonic class power and culture, feminist geographers have exposed the role of the masculine gaze in geographical knowledge production. Drawing on feminist psychoanalytical theory, Gillian Rose has shown that, similar to other sciences, the central categories of knowing in geography became masculinist despite the supposedly gender neutral conceptual tools, methodologies (e.g., fieldwork), and subject matter (e.g., landscape). That supposed absence of gender from the discipline was due to the fact that women were absent as both subjects and objects of research.

Yet, concepts of femininity and masculinity have shaped the ways in which geographers have both analyzed and represented the world around them. In particular, the visual examination of landscapes has long been an important analytical tool in geography. This method was the forte of the first modern geographers who were mainly men engaged into scientific exploration within the context of Western culture, a culture in which nature is constructed as feminine in art, literature, and science. On the one hand, nature is seen as a nurturing mother to be awed and adored, and, on the other hand, she is an unruly and seductive female. Furthermore, the unknown territory was associated with the feminine (e.g., a passive object to be discovered, chaos to be ordered, an image to be seen) while the knower clearly possessed the supposed masculine qualities (e.g., an active and rationalizing subject, a looking and ordering master).

Early geographical inquiry depended on fieldwork during which direct visual examination and observation of landscape features served as a fundamental method of discovery. The traveling geographer became a hero confronting dangers during his scientific explorations. But in addition to the quest for knowledge, Rose argues, visual observation also gave pleasure, making the geographer's relationship with the landscape profoundly ambivalent. Discovery combined fear, a desire to conquer, adoration, and attraction. Resolving this ambiguity required maintaining a distance between the scientist and the landscape. That intended distancing, in turn, reinforced the role of vision in the process of knowing. According to Rose, the distanced male gaze played a double role in the foundation of geography as a masculine discipline – as an analytical tool to study landscapes and as a source of pleasure for the geographer.

The masculinist geographic imagination also fed the imperial ambitions of the West. Constructing colonized landscapes as feminine played a key role in sustaining imperial practices by representing both women and landscapes as sites for colonization. Here, male colonizers engaged in heroic conquests of new unknown lands that were also identified with indigenous women. This identification made both lands and their people subject to colonial desire and control. Furthermore, mastery of colonial lands was supported by the creation of maps in which the acquired territories were represented as known and ordered through the lens of colonizers' knowledge. In the case of Ireland, constructing its territory as female made it naturally subject to control by British colonial power, the state, and the Church and also helped sustain a male nationalism.

Although the academy and geography have been significantly transformed by feminist scholarship over the last decades, the masculinist nature of the discipline continues to manifest itself in commonly used visualization practices. Many maps published by geographers and especially those created with a GIS, for example, continue to express the supposedly objective vision of a scientist. Furthermore, Gillian Rose has recently examined contemporary visual practices that geographers use to deliver their findings publicly. Because geography's relationship to images is often taken for granted, the effects of these strategies are ironically overlooked. Rose focused on the role of what she called "disciplinary visualities" in the production of geographical knowledge or how geographers use images (as opposed to a focus on the meaning of the images themselves). She found that particular uses of images (in her example slides) in academic presentations (including image display, audiencing, and space) continue to produce specific notions of truth and position geographers relative to the world in specific ways that yield authority.

Feminist Visualization

Feminist visualization may be defined as the use of images in geographic research that is inspired by feminist sensibility and politics. While feminist geographers continue to deconstruct the hegemony of looking in order to create 'multivocal' spaces where different voices can be heard, they also made an effort to appropriate looking and transform it into a nonmasculinist practice of knowledge production. They began to actively use visual methodologies and images in their work as well as to reflect on the intersections of power that arise during this process.

First, feminist scholars complicated their own accounts of visual representation and exploration. They problematized gaze and vision as solely masculine and representations of the landscape as solely feminine. By disrupting these linkages, they created openings for reclaiming the authority of vision and its use in feminist geographic research. In addition, they created room for responsible and accountable feminisms that do not collapse differences between women but allow them to look and speak from diverse locations. For example, Catherine Nash analyzed work by feminist artists who reversed gender power hierarchies by equating the male body with the landscape and positioning women as observers taking pleasure in a nonexploitative looking.

Second, feminist geographers have employed visual images such as photographs of places and people, drawings, and maps as crucial strategies of representation in their work – in addition to textual analysis. For example, cultural and historical geographer Mona Domosh interprets historical photographs from a feminist perspective to construct linkages between gender relations and urban space in Boston and New York. Also, Gillian Rose has recently analyzed practices of viewing family photographs to explore the 'emotional geographies' of mothers.

Third, feminist geographers turned to cartography as a means to reclaim woman's distinct economic, social, and cultural experiences. In other words, women finally became the object of systematic mapping and began authoring maps themselves. A great example is the 2003 'The state of women in the world' produced by Joni Seager in collaboration with a group of scholars. Having survived multiple editions, this atlas displays country-level socioeconomic indicators that specifically describe women's achievements and struggles worldwide.

Finally, in recent years, Mei Po Kwan defined feminist visualization as a method for feminist geographers specifically using GIS. It is in this capacity that it is becoming increasingly known. While geography has a long-standing relationship with technology, it is technologies for the production of cartographic images that have played an especially important role in shaping geographic visualization practices. Although cartographic practices have always been central to the field of geography, until recently the world of mapping and GIS has remained largely separated from the world of feminism. While feminist geographers made an outstanding contribution to the discipline in the last two and a half decades, they did not directly engage cartography or GIS until the end of the 1990s. In most feminist and post structuralist writings, mapping and maps are used metaphorically reflecting the recent spatial turn in social sciences and the humanities. This relatively recent direct engagement with mapping and GIS, discussed below, is nonetheless potentially very fruitful given the significance of feminist analyses of vision and the centrality of visualization for geospatial technologies.

From Map Communication to Geovisualization

Visualization of spatial information has been inseparable from cartography and mapmaking. Today, it constitutes a core functionality of GIS and other spatial and nonspatial multimedia technologies. As such, digital visualization provides an entry point for feminism into a field of geospatial technologies. Before the advent of digital data processing, spatial data could only be visualized in the form of paper maps and graphs, the production of which was very labor and cost intensive and required special ized training and expertise. Its primary goal was to communicate map data to the public as Robinson and Petchenik explained in their 1976 map communication model, to convey information from the scientist and/or a cartographer (i.e., author) to the map user (i.e., reader). The meaning of visualization has changed as digital datasets and analytical tools for their analysis and display (e.g., GISs) have rapidly spread throughout the 1990s and geocomputational power has dramatically increased.

The first major consequence of these innovations in cluded the use of mapping not only for the presentation of a final product but also as a tool for data exploration and analysis (e.g., exploratory spatial data analysis or ESDA). Data visualization (or geovisualization or carto graphic visualization) became an integral part of the research process (as in Allan MacEarchren's concept of cartography cubed or C^3) along with spatial analysis itself. The second was the unforeseen democratization of the practice of mapping that has transformed the map user into a cartographer her/himself. This shift is in creasingly possible as more software tools and data have become accessible to a wide range of users both in terms of design and cost. The third is that the notion of spatial data and its analysis have also changed. In addition to numerical datasets, today's technologies not only provide us with radically new symbolizations and color schemes but also incorporate and visualize various qualitative information such as nonmap images and pictures, three dimensional (3 D) representations, sounds, voices, and movies. These developments have been recently enriched by the ideas about representation, image, and visuality stemming from feminist, critical, post structuralist, and postcolonial scholarship. While these ideas have gained prominence in the social sciences, including human geography, some time ago, their encounter with GIS based geographic visualization is more recent yet is already and increasingly fruitful.

Critical Cartography and GIS

Despite the turn toward cartographic visualization as an important scientific exploration practice as well as the attempts, similar to other post positivist scientific trends, to, for example, incorporate uncertainty, visualization researchers have not questioned the production of knowledge through mapping *per se*. The interrogation of maps as imbued with and mediating power relations originated in the work of the historian of cartography J. B. Harley. Drawing on Foucault and Derrida, Brian Harley laid out an agenda for deconstructing the map as a power knowledge practice as opposed to a scientific fact and initiated what is today referred to as 'critical cartography'. In his work, maps emerged as social con structs employed by those in power to sustain relations of domination, especially those of empire. More recently, critical cartography has been complemented by a grow ing critical GIS literature that examines the societal ef fects of GIS technologies and the mapping practices it enables. As GIS expanded in the 1990s, human geog raphers began to critically examine the power relations embedded within its design, application, use, and the images it produces. In their view, GIS was exclusively a positivist technology that supported established hier archies of social power. Its social impact was enhanced by the unprecedented rhetoric of the digital map in creation of which visual and computing power powerfully combined.

More recent critics of geospatial technologies no longer read them as incompatible with nonpositivist practices of knowledge production. Maps and GIS dis plays are no longer seen as only suitable for the visual ization of quantitative scientific data, bearing only the message from those in power and serving solely tech nocratic, patriarchal, imperialistic, capitalist, and mili tarist goals. Instead, these authors argue, the meanings of maps and of geospatial technologies are actively constructed through contradictory social practices. Moreover, the control over technology, and therefore, representations that it enables, increasingly (also not without setbacks and contradictions) shifts to those out side the established networks of power. For example, map readers have been recognized as important as map authors in the construction of a map meaning while the democratization of GIS is collapsing this difference altogether. Not only does GIS continue to serve the holders of power, it also can empower communities as the vibrant research on public participation GIS (PPGIS) demonstrates.

Feminist Critiques of GIS

In their critique of vision, feminist scholars have emphasized the role of visual technologies in enabling particular types of knowledge production. In Donna Haraway's opinion, these technologies (e.g., satellites and microscopes) are prosthetic devices that people increasingly use to augment their vision of the world. Her point, however, was that these technologies, contrary to masculinist scientific claims, do not make vision more objective; rather, they expand what a human eye can see. They help, therefore, to incorporate additional – although equally subjective – capacity to see, this time assisted by the machines. Therefore, Haraway argues, it is important to understand and master visual technologies because it could both enhance partial objectivity of a situated (feminist) knowledge and show the embodied character of masculine science. Furthermore, understanding technology also enables feminist scholars to practice its use reflectively and responsibly such that the very basis for domination and social hierarchies is undermined as in the new world of information that Haraway describes in her *Cyborg Manifesto*.

As the discourses of feminism and GIS have recently begun to speak to each other, feminist geographers have fruitfully addressed, among other things, the visual power of GIS. They proceeded in two ways – by deconstructing this technology's visual practices and by appropriating its visual power for the purposes of feminist research. In the first instance, and similar to social theorists in the 1990s, they provided excruciating critiques of mapping and GIS as being the tools for masculinist – and positivist – science and a male dominated professional field. Similarly, these early critiques came from the outside of the GIS field *per se*. Cartography, GIS, and remote sensing (RS) were exposed as having the same masculinist biases as quantitative and spatial science methods (e.g., reliance upon a Cartesian grid to organize space, preference for quantitative data, reduction of lived worlds to a pixel or a point) that do not capture but marginalize women's experiences. But additionally, viewing territories 'from above' with help of maps and commercial satellite images and aerial photos provides the viewer with a sense of mastery and in the case of Third World countries contributes to their appropriation via satellite imagery. Later feminist critiques, however, originated from a group of feminist geographers who were trained in GIS themselves. While they challenged the masculinist and positivist nature of GIS technology, they refused to discard GIS altogether. Instead, they started developing alternative uses of GIS and worked to create alternative visualizations, which is more difficult than critique.

Feminist Visualization in GIS

Among the feminist geographers who worked with GIS and began rethinking the field in feminist terms, Mei Po Kwan was the first to claim GIS as specifically suitable for feminist visualization and, therefore, as a method for feminist geographic research. In a key article published in the *Annals of the Association of the American Geographers*, she drew on work of Donna Haraway, Gillian Rose, and other feminist scholars as well as science studies literature, to argue that GIS is not antifeminist by its nature; rather, prevailing GIS practices define it as a masculinist and positivist technology. Her's and other feminist geographers' work demonstrated that GIS visualization, despite its current limitations, can address several concerns in feminist research. First, it makes visible women's worlds and the everyday practices of people, something that conventional GIS does not address. Second, it uses nonstandard datasets that are often generated through in depth fieldwork and, therefore, can answer questions that commercial datasets cannot. Third, it pioneered the incorporation of qualitative and nonquantitative data into GIS. In this way, GIS becomes a key method in mixed methods projects and can be used in critical geographic research. Fourth, feminist visualizations also can incorporate sound and movement in order to capture complex human experiences and challenge the primacy of vision itself.

The author ends this article with examples from feminist GIS based visualization that illustrate the above points. In a number of studies, Mei Po Kwan produced spatial representations of women's daily lives using space–time geography. Kwan combined Hagerstrand's ideas on mapping the daily paths of individuals in space and time with a feminist focus on gendered experiences. Information about women's daily travels was collected from dairies written by the research participants. Kwan combined this data with information from commercial datasets representing urban space such as road networks and urban opportunities. She then analyzed the data in a GIS enhanced by dedicated algorithms. The result is a series of 'body maps' that represent the life paths of women within the urban environment (**Figure 1**).

Kwan showed that while women's life paths are very different from those of men, women's space–time experiences also differ from that of other women depending upon, for example, class and race. The daily paths of African American women in **Figure 1** are clearly the most restricted thereby limiting their access to urban opportunities. By creating this visualization in addition to geocomputational analysis, Kwan made ready for visual examination the striking differences in space–time activities. This type of work would not be possible without either GIS or feminist scholarship.

In another example, Kwan combined image, sound, voice, and movement in order to visualize in a multimedia environment the emotional experiences of Muslim women after 11 September. In particular, Kwan models the daily movement of these women through urban space in a 3 D GIS. The route taken by one woman is presented as though through her eyes and accompanied by this woman's voice as she talks about her experiences of fear in the public spaces she visits on a daily basis.

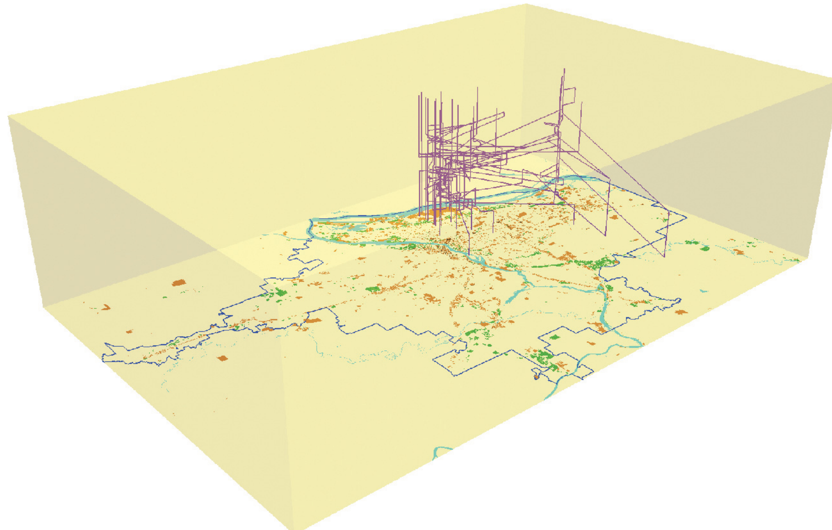

Figure 1 The time space paths of a sample of African-American women in Portland, Oregon. Reproduced from Kwan, M.-P. (2002). Feminist visualization: Re-envisioning GIS as a method in feminist geographic research. *Annals of the Association of American Geographers* 92(4), 645 661, with permission from Blackwell.

Different parts of the 3 D model of the city are color coded to reflect the level of anxiety experienced by the respondent.

The author's own work on the multiple economies of Moscow households employs GIS to visualize and analyze their changing the daily economic practices of households in the context of rapid urban privatization. These practices are gendered and many of them are informal and not accounted for in official statistics (e.g., domestic work or help received through social networks). The author has collected information about these practices through qualitative interviews and using this data visualized the diverse economic spaces of Moscow households both before and after the transition from state socialism. On maps comparing these two periods, each household location was represented by a circle whose size and color corresponded to the total number and composition of economies in which the household participated. The map in **Figure 2** is similar but focuses just on the informal economies (both mon etized and nonmonetized resources) of households.

Clearly, informal economic practices were prominent in the economic landscape of Moscow both under the Soviet system and after but their role remains ignored by conventional economic approaches and transition theories as well as state sponsored data collections. Revealing and mapping the multiple economies of households is a feminist issue because the lion's share of informal and unpaid work is domestic production and is carried out by women. Visualizing these subordinated economies through mapping has ontological power; mapping makes them present in the landscape, it pos itions them in space, and shows what they are like. Once visible, they exist and can no longer be ignored.

Another example illustrates how GIS visualization can empower communities in partnership with feminist geographers. Sara McLafferty describes how women on Long Island, New York organized to draw attention to the high incidence of breast cancer in their communities which they believed was related to environmental factors. They independently collected information about inci dences of illness and showed their location on a 'pin map'. Then researchers from Hunter College transferred this information into a GIS that allowed querying the spatial distribution of breast cancer cases relative to a number of environmental factors. As a result, previous conclusions about the individual characteristics of respondents as risk factors were revised. A powerful visual and technology based statement about the problem pressed the author ities to undertake further research.

Finally, feminist artists have also turned to the powerful visuality of GIS and RS images. For example, Ursula Biemann has created multi media (video/audio) narratives about women transported and trafficked from Eastern Europe and South east Asia to countries where demand exists for their sex work (e.g., Europe and the US). In these flows produced by globalized capitalism, these women are especially objectified as commodities and often deprived the basic human rights and abused. In order to re contexualize their experiences, Biemann combined highly personal stories of these women with the remotely sensed images showing countries and places they came from and went to. The global flows of women fueled by a transnational capitalist economy demanding their bodies involve unprecedented numbers of voluntary and forced migrants. The artist thinks of her work as a counter geography that maps 'female geobodies' and emphasizes that women's personal lives are intimately

Circle size corresponds to the number of informal resources at each household

Monetary resources: additional and informal wages, help with money from networks

Nonmonetary internal resources: domestic production of goods, food, and services (housework)

Nonmonetary network-based resources: help in kind, with childcare, and other services

Figure 2 Informal resources and households, 1989 95. Reproduced from Pavlovskaya, M. E. (2004). Other transitions: Multiple economies of Moscow households in the 1990s. *Annals of the Association of American Geographers* 94(2), 329 351, with permission from Blackwell.

connected to seemingly impersonal global circuits of capital. These connections, however, are made invisible by scientific representations of the Earth, such as those originating from satellite imagery. These images ignore and mask the gendered experiences that are fun damental to the functionings of the transnational econ omy. The description and images from Biemann's work can be found on the Internet.

Conclusion

In conclusion, feminist visualization has thrived in a number of spaces, including prominently the field of geography and geospatial technologies. It has roots in feminist critiques of science and the role of vision in scientific research and particularly in geography, a very visual discipline. The rise of visual geospatial technolo gies such as GIS has crucially contributed to its devel opment in geography. Critical cartographers, critical GIS scholars, and feminist geographers have all critically examined these technologies and representations that they enable to create. Feminist geographers, however, made a particularly fruitful effort to change GIS and use it to create visualizations inspired by feminist scholarship.

See also: Critical Cartography; Critical GIS; Feminism, Maps and GIS; Geovisualization; Postcolonialism/ Postcolonial Geographies; Poststructuralism/ Poststructuralist Geographies; Visuality.

Further Reading

Cosgrove, D. E. (2001). *Apollo's Eye: A Cartographic Genealogy of the Earth in the Western Imagination*. Baltimore, MD: John Hopkins University Press.

Edsall, R. M. (2003). Design and usability of an enhanced geographic information system for exploration of multivariate health statistics. *The Professional Geographer* 55(2), 146 160.

Haraway, D. (1991). Situated knowledges: The science question in feminism and the privilege of partial perspective. In McDowell, L. & Sharp, J. (eds.) *Space, Gender, Knowledge: A Reader for Feminist Geographers*, pp 183 201. London: Edward Arnold.

Harvey, F., Kwan, M. P. and Pavlovskaya, M. E. (eds.) (2005). *Special Issue: Critical GIS*. *Cartographica* 40, 4.

Kwan, M. P. (2002). Feminist visualization: Re envisioning GIS as a method in feminist geographic research. *Annals of the Association of American Geographers* 92(4), 645 661.

Kwan, M. P. (2002). *Special Issue: GIS and Feminism. Gender, Place and Culture: A Journal of Feminist Geography* 9(3), 261 262.

Kwan, M. P. (2007). Affecting geospatial technologies: Toward a feminist politics of emotion. *The Professional Geographer* 59(1), 22 34.

McLafferty, S. L. (2002). Mapping women's worlds: Knowledge, power, and the bounds of GIS. *Gender, Place and Culture: A Journal of Feminist Geography* 9(3), 263 269.

Pavlovskaya, M. E. (2002). Mapping urban change and changing GIS: Other views of economic restructuring. *Gender, Place and Culture: A Journal of Feminist Geography* 9(3), 281 289.
Pavlovskaya, M. E. (2004). Other transitions: Multiple economies of Moscow households in the 1990s. *Annals of the Association of American Geographers* 94(2), 329 351.
Pavlovskaya, M. E. and St. Martin, K. (2007). Feminism and GIS: From a missing object to a mapping subject. *Geography Compass* 1(3), 583 606.
Rose, G. (1992). Geography as a science of observation: The landscape, the gaze, and masculinity. In Driver, F. and Rose, G. (eds.) *Nature and Science*: *Essays in the History of Geographical Knowledge*, Research Series, No. 28, pp 8 18.
Rose, G. (2003). On the need to ask how, exactly, is geography "visual"? *Antipode* 35, 212 221.
Schuurman, N. (2002). Women and technology in geography: A cyborg manifesto for GIS. *The Canadian Geographer* 46(3), 262 265.
St. Martin, K. and Wing, J. (2007). The discourse and discipline of GIS. *Cartographica* 42(3), 235 248.
Sui, D. Z. (2000). Visuality, aurality, and shifting metaphors of the geographical thought in the late twentieth century. *Annals of the Association of American Geographers* 90(2), 322 343.

Relevant Websites

http://www.geobodies.org
Ursula Biemann's gender and geography site: artwork by Ursula Biemann.
http://geog www.sbs.ohio state.edu
Website with geovisualization images by geographer Mei Po Kwan.

Voluntary Sector

C. Milligan, Lancaster University, Lancaster, UK

Introduction

Renewed interest in the voluntary sector, particularly since the 1980s, has generated a variety of terms to describe it. Examples include the voluntary and com munity sector, nongovernmental organizations (NGOs), the third sector, the informal sector, the independent sector, organized civil society, and the social economy. Each term has particular currency in differing national, social, policy, and academic contexts. Attempts to define the voluntary sector have also been the subject of much debate. However, while many attempts have been made to refine these definitions, at its broadest level it is widely accepted that it refers to that sector of society, which encompasses formal, nonprofit distributing organizations that are both self governing and constitutionally in dependent of the state. Though voluntary organizations may employ paid staff and receive funding from the state, their remit is to act for public rather than shareholder benefit.

Like the private and statutory sectors, the voluntary sector covers a vast range of activities from the social and cultural to the environmental and political. By far the largest grouping of voluntary activities encompasses health and social welfare. It is within this field of activity that the earliest references to geographies of voluntary activity can be found. During the 1970s and 1980s, work by Julian Wolpert, Jennifer Wolch and others examined the spatial dimensions of voluntary activity. Based on studies undertaken in the United States, they demon strated that active volunteering was undertaken primarily by people residing in more affluent communities and that their employers were also important philanthropic donors. Economic restructuring in the 1970s and 1980s, combined with the flight of social capital from these more affluent communities, however, resulted not only in a strategic withdrawal of the volunteer sector but also in a reduced tax base for the support of public sector wel fare services within inner city neighborhoods. The re location of employers and employees left a residual population with increased welfare needs, but with a sig nificantly reduced public and voluntary sector infra structure with which to meet those needs. This demonstrates the importance of understanding how changing social and economic influences contribute to the spatial patterning of voluntary activity. The 'volun tary sector rich and voluntary sector poor' explanation for the uneven spatial distribution of voluntary sector activity has provided a baseline thesis against which much subsequent work on the geography of the voluntary sector has been examined and assessed.

The following sections explain how spatial differences in the development of voluntary activity occur; an under standing of the reasons behind these variations and the implications for people and places are important themes for geographical research at a range of spatial scales.

Spatial Distribution of Voluntary Activity: The Global Context

The growth of political and academic interest in volun tarism has mainly been associated with the restructuring of welfare policies in industrialized countries. Recently, however, interest has focused on the rise in the global not for profit sector. This sector is now a major global economic force whose income – were it to be combined – would make it the seventh largest economy in the world. So great is its impact, that it has been suggested that the global rise of the voluntary not for profit sector in the late twentieth and early twenty first centuries could prove to be as significant a development as the rise of the nation state at the end of the nineteenth and early twentieth centuries.

Within development studies, voluntary organizations are more commonly referred to as NGOs and can range from large, international charities – mainly Northern based (NNGOs) – such as Oxfam, CARE, World Vision, and Médicins Sans Frontières, to national and small grassroots or community based self help groups that have emerged in the South (SNGOs). This sector has more than doubled in size since the 1970s in an effort to fill the humanitarian vacuum left by the corporate sector and nation states. NGOs play an important role in pur suing activities that relieve suffering, provide basic health and social services, and which promote the interests of the poor in developing countries. Organizations fall into two main categories: (1) operational organizations, which design and implement development related projects or which undertake emergency work and disaster relief (ranging from the effects of natural disasters, war, or pandemics) (2) lobbying or advocacy groups whose pri mary purpose is to promote or defend a specific cause.

The global rise of NGOs has raised important ques tions about the role of voluntarism in a world where governments are retreating from health and welfare responsibilities and where solutions to global poverty continue to prove elusive. Key areas of interest for those working within the field have included: the shifting relationship between NNGOs and SSGOs as they move away from donor–recipient models to partnership and capacity building approaches; the shift from indirect to direct funding relationships with SNGOs by bilateral and multilateral official donors; and the potential crisis of identity and legitimacy among NNGOs as SSGOs become increasingly effective in undertaking activities previously carried out by organizations from the North. How these changes facilitate or constrain efforts to improve access to health and welfare provision and the reduction of health inequalities at global scale form critical areas of study for those working in the fields of development and health geography.

The Comparative Context

The North/South debate is only one aspect of the global rise of voluntarism. Studies have also begun to map this field in a comparative global context. Jennifer Wolch is one of the few geographers to have undertaken a cross national comparison. In contrasting the trajectory of voluntary sector development across the USA and UK in the 1980s, she revealed how the rapid expansion of the post war welfare state in the USA created an upsurge in voluntary sector development that contrasted strongly to the UK – where the centralized welfare state acted to limit local voluntary sector development. Important comparative research has also been undertaken by the Centre for Civil Society at Johns Hopkins University, in the United States. Voluntarism has been compared across 35 countries, including advanced industrial, developing, and transitional countries in central and Eastern Europe. Differences in the scale of organizational activity across countries have been mapped as well as variations in their reliance on volunteers. Comparisons across nation states demonstrate that though the voluntary sector is a sizeable force in a wide range of countries, the overall scale of voluntary sector activity and its proportion of the economically active workforce varies greatly – ranging from a high of 14% in the Netherlands to a low of 0.4% in Mexico. Furthermore, though the voluntary sector is proportionally 3 times larger in developed versus developing countries, there are significant variations even among less well developed countries. So while the overall scale of volunteering is higher in developed countries, there appears to be no systematic difference. So, for example, proportionally less than 10% of voluntary organizations rely on volunteers in Egypt in comparison to over 75% in Tanzania and Sweden.

The mapping out of variations in the scale of voluntary activity across countries, the identification of what sort of activity is occurring, and where volunteers are most active are important starting points in elucidating why global variations in voluntarism occur. Tentative explanations for these differences have focused on different social, political, and historical traditions, but as yet there has also been no sustained attempt to theorize why these variations occur. So despite its growing relevance to global health and welfare, and the unique conceptual and methodological insights geographers could bring to bear on the topic, work focusing on the relationship between voluntarism, people, and place in a global context is still limited.

Voluntarism: Place and the Political Context

Place and politics have been central to much critical geographical research on the voluntary sector. At the national scale, different political contexts influence the extent and form of voluntary activity that is likely to develop. In classic welfare states, such as Nordic countries, the extent of state funded and provided welfare has meant that there has been little room for the growth of voluntary provider organizations. This in turn has led to a greater development of grassroots organizations focused around advocacy and lobbying activities. In the old Soviet regime in Eastern and Central Europe, health and welfare was also viewed as the role of the state, but here, lobbying and advocacy roles were also discouraged. As a consequence, the presence of voluntary health and welfare organizations in these countries was severely limited until the fall of the Soviet regime in the late 1980s and early 1990s. Conversely, in Western European countries the tendency toward more social democratic forms of government has led to the development of more widespread welfare partnerships between the state and voluntary organizations, so rather than displacing the voluntary sector it has encouraged its growth.

Different forms of government and their approach to welfare are, thus, significant in influencing how and where voluntary welfare develops and flourishes. This has been particularly evident from the 1980s onwards, as governments in advanced capitalist states have increasingly turned to nonstate actors in their efforts to address the severe and ongoing fiscal pressures associated with public provision of health and welfare services. The fostering of a radically pluralized social economy, in which the voluntary sector plays a central role, has offered governments the prospect of being able to address seemingly intractable problems of maintaining health and welfare provision through the promotion of service delivery beyond the state.

Beyond the national political context, geographers have also demonstrated how regional and local government can influence the differential development of voluntary welfare services. Work by Brown on AIDS organizations in Vancouver in the 1980s revealed how the refusal of the provincial government to respond to welfare issues faced by a specific (gay) community impelled the community and its supporters to respond to the developing crisis by establishing their own voluntary provision in specific parts of the city. Similarly, in seeking to make sense of the uneven development of voluntary welfare services for the frail elderly in Scotland in the late 1990s, Milligan noted the importance of under standing place based differences in local politics. The approach of a traditional Labour local authority in a metropolitan area, with its commitment to public service provision, was contrasted with that of a more conserva tive rural local authority keen to dismantle traditional public services for the frail elderly in favor of contracted services purchased from the voluntary and private sec tors. Recent work in New Zealand and Canada has fur ther demonstrated how local and regional government can impact significantly on how and where voluntary sector development occurs and what form it takes.

'The Shadow State'

Theoretically, geographers have drawn on the 'shadow state' thesis to examine the interplay between politics and place and the implications of a growing dependence on voluntary health and welfare activity for both the or ganizations themselves and the communities they serve. The term was first coined by Jennifer Wolch in 1989. It refers to voluntary organizations that have collective service responsibilities that were previously borne by the public sector but which are now provided and adminis tered outside of traditional democratic politics. These organizations are seen to be controlled in both formal and informal ways by the state. Geographers have high lighted the tensions between the progressive elements of voluntarism and the potential dangers posed by the development of a 'shadow state'. A wide range of studies covering topics such as AIDS, mental health, the frail elderly, immigration, community governance, and the social economy have clearly shown how the involvement of the voluntary sector in the mixed economy of welfare provision is proving vital for filling gaps and enhancing the responsiveness of welfare service delivery to local needs. As these studies illustrate, however, the increased dependence of voluntary organizations on state contracts and grants, together with a growth in state regulation and administrative oversight, can act to reinforce state control over welfare provision and increase state penetration into the activities of everyday life. For Wolch, the growth of the 'shadow state' represents a profound danger for society, associated with the ways in which cutbacks in voluntary activity may be linked to an erosion of basic entitlements.

The dialectic of state voluntary sector in(ter) dependence in social welfare has been the subject of significant debate within geography but also more widely across the social sciences. Cutbacks in state funding and the resultant fiscal squeeze faced by voluntary welfare providers in the USA have led to an increase in fees and charging as a means of replacing lost revenues and a concomitant incursion of the for profit sector into traditional areas of voluntary action. Growing com petition for limited resources has led many organizations across a range of advanced capitalist states to seek to professionalize their services to meet increased regu latory demands. 'Shadow state' development thus repre sents a paradox for voluntary organizations, in that to remain in opposition to the state is likely to reduce their ability to deliver the extent of welfare services they would be able to deliver if working more closely with the state. Yet the resultant organizational restructuring and change to working practices and modes of operation resulting from increased professionalization and regu lation has effectively distanced the administrative centers of voluntary organizations from the communities they were set up to serve.

Relocating the Shadow State: De-centering and Renewal

Faced with this paradox, voluntary organizations have found themselves subjected to criticisms about how they should or have responded to political and economic change in order to survive. Debate has arisen around whether the voluntary sector should seek 'renewal' or 'de centering'. 'Renewal' involves a shift away from tra ditional voluntary ideals and a jettisoning of the drive to maintain independence in favor of partnership working with the state and private sectors in the bid to solve health and welfare problems. Geographers have demon strated how local voluntary and community organizations have been affected by the shift from government to governance. The creation of new geographically bounded communities with fiscal and decision making powers has resulted in the creation of areas of 'winners' and 'losers' and the potential marginalization of the welfare needs of nongeographically bounded communities (such as older people and the disabled). In addition, geographers have demonstrated the emergence of new hybrid spaces of governance that operate between the state and civil society. De centering involves voluntary organizations actively relocating to the margins, consciously moving away from partnership working or holding privileged places in relation to dominant institutions and powerful groups. By maintaining independence from the state, de centering is seen to provide an alternative means of addressing the profound problems that confront many societies today. Only by de-centering, it is argued, can we create a new, more inclusive social contract.

Neither of these polarized responses are problem free. On the one hand, partnership working can be fraught for voluntary organizations where it is formed and performed on an uneven playing field. On the other hand, by opting to remain on the margins in order to act as 'spaces of resistance', geographers have demonstrated that voluntary organizations are likely to find themselves unable to offer the level and amount of welfare services that they could have done had they embraced a mode of partnership working. Other observers maintain that, in reality, what we are seeing is a 'bifurcation' of the welfare voluntary sector into 'grassroots' organizations that are genuinely voluntary and happily remote from government at one end of the spectrum and much larger 'corporatist' welfare organizations that are voluntary in name only, and which are linked to government and business and subject to all sorts of rules and regulations at the other.

Place and Voluntary Activity

Regardless of differing interpretations of how and why welfare reform has been implemented, governments in many advanced capitalist countries have sought to rely to a greater extent on the voluntary sector as a means of buttressing the core services of state welfare and obtaining greater flexibility and responsiveness to welfare needs. With social welfare delivery increasingly being channelled through nonstate actors, critical questions arise about the spatially uneven structure of the voluntary sector, the factors that facilitate or constrain the development of voluntary activity in particular places and the potential impact on those dependent on social welfare services.

Though significant differences exist in the scope, structure, role, and funding of the voluntary sector in different countries, to a large extent, voluntary sector development is country specific and reflects differences in their social, cultural, political, and economic histories. While previous sections have illustrated how structural influences on voluntary activity can operate at both the national and local levels, voluntary action commonly develops in response to much more localized needs and interests that change over time and space. At the intranational level, geographers have uncovered a range of additional influences that have been instrumental in influencing the interrelationship between people, place, and the voluntary sector and its spatial distribution. These include:

1. variations in need, creating differing pressures for the development of voluntary activity;
2. differences in culture and socioeconomic circumstances, affecting people's willingness and ability to support the development of local voluntarism;
3. differing histories of philanthropic giving within places;
4. recent histories of change in localized welfare policies that have targeted or removed resources from local communities and places;
5. the availability and accessibility of an appropriate physical infrastructure; and
6. the cult of the individual – where identified 'champions' or social entrepreneurs have been instrumental in promoting the development of a particular service or who have been active in developing resource-seeking skills in an area.

All the above have been cited by geographers as explanations for why voluntary activity develops in different places at different times. As a consequence, the geographies of voluntarism are often uneven.

Sociocultural Approaches to Place and Voluntarism

An emergent strand of work within this field takes a sociocultural approach to the geography of voluntarism. The emphasis here is on how particular sets of (voluntary) organizational spaces emerge within places and how they are experienced and performed by: (1) those who both work within them; and (2) by those who seek to avail themselves of the services they offer. One strand of this work has focused on how voluntary effort can act to change the landscape of the city. Research in Australia has demonstrated how, through the pooling of voluntary resources and expertise, migrant communities have been able to develop communal facilities that address specific cultural needs that would otherwise go unmet. Over time, this results in the emergence of a voluntary and community landscape that is reflective of the cultural community it represents. Other work taking a sociocultural approach to the geography of voluntarism has focused on the 'doing' of organizational space. Three areas of concern include: (1) how the experiential characteristics of organizations are produced and reproduced over time; (2) on how the organizational spaces of voluntary organizations are negotiated and interpreted through social practices and discourses; and (3) on how embodied practices form, and are formed by, particular voluntary spaces at particular times. Examples are drawn from community drop-in centers in New Zealand and charity shops in the UK, the latter illustrating the tensions that have arisen for volunteers, customers, and organizations as once relatively informal (domestic) sites of voluntary activity become recast as professional, work-inflected sites. This also reveals the problematic inherent in government assumptions that the voluntary sector is a fruitful site through which to promote active citizenship. This approach to voluntarism has particular methodological implications in that it lends itself to a specifically ethnographic form of enquiry.

Volunteering

A further defining feature of the voluntary sector is its engagement with volunteers – that is, people who give of their time to help others, without obligation and for no financial recompense – whether in the day to day operations of voluntary organizations, on their management boards, or in both contexts. The act of volunteering requires further clarification. It can occur 'formally' as an activity that takes place within the structure of a voluntary organization, or 'informally', where an individual performs similar unpaid work without a formal organization. Examples of informal volunteering include caring for a friend, spouse, or other family member, shopping for a neighbor, helping at school functions or youth groups. It can also include volunteering outside of a formal voluntary organization but within the structure of a formal private or statutory organization – for example as a school or hospital governor, a parish or community councillor, or as a volunteer firefighter or auxiliary policeperson. Throughout the 1990s and early twenty first century, the desire of many governments in advanced capitalist economies to embrace 'third way' approaches to the delivery of health and social welfare and their growing concerns about increased disaffection in society has drawn their attention to how active citizenship might be promoted through volunteering. Though as this section on 'volunteering' illustrates this approach may prove more problematic than governments anticipate – a view that has been reinforced by work in Scotland and elsewhere, that has suggested that the drive to professionalize may in fact be reducing the space to volunteer.

People and Place

Geographical perspectives on volunteering encompass two main strands of work: firstl, studies drawing on large scale data sets that have examined the patterns and determinants of 'who volunteers, where, and why'. Using sophisticated techniques such as multilevel modeling as well as the interrogation of national survey data, geographers have sought to identify whether the propensity to volunteer was more likely to be influenced by the characteristics of people or of places. Work in this field is in its infancy, and as yet there is little consensus about the scale at which spatial variations in volunteering can be determined. Analyses of formal and informal volunteering in both the UK and North American contexts, for example, have uncovered significant variations in who undertakes formal and informal volunteering in particular places. On the one hand, this has been attributed to variations in the historical development of state–civil society relations. On the other, variations are claimed to be largely a function of the compositional characteristics of people in particular areas and differences in the social class make up of local neighborhoods. What both these oppositional views agree on, however, is that where geography is most likely to matter in relation to who volunteers, where, and why, is at the local and sub regional levels.

Place and Gender

The second strand of work on volunteering has focused around geographical explanations for the 'gendered nature of volunteering' – specifically the role of women in voluntary activity. As the embodiment of the altruistic spirit, women are seen to have the caring mentality required for volunteer activity, so legitimizing their suitability for volunteer work. While large scale data sets confirm that women have a greater tendency to volunteer than men, work in this field also focuses on the relationship between the nature of that voluntary work and the place in which it occurs. Most geographical research around women's volunteer work focuses on rural areas and demonstrates that women's participation in voluntary activity in rural areas is a widespread and significant part of their workload. Work in the UK, Australia, and New Zealand has drawn out the links between rural ideology and the sorts of gender roles and relations that are inherently supported and sustained by many rural based voluntary organizations. Three key themes can be identified: first, they draw out the importance of women's volunteering activity in supporting and sustaining rural communities. Helping out at rural schools, caring for frail older people, providing rural transport to hospitals or GPs, and meals on wheels services provide examples of where women's voluntary activity appears to be directly replacing state welfare provision and maintaining social support. Second, women's contribution to volunteer work in the rural labor market has been identified as a third sphere of work that disrupts the binary division between public and private forms of work. Women's participation in voluntary work has also been shown to give rise to unanticipated tensions in rural communities. On the one hand, it provides an important opportunity for rural women to gain influence and power within the community (e.g., as local voluntary sector organizers or Chairpersons). On the other hand, expectations about women's willingness to participate can create social pressures to contribute leading to a sense of unwanted obligation. Finally, women's volunteering is seen to capture the 'spirit' of rural living, linking closely to valued attributes of rural communities such as small size, self sufficiency, and a willingness to assist others. By implication, women's engagement in voluntary activity not only benefits the community and helps to preserve traditional rural society, it also contributes to the con ceptualization and representation of rurality.

Issues

As will be clear from the above sections, voluntarism is a relatively new sphere of geographical interest. Nevertheless, geographers are making a significant contribution to the understanding of where and why spatial variations in voluntary activity in health and welfare occur. Geographies of voluntarism have been most fruitful at the national and local levels and have demonstrated the importance of understanding the influences of differing political, economic, and socio cultural contexts over time and place. Theoretically, work on the shadow state has also had an important influence on voluntary sector research, both within geography and more widely within the social sciences. Comparative work at the international and global scale will be an important future development for this field of geography.

See also: Health Geography; Health Services Restructuring; Welfare Reform.

Further Reading

Brown, M. (1997). *Replacing citizenship: AIDS activism and radical democracy*. London: Guilford Press.

Conradson, D. (2003). Doing organisational space: Practices of voluntary welfare in the city. *Environment and Planning A* 35, 1975 1992.

Fyfe, N. and Milligan, C. (2003). Out of the shadows: Exploring contemporary geographies of the welfare voluntary sector. *Progress in Human Geography* 27, 397 413.

Little, J. (1997). Constructions of rural women's voluntary work. *Gender, Place and Culture* 4, 197 209.

Milligan, C. (2001). *Geographies of Care: Space, Place and the Voluntary Sector*. Aldershot: Ashgate.

Milligan, C. and Conradson, D. (eds.) (2006). *Landscapes of Voluntarism: Health, Welfare and Governance*. Bristol: Policy Press.

Salamon, L., Sokolowski, S. W. and List, R. (2003). *Global Civil Society: An Overview*. Baltimore, MD: Centre for Civil Society Studies, The Johns Hopkins University.

Wolch, J. R. (1990). *The Shadow State: Government and Voluntary Sector in Transition*. New York: The Foundation Centre.

Wolpert, J. (1977). Social income and the voluntary sector. *Papers of the Regional Science Association* 39, 217 229.

Wolpert, J. and Reiner, T. (1985). The not for profit sector in stable and growing regions. *Urban Affairs Quarterly* 20, 487 510.

Relevant Websites

http://www.istr.org/pubs/voluntas/index.htm
International Society for Third Sector Research.

http://www.ncvo vol.org.uk/publications/
National Council for Voluntary Organization.

http://www.spa.ucla.edu/ccs/webfiles/template1.cfm?page =res main.cfm&mid=3
School of Public Affairs UCLA.

http://www.jhu.edu/~ccss/
The Johns Hopkins University.

von Humboldt, A.

A. Buttimer, University College Dublin, Dublin, Ireland

von Humboldt, Alexander (1769–1859)

Scientist and explorer, Alexander von Humboldt (**Figure 1**) was born on 14 September 1769 in Schloss Tegel, near Berlin, second son of Commandant Alexander Georg von Humboldt, a Prussian army officer and Elizabeth de Colomb, descendant of a French Huguenot family. Throughout his life, Alexander enjoyed the affection and support of his brother Wilhelm von Humboldt, celebrated founder of the Humboldt University in Berlin. Privately tutored, he was exposed to a diverse range of intellectual interests in childhood, three of which continued to evoke passion throughout his lifetime: natural history, geographical exploration, and the social egalitarian ethos of the French Revolution. Polymath scholar and celebrated patriarch of European geography, Alexander von Humboldt's career details and voluminous writings were to retain an enduring appeal in both academic and public life. He died on 6 May 1859 in Berlin.

Humboldt's studies at Frankfurt on the Oder, Göttingen, Hamburg, and Freiburg between 1787 and 1792 introduced him to physical and biological sciences. His first publication *Mineralogische Beobachtungen über einige Basalte am Rhein* (1790) was based on a journey along the Rhine with Georg Forster. A visit to Paris in the same year reinforced his republican idealism and desire to travel to the Tropics. Appointed assessor and later Director of Mines in Bayreuth (1792), he was concerned for the social conditions of the miners, setting up a school of mines and seeking to improve living standards. He conducted research on underground vegetation and in his *Florae Friburgensis Specimen* (1793) he speculated on the idea of 'life force' which he shared later with the humanists of Weimar where he met Goethe in 1794. A committed empiricist, Humboldt conducted experiments on muscular irritability even on his own body (1797), and encouraged Goethe's research on optics and colors. He discovered the local magnetic declination of rocks and between 1792 and 1797 traveled to salt mining regions of the Alps and northern Italy. With Leopold von Buch, a colleague in geology, he made daily measures at the same hours of barometric pressure, temperature, humidity, oxygen, and carbon dioxide and electric charge of the air. These served as bases for his proposed program for meteorological observations which later became standard practice.

His mother died in 1796, leaving him a substantial fortune which enabled him to pursue his life interests and later to publish the results of his research. In 1797 he joined his brother Wilhelm in Paris where he met several famous scientists including Aimé Bonpland, a botanist who shared his eagerness for scientific travel. With the welcome permission of King Charles IV of Spain they set out to explore the New World. Traveling via Spain they arrived in Tenerife in June 1799 where they climbed the volcanic peak of Teide, witnessing lava streams, sulfurous vapor, and spectacular landscapes including the famous 'Dragon Tree' whose circumference was 45 ft. Following the route of Columbus across the Atlantic they arrived in Cumana, Venezuela, 1 July 1799.

For 5 years they traveled together over a distance of 6000 miles in hitherto little known regions of the New

Figure 1 Alexander von Humboldt (Julius Schrader Oil on canvas 1859). Source: Halina Nelken (1980). *Alexander von Humboldt. His Portraits and their Artists. A Documentary Iconography*. Berlin: Dietrich Reimer Verlag 1980, Seite 169.

World. They explored the Orinoco country, collecting samples of hitherto unknown flora and fauna, and even tually discovering the Casiquiare channel which linked the drainage systems of Amazon and Orinoco. This meant a triumphal debunking of previous ideas, for example, those of Buache, that a vast mountain chain separated the two river systems, and that the source of the Orinoco was on the eastern slopes of the Andes. In November 1800 they set out for Cuba, intending to send samples back to Europe, studied life on the island, and returned after a few months to Cartagena where they undertook an arduous journey along the Cordilleras from Colombia to Peru with stops in Bogota and Quito. On 9 June 1802 they ascended Chimborazo to an elevation of 18 893 ft, a record at the time, and along the way they carefully documented altitudinal variations in vegetation cover and other biophysical features. These data would eventually serve in the design of the famous graphic representations contained in reports of his explorations. On a long detour to the sources of the Amazon they learned much about the cultures, languages, folkways, and economies of native populations and listened to accounts of pre Colombian times. At Callao (Peru), Humboldt measured the velocity and temperatures of the ocean current off the west coast of South America which now bears his name.

These Andean explorations led to a virtual revolution in geological theory and a transcendence of 'Neptunist' ideas such as those Humboldt had learned in Göttingen. Here was evidence of geological faults and of the igneous origin of granite and other rocks. The entire high ground around Quito, Humboldt suggested, was one single vol canic system, the volcanoes of Cotopaxi and Pichincha acting as vents. And in Mexico, the recently formed volcano Jurillo enabled further opportunity to test his theses on vulcanism. After a year in Mexico they set sail once again first to Cuba and then to Philadelphia, where they were welcomed by the American Philosophical Society and later by President Jefferson in Washington DC. The information they carried with them on New Spain was especially welcome at this time when details on lands west of the Mississippi, recently acquired through the Louisiana Purchase, were little known in Washington. They returned to France in September 1804.

After a short trip to Mount Vesuvius with Gay Lussac and C L von Buch, and a sejour of two and a half years in Berlin, where he wrote *Ansichten der Natur* (1808), Humboldt returned to Paris where the next 21 years were spent on editing the results of his travels. Eagerly welcomed in scientific academies, he preferred Paris as a place for work and the French language as the best medium for communicating geographical ideas. He turned down an appointment as Prussian Minister of Public Instruction in 1810, but did participate as scientific advisor in various official delegations by the King of Prussia to European countries. This was a highly productive phase in Humboldt's career. In 1804 he presented his thesis on the decreasing intensity of Earth's magnetic force from poles to equator, lectured on the geography of plants, and engaged some of Europe's most outstanding artists and scientists to help edit his notes and sketches made during his years in the tropics. The 23 volume work *Voyage de Humboldt et Bonpland* (1805–34), jointly with Bonpland, contains masterpieces of graphic, scientific, and literary discourse, thematically spanning the full gamut of geographical interest, from rock for mations to population analyses, resources and trade, and a proposed route for the Panama Canal. Central to the entire collection was the work on plants and the *Tableau physique des Andes et pays voisins*, an impres sive graphic summary of the spatiotemporal and alti tudinal interconnectedness of terrestrial and atmospheric phenomena.

In 1827, summoned by the king, he returned to Berlin. The idea of *Kosmos* took shape during public lectures delivered 1827–29. He was actively involved in *Gesellschaft der Naturforscher und Ärzte* and was a founder member of *Gesellschaft für Erdkunde zu Berlin* (1828). At the invitation of Tsar Alexander of Russia (1829) he visited Central Asia, journeying from Estonia and St. Petersburg to the Chinese border, returning via the Caspian Sea. This enabled him to complete meteoro logical data for the world isothermal map, to refine a theory on the orographic configuration of the central Asiatic mountain systems and tablelands, and to discover diamonds in the gold mines of the Urals. He persuaded the Russian government to establish a line of magnetic and meteorological stations across northern Asia, and in subsequent correspondence with the President of the British Royal Society, secured the establishment of a wide range of similar stations throughout the Com monwealth. After 1830 he was sent on various diplomatic missions, promoting scientific education, and working on *Kosmos*. The idea for the book *Kosmos* was initially con ceived as a *Physique du monde*, treating all forms of life and their relations to physical environments, with par ticular emphasis on the Andes. The first two volumes were published 1845 and 1847, the third and fourth between 1850 and 1858, and the fifth (posthumously) in 1862.

Humboldt's ideas and work are best appreciated when placed in the context of late eighteenth and nineteenth century ideas, a time that witnessed dramatic tensions within academy and society. The tropical world posed new horizons: issues of health and disease, of racial diversity, commercially viable commodities and colonial administration were eagerly debated. How to understand 'nature' became a burning issue, scholarly opinion sharply divided between *Encyclopédistes*, who believed in the power of science and scientifically based technology and politics to bring about a better world, and 'Nature philosophers' who emphasized spiritual and esthetic dimensions of nature experience. Both Goethe and Humboldt sought ways to transcend the antipathies between the sciences and the humanities, each learning much from the other. There were other quarrels in the air about the so called 'New World'. Buffon argued that because of their 'geological immaturity' and their hot humid climes and swamplands, the Americas could only have native people who were morally and physically degenerate. Others argued that New World 'noble savages' probably lived in Edenic harmony with nature. Humboldt's work transcended these divides: while always placing humanity within its multidimensional contexts, he confounded both the environmental determinists and Romantic idealists.

This New World was also undergoing change at the time of his visits there. Already there were rumblings of Creole inspired independence from colonial powers. Humboldt's daring and empirically accurate accounts were welcomed by leaders of such movements. They were also heeded by European commercial interests eager to exploit New World riches once independence would dawn. The *Political essay on the kingdom of New Spain* (1811), with its detailed accounts of Mexican population and its ways of life set in the context of their biophysical environment, stands out as a classic of human geography. Access to data from the census of 1793 enabled detailed analyses of demographic, socioeconomic, and political aspects of everyday life. Availing of local sources of environmental knowledge and memory, too, he presented a picture of indigenous society which contrasted markedly with contemporary European perspectives. Contrary to general opinions of the day, he claimed that Mexico's wealth lay in agriculture, not mining; that the relative vulnerability of agrarian livelihoods to climatic or other hazards was due to lack of investment, social inequalities, and overdependence on a single crop of maize. The socially engineered marginalization of native populations over centuries, coupled with the destruction of 'hieroglyph' sources of traditional wisdom, could shed light on the apparent misery and melancholy which characterized their present state. Mexico City itself, its botanical gardens, monuments, and educational institutions were especially striking: education, he argued, being a force toward leveling differences among social classes. Social equality was also central in his *Essai politique sur l'ile de Cuba* (1826), a powerful harangue against slavery. The influence of these two works on the young Simon Bolivar made him a hero of the liberation movement throughout South America.

Humboldt's American expeditions were revolutionary. Previous explorers had reported on exotic phenomena but Humboldt added accurate measurements, scientific explanations, and results were presented in a variety of graphic, tabular, and cartographic modes. Analytically, he took the study of 'nature' out of the laboratory into the landscape. While each element demanded specialized analytical attention in order to establish its own specific rules of order and organization, the most important challenge was that of integrating all these knowledges. This challenge was addressed at length in his *Kosmos*, and earlier in his brilliantly succinct *Essai* on the geography of plants (1805).

By any standard, Humboldt's *oeuvre*, spanning over 50 years, was remarkable. His enduring legacy, however, consists in not only his voluminous *faits accomplis* but also in the poetics of his discourse. Throughout his work there is an obvious sensitivity to questions of scale, the need for comparative study, and for critical reexamination of conventional theories on the age of the Earth, the sizes and heights of the continents, the lower limits of the snowline in different parts of the Earth, and flora and fauna which carpet its surface. Geographically speaking, it was not just the locations of phenomena which mattered for him, it was their distributions in multidimensional space and time. The most famous instrument for global comparative analyses were the 'isolines', of which the global map of isotherms became the most famous and served as background frame within which a range of other distributions such as plants and animals, livelihoods, patterns of health, and disease which were displayed in the *Berghaus Physikalicher Schul Atlas* (1850).

A second feature of Humboldt's work is a vivid awareness of temporality. All phenomena on planet Earth were in process of evolution, however diverse their temporal rhythms. Among those who explicitly acknowledged the influence of his work are Louis Agassiz, Charles Darwin, and Charles Lyell. Agassiz found his work on vulcanism and on climate change to be truly liberating. Lyell continued to work at the interface of geology and biology, using the isothermal lines as basis for his own investigations into climate change. Questions of changes over time, slow moving or abrupt, as sketched in the *Personal narrative of travels to equinoctial regions of America during the years 1799–1804*, greatly impressed the young Darwin as he embarked on his own voyages. Humboldt raised questions of global environmental changes using records of plant and animal domestication associated with sedentary agriculture. On conventional European theories about 'phases' in the evolution of livelihoods, however, he was skeptical. The assumed evolution of livelihoods from hunter gatherer to pastoral to sedentary agriculture was not applicable in the New World: here, in the absence of large domesticated animals, agriculture appeared to have developed directly from gathering to sedentary livelihoods.

Thirdly, from early years on, Humboldt was inquisitive about social aspects of all living forms – plants, animals, and humans. In the geography of plants distinctions were made, as among animals, between two different ways of life, some growing separately, others assembling in societies, “like ants and bees, covering vast expanses and excluding all heterogeneous species” (Humboldt, 1805: 15–16). On human societies, his views were usually quite explicit. He uncovered treasures of indigenous South American civilizations, their languages, arts, ways of life, belief systems, and political structures. He also freely criticized the Eurocentrism of conventional histories and, like Benedictus de Spinoza, acknowledged the reflexivity in all scientific endeavor. Slavery he regarded as intolerable, sparing no irony on white plantation owners and their treatment of servants. Enlightenment convictions are also evident in his visions for the future – including the prospect of political emancipation from colonial powers.

Finally, Humboldt’s graphic language and ingenious renderings of landscape sought to reveal not only the material (ecological) but also the esthetic and ethical dimensions of humanity’s relationship to the natural world. Landscape thus became a legitimate object of enquiry for both natural science and humanities: evoking not only understandings of regional character and people’s sense of place, but also, via his celebrated *Tableau*, a powerful template representing the interconnectedness of living forms. Holding firmly to the educational value of landscape pictures as catalysts for intercultural communication and mutual understanding, he encouraged artists to sketch regional images based on the physiognomy of plants. Artists such as Frederick E Church eagerly responded to this challenge and a new genre of landscape painting evoked public awareness of nature, eventually leading to the establishment of America’s first national parks of Yellowstone and Yosemite.

The impact of Humboldt’s ideas varied from one country to another, and from one period to another. Eagerly welcomed initially in the New World, responses in the Old were not unqualified. Some found his language to be excessively ornate, his blending of scientific and literary styles confusing. His ideas were better understood in the Latin world, in parts of continental Europe, Scandinavia, and Russia than they were in Anglo circles. Initial enthusiasm expressed in nineteenth century United States was followed by virtual oblivion in the early twentieth. During his own lifetime he was widely acclaimed by artists and poets; his public lectures drew audiences from all classes of society. A glance at the number of Earth features which today still bear the name of Humboldt bear witness to the charm and eloquence of this great scientific traveler and the enormous public appeal of his life and work. His name is commemorated in statues and street names, myths and monuments, as well as in a river, a lake, a glacier, a mountain range, a lily, an ocean current, several ships, an Internet website, and at least one Salzburg café.

It may have been in the realms of natural history, botany, and sylviculture, particularly in Russia and in Spanish speaking realms, that his works were especially influential. In Germany, it seems, the global isothermal map, and the project of continuous measurements of meteorological, hydrological, and geomagnetic conditions aroused most applause and continued research. Within geography, Humboldt’s impact has been less than one might have expected. Writing in the period before the subject became institutionized as an academic discipline, he did not cultivate a ‘school’ of devoted disciples, nor did he produce a textbook for the subject. As a lone scholar he wrote in a manner that aimed at opening readers’ minds to new ways of knowing and seeing nature and culture, and was articulately critical of Eurocentrism in views held about the native peoples of the New World. When geography achieved academic status within universities, especially those of imperial and colonizing societies, during the closing years of the nineteenth century, it was Carl Ritter’s *Erdkunde*, penned from an armchair, rather than Humboldt’s *Cosmos* or *Voyage* that provided the canonical texts for regional geography.

At the dawn of the twenty first century, when issues of humanity and environment reclaim attention, and the subject of geography is experiencing an institutional restructuring, Alexander von Humboldt’s ideas could again merit serious study. His legacy provides unquestionable rationale for again exploring dialogue among specialists of geography’s so called humanistic and scientific branches. Understanding the humanized landscapes of Planet Earth involves more than scientific mastery of various geophysical, biological, political, or cultural processes; it involves lived experience, esthetics, emotion, and shared concern for wiser ways of dwelling, from local to global scales.

See also: Environmentalism; Regional Geography I.

Further Reading

Beck, H. (1959, 1961). *Alexander von Humboldt. Biographie in zwei Bänden*. Wiesbaden: Fr. Steiner Verlag.

Biermann, K. R. (ed.) (1989). *Alexander von Humboldt. Aus meinem Leben. Autobiographische Bekenntnisse*. Leipzig: Urania Verlag.

Botting, D. (1973). *Humboldt and the cosmos*. London: Michael Joseph.

Buttimer, A. (2001). Beyond Humboldtean science and Goethe’s way of science: Challenges of Alexander von Humboldt’s geography. *Erdkunde* 55(2), 105 120.

de Terra, H. (1955). *Humboldt*. New York: A Knopf Inc.

Geiger, L. (1909). *Goethe’s Briefwechsel mit Wilhelm und Alexander v. Humboldt*. Berlin: Erschienen bei Hans Bondy MCMIX.

Geographical Review (2006). Special Issue on *Humboldt and the Americas* 96, 3.

Kellner, L. (1963). *Alexander von Humboldt*. London and New York: Oxford University Press.

Minguet, Ch (1969). *Alexandre de Humboldt. Historien et géographe de l'Amérique Espagnole* (*1799 1804*). Thèse pour le doctorat čs lettres à l'Université de Paris (Sorbonne), Paris, Faculté des Lettres et des Sciences Humaines.
Von Humboldt, A. (1845 1862). *Kosmos, Entwurf einer physischen Weltbeschreibung*. 5 vols. Stuttgart: Cotta. English translation (1848 1858) by Otté, E. C. as *Cosmos: A sketch of a physical description of the universe*. London: Henry G. Bohn.

Relevant Websites

http://www2.ku.edu/~maxkade/humboldt.htm
The University of Kansas.
http://www.uni potsdam.de
University of Potsdam, Germany.

Vulnerability

B. Wisner, Benfield UCL Hazard Research Centre, London, UK

Glossary

At Risk Term in public health and epidemiology for statistical likelihood of suffering a specific morbidity.
Capacity Ability to protect oneself from a threat, to overcome or reverse vulnerability.
Coping To adjust one's behavior in order to minimize the impact of a hazard.
Hazard An event or hazard in nature, the built environment, or social life that may cause harm to persons, assets, or livelihoods.
Human Development Index (HDI) Measurement developed by United Nations Development Program (UNDP) and used in its *Human Development Report* since 1990, composed of national longevity, literacy, school enrollment, gross domestic product per capita, and an indicator of purchasing power.
Mitigation Action taken to reduce the potential impact of a hazard; to reduce potential losses.
Participatory Action Research (PAR) A goal-directed problem-solving approach that builds on local knowledge and experience with or without the admixture of external specialist knowledge.
Resilience The ability of a system to return to a prior state after having experienced a shock.
Risk The coincidence of hazard and vulnerability, with the potential for harm depending on the degree of self-protection and social protection at work in a given situation.
Self-Protection The total coping and capacity mobilized and acted upon by an individual or household in the face of a hazard.
Social Protection Sum of mitigation activities and provisions provided by collectivities, especially by the state, as a matter of routine governance, including health and other services, infrastructure, and early warning.

Introduction

Development studies specialist, Robert Chambers popularized the term 'vulnerability' in the analysis of rural poverty in his 1983 book, *Rural Development: Putting the Last First*. It came as one of five elements that interlocked with each other, producing what he termed a 'ratchet effect' or 'deprivation trap': a condition of 'integrated rural poverty' that is very difficult to extract oneself from. The other elements were political powerlessness, physical weakness (ill health), isolation, and income poverty. Contemporary users of the term vulnerability in the literature of disaster risk reduction, as well as climate change might find Chambers' usage odd. Isn't vulnerability precisely a matter of what Piers Blaikie and Harold Brookfield called "ecological, economic, and political marginality" in *Land Degradation and Society*, published in 1987? At a time when disciplines and knowledge/practice communities need to cooperate more than ever to understand and combat interlocking threats – poverty, disease, climate change, violence, new and emergent environmental and natural hazards – the conceptual differences that underlie different uses of this term still separate scholars and practitioners.

Many Facets and Faces of Vulnerability

There are many hyphenated forms in using or modifying the noun vulnerability in four literatures (development studies, climate change, public health, hazards and disaster research), such as economic, political, cultural, and educational. Aysan listed ten combinations of adjectives and the noun vulnerability in her keynote address that appears in the 1993 collection *Natural Disasters: Protecting Vulnerable Communities:*

- lack of access to resources (material/economic vulnerability);
- disintegration of social patterns (social vulnerability);
- degradation of the environment and inability to protect it (ecological vulnerability);
- lack of strong national and local institutions (organizational vulnerability);
- lack of public awareness (attitudinal and motivational vulnerability);
- limited access to political power and representation (political vulnerability);
- certain beliefs and customs (cultural vulnerability); and
- weak buildings or weak individuals (physical vulnerability).

Notions of Vulnerability in Five Research and Policy Communities

Knowledge and practice communities have been using the term vulnerability for the past 25 years. They use the term in different ways, but these communities need more than ever to cooperate with each other to understand and address threats that cannot be neatly allocated to categories such as 'climate', 'development', 'conflict', or 'health'.

Hazards and Disaster Research

Beginning in the 1980s, the terms vulnerable and vul nerability turned up more and more frequently in writing about hazards and disasters. Literally dozens of authors use the term and related ones such as resilience, mar ginality, susceptibility, adaptability, fragility, and risk. At the beginning of the International Decade for Natural Disaster Reduction (IDNDR, 1990–99), there was still a strong engineering bias in the leadership of this initiative, so vulnerability was often reduced simply to exposure to seismic risk and residence or presence in a fragile structure. As some put it at the time, "earthquakes don't kill people, buildings do."

By mid IDNDR, there was a strong call for community participation in defining risks and risk reduction and expansion of the IDNDR to include the hazard of drought. These modifications of the international decade program ushered in a shift in how vulnerability was understood. There was more emphasis on household livelihoods, the dynamics of these survival strategies, and household coping patterns. One common definition at the time, put forward by Piers Blaikie and his colleagues in the 1994 edition of *At Risk*, ran "the characteristics of a person or group in terms of their capacity to anticipate, cope with, resist, and recover from the impact of a natural hazard."

What also emerged from work with policymakers and planners at the time, and since, is that concepts and definitions are only a beginning. The difficult work comes when one tries to apply the concept of vulnerability to concrete situations. Planners, community activists, and others need to be able to specify as precisely as possible the characteristics of groups of people, households, and individuals that make them vulnerable to disasters. Much current writing on vulnerability is too general to help in this regard. For example, vulnerability is sometimes simply identified with poverty.

While there is a strong correlation between income and access to resources with the ability of people to protect themselves and, especially, to recover after disasters, the straight forward identification of 'the poor' as vulnerable does not help planners and activists formulate short and medium term plans and demands. First of all, all persons at the same level of income do not suffer equally in disaster situations nor do they encounter the same handicaps during the period of recovery. For instance, in West Africa during drought in the 1980s, nomads who may actually have been 'less poor' than sedentary farmers actually suffered more, as measured by death and dislocation. Furthermore, the elimination of poverty is a long range goal requiring social justice and equity, income and resource redistri bution, possibly the creation of a social or family wage, and economic democracy. Disaster mitigation need not wait for the achievement of such goals.

Humanitarian Assistance

Humanitarian assistance and the study groups that have been convened to understand and improve its practice focus mainly on response to emergencies and recovery from them. In many conflict and post disaster situations large numbers of people need assistance in a short period of time. Displacement is common, and large camps for refugees or internally displaced persons in particular, call for quick, simple ways of triaging people and prioritizing needs. It is therefore not surprising that in the vocabulary of humanitarian assistance vulnerable means to subject to immediate or imminent harm. Thus children may be screened by mid upper arm circum ference as a rough guide to their nutritional status, and some sent directly to nutritional rehabilitation units. In other cases classes of people – women and girls – are correctly viewed as vulnerable to sexual predation in refugee camps, and special protections are usually provided.

The term vulnerable in these cases is utilitarian and, though useful for their purpose, does not attempt to unravel or analyze the reasons why people are dis placed in the first place: the root causes of underlying patterns of vulnerability and blocked or defeated capacity.

Public Health

Individuals with certain characteristics are denomi nated 'at risk' in epidemiological models. The term vulnerability also occurs in retrospective studies of situations as wide ranging as political violence and the risk of injury in Karachi, Pakistan; health status of HIV AIDS orphans in Cape Town, South Africa; and mental health problems in the United States. For example, males working as migrant mine workers in South Africa are, for example, at risk of contracting HIV AIDS, as are their wives at home in rural areas of surrounding countries. This notion of being 'at risk' or having 'risk factors' shares characteristics with the utili tarian approach of humanitarian assistance just discussed. Both are highly focused on the relationship of individuals with certainly measurable characteristics (age, sex, etc.) to specific, definable threats. Although the so called epidemiological triangle (agent–host–environment) pro vides a broader causal and analytical context, there is no interest in probing why an individual (e.g., host of disease) has the risk factor.

Climate Change

About a decade ago, the Intergovernmental Panel on Climate Change (IPCC) defined vulnerability as "the extent to which climate change may damage or harm a system." Although some climate researchers have sub sequently focused on households by adopting a livelihood approach, mainstream climate science still takes a sys tems focus, a perspective reaffirmed by the IPCC in its Fourth Report in 2007. Moreover, discourse on vulner ability takes place at basin, world region, and national scale. A major conference in early 2008 held at the Peace University in the Hague featured a discussion on 'coping networks' at basin, regional, and national level that have been organized by the Cooperative Programme on Water and Climate.

A focus on system has its place in scientific and policy discussions, but it should not be confused with analysis that seeks to identify the needs of ordinary people. Certainly it is true that rain fed farming in the semiarid tropics or cities of a certain age in earthquake zones are vulnerable. However, such statements do not help set the agendas of planners and activists. Quite often it is not the system of production or habitation *per se* that is vulnerable to environmental hazards, but persons or households within those systems who lack the resources to mobilize the defenses such systems already have against hazards.

This is best documented in the case of farming and herding. There exist many traditional forms of coping with hazards such as drought, hail, frost, insect pests. Studies of vernacular architecture reveal a range of de sign and construction features that can serve as the basis of low cost measures to increase the safety of housing. Those concerned with mitigation of disaster at the grassroots need to be able to understand why certain members of the community are unable to avail them selves of this knowledge and practice. In other words, they need to be concerned with the vulnerability of people, not of systems.

Another problem is created by spatial scale. The concept of vulnerability is sometimes used to charac terize cities or metropolitan regions, whole nations, even regions of the world. This may be useful in the context of the political struggle for resources; for instance, in ar guing in a parliamentary context that such and such a city is more vulnerable than another and requires more of the national budget. It is also possibly useful in the bureaucratic allocation of resources to nations or groups of nations by multilateral or bilateral donors. For such purposes a number of tool have been created such as the Disaster Hotspots database and maps and UNDP's national scores on its Disaster Risk Index. However, there is a dangerous fallacy lurking. It is tempting to assume that reducing vulnerability at, say, the national level automatically achieves reduced vulnerability among so cial groups, households, and individuals in that nation. That is certainly not the case.

Identification of 'national' or 'societal' vulnerability can simply mean revealing that certain countries import a large share of their food or are geographically situated with long, low coastlines, are frequently in the path of tropical storms, or in regions of high rainfall variability. One can imagine steps being taken to reduce 'national' vulnerability in such situations that have a variety of outcomes. Adoption of a 'production first' strategy using high yielding seeds might be proposed as a way of 'drought proofing' a nation. However, food availability alone does not ensure adequate consumption by all persons. In fact, such a strategy could actually leave some groups such as small farmers worse off, that is, leave 'households' more vulnerable if it caused a shift in gov ernment credit away from the small farm in favor of the large scale producer. In a similar way, national scale adaptation to climate change might take the form of building large dams that displace many poor farmers whose household vulnerability increases as a result.

Development Studies

In development studies one encounters a more holistic use of the notion of vulnerability. This article began with Robert Chambers, and his fivefold analysis of the 'deprivation trap' which holds all the elements that have subsequently been expanded and nuanced. So 'income poverty' has been analyzed as a feature of livelihood security and insecurity. In subsequent work Chambers and others studied seasonality and health, and also access to natural resources, as factors also affecting vulnerability. This work foreshadowed some of the analysis some contemporary climate change researchers have produced, and has tended to merge with hazards research as well.

Today development studies commonly uses a liveli hood framework to understanding rural and urban pov erty and escape from poverty. This approach looks holistically at the strategic use of a variety of 'capitals' (economic, natural, social, human, political) in securing a livelihood over time. In this context, the notion of 'vul nerable livelihoods' is often used. Within this framework, vulnerability is the opposite of 'resilience', or the ability to return quickly and easily to an earlier state before a shock.

Vulnerability and Daily Life

The daily life of many people alive today is a permanent emergency. Disasters can be interpreted as "the extreme situation which is implicit in the everyday condition of the population," in the world of Alex Baird and his colleagues in the early days of the Disaster Research Unit at Bradford University in UK. Disasters "bring to the surface the poverty which characterizes the lives of so many inhabitants," as Jorge Hardoy and David Sat terthwaite put it in their important 1989 book, *Squatter Citizen*.

The challenge is to create ways of analyzing the vul nerability implicit in daily life. Planners and activists require clear and distinct ways of articulating the vul nerability of urban and rural people, thereby putting their situation on the political agenda and keeping it there. When one maps the physical hazard of landslide or flood, one measures very specific parameters: length and angle of slope, soil type, rainfall intensity, the flow of rivers, size of flood plain, etc. We take such measurement for granted, forgetting that the specific measurements are dictated by working models of how land slides and flood occurs. These models are backed up by a large body of knowledge about the physical world. Hydraulics is ul timately dependent on physics, as is slope dynamics. We tend to take this all for granted because our culture generally accepts the hegemony of Western empirical science, if for no other reason than because of its practical achievements in applied technology over the last few centuries.

If one is to 'map' vulnerability to disaster, then, one must also choose measurements in a pragmatic way. Yet, indicators of vulnerability cannot be chosen at random from daily life. They must be chosen by reference to assumptions made about social causality and the linkages that actually exist among whatever 'parts' we believe constitute society. What we believe are important char acteristics of vulnerable households at the 'micro' level must coherently relate to beliefs we hold about society as a whole (the 'macro' level) and whatever mediating en tities we believe link 'macro' and 'micro'. This social data (at the grassroots or micro level) will stand in a similar relationship of give and take with a variety of models of disaster occurrence. Models in turn will refer back to social theories. Here, of course there is a major differ ence. While most people are unaware of the fact that scientific theory is not unitary and unquestioned, even at the fundamental level of physics, most are abundantly aware that social theory is a matter of controversy.

Most work in hazards research, disaster response, and risk assessment assumes the validity of conservative or liberal social theories in which society is believed to be made up of individuals who optimize their perceived interests more or less rationally in the context of insti tutions such as governments and markets that are more or less successful in harmonizing the interests of large numbers of individuals. There are alternative understandings of social existence that are relevant to vulnerability analysis.

For instance, it is equally possible to believe that society is made up of groups whose interests cannot necessarily be made harmonious but are in conflict. Over long periods of time the daily acting out of these con flictual relations gives rise to institutions, and these institutions – in turn – influence the way groups identify themselves and their interests. Over the last few hundred years the dominant groups have been white male owners of capital and technology. They have influenced power fully the way others in society define themselves, their place in the world, and – of particular relevance to disaster research – just what words like 'progress', 'devel opment', and 'risk' mean. In this view of society, power takes the form of ownership of resources and monopoly of lethal force, and also the ability to define agendas. The powerful in society discipline not only the bodies, but also the minds and imaginations of the majority of people.

Thus individual action or agency – in particular, in dividual hazard perception and choice of behavior in the face of hazards – is constrained not just by imperfect information (the type of constraint recognized by rational choice theory), but by the whole complex give and take relationship between agency and structure.

Assessing Vulnerability

There is an increasingly common approach to com munity based disaster management that has at its core a method of self assessment of coping and capacity. This method has developed slowly over the past 40 years, ever since development workers first began noticing the phenomenon of differential vulnerability/capacity in the face of natural hazards. In brief, it emerged as some of us began to synthesize the field observations that were coming in from different parts of the world: the Sahel famine (1967–73), the 1970 cyclone in Bangladesh, and Hurricane Fifi in Honduras (1974), for example. There were several common elements in all of these obser vations. Chief among them were:

- Death, injury, loss, and the ability to recover (i.e., vulnerability) were highly associated with livelihoods (their nature and their security).
- Vulnerability was not only an economic matter, but depended also on location and access to political power.
- Vulnerability was not homogeneous in 'communities', but varied widely.
- Capacity also existed. Farmers had coping strategies that relied on indigenous technical knowledge, social networks, and alternative income generating activities.
- National government officials did not understand or trust such capacities, and national counter disaster strategies generally came from the top down (if they existed at all in marginal, peripheral zones). On the whole these made the situation worse.

Responding to these observations, an approach was developed in the 1980s and 1990s for defining and analyzing vulnerability and capacity that linked these concepts to the livelihoods, locations, and ecological conditions of households, to political access and 'voice', and to local knowledge and social relations.

From a practical point of view, nongovernmental organizations (NGOs) and other development institutions simplified the approach by creating taxonomies of 'vulnerable groups': for example, women, children, elderly people, people living with disabilities and chronic health problems, ethnic and religious minorities, and illegal immigrants. This taxonomic approach is still quite common, and manifests as 'check lists' used by various NGOs in their efforts to provide assistance in conflict, disaster, and other humanitarian situations.

However, an alternative to the taxonomic approach has slowly developed. Built on the principles of participatory action research (PAR), the alternative aims to build enough trust, common purpose, and motivation among a group of people so that they can define their own vulnerabilities, using a variety of simple tools (hazard mapping, time budgets, problem trees, wealth ranking, and so on) and ask key questions (e.g., what are our strengths/opportunities/weaknesses/threats?). This is the form of self assessment at community level that is the basis of community based disaster management, which is very different from global assessment and measurement methodologies.

Self assessment is proactive: it does not focus solely on hazards and vulnerability, but also on capacities. It takes a problem solving perspective. The approach is situational: it is place and group specific. It takes into account specificities, change, and surprise. It is therefore a special case of what is more formally known as 'adaptive planning'.

What groups in the Philippines, Bolivia, and Zimbabwe (among other places) are doing is also dialogical because there is no 'expert' or 'teacher'. The facilitator seeks to understand the reality on the ground and find the way forward together with the participants. In this way, it is also natural to begin with local knowledge (e.g., of soil, weather, pests, ocean tides, and storms, etc.). Outside knowledge may well be brought into the mix, but as knowledge that is added, and not as a replacement for the vernacular system of understanding.

Specific versus Comprehensive Vulnerability Indexing

No single measurement or indexing scheme suits the needs of every user. At the local level most of the national and subnational level indicators are not useful. Those based on indices of corruption and human rights, as well as those based on the Human Development Index (HDI), as well as those that concern the likely fiscal consequences for the national treasury of a large event are also not relevant at the local level. At the municipal level the decision makers are more likely to be concerned not with populations in a statistical manner, but with specific groups of people in specific locations, who live in specific ways. More often than not, these groups/places/livelihoods are directly or indirectly represented on the municipal council or by civil society in a more face to face and direct manner than at subnational and national level.

Given the division of labor in governmental administration and, to a large extent also within nongovernmental and international organizations, problem orientation or theme, as well as scale, influences the kind of indictor chosen. For example, if food security is the primary focus, then the kinds of factors used by the World Food Programme (WFP) in their Vulnerability Analysis Mapping (VAM) may be appropriate:

- the physical availability of food stocks,
- people's ability to produce or purchase food, and
- people's ability to achieve an acceptable level of nutrition.

WFP's website states that VAM products try to predict the likelihood of a population maintaining food security by looking at how exposed to risk it is, and how well it can protect its food security when challenged by a hazard, anthropogenic (including violent conflict) or natural.

If the concern is broader than simply food security, and, for example, a local government wants to anticipate, plan for, and reduce the risk of high impact extreme events, indicators of the following sort are the kind that have been found useful:

- Numbers of people living in exposed locations (flood plain, steep slopes, unprotected coastal area).
- How these people are organized (or not) and how they communicate; what resources, skills, and knowledge they have; savings, insurance, dependence on remittances, on employment/self employment, on accessibility to markets, dependency on infrastructure (such as electricity).
- Numbers and locations of spatially remote people or people who have spatially constrained evacuation routes (seasonal mud, poor or nonexistent roads, no access to telephones or other forms of information and communication technology).
- Numbers and locations of schools in remote and/or exposed locations and whether these schools have disaster plans.
- Numbers and locations of healthcare facilities in remote and/or exposed locations and whether these healthcare units have disaster plans.

The first three points could be treated together by dividing total number of persons or households into categories by level of organization (e.g., well organized, moderately well organized, unorganized) and resources/skills/knowledge (e.g., high, medium, and low). The categorization would, of course, be done locally both by experts and by focus groups.

Issues

Why Study Vulnerability?

In the twenty first century, after nearly seven failed 'decades of development' since the end of World War II, all researchers should ask some simple questions. Why am I studying vulnerability? Who will benefit from this study? What responsibility do I have to the vulnerable people who are willing to share their lives and knowledge with me?

These questions imply the recognition of the horizontal and vertical multidimensionality and com plexity of vulnerability and the diversity of stakeholders, decision makers, and interest groups at different levels who act according to their political mandate and material interests. That means, for example, that the Social Pro tection Unit of the World Bank or a governmental institution of a developing country or a local envi ronmental justice movement looks at vulnerability from different perspectives. Some focus more on the coping or the exposure side, for example. Diverse groups of decision makers see vulnerability reduction in a specific context such as spatial planning, poverty alleviation, disaster preparedness, or political empowerment.

There are power dynamics at work. Interests are at stake in any assessment of vulnerability. Although the UN and other organizations emphasize 'community participation' and co responsibility for the creation of a 'culture of prevention', in fact, in most places empowerment of communities to protect themselves and to demand social protection from government is seldom a by product of vulnerability assessment.

One might also ask, "What do decision makers need to know about vulnerability?" This question is sensitive and can even be a dangerous one. Politicians usually take what they want from any report or assessment. It might be futile to debate whether it is sufficient to provide decision makers with aggregated overview information, or better to try to communicate in depth assessments and detailed research outcomes. The latter might enable political decision makers to make better judgments, but such a desirable outcome depends on the entire political system, not on whether a certain key political figure has more or less information available. Even worse, some times politicians use the information they get for their own purposes. In this context it is important to lay open the assumptions constituting aggregated data or any as sessment tool for measuring vulnerability. Thus, special attention needs to be paid to the process of the devel opment of knowledge based vulnerability indicators and how to smoothly translate them into politically applic able assessment tools. To ensure that this is understood by the general public and not just by political decision makers, a simple language should be used, including the local vernacular.

Can Research Communities Cooperate?

The five research communities discussed earlier have grown in parallel with only a little cross fertilization. This is wasteful of time and resources. Cooperation needs to accelerate among those working on vulnerability from the point of view of development studies, hazard research, humanitarian assistance, climate change, and public health. The difference in usage of the term vul nerability discussed above is not accidental. It reflects underlying differences in assumptions about society, history, and the nature of science. These underlying differences need to be drawn out and discussed explicitly and frankly so that a combined effort can address the wide array of threats to humanity that the twenty first century has brought with it.

See also: Climate Change; Environmental Hazards; Indigenous Knowledges; Livelihoods; Nongovernmental Organizations; Participation; Poverty.

Further Reading

Anderson, M. (1999). *Do No Harm: How Aid Can Support Peace or War*. Boulder, CO: Lynne Rienner.
Aysan, Y. (1993). Keynote. In Merriman, P. A. & Browitt, C. W. A. (eds.) *Natural Disasters: Protecting Vulnerable Communities*, pp 1 24. London: Thomas Telford.
Birkman, J. and Wisner, B. (2006). UN EHS Source No. 5. Measuring the Un Measurable: The Challenge of Vulnerability. Bonn: United Nations University Institute for Environment and Human Security.
Blaikie, P. and Brookfield, H. (1987). *Land Degradation and Society*. London: Longman.
Cannon, T. (2000). Vulnerability and disasters. In Parker, D. J. (ed.) *Floods*, vol. 1, pp 45 55. London: Routledge.
Chambers, R. (1983). *Rural Development: Putting the Last First*. London: Longman.
Chambers, R. and Conway, G. (1991). *Sustainable Rural Livelihoods: Practical Concepts for the 21st Century*. Discussion Paper 296, Falmer: Institute for Development Studies.
Dow, K. (1992). Exploring differences in our common future(s): The meaning of vulnerability to global environmental change. *Geoforum* 23, 417 436.
Enarson, E. and Morrow, B. (eds.) (2004). *Through Women's Eyes: Gender Terrain of Disaster*. Miami, FL: International Hurricane Center.
Hardoy, J. and Satterthwaite, D. (1989). *Squatter Citizen*. London: Earthscan.
Shipper, L. and Pelling, M. (2006). Disaster risk, climate change and international development. *Disasters* 30(1), 19 38.

Wisner, B. (2004). Assessment of capability and vulnerability. In Bankoff, G., Frerks, G. & Hilhorst, T. (eds.) *Vulnerability: Disasters Development and People*, pp 183 193. London: Earthscan.
Wisner, B., Blaikie, P., Cannon, T. and Davis, I. (2004). *At Risk: Natural Hazards, People's Vulnerability and Disasters* (2nd edn.). London: Routledge.

Relevant Websites

http://www.gdnonline.org
Gender and Disaster Network.
http://www.odihpn.org
Humanitarian Practice Network.
http://www.ids.ac.uk
Institute of Development Studies: Livelihoods Connect.
http://www.ipcc.ch
Intergovernmental Panel on Climate Change.
http://www.interragate.info
In TERRAgate: Coalition for Global School Safety (COGSS).
http://www.islandvulnerability.org
Island Vulnerability.
http://www.proventionconsortium.org/?pageid=39
ProVention Consortium: Community Risk Assessment Tool Kit.
http://www.radixonline.org
Radix: Radical Interpretations of Disaster and Radical Solutions.
http://www.riskred.org
Risk Reduction Education for Disasters.
http://www.grid.unep.ch
UNEP Activities, Early Warning: Disaster Risk Index.
http://hdr.undp.org
United Nations Development Programme: Human Development Report 2007/2008, Fighting Climate Change: Human Solidarity in a Divided World.
http://hdr.undp.org
United Nations Development Programme: Human Development Report Office.
http://www.unisdr.org
United Nations International Strategy for Disaster Reduction.
http://geohotspots.worldbank.org
World Bank Disaster Hotspots Database.
http://www.wfp.org
World Food Programme, What We Do: Vulnerability Analysis Mapping (VAM).
http://www.who.int
World Health Organization Health Action in Crises.

W

War

V. Mamadouh, University of Amsterdam, Amsterdam, The Netherlands

Glossary

Conflict A competitive or opposing action of political actors, for example, a conflict between two states.

Civil War A war between two groups within the same state (generally with the intention of one group to become a state of its own while the other defends the existing state).

Defensive War A war to defend the own territory and population against foreign aggression.

Interstate War or International War A war between two states.

Just War A war that is seen as just, a war in which international regulations are respected.

Military Of, or related to, war and war-waging institutions, soldiers, and weaponry.

National Liberation War A war to obtain independence from foreign rule.

Peacekeeping The enforcement and supervision of a truce between two conflicting parties (two states or two groups within one state) by an external, generally international force.

Preemptive War A war to prevent an imminent aggression towards the own territory and population.

Resource War A war or conflict around natural resources.

War Open and declared armed conflict between political groups, generally between states.

World War A war in which most states in the world are engaged in.

Historical Shift in the Study of War

Historically geography has long been closely related to the waging of war: geographical knowledge about foreign places and people was often gathered during military campaigns and conquests, while this knowledge was primarily used to conquest new territory and control and rule it after war. Geographers were often part of military or imperial agencies. Nowadays most geographers working on war and related issues tend to distance themselves from these institutions and to promote peace and conflict resolution, rather than to take side in a particular war.

Since the institutionalization of geography as an aca demic discipline at the end of the nineteenth century, the ways geographers consider war have drastically shifted. This shift reflects changing ideas about war waging in (Western) society. In the past, when war was seen as a normal practice of foreign policy, geographers were prone to offer their services and provide knowledge and insights to policymakers and war wagers. 'Geography as an aid to statecraft' was the well known motto of Sir Halford Mackinder, the first reader in geography at the University of Oxford (1887). Warfare being an important branch of statecraft, it was an important field of applied geography.

The Great War (1914–18), or World War I as we know it now, was a turning point in Western thinking about war. The scale of the war had been profoundly disturbing, both the number of casualties (about one fourth of the 1891–1895 male cohort in France!) and the number of parties involved, including colonies of European states and states outside Europe, like Japan and later China, Brazil, and the USA. The eventual involvement of the Americans was key to the post war peace negotiations. President Woodrow Wilson promoted new principles for international relations (enunciated in the so called *Fourteen points*) that resulted in the establishment of the League of Nations and the prevention of war as a col lective project of member states. Geographers played an important role at the peace conference, as advisers for the redrawing of political borders in the states that were established in the defeated multinational empires.

Still, the League failed to prevent the rearmament of Germany and a new world war started in Europe in 1939. After World War II (1939–45) and its devastating impact on civilian populations, the League was replaced by the Organization of the United Nations and war truly be came an exceptional, condemnable state behavior, justi fiable only in last resort. A symbol of this change is the re naming of most Departments (or Ministries) of War around the world into Departments (or Ministries) of Defence. After World War II geographers took a distant stance towards the analysis of war, its causes and its ef fects, and when they ambitioned the application of geo graphical knowledge, they aimed at conflict prevention and conflict resolution. Their moral condemnation of war as social activity brought about intellectual neglect: war has no entry in the *Dictionary of Human Geography* edited by Ron Johnston, Derek Gregory, Geraldine Pratt, and Michael Watts (fourth edition published in 2000).

Still geographers have been engaged with wars in different ways. This article presents three broad groups of geographical perspectives: military geographies of war, human geographies of war, and the military, and political geographies of war and peace. Before proceeding it is necessary to scrutinize the term 'war' in more detail.

Definitions of War

At first sight, every one knows what war is. It is however difficult to define war and to distinguish it from other forms of political violence. When is an armed conflict a war? What are the differences between war, guerrilla, terrorism, occupation, and repression? Neither is peace easily defined as the simple absence of war and the freedom from insecurity.

A war could be defined as a state of open, armed, conflict between political groups. It involves the use of force. Generally war is reserved for conflicts of a certain size, for example, the amount of 1000 casualties is often used as a threshold between a militarized conflict and a full scale war. The two main categories of wars are interstate wars – wars between states, and civil wars – wars between groups within the same state. Just any use of force in a conflict is not necessarily a war. Policing activities including the 'occupation' of foreign territories or the 'repression' of subversive political movements are gener ally not seen as war if they are understood as the legit imate intervention of the state. Likewise, 'guerrilla' (from the diminutive of *guerra* – war in Spanish) is used for a conflict in which local paramilitary groups are operating in raids against military or police forces. It points at actors and tactics that differ from a war situation. 'Terrorism' pertains to the use of violence, or the threat of using violence, to intimidate citizens, generally to achieve pol itical goals. Terrorism can not only be used by political groups to pressure states, but also by states against the population of another state or against its own population to prevent political opposition. Accidental civilian deaths during wars are generally not seen as terrorism. A conflict can progress from one type to the other, for example, think of a liberation movement that begins as a guerrilla against the state it sees as illegitimate, uses terrorism to weaken that state, stirs a civil war, and finally obtains independ ence. The reverse can also be true, as the recent devel opments in Iraq show: an interstate war can evolve into a civil war when part of the population does not recognize the post war regime.

Interstate wars are regulated by international laws regarding the declaration of war (*jus ad bellum*) and the conduct of war (*jus in bello*), such as the appropriate (i.e., proportionate) use of violence, the treatment of soldiers, civilians, and prisoners. Among the most important international treaties regulating warfare are the *United Nations Charter* (1945), the *Geneva conventions* (1949) and the additional protocols (1977 and later), and the *Inter national Criminal Court Treaty* (1998). The later court, es tablished in 2002 in The Hague, can prosecute individuals for genocide, crimes against humanity, war crimes, and the crime of aggression. However major geopolitical actors, notably the United States, but also Russia, China, India, and Iran, did not sign this treaty. By contrast, civil wars imply that the legitimacy of the state is contested: there are no similar 'rules of war'; and guerrilla and terrorism are defined by the use of unlawful force.

The term 'just war' is used to characterize a war waged according to the rules: that is a war waged for just reasons (righting some wrongs such as the illegal appropriation of territory and violation of human rights) that is, *jus ad bellum*, and conducted properly (discrimination, pro portionality, minimum force) that is, *jus in bello.* Recent debates about the ethics of war also proposed rules for the proper ending of war, that is, *jus post bellum.* Whether a state does or does not comply with these rules is often contested.

An important distinction in this context is the dis tinction between defensive war and preemptive war: the first describes a legitimate reaction against aggression by another state (self defence according to Article 51 of the UN Charter), while a preemptive war attempts to defeat an imminent offensive or invasion. Some argue that a preemptive war is also self defence as it aims at gaining a strategic advantage in an imminent war that is believed to be unavoidable. Others see a preemptive war as an act of aggression. Therefore different interpretations coexist regarding, for example, the American war on Saddam Hussein's Iraq in March 2003.

In addition, a number of common terms qualifying war should be explained. These notions generally reflect the scope, the goals, and the specifics of warfare. Local war, regional war, and world or global war, qualify the scale of the war theatre and the actors involved. Prot agonists are not always directly involved in the war: a proxy war is a war in which two powerful states are not involved directly, but through third parties. The Cold War was the period of conflict between the USA and the Soviet Union, when direct open conflict between the two superpowers was avoided, despite open hostility. An independence war, or national liberation war, is a war to secure independence from foreign rule; a secession war aims at seceding from an existing state. Nowadays less common, are dynastic wars (between proponents of two candidates to the succession of a deceased monarch), colonial wars (between colonial powers to control colonies), and trade wars (in the literary meaning of the world war, such as the Opium Wars in the nineteenth century). Tactics and the military hardware employed are other elements to characterize a war. Warfare can be naval or aerial and conventional or unconventional (implying clandestine, covert operations). Conventional warfare can also mean the use of conventional weapons, as opposed to ABC warfare, that is using atomic (i.e., nuclear), biological, or chemical weapons. For information warfare in the age of the Internet, new terms such as cyberwar and netwar have been coined.

Finally the metaphorical use of the term 'war' needs to be noticed, as it has also been used in foreign policy to strengthen collaborative efforts under US leadership, noticeably in the War on drugs and the War on terror.

Military Geographies of War

As explained above, since World War II academic geographers have avoided being involved in contributing to the application of geographical knowledge to the waging of war. There is of course a lively, if not visible, community of geographers in military circles, especially at military and naval academies. These geographers play an important role in the education of the military personnel across the world. Military geographies of war apply geographical knowledge to wage war at different scales: tactics and knowledge of terrains and climate at the local level (how to win a battle), strategies and logistics at the regional level (how to win a war), and geostrategy and geopolitics at the global scale (how to secure global support).

Military geographies are closely linked to technological developments as both the available and the necessary geographical knowledge change with technologies. The logistics of a naval warfare differ from that of aerial warfare or of terrestrial operations. They also changed with the range of the weaponry and with the energetic autonomy of ships, submarines, and planes. The use of aerial bombing – first in colonial wars by the British in Iraq in the 1920s, by the German *Luftwaffe* for Franco's troops in Gernika/Guernica in 1937 during the Spanish Civil War, and later at a large scale against many European and Japanese cities during World War II – implied a major shift regarding the geographical knowledge needed for that terrain: not anymore the battlefields and their physical components, but rather existing cities and infrastructures, and the people and activities inside them.

The nuclear weaponry used at the end of World War II dramatically changed the scope of warfare as it could bring about mass destruction. The precarious equilibrium between the two superpowers during the Cold War was often explained by the fact that they would not use nuclear weapons because that was synonymous with a 'mutual assured destruction' (for which the acronym was appropriately MAD). The proliferation of nuclear weapons and the number of states holding them, however, reintroduced insecurity, even for MAD believers.

Technological improvements in the field of information and communication technologies also impacted dramatically on military geographies: think in the past of the telegraph, weather broadcast, radar engineering, and aerial photography, and more recently of Geographical Information Systems and Science (GIS), Global Navigation Satellite System (GNSS) such as Global Positioning System (GPS), and remote sensing.

In the post Cold War period, the role of the military also dramatically changed. Military operations shifted from conventional operations to defend or invade a specific territory, towards overseas international interventions for peace enforcement, peacekeeping, and more recently reconstruction. The military is deployed overseas, often in an international force, such as UN peace forces. They operate in urban areas, rather than in battlefields and trenches, and they interact mostly with civilians. In the American military jargon such operations are known as 'military operations other than war' (or MOOTW). They include humanitarian assistance, disaster relief, and the evacuation of noncombatants, but also support to civil authorities, arms control, enforcement of sanctions, operations in the war on drugs and the war on terrorism, support of insurgency or alternatively counter insurgency (depending on US interests), strikes, and raids. It goes without saying that geographical knowledge of the places they have to intervene in is of eminent value to the military for the effectiveness of these operations in peacetime.

Human Geographies of War (and the Military)

Diametrically opposed to these military geographies are the approaches of geographers who study the impact of war on human societies. These geographers aim at disclosing the longlasting and devastating effects of war activities. Their work forms a varied body of human geographies of war that has been recently expanded to the military in general, not only the military in war operations.

Human geographies and historical geographies of the landscapes of war have demonstrated the long lasting problems in regions where wars have been waged. Almost a century after the Great War, explosives remain a liability for farmers on the lands of the former trenches in the North of France. Prior to major infrastructural works in European cities, remaining bombs regularly have to be dismantled. More recent wars have an even more tragic legacy, especially those in which anti personnel mines and cluster bombs have been widely used. Landmines and explosive remnants kill and injure thousands of people every year and has considerably slowed down the reconstruction of the affected regions. The UN has supported the Mine Ban Treaty (or *Convention on the Prohibition of the Use, Stockpiling, Production, and Transfer of Antipersonnel Mines and on their Destruction*) signed in Ottawa in 1997 (but then again key geopolitical actors like the USA, Russia, and China are among the nonsignatories) and deploys the UN Mine Action Service to carry out de mining activities and provide mine risk education and assistance in affected countries.

Another aspect that geographers studied extensively is the destruction of cities by aerial warfare, especially the bombing of British cities by the German air force and the bombing of German and Japanese cities by the allied air forces. Cities have also been the targets of systematic destruction: snipers in Sarajevo in the 1990s to destroy the multiethnic Bosnian city, Israeli bulldozers in Jenin in the 2000s to destroy Palestinian urbanity. The term 'urbicide' has been coined for this deliberate attack on cities and the pluralism they represent.

Wars generate specific landscapes, but not only landscapes of destruction. These are memorial landscapes with war cemeteries and war monuments, generally located close to the battlefields where the soldiers fell, or near military academies, or in the places of origins of the soldiers. These memorials have been important icons of the sacrifice of veterans for the nation and powerful symbols of national identity. The smallest French village generally has a memorial for those who died for the nation, and ceremonies are held on Remembrance Days. In capital cities, national war monuments like the Vietnam Veterans Memorial in Washington, DC, or Tombs of the Unknown Soldier attract large flows of visitors. Finally wars' legacy includes museums commemorating specific war episodes. Some places have become memorials regularly visited by veterans and relatives of the fallen. Others became major tourist attraction, such as the tour in Normandy along the beaches and the battles of Operation Overlord (still known as D Day, that is 6 June 1944). Slightly different but obviously related are the memorials and museums commemorating civilian victims of war, like the Peace Memorial in Caen (in Normandy also) or the Hiroshima Peace Memorial Park, the Homomonument and the Anne Frank House in Amsterdam, several concentration and extermination camps that have been preserved for that purpose, and Holocaust memorials and museums, sometimes located far outside Europe.

Other war generated landscapes are the borderlands resulting from the territorial arrangements agreed on at the end of the conflict, including the effects of civil war on the residential patterns of different groups. Some geographers have been involved in border drawing and territorial arrangements, others have studied how shifting state borders and border fortifications have affected the development of places and their communities in the affected borderlands.

Human geographies of the impact of war have been broadened to consider militarism and the preparation of war, especially to research the impact of military bases, armament production sites, and training sites. Certain local economies depend heavily on defence contracts for their industries or on the military as the single major employer. Changes of military policies (cut backs in military spending, base closures) and war operations overseas have a direct and deep impact on these communities. Training sites have been seriously affected by the ecological consequences of military activities: nuclear tests, bombing exercises, and shooting trainings are important contaminators.

Last but not least, feminist geographies have recently opened new directions in the investigations of the geographies of the militarization of everyday life. The military draft has long been, and still is in many countries, the most direct and obvious connection between citizens and the military ambitions of the state they belong to. Feminist geographers analyze the impact of the military not just on the lives of soldiers, but also on their families and the civilians living and/or working around the military bases.

Political Geographies of War and Peace

The third and last category of geographies of war consists of political geographies of war and peace. These are geographical approaches that try to explain war as a specific type of relations between states. Political geographies of wars consider spatial patterns of war and peace and disclose the factors that might explain these patterns. Some geographers have studied the diffusion of conflict. Are there regional clusters of conflicts? Is a war likely to spread to neighboring states? Is a state with a large number of neighbors more likely to wage war, than a state with fewer borders or a state on an island?

Among other geographical factors that might explain the frequent occurrence of war, resources are probably the most often discussed. The term 'resource wars' has been used to qualify wars revolving around resources such as oil, water, diamonds, and so on. According to Le

Billon there are three perspectives on resource wars. First, classical geopolitical approaches simply look at the state's level when researching the correlation between the presence of resources and armed conflicts. They frame conflict as scrambles in the competition for resources and neglect other aspects of these wars. Second, political economical approaches underline issues of resource scarcity and dependence. The dependence on resources can weaken a state and makes it vulnerable for conflict; it can motivate conflict between competing parties wanting to control the resources and therefore it increases the risk of war; and it generates opportunities to wage war as resource revenues can be used to finance weapons and manpower for the hostilities. Third, political ecological approaches adopt multiscalar analyses to research the relations between resource access and control, and vari ous forms of violence. They pay more attention to the complex relations between the control of resources and other motives in such conflicts.

Finally a large amount of studies have demonstrated the role of geographical representations and imaginations in the justification of war. Starting as a critique of classical geopolitics and its unquestioned use of geographical knowledge to justify territorial expansion and war, the ap proach known as critical geopolitics has become an im portant subdiscipline of political geography. It questions the geographical representations that are used by policymakers, politicians, and statesmen to justify their policies and mo bilize popular support. Notions such as geopolitical codes or visions are employed to describe these representations. Another important school of work stems from cultural studies, cultural geographies, and postcolonial studies and questions geographical imaginations (Orientalism, white ness) behind such geographical representations.

It is not just the elites' views that are important. Critical geopolitics scrutinizes the representations ar ticulated in mass media. Some geographers have studied grassroots initiatives for peace and conflict resolution, whether local, national, or transnational. More recently feminist political geographers have disclosed the con nections between geopolitical representations and the everyday life of ordinary people.

Finally, following changes in global politics and in the US foreign policies, political geographers focus less on conventional interstate relations, and more on phe nomena such as global governance, humanitarian inter ventions, transnational terrorism, and counterterrorism.

Challenges and New Directions

Despite the fact that war is certainly no usual topic for geographers, a growing body of geography literature is dealing with it. In this article, three geographical per spectives on war were presented. Military geographies exist wherever there is military training. By contrast, social and political geographies of war are most de veloped in the Anglo American geography. This is to be explained partly by the mere size of the US military and the role of the US in the world politics (and to a lesser extent its main ally the UK) and partly by the strength of the political economy and quantitative approaches but also by the feminist and poststructuralist approaches in Anglo American geography. Despite the seminal critical work of Yves Lacoste on the role of geographical knowledge in war making in the mid 1970s, French geographers did not engage as much with war as their English speaking counterparts. In addition to this geographical divide, the gap between these three geographical perspectives, especially between military and academic geographies, is huge. This is partly the result of the institutional separation of geography in military circles and geography in academic circles. Unlike other policy domains, the practicalities of defence and security policies have been cautiously ignored by academic geographers since World War II. New de velopments in warfare, international relations, and global governance underline this gap. The need for mutual engagement is greater than ever. While the new role of the military warrants critical scrutiny from academic geographers, the military also needs more and more human and political geographical knowledge to carry out its new tasks in the most ethical way possible.

See also: Borderlands; Cold War; Critical Geopolitics; Ethnic Conflict; Geopolitics; Geopolitics and Religion; Military and Geography; Military Geographies; Postconflict Geographies; State; Superpower; Territory and Territoriality; War, Historical Geography and.

Further Reading

Clout, H. D. (1996). *After the ruins: Restoring the countryside of northern France after the Great War*. Exeter: University of Exeter Press.
Enloe, C. (2000). *Maneuvers; the international politics of militarizing women's lives*. Berkeley: University of California Press.
Flint, C. (ed.) (2004). *The geography of war and peace: From death camps to diplomats*. New York: Oxford University Press.
Graham, S. (ed.) (2004). *Cities, war and terrorism; towards an urban geopolitics*. Malden, MA: Blackwell.
Gregory, D. (2004). *The colonial present: Afghanistan, Palestine, Iraq*. Malden, MA: Blackwell.
Hewitt, K. (1983). Place annihilation: Area bombing and the fate of urban places. *Annals of the Association of American Geographers* 73(2), 257 284.
Johnson, N. C. (1999). The spectacle of memory: Ireland's remembrance of the Great War, 1919. *Journal of Historical Geography* 25(1), 36 56.
Kirby, A. (ed.) (1992). *The Pentagon and the cities*. Newbury Park, CA: Sage.
Kliot, N. and Waterman, S. (eds.) (1991). *The political geography of conflict and peace*. London: Belhaven.
Lacoste, Y. (1976). *La géographie, ça sert, d'abord, à faire la guerre*. Paris: François Maspéro.

Lacoste, Y. (1977). The geography of warfare. An illustration of geographical warfare: Bombing of the dikes on the Red River, North Vietnam. In Peet, R. (ed.) *Radical geography*, pp 244 262. London: Methuen.
Le Billon, P. (2007). Geographies of war: Perspectives on 'Resource Wars'. *Geography Compass* 1(2), 163 182.
Myers, G., Klak, T. and Koehl, T. (1996). The inscription of difference: News coverage of the conflicts in Rwanda and Bosnia. *Political Geography* 15(1), 21 46.
O'Loughlin, J. (1986). Spatial models of international conflicts: Extending current theories of war behavior. *Annals of the Association of American Geographers* 76(1), 63 80.
O'Loughlin, J. and van der Wusten, H. (1993). Political geography of war and peace. In Taylor, P. J. (ed.) *Political geography of the twentieth century*, pp 63 113. London: Belhaven.
Ó Tuathail, G. (2002). Theorizing practical geopolitical reasoning: The case of the United States response to the war in Bosnia. *Political Geography* 21(5), 601 628.
Palka, E. J. and Galgano, F. A. Jr. (eds.) (2000). *The scope of military geography: Across the spectrum from peacetime to war*. New York: McGraw Hill.
Palka, E. J. and Galgano, F. A. Jr. (eds.) (2005). *Military geography: From peace to war.* New York: McGraw Hill.
Pepper, D. and Jenkins, A. (eds.) (1985). *The geography of peace and war*. Oxford: Basil Blackwell.
Smith, D. and Bræin, A. (2003). *The atlas of war and peace* (4th edn). London: Earthscan.
Woodward, R. (2004). *Military geographies*. Oxford: Blackwell.

Relevant Websites

http://milgeo.blogspot.com
Association of American Geographers (AAG) Military Geography Specialty Group.
http://www.fsu.edu
Association of American Geographers (AAG) Political Geography Specialty Group, Florida State University.
http://www.cas.muohio.edu/igu cpg
Commission on Political Geography of the International Geographical Union (IGU).
http://www.sipri.org
Databases of the Stockholm International Peace Research Institute (SIPRI).
http://privatewww.essex.ac.uk
Kristian Gleditsch's Expanded War Data.
http://correlatesofwar.org
The Correlates of War (COW), Project.
http://www.fsu.edu
The Issue Correlates of War (ICOW) Project at Florida State University.
http://new.prio.no
UCDP/PRIO Armed Conflicts Dataset, International Peace Research Institute Oslo (PRIO) & Uppsala Conflict Data Program (UCDP).

War, Historical Geography and

N. C. Johnson, Queen's University, Belfast, Northern Ireland

Glossary

Bruno Latour A French historian of science who did the pioneering work in actor-network theory.
Center of Calculation An institution where diverse sources of information are gathered together, rearranged, and recalculated to produce new knowledge.
Friedrich Ratzel German geopolitical thinker who developed the theory that the state was a living organism and followed organic patterns of growth.
Geopolitics The analysis of the relationships between geography and the development of the state.
Military Geography The application of geographic tools, information, and techniques to military problems.
Royal Geographical Society Founded as a learned society for the pursuit of geographical knowledge in 1830 under the name the Geographical Society of London and was granted a Royal Charter in 1859.

Introduction

> Patriots always talk of dying for their country and never of killing for their country. (Bertrand Russell)

> The quickest way to end a war is to lose it. (George Orwell, 1946, Second thoughts on James Burnham. *Polemic.*)

The two quotations with which this article on war begins point to the paradoxical and contradictory nature of the use of large scale physical confrontation to resolve human conflicts. Bertrand Russell and George Orwell each share a skepticism about the ideological, strategic, and moral narratives that underpin the theory and practice of war. Nonetheless, the presence of war historically and the continued use of force across a range of territories around the globe today points toward the continuing appeal of war in the resolution of human difference. The contemporary 'war on terror' precipitated by the 11 September 2001 destruction of the World Trade Center in New York and the many other regional conflicts that are taking place in various locations around the world continue to engage the minds of geographers both at the empirical and conceptual level. This extends a long historical interest in war from a spatial perspective. Whether it is the control of territory (e.g., land), resources (e.g., oil), or cultural values (e.g., religion) that provokes wars geography has been implicated in important ways in their articulation and prosecution as recent attempts at ethnic cleansing in Bosnia illustrate.

War, however, has a much longer genealogy in the history of human societies than the last few decades and it has a significantly deeper pedigree in the history of geographical thought. For more than a century war has been the subject of investigation among geographers and the historical geography of conflict has occupied an important position within the discipline. Geography and geographers have not just studied the theories and practices of war, they have also played an active role in the definition and execution of military strategy. There has been a history of 'applied military geography' in the strategic, diplomatic, and practical dimensions of war. The historical geography of war thus can be organized into at least four traditions of enquiry. First, there is a military geography where the techniques and theories of geographers have been taught in military academies to train officers in the practice of warfare. Moreover during wartime geographers themselves in the service of the state have provided this geographical expertise and research. Geographical knowledge, consequently, has been militarized outside the conventional sites of military education. Second, there is a tradition of geopolitics which investigates the relationships between geography and international relations. Here the focus of attention is twofold: geographers investigate historically the contributions of particular geopolitical thinkers to war and strategic affairs. For instance, the role of Sir Halford Mackinder to British military thinking has been the subject of numerous studies. Additionally there is a critical focus on the broader causes of war and the role that geopolitical thinking plays within a wider international system of strategic affairs. Third, there is a tradition of examining war as an agent of landscape and social change. Here the emphasis is placed on the effects of war on the economic, demographic, gender, and cultural landscapes of particular communities. Fourth, there is an extensive literature which examines the role of war commemoration in the cultivation of political identities. This line of enquiry mainly focuses on the role of space and place in the debates surrounding the making of sites of war memory. In the what follows, key features of these four strands of research are briefly highlighted.

Military Geographies and War

There is a considerable literature on the relationship between geography and military practices. Such military geography has been defined as the application of geo graphic information, tools, and techniques to military problems. Typically these include skills such as surveying, mapmaking and map reading, regional description, settlement patterns, and demography. While these meth ods have been taught in military academies such as the United States Military Academy at WestPoint, during war geographers have been directly mobilized to conduct practical training and to carry out military research. Professional geographers have been overtly and indirectly employed by the armed forces to assist in the prosecution of war strategy. Historical geographers have investigated in detail the role that military geography has played in a variety of wartime situations.

During World War II, for instance, the part played by the Royal Geographical Society (RGS) in the develop ment of intelligence services has been analyzed. As the oldest and most prestigious geographical society of its time the RGS was to play a noteworthy role in British strategic planning and intelligence gathering in the period 1914–19. In the War Office among the numerous intelli gence units Military Operations 4 (MO4) was responsible for cartography and topography. It was known as the Geographical Section of the General Staff. It was assigned the task of compiling maps of territories around the world (except for India where maps were produced by the India Office) and with the provision of geographical infor mation on population, social conditions, and borders. The President of the RGS, Douglas William Freshfield, put the resources and personnel of the society at the disposal of the War Office and soon after war was declared in 1914 the RGS was charged with the compilation of an index of place names for the large scale maps of France and Bel gium. It was also charged with producing a 1:500 000 wall map of the United Kingdom detailing towns, rivers, county boundaries, and railways. This work established the significance of geography and the RGS in the war effort and it put the Society in an important strategic position *vis à vis* the War Office. It was the principal source of cartographic and geographical expertise.

The RGS went on to produce a series of 1:1 000 000 maps of Europe although the static nature of trench warfare made the usefulness of these maps more ques tionable. The anticipated mobile war of sweeping infan try attacks and cavalry charges was replaced by a more stationary affair thus rendering the utility of these maps more limited. As the war expanded into new territories of conflict the RGS responded. They began to produce maps of the Italian Alps in 1915 and as the war in the East took on more strategic significance the RGS abandoned the production of the 1:1 000 000 maps of Europe and instead began to compile an index of place names for the War Office's large scale maps of the Ottoman Empire. It also began work on the production of 1:1 000 000 sheets of Turkish lands. The RGS had fewer accurate maps or information on the Ottoman Empire compared to the Western Front and thus had to rely on gathering infor mation from experts, such as David George Hogarth. He was the leading expert on the archaeology and geography of the Middle East and a member of the RGS Council. He was also a personal friend and colleague of T. E. Lawrence who provided the RGS with plenty of geo graphical information gathered in the field when he was in Damascus and its surroundings.

Hogarth highlighted the significance of having detailed knowledge of the Middle East especially as Allied at tempts to infiltrate Ottoman lands were thwarted. The Admiralty decided to gather their own intelligence which might aid a naval attack on Turkey and they recruited geographers to carry out some of this work. In particular the Admiralty produced a series of geographical hand books (known as the Admiralty Handbooks) through its Naval Intelligence Department and these handbooks were prepared by twenty academics working full time at the RGS's headquarters in Kensington Gore, West London.

The role of the RGS was cartographic and infor mational in focus, providing detailed maps and de scriptions of various theaters of war. However, the function of professional geography was not confined to the conflict. Afterwards the RGS also played a part in the peace negotiations at Versailles. The 1:1 000 000 maps proved indispensable in 1919 and the work of British geographers in producing this data is well acknowledged. Having said that the views of key RGS actors did not prevail in the negotiations surrounding the carving up of the Ottoman territories. Geographer's principal role was in the production of cartographic knowledge rather than on larger geopolitical questions. Unlike other states which did take advice from geographers in developing their post war strategic objectives British geography's role overall proved to be technical rather than strategic and applied rather than theoretical.

The strategic role of American geographers in World War II has been extensively analyzed. While military in telligence gathering is as old as war itself World War II represented a major expansion in the scale and bureau cratization of war related research. In America, in par ticular, this took two forms: the application of science and technology to the development of ever more sophisticated and lethal military weaponry. The Manhattan Project is an exemplar of this practical application of science to military objectives. The second area of research was in the social sciences where the methods of social science were used to assemble and analyze existing information that would aid military strategy. This found formal expression in a Presidential Order that created the US Office of the Coordination of Information (OCI) in 1941. This was reorganized in 1942 into the Office of Strategic Services (OSS), the forerunner to the national intelligence agency (the Central Intelligence Agency, CIA) established in

1947. The task of the OSS was not to invent something completely new as the natural scientists were doing in the Manhattan Project but to take existing information and assemble it in ways that would help military intelligence and planning. The OSS was to become what has been described by Bruno Latour as a "centre of calculation." The library, archive, and map collection were the primary data sources for these social scientists and geographers played a crucial role in the OSS.

Richard Hartshorne, one of the most important voices in pre war American geography, occupied the key Chair of the Projects Committee and was Chief of the Geog raphy Division with the OSS. At its height the OSS directly employed 129 geographers mainly in its Re search and Analysis Branch (R&A) where most of the social science research was conducted. At the end of the war the R&A had accumulated a formidable amount of data including 300 000 photos, 300 000 intelligence re ports and over 2 million maps. The personnel working in the R&A came from across the social sciences (but mainly from history, geography, and economics) and initially work was organized around academic disciplines (e.g., geography division). This was altered in 1943 when research was assembled around the four geographical areas that were considered the theaters of war (e.g., Europe Africa, USSR). Work was then carried out as a cross disciplinary exercise with social scientists working together and sharing information in the production of their reports. Hartshorne was central in devising the form that these reports would take and in ensuring that they would be objective.

Nonetheless, in time the geographers working in the R&A felt that their skills were not always adequate for the job particularly in comparison to other social scien tists. Some of the shortcomings included a lack of knowledge of other languages, insufficient detailed sys tematic knowledge, superficiality of information, and lack of training in statistical techniques. These perceived in adequacies would have implications for the development of geographical education after the war. Nonetheless, during the war geographers contributed to two main areas of research. First, they were central to the pro duction of the Joint Army and Navy Intelligence Studies (JANIS). These volumes resembled regional monographs following a standard format which outlined the physical geography of the region, information on landing beaches, transport, port facilities, cities and towns, and forms of government. They were highly descriptive pieces, nar rowly focused and written in a dry style. The second and most significant role of geographers was in the Map Division. Much of its work rotated around the cataloging of existing map collections so that intelligence infor mation could be quickly accessed and provided. The Map Division answered over 50 000 requests for information during the war and it produced over 8000 new maps. The role of geography in the pursuit of military power was well established by 1945 in the United States.

Overall one of the key threads in a historical geog raphy of war is the role of geographers in the prosecution of war. The connection between the establishment of geography as an academic discipline in the nineteenth century and its potential to aid state strategic objectives was closely linked. Geography was seen to be able to bring practical skills and knowledge to military ends and thus was a useful field of enquiry. Mapmaking and in terpretation, regional writing and description of topo graphical features were all important techniques for the army and it is during conflict that the potential of geography to bring applied knowledge to the field of battle was made large.

However, it was not just in the practices of war that geography could yield productive knowledge, geography has historically also contributed to the development of theoretical approaches to warfare and it is to this geo political history that we now turn.

Geopolitics and War

After World War II geopolitics became both an un fashionable term and an unpopular research focus within geography. This stemmed from the fact that geopolitics, that is, the analysis of the relationships between geog raphy and the practice of statecraft, had become strongly associated with some of the key thinking in the execution of policy in Nazi Germany. Although in recent decades the presumed closeness between the ideas of political geographers and Nazi politicians and military men has been queried, immediately after the war geopolitics lost its intellectual rigor and was marginalized within the discipline. Over the past 30 years, however, there has been a renewed interest in both the role of geopolitics in the history of geographical thought and in the study of contemporary international relations from a geographical perspective. It is in the former area of geopolitics that historical geographers have been interested and work has concentrated on the causes of war, the role of war within a wider system of international diplomatic and strategic affairs, and the influences of the geographical environ ment in the conduct of war. Historical geographers have been particularly focused on assessing the theoretical contributions to war of some key geopolitical thinkers of the late nineteenth and first half of the twentieth cen turies. The term geopolitics itself has Greek origins: the *Ge* or *Gaia*, the Goddess of the Earth, and *polis*, the city state of classical Greece. The etymological roots of the word thus bring together humankinds' terrestrial home and its political organization and control.

Although the coining of the term geopolitics is asso ciated with the Swedish political scientist, Rudolf Kjellén

(1864–1922), the founding father of geopolitical thinking is considered to be the German, Friedrich Ratzel (1844–1904). Born in Karlsruhe, Ratzel initially studied the nat ural sciences before moving toward geography. He re garded geography as a synthesis of the natural and human world in which the nature and distribution of human ac tivities would be explicated. He held the Chair of Geog raphy at the University of Munich and then at the University of Leipzig and it is here that he developed his ideas about human geography and particularly geopolitics. In 1897 he published his seminal work *Politische Geographie* and this book would have a profound impact across Eur ope but especially in Germany in the first decades of the twentieth century. Ratzel's approach was to apply the knowledge he had gained in the natural sciences to understanding human activity and thus to produce a theory of how the geographical dimensions of a state in fluenced state behavior. Central to his thesis was the claim that the state was an organism, similar to those found in the natural world, and that its behavior could be best understood in these terms. A state, he claimed, had both physical and human geographical components and al though the state was a biological entity, it was also a spatial entity. The concepts of territory and location were im portant for Ratzel's thesis and he developed this by sug gesting that as the state was an organic entity it was subject to laws of territorial growth and development. In so doing, he was deploying the vocabulary of evolutionary theory to account for changes in the social and political world.

Ratzel's work laid the foundations for a whole variety of subsequent geopolitical theories that developed in the early part of the twentieth century. Some of this thinking informed and influenced the practice of war and geog raphers have assessed the intellectual, social, and cultural roots of these thinkers. They include the British geog rapher, Halford Mackinder; the American, Isaiah Bow man; and the German, Karl Haushofer. Together these studies have emphasized the relationship between the biographical and the contextual in the evolution of this body of geopolitical theory. These studies have been paralleled by a revived interest in statecraft in general and this body of work has been presented as critical geopolitics. Using social theories of power and the nat uralization of knowledge critical geopolitics has chal lenged conventional understandings of conflict by highlighting and analyzing the role of discourse in the execution, presentation, and re presentation of war in official and unofficial media.

Social and Popular Effects of War

While much of the discussion of geographers' contri bution to a historical analysis of war has focused on theories underpinning the practice of war and military strategy, geographers have also devoted time to investi gating the impact of war on landscape and social struc ture. Within this body of literature are historical geographies which consider the transformative effects war can have on the physical, as well as social landscape. For instance, there have been investigations into the re construction of France after World Wars I and II. Here the emphasis has been on the role of the state and in dividual communities in reconstructing livelihoods and landscapes in areas lying within the war zone. These include, for example, rural reconstruction in the case of the Western Front in France after World War I and a discussion of the legal, economic, and social difficulties of establishing title to land abandoned during the war. The redesign of the urban landscape has also been the subject of inquiry particularly where large sections of cities have been destroyed during carpet bombing campaigns.

As well as looking at the effects of war on landscape geographers have also examined the cultural and social consequences of war activity. Studies have included examinations of the impact of war, for example, on post war gender relations. How do the demographic changes brought about by war have an impact on family life, the gendering of the labor market, family planning, and the social geography of the city? Historical geographers have also examined the impact of war on medical geographies. Such studies have considered the epidemiology and spread of contagious illness through armies; the spread of cholera among British troops during the Crimean War, for instance, has been investigated and the disease's march has been tracked across time and space. Similarly, the effect of war on the spread of illnesses among war refugees has been analyzed. Such studies provide stat istical models of the geographical spread of maladies and map how the state and local communities responded to such epidemics. They emphasize the impact of war on the health of a society among combatants and the civilian population and they provide valuable quantitative data on the spread and containment of infectious diseases in times of conflict.

War and Memory

An area in which historical geographers have contributed substantially to our knowledge has been in the analysis of post war landscapes of remembrance. While wars have important geopolitical and social effects they also have major cultural impacts. In particular, societies have had to make sense of the huge loss of life that often accompanies war. Part of this process has involved the development of numerous post war memory practices. The need to re member and honor the dead, as well as provide solace and healing to the living has meant that states, com munities, and individuals have invested enormous financial and emotional energy in creating sites of memory. These sites, although widespread across the world, vary considerably from region to region.

Official public memory and unofficial private memory have regularly interconnected in the evolution of these memorial landscapes and rituals. While the loss of loved ones in war is a very personal and private experience, it is also a collective and public one as this loss is shared with many others within a state. Consequently around the globe, under a wide variety of different political regimes, public memorials to the war dead have been created and rituals performed around such sites. Historical geographers have been particularly interested in investigating the spatialities of these war memory sites and analyzing how place, representation, and interpretation interact in the making of the social memories of war.

Memorials vary considerably in scale and iconography. They can be simple commemorative plaques placed in towns and villages listing local people who were killed, or they can be colossal national monuments located in prestige positions within capital cities (e.g., the Cenotaph in Whitehall, London). Annual pilgrimages to these memorials and the laying of wreaths at their foot reinforces the states' recognition of the importance of war and the losses it incurs, but these occasions also enable the mass of the population to participate openly in a national event. The tone of the inscriptions on war memorials replicates those found on headstones in cemeteries, but unlike the latter they tend to conceal the class and gender divisions of the people they represent. In this sense they function as 'nationalized' monuments. Geographers have interpreted the debates surrounding the location, design, rituals, and reception of specific landscapes of memory. They have been especially concerned with the symbolic role of such sites in the articulation of a variety of political identities and have emphasized that these sites often generate contradictory and contested sets of meanings to the remembering public.

In the case of Auschwitz in Poland, where the Nazi death camp has been transformed into a permanent memorial site and museum to the Holocaust, there has been continuous controversy over the meaning of the space. The competing interpretations by Polish nationalists, communists, the Catholic Church, and the Jewish population, who were the predominant group murdered in the camp, have influenced the overall representational practices and have been the object of contested cultural meaning. For Polish nationalists, Auschwitz stands as a colossal monument to their suffering under German occupation and it has been symbolically and ritually scripted through a Polish Catholic lens. This found expression in the speeches made by Cardinal Karol Wojtyla (later Pope John Paul II) at the site emphasizing the role of Polish Catholics in resisting Nazi aggression and in the opening of a Carmelite convent on the site in the 1980s. These acts have been seen to diminish the significance of the location as a center for the extermination of Europe's Jews and this dilution of Jewish history at the camp transforms the space from being a holocaust memorial space to a much broader space of Polish national remembrance.

Several studies of war monuments have examined the links between these memorial spaces and the geographies of national identity. In some instances involvement in a war has provided a consolidating narrative around which a population can cultivate a sense of national belonging. The image of a collective sacrifice and communal loss can unite a population around a common set of narratives. For instance, Australian and New Zealand memorials to World War I, particularly those dedicated to their soldiers' role in the Gallipoli campaign, have been important in the articulation of Antipodeans' images of national identity. Memorial sites can also become important as places of reconciliation where civil war is concerned. In the collective memorialization of the American Civil War the erection of public statues helped to mask the serious divisions that the war represented in relation to slavery and instead many of the monuments generated a process of sectional reconciliation that united northern and southern white populations. Finally the racial, gender, and class dimensions of war memorials have also been analyzed by geographers as they reflect and refract images of national identity.

Together historical geography has made a notable contribution to our understanding of the theories and practices underpinning the exercise of war.

From geopolitics to commemorative spaces geographers have deepened our critical understanding of war. By focusing on the spatialities and the landscapes in which war takes place, geographers have highlighted the complex relationships between theories and practices in the planning, expedition, and remembrance of war. Critical geopolitics has helped deconstruct the geographical narratives underpinning war and has challenged discourses that naturalize conflict through the languages of space, territory and identity. Moreover, a geography of war not only offers reflections on the immediate conflicts in question but throws light on broader societal issues including questions of health and welfare, military spending, cultural politics/diplomacy, violence, and gender.

See also: Ethnicity and Resistance, Historical Geographies of; Nationalism, Historical Geography of.

Further Reading

Barnes, T. (2006). Geographical intelligence: American geographers and the research and analysis in the Office of Strategic Services 1941 1945. *Journal of Historical Geography* 32, 149 168.

Charlesworth, A. (1994). Contesting places of memory: The case of Auschwitz. *Society and Space* 12, 579 593.
Clout, H. D. (1996). *After the ruins: Restoring the countryside of northern France after the Great War*. Exeter: Exeter University Press.
Clout, H. D. (2004). Ruins and revival: Paris in the aftermath of the Second World War. *Landscape Research* 29, 117 139.
Heffernan, M. (1996). Geography, cartography and military intelligence: The Royal Geographical Society and the First World War. *Transactions of the Institute of British Geography* 21, 504 533.
Hyndman, J. (2004). Revisiting Mackinder 1904 2004. *The Geographical Journal* 170, 380 383.
Johnson, N. C. (2003). *Ireland, the Great War and the geography of remembrance*. Cambridge: Cambridge University Press.
Ó Tuathail, G. (1996). *Critical geopolitics*. London: Routledge.
Parker, G. (1998). *Geopolitics: Past, present and future*. London: Pinter.
Small Raynor, M. and Cliff, A. D. (2004). The geographical spread of cholera in the Crimean War: Epidemic transmission in the camp systems of the British army of the east, 1854 55. *Journal of Historical Geography* 30, 32 69.
Smith, N. (2003). *American empire: Roosevelt's geographer and the prelude to globalization*. London: University of California Press.
Woodward, R. (2005). From military geography to militarism's geographies: Disciplinary engagements with the geographies of militarism and military activities. *Progress in Human Geography* 29, 718 740.

Relevant Websites

www.ordnancesurvey.co.uk
Britain's National Mapping Agency.
www.cwgc.org
Commonwealth War Graves Commission.
www.oxforddnb.com
Oxford Dictionary of National Biography.
www.rgs.org
Royal Geography Society.

Waste Management

M. Watson, University of Sheffield, Sheffield, UK

Glossary

Energy from Waste Any waste-treatment method that recovers energy, whether electricity or heat, from waste materials, either directly or through the production of fuels. Technologies include incineration, pyrolysis, and methane capture.

Environmental Justice Concern over iniquitous distribution of environmental goods or bads and extending to equitable procedures of decision making over environmental issues.

Landfill Method of solid-waste disposal by filling low-lying land areas. Where regulated, waste is covered with layers of soil and the site is subject to pollution control measures.

Proximity Principle Requires that waste should be dealt with as close as possible to its source of production.

Recycling Means using materials from waste products to make new products.

Reuse Means using a product again after its first use in its original form, whether for the same or a different purpose.

Waste Hierarchy Hierarchical ranking of waste management options based on environmental impact, typically ranking reduction over reuse, over recycling, over recovery, with disposal the last resort.

Introduction

Waste is inevitable. However, there has never been any precedent for the prodigious amounts of contemporary global waste generation. Growing amounts of waste and the diverse spaces and ways in which it is managed have critical implications across scales of analysis, from resource efficiency and climate change to social inequalities and human health effects.

Human geography, like the social sciences more generally, has had little to say about the massive matter of waste until recent years. Waste has become an increasingly pressing environmental and political issue, and in the process has come into the view of human geographers. This developing body of work has begun to map the complexities of waste, addressing themes from international shipments of hazardous waste to the details of householder recycling habits, and from the challenges of governing the rapidly shifting terrain of waste management to environmental justice.

What Is Waste?

For the 1989 Basel Convention, "[w]astes are substances or objects which are disposed or are intended to be disposed or are required to be disposed of by the provisions of national laws." This captures the common understanding of waste as matter to be rid of. But moving beyond broad definitions such as this, either practically or conceptually, raises some key issues.

First, operational definitions of waste, and classifications of different wastes, vary widely between territories. It can be classified in terms of its origin, from nuclear waste to sewage to mining spoil to household refuse, or in terms of its composition, or the methods of collection and management it is undergoing, or its level of toxicity or hazardousness. Just what falls into different measures of waste inevitably varies, as does the reliability of those measures, in part reflecting the range of collection, treatment, and disposal paths waste materials can travel. It is consequently very difficult to compare waste production and management across different territories, or to gain any overall measure of global waste production. There are, however, indications of its scale. For example, the Organisation for Economic Cooperation and Development (OECD) estimated that its 25 member countries produced approximately 4 billion tonnes of waste in 2001.

Second, defining waste as matter to be disposed immediately demonstrates the situational basis for classifying matter as waste. The vast bulk of materials that fall into the classification of waste could remain as resources, even commodities, if only they were in the right place at the right time. An object becomes waste when it crosses a socially contingent line where its apparent value is less than the apparent costs of keeping it. So a wood product manufacturer might recognize the potential economic value of the wood waste generated but not be able to find an economically viable market for it. But the ways in which objects can end up on the wrong side of the waste line are not always matters of self evident economics. The extent to which materials are wasted can be seen as the fallout of the ways in which we organize social systems of production and consumption, and of the organization of how we accomplish our everyday lives. From a cultural perspective, it is clear that the value of objects, and the costs of keeping them, can be emotional and symbolic as well as pragmatic and economistic. Wasting materials may be inevitable practically in the processes of social reproduction, but from anthropological perspectives it can also be argued that the symbolic processes of delineating resource from waste, possessions from refuse, are inevitable aspects of social existence. The situational and essentially social basis through which matter becomes waste adds to the complexity of mea suring waste production; it also provides the grounds for changing the social processes which lead to the pro digious production of waste.

Why Does Waste Matter?

Waste matters, first, because it is a waste. The annual disposal of billions of tonnes of material resources is self evidently counter to any understanding of environmental sustainability. A recent study on resource flows in the United Kingdom (UK) economy estimated that, after 6 months, as little as 2% of input resources by mass will be retained in the economy, the remaining 98% will be waste, indicating the enormity of material inefficiencies in the economy. A report on material throughputs for a number of industrialized nations calculated that material outputs to the environment from economic activity in the United States equated to 25 metric tons per person per year.

Second, waste disposal is a major source of local pollution. This has been the main concern driving waste regulation since the nineteenth century and was the basis for the earliest legislation applying to waste, such as in the UK 1848 Public Health Act. Different forms of waste and of treatment and disposal cause pollution to surface water, groundwater, soils, and air. Even in coun tries with the tightest regulations, there remain sub stantial causes for concern over local pollution; and globally, many territories lag far behind in legislation and enforcement on pollution from waste management pro cesses, and indeed on management infrastructures to be regulated.

Third, waste management processes are a major source of greenhouse gas emissions. Decomposition of organic matter in landfill sites produces methane and together with carbon dioxide emissions from the in cineration of waste, waste management was estimated to account for as much as 8% of developed countries' greenhouse gas emissions in 2002.

Fourth, waste matters because it raises serious issues of social justice. Crudely, the rich are disproportionately responsible for waste production; the poor suffer dis proportionately from the negative effects of waste man agement and disposal. The effects of hazardous waste dumping were the cause of the protests now seen to mark the emergence of the Environmental Justice Movement. That systematic inequalities exist around waste on a global scale was amply demonstrated in the late 1980s. With increasingly strict regulation of hazardous waste disposal in developed countries, the exporting of hazardous waste to developing countries and Eastern Europe developed dramatically, taking advantage of their looser regulation of hazardous waste disposal. Although the 1989 Basel Convention tightened regulation of transboundary shipment of hazardous wastes from OECD to non OECD countries, global trade in wastes continues. This is exemplified by the growth of shipbreaking in Asia after environmental and health and safety legislation made it uneconomic in Europe, where the industry was concentrated up to the 1970s.

The social justice implications of waste management are most apparent where a rich community or nation is dumping evidently dangerous waste on a relatively powerless community or nation. This does not mean that social justice implications are absent in less acute situ ations. Studies of the sociospatial distribution of waste management facilities in the UK have found them to be disproportionately concentrated in areas of relative deprivation.

Managing Waste across Time and Space

To the extent that there has always been waste, there has always been waste management. For most of human history, this has proceeded largely without either insti tutionalization or regulation. Before industrialization, material economies typically operated on a much more local scale. Overall throughput of materials was infini tesimal in comparison to developed countries today. Finally the range of materials circulating in the economy was far smaller. These characteristics enabled local material economies which kept a higher proportion of materials in productive circulation. Industrialization and urbanization were the twin drivers of the need for waste management to become a significant matter of public administration. By the mid eighteenth century in Lon don, there were calls for improved coordination of waste collection and disposal, but it was the mid nineteenth century and the development of the public health movement which gave the impetus to public reform of urban waste management in industrialized countries.

The role of urbanization and industrialization in the massive growth of waste production highlight how waste is a result of spatial and scalar aspects of social organ ization. Industrialization not only massively increased the throughput of materials in the economy but also led to processes of centralization and spatial specialization. As circuits of material flows operate on larger scales and become more specialized, the costs of keeping materials productive in the economy increase. Industrialization also led to increasing diversity of materials in the econ omy, particularly from the mid twentieth century as the increasing complexity, and specialization of materials and technologies brought with it increasing difficulties of materials recovery and reuse, and problems of toxicity and safe waste disposal. Urbanization increases the challenges of waste management first through the concentration of people producing waste and the concomitant lack of space for dealing with it, raising substantial practical difficulties in managing the waste produced by households and urban businesses. Urbanization also entails spatial concentration of consumers, spatially isolated from the diverse chains of production and supply which bring consumer goods to converge in the city, adding to the difficulties of returning waste materials to productive use.

Ironically, in countries like the UK, many economically self sustaining reuse and recycling activities met their final demise just as environmental consciousness began to emerge in the 1960s and 1970s. This period saw the fullest embedding of the paradigm which shaped the twentieth century waste management in industrialized countries. This can be termed the disposal paradigm, characterized by a linear flow of materials from primary extraction through processing, manufacturing, use, and ending in disposal (**Figure 1**). Under this framing, waste is worthless material to be rid of at least cost and within the prevailing environmental regulatory regimes. These regulatory regimes are driven by concerns over pollution as a potential source of human health effects, with little if any attention paid to issues of resource efficiency or broader issues of environmental sustainability. Concomitantly, waste management is framed as a technical, end of pipe service operation, collecting and disposing of waste, characteristically through landfill and incineration.

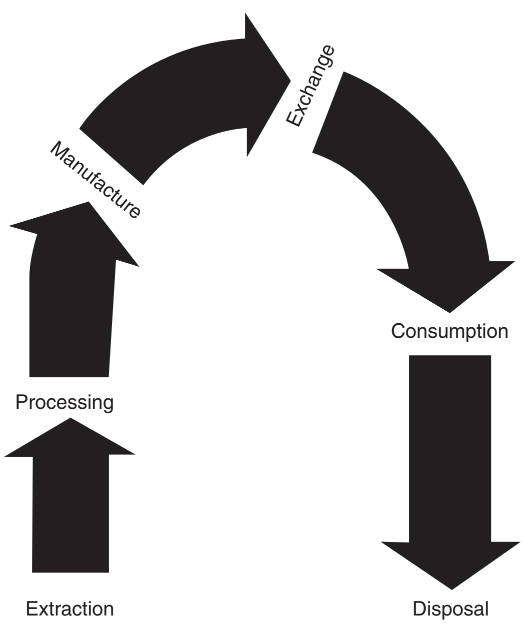

Figure 1 Materials flow in the disposal paradigm.

Shifting Paradigms?

The last decades of the twentieth century saw the gradual emergence of the grounds for a new paradigm of waste management – what might be awkwardly termed the sustainable resource management paradigm. The initial impetus was a rapid tightening of the established central concern of pollution control. The late 1960s and early 1970s saw rapidly increasing awareness of environmental toxicity and policy responses to it. Successive rounds of legislation at national, international, and global level rapidly increased the regulatory standards for waste management, recognizing the harmful effects from waste treatments, such as leachate from landfill contaminating water courses and carcinogens and other toxins from incineration. Although enforcement has consistently lagged behind regulation, key waste management technologies have had to meet increasingly stringent environmental standards.

The 1970s saw the codification of the foundational principles of sustainable waste management. In 1975, the European Economic Community (EEC) Waste Framework Directive formalized the waste hierarchy and the principles of proximity and of self sufficiency. The waste hierarchy (**Figure 2**) specifies the prioritization of approaches to waste. At the top of the hierarchy is the requirement to reduce waste at source, such as through efficiencies of production or extending product life spans. Where waste cannot be prevented, products should be reused. Where that is not possible, materials should be recycled. These top three options – to reduce, reuse, and recycle – together should minimize the amount of materials entering the final waste stream. Where materials do fall through these levels, value should be recovered, such as through recovering energy from waste through incineration. Incineration without energy recovery and landfill languish at the bottom of the hierarchy as last resort options. The proximity principle is that waste should be managed as close as possible to its source; the self sufficiency principle that a (unspecified) geographic

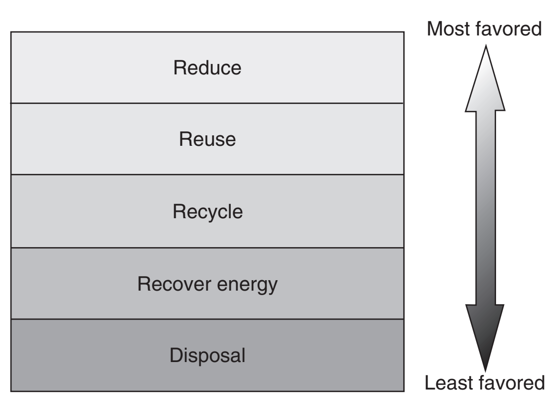

Figure 2 The waste hierarchy.

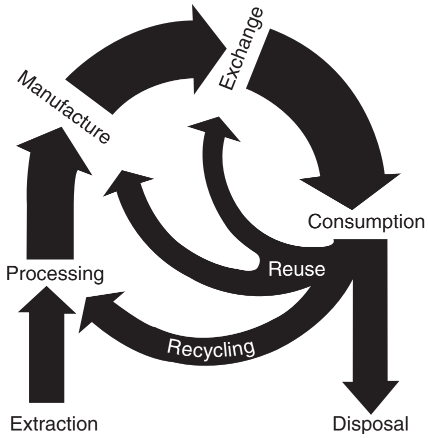

Figure 3 Materials flow in the sustainable resource management paradigm.

area should be self sufficient in waste management. These principles, particularly the waste hierarchy, can be seen as the basis of the emergent sustainable resource management paradigm (**Figure 3**).

For more than two decades after these principles were first expressed as international policy, there was little evidence of the juggernaut of waste growth even begin ning to slow, much less of the step change required to shift toward greater sustainability. Fresh impetus was given to moves toward more sustainable management in the 1990s, when the links between waste management and climate change reached the political agenda. Nevertheless, commitments to the waste hierarchy and proximity and self sufficiency principles have remained more visible as declarations of principle than as recog nizable features of waste management practice. There are some basic reasons for this. The principles of proximity and of self sufficiency are clearly very open to inter pretation and contestation over basic issues of space, scale, and boundaries. For example, the principles of self sufficiency have been used to conflicting ends in sub regional disputes over waste planning in the UK, as rural districts which have historically imported urban waste for landfill argue that those urban authorities should be self sufficient; while the urban authorities can argue that the principles apply at a regional level and therefore legit imate continuing the disposal of waste outside the city. Arguments are also made for the sustainability of actions which clearly contravene these principles. In Western Europe, rapidly increasing volumes of plastics collected for recycling, combined with a high demand for materials and low cost of processing labor in Asia, led to a massive increase in exports of plastics for recycling in the early years of the twenty first century. The trade is defended against critiques based on the proximity principle by pointing out that the plastic was going to Asia on ships that would otherwise be empty after bringing imported commodities. This exemplifies the room for negotiation left open by the flexibility of ill defined declarations of principle.

The waste hierarchy seems less immediately open to negotiation. In principle, the minimum of materials should pass down through each level of the hierarchy. Yet, despite both the OECD and the EEC adopting the waste hierarchy in the mid 1970s, 49% of municipal waste in the European Union (EU) was going into landfill in 2005. This proportional amount of land filling would not be so significant if reduction and reuse had served to reduce the total amount of material going for disposal. However, waste volumes have continued to grow with or ahead of growth in gross domestic product (GDP). In the EU and the European Free Trade Association (EFTA) between 1990 and 1995, GDP increased by 6.5%, but total waste generation increased by 10%. Between 1995 and 2003, the GDP of the nations that became the EU25 increased by 19%, and the generation of municipal solid waste increased by the same percentage. Continuing growth in amounts of waste often outbalance advances in waste management. For example, while the percentage of waste going to landfill may reduce in a country, the in creasing total volume can still mean an increase in ab solute amounts of waste going to landfill. That it is easier to increase proportions of waste going to recycling than it is to reduce the overall amount of materials becoming wastes in the first place is indicative of the profound challenges which face the governing of waste manage ment as it shifts toward environmental sustainability.

Governing the Shift to Sustainable Waste Management

The challenges facing the governing of waste manage ment, from international policymaking to public par ticipation in local decision making, have been one of the dominant concerns of recent work on waste management in human geography, and social science more generally. The key problematic for the governing of sustainable waste management can be related to the institutional, political, and cultural framing of waste under the disposal paradigm. If waste is seen as matter to be rid of in the most cost effective way, then the idea of waste manage ment takes a particular shape. It can be an essentially technical, end of pipe operation concerned with taking whatever matter falls into the waste stream on a linear journey from its collection to its point of disposal. Paradigms in policy are generally seen primarily as sets of ideas which define the problem to be addressed and how to address that problem. However, it is important to recognize that paradigms are also embedded in, and re produced by, the institutions, infrastructures, and work ing cultures through which those paradigms are operationalized. So the disposal paradigm is comprised and reproduced by the institutions which have evolved, often in haphazard fashion, to serve that paradigm, from international government down to the relevant service division of a local authority. It also involves the disposal infrastructures, such as landfill sites and incinerators, often representing huge investments and long term contractual obligations for local authorities. It extends too to the vehicles collecting the waste, to working practices of waste management operatives, and even to the everyday waste practices of people in their work and at home, in the routinized decisions about when a ma terial or object becomes waste and then what should happen to it. The policy ideas of the disposal paradigm flow from these distributed institutional and practical arrangements of waste management as much as the ar rangements flow from the ideas. A change in ideas at the level of international waste policy, then, cannot be ex pected to result in immediate change to the paradigm as expressed in and reproduced by the infrastructures and practices of waste management.

National and international structures of governance, such as in departmental structures and ministerial port folios, divide up responsibilities and powers to act ac cording to particular, historically contingent, framings of the problems of governing. Waste management can fit much more straightforwardly within the boundaries of governmental responsibility when it is a technical end of pipe service than it can when it is a matter of intervening in situations and processes of production and con sumption. The challenges of moving toward sustain ability in waste management for governing structures and processes are in fact common to many environmental issues. Recognizing the interdependence of issues previ ously divided into different policy silos, with each other and with environmental concerns, has lead to successive calls for, and action toward environmental policy inte gration (EPI). Some geographers have used waste spe cifically to interrogate ideas of, and progress toward, EPI from international to local scales of governing.

The difficulties of making progress carry through, from policy principles to practice on the ground, are exemplified by analyses of initiatives and progress in relation to municipal waste management. Recent years have seen rapid increases in rates of recycling of munic ipal waste in many European countries, yet, despite EU commitments to the waste hierarchy, little evidence of progress in reducing the rate of growth in waste pro duction. It can be argued that the prioritization of re cycling in policy responses to the changing waste agenda reflects how well recycling fits into the existing structures and competencies of waste management, in comparison to the activities involved in pursuing waste reduction and reuse. Recycling can involve enrolling householders to new domestic waste management practices, such as cleaning and separating wastes, and involves getting recyclable materials to markets, rather than to a landfill site. Nevertheless, it involves a relatively small section of the overall cycle of materials flow through the economy. Getting material from its point of collection to its point of next use remains a largely technical set of activities that fits relatively well in the structures of governing and of operation inherited from the waste disposal paradigm. Conversely, waste minimization and reuse involve interventions into a far broader range of processes and situations, demanding active relationships with a wide range of different social actors and very different com petencies from those involved in moving bin lorries around efficiently. Pursuing reduction and reuse there fore demands much greater institutional reorientation, arguably to the highest levels of governing.

Relating closely to themes of governing, issues of public participation in waste strategies and in planning waste infrastructures have been a significant concern of human geographers engaging with waste. Waste man agement infrastructures can be framed politically as matters of regional or national strategic importance and are also often major investments for large corporations. Given the tendency for waste management infra structures to be situated in socioeconomically deprived areas, there are very clear grounds to be concerned for the representation of local populations in waste strategies and planning decisions. The shifting waste agenda has to some extent changed the nature of concerns over public participation. First, increasing rates of recycling and composting tend to demand a larger number of smaller waste infrastructures with low levels of environmental risk – such as bottle banks and local material recycling facilities. While the justice implications of a bottle bank in a middle class suburb may be far less apparent, there can still be considerable opposition through the planning process to local authorities putting such infrastructures in place. Second, however, the opening up of the waste agenda, from a technical service to a much more broadly defined social and economic concern, has given local communities and activist groups the basis for critiquing waste management strategies at a more fundamental level. For example, a campaign group in Newcastle, in the northeast of England started life in opposition to an energy from waste incinerator, but quickly broadened to carry out a long term deliberative process resulting, in 2003, in an alternative waste strategy for the city, based on a shift from technologies of disposal to people centered resource recovery.

Attitudes, Values, and Everyday Waste Management Practices

A focus for some geographical work on waste manage ment has been the attitudes and values of people to issues like recycling and composting. In particular, the gap between how much people report they recycle in surveys and what the statistics for recycling show gets recycled (rather less than that is reported) has sparked a number of studies. What emerges is that recycling behavior is often less about individual attitudes, knowledge, or values than it is about the availability of appropriate means of recycling such as multimaterial doorstep collections rather than centralized facilities requiring individuals to bring waste, and the social normalization, both at street level and in wider society, of recycling behavior.

A focus on the small scale practices of domestic space has also begun to take analyses of waste into new terrain. When researchers move beyond looking at what people do with waste to study the everyday processes through which waste is produced, analysis gets closer to recognizing that waste is the fallout of the social and economic processes that converge in the practices through which everyday life is accomplished. Waste management is far from simply a technical operation. It is part of complex systems, reaching from international governance through processes of production and to the mundane everyday practices of consumption and the cultural norms of disposal.

Geography and the Future of Waste Management

It can be argued that currently geographers have fallen behind the progress of policy debates on waste management. On waste, geography, along with other social science disciplines, has tended to continue focusing on those aspects of society which are recognizably about waste. This is entirely unsurprising as the waste management processes being studied empirically have yet to escape from the disposal paradigm. Indeed, there is a foundational paradox to be overcome. To frame the issues involved in waste management as being about waste is to remain trapped by increasingly anachronistic understandings. Moving toward sustainability demands that we find ways to stop materials becoming waste, and few of those ways will be discovered by looking at waste management.

This is where the policy agenda is ahead of waste research, in human geography at least. In policy statements, waste as a theme is increasingly placed within the Sustainable Consumption and Production agenda, in recognition that the foundational challenges of waste are not to be met in waste management but in the material flows, processes, and practices throughout society from which wastes are the fallout. Forging a sustainable materials economy requires analytical perspectives that can encompass the complex political, technical, material, and cultural systems involved; which can travel across scales of analysis from the global processes of trade and environmental impact to the micro social scale of domestic and commercial practices. Perhaps a future encyclopedia of human geography will have a section on the discipline's central contribution to the forging of a sustainable materials economy, rather than one on waste management.

See also: Climate Change; Consumption; Environmental Justice; Environmental Policy; Natural Resources; Scale; Sustainability.

Further Reading

Bulkeley, H., Watson, M. and Hudson, R. (2007). Modes of governing municipal waste. *Environment and Planning A* 39, 2733 2753.

Darier, E. (1996). The politics and power effects of garbage recycling in Halifax, Canada. *Local Environment* 1(1), 63 86.

Davies, A. R. (2003). Waste wars public attitudes and the politics of place in waste management strategies. *Irish Geography* 36(1), 77 92.

Gandy, M. (1994). *Recycling and the Politics of Urban Waste*. London: Earthscan.

Murray, R. (2002). *Zero Waste*. London: Greenpeace Environmental Trust.

O'Brien, M. (2007). *A Crisis of Waste? Understanding the Rubbish Society*. London: Routledge.

Pellow, D. N. (2002). *Garbage Wars: The Struggle for Environmental Justice in Chicago*. Cambridge, MA: Massachusetts Institute of Technology.

Petts, J. (1995). Waste management strategy/development: A case study of community involvement and consensus building in Hampshire. *Journal of Environmental Planning and Management* 38, 519 536.

Scanlan, J. (2004). *On Garbage*. London: Reaktion Books.

Strasser, S. (2000). *Waste and Want: A Social History of Trash*. New York: Owl Books.

Watson, M., Bulkeley, H. and Hudson, R. (2008). Unpicking environmental policy integration with tales from waste management. *Environment and Planning C: Government and Policy* 26, 481 498.

Relevant Websites

http://www.basel.int/
Comprehensive website of the Basel Convention on the Control of Transboundary Movements of Hazardous Wastes and their Disposal.

http://ec.europa.eu/environment/waste
Portal to information, statistics and policies relating to waste in the European Union.

http://www.unep.fr/scp/
Web site of the United Nations Environment Programme Sustainable Consumption and Production Branch, including a section on waste.

http://www.thewasteoftheworld.org/
Website of *Waste of the World*, a five year research programme funded under UK Economic and Social Research Council's Large Grant Scheme. It brings together researchers in geography, anthropology and materials science from the University of Sheffield, Durham University, University College London and Goldsmiths College London, and connects the UK with South Asia (particularly India and Bangladesh), as well as the US, Europe and Kazakhstan.

Water

Fiona Allon, Centre for Cultural Research, University of Western Sydney, Australia

Glossary

Actor-Network Theory A theoretical approach that addresses the mutual constitution of technologies and their consumers and users in networks, and which emphasizes the role such technologies and technical objects play as 'actors' or 'agents' capable of defining frameworks of action.

Big Water A term used to describe modernity's 'big' hydro-engineering projects in the twentieth century (dams, sewerage networks, piping, and plumbing systems) and the wider social relations they embodied. This large-scale water resources development established massive infrastructures for hydro-energy, powerful water/hydraulic bureaucracies, and the technological and institutional systems for water delivery in most Western industrial societies, including state–citizen relationships. It is also referred to as the 'hydraulic mission': an approach to water resources dominated by the engineering profession and technical sciences, and undertaken by colonial and national water agencies.

Demand Management A strategy focused on 'consumption' rather than 'supply' and used primarily as a way of reducing the domestic consumption of water through measures such as low-flow showerheads, water restrictions, fines, etc.

Everyday Water A term used to indicate a paradigm shift in the understanding and analysis of water, from approaches dominated by scientific disciplines associated with natural resource management toward a greater focus on 'ordinary' cultures, settings, and environments such as domestic homes and gardens, and including the cultural conventions, habits and expectations influencing residential water use and consumption.

Hydrology A scientific discipline concerned with the study of water, including the hydraulic cycle and water resources.

Natural Resource Conceptualization of natural phenomena as raw materials able to be used, exploited, and commodified in an efficient and rational manner.

Nature/Culture Dichotomy Conceptualization of natural and cultural phenomena as separate, divided, and easily distinguishable.

Socio-Technical A theoretical approach that seeks to overcome the division between the natural sciences and the social sciences to define society in a way that includes both human and nonhuman elements bound together in interaction.

Introduction/Overview

Water is essential to life and to human existence. It is not only indispensable to the human metabolism, but also to the very sustainability of societies, cities, and nations. It is a social 'agent' central to agricultural food production, and to industrial, economic, household, recreational, and environmental activities. When it is available, water is ubiquitous in the practices of everyday life: we drink it, wash, clean, cook, and bathe with it, use it for work and leisure, and need it to maintain domestic homes and gardens.

Water is also an integral feature of physical geographies and the natural environment. It is present in our lives as rainfall, floods (and droughts), river systems, wetlands, seas and oceans, lakes, dams, and catchment areas. This is a 'natural' world that is in constant transformation, with the ongoing entanglements of human and natural forms and processes continuously producing new landscapes and geographies. Water traverses both nature and society; it has, simultaneously, an ecological presence, a critical life sustaining role, and a host of cultural and spiritual uses and meanings.

Every culture has devised hydraulic methods for storing, controlling, or using water in some way: wells, mills, pipes, taps, drains, irrigation systems, dams, and so on. It is an essential part of many areas and activities of human life, a 'total social fact'. Water, therefore, is rarely meaningful in and of itself – simply as 'water' – but rather is defined by the variability of its social, cultural, and ontological meanings, and the highly diverse ways people know and live with water. Whether it is seen in terms of its scientific properties as a compound of hydrogen and oxygen (H20), or as a substance enabling the technical flows of the hydrological cycle, a means of generating electricity, as the lifeblood of place and community, or simply as a contested resource, water acquires meaning and assumes utility through its interactions with the people who use and live with it. Water, like nature in general, does not have one uniform or universal value or meaning. Its meanings and values are diverse, constantly changing, and complex, derived from the very different contexts in which it exists and is used. Water's diversity is its most important characteristic: a limited resource with unlimited meaning.

Recognizing this variability shifts emphasis to the complex interconnections and relationships between water, place, people, and the nonhuman world. In many communities water has social, cultural, as well as ecological significance, and is a crucial part of environmental belonging and local regimes of resource management. For indigenous communities especially, lands and waters are in a continual and interconnected process of mutual transformation. For Native North Americans and First Nations people, for example, water sources are not only sacred, spiritual sites but also used to sustain indigenous economies such as fishing, and thus have become an important part of Native Title land claims. However, many indigenous groups must also contend with severe problems associated with both the quantity and quality of water supplied to communities. In 2006, over 75 First Nation communities in Canada were advised to boil drinking water due to health risks associated with water contamination and poor quality. Around the world such examples are multiplying. Political struggles, ecological problems, and social tensions are also increasing as more and more people are engaged in conflicts over water.

More than one billion people lack access to clean drinking water, and up to 50% of residents in the large urban centers of developing countries cannot easily obtain clean and affordable water. Worldwide, growing numbers of people are enduring some form of water scarcity. In the developing world in particular, questions of water supply, access and management (both in quantitative and qualitative terms), have become the key issues on local and global political and environmental agendas. Much of Africa is chronically short of potable water while in other parts of the world access to water involves constant contestation and struggle. In the Middle East and in the Americas competition for access to water is intensifying, while in Europe there are increasing problems with water supply and control.

Yet, in other settings, water has become such an essential, ubiquitous part of daily life that its presence is relatively taken for granted. Seen as a natural and normalized part of modern living, its constant, unvarying supply is now expected and unquestioned. At the same time, irrespective of levels of socioeconomic wealth, cities are getting increasingly ‘thirsty’. Many countries and regions are facing acute water shortages, and are implementing a series of measures to cope with the prospect of a lack of supply in the future, including demand management measures such as water restrictions and price increases, as well as new large scale engineering megaprojects. While reduced rainfall due to drought and climate change has contributed to current water shortages, it is apparent that ‘water scarcity’ is only a small part of a much bigger picture. For many commentators, many water problems are not due simply to a lack of water, but rather to a lack of effective water management and governance.

Water Resources

No other natural resource is more contested than water. The circulation of water produces not only physical geographies and material landscapes, but also symbolic and cultural landscapes of power. As the work of geographers such as Erik Swyngedouw shows, wherever there are flows of water, there will also be flows of money and power. Control over, and access to, water is a highly contested terrain, involving conflicts and struggles along class, gender, ethnic, and cultural lines. These multiple power relations occur on interrelated and overlapping geographical and spatial scales, ranging from the scale of the body and the intimate spaces of the household, to the complex distribution and supply networks of cities and regions, to the geo political and global scale of highly unequal, uneven development. The countless conflicts arising from competing interests, stakeholder groups, values, ideologies, narratives, and worldviews are also played out across scales ranging from the level of the household and small distribution systems to the global level.

Contemporary struggles over water – access to river basins and waterholes, demands for safe and affordable drinking water, native title land claims, the political resistance to megadams and water privatisation schemes, and disputes over ‘water rights’ – reveal the contested and political nature of water, and demonstrate that water resources are complex, conflict ridden sites of power, economic control, and often territorial dispute. Problems around water, therefore, cannot simply be reduced to the question of who does or does not have access to water, but rather involve a much wider field of relations, including the myriad meanings associated with nature and a geographic imagination of power and control. This includes the perception of water as a (scarce or abundant) ‘resource’.

Naming water as a resource constructs it as a raw material waiting to transformed and controlled by human labour and technology, as a substance ‘external’ to human societies and best apprehended through instrumental reason and rationality, and through a logic of development. Throughout much of the twentieth century the state, with its large, bureaucratic, and centralized control systems and regulatory authorities, was considered the most rational entity for managing essential resources and services like water. The development of state administered and coordinated water supply and infrastructure was deemed essential for the continual growth and progress of national economies, while the development of the hydrological sciences and related disciplines such as geology and geography also contributed to this view of water as an important resource.

Since the 1980s the institutions, politics, and culture of water supply have radically changed. The privatization of water utilities is increasingly promoted as the solution to today’s ‘water crisis’ and the supposed inefficiency of the public sector and cumbersome state water bureaucracies. With the construction of private water markets, water is taken out of the public sphere and becomes subject to a host of neo liberal economic imperatives and pricing mechanisms. In many regions, water management is no longer an issue to be handled by national govern ments, and is instead controlled at the level of global financial markets by institutions such as the World Bank and IMF.

Alongside these shifts, the 'scarce' character of water resources has consistently been emphasized by water authorities, private enterprise, and environmentalists alike, and has been successful in creating an under standing of water as a valuable resource. Within these debates 'nature' continues to be imagined as unpredict able, as the source of 'crisis' and 'scarcity', and the in crease in water's exchange value is seen as the necessary social response to an inevitable natural phenomenon (drought, water shortages, and climate change).

Once regarded as a 'common' or 'public good', an essential service to which citizens were entitled, water is now constructed as a precious, valuable 'commodity' to be bought by consumers, and one best managed through the market economy. Yet, as many economic geographers have noted, scarcity (along with other terms of value) is an economic concept, socio economically and politically constructed, rather than a feature of nature. The 'scarcity' of water, therefore, is not simply a result of its limited availability in nature, but rather is produced in economic discourse and the categories of thought through which we understand nature as a resource.

One growing area of research highlights these con nections between discursive practices and material–en vironmental conditions. It shows how natural phenomena such as drought, for example, are discursively con structed. In the UK the 'Yorkshire Drought' became one such instance, a public media event that was not only a result of a lack of rainfall, but a much broader response to the ecological and cultural dimensions of socio economic restructuring and the privatization of the English water industry. This article reveals how the very notion of a 'water crisis' produces a particular narrative that in turn affects how people interpret, understand, and create re sponses to it. A water crisis, therefore, must be seen as simultaneously natural and human made, a 'socio nat ural' hybrid that involves interactions between me teorological and climatological patterns, and political, economic, regulatory, and consumption practices (e.g., the conflicts over the urban water supply in cities such as Athens, Atlanta, Buenos Aires, Cochabamba, and Manila).

Big Water: Water and Modernity

The historical geography of water circulation highlights the political and economic discourses and practices, as well as social and cultural power relationships that ac tually become 'built, etched, or baked into' the steel and concrete of the technological and ecological structures of the water system itself. To illustrate this, we could take the example of modern hydrological engineering systems organized by large, centralized decision making bureau cracies. Such systems are still the basis of current infra structures for water delivery throughout most Western industrial societies. These large scale projects that de veloped during the twentieth century have been termed 'Big Water', and are typified by the creation of new landscapes: giant reservoirs and catchment areas, moun tains pierced with pipelines and aqueducts, and of course, massive dams, pipe networks, pumps, and central treat ment plants. As one of the quintessential products of modernity's 'engineering era', Big Water embodied long standing fantasies of mastering nature, including the dream of 'making the desert green', and of creating abundant, always available, flows of water.

The building of dams – the Hoover Dam on the Arizona–Nevada Border, the Marathon Dam in Athens, and the Snowy Mountains Scheme in Australia, to give a few iconic examples – became a key expression of Big Water's vision of technological modernity. At the same time, large scale water resources development and complex networks such as the urban water supply be came the basis of urban expansion and, together with the 'harnessing' of the power of water for energy sources such as hydroelectricity, became a guarantee of a country's status as a modern, modernizing industrial nation. The construction of big dams and massive water infra structures all over the world in the twentieth century represents the apogee of modernity's Promethean project of conquest and control. It was also the pinnacle of the so called 'hydraulic mission', an approach to water re sources controlled by the engineering profession and technical sciences, guided by the domination of nature, undertaken by powerful colonial and national water agencies, and resulting in around 50 000 large dams and 280 million hectares of irrigated land.

At the beginning of the twenty first century, both faith in technological progress and the optimistic belief in a steady flow of always available water have faded. Many dams once seen as engineering triumphs are now re garded as ecological disasters. A survey of the world's 45 000 biggest dams found their construction had displaced around 40–80 million people, majority of whom were poor and vulnerable. Many contemporary dam projects are highly controversial, and international social move ments have been formed to highlight the environmental and social costs to local communities. These struggles have included indigenous people and ethnic minorities protesting against the loss of lands, livelihoods, and cul tural heritage, and claiming the right to retain traditional social rituals and forms of collective water management. Recent examples include: The Itapu Dam bordering Paraguay, Brazil, and Argentina, the Tucuri Dam in Brazil, the Chixoy Dam in Guatemala, the Malaysian

Pergua Dam, the Namarda and Sardar Sarovar Dams in India, the Biobia in Chile, the Arun Dam in Nepal, and the Ilisu Dam in Turkey. In Cochabamba, Bolivia, how ever, local support for a big dam project became part of a wider struggle to resist water privatization and World Bank led neo liberal water reforms, producing new and sometimes conflicting constructions of regional and na tional identity and modernity. This also resulted in Cochabamba's highly controversial 'water wars' (Guerra del Agua), which are celebrated as a victory for local communities and social movements against privatization and have become emblematic of the anti water privat ization campaigns and protests that have occurred around the world since the 1990s.

Despite widespread criticism, development 'mega projects' such as new dams, pipelines, desalination plants, and even membrane covered canals are still promoted as the answer to current water problems. In China, for ex ample, the Three Gorges Dam promised 'water and electricity for all', and went ahead even though 700 000 people were displaced by the project. In Spain, the country's national hydrological plan is one of the most ambitious dam projects ever undertaken, and in the UK new reservoirs (such as the Upper Thames Reservoir) are being built to support new urban housing developments. Meanwhile in Australia some states (e.g., New South Wales and Western Australia) are building large, expen sive, and energy intensive desalination plants to support their urban water networks.

Big Water, however, refers not only to the giganticism of the engineering projects themselves, but also to a whole network of relationships between human beings and nature, nature and the city, the state and citizens, and domestic users and technical authorities/experts. Within these massive infrastructures, water is controlled, circu lated, standardized as 'potable', supplied to users, and also disposed of and eliminated, often after an all pur pose, one time use. The actual system of water man agement – the dams, treatment plants, pipes, drains, water meters, pricing mechanisms, and taps – not only controls the production, circulation, and delivery of water, but also guides the social and human interactions associated with its use, along with the meanings and values attached to its status as a tamed and plentiful resource.

These large scale systems and government owned utilities delegated with responsibility for supplying the continuous flows of water across the modern city went hand in hand with a dispersed system of domestic users who naturally came to expect the ready availability of water in their households. Simply by turning a tap the consumer accessed a seemingly constant flow of water whose source (both physically and symbolically) was far removed from the urban environments where it was ac tually used, but whose presence symbolized the engin eer's promise of a rational, comfortable, modern life.

Everyday Water

Worldwide, the growing recognition of the scale of the current water crisis is shifting the focus of public dis courses from questions of production and supply (Big Water) to finding ways of reducing domestic water con sumption through strategies of 'demand management'. Alongside this shift has emerged a growing practical and theoretical interest in water's role in the activities of daily life. This is the terrain of Everyday Water, where values are rendered meaningful through daily practice, and where consumption patterns, habits, and routines are collectively acted out, maintained, and subject to change.

Recently, human and cultural geographers have drawn attention to the diverse kinds of relationships people have with water. This research has highlighted the wider cultural, social, and importantly, affective and emotional connections to water, and has emphasized the social spaces – from suburban homes and gardens to local communities – where the consumption of water occurs. This work situates water as part of a broader set of consumption practices associated with the space of the home, the garden, and suburban living. Much of this work concentrates on the kinds of socio cultural variables that influence water consumption. Related work has also given increased attention to the human dimensions of natural resource and environmental management, highlighting the importance of recognizing the variability of social cultural values and practices (including Indigenous val ues) in relation to complex 'social natures' such as water.

The detailed analysis of specific 'waterscapes' in re cent research shows the way in which natural and social geographies are always intertwined, and how social re lations (gender, household, class, etc.) interact with en vironmental resources, conditions, and uses. The vast systems of water engineering underpinning the modern city, for example, not only went hand in hand with the rise of new social spaces and social relations associated with water (the kitchen, bathroom, toilet, pipes and sewers, irrigated gardens, etc.), but also a new concept of domesticated water, tied to personal and urban (civic) cleanliness, hygiene, and sanitation. Today, access to, and the display of, water (in swimming pools, gardens, water features, etc.) functions more than ever as a symbol of cultural capital and distinction, and as a key indicator of wealth and social power. Recent work on southeastern Turkey, for example, demonstrates how new and emer gent irrigated waterscapes not only represent physical and symbolic signs of 'progress', but also produce new labor, gender, and social status geographies.

Everyday water consumption is not experienced as the use of 'X' number of liters of the anonymous resource 'H_2O'; rather, it is always already entangled with the consumer's expectations and habitual enjoyment of the services, technologies, and experiences water makes possible. This underscores the importance of investi gating the ordinary dimensions of those rituals of water use that have become routinized as 'inconspicuous con sumption'. This is the practical, nondiscursive realm of embodied water consumption, a set of actions, behaviors, and habits normalized through everyday use, which generally operate without cognitive thought and re flection. The social spaces involved with domestic water use and consumption are the sites not only where par ticular habits are learned, but where strategies of demand management are actively focused: low flow shower heads and reduced shower times, reduced toilet flushing, and the use of recycled water.

It is also important to draw attention to the sensory experiences of water, and its direct physical connection to the body. Many experiences of water involve physical contact, and are associated with feelings of bodily pleasure, immersion, and relaxation. Water is often as sociated with mythology, and with sacred and ritual uses, and cleanliness, purity, health, and naturalness are some of the more obvious symbolic meanings attached to it. This symbolic power of water has, for example, enabled the mineral water industry to become a successful global enterprise.

A focus on locally specific values and practices around water enables a multidimensional picture of everyday 'water cultures' as practical, lived experience, involving physical spaces, objects, and bodies. The importance of this approach has been recently recognized, and is re flected in much contemporary work that aims to produce situated knowledges about community attitudes to water and actual practices and patterns of water use, conser vation, and change.

Water Cultures/Natures/Networks

Because it is seen as the 'stuff' of social and biological life, water is assumed to be the most natural of all natural re sources. Yet how natural is water? Water may be present in, and produced from, nature, but it is also farmed and col lected, purified, and chemically treated, pumped and piped, transformed into power and electricity, used by households, agriculture and industry, bought and sold, manipulated for public display in water features like fountains and cascades, captured in swimming pools and reservoirs, consumed and eliminated, and turned into sewage which is pumped into the ocean to evaporate into the air, form into clouds and fall as rain, to start the process again.

Drawing on new theoretical perspectives such as actor network theory, researchers have begun to outline the way that human and nonhuman actors are entangled in diverse networks that, although not purely 'natural', are also 'more than human'. Within these networks both water, and the systems in which it exists, can be recognized as 'agents'; in other words, as having the capacity and agency to influence social life and inter actions, and to 'script' certain human uses and actions. Rather than seeing agency solely as the property of in tentional human actors, this approach recognizes that water flows and systems have already defined users and their roles in certain limited ways. Water, in this sense, plays very complex 'roles' across different environments, and has many varied interactions with different types of technological systems of delivery.

Water transcends, literally flows over, the division between the natural world and the human world – the nature/culture dichotomy that is such an established feature of Western (and much modern geographic) thought. Human engagement and interaction with water is one of the best examples of the way in which nature and culture are always part of each other, forming hybrid combinations of the human, the nonhuman, and the technological across a broad range of scales: from kit chens and bathrooms to agriculture and industry. As Swyngedouw explains, water is a kind of 'cyborg': "as it flows from spaces of production to spaces of con sumption, it undergoes changes in its physical, socio political and cultural character. When it reaches our faucet in the form of potable water it is neither a purely natural nor purely a human construction." Such new, interdisciplinary perspectives on water, combining in sights from both actor network theory and posthumanist philosophy, highlight 'human–nature relationality'. This relational view asserts that "phenomena do not have properties in themselves but only by virtue of the their relationships with other phenomena." Water, in essence then can be seen as a network of interwoven processes that are simultaneously human and nonhuman, natural and non natural.

See also: Ecology; Environmental Justice; Gardens and Gardening; Natural Resources; Nature-Culture; Public Good; Sustainability; Water Management.

Further Reading

Allon, F. (2006). *Special Issue on Water Resources & Discourses: Dams, Plants, Pipes and Flows: From Big Water to Everyday Water. Reconstruction*. 6.3.

Allon, F. and Sofoulis, Z. (2006). Everyday water: Cultures in transition. *Australian Geographer* 37(1), 45 55.

Askew, L. E. and McGuirk, P. M. (2004). Watering the suburbs: Distinction, conformity and the suburban garden. *Australian Geographer* 35, 17 37.

Bakker, K. (2003). A political ecology of water privatisation. *Studies in Political Economy* 70, 35 58.

Bakker, K. (2003). *An Uncooperative Commodity: Privatizing Water in England and Wales*. Oxford: Oxford University Press.

Bakker, K. (2008). The ambiguity of community: Debating alternatives to private sector provision of urban water supply. *Water Alternatives* 1(2), 236 252.

Castree, N. (2005). *Nature*. London: Routledge.

Gibbs, L. (2006). Special Issue: Applied Natures Cultural Engagements with Australian Environmental Management: Valuing Water: Variability and the Lake Eyre Basin, Central Australia. *Australian Geographer* 37(1), 73 86.
Harris, L. M. (2006). Irrigation, gender, and social geographies of the changing waterscapes of southeastern Anatolia. *Environment and Planning D: Society and Space* 24, 187 213.
Haughton, G. (1999). Analysing the multiple constructions of drought. *Transactions of the Institute of British Geographers* 24(3), 373 378.
International Water Management Institute (2006). Comprehensive assessment of water management in agriculture. www.iwmi.cgiar.org
Kaika, M. (2005). *City of Flows: Modernity, Nature, and the City*. London: Routledge.
Latour, B. (1993). *We Have Never Been Modern*. London: Harvester Wheatsheaf.
Laurie, N. and Marvin, S. (1999). Globalisation, neoliberalism, and negotiated development in the Andes: Water projects and regional identity in Cochabamba, Bolivia. *Environment and Planning A* 31(8), 1401 1415.
McCully, P. (2001). *Silenced Rivers: The Ecology and Politics of Large Dams*. London: Zed Books. New York: St.Martin's Press
Molle, F., Mollinga, P. P. and Meinzen Dick, R. (2008). Water, politics and development: Introducing water alternatives. *Water Alternatives* 1(1), 1 6.
Parr, J. (2005). Local water diversely known: Walkerton Ontario, 2000 and after. *Environment and Planning D: Society and Space* 23, 251 271.
Shove, E. (2003). *Comfort, Cleanliness and Convenience: The Social Organization of Normality*. Oxford: Berg.
Sofoulis, Z. and Allon, F. (2005). *Everyday Water: Values, Practices, Interactions*. University of Western Sydney: Centre for Cultural Research.
Sofoulis, Z. (2005). Big Water, Everyday Water: A socio technical perspective. *Continuum: Journal of Media and Culture* 19(4), 445 463.
Strang, V. (2004). *The Meaning of Water*. Oxford: Berg.
Swyngedouw, E. and Kaika, M. (2002). Urban Water: A political ecology perspective. *Built Environment* 28(2), 124 137.
Swyngedouw, E. (2004). *Social Power and the Urbanization of Water*. Oxford: Oxford University Press.
Toussaint, S., Sullivan, P. and Yu, S. (2005). Water ways in Aboriginal Australia: An interconnected analysis. *Anthropological Forum* 15(1), 61 74.

Relevant Websites

http://www.dams.org/
The World Commission on Dams.
http://www.water alternatives.org
Water Alternatives: An Interdisciplinary Journal on Water, Politics and Development.
http://www.worldwaterwars.com/
Water Wars Worldwide.

Water Management

D. Kay, Aberystwyth University, Aberystwyth, UK

Glossary

Bathing Water Directive (BWD) Standards for bathing water and a management system set by the European Union.

Biochemical Oxygen Demand (BOD) The oxygen in $mg\,l^{-1}$ removed from a water sample stored at 20 °C for 5 days in a dark incubator.

Clean Water Act (CWA) North American legislation covering water and wastewater management (see EU Water Framework Directive for the EU equivalent).

Coliform A group of generally harmless gut bacteria used to indicate fecal contamination of the environment, for example, *Escherichia coli*.

Enterococci A group of generally harmless gut bacteria used to indicate fecal contamination of the environment, for example, intestinal enterococci.

European Union (EU) A community of 24 states with common economic and environmental policies as defined in directives.

Fecal Indicator Organism (FIO) Microbial species which indicate fecal contamination (see coliform).

Food and Agriculture Organization (FAO) A United Nations agency devoted to enhancing food availability worldwide through improving the agricultural system.

Hazard Assessment Critical Control Point (HACCP) A risk identification and management system designed to provide early warning of potential problems and facilitate appropriate interventions to prevent health risks to consumers.

Headfall The vertical distance a riverbed changes over its length which provides an indication of its resource potential for power generation by HEP.

Hydroelectric Power (HEP) Power generation facilities using river water to produce electricity.

No Observed Adverse Effects Level (NOAEL) The highest known value of a level of a chemical (generally expressed in $mg\,kg^{-1}$ body) which produced no measurable adverse effects in a test animal.

Program of Measures (POM) A strategy designed to ensure that 'water bodies' defined under the EU WFD achieve 'good' water quality or ecological status.

Suspended Solids (SS) Suspended particulate material in water, generally greater than 0.45 microns in size.

Total Maximum Daily Loads (TMDL) The maximum flux of pollution which would not cause a water quality standard to be exceeded.

United Nations Educational Scientific and Cultural Organisation (UNESCO) A United Nations agency devoted to science and education.

Water Framework Directive (WFD) EU legislation designed to underpin water and wastewater management at the catchment scale (see Clean Water Act for the US equivalent).

World Health Organization (WHO) A United Nations agency devoted to improvement in human health at the global scale.

Global Freshwater Quantity

Approximately 2% of the global water cycle is freshwater and 86% of this resource is effectively inaccessible as ice caps and glaciers which have very long recycling times. The principal 'resource' available to society is found in rivers, lakes, and shallow groundwater which makes up approximately 0.3% of the total global water volume Thus, a tiny proportion of the global water store can be considered as a usable resource available to society.

The United Nations Food and Agriculture Orga nization (FAO) compiled the most recent global estimate of water resource availability by country and region. They calculate a global annual water resource availability of 43 750 $km^3\,yr^{-1}$. It should be noted that such assess ments are far from precise because the data sources used are not of uniform spatial coverage and quality. Indeed, many governments consider country specific information to be strategically important and, hence, confidential.

Notwithstanding these problems, the global assessment uncovers a pattern of large differences in water availability per person ranging from 10 $m^3\,yr^{-1}$ in Kuwait to 100 000 $m^3\,yr^{-1}$ in Canada (**Figure 1**). The global average is 7243 $m^3\,yr^{-1}$ which can be split into America 24 000 $m^3\,yr^{-1}$, Europe 9300 $m^3\,yr^{-1}$, Africa 5000 $m^3\,yr^{-1}$, and Asia 3400 $m^3\,yr^{-1}$.

Some 33 countries worldwide depend on their neighbors for over 50% of their renewable water resources (termed their 'dependency ratio') and this metric provides one, albeit crude, index of water conflict potential. UNESCO has specifically addressed these complex interactions of societal and sociodemo graphic drivers which impact on conflict initiation and resolution.

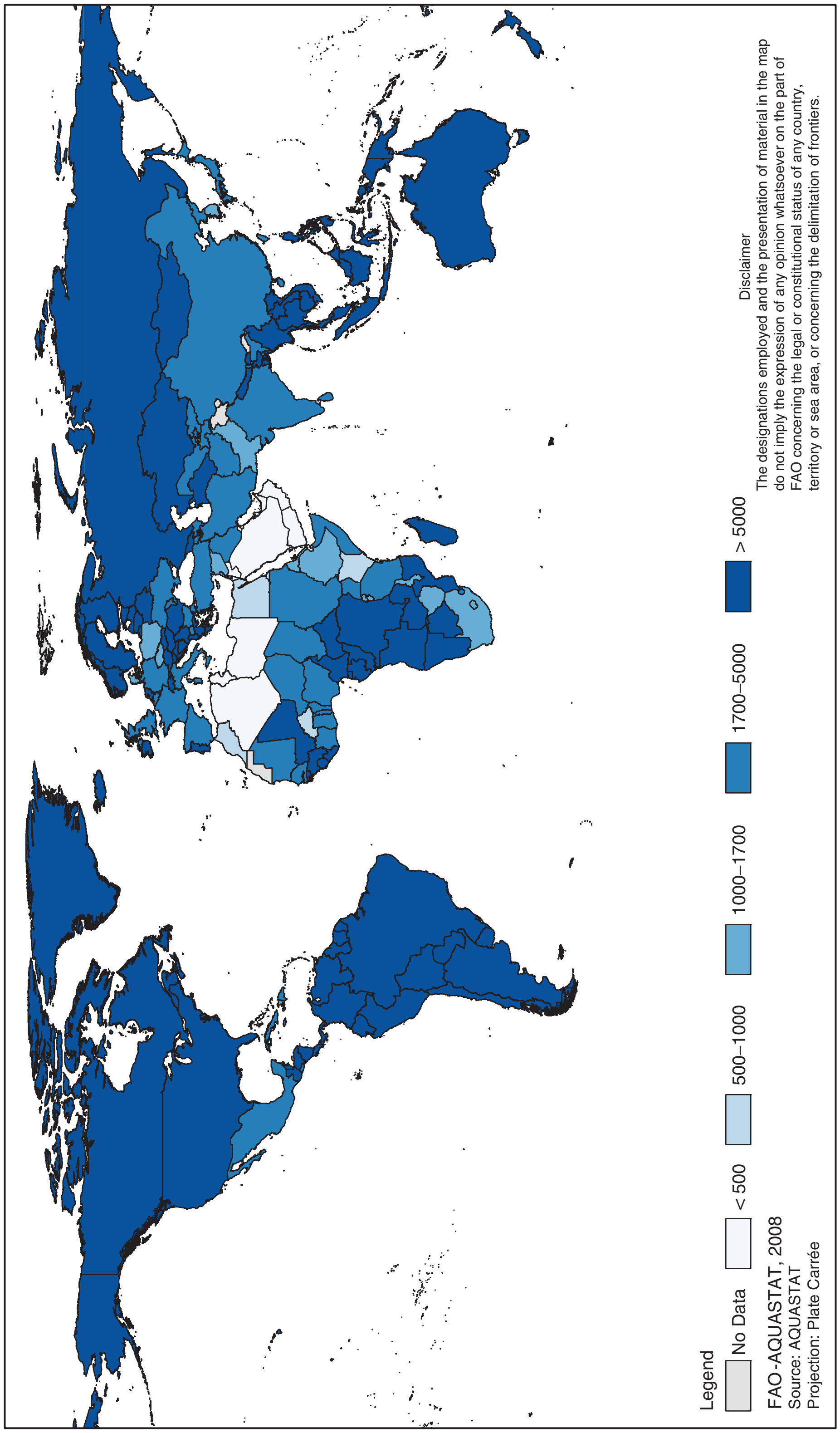

Figure 1 Tota actua renewab e water resources per nhab tant (m^3/yr^{-1}) (Actua renewab e surface water and groundwater resources per nhab tant (n 2005)).

Transboundary Resource Management

Water Wars

A high dependency ratio is seen in many areas of unresolved military conflict. The Middle East has, perhaps, been the most studied region in this regard. Here, the first action of the early Fateh organization was to attack a major water resource diversion project under construction by Israel. This involved a new pumping station designed to take water from the western shore of Lake Galilee and supply it through a system known as the 'National Water Carrier' to the Negev Desert. Darwish provides insight into the importance of water as a source of conflict in this region using the words of General Ariel Sharon, who later became the Israeli defence minister and stated:

> People generally regard 5 June 1967 as the day the six day war began. That is the official date. But, in reality, it started two and a half years earlier, on the day Israel decided to act against the diversion of the Jordan. (Darwish, 1994)

In effect, Sharon was attributing a major cause of the 6 day war to the 1964 Arab Summit in Amman which decided to divert the Jordan and reduce Israel's access to water.

Equally contentious upstream diversion is seen in Turkey's impoundment of the Euphrates by closure of the Ataturk Dam in 1990 which, it has been suggested, was designed to demonstrate Turkish power over Syrian water resources and, thus, produce a more pliant downstream neighbor less inclined to offer support to Kurdish fighters resident in Southeastern Anatolia. The immediate effect was to unify the downstream states of Syria and Iraq which drew up joint military plans to strike at Turkey which resulted in a reestablished flow in the Euphrates after 3 weeks. However, Turkish plans to utilize both the Euphrates and Tigris headwaters remain and conflict over this transboundary resource has been one of the drivers of Syrian historical support for the fighters in the Kurdish Workers Party (PKK) who consider the main areas of Turkish water resource development, known as the GAP project, as their homeland.

International Cooperation in Transboundary Resource Allocation

Perhaps the best examples of transnational cooperation is seen where nations of similar socioeconomic structure negotiate integrated use of transboundary resources. The Columbia River, which discharges to the US Pacific coast, has only 13% of its hydrological catchment draining Canadian territory. The US portion of the river was highly regulated during the early years of the twentieth century when 28 major projects and 11 200 km of irrigation feeder canals were constructed during the Theodore Roosevelt administration. Later, in the 1930s, some 30 US impoundments were constructed for irrigation and hydroelectric power generation. This intensive impoundment activity on the US reaches of the Columbia Basin, utilized some 238 m of the available 537 m of headfall south of the Canadian border producing 9.9 km^3 of US storage. This also reduced the risk of downstream flooding which had major impacts in the town of Dalles in 1894 and 1948 when discharges of 35 118 $m^3 s^{-1}$ and 28 604 $m^3 s^{-1}$ had been experienced respectively. The channel capacity at this site was 22 656 $m^3 s^{-1}$. It was estimated that an additional 13.6 km^3 of storage would be needed further to reduce a flood of 1894 proportions to manageable levels.

Attention was therefore focused on the Canadian Columbia to the north which generated some 30% of the total flow with a headfall of some 415 m over a river distance of 772 km. The headfall and impoundment sites made the Canadian catchment an excellent candidate for major development. Yet, despite the cultural integration of these two close neighbors and firm allies, achieving a treaty to effect this resource development took some 20 years of, often acrimonious, negotiation within the International Joint Commission (IJC), resulting in the Columbia River Treaty which was ratified in 1964. Key elements of this treaty were agreements on sharing of downstream benefits resulting from enhanced US hydroelectric power (HEP) generation and experienced by newly flood protected US communities which resulted from flow control structures in Canada. Discussion on the long term security and allocation of these benefits within and between provincial and national administrations in Canada were a cause of political conflict and delay between the international treaty signing at the federal level in 1961 and its ratification in Ottawa on 16 September 1964.

The treaty required Canada to produce 19.12 km^3 of usable storage in three large impounded reservoirs behind Duncan, Arrow (now Hugh Keenleyside), and Mica dams. Canada was to receive 50% of the downstream power benefits and 50% of the downstream flood control benefits up to 2024. Associated ecosystems (and in particular concerns for migratory fisheries) were generally not considered high priorities at the time of treaty negotiation. However, mechanisms have been built into management of this system through annual agreement on 'detailed operating plans' (DOPs) which have increasingly incorporated conservation and ecosystem concerns as societal appreciation of these issues has progressed since 1964.

Transnational Management Structures

The need for strong transnational basin management agencies, with executive responsibility and enforcement powers, has been recognized for many years. However, even the best practice example of the Columbia system did not exhibit this structure. Rather, the IJC facilitated the treaty negotiations. It was not, and is not, a basin management agency. If two neighbors such as the US and Canada cannot achieve such coordination and basin-wide management, there is little chance that nations with divergent political traditions and economies, particularly with a history of conflict, would be willing to release national resource control and management to an agency which might be less willing to deploy the political pressure derived from control of key resources. Thus, the ‘ideal’ of strong river basin management of transboundary resources remains an elusive goal.

Water Quality Considerations

Surface water systems have multiple uses throughout the globe. In addition to providing water for drinking, crop irrigation, and industry, rivers and lakes are used to transport anthropogenic pollution away from human populations. This pollution can enter water resource systems as ‘point’ discharges, generally via a discharge pipe from industrial and/or urban activities or as ‘diffuse’ pollution from many sources within the hydrological catchment area. This ‘effluent transport’ function of rivers, in particular, provides a major societal benefit, but potential hazard to downstream ecosystems and human health.

The presence of biological, chemical, thermal, and potentially radiological pollution of global water resources effectively reduces the utility of the available water volume and, in the developed nations at least, attention has focused on management of water ‘quality’ to maximize the ‘quantity’ of the available resource.

There are many measures of water quality which cover microbiological and chemical parameters in surface fresh waters. It is impossible to test for every potential contaminant in any water system and, to overcome this, key parameters are measured which suggest, or index, a level of risk to impacted ecosystems and/or users of the water resource for drinking, irrigation, and/or industrial purposes.

Microbiology

The use of indicator parameters is perhaps best exemplified by microbiological assessment. Here, the principal water quality constituents that cause problems to society are ‘pathogenic’ (i.e., disease causing) bacteria, viruses, and protozoa. Many of these, particularly the pathogenic viruses, cannot be measured in environmental waters and, even if they could be quantified accurately, proven absence of one pathogen could not be used to reassure the water consumer that other pathogens were not present. Thus, ‘indicator species’ are used which are generally harmless, cheaper to measure than the pathogens, and indicative of general contamination risk by ‘potential’ pathogen carriers. In effect, these sentinel species ‘indicate’ a route of contamination and thus a ‘potential’ risk; they do not, however, prove any ‘actual’ and ‘immediate’ risk to water consumers and/or bathers in recreational waters. The principal microbial indicator species which have been used, to date, are the coliform and enterococci bacteria which are excreted by all warm-blooded animals.

In this context, the water resource would be providing a disease transmission pathway to facilitate infection by what is known as the ‘fecal–oral’ route. This transmission pathway accounts for 1.8 million deaths per annum through diarrheal illness which mainly affects infants in the developing nations. World Health Organization (WHO) estimates that 88% of this disease burden is attributable to the interrelated causes of contaminated water and poor sanitation which, together, offer ideal conditions for ‘fecal–oral’ transmission. A great deal of attention has been focused recently on the individual contribution of specific exposures, such as recreational waters, and the potential ‘health gain’ of interventions such as the provision of sanitation systems, clean water supplies, or personal hygiene interventions such as hand washing. However, international, carefully controlled trials in this area are very difficult, if not impossible, and intervention studies reported, to date, have produced an evidence base which is, at best, ‘qualitative’.

Chemistry

Chemical parameters of water quality also exhibit a similar reliance on key indicators which have been developed to assess the quality of environmental waters. Early water management concerns in the industrialized nations focused on river water quality impacted by untreated sewage effluent. Domestic sewage principally comprises carbon in an organic form. When large quantities of this waste enter a surface water ecosystem such as a river, the ecological ‘carbon subsidy’ (i.e., introduced carbon that has not been produced by plants which use photosynthesis to fix carbon into plant material) causes an imbalance in the system’s oxygen status. In effect, ‘respiration’ by organisms using the carbon subsidy draws oxygen from the water. This can cause severe ecological disruption and kill fish, particularly, the most valuable ‘game’ species (e.g., salmon and trout) which need high oxygen saturation. However, the impact on the human population living by the impacted river is less subtle. Severe deoxygenation makes the river stink. This condition has been well documented for the River Thames in

London, UK, where the smells were so bad that the Houses of Parliament used to resort to hanging dis infected sheets against windows facing the river. At the time, it was believed that disease was transmitted by smells and vapors or 'miasma' and the cholera outbreaks in UK cities in the mid nineteenth century generated the political impetus to address this problem with major Victorian water resource developments in the UK uplands.

Early sewage engineers sought to improve the quality of riverine ecosystems by allowing the sewage solids to settle, then oxygenating the remaining liquid as it trick led over a matrix of stone or charcoal (to increase surface area). Alternatively, it was treated in tanks fitted with oxygen diffusers or paddles to agitate the liquid waste. A test was needed to assess the likely impact of the resultant effluent on river systems. The 'biochemical oxygen de mand' (BOD) test was developed and it gives a measure of how much of the river's dissolved oxygen would be 'demanded' by the effluent from the sewage treatment plant. In addition to the oxygen demand of the effluent, its clarity (or light transmission) also affected the esthetics and ecology of the riverine receiving water and, thus, the first chemical water quality standards combined BOD and the suspended solids (SS) as a measure of the quality of a treated effluent discharged to a river system. Because these measurements were used to characterize flows from sewage treatment plants, they have become known as the 'sanitary parameters' although, strictly speaking, neither has any public health, that is, 'sanitary' significance.

Discharge standards for sewage works effluents are mainly regulated worldwide through these two pa rameters. Commonly, a standard for riverine discharges of 20:30 is applied, which means a sewage plant must achieve a standard of $20\,\mathrm{mg\,l^{-1}}$ BOD and $30\,\mathrm{mg\,l^{-1}}$ SS which was first suggested by the UK Royal Commission on environmental pollution.

Water Quality Standards

Drinking water quality criteria are also suggested in international guidelines produced by WHO which are adapted for regional application in, for example, North America and Europe. These contain numerical standards for a wide range of chemical, esthetic, and micro biological parameters. The numerical water quality cri teria for individual parameters are, wherever possible, based on scientific evidence and designed to reduce the risk to health from adverse water quality to 'acceptable' levels. The aim for most chemical standards is to define a 'no observed adverse effects level' (NOAEL) in the ani mal/human test population which can be expressed in, for example, milligrams of chemical per kilogram animal body weight. This is then translated into a water quality standard by estimating the daily ingestion level in terms of water consumed and the acceptable concentration of the chemical in the treated water. Most standards have a considerable safety margin which is much wider for carcinogenic chemicals than for noncarcinogens. The standards design process will seek to protect susceptible segments of the population. One such group is bottle fed infants for whom drinking water is used to make up in fant formula producing a disproportionate water intake per unit body weight together with any chemical con tamination. Added to the fact that neonates may be more susceptible to potential toxins because of an immature metabolism, and rapid tissue growth, protection of this group often forms a central driver in the definition of water quality standards.

WHO designs these standards through extensive ex pert consultation using a process of temporary technical advisers who come together over a period of perhaps a decade and evaluate the scientific evidence available for each parameter before the resultant water quality criteria are published. The WHO technical advisers serve as independent scientists and have no political or com mercial affiliations. However, WHO guidelines have no legal weight. It is the responsibility of national authorities or supranational groupings, such as the EU, to select appropriate WHO guidelines and translate these into local and legally enforceable standards. In effect, the WHO provides an internationally accepted system of independent peer review of the scientific and public health evidence base.

The most recent application of this process is seen in the water quality standards outlined in the new EU Bathing Water Directive (BWD) published in 2006. These new microbiological standards are legally binding in member states of the EU and replace the earlier directive published by the Commission in 1976. The 1976 BWD set microbiological standards for bathing waters which have become the basis for Blue Flag and other seaside awards in Europe. A House of Lords Select Committee enquiry was informed in 1994 that the UK alone had invested over £10 000 million ($\sim$20 000 000 000 USD) to achieve the 1976 standards at coastal bathing sites. However, it is widely accepted that these early water quality criteria have no scientific evidence base, hence the need for a redesign leading to the 2006 BWD.

New Geographical 'Management' Perspectives

Water management has undergone a recent paradigm shift, driven in large part by the geographical community and its concepts. This has been guided by a simple but radical concept of holistic water and wastewater management focused on the drainage basin as the central management unit. This 'ecological' perspective is exemplified by the US Clean Water Act (CWA), 1977, and the EU Water Framework Directive (WFD), 2000. Both require a river basin management approach to achieving water quality standards at some defined point of use. This could, for example, be an abstraction point for drinking water, a bathing water location, or simply an ecosystem needing protection. The management agency is responsible for assessing and regulating all upstream pollution sources, from both point and diffuse sources, to ensure the water resource use is not compromised. This differs radically from past approaches where multiple upstream dischargers could be given a uniform 'consent' based on a defined effluent quality (see the 20:30 standard for BOD:SS outlined above) and an agreed effluent volume per day.

In the US, the management starting point of CWA is the identification of 'impaired' waters: that is, those which do not achieve water quality standards for a particular use. For each identified impaired waters, the regulatory agency is required to undertake a 'total maximum daily load' (TMDL) estimate which should quantify the sources and fluxes of specific pollutants causing the use location to be impaired. The TMDL study is then used to develop a catchment management plan with stakeholders designed to bring impaired waters into compliance.

In Europe, a directly parallel development can be seen in the WFD. This was developed when Sir Ken Collins (a geographer trained at Glasgow University, UK) chaired the EU Parliament Environment Committee and it is commonly termed in Brussels 'the UK Directive'. The WFD requires member states to examine all their drainage basins and, following a formal public participation stage with stakeholders, to define 'pressures and impacts' for all water bodies (i.e., rivers, lakes, groundwaters, and coastal seas). They are then required to develop a monitoring strategy and a 'program of measures' to ensure that appropriate water bodies achieve good ecological condition and water quality status as defined by daughter directives protecting drinking, bathing, and shellfish harvesting waters.

This policy agenda requires new and complex geographical understanding to underpin credible 'programs of measures'. First, the understanding of diffuse and point source pollution contributions from different sources is not well developed for parameters such as the fecal indicator bacteria impacting on bathing and shellfish harvestings waters. Second, the 'apportionment' of pollutant loadings between sources is pollutant specific; it is also highly dynamic and can change dramatically between low flow conditions and high flow events. Thus, knowledge of the timing of noncompliant events, when water quality deteriorates, may be crucial in determining the reasons and sources of the key pollutant loadings. Thus, the geographical dimensions of dynamic catchment systems and societal involvement in problem definition and resolution are central to this new agenda.

Perhaps the most radical paradigm shift in contemporary water management introduced in new WHO guidelines, covering drinking and recreational waters, and carried forward into related EU directives, is the requirement for predictive risk assessment to underpin real time risk management through public information dissemination. At its simplest level, this is a qualitative process which parallels tools developed in the food processing industry. Here, product testing is generally destructive and therefore has to be highly selective. Thus, other means have been developed to reduce risk to the consumer. 'Hazard assessment critical control point' (HACCP) analysis has been developed to address this problem and this involves examination of the total production process from raw material acquisition to delivery of a food product to the consumer plate. Risks of contamination are identified and procedures for their control are designed and implemented. In the water management arena, a very similar approach was first suggested by WHO in new guidelines for drinking water quality published in 2004. The approach is termed 'water safety planning' which charts the process of water delivery from the water gathering ground to the consumer tap, and identifies all risks to water supply quantity or quality including contamination hazards at each stage of the water collection, treatment, and supply system.

The 'water safety planning' approach, suggested by WHO, is technically demanding and may appear highly dependent on expert judgment and therefore seen as essentially technocentric. However, a central aim of the WHO in suggesting this approach was to provide a tool which developing nations could deploy where no facilities were available for any traditional water quality analysis. This predictive and qualitative risk assessment does require training but is being applied in developing nations following capacity building efforts guided by the WHO.

Predictive modeling of this type is not a new concept. It has been common in the US to post 'advisory' notices on beaches after significant rainfall events when water quality would be expected to deteriorate. What is new, and contentious, is the development of more complex, multivariate, and hydrodynamic models by the geographical community and the incorporation of this process into legislation such as the 2006 BWD which explicitly allows 15% of samples to be discounted where an appropriate predictive system exists. In the EU, Scotland is leading developments in this area with predictive modeling driving electronic signage systems which are used to inform the public in real time of adverse water quality. In conceptual terms, this can be considered a development of the HACCP principle taken to its logical conclusion.

Conclusions

Thus, water management across several sectors is moving to a new paradigm, largely driven by WHO initiatives, in which geographers are central to the new agendas of hazard assessment, catchment focused holistic manage ment, and real time modeling and prediction to underpin public health protection.

These developments are driven by relatively simple concepts of drainage basin hydrology, biogeochemical cycles, and disease prevention. The novel aspect of the new paradigm could be described as little more than 'common sense' in acknowledging that these elements are connected and interdependent. Water regulation, based on this paradigm, is now in force in North America and Europe where significant improvements in water quality, and resource allocation, should be possible in the me dium term.

There is however, a hugely more intractable challenge in management of water allocation and quality in trans boundary river basins. At worst, military conflict as a water dispute resolution mechanism is still in evidence and there is little sign, worldwide, of integrated basin wide management of water resources to ensure sufficient quantity and quality of supply for disadvantaged and politically weak nations and subpopulations. Thus, academic and agency effort is still devoted to quantifi cation of the size of the health burden from water and sanitation in developing nations long after such problems should have been resolved using nineteenth century technology. It is most often the societal and economic infrastructure that halts progress or causes regression to water scarcity, water related disease outbreaks, and/or prevents progress.

There is a significant and growing role for human geographers in water resources management. The prin cipal contributions are required in development and political science; that is, better to understand the need structures of communities where water supply and sani tation infrastructure is planned and also the international context of conflict avoidance and resolution which is essential in providing a stable framework for imple mentation. In developed nations, human geographers are playing a pivotal role in researching, facilitating, and promoting public and stakeholder participation in river basin planning. This is a legal obligation on the regulators who do not, generally, have a strong social science cul ture or academic background.

See also: Health Geography; River Basin Development; Water.

Further Reading

Ashok, S. (2004). *Managing Water Conflict: Asia, Africa, and the Middle East*, 234p. London: Routledge.

Blackstock, K. and Richards, C. (2007). Evaluating stakeholder involvement in river basin planning: A Scottish case study. *Water Policy* 9(5), 493 512.

Crowther, J., Kay, D. and Wyer, M. D. (2001). Relationships between microbial water quality and environmental conditions in coastal recreational water: The Fylde coast, UK. *Water Research* 35, 4029 4038.

Daoudy, M. (2008). A framework for negotiation in power asymmetry: Syria and Turkey's negotiations over the Euphrates and Tigris waters. *Paper Presented at the Annual Meeting of the ISA's 49th Annual Convention,* Bridging Multiple Divides. http://www.allacademic.com/meta/p252694 index.html (accessed Jun. 2008).

Darwish, A. (1994). Water wars. A lecture given to the Geneva conference on environment and quality of life. June 1994.

Fewtrell, L., Kaufmann, R. B., Kay, D., Enanoria, W., Haller, L. and Colford, J. M. Jr. (2005). Water, sanitation and hygiene interventions to reduce diarrhoea in developing countries: A systematic review and meta analysis. *Lancet Infectious Diseases* 5, 42 52.

Fisher, F. M. (2005). *Liquid Assets: An Economic Approach for Water Management and Conflict Resolution in the Middle East and Beyond*. Washington, DC: Resources for the Future.

Gray, N. F. (1997). *Water Technology,* 548pp. London: Arnold.

Hunter, P. R. (1997). *Waterborne Disease Epidemiology and Ecology,* 372pp. Chichester: Wiley.

Jones, J. A. A. (1997). *Global Hydrology,* 399pp. London: Longman.

Jonsson, A. (2005). Public participation in water resources management: Stakeholder voices on degree, scale, potential, and methods in future water management. *Ambio* 34(7), 495 500.

Kay, D., Bartram, J., Pruss, A. *et al*. (2004). Derivation of numerical values for the World Health Organization guidelines for recreational waters. *Water Research* 38, 1296 1304.

Kay, D., Stapleton, C. M., Wyer, M. D., McDonald, A. T. and Crowther, J. (2006). Total maximum daily loads (TMDL). The USEPA approach to managing faecal indicator fluxes to receiving waters: Lessons for UK environmental regulation? In Gairns, L., Crighton, C. & Jeffrey, B. E. (eds.) *Agriculture and the Environment VI; Managing Rural Diffuse Pollution. Proceedings of the SAC/SEPA Biennial Conference. Edinburgh International Water Association, Scottish Agricultural College, Scottish Environmental Protection Agency*, pp 23 33. Edinburgh: Scottish Agricultural College, Scottish Environmental Protection Agency.

Kliot, N. (1994). *Water Resources and Conflict in the Middle East,* 368pp. New York: Routledge.

McDonald, A. T. and Kay, D. (1986). *Water Resources Issues and Strategies,* 284pp. London: Longman.

Newson, M. (1997). *Land Water and Development, Sustainable Management of River Basin Systems*, 423pp. London: Routledge.

Orr, P., Colvin, J. and King, D. (2007). Involving stakeholders in integrated river basin planning in England and Wales. *Water Resources Management* 21(1), 331 349.

Pruss, A., Kay, D., Fewtrell, L. and Bartram, J. (2002). Estimating the burden of disease from water, sanitation and hygiene at a global level. *Environmental Health Perspectives* 110(5), 537 542.

Pruss, A., Kay, D., Fewtrell, L. and Bartram, J. (2002). Estimating the Burden of Disease from Water, Sanitation, and Hygiene at a Global Level. http://www.who.int/quantifying ehimpacts/global/en/ArticleEHP052002.pdf (accessed Jun. 2008).

Tamas, P. (2003). Water Resource Scarcity and Conflict: Review of Applicable Indicators and Systems of Reference. http://unesdoc.unesco.org/images/0013/001333/133307e.pdf (accessed Jun. 2008).

Wirsing, R. G. and Jasparro, C. (2007). River rivalry: Water disputes, resource insecurity and diplomatic deadlock in South Asia. *Water Policy* 9(3), 231 251.

Relevant Websites

http://www.allacademic.com
A Framework for Negotiation in Power Asymmetry: Syria and Turkey's Negotiations over the Euphrates and Tigris Waters.

http://www.fao.org
AQUASTAT Report on Global Water Resources, Food and Agriculture Organisation.

http://www.who.int
Burden of Disease and Cost effectiveness Estimates, WHO.

http://eur lex.europa.eu/LexUriServ/site/en/oj/2006/l 064/l 06420060304en00370051.pdf
Directive 2006/7/ec of the European Parliament and of the council.

http://eur lex.europa.eu/LexUriServ/site/en/oj/2000/l 327/l 32720001222en00010072.pdf
Directive 2000/60/ec of the European Parliament and of the council of 23 October 2000.

http://www.aber.ac.uk
Epibathe, Aberystwyth University.

http://www.fao.org
For an analysis of the data quality contained within the FAO Aquastat system, Food and Agriculture Organization.

http://www.fao.org
Global maps produced as part of the AQUASTAT initiative are shown, Food and Agriculture Organisation.

http://www.who.int
Guidelines for safe Recreational Waters Volume 1 Coastal and Fresh Waters, World Health Organization.

http://www.who.int
Health based Monitoring of Recreational Waters, World Health Organization.

http://www.epa.gov
US Clean Water Act, U S Environmental Protection Agency.

http://www.who.int
Water Recreation and Disease. Plausibility of Associated Infections: Acute Effects, Sequelae and Mortality, WHO.

http://www.who.int
WHO Guidelines for drinking water quality, World Health Organization.

Waterfront Development

M. Davidson, University of Western Sydney, Sydney, NSW, Australia

Glossary

Brownfield A piece of previously used land that requires some remediation of contaminants, pollutants, debris, or hazardous substances in order for it to be reused and redeveloped.

Containerization A shipping method which uses standardized containers to store and transport goods and materials.

Gentrification A complex process whereby a group of middle-class people move into a previously working-class community, causing substantial neighborhood reinvestment and the displacement of existing working-class residents.

Place Marketing The marketing activities which seek to influence a target audiences perceptions about a particular place in a positive way.

Postindustrial City A city that has undergone industrial decline and now has an employment profile that is characterized by the tertiary service sector.

Urban Development Corporation A quasi-public authority that is established by government in order to generate and direct urban development.

Waterfront A space where water (i.e., river, lake, sea, and ocean) meets with urbanized land, creating a unique spatial interface.

Introduction: A Historical Perspective

The earliest forms of waterfront development occurred as various societies began to utilize waterborne transit. This meant that up until 50 years ago, waterfront urban development largely occurred where physical relief was conducive to harborage and shelter. For example, the Roman settlement of Londinium – later to become London, UK – was established in AD 50 at the last bridgeable point on the river Thames. Subsequently, many more urban settlements have been established along favorable waterfront areas, such as Falmouth (UK), Bahia (Brazil), and Mumbai (India), where particular sections of riverside and coastline have provided ports and natural harbors suitable for maritime activities.

As international trade developed apace from the fourteenth century onward, waterfront cities witnessed significant growth as mercantile activities produced burgeoning urban economies. As the capital of the first imperial trading nation, Amsterdam in the Netherlands developed from a fishing village founded in the twelfth century to a city of 200 000 people in 1700. Other major waterfront trading cities established during this period included the Mediterranean cities of Naples, Venice, and Marseille. As European nations developed imperial trading routes, many other port cities developed. The British Empire's growth saw major ports established in Mumbai (India), Cape Town (South Africa), and Sydney (Australia). As the central point within this globalized trading network, London's port grew to stretch some 11 miles along the River Thames.

Waterfront space was also developed during this period for military and strategic reasons. Port cities were centers of economic and political power, and therefore required means of protection during a period of competitive imperial expansion and mercantilism. For example, to the east of London from the sixteenth century onward, a large stretch of the Thames riverside housed a naval dockyard and munitions storage facility. Originally located to protect the capital from invasion via the Thames estuary, the facility, which became Royal Arsenal, eventually grew to a 1200 acre site, employing 72 000 people during the early twentieth century.

Industrial Waterfronts

Although imperial trade and military expansion lie at the origins of much urban waterfront development, it is the industrial development which took place in waterfront areas during the nineteenth and twentieth centuries that has left the greatest legacy for contemporary urbanism.

The scaling up of production involved in industrialization had the commensurate effect of expanding city ports. As industrialization brought with it increased demand for raw materials and new streams of export products, port facilities had to be expanded. Nowhere was this more striking than in London, UK, where dock construction continued throughout the nineteenth and early twentieth century so that most of the eastern section of the metropolitan Thames was devoted to shipping and related mercantile activity. Starting with the construction of West India dock in 1802, and followed by London (1805), East India (1805), Surrey (1807), St. Katherine (1828), West India South (1829), Royal Victoria (1855), Millwall (1868), Royal Albert (1880), and King George V (1921) docks, the scale of development was vast and played a central role in enabling London's economy to become both prosperous and diverse.

Industrial production methods were also used in port and shipyard construction; building larger ships meant constructing larger dry docks in which to build them. In Britain, the rivers Thames (London), Mersey (Liverpool), Tyne (Newcastle), Wear (Sunderland), and Clyde (Glasgow) were all utilized to build large, ocean going ships. In Belfast, Northern Ireland, the Harland and Wolff shipyard was established in 1861 when Edward James Harland and Hamburg born Gustav Wilhelm Wolff founded a small shipyard on Queen's Island. As Belfast and the wider UK developed into the globe's largest industrial center, the Harland and Wolff shipyard developed apace, building larger and larger ships, in cluding the Titanic between 1909 and 1911. During this period, the Harland and Wolff shipyard employed 15 000 people working 49 hr weeks. Harland and Wolff also built the world's largest dry dock in Glasgow, Scotland – a facility which measured 840 ft long × 240 ft wide.

As processes of globalization and industrialization drove waterfront industry growth and dock construction, the supporting physical infrastructure which surrounded also had to be expanded. Dockland areas gradually be came larger and larger as vast railway terminals were constructed on waterfronts to transport incoming ship freight. In addition, huge warehouses were also built to store the vast amounts of goods that flowed into port cities. This resulted in many cities losing their water fronts as public spaces. Waterfront areas simply became the domain of heavy industry, rarely seen by most city residents. In Toronto, Canada, this port infrastructure dislocated the city from the Lake Ontario shorefront. Railways ran parallel to the shore, and in front of these port facilities hugged the waterfront. As a consequence, when the Toronto waterfront began to be targeted for redevelopment during the 1960s by city government, increasing accessibility to the waterfront quickly became a priority.

With the growth of port facilities and related infra structure, waterfront urban areas required large labor forces. As a consequence, residential communities reliant upon dock related employment grew in areas adjacent to the waterfront. These often developed distinctive polit ical and social cultures, since the demands and practices of port related labor were quite particular. For example, the casual work practices of dockers created distinct so cial networks and political practices. Quayside workers in cities such as London operated under a complex and informal employment system, whereby workers would be recruited on a needs basis every morning by foremen. Such an employment system meant that dockers often faced great levels of periodic hardship as recessions, tides, and weather combined to create insecure labor demand. This also affected social practices and networks, with it being imperative that dockers know their foreman to ensure their labor was recruited. Such an obvious and unmediated relationship between labor and capital also contributed to dock communities being places of radical politic ferment. Port communities in London and Liv erpool (UK), New York and Boston (US), and Brisbane (Australia) all became sites of political radicalism and trade union power.

In addition to shipping and port activities, the nine teenth and early twentieth century waterfronts were also appropriated by industries which relied upon water re sources. Such industries included textiles, dyes, cable making, breweries, various milling industries, steel, and energy. Throughout this period, many cities developed coal fired power stations along waterfront areas, because the interface of water and land provided an efficient means to transport the bulk materials required to gen erate power. Many of these power stations have left, and continue to leave, indelible marks on the urban land scape, both in terms of impressive built structures and contaminated land. Although some of these early twen tieth century power stations continue to operate using new energy production technologies, others have been converted to adaptive uses. Examples include the Bank side and Battersea Power Stations in London, both de signed by architect Sir Giles Gilbert Scott. The Bankside building is now the Tate Modern art gallery, while Bat tersea Power Station is undergoing redevelopment for conversion into residential and commercial usage.

Around the globe during the nineteenth and twentieth centuries, many waterfront cities, such as Toronto, Bos ton, Baltimore, Liverpool, Barcelona, Bilbao, Santiago, and Osaka, developed their waterfronts for the types of industrial purposes described above. However, as urban economies began to make a postindustrial transition, many of these once highly productive waterfront spaces went into decline and became redundant.

Industrial Decline

As industrialized nations such as the UK and USA underwent industrial decline in the second half of the twentieth century, many of their urban waterfronts fell into dereliction. The massive industrial and trading complexes that once dominated urban waterfronts be came landscapes of despair. As capital fled from urban waterfronts, the communities that were left behind often faced high unemployment, economic stagnation, and growing social problems.

Many of the ports which were built in the nineteenth and early twentieth century became unprofitable and eventually redundant in the late twentieth century as the shipping industry began to use containers to transport goods. Container shipping, a method developed by Malcolm McLean of the Pan Atlantic Steamship Cor poration, USA, greatly reduced the costs of handling goods at ports and simplified the logistics of shipping, drastically cutting the labor and time involved compared to the previously used bulk break method of shipping. The economic and social effects of the transition to container based shipping were drastic. Container shipping required large areas of land to store and organize shipping containers, large docks to hold container ships, and different labor practices. This meant that the many waterfront industries and communities that developed during the previous decades quickly became unprofitable and obsolete. This decline created a host of complex problems for the affected cities, many of which are still felt today.

As large dock facilities and surrounding communities fell into a decline, city and national governments were presented with the question of how to deal with the resulting problems. Initially, many governments attempted to reform and restructure port industries to make them more competitive. In London, the national government unified the city's docks and wharves under a single jurisdiction – the Port of London Authority (PLA) – in an attempt to remove internal wrangling and restructure complex operations. Yet, faced with an archaic labor system and restrictive legislation, the PLA ultimately abandoned its attempts to regenerate London's docks, and in 1969, it instigated the movement of these operations east to a new container port in Tilbury, Essex. This story of a failed attempt to regenerate urban docks was repeated elsewhere. In Toronto during the 1950s and 1960s, the city government responded to increased competition and changing shipping practices with a plan to expand port facilities, create a new outer harbor, and begin a process of redevelopment in established ports. Initial reactions to waterfront decline were therefore often focused upon restructuring existing industries and maintaining these spaces as sites of industry.

As various government bodies attempted to reverse this decline, a number of political conflicts arose. The containerization of shipping demanded that labor practices within docks changed. Working hours, employment contracts, places of work, and labor requirements were all being dramatically altered. As a consequence, the dockers unions were brought into conflict with managers and political representatives. In New York City during the 1960s, the New York Shipping Association introduced a computer based hiring system, developed by IBM, to reorganize dock labor. This system had the effect of rationalizing labor recruitment and working hours, and changed the organization of work units. The latter caused discontent within the dockers unions as it threatened to reduce labor demand and increase the risk of workplace accidents. This was repeated in other cities, such as Liverpool, UK, as difficult negotiations were undertaken in order to maintain port facilities. However, these did little to stem the processes of decline, which finally resulted in the complete abandonment of many urban waterfronts by port activities.

Economic decline and abandonment of city ports plunged the communities which relied upon dock related employment into severe distress. In London, between 1971 and 1981, the population in the docklands area fell from 48 352 to 39 429, due in large to out migration. By 1981, the unemployment rate in the area was running at 17.8%. However, even after a decade of decline, between 1978 and 1983, a further quarter of all jobs in the area (13 000) were lost and another 5100 acres of land fell into dereliction. The same processes of decline occurred in other port cities, with similar consequences for working class communities. In New York City, the piers of Manhattan and Brooklyn were abandoned for the container port in Newark/Elizabeth, New Jersey. While in Dublin, Ireland, all port activities transferred 1 mile down the River Liffey, from the central city to a container shipping facility at Alexander Quay, reducing labor demand and dislocating dockers from traditional sites of employment.

As industries abandoned urban waterfronts, much of the land fell into dereliction. Furthermore, as a consequence of the character of many dock and waterfront industries, such as gas stations, chemical plants and oil refining, large swathes of land were left highly contaminated. For example, the Union Carbide chemical plant on Rhodes Peninsula, Sydney, Australia, dumped dioxins and other chemical by products in surrounding wetlands up until 1970. As a consequence of such practices, many waterfront sites have required significant expenditure on remediation to make them fit for reuse.

The Return to the Waterfront

The Explosion of Waterfront (Re)Development

As waterfront cities began to develop postindustrial urban development strategies throughout the 1970s, 1980s, and 1990s, urban waterfronts became central to urban renewal and regeneration throughout the world. In Sydney, Australia, the once derelict docklands of Darling Harbour have been transformed into a showpiece for the city and a hub of tourism. Baltimore's Inner Harbor has been transformed from a symbol of industrial decline into one of postindustrial urbanism, with all of the related connotations of gentrification, spectacle, social polarization, and inequity. In Melbourne, Australia, the urban waterfront has been redeveloped to represent the city's globally orientated, economically focused, consumption based urban strategy. Hong Kong's waterfront has been redeveloped and substantially extended through a reclamation project that has involved building a new urban infrastructure, consisting of an airport, railways, motorways, tunnels, and bridges. In Toronto, the landscape of industrial and rail terminal facilities that once lined the shorefront of Lake Ontario has, and continues to be, reinvented. In London, the docklands have been remade into a postindustrial space where remnants of past activities now only exist in the form of postmodern pastiche, where quaysides support the balconies of luxury apartments and cranes exist as artifacts.

Urban Development Agendas and Waterfronts

Waterfront redevelopment has been at the forefront of some of the urban redevelopment schemes which have come to characterize postindustrial urban renewal, such as London Docklands and Battery Park City, New York City. Under postindustrial urban policy agendas that are focused upon consumption and the attraction of capital, waterside redevelopment schemes have followed similar patterns. Waterside office based redevelopment programs have aimed to attract a global clientele of transnational corporations to the city. The development of waterside leisure spaces has catered to the consumption desires of the professional middle classes. Finally, waterside residential development has focused upon providing residencies for the postindustrial metropolis' burgeoning ranks of wealthy professionals. While there have been exceptions to this trend, notably in Vancouver's False Creek South neighborhood where a liberal city administration constructed a waterfront neighborhood based upon principles of livability, the redevelopment of deindustrialized waterfronts has overwhelmingly reflected the neoliberal reinvention of urban policy and related issues of inequity and questionable political representation.

Case study: Battery Park City, New York City. The redevelopment project which transformed Lower Manhattan's Hudson River waterfront from a collection of derelict piers into a beacon of global capital has become a symbol of both postindustrial waterfront development and urban renewal. The redevelopment of this 92 acre slice of Manhattan began in 1968 when New York State established the Battery Park City Authority, an urban development corporation. The initial plans for this space proposed by the New York City Department of Marine and Aviation had wanted to redevelop the port facilities along this stretch of waterfront, creating an updated dock and warehouse facility for the island of Manhattan. However, under New York State Governor Nelson Rockefeller, a series of master plans were developed which envisaged this waterfront area becoming a pedestrianized park space, interspersed with both private and subsidized futuristic housing. The plan received a warm reception in the city with architecture critic Ada Louise Huxtable commenting in the *New York Times*: "Is this any way to plan a city? You bet it is."

New York City's 1976 financial crisis led to the plans of Nelson Rockefeller's government being abandoned for a much more private capital orientated strategy. This change to the Battery Park waterfront redevelopment scheme represents a much wider reorientation in urban policy. In this recent examination of neoliberalism, David Harvey traces the emergence of the current ascendancy of the New Right over urban (re)development, and more generally social policy, to New York City's financial crisis. Harvey argues that, at this point in time, global financial actors, such as Citibank chairman Walter Wriston, realized significant leverage and influence over urban and social policy, which resulted in them forcing a vast scaling back of the Keynesian welfare state apparatus. As the major urban development project in North America at this time, Battery Park City's subsequent neoliberal reorientation has therefore become symbolic of the more global shift in urban redevelopment practice which has occurred over the past 30 years.

The final 1979 master plan for Battery Park City, designed by Alexander Cooper and Stanton Eckstut, dropped the ambitious and socially inclusive plans of the early 1970s and adopted what the Battery Park City development corporation was to call 'a hard nosed, realistic approach'. This approach was physical redevelopment led, involved little local community participation, provided no affordable housing, and focused upon private market imperatives. In 1980, construction on the site began and continued throughout the decade. Eventually, the development included the World Financial Center office complex and a series of private residential towers. Although the area was heavily affected by the attacks of 11 September 2001, on the neighboring World Trade Center, development in the area continues today with a new luxury hotel and apartment complexes being built on the reclaimed land (**Figure 1**).

Why Is Postindustrial Waterfront Redevelopment Distinctive?

The unique dynamic created by the interface of water and urban space has greatly influenced the nature of waterfront redevelopment in the postindustrial era, just as the same interface dictated the usage of waterfronts for shipping and industrial purposes in previous decades. However, it is important to recognize that this type of renewal is not disconnected from general urban policy directions. The same themes of private sector led development, urban development corporations, circumvention of planning protocol, and lack of public accountability that characterizes neoliberal urban redevelopment also characterizes most waterside redevelopment schemes. With waterside redevelopment schemes reflecting general postindustrial urban redevelopment themes, it is therefore appropriate to ask whether waterside redevelopment should be considered as a distinctive form of urban development. In answer to this question, we can identify a number of important

Figure 1 An image showing the continued development of Battery Park City, New York City, 2006.

factors that distinguish waterfront development from general urban development. These include the political significance that is imbued upon waterside locations, the economic potential that is bound up in disinvested waterside spaces, and the sociocultural value that is associated with urban waterside locations.

Political significance

The distinct physical qualities of urban waterside spaces have meant that in recent decades, local political regimes have made them redevelopment priorities. Waterfront sites which had previously been used for port and industrial activities were often in central city locations. Therefore, as a result of decline, many cities found themselves with large areas of devalorized land in prime central city locations. Furthermore, due to their location, these sites are also highly visible, often close to the central business district, and able to be panoramically gazed upon from opposing riverbanks and adjacent foreshores. This has meant that waterfront sites have become important spaces within the context of place marketing redevelopment agendas, where highly visible demonstrations of postindustrial renewal and consumption are central to determining urban prosperity.

Place marketing has become a key tenet of local economic development in an era where cities are constantly hierarchically ranked by external agents. Cities must actively pursue and construct positive imaginative geographies to ensure that they become and remain 'hot spots'. A negative place image can prove unattractive to potential investment, since negative imagery can indicate the city's economic decline through falling consumption. The redevelopment of highly visible urban waterside sites has become a key mechanism by which positive place images are constructed. For example, in Singapore, the waterside was chosen for redevelopment specifically with the purpose of demonstrating and displaying the city state's global city ambitions. The redevelopment of waterside spaces for place marketing purposes has often resulted in city planning authorities and development corporations overriding democratic planning procedures and local interests in order to bring perceived citywide economic benefits.

Economic value

In addition to their significance in place marketing development agendas, brownfield waterside sites are also distinctive urban spaces because of their potential economic value. Waterside development can offer substantial premiums to developers, land owners, and local governments. The potential returns on investment can be as much as 40–60% higher for waterside residential units compared to the equivalent units without waterfront views. Releasing and redeveloping devalorized and derelict waterside land can therefore offer hefty windfall profits.

The return of capital to waterfront locations has therefore often been paralleled by a movement of people to the waterfront. Urban waterfronts which were once undesirable brownfields have now become some of the most desired pieces of real estate around the globe. Examples include waterfront areas of Sydney (Australia), Baltimore (USA), and Prague (Czech Republic). A significant consequence of this has been the widespread gentrification of urban waterfront space.

Case study: London Riverside. From almost any vista in London, the signs of the Thames' postindustrial transformation are obvious. A gaze toward East London encounters a dynamic landscape of continually growing skyscrapers huddling around the Thatcherite government driven 1980s Docklands tower of One Canada Square. In Central London, the new London government has established its Foster and Partners designed egg shaped offices on the banks of the Thames and the Tate Modern art gallery has attracted 4 million tourists to the Southbank each year since opening in 2000. In addition, all along London's waterfront, a vast swath of residential development has taken place since 2000, making the riverside one of the most desired residential spaces in the city.

London's newly (re)established metropolitan government, the Greater London Authority, has embraced the waterfront's reinvention by developing a strategic planning vision for it: The Blue Ribbon Network. This planning policy is intended to stimulate the redevelopment of what the metropolitan government considers areas of prime and underutilized industrial, residential, and transit resources. While this policy adheres to a wider neoliberal politics by not directly intervening in reducing poverty and alienation, it has found utility in the riverside as a social policy resource. It is seen as a place where various social groups will want to congregate and mix, generating cross class and cross cultural social interactions which will in turn, it is thought, help reduce the city's social problems. The Blue Ribbon Network vision therefore provides an example of how waterfront spaces have become politically valued and how certain sociocultural values are associated with them. In order to achieve its vision, the metropolitan government has encouraged the already accelerating rates of waterfront redevelopment.

Major corporate residential property developers have continued to answer the calls of the metropolitan government as they have rushed to build luxury apartments along the Thames (**Figure 2**). Developers have been met by substantial demand for their products as London's professional elites have proven eager to become riverside

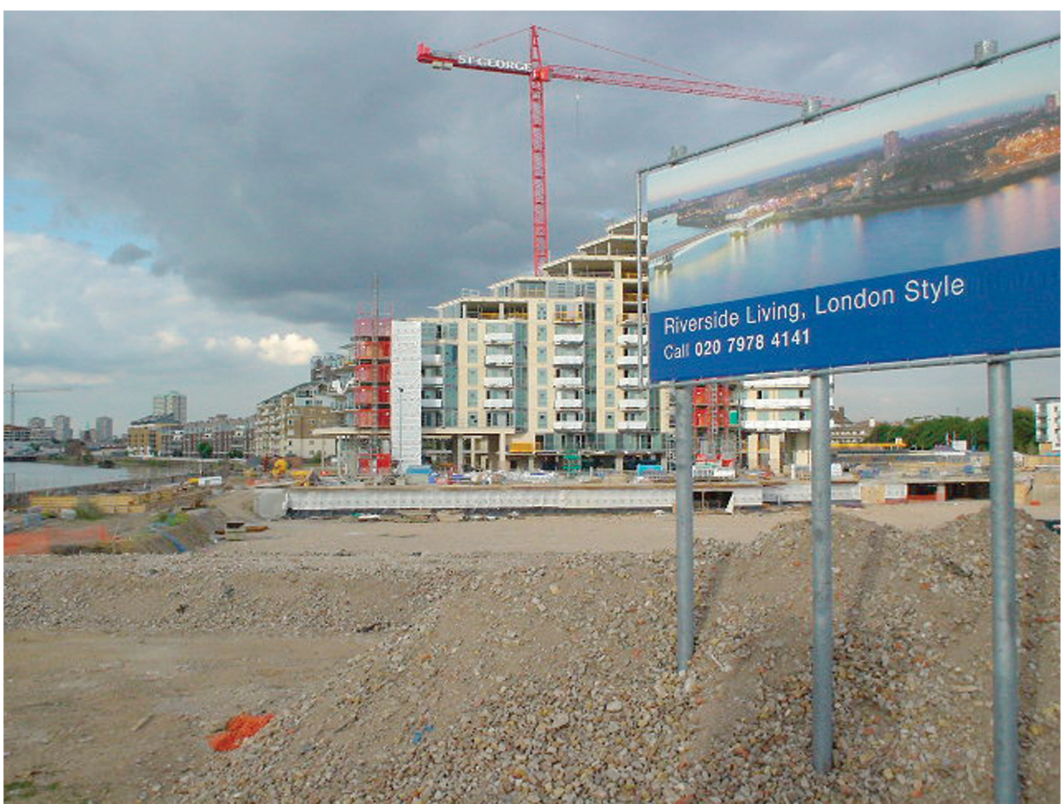

Figure 2 Ongoing residential waterfront development.

residents. The collection of political promotion, large scale corporate development, and burgeoning demand for 'riverside living' has led to a widespread gentrification of the waterfront. Although processes of neighborhood change along the riverside are still in their early stages, those working class waterfront communities which have received a large influx of middle class people now living in newly built luxurious riverside apartments appear to be witnessing significant gentrifying, and as a result dis placing, changes in their neighborhoods.

Sociocultural value

Although it is only implicitly referred to in many ac counts of urban waterside redevelopment, it is clear that much of the political and economic significance of these spaces is related to the sociocultural value placed upon them. Waterside spaces are often part of the city's valued collection of open spaces, comparable to the parks and gardens which are valued for their sensory and physical qualities. Many of the debates which surround issues of public access to redeveloped waterfronts bear witness to the perceived communal value of urban waterside space.

In an era of neoliberal urban development, the spatial and sensory qualities of waterfront areas have been utilized in order to generate consumption. In Melbourne, Australia, the city's Southbank waterfront region has been extensively redeveloped into a space of (middle class) consumption. Here, carefully managed urban design techniques and omnipresent policing controls manage this waterfront space filled with manufactured spectacles. The same commodification of waterfront space can be witnessed in other cities, such as Copenhagen (Den mark), Sydney (Australia), and Boston (USA), where parks and promenades have been provided alongside expensive hotels and retail stores in an effort to produce an attractive and consumption efficient urban space.

See also: Gentrification; Local Economic Development; Neoliberalism; Postmodern City.

Further Reading

Breen, A. and Rigby, D. (1996). *The New Waterfront: A Worldwide Urban Success Story*. London: Thames and Hudson.

Chang, T., Huang, S. and Savage, V. (2004). Globalization and urbanization in Singapore. *Urban Geography* 25(5), 413 436.

Davidson, M. and Lees, L. (2005). New build 'gentrification' and London's riverside renaissance. *Environment and Planning A* 37(7), 1165 1190.

Dovey, K. (2005). *Fluid City: Transforming Melbourne's Urban Waterfront*. London: Routledge.

Fainstein, S. (1994). *City Builders: Property, Politics, and Planning in London and New York*. Oxford: Blackwell.

Gordon, D. (1997). *Battery Park City: Politics and Planning on the New York Waterfront*. New York: Routledge.

Goss, J. (1996). Disquiet on the waterfront: Reflections on nostalgia and Utopia in the urban archetypes of festival marketplaces. *Urban Geography* 17(3), 221 247.

Harvey, D. (2005). *A Brief History of Neoliberalism*. Oxford: Oxford University Press.

Hoyle, B., Pinder, D. and Husain, M. (eds.) (1988). *Revitalising the Waterfront: International Dimensions of Dockland Redevelopment*. New York: Belhavan Press.

Jauhianen, J. (1995). Waterfront redevelopment and urban policy: The case of Barcelona, Cardiff and Genoa. *European Planning Studies* 3(1), 3 23.

Ley, D. (1980). Liberal ideology and the postindustrial city. *Annals of the Association of American Geographers* 70(2), 238 258.

Marshall, R. (2001). *Waterfronts in Postindustrial Cities*. London: Spon Press.

Stevens, Q. and Dovey, D. (2004). Appropriating the spectacle: Play and politics in a leisure landscape. *Journal of Urban Design* 9(3), 351 365.

Waitt, G. (2004). Pyrmont Ultimo: The newest chic quarter of Sydney. In Bell, D. & Jayne, M. (eds.) *City of Quarters: Urban Villages in the Contemporary City*, pp 15 36. Aldershot: Ashgate.

Relevant Websites

http://www.bm30.es/socios/instituciones/bilpr uk.html
A site documenting waterfront development in Bilbao, Spain.

http://www.batteryparkcity.org
Battery Park City Authority, New York City.

http://www.concordpacific.com/
Concord Pacific, major waterfront developers in Vancouver.

http://www.hudsonriverpark.org
Governing body for Manhattan's Hudson Riverside redevelopment.

http://www.lddc history.org.uk/
History of the London Docklands Development Corporation.

http://www.tate.org.uk/modern/
London's Tate Modern, an example of waterfront renewal.

http://www.museumindocklands.org.uk
Museum in Docklands, working class history in London's Docklands.

http://www.barangaroo.com/
Sydney's latest waterfront redevelopment site, Barangaroo.

http://www.london.gov.uk/thelondonplan/blue ribbon/
The Greater London Authority's Blue Ribbon Network.

http://www.docklands.com/
Victoria's Melbourne Docklands redevelopment website.

Watts, M. J.

S. Batterbury, University of Melbourne, Melbourne, VIC, Australia

Michael J Watts (**Figure 1**) has made contributions to scholarship, and to the discipline, across a broad canvas. His work contributes to theory, while drawing on the arguments of a wide range of intellectuals and activists. It transcends scale, and it links historical events. Although strongly identified with a Marxist tradition and with political economy, his work since the 1970s has extended from the human ecology of agrarian societies in Africa, to the theorization of culture and difference, and the pros ecution of war under contemporary capitalism.

Watts began his geographical interests when growing up in a small village between Bath and Bristol in the UK in a rural environment that fed his lifelong interest in agrarian landscapes, the commons, and their occupants. He excelled in school and later at University College, London, where he also dabbled in the student politics of the late 1960s and early 1970s. His masters and PhD at Michigan were supervised by Barney Neitschmann. From 1979 he taught geography and development studies at University of California, Berkeley, becoming the 'class of 1963' professor.

Watts' defining work was on the political economy of food security, famine, and agrarian labor in northern Nigeria, published as *Silent Violence* in 1983. Household level fieldwork was combined with extensive historical and archival research. The central argument was that the Hausa speaking rural peasantry north of Kano were historically disadvantaged not by drought, resource scarcity, or population pressures, but more importantly by the social convulsions wrought by the extractive British colonial regime. Watts was one of the first to offer a sustained social and historical explanation for food insecurity and famines and, with Amartya Sen's work, this has led to considerable debate among scholars and pol icymakers. Watts described a rural peasantry struggling against oppression, hierarchies of power, and disadvan tageous markets, such that reciprocal networks that for merly operated during times of food shortage were eroded under colonial rule. His theoretical challenges to the ahistorical reasoning offered by cultural ecologists and natural hazards researchers (which labeled peasants as 'adaptive') were prescient, but the tone of his chal lenges was too severe for some. For example, Michael Mortimore, who worked in northern Nigeria for 27 years, remains less critical of the emergence of contemporary market relations and the prospects for continued farmer innovation in the region, and has persisted with studies of adaptation.

Figure 1 Michael Watts in Nigeria.

The agrarian theme has continued in less fieldwork intensive studies carried out in other locales. Since the 1980s these have included studies of gender relations in Gambian rice farming (with Judith Carney), the workings of contract farming systems and the rise of agribusiness (with Peter Little and David Goodman, respectively), the commercial transformation of chicken farms in the US, de collectivized agriculture in Vietnam, power relations in agricultural biotechnology, and land reform and agrarian strategy in Kerala, India. He has also worked on *West of Eden*, a project on the legacy of the 1960s coun tercultural movements and communes in northern California (with Iain Boal).

Watts' interest in rural landscapes and the spatial expressions of capitalist practices has led to new 'geog raphies of commodities' via an exploration of the conflicts these generate. Watts has examined processes of oil ex traction and the rapid speed of change, the massive wealth, ethnic animosities, corruption, and protest re sulting from oil's value. In northern Nigeria, he traced antimodernist Islamic movements to discontent among poor migrants to Kano and their perception that 'bad' oil money and state revenues and services were unequally apportioned. His studies of protest in Ogoni oil producing territories on the Niger Delta have de tailed the history of extreme political marginalization and local conflict with oil corporations under 'petro capitalism'. Oil, he argues, with its possibilities for global profit and thus control and corruption permeates politics, economies, and images. *Violent Environments* compares several examples of resource driven conflicts, rejecting Malthusian claims of scarcity driven violence and argu ing instead that the new politics of identity and ethnic conflict is rooted in local histories and social relations yet connected to wider processes of material transformation.

Power, for Watts, is also imbued with important symbolic meanings, as are new forms of environmental and political 'control' and governance. In the 1990s his collaborations with Alan Pred extended a 'material' cri tique of capitalism to the struggles for meaning, locating a range of historical examples to particular landscapes and struggles. Recent collective work by Retort (a San Francisco Bay area collective of activists and scholars) links a situationist concern with spectacle in modern society to a hard hitting critique of the war in Iraq and US imperialism in the 'poisonous epoch we are living through'. Retort identify the naivety of US attacks on the Middle East and argue that "The state has behaved like a maddened beast…" Military reprisals were motivated by the attack on the World Trade Center in 2001, itself set in a context of symbolic and material domination of everyday life by neoliberal policies. These have had major geopolitical implications, occluding many social relationships and communal bonds. Neoliberalism has, they note, economically marginalized the very regions from which terrorism is now emanating.

Through these and other studies, Watts has contrib uted to the interdisciplinary fields of political ecology and development studies, making defined contributions to each and working with a range of collaborators over three decades. For example, *Liberation Ecologies*, a project begun with Richard Peet, defines a particular form of theoretically informed political ecology that traces the environmental and social impacts of contests over natural resources, through ideas and discourses about those re sources, and it appraises the struggles to secure liveli hoods that persist under neoliberalism, largely in developing countries.

Watts has said of his commitment to political econ omy that "history does matter, and I can think of no better way of preparing for the future than, in part, by looking backward." Yet his commitment to geography is strong. It includes numerous introductory and peda gogical texts, including the *Dictionary of Human Geography*, a lifetime of teaching, and supervision of over 75 PhD students and postdoctoral scholars.

See also: Africa; Marxism/Marxist Geography I; Marxism/Marxist Geography II; Peasant Agriculture; Political Ecology; Postcolonialism/Postcolonial Geographies; Pred, A.; Radical Geography.

Further Reading

Boal, I., Clark, T. J., Matthews, J. and Watts, M. J. (2005). *Afflicted Powers: Capital and Spectacle in a New Age of War*. London: Verso.
Goodman, D. and Watts, M. J. (eds.) (1997). *Globalizing Food: Agrarian Questions and Global Restructuring*. London: Routledge.
Johnston, R. J., Pratt, G., Gregory, D. and Watts, M. J. (eds.) (2000). *Dictionary of Human Geography* (4th edn.). Oxford: Blackwell.
Little, P. D. and Watts, M. J. (eds.) (1994). *Living Under Contract: Contract Farming and Agrarian Transformation in Africa*. Madison, WI: University of Wisconsin Press.
Mortimore, M. J. (1989). *Adapting to Drought: Farmers, Famines and Desertification in West Africa*. Cambridge: Cambridge University Press.
Peet, R. and Watts, M. J. (eds.) (2004). *Liberation Ecologies: Environment, Development, Social Movements, 2nd edn*. London: Routledge.
Peluso, N. and Watts, M. J. (eds.) (2001). *Violent Environments*. Ithaca, NY: Cornell University Press.
Perrault, T. (2004). Michael Watts. In Hubbard, P., Kitchin, R. & Valentine, G. (eds.) *Key Thinkers on Space and Place*, pp 323 329. London: Sage.
Pred, A. and Watts, M. J. (1992). *Reworking Modernity: Capitalisms and Symbolic Discontent*. New Brunswick, NJ: Rutgers University Press.
Watts, M. J. (1983). *Silent Violence: Food, Famine and Peasantry in Northern Nigeria*. Berkeley, CA: University of California Press.
Watts, M. J. (2005). Righteous oil? Human rights, the oil complex and corporate social responsibility. *Annual Review of Environment and Resources* 30, 373 407.

Relevant Websites

http://globetrotter.berkeley.edu
Institute of International Studies, University of California, Berkeley: Harry Kreisler interviews Michael Watts, 2001.

Welfare Geography

R. Lee, Queen Mary, University of London, London, UK
C. Philo, University of Glasgow, Glasgow, UK

Glossary

Social Injustice The existence in most societies of 'winners' and 'losers' as a result of the operation of social, economic, and political relationships over which they have little, if any, substantive control. It involves the distinction within societies between those able to wield power and acquire wealth, status, and well-being and those who are not. Although it is possible to conceptualize degrees of social injustice in abstract terms, geographical and historical responses to it reflect a politics of awareness and responsiveness to inequalities in material conditions which may engender profound political and social disruption in the attempt to create a more 'just' society.

Spatial Inequality Spatial variations in and proximity to social 'goods' (such as access to knowledge, employment, income, housing, education, or healthcare) and 'bads' (such as pollution, deprivation, political repression, or lack of knowledge) such that people in different places experience different levels of 'goods' and 'bads'. Spatial inequality is bound up with, and engendered in part by, social injustice.

Territorial Social Indicators Data on socioeconomic variables (e.g., to do with employment, income, physical and mental health, access to services, etc.) that are spatially referenced to particular territorial units (places, areas, and regions), perhaps at different spatial scales, and whose measurement can reveal significant spatial inequality.

Welfare Geography An approach to the study of human geography that centralizes the welfare of human beings, taking seriously spatial inequality in well-being and life chances occasioned by the simple fact of where people live and work relative to one another; this approach describes, purports to explain, and offers normative evaluations of this spatial inequality.

Welfare Geography: Origins

As elaborated at length in other articles, the 1960s in Britain, North America, and to an extent elsewhere witnessed a new era of (human) geography as spatial science or locational analysis: a highly 'technical' version of the discipline, predicated on quantitative methods, formal modeling (often based on neoclassical economics) and the search for supposedly fundamental spatial laws of human behavior and organization (with a loosely positivist cast). By the close of the 1960s, however, and again elaborated in other articles, there was growing resistance to this form of human geography, one aspect of which was an increasing clamor to take seriously all manner of 'social' and 'political–economic' issues and problems rarely if ever considered by spatial scientists. It can be argued that this clamor eventually led to the widespread adoption of a 'radical geography' perspective, specifically one with Marxist influences. But it can also be argued that a key staging post in this development – and one with its own decisive impacts and legacies – was something called 'welfare geography', a perspective that was most assertive in placing on the agenda questions of spatial inequalities in social welfare or well being.

The British geographer, David M. Smith, was the prime mover of this development. A highly accomplished practitioner of locational analysis, whose 1971 text on *Industrial Location* remains a classic of its kind (a second edition was published in 1981), Smith was clearly affected by the critical mood that was sweeping through (certain sections of) the academic geographical community by the early 1970s. Following his attendance at the 1971 Annual Meeting of the Association of American Geographers in Boston, when this mood first began to gain a high profile, he authored a short paper on radical geography as the coming 'revolution' and, in so doing, indexed a dazzling variety of social phenomena that 'should' become the objects for sustained human geographical enquiry. As he said:

> However we define geography … there are now (and probably always have been) sound logical reasons for studying many social phenomena that we have traditionally tended to ignore. These would include such matters as racial segregation, poverty, hunger, infant mortality, morbidity, drug addiction, mental illness, suicide, illegitimacy, sexual deviance, welfare services, medical care, crime, justice, and the incidence of areas or regions of social deprivation, and so on. Many of these phenomena show more extreme areal variation, and are stronger criteria of areal differentiation, than some of the physical and economic criteria to which we have traditionally attached so much importance. (Smith, 1971: 156)

Smith made an intriguing attempt to position this focus on social phenomena as a logical extension of a Hartshornian approach to the discipline, one preoccupied with the criteria for differentiating world regions from one another, but one that accorded little role in this regard to anything remotely 'social'. Revealingly, he here displayed some willingness to allow the coming 'revolution' to refrain older themes within academic geography, a point of some moment when it comes to reassessing his contribution. More importantly, though, he demanded an alertness to social phenomena indelibly bound up with the suffering of many human beings: in short, he called upon geographers to turn their skills to alleviating the social problems faced by many human 'casualties' of how 'society' (and, crucially, 'economy') usually operates.

This was, therefore, a critical orientation that in effect highlighted the complex intersections between social injustice and spatial inequality: between the fractures of power, status, wealth, welfare, and well being (cleaving apart the 'haves' and the 'have nots'), seemingly arising in all societies (whether capitalist, socialist, or however organized) and the differentiation between places on the ground (creating 'richer' places in contradistinction, but structurally related to, much 'poorer' places) at a range of geographical scales from the city to the region (and ultimately configuring such sociospatial relations at a global scale). Here too, it might be stated, lay the seeds of a persuasive new position within human geography: one taking its very starting point as the inequities rife in the spatial distribution of well being and social problems, producing a highly uneven welfare surface full of peaks ('winners') and troughs ('losers'). This position betrayed continuities with older geographical traditions, including regional geography. It also brought spatial science into play in the attempt to create objective measures of inequality and even to use the more formal tools of neoclassical economics to characterize the exact contours of different welfare surfaces. At the same time, however, it betrayed a deep seated urge to alter the focus of human geography – Smith called it "an alternative framework for human geography" – and in so doing to introduce a thoroughly politicized – he would later position it as a 'moral' or ethically charged – vision into the heart of human geographical enquiry. Such a position, albeit elaborated in various directions by Smith himself as well as by many others, not all of whom would necessarily recognize the label, came to be identified as 'welfare geography', most obviously perhaps in Smith's own 1977 text *Human Geography: A Welfare Approach* and then successfully popularized by him in his 1979 text *Where the Grass is Greener: Living in an Unequal World.*

David Smith was, it should be underlined, not alone in developing this perspective. Other human geographers in the late 1960s and early 1970s were starting to put inequality near the center of their concerns. An early British example is the 1971 survey of *Regional Variations in Britain*, subtitled 'Studies in economic and social geography', by B. E. Coates and E. M. Rawstron, which used the 'geographical counties of the UK' as the basis for mapping out, describing and, to an extent, explaining variations in a range of key economic and social indicators (from personal incomes and employment rates through medical and dental services to morbidity and education). The authors proposed that they were refining 'the regional idea' by using statistical evidence relevant to planning and policy. They drew explicitly upon the emerging 'scientific' – we could say spatial scientific – orientation demanded by the likes of Chorley and Haggett in their famous 1967 collection *Models in Geography.* (And the manner in which a certain resolution was here offered to tensions between the ostensibly opposed regional and spatial scientific approaches to geography, which can also be seen in Smith's studies, is itself something that might repay further exploration by historians of the discipline. Eric Rawstron was David Smith's undergraduate teacher at the University of Nottingham and Smith subsequently succeeded him as professor of geography at Queen Mary, University of London.) What Coates and Rawstron also indicated, though, was an interest in "the future well being of Britain," and an appreciation of "spatial variations that are unjust, harmful and inefficient" and which lead some to "suffer badly because they happen to live and work in particular areas." Six years later, Coates returned to these themes, this time with Ron Johnson and Paul Knox, in their better known 1977 text *Geography and Inequality,* seeking to use 'objective' social indicators to trace spatial inequalities across the UK, US, France, India, and other world regions in the likes of housing, education, nutrition, and health care. As one reviewer noted, while some on the political 'right' may see inequality as "the winder of the springs of incentive that power the capitalist system," Coates *et al.*, here "establish their position early: inequality is bad". Moreover, these authors drew extensively on Smith's contributions, "including advocacy of using the concepts of welfare economics in a spatial context."

In North America, there was also an emerging sense that human geographers should be investigating the geographies of inequality, and perhaps the clearest index was a number of papers appearing in *Antipode: A Radical Journal of Geography* during its earliest years (when the journal was itself still very much an 'underground', relatively anti establishment vehicle). In the opening two issues in 1969 and 1970, for instance, Richard Morrill, a spatial scientist of repute, outlined the possible role of geography in 'the transformation of society', insisting that 'we' have a responsibility to ask questions about the 'rightness' or otherwise of existing 'spatial patterns', many of which currently exhibit extreme imbalances between 'groups and areas' in terms of job availability and related income levels. He duly called upon geographers to be active in combating spatial inequalities in employment and resources, wealth and poverty, discrimination and racism, pollution, and the (mis)treatment of 'scenic and historic landscapes'; at base, to ensure that short term economic gain, dominated by selfishness and greed (in all sorts of societies), is countered with the injunction "[s]patial inequality in income, services and justice must be alleviated." This said, Morrill deliberately stopped short of advocating social 'revolution', "[a] simple Marxist type change," and argued instead for a system allowing greater 'social control' over 'private property' and mitigating the effects of economic power on the least advantaged groups within society. Many pieces in the early issues of *Antipode* echoed Morrill's plea to re search spatial inequality, notably the 1970 special issue on 'The geography of American poverty' edited by Richard Peet. This combined a substantial use of stat istical and cartographic methods with the plea for readers to look beneath "America's chrome surface" in order to "re discover lower class America" in the spaces of day laborers, the places of the ghetto or even the "fields of wine bottles" on the way to "the Salvation Army hostel" where "old men die in poverty." Such sentiments were, in spirit and purpose, as well as in execution, arguably very close to those of Smith as he formulated his welfare geography with a search for objective evidence and proof (of inequality) driven by a fundamentally human (polit ical, moral, and ethical) commitment to progressive social change. It is now time to provide a narrative introduction to Smith's welfare geography.

Welfare Geography: Concepts and Criticisms

That the notion of welfare – like that of the welfare state – is the subject of constantly recurring debate, or that areas of concern and critical scholarship such as welfare geog raphy emerge, is testament not only to altruism or moral concern but to the transformative and unequal dynamics of politics and society and, in that context, to the signifi cance, even 'naturalization', of economics in social life. The taken for granted acceptance of economy is clearly demonstrated in the spaceless and timeless world of wel fare economics which purports to study – at the level of society as a whole – the criteria for judging, and desir ability (or otherwise) of, the economic circumstances arising from the behavior of consumers and the con sequent desirability of changing those circumstances. Thus, and notwithstanding the normative issues involved in assessing different economic states of society, the question of welfare is internalized within economy and framed as a technical question of states of economy in which no individual can be made better off 'without' making another worse off (so called Pareto efficiency). Furthermore, the restriction of causative influence to consumers points to another indication of the limitations of this approach which, nevertheless, makes claim to assess the entire economic state of society.

By contrast, welfare geography is more clearly nor mative and more narrowly defined. For David Smith, it was initially concerned, in a positive sense, with 'who gets what, where and how' (thereby inserting the 'where' into the previously a spatial economists' and political scientists' formulations of 'who gets what and how'). This means that, first and foremost, Smith's welfare geography demanded attentiveness to the extant 'geography of welfare', neces sitating the detailed description, preferably by using ob jective measures, of inequalities in everything from, say, variations in infant mortality between different US states to the differential accessibility of people resident in different neighborhoods to hospitals in Sydney, Australia (to cite two examples from Smith's own writing). In introducing these ideas to human geography, Smith re vealed a landscape of enquiry encompassing the major social circumstances impacting upon the lives and life chances of people. More technically, he argued that the general level of welfare in a spatially disaggregated society can be thought of empirically as a function of the sum of levels of living or social well being across the territorial units comprising that society. Social well being may then be defined in terms of the societal 'goods' (i.e., benefits) and 'bads' (i.e., problems), measured through territorial social indicators, consumed or experienced as utilities or 'dis'utilities by people living in different territories, and set relative to (i.e., above or below) the general level.

Note that here, as in mainstream economics, the emphasis still remained on consumption: the welfare ef fects of production, which clearly affect the working population and those otherwise affected directly or in directly by production, become subsumed into the consumption of goods and bads. Such a focus on con sumption was later to be challenged, not least by Smith himself, in a critique of a welfare geography modeled essentially on welfare economics and, thereby, continuing to be aligned with neoclassical economic analyses which could not readily move beyond 'description' to furnishing more in depth 'explanations' of the derivation of in equalities in the welfare landscape of consumption and experience. As Smith put it:

> The early preoccupation with descriptive research in welfare geography subsequently gave way to more pro cess orientated work on the question of how inequality arises. The abstract formulation of welfare economics based in neo classical economics was found impotent as a basis for explanatory analysis, and alternatives such as Marxian economics have been useful sources of guid ance. (Smith, 2000: 898)

The consequence was that researchers needed to consider "the operation of the economic social political system as an integrated whole in order to reveal its general tendencies," and a specific proposal was that the inherent injustices of capitalism – and also perhaps certain in built tendencies to inequality even under socialism – should be brought center stage in a welfare geography with explanatory ambitions. Further, what has been realized more recently is that explaining the uneven geography of welfare must go beyond Marxian political economy to encompass a range of more sociocultural process – bound up with, for instance, colonialism, Orientalism, the dynamics of inclusion and exclusion – notwithstanding the profound importance of the dynamic geographies of circuits of value in shaping life chances.

What Smith also required of welfare geography from the outset was the additional and crucial step of critical normative analysis of such distributions of societal goods and bads through the lenses of moral and ethical geographies. This step entailed judging the gaps between what 'is' and what 'ought' to be, between the current state of a society and what might be some preferred alternative state of society which, in turn, demanded reasoned assessment of what might be morally/ethically a 'better' distribution, perhaps one that could be more socially and spatially equitable. In fact, Smith distinguishes between "ethics as moral theory, and morality as practical action." Such a readiness to judge, assess, and envisage alternatives was always present in Smith's welfare geography, and he has since stated that achieving "[a] better state of affairs with respect to any or all of the criteria of what, how, for whom and where is a welfare improvement." However, and reflecting what Smith has tackled in his more recent work on 'moral geographies', even more difficult ethical assessments of who gets what, where and how, raise highly complex practical and philosophical issues, not least those of relativism: i.e., whether 'goodness' or 'badness' of social well being can be judged only relative to specific geographical and temporal contexts. In a world of globalization, in which the nature of space is constantly disrupted through the intersection of relational and absolute/territorial formations of place, this is a fluid and constantly changing question. As such, the Enlightenment possibility of universals comes head to head with the alleged postmodern concern to decenter and for difference; and it is telling that Smith's conclusion to his 1994 text *Geography and Social Justice* included explicit attention to 'the possibility of universals'.

Some critical remarks can now be added, not only ones partially acknowledged in Smith's own self critiques but ones also reflecting certain concerns of the present authors. Leaving aside the presumption that 'welfare' has to be written in terms of 'economy' rather than, for example, being framed by what it is to be human or, more generally, by questions about what it means to have a life or to exist, a major difficulty with the conceptions underlying Smith's welfare geography – at least as initially conceived – is that welfare can never be singular, geographically even or merely material. This is not only because of the well known distinction between absolute and relative concepts of well being – with all the dynamic geographies of welfare that such distinctions imply – but also, rather, because it reflects the fact that all of the terms in Smith's nostrum are constantly changing in space and time. If 'who' refers to an individual, then are we talking about an individual constructing (or not) an identity in what exact sets of social relations? And if 'who' is a 'collective', how is this defined and limited? Is 'who' restricted to human beings? Does 'gets' refer to private or social receipt, and does it imply ownership? 'What' clearly shifts as, for example, economies change from material to knowledge based circuits of value, while 'where' also transforms constantly in the context of the dynamic interactions between absolute and relational space.

At one extreme, moreover, 'how' may be centralized and collective or, at another, privatized and individualized. How, for example, can/should welfare be thought and politically acted upon in a world in which economies are driven primarily by minds rather than bodies; where education, health, social services, and housing are framed in terms of economic competitiveness; in which the nation state becomes a market enabling state; where an ideology of privatization predominates in which individuals find identities in terms of private material success; in which exclusion and failure are seen as inevitable and to be avoided by individuals rather than confronted socially; and in which the solution of social problems is seen to lie solely through individual responses; but in which, too, markets may be seen as part of the solution while collective responses outside the state teach directly that individualism is not only damaging but impossible (think climate change and the inevitability of death)? That these issues – and many more – remain at the forefront of political debate, and address profoundly significant questions about the nature of human being, demonstrates the far seeing vision that lay behind the emergence of welfare geography, notwithstanding its tendency to abstract from the sociopolitical processes sustaining particular notions and social relations of value and the transformed geographies to which they give rise and through which welfare comes to be defined and redefined.

Welfare Geography's Legacies

In formulating welfare geography as he did, Smith reintroduced the 'human' into human geography in a very profound sense as a concern for the conditions – economic and social – of social well or ill being. This was a major achievement: it not only challenged prevailing asocial views of content and method in human geography, but focused on profoundly significant aspects of social life – poverty, racism, ill health, discrimination, and exclusion – that stunt, limit, and distort the lives of groups and individuals affected by them and that, up to that point, had largely been ignored in geography previously, despite their highly uneven incidence across worldly space. This substantive reorientation was vital, and there is no doubt that any number of research projects, monographs, and policy interventions by human geographers ever since – for example Sarah Curtis's 1989 study *The Geography of Public Service Provision* or Chris Philo's 1995 edited collection *Off the Map: The Social Geography of Poverty in the UK*, to name two British examples – stand in the horizon set by welfare geography. But it was not merely in opening up the geography of welfare as a new arena for research, filled with new objects for enquiry, that Smith's contribution mattered and continues to matter today. For what also registers is the dual attempt to go beyond describing uneven welfare landscapes to explaining them and to evaluating them according to stated principles, political, and/or ethico moral, which – however awkwardly – commit the (welfare) geographer to a critical stance often at odds with the *status quo* (not only under capitalism, in its market and neoliberal forms, but also under other societal regimes).

It is true that the term 'welfare geography' is not used all that often today, and many who practice what surely can, with legitimacy, be positioned in the heritage of Smith's welfare geography may be unaware either of these deeper roots or of their salience within the seismic rethinking of human geography that occurred in the late 1960s through the 1970s. In retrospect, it seems obvious that Smith's redirection of human geography contributed in a very significant way to the rise and shaping of radical and critical geographies, and Smith is probably too modest when he reflects that "the welfare approach [ie. his own] soon merged with other lines of critical inquiry within geography directed towards the fundamental problem of inequality." That said, the radicalism of what he advanced can easily be undervalued. This may be because of Smith's own self reflexive modesty, especially in noting that his opening version of welfare geography, with its anchorage in welfare economics, needed to be supplanted by a more 'Marxian' version attuned to political economy (and hence to a more holistic view pf political economy).

Tellingly, just such a move runs through David Harvey's massively influential 1973 text *Social Justice and the City*, as he moves from the book's Part One on 'Liberal formulations' – discussing at length the spatial (re)distribution of income in urban systems, and hence suggesting an approach akin to Smith's welfare geography (but not referencing Smith) – to its Part Two on 'Social formulations' with a classical Marxist theoretical armature. The implication could be that the former had already been tried and found wanting, and that little was to be gained from dallying with "what can only be termed 'moral masturbation' of the sort which accompanies the masochistic assemblage of some huge dossier on the daily injustices to the population of the ghetto, over which we beat our breasts and commiserate with one another before retiring to our fireside comforts" (Harvey 1973: 145). It is probable that Harvey had the likes of welfare geography in his sights with this remark, but it is one that is arguably too harsh – too impatient with the ongoing need for careful empirical reconstruction of countless welfare landscapes tyrannized by the fractures of power, status, and wealth; too hasty in denouncing the motives and convictions of Smith and others; too unappreciative of the platform that welfare geography created for the building of other, maybe ultimately more challenging possibilities for critical scholarship; and possibly too simplistically optimistic about the possible ways in which social injustice may be overcome. After all, Harvey needed just such a platform on the way to his Marxist geography, just as, it can be claimed, all manner of recent critical approaches within human geography still owes something to the substantive and normative provocations of welfare geography.

See also: Marxism/Marxist Geography I; Marxism/Marxist Geography II; Positivism/Positivist Geography; Radical Geography; Regional Geography I; Spatial Science.

Further Reading

Chorley, R. and Haggett, P. (eds.) (1967). *Models in Geography*. London: Methuen.

Coates, B. E., Johnston, R. J. and Knox, P. L. (1977). *Geography and Inequality*. Oxford: Oxford University Press.

Coates, B. E. and Rawstron, E. M. (1971). *Regional Variations in Britain: Studies in Economic and Social Geography*. London: B.T. Batsford Ltd.

Curtis, S. (1989). *The Geography of Public Welfare Provision*. London: Routledge.

Harries, K. D. (1978). Review of Coates et al's Geography and Inequality. *Geographical Review* 68, 108 109.

Harvey, D. (1973). *Social Justice and the City*. London: Edward Arnold.

Lee, R. (1979). The economic basis of social problems in the city. In Herbert, D. T. & Smith, D. M. (eds.) *Social Problems and the City: Geographical Perspectives*, pp 47 62. Oxford: Oxford University Press.

Lee, R. (1989). Urban transformation: From problems 'in' to problems 'of' the City. In Herbert, D. T. & Smith, D. M. (eds.) *Social Problems and the City: New Perspectives*, pp 60 77. Oxford: Oxford University Press.

Lee, R. (2006). The ordinary economy: Tangled up in values and geography. *Transactions of the Institute of British Geographers* 31, 413 432.

Lee, R. and Smith, D. M. (eds.) (2004). *Geographies and Moralities: International Perspectives on Development, Justice and Peace*. Malden, MA: Blackwell Publishing.

Morrill, R. (1969). Geography and the transformation of society. *Antipode* 1(1), 6 9.

Morrill, R. (1970). Geography and the transformation of society: Part II. *Antipode* 1(2), 4 10.

Peet, R. (ed.) (1970). *Theme issue dedicated to 'Geography of American poverty'*. *Antipode* 2(2), i 106.

Peet, R. (1970). Preface. *Antipode* 2(2), iv.
Philo, C. (ed.) (1995). *Off the Map: The Social Geography of Poverty in the UK*. London: Child Poverty Action Group.
Smith, D. M. (1971). *Industrial Location: An Economic Geographical Analysis*. London: Wiley.
Smith, D. M. (1971). Radical geography: The next revolution. *Area* 13, 153 157.
Smith, D. M. (1973). *The Geography of Social Well Being in the United States: An Introduction to Territorial Social Indicators*. New York: McGraw Hill.
Smith, D. M. (1974). Who gets what 'where', and how: A welfare focus for human geography. *Geography* 59, 289 297.
Smith, D. M. (1977). *Human Geography: A Welfare Approach*. London: Edward Arnold.
Smith, D. M. (1979). *Where the Grass Is Greener: Living in an Unequal World*. London: Penguin.
Smith, D. M. (1994). *Geography and Social Justice*. Oxford: Blackwell.
Smith, D. M. (2000). Welfare geography. In Johnston, R. J., Gregory, D., Pratt, G. & Watts, M. (eds.) *The Dictionary of Human Geography* (4th edn.), pp 897 898. Oxford: Blackwell.

Welfare Reform

D. Conradson, University of Southampton, Southampton, UK

Glossary

Neoliberalism An economic and political philosophy in which the market is seen as the most efficient way to allocate goods and services. As in classical liberalism, welfare is understood to primarily be the responsibility of individuals and families rather than the social collective or the state.

New Public Management A style of public sector management modeled on private sector practice. Activities are to be scrutinized with regard to their measurable contribution toward defined outcomes, and organizational units may be disaggregated so as to facilitate enhanced efficiency.

Privatization The transfer of public sector functions or assets to the private sector. It is generally assumed that competition and the profit imperative will incentivize efficiency and reduce overall costs.

Retrenchment The withdrawal of particular elements of statutory welfare provision, or the tightening of their eligibility criteria, such that the comprehensiveness of support available to citizens is reduced.

Social Polarization A widening gap between the relatively poor and relatively rich in a given society. A range of socioeconomic data are used to measure this process, including those pertaining to income, educational attainment, and health outcomes. The related phenomenon of sociospatial polarization describes the divergence between relatively deprived and relatively affluent areas over time.

Introduction

Throughout history, human societies have developed a variety of collective strategies for meeting the welfare needs of their members. These arrangements have usually sought to secure the basic requirements for life, such as food and shelter, but in some cases have extended to more complex goods such as education and healthcare. Families and local communities have typically been a first force in such provision. In post agrarian societies, however, their efforts have often been supplemented by contributions from the government, voluntary sector, and even businesses. The resulting diversity of actors is commonly described as a mixed economy of welfare.

Welfare reform is the process of adjusting the mix and contributions of the actors involved in social security provision. In practice, we can distinguish between incremental changes to welfare systems – what might be described as the tinkering of contemporary statecraft – and more thorough going processes of reform. The most recent instances of major welfare reform in Western countries have been associated with the rise of neoliberal modes of governance. During the 1980s and 1990s, the United States, Britain, Canada, Chile, Australia, and New Zealand embarked upon programs encompassing benefit cuts, privatization, the introduction of new public management (a style of public sector management premised on private sector practice), and increased reliance upon nonstatutory organizations for service delivery. Reflecting the importance of statutory welfare provision for many people's livelihoods, these changes have had substantive implications for social well being.

The Context for Welfare Reform

In advanced liberal democracies, the sustainability of a particular set of welfare arrangements is usually linked to its affordability, its publicly perceived effectiveness in addressing social need, and the popularity of the philosophical perspectives upon which it rests. If any of these elements comes under pressure, then existing welfare commitments are likely to be subjected to enhanced scrutiny by both politicians and the public.

With regard to affordability, it is clear that welfare provision represents a significant financial commitment for governments. The majority of Organization for Economic Cooperation and Development (OECD) nations spend over 15% of gross domestic product (GDP) on welfare services and support, for instance, while in the Scandinavian countries such expenditure may approach 30%. The sustainability of these arrangements depends upon states generating sufficient income to cover their costs, whether through taxation or investment returns. At the same time, welfare costs are often inclined to grow, with relevant factors including macroeconomic downturns that generate increased number of claimants, general inflation, and the financial implications of an aging population. When economic productivity declined in many Western nations during the 1970s and unemployment subsequently rose, rising welfare costs thus spurred many governments to reconsider their social security commitments. Common strategies for reducing expenditure included withdrawing selected benefits and services, along with narrowing their eligi bility criteria.

As Pierson has argued in his resilience thesis, however, governments are often constrained when implementing programs of welfare reform. Overly radical or extensive changes may provoke an electoral backlash, undermining popular support and contributing to political vulner ability. Interest groups such as families and pensioners may mobilize themselves to resist policy change, en gaging in activities ranging from written petitions through to active protest. Even with strong political and fiscal motives for cost containment, welfare expenditure may therefore be difficult to curb to practice. In Britain, such spending actually rose slightly during the 1980s, despite Conservative reforms intended to reduce it. In part, this reflected governmental hesitancy over imple mentation, with an eye to its potential political fallout, but it was also the result of a large growth in un employment claimants, a phenomenon that more than offset the savings achieved through retrenchment.

Despite these constraints, periods of economic stress often force governments to reappraise the philosophies which guide their operation and policies. In the arena of welfare, social security practices in other settings may come to be viewed as conceptual resources, suggesting alternative approaches or even potential solutions to welfare dilemmas at home. During the 1970s, the com bination of a global economic downturn and multiple instances of national stagflation accordingly prompted widespread reconsideration of Keynesian interventionism. Proponents of neoliberalism argued that the social democratic welfare had become unsustainably expensive, was implicated in client dependency, and in any case eroded personal liberty. While the latter arguments had been advanced in theoretical terms by Hayek and Fried man, it was not until the 1980s and 1990s that a range of right of center governments implemented them politic ally. The canonical examples are the Reagan and Thatcher administrations in the United States and Britain respect ively, but the governments of Australia, Canada, Chile, and New Zealand also enacted similar changes.

Since the late 1990s, the political landscape has moved on again, with so called third way administrations emer ging in a number of Western countries. Britain and New Zealand have been notable examples, with their national governments reintroducing elements of redistributive taxation and progressive welfare policy, while remaining broadly committed to economic liberalism. The term 'third way' is itself contested, but commentators have also debated whether these developments represent a genuine move beyond neoliberalism or are simply its latest variant. Whatever the case, the return of a comprehensive social democratic welfare state is generally thought unlikely, if only because the post war conjunction of Keynesian ideals with strong economic growth and near full employment is now recognized as being highly historically specific. Future welfare systems are likely to continue to draw upon a range of statutory and nonstatutory actors, each making different contributions and sometimes operating in partnerships. In view of their aging populations, Western governments will face challenges regarding state pensions and care of the elderly, but these too are likely to be addressed through multisectoral solutions.

Welfare Reform: Neoliberal Initiatives

Across the OECD, the neoliberal reforms of the 1980s and 1990s encompassed both economic liberalization and welfare retrenchment. Although policies differed be tween countries, several common forms of welfare ini tiative can be identified.

Cutting social security programs and payments was a widespread strategy for reducing welfare expenditure. A number of governments closed support and training programs, while others reduced the financial value of benefits or restricted their eligibility. Criteria such as age, household formation, and citizenship were used to dif ferentiate between deserving and undeserving claimants. Those who diverged from traditional nuclear family ar rangements, such as independent youth or sole parents, were not always well supported. The age at which a person could receive housing benefit or a state pension was in some instances raised. Some highly skilled mi grants were allowed recourse to particular state benefits, while asylum seekers and refugees were forced to rely upon informal community networks and not for profit organizations. Time limits were also imposed on benefits. Governments might require a person to have been un employed for a certain period before they became eli gible for welfare support, for instance, while the duration of payments might also be restricted.

A second dimension of neoliberal welfare reform has been privatization. In addition to the sale of nationalized industries, such as rail and telecommunications, a num ber of governments privatized welfare related assets and functions. In post war Britain, for example, state owned housing had enabled citizens to rent accommodation at affordable, often below market rates. Tenants were thereby afforded a degree of protection from the fluc tuations and speculation inherent to housing markets. Under the 1980 Right to Buy Act, however, much of this council housing was released for sale, with the accom panying discourses promoting the benefits of private ownership. The take up of this offer varied considerably, with strong sales in southern England and rural areas but fewer purchases in northern and de industrializing areas. Some council housing areas were residualized in this way, further intensifying their relatively deprived position within local regions.

A third dimension of welfare reform, and one that has also often involved the private sector, has been the contracting out of public services. In a search for cost savings and enhanced efficiency, many public sector organizations put a selection of their activities and functions, such as cleaning and refuse collection, out for tender to external organizations. Where private sector contractors have been sought, the hope has been that competition between potential providers would generate cost savings. In some cases these arrangements have reduced service costs, but debate exists over whether they sufficiently prioritize quality and safety relative to profit generation. Voluntary sector organizations have also been involved in the delivery of public services, particularly with respect to welfare and social care. Whether private or voluntary sector agencies undertake the work, managing the relevant contracts has financial and personnel implications for public sector funders. Assessing the costs and benefits of contracting out is thus more complex than a simple calculation of headline expenditure might suggest.

A fourth dimension of reform, spanning the domains of both welfare and healthcare, has been psychiatric deinstitutionalization. The impetuses and timeframe here are broader than processes of neoliberal reform, but deinstitutionalization warrants consideration because its outcomes have often been interwoven with deprivation, social exclusion, and processes of welfare retrenchment. Deinstitutionalization gathered momentum during the 1960s and 1970s, as many Western governments closed psychiatric institutions in response to stringent critiques regarding their therapeutic limitations and expense. Psychiatric in patient wards remain, but many people with mental illness are now treated via an array of community services, including outpatient clinics, day hospitals, and hostels. This move to 'care in the community' has received mixed evaluations. In particular, community services and the supply of informal care have not always been sufficient to meet the needs of people with severe mental illness. This has caused significant difficulties in some communities, and a number of geographers have accordingly examined the consequences of deinstitutionalization. Drawing on research undertaken in California and Ontario during the 1980s, Michael Dear and Jennifer Wolch's *Landscapes of Despair* explored the formation of ghettos of service dependent populations and possible policy responses. In *Bedlam on the Streets*, Caroline Knowles highlighted the complexities many individuals face in negotiating the post asylum service landscape of Canadian cities. Hester Parr and Chris Philo have provided richly textured social, cultural, and historical geographical accounts of mental illness in Britain, while Robin Kearns and colleagues have examined the dynamics of post asylum landscapes in New Zealand and beyond.

Fifth, welfare to work policies have come to the fore in a range of industrialized nations. These initiatives have overturned the social democratic notion of social security as citizenship entitlement, repositioning it as just one element within a broader portfolio of rights and responsibilities. An important argument has been that statutory welfare arrangements have become complicit in reproducing rather than alleviating welfare dependency. In response, workfare policies have sought to incentivize and reward paid employment, utilizing devices such as tax credits. Unemployment benefits have been firmly recast as interim and transitional support only, with employment framed as the expected and normal human condition. These philosophical shifts have been discursively marked in numerous ways. In Britain, what was once unemployment benefit is now 'Jobseekers Allowance', for instance, while the Department of Social Security has been renamed the Department of Work and Pensions. These linguistic changes, paralleled in many other countries, reflect a broader movement from social democracy to market liberalism.

Finally, neoliberal welfare reform has seen growing governmental engagement with voluntary organizations as service providers. While community and voluntary agencies have long offered social support in Western countries, the post war welfare state tended to displace and somewhat marginalize their contributions within the mixed economy of care. Yet their involvement grew during the 1980s and 1990s, both as a result of emergency responses to increased social need (itself reflecting worsening economic conditions and statutory welfare retrenchment) and because government agencies were directly engaging nonstatutory agencies to provide services on a contractual basis. Many voluntary organizations found the accountability and compliance requirements of these contracts to be disproportionately onerous, however, and some experienced tension between maintaining fidelity to their core objectives and meeting the expectations of funders. These concerns were encapsulated in Wolch's notion of the 'shadow state', a term she used to describe those voluntary organizations whose autonomy had become compromised in the process of meeting the terms of statutory funding contracts. Such organizations, she suggested, were often effectively functioning as agents of the state. In response to such criticisms, from the late 1990s a number of Western governments endeavored to improve their contractual funding relationships with voluntary organizations. These initiatives have sought to develop more egalitarian forms of partnership, with improved channels for dialog and, in some cases, longer funding time frames. Such maneuvers also reflect the broader interest Western governments have in working with voluntary organizations as a way of engaging citizens, promoting well being and developing social capital.

Geographical Dimensions of Welfare Reform

Geographical research into neoliberal welfare reform has highlighted a series of issues. Most notably, welfare retrenchment has been linked to increased sociospatial polarization. A number of studies have demonstrated that welfare reforms implemented in the context of broader economic restructuring have tended to impact disproportionately upon already marginalized groups, including female headed households, minority ethnic communities, youth, and the elderly. In the United States, Glasmeier's work has highlighted the extent to which these social groups and their neighborhoods have been left behind, in relative if not absolute terms, despite the general trend of rising wealth in recent years. Similar socio spatial polorization is evident in Lupton's examination of 12 relatively deprived areas in Britain, undertaken as part of a longitudinal evaluation of an area based poverty intervention program. Even where case study areas recorded improved employment, education, and health outcomes, the 'rate' of this improvement was consistently less than the national median. Le Heron and Pawson's consideration of the impacts of neoliberal restructuring in New Zealand paints a similar picture of increased unemployment and social exclusion in a number of communities, many of which were already economically or geographically marginal. In short, some of the most disadvantaged social groups and places have often been disproportionately impacted by neoliberal reforms.

For those seeking assistance in particular places, a second finding is that neoliberal reforms have tended to increase the spatial unevenness of welfare provision. The principle of uniform access to support and services was central to many post war welfare states, even in those communities where it remained more aspiration than reality. In contemporary welfare arrangements, the diversity of actors involved has made it more difficult to achieve the level of planning and coordination necessary for such equity. Unevenness has also sometimes followed from the devolution of policy responsibilities to smaller scales of government. In the United States, responsibility for some elements of welfare policy has shifted from the federal to the state level, for instance, fostering new forms of geographic differentiation in terms of the assistance available to citizens. A further factor in increasing the unevenness of service provision has been the growing involvement of nonstatutory organizations in service delivery. Where businesses are concerned, this unevenness typically reflects the fact that firms locate in areas where they are able to generate a profit. In relatively deprived or geographically peripheral locations, there may therefore only be one or two businesses willing to compete for a home care or personal services contract. Equally, in some disadvantaged areas, market conditions may fail to support any viable private sector care providers at all. In the absence of state intervention, significant gaps in service provision may then arise. For voluntary welfare organizations, spatial unevenness emerges not as a result of the variable potential to make a profit, but rather because of the differing levels of motivated staff, volunteers, and charitable resources within particular communities. The emergency relief and drop in centers evident in some cities may thus be largely absent in others, despite similar levels of deprivation and need. Voluntary sector unevenness may also emerge as a result of variations in contracting out practices between different divisions of local government. Research undertaken by Fyfe and Milligan in the Scottish city of Glasgow identified a correspondence between the geography of local government administrative units and the relative presence or absence of voluntary welfare services in particular areas for instance. Whatever its origin, such unevenness of provision matters at a time when many industrialized nations are emphasizing the potential of nonstatutory organizations to meet local welfare needs.

A third research finding is that the effectiveness of state welfare programs varies in relation to the composition and characteristics of local communities. In Sunley and Martin's study of the New Deal for Young People, a workfare program designed to address youth unemployment in Britain, significant variations were observed in the outcomes across the country. In rural or relatively prosperous regions, where casual and relatively unskilled work was easier to access, the scheme generally experienced greater success. In deprived urban communities, the capacity of local employment to function as a route out of deprivation was comparatively limited. While such geographic variation may be unsurprising, it nevertheless challenges the aspatial assumptions of much contemporary welfare policy. Variations in local labor markets, along with other community characteristics, may influence the success of welfare policies in significant ways.

Fourth, geographers have traced the circulation of welfare policy between places and across scales. There has been widespread interest in places such as Wisconsin, for example, with policymakers from elsewhere examining its well known workfare program in search of potentially relevant insights. In considering the transnational knowledge flows that such interest generates, Jamie Peck has written of the "geopolitical etymology of workfare." Policy transfer is of course never a matter of straightforward copying, as practitioners are typically selective in their adoption of ideas from particular settings, while implementation inevitably involves adjustments for local conditions. There will also be constituencies who require persuasion regarding the merits and practicability of the imported ideas. Even within nations, the transfer of policy ideas and practices between different scales involves such negotiation. With reference to the United States, Peck has thus argued that welfare reform should not be understood as a top down process, implemented by Washington, but rather as the product of a "complex set of political interactions between the federal, state and local levels."

Conclusion

In work that spans North America, Europe, and Australasia, geographers have been active in tracing the impacts of welfare reform in recent decades. In examining the shift from social democratic through to neoliberal and, in some cases, third way modes of governance, their research has examined the emplaced consequences of benefit cuts, privatization, deinstitutionalization, public sector reform, and the growing engagement with non statutory organizations as service providers. We now know more about the connections between neoliberal modes of governance and processes of sociospatial polarization, as well as the increased unevenness of welfare provision itself. While in some settings this unevenness has been ameliorated by the emergence of third way politics or the activities of voluntary organizations, it has by no means gone away. Indeed, in most OECD nations researchers continue to identify social groups who remain relatively marginalized despite the welfare systems that are intended to support and assist them. One might think here of the long term unemployed, those who are homeless, and people with serious mental illness, among others. In order to assist these individuals, and because new social forms of social need regularly arise, it is clear that welfare reform remains on the political agenda for many governments. As long as that is the case, it will remain important for geographers to consider its consequences for individuals and social groups in particular places.

See also: Health Geography; Health Services Restructuring; Health Systems and Health Services; Inequality; Mental Health; Poverty; Voluntary Sector; Welfare Geography.

Further Reading

Cope, M. and Gilbert, M. (2001). Geographies of Welfare Reform. *Urban Geography* 22(5), 385 390.

Dear, M. and Wolch, J. (1987). *Landscapes of Despair: From Deinstitutionalization to Homelessness*. Cambridge: Polity Press.

Dorling, D., Rigby, J., Wheeler, B. *et al*. (2007). *Poverty, Wealth and Place in Britain, 1968 to 2005*. Bristol: Policy Press.

Fyfe, N. and Milligan, C. (2003). Out of the shadows: Exploring contemporary geographies of voluntarism. *Progress in Human Geography* 27(4), 397 413.

Glasmeier, A. (2006). *An Atlas of Poverty in America: One Nation, Pulling Apart, 1960 2003*. New York: Routledge.

Handler, J. F. (1995). *The Poverty of Welfare Reform*. New Haven, CT: Yale University Press.

Joseph, A. E. and Kearns, R. A. (1996). Deinstitutionalisation meets restructuring: The closure of a psychiatric hospital in New Zealand. *Health and Place* 2(3), 179 189.

Knowles, C. (2000). *Bedlam on the Streets*. London: Routledge.

Le Heron, R. and Pawson, E. (eds.) (1996). *Changing Places in New Zealand*. Auckland: Longman.

Lupton, R. (2003). *Poverty Street: The Dynamics of Neighbourhood Decline and Renewal*. Bristol: Policy Press.

Mohan, J. (1999). *A United Kingdom? Economic, Social and Political Geographies*. London: Arnold.

Parr, H. (1997). Mental health, public space and the city: Questions of individual and collective access. *Environment and Planning D: Society and Space* 15(4), 435 454.

Peck, J. (1998). Workfare: A geopolitical etymology. *Environment and Planning D: Society and Space* 16(2), 133 161.

Peck, J. (2001). *Workfare States*. London: Guilford Press.

Philo, C. and Wolch, J. (2001). The 'three waves' of research in mental health geography: A review and critical commentary. *Epidemiologia e Psichiatria* 10(4), 230 244.

Pierson, P. (1996). The new politics of the welfare state. *World Politics* 48, 143 179.

Sunley, P., Martin, R. and Nativel, C. (2006). *Putting Workfare in Place: Local Labour Markets and the New Deal*. Oxford: Blackwell.

Wolch, J. R. (1990). *The Shadow State: Government and Voluntary Sector in Transition*. New York: Foundation Centre.

Wellbeing

S. C. MacKian, Open University, Milton Keynes, UK

Glossary

Eudaimonia A concept first used by Aristotle to indicate human 'happiness' or 'flourishing'. Stemming from the Greek 'eu' (happy, well, harmonious) and 'daimon' (the individual's spirit).

Health "A state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity" (World Health Organization (WHO) definition).

Welfare Physical and mental health and happiness. A system of help given to support those who need financial, medical, or social assistance.

Well-Being A good or satisfactory condition of existence; a state characterized by health, happiness, and prosperity. The state of feeling healthy and happy.

Introduction

Well being is a concept that permeates all aspects of contemporary society, from the individual to the work place, the hospital to the national economy. To achieve personal well being is one of Western society's key as pirations. However, it is not a concept we can all agree upon – it is inherently subjective, relational, and con tested – and must be understood within the context in which it emerges.

History of Understandings of Well-Being

Historically the search for well being was a central part of what it meant to be human. As is frequently the case, we can turn to Aristotle for an insight into the philo sophical underpinnings to the holistic nature of well being. Aristotle's notion of 'eudaimonia' (best translated as 'human flourishing') represented the ultimate human good. However, contrary to key elements we might as sociate with contemporary understandings of ultimate human good, eudaimonia goes beyond the trappings of wealth, power, and knowledge, to focus on the virtues that underpin a life premised on spiritual, emotional, and material well being. It highlights strength of character, including courage, honesty, and pride, together with mutually beneficial rationality.

The underlying belief in the need to achieve 'balance' in the art of living continued to underpin philosophical and popular understandings of well being, illustrated by the continued popularity of sanctuaries of rest and re laxation such as spas, from the Roman times onward. With the advent of scientific medicine however, the ad vancement of the biomedical paradigm and the birth of the Enlightenment subject, this holistic notion of well being was lost as the body became divorced from the mind. As biomedical belief took over from folklore, superstition, and religion as sources for the maintenance and promotion of health and well being, disease of the 'body' became the primary focus and well being of the 'person' was increasingly overlooked. Furthermore, ana tomical specialisms subdivided the body and mind so that each instance of disease could be isolated, measured, and diagnosed. This was arguably further compounded in the latter part of the twentieth century with spiritual breakdown and falling formal religious participation, and well being became increasingly something that could be bought at the counter or obtained from the doctor instead. Western societies had lost the art of locating well being at the heart of human existence. As it is fundamentally this divorce in Western rational scientific thinking that has resulted in subsequent confusion about what well being is and how to comprehend, measure, or enhance it, this discussion focuses mainly on developing understandings of well being in Western societies; al though as will become clear the relation to more holistic understandings of well being rooted in Eastern philo sophical and spiritual teachings has been an ongoing undercurrent.

In Western society by the twentieth century, or 'high modern' period, well being was predominantly under stood, conceptualized, and produced through the insti tutions of national government; the British Welfare State being the ultimate embodiment of this. As the apparent success of the Welfare State took hold and the worst excesses of 'illness, ignorance, disease, squalor, and want' were eradicated, concerns over well being were relocated to wider structures of nation and economy. Discussion grew in relation to the 'economic well being' of countries alongside the individual right to financial security with the growth of market liberalism. At the root of this was the economic imperative that more wealth would lead to improved well being, and that material well being was therefore the solution to many of society's wider issues and malaises. So much so that social, emotional, or vir tuous understandings of well being in a more holistic sense were almost completely lost, as illustrated by the successful US presidency campaign in the 1990s by Bill

Clinton with the infamous slogan "It's the economy stupid."

However, there emerged a growing recognition that GDP cannot alone stand for a measure of well being. Although well being appears to increase with increased wealth, this is only to a certain level, and it was progressively recognized that well being and happiness do not correlate simply with material advancement. This led to the development of a number of 'indexes' to measure well being, and by the end of the twentieth century, well being in a wider sense was firmly on the political and social agenda. In 2005 Tony Blair, the then UK Prime Minister, turned Bill Clinton's slogan on its head, vowing to find out how 'happy' his country was while espousing "It's *not* just the economy stupid." Even the leader of the opposition, David Cameron, spoke of the need to make people happier and to refocus from 'GDP to GWB' (general well being). At the start of the twenty first century, well being is thus a central political challenge of our times, a sign of how used to continuous economic growth Western society has become, and how aware we now are of the sometimes hollow benefits that accompany it.

Alongside this changing political climate alternative understandings of well being were being promoted at the margins by the New Age Movement, urging us to leave behind the stresses of mainstream society and develop the mind, body, and spirit holistically in order to achieve well being. This was essentially a Western reappropriation of Eastern traditions and Pagan histories, reapplied in a postcolonial, postmodern interpretation of well being for groups finding themselves marginal to the central drives of Western capitalism. It heralded the changing conceptualization of well being that was to come. The rise of reflexive individualization and self help was beginning to elevate the individual, allowing the self to take center stage in discourses of well being by the close of the twentieth century. Well being from this perspective became firmly located in the individual body, with diet, exercise, and supplements being key components in the regime of the responsible individual who was to maintain his/her own well being, no longer relying on state institutions.

As understanding of well being has become more sophisticated there has been a proliferation of studies and avenues for exploration and hence a significant variety of contexts within which discourses of well being are being constructed. The well being of 'women', for example, has come to be seen not just as an issue of dignity and human rights, but as an essential element in global development. Thus the Millennium Development Goals identify women's well being as a central component in the eradication of poverty. The well being of 'children' is also of growing international concern, and 'environmental' well being continues to be high on international, national, and local agendas. Since the 1992 Rio Earth Summit it has been increasingly recognized that the well being of the environment influences human well being, and should therefore be of international political concern. At the 'individual' level, the turn of the twenty first century has seen a renewed interest in the mind, body, spirit triadic conceptualization of well being. With the advent of 'risk society' and increasing opportunity and necessity for self reflexivity, the rise of 'self help' literature and the expansion of therapeutic 'experts' can be seen as further evidence of the search for a meaningful, knowable focus in a disorientating world.

Elements of Well-Being: Happiness, Health, and Prosperity?

The difficulty in defining well being is in the nature of the factors it hinges upon. Well being might be understood as 'happiness, health, and prosperity'. However, 'health' itself is a term even the WHO has experienced difficulty in pinning down, and how do we define such relative concepts as 'happiness'? Alternatively well being can be conceptualized as a mixture of meeting basic needs and simple desires. Yet again we face the problem that need is a concept that has troubled academics and policymakers alike for centuries, and one person's desires can impinge upon another's well being. It might be more helpful to think of well being therefore as hinging upon some measurable notions of 'welfare' together with some self ascribed understandings of 'contentment'. However, a third element must be added to the equation, as without a sense of 'dignity' the procurement of the other two may prove to be a hollow victory.

Thus well being is a slippery concept which defies easy definition and varies in time and space. Dictionary definitions suggest it hinges on the state of being well, happy, or prosperous. However, an individual can certainly experience a profound sense of well being without being prosperous, and health is not a prerequisite for well being. 'Subjective well being' is generally premised on three components: the presence of pleasant emotions, the relative absence of unpleasant emotions, and personal judgments about satisfaction. Subjective well being and the judgments of satisfaction attached to it, can relate to specific conditions, such as marriage, home, or work, and can also involve explicit goals or values. Hence the inherent complexity across and within societies, and even within individuals over time and space. 'General well being' may on the other hand relate to some assessment of attributes to be found within a society or institution that are taken as proxy measures of 'satisfaction'. Without wanting to pass judgment on the relative value, weighting, or contribution of each element, **Table 1** makes

Table 1 Elemental table of well-being

Element	*Key attributes*	*Explanation*
Physical	Health, illness, fitness, disability	Although health is only one possible, and not necessary, condition of well-being, the body nonetheless represents a key site for the enactment, performance, and securing of well-being. Hence, e.g., the notion of designing 'therapeutic landscapes' to enhance well-being and hasten recovery from ill-health and surgery. Nonetheless it is undoubtedly possible to experience a sense of well-being during periods of acute illness. People in the situation of long-term illness often end up with a profound sense of well-being on many levels because they have more to come to terms with and have been forced to develop ways of being which most people have never had to contemplate. Thus a subjectively assessed sense of well-being can transcend physiological state. This is also being acknowledged in studies of health and place which are seeking increasingly to move notions of 'health' beyond medicalized constructions.
Social	Relationships at personal, community, and society level	Social connectedness can lead to both individual well-being and community or social well-being. This aspect is reflected in an increasing interest in interconnectedness, interdependence, and notions of 'social capital' which have pushed social well-being center-stage. Here the argument is that weakening of social relations has led to a degradation of social networks and the benefits they accrue. This discourse has also infiltrated educational agendas with national governments, local authorities, and individual schools and workplaces recognizing and acting upon the significance of well-being to the smooth running of institutions, communities, and families.
Economic	Consumption, wealth, material advancement, career, work	Economic comfort and good health do not automatically equate with a sense of well-being. Conversely, some who experience what has been identified as a quantifiably poor quality of life may still feel a sense of well-being. The idea of economic well-being can be tied to the national economy, or focus in upon the individual self, particularly in the context of the workplace. Increasingly in a Western context the workplace is being seen as a place where the individual should not have his/her well-being undermined-both for the employees' dignity and for the economic success of the employer. Indeed measures to promote well-being in the workplace, from Tai Chi classes to the creation of therapeutic outdoor spaces, have been adopted as one strategy toward increased productivity in the workplace.
Political	Democracy, dignity, choice, influence	The right to dignity, choice, and an opportunity to make change is a central part of well-being. The well-being agenda is now also a big part of political strategy. Politicians and think tanks in countries as far apart as England, Finland, and Australia are bringing well-being and happiness to the political arena, while others, e.g., Bhutan have always held it centrally. From implicitly addressing it-through, e.g., health, wealth, and housing-to explicit use of the term 'well-being' and a wider sense of 'happiness' itself.
Environmental	Home, living, nature, global	Living environment has long been seen as an issue for health and well-being, from the pioneering work of people like Edwin Chadwick. Although the negative effects of poor housing on well-being continue to be documented, even relatively good-quality housing is now being highlighted as potentially damaging to inhabitants well-being. New housing developments, e.g., have been shown to undermine the well-being of young families with lack of gardens, parks, and play areas. Wider reflections of the importance of environment to well-being are reflected in the development of such measures as NEF's Happy Planet Index.
Emotional	Self-image, personal growth, self-esteem	The Western individual, emerging through the subjective turn and consolidated through the reflexivity of contemporary society is at the heart of current conceptualizations of well-being. Here well-being implies a sense of agency and the ability to live a life that is experienced as meaningful. However the obsessive consumption driven individualization permeating Western society also means the widespread promotion of regimes of health, well-being, mind, and the body beautiful are frequently motivated by fear campaigns that create false needs and induce feelings of personal and emotional ineptitude, thereby undermining emotional well-being.
Spiritual	Spirituality, faith, inner peace, life path	Historically spiritual awareness and development was a fundamental part of well-being, be this through self-exploration, holistic treatments for illness, or as part of organized institutional religion. However, with the rise of rationality and modernity, a practical drive to find well-being rooted in the empirically observable physical body lost this edge. To some extent however spiritual well-being is reemerging as an essential part of the overall mix. Western society now sees a mix of serious spiritual path finding, as well as an element of susceptibility to the latest consumer trend in the search for spiritual fulfilment.

NEF New Economics Foundation.

some attempt to pick apart the notion of well being, so we can scrutinize each element in a little more detail.

How Do We Measure Well-Being?

The history of well being appears to be one of obsession with measuring the seemingly immeasurable. However, by employing a multimethod approach researchers are able to develop adequate empirical representations of well being. At an individual level, the most common psychological tools available are self report surveys, many of which have been developed for use in clinical assessments in a variety of contexts (such as the General Well Being Index). At the collective level a number of models, indexes, or approaches have been developed to measure well being. A selection of these are summarized in **Figure 1**.

Relevance and Significance of Well-Being: Why Does Well-Being Matter?

Since the days of Charles Booth mapping of poverty, health and other objective indicators of quality of life have had clear applications for policymakers and practitioners. However, these measures may miss the varying degrees of resilience and emotional thriving that individuals demonstrate. It is often this individual response to and judgment of circumstances which influence the degree to which any one individual copes with the same environmental factors – even in crisis situations. Hence objective measurements should not be taken alone in quality of life assessments and this is where the notion of well being adds substantially to the quality of life debate.

While the obvious benefits of adequate income, good health, and solid social relationships might be clear, research suggests that a sense of subjective well being can

Gross National Happiness

- A term coined and developed in Bhutan from the 1970s to reflect the importance of spiritual development alongside material development in order to secure human well-being. Based on four principles:
  - Equitable and sustainable socioeconomic development
  - Preservation and promotion of cultural values
  - Conservation of the natural environment
  - Good governance

The Ottawa Charter

- In 1986 an international health promotion conference identified three key components to securing physical, mental, and social well-being. The ability to:
  - Identify and realize aspirations
  - Satisfy needs
  - Cope with or change the environment

Capabilities Approach

- An emerging approach to poverty in the1990s which emphasizes the importance of:
  - Practical choice and agency
  - Beyond simple notions of availability or access to resources or utilities

Human Development Index

- Developed in Pakistan in 1990 measuring average achievements in a country based on:
  - Life expectancy
  - Literacy rates and education enrolment
  - Standard of living

Genuine Progress Indicators (GPIs)

- Idea first developed in California in 1995, and now taken on board in a number of different contexts (e.g., the Canadian Index of Well-being, GPI Atlantic, GPI Online Australia). Originally based on 26 social, economic, and environmental variables covering:
  - Indicators and measures of progress
  - Assessment of economic value of nonmarket social and environmental assets

Happy Planet Index

- Developed by the UK New Economics Foundation (NEF) in 2006, linking environmental well-being explicitly with social well-being. NEF assessed the ecological efficiency with which human well-being was being achieved. This combines a measurement of:
  - Ecological footprint
  - Measures of life satisfaction
  - Life expectancy

Figure 1 Well-being indexes.

in itself influence well being and success in other areas. Someone who is happy most of the time is likely to have more self confidence, perform better economically, and may even live longer. So what is the wider significance of well being? In relation to this it is worth expanding a little upon three main areas.

Individuals and the 'Well-Being Industry'

In contemporary interpretations, self responsibility and self actualization emerge as key ideals in the procure ment of well being, and there is a growing social struc ture and commercial network emerging to accommodate that. Lash and Urry describe the rise of 'reflexivity en hancers' – personal trainers, personal shoppers, stylists, dieticians, Tai Chi instructors, health spas; all those who offer the promise of luxurious self improvement and an enhanced sense of well being with the swipe of a credit card. The self is very much a project to be worked on and these individuals and institutions provide the promise of achieving soul satisfaction. The pinnacle perhaps being the 'life coach' – the expert charged with the total physical, mental, and spiritual rescue of the failing in dividual. In short, a thriving well being industry has been born in response to the modern search for our very own eudaimonia.

Politics: A New Development Agenda?

Beyond a certain level, an increase in material wealth does not lead to an equal rise in overall well being. Thus, economic prosperity does not always impact positively on the subjective well being of a nation. Progress and eco nomic opportunity for some may lead to misery in other aspects of their lives and social or environmental deg radation more generally. For example, the rise of private car use although ensuring freedom of movement in privacy and comfort for large sections of the world's population, has also led to parental paranoia and child hood obesity as the street became transformed into a car dominated space of fear. At global levels it has con tributed to damaging the Earth's atmosphere and is a factor in increasingly polarized wealth.

In recognition of such trends, a wider 'well being agenda' is progressively underpinning government policy drives in wealthy industrialized countries and political discourse around well being has grown. From health and education to work–life balance and responsible con sumption patterns, political rhetoric and structures are opening up to facilitate choices and actions aimed at enhancing individual well being. Politically understand ings of well being are also therefore hinged upon indi vidual agency and responsibility, and the number of areas where we might exercise that responsibility is growing.

Moral Imperatives: Individual Choices in a Global World

The British political system invented the Welfare State and in some senses may therefore consider itself to have the moral upper hand in terms of enhancing individual and national well being. Yet, it continues to contribute to international humanitarian atrocities, including war, arms trading, and environmental degradation, with apparent scant regard for well being beyond its own borders. Thus, there are inevitably times where one person's well being may hinder that of another. Britain is far from alone in setting such double standards. Nonetheless, a growing exploration of well being and heightened political awareness can only serve to raise the profile of such in consistencies and offer opportunities for individuals and civil society to respond. The rise of movements such as Reclaim the Streets, Guerrilla Gardening, and Green peace all offer opportunities to highlight social and en vironmental issues affecting well being, together with suggesting avenues for action. Thus, the 'responsible in dividual' does not relate solely to self focused choices, but can also involve choices made through a moral sense of responsibility to others, and can therefore act as an agent of change.

The Future of Well-Being?

Well being as a concept and goal has been with us since the earliest philosophical and moral debates, and con tinues to transform through political appropriations and changing social aspirations. As the world changes, so will our requirements, expectations, and measurements of well being. Perhaps we will become more in tune with the likes of the Bhutanese in realizing that with advances in modernization comes the threat of loss of other fun damental components of well being, such as tranquility, happiness, and spirituality. Around the world most studies report an element of subjective well being among populations regardless of, and often in spite of, economic or environmental circumstances. Thus, out of all the elements of well being discussed it is perhaps the sub jective one, and ultimately the one most difficult to measure satisfactorily, which should be our key concern.

Continued advances in information technology, bio genetics, and the automation of society are likely to raise new questions, in particular around our understanding of relative and experiential rather than absolute notions of well being. With the use of new technologies, maps of physical and emotional well being are already being produced in a variety of inventive ways. Physiological measurements of heartbeat, skin response, and neuro logical activity have been mapped to plot the 'well being' of people places. Such work has the potential to tell us a lot more about what well being 'really' looks like and offers hope for innovative and useful understanding in the future.

The likelihood of well being emerging as something we can conceptualize simply and definitively remains unlikely, but the growth of political interest in the con cept means it is likely to be increasingly central in both economic and social analysis. At the time of writing, major contemporary texts in human geography omit well being from the index completely. It is unlikely such an omission can continue. Geographical variations in interpretations, cultural experiential factors, and the uniqueness of individual subjectivity all ensure that the difficult task of nailing this nebulous concept will con tinue for sometime yet, and geographers should be playing a key part in that undertaking.

The growth of academic work on reenchantment and spirituality, the continued popularity of grassroots activ ism addressing global imbalances, and a vastly expanding and thriving well being industry, are all indicative that well being is set to remain on the agenda and continue to raise interest. It is now more than ever a concept pushing at the doors not only of the philosophers, but of world leaders, policymakers, and academics alike.

See also: Affect; Emotional Geographies; Health Geography; Therapeutic Landscapes.

Further Reading

Adam, B., Beck, U. and van Loon, J. (eds.) (2000). *The Risk Society and Beyond Critical Issues for Social Theory*. London: Sage.
Alkire, S. (2005). *Valuing Freedoms: Sen's Capability Approach and Poverty Reduction*. Oxford: Oxford University Press.
Crossley, N. (2001). *The Social Body: Habit, Identity and Desire*. London: Sage.
Deneulin, S. (2006). *The Capability Approach and the Praxis of Development*. Basingtoke: Palgrave.
Diener, E. and Eunkook, M. S. (eds.) (2000). *Culture and Subjective Well being*. London: MIT Press.
Gaston, J. and Vogl, L. (2005). Psychometric properties of the General Well Being Index. *Quality of Life Research* 14(1), 71 75.
Giddens, A. (1991). *Modernity and Self Identity: Self and Society in the Late Modern Age*. Cambridge: Polity Press.
Griffin, J. (2002). *Well being: Its Meaning, Measurement and Moral Importance*. Oxford: Clarendon.
Lash, S. and Friedman, J. (1992). *Modernity and Identity*. London: Blackwell.
Lash, S. and Urry, J. (1994). *Economies of Signs and Space*. London: Sage.
McCready, S. (ed.) (2001). *The Discovery of Happiness*. Naperville, IL: Sourcebooks.
Putnam, R. (2000). *Bowling Alone: The Collapse and Revival of American Community*. New York: Touchstone.
Raz, J. (2004). The role of well being. *Philosophical Perspectives* 18(1), 269 294.
Sumner, W. (1996). *Welfare, Happiness, and Ethics*. Oxford: Clarendon Press.
Taylor, C. (1989). *Sources of the Self: The Making of the Modern Identity*. Cambridge, MA: Harvard University Press.
Williams, A. (ed.) (1999). *Therapeutic Landscapes: The Dynamic between Place and Wellness*. New York: University Press of America.

Relevant Websites

http://biomapping.net
BioMapping.
http://www.gpiatlantic.org
GPIAtlantic, Genuine Progress Index for Atlantic Canada.
http://www.gpionline.net
Gpionline.
http://www.greenpeace.org
Greenpeace International.
http://rts.gn.apc.org
Reclaim the Streets.
http://www.atkinsonfoundation.ca
The Atkinson Charitable Foundation: Canadian Index of Wellbeing (CIW).
http://www.guerrillagardening.org
The Guerilla Gardening.
http://www.happyplanetindex.org
The Happy Planet Index.
http://www.worklifesupport.com
Worklife Support.

Wetlands and Reclamation

M. G. Hatvany, Université Laval, Quebec City, QC, Canada

Glossary

Ahistoric Lacking historical perspective or context.
Drainage Lowering groundwater level by gravity or mechanical means.
Marais Salant French term for salt marsh transformed into a salt-producing basin using solar energy to evaporate seawater. Often erroneously confused with *marais salé* (salt marsh).
Mitigation Replacing the lost functions of a disturbed environment with those of another natural or purposely engineered environment.
Polder Dutch term for reclaimed wetland water management unit where groundwater levels are controlled by dikes, ditches, and/or pumps.
Reclamation Word erroneously used to refer to the initial transformation of wetlands for agricultural/urban use through drainage. In Holland, where peatlands were initially drained, they needed to be reclaimed (drained a second time) because of subsidence. Reclamation is now synonymous with initial and secondary wetland development and implies the upgrading of extensively used land to make it fit for a higher intensity of land use.
Restoration The act of restoring a disturbed environment to a former natural state.
Subsidence The shrinking, consolidation, and oxidization of formerly saturated peatland due to water loss after drainage.
Terpen/Wierden Dutch terms for ancient wetland-dwelling mounds.
Wasteland Perceived to be marginal land in both setting and economic productivity.
Wetland A transitional environment (ecotone) between terrestrial and aquatic environments that is inundated or saturated by surface or groundwater at a frequency or duration to support vegetation adapted to life in saturated soils.

Introduction

Wetlands are ecotones (transition zones) between terrestrial and aquatic environments. They make up a myriad of landforms that are inundated or saturated by water, part or all of the year, and support specialized vegetation adapted to such conditions. In the US alone there are more than 90 names used for wetlands, and while some are familiar (bog, bottomland, fen, moor, mangrove, marsh, peatland, tundra, swamp, etc.), differentiating between others is a specialized science. Neither land nor sea, human interaction with wetlands often seems a study in contradictions. Wetlands have frequently sported two opposing facades: on the one hand, the marshes of Mesopotamia are thought to be the inspiration for the Garden of Eden in the Old Testament. On the other hand, the mythical battle between Hercules and the Hydra in the Greek swamps of Lerna is viewed as a metaphor for the struggle to reclaim wetlands and make an insalubrious environment more amenable. Like all landscapes, human perception and interaction with wetlands have changed as a result of alterations in demography, economy, environment, and technology. In response to widespread destruction of wetlands through reclamation in the past 150 years, wetland preservation, mitigation, and restoration have become major societal preoccupations seeking to reverse destruction and promote biodiversity and water functions (**Figure 1**).

Physical Milieus

It is estimated that about 6% of the world's land surface (7–8 million km^2) is wetland. The significance of this area is magnified by the fact that wetland ecosystem productivity is generally high compared to that of many other aquatic and terrestrial ecosystems. Far from the popular image of wasteland, many wetlands contribute a far greater proportion than other landscapes to overall biological productivity and water resource functions. They are found throughout the world, in humid cool areas as bogs, fens, and tundra, as riparian wetlands along rivers and streams, as seasonally flooded bottomland in forest and swamp, in the estuaries of major rivers, along temperate, tropical, and subtropical coastlines as marshes and mangroves, and in arid areas as periodically flooded salt flats. Well known wetlands linked to human history in the Old World include those of coastal Holland, the Fenlands (England), the Po River Delta (Italy), the Danube Delta (Romania and Ukraine), the Mekong Delta (Vietnam), the Nile River and Delta (Egypt and central Africa), the Mesopotamian Marshes (Iraq and Iran), and the Yellow and Yangtze basins (China). In the New World, the list is no less impressive, including the Bay of Fundy (Canada), the New Jersey Meadowlands, Everglades, and Mississippi River Delta (USA), the Orinoco Delta (Venezuela), the Pantanal (Brazil), and the Swan River estuary (Australia).

Because of the generally high vegetal productivity of wetlands, the standing stock of animal biomass is also

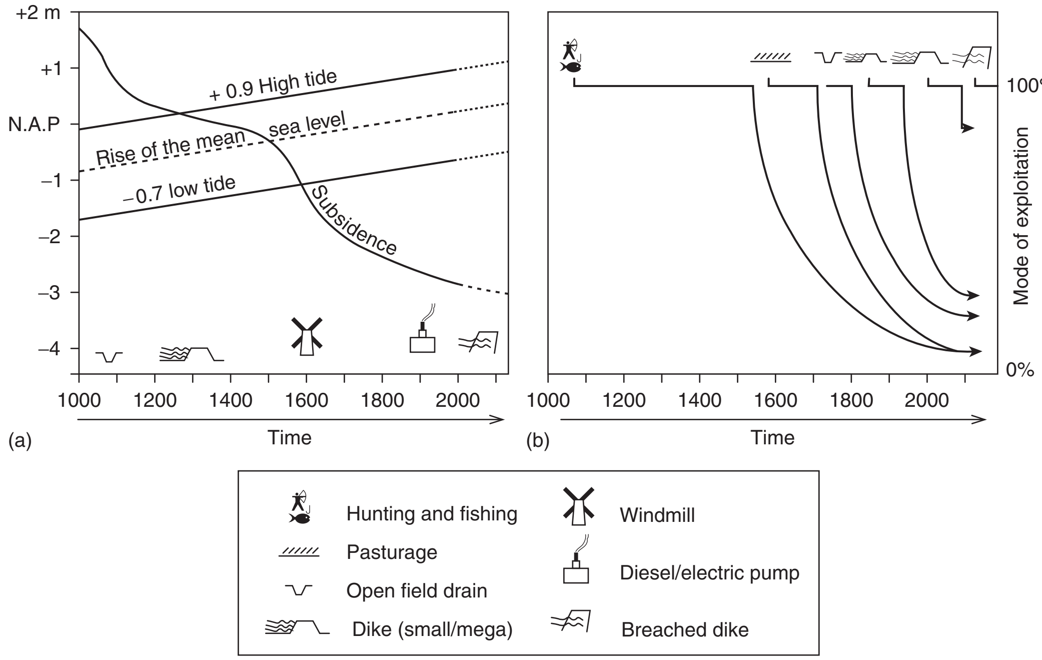

Figure 1 (a) Technical innovations and lowering of the surface of coastal fenlands in Europe. (b) Technology and models of exploitation and transformation of coastal wetlands in Northeast US and Canada (excluding Acadia).

correspondingly high. While rates of production and decay vary, part of the floral and faunal biomass produced each year in wetlands falls to the surface as detritus. In the underlying anaerobic conditions, partial decay of the detritus results in thick strata of peat. On the surface, organic productivity is expressed in masses of crawling, swimming, and burrowing insects that digest and trans form tons of vegetable matter, establishing a rich food web attracting stocks of fish and waterfowl seeking food, refuge, and safe nurseries for reproduction. At the apex of this food web have been human beings eager to exploit the abundant flora and fauna, and noting the agricultural potential of the underlying peat.

Contrasting Perceptions

Coexistence

Only recently has modern society begun to recognize what premodern humans intrinsically understood: wet lands are fecund places. While precocious groups like the National Trust in Britain began preserving wetlands as early as 1899 (Wicken Fen), and hunter supported or ganizations like Ducks Unlimited in the US and Canada began to promulgate continental scale wetland conser vation in 1937, the first scientific conferences on wetlands began only in the late 1950s. A decade later, during the emergence of worldwide environmental movements, John and Mildred Teale's *Life and Death of the Salt Marsh* (1969) translated new scientific understandings of wetland fecundity into laymen's terms. Similar texts followed in scientific and popular journals. The prevailing discourse underlined the need for public understanding of the worldwide decrease in wetland area in the past 150 years as a result of reclamation. In the US, it was pointed out, more than half of the wetlands had been lost since 1850, while in the Old World the percentage was greater.

Focusing on wetland disappearance since industrial ization, most recent wetland literature has been funda mentally ahistoric, viewing humans as environmental destructors, culturally incapable of envisioning wetlands as anything other than waste. "Many European land drainage practices," one leading wetland book states, "were imported to the Americas. While much of the Old World landscape had been drained for centuries … wetland loss has occurred in the New World mostly since the mid 19th century. Before then, wetlands were usually viewed with disdain and fear." A more nuanced reading of the primary sources, however, tells a different story – one where humans have not always been culturally pre disposed to destroy wetlands. As interdisciplinary re search has gained momentum, a new interpretation has emerged illustrating that indigenous peoples understood and practiced long term sustained coexistence with wetland environments. Contrasting the indigenous Maori of New Zealand with the first Europeans to ascend the Thames River in 1769 and see the extensive wetlands of the Hauraki Plains, it is now understood that what was for nearly a millennia a *tapu* (sacred) food rich region of wetland forest and swamp to the Maori was but an uninhabited wilderness to Europeans. Harking back to John Locke's labor theory of value, this clash in cultural perceptions of the landscape is explained as a result of the European conviction that land "'left wholly to nature, that hath no improvement of Pasturage, Tillage, or Planting,' is called … waste … Land as patently fertile as this, was predestined to be cultivated – or forfeited."

Thanks to such revisionary geography and history, a clearer picture is emerging that indigenous coexistence with wetlands was the norm in the New Worlds. Floral and faunal exploitation took place over hundreds and thousands of years with limited negative impacts on the environment. Research suggests that limited demographic pressure, simple technology, transhumance, a subsistence oriented economy, and the use of religious restrictions are all factors explaining the ability of indigenous peoples to practice long term coexistence with wetlands. New research is illustrating that the early inhabitants of the Old World may also have had a more benign relationship with wetland environments than heretofore recognized. Some 5000 years before the present (BP), in ancient Mesopotamia, the lives of marsh dwellers on the Tigris and Euphrates Rivers were characterized by reliance on the renewable resources of the marsh – hunting waterfowl, fishing, building homes, and woven materials from reeds (*Phragmites*), and pasturing water buffalo. In northwestern Europe around 2500 BP, the first inhabitants of the peatlands and salt marshes between Belgium and Denmark gleaned their subsistence from hunting, fishing, pasturage, and exploiting the flora. To escape periodic flooding, they built massive dwelling mounds in the wetlands called *terpen* and *wierden*.

Postmodern studies make much of the lexicon of swampspeak: words like 'dismal', 'dreary', 'desolate', 'gloomy', 'miasma', and 'disease ridden' are so often used to describe wetlands in literature. What becomes clear is that while there is some validity in the application of such words to wetlands, they are more often derived from the secondary sources of individuals who had little direct (territorial) contact or understanding of wetland ecosystems. Primary sources, on the other hand, written by observers enjoying frequent contact with wetlands tell a different story. Throughout the Medieval period from the fall of Rome to the sixteenth century in the Gulf of Poitou in France, and the Fenlands of England, there were hundreds of villages where the law treated wetlands as common land and natural resources were shared. Drawing on primary sources, historical geographers have documented centuries of sustained wetland exploitation based on a diversity of products: fishing (especially eels), waterfowl hunting, pasturing of domestic animals on marsh grasses (*Juncus* sp.), reed cutting (*Phragmites*), peat harvesting for fuel (*Sphagnum*), and salt production. Such communities were highly independent and wary of outside interference even into the nineteenth century, as suggested by P. H. Emerson's and T. F. Goodall's photographic documentary *Life and Landscapes on the Norfolk Broads* (1882), and more recently in Graham Swift's *Waterland* (1992).

European expansion overseas in the preindustrial period (before 1850) brought British, Dutch, French, Germanic, and Swedish colonists from northwest Europe into direct contact with wetland environments in the New World. A comparative analysis of Northeastern North America (eastern Canada, coastal New England, the Middle Atlantic) illustrates that the natural grasses, fish, and waterfowl of the salt marshes, brackish marshes, and flooded river meadows, as in the Old World, played an important role in the colonial subsistence economies of Dutch, English, French, Germanic, Irish, Scottish, and Swedish settlers from the 1600s until industrialization in 1850. Before dense colonial forest could be sufficiently cleared for agricultural production, the fish and waterfowl trapped and hunted in the wetlands of the Northeast were viewed as manna from heaven, while domestic animals from the 1600s to the 1940s were pastured and fed marsh and dune grasses (*Spartina patens*, *Juncus* sp., *Scirpus* sp.). It is this sense of coexistence with the wetland landscape of New England that comes to the fore in Sarah Orne Jewett's *A Marsh Island* (1885). During the first two and a half centuries of colonization in the Northeast, marsh hay harvests were of vital importance to the overall grassland economy, and as late as the 1890s, marsh grasses were the focus of study by agricultural colleges to determine their influence on taste (butter and cheese products) and nutritional value (from 10 to 20% less nutritious than upland hay).

Reclamation

Wetland coexistence in the Old and New Worlds was possible as long as demographic pressure remained low, and economic production was local and oriented toward reproduction of the family. The earliest deliberate reclamation of wetland ecosystems for human needs began around 4000 BP in the mega river basins and deltas of the monsoon lands of Asia, where elaborate systems of water control emerged around wet rice cultivation. Slightly later, the establishment of sedentary farmers in Mesopotamia led to the emergence of flood control technology (dikes and drains) for economic reasons, followed by similar developments in the Nile Valley and its Delta. All over the Mediterranean basin, sophisticated engineering plans to dike, drain, and transform wetlands into dry agricultural land for burgeoning populations appears in the earliest Roman records – the Pomptine Marshes, the Po basin, the lower Rhine, and the fenlands of England.

Little is known about wetland management after the fall of Rome. Religious ascetics established monasteries in the wetlands, and Beowulf mentions them as places of refuge for outlaws. After the year AD 1000, because of rising demographic pressure, farmers in the Netherlands began digging open ditches to transform the peat bogs into agricultural land. Gravity drains in fan like and feather like patterns were dug to drain raised peat domes toward naturally flowing rivers and purposely dug canals. By AD 1300, almost all the fenlands in Holland had been transformed into a densely populated agricultural region. At the same time, local water management organizations emerged to establish control over maintenance and operation of dams, sluices, and drainage canals. Lower groundwater levels, however, brought about land subsidence as the underlying peat shrunk, consolidated, and oxidized. When combined with rising sea levels, subsidence (up to 4 m) made gravity drainage insufficient. Huge amounts of land were lost as they reflooded, and not until the introduction of wind powered pumps (windmills) after 1400 did reclamation of the lost lands begin. It is suggested that this process of reclaiming wetlands a second time gave rise to the verb ‘to reclaim’, now synonymous with initial and secondary wetland reclamation.

The Dutch experience is representative of the immense European expansion of agricultural land through forest clearance and wetland reclamation in the High Middle Ages (AD 1200–1350) as demographic pressure mounted and land hungry monasteries and royal treasuries looked to forests and wetlands as potential land for reclamation. In contrast to those who coexisted with wetlands, reclamation was viewed by the nobility and clergy as a means of enlarging the human ecumene. In the wetlands of France’s Gulf of Poitou, and in the English Fenlands, reclamation was often directed on an ‘ad hoc’ basis by monasteries granted wetlands by the nobility in return for drainage and sharing the resulting revenues. The monasteries organized sufficient local workers (using simple hand tools and draft animals) to build dikes, and dig drains and canals that relied on gravity. This work was coupled with dams and tidally and manually controlled sluice gates at river junctures to control water movement seaward.

Where did this pre 1500 reclamation technology originate? Was it locally developed, diffused from places like Holland, or a holdover from ancient Roman? Some suggest that the origins of certain technology, like the tidal sluice gate, can be traced back to Rome. The salt producing basins constructed in the tidal wetlands of the Gulf of Poitou (*marais salants*) appear to have Roman origins too. Windmills, while not a Dutch invention, were first used by the Dutch for draining in 1408 and then diffused to England, France, and elsewhere. Other techniques, like open drains and canals may be part of a larger shared European maritime culture. What is clear is that with the end of religious wars, the consolidation of nation states, European overseas commercial expansion, and the emergence of mapping, and later the collection of statistical information, large and often sophisticated state and capitalist supported movements to modernize agriculture began sweeping northwest Europe after 1500. The result was the birth of hundreds of small and large scale wetland reclamation projects intent on a higher intensity of land use.

This economic and ecological transformation of wetlands was met with resistance, especially by those dispossessed of access to common wetland resources. When the English monarch James I announced his intention to support draining the Fens, and Dutch engineers like Cornelius Vermuyden were invited to consider the matter, he declared that the honor of his kingdom “would not suffer any longer … to bee abandoned to the will of the Waters, nor let it lye waste and unprofitable.” In response, Fen dwellers (as would occur elsewhere on the Continent in a like manner) testified before court officials that from time immemorial their ancestors had gleaned their livelihood from the seasonal rhythm of the wetlands. Dikes were torn down and sluice gates sabotaged in protest, but to little avail. Like the enclosure of commons everywhere, wetlands were being incorporated into the system of modern intense agriculture.

There were two forms of enclosure: that of the well known open fields through parliamentary acts to create more compact farms that were easier and cheaper to work, and the less well known enclosure and reclamation of the commons and waste. In both cases, the aim was to convert the countryside to more intense and profitable use. Between 1780 and 1880, about 2 million ha of forest, marsh, and seasonally utilized land was enclosed, drained, or otherwise reclaimed in Britain – followed slightly later in continental Europe. As drained peatlands subsided, gravity drains were replaced by windmills, which in turn evolved from using scoop wheels to more efficient Archimedean screws. When these technological advances no longer sufficed, reclamationists turned to steam pumps in the nineteenth century, to diesel pumps in the 1920s, and finally to electric pumps in the 1960s. With each technological advance, combined with new techniques like underground drainage, additional lands were reclaimed until one visitor to the English Fenlands in 1890 could remark how the diverse wetland ecology had been radically simplified to serve the purposes of intense agriculture: “all is gone – reeds, sedges, the glittering water, the butterflies, the gypsies, the bitterns, the wild fowl, and in its place … a dreary flat of black arable land, with hardly a jack snipe to give it charm.”

The timing of reclamation in the New World would begin several centuries later than in Europe. While there are examples of early and quite sophisticated wetland reclamation, such as the Acadian and Swedish diking on the Bay of Fundy and Delaware Bay as early as the 1640s and 1650s, and some reclamation around burgeoning colonial cities like Boston, New York, Philadelphia, and Baltimore, coexistence with wetlands was more com monly the norm until forest clearance reached its max imum on farms around the mid nineteenth century, and the production of hay became the dominant crop to be channeled, no longer into family subsistence activities, but market production of dairy goods for nascent cities and overseas consumers. By mid nineteenth century, the seasonal exploitation of wetlands was coming to be viewed as a supplemental practice, no longer vital for subsistence. More importantly, growing demographic pressure and the agricultural revolution in production begun in Europe in the 1780s was being felt in North America by the early decades of the nineteenth century, pushing and pulling premodern farmers into the emer ging modern, global market place.

Mixed extensive farming disappeared, replaced by intensive, market oriented production. In 1850 the US government would pass the Swamp Lands Act, granting wetlands to the individual states as a means to promote reclamation. In his *Tide Marshes of the United States*, Nesbit (1885) wrote a reappraisal of this changing perception of wetlands: "we are coming rapidly to a time when other conditions will prevail … The day of unlimited cheap lands is fast passing. Unless our growth meets a serious check there will soon be little available farm land… [necessitating] the reclamation of millions of acres of rich marsh land." Once the modernist vision of wetland values took hold among elite improvers, common farmers on the East, South and West coasts, and on the prairies, followed and wetland loss quickly escalated. Limited diking and draining in the 1850s using horse drawn equipment gave way to mega dikes, drains, and canals constructed with steam and mechanical shovels and financed through massive government subsidies until the 1930s. Following Word War II, even more extensive wetland reclamation took place for highway construction and urban expansion.

This Western model of wetland reclamation was dif fused to Africa and Asia, especially after 1945, where the swampy grasslands at the base of the Himalayas to the estuaries of Thailand, Cambodia, and Vietnam were significantly reduced in size because of population growth and pressure to raise more crops. As part of China's Cultural Revolution in the 1950s, Mao Tse Tung instituted a national program to reclaim China's 65 million ha of wetland (about 10% of the world's total) to increase production of wheat and other crops. But changing demography, economy, and technology are not the sole reasons for the loss of wetlands at this time. Extensive damming of the Tigris Euphrates headwaters by Turkey and Syria beginning in the 1950s for water and hydroelectric development had a negative impact on the 15 000 km^2 Mesopotamian marshlands of Iraq and Iran. Nevertheless, the loss of nearly 90% of those same marshes came about only in the 1990s, when Saddam Hussein ordered their reclamation as political retribution against the 300 000 Marsh Arabs who had risen up against his rule.

Preservation and Restoration

Not until the late 1960s, at the juncture of the industrial and postindustrial eras, did books like *Life and Death of the Salt Marsh*, and international accords like the Ramsar Convention on Wetlands (1971), raise questions re garding the reclamation of wetlands and the need to preserve them, not only in intellectual circles, but also among average citizens. Reevaluating and challenging the easily quantified economic benefits of reclamation with the much less visible benefits of preserving wetland ecosystem functions, biodiversity, and water functions, was contentious and inspired heated debate between technocrats and ecologists over their differing visions of the value of wetlands. For example, during the 1970s, the US Army Corps of Engineers and the Bureau of Rec lamation all encouraged wetland reclamation, while simultaneously the Department of Interior (particularly the US Fish and Wildlife Service) encouraged their protection. Only in the 1980s, as environmental edu cation and public recognition of the value of wetlands in their natural state mounted, were government policies like 'No net loss' in the US, and similar policies in Canada and Europe, adopted to provide comprehensive strategies for wetland preservation.

Since the 1990s, postmodern thinking has questioned the wisdom of past reclamation and its overall economic, social, and environmental benefits to society. As water scarcity, wildlife disappearance, and climate change be come increasingly contentious political issues, the breaching of dikes and the reflooding of former wetlands has become an increasingly popular option. In what has been called one of the world's largest tidal marsh res toration projects, ecologists in the late 1990s began breaching the dikes of former salt marsh farms on the Delaware Bay as part of a mitigation project to offset nekton (krill, small fish) loss through cooling at a nearby power plant. Breaching dikes and reflooding former wetlands at Abbots Hall Farm in Britain has been pur sued with the goal of demonstrating how farming and nature restoration can work side by side while reducing the risk of flooding and improving coastal landscape conservation. In China, decades of reclamation have re sulted in half of the country's mudflats disappearing, draughts where wetlands used to conserve water, in creased erosion and flooding previously held in check by wetland vegetation, and worsened water pollution. Losses of these ecological benefits from the destruction of wet lands have recently been calculated at 4–8% of China's

GNP. As a result, since 2000, the Chinese government has implemented an action plan to protect China's re maining wetlands and begin restoring key wetlands.

As with the many changes that have occurred over the past 1000 years in human perception and interaction with wetlands, current efforts to achieve preservation, miti gation, and restoration of wetlands are all indicative of a societal turn toward a more holistic understanding of the human place in the environment. In an age when the discourse of sustainable development dominates gov ernment agendas, this new approach to wetlands can be interpreted as the rediscovery of aspects of the coexist ence with wetlands that once existed before modern ization reoriented and distanced intimate knowledge of the impact of human activities on wetland environments toward a much less aware state where the health, and even the very existence of the wetlands, was secondary to issues of intense production and consumption.

See also: Agrarian Transformations; Conservation and Ecology; Environment, Historical Geography of; Field Systems and Enclosure; Nature, Historical Geographies of; Sustainable Development; Technological Change; Territory and Territoriality; Water Management.

Further Reading

Baron Yellès, N. and Goeldner Gianella, L. (2001). *Les marais maritimes d'Europe atlantique*. Paris: Presses Universitaires de France.

Bleakney, J. S. (2004). *Sods, Soil, and Spades: The Acadians at Grand Pré and Their Dykeland Legacy*. Montreal, QC: McGill Queen's University Press.

Casagrande, D. G. (1997). *Restoration of an Urban Salt Marsh: An Interdisciplinary Approach*. New Haven, CT: Yale University Press.

Danner, H. S., Renes, J., Toussaint, B., van de Ven, G. P. and Zeiler, D. (eds.) (2005). *Polder Pioneers: The Influence of Dutch Engineers on Water Management in Europe, 1600 2000*. Utrecht: Labor Grafimedia.

Darby, H. C. (1940). *The Medieval Fenland*. Cambridge: Cambridge University Press.

Giblett, R. J. (1996). *Postmodern Wetlands: Culture, History, Ecology*. Edinburgh: Edinburgh University Press.

Hatvany, M. G. (2003). *Marshlands: Four Centuries of Environmental Change on the Shores of the St. Lawrence*. Sainte Foy, QC: Les Presses de l'Université Laval.

Hoeksema, R. J. (2006). *Designed for Dry Feet: Flood Protection and Land Reclamation in the Netherlands*. Reston, VA: ASCE Press.

Meindl, C. (2003). Finding our way into swamps. *The North American Geographer* 5(1 2), 154 161.

Mitsch, W. J. (1994). *Global Wetlands: Old World and New*. New York: Elsevier.

Park, G. (1995). *Nga Uruora/the Groves of Life: Ecology and History in a New Zealand Landscape*. Wellington: Victoria University Press.

Prince, H. (1997). *Wetlands of the American Midwest: A Historical Geography of Changing Attitudes*. Chicago, IL: University of Chicago Press.

Rippon, S. (2000). *The Transformation of Coastal Wetlands: Exploitation and Management of Marshland Landscapes in Northwest Europe during the Roman and Medieval Periods*. Oxford: Oxford University Press.

Sullivan, R. (1999). *The Meadowlands: Wilderness Adventures on the Edge of a City*. New York: Doubleday.

Swift, G. (1992). *Waterland*. London: Pan Books.

Relevant Websites

http://www.essexwt.org.uk
Abbotts Hall Farm, Essex Wildlife Trust.

http://www.ducks.org
Ducks Unlimited (North America).

http://www.edenagain.org
Eden Again (Iraq).

http://www.ramsar.org
Ramsar Convention on Wetlands.

http://www.wicken.org.uk
Wicken Fen Vision (Britain).

Whiteness

L. Peake, York University, Toronto, ON, Canada

Glossary

Critical Race Theory Refers to a Western school of thought that emphasizes the socially constructed nature of race and the necessity of antiracism. Scholars generally explore the links between racial difference and the law, labor movements, education, literary canons, political systems, history, and identity.

Miscegenation The practice of sexual relations between people of different ethnic groups and different skin colors that leads to the birth of children. Its usage in English, at least, has a pejorative tone.

New Racism Subtle forms of racism that have become increasingly common since the late twentieth century, in which the focus is on emphasizing cultural differences rather than biological hierarchies of race.

Introduction

What Is Whiteness?

Whiteness is a somewhat difficult concept to explain, with some according it the status of a meta narrative, as the defining condition of being human, and others denying its existence. How can there be such widely divergent views? Answers to this question may be forthcoming by asking, 'what kind of category is whiteness?' Whiteness, as many antiracists have argued, like all categories, may be an invention, but it is still a cultural construction; in other words, it is a way of being in the world, a set of cultural practices to be drawn upon, often not named as 'white' by white people, but looked upon instead as normal or natural. Positioned as the overarching (and invisibilized) norm from which all other racialized categories are perceived and understood, whiteness operates as a position of structural advantage and privilege and as an epistemological 'standpoint' or place from which to look at oneself, others, and society. As numerous scholars have noted, this epistemological supremacy of whiteness, in many contexts, is signaled not only by discourses of normalcy, but also those of cleanliness, beauty, freedom, purity, and morality.

Such discourses are ideologically associated with bodies that, in certain times and places, are considered white, assigning them political, economic, social, and psychological privileges. Whiteness, after all, draws its power from the geographically and historically different meanings ascribed primarily to skin color but also to facial type and color and texture of hair. A further appreciation afforded by the embodiedness of whiteness is the understanding that it is always produced and performed not only in relation to race and ethnicity (specifically those who are visibly non white), but also to class, gender, sexuality, and other aspects of bodily identity. Thus, whiteness can be understood as a historically and geographically specific social construct that has been accorded a privileged position at the center of racialized and embodied systems of social stratification and differentiation. Its contextual specificity though indicates that whiteness is never a stable bodily signifier; rather, the social meanings it conveys are always contingent. But this is not to deny that it has real effects on people's lives – white skin privilege is the other side of racism – and it is invariably deployed to augment oppressive relations of power.

There have been some serious dissensions to this view of whiteness, especially from right wing theorists and white supremacists who argue that race, and the notion of a white race, has a basis in biology, both in terms of phenotype and genotype. Indeed, scientists have claimed to identify ostensibly biological and discrete categories of races such as Caucasian (read 'white'). Whereas such categories of race are now critically exposed for their racializing politics, they still have their proponents and have been taken as real and natural by millions of people across the globe. The consequences of these views are self evident, from the Nazi inspired notion of the Aryan nation in Europe, the system of apartheid in South Africa, to the Ku Klux Klan in the United States, from physical extermination and slavery to systems of advantage to white people in everyday life (such as access to housing, education, and employment, especially to positions of power), and to life expectancy.

Critics of the notion of discrete races of whites and people of color have existed since the 1800s. They have argued that genetically distinct races cannot be recognized and that the different ways in which people live their lives do not correspond to racialized differences, but rather to societal distinctions. Moreover, it is the contingency with which phenotypical features such as skin color, hair type, and facial physiognomy have been used to indicate differences between peoples that makes the idea of races untenable. In the present day, these tenets are taken for granted by academics in the fields of social science and the humanities. Beyond this, however, it becomes complicated.

Calls for the transcendence of race have come from post race theorists, such as Paul Gilroy, who argue for a 'post racial humanism'. While they agree that processes of racialization (through which categories of race are constructed) and racism (the outcome of acceptance of racial categories) are discursive and material realities, race is not and serves only to divide people, thus having no useful purpose, particularly in terms of liberty, global humanism, and collectivity. But if the cultural reality of race is to be dismissed, then what is it about race that exists? Indeed the denial of race is also a strategy of the New Racism, in which racism is expressed in populist euphemisms about cultural differences rather than in crudely overt themes about biologically distinct races. Arguments for the denial of post racial humanism are emerging from proponents of the material turn, taking place in the social sciences, who argue that there is increasing recognition that social construction *per se* is not enough to understand the machinations of race and that the materiality of racial categories has been disavowed. Physical differences, even if only surficial, do exist and race, some claim, may best be understood as 'culturally embedded phenotype', although concerns about this claim are also arising. Arguments about something being only socially constructed are indeed essentialist because they argue for a different material reality that is divorced from the material. So, the significance of the material turn is important, but not if it is employed to create a similar divide between the material and the nonmaterial, in other words, just flipping the coin around to reflect a new form of essentialism.

Origins of the Term Whiteness, Whiteness Studies, and Interests in Whiteness in the Discipline of Geography

The concept of whiteness cannot be divorced from its origins in modernism; it is generally agreed that the concept of whiteness entered the English lexicon in the seventeenth century as a result of colonialism and the need for West European nations to distinguish themselves from the colonized and, most especially, slaves. Whiteness quickly became a legal category, encoded in a variety of laws and conferring differential status. Colonial rule not only comprised stripping colonies of their resources and the enslavement of indigenous groups as well as the development of the slave mode of production, it also produced ways of knowing and being of the colonized that still persist today. This 'epistemic violence' rationalized colonial domination from the standpoint of the West while also formulating the colonized as irreducibly other. Thus, the discursive construction of a range of colonized cultural and racial others also led to the production of the supposedly stable and unmarked white West European self. In the United States and Canada, the construction of whiteness was an effort to distance slave owners from slaves. By the eighteenth century, whiteness had become well established as a racial term. And the nineteenth century in the United States, during the post reconstruction era known as 'Jim Crow' (whereby whites could legally separate themselves from blacks), was the period in which the cultural hegemony of whiteness was secured.

Radical black intellectuals, descendants of slaves and freed people from the Caribbean and the United States, such as Frantz Fanon, Edouard Glissant, C. L. R. James, W. E. B. Du Bois, James Baldwin, Frederick Douglass, Richard Wright, Ralph Ellison, Wilson Harris, and Sylvia Winter, began effectively writing about this white Western self in the late nineteenth century, long before the emergence of Whiteness Studies. Some of the key ideas that they, and more current Critical Race theorists, have alerted us to are that whiteness is a taken for granted category; that whiteness is socially and culturally constructed; that whiteness is, like other racial categories, differentially produced according to time and space, and therefore, never a stable sign or identity; that whiteness is a relational marker – it needs the 'other' to make sense and to become 'taken for granted' and invisible; that the normalization of whiteness makes its effects difficult to see; and that, while normalized and seemingly invisible, whiteness impacts upon how non white people understand themselves.

Whiteness Studies, popular mostly in North America and the UK, is a controversial field of study because it teeters on the brink of putting white people back at the center at the expense of non white people and hence of maintaining white privilege. As a result of this controversy, some have argued that the focus of attention should not be whiteness *per se*, but rather antiracism. Others argue that centering 'whiteness' reproduces the very conditions that produced 'blackness', and that it seeks to deconstruct. While Whiteness Studies started to develop in the early 1980s the field is still in a nascent stage with its presence being extremely uneven within universities and across disciplines. **Table 1** highlights the major topics that have been investigated by theorists of Whiteness Studies. The field is closely associated with Critical Race Theory, and inherits from it a focus on the legal and historical construction of white identity.

The emergence of the study of whiteness in Geography, especially European Geography, cannot be separated from the discipline's origins in the building of empires and imperialist projects, and *inter alia*, renderings of 'otherness', race, racialization, and racism. European colonial and imperial geographies of the late nineteenth and twentieth centuries led to sporadic studies by white male British geographers in the early twentieth century into, for example, the effects of tropical climates on white skin. Such studies formed part of a broader research agenda on the anxieties invoked by the increasing awareness of the limits of white rule and authority over

Table 1 Topics in Whiteness Studies

White supremacy (and its relationship to globalization), white privilege, racism
The historical processes through which white racial identities have been created by processes of segregation economic, social, political, and always spatial, for example, slavery, colonial settlement, apartheid, citizenship, and industrial labor
The role of modernism, the Enlightenment, Orientalism and the trans-Atlantic in producing whiteness
White ethnicities; how certain groups have come to be considered white
The nature of white identity, and increasingly the problematization, or unpacking, of the concept of whiteness itself; how whiteness has to be theorized in a racialized context in which white identities are privileged over others
Possible processes of social change in white identities

colonized nations. Apart from these studies, however, whiteness did not receive any critical attention from geographers until the early 1990s. This is surprising given the interest shown by geographers in the work of Fanon and in Orientalism and Said's work. In part, current interest stems from the reflexive turn in the social sciences in which dominant categories are being critically reexamined. As the norms against which 'difference' was identified came under investigation, there has been increasing interest in whiteness, especially in the context of Critical Race Theory. But this is not to imply that studies of whiteness have been wholeheartedly embraced by geographers; uptake has been slow and sporadic. It has been the last of the so called meta narratives to be deconstructed as studies of class, sexuality, and gender still continue to occupy the interests of most critical geographers. This is disappointing, but hardly surprising given the overwhelming whiteness of practitioners of Geography in the Western academy, leading most recently to critiques of the whiteness of the discipline of Geography itself and to insisting on the necessity of increasing the diversity of racialized identities among university teachers.

The remainder of this entry turns its attention to the major contributions that geographical analyses have made to studies of whiteness including historical geographies of whiteness, the role of whiteness in discourses of nation, whiteness and the production of racialized landscapes, the idea of embodied whiteness, and processes of social change in white identities.

Historical Geographies of Whiteness

Geographers, not surprisingly, have emphasized the need to study how whiteness has been constructed across time and space. Whiteness, despite its ubiquity, cannot be theorized as a universal category that operates in a way that is transhistorical and aspatial. Rather, its investigation is an 'empirical question', in that its study must be specifically positioned in the geographic and social systems of racialization within which it operates and through which it is produced. Hence, one area of research that has received attention is that of uncovering historical geographies of whiteness.

Most research on historical geographies of whiteness has centered on the United States, where it is claimed that racialization needs to be understood from a normative and relational position of whiteness that situates non white people in and out of place. Whiteness has been positioned as a tightly policed dualism between African Americans and Americans of European origins although, increasingly, studies are addressing the ways in which groups such as Hispanic Americans and newer immigrant groups are differentially placed in relation to, and are disrupting, this binary racialized divide. This has been illustrated by studies of the large scale non European immigration to cities in the United States in the twentieth century that have revealed not only increasing economic prosperity for the earlier (white) arrivals, but also that this upward mobility has reduced intra white differences, as white ethnic differences became blurred (although not fully erased), leading to increased intermarriage and a reduction in residential segregation between different white ethnicities. However, African Americans have not benefited from this process; rather, other lighter skinned newer arrivals have been the beneficiaries in terms of reduced levels of residential segregation and job opportunities.

The geographical contingency of whiteness has also been investigated in the United States through the phenomenon of 'passing', in which a non white subject presents her or himself as white. The performance of passing – often understood as a form of escape by its practitioners – highlights the anxiety around threats to the purity of whiteness. Passing is not only about the metaphorical passage from a subordinated identity to one of privilege and status; it also involves geographical movement through material environments, from ones where one's 'true' identity is known to one where it is not.

Increasingly, geographers are also problematizing whiteness so that it is not only associated with a white, Western racialized ethnicity. Hence, recent work has started to assess the impact of whiteness outside Anglo America, specifically investigating the emergence of a white modernity in non Western societies in the twentieth century. It has been maintained, for example, that the development of white identities in Latin America cannot be conceptualized as simply reflecting the imposition of 'Western values' on 'non Western societies'. Studies in Latin America and the Caribbean have shown that, unlike the racialized binary of the United States, there is a fluidity to racial formations based on relational notions of the darkness/lightness of skin color and hair type. Despite these differences between systems of racialization in the United States and Latin America, a white norm dominates them. In Latin America, the desire for whiteness – as the way to secure social and economic advancement – has been thought to be a product of the disciplining discourse of whiteness that functions by in stilling a sense of inferiority within its 'other', what has been referred to as a 'colonization of the mind'.

The Role of Whiteness in Discourses of Nation

Throughout the twentieth century, whiteness has been reified and normalized as a racial and cultural identity. Its durability is a function of its close relationship with modernity, in that the 'white race' has been identified as its physical crucible. Whiteness Studies in the Australian context, for example, have examined the ways in which indigenous Australians were marginalized in the wake of European colonization, as whiteness came to be defined as central to Australian identity. From the late nineteenth century until 1973, the Government of Australia re stricted all permanent immigration to the country by non Europeans under its White Australia policy. Immi gration inspectors were empowered, for example, to ask immigrants to take dictation from any European lan guage as a test for admittance, a test used in practice to exclude people from Asia. Similar processes of en couraging white immigration also operated in Latin America. Currently in non Western countries, it is not the physical presence of white people that is encouraged, but rather ways of becoming white, exemplified through the close association between consumer identities and white identities. Whiteness is being codified as a lifestyle with the white body – white skin, straight blonde hair, and blue eyes – being the ideal to be emulated. Who you are – the national identity you take on – not only de pends upon where you are but increasingly on who you look like. Thus, whiteness is increasingly being per formed by non white bodies enacting themselves or passing themselves as white, or what can be seen as a 'colonization of bodies'.

The study of the role of whiteness in nation making has been conducted mainly in white Western countries, including the white settler nations of Canada and Aus tralia. The conflation of whiteness and nation has taken place through nationhood being tied to blood purity, such that citizenship has been demarcated by those with un tainted blood, or the absence of the 'other'. Scholars have thus analyzed whiteness as a gendered category that is inflected by discourses of nation and culture – it is women's bodies that bear the burden of reproducing the nation. They have also shown the ways in which white women at times have been complicit with colonial and imperial practices across the globe. However, studies are now taking whiteness beyond an identification solely with white (Anglo) ethnicities, for example, juxtaposing whiteness with the racialized identities of indigenous peoples and considering the role played by whiteness in the post /neocolonial contexts within which indigenous struggles are taking place.

Geographers have also engaged in studies of nation ally excluded groups of whites, showing how they (e.g., the Irish and Italians) have moved in and out of whiteness (and the nation and the polity). The origins of the term 'white trash' may come from a racially segregated past, but modern usage of the term places emphasis on the word 'trash' or the labeling of certain categories of whites as socially worthless and who undermine discourses of natural white superiority. For example, it has been shown how rounds of welfare restructuring have resulted in the white working class in Britain being differentiated into the respectable working class that aims to improve itself as distinct from the white underclass; the latter group is racialized by its association with the black 'ghetto' underclass, and is seen as socially excluded and at odds with the image of a progressive British identity seeking to recover its economic and political leadership status in a rapidly changing world. Constructions of whiteness, at the scale of the nation state, then speak not only to dif ferent varieties of whiteness but also to a politics of racialized oppression.

Oppressive, essentialist, and homogenous versions of the nation are on the rise in Western Europe and the United States, as is the linking of far right groups with a white identity and white nationalism in Australia. In the UK, the current backlash against asylum seekers and immigration also reveals the persistence of racist dis courses of national belonging. Geographers have sought to show how these discourses have played out in official/ institutional definitions of whiteness, and how these have changed over the years, often resulting in very little correlation between census categories and current na tionalist discourses of whiteness. For example, the US Census currently defines white people as those having origins in any of the original peoples of Europe, the Middle East, or North Africa, while in the UK white usually refers only to people of European origin, and in Australia the census does not record ethnic or racial origin.

Whiteness and the Production of Racialized Landscapes

Given that we live in an increasingly racialized world we might think of certain places as being imbued with whiteness – the golf course, the boardroom, even whole countries (e.g., the portrayal of Canada as the Great White North, Denmark, Sweden). Yet white places, like white people, have largely escaped the geographical gaze.

This is slowly beginning to change and geographers have started to investigate the ways in which whiteness is imbricated in the production of racialized landscapes (**Figure 1**).

White people (but not all of them) have been shown to have the privilege of access to safe neighborhoods with 'good' schools and services. Studies have shown, for example, how whites in the United States have gained access to cleaner urban environments by moving away from industrialized areas and into suburbs. Thus, processes of suburbanization and decentralization are instances of white privilege, revealing how the virtually total freedom of movement of white people is predicated on the spatial entrapment of non white people. This differential mobility can be institutionalized, as in landscapes of apartheid, or naturalized, as in journeys to work or to homes in the suburbs, producing spatialities through which whiteness and white privilege are maintained and rendered as natural. Hence, geographers are adding to the understanding of landscapes as a dominant way in which white power is played out.

Landscape imagery, cultural and natural, iconic and vernacular, has also been shown to be evoked in the services of racialist discourses of belonging, constructing who is in and out of place. Research has shown how commonplace images, embedded in the vernacular landscape– such as billboards – and the realm of the visual – soap shows, TV adverts – reflect a nation's desire for whiteness. Studies of specific places have also been employed to show how landscape, as a cultural practice, can anchor white collective memories. For example, rural states such as Vermont have been represented as iconic white spaces in the United States, geographical imaginary, associated with the images of Norman Rockwell, the poetry of Robert Frost, and the figure of the Yankee, enacting white values, mores, and esthetic tastes.

Embodying Whiteness

Studies of white embodiment are a legacy of both the embodied politics of feminism and the writings of Frantz Fanon who, *inter alia*, demonstrated that whiteness was most effective when not only white people but also non white people embodied whiteness. Geographers have only recently started to engage in studies of white embodiment showing how skin color and the form of the body mark who belongs to the polity and the nation. Studies have shown how modern citizenship, while ostensibly universal, has embodied subjects, always raced and gendered, and how through various acts of inclusion and exclusion certain types of bodies come to represent certain types of subjects. Most recently, studies have analyzed representations and bodily performances of white identities, and have also portrayed how bodies are able to become white. As mentioned above, because race is bound to nation by notions of racialized reproduction, women's bodies act as markers of the limits of race and nation. Research in Latin America has shown that racially marked women are likely to claim whiteness to enhance their marriage prospects, to give their children more opportunities in life, and to retract unwanted sexual advances; while indigenous and black women are viewed as promiscuous and thus, vulnerable to sexual predators, white women are considered virtuous.

Figure 1 Scenery from the Second City TV program (photograph by Dave Prescott).

Traditionally, interest in the phenomenon of claiming white embodiment originated in studies of racial passing or miscegenation. In Latin America, for example, white *mestizos*, that is, people who may self identify as white, but who are located within nationalist ideologies of racial and cultural mixing, are popularly perceived to be gradually evolving from indigenous identities into more 'civilized' (read 'white') bodies. In other words, misce genation is not so much about mixing as it is about a progressive whitening of the population, referred to as the process of *blanqueamiento*. More recently, studies have come to engage with visual political economies of ap pearance. In Latin America, the ways in which chemical processes and physical practices such as skin lightening, hair straightening, and wearing of Western styles of clothing have been employed in white identity formation have been emphasized, whereas in the South Asian context plastic surgery to change eye shape is being in vestigated (**Figure 2**).

Other geographers have called for studies of bodily experiences of spatialities of whiteness. In other words, how does whiteness give meanings to places and how does it affect the embodied experiences of white and non white people moving through non white and white places? Interrogations of how bodies identified as white perform such identities can demonstrate how whiteness is reproduced and presumably stabilized as a category.

Processes of Social Change in White Identities

Deconstructing whiteness is not just about confronting the geographies it produces; it is also, as mentioned earlier, about the discipline of Geography itself. The psychological and emotional energy expended by non white faculty in all white environments testifies to the fact that whiteness cannot be just another field of aca demic endeavor. It is about real bodies in real places and it is about an antiracist politics whose effects must be named. Whiteness Studies have investigated the extent to which white identities can be situated within progressive antiracist politics. For example, contributors to the US based journal *Race Traitor* (a journal devoted to the 'abolition of whiteness') argue that the effects of white privilege are kept in place by those who benefit from passively accepting its systemic advantages. Only iden tities that seek to transcend or defy that privilege, they argue, are effectively antiracist and thus, "treason to whiteness is loyalty to humanity."

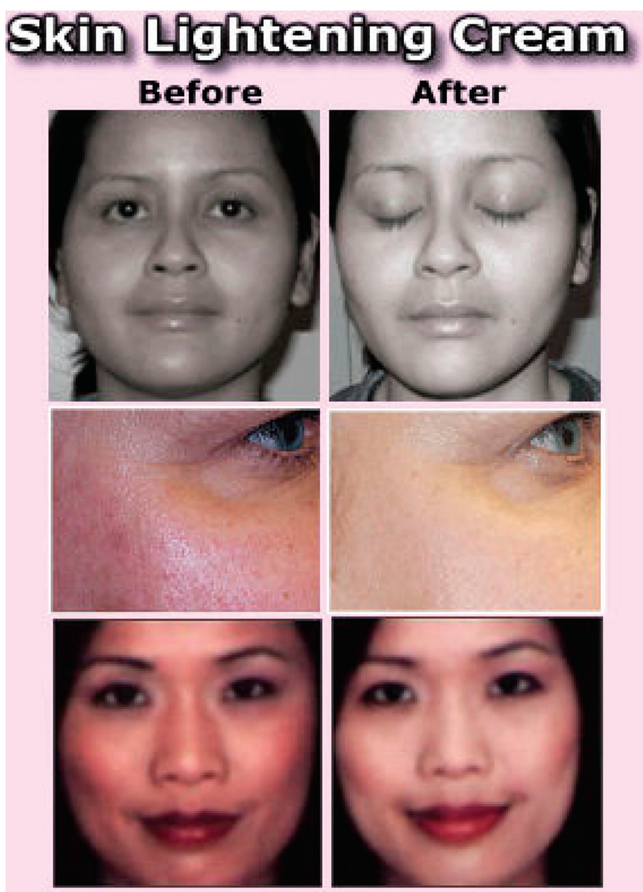

Figure 2 Adverts for skin lightening cream (taken from eBay).

Recently, geographers – both white and non white – have started to examine the implications of white privilege and of the white norms that inform the practices of the discipline of Geography itself, which, in the Anglo American sphere, has been referred to as being nearly as white an institution as the Supreme Court of the United States. Whiteness oppresses when it operates as the invisible regime of normality, and thus making whiteness visible is a principal goal of antiracist pedagogy. Critiques of the whiteness of the discipline reveal not only its skewed intellectual production but also its institutional racism. It has been argued that the discipline of Geography has deeply entrenched biases in favor of the white majority in evaluation procedures, curricula, and power relations. Moreover, given that the vast majority of its practitioners are white, there has been no impetus for change; the current racial hierarchy suits the interests of white people. It is also the case that for white people, race and racism are not daily facts of life; for white people thinking about race is optional – the effects of white privilege are then rendered invisible or normal and consequently, non white people can be marked as 'abnormal'. Until serious consideration is given as to how to diversify the academic body/bodies of Geography/geographers, whiteness will not receive the serious unpacking it deserves.

See also: Anglo-American/Anglophone Hegemony; Apartheid/Post-Apartheid; Diaspora; Enlightenment Geography; Ethnicity; Historical Geography; Imperialistic Geographies; Indigenous Mapping; Mapping, Non-Western; Postcolonialism/Postcolonial Geographies; Race; Racism and Antiracism; Transatlantic.

Further Reading

Bonnett, A. (2000). *White identities: Historical and international perspectives*. Harlow: Pearson.

Bonnett, A. and Nayak, A. (2003). Cultural geographies of racialization The territory of race. In Anderson, K., Domosh, M., Pile, S. & Thrift, N. (eds.) *Handbook of cultural geography*, pp 300 312. London: Sage Publications.

Carby, H. (1997). White women listen! Black feminism and the boundaries of sisterhood. In Mirza, H. S. (ed.) *Black British feminism: A reader*, pp 45 53. London: Routledge.

Dyer, R. (1997). *White*. London: Routledge.

Dwyer, O. and Jones, J. III (2000). White socio spatial epistemology. *Social and Cultural Geography* 1(2), 209 222.

Fanon, F. (1967). *Black skin, white masks*. London: Grove Press.

Frankenberg, R. (1994). *White women, race matters: The social construction of whiteness*. Minnesota: Minneapolis University Press.

Frankenberg, R. (ed.) (1997). *Displacing whiteness*. Durham: Duke University Press.

Hill, M. (1997). *Whiteness: A critical reader*. New York: NYU Press.

Ignatiev, N. (1996). *How the Irish became white*. London: Routledge.

Jackson, P. (1998). Constructions of 'whiteness' in the geographical imagination. *Area* 30(2), 99 106.

Kobayashi, A. and Peake, L. (2000). Racism out of place: Thoughts on whiteness and an antiracist geography in the new millennium. *Annals of the Association of American Geographers* 9(2), 392 403.

Lambeth, D. (2002). *The Master Subject: White Identities and the Slavery Controversy in Barbados, 1780 1834*. PhD Thesis, Department of Geography, Cambridge University

Lipsitz, G. (2006). *The possessive investment in whiteness: How white people profit from identity politics*. Philadelphia, PA: Temple University Press.

Morrison, T. (1990). *Playing in the dark: Whiteness and the literary imagination*. New York: Vintage.

Peake, L. and Ray, B. (2001). Racializing the Canadian landscape: Whiteness, uneven geographies, and social justice. *The Canadian Geographer* 45(1), 180 186.

Pulido, L. (2002). Reflections on a white discipline. *The Professional Geographer* 54(1), 42 49.

Roediger, D. R. (1991). *The wages of whiteness: Race and the making of the American working class*. New York and London: Verso.

Roediger, D. R. (1994). *Towards the abolition of whiteness*. London: Verso.

Relevant Websites

www.euroamerican.org
Academic Studies of Whiteness and White Culture (Provides links to academic studies on whiteness), Centre for the study of white American culture.

Wilderness

C. R. Warren, University of St. Andrews, St. Andrews, UK

Wilderness: Origins and Evolution of the Concept

During its long history, the concept of wilderness has been invested with diverse and contrasting meanings, meanings which have become progressively more com plex and contested in modern times. Wilderness today is neither simply a biophysical reality nor a conceptual abstraction but a blend of the two; 'real world' wilder nesses and ideas of wilderness exist in a state of co dependence. It is only in recent times that wilderness has become a battleground – conceptually, politically, emo tionally, and sometimes literally. For much of history, the term was a straightforward and largely pejorative de scription of places which were regarded as fearsome, hostile, and/or unproductive. Thus, in the Biblical nar rative, desolate wilderness is the antithesis of the Garden of Eden, providing the appropriately unforgiving setting for the temptations of Christ. Although wilderness was also sometimes regarded as a sanctuary or place of refuge, the word retained its mainly negative connotations in Euro American cultures well into the nineteenth century, an era when the notion of wilderness preservation would have been unthinkable.

The origins of the word itself reflect its history as a derogatory descriptor. Derived from the Anglo Saxon *wildeor*, a term for wild or savage beasts, it broadened to encompass the wild, untamed, barren landscapes in which such beasts lived. The word implied that such areas were 'self willed' or 'wilful'. Such places were defined in contradistinction to cultivated, humanized, unthreatening countryside; without tamed landscapes there could be no untamed wilderness. The concept was meaningless for nomadic hunter gatherers. Only with human control of nature did uncontrolled nature become a category in need of a name. Civilization was almost synonymous with the conquest and subjugation of nature; wild was useless and bad, while tamed was useful and good. Conceptual (and often literal) fences were erected separating the civilized from the wild. This outlook characterized co lonial projects around the world. Most notably, it dom inated and drove the unprecedented, continental scale transformation of North America during the nineteenth century; the chief task for the frontiersmen of the 'Wild West' was to master nature and make it serve human ends. Wilderness was a threatening adversary to be overcome and a natural resource to be exploited.

In modern times wilderness has undergone a con ceptual revolution, and has emerged, newly minted, as an almost hallowed description of areas seen as precious and valuable because of their perceived natural qualities. Having largely shed its negative connotations, the word has become a term of high praise: the wilder the better. What was worthless has become priceless, and the place of satanic temptation has become a sacred temple. Where once the land's highest calling was productive use for society, now the greatest accolade that can be conferred on an area is to be selected for *non* use and set aside for conservation. Aldo Leopold, founder of the Wilderness Society, explicitly inverts the former utilitarian belief that society invests nature with value by 'improving' and 'developing' it, arguing instead that "the raw wilderness gives definition and meaning to the human enterprise." Similarly, whereas the natural world used to be seen as the servant of civilized society, now it has been elevated by some to a position of mastery. The land should again be given the freedom to be 'self willed'. This radical perceptual U turn has been an integral part of the evo lution of Western attitudes to nature more generally.

The progressive reinvention of wilderness as some thing to be treasured – 'unspoiled nature' – paralleled the increasing urbanization and industrialization of Western society, with all the associated spoliation of nature. The perception of wild nature as a threat was supplanted by the recognition that nature itself was threatened by in dustrialized society. Its dwindling extent gave it rarity value, making it suddenly appear as an 'endangered species' in need of protection. Increasingly affluent, urban societies, separated from the daily struggles of coaxing a living from the land, came to see nature as an appealing 'other', a place of peace and beauty, a space for recreation and refreshment for body and soul. As Leopold (1949) wryly put it, "wild things … had little human value until mechanisation assured us of a good breakfast." Taking the lead in articulating and promoting this new view of wilderness was a group of Romantic writers, thinkers, and proto conservationists, together with landscape artists and photographers, who developed the esthetic appre ciation of wild places, a perception that is now so dom inant as to be regarded as 'natural'. Wordsworth, Ruskin, Goethe, and Rousseau in nineteenth century Europe, and John Muir, Henry Thoreau, Ralph Waldo Emerson, Aldo Leopold, and Ansel Adams in nineteenth and early twentieth century America were particularly influential in revalorizing nature and wilderness as precious com modities (**Figure 1**).

A belief in the value of wilderness led naturally to a desire to protect it, and this, in turn, gave birth to the

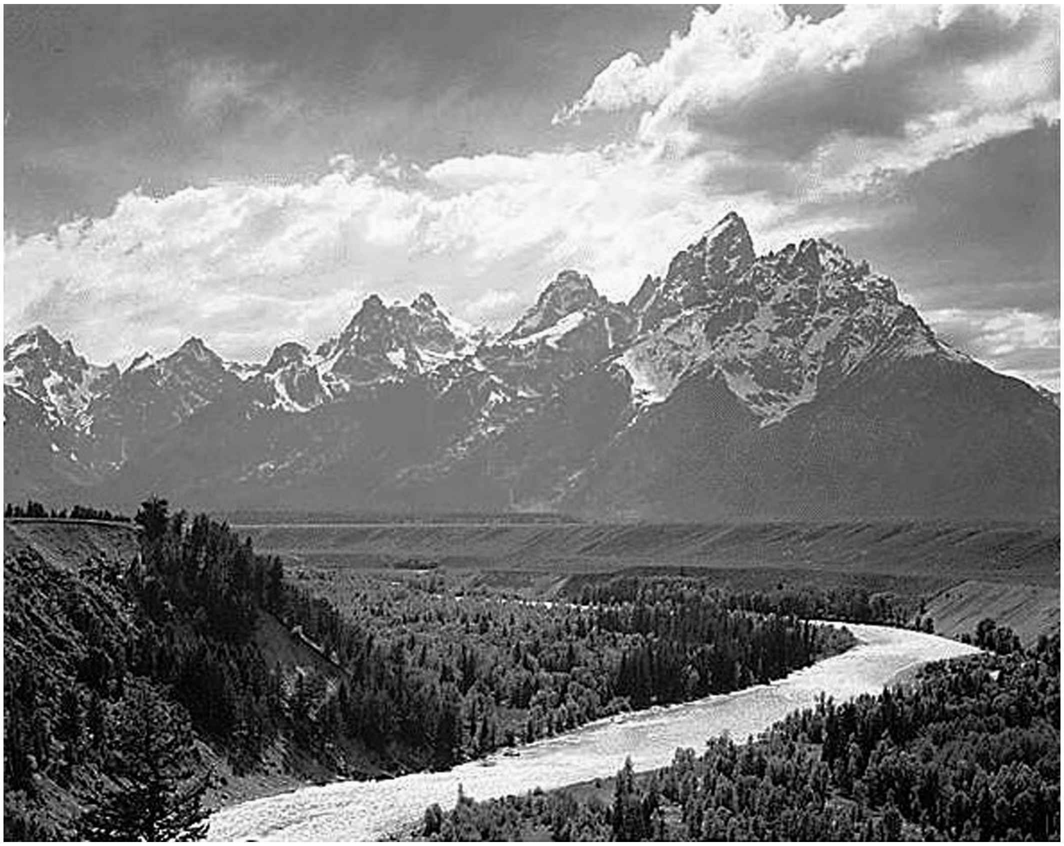

Figure 1 Grand Teton National Park, Wyoming, by Ansel Adams.

national park movement. This movement, inspirationally led by John Muir, campaigned successfully for the es tablishment of the first parks at Yellowstone (1872) and Yosemite (1890) in western USA. National parks have gone on to become flagship aspects of nature conser vation and symbols of national pride worldwide, and protected areas of all kinds now cover some 10% of the Earth's land surface. The landscape protection movement is a broad church, incorporating a diverse range of mo tivations and beliefs ranging from the anthropocentric to the ecocentric, and from utilitarian values to the intrinsic values of 'deep ecology'. The conservationist end of this spectrum, heir to Gifford Pinchot's 'wise use' approach, permits some human utilization of resources, whereas the preservationist end disparages the idea that nature should be seen primarily as a human resource; by investing wilderness with intrinsic value (irrespective of its use value to humans) it challenges a key foundation of Western value systems, namely that moral standing is a strictly human quality. Nevertheless, despite such di versity, a belief that wild places are valuable is near universal.

As this conviction came to be shared ever more widely in Western society in the later twentieth century, so the idea of wilderness grew in power and influence both socially and politically, and, indeed, in emotional and spiritual ways; Muir's 'wilderness theology', a belief in the sacredness and healing power of wild places (for body, mind, and spirit) now has many adherents. (The internet site Wilderness.net, for example, describes wil derness areas as "places where one can retreat from civilization, reconnect with the Earth, and find healing, meaning and significance.") The defense of wilderness, especially in 'new world' countries, rose to become the environmental movement's first order goal, despite the fact that formal definition of the concept has always proved elusive and controversial. As Nash observes, "the term is so heavily freighted with meaning of a personal, symbolic, and changing kind as to resist easy definition." Arguments for the protection of wilderness come more easily to hand, however. Four that are frequently made have been characterized by Godfrey Smith as the cath edral argument, the laboratory argument, the silo argu ment, and the gymnasium argument. Respectively, these refer to the value of wilderness for esthetic pleasure and spiritual renewal; for scientific research into the planet's life support systems; for the provision of a stockpile of genetic diversity; and for athletic or recreational activ ities. To these human interest arguments, which have long been made and which some now find inadequate, can be added two others: the value of large wilderness areas in the maintenance of ecological functioning, and the intrinsic value case that such areas are valuable in and of themselves.

The passing of the US Wilderness Act of 1964 was the culmination of a century of conservation advocacy and activism, marking perhaps the apogee of confidence in the traditional, Romantic concept of wilderness. This act formalized the modern Western concept of wilderness, famously defining it as:

> … an area where the earth and its community of life are untrammeled by man, where man himself is a visitor who does not remain. … [It is] an area of undeveloped Federal land retaining its primeval character and influ ence, without permanent improvements or human habitation, which is protected and managed so as to preserve its natural conditions. … [It] generally appears to have been affected primarily by the forces of nature, with the imprint of man's work substantially unnoticeable.

Since that time, wilderness advocates have had to contend with a growing chorus of criticism.

Deconstructing the Idea of Wilderness

In recent decades the idea and practices of wilderness protection have come under sustained attack from social scientists and others. Four main critiques have been de veloped. The first takes issue with the claim that wil derness areas are pristine and free of human modification. The argument, quite simply, is that a human signature can be found from pole to pole, and from mountain top to deepest ocean, even if only in trace form. If Earth is nowhere pristine, then, according to an absolutist defin ition, true wilderness is extinct. This was the line de veloped by McKibben in *The end of nature*:

"Those Coke cans are everywhere. We're every where. … It's mostly us now."

Given that the hydrological cycle, the atmosphere, and the carbon cycle have all been altered, there is no place where culture has not colonized nature. We live in the 'Anthropocene'. This first criticism is hardly new, dating back as it does to Karl Marx. It is readily acknowledged by many conservationists, especially in Europe where the influence of people has been so ancient and pervasive that few if any places are entirely unmodified. Most accept that naturalness exists on a sliding scale, and that wil derness cannot be defined in absolute terms. Other cri tiques, however, have proved more controversial.

A second argument grows from the historical reality sketched above: since the meanings ascribed to wilder ness have altered profoundly, wilderness is as much an idea as it is a biophysical reality. It exists 'in here', in our minds, as much as 'out there' in the environment. Some have gone further, arguing that wilderness is a wholly cultural construct. This argument accepts, of course, that the material world exists but contends that wilderness has no existence independent of us. In his famously contro versial essay 'The trouble with wilderness; or, getting back to the wrong nature', Cronon (1996: 69–90) works out this critique at some length, arguing that wilderness is "entirely the creation of the culture that holds it dear," existing merely as a counterpoint to civilization. Further, he accuses it of serving as:

> … the unexamined foundation on which so many of the quasi religious values of modern environmentalism rest … [Wilderness represents] the false hope of an escape from responsibility, the illusion that we can somehow wipe clean the slate of our past and return to the *tabula rasa* that supposedly existed before we began to leave our marks on the world. (Cronon, 1996: 69 90)

The cultural construction critique, if upheld, robs wilderness of its 'otherness' and its claimed objective reality, arguing instead that it consists of a projection of (evolving) societal values onto the physical environment. It is thus revealed as a specific cultural category, not a 'scientific label' for particular assemblages of material phenomena.

Third comes the accusation of misanthropy, the ar gument that conceptions of people free wilderness exalt nature to the exclusion (literally) of people. This is an accusation which has also been leveled at the conservation movement more generally, especially at its more eco centric subsets such as 'deep ecology'. The conviction that natural, wild areas are of greater value than human altered ones pervades conservation thinking and popular discourses about the environment. 'Unspoiled' ecosystems are seen as the gold standard for nature, the reference condition for ecological 'integrity' or 'health', and only wild places are worth cherishing. Wilderness areas are accorded high value precisely because of their natural ness, with 'natural' typically carrying the force of a moral imperative. Such a value system, within which wilderness is sacred and humanity is profane, was elo quently articulated in the nineteenth century by John Muir who saw humans as the only unclean animals and regarded nature as his temple. This outlook retains con siderable power today. However, this value system can all too easily give rise to misanthropy, a negative view of people and communities and of their efforts to improve their lot. Human beings come to be seen as an in convenient nuisance, a spanner in the finely tuned workings of sublime nature. Such a view has been roundly criticized by many. Brown, for example, believes that:

> A romantic urban based wilderness fantasy has no place for human beings to make a living. It is dangerous

> because it leaves no place for discovering what a sustainable human place in nature might be. Furthermore, it sets an arbitrarily high standard against which cultural and productive landscapes are regarded as inferior. (Brown, 1997: 191 198)

Thomas argues similarly that “nature parks and conservation areas … are fantasies which enshrine the values by which society as a whole cannot afford to live.”

A fourth and final criticism of the wilderness idea, also related to questions of the place of people in nature, is the charge that the wilderness concept is tarred with the brush of ethnic cleansing. Many of the archetypal wilderness areas which inspired early conservationists were devoid of people not because this was their natural state but because their native inhabitants had been removed, either by force or by introduced diseases or both. The recognition that most such places have a long history of human habitation by indigenous peoples who regard them as ancestral homelands has further destabilized cherished, frontier born notions of wilderness. This fourth critique charges that the designation of such locations as national parks or wilderness areas amounts to the airbrushing of First Nations out of history, a process which can lead to the erasure of these peoples’ land rights. Further, the conception of pristine, people free wilderness denies the humanness of the peoples who lived there for countless generations, turning them instead into ‘savages’ (or, at best, ‘noble savages’). The European colonial experience was seminal in the formation of Western environmental attitudes, and, reciprocally, Western narratives of wilderness were woven through colonial projects around the world. Callicott and Ybarra (2001) argue that “the wilderness idea plays a crucial role in masking colonial American genocide and ethnic cleansing.” Such criticisms relate most directly to the American West where, beginning with Yellowstone, the creation of national parks did sometimes involve the expulsion of native people, but they also apply in many other places, both in the developing and developed world. The Highlands of Scotland, for example, are celebrated and protected for their wildland qualities, and yet the emptiness of many glens is the result of forced evictions during the infamous Clearances of the early nineteenth century. The ‘people free wilderness’ concept thus stands accused of being an elite European/American construction with strong imperialist and colonialist associations, a moral menace which continues to threaten the sustainable livelihoods of indigenous peoples and local communities.

Wilderness began as a word with deeply negative connotations, but then, following its nineteenth century rebranding, it enjoyed a century or more of entirely positive press. It now seems as if its journey is coming full circle as a consequence of the critical attacks of recent years. In some quarters, at least, wilderness is now a dirty word once more.

From Wilderness to Wildness and Re-wilding

The combined weight of the above arguments has led to the abandonment of earlier, more fundamentalist, reified thinking at least in academic circles, and to the dismissal of the concept as ‘the wilderness fallacy’ or ‘the wilderness illusion’. With the increasing acceptance that the purist formulation of wilderness (as both pristine and people free) is indefensible, a subtle but significant shift of language and concept has commenced, replacing ‘wilderness’ with ‘wildness’. Even though wilderness *sensu stricto* may be effectively extinct, and despite the fact that the concept has been unmasked as a cultural construct, it is nevertheless true that vast tracts of the globe still evoke a profound sense of ‘wildness’, a sense of ‘nature in charge’. These are the places where the landscape as a whole has not been humanized, even if some human strands are woven through the landscape tapestry. For many people, it is this perception of wildness – of nonhuman processes being in the ascendant – which invests areas with value and speciality. Interestingly, the ‘founding fathers’ of the wilderness movement used both ‘wilderness’ and ‘wildness’ interchangeably. Thus Thoreau’s famous dictum, adopted as the motto of the influential Sierra Club, is “in wildness lies the preservation of the world,” and Muir expressed the belief that “in God’s wildness lies the hope of the world, the great fresh unblighted, unredeemed wilderness.” These men regarded wildness as an attribute of *Homo sapiens* as well as of other animal species, and they articulated a proto ecological philosophy of holistic interconnectedness between people and nonhuman nature.

However, the subsequent evolution of the wilderness concept progressively came to exclude people. This has created one of the greatest conceptual problems with today’s dominant vision of wilderness, namely that it institutionalizes the Western conceptual separation of people and nature, leaving no place for human beings in nature. With the mounting evidence of the longstanding, profound, and pervasive influence of human societies on the environment has come the realization that many ‘natural’ areas are in fact ancient cultural landscapes, the product of millennia of human–environment interactions. Natural and artificial bleed into each other along a complex continuum. There is therefore no place for simplistically equating ‘natural’ with ‘morally best’; to do so is to commit the naturalistic fallacy. The idea that wilderness areas should be free of people seems to have less to do with ecology than with the persistent pre Darwinian view that humanity is separate from nature, and with the Romantic ideal of a primaeval nature independent of 'man', concepts which are now widely regarded as false. The renewed emphasis on 'wildness' in place of 'wilderness' acknowledges the need to find a middle way, to find ways of reintegrating people and nature which preserve and promote the health of both.

Despite the overlapping meanings of 'wilderness' and 'wildness', arguably the latter is a broader category, a quality which can be encountered in many places outside wilderness areas even though it exists in its most concentrated, 'pure', and spatially extensive form within them. Wildness is even more challenging to define than wilderness, and distinguishing between them is not easy because 'wilderness' itself has many shades of meaning. For example, the US Eastern Wilderness Areas Act of 1975, in using a definition of wilderness which is more accommodating of human influence than the 1964 Act, travels some distance along the wilderness–wildness continuum. Equally, some European definitions of wildness are only distinguishable from 'new world' conceptions of wilderness in terms of spatial scale. But although the two terms overlap substantially, adopting the more flexible notion of wildness has several advantages. It avoids many of the fraught connotations of 'wilderness' while preserving the essence of the speciality which many cherish. Secondly, it escapes the confines of a spatially defined topographic entity ('wilderness areas'), replacing it with a concept which is more spatially neutral. Wildness can exist anywhere and everywhere; a migrating bird in a city is as wild there as it would be anywhere else. Thirdly, it helps to overcome the troublesome dualism between nature and culture. Distinguishing between the two terms creates a place for tutored human intervention that does not diminish or destroy natural value because human action now belongs in the context of the natural. Perhaps Thoreau coined his famous aphorism better than he knew.

The latest addition to wilderness debates is the concept of re wilding. Historically, the primary response to the threats posed by economic development has been preservation, but in the late twentieth century a more proactive response emerged, with attempts being made to right the perceived wrongs of the past by restoring areas to their 'natural' (or at least 'historic') state. Dams are removed, native ecosystems reestablished, lost species reintroduced, and drained wetlands flooded once more. The motivations for such restorative action are diverse (ecological, esthetic, social, economic) but frequently include an explicit moral imperative arising from the belief that, as the perpetrators of the damage, we bear an ethical responsibility to put things right. The trend for such ecological restoration or re wilding has gathered considerable momentum, and proposals have become ever more ambitious in scope and scale. For example, the so called 'Pleistocene re wilding' of North America envisages the introduction of large African mammals as proxies for the species which were lost after the last Ice Age. Its advocates claim that significant environmental, economic, and social benefits would flow from this change in conservation's ethos (a shift from reactively managing extinction to proactively restoring natural processes), but the proposals have, unsurprisingly, generated sharp controversy.

Irrespective of the pros and cons of specific re wilding projects, the very concept of ecological restoration is itself contested. 'Restored' ecosystems are dismissed by some as fakes and artefacts because, despite the superficial appearance of wildness and the restored dominance of nonhuman processes, such areas are human creations and therefore not authentic. The concept is also afflicted by a range of fraught questions about boundaries and baselines, both in space and time. For example, how should the restoration area be chosen and delimited? Which of the many historical states of naturalness should be (subjectively) selected as the template defining the goals of restoration? Such questions cannot be answered unambiguously because striving for historical fidelity unavoidably involves arbitrary spatiotemporal choices, privileging certain historical geographies over others. In this regard, debates over re wilding resonate strongly with the dilemmas surrounding native and alien species.

Whither Wilderness or Wither Wilderness?

Does the idea of wilderness have a future? Or will it wither away under the weight of critical attack and be abandoned as a Romantic, sentimental anachronism? Debates about wilderness are situated within broader deliberations about nature, culture, and the place of *Homo sapiens* in the 'natural' world, debates which have been fertile ground for geographers. Prominent within these explorations has been the reconfiguring of conceptions of nature and naturalness alluded to above. One consequence is that a belief that nature and wilderness are culturally constructed concepts is now widely shared in academic circles (though not uncontroversial). Moreover, the equating of 'valuable' with 'undisturbed' has been challenged by the recognition that all ecosystems are hybridized palimpsests of human and nonhuman influences. This historical reality means not only that people are a functional component of many landscapes, but that the concept of a 'pristine' (human free) past has little relevance for conservation management. Notwithstanding the rearguard action of conservationists who argue that maintaining the conceptual separation between nature and culture is important (for fear that its collapse would compromise conservation efforts), it is clear that it is a modernist distinction that is rapidly dissolving. Given that traditional conceptions of wilder ness are intimately bound up with this distinction, it is tempting to conclude that the collapse of one will bring down the other. Equally, for the reasons set out above, there are grounds for suggesting that the twentieth century understanding of wilderness will be supplanted by wildness during the twenty first.

However, it may be premature to draft an obituary for wilderness. Some conservationists refuse to accept that the concept is discredited, and point out that it is alive and well in practice. There is also a large and burgeoning popular literature which, taking no heed of the critical debates, continues to assume a Cartesian separation be tween nature and culture, and to promulgate the semi religious reverence for wilderness as a source of healing and wholeness that was developed by Muir and Leopold. Consequently, despite the critical mauling, the concept of wilderness may retain its vernacular and sociopolitical potency for some time to come. Moreover, there are those who argue that despite its troubling genealogy and connotations, the wilderness idea remains a potent and on the whole positive myth, and that to abandon it now would be to abandon a powerfully resonant argument for the defense of nature. As the world's ecosystems become increasingly humanized, the significance of wilderness may actually increase. Whereas critics have accused wilderness of entrenching the separation of people and nature, it can also be argued that the wilderness idea provides a matrix for reintegration, and that wilderness preservation offers a route to eco salvation. In Oels chlaeger's words, it is an idea which undergirds "a new paradigm for understanding humankind as embodying natural process grown self conscious." Whether the ecological humility that this implies is rediscovered in the name of wilderness or wildness is perhaps less im portant than that the rediscovery takes place.

See also: Conservation and Ecology; Ecology; Environment, Historical Geography of; Environmentalism; Landscape; National Parks; Nature; Nature, Historical Geographies of; Nature, History of; Nature, Social; Nature-Culture.

Further Reading

Brown, N. (1997). Re defining native woodland. *Forestry* 70, 191 198.
Callicott, J. B. and Nelson, M. (eds.) (1998). *The great new wilderness debate*. Athens: University of Georgia Press.
Callicott, J. B. and Ybarra, P. S. (2001). The Puritan origins of the American wilderness movement. Research Triangle Park, NC: National Humanities Center, http://nationalhumanitiescenter.org/tserve/nattrans/ntwilderness/essays/puritan.htm (accessed Nov 2007).
Cronon, W. (1996). The trouble with wilderness; or, getting back to the wrong nature. In Cronon, W. (ed.) *Uncommon ground: Rethinking the human place in nature*, pp 69 90. New York: W.W. Norton and Co.
Donlan, J., Berger, J. and Bock, C. E. (2006). Pleistocene re wilding: An optimistic agenda for twenty first century conservation. *The American Naturalist* 168, 660 681.
Elliot, R. (1997). *Faking nature: The ethics of environmental restoration*. London: Routledge.
Gobster, P. H. and Hull, R. B. (eds.) (2000). *Restoring nature: Perspectives from the social sciences and humanities*. Washington, DC: Island Press.
Godfrey Smith, W. (1979). The value of wilderness. *Environmental Ethics* 1, 309 319.
Grove, R. H. (1995). *Green imperialism: Colonial expansion, tropical island Edens, and the orgins of environmentalism, 1600 1860*. Cambridge: Cambridge University Press.
Guha, R. (1994). Radical American environmentalism and wilderness preservation: A Third World critique. In Gruen, L. & Jamieson, D. (eds.) *Reflecting on nature*, pp 241 251. New York: Oxford University Press.
Hay, P. (2002). *A companion to environmental thought*. Edinburgh: Edinburgh University Press.
Leopold, A. (1949). *A Sand County almanac, and sketches here and there*. New York: Oxford University Press.
McKibben, W. (1999). *The end of nature* (3rd edn). New York: Anchor Books.
Nash, R. F. (2001). *Wilderness and the American mind* (4th edn). New Haven: Yale University Press.
Oelschlaeger, M. (1991). *The idea of wilderness: From prehistory to the age of ecology*. New Haven: Yale University Press.
Thomas, K. (1983). *Man and the natural world: Changing attitudes in England 1500 1800*. London: Allen Lane.
Whatmore, S. and Thorne, L. (1998). Wild(er)ness: Reconfiguring the geographies of wildlife. *Transactions of the Institute of British Geographers* 23, 435 454.

Relevant Websites

http://www.iucn.org
The International Union for the Conservation of Nature.
http://www.johnmuir.org
The John Muir Global Network.
www.sierraclub.org
The Sierra Club, a long established wilderness advocate.
http://www.wilderness.org
The Wilderness Society.
http://www.wilderness.net
Wilderness.net, a site devoted to wilderness promotion.

Wilson, A.

G. P. Clarke, University of Leeds, Leeds, UK

Wilson, Alan (1939–)

Alan Wilson was knighted in 2001 for his services to UK higher education after a career which has spanned five decades. In part, this was recognition for his work in the late 1990s, when, having served in many areas of uni versity administration, Alan Wilson became vice chan cellor at the University of Leeds. It is difficult to recount his legacy at Leeds in a few brief words. However, in his 13 years as vice chancellor, he oversaw unprecedented growth and transformation. Student numbers increased from 12 000 to 31 500, turnover increased from £100 to £320 million, and research income increased by more than 400% to £71 million. Following his departure as vice chancellor he was appointed as the first director general for higher education by the UK government, where he operated as key advisor to various UK edu cation ministers.

Wilson was born in Bradford in 1939, and educated at Queen Elizabeth Grammar School, Darlington and Corpus Christi College, Cambridge. With a background in mathematics and physics, he began working in the socioeconomic sciences in the mid 1960s at the Ministry for Transport and then the Centre for Environmental Studies in London. He was intrigued by the links be tween the physical sciences and the human sciences, especially how people moved around cities in their everyday interactions between home and job, school, retail centre, etc. Indeed, his first major contribution to the field of urban and regional planning came with his interest in what was then known as the gravity model, a model that linked was the physical and human sciences closely. The gravity model is essentially a flow model – in physics it was derived from work by Isaac Newton re lating to the magnetic force between two planets, which Newton argued to be a function of the size of the planets (positive impact) and the distance between them (nega tive impact). However, Alan Wilson was concerned with the weak theoretical foundations of the gravity model in terms of modeling people in cities and set about to prove that the same principles could be arrived at from an economics perspective. Thus, he demonstrated that it was possible to derive interaction or flow models from first principle using entropy maximizing methods (models which produced the most likely outcomes based on the average or most likely set of generated trips). Using this model he quickly derived the family of spatial interaction models that would form the heart of many of the future mathematical models of trip movement in the city.

During the 1970s, Wilson worked on a variety of different types of interaction model. The retail model (estimating trip patterns for shopping trips) was often used as the exemplar for new theoretical extensions and developments but he quickly opened up new application areas. Working with Martyn Senior he developed models for residential location and working with Phil Rees he wrote a major book on 'spatial population analysis' which developed a variety of models for population change and migration. The 1980s saw further extensions into health care (working with Martin Clarke in particular) and agriculture and industrial location (with Mark Birkin).

It was in the 1970s that the first major critiques of urban modeling appeared in the literature. These cri tiques were primarily aimed at large scale urban models of the sort pioneered by Ira Lowry in 1964. Such models consisted of a number of subsystems linked together to form comprehensive models of cities and regions. The criticisms were based on the problems of large scale models in general and around the static nature of many models. Also, such models were criticized for ignoring the behavior of individual agents in the city. Such cri tiques found favor throughout the discipline of human geography and led a generation of young researchers to turn their backs on modeling and look to new areas of critical human geography. This included a number of now renowned researchers who had started their careers as quantitative geographers (e.g., Andrew Sayer, David Harvey, Doreen Massey) but who now explored Marxist and critical social theories.

The upshot of this movement away from models did lead to at least a partial rethink about the use of quantitative methods from within the modeling community. The focus for Alan Wilson would shift to urban dynamics and a concern to make models more accessible and useful. Advances in dynamic spatial interaction models were driven by a key paper in 1978 published in *Environment and Planning A* by Alan Wilson in collaboration with a US planner Brit Harris. It was argued in this paper that a dynamic mechanism could be added to the spatial interaction model to introduce some element of retailer response to market conditions (the retail model being again used as a source of numerical and theoretical experiments). Working closely with Martin Clarke, such research led to the discovery that small changes in key parameter values could lead to major structural change in the systems being modeled. Given that such parameter change could flip systems from one state to another, they were partly derived from bifurcation or catastrophe theory, theorems, again popular in the hard sciences.

Given the widespread skepticism expressed about modeling within the discipline, Alan Wilson was deter mined that the kitbag of techniques could be useful in real world settings. In addition to the family of spatial interaction models he had been working with his research team on powerful optimization or mathematical pro gramming models that could help to site facilities in the best locations, given a number of constraints. Although an applied modeling book was published with colleagues in 1977, it was in the mid 1980s that Alan Wilson and Martin Clarke set out on a mission to find new appli cation areas. Aided by Mark Birkin and Graham Clarke, they began to interest potential clients in both the public and private sectors. The latter seemed especially recep tive to mathematical models, especially in the retail sector. This work took off when the UK newsagents W. H. Smiths and Toyota GB each bought national geographic information systems (GIS) and modeling systems which would allow them to look at the performance of existing stores/dealerships and estimate how much revenue new stores/dealerships would be likely to raise. By the early 1990s, GMAP (geographical modeling and planning) had been formed and turnover levels since the mid 1980s reached over £50 million; staffing levels were often over the 100 mark. Major blue chip clients over the years included Ford, Exxon/Mobil, BP, Barclays, Sainsbury's, Asda, Thomas Cook, and HBOS, making this probably the largest geographical consultancy in the world. The benefits of this work (to research, the university, geog raphy, and students) have arguably been massive and led to a major new applied book by the four authors in 1996.

Another key aspect of making models more accessible to end users was the need to improve the outputs. In the mid 1980s Alan Wilson worked with colleagues at Leeds and in Italy to produce a framework for a new set of performance indicators which would measure levels of retail provision and catchment area size and structure. Graham Clarke worked with Alan Wilson on the devel opment of these new performance indicators and went on (with colleagues) to plot these provision indicators for the cities of Leeds, Bradford, and Cardiff to show the extent to which the top food retailers had avoided low income areas (now labeled 'food deserts') creating prob lems of access to fresh fruit and vegetables and therefore healthy diets (there is an obvious link between poor diets and problems relating to obesity and type II diabetes).

Despite the success of the applications of the various subsystem models, it has proved more difficult to build and operationalize a comprehensive model of the Lowry type. Better data, faster and more powerful computers, and possibly a better understanding of how cities work make such models more than a remote possibility. However, the complexity of linking sectors and in corporating the behavior of various actors (especially endogenously) has proved difficult to solve. Nonetheless, new frameworks and models do appear from time to time and Alan Wilson continues to examine this area in his most recent publications. If such a dynamic model could be built for a modern city, this would indeed be a fitting tribute to Alan Wilson's huge contribution to human geography.

See also: Planning, Urban; Regional Development Models; Regional Planning and Development Theories; Urban Modeling; Urban Planning and Human Geography.

Further Reading

Bertuglia, C. S., Clarke, G. P. and Wilson, A. G. (1994). *Modelling the City: Performance, Policy and Planning*. London: Routledge.
Birkin, M., Clarke, G. P., Clarke, M. and Wilson, A. (1996). *Intelligent Geographical Information Systems*. Cambridge: GeoInformation International.
Harris, B. and Wilson, A. G. (1978). Equilibrium values and dynamics of attractiveness terms in production constrained spatial interaction models. *Environment and Planning, A* 10, 371 388.
Rees, P. H. and Wilson, A. G. (1977). *Spatial Population Analysis*. London: Edward Arnold; New York: Academic Press.
Wilson, A. G. (1970). *Entropy in Urban and Regional Modelling*. London: Pion (Russian translation, Nauka, Moscow, 1978).
Wilson, A. G. (1974). *Urban and Regional Models in Geography and Planning*. New York: Wiley.
Wilson, A. G. (1981). *Catastrophe Theory and Bifurcation: Applications to Urban and Regional Systems*. London: Croom Helm; Berkeley, CA: University of California Press.
Wilson, A. G. (2000). *Complex Spatial Systems: The Modelling Foundations of Urban and Regional Analysis*. Harlow: Prentice Hall.
Wilson, A. G, Coelho, J. D., Macgill, S. M. and Williams, H. C. L. (1981). *Optimisation in Locational and Transport Analysis*. New York: Wiley.
Wilson, A. G, Rees, P. H. and Leigh, C. M. (eds.) (1977). *Models of Cities and Regions*. New York: Wiley.

Relevant Websites

http://www.gmap.co.uk/pages/gmap
GMAP Consulting.
http://www.casa.ucl.ac.uk/people/person.asp?ID=44
UCL Centre for Advanced Spatial Analysis (CASA).

World/Global Cities

B. Derudder, Ghent University, Gent, Belgium

Glossary

Global City A city that contains large clusters of internationally orientated producer services firms.

Globalization The expansion of transnational integration and interdependence in the economic, social, technological, cultural, and political spheres.

Multinational Corporation A company that has its facilities and other assets in at least one country other than its home country.

Network Society A society in which a combination of social and technological networks shapes its prime mode of organization and most important structures at all levels.

New International Division of Labor Economic, social, and spatial changes in the demand for and organization of labor that began in the early 1970s, and which led to the internationalization of production.

Producer Services Highly specialized and bespoke forms of service activity such as finance, accounting, advertising, property development, and law; these services are primarily sold to a clientele of business firms and government institutions.

World Cities Cities that function as the economic, geopolitical, and ideological-symbolical powerhouses of the capitalist world-system.

Definition

World cities or global cities are the command and control centers of the global economy, that is, they are nodal points that function as organizing centers for the interdependent skein of material, financial, and cultural flows that together sustain contemporary globalization. The most commonly identified examples of world/global cities include major centers such as London, New York, Tokyo, Paris, and Hong Kong (**Figure 1**).

Background

The idea of a world economy articulated through certain cities is longstanding, but the specific suggestion of a set of cities having a strategic role in coordinating and controlling the world's economy is relatively new. Contemporary research on world/global cities can be traced back to a handful of authors writing in the 1970s and 1980s, whereby – in spite of some earlier contributions – its theoretical raw materials are regularly traced back to John Friedmann's 'World city hypothesis'. In this article, the author identifies world cities as centers that control and articulate the 'new international division of labor' created by multinational corporations. This unambiguous conceptual connection between 'world city formation' and the then much debated emergence of a 'new international division of labor' already points to what would become a major, recurring characteristic of the world/global cities literature, that is, the fact that it has always been very quick in absorbing new theories, methodologies, and intellectual frameworks developed in other literatures. Major examples include the intrusion of vigorous and wide ranging debates, such as those on the potential of post structuralist thinking and on the role

Figure 1 Hong Kong as a world city.

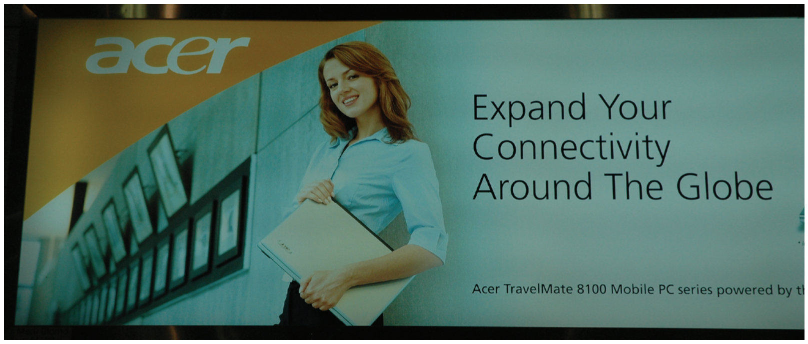

Figure 2 The rise of a 'network society'.

and relevance of national states in a global era. Taken together, this has led to a situation in which the world/global cities research has turned into a laboratory for the social sciences as a whole, whereby major figures in human geography have made significant contributions.

Research on world/global cities is founded on two key observations. First, the worldwide distribution of economic activity necessitates strategic control functions that are found in a limited number of locations: globalization in its various guises has led to increased geographical complexity, which calls for some control points to ensure the smooth functioning of the global system. In other words, world/global cities contain a disproportionate quantity of strategic agents in the global system (e.g., headquarters of multinational corporations and international institutions, and specialized and internationalized business firms), and this functional concentration mirrors the deepening of processes of global integration and interdependence.

Second, the exercise of this strategic control is accomplished through the capacity of these world/global city agents to network across space. From the 1970s onward, two distinct advanced technologies, computers and communications, combined to create a new enabling infrastructure for global organization. As it became more pervasive and sophisticated, this global infrastructure implied that spatial organization became increasingly conceptualized through networks (in which interaction is defined by simultaneity) to the detriment of territoriality (in which interaction is defined by proximity). Envisaging world/global cities in the context of a 'network society' particularly gained momentum in the 1990s, when influential authors like Manuel Castells asserted that globalization processes are basically all about transnational processes operating through numerous networks.

Taken together, it is clear that the upsurge of interest in world/global cities has been instrumental in rooting urban geography in the debates that have dominated the social sciences from the 1980s onward. A straightforward interpretation of the remarkable success of this research field, therefore, is that it directly confronts the commonplace observation that contemporary urban life cannot be properly understood without making some reference to 'globalization' and the coming into existence of a 'network society' (**Figure 2**). When combining both buzzwords, it can be noted that world/global cities are commonly conceived within the framework of a global network of cities, which provides geographers with an alternative spatial skeleton for mapping and understanding contemporary globalization.

Research Agendas: 'Cities as Systems Within Systems of Cities'

In tune with the urban geography literature at large, world/global cities research can be conceived as involving two more or less distinct research agendas, the first focusing on the position of cities in an overarching system, and the second on shifting socioeconomic patterns within cities.

The first research agenda emphasizes that world/global cities provide an interface between various scales from the global to the local, as they contain economic, sociocultural, and institutional settings that facilitate the articulation of regional and metropolitan resources and impulses into globalizing processes, while, conversely, mediating the impulses of globalization to regional and local settings. As such, there are several, heavily interconnected functional components related to world/global city formation:

- they are the sites of concentrations of headquarters of multinational, transnational, and national enterprises (e.g., the concentration of corporate headquarters in the New York metropolitan area);
- they are the sites of clusters of specialized business services such as finance, accounting, advertising, property development, and law, especially those which are international in scope (e.g., the clustering of internationally oriented business service firms in Lower Manhattan);

- they are the sites of most of the leading global markets for commodities, commodity futures, investment capital, foreign exchange, equities, and bonds (e.g., the New York Stock Exchange); and
- they are the sites of nongovernmental and inter governmental organizations with an international scope (e.g., the headquarter of the United Nations in New York).

The second research agenda deals with the alleged link between a city's global connectivity and its internal economic and social structure. One of the major theses advanced by leading researcher Saskia Sassen is that growth of the financial and business services sector has led to a deepening of the gap between rich and poor in world/global cities. On the one hand, these sectors imply the formation of well paid jobs, such as managers and executives. On the other hand, there are also a large number of low paying jobs in unskilled positions that cater to this elite. While the former group enjoys a soaring income through high wages and excessive bo nuses, the latter struggle in their minimum wage jobs, which together lead to massive social polarization and ensuing gentrification in world/global cities. It is thereby often emphasized that the highbrow jobs are often filled by expatriates, while the low paid jobs are typically filled by immigrants, hence the direct connections between research on world/global cities, on the one hand, and research on international migration patterns, on the other. This 'polarization thesis' has, however, been con tested. Chris Hamnett, for instance, has called for a more comprehensive approach to understanding shifting socioeconomic structures in world/global cities.

Cities in Globalization: Different Conceptualizations

Although there are most certainly a number of common vantage points in the global/world cities literature at large, this is in practice a very large research field with a blurry perimeter. Apart from the consensus that the study of contemporary urbanization needs to recognize the importance of 'globalization' and 'networks' in one way or the other, there is sometimes very little that connects the various strands of investigation. For one thing, the various approaches are loosely united in their observation that cities such as New York and London somehow derive their functional importance from a privileged position in transnational networks of capital, information, and peo ple. Furthermore, most researchers take a similar ap proach in that they: (1) point to the effects of the increasing internationalization of trade and production, (2) give some weight to the increased service intensity in all sectors of economic life, and (3) identify the relevance of recent evolutions in telematics. However, at the same time, it is quite obvious that an all purpose approach rings rather hollow. Acclaiming the relevance of studying cities in the context of a transnational urban network may be acceptable for pedagogic or even heuristic reasons, but it most certainly needs further specification if one wishes it to have some analytical value within the scientific debate.

One major example is that the terms 'global city' and 'world city', although often used as synonyms, refer to different concepts. There are many other terminologies for describing these new metropolitan forms, but here it suffices to note that there are specific conceptualizations that merit specific treatment. Arguably, the most im portant approaches are those developed by John Fried mann (world cities) and Saskia Sassen (global cities), together with recent attempts by leading geographer, Allen Scott, to connect the world/global cities literature with research on the 'new regionalism' in the global economy through his self styled 'global city region' concept.

The world city concept can be traced back to John Friedmann's seminal 'World city hypothesis'. This text framed the rise of a transnational urban network in the context of a major geographical transformation of the capitalist world economy. This restructuring, most commonly referred to as the 'new international division of labor', was basically premised on the international ization of production and the ensuing complexity in the organizational structure of multinational enter prises. This increased economic geographical complex ity, Friedmann argues, requires a limited number of control points in order to function, and world cities are deemed to be such points. His basic contention, then, is that capital assigns spatial dominance to the city, that transnational capital allocates control to just a select few of those cities, and that without these 'world cities' the global system of economic relations would simply be impossible.

Friedmann attempts to give theoretical body to his framework for research by implicitly referring to Im manuel Wallerstein's world systems analysis, in which capitalism is envisaged as a system that involves a hier archical and a spatial inequality of distribution based on the concentration of relatively monopolized and there fore high profit production in a limited number of 'core' zones. The division of labor that characterizes this spatial inequality is materialized through a tripolar system consisting of core, semi peripheral, and peripheral zones. Friedmann's prime purpose, then, is to build an analytical framework that deflects attention from the role of terri torial states in the reproduction of spatial inequality in the global economy.

Apart from being the economic power houses of the capitalist world system, world cities are also locales from which other forms of command and control are exercised, for example, geopolitical and/or ideological symbolical control over specific (semi)peripheral regions in the world system – Miami's control position over Central America is a case in point here. Having said this, it is clear that Friedmann believes that the economic variable dominates all other attempts at explanation, and hence the focus on headquarters of multinational corporations in a significant segment of the research based on his seminal contribution.

The global city concept, in turn, can be traced back to Saskia Sassen's book, *The Global City*, which has arguably become the key text in this literature. Criticizing Friedmann for his failure to properly explain the functional concentration of key agents in selected cities, Sassen proposes to look afresh at the functional centrality of cities in the global economy, and she does so by focusing upon the attraction of producer service firms to major cities that offer knowledge rich and technology enabled environments. In the 1980s and 1990s, many such service firms followed their global clients to become important multinational corporations in their own right, albeit that service firms tend to be more susceptible to the agglomeration economies offered by city locations. These emerging producer service complexes are at the root of her new concept of urban centrality in the global economy: global cities. Sassen's conceptualization thus advocates a shift of attention to the advanced servicing of worldwide production, which implies a shift in focus from formal command power in the world system to the day to day practice of global control: the work of producing and reproducing the organization and management of a global production system and a global marketplace for finance.

Although New York, London, and Tokyo do not exhaust the framework for research on global cities, Sassen emphasizes that just a handful of cities can be said to have a new strategic role that is a product of them acting as the most important centers for financial markets, the production of financial innovations, and outsourced specialized business services. The global city network, then, is primarily created trough the transnational, city centered spatial strategies of producer service firms, an idea later picked up and extended by Peter Taylor and his colleagues of the Globalization and World cities group and network (GaWC). Their contention is that such producer services firms have created worldwide office networks covering major cities in most or all world regions, and it is exactly the myriad of connections between these service complexes that give way to the formation of a transnational urban network. In contrast to Friedmann, Sassen argues that this network results in a geography of centrality that may well cut across existing territorial core/periphery divides (**Figure 3**).

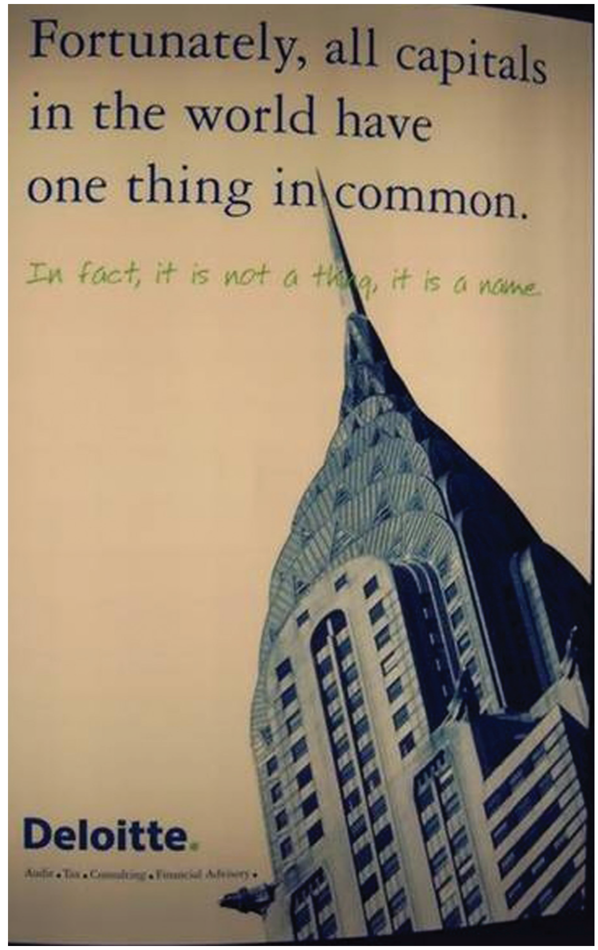

Figure 3 Global city-orientated location strategies of producer services firms.

Battling the Lack of Empirical Evidence

A key research theme in the research on world/global cities has been the empirical identification of these cities (**Table 1**). At one level, this simply relates to the perennial (and possibly irrelevant) discussion on which cities may or may not be considered to be 'global cities' or 'world cities'. But there is a broader and more important theme to the empirical literature, in that the problems encountered in such exercises bring to light the apparently inevitable data problems when doing empirically sustained globalization research. The most commonly voiced problem is the paucity of data on actual transnational relations between cities: few of the available data sources on cities reveal anything about the flows and interdependencies that lie at the heart of this literature.

In the past few years, however, there has been a proliferation of empirical studies that have explicitly sought to rectify this empirical conundrum. Researchers have thereby relied on a wide variety of data sources to gauge transnational intercity relations. Generally

Table 1 Main empirical approaches in world/global cities research

Key agents	*Corporate organization*		*Infrastructure*	
	Producer service firms	*Multinational enterprises*	*Telecommunications*	*Physical transportation*
Indicator of transnational inter-city relations	Surrogate measure for flows of information, knowledge, directions,… between two cities	Number of corporate headquarters subsidiary ties between two cities	Internet backbone bandwidth between two cities	Number of business passengers traveling between two cities

speaking, the production of these measures has been premised upon two foundations, which can respectively be labeled: (1) the corporate organization and (2) the infrastructure approach. Further distinctions can be made on the basis of the types of firms or infrastructures involved.

The corporate organization approach for measuring relations between key cities in the global economy starts from the observation that these relations are primarily created by firms pursuing transnational location strat egies. A leading example is the research pursued by GaWC researchers, which starts from Sassen's assump tion that advanced producer service firms 'interlock' cities through their intrafirm communications of infor mation, knowledge, plans, directions, advice, etc., to create a network of global service centers.

The infrastructure approach, in turn, focuses on a series of enabling infrastructures that underpin border crossing urban networks. The gist of this approach is the observation that advanced telecommunication and transportation infrastructures are unquestionably tied to key cities in the global economy: the most important cities also harbor the most important airports, while the extensive fiber backbone networks that support the Internet have equally been deployed within and between major cities. These enabling (tele)communication and transportation networks are the fundaments on which the connectivity of key cities is built, and it is therefore no surprise that the geography of these networks has been used to invoke a spatial imagery of a transnational urban network (**Figures 4** and **5**).

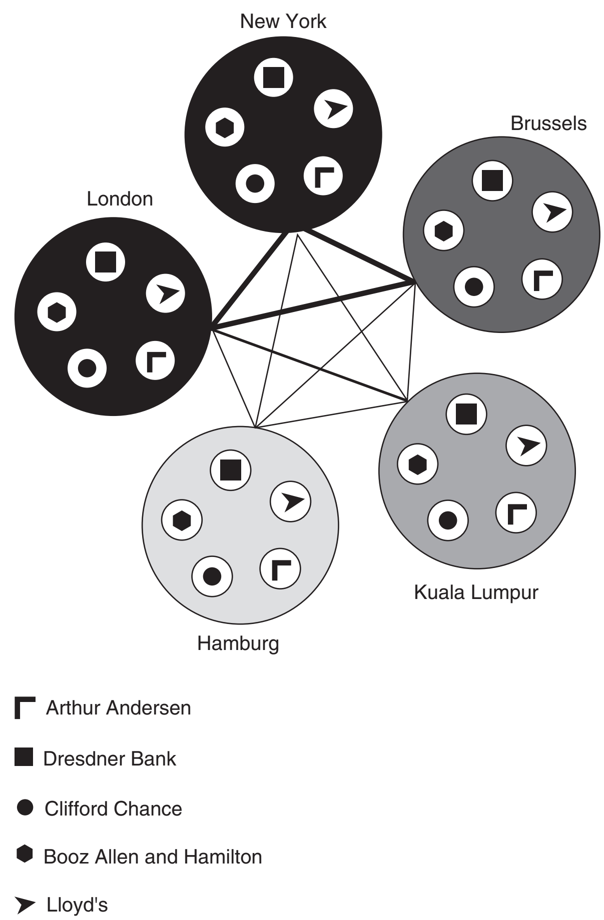

Figure 4 Producer services firms as 'interlockers' in the network of global cities.

Critiques and Ensuing Research Agendas

Virtually all of the commonly voiced assumptions in mainstream world/global cities research have been criticized from different quarters. Four fundamental criticisms stand out, the first being Hamnett like critiques of Sassen's 'polarization thesis'. The second critique is that the world/global cities literature has led to a dis proportionate attention for a small number of metro politan centers to concomitant neglect of all other cities. The most trenchant critique along these lines is by Jenny Robinson, who has complained that, because of the dominant approaches in global/world cities research, millions of people and hundreds of cities are dropped off the map in urban studies. This exclusion is from two 'maps': (1) the geographical map of world/global cities wherein most cities in the 'South' are missing and (2) the conceptual map of world/global cities which focuses on a narrow range of global economic processes so that myriad other connections between cities are missing. Robinson points out that all cities experience con temporary global processes in one way or the other, so that globalization cannot be construed as affecting just a few privileged cities. This critique has led to a broad ening of the scope of the literature, as exemplified by the postulation of such ideas as 'globalizing cities' or 'cities in globalization'.

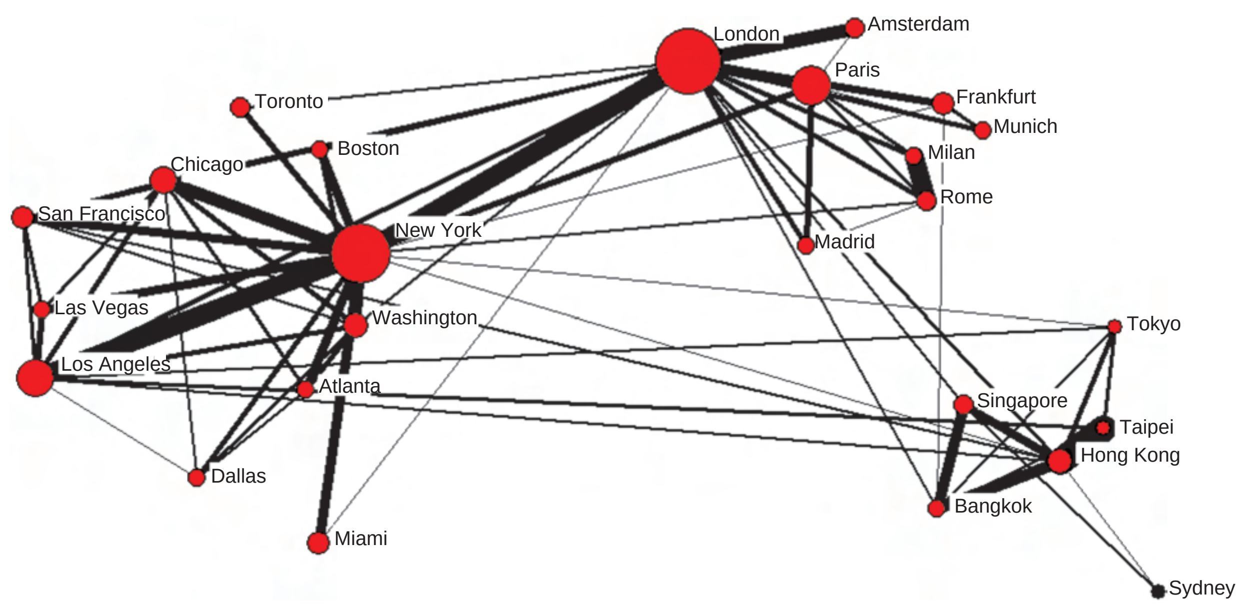

Figure 5 Airline networks as enabling infrastructures.

The third critique relates to its rather underdeveloped urban theoretical underpinnings. The dominant arguments in this research field developed from efforts to try and make sense of contemporary economic processes – recall that the two seminal contributions are by Friedmann figuring out implications of the 'new international division of labor', and by Sassen trying to understand 'the composition of globalization'. Neither Friedmann nor Sassen begins with, or develops, a theory of cities *per se*. What cities are and how they relate to one another are questions that are left unexamined except as a vague hierarchical premise. It is the latter that leads to the privileging of the few that is criticized by Robinson. Peter Taylor has recently argued that this situation can be rectified through a confrontation of some of the ideas developed by Sassen with those of Jane Jacobs, who treats cities as key economic entities. For Jacobs, vibrant cities expand economic life in ways that diversify economic processes within the city that in turn lead to complex relations with other cities. Following this line of argument, Taylor argues that the use of advanced producer services to define a network of world/global cities does not necessarily mean that they encompass most or even a sizeable proportion of the myriad complex flows between cities, but rather that they are, as cutting edge industries, critical indicators of vibrancy in the Jacobs' sense: where there are concentrations of advanced producer services, there is manifest expansion of economic life, and they are therefore integral to 'cityness'.

The fourth major critique relates to the underdeveloped attention for the continued importance of territorial states in globalization processes. An implicit assumption of some world/global cities research is that cities and states are emblematic of the two types of spaces that geographers deal with, networks and territoriality respectively. It is assumed that networks of cities and mosaics of states depict two very different ways of imagining the spatial organization of the world, forcing a choice between city centric and state centric positions. As a consequence, it is often believed that world/global cities research subscribes to the view that the demise of national state power is imminent due to the purportedly borderless, politically uncontrollable forces of global economic integration. Although some critics have caricatured the positions of world/global cities researchers in this debate, the question of state involvement on various scales has indeed largely been neglected. One of the scholars that has moved beyond this stalemate is Neil Brenner, who interprets world/global city formation as one among many tendencies which can be identified in contemporary social change. Focusing on Western Europe, Brenner shows that during the last decade national states are being qualitatively transformed (rather than dismantled); wedding state theory with the basic tenets of world/global cities research, he positions the latter within the fundamental rescaling of state spatial politics, thus allowing for a more sensitive treatment of 'the state' than is commonly the case in world/global cities research.

See also: Globalization and Transnational Corporations; Networks, Urban; Taylor, P.; Transport and Globalization.

Further Reading

Abu Lughod, J. L. (1999). *New York, Chicago, Los Angeles: America's Global Cities*. Minneapolis, MN: University of Minnesota Press.

Beaverstock, J. V., Smith, R. G. and Taylor, P. J. (2000). World city network: A new metageography? *Annals of the Association of American Geographers* 90, 123 134.

Brenner, N. (2004). *New State Spaces: Urban Governance and the Rescaling of Statehood*. Oxford: Oxford University Press.

Brenner, N. and Keil, R. (eds.) (2006). *The Global Cities Reader*. Routledge: London.

Castells, M. (2000). *The Rise of the Network Society* (2nd edn.). Oxford: Blackwell.

Derudder, B. (2006). On conceptual confusion in empirical analyses of a transnational urban network. *Urban Studies* 43, 2027 2046.

Derudder, B. and Witlox, F. (2005). An appraisal of the use of airline data in assessing the world city network: A research note on data. *Urban Studies* 42, 2371 2388.

Friedmann, J. (1986). The world city hypothesis. *Development and Change* 17, 69 83.

Hamnett, C. (2003). *Unequal City: London in the Global Arena*. London: Routledge.

Knox, P. L. (2002). World cities and the organization of global space. In Johnston, R. J., Taylor, P. J. & Watts, M. J. (eds.) *Geographies of Global Change* (2nd edn.) pp 328 338. Oxford: Blackwell.

Marcuse, P. and van Kempen, R. (eds.) (2002). *Globalizing Cities: A New Spatial Order?* Oxford: Blackwell.

McCann, E. J. (2004). Urban political economy beyond the ‘global city’. *Urban Studies* 41, 2315 2333.

Robinson, J. (2002). Global and world cities: A view from off the map. *International Journal of Urban and Regional Research* 26, 531 554.

Samers, M. (2002). Immigration and the global city hypothesis: Towards an alternative research agenda. *International Journal of Urban and Regional Research* 26, 389 403.

Sassen, S. (2001). *The Global City: New York, London, Tokyo* (2nd edn.). Princeton, NJ: Princeton University Press.

Sassen, S. (ed.) (2002). *Global Networks, Linked Cities*. London: Routledge.

Smith, N. (2002). New globalism, new urbanism: Gentrification as global urban strategy. *Antipode* 34, 434 457.

Taylor, P. J. (2004). *World City Network: A Global Urban Analysis*. London: Routledge.

Taylor, P. J., Derudder, B., Saey, P. and Witlox, F. (eds.) (2007). *Cities in Globalization: Practices, Policies, and Theories*. London: Routledge.

Relevant Websites

http://www.lboro.ac.uk/gawc
GaWC, Loughborough University.

World-System

K. Terlouw, Utrecht University, Utrecht, The Netherlands

Glossary

Antisystemic Movements These are both social labor movements and nationalist anticolonial movements, that want to replace the modern world-system with a more equitable social system.

Hegemony This is the temporary situation when the combination of economic, financial, political, and military superiority gives one state a dominant position in the world-system.

Luxury Goods These were exchanged between social systems, but neither was their production organized for this trade, nor their consumption had a significant influence on the functioning of the other social system.

Mini-Systems Small social systems, where cultural, political, and economic boundaries mostly overlap.

Semiperiphery States and other territorial entities that are in between the core and periphery. They trade core-like products with the periphery and peripheral products with the core. They have the most active states that intervene in the economy to improve their position in the world-system further.

Semiproletarian Households These households pool multiple income sources from several members over a long period. Part of these income sources are market related and are part of the world-system. Another part of these income sources are based on self-support, like subsistence farming. Through this partial self-sufficiency they can survive despite the low income from the market sector. The members of the household are not necessarily family and do not necessarily live together, but frequently do both.

TimeSpace Time and space are not separate entities, but are linked. For every kind of social time, there exists a particular kind of social space. The eternal TimeSpaces are explanations regardless of time and space. Structural TimeSpaces have their own dynamic of social change, like our world-system. Transformational TimeSpaces are the unique and infrequent episodes when one structural TimeSpace replaces another. There are many, partly overlapping cyclico-ideological TimeSpaces through which the world-system functions. Episodic TimeSpaces are the specific place and moment of very visible events that cannot be analyzed without the other TimeSpaces.

World-Empire Centrally controlled area, with a spatial division of labor based on political power between different regions with different cultures.

Introduction

Immanuel Wallerstein formulated the world systems approach in the 1970s as an alternative to state centered approaches to development. Both the then dominant modernization theory and Marxism equated state with society and studied social change from a state centered evolutionary perspective. Dependency theorists like Raúl Prebisch and André Gunder Frank criticized this and explained the lack of development in peripheral states through their unequal relations with core states. The world systems approach builds on this, but shifts the unit of analysis to the total network of these relations. This world system has its own dynamic; it is more than the sum of its parts. Globalization is the newest phase of the development of this centuries old world system. However, most participants in the globalization debate are unwilling to shift the unit of analysis to the global scale. They analyze the external challenges to the way of life of communities, without making them parts of a global system with a dynamic of its own (**Figure 1**).

The Modern World-System as a Whole: Rise and Demise

Understanding our present world system requires ana lyzing its emergence in the sixteenth century and its development over the centuries. After sketching the historical development of the present world system, the subsequent sections theorize on the world system as a whole and its constituent parts.

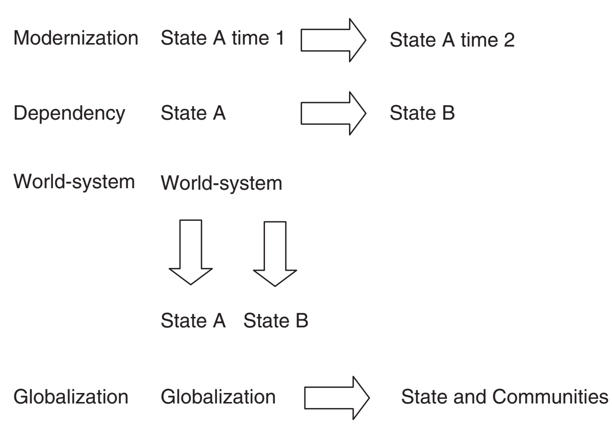

Figure 1 Different ways to explain social change.

The Emerging World-System

The world system is not a recent phenomenon, but emerged in the sixteenth century. The core in Northwest Europe profited from trade in agricultural products with Eastern Europe (grain and timber) and the Americas (gold, silver, and sugar). This enabled their economies to concentrate on highly profitable industries. The development in the peripheries suffered from their specialization in agricultural commodities with low profit margins. The relations with Northwest Europe also hindered the emergence of an independent class of entrepreneurs and weakened the peripheral states.

Asia and Africa were then still outside the world system. The trade with Asia was in luxury goods whose production in Asia and consumption in Europe did not significantly influence the functioning of both economies. Not the world system, but the Chinese empire determined the terms of trade for silk and porcelain, forcing the Europeans to pay with silver and gold.

Stagnation and Consolidation

After its emergence in the long sixteenth century (±1450–1650), our world system stagnated. Production and trade diminished. Although regional self sufficiency increased and the international division of labor contracted, the capitalist world economy did not disappear. The many wars in this period (±1600–1750) strengthened the core states and institutionalized the interstate system of competing sovereign states. This consolidated the modern world system and created the conditions for a renewed expansion in the next period, for instance, by integrating their territories into national markets (**Figure 2**).

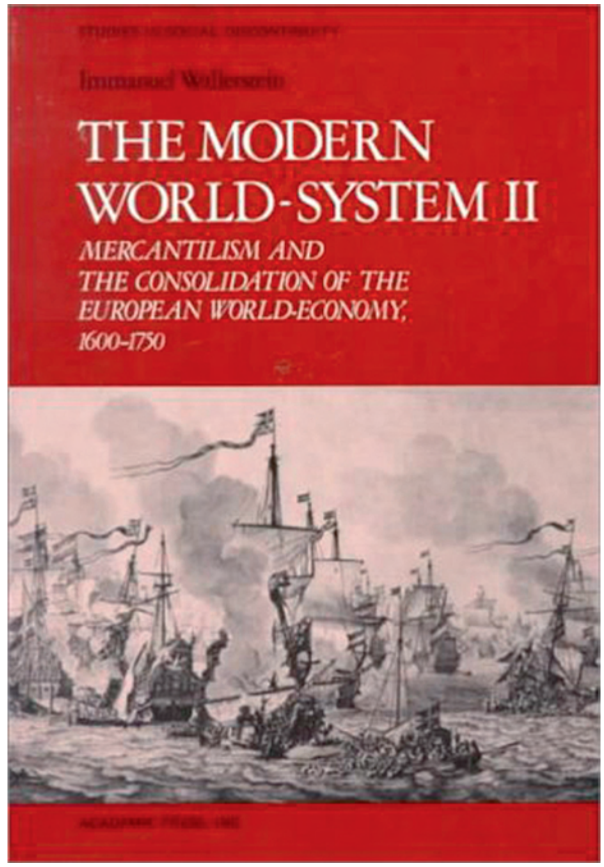

Figure 2 Frontcover. Wallerstein, I. (1980). *The Modern World-System II*. New York: Academic Press.

Expansion and Intensification

The revival of trade from the mid eighteenth century onwards signaled a period of renewed expansion of the world system. It expanded and finally incorporated the entire globe at the end of this period (±1900). Europe's military power opened up Asian economies and imposed favorable terms of trade. The world system expanded in Africa through colonialism.

This spatial expansion went hand in hand with the further intensification of the world system. The economic specialization matured to such a degree that many characterized it as an industrial revolution. Although agriculture no longer dominated the economy and the role of wage laborers increased, there was no fundamental break in the functioning of the world system. The fortunes of different groups and areas within the world system changed, but the working of the market based capitalist world economy remained the same.

Consolidation and Globalization

After its glorious nineteenth century, the world system entered into a more challenging period. Economic stagnations and World Wars characterized the first half of the twentieth century. The spatial division of labor changed. The European core states lost their dominant position in the world system to new core states. The center of the world system shifted further westward. The USA and later Japan overtook the West European economies.

More fundamental than this shift in focus within the world system were the socialist revolutions in Russia and China. They challenged the operating principles of the world system and tried to create an alternative system. Despite their failure, the strengthening of their state helped them to improve their position in the world system.

De colonization also initially undermined the political dominance of the core states. However, the loosening of the privileged links between individual core states and peripheries strengthened the economic integration of the world system. The more recent globalization further intensified integration and spread development toward some parts of the former periphery in Asia.

The Modern World-System Compared to Others

Our world system is not the only possible world system. According to Wallerstein there existed other social sys tems in history. Others question the adequacy of Wal lerstein's approach to our society. According to some it is much older than the sixteenth century, while others divide it into different world systems.

Wallerstein's Historical Systems

Mini systems are very small social systems, with a single economy, culture, and polity. They exhibit hardly any division of labor. These 'tribal' societies have now vir tually disappeared.

World systems are obviously much larger than mini systems, but only our system spans the globe. World systems have more developed economies. Social in equalities increase and a spatial division of labor between core and peripheral areas is established. Cultural differ ences institutionalize these inequalities between regions.

The hyphen indicates that world systems are not sys tems 'of' the world, but systems that 'are' a world con taining all essential processes for social change. Only in mini systems, do economic, political, and cultural bound aries overlap. Wallerstein chooses economic exchanges to delimit social systems, because these relations have the biggest influence on other spheres of human activity.

World empires and world economies are types of world systems which differ in political structure. World empires were controlled from the center, whereas a world economy is economically integrated, but politically divided. Political redistribution characterizes world empires, while market forces regulate a world economy. The Roman, Chinese, Ottoman, and Inca empires are ex amples of world empires. Our world system is the most successful example of a world economy (**Table 1**).

Table 1 Characteristics of different historical systems

	Mini system	*World system*	
		World empire	*World economy*
Economy	Economic unit	Economic unit	Economic unit
Polity	Political unit	Political unit, with centralized authority	Competing
Culture	Homogeneous culture	Encompasses several cultures	Big cultural differences
Social division of labor	Slight social division of labor	Social division of labor, mainly between rulers and peasants	Social division of labor, mainly between employers and proletariat
Spatial division of labor	No spatial division of labor	Spatial division of labor, mainly between city and countryside	Spatial division of labor between core and periphery
Area	Very small	Big	Big

Source: Terlouw, K. (1992). *The Regional Geography of the World System: External Arena, Periphery, Semiperiphery, Core*, p 55. Utrecht: KNAG.

Other Types of World-Systems

Wallerstein's choice for an economic division of labor to delimit world systems is disputed **Figure 4**). Instead of an enduring dense economic network, irregular and one sided political and cultural influences can also de cisively affect social change.

André Gunder Frank argued that the world system emerged 5000 years ago. Then, for the first time in his tory, systematic plunder of surplus emerged in Meso potamia and Egypt. These, like most subsequent civilizations, heavily influenced each other. Culturally, they learned from each other and politically they com peted. In this millennial systemic connection between Asia, Europe, and Africa, Asia was dominant except for a short interlude in the nineteenth and early twentieth century. This very comprehensive world system draws attention to the neglected early instances of global in fluences. However, combing the whole of history into a single world system neglects the consequences of the unprecedented intensity of the more recent worldwide relations **Figure 3**).

Fernand Braudel deconstructs Wallerstein's single capitalist world system into a succession of world sys tems and gives capitalism a much more limited role. Capitalism coexists with other modes of production in a world system. Braudel employs a hierarchical model of material production. At the lowest level, he places the autarchic subsistence economy of households. Capitalism never really penetrated these regionally rooted ways of life. The trade between different regional markets is the second layer. Not the market, but the law of the jungle governs the layer of capitalism on top of this market economy. Capitalism never organizes the whole economy, but dominates the economy from strategic points, con trolling key sectors of accumulation through monopolies **Figure 5**).

Already in the late Middle Ages a capitalist world system emerged around Venice according to Braudel. After its crisis, a new Antwerp based world system emerged, that restructured existing, rather than created new, trade relations. The same happened in subsequent

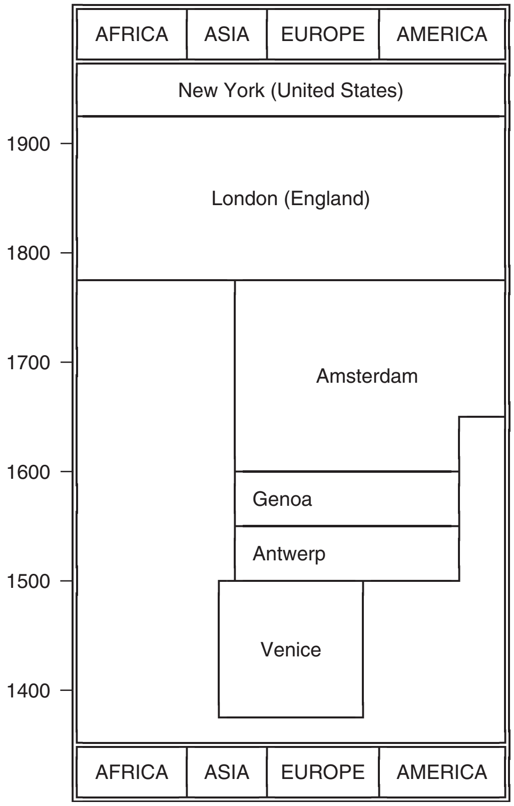

Figure 3 Frank's extended world-system in space and time. Source: Terlouw, K. (1992). *The Regional Geography of the World-System: External Arena, Periphery, Semiperiphery, Core*, p 62. Utrecht: KNAG.

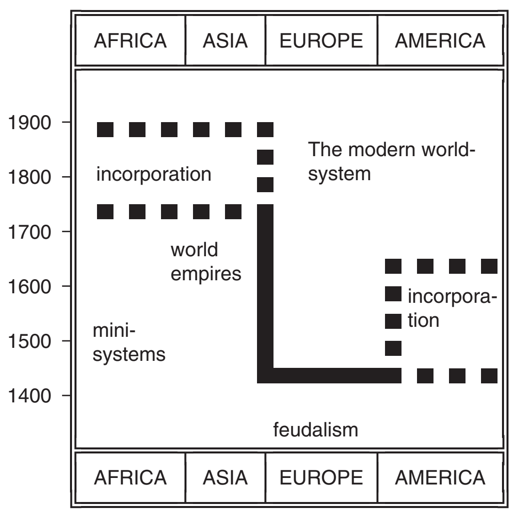

Figure 4 Wallerstein's world-system in space and time. Source: Terlouw, K. (1992). *The Regional Geography of the World-System: External Arena, Periphery, Semiperiphery, Core*, p 56. Utrecht: KNAG.

transitions between world systems. Each had a different spatial structure and organization. Every new world system emerged around a new central city. Their relative power over the world system differed. England was the

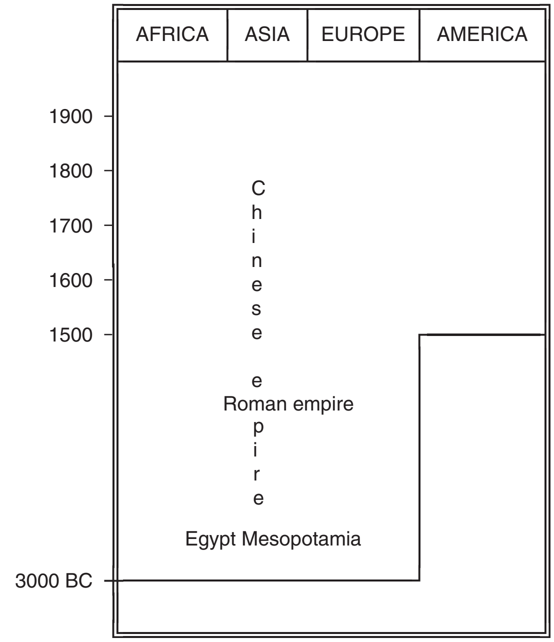

Figure 5 Braudel's world-systems and their core in space and time. Source: Terlouw, K. (1992). *The Regional Geography of the World-System: External Arena, Periphery, Semiperiphery, Core*, p 53. Utrecht: KNAG.

most dominant. Genoa was the least conspicuous; its control over the world system was restricted to its dominant position in the credit system. In contrast to previous world systems, those organized from London and New York based their dominance upon the competitive industries in their region (**Table 2**).

Recent research on world cities is related to Braudel's conceptualization of world systems. Especially Peter Taylor's theoretical and empirical work on the global network of cities provides new insights into a worldwide system with not states but cities as its building blocks.

TimeSpaces

Time and space are essential for the functioning of the world system. These are not separate entities, but form one dimension. The concept TimeSpace expresses this connection. After discussing the different TimeSpaces, we briefly compare the different conceptualizations of world systems discussed above.

Characteristic for the 'eternal TimeSpace' are explanations regardless of time and space. This search for general laws of behavior has until recently dominated the social sciences. They conceptualized social change as eternal progress everywhere. Time and space were irrelevant for them.

Space and time in this Newtonian worldview were only the here and now of observable facts caused by outside forces. The specific place and moment form the

Table 2 Braudel's world systems and their cores

City	*Region*	*Period*
Venice		1378 1498
Antwerp		1500 1569
Genoa		1557 1627
Amsterdam		1585 1773
London	England	1773 1929
New York	United States	1929 present

Source: Terlouw, K. (1992). *The Regional Geography of the World System: External Arena, Periphery, Semiperiphery, Core*, p 53. Utrecht: KNAG.

'episodic TimeSpace'. The less scientific humanities studied these events. History studied time as the events of the past, while geography studied space as unique and separate regions.

The world systems approach criticizes the search for universal laws, but also wants to do more than just ana lyzing particular events and places. It therefore focuses on the TimeSpaces between these extremes. This vio lates the traditional disciplinary divisions. Not sur prisingly, the world systems approach has been severely criticized from within these disciplines. Historians criti cize Wallerstein's use of historical sources and sweeping generalizations. Economists criticize the negative role ascribed to international trade for the periphery and focus on internal obstacles for economic development. However, geographers are more interdisciplinary and, for instance, incorporate parts of world system thinking into their textbooks.

The world systems approach focuses on the inter mediate TimeSpaces. Each world system forms a 'structural TimeSpace', with fundamentally different operating principles and developmental path. Expanding borders and a developing spatial division of labor trans forming the global environment are spatial aspects of the structural TimeSpace of our capitalist world system. These structures are quite persistent and change only gradually.

A specific structural TimeSpace defines a world sys tem; it functions through successive 'cyclico ideological TimeSpaces'. As its composite name suggests, these are diverse and partly overlapping. It combines cyclical time with ideological space. Cycles are, for instance, the four periods discussed above, or the economic and political long waves discussed below. Through their repetition, the world system functions and develops in its structural TimeSpace. For instance, the structural competition be tween states generates cycles with alternating periods of rivalry and peace, which structurally increase state power over time. Spatial and ideological divisions like those between East and West during the cold world war and between North and South during de colonization char acterize different cyclico ideological TimeSpaces. The rise and decline of specific regions and states, under circumstances that differ between period and zone, also shape these TimeSpaces.

The transition from one structural TimeSpace to another constitutes a 'transformational TimeSpace'. They are unique occurrences at the right time and place when one structural TimeSpace succeeds another. These are the rare moments when free will can determine the fu ture organization of society. Our world system was the new structural TimeSpace that emerged out of the crisis of feudalism in Northwestern Europe. According to Wallerstein, the current crisis of our world system her alds the beginning of a new transformational TimeSpace.

The differences between the world systems discussed above relate to these TimeSpaces. Wallerstein's world system can also be seen as a cyclico ideological Time Space within Frank's world system, while Frank's world system comes close to an eternal TimeSpace from Wallerstein's perspective. Braudel's world systems are cyclico ideological TimeSpaces within Wallerstein's world system.

The Functioning of the Modern World-System

Unequal exchange between core and periphery, with different socioeconomic processes, regulated by the strong core states characterizes the structural TimeSpace of the world system. The changing location and char acter of core–periphery relations are part of the cyclico ideological TimeSpaces within the world system.

Core and Periphery

The exploitation of the periphery by the core character izes the division of labor within the modern world system. In the core, high wages, advanced technology, and a di versified production dominates. In contrast, the periphery has low wages, rudimentary technology, and a simple production mix. In world trade, low priced peripheral products are exchanged with high priced products from the core. The strength of core states enforces this unequal exchange.

The initial differences in wealth in the agricultural division of labor were relatively small, but later developed into huge gaps. The spatial division of labor between core and periphery is always present in the structural Time Space of the world system. The types of goods exchanged vary between different cyclico ideological TimeSpaces. After the initial, mostly agricultural, division of labor came the unequal exchange between industrial products and peripheral raw materials. Nowadays, many industrial products come from the periphery, and the core profits from its dominance of the high tech service sector.

The exploitation of the periphery by the core is also based on a different class structure. Wages in the periphery can be so low, while workers as members of larger income sharing households have other means of subsistence. These semiproletarian households augment income from the market with other sources, like subsistence farming. In addition, migration of young workers from rural regions to cities shifts the reproduction costs away from the com mercialized sector in the cities to their rural homeland. Semiproletarian households extend beyond the living con ditions of the here and now; they span generations and extend in space.

The workers in the core use their much stronger bargaining position to increase their pay. Their income is large enough to sustain their households over time. **Figure 6** depicts the relation between exploitation be tween classes (vertical), and unequal exchange in the modern world system (horizontal).

The spatial separation of core and periphery charac terizes our world system. Spatial preconditions, like re sources and other production potentials, were initially important. Spatial separation facilitated the development of the different production structures in core and pe riphery. The different class structures are part of this. The spatial coincidence of social divisions with differ ences in wealth hinders the development of a class based consciousness and helps the interpretation of spatial differences in racist terms. While the increased trans portation opportunities make distance less relevant for the spatial separation between core and periphery, core states increasingly use border controls to hinder migration to their more profitable labor markets. Although changing in character over time, the spatial separation of core and periphery secures the world system (**Figures 7–9**).

The semiperiphery also stabilizes our world system. Semiperipheral states depolarize the relation between core and periphery. The possibility of joining the core appeases the semiperiphery. Its political and economic power is clearly subordinate to the core, but unlike the periphery, it has the possibility of resisting exploitation. In many core states, institutional fixes and congestion stifle development. In the periphery, the absence of good government, services, and skilled labor hamper devel opment. The semiperiphery maximizes the possibility and necessity for development giving it the strongest motivation for state intervention in the economy. Semi peripheral states are the most active in transforming their socioeconomic structure and in protecting their economy from external competition.

Economic cycles generate changes in the location of core and periphery and the products exchanged. These cycles of about 50 years have their origins in the way supply and demand are determined in the modern world system. The constant drive for more profits of individual firms causes production to rise continually. The demand for goods is the result of a long term compromise be tween labor and capital in the nation states. In a few decades, the steadily growing supply of goods reaches the limits of the relatively constant demand for goods. This results in decades of economic stagnation.

Reducing production costs is the first measure in combating the declining profits during an economic stag nation. Increasing workload and pay cuts reduce labor costs. Transferring production from the core to regions with a lower wage level in the semiperiphery also lowers

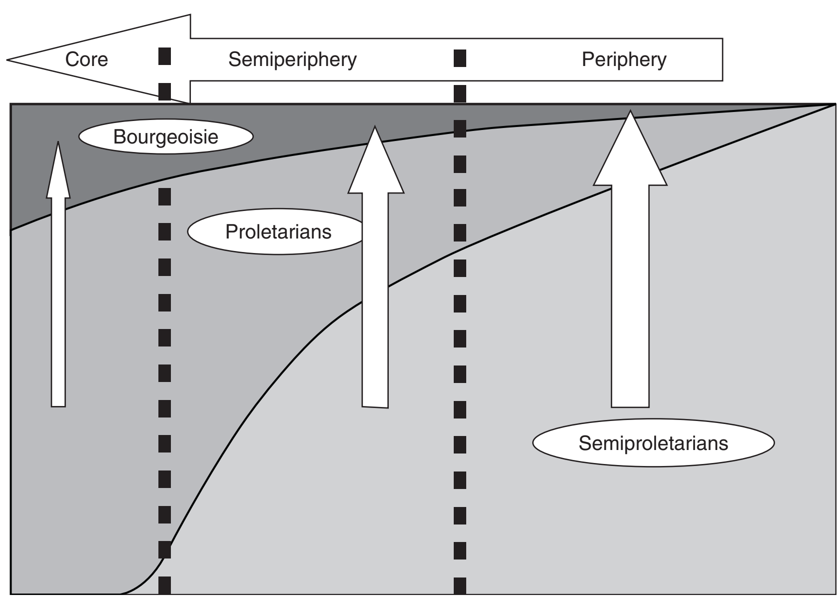

Figure 6 Exploitation in the modern world-system.

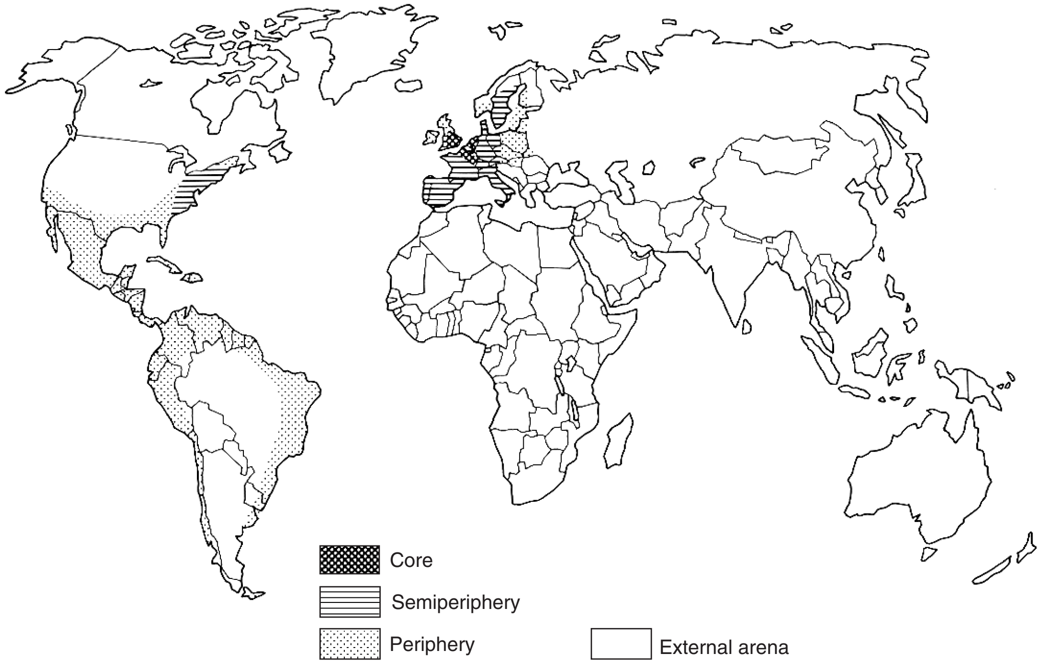

Figure 7 The world-system. According to Wallerstein as constructed by Terlouw, C. P. (1985). *Het Wereldsysteem: Een Interpretatie van het Werk van I.M.* Wallerstein. Rotterdam: CASP, p 44.

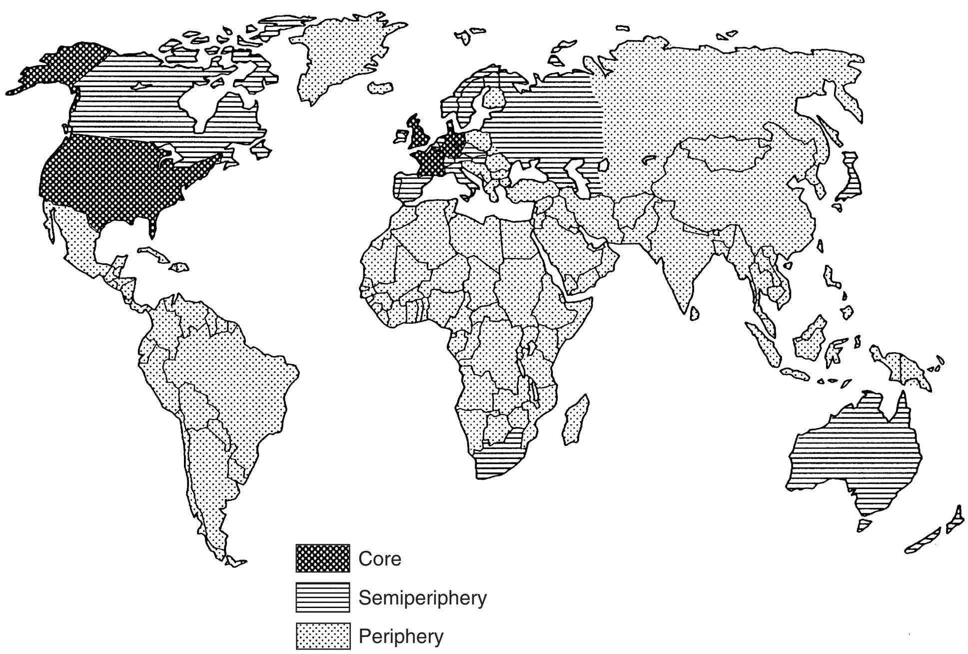

Figure 8 The world-system: *c.* 1900. According to Wallerstein as constructed by Terlouw, C. P. (1985). *Het Wereldsysteem: Een Interpretatie van het Werk van I.M.* Wallerstein. Rotterdam: CASP, p 45.

costs. This also cuts the more hidden production costs, like environmental protection measures. This combination of lax regulation and strong developmental pressures makes the semiperiphery vulnerable to social, cultural, political, and ecological contradictions. It makes the semiperiphery the most dynamic part of the world system.

In the core, the development of new technologies increases in the later phases of an economic stagnation. New industries with high profit levels restore the position of the core in the world system.

However, all these measures are only successful in the short run and for specific areas in the world system. Only

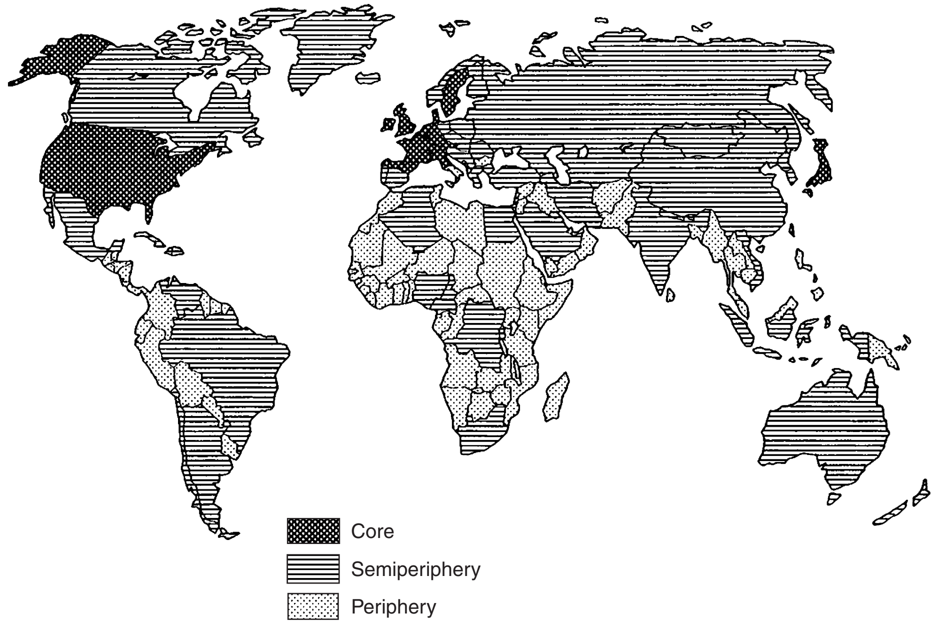

Figure 9 The world-system: *c*. 1980. According to Wallerstein as constructed by Terlouw, C. P. (1985). *Het Wereldsysteem: Een Interpretatie van het Werk van I.M.* Wallerstein. Rotterdam: CASP, p 46.

expanding worldwide demand can overcome an eco nomic stagnation. For instance, the spread of wealth away from the core revives economic growth. As discussed above, the production of traditional core products shifts toward the semiperiphery. This stimulates the growth of the proletariat at the expense of the semiproletariat.

In the past, incorporating new areas restored the profitability of the spatial division of labor for the core by adding new peripheries. This is no longer possible in our global world system. The world system is now entering a structural crisis due to the success of the market economy in expanding in space and into social relations.

The State and the World-System

States are crucial for the functioning of the modern world system. The Habsburgs' attempt to transform the emerging world economy into a world empire was de feated by other states. In the Treaty of Westphalia of 1648, states mutually acknowledged their sovereignty within their territory. This institutionalized an inter national system of states that provides the necessary freedom and support for capitalism to develop.

The state guarantees property rights and regulates contracts. This gives entrepreneurs the necessary security to operate and motivates them to reinvest their profits. Strong states can further effectively support their entre preneurs in the competition with other core states and in the exploitation of the periphery. Entrepreneurs in un successful states are pushed away from the profitable core business and languish in peripheral activities or relocate to stronger core states. This further improves the economy in the core, which gives them the resources to increase their state power. Economic and political strength thus reinforce each other giving the state an ever increasing importance.

Political geographers criticize this focus on the state by the world system approach. Political geography's focus on power relations in space shifted from the state to other scales, including the world system. Peter Taylor pioneered this shift in the focus of political geography away from the nation state. He also stressed the im portance of the subnational scale. Political geographers increasingly study cities and regions, and their changing relations to other scales. These are not just lower ad ministrative levels of government, but involve a wide range of urban and regional actors outside the state. The worldwide network of world cities and urban regions and the local or regional embedding of global forces are now often studied.

Hegemony

States always compete with each other. Sometimes the warfare is so intense and widespread that it escalates into world wars. These end with a peace agreement (1648 treaty of Westphalia, 1815 Vienna congress, 1945 Yalta confer ence), which heralds a new period of relative peace under the hegemony of a leading member of the winning coali tion not directly affected by warfare. This was the case for the United Provinces in the mid seventeenth, the United Kingdom in the mid nineteenth, and the United States in the mid twentieth century. Despite the differences between

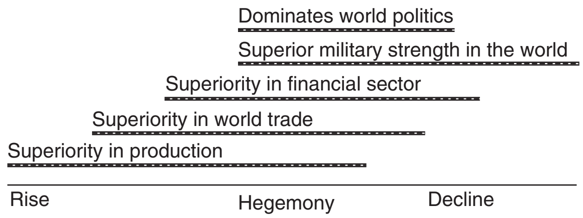

Figure 10 Hegemony: rise and decline.

these three states, they all dominated the world system in a similar way. All hegemonic states had the most advanced economies of their time. Their entrepreneurs were even competitive on the home markets of their rivals. Besides production and trade, the hegemonic state is also superior in the financial sector. It has the highest rate of return on capital and it lends money to others. Consequently, a hegemonic state also dominates world politics. The dominant, though not necessarily unrivalled, military strength of the hegemonic state, especially in sea power, is instrumental in maintaining its dominant position in world politics (**Figure 10**).

A hegemonic state makes the world economy operate more freely. Through its economic superiority, a hegemonic state profits from the unrestrained functioning of the world market. The hegemonic state cracks down on states that try to compensate their economic weakness through state intervention. A period of hegemony is therefore a period of relative free trade.

Nevertheless, hegemony is not a stable condition of the world system. In decades, other core states undermine the dominant economic position of the hegemonic state. First, the liberal ideology of the hegemonic state allows the new technologies developed in the hegemonic state to spread to other states. Second, the production costs in the hegemonic state rise because the workers negotiate increasingly better deals with the bourgeoisie in the hegemonic state, who appease their workers in order to capitalize on their hegemonic advantage over other states.

The decline of hegemony becomes first apparent in production. This was the case in the late 1960s for the USA. Through their multinationals, the USA has dominated world trade much longer. Financially the position of the USA is still strong, but no longer dominant. The power of the US military is still unrivalled, but without the support of the other aspects of hegemony and the increasing strength of other states, it is less successful politically as before.

Increasing State Power

The competition between states not only structures the world system and cycles between rivalry and hegemony, it also generates a trend of increasing state power. Like the expansion of the market economy, the outcomes of these cyclico ideological TimeSpaces undermine the structural TimeSpace of the world system and herald its transition into a transformational TimeSpace. The escalating destructive force of military power reflects this trend in international relations. Our world will not survive a world war with nuclear weapons. The state also becomes too powerful within its territory.

States always intervened in their economy to finance their growing state machinery. However, this growing state influence depends more and more on popular support. This transformed territorial states into nation states that could mobilize its population for war. The Napoleonic draft was a powerful innovation that helped the straggling core state France. Other states copied it later. Mobilization also involved creating a national identity through education and later by providing state welfare.

Increasing state power is also a reaction to resistance from below. In the core states, the increasing proletarization in the nineteenth century was the basis for organized workers resistance. Their initial goal was to overthrow the capitalist system. This first instance of what Wallerstein labels antisystemic movements was eventually absorbed in the world system. The goals of these social movements became reformist and focused on gaining power in individual nation states. Better pay and a welfare state appeased, after many decades of struggle, the workers in core states.

Revolutionary movements came to power in some semiperipheral states. As communist states, they partially withdrew from the world system to protect their state centered developmental strategy typical for semiperipheral states. In the periphery, anticolonial movements resisted exploitation in the world system. These national movements increasingly focused on the independence of their colony and the formation of their own state.

By focusing on state power, rather than the world system as a whole, all types of antisystemic movements adapted to the workings of the world system. They thus strengthened the world system, but only in the short term. The cost of the appeasement of exploited groups reduces the profitability of the world system for the bourgeoisie. The rising costs for wages and taxation did not hurt the bourgeoisie so much during the long period of economic growth after World War II. When that ended in the 1970s, the neoliberalism of Reagan and Thatcher tried with only partial success to redress the squeeze on the bourgeoisie. The financial speculations since the 1990s temporarily restored the profitability of the world system for the bourgeoisie.

See also: Capitalism; Core-Periphery Models; Hegemony; Modernization Theory; North-South; Spatial Division of Labor; Taylor, P.; Third World; Uneven Development; World/Global Cities.

Further Reading

Arrighi, G. (1994). *The Long Twentieth Century: Money, Power and the Origins of Our Times*. New York: Verso.

Braudel, F. (1986). *The Perspective of the World: Civilization and Capitalism 15th 18th Century*. New York: Harper & Row.

Chase Dunn, C. (1989). *Global Formation: Structures of the World Economy*. Oxford: Blackwell.

Flint, C. and Taylor, P. (2007). *Political Geography: World Economy, Nation State and Locality*. Harlow: Prentice Hall.

Frank, A. G. (1998). *Reorient: Global Economy in the Asian Age*. Los Angeles: University of California Press.

Goldstein, J. S. (1988). *Long Cycles: Prosperity and War in the Modern Age*. New Haven: Yale University Press. http://www.joshuagoldstein.com/jgcycle.htm (accessed on January 2008).

Shannon, R. T. (1989). *An Introduction to the World System Perspective*. Boulder: Westview.

Taylor, P. (2004). *World City Network: A Global Urban Analysis*. London: Routledge.

Terlouw, C. P. (1985). *Het Wereldsysteem: Een Interpretatie Van het Werk van I.M. Wallerstein*. Rotterdam: CASP.

Terlouw, K. (1992). *The Regional Geography of the World System: External Arena, Periphery, Semiperiphery, Core*. Utrecht: KNAG.

Terlouw, K. (2002). The semiperipheral space in the world system. *Review, A Journal of the Fernand Braudel Center* 25, 1 22.

Wallerstein, I. (1974). *The Modern World System: Capitalist Agriculture and the Origins of the European World Economy in the Sixteenth Century*. New York: Academic Press.

Wallerstein, I. (1980). *The Modern World System II: Mercantilism and the Consolidation of the European World Economy 1600 1750*. New York: Academic Press.

Wallerstein, I. (1989). *The Modern World System III: The Second Era of Great Expansion of the Capitalist World Economy 1730 1840s*. New York: Academic Press.

Wallerstein, I. (2000). *The Essential Wallerstein*. New York: New Press.

Wallerstein, I. (2004). *World Systems Analysis: An Introduction*. Durham: Duke University Press.

Relevant Websites

http://www.binghamton.edu/fbu
Fernand Braudel Cente, Binghamton University.

http://www.lboro.ac.uk/gawc
Globalization and World Cities, Loughborough University.

http://www.irows.ucr.edu/
Institute for Research on World Systems.

http://journals.cambridge.org/action/displayJournal?jid=JGH
Journal of Global History, Cambridge Journals.

http://muse.jhu.edu/journals/jwh/
Journal of World History, Project MUSE.

http://jwsr.ucr.edu
Journal of World Systems Research.

http://www2.asanet.org/sectionpews/
Political Economy of the World System section of the American Sociological Association.

http://www.forumsocialmundial.org.br/
World Social Forum Mundial, Forum Social Mundial.

http://wsarch.ucr.edu/
World Systems Archive.

Wreford Watson, J.

G. M. Robinson, Kingston University, Kingston upon Thames, UK

Glossary

Image Region The image of a particular region or landscape held in people's minds. People's sense of place.
Mental Distance Distance as perceived in the human mind.
Social Distance The distance between social groups. Social groups may be separated by a small physical distance, but in terms of social distance there may be a huge gulf between them.
Social Region The identification of different regions of the earth's surface according to associations of social phenomena related to the total environment. How different people build their own tastes and prejudices into systems of land development and settlement.

Wreford Watson, James (1915–90)

Having graduated in 1935 from the University of Edinburgh in Scotland, Wreford Watson emigrated to Canada in the late 1930s where he became the country's Chief Geographer. He wrote widely about Canada and the United States for the next four decades. Working initially in the 1940s on Hamilton, Ontario, Watson applied ideas from the Chicago school of social ecology to gain greater understanding of what he termed "social regions." He stressed the need to recognize social distance as a major component within regional differentiation. He not only recognized distinctive social regions within cities, but also translated this concept to a larger scale when working on regional differentiation within North America. Here his work emphasized social problems associated with multiracial development, the 'energy crisis', conservation of resources, urban decay, and suburban sprawl, while also emphasizing the importance of local cultures in shaping places and landscapes. He used the language of the social ecologists to recognize urban transition zones, cultural shatter belts, and distinct zones associated with the different socioeconomic groups within Hamilton. He related divisions within society to the operation of commercial and administrative institutions that shaped the fundamental character of the city. This was part of pioneering work that helped develop social geography as a systematic specialization, evolving from the concept of social distance.

He was based back in Scotland from 1954 when he was appointed to the Chair of Geography at Edinburgh. His inaugural lecture – 'Geography: a discipline in distance' – set out his views on how regions should be properly analyzed by geographers. He stressed the need for geographers to be aware of the evolution of phenomena over time, so that the dynamics of regions could be properly understood. An integral part of this was the introduction of ideas on time distance, cost distance, and social distance as the keys to understanding spatial distributions. He argued that a vital part of geographical enquiry was to understand how these distances changed through time, and hence the conception of regions, for example, as dynamic entities. In his own work this meant engaging with the historical–cultural context, as for example in his vivid descriptions of settlement and development in Canada. Its strong cultural contrasts and rich mosaic of cultures formed the subject for Watson to study the way in which different groups built their own tastes and prejudices into systems of land development and settlement. As part of this he developed systematic work on social distance and how the influence of people's images of places and their environment was translated into 'geographical reality'.

His *North America, Its Countries and Regions* (1963) was a standard text from the regional tradition, but written in an unusual literary style and containing new insights on the social character of urban areas and the importance of people's images of place as key determinants of their decision making. Indeed, Watson retained and developed a focus upon people and places that was quite unfashionable during Geography's Quantitative Revolution. He argued that it was the minds of people that offered fertile grounds for investigation, with the concept of people's sense of place providing a focal point for research. He expressed this cogently in the opening to his book *Social Geography of the United States* (1979: 1): 'The geography of any country is what people see in it, want from it and do with it…'. Hence, he argued that geographers should be concerned with how regional differentiation was interpreted in the daily lives of ordinary citizens. In work on the remote Canadian industrial town of Kirkland Lake in Northern Ontario, he recognized the presence of alienation, injustice, and inequality within the community, in championing a social geography stretching well beyond the parameters of spatial patterns, regional boundaries, and narrow ideologies. His work contained some of the key ideas later embodied within humanistic geography: sense of place, the importance of human experience, and the use of literature as source material.

Wreford Watson also has a genuine claim to the title of 'geographer poet', having been awarded the Governor General of Canada's Medal for Poetry in 1954. In his geographical writing he drew extensively upon literature and poetry as 'natural' sources to be utilized by geog raphers. He believed in the importance of literary views of place as these underlay 'image regions' by representing mental images of place. He endeavored to show in par ticular how images from literature helped shape the views of people and how people transformed landscapes to fit a particular image. So he drew on the writings of Mark Twain, William Faulkner, Mike Fink, and John Steinbeck to substantiate the argument that freedom loving Americans built up a market landscape in which the ethos of private enterprise and individual competition dominated. He drew extensively on this literature to evoke a sense of place, claiming in his Presidential Ad dress to the Institute of British Geographers in 1983 that he had never written an article or book without an appeal to literature. This was literature as an essential geo graphical source material, and he regarded the appeal to literature as representing the 'soul' of the discipline.

Wreford Watson is remembered for the quality of his writing and ability to convey a sense of place, partly through his use of poetry and literature. He made major statements on the nature of social geography and con cepts of distance. Had he worked in prominent uni versities in the US or England rather than in Canada and Scotland, his ideas might have achieved even greater attention as they were certainly ahead of their time. His thoughts on image regions and social distances attained greater currency in the 1970s, though his abiding at traction to the notion of the region as a central pillar of geographical thought was largely a product of his pre 1939 geographical education.

See also: Literature; Regional Geography II; Sense of Place.

Further Reading

Johnston, R. J. (1993). The geographer's degree of freedom: Wreford Watson, post war progress in human geography, and the future of scholarship in UK geography. *Progress in Human Geography* 17, 319 332.

Robinson, G. M. (1991). An appreciation of James Wreford Watson, with a bibliography of his work. In Robinson, G. M. (ed.) *A Social Geography of Canada*, pp 492 506. Toronto: Dundurn Press.

Robinson, G. M. (1996). James Wreford Watson. *Geographers Biobibliographic Studies* 17, 106 116.

Watson, J. W. (1951). The sociological aspects of geography. In Taylor, G. (ed.) *Geography in the Twentieth Century: A Study of the Growth, Fields, Techniques, Aims and Trends*, pp 463 499. New York: Philosophical Library.

Watson, J. W. (1953). *General Geography*. Toronto: Copp Clark.

Watson, J. W. (1963). *North America, Its Countries and Regions* (2nd. edn. 1968). London: Longman Green.

Watson, J. W. (1968). *Canada, Its Problems and Prospects*. Don Mills, Canada: Longman.

Watson, J. W. (1979). *Social Geography of the United States*. London: Longman.

Watson, J. W. (1982). *The United States: Habitation of Hope*. London: Longman.

Watson, J. W. and Sissons, J. B. (eds.) (1964). *The British Isles, a Systematic Geography*. London: Thomas Nelson.

Watson, J. W. and Robinson, J. W. (1978). *The Canadians, How They Live and Work*. Newton Abbot: David & Charles.

Relevant Websites

library.mcmaster.ca
Archive of Watson material at McMaster University, Hamilton, Canada.

Wright, J. K.

I. M. Keighren, University of Edinburgh, Edinburgh, UK

Glossary

Choropleth Map A thematic map in which a sequence of colors, shades, or patterns is applied to areas in order to represent data such as rates, densities, or percentages.

Dasymetric Map A choropleth map in which the shape and position of boundaries are altered to reflect underlying geographical distributions, such as land use or topography, rather than arbitrarily defined enumerative or administrative zones.

Wright, J. K. (1891–1969)

In a six decade research career, John Kirtland Wright made important contributions to academic and pro fessional geography in the US and internationally. His interests ranged from exploration and cartography, through bibliography and historiography, to quantitative techniques and the geographies of knowledge. These enthusiasms were united by a desire to elucidate the influence of human nature upon the conception of ideas. His work on imagination and perception prefigured re search in humanistic and phenomenological geography.

The emergence of Wright's geographical interests paralleled the discipline's establishment in the US. He was born in Cambridge, Massachusetts, on 30 November 1891, the son of a Harvard University professor. There, Wright encountered geographers central to the discip line's development, including Russian anarchist, Pyotr Kropotkin, and geomorphologist, William Morris Davis. Davis encouraged Wright's geographical passions, and gave him a drawing board to practice mapmaking and geomorphological sketching.

Wright honed his cartographical skills mapping an imaginary country, Cravay. The fictional nation was a focus for Wright's geographical imagination, and he spent long periods charting Cravay's topography and history. Beyond this illusory realm, Wright's geographical inter ests were exercised by New England's Mahoosuc Mountains, which he explored during childhood sum mers. The uncharted range was, for Wright, an alluring *terra incognita*, and engendered both sublimity and rev erie. In mountain diaries Wright recorded his emotional reactions and aesthetic impressions during ascents, and noted how these changed according to climate or the influence of recently encountered literature and music. The appeal of the unknown, and the relationship be tween perception and emotional response to place, be came later for Wright important concerns.

As a student at Harvard, Wright studied physical geography, geology, and climatology before turning to history for graduate work. Under the guidance of Charles H. Haskins, Wright researched the plurality and social construction of geographical knowledge in Europe dur ing the Middle Ages. While in Paris for archival work, he attended lectures at the Sorbonne from leading regional geographers, including Emmanuel de Martonne, Albert Demangeon, and Lucien Gallois. Upon his return to the US in 1920, Wright was appointed librarian at the American Geographical Society, where he remained until 1956.

At the Society, Wright – in addition to completing his doctoral thesis, published as *The Geographical Lore of the Time of the Crusades* (1925) – undertook a number of bibliographical Projects. Among these was his Research Catalogue, which grouped the Society's collections ac cording to an innovative topical and regional classifi cation. The catalog formed the basis to an important bibliography, *Aids to Geographical Research* (1923), and was used until 2005 as the classificatory model for the Society's *Current Geographical Publications*.

As a contributor of unsigned reviews and notes to the Society's journal, *Geographical Review*, and editor of its research series, Wright engaged the discipline's entire scope. His most impassioned work related to the history of geography. Wright outlined a perspective which put the conception of knowledge at its core. He called for an appreciation of the history and geography of geo graphical thought – an approach which considered the role of location and environment in shaping knowledge about the world. Wright argued that where claims to knowledge were made mattered to what those claims were. Seemingly redolent of an unfashionable environ mental determinism, Wright's plea was largely ignored. Only later were the questions of perception and sub jectivity with which his papers dealt reassessed by hu manistic geographers including David Lowenthal and Yi Fu Tuan.

During the 1930s and early 1940s, Wright pioneered work in quantitative and statistical geography. He applied mathematical techniques to cartographical analysis, and advanced novel methods of mapping population distri butions. In collaboration with Osborn M. Miller, Wright popularized choropleth and dasymetric mapping of statistical data. Like his earlier work on the history of geography, Wright's quantitative geography had a limited impact upon contemporaneous research. Many of his statistical analytical techniques, particularly those deal ing with the concentration and dispersion of populations, were developed independently by quantitative theorists during the 1960s.

Between 1938 and 1949, Wright served as director of the American Geographical Society. Despite adminis trative burdens, he revisited themes of perception, sub jectivity, and lore. Where his earlier concern had been to describe the geography of knowledge, his later work dealt with the social aspects of knowledge making. In papers on the role of human nature in geography, Wright examined how psychology, behavior, motivation, and bias influence the conception of knowledge. Lore and erro neous ideas were, for Wright, as significant as scientific geographical knowledge; all were illustrative of a sub jective and partial process.

Wright's perspective on geographical perception, im agination, and subjectivity was articulated most fully in his 1946 presidential address to the Association of American Geographers – '*Terrae incognitae*: The place of imagination in geography'. Wright's paper touched upon his enduring passions – landscape perception, the allure of the unknown, the influence of imagination upon the production of geographical knowledge – and proposed a novel subfield of geographical inquiry: geosophy. Wright envisioned geosophy as the multidisciplinary study of geographical knowledge and perception – an attempt to understand the various social, cultural, and psychological factors that influence geographical thought. Again, Wright's conviction elicited limited initial reaction. During the late 1950s and early 1960s, however, his paper, and aspects of his research agenda, informed emerging work on behavioral geography. In retirement, and until his death on 24 March 1969, Wright supported this humanistic trend, and corresponded with its leading proponents.

It is only in retrospect that the significance of much of Wright's work is apparent. His central role in the pro fessional apparatus of geography in the US distanced him from the discipline's academic mainstream. Having never held a faculty position, Wright also lacked intellectual disciples. His attention to the geography of geographical knowledge, to quantitative research and statistical methods, and to perception and imagination in geog raphy can all be seen, however, to have prefigured im portant later work. Wright occupied a broad intellectual hinterland, and, as a function of his idiosyncratic geo graphical passions, made important contributions to the professionalization of geography in the US, to historio graphical work in geography, and to the emergence of humanistic themes.

See also: Geography, History of; Historical Geography; Humanism/Humanistic Geography; Landscape Perception; Lowenthal, D.; Tuan, Y.-F.

Further Reading

Bowden, M. J. (1970). John Kirtland Wright, 1891 1969. *Annals of the Association of American Geographers* 60, 394 403.

Koelsch, W. A. (2002). John Kirtland Wright, 1891 1969. In Armstrong, P. H. & Martin, G. J. (eds.) *Geographers: Biobibliographical Studies*, vol. 22, pp 169 181. London: Continuum.

Lowenthal, D. and Bowden, M. (eds.) (1976). *Geographies of the Mind: Essays in Historical Geography in Honor of John Kirtland Wright*. New York: Oxford University Press.

Wright, J. K. (1952). *Geography in the Making: The American Geographical Society 1851 1951*. New York: American Geographical Society.

Wright, J. K. (1966). *Human Nature in Geography*. Cambridge, MA: Harvard University Press.

Y

Youth/Youth Cultures

L. Holt, University of Reading, Reading, UK

Glossary

Apollonian Discourses A romantic view of childhood, which suggests that children are inherently good and born with natural talents and potentials which need to be nurtured and developed. Provided children are not corrupted by negative influences and bad treatment, they will achieve their immanent potential.

Counter-Cultural The distinctive youth subcultural practices that directly challenge mainstream values and norms.

Dionysian Discourses These suggest that children are born unruly and inherently, naturally bad. The evil natural tendencies of childhood need to be overcome through strict discipline.

Hegemony It expresses the cultural dominance of one set of ideas over others. It tends to be associated with Marxism and class-based cultural politics, with the emphasis on the dominance of middle-class over working-class values. Attributed to Antonio Gramsci.

Subcultures They also emphasize the distinctiveness of youth cultures, focusing upon identifiable subgroups such as Punks or Goths.

'Tribal' Youth Cultures They present youth cultures as a tribe apart from mainstream culture. They emphasize the distinctiveness of youth cultural practices and mores.

Defining Youth – At Risk and Risky

Youth is an embodied social construction attached to young people. It is sociospatially specific, with youth meaning different things in varying times and spaces. Youth is the category ascribed to individuals who are perceived to be neither still children nor yet fully adult. The notion of youth is a relatively new concept historically, and it does not exist in a comparable form in all societies. Studies of youth developed largely with reference to the Global North, and youth is therefore arguably a colonizing construct, although research into young people's experiences is increasingly diverse in its geographical focus.

In many national contexts of the Global North, youth and young people are viewed as a liminal group – on the cusp of adulthood and childhood, difficult to place within either adult or child spaces. Youth is almost by definition 'out of place' and troubling. Young people are often constructed in dominant discourses as simultaneously and paradoxically at risk, and as a risk to society. They are variously defined as either angels or devils – innocents in need of protection or agents of disruption to the social order. The conflicting discourses that cast young people as 'angels' or 'devils' are associated with age, class, race/ethnicity, gender, and other factors. As children age and become cast in the category 'youth', rather than 'child', they are more often categorized under Dionysian (devil like) discourses as problematic, rather than under Apollonian frameworks (angel like) as at risk.

Tribal Youth Subcultures and Resistance

Studies of youth cultures emerged in the 1960s and 1970s. In the UK context, research into youth cultures was a primary concern for the influential Centre for Contemporary Cultural Studies (CCCS) at the University of Birmingham. Indeed, studies of youth cultures inspired the new cultural geographies both directly and indirectly. Early studies of youth gangs in the USA by scholars such as James Duncan and David Ley, along with the work of the CCCS, were key stimuli for the development of new cultural geographies. By contrast to earlier types of cultural geography, 'new' visions, tied to the 'cultural turn', contend that places contain a multiplicity of cultures, which often conflict, rather than being defined by one 'holistic' culture. Emphasis is also placed on individuals as agents of cultural reproduction, as opposed to structuralist superorganic views that conceptualize culture as preexisting individuals and specific social interactions. Geographers emphasize the conflicts and differentiations between mainstream and youth subcultures, thereby critiquing the causal association of an individual place with a particular, identifiable, culture. The centrality of studies of youth to the emergence of social and cultural geography has been somewhat overlooked and/or neglected in recent years.

Early research into youth cultures, conducted in the 1960s and 1970s, arguably unintentionally reproduced the idea that young people are a distinct social group who are troubling to society. These studies were self consciously counterposed to the original studies of youth, which can be traced to post World War II criminology and fears about public disorder. However, early youth culture studies, by taking as their departure point these early studies of troubling youth, inadvertently reproduced aspects of this earliest research. The new cultural studies of youth drew upon ethnographic methods to investigate the experiences of young people rather than objectifying young people from an external standpoint, as did the earlier criminological studies. However, these pioneering studies of youth cultures focused upon countercultural and subcultural practices as a form of resistance against the mainstream, and reproduced the notion of youth cultures as 'tribal', distinct from mainstream cultures. As the name suggests, studies of 'tribal' youth cultures imply that young people's social and cultural practices are a culture apart from the rest of the society. Research in this vein explored a variety of subcultural groups, including Mods, Punks, and violent gangs, and focused upon how their norms, values, and practices differed from the mainstream. Such studies therefore emphasized the distinctiveness of youth cultures. The role of subcultures as working class forms of resistance was an important theme in this research, which often prioritized class based power struggles.

The focus on alternative, rather than mainstream, youth cultures implicitly reproduces the perception of young people as distinct from, other to, and a potent threat to, broader society. Ultimately, such early studies could be viewed as essentialist, emphasizing the cohesiveness of youth subcultures, as opposed to reflecting upon the differentiations within them. It is easy to identify these limitations in retrospect. These critiques are tied to those of much radical scholarship of the 1970s. However, it should not be forgotten that these pioneering studies of youth culture have been influential, both forging new areas of research interest and contributing to the turn toward the cultural and social in the social sciences and geography.

The tradition of research about youth subcultural practices continues within geography and the social sciences more generally. Recent research has included ethnographies of distinct subcultural groups, such as Goths, clubbers, and students. These later studies are more sensitive to the interconnections between subcultures and broader societal social and cultural processes. However, they continue to focus upon the distinctiveness of the subcultural group vis à vis the mainstream. There is a danger of these studies tending toward essentialism by highlighting the difference between subcultures and the mainstream rather than fully reflecting upon the differentiations within these youth cultures. This danger can be overcome by careful consideration of differences 'within' and between subcultural groups.

Studies of 'tribal' youth cultures, which emphasize the distinctiveness of these subcultural groups, are contextualized within the greater attention paid to the powerfulness of cultures since the inception of the new cultural geographies. The idea of countercultures suggests that dominant cultures can be contested, and possibly even transformed. In early studies, power was often expressed via the concepts of hegemony and counter hegemony. Although often tied to, and having its roots within, Marxist inspired understandings of culture as underpinned by the superstructure of the capitalist political economy, notions of hegemony and counter hegemony do not necessitate a Marxian, economically reductionist view of culture. In comparison to contemporary debates about power, however, hegemony and counter hegemony may appear somewhat simplistic and reliant upon an apparent dichotomy between the 'oppressor' and the 'oppressed'. Despite this critique, endeavors move beyond this dualism, which are tied to more complex ideas of power, and pose political problems of how to build resistive alliances. Ultimately, even the most post structuralist of scholars are challenged by this impasse. The emphasis on power relations and the resistive potential of youth subcultural groupings has influenced research on new social movements (NSMs). Although diverse, NSMs are often associated with youth.

The tendency for research to portray youth subcultures as distinct from broader society is arguably associated with the dominance of ethnographic approaches. Indeed, ethnographic research may be more sensitive to the 'life worlds' of the research participants than broader contexts. If it is accepted that 'structures' are reproduced by individuals through, among other things, unreflective everyday sociospatial practices, individuals may not fully perceive the structural constraints and opportunities within which their lives are contextualized. However, examples can be found, of, critical ethnographies of young people that examine how everyday practices reproduce broader social–spatial patterns and processes, with pivotal work by Willis and Katz notable in this tradition. The challenge remains of how to produce ethnographic research that explores the complex relationship between subcultures, and broader cultures and sociospatial relations, without being either structurally reductive or imposing the researchers' *a priori* perceptions of structural constraints.

Critiquing Tribal Youth: Heterogeneous Cultures and Broader Connections

Despite critiquing the association of individual places with identifiable cultures, the early studies cast youth cultures as implicitly holistic and undifferentiated, and aspects of this tendency can be reproduced in contemporary studies. There is a danger that youth cultures are also viewed as inclusive, as a bulwark against the hegemony of mainstream society. Scholars have suggested that some studies of youth cultures have paid limited attention to the differences and power relations within youth subcultures, according to gender, ethnicity, and sexuality, and so on. Feminists critique much research about youth cultures for naturalizing youth as masculine. The gendering of youth as masculine has been explicitly contested by exploring females' experiences. In line with broader changes in feminist scholarship, critical attention has also been increasingly paid to how youth cultures reproduce both female and male gendered identities in particular social settings. These studies move far beyond the uncritical acceptance of youth as gendered male.

Scholarship increasingly explores the multiple lines of differentiation and power within youth cultures, along the lines of gender, race, sexuality, and disability. There is an imbalance in the existing literature here, with a greater level of attention paid to concerns of gender, race, and to a lesser extent, sexuality, and with limited concern with youthful experiences of disability. Although some recent studies have been sensitive to class and socioeconomic differentials, overall, class differences have also been somewhat neglected in recent decades. This relative neglect counterposes earlier research which often focused almost exclusively on class differentiations, and was arguably economically reductive. Contemporary studies of youth cultures could often take more seriously the myriad differentiations of youth, including classed identities. These lessons could be applied even when this is not an explicit object of study, by teasing out more fully how the experience of membership of a youth subcultural group, for instance, is mediated by other axes of power. The work by Linda McDowell on masculine identities in economically restructured cities, and that by Sarah Holloway and Gill Valentine on young people's use of information and communication technology point to how this can be achieved.

The move toward exploring the differentiations of youth has often been associated with a change in terminology, which is reflective of, and connected to, more critical understandings of power, space, and place. Rather than examining 'youth cultures' *per se*, geographers increasingly express the object of their study as the geographies of young people. There are at least two shifts which this changing terminology represents. First, a move toward expressing more fully that young people as individuals and collectives reproduce, and are positioned within, multiple cultures which might conflict with each other. Second, a move away from reifying culture, toward exploring how cultural norms, values, meanings, and interpretations are intimately interconnected with the social, political, and economic, so as to be often indistinguishable. Thus, cultures are always also social, economic, and political. Such a move suggests, and sometimes explicitly draws upon, more complex conceptualization of power relations. Rather than oppositional, power relations are increasingly understood as complex and intersecting. Greater attention has been paid to the ways in which young people's cultures connect with, and can reproduce, dominant societal norms, values, and power relations along the axes of gender/sex, sexuality, and ethnicity/race. It is also more widely appreciated that power is both constructive of subject/agents along with setting limits. A final, and related point, is that power is also increasingly understood to operate subconsciously through interconnected embodied individuals. Thus, youth subcultures might specifically set themselves up as countercultural and to contest the mainstream. However, such subcultures are not distinct, but connected to broader sociospatial, cultural, economic, and political processes, which are often reproduced through subconscious or unreflexive practices. Studies of the reproduction of distinction through the consumption of specific youth cultural commodities are key exemplars here.

The increasing appreciation that youth cultures are both differentiated and connected to broader sociocultural processes has been tied to the development of more nuanced reconceptualizations of space/place. The work of Cindi Katz and Doreen Massey have proved influential in understanding the spaces of youth cultures. Places/spaces are increasingly understood as both specific moments for the expression of agency and connected to (and capable of transforming) broader sociospatial and cultural networks and relationships reproduced at a variety of intersecting (rather than hierarchically nestled) scales. Such complex readings of place help to conceptualize how youth cultures can simultaneously be specific and self consciously resistive, and yet often reproduce more generalized sociospatial processes operating at a variety of spatial scales.

From Studies of Youth Culture to Young People

The move toward more nuanced and complex understandings of youth culture has also been associated with a shift in focus away from studies of youth toward the bourgeoning of studies of geographies of children and young people. Such a focus is stimulated in part by dialog with studies of both youth culture and the 'new social studies of childhood'. The journal *Children's Geographies* has developed as a crucial forum for the publication of research in this area. Despite the common root of studies of children and youth cultures in geography, critical evaluations of youth cultures have often been implicit rather than explicit foci of study within this emerging 'subdiscipline'.

Within geographies of children and youth, there is a paradigmatic orthodoxy that places children and young people as social and cultural agents at the center of research. Although this centrality of agency may trace its lineage to early studies of youth culture, this body of research is generally not framed as youth cultural studies. In addition, partly due to a perceived neglect of the 'middle' years of childhood, a significant proportion of recent studies focus upon the experiences of younger children who arguably do not fall within the category of youth.

Contesting Youth: Academic Perspectives and Societal Shifts

At the same time as geographical studies have extended the timeframe of youth downward into childhood, academic debates and societal shifts have begun to question how youth can be conceptualized as a distinctive category. There are two interrelated societal shifts that have instigated this challenge: first, changing life course trajectories and shifting youth transitions, and second, a perceived reduced distinctiveness of youth cultures.

There is increasing uncertainty about the distinction between the categories of 'childhood', 'youth', and 'adulthood', with the moment of transition between one stage and the next increasingly difficult to identify. (Of course, these transitions have always been contested and fuzzy). Geographers and others have pointed to the challenge posed by increasingly complex biographies and extended/protracted and more convoluted transitions between youth and adulthood. For instance, within national contexts across the globalized world, social reproduction and the move to the labor market has been disrupted. Life courses do not follow the same predictable trajectories in formerly industrialized nations from school and education to a particular type of work, guaranteed for life. Similarly, relationships and kinship networks are perceived to be more fluid and complex. Marriage, leaving home, parenthood, sexual relationships, and economic independence are no longer chronologically predictable rights of passage from youth to adulthood. Increasing uptake of higher education, and reductions in state financial support for students within the Global North, has led to complex transitions between youth and adulthood, along with producing a distinct and geographically significant subgroup of predominantly young students.

Attending a higher education institution is almost a right of passage for many middle class youths. At the same time, this bourgeoning of attendance at higher education, with its associated distinctive lifestyles and cultural preferences, is highly class specific. The move toward the unpredictable risk society, which opens increasing opportunities for some and reduces the predictability of the life course, has by no means eliminated socioeconomic differences and intergenerational inequalities. Rather, the transitions between youth and adulthood are differentiated on socioeconomic grounds.

At the same time as transitions between childhood, youth, and adulthood are becoming more difficult to pinpoint, there is increasing acceptance that youth cultures are becoming less distinctive subcultures and more reflective of the mainstream. Indeed, the shift from exploring countercultural practices toward a greater appreciation of youth cultures as porous and connected to broader societal trends and processes is associated both with changing scholarship traditions and societal shifts in the ways that youth is played out.

Some Potential Future Directions

Despite the many contributions of studies of youth cultures to human geography, there remain areas of study that have been underdeveloped and which commentators have recently argued could be more fully explored. Four potential areas are outlined below. First, although the attention to youth cultures has been pivotal in the development of the distinct area of interest of geographies of children and/or young people, arguably, studies of 'youth cultures' and geographies of children have become somewhat disentangled. The effect of this has been to polarize studies of children and young people and those of older youth. This is problematic as commonalities and continuities, along with divergence in the experience of younger children and those closer to adulthood, could be more fully explored. Second, greater attention could be paid to the commonalities and differences of the experiences of young people across a variety of national contexts, in both the Global North and South. Although a significant body of recent work has begun to address this gap, there is still further to go to address the relative ethnocentricity of existing research. Third, studies of youth could more fully explore how elongated and more complex transitions between childhood, youth, and adulthood are played out, for instance, by reference to emerging student populations and the urban change that this stimulates. Finally, further attention could be paid to continuities across the life course, rather than implicitly suggesting that transitions are points of rupture and departure. Key here is the need to pay attention to both the experiences of children and young people in their own right and reflect upon how these experiences in fluence adult futures.

Further Reading

Aitken, S. C. (2001). *Geographies of Young People*. London: Routledge.

Chatterton, P. and Hollands, R. (2003). *Urban Nightscapes: Youth Cultures, Pleasure Spaces and Corporate Power*. London: Routledge.

Hall, S. and Jefferson, T. (eds.) (2006). *Resistance Through Rituals: Youth and Sub cultures in Post War Britain* (2nd edn.). London: Routledge.

Holloway, S. L. and Valentine, G. (eds.) (2000). *Children's Geographies: Playing, Living, Learning*. London: Routledge.

Holloway, S. L., Valentine, G. and Bingham, N. (2000). Institutionalising technologies: Masculinities, femininities, and the heterosexual economy of the IT classroom. *Environment and Planning A* 32, 617 633.

Hopkins, P. and Pain, R. (2007). Geographies of age: Thinking relationally. *Area* 39, 287 294.

Jeffries, C. and Dyson, J. (eds.) (2008). *Telling Young Lives: Portraits in Political Geography*. Philadelphia, PA: Temple University.

Katz, C. (2004). *Growing up Global: Economic Restructuring and Children's Lives*. Minneapolis, MN: University of Minneapolis Press.

Malbon, B. (1999). *Clubbing: Dancing, Ecstasy and Vitality*. London: Routledge.

McDowell, L. (2003). *Redundant Masculinities*. Oxford: Blackwell.

McRobbie, A. (2007). *Displacement Feminism*. London: Sage.

Nayak, A. (2003). *Race, Place and Globalisation: Youth Cultures in a Changing World*. Oxford: Berg.

Panelli, R., Punch, S. and Robson, E. (eds.) (2007). *Global Perspectives on Rural Childhood and Youth: Young Rural Lives*. London: Routledge.

Ruddick, S. (1996). *Young and Homeless in Hollywood: Mapping Social Identities*. London: Routledge.

Sibley, D. (1995). *Geographies of Exclusion*. London: Routledge.

Skelton, T. and Valentine, G. (eds.) (1998). *Cool Places: Geographies of Youth Cultures*. London: Routledge.

Smith, D. P. and Holt, L. (2007). Studentification and 'apprentice' gentrifiers within Britain's provincial towns and cities: Extending the meaning of gentrification. *Environment and Planning A* 39, 142 161.

Thornton, S. (1996). *Club Cultures: Music, Media and Subcultural Capital*. Middletown, CT: Wesleyan University Press.

Valentine, G. (2003). Boundary crossings: Transitions from childhood to adulthood. *Children's Geographies* 1, 37 52.

Willis, P. (1974). *Learning to Labour: How Working Class Kids get Working Class Jobs*. Farnborough: Saxon House.

INDEX

Notes

Abbreviations

GIS – geographic information system
GPS – global positioning system

Cross-reference terms in italics are general cross-references, or refer to subentry terms within the main entry (the main entry is not repeated to save space). Readers are also advised to refer to the end of each article for additional cross-references – not all of these cross-references have been included in the index cross-references for space reasons.

The index is arranged in set-out style with a maximum of three levels of heading. Major discussion of a subject is indicated by bold page numbers. Page numbers suffixed by T and F refer to Tables and Figures respectively. *vs.* indicates a comparison.

This index is in word-by-word order, whereby spaces within index headings are given value in the alphabetization e.g 'car ownership' precedes 'carbon'; hyphenated words are listed after similar words with spaces e.g. 'African-Americans' comes after 'African National Congress'. Prefixes and terms in parentheses are excluded from initial alphabetization.

Where index subentries and sub-subentries pertaining to a subject have the same page number, they have been listed to indicate the comprehensiveness of the text.

9/11 terrorist attack (US) see World Trade Center, terrorist attack.

A

C

D

E

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y